



# Electro-facies classification based on core and well-log data

Reda Al Hasan<sup>1</sup> · Mohammad Hossein Saberi<sup>2</sup> · Mohammad Ali Riahi<sup>3</sup> · Abbas Khaksar Manshad<sup>4</sup>

Received: 16 July 2022 / Accepted: 18 June 2023 / Published online: 13 July 2023  
© The Author(s) 2023

## Abstract

Facies studies represent a key element of reservoir characterization. In practice, this can be done by making use of core and petrophysical data. The high cost and difficulties of drilling and coring operations coupled with the time-intensive nature of core studies have led researchers toward using well-log data as an alternative. In the Teapot Dome Oilfield, where core data are limited to those from only a single well, we used well-log data for reservoir electro-facies (EF) studies via two unsupervised clustering methods, namely multi-resolution graph-based clustering (MRGC) and self-organizing map (SOM). Satisfactory results were obtained with both methods, distinguishing seven electro-facies from one another, where MRGC had the highest discriminatory accuracy. The best reservoir quality was exhibited by electro-facies 1, as per both methods. Our findings can be used to avoid some time-intensive steps of conventional reservoir characterization approaches and are useful for prospect modeling and well location proposal.

**Keywords** Multi-resolution graph-based clustering (MRGC) · Self-organizing map (SOM) · Electro-facies · Lithofacies

## Abbreviations

### Abbreviations

AHC	Ascendant hierarchical clustering
ANN	Artificial neural network
ANNs	Artificial neural networks
BMU	The best matching unit
DC	Dynamic clustering
DEN	Density log
DT	Sonic log
GR	Gamma ray log

H-D	Higher-dimensional
KNN	The k-nearest neighbor
LLD	Resistivity log
LF	Lithofacies
L-D	Low-dimensional
MRGC	Multi-resolution graph-based clustering
NPHI or PHIN	Neutron log
PE	Photoelectric log
PRB	The Powder River Basin
RHOZ	Density log
SOM	Self-organizing map
TSF	Tensleep sandstone formation

✉ Mohammad Hossein Saberi  
mh.saberi@semnan.ac.ir  
Reda Al Hasan  
realhasan@albaath-univ.edu.sy  
Mohammad Ali Riahi  
mariahi@ut.ac.ir  
Abbas Khaksar Manshad  
khaksar@put.ac.ir

- <sup>1</sup> Master of Petroleum, Faculty of New Sciences and Technologies, Department of Petroleum Engineering, Semnan University, Semnan, Iran
- <sup>2</sup> Petroleum Engineering Department New Technology Campus, Semnan University, Semnan, Iran
- <sup>3</sup> Institute of Geophysics, University of Tehran, Tehran, Iran
- <sup>4</sup> Petroleum Engineering at the Petroleum University of Technology, Ahvaz, Iran

## List of symbols

$D(x, y)$	Distance function
$D_{ij}$	Distance between input vector and communication weights
$\text{Dist}(X, Y)$	Euclidean distance between two samples $x$ and $y$
KRI	Kernel representative index
$M(x, y)$	Neighborhood function
$N$	Total number of data points
NI or $I$	Neighboring index function
$P$	Total number of properties
$s(x)$	Sum of the weighted ranks for a given measurement at point $x$
$T$	Maximum number of iterations or training length

$t$	Iterations number of the training algorithm
$W_{ij}$	Communication weights between the map unit $j$ and the input sample $i$
$x$	$m$ Th nearest neighbor of $y$ ( $m \leq N - 1$ )
$X_k$	Input vector
$\alpha(t)$	Learning rate
$\alpha_0$	Initial training rate
$a$	Smoothing parameter greater than zero

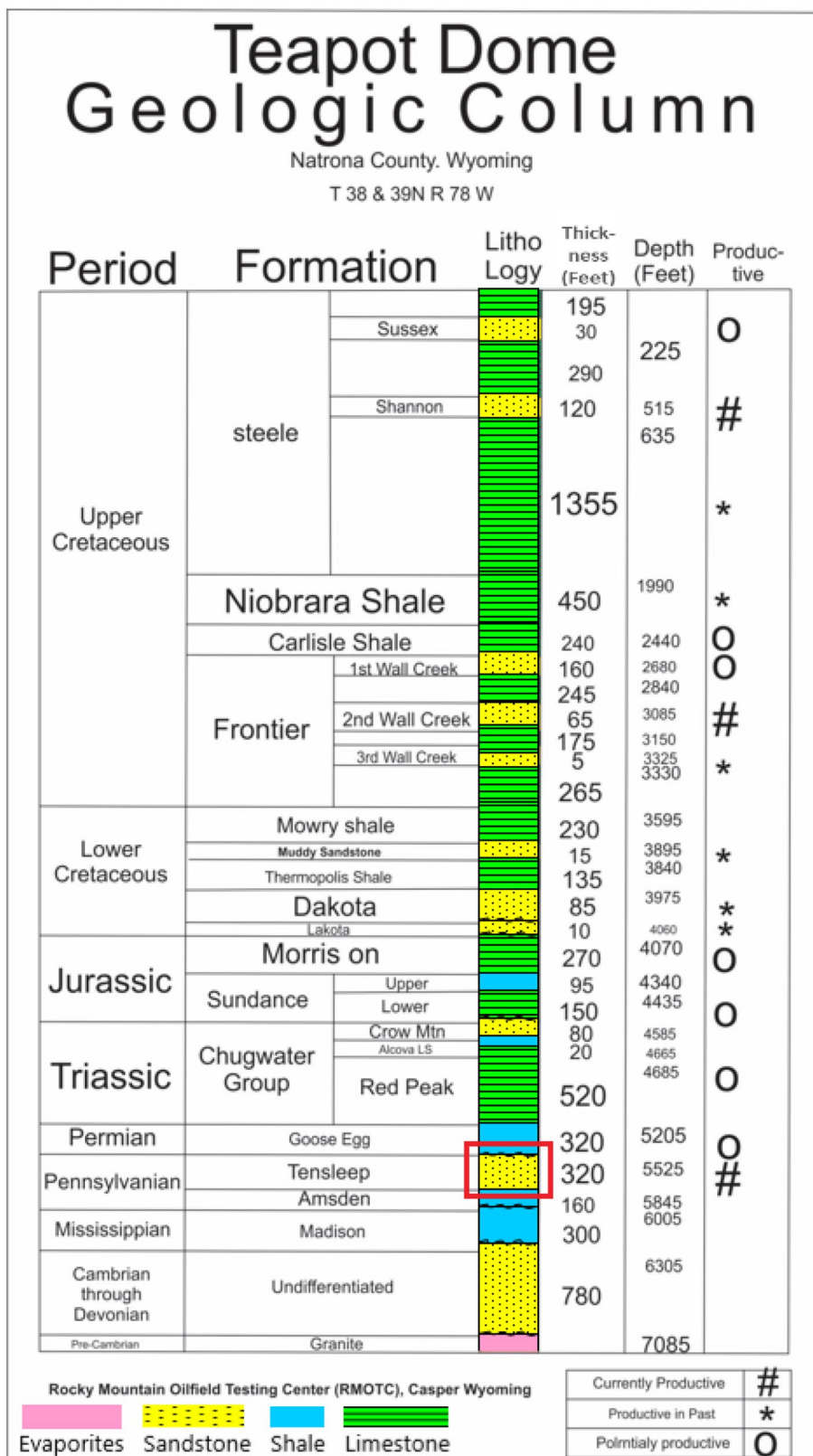
## Introduction

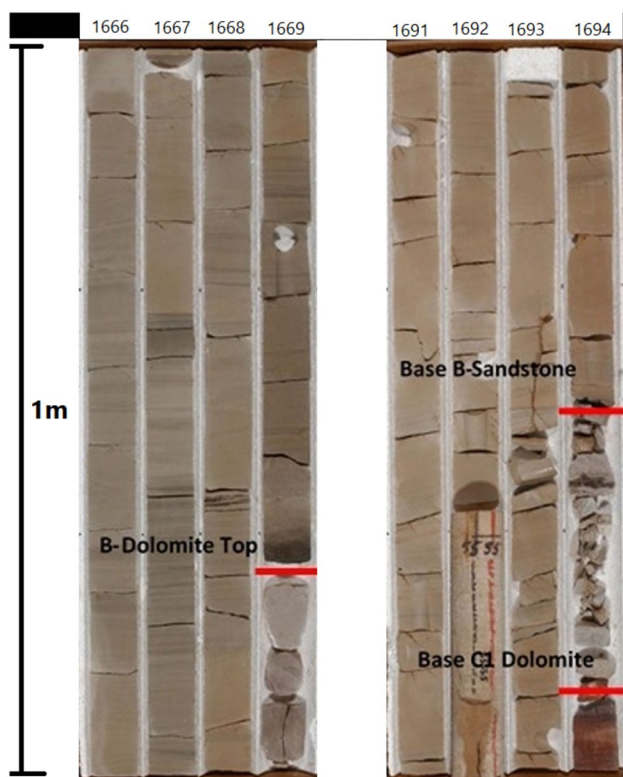
Electro-facies (EF) study is a basic step of any reservoir characterization effort. Conventionally, it has been done based on a combination of core and petrophysical (i.e., well-log) data. In the meantime, the relatively high cost of coring and difficulties of core study as the number of cores grow larger have boosted the importance of well-log data for lithofacies studies. Well logs have been used for reservoir characterization and geological evaluation. Manual analysis of well-log data is however difficult to perform due to the relatively large amount of such data. EF studies have been emerged to address this problem. An EF refers to a series of log responses that characterize a particular rock unit as is distinctive of other rock units (Serra et al. 1982). An EF usually captures one or more reservoir properties as log responses are measurements of physical rock properties. The concept of EF was originally used by geologists to identify rock units of similar properties and hence mature prospects of hydrocarbon, coal, minerals, etc. The term (i.e., EF) was however first coined by Gressly (1838). By definition, EF refers to a specific set of properties for a sedimentary rock unit. These characteristics initially included only geological and fossilological features, including color, stratification, composition, texture, sedimentary structures, and fossil appendages. Later on, a better and more comprehensive definition of EF was proposed by Selley (1976): A sedimentary EF characterizes a complex of sediments or sedimentary rocks with particular lithology, geometry, fossil appendages, sedimentary structures, and paleo-stream patterns, which can be discriminated from other sediments. Lithological-based facies are referred to as lithofacies (LF) (Selley 1986). Most geologists have asserted that sedimentary facies represent geological units (Embry and Johannessen 1993; Odezulu et al. 2014; Sisinni et al. 2016; Tomassetti et al. 2018). LFs reflect significant reservoir parameters, with each LF characterizing a distinguished stratum from the others (Kadhim et al. 2015). This highlights the role of LF classification in reservoir characterization (Avanzini et al. 2016). Previous research works have demonstrated that proper arrangement of LFs facilitates the interpretation of vertical distribution of geological events such as sedimentary sequences and

evolutions (Serra and Abbott 1982), not to mention reservoir properties distribution (Shoghia et al. 2020). LF studies have been done by core, well-log, and seismic data. Core data can provide accurate information about sedimentary processes, but the cost and time intensiveness of coring and core data analysis limits their availability in many cases. As a workaround, core data have been combined with traditional well-log data to obtain the best description of the LFs (Jarvie et al. 2007; Loucks and Ruppel 2007; Dong et al. 2015). This hybrid approach with well-log and core data has been proven to be effective (Bishop 1995; Lim and Kim 2004; Wang 2012; Bhattacharya et al. 2016; Schlanser et al. 2016). Well logs represent many rock characteristics that are well correlated to core data for geological interpretation (Rider and Kennedy 2011). Among other well-log data, gamma ray (GR), density (RHOB), sonic (DT), porosity (NPHI), and photoelectric index (PE) logs have been acknowledged as the best to describe underground rocks (Davis et al. 1997; Qing and Nimegeers 2008). Many statistical algorithms have been used for EF characterization; these include support vector machines (SVMs) (Vapnik et al. 1997; Smola and Schölkopf 2004), artificial neural networks (ANNs) (Liu et al. 1992; Dubois et al. 2007), and multi-resolution graph-based clustering (MRGC) (Ye and Rabiller 2000; Wu et al. 2020).

In Iranian carbonate reservoirs, researchers have used various mathematical and statistical concepts for EF studies. An EF study of Darian Formation was performed by MRGC at 22 wells penetrating this formation. The reservoir quality corresponding to different EFs was then evaluated using core data. 3D modeling was performed, and sequential simulation was applied as a geostatistical approach. Results confirmed the consistency of the 3D models built by the petrophysical logs over the Darian Formation, indicating the effectiveness of EF studies for reservoir characterization (Mehmandosti et al. 2017). EF classification in reservoir zones has been practiced by various researchers. For instance, Pabakhsh et al. (2012) used the MRGC for estimating photomechanical properties and hence distinguish between different lithological formations. Kumar and Kishore (2006) used an ANN to estimate EFs in a carbonate/clastic reservoir. In another work, Bahar et al. (1999) introduced different clustering methods and used them to identify reservoir EF in a carbonate reservoir. Ye and Rabiller (2000) used MRGC for classification of EFs. Using different clustering methods including MRGC, self-organizing map (SOM), dynamic clustering (DC), ascending hierarchical clustering (AHC), and ANN, Khosbakht and Mohammadnia (2012) achieved acceptable permeability prediction accuracy. Mahmoudi et al. (2011) utilized multivariate cluster analysis at Western Salt Well No. 1 (Bandar Abbas, Iran) to identify different EFs and zonate the reservoir accordingly. Hemmati et al. (2016) employed four clustering approaches to zonate a carbonate reservoir in southern Iran based on geological/petrophysical

**Fig. 1** Geologic column of the Teapot Dome and the study formation inside the red rectangle





**Fig. 2** An image of a core box from Well 48-X-28 showing the top of B-dolomite, the base of B-sandstone, and C1-dolomite (Kamran Jafri et al., 2015)

facies. The best clustering performance was obtained with AHC and MRGC although AHC marginally outperformed the MRGC.

In the Teapot Dome Oilfield, Natrona County, Wyoming, the vast area of the field coupled with the large number of drilled wells complicate the study of facies using core data. Accordingly, we alternatively opted for well logs to undertake EF assessments. To increase the reliability of the results, several logs, rather than one, were used in the present study. These included GR, DT, RHOB, and NPHI, among others.

The main goal of this study was to find the best reservoir quality in Teapot Dome Oilfield based on EF classification on well-log data with the help of MRGC and SOM. The results were then verified and compared to the corresponding core data. In the proposed methodology, our findings can be used to avoid some time-intensive steps of conventional reservoir characterization approaches and are useful for prospect modeling and well location proposal, especially when reservoir heterogeneity is not significant.

The rest of this article is managed as follows. Section “[Geological setting](#)” presents the geological setting of the study area. Section “[Materials and methods](#)” elaborates on the used data and the proposed methodology, with the

results presented and discussed in Section “[Results and discussion](#).” Final conclusions are drawn in Section “[Results and discussion](#)”.

## Geological setting

The Teapot Dome is a faulted dome structure within the Salt Creek Anticline in the southwestern sector of the Powder River Basin (PRB), Natrona County, Wyoming. It is a part of the Basin-Margin Anticline Play of the PRB petroleum province (Dolton et al. 1995), where a Precambrian basement is overlaid by Paleozoic strata composed of relatively thin interbedded sequences of sandstones (dune/interdune origin), dolomites, limestone, and evaporites with marine origin. Sedimentary strata of the Cretaceous age range from fluvial sandstones and shales to marine shales and sandstones.

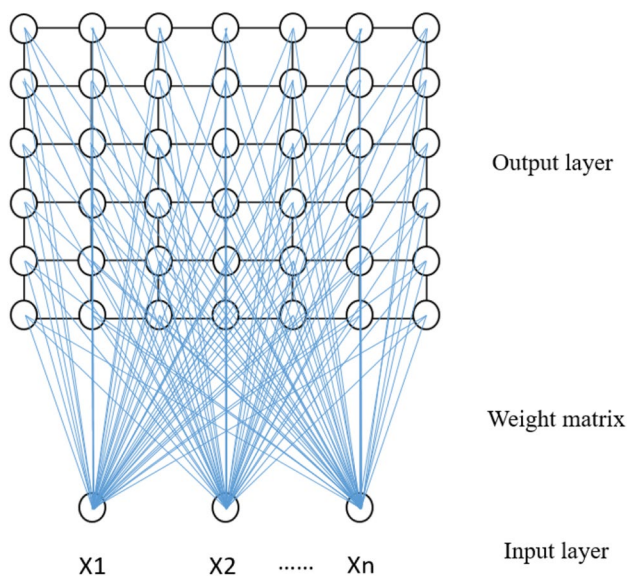
The top two hydrocarbon-producing reservoirs in the Teapot Dome are the Shannon Sandstone and Second Wall Creek Sands of the Upper Cretaceous Cody Shale and Frontier formations, respectively. Hydrocarbons have also been produced from other Upper and Lower Cretaceous formations including Niobrara and Steele Shale and also from Lower Cretaceous non-marine sands of the Thermopolis Shales, the Muddy Sandstone, and the Dakota Sandstone (Dennen et al. 2005).

Being a major hydrocarbon-producing horizon in the Teapot Oilfield, Wyoming, the Pennsylvanian Tensleep Sandstone Formation (TSF) is partially eolian in origin. The TSF consists of marine carbonate/dolomite beds and porous and permeable eolian cross-bedded sandstones of dune and interdune origin. The siliciclastic units comprise the dominant part of the TSF with alterations of dolomite (Dolton et al. 1995, Kamran Jafri et al. 2016). Figure 1 shows the geologic column of the Teapot Dome.

## Materials and methods

### Details of the data used

The core data from Well 48-X-28 were only available for the depth range of 5300–5653 m from mean sea level (MSL)—i.e., only for TSF, and the data from this well were subsequently used for testing the results. Core images and descriptions were used to describe the A- and B-dolomite and B-sandstone units in the TSF, as per a previous study by Kamran Jafri et al. (2016). Figure 2 shows an image of a core box containing core samples from top of B-dolomite, base of sandstone-B and C1-dolomite at Well 48-X-28. Given the wide extension of the study area and



**Fig. 3** A 2D SOM: all nodes of the map are directly connected to the input vector, with no inter-node connection (Kohonen 2000)

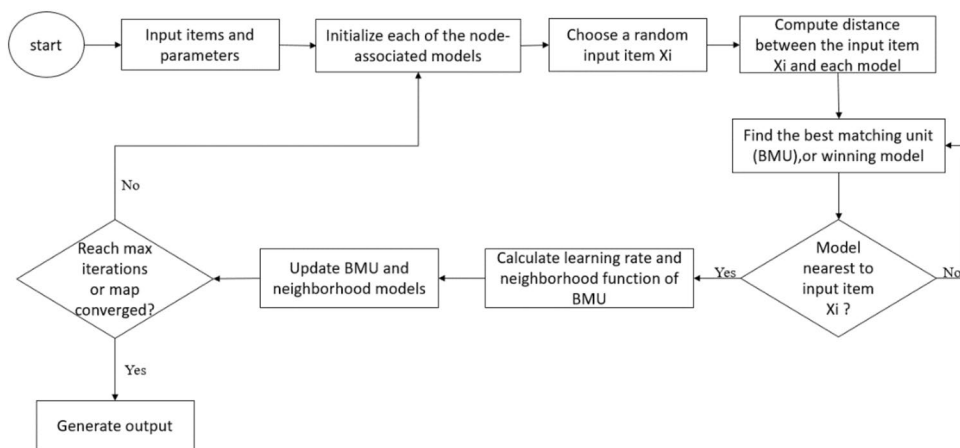
practical infeasibility of studying core data at the many wells drilled into the Teapot Dome, well-log data were focused in this work. In this study, a total of 45 wells were selected where the required well logs for facies study were available. Indeed, well selection was done in such a way to ensure the availability of resistivity (LLD), GR, RHOB, NPHI, and DT logs in most of the wells. LLD measurements can be used to distinguish between hydrocarbon-bearing and water-bearing areas, not to mention its applicability for porosity and permeability evaluations. Among other applications, GR log data help determine the shale (clay) volume ( $V_{shale}, V_{clay}$ ) in sandstone reservoirs containing uranium minerals, potassium feldspar, mica, and/or glauconite, differentiation between radioactive and oil

shale reservoirs, rock source assessment, potash deposit assessment, and geological correlations. A sonic (DT) log serves as an indication of porosity by measuring the interval transit time ( $t, \Delta t, \text{ or } Dt$ ) for a compressed acoustic wave traveling through the formation along the axis of the well, thus indicating the porosity in the rock. RHOB log data contribute to identification of evaporite minerals, gas-bearing zones, and hydrocarbon density, highlighting sandstone reservoirs. NPHI log records porosity data by measuring the hydrogen content of the formation. In a clean (i.e., shale-free) formation where the pore space is filled with water or oil, the NPHI log measures the fluid-filled porosity (Asquith et al. 2004).

### Clustering methods

Clustering has been acknowledged as a powerful approach to data analysis. First introduced in 1935, it has been developed into numerous aspects with a handful of different applications. Multivariate analysis aims at understanding and describing possible associations among multiple variables. In presence of complex relationships among different parameters, single-parameter statistics tend to ignore variations of other variables. Hence, different methods have been developed to handle multivariate data. Properties of univariate and bivariate datasets can be easily extracted by examining the corresponding 2D histogram/plot. A 3D representation is however necessary to analyze a trivariate dataset, which can be developed by compiling multiple 2D histograms of data points. As the number of dimensions increases, these simple solutions may no longer work and one needs to reduce the problem dimensionality before it can be solved properly. Several methods have been used to group similar features (i.e., variables) considering the studied dataset, so as to achieve reduced dimensionality of the problem. Such methods are generally referred to as clustering (Webb 2002).

**Fig. 4** Workflow of a SOM



The basic idea behind the clustering is that visual interpretation of a large problem can be quickly achieved by clustering the entire dataset into distinctive clusters. In this respect, a cluster refers a set of objects with maximum similarity to one another and maximum distinctiveness to other clusters. Herein, the similarity can be measured by different criteria. An example of similarity criterion is the inverse distance between different objects, so that the higher the inverse distance of the objects, the higher their similarity.

The followings are some of the conventional clustering models:

1. Connectivity models (e.g., hierarchical clustering), where model is built based on a distance criterion.
2. Centroid models (e.g., K-means clustering), where each cluster is represented by an average vector.
3. Distribution models, where clusters are modeled using statistical distributions.
4. Density models, where clusters (e.g., areas) are distinguished by density.

Clustering algorithms can be classified based on their clustering model. So far, more than 100 clustering algorithms have been presented. The following subsections explain the two clustering algorithms used in this study.

### Clustering by SOM

SOM is an unsupervised ANN that produces a low-dimensional (L-D) graph called a “map.” It is distinctive of other neural networks in that it uses a neighborhood function to maintain topological properties of the input space, reflecting an L-D representation of high-dimensional (H-D) input data. In other words, SOM maps an H-D input to the corresponding L-D map. This is done by finding the node with the closest weight vector to the input vector and assigning the coordinates of that node on the map to the corresponding input vector. Similar to other neural networks, SOM works in two phases: training and mapping (i.e., automatic classification of unseen input vectors) (Fig. 3).

### Mechanism of the SOM algorithm

An SOM algorithm goes through an iterative process involving vector measurements through the following steps:

1. Initialize the weight for each output node.
2. Select a random vector from the training data and take it to the SOM.
3. Calculate the distance between the input vector and communication weights of each output node as

4.  $d_{ij} = \|x_k - w_{ij}\|$  to find the best matching unit (BMU).
5. Calculate the neighborhood radius around the BMU using the desired neighborhood function. The size of this neighborhood decreases with increasing the algorithm time.
6. Adjust communication weights of the nodes in the BMU neighborhood to make them closer to the BMU. For this purpose, further adjustment must be applied to closer, rather than farther, nodes to the BMU.
7. Repeat Steps 2 to 6 iteratively until the algorithm converges (i.e., the weight vectors exhibit no significant change).

BMU calculation is based on the Euclidean distance, as the similarity criterion, between the weight vectors of the output nodes and the values of the input vectors. The neighborhood radius shrinks as the training algorithm proceeds, eventually encompassing the BMU alone. The neighborhood is usually determined by a Gaussian or exponential function so that nodes closer to the winning BMU are more affected than the farther nodes. The learning rate is set by an exponential function to ensure SOM convergence (Cai et al. 2019).

$$\alpha(t) = \alpha_0 \left(1 - \frac{t}{T}\right) \quad (1)$$

In Eq. (1),  $\alpha_0$  indicates learning rate,  $t$  is the iteration number of the training algorithm, and  $T$  is the maximum number of iterations (i.e., training length).

Then in the mapping step, SOM automatically categorizes unforeseen input vectors.

An ANN is applied in two stages:

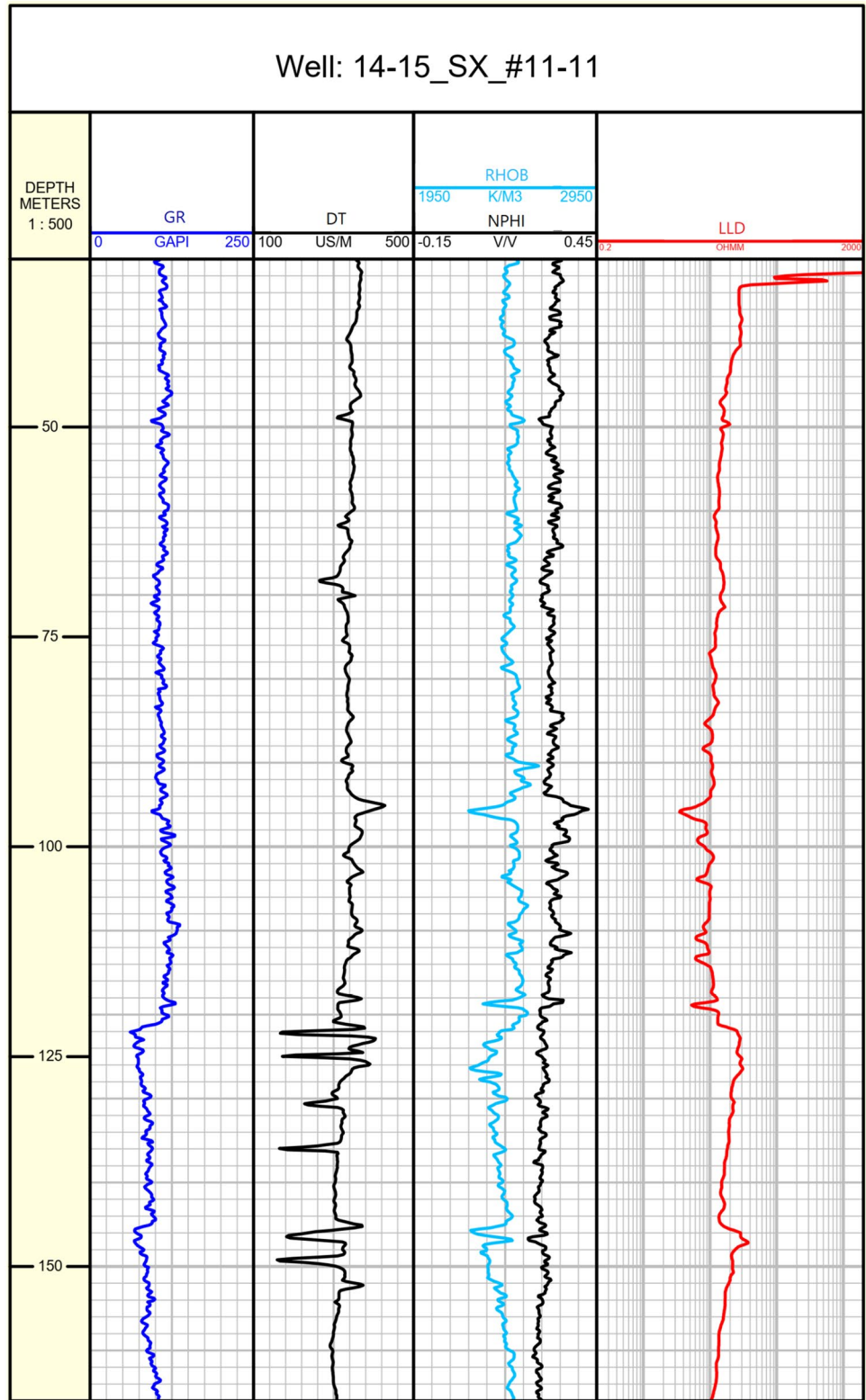
1. Determining the entries from the record dataset.
2. Designing an error backpropagation neural network with an appropriate training algorithm.

An error backpropagation network is a training monitoring tool that feeds inputs to the network and compares the error between the target output and the generated output by the training dataset.

The error of the network is then backpropagated, and the weights are adjusted through multiple iterations. The training process terminates the calculated output which is close enough to the designed output. In many cases, training algorithms have been optimized by updating the weights and biases (Sefidari et al. 2014; Cai et al. 2019), and Fig. 4 shows the workflow of a SOM.

The network efficiency criterion was set to be the correlation coefficient of the considered feature between the modeling results and expected data, which was supposed

**Fig. 5** GR, DT, RHOB, NPHI, and LLD logs along Well 14-15-SX in Teapot Dome and demonstration of the depth matching



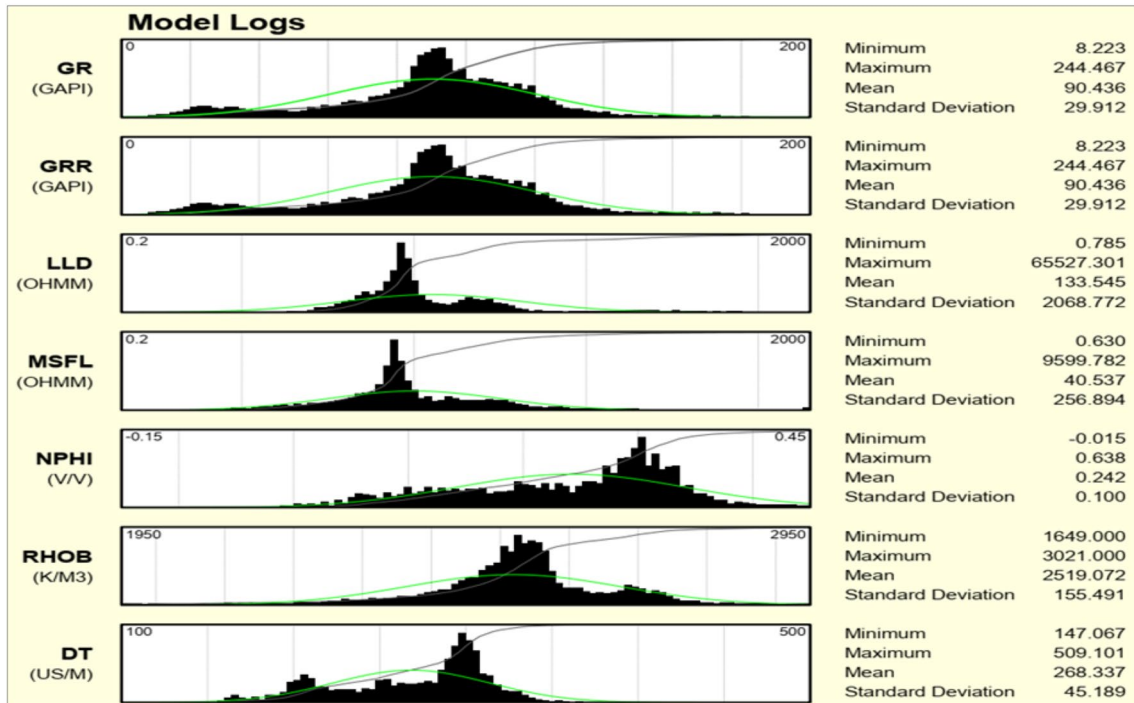
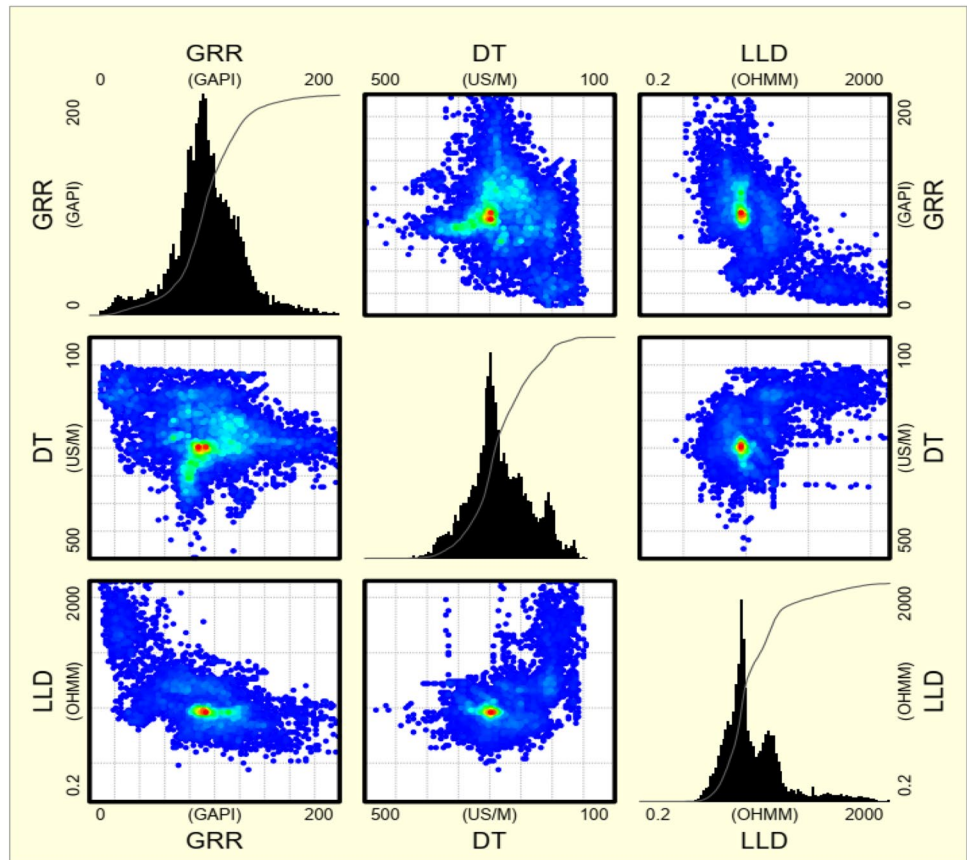


Fig. 6 Histograms of GR, DT, RHOB, NPFI, and LLD logs for the studied wells

Fig. 7 Cross-plots of GR, DT, and LLD logs for the studied wells





to be minimal (Kohonen 1998; Sefidari et al. 2014, Hemmtin et al. 2016; Cai et al. 2019, de Passos et al. 2020).

**Clustering by MRGC**

Conventional clustering algorithms suffer from a number of limitations. First, the number of clusters to be distinguished by the algorithm must be known. Second, they are highly sensitive to initial conditions and distinctions among data points. Third, they are practically not robust to data discrepancy (Mourot and Bousghiri 1993). As a modern-generation algorithm, MRGC has been shown to be free of such limitations and superior to conventional clustering algorithms.

K-nearest neighborhood (KNN) is a clustering algorithm that considers a fixed (i.e., k) number of neighboring data points rather than a fixed neighborhood in space (Dubois et al. 2005). This approach comes with particular advantages. Being easy to implement, this method allows you to record and examine clusters of small size and very different densities. The fact that the power of estimation by KNN is

still exponentially proportional to the number of data points has limited its application for clustering purposes. The KNN classifier is usually based on the Euclidean distance between an experimental sample and a training sample. The Euclidean distance between two samples *x* and *y* is defined as follows:

$$\text{Dist}(X, Y) = \sqrt{\sum_{i=1}^N (X_i - Y_i)^2} \tag{2}$$

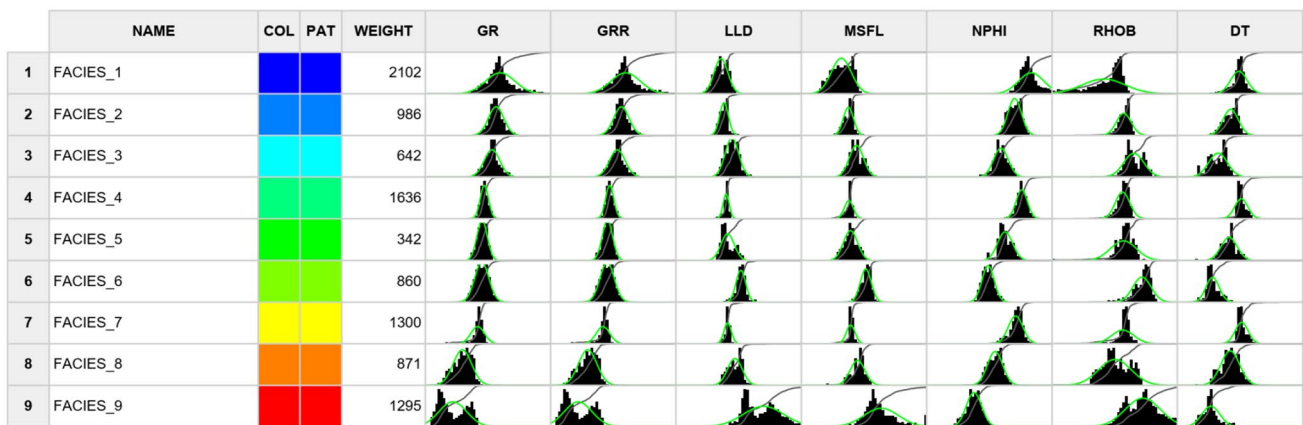
Then the samples are sorted by their distance to the neighbors and the k-nearest neighbors are identified.

Graphical classification methods are known to be suitable for analyzing low-dimensional small datasets. They are generally robust to different batch sizes (Aghchelou et al. 2013).

MRGC is a nonparametric method that combines the KNN method with the graphical approach to enjoy benefits of both for clustering datasets of any dimension and complex structure.

**Table 1** Results of SOM clustering for the studied wells

Facies	Weight	GR (GAPI)	GRR (GAPI)	LLD (OHMM)	MSFL (OHMM)	NPHI (V/V)	RHOB (K/M3)	DT (US/M)
1	2102	112.6	112.6	6.73	5.41	0.31	2446.56	292.02
2	986	107.66	107.66	8.46	7.14	0.27	2498.03	274.07
3	642	99.6	99.6	13.31	11.73	0.21	2576.58	244.24
4	1636	99.39	99.39	8.15	6.71	0.29	2483.24	292.6
5	342	90.03	90.03	12.71	10.41	0.23	2526.39	269.09
6	860	78.2	78.2	25.63	20.38	0.16	2600.23	234.9
7	1300	84.85	84.85	10.15	8.4	0.27	2506.88	292.29
8	871	71.32	71.32	19.91	15.25	0.2	2539.06	263.82
9	1295	58.16	58.16	47.52	32.71	0.13	2608.41	228.93



**Fig. 8** Logs of the EFs identified by SOM clustering based on the input logs

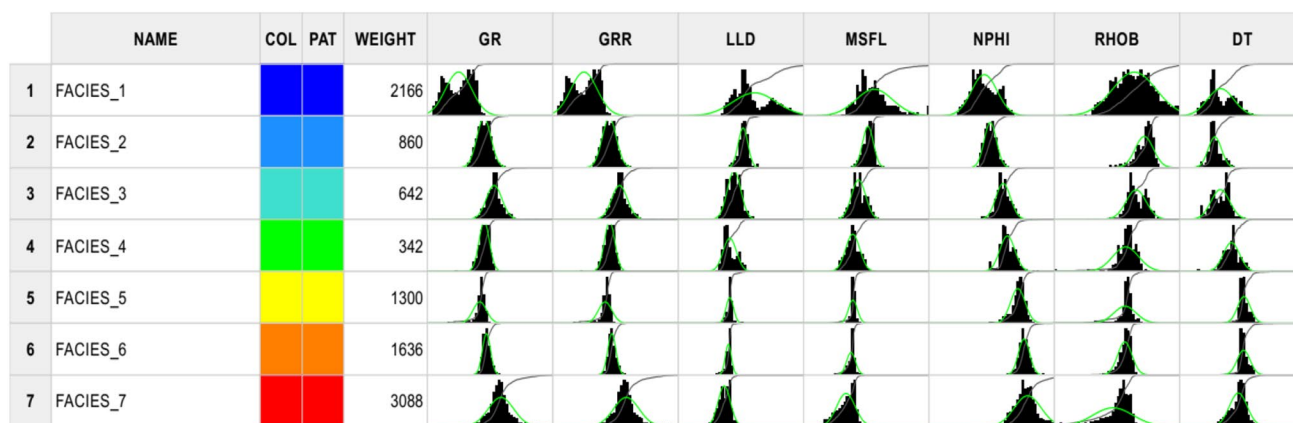


Fig. 9 The EFs identified by the SOM clustering after the merging process

### MRGC: why and how?

Facies analysis is very important to determine reservoir characteristics. Two nearby points along a wellbore may render geologically far different from one another—this issue has been referred to as dimension problem. MRGC offers good capabilities for identifying such geological (i.e., facies) contrasts. Unlike conventional clustering algorithms, this method does not need a previous knowledge of the data structure and number of clusters. Indeed, the optimum number of clusters is herein determined automatically, and this task can be adjusted, according to specific requirements of the problem at hand, by setting the neighbor index (NI) and kernel representative index (KRI) (Ye and Rabiller 2000).

NI is based on the weighted classification of a given measurement point  $x$  concerning all other measurement points  $y$ . Accordingly, a high NI indicates easily distinguishable points (for more information on NI and KNN, see Ye and Rabiller 2001). The number of clusters (i.e., facies in this work) can be easily determined as follows:

$$s(x) = \sum_{n=1}^{N-1} \exp(-m/a) \quad (3)$$

$$S_{\min} = \text{Min}_{i=1,N} \{S(x_i)\} \quad (4)$$

$$S_{\max} = \text{Max}_{i=1,N} \{S(x_i)\} \quad (5)$$

$$NI(x) = \frac{S(x) - S_{\min}}{S_{\max} - S_{\min}} \quad (6)$$

The KRI combines NI ( $x$ ) with a neighborhood function and a distance function. By NI ( $x$ ), the kernel of a cluster can be identified. The number of neighbors,  $M$ , helps create

groups of equal size while groups of equal volume can be obtained by the distance (for more information on KRI, see Ye and Rabiller 2001). KRI is expressed as follows:

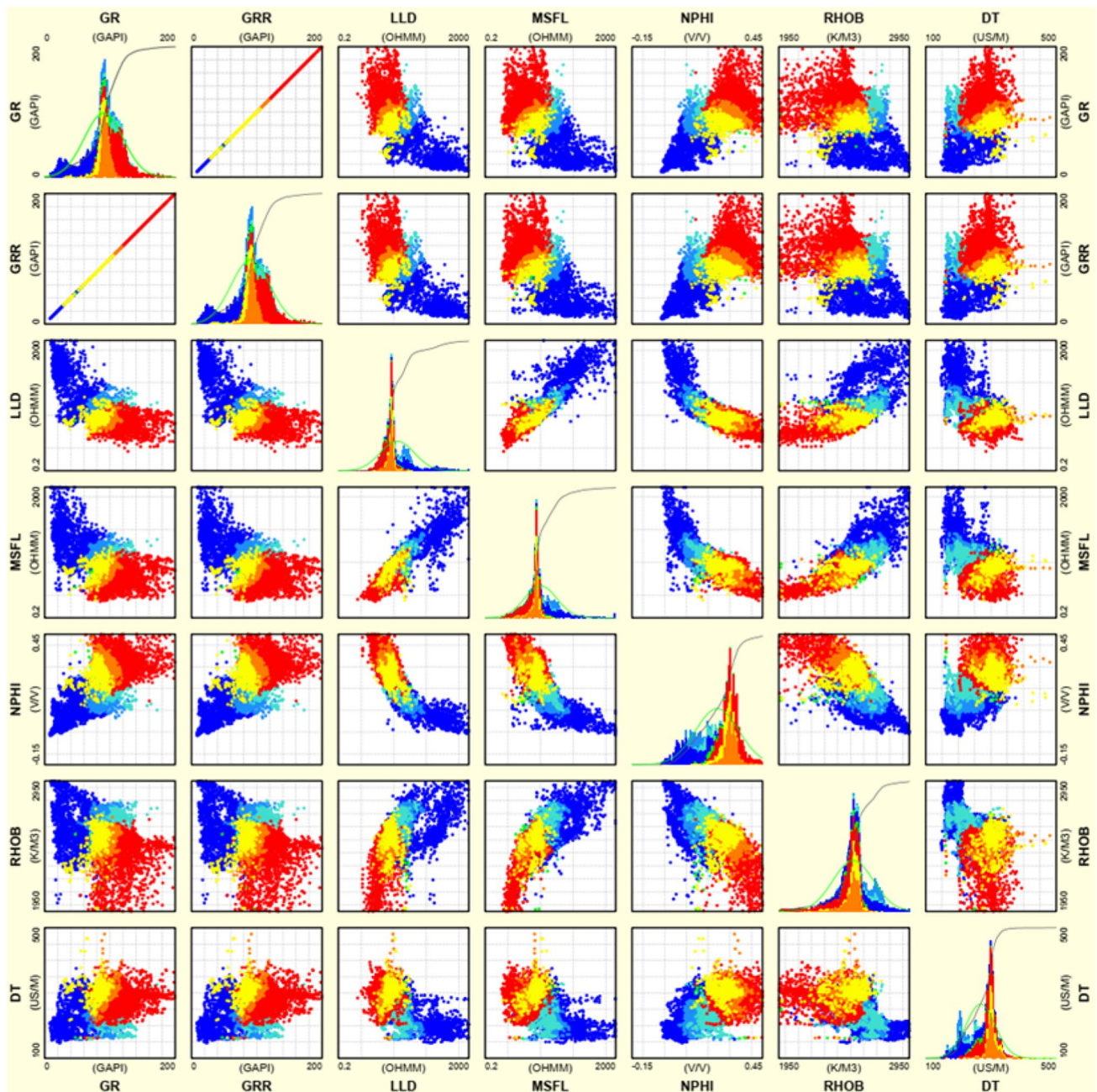
$$KRI(x) = NI(x) \times M \times D \quad (7)$$

Main advantages of MRGC are listed below:

- Ability to identify natural patterns within graph data, representing the facies arrangement.
- No need to previous knowledge of the dataset.
- Automatic determination of the optimal number of clusters (i.e., facies).
- Capability of handling real-time data with complex structure.
- With adjustable parameters, it produces consistent results.
- No theoretical limitation in the number of dimensions, points and categories (Tian et al. 2016; Shi et al. 2017; dos Passos et al. 2020).

## Results and discussion

Well-log data clustering was performed by SOM and MRGC in Geolog software. For this purpose, depth matching was performed based on anomalous log readings and necessary corrections were made. Figure 5 shows the GR, DT, RHOB, NPHI, and LLD logs recorded at Well 14-15-sx in the Teapot Dome together with the depth matching results. Once finished with the initial data verification, facies clustering was practiced using the Facimage module in the Geolog. To this end, we opted for a well that provided good information about the field. Subsequently, feature (i.e., log) selection was performed to identify the training data. Accordingly, GR, DT, RHOB, NPHI, and



**Fig. 10** Cross-plots of the logs used for final SOM clustering

LLD logs were selected and their histograms were checked to see the data distribution and frequency range (Fig. 6). Results indicated the data normality and suitability for training the algorithm.

Figure 7 shows cross-plots of different logs. On this figure, different colors indicate density of the data points. This figure shows that particular logs exhibit similar trends that are well correlated to one another. According to this figure, the highest density of data points was seen to exhibit GR values of 60–120 API and DT values of 200–350  $\mu\text{s/m}$ ,

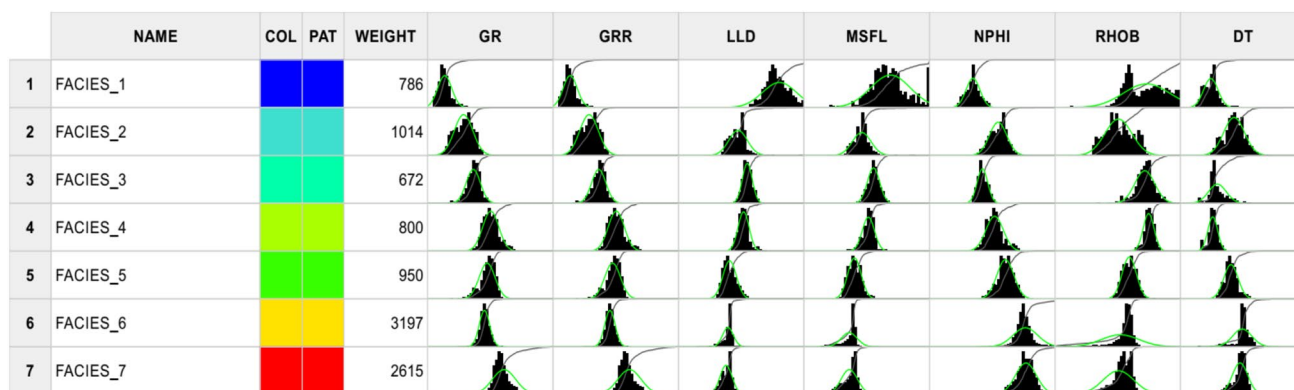
indicating good reservoir quality when GR readings are small.

### Results of SOM

The 2D SOM clustering method was applied to distinguish between different clusters in the data. In this method, horizontal and vertical axes are defined to determine the number of clusters. In this work, input data to the SOM clustering included the horizontal and vertical coordinates

**Table 2** Results of MRGC for the studied wells

Facies	Weight	GR (GAPI)	GRR (GAPI)	LLD (OHMM)	MSFL (OHMM)	NPHI (V/V)	RHOB (K/M3)	DT (US/M)
1	404	26.29	26.29	560.95	248.75	0.05	2816.93	191.49
2	382	28.32	28.32	187.33	49.83	0.06	2534.12	202.22
3	1014	57.35	57.35	15.78	13.77	0.18	2455.48	270.87
4	672	72.81	72.81	31.04	32.78	0.1	2665.26	217.67
5	800	100.7	100.7	23.66	22.49	0.16	2696.15	202.82
6	950	94.82	94.82	8.48	8.08	0.21	2541.57	260.87
7	360	118.49	118.49	3.74	2.08	0.38	2252.92	295.22
8	2512	91.52	91.52	8.35	6.97	0.3	2521.24	305.81
9	2255	122.38	122.38	7.14	6.48	0.3	2488.05	288.97
10	685	87.22	87.22	5.07	2.3	0.34	2281.75	263.87

**Fig. 11** The EFs identified from the MRGC after the merging process

of each group in the network space. A training algorithm was developed to structure SOM ensembles in such way to represent the entire dataset and associated weights at each iteration. At each iteration, a horizontal vector was randomly selected from the dataset and its distance to all weighted vectors of the network was calculated. Therefore, after the training phase, we ended up running the SOM to obtain a model of 9 distinctive facies. Table 1 displays the results of SOM clustering at the studied wells (Fig. 8).

In the next step, EFs of similar characteristics were merged to prevent cluster overgrowth. After reviewing the clustering results, it was found that EF1, EF2, EF3, and EF5 are similar enough to be merged into a single EF. Following this procedure, final number of EFs was reduced to 7. Figure 9 shows the identified EFs after the merging process, and Fig. 10 shows cross-plot of the logs used for the final clustering. From the cross-plots of Fig. 10, it is evident that EF1 and EF7 provide the best and worst reservoir qualities, respectively. On the other hand, EF2, EF3, EF4, EF5,

and EF6 exhibited similar characteristics corresponding to medium reservoir quality. Ultimately, KNN clustering was deployed to generalize the results to all intervals of the well.

## Results of MRGC

MRGC is an unsupervised clustering algorithm for identifying similar areas, categories, or facies. Some unsupervised clustering algorithms (e.g., SOM) require that the number of final clusters is known. MRGC, however, works based on the number of dimensions of the search space, where data distribution density determines the number of clusters. Upon clustering, the produced cross-plots indicate different clusters (i.e., facies) clearly, as shown by different colors. Afterward, one can see the number of facies and distribution of different logs with respect to each facies (i.e., number of samples to which each facies is assigned and average values of each log against that facies). Optimal models were obtained with 10, 14, 18, 20, and 23 facies. Given the geological setting of the

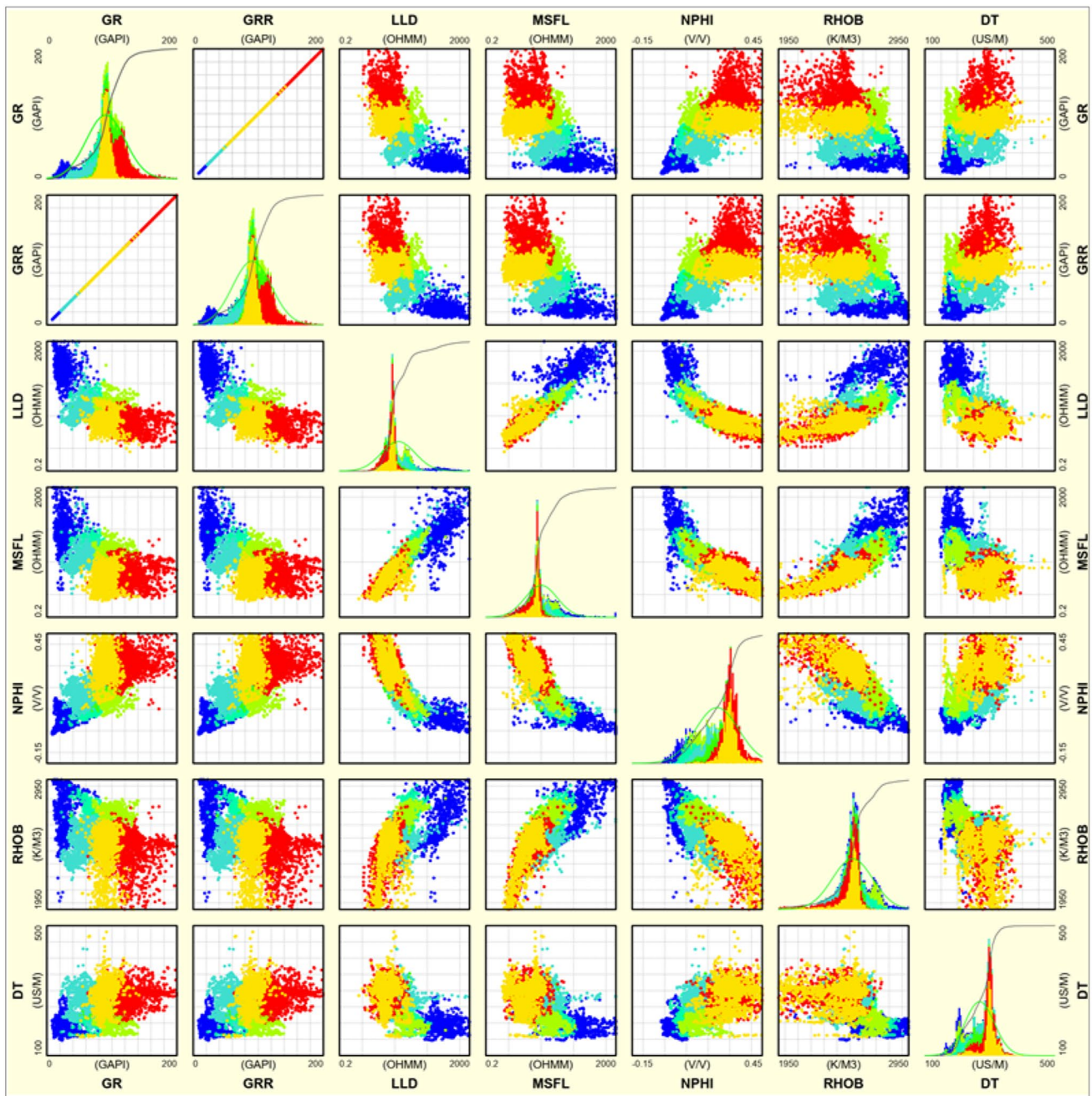


Fig. 12 Cross-plots of the logs used after implementing the MRGC method

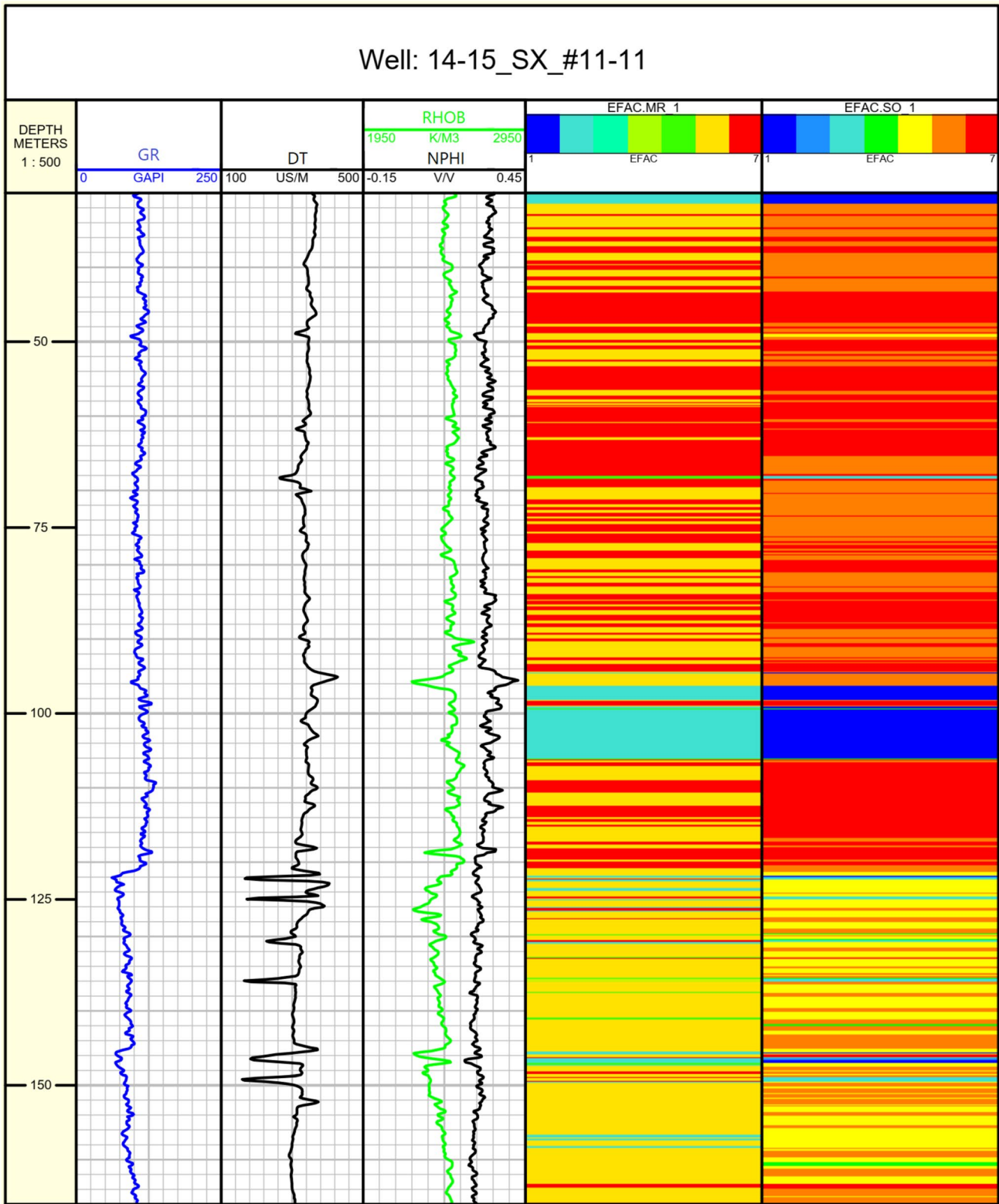


Fig. 13 Comparing EF clustering outputs with petrophysical logs at Well 14-15-SX

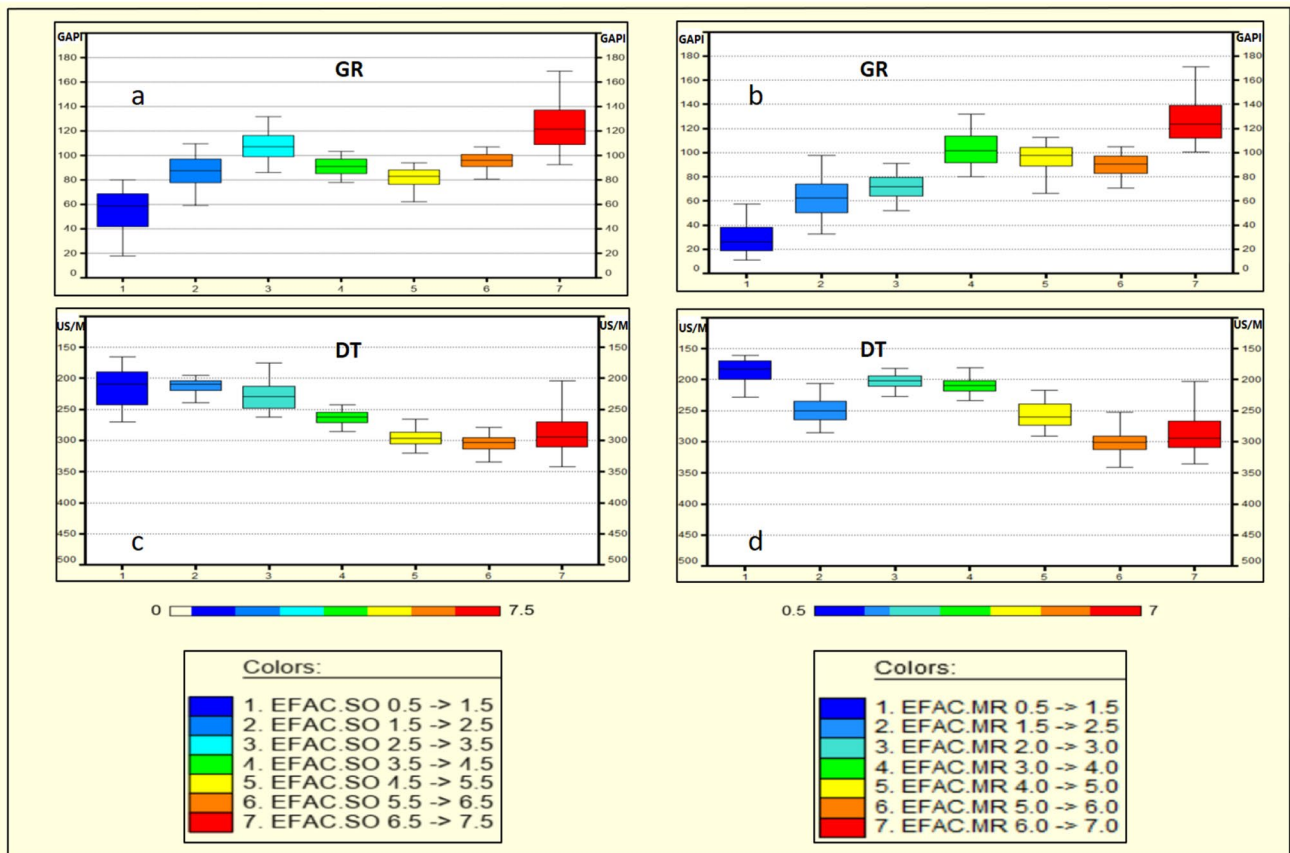


Fig. 14 a, b Boxplots of GR log for the EFs obtained from the SOM and MRGC algorithms, and c, d boxplots of DT log for the EFs obtained from the SOM and MRGC algorithms

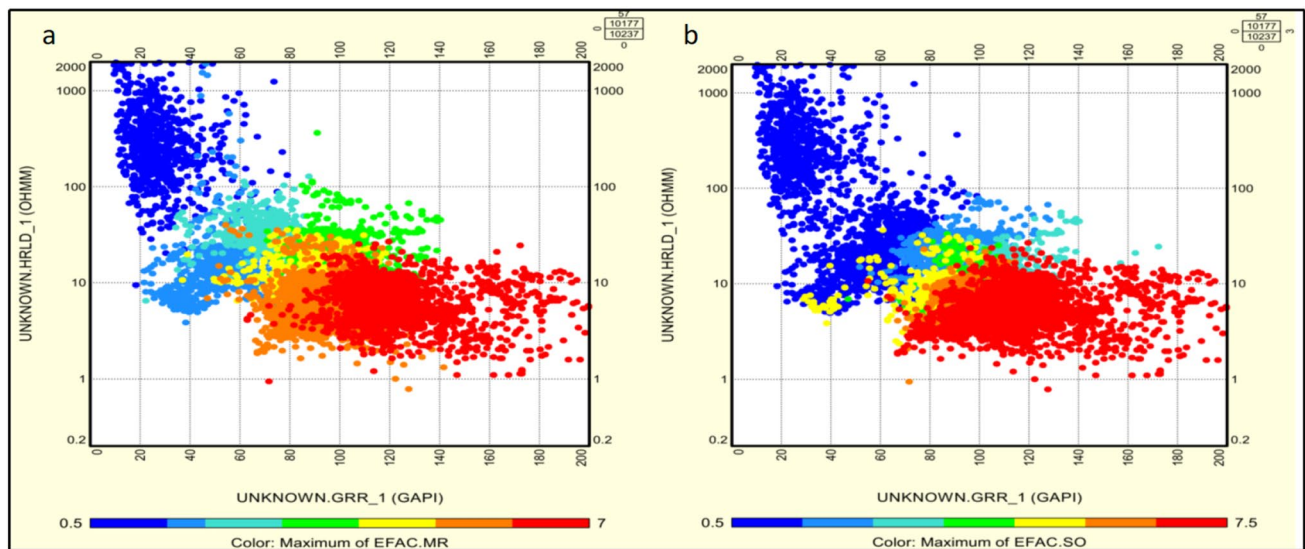
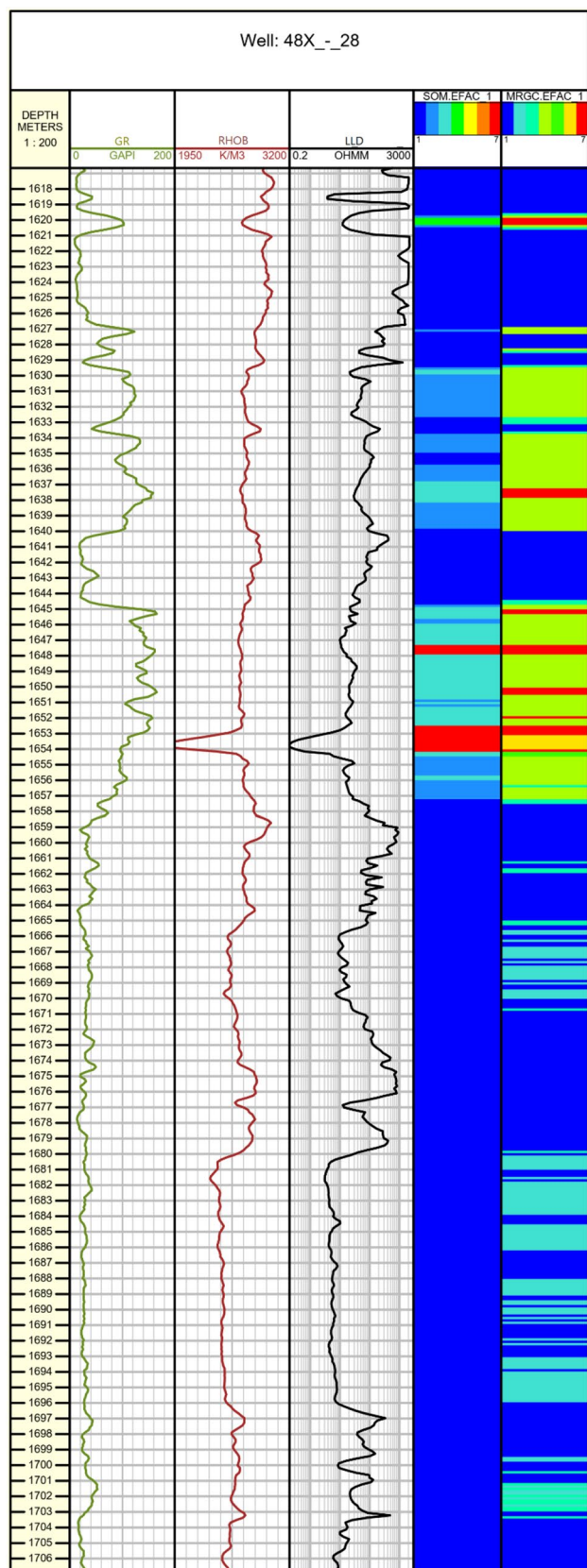


Fig. 15 Cross-plots of GR versus LLD for the EFs resulting from a MRGC and b SOM



**Fig. 16** Comparing the MRGC- and SOM-derived EFs against petrophysical logs at Well 48x-28

study area and previous studies, we ended up developing a 10-EF model. Table 2 displays the results of the MRGC for the studied wells.

As in the SOM method, the EFs of similar characteristics were merged to prevent cluster overgrowth. This led to a reduction of the number of facies to 7. Figure 11 shows the EFs after the merging process, and Fig. 12 shows cross-plots of the logs used for the clustering. From this cross-plot, it is evident that EF1 and EF2 refer to the best reservoir while the worst reservoir quality is associated with EF6 and EF7. On the other hand, EF3, EF4, and EF5 exhibited similar characteristics corresponding to medium reservoir quality.

The KNN algorithm was then utilized to propagate the results to the entire well column. In Fig. 13, one can see and compare the set of identified EFs on petrophysical logs. The leftmost column indicates the depth in m. The GR and DT logs are visualized along Tracks 1 and 2, respectively, while Track 3 hosts RHOB and NPHI logs. The two columns on the right demonstrate the results of EF clustering through MRGC and SOM algorithms, respectively.

### Prioritization of EF logs based on reservoir quality

Reservoir quality is known to be controlled by reservoir porosity, permeability, and shale volume. Accordingly, any geological process that improves either of these parameters can contribute to reservoir quality. Such processes can be physical or chemical. The generated EF logs were examined and prioritized in terms of reservoir quality. For analyzing the EF logs, interpretations were made on each final EF based on the number of EFs merged to obtain that final EF. Separate analysis for EF can be done using box diagrams. Figure 14 shows variations of GR and DT logs in different EFs identified by the SOM and MRGC methods. Indeed, the lower the GR reading and/or the higher the DT reading, the higher the reservoir quality. Accordingly, EF1 and EF2 in Fig. 14a refer to the best reservoir quality, as indicated by their low GR readings (i.e., small shale volume), while EF7 marks the lowest reservoir quality due to its high GR readings. In Fig. 14b, however, the best and worst reservoir qualities are shown by EF1 and EF7, respectively. On Fig. 14c, DT values range from 200 to 300 US/M, while the corresponding DT readings to Fig. 14d fall in the range of 150–300 US/M.

From the above, we find that EF1 exhibits the best reservoir quality by both methods, while EF6 and EF7 showed the worst reservoir quality with the SOM and MRGC methods, respectively. In professional communities, it is usually common to classify different facies as either good or bad ones. To provide a reliable standard for drilling and verify the results in terms of correctness and practicality, Fig. 15 shows a cross-plot of GR versus LLD. This figure validates the results.



**Fig. 17** Photographs of core box from Well 48x-28. On this figure, lithofacies (a), (b), (c), (d), (e), (f), and (g) correspond to EF1, EF2, EF3, EF4, EF5, EF6, and EF7, respectively

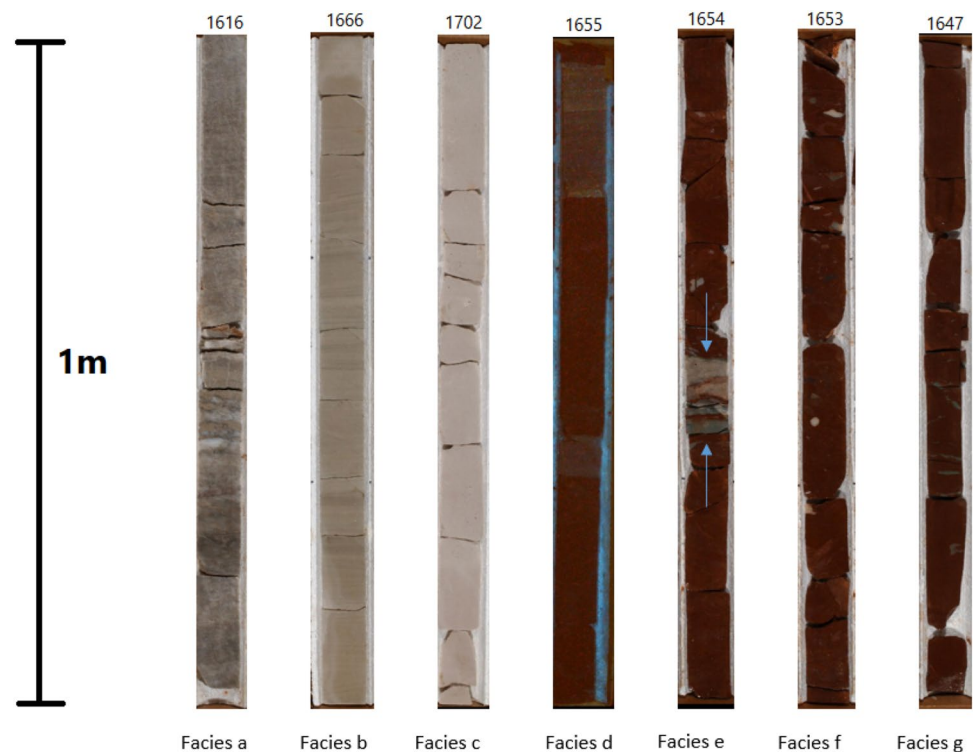


Figure 16 depicts the depth, well logs (GR, LLD, and RHOB), and the results of MRGC and SOM at Well 48x-28. By comparing the MRGC- and SOM-derived EFs with the logs, one may see that the MRGC outperformed the SOM. Since core data were available from the mentioned well in a depth interval of 1615–1723 m, Fig. 17 presents photographs of the core box corresponding to different EFs. On this figure, facies (a) corresponds to EF1, which has the highest porosity coupled with low GR and is encountered in a wide range of depths (1289–1304 m, 1660–1757 m); facies (b) corresponds to EF2, exhibits a GR close to facies (a), and is abundant in the depth ranges of 870–885 m and 1319–1332 m; facies (c) corresponds to EF3, exhibits medium reservoir quality, and occurs in depth ranges of 152–163 m and 1703–1704 m; facies (d) corresponds to EF4 and refers to a rock of lower reservoir quality than EF3, being found in the depth ranges 1308–1312 m, 1430–1559 m, and 1605–1615 m; facies (e) corresponds to EF5 with a medium-to-weak reservoir quality in terms of porosity and GR and occurs in the depth ranges of 757–813 m, 887–930 m, and 1204–1303 m; facies (f) corresponds to EF6 with poor reservoir quality and is found in the depth ranges of 163–405 m, 609–632 m, 1176–1197 m, and 1398–1319 m; and facies (g) corresponds to EF7 with extremely poor reservoir quality and occurs in the depth ranges of 472–753 m, 828–869 m, 931–1171 m, and 1398–1402 m. From Fig. 16, we find that the MRGC method gives a better result than SOM,

and this can be clearly seen by matching the records, for example the facies at a depth of 1694–1703 m.

## Conclusions

Data compilation is the basis of modeling and classification algorithms. A clustering algorithm identifies different clusters of similar properties in a large set of data and tries to maximize the similarity within each cluster while minimizing it between different clusters. The SOM algorithm maps an H-D input space to a L-D map space. The MRGC algorithm is a nonparametric method that combines the KNN method with the graphical classification techniques. Based on the results, the following conclusions were drawn:

- Considering the conditions of the research problem, we focused on a particular set of well logs, including GR, RHOB, DT, NPHI, and LLD.
- Necessary corrections were made to raw data and the Facimage module in the Geolog software was utilized to implement SOM and MRGC algorithms.
- With both methods, we ended up with 7 EFs, with the results verified by cross-plots of GR versus RT for the identified EFs coupled with photographs of the corresponding core boxes.

- With both algorithms, EF1 showed the best reservoir quality, as shown by low GR coupled with high NPHI readings. At the other end of the spectrum, EF6 and EF7 were associated with the poorest reservoir quality, as indicated by high GR coupled with low NPHI readings.
- Comparing the results of the two algorithms with well logs, it was found that MRGC outperformed the SOM in terms of accuracy.
- Using the core data, actual lithofacies corresponding to the identified EFs were delineated.
- EF1 exhibited the best reservoir quality in terms of porosity and GR and occurred in the depth ranges of 1289–1305 m and 1659–1452 m. However, EF7 ended up with the poorest reservoir quality and occurred in the depth ranges of 472–753 m, 827–869 m, 931–1172 m, and 1397–1402 m.

**Funding** This research does not have any fund from the government sector and it was done only with the support of Semnan University.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aghchelou M, Hemmati Ahoei HR, Nabi-Bidhendi M, Rahimi Bahar AA (2013) Lithofacies estimation by multi-resolution graph-based clustering of petrophysical well logs: a case study from one of the Persian Gulf's gas fields, Iran. *Iran J Geophys* 7(4):11–30. <https://doi.org/10.2118/162991-MS>
- Asadi Mehmandosti E, Mirzaee S, Moallelemi SA, Arbab B (2017) Study and three-dimensional modeling of the Dariyan Formation Electrofacies by using Geostatistics, in one of the Persian Gulf Oilfields Kharazmi. *J Earth Sci* 3(1):25–44
- Asquith GB, Krygowski D, Gibson CR (2004) Basic well log analysis, vol 16. American Association of Petroleum Geologists, Tulsa
- Avanzini A, Balossino P, Brignoli M, Spelta E, Tarchiani C (2016) Lithologic and geomechanical facies classification for sweet spot identification in gas shale reservoir. *Interpretation* 4:SL21–SL31. <https://doi.org/10.1190/INT-2015-0199.1>
- Bahar M, Johnstone AH, Hansell MH (1999) Revisiting learning difficulties in biology. *J Biol Educ* 33(2):84–86. <https://doi.org/10.1080/00219266.1999.9655648>
- Bhattacharya S, Carr TR, Pal M (2016) Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: case studies from the Bakken and Mahantango-Marcellus Shale, USA. *J Nat Gas Sci Eng* 33:1119–1133. <https://doi.org/10.1016/j.jngse.2016.04.055>
- Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford, p 482
- Cai H, Wu Q, Ren H, Li H, Qin Q (2019) Pre-stack texture based semi-supervised seismic facies analysis using global optimization. *J Seism Explor* 28:513–532
- Davis R, Fontanilla J, Biswas P, Saha S (1997) Lithology, lithofacies, and permeability estimation in Ghawar Arab-D reservoir. Middle East Oil Show, Society of Petroleum Engineers 37702. Bahrain 15–18:15–17. <https://doi.org/10.2118/37702-MS>
- Dennen K, Burns W, Burruss R, Hatcher K (2005) Geochemical analyses of oils and gases, naval petroleum reserve No. 3, Teapot Dome Field, Natrona County, Wyoming. US Geological Survey Open-File Report, 1275, 69. USGS Open-File Report 2005–1275. BiblioBazaar (2013)
- Dolton GL, Fox JE (1995) Powder River Basin Province (033). In: Gautier DL, Dolton GL, Takahashi KI, Varnes KL
- Dong T, Harris NB, Ayranci K, Twemlow CE, Nassichuk BR (2015) Porosity characteristics of the Devonian Horn River shale, Canada: insights from lithofacies classification and shale composition. *Int J Coal Geol* 141:74–90. <https://doi.org/10.1016/j.coal.2015.03.001>
- Dos Passos FV, Braga MA, Carelli TG, Plantz JB (2020) Electrofacies classification of ponta grossa formation by multi-resolution graph-based clustering (MRGC) and self-organizing maps (SOM) methods. *Braz J Geophys* 38(1):52–61. <https://doi.org/10.22564/rbgf.v38i1.2>
- Dubois MK, Bohling GC, Chakrabarti S (2005) Comparison of four approaches to a rock facies classification problem. *Comput Geosci* 33:599–617. <https://doi.org/10.1016/j.cageo.2006.08.011>
- Dubois MK, Bohling GC, Chakrabarti S (2007) Comparison of four approaches to a rock facies classification problem. *Comput Geosci* 33:599–617. <https://doi.org/10.1016/j.cageo.2006.08.011>
- Embry AF, Johannessen EP (1993) T–R sequence stratigraphy, facies analysis and reservoir distribution in the uppermost Triassic–Lower Jurassic succession, western Sverdrup Basin, Arctic Canada. Elsevier, Amsterdam. <https://doi.org/10.1016/B978-0-444-88943-0.50013-7>
- Gressly A (1838) Observations géologiques sur le Jura Soleurois. *Neue Denkschr. Allg Schweiz. Ges Naturw* 2:1–112
- Hemmati N, Nazari F, Tabatabai R, Hashem S (2016) Determination of electro-facies using clustering methods in one of the carbonate reservoirs in southern Iran. *Sci Mon Oil Gas Explor Prod* 137:67–77. <https://doi.org/10.1016/j.marpetgeo.2018.06.004.inPersian>
- Jafri MK, Lashin A, Ibrahim E, Naeem M (2016) Petrophysical evaluation of the Tensleep Sandstone formation using well logs and limited core data at Teapot Dome, Powder River Basin, Wyoming, USA. *Arab J Sci Eng* 41(1):223–247. <https://doi.org/10.1007/s13369-015-1741-7>
- Jarvie DM, Hill RJ, Ruble TE, Pollastro RM (2007) Unconventional shale-gas systems: the Mississippian Barnett Shale of north-central Texas as one model for thermogenic shale-gas assessment. *Am Asso Petrol Geol Bull* 91:475–499. <https://doi.org/10.1306/12190606068>
- Kadhim F, Samsuri A, Alwan H (2015) Determination of lithology, porosity and water saturation for Mishrif carbonate formation. *Int J Geol Environ Eng* 9:1025–1031
- Khoshbakht F, Mohammadnia M (2012) Assessment of clustering methods for predicting permeability in a heterogeneous carbonate

- reservoir. *J Pet Sci Technol* 2(2):50–57. <https://doi.org/10.3997/2214-4609.20145462>
- Kohonen T (1998) The self-organizing map. *Neurocomputing* 21(1–3):1–6. [https://doi.org/10.1016/S0925-2312\(98\)00030-7](https://doi.org/10.1016/S0925-2312(98)00030-7)
- Kohonen T, Kaski S, Lagus K, Salojärvi J, Honkela J, Paatero V, Saarela A (2000) Self-organization of a massive document collection. *IEEE Trans Neural Netw* 11:574–585. <https://doi.org/10.1109/72.846729>
- Kohonen T (2000) The self-organizing map, [https://www.saedsayad.com/clustering\\_som.htm](https://www.saedsayad.com/clustering_som.htm)
- Kumar B, Kishore M (2006) Electrofacies classification—a critical approach. In: 6th international conference & exposition on petroleum geophysics, New Delhi, India, pp 822–825
- Lim JS, Kim J (2004) Reservoir porosity and permeability estimation from well logs using fuzzy logic and neural networks. In: society of petroleum engineers asia pacific oil and gas conference and exhibition, Perth, Oct. 18–20. <https://doi.org/10.2118/88476-MS>
- Liu RL, Zhou CD, Jin ZW (1992) Lithofacies sequence recognition from well logs using time-delay neural networks. 33rd Annual Logging Symposium of Society of Petrophysicists and Well Log Analysts, Oklahoma, Jun. 14–17, ID: SPWLA-1992-L
- Loucks RG, Ruppel SC (2007) Mississippian Barnett Shale: litho facies and depositional setting of a deep-water shale-gas succession in the Fort Worth Basin, Texas. *Am Asso Petrol Geol Bull* 91:579–601. <https://doi.org/10.1306/11020606059>
- Mahmoudi HS, Harami M, Mehrgini B (2011) Identification and zoning of EF using multivariate cluster analysis method in the western salt well 1 of Bandar Abbas region. In the first national conference of Iranian geology, <https://civilica.com/doc/117628/>. (in Persian)
- Mourot G, Bousghiri S, Ragot J, (1993) Pattern recognition for diagnosis of technological systems: a review: international conference on systems IEEE, pp 275–281. <https://doi.org/10.1109/TAES.2009.5259178>
- Odezulu CI, Olawole S, Saikia K, Mento D (2014) Effect of sequence stratigraphy based facies modeling for better reservoir characterization: a case study from powder river basin. In: paper presented at the abu Dhabi international petroleum exhibition and conference. UAE, Abu Dhabi 2014/11/10/. <https://doi.org/10.2118/171746-MS>
- Pabakhsh M, Ahmadi K, Riahi MA, Abbaszade Shahri A (2012) Prediction of PEF and LITH logs using MRGC approach. *Life Sci J* 9(4):974–982
- Qing H, Nimegeers AR (2008) Lithofacies and depositional history of Midale carbonate-evaporite cycles in a Mississippian ramp setting, Steelman-Bienfait area, southeastern Saskatchewan Canada. *Bull Can Pet Geol* 56:209–234. <https://doi.org/10.2113/gscpgbull.56.3.209>
- Rider MH, Kennedy MK (2011) *The Geological Interpretation of Well Logs* (3rd edition). Rider-French Consulting Limited, Scotland, p 440
- Schlanser K, Grana D, Campbell-Stone E (2016) Lithofacies classification in the Marcellus Shale by applying a statistical clustering algorithm to petrophysical and elastic well logs. *Interpretation* 4:SE31–SE49. <https://doi.org/10.1190/INT-2015-0128.1>
- Sefidari I, Kadkhodai A, Sharifi M (2014) Comparison of self-organizing map methods and cluster analysis to evaluate the amount of organic carbon in hydrocarbon-containing formations using intelligent systems. Faculty of Geology, Campus of Sciences, University of Tehran, 23(75):117–130. (in Persian) <https://www.sid.ir/FA/Journal/ViewPaper.aspx?id=231761>.
- Selley RC (1976) *An introduction to sedimentology*. Academic Press, London, p 408
- Selley RC (1986) *Ancient sedimentary and environment and their subsurface diagnosis*. Chapman and hall, London. 3rd ed. p 317
- Serra O, Abbott HT (1982) The contribution of logging data to sedimentology and stratigraphy. *Soc Pet Eng J* 22(1):117–131. <https://doi.org/10.2118/9270-PA>
- Shi X, Cui Y, Guo X, Yang H, Chen R, Li T, Meng L (2017) Logging facies classification and permeability evaluation: multi-resolution graph based clustering. In: SPE annual technical conference and exhibition. OnePetro, <https://doi.org/10.2118/187030-MS>
- Shoghi J, Bahramizadeh-Sajjadi H, Nickandish AB, Abbasi M (2020) Facies modeling of synchronous successions—A case study from the mid-cretaceous of NW Zagros, Iran. *J Afr Earth Sci* 162:103696. <https://doi.org/10.1016/j.jafrearsci.2019.103696>
- Sisinni V, McDougall N, Guarnido M, Vallez Y, Estaba V (2016) Facies modeling described by probabilistic patterns using Multi-point statistics an application to the kfield. *Libya Am Assoc Petrol Geol*. <https://doi.org/10.1190/ice2016-6320106.1>
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14:199–222
- Tian Y, Xu H, Zhang XY, Wang HJ, Guo TC, Zhang LJ, Gong XL (2016) Multi-resolution graph-based clustering analysis for lithofacies identification from well-log data: case study of intra-platform bank gas fields, Amu Darya Basin. *Appl Geophys* 13(4):598–607
- Tomassetti L, Petracchini L, Brandano M, Trippetta F, Tomassi A (2018) Modeling lateral facies heterogeneity of an upper Oligocene carbonate ramp (Salento, southern Italy). *Mar Petrol Geol* 96:254–270. <https://doi.org/10.1016/j.marpetgeo.2018.06.004>
- Vapnik V, Golowich SE, Smola AJ (1997) Support vector method for function approximation, regression estimation and signal processing. In: Mozer, M.C., Jordan, M.I., and Petsche, T. eds *Proceedings of the 1997 conference on Advances in neural information processing systems*. MIT Press, Cambridge, pp 281–287
- Wang G (2012) Black Shale lithofacies prediction and distribution pattern analysis of Middle Devonian Marcellus Shale in the Appalachian basin, Northeastern USA. Ph.D. Thesis, West Virginia University, Morgantown, USA, p 216
- Web AR (2002) *Statistical pattern recognition*. Wiley, New Jersey
- Wu X, Geng Zh, Shi Y, Pham N, Fomel S, Caumon G (2020) Building realistic structure models to train convolutional neural networks for seismic structural interpretation. *GEOPHYSICS* 85:WA27–WA39
- Ye SJ, Rabiller P (2000) A new tool for electrofacies analysis: Multi-Resolution Graph Based Clustering: SPWLA 41st Annual Logging Symposium, Dallas, Texas, USA, Jun 4–7
- Ye SJ, Rabiller PJ (2001) U.S. Patent No. 6,295,504. Washington, DC: U.S. Patent and Trademark Office

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.