Visualization of the sequestered carbon-dioxide plume in the subsurface using unsupervised learning

Keyla Gonzalez[1,1] and Siddharth Misra[1,1]

[1]Texas A&M University

November 30, 2022

Abstract

Subsurface sequestration of carbon dioxide (CO2) requires long-term monitoring of the injected CO2 plume to prevent CO2 leakage along the wellbore or across the caprock. Accurate knowledge of the location and movement of the injected CO2 is crucial for risk management at a geological CO2-storage complex. Conventional methods for locating/assessing the injected CO2 plume in the subsurface assume a geophysical model, which is specific and may not be applicable to all types of CO2-injection reservoirs and scenarios. We developed an unsupervised-learning-based visualization of the subsurface CO2 plume that adapts and scales based on the data without requiring an assumption of the geophysical model. The data-processing workflow was applied to the cross-well tomography data from the SECARB Cranfield carbon geo-sequestration project. A multi-level clustering approach was developed to account for data imbalance due to the absence of CO2 in the large portion of the imaged reservoir. The first level of clustering differentiated CO2-bearing regions from the non-CO2 bearing regions and achieved a silhouette score of 0.85, a Calinski-Harabasz index of 160666, and a Davies-Bouldin index of 0.43, which are indicative of high quality, reliable clustering. The second level of clustering further differentiated the CO2-bearing regions into regions containing low, medium, and high CO2 content. Overall, the multi-level clustering achieved a silhouette score, Calinski-Harabasz index, and Davies-Bouldin index of 0.74, 59656, and 0.32, which confirm the high quality and reliability of the newly proposed unsupervised-learning-based visualization. Three distinct clustering techniques, namely k-means, mean-shift, and agglomerative, generated similar visualizations. In terms of the adjusted Rand index, the similarity of clusters identified by the three distinct clustering techniques is around 0.98, which indicates the robustness of the cluster labels assigned to various regions of the CO2-injection reservoir. Further, we find certain geophysical signatures, such as Fourier transform and wavelet transform, to be highly relevant and informative indicators of the spatial distribution of CO2 content.

Visualization of the sequestered carbon-dioxide plume in the subsurface using unsupervised learning

Keyla Gonzalez^ and Siddharth Misra^*

^Harold Vance Department of Petroleum Engineering, College of Engineering, Texas A&M University, College Station, Texas, USA

**Department of Geology and Geophysics, College of Geosciences, Texas A&M University, College Station, Texas, USA*

Abstract

Subsurface sequestration of carbon dioxide (CO_2) requires long-term monitoring of the injected CO_2 plume to prevent CO_2 leakage along the wellbore or across the caprock. Accurate knowledge of the location and movement of the injected CO_2 is crucial for risk management at a geological CO_2-storage complex. Conventional methods for locating/assessing the injected CO_2 plume in the subsurface assume a geophysical model, which is specific and may not be applicable to all types of CO_2-injection reservoirs and scenarios. We developed an unsupervised-learning-based visualization of the subsurface CO_2 plume that adapts and scales based on the data without requiring an assumption of the geophysical model. The data-processing workflow was applied to the cross-well tomography data from the SECARB Cranfield carbon geo-sequestration project. A multi-level clustering approach was developed to account for data imbalance due to the absence of CO_2 in the large portion of the imaged reservoir. The first level of clustering differentiated CO_2-bearing regions from the non-CO_2 bearing regions and achieved a silhouette score of 0.88, a Calinski-Harabasz index of 271145, and a Davies-Bouldin index of 0.30, which are indicative of high quality, reliable clustering. The second level of clustering further differentiated the CO_2-bearing regions into regions containing low, medium-low, medium-high, and high CO_2 content. Overall, the multi-level clustering achieved a silhouette score, Calinski-Harabasz index, and Davies-Bouldin index of 0.68, 86750, and 0.46, which confirm the high quality and reliability of the newly proposed unsupervised-learning-based visualization. Three distinct clustering techniques, namely k-means, mean-shift, and agglomerative, generated similar visualizations. In terms of the adjusted Rand index, the similarity of clusters identified by the three distinct clustering techniques is around 0.98, which indicates the robustness of the cluster labels assigned to various regions of the CO_2-injection reservoir. Further, we find certain geophysical signatures, such as Fourier transform and wavelet transform, to be highly relevant and informative indicators of the spatial distribution of CO_2 content.

Keywords: *Carbon Sequestration; Unsupervised Learning; Statistical Tests; Clustering; Visualization*

1. Introduction

CO_2 concentrations have increased from 280 ppm (pre-industrial) to 410 ppm in 2018 (NOAA Earth System Research Laboratories, 2020). The role of geo-sequestration or geological carbon storage has been identified as an essential step towards reducing CO_2 emissions into the atmosphere. According to Rackley (2017), by 2050, capture and storage of 730-1100 Gt CO_2 are needed to be removed to maintain CO_2 concentrations below 450 ppm. And to have a perspective on this quantity, this is equivalent to emissions from 143 million vehicles for one year.

Geological carbon storage is an important technology for reducing the anthropogenic CO_2 content in the atmosphere. Numerous industries are developing technologies and strategies to mitigate the high CO_2 emissions specific to their industries. Carbon geo-sequestration involves the injection of CO_2, at supercritical conditions, into an underground geological formation. Geological storage sites include oil and gas reservoirs, unused saline aquifer, and coal seam. In particular, CO_2 enhanced oil recovery (EOR) and the injection of CO_2 into saline aquifers are promising solutions due to global storage potential and high economic value. EOR is considered a key methodology for CO_2 reuse along with storage, while saline aquifers display the potential to store one Mt of CO_2 per year. Hence, new CO_2 sequestration projects are needed to be developed to address this goal. In addition, investments in carbon capture and storage have increased, expecting to worth $3.5 billion by 2025 (Markets and Markets, 2020). This could establish the geo-sequestration and carbon capture industry as one of the most attractive ones.

Geological carbon sequestration involves four stages: (1) selection of a subsurface geological site, (2) CO_2 transport and injection into the geological site, (3) monitoring and verification of the CO_2 storage during the injection and over the long term, and (4) long-term risk management of the entire geological CO_2-storage complex. Each stage exhibits large uncertainty, requiring continuous monitoring and optimization. Machine learning tools could be implemented to improve the efficiency and efficacy of these four stages. Monitoring and verification of CO_2 movement in the subsurface can be classified based on a specific purpose, such as injection monitoring, CO_2 plume location and movement, or ground movement detection (Rackley, 2010). Lastly, a detailed risk assessment needs to be addressed to identify potential CO_2 leakage and prevent any escape over hundreds of years. Environmental impact and consequences of adverse leakage need to be evaluated for designing an appropriate mitigation procedure.

Diverse geophysical technologies are used in the lifecycle of CO_2 geo-sequestration. A detailed geophysical characterization is needed to define storage capacity, injection design (rates and pressures), and caprock integrity (Smit et al., 2014). After CO_2 injection starts, the location and movement of CO_2 become a vital factor for risk management. Carbon sequestration requires ongoing monitoring and verification of long-term storage. Time-lapse seismic, electrical tomography, and gravitational survey are some of the main technologies that can be used for tracking the evolution of the CO_2 plume (Rackley, 2010).

Geochemical and pressure monitoring along with high-resolution acoustic imaging can be used for detecting a potential leakage (Davis et al., 2019). Moreover, CO_2 leakage to the surface can be detected using airborne electromagnetic sensing, which provides a spectroscopy image of atmospheric gases (Rackley, 2010). Another important consideration is the potential ground movement during CO_2 injection due to excessive pressure buildup. Tiltmeters and satellite-based (InSar) tools are required to monitor the geomechanical deformation over time. These tilt measurements provide a high-resolution detection at microradian-scale, while InSar detects surface deformation at millimeter scale (Rackley, 2010). Similarly, geophones can be used for monitoring fault activation due to the pressure buildup, being this a vital aspect for geomechanical stability.

Effective geo-sequestration projects involve the selection of a suitable geological site, transport, injection, surveillance, verification, and assessment of long-term CO_2 storage. Due to the complexity of each process, large uncertainties are encountered in a geo-sequestration project. Major initiatives to improve includes the analysis of risk management for the development of leakage detection systems, work for effective CO_2 monitoring to understand the behavior of injected carbon, ground motion to predict the extent of pressure change and potential induced seismicity and focus on geochemical research to analyze the chemical interaction between CO_2 and in-site fluids.

1.1. Use of machine learning in carbon sequestration

Over the last few years, machine learning has served as a tool to assist the ongoing subsurface monitoring and injection process. Machine learning can address the high uncertainty in the long-term spatiotemporal evolution associated with carbon plume migration in the subsurface. Areas with potential growth involve computer vision and unsupervised learning approaches. The necessity for obtaining insights into CO_2 processes is detrimental to the feasibility of carbon storage, and this could be solved by the application of machine learning techniques. In addition, methodologies have been proposed to manage the urgency of rapid CO_2 knowledge, such as real-time visualizations, real-time forecasts, and rapid prediction models (National Energy Technology Laboratory, n.d.).

For real-time CO_2 plume visualization, MacLennan (2020) used deep learning on electromagnetic data. Real-time forecasting using machine learning help optimize storage capacity and fluid/rock contact, e.g. CO_2 content, reservoir pressure evolution, and plume migration. Haghighat (2013) reported a leakage detection system using trained neural networks for real-time location and quantification of CO_2. The use of pressure data provided direct information on pressure changes. Seismic, well logs, and core data can be integrated with rock physic models for CO_2 saturation prediction. Machine learning helps in the assessment of induced seismicity during CO_2 injection. For example, microseismic signatures can be used as a potential indicator of fault reactivation. As a result, risk systems can be implemented to deploy effective mitigation strategies under seismic hazards (He et al., 2020).

1.2. Motivation and originality of our study

Conventional method for locating/assessing the injected CO_2 plume in the subsurface assumes a geophysical model to assess the spatial distribution of CO_2 content. The assumed geophysical model may not be applicable to all types of CO_2-injection reservoirs and scenarios. We developed a novel and reliable unsupervised learning methodology, based on multi-level clustering, for the visualization of the CO_2 plume in the subsurface. This approach is adaptive and scalable without incorporating a pre-defined geophysical model. The new workflow was applied on the cross-well tomography data from the SECARB Cranfield project. Multi-level clustering approach was developed to account for data imbalance due to the absence of CO_2 in the large portion of the imaged reservoir. The silhouette score, Calinski-Harabasz index, and Davies-Bouldin index are used to assess the quality and reliability of clusters generated using multi-level clustering. Moreover, independent clusters were generated using three clustering algorithms, called k-means, mean-shift, and agglomerative, and then evaluate them in terms of adjusted Rand index and homogeneity to assess the similarity of clusters identified by the three distinct clustering techniques. These tests the robustness of the cluster labels assigned to various regions of the CO_2-injection reservoir. Finally, we develop a methodology to discover certain geophysical signatures that are relevant and informative indicators of the spatial distribution of CO_2 content in the subsurface. For a safe, long-term storage of carbon dioxide in the subsurface, there is an urgent need for discovering new geophysical signatures that facilitate real-time CO_2 visualization and CO_2 leakage detection. This first-of-its-kind study that extensively tests the reliability and robustness of unsupervised learning methods for detecting and identifying the injected carbon dioxide in the subsurface thousands of feet below the ground. The current study provides for the first time an extensive implementation of both multiple evaluation metrics and multiple clustering methods to reliably visualize the subsurface CO_2 plume. A similar visualization of fractures was achieved by Misra et al. (2019) and Chakravarty et al. (2021).

2. CO_2 plume in the Injection Reservoir

2.1. CO_2 injection and plume migration

CO_2 behaves as both liquid and gas at critical pressure (1070 psi) and critical temperature (87.8 °F). The injection of CO_2 is generally performed at a supercritical condition. As the CO_2 reaches the supercritical condition, the CO_2 starts behaving like gas with liquid density. The process of injection occurs through non-corrosive injection wells where materials need to be carefully picked to maintain the well integrity. Subsequently, the injected CO_2 is continuously monitored to assess the reservoir response and regular pressure changes. Pressure and temperature gauges are usually installed to surveille the injection progress and identify any potential well problems (Rackley, 2010).

The CO_2 plume is defined as the volume of carbon dispersed in the reservoir. CO_2 can be immiscible or miscible in presence of other fluids. For instance, water and CO_2 are

immiscible while CO_2 and natural gas are miscible. Under immiscible fluids, CO_2 needs to be injected at a higher-pressure rate to displace the in-situ reservoir content. Once the injection stops, CO_2 migrates to the base of the caprock.

Over a longer period, CO_2 is trapped by capillary forces preventing the movement of the carbon molecules across the caprock above the injection reservoir. Furthermore, the connate water starts dissolving the CO_2 due to chemical interaction, allowing more storage space in the rock; however, the dissolution is generally a slow process depending on the CO_2 and water ratio. These chemical reactions can also modify the porosity and permeability of the formation. Lastly, mineral reactions may occur decreasing the size and connectivity of the pores. In short, for effective and efficient CO_2 geo-sequestration, CO_2 could be trapped in the injection reservoir by four mechanisms: structural traps, capillary forces, solubility, and mineral reactions.

2.2. Need for CO_2 plume monitoring

Monitoring of CO_2 plume is crucial for tracking its movement and behavior in the subsurface. Monitoring confirms the injection process and CO_2 distribution. This is vital for risk assessment and mitigation strategies. For instance, the identification of early leakages could lead to the prevention of groundwater contamination. Reliable monitoring can also assess the effect of geomechanical changes and induced seismicity because the injection of CO_2 (at high-pressure rates) can enhance the movement of the subsurface, increasing the possibility of leakage. More specifically, the surveillance data permits the validation and modeling of CO_2 growth, allowing us to predict the behavior of long-term CO_2 storage. Besides, carbon storage becomes riskier over time. Data-driven frameworks could be established to reduce the uncertainty in plume location and migration.

2.3. Monitoring techniques

Geophysical data analysis has a crucial role to play in carbon storage. A diverse number of geophysical technologies can be used in the lifecycle of CO_2 monitoring. A wealth of knowledge already exists in this area due to their current application in the oil and gas industry, which was rapidly expanded to geo-sequestration. Monitoring and verification of CO_2 movement are classified according to their specific surveillance goal. For CO_2 plume, the measurement techniques are summarized in table 1, to understand the variability of techniques. Time-lapse seismic is considered the most effective tool due to the high contrast of CO_2 acoustic impedances. Pre- and post-injection seismic are commonly acquired to provide an image of the change of fluids over time.

Geophysical monitoring method	Physical Principle
Seismic	Time-lapsed seismic
	Crosswell seismic
	Vertical seismic profile
	Microseismic
Gravimetry	Time-lapsed gravimetry
Electric and electromagnetic	Electric resistance tomography
	Crosswell resistivity
	Electric spontaneous potential
Remote sensing	Satellite interferometry
	Airborne electromagnetic

Table 1. Geophysical methods for CO_2 plume monitoring according to their physical principles.

2.4. Crosswell seismic imaging

Crosswell seismic is an effective tool to monitor supercritical CO_2 movement. This technique involves the use of downhole array of seismic sources and receivers (figure 1). Both are placed in adjacent wells to transmit and capture high-frequency seismic waves. As the source and receivers move, the process is repeated multiple times to image the subsurface properties and their variations. The higher frequency wave transmission provides detailed information of thin reservoirs, having a thickness from 3 to 33 feet, at interwell distances of 33 to 330 feet.

Crosswell seismic data can be processed into two fundamental measurements: tomography and reflection imaging. Crosswell tomography uses direct-wave traveltimes to image the subsurface seismic velocity variation in the inter-well region. However, the cross-well traveltimes are a small fraction of the total seismic information recorded during a cross-well recording. Reflection imaging uses the later arriving reflection events extracted from the whole complex wavefield to map the presence of high-contrast inter-well and surrounding regions. Velocity variations in the inter-well region can serve as indicators of fluids movement. As the CO_2 is injected into a reservoir, the velocities of the regions containing CO_2 generally decrease. The differences between time-lapsed tomography can provide a direct image of the CO_2 plume. It also captures the degree of velocity change which can be associated with the concentration levels of CO_2 in the injection reservoir.

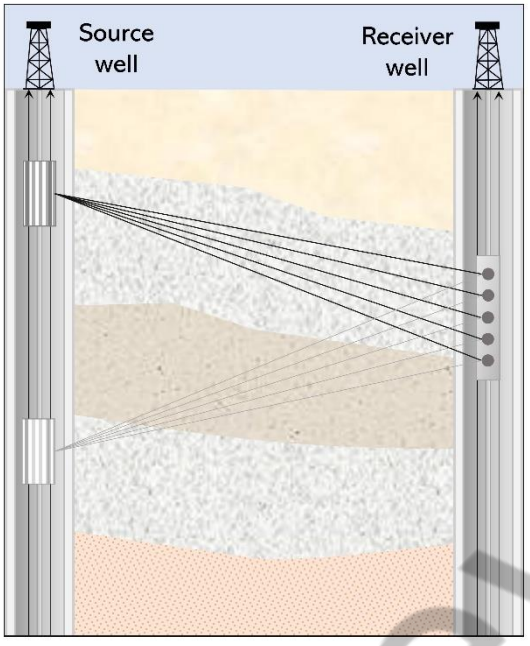


Figure 1. Crosswell survey scheme on a source-receiver profile where the transmissions of seismic waves from the source well are captured at the receiver well. This process is repeated as the seismic source and receivers move in both the wellbores resulting in a high-resolution mapping of the subsurface properties and velocity variations.

3. Methodology

3.1. SECARB Cranfield carbon geo-sequestration project

The Southeast Partnerships (SECARB) Cranfield Project was a commercial CO_2-EOR program located at Cranfield field in Mississippi. This project was designed to establish the feasibility of long-term CO_2 storage at low risk. It also serves to set up strategies for stacked storage, where EOR infrastructure can be used to inject CO_2 above and below EOR operations (Hovorka, 2013). The project started with the CO_2 injection in the fluvial sandstones of the Tuscaloosa formation on an unused saline aquifer.

The research was divided into four stages called: 1) phase-two, 2) high volume injection test, 3) detailed area of study, and 4) near-surface observatory area (Hovorka et al., 2011). The SECARB project conducted effective subsurface monitoring to evaluate the behavior and permanence of carbon dioxide. Different monitoring techniques were used according to their specific research goal. The Cranfield project focused its analyses on three main goals: risk management, CO_2 plume prediction, and pressure impact. Time-lapsed seismic, electromagnetic, and tracer chromatography measurements were techniques used for CO_2 plume prediction. For environmental assurance, pressure temperature, and groundwater and soil gas analysis were frequently applied (Hovorka et al., 2011).

3.2. Dataset description

Figure 2 displays the schematic representation of the well's location and acquisition design. The depth of interest corresponds to a range of 10,400 and 10,510 feet, where the supercritical CO_2 condition can be met. The dataset used in our study consists of two time-lapsed crosswell tomography that was collected between three wells at an interwell distance of 229 feet for the first profile (F1-F2) and 98 feet for the second profile (F2-F3). These profiles were acquired before and after the injection stage using a 10-level hydrophone array. The data was recorded on both profiles at a time difference of 10 months. The data acquired was processed into two components, reflection imaging, and tomography. The latter provides a seismic velocity map of the subsurface properties and in-situ fluids.

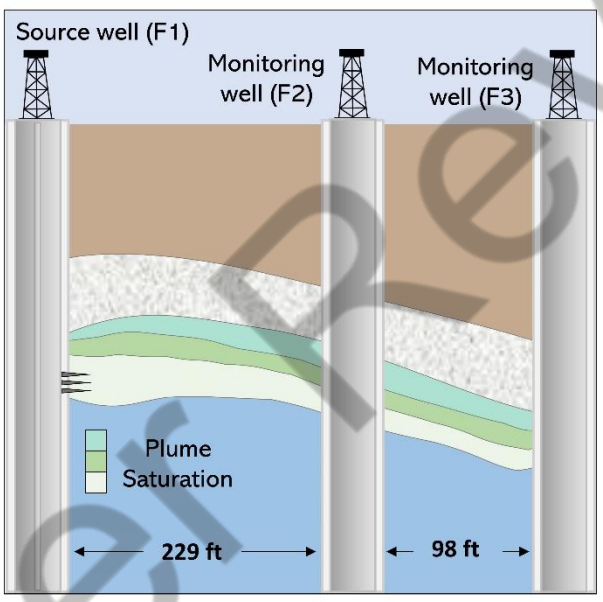


Figure 2. Schematic representation of the study area showing the side view (left) and top view (left) of the crosswell survey.

Due to the high correlation of seismic velocity and CO_2 content, the difference between the pre- and post-injection cross-well tomography is used to create the unsupervised learning workflow for the CO_2 visualization. The difference between the change of velocity under CO_2 injection is a strong indicator of the spatial distribution of CO_2 plume in the subsurface. Figure 3 displays the percentage change of velocity from the above-mentioned crosswell tomography difference. Values close to zero represent zero to low CO_2 content while values close to 14 indicate a high CO_2 concentration.

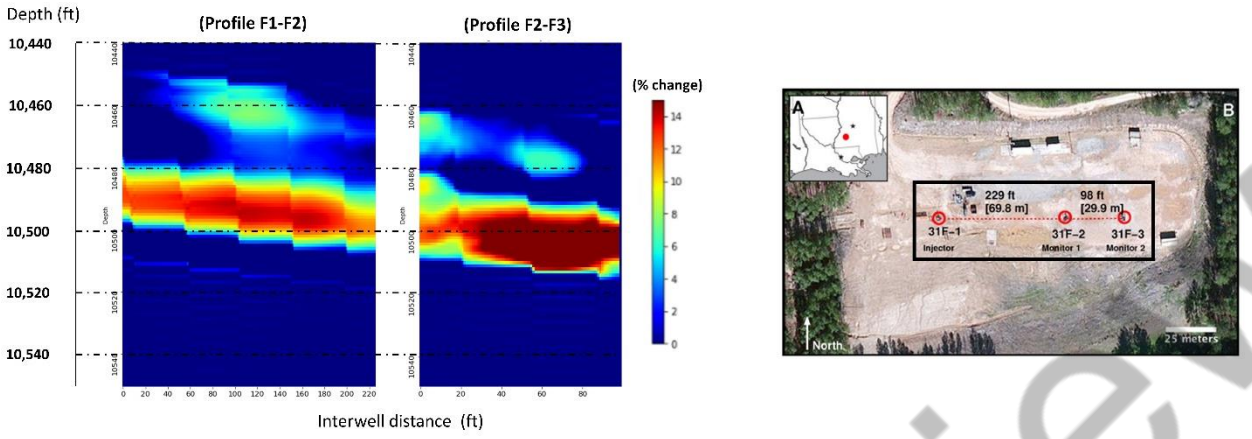


Figure 3. Left: Crosswell tomography image obtained after the data processing of pre- and post-injection profiles. Right: Study site of Cranfield field using one injection well (F1) and two monitoring wells (F2 and F3). Our analysis considers the entire region between F1 and F3 that includes the F1-F2 and F2-F3 profiles.

3.3. Unsupervised learning workflow for CO_2 plume visualization

A novel workflow is proposed for the visualization of CO_2 content and spatial distribution in the subsurface (figure 4). The workflow process cross-well tomography data to finally generate a cluster label for each region in the injection reservoir. The cluster labels represent the levels of CO_2 saturation at various locations ranging from negligible CO_2 content to high CO_2 content. First, the time-lapsed tomography data is converted into a change of velocity image. The image covers a depth range spanning from 10,440 to 10,550 feet, and an interwell distance of 229 feet between F1 and F2 and 98 feet between F2 and F3. The velocity change at any location is represented as pixel intensity values (figure 5). The velocity-change images contain 54776 pixels (F1-F2 profile) and 40145 pixels (F2-F3 profile). Pixels intensity close to zero are linked to low-velocity changes due to limited intrusion of CO_2 in that region, while intensity between 60 and 255 correspond to medium and high-velocity changes due to the intermediate amount of CO_2 intrusion. More details are provided in figure 5. The most important step is the creation of new geophysical signatures and statistical parameters that can be used to achieve a reliable and consistent spatial clustering of the injection reservoir based on CO_2 content. Similar robust workflows for regression tasks are presented by Li and Misra (2021) and Osogba et al. (2020) and those for classification tasks are presented by Ganguly et al. (2020). However, these models were developed for a known target where the labels are available for the model training. In particular, Li and Misra (2021) and Osogba et al. (2020) work do not require the extraction of features due to the acquisition of multiple well logs. On the other hand, Ganguly et al. (2020) propose a semi-supervised model where pixels were labeled for image segmentation. Another key difference is the resolution of the acquired datasets. The well-logs and SEM images are 2 cm - 2 m and 1 - 20 nm respectively while for the research crosswell seismic it is around 1 -100 m.

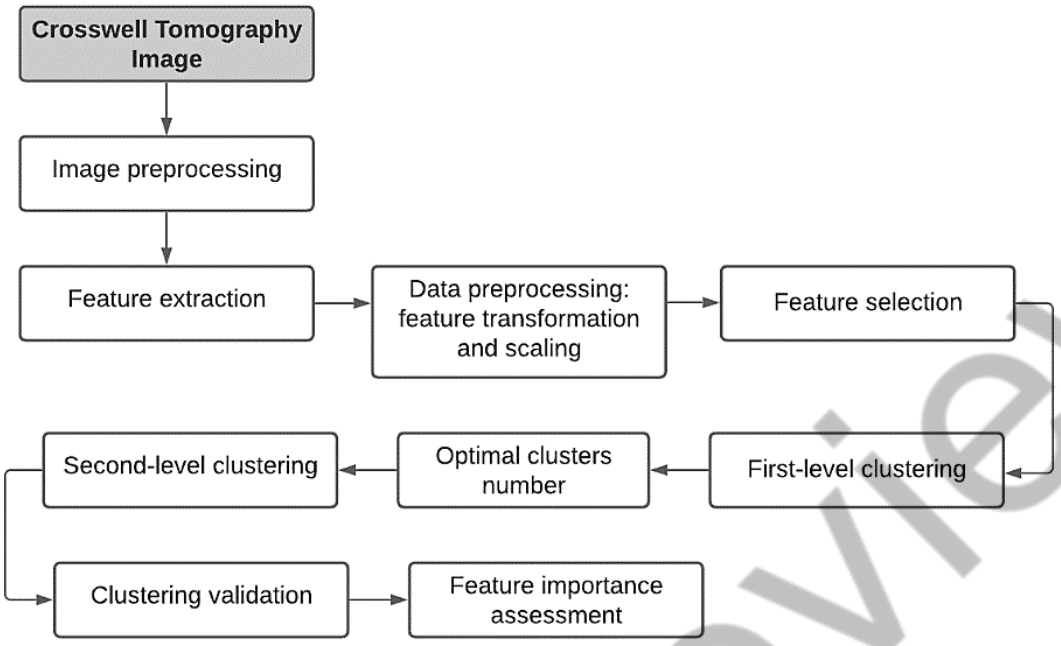


Figure 4. Flowchart of the proposed unsupervised learning-based visualization of the CO_2 plume in the subsurface.

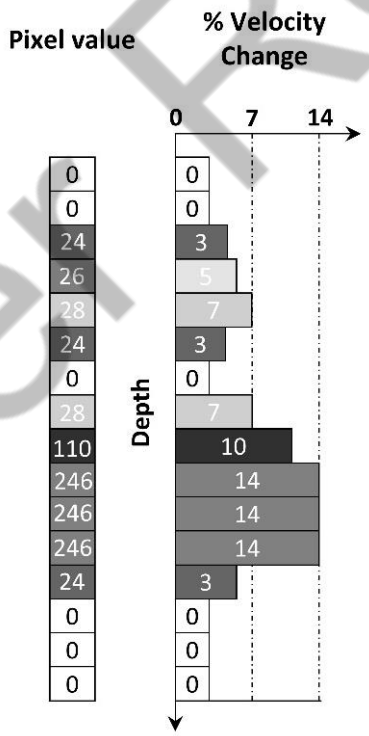


Figure 5. Representation of crosswell seismic tomography using pixel intensity. The velocity change at any location is due to the difference in velocity between pre- and post-CO_2 injection. Higher pixel intensity corresponds to higher CO_2 content, whereas pixel intensity of zero corresponds to negligible CO_2 content.

To identify the local information of a specific object, different aspects of an image need to be extracted. An image can be seen as a set of connected regions where unique characteristics are observed (e.g. shape, edges, intensity, texture, noise). The use of multiple features separates the CO_2 main attributes to process the clustering at a deeper and more efficient level. It incorporates all the extracted information on a single model, improving their accuracy and giving a more interpretable feature description. For purposes of feature engineering, the pixel intensities of each region and its neighboring pixels are processed using several feature extraction techniques. Informative, relevant, and independent features help build robust unsupervised learning models. Fourteen features were extracted from the velocity maps. Each extracted feature represents a specific characteristic of the velocity change in a region with respect to its neighboring regions. Table 2 compiles the extracted features and their descriptions. Robust scaler and power transformation were then applied on the features to achieve a Gaussian-like distribution and to standardize each feature within a unique, common range. Histograms and scatter plots were used to evaluate the pre-processing steps. Statistical tests were performed to select an appropriate set of features, which include those that exhibit low multi-collinearity with other features and high statistical importance for the desired spatial clustering. Efficacy of similar feature extraction in improving the robustness of the data-driven methods was demonstrated by Wu and Misra (2019) and Misra and Wu (2020). For this study, after the entire preprocessing step, the following nine features (also explained in Table 2) were used to build the two-level clustering approach for the desired visualization:

- Pixel intensity
- GLCM ASM
- GLCM correlation
- GLCM dissimilarity
- Local binary pattern
- Wavelet transform
- Fast-Fourier transform
- Edges
- Hxx (Hessian matrix)

Feature	**Description**
Gray-Level Co-Occurrence Matrix (GLCM)	Statistical analysis of spatial relations between pixels. Statistical methods include contrast, dissimilarity, homogeneity, energy, correlation, and ASM.
Fast-Fourier transform	Transformation of the image from spatial to the frequency domain. Low and high pass filters permit to pass certain image frequencies.
Linear binary pattern (LBP)	Texture operator which labels pixels based on the intensity of the central point
Sobel (Edges)	Gradient of pixels intensity for edge detection. It captures sharp changes in intensity due to even edges.

Wavelet transform	Time-frequency analysis for selection of suitable frequency band. It is commonly used to remove noisy signals.
Hessian matrix	Second-order derivative of the Gaussian kernel for region detector. It is applied in the Hxx, Hxy, and Hyy direction. It is suited for detecting local structures, like blobs and ellipsoids, where there exist odd edges.

Table 2. Brief description of the 9 features extracted from the pixel intensity representing the velocity change due to CO_2 injection. These features are used to build the two-level clustering approach for the desired visualization.

3.4. Unsupervised clustering

Unsupervised learning is a type of machine learning technique that enables us to detect unknown patterns and structures in the data and generate new insights from the data. Clustering is an unsupervised learning method that splits the dataset into clusters or groups. These clusters represent a subset of samples that belong to similar high-density regions in the feature space. In this study, k-means was used to accomplish spatial clustering. k-means successfully discovered hidden patterns of CO_2 presence, content, and distribution from the new geophysical signatures that were extracted from the map of velocity change. For a safe, long-term storage of carbon dioxide in the subsurface, there is an urgent need for discovering new geophysical signatures that facilitate real-time CO_2 visualization and CO_2 leakage detection. K-means enabled us to find patterns that can facilitate the CO_2 plume visualization. The clustering process begins by analyzing the preprocessed dataset, as shown in figure 4.

Imbalance in the data is a critical problem for any clustering method. Therefore, a novel approach was designed to handle the disparity of samples belonging to the various categories. In our dataset, as in most other CO_2 injection reservoirs, the volume of CO_2 corresponds to only a small portion of the reservoir volume. This results in data imbalance. To avoid the adverse consequences of the imbalanced dataset, we used a multi-level clustering to first distinguish the regions containing CO_2 from those without CO_2. These clusters can be further processed to identify regions having low, medium-low, medium-high, and high CO_2 content. K-means, agglomerative and mean-shift clustering methods were deployed to test the consistency and reliability of the clusters. Each clustering technique has a distinct underlying principle and assumptions. For instance, k-means clustering groups samples such that samples belonging to a cluster are closer to a common cluster centroid as compared to other cluster centroids. On the other hand, mean-shift clustering assumes that samples belonging to a cluster are closer to a common mode as compared to other modes. Further, agglomerative clustering recursively merges nearby clusters pairs into a hierarchical structure such that smaller clusters are merged into a bigger cluster. A comparative study on these methods was done by Chakravarty et al. (2021).

3.5. Validation of the Spatial Clustering

An important requirement for robust spatial clustering is to determine the number of clusters in the dataset. The number of clusters should be consistent and reliable. To that end, we used the elbow plot, silhouette score, Davies-Bouldin index, and Calinski-Harabasz index that confirmed the existence of five clusters in the dataset. The optimal cluster number is defined according to four scoring metrics with the purpose of generating dense and well-separated clusters. For instance, a silhouette score close to one indicates a perfect performance, while for Davies-Bouldin the best values are close to zero. An optimal number of clusters and consistency/reliability were also validated by evaluating the similarity of spatial clustering computed by three different clustering methods. K-means, agglomerative and mean-shift clustering were compared using the adjusted rand score and homogeneity score.

Finally, each cluster was analyzed to evaluate the CO_2 distribution and features importance. A frequency histogram was carried out to investigate cluster distributions and levels of CO_2 represented by each cluster. For the feature importance, different statistical tests were applied. The first test is the ANOVA or analysis of variance F-test that calculates the ratio of the variance of the group means to the within-group variances. A large value of ANOVA F-test indicates that the uniqueness of the feature. A second test was performed called mutual information. This analysis estimates the statistical dependence or joint probability between feature and target. In addition, Kendall's Tau correlation coefficient is estimated to measure the association between the cluster labels and the features. All these tests help us identify unique and important features for the desired spatial clustering. In other words, these tests quantify if the feature is significantly different among the clusters.

Unlike the above-mentioned tests, this section quantifies if a feature is significantly different between two cluster. Post-hoc test was performed to estimate the statistical difference of a feature between two distinct clusters. We used Tukey's honestly significant difference (Tukey's HSD) test that analyzed the mean differences of the feature values of the cluster means to determine the statistical differences between clusters. This test reveals which features strongly relate with the pair-wise distinctness of clusters. Tukey's HSD test is computed using the following equation:

$$HSD = \frac{M_i - M_j}{SE} \quad (1)$$

where M_i and M_j correspond to the means of two clusters, and SE is the standard error of the sum of means.

4. Two-level clustering

4.1. Design of the multi-level clustering

In this work, the two crosswell tomography were processed using a two-level clustering scheme to handle the imbalanced nature of the dataset, where there is more information from regions without any CO_2 content. An imbalanced dataset introduces a bias towards the samples belonging to the majority category. Hence, for the imbalanced dataset, a single-level clustering cannot find well-separated and dense clusters, i.e. high-quality clusters. We tested various clustering techniques, such as agglomerative, k-means, and mean-shift clustering. K-means was selected as the best clustering method for generating the final clusters, which qualitatively represent the CO_2 content (negligible, low, medium-low, medium-high, and high) in the CO_2-injection reservoir. Agglomerative and mean-shift clustering were used to validate the clustering results generated by the k-means clustering. All these clustering techniques are based on distinct mathematical/statistical assumptions and formulations. Hence, consistency among the clusters obtained from these distinct clustering methods ensures the reliability/quality of the clusters. To further assess the quality of the clusters, silhouette score, Davies-Bouldin index, and Calinski-Harabasz index were computed (see Appendix A for details) for both the levels of clustering. In general, such scores/indices represent the dissimilarity between clusters and the similarity within clusters.

The first level of clustering differentiates the regions that contain CO_2 from those that do not contain any CO_2 (figure 6). In other words, the first level of clustering is based on the presence or absence of CO_2 in a specific region of the CO_2-injection reservoir. The first-level clusters serve as inputs for the second level clustering. The second level of clustering was applied to regions where CO_2 is present, as indicated by the first level of clustering. Second-level clusters qualitatively represent four degrees of CO_2 content, i.e. low, medium-low, medium-high, and high (figure 6). Silhouette score, Davies-Bouldin index, and Calinski-Harabasz index enable the selection of the optimal number of clusters for each level of clustering. These scores/indices indicate that the best quality clusters are obtained when the first level of clustering finds two clusters and then the second level of clustering finds four clusters for the regions where CO_2 is present.

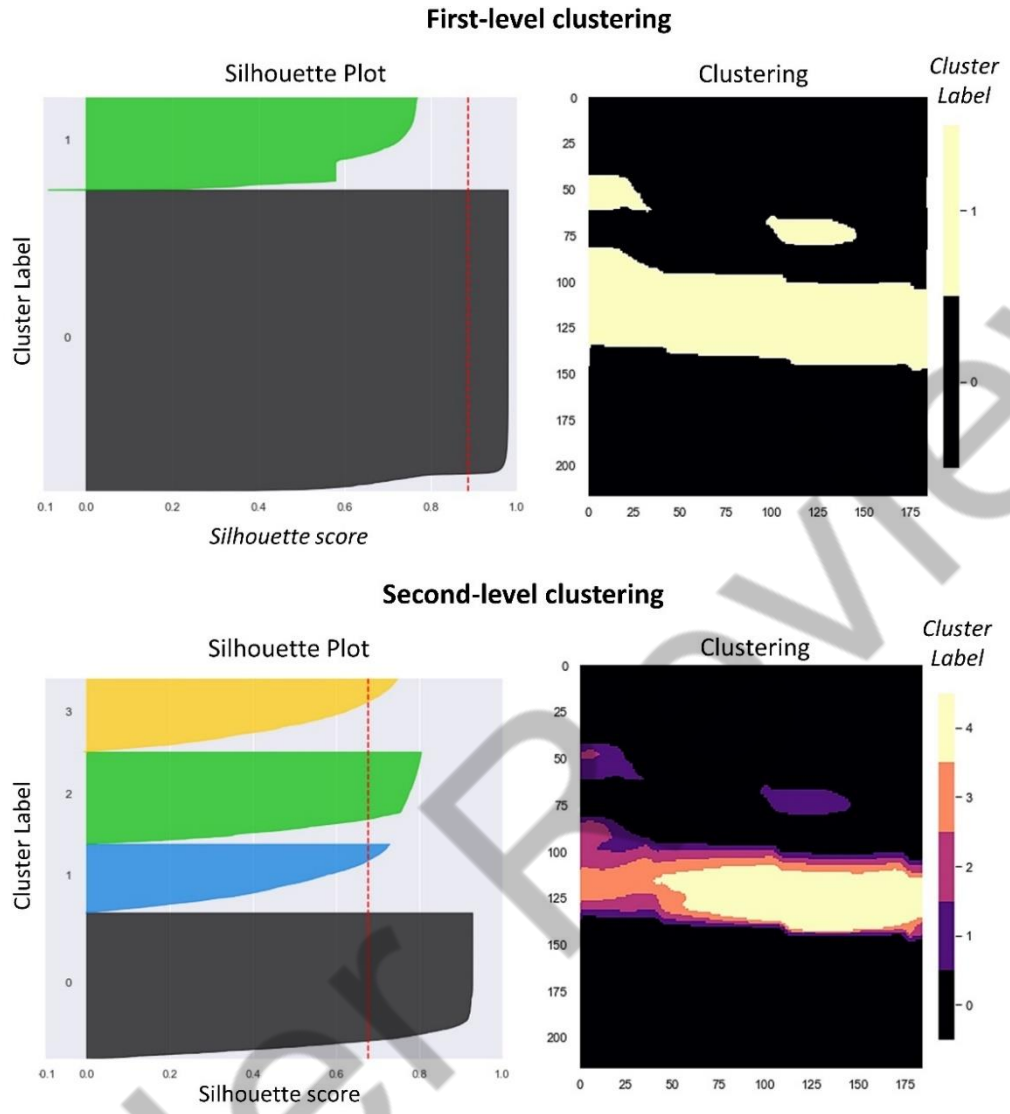


Figure 6. Top Left: Silhouette plot for two clusters obtained by the first level of clustering. Top Right: The two clusters correspond to regions where CO_2 is absent (cluster 0) and where CO_2 is present (cluster 1) as obtained by the first level of clustering. Bottom Left: Silhouette plot for four clusters obtained by the second level of clustering. Bottom Right: The five clusters correspond to regions where cluster 0, 1, 2, 3, and 4 represent negligible, low, medium-low, medium-high, and high CO_2 content, respectively, as obtained by the two-level clustering.

5. Results and discussion

5.1. Validation of clustering methods

We validate the cluster predictions by evaluating the densities and separations of clusters. To assess the clustering performance, silhouette scores, Davies-Bouldin index, and Calinski-Harabasz index were computed at the two levels of two-level clustering. The coefficients indicate the degree of similarity between and within clusters. A lower value of

the Davies-Bouldin index indicates a higher quality clustering, whereas for the Calinski-Harabasz index and silhouette score a higher value indicates a higher quality clustering.

At the first level, two clusters were established to detect the presence/absence of CO_2. As shown in table 3, the silhouette score for the cluster representing the absence of CO_2 is close to 1 and the median value of the silhouette score for the first level of clustering is 0.88. The Davies-Bouldin index values are close to zero, indicating accurate segregation of the regions based on the presence/absence of CO_2.

Profile	**Clustering level**	**Score**		
		Silhouette	*Davies-Bouldin*	*Calinski-Harabasz*
F2-F3	First-level clustering	0.88	0.30	271145
	Second-level clustering	0.68	0.46	86750
F1-F2	First-level clustering	0.89	0.24	573174
	Second-level clustering	0.59	0.51	51335

Table 3. Silhouette score, Davies-Bouldin index, and Calinski-Harabasz index for first and second levels of the two-level clustering used to differentiate the regions in the CO_2-injection reservoir between wells F1 and F2 (F1-F2 profile) and that between wells F2 and F3 (F2-F3 profile).

After the first level of spatial clustering based on CO_2 presence, a second level of clustering was performed. We utilized the above-mentioned methods to evaluate the model efficiency and the optimal number of clusters. With a total of five clusters, the silhouette and Davies-Bouldin scores achieved values of 0.68 and 0.46, confirming the need for a two-level clustering (table 3). Calinski-Harabasz was also estimated to show a good agreement with the other two clustering scores.

For this work, the final five clusters represent the five levels of CO_2 content ranging from negligible, low, medium-low, medium-high to high. A total of five clusters were established, providing an image of the CO_2 plume after nearly 10 months of CO_2 injection into the subsurface reservoir.

5.2. Traditional clustering vs. two-level clustering

In this section, we compare the traditional clustering and proposed multi-level clustering. The comparison highlights the impact of this novel approach. Traditional clustering consists of a one-level partitioning of the extracted data. Five clusters were predefined to analyze the clustering behavior and compare it with the proposed methodology.

Traditional/single-level k-means overestimates regions with high CO_2 content. In addition, the reliability of clusters obtained using single-level clustering are not reliable because of the low separation and low density of the certain clusters, resulting in a higher uncertainty in the CO_2 location and over estimation of regions with high CO_2 content. The proposed two-level clustering using k-means has better reliability as compared to the single-level/traditional clustering (figure 7).

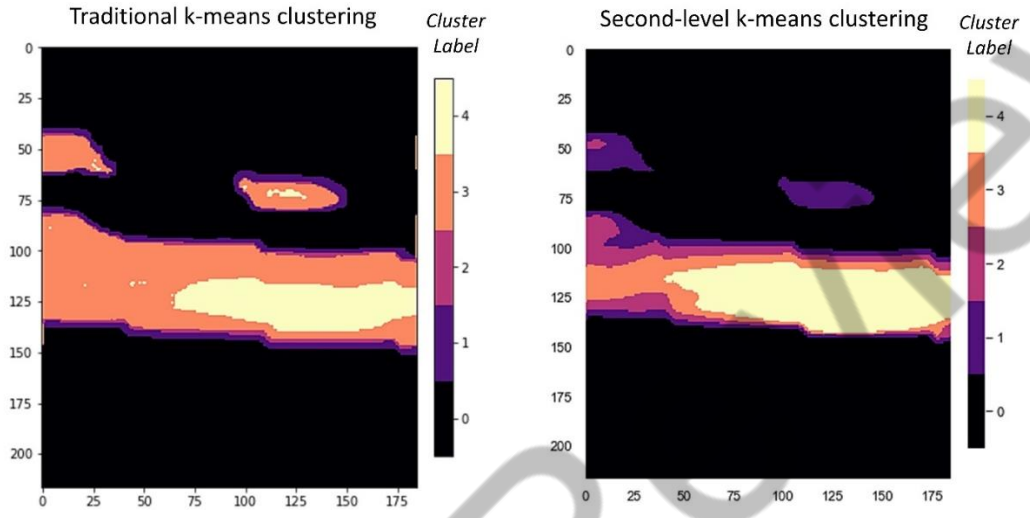


Figure 7: Comparison of the clustering results for the single-level/traditional k-means versus two-level k-means. The results are reasonable for the non-CO_2 cluster; however, the spatial clustering based on the content of CO_2 displays high discrepancies (i.e. for clusters 1, 2, 3, and 4).

5.3. Consistency of spatial clustering obtained using different clustering methods

A comparative analysis was conducted to confirm the consistency of cluster labels obtained using the two-level k-means clustering. As shown in figure 8, we compare the spatial clustering obtained using k-means against those obtained using mean-shift and agglomerative clustering methods. Mean-shift clustering aims to cluster data points based on the discovering of the modes in a data distribution, while agglomerative clustering groups the samples in hierarchical structure based on the similarity and recursive cluster merging of similar clusters to obtain larger clusters. Figure 8 qualitatively evaluates the consistency between the cluster labels assigned using k-means, agglomerative, and mean-shift clustering methods.

Quantitative evaluation of the consistency and robustness of clusters are presented in Table 4 using two pair-wise scores, namely adjusted random score and homogeneity score. Adjusted random score estimates the similarity between two clustering results while ignoring permutations. The homogeneity score evaluates the clusters labeling based on the principle of clusters containing only a single class. The quantified similarities range from 0.90 to 0.99, wherein 1 represents a perfect match between spatial clustering obtained using different methods. Each clustering technique has distinct assumptions and underlying principles to achieve the clustering. The extremely high consistency among the cluster labels confirms the robustness of the proposed unsupervised learning

workflow and the reliability of specific geophysical signatures/features used for the desired visualization of the CO_2 plume in the subsurface.

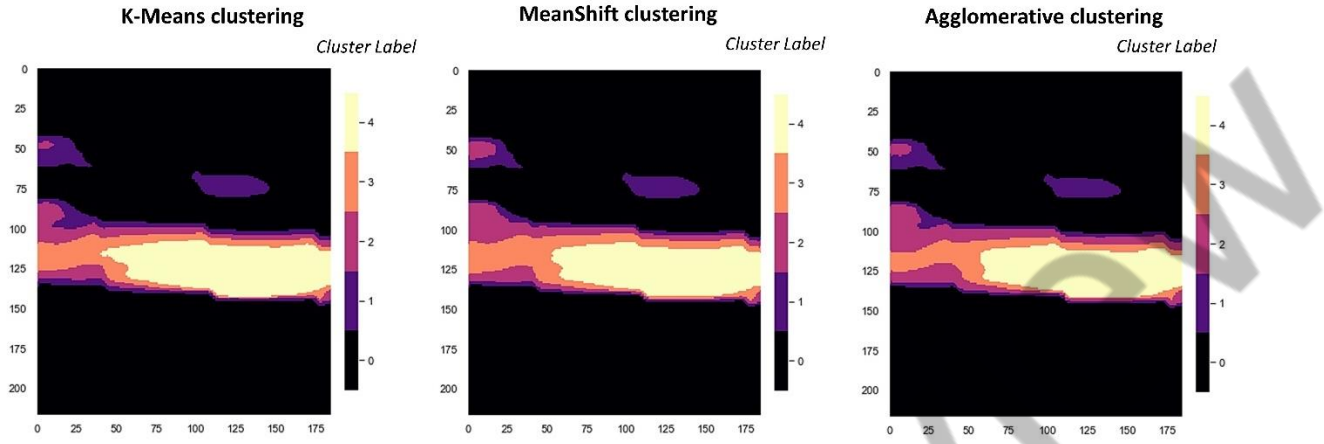


Figure 8. Spatial clustering was obtained using the two-level k-means, mean-shift, and agglomerative clustering. The similarity in the spatial clustering reinforces the consistency and robustness of the proposed workflow. The comparison is better quantified in Table 4.

Clustering comparison methods	**Adjusted random score**	**Homogeneity score**
K-means and agglomerative	0.979	0.899
K-means and mean-shift	0.989	0.924
Mean-shift and agglomerative	0.982	0.905

Table 4. Comparison of the similarities between spatial clustering obtained using the two-level k-means, mean-shift, and agglomerative clustering methods. Scores close to one indicates high similarity between the results from different clustering methods.

5.4. Analysis of the Robustness and Reliability of the Spatial Clustering

The above-mentioned workflow assigns a specific cluster label to a specific region of the injection reservoir. To investigate the robustness and reliability of the cluster labels in representing the presence and content of CO_2, we perform several statistical tests. In this section, we will assign a physically consistent CO_2-content indicator to each cluster label. As a result, statistical cluster labels are converted to physically meaningful labels. Figure 9 shows the histogram of the cluster labels. Regions without any CO_2 are represented by Cluster 0, which is the dominant cluster in the injection reservoir. Cluster 0 accounts for approximately 76.3% of the data. Clusters 1, 2, 3, and 4 represent regions containing low, medium-low, medium-high, and high CO_2 content. Regions containing low, medium-low, medium-high, and high are equivalent to 5.7%, 4.3%, 4.6%, and 9.1% of the data,

respectively. Similar behavior and trends are also present in the silhouette plots of figure 6, where the thickness of each cluster represents the number of data points belonging to a particular cluster.

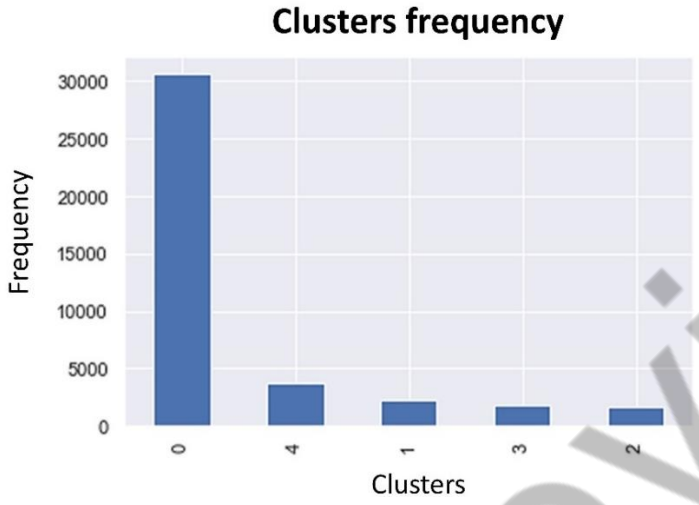


Figure 9. Clusters histograms. Cluster labels were generated by the two-level k-means clustering of the nine extracted features. Cluster 0 is associated with no-CO_2 whereas clusters 1, 2, 3, and 4 with various levels of CO_2 content.

To further investigate the uniqueness of clusters, we estimate the Euclidean distances between cluster centers. This represents the extent of dissimilarity among various clusters. The estimation is presented in table 5, where larger distances correspond to larger dissimilarities. Cluster 4 (high CO_2) is farthest from Cluster 0 (non-CO_2) and Cluster 1 (low CO_2). Cluster 1 (low CO_2) is relatively closer to Cluster 0.

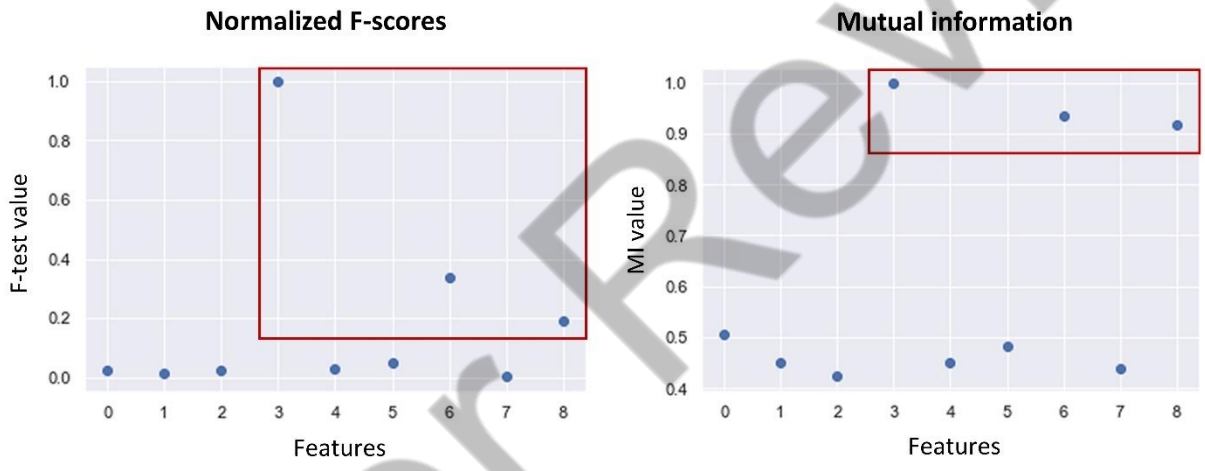
Distances between cluster centers					
Clusters	0	1	2	3	4
0	0.00				
1	186.18	0.00			
2	744.08	558.05	0.00		
3	1418.51	1232.54	674.49	0.00	
4	2034.72	1848.80	1290.77	616.30	0.00

Table 5. Euclidian distances between cluster centers, where Cluster "0" indicates regions that do not contain any CO_2, while Clusters "1", "2", "3", and "4" indicate the regions that contain low, medium-low, medium-high, and high CO_2 content, respectively.

5.5. Most informative, relevant, and discriminative features

In this section, we identify the features that are the most informative, discriminative, and relevant for achieving robust clustering-based visualization of CO_2 content. To that end,

ANOVA (analysis of variance) F-test and mutual information values were computed to determine the strength of the association between a feature and the clusters (Figure 10). Mutual information quantifies the mutual dependence between a feature and a cluster. In other words, it measures the amount of information obtained about the clusters when a specific feature is implemented in the clustering. Mutual information for a discrete target variable was used in this study, which is based on the entropy estimation of features and target. Meanwhile, ANOVA F-test compares the variances between groups and within groups. This is a specific statistical test that allows the analysis of multiple clusters to determine the features that exhibit significant variation across the clusters. Since we only have one target, a one-way ANOVA was applied. High values of ANOVA F-Test and mutual information indicate that pixel intensity, fast-Fourier transform coefficients, and wavelet transform coefficients are the most discriminative, informative, and relevant features.



Number of features	Feature
0	GLCM ASM
1	GLCM Correlation
2	GLCM Dissimilarity
3	Fast-Fourier transform
4	LBP
5	Edges
6	Wavelet transform
7	Hxx
8	Pixels

Figure 10. Values of normalized ANOVA F-test and mutual information to determine the most informative and relevant features. Fast-Fourier transform, wavelet transform, and pixel intensity are the most discriminative, informative, and relevant features for the clustering-based visualization of CO_2 content

Following that, Kendall's τ correlation is used to quantify the strength of ordinal association between a feature and the cluster. This non-parametric method was designed for a categorical target, such as clusters. As non-parametric test, this correlation method does not require assumptions of the underlying distributions in data and can be used for non-gaussian distributions. Moreover, it uses the concept of concordance/discordance of sample pairs to evaluate the pair-wise relationship. A strong association displays values close to 1 or -1 whereas values close to zero a weaker relation. As shown in table 6 the strongest correlations are linked to pixel and wavelet transform.

Feature	**Kendall's tau score**
GLCM ASM	0.64
GLCM Correlation	0.60
GLCM Dissimilarity	0.62
Fast-Fourier Transform	0.63
Linear Binary Pattern	0.68
Sobel (Edges)	0.68
Wavelet Transform	**0.96**
Hxx (Hessian matrix)	0.26
Pixels Intensity	**0.96**

Table 6. Correlation coefficients of clusters and features computed using Kendall's tau for F2-F3 profile. The correlation score displayed wavelet transform and pixels intensity as the most impactful features with values of 0.96.

All statistical analyses indicate that pixel, fast-Fourier, and wavelet transform are the features that are the most significant for differentiating the spatial regions based on CO_2 presence and distribution. Besides, a further analysis was conducted to determine the statistical significance among clusters for the three features. Post-hoc "Tukey HSD" test was implemented to identify the mean difference between clusters with respect to the most significant features. Table 7 summarizes the statistical mean difference for pixel, fast-Fourier, and wavelet transform. Fast Fourier transform is the most informative feature among the three to differentiate any two cluster. Cluster 4 (high CO_2 content) is the most distinct from both Clusters 0 and 1, and Clusters 0 and 1 are the most similar.

Feature	Clusters being compared		Mean difference
Fast-Fourier Transform	Cluster #	Cluster #	
	0	1	179.05
	0	2	716.16
	0	3	1365.86
	0	4	1957.53
	1	2	537.11
	1	3	1186.80
	1	4	1778.48
	2	3	649.69
	2	4	649.69
	3	4	591.67
Wavelet Transform	Cluster #	Cluster #	
	0	1	45.39
	0	2	181.06
	0	3	344.74
	0	4	496.99
	1	2	135.67
	1	3	299.35
	1	4	451.60
	2	3	163.68
	2	4	315.93
	3	4	152.25
Pixels	Cluster #	Cluster #	
	0	1	21.10
	0	2	91.54
	0	3	174.66
	0	4	249.91
	1	2	70.44
	1	3	153.56
	1	4	228.81
	2	3	83.12
	2	4	158.37
	3	4	75.25

Table 7. Tukey HSD for post hoc analysis of the significance of the features, namely fast-Fourier transform, wavelet transform, and pixels, for purposes of clustering the regions based on CO_2 content and presence. Mean differences (in the final column) between clusters indicate the significance of the features with respect to the distinctiveness of the clusters. Cluster “0” indicates regions that do not contain any CO2, while Clusters “1”, “2”, “3”, and “4” indicate the regions that contain low, medium-low, medium-high, and high CO_2 content, respectively. Fast Fourier transform is the most significant feature. Cluster 4 is the most distinct from both Clusters 0 and 1, and Clusters 0 and 1 are the most similar.

In addition, to determine the relationship between signatures and clusters, we generated boxplots on the most significant features (figure 11). They were established to associate the levels of CO_2 content with the feature signals. Cluster 0 corresponds to the non-CO_2 while clusters “1”, “2”, “3”, and “4” represent low, medium-low, medium-high, and high CO_2 content. The boxplots confirm that signatures of features can clearly differentiate levels of CO_2 content which are linked to a specified range. Figure illustrates that fast Fourier transform is a feature that as easily differentiable feature distribution for each cluster. Wavelet transform as a feature has a large overlap between clusters 2 and 3 and between clusters 1 and 2.

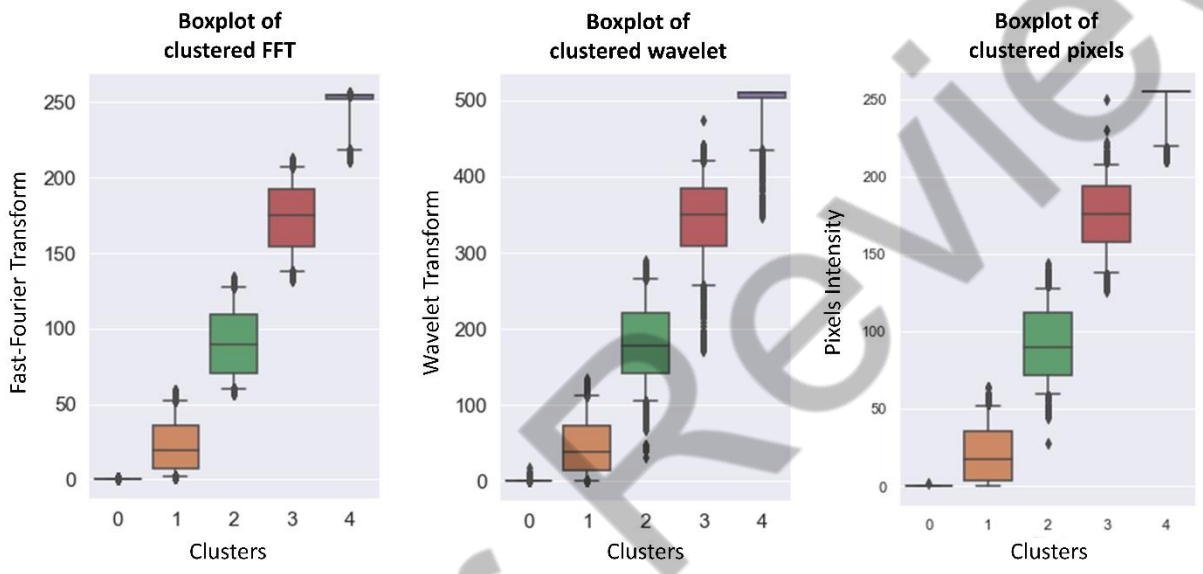


Figure 11. Boxplot of clustered fast-Fourier transform, wavelet transform, and pixels intensity. Boxplots were defined for a low 5th percentile and a high 95th percentile. For fast-Fourier transform, values of 0 were associated with non-CO_2, ~1-59 to low CO_2, ~60-138 medium-low CO_2, ~139-218 medium-high CO_2, and ~219-255 to high CO_2. For wavelet transform, values of 0 were associated with non-CO_2, ~1-109 to low CO_2, ~110-259 medium-low CO_2, ~260-435 medium-high CO_2, and ~436-510 to high CO_2. For pixels, values of 0 were associated with non-CO_2, ~1-59 to low CO_2, ~60-137 medium-low CO_2, ~138-219 medium-high CO_2, and ~220-255 to high CO_2.

6. Conclusions

Unsupervised learning enabled the visualization of carbon dioxide (CO_2) plume in a subsurface reservoir developed for carbon geo-sequestration. The workflow for CO_2 visualization incorporates feature extraction, feature selection, and two-level clustering. Four statistical tests, namely F-test, mutual information, Tukey’s HSD, and boxplot analysis, were performed to determine new geophysical signatures are discovered that are suitable for detecting CO_2 presence and content. Fourier transform and wavelet transform are the most relevant and informative features for the desired spatial clustering. The proposed two-level clustering approach is suitable for imbalanced data, which is common for a CO_2 injection reservoir, where most of the regions do not contain CO_2 and there exists variability in the spatial distributions of connected pores and CO_2. An important requirement for robust spatial clustering is to determine physically consistent

and optimal number of clusters present in the dataset. To that end, elbow plot, silhouette score, Davies-Bouldin index, and Calinski-Harabasz index indicated the existence of five robust clusters. Adjusted random score and homogeneity score confirm the robustness of the proposed unsupervised learning workflow and the reliability of specific geophysical signatures/features used for the desired visualization of the CO_2 plume in the subsurface. The use of unsupervised learning provides a fast, data-driven, qualitative approximation of CO_2 content, distribution, and presence, which serves as a substitute to rock-physics models that have inherent parametric and geophysical assumptions. Such fast real-time visualizations facilitate the assessment of safe long-term storage of carbon dioxide in the subsurface. The proposed data-driven application can be extended to many CO_2 geo-sequestration scenarios at varying conditions.

Acknowledgments

This work was supported by the Texas A&M Energy Institute funded through the Convergence Research Incubator. The data was collected from the National Energy Technology Laboratory (NETL) Energy Data eXchange.

References

Arthur, D., & Vassilvitskii, S. (2006). k-means++: The advantages of careful seeding. Stanford.

Caliński, T., and Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics, 3(1), 1–27. https://doi.org/10.1080/03610927408827101

Chakravarty, A., Misra, S., & Rai, C. S. (2021). Visualization of hydraulic fracture using physics-informed clustering to process ultrasonic shear waves. International Journal of Rock Mechanics and Mining Sciences, 137, 104568.

Davies, D., and Bouldin, D. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Davis, T., Landrø, M., and Wilson, M. (Eds.). (2019). Geophysics and Geosequestration. Cambridge: Cambridge University Press. doi:10.1017/9781316480724

Ganguly, E., Misra, S., & Wu, Y. (2020, October). Generalizable Data-Driven Techniques for Microstructural Analysis of Shales. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Haghighat, S., Mohaghegh, S., Gholami, V., Shahkarami, A., and Moreno, D. (2013). Using Big Data and Smart Field Technology for Detecting Leakage in a CO2 Storage Project. Paper presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, USA. doi: https://doi.org/10.2118/166137-MS.

He M, Li Q and Li X (2020). Injection-Induced Seismic Risk Management Using Machine Learning Methodology – A Perspective Study. Front. Earth Sci. 8:227. doi: 10.3389/feart.2020.00227

Li, H., & Misra, S. (2021). Robust machine-learning workflow for subsurface geomechanical characterization and comparison against popular empirical correlations. Expert Systems with Applications, 177, 114942.

MacLennan, K., Ganssle, G., Chen, J., Stone, K., and Yua, H. (2020). Rapid imaging of CO2 storage using deep learning with 4D electromagnetic data. SEG Technical Program Expanded Abstracts 2020. doi: 10.1190/segam2020-3418129.1

Misra, S., Chakravarty, A., Bhoumick, P., & Rai, C. S. (2019). Unsupervised clustering methods for noninvasive characterization of fracture-induced geomechanical alterations. Machine Learning for Subsurface Characterization, 39.

Misra, S., & Wu, Y. (2020). Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking (pp. 289-314). Gulf Professional Publishing.

National Energy Technology Laboratory. (n.d.). About SMART Initiative. Retrieved from https://edx.netl.doe.gov/smart/about-smart/

Osogba, O., Misra, S., & Xu, C. (2020). Machine learning workflow to predict multi-target subsurface signals for the exploration of hydrocarbon and water. Fuel, 278, 118357.

Hovorka, S.D. (2013). Three-Million-Metric-Ton-Monitored Injection at the SECARB Cranfield Project—Project Update. Energy Procedia, 37, 6412-6423.

Hovorka, S.D., Meckel, T.A., Trevino, R.H., Lu, J., Nicot, J., Choi, J., Freeman, D., Cook, P., Daley, T.M., Ajo-Franklin, J., Freifeild, B.M., Doughty, C., Carrigan, C.R., La Brecque, D., Kharaka Y.K., Thordsen, J.J., Phelps, T.J., Yang, C., Romanak, K.D., Zhang, T., Holt, R.M., Lindler, J.S., Butsch, R.J. (2011). Monitoring a Large Volume CO 2 Injection: Year Two Results from SECARB Project at Denbury's Cranfield, Mississippi, USA. Energy Procedia 4: 3478 – 3485.

Markets and Markets. (2020). Carbon Capture, Utilization, and Storage Market by Service (Capture, Transportation, Utilization, Storage), End-Use Industry (Oil & Gas, Iron & Steel, Cement, Chemical & Petrochemical, Power Generation), and Region - Global Forecast to 2025. Retrieved from https://www.marketsandmarkets.com/Market-Reports/carbon-capture-utilization-storage-market-151234843.html

National Energy Technology Laboratory. (2020). About SMART Initiative. Retrieved from https://edx.netl.doe.gov/smart/about-smart/

NOAA Earth System Research Laboratories. (January 6, 2021). Trends in Atmospheric Carbon Dioxide. Retrieved from https://www.esrl.noaa.gov/gmd/ccgg/trends/gl_trend

Rackley, S. A. (2010). Carbon capture and storage. Butterworth-Heinemann/Elsevier. First Edition. ISBN: 9781856176361

Rackley, S. A. (2017). Carbon capture and storage. Butterworth-Heinemann/Elsevier. Second Edition. ISBN: 9780128120415

Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Smit, B., Reimer, J., Oldenburg, C., and Bourg, I. (2014). Introduction to carbon capture and sequestration (Vol. 1, The Berkeley lectures on energy). Singapore: World Scientific Publishing Pte.

Wu, Y., & Misra, S. (2019). Intelligent image segmentation for organic-rich shales using random forest, wavelet transform, and hessian matrix. IEEE Geoscience and Remote Sensing Letters, 17(7), 1144-1147.

Appendix A. Clustering evaluation performance

The mathematical formulations of the clustering performance methods are presented in this section. Three evaluation approaches were used to evaluate the optimal clustering number. In this work, the silhouette score, Davies-Bouldin, and Calinski-Harabasz indexes were applied.

The silhouette score (Rousseeuw, 1987) is given as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad -1 \leq s(i) \leq 1 \tag{1}$$

Where $a(i)$ represents the average distance of each point on the same cluster and $b(i)$ the average distance to the nearest other cluster. The best performance, a score equals to 1, is achieved under lower distances of $a(i)$ and higher distances of $b(i)$. This implies a lower dissimilarity within clusters and a higher dissimilarity between them. An intermediate case can occur when a value close to zero is obtained. In this case, each point has an equal distance from both clusters; therefore, it can be assigned to either one of them. The worst performance would be with an $s(i)$ close to -1 where clustered data points are clearly misclassified.

The Davies-Bouldin index (Davies and Bouldin, 1979) computes the average similarity within clusters and between. The mathematical formulation is defined as:

$$R_{ij} = \frac{S_i + S_j}{M_{ij}}, \qquad \bar{R} = \frac{1}{N} * \sum_{i=1}^{N} \max\left(R_{ij}\right) \tag{2}$$

S_i and S_j corresponds to the average distance between each point and their respective centroid cluster, and M_{ij} the distance between the cluster's centroids. The optimal number will be the one that minimizes the similarity between clusters ($\bar{R}$). It is a non-negative index with the lowest possible score equals to zero.

The Calinski-Harabasz index (Calinski and Harabasz, 1974) is characterized as the following equation:

$$s = \frac{\left[\frac{\sum_{k=1}^{K} n_k \left||c_k - c\right||^2}{K-1}\right]}{\left[\frac{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left||d_i - c_k\right||^2}{N-K}\right]} = \frac{\left[\frac{BGSS}{K-1}\right]}{\left[\frac{WGSS}{N-K}\right]} \tag{3}$$

It represents the ratio of within-cluster (WGSS) and between-cluster sum of square (BGSS). n_k is the number of points per cluster, and c_k and c the cluster and global centroids. A higher value indicates a denser and well-separated cluster, being this the most optimal one.

Appendix B. Thresholding methods vs k-means clustering

The application of both algorithms is designed to retrieve regions or groups of similar characteristics and properties. Nevertheless, key differences are displayed in the proposed research for CO_2 plume characterization. To validate the use of unsupervised clustering, a thresholding model was established called multi-Otsu. The multi-Otsu thresholding is a multilevel method used to separate pixels based on their intensity level. The selection of threshold regions was developed by Liao et al. (2001) at the maximum between-class variances.

In this study, the multi-Otsu thresholding was implemented to create five unique regions from the CO_2 image. As shown in figure 12, four threshold values were determined at a pixel intensity of 30, 90, 154, and 218. The results are observed on the two-dimensional image with the proposed regions, observing the CO_2 plume shape and their respective CO_2 saturation. Taking these clusters, a comparison between two-level k-means, meanshift, and agglomerative clustering was performed using the adjusted random and homogeneity scores. Table 8 presents the estimated values between k-means, agglomerative, meanshift, and multi-Otsu thresholding where a value equal to one represents a perfect match. From the pair-wise comparison, we can acknowledge the lowest score for the multi-Otsu comparison on each of the two-level clustering approach, achieving average adjusted random of 0.867 and a homogeneity score of 0.834.

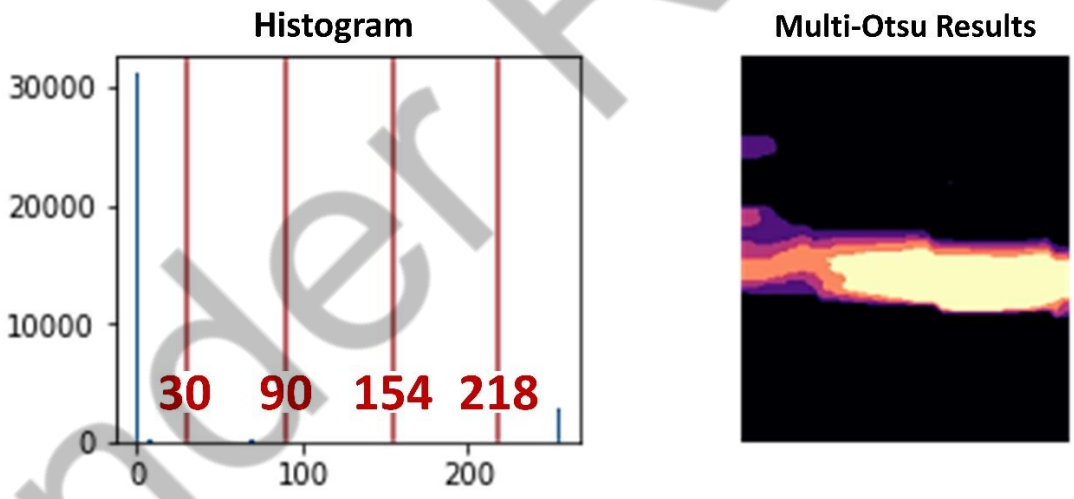


Figure 12. Thresholds of pixel intensity using multi-Otsu algorithm. On the left, a histogram was displayed to observe the defined regions of thresholding method. On the right, five regions were set based on the thresholds of 30, 90, 154, and 218 intensities.

Clustering comparison methods	Adjusted random score	Homogeneity score
K-means and Agglomerative	*0.979*	*0.899*
K-means and Meanshift	*0.989*	*0.924*
K-means and ***Multi-Otsu***	***0.865***	***0.754***
Multi-Otsu *and Meanshift*	***0.866***	***0.877***
Multi-Otsu *and Agglomerative*	***0.871***	***0.870***

Table 8. Comparison of two-level clustering using K-means, meanshift and agglomerative clustering, and multi-Otsu thresholding. A pair-wise score close to one indicates a high similitude.

Discrepancies between multi-Otsu are further analyzed in figure 13 with their respective pair-wise scores. From the multi-Otsu results, a higher non-CO_2 content is observed which reduces the levels of CO_2 at multiple regions. The decrease will incorporate a lower area of CO_2, generating potential false predictions for non-CO_2. The unbalanced nature of the dataset has not been taken into account which can generate misleading regions due to the different distribution of classes.

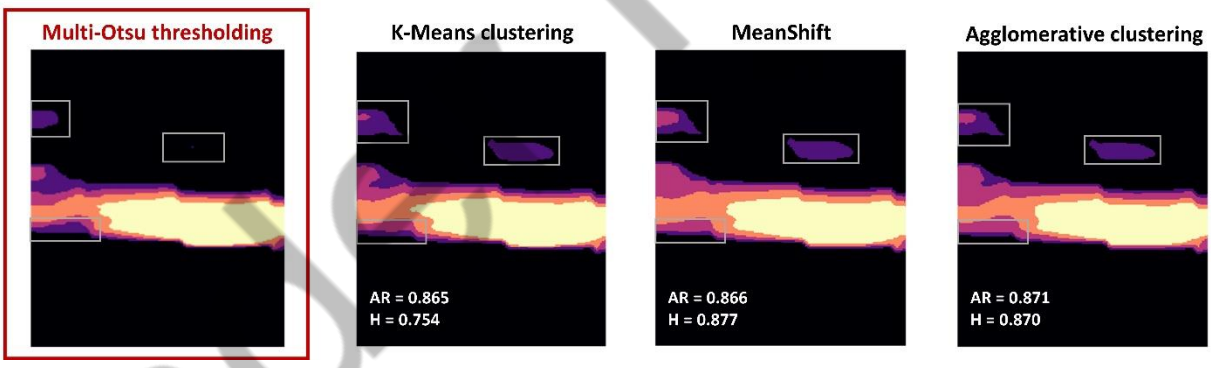


Figure 13. Pair-wise comparison of multi-Otsu regions with two-level clustering k-means, meanshift, and agglomerative clustering. “AR” represents adjusted random and “H” homogeneity scores. Visual differences are distinguished at grey rectangles to observe the decrease of CO_2 content at various regions.

Problems with thresholding methods are linked to the over-strictly rigid thresholds. The selection of their optimum value should be set based on prior knowledge to be the most informative, being this a biased consideration and time-consuming process. Furthermore, thresholding is based on a unique feature which will describe an entire phenomenon based on only one characteristic at a particular spatial time. On the other hand, unsupervised clustering considers a set of local information of distinctive characteristics such as edges, shape, and texture which extracts the target at a more efficient and deeper level. Another key drawback of threshold methods is the static component of the threshold

regions. Hence, new thresholds are needed to be defined for each new image. Errors commonly appeared under changing conditions due to factors of variable noise levels or different statistical distributions. Unsupervised clustering does not require performing a prior statistical investigation for each new data. It is flexible to dynamic objects and therefore does not need to set an arbitrary threshold, considering only the distances between observations.

Appendix C. Two-level k-means generalization

A good data clustering is achieved under different data conditions while preserving their clustering performance. This will indicate well-defined and separated clusters of simple clustering boundaries. The clustering generalization can be investigated through the changes in data shape, density, and quality. Under this scenario, generalization two tests were designed based on their change of size and noise level. The first case incorporates a Gaussian noise to the original dataset where a random set of disturbances were integrated. On the other hand, the second test is applied to a small portion of the original image to examine the influence of the change of data size. As illustrated in figure 14, the Gaussian noise does not impact the clustering results at either first or second-level, retaining the same clustering results at lower data quality. Furthermore, for the second test (figure 15) the portion of data is clustered at ~1/2 of the original image where clusters did not get affected at both levels. To evaluate the k-means generalization, a pair-wise similarity measure between second-level clustering and these two tests was estimated. Using the adjusted random metric, a value of 0.951 for the noisy case while for the portion of data a value of 0.961 where scores close to one indicate a high pair-wise similarity. Hence, our proposed workflow is generally applicable to noisy and changing size datasets without altering the goodness of clustering.

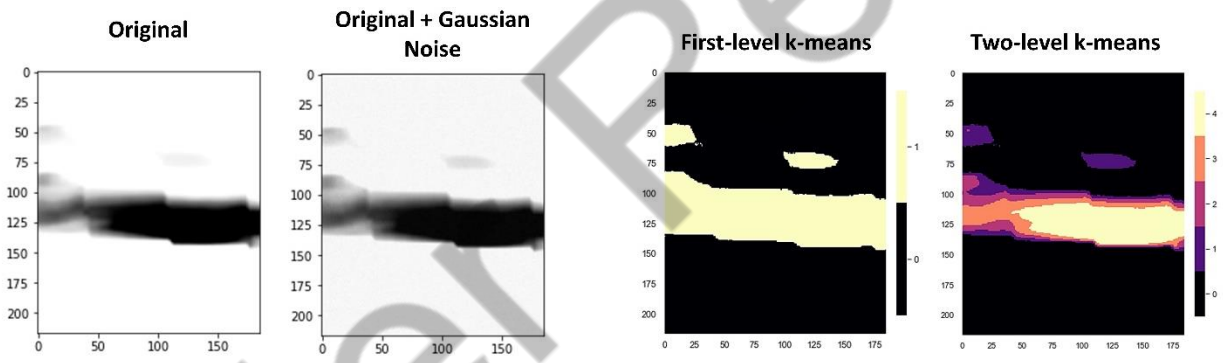


Figure 14. Spatial clustering of the noisy image for the first and second-level clustering. The Gaussian noise was designed with a standard deviation of 2.5, adding random samples to the original distribution.

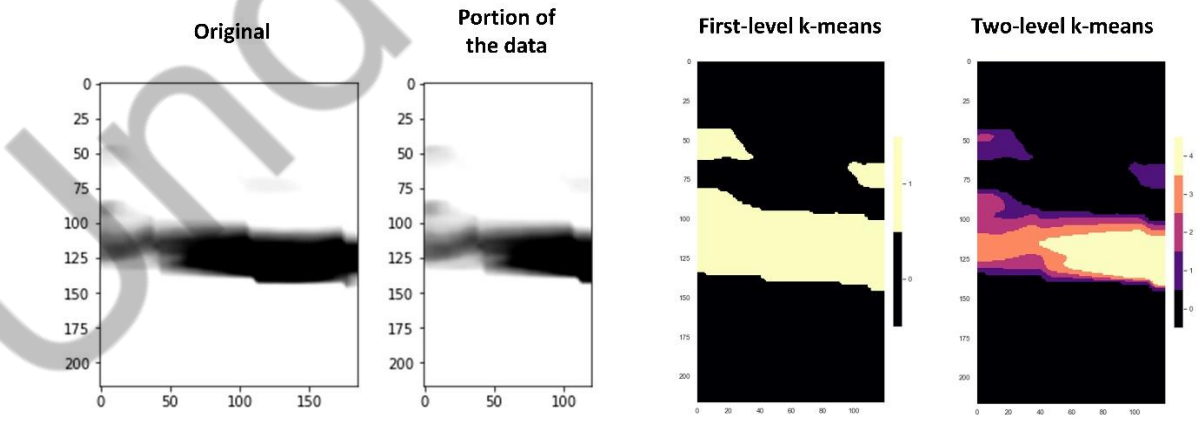


Figure 15. Two-level k-means clustering using a portion of the original dataset. This new image corresponds to ~half of the original one for the multi-level clustering. The stability of the clustering results was examined to validate the generalization of the proposed workflow.

In addition, to ensure consistent clustering results, the k-means++ convergence method was used. This initialization technique first selects the cluster's center at random and later weights their data on their squared distance from the closest center (Arthur and Vassilvitskii, 2006). The algorithm ran 10 times with different centroid seeds and a convergence tolerance level of 1e-4. Furthermore, with this set of hyperparameters, we iterated the proposed two-level clustering for 20 and 50 runs to evaluate the variance of the clustering results. Figures 16 and 17 display the Davies-Bouldin, Calinski-Harabasz, and silhouette scores for the first and second-level clustering. Based on the steady internal measures, stability for both clustering levels under the 20 and 50 runs was achieved. Thus, the initialization will not impact the performance of multi-level clustering for this dataset.

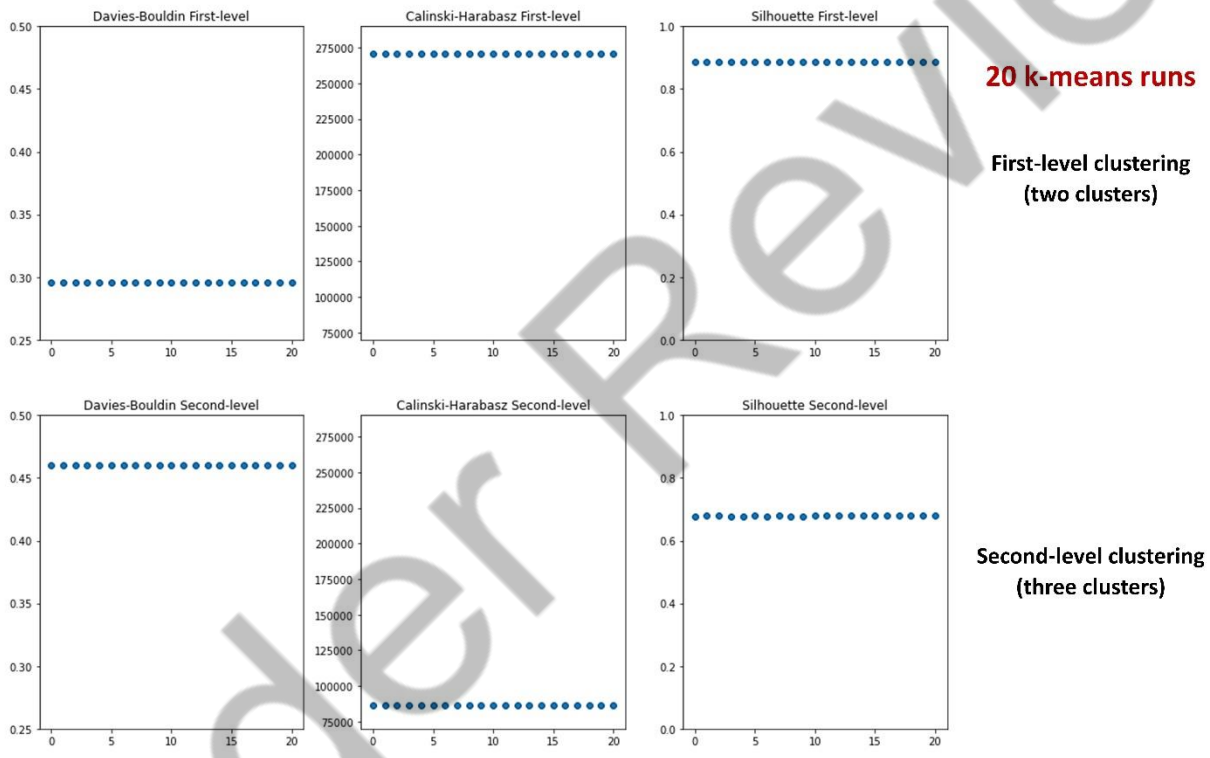


Figure 16. Internal clustering measures of Davies-Bouldin, Calinski-Harabasz, and silhouette at 20 k-means runs. The first and second-level k-means clustering were implemented using k-means++ initialization, the convergence tolerance level of 1e-4, and 300 maximum iterations for a single run.

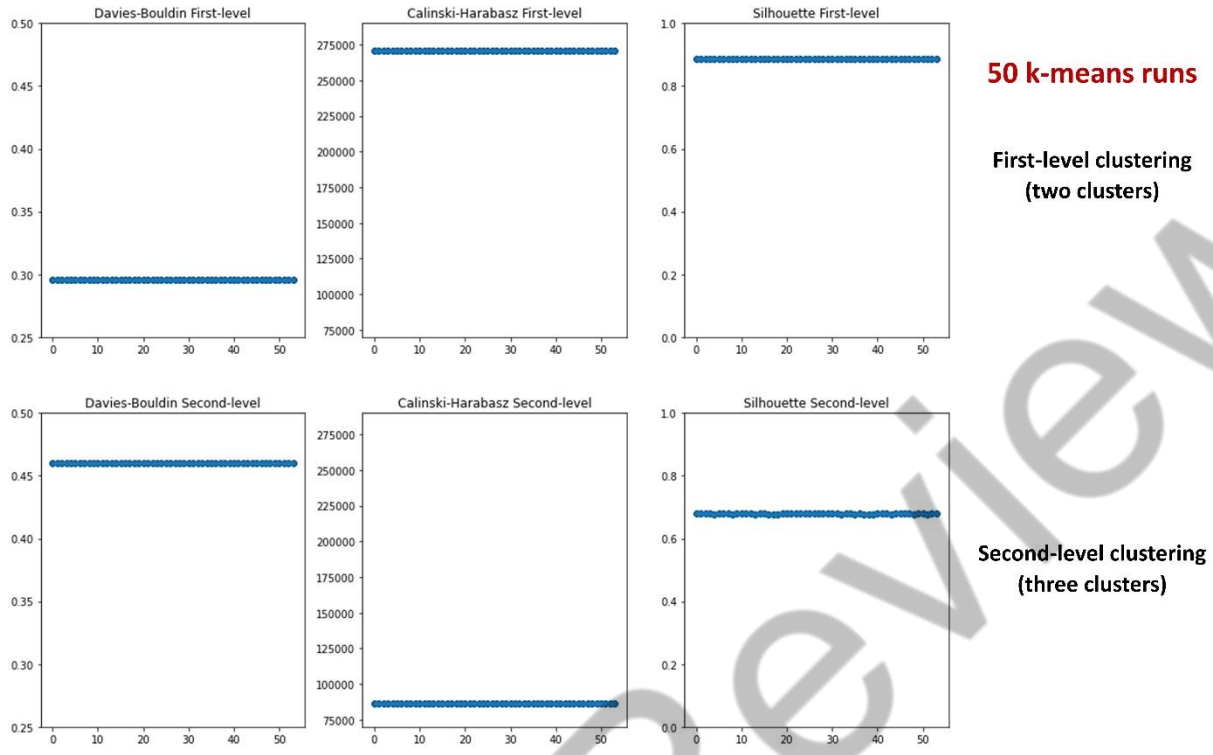


Figure 17. Davies-Bouldin, Calinski-Harabasz, and silhouette scores at 50 k-means runs. Consistent and reliable measures were obtained for each clustering level at the pre-defined hyperparameters.