An integrated Shannon Entropy and reference ideal method for the selection of enhanced oil recovery pilot areas based on an unsupervised machine learning algorithm

S. Mahdia Motahhari, Mehdi Rafizadeh, S. Mahmoud Reza Pishvaie, Mohammad Ahmadi

Oil & Gas Science and Technology – Rev. IFP Energies nouvelles **76**, 82 (2021)

https://doi.org/10.2516/ogst/2021061

Available online at:
ogst.ifpenergiesnouvelles.fr

REGULAR ARTICLE

OPEN ACCESS

An integrated Shannon Entropy and reference ideal method for the selection of enhanced oil recovery pilot areas based on an unsupervised machine learning algorithm

S. Mahdia Motahhari[1], Mehdi Rafizadeh[1,*], S. Mahmoud Reza Pishvaie[2], and Mohammad Ahmadi[1]

[1] Department of Petroleum Engineering, Amirkabir University of Technology, Tehran 1591634311, Iran
[2] Department of Chemical and Petroleum Engineering, Sharif University of Technology, Tehran 1458889694, Iran

Received: 12 January 2021 / Accepted: 10 November 2021

Abstract. Pilot-scale enhanced oil recovery in hydrocarbon field development is often implemented to reduce investment risk due to geological uncertainties. Selection of the pilot area is important, since the result will be extended to the full field. The main challenge in choosing a pilot region is the absence of a systematic and quantitative method. In this paper, we present a novel quantitative and systematic method composed of reservoir-geology and operational-economic criteria where a cluster analysis is utilized as an unsupervised machine learning method. A field of study will be subdivided into pilot candidate areas, and the optimized pilot size is calculated using the economic objective function. Subsequently, the corresponding Covariance (COV) matrix is computed for the simulated 3-D reservoir quality maps in the areas. The areas are optimally clustered to select the dominant cluster. The operational-economic criteria could be applied for decision making as well as the proximity of each area to the center of dominant cluster as a geological-reservoir criterion. Ultimately, the Shannon entropy weighting and the reference ideal method are applied to compute the pilot opportunity index in each area. The proposed method was employed for a pilot study on an oil field in south west Iran.

1 Introduction

Maximizing the Recovery Factor (RF) and economic profit are the main goals of Enhanced Oil Recovery (EOR) methods. Pilot design plays a vital role in the road map of EOR planning and risk reduction. Therefore, EOR pilots are conducted to reduce uncertainty in EOR performance [1–4]. In the last 50 years, a significant number of EOR field pilots has been studied and the gained experience has been used in the oil and gas industry [5]. The results of the implementation of EOR pilots may increase consciousness regarding hydrocarbon field behavior and EOR scenarios designed for entire hydrocarbon field.

Pilot area selection is one of the imperative factors in designing an EOR [6]. The objective of the pilot location studies is determining how to narrow down the pilot candidate areas from the field extent to only an optimum area of interest.

Three methodologies have been presented in the literature for pilot area selection. The first category presents some qualitative insights regarding the pilot area, such as the necessity of using predictive reservoir simulation models [4, 7–10]. These studies do not provide any quantitative methods to select a pilot area. In the second category, the reservoir area is filtered to select the pilot area according to some merit parameters [11–13]. These studies are based on reservoir past performance. In the final category, the reservoir is divided into a number of areas and the pilot area is selected on the basis of a semi-quantitative approach. In a previous study [14], a hydrocarbon field was divided into six square kilometers parts and each part labeled as a locator. Consequently, locators were classified based on permeability patterns and the number of water-invaded zones. Permeability patterns were described as the combination of the average permeability of the different layers. The average value was marked per layer as shown in Figure A1 (Appendix A). The locators were chosen as pilot areas from a class that has a maximum number of dominated pattern.

According to other studies [15, 16] that have been carried out, the field was divided into six areas as shown in Figure A2 (Appendix A). The size of the areas can be arbitrary. A semi-quantitative approach was chosen to rank each of the six areas for pilot implementation. A list of selection criteria, including the qualitative ranking (low, medium, and high) in each area is shown in Figure A3 (Appendix A). Although this approach tries to determine the part with the best properties, each selected part is not representative of the whole field. Moreover, predefined

* Corresponding author: mehdi@aut.ac.ir

patterns such as permeability patterns were applied to identify dominated pattern in this category.

In addition to the methodological approaches, criteria based on reservoir-geology and operational-economic issues could be considered. According to the geological criteria of the reservoir, the pilot area should have alike behavior to the reservoir to be an appropriate representative and achieve the EOR goals [4, 12, 14]. However, an area with moderate to good reservoir properties is another selection criterion. For example, having remaining oil saturation in the range of 0.2–0.5 could lead to an appropriate pilot area [11]. According to other studies [7, 8], high remaining oil saturation has been applied to locate pilot areas. In addition, a lower ratio of vertical to horizontal permeability and higher lateral permeability continuity have also been a basis for pilot area selection [15, 16].

Operational-economic criteria refer to uncertainty and economic aspects of pilot area selection. An area with a higher accuracy in analysis of pilot production performance is suitable from an operational point of view. Moreover, a lower cost of pilot implementation is desirable too. The level of knowledge, the interaction of pilot and production area, the distance between a candidate area and surface facilities and the number of existing applicable wells for running a pilot are all important operational-economic criteria. For example, a pilot area could be selected based on the quality and quantity of the available petrophysical data [4, 14, 15]. In addition, less interaction between the pilot and production areas could lead to a more reliable interpretation of the gathered data [4, 11–13]. Besides, the closeness of the pilot to existing facilities could lead to economic aspects [12–14]. Finally, existing wells may be used to reduce pilot implementation costs [13]. The pilot operation will assist in receiving relevant operational and technical experience and establishing required quality control [17].

It is therefore apparent that having a systematic and quantitative approach is necessary to choose the optimum pilot area. Such an approach should consist of a better representative of reservoir features along with operational-economic criteria. In this paper, a methodology is presented to select the optimal size and location of the pilot for an EOR project. To do so, a pilot opportunity index is introduced. The proposed index is defined based on integration of reservoir quality maps (from reservoir simulations), fuzzy clustering, Shannon entropy, Analytical Hierarchy Process (AHP) and the Reference Ideal Method (RIM) as a multi-criteria decision-making method. Finally, the presented method is utilized to a real reservoir case and the optimum pilot area is thereby determined.

2 Methodology

The three key steps for choosing the location and size of a pilot established are:

1. Pilot candidate areas specification;
2. Reservoir representativeness quantification and calculating its value for each pilot candidate area;
3. Pilot opportunity index calculation and selecting the optimum pilot region.

Figure 1 illustrates the workflow of the planned scheme.

2.1 Pilot candidate areas

In the model of hydrocarbon reservoir, the field is segmented into pilot candidate regions of an equal size. Then, the five-spot pattern are utilized to locate wells in each region.

Afterwards, a sensitivity analysis is performed for the Net Present Value (NPV) in regard to the production and injection wells distance, which reflects the size of the pilot. The NPV is the most commonly used economic objective function and is shown in equation (1):

$$\mathrm{NPV}(r,\ T) = R(t_0) + \int_{t_1}^{T} \left[P_0 \times q_0(t) - C_{\mathrm{w}}^{\mathrm{prod}} \times q_{\mathrm{w}}^{\mathrm{prod}}(t) - C_{\mathrm{w}}^{\mathrm{inj}} \times q_{\mathrm{w}}^{\mathrm{inj}}(t) - C_0 \times q_0(t)\right](1+r)^{-t}\mathrm{d}t, \quad (1)$$

$$R(t_0) = C_{\mathrm{facility}} + n_{\mathrm{well}}\, C_{\mathrm{well}}, \quad (2)$$

where t_1 is the start time of production, T is the overall time of econometry, P_{o} is the unit price of oil, q_{o} is the oil production rate, $C_{\mathrm{w}}^{\mathrm{prod}}$ is the unit cost of handling the produced water, $C_{\mathrm{q}}^{\mathrm{prod}}$ is the water production rate, $C_{\mathrm{w}}^{\mathrm{inj}}$ is the unit cost of injecting water, $q_{\mathrm{w}}^{\mathrm{inj}}$ is the water injection rate, C_{o} is the unit operating cost per barrel of produced oil and r is the discount rate. $R(t_0)$ is the initial capital expenditure, as shown in equation (2), where C_{facility} is the cost of facility installation, n_{well} is the entire number of injection and production wells and C_{well} is the drilling cost of a well. Consequently, the optimal pilot size has the highest NPV. Finally, the size and number of pilot candidates are determined.

2.2 Impression of reservoir representativeness

Each candidate region has a reservoir dynamic behavior. Considering all areas, it is possible to classify the behavior. The class which has the highest number of areas is the representative of reservoir behavior. The behavior of each area is characterized by two arrays of cell RF values and the covariance matrix of the residual oil volume in cells positioned in each candidate region.

2.2.1 Quality maps

The various static and dynamic parameters could be combined together to form the quality maps. Due to dynamic parameters such as oil saturation, the quality maps change during the production period [18, 19]. The quality map is applied to solve decision making problems as an auxiliary analysis with parameters like cumulative oil production [20].

The 3-D time-dependent quality maps of residual oil volume is computed by using equation (3) to aggregate various parameters:

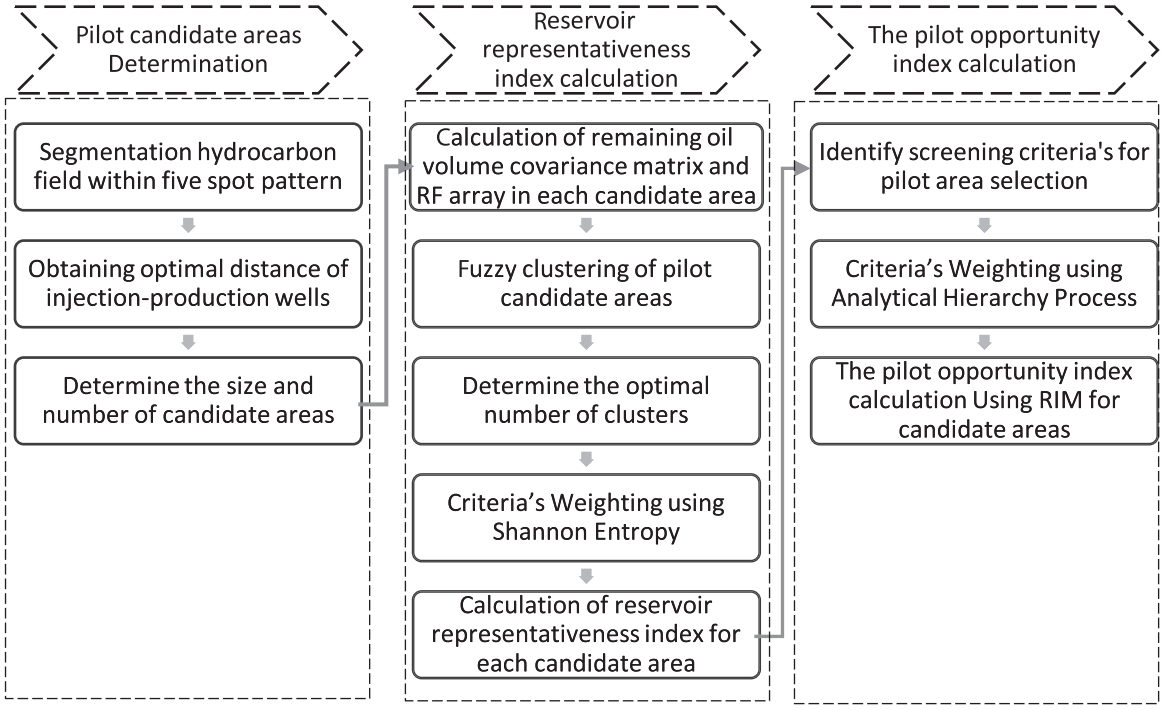

Fig. 1. Workflow of the planned scheme.

$$V_o(i,j,k,t) = A_{i,j} \times h_k \times \emptyset_{i,j,k} \times \mathrm{So}_{i,j,k,t}, \tag{3}$$

where $A_{i,j}$, h_k, $\emptyset_{i,j,k}$ and $\mathrm{So}_{i,j,k,t}$ are the surface area, the height, the porosity and the simulated oil saturation of cell (i, j, k) in the given time period, respectively. Equation (4) is applied to calculate the RF in each cell based on these quality maps:

$$\mathrm{RF}(i,j,k) = [V_o(i,j,k,t_s) - V_o(i,j,k,t_e)] \div V_o(i,j,k,t_s), \tag{4}$$

where t_s and t_e are the beginning and termination times of the simulation, correspondingly.

Equation (5) is applied to illustrate the covariance matrix of the residual oil volume for a typical candidate region. For instance, $\mathrm{Vo}_{\mathrm{cell}\#,1}$ is a set of annual residual oil volume sequence from t_s to t_e in the first cell of the area:

See the equation (5) bottom of the page

2.2.2 Clustering

The number of clusters of reservoir behaviors could be specified by Fuzzy clustering. Clustering is a subset of unsupervised machine learning [21] and data mining [22] methods. It is used to assess the flow field in hydrocarbon reservoirs [23–25].

The similar objects are placed in a same group using clustering analysis. Equation (6) [26] is utilized to define the objective of fuzzy clustering. The number of clusters and cluster center are signified by c and c_i. A multi-dimensional space is denoted by X that includes n candidate areas. For the jth candidate region in the cluster i, the membership degree is indicated by u_{ij}.

Furthermore, v and d_{ij} are identified as the fuzzifier and the Euclidean distance of the jth candidate region from the c_i.

$$J(X, C, U) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{v} d_{ij}^{2}, \tag{6}$$

$$\mathrm{COV} = \begin{bmatrix} \mathrm{COV}(\mathrm{Vo}_{\mathrm{cell}\#1}, \mathrm{Vo}_{\mathrm{cell}\#1}) & \mathrm{COV}(\mathrm{Vo}_{\mathrm{cell}\#1}, \mathrm{Vo}_{\mathrm{cell}\#2}) \cdots & \mathrm{COV}(\mathrm{Vo}_{\mathrm{cell}\#1}, \mathrm{Vo}_{\mathrm{cell}\#e}) \\ \mathrm{COV}(\mathrm{Vo}_{\mathrm{cell}\#2}, \mathrm{Vo}_{\mathrm{cell}\#1}) & \mathrm{COV}(\mathrm{Vo}_{\mathrm{cell}\#2}, \mathrm{Vo}_{\mathrm{cell}\#2}) \cdots & \mathrm{COV}(\mathrm{Vo}_{\mathrm{cell}\#2}, \mathrm{Vo}_{\mathrm{cell}\#e}) \\ \vdots & \vdots \quad \ddots & \vdots \\ \mathrm{COV}(\mathrm{Vo}_{\mathrm{cell}\#e}, \mathrm{Vo}_{\mathrm{cell}\#1}) & \mathrm{COV}(\mathrm{Vo}_{\mathrm{cell}\#e}, \mathrm{Vo}_{\mathrm{cell}\#2}) \cdots & \mathrm{COV}(\mathrm{Vo}_{\mathrm{cell}\#e}, \mathrm{Vo}_{\mathrm{cell}\#e}) \end{bmatrix}. \tag{5}$$

$$(\forall\, i; 1 \le i \le c) \quad \text{and} \quad \left(0 \le u_{ij} \le 1\right) \quad \text{and} \quad \left(\forall\, j;\, 1 \le j \le n : \sum_{i=1}^{c} u_{ij} = 1\right),$$

$$u_{ij} = \frac{d_{ij}^{\frac{2}{1-v}}}{\sum_{k=1}^{c} d_{kj}^{\frac{2}{1-v}}}, \tag{7}$$

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{v} x_j}{\sum_{j=1}^{n} u_{ij}^{v}}. \tag{8}$$

The optimal number of clusters is determined using two cluster validity indices: (1) XB (Xie and Beni) [27] and (2) fuzzy-silhouette [28] as illustrated in equations (9) and (10)–(14), correspondingly.

The higher square distance between cluster centers (a_j) is more favorable, as well as lower mean square distance between a candidate area and its cluster center (b_j). Therefore, the lower and higher values of the *XB* and fuzzy-silhouette are the result of high-quality clustering:

$$XB = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^2 d_{ij}^2}{n \times \min_{\tilde{j}\neq j} d_{\tilde{j}j}^2}, \tag{9}$$

$$a_j = \min\left\{\frac{\sum_{k=1}^{n} \text{intra}_i(j,k) d_{jk}}{\sum_{k=1}^{n} \text{inter}_{rs}(j,k)};\quad \sum_{k=1}^{n} \text{intra}_i(j,k) > 0, 1 \le i \le c\right\}, \tag{10}$$

$$b_j = \min\left\{\frac{\sum_{k=1}^{n} \text{inter}_{rs}(j,k) d_{jk}}{\sum_{k=1}^{n} \text{inter}_{rs}(j,k)};\quad \sum_{k=1}^{n} \text{inter}_{rs}(j,k) > 0, 1 \le r < s \le c\right\}, \tag{11}$$

$$\text{intra}_i\left(d_{jk}\right) = \left(u_{ij} \bigwedge u_{ik}\right), \tag{12}$$

$$\text{inter}_{rs}\left(d_{jk}\right) = \left(u_{rj} \bigwedge u_{sk}\right) \vee \left(u_{sj} \bigwedge u_{rk}\right), \tag{13}$$

$$\text{Sil}(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{b_j - a_j}{\max\left\{a_j, b_j\right\}}. \tag{14}$$

2.2.3 Reservoir Similarity Index

For a candidate area, the intensity of the reservoir representativeness is indicated by the Reservoir Similarity Index (RSI). Therefore, the RSI is composed of the two normalized distances: 1) *d*(RF array, RF-center) and 2) *d*(COV array, COV-center). The RF-center and COV-center are the centers of the clusters that have the maximum size after conducting clustering on the RF array and the covariance matrix of the residual oil volume, correspondingly. The RF array and COV array are the Recovery Factor and Covariance matrix arrays of the residual oil volume of the candidate area. The weights of the two normalized distances are w_1 and w_2. The RSI is calculated as the sum of the product of weight and the normalized value for multiplicative inverse of the two criteria using equation (15):

$$\text{RSI} = w_1 \times \frac{1}{d(\text{RF array, RF} - \text{center})} + w_2 \times \frac{1}{d(\text{COV array, COV} - \text{center})}. \tag{15}$$

The Shannon entropy technique is utilized to determine the w_1 and w_2.

2.2.4 Shannon entropy

The information content of an event as the concept of entropy was introduced by Shannon in information theory [29]. Shannon entropy is used to optimize data acquisition of the hydrocarbon fields [30]. Furthermore, the Shannon entropy can quantity the relative importance of criteria in a multi-criteria problem [31]. The more important criteria have greater dispersion and their effects are higher on the RSI value. Considering m alternatives and n criteria, a decision matrix **S** of $m \times n$ elements is shown in equation (16), where s_{ij} is the *j*th criterion value for the *i*th alternative (candidate area).

Equation (17) is applied to normalize the decision matrix **S**. Then, equation (18) is utilized to compute the Shannon entropy. Afterward, equation (19) is employed to calculate the weight of each criterion:

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{bmatrix}, \tag{16}$$

$$\bar{s}_{ij} = s_{ij} \Big/ \sum_{i=1}^{m} s_{ij} \quad (i = 1, ..., m; j = 1, ..., n), \tag{17}$$

$$E_j = -\frac{1}{\ln m}\sum_{i=1}^{m} \bar{s}_{ij}, \quad \text{j} = 1, ..., n, \tag{18}$$

$$w_j = \frac{1 - E_j}{\sum_{j=1}^{n}\left(1 - E_j\right)}. \tag{19}$$

2.3 Pilot opportunity index

The attractiveness of pilot implementation in a candidate area is indicated by the Pilot Opportunity Index (POI). Two significant features for POI computations are reservoir

geological and operational-economic. Reservoir geological factors correspond to the reservoir representativeness intensity. Operational-economic factors refer to the reliability of the pilot performance results and the implementation cost of the pilot.

Therefore, the POI is based on the RSI, the level of knowledge, the interference between internal and external production of a candidate area, the distance from the surface facilities to candidate area, and the number of existing applicable wells for running a pilot in an area. Hence, the RIM in multi-criteria decision-making methods is utilized to quantify the POI. Moreover, criteria weights are calculated based on the Analytical Hierarchy Process (AHP). The higher the POI is, the more desirable the pilot implementation in an area is. The pilot area is an area with the highest POI value.

2.3.1 Decision criteria

The criteria to evaluate the POI for each candidate area include:

- RSI: In a candidate area, this criterion shows the value of the reservoir representativeness. Higher values of this criterion are preferable.
- Interference between internal and external production of a candidate area: Pilot production performance could be affected considerably by nearby existing production wells. Minimizing this interference lead to a more accurate pilot production performance analysis. The drainage radius, the distance from a well where its normal pressure gradient approaches zero (*i.e.* zero fluid flux), is applied to quantify this issue [32]. In other words, there would be pressure interference in case the distance between two wells is less than two times the drainage radius. Therefore, to determine the interaction between the candidate area and the existing wells, two criteria are established, namely, the number of existing wells around a candidate area with a distance less than two times the drainage area, known as "interfered wells", and the average distance between these wells and a candidate area. A lower number of interfered wells and higher values of the average distance are more favorable.
- Level of knowledge: A higher quality and quantity of available data in a candidate area lead to a lower uncertainty in pilot design and evaluation. The variogram range, the distance limit beyond which the data are no longer spatially correlated, is applied to quantify this criterion [33]. The properties in cells that contain an existing well are known while the properties of other cells can be estimated by the known cell value. The estimation is possible in the case of the spatial distance between known and unknown cells being smaller than the variogram range. Furthermore, the error of estimation decreases as the distance decreases. Therefore, to determine the level of knowledge in a candidate area, two criteria are established, namely, the total number of existing wells inside and outside an area with a distance less than the variogram range to an area, known as "adjacent wells", and the average distance between these wells and a candidate area. A higher number of total wells and lower values of the average distance are more favorable.
- Distance between a candidate area and surface facilities: the lower the distance, the lower the cost. The closest route from areas to the surface facilities is the most desirable path.
- Number of existing applicable wells for running a pilot in an area: This is the number of existing wells in a candidate area that matches the five-spot pattern in terms of location. Higher values of this criterion are more preferable due to less capital expenditure to drill new wells that fit the pattern in an area.

Table 1. Intensity and explanation for comparing criteria in a matrix of pairwise comparisons [34].

Intensity of importance	Definition
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme importance

2.3.2 AHP

Analytical Hierarchy Process (AHP) combines mathematics and intuition to express the comparative importance of each criterion in a rational and consistent approach [34, 35]. Table 1 is applied to make the pairwise comparisons matrix among criteria.

In the pairwise comparisons matrix, i rows are compared with the j column for n number of criteria such that the main diagonal is equal to one. The results of the pairwise comparison on criteria are summarized in an $(n \times n)$ matrix in which elements w_{ij} $(i, j = 1, 2, \ldots n)$ are shown by equation (20):

$$w = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}. \tag{20}$$

The pairwise comparisons matrix is normalized by dividing the value of each element with the sum of the corresponding column. After normalization, the relative weight of each criteria is calculated by computing the average of each row.

The accuracy of the AHP output is related with the consistency of the pairwise comparison judgment. A Consistency Ratio (CR) is calculated by dividing the Consistency Index (CI) for the judgment by the Random Consistency Index (RI) for the corresponding random matrix using equations (21) and (22). The RI is in accordance with the degree of consistency, as shown in Table 2 for a different number of criteria. Moreover, the Cl is based on the largest eigenvalue (λ_{max}) and the number of criteria

Table 2. Random consistency indices (RI) for n criteria [34].

n	RI	n	RI	n	RI	n	RI	n	RI
1	0	4	0.89	7	1.35	10	1.49	13	1.56
2	0	5	1.11	8	1.4	11	1.51	14	1.57
3	0.52	6	1.25	9	1.45	12	1.54	15	1.58

(n), as shown in equation (21). The weighted sum vector is calculated with a multiplication of the pairwise comparisons matrix of criteria in the vector of relative magnitudes. Then, the consistency vector is computed by dividing the magnitudes of the weighted sum vector of criteria by the vector of relative magnitudes. The largest eigenvalue $(\lambda_{\max})$ is calculated by averaging elements of this vector:

$$\mathrm{CI} = \frac{\lambda_{\max} - n}{n-1}, \tag{21}$$

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}}. \tag{22}$$

The CR above 0.1% or 10% indicates that the pairwise comparisons should be revised, while smaller numbers mean comparison are more consistent.

2.3.3 RIM

Multi Criteria Decision Making (MCDM) is applied to solve selection issues including multiple criteria and alternatives for petroleum field development problems [36]. The RIM [37] multi-criteria decision-making technique is applied to prioritize the alternatives (candidate areas) with respect to the criteria outlined in Section 2.3.1.

Equation (24) is applied to normalize the decision matrix **S**, as illustrated by equation (23). The range of values and ideal values for each criterion are shown by $[A, B]$ and $[C, D]$, respectively. Equation (25) is utilized to compute the weighted normalized matrix. Equation (26) is applied to calculate the distances between the ideal alternatives and a candidate area. In conclusion, equation (27) is employed to compute the score of each area:

See the equation (23) bottom of the page

$$f(s,[A,B],[C,D]) = \begin{cases} 1 & \text{if } s \in [C,D] \\ 1 - \frac{d_{\min}(s,[C,D])}{|A-C|} & \text{if } s \in [A,C] \wedge A \neq C \\ 1 - \frac{d_{\min}(s,[C,D])}{|D-B|} & \text{if } s \in [D,B] \wedge D \neq B \end{cases} \tag{24}$$

$$Y = \begin{bmatrix} f(s_{11},[A,B]_1,[C,D]_1) & f(s_{12},[A,B]_2,[C,D]_2) \cdots & f(s_{1n},[A,B]_n,[C,D]_n) \\ f(s_{21},[A,B]_1,[C,D]_1) & f(s_{22},[A,B]_2,[C,D]_2) \cdots & f(s_{2n},[A,B]_n,[C,D]_n) \\ \vdots & \vdots \quad \ddots & \vdots \\ f(s_{m1},[A,B]_1,[C,D]_1) & f(s_{m2},[A,B]_2,[C,D]_2) \cdots & f(s_{mn},[A,B]_n,[C,D]_n) \end{bmatrix}, \tag{23}$$

$$\acute{Y} = Y \otimes W = \begin{bmatrix} y_{11}.w_1 & y_{12}.w_2 \cdots & y_{1n}.w_n \\ y_{21}.w_1 & y_{22}.w_2 \cdots & y_{2n}.w_n \\ \vdots & \vdots \quad \ddots & \vdots \\ y_{m1}.w_1 & y_{m2}.w_2 \cdots & y_{mn}.w_n \end{bmatrix}, \tag{25}$$

$$I_i^+ = \sqrt{\sum_{j=1}^{n} \left(\acute{y}_{ij} - w_j\right)^2},\ I_i^- = \sqrt{\sum_{j=1}^{n} \left(\acute{y}_{ij}\right)^2} \quad (i = 1,...,m; \mathrm{j} = 1,...,n), \tag{26}$$

$$R_i = \frac{I_i^-}{I_i^+ + I_i^-} \quad (0 < R_i < 1; i = 1,...,m). \tag{27}$$

3 Results and discussion on a case study

The studied field has been producing since 2006 and 31 wells have been drilled so far. The field's oil in place is around 4 billion standard barrels. The number of grids in the three main directions x, y and z is $88 \times 275 \times 22$, respectively. The field has 15 different rock types. Variogram model and its parameters are calculated for the porosity as shown in Figure 2. The variogram range is equal to 1500 m. Furthermore, drainage radius is 320 m according to interpretation of well test in the field.

A history-matched model is applied to predict production scenarios by reservoir numerical simulation. In case of enhancing the history match, it is proposed to use the integration of multi-start simulated annealing with genetic algorithm [38].

Waterflooding is planned in the field to increase production from 20,000 to 110,000 barrels per day. Moreover, the maximum allowable bottom hole pressure in injection wells and the minimum flowing bottom hole pressure in production wells are 5880 and 3290 psia, correspondingly.

The pilot project is required prior to the design and field implementation of EOR. Hence, it is necessary to determine

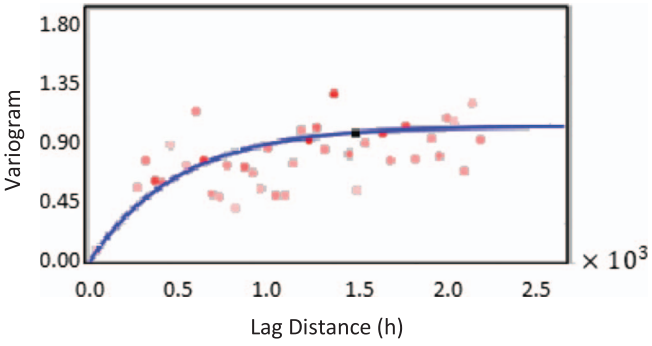

Fig. 2. The variogram model of porosity values.

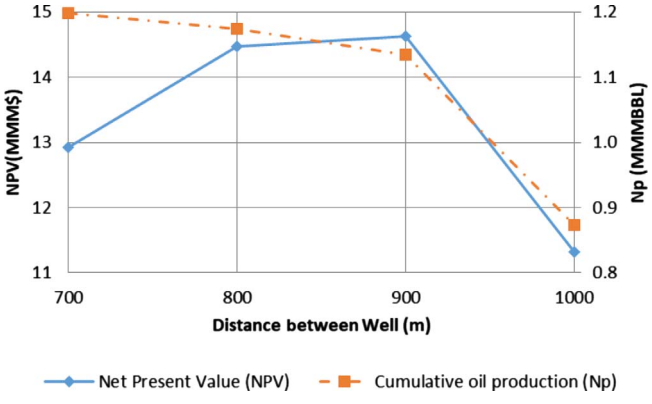

Fig. 3. Determining pilot size by sensitivity analysis.

the size and location of the pilot area. The method established in this research was utilized to choose the best pilot region for this field.

Initially, the reservoir model is subdivided into equal size pilot candidate areas. In each area, production and injection wells are located according to a five-spot well pattern. The sensitivity analysis on the NPV is utilized to determine the optimum pilot dimensions. Data to compute the NPV via equations (1) and (2) are illustrated in Table B1 (Appendix B). The result of varying distance between production and injection wells, which reflects the pilot size on the NPV and cumulative oil production, is illustrated in Figure 3. There is the highest of NPV in case of well distance is 900 m.

The positions of the injection and production wells, along with existing wells in the reservoir, are illustrated in Figure 4 according to this distance. Moreover, reservoir segmentation resulted in 72 candidate regions, as illustrated in Figure A4 (Appendix A). Therefore, the number of grids for a candidate area in the directions x, y and z is $9 \times 9 \times 22$, respectively. If an area does not contain all the 22 reservoir layers, it is removed. Therefore, six regions were removed. Consequently, there are 66 valid candidate regions in the case study reservoir.

Running of the production and injection well pattern in the simulator leads to oil saturation computation. With equation (3), the product of the pore volume in oil

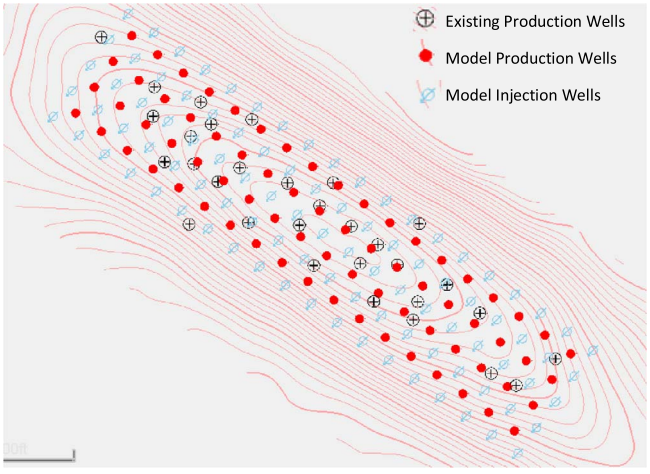

Fig. 4. Five-spot patterns in the case study reservoir.

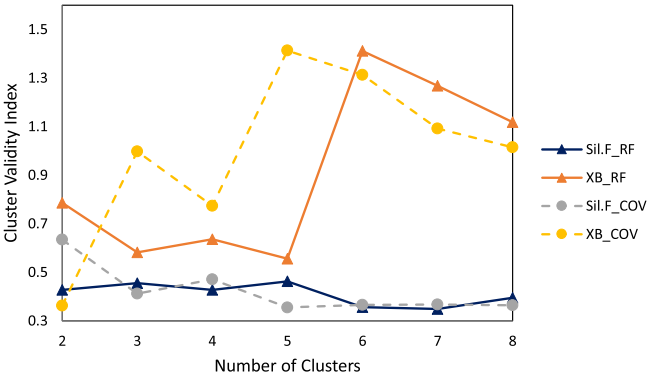

Fig. 5. Determining the optimal number of clusters with cluster validity indexes.

saturation resulted in the quality maps of the residual oil volume as shown in Figure A5 (Appendix A). By using equations (4) and (5), 1782-member array of the recovery factor and 1588653-member array of the covariance between cells are allocated to an area. By using equations (6)–(8), the candidate areas are clustered. The values of the hyper-parameter (fuzzifier v) and convergence criterion are 1.5 and 1e–9 in fuzzy clustering, respectively.

By using equations (9)–(14), fuzzy clustering of the candidate areas are validated as shown in Figure 5. According to Figure 5, if the number of clusters is considered to be two and five, the clustering of the covariance between cells (Sil.F_COV and XB_COV) and the recovery factor (Sil.F_RF and XB_RF) arrays are optimal, accordingly.

According to the preceding graph, clustering of candidate regions leads to two and five clusters for covariance matrixes and recovery factor arrays, respectively. Figure 6 shows membership of clusters based on the recovery factor arrays. Moreover, Figure 7 shows membership of clusters based on the covariance matrixes. Each point on these two charts corresponds to a candida region. Additionally, the * symbol shows centers of the clusters. Also,

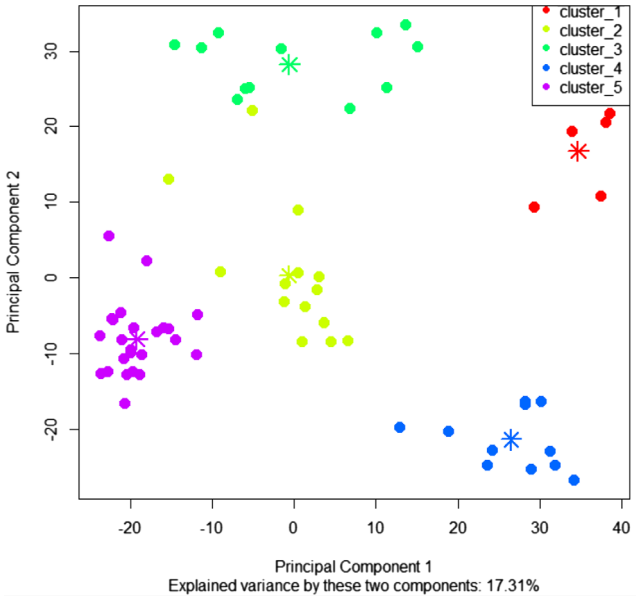

Fig. 6. Clustering result using the color-coding graph for the RF arrays.

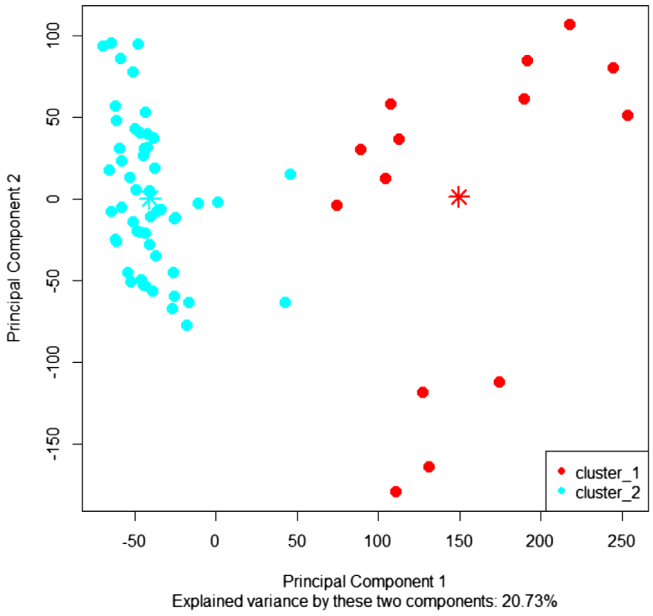

Fig. 7. Clustering result using the color-coding graph for the COV arrays.

Tables B2 and B3 (Appendix B) show results of clustering such as cluster size. The dominant clusters are cluster 5 and 2 for recovery factor and the covariance matrixes, respectively.

Figure A6 (Appendix A) shows the distribution map of the clusters in the field with reference to Figures 6 and 7. The candidate areas in the same cluster are close to each other with respect to location. Therefore, there is a spatial relationship between the reservoir behavior, which is characterized by the covariance of the residual oil volume and the RF, and the position in the field.

The covariance-center and RF-center are utilized to show the cluster centers. Averaging all regions in dominant clusters leads to determination of these centers. Afterward, the distance from these two centers to candidate areas is computed as displayed in Table B4 (Appendix B).

Prior to computing the RSI value of each region, the weights of two criteria: (1) the covariance-center distance from an area and (2) the RF-center distance from an area are determined. Using equations (16)–(19) besides the criteria values in Table B4 (Appendix B), the weight of the first criterion is 0.313 while the weight of the second criterion is 0.686. The sum of product of weight and normalized value for multiplicative inverse of the two criteria has resulted in a RSI value that is shown in Figure A7 (Appendix A). For instance, the candidate area numbers 29, 34 and 33 have the highest reservoir representativeness intensity with RSI values equal to 0.99, 0.92 and 0.82, respectively.

To calculate the POI of each candidate area, the criteria of RSI, the distance from the surface facilities to candidate area, the number of interfered wells, the mean distance from these wells to candidate area, number of existing applicable wells for running a pilot, number of adjacent wells, and finally the mean distance from these wells to candidate area have been considered with the weight factors of 0.33, 0.145, 0.145, 0.055, 0.055, 0.09 and 0.18, respectively, according to the AHP weighting method. The pairwise comparison matrix (Eq. (20)) and weights of the criteria are shown in Table B5 (Appendix B). In this method, every pair of criteria is compared with each other with respect to the expert judgments. In each comparison this question must be answered: which criterion is more important and how much? Table 1 is applied to express the rating preferences between each pair of criteria. For instance, experience and judgment strongly favor the RSI over the number of adjacent wells with importance intensity value equal to 5 as shown in Table B5 (Appendix B). Furthermore, the consistency ratio is computed to verify the reliability between the pairwise judgements using equations (21) and (22). The values of the largest eigenvalue (λ_{max}), the random consistency index (see Tab. 2), the Consistency Ratio (CR), and the Consistency Index (CI) are 7.06, 1.35, 0.008, and 0.011, respectively. The Consistency Ratio of 0.011 shows that the pairwise comparisons are consistent.

The POI is computed for each region utilizing the Reference ideal method via equations (23)– (27) and the aforementioned criteria values, as shown in Figure A7 (Appendix A). For example, the values of criteria that contributed to the POI for candidate area numbers 8, 45, 34 and 29 are shown in Table 3.

The values of calculated POI in these regions are equal to 0.73, 0.66, 0.64 and 0.59, accordingly. Therefore, Pilot execution in these regions is relatively desirable. In Table 3, the average distance between interfered wells and a candidate area is indicated by the expression of NA when the number of interfered wells is equal to zero.

According to Table 3 and Figure A7 (Appendix A), for example, although candidate area number 29 has the highest amount of RSI value and its distance between a candidate area and surface facilities is relatively low in

Table 3. Weight factors of criteria and their values for candidate areas numbers 8, 45, 34 and 29 to calculate POI values.

Criteria name	Weight factor (%)	Candidate area number			
		8	45	34	29
RSI value	33	0.45	0.47	0.92	0.99
Number of interfered wells	14.5	0	2	0	2
Average distance between interfered wells and a candidate area (m)	14.5	NA	472	NA	524
Number of Adjacent wells	5.5	2	5	2	6
Average distance between Adjacent wells and a candidate area (m)	5.5	1005	784	814	1040
Distance between a candidate area and surface facilities (km)	9	12.7	11.1	6.9	10.1
Number of existing applicable wells for running a pilot	18	1	1	0	0

Table 4. Ranking of candidate areas numbers 8, 45, 34 and 29 for different weight criteria scenarios.

Candidate area number	Candidate areas ranking by scenarios							
	S-1	S-2	S-3	S-4	S-5	S-6	S-7	S-8
8	1	1	1	1	24	2	24	2
45	2	4	32	14	3	1	19	3
26	3	6	9	40	14	10	31	5
34	4	3	2	2	28	4	1	16

comparison with the other candidates (*i.e.* 8 and 45), candidate area number 8 has the highest POI value and it is suggested for running the pilot because candidate area number 29 does not have any existing applicable wells for the implementation of a pilot. Furthermore, candidate area number 29 has two interfered wells that considerably reduce the accuracy of interpretation of gathered data after running a pilot in it.

Figure A8 (Appendix A) shows contour maps of the RSI and POI values on top of the reservoir. There is a spatial relationship between the covariance matrix for the residual oil volume as well as recovery factor and location in the field as shown in the figure. Therefore, there is a spatial relationship between the RSI and the location in the field because RSI is based on covariance matrix for the residual oil volume as well as recovery factor.

Figure A9a (Appendix A) shows assigned numbers of candidate regions. In Figure A9b (Appendix A), Ranking of each region derived from RSI is shown. Moreover, the pilot area number 29 has the highest reservoir representative intensity value. In addition, in Figure A9c (Appendix A), each candidate region is ranked according to its POI value. Therefore, the candidate area number 8 has the first priority to be selected as the pilot area. It should be noted that the POI value is the most influential parameter for selecting the best candidate area, because it already covers RSI and other factors.

A sensitivity analysis is performed to evaluate how changes in the weights assigned to the criteria would change the ranking of the pilot candidate areas. Table B6 (Appendix B) shows sensitivity analysis by preference of certain criteria through 7 scenarios. In Table B6 (Appendix B), the A–G is a criteria name as shown in Table B5 (Appendix B). The S-1 is a base case scenario in which the weight of the criteria is determined using AHP method, as shown in Table B5 (Appendix B). In scenarios S-2 to S-8, the weight of criteria A to G is increased by 0.2 compared to the base case scenario, respectively. For instance, the weight of the RSI has increased by 0.2 while the ratio of the weight is the same for other criteria [39] in scenario S-2 as shown in Table B6 (Appendix B).

By making a comparison between the rankings of each candidate area number (*e.g.* 8, 45, 26 and 34) in different scenarios, it is demonstrated that candidate area number 8 is ranked first and second in six out of eight scenarios as shown in Table 4. Candidate area number 8 could be selected as pilot area based on the rankings obtained by using different weight of criteria. Therefore, candidate area number 8 is clearly dominant compared to other candidate areas.

4 Conclusion

Based on this research, six important conclusions have been found:

1. An innovative quantitative and systematic approach consisting of reservoir-geology and operational-economic criteria has been presented.
2. A cluster analysis as an unsupervised machine learning method could be utilized to quantify reservoir representative intensity values.
3. MCDMs as decision-making methods have been utilized to integrate the factors, *i.e.*, the distance from the surface facilities to candidate area, the number

of interfered wells, the mean distance from these wells to candidate area, number of existing applicable wells for running a pilot, number of adjacent wells, and the mean distance from these wells to candidate area, in addition to RSI criteria with each other, and thereby, a new index called the POI has been presented.

4. A model of reservoir is subdivided into 66 pilot candidate regions in accordance with optimized pilot size. Then, corresponding covariance matrix and recovery factor arrays of the areas are optimally clustered in 5 and 2 clusters based on simulated annual 3-D reservoir quality maps. A hybrid integration of the AHP-Shannon Entropy and RIM resulted in the highest value of POI in area number 8. Candidate area number 8 has been selected as pilot area in this case study since it has the POI value of 0.73. There is no interfering well affecting the reliability of pilot results in the area. Furthermore, it has one existing applicable well for running a pilot that reduces execution of pilot cost.
5. Finally, the proposed procedure has the capability to be extended to all enhanced oil recovery processes. Several EOR processes such as chemical flooding and miscible gas can be modeled to calculate remaining oil volume covariance matrix and RF array. Also, other screening criteria could be identified to determine POI according to the combination of the Shannon Entropy- AHP and MCDM methods.
6. The validity of the results taken from the presented methodology has been assessed by the following stages: (1) numerical reservoir modeling, (2) fuzzy clustering and (3) criteria weighting. In order to validate the simulated annual remaining oil volume, a history-matched reservoir model is utilized in this study. In addition, the validity indices of fuzzy-silhouette and Xie-Beni are utilized to validate the best cluster numbers. Furthermore, the criteria weighting results are validated by the Consistency Ratio (CR) in the AHP method.

Moreover, the proposed procedure could be developed using other criteria and well patterns.

References

1 Hu W. R. (2008) Necessity and feasibility of PetroChina mature field redevelopment, *Pet. Explor. Dev.* **35**, 1, 1–5.

2 Al-Qenae A., Chetri H., Kumar P.R., Orjuela J. (2018) Tracking the performance of strategically significant EOR Pilot: Zooming into inter-well connectivity, in: *Abu Dhabi International Petroleum Exhibition & Conference*, OnePetro

3 Chen P., Selveindran A., Kumar C., Saloma Y., Bose S., Balasubramanian S., Thakur G. (2019) CO_2-EOR and carbon storage in Indian oilfields: from laboratory study to pilot design, in: *SPE Western Regional Meeting*, OnePetro.

4 Teletzke G.F., Wattenbarger R.C., Wilkinson J.R. (2010) Enhanced oil recovery pilot testing best practices, *SPE Reserv. Evaluation Eng.* **13**, 01, 143–154.

5 Liu Z.X., Liang Y., Wang Q., Guo Y.J., Gao M., Wang Z.B., Liu W.L. (2020) Status and progress of worldwide EOR field applications, *J. Pet. Sci. Eng.* **193**, 107449.

6 Babadagli T. (2020) Philosophy of EOR, *J. Pet. Sci. Eng.* **188**, 106930.

7 Sandoval J.R., Pérez H., Maya G., Castro R., Muñoz E., Colmenares K., León J., Sánchez F., Villadiego D., Manrique E., Romero J., Izadi M. (2010) Dina Cretáceos Field chemical EOR: from screening to pilot design, in: *SPE Latin American and Caribbean Petroleum Engineering Conference*, OnePetro.

8 Chen X., Feng Q., Wu X., Zhao G. (2016) A pilot numerical simulation case study for chemical EOR feasibility evaluation, *J. Pet. Explor. Prod. Technol.* **6**, 2, 297–307.

9 Alfarge D., Wei M., Bai B., Alsaba M. (2018) Lessons learned from IOR pilots in Bakken formation by using numerical simulation, *J. Pet. Sci. Eng.* **171**, 1–15.

10 Taqi F., Ahmad K., Garcia J.G., Zhang I., Zijlstra E., Ayyad H., Sullivan M. (2019) Interference pressure transient test for permeability anisotropy evaluation in shallow unconsolidated reservoir undergoing EOR polymer flood pilot, in: *SPE Kuwait Oil & Gas Show and Conference*, OnePetro.

11 Chai C.F., Adamson G., Lo S.W., Agarwal B., Ritom S., Du K., Azizan N. (2011) St. Joseph Chemical EOR Pilot – a key de-risking step prior to offshore ASP full field implementation, in: *SPE Enhanced Oil Recovery Conference*, OnePetro.

12 Ozen O., Wahlheim T.A., Attia T., Barrios L., Bin Majid M. N., Wilkinson J. (2014) Dukhan field CO_2 injection EOR pilot: Reservoir modeling and planning, in: *International Petroleum Technology Conference*, OnePetro.

13 Ali H.A., Musa T.A., Doroudi A. (2015) Chemical enhanced oil recovery pilot design for Heglig Main Field-Sudan, in: *SPE Saudi Arabia Section Annual Technical Symposium and Exhibition*, OnePetro.

14 Saniez J., VandeBeuque S., Ekpenyong D.E., Bastos N., Wantong P., Salley B., Al-Yafei A. (2012) State of the art of geoscience and reservoir integrated study for EOR CO_2 Pilot Implementation: example of a Giant Carbonate Reservoir of Arabian Gulf UAE, in: *Abu Dhabi International Petroleum Conference and Exhibition*, OnePetro.

15 Al-Dhuwaihi A.S., Abdullah M.B., Tiwari S., Al-Murayri M. T., Al-Mayyan H., Shahin G.T., Shukla S. (2017) Fit-for-purpose chemical EOR ASP modeling strategy to guide pilot development decisions for a giant reservoir in North Kuwait, in: *SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition*, OnePetro.

16 Sharma S., Kamal D., Al-Maraghi E., AlMahrooqi S., Winkler M. (2016) Miscible gas EOR pilot design decisions driven by linking EOR performance parameters to uncertainties – a Kuwait Field Example, in: *SPE EOR Conference at Oil and Gas West Asia*, OnePetro.

17 Prasad D., Pandey A., Kumar M.S., Koduru N. (2014) Pilot to full-field polymer application in one of the largest onshore field in India, in: *SPE Improved Oil Recovery Symposium*, OnePetro.

18 Da Cruz P.S., Horne R.N., Deutsch C.V. (2004) The quality map: a tool for reservoir uncertainty quantification and decision making. *SPE Reserv. Evaluation Eng.* **7**, 01, 6–14.

19 Martini R.F., Schiozer D.J., Nakajima L. (2005) Use of quality maps in reservoir management, *J. Braz. Soc. Mech. Sci. Eng.* **27**, 4, 463–468.

20 da Cruz Schaefer B., Sampaio M.A. (2020) Efficient workflow for optimizing intelligent well completion using production parameters in real-time, *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **75**, 69.

21 James G., Witten D., Hastie T., Tibshirani R. (2017) *An introduction to statistical learning with applications in R*, Springer.

22 Gorunescu F. (2011) *Data mining: concepts, models and techniques*, Vol. **12**, Springer Science & Business Media.

23 Mohaghegh S.D. (2020) Subsurface analytics: contribution of artificial intelligence and machine learning to reservoir engineering, reservoir modeling, and reservoir management, *Pet. Explor. Dev.* **47**, 2, 225–228.

24 Hu J.I.A., Lihui D.E.N.G. (2018) Water flooding flowing area identification for oil reservoirs based on the method of streamline clustering artificial intelligence, *Pet. Explor. Dev.* **45**, 2, 328–335.

25 Vaseghi F., Ahmadi M., Sharifi M., Vanhoucke M. (2021) Generalized Multi-Scale Stochastic Reservoir Opportunity Index for enhanced well placement optimization under uncertainty in green and brownfields, *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **76**, 41.

26 Borgelt C. (2013) Objective functions for fuzzy clustering, *Comput. Intell. Intell. Data Anal.* 3–16.

27 Xie X.L., Beni G. (1991) A validity measure for fuzzy clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 8, 841–847.

28 Subbalakshmi C., Krishna G.R., Rao S.K.M., Rao P.V. (2015) A method to find optimum number of clusters based on fuzzy silhouette on dynamic data set, *Proc. Comp. Sci.* **46**, 346–353.

29 Shannon C.E. (1948) A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 3, 379–423.

30 Abellan A., Noetinger B. (2010) Optimizing subsurface field data acquisition using information theory, *Math. Geosci.* **42**, 6, 603–630.

31 Zhang Y., Li P., Wang Y., Ma P., Su X. (2013) Multiattribute decision making based on entropy under interval-valued intuitionistic fuzzy environment, *Math Prob. Eng.* **2013**, 526871.

32 Chaudhry A. (2004) *Oil well testing handbook*, Elsevier.

33 Pyrcz M.J., Deutsch C.V. (2014) *Geostatistical reservoir modeling*, Oxford University Press.

34 Saaty T.L., Peniwati K. (2013) *Group decision making: drawing out and reconciling differences*, RWS Publications.

35 Samad A.M., Hifni N.A., Ghazali R., Hashim K.A., Disa N. M., Mahmud S. (2012) A study on school location suitability using AHP in GIS approach, in: *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, IEEE, pp. 393–399.

36 Li Q., Zhang J., Deng B., Chang J., Li H., Liu S., Xu X. (2011) Grey decision-making theory in the optimization of strata series recombination programs of high water-cut oilfields, *Pet. Explor. Dev.* **38**, 4, 463–469.

37 Cables E., Lamata M.T., Verdegay J.L. (2016) RIM-reference ideal method in multicriteria decision making, *Inform. Sci.* **337**, 1–10.

38 Maschio C., Schiozer D.J. (2019) Integration of geostatistical realizations in data assimilation and reduction of uncertainty process using genetic algorithm combined with multi-start simulated annealing, *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **74**, 73.

39 Alinezhad A., Amini A. (2011) Sensitivity analysis of TOPSIS technique: the results of change in the weight of one attribute on the final ranking of alternatives, *J. Optim. Indus. Eng.* **4**, 7, 23–28.

Appendix A

Pattern	Permeability (md)	Layer 1	Layer 2	Layer 3	Layer 4
Alpha	1				
	10				
	100				
	1000				
Beta	1				
	10				
	100				
	1000				
Gamma	1				
	10				
	100				
	1000				
Delta	1				
	10				
	100				
	1000				

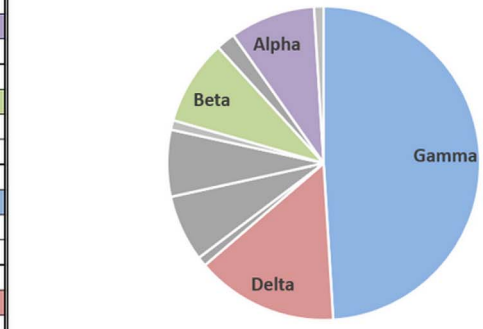

Fig. A1. Permeability ranges of the most represented patterns of one of the reservoirs and the associated pie chart showing their areal distribution on the field [14].

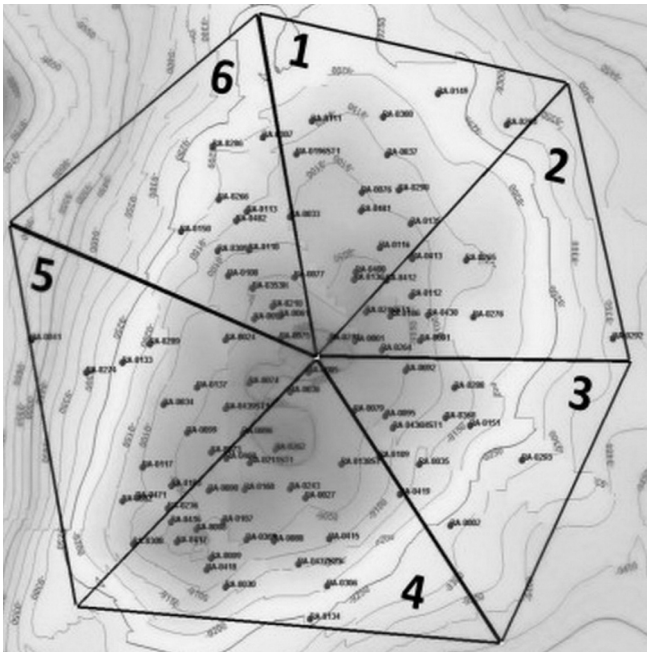

Fig. A2. Areal division of the reservoir into six areas [15].

Area	Zones	Remaining Oil	Cum Oil Production	Cum Water Injection	Cum WOR	RF	Injectivity	Distance between Inj & Prod	Sweep
1	Zone 1	M	H	H	L	H	M	H	M
	Zone 2	H	M	M	L	M	M	H	M
	Zone 3	M	M	L	L	H	M	H	H
	Zone 4	M	L	L	L	M	M	H	H
	Zone 5	H	M	M	M	M	M	H	H
2	Zone 1	M	M	M	L	M	M	L	L
	Zone 2	H	H	H	M	M	M	L	L
	Zone 3	L	L	M	M	M	M	L	L
	Zone 4	L	M	L	M	H	M	L	L
	Zone 5	H	M	L	L	M	M	L	L
3	Zone 1	H	M	L	M	L	M	M	M
	Zone 2	H	M	H	H	M	M	M	L
	Zone 3	L	L	L	M	M	M	M	L
	Zone 4	M	L	L	M	M	M	M	L
	Zone 5	H	H	L	L	M	M	M	H
4	Zone 1	H	M	M	H	M	H	M	L
	Zone 2	H	H	H	H	M	H	M	H
	Zone 3	M	L	L	M	M	H	M	H
	Zone 4	L	L	NONE	M	H	H	M	H
	Zone 5	H	M	L	L	M	H	M	H
5	Zone 1	H	L	H	M	M	H	M	M
	Zone 2	M	H	NONE	M	H	H	M	H
	Zone 3	L	M	M	L	H	H	M	H
	Zone 4	L	M	M	L	H	H	M	H
	Zone 5	H	H	L	L	H	H	M	H
6	Zone 1	M	M	M	M	M	H	M	M
	Zone 2	H	H	H	H	M	H	M	H
	Zone 3	L	M	M	H	M	H	M	H
	Zone 4	L	M	M	H	M	H	M	H
	Zone 5	H	L	L	M	L	H	M	H
Property Description		H>100 MMSTB	H - High Cum Oil Production, Cum Water Injection			H >40 %	H - High Injectivity	H> 1000 meter	H - High Sweep
		M: 80-40 MMSTB	M - Medium Cum Oil Production, Cum Water Injection			M:20-40 %	M - Medium Injectivity	M: 1000-500 meter	M - Medium Sweep
		L<40 MMSTB	L - Low Cum Oil Production, Cum Water Injection			H <20 %	L - Low Injectivity	L< 500 meter	L - Low Sweep

Fig. A3. Qualitative ranking of the selection criteria in each area [16].

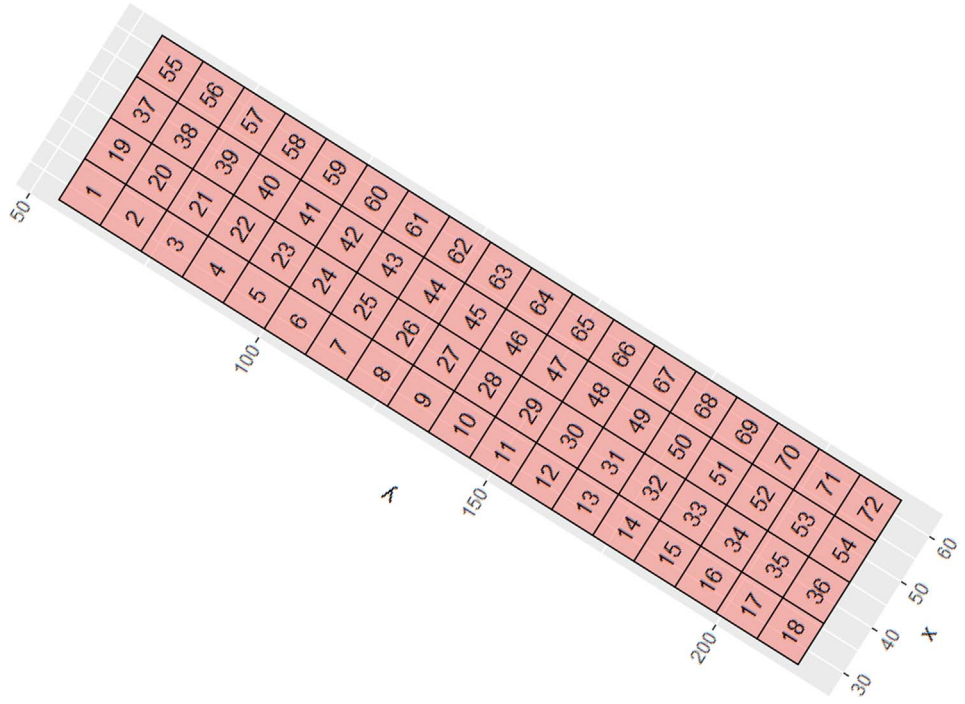

Fig. A4. Segmentation of the reservoir with allocated area numbers.

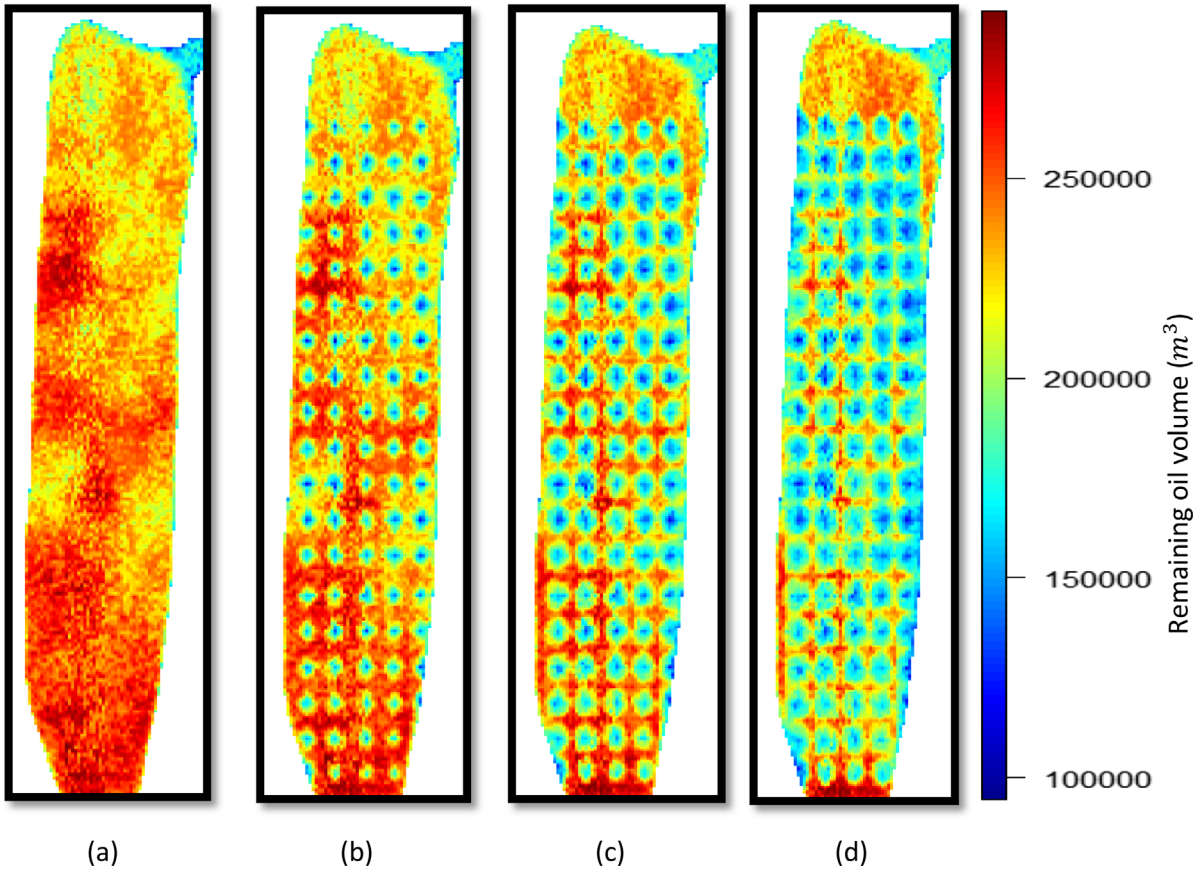

Fig. A5. Remaining oil volume maps at (a) beginning, (b) 10, (c) 20, and d) 30 years after waterflooding for the 15th layer of the reservoir.

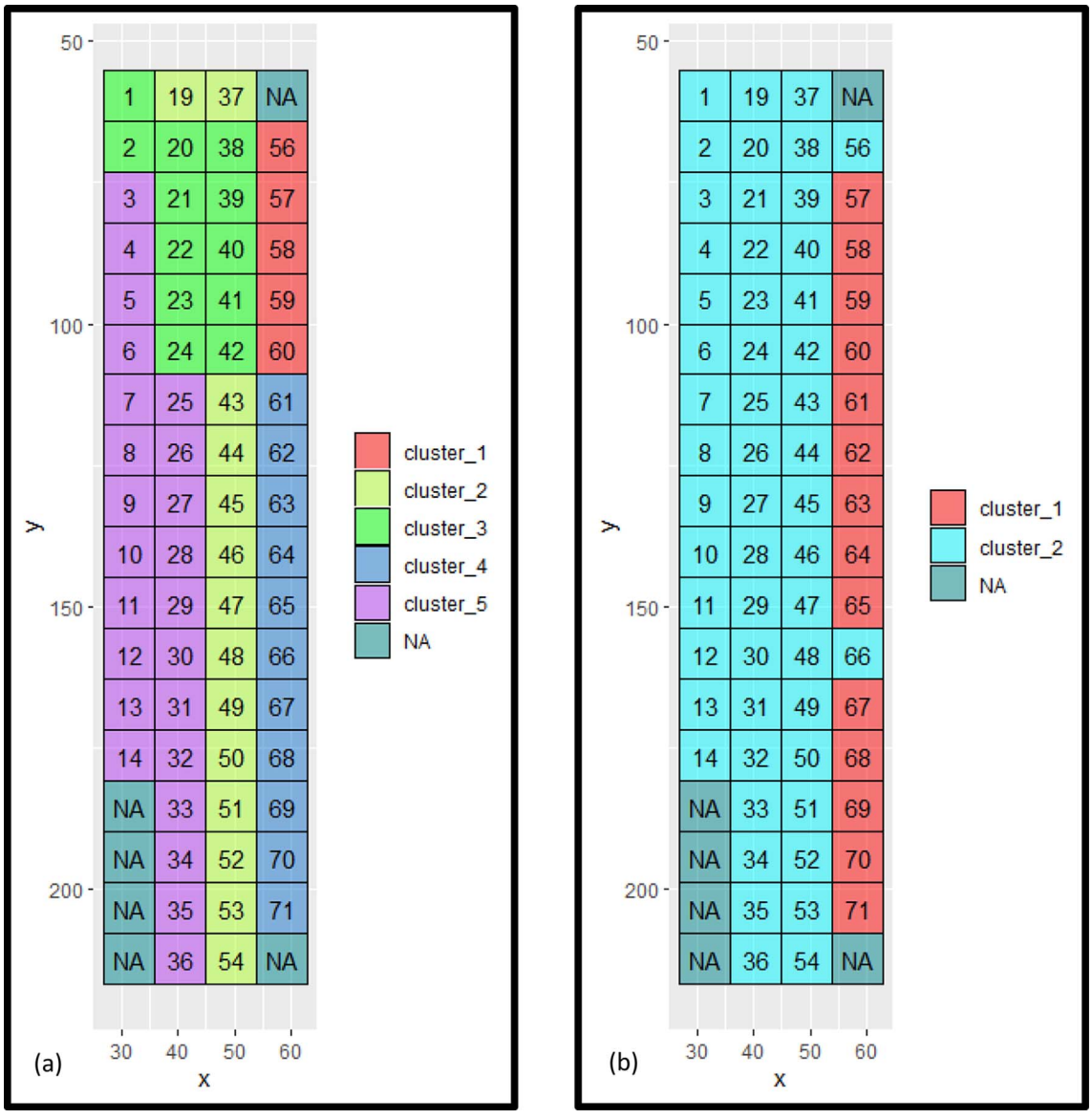

Fig. A6. The distribution map of the clusters in the field based on: (a) RF arrays. (b) COV arrays.

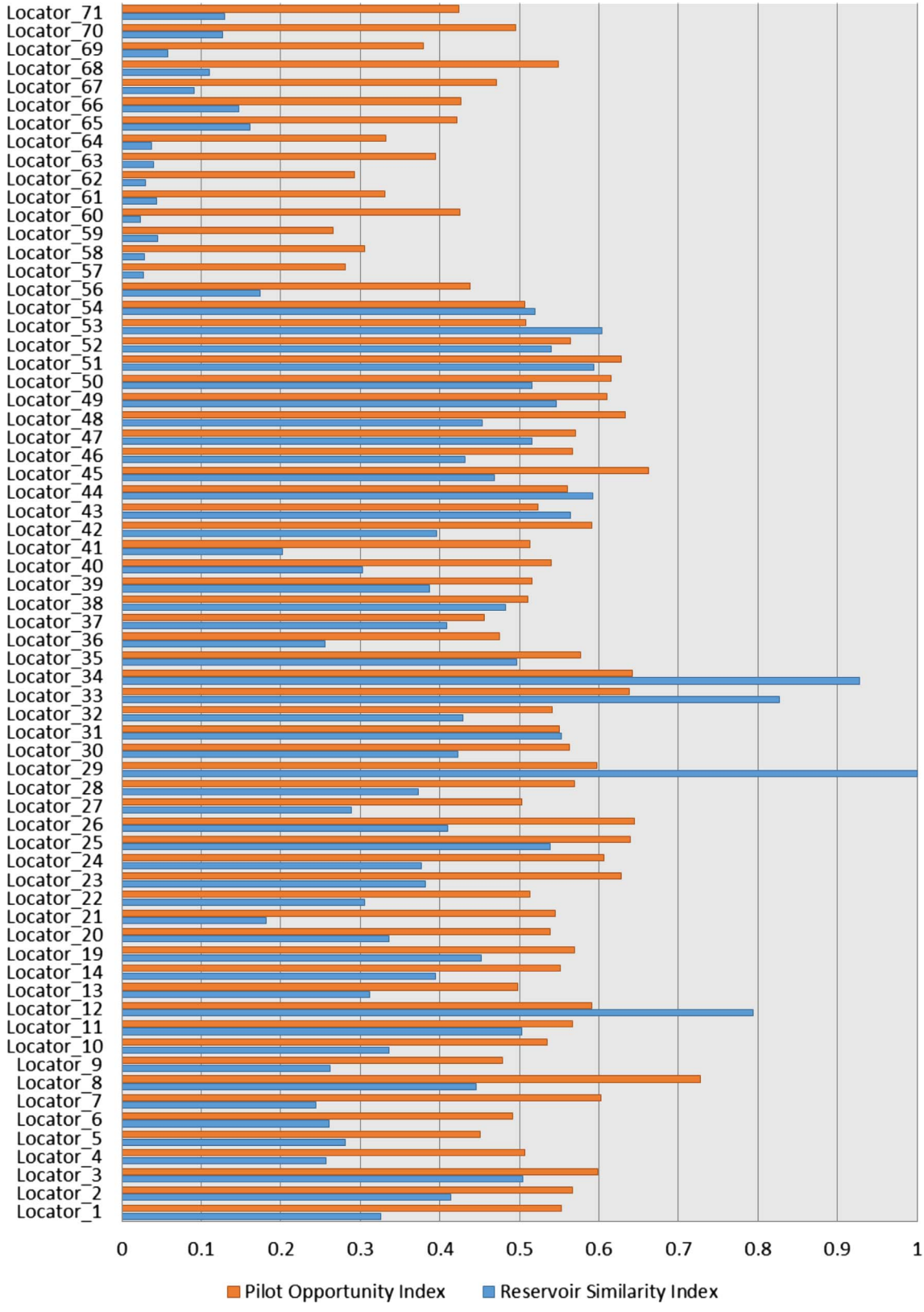

Fig. A7. Values of RSI (blue) and POI (red) in each candidate area.

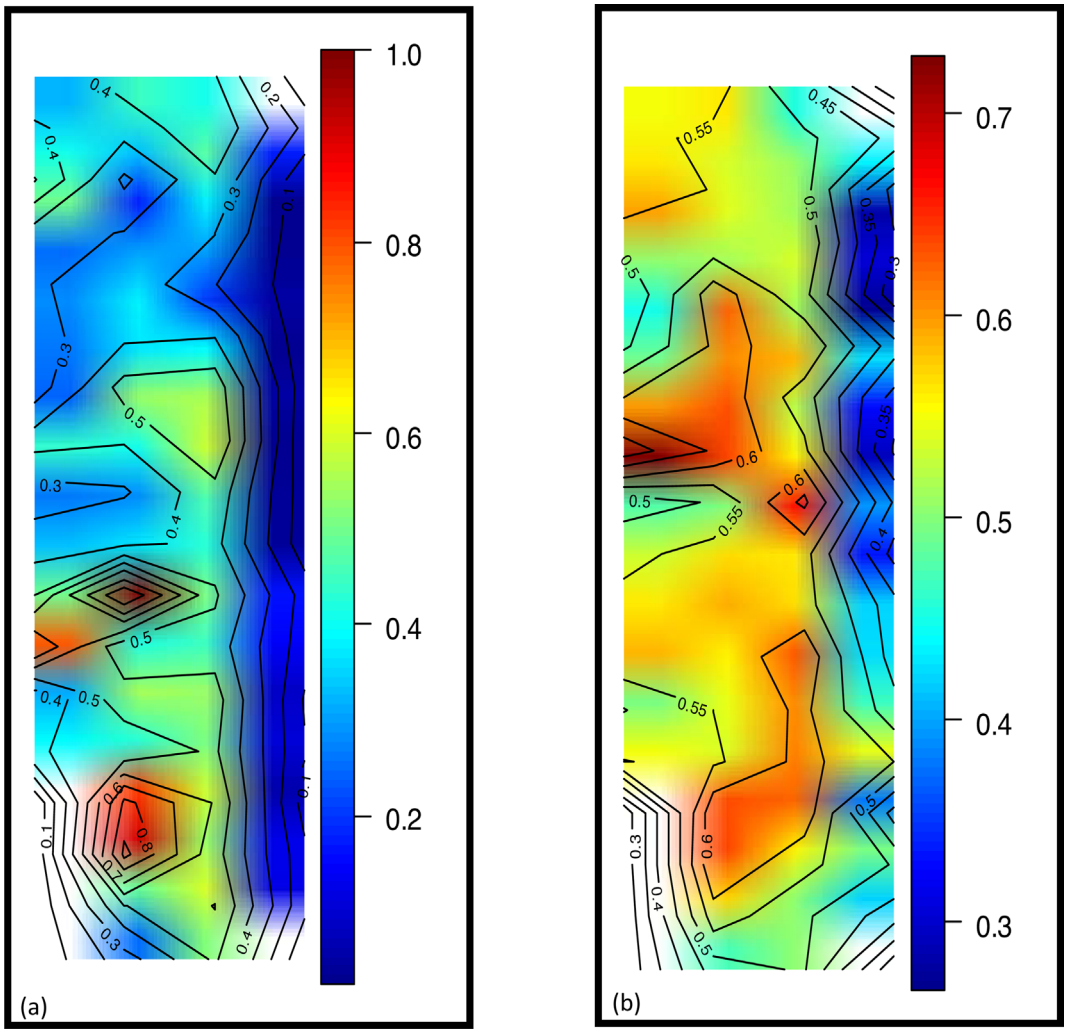

Fig. A8. Contour maps of the: (a) RSI and (b) POI values.

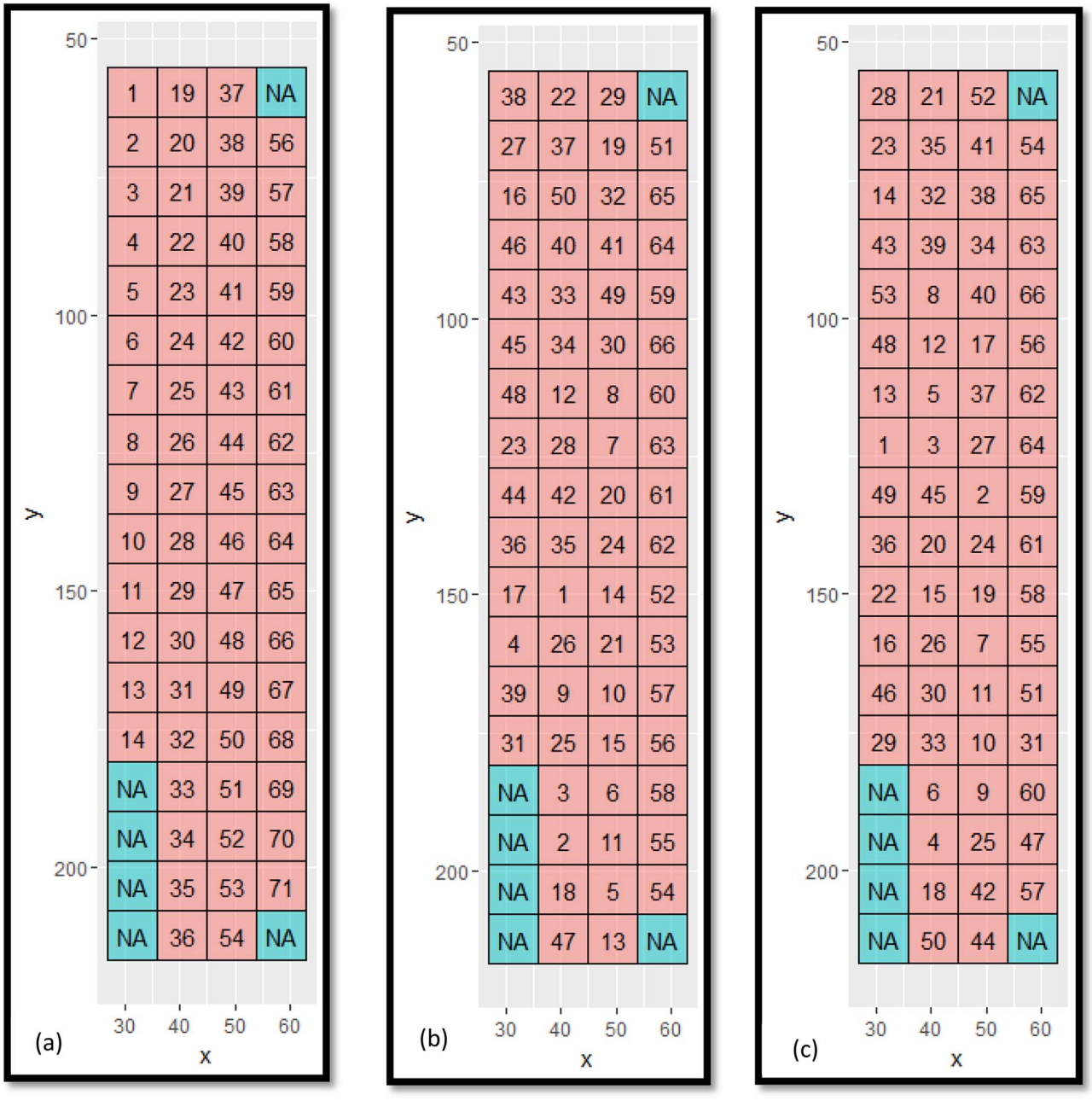

Fig. A9. (a) Number of candidate regions. (b) Ranking of each region derived from RSI. (c) Ranking of each region derived from POI.

Appendix B

Table B1. Data for determining pilot size.

1	Price of oil ($/BBL)	55
2	Discount rate (%)	10
3	Cost of drilling (MM$/per well)	12
4	Cost of water injection ($/BBL)	5
5	Cost of oil production ($/BBL)	1
6	Cost of handling produced water ($/BBL)	2
7	Cost of surface facilities (MM$)	300
8	Field Production rate (MMBBLD)	110
9	Voidage replacement ratio	1
10	Maximum water cut (%)	50

Table B2. Clustering outcomes for recovery factor arrays.

	Size of cluster	Average membership degree	Matrix of Euclidean distance				
			Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	5	0.99	0				
Cluster 2	14	0.95	52.25	0			
Cluster 3	12	0.94	40.57	39.83	0		
Cluster 4	11	0.98	43.32	41.75	57.01	0	
Cluster 5	24	0.97	60.46	33.09	41.51	48.26	0

Table B3. Clustering outcomes for covariance arrays.

	Size of cluster	Average membership degree	Matrix of Euclidean distance	
			Cluster 1	Cluster 2
Cluster 1	14	0.94	0	
Cluster 2	52	0.98	426.99	0

Table B4. Distance between candidate areas and center of dominant cluster for Recovery Factor (RF-center) and covariance between cells (covariance-center) arrays.

Number of areas	Distance of area from the Recovery Factor center	Cluster number (Recovery Factor)	Distance of area from the covariance center	Cluster number (Covariance between cells)	Number of areas	Distance of area from the Recovery Factor center	Cluster number (Recovery Factor)	Distance of area from the covariance center	Cluster number (Covariance between cells)
1	50.34	3	202.35	2	38	47.49	3	147.01	2
2	47.53	3	168.02	2	39	50.31	3	176.48	2
3	19.67	5	169.25	2	40	56.77	3	210.84	2
4	23.54	5	390.09	2	41	59.72	3	278.19	2
5	25.72	5	293.87	2	42	56.77	3	170.94	2
6	24.93	5	340.89	2	43	32.08	2	135.37	2
7	24.76	5	394.32	2	44	33.48	2	128.45	2
8	23.37	5	182.44	2	45	37.93	2	155.48	2
9	23.24	5	386.37	2	46	38.02	2	167	2
10	23.87	5	248.18	2	47	40.43	2	141.44	2
11	20.11	5	168.57	2	48	44.75	2	156.5	2
12	18.98	5	106.56	2	49	41.07	2	134.15	2
13	21.69	5	305.74	2	50	41.13	2	141.29	2
14	26.08	5	198.72	2	51	35.65	2	126.69	2
19	58.04	2	152.73	2	52	34.43	2	139.18	2
20	44.01	3	200.83	2	53	32.84	2	126.45	2
21	46.7	3	313.23	2	54	38.63	2	141.67	2
22	44.67	3	215.84	2	56	58.08	1	307.63	2
23	45.64	3	180.48	2	57	64.01	1	655.31	1
24	45.92	3	182.21	2	58	69.51	1	606.88	1
25	20.74	5	155.18	2	59	68.97	1	542.08	1
26	23.47	5	199.21	2	60	66.26	1	659.83	1
27	21.96	5	346.04	2	61	57.94	4	623.11	1
28	21.77	5	232.54	2	62	61.46	4	730.42	1
29	18.74	5	85.42	2	63	57.54	4	684.94	1
30	24.42	5	189.89	2	64	58.18	4	714.6	1
31	22.14	5	148.83	2	65	53.55	4	325.81	1
32	23.31	5	190.46	2	66	56.1	4	341.08	2
33	23.79	5	98.65	2	67	58.18	4	436.99	1
34	20.89	5	89.98	2	68	51.23	4	417.33	1
35	20.58	5	169.97	2	69	52.51	4	625.73	1
36	27.71	5	310.31	2	70	47.77	4	393.38	1
37	77.19	2	163.24	2	71	43.06	4	418.46	1

Table B5. The pairwise comparison matrix and weights of the criteria using AHP.

Criteria	A	B	C	D	E	F	G	Weights (%)
(A) RSI value	1	3	3	5	5	4	2	33
(B) Number of interfered wells	1/3	1	1	3	3	2	1	14.5
(C) Average distance between interfered wells and a candidate area	1/3	1	1	3	3	2	1	14.5
(D) Number of Adjacent wells	1/5	1/3	1/3	1	1	1/2	1/3	5.5
(E) Average distance between Adjacent wells and a candidate area	1/5	1/3	1/3	1	1	1/2	1/3	5.5
(F) Distance between a candidate area and surface facilities	1/4	1/2	1/2	2	2	1	1/2	9
(G) Number of existing applicable wells for running a pilot	1/2	1	1	3	3	2	1	18

Table B6. Scenarios with different criteria weights.

Scenarios	Criterion weight						
	A	B	C	D	E	F	G
S-1) base case	0.33	0.145	0.145	0.055	0.055	0.09	0.18
S-2) increasing of criterion (A)	0.53	0.10	0.10	0.04	0.04	0.06	0.13
S-3) increasing of criterion (B)	0.25	0.345	0.11	0.04	0.04	0.07	0.14
S-4) increasing of criterion (C)	0.25	0.11	0.345	0.04	0.04	0.07	0.14
S-5) increasing of criterion (D)	0.26	0.11	0.11	0.255	0.04	0.07	0.14
S-6) increasing of criterion (E)	0.26	0.11	0.11	0.04	0.255	0.07	0.14
S-7) increasing of criterion (F)	0.26	0.11	0.11	0.04	0.04	0.29	0.14
S-8) increasing of criterion (G)	0.25	0.11	0.11	0.04	0.04	0.07	0.38