



基于深度学习的地层沉积正演模拟代理模型构建与应用

刘彦锋, 段太忠, 龚伟, 廉培庆, 张文彪, 黄渊

引用本文:

刘彦锋, 段太忠, 龚伟, 廉培庆, 张文彪, 黄渊. 基于深度学习的地层沉积正演模拟代理模型构建与应用[J]. 沉积学报, 2023, 41(3): 791-803.

LIU YanFeng, DUAN TaiZhong, GONG Wei, LIAN PeiQing, ZHANG WenBiao, HUANG Yuan. Construction and Application of a Proxy Model for Stratigraphic Forward Modeling Based on Deep Learning[J]. *Acta Sedimentologica Sinica*, 2023, 41(3): 791-803.

相似文章推荐 (请使用火狐或IE浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

构造掀斜主导的断陷湖盆缓坡层序“源—汇”正演模拟定量研究

Source-to-sink Quantitative Stratigraphic Forward Modeling on the Tilted Hanging-wall Sequence Architecture of a Tectonically-driven Lacustrine Rift Basin

沉积学报. 2022, 40(6): 1481-1493 <https://doi.org/10.14027/j.issn.1000-0550.2022.061>

碎屑锆石UPb年代学定量物源分析的基本原理与影响因素

Principles and Biases of Quantitative Provenance Analysis Using Detrital Zircon U-Pb Geochronology: Insight from modern river sands

沉积学报. 2022, 40(6): 1599-1614 <https://doi.org/10.14027/j.issn.1000-0550.2022.099>

苏北东台坳陷阜宁组沉积演化数模物源分析

Provenance System of Funing Formation in Dongtai Depression, Subei Ba-sin: Insight from 3D stratigraphic forward modeling

沉积学报. 2018, 36(4): 722-730 <https://doi.org/10.14027/j.issn.1000-0550.2018.060>

塔里木盆地顺南地区中—下奥陶统正演沉积数值模拟

A Three-dimensional Numerical Sedimentary Simulation Model for the Lower-Middle Ordovician of Shunnan Area, Tarim Basin

沉积学报. 2017, 35(1): 117-123 <https://doi.org/10.14027/j.cnki.cjxb.2017.01.012>

中国沙漠物源研究:回顾与展望

Provenance Studies of Chinese Deserts: Review and Outlook

沉积学报. 2015, 33(6): 1063-1073 <https://doi.org/10.14027/j.cnki.cjxb.2015.06.001>

文章编号: 1000-0550(2023)03-0791-13

DOI: 10.14027/j.issn.1000-0550.2021.152

基于深度学习的地质沉积正演模拟代理模型构建与应用

刘彦锋¹, 段太忠¹, 龚伟², 廉培庆¹, 张文彪¹, 黄渊¹

1. 中国石化石油勘探开发研究院, 北京 102206

2. 北京师范大学, 北京 100091

摘要 地质沉积正演模拟方法能更真实地刻画地下地质体的分布规律, 比传统的基于地质统计学的建模更有优势, 但是条件化难度大, 使其应用在实际油藏地质建模时面临较大挑战。地质沉积反演模拟提高了该方法的实用性, 基于模拟结果与观测数据定量比较, 地质沉积反演模拟在全局优化算法的驱动下不断修正地质沉积正演模拟输入参数, 使模拟结果与观测数据吻合度达到最佳。由于反演系统优化参数多, 非线性极强, 收敛难度大, 需要的迭代次数多, 单次正演模拟耗时长, 导致该方法效率较低。地质沉积反演模拟和深度学习算法中的生成对抗网络相结合, 提出了构建地质沉积正演模拟代理模型的方法。以大量的碳酸盐岩地质沉积模拟的合成模型作为样本, 通过神经网络训练, 形成地质沉积正演模拟器的代理模型, 然后将其代入地质沉积反演模拟系统, 避免了直接地质沉积正演模拟的长耗时, 加快了反演模拟的收敛速度。通过巴哈马西缘碳酸盐岩地质沉积模拟验证了该方法的可行性, 采用学习后的生成网络进行沉积反演模拟, 反演效率得到了大幅提升。尽管本文展示的是二维实例, 也有望扩展应用在三维模型上。

关键词 生成对抗神经网络; 代理模型; 深度学习; 地质沉积过程正演模拟

第一作者简介 刘彦锋, 男, 1986年出生, 博士, 副研究员, 开发地质学, E-mail: liuyf.syky@sinopec.com

中图分类号 P628⁺.3 **文献标志码** A

0 引言

地质沉积过程正演模拟比传统的基于统计学的建模能更真实地刻画地下地质体的分布规律, 但条件化方面仍需改善。地质沉积过程正演模拟以质量守恒为基础, 考虑了沉积物的供给、产生、搬运、剥蚀和再沉积等地质作用, 模拟结果符合地质规律^[1]。正演模拟的输入参数有初始地形、海平面曲线、沉积物搬运系数、沉积物供给曲线、基底沉降曲线、碳酸盐岩生产的相关系数, 以及这些系数随时间和空间的变化等^[2]。由于这些输入参数很难通过现有的观测手段获得, 使得模型存在较大不确定性, 与井数据和地震解释数据吻合难度大, 导致该建模技术在实际油藏地质建模中应用推广缓慢。

地质沉积反演模拟技术提高了地质沉积过程模拟的实用性。Lessenger *et al.*^[3], Cross *et al.*^[4]提出了地质沉积反演模拟的基本框架, 认为地质沉积反演模

拟包括地质沉积正演、模拟结果和观测数据之间的比较, 以及自动调整正演参数的最优化算法。基本思路是: 以有限的、能反映地层分布特征的观测数据, 比如井上钻遇的沉积微相、地震层位解释等, 反推地质沉积正演模拟的输入参数, 实现模拟结果与观测数据吻合, 然后进行地层空间展布的预测。地质沉积模拟的基本过程是从一组初始的过程参数开始执行正演模拟, 将模拟结果与观测数据比较, 优化算法按照一定的策略调整过程参数, 再次执行正演模拟, 直到模拟结果与观测数据的吻合度达到阈值。前人在这些方面做了很多研究, 涵盖正演模拟技术、比较技术和优化算法等^[5-11]。

在反演系统中, 通常正演模拟的计算量很大, 是最耗时的计算部分。代理模型是工程问题中常用的一个优化方法。当实际问题计算量很大、不容易求解时, 可以使用计算量较小、求解速度快的代理模型替代原模型, 加速优化过程, 比如多项式法、二次响

收稿日期: 2021-09-10; 修回日期: 2021-11-12; 录用日期: 2021-12-22; 网络出版日期: 2021-12-22

基金项目: 中国石化科技部项目(P21038-3, P20077kxjgz); 中科院先导A类项目(XDA14010204) [Foundation: SINOPEC Science and Technology Department, No. P21038-3, P20077kxjgz; CAS Pilot Class A, No. XDA14010204]

应曲面和克里金方法等。尽管代理模型是复杂系统反演的常见处理方式^[12-13],但针对沉积模拟代理模型的相关研究未见报道。

与油气藏渗流、地下水流动、数值天气预报等模型类似,地层沉积反演模拟也是一个复杂的非线性系统,本质上求解这些复杂系统都是大规模高度非线性问题的优化问题,需要借助高效的优化算法反复进行迭代求解,其中单次正演模拟求解耗时长、非线性强是影响问题快速求解的主要瓶颈。对于这类问题,代理模型是用另外一种更快的数学方法产生与原模型误差最小的结果,以达到快速迭代。目前,已经有一些替代正演模拟的方法,这些模型替代非线性较弱的模型时效果较好,对复杂性强的模型仍无法替代^[14],深度神经网络具有很强的模式表达能力,为代理模型构建提供了新的途径,也为快速的地层沉积反演模拟提供了可能。

最近几年发展迅速的生成对抗神经网络方法具有很强的模型生成能力,已经用于模拟复杂地质模型的生成和油气藏数值模拟中替代模型的建立。神经网络主要分为判别网络和生成网络,判别网络建立高维参数向低维参数的映射,常用于分类、聚类和判别问题等;生成网络建立低维参数向高维参数的映射,常用于回归和模型生成问题。复杂三维地质建模可以通过生成网络实现。常见的生成网络有贝叶斯网络、变分自编码器、玻尔兹曼机、生成对抗网络等,其中生成对抗网络的应用最广泛。

相比于其他方法,生成对抗网络具有更强的模式生成能力,可以生成更逼真的模型,但存在样本需求量的问题,不适合样本少的情况。对于地下地质类问题,难以获取真实的三维模型作为样本,通过正演模拟方法合成样本是深度学习技术在该领域落地应用的主要方式。在沉积反演模拟系统中,样本通过正演模拟器生成,不存在样本不足的问题。使用生成对抗网络,有望提高反演的收敛速度,解决目前的效率问题。

本文在地层沉积反演模拟基本框架、生成对抗网络基本原理介绍的基础上,提出了基于生成对抗网络的地层沉积代理模型构建方法,并通过实例研究说明该方法的可行性,最后讨论了该方法存在的局限性和未来发展的方向。

1 生成对抗网络代理模型构建方法

1.1 地层沉积反演模拟基本框架

地层沉积反演模拟包括生成三维模型的正演模拟器、模拟结果与观测数据比较的误差计算器,调整正演模拟输入参数的全局优化器。优化器是整个系统的核心驱动力,它根据误差计算器的响应不断调整正演模拟输入的参数,使模拟结果与观测数据吻合度最高。

地层正演模型的输入参数考虑了边界条件和初始条件,它们描述了可容空间、沉积物剥蚀、供给、生产、搬运、堆积以及压实作用等。输出结果是模拟的地层和一系列的古环境条件,比如整个演化历程的古地貌、沉积间断等。区别于碎屑岩地层,碳酸盐岩沉积体系的形成、演化和消亡,除了受构造活动、海平面变化、气候条件、海洋环境和水动力条件等多种作用于碎屑岩沉积体系的共同因素的直接或间接的控制外,还受体系中生物与生态因素的重要控制^[15-17]。此次采用了自主研发的碳酸盐岩地层沉积正演模拟方法,它考虑了最新的碳酸盐岩工厂、生态可容空间和层序地层学等基本原理,建立了基于环境能量和生物碳酸盐产能的地层沉积正演模拟方法和软件系统,可以模拟台缘带、缓坡、孤立台地以及复杂地形下的碳酸盐岩沉积体系等^[18],限于篇幅问题,具体的正演模拟方法不再展开。

模拟结果与观测数据的比较方法依赖于观测数据的多少及类型,往往是较少的观测资料,比如少量的几口钻井资料、分辨率较低的地震解释资料等。为了降低地震资料引起的不确定性,本文选择井上的解释的沉积相与实际模拟的过井处沉积相的均方根误差作为反演的目标函数。对于复杂非线性且不可求导的非线性系统,需要采用启发式优化算法,比如模拟退火、差异演化、粒子群算法等,通过多参数下的优化算法收敛性能对比,基于数值梯度的复合多重形优化算法(SCE-UA)在多参数反演方面具备更多优势^[19]。

由于整个地层沉积反演模拟系统非线性极强,收敛难度大,需要的迭代次数很多,导致效率较低。提出了基于生成对抗网络沉积模拟代理模型构建方法,替代其中的正演模拟过程,充分利用神经网络全程可微分的性质,提高整个反演系统的收敛性能。

1.2 生成对抗神经网络方法

生成对抗神经网络为地质建模提供了新方法。随着深度学习等人工智能技术的发展,地质建模技术也朝着智能化方向快速发展。由于生成神经网络与地质建模之间的高度相似性,这项技术在地质建模领域快速落地应用。在多层神经网络框架下,地质建模是一个生成问题,油气藏地质建模通常是输入稀疏的、高精度的井点数据,在稠密的、相对低精度的地震数据约束条件下建立高精度三维网格模型,可以视为生成神经网络问题,其中由 Goodfellow *et al.*^[20]提出的生成对抗神经网络(Generative Adversarial Networks, GAN)是解决这类问题的主要方式。

生成对抗网络是一种适用于任意形式复杂分布的无监督对抗式深度学习方法。它一般由两个模块构成:生成模型(Generative Model, 简称为G)和判别模型(Discriminative Model, 简称为D)组成,生成模型也称为生成器,判别模型也称为判别器。生成模型G根据一系列特征,可以是数值或者字符串等,来生成一组数据,生成的数据可以是图片、影像、声音等,也可以是地质模型。生成的数据和真实数据相比要尽可能相似。判别模型D判断生成模型输出的结果是否和真实数据相似,其目标是将生成数据和真实数据尽可能分开。G和D两个模型在对抗性训练中共同进步,G生成的数据越来越逼真,D的辨别能力越来越强。GAN训练终止的条件和一般的网络有所区别,并非以损失函数达到某一阈值作为终止条件,而是人工判断生成模型的结果逼真程度是否还在持续改善。如果持续多轮迭代不再改善则终止训练。

生成对抗网络是目前深度学习地质建模中主要采用的神经网络^[21],为替代模型建立提供了新途径。对于标准的生成对抗网络,需要输入一个低维的随机数序列得到高维的网络层,与非条件地质建模过程较为相似,即通过变差函数、训练图像或目标体长宽高

等输入参数得到一系列三维模型实现,这些实现不考虑与井点数据或地震约束数据的吻合情况,但能够体现较为合理的地质模式。生成对抗网络深度学习体现了类似过程,通过学习地质模式,利用训练后的生成器可以得到满足地质认识的模拟结果。由于神经网络固有的线性化性质,按照这种方式建立的代理模型,具备更好的反向求导和快速收敛特征。

1.3 基于生成对抗网络的代理模型构建方法

深度学习模型的训练需要大量的样本,不断扰动沉积正演模拟器的输入参数,可以获得大规模沉积模拟模型作为深度学习的样本。沉积模拟的输入包括初始地形、碳酸盐岩产率、沉积物搬运、海平面曲线、构造沉降等相关参数,改变任意的输入参数都可能产生不同的结果。采用生成对抗神经网络框架进行训练,其中生成器接收与沉积模拟一样的输入参数,产生二维或三维模拟结果,判别器输入二维或三维模型,输出真或假,其中训练后的生成器即为最终需要的沉积模拟代理模型(图1)。具体构建步骤如下。

步骤一:分析具体的地质问题,搭建反演模拟基本框架。根据钻井和地震解释的地层特征,结合区域地质特征,确定目标地区可能的地层特征和沉积相展布规律。通过反复的正演模拟测试,确定初始地形和构造沉降的基本特征,以及沉积物搬运、碳酸盐岩产率、水体能量相关参数的数值区间,为样本生成提供基础。

步骤二:基于沉积正演模拟器生产大规模样本。选择对模拟结果敏感的参数,且尽量涵盖更大的解空间,通过超立方采样的方式,产生不同的参数组合,开展正演模拟,建立样本库。一个样本包括输入的参数和正演模拟的结果,其中模拟结果是二维或三维的相模型,通过离散数值(比如a1、b1、b2、c1、c2、c3、c4共7种相类型)的矩阵表达。随机选择其中的80%作为训练集,20%作为测试集。

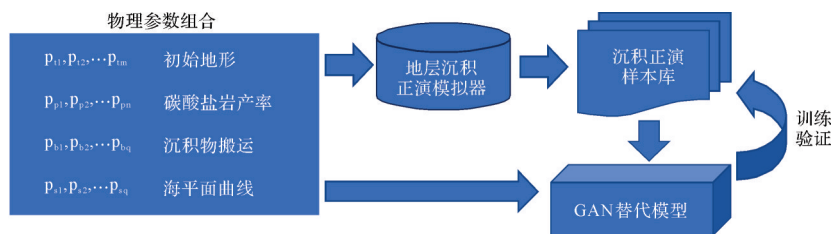


图1 基于生成对抗网络的代理模型构建方法

Fig.1 Proxy model construction method based on generative adversarial network

步骤三:基于生成对抗网络的代理模型构建。根据样本数据特征及生成对抗网络的要求设计生成网络和判别网络的结构,为了更好地捕捉地质模式并增加效率,采用全卷积式网络,即网络的核心处理单元均采用卷积和转置卷积层,每个卷积层和转置卷积层配套标准化层和非线性层,不采用全连接层。在训练集上对网络进行训练,在测试集上对训练后的生成器和判别器进行测试,如果生成器或判别器的效果不再改进,则停止训练。如果训练后网络无法生成合理的地质模型则调整网络结构,甚至回到步骤二生成更多的样本,直到生成器满足要求。

步骤四:基于深度学习代理模型的地层沉积反演模拟。将步骤三训练得到的代理模型(即对抗神经网络中的生成器)带入地层沉积反演模拟系统,以计算结果与观测数据的误差最小为目标,采用全局优化器不断调整代理模型的输入参数,直到得到最优结果。将反演得到的最优参数带入代理模型,得到最终的地质模型。

地层沉积正演模拟器有很强的灵活性。理论上,可以构建代表任意参数个数和参数区间的代理模型,但它需要样本太多、网络的深度也太大,训练难度大,甚至不可能。但是,针对一定的情况,参数空间会大大减小,需要的样本也会减少。针对常见的几种碳酸盐岩沉积环境类型,比如斜坡、镶边台地、孤立台地等类型,可以训练出有针对性、同时又不失一般性的神经网络代理模型。下面将以巴哈马滩碳酸盐岩沉积为例,介绍地层沉积模拟代理模型构建方法及其在反演模拟系统中的应用。

2 实例应用

2.1 巴哈马滩反演模型基本框架搭建

巴哈马滩是国内外研究碳酸盐岩地层沉积的热点地区,在其西缘斜坡有几口科学钻探井和实测的地震剖面,前人在该地区开展了大量研究^[22-24]。台地整体形态、初始地形、构造沉降速度、沉积演化过程等都有比较清楚的认识,也有沉积过程定量模拟的相关研究,但都侧重于沉积过程的控制因素分析,缺少精细的沉积过程建模研究。结合地质分析和前人研究成果^[25-26],用地层沉积正演的方法检测地区地震解释剖面,初步获取该地区碳酸盐岩海平面变化、构造沉降、碳酸盐岩产率等参数,搭建了该地区的地层沉积过程模拟基本框架。沉积模拟范围为 32 km×

32 km,时间跨度为 5.3 Ma,平面网格数为 129×129,总时间步数为 100。

通过正演模拟的参数尝试,获取了主要参数的分布区间,为生成大量样本和代理模型构建奠定了基础。地层沉积正演模拟要输入的参数有初始地形、构造沉降、海平面变化曲线、碳酸盐岩产率相关参数、沉积物搬运参数等 34 个。根据对该地区前期的资料分析认为初始地形和构造沉降的不确定性较低,而海平面曲线、碳酸盐岩产率和沉积物搬运的相关参数不确定性较大,建立反演模拟系统时确定性较强的参数给了较小的变化区间,不确定性较大的参数给了较大的变化区间(表 1)。这 34 个参数中,第 1~4 个参数与初始地形有关(沿着斜坡方向均匀分布的 4 个点,用来插值生成初始地形),第 5~11 个参数与海平面曲线有关(分别是表示三级和四级海平面旋回的正弦函数的振幅、周期和相位,以及海平面基准值),第 12~16 个参数与构造沉降有关(沿着斜坡方向均匀分布的 4 个点,用来插值生成构造沉降面),第 17~20 个参数与沉积物搬运有关(分别是势能和动能在 X 和 Y 方向的搬运系数),第 21~26 个参数与产率有关(分别是势能产率最大幅度、势能产率递减系数、透光带厚度,动能产率系数,动能产率基准值,势能产率权重),第 27~32 个参数与水体动能有关(动能幅度、波浪能下降系数、地形消浪能系数、浪基面、风能系数),第 33 个参数为地层挠曲系数,第 34 个参数为生物群落系数,各个参数更详细的介绍可以参考文献[18]。

2.2 基于沉积正演模拟的大规模样本库建立

纳入全部 34 个可调整的物理参数,并通过随机抽样生成了大量的样本,构建了深度学习所需的数据库。样本为沉积相剖面相模型及相对应的 34 个输入参数,沉积相取值 0~7,其中 0 表示背景值,数字 1~7 对应图 2 中图例中的 a1、b1、b2、c1、c2、c3、c4,表示不同的相类型,该相类型的划分方案同时考虑了沉积物的水体深度和水体能量。模型以二维矩阵形式表示,矩阵尺寸为 100×129,建立了地层沉积过程正演模型样本库。通过人工和自动筛选,去除模拟结果明显不真实不合理的样本后,该数据库共计纳入训练样本 416 079 个,约占总样本的 83%;测试样本 83 167 个,约占总样本的 17%。训练样本和测试样本相互独立,保证深度学习模型不会陷入过拟合。从训练样本中随机抽取几个样本(图 2),可以看出它们

表1 沉积反演模拟参数区间

Table 1 Parameter range of inverse stratigraphic modeling

序号	参数名字	最小值	最大值	序号	参数名字	最小值	最大值
1	InitTopoElev1	-267.00	-67.08	18	PDY1	30 000.00	99 972.00
2	InitTopoElev2	-341.00	-141.08	19	KDX1	200.00	599.84
3	InitTopoElev3	-467.50	-267.08	20	KDY1	200.00	599.84
4	InitTopoElev4	-632.00	-450.07	21	Aprod1	1.30	5.00
5	AmpSeaL0	40.00	79.98	22	Kprod1	0.04	0.12
6	PerSeaL0	230.00	699.81	23	Wprod1	15.00	44.99
7	PhaseSeaL0	-0.50	1.00	24	kEng1	0.50	1.50
8	AmpSeaL1	10.00	49.98	25	bEng1	0.20	0.60
9	PerSeaL1	10.00	99.96	26	pPr1	0.10	0.90
10	PhaseSeaL1	-1.00	1.00	27	kV1	10.00	39.99
11	SeaLevelDatum	-50.00	39.96	28	kW1	0.02	0.20
12	SubsRate1	-0.30	0	29	kR1	0.01	0.06
13	SubsRate2	-0.30	-0.05	30	kH1	0.05	0.17
14	SubsRate3	-0.55	-0.01	31	WaveBase1	20.00	59.98
15	SubsRate4	-0.65	-0.05	32	kL1	0	0
16	SubsRate5	-0.87	-0.05	33	DFlex	1E+24	4.998E+26
17	PDX1	30 000.00	79 980.00	34	rkM1	0	0.10

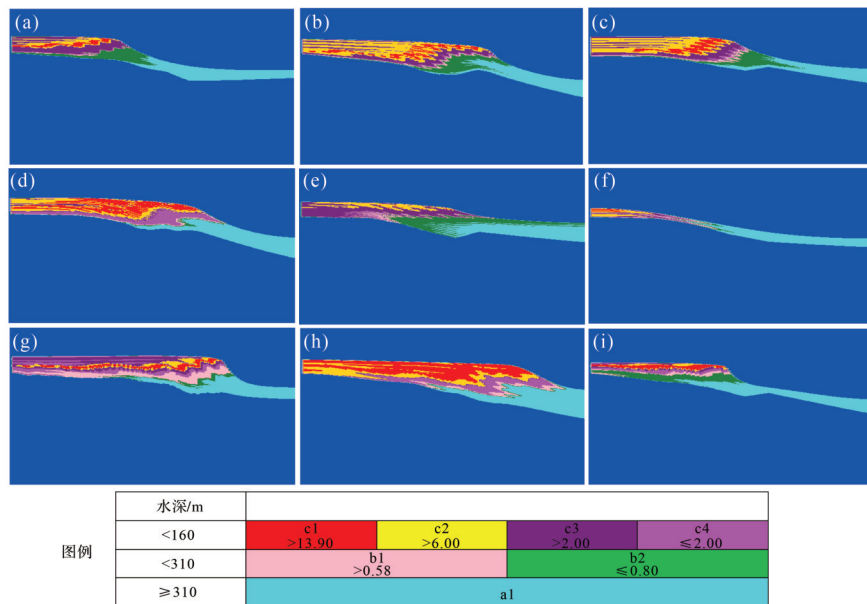


图2 训练集随机抽取的9个训练样本

Fig.2 Nine samples randomly selected from training set

具备不同的沉积相结构样式,体现了样本包含的沉积模式的多样性,这9个样本采用的正演模拟参数具体见附表1。

2.3 基于生成对抗网络的代理模型构建

通过测试,选择了改进后的卷积对抗神经网络GLS-GAN^[27]。原始GAN要最小化生成分布与真实分布的KL散度,同时要最大化两者的JS散度,在数值上会导致梯度不稳定(通常存在梯度消失和梯度爆

炸的可能),并且KL散度具有不对称性,降低了生成器的多样性,会造成模式坍塌问题。GLS-GAN具备损失敏感的自适应损失函数,在训练过程中自动调整损失函数,如果某些区域的真实样本和生成样本已经很接近,则生成器的优化重点就转移到真实样本和生成样本依然差异很大的区域,给GAN提供“按需分配”的建模能力,解决了原版GAN及其衍生算法的梯度消失和模式坍塌问题。

为了保持物理参数和正演结果的一一对应关系,我们在训练生成器时还需要考虑逐个像素的均方误差(Mean Square Error, MSE)。整个网络均由卷积 Conv2D 单元和转置卷积 ConvTranspose2d 单元组成,不包含全连接层,每个单元配套一个批量标准化层 BatchNorm2d 和非线性化层 ReLU,通过大量测试推荐采用5~8个卷积层,生成网络和判别网络的具体结构如图3。

利用上述416 079个训练样本,进行56轮次的迭代训练,取得了比较理想的效果。将不同训练阶段的网络应用在测试集的一套参数上,可以看出不断优化的模拟效果,相序关系变得更加合理(图4)。把最终训练后的网络应用在测试集,随机抽取几个模型,都具备合理的沉积相分布模式(图5),这6个样本的输入参数见附表2。可见相同输入参数下LS-GAN模型的输出与地层沉积过程模型的输出—

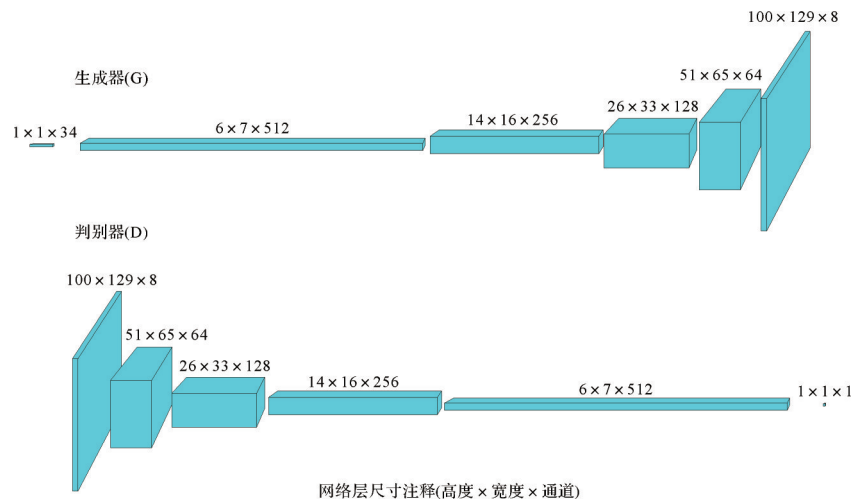


图3 生成对抗网络的结构(生成器和判别器)

Fig.3 Structure of generative adversarial network (generator and discriminator)

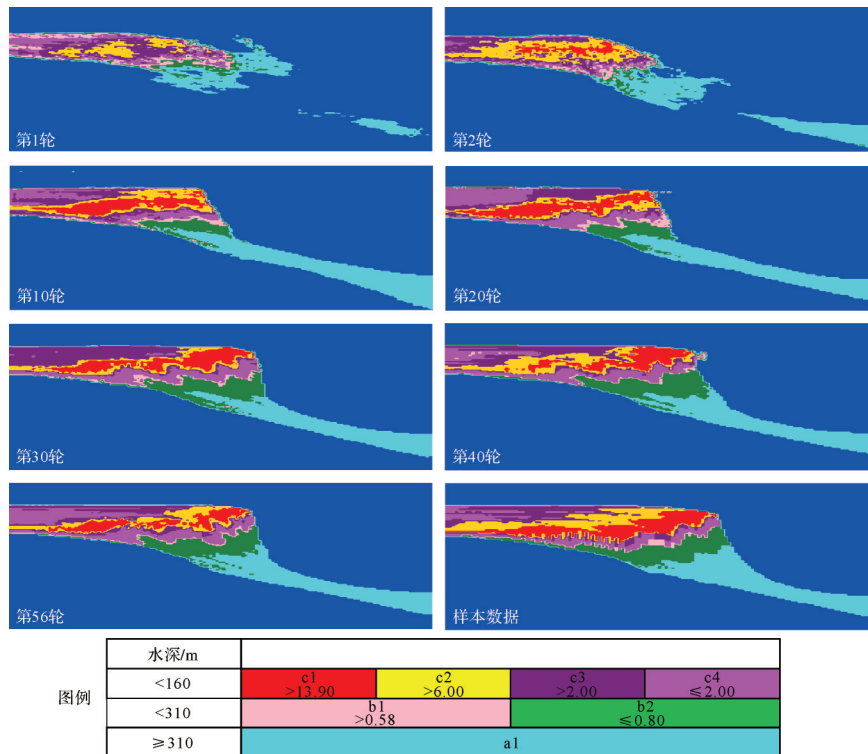


图4 不同训练轮次的网络在测试集某一个样本上的模拟效果

Fig.4 Simulation results of networks different training rounds on a sample of the test set

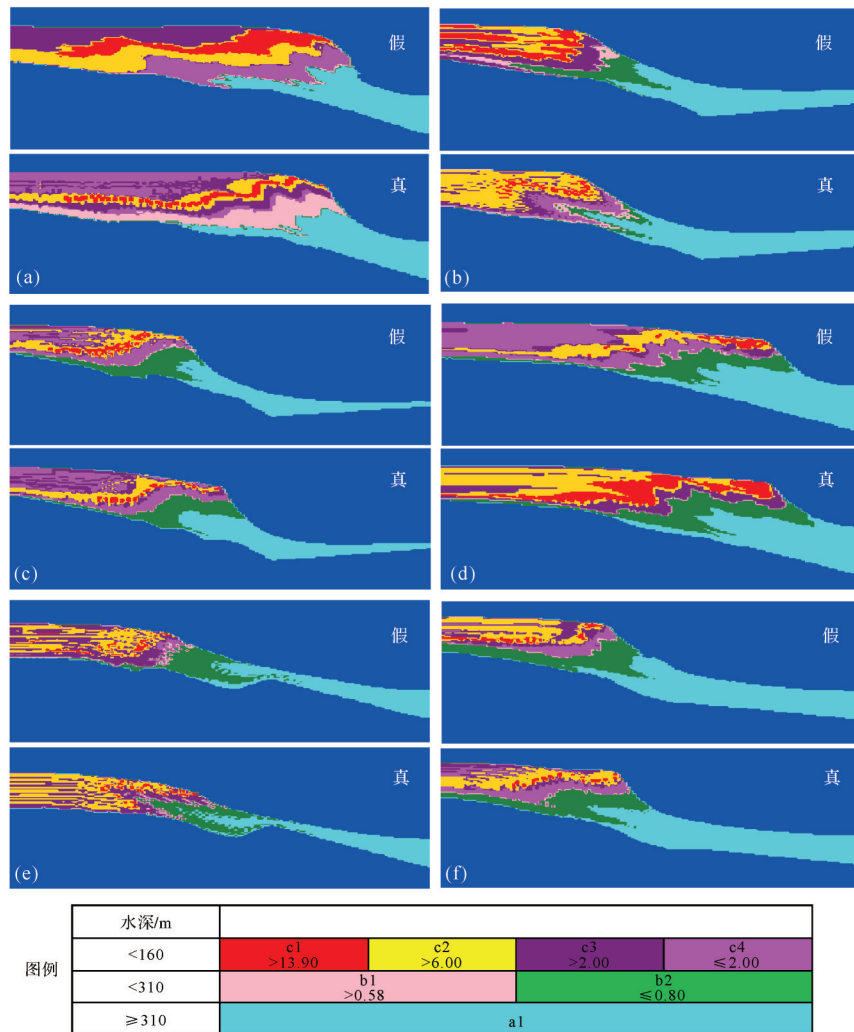


图5 测试集中随机抽取的6组神经网络生成器模拟(假)与传统正演模拟(真)比较
Fig.5 Comparison between generator simulation (false) and conventional forward simulation (true) on randomly selected 6 samples from test set

致性很强,虽然没有达到完全相同,但是基本实现了合理的相带组合特征。通过34个参数生成符合条件的相模型本身是个难度很大问题,GLS-GAN很大程度上改善了模式坍塌问题,但训练难度大、收敛过程不稳定是GAN类方法自身的属性,无法完全避免。另外,网络结构、样本数量,样本质量等也会影响最终的效果。目前认为形成的替代模型可以用于下一步的反演模拟。

2.4 基于代理模型的反演模拟

在开始真实观测数据反演模拟之前,先从测试集中随机选择样本进行参数反演。该样本的参数及其沉积过程模型模拟结果都是已知的,可以用于验证参数反演是否有效。在水平方向等间距设置了6个虚拟测井,作为反演系统的观测数据。用观测数

据与模拟结果的均方根误差作为目标函数,优化器采用上述提到的SCE-UA算法^[19]。

用SCU-UA进行了500 000次迭代优化,得到了预期的效果(图6)。图6a展示了参数的真实值(灰色虚线)和SCE-UA算法得到的最优参数值(蓝色实线),图6b展示了SCE-UA优化过程中目标函数的变化,图6c展示了替代模型LS-GAN给出的地层沉积相,其中虚拟测井位置上的沉积相用真实值(即沉积过程模型模拟值)代替。模拟结果表明,随着优化过程的进行,目标函数稳定下降并不再降低,表明SCE-UA算法已经收敛。然而反演所得到的参数值与参数的真实值有较大差异,且互不相同。这一结果表明地层沉积过程的参数反演是一个高度复杂的问题,存在多个局部极小值。

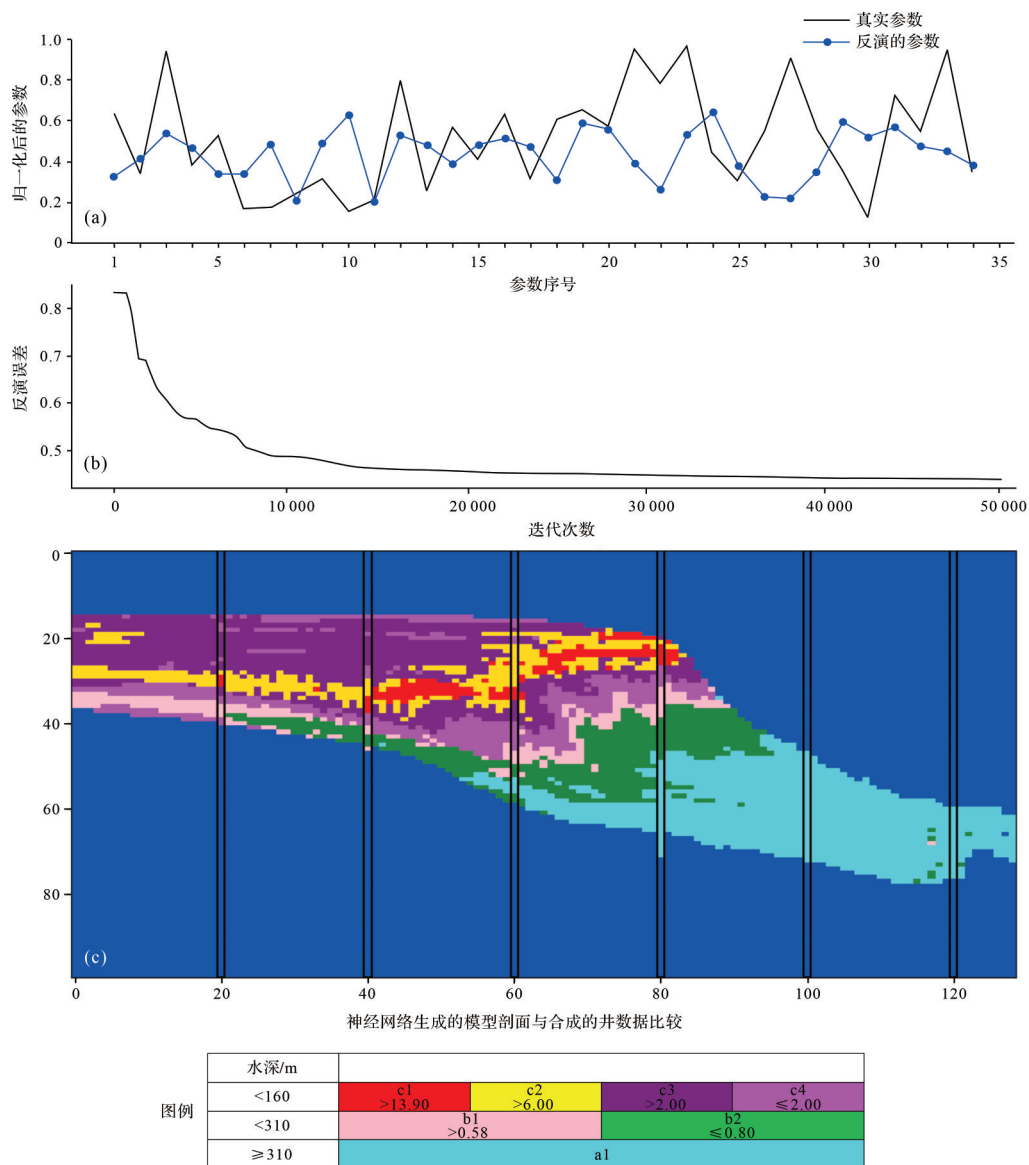


图6 合成样本反演结果

Fig.6 Inverse result on synthetic sample

采用类似的方法进行真实观测数据的反演。将代理模型加入地层沉积反演模拟系统,替换传统的地层正演模拟器,实现参数反演。选择4口科考钻井的人工解释的沉积数据为观测数据,将钻井穿过网格的像素均方根误差为目标函数,SCE-UA为反演优化算法,进行参数反演。对于真实数据,我们并不知道测井以外区域的沉积相,也不知道真实的物理参数的值,因此只能靠专家经验判断结果是否合理。从反演结果可以看出,使用替代模型反演得到的地层沉积相的层序结构明显,分布较为合理(图7)。与基于正演模拟器的反演方法相比,结果一致性较强,效率有大幅提升(图8)。

3 适应性分析

相比传统的SCE-UA算法驱动下调直接采用正演模拟器反演,基于深度学习代理模型的反演在效率方面有大幅提升。选取全部34个参数参与优化,为了使SCE-UA算法充分收敛,使用SCE-UA直接优化地层沉积过程模型往往需要不间断执行一个月。相比之下,基于神经网络代理模型的优化过程在普通计算机上仅需数分钟即可完成。基于深度学习代理模型的地层沉积反演模拟在效率方面有了大大提高,但也存在一定的局限性。

代理模型依赖于正演模拟器,而且需要一定的限制条件。代理模型缺乏正演模拟器所拥有的模拟

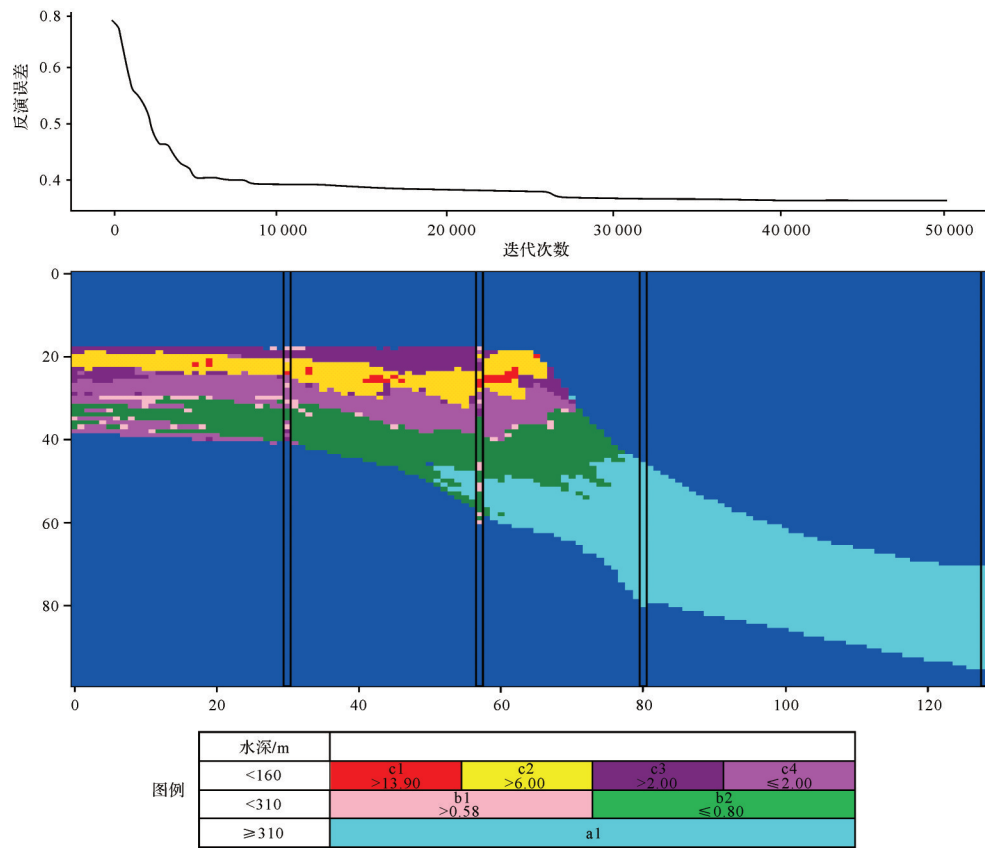


图7 真实数据反演模拟结果
Fig.7 Inverse result on actual observed data

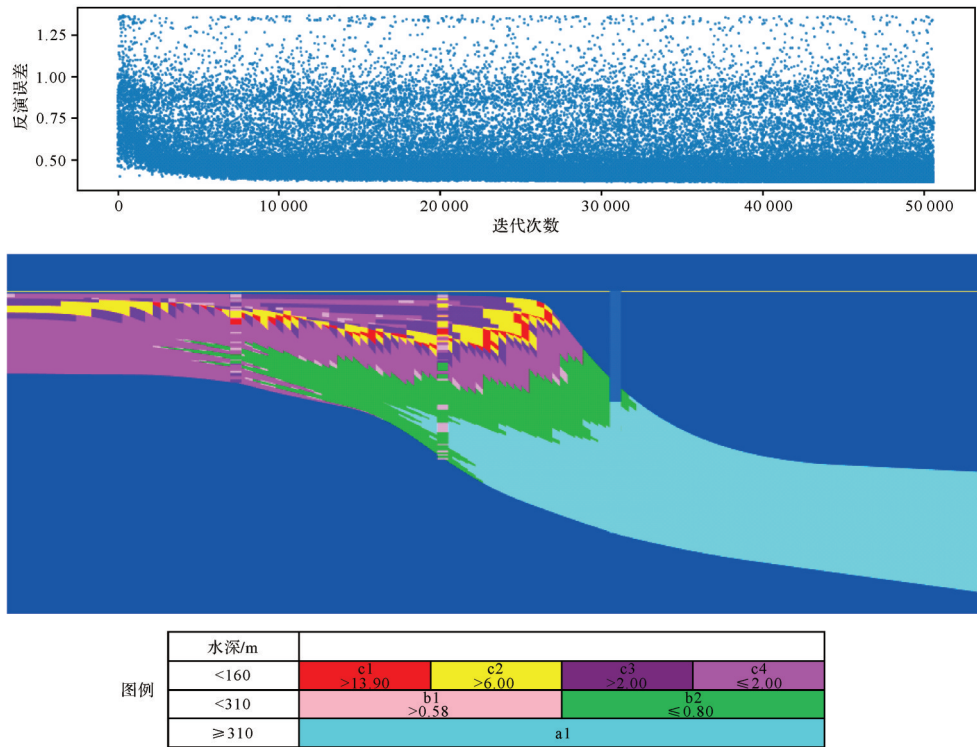


图8 基于正演模拟器的反演模拟结果
Fig.8 Inverse result by conventional forward simulation

能力完备性,尽管生成样本时已经考虑了模型的泛化性能,但是由于沉积模拟的参数根本无法穷尽,特别是初始地形和构造沉降参数,因此只能构建限定条件下的模型代理。这也是代理模型方法逊于传统反演系统的地方。

基于深度学习代理模型构建可以进一步推广应用。本文实例研究展示的是斜坡和镶边类型的碳酸盐岩沉积在二维剖面上的代理模型,其实也可以进行孤立台地、潟湖等更复杂地形条件的下的代理模型构建。也有望推广到三维沉积模拟,不过神经网络的结构、需要的样本数量等具体参数也需要进一步探索。本文展示的是针对碳酸盐岩地层沉积正演模拟的深度学习代理模型构建方法,理论上该方法可以用于其他领域的正演模拟替代模型构建,比如油气藏渗流模型、地下水流动模型等。

4 结论

(1) 地层沉积过程反演模拟系统是在优化器的驱动下,不断调整正演模拟输入参数,提高模拟结果与观测数据的吻合程度,大大增加了沉积过程模拟建模方法的实用性,但是存在反演系统非线性强、耗时长、效率较低的问题。

(2) 生成对抗网络具有较强的模式生成能力,在一定条件下,利用该方法建立的碳酸盐岩地层沉积过程模拟系统的代理模型,具有较强的泛化能力。

(3) 将代理模型带入沉积模拟反演系统,替换正演模拟器,可以大大提高反演的效率。通过巴哈马滩西缘斜坡现代沉积的反演模型实例,验证了该方法的可行性。

参考文献(References)

- [1] Tetzlaff D M, Harbaugh J W. Simulating clastic sedimentation [M]. New York: Springer, 1989.
- [2] Burgess P M, Allen P A. A forward-modelling analysis of the controls on sequence stratigraphical geometries [J]. Geological Society, London, Special Publications, 1996, 103(1): 9-24.
- [3] Lessenger M A, Cross T A. A stratigraphic inverse simulation model[C]//Proceedings of annual meeting of the American association of petroleum geologists. Dallas: AAPG, 1991.
- [4] Cross T A, Lessenger M A. Construction and application of a stratigraphic inverse model[M]//Harbaugh J W, Watney W L, Rankey E C, et al. Numerical experiments in stratigraphy: Recent advances in stratigraphic and sedimentologic computer simulations. Oklahoma: SEPM, 1999: 69-83.
- [5] Bowman S A, Vail P R. Interpreting the stratigraphy of the Baltimore canyon section, offshore New Jersey with PHIL, a stratigraphic simulator[M]//Harbaugh J W, Watney W L, Rankey E C, et al. Numerical experiments in stratigraphy: Recent advances in stratigraphic and sedimentologic computer simulations. Oklahoma: SEPM, 1999: 117-138.
- [6] Tetzlaff D M. Input uncertainty and conditioning in siliciclastic process modelling[J]. Geological Society, London, Special Publications, 2004, 239(1): 95-109.
- [7] Imhof M G, Sharma A K. Quantitative seismostratigraphic inversion of a prograding delta from seismic data [J]. Marine and Petroleum Geology, 2006, 23(7): 735-744.
- [8] Charvin K, Gallagher K, Hampson G L, et al. A Bayesian approach to inverse modelling of stratigraphy, Part 1: Method [J]. Basin Research, 2009, 21(1): 5-25.
- [9] Falivene O, Frascati A, Gesbert S, et al. Automatic calibration of stratigraphic forward models for predicting reservoir presence in exploration [J]. AAPG Bulletin, 2014, 98(9): 1811-1835.
- [10] Sacchi Q, Weltje G J, Verga F. Towards process-based geological reservoir modelling: Obtaining basin-scale constraints from seismic and well data [J]. Marine and Petroleum Geology, 2015, 61: 56-68.
- [11] Sharma A K. Quantitative stratigraphic inversion[D]. Virginia: Virginia Polytechnic Institute and State University, 2006.
- [12] Chan S, Elsheikh A H. Parametrization of stochastic inputs using generative adversarial networks with application in geology [J]. Frontiers in Water, 2020, 2: 5.
- [13] Ren X L, Li X Y, Ren K J, et al. Deep learning-based weather prediction: A survey [J]. Big Data Research, 2021, 23: 100178.
- [14] Gong W, Duan Q Y. An adaptive surrogate modeling-based sampling strategy for parameter optimization and distribution estimation (ASMO-PODE) [J]. Environmental Modelling & Software, 2017, 95: 61-75.
- [15] Pomar L. Ecological control of sedimentary accommodation: Evolution from a carbonate ramp to rimmed shelf, Upper Miocene, Balearic Islands [J]. Palaeogeography, Palaeoclimatology, Palaeoecology, 2001, 175(1/2/3/4): 249-272.
- [16] Pomar L. Carbonate systems [M]//Scarselli N, Adam J, Chiarella D, et al. Regional geology and tectonics. Amsterdam: Elsevier, 2020: 235-311.
- [17] Pomar L, Haq B U. Decoding depositional sequences in carbonate systems: Concepts vs experience [J]. Global and Planetary Change, 2016, 146: 190-225.
- [18] 段太忠,王光付,廉培庆,等. 油气藏定量地质建模方法与应用[M]. 北京:石油工业出版社,2019. [Duan Taizhong, Wang Guangfu, Lian Peiqing, et al. Quantitative geological modeling method and application of oil and gas reservoir [M]. Beijing: Petroleum Industry Press, 2019.]
- [19] Duan Q Y, Sorooshian S, Gupta V K. Optimal use of the SCE-

- UA global optimization method for calibrating watershed models [J]. *Journal of Hydrology*, 1994, 158(3/4): 265-284.
- [20] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C]//Proceedings of the 27th international conference on neural information processing systems. Montreal, Canada: MIT Press, 2014.
- [21] 刘彦锋, 张文彪, 段太忠, 等. 深度学习油气藏地质建模研究进展[J]. *地质科技通报*, 2021, 40(4): 235-241. [Liu Yanfeng, Zhang Wenbiao, Duan Taizhong, et al. Progress of deep learning in oil and gas reservoir geological modeling[J]. *Bulletin of Geological Science and Technology*, 2021, 40(4): 235-241.]
- [22] Betzler C, Reijmer J J G, Bernet K, et al. Sedimentary patterns and geometries of the Bahamian outer carbonate ramp (Miocene-Lower Pliocene, Great Bahama Bank) [J]. *Sedimentology*, 1999, 46(6): 1127-1143.
- [23] Ginsburg R N. Subsurface geology of a prograding carbonate platform margin, great Bahama bank: Results of the bahamas drilling project[M]. Oklahoma:SEPM, 2001.
- [24] Ginsburg R N. The Bahamas drilling project: Background and acquisition of cores and logs[M]//Ginsburg R N. Subsurface geology of a prograding carbonate platform margin, great Bahama bank: Results of the Bahamas drilling project. Oklahoma: SEPM, 2001.
- [25] Eberli G P, Kendall C G S C, Moore P, et al. Testing a seismic interpretation of Great Bahama Bank with a computer simulation [J]. *AAPG Bulletin*, 1994, 78(6): 981-1004.
- [26] Eberli G P, Anselmetti F S, Kroon D, et al. The chronostratigraphic significance of seismic reflections along the Bahamas Transect[J]. *Marine Geology*, 2002, 185(1/2): 1-17.
- [27] Qi G J. Loss-sensitive generative adversarial networks on lipschitz densities [J]. *International Journal of Computer Vision*, 2020, 128(5): 1118-1140.

Construction and Application of a Proxy Model for Stratigraphic Forward Modeling Based on Deep Learning

LIU YanFeng¹, DUAN TaiZhong¹, GONG Wei², LIAN PeiQing¹, ZHANG WenBiao¹, HUANG Yuan¹

1. Petroleum Exploration and Development Research Institute of SINOPEC, Beijing 102206, China

2. Beijing Normal University, Beijing 100091, China

Abstract: Stratigraphic forward modeling (SFM) describes the subsurface geological bodies distribution more realistically than traditional geostatistical modeling, but it is difficult to condition and therefore challenging to apply to practical reservoir geological modeling. Inverse stratigraphic modeling (ISM) improves the practicability of the method. Based on a quantitative comparison between simulated results and observational data, ISM uses a global optimization algorithm to continuously modify the SFM input parameters to find a best fit between simulated and observed data. However, ISM tends to be inefficient and time-consuming, because it has many optimization parameters, strong nonlinearity and long time-consuming of single iteration, and requires a large number of iterations. In this study, ISM was combined with a deep learning algorithm called generative adversarial network (GAN), to construct a stratigraphic forward modeling proxy model. A proxy SFM model based on a large number of synthetic samples formed by neural network training is substituted into the ISM, avoiding the long time-consuming of single iteration, to accelerate the convergence speed of the inversion simulation. The feasibility of the method was verified by application to carbonate stratigraphy on the western margin of the Bahamas. Using the learnt network generated for sedimentation inversion simulation greatly accelerates convergence speed. Although this study focuses 2-D examples, it is expected that the method can be extended to 3-D models.

Key words: generative adversarial network; proxy model; deep learning; stratigraphic forward modeling

附表1 训练集随机抽取的9个样本实例的参数

Attached Table 1 Parameters of 9 samples randomly selected from training set

参数名称	样本A	样本B	样本C	样本D	样本E	样本F	样本G	样本H	样本I	样本J
InitTopoElev1	-241.740	-239.795	-240.203	-189.967	-210.110	-237.965	-212.366	-176.158	-213.844	-228.109
InitTopoElev2	-304.687	-308.444	-288.532	-302.394	-315.569	-316.927	-311.806	-313.322	-299.977	-290.627
InitTopoElev3	-443.803	-420.896	-440.692	-372.690	-426.142	-429.549	-419.037	-445.090	-444.184	-438.012
InitTopoElev4	-608.549	-594.739	-606.268	-608.710	-611.325	-600.513	-599.063	-603.759	-597.293	-608.271
AmpSeaL0	47.993 0	49.966 8	49.034 4	52.207 7	46.511 8	47.745 1	51.338 5	63.144 5	46.858 5	50.374 2
PerSeaL0	415.966	333.573	350.251	319.018	288.357	427.918	302.105	338.303	314.502	308.967
PhaseSeaL0	-0.054 27	-0.013 71	-0.303 06	-0.270 38	-0.078 13	-0.331 04	-0.128 83	0.177 19	-0.295 01	-0.309 59
AmpSeaL1	19.470 5	15.821 2	15.208 0	14.819 3	18.243 8	19.648 7	29.210 7	15.980 5	15.315 4	17.789 8
PerSeaL1	36.966 0	50.087 9	29.274 7	25.920 3	28.986 7	29.643 4	35.675 9	30.717 8	37.475 2	27.578 2
PhaseSeaL1	-0.440 86	-0.645 22	-0.632 37	-0.568 44	-0.768 05	-0.692 32	-0.678 35	-0.655 46	-0.690 46	-0.746 04
SeaLevelDatum	-39.560 5	-38.872 9	-37.119 8	-34.170 6	-34.217 7	-38.279 3	-39.223 3	-37.920 3	-39.168 8	-27.961 8
SubsRate1	-0.243 89	-0.247 15	-0.253 77	-0.258 66	-0.258 47	-0.228 18	-0.243 00	-0.236 90	-0.242 21	-0.195 93
SubsRate2	-0.260 98	-0.219 83	-0.251 05	-0.240 68	-0.228 05	-0.246 30	-0.163 37	-0.237 58	-0.260 05	-0.249 80
SubsRate3	-0.370 92	-0.417 32	-0.476 49	-0.453 03	-0.408 79	-0.467 76	-0.427 09	-0.454 19	-0.458 59	-0.487 05
SubsRate4	-0.563 53	-0.496 77	-0.527 97	-0.487 82	-0.444 86	-0.581 18	-0.539 59	-0.519 66	-0.311 60	-0.544 13
SubsRate5	-0.752 94	-0.781 76	-0.671 17	-0.763 93	-0.687 47	-0.667 36	-0.593 09	-0.723 55	-0.767 86	-0.653 68
PDX1	37 175.7	43 489.0	39 140.9	35 975.7	37 643.6	36 044.0	36 179.2	40 122.3	40 944.5	52 077.4
PDY1	57 366.0	54 775.4	44 088.4	48 277.3	44 436.4	40 488.0	39 890.3	40 007.2	38 728.0	38 879.8
KDX1	275.378	266.399	245.489	256.212	251.925	276.658	245.710	255.417	396.307	275.302
KDY1	262.749	249.093	278.053	312.542	309.070	284.470	270.730	259.157	435.309	438.423
Aprod1	2.617 08	2.159 54	2.729 96	2.279 54	2.192 18	1.835 66	1.766 42	3.566 48	2.022 24	1.734 97
Kprod1	0.072 577	0.049 217	0.049 603	0.051 510	0.052 695	0.053 192	0.049 730	0.059 071	0.049 713	0.049 448
Wprod1	18.598 8	19.906 0	22.985 8	22.905 6	26.301 7	42.079 8	23.508 3	24.068 4	23.353 3	18.509 9
kEng1	0.671 643	0.738 103	0.727 784	0.666 155	0.819 835	0.799 826	0.620 340	0.714 996	0.923 437	0.725 410
bEng1	0.271 113	0.282 144	0.299 423	0.336 546	0.261 021	0.317 633	0.246 501	0.282 42	0.257 869	0.391 195
pPr1	0.206 796	0.196 949	0.268 695	0.204 356	0.305 793	0.231 975	0.261 185	0.364 455	0.210 016	0.258 421
kV1	28.296 3	15.356 4	14.039 5	13.641 2	17.244 1	15.175 5	15.189 6	15.865	20.368 9	16.541 7
kW1	0.058 621	0.051 851	0.061 631	0.044 944	0.053 190	0.125 966	0.043 592	0.060 993	0.050 890	0.041 966
kR1	0.022 455	0.024 601	0.025 026	0.034 426	0.015 812	0.016 974	0.016 930	0.017 06	0.017 712	0.016 261
kH1	0.063 897	0.065 008	0.120 705	0.069 272	0.063 445	0.077 070	0.137 212	0.083 297	0.077 088	0.076 075
WaveBase1	28.713 3	27.629 2	30.370 3	25.878 6	31.746 9	28.013 1	27.847 9	24.5619	30.966 8	31.201 4
kL1	0.000 107	4.65E-05	4.53E-05	6.79E-05	5.09E-05	4.26E-05	6.15E-05	4.73E-05	5.87E-05	4.86E-05
DFlex	1.19E+26	7.95E+25	7.63E+25	6.58E+25	1.13E+26	5.80E+25	6.33E+25	5.81E+25	7.15E+25	1.94E+26
rkM1	0.016 568	0.027 596	0.031 403	0.012 224	0.031 896	0.017 919	0.015 168	0.018 852	0.014 158	0.034 389

附表2 测试9随机抽取的6个样本
Attached Table 2 Parameters of 9 samples randomly selected from test set

参数名称	样本A	样本B	样本C	样本D	样本E	样本F
InitTopoElev1	-235.872	-211.007	-240.939	-226.946	-216.507	-134.497
InitTopoElev2	-294.878	-317.912	-305.656	-283.186	-314.672	-294.664
InitTopoElev3	-440.537	-404.218	-405.011	-363.598	-402.461	-429.075
InitTopoElev4	-611.515	-604.388	-572.029	-611.646	-603.645	-602.346
AmpSeaL0	45.907 7	44.462 8	54.435 1	44.457 3	45.539 2	45.516 2
PerSeaL0	302.895	297.319	283.025	327.798	305.036	285.534
PhaseSeaL0	-0.313 27	-0.268 81	-0.295 00	-0.265 74	-0.125 55	-0.318 12
AmpSeaL1	31.055 7	16.873 3	14.874 9	16.619 7	24.722 8	16.481 8
PerSeaL1	22.732 1	24.420 0	21.763 3	42.429 3	51.685 6	38.696 4
PhaseSeaL1	-0.751 93	-0.711 99	-0.647 79	-0.664 00	-0.199 81	-0.670 46
SeaLevelDatum	-23.082 0	-29.752 5	-33.811 0	-35.430 4	-37.047 0	0.349 351
SubsRate1	-0.236 21	-0.122 30	-0.224 62	-0.231 28	-0.232 31	-0.252 31
SubsRate2	-0.248 49	-0.230 61	-0.260 81	-0.253 01	-0.245 30	-0.262 11
SubsRate3	-0.450 76	-0.489 44	-0.455 16	-0.483 23	-0.454 18	-0.485 23
SubsRate4	-0.572 49	-0.528 56	-0.519 46	-0.463 54	-0.536 64	-0.538 34
SubsRate5	-0.802 85	-0.600 26	-0.794 73	-0.775 51	-0.636 48	-0.770 71
PDX1	39 367.2	51 846.1	50 318.1	48 834.0	40 938.0	37 031.7
PDY1	49 483.2	38 700.8	41 724.9	46 108.6	40 312.3	40 762.6
KDX1	303.702	319.104	317.759	346.898	267.878	261.639
KDY1	262.673	258.522	300.591	355.141	252.191	270.105
Aprod1	2.488 89	1.720 79	2.116 18	1.935 73	1.895 37	1.905 39
Kprod1	0.078 654	0.057 464	0.058 098	0.071 617	0.050 401	0.054 544
Wprod1	18.519 7	19.082 2	21.298 2	27.134 6	19.764 0	18.834 8
kEng1	0.682 237	0.835 305	0.670 850	0.660 123	0.678 857	0.826 163
bEng1	0.348 771	0.268 214	0.268 690	0.302 373	0.271 550	0.259 995
pPr1	0.345 856	0.399 625	0.372 746	0.233 566	0.392 512	0.267 308
kV1	16.119 3	17.201 6	19.910 2	14.708 3	16.862 4	15.730 3
kW1	0.063 514	0.042 187	0.049 663	0.048 818	0.040 613	0.063 422
kR1	0.016 335	0.018 669	0.019 181	0.035 421	0.023 036	0.015 796
kH1	0.074 975	0.070 531	0.065 382	0.063 510	0.086 667	0.063 957
WaveBase1	26.208 4	36.326 0	34.947 2	24.898 5	26.646 7	33.491 0
kL1	0.000 105	0.000 138	8.20E-05	4.97E-05	0.000 124	4.43E-05
DFlex	1.34E+26	1.17E+26	9.29E+25	6.42E+25	8.00E+25	1.58E+26
rkM1	0.039 695	0.012 137	0.030 700	0.020 077	0.014 142	0.022 487