COMPARING PIEDMONT AND COASTAL PLAIN NEAR-SURFACE SUBSTRUCTURES USING P-WAVE REFRACTION AND SURFACE WAVE DISPERSION

Neha Narayan
Advisor: Dr. Mong-Han Huang
GEOL394
April 28, 2026

DEPARTMENT OF
GEOLOGICAL, ENVIRONMENTAL,
AND PLANETARY SCIENCES

ABSTRACT

The near-surface substructure informs how land can be used. Groundwater pumping, proximity to the coast, subsidence, and other hazards in the substructure may influence how safe land is for use. This study utilized near-surface geophysics to probe the subsurface in regions above and below the Atlantic Seaboard Fall Line to determine their material structures and saturation level. The Fall Line indicates a transition from stronger rock to unconsolidated sediments and a decrease in elevation going from the Piedmont to the Coastal Plain, which indicates a possible contrast in substructure and subsurface water storage. A site in the Piedmont, a University of Maryland (UMD) Agricultural Facility in Clarksville, MD, and a site in the Coastal Plain, the United States Geological Survey's West Point Land Motion Observatory in West Point, VA, were considered. The Clarksville location contains groundwater wells, and the West Point location is known to be subsiding. Active source data were collected and processed using a Transdimensional Hierarchical Bayesian (THB) framework that integrates p-wave refraction and surface wave dispersion to determine the seismic velocity structure and v_p/v_s ratio over depth. In Clarksville, the v_p/v_s ratio was 3.9 for the soil layer, 2.6 and 1.8 for the deeper and more consolidated layers, and 2.1 for the bedrock layer. The groundwater table appeared to be at ~12.5 m, which aligned with groundwater well measurements. In West Point, the v_p/v_s ratio was 5.7 for the first layer, and 2.2 for the remaining depth. The groundwater table was located at ~3 m, which somewhat aligns with the Surficial aquifer at a depth of 5 m. Moreover, the Clarksville survey found bedrock at ~33 m, but the West Point survey was not able to reach the bedrock layer. As there was a thicker layer of unconsolidated, porous sediments, and higher v_p/v_s ratios, it was inferred that the West Point field site holds more subsurface water. The modified THB1D framework serves as a tool that can be applied to any active source seismic survey to constrain the subsurface layers and infer their porosity.

PLAIN LANGUAGE SUMMARY

Hazards in the subsurface, the layers of material underneath the Earth's surface, influences whether and how land can be used. This study used shallow geophysical field methods to determine the different layers of materials in the ground and how much water they can hold. This is especially crucial in regions that experience groundwater pumping or are in close proximity to the coast. The Atlantic Seaboard Fall Line in the eastern United States signifies a difference in elevation, and possible differences in sediment and groundwater structures, going from the Piedmont region to the Coastal Plain region on the easternmost side of the United States. This study considered two locations, one in the Piedmont region and one in the Coastal Plain region to determine the different subsurface layers and their water content. The Piedmont site is in Clarksville, MD, a University of Maryland (UMD) Agricultural Facility that has groundwater wells. The Coastal Plain site is in the United States Geological Survey's West Point Land Motion Observatory in West Point, VA, a region that is known to experience land sinking over time. Active source seismic data can be inverted to determine the seismic velocities

in different layers. These seismic velocities can be compared to known material seismic velocities to infer the layers. Active source data were collected from both field sites and implemented in an iterative algorithm that proposes velocity models until the data and model agree. Velocity models were created for both field sites, and the groundwater depth was estimated for both sites. The West Point field site had shallower groundwater and more loose sediments than Clarksville. These findings indicated that the West Point site is able to hold more groundwater than the Piedmont site. This study also produced code that can be used with any active source data set to determine the subsurface layers and their water content.

TABLE OF CONTENTS

1. Introduction……4
    1.1 Field sites & geology……4
    1.2 Near-surface geophysics……6
    1.3 Inferring from seismic velocities……...6
2. Data Collection……8
3. Analysis Methods……...10
4. Results & Discussion……14
    4.1 Uncertainty……...19
5. Conclusions and Broader Implications……..20
6. Future Work……...21
Acknowledgements……21
Bibliography……..22
Appendix……24

1. INTRODUCTION

Before land is used for construction or development, it is often surveyed for hazards to determine if it is suitable for use. Groundwater pumping, proximity to the coast, subsidence, and perhaps particular subsurface materials can influence how land is used. Groundwater is water underneath the surface that may already exist in the subsurface, or has permeated into the ground over time from precipitation and/or through lateral subsurface flow. It may be used as a source of freshwater through pumping it out from groundwater wells, so there is also a need to model where it is located and ensure that it is extracted safely and effectively. Modeling the different layers in the subsurface and determining the concentration of water in these layers can contribute to land surveying for hazard assessment. This study develops a computational method to constrain the subsurface layers to better model their porosity and compares two field sites in different geological locations on the east coast of the United States.

1.1 Field sites & geology

The two field sites investigated are in the Piedmont and Coastal Plain, which are above and below the Atlantic Seaboard Fall Line (Figure 1). The Fall Line indicates a transition from stronger rock to unconsolidated sediments and a decrease in elevation going from the Piedmont to the Coastal Plain. This change in elevation varies along the Fall Line, is on the order of a couple hundred meters, and is attributed to fault systems in the region. For example, the Brandywine fault system in southern Maryland and the Stafford fault system in Virginia both run nearly parallel to the Fall Line. These systems, among others, interacted with each other to displace the crystalline bedrock and coastal plain strata (Mixon, 1977).

The Piedmont field site is in Clarksville, MD, a University of Maryland agricultural facility that contains groundwater wells. The bedrock in the eastern Piedmont is mostly composed of schist, gneiss, and gabbro (Edwards, 1981). Below the Central Maryland Research and Education Center (CMREC) Clarksville Facility, there are known to be marble, schist, gabbro, granite, and some pegmatite formations. The facility is near the unexposed contact between the Loch Raven Formation and the Cockeysville Marble (Appendix A). Groundwater in the Piedmont originates from precipitation that reaches the groundwater table, a layer where sediments are saturated with water, or fractures in the rock that eventually enter aquifers or groundwater wells (Nutter and Otton, 1969). In the CMREC facility, the groundwater wells are drilled into schist, saprolite (weathered bedrock), and alluvium. Groundwater well depths are periodically measured, and there are six wells near the Clarksville survey location. The survey was conducted by one of the wells, HO Cd 25, which is drilled into schist. From February 2024 to November 2025, the groundwater depth has decreased from 13.7 m to 12.5 m below the surface.

The Coastal Plain field site is in West Point, VA, and is the United States Geological Survey's (USGS) West Point Land Motion Observatory. This site is about 30 miles west of the Chesapeake Bay coast and has deeper bedrock than in the Piedmont. Above the bedrock, there

are unconsolidated sediments such as sand, gravel, silt, and clay (Edwards, 1981). Some groundwater is stored in these sediments as there is more pore space. West Point experiences groundwater pumping, and the region is also known to be subsiding, or slowly sinking over time. In West Point, the Surficial aquifer is at a depth of ~5 m, and the Yorktown-Eastover aquifer is below at a depth of ~50 m (Appendix B). The USGS has deployed an extensometer, which measures land motion, on the observatory (U.S. Geological Survey, Virginia and West Virginia Science Center, 2025). This site is considered in collaboration with Dr. Sarah Stamps from Virginia Tech. Stamps is using vertical land motion observations, measured using Global Navigation Satellite Systems data, to investigate how changes in groundwater and aquifers can influence sea level rise and land subsidence in the Virginia Coastal Plain. Stamps also utilizes near-surface geophysical methods to determine the subsurface layers.

For this study's basis, it was hypothesized that higher v_p/v_s ratios are present in West Point than in Clarksville, indicating unconsolidated material with high porosity in a depositional environment and increased water storage. The null hypothesis states that the v_p/v_s ratio does not vary between West Point and Clarksville, with no difference in water storage above bedrock.

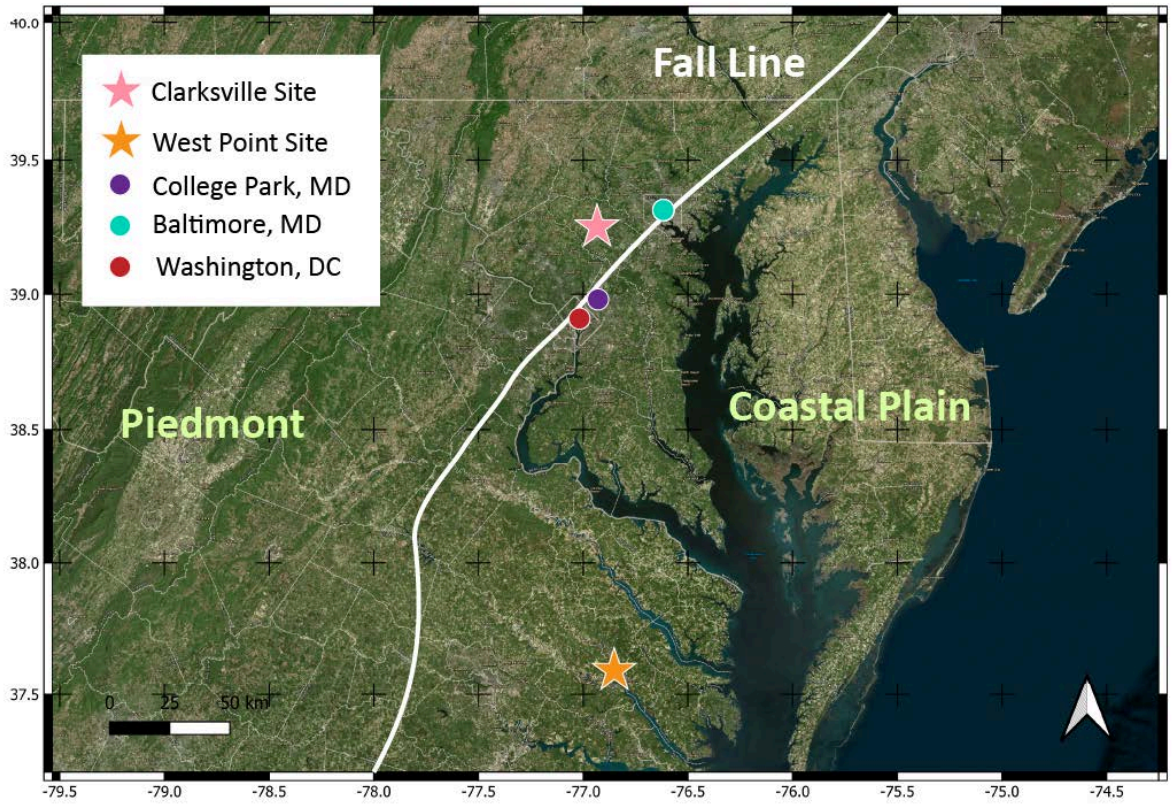


Figure 1: Map showing Clarksville (Piedmont) and West Point (Coastal Plain) field sites. The Atlantic Seaboard Fall Line is roughly drawn in white.

1.2 Near-surface geophysics

This study used near-surface geophysics to characterize the subsurface structure at the two study sites. One of these methods is active source seismology, which involves creating a controlled seismic source, such as a hammer swing on a metal plate, and recording the resulting seismic waves on an array of seismometers, or geophones. Seismic waves propagate as p-waves (primary/pressure), s-waves (secondary/shear), and surface waves that produce the shaking that is felt. P-waves are the fastest, s-waves are slower, and surface waves are the slowest waves. This study used vertical component seismometers, which are typically placed in a line, and a seismic source, or a "shot", is produced at each seismometer location. It is possible that a shot is not strong enough to produce a detectable signal, in which case multiple shots can be taken at the same location and stacked, or added, to produce a more visible signal. Although active source surveys are able to produce controlled seismic waves that can be reliably detected, there are sources of anthropogenic and natural noise in the data. Some examples of noise include pedestrians walking, moving automobiles, and the ocean waves. However, some of this noise can be mitigated during data analysis by applying filters to the waveforms to block out frequencies associated with noise. After active source data is collected, the first p-wave arrival, or the first motion, is manually selected for each receiver for each shot. These "picks" produce a time vs distance, or travel-time, curve. An inversion is performed on the travel-time curve using a forward (predicted) model of the p-wave velocity (v_p) as a function of depth. This forward model is adjusted iteratively to minimize the residuals, resulting in a v_p vs depth profile.

Surface waves are generated when p and s-waves reach the surface. There are two types of surface waves: Rayleigh waves that propagate elliptically and Love waves that propagate laterally. Surface waves can travel at different speeds depending on their frequency. High-frequency waves are able to resolve fine, shallow structures, but are attenuated as they propagate deeper. Low-frequency waves resolve structures with lower spatial resolution, but are able to penetrate deeper into the ground with lower attenuation in comparison to higher frequency waves. Active source data can be interpreted to characterize the dispersion. When a shot is performed, each geophone receives a waveform. A bandpass filter highlights particular frequency ranges, and the cross-correlation is calculated between geophones to estimate time delays, which are used to determine the surface wave phase velocities. The energies corresponding to different possible phase velocities are plotted for each frequency, and a line can be drawn connecting the areas of high energy, creating a dispersion curve. It is possible for there to be multiple phase velocities for a particular frequency, where the slowest phase velocity is the fundamental mode, and the remaining are higher modes, which are seen as additional dispersion curves drawn above the fundamental mode. Dispersion curves can constrain s-wave velocities (v_s). Similar to the p-wave inversion, the dispersion curve is inverted using a forward (predicted) model of the s-wave velocity. A dispersion curve is created for the forward model and compared with the dispersion curve from the data, and the predicted dispersion curve is iteratively adjusted to minimize the residuals. This results in a v_s vs depth profile.

1.3 Inferring from seismic velocities

Both v_p and v_s structures can provide insight into the different layers that exist beneath the surface, and a significant increase of velocity can indicate the presence of bedrock. In previous studies, the bedrock structure has been mapped using similar methods. Uecker et al. (2023) utilized p-wave refraction and performed a two-dimensional inversion to model the subsurface of some valleys, particularly looking for lateral variations in bedrock, in the Wyoming Laramie Mountain Range, mostly composed of granite, with crystalline bedrock lying below. It is worth noting that p-wave arrival times are not the only method to model the substructure, as seen in Eppinger et al. (2024) that also studied the Laramie Mountain Range but used full waveform tomography using first arrival phase velocities and amplitudes. Previous work has also characterized v_p and v_s in regions close to this study, such as the Piedmont and Coastal Plain regions in South Carolina (Odum et al., 2003) and regions experiencing compression similar to the Piedmont (Ma et al., 2021). Additionally, surface wave dispersion has been used to determine v_s profiles, including multimodal dispersion curves from the Virginia Coastal Plain (Guo et al., 2023).

As described previously, travel-time data and surface wave dispersion can be used to develop a v_p and v_s model for the subsurface. One method that uses this data is Bayesian statistics, particularly, Transdimensional Hierarchical Bayesian (THB) frameworks. This method begins with a proposed initial model, perturbs the model, forward calculates the predicted observation, compares it with the actual observation and determines whether to update the model to the proposed model or not. This process typically repeats for a fixed number of iterations. Previous work, such as Huang et al. (2021) and Kolb and Lekić (2014), has used THB frameworks to characterize the critical zone and better interpret seismic data, respectively. This method can be applied to the subsurface model for the p-wave and s-wave velocity structures. Gao et al. (2019) constrained v_s using surface wave dispersion and a computational method that employed a THB framework and receiver functions. This work was also able to account for anisotropy in the subsurface.

After determining the subsurface velocity models, the bedrock depth, groundwater table depth, the porosity can be inferred from changes in velocity. v_s depends on the shear modulus, G, which is impacted by the bedrock and will have a significantly higher velocity than loosely consolidated sediments, and groundwater will be indicated by an increase in the v_p/v_s ratio, since saturated sediments are denser, which increases v_p. In order to determine the depth of the groundwater table, studies such as Grelle and Guadagno (2009) have used active source investigations to compare v_p and v_s with a water seismic index, which was developed from how p and s-waves propagate in unsaturated and saturated media. Seismic wave velocity measurements can be used to determine sediment porosity, the amount of empty space (pores) in the sediments. This has been studied by comparing s-wave velocity with borehole measurements (Zhang et al., 2023) and calculating porosity from seismic velocities (Gu et al., 2020).

2. DATA COLLECTION

On October 24th, 2025, active source data were collected at the UMD CMREC Facility in Clarksville, MD. This was performed with Dr. Mong-Han Huang, post-doctoral researcher Dr. Kat Dapre, and undergraduate student Claire Fagan from the UMD Active Tectonics Lab. The survey line was 200 m long (Figure 2). Starting at 108 m, every 4 m, a vertical component geophone was placed (Figure 3), with a total of 24 geophones. At 68, 88, 100, 108, 120, 132, 144, 156, 168, 180, 192, and 200 m, a 10-pound sledgehammer was swung downwards on a 1' by 1' metal plate with ½" thickness. To increase the signal-to-noise ratio, four shots were taken at each shot location and stacked together. Electrodes for electrical resistivity tomography (ERT) were placed every 2 m as well, but ERT data were not collected.

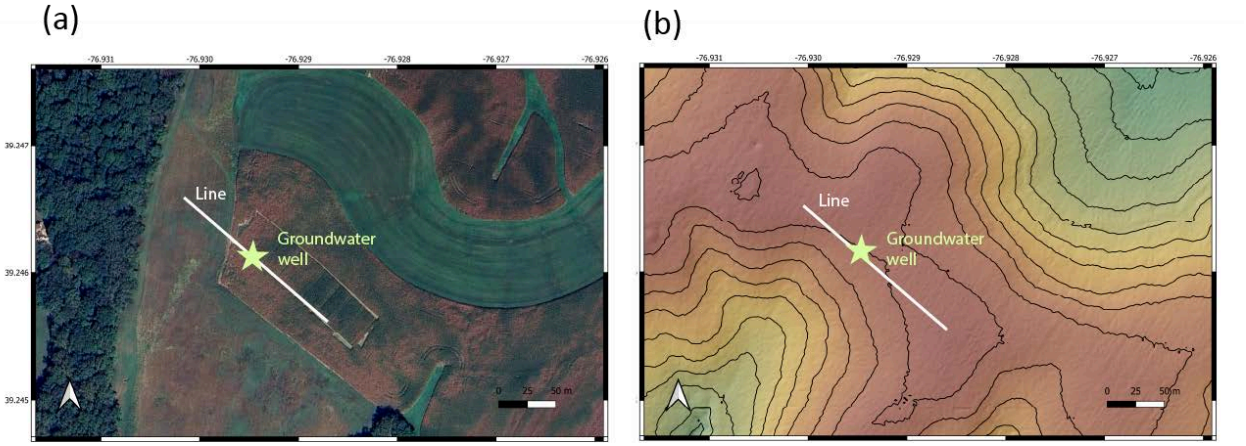


Figure 2: Map of survey location (seismic line) in Clarksville and approximate location of groundwater well shown on **(a)** Google Maps and **(b)** LIDAR image.



Figure 3: Clarksville field site and student performing active source survey. Geophones spaced every 4 m shown to the right of the line, and ERT line with 2 m spacing between electrodes shown to the left of the line. In the top right corner, a groundwater well (brown) can be seen.

On November 1st and 2nd, 2025, active source and ERT data were collected at the USGS West Point Land Motion Observatory. This was performed with Dr. Mong-Han Huang and undergraduate student Maggie Andrews from the UMD Active Tectonics Lab, in collaboration with Dr. Sarah Stamps, Dr. Steve Holbrook, Dr. Sean Bemis, and undergraduate student Jasmine Floyd from Virginia Tech. The first survey line was 400 m long, spanning the long side of the field (Figure 4). Every 1 m, a STRYDE geophone sensitive to vertical motion was placed, for a total of 400 geophones (Figure 5). A 15-pound sledgehammer was swung on a 6” by 6” metal plate with 1” thickness every one meter. At every 10 m, three shots were performed and stacked, and at the remaining receiver locations, only one shot was performed. For the ERT line, every 2 m, an electrode was placed. Initially, this was at the first 128 m of the line, since there were only 64 electrodes total. These 64 electrodes are split into two groups of 32. After the first ERT survey was conducted from 0 to 128 m, the first set of 32 electrodes was moved past the second set of 32, so the survey could be conducted from 64 to 192 m. This was done once more to conduct a survey from 128 to 190 m. The second survey line was 140 m, spanning the horizontal side of the field (Figure 4). Two geophones were placed every 1 m. These geophones are only sensitive to vertical motion, so one was placed vertically, and the other was placed in a holder to orient it horizontally to capture shear motion (Figure 5). One shot was taken every meter, and every 10 m, three shots were taken and stacked together. For the ERT, a single 128 m line was surveyed, with 64 electrodes spaced every 2 m.

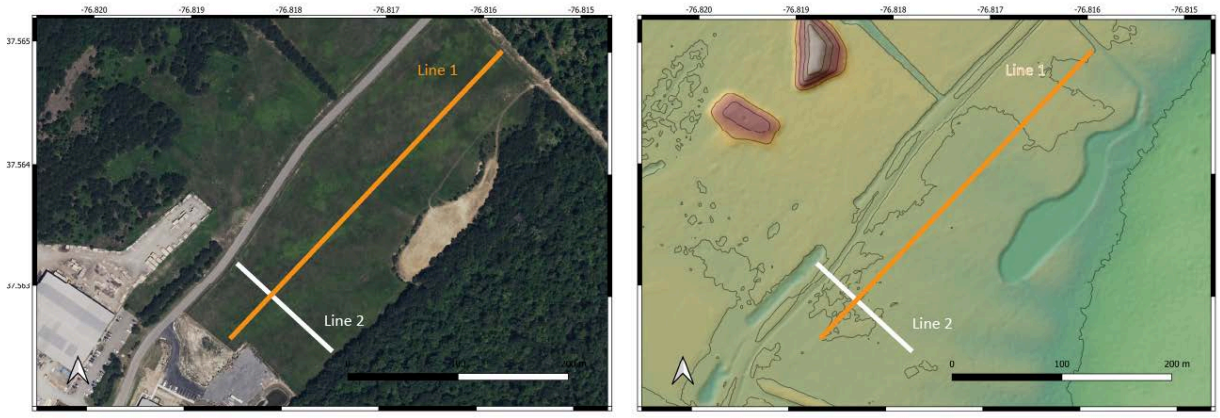


Figure 4: Map of survey location (seismic lines) in West Point shown on (a) Google Maps and (b) LIDAR image.



Figure 5: West Point field site. Orange geophones mark the side of the line used for the active source seismic survey. Geophones spaced every 1 m, and electrodes were placed every 2 m on the other side of the line. Second image shows geophones oriented vertically and placed in holders horizontally.

3. ANALYSIS METHODS

The data from both active source surveys was first uploaded into PICKWIN (Geometrics), a software package that visualizes seismic waveform data. The first motion, or p-wave arrival, was picked for each data file. Examples of p-wave picks for the Clarksville field site are shown in Figure 6 for shots taken at 120 m and 200 m along the line. The Clarksville site did not have significant anthropogenic and natural noise, so filters were not required to make the first arrivals visible. Surface wave dispersion curves were generated for shots taken on the line using a MATLAB script developed by (Huang) The same was done for the West Point data, but p-wave arrivals were picked using a simple plotting and picking script in MATLAB. P-wave arrivals were picked for the first 10 shots from line 2 from the survey (140 m), and dispersion curves were generated for these as well. Similarly, there were no significant anthropogeneric or natural sources of noise during the survey, so filters were not necessary to visualize the p-wave arrival.

The Bayesian analysis was developed from the THB1D MATLAB script (Huang et al., 2021) to produce the seismic velocity profiles. It proposes both p-wave and s-wave velocity models of the substructure. Then, it adds a layer, deletes a layer, changes the p-wave velocity, changes the s-wave velocity, adds noise to the travel time curve, or adds noise to the dispersion curve. It then produces a travel time curve and surface dispersion curve, and compares these curves to those of the data. If the model meets the specified conditions, it keeps the model, updates the proposed model, and begins the process again. If the model does not meet the

specified conditions, it rejects the model, returns to the previously proposed model, and begins the process again (Figure 7).

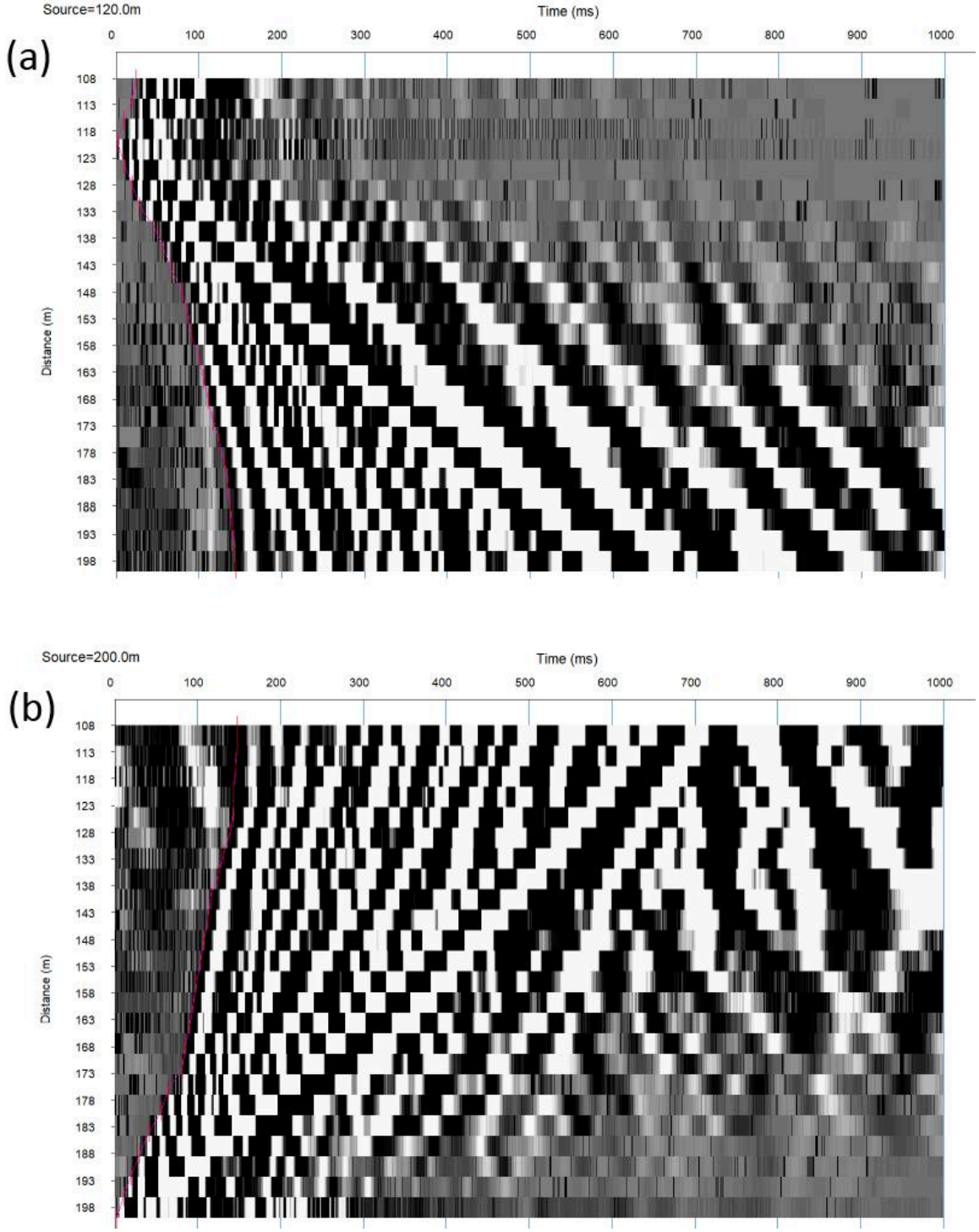


Figure 6: P-wave picks at Clarksville for a shot at **(a)** 120 m and **(b)** 200 m along the survey line. The survey site did not have significant noise, so no filters were applied to pick the first arrivals.

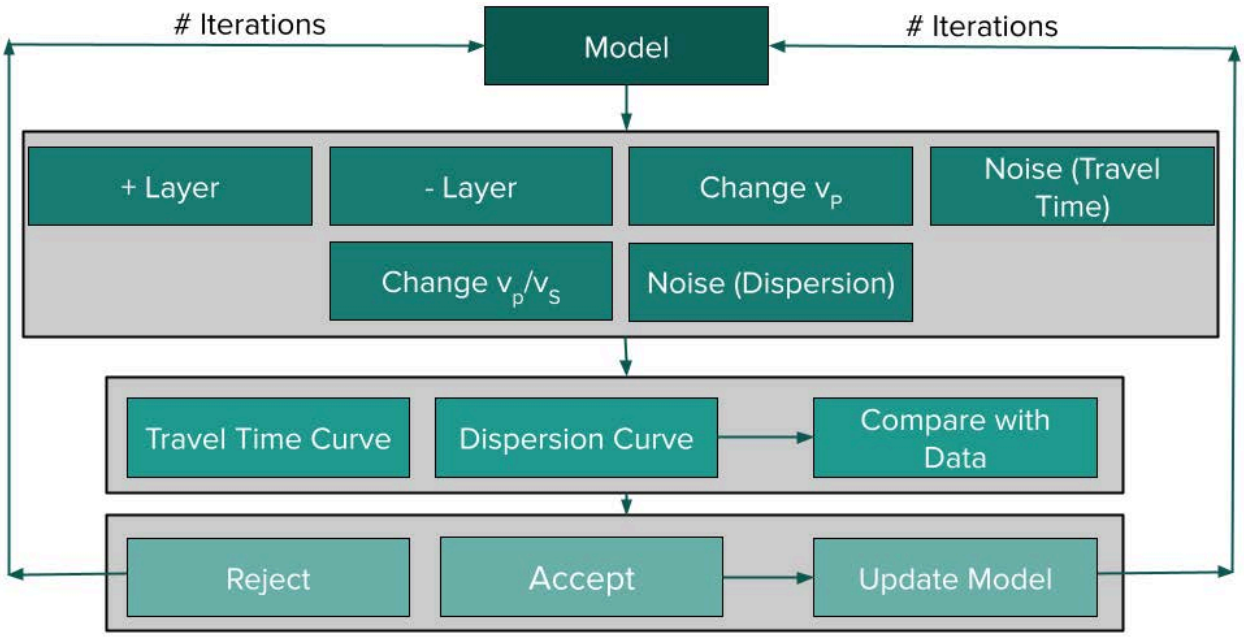


Figure 7: Flowchart showing THB1D functionality with dispersion and v_s included.

In order to constrain v_s using observed surface wave dispersion, a MATLAB script was written to forward compute the dispersion curve. The script first calls Geopsy, a package that helps interpret seismic data. It takes in an input model consisting of a predicted number of layers, v_p, v_p/v_s, and layer density. It is also possible to specify the number of modes (one for only the fundamental mode, two for the fundamental and first mode, and so on). In this study, only the fundamental mode is used for the inversion, since the higher modes were not clearly visible. It then computes dispersion curve points for the proposed model, frequency and phase velocity. The MATLAB script reads this output file, and plots the phase velocity over frequency. Figure 8 shows the logic used to constrain v_s. THB1D was run with the Clarksville data to produce a preliminary v_p model (Figure 8a). The 1D structure was estimated into finite layers, and a v_p/v_s ratio of 1.8 was assumed to produce a v_s profile (Figure 8b). Then, these estimated velocity profiles were input into the forward prediction script to produce dispersion curves (Figure 9). The predicted dispersion curves generally aligned with the data, but with the implementation in THB1D, this misfit would be minimized.

Then, THB1D was edited to include the v_p/v_s ratio as a prior. The script changes the v_p/v_s ratio and predicts the misfit between the theoretical and observed dispersion curve. In THB1D, the function "UpdateMod" performs the change to the model after the component of the model to be changed is selected (Appendix C). v_p and v_s are modified separately, but are constrained such that the v_p/v_s ratio does not go outside of the range established in the priors. In order to make the comparison between the theory and data for dispersion, a MATLAB script that computes dispersion curves was used to generate dispersion curves for each of the shots taken in the field

sites. There were 13 shots used for Clarksville, and 10 for West Point. This can be performed for any field site, and the resulting data file can be uploaded into the THB1D script. Assuming the site subsurface is expected to be fairly uniform, only a few shots (~10) are needed to produce reasonable results. The dispersion plot has an amplitude ranging from zero to one for different frequency and phase velocity pairs. The theoretical dispersion curve points produced by Geopsy are set to have a value of one. Thus, the misfit between the theory and data is the amplitude from the data subtracted by one. In summary, v_p is constrained by minimizing the misfit between the traveltime curve from the data and model, and v_s is constrained by minimizing the misfit between the dispersion curve from the data and mode.

A similar Bayesian inversion in MATLAB (Huang) was adapted to estimate the porosity and saturation of the layers. This script initially varied the porosity, saturation, and mineral concentration, but was edited to only vary the porosity and saturation (Appendix D). The script takes the v_p and v_s values at given depths and calculates the porosity and saturation values at different depths based on the known elastic moduli (bulk modulus K and shear modulus G) of two minerals that are expected to be in the layers. The v_p and v_s profiles produced by THB1D for Clarksville and West Point were imported into this script. The approximate concentration of the two minerals in the rock are fixed. Since the groundwater well along the Clarksville seismic line was dug into schist, 0.3 quartz and 0.2 biotite mica concentrations were used. For West Point, the bedrock was expected to be much deeper, so minerals commonly found in coastal plain sediments, quartz and feldspar, were used at 0.3 and 0.2 concentration respectively. However, this script assumes that the material is solid rock, such as bedrock. Thus, calculations done for shallower layers like soil or weathered rock may be less accurate.

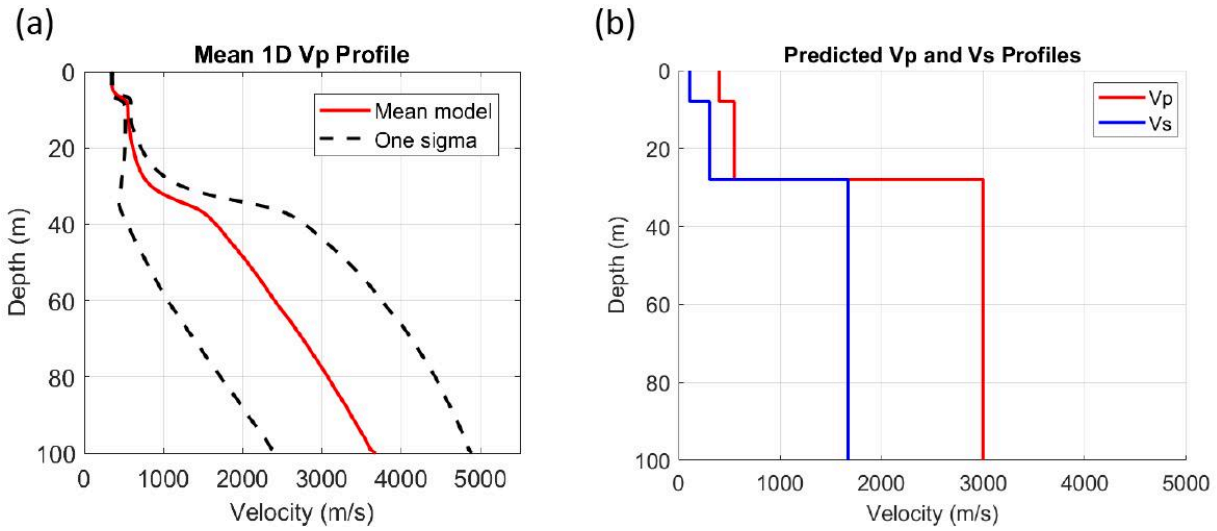


Figure 8: (a) Preliminary 1D v_p velocity model for Clarksville from THB1D. **(b)** 1D v_p and v_s vs depth with layer number, layer thickness, and v_p roughly estimated from (a) and assuming v_p/v_s = 1.8.

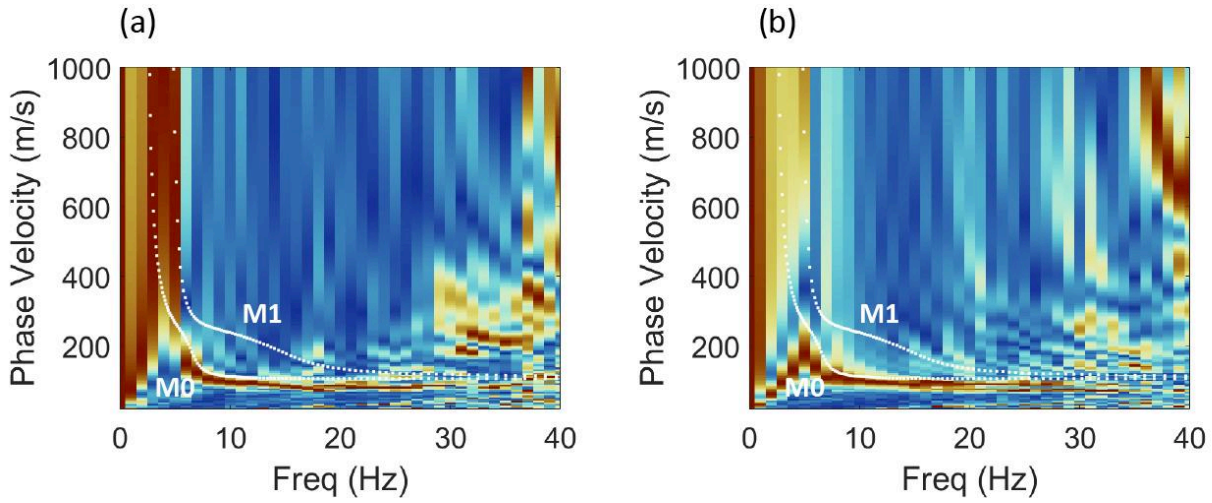


Figure 9: Fundamental (M0) and first (M1) modes of predicted dispersion curve developed from a predicted model of subsurface (Figure 8) plotted against surface wave dispersion plot from Clarksville for **(a)** shot at 120 m and **(b)** shot at 200 m.

4. RESULTS & DISCUSSION

Both inversions were run for 50,000 iterations with the same priors (Appendix C). For each field site, the v_p and v_s profiles were plotted along depth, in addition to the v_p/v_s ratio. During the inversion process, every 10th model was saved, for a total of 5,000 saved models. Each of these saved models was plotted to show the velocity evolution for both v_p and v_s. The final travel time curve fits and dispersion curve fits were also plotted. The final travel time curve plots show the time and offset points and the final fit. The dispersion curve plot shows the average of all dispersion images for every active source shot with the final dispersion curve fit and their standard deviation.

For Clarksville, four layers were clearly present in the final model (Figure 10)**.** Table 1 shows the thickness, v_p, v_s, and v_p/v_s of these layers. Based on the low v_p (<400 m/s), the first layer is likely the soil layer. The v_p of the second layer was 507 m/s with a v_p/v_s ratio close to 2.6 m/s, indicating unconsolidated sediment deposits below the soil layer. The next layer with a v_p of 676 m/s is a larger bed of slightly more consolidated sediments. The fourth layer is likely the bedrock, with a depth ~33 m. The v_p/v_s ratio was higher in the first layer than the bottom two layers, which remains around 2. The fit of the velocity models is better understood in Figures 11 and 12**.** The velocity evolution in Figure 11 indicates that the layers were constrained fairly quickly, with some variation in the velocity of the final layer. The travel time fit and dispersion curve fit in Figure 12 also indicate that the layers were constrained. The misfit in the travel time across the travel time curve was ~0.01 s, and the misfit in the phase velocity in the dispersion curve was ~110 m/s below 7 Hz and 25 m/s for the remaining curve. The porosity and saturation were also plotted over the depth from the rock physics model. As mentioned previously, the rock physics model assumes the layer is bedrock and cannot properly model soil and saprolite. Figure

10c shows porosities between 80-100% for the first 15 m. These values are higher than the expected porosities for soil, but the first 15 m likely consists of soil, and saprolite, which the rock physics model cannot constrain accurately. The groundwater well along the survey line, HO Cd 25, had a measured water depth of ~12.4 m on the day of the active source survey. Around this depth, there is a slight increase in both v_p and v_s and a decrease in the v_p/v_s ratio, which may indicate the groundwater table. It is expected that the v_p/v_s ratio would increase with higher water content, but this model shows a slight decrease in the ratio, which may be due to the slight decrease in porosity seen at this depth. At bedrock depth (~33 m) the porosity begins to decrease, and the saturation begins to increase. It is worth noting that from 35 to 50 m, the velocity models may not be as well constrained since the seismic line was short, less geophones were used, and the spacing between the geophones was large.

For West Point, there appeared to be a ~1 m thick soil layer with varying velocities. However, there were two clearly present layers (Figure 13). Table 2 shows the thickness, v_p, v_s, and v_p/v_s of these layers. The first of these layers was a thin 3 m layer with a fairly high v_p/v_s ratio of 5.7, indicating a significantly saturated soil layer. The bottom layer that continues beyond the depth of the model still had a fairly low v_p of 1430 m/s, which indicates that the survey did not reach the bedrock layer. The velocity evolution in Figure 14 shows that the layers were quickly constrained, and the travel time curve and dispersion curve in Figure 15 visually agreed with the data. The resolution of the dispersion image is not as high as the Clarksville data, so the code was not able to fit the dispersion curve for frequencies below 10 Hz. However, there is a good fit for the remaining dispersion curve. The misfit in the travel time across the travel time curve was ~0.05 s, and the misfit in the phase velocity in the dispersion curve was ~250 m/s below 10 Hz and 25 m/s for the remaining curve. While the rock physics model is not entirely accurate for unconsolidated sediments, there is an increase in both porosity and saturation, which is expected for the location. The Surficial aquifer (shown in Appendix B) is at a depth of ~5 m, and that may be seen in this model at ~3 m. The discrepancy in depth is likely attributed to how the p-wave arrivals were picked, as seen in Figure 14, where the picks at the beginning of the travel time curve are quite scattered.

These findings suggest differences in subsurface water storage between the two field sites. The Clarksville survey was able to reach bedrock, but the West Point survey was not able to, indicating that the bedrock is likely much deeper below the unconsolidated coastal plain sediments. At the depths of 10-33 m, v_p is much lower in Clarksville than it is in West Point, and the v_p/v_s ratio is always higher in West Point. These indicate that the West Point field site is able to hold more groundwater.

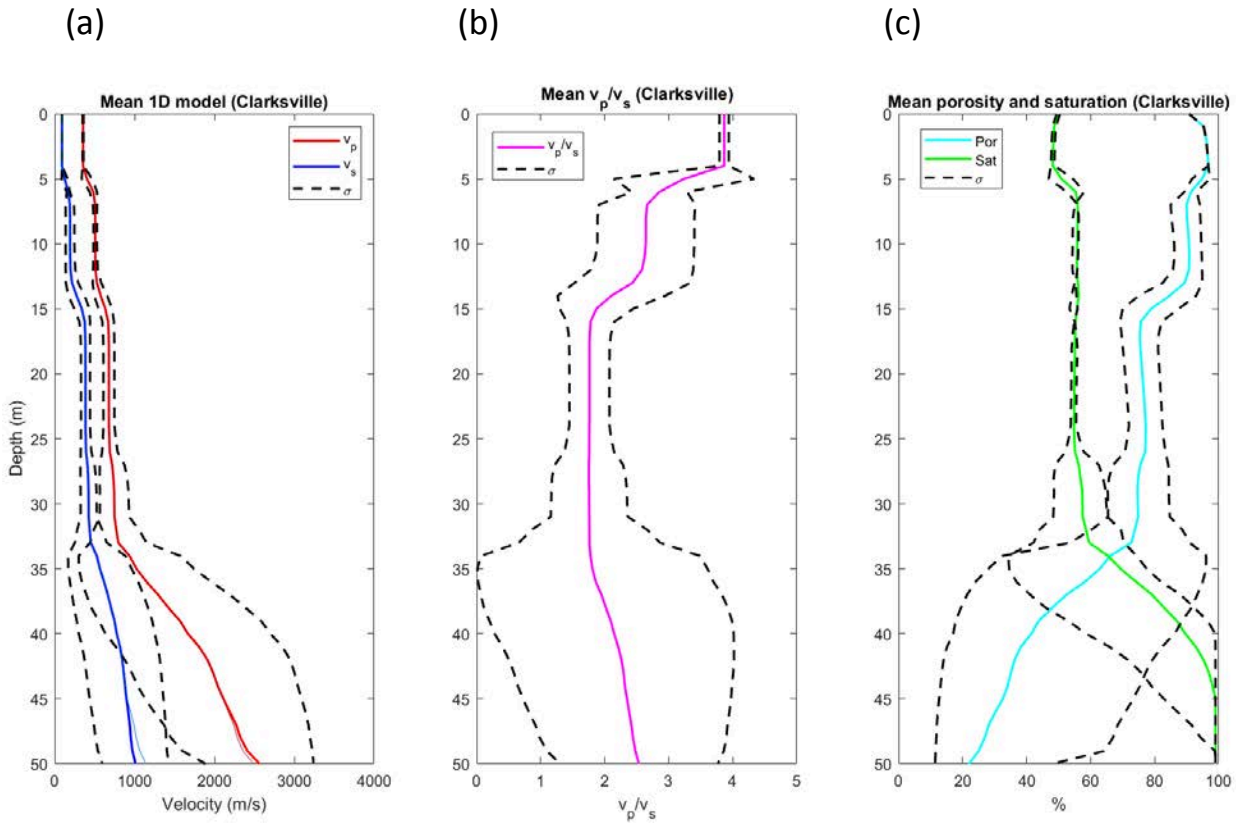


Figure 10: Clarksville 1D models. **(a)** v_p and v_s vs depth, **(b)** v_p/v_s vs depth, **(c)** porosity and saturation vs depth.

Table 1: Clarksville layers, including depth, v_p, v_s, and v_p/v_s estimated from Figure 10. Final layer (bedrock) has no thickness, as it is expected to continue.

Layer	Thickness (m)	v_p (m/s)	v_s (m/s)	v_p/v_s
1	5	355	91	3.9
2	8	507	198	2.6
3	18	676	385	1.8
4	N/A	1913	846	2.1

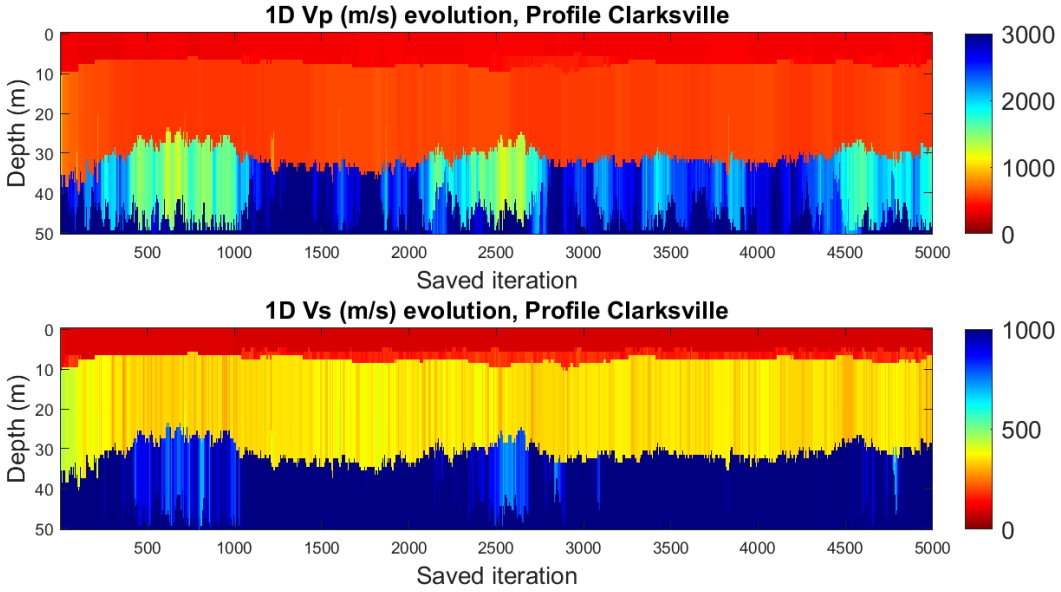


Figure 11: Clarksville velocity model evolution for v_p and v_s.

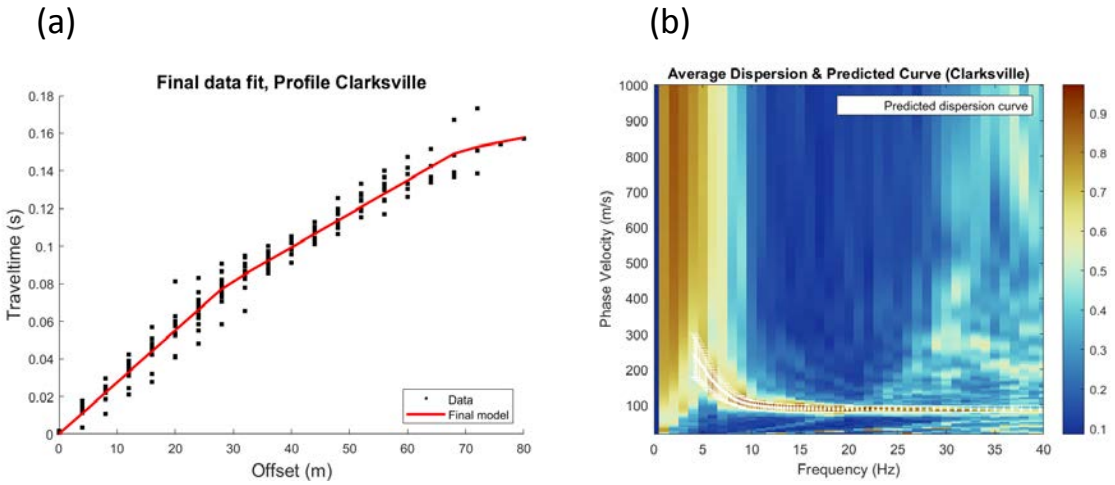


Figure 12: Clarksville **(a)** travel time and offset points plotted with final travel time curve fit and **(b)** average dispersion image plotted with final dispersion curve fit.

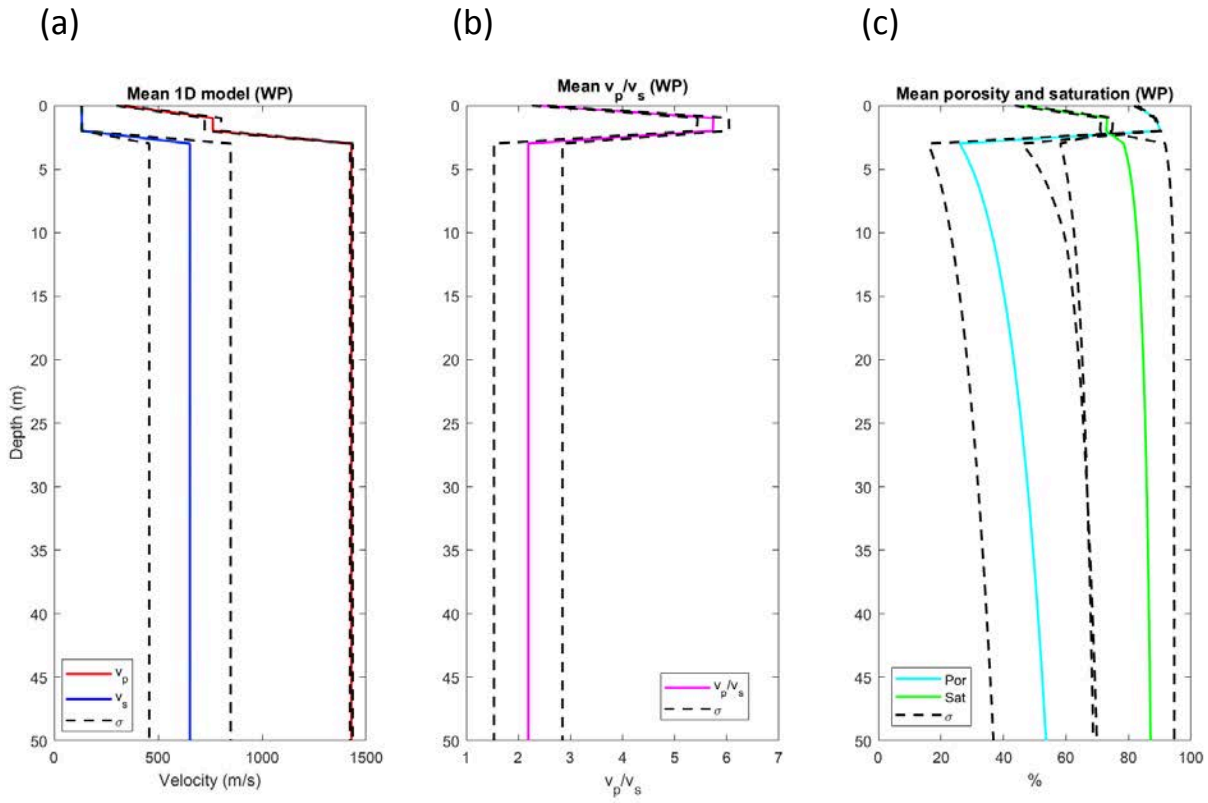


Figure 13: West Point 1D models. **(a)** v_p and v_s vs depth, **(b)** v_p/v_s vs depth, **(c)** porosity and saturation vs depth.

Table 2: West Point layers, including thickness, v_p, v_s, and v_p/v_s estimated from Figure 13. Final layer has no thickness, as it is expected to continue.

Layer	Thickness (m)	v_p (m/s)	v_s (m/s)	v_p/v_s
1	3	764	133	5.7
2	N/A	1430	653	2.2

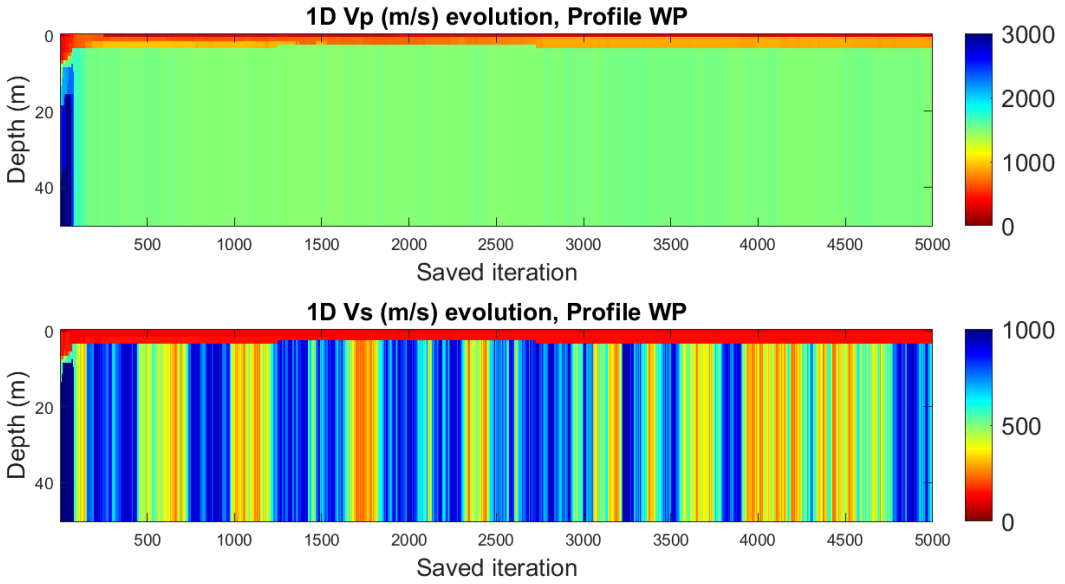


Figure 14: West Point velocity model evolution for v_p and v_s.

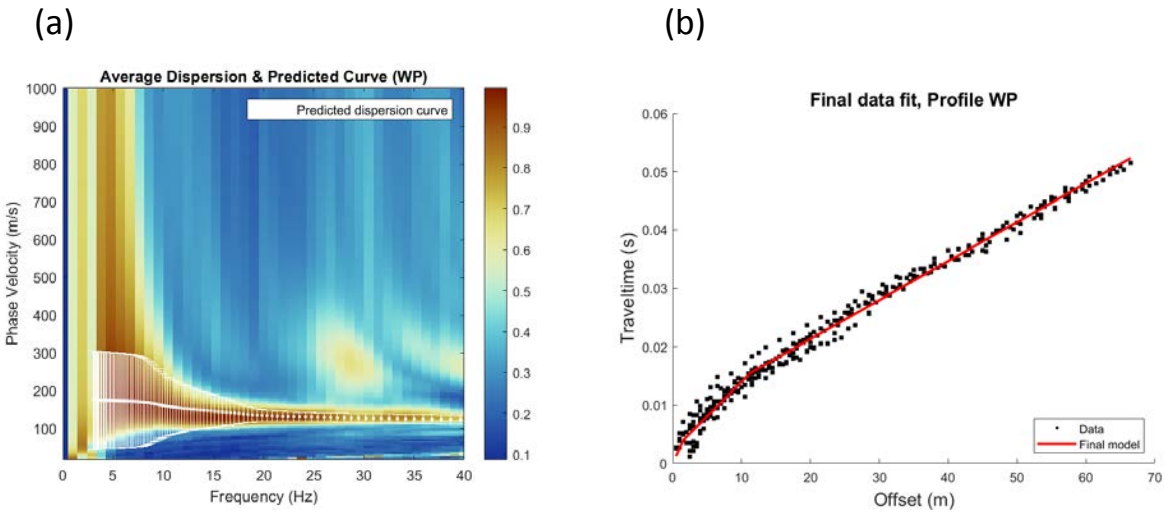


Figure 15: West Point **(a)** travel time and offset points plotted with final travel time curve fit and **(b)** average dispersion image plotted with final dispersion curve fit.

4.1 Uncertainty

THB1D only begins calculating statistics after a certain number of iterations, which is initialized at the beginning of the inversion. In these inversions, the script began performing statistics after 2,500 runs. This prevented statistics from being computed using velocity models with a lot of variability that were not well constrained. THB1D calculates the standard deviation of the velocity values, which were shown in the plots in Figures 10 and 13. For the uncertainty in the v_p/v_s ratio, the error propagation formula for division was used, where σ_{vp} and σ_{vs} were the standard deviations in v_p and v_s:

$$\sigma_{v_p/v_s} = \frac{\overline{v_p}}{\overline{v_s}} \sqrt{(\frac{\sigma_{v_p}}{\overline{v_p}})^2 + (\frac{\sigma_{v_s}}{\overline{v_s}})^2}$$

The porosity and saturation inversion also computes the standard deviation of the porosity and saturation values. However, since only the porosity and saturation values were adjusted iteratively, there was not much variation in the final porosity and saturation values, which led to very low standard deviations. To better capture variability in the calculations, the porosity and saturation inversion was run using the velocity model plus and minus the calculated standard deviation from the THB1D inversion. These porosity and saturation profiles were used as the upper and lower uncertainty bounds. Additionally, the values were calculated by assuming the underlying rock is mostly composed of two minerals, and the model assumes solid bedrock, not soil, saprolite, and other unconsolidated sediments. This model provides a general guide of porosity and saturation over depth, but the values themselves are not entirely accurate.

5. CONCLUSIONS & BROADER IMPLICATIONS

This study compared the subsurface of two field sites, one in the Piedmont, and one in the Coastal Plain region on the eastern United States coast using a Bayesian inversion. It was expected that the Piedmont field site would have lower v_p/v_s ratios than the West Point field site. The inversion showed that the v_p/v_s ratio in Clarksville was 3.9, 2.6, 1.8, and 2.1 for the four constrained layers, from shallowest to deepest. The groundwater level was at a depth of ~12.5 m below the surface according to the measured groundwater well depth, which appeared to be present in the velocity model from an increase in v_p and a decrease in the v_p/v_s ratio at that depth. In West Point, the v_p/v_s ratio was 5.7 and 2.2 for the two constrained layers. The groundwater table level was expected to be ~5 m below the surface, but was seen in the velocity model ~3 m. These results indicate that the subsurface layers in West Point have a higher water content than in Clarksville. Furthermore, the West Point survey was unable to reach the bedrock, whereas the Clarksville survey found the bedrock at a depth of ~33 m. Both velocity profiles were shown to 50 m in depth, but the West Point profile does not show any bedrock, indicating that the layers are primarily soil and unconsolidated coastal plain sediments. These results indicate that the Coastal Plain field site is able to store more groundwater.

Seismic surveys are not as intrusive as drilling and are able to penetrate deeper into the subsurface. Regions with groundwater storage or along the coast may be susceptible to land subsidence or hazards related to sea level rise, and modeling subsurface layers and their water content can aid in site characterization efforts. Using both v_p and v_s to constrain subsurface layers is particularly valuable to determine the presence of water in subsurface layers. This study produced a tool that uses seismic data to determine the subsurface layers and their water content in one dimension. Only p-wave picks are necessary to use the code, so it can be applied to any active source dataset. This inversion was also supplemented with preliminary porosity and saturation calculations from another Bayesian inversion to better characterize the water content

in the subsurface layers, which only requires a velocity model with v_p and v_s. Ultimately, these scripts were able to constrain subsurface layers and the groundwater depth of two field sites with differing groundwater storage, and they could be applied to other field sites.

6. FUTURE WORK

The West Point survey had two lines, one much longer than the other (400 m). The results in this study were determined using the shorter line. A longer line may be able to constrain deeper into the subsurface, so future work could involve running the analysis on this line. It would also be beneficial to use THB1D on different field sites between the Piedmont and Coastal Plain, since the subsurface does vary even within the regions. It would be particularly interesting to perform seismic surveys in a line running roughly perpendicular to the Fall Line to see a change in bedrock depth and subsurface water storage along the transition. Furthermore, THB1D is only able to determine velocity structures in one dimension and cannot account for changes in topography. THB2D (Huang et al., 2021) constrains v_p using a similar Bayesian framework in two dimensions. Future work could involve implementing fitting to dispersion into THB2D. This would be particularly useful in regions without flat topography, or where lateral variations in water storage are expected. One possible application of this is another survey location at the Clarksville field site. In 2024, an active source seismic survey was conducted nearly perpendicular to the survey in this study. The seismic line ran along multiple groundwater wells. Here, it would be worthwhile to see how the v_p/v_s ratio varies along the cross section. Finally, the rock physics model used to determine the porosity and saturation assumes the material is bedrock, but equations to model soil and weathered bedrock would help constrain porosity and saturation for shallower layers.

ACKNOWLEDGEMENTS

Thank you to the Clarksville field team (Dr. Mong-Han Huang, Dr. Kat Dapre, Claire Fagan) and the West Point field team (Dr. Sarah Stamps (VT), Dr. Steve Holbrook (VT), Dr. Sean Beamis (VT), Jasmine Floyd (VT), Dr. Mong-Han Huang, Maggie Andrews).

Thank you to Phil Piccoli for supporting us throughout the program, fellow senior thesis students for the camaraderie, and the Senior Thesis Grant for funding my trip to the Seismological Society of America Annual Meeting.

Finally, a huge thank you to Mong-Han for being an incredible and inspiring mentor.

UMD HONOR PLEDGE

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination.
Neha Narayan, 4/28/2026

BIBLIOGRAPHY

Campbell, E.M., 2000, Geologic map of Virginia, Historic Rivers Chapter of the Virginia Master Naturalists, https://historicrivers.org/docs-links/geology/geologic_map_va.pdf. (accessed Dec 2025)

Edwards, J., 1981, A brief history of the geology of Maryland, Maryland Geological Survey, http://www.mgs.md.gov/geology/geology_of_maryland.html. (accessed Nov 25)

Eppinger, B.J., Holbrook, W.S., Liu, Z., Flinchum, B.A., and Tromp, J., 2024, 2D near-surface full-waveform tomography reveals bedrock controls on critical zone architecture: Earth and Space Science, v. 11, p. e2023EA003248, https://doi.org/10.1029/2023EA003248.

Gao, G., Cunningham, E., Lekić, V., 2019, Spurious low velocity zones in joint inversions of surface waves and receiver functions, Geophysical Journal International, v. 219, p. 1032-1042, https://doi.org/10.1093/gji/ggz345.

Geometrics, SeisImager/2D, https://www.geometrics.com/software/.

Grelle, G., and Guadagno, F.M., 2009, Seismic refraction methodology for groundwater level determination: "Water seismic index," Journal of Applied Geophysics, v. 68, p. 301-320, https://doi.org/10.1016/j.jappgeo.2009.02.001.

Gu, X., Mavko, G., Ma, L., Oakley, D., Accardo, N., Carr, B.J., Nyblade, A.A., Brantley, S.L., 2020, Seismic refraction tracks porosity generation and possible CO2 production at depth under a headwater catchment, Proceedings of the National Academy of Sciences of the United States of America, v. 117, p. 18991-18997, https://doi.org/10.1073/pnas.2003451117.

Guo, W., Li, Z., Zhao, S., and Chen, X., 2023, Characterizing the shallow structure with the multimodal dispersion curves and the body wave refraction travel times from deep seismic sounding data, Frontiers in Earth Science, v. 10, https://doi.org/10.3389/feart.2022.1025935.

Howard County GIS, 2004, Geologic formations, https://data.howardcountymd.gov/mapgallery/misc/Geology22x34countywide.pdf. (accessed Oct 2025)

Huang, M.-H., Hudson-Rasmussen, B., Burdick, S., Lekić, V., Nelson, M.D., Fauria, K.E., and Schmerr, N., 2021, Bayesian seismic refraction inversion for critical zone science and near-surface applications: Geochemistry, Geophysics, Geosystems, v. 22, p. e2020GC009172, https://doi.org/10.1029/2020GC009172.

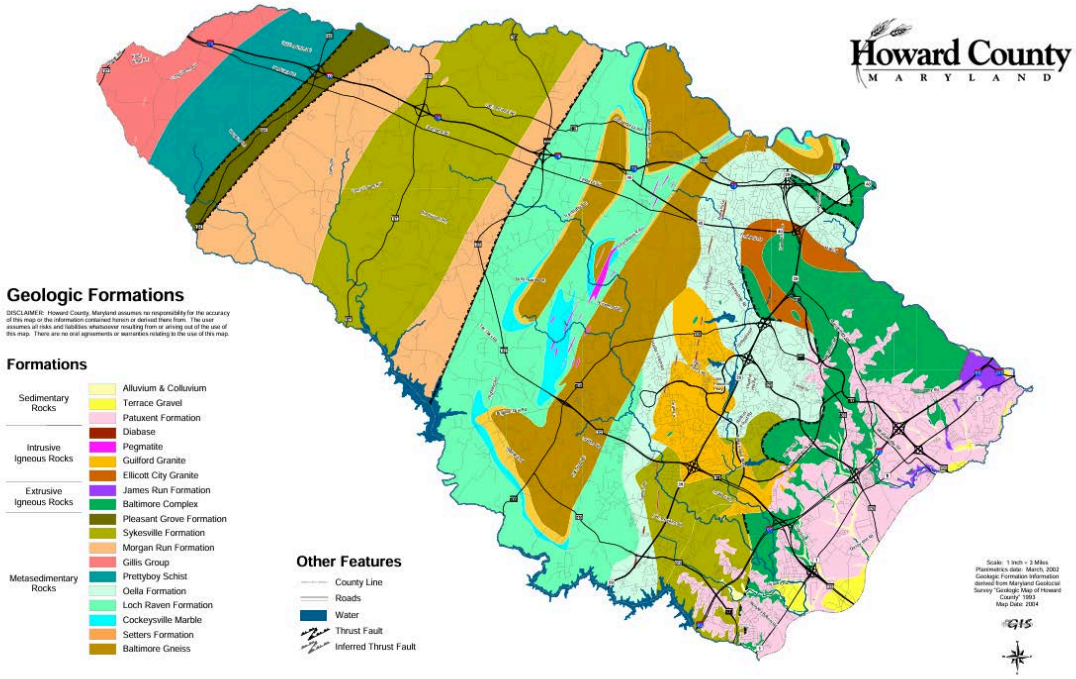
Kolb, J.M., and Lekić, V., 2014, Receiver function deconvolution using transdimensional hierarchical Bayesian inference: Geophysical Journal International, v. 197, p. 1719-1735, https://doi.org/10.1093/gji/ggu079.

Ma, L., Oakley, D., Nyblade, A., Moon, S., Accardo, N., Wang, W., Gu, X., Brubaker, K., Mount, G.J., Forsythe, B., Carr, B.J., and Brantley, S.L., 2021, Seismic imaging of a shale landscape under compression shows limited influence of topography-induced

fracturing, Geophysical Research Letters, v. 48, p. e2021GL093372, https://doi.org/10.1029/2021GL093372.
McFarland, E.R. and Bruce T.S., 2006, Hydrogeologic sections DD-DD', ED-ED', FD-FD' in the Virginia Coastal Plain, https://pubs.usgs.gov/pp/2006/1731/PDF_plates/PP1731plate4.pdf (accessed Apr 2026).
Mixon, R.B. and Newell, W.L., 1977, Stafford fault system: Structures documenting Cretaceous and Tertiary deformation along the Fall line in northeastern Virginia: Geology, v. 5, p. 437-440. https://doi.org/10.1130/0091-7613(1977)5%3C437:SFSSDC%3E2.0.CO;2
Nutter, L.J., and Otton, E.G., 1969, Ground-water occurrence in the Maryland Piedmont: Maryland Geological Survey Report of Investigations No. 10, p. 56, http://www.mgs.md.gov/publications/report_pages/RI_10.html.
Odum, J.K., Williams, R.A., Stephenson W.J., and Worley, D. M., 2003, Near-surface S-wave and P-wave seismic velocities of primary geological formations on the Piedmont and Atlantic Coastal Plain of South Carolina, USA, U.S. Geological Survey, https://doi.org/10.3133/ofr0343.
Uecker, R.K., Flinchum, B.A., Holbrook, W.S., and Carr, B.J., 2023, Mapping bedrock topography: a seismic refraction survey and landscape analysis in the Laramie Range, Wyoming. Frontiers in Water, v. 5, p. 1057725, https://doi.org/10.3389/frwa.2023.1057725.
U.S. Geological Survey Virginia and West Virginia Science Center, 2025, In West Point, Virginia, a new "Subsidence Superstation" measures changing land surface: U.S. Geological Survey News Release, https://www.usgs.gov/centers/virginia-and-west-virginia-water-science-center/news/west-point-virginia-a-new-subsidence (accessed Nov 2025)
Zhang, Y., Song, S., Yao, Y., Cai, J., Zhang, S., and Zhang, C., 2023, Porosity and permeability estimation using seismic wave velocity along Yangtze River embankment: Journal of Applied Geophysics, v. 213, p. 105039, https://doi.org/10.1016/j.jappgeo.2023.105039.

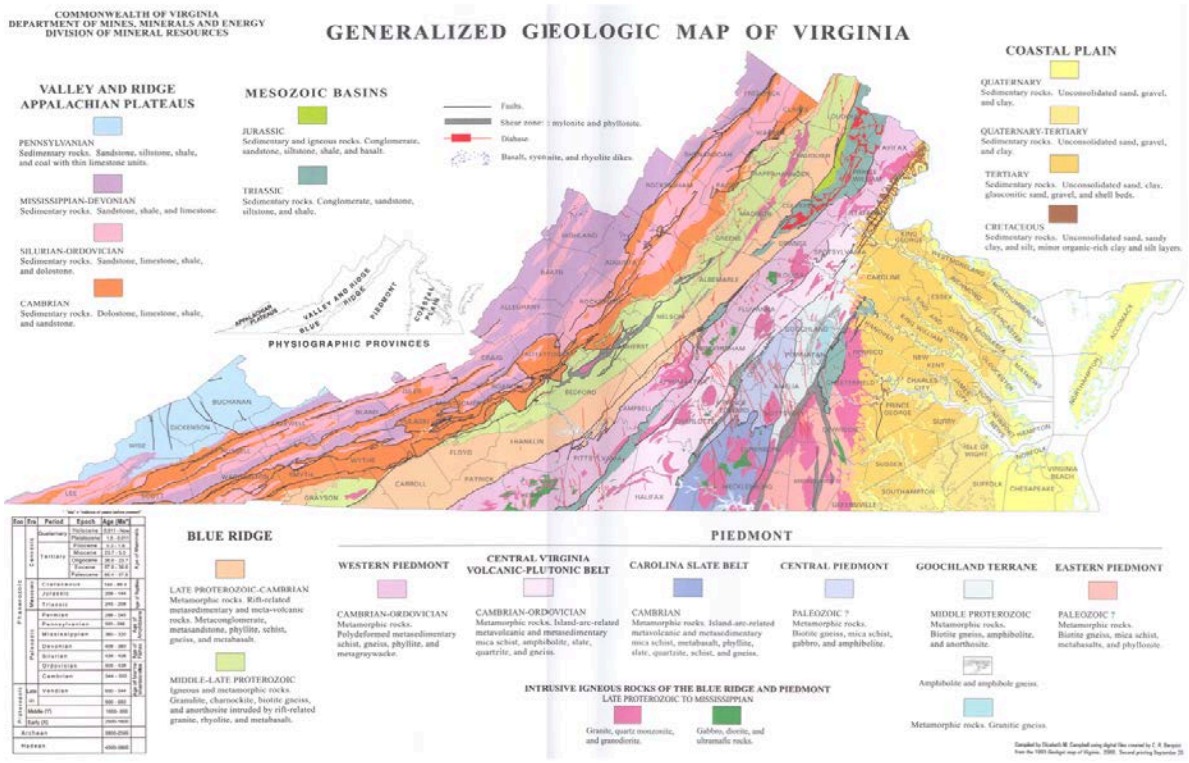
APPENDIX

A. Howard County Geologic Formations Map (Howard County GIS, 2004)

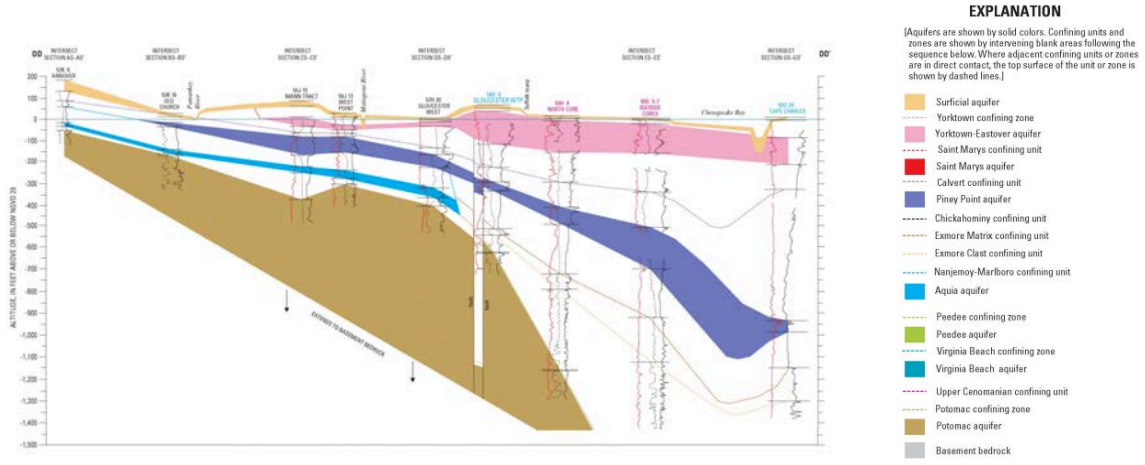


B. Virginia and West Point Geology

a) Virginia Geological Map with Coastal Plain at the rightmost side of the map (Campbell, 2000)



b) Hydrogeologic cross sections of the Virginia Coastal Plain (McFarland and Bruce, 2006)



C. Updated THB1D Inversion Code

a) Main Script (edited portions bolded)

```
clear
close all
tic
% Add subdirectory to matlab path
addpath(genpath('./'))
%working_fold='Shots_Summer2016'; % name of folder working in
%fname = 'Clarksville_20251024.new'; % Output model folder name
fname = 'WP_201_picks.mat';
disp_file = "WP_201_disp_data.mat"; % dispersion data name
% min and max frequencies used for dispersiono inversion - depends on site
minFreq = 3; % try 3 for WP, was 4 for clarksville
maxFreq = 40;
PlotDepth = 60; % this is only for plotting
maxiter = 5e4; % Maximum number of model iterations
hier = 0;  % Hierachical?  Currently doesn't work.
datsav = 10; % Save model every this many steps
burn = 0.5*maxiter/datsav;% 50;  % Start doing stats after this many saved models
% Define file folder and file name
tmp = dir('./data/');
%file_name = 'WP_201_picks.mat'; % input file name
file_name = 'WP_201_picks.mat';
% Range for velocities and interface depths
prior.v = [100 4000]; % range of velocities (m/s)
prior.vpvs = [1.6 6.2]; % range of vp/vs ratio
prior.vs = [30 2500];
prior.h = [0 50]; % Thickness of layer (m)
% "Proposal Sigmas"
% Standard deviations for proposing changes to model
psig.v = 100; % vary for vp
psig.vpvs = 0.2; % vary for vp/vs
psig.h = 2; % range for thickness
psig.d = 100; % range for vp during change option
psig.n = .002; % range for arrival time noise
psig.ndisp = .2; % range for dispersion noise
% Initial velocity and depth to top of layer
v0 = [300  1000  2000  2500  3000  3500 ];
vpvs = [2 2 2 2 2 2];
vs0 = v0./vpvs;
h0 = [0  8  16  24  32  40 ];
%%
[profile{1:length(tmp)}] = deal(tmp.name);
profile = profile(~cellfun(@isempty,strfind(profile,file_name)));
clear tmp
%%
% Iterate over lines
ipro = 1;
clear Tmm ensemble
pronum = profile{ipro};
pronum(end-11:end) = [];  % Profile number for plotting
% Regular depth grid for plotting
Nzgr = prior.h(2)+1;
zgr = linspace(prior.h(1),prior.h(2),Nzgr);
% Load data
load (['./data/' profile{ipro}]);
load(disp_file); % load dispersion data
fprintf('Loaded %s\n',['./data/' profile{ipro}])
```

```
% change back to master
%Master=combined;
T= Master(:,1); % Traveltimes
X = abs(Master(:,2)-Master(:,3)); % Distanecs
X(T==0)=[];
T(T==0)=[];
xx=sort(unique(X));
% Calculate standard deviation of picks  all at once
devs = [];
for m=1:length(xx)
   tmp=T(X==xx(m));
   if length(tmp)>=3
       devs = [devs; tmp-mean(tmp)];
   end
end
Dsig = sqrt(sum(devs.^2)/(length(devs)-1)); % for vp
Ndata=length(T); % Number of picks
Nsig = 1; % Number of noise variables
% computing p-wave arrival
v=v0;
vs=vs0; % for dispersion
h=h0;
dh=diff(h);
p = v.^-1;  % Slowness in each layer
Nlay = length(v); %Initialize number of layers
% Compute traveltimes based on model
tau=zeros(size(v));
for mm=1:length(v)
   for nn=1:mm-1
       tau(mm) = tau(mm) + 2*sqrt(p(nn).^2-p(mm).^2)*dh(nn);
   end
   Tmm(:,mm) = tau(mm)+p(mm)*X;
end
Tm=min(Tmm,[],2);
% Calculate error function and weighted error function
E=sum((T-Tm).^2./Dsig.^2);
E0=sum((T-Tm).^2);
E0 = E0/length(T)*1e3;
% computing dispersion; first generate layer model in a text file
N_layer = length(dh);
thickness = dh;
N = length(h);
vp = v(1:N-1);
vss = vs(1:N-1) ;
vp_bottom = v(N);
vs_bottom = vs(N);
% create an output model txt file
fid = fopen('test.model','w');
fprintf(fid,'%s\n',num2str(N_layer));
for i = 1:N_layer
   fprintf(fid,'%s %8.2f %8.2f 2500\n',num2str(round(thickness(i))),vp(i),vss(i));
end
fprintf(fid,'0 %f %f 2500\n',vp_bottom,vs_bottom);
fclose(fid);
% call geopsy
system(['gpdc test.model -R 1 -min ' num2str(minFreq) ' -max ' num2str(maxFreq) ' >
output.model']);
% read output
A = readlines('output.model');
A = A(~startsWith(A, '#'));
N = length(A)-1;
clear freq velo
```

```
for i = 1:N
   tmp = sscanf(A(i),'%f');
   freq(i) = tmp(1);
   velo(i) = 1/tmp(2);
end
% calculate error for dispersion
% number from 0-1 as the misfit (dispersion curve we
% make in geopsy has 1, 0-1 val is the deviation)
% take geopsy disp curve points and look for them in each dispersion image
nPoints = length(freq);
nDisps = size(disp_image,2); % number of dispersion files from data
% get the closest freq and vel values
row = zeros(1,nPoints); % freqs
col = zeros(1,nPoints); % vels
for i = 1:nPoints
   [~,row(i)] = min((f-freq(i)).^2); %freq
   [~,col(i)] = min((ct-velo(i)).^2); %velocity
end
% get the matching dispersion amplitudes
for d = 1:nDisps
   for i = 1:nPoints
       amps = disp_image{d};
       dataVel(i,d) = amps(col(i),row(i));
   end
end
A=dataVel(:);
A(isnan(A)) = 0;
A(A==0)=[]; % removing zeroes
%%
Am=ones(length(A),1); % amplitude of model (from geopsy, all points will be 1)
devs = A-Am;
Dsigd = sqrt(sum(devs.^2)/(length(devs)-1)); % for dispersion? check if this is a good way
% error functions for dispersion
Ed = sum((A-Am).^2./Dsigd.^2);
Ed0 = sum((A-Am).^2);
Ed0 = Ed0/length(A);
% Interpolate onto regular grid
vgr = zeros(round(maxiter/datsav),Nzgr); % Velocity(z) vp
vsgr = zeros(round(maxiter/datsav),Nzgr); % Velocity(z) vs
discgr = zeros(round(maxiter/datsav),Nzgr); % Discontinuity(z)
% Initialize ensemble
ensemble(round(maxiter/datsav)) = struct();
cnt=0;
%% Iterate model
for m = 1:maxiter
   if rem(m,200) == 0
       disp(['iteration ' num2str(m) ])
   end
   % Choose an operation for updating model
   oper = RandomOperRef(hier,Nlay);
   % Update random layer
   [Tm2,v2,vs2,h2,Dsig2,Dsigd2,delv] = tmp_UpdateMod(oper,X,v,vs,h,Dsig,Dsigd,psig,prior);
   % Recalculate error function
   E2=sum((T-Tm2).^2./Dsig2.^2);
   E02=sum((T-Tm2).^2);
   % Update dispersion model
   % computing dispersion; first generate layer model in a text file
   dh2 = diff(h2);
   N_layer = length(dh2);
   thickness = dh2;
   N = length(h2);
   vp = v2(1:N-1);
```

```
    vss = vs2(1:N-1) ;
    vp_bottom = v2(N);
    vs_bottom = vs2(N) ;
    % create an output model txt file
    fid = fopen('test.model','w');
    fprintf(fid,'%s\n',num2str(N_layer));
    for i = 1:N_layer
        if thickness(i) < 1
            thickness(i) = 1;
        end
        fprintf(fid,'%s %6.1f %6.1f 1500\n',num2str(round(thickness(i))),vp(i),vss(i));
    end
    fprintf(fid,'0 %6.1f %6.1f 2500\n',vp_bottom,vs_bottom);
    fclose(fid);
    % call geopsy
    system(['gpdc test.model -R 1 -min ' num2str(minFreq) ' -max ' num2str(maxFreq) ' >
output.model']);
    % read output
    AB = readlines('output.model');
    AB = AB(~startsWith(AB, '#'));
    N = length(AB)-1;
    clear freq velo
    for i = 1:N
        tmp = sscanf(AB(i),'%f');
        freq(i) = tmp(1);
        velo(i) = 1/tmp(2);
    end
    if N == 0
        velo = 0;
        freq = 3;
    end
    global_vel(m,:) = velo;
    nPoints = length(freq);
    % get the closest freq and vel values
    row = zeros(1,nPoints); % freqs
    col = zeros(1,nPoints); % vels
    for i = 1:nPoints
        [~,row(i)] = min((f-freq(i)).^2); %freq
        [~,col(i)] = min((ct-velo(i)).^2); %velocity
    end
    % get the matching dispersion amplitudes
    for d = 1:nDisps
        for i = 1:nPoints
            amps = disp_image{d};
            dataVel(i,d) = amps(col(i),row(i));
        end
    end
    A2=dataVel(:);
    A2(isnan(A2)) = 0;
    A2(A2==0)=[]; % removing zeroes
    Am=ones(length(A2),1); % amplitude of model (from geopsy, all points will be 1)
    % recalc error for dispersion
    Ed2=sum((A2-Am).^2./Dsigd2.^2);
    Ed02=sum((A2-Am).^2);
    Ed02 = Ed02/length(A);
    E02 = E02/length(T)*1e3;
    % % ignore P arrival times
    % E2 = Ed2;
    % E02 = Ed02;
    dE=E2-E; % Change in error (vp)
    dEd=Ed2-Ed; % change in error (dispersion)
    %%%% combine arrival and dispersion misfit
```

```
    dE = dE + dEd;
    dEd = dE;
    %%%%%
    % Decide whether to accept or reject model
    keep = AcceptItRef(oper,dE,dEd,psig,delv,prior,Dsig,Dsig2,Dsigd,Dsigd2);
    % If we accept the new model update values
    if keep>=rand(1)
        E=E2;
        E0=E02;
        Ed=Ed2;
        Ed0=Ed02;
        Dsig=Dsig2;
        Dsigd=Dsigd2;
        Tm=Tm2;
        A=A2;
        v=v2;
        vs=vs2;
        h=h2;
        Nlay = length(v);
    end
    % Save model values and regular intervals
    if  mod(m,datsav)==0
        cnt=cnt+1;
        ensemble(cnt).v=v;
        ensemble(cnt).vs=vs;
        ensemble(cnt).h=h;
        ensemble(cnt).Dsig=Dsig;
        ensemble(cnt).Dsigd=Dsigd;
        ensemble(cnt).E=E;
        ensemble(cnt).E0=E0;
        ensemble(cnt).Ed=Ed;
        ensemble(cnt).Ed0=Ed0;
        ensemble(cnt).Nlay=Nlay;
        %display(num2str(v))
        ensembleE0(cnt) = E0;
        ensembleEd0(cnt) = Ed0;
        ensembleTm(cnt,:) = Tm';
        % Find regular grid values
        vtmp=zeros(1,Nzgr);
        disctmp=zeros(1,Nzgr);
        for ih=1:length(h)
            vtmp(zgr>=h(ih))=v(ih);
        end
        vstmp=zeros(1,Nzgr);
        for ih=1:length(h)
            vstmp(zgr>=h(ih))=vs(ih);
        end
        for ih=2:length(h)
            disctmp(find(zgr<=h(ih),1,'last'))=1;
        end
        vgr(cnt,:)=vtmp;
        vsgr(cnt,:) = vstmp;
        discgr(cnt,:)=disctmp;
        if  mod(m,1000*datsav)==0
            fprintf('Saved %d models\n',cnt);end
    end
end
%%
% Compute basic stats for model
meanv = mean(vgr(burn:cnt,:));
stdv = std(vgr(burn:cnt,:));
meanvs = mean(vsgr(burn:cnt,:));
```

```
stdvs = std(vsgr(burn:cnt,:));
% Gather all predicted dispersion wanteed
gather_disp = global_vel(burn*datsav:datsav:end,:); %all dispersion velocity during iterations
%discs = sum(discgr(burn:cnt,:))/(cnt+1-burn);
% Create output directories
if ~exist(sprintf('./models/%s/',fname),'dir')
   mkdir(sprintf('./models/%s/',fname));
end
if ~exist(sprintf('./figs/%s/',fname),'dir')
   mkdir(sprintf('./figs/%s/',fname));
end
%%% create a table of Vp and Vs
clear Vmodel
Vmodel.depth = zgr';
Vmodel.Vp = meanv';
Vmodel.Vs = meanvs';
Vmodel.VpStd = stdv';
Vmodel.VsStd = stdvs';
Vmodel = struct2table(Vmodel);
save(['./models/' fname '/VelocityModel.mat'],'Vmodel');
writetable(Vmodel,['./models/' fname '/VelocityModel.csv']);
%% Plot all of the figures
Plot_All_Figs
%% plot dispersion figs
% add to plot_all_figs when finalized
% dispersion curve evolution? - add that with model evolution above
% maybe
% final vs profile
% plot the average dispersion as an image
ave_image = zeros(size(disp_image{1}));
for ii = 1:nDisps
   ave_image = ave_image + disp_image{ii}/nDisps;
end
figure;imagesc(f,ct,ave_image)
hold on
% plot(freq,(gather_disp),'k')
errorbar(freq,mean(gather_disp),std(gather_disp),'w.')
xlim([0 40])
set(gca,'ydir','normal')
% colormap(crameri('roma'))
title('Average Dispersion & Predicted Curve (Clarksville)')
xlabel('Frequency (Hz)')
ylabel('Phase Velocity (m/s)')
legend('Predicted dispersion curve')
%%
% figure
% set(gca,'Fontsize',14);box on
% [Xsort, isort] = sort(X);
% Tmsort = Tm(isort);
% pl1 = scatter(X,T,100,'k.'); hold on;
% pl2 = plot(unique(Xsort),unique(Tmsort),'r','Linewidth',2);
% title(sprintf('Final data fit, Profile %s',pronum),'Fontsize',14)
% xlabel('Offset (m)','Fontsize',14);ylabel('Traveltime (s)','Fontsize',14)
% legend([pl1, pl2],{'Data','Final model'},'Location','SouthEast')
figure
subplot(2,1,1);plot(1:datsav:datsav*cnt,ensembleE0)
ylabel('misfit (ms)')
title('Travel time curve')
set(gca,'yscale','log','xscale','log')
subplot(2,1,2);plot(1:datsav:datsav*cnt,ensembleEd0)
ylabel('misfit (power)')
xlabel('iterations')
```

```
title('Dispersion')
set(gca,'yscale','log','xscale','log')
toc
```

b) Function to change model, UpdateMod (edited portions bolded)

```
function [Tm2,v2,vpvs2,h2,Dsig2,Dsigd2,delv] = UpdateMod(oper,X,v,vpvs,h,Dsig,Dsigd,psig,prior)
   v2=v;
   vpvs2=vpvs; % keep for return
   h2=h;
   Dsig2=Dsig;
   Dsigd2=Dsigd;

if strcmp(oper,'changeV')    %CHANGE V
   %%
   nind = randi(length(v));

   v2=v;
   vpvs2=vpvs; % keep for return
   h2=h;
   Dsig2=Dsig;
   Dsigd2=Dsigd;

   vtmp=[prior.v(1) v prior.v(2)];

   delv = psig.d*randn(1);
   delv = max([delv vtmp(nind)-vtmp(nind+1)]);
   delv = min([delv vtmp(nind+2)-vtmp(nind+1)]);

   v2(nind) = v2(nind) + delv;

   %%

   dh=diff(h2);

   p = v2.^-1;


   tau=zeros(size(v2));
   for mm=1:length(v2)
       for nn=1:mm-1
           tau(mm) = tau(mm) + 2*sqrt(p(nn).^2-p(mm).^2)*dh(nn);
       end

       Tmm(:,mm) = tau(mm)+p(mm)*X;
   end


   Tm2=min(Tmm,[],2);


elseif strcmp(oper,'changeps')
   %% change vpvs ratio (same as first half of changeV)
   nind = randi(length(v));

   vpvs2=vpvs;
   Dsigd2=Dsigd;
```

```
    pstmp=[prior.vpvs(1) vpvs prior.vpvs(2)];


    delv = psig.vpvs*randn(1);
    delv = max([delv pstmp(nind)-pstmp(nind+1)]);
    delv = min([delv pstmp(nind+2)-pstmp(nind+1)]);

    vpvs2(nind) = vpvs2(nind) + delv;
    %% old, keep for func returns
    v2=v;
    h2=h;

    dh=diff(h2);
    p = v2.^-1;
    tau=zeros(size(v2));
    for mm=1:length(v2)
        for nn=1:mm-1
            tau(mm) = tau(mm) + 2*sqrt(p(nn).^2-p(mm).^2)*dh(nn);
        end
        Tmm(:,mm) = tau(mm)+p(mm)*X;
    end
    Tm2=min(Tmm,[],2);

elseif strcmp(oper,'move')      %MOVE LAYER
    %%
    nind = 1 + randi(length(v)-1);


    v2=v;
    vpvs2=vpvs; % keep for return
    h2=h;
    Dsig2=Dsig;

    htmp=[h prior.h(2)];


    delv = psig.h*randn(1);
    delv = max([delv htmp(nind-1) - htmp(nind)]);
    delv = min([delv htmp(nind+1) - htmp(nind)]);
    h2(nind) = h2(nind) + delv;

    dh=diff(h2);

    p = v2.^-1;


    tau=zeros(size(v2));
    for mm=1:length(v2)
        for nn=1:mm-1
            tau(mm) = tau(mm) + 2*sqrt(p(nn).^2-p(mm).^2)*dh(nn);
        end

        Tmm(:,mm) = tau(mm)+p(mm)*X;
    end


    Tm2=min(Tmm,[],2);

elseif strcmp(oper,'birth')     %ADD LAYER
    %%
    newh = prior.h(1) + rand*prior.h(2);
```

```
    htmp = [prior.h(1) h prior.h(2)];

    %ioldv = find(htmp>newh,1,'first');
    ioldv = find(htmp<newh,1,'last');

    Dsig2=Dsig;

    vtmp=[prior.v(1) v prior.v(2)];
    vpvstmp = [prior.vpvs(1) vpvs prior.vpvs(2)];
    % for a new Vp layer
    delv = abs(psig.v*randn(1));

    delv = max([delv vtmp(ioldv)-vtmp(ioldv+1)]);
    delv = min([delv vtmp(ioldv+1)-vtmp(ioldv)]);

    newv = vtmp(ioldv)+delv;
    % for a new VpVs layer
    delv2 = abs(psig.vpvs*randn(1));

    delv2 = max([delv2 vpvstmp(ioldv)-vpvstmp(ioldv+1)]);
    delv2 = min([delv2 vpvstmp(ioldv+1)-vpvstmp(ioldv)]);

    newvpvs = vpvstmp(ioldv)+delv2;


    h2 = [h newh];
    v2 = [v newv];
    vpvs2 = [vpvs newvpvs];

    [h2,ih]=sort(h2);
    v2=v2(ih);
    vpvs2 = vpvs2(ih);


    dh=diff(h2);

    p = v2.^-1;

    tau=zeros(size(v2));
    for mm=1:length(v2)
        for nn=1:mm-1
            tau(mm) = tau(mm) + 2*sqrt(p(nn).^2-p(mm).^2)*dh(nn);
        end

        Tmm(:,mm) = tau(mm)+p(mm)*X;
    end


    Tm2=min(Tmm,[],2);


elseif strcmp(oper,'death')     %REMOVE LAYER
    nind = 1 + randi(length(v)-1);


    v2=v;
    h2=h;
    Dsig2=Dsig;
    vpvs2=vpvs; % keep for return

    delv = v2(nind-1) - v2(nind);
```

```
    v2(nind)=[];
    h2(nind)=[];
    vpvs2(nind)=[];


    dh=diff(h2);
    p = v2.^-1;


    tau=zeros(size(v2));
    for mm=1:length(v2)
        for nn=1:mm-1
            tau(mm) = tau(mm) + 2*sqrt(p(nn).^2-p(mm).^2)*dh(nn);
        end

        Tmm(:,mm) = tau(mm)+p(mm)*X;
    end


    Tm2=min(Tmm,[],2);

elseif strcmp(oper,'noise')     %CHANGE NOISE

    v2=v;
    vpvs2=vpvs; % keep for return
    h2=h;

    delv=psig.n*randn(1);

    Dsig2=Dsig+delv;

    dh=diff(h2);
    p = v2.^-1;

    tau=zeros(size(v2));
    for mm=1:length(v2)
        for nn=1:mm-1
            tau(mm) = tau(mm) + 2*sqrt(p(nn).^2-p(mm).^2)*dh(nn);
        end

        Tmm(:,mm) = tau(mm)+p(mm)*X;
    end

    Tm2=min(Tmm,[],2);

elseif strcmp(oper,'noisedisp')
    %% new
    vpvs2=vpvs; % keep for return
    delv = psig.ndisp*randn(1);
    Dsigd2=Dsigd+delv;
    %% old, keep for func returns
    v2=v;
    h2=h;

    dh=diff(h2);
    p = v2.^-1;


    tau=zeros(size(v2));
    for mm=1:length(v2)
        for nn=1:mm-1
            tau(mm) = tau(mm) + 2*sqrt(p(nn).^2-p(mm).^2)*dh(nn);
```

```
        end

        Tmm(:,mm) = tau(mm)+p(mm)*X;
    end


    Tm2=min(Tmm,[],2);

else
    display('Thats not a thing')

end
```

D. Porosity and saturation inversion code (edited portions bolded)

```
%% Porosity and saturation inversion
% In this code, we assume known Vp and Vs measurements, we test the range of predicted
%  porosity and saturation with a varied model parameters (in Archie and
%  rock physics model)
%  In this version, we assume a fixed q and f and search for por and sat,
%  and then randomly set another set and so on
clear
close all
%% observation
% read velocity models from csv
data = readtable('clark_model.csv');
%data = readtable('wp_model.csv');
depths = data{:,1};
Vps = data{:,2}; Vss = data{:,3};
vp_std = data{:,4}; vs_std = data{:,5};
Vps = Vps + vp_std; Vss = Vss + vs_std;
Vps = Vps - vp_std; Vss = Vss - vs_std;
% calc vp/vs and sigma
vpvs = Vps ./ Vss;
vpvs_sigma = vpvs .* sqrt((vp_std./Vps).^2 + (vs_std./Vss).^2);
% por sat sigmas
por_plus_std = load("por +std.txt");
por_minus_std = load("por-std.txt");
sat_plus_std = load("sat +std.txt");
sat_minus_std = load("sat-std.txt");
%% model prior range
% prior
prior.sat = [0 0.99];
prior.por = [0.01 0.99];
% 1-sigma of perturbation
psig.sat = 0.05;
psig.por = 0.05;
% initial porosity and saturation
por0 = 0.3;
sat0 = 0.7;
% inversion prep
% number of total iteration (of model parameters)
% number of iterations
iteration = 2e5;
datsav = 1000;
% burn-in
burn = 0.7;
%%
tic
```

```
for i = 1:length(depths)
   cnt=0;
   obs_Vp = Vps(i);
   obs_Vs = Vss(i);
   obs_depth = depths(i) + 0.1;
   dep0 = obs_depth;
       % pick values based on prior studies
       q0 = 0.3; % quartz
       f0 = 0.2; % feldspar (or whatever is in the calc function)
       %% predicted values
       % first, estiamte the misfit of Vp and Res
       [Vp0,Vs0] = get_vpvs(por0,sat0,dep0,q0,f0);
       E_Vp = (Vp0-obs_Vp).^2;
       E = E_Vp+E_Vs;
       %%
       por = por0;
       sat = sat0;

       for itr = 1:iteration
           oper = RandomOperation;
           [por2,sat2,q2,f2] = UpdateModel(oper,prior,psig,por,sat,q0,f0);
           [Vp2,Vs2] = get_vpvs(por2,sat2,dep0,q0,f0);
           E_Vp2 = (Vp2-obs_Vp).^2;
           E_Vs2 = (Vs2-obs_Vs).^2;
           E2 = E_Vp2+E_Vs2;
           dE = E2 - E;
           X = 1;
           keep = min(1,X*exp(-(dE)/2));
           if keep >= rand(1)
               E = E2;
               E_Vp = E_Vp2;
               E_Vs = E_Vs2;

               Vp = Vp2;
               Vs = Vs2;
               por = por2;
               sat = sat2;
               q = q2;
               f = f2;

           end
           % only start saving after burn-in
           if mod(itr,datsav) == 0 && itr > iteration * burn
               cnt = cnt + 1;
               ensemble(i,cnt).por = por;
               ensemble(i,cnt).sat = sat;
               ensemble(i,cnt).q = q;
               ensemble(i,cnt).f = f;
               ensemble(i,cnt).Vp = Vp;
               ensemble(i,cnt).Vs = Vs;
               ensemble(i,cnt).E = E;
               ensemble(i,cnt).E_Vp = E_Vp;
               ensemble(i,cnt).E_Vs = E_Vs;
           end
        end

end
toc

%% store data for plotting
nDepth = length(depths);
Vp_mean = zeros(nDepth,1);
```

```
Vp_std = zeros(nDepth,1);
Vs_mean = zeros(nDepth,1);
Vs_std = zeros(nDepth,1);
por_mean = zeros(nDepth,1);
por_std = zeros(nDepth,1);
sat_mean = zeros(nDepth,1);
sat_std = zeros(nDepth,1);
for i = 1:nDepth
   Vp_mean(i) = mean([ensemble(i,:).Vp]);
   Vp_std(i) = std([ensemble(i,:).Vp]);
   Vs_mean(i) = mean([ensemble(i,:).Vs]);
   Vs_std(i) = std([ensemble(i,:).Vs]);
   por_mean(i) = mean([ensemble(i,:).por]);
   por_std(i) = std([ensemble(i,:).por]);
   sat_mean(i) = mean([ensemble(i,:).sat]);
   sat_std(i) = std([ensemble(i,:).sat]);
end

%% plot
figure;
% vp & vs
subplot(1,3,1);
plot(Vps, depths, 'r', 'LineWidth', 1.5); hold on;
plot(Vss, depths, 'b', 'LineWidth', 1.5);
plot(Vss+vs_std,depths,'k--','LineWidth',1.5); hold on;
plot(Vss-vs_std,depths,'k--','LineWidth',1.5);
plot(Vps+vp_std,depths,'k--','LineWidth',1.5);
plot(Vps-vp_std,depths,'k--','LineWidth',1.5);
plot(Vp_mean, depths);
plot(Vs_mean, depths);
set(gca,'YDir','reverse');
title('Mean 1D model');
xlabel('Velocity (m/s)'); ylabel('Depth (m)');legend('v_p','v_s','\sigma');
% vp/vs ratio
subplot(1,3,2);
plot(vpvs, depths, 'm', 'LineWidth', 1.5); hold on;
plot(vpvs+vpvs_sigma, depths, 'k--', 'LineWidth',1.5);
plot(vpvs-vpvs_sigma,depths,'k--','LineWidth',1.5)
set(gca,'YDir','reverse');
title('Mean v_p/v_s');
xlabel('v_p/v_s'); legend('v_p/v_s','\sigma')
% porosity & saturation
subplot(1,3,3);
plot(por_mean.*100, depths, 'c', 'LineWidth', 1.5); hold on;
plot(sat_mean.*100, depths, 'g', 'LineWidth', 1.5);
plot(por_plus_std.*100, depths, 'k--', 'LineWidth',1.5);
plot(por_minus_std.*100, depths, 'k--', 'LineWidth',1.5);
plot(sat_plus_std.*100, depths, 'k--', 'LineWidth',1.5);
plot(sat_minus_std.*100, depths, 'k--', 'LineWidth',1.5);
set(gca,'YDir','reverse');
title('Mean porosity and saturation');
xlabel('%'); legend('Por','Sat','\sigma');
set(gcf,'color','w');

%% rock phys model
function [Vp,Vs] = get_vpvs(por,sat,depth,q,f)
% Vp = get_vp(por,sat,depth,Qz concentration, feldspar concentration);
% I. setup elastic moduli of major minerals.
% See
(https://www.subsurfwiki.org/wiki/Elastic_properties_of_common_sedimentary_rock_forming_minera
ls)
% Flinchum et al (2018): Sherman granite
```

```
% mineral: K(GPa) G(GPa) rho(g/cm^3)
% microcline: 55.4 28.1 2.56
% Qz: 36.6 45 2.65
% Plag 75.6 25.6 2.63
% Perthite 46.7 23.65 2.54
% biotite 50.4 27.4 3.05
% for more values
% Model Parameters
%feld_bulk = 37.5e9;%20-30%  [Pa]
%feld_shear = 15e9; % [Pa]
quartz_bulk = 37e9; %30-50%   [Pa]
quartz_shear = 44e9; %   [Pa]
clay_bulk = 52.3e9;%2.9;   [Pa]
clay_shear = 31.7e9;%7;   [Pa]
% this is biotite mica, not feldspar
feld_bulk = 50.4e9; %20-30%  [Pa]
feld_shear = 27.4e9; % [Pa]
% e is an empirical parameter typically ranging from 2 to 5 (A. Brie, F. Pampuri, A. F.
Marsala, O. Meazza, in SPE-30595-MS (Society of Petroleum Engineers, Dallas, Texas, 1995);Gu
et al 2020)
e = 5;
% density
% rho_ss = 2650;
rho_ss = 2700; % density of sandstone [kg/m^3]
rho_h2o = 1000; % density of water
%critical porosity
phi = .6; % .4;
n = 10; %contact points (typical range: ???)
g = 9.8; %gravity constant
% q = 0.2; %Qz concentration
% f = 0.4; %Feldspar concentration
c = 1 - q - f; %Clay concentration
b_ss = 0.5* (((q*quartz_bulk) + (f* feld_bulk) + (c*clay_bulk))+...
   ((q/quartz_bulk)+(f/feld_bulk)+(c/clay_bulk))^-1); %Gu 2020 Eq S1
s_ss = 0.5* (((q*quartz_shear) + (f* feld_shear) + (c*clay_shear))+...
   ((q/quartz_shear)+(f/feld_shear)+(c/clay_shear))^-1); %Gu 2020 Eq S1
K_f = 2.2e9; %bulk modulus water [Pa]
K_g = 101e3; %bulk modulus air [Pa]
% Poisson's ratio
mu_ss = (3*b_ss - 2*s_ss)/(6*b_ss + 2*s_ss); %Gu 2020 Eq S4
por_var = por;
sat_model = sat;
rho_b_dry = (1-por_var)*rho_ss; % density of the dry model with different porosity
P_eff_dry = rho_b_dry.*depth.*g; % pressure of the dry model (rho g h)
%                 P_eff_dry = rho_b_dry.*abs(ElevFMM(z,1)-z).*g; % pressure of the dry model
(rho g h)
K_hm = (((n^2*(1-phi)^2*s_ss^2)/(18*pi^2*(1-mu_ss)^2)) .*...
   P_eff_dry).^(1/3); %Gu 2020 Eq S2
G_hm = ((5-4*mu_ss)/(5*(2-mu_ss)))*...
   ((((3*n^2*(1-phi)^2*s_ss^2)/(2*pi^2*(1-mu_ss)^2)) .* P_eff_dry).^(1/3)); %Gu 2020 Eq S3
if por_var == phi
   K_dry = K_hm;
   G_dry = G_hm;
elseif por_var < phi % the lower bound (G_dry is based on Gu and different from Flinchum et
al. 2018)
   Z2 = (G_hm./6).*((9.*K_hm + ...
       8.*G_hm)./(K_hm + 2.*G_hm));
   K_dry_L = ((((por_var/phi)./(K_hm + ...
       ((4/3).*G_hm)))+ ((1-(por_var/phi))./(b_ss + (4/3).*G_hm))).^-1) - ((4/3).*G_hm);
   G_dry_L = ((((por_var/phi)./(G_hm + ...
       Z2))+ ((1-(por_var/phi))./(s_ss + Z2))).^-1) - Z2;
   K_dry = K_dry_L;
```

```
    G_dry = G_dry_L;
else % the upper bound (when porosity > critical porosity)
    Z2 = (G_hm./6).*((9.*K_hm + ...
        8.*G_hm)./(K_hm + 2.*G_hm));
    K_upper_bound = (((((1-por_var)/(1-phi))./(K_hm + ...
        ((4/3).*G_hm)))+ (((por_var-phi)/(1-phi))./((4/3).*G_hm))).^-1) - ((4/3).*G_hm);
    G_upper_bound = (((((1-por_var)/(1-phi))./(G_hm + ...
        Z2))+ (((por_var-phi)/(1-phi))./(Z2))).^-1) - Z2;
    K_dry = K_upper_bound;
    G_dry = G_upper_bound;
end
Vp_dry = sqrt((K_dry+(4/3.*G_dry))./rho_b_dry);% dry condition (not used here)
Vs_dry = sqrt(G_dry./rho_b_dry);% dry condition (not used here)
% saturated state
S_g = 1 - sat_model; % gas saturation
rho_b = (1-S_g)*por_var*rho_h2o + (1-por_var)*rho_ss; %saturated version of the density term
k_f_var = ((K_f - K_g)*(1-S_g)^e) + K_g; %Gu 2020 Eq S8
K_sat = b_ss .* ((por_var.*K_dry-(((1+por_var)*k_f_var.*K_dry)./b_ss)+ ...
    k_f_var)./(((1-por_var)*k_f_var)+(por_var.*b_ss)-(k_f_var.*K_dry./b_ss))); %Gu 2020 Eq S7
Vp_hm_sat = sqrt((K_sat+(4/3.*G_dry))./rho_b); %Gu 2020 Eq S9 (with saturated density)
Vs_hm_sat = sqrt(G_dry./rho_b);
Vp = Vp_hm_sat;
Vs = Vs_hm_sat;
end

function oper = RandomOperation
nxt = randi(20);
if nxt<11
    oper = 'por';
elseif nxt<21
    oper = 'sat';
end

function [por2,sat2,q2,f2] = UpdateModel(oper,prior,psig,por,sat,q,f)
por2 = por;
sat2 = sat;
q2 = q;
f2 = f;
if strcmp(oper,'por')
    tmp = por + psig.por*randn(1);
    if tmp >= min(prior.por) && tmp <= max(prior.por)
        por2 = tmp;
    end
elseif strcmp(oper,'sat')
    tmp = sat + psig.sat*randn(1);
    if tmp >= min(prior.sat) && tmp <= max(prior.sat)
        sat2 = tmp;
    end
end
end
```