

# 生产数据的整合与初步分析在钻井中的应用实例

李 谦<sup>1</sup>, 周长春<sup>1</sup>, 朱海燕<sup>2</sup>, 鲁柳利<sup>3</sup>

(1. 成都理工大学环境与土木工程学院, 四川 成都 610059; 2. 成都理工大学能源学院, 四川 成都 610059;  
3. 成都工业学院大数据与人工智能学院, 四川 成都 611730)

**摘要:** 钻井过程中的生产数据是推动产业发展的重要驱动力,也是未来人工智能在钻井行业应用的基础。当前国内外行业巨头均已开始建立生产数据的收集与分析平台,但普通生产一线作业的数据整合与分析仍未引起重视。本文以采集自南海某区域10口井44种不同参数共21912条数据为例展现了生产数据从采集至定量分析的全流程。通过图像点采算法与数据插值补齐算法,实现不同格式数据的矩阵化整合。经过标准化与可视化的处理,可完成对整合数据的定性分析,明确生产数据的规律与趋势。在此基础上,通过统计分析、相关性分析与因子分析,可获得数据特征值,也能明确不同数据间的相互关系。分析结论实现了数据的分组与降维,在保证后续数据建模、人工智能等分析精度的同时降低了建模复杂度。

**关键词:** 钻探数据;矩阵化整合;标准化;可视化;统计分析;相关性分析;因子分析;人工智能

**中图分类号:** P634;TE242 **文献标识码:** A **文章编号:** 2096-9686(2021)S1-0085-11

## Application of integration and preliminary analysis of production data in drilling

LI Qian<sup>1</sup>, ZHOU Changchun<sup>1</sup>, ZHU Haiyan<sup>2</sup>, LU Liuli<sup>3</sup>

(1. College of Environment and Civil Engineering, Chengdu University of Technology, Chengdu Sichuan 610059, China;

2. College of Energy, Chengdu University of Technology, Chengdu Sichuan 610059, China;

3. School of Big Data and Artificial Intelligence, Chengdu Technological University, Chengdu Sichuan 611730, China)

**Abstract:** The production data is an important driving force to promote the development of the drilling industry, and it is also the basis for the future application of artificial intelligence in the drilling industry. At present, all the drilling industry giants, either domestic or abroad, have begun to establish production data collection and analysis platforms. But the data integration and analysis for ordinary drilling production operations have still not attracted attention. The proposed paper takes 21912 data collected from 10 wells in a certain area of the South China Sea with 44 different parameters as an example to show the whole analyze process using production data, which from collection to quantitative analysis. The matrix integration of data in different formats can be realized from the image point sampling algorithm and data complementation algorithm. According to standardization and visualization processing, qualitative analysis of integrated data can be completed, and the law and trend of production data can be clarified. Based on the statistical analysis, correlation analysis and factor analysis, the data characteristic values can be obtained, meanwhile the interrelationship between different parameters can be clarified. Realizing the parameters grouping and dimensionality reduction, the accuracy of further data modelling can be ensured with the reduced modeling complexity.

**Key words:** drilling data; matrix integration; standardization; visualization; statistical analysis; correlation analysis; factor analysis; artificial intelligence

收稿日期: 2021-05-31 DOI: 10.12143/j.ztgc.2021.S1.013

基金项目: 中海石油(中国)有限公司项目“南海西部油田上产2000万方钻完井关键技术研究”子课题“乐东10区超高温高压井综合提速技术研究”(编号: CNOOC-KJ135ZDXM38ZJ05ZJ); 四川省科技支撑计划应用基础研究项目“四川深层页岩气产能大数据挖掘和智能评估方法研究”(编号: 2021YJ0360)

作者简介: 李谦,男,汉族,1987年生,讲师,地质资源与地质工程专业,博士,从事钻井智能化分析与应用研究工作,四川省成都市成华区二仙桥东三路1号,liqian2014@cdut.edu.cn。

引用格式: 李谦,周长春,朱海燕,等. 生产数据的整合与初步分析在钻井中的应用实例[J]. 钻探工程, 2021, 48(S1): 85-95.

LI Qian, ZHOU Changchun, ZHU Haiyan, et al. Application of integration and preliminary analysis of production data in drilling[J]. Drilling Engineering, 2021, 48(S1): 85-95.

## 0 引言

随着大数据、人工智能、5G等相关技术的飞速发展,数据已成为当前工业生产各领域发展的重要驱动力。而钻探钻井作业面对未知地层进行施工时,出于安全和成本的考虑,对数据的需求更不可同日而语。无论是石油钻井还是地质钻探行业,随着资源勘探的深入与钻探钻井技术的发展,生产一线均已累积了海量数据,为即将到来的大数据与人工智能的应用奠定了坚实的基础。

目前钻探钻井行业对数据的使用和分析处于蓄势待发的状态。来源于生产一线的数据内容繁多,数据形态各异,包含各类静态结构化表格、视频、图片与各种不同的工业数据标准。同时不同来源的数据量变化巨大,各类数据生成周期不同,采集方式各异,导致数据可管理性极差,难以用统一的数据模型容纳所有的完钻井数据<sup>[1]</sup>。为了改变这种现状,当前国内外钻井相关行业巨头均开始研发以数据分析、共享为主要目的相关平台。康菲公司开发了IDW(Integrated Data Warehouse)大数据分析平台,以多学科集中式数据存储中心为理念,实现了不同功能的数据仓库(包括地质、钻完井、油藏工程、生产、运营、财务等)的整合与跨功能集成<sup>[2]</sup>。科威特国家石油公司针对油田产能预测与生产参数优化,建立了基于数据分析和人工智能的油田智能工作平台<sup>[3]</sup>。英国石油公司(BP)的Sandy平台,打通了上下游业务,实现决策自动化<sup>[4]</sup>。壳牌与微软共同研发Geodesic平台,为水平井定向控制,并筒钻探数据处理算法提供计算支持<sup>[5]</sup>。埃克森美孚联合XTO公司与微软合作开发了收集数据、整合解决方案综合平台<sup>[6]</sup>。俄罗斯天然气公司(GASPROM)提出数字化转型计划(DT),针对全行业12类项目的数字

化提出了战略性要求<sup>[7]</sup>。雪佛龙与斯伦贝谢公司以实现技术一体化为目的开发了DELFI信息平台,整合历史数据,为各类相关专业操作和程序提供接口<sup>[7-8]</sup>。贝克休斯公司研发了BHC3平台,破除了钻井施工现场不同信息间的数据壁垒,实现了现场工况与决策方案的可视化<sup>[9]</sup>。哈里伯顿与微软合作推出了DecisionSpace 365云平台,实现了地质、钻井、油藏、生产等数据资料的快速更新与一体化分析<sup>[10]</sup>。国内石油行业主要由三桶油领衔,中石油推出了勘探开发梦想云平台,包含了大量的数据,实现了勘探开发、经营、管理、决策的一体化<sup>[11]</sup>;中石化也在逐步打造覆盖全公司的EPBP岗位业务信息化系统,基于数据授权共享机制,实现关键业务数据的共享与应用<sup>[1]</sup>;中海油则推出了智能油田技术平台,用以实现智能油田建设与勘探开发的数据治理<sup>[7]</sup>。同时地质行业自2017年起自然资源部也开始主推“地质云”国家地质大数据共享服务平台,就钻探领域已包含超过95万个钻孔地层数据和29万多米的钻孔岩心图像,但钻探数据涉及较少<sup>[12]</sup>。

尽管各大油企和油服均在企业层面推出了数据采集、分析相关整合服务平台,同时各大钻探钻井相关科研所在数据处理分析、甚至机器学习、人工智能方向持续投入大量的研发精力。但是针对生产一线,尤其是距石油钻井还有一定差距的地质钻探生产一线,对现场数据的分析利用仍处在较为初级的水平。故本文以南海某区块相关钻井数据为例,在收集10口井合计5大类44种不同参数(见表1)共21917条数据后,展示了如何通过基本数据处理(不使用人工智能的前提)对现场的工况进行初步识别与分析。

表1 收集参数列表

| 参数类型  | 参数名称/缩写   |
|-------|---|
| 井眼位置  | 井号/No.(指代井眼位置坐标);深度/D   |
| 施工工艺  | 井径/d;钻速/ROB;钻压/WOB;大钩载荷/WOH;转速/RPM;扭矩/T;泵压/SPP;泵量/Q;钻时/BT;泵时/PT;钻井液入口密度/MI;钻井液出口密度/MO;钻井液入口温度/TI;钻井液出口温度/TO                           |
| 钻井液性质 | 屈服值/YP;密度/MW;漏斗粘度/MV;塑性粘度/PV;3转读数/D3;6转读数/D6;10s静切力/SS;10min静切力/SM;滤矢量/FL;泥饼厚度/MT;pH;N;K;氯离子含量/CLC;钙离子含量/CAC;膨润土含量/SOC;固相含量/SO;含沙量/SA |
| 地质情况  | 地震速度/EV;孔隙压力/PP;破裂压力/FP;上覆压力/OP;岩性/TYP  |
| 钻头参数  | 喷嘴数量/NN;喷嘴等效直径/NS;钻头内排磨损分级/WI;钻头外排磨损分级/WO;提速钻具使用/ST   |

### 1 数据采集:生产数据的矩阵化整合

数据的矩阵化整合是在完成数据收集后进行挖掘分析的必要环节。矩阵化整合数据的主要目的在于通过一定的技术手段,将所有待分析数据转化整合为同一个数据矩阵。以钻井数据为例,整合后的数据矩阵应以参数为列,深度为行。令整合后的参数数量为  $m$ ,深度为  $D$ ,不同参数间最小的深度采样间隔为  $d$ ,则转化后的数据矩阵应为  $m$  列  $\times D/d$  行。但生产现场主要的问题在于数据格式并不统一,同时可能包含多种数据格式,需要通过不同的步骤进行数据的矩阵化整合。常见的整合方式包括图像点采和数据补齐算法。

#### 1.1 图像点采算法

由于各种不同的原因,生产现场的数据可能仅有数据曲线图片,而缺失相关的原始数据。据此研发图像点采算法,其主要目的在于根据图片刻度比例,通过计算获得原始图片文件中的初始数据。图像点采算法的核心在于建立图片像素点坐标与图示坐标系之间的比例关系,通过特征值的像素点坐标

与图示坐标系坐标,反算得到指定位置的图示坐标系坐标。以图 1 所示的地层孔隙压力曲线为例,首先确定 3 个特征点在图示坐标系与屏幕坐标系中的坐标如表 2 所示。基于该坐标关系,可分别计算单位像素在图示坐标系中对应的 X 轴尺寸  $U_X$  与 Y 轴尺寸  $U_Y$  分别如式(1)所示。

$$\begin{cases} U_X = \frac{A_{X_{end}} - A_{X_0}}{D_{X_{end}} - D_{X_0}} \\ U_Y = \frac{A_{Y_{end}} - A_{Y_0}}{D_{Y_{end}} - D_{Y_0}} \end{cases} \quad (1)$$

在获取图片在屏幕的位置后,鼠标点选指定点将获得该点在屏幕上的像素点坐标为  $(D_X, D_Y)$ ,结合式(1)与表 2,可计算该点在图示坐标系中的坐标值  $(A_X, A_Y)$  如式(2)所示。联合式(1)、式(2),可计算获得指定图片上图线任意点计算结果如图 1 所示。

$$\begin{cases} A_X = U_X(D_X - D_{X_0}) \\ A_Y = U_Y(D_Y - D_{Y_0}) \end{cases} \quad (2)$$

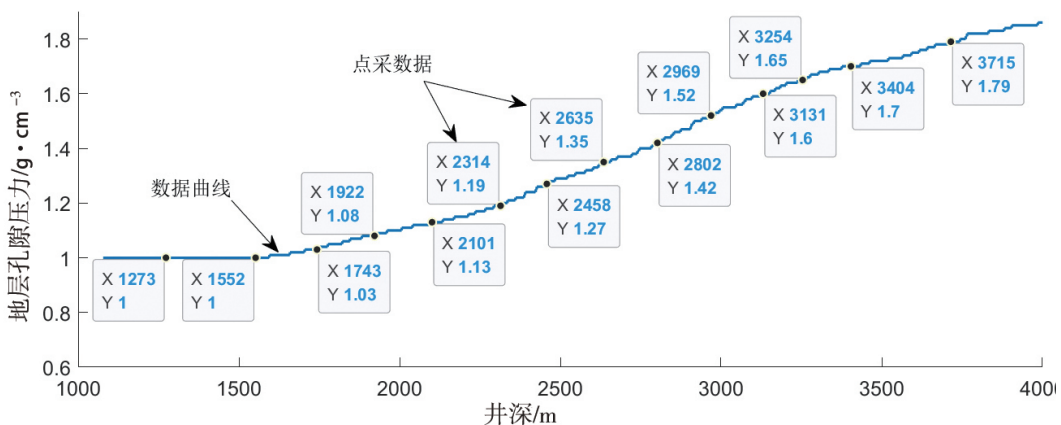


图 1 图像点采算法获取原始数据实例

表 2 图像点采算法特征点坐标

|       | 原点                                | X轴终点                                  | Y轴终点                                  |
|-------|-----------------------------------|---------------------------------------|---------------------------------------|
| 图示坐标系 | $(A_{X_0} = 1000, A_{Y_0} = 0.5)$ | $(A_{X_{end}} = 4000, A_{Y_0} = 0.5)$ | $(A_{X_0} = 1000, A_{Y_{end}} = 1.9)$ |
| 像素点坐标 | $(D_{X_0}, D_{Y_0})$              | $(D_{X_{end}}, D_{Y_0})$              | $(D_{X_0}, D_{Y_{end}})$              |

注:像素点坐标因随图片摆放位置而改变,故以字母代替

#### 1.2 数据插值补齐算法

对同一口钻孔/钻井而言,不同类型的数据采集不可避免地存在不同数据密度。例如布设了实时数据监测的钻孔/井眼可能以每米 1 条的数据密度获得工程参数(钻压、转速、扭矩等)。但对于每日定时

测量的钻井液性能而言,其记录数据密度应为每日 1~2 条数据(以测量频率为准)。对某些地质或测井资料而言,通常的数据形式为测井曲线,需要使用式 2 所示图像点采算法获得数据。其数据密度直接取决于采集点数量,可能达到数十米 1 条数据。据

此,数据插值补齐算法的主要目的在于以最密的数据密度为基准,补充缺失数据,对齐来源不同的数据矩阵。

选择常见的拉格朗日插值法作为基本的数据插值补齐算法,其基本原理在于对实际数据中的某物理量的 $k$ 个离散观测值 $(x_k, y_k)$ ,均可找到一个多项式 $y(x)$ ,使其恰好各点 $x_k$ 均取到观测值 $y_k$ ,据此其基本公式列举如式3所示<sup>[13]</sup>。经计算,数据插值补

齐的效果如图2所示,将图1所示原始数据曲线读取61条数据后进行插值补全,补全后达到3000条数据。尽管从图2所示的放大图上可知,细节上插值补全曲线与真实曲线仍存在差异(由于使用多项式代替未知的原始函数所致),但整体趋势上经插值补全后的数据曲线与原始数据曲线几乎吻合。

$$y(x) = \sum_{k=1}^n y_k \left( \prod_{\substack{j=1 \\ j \neq k}}^n \frac{x - x_j}{x_k - x_j} \right) \quad (3)$$

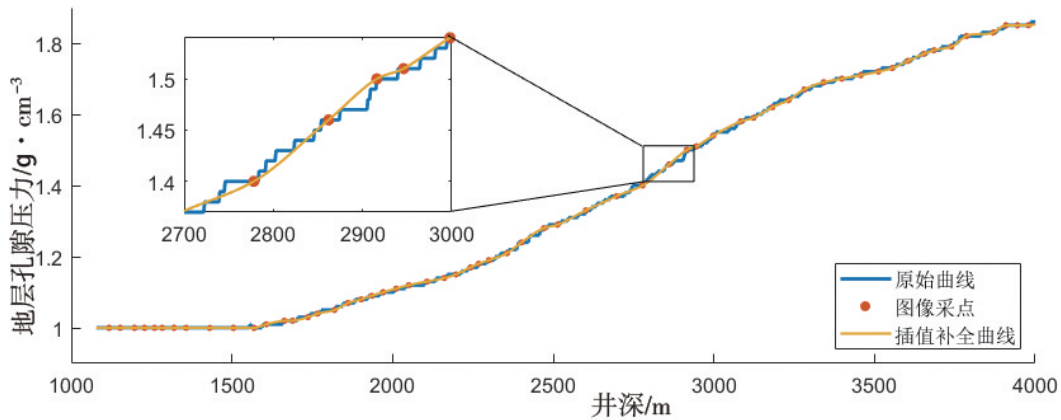


图2 数据插值补全曲线与原始数据曲线的对比

## 2 数据定性分析:生产数据标准化与可视化

### 2.1 数据标准化

生产实际现场采集的数据存在巨大的数量级差异,通过箱式图绘制本例中44组参数的数量级分布范围如图3所示,发现参数的数量级分布范围为 $10^{-1} \sim 10^3$ 。如此巨大的数量级差异,将为后续定性分析和定量分析均带来误差隐患。针对定性分析,大数量级的数据曲线将明显掩盖小数量级数据曲线的波动,导致对工况和相关程度的误判。而定量分析中数量级的差异则会直接影响建模精度。若关键参数 $A$ 影响程度大,但数量级较小,同时存在另一参数 $B$ 影响程度较小而数量级巨大,则在建模数学运算中非关键参数 $B$ 的波动将会显著掩盖关键参数 $A$ 的影响,对建模精度带来较大误差<sup>[14]</sup>。

数据标准化是消除各数据间数量级与单位差异的主要方法,其核心策略在于通过计算将所有数据压缩到同一区间。本例中采用min-max规范法进行数据的标准化,其计算核心如式4所示,通过该方式可消除所有数据单位,并将值域压缩至 $[0, 1]$ 区间。

完成标准化前后的数据曲线对比如图4所示,可见在没有改变数据波动趋势的前提下,将所有数据压缩到同一区间可以更明显地观察到数据的变化幅度。

$$x_{new} = \frac{x_{old} - \min(x_{list})}{\max(x_{list}) - \min(x_{list})} \quad (4)$$

式中: $x_{new}$ ——完成标准化的数据; $x_{old}$ ——标准化前的原始数据; $x_{list}$ ——标准化前的整个数据列。

### 2.2 数据可视化

数据可视化是针对生产数据进行定性分析的主要手段。数据可视化在于通过图形化的手段,对数据进行多维度观察,从而方便对数据进行更深入的分析,以求寻找数据变化的趋势与规律。在本文涉及的相关数据中,可观察的规律列举(但不限于)如下:

(1)寻找规律:如图5所示,在将收集的10口井眼钻速数据可视化后,发现该区域井眼在达到一定深度后(2500~3500 m)会出现明显的钻速急剧下降。并非缓降的钻速曲线,说明在该区域的钻井作业中存在共性问题,引起了钻速突变。由此可见,数



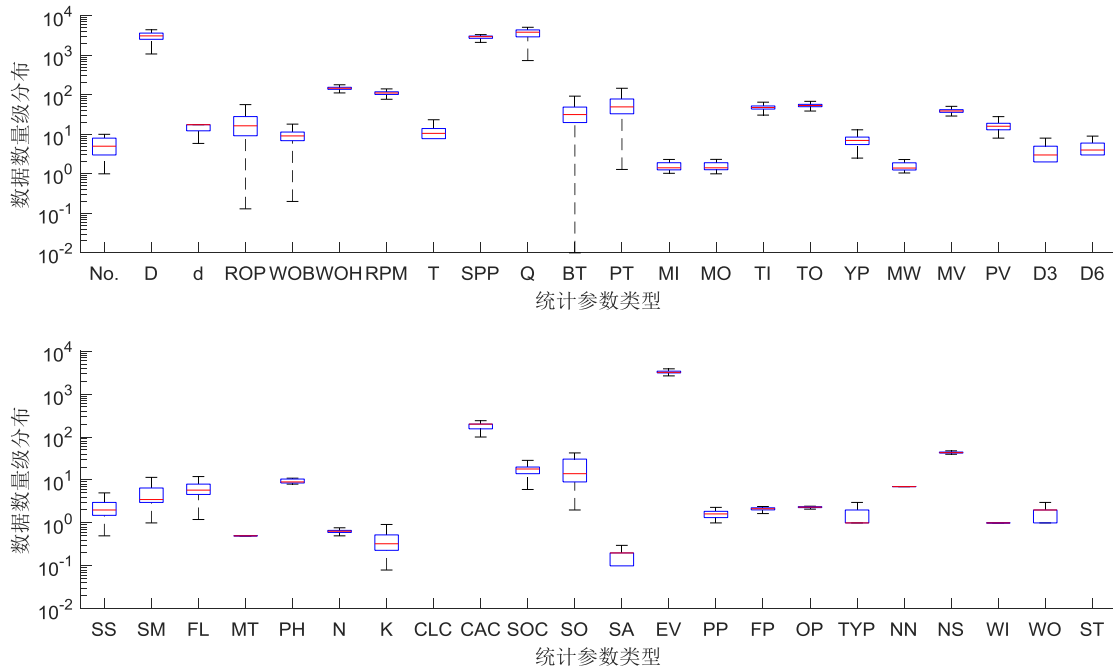


图3 统计所有参数数量级分布

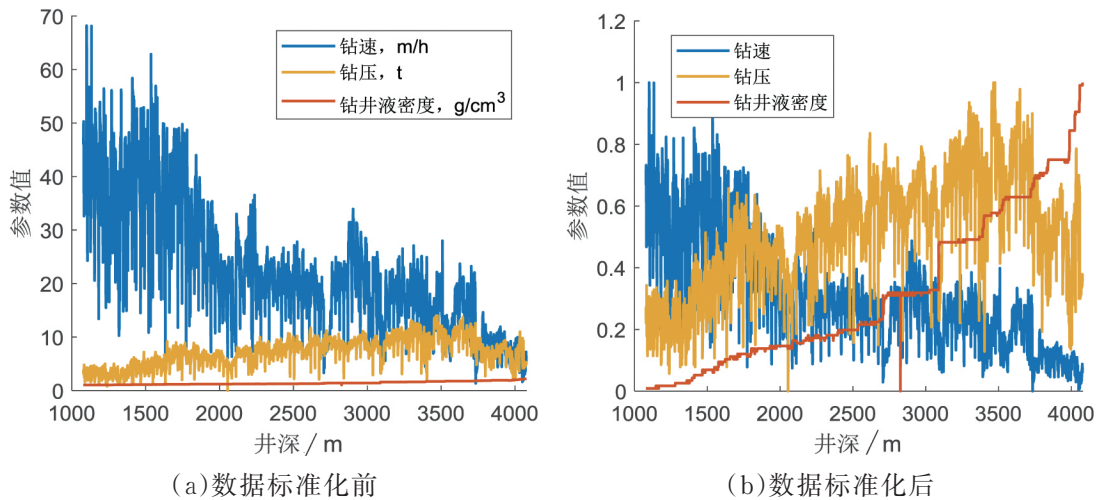


图4 数据标准化前后对比

据可视化可显著的发掘和暴露生产过程中的潜在问题。

(2)定性寻找影响因素:将10口井钻速曲线首尾相连,并将钻压、转速、泵量、钻井液密度等常规规程参数以同样的形式绘图进行对比如图6所示,可见常规理论中影响较大的钻压、转速的变化对钻速的变化影响并不明显,说明钻速的突降并非由于操作规程参数的影响而导致。与之对应,在多个钻速突降点上(图6黑色虚线),呈现出明显的钻速突降

伴随着泵量的突降与钻井液密度的大幅度提升,由此可得,泵量与钻井液密度的变化可能是引发钻速突降的主要原因之一。

### 3 数据定量分析:统计、相关性与因子分析

#### 3.1 统计分析

统计分析是生产数据定量分析的必要方法之一。基于对生产数据的统计,可充分了解生产过程和对象的特征。如图7所示,对本区域的岩性进行

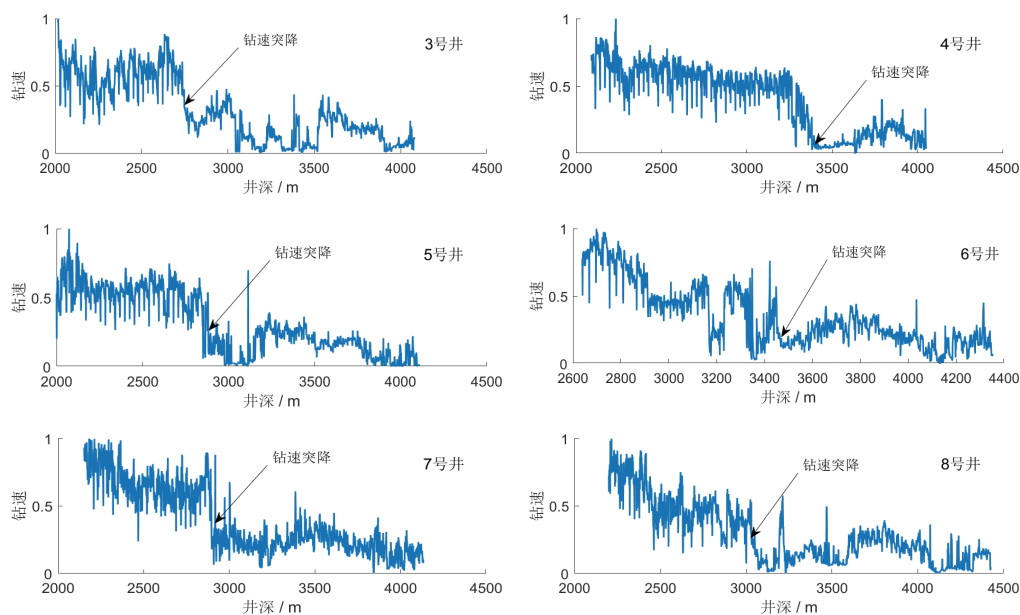


图5 基于数据可视化发现同井区均存在明显的钻速突降点

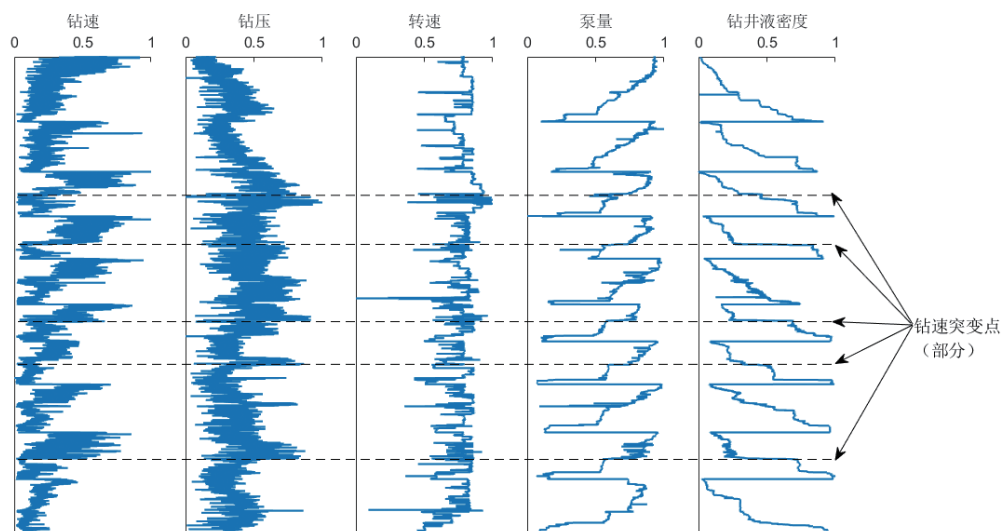


图6 数据曲线对比寻找钻速下降的影响因素

统计分析,结果显示该区域主要岩性由泥岩、粉砂质泥岩、泥质分砂岩与中砂岩构成,4种岩性合计占比超过94.5%。同时可以看出,占比最重的4种岩性中前3种都含有泥质,纯泥岩占比更是达到65%,故可认为该地层属于巨厚泥岩层,与传统油气钻井砂泥互层、占比接近的地层情况并不相似。基于此,在附近区域进行后续施工时必须对泥岩进行有针对性的设计。

结合如图6所示定性分析结论,考虑钻井液泵量和钻井液密度可同时以静液柱压力进行描述,故

将数据中所有含泥岩性的地层数据提出,绘制钻速与静液柱压力统计分析如图8所示。据此可知,在该区域内存在两个明显的钻速突降点。第一个钻速突降点位于静液柱压力约44.5 MPa附近,此时钻速将突降至30 m/h以下。第二个钻速突降点位于静液柱压力约87 MPa附近,钻速会二次突降至10 m/h以下。由此可见,通过定量统计分析结合可视化的图形显示,能够显著地观察相应的规律与对应的特征性数据。

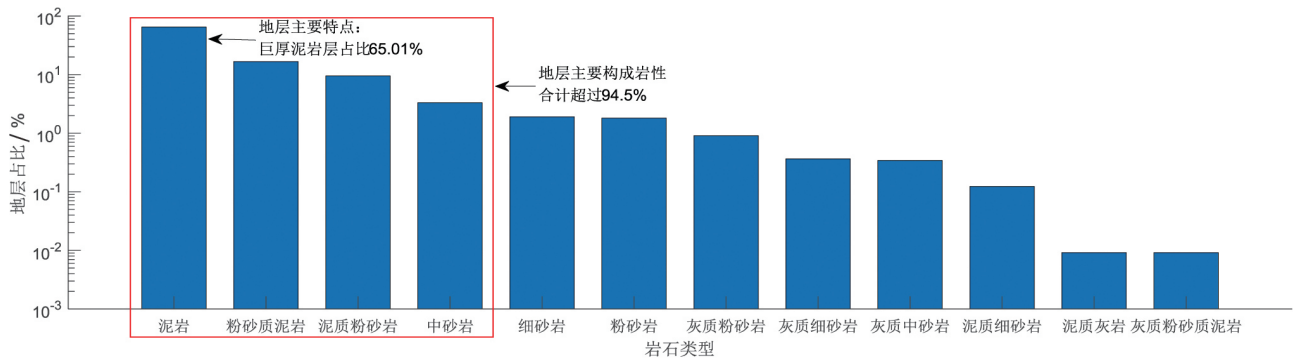


图 7 分析区域不同岩性的统计结果

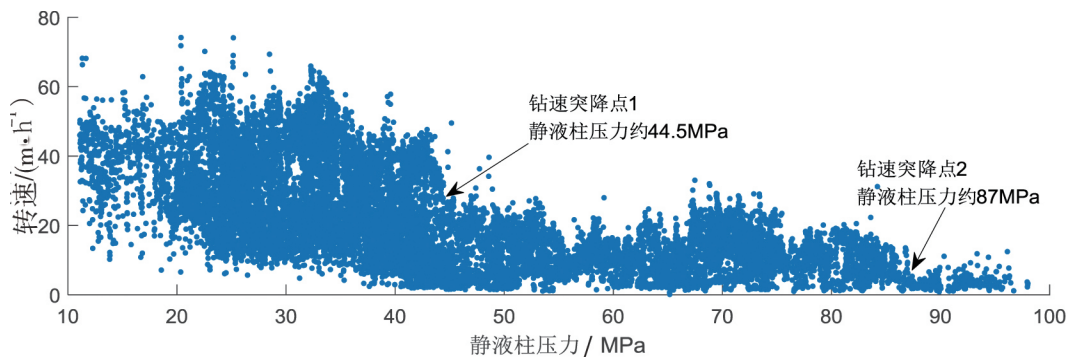


图 8 分析区域泥岩地层钻速与静液柱压力的统计关系

### 3.2 相关性分析

基本的数据可视化只能获取数据变化的基本规律与潜在影响因素分析,不能完全满足实际生产的需要。在此基础上,还需要定量计算数据之间的相关性,用以验证和精确判定前期定性分析的结论。本例使用 Person 相关系数法进行相关性的计算,则变量  $a, b$  的相关性系数  $\rho_{ab}$  计算如式 5 所示,计算结果区间为  $[-1, 1]$ 。计算结果的正负指代两个参数之间的正负相关性,当计算结果  $\rho_{ab}$  的绝对值越靠近 1,则说明变量  $a, b$  的相关性越高。相反当计算结果  $\rho_{ab}$  的绝对值越靠近 0,则说明变量  $a, b$  之间越不相关<sup>[15]</sup>。

$$\rho_{ab} = \frac{\text{cov}(a, b)}{\sigma_a \sigma_b} = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^n (b_i - \bar{b})^2}} \quad (5)$$

式中:  $\text{cov}(a, b)$ ——变量  $a, b$  的协方差矩阵;  $\sigma_a, \sigma_b$ ——变量  $a, b$  各自的标准差;  $a_i, b_i$ ——变量  $a, b$  数据集中第  $i$  个变量值;  $\bar{a}, \bar{b}$ ——变量  $a, b$  平均值;  $n$ ——变量  $a, b$  的数据集大小。

分别计算表 1 所示除钻速外所有变量与钻速的相关性,计算结果如图 9 所示,按照相关性系数的计算结果可将本例 43 种参数划分为 3 个类别:

(1)高相关性参数组(12 种参数,占有参数种类的 27.9%),该组内所有参数与钻速相关性较高 ( $|\rho_{ab}| \geq 0.6$ ),计算结果全部位于  $[0.6, 0.7]$  之间。

(2)中相关性参数组(15 种参数,占有参数种类的 34.9%),该组内所有参数与钻速相关性中等 ( $|\rho_{ab}| \in [0.3, 0.6)$ ),其中有 9 种相关性系数位于  $[0.5, 0.6]$  之间。

(3)低相关性参数组(16 种参数,占有参数种类的 37.2%),该组内所有参数与钻速相关性较低 ( $|\rho_{ab}| < 0.3$ ),其中有 8 种相关性系数  $< 0.1$ 。

单从相关性进行分析,并非现场所有数据都与钻速高度相关,因此在后续建模分析中也可有针对性的使用高相关性数据,以求获得更精确的结果。同时由图 9 也可看出,基于数据可视化定性判定的泵量与钻井液密度确实与钻速存在高相关性,从定量分析的角度验证了数据可视化定性分析的准确性。除此以外,还发现钻井液失水量 FL、钻柱扭矩 T、地层孔隙压力 PP、地层上覆压力 OP 与地层破裂

压力FP均较高,因此可判定该区域10口井钻速突降的主要原因是由地层和钻井液性能综合导致的。

其中调节控制钻井液的失水量和密度能有效的提高钻速。

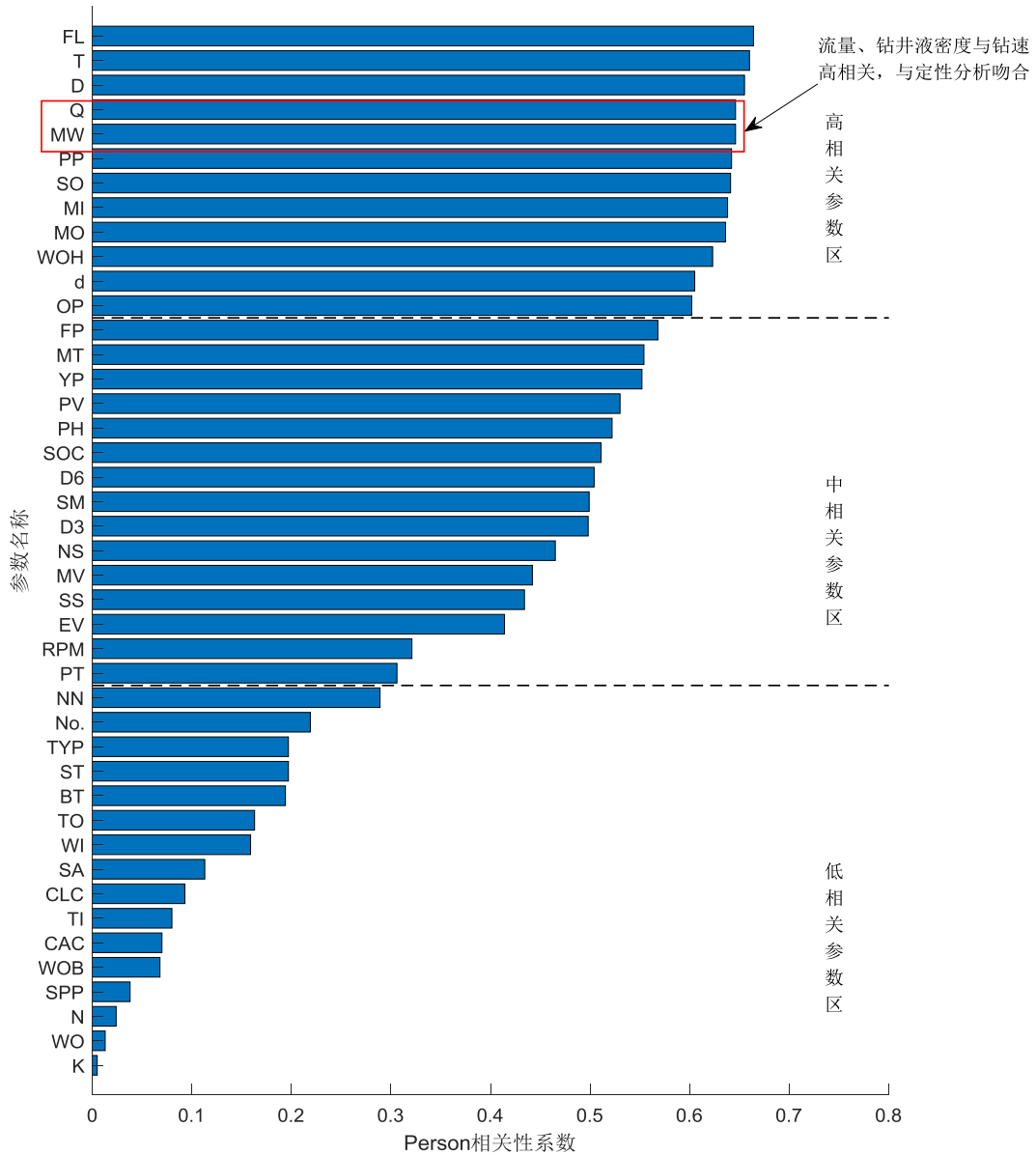


图9 各参数与钻速的相关性分析结果

### 3.3 因子分析

在未明确影响钻速准确因素的前提下,收集生产资料将面对需要分析变量参数过多的问题,如本例所示高达43种相关变量。过多变量可通过因子分析算法进行分组和降维,从中寻找内部相关性较高的变量组合。因子分析的核心原理在于通过准确计算变量间的相关性,将变量间按照两两之间的相关性进行分组,使同组内的变量相关性较高,而不同组的变量之间相关性较低。同时,计算分析通过一

个不可观测的基本结构来描述和定义划分好的每一个变量组合,这个基本结构称之为公共因子,如式6所示<sup>[16]</sup>:

$$\begin{cases} X_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \epsilon_1 \\ X_2 = l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \epsilon_2 \\ \vdots \\ X_n = l_{n1}F_1 + l_{n2}F_2 + \dots + l_{nm}F_m + \epsilon_n \end{cases} \quad (6)$$

式中： $X_1, X_2, \dots, X_n$ ——原始可观测的变量； $F_1, F_2, \dots, F_m$ ——公共因子，不可观测变量；



$\epsilon_1, \epsilon_2, \dots, \epsilon_n$ ——特殊因子,用于衡量测量误差等因素; $(l_{ij})_{n \times m}$ ——因子载荷矩阵,用以表示第*i*个变量对第*j*个因子的影响系数。

基于公共因子的定义,经过如因子旋转等数学变换就可以得到对应的,相对独立的综合指标,如式7所示<sup>[16]</sup>:

$$\begin{cases} f_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n \\ f_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2n}X_n \\ \vdots \\ f_m = a_{m1}X_1 + a_{m2}X_2 + \dots + a_{mn}X_n \end{cases} \quad (7)$$

式中: $f_1, f_2, \dots, f_m$ ——转换后综合指标; $(a_{ij})_{n \times m}$ ——因子得分矩阵,用以表示第*i*个变量在第*j*个指标中所占比例。

因子分析理论的核心在于使用有限个公共因子

(低维)取代原有多维参数,因此在选择公共因子数量时不可避免地可能产生原始数据信息丢失的问题。通过公共因子贡献率(包含原有数据信息的信息量)可直观展示这种效应。依次计算从原始43种参数中提取不同公共因子数量后的贡献率如图10所示。由图10左,43个原始参数最多可以提取等量(43个)的公共因子,每个公共因子的贡献率随因子编号的增加逐级递减。第1个公共因子包含最多的原始数据信息,贡献率达34.7%。当因子编号超过20时,每个因子包含的原始数据信息不足1.5%。由图10右可知,前20个公共因子的贡献率累积已经超过原始数据信息量的90%,故在后续的建模分析过程中可考虑仅使用前20个公共因子。尽管没有包含后23个公共因子,但丢失的原始信息不足10%,这种误差在工程应用领域是可接受的。

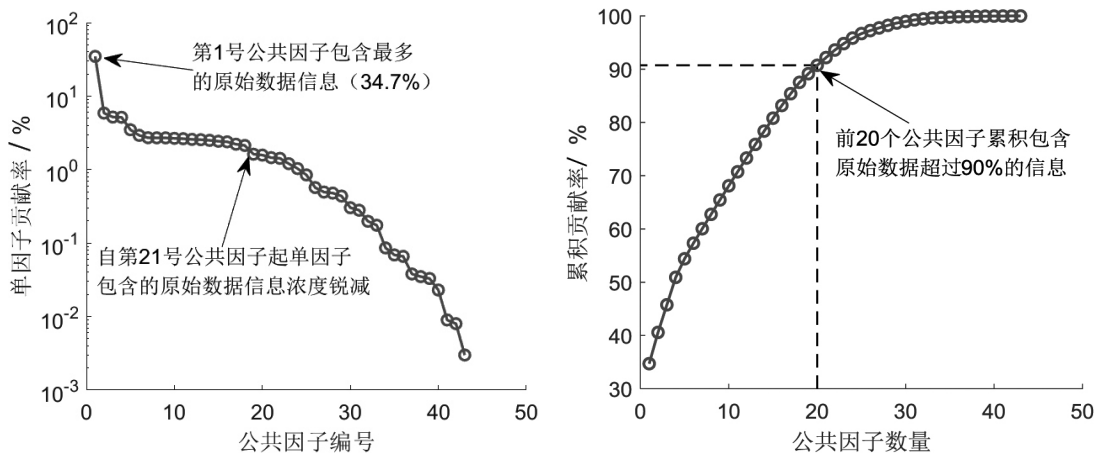


图10 因子分析计算结果

基于计算得到的载荷矩阵 $(l_{ij})_{43 \times 20}$ ,载荷矩阵中每个公共因子对列中超过0.8的参数值可确定为主要影响参数,据此可得如图11所示的各公共因子与影响程度最大的原始参数对应图谱。基于该图谱,可明确获知各公共参数的主要影响参数,从而得到简化的公共因子物理含义。以前4个单因子贡献率超过5%,合计贡献率超过50%的公共因子为例进行解释如下:

公共因子1:主要影响参数包括井径D、井深d、泵量Q、钻井液密度MI/MO/MW、钻井液pH值、固相含量SO、地层孔隙压力PP、地层破裂压力FP与地层上覆压力OP。由此可见,除井径和井深两个显而易见的影响因素外,最大的影响因素仍然是以

密度为主的钻井液性能与地层压力,该结论也与定性分析和相关性分析的结论吻合;

公共因子2:主要影响参数为钻井液进出口温度TI与TO,说明该地层温度也具备一定的影响;

公共因子3:主要影响参数为钻进时间BT与泵送时间PT,泵时在一定程度上会影响钻井液当量密度,进一步说明钻井液密度对钻进的影响;

公共因子4:主要影响因素为井眼编号No.(指代井眼位置与对应地层条件)和钻井液钙离子含量CAC,由于此处钙离子含量为钻井液出口数据,在一定程度上也包含了地层信息,故公共因子4可认定为与地层性质相关。

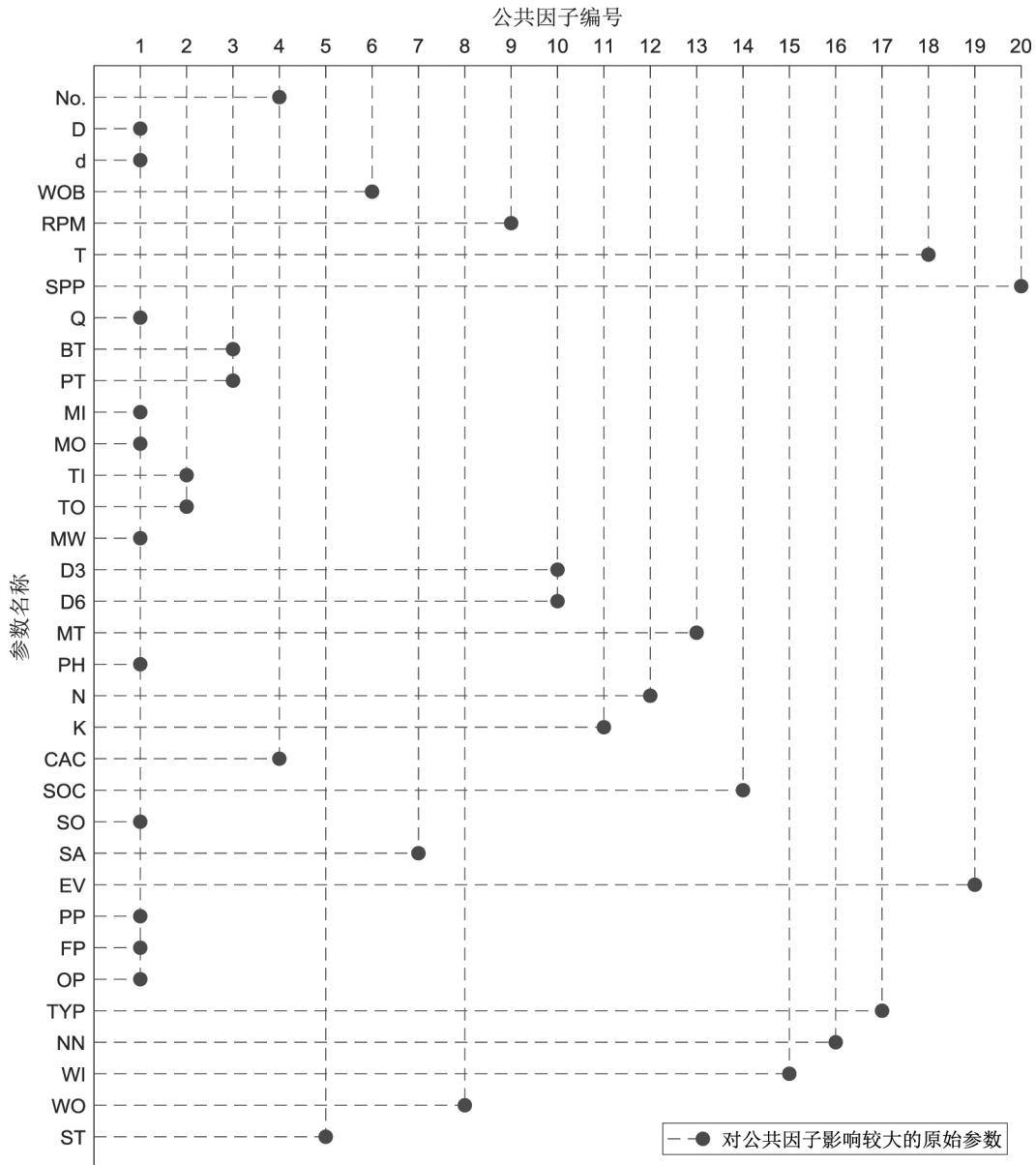


图 11 公共因子与原始参数的对应图谱

#### 4 结论

对实际生产数据的整合与初步分析是提高生产效率的有效途径,同时也是进行后续大数据建模、人工智能甚至自动施工的必要基础。本文以南海某区块相关钻井数据为例,从数据采集、数据定性分析和定量分析3个完整流程对生产数据的初步分析进行了举例与展示,主要结论如下:

(1)面对不同的数据格式,可通过图像点采算法实现图像数据向矩阵化数据的转换;同时针对不同的数据密度,也可通过数据插值补齐算法进行统一。

(2)数据标准化算法能够消除不同数据参数间

的数量级与单位差异,避免后续建模分析的误差;而数据可视化则可辅助定性地快速寻找数据规律与趋势。

(3)就定量分析而言,直接的统计分析可快速了解数据本质,获得特征性参数;相关性分析算法则可量化不同数据参数间的相关关系,为后期建模选择参数奠定基础;同时因子分析算法可对大量的数据参数进行分组归类,寻找不同数据参数间的公共因子并实现高维数据的降维,降低后期建模分析的难度。

## 参考文献:

- [1] 耿黎东. 大数据技术在石油工程中的应用现状与发展建议[J]. 石油钻探技术, 2021, 49(2): 72-78.
- [2] Zborowski M. How ConocoPhillips solved its big data problem [J]. Journal of Petroleum Technology, 2018, 70(7): 21-22.
- [3] Al-Subaiei D, Al-Hamer M, Al-Zaidan A, et al. Smart production surveillance: production monitoring and optimization using integrated digital oil field [C]. SPE Kuwait Oil and Gas Show and Conference, 2019.10.
- [4] New AI technology, Sandynicknamed, to accelerate projects [EB/OL]. (2019-01-28) <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/news-and-insights/press-releases/bp-invests-in-new-artificial-intelligence-technology.pdf>.
- [5] Shell expands strategic collaboration with Microsoft to drive industry transformation and innovation [EB/OL]. (2018-09-20) <https://news.microsoft.com/2018/09/20/shell-expands-strategic-collaboration-with-microsoft-to-drive-industry-transformation-and-innovation/>.
- [6] ExxonMobil to increase Permian profitability through digital partnership with Microsoft [EB/OL]. (2019-02-22) [https://corporate.exxonmobil.com/news/newsroom/news-releases/2019/0222\\_exxonmobil-to-increase-permian-profitability-through-digital-partnership-with-microsoft](https://corporate.exxonmobil.com/news/newsroom/news-releases/2019/0222_exxonmobil-to-increase-permian-profitability-through-digital-partnership-with-microsoft).
- [7] 匡立春, 刘合, 任义丽, 等. 人工智能在石油勘探开发领域的应用现状与发展趋势[J]. 石油勘探与开发, 2021, 48(1): 1-11.
- [8] DELFI cognitive E&P environment [EB/OL]. (2021-05-18) <https://www.software.slb.com/delfi>.
- [9] AI by BakerHughesC3. ai [EB/OL]. (2021-05-18) <https://www.bakerhughes.com/ai-bakerhughesc3ai>.
- [10] Manage information and help turn data into action with DecisionSpace® 365 [EB/OL]. (2021-05-18) <https://www.halliburton.com/en/software/decisionspace-365-information-management>.
- [11] 中石油发布勘探开发梦想云平台 [EB/OL]. (2018-11-27) [http://www.xinhuanet.com/2018-11/27/c\\_1123775741.htm](http://www.xinhuanet.com/2018-11/27/c_1123775741.htm).
- [12] 张志伟. 国内外岩芯数字化信息发布平台建设进展 [J]. 地质论评, 2020, 66(2): 493-498.
- [13] JEON G. Lagrange interpolation for up sampling [J]. International Journal of Multimedia and Ubiquitous Engineering, 2015, 10: 339-350.
- [14] LI Q, LI J P, DUAN L C, et al. Prediction of rock abrasivity and hardness from mineral composition [J]. International Journal of Rock Mechanics and Mining Sciences, 2021, 140: 104658.
- [15] 向东进, 李宏伟, 刘小雅. 实用多元统计分析 [M]. 武汉: 中国地质大学出版社, 2006.
- [16] WENDLER T, GRÖTTRUP S. Factor analysis [M] // Data Mining with SPSS Modeler. Springer International Publishing, 2021: 547-622.