A semi-automatically identifying lithofacies approach from wireline logs using fuzzy clustering

Duy Thong Kieu
Hanoi University of Mining and Geology
kieuduythong@humg.edu.vn

Quang Man Ha
PetroVietnam Exploration Production Corporation
manhq@pvep.com.vn

Viet Dung Bui
Vietnam Petroleum Institute
dungbv@vpi.pvn.vn

Huy Hien Doan
Vietnam Petroleum Institute
hiendh.epc@vpi.pvn.vn

Nguyen Binh Kieu
Petro Explorers Inc.
binhkieu@petroexplorers.com

SUMMARY

Lithofacies is important for reservoir evaluation. In this work, we present a workflow to define the lithofacies from wireline logs. The workflow includes three phases: in the first phase the boundaries are automatically defined from wireline logs by using the recurrence technique; in the second phase, we extract the data set from wireline logs within the boundaries and put it in a modified fuzzy c-means clustering process. The second phase results are analysed to identify the facies. Noting that our workflow can be automated if we have an available dataset including labels to build a prediction model in the third phase. We apply our workflow to a data set in Nam Con Son basin, Vietnam. The results are comparable with core data.

Key words: lithofacies, fuzzy c-means, supervised learning, Nam Con Son Basin.

INTRODUCTION

Lithology identification is essential in reservoir evaluation. The most direct approach to identifying lithofacies is by observation of cores. However, cores are usually limited because coring is high-cost, time-consuming, and missing. Therefore, lithofacies definition from wireline logs is needed. The lithofacies will determine both the values and shape of the wireline logs. Therefore, we exploit both the values and morphological features of the well logging data in the prediction lithofacies.

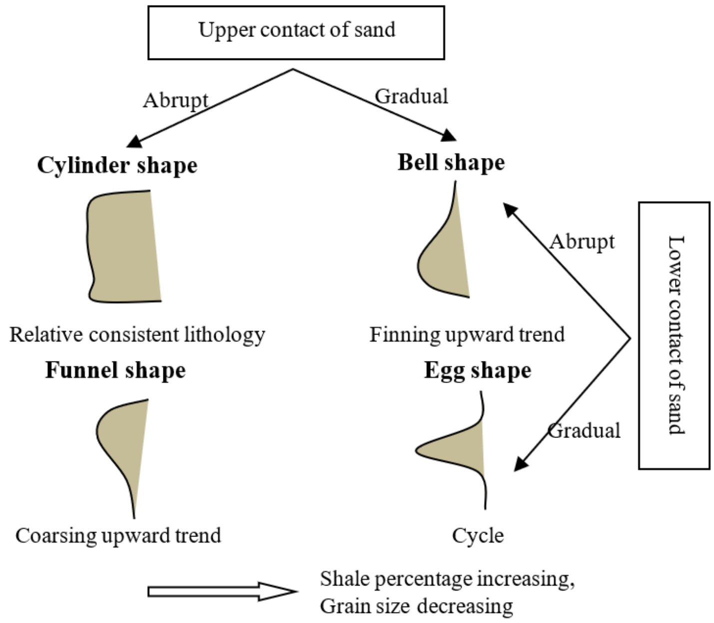

Figure 1. Basic log shapes (GR or SP) and possible facies used for interpretation (Modified from Serra (1986)).

The most useful log for lithological determination is Gamma Ray (GR) or Spontaneous Potential (SP). GR measures natural gamma-radiation emitted by sediments based on the concentrations of mainly potassium, uranium, and thorium. They are commonly taken as indicators of sediment composition since high GR values normally correspond to high clay content in clastic successions. Figure 1 shows several GR shapes that are used as a basis for identifying the sandstone depositional environments following as follows: Blocky (stable low GR values), Bell (upward increase in GR values), Funnel (upward decrease in GR values), Symmetrical (coarsening-upward overlain by fining-upward trend), Serrated (fluctuating GR values). Additionally, the other logs can provide complementary information about the facies as well which is defined as electrofacies (Serra, 1986). Thus, different from the work of Song et al. (2020) which used only SP logs for defining lithofacies, we exploit GR logs, resistivity (LLD log), and density (RHOB log).

We will present a workflow that enables us to semi-automatically define lithofacies, and then apply this workflow to a data set in Nam Con Son basin, Vietnam. It is worth noting that our workflow will automate the lithofacies definition if we have enough label data for building prediction models. In this work, we have not yet gotten label data, thus we utilise an unsupervised learning technique namely fuzzy c-means to identify “pseudo-lithofacies” (clusters) from well logs data.

METHODOLOGY

Workflow

The unsupervised machine learning technique is widely used for clustering, which does not require labelled data. The definition of cluster using the point-to-point relationship of the wireline logs like the work of Asante-Okyere et al. (2019) may lose morphological features of the logs that contain important lithological information (Ao et al., 2022; Song et al., 2020). In this work, we analyse both values and morphological features of well logging data. To extract the shape of the log segment, we treat the well logging data like time series.

Our workflow is presented in Figure 2. The workflow includes three phases. In phase 1, we applied the recurrence techniques (Zaitouny et al., 2020) to automatically define the boundaries (Figure 3). The wireline logs within boundaries will separate into log segments (Figure 4). The log segments will be put in process of phase 2 for clustering definition. Phase 3 will be the lithofacies definition from the clustering results.

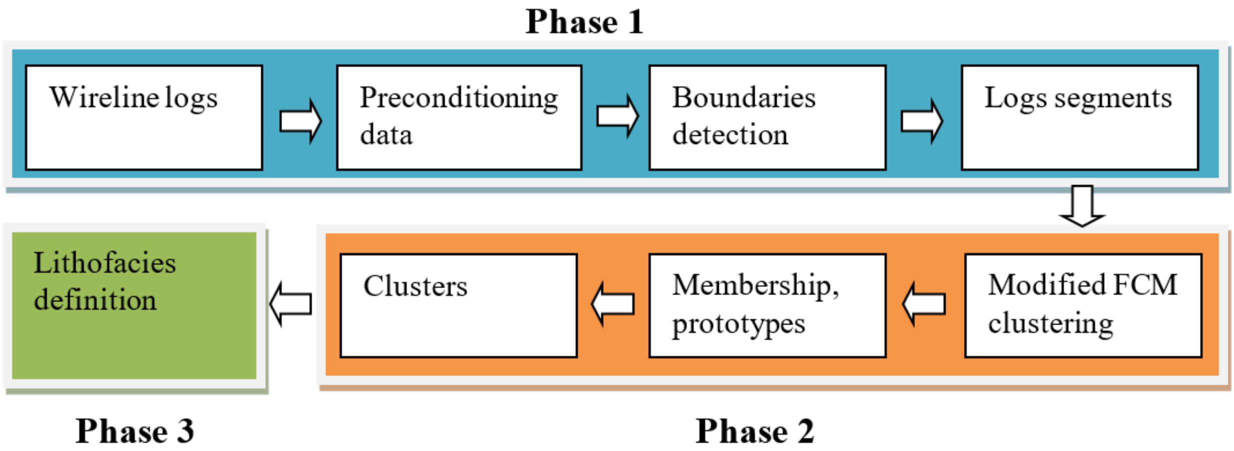

Figure 2. Workflow of lithofacies identification from wireline logs including three phases.

Clustering of the wireline logs

Fuzzy c-means clustering

In this research, we utilize the fuzzy c-means (FCM) clustering algorithm (Bezdek et al., 1984). This is an unsupervised learning technique that classifies N samples of dataset X[x_j, j=1, N] into C clusters by minimizing the following objective function:

$$\emptyset_{FCM} = \sum_{j=1}^{N}\sum_{k=1}^{C} u_{jk}^{m} D(x_j, v_k) \text{ subject to } \sum_{k=1}^{C} u_{jk} = 1, \quad (1)$$

where m is the fuzziness parameter, v_k is the centre of the kth cluster, u_{jk} is the membership degree, and d_{jk} is the distance of the jth data to the kth cluster, in Euclidian Norm

$$d_{jk} = \|x_j - v_k\|_2^2 \quad (2)$$

The objective function (equation 1) is minimized by an iteration process by updating a prototype v_k (equation 3) and membership degree (equation 4).

$$v_k = \frac{\sum_{j=1}^{N} u_{jk}^{m} x_j}{\sum_{j=1}^{N} u_{jk}^{m}}. \quad (3)$$

$$u_{jk} = \left\{\sum_{i=1}^{C} \left[\frac{d_{jk}^{2}}{d_{ik}^{2}}\right]^{\frac{1}{q-1}}\right\}^{-1} \quad (4)$$

It can be seen that the FCM objective function (1) is a weighted sum of errors as a data set X is replaced by the vector centre v. The centre uses (prototypes) and, each cluster might represent each rock unit or prototypes can assist to define lithofacies.

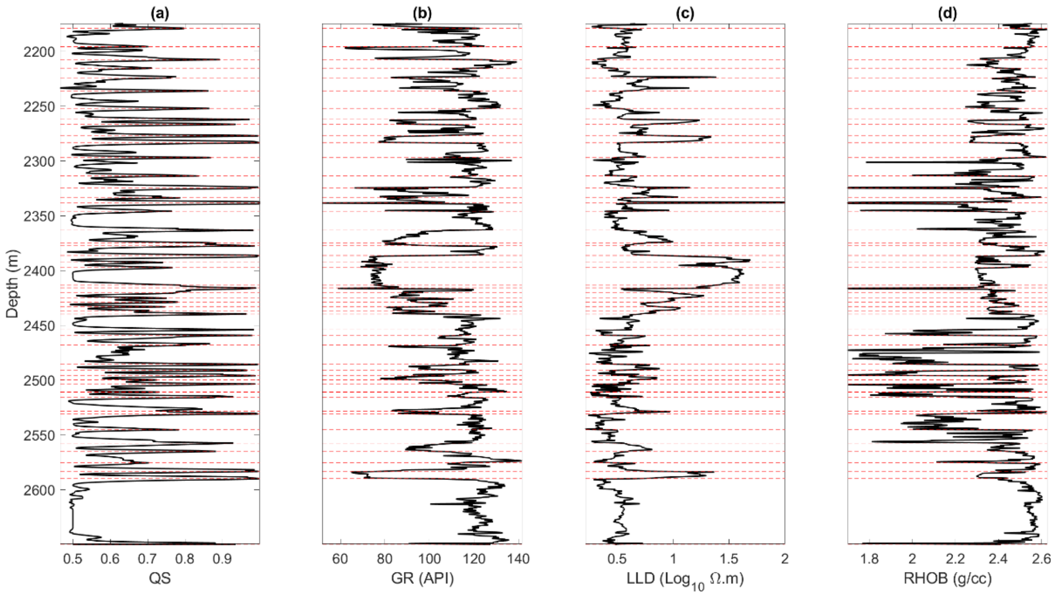

Figure 3. An example of automatic boundary identification (dashed red line) (a). The boundaries are defined by the quadrant scan (QS) method proposed by Zaitouny et al., (2020) using four logs GR (b), LLD (c), and RHOB (d). The peaks of QS indicate significant transitions corresponding to boundaries.

Modified fuzzy c-means clustering

The conventional input data for FCM clustering is discrete samples. However, using the point-to-point relationship of the wireline logs may loss of morphological features of the logs that contain important lithological information (Ao et al., 2022; Song et al., 2020). To keep the morphological features of the logs, we divide the logs into subsets based on the boundaries that are previously defined (Figure 4). Thus, the input data for clustering are log segments.

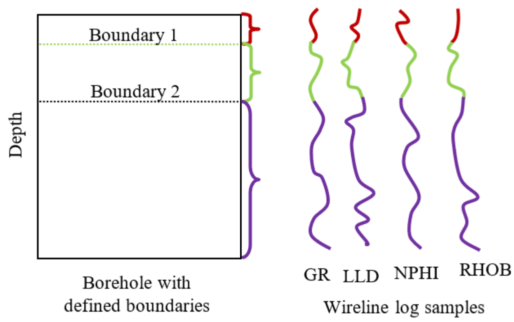

Figure 4. Scheme of sampling input data for modified FCM clustering. We can segment wireline logs into subsets of logs following defined boundaries.

We modified the conventional FCM clustering. Noting that, the major issue of FCM clustering is how we define prototypes or vector centre v and then how to calculate the error (equation 2). In this work, we initially generate a fixed length of prototypes and then compute errors, d_{jk}, between log segments and the prototype by using the dynamic time wrapping techniques (Figure 5). The prototypes and membership degree are updated like in the conventional process using equations 3 and 4, respectively.

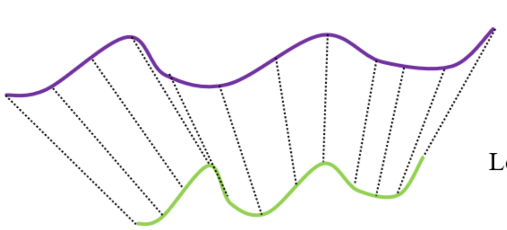

Figure 5. Scheme of dynamic time warping that measures the similarity between two log segments.

APPLICATION

Study area

The study area lies in the SE of the Nam Con Son basin, Vietnam. The Nam Con Son Basin (NCSB) is located in the southeastern continental shelf of Vietnam. Although there were likely tectonic influences affecting this area in the pre-Tertiary, the first rifting event forming the NCSB was occurred in the late Eocene to early Oligocene and associated with the onset North - South opening of the East Vietnam Sea. From seismic evidence, there appears to be a stratigraphic interval that could be interpreted as Eocene to Early Oligocene. They are marked by a major breakup unconformity at the top of Early Oligocene. This early rifting was followed by thermal sag associated with the post-rift phase. During the Early Miocene, SW propagation of East Vietnam Sea floor spreading to the NE of the basin and other regional tectonic forces impacted the NCSB eventually leading to a second rifting phase followed by a regional uplift event in late Middle Miocene as evidenced by the Mid-Miocene Unconformity (MMU). The post rift section during late Miocene to present is marked by thermal sag and the progradation of the Paleo-Mekong Delta into the basin.

Number clusters

In the FCM clustering technique, the number cluster is initially defined, the question is how many clusters are a good option? We tested several numbers of clusters and used Normalize classification entropy (NCE) index (Roubens, 1982) and Xie and Beni (XB) index (Xie and Beni, 1991) as cluster validity. The results show that 4 is the optimal number of clusters (Figure 6).

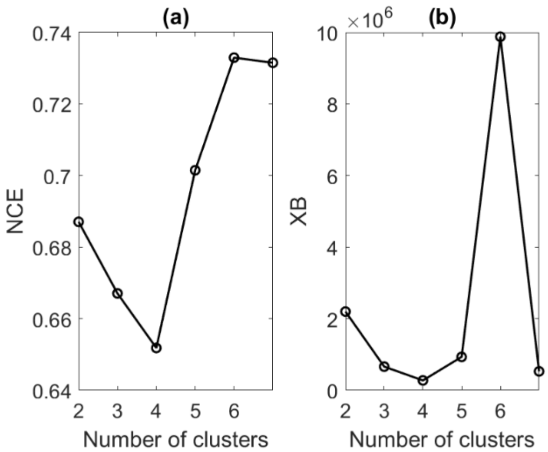

Figure 6. Various NCE and XB indexes with the number of clusters. The smallest NCE and XB value at 4 clusters indicates that the optimal option for the data set is 4 clusters.

Results

Both value and shape of the prototypes distinguish the geological characters (Figure 7 and Figure 8). Regarding the values of clusters. Cluster 1 has the smallest values of GR, but the highest values of resistivity and density in the middle. The histogram of the data shows that this cluster is separated from the others regarding GR and resistivity, but it is indistinguishable from the density of clusters 2 and 3. The values of the wireline log indicate that this cluster may relate to sandstone. Cluster 2 is defined by the second lowest GR values and the second highest values of resistivity. This cluster is likely shale and sandstone. The GR values of cluster 3 and 4 is almost the same, however, they are separate in resistivity values, particularly, they are opposite in density, while the density of cluster 3 is the smallest, the values of cluster 4 are the highest, thus cluster 3 and 4 is likely shale with a higher percentage of shale in cluster 4. The shape of prototypes may relate to the sedimentary environment (Figure 8).

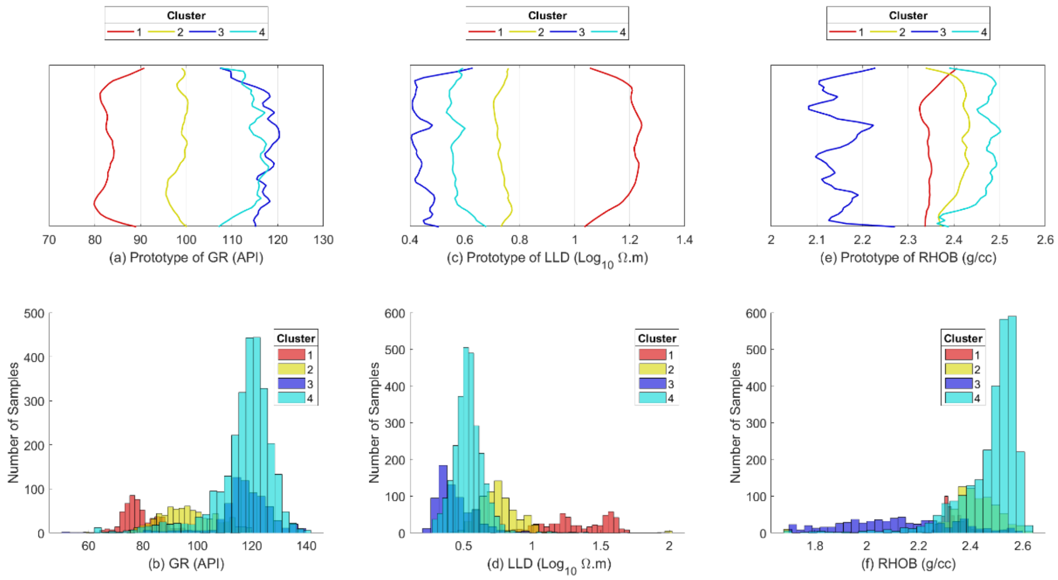

Figure 7. Prototype and histogram of the clusters. The clusters may relate to lithofacies

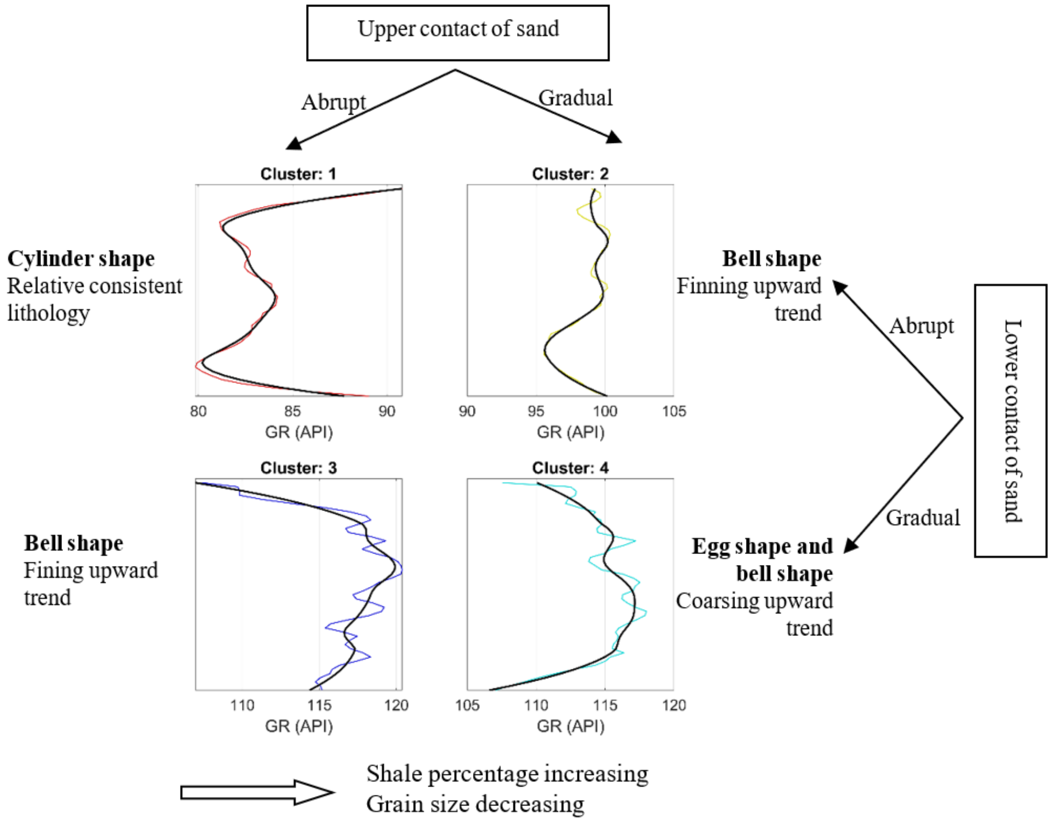

Figure 8. GR prototype of each cluster, the bold black lines show the smoothness of the prototypes. It can be seen that cluster 1 closely has a cylinder shape, cluster 2 has a bell shape, cluster 3 has a horizontal opposite bell shape, and cluster 4 looks like in between a bell and egg shape, but on the opposite side. According to the basic interpretation guide (Figure 1), we can divide the clusters corresponding to shapes and facies.

The core porosity of cluster 1 is the highest and straight-line trend (Figure 9c and e). This matches the cylinder shape of the prototype of cluster 1(Figure 8). There is a sharply different porosity between cluster 1 and cluster 4 (Figure 9e) and between cluster 2 and cluster 4 (Figure 9b) indicating the difference between environment, possibly between sand and shale. Coarsening upward of cluster 4 possibly relates to an increasing porosity trend (Figure 9b). The lowest values of porosity in clusters 3 and 4 show the percentage of shale (Figure 9d).

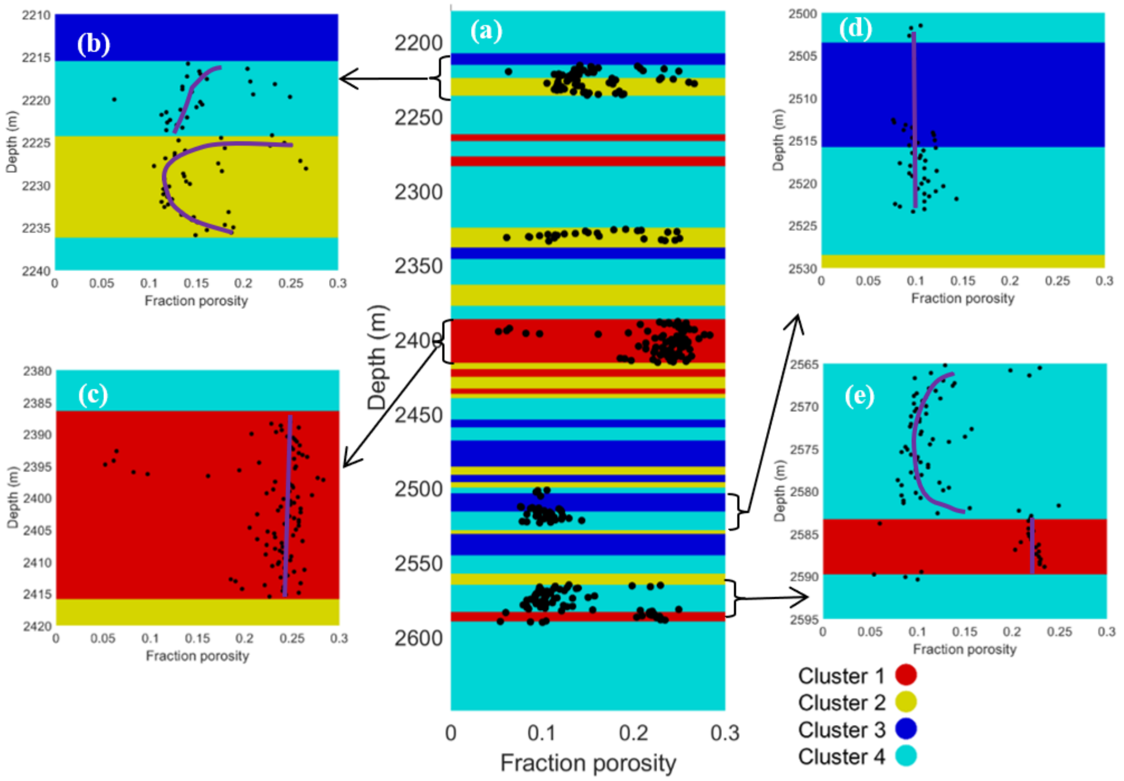

Figure 9. Clustering results superimposed by coring with porosity information (back dots) for the well (a). Zoom in segments of the well with core data, the bold violet lines show a trend of porosity of cores.

CONCLUSION

We present a workflow of the semi-automatic definitions of lithofacies. The workflow includes three phases: the definition of boundaries to separate wireline logs into log segments, then the log segments are the input of a modified fuzzy c-means clustering to define the prototypes of the cluster, according to the values and shape of the prototypes we can indicate the lithofacies. We apply the workflow to a well logging dataset in Nam Con Son Basin, Vietnam. Although the core data is not fully supporting our work, we have only the core porosity data, and the information of the core supports our statements of the facies. This work can be upgraded to the automatic process if we have got full datasets including labels for training.

ACKNOWLEDGMENTS

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.04-2021.23

REFERENCES

Ao, Y., Lu, W., Hou, Q., and Jiang, B., 2022, Sequence-to-sequence borehole formation property prediction via multi-task deep networks with sparse core calibration: Journal of Petroleum Science and Engineering, v. 208, p. 109637.

Asante-Okyere, S., Shen, C., Yevenyo Ziggah, Y., Rulegeya, M., and Zhu, X., 2019, A Novel Hybrid Technique of Integrating Gradient-Boosted Machine and Clustering Algorithms for Lithology Classification: Natural Resources Research.

Bezdek, J. C., Ehrlich, R., and Full, W., 1984, FCM: The fuzzy c-means clustering algorithm: Computers & Geosciences, v. 10, no. 2-3, p. 191-203.

Roubens, M., 1982, Fuzzy clustering algorithms and their cluster validity: European Journal of Operational Research, v. 10, no. 3, p. 294-301.

Serra, O., 1986, Fundamentals of Well-log Interpretation: The interpretation of logging data, Elsevier.

Song, S., Hou, J., Dou, L., Song, Z., and Sun, S., 2020, Geologist-level wireline log shape identification with recurrent neural networks: Computers & Geosciences, v. 134, p. 104313.

Xie, X. L., and Beni, G., 1991, A validity measure for fuzzy clustering: Pattern Analysis and Machine Intelligence, IEEE Transactions on, v. 13, no. 8, p. 841-847.

Zaitouny, A., Small, M., Hill, J., Emelyanova, I., and Clennell, M. B., 2020, Fast automatic detection of geological boundaries from multivariate log data using recurrence: Computers & Geosciences, v. 135, p. 104362.