

文章编号: 1000-0747(2021)01-0175-10

DOI: 10.11698/PED.2021.01.16

基于多变量时间序列及向量自回归机器学习模型的水驱油藏产量预测方法

张瑞, 贾虎

(“油气藏地质及开发工程”国家重点实验室 西南石油大学, 成都 610500)

基金项目: 霍英东教育基金会高等院校青年教师基金(171043); 四川省杰出青年科技人才项目(2019JDJQ0036)

摘要: 提出了一种基于多变量时间序列(MTS)及向量自回归(VAR)机器学习模型的水驱油藏产量预测方法, 并进行了实例应用。该方法在井网分析的基础上通过MTS分析对注采井组数据进行优选, 并将井组内不同采出井产油量及注入井注水量作为彼此相关的时间序列, 通过建立VAR模型从多个时间序列中提取出相互作用规律, 挖掘注采井间流量的依赖关系从而进行产量预测。水驱油藏历史生产数据分析结果表明, 与数值模拟历史拟合结果相比, 机器学习模型产量预测结果具有更高精度, 同时不确定性分析提升了预测结果的安全性。通过脉冲响应分析对注入井的采油贡献量进行评价, 可为注水开发方案调整提供理论指导。图10表5参28

关键词: 水驱油藏; 产量预测; 机器学习; 多变量时间序列; 向量自回归; 不确定性分析

中图分类号: TP392

文献标识码: A

Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs

ZHANG Rui, JIA Hu

(State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation of Southwest Petroleum University, Chengdu 610500, China)

Abstract: A forecasting method of oil well production based on multivariate time series (MTS) and vector autoregressive (VAR) machine learning model for waterflooding reservoir is proposed, and an example application is carried out. This method first uses MTS analysis to optimize injection and production data on the basis of well pattern analysis. The oil production of different production wells and water injection of injection wells in the well group are regarded as mutually related time series. Then a VAR model is established to mine the linear relationship from MTS data and forecast the oil well production by model fitting. The analysis of history production data of waterflooding reservoirs shows that, compared with history matching results of numerical reservoir simulation, the production forecasting results from the machine learning model are more accurate, and uncertainty analysis can improve the safety of forecasting results. Furthermore, impulse response analysis can evaluate the oil production contribution of the injection well, which can provide theoretical guidance for adjustment of waterflooding development plan.

Key words: waterflooding reservoir; production prediction; machine learning; multivariate time series; vector autoregression; uncertainty analysis

引用: 张瑞, 贾虎. 基于多变量时间序列及向量自回归机器学习模型的水驱油藏产量预测方法[J]. 石油勘探与开发, 2021, 48(1): 175-184.

ZHANG Rui, JIA Hu. Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs[J]. Petroleum Exploration and Development, 2021, 48(1): 175-184.

0 引言

近年来, 随着人工智能技术在石油工业领域的应用与推广, 油田进入信息化、智能化时代^[1-2]。油田开发的各个环节都积累了大量的历史生产数据, 且数据形式多样、结构复杂, 传统的分析预测方法已经无法满足需求, 因此, 需要发展一种智能化生产数据分析

预测方法。

合理地引入人工智能方法, 建立适应于油田数据的机器学习模型, 是进行油田数据智能分析的关键。2004年Li等^[3]提出了基于支持向量机(SVM)技术的油藏预测方法。2005年邢明海等^[4]建立了基于模糊神经网络的油田产量预测模型。2009年Weber等^[5]通过线性及非线性回归方式表征注入井和生产井间连通

性,在此基础上建立了容抗模型(CRM),并通过该模型确定使油藏产出最大化的注水方案。2011年Anifowose^[6]回顾了人工智能技术在石油工程中的应用,并总结了人工智能技术在油藏建模及产量管理等方面的用途。2012年Gherabati等^[7]将任意注采井在任意地层内的位置作为模型中节点,以此建立神经网络模型,并通过将模拟结果与实际生产数据之间的均方差最小化完成模型拟合。2012年Salazar等^[8]通过叠加原理将衰减曲线与容抗模型相结合,对任意生产井的产量进行预测。2013年Holdaway^[9]建立了集引导、聚类和数据挖掘于一体的概率模型来进行产量预测。2013年张方舟等^[10]将BP神经网络与灰色网络的优点相结合,提出了基于改进灰色网络的产量预测模型。2014年Gupta等^[11]提出了基于数据挖掘技术和时间序列分析的产量预测方法。2014年Zhou等^[12]采用多变量回归分析、聚类分析以及主成分分析等数据挖掘技术来识别生产井的产能与地质工程参数的相关性。2016年Li等^[13]以产液剖面及注入剖面数据为目标变量,以渗透率、储集层厚度、注采井距、注入速度和产液量作为影响因素,建立了以支持向量机为核心的多层水驱油藏注水分配模型。2016年Cao等^[14]利用地质模型、生产历史数据和生产制度等信息,采用机器学习算法对非常规油气储集层中老井和新井的生产现状进行了预测。2016年赵辉等^[15]先将油藏分层离散成一系列由井间传导率和连通体积等参数表征的井间连通单元,再建立针对水驱开发油藏的多层注采井间连通性反演模型,最后给出了相应的模型参数反演方法。2017年Maqui等^[16]通过将传统数值模拟方法与纯数据预测方法相结合,将流场转化为适合于最优化方法的注采井网络连接图。2017年武志军^[17]基于Spark建立油藏数据处理与分析方法,利用数据挖掘技术对油藏数据进行挖掘,分析油田产量数据与油藏参数之间的关系。2017年Martin等^[18]提出了一种采用两步机器学习方案进行产量预测的自动化数据驱动方法。2018年贾虎等^[19]基于流线聚类人工智能的方法,开展流场精细化描述,找出潜在优势流场。2019年Khan等^[20]采用人工神经模糊推理系统(ANFIS)和SVM算法建立了一种原油流量估算经验模型,预测人工气举井的产油率。2019年Noshi等^[21]探讨了梯度增强树(GBT)、Adaboost和支持向量回归(SVR)3种机器学习算法在产量预测方面的潜在应用。

可以看出,前人主要通过建立预测模型来解决油田生产数据分析预测的问题。然而,井数众多的大型

油田往往生产历史较长,井史比较复杂,并且不同井组的流场在不同时刻差异极大,因此有必要建立可对复杂多过程时间序列进行精确预测的模型,以准确表征井组内采出井产油量与注入井注水量之间的关系。本文提出一种基于多变量时间序列(MTS)及向量自回归(VAR)机器学习模型的水驱油藏产量预测方法,在井网分析的基础上通过MTS分析对注采井组数据进行优选,并将井组内不同采出井产油量及注入井注水量作为彼此相关的时间序列,建立VAR模型,从多个时间序列中提取出相互作用规律,挖掘注采井间流量的依赖关系,通过模型拟合来预测生产井产量,并采用不确定性分析及脉冲响应分析方法进行评价。

1 理论基础

1.1 多变量时间序列分析

时间序列是按发生时间顺序排列的具有相同统计指标的一系列值,如果某一变量可以进行时间序列观测,并且过去的的数据包含该变量未来变化的信息,则可以使用过去观测数据的某个函数对该变量的未来值进行预测^[22]。水驱油藏历史生产数据中就存在时间序列观测数据,如采出井产油量,并且过去的产油量数据包含了未来产油量的变化信息,因此可以利用产油量历史数据去预测采出井未来产油量。在处理历史生产数据变量时,通常一个变量不仅依赖于该变量的过去值,而且还与其他变量在时间上相关。例如,水驱油藏中某一口采出井的产油量变化规律可能还依赖于多口注入井注水量的变化。针对井数众多的水驱油藏的历史生产数据,单变量时间序列分析难以挖掘多口注采井之间的数据关系,因此需要进行多变量时间序列分析。

MTS分析的关键是相关性分析。相关性分析是指对两个或多个变量之间的相互依存关系进行分析,从而衡量变量之间的相关程度及相关方向,挖掘变量之间的内在关系^[23]。相关系数是衡量两个变量之间相关程度的指标,其中Pearson相关系数用来衡量两个变量之间的线性相关程度^[24]。设有 n 个关于 a 和 b 两个变量的数据对 (a_q, b_q) , $q=1, 2, \dots, n$,则Pearson相关系数的计算公式为:

$$r = \frac{\sum_{q=1}^n \left(a_q - \frac{1}{n} \sum_{q=1}^n a_q \right) \left(b_q - \frac{1}{n} \sum_{q=1}^n b_q \right)}{\sqrt{\sum_{q=1}^n \left(a_q - \frac{1}{n} \sum_{q=1}^n a_q \right)^2} \sqrt{\sum_{q=1}^n \left(b_q - \frac{1}{n} \sum_{q=1}^n b_q \right)^2}} \quad (1)$$

(1) 式中 r 的取值范围为 $[-1, 1]$; r 的符号表示变

量之间的相关方向, $r > 0$ 表示线性正相关, $r < 0$ 表示线性负相关; $|r|$ 的大小可用来定量分析变量之间的线性相关程度。Pearson 线性相关程度评价标准为: $|r|=0$ 表示线性无关, $0 < |r| < 0.3$ 表示低度线性相关, $0.3 \leq |r| < 0.8$ 表示中度线性相关, $0.8 \leq |r| < 1.0$ 表示高度线性相关, $|r|=1.0$ 表示完全线性相关。

对于水驱油藏注采井流量而言, 可以将不同采出井产油量与注入井注水量作为多变量时间序列, 并利用 Pearson 相关系数进行相关性分析。依据 (1) 式可以推导出针对多口注采井流量的 MTS 相关性分析公式:

$$R_{ij} = \frac{\sum_{t=1}^N \left(y_{i,t} - \frac{1}{N} \sum_{t=1}^N y_{i,t} \right) \left(e_{j,t} - \frac{1}{N} \sum_{t=1}^N e_{j,t} \right)}{\sqrt{\sum_{t=1}^N \left(y_{i,t} - \frac{1}{N} \sum_{t=1}^N y_{i,t} \right)^2} \sqrt{\sum_{t=1}^N \left(e_{j,t} - \frac{1}{N} \sum_{t=1}^N e_{j,t} \right)^2}} \quad (2)$$

为了增强机器学习模型训练过程的稳定性以及产量预测结果的合理性, 可以利用 MTS 相关性分析结果, 对水驱油藏注采井流量数据进行优选。依据 Pearson 线性相关程度评价标准, 并结合 (2) 式计算的 Pearson 相关系数, 可以得到如下注采井流量数据优选公式:

$$\begin{cases} M_i = \max(|R_{i1}|, |R_{i2}|, \dots, |R_{ik_i}|) \\ M_j = \max(|R_{1j}|, |R_{2j}|, \dots, |R_{k_p j}|) \\ I = \{ \alpha | M_\alpha \geq 0.3, M_\alpha \in \{M_1, M_2, \dots, M_{k_i}\} \} \\ P = \{ \beta | M_\beta \geq 0.3, M_\beta \in \{M_1, M_2, \dots, M_{k_p}\} \} \end{cases} \quad (3)$$

(3) 式所代表的数据优选过程可以确保某一口采出井 (注入井) 流量至少与一口注入井 (采出井) 流量之间存在中度以上的线性相关关系, 有利于机器学习模型挖掘注采井间流量的依赖关系。

1.2 向量自回归模型

机器学习模型按照可使用的数据类型分为两大类: 监督学习模型和无监督学习模型, 其中监督学习模型主要包括用于分类和用于回归的模型。向量自回归模型是一种用于回归的机器学习模型, 该模型主要用来处理多变量时间序列数据, 通过 VAR 模型可以挖掘多个时间序列数据之间的相互线性依赖关系。VAR 模型是自回归模型与方程联立的结合与延伸, 是一种非结构化的多方程回归模型, 模型中任意变量均可通过其自身的滞后值、其他相关变量、常数项及误差项组成的方程表示, 并可以通过迭代对其未来值进行预测^[25]。

注采井间流量的 VAR 模型并不以油藏工程理论为

基础, 而是基于注采井流量数据的统计性质来建立。首先将不同采出井产油量与注入井注水量作为彼此相关的时间序列, 并构造如下向量:

$$\begin{cases} \mathbf{Y}_t = [y_{1,t}, y_{2,t}, \dots, y_{i,t}, \dots, y_{k_p,t}]^T \\ \mathbf{E}_t = [e_{1,t}, e_{2,t}, \dots, e_{j,t}, \dots, e_{k_i,t}]^T \end{cases} \quad (4)$$

然后通过注采井间流量的相互关系可以建立如下 VAR 模型:

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{Y}_{t-1} + \mathbf{A}_2 \mathbf{Y}_{t-2} + \dots + \mathbf{A}_p \mathbf{Y}_{t-p} + \mathbf{B}_0 \mathbf{E}_t + \mathbf{u}_t \quad (5)$$

预测模型以当月采出井产油量 $\mathbf{Y}_t$ 、以往 $p-1$ 个月的采出井产油量 ($\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \dots, \mathbf{Y}_{t-p+1}$) 以及未来 1 个月注入井的注水量 $\mathbf{E}_{t+1}$ 作为输入, 然后预测未来 1 个月的采出井产油量 $\hat{\mathbf{Y}}_{t+1}$, 最后通过迭代对未来采出井的产油量进行预测, 迭代公式如下:

$$\hat{\mathbf{Y}}_{t+1} = \hat{\mathbf{c}} + \hat{\mathbf{A}}_1 \mathbf{Y}_t + \hat{\mathbf{A}}_2 \mathbf{Y}_{t-1} + \dots + \hat{\mathbf{A}}_p \mathbf{Y}_{t-p+1} + \hat{\mathbf{B}}_0 \mathbf{E}_{t+1} \quad (6)$$

(6) 式中 $\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2, \dots, \hat{\mathbf{A}}_p, \hat{\mathbf{B}}_0$ 为估计参数矩阵, 可以通过联立线性方程组进行求解, 如 (7) 式所示。

$$\mathbf{G} = \mathbf{DZ} + \mathbf{U} \quad (7)$$

其中

$$\mathbf{G} = [\mathbf{Y}_p, \mathbf{Y}_{p+1}, \dots, \mathbf{Y}_m] = \begin{bmatrix} y_{1,p} & y_{1,p+1} & \dots & y_{1,m} \\ y_{2,p} & y_{2,p+1} & \dots & y_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k_p,p} & y_{k_p,p+1} & \dots & y_{k_p,m} \end{bmatrix}$$

$$\mathbf{D} = [\mathbf{c}, \mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p, \mathbf{B}_0]$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \mathbf{Y}_{p-1} & \mathbf{Y}_p & \dots & \mathbf{Y}_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}_0 & \mathbf{Y}_1 & \dots & \mathbf{Y}_{m-p} \\ \mathbf{E}_p & \mathbf{E}_{p+1} & \dots & \mathbf{E}_m \end{bmatrix}$$

$$\mathbf{U} = [\mathbf{u}_p, \mathbf{u}_{p+1}, \dots, \mathbf{u}_m]$$

最后通过最小二乘法进行求解, 总参数矩阵 $\mathbf{D}$ 的最小二乘解 $\hat{\mathbf{D}}$ 可以表示成如下形式:

$$\hat{\mathbf{D}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{G} \quad (8)$$

对于 (5) 式给出的 VAR 模型, 其中重要的参数为滞后阶数 p , 其代表了假设最多以往 p 个月的采出井产油量 ($\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \dots, \mathbf{Y}_{t-p}$) 将影响当月采出井产油量 $\mathbf{Y}_t$ 。滞后阶数的选择将直接影响到 VAR 模型的稳定性和准确性, 滞后阶数过大会导致估计参数增多, 影响模型参数估计的有效性; 滞后阶数过小会导致误差项自相关, 影响模型参数估计的一致性。滞后阶数一

般依据信息准则来确定。首先利用 (9) 式计算 VAR 模型在滞后阶数为 p 时的似然估计值 L , 然后结合似然估计值及模型参数, 可得到如 (10) 式所示的 3 个信息准则^[26-28]。最后选取使 3 种信息准则评价价值均较小的滞后阶数作为最佳的滞后阶数。

$$L = \lg \left| \left(\mathbf{Y}_t - \hat{\mathbf{Y}}_t \right)^T \left(\mathbf{Y}_t - \hat{\mathbf{Y}}_t \right) / m \right| \quad (9)$$

$$\begin{cases} Q_{aic} = -2 \ln L + 2k \\ Q_{bic} = -2 \ln L + k \ln m \\ Q_{hqic} = -2 \ln L + k \ln (\ln m) \end{cases} \quad (10)$$

尽管滞后阶数分析可以保证 VAR 模型的稳定性和准确性, 但当给定的数据量较小时, 预测结果仍具有一定误差, 而不确定性分析可估计以往数据导致的模型误差大小, 并给出模型对于预测结果的把握性大小, 从而进行更为安全的预测, 即在预测不确定性较大时谨慎决策。因此, 对模型预测结果进行不确定性分析十分必要。对预测结果不确定性大小的推导主要基于将原本的时间序列过程考虑为随机过程^[25], 其简洁的矩阵表示如下:

$$\bar{\mathbf{u}}_t = \left(\mathbf{Y}_t - \hat{\mathbf{Y}}_t \right) \left(\mathbf{Y}_t - \hat{\mathbf{Y}}_t \right)^T / m \quad (11)$$

同时考虑到参数的持续影响, 未来某一时刻的累计影响矩阵与估计预测误差存在如下关系:

$$\mathbf{y}(h) = \sum_{t=1}^h \Phi_t \bar{\mathbf{u}}_t \Phi_t^T = \begin{bmatrix} [\sigma_1(h)]^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & [\sigma_{k_p}(h)]^2 \end{bmatrix} \quad (12)$$

其中

$$\Phi_t = \begin{cases} \mathbf{I}_{k_p} & t=1 \\ \sum_{s=1}^{\min(p,t)} \Phi_{t-s} \hat{\mathbf{A}}_s & t>1 \end{cases}$$

VAR 模型不仅可以进行不确定性分析, 也可以通过脉冲响应分析对系统进行评价^[25], 即在模型拟合完成后可通过脉冲响应分析对注入井进行评价, 评价依据是将任意注入井日注水量增加固定数值 (通常为 1 m^3), 观察其对整个模型系统影响, 注入井的脉冲响应分析公式可以写成如下形式:

$$\mathbf{Z}_t = \begin{cases} \mathbf{I}_{k_i} \hat{\mathbf{B}}_0 & t=1 \\ \sum_{s=1}^{\min(p,t)} \mathbf{A}_s \mathbf{Z}_{t-s} & t>1 \end{cases} \quad (13)$$

2 模型框架

利用 MTS 及 VAR 机器学习模型可以对水驱油藏

采出井的产油量进行预测, 其模型框架如图 1 所示, 主要包括以下 5 个模块。

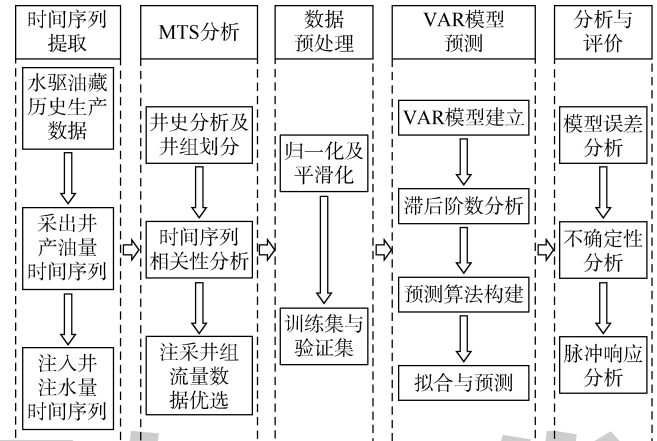


图 1 MTS 及 VAR 机器学习模型框架图

①时间序列提取模块。将水驱油藏历史生产数据作为模型的输入, 并根据产油量与注水量数据标签以及时间索引提取采出井产油量及注入井注水量时间序列。

②MTS 分析模块。在井史分析的基础上进行井组划分, 从所有注采井时间序列中截取井组开发时间段内的时间序列, 再利用 (2) 式进行 MTS 相关性分析, 并应用 (3) 式对注采井组流量数据进行优选。

③数据预处理模块。对优选后的注采井流量数据进行归一化及平滑化处理, 增加数据的平稳程度从而降低拟合难度, 并将数据划分为训练集与验证集。

④VAR 模型预测模块。该模块为模型的核心模块, 大致可分为 4 个步骤: 首先, 按照 (5) 式建立 VAR 模型; 其次, 根据 (10) 式给出的信息准则进行滞后阶数分析并确定最佳滞后阶数; 然后, 按照 (6) 式构建采出井产油量预测算法; 最后, 利用训练集数据进行 VAR 模型拟合并预测验证集数据。

⑤分析与评价模块。首先进行 VAR 模型误差分析, 以验证集预测数据与验证集实际数据之间的误差作为评价标准, 确保预测结果的准确性, 然后利用 (12) 式进行不确定性分析, 提升预测结果的安全性, 最后利用 (13) 式进行脉冲响应分析, 进而对注入井的采油贡献量进行评价。

3 实例应用

以国内大港油田某区块的断块油藏作为研究对象, 断裂主控区域内共有 105 口采出井和 66 口注入井, 该区域断层发育, 储集层非均质性较强, 建立的油藏

地质模型共有 54 个小层，由于油藏地质模型复杂，油藏数值模拟方法收敛性较差，历史拟合比较困难，精度难以保证。本文利用该断块油藏的历史生产数据，根据月度日产油量与日注水量数据标签以及时间索引，提取采出井产油量及注入井注水量时间序列，一共提取了 171 组数据，每组数据共 573 个时间样点，

时间范围从 1970 年 7 月到 2018 年 3 月，图 2 给出了采出井西 42-7-3 井和注入井西 43-67 井两口井的历史生产数据。下面在井组分析的基础上，利用 MTS 及 VAR 机器学习模型对该水驱油藏的注采井组流量时间序列进行处理，并对采出井的产油量进行预测与分析，同时对注入井的采油贡献量进行评价。

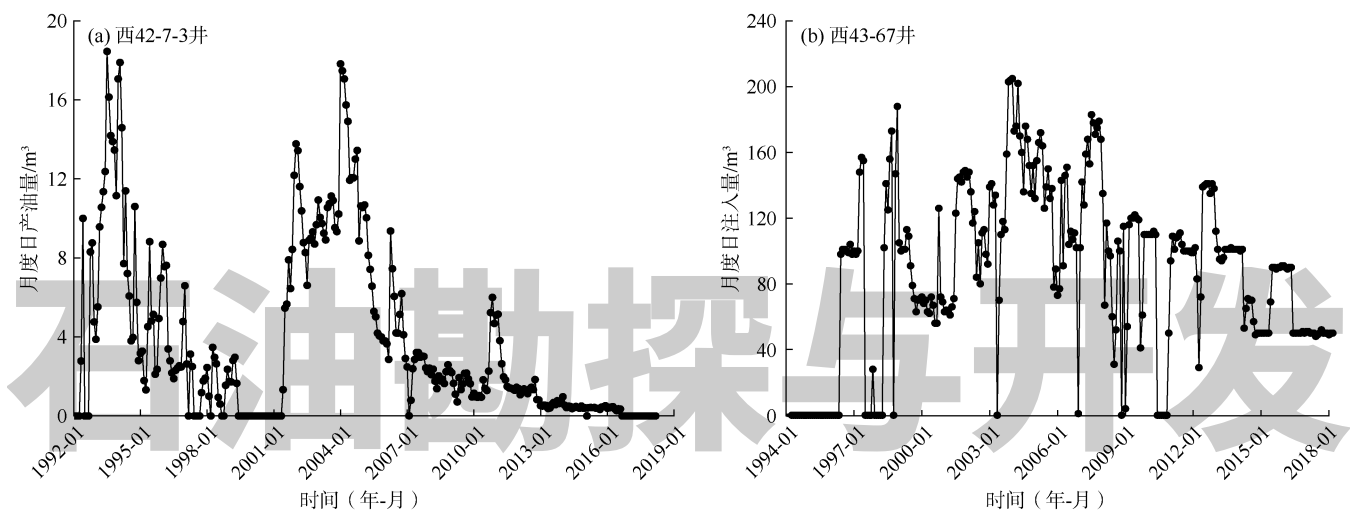


图 2 采出井西 42-7-3 井与注入井西 43-67 井的历史生产数据

3.1 MTS 相关性分析

在井史分析的基础上进行井组划分，为了降低开发调整与措施对生产数据的影响，同时保证预测结果的可靠性，主要考虑了以下井史：①加密，井组内不增加新井；②调层，主力开发层不调整；③转注，井组开发时间段内所有井不存在转注；④关井，允许短暂的关井。表 1 给出了井史分析基础上井组划分结果，共选出了 10 个井组，每个井组的注采井数、主力开发层、开发时间段不尽相同。

表 1 井史分析基础上井组划分结果					
井组 编号	注入 井数	采出 井数	主力开采层 编号	开发时间段 (年-月—年-月)	数据量
1	2	4	27	1975-02—1987-03	146
2	2	5	27	1976-12—1982-09	70
3	3	5	27	1992-01—1998-03	75
4	2	4	33	1992-07—1997-06	60
5	4	6	27	1976-01—1992-08	200
6	3	5	25	2004-09—2018-03	163
7	3	6	27	1998-06—2009-04	131
8	2	4	37	2010-11—2017-02	76
9	3	7	33	2006-04—2018-03	144
10	3	5	27	2003-08—2018-03	176

利用 (2) 式对提取出的 10 个井组内采出井产油量时间序列与注入井注水量时间序列进行相关性分析，得到注采井组流量时间序列相关图，其中 4 个井

组的时间序列相关图如图 3 所示。结合 Pearson 线性相关程度评价标准，从图 3 中可以看出，6 号井组注采井间流量相关性整体较强，4 号和 10 号井组的注采井间流量相关性整体较差。图 4 为 6 号井组中相关性最高的一对注采井的流量时间序列图，其相关系数为 0.86，属于高度线性相关，从时间序列图中可以看出注采井流量表现出同增同减的情况。对于 4 号井组和 10 号井组，由于两个井组处于断裂带附近，地层中的流体流动规律复杂，因而井组内的注采井间流量相关性整体较差。

利用 (3) 式对注采井组数据进行优选，保留相关性较高对应注采井组的数据，表 2 给出了注采井组数据优选后的结果。用 $|R_{ij}|$ 的平均值来表征不同井组的相关性程度，可以发现不同井组的相关性存在一定差异。另外，与表 1 中数据对比可知，优选后部分相关性不高的井组将不再进行产量预测分析，如 4 号井组和 10 号井组，同时优选后一些井组中部分相关性不高的注采井将不再被考虑，如图 3 所示 1 号井组中的采出井西 47-4-2。经注采井组数据优选后便可有针对性地对注采井网系统建立 VAR 模型。

3.2 VAR 模型拟合与预测

为了保证模型拟合与预测结果的稳定性，对优选后的注采井组流量数据进行归一化及平滑化处理，增加数据的平稳程度，并选取井组开发时间段最后 10 个

月的数据作为验证集，剩下的数据作为训练集。在对数据进行预处理后，对 8 个井组进行 VAR 模型拟合与预测，重点讨论整体相关性最高的 6 号井组结果，图 5 给出了 6 号井组的井位图。进行模型拟合之前，需要先进行滞后阶数分析来确定最佳滞后阶数。利用 (9)

式和 (10) 式计算了滞后阶数取不同值时对应的 AIC、BIC 以及 HQIC 信息准则评价值。表 3 给出了 6 号井组的 VAR 模型滞后阶数分析结果，可以看出，3 种信息准则评价值最小时对应的滞后阶数均为 2，因此可以选择 2 作为最佳滞后阶数。

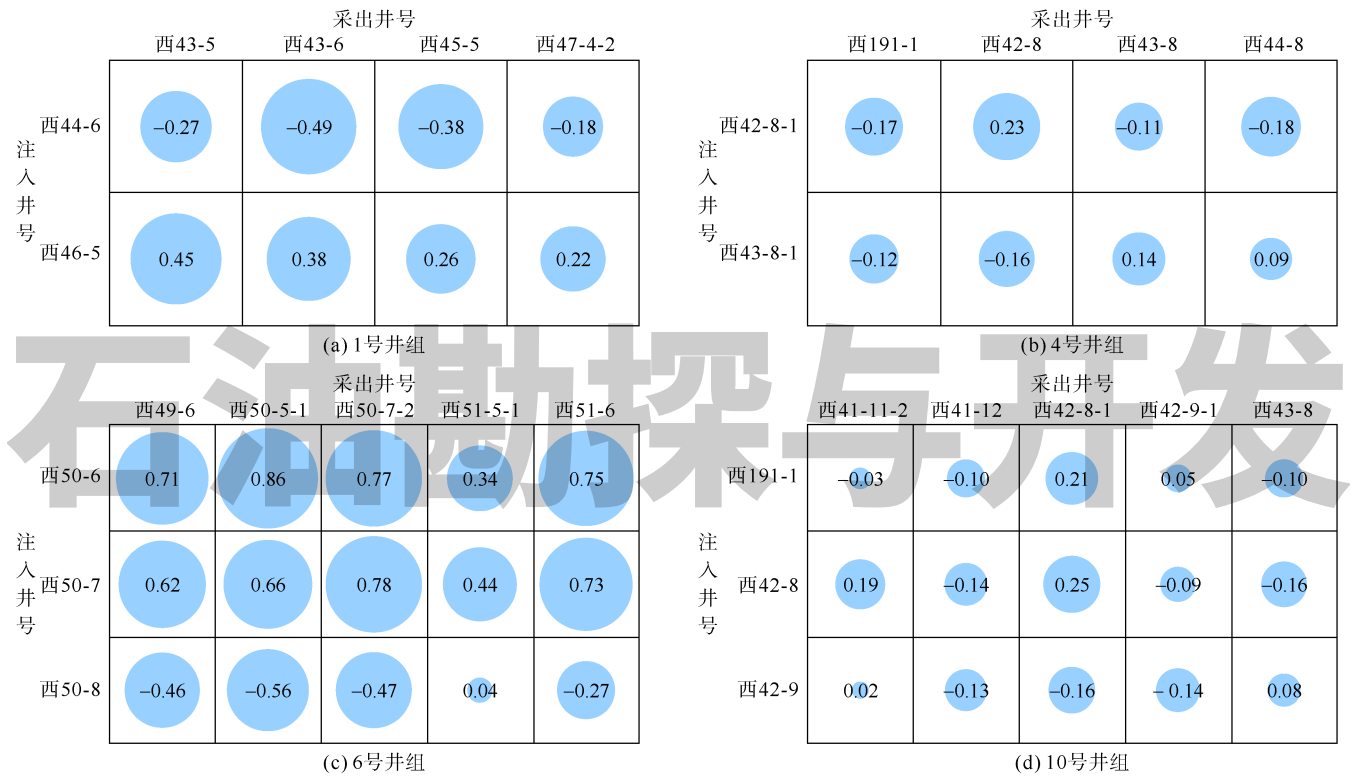


图 3 注采井组流量时间序列相关图（图中蓝色圆形中的数值为相关系数，圆形的大小代表相关系数绝对值的大小）

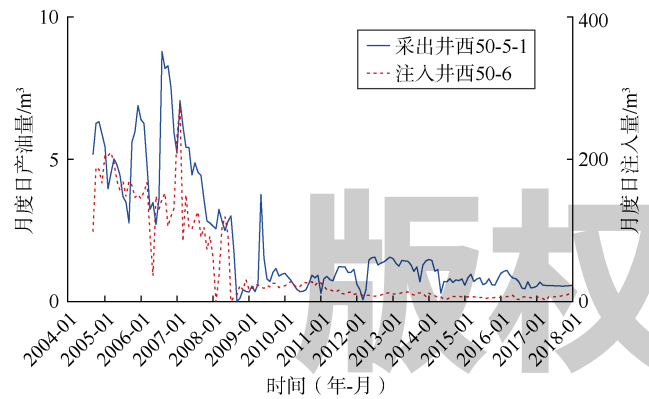


图 4 6 号井组中相关性最高的一对注采井的流量时间序列图

最佳滞后阶数确定后，即可开始模型拟合与预测。图 6 展示了 6 号井组中西 50-5-1 井和西 50-7-2 井两口采出井验证集数据的预测结果，可以看到预测日产量与实际日产量基本吻合，两口井每个产量数据点相对误差的平均值分别为 0.036 49 和 0.012 60，预测数据与实际数据之间的偏差较小，说明了预测结果的准确性。

分析 6 号井组中所有采出井验证集数据的预测情

表 2 优选后的注采井组数据			
井组编号	注入井数	采出井数	$ R_{ij} $ 的平均值
1	2	3	0.329
2	2	5	0.495
3	2	5	0.527
5	3	5	0.421
6	3	5	0.564
7	2	4	0.481
8	2	4	0.548
9	3	6	0.355

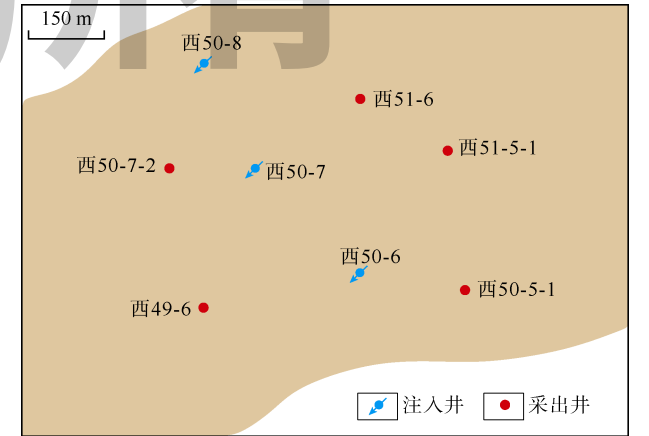


图 5 6 号井组的井位图

表 3 VAR 模型滞后阶数分析结果

滞后阶数	AIC 信息准则评价	BIC 信息准则评价	HQIC 信息准则评价
0	-14.75	-14.34	-14.58
1	-27.15	-26.22	-26.77
2	-29.14	-27.69	-28.55
3	-29.05	-27.09	-28.25
4	-28.96	-26.49	-27.96
5	-29.13	-26.14	-27.91

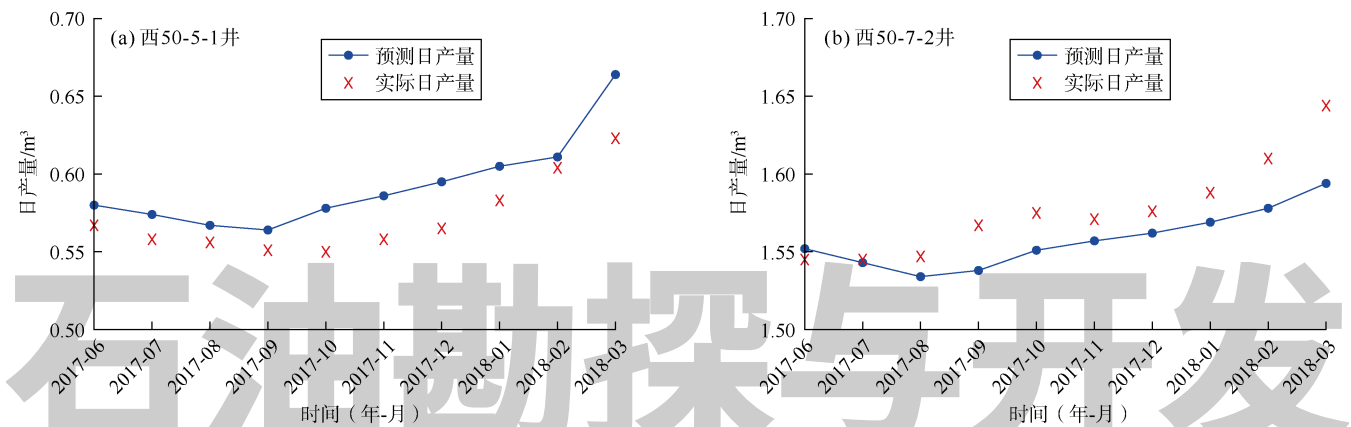


图 6 6 号井组中采出井西 50-5-1 井和西 50-7-2 井验证集数据的预测结果

况，并与数值模拟历史拟合结果进行对比分析，结果如表 4 所示。可以看出，除采出井西 51-5-1 井的机器学习预测结果较实际结果相对误差较大，其他采出井预测结果的精度都比较高。由图 3 中 6 号井组的时间序列相关图分析可知，采出井西 51-5-1 井与注入井之间的相关性不高，所以对机器学习模型的预测精度有一定影响。对误差较大的采出井西 51-5-1 井进行不确定性分析，结果如图 7 所示，发现虽然该井的预测误差较大，但其预测不确定性范围基本包含了实际数据，确保了预测结果的安全性。

表 4 6 号井组中所有采出井产量预测结果

井号	实际平均日产量/m ³	机器学习模型预测平均日产量/m ³	历史拟合平均日产量/m ³	机器学习相对误差	历史拟合相对误差
西 49-6	2.75	2.58	3.29	0.066	0.196
西 50-5-1	0.59	0.57	1.37	0.035	1.404
西 50-7-2	1.55	1.57	1.43	0.013	0.077
西 51-5-1	2.79	2.48	3.38	0.125	0.211
西 51-6	3.49	3.25	3.98	0.074	0.140

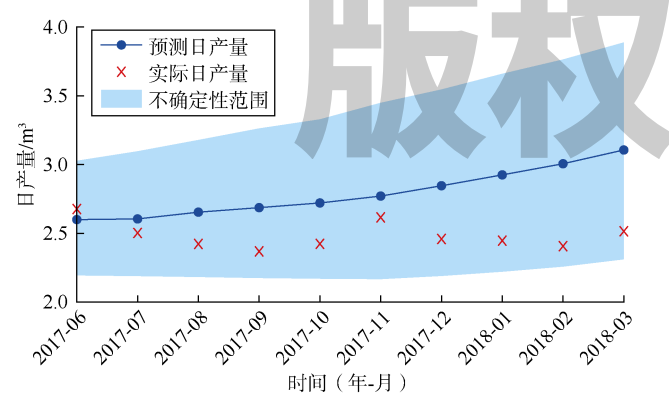


图 7 6 号井组中采出井西 51-5-1 井预测结果不确定性分析

对比表 4 中机器学习和历史拟合结果可以发现，历史拟合整体相对误差较大，其中采出井西 50-5-1 井

的历史拟合相对误差最大，结合该井的历史拟合曲线（见图 8）可以发现，由于油藏模型复杂，开发中期历史拟合精度较高，但开发早期和后期历史拟合较困难，特别是后期历史拟合结果产量有所上升，与实际产量变化有较大差异，因此相对误差较大，相比而言，该井的机器学习结果相对误差非常小，结合该井的机器学习曲线（见图 8）可以发现，尽管在历史拟合较困难的开发早期阶段产量数据波动较大，但数据的预处理增加了数据的平稳程度，保证了 VAR 模型拟合与预测的稳定性；在开发的中后期，产量数据比较稳定，VAR 模型能很好地拟合及预测产量的变化趋势，预测效果更好。

机器学习模型的适应性主要取决于数据的数量和质量两个方面。表 5 给出了 8 个井组所有采出井累计

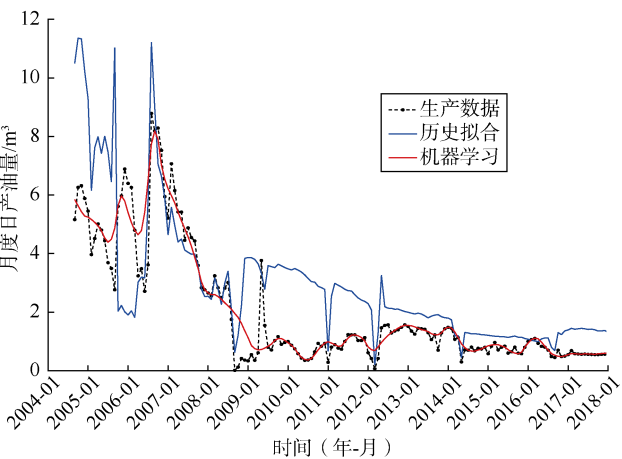


图8 6号井组中采出井西50-5-1井产量数据拟合结果

产量预测结果,可以发现3、6、8号井组的相对误差较小,1、9号井组的相对误差较大,结合表1中数据量分析可以发现,数据量的多少对VAR模型的拟合预测精度影响并不显著;相比而言,数据的质量是保证VAR模型可靠性和可用性的关键,结合表2中 $|R_{ij}|$ 的平均值分析发现,总体而言 $|R_{ij}|$ 的平均值越大,井组的机器学习模型预测精度越高,说明该参数可以用来评价不同井组机器学习模型的适应性。同时对比1、9号井组 $|R_{ij}|$ 的平均值发现,虽然9号井组 $|R_{ij}|$ 的平均值要比1号井组大,但9号井组的相对误差较大,同时从表2中发现,9号井组的注采井数明显多于1号井组,说明注采井数也会对井组累计产量预测结果产生一定影响。对比表5中机器学习和历史拟合结果可以发现,机器学习模型预测结果整体相对误差比历史拟合要小,说明VAR模型对井组累计产量的预测效果更好。

3.3 注入井脉冲响应分析

通过脉冲响应分析可以对注入井的采油贡献量进行评价,具体方法是通过假设某注入井当月平均日注

表5 各井组所有采出井累计产量预测结果

井组 编号	实际累计 产量/ m^3	机器学习 模型预测 累计产量/ m^3	历史拟合 累计产量/ m^3	机器学习 相对误差	历史拟合 相对误差
1	265.1	216.7	209.9	0.183	0.208
2	201.4	232.2	236.7	0.153	0.175
3	173.3	194.1	200.6	0.120	0.158
5	155.6	181.0	185.3	0.163	0.191
6	117.0	104.5	134.5	0.107	0.150
7	113.6	129.8	94.7	0.143	0.166
8	118.9	132.5	101.5	0.114	0.146
9	89.6	71.3	109.8	0.204	0.225

入量增加 1 m^3 ,观察其对其他采出井影响,计算公式如(13)式所示。图9展示了6号井组中西50-6井和西50-7井两口注入井对全体采出井采油量影响,例如,影响为 0.05 m^3 代表使全体采出井日采油量共增加 0.05 m^3 ,影响为 -0.01 m^3 代表使全体采出井日采油量共减少 0.01 m^3 ,以此类推。可见影响量随着时间的延续在逐渐减小。图10给出了这两口注入井随时间延续的累计影响,其结果代表一种长期效应,可以看出两口井的累计影响截然相反,说明两口注入井在短期内的驱油潜能存在明显差异。因此可以将2018年3月不同注入井对全体采出井采油量的累计影响作为评价指标来评价注入井的开发潜能,西50-6井、西50-7井、西50-8井这3口注入井的累计影响分别为 0.916 , -0.156 , -0.055 m^3 ,可见注入井西50-6井的累计影响较大,说明具有一定的驱油潜能,而西50-7井和西50-8井这两口井的累计影响为负值,说明对所有井的产量产生了一定的负面影响,其中西50-8井累计影响虽然也为负值,但其绝对值很小,说明该井对所有井产量变化的负面影响较小。在实际生产中,方案决策者可以根据注入井脉冲响应分析结果对井组进行注水优化调整,提高水驱油藏的产量。

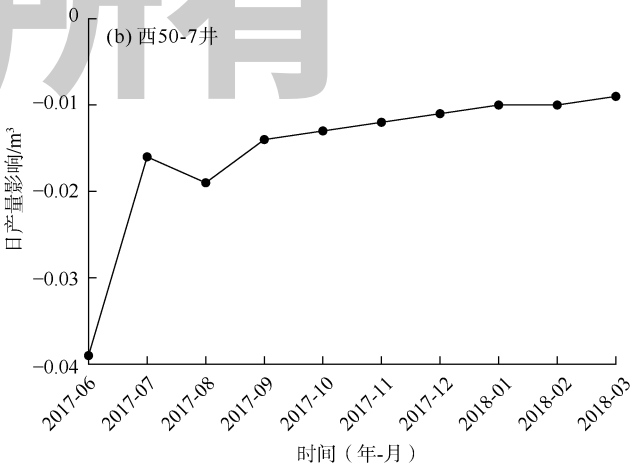
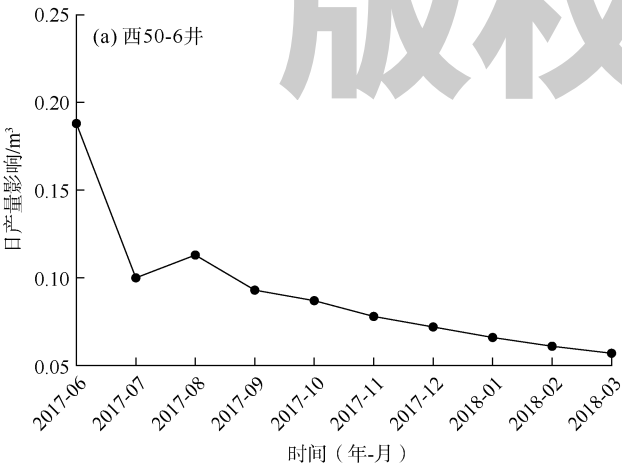


图9 6号井组中注入井西50-6井和西50-7井对全体采出井采油量影响

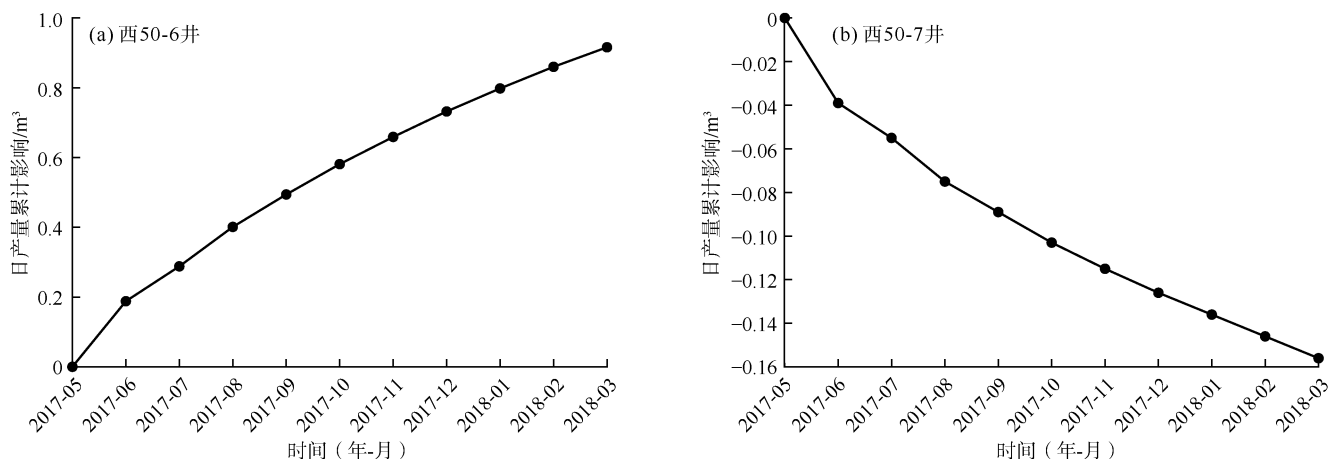


图10 6号井组中注入井西50-6井和西50-7井对全体采出井采油量累计影响

4 结语

本文提出了一种基于多变量时间序列及向量自回归机器学习模型的水驱油藏产量预测方法,该方法以井网分析为基础,利用机器学习模型从多个时间序列中提取出相互作用规律,挖掘注采井间流量的依赖关系从而进行产量预测,同时采用脉冲响应分析方法对注入井注水效果进行评价,为进一步注水开发调整提供理论指导。本文方法主观性较少,无需依靠专家经验即可进行评价,且可进行不确定性分析,确保预测结果的安全性与准确性。需要注意的是,对于部分井组,由于地下流体流动规律复杂,机器学习方法并不完全适用,同时部分相关性不是很高的井组的预测结果精度欠佳,因此有必要将其与油藏工程方法相结合来进行综合评价。

符号注释:

a, b ——变量; $A_1, A_2, \dots, A_p$ ——第1阶,第2阶, ..., 第 p 阶采出井参数矩阵,用于描述不同采出井之间的流量关系; $\hat{A}_1, \hat{A}_2, \dots, \hat{A}_p, \hat{B}_0$ —— $A_1, A_2, \dots, A_p, B_0$ 的估计值; B_0 ——注入井参数矩阵,用于描述注采井之间的流量关系; c ——偏差因子向量; $\hat{c}$ —— c 的估计值; $e_{j,t}$ ——第 j 口注入井 t 时刻的注水量, m^3/d ; E_t —— t 时刻的注入井注水量向量, m^3/d ; i, j ——采出井和注入井序号; I, P ——优选后的注入井和采出井序号集合; I_{k_i} ——注入井对角单位矩阵; I_{k_p} ——采出井对角单位矩阵; k ——模型参数的个数; k_i, k_p ——注入井和采出井总数; L ——似然估计值; m ——可观测数据量; M_i ——第 i 口采出井与所有注入井 Pearson 相关系数绝对值的最大值; M_j ——第 j 口注入井与所有采出井 Pearson 相关系数绝对值的最大值; n ——数据对个数; N ——总时间步数; p ——滞后阶数,表征最多以往 p 个月的采出井流量可

对现在采出井流量产生影响; q ——数据对序号; Q_{aic} ——AIC 信息准则评价值; Q_{bic} ——BIC 信息准则评价值; Q_{hqic} ——HQ 信息准则评价值; r ——Pearson 相关系数; R_{ij} ——第 i 口采出井产油量与第 j 口注入井注水量之间的 Pearson 相关系数; s , t ——时间索引; u_t —— t 时刻的残差值向量, $u_t = Y_t - \hat{Y}_t$, m^3/d ; $\bar{u}_t$ ——平均预测误差矩阵, m^3/d ; $y(h)$ ——未来 h 时刻的累计影响矩阵, m^3/d ; $y_{i,t}$ ——第 i 口采出井 t 时刻的产油量, m^3/d ; Y_t —— t 时刻的采出井产油量向量, m^3/d ; $\hat{Y}_t$ —— Y_t 的估计值, m^3/d ; Z_t ——注入井在未来 t 时刻的采油贡献量, m^3 ; α, β ——优选后的注入井和采出井序号; $\sigma_i(h)$ ——第 i 口采出井在未来 h 时刻的估计预测误差, m^3/d ; Φ_t —— t 时刻的累计影响参数矩阵。

参考文献:

- [1] 林伯韬, 郭建成. 人工智能在石油工业中的应用现状探讨[J]. 石油科学通报, 2019, 4(4): 403-413.
LIN Botao, GUO Jiancheng. Discussion on current application of artificial intelligence in petroleum industry[J]. Petroleum Science Bulletin, 2019, 4(4): 403-413.
- [2] 贾爱林, 郭建林. 智能化油气田建设关键技术与认识[J]. 石油勘探与开发, 2012, 39(1): 118-122.
JIA Ailin, GUO Jianlin. Key technologies and understandings on the construction of smart fields[J]. Petroleum Exploration and Development, 2012, 39(1): 118-122.
- [3] LI J, CASTAGNA J, LI D, et al. Reservoir prediction via SVM pattern recognition[C]//SEG Technical Program Expanded Abstracts 2004. Tulsa: Society of Exploration Geophysicists, 2004: 425-428.
- [4] 邢明海, 陈祥光, 王渝. 应用模糊神经网络预测油田产量[J]. 计算机仿真, 2005, 22(2): 187-190.
XING Minghai, CHEN Xiangguang, WANG Yu. Oilfield output forecast using fuzzy neural network[J]. Computer Simulation, 2005, 22(2): 187-190.
- [5] WEBER D, EDGAR T F, LAKE L W, et al. Improvements in capacitance-resistive modeling and optimization of large scale reservoirs[R]. SPE 121299, 2009.
- [6] ANIFOWOSE F A. Artificial intelligence application in reservoir

- characterization and modeling: Whitening the black box[R]. SPE 155413, 2011.
- [7] GHERABATI S A, HUGHES R G, ZHANG H, et al. A large scale network model to obtain interwell formation characteristics[R]. SPE 153386, 2012.
- [8] SALAZAR M, GONZALEZ H, MATRINGE S, et al. Combining decline-curve analysis and capacitance-resistance models to understand and predict the behavior of a mature naturally fractured carbonate reservoir under gas injection[R]. SPE 153252, 2012.
- [9] HOLDAWAY K R. Data mining methodologies enhance probabilistic well forecasting[R]. SPE 167428-MS, 2013.
- [10] 张方舟, 严胡勇, 杨立全, 等. 改进型灰色神经网络模型在油田产量中的应用[J]. 计算机技术与发展, 2013, 23(6): 241-244.
ZHANG Fangzhou, YAN Huyong, YANG Liquan, et al. Application of improved grey neural network model to oil yields[J]. Computer Technology and Development, 2013, 23(6): 241-244.
- [11] GUPTA S, FUEHRER F, JEYACHANDRA B C. Production forecasting in unconventional resources using data mining and time series analysis[R]. SPE 171588-MS, 2014.
- [12] ZHOU Q, DILMORE R, KLEIT A, et al. Evaluating gas production performances in Marcellus using data mining technologies[J]. Journal of Natural Gas Science and Engineering, 2014, 20(1): 109-120.
- [13] LI J, LEI Z, LI S, et al. Optimizing water flood performance to improve injector efficiency in fractured low-permeability reservoirs using streamline simulation[R]. SPE 182779, 2016.
- [14] CAO Q, BANERJEE R, GUPTA S, et al. Data driven production forecasting using machine learning[R]. SPE 180984, 2016.
- [15] 赵辉, 康志江, 孙海涛, 等. 水驱开发多层油藏井间连通性反演模型[J]. 石油勘探与开发, 2016, 43(1): 99-106.
ZHAO Hui, KANG Zhijiang, SUN Haitao, et al. An interwell connectivity inversion model for waterflooded multilayer reservoirs[J]. Petroleum Exploration and Development, 2016, 43(1): 99-106.
- [16] MAQUI A F, ZHAI X, SUAREZ N A, et al. A comprehensive workflow for near real time waterflood management and production optimization using reduced-physics and data-driven technologies[R]. SPE 182696, 2017.
- [17] 武志军. Spark 环境下基于数据挖掘的油田产量预测技术研究与应用[D]. 青岛: 中国石油大学(华东), 2017.
WU Zhijun. Research and application of oilfield production forecast technology based on data mining in Spark environment[D]. Qingdao: China University of Petroleum, 2017.
- [18] MARTIN E, WILLS P, HOHL D, et al. Using machine learning to predict production at a Peace River thermal EOR site[R]. SPE 180984, 2017.
- [19] 贾虎, 邓力琿. 基于流线聚类人工智能方法的水驱油藏流场识别[J]. 石油勘探与开发, 2018, 45(2): 312-319.
JIA Hu, DENG Lihui. Oil reservoir water flooding flowing area identification based on the method of streamline clustering artificial intelligence[J]. Petroleum Exploration and Development, 2018, 45(2): 312-319.
- [20] KHAN M, ALNUAIM S, TARIQ Z, et al. Machine learning application for oil rate prediction in artificial gas lift wells[R]. SPE 194713, 2019.
- [21] NOSHI C, EISSA M, ABDALLA R, et al. An intelligent data driven approach for production prediction[R]. SPE 29243-MS, 2019.
- [22] 杨海民, 潘志松, 白玮. 时间序列预测方法综述[J]. 计算机科学, 2019, 46(1): 21-28.
YANG Haimin, PAN Zhisong, BAI Wei. Review of time series prediction methods[J]. Computer Science, 2019, 46(1): 21-28.
- [23] 梁吉业, 冯晨娇, 宋鹏. 大数据相关分析综述[J]. 计算机学报, 2016, 39(1): 1-18.
LIANG Jiye, FENG Chenjiao, SONG Peng. A survey on correlation analysis of big data[J]. Chinese Journal of Computers, 2016, 39(1): 1-18.
- [24] PEARSON K. Mathematical contributions to the theory of evolution (III): Regression, heredity, and panmixia[J]. Philosophical Transactions of the Royal Society of London. Series A, 1896, 187: 253-318.
- [25] LÜTKEPOHL H. New introduction to multiple time series analysis[M]. Berlin: Springer Science & Business Media, 2005.
- [26] AKAIKE H. Information theory and an extension of the maximum likelihood principle[R]. Budapest: 2nd International Symposium on Information Theory, 1973.
- [27] SCHWARZ G. Estimating the dimension of a model[J]. Annals of Statistics, 1978, 6: 461-464.
- [28] HANNAN E J, QUINN B G. The determination of the order of an autoregression[J]. Journal of the Royal Statistical Society. Series B (Methodological), 1979, 41: 190-195.
- 第一作者简介: 张瑞 (1994-), 男, 湖北武汉人, 西南石油大学在读博士研究生, 主要从事油藏工程与数值模拟方面的研究工作。地址: 四川省成都市新都区西南石油大学石油与天然气工程学院, 邮政编码: 610500。
E-mail: zhangruixiao@163.com
- 联系作者简介: 贾虎 (1983-), 男, 湖北武汉人, 博士, 西南石油大学教授, 主要从事油田化学与提高油气采收率方面的研究工作。地址: 四川省成都市新都区西南石油大学石油与天然气工程学院, 邮政编码: 610500。
E-mail: tiger-jia@163.com
- 收稿日期: 2020-02-17 修回日期: 2020-10-15

(编辑 胡苇玮)