UNIVERSITY OF OKLAHOMA GRADUATE COLLEGE

IT IS NOT OUR FAULT YET: MULTI-ATTRIBUTE AND MACHINE LEARNING STUDY FOR IMPROVED UPPER BASEMENT FAULT DETECTION IN AN AREA OF CARBON CAPTURE UTILIZATION AND STORAGE: DECATUR, ILLINOIS, USA

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By Hy Gia Tran

Norman, Oklahoma

2025

IT IS NOT OUR FAULT YET: MULTI-ATTRIBUTE AND MACHINE LEARNING STUDY FOR IMPROVED UPPER BASEMENT FAULT DETECTION IN AN AREA OF CARBON CAPTURE UTILIZATION AND STORAGE: DECATUR, ILLINOIS, USA

A THESIS APPROVED FOR THE

SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Heather Bedle, Chair

Dr. Matthew Pranter

Dr. Hao Hu

To those who constantly try

to learn and grow from their faults

ACKNOWLEDGEMENTS

Firstly, I want to express my immense gratitude to my wonderful advisor Dr. Heather Bedle. Her energy, dedication, and passion for her craft and students are the reason why I wanted to come to the University of Oklahoma in the first place. Her commitment to making her students feel welcome in a foreign place is something I will never forget. Her mentorship provided me with wonderful opportunities and experiences— professionally and academically.

To Dr. Matthew Pranter and Dr. Hao Hu, thank you from the bottom of my heart for serving on my committee. I appreciate the mentorship and guidance from Dr. Pranter coming from both a Teaching Assistant perspective and as a student.

To the School of Geosciences at the University of Oklahoma, the Mewbourne College of Earth and Energy, the Graduate College, the Society of Exploration Geophysicists, and the Association of Southeast Asian Nations. Not only did they provide scholarships, awards, and waivers, but they also helped ease my financial burdens. They also gave me the tools to grow professionally and meet an immensely diverse community of people with their own beautiful stories.

To the friends I have made, Issa Al Aamri, Montasir Akif, Marcus Maas, Elayne Maas, Dr. David Lubo, Danial Mansourian, April Moreno-Ward, Manoj Thapa, Rui Zhai, Grace Barber, Gabriel Beltran, Jared Brown, Faith Grayson, Morgan Levrets, Hilmi Putra, Ruhi Sahin, Yasin Uzum, Alexandro Vera, Jose Viteri, and many many more. Thank you all for the wonderful adventures and experiences, whether it be working in 1010 on late nights or little occasions together.

And last but not least, I owe everything to my family. My parents and grandma, who have worked and sacrificed everything for me to get a good education. To my sister and cousin, who are starting or about to start their college journeys. To my partner, who has been there for me regardless of anything. I hope I have made you all proud.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Identifying and interpreting faults is critical in many geological prospects, from traditional well planning and operations to energy transition ventures such as Geothermal and carbon capture, utilization, and storage (CCUS). Effective geohazard analysis can reduce the operation time, cost, and uncertainty. More importantly, with new technologies such as CCUS, where public perception and support are sensitive, it is important to decrease analysis uncertainty and prevent large-scale reactivation events.

To improve the accuracy of fracture delineation, a machine learning (ML) and multi-attribute approach was employed in Decatur, Illinois— where microseismicity has been induced in the rhyolitic basement related to CCUS. Due to the poor seismic imaging resolution of the basement and the potential presence of sub-seismic faults, traditional geometric attributes (e.g., coherence and curvature) are not sufficient alone. Structure-oriented filter (SOF) is utilized before the application of the multi-attributes which resulted in increased fault confidence and connectivity. Seismic attributes candidates are gray level co-occurrence matrix entropy (GLCM), fault enhancement of energy ratio similarity (ERS), most positive curvature (k1), most negative curvature (k2), and aberrancy. For ML, a pre-trained convolutional neural network (CNN) was utilized while generative topographic mapping (GTM) was calculated using the aforementioned candidate attributes. Our findings show ERS and CNN can effectively map larger-scale fault patterns where curvature and aberrancy are better for faults seen as flexures or folds. However, only some faults detected from ML and attributes coincided with the location of microseismic events. This is most likely due to a combination of factors such as the poor quality of seismic image at the basement, high viscoelastic attenuations at deep depths, and small vertical displacement of sub-seismic faults.

IT IS NOT OUR FAULT YET: MULTI-ATTRIBUTE AND MACHINE LEARNING STUDY FOR IMPROVED UPPER BASEMENT FAULT DETECTION IN AN AREA OF CARBON CAPTURE UTILIZATION AND STORAGE: DECATUR, ILLINOIS, USA

INTRODUCTION

The characterization of subsurface faults is an essential component in many geologic projects. Knowing where faults are not only means understanding and identifying hydrocarbon traps and migration, but it also means preventing fluid leakage (heat, CO2, wastewater, mud), economic loss, groundwater contamination, and induced seismicity.

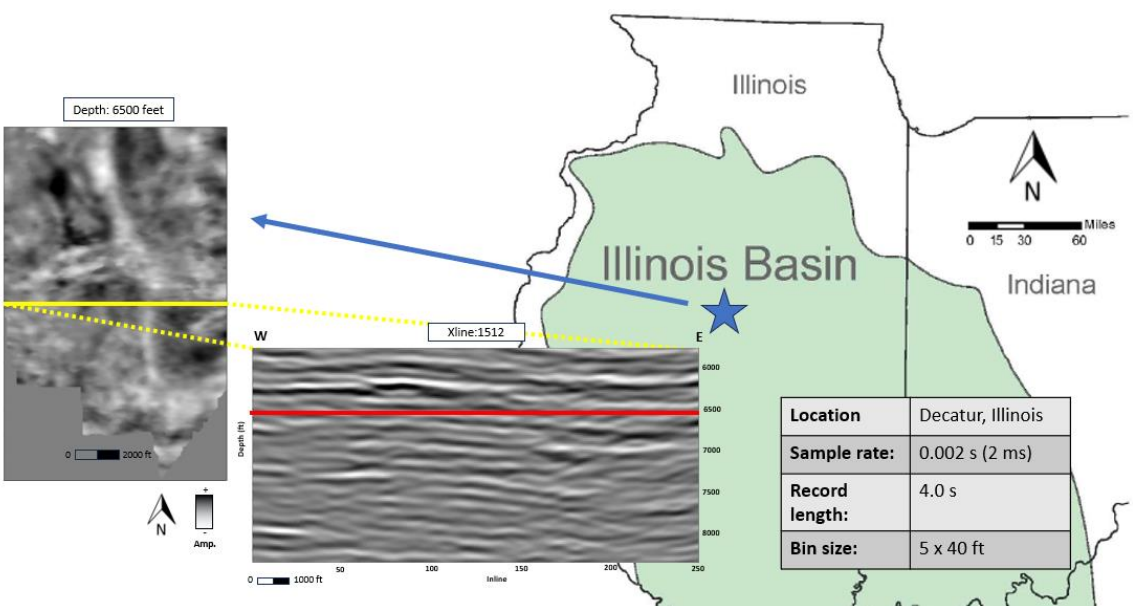


Figure 1. Illinois Basin seismic amplitude volume at 6500 feet depth slice. The crossline displayed is at 1512 with the red line indicating the depth slice of 6500. The incorporated table illustrates the sample rate of 2 milliseconds, record length of 4.0 seconds, and bin size of 5 x 40 feet.

This study focuses on the Illinois Basin-Decatur Project (IBDP), a CCUS site operated by Archer-Daniels-Midland (ADM). According to a 2024 report by the Congressional Research Service, the Illinois Basin is one of four Energy Protection Agency (EPA) federal-level class VI wells in the United States (Figure 1), meaning these wells are approved to inject carbon dioxide into the Earth for storage and sequestration. The IBDP's injection target is the lower Mt. Simon reservoir, an upper Cambrian sandstone, overlaying an unconformable Precambrian basement (Willman et al., 1975). From November 2011 to 2014, the IBDP successfully injected 1.1 million tons of CO2 into their first well. As of late October 2024, it has injected an additional 3.4 million tons and the project is expected to continue for a few more years.

However, approximately 2 months after the initial injection in the first carbon capture and storage well (CCS1), microseismic events were detected by subsurface borehole arrays and processed by surface data acquisition units (NETL, 2024; Babarinde et al., 2021). Then later additional microseismic events were detected after the injection in carbon capture and storage well 2 (CCS2). Microseismic events are defined as very weak earthquakes with a magnitude of 2 or less (Kamei et al., 2015). The magnitudes detected range from -3 to upwards of 1.147 magnitudes. In the Illinois Basin, these induced seismicity events formed long cluster patterns in the Mt. Simon sandstone, Argenta, and upper Precambrian basement. After new analysis and updated horizon interpretations of reprocessed seismic data, more than 80 to 90% of the events occurred within the rhyolitic basement (Bauer et al., 2019). This suggests that preexisting weaknesses such as faults and fractures were present, and that larger-scale reactivation may be possible. Despite the fears and worries, it is also important to note that no large reactivation or any that is noticeable by the public as of 2025.

Fear of induced seismicity is based on the occurrence of large-magnitude events due to wastewater injection. For instance, in 2011 to 2012, Oklahoma experienced a case of induced seismicity, resulting in a 5.6 magnitude earthquake that destroyed 14 buildings and injured 2 people (Song et al., 2023). Being a young technology very few people are familiar with it. Curry et al. (2004) found that almost 95% of people are unfamiliar with CCUS. In contrast, a 2023 survey conducted at the University of Oklahoma found this number to drop to about 34.8% in the United States, making CCUS's unfamiliarity higher than other energy sources such as solar, wind, and nuclear (Bedle et al., 2024). More specifically, the IBDP is situated about 5 miles south of the Mahomet aquifer which provides over 840,000 people with water. If reactivation leads to the contamination of surface and groundwater the public may be more mistrustful of this new technology. Therefore, to prevent harm to CCUS's public perception and the continued governmental support, this study proposes potential methods to derisk and minimize uncertainty to prevent the risk of large-scale reactivation. We aim to do this by identifying potential preexisting faults and fractures utilizing a combination of multi-attribute analysis, machine learning, and overlaying them with microseismic recordings. These fault enhancement methods could also help with derisking other areas of interest, from traditional drilling to wastewater injection and geothermal. For CCUS, identifying basement faults means avoiding potential issues such large-scale reactivation, air contamination, surface and groundwater contamination, public mistrust, and decrease government funding.

DATA

Seismic data

The IBDP high-resolution 3-D seismic survey was acquired in January 2015 and reprocessed in August 2019. This was done to improve the resolution needed to identify shallower reservoir faults. The seismic volume has a bin size of 5 x 40 ft, a record length of 4 seconds, and a sample increment of 2 milliseconds. The small bin may potentially cause issues such as the blurred stair-stepping pattern seen in the seismic. Initial observations of seismic data (Figure 1) indicate the data is very blurred and lacks detail. This could be caused by a variety of factors such as high attenuations of seismic waves associated with deep depths. It is difficult to discern between stratigraphic and structural features on the depth slice. However, there are some indications of discontinuous reflectors and abrupt changes in dip within the crossline indicative of faults. The nature of high-resolution seismic data means high-frequency waves cannot effectively penetrate the basement. Meaning, high resolution seismic is great in shallower intervals but gets progressively worse with depth. Additionally, the basement geology does not help. Igneous rocks shield and absorb seismic waves, leading to a blurred seismic response making it difficult to image faults. Normally, the high density of igneous rocks would lead to a strong basement reflector (SBR). However, despite being reprocessed, our area of interest (AOI) lacks a SBR making it difficult to image map and track faults (Williams-Stroud et al., 2020). The presence of a large basement reflector requires large acoustic impedance contrasts, which means the rock layer above and the basement must have large density differences. The IBDP's basement is comprised of granite and rhyolite which is still denser than the overlaying sandstone. However, compared to counterpart basement types like gneiss-migmatite, granitic-rhyolite is comparatively less dense.

Microseismic data

The IBDP microseismic data was acquired utilizing a network of multiple downhole subsurface arrays deployed in 3 wells each for CCS1 and CCS2, respectively— the injection wells (CCS1 and CCS2), the Geophysical monitoring well (GM1 and GM2), and a verification well (VW1 and VW2, Figure 2). On the surface, three data recording and processing stations were present. For the CCS and GM wells, data was recorded using data acquisition units (DAUs) at a field office while data from VW was recorded at the wellhead (Will et al., 2015). All data is integrated in real-time using a microseismic data acquisition system and software. There are also an additional five subsurface arrays installed by the Illinois State Geological Survey seismometer stations. The microseismic responses for the first round of injections in CCS1 were first recorded 2 months after injection and appeared as linear clusters. These responses started in 2011 and ended in 2015, coinciding with the completion of carbon injection into CCS1. The second round of microseismic responses occurred after injection into CCS2— lasting from 2015 to 2018.

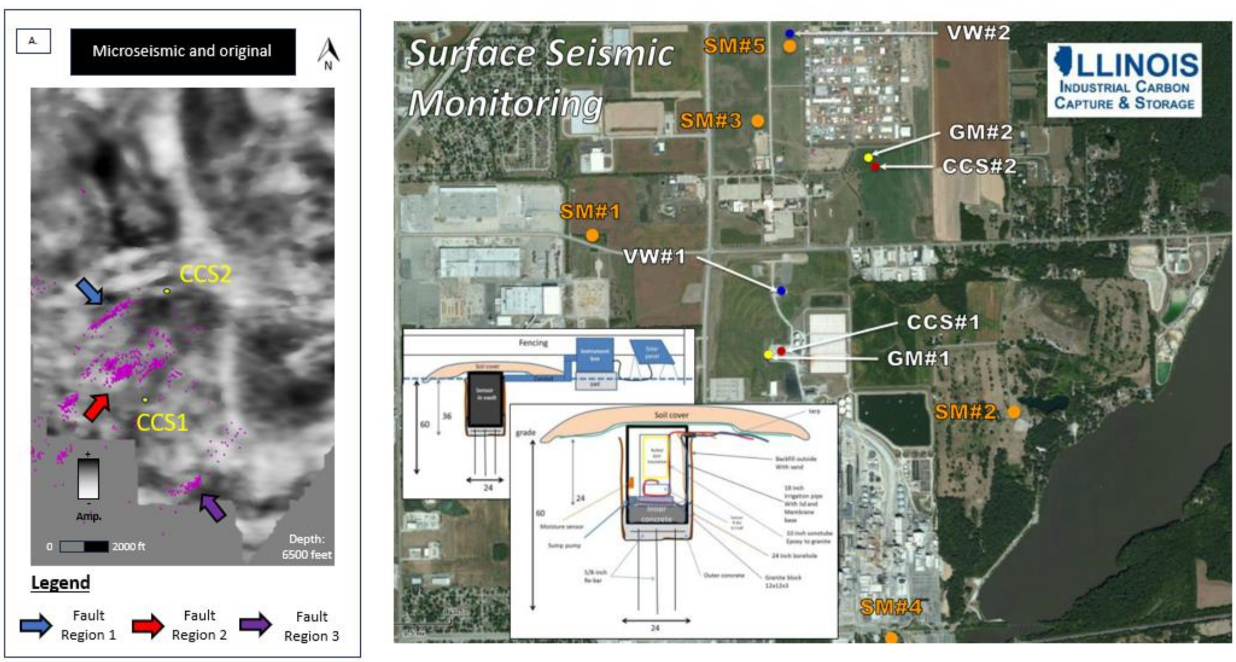


Figure 2. left) Microseismic clusters and CCS injection sites 1 and 2, overlayed with seismic amplitude of depth slice of 6500 feet. Right) Surface seismic monitoring array locations, along with sites of CCS injections.

GEOLOGICAL SETTING

Understanding the geological context means understanding the uncertainty that comes with a new AOI. Within the context of the IBDP we are focused on the Illinois Basin, a large cratonic feature that extends through south-central portion of Illinois, Indiana, and Kentucky. The Illinois Basin's depositional history can be divided into two major sections: 1) the 7 km thick Paleozoic sedimentary section, ranging from early or middle Cambrian to early Permian, this section is comprised of the Mt. Simon and Argenta, and 2) the Precambrian rhyolitic basement (Collinson et al., 1988). This study focuses on the upper basement region which has a history of faulting, uplift, and folding. The upper basement region as we define it includes the lower Mt. Simon reservoir, parts of the Argenta formation, and the Precambrian rhyolitic basement.

The Mt. Simon is an approximately 457-meter sandstone broken up into three sections (upper, middle, lower) and five units (A to E) based on the environment of deposition (Leetaru and Freiburg et al., 2014, 2009; Robert Will et al., 2014). The target reservoir for CO2 injection is the lower Mt. Simon Unit A. This high-quality reservoir consists of fluvial fine to coarse arkosic sandstone and conglomerates. Overlaying the lower Mt. Simon is the pre-Mt. Simon or the Argenta in some sources (Freiburg et al., 2014). The Argenta formation is a low porosity and permeability marine deposit. Both geologic formations also have sealing formations for the prevention of CO2 migration. Finally, underlying them all is the Precambrian rhyolitic basement which is comprised of a thin rhyolite surface, underlain by granodiorite and granite (Robert Will et al., 2014). As previously mentioned, this basement composition is likely why there is not a SBR. In short, low-density differences between sandstone layer above (Mt. Simon) and below (basement), result in low acoustic impedance contrasts.

Precambrian rhyolitic basement (PCB) is a heterogeneous complex that forms part of the Granitic-Rhyolite province (McBride et al., 2010). The late Precambrian to early Cambrian rifting caused by the breakup of the Rodinia is proposed to be the PCB's origin (Freiburg et al., 2020). The basement band extends from Mexico to Quebec and is exposed in the St. Francois Mountains in Missouri. The rocks are dated to be about 1480-1460 million years old (Thomas et al., 2012), which indicates the basement unconformity under Decatur may have almost 1 billion years of missing time. Regional and local drill-hole data from a study from McBride et al. (2010) have supported this claim.

Again, the basement is where about 80 to 90% of detected microseismic events are located, with magnitudes from approximately -3 to the largest being around 1.16 and 1.147 (Bauer et al., 2019). Moreover, natural seismicity is prevalent in the southern band of the basin potentially caused by Proterozoic rift structures such as the Reelfoot Rift-rough Creek Graben systems (Tuttle et al., 2002; McBride et al., 2007). After the acquisition of reprocessed seismic data, the Decatur basement was reevaluated to be approximately 2054 meters or 6243 to 6738 feet in depth (Williams-Stroud et al., 2020). Based on these findings, we analyzed attributes along these depths. It is important to note that seismic signals may not be able to effectively penetrate and detail the basement. This in combination with the igneous properties of the PCB makes it extremely difficult to image and risk.

METHODOLOGY

Orientation within a new dataset is a foundational component of any new project. Our study focuses on the upper basement region which we define as the lower Mt. Simon, Argenta, and Top of the basement. This means mapping our target horizon and cropping our data to approximately 5600 feet to 8000 feet. Access to microseismic data allows us to quality-check our crop to include all microseismic values within the survey. The IDBP microseismic data allows us to validate the usefulness of various attributes and machine learning (ML) outputs.

In noisy or low-frequency data the post-stack conditioning step is essential to remove noise. Interpreters have access to a wide variety of filters such as spectral balancing (Marfurt, 2018; Li et al., 2021), acquisition footprint removal filter (Falconer and Marfurt, 2008), and structure-oriented filter (Luo et al., 2002; Fehmers and Hocker, 2003). Due to its smoothing and edge preservation and success in previous structure enhancement work, we choose to process our data through one-time structure-oriented filter (Hale, 2009; Marfurt, 2018; Blanco et al., 2024; Florez and Bedle, 2024).

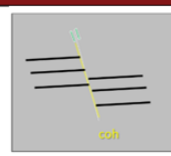
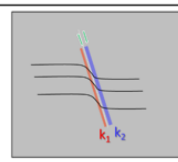
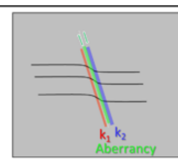
Attribute	Literature	Importance?
outer product similarity	Bahorich and Farmer, 1995; Marfurt et al. 1998 	Measure the similarity of neighboring traces in the inline and crossline direction within an analysis window and considers dip
Sobel filter		
energy ratio similarity		
most positive curvature (k1)	Chopra and Marfurt, 2011 	Map subtle sub-seismic faults that lack a clear reflector, rather seen as a fold.
most negative curvature (k2)		
aberrancy max magnitude	Gao et al. 2013 	Derivative of curvature. Optimal for faults with throw below seismic resolution
aberrancy max azimuth		
GLCM Entropy	Gao et al. 2009	A texture attribute that measures of chaos or disorder. Ideal for ML input.

Table 1: Seismic attributes, the importance of each, and literature overview. Each attribute was selected based on the potential qualities they contribute to the interpreter or machine learning algorithms. For example, aberrancy mitigates the pitfall of similarity by highlighting potential faults and features below seismic resolution.

Then we calculate a suite of geometric attributes: coherence, curvature, and total aberrancy (Table 1). For the coherence workflow branch, we chose to refine our coherence results and compute fault enhancement (Figure 3). Fault enhancement improves local planar features calculated by coherence attributes. After this, fault skeletonization is employed which is based on the Artificial Ant algorithm (Randen et al., 2001), edge-detection (Zhang et al., 2014), and the Hale fault construction technique (Wu and Hale, 2015). Ant tracking and its family members allow for the extraction of fault surfaces or networks from previously calculated coherence attributes.

Then, we ran both unsupervised and supervised ML (Table 2). For supervised ML we computed a convolutional neural network (CNN) utilizing our input data and a prebuilt fault training model. A prebuilt model is ideal, especially in the basement, where faults are difficult to visualize or map.

For unsupervised we employed generative topographic mapping (GTM), first introduced by Bishop et al. (1998). In the past, GTM has been a successful technique for both facies characterization and fault enhancement (Roy et. al., 2014, Florez, 2023; Florez and Bedle, 2024, Kumar et al., 2025).

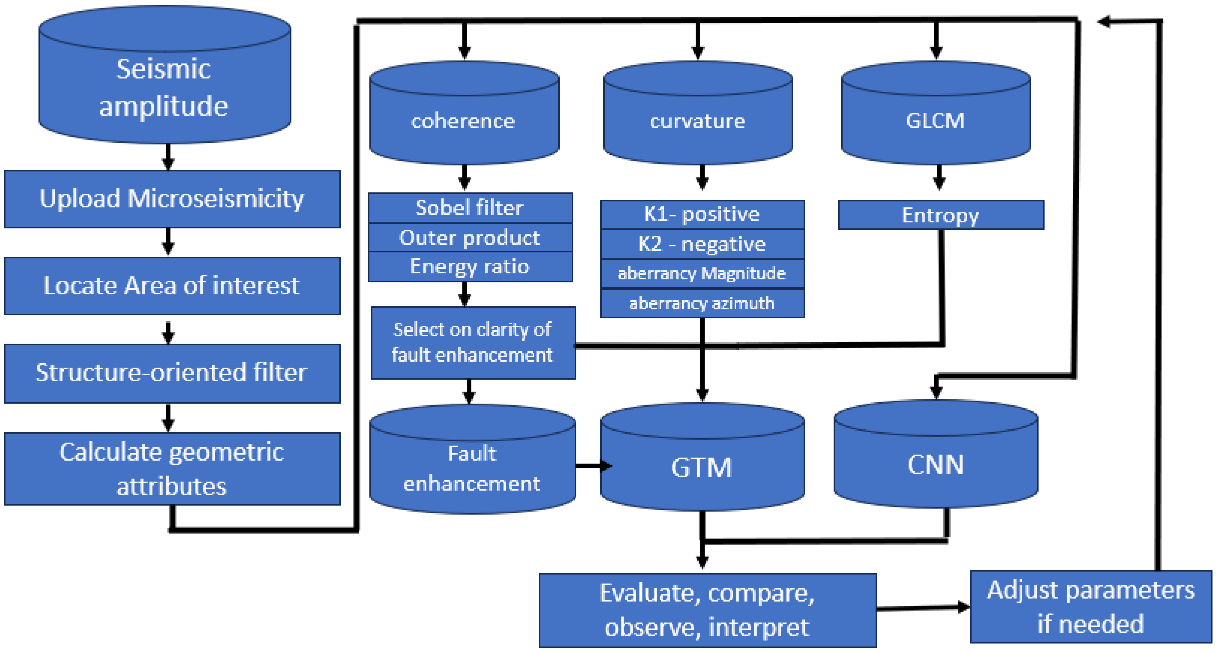


Figure 3: Proposed workflow for fault enhancement and identification. Seismic amplitude is analyzed with additional microseismic data. Then, a target AOI is located. First, a structure-oriented filter is calculated then a multi-attribute suite of coherence, curvature, and GLCM. Finally, machine learning algorithms are tested (GTM, and CNN). Results are evaluated and compared, and parameters are adjusted if needed.

Attribute	Literature	Importance?
Convoluted Neural Network	Wu et al. 2018; Wu et al. 2019; Qi et al. 2019	Supervised ML methods based on training data
Generative topographic mapping (GTM)	Bishop et al. 1998	Non-Linear reduction technique to find patterns between data points
Self-Organizing Maps (SOM)	Hussein et al. 2020; Perico et al. 2021; Qi et al. 2022	Projection technique that allows to cluster features of similar nature
Probabilistic neural network (PNN)	Lubo-Robles et al. 2021; Mora et al. 2022	Better for shallower depths, where faults can be clearly mapped

Utilized   Considered in workflow

Table 2: Supervised and unsupervised machine learning algorithms utilized and considered. Red represents algorithms utilized with literature and importance. Gray are algorithms that were also considered.

Structure-oriented filter (SOF)

A structure-oriented filter is a well-recognized addition to many structure enhancement workflows. This post-stack data conditioning technique preserves signals parallel to structural dip and rejects random and cross-cutting noise to improve the imaging of seismic structures (Fehmers and Hoecker, 2003; Luo et al., 2002; Marfurt et al., 2018). The low resolution and noise attributed to lower frequency wavelengths in the PCB present challenges for fault identification on seismic amplitude data. Furthermore, excess noise may affect geometric attribute calculations and therefore ML results.

Both versions of SOF proposed by Fehmers and Hoecker (2003) and Luo et al. (2006) are integrated. The difference between our SOF and these previous studies is that ours utilizes a computed similarity constraint to measure the smoothness or presence of an edge. Secondly, more can be compared and contrasted. Two SOF filter options were selected, principal component and mean, to determine the optimal input for calculating geometric attributes. After filtering, the outputted noise is analyzed to ensure that the rejected data is acceptable and no valuable data is removed.

Seismic attributes

Geometric attributes are well studied and utilized in both industrial and academic settings and are powerful tools for visualizing and decreasing the uncertainty of faults and other structural features. Based on previous studies, we selected three types of coherence (outer product, Sobel filter, and energy ratio similarity), fault enhancement, fault skeletonization, k1-most positive principal curvature, k2-most negative principal curvature, aberrancy azimuth, and aberrancy magnitude (Marfurt et al., 1998; Chopra and Marfurt, 2011; Gao et al., 2009; Gao et al., 2013; Hussein et al., 2021; Perico et al., 2023; Florez, 2023; Florez and Bedle 2024). All of these attributes are calculated along a structural dip.

Our first group of attributes tested is coherence. Originally developed at Amoco in the 1990s, coherence compares waveforms along structural dips within an analysis window. Windows where data is similar is “coherent” and is traditionally plotted as white, whereas “discordant” areas are plotted as black. These discontinuities are useful when mapping or identifying structures within the seismic such as faults. Fault enhancement is an extension of coherence because it requires a calculated coherence volume as its input. Then, a fault skeletonization calculation can further lineate faults detected. Ideally, these post-coherence steps should improve fault imaging by its predecessor.

Curvature is broken up into two main components, the k1 most positive curvature and the k2 most negative curvature. When analyzed independently, k1 most positive curvature aids in identifying anticlinal or dome structures while k2 most negative curvature aids in identifying synclines and bowls. When co-rendered, the attribute pair potentially helps outline faults that are characterized by folded reflectors on each side and sub-seismic faults (Chopra and Marfurt, 2011).

Aberrancy is the derivative of curvature introduced by Gao et al. (2013). Instead of measuring folds and providing an outline of potential sub-seismic fault zones, aberrancy pinpoints and maps planes of these potential faults. Unlike larger faults characterized by a large vertical offset or discontinuity, sub-seismic faults are subtle with vertical offsets below the seismic resolution. Some studies that have aberrancy to better lineate subtle faults include: Salazar, 2024; Salazar and Bedle, 2024; Kumar et al., 2024.

Entropy belongs to a family of attributes called gray level co-occurrence matrix (GLCM) and is calculated along structural dip. In short, these attributes analyze the "texture" of the seismic amplitude utilizing an analysis window like a finger to "feel" it. Members of GLCM were employed by Gao et al. (2004, 2007, 2009, 2011) within self-organizing maps. Other studies that have used GLCM as it pertains to unsupervised ML include Matos et al. (2011), Yenugu (2010), and Roy et al. (2011). For our purposes, we will be inputting it into the generative topographic mapping algorithm. GTM works similarly to SOM and tends to be a more 'powerful' clustering algorithm. The entropy attribute was selected because it measures disorder or roughness. Therefore, higher readings of entropy are usually associated with faults.

Machine Learning

Four machine learning algorithms were considered for this study, probabilistic neural network (PNN, Lubo-Robles et al., 2021; Mora et al., 2022; convolutional neural network (CNN, Wu et al., 2019; Qi et al., 2020), self-organizing maps (SOM, Hussein et al., 2021; Perico et al., 2021; Qi et al., 2022), and generative topographic mapping (GTM, Bishop et al., 1998). PNN is better for shallower intervals where faults are easily mapped (Table 2). CNN's prebuilt model allows the interpreter to apply it to areas where interpreters may have difficulty visualizing faults, such as the basement. For this reason, CNN was selected over PNN. GTM was selected over SOM due to its improvements over its predecessor. GTM and SOM were employed by Florez (2023) and Florez and Bedle (2024) to better visualize and lineate hidden faults in the Kevin Dome and Oklahoma areas. In those studies, GTM was able to output slightly sharper lineaments than SOM (Florez, 2023).

Generative topographic mapping (GTM)

Generative topographic mapping was introduced by Bishop et al. (1998) and is a non-linear dimension reduction technique, similar to that of Self-organizing maps (SOM), that improves on SOM's shortcomings. GTM takes the relationships and points between inputs (multiple attributes) and distributes them on a 2D space then a N-dimensional space. Within the N-dimension, these points become vectors and are assigned a probability based on the Gaussian probability function. During iteration, EM will be used to accommodate vectors, then the Bayes theorem will be utilized to the final probabilities of the data vectors are calculated and then projected back onto the 2D space (Roy et al., 2013). In this study, because the goal was to characterize and improve the detection of faults, the geometric attributes: Coherence (energy ratio similarity), k1 most positive curvature, k2 most positive, entropy, and aberrancy magnitude were selected. These geometric attributes were chosen based on three factors: how well they identify or enhance faults, how each attribute functions mathematically, and previous literature (Marfurt et al., 1998; Chopra and Marfurt, 2011; Gao et al., 2009; Gao et al., 2013; Hussein et al., 2021; Perico et al., 2023; Florez, 2023; Florez and Bedle 2024). How effective an attribute performs may be different for each AOI due to varying data quality and subject to the geoscientist's interpretation. Therefore, it is crucial for other factors to be considered. For example, an attribute like entropy may not be the most optimal result for the interpreter, but mathematically it makes sense to include it within GTM to maximize results. Attribute selection for GTM calculation will be discussed later.

In short, GTM utilizes models based on Gaussian and expectation maximization (EM) to find correlations, patterns, and probabilities between data points. Based on the performance of GTM over SOM in previous studies, GTM is selected because it is able to better lineate faults (Florez, 2023).

Convolutional Neural Network (CNN)

Mapping faults with seismic data is very time-consuming and is difficult if the data quality is low. Seismic attributes are commonly used to mitigate these pitfalls. Coherence, curvature, and aberrancy all perform very well for improving fault detection and therefore aid in fault mapping. However, these methods can be further complemented with a supervised machine learning approach— CNN. This algorithm is broken up into two components— training and prediction. The training is provided by a pre-established model created/quality controlled by subject matter experts and developed on synthetic seismic data, fault-labeled images, and field data (Wu et al., 2019). A pre-established model is selected because we want to enhance faults in areas where an interpreter may not see a clear fault. Compared to other ML algorithms, CNN has a relatively low computational time and is easy to use if there is a pre-established training model/dataset. CNN's strengths lay in the ability to increase fault connectivity and measure fault probability and therefore minimize uncertainty. However, due to its reliance on pre-established training data, CNN is good at picking faults with large vertical displacements and may not be very good at picking up sub-seismic faults and fractures— unless it is trained to do so.

An argument can be made that CNN's inclusion would lessen the impact of attributes in an attribute-focused workflow. However, due to its low computational time and ease of use, attributes can be used to quality-check CNN results and vice versa. Attributes and machine learning can complement one another. If multiple attributes or methods tell you a similar story independently, the bigger picture becomes clearer. Likewise, if multiple outputs tell you different stories, you are either missing a piece of the puzzle or something is wrong.

RESULTS

Apart from SOF, results are presented in depth slices. In previous fault interpretation research, faults appeared sharper in depth slices compared to horizon slices (Carvajal et al., 2023). The depth slice of 6500 feet was chosen because it is between the Decatur basement reevaluation of 6243 and 6738 feet using the 2019 reprocessed seismic data (Williams-Stroud et al., 2020). A combination of arbitrary lines and crosslines were selected based on microseismic clusters and potential fault zones, for cross-sectional analysis (Figure 4). Fault regions 1-3 (FR1-3) indicated by the blue, red, and purple arrows are selected for microseismic and attribute comparison because they appear to be the main clusters recorded from microseismic. The microseismic clusters in FR1 to FR3 strike in the northeast to southwest direction (Figure 4).

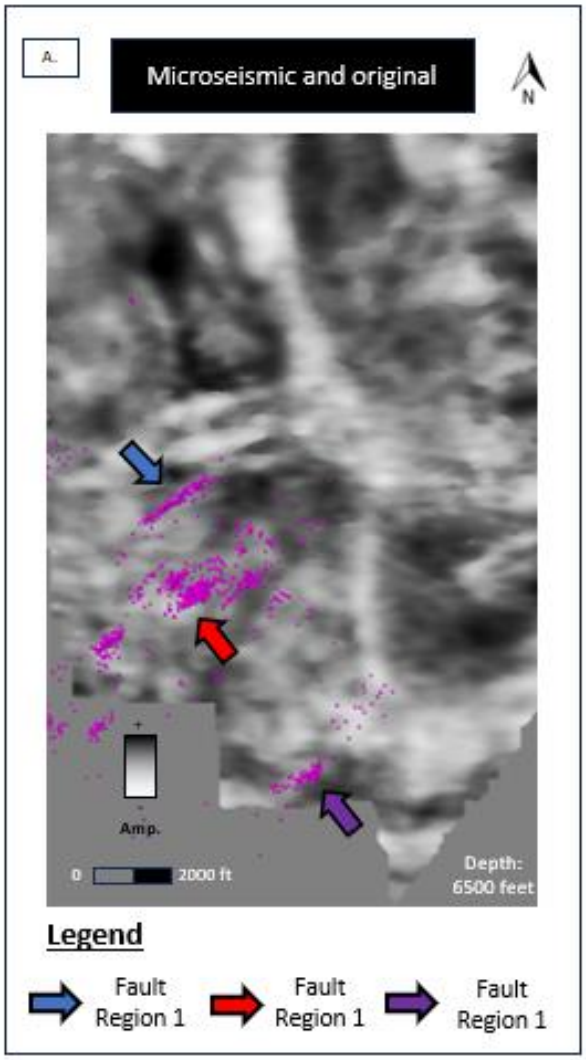


Figure 4. Original seismic amplitude on the depth slice of 6500 feet overlayed with microseismic data. Arrows displayed indicate fault regions (FR1, FR2, FR3) that may potentially be overlayed with microseismic data.

Structure-oriented filter

The SOF results are presented in Figure 5 on crossline 1512. The amplitude discontinuities and offsets created by potential faults are less sharp pre-SOF juxtaposed to post-SOF. Post-SOF, Fault connectivity is increased to deeper intervals and the discontinuity of reflectors is improved compared to pre-SOF (Figure 5-A, 5-B). This is especially evident in areas marked by the green arrow. The noise removed is examined for any potential indicators that may negatively impact attribute calculations and interpretation (Figure 5-C). This is a crucial step as it allows interpreters to gauge the aggressiveness and effectiveness of the filter. Analysis of the noise output revealed important fault-related features are not eliminated. Instead, the removed noise coincided with potential fault zones and increased fault interpretation confidence to deeper intervals. Of course, filters can always be more or less aggressive. However, because our focus is purely on fault enhancement, we are not too concerned about the noise that is removed parallel to amplitude. Additionally, the noise parallel to amplitude appears to be evenly removed throughout the seismic volume so any attribute calculations should not be affected.

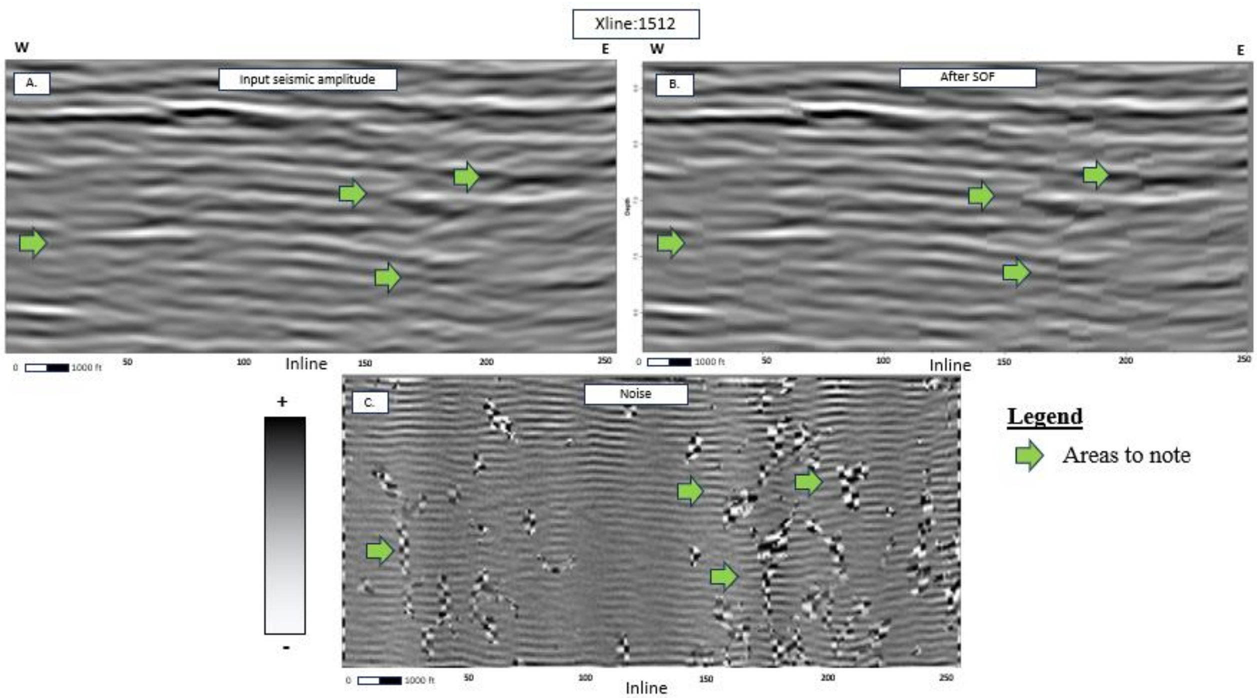


Figure 5. The effects of structure-oriented filter A) Input seismic amplitude before SOF B) After SOF C) Noise SOF removed. Green arrows represent arrows to note and compare with each output. Overall, SOF increases fault connectivity and confidence.

Seismic Attributes

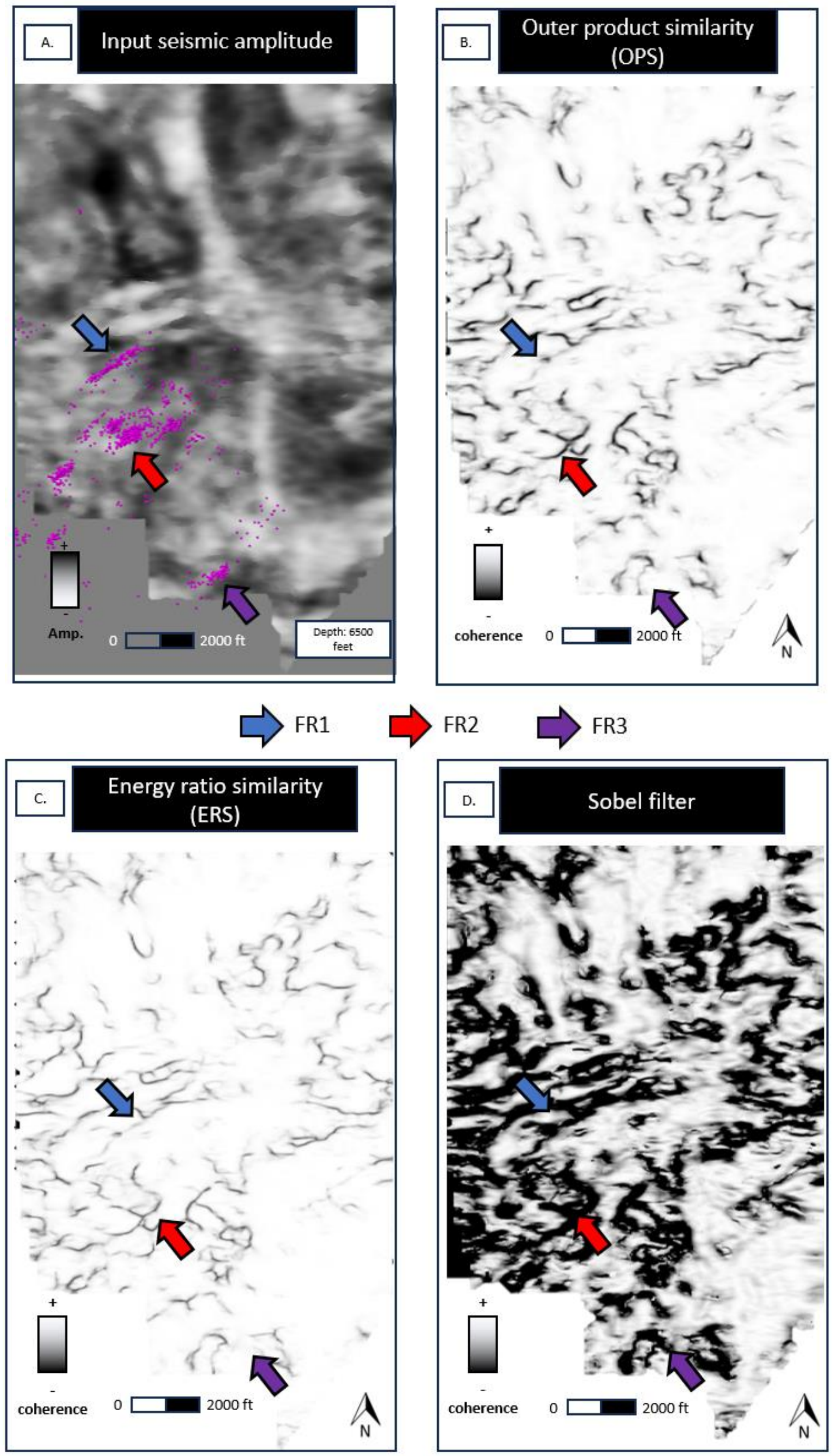


Figure 6: A) Comparison of microseismic overlay with seismic amplitude with microseismic data (purple) B) outer product similarity (OPS) and C) energy ratio similarity (ERS) D) Sobel filter. Blue Arrow represents fault region 1, Red arrow represents fault region 2, and Purple arrow represents fault region 3.

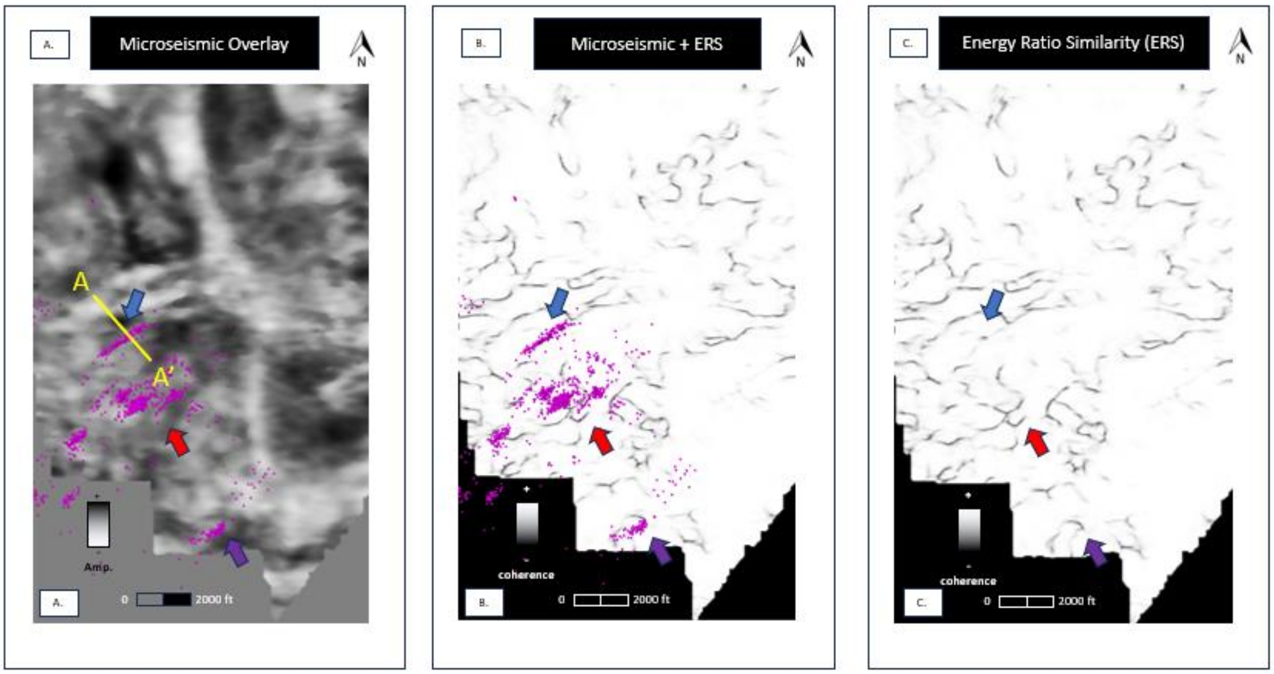


Figure 7: A) Comparison of microseismic overlay with seismic amplitude with arbitrary line A to A' B) Microseismic and energy ratio similarity (ERS) and C) energy ratio similarity. Blue Arrow represents fault region 1, Red arrow represents fault region 2, and the purple arrow represents fault region 3.

The three main coherence outputs are displayed in Figure 6. Energy ratio similarity (ERS) and outer product similarity (OPS) yielded similar results they may be difficult to distinguish. However, high coherence responses in OPS appeared slightly thicker and less sharp in contrast to ERS. Sobel filter coherence responses are more "noisy", chaotic, and less lineated. Instead, they appear in thick-large zones. Sobel filter's response makes it extremely difficult to discern the features on the depth slice. The selection of the ideal coherence attribute for GTM calculation is based on how well that attribute enhances the fault. Attribute selection for ML purposes will be discussed later. Overall, it is difficult to discern whether features are stratigraphic or structural in origin, especially in Sobel filter. The blue arrows in Figures 6 and 7 indicate a potential lineation captured on all three coherence attributes. The overlap of microseismic data and coherence in FR1 further supports a potential lineament associated with faulting (Figure 7). This is further examined in an arbitrary line to confirm if there is an indication of faults (Figure 8). The discontinuous reflectors in the seismic input suggest that there is indeed a fault in FR1. The fault has high vertical displacement at a depth of approximately 6,300 to 6,600 feet then it becomes shallower beyond 6,700 feet. Analysis revealed that the fault overlaps the microseismic response in FR1 in both depth and cross-section slices. For fault regions 2 and 3, there is also some overlap in the microseismic. However, coherence values are not as prominent. Further examination of the fault in FR1 is conducted through the comparison of shallower depth slices. The depth slices range from 6,300 feet to 6,500 feet. This range was chosen to coincide with the upper basement region where microseismic values were recorded, to further visualize the fault in FR1, and show the potential connectivity of this fault at shallower intervals (Figure 9). Lower coherence responses were exhibited at 6,300 feet then gradually increased all the way to 6,460 feet, and then later dimming out at 6,500 feet.

Finally, fault enhancement and skeletonization is the natural extension of coherence that preserves the main signal of the fault. Enhancement was tested and compared on all three coherence types. Again, OPS and ERS had similar results so only OPS's output is displayed. Despite Sobel filter not having the best raw results, the fault enhancement of Sobel filter proved to be the most informative of the overall story and structure of the AOI (10). The thick similarity values that appeared in Sobel filter are now better lineated after fault enhancement (Figure 6,10).

Both fault enhancement products had regions that overlapped with microseismicity, particularly in FR1.

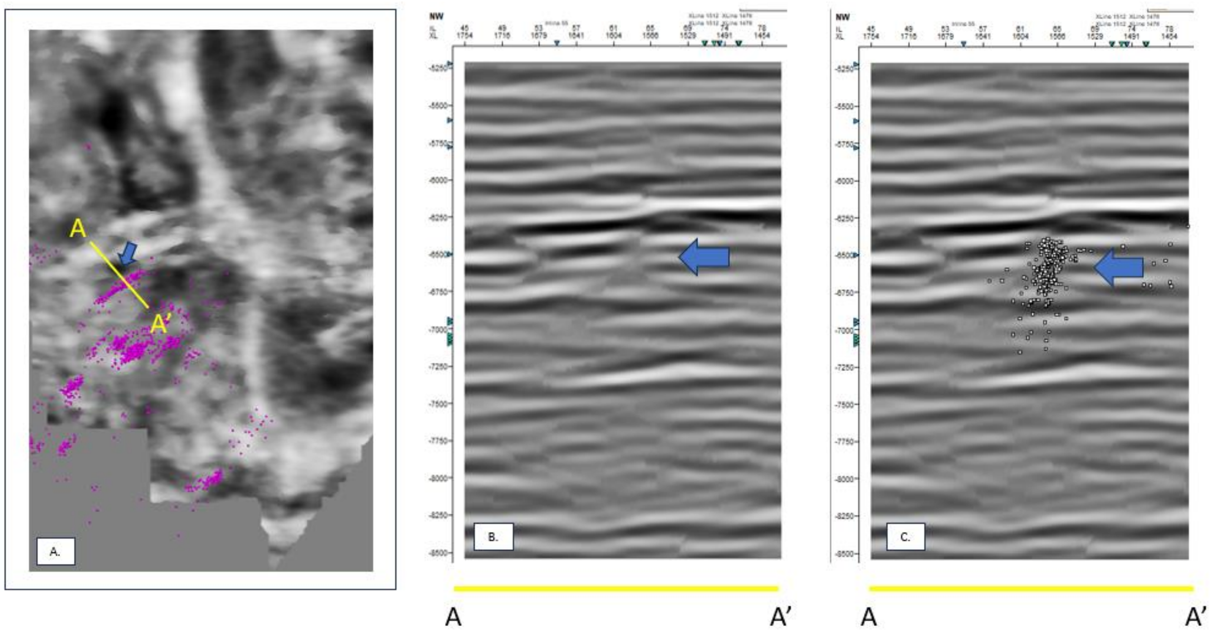


Figure 8: A) Original seismic amplitude with microseismic data overlay, arbitrary line, and fault region 1 (blue arrow) B) Arbitrary line cross-section with no microseismic c) Arbitrary line with microseismic data with a projection 250 feet. Arbitrary line analysis revealed there is a potential fault on A to A' due to evident reflector discontinuity and microseismic overlap.

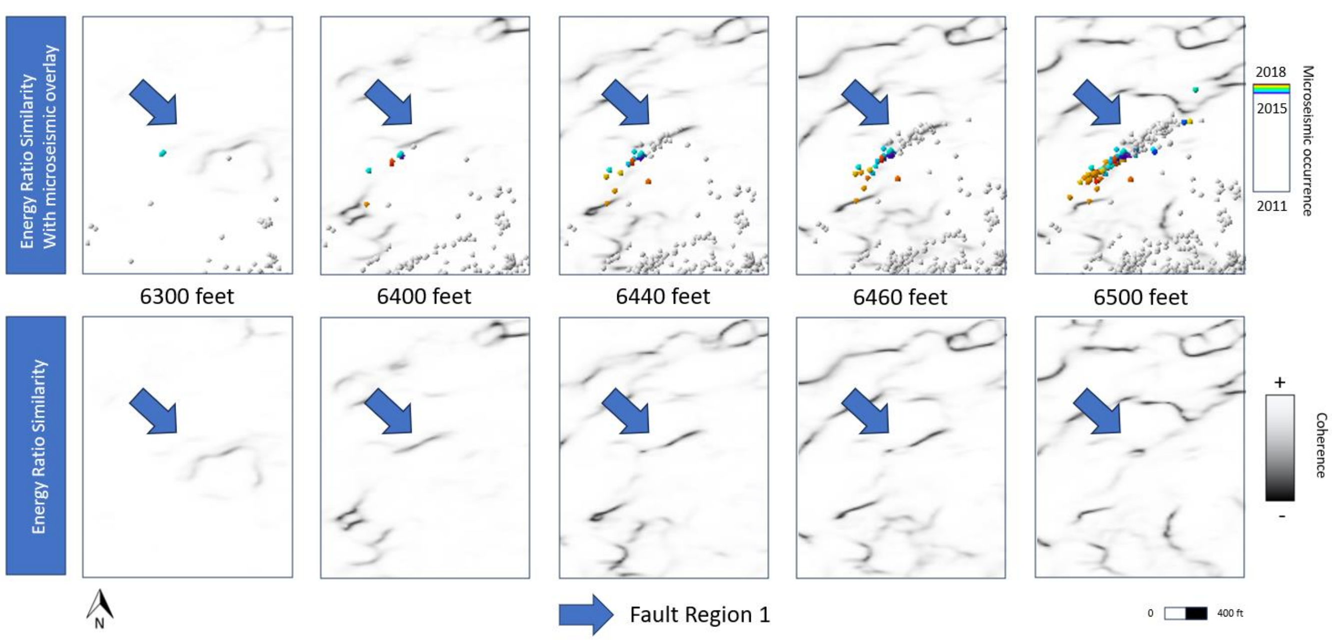


Figure 9: Analysis of Fault Region 1 (FR1) with varying depths from 6300 feet to 6500 feet. The top row shows the overlap between microseismicity recordings and energy ratio similarity, and the bottom row shows only energy ratio similarity. Blue arrow represents Fault Region 1 (FR1) which is marked in previous figures. The lineation in FR1 strongly corresponds to microseismic recordings in the given depth range of 6300 to 6500 feet.

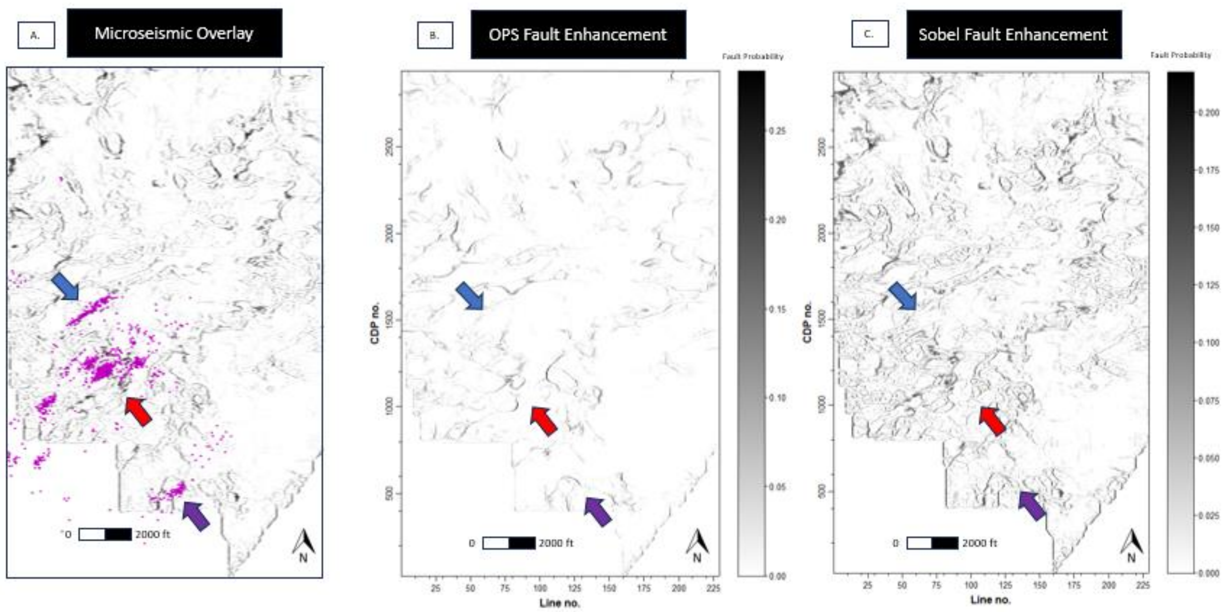


Figure 10: Comparison of different products of fault enhancement A) Microseismic overlaid on top of Sobel filter + fault enhancement B) Outer product similarity + fault enhancement C) Sobel filter + fault enhancement.

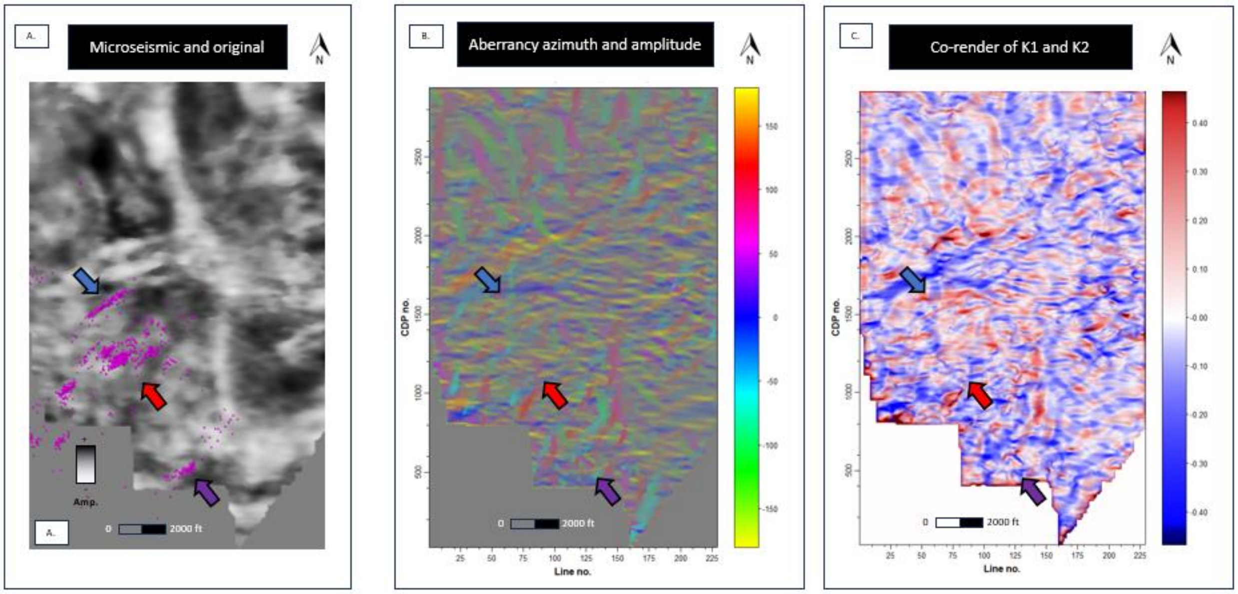


Figure 11: Comparison of microseismic overlay and curvature-based attributes A) Original seismic amplitude with microseismic recordings B) Co-render of aberrancy azimuth and magnitude C) Co-render of most positive curvature (k1) and most negative curvature (k2).

Curvature highlighted areas that are not captured with coherence. Negative and positive curvature has proved effective in previous studies in mapping channel edges and potential fault zones, but it is difficult to interpret without co-rendering (Florez, 2023). The co-render of k1 most positive curvature and k2 most negative curvature is shown in Figure 11. Co-renders help interpreters with the bigger picture as k1 most positive curvature is strongly associated with anticlines and k2 most negative is associated with synclines. When high curvature responses are next to lower curvature responses in areas with no coherence, this suggests there may be low vertical displacement and thus potential faults. It is important to note that curvature correlates with some microseismic recordings even in areas where there is high coherence.

Furthermore, aberrancy, the derivative of curvature, provides additional detail not captured within coherence or even curvature. If curvature provides the channel and fault edges, aberrancy would be the fault or channel itself. Like curvature, aberrancy appears and overlaps with areas of high microseismicity. The main areas attributed with microseismic have an approximate aberrancy azimuth of 180, which results in the yellow color. Beyond the three FRs, both curvature and aberrancy outputs completely paint the seismic data with color. In contrast, the similarity suite only highlights regions near the FRs.

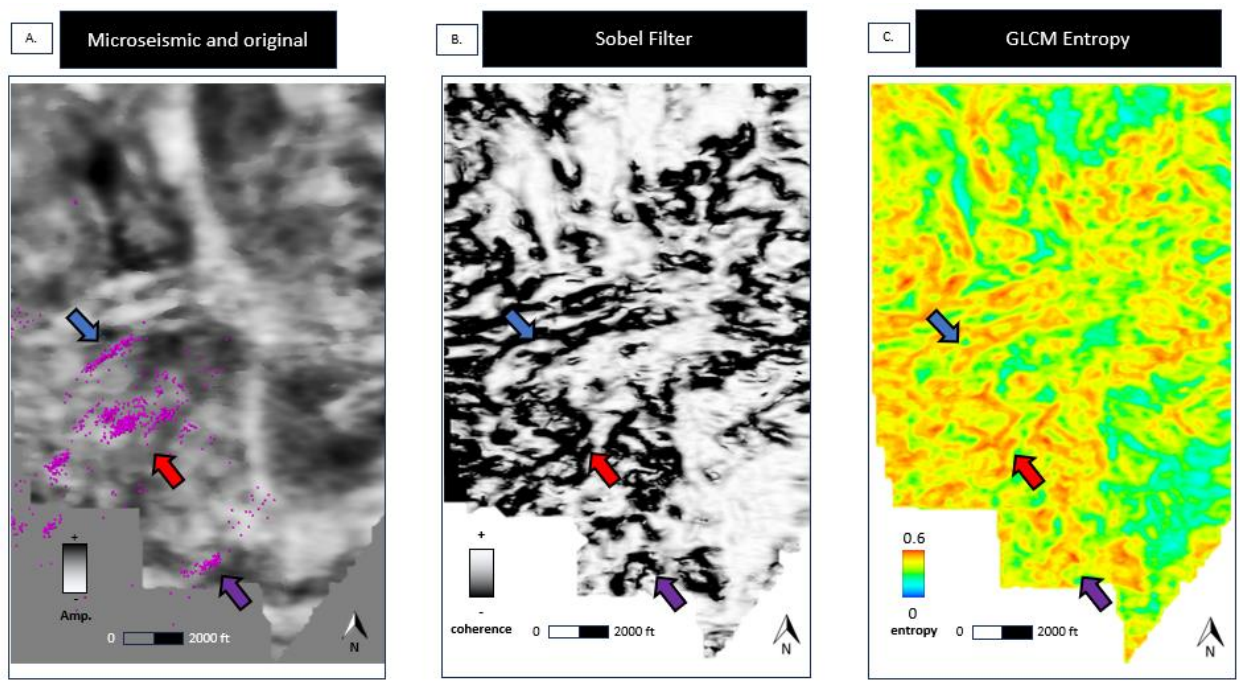


Figure 12: Comparison of microseismic overlay and curvature-based attributes A) Original seismic amplitude with microseismic recordings B) Sobel filter C) GLCM Entropy with color scale range set to 0.6 to 0.

Entropy highlights areas with a high disorder or chaos— indicated by colors mapped in orange and red (Figure 12). Interestingly, entropy exhibits similar patterns to Sobel filter with areas of low coherence overlapping with high entropy regions. Even though entropy and coherence do not directly correlate to faults, this overlapping relationship indicates there is a correlation between high disorder and low coherence. This makes sense given that faults tend to have a higher entropy and lower coherence. Visually, entropy does not bring much improvement as it pertains to FR1-FR3 in Figure 12, compared to coherence or curvature (Figures 7,10,11).

Moreover, there are also no new major lineaments or new areas of note that have not already shown up on curvature and similarity, respectively. Regardless, the additional measure of disorder from entropy may help improve GTM results, making it a better input rather than an attribute used for direct visual interpretation.

Machine Learning

Convolutional Neural Network (CNN)

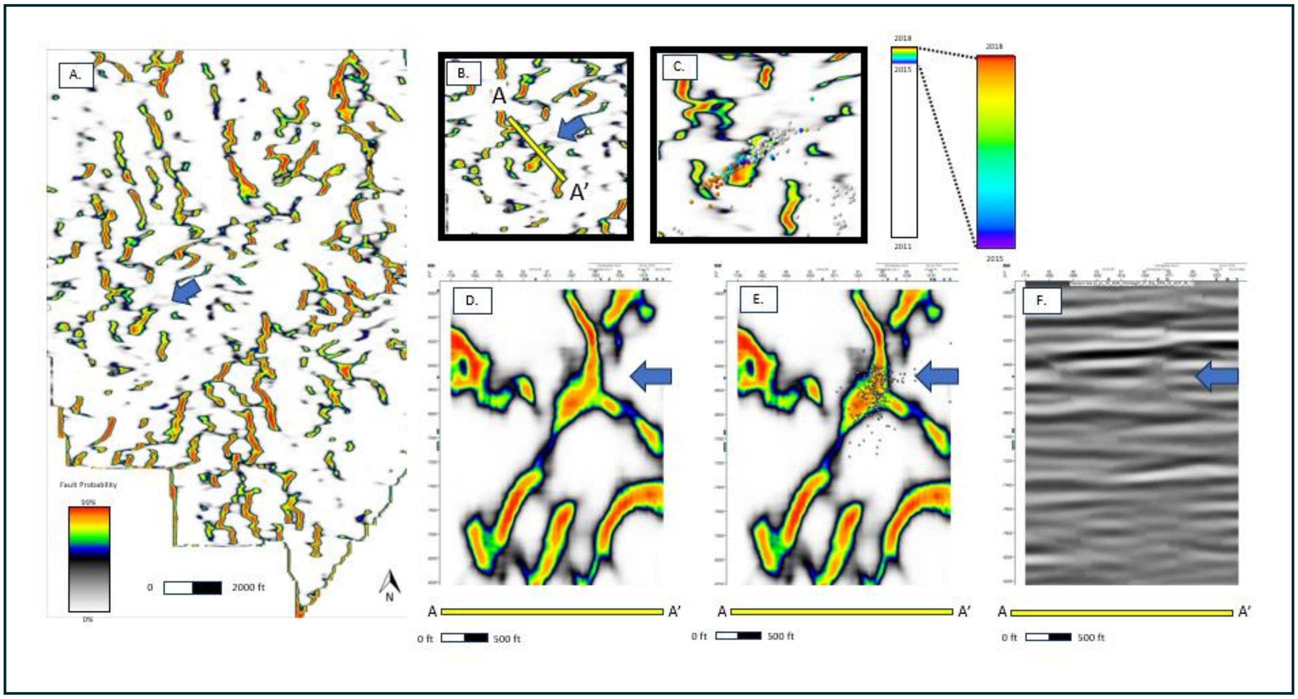


Figure 13: Convolutional Neural Network results A) CNN B) Magnified view of arbitrary line A to A' C) Microseismic overlay with time index D) Arbitrary line cross-section E) Arbitrary line cross-section with microseismic overlay F) Original seismic amplitude

Although multi-attribute results and calculations are independent of CNN, both workflows yielded similar results. CNN provides the broader story because it directly measures fault probability and therefore potential risk zones, while attributes provide finer detail for quality assurance (Figure 13). Traditionally, CNN is displayed in a black-white color scale with white being low fault probability and black being high fault probability. In our case, the color scale "interleave" was selected to bring more detail and complement fault probability. Fault probability is outputted in a rainbow with red being the highest chance of faults, followed by a black-white color scale for probabilities below 50%.

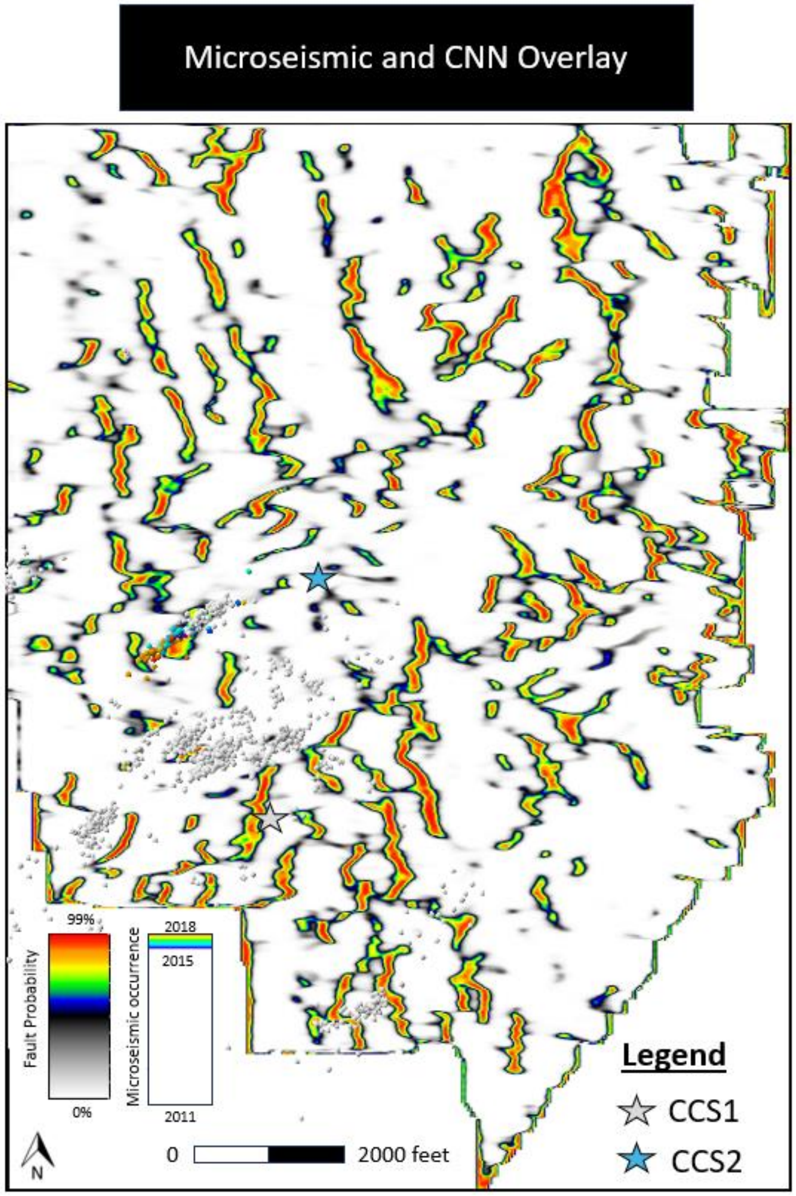


Figure 14: Overview of Convolutional Neural Network with microseismic overlay. Stars indicate where CCS injection sites are. The microseismic occurrence color scale is based on the timing of microseismic events. Microseismic points in white occurred from 2011 to 2015 whereas points in rainbow occurred from 2015 to 2018.

An examination of the arbitrary line in Figure 8, compared to its CNN counterpart in Figure 13 part D-F, revealed CNN also detected faults that are supported by analysis of the original seismic amplitude. Microseismic data with a maximum projection of 250 feet was overlayed on top of our products to evaluate if these clusters support the ML calculations. Like the attribute results in Figures 7,8 and 11, microseismic recordings overlapped with CNN. Interestingly, about half of the overlapping clusters in FR1 occurred after the initial injection and coincided with the second wave of CCUS (Figure 13-C). An overview of CNN overlayed with the microseismic data reveals that most events happened from 2011 to 2015, coinciding with injection into CCS1 (Figure 14). This makes sense given that CCS1 occurred for a longer period. In contrast, events triggered by CCS2 only occurred near FR1 and strikes southwest— indicating reactivation in this direction.

Generative topographic mapping (GTM)

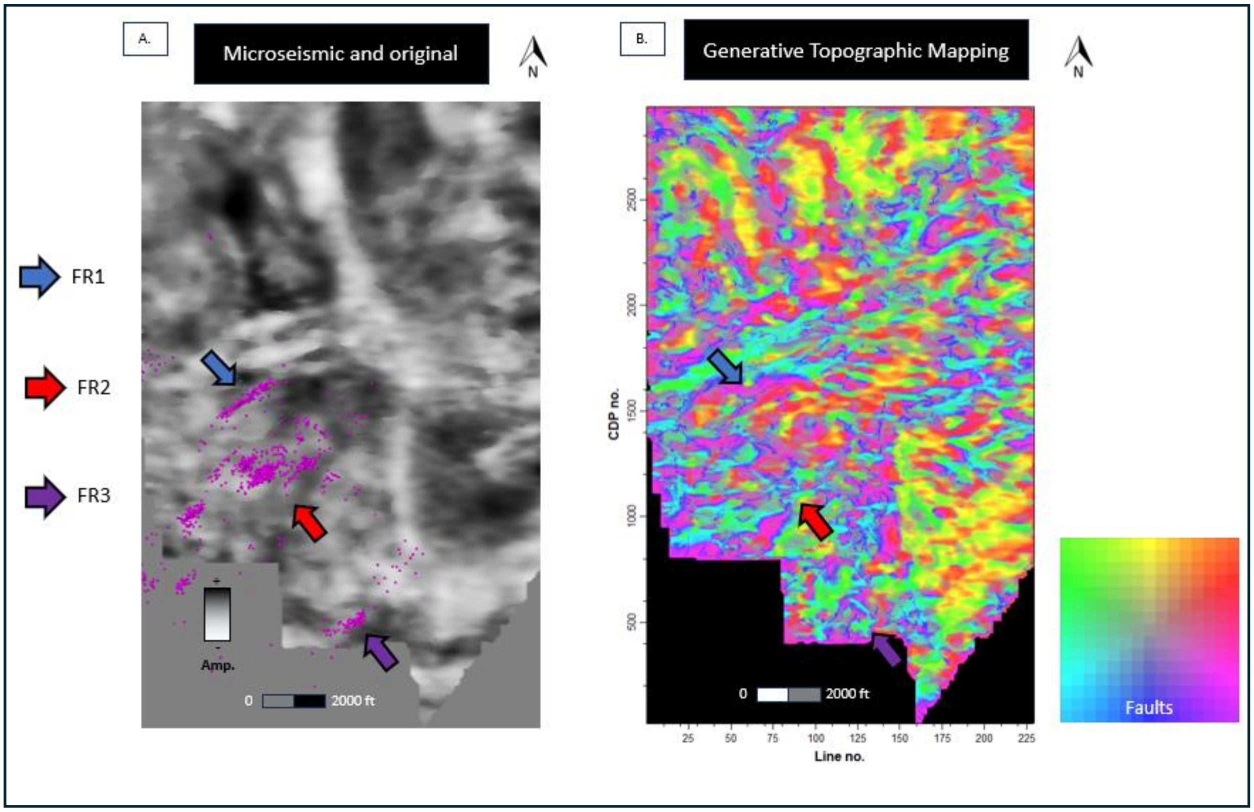


Figure 15: Generative topographic mapping A) Original seismic amplitude with microseismic recordings B) Generative Topographic mapping results using aberrancy total, k1 most positive curvature, k2 most negative curvature, entropy, and fault enhancement as inputs.

GTM yielded similar results to the previous sections (Figure 15). Clusters are assigned varying color values based on the relationship between attributes. The area in dark blue correlates to the fault enhancement while the other areas correspond to the other inputs (mainly k1, k2, and aberrancy). It is hard to discern where entropy is contributing because those results were difficult for an interpreter to identify faults. There is a slightly sharper and more connected lineament for FR1 in the GTM. In FR2 and 3, respectively, the liniments also appear slightly sharper but the difference is not significant enough to warrant a GTM calculation.

DISCUSSION

The primary motivation is to enhance basement faults to test if they could potentially coincide with microseismic activation due to CO_2 injection, therefore decreasing uncertainty.

These methods and workflows are also highly transferable and foundational to other areas of interest that deal with fault geohazard analysis. From conventional and unconventional drilling projects to wastewater injection and geothermal projects. The main challenge with this dataset is the low quality of seismic for the Precambrian basement. This low quality can be attributed to both the depth and the high-resolution data acquisition. High resolution data is exceptional for shallower intervals where immense amounts of details can be revealed. However, these high frequency wavelengths cannot penetrate deep due to wavelength attenuation caused by the Earth. Therefore, to maximize and condition the available post-migrated data, applying filters such as SOF is an essential component before applying attributes and ML. SOF is a widely used technique in both academia and industry that sharpens faults and edges, thus improving attribute calculations and subsurface interpretation.

Structure-oriented filter

In Figure 5, we observe a dramatic difference between pre-SOF vs post-SOF, where faults are better lineated and connected. Faults are better lineated on the crossline and inline slices, albeit with a "stair-stepping" effect. This effect is most likely due to low basement resolution, bin size, and low data quality, but regardless there is still an improvement. SOF makes fault interpretation easier and increases confidence when tracking faults deeper or shallower. In addition to looking at pure SOF results, the analysis of SOF noise removed is highly crucial in this workflow (Figure 5). Evaluating noise allows the interpreter to see how effective or ineffective the filter was. In our case, SOF noise lined up extremely well with fault offset and discontinuities in the original seismic. This makes sense and increases confidence in our results because SOF tends to target noise around structural elements like faults. In theory, improvements in SOF mean better attribute results. An in-depth comparison of attribute calculations with or without SOF was not explored in this paper and could be a potential future work.

Attributes

Coherence attributes are a foundational component of many multi-attribute workflows. They are exceptional when trying to understand the bigger structural and stratigraphic features of an AOI. This is especially true in AOIs with high data resolution and quality. However, this is not the case with the PCB. Figure 5 illustrates that coherence struggles with distinguishing faults from stratigraphic features like channels. Moreover, it has difficulty in enhancing any faults at all, apart from the region indicated by the blue arrow. The PCB not only suffers from low data resolution caused by attenuation of high frequency waves, but the sub-seismic faults or faults with low vertical displacement present make it difficult for coherence to pinpoint faults. Therefore, to fully maximize coherence it is important to run it through fault enhancement workflow or test other attributes (curvature, aberrancy, etc).

Fault enhancement improves products from its predecessor as it better lineates some features. Enhancement or skeletonization is something to consider when data quality is low or additional steps need to be taken to improve the similarity products. Additionally, it is also computationally inexpensive.

The information from curvature (k1, k2) and aberrancy support the results from coherence and the microseismic recordings (Figure 11). This suggests that there are faults in FR1, 2, and 3 that may be linked to reactivation. Additionally, curvature and aberrancy also provide interpreters with information in areas of high coherence. Without the additional calculation of these two attributes, interpreters may run the risk of missing faults with a low vertical displacement. This is especially important when considering new plays and prospects.

Entropy is better used as an additional input into ML algorithms like GTM rather than an attribute used for interpreter visualization. In Figure 12, we see a strong resemblance between Sobel filter and entropy, where areas of low coherence overlap with areas of high entropy. This makes sense given that entropy and disorder mean less coherence and therefore higher potential for faults. However, there is not a significant visual improvement between the attributes. Entropy does not add any additional lineaments in FR1-FR3, nor does it capture additional features in our AOI.

A Different Perspective

An essential and sometimes overlooked component of seismic interpretation is analyzing the seismic data from a different perspective, whether it be through corendering the data, adjusting color scales, or examining different depth and cross-sectional slices. This is especially important in AOIs where data quality is not optimal, and faults appear as less-defined lineaments. Cross-sectional analysis provides interpreters an easy way to check whether a lineament is a fault or not. In our case, an arbitrary line provides a more targeted cross-section (Figure 8). The evaluation confirmed the lineament in FR1 is indeed a fault due to the discontinuity and offset of reflectors. Moreover, it also revealed a powerful overlap between the fault and microseismic recordings, suggesting reactivation occurred along this fault plane.

Another way to analyze a specific fault is by simply checking its connectivity in shallower depth slices. By analyzing the evolution of the fault with microseismic data and depth, we notice that events taking place from 2011 to 2015 correlate to the low coherence areas of the fault at 6400 to 6500 feet in depth (Figure 9). This lower coherence area strikes towards the northeast. This suggests may be some relationship between areas of very low coherence and fault reactivation potential. Areas with a lower coherence could be weaker and thus activated first, coinciding with the initial wave of CCUS injections. Only after the second wave of injection from 2015 to 2018 does reactivation occur in the southwestern portion of the fault, where there is higher coherence. This illustrates that with more carbon injections, the fault is reactivating and striking in the southwest direction. Knowing where these potential weak zones and reactivation trends is crucial to further planning and injections. It aids geoscientists and engineers in determining well placement and decide if further injection should take place. And sometimes all it takes is looking at things from a different perspective.

Generative Topographic Mapping

Attribute selection is the most foundational step in any unsupervised ML. Attributes were selected on three criteria: how well they identify or enhance faults, how each attribute functions mathematically, and previous literature (Marfurt et al., 1998; Chopra and Marfurt, 2011; Gao et al., 2009; Gao et al., 2013; Hussein et al., 2021; Perico et al., 2023; Florez, 2023; Florez and Bedle 2024). Attributes considered and not utilized in the current workflow are multispectral coherence and aberrancy total azimuth. Aberrancy total azimuth is not utilized because it does not make sense to consider direction for fault enhancement. Multispectral coherence is not utilized because its results are not good, and it may be redundant as an additional coherence input.

In studies involving facies characterization, volumetric ML techniques (GTM, SOM, etc.) are effective in identifying varying facies (La Marca et al., 2024). However, for fault identification and enhancement, improvements made by GTM are not significant. For the Illinois Basin, low improvement between GTM and multi-attribute results is most likely due to the poor quality of the seismic at the basement, high attenuations at deep depths, and small vertical displacement of sub-seismic faults. Machine learning is reliant on good data and therefore good inputs. So, in AOIs of low data quality GTM will yield lower quality results. Additionally, another factor that may lead to low result improvement when employing GTM for fault enhancement and identification is the nature of the attributes themselves. Many geometric attributes employed are not a direct measure of whether or not a feature is a fault. There are a lot of potential relationships to be found in each attribute that may be tied to varying types of structural features not just faults. For example, coherence measures how similar seismic traces are to one another within a specific analysis window. Therefore, a feature seen within coherence could also be a channel edge. To mitigate this pitfall, fault enhancement was used as an input to GTM instead of a simple coherence calculation.

Due to the relatively low computational time, it would not be detrimental to test volumetric methods if the recommended geometric attributes are already calculated. However, if the goal is to improve preexisting results, GTM may not be worthwhile for fault identification or enhancement.

Convolutional Neural Network

CNN directly measures fault probability, similar to the fault enhancement attribute. Along with the calculations of the aforementioned suite of geometric attributes, it provides geoscientists and engineers with an additional way to grade the risk of a play. CNN is also computationally inexpensive. Depending on the volume of the seismic data set it could take up to 15 hours. This is a relatively good timeframe if the interpreter wants a big picture summary of the area.

However, even the best CNN or ML results are nothing without geologic context, the calculation of well-studied attributes, and the input of subject matter experts. In the IBDP, CNN fault probability is outputted in areas where faults appear in the seismic amplitude and coherence, respectively (Figure 13). These faults are proven to exist upon examining the seismic amplitudes on an arbitrary line. Discontinuous reflectors suggest there is a fault in the area. Additionally, microseismic clusters also appear on the probability output by CNN. Again, this is checked on both the depth slice of 6500 and the arbitrary line (2,1). This illustrates that CNN can be trusted within the context of the Illinois Basin. Additionally, examination of microseismic cluster propagation may reveal potential weaknesses to be aware of. For example, in Figure 14 there is a southwest trending cluster after injection into CCS2. This most likely indicates that there in some weaknesses in FR1. This makes sense due to the timing of events and the proximity of the microseismic cluster to CCS2.

Careful analysis of fault probability and microseismicity recordings allows the IBDP to monitor these areas. This may help prevent or mitigate risk caused by leakage into the air, ground, or water. More specifically, it could prevent leakage into the Mohomet Aquifer 5 miles north of the project. This aquifer is crucial in supplying a growing population of 840,000 people with water. Incident prevention will keep the populous safe, maintain a positive or neutral public perception towards CCUS, and thus continue government and industry support. It is important to note that CNN may not be as effective in other AOIs due to a differing context thus varying uncertainty. Therefore, it is recommended to calculate and analyze trusted attributes first then machine learning. CNN, in combination with other attributes, can provide a less intensive way to understand the overall fault networks of an AOI. A trained CNN model that is refined to an AOI will perform better than a pre-trained model. For our purposes, the goal is to enhance faults that may be hidden.

Fault damage zones are defined as a region of deformed rocks around a fault surface which results in interaction, propagation, and build-up along faults (Kim et al., 2004). Due to the complex network of fractures attributed to damage zones, they are critical in the energy sector. Understanding these zones is crucial to understanding hydrocarbon, CO_2, and water migration pathways and therefore aids in better risking an AOI. We hypothesize that fault probability outputted by CNN could potentially be linked to damage zones. This may be difficult to test without seismic modeling of damage zones and comparing the extent of modeled zones to fault probability outputs. Damage zones are rarely explored in 3D seismic because they are at the limit of seismic resolution. However, studies such as Liao et al., 2019 found the effectiveness of coherence in the characterization of damage zones and their features. A potential approach to test this hypothesis would be a marriage between Liao et al., 2019 and Botter et al., 2016, where the author would employ a combination of seismic modeling, field data, seismic attributes, and correlation of seismic amplitude to relate to fault damage.

Looking at the present

On September 13th, E&E News published that permit violations in the Illinois Basin were issued by the Energy Protection Agency (EPA). These violations pertain to the leakage of over 8,000 tons of CO_2 in one of the monitoring wells (MW) into unauthorized zones. Fortunately, for the approximately 850,000 residents who rely on the Mahomet Aquifer, their groundwater was safe. As of February 2025, there have been no specifications on which MW leaked. Based solely on our study, we believed the possible leakage route could have come from MW2 due to its proximity with CCS2. Additionally, according to our microseismic data, there has been reactivation near CCS2 as recently as 2018, due to injection (Figure 2, 13). However, according to Archer-Daniels-Midland (ADM), the leak was caused by pipe corrosion. Although it is not explicitly stated, we hypothesize that the corrosion could be due to the mix of incompatible but common fluids, such as CO_2 and water, within the pipe wall. The presence of CO_2 and water could potentially produce the corrosive carbonic acid and hydrates that can block flow (Jacobson, 2014).

Although this incident is not caused by the fault of geoscientists, it highlights an important lesson. A reminder that this technology is young and therefore issues are inevitable. However, if we want CCUS/CCS to be one of the avenues to combat anthropogenic climate change, we cannot let the fear of failure and leakage deter us. In other words, we must continue to test and find faults— technologically, geologically, or as a whole, to improve and learn.

Recommendations for the future

For seismic data interpretation, there is an almost unlimited combination of seismic attributes and machine learning algorithms a geoscientist may employ. The attribute selection stage is a foundational step for any interpreter, which means not only knowing the function, purpose, and importance of each attribute but fully understanding our study motivation overall. This section will review our workflow as well as provide recommendations on how to streamline it, potentially saving interpreters time and money.

Bad data quality, especially onshore data, is no news to geoscientists therefore, making the most of resource availability is key. As previously stated, the Illinois Basin's high-resolution 3D seismic data means it is great for shallower intervals but gets progressively worse with depth. This makes sense given the high attenuation of high-frequency seismic waves. Additionally, in terms of geology, the igneous basement rocks shield and absorb seismic waves leading to a blurred response, making it difficult to image smaller features.

Therefore, we highly recommend pre-conditioning the data with SOF before the application of seismic attributes. Applying an SOF is an essential step for low-quality seismic data and is especially crucial for fault identification workflows. Not only is SOF is great for pre-conditioning, it also makes it easy to interpret a seismic amplitude volume.

For noisy data, we suggest applying SOF before calculating seismic attributes because of its ability to reject cross-cutting noise and preserve signals parallel to the structure. Again, we see this in Figure 5 where SOF provides a dramatic difference between original seismic amplitude and post-SOF. Not only is SOF great for pre-conditioning but it can used to map faults as well—it increases fault confidence, decreases uncertainty, and improves connectivity. Moreover, analyzing the rejected noise from SOF is highly important as it helps us determine if we should increase or decrease the aggression of our filter. A filter that is too aggressive or removes too much noise and structure may risk making attribute calculations worse. Even for high-quality datasets, SOF still could be a powerful asset that may help lineate and sharpen features even more so. If an interpreter is looking to improve their results to a preexisting seismic dataset, applying an SOF before an attribute calculation may be a potential route.

SOF is a great filter and analyzing it independently is good. However, synthesizing all our tools and sources is how we bring the geologic story together. Seismic attributes and ML (CNN and GTM) provide that additional component. The original geometric attributes we tested are GLCM entropy, the similarity suite (energy ratio similarity, out product similarity, Sobel filter), fault enhancement of similarity, k1, k2, total aberrancy azimuth, and total aberrancy magnitude. All algorithms selected have their specific role to play. However, we found some do not have to be calculated if a refined workflow is desired.

For attributes, the main algorithm that can be eliminated is GLCM entropy. Entropy's measure of chaos is a good additional input into clustering algorithms like GTM and SOM. In the Illinois Basin, analyzing entropy independently does not result in additional features or lineaments that have not already been visualized in curvature, aberrancy, or similarity. Therefore, removing entropy from the workflow will save interpreters some computational time.

Other candidate attributes could be retained because each of them have specific strengths. For example, aberrancy is great for faults below resolution, curvature is great for faults with offsets seen as folds or reflectors, and similarity is for traditional faults with large offsets. Fault enhancement is the only additional attribute that has some overlap in function. Because it is the natural extension of similarity it also measures faults with large offsets. So, if similarity results are optimal after analyzing them from different perspectives and the interpreter is satisfied with their results, fault enhancement is not a required calculation. On the flip side, it can be argued that fault enhancement better lineates and decreases the uncertainty of those same faults. Therefore, if improvements of similarity products is desired then fault enhancement would be a potential route. Overall, we believe fault enhancement can be recommended but it is optional if similarity products are sufficient and saving time is desired.

To streamline the workflow, the ML algorithm we would remove would be GTM. GTM is good for identifying varying seismic facies as evident in previous studies (La Marca et al., 2024). However, like entropy, the products of GTM are not a significant improvement when compared to CNN or even the multi-attribute results. The latter methods of CNN and multi-attributes perform better and are computationally less expensive. Proper attribute selection take a lot time is essential for an effective GTM output. Despite careful attribute selection taken in this study, there were not major improvements in results. This is most likely due to low data quality and each attribute not being a direct measure of faults but more a potential measure. This leads to GTM results being an array of distracting colors with only one of those colors being fault lineaments. This is not helpful when the only features you are looking for are faults.

In contrast, CNN's prebuilt model allows the interpreter to simply input their original seismic volume with or without pre-conditioning and return a product in a decent amount of time. Even without the calculation of seismic attributes, the quick ML result from CNN allows interpreters to have a good and generalized idea of the geologic story. Regardless of whether you calculate multi-attributes first then CNN or vice versa, the key takeaway is you must calculate both. Attributes are the methods and algorithms we trust due to their foundation in previous literature and are a good basis of comparison. We highly recommend the use of CNN due to its low computational time, ease of use, and the potential to improve its results.

To summarize, if time were an issue, the algorithms we would keep would be SOF, similarity suite, both curvatures, aberrancy total azimuth, and aberrancy total magnitude. As for machine learning, CNN would be retained. The beauty of science is that each study has its own set of suggestions for improvement, and it is up to the next scientist to make their own choices.

CONCLUSIONS

Coherence attributes is the foundational step to many geologic interpretations of seismic data. However, due to the restrictions of low-quality basement onshore data and the presence of sub-seismic faults, coherence alone is not sufficient. We employed a ML approach with multi-attributes in Decatur, Illinois where microseismicity related to CCUS had been induced in the basement.

We found that applying SOF sharpened faults and removed noise, therefore improving the results of multi-attributes and ML. The improvement in SOF allows interpreters to increase confidence in the connectivity of faults and decrease uncertainty. Broadband coherence was largely unsuccessful in enhancing most fault networks. However, some overlap exists between microseismicity recordings and coherence, especially in FR1. The overlay between microseismicity, coherence, and seismic amplitude results is further complemented by examining additional depth slices. This reveals the microseismic history and a potential southwest striking reactivation pattern upon further carbon injection.

As expected, entropy did not aid in visual interpretation. However, it was utilized as an input for GTM. Curvature and aberrancy were found to be more sensitive to faults with sub-seismic vertical displacement, capturing potential faulted regions that coherence could not. This is due to the curvature suite measuring curves and flexures versus dissimilarity. Therefore, without curvature and aberrancy, interpreters may risk missing faults with a low vertical displacement.

GTM was tested utilizing the aforementioned geometric attributes. For enhancing basement faults, GTM did not capture anything new or improve upon the multi-attribute calculations. This makes sense given the noise and the low performance of GTM for fault enhancement in other datasets.

In contrast, supervised ML had better results. Despite using a pre-trained model, CNN mapped faults that were similar to potential lineaments picked up by coherence and overlapped with microseismic data. Independently, ML is nothing without the input of subject matter experts and geologic context. But with the addition of a similar result within coherence and strong microseismic correlation, CNN proved effective.

The multi-attribute and ML workflow is highly transferable to other geological prospects such as well planning and operations to energy transition ventures such as Geothermal and CCUS. Geohazard analysis can reduce the operation time and cost. More importantly, with new technologies such as CCUS, where public perception and support are sensitive, it is important to decrease uncertainty and prevent large-scale reactivation events. CO2 leakage may not be our fault yet, but it can be. So, we as scientists— as storytellers, must continue to play our part in mitigating risks and doing the best we can.

REFERENCES

Bauer, R. A., Will, R., Greenberg, S., Whittaker, S. G., 2019, Illinois Basin- Decatur Project, Geophysics and Geosequestration.

Bedle, H., Beutel, A.M., Garneau, C.R.H. (2024) SPEER23 Survey

Report: Investigating Climate, Weather, and Energy Attitudes in the United States through the Lens of Social and Psychological Factors.

Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, The generative topographic

mapping: Neural Computation, 10, 215–234.

Blanco D. V. P., Bedle H., Marfurt, K., 2024, Post migration seismic data conditioning through merge datasets: enhancing accuracy and insight, IMAGE extended abstract.

Blanco D. V. P., 2024, Post-migration seismic data conditioning methods on a merged dataset. MS Thesis, University of Oklahoma. p. 1-61. https://shareok.org/items/da7907c3-c708-4836-b11b-d9b413f4f95f

Botter, C., Cardozo, N., Hardy, S., Lecomte, I., Paton, G., Escalona, A., 2016, Seismic characterization of fault damage in 3D using mechanical and seismic modeling, Marine and Petroleum Geology, Volume 77, p. 973-990, ISSN 0264-8172, doi: https://doi.org/10.1016/j.marpetgeo.2016.08.002.

Collinson, C., Sargent, M.L., and Jennings, J.R., 1988, Illinois Basin region, in Sloss, L.L., ed.,

Sedimentary cover—North American craton, U.S.: Boulder, Colorado, Geological Society of America, Geology of North America, v. D-2, p. 383–426.

Chopra, S., and K. J. Marfurt, 2011, Interesting pursuits in seismic curvature attribute analysis: CSEG Recorder, 36, 40 – 50.

Chopra, S., and K. J. Marfurt, 2018, Coherence attribute applications on seismic data in various Guises – Part 2: Interpretation, 6 (3), T531 – T54, doi: 10.1190/INT-2018-0007.1

Curry, T. & Reiner, David & Ansolabehere, S. & Herzog, H.. (2005). How aware is the public of carbon capture and storage? in Rubin E S, Keith D W and Gilboy C F (eds), Proceedings of the Seventh International Conference on Greenhouse Gas Control Technologies: GHGT7. IEA GHG, Cheltenham, UK. 1. 1001-1010.

Falconer S. and K. J. Marfurt, 2008, Attribute-driven footprint suppression: SEG Technical Program Expanded Abstracts, 2667-2671.

Fehmers, G. C., and C. F. W. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, 68 (4), 1286–1293, doi:10.1190/1.1598121

Fehmers, G. C., and C. F. W. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, 68 (4), 1286 – 1293, doi: 10.1190/1.1598121

Florez, D. K. S., 2023. Application of Seismic Attributes and Unsupervised Machine Learning Methods for identification of Hidden Faults in the Basement and Carbonate Rocks. MS Thesis, University of Oklahoma, p1-97. https://shareok.org/handle/11244/340047.

Florez, D. K. S. and H. Bedle. 2024. Application of novel geometric seismic attribute for enhancing fault visualization in areas of potential carbon capture and storage. Leading Edge, 43 (3) (2024), pp. 165-175, doi: 10.1190/tle43030165.1

Gao, D., 2004, Texture model regression for effective feature discrimination: Application to seismic facies visualization and interpretation: Geophysics, 69, 958-967.

Gao, D., 2007, Application of three-dimensional seismic texture analysis with special reference to deep-marine facies discrimination and interpretation: An example from offshore Angola, West Africa: AAPG Bulletin, 91, 1665-1683.

Gao, D., 2009, 3D seismic volume visualization and interpretation: An integrated workflow with case studies: Geophysics, 74, W1–W12.

Gao, D., 2011, Latest developments in seismic texture analysis for subsurface structure, facies, and reservoir characterization: A review: Geophysics,

Gao, D., 2013, Integrating 3D seismic curvature and curvature gradient attributes for fracture characterization: Methodologies and interpretational implications: Geophysics, 78 (2), O21 – 031, doi: 10.1190/geo2012-0190.1

Hale, D., 2009, Structure-oriented smoothing and semblance, CWP Report 635

Hussein M., R. R. Stewart, and J. Wu, 2021, Which seismic attributes are best for subtle fault detection?: Interpretation, 9 (2), T299–T314

Leetaru, H.E. and Freiburg, J.T. (2014), Litho-facies and reservoir characterization of the Mt Simon Sandstone at the Illinois Basin – Decatur Project. Greenhouse Gas Sci Technol, 4: 580-595. doi: 10.1002/ghg.1453

Lubo-Robles, D., T. Ha, S. Lakshmivarahan, K. J. Marfurt, and M. J. Pranter, 2021, Exhaustive probabilistic neural network for attribute selection and supervised seismic facies classification: Interpretation, 9, no. 2, T421–T441, doi: 10.1190/INT-2020-0102.1.

Jacobson, G. A., 2014, Pipeline Corrosion Issues Related to Carbon Capture, Transportation, and Storage. NaCE International, Volume 53, Issue 5, p. 24-31. https://www.icmt.ohio.edu/Articles/pdfs/May-MP-CCTS-feature.pdf

Jones C., Angela, et al. "Class VI Carbon Sequestration Wells: Permitting and State Program Primacy." Congressional Research Service. Retrieved from: https://crsreports.congress.gov/product/pdf/R/R48033

Kim, Y. S., Peacock, D. C. P., Sanderson, D. J., 2004, Fault Damage Zones, Journal of Structural Geology, Volume 26, Issue 3, p. 503-517. Doi: https://doi.org/10.1016/j.jsg.2003.08.002

Kumar, P. C., Saikia, P. P., Bedle, H., Sain, K., 2025. Unsupervised machine learning models applied to basement faults: An example from the Dibrugarh region, NE India, Journal of Asian Earth Sciences, Volume 280, 106446, ISSN 1367-9120, doi: https://doi.org/10.1016/j.jseaes.2024.106446.

La Marca, K., Bedle, H., Stright, L., Marfurt, K., 2024. Uncertainty assessment in unsupervised machine learning methods for deepwater channel seismic facies using outcrop-derived 3D models and synthetic seismic data, Interpretation 12, Issue 3. doi: https://doi.org/10.1190/INT-2023-0094.1

Li, F., et al. "Revisit seismic attenuation attributes: Influences of the spectral balancing operation on seismic attenuation analysis." Interpretation 9, no. 3 (2021): T767–T779.

Luo, Y., M. Marhoon, S. Al Dossary, and M. Alfaraj, 2002, Edge-preserving smoothing and applications: The Leading Edge, 21 (2), 136 – 158, doi: 10.1190/1.1452603

Marfurt, K. J. Seismic Attributes as the Framework for Data Integration Throughout the Oilfield Life Cycle. Society of Exploration Geophysicists, 2018

McBride, J.H., Kolata, D.R., Sargent, M.R., and Hildenbrand, T.G., 2010, The Precambrian crust of Illinois—The geological and geophysical evidence, in Kolata, D.R., and Nimz, C.K., eds., The Geology of Illinois: Champaign, Illinois State Geological Survey, p. 123–135.

McBride, J.H., Leetaru, H.E., Keach, R.W., II, and McBride, E.I., 2016, Fine-scale structure of the Precambrian beneath the Illinois Basin: Geosphere, v. 12, no. 2, p. 585–606, doi:10.1130/GES01286.1

Mora, J.P., H. Bedle, and K.J. Marfurt, 2022, Fault enhancement using probabilistic neural networks and Laplacian of a Gaussian filter: A case study in the Great South Basin, New Zealand: Interpretation, 10, SC1-SC15, doi: 10.1190/INT-2021-0127.1

Oladipupo Babarinde, Roland Okwen, Scott Frailey, Fang Yang, Steven Whittaker, Dustin Sweet, 2021, A workflow to assess the efficacy of brine extraction for managing injection-induced seismicity potential using data from a CO2 injection site near Decatur, Illinois, International Journal of Greenhouse Gas Control, v. 109, doi: 10.1016/j.ijggc.2021.103393.

Perico, E. 2021, Application of seismic attributes and machine learning clustering techniques to the characterization of faults in a post-salt reservoir, Jubarte Field (Campos Basin). Master Thesis. University of Oklahoma

Perico, E., Bedle, H., Buist, B., Damasceno, A. C., 2023. Fault characterization in a postsalt reservoir interval, Jubarte Field (Campos Basin), using seismic attributes and machine learning. Interperation.

Randen, T., S.I. Pedersen, and L. Sønneland, 2001, Automatic extraction of fault surfaces from three-dimensional seismic data: Annual International Meeting, SEG, Expanded Abstract, 551–554.

Rie Kamei, Nori Nakata, and David Lumley, 2015, Introduction to microseismic source mechanisms, The Leading Edge, 34:8, 876-880 doi: 10.1190/tle34080876.1

Robert Will, Valerie Smith, Hannes E. Leetaru, Jared T. Freiburg, Donald W. Lee, 2014, Microseismic Monitoring, Event Occurrence, and the Relationship to Subsurface Geology, Energy Procedia, v. 63, p. 4424-4436, doi: 0.1016/j.egypro.2014.11.478.

Roy, A., M. Matos, and K. J. Marfurt, 2011, Application of 3D clustering analysis for deep marine seismic facies classification – an example from deep water northern Gulf of Mexico: to be presented at the GCSSEPM 31st Annual Bob. F. Perkins Research Conference

Roy, A., 2013. Latent space classification of seismic facies. The University of Oklahoma. Ph.D. dissertation. University of Oklahoma.

Roy, A.S. Romero-Peláez, T.J. Kwiatkowski, K.J. Marfurt, 2014. Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico. Interpretation, 2 (1) (2014), Article SA31-SA47, 10.1190/INT-2013-0077.1

Sherilyn Williams-Stroud, Robert Bauer, Hannes Leetaru, Volker Oye, Frantisek Stanek, Sallie Greenberg, Nadege Langet; Analysis of Microseismicity and Reactivated Fault Size to Assess the Potential for Felt Events by Injection in the Illinois Basin. *Bulletin of the Seismological Society of America* 2020;; 110 (5): 2188–2204. doi: 10.1785/0120200112

Thomas, W.A., Tucker, R.D., Astini, R.A., and Denison, R.E., 2012, Ages of pre-rift basement and synrift rocks along the conjugate rift and transform margins of the Argentine Precordillera and Laurentia: Geosphere, v. 8, p. 1366–1383, doi: 10.1130 /GES00800.1.

Tuttle, M.P., Schweig E.S., Sims J.D., Lafferty R.H., Wolf L.W., Haynes M.L., The earthquake potential of the new madrid seismic zone, Bull. Seismol. Soc. Am., 92(6) (2002), pp. 2080-2089, doi: 10.1785/0120010227

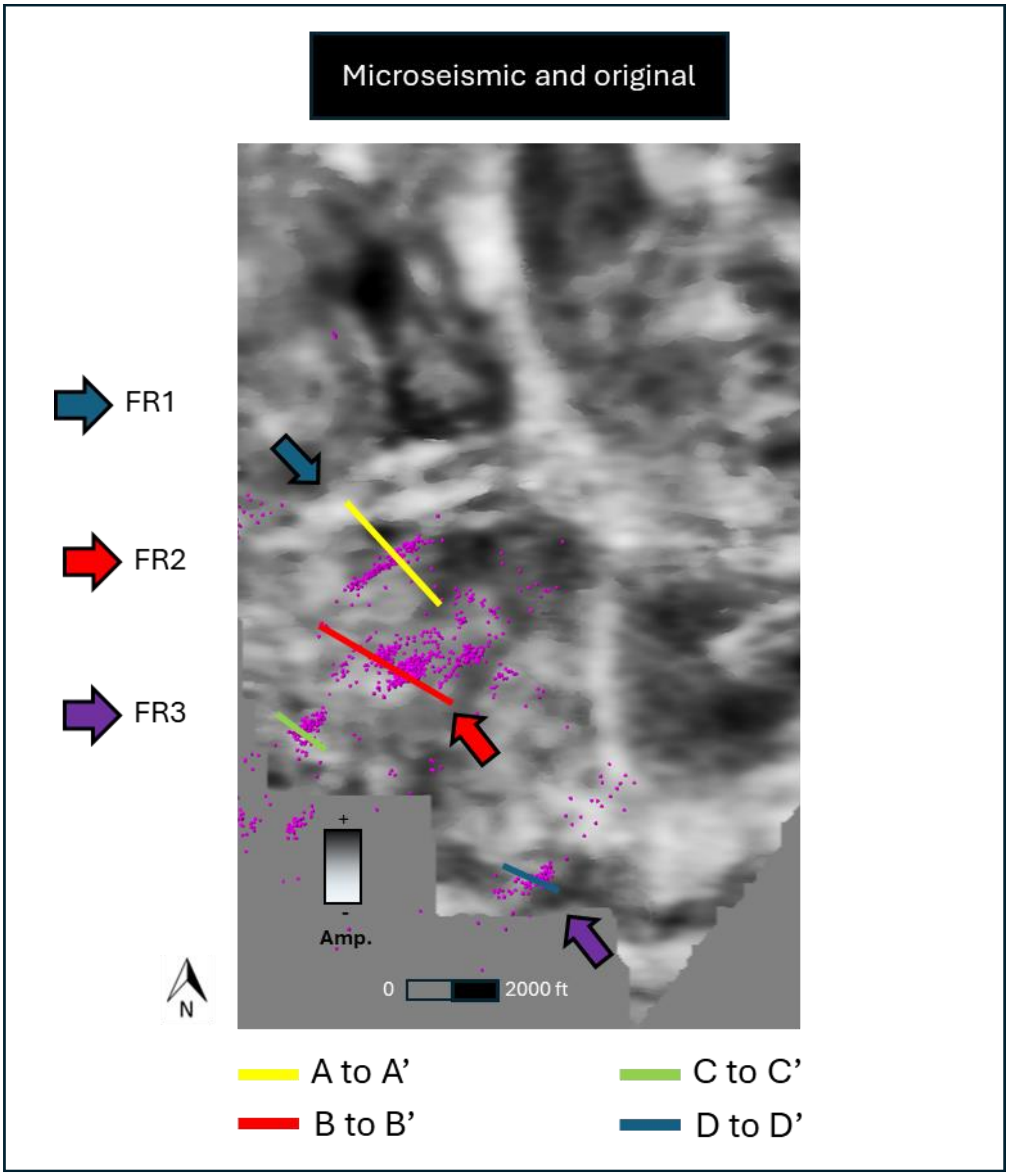
Wu, X., and D. Hale, 2015, 3D seismic image processing for faults: Geophysics, 81, 1-11.

Zhang, B., Y. liu, M. Pelissier, and N. Hemstra, 2014, Semiautomated fault interpretation based on seismic attributes: Interpretation, 2, SA11-SA19.

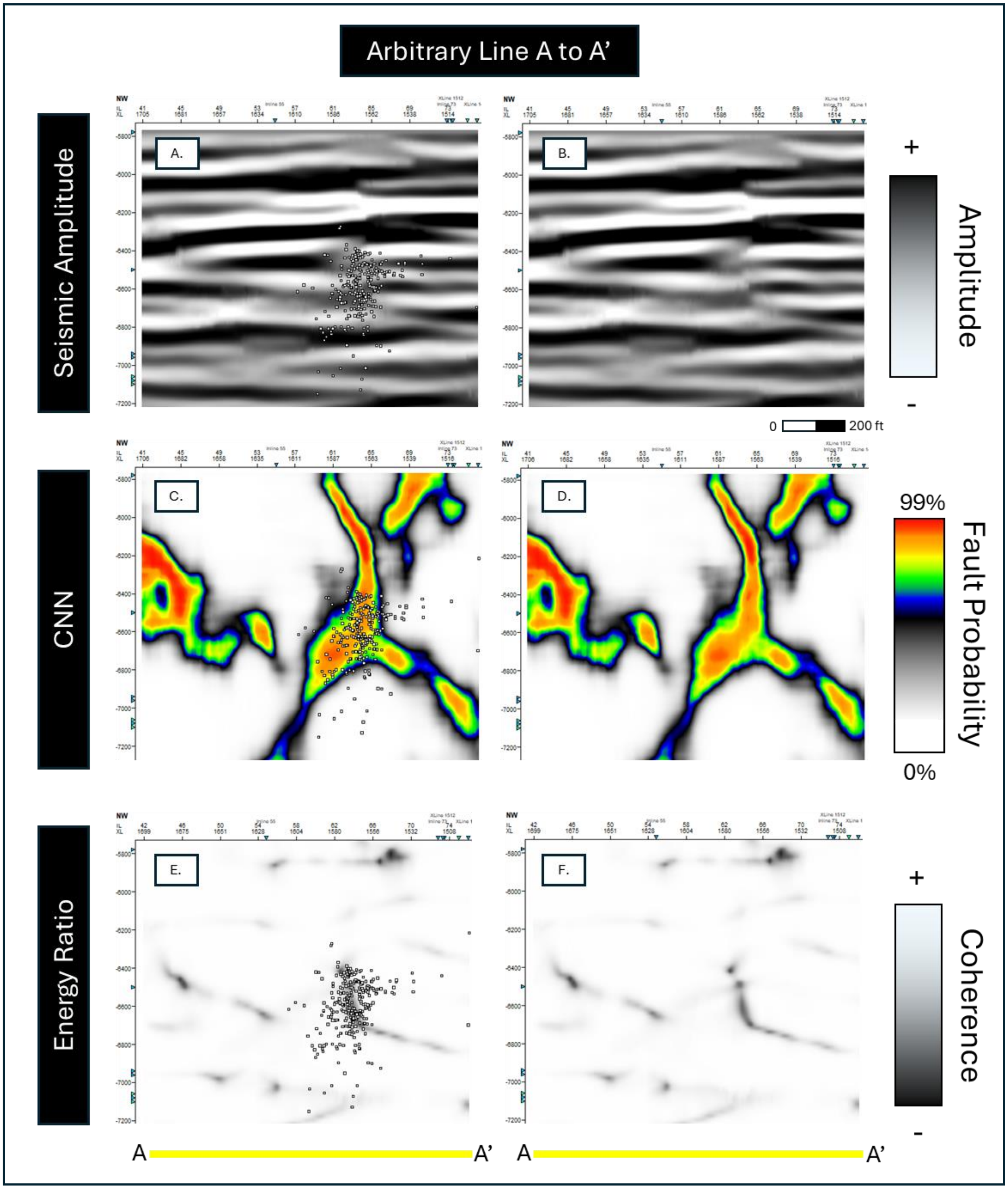
Wu, X., L. Liang, Y. Shi, and S. Fomel, 2019, FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation: Geophysics, 84, no. 3, IM35–IM45, doi: 10.1190/geo2018-0646.1.

Yenugu, M., K. J. Marfurt, and S. Matson, 2011, Seismic texture analysis for reservoir prediction and characterization: The Leading Edge, 29, 1116-1121.

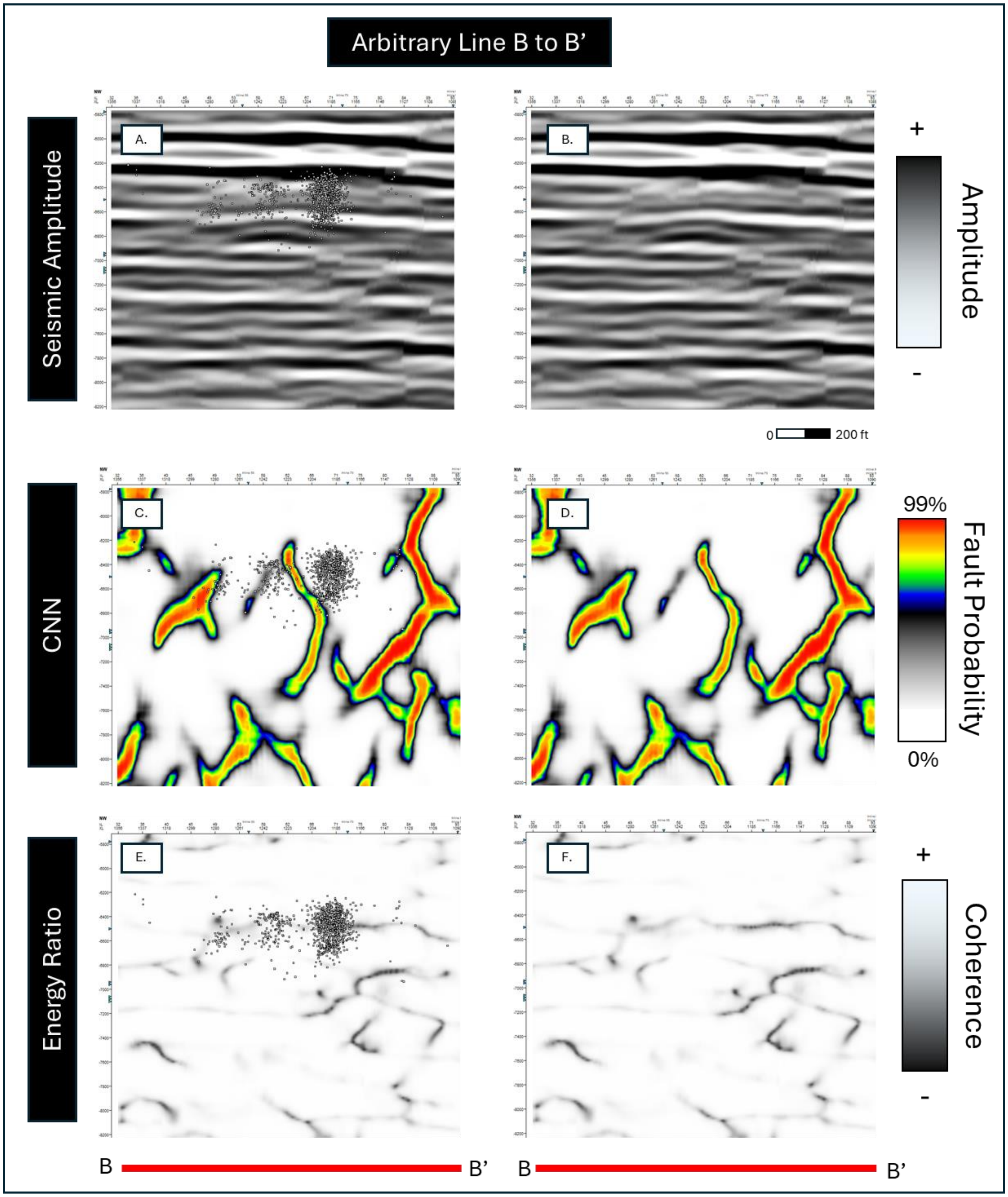
APPENDIX



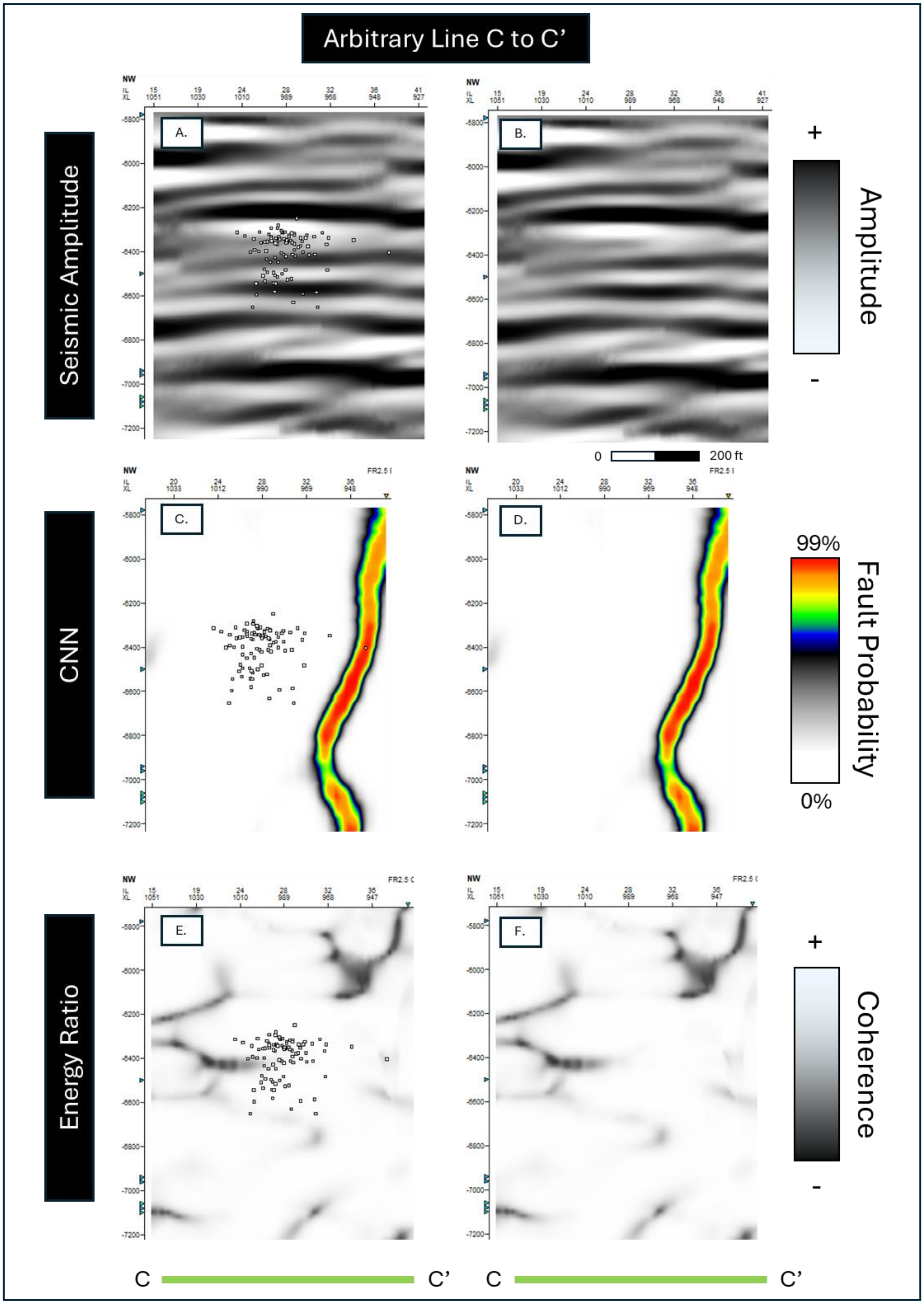
Appendix. Figure 1. Overview of the original seismic amplitude with microseismic clusters in purple. Fault regions 1 to 3 are visualized the same as before. The blue arrow represents fault region 1 (FR1), the red arrow represents (FR2), and the purple arrow represents (FR3). Four arbitrary lines are highlighted: A to A' (yellow), B to B' (red), C to C' (Green), and D to D' (Blue).



Appendix. Figure 2. Arbitrary line A to A' cross sectional slices. Sections A, C, and E represent seismic amplitude (post-SOF), CNN, and energy ratio similarity (ERS) with microseismic clusters, respectively. Microseismic clusters have a projection of 250 feet. Sections B, D, and F are seismic amplitude (post-SOF), CNN, and ERS with microseismic clusters.

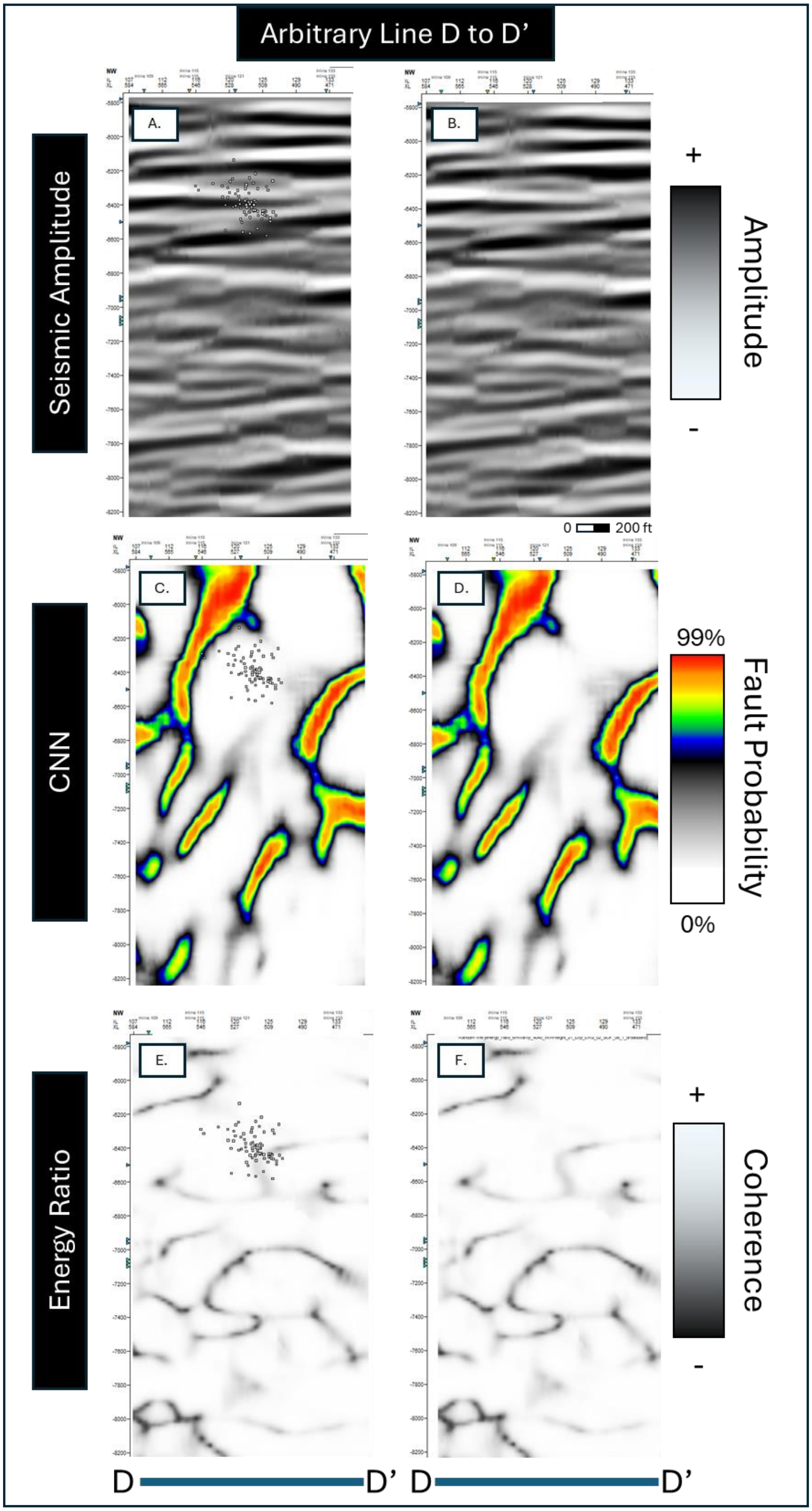


Appendix. Figure 3. Arbitrary line B to B' cross sectional slices. Sections A, C, and E represent seismic amplitude (post-SOF), CNN, and ERS with microseismic clusters, respectively. Microseismic clusters have a projection of 250 feet. Sections B, D, and F are seismic amplitude (post-SOF), CNN, and ERS with microseismic clusters.

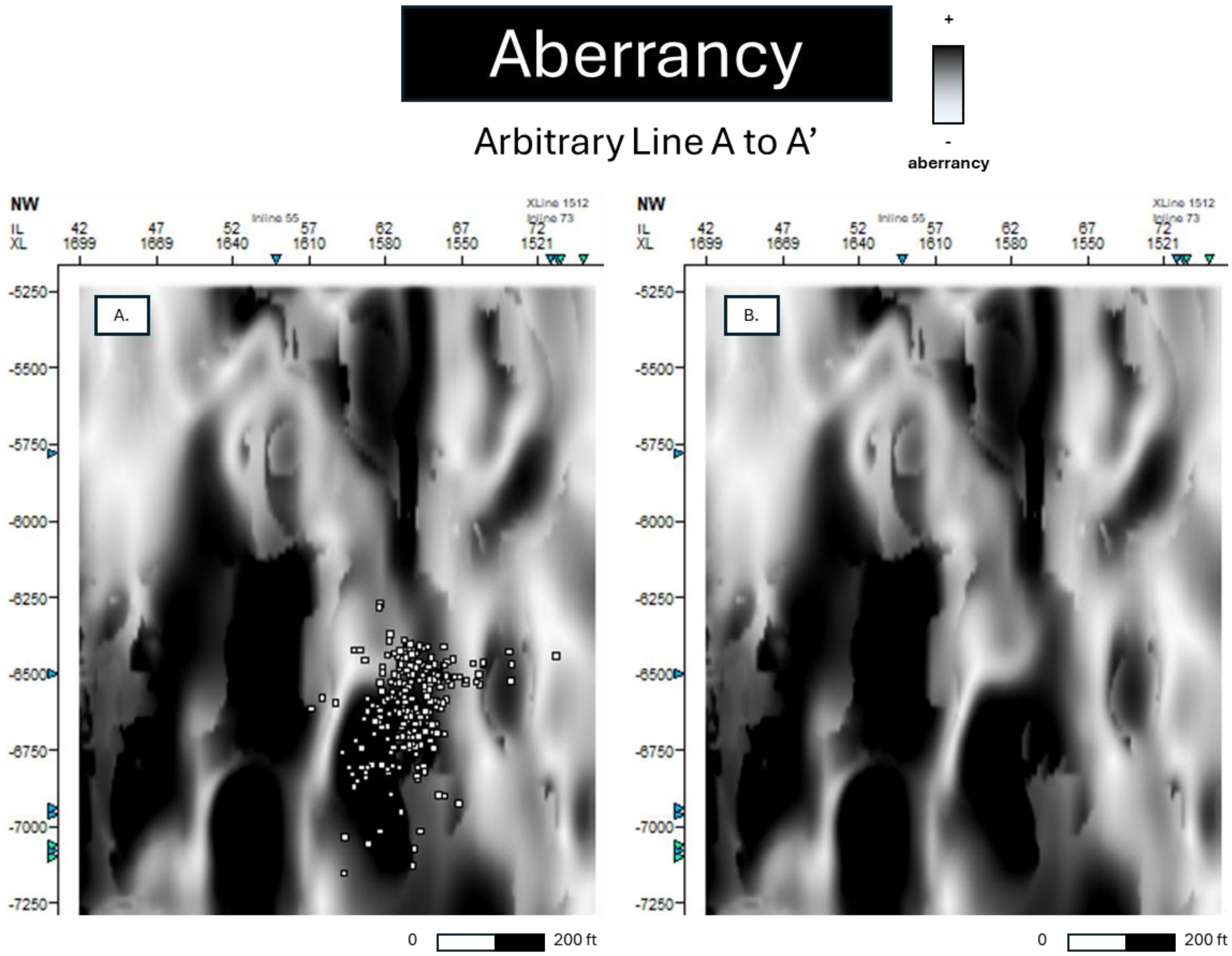


Arbitrary Line C to C’
Seismic Amplitude
CNN
Energy Ratio
A.
B.
C.
D.
E.
F.
+
Amplitude
-
0
200 ft
99%
Fault Probability
0%
+
Coherence
-
C
C’
C
C’

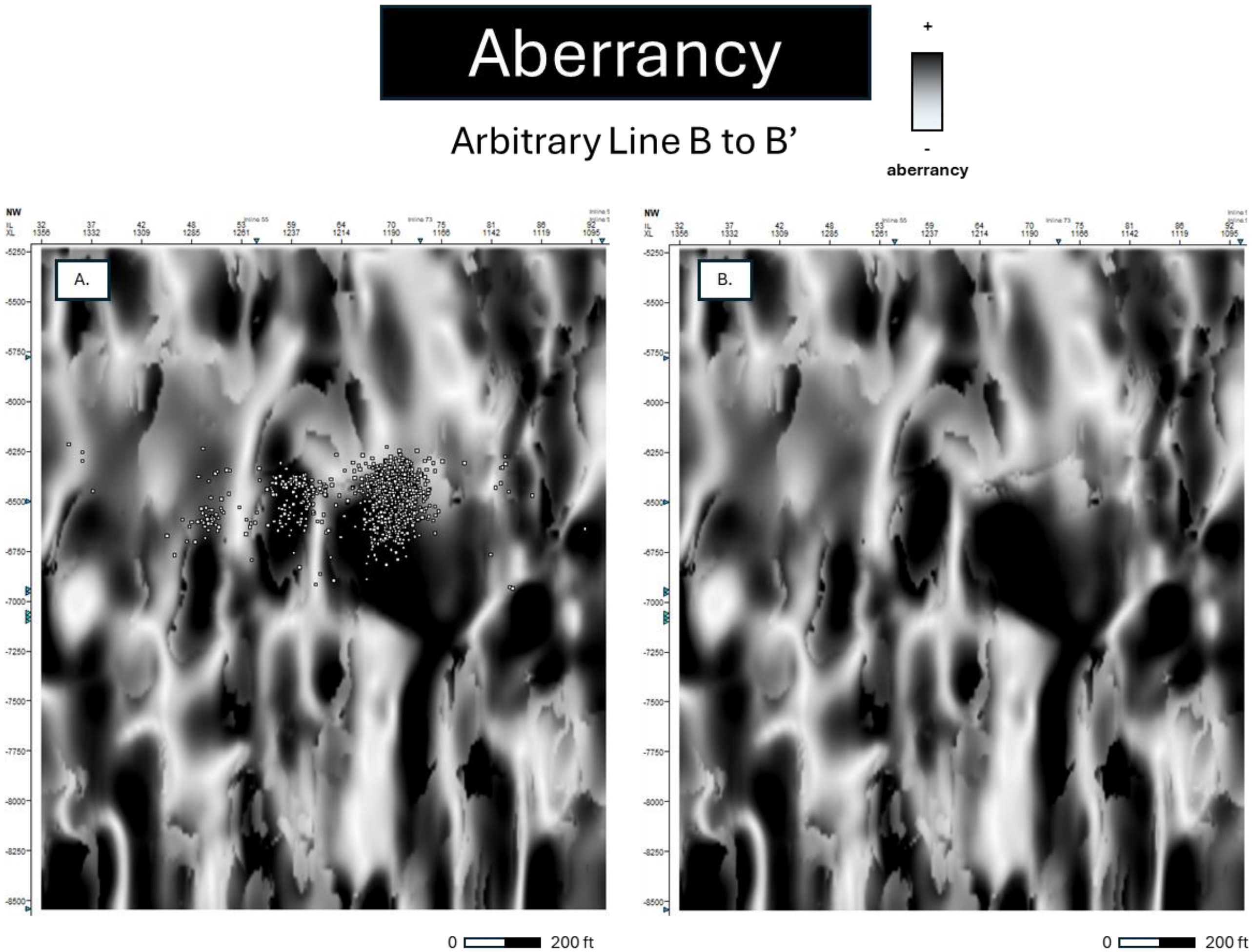
Appendix. Figure 4. Arbitrary line C to C' cross sectional slices. Sections A, C, and E represent seismic amplitude (post-SOF), CNN, and ERS with microseismic clusters, respectively. Microseismic clusters have a projection of 250 feet. Sections B, D, and F are seismic amplitude (post-SOF), CNN, and ERS with microseismic clusters.


Arbitrary Line D to D’
Seismic Amplitude
CNN
Energy Ratio
A.
B.
C.
D.
E.
F.
+
Amplitude
-
0 200 ft
99%
Fault Probability
0%
+
Coherence
-
D
D’
D
D’

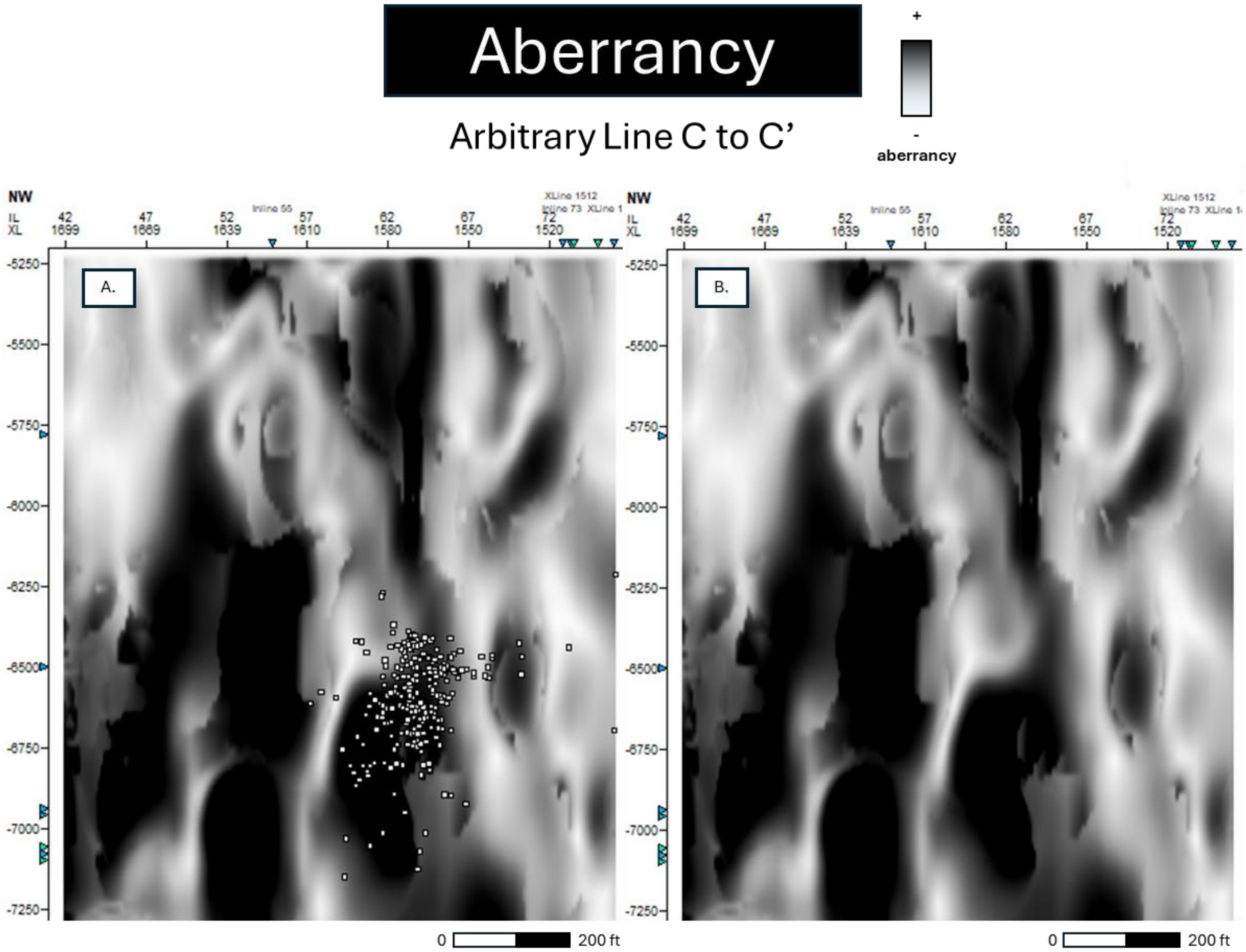
Appendix. Figure 5. Arbitrary line D to D' cross sectional slices. Sections A, C, and E represent seismic amplitude (post-SOF), CNN, and ERS with microseismic clusters, respectively. Microseismic clusters have a projection of 250 feet. Sections B, D, and F are seismic amplitude (post-SOF), CNN, and ERS with microseismic clusters.



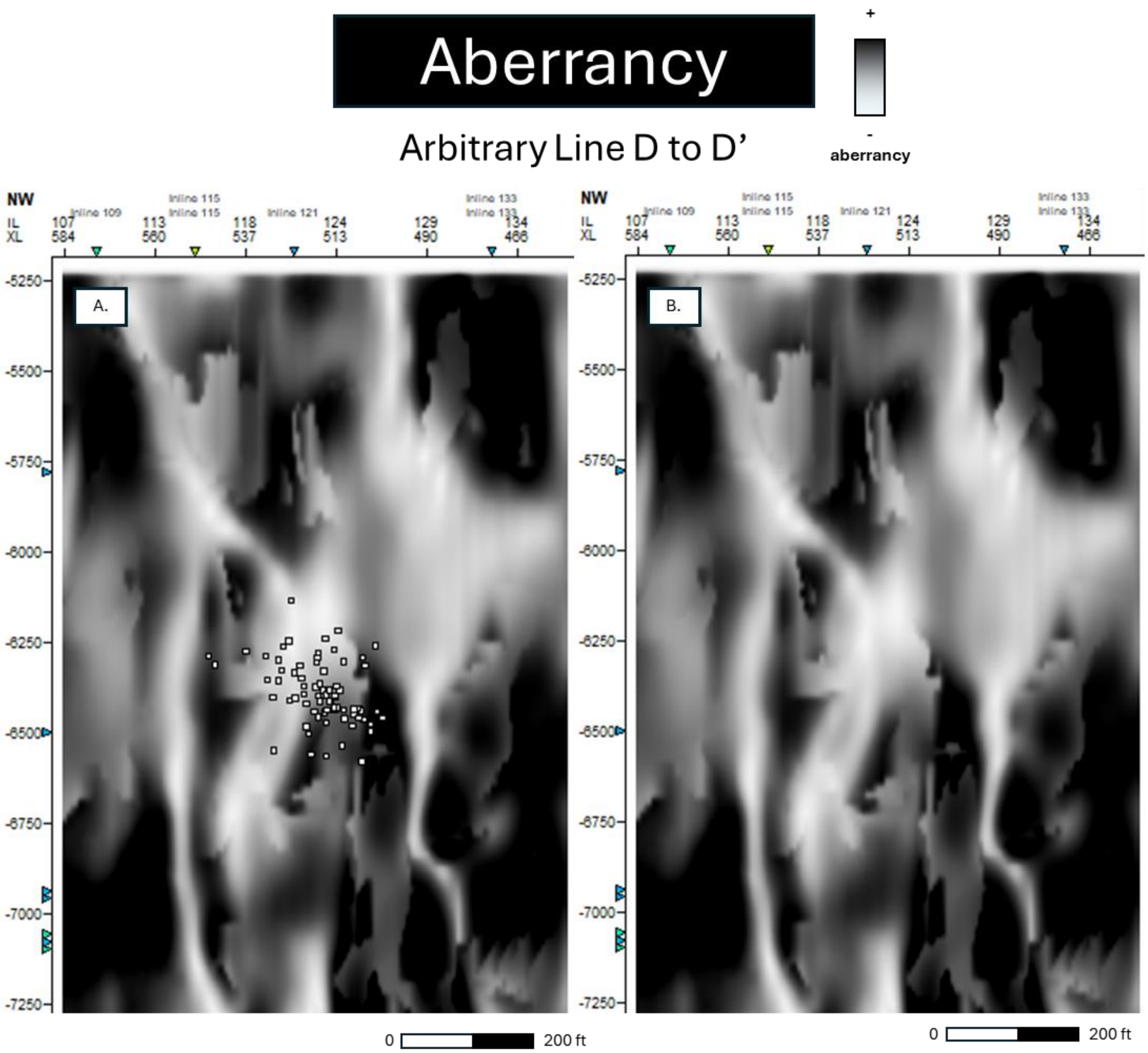
Appendix. Figure 6. Cross-sectional view of total aberrancy magnitude from arbitrary line A to A'. A) is the view with microseismic clusters, and B) is the view without microseismic clusters. Clusters are concentrated in regions of high aberrancy or black color.



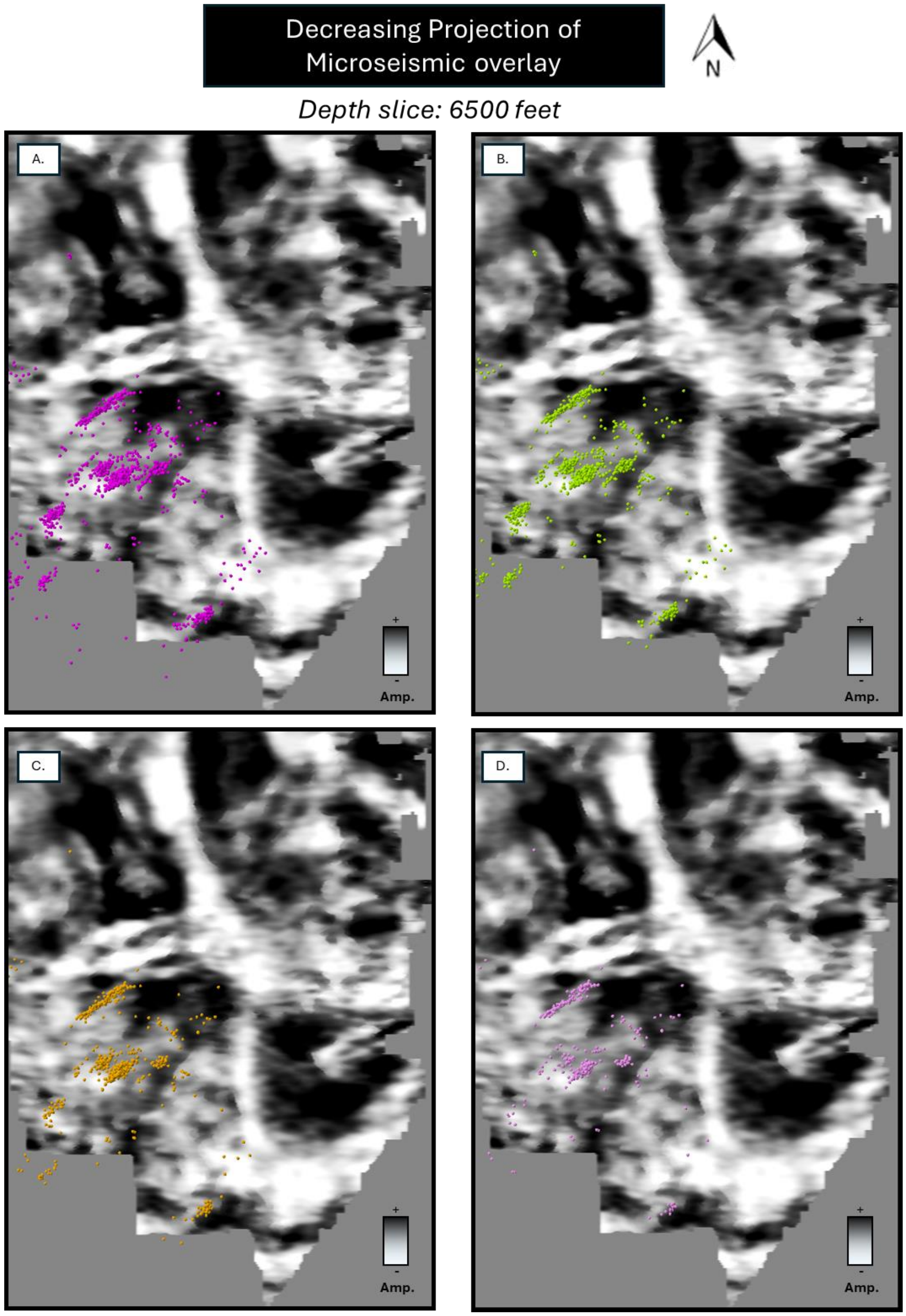
Appendix. Figure 7. Cross-sectional view of total aberrancy magnitude from arbitrary line B to B'. A) is the view with microseismic clusters, and B) is the view without microseismic clusters. Clusters are concentrated in regions of high aberrancy or black color.



Appendix. Figure 8. Cross-sectional view of total aberrancy magnitude from arbitrary line C to C'. A) is the view with microseismic clusters, and B) is the view without microseismic clusters. Clusters are concentrated in regions of high abberancy or black color.



Appendix. Figure 9. Cross-sectional view of total aberrancy magnitude from arbitrary line A to A'. A) is the view with microseismic clusters, and B) is the view without microseismic clusters. Clusters are NOT concentrated in regions of high but in regions of moderate aberrancy or dark gray color.



Appendix. Figure 10. An overview of what the depth slice of 6500 feet may look like if the microseismic projection is decreased. A) All microseismic readings are considered regardless of depth. These values range from 5,943 to 6500 feet, resulting in a projection of 557 feet. B) Projection is decreased to 200 feet. Microseismic range (MSR) starting at 6300 to 6500 feet. C) Projection is decreased to 100 feet. MSR is 6400 to 6500 feet. D) Projection is decreased to 50 feet. MSR is 6450 feet to 6500 feet.

Additional arbitrary lines were selected to test if there is a potential overlap between microseismic recordings, ML, and attribute calculations (Appendix, Figure 1). Based on the evident microseismic overlap in FR1, the cross-sectional arbitrary lines of post-SOF seismic amplitude, CNN, and ERS were selected (Appendix, Figure 2). SOF provides a basis of comparison to visualize and observe for any evidence of faults, i.e., abrupt discontinuity of reflectors, seismic amplitude die out, or seismic amplitude offsets. In appendix Figure 2, there is a strong correlation between the fault on arbitrary line A to A' and microseismic recordings. This fault is detected by both CNN and energy ratio, evident by the warm spectrum of the fault probability color bar (Figure C-D) and the low-coherence black of energy ratio (Figure 2-E, F). Out of all the arbitrary slices through the microseismic clusters, A to A prime shows the most correlation between results and recordings. All other arbitrary lines exhibit little to no overlap between microseismic clusters and ML-attribute results (Appendix, Figure 3-5).

The hypothesis to why this is, may be due to the nature of CNN and ERS. CNN is trained on synthetics and not on the Illinois Basin, so it will not detect everything. In theory, CNN results can be improved if trained on the Decatur dataset. However, this training may be limited and time-consuming given the difficulty of picking hidden faults. As seen in Figure 10, ERS is better for faults with larger displacements, so oftentimes, it leaves large areas of high coherence. Unlike curvature or aberrancy, ERS cannot detect faults seen as folds or flexures (Figure 11). Therefore, it is crucial to also analyze aberrancy. Appendix Figures 6 to 8, illustrate that there is strong correlation between microseismic clusters and areas of high aberrancy. The only exception would be appendix figure 9, which has clusters in areas of moderate to low aberrancy. Despite the strong correlation between aberrancy and microseismic, it can be difficult to interpret. Areas of high aberrancy appear as large dark blobs versus the more lineated features of ERS or fault probability zones of CNN. Regardless, it provides an additional piece of the story that can be used in tandem with ERS and CNN results. In the case of the Illinois Basin, aberrancy manages to find potentially high risk regions that geoscientists may need to be aware of.

Decreasing the projection of microseismic clusters on cross-sectional slices lineate faults or lineaments better. This is also tested on the depth slice of 6500 (Appendix, Figure 10). Lineaments created by microseismic on the depth slice of 6500 feet appear to be a lot sharper, comparing Figures 10-A and 10-D. It appears that most microseismic recordings around the depth of 6300 feet to 6500 feet are localized around FR1 or the main fault detected in CNN and ERS. This could potentially explain why FR1 shows the strongest correlation. Not only does FR1 have the most microseismic recordings near the depth of 6500, but it also has a strong vertical displacement. Combined, this may explain why FR1 exhibits high correlation between microseismic, CNN, and ERS.