

特色专题

基于机器学习的流体深部地质封存潜力智能预测方法

杨占^{1,2}, 李祥^{1,2}, 张啸^{1,2}, 陶盛^{1,2}, 杜松^{1,2*}

摘要 提出一种融合运行监测数据与储层结构特征的深部流体封存潜力智能预测方法框架,通过引入机器学习与时序特征驱动建模,可有效识别封存阶段并精准预测注入压力演化。该方法结合长期注入压力、流量高频监测数据与岩芯微观孔隙和微裂隙结构信息,构建阶段标签驱动的分段预测机制,并利用多模型对比与交叉验证策略提升预测稳健性。在高矿化度流体深部封存工程场景中,基于18个月现场运行数据(累计注入量超过150万t)验证了该方法。结果表明,该智能预测流程可有效识别突破期与填充期等关键运行阶段,显著提升压力预测精度,最优模型的平均绝对百分比误差低至0.6,同时降低对复杂机理模型与大规模标注样本的依赖。在不同时段、不同运行阶段数据条件下均表现出良好的泛化能力与稳定性,明显提高了运行状态评估与调控决策效率。该方法具有良好的可扩展性,可为多类型深部流体注入与封存工程的智能评估与安全运行提供技术支撑。

关键词 深部地质封存; 深地流体; 孔隙压力; 机器学习; 深度学习

近年来,CO₂排放持续增长,已成为全球气候变化的主要驱动因素。为应对气候风险,多个国家相继提出碳中和目标。然而,当前全球能源结构中化石能源占比仍超过80%,实现减排目标亟需源头减排和末端封存协同推进,其中碳捕集与封存(carbon capture and storage, CCS)被广泛认为是实现大规模负碳的重要技术路径^[1]。研究表明,到2030年CCS年减排潜力可达1亿~4亿t CO₂,至2050年有望提升至30亿~68亿t CO₂^[2]。作为CCS体系中的关键环节,深部地质封存技术通过将流体注入1500~3500 m深部地层中,依托盖层封闭性,实现流体与生物圈的长期隔离,

是目前最具应用前景的负碳方式之一^[3]。

在流体深部地质封存过程中,注入压力是刻画流体运移行为、储层应力演化及封存潜力变化的关键观测指标,对封存安全与运行效率具有直接影响。然而,传统机理模型对复杂地层条件和非线性压力演化过程刻画能力有限;同时,超临界CO₂具有较高可压缩性,易受相态变化与溶解效应影响,导致压力响应对真实孔隙应力变化的指示性受限,从而增加运行监测与调控难度^[4]。此外,CO₂封存多为小试规模为主,长期连续运行数据相对匮乏,制约了数据驱动预测方法的系统构建与验证^[5-6]。因此,亟需发展一种融合运行监测数据与地质结构特征的智能预测方法,以实现深部封存过程压力响应与封存潜力的高精度刻画。

相比之下,高矿化度矿井水在储层-盖层配置要

1. 中国煤炭地质总局勘查研究总院,北京 100039

2. 矿山透明地质与数字孪生技术国家矿山安全监察局重点实验室,北京 100039

收稿日期: 2025-11-21; 修回日期: 2026-02-10

基金项目: 国家重点研发计划项目(2023YFC3012104); 鄂尔多斯市标志性创新团队项目(TD20232305)

作者简介: 杨占,工程师,研究方向为深部地质封存下新污染物迁移、机器学习、图像融合,电子信箱: 1810810@tongji.edu.cn; 杜松(通信作者),正高级工程师,研究方向为流体深部地质封存技术、深部地层流体渗流规律,电子信箱: dusongpx@gmail.com

引用格式: 杨占,李祥,张啸,等. 基于机器学习的流体深部地质封存潜力智能预测方法[J]. 科技导报, 2026, 44(6): 68-75; doi:10.3981/j.issn.1000-7857.2025.12.00122

求上与 CO_2 封存具有一致性,但其可压缩性较低,注入压力对孔隙压力变化更为敏感,且与地层原生水理化性质相近,地球化学扰动风险较低^[7-8]。同时,该类流体来源稳定、规模可观,更适合开展百万吨级长期连续注入试验,为构建数据驱动的压力响应分析与预测方法提供了理想应用场景^[9]。

基于上述需求,提出一种融合运行监测数据与储层微观结构特征的深部流体封存潜力智能预测方法框架,引入机器学习与深度学习技术构建阶段识别与压力预测模型,并在高矿化度流体封存工程场景中开展了验证。结果表明,该方法能够有效识别关键运行阶段并实现高精度压力预测。所提出的方法具有良好的可扩展性与可迁移性,后续可推广应用于 CO_2 地质封存、深井回灌、高盐水处置及其他深部流体注入工程的运行评估与风险预警。

1 研究方法

1.1 试验区概况

本研究试验区位于黄土高原地区东胜梁南侧,处于区域地表分水岭附近(图1)。区内地形总体呈北高南低、西高东低特征,海拔范围为 1223.6~1398.7 m,最大高差约 175 m,整体起伏约 90 m。区域气候属温带大陆性干旱、半干旱气候,年降水量为 194.7~531.6 mm,平均约 396.0 mm,降水主要集中在 7~9 月。

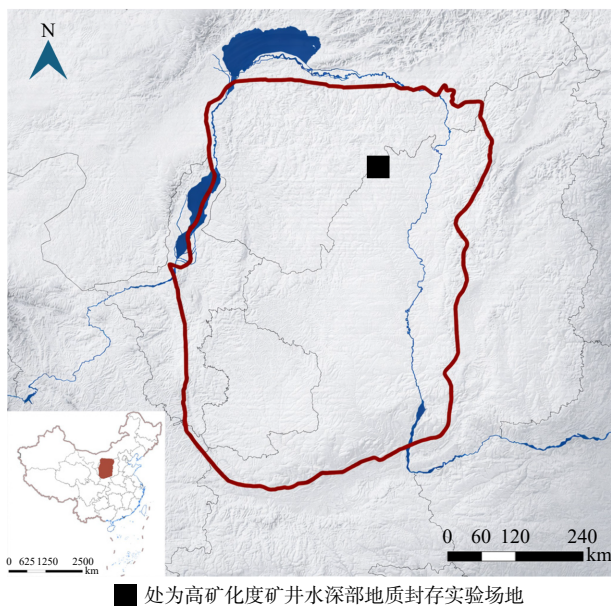


图1 鄂尔多斯盆地示意

1.2 井位布设与注入系统

试验共布设 1 口注入井和 1 口监测井。注入井位于 $\text{N}39^\circ13'36''$ 、 $\text{E}109^\circ53'48''$,井口高程 1250 m,井深 2240 m;监测井位于 $\text{N}39^\circ11'54''$ 、 $\text{E}109^\circ52'35''$,井口高程 1263 m,井深 2247 m。注入系统采用柱塞泵,单泵最大工作压力 12 MPa,最大流量 $130 \text{ m}^3/\text{h}$,共安装 1 台。柱塞泵出口压力可直接反映储层孔隙压力变化,是评估封存过程压力响应的重要参数。

1.3 岩芯采集与微观表征

为表征储层微观结构特征,在刘家沟组和石千峰组采集岩芯,其中刘家沟组 10 块、石千峰组 9 块,取芯深度主要分布在 1820~2113 m,覆盖两套地层的中部及中下部。代表性砂岩样品经切割、干燥和镀金处理后,采用扫描电子显微镜(scanning electron microscope, SEM)对孔隙形态和微观结构进行观察与分析。

1.4 监测数据处理与模型构建

研究采用 2023 年 1 月至 2024 年 10 月柱塞泵运行数据,累计运行 18 个月。首先剔除低于 4 MPa(启泵阶段)和高于 12 MPa(极限工况)的异常数据。针对原始数据采样频率高、冗余度大的特点,采用滑动平均-降采样策略,以 20、40、60 min 为窗口对原始数据做均值重采样,结果表明 40 min 重采样在保持主要动力学特征的同时有效降低噪声,因此选择用于后续研究。

在模型训练前,对数据进行预处理:采用滑动平均法(窗口宽度 $N=5$)平滑数据以削弱短期波动;按月识别并剔除超过 1.5 倍四分位距(interquartile range, IQR)的异常值;删除缺失数据后对样本进行零均值标准化处理。最终数据按时间顺序划分为训练集(前 80%)和测试集(后 20%)。

为预测深部地质封存过程中泵压变化特征,构建并对比了自回归积分滑动平均模型(autoregressive integrated moving average model, ARIMA)、极端梯度提升(extreme gradient boosting, XGBoost)和长短期记忆网络(long short-term memory, LSTM)3 类时间序列模型。ARIMA 作为典型统计时序模型,适用于刻画线性与平稳序列结构;XGBoost 作为树模型驱动的集成学习方法,在非线性关系建模与特征交互刻画方面具有优势;LSTM 作为深度学习时序模型,能够通过

记忆单元捕捉长期时间依赖特征。3类模型在建模机理与特征学习方式上具有互补性,能够形成多范式对比框架,从而更全面评估不同智能方法在深部流体注入压力预测场景下的性能差异与适用范围。模型训练与计算均基于 Python 3.9 平台,采用 Scikit-learn 和 PyTorch 等开源库实现。

1.5 评价指标

采用均方根误差(root mean square error, RMSE)、平均绝对百分比误差(mean absolute percentage error, MAPE)和最大误差(max error)对模型性能进行评价,分别表征预测精度、相对误差水平及极端工况下的预测能力。

2 研究结果

2.1 岩芯分析

深部地质封存对储层与盖层的基本要求与 CO₂ 地质封存类似,需依托低渗透性盖层实现流体的长期封隔^[10]。本研究选取的储层为下三叠统刘家沟组 and 上二叠统石千峰组河流相砂岩,整体表现为低孔-低渗储层特征(渗透率 $(4.0\sim 7.0)\times 10^{-16} \text{ m}^2$),其上覆泥岩及砂质泥岩层系构成有效隔水层,为封存安全提供地质保障^[11]。

岩芯的 SEM 分析结果表明,2套储层砂岩的储集空间主要由粒间孔、粒内溶蚀孔和微裂隙3类孔-裂隙结构组成(图2)。粒间孔尺度介于百纳米至微米级,多呈不规则形态,受颗粒粒径、堆积方式、压实作用及胶结物分布共同控制。石千峰组砂岩中可见最大粒间孔径达百微米以上,刘家沟组砂岩粒间孔相对

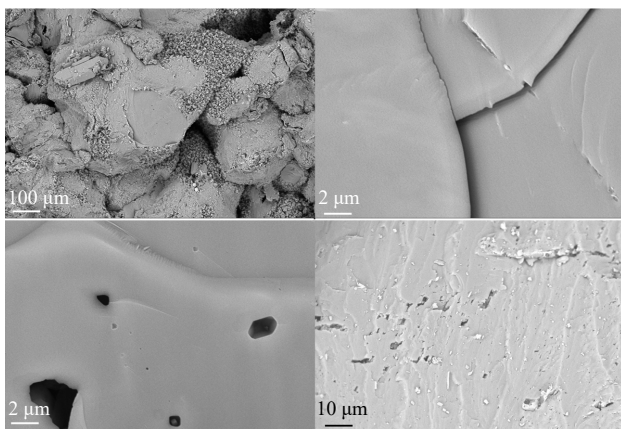


图2 刘家沟组和石千峰组中岩芯 SEM 图像

偏小,分布较为连续。粒内溶蚀孔以长石溶蚀孔为主,孔径多小于 10 μm,部分溶蚀孔与粒间孔或微裂隙连通,对有效孔隙度具有重要贡献^[12]。溶蚀孔的发育程度受孔隙流体性质、原始孔隙结构及碎屑矿物组成影响显著。微裂隙广泛发育于2套地层中,尺度以纳米至微米级为主,多小于 10 μm,既包括沿矿物解理发育的粒内裂隙,也包括切穿颗粒的粒间裂隙,形态多呈不规则或弧形,反映了地层原位应力对脆性矿物(如长石)的改造作用^[13]。微裂隙为流体运移提供了优先通道,是控制渗透性演化的重要结构单元^[14]。

在注入过程中,井周围砂岩及相邻泥岩层孔隙水压力逐步升高,诱发微裂隙萌生和原有孔隙的扩展,导致近井区孔隙度与渗透率阶段性增强^[15]。由于该过程在空间上具有非定向性,注入初期泵压与流量波动显著,压力-流量关系高度离散,对应“突破期”特征。当孔-裂隙体系逐渐连通并趋于稳定后,近井区逐步达到饱和状态,流动阻力主要受地层固有渗透性控制,泵出口压力趋于平稳,流量波动减弱,压力-流量数据明显收敛,进入“填充期”。随后,压力前缘向远井区扩展,再次诱发新的微裂隙和孔隙发育,导致压力与流量重新出现波动,进入新一轮“突破期”。上述突破期与填充期的交替演化,共同构成了注入压力随时间呈现的阶段性振荡特征。

2.2 压力和流量演化分析

图3(a)~(i)展示了注入期间每2个月注入泵压力与流量的变化特征。整体来看,注入压力主要分布在 4~9 MPa,流量范围为 25~125 m³/h。注入初期第1~2个月(图3(a)),压力-流量数据点高度离散,随着压力升高,流量波动显著增强,表明井周围储层尚未形成稳定的流动通道,注入过程以局部突破和流向调整为主。至第3~4个月(图3(b)),压力逐步升高至 5.0~6.5 MPa,流量同步增加至 75~125 m³/h,反映出近井区孔-裂隙体系逐渐连通,注入能力明显提升^[16]。

在第5~6个月(图3(c)),压力与流量均趋于稳定,压力主要集中在 6.5~7.5 MPa,流量集中于 105~120 m³/h,对应注入系统的首个稳定运行阶段。随后,第7~8个月(图3(d))压力-流量数据再次出现离散特征,压力分布在 6.2~7.5 MPa,而流量明显下降至约 75 m³/h,表明注入压力前缘向远井区推进,诱发新的孔、裂隙发育与流动路径重构。第9~10个月(图3(e))

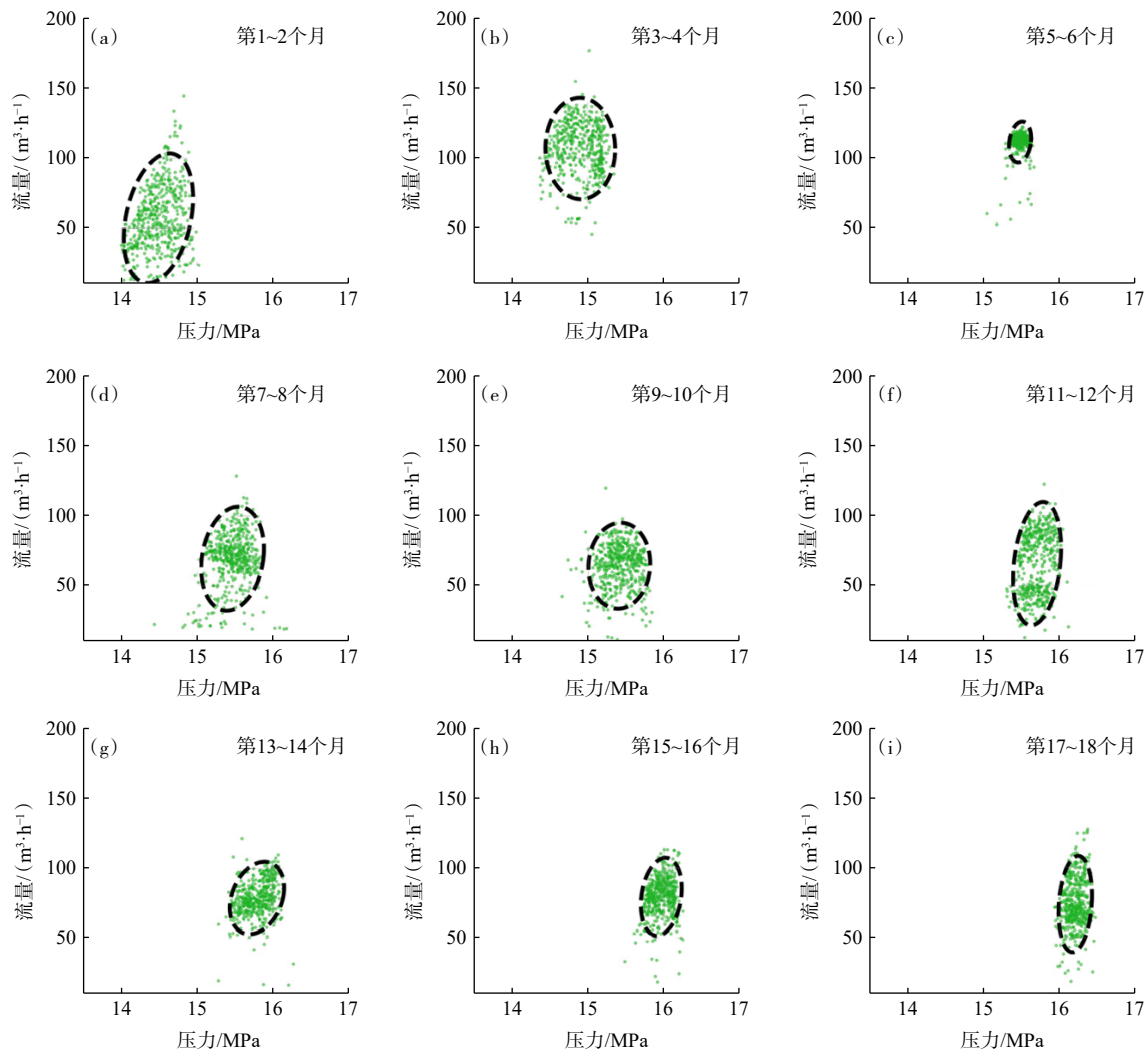


图3 封存期间流量-压力分布情况

压力波动幅度较小,但流量呈现中等幅度下降。第11~12个月(图3(f))期间,流量先降低后逐步恢复,注入压力持续升高至6.9~8.0 MPa。

在第13~18个月(图3(g)~(i)),流量基本稳定在70~85 m³/h,而注入压力则持续升高并逐步接近9 MPa,表明注入范围进一步扩大,井周围及更远距离地层孔隙与微裂隙不断发育,系统整体渗流阻力逐渐增加。对18个月全时段数据进行核密度分析,可识别出3个高密度区,其中最大高密度区对应第5~6个月的稳定期,标志着首轮“突破-填充”过程的完成。随后在约8 MPa附近出现第二个稳定高密度区,伴随流量由110~120 m³/h下降至70~78 m³/h,反映新一轮突破期启动并逐步进入填充阶段。

基于压力-流量数据的离散特征及时间演化规律,注入过程可划分为2个注入循环:周期1(第1~

6个月)和周期2(第7~18个月)。周期1结束后,井周围地层趋于饱和,后续注入需克服更大范围内的地层孔隙压力,促使周期2持续时间显著延长,并对应更大的累计注入量。

结合箱线图进一步分析(图4),第5~6个月及第17~18个月泵出口压力箱体显著收敛,基于1.5IQR明显减少,验证了2个注入循环的阶段特征。受储层

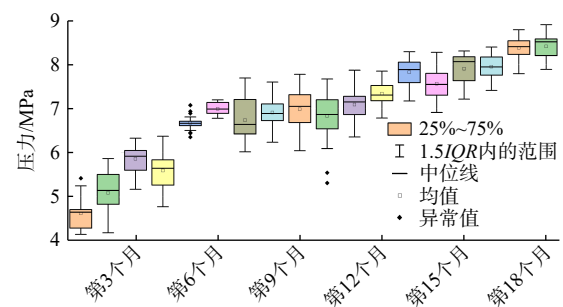


图4 各月份压力变化情况

非均质性和各向异性影响,注入过程并非单调推进:当某一方向率先发生流体突破后,局部孔隙压力升高,迫使流体重新分配流向。基于此,周期1可细分为阶段1a(第1~3个月)和阶段1b(第4~6个月);周期2进一步划分为阶段2a(第7~9个月)、阶段2b(第10~13个月)和阶段2c(第14~18个月)。

从流量特征看,周期1中位流量波动范围较大,介于40~115 m³/h,体现出早期突破期流动通道不稳定的特征;而周期2整体更加稳定,中位流量维持在65~90 m³/h。值得注意的是,在周期2中,尽管流量保持相对稳定,注入压力仍持续升高,表明注入影响半径不断扩大,孔-裂隙体系在更远距离逐步发育,同时累积孔隙压力不断增加。该阶段性划分较好地刻画了深部地质封存过程中流体分布与地应力之间的耦合演化关系。

为实现对注入过程压力变化的可解释预测,并考虑不同运行状态下压力-流量关系的显著差异,将前述“循环-阶段”结果进一步形式化为可用于机器学习建模的阶段标签。具体而言,首先依据压力-流量散点离散度及其稳定高密度区特征,将全时段划分为2个注入周期:周期1(第1~6个月)与周期2(第7~18个月)。在此基础上,结合月尺度压力箱线图的收敛性以及压力序列的单调趋势,将周期进一步细分为5个阶段:第1~3个月、第4~6个月、第7~9个月、第10~13个月与第14~18个月。

从物理意义上,上述阶段可归纳为2类运行状态:突破期与填充期。突破期对应压力-流量数据离散、流量波动较强的区间,指示近井区或压力前缘处孔-裂隙通道的逐步贯通与流向重构;填充期对应数据点收敛、流量波动减弱的区间,反映近井区趋于饱和和后流动阻力主要由地层固有渗透性控制。基于该定义,后续采用2种互补策略:其一,在全数据集上训练统一模型并评估总体预测能力;其二,按“突破/填充”进行分阶段建模与对比分析,以检验阶段差异对最优时间窗、特征贡献及模型泛化能力的影响,并为现场实时调控提供更具针对性的预测结果。

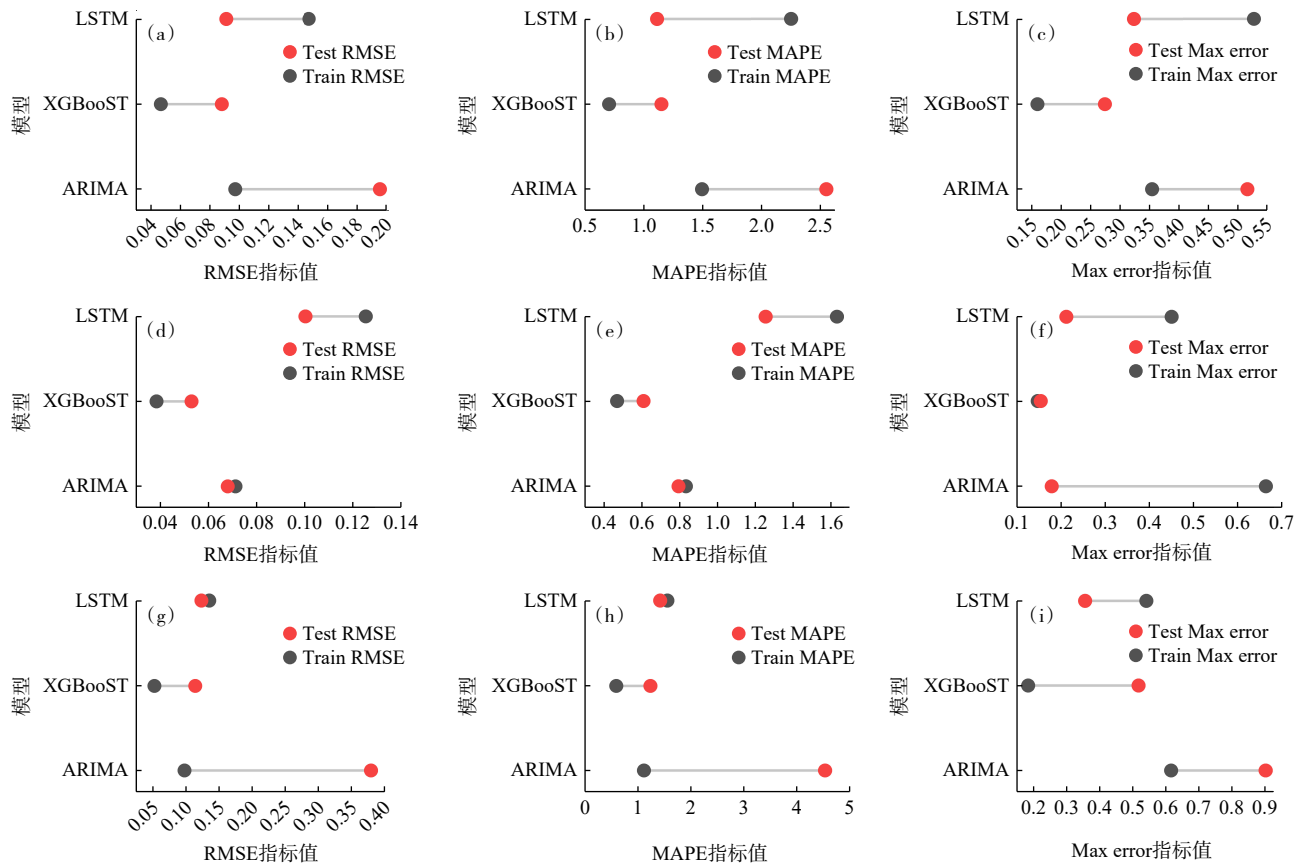
2.3 模型训练和比较

为揭示注入过程中压力-流量数据中潜在的时间序列特征,基于实际工程监测数据,对传统统计模型、机器学习模型和深度学习模型的预测性能进行系统

对比分析。ARIMA是一种基于时间序列统计特性的预测方法,通过利用序列的自相关性和趋势性进行建模,在时间序列预测中得到广泛应用^[17]。本研究针对每一运行阶段,选取80%的数据作为训练集,其余20%作为测试集,并固定滑动窗口长度为5,即模型预测基于前5个时间步的历史信息。在一次差分($d=1$)后,将非平稳序列转化为平稳序列,仅采用自回归项对历史信息的影响进行刻画。结果表明,ARIMA模型在训练集中的拟合效果较好,各阶段RMSE均小于0.1,MAPE约为1.0(图5)。然而,在测试集中模型误差显著放大,预测性能明显下降(图6),部分阶段的误差达到训练集的数倍。这主要源于深部地质封存过程中注入压力随时间呈现出复杂的波动上升特征,传统统计模型难以充分识别其中的非线性变化与突变行为。因此,与机器学习和深度学习方法相比,ARIMA在多个时间段的预测精度相对不足。

XGBoost是一种基于决策树的集成学习方法,通过逐步引入弱学习器以降低整体预测误差,已被证明在处理非平稳数据和刻画复杂非线性关系方面具有优势^[18-19]。在本研究中,XGBoost模型能够较好地学习训练集中的时序特征,并在测试集中表现出较高的预测精度^[20]。多个阶段中,其误差指标均低于ARIMA和LSTM模型(图5)。但同时也发现,在几乎所有阶段,测试集误差普遍高于训练集,表明模型在一定程度上存在对训练数据噪声的过拟合现象。为降低单次训练-测试划分带来的随机性,引入5折交叉验证方法对模型泛化能力进行评估。结果显示,交叉验证得到的RMSE介于训练集与测试集之间(图5),表明XGBoost在各阶段具有较为稳定的泛化性能。然而,由于不同时间段数据分布存在显著差异,模型预测效果仍受到阶段特征影响,在数据波动较小的填充期表现较优,而在突破期(如第7~9个月)MAPE和最大误差明显增大。

本研究采用的LSTM模型由2层LSTM网络和1层全连接层组成,使用Adam优化器,在每一阶段迭代训练100次以降低预测误差。结果表明,尽管在部分阶段训练集中受样本数量影响,LSTM未取得最优拟合效果,但其在测试集中的预测误差相对较低,表现出更强的泛化能力。这主要得益于LSTM能够有效捕捉时间序列后段的趋势特征和长期依赖关系^[21]。



(a)(b)(c) 模型在阶段 1a 的运行效果; (d)(e)(f) 模型在阶段 1b 的运行效果; (g)(h)(i) 模型在阶段 2a 的运行效果

图 5 3 种模型在周期 1 中的评价结果

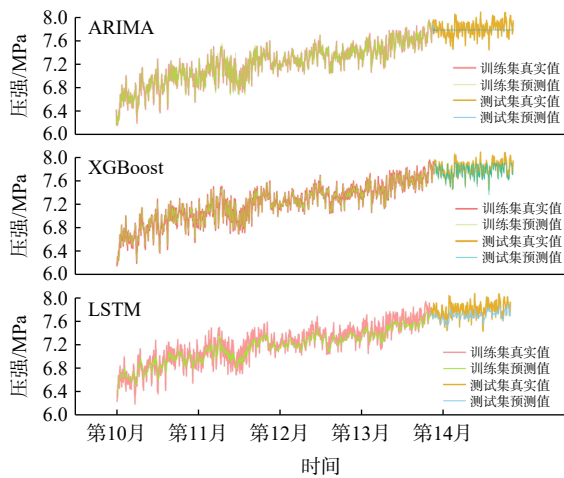


图 6 第 10~14 个月 3 种模型的训练预测情况

总体来看, LSTM 在不同时间段和不同数据分布条件下表现更加稳定, RMSE 集中在 0.1 左右, MAPE 约为 1.0。尽管深度学习模型对数据规模和计算复杂度要求较高,在短期预测中可能略逊于树模型,但其在结构扩展和模型优化方面具有更大潜力^[22]。通过增加训练样本、扩充网络规模或引入注意力机制, LSTM

有望进一步提升对长期压力演化过程的预测能力^[23]。

2.4 研究局限性和未来展望

本研究仍存在一定局限性。首先,当前监测系统仅布设 1 口距离注入井约 150 m 的观测井,虽然压力响应的耦合特征能够从整体上反映储层运行稳定性,但在空间分辨率上仍不足以精细刻画多尺度压力传导过程及流体运移路径分布特征。未来研究有必要构建多观测井联合监测体系,形成近井-中场-远场多尺度监测网络,以提高对压力传播范围与封存影响半径的综合识别能力。

其次,本研究监测指标主要基于压力与流量数据,对流体地球化学特征的联合约束仍相对有限。后续可将水化学参数(如总溶解性固体及关键离子比值)与压力、温度等运行参数进行多源融合分析,从水力一致性与地球化学一致性 2 个维度协同验证储层封闭性与流体稳定性。

此外,本研究监测时长为 18 个月,仍属于中期运行尺度,尚不足以独立验证深部封存的长期安全性。

未来将结合研究区地质与水文地质条件,进一步发展长期连续监测与智能评估方法,构建 10 年尺度以上的运行安全动态评价体系,为深部流体封存技术的规模化推广与长期安全运行提供持续数据支撑与方法保障。

3 结论

以注入泵出口压力为关键监测指标,分析了深部地质封存过程中储层应力响应特征,为理解地下流体封存过程中的地球力学行为提供了新的认识。结合前期地质调查成果与实际运行期间泵压变化特征,推断注入后流体在储层中的运移主要受裂隙导流和孔隙扩散机制共同控制。基于 18 个月连续运行过程中获取的 3 亿余条泵运行监测数据,对注入过程进行了系统分析,结果表明,可依据压力-流量数据的离散程度、月均压力的单调性及其整体演化趋势,将封存过程合理划分为突破期与填充期。引入阶段标签开展分阶段模型训练后,压力预测精度得到显著提升。

在模型对比中,XGBoost 在短期预测方面表现最优,尤其在压力波动较小的填充期,其预测精度最高(RMSE 为 0.09)。尽管 LSTM 在短期预测精度上略低于 XGBoost,但其内部记忆机制在刻画时间序列长期依赖关系方面具有优势。随着监测数据的持续积累,LSTM 在长期压力演化预测中的能力有望进一步增强,这对于深部地质封存的长期安全运行与可持续实施具有重要意义。

尽管岩芯分析能够直观揭示储层岩石的微观结构特征,但其在刻画原位封存过程的动态演化方面仍存在局限。相比之下,本研究发现,长期压力与流量监测为认识深部地质封存过程提供了一种新的研究视角。通过对压力随时间变化特征的分析,识别出压力周期性升高与离散增强的演化模式,该特征指示了储层孔隙与微裂隙的逐步发育过程。这一发现不仅与既有理论认识相一致,也在工程数据层面上首次验证了深部地质封存过程中孔-裂隙动态演化机制,为理解流体封存的孔隙尺度约束机理提供了新的实证依据。

此外,本研究构建的机器学习预测模型在实际工程场景中表现出良好的适用性,未来可进一步拓展用于封存容量评估、流体运移前缘预测以及封存过程的

动态风险评估,为深部地质封存工程的智能化运行与决策支持提供技术支撑。

参考文献(References)

- [1] Massarweh O, Abushaikh A S. CO₂ sequestration in subsurface geological formations: A review of trapping mechanisms and monitoring techniques[J]. *Earth-Science Reviews*, 2024, 253: 104793.
- [2] 赵欣豪,刘会虎,丁海,等. 二氧化碳地质封存工程实例及技术分析[J]. *中国煤炭地质*, 2024, 36(10): 66-69.
- [3] Song Y, Jun S, Na Y, et al. Geomechanical challenges during geological CO₂ storage: A review[J]. *Chemical Engineering Journal*, 2023, 456: 140968.
- [4] Babarinde O, Schwartz B, Meng J, et al. An overview of geological carbon sequestration and its geomechanical aspects[J]. *Geological Society, London, Special Publications*, 2023, 528(1): 61-72.
- [5] Zhao W, Jiang Z, Jing T, et al. Integrated environmental-economic assessment of CO₂ storage in Chinese saline formations[J]. *Water*, 2025, 17(15): 2320.
- [6] Lin K, Wei N, Zhang Y, et al. Hydro-mechanical interactions in CO₂ storage: Critical parameters influencing coal mine at Shenhua's CCS site[J]. *Engineering Geology*, 2025, 352: 108087.
- [7] Pumjan S, Long T T, Loc H H, et al. Deep well injection for the waste brine disposal solution of potash mining in north-eastern Thailand[J]. *Journal of Environmental Management*, 2022, 311: 114821.
- [8] Du S, Li D, Yang Z, et al. Detection of high-salinity water in deep geological sequestration using the wide-field electromagnetic method[J]. *Journal of Earth Science*, 2025. doi: 10.1007/s12583-025-0224-3.
- [9] Luboń K, Tarkowski R. Influence of capillary threshold pressure and injection well location on the dynamic CO₂ and H₂ storage capacity for the deep geological structure[J]. *International Journal of Hydrogen Energy*, 2021, 46(58): 30048-30060.
- [10] Du Z, Song J, Du S, et al. Numerical modeling of geological sequestration of brine wastewater due to coal mining in the Ordos Basin, China[J]. *Science of The Total Environment*, 2024, 908: 168580.
- [11] Liu Y, Bu Y, Du S, et al. Reservoir characteristics and regional storage potential evaluation of deep well injection and storage of high-salinity water in coal mines in the Ordos Basin[J]. *Processes*, 2025, 13(2): 579.
- [12] Che Q, Du S, Zhang D, et al. Monitoring-based study of migration characteristics of highly saline mine water during deep well injection and storage in the Ordos Basin[J].

- Processes, 2025, 13(2): 494.
- [13] Zoback M D. Reservoir geomechanics[M]. Cambridge: Cambridge University Press, 2007.
- [14] Lin Y, Wang S, Qiao J, et al. Chemical characteristics, formation mechanisms, and geological evolution processes of high-salinity coal reservoir water in the Binchang area of the southern Ordos Basin, China[J]. *International Journal of Coal Geology*, 2024, 291: 104574.
- [15] Lu C, Cheng W, Yin H, et al. Study on inverse geochemical modeling of hydrochemical characteristics and genesis of groundwater system in coal mine area: A case study of Longwanggou Coal Mine in Ordos Basin[J]. *Environmental Science and Pollution Research*, 2024, 31(11): 16583–16600.
- [16] Tsalidis G A, Tourkodimitri K P, Mitko K, et al. Assessing the environmental performance of a novel coal mine brine treatment technique: A case in Poland[J]. *Journal of Cleaner Production*, 2022, 358: 131973.
- [17] Sharma A K, Punj P, Kumar N, et al. Lifetime prediction of a hydraulic pump using ARIMA model[J]. *Arabian Journal for Science and Engineering*, 2024, 49(2): 1713–1725.
- [18] Chen T, He T, Benesty M, et al. Xgboost: Extreme gradient boosting[J]. *R Package Version*, 2025, 42(1): 1–4.
- [19] Lan D, He P, Qi Y, et al. Machine learning and hyperspectral imaging-aided forecast for the share of biogenic and fossil carbon in solid waste[J]. *ACS Sustainable Chemistry & Engineering*, 2023, 11(10): 4020–4029.
- [20] Yang Z, Zhang H, Lu F, et al. A novel high-throughput detection method for plastic debris in organic-rich matrices based on image fusion[J]. *Analytical Chemistry*, 2024, 96(15): 6045–6054.
- [21] Landi F, Baraldi L, Cornia M, et al. Working memory connections for LSTM[J]. *Neural Networks*, 2021, 144: 334–341.
- [22] Luo J, Gong Y. Air pollutant prediction based on ARIMA-WOA-LSTM model[J]. *Atmospheric Pollution Research*, 2023, 14(6): 101761.
- [23] Zha W, Liu Y, Wan Y, et al. Forecasting monthly gas field production based on the CNN-LSTM model[J]. *Energy*, 2022, 260: 124889.

An intelligent prediction method for deep fluid storage potential based on machine learning

YANG Zhan^{1,2}, LI Xiang^{1,2}, ZHANG Xiao^{1,2}, TAO Sheng^{1,2}, DU Song^{1,2*}

1. General Prospecting and Research Institute of China National Administration of Coal Geology, Beijing 100039, China

2. Key Laboratory of Transparent Mine Geology and Digital Twin Technology, National Mine Safety Administration, Beijing 100039, China

Abstract An intelligent prediction framework for deep fluid storage potential is proposed by integrating operational monitoring data with reservoir structural features. By introducing machine learning and time-series feature-driven modeling strategies, the proposed workflow enables automated identification of injection pressure evolution patterns and accurate prediction of storage-stage responses. The method combines long-term high-frequency pressure-flow monitoring data with pore-fracture structural information derived from core-scale analysis, and establishes a stage-labeled, phase-wise prediction mechanism to improve model robustness through multi-model comparison and cross-validation. The framework was validated in a deep injection and storage engineering scenario involving highly saline fluids, using an 18-month field dataset with a cumulative injection volume exceeding 1.5 million tonnes. Results demonstrate that the proposed intelligent workflow can effectively distinguish key operational stages, including breakthrough and filling phases, and significantly enhance pressure prediction accuracy, with the best model achieving a MAPE as low as 0.6. The method reduces dependence on complex mechanistic models and large labeled datasets, and shows strong generalization across different operational periods and stage conditions. Owing to its scalability and transferability, the proposed approach provides technical support for intelligent assessment and safe operation of various deep fluid injection and storage projects.

Keywords deep geological storage; deep geological fluid; pore pressure; machine learning; deep learning ●



(责任编辑 傅雪)