Monitoring the CO_2 Plume Migration during Geological Carbon Storage using Spatiotemporal Clustering

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

Precision monitoring of the subsurface carbon-dioxide plume ensures long-term, sustainable geological carbon storage. Carrigan et al. (2013) and Yang et al. (2014) showed that electrical resistivity tomography (ERT) can accurately map the evolution of the CO_2 saturation during geological carbon storage. To better monitor the CO_2 plume migration in a storage reservoir, we develop an unsupervised spatiotemporal clustering to process the CO_2 saturation maps derived from the ERT measurements acquired over 80 days by Carrigan et al. (2013). Using dynamic time wrapping (DTW) K-means clustering, four distinct clusters were identified in the CO_2-storage reservoir. The four clusters exhibit a Davies-Bouldin (DB) index of 0.71, a Calinski-Harabasz (CH) index of 262791, and a DTW-silhouette score of 0.58. Unlike traditional clustering methods, the DTW K-means incorporates a temporal distance metric. Traditional clustering methods, such as Euclidean K-means, agglomerative and meanshift clustering, exhibit a lower performance with DB index of 0.83, 0.95, and 1.01, respectively, and CH index of 157866, 131593, and 69438, respectively. Subsequent statistical analysis indicates that contrast stretching and fast-Fourier transform are strong geophysical signatures of the spatiotemporal evolution of CO_2 plume. We also identified a strong correlation between injection flow rate and the spatial evolution of regions with high CO_2 content. Finally, the previously computed spatiotemporal clusters were further clustered to discover distinct temporal sequences emerging with respect to the overall CO_2 plume distribution in the subsurface. Six distinct temporal clusters of CO_2 plume evolution were detected over a period of 2 months. A tensor-based feature extraction was critical for processing the ERT data acquired over 80 days to capture both the temporal and spatial components relevant to the evolution of CO_2 plume in the storage reservoir.

Keywords: *Carbon Storage; Unsupervised Learning; Spatiotemporal; Clustering; Subsurface Monitoring*

1. Introduction

Subsurface monitoring of geological CO_2 storage is critical to ensure storage integrity and efficiency. To identify and map potential pathways of CO_2, appropriate imagining technologies are required. According to Davis et al. (2019), these can be divided into three main categories: surface, near-surface, and subsurface techniques. For subsurface mapping, a set of monitoring tools can be used to establish the CO_2 location by providing a spatial image of the migrated CO_2. A critical aspect of any CCS project is the real-time location of the injected CO_2 and its migration. Numerous processes can affect the CO_2 plume evolution such as geological heterogeneity, interfacial tension, geological structures, leaky pathways, and gravity forces. For instance, geological heterogeneity can significantly reduce the CO_2 injection capacity due to regions of low permeability. Geological structures such as dips can also affect the migration of CO_2 by obtaining a condensed supercritical fluid at the bottom of the seal. Hence, the understanding of these dynamic systems is vital to establish a safe long-term carbon storage and reduce potential CO_2 leakage.

Machine learning (ML) algorithms are ideal for uncovering hard-to-detect relationships between the fluid flow processes and the monitoring signals in complex and dynamic physical systems. Haghighat et al. (2013) demonstrated the use of ML using modeled CO_2 leakage and real-time pressure data. Ni and Benson (2020) developed an unsupervised clustering model to identify capillary flow regimes on five CO_2 coreflooding datasets. Pires de Lima et al. (2019) used convolutional neural networks to model the amount of leaked CO_2 using synthetic pressure and seismic data. Traditional studies of CO_2 systems while using ML are often comprised of synthetic datasets and limited subsurface measurements, being key challenges to characterize CO_2 movement. Moreover, subsurface imaging must be acquired for an extensive period of time in order to incorporate the respective spatial plume change. Therefore, spatiotemporal approaches are needed to be investigated to examine the injected CO_2 growth.

The main purpose of this work is to investigate the evolution of CO_2 plume and associated fluid-flow mechanisms by processing the electrical resistivity tomography (ERT) of a geological carbon-storage reservoir using unsupervised learning methods. Dynamic time wrapping and k-means clustering were implemented to identify spatiotemporal patterns of CO_2 plume migration. A novel tensor-based feature extraction processes the ERT data acquired over 80 days to capture both the temporal and spatial components relevant to the evolution of CO_2 plume. To validate the quality of obtained clusters, Calinski-Harabasz, Davies-Bouldin, and silhouette scores were calculated. In addition, a comparison between traditional and tensor-based feature extraction was performed to corroborate the improvement of clustering performance when using tensor-based feature extraction. A key aspect in this work is the uncovering of the most informative features associated with the spatiotemporal evolution of the CO_2 plume. Hence, statistical tests such as ANOVA f-test, interquartile range, and post-hoc Tukey's were applied. Lastly, we develop a second "temporal" clustering of the daily clustered results to gain insights into the stages of CO_2 plume development where distinctive fluid mechanisms can be recovered for further analysis.

1.1. Use of unsupervised methods for geological carbon storage

Use of supervised learning on geophysical field data is constrained by the availability of targets/labels associated with measured signals. Consequently, unsupervised learning methods are valuable for analyzing geophysical data. Unsupervised learning help discover hidden relationships, generate new insights, and identify new patterns when working with datasets acquired in complex systems. Clustering partitions and categorizes the data in terms of similarity, dissimilarity, distance and density of samples and corresponding clusters.

Spatiotemporal clustering is an extension of the spatial one where the time dimension is introduced. The spatiotemporal clustering will categorize the objects of similar spatial and temporal behaviors into well-defined groups. Spatiotemporal clustering has been implemented in traffic management, image processing, molecular biology, satellite measurements, seismology, medical imaging, and real-time monitoring. For instance, Liu et al. (2018) developed a spatiotemporal model to extract patterns of land

surface temperatures using satellite images. Chen et al. (2015) examined the dynamic behavior of water bodies based on their change in size, shape, and signal. Anomaly detection has also shown the use of spatiotemporal clustering for marine remote sensing measurements (Liu et al., 2018) and traffic sensors (Cao et al., 2018).

Spatiotemporal data contain signatures that evolve over time on a spatial domain. The spatial component relates to the location of a specific property or process, while the temporal component relates to the time dependent variation in the property or process. Challenges with unsupervised clustering of spatiotemporal datasets include: 1) lack of a clear separation/boundaries between clusters because the samples continuously vary in space and time, 2) data could contain redundant information within each temporal image, and 3) spatiotemporal changes can lead to computational complexity due to the dynamic states.

Carbon sequestration requires an on-going surveillance to ensure a safe long-term storage. In particular, subsurface monitoring can be used to establish the location and evolution of the injected CO_2. Time-lapse mapping of the CO_2 signals from the subsurface is commonly acquired to investigate the CO_2 migration over an extensive period of time. Spatiotemporal clustering has the capability of grouping large amounts of geophysical signals for the real-time detection of moving CO_2 plume.

2. Methodology

2.1. Use of electrical resistivity tomography (ERT) for geological carbon storage sites

ERT is a geophysical method that involves the application of a direct electric current at a certain point in the subsurface to measure the electric potential at various other locations in the subsurface. In our study, ERT measurements were performed daily for several months across two wells drilled into a geological CO_2 storage site. The crosswell ERT source/sensor or transmitter/receiver configuration of electrodes is displayed in Figure 1. The electrodes are placed along the two monitoring wells at a specific electrode spacing. The ERT measurement involves the injection of electrical current from one of the electrodes in the first well (well F2). The current travels through the carbon-storage reservoir and finally reaches to the electrodes in the second well (F3),

resulting in a voltage gradient across the electrodes in the second well. These potential/voltage differences constitute the raw ERT dataset. The two monitoring wells F2 and F3 are 98 feet away from each other. The current injection in well F2 and the voltage measurements in well F3 are repeated across multiple electrodes over multiple days resulting in the time-lapse measurements of the changes in the electric-current pathways as the CO_2 plume moves in the storage site. All this generates the time-lapse ERT data that is subsequently processed to obtain a spatiotemporal resistivity map of the geological storage site. Next, the resistivity maps are inverted to obtain the CO_2 saturation map between the wells.

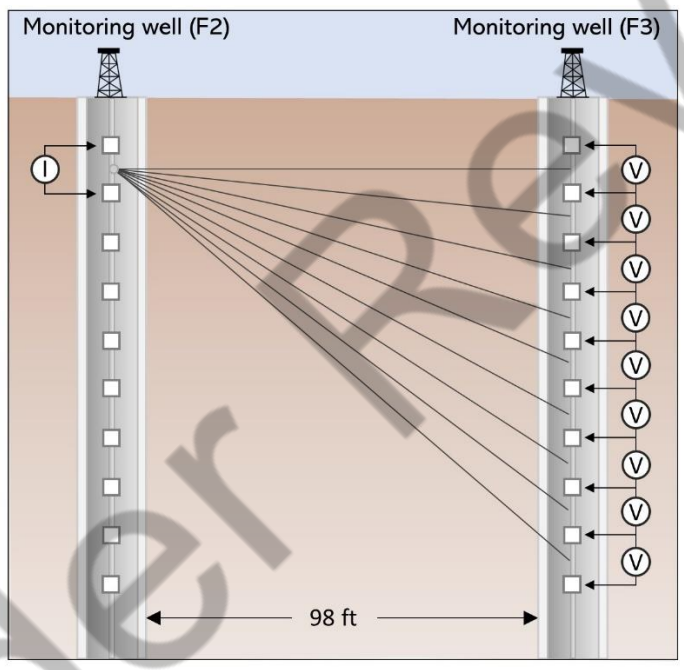


Figure 1. Schematic representation of the crosswell ERT acquisition, where the electric potential measured across multiple electrodes in the well F3 as a result of the current injection in one of the electrodes in the well F2. Wells F2 and F3 are 98 feet in distance. I = electric current injected in well F2, V= measured voltage gradient in well F3.

Electrical techniques are particularly suitable for monitoring the displacement of fluids and changes in fluid properties in the subsurface. ERT is suitable for monitoring changes in chemical and physical properties of the subsurface due to the interaction of the injected CO_2 with the connate fluids, such as brine and oil, especially the effects of fluid displacement. To investigate the evolution and movement of CO_2, crosswell ERT is ideal. The supercritical CO_2 can be easily visualized due to the high resistivity contrast

with respect to the brine reservoirs, which is approximately five times more resistive than their surroundings. The resulting ERT provides an image of the location of CO_2 plume content at a particular time and at a certain distance. The use of crosswell designs improves the estimation of CO_2 saturations and extends the range to a higher imaging resolution (Yang et al., 2014).

2.2. Dataset description

In this work, the dataset was retrieved from the Southeast Regional Carbon Sequestration Partnership (SECARB), which is considered one of the main geo-sequestration programs across the United States (SECARB, 2017). In particular, the Cranfield project from SECARB was created to investigate carbon sequestration at a CO_2-EOR field. The overall goal is to demonstrate the feasibility of CO_2 injection and safe long-term storage. An extensive monitoring program was conducted to evaluate the movement and location of supercritical CO_2 at deep and near-surface scales (Carrigan et al., 2013; Yang et al., 2014). Measurement techniques such as time-lapse seismic, crosswell seismic and ERT, pressure, groundwater, and soil-gas monitoring were acquired to assess CO_2 distribution.

The dataset utilized in this study contains 80 daily ERT measurements collected from 21 December of 2009 to 12 March of 2010. The ERT acquisition was configured in two monitoring wells, wells F2 and F3, with the placement of 14 electrodes in well F2 and 7 electrodes in well F3 (figure 2). The electrodes in well F2 inject current. The electrodes in well F3 record the voltage differences. The injection well is set to a distance of approximately 229 feet from well F2 and 327 feet from well F3. On the other hand, the reservoir depth is set at a range of ~10449.5 to ~10521.5 feet at a thickness of 72 feet (figure 3). Each ERT map in this study represent an interwell distance of 32 meters between the monitoring wells F2 and F3 along the x-axis and a reservoir thickness of 57 meters along the y-axis. In Figure 2, the *x*-coordinates represent the distance between well F-2 and F-3, and the top and bottom *y*-coordinates represent the first and last electrodes. For purposes of clustering, we focus on a subset of the original ERT map comprising 25 meter along *y* axis (i.e. depth) and 30 meters along *x* axis (interwell distance).

Processing of the data was previously carried out by Carrigan et al. (2013) where three major steps were performed:

1) Removal of noisy data points using multiple thresholds.
2) Modeling baseline measurements to construct a reference dataset.
3) Time-lapse inversion to obtain CO_2 saturations from the daily resistivity changes. The processing approach was based on the ratio inversion scheme and Archie's law.

The raw resistivity is converted to carbon saturation by applying the following equation:

$$S_{CO_2} = 1 - \left(\frac{\rho_0}{\rho}\right)^{\frac{1}{n}} \qquad (1)$$

where ρ_0 the baseline resistivity, ρ the measured inverted resistivity and n the saturation exponent. The baseline dataset was acquired before the CO_2 injection. The CO_2 data are presented as percentages in Figure 3, where a value of zero is the lowest response and 25 is the highest saturation observed in the 80-day window.

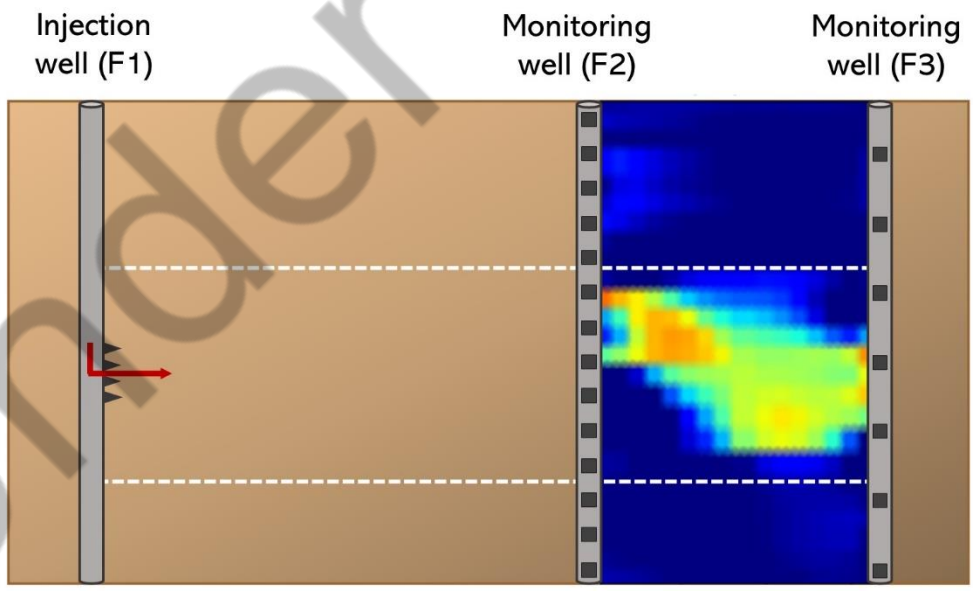


Figure 2. Diagram of the injection well (F1) and two monitoring wells (F2 and F3). Well "F2" has 14 electrodes and well "F3" has 7 electrodes.

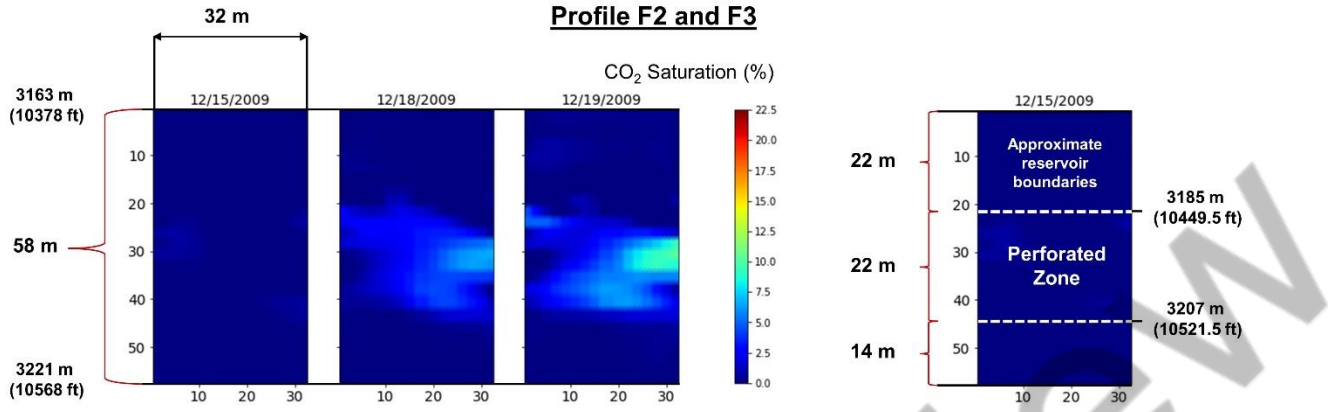


Figure 3. Three days of CO_2 saturation maps between wells F2 and F3. The CO_2 saturation map is derived from ERT measurements. The illuminated reservoir volume of interest for purposes of CO_2 monitoring has a thickness of ~22 m and a length of ~32 m. The CO_2 saturation ranges from 0 to 22.5 %.

2.3. Spatiotemporal clustering of the electrical resistivity tomography of a geological CO_2 storage site

Conventional method for locating/assessing the injected CO2 plume in the subsurface assumes a geophysical model to assess the spatial distribution of CO2 content. The geophysical model is specific to sensor configuration, sensor type, engineering design, and reservoir properties. The assumed geophysical model may not be applicable to all types of CO2-injection reservoirs and scenarios. We developed a novel and reliable unsupervised learning methodology, based on spatiotemporal clustering, for the visualization of the CO2 plume in the subsurface. This approach is adaptive and scalable without incorporating a pre-defined geophysical model. The unsupervised learning facilitates sensor-agnostic, geophysical-model-free, rapid monitoring of the CO2 content and distribution in subsurface. Appendix A provides more information on the efficacy of the proposed unsupervised learning as compared to that of traditional thresholding based data partitioning method.

Few benefits of using unsupervised methods as compared to traditional thresholding methods or geophysical models include:

1. Unsupervised learning workflow can process sensor data irrespective of the sensor types, transmitter/receiver configurations, sensor-data processing, engineering

designs, CO2 injection schedules, and geological properties of the CO2 injection reservoir.

2. Rapid spatiotemporal monitoring of CO2 plume movement can be achieved for any type of geophysical data acquired from any type of geological carbon storage site without requiring a specific assumption of the geophysical model or specific source-sensor configuration.

3. The unsupervised learning workflow will allow pathways for improved assimilation of expert domain knowledge in form physical interpretations and the infusion physical principles.

The proposed workflow is designed to monitor the spatiotemporal alterations in a geological carbon-storage reservoir due to the CO_2 plume migration. The workflow also identifies few geophysical signatures that are most relevant to the spatiotemporal evolution of the CO_2 plume. The workflow also uncovers specific fluid-transport processes corresponding to the CO_2 plume migration. The proposed workflow is presented in figure 4. 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). The key steps implemented in the workflow are summarized as follows:

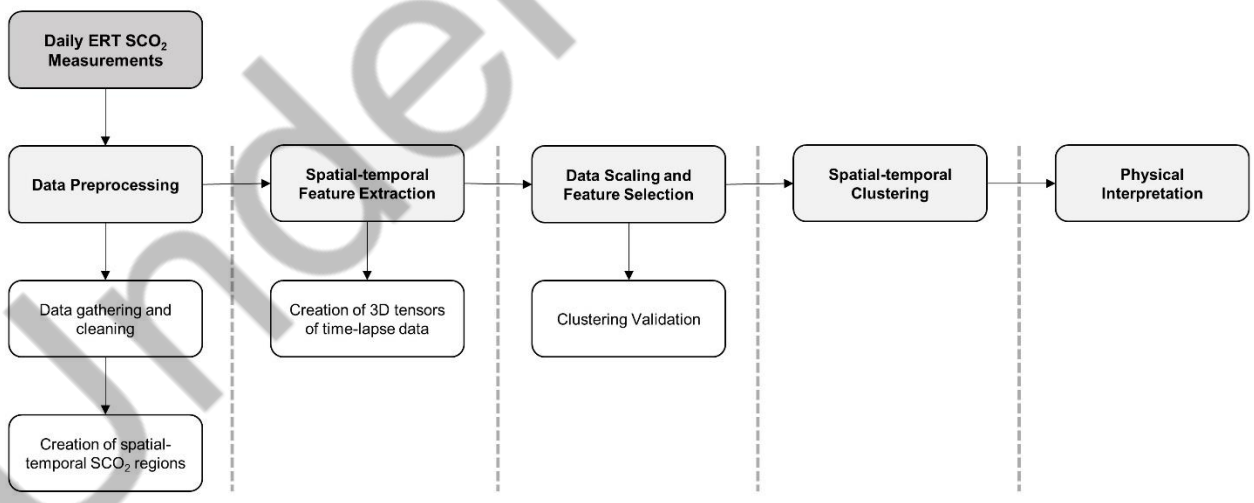


Figure 4. Workflow used for the spatiotemporal clustering of ERT for monitoring the CO_2 plume migration. The workflow involves six important steps to process, implement, evaluate, and validate the CO_2 plume migration between the two monitoring wells. The workflow is explained in section 2.3.

Step 1: Data preprocessing: The dataset was gathered and cleaned by identifying the missing and irrelevant data points to increase the robustness and generalization of the data-driven model.

Step 2: Spatiotemporal feature extraction: For each region in the storage reservoir, the features are developed by considering the previous, current, and subsequent states of the region of interest and its neighboring regions. Statistical parameterization compresses temporal-information of the features to single value. In total, for each region at a specific time, 12 features were extracted from a 3D feature tensor centered at the time and region of interest.

Step 3: Data scaling and feature selection: MinMax scaling was applied to transform the features to a comparable scale for the unsupervised clustering algorithm. MinMax scaler estimation is presented as:

$$X_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

where "min" and "max" represent the feature range and "*x*" the feature vector. This standardization will transform the samples to range from 0 and 1. From the extracted features, we selected the most impactful ones based on their pairwise correlations and the association with the moving spatial clusters.

Step 4: Spatiotemporal Clustering. K-means clustering was implemented using the time series distance metric: dynamic time warping (DTW). The optimal number of clusters was determined using the internal metric scores of Davies-Bouldin, Calinski-Harabasz, and DTW silhouette. For evaluating the DTW k-means approach, three "traditional" clustering algorithms were employed: agglomerative, mean-shift, and Euclidean k-means. We established their performance and compared them to each other with the internal measures of Davies-Bouldin and Calinski-Harabasz.

Step 5: Geophysical Signatures. Statistical analyses were developed to validate the clustering results and retrieve the features that best describe the CO_2 content and its distribution. For this, ANOVA (analysis of variance) and post-hoc Tuckey's test were applied to capture the statistical differences of selected features for various clusters. All

this led to the identification of relevant and informative geophysical signatures strongly associated with the CO_2 plume migration.

Step 6: Physical interpretation. Using the wellbore measurements (temperature, pressure, and flow rate) of the injection well, we compared the daily clusters count with the injection phases. The goal is to correlate the behavior of each cluster to the wellbore responses by examining their temporal changes. Subsequently, a second clustering model was developed to temporally categorize the previous spatiotemporally clustered maps. The data is reshaped to a daily format to discover the distinct CO_2 fluid regimes.

2.4. Feature extraction

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. Twelve features were extracted from the tensors created out of the ERT maps. Each extracted feature represents a specific characteristic of the resistivity change in a region with respect to its neighboring regions.

To reliably incorporate the temporal and spatial changes of moving CO_2 plume, we adopt a tensor-based feature extraction approach. The extraction procedure begins by splitting the original 25m-by-30m CO_2 map for each day into 30 regions. Each region is 5m by 5m (figure 5). For a given location, 5m-by-5m regions for the three consecutive days, namely previous, current and next day, are combined to build a tensor, as illustrated in figure 6. The 3D tensor corresponding to a certain location has a shape of 3x5x5, which incorporates the temporal responses of the previous, current, and next days. Subsequently, the feature extraction was performed on each slice of the 3D array and the final representation of the feature was obtained by averaging the three values for each

spatial location in the 5m-by-5m region. This 3D tensor attempts to capture the spatiotemporal flow dynamics; thereby by enhancing the monitoring of the CO_2 dynamics. The tensor-based feature extraction was conducted and repeated for each of the 30 regions constituting the 25m-by-30m CO_2 storage reservoir. In doing so, each sample in space and time has 12 distinct features. In total, there were 78×25×30 samples that will be subsequently clustered based on the 12 features per sample.

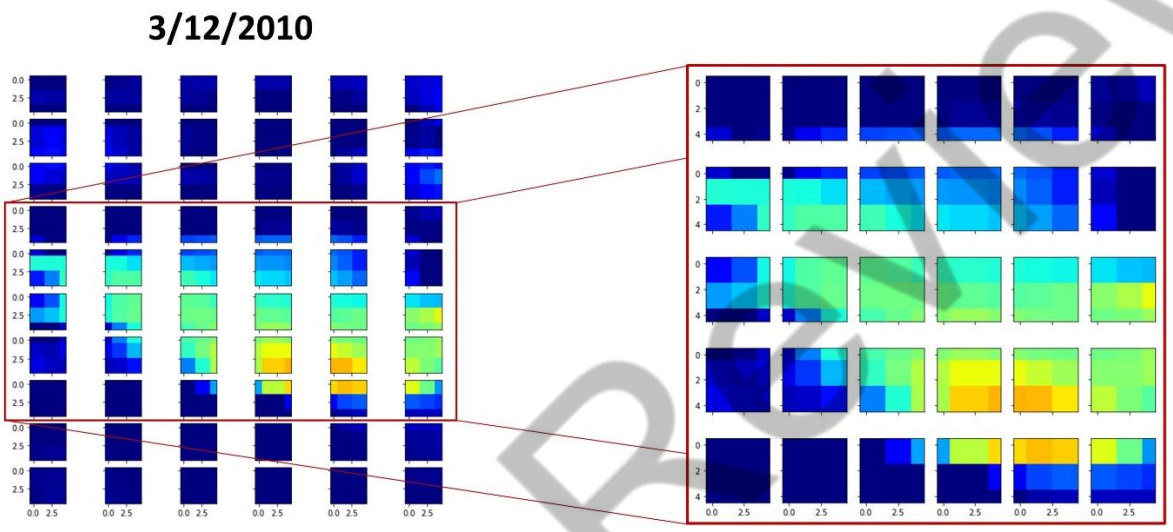


Figure 5. The ERT map of the 25m-by-30m storage reservoir for each day is split into 30 regions. Creation of 30 5m-by-5m regions out of the 25m-by-30m reservoir for the carbon storage. ERT data specific to a 5m-by-5m region for three consecutive days constitutes a 3D tensor for the subsequent tensor-based feature extraction, illustrated in figure 6.

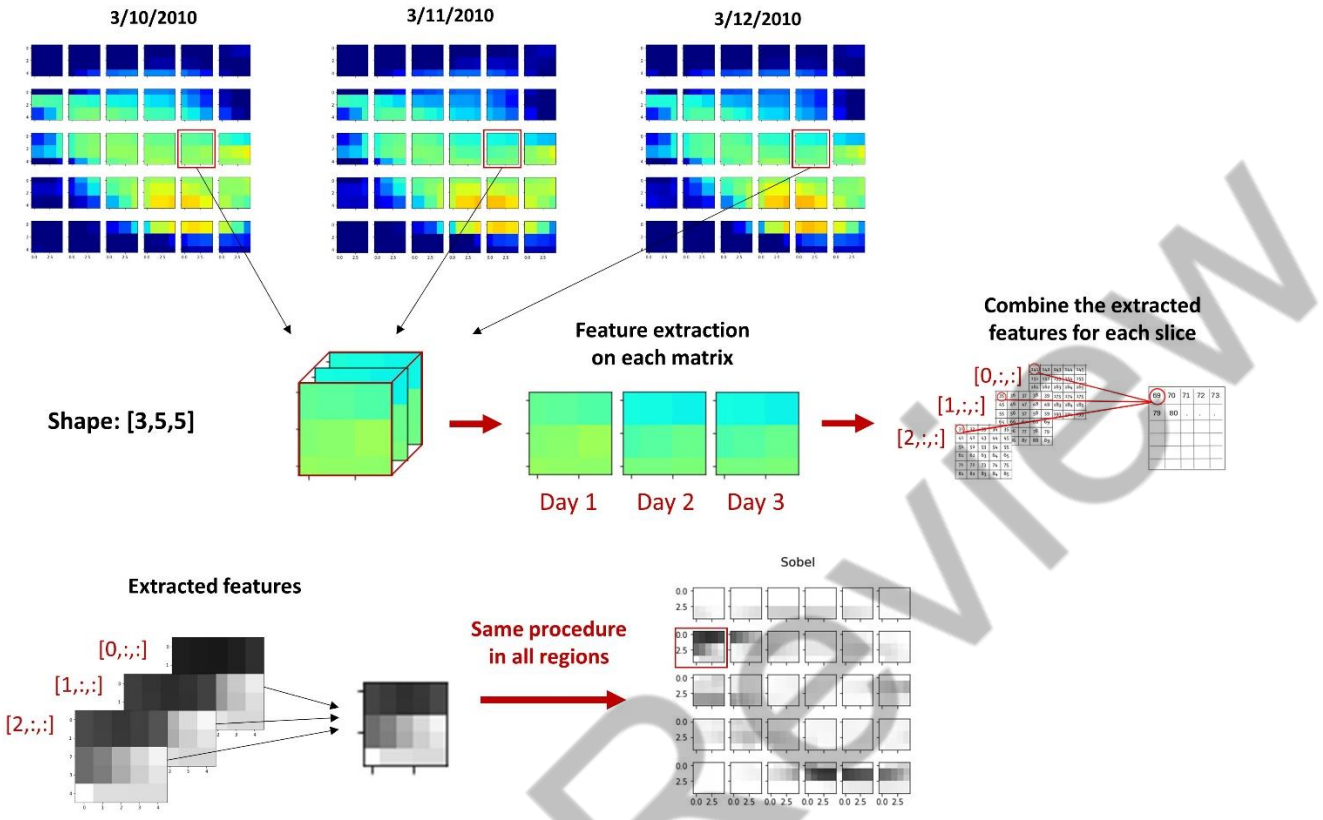


Figure 6. Tensor-based feature extraction design for the 3D tensor of 3x5x5 dimension. To account for the temporal and spatial components of ERT measurements, four steps were implemented: 1) regions are arranged in 3D tensors, 2) feature extraction on each tensor slice, 3) transformation of 3D to a 2D array by averaging, and 4) repetition of the process on all tensor regions.

Several feature extraction methods were applied on the 3 slices in the 3D tensor. Each feature extraction generates three slices per region, which are then averaged into a 2D array for each feature (figure 6). The 12 feature extraction applied on the 3D tensors were: Sobel, gray-level co-occurrence matrix, fast-Fourier transform, linear binary pattern, Hessian matrix, the difference of Gaussians, structure tensor, entropy, height bellow, height above, contrast stretching, and CO_2 saturation change (explained in Table 1). 5 out of 12 features were selected for the subsequent clustering and cluster analysis. The 5 features were selected based on their statistical correlation with the clusters and the sensitivity of the clustered objects to the variations in the features.

Table 1. Brief description of the 12 feature extraction techniques that were applied on the 3D tensors created from the ERT data.

Feature	Description
Sobel (Edges)	Gradient of pixels intensity for edge detection. It captures sharp changes in intensity due to even edges.
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
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.
Difference of Gaussians	Difference between a high-blurred and a low-blurred version of the original image. Two Gaussian kernels are applied at different standard deviations.
Structure tensor	Estimation of the weighted sum of squared differences in a centered pixel window. It captures the orientation and size of image gradients.
Entropy	Statistical measure computed using base 2 logarithm. It represents the level of complexity of a neighboring region around a corresponding pixel.
Height below	Difference between the highest elevation point and current point. It is used to enhance the highest pixel values.
Height above	Difference between the lowest elevation point and current point. It is used to enhance the lowest pixel values.
Contrast stretching	Image enhancement method that spreads out the most frequent intensity values. It rescales the pixels that fall within the 2nd and 98th percentiles.
SCO_2 difference	Difference between the saturations of CO_2 for the previous date and current date. It serves to identify the daily changes of the signal.

2.5. Dynamic time wrapping and k-means

In this study, we used dynamic time wrapping (DTW) as the distance metric for the spatiotemporal clustering. DTW evaluates the optimal aligning between two time-dependent signals (Muller, 2007). This is a technique suitable for time-series datasets due to the ability to capture the similarity and dissimilarity of temporal distances that do not line up in time, speed, or length. In other words, they are invariant to time shifts. These pairwise distances are warped in a nonlinear regime to approximate both sequences. The use of k-means clustering and DTW can be divided into two major steps where the algorithm is going to:

- First, arrange the time-series to similar shapes by using DTW.
- And second, compute the clusters centers with respect to DTW. This will provide an average shape of signals without getting affected by their temporal shifts in their respective cluster.

2.6. Physical interpretation of spatiotemporal clusters

To provide a physical context to the spatiotemporal clusters, we examined the association between cluster counts and the wellbore measurements acquired in the CO_2 injection well (figure 7). In addition, the maps containing the spatiotemporal clusters were temporally clustered to uncover the temporal patterns pertaining to specific CO_2 flow regime. As illustrated in figure 8, the clustered maps were arranged sequentially and then temporally clustered. For the initial spatiotemporal clustering using DTW KMeans, there were 78×25×30 samples representing each unit in space and time, where each sample is described using 12 features. For the second temporal clustering using KMeans, there were 78 samples representing the 78 consecutive days, where each sample is described using 25×30 features, i.e. the cluster number assigned to each grid point in the daily ERT map. The temporal clusters obtained after second clustering represents the stages of CO_2 plume development that can be correlated with distinctive fluid transport mechanisms.

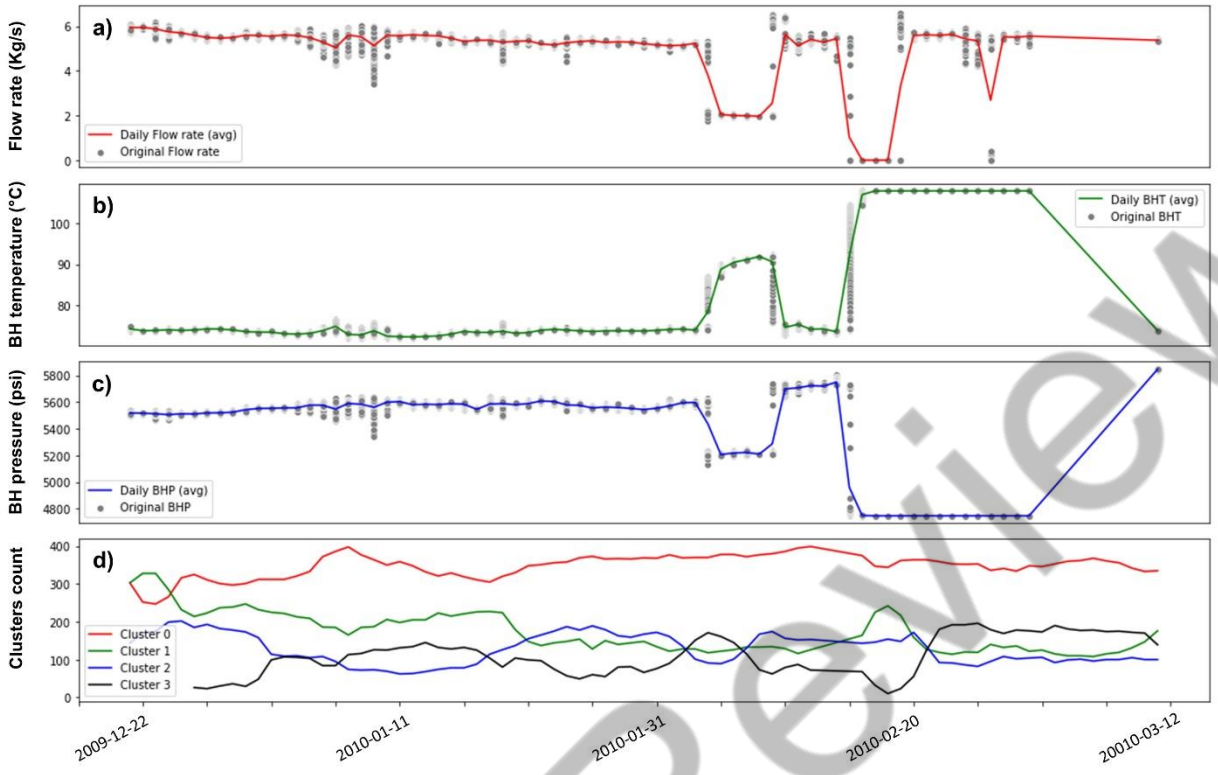


Figure 7. Wellbore measurements: a) flow rate, b) temperature and c) pressure measured in the CO_2 injection well; and d) daily clusters occurrences. We investigated the correlations between changes in the counts of each cluster with the wellbore measurements in Section 3.5.

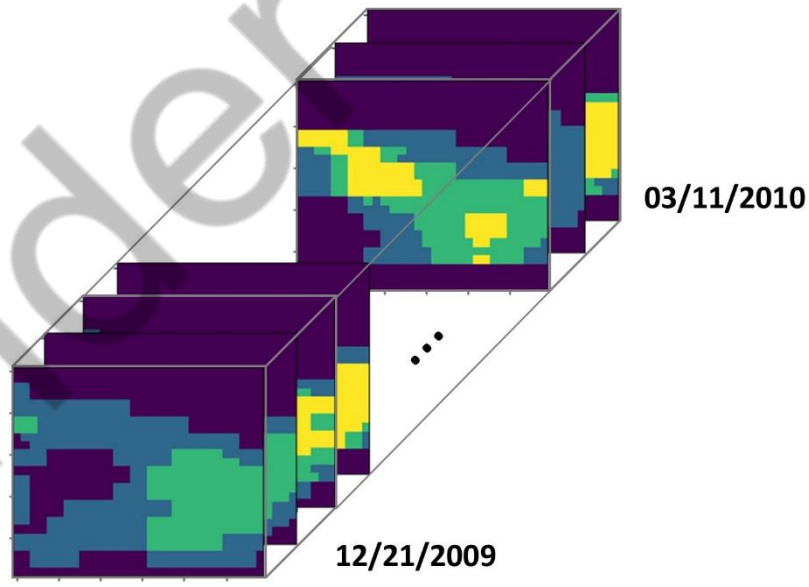


Figure 8. For the second temporal clustering using KMeans, there were 78 samples representing the 78 consecutive days, where each sample is described using 25×30 features, i.e. the cluster number assigned to each grid point in the daily ERT map. We investigated the temporal clustering in Section 3.6.

3. Results

Continuous rapid monitoring of injected CO_2 is necessary to verify the long-term storage and identify potential leakage paths. Data-driven models facilitate rapid monitoring. Unsupervised clustering methods can categorize the subsurface signals based on the content, distribution and movement of CO_2 plume (as shown by Gonzalez and Misra, 2021). Domain expertise is needed to derive novel fundamental insights from the unsupervised spatiotemporal clusters (Chakravarty et al., 2021).

3.1. Validation of the spatiotemporal clusters

Davies-Bouldin, Calinski-Harabasz, and DTW-Silhouette scores are used to evaluate the performance of the proposed spatiotemporal clustering. An important requirement for robust 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.

Clustering metrics, such as Davies-Bouldin, Calinski-Harabasz, and DTW-Silhouette scores, called internal validation measures. These metrics assess the goodness of the data partitioning without the use of any external information. These scores serve as heuristic tools to evaluate the clustering performance. These scores are based on the compactness (similarity) of each cluster and the separation (distinction) between the clusters. A higher value of Calinski-Harabasz score and lower value of Davies-Bouldin score indicates a robust clustering. For the DTW-Silhouette score, the range of performance is set from -1 to 1, where 1 represents the best clustering performance.

Davies-Bouldin index is the maximum of the ratios of sum of within-cluster scatters for the two clusters in a cluster pair to the inter-cluster distance between the two clusters in the cluster pair. A lower Davies-Bouldin index indicates good clustering. Calinski-Harabasz index is also known as the variance ratio criterion. Calinski-Harabasz index is the ratio of the weighted variance of cluster centers with respect to a global cluster center to the weighted sum of within-cluster variances. This index does not have an upper acceptable cut-off. Higher values of Calinski-Harabasz index indicates good clustering.

In the DTW K-means implementation, four robust clusters were identified using the Davies-Bouldin and Calinski-Harabasz scores. Figure 9 displays the ERT-based CO_2 saturation and the corresponding spatiotemporal clusters for a specific day in the time-lapse dataset. Clusters are primarily related to CO_2 content, ranging from zero (0) to high (3). In addition, unlike simple thresholding, the spatiotemporal clusters also consider the temporal changes in a specific spatial region due to the spatiotemporal feature extraction that was implemented prior to the clustering. Unlike traditional threshold-based methods to categorize spatial data, the proposed spatiotemporal clustering takes into account the rate of change of CO_2 as a function of distance and time, and the regional as well as temporal distribution of the fluids. The proposed clustering better honors the regional and temporal variations and movement of the CO_2 plume as compared to the simple threshold-based categorization methods.

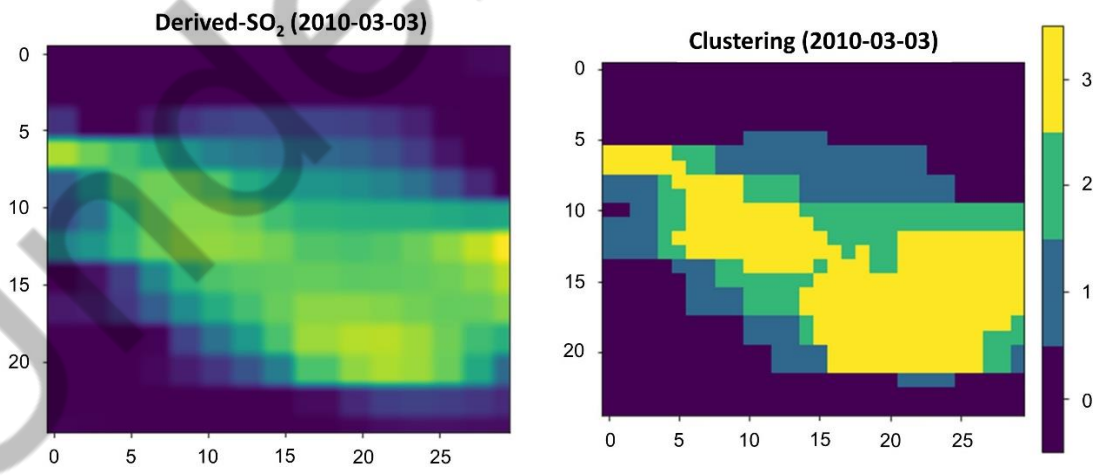


Figure 9. Left: Distribution of CO_2 content as derived from the ERT measurement on 03/03/2010**.** Right: The spatiotemporal clusters computed using the DTW K-means clustering. Clusters strongly correspond to no-CO_2 (cluster 0), low CO_2 (cluster 1), medium CO_2 (cluster 2), and high CO_2 (cluster 3). Injection well

"F2" is set on the left at a distance of 229 feet from the "F1" monitoring well, and “F3” on the right at a distance of 327 feet.

As shown in Figure 9, the clusters allow us to define the shape and boundaries of the CO_2 movement in the subsurface storage reservoir. The optimal number of clusters is four as confirmed by a Davies-Bouldin index of 0.71, a Calinski-Harabasz of 262791, and a DTW-silhouette score of 0.58. These clusters do not merely represent regions; rather the clusters represent objects evolving in space and time. The scores support the effectiveness of the spatiotemporal clustering. With the time-lapse clustered data, the evolution of CO_2 can be distinguished where an early stage only displays clusters of “0” (non-CO_2), “1” (low-CO_2), and “2” (medium-CO_2). The shapes of clusters continuously change, being cluster “3” (high-CO_2) the one with the highest spatiotemporal change of CO_2 occurrences and displacements. We also can distinguish slow access of certain reservoir regions from the clustered plume shape where movement is observed at the clusters “1”, “2”, and “3”. These moving clusters will serve to extract further information at a much more granular level (see Appendix B for details). Appendix B presents the spatiotemporally clustered ERT maps for days 1, 10, 20, 30, 40, 50, 60, 70, and 78. The spatiotemporal clusters are within a 25m-by-30m region between the F2-F3 monitoring wells. From days 1 to 30, a steady increase of high CO_2 is observed and linked to the early development of the CO_2 plume. From days 40 to 78, the shape of the CO_2 plume drastically changes with a clear CO_2 movement at all levels.

3.2. Comparison of DTW K-means against three traditional clustering methods

To test the efficacy of DTW K-means on spatiotemporal datasets, the performance of DTW K-means was compared against three traditional clustering methods, namely K-means with Euclidean distance, agglomerative, and mean-shift clustering methods. All clustering methods processed the same tensor-based features to obtain 4 clusters. K-means, meanshift and agglomerative display poor clustering performance as presented in figure 10. Qualitatively, meanshift clustering fails to adequately describe the spatial distribution of CO_2 content. Clusters obtained using the agglomerative and Euclidean k-means clustering are similar during the late-time CO_2 storage but very different during the early-time CO_2 storage (figure 11). Specifically, agglomerative displays a drastic

reduction of cluster “0” (non-CO_2) and an increase of cluster “1” (low-CO_2), which indicates the poor temporal performance at certain injection phases.

In addition, Davies-Bouldin and Calinski-Harabasz scores were calculated to quantitatively compare the clustering performance and validate the efficacy of DTW-KMeans as compared to traditional clustering approaches. The quantitative comparison aims to find the most appropriate clustering approach based on the lowest Davies-Bouldin index and the highest Calinski-Harabasz index. Table 2 presents the scores of each method, wherein DTW K-means achieves the highest clustering performance, followed by Euclidean k-means, and agglomerative clustering. Meanshift clustering displays the worst scores in the clustering task. In summary, the use of temporal distances, such as DTW, in KMeans along with tensor-based feature extraction can capture the variations in CO_2 with time and distance; thereby enhancing the spatiotemporal clustering of moving objects.

Table 2. Internal metric scores for the clustering using various algorithms, namely DTW k-means, Euclidean k-means, meanshift, and agglomerative techniques. A low Davies-Bouldin index and high Calinski-Harabasz index indicates a better clustering performance. DTW k-means has the lowest Davies-Bouldin and highest Calinski-Harabasz score. DTW k-means is best performing method while Meanshift is the worst performing method for the ERT-based monitoring of CO_2 plume movement.

Clustering algorithms	**Score**	
	Davies-Bouldin	*Calinski-Harabasz*
DTW K-means	0.71	262791
Euclidean K-means	0.83	157866
Agglomerative	0.95	131593
Meanshift	1.01	69438

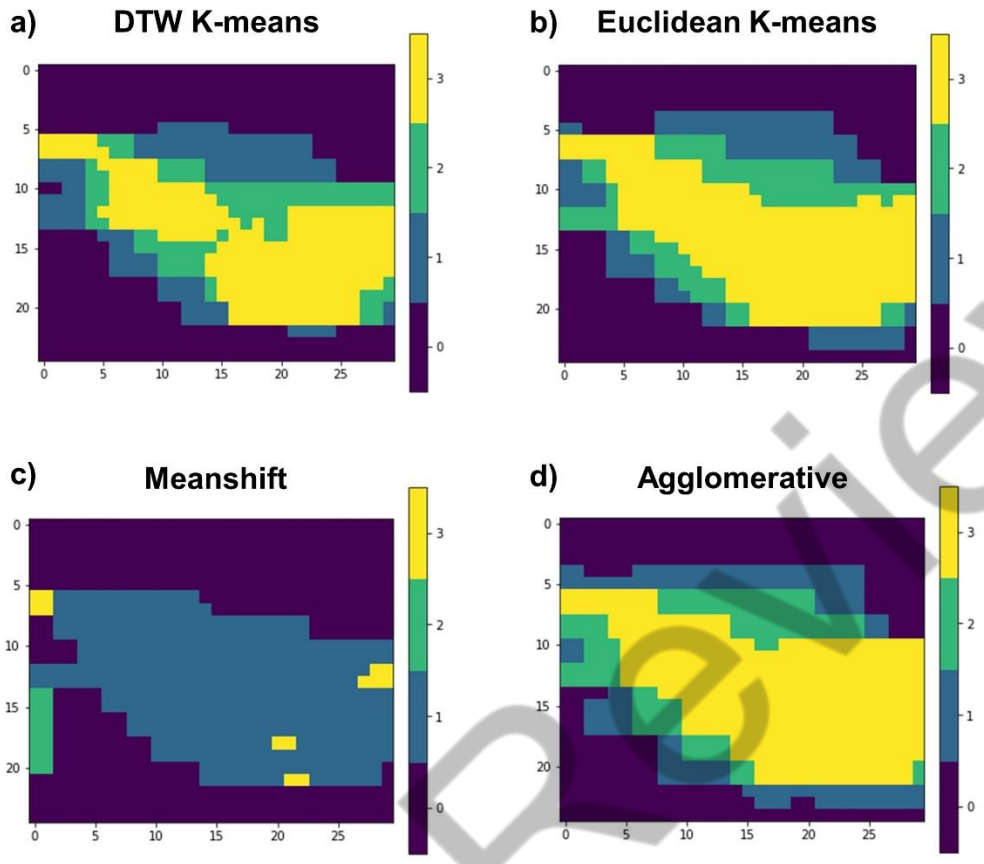


Figure 10. Clusters obtained using a) DTW k-means, b) Euclidean k-means, c) meanshift and d) agglomerative clustering. Qualitatively k-means and agglomerative display a similar behavior of low migration movements, while meanshift display a poorly clustering performance. DTW k-means provides a better granularity.

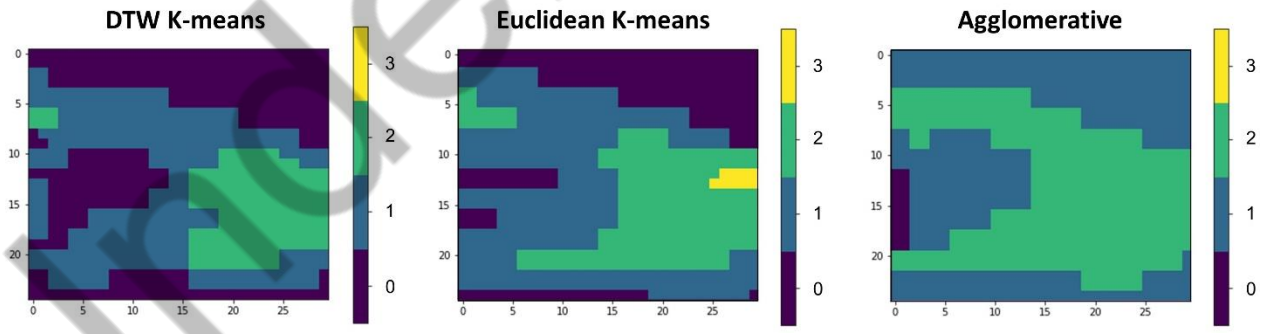


Figure 11. DTW k-means, Euclidean k-means and agglomerative clustering results for "2009-12-21", at the very initial stage of geological storage. Agglomerative displays 3 classes of non-CO_2 (cluster 0), low CO_2 (cluster 1), and medium CO_2 (cluster 2), while Euclidean k-means shows high CO_2 (cluster 3) away from the injection site. Differences are also linked to non-CO_2 and low CO_2 content.

3.3. Comparison between traditional and tensor-based feature extraction

We implement a tensor-based feature extraction to ensure all the samples contain sufficient spatial and temporal aspects of the fluid movement. To further validate and investigate the use of tensor-based feature extraction, a comparison between tensor-based and "traditional" feature extraction is conducted in this section. The traditional approaches extract only spatial features for each daily image, without considering both the temporal and regional variations of the time-series dataset. In figure 12, clustering performance using the tensor-based features is compared against that using traditional features for three days.

For date "2009-12-25", which is the early stage of CO_2 injection, the clustering using traditional features displays high degree of tortuous paths with several isolated regions without CO_2 (figure 12 right). In addition, the clustering using traditional features produces regions of high CO_2 during the early stage of CO_2 injection. All these observations from the right-hand side subplots in figure 10 indicates the unreliable clustering obtained using traditional features. For "2010-02-10", which represents the mid-stage of CO_2 injection, the shapes of the clusters obtained using traditional and tensor-based features are distinct. The clustering using traditional features exaggerates the presence of cluster 3 (high CO_2), as shown in the bottom two rows of subplots on the right in Figure 11. Lastly, for the final time-lapse results on the 90^{th} day ("2010-03-08") of ERT measurements, the traditional model fails to separate the CO_2 levels. Clusters display an overestimation of high CO_2 and unclear boundaries at all levels.

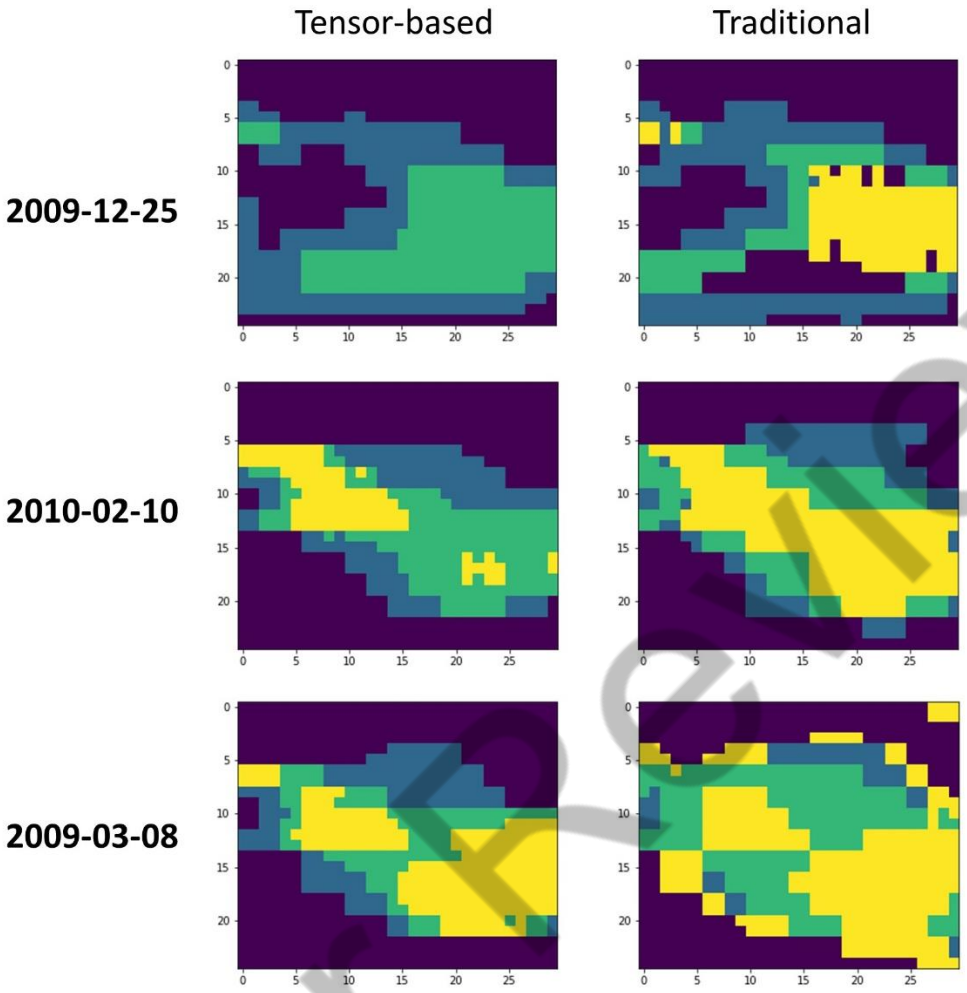
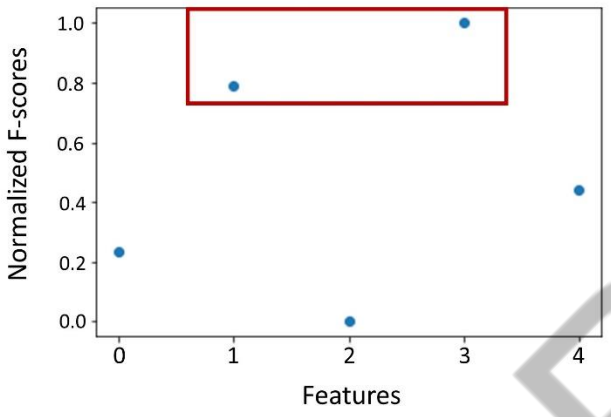


Figure 12. DTW k-means clustering using tensor-based features and traditional spatial features. The clustering performance using traditional features is lower than that using tensor-based features. Injection well "F2" is set on the left at a distance of 229 feet from the "F1" monitoring well, and “F3” on the right at a distance of 327 feet.

3.4. Discovering new geophysical signatures for monitoring geological carbon storage

In this section, we examine the relevance and association of various ERT-based features with the spatiotemporal clusters. By doing this, we discover the most-informative geophysical signatures that can assist the monitoring of geological carbon storage. The statistical significance of these geophysical signatures was tested using one-way ANOVA (analysis of variance) (figure 13). The ANOVA F-test measures the statistical differences between the means of the features of two or more groups. A low F-test value for a feature *X1* indicates that the samples in various clusters have nearly similar values for the feature *X1*; in other words, the clusters overlap along the feature *X1*. A high F-test value for a

feature *X2* indicates that the samples belonging to different clusters will have different values for the feature *X2*; in other words, the clusters are well separated along the feature *X2*. A high F-test for the ERT-based monitoring of geological carbon storage reservoir, contrast stretching and fast-Fourier transform are the most-informative geophysical signatures that capture both the CO_2 content and its displacement. According to the ANOVA F-test, contrast stretching and fast-Fourier transform are the two features whose means show the highest statistical differences between the 4 spatiotemporal clusters.



Feature Index	Feature Name
0	Structure tensor
1	Fast-Fourier transform
2	SCO_2 difference
3	Contrast stretching
4	Height bellow

Figure 13. One-way ANOVA test was implemented to compute the normalized F-test values. The normalized F-test values enabled the discovery of the most-informative ERT-based feature that can be used to monitor the geological carbon storage. Contrast stretching and fast-Fourier transform have the highest F-test value. This confirms that contrast stretching and fast-Fourier transform are the most-informative, relevant geophysical signatures for the ERT-based monitoring of geological carbon storage.

Post-hoc Tukey's Honest Significant Difference (HSD) test was conducted to analyze the pairwise cluster differences for each feature (table 3). Tukey HSD test is performed after the ANOVA F-test. First, ANOVA F-test is used to identify the most-informative features. Following that, Tukey's HSD test computes the pairwise differences between clusters for each feature. A large difference between two clusters *C1* and *C2* for a feature *X1* indicates that the samples in cluster *C1* and the samples in cluster *C2* have drastic separation along the feature X1; in other words, the values of feature *X1* for samples in *C1* and *C2* are very different, i.e. distinct.

The Tukey's HSD test was performed on the two features with the highest ANOVA F-test values. The two features are contrast stretching and fast-Fourier transform. Clusters "0" and "3" are the most different. This large difference supports our claim that

cluster “0” represents the absence of alteration due to the injected CO_2, representing absence of CO_2. Along those lines, cluster “3” represents regions of large alterations due to the injected CO_2, which corresponds to high CO_2 content. Conversely, clusters “2” (medium-CO_2) and “3” (high-CO_2) reveal the lowest difference and highest similarity among other cluster pairs. Difference between clusters “0” and “1”, and that between clusters “1” and “2” is more than that between clusters “2” and “3”. The test also served to establish contrast stretching, as the strongest feature due to their high clusters difference.

Table 3. Tukey HSD for post hoc analysis of the two most-informative features (contrast stretching and fast-Fourier transform) and CO_2 saturation. Mean difference of a feature for a cluster pair indicates the significance of that feature in differentiating the two clusters. Cluster “0” indicates non-CO_2 content, and clusters “1”, “2”, and “3” their respective level of saturation (low, medium, and high).

Feature	**Clusters being compared**		**Mean difference**	**Normalized difference**
Contrast stretching	Cluster #	Cluster #		
	0	1	834395206	0.45
	0	2	158820770	0.85
	0	3	1867909636	1
	1	2	753812495	0.4
	1	3	1033514430	0.55
	2	3	279701934	0.15
Fast Fourier Transform	Cluster #	Cluster #		
	0	1	5.89	0.41
	0	2	11.09	0.76
	0	3	14.51	1
	1	2	5.2	0.36
	1	3	8.62	0.59
	2	3	3.42	0.24
CO_2 Saturation	Cluster #	Cluster #		
	0	1	5.77	0.40
	0	2	10.92	0.75
	0	3	14.50	1.00
	1	2	5.15	0.36
	1	3	8.73	0.60
	2	3	3.58	0.25

To confirm the usefulness of the two above-mentioned features as the most-informative geophysical signatures for ERT-based monitoring of geological carbon storage, we use boxplots and 2D KDE plot. In the boxplots (figure 14 left and center), the colored boxes represent the interquartile range, where 50% of the data resides. For the two dominant features, i.e. contrast stretching and fast-Fourier transform, the interquartile range do not overlap, confirming the significance of these two features in differentiating the clusters. The whiskers in the boxplot indicate 5th and 95th percentile, where 90% of the data resides. Considering the whiskers, contrast stretching provides a better separation of the spatiotemporal clusters and indicates a slight similarity/overlap between clusters “2” and “3”. The box plots validate the large statistical difference between clusters “0” and “3”. Considering the whiskers for fast-Fourier-transform, clusters overlap is higher and their separation is lower as compared to that for contrast stretching. Finally, the 2D-KDE (figure 14 right) shows the distribution of the clusters in a two-dimensional feature space made of contrast stretching and fast-Fourier transform. The 2D-KDE plot confirms the strengths of these two features in differentiating the spatial regions and time periods based on CO_2 content, distribution, and displacement. All these plots validate the usefulness and relevance of contrast stretching and fast-Fourier transform as the geophysical signatures for monitoring the CO_2 plume migration.

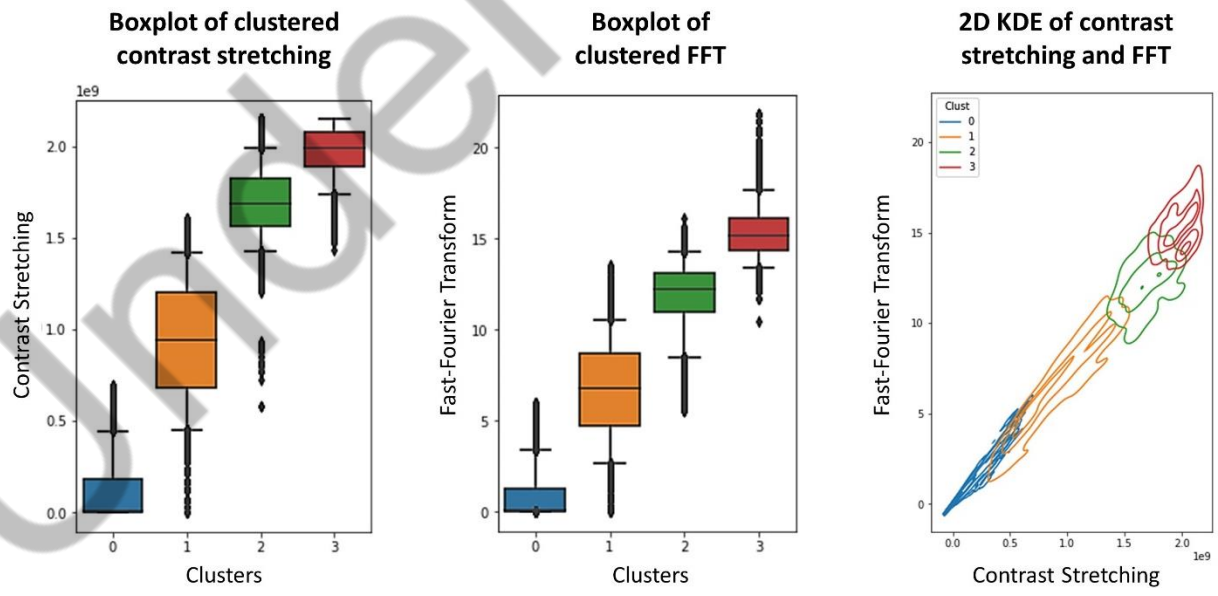


Figure 14. Left and center: Boxplots of clustered contrast stretching and fast-Fourier transform (FFT). The x-axis represents the cluster numbers. The whiskers on the boxplots are defined for a low 5th percentile

and a high 95th percentile. Right: 2D-KDE of contrast stretching and FFT. Overlapping of clusters can be observed on the 2D-KDE and boxplots due to the CO_2 plume migration. The bivariate KDE is mapped with their respective classes to show their conditional distribution. The outermost represents the high-CO_2, the middle contour the medium (cluster 2) and low (cluster 1) -CO_2, and the innermost the non-CO_2 (cluster 0).

3.5. A new correlation between injection rate and the spatiotemporal cluster count

The physical relevance of the clusters is investigated by comparing the cluster counts against the pressure-rate-temperature measurements at the injection well. Figure 15 displays the daily flow rate, temperature, pressure, and the daily count of cluster “3” (high CO_2). Strong correlation in form of simultaneous decrease/increase of cluster “3” count and flow rate are observed for dates ranging from “2010-01-30” to “2010-03-09”. Such a correlation can be attributed to the drainage and imbibition processes associated with CO_2 plume migration. The first decrease in flow rate from 5.2 kg/s to 2 kg/s leads to a reduction in cluster “3” count from 171 to 62. A subsequent increase of flow rate to a value of 5.6 kg/s leads to an increase in the cluster “3” count to 82. The following flow rate decrease to 0 kg/s leads to a drop in cluster “3” count to 10. Finally, with a significant increase in flow rate from 0 kg/s to 5.5 kg/s is associated with increase in cluster “3” count from 10 to 190. This final stage is followed by a decrease of flow rate to 2.6 kg/s and a drop in cluster count to 169 respectively. Existence of such correlation is a notable discovery.

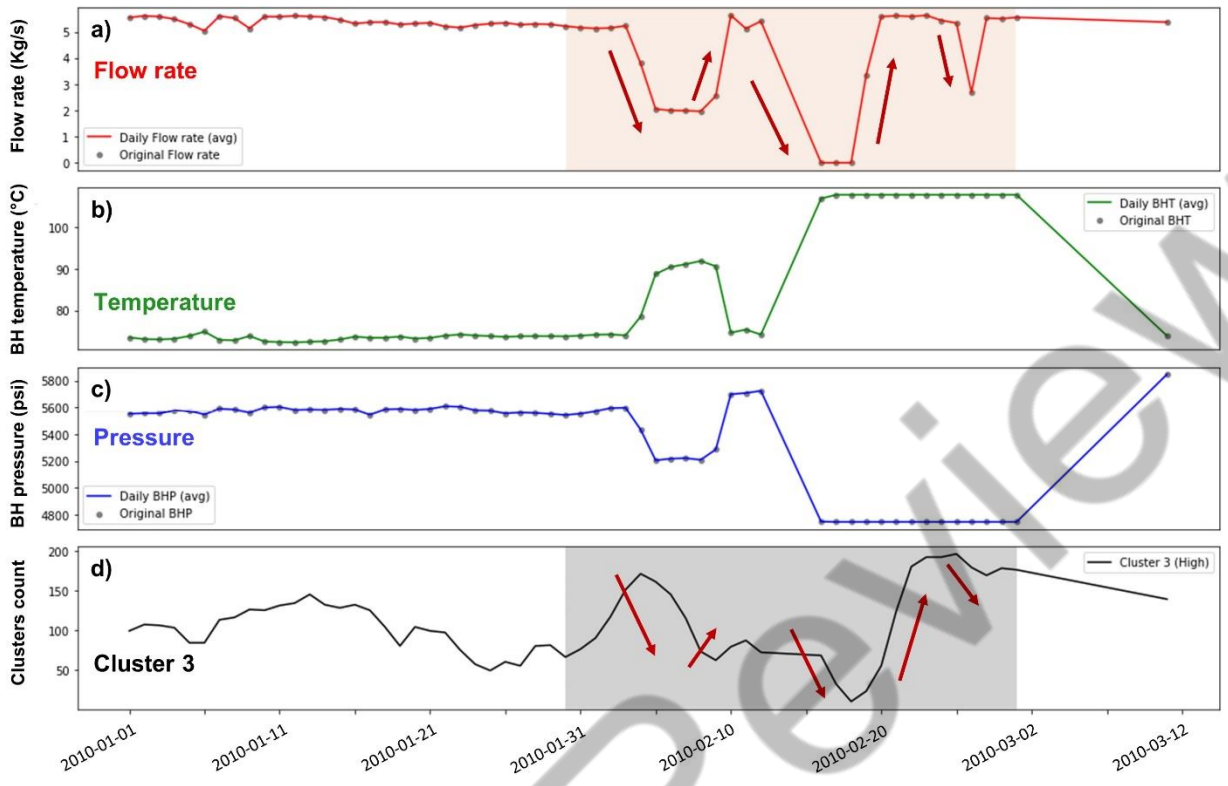


Figure 15. Wellbore measurements: a) flow rate, b) temperature and c) pressure of injection well; and d) daily count of cluster 3 (high SCO_2). The gray background corresponds to the dates from "2010-01-30" to "2010-03-09" where a change of all measurements is observed.

3.6. A novel temporal clustering following the spatiotemporal clustering

Temporal clustering of 78 days of spatiotemporally clustered ERT maps identified specific sequences of flow regimes developing in the CO_2 injection reservoir. These temporal clusters are labeled as T0 to T5. Figure 16 shows the six temporal clusters. An optimal number of 6 clusters was determined by using Davies-Bouldin and Calinski-Harabasz indices. A physics-based interpretation of the temporal clusters is presented in table 4, where each cluster is associated with distinct CO_2 plume behavior. Each cluster displays a different number of daily clustered images: cluster "T0" comprises the first 4 days, cluster "T1" contains 11 days, cluster "T2" contains 16 days, cluster "T3" contains 9 days, cluster "T4" contains 20 days, and cluster "T5" contains the final 18 days. These temporal clusters can be associated with stages of CO_2 plume evolution during the geological carbon storage, such as the plume early development, changes of plume shape, increases/decreases of CO_2 saturation, and changes in certain plume

constituents. Such observed changes over time can be associated with specific dominant CO_2 flow mechanisms. For instance, movement at the top or bottom of CO_2 may be related to gravitational and buoyancy forces, or capillary to geological heterogeneity. Lastly, the temporal clusters exhibit concordance with the injection well conditions. Unique sequences of temporal behaviors associated with CO_2 flow regimes were uncovered and linked to the wellbore measurements at the injection well (figure 17). Drastic changes in the injection well lead to the flow regimes being captured by cluster "T4", where the drastic displacement and reduction of the high-CO_2 regions are observed.

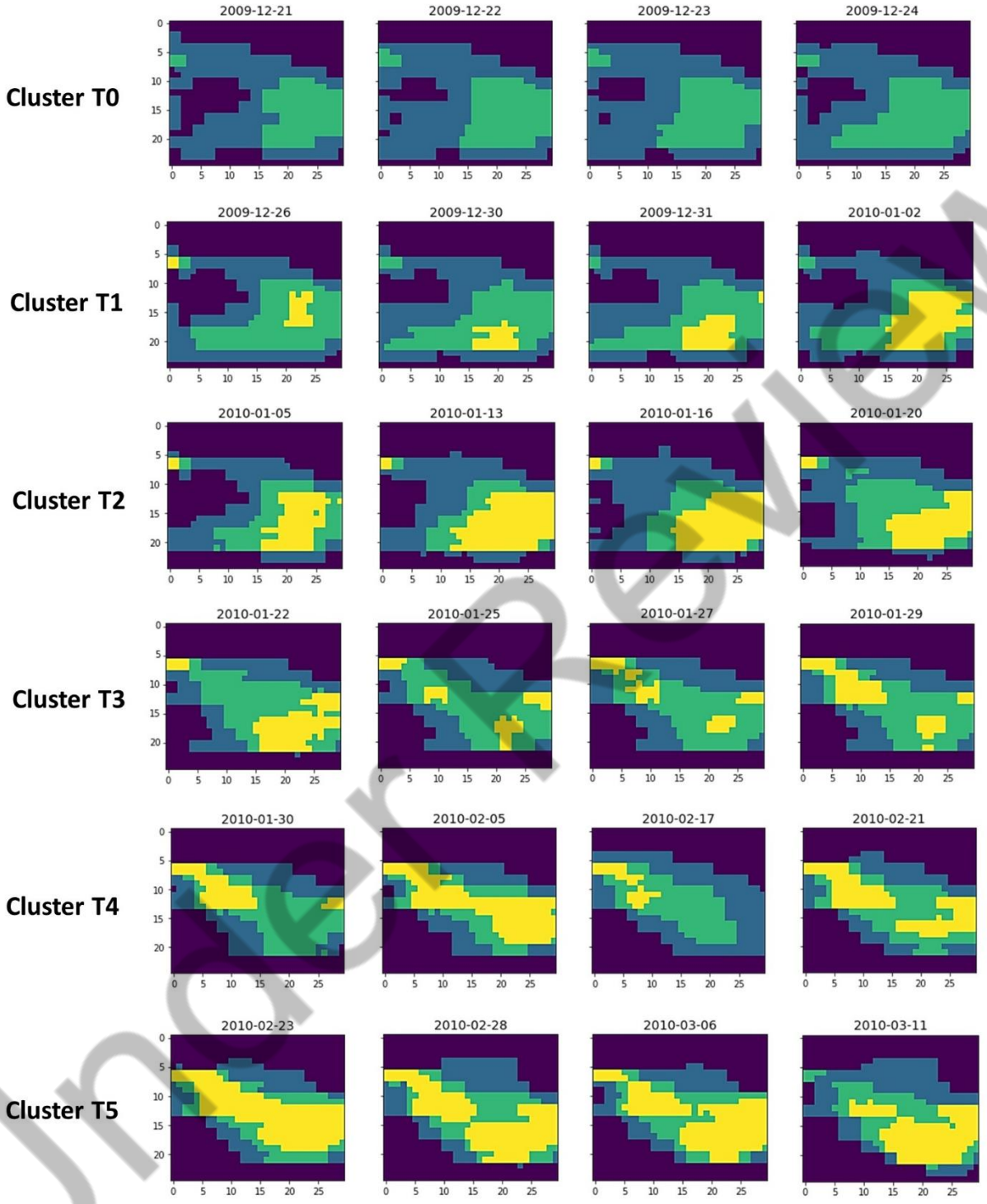


Figure 16. Temporal clusters "T0 to T5"" obtained by clustering the spatiotemporally clustered ERT maps shown in Figure 8. Six temporal clusters were determined that correspond to specific CO_2 flow regimes. "T0", "T1", "T2", "T3", "T4", and "T5" contains 4, 11, 16, 9, 20, and 19 maps corresponding to consecutive 78 days.

Table 4. Physics-based interpretation of the temporal clusters derived from the spatiotemporally clustered ERT Maps. The temporal clusters are shown in Figure 16.

Temporal clusters	Clustering analysis
Cluster T0	Dominance of spatiotemporal clusters 1 and 2 (low and medium CO_2 saturation)
Cluster T1	Occurrence and increment of spatiotemporal cluster 3 (high CO_2 saturation)
Cluster T2	No significant changes in the spatiotemporal clusters
Cluster T3	Slow spatial movement of cluster 3 and subsequent changes in the spatiotemporal cluster shapes
Cluster T4	Quick fluctuations of spatiotemporal cluster 3 where multiple changes of high-CO_2 saturations are observed.
Cluster T5	Dominance of spatiotemporal cluster 3.

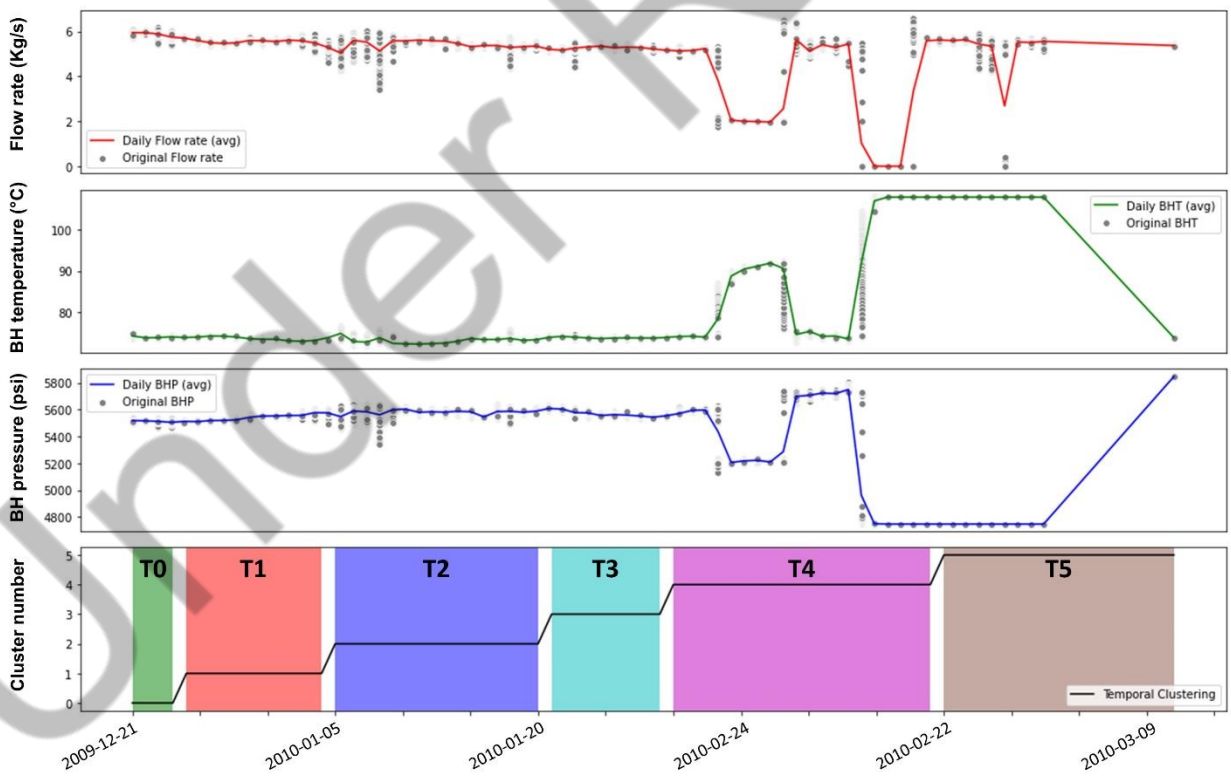


Figure 17. Comparison of the wellbore measurements against the temporal clusters “T0 to T5” obtained by processing the spatiotemporally clustered ERT maps. The temporal clusters indicate flow regimes in the

carbon-storage reservoir that seem to correlate with the drastic changes in flow rate, temperature, and pressure. Detailed description of the temporal clusters T0 to T5 is presented in Table 4.

4. Conclusions

Spatiotemporal clustering processed daily electrical resistivity tomography (ERT) maps to visualize the CO_2 plume migration during a geological carbon storage. The ERT data is made available by SECARB (2017) and well described by Carrigan et al. (2013) and Yang et al. (2014). A novel unsupervised learning workflow was designed to incorporate spatial and temporal components of the moving CO_2 content. Creation of neighboring regions, tensor-based feature extraction, and dynamic time warping (DTW) K-means are crucial for the proposed workflow. DTW as distance metric was used to compare temporal sequences of similar shapes where daily changes of spatial signals are associated with the evolution and movement of CO_2 plume. A comparison between agglomerative, meanshift, and Euclidean k-means confirmed the efficacy of DTW distance metric for the spatiotemporal clustering. In the DTW K-means implementation, four robust clusters were identified using the Davies-Bouldin and Calinski-Harabasz scores. The optimal number of clusters is four as confirmed by a Davies-Bouldin index of 0.71, a Calinski-Harabasz of 262791, and a DTW-silhouette score of 0.58. These clusters do not merely represent regions; rather the clusters represent objects evolving in space and time. Clusters are primarily related to CO2 content, ranging from zero (0) to high (3).

Unlike simple thresholding, the spatiotemporal clusters also consider the temporal changes in a specific spatial region due to the spatiotemporal feature extraction that was implemented prior to the clustering. The use of temporal distances, such as DTW, in KMeans along with tensor-based feature extraction can capture the variations in CO2 with time and distance; thereby enhancing the spatiotemporal clustering of moving objects. We implement a tensor-based feature extraction to ensure all the samples contain sufficient spatial and temporal aspects of the fluid movement. We discovered a strong correlation between cluster “3” count and CO_2 injection rate. Existence of such correlation is a notable discovery.

ANOVA F-test, Tukey's HSD, and kernel density estimation plots established contrast stretching and fast-Fourier transform as the most informative geophysical signatures. According to the ANOVA F-test, contrast stretching and fast-Fourier transform are the two features whose means show the highest statistical differences between the 4 spatiotemporal clusters. The box plots validate the large statistical difference between clusters "0" and "3". For fast-Fourier-transform, clusters overlap is higher and their separation is lower as compared to that for contrast stretching. The 2D-KDE plot confirms the strengths of these two features in differentiating the spatial regions and time periods based on CO_2 content, distribution, and displacement.

5. 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. Special acknowledgments to Dr. Xianjin Yang for his guidance and support through this research.

6. References

Cao, N., Lin, C., Zhu, Q., Lin, Y., Teng, X., and Wen, X. (2018). Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data. IEEE Transactions on Visualization and Computer Graphics, vol. 24, no. 1, pp. 23-33. doi: 10.1109/TVCG.2017.2744419.

Carrigan, C. R., Yang, X., LaBrecque, D. J., Larsen, D., Freeman, D., Ramirez, A. L., Daily, W., Aines, R., Newmark, R., Friedmann, J. and Hovorka, S. (2013). Electrical resistance tomographic monitoring of CO2 movement in deep geologic reservoirs. International journal of greenhouse gas control, 18, 401-408. https://doi.org/10.1016/j.ijggc.2013.04.016.

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.

Chen, X.C., Faghmous, J.H., Khandelwal, A., Kumar, V. (2015). Clustering dynamic spatio-temporal patterns in the presence of noise and missing data. In: International joint conference on artificial intelligence, pp 2575–2581

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.

Gonzalez, K., & Misra, S. (2021). Visualization of the sequestered carbon-dioxide plume in the subsurface using unsupervised learning. https://www.essoar.org/doi/abs/10.1002/essoar.10507269.2

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.

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.

Liu, H., Zhan, Q., Yang, C., and Wang, J. (2018). Characterizing the Spatio-Temporal Pattern of Land Surface Temperature through Time Series Clustering: Based on the

Latent Pattern and Morphology. Remote Sens. 10, 654. https://doi.org/10.3390/rs10040654

Liu, J., Xue, C., He, Y., Dong, Q., Kong, F., Hong, Y. (2018). Dual-constraint spatiotemporal clustering approach for exploring marine anomaly patterns using remote sensing products. IEEE J Sel Top Appl Earth Obs Remote Sens 11(11):3963–3976. https://doi.org/10.1109/jstars.2018.2873216

Müller, M. (2007). Dtw-based motion comparison and retrieval. Information Retrieval for Music and Motion, 211-226.

Ni, H., & Benson, S. M. (2020). Using unsupervised machine learning to characterize capillary flow and residual trapping. Water Resources Research, 56(8), e2020WR027473.

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.

Pires de Lima, R., and Lin, Y. (2019). Geophysical data integration and machine learning for multi-target leakage estimation in geologic carbon sequestration. SEG Technical Program Expanded Abstracts: 2333-2337.

SECARB, 2017. Lawrence Livermore National Laboratory, Southern States Energy Board, National Energy Technology Laboratory, SECARB Electrical Resistance Tomography Data, 2017-12-01, https://edx.netl.doe.gov/dataset/secarb-electrical-resistance-tomography-data, DOI: 10.18141/1468466

Yang, X., Chen, X., Carrigan, C.R. & Ramirez, A.L., 2014. Uncertainty quantification of CO2 saturation estimated from electrical resistance tomography data at the Cranfield site, Int J Greenh Gas Con, 27, 59-68.

Zhou, Q., Yang, X., Zhang, R., Hosseini, S. A., Ajo-Franklin, J. B., Freifeld, B. M., Daley, T., and Hovorka, S. (2020). Dynamic Processes of CO2 Storage in the Field: 1. Multiscale and Multipath Channeling of CO2 Flow in the Hierarchical Fluvial Reservoir at Cranfield, Mississippi. Water Resources Research, 56, e2019EF001360. https://doi.org/10.1029/2019WR025688

Appendix A. Thresholding Methods vs. k-Means Clustering for Monitoring the Spatiotemporal Distribution of Carbon Dioxide in the Geological Carbon-Storage Reservoir

Both thresholding and clustering methods can retrieve regions exhibiting similar flow characteristics and seismic properties. A thresholding method, called the multi-Otsu thresholding, is compared with k-Means clustering. The multi-Otsu thresholding is a multilevel method used to separate pixels based on their intensity level. The selection of threshold regions was developed as the maximum between-class variances.

In this study, the multi-Otsu thresholding was implemented to create five unique regions from the CO_2 image. Four threshold values were determined at a pixel intensity of 30, 90, 154, and 218. The threshold-based clusters are compared against the clusters computed using the two-level k-means, meanshift, and agglomerative clustering. The comparison was performed using the adjusted random and homogeneity scores. Table A1 presents the estimated values of similarities between the clusters computed using the two-level k-means, agglomerative, meanshift, and multi-Otsu thresholding, where a value equal to one represents a perfect match. Based on the pair-wise comparisons, the lowest scores were obtained for the multi-Otsu comparisons, achieving an average adjusted random of around 0.87 and an average homogeneity score of around 0.84.

Table A1. Comparison of two-level clustering using K-means, meanshift and agglomerative clustering, and the multi-Otsu thresholding. A pair-wise score close to one indicates a high similitude between two clustering methods. Multi-Ostu thresholding has lower similarity with other clustering methods.

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***

Need of feature extraction for the spatiotemporal characterization

Threshold based methods are not flexible. The selection of optimum thresholds requires prior knowledge to be the most informative. Rigidity of the thresholds lead to bias and is a time-consuming process. Furthermore, thresholding generally considers one or two features that are not adequate to describe a spatial phenomenon. Feature extraction

helps account for statistical information of the neighboring regions, rate of change of pixel intensity, presence of certain local structures, and many other spatial characteristics that cannot be accounted for by using only pixel intensity. Pixel intensity only focus on the magnitude at a specific location, while feature extraction allows for a much broader focus on spatial changes, variations, similarities, and their rates of change at specific location, in a region, and within regions. As compared to the use of only pixel intensity, feature extraction significantly improves the spatial characterization. Feature extraction leads to identification of generalizable geophysical signatures and patterns that can improve the scalability and generalization of the clustering methods. Pixel intensity cannot describe the entire spatial phenomenon, while feature extraction allows for a more generalized description of a spatial phenomenon. Most importantly, Figure 10 confirms the higher importance of extracted features, such as Fourier Transform and Wavelet Transform, as compared to the pixel intensity.

Need of unsupervised clustering for spatiotemporal characterization

Unsupervised clustering considers a set of local information of distinctive characteristics such as edges, shape, and texture, which extracts the desired distinctive characteristics at a more efficient and deeper level. Another key drawback of threshold method is the static component of the threshold regions. Hence, new thresholds are needed to be defined for each new image, new field, new sensors, and new configurations. Errors commonly appears 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.

Few benefits of using unsupervised methods as compared to traditional thresholding methods or geophysical models include:

1. Unsupervised learning workflow can process sensor data irrespective of the sensor types, transmitter/receiver configurations, sensor-data processing, engineering designs, CO2 injection schedules, and geological properties of the CO2 injection reservoir.
2. Rapid spatiotemporal monitoring of CO2 plume movement can be achieved for any type of geophysical data acquired from any type of geological carbon storage site without requiring a specific assumption of the geophysical model or specific source-sensor configuration.
3. The unsupervised learning workflow will allow pathways for improved assimilation of expert domain knowledge in form physical interpretations and the infusion physical principles.

Appendix B. Spatiotemporal evolution of CO_2 content and distribution

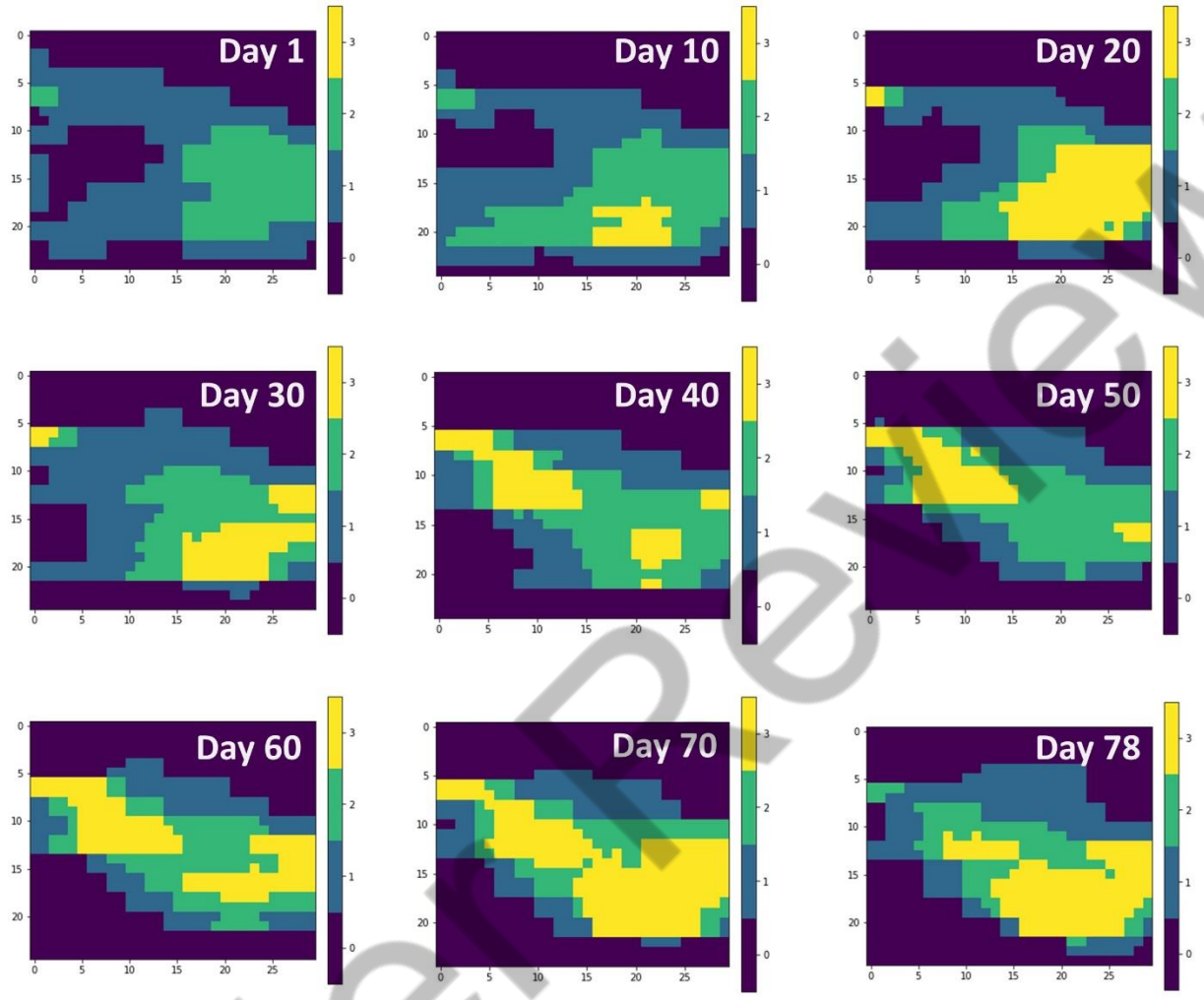


Figure B1. Spatiotemporally clustered ERT maps for days 1, 10, 20, 30, 40, 50, 60, 70, and 78. The spatiotemporal clusters are within a 25m-by-30m region between the F2-F3 monitoring wells. From days 1 to 30, a steady increase of high CO2 is observed and linked to the early development of the CO2 plume. From days 40 to 78, the shape of the CO2 plume drastically changes with a clear CO2 movement at all levels.