Freshwater-Saltwater Boundaries in Submarine Aquifers

Seth Reichelt
Department of Geophysics, Colorado School of Mines
Golden, CO 80401

Abstract

This study examines submarine groundwater offshore of Martha's Vineyard, Massachusetts, where an estimated 100 km^3 of freshwater has been identified. It focuses on enhancing the seismic interpretation of freshwater-saltwater boundaries within the aquifer by analyzing variations in locally collected seismic and electromagnetic (EM) data. By integrating industry-standard seismic interpretation techniques with machine learning methods, this research aims to improve the detection of transitions at freshwater-saltwater interfaces.

1 Introduction

Submarine groundwater systems hold vast quantities of offshore freshened groundwater ($\approx$ 300–500$\times 10^3 km^3$) distributed beneath the continental shelves worldwide (5). These aquifers, are a vital resource for addressing freshwater scarcity, especially in water-stressed coastal cities. Offshore of Martha's Vineyard, Massachusetts, an estimated 100 km^3 of freshwater reserves have been identified. To put this in perspective, one cubic kilometer of water is equivalent to more than 264 billion gallons—enabling this 100 km^3 aquifer to meet New York City's water demands for 72 years or the entire United States' consumption for 82 days based on current usage rates (15; 14).

The geophysical exploration of submarine aquifers is critical for understanding their potential as sustainable freshwater resources. Although groundwater comprises 97 percent of the Earth's freshwater, it remains poorly understood, with substantial gaps in its mapping and characterization. This study examines the submarine freshwater aquifer system offshore of Martha's Vineyard, where seismic investigations conducted in 2014 provided insights into the region's stratigraphy and glacial history. Using reflection seismic data, the study predicted freshwater distributions up to 30 km offshore, influenced by Pleistocene glaciations (11). In contrast, subsequent electromagnetic (EM) surveys along the same line in 2019 identified freshwater reserves extending as far as 90 km offshore, revealing a discrepancy in the interpreted extent of these aquifers (3).

Seismic data, traditionally employed for stratigraphic interpretation, has been underutilized for identifying hydrologic boundaries, such as those between freshwater and saltwater-saturated media. This is largely due to the subtle density contrasts that produce

minimal seismic responses. This research addresses this limitation by integrating theoretical frameworks with the combined interpretation of seismic and electromagnetic (EM) data. By analyzing changes in seismic attributes—such as amplitude and acoustic impedance—at known freshwater-saltwater and saltwater-saltier water boundaries derived from resistivity data, the study aims to predict expected variations within seismic datasets. The ultimate objective is to determine whether seismic attributes alone can delineate these hydrologic boundaries without directly relying on prior resistivity information.

This approach builds on established lithologic and fluid property assumptions, utilizing industry-standard seismic analysis tools and machine learning algorithms to enhance the interpretation of seismic responses at these boundary zones.

This research focuses on detecting boundary transitions between freshwater and saltwater, as well as transitions between varying salinity levels of saltwater, in seismic data collected offshore of Martha's Vineyard, Massachusetts. By correlating seismic findings with EM results, the study aims to refine methodologies for identifying and characterizing subsurface fluid boundaries. These advancements support the sustainable management of submarine freshwater resources, emphasizing the role of seismic data as a vital tool in hydrologic investigations and addressing challenges in resource accessibility for future generations.

2 Theory

The theoretical framework presented here details seismic reflections at freshwater and saltwater boundaries using acoustic impedance (Z) and related site-specific parameters.

Acoustic impedance is derived from P-wave velocity ($V_p = \sqrt{\frac{M}{\rho}}$) and bulk density ($\rho_b$), facilitating the identification of transitions between layers based on their seismic signatures (9):

$$Z = \rho_b V_p \tag{1}$$

Furthermore, the P-wave modulus (M) establishes the relationship between bulk density and seismic velocity (9):

$$M = \rho_b V_p^2 \tag{2}$$

Bulk density (ρ_b) is estimated using porosity (ϕ), grain density (ρ_g) and fluid density (ρ_f) (6):

$$\rho_b = \phi\rho_f + (1 - \phi)\rho_g \tag{3}$$

For quartz-rich sediments, the density of the grain matrix is typically approximated as $2650\,\mathrm{kg\,m^{-3}}$ (4).

2.1 Porosity Considerations

A porosity (ϕ) value of 0.2 was applied in Equation 3, based on geological interpretations derived from processed seismic data (11). The majority of freshwater is contained within the following units:

- **Unit 2:** Paleocene/Eocene carbonate mud with a porosity of 0.2.

- **Units 5 and 6:** Late Pleistocene sand, silt, clay clinoforms, and thin clay layers, also with a porosity of 0.2.

Beneath these units lies a smaller section of Cretaceous carbonate sandstone, which displays an increased porosity of 0.3. However, to reflect the dominant geological characteristics and maintain consistency, a standard porosity value of 0.2 was used in the bulk density calculations.

2.2 Calculated Values

The theoretical bulk densities were calculated using the mean velocity values determined for known saltwater and freshwater sections. These calculations provided typical acoustic impedance and P-wave modulus values for each medium:

- **Freshwater Section**:

$$\begin{aligned} V_p &= 1894.3\,\mathrm{m/s} \\ \rho_b &= 2320\,\mathrm{kg/m^3} \\ Z_{\mathrm{fresh}} &= 4.39 \times 10^6\,\mathrm{kg/(m^2 s)} \\ M_{\mathrm{fresh}} &= 8.33\,\mathrm{GPa} \end{aligned}$$

- **Saltwater Section**:

$$\begin{aligned} V_p &= 1955.2\,\mathrm{m/s} \\ \rho_b &= 2324.8\,\mathrm{kg/m^3} \\ Z_{\mathrm{salt}} &= 4.55 \times 10^6\,\mathrm{kg/(m^2 s)} \\ M_{\mathrm{salt}} &= 8.89\,\mathrm{GPa} \end{aligned}$$

These values provide a foundational understanding of the expected acoustic impedance and P-wave modulus, while also illustrating the subtle transitions caused by small variations in density and velocity between the freshwater and saltwater sections.

2.3 Haversine Formula:

The Haversine formula was applied to align latitudinal and longitudinal spatial data with the original seismic and EM survey lines. It calculates the great-circle distance (d) between two points on a sphere, where r represents the radius of the sphere (1).

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \tag{4}$$

2.4 Amplitude Envelopes

Amplitude variations were analyzed using amplitude envelopes, defined as:

$$A_{\text{envelope}} = \sqrt{A^2 + H(A)^2} \tag{5}$$

where A represents the amplitude, and $H(A)$ is the Hilbert transform of A. This approach emphasizes energy variations, which are critical for identifying fluid transitions (7).

2.5 Correlation Coefficients

The Pearson correlation coefficient (r) was computed to quantify the relationship between seismic amplitude and resistivity:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{6}$$

Here:

- x: Seismic amplitude values.
- y: Resistivity values.

This statistical measure evaluates the strength and direction of the linear relationship between the variables (12).

2.6 Machine Learning Interpretation

Machine learning techniques were applied to analyze boundary relationships based exclusively on seismic data.

2.6.1 K-Means Clustering

Amplitude envelopes were clustered into three groups using K-Means clustering ($k = 3$). This unsupervised learning method organizes data into clusters based on inherent similarities within the dataset (13).

2.6.2 Random Forest Classification

A Random Forest classifier was trained on seismic features to predict boundary probabilities. The input features and labels were defined as follows:

- Features: Frequency and amplitude envelopes.
- Labels: Known boundary regions (freshwater-saltwater and saltwater-saltier water).

Boundary probabilities (P_{boundary}) were calculated and visualized using the following formula:

$$P_{\text{boundary}} = \frac{1}{n_t} \sum_{i=1}^{n_t} p_i \tag{7}$$

where n_t is the number of trees, and p_i denotes the probability predicted by the i-th tree. This supervised learning approach integrates predictions from multiple decision trees to enhance classification accuracy and mitigate overfitting (10).

3 Experiments

3.1 Seismic Processing

The initial processing of the seismic and velocity data involved smoothing the velocity model using a low-pass filter to minimize ambient noise. The velocity data was then used to convert the seismic data into the time domain, facilitating their alignment and enabling a comprehensive analysis of the depth relationships within the dataset.

To improve the visibility of amplitude variations in the saltwater and freshwater boundary zones, automatic gain control (AGC) was applied. This technique amplified lower or less discernible amplitude variations, enhancing the clarity of the amplitude data. While it is unclear whether this step was part of the original survey, its implementation in this study ensured a more accurate interpretation of the minor amplitude transitions predicted at the freshwater-saltwater boundaries.

Using the provided location file, distances along the seismic line were calculated using the Haversine formula (Equation 5). The initial data location was referenced as the shoreline, with calculations extending up to 100 kilometers offshore to encompass the well-defined freshwater and saltwater boundaries within this domain. These calculations facilitated precise mapping of the seismic line.

After processing the depth-domain data, a contour plot was generated, with depth represented on the y-axis, distance on the x-axis, and amplitude variations depicted as contours. This plot closely aligned with patterns observed in the published data, confirming the accuracy of the data processing. To gain further insight into subsurface transitions, detailed analyses of known boundary zones were conducted by overlaying the seismic data with electromagnetic (EM) results. This approach enabled a clearer understanding of the boundary dynamics. The filtered, automatic gain-controlled seismic data is shown in Figure 1.

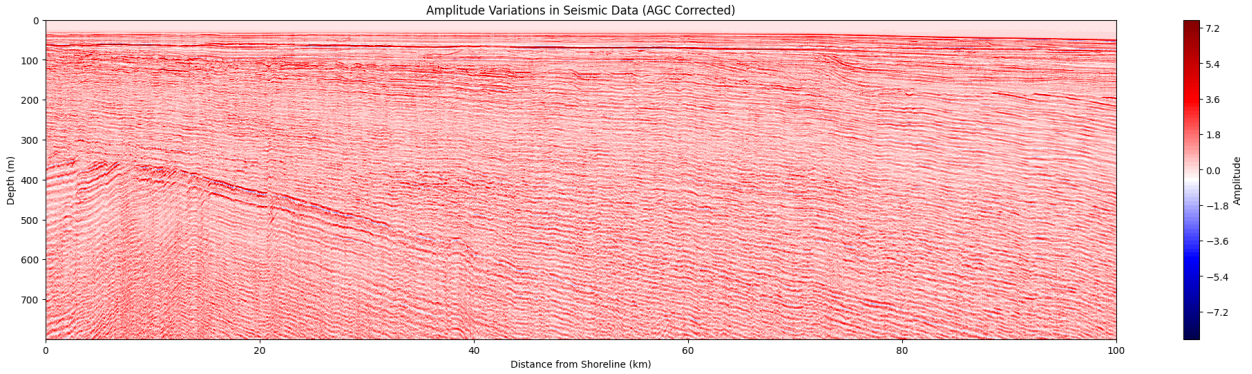

Figure 1: Automatic Gain Control Corrected Seismic Data. Sections with a sharper red contour indicate areas inclusive of higher amplitude contrasts and potential stratigraphic boundary zones. This seismic section effectively matched that previously interpreted in the initial 2014 seismic survey.

3.2 Acoustic Impedance Analysis

Acoustic impedance (Z) was calculated for the entire seismic section by discretizing the space into a thousand velocity points along the seismic line. These velocity values were multiplied by the previously calculated bulk densities of the freshwater and saltwater zones, as detailed in Section 2.2. The acoustic impedance calculations were applied to constrained depth and distance windows corresponding to known freshwater-saltwater and saltwater-saltier water boundaries. These boundaries were identified by overlaying the seismic AGC plot with the interpreted resistivity data from 2019 (3). This analysis provided a foundational understanding of the theoretical relationships between velocity, density, and seismic responses within the dataset.

3.3 Amplitude Analysis

To better understand the windows identified in the impedance analysis, wiggle plots were generated for each constrained section. Amplitude envelopes (Equation 5) for these sections were calculated to enhance the detection of subtle seismic variations across depth ranges. These envelopes utilized the Hilbert transform to provide a clearer visualization of energy distributions within the seismic data, facilitating the identification of potential boundary-related features.

The amplitude envelopes were analyzed using K-Means clustering (Equation 7), segmenting the data into three distinct groups: high, moderate, and low amplitudes. High values were associated with potential boundaries or transition zones.

Heatmaps of the amplitude envelopes were also generated, revealing consistent patterns at depth ranges corresponding to transition zones. Depth-segmented clustering using K-Means further validated these transitions, highlighting specific regions of interest.

3.4 Integration with Resistivity Data

The EM resistivity dataset was aligned with the seismic data using the Haversine formula (Equation 5), enabling direct comparisons along the seismic line. Resistivity values were

overlaid on amplitude envelope heatmaps and K-Means clustering results to validate the identified boundaries. Consistent decreases in resistivity were observed at depth ranges corresponding to the clustered amplitude envelope transitions.

Correlations were analyzed using both the EM dataset and the seismic amplitude envelopes within constrained windows corresponding to the saltwater-freshwater and saltwater-saltier water boundaries. Correlation coefficients (Equation 6) were calculated to quantify the relationship between seismic amplitude and resistivity.

3.5 Machine Learning Application

To extend these interpretations, a Random Forest classifier (Equation 8) was trained on windows defined from the integrated seismic and EM datasets. The model was then applied to predict boundary probabilities across the entire seismic section, relying exclusively on seismic data.

This integration of theoretical calculations, clustering techniques, and machine learning models established a robust framework for seismic interpretation. It facilitated a deeper understanding of subsurface fluid transitions and lithologic changes.

4 Discussion

4.1 Observed Acoustic Impedance Transitions

Within the known freshwater-to-saltwater transition zone, observed acoustic impedance values ranged from approximately:

$$3.844 \times 10^6 \,\mathrm{kg/m^2s} \,\mathrm{to}\, 3.86 \times 10^6 \,\mathrm{kg/m^2s}.$$

These values were mapped across the entire section and verified to be consistent within the first 25 meters when compared to the known electromagnetic (EM) data. However, this consistency diminished beyond this range, as shown in Figure 2a. This decline is likely attributed to stratigraphic and sedimentary variations within the boundary zone, where lithologic heterogeneity and porosity changes disrupt the otherwise distinct impedance contrast.

The calculated acoustic impedance in this zone falls below the theoretically derived value from Section 2.2, indicating variations in bulk density and P-wave velocity compared to the homogeneous assumptions made in the theoretical calculations.

For the known saltwater-to-saltier water transition zone, acoustic impedance values ranged from:

$$4.474 \times 10^6 \,\mathrm{kg/m^2s} \,\mathrm{to}\, 4.49 \times 10^6 \,\mathrm{kg/m^2s}.$$

This transition was identified in the lower portion of the freshwater section, bordering Unit 1 (Cretaceous carbonate sediments), which exhibits higher porosity, as discussed in Section 2.1. Unlike the freshwater-to-saltwater boundary, this transition zone displayed a more consistent alignment with theoretical values ($Z_{\mathrm{salt}} = 4.55 \times 10^6 \,\mathrm{kg/m^2s}$).

As shown in Figure 2b, this boundary followed a curved, loosely defined but observable pattern, remaining relatively consistent for approximately the first 50 kilometers offshore. Slight deviations in the observed impedance are attributed to lithologic changes within the

transition zone, where variations in P-wave velocity and bulk density disrupt the sharpness of the boundary.

While the calculated impedance transitions provide valuable insights into boundary dynamics, their alignment with observed patterns across broad sections underscores the limitations of theoretical models when applied to heterogeneous subsurface environments.

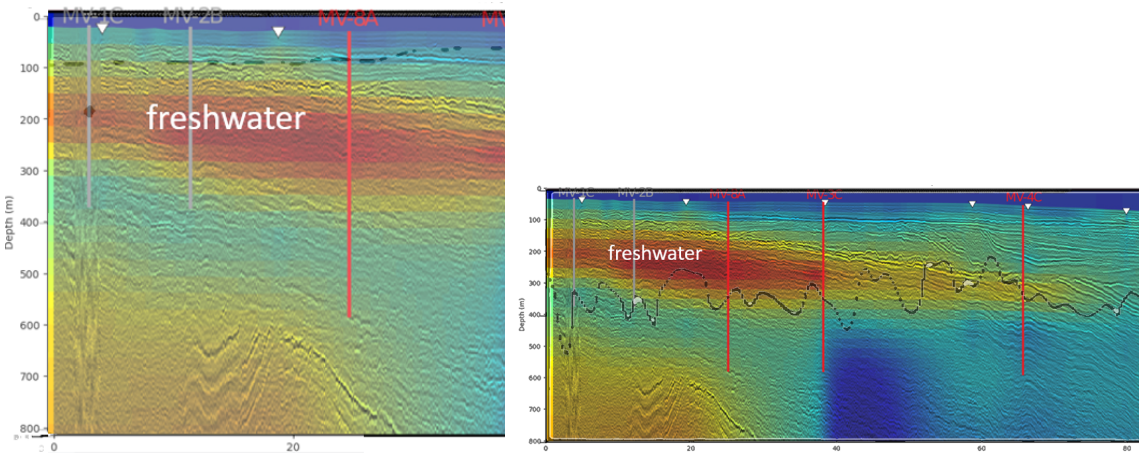

Figure 2: (a) Overlaid acoustic impedance transition range for the freshwater-to-saltwater boundary, and (b) transition range for the saltwater-to-saltier water boundary. The horizontal axis represents the distance from the shoreline, while the darker red contour indicates low resistivity, characteristic of the freshwater zone. In (a), the faint black line at the top diminishes beyond the known freshwater zone approximately 25 kilometers from the shoreline. In (b), the faint black line denotes the observed sporadic transition range, adjacent to the lower stratigraphic section.

4.2 Amplitude Inspection and Correlation Coefficient Analysis

The same constrained windows identified in the acoustic impedance analysis were used to analyze amplitude changes. To enhance visualization of the seismic data, wiggle plots were scaled at every second trace, improving the visibility of subtle trends. Due to the complexity of these trends, amplitude envelopes were calculated using Equation 5, providing a clearer representation of positive amplitude increases and revealing consistent patterns.

Using the machine learning technique of K-Means clustering ($k = 3$) to analyze trends, the freshwater-to-saltwater transition zone exhibited high cluster values within the 90 to 95-meter depth range. The amplitude envelopes were subsequently plotted as heat maps, as shown in Figure 3a, further confirming a transition zone coinciding with both the high cluster values and known freshwater-to-saltwater boundary. This boundary was validated by overlaying the seismic amplitude data with resistivity cross-sections constrained to the same windows.

To further quantify the relationship between seismic amplitude and resistivity, Pearson correlation coefficients were calculated using Equation 6. These coefficients were divided into five depth segments across the selected window, with the maximum negative correlation coefficient of $r = -0.46$ observed within the 92.5 to 101.25-meter depth range. This result indicates an inverse relationship between seismic amplitude and resistivity in this region. The negative correlation aligns with theoretical expectations: increasing seismic amplitudes (indicative of stronger reflections) correspond to decreasing resistivity values. This behavior occurs because higher salinity leads to reduced resistivity due to increased ionic concen-

tration, while regions with stronger seismic reflections exhibit higher amplitude envelopes, consistent with theoretical predictions (2; 8).

A similar analysis was conducted for the saltwater-to-saltier water boundary. Wiggle plots and amplitude envelopes revealed consistent patterns within an assumed transition zone. Using K-Means clustering and the heat map visualized in Figure 3b, a distinct trend was observed within the expected depth range. Verification against resistivity cross-sections further confirmed a valid transition zone, characterized by a gradual decrease in resistivity values corresponding to increased salinity.

Correlation coefficients calculated for this boundary yielded a maximum negative value of $r = -0.28$. This lower correlation, compared to the freshwater-to-saltwater boundary, reflects the reduced fluid and lithologic contrast between the saltwater and saltier water sections. Nevertheless, the analysis effectively identified specific depth ranges with distinct transitions, providing a reliable framework for interpreting subsurface properties.

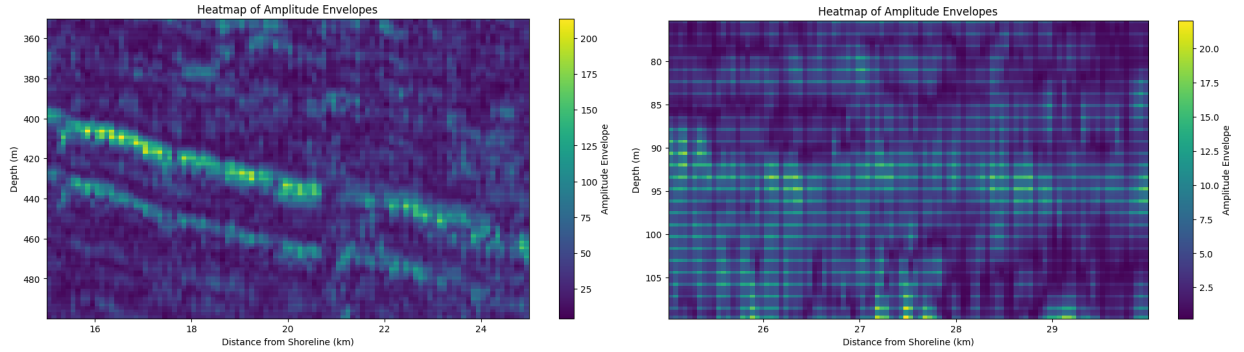

Figure 3: (a) Heat map of amplitude envelopes within the saltwater-to-saltier water transition zone, verified against resistivity data. Yellow regions indicate potential transitional boundaries, correlating with K-Means clustering results and correlation coefficients. (b) Heat map of the freshwater-to-saltwater boundary, consistent with resistivity data, K-Means clustering results, and correlation coefficient ranges.

The same steps were applied to the expanded, full seismic section, covering a depth range of 0 to 800 meters and a distance range of 0 to 100 kilometers offshore. Amplitude envelopes were calculated and visualized as heat maps, revealing significant amplitude increases at stratigraphic boundaries. These pronounced increases in the amplitude envelope aligned with interpretations from earlier seismic studies (11), highlighting substantial changes in lithologic properties and discounting the more subtle fluid change boundaries previously analyzed.

Correlation coefficients were calculated for 40 depths across the entire seismic section to evaluate relationships between amplitude envelopes and resistivity changes. The results showed sporadic trends with less consistency compared to the constrained windows. While known transitional, constrained sections exhibited strong negative correlations, broader sections showed more variable and less pronounced correlation coefficients at the identified depth ranges. This variability can be attributed to stronger increases over stratigraphic boundaries, which augment the higher negative correlation trends observed in the smaller windows. These inconsistencies likely result from subsurface lithologic heterogeneity and the influence of stratigraphic transitions extending beyond fluid-related boundaries.

The constrained windows provided a clear framework for identifying transitions be-

tween fluid boundaries, with correlation coefficients ranging from $r = -0.2$ to $r = -0.5$. To identify similar correlation coefficients across the full section, the seismic and EM data were divided into a 60 by 60 grid to evaluate changes in correlation coefficients across both depth and distance, compared to the earlier analysis, which focused solely on depth.

Figure 4 illustrates this correlation coefficient search, highlighting correlations within the -0.2 to -0.5 range over the expanded window. These findings reinforce the observation that values derived from zoomed-in windows of identified transition zones do not yield relationships consistent with the expanded dataset, mirroring the conclusions drawn from the expanded acoustic impedance search. However, the identified windows still provide valuable insights into seismic information aligned with fluid boundary transitions, offering a foundation for further analysis.

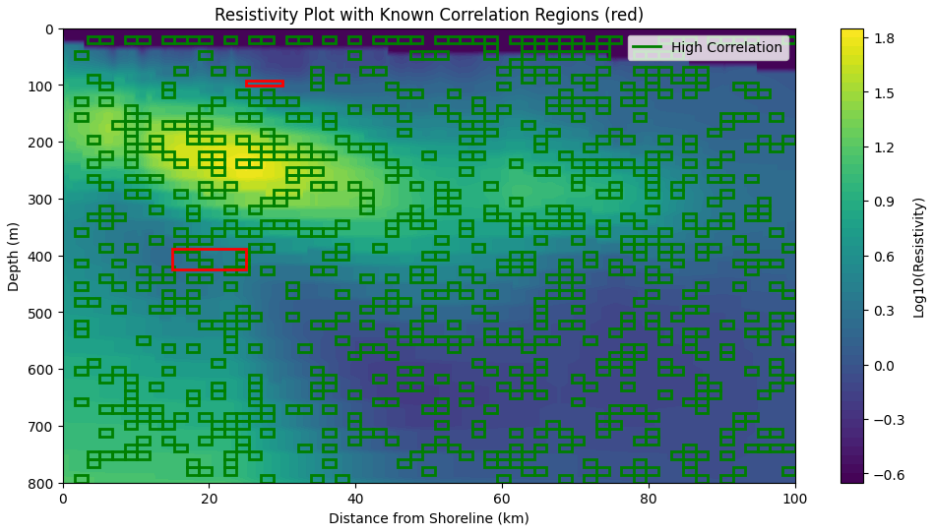

Figure 4: Correlation coefficients between resistivity and amplitude envelope data across the full surveyed section. Red boxes indicate previously defined transition zones, while green boxes highlight regions in the seismic and resistivity plots with correlation coefficients within the range of -0.2 to -0.5.

Machine Learning and Boundary Prediction Using Random Forest

Following the amplitude envelope analysis and correlation coefficient investigations, the identified transition zones served as the training data for a Random Forest classifier. These zones were defined as:

- Freshwater-to-Saltwater Boundary: Depth range of 92.5 to 101.5 meters below sea level and a distance range of 25 to 30 kilometers from the shoreline.
- Saltwater-to-Saltier Water Boundary: Depth range of 387.5 to 425 meters below sea level and a distance range of 15 to 25 kilometers from the shoreline.

Using these well-constrained and observed transition zones, the Random Forest model was trained to predict fluid boundaries based solely on seismic features extracted from the amplitude envelopes and frequencies within each identified transition window. However, the model was unable to predict strong boundary probabilities beyond the defined windows.

Within the given windows, the model produced high-probability predictions, while areas outside these windows displayed scattered patterns of low-probability zones across the section. The predicted depth and distance ranges closely matched the original input, reflecting the inherent bias in the training dataset, as only a single transition window of seismic data was used for each case.

The raw amplitude data and phase information could not be used as training data due to computational limitations. While the results indicate that the model generalized the seismic characteristics within the training zones, it was unable to confidently extrapolate them beyond the scope of the biased window information. Consequently, its predictions remained constrained by the limited scope of the input training data, as illustrated in Figure 5.

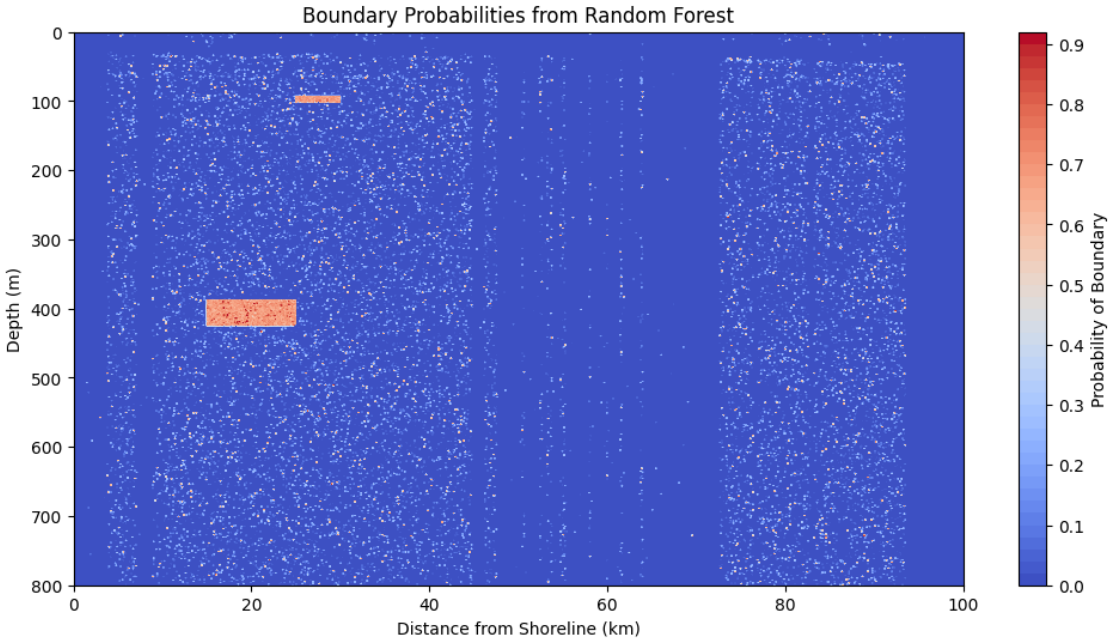

Figure 5: Boundary probabilities determined using the Random Forest Classifier. Blue areas represent regions with no probable boundaries, while darker red contours indicate a high probability of boundary presence. The small upper and larger lower boxes with increased probability correspond to the windows used as training data.

Implications for Model Improvement and Future Training

To enhance the model's capacity for boundary detection and mitigate biases, future iterations could incorporate more diverse and representative training windows, including those derived from broader depth, distance, and lithological ranges. Additionally, expanding the analysis to include neighboring seismic lines within the survey could verify whether these notable windows are consistent across the collected data. To achieve these improvements, the following strategies are recommended:

- **Acoustic Impedance Analysis:** Expand the acoustic impedance investigation to target diverse lithological transitions.

- **Amplitude Envelope Investigation:** Extend amplitude envelope analysis to include regions with subtle and complex amplitude variations. Incorporate noise quantification to better distinguish signal variations from data artifacts.
- **K-Means Clustering:** Revisit the clustering methodology by exploring alternative techniques (e.g., hierarchical clustering or Gaussian Mixture Models) to improve the segmentation of seismic amplitude variations and enhance the identification of transition zones.
- **Heatmap and Correlation Coefficients:** Utilize heatmap representations and correlation coefficient analysis with EM data to verify and refine additional identified transition zones.

By integrating these additional windows into the training dataset, the model could deliver a more robust and generalized interpretation of subsurface boundaries. This enhancement would enable the Random Forest classifier to more effectively detect boundaries in zones where resistivity data is unavailable or ambiguous, thereby increasing its utility for future seismic interpretations.

Potential Applications and Limitations

These methods establish a foundation for interpreting seismic data to identify subtle transitions between freshwater, saltwater, and saltier water boundaries. However, their applicability is constrained by the dataset's limited scope and the inherent biases in the training data. This approach is more effective in homogeneous subsurface environments, whereas the stratigraphic complexity of the dataset used in this study introduced large-scale changes that obscured minor amplitude variations.

Future work should prioritize integrating additional seismic and electromagnetic (EM) datasets collected along identical survey lines. Testing the model in controlled environments, such as sand tanks or localized studies, could also help refine its predictive capabilities. Enhancements in frequency content, data collection techniques, and processing workflows are essential for improving signal detection and resolution. Monitoring signal-to-noise ratios within transition zone windows is equally critical to ensure that subtle changes between fluid boundaries are not masked by noise.

Expanding the range of training windows to encompass diverse lithological and fluid boundary conditions and validating the method's reproducibility across different datasets and regions are critical next steps. While machine learning demonstrated potential for detecting fluid transitions, its utility was limited by biases and insufficient data diversity, highlighting the need for broader datasets to support generalized applications. Addressing these limitations would enhance the method's reliability, providing a valuable tool for seismic interpretation in appropriate geological contexts.

5 Conclusion

This study presented a comprehensive framework for integrating seismic and electromagnetic (EM) data to improve the interpretation of freshwater-saltwater and saltwater-saltier water

boundaries in a submarine groundwater system offshore of Martha's Vineyard. By employing acoustic impedance analysis, amplitude envelope calculations, and correlation coefficient investigations, key transition zones were identified that aligned with theoretical expectations and observed resistivity patterns. The findings demonstrated the utility of seismic attributes, such as amplitude variations and acoustic impedance, enhanced through automatic gain control, in delineating subtle hydrologic boundaries that are otherwise difficult to detect.

The application of machine learning, specifically a Random Forest classifier, showcased its potential to predict boundary probabilities based on constrained training zones derived from integrated seismic and EM datasets. However, the model's ability to generalize beyond the defined windows was limited by the narrow scope of input data. This highlights the importance of incorporating broader and more diverse training datasets to reduce bias and improve accuracy. While the integration of seismic analysis and machine learning provided valuable insights into subsurface fluid boundaries, the findings emphasized that seismic data alone is insufficient to resolve subtle fluid transitions without supplementary data, such as EM resistivity or high-resolution velocity measurements.

Future research should prioritize expanding the training datasets to include additional zones of interest, encompassing a wider range of lithological and fluid boundary conditions. Testing the methodologies in controlled environments, such as sand tanks or localized studies, and integrating seismic and EM datasets collected along identical lines will help refine the predictive models and validate their reproducibility across different regions. Additionally, improving frequency content, data collection techniques, and processing workflows, while monitoring signal-to-noise ratios, will enhance the resolution of detected transitions.

The methodologies developed in this study contribute to the advancement of seismic data as a tool for hydrologic boundary investigations. By addressing the limitations of current models and expanding their scope, this approach has the potential to support sustainable management of submarine freshwater resources in coastal and offshore regions, ensuring their viability for future generations.

References

[1] The Math Doctors. Distances on earth: The haversine formula, 2024. URL: https://www.themathdoctors.org/distances-on-earth-2-the-haversine-formula/.

[2] Inc. Fondriest Environmental. Conductivity, salinity & total dissolved solids. https://www.fondriest.com/environmental-measurements/parameters/water-quality/conductivity-salinity-tds/, 2023. Accessed: 2024-11-30.

[3] Chloe Gustafson, Kerry Key, and Rob L. Evans. Aquifer systems extending far offshore on the u.s. atlantic margin. *Scientific Reports*, 9:8701, 2019. URL: https://doi.org/10.1038/s41598-019-44611-7.

[4] T. William Lambe and Robert V. Whitman. *Soil Mechanics*. John Wiley & Sons, New York, NY, 1969. URL: https://books.google.com/books?hl=en&lr=&id=oRLcDwAAQBAJ&oi=fnd&pg=PA1&ots=rLw8Z-oifg&sig=ZYS8K39jQ-9fxe2tlMoN4JZkpsU#v=onepage&q&f=false.

[5] Berndt Christian Bertoni Claudia Micallef Aaron, Person Mark and Cohen Denis. Offshore freshened groundwater in continental shelf environments. pages 123–146, 2022. URL: https://link.springer.com/chapter/10.1007/978-3-030-90146-2_5, doi:10.1007/978-3-030-90146-2_5.

[6] University of British Columbia. Seismic reflection notes, 2024. URL: https://www.eoas.ubc.ca/courses/eosc350/content/methods/meth_10d/all.pdf.

[7] University of Colorado Boulder. Hilbert transform and envelope analysis, 2024. Accessed: November 29, 2024. URL: https://www.colorado.edu/physics/phys4810/phys4810_sp18/chapter3/section3-2.html.

[8] American Association of Petroleum Geologists (AAPG). Seismic data: analyzing individual reflectors. https://wiki.aapg.org/Seismic_data%3A_analyzing_individual_reflectors, 2023. Accessed: 2024-11-30.

[9] University of Saskatchewan. Reflection seismic method - lecture notes, 2019. URL: https://seisweb.usask.ca/classes/GEOL335/2019/Lectures/PDF/6a-Reflection_Seismic.pdf.

[10] Scikit-learn. Random forests explained, 2024. Accessed: November 29, 2024. URL: https://scikit-learn.org/stable/modules/ensemble.html#random-forests.

[11] Jacob Siegel, Mark Person, Brandon Dugan, Denis Cohen, Daniel Lizarralde, and Carl Gable. Influence of late pleistocene glaciations on the hydrogeology of the continental shelf offshore massachusetts, usa. *Geochemistry, Geophysics, Geosystems*, 15(12):4651–4670, 2014. URL: https://doi.org/10.1002/2014GC005569.

[12] Statology. Pearson correlation coefficient: Definition and calculation, 2024. Accessed: November 29, 2024. URL: https://www.statology.org/pearson-correlation-coefficient/.

[13] Stanford University. Clustering with k-means: A practical guide, 2024. Accessed: November 29, 2024. URL: https://cs.stanford.edu/people/ang/clustering.html.

[14] USGS. How much water is used by people in the united states? URL: https://www.usgs.gov/faqs/how-much-water-used-people-united-states.

[15] USGS. New york city water consumption summary, 2022. URL: https://webapps.usgs.gov/odrm/documents/tables/NYC_Consumption_Summary_2022.pdf.

6 Appendix

The code utilized in this project, including the scripts for data processing and overall seismic analysis, can be found in the following GitHub repository:

```
https://github.com/seth775/Freshwater_Saltwater_Boundaries.git
```