Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 21, No. 3, pp. 533–551, 2025

GRAPH REGULARZED NON-NEGATIVE MATRIX FACTORIZATION WITH OPTIMAL BASIS ESTIMATION AND ITS APPLICATION IN RECOGNIZING OIL RESERVOIR*

Kehong Yuan, Youlin Shang†, Yongsheng Dong, Haixiang Guo and Zhonghua Liu

Abstract: To solve the issue of low efficiency in interpreting multidimensional unlabeled well logging data, an improved graph regularized non-negative matrix factorization model is proposed. This model takes into account the construction of basis matrix based on low-rank characteristics of the well logging data. Firstly, the rank of basis matrix is estimated by the low-rank matrix recovery model and weighted nuclear norm optimization algorithm. Secondly, the local features of the basis matrix are represented by a spectral clustering on the clean low-rank part. The local features and the optimized low-rank value are integrated to construct the basis matrix, which appropriately reflects the structure that is latent in the data. Finally, the optimized basis matrix is employed and visualized low dimensional non-negative features are extracted for well logging data within the graph regularized non-negative matrix factorization (GNMF) framework. Experimental comparisons on three well logging datasets and their combinations demonstrate the effectiveness of the proposed model, aiding in the efficient and automatic identification of the oil-bearing in reservoirs with unlabeled well logging data.

Key words: *non-negative matrix factorization, low rank matrix recovery, feature extraction, reservoir*

Mathematics Subject Classification: *97M10, 90B50, 68T10*

1 Introduction

Reservoir identification is an important aspect of reservoir evaluation, reservoir description, and real-time drilling monitoring. The identification of oil-bearing formation is one of the most important aspect in identifying and evaluating oil and gas reservoirs in the process of oil exploration and development [29]. It means to recognize the type of reservoir according to correlative well logging features for each data. It reflects the results and application value of well logging interpretation, which affects the efficiency and success rate of oil exploration[19]. From the view of machine learning, it is a pattern recognition problem. The task of recognition relies on the selection of model and parameter optimization. However, the observed well logging data is high-dimensional and often contains noise, which not only causes computational and storage pressure, but also reduces the performance of model [3]. Therefore, in order to improve the efficiency of identification and obtain key well logging feature representations with low rank structure from high-dimensional well logging data, it is particularly

*This research was supported by Henan Province Natural Science Foundation Youth Science Fund (No.252300420898), Key Technologies Research and Development Program of Henan Province (Nos.252102240125, 242102210102) and National Natural Science Foundations of China (No.12071112).

†Corresponding author.

DOI: https:// doi.org/10.61208/pjo-2025-026

important to identify the oil content of reservoirs. Dimensionality reduction is widely used to solve this problem.

In the past few decades, dimensionality reduction has attracted widespread attention and many valuable research results have been published [25, 30, 33]. According to whether the class information of the original samples has been utilized, dimensionality reduction methods can be divided into three categories: supervised, unsupervised, and semi supervised dimensionality reduction methods [23]. The well-known dimensionality reduction methods include linear discriminant analysis (LDA) [1] and non-negative matrix factorization (NMF)[13], in which LDA is a supervised method and NMF is an unsupervised learning method. Overall, supervised methods typically achieve better performance than unsupervised methods due to the use of label information from training samples [23]. Therefore, in the identification of reservoir oil content, references [19, 8, 18] and others use label information, based on different models and evaluation criteria, to select low dimensional features with clear physical meanings. However, the heterogeneity of reservoirs leads to the multiplicity and uncertainty of petroleum geological problems, making it difficult to obtain a large amount of labeled data [11]. Furthermore, the cost of obtaining geological data is often high, so the data obtained is mostly "small samples" or unlabeled data. Therefore, it is meaningful to study the low dimensional feature representation of well logging data under unsupervised conditions. However, to our knowledge, there is very little research in this area. Due to the non-negative and low rank characteristics of well logging data, NMF can easily obtain non-negative low dimensional features of the data that align with practical applications.

A novel GNMF model is proposed for extracting accurate low dimensional features of well logging data. The innovation of this paper is mainly reflected in the following points:

1. A novel model has been proposed to improve traditional NMF and enhance the quality of feature representation. In the newly designed model, NMF integrates low rank matrix recovery to eliminate the potential impact of logging noise data and provide structural information that is latent in the data.

2. The rank of basis matrix is accurately estimated in the low rank matrix recovery model (LRMR) and an accurate optimization algorithm. In addition, we design experiments to verify the reasonableness and effectiveness of the estimated rank value.

3. The local features of the basis matrix are further optimized. Based on accurately restored low-rank information, the partition is performed by spectral clustering algorithm represents the local characteristics of the reservoir. In addition, accompanied by the basis matrix optimized in structure and local features, we also propose an improved GNMF.

4. The effectiveness of this method has been verified through experiments on real data from Jianghan Oilfield, and its performance is superior to comparative methods.

The rest of this paper is organized as follows. In Section 2, we briefly review the related works. NMF and its variants are introduced in Section 3. We describe the optimal estimation of the base matrix in detail in Section 4 and give an improved GNMF model with optimal estimation. Experimental results and analysis on three real well logging data and their combinations are demonstrated in Section 5. Finally, we briefly conclude this paper in Section 6.

2 Related Works

Well logging data is influenced by various factors such as formation porosity, well fluid composition, lithology, etc, during the collection process, manifested as the superposition of real data and noise data [22]. In fact, potential noise in the data will definitely affect the performance of subsequent processing, such as clustering and classification. Wright et al. [36] proposed robust principal component analysis (RPCA) and demonstrated that low rank matrix A can be effectively and accurately recovered from damaged observations $X = A + E$, where E is an unknown error matrix with sparsity. The low-rank part usually contains the main information and structure in the data, which is necessary for tasks such as dimensionality reduction and pattern recognition. Therefore, in order to accurately recover the low rank part of the data, many scholars have conducted extensive research. Lin et al. [21] proposed the inexact augmented Lagrange multipliers (IALM) algorithm and efficiently recovered low-rank matrix of the data. In order to improve the flexibility of the lowest kernel norm, Gu et al. [7] proposed the weighted kernel norm minimization model (WNNM) by establishing weights based on the relationship between singular value sizes. Oh et al. [26] proposed partial sum of singular values (PSSV) minimization based on prior rank information. He et al. [9] proposed an improved model with rank constraints based on prior rank information. How to choose an appropriate algorithm based on actual scenarios to accurately recover the low rank part is a key issue in low rank matrix recovery. That is to say, the performance and information contained in the low-rank part depending on the optimization algorithm suitable for the scenario.

The core of existing methods is to find a low rank matrix that approximates the original data matrix, in order to estimate the low rank structure of the original data. Although determining the rank of a matrix is a challenging task, it often serves as a predetermined input for some methods, affecting the performance of the algorithm [20]. As a low-rank approximation technique, NMF is exactly that way, where the rank of the basis matrix needs to be pre-set, attempting to express each object as a linear combination of several basis vectors. In order to enhance the representation ability of the basis matrix, Wild et al. [35] proposed a K-means method for initializing the basis matrix, which replaces random initialization with class center vectors to improve decomposition performance. Zhang et al. [40] introduced the Ng-Jordan-Weiss (NJW) spectral clustering algorithm into NMF, reducing the potential impact of shape distribution and high dimensionality in the data space. Li et al. [16] considered the potential impact of noise in the raw data and incorporated NMF into the RPCA framework, proposing a non-negative low rank matrix factorization.

Although NMF can effectively achieve low dimensional representation of observed data and furtherly obtain good clustering performance through low-rank factorization, there are still some limitations. Firstly, the NMF model is mainly used to learn the low rank representation of observed samples in the basis space of the original data vector space, and it is still essentially a representation-based method. The base for representation requires a suitable basis space for local structural features [15]. Due to the potential impact of noise, it is difficult to directly compute the intrinsic dimensionality that reflect the structure of the original observed data. After all, external disturbance universally exists in industrial systems, which will degrade the system performance and even destroy the stability of the system once the anti-disturbance question is not seriously treated [32]. Well logging data is just like this, the potential noise may change the data structure and disrupt the low-rank property of the data, thereby degrade the performance of matrix factorization algorithms [20]. If applied directly to the original data and subjectively set the rank of the basis matrix, it can affect the quality of the features, which in turn influences the performance of clustering or classification tasks. Secondly, the local feature information of the data plays an important role in NMF, which is ignored due to its random setting. Basis vector corresponds to a certain local feature, it can be textures in images or topics in text. However, from the perspective of a linear space, basis vectors are linearly independent, meaning they cannot represent each other. In practical applications, these features are closely related to the low-rank structural information of the observed samples. In this proposed method, for feature extraction of well logging data, the rank and local features of the NMF basis matrix are optimized and estimated, respectively. Firstly, the rank of the basis matrix is estimated in the low rank part of the well logging data, significantly reducing the potential impact of noise. The presence of noise changes the true low rank structure of the vector space of well logging observation data. The intrinsic structure of well logging data should be estimated in the low-rank part after separating noise. Meanwhile, the obtained low rank value is helpful to extract the appropriate number of local feature. The effects of parameter variations cannot be ignored in many practical physical systems [31]. Secondly, the basis vectors are represented by the cluster centers of the low-rank part of the well logging data, which can enhance the accuracy of the reconstructed data. Basis vectors represent local features of the reservoir through class centers under clean data partitioning, it is more objective and interpretable. Integrating the representation of local features and the rank of the matrix, a basis matrix is constructed to appropriately reflect the structure that is latent in the well logging data. Finally, the constructed basis matrix is employed by GNMF and the robust and efficient low dimensional features can be obtained for reservoir oil identification. The experimental results of real well logging data show that our proposed improvement method is effective and superior to the four related algorithms.

3 NMF and Its Several Variants

Because the main work is carried out on the basis matrix of NMF, it is necessary to review NMF and its several variants. NMF algorithm [13] is an effective method for matrix low-rank approximation, which obtains the latent features of the data by performing non-negative factorization on a non-negative matrix. This approach learns a basis matrix that allows the original data to be faithfully reconstructed under the non-negative linear combination of its local features [16]. For a given $X = [x_1, \cdots, x_n] \in R^{m\times n}$, each column of X is a sample vector. NMF tries to find two non-negative matrices $U = [u_{ik}] \in R^{m\times r}$ and $V = [v_{ik}] \in R^{n\times r}$ to approximated the original matrix. That is $X \approx UV^T$, where U is the basis matrix, and V is the coefficient matrix. From a mathematical perspective, each observed sample vector x_i can be represented as a linear combination of the columns of U, weighted by each column of V. Thus, V can be regarded as a projection of the corresponding observed sample vector in X according to the basis matrix U, and it is a compressed approximation of the original matrix. In fact, the basis vectors represent fundamental features in the data space, while the coefficient vector controls the way how to combine these basis vectors.

To solve the low-rank approximation, the following minimize the objective function is performed.

$$\begin{aligned} &\min_{U,V} f(U,V) = \min_{U,V} \left\| X - UV^T \right\|_F^2 \\ &s.t. U > 0, V > 0 \end{aligned} \tag{3.1}$$

where, $\|\cdot\|_F$ denotes the matrix Frobenius norm. Lee and Seung presented the following updating rules[13]:

$$u_{ik}^{t+1} \leftarrow u_{ik}^t \frac{(XV)_{ik}}{(UV^TV)_{ik}} \tag{3.2}$$

$$v_{kj}^{t+1} \leftarrow v_{kj}^t \frac{(X^TU)_{kj}}{(VU^TU)_{kj}} \tag{3.3}$$

In NMF, the representation of the observed data is obtained under a low-rank basis matrix and non-negative constraint, so the low dimensional features can be well represented.

This kind of representation can give a good interpretation of the data in the related applications due to their values are non-negative naturally, such as signal processing [14], pattern recognition [37], clustering [10, 4]. Several NMF variants methods are given as follows.

Graph regularized non-negative matrix factorization (GNMF) [2]. GNMF considers the geometric structure in the data by constructing an affinity graph, and its model is given as follows.

$$\begin{aligned} &\min_{U,V} f(U,V) = \min_{U,V} \left\| X - UV^T \right\|_F^2 + \lambda_1 Tr(V^T L_z V) \\ &s.t. U > 0, V > 0 \end{aligned} \tag{3.4}$$

where λ_1 is the tradeoff parameter of the Laplacian regularization. $L_z = D - S$ is the graph Laplacian matrix. S is a symmetric weight matrix and D is a diagonal matrix whose entries are column sums of S.

NMF with spectral clustering initialization enhancer (NJW-NMF) [40]. The model introduces the spectral clustering center to initialize the basis vector for NMF. In NJW-NMF, both the rank and the basis vector of matrix are optimized simultaneously by clustering the original data.

4 The Estimation of Basis Matrix

4.1 Motivation

Though NMF can obtain non-negative representation of the observed samples, the reconstruction relies on the setting of basis matrix. A good approximation can only be achieved if the basis matrix appropriately reflects the underlying structural features of the data [12]. Accurately expressing mathematical model is conducive to correctly representing the dynamic behavior of the system [28]. In NMF, the rank of basis matrix is pre-set subjectively, which cannot reflect the true intrinsic dimension. However, the rank of basis matrix will affect the performance of the matrix factorization [6] and determine the number of feature to be extracted. So how to choose the rank is very important. Unfortunately, determining the rank is a challenging task. In addition, most existing works directly apply NMF on high-dimensional data for computing the effective representation of raw dataset, and the potential noise contained in the raw data is ignored.

In real world applications, well logging data always contains noise, which disrupts the low-rank structure of the data. Inspired by the low-rank matrix recovery (LRMR) with ℓ_1 norm, the well logging data can be divided into low-rank part and sparse part. The motivation behind low-rank is from the observation that a part of principal components of a matrix usually contain most of the information [17]. The essential information of the well logging data is hidden in low-rank part. Low-rank part plays a key role in the estimation of rank and it is also the basis for estimation. That is to say, the accuracy of the estimation for the rank heavily relies on the low-rank part obtained from the optimization algorithms. We aim to use accurate optimization algorithm to precisely recover the low-rank part of well logging data and construct basis matrix. It is well-known that the low-rank part in LRMR can well recovery under the conditions which is the calculated low rank value is not too high and the noise is sufficiently sparse [27]. Therefore, an adaptive recovery optimization algorithm WNNM is adopted in the proposed algorithm. The rank and local feature of basis matrix are estimated and represented in an appropriate manner based on the low-rank part, respectively. We also consider the construction for the local features based on low-rank part. So, the rank estimation and local feature representation are explored together to optimize basis matrix in NMF for furtherly achieving accurate feature extraction.

4.2 The estimation of rank in basis matrix

Well logging data is influenced by various factors such as formation porosity, well fluid composition, and lithology during the collection process, manifested as the superposition of real data and noise data [22]. Due to the presence of potential noise disrupting the low-rank structure of well logging data, the low-rank data contains the common properties of the reservoir. Thus, the observed data X can be approximated by $X = A + E$, where A denotes the low-rank part of the data, and E denotes the sparse noise. The low-rank part contains the essential information of the original well logging data.

In LRMR, the rank of the low-rank part approximates the true rank of the original matrix. As we know, NMF is a matrix factorization algorithm that each sample vector can be represented as a linear combination of the column of the basis matrix. According to [16], we know that if the rank of X is given, then there are two matrices U and V, let $X = UV^T$. That is to say, both the basis matrix U and observed data matrix X have the same rank value r. Therefore, we can use the rank of low rank matrix obtained from LRMR to approximate the rank of basis matrix in NMF in the case of exact recovery. This kind of exact recovery requires satisfying relatively strict conditions, such as low rank value is not too high, the noise is sufficiently sparse and the optimization algorithm used for recovery has a certain degree of adaptability.

According to the property of matrix rank in linear algebra, the rank value does not exceed 6 for each well logging dataset. If we consider the mathematical meaning of rank, which indicates the number of linearly independent columns or rows in a matrix, the rank value does not exceed 4. However, it is important to note that even with a low rank, there is still uncertainty in determining the value of the rank. That is the reason that we employ the WNNM algorithm to exactly recovery low-rank part and estimate the rank precisely, which it reveals the underlying structure and essential information within the data. It is helpful in analyzing the underlying structure and patterns of the data. The model of LRMR can be summarized as follows:

$$\begin{aligned} &\min_{A,E} \|A\|_* + \lambda_2 \|E\|_1 \\ &s.t. X = A + E \end{aligned} \tag{4.1}$$

where $\|A\|_*$ denotes the nuclear norm of low rank matrix A, calculated as $\|A\|_* = \sum_i^n \delta_i$, δ_i is the i-th singular value of matrix A, n is the number of sample in data, E is the noise matrix, $\|E\|_1$ is the ℓ_1 norm of matrix E, defined as the sum of absolutes of all entries, and λ_2 is a penalty parameter for balancing the rank function and the ℓ_1 norm of matrix E. It is the goal of LRMR that estimating the low-dimensional subspace via finding a low rank matrix A, and the rank of A can be an optimal estimation for the target dimension of the subspace.

Several optimization algorithms have been used to solve LRMR model, such as iterative thresholding, accelerated proximal gradient, augmented Lagrange multiple (ALM) algorithms, inexact augmented Lagrange multiple (IALM) [21], and weighted nuclear norm minimization (WNNM) [7]. In WNNM, due to assigning different weights to different singular values, it is possible to better control the low-rank approximation of the matrix. Thus, we employ the WNNM algorithm to recovery the low-rank part for the well logging data. It is used to estimate the intrinsic dimensionality for well logging data and its implementation steps can be described as follows:

Step 0: Set well logging data matrix $X \in R^{m\times n}$, weight vector ϖ, initial parameters $\mu_0 > 0, \rho > 1, \theta > 0$, $X_0 = X, L_0 = 0$ and the iterative number $k = 0$;.

Step 1: Set the termination conditions for the algorithm: $\|X - A_{k+1} - E_{k+1}\|_F / \|X\|_F > \theta$

Step 2: Update sparse matrix: $E_{k+1} = \arg\min\limits_{E} \|E\|_1 + \frac{\mu_k}{2} \left\|X + \mu_k^{-1} L_k - A_k - E\right\|_F^2$;

Step 3: Update low-rank matrix: $A_{k+1} = \arg\min\limits_{A} \|A\|_* + \frac{\mu_k}{2} \left\|X + \mu_k^{-1} L_k - E_{k+1} - A\right\|_F^2$;

Step 4: Update Lagrange multiple matrix:$L_{k+1} = L_k + \mu_k \left[X - A_{k+1} - E_{k+1}\right]$;

Step 5: Update parameter:$\mu_{k+1} = \rho\mu_k$;

Step 6: If the termination condition is met, then calculate the $R = rank(A_k)$ and output (A_k, E_k, R); otherwise,let $k \leftarrow k + 1$ and go to step2.

The estimated results need to be verified. As we know, the kernel norm is approximation of the rank function in model (4.1). The performance of optimization algorithm is another factor that affects the quality of the recovery part. Therefore, it is necessary to verify the estimated results. We test the rationality of the estimated results, according to its subsequent impact. As we know, good approximation depends on the appropriate structure of the basis matrix, and good features can be regarded as one of manifestation of this approximation. Under the same conditions, we can quantify the quality of these features, such as clustering them in the same classifier. In fact, we employ the K-means classifier to quantitatively evaluate the feature quality of different low rank value to demonstrate the validity of the estimated value. If the estimated rank and the prior category information are denoted as R and k, respectively, considering computational errors, we set up a candidate set $S = \{R - 1, R, R + 1, k\}$.

4.3 The estimation of local features

Rank is an important factor in constructing the basis matrix, as it indicates the number of basis vectors. Another factor, namely local features, is also an important component in constructing the basis matrix. For example, these features can be textures in image recognition or topics in text recognition. To quantitatively describe the local feature of the reservoir, we need to represent the typical features of various reservoirs in vector form. The different vectors composed of these elements correspond to different types of reservoirs, such as oil layer, inferior oil layer, water layer, dry layer. According to [5], local feature and the "class center" in clustering problem is equivalent. And NJW is more effective than principal component analysis (PCA), fuzzy c-means (FCM) and K-means in initializing local feature, in extracting feature of mechanical faults [40]. Therefore, we employ NJW algorithm to obtain the average feature performing on the low-rank part.

4.4 GNMF with optimal basis matrix

The construction and estimation of the basis matrix provide NMF with the appropriate structure and features. It is beneficial to extract appropriate low-dimensional features, due to its accurate and appropriate linear structural information. However, the intrinsic geometric information of the data is ignored. To utilize the geometric structural information of the data space, a kNN-based graph is constructed to encode the relationship of the samples. The constructed basis matrix can be employed in the GNMF framework. Thus, the more comprehensively the intrinsic structural information of the data is utilized, the better the quality of the extracted features. Figure 1 shows the process of feature extraction based on GNMF with optimal basis matrix.

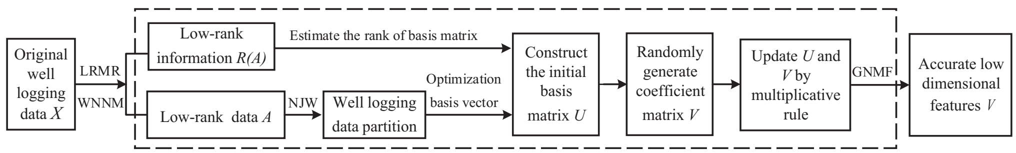

Figure 1: The process of feature extraction based on GNMF with optimal basis matrix

The proposed model integrates the advantages and ideas of the above two variants of NMF. It is implemented within the GNMF framework, and its basis vectors are represented by NJW class center. However, there are two differences with the two variants of NMF. Compared with GNMF, the proposed model can extract accurate low-dimensional features and have good clustering performance because the initial basis matrix is optimized in the rank and basis vectors. Compared with NJW-NMF, the proposed model provides a more optimized initial basis matrix. Its rank is estimated in the low-rank part separated from original well logging data. And its basis vectors are represented by NJW's class center partitioning applied to low-rank data.

5 Experimental Result

As a class of popular data representation method, NMF can be used in a variety of fields. We want to utilize it in the oil-bearing recognition to obtain the key low-dimensional feature. Our goal is to construct a basis matrix with an accurate structure for GNMF model to obtain appropriate approximation. The structure and feature representation of the basis matrix come from the low-rank part recovered from original well logging data. The effectiveness of the proposed method is verified on the real data oilsk81, oilsk83, oilsk85 wells and their combinations of Jianghan oil fields in China. Particularly, once the proposed algorithm learns the low-dimensional representation V of the raw dataset X in the proposed method, we employ K-means algorithm on V to produce the cluster label, followed by the permutation mapping function [24]. The experimental setting and results are presented in the following sections.

5.1 Data description

Experiments are performed on three well logging data oilsk81, oilsk83, oilsk85 and their combinations from Jianghan oil fields in China (see Table 7-9 in Appendix). Table 1 shows the statistic description of the real data. There are six well logging features and four class information in recognizing oil-bearing formation. The well logging features are acoustic travel time (AC), compensated neutron logging (CNL), resistivity (RT), porosity (POR), oil saturation (SO), and permeability (PERM), respectively. Class information includes dry layer, water layer, oil layer, and inferior oil layer, respectively. The quality of data is good without any missing values. The experiment is implemented in the Matlab R2018b environment.

Table 1: Statistic description on the well logging datasets

Dataset	Features	Samples	Classes
oilsk 81	6	31	4
oilsk 83	6	50	4
oilsk 85	6	34	4
oilsk 81-83	6	81	4
oilsk 81-85	6	65	4
oilsk 83-85	6	84	4
oilsk 81-83-85	6	115	4

5.2 Evaluation Metrics

The purpose of dimensionality reduction is to eliminate noise and remain meaningful features, making the dataset easier to manage and understand, and improving the accuracy of predictive models accordingly. And quantitative evaluation metrices for assessing clustering performance include external and internal indices. They are very useful for evaluating the quality of clustering. Accuracy (ACC) and normalized mutual information (NMI) [16, 38, 34, 39] are two commonly used metrics in external evaluation. Good recognition results benefit from the quality of feature extraction and the good design of the classifier. Both ACC and NMI are in the range of [0,1], and the larger value of ACC or NMI, the better the clustering performance. Additionally, it means that well logging data can be better segmented by K-means in new feature space, and the quality of extracted features is higher. For a dataset X with n samples, ACC is calculated as Eq. (5.1)

$$ACC = \frac{\sum_{i}^{n} \delta(y_i, map(r_i))}{n} \tag{5.1}$$

where, y_i and r_i denote the cluster label obtained by clustering algorithm and the true label of sample x_i, respectively. Under the condition of $x = y$, then $\delta(x, y) = 1$, otherwise, $\delta(x, y) = 0$. $map(\cdot)$ is a permutation mapping function, used to map each prediction cluster label r_i to the equivalent label according to the distribution of the true label. When the predicted cluster label K' is obtained and the true label K is given, NMI is defined as follows:

$$NMI(K, K') = \frac{MI(K, K')}{\max(H(K), H(K'))} \tag{5.2}$$

where, $H(K), H(K')$ denote the entropy of labels K and K', respectively. Mutual information $MI(\cdot)$ is calculated as follows:

$$MI(K, K') = \sum_{s\in K}\sum_{t\in K'} p(s,t)\log_2\left(\frac{p(s,t)}{p(s)p(t)}\right) \tag{5.3}$$

where, $p(s, t)$ denotes the joint probability distribution of s and t, $p(s)$ and $p(t)$ are the marginal probability of s and t, respectively.

5.3 Experiments on real well logging data

In this section, we choose three real well logging data, oilsk81, oilsk83, oilsk85 and their combinations to prove the effectiveness of the proposed method. Table 2 shows the estimated rank results for oilsk81, oilsk83, oilsk85 and their combinations.

Table 2: Rank estimation with IALM and WNNM algorithm

Dataset	IALM	WNNM	Samples
oilsk 81	3	3	31
oilsk 83	3	3	50
oilsk 85	2	3	34
oilsk 81-83	3	3	81
oilsk 81-85	3	3	65
oilsk 83-85	3	3	84
oilsk 81-83-85	3	3	115

It can be seen that when estimating the rank of the observation data matrix, WNNM is more stable than IALM algorithm and it does not increase with the number of logging data. It reflects that the intrinsic feature number is much lower than the observed attribute number, and the multidimensional well logging data lie in a low-dimensional linear space. However, due to differences in algorithm performance, the calculation results still show some fluctuations. According to the number of prior categories k and possible algorithm errors, the candidate solution is set as $S = \{2, 3, 4\}$.

The evaluation are performed by standard NMF algorithm with different rank in the three real logging data. In addition, two metrics, i.e., ACC and NMI are selected to quantitatively evaluate clustering performance. Table 3 shows the statistical results of independent experiments conducted 20 times with standard NMF algorithm for the three well logging datasets with different rank. The bold numbers in the Table 3 highlight the optimal values for the rank and clustering results.

Table 3: Clustering performance of NMF with different embedding dimensions.

	oilsk81		oilsk83		oilsk85	
R	ACC	NMI	ACC	NMI	ACC	NMI
2	89.5 ± 6.6	84.6 ± 5.7	81.6 ± 4.2	73.5 ± 4.4	65.7 ± 5.9	54.7 ± 1.4
3	$\mathbf{95.5 \pm 2.4}$	$\mathbf{89.6 \pm 4.8}$	$\mathbf{88.3 \pm 2.4}$	$\mathbf{75.5 \pm 4.0}$	$\mathbf{77.6 \pm 7.5}$	$\mathbf{69.3 \pm 5.6}$
4	91.2 ± 5.4	85.5 ± 5.4	83.1 ± 6.8	74.3 ± 4.8	75.4 ± 7.4	68.1 ± 8.6

Firstly, it is completely consistent in the issue of rank between calculation results by WNNM optimization algorithm and clustering effectiveness of NMF, which shows two characteristics of having the highest average value and relatively small standard deviation. It indicates that model of LRMR with WNNM optimization algorithm can achieve precise recovery, and the embedded dimension of well logging data can be accurately estimated. We notice that the estimated result is consistent with the parameter dimension commonly chosen by experts in the field. The parameters POR, PERM and SO are commonly used in combination with the upper and lower limits of electrical properties of other logging indicators to determine reservoir categories [41]. Therefore, the method of estimating the embedded features in well logging data has good interpretability and high precision. By mathematically constructing the intrinsic structure of the original data and accurately representing low-dimensional features, it reduces the randomness and subjectivity in setting the feature dimensions for NMF or GNMF.

Due to good adjustment function for singular values, the WNNM algorithm can obtain the accurate low-rank part of well logging data. It reveals the optimal number of latent features in multidimensional well logging data. However, the low-rank part can not only reveal the intrinsic structure of the data but also express its local features. NJW algorithm is employed to partition the low-rank part to represent the local feature of the reservoir. Then three local features are selected to construct the basis matrix for GNMF. Figure 2 illustrates the spatial geometric distribution of the low-dimensional features extracted from the proposed method. The (a), (b), (c), (d), (e), (f) and (g) in Figure 2 correspond to the oilsk81, oilsk83, oilsk85, oilsk81-83, oilsk81-85, oilsk83-85, and oilsk81-83-85 datasets, respectively.

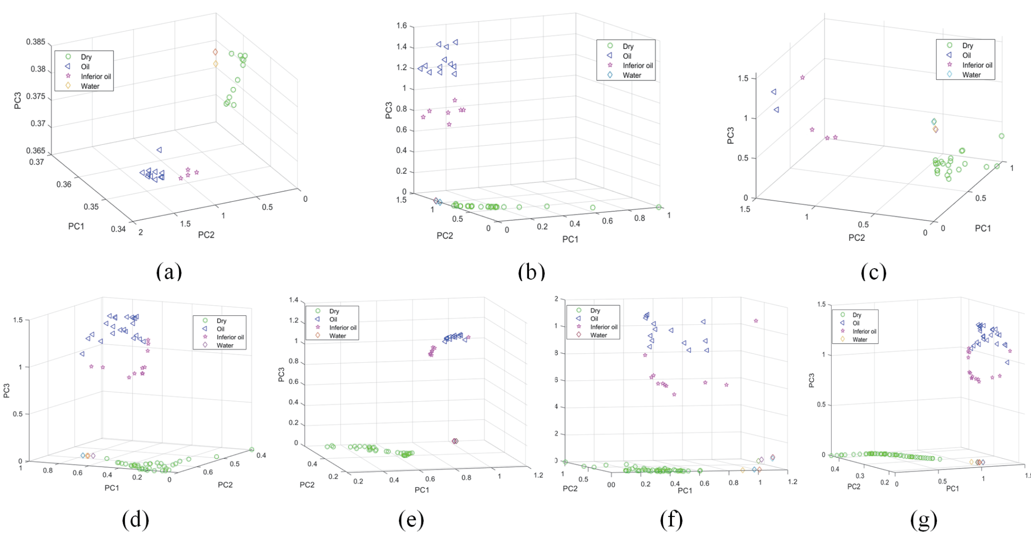

Figure 2: The visualization of well logging data via extracted low dimensional features

Firstly, it can be seen in Figure 2 that under the extraction of three features from well logging data, samples of oil layer, inferior oil layer, dry layer, and water layer are distributed in different spatial regions, showing the characteristic of compact distribution of samples from the same reservoir type and significant separation of samples from different reservoir types. The new low-dimensional feature space makes it easier to distinguish well logging data. On the one hand, it is due to the GNMF model capturing the local geometric structure information of well logging data, and on the other hand, it is due to the complete representation of the intrinsic structure within the matrix factors. Secondly, due to significant differences in features, the oil layer, inferior oil layer, dry layer, and water layer are completely separated in the feature space, which is beneficial for making decisions about whether underground reservoirs contain oil. Finally, in the feature space, although the data of the oil layer and poor oil layer exhibit good separability, there are still individual misclassified data points. This is because the interpretation of the inferior oil layer itself is a compromise concept, which is a method of interpretation that maximally retains oil-bearing reservoirs while ensuring the non-omission of oil layers. In conclusion, in the extracted feature space, multidimensional well logging data contains sufficient discriminative information and exhibits a clear spatial geometric distribution.

To quantitatively demonstrate the feature quality extracted by the proposed algorithm in this paper, some related algorithms such as NJW, NMF, GNMF, NJW-NMF and Ours, are selected to compare based on two evaluation metrices under the K-means clustering. As we know, NJW-NMF model employs the clustering centers obtained from NJW algorithm as the initial basis matrix for NMF model. That is to say, the rank of basis matrix is selected as 4 in NMF model for the well logging data. Similar to this, we set the same rank value of basis matrix in NMF and GNMF model, and the local features are randomly generated. In our proposed model, the rank and local features in basis matrix are calculated and represented from low-rank parts which is separated by LRMR model and WNNM algorithm. Each algorithm is performed independently for 20 times, and the ACC and NMI values are recorded to calculate the corresponding means and standard deviations. The results are presented in Table 4-6. The bold entries in the Table 4-6 highlight the best comparative results.

Table 4: Comparison of clustering performance with different algorithms.

	oilsk81		oilsk83		oilsk85	
Algorithms	ACC	NMI	ACC	NMI	ACC	NMI
NJW	74.1 ± 8.1	51.9 ± 6.1	79.8 ± 3.9	72.8 ± 1.1	61.8 ± 0.0	53.1 ± 2.6
NMF	91.2 ± 5.4	85.5 ± 5.4	83.1 ± 6.8	74.3 ± 4.8	75.4 ± 7.4	68.1 ± 8.6
GNMF	93.2 ± 7.6	88.2 ± 6.3	88.5 ± 8.4	81.7 ± 8.1	82.1 ± 7.1	72.7 ± 4.1
NJW-NMF	92.2 ± 8.6	89.1 ± 9.1	86.3 ± 11.5	77.5 ± 10.9	81.2 ± 9.6	76.4 ± 11.5
Ours	$\mathbf{96.8 \pm 0.0}$	$\mathbf{90.5 \pm 0.0}$	$\mathbf{90.5 \pm 5.2}$	$\mathbf{82.1 \pm 4.1}$	$\mathbf{85.3 \pm 0.0}$	$\mathbf{78.2 \pm 0.0}$

Table 5: Comparison of clustering performance with different algorithms.

	oilsk81-83		oilsk81-85	
Algorithms	ACC	NMI	ACC	NMI
NJW	70.1 ± 5.4	50.4 ± 4.6	67.3 ± 3.9	59.3 ± 3.3
NMF	81.3 ± 7.9	78.7 ± 4.3	81.1 ± 7.1	72.1 ± 5.6
GNMF	83.2 ± 6.4	79.4 ± 7.2	84.6 ± 2.8	74.7 ± 8.6
NJW-NMF	82.2 ± 7.2	80.6 ± 8.3	82.7 ± 9.8	73.3 ± 9.8
Ours	$\mathbf{89.8 \pm 4.5}$	$\mathbf{81.4 \pm 3.2}$	$\mathbf{87.3 \pm 4.6}$	$\mathbf{83.4 \pm 2.5}$

Table 6: Comparison of clustering performance with different algorithms.

	oilsk83-85		oilsk81-83-85	
Algorithms	ACC	NMI	ACC	NMI
NJW	59.4 ± 1.6	52.8 ± 3.7	58.3 ± 2.4	51.6 ± 2.9
NMF	73.6 ± 4.7	70.4 ± 7.9	75.4 ± 2.7	69.8 ± 1.4
GNMF	81.8 ± 6.5	74.1 ± 3.4	80.6 ± 4.3	72.7 ± 2.6
NJW-NMF	79.8 ± 7.8	72.3 ± 10.6	78.6 ± 3.7	71.8 ± 2.4
Ours	$\mathbf{84.8 \pm 1.5}$	$\mathbf{76.1 \pm 1.6}$	$\mathbf{82.7 \pm 4.9}$	$\mathbf{74.3 \pm 2.8}$

First, it can be observed from Table 4-6 that the proposed algorithm in this paper outperforms other relevant algorithms in terms of both ACC and NMI metrics in the three well logging datasets, which demonstrates that the algorithm has a high recognition accuracy in identifying oil-bearing reservoirs. In comparing algorithms, GNMF outperforms NMF, demonstrating that the enhanced local geometric structural information in well logging data improves the discriminative capability of feature representation. GNMF outperforms NJW-NMF, indicating that in enhancing the representation capability of NMF, mining and utilizing the local geometric structural information between data is superior to optimizing the intrinsic structure representation of the basis matrix. NJW-NMF outperforms NMF, indicating that the initialization of the basis matrix can enhance the representational capability of feature extraction. In conclusion, establishing a matrix with an appropriate structure by mining and utilizing the hidden low-rank information and category structural information in well logging data can obtain a good approximation for low rank factorization model. It can enhance the representational capability of feature extraction in the GNMF model and thereby improve the recognition ability of well logging data in low-dimensional feature subspace.

6 Conclusions

To achieve feature extraction from multi-dimensional well logging data, this paper proposes a modified GNMF model with optimal basis matrix based on the low-rank part recovered accurately by WNNM algorithm, which integrates the non-negativity and low-rank characteristics and structural information of data. The intrinsic embedding dimension of well logging data is estimated in the low-rank part obtained from the LRMR model and WNNM algorithm, and it is used to construct the rank of basis matrix. Additionally, the basis vector of basis matrix, that is local feature, is represented by the class center obtained from NJW algorithm in the low-rank part. The visualized spatial distribution and comparison experiment validate the effectiveness of the proposed model and algorithm in well logging feature extraction. It is beneficial to automatically recognize the oil-bearing of reservoir by virtue of well logging data. This precise estimation is achieved by the WNNM algorithm under the assumptions of small rank values and sparse noise. However, the separation of the low-rank part for the original well logging data is based on the assumption of sparse noise in low-rank matrix recovery. Meanwhile, the estimation of rank relies on the accurate recovery for the low-rank part obtained from optimization algorithm. This somewhat limits its general applicability in practical scenario. When the structural distribution and the types of noise in the data become complex, estimating the embedded low-rank structure remains a challenging task.

References

[1] P. Belhumeur, J. Hespanha and D. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (1997) 711–720.

[2] D. Cai, X. F. He, J.W. Han and T.S. Huang, Graph regularized nonnegative matrix factorization for data representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (2010) 1548–1560.

[3] H. Chen, F.P. Nie, R. Wang and X.L. Li, Unsupervised feature selection with flexible optimal graph, *IEEE Transactions on Neural Networks and Learning Systems* 35 (2022) 2014–2027.

[4] P. Deng, T.R. Li, H.J. Wang, D.X. Wang, S.J. Horng and R. Liu, Graph regularized sparse non-negative matrix factorization for clustering, *IEEE Transaction on Computational Social Systems* 10 (2023) 910–921.

[5] C. Ding, X.F. He and H.D. Simon, On the equivalence of nonnegative matrix factorization and spectral clustering, in: *Proceedings of the SIAM International Conference on Data Mining*, 2005, pp. 606–610.

[6] B. Gao and P.A. Absil, A Riemannian rank-adaptive method for low-rank matrix completion, *Computational Optimization and Applications* 81 (2022) 67–90.

[7] S.H. Gu, Q. Xie, D.Y. Meng, W.M. Zuo, X.C. Feng and L. Zhang, Weighted nuclear norm minimization and its applications to low level vision, *International Journal of Computer Vision* 121 (2017) 183–208.

[8] H.X. Guo, X.W. Liao, K.J. Zhu, C. Ding and Y.H. Gao, Optimizing reservoir features in oil exploration management based on fusion of soft computing, *Applied Soft Computing* 11 (2011) 1144–1155.

[9] R. He, Z.Q. Xu, S.Q. Wu and M. Jia, Fast robust component analysis with rank constraint and applications, *Acta Electronica Sinica* 51 (2023) 1448–1457.

[10] R. Hedjam, A. Abdesselam and F. Melgani, NMF with feature relationship preservation penalty term for clustering problems, *Pattern Recognition* 112 (2021): 107814.

[11] L.C. Kuang, H. Liu, Y.L. Ren, K. Luo, M.Y. Shi, J. Su and X. Li, Application and development trend of artificial intelligence in petroleum exploration and development, *Petroleum Exploration and Development* 48 (2021) 1–10.

[12] D. Lee and H. S. Seung, Algorithms for non-negative matrix factorization, in: *NIPS'00: Proceedings of the 14th International Conference on Neural Information Processing Systems*, 2000, pp. 535–554.

[13] D. Lee and H. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* 401 (1999) 788–791.

[14] J.M. Li, R.Y. Wang and L.Z. Pan, An enhanced EMG-driven Musculoskeletal model based on non-negative matrix factorization, *Biomed Signal Process Control* 79 (2023): 104178.

[15] L. Li and Y. J. Zhang, A survey on algorithms of non-negative matrix factorization,*Acta Electronica Sinica* 36 (2008) 737–743.

[16] X.L. Li, G.S. Cui and Y.S. Dong, Graph regularized non-negative low-rank matrix factorization for image clustering, *IEEE Transactions on Cybernetics* 47 (2016) 3840–3853.

[17] X.L. Li, H.Y. Zhang and R. Zhang, Matrix completion via non-convex relaxation and adaptive correlation learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (2022)1981–1991.

[18] Y.J. Li, H.X. Guo, X. Liu, Y.N. Li and J.L. Li, Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data, *Knowledge-Based Systems* 94 (2016) 88–104.

[19] Y.N. Li, H.X. Guo, X. Liu and Y. Li, The feature selection algorithm based on self-adaptive differential evolution in the application of oil reservoir identification, *Systems Engineering-Theory and Practice* 35 (2015) 2968–2979.

[20] Z.H. Li, F.P. Nie, R. Wang and X.L. Li, Robust rank-one matrix completion with rank estimation, *Pattern Recognition* 142 (2023): 109637.

[21] Z.C. Lin, M.M. Chen and Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, Available:http://arXiv:1009.5055 (2010).

[22] Z.F. Liu, X.G. Zou, X.H. Tang, Z.H. Wei, W.Z. Liu and R.Q. Li, Research on denoising method of logging dada based on modified low rank matrix recovery model, *Uranium Geology* 38 (2022) 336–344.

[23] Z.H. Liu, J.J. Wang, G. Liu and L. Zhang, Discriminative low-rank preserving projection for dimensionality reduction, *Applied Soft Computing* 85 (2019): 105768.

[24] L. Lovász and M. Plummer, *Matching Theory*, AMS Chelsea Publishing, 2009.

[25] X.A. Ma, H. Xu and C.H. Ju, Class-specific feature selection via maximal dynamic correlation change and minimal redundancy, *Expert Systems with Application* 229 (2023): 129045.

[26] T.H. Oh, Y.W. Tai, J.C. Bazin, H. Kim and I.S. Kweon, Partial sum minimization of singular values in robust PCA: Algorithm and applications, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (2015) 744–758.

[27] Y.G. Peng, A. Ganesh, J. Wright, W.L. Xu and Y. Ma, RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (2012) 2233–2246.

[28] Z. L. Peng, X.N. Song and V. Stojanovic, Hysteresis quantified control for switched reaction diffusion systems and its application, *Complex & Intelligent Systems* 9 (2023) 7451–7460.

[29] Y.L. Ren, R.B. Gong, Z. Feng and M.C. Li, Valuable data extraction for resistivity imaging logging interpretation, *Tsinghua Science and Technology* 25 (2019) 281–293.

[30] S. Solorio-Fernández, J. Carrasco-Ochoa and J. Martínez-Trinidad, A survey on feature selection methods for mixed data, *Artificial Intelligence Review* 55 (2022) 2821–2846.

[31] X.N. Song, Y.L. Song, V. Stojanovic and S. Song, Improved dynamic event-triggered security control for TS fuzzy LPV-PDE systems via pointwise measurements and point control, *International Journal of Fuzzy Systems* 25 (2023) 3177–3192.

[32] P. Sun, X.N. Song and V. Stojanovic, Composite adaptive finitetime fuzzy control for switched nonlinear systems with preassigned performance, *International Journal of Adaptive Control and Signal Processing* 37 (2023) 771–789.

[33] C.F. Tsai and Y.T. Sung, Ensemble feature selection in high dimension, low sample size datasets: Parallel and serial combination approaches, *Knowledge-Based Systems* 203 (2020): 106097.

[34] J. Wen, X.Z. Fang, Y. Xu, C.W. Tian and L.K. Fei, Low-rank representation with adaptive graph regularization, *Neural Networks* 108 (2018) 83–96.

[35] S. Wild, J. Curry and A. Dougherty, Improving non-negative matrix factorizations through structured initialization, *Pattern Recognition* 37 (2004) 2217–2232.

[36] J. Wright, A. Ganesh, S. Rao, Y. Peng and Y. Ma, Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization, *Advances in Neural Information Processing Systems* 22 (2009) 2080–2088.

[37] L.J. Yang, L.L. Yan, X.G. Wei and X.H. Yang, Label consistency-based deep semi-supervised NMF for tumor recognition, *Engineering Applications of Artificial Intelligence* 117 (2023): 105511.

[38] M. Yin, H.Y. Wu, S.L. Xie and Q.Y. Yang, Self- attention adversarial based deep subspace clustering, *Acta Automatica Sinica* 48 (2022) 271–281.

[39] K.H. Yuan, Y.L. Shang, H.X. Guo, S.F. Zang and Z.H. Liu, A precise and adaptive graph regularized low rank representation model for recognizing oil-bearing, *Journal of Control Engineering and Applied Informatics* 25 (2023) 65–74.

[40] Y. Zhang, B.P. Tang and L. Deng, Mechanical fault diagnosis based on non-negative matrix factorization with spectral clustering initialization enhancer, *Chinese Journal of Scientific Instrument* 34 (2013) 2806–2811.

[41] K.J. Zhu, A study of soft-computing theory and method in petroleum exploration information management, China University of Geosciences, 2007.

Manuscript received 31 May 2024
revised 30 August 2024
accepted for publication 11 September 2024

Kehong Yuan
College of Information Engineering
Henan University of Science and Technology
Luoyang 471023, China

School of Mathematics and Statistics
Luoyang Institute of Science and Technology
Luoyang 471023, China
E-mail address: yuankh2006@126.com

Youlin Shang
School of Mathematics and Statistics
Henan University of Science and Technology
Luoyang 471023, China
E-mail address: mathshang@sina.com

YongSheng Dong
College of Information Engineering
Henan University of Science and Technology
Luoyang 471023, China
E-mail address: ysdong@haust.edu.cn

Haixiang Guo
College of Economics and Management
China University of Geoscience, Wuhan, 430074, China
E-mail address: faterdumk0732@sina.com

Zhonghua Liu
School of Information Engineering
Zhejiang Ocean University, Zhoushan, 316022, China
E-mail address: lzhua_217@163.com

A Section Title of First Appendix

Table 7: Log explanation of oilsk81 well

Layer	AC	CNL	RT	POR	SO	PERM	Conclusion
1	195	7.5	13.0	6.0	0	0	Dry
2	225	10.0	7.3	11.0	0	0	Water
3	230	14.0	5.5	12.0	0	0	Water
4	220	9.0	25.0	9.0	56	1.3	Oil
5	225	8.0	30.0	9.0	58	2.3	Oil
6	210	7.0	26.0	6.0	0	0	Dry
7	220	8.0	26.0	10.0	60	2.4	Oil
8	225	9.0	30.0	10.0	62	2.5	Oil
9	195	4.0	36.0	5.5	0	0	Dry
10	220	9.0	30.0	9.0	61	1.7	Oil
11	217	7.5	50.0	8.0	55	1.1	Oil
12	210	6.0	130.0	7.0	48	0.7	Inferior oil
13	195	4.0	100.0	5.0	0	0	Dry
14	195	4.0	70.0	5.0	0	0	Dry
15	200	6.0	90.0	6.0	0	0	Dry
16	200	4.0	130.0	6.0	0	0	Dry
17	200	4.0	90.0	5.0	0	0	Dry
18	215	9.0	25.0	9.0	54	1.6	Oil
19	195	4.0	70.0	4.0	0	0	Dry
20	200	6.0	55.0	6.0	0	0	Dry
21	200	4.0	100.0	5.0	0	0	Dry
22	240	13.5	12.0	12.0	40	2.4	Oil
23	212	8.0	36.0	8.0	60	1.5	Oil
24	197	6.0	50.0	6.0	0	0	Dry
25	202	6.0	55.0	7.0	52	0.8	Inferior oil
26	195	4.5	50.0	6.0	0	0	Dry
27	203	5.0	45.0	7.0	46	0.6	Inferior oil
28	195	6.0	50.0	6.0	0	0	Dry
29	210	7.5	20.0	8.0	57	1.2	Oil
30	201	6.0	16.0	7.0	40	0.4	Inferior oil
31	213	9.5	12.0	9.0	61	2	Oil

Table 8: Log explanation of oilsk83 well

Layer	AC	CNL	RT	POR	SO	PERM	Conclusion
1	225	10	4	10	0	0	Water
2	226	10	5	10.5	0	0	Water
3	220	8.5	6.6	9.5	0	0	Water
4	235	12	8.8	10	32	0.4	Inferior oil
5	226	13	8	9	35	0.2	Inferior oil
6	202	10	11	7	0	0	Dry
7	209	12	30	3	0	0	Dry
8	198	8	46	4	0	0	Dry
9	178	0.8	600	1.5	0	0	Dry
10	220	9	35	10	52	1.8	Oil
11	205	6	58	8	36	0.5	Inferior oil
12	216	8.3	40	10	55	2.6	Oil
13	197	3.5	120	4	0	0	Dry
14	236	11	17	9	51	1.2	Oil
15	213	6	40	5	0	0	Dry
16	235	10	30	9.5	52	2.5	Oil
17	202	6	60	5	0	0	Dry
18	206	7	40	8	50	1.6	Oil
19	192	4	130	3	0	0	Dry
20	210	8	40	7.6	53	2.2	Oil
21	205	7.5	50	7	36	0.7	Inferior oil
22	208	5	18	7	35	0.8	Inferior oil
23	225	7	15	9	50	1.2	Oil
24	190	2	53	3	0	0	Dry
25	212	5	30	7	30	0.5	Inferior oil
26	200	4	40	2	0	0	Dry
27	201	4	46	2.9	0	0	Dry
28	195	3.5	100	3	0	0	Dry
29	199	11	40	1	0	0	Dry
30	188	3.8	400	2	0	0	Dry
31	197	6	280	3	0	0	Dry
32	200	6	105	5	0	0	Dry
33	196	6	190	3	0	0	Dry
34	210	11	60	8.5	62	2.6	Oil
35	209	9	48	8	52	1.6	Oil
36	185	1.6	70	1	0	0	Dry
37	188	4	70	2	0	0	Dry
38	203	8	27	7	40	0.8	Inferior oil
39	192	5.5	98	3	0	0	Dry
40	190	4	100	2	0	0	Dry
41	191	4.3	105	3	0	0	Dry
42	188	5	70	2	0	0	Dry
43	210	8.3	30	8	60	4	Oil
44	185	3.9	85	1	0	0	Dry
45	190	5	23	4	0	0	Dry
46	211	9.5	10	7.5	61	4.3	Oil
47	199	5.2	14	2	0	0	Dry
48	205	8	12	4	0	0	Dry
49	200	5	18	3	0	0	Dry
50	211	8.5	9	7.5	50	5	Oil

Table 9: Log explanation of oilsk85 well

Layer	AC	CNL	RT	POR	SO	PERM	Conclusion
1	225	15.1	10.5	10.7	0	3.2	Water
2	224	13.4	16	10.5	0	2.9	Water
3	200	11.9	23	4.8	0	0	Dry
4	230	13	8.5	11.3	0	3.5	Water
5	245	15.7	12	14.8	48	8.1	Inferior oil
6	230	17.5	0	11.3	0	3.8	Water
7	203	7.2	18	5.2	0	0	Dry
8	201	8.1	20	4.8	0	0	Dry
9	208	6.6	16	6.8	35	1	Inferior oil
10	205	9	36	6.1	39	0.9	Inferior oil
11	200	8.1	33	5	0	0	Dry
12	195	9.8	34	3.8	0	0	Dry
13	175	12.4	360	0.1	0	0	Dry
14	190	11.1	100	0.3	0	0	Dry
15	200	14	50	5	0	0	Dry
16	195	12.9	90	3.8	0	0	Dry
17	199	11.5	100	4.7	0	0	Dry
18	190	16.6	100	2.7	0	0	Dry
19	180	8.9	300	0.5	0	0	Dry
20	230	13	40	11.8	59	3.5	Oil
21	200	14.6	160	4.9	0	0	Dry
22	215	12.1	80	8.4	60	2.2	Oil
23	188	8.6	90	2.3	0	0	Dry
24	188	11.3	150	2.3	0	0	Dry
25	200	11.5	165	5	0	0	Dry
26	190	9.5	180	2.7	0	0	Dry
27	198	10.8	60	4.5	0	0	Dry
28	195	9.8	90	3.8	0	0	Dry
29	193	9.4	35	3.4	0	0	Dry
30	195	10	32	3.8	0	0	Dry
31	195	11.6	390	3.8	0	0	Dry
32	197	8.8	100	4.3	0	0	Dry
33	207	8.6	60	6.6	46	1.8	Inferior oil
34	185	10.3	100	1.6	0	0	Dry