地下模型编程：
前沿大型语言模型原生理解并编辑动态油藏模型

Yanchen Song*、Huizhu Xiang、Zhengdong Lei、Yuezhong Wang、Xingbang Li u

中国石油天然气集团有限公司石油勘探开发研究院，中国
北京 100083

https://github.com/syclong/llm-reservoir-engineering-benchmarks

arXiv:2611.29612v1 [cs.CE] 22 Nov 2026

摘要

非均质多孔介质中的多相流模拟是烃类采收、地质碳封存和地热开采等领域决策的基础。这些模拟依赖于指定网格拓扑、岩石和流体性质以及井控参数的输入文件。然而，人工智能在很大程度上将地下物理模拟文件视为黑箱数值对象，依靠统计代理模型或在数百次正向运行中进行盲目优化。在此，我们提出地下模型编程这一工作流程：推理型大语言模型（LLM）读取并修改原始模拟数据卡，其方式类似于油藏工程师检查源代码。无需数值梯度或外部优化算法，大语言模型便能直接根据物理参数推断三维流动连通性、能量耗竭和流体突破机制，并据此提出有针对性的修改。在四个复杂度递增的基准测试中，使用开源多相模拟器进行端到端评估，零工具全文件推理在静态三维井间连通性解释任务中达到了 100% 的准确率，在单次处理中正确分类了全部六种储层层位井对标签，以及开放完井之间的全部三种流动路径标签。在 PUNQ-S3 上，单次提案修改了 1,761 个活跃网格中的 54 个（3.1%），并使井均衡含水率均方根误差降低了 71.9%。在 SPE10 扇区模型上，五个大语言模型提案中表现最佳者在十年内实现了 1.795 million sm^3 的累积产油量，较进行 25 次评估的粒子群优化基线高出 8.9%。扩展至包含 533,403 个网格单元的 UNISIM-IV-2026 组分模型后，经过五轮基于脚本的编辑，模型识别出了主要的注采井通道，将油田与井的归一化均方根误差从 37.98% 降至 11.62%（降低 69.4%），并使油田产量误差降低了 73.9%；最后一轮的改进来自沿既有注采井连通路径进行的孔隙度修改。由于所有模型输入、修改和模拟器输出均已存档，每项结果都可以

可以直接追溯到其潜在的物理修改。仅通过一至五次运行，以具有物理可审计性的因果干预取代穷举式随机搜索，为自主油藏开发、快速历史拟合、旧模型诊断和耦合多物理场工程提供了可推广的基础。

关键词：地下模型编程；大语言模型；油藏数值模拟；历史拟合；油田开发优化；机制推理

1 引言

油藏数值模拟是实现最优烃类采收率、地质碳捕集、利用与封存（CCUS） [1], 以及深层地热能开发[2], 的基础计算引擎，其核心是求解控制三维非均质地质网格中多相流体流动的非线性偏微分方程耦合系统 [3]。然而，在蓬勃发展的科学人工智能（AI for Science）领域，人工智能在地下工程中的当前应用主要仍局限于外围辅助工作流，例如训练数据驱动的替代模型或代理模型来模拟模拟输出[4, 5], 、开展基于文献的领域知识问答 [6], ，或编写用于自动化数据预处理与后处理的脚本 [7, 8]。相反，原始模拟数据卡本身完整封装了地质与物理现实——包括三维角点网格几何、空间

*通讯作者：Yanchen Song（sycindex@gmail.com）。

孔隙度和渗透率张量分布、流体相平衡与PVT性质，以及动态井生产计划——长期以来一直被视为大语言模型（LLM）无法解读的数值“黑箱”。由于缺乏原生解读和编辑这些底层物理文件的能力，历史拟合和油田开发方案规划等核心油藏管理任务仍严重依赖工程师凭经验反复试错，或依赖粒子群优化（PSO）[9, 10] 和集合卡尔曼滤波（EnKF） [11, 12]等不考虑物理规律的随机算法。这些传统方法需要进行数百甚至数千次计算成本高昂的正向模拟，缺乏机理直觉，且常常停滞于次优局部极值 [13]。

这一现状引出了一个基础性且前沿的科学问题：前沿大语言模型能否超越对话和辅助脚本生成的表面局限，像领域工程师和软件开发人员阅读源代码那样，原生解读原始模拟数据卡——直接解析三维空间坐标、拓扑网格连通性和多相流机制，自主执行高保真参数反演并制定最优开发决策？换言之，不依赖外部图形化前后处理软件，也不启动大规模随机搜索，大语言模型能否仅通过读取原始模型文件，推断地下流体的原位动态流动状态（例如，辨识井间砂体连通性、识别注水窜流条带 [14], 或诊断油藏能量衰竭），随后以程序化的精度重构数值模拟数据卡?

本研究通过严谨的物理模拟基准测试，首次证明前沿大语言模型具备原生读取与编辑动态储层模型的新兴能力，推动地下科学计算从盲目的黑箱试错迈向由物理认知驱动的“地下模型编程”。围绕这一突破，我们建立了一个四层分层评估基准，涵盖静态连通性解释、单步历史拟合、十年油田开发优化，以及五轮闭环三维组分属性重建。至关重要的是，所有模型生成的编辑结果均与高性能开源油藏数值模拟器相结合，进行端到端物理验证。具体而言，我们的主要发现与贡献有三点：

- **大语言模型在动态油藏模型上原生读写能力的首次实证展示：**我们证明，前沿大语言模型能够直接操控动态油藏模型，其方式类似于编写和重构软件代码。对于包含数万个网格单元的完整模拟数据卡，模型仅凭原始数据卡文件，通过数据卡上下文推理，便能直接完成三维井间连通性推断、单步参数反演和油田开发规划。对于拥有数百万个角点网格单元的工业规模资产，模型则借助轻量级诊断工具生成结构化补丁，直接就地编辑底层属性文件。这些发现揭示了“原生油藏模型编程”的雏形，展现出远超人类启发式预期和传统算法的推理效率，并为地下优化和物理可解释性带来变革性影响。

- **通过开源仿真引擎建立严格的物理验证协议：**大语言模型编辑或合成的每个数据卡，都会直接在开源油藏数值模拟器中运行，并以真实的多相生产拟合误差和十年累计采油量作为客观量化指标。该协议确保所有模型修改都严格遵循多相流的控制偏微分方程，从而提供无妥协、基于物理闭环的基准测试，评估模型真正的物理推理能力。

- **阐明机制推理的优势，以重塑地下工程工作流程：**通过剖析模型的推理轨迹和迭代参数调整，我们揭示了大语言模型如何利用其内化的储层物理先验——例如见水时间、射孔生产剖面和窜流通道识别——以显著少于传统随机优化算法的模拟运行次数，实现有针对性的单调收敛。这为自主油田开发规划、小样本历史拟合以及老旧油藏资产的白盒病态审计奠定了变革性的方法论基础。

2 方法与实验设置

2.1 油藏模拟文件、概念定义与交互模式

油藏数值模拟输入文件定义了地下资产的物理和作业状态，包括三维网格几何结构、岩石孔隙度和渗透率分布、流体性质、初始平衡状态以及井作业计划。紧凑型模型包含在单个文本文件中，而大型三维模型则会将大量网格坐标和岩石属性数组拆分到多个配套数据文件中。

基于这种表示方式，我们将地下模型编程定义为一种范式：它将油藏数值模拟文件视为可编程的计算对象。在这一范式下，地下流动行为（如三维网格连通性、流体驱替和压力衰竭）通过直接读取原始模型文件来解读；动态生产响应则通过直接编辑岩石属性场、井完井和操作计划来改变，完全独立于粒子群优化（PSO）或集合卡尔曼滤波（EnKF）等传统数值优化器。在这里，“程序员”是一个通用型大语言模型（LLM），它直接依据自然语言工程指令进行操作，无需数值梯度或伴随灵敏度。

为系统评估大语言模型在物理复杂度逐步增加的各项任务中的能力与局限，我们设计了四项递进式基准实验：静态井间连通性解读（实验01）、单步黑油历史拟合（实验 02）、多轮开发规划（实验 03）和多轮组分历史拟合（实验04）。尽管理论上，具有足够长上下文窗口的大语言模型可以读取任何储层模型的原始文本，但考虑到当前的上下文长度限制、计算效率，以及在庞大坐标数组上的编辑稳定性，我们采用两种互补的交互模式来实现这四项任务（图 1）：

- **模式 A — 全油藏文件直接交互：**对于由数千个网格单元组成的紧凑型单文件模型（实验 01–03），包含所有网格、岩石、流体、初始状态和井计划定义的整个模拟文件会直接放入大语言模型提示中。在静态连通性解释（实验01）和单步历史拟合（实验 02）中，大语言模型不使用任何外部工具，仅根据文本生成连通性判断或岩石属性修改。在多轮开发规划（实验 03）中，模型在五轮迭代中仅根据模拟器返回的生产曲线和饱和度/压力场，更新井位和压力计划。

- **模式 B——基于脚本的属性场编辑：**对于大型多文件组分模型（以 UNISIM-IV [15], 为例，该模型包含 19 个文件，采用角点网格，网格单元总数为 533,403 个，其中活跃网格单元有 72,267 个，并包含数百万个空间坐标和属性条目），只要上下文足够，理论上仍可摄取全文，但在当前上下文限制下会产生大量令牌开销，并降低空间编辑精度。在此情境（实验04）中，仅使用标准文件与命令行环境，且不借助预构建的油藏诊断软件，大语言模型会编写自己的 Python 脚本，根据模拟器输出分析三维网格连通性、水到达时间和射孔流量，然后编写有针对性的编辑脚本，在不使用任何外部优化器或参数扫描的情况下，更新三维渗透率和孔隙度文件。

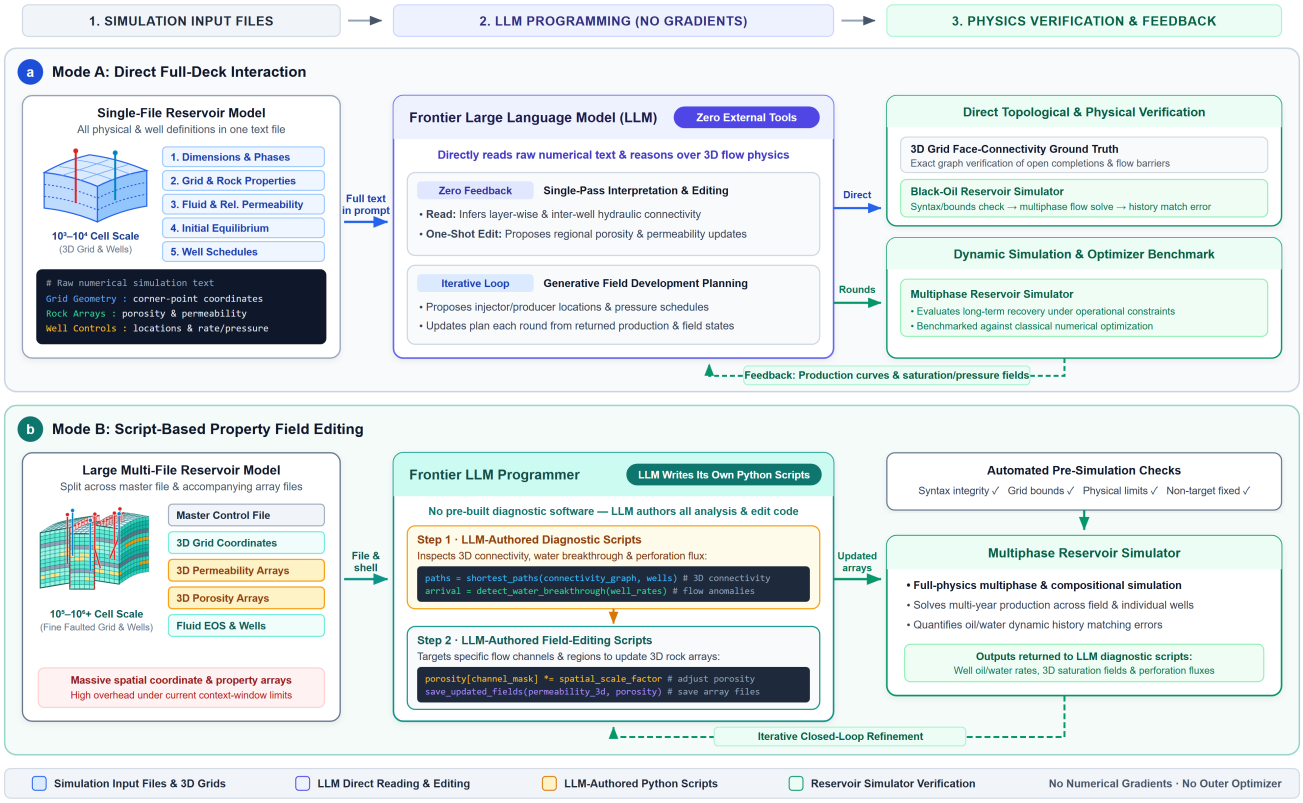

图 1：地下模型编程的概念工作流程，展示了全油藏文件直接交互（模式 A）以及与油藏数值模拟器耦合的自主脚本式属性场编辑（模式 B）。

2.2 模拟器、模型与评估设置

所有模型输出均使用开源油藏模拟器，直接根据多相流物理进行评估。为匹配不同模拟模型的关键字兼容性，各项任务会使用不同的互补模拟器：OPM Flow（v2026.04） [3] 在实验 02 中进行黑油历史拟合，而 JutulDarcy [16] 则在实验 03 中模拟十年开发规划，并在实验 04 中进行五轮组分历史拟合。对于实验 01中的静态解释，推断出的连通性会直接与基于相邻可渗透活跃网格单元和流动屏障得出的精确三维几何连通性进行核验。

不使用数值梯度或外部优化算法，我们根据任务需求选择了两种推理模型：旗舰模型 gpt-6-astra 负责零工具、全牌组任务（实验 01–03），而成本效益更高的 gpt-6.1-sol 则负责多轮、基于脚本的工作流程（实验04），以测试不同模型层级在此能力上的表现。每项拟议修改都会通过关键字语法、网格边界和物理参数限制的自动化检查，同时保持非目标设置不变。在实验 03 中，我们在相同物理约束下比较大语言模型规划 [10] 与粒子群优化（PSO） [17] 在 5 次和 25 次评估预算下的表现。表1 总结了全部四项基准测试中的油藏模型、网格尺度、受测大语言模型、交互模式和评估指标。

表 1： 四个储层模拟基准概述。

实验	储层模型与网格尺度	测试的大语言模型	交互 Mode	任务目标	模拟器与评估指标
实验01	合成三井背斜（19 × 28 × 3；总计 1,596 / 1,146 活跃网格）	gpt-6-astra	模式 A：直接全牌组（只读，零工具）	推断逐层以及井间连通性	无（静态几何检查）；连通性准确率超过相邻渗透单元
实验 02	PUNQ-S3 [13] (19 × 28 × 5；共 2,660 个 / 1,761 个活跃网格)	gpt-6-astra	模式 A：直接全牌组 (单轮编辑，零工具)	单步孔隙度和方向性渗透率更新	OPM Flow 2026.04 [3]; 油田/油井含水率以及油/水流量误差
实验 03	SPE10扇区模型 [10, 18](60 × 50 × 1；3,000 个活跃网格单元）	gpt-6-astra	模式 A：直接全牌组（5 轮次规划）	优化 4 口井井位和 5 个井底压力阶段历时 10 年	JutulDarcy [16]; 10 年累计石油产量产量与 5 次和 25 次评估粒子群优化 [17]
实验04	UNISIM-IV 组分模型 [15](47 × 39 × 291；533,403 角点网格 / 72,267 活跃网格，跨越 19 文件）	gpt-6.1-sol	模式 B：基于脚本的编辑（5 轮次）	迭代式三维渗透率和孔隙度场历史拟合	JutulDarcy [16]; 综合油田和井归一化产率误差（2,891天内）

3 实验结果

直接将大语言模型与多相流模拟器进行对比评估，结果表明，前沿推理模型能够解读原始储层模拟输入、诊断流动瓶颈，并以符合物理规律的精度修改三维岩石属性。下文将介绍静态连通性解读（实验01）、单步黑油历史拟合（实验 02）、十年开发规划（实验 03）和多轮组分场校准（实验04）的结果。

3.1 Zero-Tool3D 井间连通性解释（实验01）

仅提供单文件储层模拟原始输入，且没有任何外部可视化或网格解析工具，gpt-6-astra 在所有构造、流体、完井和井间连通性任务中均达到 100% 准确率（图 2 和表 2）。仅凭角点坐标和表格式岩石属性，该模型便准确重建了三层背斜穹隆、精确的网格尺寸、流体界面，以及所有高渗透率储层砂体的空间几何形态。

重要的是，该模型没有依赖简单的视觉或数值捷径——例如，假定同一层中的井必然连通，或将微量背景渗透率视为可导流。正如 W3–W1–W2构造剖面（图 2a）所示，上部K1层形成连续的斜向砂通道，中部K2层构成致密的封隔屏障，而下部K3层则分为两个独立砂体。模型正确识别出，零传导率边界屏障将这些砂体隔开，并封闭穿过K2层的垂向界面。尽管 W3井在上部和下部砂层均有射孔，但井筒内的窜流已被禁用，且其下部砂体与 W2隔离。因此，模型准确判断了井间连通性：只有 W1 和 W3 通过上部K1砂层连通；而 W2（仅在隔离的下部K3砂体中射孔）与 W1 和 W3 均保持水力不连通。所有定量流体界面、井位、运行速率和压力目标，以及365天计划也都无误地提取出来（图 2b）。

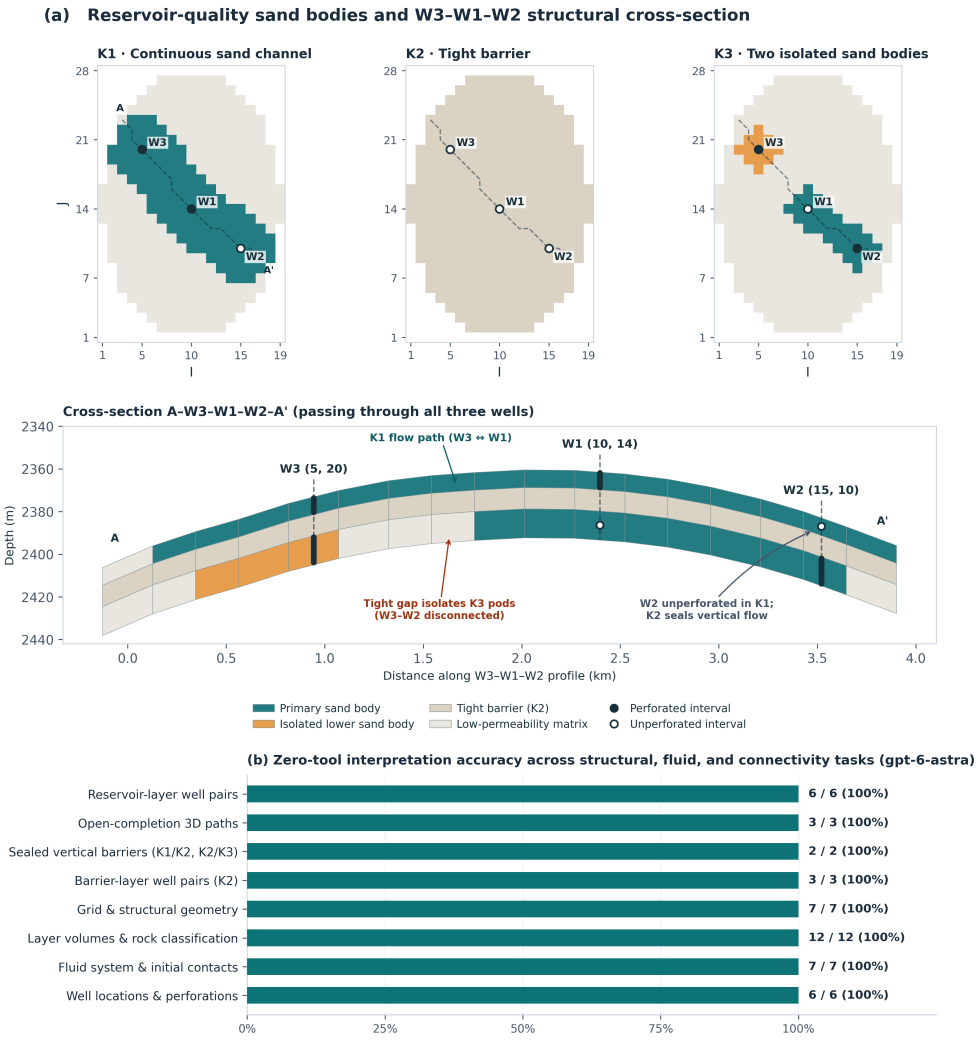

图2： 实验01中的静态三维井间连通性及零工具解读准确率。(a) K1–K3各层的平面砂体分布，以及穿过三口井的西北—东南向构造剖面 A–W3–W1–W2–A’；图中显示W3与W1之间K1层连通，K2层起垂向封隔作用，两个K3砂体彼此紧密分隔，并标示了已射孔（实心）与未射孔（空心）的完井。(b) gpt-6-astra在构造、流体、完井和连通性评估各组中的逐项准确率（所有任务均为100%）。

表2　： 无工具静态储层解释（实验01）项目级验证摘要　　.

验证组	正确数 / 总数	准确率 (%)	关键验证发现
储层层位井对（主要）	6 / 6	100.0	精确区分连续K1通道与分裂K3砂体
裸眼完井三维路径	3 / 3	100.0	W1–W3通过K1连通；W1–W2和W2–W3不连通
封闭垂向层隔挡层	2 / 2	100.0	识别出的封隔层分隔 K1/K2 和 K2/K3界面
隔挡层井对 (K2)	3 / 3	100.0	已标记致密中间层 K2 中的所有井对不连通
网格与构造几何	7 / 7	100.0	19 × 28 × 3网格（间距 180 米）；中央背斜顶部 W1 下方
层体积与岩石类型	12 / 12	100.0	储层与非储层砂体的精确数量以及各层厚度
流体系统与接触面	7 / 7	100.0	含溶解气的黑油；基准面 2,355 m（234.46 巴）；油水接触面 2,405 m
井位与射孔	6 / 6	100.0	以下井的准确井位及射孔层段 W1、W2 和 W3

3.2 PUNQ-S3 上的单步历史拟合（Exp02）

当被要求在单次零工具推理轮次中校准 PUNQ-S3 黑油模型时，gpt-6-astra 提出了 16 项局部属性修改，仅修改了 1,761 个活跃网格单元中的 54 个（3.1%）——更新了 44 个网格单元的孔隙度和 11 个网格单元的渗透率，其中有一个网格单元重叠，位于 PRO-4 底部完井段（图 3b）。

表3： PUNQ-S3（实验 02）单步编辑前后油田尺度及逐井历史拟合误差对比。

范围 / 井	含水率均方根误差（pp）之前 → 之后	石油产量均方根误差（sm^3/天）之前 → 之后	产水量均方根误差 (sm^3/天) 之前 → 之后	含水率相对变化
均衡平均值 (主要)	7.2333 → 2.0346	33.30 → 12.10	8.78 → 2.48	−71.87%
全油田总计	3.8737 → 1.2574	109.47 → 33.65	26.81 → 5.15	−67.54%（产水量 −80.80%）
PRO-4	16.5182 → 3.0210	20.90 → 3.82	20.90 → 3.82	−81.71%
PRO-12	6.6777 → 1.7604	71.03 → 31.12	5.81 → 2.66	−73.64%
PRO-20	5.7279 → 2.3093	5.73 → 2.31	5.73 → 2.31	−59.68%
PRO-11	3.4299 → 2.7037	4.34 → 3.42	4.34 → 3.42	−21.17%
PRO-5	1.9595 → 1.4960	31.75 → 1.89	4.17 → 1.89	−23.65%
PRO-15	0.6207 → 1.3672	34.91 → 4.55	0.48 → 0.95	+120.25%（石油产量均方根误差 −86.98%）
PRO-1	0.0105 → 0.0142	0.013 → 0.018	0.013 → 0.018	+35.44% (+0.0037 个百分点)

注意：井均衡含水率均方根误差是对 7 口生产井中 274 个生产井/层段配对取平均值；产量均方根误差涵盖全部 364 个配对层段。第16.5年时，油田累积产油量低估从 17.77%（287 万 sm^3）降至 5.47%（330 万 sm^3，观测值为 349 万 sm^3）。

在 OPM Flow 中对 16.5 年的生产历史（1967–1983 年）进行模拟后，这项单步调整降低了所有油田尺度误差以及经过平衡处理的井级历史拟合误差（图 3a 和表 3）。主要的平衡井含水率误差下降了 71.9%（均方根误差从 7.23 个百分点降至 2.03 个百分点），而全油田油水产量误差分别降低了 69.3% 和 80.8 %。因此，模拟的累积产油量大幅接近历史基准（3.30 对 3.49 百万 sm^3实测值），累积采收率低估从 17.8% 降至 5.5%。

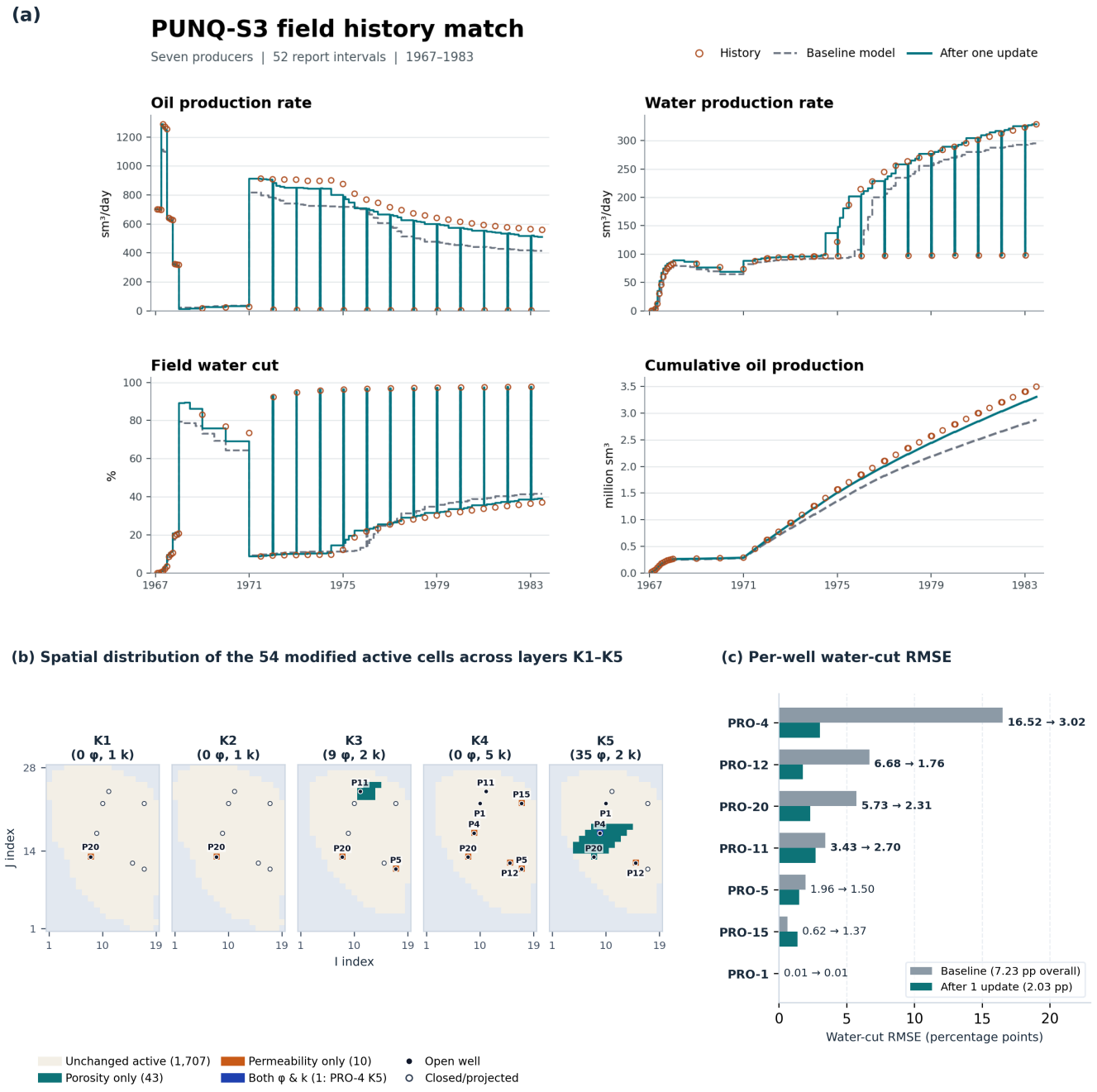

图 3： PUNQ-S3（实验 02）上的单步历史拟合性能与空间修改定位。(a) 在 OPM Flow（v2026.04）中，对比历史观测值、基线和单步候选方案的全油田石油产量、产水量、含水率及累积采油量。(b) 54 个已修改活跃网格相对于生产井位置的逐层分布（K1–K5）。(c) 全部 7 口生产井的单井含水率均方根误差降幅。

54个编辑单元的空间分布（图3b）及由此产生的井剖面（图3c和图4）揭示了三项明确且具有物理依据的调整。首先，为抑制PRO-4过早发生底水锥进，模型降低了底层通道区域的孔隙度，并降低了PRO-4完井段中过高的渗透率，同时提高上覆K4油层的水平渗透率，以维持石油流入，使PRO-4含水率误差降低81.7%（均方根误差从16.52降至3.02个百分点）。其次，为缓解东部和南部翼部石油产量偏低的问题，模型提高了PRO-5、PRO-12和PRO-15周围的近井水平渗透率，使三口井的石油产量均显著恢复。第三，在PRO-11周围局部降低孔隙度，并降低PRO-20井段附近的渗透率，从而遏制两口井过高的产水量。总体而言，7口生产井中有5口的含水率匹配得到改善；只有PRO-15以略高的含水率为代价，将石油产量误差降低了87.0%；而几乎不产水的PRO-1变化可以忽略不计（表3）。

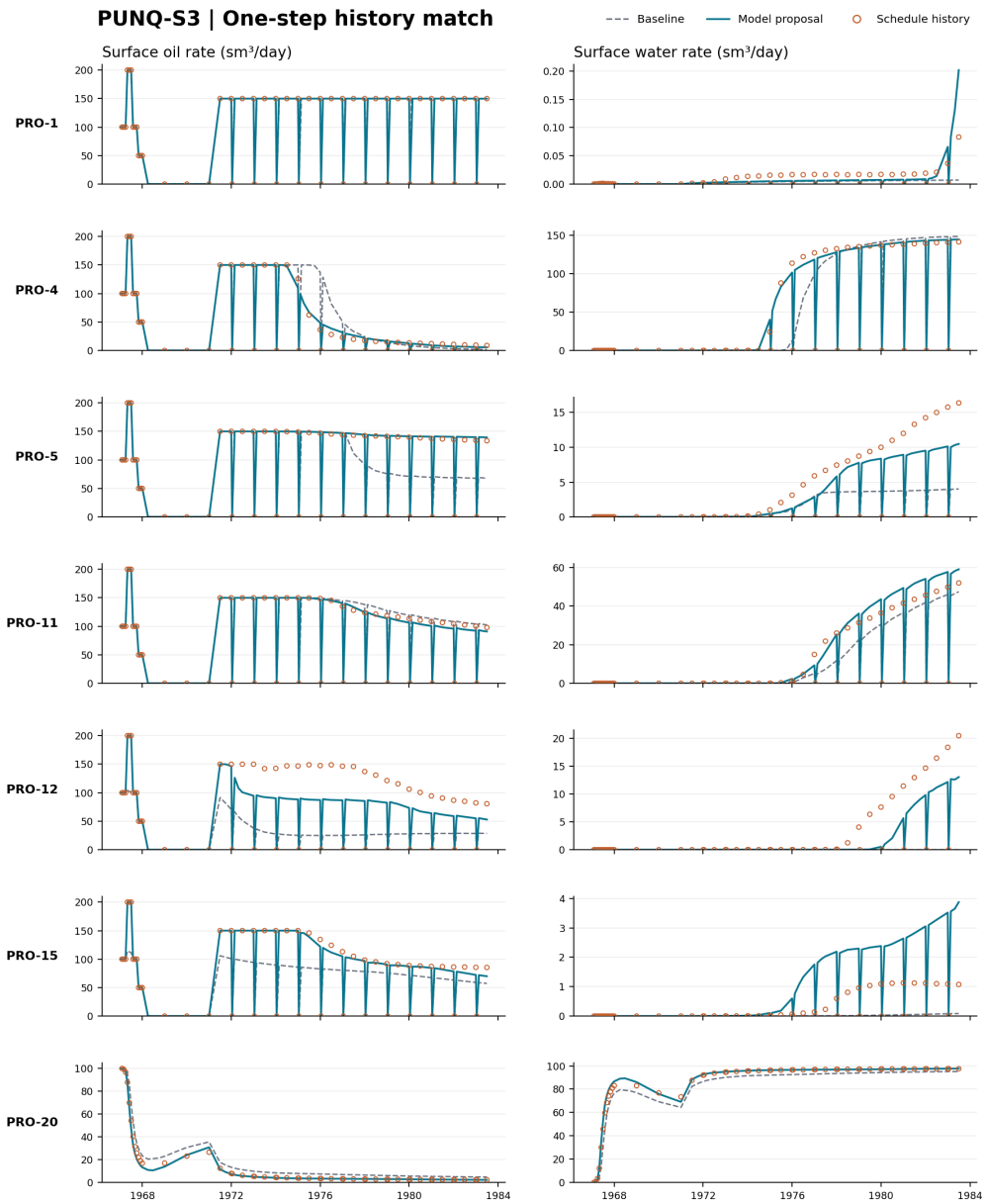

图4： PUNQ-S3（实验 02）上单步大语言模型编辑前后七口生产井各自的地面产油和产水曲线。PRO-5、PRO-12 和 PRO-15 的石油产量显著恢复，同时 PRO-4 的水突破得到显著校正。

3.3 十年开发规划与粒子群优化对比（实验 03）

在高度非均质的SPE10区域基准测试中，针对28个决策变量进行联合优化（两口注入井和两口生产井的平面位置，以及在10年时间范围内每口井分为5个阶段、每阶段持续2年的井底压力；经济性含水率关井上限为78%），gpt-6-astra仅通过5次油藏模拟便超越了同步粒子群优化（PSO）（图5a和表4）。其最佳方案（第3轮）实现了10年累积产油量179.5万sm^3，，超过了使用相同5次模拟预算的PSO最佳结果（155.8万sm^3，+15.2%）以及使用完整25次模拟预算的PSO搜索结果（164.9万sm^3，+8.9%），而所需模拟次数仅为后者的五分之一。

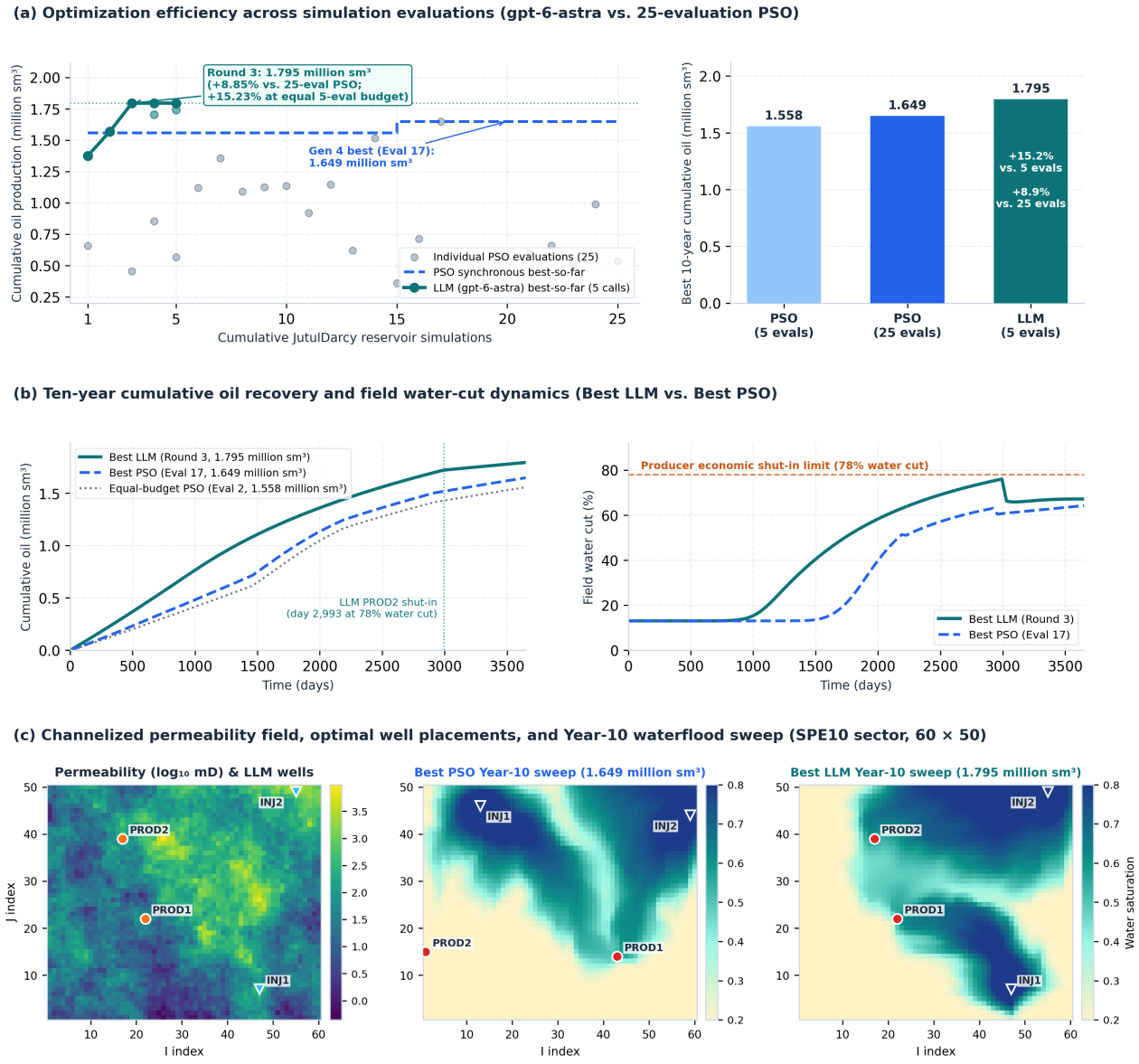

图 5： 通道化SPE10基准上的十年油田开发优化性能（实验 03）。(a) 不同模拟评估次数下的累积采油量轨迹，显示 gpt-6-astra 到第3轮时已超过 5 次和 25 次评估的粒子群优化基线。(b) 对比最佳大语言模型（第3轮）和最佳粒子群优化（第 17 次评估）的十年累积产油量及油田含水率历史。(c) 最佳粒子群优化与最佳大语言模型的对数渗透率通道结构和第 10 年含水饱和度扫掠图。

检查井位部署和最终饱和度场，可以解释为什么直接物理推理优于基于种群的搜索（图 5b–c）。在第1轮中，沿一条高渗透率条带过于密集地部署井，导致水窜迅速发生，迫使两口生产井在两年内停产（1.375 million sm^3）。在模拟器反馈的指导下，大语言模型将两口注入井分置于东部两侧相对的高渗透率角落，采用最大450巴的注入压力，并将两口生产井部署在中西部高渗透率通道内，采用最低100巴的井底压力——最大化压力降，并推动水驱沿东西向形成广泛的对角线波及（图 5c，右）。相比之下，粒子群优化的最佳方案将PROD2置于西部边缘的低渗透率区域附近（仅贡献了 ~10%的总石油），并在较低的平均油藏压力下留下大量未波及油（图 5c，中）。

表4： JutulDarcy中大语言模型gpt-6-astra（5次模拟）与粒子群优化（25次模拟）的十年开发规划方案汇总（实验 03）。

方法 / 轮次	模拟预算 / 评估	十年累积产油量（百万 sm^3）	相对于最佳 LLM	井位布局与驱替策略	关键流动行为 & 关井时机
大语言模型第1轮	1	1.375	−23.37%	聚集在高速通道	快速突破；在第 328.5 天关井 & 693.5
大语言模型第二轮	2	1.569	−12.58%	注入井向东分开；生产井移至西侧	消除了提前关井；超越了5次评估粒子群优化的最佳结果
大语言模型第3轮（最佳）	**3**	1.795	Best	东角注入井 + 中西部通道（450/100 巴）	宽幅对角扫描；PROD2 关井晚至第天 2,993
大语言模型 第4轮	4	1.704	−5.04%	PROD1 向东北移动朝向中央脊部	更早的水突破导致两口口生产井关闭
LLM 第5轮	5	1.740	−3.06%	恢复为 R3 布局采用 PROD2 分阶段压力递增	延迟 PROD2 关井直至第 3,613.5 天
粒子群优化最佳（5-评估预算，代 1)	5	1.558	−13.21%	北部注入井 + 南部/西部边缘生产井	5 次评估预算内的最佳结果（第 1 代均值：0.818 百万 sm^3）
粒子群优化最佳结果（25 次评估预算，第 4)	25	1.649	−8.13%	北部注入井 + 东部生产井；PROD2 位于狭窄的西侧边缘	PROD2 滞留（占总石油的 ~10%）；未波及边缘区

3.4 五轮次组分储层历史拟合（Exp04)

在包含 533,403 个网格单元（72,267 个活跃网格单元）的 UNISIM-IV 组分基准算例中，gpt-6.1-sol 在五轮迭代中编写并运行了自己的诊断脚本，以检查三维流动路径和水到达时间，然后再修改岩石属性场。在 2,891 天（ ~7.9 年）的生产历史期间，油田与单井综合归一化误差（归一化均方根误差）在每一轮已完成的迭代中都单调下降：从基线的 38.0% 降至第3轮的 20.8%、第4轮的 13.9% 和第5轮的 11.6%，总体相对降幅为 69.4%；油田级油水产量误差则下降了 73.9%（从 2,089 降至 545 m^3/d；图 6b 和表 5）。

这五轮并非盲目调参，而是构成了一个有条理、自我修正的诊断循环（图 6a、图 7和表 5）。在第1轮和第2轮中，运行至历史中期的部分模拟已揭示出明显的诊断趋势——表明没有必要尝试性地调整初始烃类组成，而大幅提高下部区域渗透率会导致 P15 过早发生水突破——因此两次运行都提前停止以节省计算资源，并在第3轮中回调这些调整，完成完整的 2,891 天历史模拟，综合归一化均方根误差达到 20.8%。相邻注水井 I12-W 投产后，P16 在历史中后期产水量激增，使第3轮的误差从早期略微高估转为后期严重低估；第4轮因此以沿离散的 I12-W–P16 通道增强导流能力，取代了先前的底水调整，归一化均方根误差降至 13.9%。第5轮则保持渗透率不变，降低该注采井通道沿线的孔隙度，以加快驱替前缘推进（图 9b），使 P16 的突破响应提高一倍，整体归一化均方根误差达到 11.6%。



图6： UNISIM-IV 组分历史拟合基准（实验04）中的多轮误差收敛。(a) 基线和第1至第3轮在共享的1,281天时间窗口内的产水量均方根误差，显示 P14 的误差迅速改善，并纠正了第2轮在 P15 过早产水的问题。(b) 已完成轮次中，全时段（2,891天）的综合油田/井归一化均方根误差（从37.98%降至11.62%）以及油田/P16产量均方根误差单调下降。(c) 从基线到第5轮，全历史单井误差降低。

表5： UNISIM-IV（实验04）中候选修改的多轮演进、模拟覆盖范围及全时段历史匹配误差。

轮次	物理修改策略	模拟时间窗	油田油/水产量均方根误差（m^3/天）	综合归一化均方根误差	诊断结果
基线	无（初始储层模型）	2,891天（516步）	2,089.43	37.9776%	突破延迟发生于P14–P16；油田水 NRMSE 51.23%
第1轮	近井渗透率，底水通道 & 探索性组分 test	停止日 1,912 *	—（1,281-天：155.82 对比 195.96）	—	早期P14改善 (218.39 → 170.77)；显示组成变化 不必要
第 2 轮	继承的R1渗透性恢复后的归一化组成；强劲P14–P16下部区域增产	停止日 1,446 *	—（1,281-天：125.67 对比 195.96）	—	P14误差降至62.77，但导致过早出现P15处见水（127.93）和P16
第3轮	缩减过度的P15/P16下部区域渗透率和孔隙度修改	2,891天（516步）	941.81	20.7914%	消除了P15过早出水水（13.98）；P16后期出水仍然低估
第4轮	恢复的 P16 底水编辑；增强离散 I12-W–P16 通道	2,891天（516步）	721.00	13.9150%	捕获的注入井–生产井扫掠路径；P16 全周期均方根误差 634.23
第5轮	固定渗透率；仅降低孔隙度沿 I12-W–P16路径（1,120 个网格单元）	2,891天（516步）	544.72	11.6202%	最佳模型（−69.40% 归一化均方根误差）；加速驱替前缘至 P16（均方根误差 446.06）

*第1轮和第2轮在部分时间窗诊断明确物理趋势后，均无求解器错误地顺利停止；在历史液量控制下，以 m^3/d 为单位的全油田油水产量均方根误差相等。

表6： UNISIM-IV（实验04）中油田尺度及生产井层面的全时段（2,891天）历史拟合性能。

井 / 范围	油/水产量均方根误差 (m^3/天) 基线 → 第5轮	产量误差变化	含水率均方根误差 (pp) 基线 → 第5轮
全油田总计	2,089.43 → 544.72	−73.9%	—
P11	12.24 → 3.75	−69.4%	0.23 → 0.07
P12	43.49 → 9.48	−78.2%	0.87 → 0.20
P13	13.58 → 5.81	−57.2%	0.29 → 0.12
P14	918.63 → 277.92	−69.7%	20.16 → 6.11
P15	331.17 → 150.53	−54.5%	6.36 → 2.79
P16	1,046.45 → 446.06	−57.4%	23.98 → 10.19

注：油田与井的归一化均方根误差降低了 69.4%（从 37.98% 降至 11.62%），活跃样本的井含水率均方根误差从 12.90 降至 4.91 个百分点。在含水率达到 1% 时，第5轮中 P14–P16 的突破时间分别为第 1,101、1,271 和 1,465 天，而观测到的时间分别为第 1,101、1,240 和 1,358 天（基线则延迟至第 2,126、1,674 和 1,747 天）。

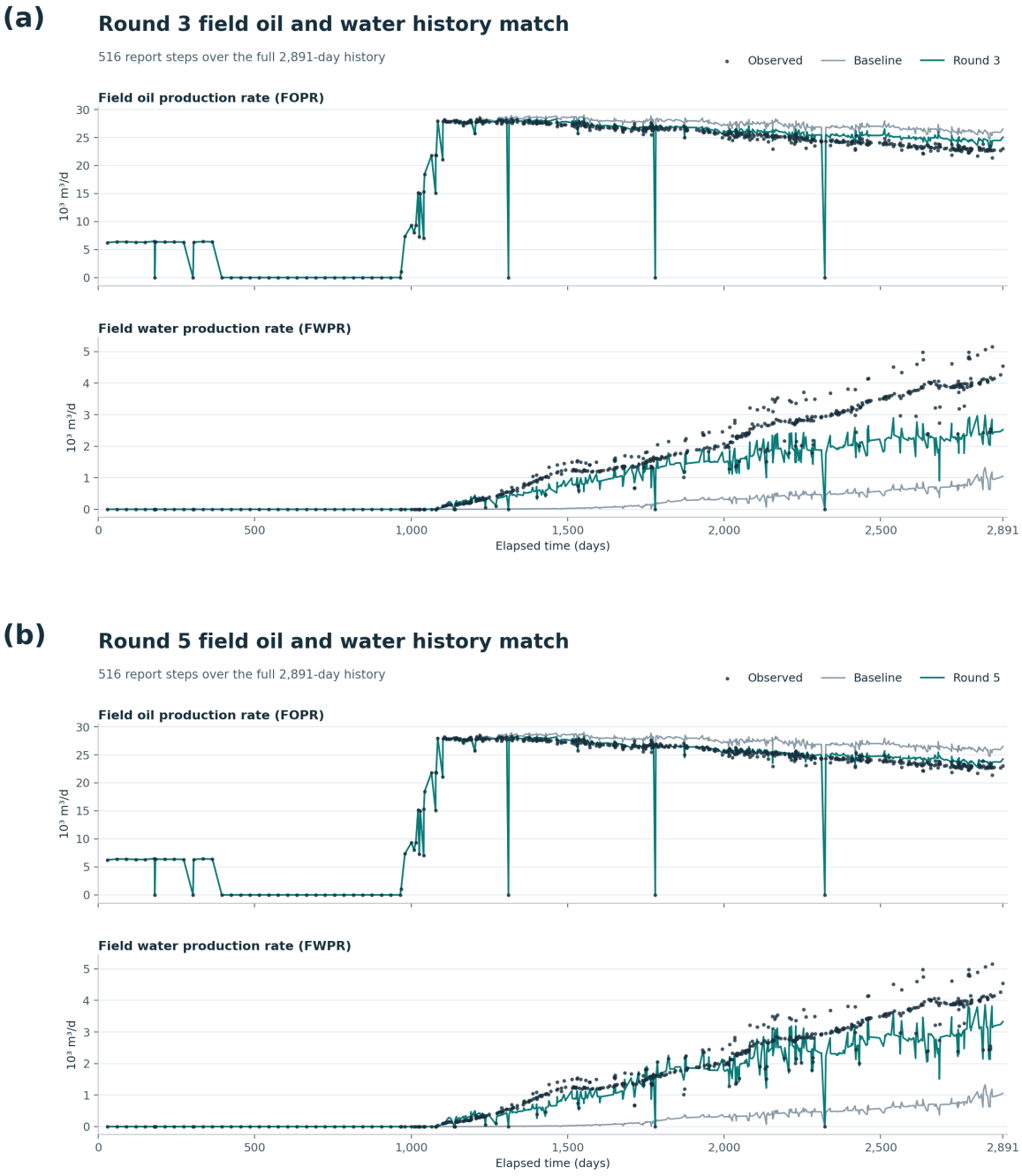

图7： 实验04中全时段（2,891天，516步）全油田油水产量历史拟合，对比（a）第3轮（综合归一化均方根误差 20.79%）和（b）第5轮（综合归一化均方根误差 11.62%）与观测历史及基准模型。

到第5轮时，低含水率井（P11–P13）与油田历史数据高度吻合（< 10 m^3/d 产量误差和 ≤ 0.2 个百分点的含水率误差），而高含水率生产井（P14–P16）的产量误差降低了55–70%，且水突破时间准确（图8 和表6）。独立审计还确认，每个网格单元中的组分摩尔分数均严格满足和为一归一化约束（图9a）。剩余差异主要集中在P16和P14的后期产水量上，这反映了五轮迭代限制，以及在固定地质分层条件下全油田反演的不唯一性。

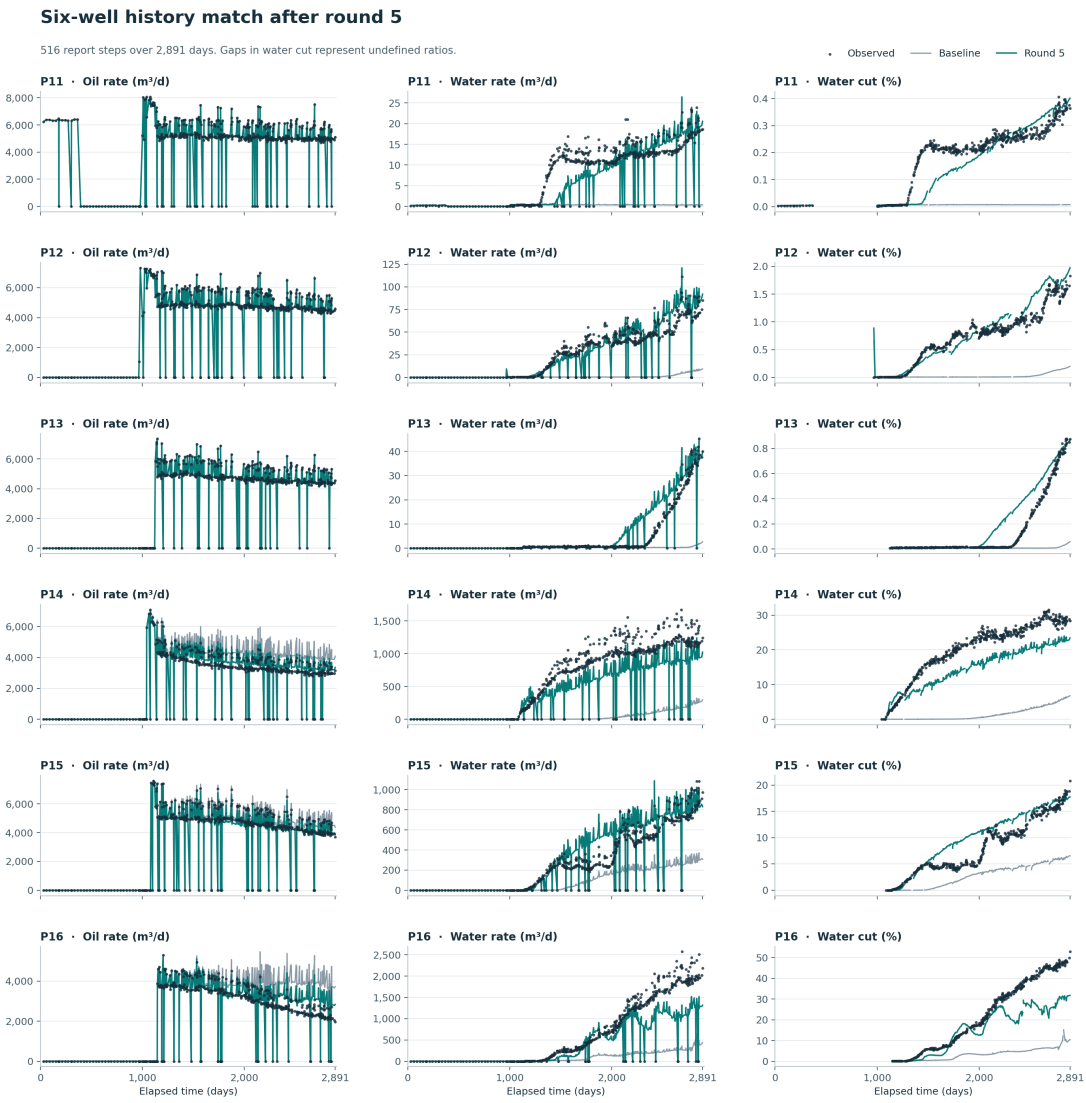

图8： UNISIM-IV（实验04）中六口单井生产井（P11–P16）的石油产量、产水量及含水率历史拟合在轮次5后历经2,891天的表现。

4 讨论：油藏工程中的机制性洞见与范式转变

在本研究中，我们展示了前沿大语言模型（大语言模型）原生读取和编辑动态储层模拟模型的新兴能力，并由此确立了“地下模型编程”。在我们的全部实验中，推理轨迹表明，大语言模型不会将离散的储层网格单元和模型条目视为彼此孤立的数值符号；相反，它们构建了连贯的三维物理直觉——区分封闭屏障与连续通道，权衡储层压力支撑与窜流风险，并将流动传导能力与孔隙体积波及动态分离开来。同样重要的是，与仅提供总体误差指标相比，提供清晰的物理反馈——例如油井见水时间和流体产量变化——能够在高维参数空间中实现更有针对性的调整。尽管海量油藏文件的上下文容量以及迭代数值模拟的计算运行时间仍存在实际瓶颈，但这些能力预示着地下可解释性和数值优化将发生深刻转变，并指向下文所述的四项主要未来应用。

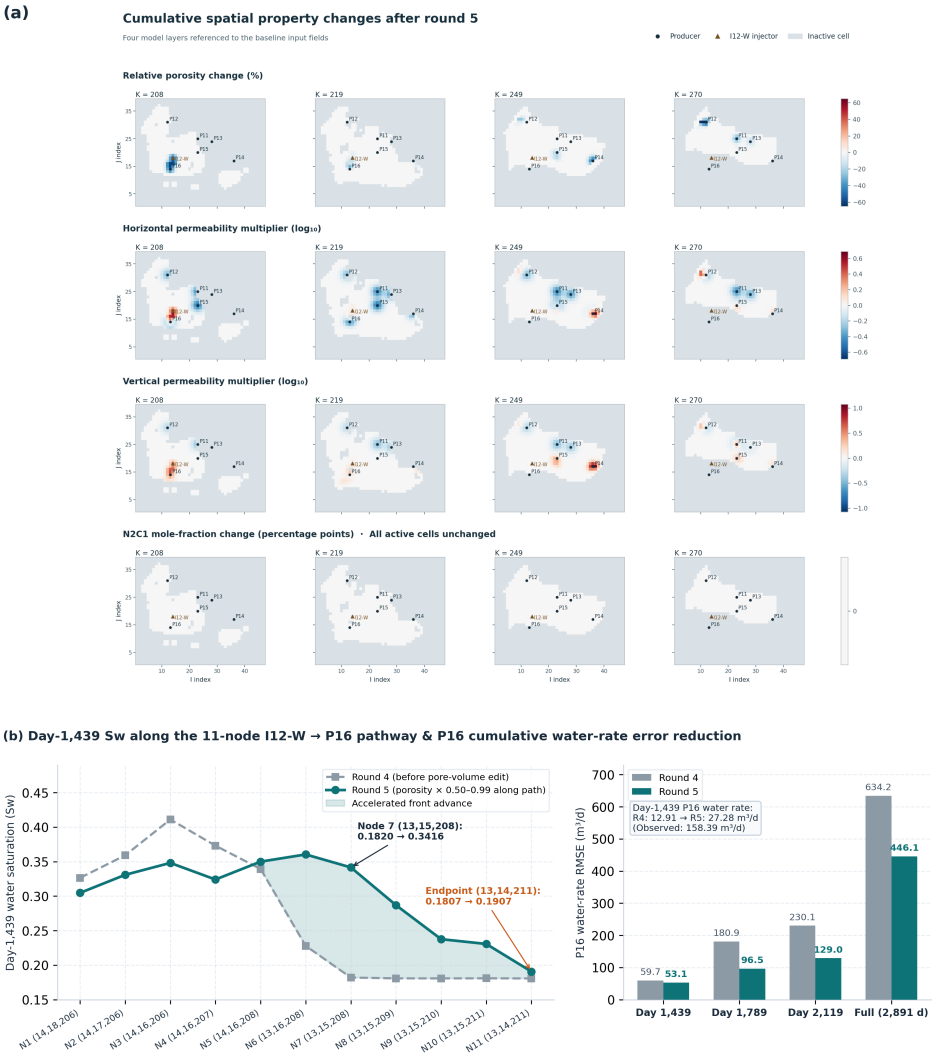

图9： 实验04中多轮油田编辑的空间定位与机理验证。（a）相对于基线，四个代表性层中孔隙度、水平渗透率、垂向渗透率和归一化初始轻组分摩尔分数的累积变化。（b）仅由第5轮的孔隙度降低所触发的、沿11节点离散I12-W至P16路径推进的第1,439天含水饱和度推进，以及P16在各诊断窗口中的产水量均方根误差逐步降低。

4.1 Gene开发方案的迭代综合与物理协同优化 s

传统油气开发规划涵盖井网、井型、驱替策略和分阶段控制，但仍受正向模拟成本高以及人工前处理和后处理障碍的限制。在日常工程实践中，团队往往依靠局部经验，仅提出少数几个候选布局，再被动地进行模拟验证，导致大量设计空间未得到充分探索，也难以预判强非均质储层中复杂的井间干扰 [9, 10]。通过原生解析三维构造高部位、渗透率结构和剩余油饱和度，大语言模型能够内化核心工程先验，例如使注采井网与主流线对齐，并控制压降压力以抑制早期窜流，从而高效合成符合物理规律的开发方案。结合高速油藏数值模拟器，这项能力构建起“认知合成—快速模拟—物理反馈”的闭环优化引擎，推动油田开发规划从经验性试错对比转向物理引导的生成式探索。

4.2 物理引导的少样本历史拟合与地质合理反演

主流的集合式和进化式反演方法——例如集合卡尔曼滤波（EnKF）、平滑器变体和启发式群体搜索——通常需要$O(10^2)$–$O(10^3)$ 次正向模拟，并采用分散的参数调整，往往会产生扭曲地质连续性的非物理“过拟合噪声” [9–13]。原生模型编程使大语言模型成为具备因果推理能力和先验知识的反演引擎：它能将异常的水突破时间和产水量差异直接归因于特定的高渗透率窜流通道或底水锥进区（实验 02 和实验04）。这种机理诊断将校准模拟预算从$O(10^3)$ 压缩至$O(10^1)$ 量级，同时保留更新后储层模型的地质连续性和物理合理性。

4.3 遗留储层模拟数据卡的自动化白盒审计与知识恢复

石油工业数十年来积累了大量遗留储层模拟输入文件，这些文件历经多个团队依次编辑。许多模拟数据卡包含未记录的“历史补丁”——例如，为拟合过去的压力而强行设置任意局部传导率乘数、不符合物理规律的垂向渗透率覆盖值，或相互冲突的井控开关——这些设置会在不知不觉中损害预测的可靠性。大语言模型能够原生读取底层模拟数据卡，因此可充当自动化白盒审计器，直接扫描原始文件，以标记隐藏的物理违规、诊断数值收敛停滞、恢复未记录的修改意图，并生成整洁且通过守恒验证的修复补丁。

4.4 面向碳捕集、利用与封存、地热能及多物理场能源系统的通用原生模拟接口

在新兴的能源转型工作流中，包括碳捕集、利用与封存中的长期 CO_2 羽流追踪和盖层完整性审计[1], 、深层地热储层中的热突破管理 [2], 以及地下储氢，建模通常需要在黑油、组分、热采和反应运移模拟器之间耦合多相流、地球化学、热传输和地质力学 [1–3, 16] ，而这些模拟器的输入语法各不相同且僵化。由于地下模型编程直接作用于底层物理输入文件，无需依赖脆弱的、特定于软件的 API 封装，前沿大语言模型得以建立一个通用的、与语法无关的接口，实现能源转型领域跨物理场工作流的无缝编排和参数同化。

5 结论

我们证明，前沿大语言模型（大语言模型）能够原生读取和编辑动态油藏模拟模型，从而确立了“地下模型编程”这一以物理学为基础的计算范式。我们在四项复杂度逐步提升的基准测试中，直接使用开源多相模拟器进行评估：大语言模型推理在零工具条件下实现了三维连通性解释的 100% 准确率（实验01）；通过单次零工具编辑，仅修改 3.1% 的活跃网格，将 PUNQ-S3 井间平衡含水率误差降低了 71.9%（实验 02）；在 SPE10 开发规划任务中，仅用五次模拟便胜过了经过 25 次评估的粒子群优化（+8.9%）（实验 03）；在包含 72,267 个活跃网格的 UNISIM-IV 组分模型上，经过五轮优化，将油田与井的综合归一化误差降低了 69.4%（实验04）。

通过直接基于底层模型物理原理进行推理，而非依赖盲目的试错，Subsurface Model Programming 将人工智能推理与基于物理的油藏模拟连接起来。尽管扩展长上下文容量、提高模拟器运行效率以及联合处理地质不确定性仍是重要的前沿课题，但这一范式为自主开发规划、少样本历史拟合、对遗留模拟资产进行白盒审计，以及更广泛的能源转型应用奠定了灵活而坚实的基础，其中包括 CO_2 地质封存（碳捕集、利用与封存）和深层地热系统。

参考文献与注释

[1] H. Class、A. Ebigbo、R. Helmig、H. K. Dahle、J. M. Nordbotten、M. A. Celia、P. Audigane、M. Darcis、J. Ennis-King、Y. Fan、B. Flemisch、S. E. Gasda、M. Jin、S. Krug、D. Labregere、A. Naderi Beni、R. J. Pawar、A. Sbai、S. G. Thomas、L. Trenty 和 L. Wei。关于地质构造中 CO_2 封存相关问题的基准研究。《计算地球科学》，13(4):409–434，2009。doi:10.1007/s10596-009-9146-x。[2] K. Pruess。TOUGH代码——用于可渗透介质中多相流和运移过程的一系列模拟工具。《包气带期刊》，3(3):738–746，2004。doi:10.2136/vzj2004.0738。[3] A. F. Rasmussen、T. H. Sandve、K. Bao、A. Lauser、J. Hove、B. Skaflestad、R. Klöfkorn、M. Blatt、A. B. Rustad、O. Sævareid、K.-A. Lie 和 A. Thune。Open Porous Media Flow 油藏模拟器。《计算机与应用数学》，81:159–185，2021。doi:10.1016/j.camwa.2020.05.014。[4] Z. Zhong、A. Y. Sun、B. Ren 和 Y. Wang。一种用于不确定条件下油藏产量预测的基于深度学习的方法。《SPE期刊》，26(3):1314–1340，2021。doi:10.2118/205000-PA。[5] Z. Liu 和 F. Zhou。借助经模拟训练的代理模型，大语言模型智能体从分层知识图谱中检索多目标油藏开发方案。《Scientific Reports》，2026。doi:10.1038/s41598-026-78530-1。[6] H. Pan、J. Liu、B. Gong、Y. Zhu、J. Bai、H. Huang、Z. Fang、H. Jing、C. Liu、T. Kuang、Y. Lan、T. Wang、T. Xie、M. Cheng、B. Qin 和 Y. Shen。大语言模型在油藏动态分析中的构建与初步应用。《石油勘探与开发》，51(5):1357–1366，2024。doi:10.1016/S1876-3804(25)60546-5。[7] I. Matejka、W. Laube 和 D. Perschke。用于自主生成和执行油藏模拟数据文件的对话式用户界面。载于：SPE年度技术会议与展览，美国休斯敦，2026。doi:10.2118/234136-MS。SPE-234136-MS。[8] L. Samigullin、S. Shumilin 和 E. Burnaev。油藏历史拟合的多智能体方法。arXiv预印本 arXiv:2605.15028，2026。doi:10.48550/arXiv.2605.15028。[9] J. E. Onwunalu 和 L. J. Durlofsky。应用粒子群优化算法确定最优井位和井型。《计算地球科学》，14(1):183–198，2010。doi:10.1007/s10596-009-9142-1。[10] T. D. Humphries、R. D. Haynes 和 L. A. James。井位部署与控制联合优化的同步和序贯方法。《计算地球科学》，18(3–4):433–448，2014。doi:10.1007/s10596-013-9375-x。[11] G. Nævdal、L. M. Johnsen、S. I. Aanonsen 和 E. H. Vefring。利用集合卡尔曼滤波进行油藏监测与持续模型更新。《SPE期刊》，10(1):66–74，2005。doi:10.2118/84372-PA。[12] D. S. Oliver 和 Y. Chen。油藏历史拟合的最新进展：综述。《计算地球科学》，15(1):185–221，2011。doi:10.1007/s10596-010-9194-2。[13] F. J. T. Floris、M. D. Bush、M. Cuypers、F. Roggero 和 A.-R. Syversveen。产量预测不确定性的量化方法：比较研究。《石油地质科学》，7(Suppl.):S87–S96，2001。doi:10.1144/petgeo.7.S.S87。[14] D. Li、J. Yang 和 D. Lu。基于瞬态压力分析的窜流层识别：现场案例研究。《石油勘探与生产技术期刊》，6(1):63–72，2016。doi:10.1007/s13202-015-0168-8。[15] V. E. Botechia、M. G. Correia、V. S. Rios、A. Davolio、S. M. G. Santos、J. C. V. Hohendorff Filho 和 D. J. Schiozer。UNISIM-IV：高 CO_2 含量轻质油碳酸盐岩油藏的基准方案。《巴西石油与天然气期刊》，16(4):149–160，2022。doi:10.5419/bjpg2022-0012。[16] O. Møyner。JutulDarcy.jl——一种基于自动微分、完全可微的高性能油藏模拟器。《计算地球科学》，29(4):30，2025。doi:10.1007/s10596-025-10366-6。[17] M. Clerc 和 J. Kennedy。粒子群——多维复空间中的爆炸、稳定性与收敛性。《IEEE进化计算汇刊》，6(1):58–73，2002。doi:10.1109/4235.985692。[18] M. A. Christie 和 M. J. Blunt。第十届SPE对比求解项目：上尺度技术比较。《SPE油藏评价与工程》，4(4):308–317，2001。doi:10.2118/72469-PA。