

# 主成分分析法在碳酸盐岩岩性识别中的应用

## ——以 YH 地区寒武系碳酸盐岩储层为例

刘爱疆<sup>1,2</sup>, 左烈<sup>3</sup>, 李景景<sup>4</sup>, 李瑞<sup>1,2</sup>, 张玮<sup>2</sup>

(1. 成都理工大学 油气藏地质及开发工程国家重点实验室, 四川 成都 610059; 2. 成都理工大学 地球物理学院, 四川 成都 610059; 3. 中国石油 新疆油田分公司 陆梁油田作业区, 新疆 克拉玛依 834000; 4. 中国石油 新疆油田分公司 勘探开发研究院, 新疆 克拉玛依 834000)

**摘要:**在复杂岩性油气藏储层评价中,利用测井曲线直接判别复杂岩性一直存在着多解性和不确定性。利用测井曲线构建主成分变量,选取累积方差贡献率大于 85% 的主成分代表输入的多维测井信息,保证在原始信息损失最小的情况下,以少量综合变量取代原有测井信息,简化数据结构,从而解决复杂的碳酸盐岩岩性识别的难题。文中根据 YH 地区寒武系碳酸盐岩储层实测的自然伽马测井(GR)值、中子孔隙度测井(CNL)值、密度测井(DEN)值、声波时差测井(AC)值和深侧向电阻率值(RT),构建  $F_1$ — $F_5$  五个综合变量,其中  $F_1$  和  $F_2$  两个主变量占总贡献率的 87.73%,可以有效地代替原始的多维测井信息。实例表明,通过岩心薄片分析定名的岩性数据,根据  $F_1$  和  $F_2$  主成分的交会图分析,可以有效地识别碳酸盐岩的岩性。该方法在 YH 地区寒武系碳酸盐岩储层岩性识别中取得了良好的应用效果。

**关键词:**测井曲线;主成分分析;岩性识别;碳酸盐岩;寒武系

**中图分类号:**TE121.1

**文献标识码:**A

## Application of principal component analysis in carbonate lithology identification: a case study of the Cambrian carbonate reservoir in YH field

Liu Aijiang<sup>1,2</sup>, Zuo Lie<sup>3</sup>, Li Jingjing<sup>4</sup>, Li Rui<sup>1,2</sup>, Zhang Wei<sup>2</sup>

(1. State Key Laboratory of Oil and Reservoir Geology and Exploitation, Chengdu University of Technology, Chengdu, Sichuan 610059, China; 2. College of Geophysics, Chengdu University of Technology, Chengdu, Sichuan 610059, China; 3. Liliang Field Operation District, PetroChina Xinjiang Oilfield Company, Karamay, Xinjiang 834000, China; 4. Exploration and Development Research Institute, PetroChina Xinjiang Oilfield Company, Karamay, Xinjiang 834000, China)

**Abstract:** Log-based direct identification of complex lithology is still challenging in complex reservoir evaluation due to the ambiguity and uncertainty of results. The principal component variables constructed based on logging curves and having an accumulative variance contribution ratio more than 85% were selected to represent the inputs of multidimensional variables. This method can simplify the data structure and solve the problem of carbonate lithology identification through replacing the original multidimensional variables with a small amount of comprehensive variables while keeping the original information as much as possible. Five comprehensive variables including  $F_1$  –  $F_5$  are constructed by using the measured GR, CNL, DEN, AC and RT of the Cambrian carbonate reservoirs in YH field. Among these variables, the  $F_1$  and  $F_2$  principle variables account for 87.73% of the total contribution ratio, so they can replace the multidimensional variables efficiently. The case study shows that the lithology of carbonates can be effectively identified through analysis on crossplot of  $F_1$  and  $F_2$ . This method has been successfully applied to identify the lithology of the Cambrian carbonates in YH field.

**Key words:** well log, principal component analysis, lithology identification, carbonate rock, Cambrian

在油田实际应用中,岩性识别一直是储层评价的重要工作之一,是求解储层参数的基础<sup>[1]</sup>。测井信息可以确定储层岩性,储层流体性质和储层参数<sup>[2]</sup>,因此,测井信息在碳酸盐岩复杂储层识别中具有举足轻

重的作用。在两种岩性交替出现的地层中识别岩性相对简单;但当地层中的岩性比较复杂时,如含多种岩性的碳酸盐岩储层,其形成和发育的过程中要经过漫长而复杂的成岩作用,受到多期构造运动的叠加和改造,

储层非均质性非常严重,其岩性识别存在着较大的困难<sup>[3-4]</sup>,目前常用多种岩性交会图分析的方法<sup>[5]</sup>,识别复杂岩性。这种类型的交会图,在复杂的碳酸盐岩储层和火成岩储层中,能够采用直接测量的测井变量识别复杂地层的岩<sup>[6-7]</sup>,但只应用两种直接的测井变量,忽略了更多与岩性有关的变量;或者两个变量存在着明显的相关性,使交会图的精度有所降低。主成分分析是通过降维的思想,将多个原始测井变量转换成几个独立的综合性测井变量<sup>[8]</sup>,仅利用几个重要的主成分就能够代表多个测井变量之间的变化规律,这样就可以尽量减少信息的损失。本文采用主成分分析的方法,综合多种测井变量,选取代表大部分测井信息的主成分变量,然后再采用交会图方法识别岩性,有效地解决交会图分析中存在的问题。

## 1 研究区概况

塔里木盆地寒武纪的气候以干燥为主,构造活动相对稳定,其岩性主要是灰岩和白云岩<sup>[9]</sup>。根据 YH 地区 209 块薄片分析资料统计,碳酸盐岩储层主要以白云岩为主,夹少量灰岩和云灰岩。白云岩主要类型有藻白云岩、泥晶白云岩、砂屑云岩、细晶白云岩、粉晶白云岩和灰质白云岩等。白云岩中白云石含量为

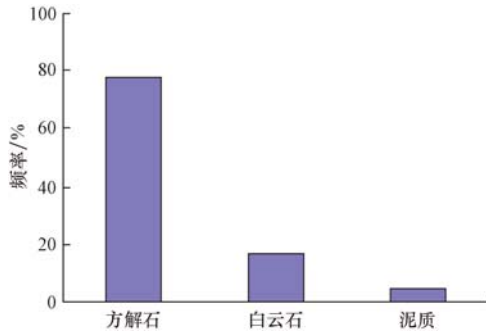


图1 YH 地区寒武系灰岩矿物成分直方图

Fig. 1 Mineral component histogram of the Cambrian limestone in YH field

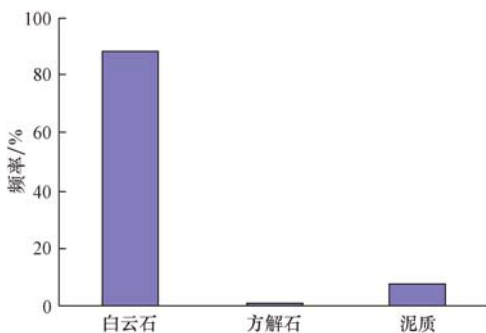


图2 YH 地区寒武系白云岩矿物成分直方图

Fig. 2 Mineral component histogram of the Cambrian dolomite in YH field

58%~100%,平均为88%;方解石含量为0~20%,平均为0.48%;泥质含量平均为7.63%;局部含硅质和黄铁矿;灰岩类型主要为泥晶灰岩、泥晶云灰岩。灰岩中方解石含量为58%~95%,平均为78%;白云石含量为1%~45%,平均为17.1%;泥质含量平均为5%(图1,图2)。

## 2 主成分分析法原理

在数据处理中,经常遇到高维数组,由于维数高、变量多,且变量间存在相关关系,因此难以抓住主要信息。为了分析这些多元数据,必须适量地减少维数<sup>[10]</sup>。主成分分析(Principal Component Analysis, PCA)就是一种数据降维的方法,它在一定程度上揭示了数据最好的解释变量的隐藏结构(主成分)。通过投影方法,将高维数据以尽可能少的信息损失投影到低维的空间,使数据降维达到简化数据结构的目的。

主成分分析的基本思路是<sup>[11-13]</sup>:设 $p$ 维随机向量 $X = (x_1, x_2, \dots, x_p)^T$ ,其均值向量为 $u$ ,协方差阵为 $V$ 。将这 $p$ 个特征向量 $x_1, x_2, \dots, x_p$ 重构成尽可能少的几个不相关的变量 $y_1, y_2, \dots, y_m$  ( $m < p$ ),并能充分反映原来 $p$ 个特征变量所反映的信息。

从 $p$ 个特征变量的 $n$ 维观测样本矩阵 $X$ 出发,求主成分的方法步骤如下<sup>[14-16]</sup>:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

1) 将原始数据进行标准化处理,即对样本集中元素 $x_{ik}$ 作变换:

$$x_{ik} = \frac{x_{ik} - \bar{x}_k}{S_k^2} \quad (i = 1, 2, \dots, n; k = 1, 2, \dots, p) \quad (2)$$

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik} \quad (3)$$

$$S_k^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \quad (4)$$

主成分分析的明显特征是每个主分量依赖于测量初始变量所用的尺度,当尺度改变时,会得到不同的特征值 $\lambda$ 。克服这个困难的方法是对初始变量进行以上标准化处理,使其方差为1。

2) 计算样本矩阵的相关系数矩阵:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (5)$$

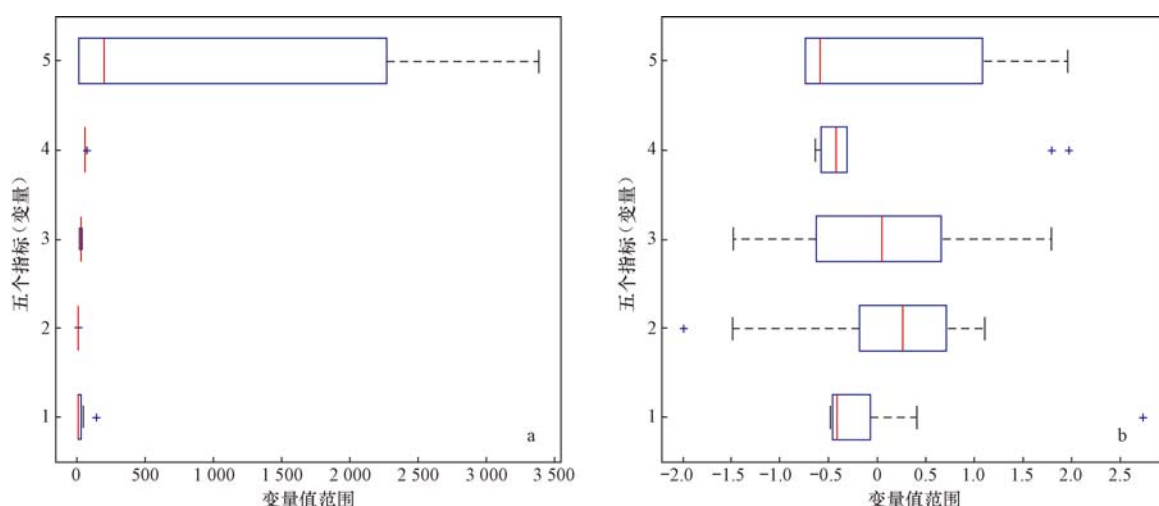


图3 原始样本数据与标准化样本数据分布范围

Fig. 3 Distribution of the original and standardized data of samples

a. 原始样本数据; b. 标准化样本数据

1. 深侧向电阻率(RT); 2. 声波时差(AC); 3. 自然伽马(GR); 4. 密度 DEN; 5. 中子孔隙度(CNL)

3) 对应于相关系数矩阵  $R$ , 用雅可比方法求特征方程  $|R - \lambda I| = 0$  的  $p$  个非负特征值  $\lambda_1 > \lambda_2 > \dots > \lambda_p \geq 0$ , 对应于特征值  $\lambda_i$  的相应特征向量为:

$$C^{(i)} = (C_1^{(i)}, C_2^{(i)}, \dots, C_p^{(i)}) \quad (6)$$

并且满足:

$$C^{(i)} C^{(j)} = \sum_{k=1}^p C_k^{(i)} C_k^{(j)} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (7)$$

4) 选择  $m$  ( $m < p$ ) 个主分量。当前面  $m$  个主分量  $y_1, y_2, \dots, y_m$  ( $m < p$ ) 的方差和占全部总方差的比例  $a = \left( \sum_{i=1}^m \lambda_i \right) / \left( \sum_{i=1}^p \lambda_i \right)$  接近于 1 时 (例如  $a \geq 0.85$ ), 选取前  $m$  个因子  $y_1, y_2, \dots, y_m$  为第 1, 2,  $\dots$ ,  $m$  个主分量。这  $m$  个主分量的方差和占全部总方差的 85% 以上, 基本上保留了原来因子  $x_1, x_2, \dots, x_p$  的信息, 由此因子数目将由  $p$  个减少为  $m$  个, 从而起到筛选因子的作用。

### 3 应用实例

利用岩石样品的自然伽马 (GR)、密度 (DEN)、声波时差 (AC)、中子孔隙度 (CNL) 和深侧向电阻率 (RT) 值作为输入参数  $X = [GR, DEN, AC, CNL, RT]$ , 以实验室 209 块岩心薄片分析岩性作为岩性判别依据 (表 1), 进行主成分分析, 可以得到的主成分特征向量的特征值和方差贡献率。

在对样本数据进行主成分分析之前, 必须对样本数据标进行准化, 即使每一个变量的均值为 0, 方差为 1 (表 2), 以消除各参数量纲和数量级的不同, 避免给计算结果带来系统误差。经过标准化后, 各样本数据的量纲和数量级在同一个范围内 (图 3), 然后根据标准化数据表计算出相关系数矩阵 (表 3)。在得出样

表 1 研究区内几种岩性的测井响应参数值

Table 1 Logging response parameter values of several lithologies in the study area

| GR/API | DEN/<br>(g · cm <sup>-3</sup> ) | AC/<br>(μs · ft <sup>-1</sup> ) | CNL/%  | RT/<br>(Ω · m) | 薄片分析<br>岩性 |
|--------|---------------------------------|---------------------------------|--------|----------------|------------|
| 21.72  | 2.72                            | 48.79                           | 20.09  | 3383.57        | 粉晶灰岩       |
| 19.56  | 2.69                            | 48.43                           | 40.53  | 2344.12        | 泥晶灰岩       |
| 24.83  | 2.68                            | 49.73                           | 138.55 | 2274.75        | 砂屑灰岩       |
| 10.92  | 2.66                            | 50.46                           | 3.88   | 5.15           | 云质灰岩       |
| 10.18  | 2.79                            | 48.62                           | 2.66   | 110.28         | 粉晶云岩       |
| 18.44  | 2.53                            | 63.86                           | 7.02   | 3.17           | 硅质云岩       |
| 32.56  | 2.75                            | 50.22                           | 2.58   | 175.39         | 泥晶云岩       |
| 15.98  | 2.76                            | 49.74                           | 4.35   | 719.41         | 细晶云岩       |
| 25.25  | 2.73                            | 49.32                           | 4.81   | 207.65         | 藻云岩        |
| 23.68  | 2.48                            | 64.99                           | 7.15   | 3.72           | 角砾岩        |

表 2 样本数据标准化值

Table 2 Standardized data of samples

| GR'   | DEN'  | AC'   | CNL'  | RT'   | 薄片分析岩性 |
|-------|-------|-------|-------|-------|--------|
| -0.07 | 0.41  | 0.21  | -0.57 | 1.96  | 粉晶灰岩   |
| 0.41  | 0.11  | -0.11 | -0.63 | 1.13  | 泥晶灰岩   |
| 2.73  | 0.01  | 0.66  | -0.42 | 1.08  | 砂屑灰岩   |
| -0.46 | -0.19 | -1.37 | -0.31 | -0.73 | 云质灰岩   |
| -0.49 | 1.11  | -1.48 | -0.60 | -0.65 | 粉晶云岩   |
| -0.38 | -1.49 | -0.27 | 1.80  | -0.73 | 硅质云岩   |
| -0.49 | 0.71  | 1.79  | -0.34 | -0.59 | 泥晶云岩   |
| -0.45 | 0.81  | -0.63 | -0.42 | -0.16 | 细晶云岩   |
| -0.43 | 0.51  | 0.72  | -0.49 | -0.57 | 藻云岩    |
| -0.38 | -1.98 | 0.49  | 1.97  | -0.73 | 角砾岩    |

表 3 相关系数矩阵

Table 3 Correlation coefficient matrix

|      | GR'    | DEN'   | AC'   | CNL'   | RT'   |
|------|--------|--------|-------|--------|-------|
| GR'  | 1.000  |        |       |        |       |
| DEN' | -0.002 | 1.000  |       |        |       |
| AC'  | 0.234  | -0.083 | 1.000 |        |       |
| CNL' | -0.201 | -0.925 | 0.084 | 1.000  |       |
| RT'  | 0.564  | 0.228  | 0.157 | -0.431 | 1.000 |

表4 特征值对应的单位化特征向量、特征值、方差贡献率

Table 4 Unit characteristic vector, characteristic values and variance contribution rate of characteristic values

| 特征向量  | 特征值    | $a_{11}$ | $a_{12}$ | $a_{13}$ | $a_{14}$ | $a_{15}$ | 方差贡献率/% | 累积贡献率/% |
|-------|--------|----------|----------|----------|----------|----------|---------|---------|
| $F_1$ | 2.2384 | 0.3198   | 0.5446   | 0.0412   | -0.6181  | 0.4663   | 51.77   | 51.77   |
| $F_2$ | 1.4979 | -0.5811  | 0.4107   | -0.5123  | -0.2720  | -0.3965  | 35.96   | 87.73   |
| $F_3$ | 0.8229 | -0.3292  | 0.2631   | 0.8557   | -0.0961  | -0.2845  | 8.06    | 95.78   |
| $F_4$ | 0.3970 | -0.6653  | -0.1313  | 0.0369   | 0.0921   | 0.7283   | 3.34    | 99.12   |
| $F_5$ | 0.0439 | 0.0953   | 0.6695   | -0.0483  | 0.7254   | 0.1185   | 0.88    | 100.00  |

本数据的相关系数矩阵  $R$  后,利用雅可比方法求取相关系数矩阵  $R$  的特征值  $\lambda_i$  及其所对应的单位化特征向量和方差贡献率(表4)。

由各特征值所对应的特征向量,可得到 YH 地区寒武系碳酸盐岩储层的主成分变换的方程式:

$$F_1 = 0.3198GR' + 0.5446DEN' + 0.0412AC' - 0.6181CNL' + 0.4663RT' \quad (8)$$

$$F_2 = -0.5811GR' + 0.4107DEN' - 0.5123AC' - 0.2720CNL' - 0.3965RT' \quad (9)$$

$$F_3 = -0.3292GR' + 0.2631DEN' + 0.8557AC' - 0.0961CNL' - 0.2845RT' \quad (10)$$

$$F_4 = -0.6653GR' - 0.1313DEN' + 0.0369AC' + 0.0921CNL' + 0.7283RT' \quad (11)$$

$$F_5 = 0.0953GR' + 0.6695DEN' - 0.0483AC' + 0.7254CNL' + 0.1185RT' \quad (12)$$

在上述转换方程中,  $F_1$ — $F_5$  相互独立,且  $F_1$  方差贡献最大,占总方差的 51.77%,  $F_2$  的贡献率为 35.96%,  $F_1$  和  $F_2$  两个主成分占总方差的 87.73%,这说明应用  $F_1$  和  $F_2$  这两个变量,可以代替原来的 5 个变量,数据信息损失很小。根据上述主成分变换模型,对研究区 209 块薄片鉴定样品进行了主成分计算,由最重要主成分  $F_1$  和次要主成分  $F_2$  对岩性进行交会图分析。在常规的自然伽马—声波时差交会图上(图4),很难区分藻云岩和灰岩。而主成分交会图上能够

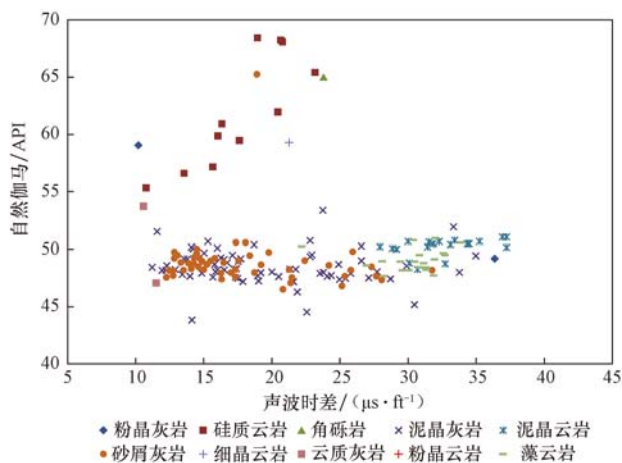


图4 自然伽马与声波时差交会识别岩性

Fig. 4 Lithology identification through GR-AC crossplotting

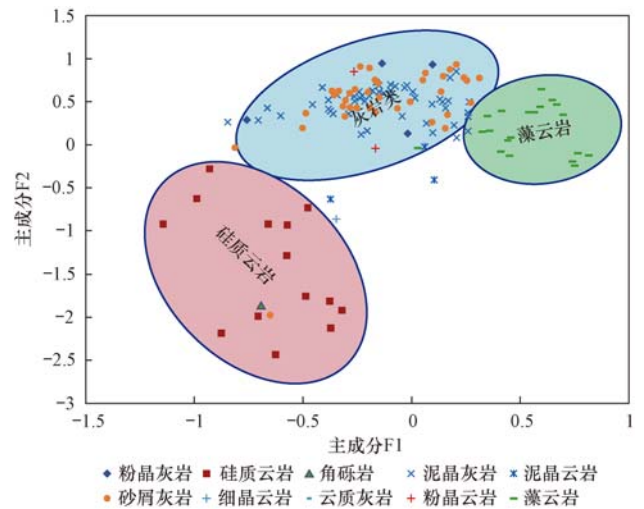


图5 主成分交会识别岩性

Fig. 5 Lithology identification through PCA crossplotting

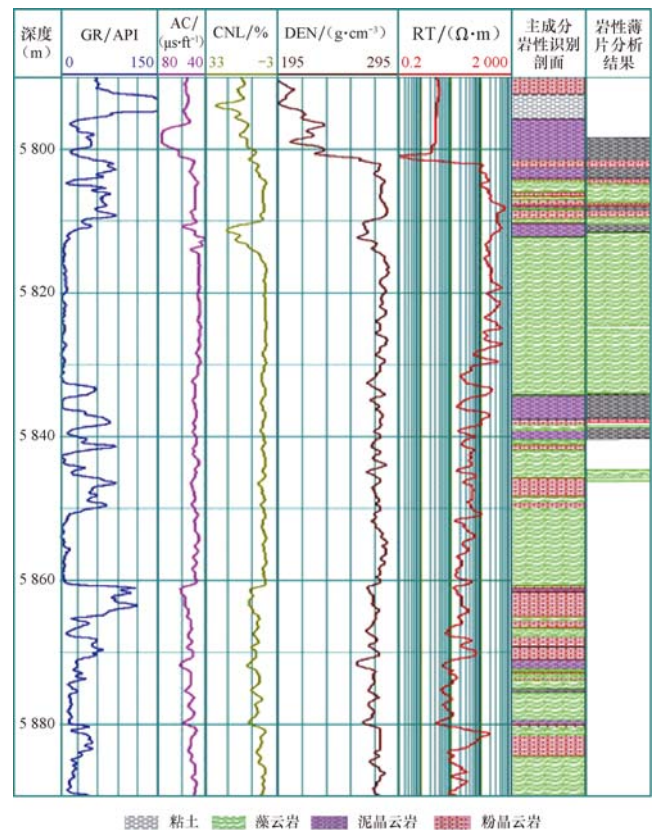


图6 研究区内某井主成分岩性识别剖面

Fig. 6 Lithology profile constructed through PCA of a well in the study area

有效的区分藻云岩、硅质云岩和灰岩类储层(图5)。根据研究区内某井的实际资料分析(图6),岩性识别符合率有很大程度的提高。

## 4 结论

1) YH地区寒武系碳酸盐岩储层岩性复杂多变,识别的关键是综合利用各种对岩性敏感的测井信息,与常规的测井交会图岩性识别方法不同,本文利用实验室岩心薄片分析资料标定测井信息,采用主成分交会图方法,把多种测井信息综合成能突出岩性特征的主成分变量,最大程度的减小了有效信息的损失。

2) 综合分析表明,选择累积方差贡献率大于85.00%的主成分变量来做交会图,能够有效的识别灰岩、藻云岩和白云岩储层,有效的解决了直接使用测井变量交会图岩性识别困难的问题。通过实例的分析,在实际应用中取得了良好的效果。

## 参 考 文 献

- [1] 张涛,莫修文. 基于交会图与模糊聚类算法的复杂岩性识别[J]. 吉林大学学报,2007,37(S1):109-113.  
Zhang Tao, Mou Xiuwen. Complex lithologic identification based on cross plot and fuzzy clustering algorithm[J]. Journal of Jilin University (Earth Science Edition), 2007, 37(S1): 109-113.
- [2] 赵军,王森,闫爽,等. 自然电位测井在低阻油气层识别中的应用——以塔里木盆地吉拉克地区三叠系低阻油气层为例[J]. 石油与天然气地质,2011,32(2):245-250.  
Zhao Jun, Wang Miao, Yan Shuang, et al. Application of self-potential logging to the identification of low resistivity reservoirs—an example from the Triassic reservoirs in Jilake area of the Tarim Basin[J]. Oil & Gas Geology, 2009, 32(2): 245-250.
- [3] 李旭兵,刘安,曾雄伟,等. 雪峰山西侧地区寒武系娄山关组碳酸盐岩储层特征研究[J]. 石油实验地质,2012,34(2):153-157.  
Li Xubing, Liu An, Zeng Xiongwei, et al. Characteristics of carbonate reservoirs of Cambrian Loushanguan Formation to the west of Xuefeng Mountain[J]. Petroleum Geology & Experiment, 2012, 34(2): 153-157.
- [4] 罗小龙,汤良杰,谢大庆,等. 塔里木盆地雅克拉断凸中生界底界不整合及其油气勘探意义[J]. 石油与天然气地质,2012,33(1):30-36.  
Luo Xiaolong, Tang Liangjie, Xie Daqing, et al. Mesozoic bottom boundary unconformity and its significance for hydrocarbon exploration in the Yakela faulted uplift, Tarim Basin[J]. Oil & Gas Geology, 2012, 33(1): 30-36.
- [5] 雍世和,张超谟. 测井资料处理与解释[M]. 山东东营:石油大学出版社,2004:80-96.  
Yong Shihe, Zhang Chaomo. Logging data processing and comprehensive interpretation[M]. Dong Ying: Press of Petroleum University, 2004: 80-96.
- [6] 陈冬,陈力群,魏修成,等. 火成岩裂缝性储层测井评价——以准噶尔盆地石炭系火成岩油藏为例[J]. 石油与天然气地质,2011,32(1):83-90.  
Chen Dong, Chen Liqun, Wei Xiucheng, et al. Log evaluation of fractured igneous reservoirs; a case study of the Carboniferous igneous reservoirs in the Junggar Basin[J]. Oil & Gas Geology, 2011, 32(1): 83-90.
- [7] 冯明刚,严丽,王雪玲,等. 元坝气田长兴组气藏有效储层物性下限标准研究[J]. 石油实验地质,2012,34(5):535-538.  
Feng Minggang, Yan Li, Wang Xueling, et al. Lower limit for physical property of effective reservoir in Changxing Formation, Yuanba Gas Field[J]. Petroleum Geology & Experiment, 2012, 34(5): 535-538.
- [8] Draper B A, Baek K, Bartlett M S, et al. Recognizing faces with PCA and ICA[J]. Computer Vision Image Understanding, 2003, 91(1): 115-137.
- [9] 肖朝晖,王招明,姜仁旗,等. 塔里木盆地寒武系碳酸盐岩层序地层特征[J]. 石油与天然气地质,2011,32(1):1-10.  
Xiao Zhaohui, Wang Zhaoming, Jiang Renqi, et al. Sequence stratigraphic features of the Cambrian carbonate rocks in the Tarim Basin[J]. Oil & Gas Geology, 2011, 32(1): 1-10.
- [10] Chattopadhyay A K, Mondal S, Chattopadhyay T. Independent component analysis for the objective classification of globular clusters of the galaxy NGC 5128[J]. Computational Statistics & Data Analysis, 2013, 57(1): 17-32.
- [11] 郭科,龚灏. 多元统计方法及其应用[M]. 四川成都:电子科技大学出版社,2003:249-251.  
Guo Ke, Gong Hao. Multivariate statistical method and application[M]. Chengdu: University of Electronic Science and Technology of China Press, 2003: 249-251.
- [12] 梅长林,范金成. 数据分析方法[M]. 北京:高等教育出版社,2006:113-125.  
Mei Changlin, Fan Jincheng. Methodology of data analysis[M]. Beijing: Higher Education Press, 2006: 113-125.
- [13] 方红卫,孙世群,朱雨龙,等. 主成分分析法在水质评价中的应用及分析[J]. 环境科学与管理,2009,34(12):152-154.  
Fang Hongwei, Sun Shiqun, Zhu Yulong, et al. Principal component analysis apply in water quality assessment[J]. Environmental Science and Management, 2009, 34(12): 152-154.
- [14] 刘小楠,崔巍. 主成分分析法在汾河水水质评价中的应用[J]. 中国给水排水,2009,25(18):105-108.  
Liu Xiaonan, Cui Wei. Application of principal component analysis method to assessment of water quality in Fenhe river[J]. China Water & Wastewater, 2009, 25(18): 105-108.
- [15] 黄传坤,王远清. 主成分分析方法的应用研究[J]. 现代商贸工业,2011,17:26-28.  
Huang Chuankun, Wang Yuanqing. Application of principal component analysis method[J]. Modern Business Trade Industry, 2011, 17: 26-28.
- [16] 吴庆岩,张爱军. 测井解释常用岩石矿物手册[M]. 北京:石油工业出版社,1998:26-115.  
Wu Qingyan, Zhang Aijun. Handbook of common logging interpretation of rocks and minerals[M]. Beijing: Petroleum Industry Press, 1998: 26-115.