

基于深度自回归神经网络的多井产量概率预测

韩江峡^{1,2}, 薛亮^{1,2*}, 位云生³, 齐亚东³, 王军磊³, 陈海洋^{1,2}, 刘月田^{1,2}

1 中国石油大学(北京)油气资源与工程全国重点实验室, 北京 102249

2 中国石油大学(北京)石油工程学院, 北京 102249

3 中国石油勘探开发研究院, 北京 100083

* 通信作者, xueliang@cup.edu.cn

收稿日期: 2024-02-23; 修回日期: 2024-04-16

国家自然科学基金(52274048)、北京市自然科学基金(3222037)和中国石油天然气股份有限公司“十四五”前瞻性基础性科技项目子课题“致密气生产规律与开发接替模式研究”(2021DJ2104)联合资助

摘要 传统产量预测方法易受到单井生产历史和模型假设条件的限制, 预测结果无法量化不确定性, 难以考虑区块其他生产井开发规律对目标井的指导作用, 无法充分利用大量相关的生产历史数据。为此, 提出一种以深度自回归神经网络为基础, 多井产量数据驱动的概率预测新模型。考虑生产时间、油/套压等动态协变量数据, 结合贝叶斯推断, 利用梯度下降算法和极大似然估计方法, 得到多井共有的广义历史—未来产量概率演化模式, 实现基于数据驱动的多井产量概率预测。利用鄂尔多斯盆地某两个区块 943 口致密气井的数据, 研究了深度自回归神经网络模型在单井预测、分类预测和总体区块产量预测上的性能。研究结果表明: 相比传统深度学习模型(LSTM), 新模型利用学习得到的广义产量概率演化模式与目标井的特定产量历史数据相结合, 形成“广义+特定”的产量概率预测方法, 平均意义上较 LSTM 模型相对误差降低了 45%。分类模型较全局模型相对误差降低了 24%, 实现了在全局模型的基础上, 进一步降低了概率预测的不确定性, 提高了特定精细分类井的预测精度。经过实际数据验证, 新模型预测精度更好, 鲁棒性更强, 可以用于油气藏多井产量预测分析。

关键词 产量预测; 多井预测; 神经网络; 致密气; 概率预测; 区块预测

中图分类号: TE332; TP18

Multiple well production rate probabilistic forecasting using deep autoregressive recurrent networks

HAN Jiangxia^{1,2}, XUE Liang^{1,2}, WEI Yunsheng³, QI Yadong³, WANG Junlei³, CHEN Haiyang^{1,2}, LIU Yuetian^{1,2}

1 State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum-Beijing, Beijing 102249, China

2 College of Petroleum Engineering, China University of Petroleum-Beijing, Beijing 102249, China

3 PetroChina Research Institute of Petroleum Exploration & Development, Beijing 100083, China

Received: 2024-02-23; Revised: 2024-04-16

Abstract Traditional production rate forecasting methods are often limited by the production history of individual wells

引用格式: 韩江峡, 薛亮, 位云生, 齐亚东, 王军磊, 陈海洋, 刘月田. 基于深度自回归神经网络的多井产量概率预测. 石油科学通报, 2024, 04: 679-689

HAN Jiangxia, XUE Liang, WEI Yunsheng, QI Yadong, WANG Junlei, CHEN Haiyang, LIU Yuetian. Multiple well production rate probabilistic forecasting using deep autoregressive recurrent networks. Petroleum Science Bulletin, 2024, 04: 679-689. doi: 10.3969/j.issn.2096-1693.2024.04.051

and assumptions of the models, leading to unquantified uncertainties in the prediction results and difficulty in considering the guidance of development patterns from other wells in the block on the target well. Additionally, they fail to fully utilize a large amount of relevant production history data. To address these issues, a new model for probabilistic production rate forecasting driven by multi-well production data is proposed, based on deep autoregressive neural networks. This model integrates dynamic covariate data such as production time and tubing/casing pressure, and employs Bayesian inference along with gradient descent and maximum likelihood estimation methods to derive a generalized historical-future production probability evolution pattern shared among multiple wells. Through data-driven approaches, it achieves probabilistic production forecasting for multiple wells. The performance of the deep autoregressive neural network model is studied using data from 943 tight gas wells in two blocks in the Ordos Basin. Results indicate that compared to traditional deep learning models like LSTM, the new model combines the learned generalized production probability evolution pattern with specific production history data of the target well, forming a “generalized + specific” production probability prediction method. On average, it reduces the relative error by 45% compared to the LSTM model. The classification model reduces the relative error by 24% compared to the global model, further reducing the uncertainty of probability prediction based on the global model and improving the prediction accuracy of specific fine-classified wells. Through validation with actual data, the new model demonstrates better prediction accuracy and stronger robustness, making it applicable for multi-well production forecasting analysis in oil and gas reservoirs.

Keywords production rate forecasting; multiple well production rate forecasting; neural networks; tight gas; probabilistic forecasting; block production rate forecasting

doi: 10.3969/j.issn.2096-1693.2024.04.051

0 引言

随着油气勘探开发不断的深入,我国的勘探重点已经从常规油气资源转向非常规油气资源,近年来致密砂岩气的勘探开发越来越受到重视^[1]。低渗致密气田资源总量占国内非常规天然气资源总量50%以上,资源总量庞大,具有良好的发展前景^[2-3]。不同于传统常规油气藏,致密气藏具有低渗透、低丰度、低产的特点,多采用多级压裂水平井或多层压裂直井的方式进行开发,并且存在“启动压力梯度”和“应力敏感效应”等现象,涉及到多场耦合和非达西渗流等复杂过程,综合外在表现为单井产量下降快,稳产期短的特点^[4-6]。

目前油气藏工程领域的产量预测方法包括以Arps为代表的经验递减模型方法^[7-9]、以不稳定渗流产量递减动态为基础的解析方法^[10-13]以及基于数据驱动的人工智能方法^[14-18]。梁倚维等^[8]建立了致密气井数值模型,通过划分致密气井各个渗流阶段响应时间,研究致密气井在各个渗流阶段递减指数变化规律,分析了4种不同因素对递减指数的影响,扩展了对Arps递减规律的认识,从而揭示了递减指数在致密气井整个预测期的变化规律。孙贺东等^[10]针对水力压裂形成的裂缝导流能力会随空间以及时间发生变化这一特点,建立考虑裂缝空间、时间双重变导流和应力敏感效应的致密气压裂直井渗流的新型数学模型,获得新模型的Blasingame现代产量递减分析典型曲线,以此为基础,讨论了裂缝空间、时间双重变导流及应力敏

感效应对Blasingame曲线形态的影响。王军磊等^[13]利用考虑气体滑脱效应的拟压力和拟时间变量,建立带有矩形封闭边界的分段压裂水平井渗流数学模型,应用Newman乘积、Laplace变换、相似流动替换及压力叠加原理求解模型以得到致密气井不稳态产量公式。对于数据驱动的人工智能方法,Zha等^[17]基于影响油井产量变化的60余个动静态因素,通过卷积神经网络(CNN)提取地质静态参数,长短期记忆神经网络(LSTM)模型构架时序预测模型,构建了CNN-LSTM模型实现了月度单井产量的精确预测。Xue等^[18]提出一种Arps+LSTM的物理数据联合驱动致密气产量预测方法,对比发现,联合驱动方法精度优于单纯Arps模型。这些方法都是以预测单井或者井组产量为背景开发的。

在这种情况下,每个给定产量时间序列的预测模型特征参数都是根据历史产量数据独立估计的。尽管存在良好且完备的自动化软件,但是在实践和研究中,通常都是手动选择模型,以考虑不同因素,然后根据拟合好的模型来预测未来的产量值,概率预测可能通过拟合分布的显示表达式来表现。这些方法主要存在以下3方面的问题:第一,从预测原理上来说,会受到单井生产历史和模型假设条件的限制,也难以考虑到区块其他生产井开发规律对目标井的指导作用。此外,这些方法未能充分利用大量相关的生产历史数据。第二,从预测效率上来说,在实际气田宏观调控部门,相比单井或者井组预测,更需要的是同时预测相关的成百上千口井的产量时间序列。各个气井生产

过程中可能会存在相互影响，或者各个单井属于同一井组不同分支井，所以并不是完全独立的，单纯利用并行的单井预测是不够的。第三，从预测形式上来说，传统单点预测并不能为后续的单井、井组、区块配产提供一个可供参考的执行区间，不能量化其不确定性，然而概率产量预测对于在正确的时间和地点提供正确的配产或生产制度调整至关重要^[19]。

在此背景下，提出了一种多井产量概率预测新模型，采用结合贝叶斯推断、梯度下降算法和极大似然估计方法的深度自回归神经网络模型^[20]对多井产量数据进行历史拟合和未来产量概率预测。新模型利用大量生产递减规律相似、相关的生产井的历史数据，得到多井共有的广义历史—未来产量概率演化模式，再结合单井特定的产量历史，形成“广义+特定”的产量概率预测方法，实现了多井产量历史建模和概率预测。避免了传统方法基于单井生产历史过拟合和模型假设条件先验不足的问题，减轻了选择和准备协变量特征以及选择模型所需的人工和劳动密集型步骤。新模型为油气藏工程领域提供了一种更准确、可靠且实用的产量预测框架，为宏观调控决策提供了更全面的信息基础。

1 深度自回归神经网络模型

深度自回归神经网络模型，是一种在时间序列上训练自回归循环神经网络的概率分布的方法，它可以有效地从大量相关的产量时间序列中学习一个全部井共有的广义历史—未来产量概率演化全局模式，可以学习复杂的模式，如现场的开关井操作和考虑协变量预测，可以进行概率预测建模。所谓全局模型即是能综合表征多井产量时间序列递减规律的一个通用模型，通过学习数据集中多井产量时间序列共同的演化模式得到的一个更深层的产量演化规律。外在表现为通过贝叶斯估计得到的产量概率分布，产量概率分布表示为一定长度的先验历史产量序列和一定长度的后验未来产量序列满足此分布。其基本原理：给定一定历史时间长度的数据，预测一定长度的未来数据，也就是序列对序列的预测。给定一个时间序列 i ，其在 t 时刻的值用 $z_{i,t}$ 表示， t_0 时刻之前的时间序列值可以表示为： $\mathbf{Z}_{i,1:t_0-1} = [z_{i,1}, z_{i,2}, \dots, z_{i,t_0-2}, z_{i,t_0-1}]$ ，需要预测 T 个时间步的时间序列可以表示为 $\mathbf{Z}_{i,t_0:T} = [z_{i,t_0}, z_{i,t_0+1}, \dots, z_{i,T}]$ 。根据贝叶斯定理，利用已知的先验概率和观测数据来估计未知的后验概率，在这里也就是根据先验的产量观测数

据来估计未来的后验产量分布，定义 $[1, t_0 - 1]$ 为条件范围， $[t_0, T]$ 为预测时间范围。于是对未来的时间序列预测时，有条件概率 P ，则可以表示：

$$P(\mathbf{Z}_{i,t_0:T} | \mathbf{Z}_{i,1:t_0-1}, \mathbf{X}_{i,1:T}) \quad (1)$$

式中 t_0 表示开始预测的时刻； $\mathbf{Z}_{i,t_0:T}$ 表示预测变量； $\mathbf{X}_{i,1:T}$ 表示到 T 时刻的协变量数据。

假定协变量数据是已知的，则 P 的概率分布可以表示为：

$$\begin{aligned} Q_{\Theta}(\mathbf{Z}_{i,t_0:T} | \mathbf{Z}_{i,1:t_0-1}, \mathbf{X}_{i,1:T}) \\ = \prod_{t=t_0}^T Q_{\Theta}(z_{i,t} | \mathbf{Z}_{i,1:t-1}, \mathbf{X}_{i,1:T}) \\ = \prod_{t=t_0}^T \ell(z_{i,t} | \theta(\mathbf{h}_{i,t}, \Theta)) \end{aligned} \quad (2)$$

式中 Θ 表示模型参数； $\mathbf{h}_{i,t}$ 表示时序神经网络中最后一个隐藏层的输出值。

此模型是自回归模型，上一个时刻的输出值会作为下一个时刻的输入值，同时作为循环神经网络，上一个时刻的隐藏网络参数 $\mathbf{h}_{i,t}$ 会作为下一个时刻的输入。 $\ell(\mathbf{h}_{i,t}, \Theta)$ 似然函数用作预测变量的分布范围，其分布是固定的，参数由神经网络确定。若选择高斯分布作为似然函数，使用样本的均值 μ 和标准差 σ 作为待估计参数，则高斯似然函数 ℓ_G 可以表示为：

$$\ell_G(z | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{ \frac{-(z-\mu)^2}{2\sigma^2} \right\}} \quad (3)$$

均值需要对 $\mathbf{h}_{i,t}$ 进行线性变换得到，标准差需要保证大于零，于是进行 Softplus 激活函数非线性变换得到：

$$\mu(\mathbf{h}_{i,t}) = \mathbf{w}_{\mu}^T \mathbf{h}_{i,t} + b_{\mu} \quad (4)$$

$$\sigma(\mathbf{h}_{i,t}) = \log\left(1 + e^{\left(\mathbf{w}_{\sigma}^T \mathbf{h}_{i,t} + b_{\sigma}\right)}\right) \quad (5)$$

式中 \mathbf{w}_{μ}^T ， \mathbf{w}_{σ}^T 表示变换权重； b_{μ} ， b_{σ} 表示变换偏置。

给定多个时间序列 $\{\mathbf{Z}_{i,1:T}\}_{i=1,\dots,N}$ 和协变量 $\mathbf{X}_{i,1:T}$ ，给定 $z_{i,t}$ 作为已知的分布，则最大化的高斯似然函数如下：

$$\mathcal{L} = \sum_{i=1}^N \sum_{t=t_0}^T \log \ell(z_{i,t} | \theta(\mathbf{h}_{i,t})) \quad (6)$$

式中 \mathcal{L} 表示最大化的高斯似然函数； N 表示样本个数。

通过最大化似然函数 \mathcal{L} 作为神经网络的概率损失函数来进行反向传播，优化网络最后得到最优的参数组合。网络模型的训练和预测过程如图 1 所示，训练和预测过程区别主要在于解码过程，训练过程中所有

数据都是已知的,所以输入是已知的观测值,而预测过程输入是通过抽样得到一个估计值 $\bar{z}_{i,t-1}$,将其输入到下一个时刻的循环体(cell)中,以此不断的迭代得到需要的预测结果。其相比于传统的循环神经网络模型,其最大的特点是多元预测和多步预测,即可以同时预测多个时间序列,多个时间步。而传统循环神经网络模型只能针对单个时间序列进行单步预测,其多步是通过滚动预测或者改变全连接层输出维度实现的。

2 方法流程及评价指标

首先根据预测任务和预测长度将每条单井产量数据分为训练集和测试集,选择生产时间或者油套压力为动态协变量数据,以进行辅助训练,根据实际任务选择静态分类特征,例如井型或产层等。第二,构建基于循环网络框架的时序基础模型,有3种选择:循环神经网络(RNN)、长短期记忆神经网络(LSTM)、门控神经网络(GRU)。这里以循环神经网络为例,说明方法流程,循环神经网络是由一个或多个循环体(cell)组成,每个循环体接受当前时刻的输入和上一时刻的隐藏状态,并输出当前时刻的隐藏状态。其特征参数为输入层到隐藏层的权重,隐藏层到隐藏层的权重以及偏置。循环神经网络后紧跟两个全连接神经网络,分别对应产量分布的均值和标准差的估计,由于需要保证标准差为大于零,需要Softplus激活函数限制输出。然后,将产量数据和动静态特征数据整理成标准格式,输入到循环神经网络的基本框架中。模型使用贝叶斯推断来对时间序列数据进行建模,通过贝叶斯推断,模型可以估计参数的后验分布,这使得模型可以提供对未来数据点的预测不确定性估计。神经网络会根据网络参数的初始值计算概率分布损失,使用梯

度下降算法计算最大化对数后验概率(或最小化负对数后验概率),从而更新模型参数,求解极大似然函数得到均值和方差的最佳值,从而估计得到一个初步概率分布函数,并且更新神经网络特征参数。往复循环上述步骤,以达到最终概率分布损失为0,最后得到最优的神经网络特征参数和产量概率分布参数,整个模型的流程图如图2所示。

后续模型性能测试采用NRMSE(Normalized root mean square error)、ND(Normalized deviation)以及覆盖率CPI(α)(Coverage of probability intervals)来评价。NRMSE描述的是预测值与真实值的以均方计算的距离,ND则描述的是预测值与真实值偏差占真实值的比例,反映偏离的幅度。CPI(α)用来评价模型概率预测误差程度,表示某概率区间内覆盖真实值的比例。三者公式如下:

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{N(T-t_0)} \sum_{i,t} (y_{i,t} - \hat{y}_{i,t})^2}}{\frac{1}{N(T-t_0)} \sum_{i,t} |y_{i,t}|} \quad (7)$$

$$\text{ND} = \frac{\sum_{i,t} |y_{i,t} - \hat{y}_{i,t}|}{\sum_{i,t} y_{i,t}} \quad (8)$$

$$\text{CPI}(\alpha) = \frac{\text{count}(\hat{y}_{(1-\alpha)/2} < y < \hat{y}_{(1+\alpha)/2})}{N} \quad (9)$$

式中 y 表示真实值; \hat{y} 表示模型预测值;count函数表示满足条件的样本个数; α 表示分位点,小数; $\hat{y}_{(1-\alpha)/2}$ 表示预测概率区间的下边界; $\hat{y}_{(1+\alpha)/2}$ 表示预测概率区间的上边界。

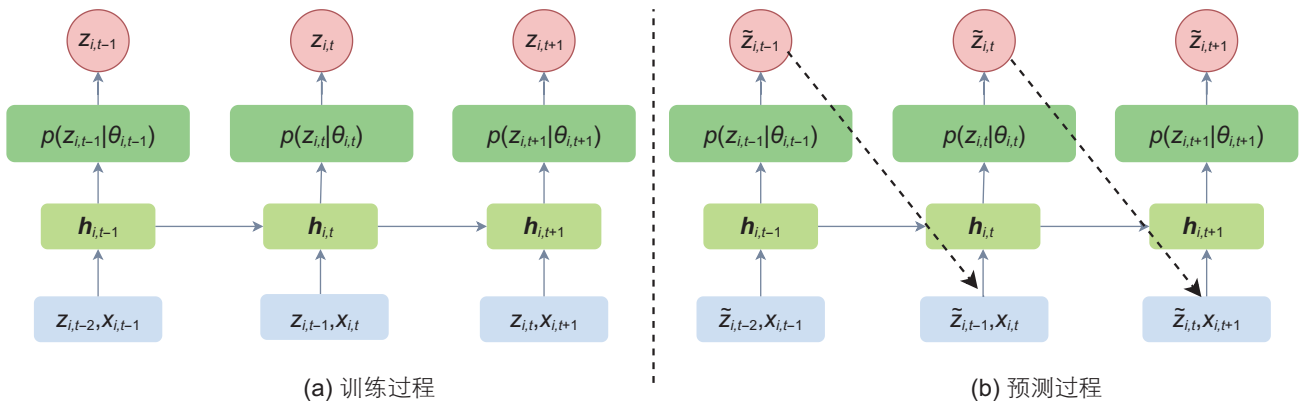


图1 深度自回归神经网络训练过程(a)和预测过程(b)

Fig. 1 Deep autoregressive recurrent networks training process (a) and prediction process (b)

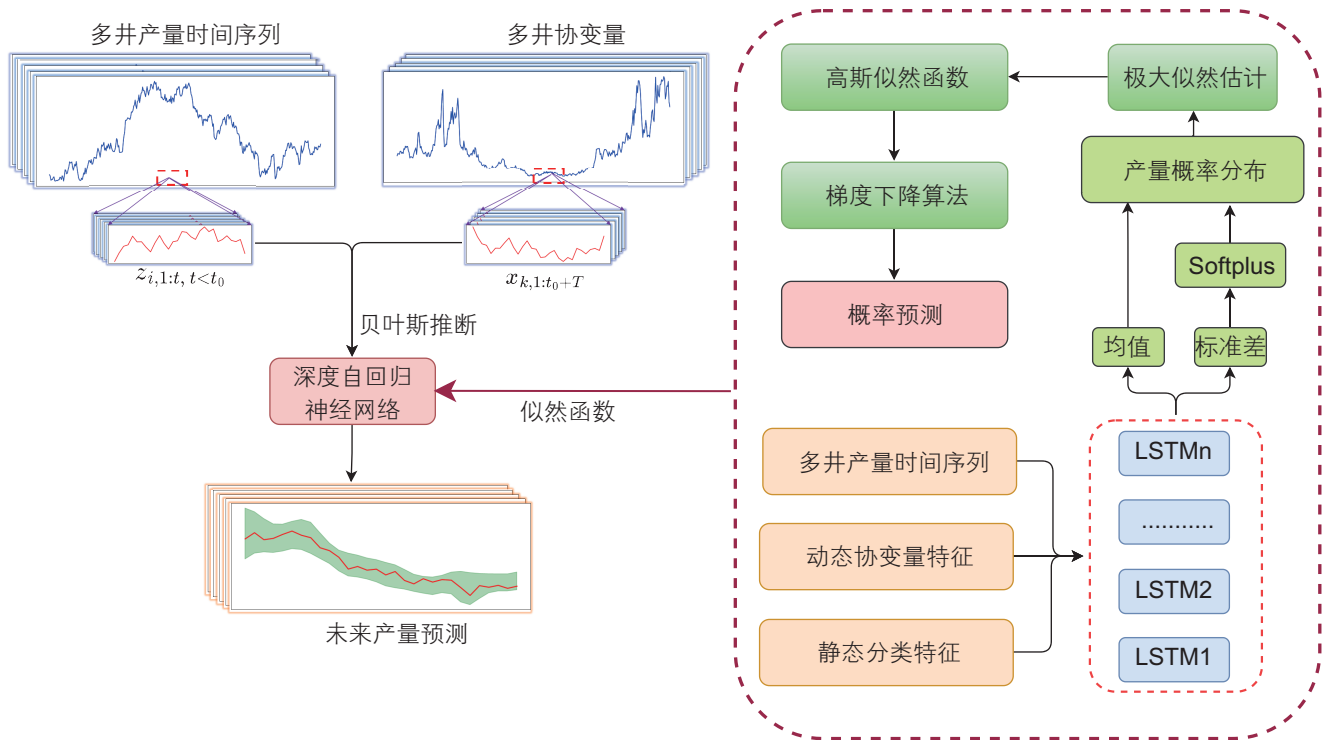


图 2 基于深度自回归神经网络的多井产量预测流程图

Fig. 2 Flow chart of multi-well production rate forecasting based on deep autoregressive recurrent networks

3 实例应用

本文使用的数据来自鄂尔多斯盆地的 943 口致密气井生产数据，其中，水平井 192 口，直井 751 口。基于深度自回归神经网络模型，分别开展了与 LSTM 模型的单井预测精度对比研究、基于气井预分类模型和未分类的全局模型预测精度对比研究、区块总体产量预测研究。

3.1 模型预测精度对比

在本节中，采用全部数据集来训练深度自回归神经网络模型，该模型的输入数据包括各井的产量历史数据以及各井每天的生产时间，网络的输出是产量的未来概率分布。使用数据集最后 30 天作为测试段，之前所有皆为训练集，模型优化器为 Adam，学习率为 10^{-3} ，梯度裁剪 (clip gradient) 极限为 10，训练过程中采用权重衰减策略，权重衰减因子为 10^{-8} 。超参数采用开源神经网络超参数调优框架 Optuna 的 TPESampler 类进行优化，并且与相同网络结构的考虑生产时间的多通道 LSTM 模型进行对比，具体的网络的超参数如表 1 所示。

选取具有代表性的 6 口井，其对比结果如图 3 和

表 2 所示。为了更加清晰的对比预测阶段结果，图中只画出了 200 天的历史数据。从预测结果上可以清晰看到深度自回归神经网络模型的预测结果更加复合真实数据的变化趋势，并且 50% 的置信区间就可以涵盖大部分真实值的变化范围。图 3d、3e、3f 展示了深度自回归神经网络在频繁开关井的上的效果，其结果表明其相比于传统的 LSTM 模型更擅长提取开关井的特

表 1 网络超参数设置

Table 1 Network hyperparameter settings

参数	本文模型	LSTM 模型
Freq	1D	1D
Context length	30	30
Prediction length	30	1
Number of layers	2	2
Number of cells	30	30
Epochs	100	100
Batch size	32	32
Output distribution	Student T Output	/

注：Freq 表示网络训练和预测的基本单位，天；Context length 表示历史长度；Prediction length 表示预测长度；Number of layers 表示网络的层数；Number of cells 表示每层网络神经元个数；Epochs 表示训练集迭代步数；Batch size 表示每个训练步骤中输入的样本数量；Output distribution 表示网络输出的概率分布函数；Student T Output 表示 T 分布。

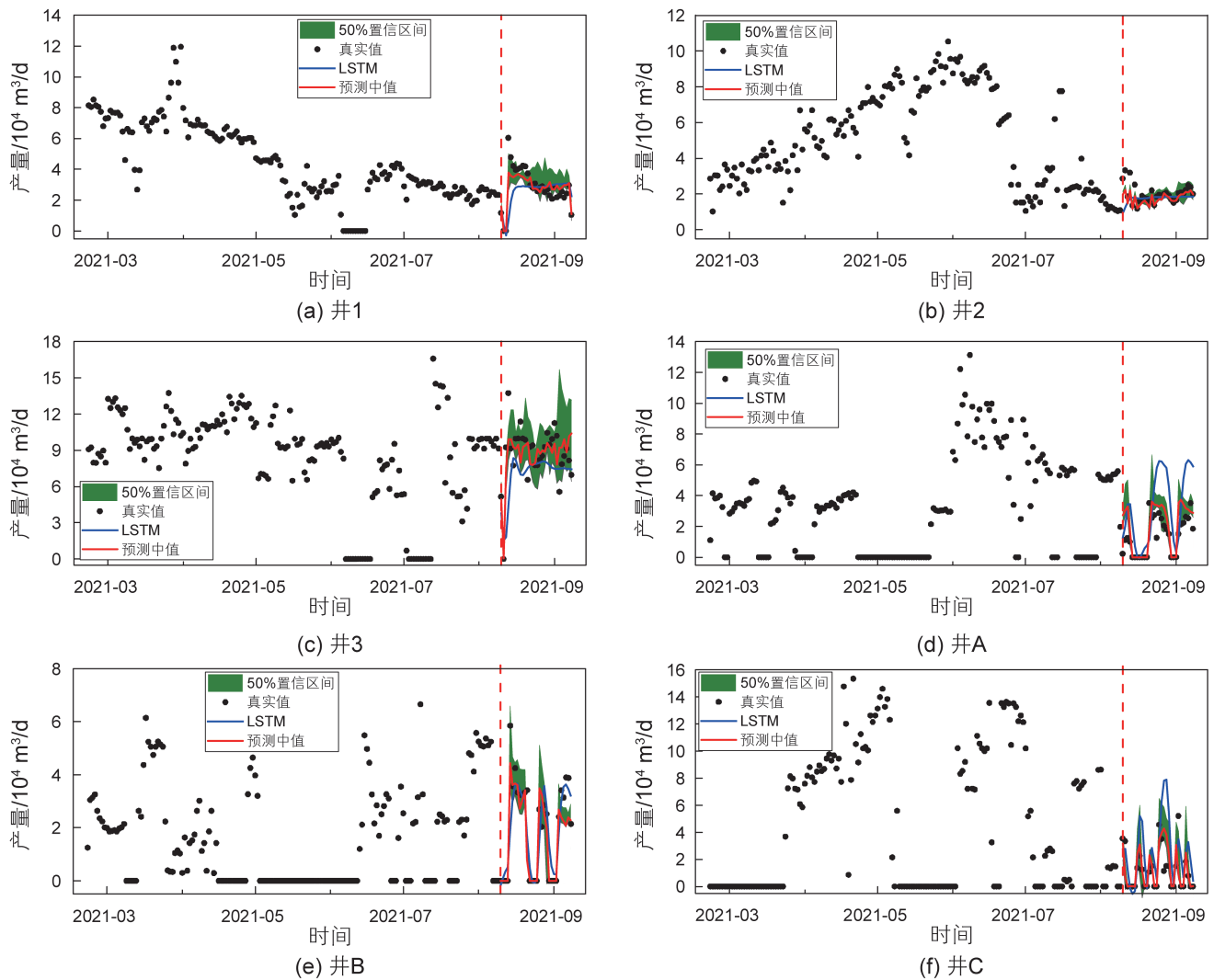


图3 深度自回归神经网络模型与LSTM模型的预测精度对比

Fig. 3 Comparison of prediction accuracy between deep autoregressive recurrent networks and LSTM model

表2 深度自回归神经网络模型与LSTM的预测指标对比

Table 2 Comparison of predictive metrics of deep autoregressive recurrent networks and LSTM

井名	模型	NRMSE	ND
Well-1	新模型(中值)	0.2293	0.1681
	LSTM	0.5478	0.3004
Well-2	新模型(中值)	0.1540	0.1351
	LSTM	0.3428	0.2108
Well-3	新模型(中值)	0.2040	0.1593
	LSTM	0.3859	0.2385
Well-A	新模型(中值)	0.8072	0.5767
	LSTM	2.3190	1.4740
Well-B	新模型(中值)	0.4085	0.2412
	LSTM	1.1398	0.3080
Well-C	新模型(中值)	0.4085	0.2412
	LSTM	2.4831	1.4094
平均值	新模型(中值)	0.3686	0.2536
	LSTM	0.8507	0.4642

征。LSTM模型通常会提供一个趋势线，但是其建模能力受到开关井的次数和模型本身的超参数影响，常常模型不收敛。从6口单井预测结果的平均意义上来说，相同网络结构的深度自回归神经网络预测结果相比LSTM模型，NRMSE可以降低57%，ND可以降低45%，表明了深度自回归神经网络模型性能的优越性。通过利用从数百口井的生产历史中获得的信息，深度自回归神经网络模型能够学习到这批井广义产量概率演化模式，将这种广义模式与目标井的特定历史数据相结合，形成了“广义+特定”的产量预测方法。

3.2 深度自回归神经网络的全局模型和分类模型精度对比

在前一节中，利用全部数据集，对比了深度自回归神经网络模型和LSTM模型在单井上的预测性能，

训练集并没有对单井进行精细分类(井型、产层、气水比等),学习的是一个更宏观的规律。实际往往一类具有相同特征的气井产量曲线具有相同的变化模式,但是传统以单井预测为背景开发的方法无法对一类产量时间序列进行建模和预测。深度自回归神经网络模型天然支持多井产量预测建模,支持分类学习,因此本节探究其用来提取某一类的产量变化特征的分类预测能力。考虑到本研究所用致密气动态数据集所在储层含气性差异大、局部气水关系复杂,单井产量变化

趋势往往受到产水的影响较大^[21]。为进一步评估模型的性能,从所有数据集中挑选满足分类指标为“直井+水气比(WGR)<0.1 m³/10⁴ m³”的井作为分类预测目标子集。为突出效果,用于对比预测(全局模型)的数据集包括分类预测目标子集,以及数量相等的其他不满足分类指标的井,如表3所示。分别利用全部数据集(全局模型)和目标子集(分类模型)进行训练,然后分别评估其在目标分类子集上的预测效果。其产量预测结果如图4和图5所示,评价指标如表4所示。由

表3 模型训练所用数据集

Table 3 Data sets used for model training

类型	范围	数量
全局模型	所有数据集	372
分类模型	直井、单层 WGR<0.1 m ³ /10 ⁴ m ³	186

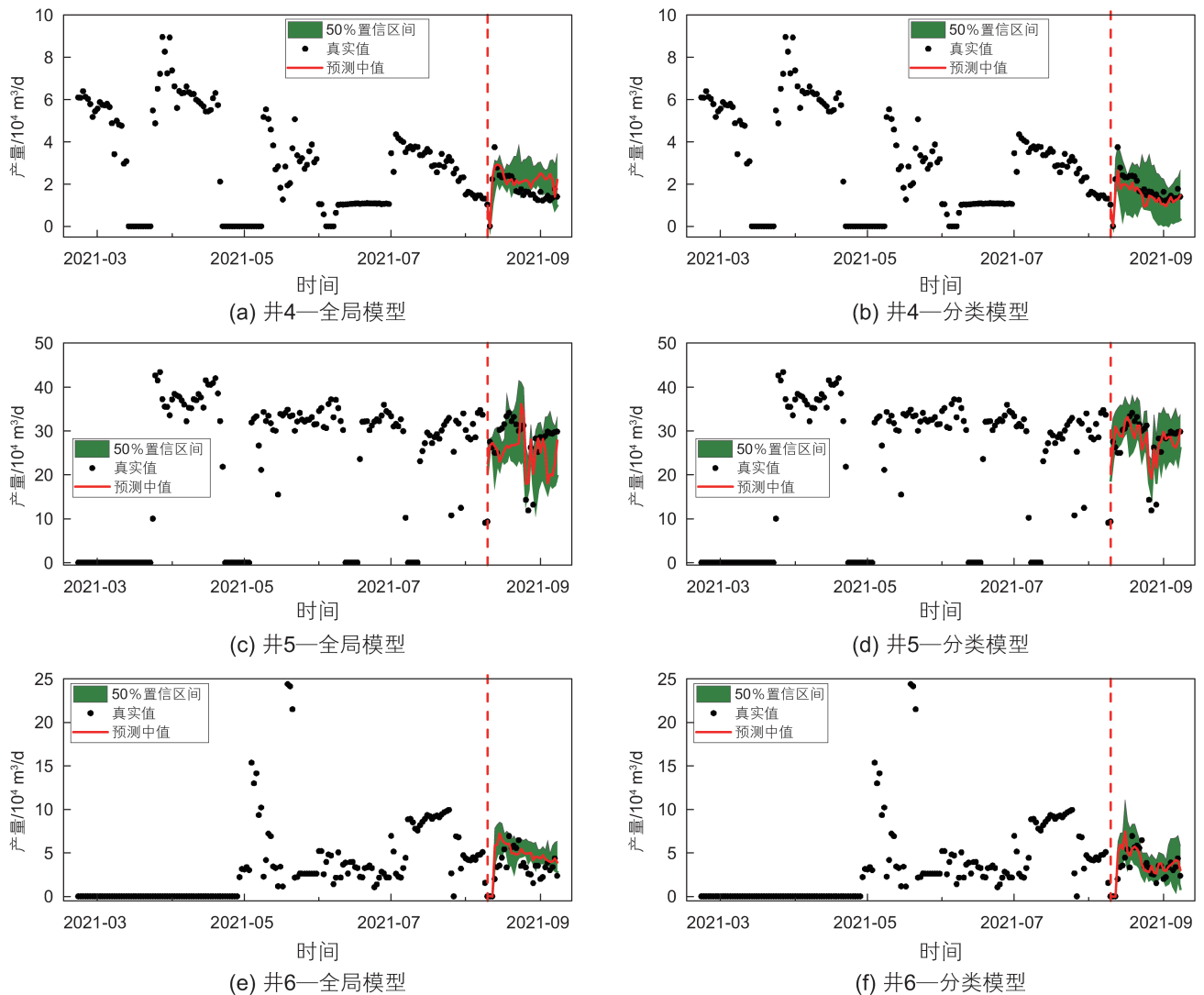


图4 全局模型和分类模型的预测精度对比

Fig. 4 Comparison of prediction accuracy between global and classification models

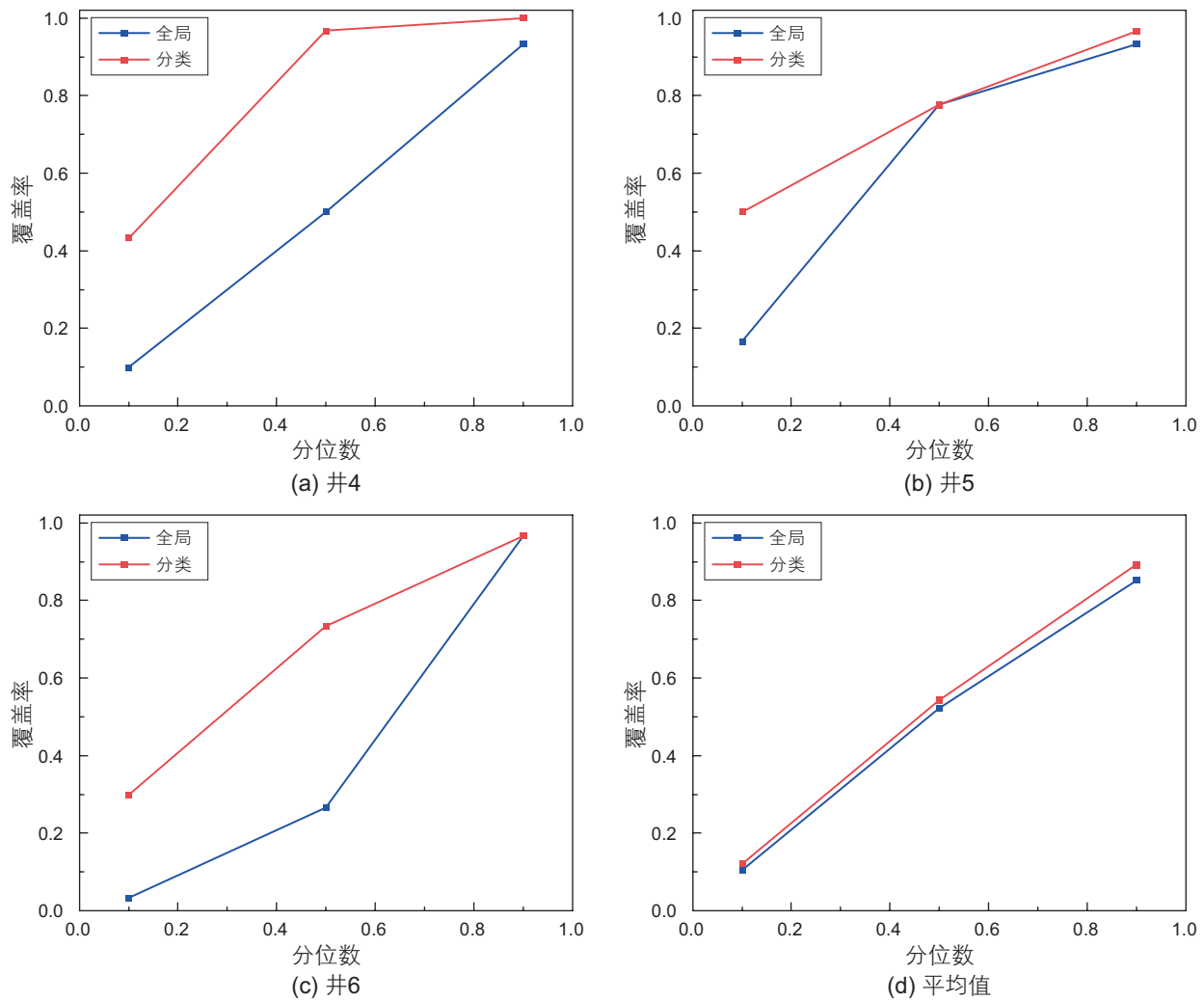


图5 全局模型和分类模型的概率区间覆盖率对比

Fig. 5 Comparison of coverage of probability intervals between global and classified models

表4 分类模型和全局模型的评价指标

Table 4 Evaluation metrics for classification and global models

井名	模型	NRMSE	ND	CPI(0.1)	CPI(0.5)	CPI(0.9)
Well-4	全局	0.3642	0.3050	0.1000	0.5000	0.9333
	分类	0.2557	0.1970	0.4333	0.9667	1.0000
Well-5	全局	0.2200	0.1840	0.1667	0.7667	0.9333
	分类	0.1490	0.1113	0.5000	0.7667	0.9667
Well-6	全局	0.5295	0.4305	0.0333	0.2667	0.9667
	分类	0.3744	0.2645	0.3000	0.7333	0.9667
所有数据集平均值	全局	0.2742	0.2145	0.1056	0.5227	0.8520
	分类	0.2221	0.1624	0.1214	0.5425	0.8927

结果可知,总体上,基于分类的深度回归神经网络模型的预测准确性上要优于全局模型,中值更接近真实值。平均意义上来说,分类模型的覆盖率CPI是高于全局模型,NRMSE指标降低了19%,ND指标降低了24%。说明分类模型能提取同一类井的产量变化,可

以在全局模型的基础上进一步降低模型的不确定性。进一步说明在相等数量的单井数据情况下,不相关井对目标井预测精度提高的不显著或反而具有相反影响。分类模型性能优于全局模型,能够灵活学习动态和静态类别特征,为“通用+特定”的产量预测模型的建

立提供了坚实基础。

3.3 区块整体产量预测

由于区块内部各气井的生产阶段不同，部分新井处于上产阶段，大多数老井处于稳产或递减阶段，传统方法无法用一个统一的模型来建模和预测，并且需要为每个单井建立一个预测模型，耗时耗力。利用深度自回归神经网络模型学习区块所有气井产量历史数据，利用贝叶斯推断得到一个历史序列—未来序列的产量概率分布，可以快速对上产新井和递减老井进行建模和预测，并且给定一定的置信区间。并且由于区块总体日产量数据是由单井数据累加得到，于是区块总体日产量概率分布可以认为与单井产量概率分布是同分布的。于是利用自动缩放算法，将总体日产量数据也作为一口单井，放到总体训练集中训练网络模

型，从而进行区块总体日产量预测。利用鄂尔多斯盆地某两个致密气区块的生产数据，其结果如图 6 和表 5 所示，相比 LSTM 模型，深度自回归神经网络模型对于区块 1 的预测结果 NRMSE 降低了 29%，ND 降低了 34%，区块 2 的 NRMSE 降低了 52%，ND 降低了 60%，并且 50% 的置信区间的覆盖率可以达到 0.6 以上，说明深度自回归神经网络模型的预测精度更高。区块 2 由于此时处于产量上升期，LSTM 模型由于是单步预测，因此其往往预测前期效果较好，后期由于滚动预测的误差累积造成误差上升。而深度自回归神经网络模型，由于是序列对序列的预测，因此其更能考虑长期和整体的预测效果。利用深度自回归神经网络对区块总体产量进行预测，可以进一步为整个气藏的未来产量预测提供支持，从而为气藏的宏观调控以及气藏、集气站和单井的配产决策提供依据和基础。

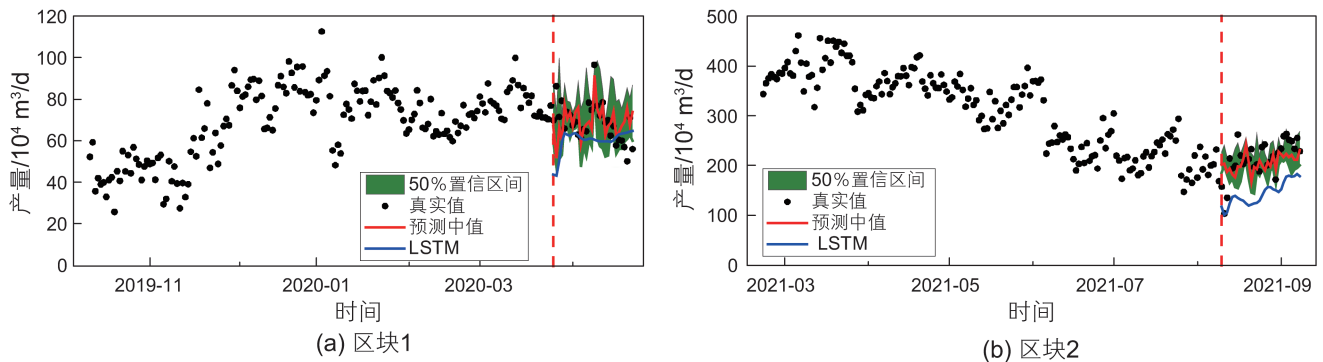


图 6 区块总体日产量预测

Fig. 6 Block overall daily production rate forecasting

表 5 区块总体日产量预测的评价指标

Table 5 Evaluation indicators for predicting overall daily production in the blocks

区块	模型	NRMSE	ND	CPI(0.1)	CPI(0.5)	CPI(0.9)
区块 1	新模型	0.1574	0.1095	0.2000	0.6667	0.9667
	LSTM	0.2204	0.1658	/	/	/
区块 2	新模型	0.1680	0.1303	0.2000	0.6333	0.9333
	LSTM	0.3494	0.3277	/	/	/

4 结论

(1) 本文提出了多井产量数据驱动的概率预测新模型，通过学习得到广义产量概率演化模式与目标井的特定产量历史数据相结合，形成了“广义+特定”的产量概率预测方法，有效突破了传统方法的限制，实现了多井多步预测。结果表明深度自回归神经网络模型的预测准确性更高，在涉及频繁开关井操作的单

井上相比于 LSTM 模型依然能保持良好稳定性和鲁棒性，6 口单井预测结果平均意义上，NRMSE 可以降低 57%，ND 可以降低 45%。

(2) 平均意义上来说，分类模型的概率预测指标覆盖率总是高于全局模型，NRMSE 指标降低了 19%，ND 指标降低了 24%。说明深度自回归神经网络模型通过提取特定精细分类井的广义产量概率演化模式，进一步降低了预测的不确定性，提高了特定精细分类井的预测精度，为“通用+特定”的产量预测模型的建立

提供了坚实基础。

(3) 基于自动放缩算法和总体区块产量与单井数据同分布假设, 将深度自回归神经网络模型应用到区块

总体产量预测上, 相比于LSTM模型避免了单步预测的误差累积效应, 两个区块的 ND 分别降低了34%和60%, 并且模型不局限于单井数量和产量具体值大小。

参考文献

- [1] 邹才能, 朱如凯, 吴松涛, 等. 常规与非常规油气聚集类型、特征、机理及展望——以中国致密油和致密气为例[J]. 石油学报, 2012, 33(2): 173–187. [ZOU C N, ZHU R K, WU S T, et al. Types, characteristics, genesis and prospects of conventional and unconventional hydrocarbon accumulations: taking tight oil and tight gas in China as an instance[J]. Acta Petrolei Sinica, 2012, 33(2): 173–187.]
- [2] 贾承造, 郑民, 张永峰. 中国非常规油气资源与勘探开发前景[J]. 石油勘探与开发, 2012, 39(2): 129–136. [JIA C Z, ZHENG M, ZHANG Y F. Unconventional hydrocarbon resources in China and the prospect of exploration and development[J]. Petroleum Exploration and Development, 2012, 39(2): 129–136.]
- [3] 孙龙德, 邹才能, 贾爱林, 等. 中国致密油气发展特征与方向[J]. 石油勘探与开发, 2019, 46(6): 1015–1026. [SUN L D, ZOU C N, JIA A L, et al. Development characteristics and orientation of tight oil and gas in China[J]. Petroleum Exploration and Development, 2019, 46(6): 1015–1026.]
- [4] 郭建春, 路千里, 刘壮, 等. “多尺度高密度”压裂技术理念与关键技术——以川西地区致密砂岩气为例[J]. 天然气工业, 2023, 43(2): 67–76. [GUO J C, LU Q L, LIU Z, et al. Concept and key technology of “multi-scale high-density” fracturing technology: A case study of tight sandstone gas reservoirs in the western Sichuan Basin[J]. Natural Gas Industry, 2023, 43(2): 67–76.]
- [5] 王永辉, 卢拥军, 李永平, 等. 非常规储层压裂改造技术进展及应用[J]. 石油学报, 2012, 33(S1): 149–158. [WANG Y H, LU Y J, LI Y P, et al. Progress and application of hydraulic fracturing technology in unconventional reservoir[J]. Acta Petrolei Sinica, 2012, 33(S1): 149–158.]
- [6] 凌云, 李宪文, 慕立俊, 等. 苏里格气田致密砂岩气藏压裂技术新进展[J]. 天然气工业, 2014, 34(11): 66–72. [LING Y, LI X W, MU L J, et al. New progress in fracturing technologies for tight sandstone gas reservoirs in the Sulige Gas Field, Ordos Basin[J]. Natural Gas Industry, 2014, 34(11): 66–72.]
- [7] 陈元千, 傅礼兵, 徐佳倩. 两类产量递减模型在预测页岩气井和致密气井中的应用与对比[J]. 油气地质与采收率, 2021, 28(3): 84–89. [CHEN Y Q, FU L B, XU J Q. Application and comparison of two production decline models in shale gas wells and tight gas wells[J]. Petroleum Geology and Recovery Efficiency, 2021, 28(3): 84–89.]
- [8] 梁倚维, 王东, 李霆钧, 等. 致密气井Arps产量递减指数变化特征[J]. 西安石油大学学报(自然科学版), 2018, 33(2): 77–81. [LIANG Y W, WANG D, LI T J, et al. Study on variation characteristic of Arps decline exponent of tight reservoir gas well[J]. Journal of Xi'an Shiyou University(Natural Science Edition), 2018, 33(2): 77–81.]
- [9] ARPS J J. Analysis of decline curves[J]. Transactions of the AIME, OnePetro, 1945, 160(01): 228–247.
- [10] 孙贺东, 欧阳伟平, 张冕, 等. 考虑裂缝变导流能力的致密气井现代产量递减分析[J]. 石油勘探与开发, 2018, 45(3): 455–463. [SUN H D, OUYANG W P, ZHANG M, et al. Advanced production decline analysis of tight gas wells with variable fracture conductivity[J]. Petroleum Exploration and Development, 2018, 45(3): 455–463.]
- [11] HE Y, CHENG S, LI S, et al. A semianalytical methodology to diagnose the locations of underperforming hydraulic fractures through pressure-transient analysis in tight gas reservoir[J]. SPE Journal, SPE, 2017, 22(3): 924–939.
- [12] CLARKSON C R R, NOBAKHT M, KAVIANI D, et al. Production Analysis of Tight-Gas and Shale-Gas Reservoirs Using the Dynamic-Slippage Concept[J]. SPE Journal, OnePetro, 2012, 17(1): 230–242.
- [13] 王军磊, 贾爱林, 何东博, 等. 致密气藏分段压裂水平井产量递减规律及影响因素[J]. 天然气地球科学, 2014, 25(2): 278–285. [WANG J L, JIA A L, HE D B, et al. Rate decline of multiple fractured horizontal well and influence factors on productivity in tight gas reservoirs[J]. Natural Gas Geoscience, 2014, 25(2): 278–285.]
- [14] 糜利栋, 顾少华, 薛亮, 等. 基于数据驱动技术的智能试井解释方法——以有水气藏产水气井为例[J]. 天然气工业, 2021, 41(2): 119–124. [MI L D, GU S H, XUE L, et al. An intelligent well test interpretation method based on data driven technology: A case study from one water-producing gas well of water-bearing gas reservoirs[J]. Natural Gas Industry, 2021, 41(2): 119–124.]
- [15] 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 L, GU S H, WANG J B, 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.]
- [16] SONG X, LIU Y, XUE L, et al. Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model[J]. Journal of Petroleum Science and Engineering, 2020, 186: 106682.

- [17] ZHA W, LIU Y, WAN Y, et al. Forecasting monthly gas field production based on the CNN-LSTM model[J]. *Energy*, 2022, 260: 124889.
- [18] XUE L, WANG J, HAN J, et al. Gas well performance prediction using deep learning jointly driven by decline curve analysis model and production data[J]. *Advances in Geo-Energy Research*, 2023, 8(3): 159–169.
- [19] 李鹏, 范倩倩, 霍明会, 等. 苏里格气田气井配产与递减率关系研究及应用[J]. *西南石油大学学报(自然科学版)*, 2020, 42(1): 126–132. [LI P, FAN Q Q, HUO M H, et al. Researches on relationship between production allocation and decline rate of gas wells in Sulige gas field[J]. *Journal of Southwest Petroleum University(Science & Technology Edition)*, 2020, 42(1): 126–132.]
- [20] SALINAS D, FLUNKERT V, GASTHAUS J, et al. DeepAR: Probabilistic forecasting with autoregressive recurrent networks[J]. *International Journal of Forecasting*, 2020, 36(3): 1181–1191.
- [21] 王继平, 张城玮, 李建阳, 等. 苏里格气田致密砂岩气藏开发认识与稳产建议[J]. *天然气工业*, 2021, 41(2): 100–110. [WANG J P, ZHANG C W, LI J Y, et al. Tight sandstone gas reservoirs in the Sulige Gas Field: Development understandings and stable-production proposals[J]. *Natural Gas Industry*, 2021, 41(2): 100–110.]

(编辑 杨雅露)