

有机地球化学参数总有机碳含量预测

赵洪帅¹, 周文², 张天择², 林亚尼²

¹成都理工大学, 四川 成都

²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

收稿日期: 2021年11月16日; 录用日期: 2021年12月15日; 发布日期: 2021年12月30日

摘要

页岩气储层中总有机碳含量(TOC)反映了页岩的生烃潜力, 连续准确获取页岩气储层TOC含量对页岩气的勘探开发具有重要意义。本文结合威远地区龙马溪组页岩的实际情况, 建立了多元回归曲线法、 $\Delta\log R$ 法、支持向量机回归算法三种预测模型, 并对该地区W1井页岩进行了TOC含量评定。结果表明: 在威远地区采用支持向量机回归算法模型其精度最高, 可以为页岩的TOC含量预测提供指导。

关键词

总有机碳含量, 威远地区, 支持向量机回归

Prediction of Total Organic Carbon Content in Organic Geochemical Parameters

Hongshuai Zhao¹, Wen Zhou², Tianze Zhang², Ya'ni Lin²

¹Chengdu University of Technology, Chengdu Sichuan

²State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Chengdu University of Technology, Chengdu Sichuan

Received: Nov. 16th, 2021; accepted: Dec. 15th, 2021; published: Dec. 30th, 2021

Abstract

The total organic carbon content (TOC) in shale gas reservoir reflects the hydrocarbon generation potential of shale. Continuous and accurate acquisition of TOC content in shale gas reservoir is of great significance for shale gas exploration and development. Combined with the actual situation of Longmaxi formation shale in Weiyuan area, this paper establishes multiple regression curve method, $\Delta\log R$ method and support vector machine regression algorithm that are used to eva-

luate the TOC content of shale in well W1 in this area. The results show that the precision of support vector machine regression model is the highest in Weiyuan area, which can provide guidance for the prediction of TOC content of shale.

Keywords

Total Organic Carbon Content, Weiyuan Area, Support Vector Machine Regression

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

页岩气是指以吸附态、游离态赋存于富有机质泥页岩及其夹层中的非常规天然气, 伴随近年来页岩气勘探开发过程中的多个重大发现, 页岩气已经在中国天然气产量增长中发挥举足轻重的作用[1] [2] [3]。总有机碳(TOC)含量是评价烃源岩品质和生烃潜力的重要参数之一, 通过间隔取样并进行实验测试分析所获得的 TOC 含量数据较为精准, 但是得到的实测数据为离散数据, 无法反映全孔或者某段目的层泥页岩的总体特征[4]。全取心方案可以获得连续的样品, 然而由于页岩质纯性脆的特点, 取心的难度和成本会提高。富有机质泥页岩储层具有典型的测井曲线特征[5], 因此如何综合连续的测井参数对 TOC 含量进行定量解释成为当下重要的研究方向。

我国页岩气勘探开发起步较晚。在通过测井曲线线性预测总有机碳含量这方面, Passey 等人(1990) [6] 利用电阻率测井曲线和声波时差测井曲线建立了 $\Delta\log R$ 方法计算 TOC。杜文凤等(2016) [7] 发现在煤系烃源岩中总有机碳含量测井响应特征与自然伽马、声波时差、密度、电阻率和中子孔隙度相关性, 建立 TOC 含量预测模型。王祥等(2020) [8] 在此基础上提出一种考虑密度因素的广义 $\Delta\log R$ 法预测总有机碳含量方法。

由于页岩气储层总有机碳含量受多种地质因素及岩性纯度和特殊矿物组分的影响, 它们之间的关系可能不是单纯的测井参数与 TOC 的线性关系。季少聪等(2018) [9] 结合三维地震数据建立测井预测 TOC 含量与井旁道地震属性之间的神经网络模型, 计算 TOC 数据体。殷梅等(2020) [4] 基于贝叶斯判别建立了总有机碳含量的自动判别分类模型。赵峦啸等(2021) [10] 在地质导向约束下通过随机森林算法生成学习网络来预测总有机碳含量。

基于上述研究现状, 笔者同时选取了常用的三种预测方法, 包括多元回归法、 $\Delta\log R$ 方法、支持向量机回归(SVR)。在工区的目的层位进行了应用, 并对这三种方法的预测结果进行评比分析。

2. 研究区地质概况

威远页岩气田位于四川盆地西南部, 面积约 $1.2 \times 10^4 \text{ km}^2$ 。区内地表为侏罗~白垩系红色地层所覆盖, 在盆地周缘广泛出露下古生界及其以下地层, 海拔在 1000 m~1200 m 之间, 西部为平原, 中部为丘陵, 海拔在 300 m~600 m 之间。威远地区下志留统龙马溪组地层经历多期构造运动见图 1, 大断层虽然不发育, 但小、微断层数量多, 利于裂缝发育带的形成[11]。

龙马溪组页岩埋深 1500 m~4000 m。威远地区发育一东西向的(水下)古隆起, 影响该区域优质页岩的厚度展布及沉积环境, 进而影响到页岩储层品质和水平井产量。威远地区五峰组页岩以深色页岩为主, TOC 普遍含量较低(TOC < 2%)。

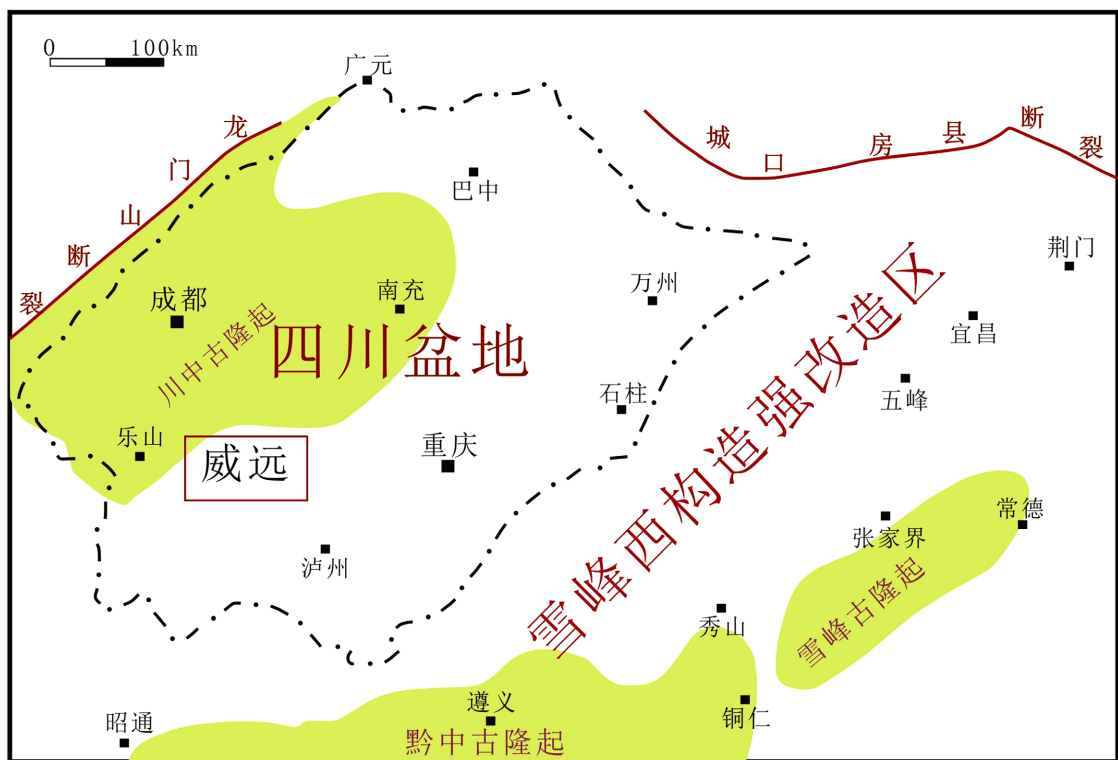


Figure 1. Schematic diagram of paleostructure of Longmaxi formation in the Sichuan Basin and surrounding areas
图 1. 四川盆地及周缘龙马溪组古构造示意图

3. 预测方法及原理

3.1. 多元线性回归法

回归分析是通过分析判断来确定相关变量之间内在关系的一种统计方法。如果存在 2 个或多个不相关自变量与因变量存在较好相关性的情况下，多元回归分析的结果通常会优于单个因素分析结果[12]。

该方法的基本原理是通过测井曲线与实测数据的相关性分析优选与 TOC 相关度高的测井参数，将这些相关性高的独立的测井参数进行多元拟合，通过建立多元回归模型来预测烃源岩中的 TOC。

本文用到的常规测井曲线主要包括声波时差(AC)、井径(CAL)、中子(CNL)、(DEN)、(GR)、(K)、(KTH)、(PE)、(RT)、(TH)、(U)共 11 条测井曲线，实验室对 W1 井龙马溪组 279 个深度点的岩样做了总有机碳含量测试分析。将实测 TOC 与对应深度下各种测井曲线做相关性分析，得到其相关系数 R^2 见图 2。

	AC	CAL	CNL	DEN	GR	K	KTH	PE	RT	TH	U	toc
AC	1											
CAL	-0.17296	1										
CNL	0.093364	0.410709	1									
DEN	-0.16899	-0.42521	0.284457	1								
GR	0.342417	-0.2315	-0.41162	-0.38692	1							
K	0.065698	0.08786	0.747931	0.503091	-0.45304	1						
KTH	0.129502	0.127461	0.747805	0.440291	-0.36419	0.894722	1					
PE	-0.14829	0.678743	0.195381	-0.48052	-0.11193	-0.01476	0.009415	1				
RT	-0.18195	-0.09262	-0.59853	-0.37715	0.343693	-0.61352	-0.63838	-0.00833	1			
TH	0.084	0.142332	0.546296	0.279323	-0.1723	0.504329	0.804662	0.039742	-0.47803	1		
U	0.138315	-0.22316	-0.67393	-0.49213	0.862416	-0.76957	-0.77476	-0.06229	0.585491	-0.54343	1	
toc	0.126427	-0.09186	-0.62181	-0.4232	0.542879	-0.61252	-0.52617	-0.0143	0.452739	-0.31343	0.64082	1

Figure 2. Correlation coefficient R^2 between TOC and each logging curve
图 2. TOC 与各测井曲线相关系数 R^2

相关性是-1 和 1 之间的统计度量,表示两组点之间线性关系的强度。相关性为 1 是一条完美的直线,而相关性为 0 则表示没有任何线性关系。从表 1 中可以看出 TOC 含量与铀相关性最强且呈正相关,相关系数为 0.64,铀含量越高,TOC 含量越高;中子孔隙度,钾,伽马,无铀伽马表现出与 TOC 含量较好的相关性,依次为-0.62, 0.61, 0.54, -0.53,将相关性大于 0.6 的三条测井曲线,与 279 个实测样本 TOC 值进行多元回归,得到计算 TOC 的三元回归表达式

$$\text{TOC} = -0.08528 \cdot \text{CNL} - 0.12486 \cdot \text{K} + 0.05364 \cdot \text{U} + 0.05364 \quad (\text{R} = 0.69, \text{R}^2 = 0.48) \quad (1)$$

将相关性大于 0.5 的三条测井曲线,与 279 个实测样本 TOC 值进行多元回归,得到计算 TOC 五元回归表达式

$$\begin{aligned} \text{TOC} = & -0.10316 \cdot \text{CNL} + 0.006833 \cdot \text{GR} - 0.381588 \cdot \text{K} + 0.009651 \cdot \text{KTH} \\ & + 0.014417 \cdot \text{U} + 2.051844 \quad (\text{R} = 0.72, \text{R}^2 = 0.52) \end{aligned} \quad (2)$$

两式比较认为,式(2)较式(1)多了 GR, KTH 两条与实测 TOC 相关性介于 0.5~0.6 的测井曲线,然而拟合效果 R 仅从 0.69 上升到 0.72,说明这两条曲线对最终拟合值的影响较 CNL, K, U 三条曲线相对小很多,故采用 CNL, K, U 三元拟合回归。

3.2. $\Delta \log R$

$\Delta \log R$ 方法由埃克森(Exxon)和埃索(Esso)公司研发,由 Passey 等提出的利用声波时差曲线和电阻率曲线重叠来定量计算 TOC 的一种方法,适用于碎屑岩和碳酸盐岩中烃源岩的定性识别和定量计算[6]。

$\Delta \log R$ 法基本原理:将孔隙度测井曲线(一般是声波时差曲线)与电阻率曲线叠加。利用声波时差和电阻率对于不同岩性层段的差异响应,将两条曲线的非烃源岩段通过调节坐标值来使其重叠并定义此非烃源岩段为基线段。该方法假定岩性基本不变,认为孔隙度和电阻率测井曲线仅仅是由有机质含量改变而引起的。 $\Delta \log R$ 即两条曲线的幅度差与 TOC 呈线性关系并且是成熟度(LOM)的函数。

依据 Passey 等[6]提出的声波,电阻率叠加计算 $\Delta \log R$ 的数学表达式为

$$\Delta \log R = \log_{10} \left(\frac{R}{R_{\text{baseline}}} \right) + 0.02 \times (\Delta t - \Delta t_{\text{baseline}}) \quad (3)$$

式中, $\Delta \log R$ 是以对数电阻率循环测量的曲线间距, R 是用测井工具测量的电阻率,单位是 $\Omega \cdot \text{m}$, Δt 是测量传输的声波时差,单位是微秒每英尺 $\mu\text{s}/\text{m}$, R_{baseline} 是电阻率曲线位于非烃源岩段基线处的电阻率值,与 $\Delta t_{\text{baseline}}$ 的值相一致。

依据 Passey 等[6]提出的 $\Delta \log R$ 从测井曲线计算富含粘土的岩石中总有机碳的经验公式为

$$\text{TOC} = \Delta \log \times 10^{(2.297 - 0.168 \times \text{LOM})} \quad (4)$$

式中总有机碳含量以重量百分比计算,单位是%, LOM 是成熟度,LOM 为 7 对应倾油干酪根成熟的开始,LOM 为 12 对应倾油干酪根过成熟的开始。

根据威 202 的实际测井数据,基于该方法可以得到预测总有机碳含量的模型,根据 $\Delta \log R$ 法预测得到的总有机碳含量见图 3。图中红色点为实测的总有机碳含量值,蓝色曲线为 $\Delta \log R$ 法预测模型预测的总有机碳含量值,可以看出该法可以一定程度上刻画泥岩中总有机碳含量的纵向变化趋势,但是无法识别异常的局部极值点。

通过把声波时差曲线(AC)与电阻率(RT)曲线进行重叠,电阻率曲线为对数刻度,刻度值为 0.035~350 ($\Omega \cdot \text{m}$),声波曲线的刻度值为 200~0 $\mu\text{s}/\text{m}$,最终确定 RT 基线值是 8.419 $\Omega \cdot \text{m}$, AC 的基线值是 79.647 $\mu\text{s}/\text{m}$ 。

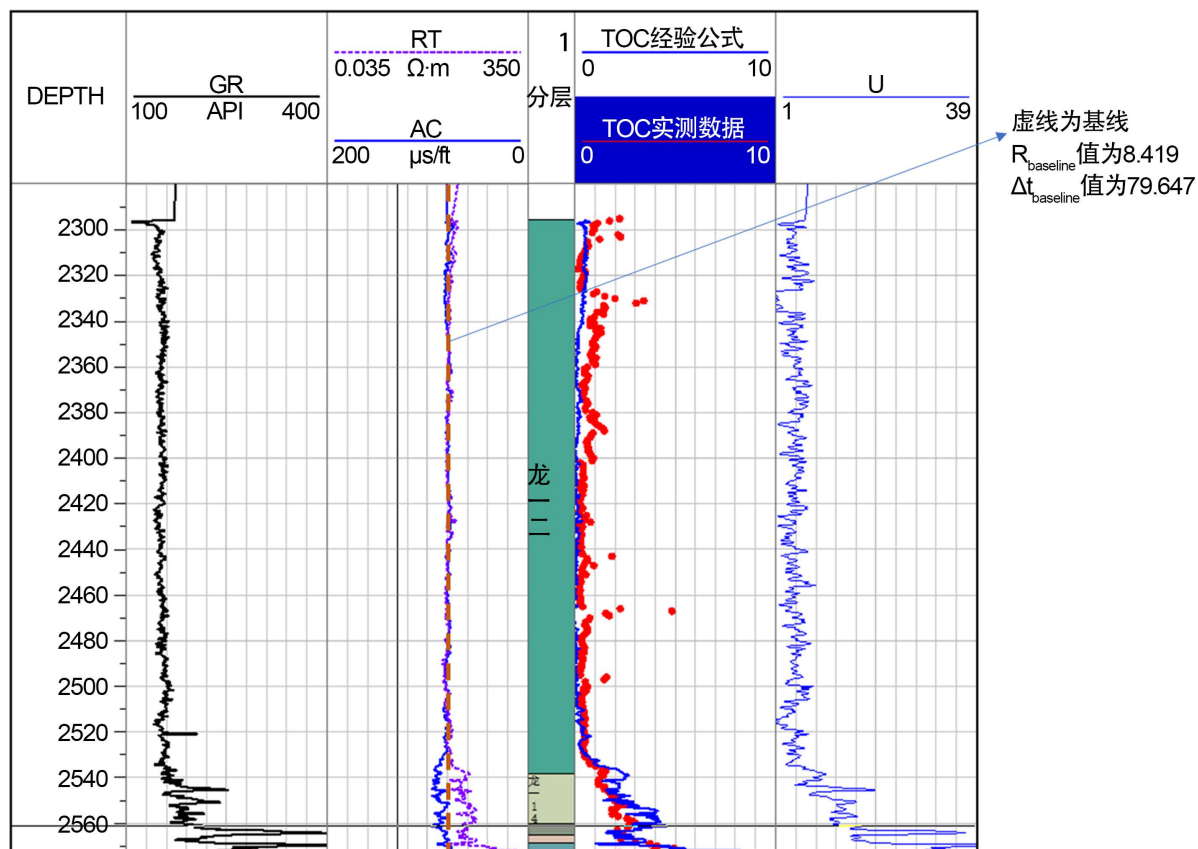


Figure 3. $\Delta \log R$ method superimposes acoustic time difference curve and resistivity curve
 图 3. $\Delta \log R$ 法叠合声波时差曲线和电阻率曲线图

3.3. 支持向量机回归法

支持向量机(Support Vector Machine, SVM)是一种机器学习算法,最早由 Vapnik 于 1995 年提出。SVM 的核心思想是通过使用核, 隐含地将数据投射到高维特征空间, 将线性算法转化为非线性算法。SVM 遵循了完备的统计学习理论基础和结构风险最小化原理(SRM), 是一种基于和的学习方法或大间隔分类器。SVM 基于严格的理论基础, 结构简单, 直观易于理解和精确实现, 训练过程等于求解一个凸二次规划问题, 仅需调节少数几个可调参数来获得全局最优解, 高效的优化算法会使 SVM 训练速度非常快。

SVR 超参数包括如下。Kernel 为指定要在算法中使用的内核类型, degree 为多项式核函数的次数, gamma 为“rbf”的核系数, coef0 为核函数中的独立项, C 为错误术语的惩罚参数。本次对该模型超参数的调节使用的是“Grid Search”, 针对每种可能参数的组合情况, 进行模型训练, 然后模型评价, 根据评价结果, 寻找取最大值时的参数取值情况。

支持向量机回归就是使得最优分类平面两侧的样本到平面的总距离最小, 从而实现输入参数的拟合。支持向量回归就是建立一个最优分类的超平面, 使得样本分类的间隔距离能够最大化, 进而等价于核函数的内风险函数在式(5)最小化问题。

$$\min_{f, \xi, \xi^*} C \sum_{i=1}^l (\xi_i + \xi_i^*) + \frac{1}{2} \|w\|^2 \quad (5)$$

式中, C 为惩罚参数, ξ 为松弛变量, ξ^* 是某 i 个松弛因子, w 为空间特征中的权向量。

本文在建立支持向量机回归模型时, 输入威远区块所有井位测井数据, 筛选了特征表征性强的测井曲线 AC, CAL, CNL, DEN, GR, KTH, PE, TH, U, 剔除了测井数据异常值, 对所有井直接测井曲线进行了标准化处理, 测井模型的核函数为高斯核函数, 建立模型时, 随机 70% 数据用以模型建立, 30% 数据用以模型验证, 通过调节超参数 C 、 γ , 使训练分数(Train score)达到 0.92, 同时测试数据(Test score)达 0.80, 将此模型应用到 W1 井得到 TOC 预测数据见图 4。

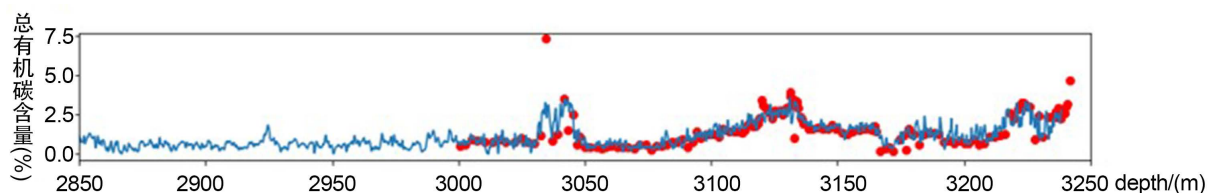


Figure 4. Effect diagram of TOC prediction by support vector regression

图 4. 支持向量回归法预测 TOC 效果图

4. 应用实例及结果分析

4.1. 误差分析

此次研究分析了三种误差, 包括平均绝对误差(MAE), 均方根误差(RMSE)和平均相对误差(MRE)。根据误差计算公式, 3 类预测结果的误差见表 1:

Table 1. Comparison of prediction errors of three models

表 1. 3 种模型预测误差比较

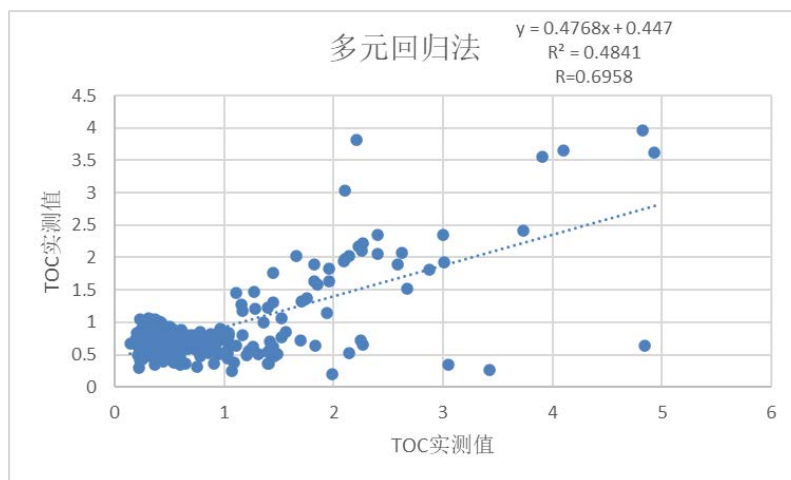
	多元回归法	$\Delta\log R$ 法	支持向量回归法
平均绝对误差(MAE)	0.4065	0.5935	0.2603
均方根误差(RMSE)	0.35	0.58	0.21
平均相对误差(MRE)	0.6275	0.5407	0.1836

相比较与多元回归法和 $\Delta\log R$ 法, 支持向量机回归法的预测结果能更好的拟合实测值, 预测误差小, 平均绝对误差 0.2603, 均方根误差为 0.21, 平均相对误差为 0.1836, 说明其测量值和真实值整体偏离最小, 预测效果更好。

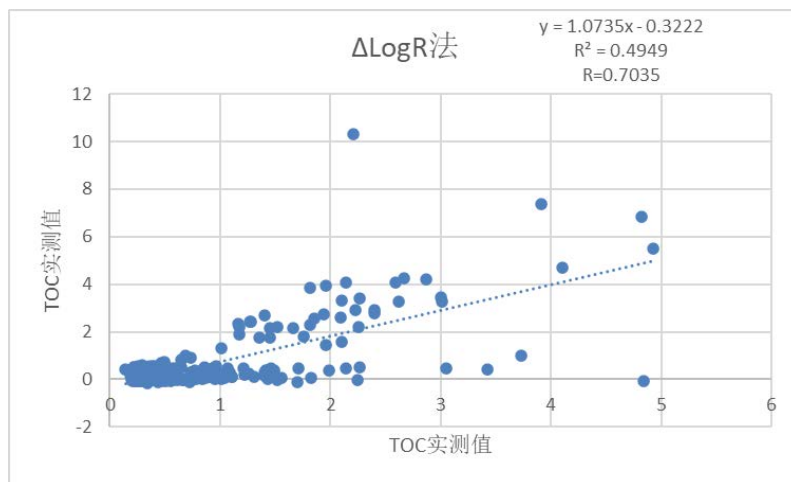
3.2. 结果分析

根据 W1 井实测资料结合测井曲线, 利用多元回归法, $\Delta\log R$ 法, 支持向量机回归法计算 TOC 并得到实测 TOC 与计算 TOC 的交会图见图 5, 由图可以看出支持向量回归的计算 TOC 与实测 TOC 相关性最好(R^2 为 0.690, R 为 0.8342)。

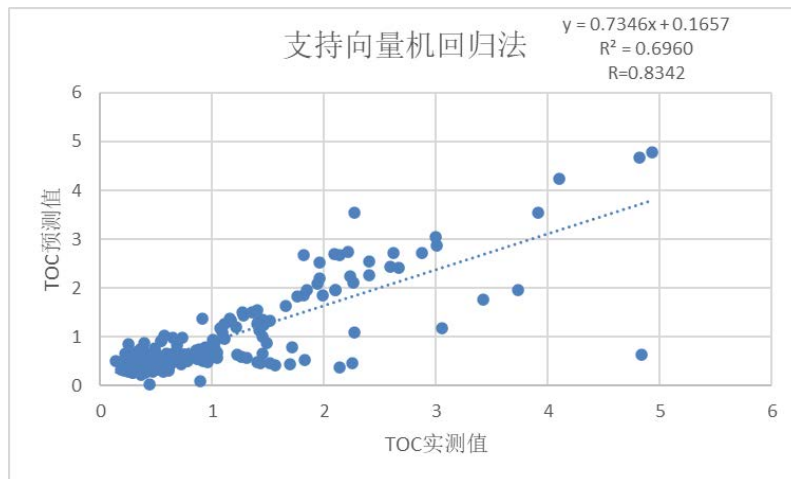
利用测井曲线, 使用不同方法对总有机碳含量进行了纵向预测对比可以看出支持向量机回归法与实测数据有更好的匹配关系见图 6, 说明支持向量机法可以更为精准的刻画总有机碳含量在空间上纵向变化规律。常规测井中所反映的总有机碳含量信息可能被地层中其他信息所干扰覆盖, 使得相关性变差, 然而借助支持向量回归的数据挖掘能力对总有机碳含量进行预测, 可以处理一些线性经验公式对局部极值点估算不足的缺陷。



多元回归法



ΔlogR 法



支持向量机回归法

Figure 5. Comparison diagram of correlation R^2 between TOC predicted value and measured value of three methods

图 5. 三种方法 TOC 预测值与实测值相关性 R^2 对比图

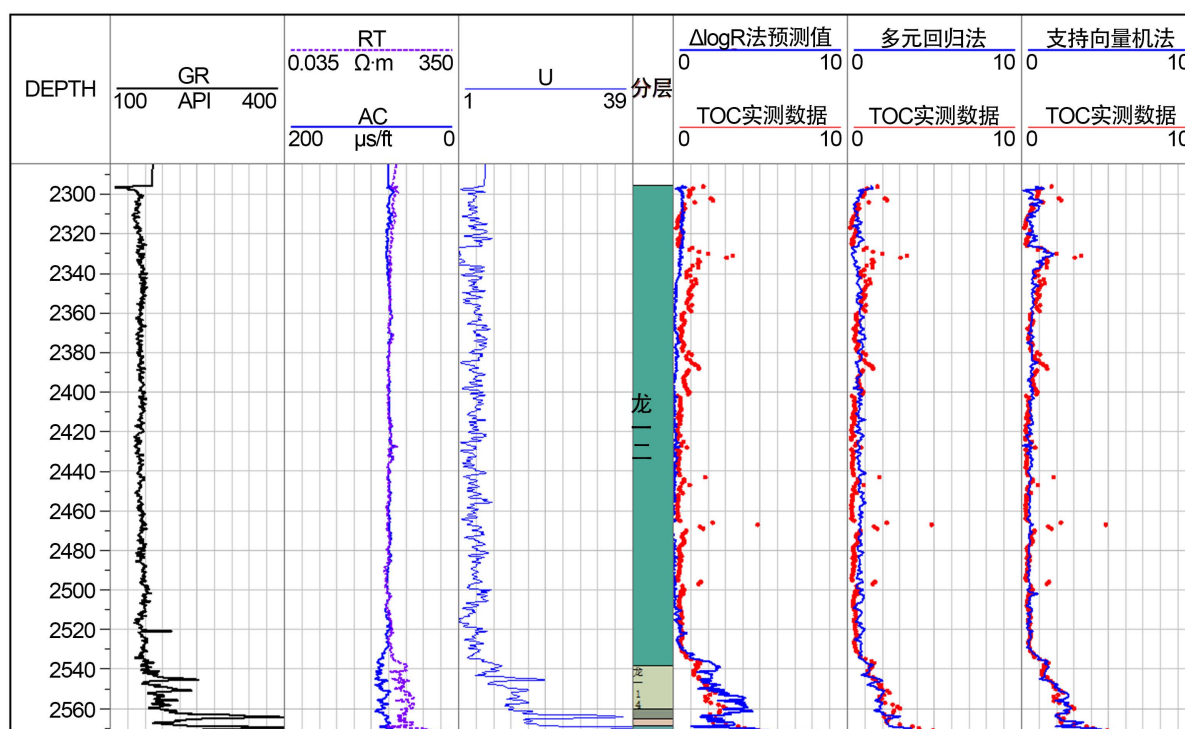


Figure 6. Comparison of prediction results of total organic carbon content of logging data by different methods

图 6. 利用不同方法进行测井数据总有机碳含量预测的结果对比

5. 结论

1) 本文以威远地区测井数据和岩石地化测试数据为基础, 使用多元回归法, $\Delta\log R$ 法和支持向量机回归法预测了威远地区 W1 井龙马溪组的总有机碳含量(TOC);

2) 根据三种方法做了做了预测 TOC 和实测 TOC 的误差分析来评价模型性能, 得到以下结论: 支持向量机回归法是这些方法中预测效果最好的方法, 平均相对误差在 20% 以内, 能够提高预测 TOC 的精度;

3) 在进行方法推广时, 应当对研究区的地质背景作足够的调查, 结合当地地区经验进行参数的选取和调整, 采取多种方法进行横向对比, 优选合适的精度较高的方法。

参考文献

- [1] 邵艳, 李卓文. 四川盆地威远地区龙马溪组页岩储层特征[J]. 地质学刊, 2016, 40(4): 624-630.
- [2] 陈孝红, 张保民, 张国涛, 等. 湖北宜昌地区奥陶系五峰组——志留系龙马溪组页岩气高产工业气流[J]. 中国地质, 2018, 45(1): 199-200.
- [3] 包书景, 李世臻, 徐兴友, 等. 全国油气资源战略选区调查工程进展与成果[J]. 中国地质调查, 2019, 6(2): 1-17.
- [4] 殷梅, 岑超, 阮娟, 等. 基于贝叶斯判别的页岩气储层总有机碳含量分类预测方法[J]. 地质学刊, 2020, 44(4): 362-369.
- [5] 熊镭, 张超谟, 张冲, 等. A 地区页岩气储层总有机碳含量测井评价方法研究[J]. 岩性油气藏, 2014, 26(3): 74-78.
- [6] Passey, Q.R., Creaney, S., Kulla, J.B., et al. (1990) A Practical Model for Organic Richness from Porosity and Resistivity Log. *AAPG Bulletin*, 74, 1777-1794. <https://doi.org/10.1306/0C9B25C9-1710-11D7-8645000102C1865D>
- [7] 杜文凤, 王攀, 梁明星, 等. 煤系烃源岩有机碳含量测井响应特征与定量预测模型[J]. 煤炭学报, 2016, 41(4): 954-963.
- [8] 王祥, 马劲风, 张新涛, 等. 一种考虑密度因素的广义 $\Delta\log R$ 法预测总有机碳含量——以渤中凹陷西南部陆相深

- 层烃源岩为例[J]. 地球物理学进展, 2020, 35(4): 1471-1480.
- [9] 季少聪, 杨香华, 朱红涛, 等. 下刚果盆地 A 区块 Madingo 组烃源岩 TOC 含量的地球物理定量预测[J]. 石油地球物理勘探, 2018, 53(2): 369-380.
- [10] 赵峦喇, 刘金水, 姚云霞, 等. 基于随机森林算法的陆相沉积烃源岩定量地震刻画: 以东海盆地长江拗陷为例[J]. 地球物理学报, 2021, 64(2): 700-715.
- [11] 王治平, 张庆, 刘子平, 等. 斜坡型强非均质页岩气藏高效开发技术——以川南威远地区龙马溪组页岩气藏为例[J]. 天然气工业, 2021, 41(4): 72-81.
- [12] 王濡岳, 丁文龙, 王哲, 等. 页岩气储层地球物理测井评价研究现状[J]. 地球物理学进展, 2015, 30(1): 228-241.