Advance approach for differentiation of Fresh water and Hydrocarbon using of AI/ML technique in integration of Saturation Height Modelling of complex clastic: Assam and Assam Arakan Basin, India.

Mohan Satpute *(Oil and Natural Gas Corporation Ltd.),* ***Manoj Kumar*** *(Oil and Natural Gas Corporation Ltd.) and* ***Shanu Batra*** *(Halliburton)*

Email ID: Satpute_MG@ongc.co.in

Keywords
Core data integration, Machine Learning, Saturation Height Modelling.

Abstracts
The permeability of the Pliocene Sandstone reservoirs of the Tipam Formation has been modelled through a thorough investigation of various petrophysical attributes. To achieve this, petrophysical zone attribute calculations were carried out on processed logs of all available wells, while facies identification was performed through conventional log studies of some key wells, and porosity-permeability relationships were established through the flow unit concept by the aid of core data available in a single well.
Hydraulic flow units were identified using the Lorenz Plot and were integrated with K-Mean Clustering supervised machine learning techniques to obtain continuous facies and permeability data for the remaining wells. Flow unit boundaries were picked using an interactive histogram and a Lorenz plot of cumulative storage capacity against cumulative flow capacity. K-Mean clustering was employed for continuous facies identification, and the resulting continuous permeability curve, derived from the hydraulic flow units, and processed porosity data computed from the multi-mineral solver were used for identifying the geometry of the pores.
The resistivity-independent saturation of the Sandstone reservoirs was computed based on the relationship between capillary pressure, water saturation, porosity and permeability which are influenced by various factors such as interfacial tension, contact angle, density difference, grain size, grain shape, sorting, and cementation. Capillary pressure analysis aided to determine the pressure required to displace fluids within the rock. Hydrocarbon zones generally exhibit higher capillary pressures compared to freshwater zones due to the immiscibility of hydrocarbons and water Capillary pressure curves were defined for the reservoirs, and the relationship between capillary pressure and water saturation was determined to compute resistivity-independent saturation for each rock class. The integration of various techniques and methodologies such as the Stratigraphic Lorenz Plot, ML techniques, K-Mean clustering, and capillary pressure curves provided an effective means of modelling permeability and computing resistivity-independent saturation, which can help in the optimal exploitation of the Geleki oil field and to identify the fresh water aquifers below the hydrocarbon zones.

Introduction

The Geleki structure in North Assam Shelf is a doubly plunging NE-SW trending anticline at the Southern junction of Upper Assam plains and Assam-Arakan fold belt. Northern part of the structure is located in alluvial plains while the Southern Part continues in the sub-thrust block of Naga thrust. The Geleki field is a multi-reservoir clastic field discovered in April, 1968 and it is one of the main producing fields of Upper Assam Basin. Recently, a thorough investigation has been carried out to re-evaluate the oil-bearing Pliocene Sandstone reservoirs of the Tipam Formation. Tipam Group of sediments belonging to Middle-Late Miocene to Pliocene age. Tipam Group is composed of (starting from bottom) Geleki Sandstone, Lakwa Sandstone, Girujan Clay and Nazira Sandstone formations. The entire Tipam Group is predominantly arenaceous with negligible shale / claystone, a product of high energy continental fluvial (braided channel) environment. In this study, petrophysical zone attribute calculation has been carried out on the processed logs of all available wells. Facies identification through detailed core studies of few key wells along with porosity-permeability relationship through RQI Methods have been performed. Hydraulic flow units (Amaefule et al 1993, Orodu et al 2009) are identified using the Stratigraphic Lorenz Plot, based on the change of flow and storage capacity slopes, and further integrated with ML Technique for getting continuous facies and Permeability in rest of the wells. The methodology and integration of machine learning technique helped to identify the similar rock classes required to effectively delineate the missed-out opportunities or the zones for further testing. Self-organizing map (SOM) clustering (Kohenen 1970) for continuous facies identification was built on core-derived facies as a calibration curve into the model, continuous permeability curve derived from HFU, and processed porosity computed from Multi mineral solver, which highlight the distribution of facies and porosity. The heterogeneity in the porosity distribution in carbonate reservoirs and its oil-bearing capacity which has been entrapped because of the non-uniform distribution of the pore spaces were also studied. Capillary pressure curves were defined for any 2-phase fluid system in each rock. The contact angle and the interfacial tension were defined for the corresponding Mercury Injection, Centrifuge, or Porous Plate methods for the Laboratory and Reservoir

using standard formula. Capillary pressure data may also be stress-corrected to reservoir conditions and clay-bound water modifications have also been performed. The all-capillary pressure data or a subset may fit one single equation to generate Leverett J function curves (Darling 2005; Wang et al., 2006). Different models can also be developed to represent various functions. For the generation of saturation equations that are not dependent on resistivity-derived saturation, the best-fit regression approach was employed. Upscaling was observed in the final saturation value, which enhances the final reserve estimation. The relevant facies have been selected that may be propagated in any number of wells for reassessing the reservoir characteristics for upscaling reserve parameters over processed log attributes. Our study provides a novel approach for evaluating resistivity-independent water saturation and helps in understanding the complex carbonate reservoirs of the Bassein formation. The integration of machine learning techniques with petrophysical and geological data has aided in identifying the similar rock classes required for further testing, improving reserve estimation, and potentially enhancing hydrocarbon recovery.

Concept of Hydraulic Flow Unit

The degree of heterogeneity in Clastic Tipam sand reservoir varies considerably. A proper approach for classifying each zone is recommended, both practically and conceptually, for the identification and characterization of facies units within it. The computation of the normalized porosity index (NPI) and reservoir quality index (RQI) is primarily used to evaluate the flow unit/flow facies (Martin et al 1997). The method is based on a modified Kozeny-Carmen equation and the concept of average hydraulic radius (Amaefule et al 1993, Orodu et al 2009). The flow zone indicator (FZI) is frequently used to categorise hydraulic flow units and rock types (Amaefule et al., 1993; Kharrat et al., 2009; Ghadami et al., 2015; Riazi, 2018; Dernaika et al., 2019). The RQI, FZI function is given as follows:

$$RQI = 0.0314 \cdot \sqrt{k\phi} \quad (1)$$

$$NPI = \emptyset_z = \phi_{1-\phi} \quad (2)$$

$$FZI = RQI\, \phi_z \quad (3)$$

Equation 3 may be used to get the FZI values of each sample based on the core permeability (k) and porosity (Ø). HFU is then obtained in order to categorize all samples. For the purpose of permeability prediction in non-cored wells, the HFU has integrated with geological features and hence it been used as petrophysical classes as input in SOM calibration.

Facies Prediction

After incorporating core porosity and permeability, initial flow unit boundaries have been adjusted using an interactive histogram and Lorenz plot of Cumulative Storage Capacity against Cumulative Flow Capacity.

Once the flow unit boundaries have been picked an output HFU curve is created. Both a numerically and a Text curve is the output.

Facies Calibration:

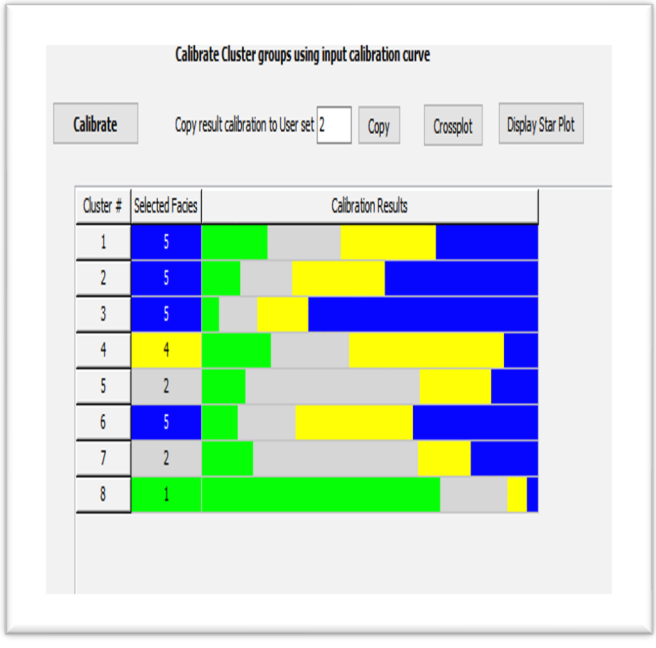
The calibration with core HFU is done by presenting the calibration data to the K-Mean Clustering. The process takes the input data one level at a time and calculates the Euclidean distance from the input vector to each node in the map. It then assigns the calibration data value to each node along with a weighting.

The first stage of Facies Clustering uses the K-mean statistical technique to cluster the data into a known entered number of clusters for this to work an initial guess has to be made of the mean value of each cluster for each input log. The initial guess can affect the results and in order to get good results the initial values should cover the total range of the logs. The initial guesses have been manually entered in the K-mean table of the Cluster Means or the automatic seeding has been applied to Seed Clusters.

The automatic seeding option works by calculating the first principal component log of the data. This will normally contain most of the variation in the data and hence is an ideal way to seed the data to give maximum coverage in the seeded values (Fig-1 Seed Clustering). Once the principal component log is calculated the data is sorted. The data is then divided up equally into the number of wanted clusters. From the data in each cluster the average value for each log is calculated and this value is used as the seed point.

#	Points	Spread	Mean	Std Dev.	Mean	Std Dev.
1	18	0.4679	1.704	0.0541	2.4989	0.03704
2	31	0.2956	1.7665	0.02539	2.4024	0.02949
3	45	0.3084	1.7377	0.02778	2.3183	0.02935
4	19	0.4355	1.7207	0.03091	2.1847	0.05427
5	31	0.3056	1.8016	0.02652	2.2551	0.02957
6	14	0.3613	1.8504	0.02613	2.446	0.03966
7	44	0.3007	1.9963	0.03012	2.4211	0.02273
8	44	0.2185	1.9512	0.01585	2.3312	0.02492
9	24	0.2068	1.9253	0.01671	2.2535	0.02235
10	33	0.2238	2.0133	0.01495	2.3264	0.02569
11	38	0.1679	1.9672	0.01437	2.2635	0.01721
12	37	0.2085	2.0023	0.0193	2.2316	0.01969
13	33	0.2099	1.9686	0.01735	2.1883	0.02185
14	33	0.2822	1.9995	0.02281	2.114	0.02782

(Fig:1 Seed Clustering)



(Fig:2 Cluster Calibration)

K-mean clustering works by assigning each input data point to a cluster. The routine tries to minimize the within-cluster sums of squares of the difference between the data point and the cluster mean value. The routine works by calculating the sum of the squares difference for a data point and each cluster mean and assigning the point to the cluster with the minimum difference. Once all the data points have been assigned to the clusters the new mean values in each cluster are calculated. Using the new mean values the routines starts again re-assigning the data to the clusters. This loop continues until the mean values do not change between loops. These then become the results.

All input log data is normalized (standardized) before starting so that each input log has the same dynamic range. The normalization is done by calculating the mean and standard deviation of the log and then normalizing the data by subtracting the mean and dividing by the standard deviation. Hence a normalized log data value of 1.0 or -1.0 will be one standard deviation.

Within each cluster the standard deviation of the distance of each data point from the cluster mean (units are standard deviation of the original data). Hence, the lower this value the tighter the data is clustered around its mean value.

Mean value of the log for the cluster in units of the input log (i.e., un-normalized).

Std Deviation of the data in a cluster for the log for each individual log, the way the data is grouped in clusters can be viewed. Large standard deviations for all clusters would indicate that this log bears little influence on which cluster the data is assigned to and could probably be excluded from the input with little change in the results.

Cluster consolidation has done completely manually by using the cross plot and log plot output to group data, or a hierarchical cluster technique used to group the data. Hierarchical clustering works by computing the distances between all clusters and then merging the two clusters closest together. The new cluster distance to all other clusters is then recomputed and the two closest clusters merged again. This process continues until we have only one cluster. The results can be plotted as a dendrogram. The dendrogram shows how the clusters were merged and the order they were merged. The numbers at the top of each branch give the merging order. The original results from the K-mean clustering are shown at the base of the plot.

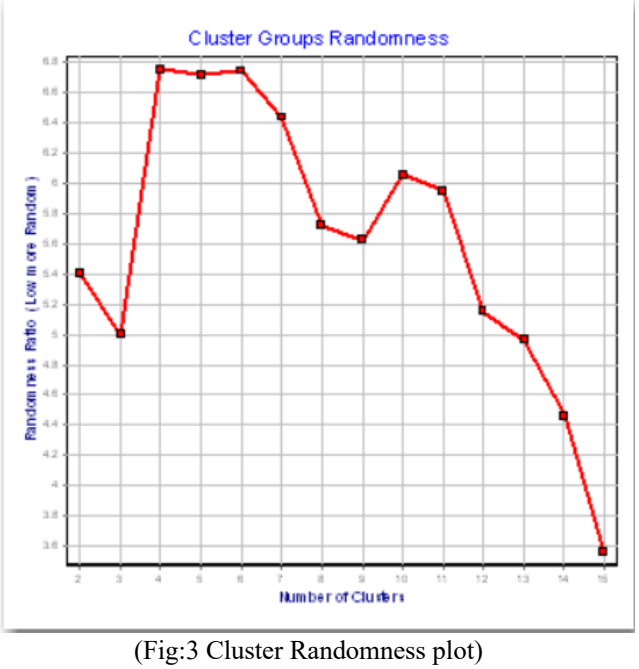
There are five different clustering methods which decide how the clusters are merged. The different methods will show considerably different results. The default method Minimize the within-cluster sum of squares distance gives good results for separating out the different log lithologies into different clusters. The five methods differ in how the distance calculation is updated after two cluster have been joined. If we assume, in the diagram below, cluster "A" and "B" have just been joined to form cluster "Z", and we need to calculate the distance of "Z" with another cluster called "C".
Minimum distance between all objects in clusters: the distance from Z to C is the minimum of the distances (A to C, B to C).

· Maximum distance between all objects in clusters: the distance from Z to C is the maximum of the distances (A to C, B to C).

· Average distance between merged clusters: the distance from Z to C is the average distance of all objects that would be within the cluster formed by merging clusters and C.

· Average distance between all objects in clusters: the distance from Z to C is the average distance of objects within cluster Z to objects within cluster C.
Minimize the within-cluster sum of squares distance: clusters are formed so as to minimize the increase in the within-cluster sums of squares. The distance between two clusters is the increase in these sums of squares if the two clusters were merged.
In general, Minimum distance between all objects in clusters will yield long thin clusters while Maximum distance between all objects in clusters will yield clusters that are more spherical. Average distance between merged clusters and minimize the within-cluster sum of squares

distance tend to yield clusters that are similar to those obtained with Average distance between all objects in clusters.
The grouping of the clusters into a known number of groups is easily done by stopping the grouping at a certain cutoff level. In the dendrogram automatically colours each grouping in a different colour. The groupings above the cutoff are in black.
It is possible to analyse the groupings to decide at which level adding another cluster gives more information or is just adding noise. The Cluster Randomness Plot provides this information.



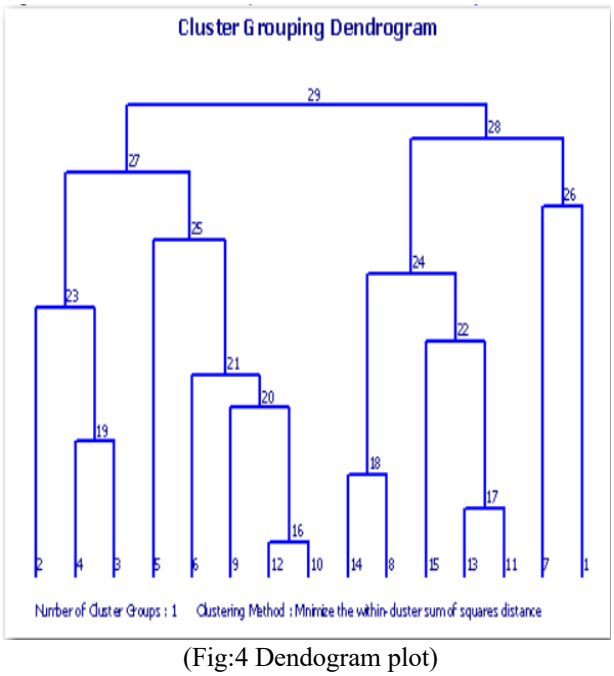
(Fig:3 Cluster Randomness plot)

The chart calculates for each number of cluster groups the perceived randomness of the data. The higher the value the less random the clusters are - i.e. more structured data. The randomness is calculated
by first calculating the average number of depth levels per cluster - i.e. the average thickness of a cluster layer. This is performed on the original log data. Then the theoretical average thickness is calculated assuming the clusters to be assigned randomly at each depth level. The randomness is the ratio of the two. A value of 1 would be totally random, higher values less random. (Fig:3 Cluster Randomness plot)

Av. Thickness = Number of depth levels / Number of cluster layers

$$\text{Random Thickness} = \Sigma\, pi / (1 - pi)$$

Where pi is the proportion of depth levels assigned to the ith cluster.



(Fig:4 Dendogram plot)

Randomness index = Av. Thickness / Random Thickness

The plot is interpreted by picking the number of clusters that are least random (highest peaks). In the above example, a cluster grouping of 6 or perhaps 10 would seem to give the most likely information into the grouping of the log data. (Ref: Multivariate Pattern Recognition and Classification Methods, Geological Log Analysis Using Computer Methods, J.H Doveton.)

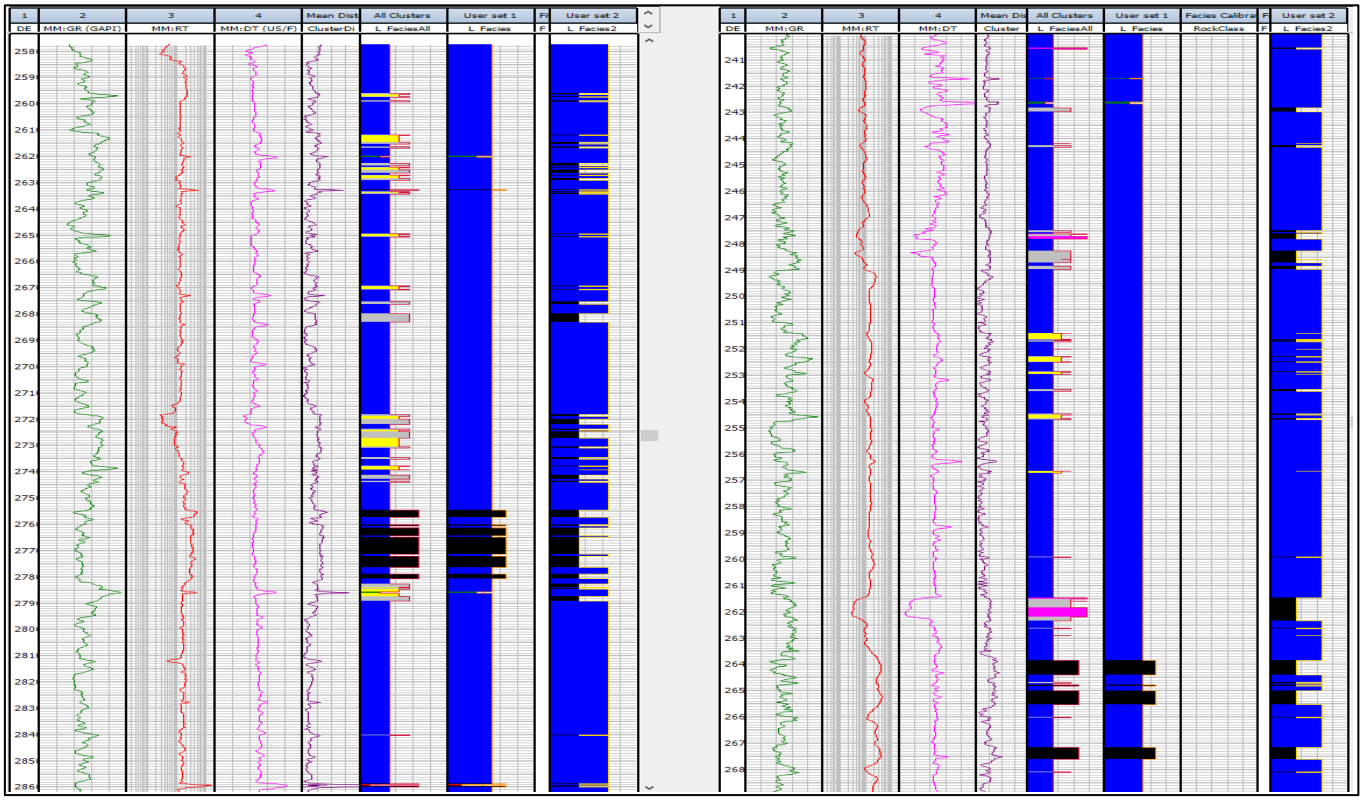
Rather than performing a cluster consolidation using Hierarchical clustering or manually grouping clusters, we have used external facies curve for cluster consolidation to group/calibrate the K-Mean clusters (Fig:4 Dendogram plot).
External facies curve is used for the calibration and does not affect the K-Mean results.

The calibration curve is a Text curve which contains the names of the facies. NOTE: the calibration curve cannot be a continuously variable curve like core permeability. The results are displayed in the table shown here in (Fig:2 Cluster Calibration).

For each input calibration data point the multi-dimensional distance from the calibration point to each K-Mean cluster point is calculated. The value of the calibration curve is then stored at each cluster with a weighting factor which is the inverse of the square of the distance of the calibration point to this cluster point. Once all the calibration data has been processed the weighted average for each calibration facies for each K-Mean cluster is calculated. For each K-

Mean cluster the facies with the highest weighting is considered the most likely result.

In the calibration table the 'Selected facies' is the facies with the highest weighting. The 'Calibration Results' column gives the relative weighting for each facies. This allows you to see how good the calibration for a particular cluster is. Clusters which show mostly one facies can be considered well calibrated. Clusters which show several facies with the same percentages would indicate poorly calibrated facies. (Fig:5 Facies output)



(Fig:5 Facies output)

Saturation Height:

The relationship between capillary pressure and water saturation is influenced by the interfacial tension between the two immiscible fluid phases present, the contact angle between the wetting phase and the rock surface, the density difference between the two fluids, as well as grain size, grain shape, sorting, and cementation. Capillary pressure curves can be defined for any 2-phase fluid system in each rock. The only variables are the Contact Angle (θ) and the Interfacial tension (σ).

As a result, capillary pressure curves from various measurement methods may be normalised to a single system. Given the understanding of these two variables in the two systems, it is also possible to convert between different fluid systems. The Contact

Angle and Interfacial Tension values for the corresponding Mercury Injection, Centrifuge, or Porous Plate methods may be defined for the Laboratory and Reservoir using standard formula. Which will be utilised to transform capillary pressure measurements obtained in the laboratory to a reservoir fluid system. PC data may also be stress-corrected to reservoir conditions and clay-bound water modifications have also been performed. The all Pc data or a subset may fit one single equation to generate Leverett J function curves. Different models can also be developed to represent various functions. For example, we have developed distinct (Fig: 8 Saturation Height Modelling) models that may be used across various porosity ranges. i.e., a model that is effective in the porosity range of 10-15%, another for the range of 15-20%, and a third for the range of 20-25% shown in (Fig-6 Regression models for sw calculation in different facies)

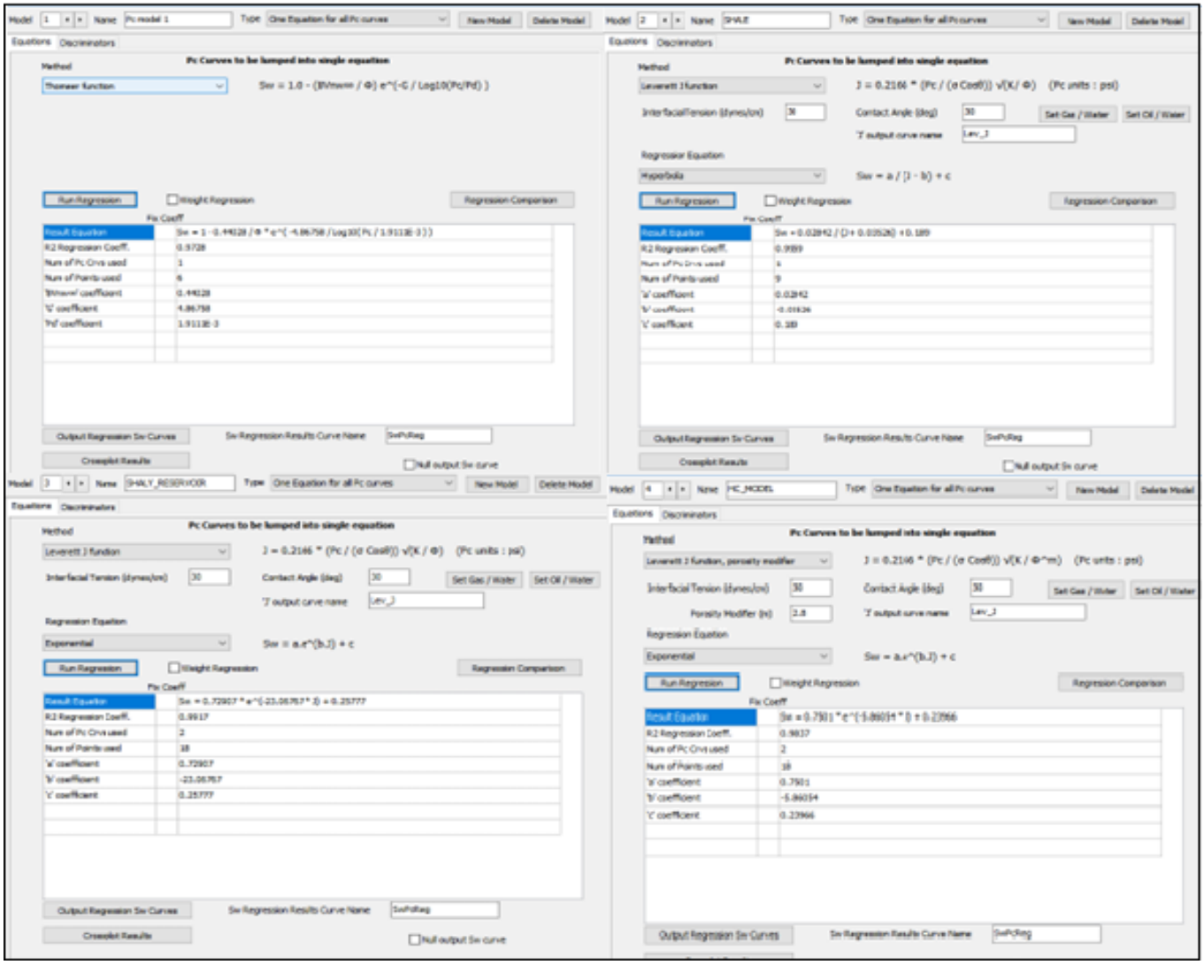
A Regression Function Comparator module in DS Petrophysics is offered to speed up the choosing process; it will evaluate each model and rank them all.

Then we may choose the appropriate one. In our study, the water saturation equations based on various rock classes obtained from HFU have been generated using a comprehensive review on Capillary function from regional free water level. For generation saturation equations that are not dependent on resistivity-derived saturation, the best

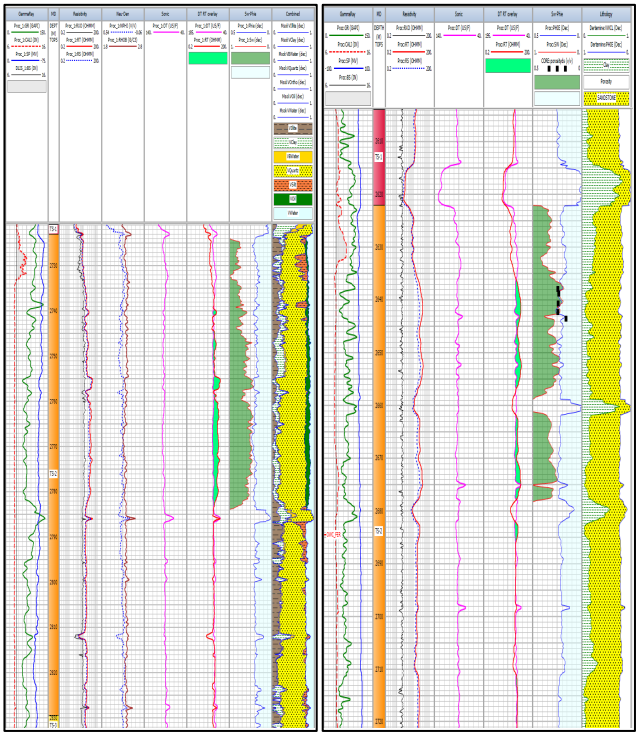
fit regression approach was employed. Upscaling was observed in the final saturation value, which enhances the final reserve estimation shown in (Fig: 8 Saturation Height Modelling)

Resistivity index and relative permeability data provided insights into the fluid flow behaviour within the rock. Hydrocarbon zones typically exhibit different resistivity index and relative permeability curves compared to freshwater zones. We have Analyzed the curves and seen the differences that indicate the presence of hydrocarbons. Regional data have been used to calibrate the analysis with available geological and petrophysical information to ensure accurate differentiation between hydrocarbon and freshwater zones using SCAL data which is eventually matching with the resistivity derived saturation.

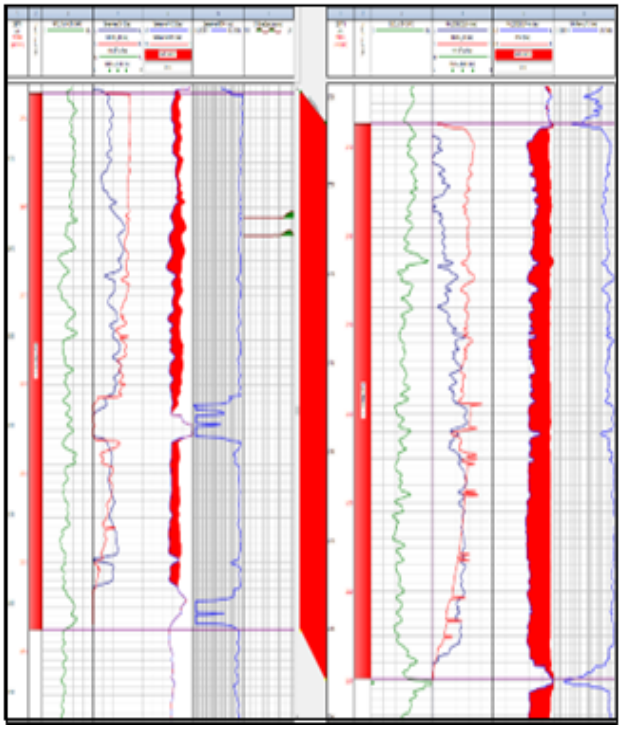
Resistivity index and relative permeability data provided insights into the fluid flow behaviour within the rock. Hydrocarbon zones typically exhibit different resistivity index and relative permeability curves compared to freshwater zones. We have Analysed the curves and seen the differences that indicate the presence of hydrocarbons. Regional data have been used to calibrate the analysis with available geological and petrophysical information to ensure accurate differentiation between hydrocarbon and freshwater zones using SCAL data which is eventually matching with the resistivity derived saturation.



(Fig-6 Regression models for sw calculation in different facies)



(Fig: 7 Resistivity derived saturation).



(Fig: 8 Saturation Height Modelling).

Conclusion

For reassessing the reservoir characteristics for upscaling reserve parameters over processed log attributes, we have selected the relevant facies that may be propagated in any number of wells. Our research introduces a novel approach for evaluating resistivity-independent water saturation. As everyone is familiar, Archie's a, m, and n parameters in highly complex heterogeneous Tipam sand are hard to define. If significant core investigations are not conducted, the usual resistivity dependent saturation parameters may not help to identify the OWC and exact hydrocarbon volume. In certain ways, the combination of SHM and AI/ML algorithms will aid in the accurate modelling of reservoir properties. Capillary pressure analysis aided to determine the pressure required to displace fluids within the rock. Hydrocarbon zones generally exhibit higher capillary pressures compared to freshwater zones due to the immiscibility of hydrocarbons and water.

Acknowledgements

The authors express their sincere gratitude to Oil and Natural Gas Corporation and the managements for providing data and granting permission to conduct the research. We would like acknowledge Mr. RBN Singh (GEOPIC, ONGC Ltd., Dehradun) for his valuable guidance and support throughout the research process. His expertise and insights were invaluable in shaping our research and helping us to overcome challenges.

References:

[1] Amaefule, J.O., Altunbay, M., Tiab, D., Kersey, D.G., and Keelan, D.K. 1993. Enhanced Reservoir Description: Using Core and Log Data to Identify Hydraulic (Flow) Units and Predict Permeability in Uncored Intervals/Wells. Paper SPE 26436 presented at the SPE Annual Technical Conference and Exhibition, Houston, 3–6 October. DOI: 10.2118/ 26436-MS.

[2] Dernaika, M., Masalmeh, S., Mansour, B., et al., 2019. Geology-based porosity-permeability correlations in carbonate rock types, 17-19 September. In: Proceedings of the SPE ReservoirCharacterisation and Simulation Conference and Exhibition. Abu Dhabi UAE.

[3] Ghadami, N., Rasaei, M.R., Hejri, S., et al., 2015. Consistent porosity–permeability modeling reservoir rock typing and hydraulic flow unitization in a giant carbonate reservoir. J. Petrol. Sci. Eng. 131, 58–69. https://doi.org/10.1016/j. petrol.2015.04.017.

[4] Kharrat, R., Mahdavi, R., Bagherpour, M.H., et al., 2009. Rock type and permeability prediction of a heterogeneous carbonate reservoir using artificial neural networks based on flow zone index approach, 15-18 March. In: Proceedings of the SPE Middle East Oil and Gas Show and Conference, Kingdom of Bahrain. https://doi.org/ 10.2118/120166-MS

[5] Lucia, F.J., Kerans, C., Jennings, J.W., 2003. Carbonate reservoir characterization. J. Petrol. Technol. 55, 70–72. https://doi.org/10.2118/82071-jpt.

[6] Kohenen, T., 1970. Self-Organizing Maps, 3rd Edition, Springer.

[7] Martin A J, Solomon S T and Hartman D J 1997 Characterization of petrophysical flow units in carbonate reservoirs AAPG Bull. 81 734–59

[8] Orodu O D, Tang Z and Fei Q 2009 Hydraulic (flow) unit determination and permeability prediction: a case study of block Shen-95, Liaohe Oilfield, North-East China J. Appl. Sci. 9 1801–16

[9] Riazi, Z., 2018. Application of integrated rock typing and flow units identification methods for an Iranian carbonate reservoir. J. Petrol. Sci. Eng. 160, 483–497.

[10] Darling, T., 2005 Well Logging and Formation Evaluation (Oxford: UK Elsevier) p 326.

[11] Wang, Y., Bandal M, Moreno J and Sakdilah M Z 2006 A systematic approach to incorporate capillary pressure-saturation data into reservoir simulation SPE Asia Pacific Oil & Gas Conf. and Exhibition SPE-101013-MS.