Automatic rock classification from core data to well logs: Using machine learning to accelerate potential CO₂ storage site characterization

Misael M. Morales*, Oriyomi Raheem, Carlos Torres-Verdín, and Michael J. Pyrcz, The University of Texas at Austin; Murray Christie and Vladimir Rabinovich, S&P Global

Summary

We develop a method for automatic rock classification (ARC) from core data to well-log scale using machine learning. Our ARC method estimates rock classes along a well to accelerate the petrophysical interpretation of well logs using core data for rapid formation evaluation and reservoir characterization. The ARC method integrates the estimated rock classes from four different unsupervised machine learning methods and compares against three different conventional physics-based methods. The benefit of this approach is its ability to rapidly estimate rock classes at the well-log scale from core data without the need for manual interpretation. We implement the new method with measurements from approximately 2400 wells in the Gulf of Mexico. Results indicate very good performance in comparison to traditional methods and can also be used to estimate the well-log scale and spatial distribution of rock classes at the basin scale to identify potential sites for CO₂ storage.

Introduction

Rock classification is a common practical approach to enhance reservoir description, evaluation, modeling, and simulation. Rock classification is also critically important to characterize potential conventional hydrocarbon reservoirs and CO₂ storage sites via identification of sweet spots based on permeable units bounded by impermeable seals (Bachu et al., 2007). The definition of rock type highly depends on the objective of the characterization. Geological rock typing is based on depositional environments or lithofacies, petrophysical rock typing aims to describe units with distinct grain and pore types, and reservoir and production rock typing aims to identify rock types as flow units. In a reservoir with negligible diagenesis, it is likely for all of these rock types to match (Ali-Nandalal and Gunter, 2003; Acosta et al., 2005). However, when dealing with highly post-depositionally altered reservoirs, it is common that geological facies, petrophysical rock types, and flow units significantly differ (Rushing et al., 2008; Xu and Torres-Verdin, 2013).

Rock type can be wholly defined as a group of rock bodies that has acceptable petrophysical regression within each group and can be spatially traceable in line with the geological framework (Neo et al., 1998). Rock classification based on routine and special core analysis has been extensively studied in the literature. Pittman (1992) implements a rock classification method based on Winland's R_{35} metric of pore throat radius distribution. Amaefule et al. (1993) implement the concept of hydraulic flow units and derive a flow zone indicator to characterize heterogeneous reservoirs. For rock classification at a reservoir or basin scale, Gunter et al. (1997) implement the Stratigraphic Modified Lorenz (SML) coefficient for characterizing reservoir flow units. Other methods for geological, petrophysical, and reservoir rock typing such as mercury injection capillary pressure (MICP) data, pore throat size distribution, and thin sections have also been studied (Al-Aruri et al., 1998; Neo et al., 1998; Clerke et al., 2008).

However, rock classification based on core measurements needs to be propagated to the uncored zones along the well, to other wells in the reservoir without core data, or to a basin scale by spatially correlating the interpreted wells. For the case of uncored wells, rock classification can be based on well logs or as a spatial distribution of cored wells. Once a classification of rock types is established along a well or spatially along a reservoir or basin, petrophysicists can infer rock properties for a complete characterization of the formation (Bennis and Torres-Verdin, 2019; Raheem et al., 2023).

We develop an automatic rock classification (ARC) method based on unsupervised machine learning methods to rapidly classify rock types from core measurements and interpolate to well log scale to detect potential CO₂ storage sites in the Gulf of Mexico. The method clusters the core data into rock types based on their flow and storage properties and interpolates to the well log scale for continuous rock classification. The ARC method serves as a quick tool for rock classification at the well log scale and can be applied to identify the spatial distribution of sweet spots at a basin scale. We implement our method on a field dataset with approximately 2400 cored wells in the Gulf of Mexico.

Method

The ARC method is designed as a tool to aid petrophysicists and reservoir engineers in the characterization and interpretation of well, field, and basin scale modeling. Using only core measurements of porosity and absolute permeability, the ARC method clusters the core measurements into distinct classes based on the properties of the core data and imputes the rock classes along the depth of

Automatic rock classification from core data to well logs using machine learning

the well. This allows for rapid core-to-log interpolation and potential multi-well formation evaluation and basin-scale modeling by tracing similar zones along multiple wells.

The three physics-based rock classification methods used to compare the results of our ARC method are Leverett's equivalent pore throat radius, Winland's R_{35}, and the Stratigraphic Modified Lorenz Coefficient. Leverett's equivalent pore throat radius method separates rock classes by marking cutoffs between different sizes of pore throats, r, where $r \cong \sqrt{k/\phi}$. This method assumes that each rock class possess the same pore throat radius despite variability in porosity or permeability, and the only distinction is an increase or decrease in number of pores. Winland's R_{35} technique assumes that pore throats are heterogeneous within a rock and can be described with a distribution function. The rock types are classified by selecting cutoffs between for the 35th percentile of the pore throat size distribution , assuming that the 35th percentile of the pore throat size distribution serves as the most meaningful metric for rock typing. However, Winland's method assumes unimodal and well-behaved distributions. The Stratigraphic Modified Lorenz (SML) coefficient quantifies the heterogeneity in a rock based on the cumulative flow ($\tilde{k}$) and cumulative storage capacity ($\tilde{\phi}$). SML cutoff values are selected by comparing the change in the magnitude of the slope, $d\tilde{k}/d\tilde{\phi}$.

The four unsupervised machine learning methods for rock classification implemented in our ARC method are K-means, bisecting K-means, Gaussian Mixture Models (GMM), and BIRCH (balanced iterative reducing and clustering using hierarchies). K-Means clustering aims to minimize inertia, or within-cluster sum of squares by separating samples into K groups of equal variances. Bisecting K-means is an iterative variant of K-means that applies a divisive hierarchical clustering strategy that iterates over batches of data. GMM implements the expectation-maximization algorithm to fit mixtures of Gaussian kernels to the data and estimate the probability of belonging to a class. BIRCH lossy compresses the data into a hierarchical tree that divides into subclusters. The subclusters are then iteratively collected into K clusters based on their within-cluster similarity and between-cluster dissimilarity.

Using the estimated rock classes from core measurements, the next step is to impute the rock classes from the core data to the well log scale by nearest neighbor interpolation. Let c_i represent the class value assigned to the ith data point, d_i represent the depth of the ith data point, $\boldsymbol{z}$ represent the depth index along the well log, and $\mathcal{C}$ represent the estimated rock class value from core data. Initialize $\boldsymbol{z} \equiv \underline{0}$. Then, for every depth along the well, z_i, we can estimate the corresponding rock class as follows:

$$c_i = \begin{cases} c_{i-1}, & if\ z_i = 0 \\ \mathcal{C}(\min|d_i - z_i|), & otherwise \end{cases}, \quad (1)$$

such that the well-scale rock classes have a well log resolution while honoring the inferred rock classes from the core measurements.

When using physics-based methods for rock classification, it is possible to obtain interpretable rock properties from the inferred classes to support any petrophysical interpretation. However, physics-based methods require prior petrophysical interpretation in order to select the optimal of the cutoff values to separate the rock classes. To overcome this limitation of physics-based methods, our ARC method uses unsupervised machine learning methods to estimate the rock classes from the core data and propagate the estimated classes to the well log scale.

Results

The first step in our ARC method is to process the core data. We have a total of 2442 wells with core measurements but select the wells with at least $N = 30$ core samples, where N is selected arbitrarily to obtain representative rock classes and correlate between wells. This filtering results in 1379 wells with at least 30 core samples of porosity and absolute permeability. Figure 1 shows the spatial distribution of the cored wells after filtering.
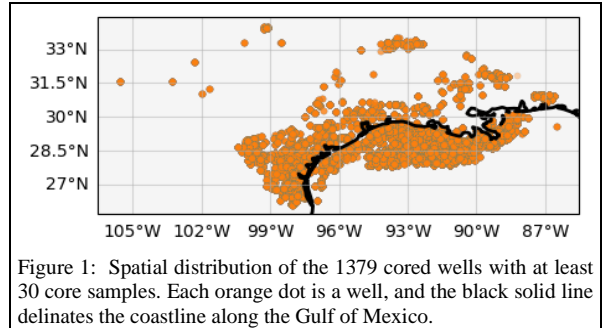

Figure 1: Spatial distribution of the 1379 cored wells with at least 30 core samples. Each orange dot is a well, and the black solid line delinates the coastline along the Gulf of Mexico.

We select a sample well arbitrarily and begin by comparing the estimated rock classes from core data for the physics-based methods and machine learning methods. We choose to classify the well data into rock types of low, medium, and high reservoir quality, for a total of $N_c = 3$ classes. Figures 2 and 3 show the crossplot of core porosity and absolute permeability, and the estimated rock classes using the physics-based methods and machine learning methods, respectively.

After rock classes are estimated from core data, we propagate the estimated rock classes to the log scale using Equation 1. The estimated rock classes at the well log scale for the sample well using the physics-based and machine learning methods are shown in Figures 4 and 5, respectively.

Automatic rock classification from core data to well logs using machine learning
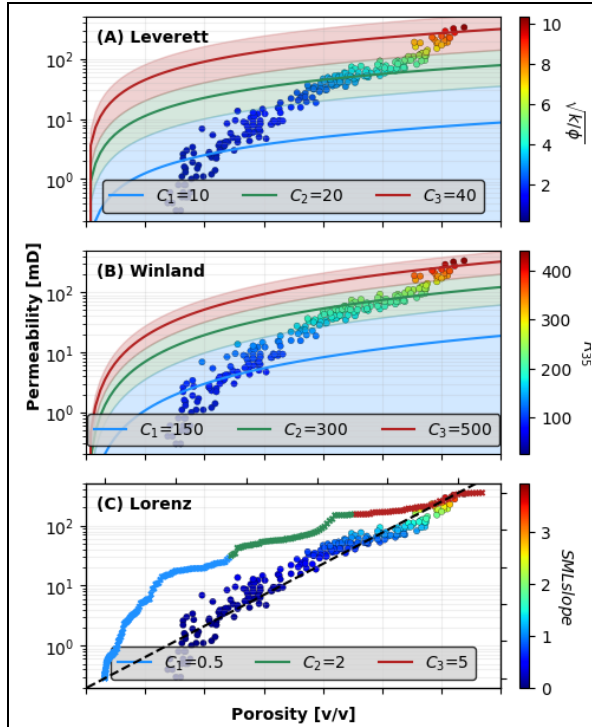

Figure 2: Crossplot of core porosity and permeability for the sample well, and the estimated rock classes using physics-based methods: (A) Leverett with cutoffs (10,20,40), (B) Winland with cutoffs (150,300,500), (C) SML with cutoffs (0.5,2,5). The colors represent low (blue), medium (green) and high (red) reservoir quality classes.

We observe good agreement between the rock classes based on the physics-based methods and the machine learning methods. Most discrepancies are located at the depths with sparse core sampling. This proves that our ARC method for rock classification based on unsupervised machine learning techniques is reliable and consistent with physics-based methods. The entire process of loading, processing, training, predicting, and saving the well-scale rock classes for all wells requires only 47.2 seconds, or approximately 0.03 seconds per well. Thus, proving our ARC method useful and efficient for rapid characterization to identify potential high quality reservoir rocks for CO_2 storage in terms of storage and flow capacity and formation thickness.

To estimate basin-scale potential CO_2 storage sites, The ARC method is implemented on all 1379 wells with at least 30 core measurements in the Gulf of Mexico. Using all four machine learning methods, we calculate the average rock class at each depth. We compute the proportion of sweet spots as the percentage of high-quality reservoir rocks in each well. Figure 6 shows the spatial distribution of sweet spots over the Gulf of Mexico using the average rock class from the machine learning methods.
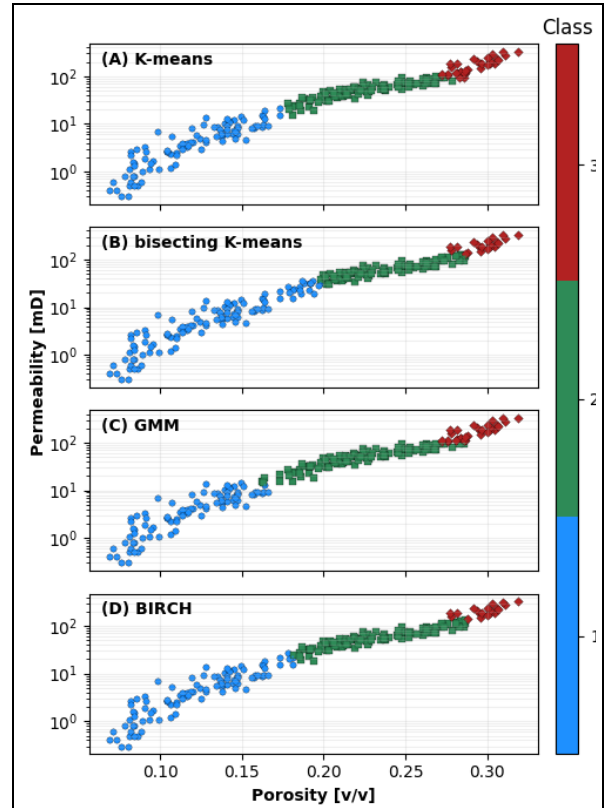

Figure 3: Crossplot of core porosity and permeability for the sample well, and the estimated rock classes using machine learning methods: (A) K-means, (B) bisecting K-means, (C) GMM, and (D) BIRCH. The colors represent low (blue), medium (green) and high (red) reservoir quality classes.

The selection of the method for rock classification, either physics-based or machine learning, is subjective based on expert knowledge. However, the ARC method provides comparatively reliable rock classes using unsupervised machine learning methods at a lower computational cost.

Further petrophysical interpretation is recommended, and the ARC method should only be used as a screening tool for rapid identification of sweet spots along the depth of a well and at a field or basin scale. Traditional physics-based methods are limited by the fact that the expert has to define cutoff values to separate rock classes, and these vary well-to-well. On the other hand, machine learning methods are automatic and computationally efficient but might violate assumptions of petrophysical properties and geologic continuity. Future work includes implementing a hybrid physics-informed machine learning framework to automatically estimate the cutoff values of the physics-based methods using machine learning techniques.

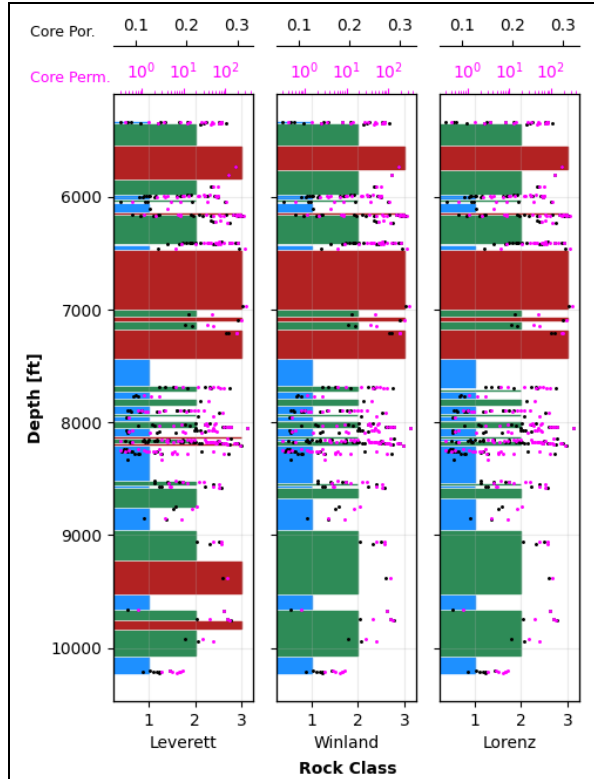# Automatic rock classification from core data to well logs using machine learning

Figure 4: Estimated rock classes at the well log scale using Leverett (left), Winland (center), and Lorenz (right) physics-based methods for the sample well. The colors correspond to the rock classes of low (blue), medium (green), and high (red) reservoir quality from the core classification. The core measurements of porosity (black) and absolute permeability (orange) are display as a function of depth along the well.

Conclusions

We developed a new method for automatic rock classification (ARC) based on core measurements scaled to the well log support. The method uses unsupervised machine learning techniques for automatic rock classification using core data from the Gulf of Mexico and compares the estimated classes to traditional physics-based techniques. Furthermore, the method allows for rapid estimation of rock classes along the well trajectory, and at a basin scale by correlating well-to-well rock classes. It can be implemented as a tool for rapid characterization of potential CO_2 storage sites, either vertically or spatially, by detecting sweet spots across high-quality reservoir rocks.
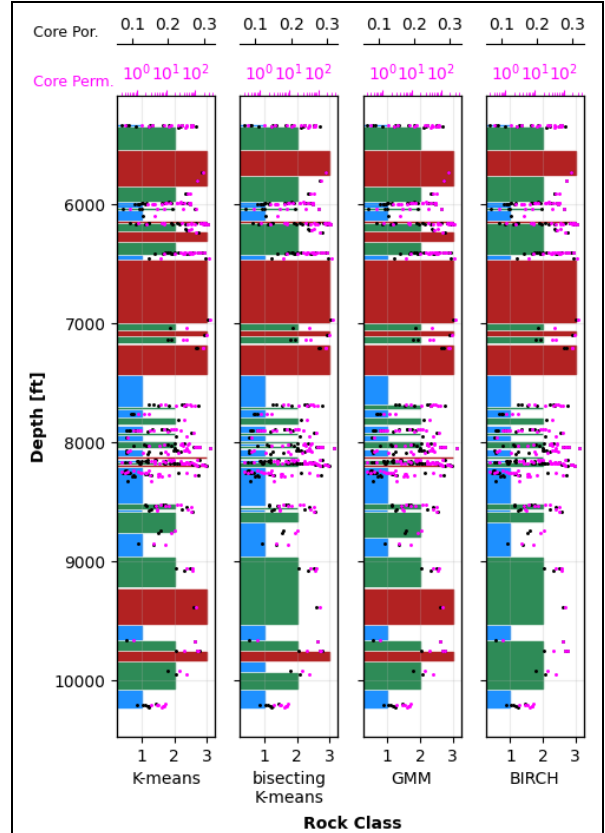
Acknowledgements

Figure 5: Estimated rock classes at the well log scale using K-means, nisecting K-means, GMM, and BIRCH machine learning methods for the sample well. The colors correspond to the rock classes of low (blue), medium (green), and high (red) reservoir quality from the core classification. The core measurements of porosity (black) and absolute permeability (orange) are display as a function of depth along the well.
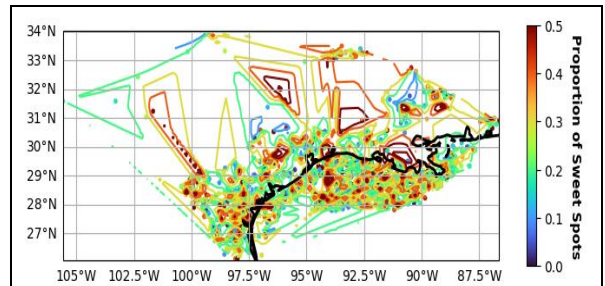

Figure 6: Spatial distribution of the proportion of sweet spots for each of the 1379 wells with at least 30 core samples in the Gulf of Mexico. The proportion of sweet spots is estimated using the mean rock class from all four machine learning methods.

Automatic rock classification from core data to well logs using machine learning

References

Acosta, L., E. Marin, E. Labastidas, J. Bello, J. Jimenez, P. Cordoba, J. C. Pascual, G. Auxiette, Y. Gou, and B. Thorsen, 2005, Reservoir Study V9 of El Furrial Field, Venezuela, *in* SPE Latin America and Caribbean Petroleum Engineering Conference: p. SPE–95047.

Al-Aruri, A., F. B. Ali, H. A. Ahmad, and S. A. Samad, 1998, Rock Type and Permeability Prediction from Mercury Injection Data: Application to a Heterogeneous Carbonate Oil Reservoir, Offshore Abu Dhabi (United Arab Emirates), *in* Abu Dhabi International Petroleum Exhibition and Conference: p. SPE–49556.

Ali-Nandalal, J., and G. Gunter, 2003, Characterising reservoir performance for the mahogany 20 gas sand based on petrophysical and rock typing methods, *in* SPE Latin American and Caribbean Petroleum Engineering Conference.

Amaefule, J. O., M. Altunbay, D. Tiab, D. G. Kersey, and D. K. Keelan, 1993, Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/wells, *in* SPE Annual Technical Conference and Exhibition.

Bachu, S., D. Bonijoly, J. Bradshaw, R. Burruss, S. Holloway, N. P. Christensen, and O. M. Mathiassen, 2007, CO2 storage capacity estimation: Methodology and gaps: International Journal of Greenhouse Gas Control, v. 1, no. 4, p. 430–443, doi:10.1016/S1750-5836(07)00086-2.

Bennis, M., and C. Torres-Verdin, 2019, Estimation of dynamic petrophysical properties from multiple well logs using machine learning and unsupervised rock classification, *in* SPWLA Annual Logging Symposium: p. D053S015R004.

Clerke, E. A., H. W. Mueller III, E. C. Phillips, R. Y. Eyvazzadeh, D. H. Jones, R. Ramamoorthy, and A. Srivastava, 2008, Application of Thomeer Hyperbolas to decode the pore systems, facies and reservoir properties of the Upper Jurassic Arab D Limestone, Ghawar field, Saudi Arabia: A "Rosetta Stone" approach: GeoArabia, v. 13, no. 4, p. 113–160.

Gunter, G. W., J. M. Finneran, D. J. Hartmann, and J. D. Miller, 1997, Early determination of reservoir flow units using an integrated petrophysical method, *in* SPE Annual Technical Conference and Exhibition? p. SPE–38679.

Neo, S., J. Asada, N. Fujita, S. Mohammed, and H. Arab, 1998, Geological framework modeling and rock type optimization for a giant oil field, *in* Abu Dhabi International Petroleum Exhibition and Conference.

Pittman, E. D., 1992, Relationship of porosity and permeability to various parameters derived from mercury injection-capillary pressure curves for sandstone: AAPG bulletin, v. 76, no. 2, p. 191–198.

Raheem, O., W. Pan, C. Torres-Verdin, and M. M. Morales, 2023, Best Practices in Automatic Permeability Estimation: Machine-Learning Methods vs. Conventional Petrophysical Models, *in* SPWLA Annual Logging Symposium: p. D041S015R001.

Rushing, J. A., K. E. Newsham, and T. A. Blasingame, 2008, Rock typing—Keys to understanding productivity in tight gas sands, *in* SPE Unconventional Resources Conference/Gas Technology Symposium: p. SPE–114164.

Xu, C., and C. Torres-Verdin, 2013, Core-based petrophysical rock classification by quantifying pore-system orthogonality with a bimodal Gaussian density function, *in* Paper SCA2013-079 presented at International Symposium of Society of Core Analysts. Napa Valley, California, September: p. 16–19.