Developing screening tools to assess fault leakage for CO_2 storage

Author: Jonathan Brunton, **E-mail:** jab121@ic.ac.uk

Github: edsml-jab121, **Project Repository:** hhtps://github.com/ese-msc-2021/irp-jab121

Imperial College London, Department of Earth Science and Engineering

MSc Environmental Data Science and Machine Learning, Independent Research Project

Supervisors: Dr. Rebecca Bell, Dr. Cedric John

2nd September, 2022

Abstract.

Captured atmospheric CO_2 may be stored in the subsurface but requires a suitable environment to be reliably trapped. While some seismic faults may provide appropriate sealing capacity to store captured CO_2, other faults may be conduits for fluid escape. Assessing a fault's leakage potential is typically a process that requires manual interpretation of both seismic reflectivity data and drill core samples from the area. One method proposed here that could improve the quality of leakage assessments is a machine learning model that is able to accurately extract faults from a 3D seismic volume and interrogate their physical properties using a high-resolution P-wave velocity model produced with full waveform inversion (FWI). FWI technology can resolve velocity anomalies within seismic data to a much finer degree than with conventional methods and may be able to reveal whether certain faults correspond to high velocity zones (a proxy for sealed locations) or low velocity zones (open for fluid leakage). The potential of 3D FWI to reveal fault zone properties for CO_2 storage applications has not been tested. This study discusses the development and validation of a K-Means machine learning model used to extract fault locations from seismic reflectivity data recovered from the Samson Dome area of the Barents Sea. Additionally, the results of this model are used to assess the capabilities of FWI by investigating the velocity structure of faults within the volume and analysing the relationships between this fault network's properties and capacity within the network to store fluid. While only an initial investigation into the capabilities of this approach, the combination of the optimized K-Means model and FWI show promising results for visual interpretation. Additionally, an alternative workflow for assessing fault leakage was successfully developed, and a preliminary discussion on the limitations of this screening tool is presented.

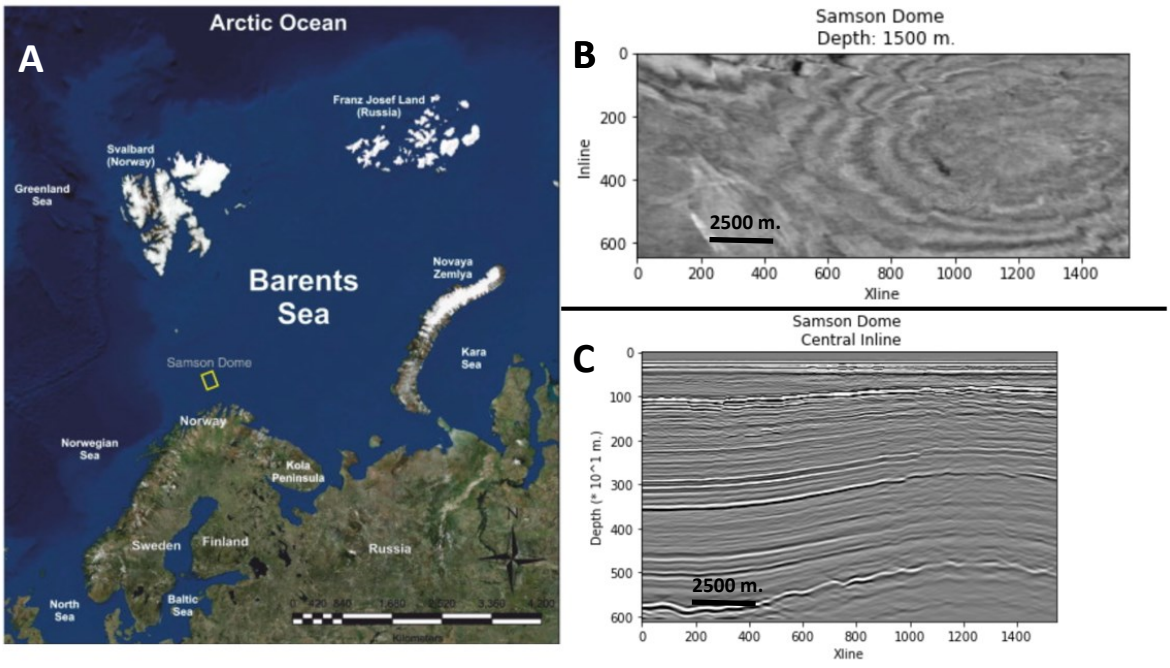

Figure 1. A.) Taken directly from Mattos et al. (2016), Figure 1.a. The location of the Samson Dome region from which the data used in this study is recovered. The Samson Dome is highlighted in the yellow box, just off the northern coast of mainlaind Norwary. **B.)** A full horizontal slice of the Samson Dome subvolume used in this study. Here we see evidence of the radial faults that dominate the area. **C.)** A vertical slice of the seismic reflectivity volume. An example of how the 'raw' seismic reflectivity data looks when used in fault interpretation.

1. Introduction.

Carbon capture and storage (CCS) has been recognized as one potential solution to mitigating the harmful effects on Earth's climate from CO_2 emissions caused by the burning of fossil fuels (Gibbins et al., 2008). As technology has become increasingly adept at being able to capture atmospheric CO_2, the most significant hurdle in CCS has become finding suitable locations where the captured CO_2 may be stored effectively. In the subsurface, geologic structures have been proven to be an effective seal for naturally occurring fluids, like CO_2, in the Earth's crust (Boot-Handford et al., 2014). Faults and fractures caused by tectonic movements may be appropriate locations in the crust that are able to provide a fluid seal that prevents injected CO_2 from escaping quickly back into the atmosphere. However, in some cases, these tectonic structures can readily form in ways that create a conduit for fluids to travel, therefore providing an opportunity for potential leakage (Mattos et al. 2016). Up to this point, investigating a fault's ability to store fluids has been a process that requires both manual interpretation of fault images and analysis of drill core samples from the region.

Background: Seismic Reflectivity Interpretation

In order to assess whether a particular fault or fault network has a capacity for storing fluid, the geologic properties and structure of the area of interest must be well understood. Seismic reflection data provides this knowledge (see Figure 1), allowing for geoscientists to acoustically image the crustal structure of an area and further interpret that area's sedimentary, magmatic, and tectonic systems (Wrona et al., 2021). With the advancement of 3D seismic technology, large subsurface geologic structures can be imaged with a high resolution (Cartwright & Huuse, 2005). In order to improve seismic reflectivity data's capability for target investigation, it is common to enhance the data by deriving or extracting a separate mathematical interpretation of the data known as a seismic attribute. For targets like faults that show up as large discontinuities in seismic images, Marfurt semblance-based coherency has proven to be a powerful attribute for the manual interpretation of these structures (Marfurt et al., 1998).

In recent years, the field of seismic interpretation has been introduced to machine learning, a computing approach that has grown increasingly popular for analysing large datasets, such as 3D seismic reflectivity. With the ability to efficiently investigate large amounts of data, machine learning techniques may be able to automate the process of manually interpreting seismic images. In geosciences, machine learning has already been used to automate the interpretation of some geologic structures, such as the detection of salts, the mapping of geologic horizons, and the classification of different sedimentary layers (Wrona et al., 2018). Likewise, these approaches have also been applied to detecting the locations of faults within a 3D volume (Wrona et al., 2021; Cunha et al., 2020; Lewis et al., 2017).

Background: Full-Waveform Inversion

Additionally, full-waveform inversion (FWI) is a technique within seismic imaging that uses data-fitting and full-wavefield model to recover increasingly high-resolution velocity data, and recently, advances in computing have allowed for FWI to be done in 3D (Virieux et al., 2009). FWI thus provides additional information for the investigation of the properties of subsedimentary targets, like faults and domes, to seismic reflection images alone (Morgan et al., 2013). The capabilities of FWI have even been displayed recently in seismic interpretation studies in combination with reflectivity data and machine learning approaches to more accurately resolve seismic horizons, specifically salt boundaries, in 3D volumes (Lewis et al., 2017). However, FWI has yet to be used to assess a particular fault's or network of faults' ability to store fluid.

Objectives

The primary goal of this study is to improve upon the workflow for assessing various faults' capacity to effectively store fluid in a large, subsurface region. This would be accomplished by both: (i) creating a machine learning model that is able to accurately extract faults from 3D seismic reflectivity volumes using semblance-based coherency, and (ii) combining the extracted fault volume resulting from (i) with FWI velocity data recovered from the same area. Currently, most workflows in industry for the automated interpretation of seismic volumes are kept private or commercially-leased, as is also the circumstance for many datasets, including both seismic reflectivity and FWI data (Wrona et al., 2021), required to perform the fluid storage screening procedure proposed here. However, both the seismic reflectivity and FWI data used in this study are freely available (Mattos et al., 2016), and are included in an open-source repository that also shares all workflows, scripts, and necessary packages used throughout (See *Code Metadata*).

The region investigated in this report is an area within the Barents Sea known as the Samson Dome. Both two datasets used, seismic reflectivity and FWI, span the same region of 644 inlines (sampled once every 25 m.), 1549 xlines (sampled once every 12.5 m.), and 611 depth slices (sampled once every 10 m.). This region is precisely located at 72.14'40.9427"N, 24.00'40.6206"E (see Figure 1). This makes for volumes that represent 2,900 km.^{3}, requiring 2.68 gigabytes each of memory for storage, and was taken as a subset of the original seismic reflectivity and FWI volumes, which contain a region roughly 2.5 times larger.

The gradual rise of the large salt dome that dominates the region is believed to be the cause behind the large network of radial faults that have been observed in the same area (Mattos et al., 2016). It is due to this network of known faults, and the novelty of the available FWI data of the area, that this is an appropriate location to develop the proposed screening tool. As aforementioned, FWI's capability in assessing fault leakage has yet to be tested, but we believe that using the combination of tools described above, we may accurately be able to determine if a fault may act as a proxy for a well-sealed location (FWI high-velocity zones) or a conduit for fluid leakage (FWI low-velocity zones).

2. Methodology.

K-Means Clustering

The proposed solution of using a machine learning algorithm to separate pixels that correspond to fault locations from the rest of the seismic coherency image falls under the computer vision field of semantic segmentation. K-means was used in this screening tool because it has been proven to be a powerful technique for interpreting 3D seismic data (Xie et al., 2020 & Ettefagh et al., 2014). Additionally, due to K-Means being a machine learning technique that can be used to cluster pixels based on their value alone, this dataset does not require any manual labeling and may be fully unsupervised (Bholawalia et

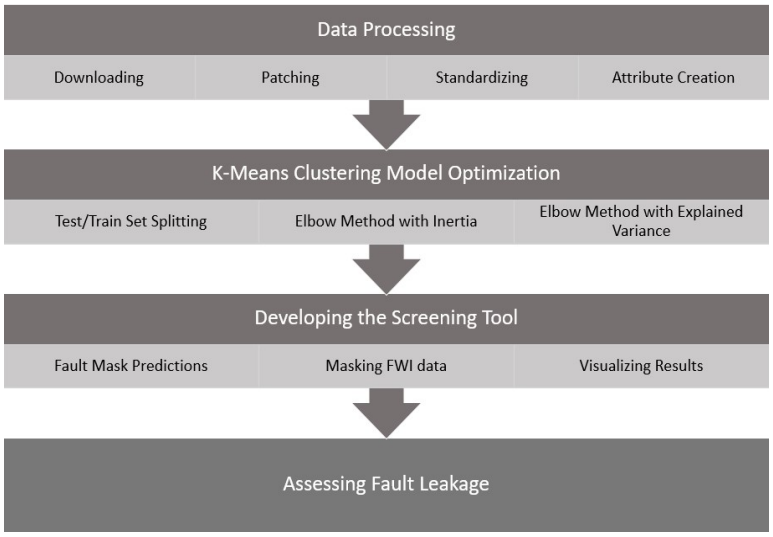

Figure 2. An overview of the workflow proposed in this study to assess fault leakage. Three major sections of work, Data Processing, K-Means Clustering Modeling, and Developing the Screening Tool are shown before being able to asses a fault's storage potential.

al., 2014). Although K-Means has been used in similar studies within geosciences to create solutions in imaging seismic profiles (Di Giuseppe et al., 2014), fault classification (Ettefagh et al., 2014), and detection of salts and other subsedimentary targets (Xie et al., 2020), using Marfurt semblance-based coherency in a K-Means algorithm to extract fault locations is a novel approach.

The specific K-Means algorithm used is that implemented by OpenCV (Bradski, 2000), and it is used to create a mask for each 2D seismic image in the dataset, with those 2D image outputs being able to be stacked back together and visualized in 3D. However, as the original horizontal image slices included too many pixels for OpenCV's K-Means implementation to process, each slice was patched into a collection of square images that would be clustered individually and repatched into one whole depth slice for final interpretation. Due to each pixel having a value that is independently evaluated by the model, the size of each image that the K-Means algorithm would see is arbitrary.

Training this unsupervised model consists of optimizing a few parameters of the K-Means algorithm. Three of the more straightforward parameters to optimize are the criteria that determine when the algorithm is satisfied with the values it has given to the pixels in the seismic image. The K-Means model has two criteria for when one run of the algorithm must complete - a maximum number of iterations and relative tolerance - which is a measure of the relative difference between images in each successive iteration, and these are the first two parameters constrained in this investigation. As the relative minimum that K-Means finds in each image can be largely dependent on a randomly determined initialization of the model, most implementations of K-Means, including the one used, perform a number of runs of the algorithm where the model is re-initialized after each run and only the model run that had the lowest average intra-cluster sum of squared distances (SSD) is selected. The number of times the model re-initializes is the final parameter that was optimized before the hyperparameter that is most influential in optimizing the K-Means model - the number of clusters the image pixels are separated into - was constrained (see *Results*).

Solution Environment

The current version of the screening tool exists as the final product of a standalone notebook, in which only file paths for desired 3D seismic reflectivity and 3D FWI volumes must be specified to the notebook scripts. After this, the notebook can be run end-to-end and completes the steps described below to produce a predicted 3D visualization tool to assess fault leakage. This notebook is able to be used locally on an Anaconda distributed environment (Jupyter), and therefore all of the programming conducted throughout this study used Python (v 3.8)(Van Rossum & Drake, 2009). An alternative version of the Python notebook that may be used on virtually distributed environments has been released. Included in this notebook is a suite of custom-built functions, all of which have had verified functionality by a custom-developed software testing package built using Pytest (Holger, 2004). Additionally, see *Code Metadata* for information on automated software testing released with the complete screening tool software package, as well as for more specific information on the specific software packages and versions that are necessary for deploying the proposed tool.

Pre-Conditioning

The proposed screening tool for assessing fault leakage begins with processing the original seismic data to prepare the images for a K-Means clustering model. Before any relevant operations were done with the 3D reflectivity data, it was standardized from its original value range of 0-255 to a range of 0-1. This is done to reduce the variance within the raw data, therefore re-

sulting in more precise numerical interpretations in the following steps. In order to evaluate the proficiency of the K-Means model and optimize it's performance, it was necessary to split the seismic images into testing, training, and validation sets. As the quality of data varies throughout the volume due to the data collection method, it was determined that the model would be trained on the bottom ~71% (1,770 m.-6110 m.) of the volume, and the top ~9% (0 m.-550 m.) of the data was kept unseen completely for potential future manual interpretation as the validation set. The ~20% (550 m.-1,770 m.) of the volume in between the training and validation sets was chosen to be the testing set as this is the region of the volume that has the highest quality data, and to prevent model bias towards relatively higher quality images, this region of the volume was kept unseen by the model until the concluding stages of the investigation. While the location of the highest quality data was previously known, this was confirmed to be an ideal testing subvolume by measuring model proficiency relative to volume depth (see Figure 3).

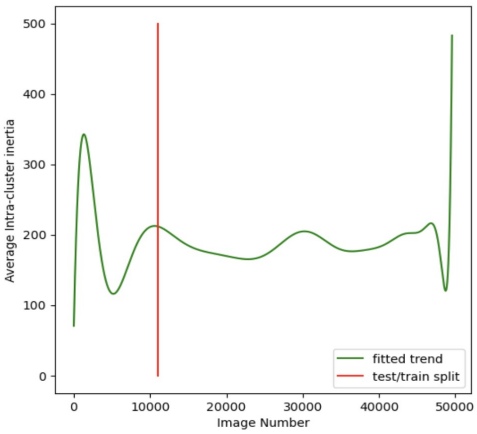

Figure 3. The average inertia across each image (ordered sequentially by depth) is shown. In green. The red line indicates the split between the sets.

Rather than using the raw reflectivity images, the model was trained on a volume derived from the previously mentioned Marfurt semblance-based coherency, which uses a moving window to assign individual pixel values based on the relative similarity/dissimilarity of the surrounding pixels in 3D (Marfurt, 1998). This is advantageous to a 2D seismic attribute as it is able to assign values to a target based on its shape in all directions, and targets - like faults - that cause more prominent discontinuities in 3D than 2D are revealed through this derivation. Figure 4 below shows the transition of a 'raw' seismic image (4a) to its Marfurt semblance interpretation (4b). The algorithm that extracts this attribute is a modified version of the algorithm constructed by Kington, J. (2016). The coherency algorithm was modified, at a computational cost, so that the moving window would see a larger 3D subvolume to assign individual pixel values in an effort to differentiate more extremely the values assigned to pixels in faults and in channels, which are two targets that share common value ranges in this attribute.

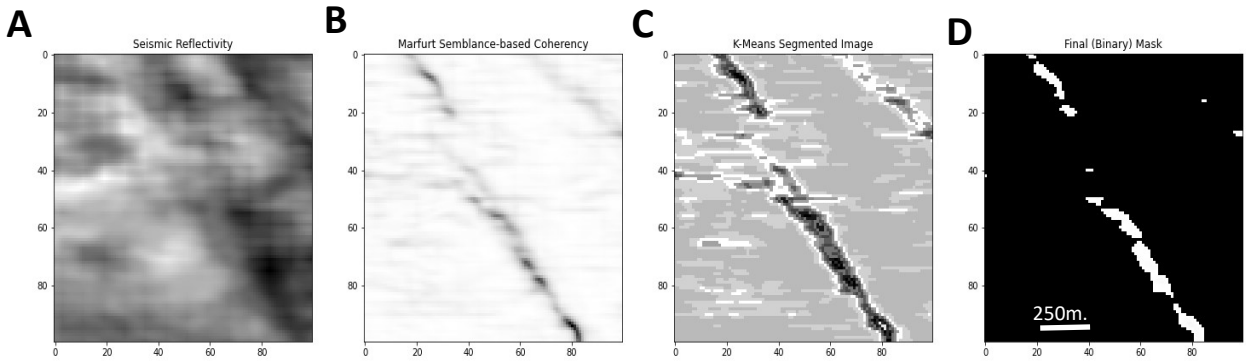

Figure 4. The transformation of a single 1.25 km. x 1.25 km. 2D image patch from seismic reflectivity (a), to its Marfurt semblance-based coherency interpretation (b) before being introduced to the K-Means model. Further, the highest confidence K-Means prediction of each pixel's cluster (c) is displayed with the final prediction(d) of which pixels belong to the image background (in black) and image foreground (in white). This specific patch displays an obvious fault at a depth of 1200 m. The x– and y-axes in each image simply show the number of pixels in the image (100x100).

3. Code Metadata.

Programming Language	Python	v. 3.x
Operating Systems	Mac, Windows, Linux	Latest versions*
Validated Notebook Platforms	Local: Jupyter Notebook Virtual: Google Colab	Latest versions*
Validated Local Environment Distributor	Anaconda Software Distribution (2021)	v. 4.11.0
Required Software Libraries	Required Libraries	Versions within.
Current Notebook	Screening_Tool	v. 1.0
Complete Software Pack-	Github Repository	v. 1.0

4. Results.

Model Training and Validation

As an unsupervised machine learning approach, typical cross-validation techniques are not an appropriate measure of model error, given that there are no ground truths for the predicted images to be evaluated against (Kodinariya et al., 2013). While the K-Means algorithm aims to minimize the mean squared error (MSE) of image pixels, the optimal MSE for unsupervised K-Means is unknown, with models approaching MSE's too low or high relative to the unknown optimal MSE being subject to overfitting and underfitting, respectively. With these challenges, K-Means models are optimized according to a process popularly referred to as the 'elbow method', which is a visualization of how the model performs for a desired evaluation metric against the number of clusters in the model. This determines how many clusters are best for segmenting the image, and for this study, which number of clusters is best at differentiating faults from other targets in a seismic image.

The most popular evaluation metric to use in the elbow method is inertia, which is a measure of the average SSD of each observation to the other pixels within its assigned cluster (Di Giuseppe et al., 2014 & Xie et al., 2020). The results of this evaluation metric across all images in the training and testing for 15 cluster values between 5-30 clusters are shown in Figure 5. A numerical interpretation of the elbow method implies the optimal number of clusters is the cluster-inertia value pair that has the largest distance from the average trend in SSD across all cluster values (Kodinariya et al., 2013). As shown, this gives an initial interpretation that 10 clusters is the optimal value for this K-Means model. Additionally, average inertia results for both the training and testing set followed similar trends, implying the results of this analysis were not biased by the location of the data. However, the relative smoothness of both trendlines describing inertia in Figure 5 raises potential ambiguity on whether any conclusions can be drawn from this model.

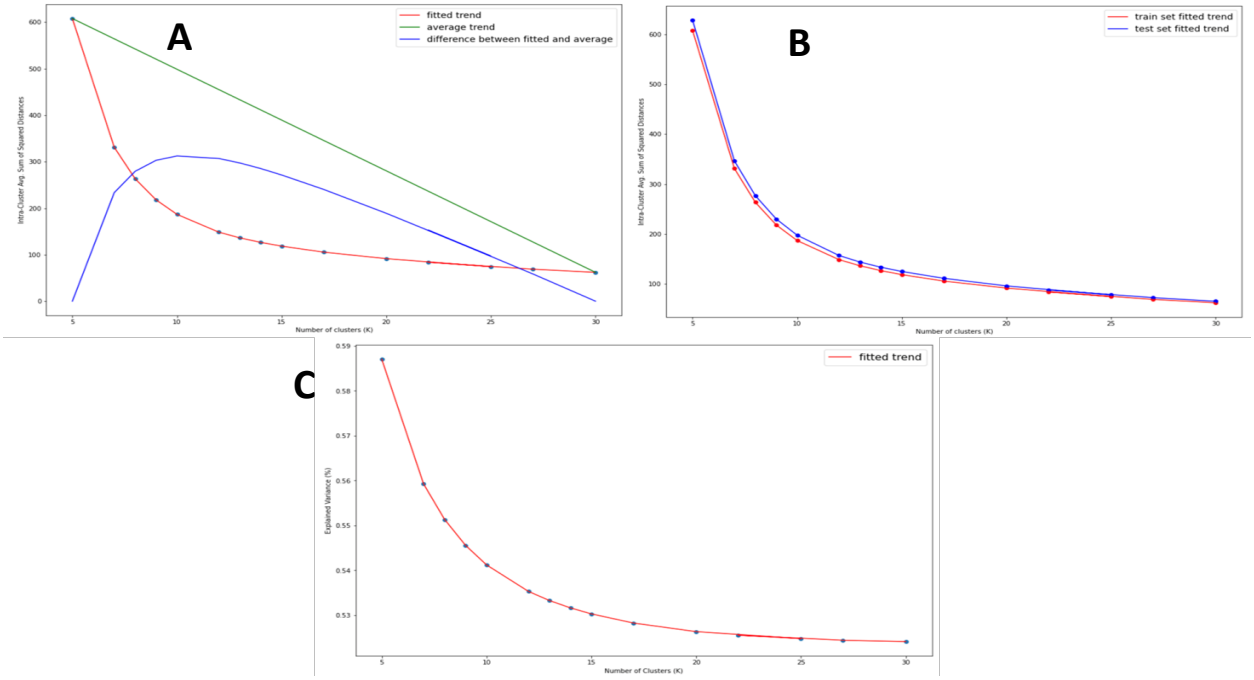

Figure 5. A.) The elbow method for the training data is shown. The red line show the fitted trend of inertia again cluster values, and the green shows the average trend. The blue line is the mathematical interpretation of the elbow-method, with the largest distance between lines being the optimal K. **B.)** The fitted trend of inertia against cluster values for both the training and testing set are shown. Similar trends indicated the model is not biased. **C.)** The elbow-method for explained variance is shown. Here, there is no upwards slanting elbow that would create a clear best cluster value.

This leads to the investigation of the second metric used to evaluate this model, explained variance. A higher score for this metric means that the model is able to explain the variation in the original data more proficiently. Explained variance is most easily explored by performing silhouette analysis of each predicted image. Silhouette analysis is a measurement of the previously described inertia of an image relative to the mean SSD of each observation in the image to all pixels in the neighbouring cluster, rather than the same cluster (Kodinariya et al., 2013 & Xie et al., 2020). This analysis provides a single score that is equal to the explained variance for each predicted image. Because K-Means algorithm assigns each cluster with equal variance, the silhouette score obtained for an entire image is also the explained variance for each cluster in the predicted image. With this, the optimal cluster value for this K-Means model can be found simply as the cluster value with the highest average explained variance.

However, as shown in Figure 5, the cluster value with the highest explained variance is 5. While the variance did not peak at the same cluster value that was predicted to be the best value with the elbow method with inertia, this trend in explained variance helps to reach the optimal model. We are now able to see that a cluster value of 10 serves as the optimal tradeoff between desirable explained variance and desirable inertia values. With the optimal model being selected, predictions of the test set images are created for model validation, with results for both inertia and explained variance also being shown in Figure 5. The test set's similar trend for both of these evaluation metrics confirm that this will be the finalized model used in this fault leakage assessment tool.

Model Parameter	Optimized Value
Maximum # of Iterations (Stopping Criteria)	10
Tolerance (Stopping Criteria)	1e-4
# of algorithm attempts (re-initialisation)	20
# of clusters – K	10

Assessing Fluid Storage Potential in Faults

After the parameters of the K-Means algorithm are finalized, a binary mask is created for each seismic image. This mask is not simply the image of each pixel with its value according to its assigned cluster, but a confidence prediction (based on thresholding inertia) of all clusters that either belong to the image background or image foreground (Bradski, 2000)(see Figure 1c-d). The image foreground, or pixels that end up receiving a value of 1 in the binary mask, is the part of the image we have trained our model to extract as a predicted fault. Through visual investigation of the predicted volume, it was found that in images where there are obvious faults relative to the rest of the image, like in Figure 6, the model performed well. However, in images where there are no faults, the model predicts noisy (large inertial values) masks with little interpretive use. Overcoming this challenge can be addressed by more thresholding, where predicted masks with high inertial values can be concluded as images with no obvious faulting and weighted accordingly when combined with FWI-recovered velocity data for fluid storage interpretation.

The predicted fault locations can now be combined with the FWI-recovered velocity volume. Figure 6 demonstrates the transformation of these two 3D volumes into a single interpretive volume, which can be achieved simply by multiplying the two datasets together. This results in no information being brought forward from the image background, while the velocities within the predicted fault locations are revealed. This allows for an automated initial interpretation of leakage potential within faults like the example shown below, where the user can see that this fault is a low velocity area (light blue), aside from a small patch with high velocity (red) data. Additionally, we can see that the areas surrounding this fault have a relatively higher velocity, which may indicate that the fault is an area of relatively high porosity compared to the surrounding rock. Based on this, FWI and the K-Means model have been able to predict and show that this particular fault is likely formed in a way that would allow for any potential injected fluid, like CO_2, to freely escape. This process of interpretation must be done at all depths within the predicted fault locations to conclude the fault's storage suitability. As well, the confidence of this prediction is evaluated once more by calculating the explained variance of the image, and this fault has a score of only 0.54 percent explained variance.

The confidence of the resulting screening tool is thus evaluated by examining the explained variance of the entire dataset. Up to this point in the study, 100x100 pixel patches (1.25 km. x 1.25 km.) of horizontal depth slices, like those seen in Figure 4 and Figure 6, have been predicted by the model. During potential future stages of evaluation, this also allows for a more proficient determination of the 3D confidence volume obtained by the model corresponding to the predictions made in the final interpretive volume.

The final stage of development for the screening tool consists of converting the image

patches back into a single 2D image and reshaped to the original dimensions. The data is now stored as a 3D array and is able to be displayed directly by Python visualization software package Mayavi (Ramachandran, P. and Varoquaux, G., 2011), which is able to display an interactive cross-section of the 3D volume for user interpretation. This allows for manual verification of the screening steps described above with a 2D image patch by enabling the user to interpret the viability of a fault as a fluid storage location at all predicted locations within the fault.

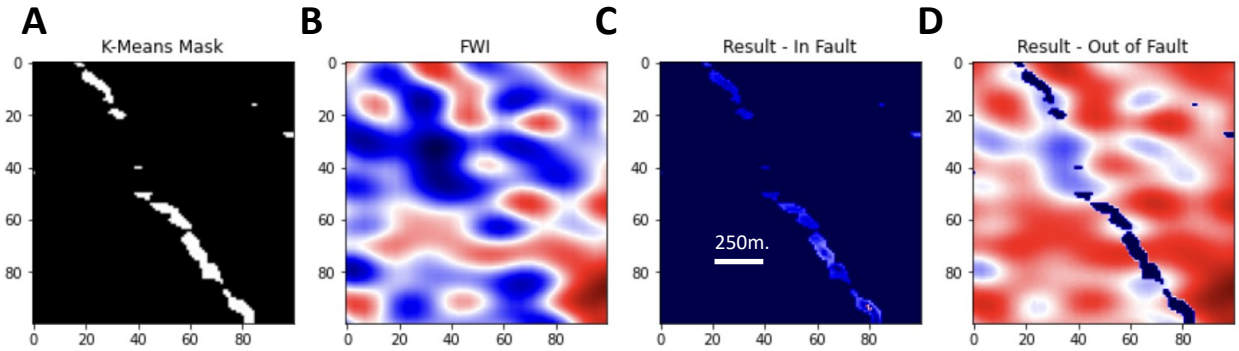

Figure 6. A.) K-Means predicted fault locations binary mask. **B.)** FWI-recovered velocity data **C.)** Using the mask to reveal the velocities of zone within the predicted fault location. **D.)** Using the mask to reveal the velocities of zones outside of the predicted fault location.

5. Discussion.

Tool Limitations

As the tool is developed in a notebook environment. and the datasets being used throughout this study are very large, there are some limitations on how much computing this tool can handle at one time. As well, in order to complete the motivation of this study, basic tools and methods used throughout this process were not optimized, and were therefore not successfully developed to be efficient. The time it takes for many of the in-built methods critical to the screening tool's workflow was investigated and documented, which can be found in full at the Github link in *Code Metadata*. With both the large amounts of memory required from the datasets used and/or created, and the time it takes to perform model optimization, it is recommended that future uses of this workflow on other datasets perform model training with high-performance computers or in batches of the desired set.

Further Steps

This workflow is only a brief exploration of how modern practices for assessing locations within a fault network for CO_2 storage could benefit from the proposed screening tool. This tool uses both a K-Means machine learning model to automate the process of manually interpreting large seismic reflectivity (potentially removing the need for corresponding local drill-core analysis) and FWI recovered velocity datasets to reveal areas within a region of the Samson Dome (location). As this is a novel approach, the primary investigation of this initial workflow was to explore the capabilities of FWI in this method. Visually, FWI looks to be a promising feature of the fluid storage screening process, which warrants further investigation with this high-resolution velocity data. However, because FWI is not introduced until the relative back-end of this workflow, the proficiency of this screening tool largely relies on the K-Means model's ability to accurately extract faults.

Performing K-Means semantic image segmentation to interpret seismic images with Marfurt semblance-based coherency is also a novel approach, and this proved to be satisfactory selection of a seismic attribute to pair with FWI velocities. Additionally, a strategy for evaluating the K-Means model was successfully derived and the optimal parameters for fitting this model to the Samson Dome dataset were found. However, there is ample opportunity for further investigation into how K-Means unsupervised clustering could be optimized for fault location extraction. While Marfurt coherency was strategically selected based on its merit (Marfurt et al., 1998), there are a wide array of existing seismic attributes that may be used in this investigation (Kington, J., 2016). These attributes may be used as the lone feature in K-Means clustering as described in this investigation, or may be used together to create an image with multiple feature dimensions for clustering. This may help address many of the challenges the model had in distinguishing data points located in faults from other targets in the image foreground, distinguishing data points in non-fault targets (like seismic channels) from the image background, and maximizing the explained variance in parts of the seismic data that may not include significant faulting.

The most significant step that can be taken to truly test the capability of this screening tool would be to manually interpret the same area using the same seismic reflectivity and FWI velocity data investigated in this study. Unfortunately, this tool was developed in large part to potentially remove the time needed for manual interpretation, and performing this independent analysis fell outside the scope of this study. Notably, the shallowest 650 m. (~5%) of this volume have been kept unseen by the model for this investigation in the future, although a portion of this data may be unusable due to poor data quality through collection. Also, this workflow is completely transferrable to other seismic regions that have corresponding reflectivity and FWI-recovered volumes as the combination of an unsupervised approach and independence in the seismic data prevents model bias (Cunha et al., 2020), which was confirmed by similar trends in the training and testing set in the elbow method.

6. Conclusions.

Carbon capture and storage technologies remain at a stage of uncertainty when evaluating subsurface prospects for CO_2 storage, and urgency is needed to address this uncertainty so that the rate of sustainable CO_2 injection may reach a satisfactory level (Lane et al., 2021). This study shows that the challenges this field of seismic interpretation faces currently through modern practices may be alleviated by the proposed workflow. Both next-generation FWI velocity data and Marfurt semblance-based coherency gave promising initial results, but more extensive manual interpretation and statistical investigation is required to fully conclude the effectiveness of the resulting screening tool.

References.

1. Haibin Di, Muhammad Shafiq, Ghassan AlRegib, Multi-attribute *k*-means clustering for salt-boundary delineation from three-dimensional seismic data, *Geophysical Journal International*, Volume 215, Issue 3, December 2018, Pages 1999–2007, https://doi.org/10.1093/gji/ggy376

2. Gibbins, J., & Chalmers, H. (2008). Carbon capture and storage. Energy policy, 36(12), 4317- 4322

3. Boot-Handford, M. E., Abanades, J. C., Anthony, E. J., Blunt, M. J., Brandani, S., Mac Dowell, N., ... & Fennell, P. S. (2014). Carbon capture and storage update. Energy & Environmental Science, 7(1), 130-189.

4. Mattos, N. H., Alves, T. M., & Omosanya, K. O. (2016). Crestal fault geometries reveal late halokinesis and collapse of the Samson Dome, Northern Norway: Implications for petroleum systems in the Barents Sea. Tectonophysics, 690, 76-96.

5. Wrona, T., Pan, I., Bell, R. E., Gawthorpe, R. L., Fossen, H., & Brune, S. (2021). 3D seismic interpretation with deep learning: A brief introduction. The Leading Edge, 40(7), 524-532.

6. Morgan, J., Warner, M., Bell, R., Ashley, J., Barnes, D., Little, R., ... & Jones, C. (2013). Next generation seismic experiments: wide-angle, multi-azimuth, three-dimensional, full-waveform inversion. Geophysical Journal International, 195(3), 1657-1678.

7. Wrona, T., Pan, I., Gawthorpe, R. L., & Fossen, H. (2018). Seismic facies analysis using machine learning. Geophysics, 83(5), O83-O95.

8. Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. Geophysics, 74(6), WCC1-WCC26.

9. Di Giuseppe, M. G., Troiano, A., Troise, C., & De Natale, G. (2014). k-Means clustering as tool for multivariate geophysical data analysis. An application to shallow fault zone imaging. *Journal of Applied Geophysics*, *101*, 108-115.

10. Ettefagh, M. M., Ghaemi, M., & Asr, M. Y. (2014, June). Bearing fault diagnosis using hybrid genetic algorithm K-means clustering. In *2014 IEEE international symposium on innovations in intelligent systems and applications (INISTA) proceedings* (pp. 84-89). IEEE.

11. Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, *105*(9).

12. Lewis, W., & Vigh, D. (2017). Deep learning prior models from seismic images for full-waveform inversion. In *SEG technical program expanded abstracts 2017* (pp. 1512-1517). Society of Exploration Geophysicists.

13. Cunha, A., Pochet, A., Lopes, H., & Gattass, M. (2020). Seismic fault detection in real data using transfer learning from a convolutional neural network pre-trained with synthetic seismic data. *Computers & Geosciences*, *135*, 104344.

14. Lane, J., Greig, C., & Garnett, A. (2021). Uncertain storage prospects create a conundrum for carbon capture and storage ambitions. *Nature Climate Change*, *11*(11), 925-936.

15. Cunha, A., Pochet, A., Lopes, H., & Gattass, M. (2020). Seismic fault detection in real data using transfer learning from a convolutional neural network pre-trained with synthetic seismic data. *Computers & Geosciences*, *135*, 104344.

16. Xie, Y., Tian, J., & Zhu, X. X. (2020). Linking points with labels in 3D: A review of point cloud semantic segmentation. *IEEE Geoscience and Remote Sensing Magazine*, *8*(4), 38-59.

17. Marfurt, K. J., Kirlin, R. L., Farmer, S. L., & Bahorich, M. S. (1998). 3-D seismic attributes using a semblance-based coherency algorithm. *Geophysics*, *63*(4), 1150-1165.

18. Cartwright, J., & Huuse, M. (2005). 3D seismic technology: the geological 'Hubble'. *Basin Research*, *17*(1), 1-20.

19. Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

20. Anaconda Software Distribution. *Conda*. Version 4.11.0, Anaconda,Oct. 2021.

21. Kodinariya, Trupti M., and Prashant R. Makwana. "Review on determining number of Cluster in K-Means Clustering." *International Journal* 1.6 (2013): 90-95.

22. Ramachandran, P. and Varoquaux, G., `*Mayavi: 3D Visualization of Scientific Data*` IEEE Computing in Science & Engineering, **13** (2), pp. 40-51 (2011)

23. Kington, J. *Semblance, Coherence, and other Discontinuity Attributes*. Github. (2016)