Identification of reservoir flow zone and permeability estimation: Review paper

Alaa S. Al-Rikaby*, Mohammed S. Al-Jawad

University of Baghdad, Department of Petroleum Engineering, Iraq

Abstract

Reservoir flow zonation is essential in evaluating heterogeneity reservoirs, constructing static/dynamic models, and reserve calculations in any reservoir characteristic. Reservoir flow zonation is accomplished through techniques and methods based on core analyses, well logging, or both, and these techniques and methods deal with statistical operations and mathematical equations. The flow zonation definition starts with assigning depositional environment and diagenetic processes of rock types by thin sections of petrographic study taken from cutting and core plugs. In the second step, electrofacies are determined based on well-log classification response using the clustering algorithm method. The well-logging used in this step included Porosity logs (neutron porosity, sonic log, and bulk density), deep resistivity (RT), nuclear magnetic resonance, and gamma-ray logs. The third step is due to determining the flow zone and the analysis of pores size distributions. The flow zone indicator value is calculated from routine core data analyses and depends on the hydraulic flow unit concept. The linking process between the above steps with seismic attributes gives certainty prediction to reservoir behavior and a more reliable model. This study reviews many approaches (deterministic and stochastic) for determining flow zonation and for different reservoir types, including conventional and unconventional, and also proposes the best approach for each by presenting the advantages and disadvantages, especially in the methods that use statistical operations as a nondeterministic method. Permeability estimation for uncored wells is also included in the study's agenda through a simplified presentation of the techniques and methods used in this scope. This study demonstrated that the K-mean approach is the most dependable and consistent method for defining reservoir electrofacies compared with the other techniques investigated of machine learning, despite some drawbacks in running this approach. In addition, Hierarchical Clustering has the problem of not being suited for bigger datasets. Moreover, it only yields the greatest outcomes in certain instances. Automated reservoir zonation has been introduced as a novel approach by previously mentioned researchers using clustering algorithm methods like (Elbow and PELT).

Keywords: Carbonate reservoirs, Flow zonation, Flow zone index (FZI), Hydraulic flow units (HFU), Sand reservoir, Tight gas reservoir

1. Introduction

Flow zonation is recognized as the most effective method for characterizing heterogeneity reservoirs. The classification techniques focus on many characteristics, including multi-scale and multi-modal pore type and size, rock textures, diagenetic changes, and static-dynamic data integrations. Flow zonation was performed utilizing many approaches according to core analyses, well logging, or both together, as seen in Fig. 1; some of which are deterministic, such as (FZI and SEM), and others are stochastic, such as (object-based, pixel-based, process-based, and other clustering analysis methods), where the heterogeneous geological environment is the only basis for diverse stochastic techniques for facies models. Amaefule introduced the flow zone indicator concept (FZI) as the first principle based on the Carman-Kozeny model to define the hydraulic flow unit (HFU). Some researchers defined flow unit zonation as a reservoir mappable portion within which geologic and petrophysical
characteristics like porosity, permeability, and bedding features that influence fluid flow are more consistent and various from other rock volume properties. The flow unit’s objectives are as follows:

1. Estimate the distribution of initial saturation in a hydrocarbon reservoir using rock-type-based capillary pressure/J-function curves.
2. Identify fluid contacts according to the distribution of rock types in the hydrocarbon reservoirs.
3. Assign rock-type-based relative permeability in the reservoir model to capture the representative fluid flow regime.

In most cases, estimating rock permeability is required for a net-pay production evaluation. The ability to accurately predict reservoir rock permeability through open hole logging in an uncored interval is critical. There is no exclusive routine open-hole logging tool that calculates permeability immediately. Therefore, the existing state of knowledge depends on many empirical relationships using either resistivity or porosity data. There are empirical methods based on the correlations among permeability, porosity, resistivity, and irreducible water saturation. Some researchers have presented empirical methods that calculate permeability by open-hole logging measurement. A nuclear magnetic resonance (NMR) log has been used in the construction of a permeability model according to the connection between the distribution of the transversal relaxation time (T2) and the distribution of pores size, especially in sandstone.

Nomenclature

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT</td>
<td>Tiab flow unit</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning electron microscopy</td>
</tr>
<tr>
<td>SMLP</td>
<td>Stratigraphic modified Lorenz plot</td>
</tr>
</tbody>
</table>

Fig. 1. Classification of reservoir flow zonation approaches.
(clastic rock). However, in tight rock and heterogeneous sand, one becomes sensitive to various rock lithofacies.12,13

The general rule for increasing the permeability by increasing porosity is violated, particularly in fractured and low porosity reservoirs at bad sorting.14 Porosity relies on the number and size of pore bodies. Permeability, however, relies upon the pore throat number cross with the pore body or the contacts per grain. Distributions of coordination numbers have a crucial impact on the topology of the pore networks and the rock flow characterizations.15 The permeability of the reservoir is influenced by the depositional environment (mineral composition and rock texture) and diagenetic processes (cementation, leaching, compaction cycles, and fracturing). The parameters that influence the rock texture include the shape and size of the grains, sorting, packing, and cementation. Grain sizes, sorting, and cementation significantly impact porous media's permeability; conversely, grain size does not influence porosity.16

This study examines many approaches (deterministic and stochastic) for determining flow zonation and for different reservoir types, which include conventional and unconventional reservoirs. It proposes the most effective approach for each by presenting the benefits and drawbacks, particularly in the methods that use statistical operations as a non-deterministic method. The study's agenda also includes permeability estimation for uncored wells via a simplified description of the methodologies and methods employed in this area.

2. Flow zonation

2.1. Concepts and methods

Flow zonation (rock type) in hydrocarbon-bearing formations, particularly heterogeneous formations, is challenging because there needs to be a comprehensive process for reservoir flow zonation. Using well logging, routine core analyses, and special core analyses (SCAL) data which considers the most efficient deterministic methods to represent the storage and flow capacity of a reservoir with observance of the rock types of numbers in hydrocarbon reservoirs and the degree of complexity of each reservoir, as there are quite a few rocks that are difficult to predict in subsequent steps.

Usually, many microfacies have been defined by sedimentologists based on microscopic studies for thin sections like scanning electron microscopy (SEM), especially for carbonate reservoirs, through examination of the features seen under the microscope and deriving some attributes to define microfacies, while flow characteristics as shown by the dynamic data are not very much controlled by this type. For example, two types of rocks may have different microfacies: fossiliferous mudstone and nonfossiliferous mudstone. However, they are not reservoir facies since they have zero petrophysical properties (porosity, permeability). They are classified into the same types of rock groups based on the porosity of the rock and its permeability. However, they belong to different rock types based on sedimentology.

Alternatively, porosity and permeability can be improved due to dissolution and fracture in non-reservoir mudstone to reach the level of grainstone. At the same time, the last can be cemented, and its petrophysical properties are reduced to those of mudstone. That is why we must integrate microfacies with diagenetic processes and similar depositional environments until rational numbers of rock types control flow and storage capacity. As a result, thin section analyses, petrophysical logging, routine core analyses, and SCAL data, together with other reservoir data (such as well testing and seismic attributes), will assist in the identification of flow zonation in a reservoir. The presence of heterogeneity in the depositional environment is acknowledged to be the single justification for various stochastic approaches for the facies model, which are subclassified into process-based, pixel-based, object-based, and clustering analysis models.2,17,18

2.2. Approaches statement based on core analyses

2.2.1. Scanning electron microscopy (SEM) technique

SEM is a deterministic technique that excites the surface electrons of a specimen by passing an electron beam across it. The system then detects changes in electrons and generates a picture. The SEM can observe specimens with a magnification of up to 50 000x.19 SEM is a crucial tool for geological investigation, and its data is distinct from other tools.

(1) SEM is a more involved technique than radiography diffraction (XRD) scanning, which may provide a phase analysis in less than an hour. Performing phase analysis with an SEM requires capturing spectra for each phase separately. A competent operator can swiftly identify minerals using grey scales to distinguish distinct phases and elemental analysis to determine their identities. XRD, however, utilizes software to do a semi-quantitative phase study on a sample more rapidly and simply than the SEM, even though both methods might arrive at identical findings.
Unlike the XRD, the SEM may also be used to infer mineral relationships.

(2) Unlike an inductively coupled plasma, which provides elemental analysis for the whole rock, a SEM may perform elemental analysis for each phase at a user-defined spot as tiny as 3 μm.

(3) The SEM has a more significant depth of focus and higher magnification than an optical microscope. Since an electron beam has a shorter wavelength than visible light, it can see things with more excellent resolution and, thus, magnification. The drawbacks of SEM over optical include complicated sample preparation, expensive operating and maintenance expenses, and the need to dry materials. Due to the SEM’s high vacuum, live samples cannot be seen similarly under an optical microscope. The SEM can only see non-living biological samples and coated with gold.

In analytical methods, SEM offers a wide variety of applications, including:

(1) Exploration and mining firms concerned with minerals identifications.
(2) Determine zonation in crystals.
(3) Particle size and modals analyses of minerals in the sample.
(4) Research on the preservation of fossils in sedimentary rocks.
(5) Examine the microscopic characteristics of the sample.

2.2.2. Testerman statistical zonation method

This approach, proposed by Testerman, does not require any estimation of the number of zones; also, the number of zone borders is regulated automatically and by a predefined termination condition. Testerman approach uses only core permeability measurements to determine zones. Testerman implemented it in two phases. These measures are separate flow unit identification and assessing flow unit continuity in nearby wells. The first phase separates each well into distinct zones or flow units. These zones have been set such that the variation of values inside each zone is minimal, but the variability between zones is maximized. This approach uses the zonation coefficient ‘R’ as a zonation criterion. Equations (1)–(3) used for each well’s zonation:

\[ W = \frac{1}{N-1} \left[ \sum_{i=1}^{L} \sum_{j=1}^{m_i} (k_{ij} - k_i)^2 \right] \]  
(2)

\[ R = \frac{B - W}{B} \]  
(3)

B: Zones variance.
L: Zones number.
W: Variance in each zone
m_i: Permeability data number in ith zone
j: Index for algebraically summation of data of each zone
i: Index for algebraically summation of zones.
N: Number of total permeability data of reservoir
k_ij: Permeability data of network (mD)
R: Zonation Coefficient range between zero (heterogeneous) & one (homogeneous)

k_i:: Summation of average permeability data in well (mD)
k_i: Average permeability data in ith zone (mD)
Zonation coefficient (R) reveals how uniformly this zonation splits the zones. The closer this coefficient is to one, the more homogenous the zones are. Therefore, the highest zonation coefficient yields the optimal two-zone subdivision of a well. Then, these stages are repeated independently for two distinct zones. The zones are divided until the providing zonation coefficient (R) in the subsequent stage is smaller than the preceding one. The second portion of the computations begins when the flow units for each well are identified. This part connects the flow units in the reservoir region from one well to another to determine flow unit continuity. The calculations are based on a statistical difference between average data in nearby well zones and the difference predicted from measuring variance values in zones, as seen in equation (4).

Tables 1–3 explain Testerman Statistical Zonation Method calculation.

\[(k_{ih} - k_{ij}) > \sqrt{0.5 \left( \frac{1}{n_h} + \frac{1}{m_i} \right) s_z(v_p)} \]  
(4)

k_h, k_i: Arithmetic permeability average for hth and ith zones in a well
n_i and n_h: Data number for hth and ith zones
s: Standard deviation for total permeability data of reservoir
z: Tabulated constant as a function of data, number of zones and probability level
v_p: Utilized to identify z as a function of probability level.
2.2.3. Rock-fabric

Lucia and Jennings\textsuperscript{21} provided an empirical relationship for carbonates as the following and shown in Fig. 2:

\begin{equation}
\log(K) = 0.4342a_0 - a_1 \log(\lambda) + (b_0 - 2.303b_1(\lambda))\log(\Phi) \tag{5}
\end{equation}

Where \(a_0 = 22.56\), \(a_1 = 12.08\), \(b_0 = 8.671\), \(b_1 = 3.603\), and \(\lambda\) is the number of rock-fabric.

Although interparticle porosity and permeability are petrophysical characteristics, rock fabric is a geological description of the pore space with petrophysical consequences. Rock fabric has a petrophysical relevance because of its interaction with interparticle porosity and permeability, as well as geological significance due to its predictability from stratigraphic geological models. Thus, rock fabric bridges the gap between petrophysics and geology, allowing for more accurate reservoir modeling.\textsuperscript{21}

2.2.4. Winland cross plot

The Winland cross-plot connection between permeability, porosity and pore throat was provided based on the measurement of mercury injection. This technique allows comparing facies-based rock types, lithology-based rock types, with petrophysical-based rock types by a generation of k/phi ratio line.\textsuperscript{3} Winland with Pittman\textsuperscript{22} published the following empirical relationship, as seen in Fig. 3, between porosity, air permeability, and pores throat radius corresponding to a mercury saturation of x%:

\begin{equation}
\log(K_{air}) = \log(a) + b(\log(\Phi)) + c\log(R_x) \tag{6}
\end{equation}

Where: (Rx) represents a pore-throat radius at x % mercury saturation (nonwetting fluid), reflecting the diageneric and depositional fabric.
2.2.5. Modified lorenz plot (MLP)

It is generated by using flow unit interpretations results, then ranking the flow capacities and replot the results, which allows identifying the petrophysical rock types of numbers for all wells. \( K = \frac{1}{2\tau^2S_{gr}^2} \times \frac{\varphi^3}{(1-\varphi)^2} \) (7)

\[ r_{hm} = \frac{r}{2} = \frac{\text{cross section area}}{\text{wetted surface area}} \] (8)

Tiab and Donaldson\(^{31}\) suggest that the term \( 2\tau^2 \) which referred to as Kozeny constant, is reality ‘variable constant’ and may vary with different hydraulic flow unit but stay constant for a specified unit. As a result, they introduced ‘variable constant’ \( K\tau \) as effective zoning factor:

\[ K = \left( \frac{1}{k\tau S_{gr}^2} \right) \times \left( \frac{\varphi^3}{(1-\varphi)^2} \right) \] (9)

\[ k\tau = F_s \times \tau^2 \] (10)

Where: \( S_{gr}^2 \) represents grain surface area/volume for solid material, \( F_s \) represent Effective pores throats shapes factor, \( \tau^2 \) represent Tortuosity.

2.3. Approaches statement based on well logging

2.3.1. Nuclear magnetic resonance (NMR)

NMR logging plays an important role in the characterization of hydrocarbon reservoirs. The main products of NMR services are partial and total porosities estimation (bulk movable fluid, clay bound water, and irreducible water), the permeability index prediction, the saturations estimation (e.g., \( S_w, S_o, S_{irr} \)), type of hydrocarbon and estimation of hydrocarbon viscosity with no environmental correction at log process. \(^{32,33}\) One of the magnificent aspects of NMR logging lies in its capability for measuring fluid saturated porosity and for distinguishing between nonproducible and producible fluid. Thus, the permeability index can be calculated using many models, such as the Coates-Timur model. \(^{34,61}\)

\[ K = \left( \frac{\varphi}{C} \right)^m \times \left( \frac{BVM}{BVI} \right)^n \] (11)

Where: \( c, m, \) and \( n \) are constants generated from core sample experimental data sets as seen in Figs. 4 and 5, (irreducible water volume, BVI) is irreducible water volume, (movable fluid volume, BVM) movable fluid volume, and (effective porosity, PHE) is effective porosity, \( PHE = BVI + BVM \).

Fig. 3. The relationship between porosity, permeability and pore-throat radius based on Winland’s equation. \(^{27}\)

Fig. 4. C determined in Coates-Timur model with \( m = 3.91, n = 0.51 \) as design, where \( C \) equal to 0.222, with coefficient of 0.948. \(^{35}\)
2.3.2. Logging while drilling formation pressure tester
This technology is utilized a novel technique and several advantages technologies that aid a better understanding of the reservoir characteristics that include an intelligent function to reduce formation shock at pressure tests in relatively low permeability and avoid the sand issue in extremely unconsolidated formations. When optimal test sequences are carried out, the precision of the pressures and mobility results is improved, and the maximum test duration may be increased to 30 min. The LWD formation pressure tester lets the user choose the optimum drawdown pressures pumping and then selects the drawdown limitation as a percentage of the maximum drawdown in four levels. This technique allows the user to restrict the pressure drawn down from the formation. The LWD formation pressure tester can apply a variable drawdown rate. At the same time, the test is being performed, and it can also be used to create three drawdowns and buildups with incredible repeatability and accurate mobility values. Note that this technology is widely used in the Middle East and is implemented by several service companies, such as Baker Hughes.

2.3.3. Seismic based methods
The seismic sequence interpretations were executed by picking fault, seismic to well-tie, and interpretations of the horizon to the top of the reservoir to generate the isochore map. This seismic data was not evident initially and was improved utilizing the next attributes: frequency filters, median filters, trace control, and difference to identify the fault. In seismic treatment, the majority of issues are often amplitude-linked. High amplitude on a display panel or map makes it more complex to view fine details. The data must be subjected to automated gain control to remedy the issues caused by large amplitudes to calculate the average value of the amplitude and a scale factor to normalize it to a set value. The variance feature was used to increase the visibility of the fault in the research region. The first step in interpreting seismic data is to emphasize a relationship between seismic reflection and stratigraphy resulting from wells data, considering the processing of high amplitudes that cause difficulty in seeing details. Well-to-seismic coupling was performed to align the reservoir’s top with the wells. The time-depth relationship was derived using check shot data for all wells. The following equation describes the frequency description of two wavelet coefficients of the same polarization separated by a time interval (\(\Delta t\)):

\[
f_n = \frac{1}{2\Delta t}
\]

Where: \(f_n\) is number of frequency and \(\Delta t\) is time interval.

2.4. Approaches statement based on core analyses and well logging
2.4.1. Hydraulic flow unit (HFU) concept
The flow zonation can be identified by using FZI and Reservoir Quality Index (RQI) calculation due to the hydraulic flow unit concept. Hydraulic flow unit determination is related to geophysical characteristics identification of the rock, like mineralogy and pore geometry. These characteristics control the reservoir’s fluid flow based on the pore throat’s
connection type. The relationship between RQI, FZI, and normalized porosity ($ø_z$) is given by Amaefule $^{3,25,37}$:

$$\varnothing_z = \frac{\varnothing}{C_0} - 1$$  \hspace{1cm} (13)

$$RQI = 0.0314 \sqrt{\frac{K}{\varnothing}}$$  \hspace{1cm} (14)

$$FZI = \frac{RQI}{\varnothing_z}$$  \hspace{1cm} (15)

$$\log RQI = \log FZI + \log \varnothing_z$$  \hspace{1cm} (16)

$$RPI = \frac{RQIrank + FZIrank}{2}$$  \hspace{1cm} (17)

Figs. 6 and 7 explain the calculations of (RQI), (FZI), and normalized porosity ($\varnothing_z$).

After identifying the HU, FZI is distributed as a continuous parameter throughout the wells, and the permeability distributions are computed based on this parameter:

$$k = 1014FZI^2 \left( \frac{\varnothing e^3}{(1 - \varnothing)^2} \right)$$  \hspace{1cm} (18)

In tight gas reservoirs, FZI is imperfectly correlated with the conventional log. So, FZI estimation may be poor from the conventional log because of substantial heterogeneity in these reservoirs. To obtain sequential FZIs for routine use, a new technique has been suggested called Schlumberger Doll Research, which is used for estimating permeability values from NMR log. $^{40}$

$$K = C_1 \cdot \varnothing m_1 \cdot T_{n2lm}$$  \hspace{1cm} (19)

Where: $T_{2lm}$ represents the logarithmic means for NMR T2 distribution in milliseconds, while $m_1$, $n_1$, and $C_1$ are parameters of the statistical model which can be derived from experimental data of the core samples. Assuming there aren’t enough core samples available, these parameters have empirical values of 4, 2, and 10, respectively.

$$\sqrt{\frac{K}{\varnothing}} = \sqrt{C_1 \cdot \varnothing^{m_1-l_1} T_{2lm}^{n_1}}$$  \hspace{1cm} (20)

$$FZI = \frac{0.0314 \sqrt{C_1 \cdot \varnothing^{m_1-l_1} T_{2lm}^{n_1}}}{\frac{\varnothing}{C_0} - 1}$$  \hspace{1cm} (21)

Sneider and King demonstrated that most petrophysical characteristics of sandstone and conglomerate are correlated with grain sizes and grain sorting, rate of rock consolidation, cementation...
index, pore size, and pore connection. In addition, they demonstrated a limited set of rock types and pore geometry that define a geologic unit. Geologic and hydraulic flow units may not match. Additionally, a geological unit may include several flow units. The Tiab (HT) flow units factor includes all geological and petrophysical characteristics listed by Sneider and King.41,42 Observe that FZI and HT are connected by the equation below:

$$HT = \frac{1}{FZI^2}$$  \hspace{1cm} (22)

$$HT = \frac{1}{K} \left( 1 - \phi \right)^2$$  \hspace{1cm} (23)

$$HT = K_{ps} \phi^1 m^2 \left( \frac{K_{ps}}{dgr} \right)^2$$  \hspace{1cm} (24)

$$K = \frac{1}{K_{ps} \phi^1 m^2 \left( \frac{K_{ps}}{dgr} \right)^2 \left( 1 - \phi \right)^2}$$  \hspace{1cm} (25)

Where: kps is permeability per pore size, kgs is permeability per grain size, and dgr is diameter/grain size.

Coates and Denoo42,43 established a relationship between permeability and free fluid index (FFI), described as porosity and hydrocarbon saturations product. It is a unit of measurement for moveable liquids, such as oil, water, and gas, and is consequently related to the flow unit.

$$K = 10^{C_{cd} \phi^4 \left( \frac{FFI}{\phi - FFI} \right)^2}$$  \hspace{1cm} (26)

$$FFI = \phi(1 - S_{wr})$$  \hspace{1cm} (27)

Where $C_{cd}$ is a correlation constant that different from a reservoir to another.

$$RQI = 3.14 \left( \frac{FFI}{\phi - FFI} \right) \sqrt{\phi^3}$$  \hspace{1cm} (28)

Take logarithm for two sides:

$$\log RQI = \log \sqrt{\phi^3} + \log I_F$$  \hspace{1cm} (29)

$$I_F = \left( \frac{FFI}{\phi - FFI} \right)$$  \hspace{1cm} (30)

Where $I_F$ is defined as flow index.

Fig. 8 Depicts three values of IF because of Reservoir Quality Index and $\sqrt{\phi^3}$ relationship. Where $C_{pp}$ is porosity-permeability correlation constant.

2.4.2. Clustering analysis

A cluster mentions a data points group that collects with each other because of certain similarities. Cluster analysis is an unsupervised classification approach to describe a set of observations that participate in common characteristics into significative subgroups, i.e., the observations are divided into distinct and uniform groups named clusters.45 There are various types of clusters, as seen in Fig. 9, including prototype-based, well-separated, shared property, graph-based, and density-based. In contrast, the significant clustering analysis involves a few techniques like K-mean, Hierarchical Clustering, Calinski Harabasz, and the technique of DBSCAN.46

2.5. K-mean (Partitioning)

The k-mean technique is one of the popular unsupervised methods of learning algorithms utilized to analyze the known clustering issue. K-means, as a simple technique, to classify recognized data to a specified number of clusters or groups by defining k
numbers of centroids, assigning data points to nearest clusters, and preserving the smallness of centroids. Defining K centroid for all clusters is the basis of the K-mean technique. The centroids placement must be exact as the various placement guides to various outputs. Once the ideal K number of observations has been identified as the center of clusters, the second stage involves computing the distance between additional observations associated with each k cluster and allocating these dissimilar observations to the closest cluster. The next step is determining the new K centroid, which will be designated as a potential location for the clusters’ centers. As this process is repeated until constant cluster centroids are reached, the points of the data sets must be connected to the new centroid. The objective function that reflects the squared error function is what the K-means algorithm seeks to minimize. The objective functions shown below is:

\[ J = \sum_{j=1}^{K} \sum_{i=1}^{n} \| x_i^j - c_j \|^2 \]  (34)

Where:
- \( J \): Objective function
- \( K \): Clusters number
- \( n \): cases number
- \( x \): case
- \( c \): centroid

Some preview studies, such as, illustrated that the appropriate number of clusters was determined by applying Silhouette analysis k-medoids (PAM) technique, as illustrated in Fig. 10. Following the determination of the number of clusters, the K-mean technique was used to split the detected clusters.

Fig. 9. Different clustering methods similar sets of data.

The electrofacies produced by K-mean represent the true vertical lithofacies distribution found by core description as seen in Fig. 11.

The popularity of K-mean clustering may be related to numerous factors. First, it is theoretically straightforward to implement. Virtually every software for data mining has implementations of it. Second, it is adaptable, meaning that almost every part of the algorithms (distance function, initialization, termination criteria, etc.) is modifiable. Third, its storage complexity is proportional to N, D, and K. In addition, disk-based variations do not need to store all points in memory. Fourth, the rate of convergence is assured to be quadratic. Fifth, it is independent of data ordering, i.e., random shuffling of the data points.

In contrast, k-means has some limitations. First, the cluster number, K, must be set beforehand. This parameter’s value may be calculated automatically using a variety of cluster validity metrics. Second, it can only identify hyperspherical cluster which are compact and well-separated. It uses a broader distance function. Third, due to its use of the squared Euclidean distances, it is sensitive to outlier and noise points since even a small number of such points may substantially affect the mean of their respective cluster. It may be remedied using outlier pruning or a more rigorous distance function. Fourth, due to the gradient descent aspect of the algorithm, it frequently converges to a local minimum of the criteria function. For the same reasons, it is sensitive to the original center pick. Inappropriate initialization results in empty clusters, slower convergence, and an increased likelihood of being
trapped in an undesirable local minimum. Except with the first issue, all these disadvantages may be eliminated by applying an adjusted initializations strategy (IM).

2.6. Calinski-Harabasz

Calinski-Harabasz's technique of clustering analysis employs an index representing 'a ratio of the
between-cluster means to the within-cluster sum of squares. The index of Calinski-Harabasz is defined mathematically.

\[
CH = \frac{B(K)/(K-1)}{W(K)/(n-K)}
\]  

(35)

Where:

- K: Clusters number.
- B(k) and W(k): between and within cluster sums of square partitions.

Calinski-Harabasz's technique has advantages. First, the scores are more excellent if the cluster were dense and well-separated, which corresponds with the conventional definition of a cluster. Second, the scores are quickly calculated. In other hand, it has some drawbacks that it is often greater for convex clusters than alternative cluster ideas, such as density-based clusters generated by DBSCAN.

For example, on Calinski-Harabasz technique of clustering analysis, Figs. 12 and 13 illustrate three identified facies by this technique.

2.7. Hierarchical clustering

Hierarchical clustering is a common clustering strategy in machine learning that is relatively old compared with the other techniques but is still frequently employed in various areas. Agglomerative and divisive techniques are the two most common methodologies for creating hierarchical groupings. The agglomerative technique treats the points as independent clusters and, at each stage, merges the nearest pair of clusters until a single meaningful group containing all the related observations is created. This method necessitates the development of a concept of cluster closeness. The dividing technique, on the other hand, begins with a predetermined number of groups. The next stage is to allocate observations to these groups, which results in each observation being assigned to the closest group. Because Ward is really an agglomerative hierarchical approach, each data point is treated as a distinct cluster. Similar clusters merge collectively at each iteration until one or more clusters are established. Hierarchical clustering produces a dendrogram that depicts the clusters' hierarchical connections, as seen in Fig. 14. Ward's approach estimates the similarity of two clusters and merges them by calculating the total square of their distances. The same objective is used in this technique as in K-means clustering. As a result, this characteristic distinguishes Ward's method as distinct from other hierarchical techniques. This may be expressed numerically as follows:

\[
\Delta(C_1,C_2) = \frac{n_{C1}n_{C2}}{n_{C1} + n_{C2}} \left[ \frac{|MC_1|}{MC_1} - \frac{|MC_2|}{MC_2} \right]^2
\]

(36)

Where:

- (C1, C2): the cost of merging the clusters C1 and C2.
- n: The points number in the cluster.
- M: Cluster's center.
The Dendrogram using the hierarchical clustering technique, like the K-mean approach, reveals various clusters, as seen in Fig. 15. These various clusters correspond to certain electrofacies congruent with the lithofacies revealed in the core analysis. The numerical results revealed that the hierarchical clustering method misclassified a few data points, whereas the K-mean strategy produced more accurate classifications.

There are several advantages of this technique. First, it is simple to implement. Second, it is capable of producing an order of things that may be beneficial for the presentation. Third, the number of clusters should be specified beforehand. Finally, smaller clusters will be formed, which may reveal data similarities. In other words, it has some drawbacks. First, in just a few instances, does the agglomerative approach provide the best results? Second, the system can never reverse a prior action; hence, a similar outcome should be guaranteed if the items were erroneously categorized at an earlier step. Third, using different distance measures to measure the distances between clusters may provide varying results. Therefore, conducting many experiments and comparing the findings is essential to improve the real results’ reliability.

2.7.1. Automated approach to reservoir zonation
Amaefule’s and Gunter’s techniques are commonly utilized for reservoirs zonation of...
hydrocarbons rocks. Nonetheless, several prevalent uncertainties influence the efficiency of these techniques, including:

1. Estimation of the FZI cluster number based on previous geologic data,
2. Estimation of the cluster number of R35 value (port sizes equal to or littler than the pore throat entered when a rock is saturated 35% a non-wetting phase), and

The involvement of humans in these phases renders the acquired findings expert dependent; that is, the results calculated by many professionals working on the same data may vary significantly. This paper presents a simple but broadly applicable mathematical technique to automate the reservoir’s zonation process and reduce the bias caused by human decision-making roles. The rock type and lithofacies were described according to Dunham’s classifications with modifications by Lokier and Al Junaibi. RQI, FZI, and HFU were calculated using Amaefule’s and Gunter’s techniques. Fig. 16 is a schematic representation of the workflow mechanism used in this mathematically based method for reservoir zonation. For simplicity, the elbow technique was utilized to determine the optimal number of clusters. The elbow technique evaluates the variance percentage represented as a function of the number of clusters; the number of clusters should be set such that inserting another cluster will not result in a more accurate data model.

Specifically, suppose the dependency of the percentages of variances illustrated by clusters has been schemed versus the number of clusters. In that case, the initial clusters will give a great deal of information, but the marginal gain will decrease at a certain point, giving the graph an incline. During this phase, the number of clusters is chosen. Also known as the F-test, the percentage of the variability described is really the ratio of the variation across groups to the overall variance. The marginal variation percentage chosen for our purposes is 0.95, corresponding to the crucial α-value of 0.05. Upon completion of clustering with the optimal number of clusters, the last step to identifying the HFUs is to locate the linear segment on the SMLP graph with an equal slope. To use it, one must be able to identify sudden changes in the slope of the SMLP graph’s point. Fig. 17 shows how the elbow technique determines the optimal number of clusters for FZI values.

The optimal cluster number is seven since the variance ratio surpasses the 0.95 thresholds beyond this amount. It indicates that other clustering results in groups whose variance ratio does not vary considerably from a statistical standpoint, resulting in overfitting.

As seen in Fig. 18, nearly 45% of FZI values were within the range of 0.37–1.65, whereas fewer than 1% fell within the scope of 150–400. A Histogram of RQI data reveals that over 95% of RQI scores are less than 1, indicating that rock type with poor reservoir quality predominates in the region under study.
Based on the SMLP, the last automated reservoir zonation process is identifying the HFUs. To do this, it is first necessary to identify the points with similar slopes and then detect the optimal number of change points to generate the lines that best match the sections of points with identical slopes. Fig. 19 depicts the variability ratio of the lines fitting to the points for each segment, computed using the PELT algorithm for various change point numbers.

This illustration illustrates the hybrid application of the elbow technique and the PELT algorithm. It demonstrates that, from a statistical standpoint, increasing the number of change points above nine...
does not appreciably improve the matched lines' accuracy.

2.8. Permeability estimation for uncored wells

The permeability of formation rocks is an essential flow property for subsurface production and injection. The number of procedures (well-logging assessment, core measurement, and well testing) generally utilized to estimate it reflects its significance. Understanding the right interrelationships between the different approaches enables significant permeability comparisons and correlations. The investigation of the correlations between permeability and petrophysical data indicated a substantial association between the permeability of the reservoir and well logging, including gamma-ray, acoustic, density, neutron, and resistivity log. The ratios of micro into shallow and shallow into deep resistivity logging (MSFL/LLS and LLS/LLD) may be used to estimate permeability. In high permeable zones, where mud filtrate may penetrate far into the formation, MSFL measurements are closer to LLS readings, but in impermeable formations, LLS is practically identical to LLD. Typically, three intelligence models are applied to estimate permeability: fuzzy logic, neuro fuzzy, and neural network. Typically, datasets are separated into training sets and testing sets. The testing set is utilized for evaluation after the intelligent models have been trained using the training set.

3. Discussion

The reservoir flow zonation (rock typing) technique combines geology, petrophysics, seismic, and
reservoir data to identify zones with comparable flow and storage capacity. In this research, all methods of rock classification were studied and evaluated. Examining sedimentary facies from accessible cores and cuts is the first step in analyzing rock classification. The sedimentary rock type will be determined via core descriptions and microscopic examinations of thin sections. According to electrofacies analysis of well logs data, petrophysical rock types are identified, and then HFU are calculated with the association of core data. The NMR log is useful in identifying rock types and permeability, particularly in tight reservoirs, since FZI estimation may be poor from the conventional log because of substantial heterogeneity in these reservoirs. Several techniques for determining flow units, including Amaefule’s approach, Lucia’s cross-plot, and Winland’s/equations. Testerman method has been shown through the zonation coefficient for identification of the range of zone heterogeneity. This study demonstrated that the K-mean approach is the most dependable and consistent method for defining reservoir electrofacies compared with the other techniques investigated in this study. It is a crucial step in reservoir model characterization that improves reservoir models’ performance in development plans and economics. Using petrophysical rock properties data such as (porosity and permeability), hydraulic flow units (HFU) are calculated. Due to the difficulties in determining the borders between distinct rock kinds, the hydraulic flow units anticipated various rock types. The NMR log is useful in identifying rock types and permeability, particularly in tight reservoirs, since FZI estimation may be poor from the conventional log because of substantial heterogeneity in these reservoirs. Several techniques for determining flow units, including Amaefule’s approach, Lucia’s cross-plot, and Winland’s/equations. All flow unit determination techniques must be tested against core data, and the approach most compatible with sedimentary rock types and petrophysical rock types of results is selected. Each reservoir rock type requires analysis of SCAL data and the assignment of representative capillary pressure and relative permeability curves.

4. Conclusions

This study reviews many approaches (deterministic and stochastic) for determining flow zonation and for different reservoir types, including conventional and unconventional and also propose the best approach for each by presenting the advantages and disadvantages, especially in the methods that use statistical operations as a non-deterministic method. Permeability estimation for uncored wells is also included in the study’s agenda through a simplified presentation of the techniques and methods used in this scope. Most methods of rock classification were studied and evaluated in this research. The sedimentary rock type will be determined via core descriptions and microscopic examinations of thin sections (e.g., SEM). According to electrofacies analysis of well logs data, petrophysical rock types are identified, and then HFU are calculated with the association of core data. The NMR log is useful in identifying rock types and permeability, particularly in tight reservoirs, since FZI estimation may be poor from the conventional log because of substantial heterogeneity in these reservoirs. Several techniques for determining flow units, including Amaefule’s approach, Lucia’s cross-plot, and Winland’s/equations. Testerman method has been shown through the zonation coefficient for identification of the range of zone heterogeneity. This study demonstrated that the K-mean approach is the most dependable and consistent method for defining reservoir electrofacies compared with the other techniques investigated of machine learning, despite the presence of some drawbacks in running this approach. Hierarchical Clustering is often used when an application demands a hierarchy. The benefit of Hierarchical Clustering is that we do not need to define the clusters beforehand. However, it could function better on massive volumes of data or big datasets. In addition, Hierarchical Clustering has the problem of not being suited for bigger datasets. Moreover, it only yields the greatest outcomes in certain instances. Automated reservoir zonation has been introduced as a novel approach by previously mentioned researchers using clustering algorithm methods like (Elbow and PELT). The study also proved that HFU, Clustering analysis, and Automated approach are more reliable than other methods and techniques because of their reliance on core analyses and well logging with the presence of lithofacies and electrofacies, which makes their accuracy great, especially in identifying the flow zone and estimating the permeability of the heterogeneous hydrocarbon reservoirs. The results of the clustering method are good. Many algorithms can be used, but they still need to reach the accuracy of the HFU approach. HFU has a physical meaning for that arrangement according to the magnitude of the permeability. In contrast, cluster analysis does not have that for grouping the data and depends on a random arrangement or specifying the centroid in some of them. We need an algorithm to put a constraint on the separation method by teaching it under supervision so that the separation between the types of rocks is based on the basis that it has meaning and is not only divided into any normal data.
Funding

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Authorship contribution statement

Alaa S. Al-Rikaby: Writing - original draft, Validation, Software, Methodology. Mohammed S. Al-Jawad: Writing - review & editing.

Conflicts of interest

There are no conflicts of interest.

References

21. Jennings JW, Lucia FJ. Predicting permeability from well logs in carbonates with a link to geology for interwell permeability mapping. SPE. 2001;71336:1–16.
24. Kazem AL, Mike OO. A Robust Approach to Flow Unit Zonation, University of Port Harcourt, Nigeria. SPE; 2005:98830.
25. El-Sayed AMA, Sayed NAE. Pore aperture size (r36) calculation from porosity or permeability to distinguish dry and producing wells. Arabian J Geosci. 2021.


