

基于灰色模型与时间序列的油井产量长短期分类预测

靳子豫¹, 李政林¹, 臧永亮², 黄伟婷¹, 范晴晴¹

1. 广西科技大学自动化学院 广西柳州 545616
2. 中国石油华北油田公司二连分公司 内蒙古锡林浩特 026000

通讯作者: Email: 59545980@qq.com

项目支持: 国家重点研发计划项目“北方砂岩型铀能源矿产基地深部探测技术示范项目”(2018YFC0604200); 教育部产学研协同育人项目“电子测量”(201702071082); 广西科技大学产学研合作项目“油气田地面计量数字采集系统研制及现场应用”(0721011901)

引用: 靳子豫, 李政林, 臧永亮, 等. 基于灰色模型与时间序列的油井产量长短期分类预测[J]. 油气井测试, 2022, 31(5): 6-11.

Cite: JIN Ziyu, LI Zhenglin, ZANG Yongliang, et al. Long-term and short-term classified prediction on well production based on grey model and time series[J]. Well Testing, 2022, 31(5): 6-11.

摘要 为解决单井产量在试采及稳产期预测计算方法精度不足、计算效率较低、适用范围较小的问题, 提出一种将预测模型划分为短期产量和长期产量的分类预测方法。短期产量预测在灰色预测模型 GM(1,1) 的基础上, 针对短期生产环境的变化增加产量转换的参数修正, 并对模型的离散过程进行优化, 从而弥补灰导数离散过程的偏差; 对于长期产量预测, 采用时间序列的方法, 结合 BP 神经网络以及短期采样特征值的提取结果, 对累积产量数据进行建模和预测。利用 2018~2020 年多个油井的产量数据信息, 进行模型训练、预测和比对, 针对模型预测结果进行精度分析。结果表明, 长期预测模型预测结果的平均相对误差为 0.091 4, 短期预测模型预测结果的平均绝对误差为 0.118 7, 计算精度高于传统 ARIMA 算法的平均相对误差 0.156 6, 可满足实际工况需求, 为油井产量的预测提供了一种新方法。

关键词 智能采油; 大数据; 短期预测; 特征提取; 长期预测; 预测模型; BP 神经网络

中图分类号: TE33 文献标识码: A DOI: 10.19680/j.cnki.1004-4388.2022.05.002

Long-term and short-term classified prediction on well production based on grey model and time series

JIN Ziyu¹, LI Zhenglin¹, ZANG Yongliang², HUANG Weiting¹, FAN Qingqing¹

1. School of Automation, Guangxi University of Science and Technology, Liuzhou, Guangxi 545616, China
2. The Erlian Filiale of PetroChina Huabei Oilfield Company, Xilinhote, Inner Mongolia 026000, China

Abstract: The calculation methods for predicting well production in pilot production and stable production stages are disadvantageous for insufficient precision, low efficiency, and small application scope. In order to solve these problems, a method that divides the prediction model into short-term production and long-term production was proposed. For short-term production prediction, based on the grey prediction model GM(1,1), the parameter correction on production conversion is added depending on the changes in short-term production environment, and the discrete process of the model is optimized, which makes up for the deviation of the gray derivative discrete process. For long-term production prediction, based on time series, and combining with BP neural network and the extraction results of short-term sampling eigenvalues, the cumulative production data were modeled and predicted. Using the production data of multiple wells in 2018~2020, model training, prediction and comparison were performed, and the accuracy of the results predicted by the model was analyzed. The results show that the average relative error of the prediction results of the long-term prediction model is 0.091 4, and the average absolute error of the prediction results of the short-term prediction model is 0.118 7. The calculation accuracy is higher than the average relative error of 0.156 6 calculated by traditional ARIMA algorithm, which can meet the needs of actual working conditions, providing a new method for predicting of well production.

Keywords: intelligent oil production; big data; short-term prediction; feature extraction; long-term prediction; prediction model; BP neural network

油气井产量预测是石油生产工程管控的重要环节, 是生产规划、制定方案、管理调控的重要依托, 该数据在石油现代化发展、石油工业物联网发展等方面都有重要的指导意义^[1]。

近些年来,随着神经网络等算法的不断发展,薛亮等^[2]利用 PSO 优化记忆神经网络算法对气井生产进行动态预测;李彦尊等^[3]运用神经网络预测页岩油气产量;周鹏等^[4]运用 weibul 模型、崔传智^[5]基于灰色模型的改进算法对油藏及产量进行了预测;寸少妮等^[6]根据产量递减规律、国亮^[7]利用多元线性回归方法进行预测。这些方法能较好地预测石油生产产量的变化趋势,但使用环境苛刻,适用度不高。

二十世纪二十年代英国统计学家 G u Yule 提出了自回归数学模型 (AR),英国统计学家 G T Walker 在此之后提出了移动平均模型 (MA),并将该模型与自回归模型进行融合分析,提出了 ARMA 模型,后来 BOX 和 JENKINS 在原有的模型基础上提出了 ARIMA 模型^[8-11]。现今的油井产量模型在考虑周期性、季节性等因素的基础上,通过差分等方式将非线性的数据进行处理,满足了在不同非线性因素影响下的数据预测^[12-14]。但时间序列预测模型针对于周期更长的数据进行预测,面对短时不确定因素带来的趋势变化无法准确的拟合测量^[15-16]。本文提出的动态灰色预测模型 GM(1,1) 可以提取出短时间内的产量信息,并利用这些信息及时更新模型,增加了短期预测的准确性。

1 原理与方法

利用时间序列分析出长期生产的趋势性、周期性,利用 BP 神经网络训练拟合时间序列模型预测结果的残差,优化本算法预测的精确度。

1.1 差分整合移动平均自回归模型 (ARIMA)

通过 AR 与 MA 模型的叠加配合 G T Walker 提出了自回归移动平均模型 ARIMA,其公式为

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (1)$$

式中: y_t 为当前时刻的结果; μ 为常量; γ_i 为自回归模型参数; y_{t-i} 为上一时刻的结果; ε_t 为当前时刻的干扰值; θ_i 为移动平均参数; $\varepsilon_{t-i} (i = 1, 2, \dots, k)$ 为过去时刻的干扰值; p 为自回归阶数; q 为移动平均阶数。

该模型需要满足以下条件:

$$\begin{cases} \gamma_i \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 \\ E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(y_s \varepsilon_t) = 0, \forall s < t \end{cases} \quad (2)$$

干扰值 ε_t 为期望为零,方差为 σ^2 的正态分布。基于上述自回归移动平均模型只能应用于平

稳序列的劣势,提出差分整合移动平均自回归模型 (ARIMA),也写作 ARIMA(p, d, q)。

若数组数据存在变化趋势,那么数据则为非平稳数据,在对非平稳数据进行预测分析之前利用差分的方法处理,其中一阶差分公式为

$$\Delta y_t = \{y_t - y_{t-1}\} \quad (t = 1, 2, \dots, k) \quad (3)$$

而后通过自相关系数 ACF 与偏自相关系数 PACF,判断该组数据的自回归阶数 p 与移动平均阶数 q 。

1.2 灰色预测动态模型

在 GM(1,1) 模型的基础上,在分析预测的过程中根据实时的生产信息,不断更新预测模型的参数,可以更加高效、准确的分析出数据的变化趋势。

设原始序列为 $y(0) = (y_1, y_2, y_3, \dots, y_n)$, 新测得的产量数据为 y_{n-1} 。

更新后的序列为

$$y(i) = (y_{i+1}, y_{i+2}, y_{i+3}, \dots, y_{i+n}) \quad (4)$$

更新后累加得到的序列为

$$y^1(i) = (y_{i+1}^1, y_{i+2}^1, y_{i+3}^1, \dots, y_{i+n}^1) \quad (5)$$

则其对应的生成的动态矩阵以及最小二乘法的公式为

$$\begin{cases} A(i) = \begin{bmatrix} -1/2(y_{i+1}^1 + y_{i+2}^1) & 1 \\ -1/2(y_{i+2}^1 + y_{i+3}^1) & 1 \\ \dots & \dots \\ -1/2(y_{i+n-1}^1 + y_{i+n}^1) & 1 \end{bmatrix} \\ B(i) = \begin{bmatrix} y_{i+2} \\ y_{i+3} \\ \dots \\ y_{i+n} \end{bmatrix} \end{cases} \quad (6)$$

$$\mu(i) = [A^T(i)A(i)]^{-1}A^T(i)B(i) = \begin{bmatrix} a_i \\ u_i \end{bmatrix} \quad (7)$$

其中计算的累加值为

$$y_{i+n+1}^1 = \left(y_{i+1} - \frac{p}{q}\right)e^{-a_i^n} + \frac{p}{q} \quad (8)$$

其中 p, q 为白化系数,通过上式解得对应参数后,即可得到新的预测值 y_{i+n+1}

$$y_{i+n+1} = y_{i+n+1}^1 - y_{i+n}^1 \quad (9)$$

本模型在运算过程中根据实时参数对计算模型实时更新,可更快速的跟随产量变化趋势,以保证预测的精确度能够满足实际井口工况。

2 建立产量预测分析模型

通过灰色预测模型对短期产量预测并提取特

征值与时间序列、神经网络相结合的方法,可以在季节、周期、趋势、信息传递等方面更好的约束预测的精度,同时简单的物理模型及其参数减少了算法复杂度。

通过对前面两种算法进行整体模型总结出

$$\begin{cases} y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \beta_t, \\ (t \geq 120 \text{ d}) \\ y_t = y_{t+1}^1 - y_t^1, \quad (3 \text{ d} \leq t \leq 10 \text{ d}) \end{cases} \quad (10)$$

通过对多口井的多组数据进行验证分析,最终定义为式(10)的形式,短期动态灰色预测模型适用于 3~10 d 的数据分析预测,对短期数据进行特征值提取带入 BP 神经网络拟合 ARIMA 模型预测结果的残差 β_t 与长期预测模型的预测结果相结合得出最终预测值 y_t ,长期动态灰色预测模型适用于两个月及以上的数据分析预测。

2.1 时间序列模型在长期油井产量中的应用

长期产量分析采用了时间序列模型与 BP 神经网络结合的方式。首先通过时间序列经过去周期、去趋势的处理后,用 ARIMA 模型拟合出产量变化,然后计算出的残差与短期产量特征值带入 BP 神经网络中进行训练,最后将 BP 神经网络的拟合后残差与时间序列模型预测结果相结合得出修正后的产量预测值。具体流程见图 1。

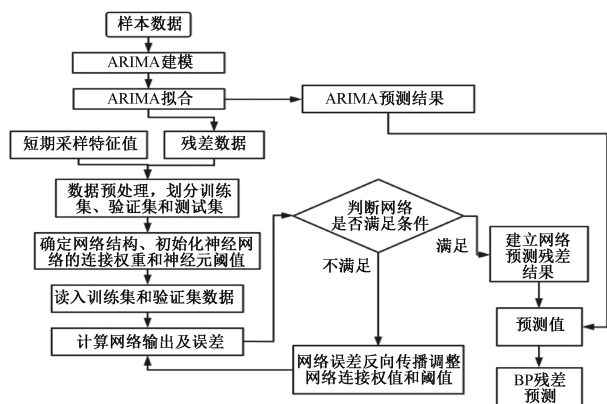


图 1 长期油井产量预测模型

Fig. 1 Long-term well production forecasting model

2.2 灰色预测动态模型在短期油井产量中的应用

油井的短期产量到生产曲线符合指数增长趋势。因此,考虑在采用灰色预测模型 GM(1,1)模型的基础上,对该模型进行改进,使得该模型在对短期数据预测时具有很好实效性。

在进行灰色模型构建时,为保证预测的准确性,需要考虑数据对于模型的实效性影响(见图 2)。

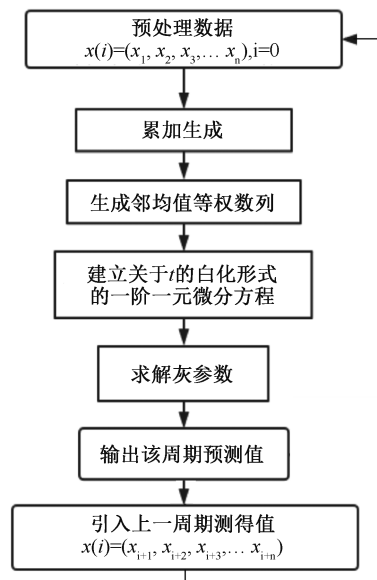


图 2 短期油井产量预测模型

Fig. 2 Short-term well production forecasting model

在实际预测过程中模型数据输入量 $x(i)$ ($i \in [4, 10]$) 根据实时测得值更新预测模型。在下次预测开始前,会用新测得的数据更新模型。

3 实例研究

为验证所提出的油井长期产量预测算法在实际的工业数据中的精确度,采用了某井开井后两年的数据进行长期产量和短期产量预测,并对预测结果进行分析。

3.1 长期产量预测结果及分析

该模型用于预测 2019 年 3 月 20 日至 2020 年 7 月 20 日未关井的油产量数据作为算法的验证实例(见表 1),得到用 BP 神经网络拟合修正过后的模型的实际累产和预测累产的曲线图以及相对误差值(见图 3)。

由表 1 可见,传统的 ARIMA 模型所得到的平均相对误差为 0.156 6,国亮法模型所得到的平均相对误差为 0.137 2,寸少妮法模型所得到的平均相对误差为 0.144 8,根据 BP 神经网络拟合残差优化的长期产量预测模型的平均相对误差为 0.091 4。根据统计学指标对比得出,长期产量预测模型的改进型灰色预测模型较为准确,预测精确度较高。经过实验验证,该预测模型在两个月以上的时间范围内均具有较高的计算精度。

3.2 短期产量预测结果分析

为验证该模型应用的广泛性,对短期产量预测结果进行了验证分析,选用了不同地区四个井口的产量数据进行验证对比(见图 4)。

表 1 长期产量预测结果误差分析				
Table 1 Error analysis of long-term production forecast results				
实际产油量 累计值/m ³	绝对误差/m ³			
	传统 ARIMA 模型	国亮法	寸少妮法	BP 神经网络 优化后模型
58.64	0.000	0.005 1	0.003 5	0.005 2
528.44	2.093 1	1.306 9	0.942 1	2.098 3
998.18	3.754 5	2.521 5	1.685 2	3.759 7
1 468.34	6.754 4	4.875 4	3.513 2	3.759 6
1 935.72	8.097 5	6.015 1	6.648 5	6.102 8
2 400.47	7.861 3	8.159 6	9.046 7	2.866 5
2 860.47	5.965 5	10.856 3	11.865 2	5.970 7
3 330.79	14.311 6	11.082 5	12.018 4	4.316 8
3 790.16	10.628 3	9.138 7	12.032 1	6.633 5
4 251.98	11.787 1	12.586 8	12.0452	1.792 3
4 690.98	-3.747 5	-4.102 9	11.968 3	-3.742 3
5 132.98	-12.476 8	5.163	11.698 9	-2.471 6
5 581.40	-12.410 1	-9.441 5	12.075 4	-7.404 9
6 040.18	-4.360	4.647 9	3.461 1	-1.354 8
6 495.08	-3.558	8.546 1	-4.684 7	2.447 2
6 948.44	-4.879	10.315 2	-9.351 2	-1.873 8
7 683.73	-95.024 2	-15.368 4	-20.864 5	-5.018 9
8 121.29	-73.969 4	-20.564 1	-15.645 1	-3.964 2
8 569.13	-59.563 5	-25.761 5	-44.864 7	-9.558 3
9 023.02	-50.546 1	-19.154 7	-35.287 9	-10.540 9
平均相对误差	0.156 6	0.137 2	0.144 8	0.091 4

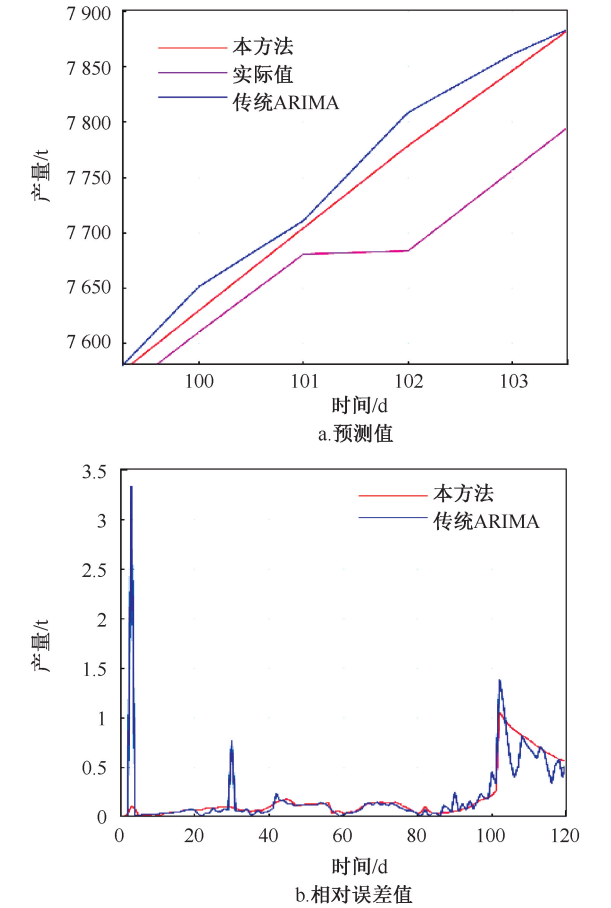


图 3 长期产量预测结果
Fig. 3 Long-term production forecasts

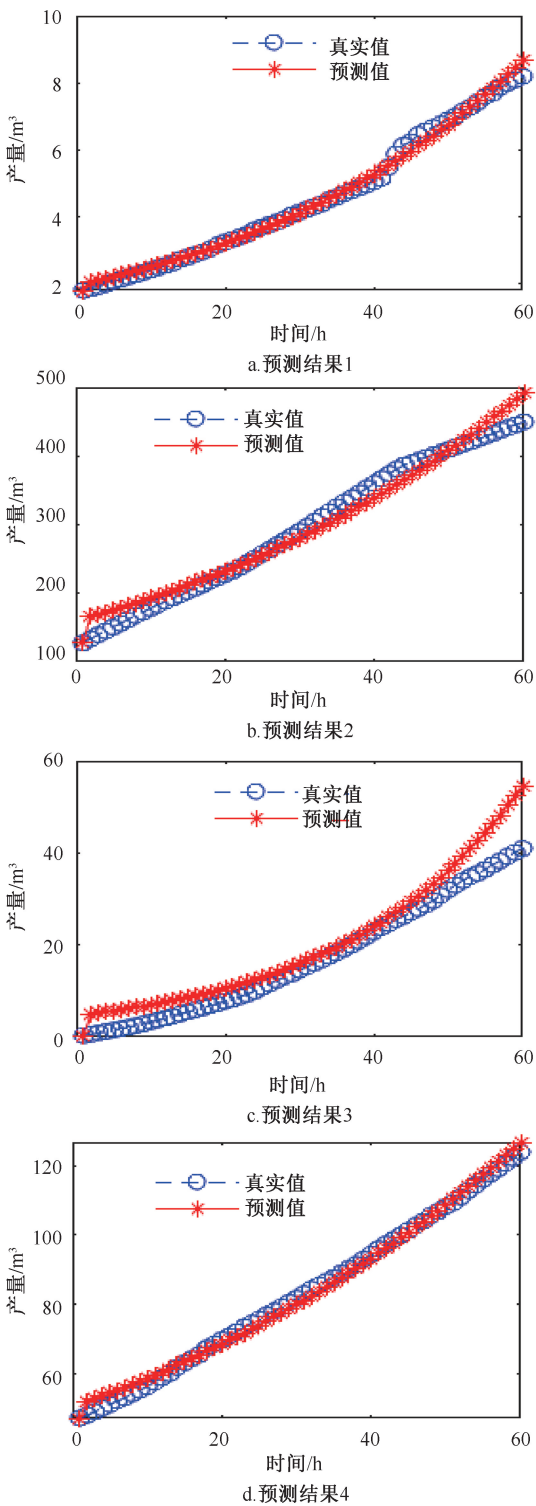


图 4 短期产量预测结果
Fig. 4 Short-term production forecasts

由图 4 可见,预测结果 2 与预测结果 3 在后期分别出现了不同程度的拟合偏移,通过对灰色模型的性能分析,结合石油生产实际工况分析得出以下几种情况:

(1)该井已进入见水期,油井见水后产量波动

较大,出现产能曲线无法快速跟随正常生产的趋势。

(2)由于采集设备的自身缘故,在采集过程中,不可避免会造成井筒积液的不断累积,额外的筒内压力会使采集的过程逐渐放缓。

(3)根据产量递减规律,考虑到实际工况中,每口井的油藏各不相同,采集到一定程度会出现油压的不同程度下降,导致产量的突变。

分别对正常状况下的数据与非正常状况下的数据进行计算分析,反映了在不同模型长度下正常工况与非正常工况分别利用动态灰色模型预测的精度情况(见表2)。

表 2 不同工况下不同测量周期的产量平均绝对误差

Table 2 Average absolute error of output for different measurement cycles under different operating conditions

测量周期 t/h	40	50	60	75	90
正常工况	0.116 9	0.117 6	0.118 7	0.119 6	0.120 3
非正常工况	38.201 6	20.428 7	14.811 5	12.432 1	10.567 7

由表2可以看出,在正常工况下,随着模型长度的不断缩小,产量预测精度会有所提高。但在非正常情况下,测量周期越小,模型的测量结果的平均绝对误差越来越大。因此,该模型可快速反映出生产情况下的非正常工况,有助于现场判断生产情况。

对预测结果1与预测结果4进行分析验证,剔除了后期干扰参数后的部分抽样数据见表3。

表 3 短期产量预测结果误差分析

Table 3 Error analysis of short-term production forecast results

实际产油量 累计值/m ³	传统灰色 模型	崔传智法	改进后 灰色模型
1.77	0	0	0
2.93	-0.068 5	-0.052 3	-0.020 3
3.01	-0.185 7	-0.095 6	-0.015 2
3.91	0.042 5	0.039 1	0.032 7
4.12	-0.002 8	0.085 2	0.042 1
4.93	0.247 1	0.210 7	-0.187 1
5.86	0.184 2	0.171 0	0.201 4
6.66	0.518 5	0.421 9	0.306 5
7.58	-0.726 9	-0.474 5	-0.262 7
平均绝对误差	0.219 6	0.172 2	0.118 7

由表3可见,传统灰色预测模型的平均绝对误差为0.219 6,崔传智法的 PSO-GM(1,1)模型所得到的平均绝对误差为0.172 2,根据实时数据更迭的动态预测模型的平均绝对误差为0.118 7。根据统计学指标,井口正常开采的情况下,改进型灰色预测模型更为准确,预测精确度更高,在3~10 d的时间范围内均具有较高的计算精度。

用油井产量长短期分类预测方法以及传统 ARIMA、国亮法、寸少妮法、薛亮法、崔传智法分别对0~38个月的生产结果进行预测比较,利用相对误差的散点分布图来进行评价(见图5)。

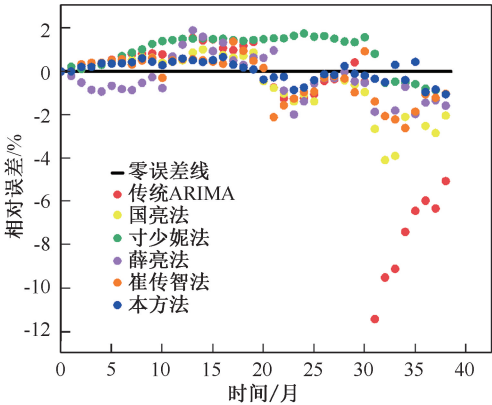


图 5 各方法预测结果对比

Fig. 5 Comparison of prediction results by method

由图5可以看出,传统的 ARIMA 算法对不同时间段内的单井产量预测效果不稳定,使用范围不广泛。BP 神经网络对短期产量的特征值提取分析,优化了时间序列预测模型在实际应用上的残差。因此,长短期分类预测模型的预测结果相对稳定,测量结果误差较小。

4 结论

(1)通过对部分井口的产量计量数据进行预处理与分析,发现长短期分类预测模型更适用于油井试油试采及前期生产阶段。

(2)石油开采初期累产曲线呈现指数上升的趋势,因此灰色预测模型的算法保证了短期内曲线拟合程度更高,产量预测精确度更好。

(3)长短期分类预测模型针对单井生产周期不同的情况下的长、短期产量预测方法预测结果更精确、避免了多元分析法的繁琐、计算效率更高,能够满足实际工况的需求。

致谢:感谢国家重点研发计划项目“北方砂岩型铀能源矿产基地深部探测技术示范项目”;教育部产学研协同育人项目“电子测量”;广西科技大学产学研合作项目“油气田地面计量数字采集系统研制及现场应用”资助。

参考文献

[1] 赵改善. 石油物探数字化转型之路:走向实时数据采集与自动化处理智能化解释时代[J]. 石油物探, 2021, 60(2):175-189.

ZHAO Gaishan. Digital transformation of petroleum geophysical exploration: towards the era of real-time data

- acquisition, automatic processing and intelligent interpretation [J]. *Geophysical Prospecting for Petroleum*, 2021, 60 (2): 175-189.
- [2] 薛亮,顾少华,王嘉宝,等. 基于粒子群优化和长短期记忆神经网络的气井生产动态预测[J]. *石油钻采工艺*, 2021, 43(4): 525-531.
- XUE Liang, GU Shaohua, WANG Jiabao, et al. Production dynamic prediction of gas well based on particle swarm optimization and long short-term memory [J]. *Oil Drilling & Production Technology*, 2021, 43(4): 525-531.
- [3] 李彦尊,白玉湖,陈桂华,等. 基于人工神经网络方法的页岩油气产量预测新技术——以美国 Eagle Ford 页岩油气田为例[J]. *中国海上油气*, 2020, 32(4): 104-110.
- LI Yanzun, BAI Yuhu, CHEN Guihua, et al. ANN method based on novel technology for production of shale oil and gas: a case study in Eagle Ford [J]. *China Offshore Oil and Gas*, 2020, 32(4): 104-110.
- [4] 周鹏,陈明,王庆勇,等. 预测油气田产量和可采储量的广义 Weibull 模型[J]. *大庆石油地质与开发*, 2017, 36(5): 93-97.
- ZHOU Peng, CHEN Ming, WANG Qingyong, et al. Generalized Weibull model for predicting the production and recoverable reserves of oil and gas fields [J]. *Petroleum Geology & Oilfield Development in Daqing*, 2017, 36(5): 93-97.
- [5] 崔传智,吴忠维,李昱东,等. 应用 PSO-改进 GM(1,1) 模型预测油田产量[J]. *数学的实践与认识*, 2018, 48(17): 119-123.
- CUI Chuazhi, WU Zhongwei, LI Yudong, et al. Application of the PSO-Modified GM (1,1) model to oilfield production forecast [J]. *Practice and Understanding of Mathematics*, 2018, 48(17): 119-123.
- [6] 寸少妮,陈涛,辛毅超,等. 产量递减规律分析及产量预测[J]. *中国石油和化工标准与质量*, 2021, 41(17): 29-30.
- CUN Shaoni, CHEN Tao, XIN Yichao, et al. Analysis of production decline law and production prediction [J]. *China Petroleum and Chemical Standard and Quality*, 2021, 41(17): 29-30.
- [7] 国亮. 改进多元线性回归模型在某油田产量预测中的应用[J]. *西安电子科技大学学报(社会科学版)*, 2009, 19(3): 71-75.
- GUO Liang. Application of improved multivariate linear regression to prediction of an oilfield [J]. *Journal of Xidian University (Social Science Edition)*, 2009, 19(3): 71-75.
- [8] IGNACIO ROJAS, HÉCTOR POMARES, OLGA VALENZUELA. *Advances in time series analysis and forecasting* [M]. Springer, Cham.
- [9] BOX GEORGE E P, JENKINS GWILYM M, REINSEL GREGORY C. *Time series analysis* [M]. John Wiley & Sons, Inc.: 2008-06-12.
- [10] CHEN Lin, LIU Zhibin, MA Nannan. Optimize production allocation for the oil-gas field basing on a novel grey model [J]. *Journal of Natural Gas Geoscience*, 2019, 4(2): 121-128.
- [11] HAMILTON JAMES D. *Time series analysis* [M]. Princeton University Press: 2020-09-01.
- [12] WANG Xiaoli. Research on the prediction of per capita coal consumption based on the ARIMA-BP combined model [J]. *Energy Reports*, 2022, 8(4): 285-294.
- [13] 杨洋,程悦菲,谯英,等. 基于时序动态分析的油井产量预测研究[J]. *西南石油大学学报(自然科学版)*, 2020, 42(6): 82-88.
- YANG Yang, CHENG Yuefei, QIAO Ying, et al. A study on oil well production prediction based in time series dynamic analysis [J]. *Journal of Southwest Petroleum University (Science & Technology Edition)*, 2020, 42(6): 82-88.
- [14] 李建周,关利军,高永海,等. 深水气井测试期间不同产量条件下水合物形成区域预测方法[J]. *油气井测试*, 2012, 21(2): 17-19.
- LI Jianzhou, GUAN Lijun, GAO Yonghai, et al. Regional prediction method of hydrate formation under different yield conditions during deepwater gas well testing [J]. *Well Testing*, 2012, 21(2): 17-19.
- [15] 丁显峰,张锦良,刘志斌. 油气田产量预测的新模型[J]. *石油勘探与开发*, 2004, 180(3): 104-106.
- DING Xianfeng, ZHANG Jinliang, LIU Zhibin. A new model for oil and gas field production prediction [J]. *Petroleum Exploration and Development*, 2004, 180(3): 104-106.
- [16] 高文君,徐君,王作进,等. 对油气田产量预测广义模型的完善与研究[J]. *石油勘探与开发*, 2001, (5): 56-59.
- GAO Wenjun, XU Jun, WANG Zuojin, et al. Improvement and Research on generalized model of oil and gas field production prediction [J]. *Petroleum Exploration and Development*, 2001, (5): 56-59.

编辑 吴志力

第一作者简介:靳子豫,男,1998年出生,2019年毕业于河北工业大学城市学院电气工程及其自动化专业,现为广西科技大学控制理论与控制工程专业在读硕士研究生,从事油气领域数据采集与分析系统的研究工作。电话:13512009551, Email:770228215@qq.com。通信地址:广西壮族自治区柳州市鱼峰区官塘大道19号广西科技大学柳东校区自动化学院,邮政编码:545616。