

引用格式:李伟,唐放,侯博恒,等.基于神经网络的南海东部砂岩油藏采收率预测方法[J].油气藏评价与开发,2021,11(5):730-735.

LI Wei, TANG Fang, HOU Boheng, et al. A method for oil recovery prediction of sandstone reservoirs in the eastern South China Sea based on neural network[J]. Petroleum Reservoir Evaluation and Development, 2021, 11(5):730-735.

DOI:10.13809/j.cnki.cn32-1825/te.2021.05.010

基于神经网络的南海东部砂岩油藏采收率预测方法

李伟¹,唐放¹,侯博恒¹,钱银²,崔传智²,陆水青山²,吴忠维²

(1.中海石油(中国)有限公司深圳分公司,广东 深圳 518067;

2.中国石油大学(华东)非常规油气开发教育部重点实验室,山东 青岛 266580)

摘要:目前南海东部砂岩油藏采收率预测多采用数值模拟和线性回归等方法,这些方法分别存在耗时长和精度低的缺点。为了快速、准确地预测油藏采收率,选择50个已开发油藏作为数据样本,在利用主成分分析对采收率影响因素进行特征提取的基础上,运用神经网络回归法,建立了适用于南海东部海相砂岩油藏的采收率预测模型。通过与支持向量机回归和线性回归两种方法建立的采收率预测模型的预测结果对比表明,神经网络回归模型预测结果具有较高的预测精度,能够快速评价此类油藏的开发潜力。

关键词:神经网络;南海东部;砂岩油藏;采收率;预测模型;主成分分析

中图分类号:TE327

文献标识码:A

A method for oil recovery prediction of sandstone reservoirs in the eastern South China Sea based on neural network

LI Wei¹, TANG Fang¹, HOU Boheng¹, QIAN Yin², CUI Chuanzhi², LU Shuiqingshan², WU Zhongwei²

(1. CNOOC Shenzhen Company, Shenzhen, Guangdong 518067, China; 2. MOE Key Laboratory of Unconventional Oil & Gas Development, China University of Petroleum, Qingdao, Shandong 266580, China)

Abstract: At present, the methods such as numerical simulation and linear regression are mostly used to predict the oil recovery of sandstone reservoirs in the eastern South China Sea, but some of them are time-consuming or with low accuracy. In order to predict the oil recovery quickly and accurately, 50 developed reservoirs are selected as the data samples. Based on the feature extraction of the influencing factors of the principal component analysis on recovery, a recovery prediction model suitable for sandstone reservoirs in the eastern South China Sea is established by the neural network. Compared with that of two methods of support vector machine regression and linear regression, the prediction results of neural network regression model have high prediction accuracy, which can evaluate the development potential of the similar reservoirs quickly.

Keywords: neural network, eastern South China Sea, sandstone reservoir, oil recovery, prediction model, principal component analysis

我国南海东部海相砂岩油藏具有构造幅度低、储层物性好、天然水体能量充足等特点,大多数油田实现了高速开发,与陆上油田的开发方式存在差异^[1],以往针对陆上水驱砂岩油藏的采收率经验公式

不再适用于海相砂岩油藏采收率的预测。此外,常见的采收率预测方法包括数值模拟^[2]和室内实验^[3-4],这些方法往往需要耗费大量的时间和人力,无法快速对油藏采收率进行预测。

收稿日期:2021-04-26。

第一作者简介:李伟(1972—),男,本科,高级工程师,主要从事油气田开发研究工作。地址:广东省深圳市南山区后海滨路(深圳湾段)3168号中海油大厦A座,邮政编码:518067。E-mail:liweil@cnooc.com.cn

通信作者简介:崔传智(1970—),男,教授、博士生导师,主要从事油气渗流理论、油气田开发技术等研究工作。地址:山东省青岛市黄岛区长江西路66号中国石油大学(华东)非常规油气开发教育部重点实验室,邮政编码:266580。E-mail:ccz2008@126.com

基金项目:国家自然科学基金项目“致密油藏多段压裂水平井时空耦合流动模拟及参数优化方法”(51974343);青岛市博士后应用研究项目“致密油藏体积压裂支撑剂分布模拟与参数优化研究”(qdy20200084)。

近年来,许多学者通过多元线性回归^[5-6]、多元非线性回归^[7]和支持向量机回归^[8]方法,针对海上水驱砂岩油藏开展了采收率经验公式研究。通常这些方法都是基于少量样本建立的预测模型,经过校正后可用于快速预测油藏采收率,但是当样本数量增多时,往往会出现拟合效果降低的情况,导致预测精度下降。因此,为了准确描述复杂的油藏系统中采收率与其影响因素之间的关系,首先针对影响因素进行主成分分析^[9-10],避免影响因素之间的相关性,降低数据冗余性,然后运用神经网络回归^[11-14]方法,建立适合南海东部海相砂岩油藏的采收率预测模型,以快速评价油藏开发潜力,指导油田生产。

1 神经网络回归原理

神经网络是一个由模仿人脑神经元的节点组成的网络^[15],输入数据经过输入层、隐藏层、输出层的仿神经元结构进行训练^[16]。神经网络利用权重值(w)模拟神经元的关联,不同输入信号的重要性用权重来表示。神经元节点会接收经过加权处理后的信号并考虑偏置(b),计算得到加权和,权重越大表示信号越重要。神经网络以该方式存储信息,从而迭代调整权重参数以减少误差,最终得到输出目标的预测模型^[17-20]。

建立一个多层神经网络模型,首先要将选择的数据样本分为特征数据集和目标数据集,作为输入样本放入输入层后按照一定比例分为训练集、验证集和测试集。然后,设置隐藏神经元的数量和选择训练算法^[21]。最后,不断地训练网络,根据数据样本的回归效果,输出符合精度要求的神经网络回归模型(图1)。研究针对的问题是数据的回归分析,因此用于函数拟合的标准网络是一个双层前馈网络,在隐藏层有一个 sigmoid 传递函数,在输出层有一个线性传递函数。

2 基于主成分分析法的特征提取

2.1 目标油藏的选取

南海东部已开发的海相砂岩油藏为本次研究对象,该区域油藏的边底水能量强,以天然能量开发为主。通过整理分析实际油田的动、静态资料,优选部分油藏作为神经网络回归的样本。选择样本主要遵

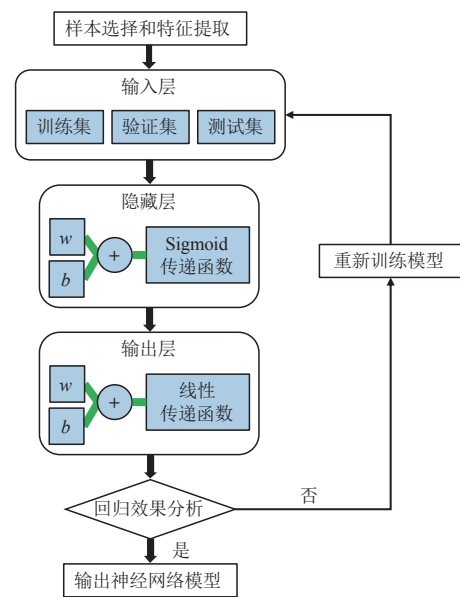


图1 神经网络回归流程

Fig. 1 Flow chart of neural network regression

循以下原则:①天然能量开发的砂岩油藏;②油藏驱动类型一致;③原油性质相近;④油藏地质特征和开发动态资料准确。

2.2 采收率的影响因素选取

采收率是反映油藏开发潜力的重要指标。研究采收率的影响因素并对采收率进行合理预测将有助于油藏开发潜力评价。采收率的影响因素包括客观因素和可控因素,客观因素一般指油藏的地质参数和流体性质,而可控因素一般指井网密度、采油速度等。

常规的采收率预测方法在实际使用过程中,往往会面临输入参数过少或过多的情况,同时每个输入参数之间可能存在相关性,导致信息重叠,增加了模型的复杂程度,还影响了预测精度。因此,针对采收率的影响因素进行优化分析是模型预测准确的前提。

根据南海东部已开发的海相砂岩油藏的统计资料,选取50个驱动类型为边水的油藏(表1),可获取的采收率影响因素包括:含油面积、油层厚度、孔隙度、含油饱和度、原始地层压力、流度、井网密度、渗透率和地质储量。由于这些参数之间可能存在相互关联,故采用主成分分析法对采收率影响因素进行特征提取,为后续建立采收率预测模型奠定基础。

表1 神经网络回归训练样本数据
Table 1 Training samples of neural network regression

油藏名称	含油面积 (km ²)	油层厚度 (m)	孔隙度 (%)	含油饱和度 (%)	地层压力 (MPa)	流度 (10 ⁻³ μm ² /mPa·s)	井网密度 (口/km ²)	渗透率 (10 ⁻³ μm ²)	地质储量 (10 ⁴ m ³)	采收率 (%)
N1	7.2	3.5	15.8	61.0	28.6	822.1	0.98	271.3	164	76.5
N2	7.5	6.4	16.2	63.4	29.0	737.9	0.79	317.3	368	67.8
N3	8.9	2.0	16.5	61.1	29.0	656.5	0.13	282.3	134	64.5
N4	7.8	2.9	22.2	63.6	18.5	196.2	0.14	1 436.4	305	47.9
N5	5.5	5.2	15.4	61.6	29.2	534.2	1.09	192.3	193	73.4
N6	9.8	5.6	21.6	60.6	14.3	138.5	1.96	637.0	681	61.9
N7	6.4	2.8	19.9	46.5	16.0	143.5	3.21	660.0	149	62.2
N8	3.7	3.4	20.5	53.2	16.4	123.7	2.70	569.0	127	51.9
N9	10.9	5.6	21.3	75.8	20.1	345.4	2.11	1 088.0	912	67.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N50	2.6	3.3	23.0	63.2	20.2	651.0	0.50	2 929.7	129	44.4

2.3 主成分分析

由于采收率各个影响因素具有不同的数量级和量纲,故首先需要对特征参数进行标准化处理,以消除数量级和量纲差异可能导致的影响。采用Z-Score标准化:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

式中: Z 为标准化处理后的参数; μ 为样本数据的均值; σ 为样本数据的标准差; X 为样本特征的观测值。

主成分分析(PCA)可以将原来多个特征信息浓缩成几个概括性特征,即降低数据特征的维度并保留原始数据的大部分信息,分析步骤为:

输入经过标准化处理后的 n 维特征数据集为 $X = \{X_1, X_2, \dots, X_n\}$,先计算出均值 $E(X)$,得到协方差矩阵 $D(X)$ 。构建出新变量 Y 为:

$$\begin{cases} Y_1 = a'_{11}X = a_{11}X_1 + a_{21}X_2 + \dots + a_{n1}X_n \\ Y_2 = a'_{12}X = a_{12}X_1 + a_{22}X_2 + \dots + a_{n2}X_n \\ \vdots \\ Y_n = a'_{1n}X = a_{1n}X_1 + a_{2n}X_2 + \dots + a_{nn}X_n \end{cases} \quad (2)$$

Y_i 是 X_i 的线性组合,且 Y 中各成分互不相关。主成分分析相当于一个条件极值问题,应满足:

$$\begin{cases} a'_i a_i - 1 = 0, & i = 1, 2, \dots, n \\ a'_i \sum a_j = 0, & i > 1, j = 1, \dots, i-1 \\ \max(a'_i \sum a_i) \end{cases} \quad (3)$$

根据式(3),计算协方差矩阵的特征值和特征向量。特征值按大小排序为 $\lambda_1, \lambda_2, \dots, \lambda_n$, a_i 为 λ_i 对应的特征向量。

依据协方差矩阵的特征值大小,选取前 k 个($k < n$)

作为主成分。通过方差解释率(或方差贡献率)来表征某一成分包含信息量的大小,其表达式为:

$$\alpha_i = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \quad (4)$$

式中: α_i 为某一主成分的方差解释率; λ_i 为某一主成分的特征值。

k 个主成分的累计方差解释率的表达式为:

$$G(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^n \lambda_j} \quad (5)$$

一般而言,当累积方差解释率大于85%时,就认为大致能反映原始数据的信息。考虑到后续建立采收率预测模型的精度,选取累积方差解释率大于95%的主成分数量。

针对南海东部海相砂岩边水油藏采收率的9个影响因素进行主成分分析(表2),最终提取得到6个主成分。该6个主成分方差解释率分别是34.19%, 23.84%, 17.71%, 10.54%, 5.23%和4.28%,累积方差解释率为95.79%,说明该6个主成分可以表达原有9个特征项共计95.79%的信息量。

3 采收率预测及影响因素分析

3.1 采收率预测模型建立

通过利用南海东部50个边水油藏的实际资料,针对含油面积、油层厚度、孔隙度、含油饱和度、原始地层压力、流度、井网密度、渗透率和地质储量等9个

表2 主成分分析结果
Table 2 Analysis results of principal component

特征编号	特征值	方差解释率(%)	累积方差解释率(%)
1	2.91	34.19	34.19
2	2.03	23.84	58.03
3	1.51	17.71	75.74
4	0.90	10.54	86.28
5	0.45	5.23	91.51
6	0.36	4.28	95.79
7	0.21	2.51	98.30
8	0.10	1.17	99.47
9	0.04	0.53	100.00

因素进行特征提取,经过主成分分析(PCA)浓缩为6个综合指标作为特征数据集,对应的采收率作为目标数据集。输入样本按照70%、15%和15%分为训练集、验证集和测试集。然后,优化隐藏神经元的数量为10。接着,通过优化训练算法选择Levenberg-Marquardt,该方法更适用于该数据样本的处理。最后,通过反复训练调整网络的权重,得到拟合度理想的神经网络回归模型,见式(6),拟合结果的相关系数为0.936 2(图2),可以看出神经网络回归模型预测采收率的相对误差基本上在15%以内(图中红色实线为45°线,黑色虚线为相对误差15%辅助线)。

$$R = \text{NeuralNetworkFunction} \left[\text{PCA} \left(A, h, \varphi, S_{oi}, P, \frac{K}{\mu}, S, K, N_o \right) \right] \quad (6)$$

式中: R 为采收率,%; A 为含油面积 km^2 ; h 为油层厚度,m; φ 为地层孔隙度,%; S_{oi} 为原始含油饱和度,%; P 为原始地层压力,MPa; K/μ 为流度, $10^{-3} \mu\text{m}^2/\text{mPa}\cdot\text{s}$; S 为井网密度,口/ km^2 ; K 为渗透率, $10^{-3} \mu\text{m}^2$; N_o 为地质储量, 10^4m^3 。

3.2 模型验证及方法对比

为了更好地对比分析基于神经网络建立采收率预

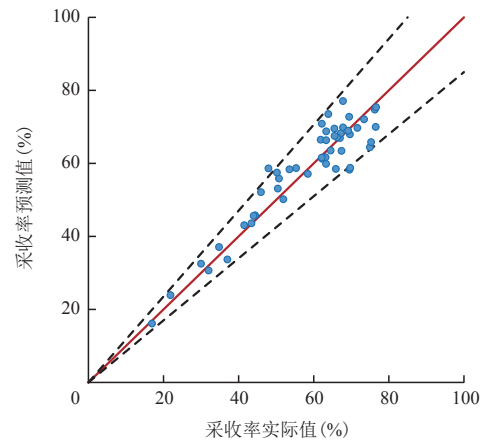


图2 神经网络回归预测采收率拟合
Fig. 2 Fitting of predicted recovery ratio by neural network regression model

测模型的优越性,笔者还采用了目前常见的支持向量机回归和线性回归两种方法建立了采收率预测模型。

除去表1用于各方法回归的50个边水油藏样本之外,遵循样本选择原则,重新选取了南海东部6个边水砂岩油藏(表3)作为模型测试样本,分别采用神经网络预测模型、支持向量机回归模型和线性回归模型对测试样本进行采收率预测。

由不同模型预测的采收率结果(表4)可以看出,基于神经网络的采收率预测模型计算出的油藏采收率相对误差均低于其他两种方法,表明神经网络模型的预测值更接近实际值,预测精度更高。

3.3 采收率影响因素分析

基于神经网络预测模型,采用正交试验设计分析影响采收率的主控因素。由于地质储量受含油面积、油层厚度、孔隙度、含油饱和度等因素控制,故分析时不做考虑。根据油藏样本数据中影响因素的变

表3 模型测试样本数据
Table 3 Model test sample data

油藏名称	含油面积 (km^2)	油层厚度 (m)	孔隙度 (%)	含油饱和度 (%)	地层压力 (MPa)	流度 ($10^{-3} \mu\text{m}^2/\text{mPa}\cdot\text{s}$)	井网密度 (口/ km^2)	渗透率 ($10^{-3} \mu\text{m}^2$)	地质储量 (10^4m^3)	采收率 (%)
H1	4.9	3.4	18.3	53.1	17.1	75.5	2.66	2 393.7	155	46.97
H2	5.2	4.4	24.2	75.6	18.2	369.9	4.42	3 587.7	402	53.87
H3	8.2	4.1	23.7	73.3	20.2	435.5	3.02	2 526.1	559	72.30
H4	9.1	4.3	25.0	71.8	22.3	534.2	3.13	3 098.3	668	69.72
H5	10.4	10.5	21.6	66.1	24.2	439.9	5.98	1 659.8	1 485	69.85
H6	2.4	2.0	21.9	58.8	19.6	284.3	1.60	1 648.7	60	64.56

化范围和集中度,确定各因素的水平(表5)。采用 $L_{32}(4^9)$ 正交表,共32组试验,然后利用预测模型计算各组试验的采收率,最后通过方差分析法确定各个因素对南海东部海相砂岩油藏采收率的影响程度(表6)。由结果可知,采收率的主要影响因素为油层厚度、井网密度、流度和渗透率。

因为南海东部砂岩油藏主要是依靠天然能量开发,所以采收率的可控因素只有井网密度。针对该区域同类型的油藏,可通过神经网络模型快速预测采收率,同时还可以分析井网密度对采收率的影响,并确定一个油藏合适的井网密度。

XJ油藏是南海东部的一个边水砂岩油藏,其生产过程中主要经历了4次井网加密(图3),井网密度最大为4.1口/km²,油藏采收率为70.86%。利用建立

的神经网络模型预测XJ油藏的采收率为75.94%,相对误差为7.17%,并计算了不同井网密度下对应的采收率(图4)。通过对比可以看出,模型预测的采收率在井网密度达到4口/km²后基本保持不变,表明该油藏合适的井网密度约为4口/km²,这与实际情况相符。

4 结论

1) 采用主成分分析方法,针对南海东部的边水砂岩油藏采收率的9个影响因素(含油面积、油层厚度、孔隙度、含油饱和度、原始地层压力、流度、井网密度、渗透率和地质储量)进行特征提取,得到6个概括性指标(即主成分),可解释原始数据95.79%的信息量。

表4 不同模型采收率预测结果对比

Table 4 Comparison of results predicted by different models

油藏名称	实际采收率 (%)	神经网络预测值 (%)	相对误差 (%)	支持向量机回归预测值 (%)	相对误差 (%)	线性回归预测值 (%)	相对误差 (%)
H1	46.97	47.23	0.55	51.77	10.22	49.60	5.60
H2	53.87	58.43	8.46	56.97	5.75	60.72	12.72
H3	72.30	63.71	11.88	56.10	22.41	60.00	17.01
H4	69.72	71.46	2.50	49.99	28.30	64.38	7.66
H5	69.85	73.55	5.30	64.42	7.77	77.84	11.44
H6	64.56	59.33	8.10	58.23	9.80	55.22	14.47

表5 影响因素水平

Table 5 Influencing factor level

水平	含油面积 (km ²)	油层厚度 (m)	孔隙度 (%)	含油饱和度 (%)	地层压力 (MPa)	流度 (10 ⁻³ μm ² /mPa·s)	井网密度 (口/km ²)	渗透率 (10 ⁻³ μm ²)
1	2	2	15	0.60	16	60	0.5	500
2	5	5	18	0.65	20	300	2.5	1 500
3	8	8	21	0.70	24	540	4.5	2 500
4	11	11	24	0.75	28	780	6.5	3 500

表6 方差分析

Table 6 Results of variance analysis

分析项	偏差平方和	自由度	F值	F临界值	显著性	分析项	偏差平方和	自由度	F值	F临界值	显著性
含油面积	124.00	3.00	2.03	9.28	不显著	地层压力	61.00	3.00	1.00	9.28	不显著
油层厚度	3 324.00	3.00	54.60	9.28	显著	流度	852.00	3.00	14.00	9.28	显著
孔隙度	174.00	3.00	2.86	9.28	不显著	井网密度	1 489.00	3.00	24.47	9.28	显著
含油饱和度	131.00	3.00	2.15	9.28	不显著	渗透率	631.00	3.00	10.36	9.28	显著

注:F值是用组间均方去除组内均方的商,与F临界值(可查阅F临界值表)相比较,若F值大于F临界值,则表示分析项的影响显著。

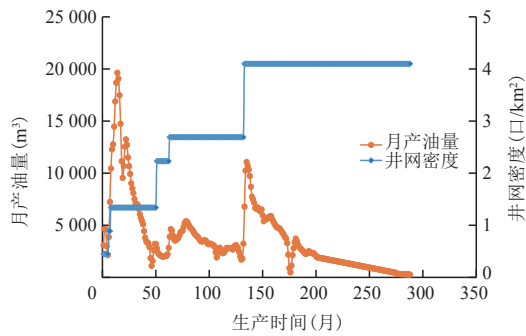


图3 南海东部XJ油藏生产历史曲线

Fig. 3 Production history curve of XJ reservoir in the eastern South China Sea

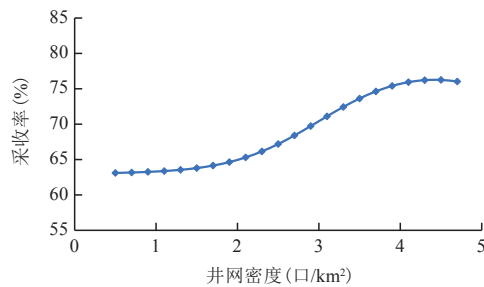


图4 模型预测采收率随井网密度变化关系曲线

Fig. 4 Relation between recovery ratio predicted by model and well pattern density

2) 基于主成分分析得到的特征数据,运用神经网络回归方法,建立了南海东部砂岩油藏采收率预测模型,与常用的支持向量机回归和线性回归两种方法预测采收率结果相比较,此模型具有更高的预测精度。

3) 基于神经网络预测模型,采用正交试验设计得到影响采收率的主控因素为油层厚度、井网密度、流度和渗透率。利用该模型可确定油藏的合理井网密度,从而指导油田生产。

参考文献

[1] 钱其豪. 海上砂岩油藏高速开发驱替特征[D]. 北京: 中国石油大学(北京), 2011.
QIAN Qihao. Marine sandstone reservoir high-speed development law[D]. Beijing: China University of Petroleum (Beijing), 2011.

[2] 王锋. 盘油田复杂断块油藏注采井网优化[J]. 西安石油大学学报: 自然科学版, 2014, 29(5): 58-60.
WANG Feng. Optimization of reasonable injection-production well pattern in complex fault block of Linpan Oilfield[J]. Journal of Xi'an Shiyou University: Natural Science Edition, 2014, 29(5): 58-60.

[3] 任耀宇, 张弦, 罗鹏飞, 等. 特低渗透轻质油藏热水驱提高采收率试验研究[J]. 非常规油气, 2019, 6(1): 69-74.

REN Yaoyu, ZHANG Xian, LUO Pengfei, et al. An experimental study of hot water flood for enhancing oil recovery in ultra-low permeability reservoir[J]. Unconventional Oil & Gas, 2019, 6(1): 69-74.

[4] 贾凯锋, 计董超, 高金栋, 等. 低渗透油藏CO₂驱油提高原油采收率研究现状[J]. 非常规油气, 2019, 6(1): 107-114.
JIA Kaifeng, JI Dongchao, GAO Jindong, et al. The existing state of enhanced oil recovery by CO₂ flooding in low permeability reservoirs[J]. Unconventional Oil & Gas, 2019, 6(1): 107-114.

[5] 冯沙沙, 王亚会, 李伟, 等. 南海东部砂岩油藏采收率经验公式的确定[J]. 石油化工应用, 2015, 34(1): 74-77.
FENG Shasha, WANG Yahui, LI Wei, et al. The determination of recovery factor empirical formula in Nanhai east sandstone reservoirs[J]. Petrochemical Industry Application, 2015, 34(1): 74-77.

[6] 段宇, 杨东东, 王美楠, 等. 渤海水驱砂岩油藏采收率经验公式研究[J]. 石化技术, 2018, 25(9): 146-147.
DUAN Yu, YANG Dongdong, WANG Meinan, et al. The research of the method on recovery efficiency empirical formula of water-drive sandstone reservoir of Bohai oil field[J]. Petrochemical Industry Technology, 2018, 25(9): 146-147.

[7] 鲁瑞彬, 刘双琪, 胡琳, 等. 水驱砂岩油田动态采收率计算新方法[J]. 西安石油大学学报(自然科学版), 2019, 34(4): 60-66.
LU Ruibin, LIU Shuangqi, HU Lin, et al. A new method for calculating dynamic recovery factor of water-flooding sandstone oilfield[J]. Journal of Xi'an Shiyou University (Natural Science Edition), 2019, 34(4): 60-66.

[8] 吴春新. 基于RS-LSSVM水驱井组采出程度计算模型——以渤海黄河口凹陷为例[J]. 新疆石油天然气, 2019, 15(3): 49-53.
WU Chunxin. Calculation model of water flooding well group recovery degree based on RS-LSSVM——Taking Huanghe River sag in Bohai as example[J]. Xinjiang Oil & Gas, 2019, 15(3): 49-53.

[9] 王敏, 陈民锋, 刘广为, 等. 主成分分析法确定海上油田水驱效果评价关键指标[J]. 油气地质与采收率, 2015, 22(2): 112-116.
WANG Min, CHEN Minfeng, LIU Guangwei, et al. Application of principal component analysis on determining the key evaluation indicators of water flooding effects in offshore oilfield[J]. Petroleum Geology and Recovery Efficiency, 2015, 22(2): 112-116.

[10] 许晓明, 李彦兰, 孙景民. 基于模糊数学的油藏干层识别研究[J]. 西南石油大学学报(自然科学版), 2019, 41(2): 45-52.
XU Xiaoming, LI Yanlan, SUN Jingmin. Study on identification of dry layers based on fuzzy mathematics[J]. Journal of Southwest Petroleum University (Science & Technology Edition), 2019, 41(2): 45-52.

[11] WANG J K, HU J H, ZHANG Y, et al. Investigation of imbibition areas during well shut-in based on mercury injection experiment and BP neural network[J]. Fuel, 2019,