

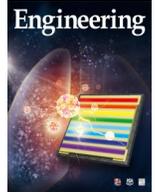


ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
Unconventional and Intelligent Oil & Gas Engineering—Article

裂缝性储层数据驱动模型证伪与不确定性量化

方军龄^{a,b}, 龚斌^{a,b}, Jef Caers^{c,*}

^a School of Earth Resources, China University of Geosciences, Wuhan 430074, China

^b Key Laboratory of Tectonics and Petroleum Resources of the Ministry of Education, Wuhan 430074, China

^c Department of Geological Sciences, Stanford University, Stanford 94305, USA

ARTICLE INFO

Article history:

Received 27 October 2021

Revised 21 February 2022

Accepted 24 April 2022

Available online 11 June 2022

关键词

贝叶斯证据学习

证伪

裂隙性储层

随机森林

近似贝叶斯计算

摘要

天然裂缝的许多特性是不确定的,如裂缝的空间分布、岩石物理特性和流体流动性能。贝叶斯定理提供了一个框架来量化地质建模和流动模拟的不确定性,从而支持储层物性预测。贝叶斯方法在裂缝性储层中的应用大多局限于合成案例。然而,在现场应用中,一个主要问题是贝叶斯先验是被证伪的,因为它不能预测油气藏的生产历史。在本文中,我们展示了如何利用全局敏感性分析(GSA)来确定先验被证伪的原因。然后,我们采用近似贝叶斯计算(ABC)方法,结合基于决策树的代理模型来拟合生产历史。我们将这两种方法应用于一个复杂的裂缝性油气藏,其中综合考虑了所有不确定因素,包括油层物理特性、岩石物理特性、流体特性、离散裂缝参数以及压力和渗透率的动态变化。我们成功地找出了证伪的几个原因。结果表明,我们提出的方法可以有效地量化裂缝性储层建模和流动模拟的不确定性。此外,关键参数的不确定性,如裂缝开度和断层传导率,得到了降低。

© 2022 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

天然裂缝性储层的储量是世界石油和天然气剩余储量的重要部分[1–2]。尽管裂缝的存在增强了流体的储存和流动能力[3–4],但它也给油田开发带来了重大挑战[1]。在储层建模和模拟中,许多裂缝特性是不确定的,如离散裂缝网络(DFN)参数[5–6]、渗透率各向异性[2]、相对渗透率和润湿性[7–8],以及裂缝渗透率对储层压力的依赖[9–10]。因此,不确定性的量化在降低裂缝性储层开发的决策风险方面起着关键作用。

从不同来源获得的信息,如三维(3D)地震测量[11–

14]、裂缝野外露头[2,15–17]、井眼成像测井[18–19]和实验室实验数据[10],有助于减少静态储层建模和动态流动模拟的不确定性。

储层预测的不确定性量化通常是通过生成受生产历史制约的多个模型来进行的[20–21],即后验概率估计。贝叶斯定理是一个常用的框架,基于一个声明的先验和似然模型,用于生成(后验)模型[22]。当用于断裂属性时,先验模型使其有可能将蒙特卡罗模拟产生的重要地质特征[23–24]强加到模型上。在蒙特卡罗方法中,不确定的变量是随机抽样的。针对我们想要的采样的分布,引入了马尔可夫链蒙特卡罗(McMC)方法。最近,McMC方法已

* Corresponding author.

E-mail address: jcaers@stanford.edu (J. Caers).

被越来越多地用于历史匹配和不确定性量化[21,25]。McMC生成了一个模型的依赖链,使得估计和采样后验分布成为可能。在马尔可夫链的每一次迭代中,都会使用一个接受规则来决定是接受提议的模型还是保留旧的模型。为了提高McMC方法的性能,已经开发了几种修正的McMC方法,包括混合McMC [26]、模式跳跃McMC [27]和平行交互McMC [28],这些方法处理了实际应用中出现的低接受度、多模式后验和高维问题。

由于McMC方法的实现是按顺序产生的,因此其效率不高。集合方法,如集合卡尔曼滤波(EnKF) [29–31]、集合平滑器(ES) [32–33]、迭代EnKF [34]和集合平滑器与多数据同化(ES-MDA) [35],最近在量化不确定性和历史匹配方面得到了关注。EnKF是一种蒙特卡罗方法,为数据同化实现了基于协方差的贝叶斯更新方案[31]。在线性系统动力学和模型参数多元高斯分布的主要假设下,先验模型参数和生产数据通过观测数据的同化被联合更新。ES方法用于避免在参数更新时重新运行模拟,而迭代EnKF则扩展了EnKF的假设,使之适用于数据和模型参数之间的高度非线性关系。ES-MDA通过引入一个膨胀的误差协方差矩阵来改进ES方法。

最近,直接预测[24,36–37]和数据空间反演[38–39]已被用来评估未来储层性能预测的不确定性,而无需生成后验储层模型。与先前不同,直接预测在(历史)数据和目标(未来预测)之间建立了一个统计模型。后验预测分布是由观测值直接估计的。数据空间反演将生产数据视为随机变量。后验数据变量分布是通过数据空间随机化最大似然法进行采样。

然而,这些方法并没有解决先验分布被证伪的情况[40]。被证伪只是意味着,无论生成了多少实现,所述模型都无法预测历史数据。贝叶斯证据学习(BEL) [41]提供了一个包括证伪的不确定性量化的协议,被越来越多地用于量化石油和天然气[24]、地下水[42]和地热[23]应用中的不确定性。与其说BEL是一种方法,不如说是一份如何应用现有方法来解决现场案例问题的协议。BEL的优点在于可以处理任何先验分布和对接任何敏感性分析方法,这使其适用于量化实践中复杂问题中存在的确定性。

许多不确定性量化过程是很耗时的,因为这个过程通常需要从地质建模和正演模拟中生成数千到数万的模型样本。在裂缝性储层中,必须考虑大量的计算成本,以便利用非结构化网格或精炼的结构化网格明确地捕捉裂缝中的详细流体流动。这一挑战使得建立代用模型(也被称为代理模型)作为计算成本较低的替代方案对实际的裂缝性储层具有吸引力。人们已经开发了许多用于历史匹配和不确定

性量化的代理模型,如多项式模型、克里格模型、具有空间填充设计的样条代理模型,以及更广泛使用的响应面代理模型[43–45]。最近,树状回归用于敏感性分析和历史匹配,如随机森林模型[46]和随机梯度提升[47],已经获得越来越多的关注。

许多关于不确定性量化的研究和方法使用合成案例(如参考文献[39]),其中先验不确定性是人工设计的,而观测数据是从模拟结果中选择的。这种简化绕过了证伪问题,因为从模拟中选择的“真相”总是在先验数据的群体中,这表明先验中观察的概率足够突出。在实际案例中,先验模型的不确定性,即通过仔细的地质、地球物理和工程分析,往往无法预测实际观测结果,这意味着先验模型被证伪[40]。先验模型被证伪的原因有很多:参数的不确定性范围可能太窄或被认为是确定性的,使用的物理模型可能不正确,或者两者的任何组合都可能发生[23,48]。

在基于先验的技术中,经常会发现证伪问题[24,49]。然而,需要一个系统的方法,来诊断先验问题,并将其调整为非证伪,即去伪存真。本文的第一个主要贡献是,我们提出了一种系统的方法,在先验模型被证伪时,该方法指出先验模型中的问题。我们结合全局敏感性分析(GSA)、降维和蒙特卡罗技术来识别问题。第二个贡献是我们采用了近似贝叶斯计算(ABC)方法,并与基于随机森林的代理模型相结合,对非证伪的先验模型进行训练,以匹配生产历史。最后,我们将此方法应用于一个复杂的裂缝性储层,所有的不确定因素都被共同考虑,包括油层物理特性、岩石物理、流体特性和裂缝特性。

现场案例是位于中国渤海盆地的一个自然裂缝变质储层,为了规划更多的油井,需要在不确定性下进行决策。不确定因素包括裂缝密度、裂缝开度、井位和空间分布的渗透率、相对渗透率和断层渗透率。问题是如何评估,更重要的是如何基于测量数据(即测井和地震数据)和历史数据(即生产速率和压力)降低储层模型的不确定性。

2. 案例分析

2.1. 现场案例的贝叶斯证据学习

BEL是一个用于不确定性量化和反演的综合框架[41]。与其说BEL是一种方法,不如说是一份基于贝叶斯推理的先验协议。它使用GSA为当前的案例提出有效的反演方法(即历史拟合),并使用观测数据(即本研究中的生产历史数据)作为证据,通过学习先验分布推断模型假设[24]。一般来说,BEL包括6个阶段[23]。

(1) 制定决策问题和定义预测变量。

- (2) 陈述先验不确定性模型。
- (3) 蒙特卡罗模拟和证伪先验不确定性模型。
- (4) 数据和预测变量的GSA。
- (5) 基于数据减少不确定性。
- (6) 后验和决策。

这项工作旨在量化储层建模和模拟过程中的不确定性；它与框架中的第2-5阶段相对应。

2.2. 地质背景和生产历史

研究区域是中国的一个海上天然断裂的潜山储层。储层由变质岩组成，渗透率极低。在被埋入地下之前，储层的顶部被风化，形成了储存和流体流动的能力。天然裂缝是由风化作用和构造运动共同作用的结果。在储层的上部地区有5个生产井，在储层的相对下部地区有4个注水井[附录A中的图S1(a)]。有两口勘探井已被钻出，但没有进入生产或注入阶段。

储层中的流体是石油和水，它们被-1880 m附近的一个恒定的油水接触(WOC)所分开[附录A中的图S1(b)]。该储层于2015年9月开始产油。收集了5口生产井的生产历史数据(附录A中的图S2)。观察数据是所有生产井的产油率(OPR)和产气率(GPR)，以及生产井P1、P2和P3的孔底压力(BHP)。直到2019年3月，所有的生产井都没有产生水。

2.3. 储层建模数据和先验的不确定性提取

在静态建模和流动模拟的许多方面，如结构建模、油层物理建模、流体建模和岩石物理建模，仍然存在不确定性。在本节中，我们描述了可用于表征储层的信息以及整个建模和模拟过程中的不确定性来源。表1中列出了量化的不确定性参数及其分布，整个建模工作流程见图1。

2.3.1. 结构不确定性

储层在纵向被分为三个区域[附录A中的图S3(a)]。由于风化程度自上而下降低，储层与非储层的比例也随之降低。根据3D地震蚂蚁追踪属性[附录A中的图S3(b)]，解释了46条断层(附录A中的图S4)。所有断层的位置都是确定的，而断层的传导性仍然是不确定的，因为断层的传导性在储层中很难测量。在这里，我们把断层导电性的不确定性简化为两种情况：完全密封和完全开放。

2.3.2. 岩石学性质的不确定性

石油物理特性，如孔隙度和渗透率，具有不确定性。在这种情况下，基质孔隙度是通过测井仪测量的。测量地震声阻抗(附录A中的图S4)以表明孔隙度的空间变化。我们应用序贯高斯模拟(co-SGS)来模拟3D属性分布，将测井解释作为“硬”数据，地震属性作为“软”数据。在矩阵属性建模过程中，不确定的建模参数包括变差函数参数，以及地震属性与井眼属性的相关系数。基质渗透率在井位上是无法测量的。因此，我们提供一个取决于过去经验的先验。基质渗透率的空间分布受到孔隙度的制约。

2.3.3. 断裂特性的不确定性

裂缝可能对流体流动有相当大的影响。现有的四口井(两口生产井和两口勘探井)利用成像测井直接测量裂缝信息，包括裂缝密度和方向。裂缝密度的空间变化可以通过地震方位角各向异性来表示(附录A中的图S4)。我们使用序贯高斯模拟(co-SGS)来模拟地震方位角各向异性方程约束下的3D断裂密度模型。几个建模参数是不确定的，如变差函数和与地震属性的相关系数。然后用裂缝密度、裂缝走向、裂缝长度分布、裂缝开度等建立一个随机的DFN模型。断裂方向受到断层线趋势的制约，而断

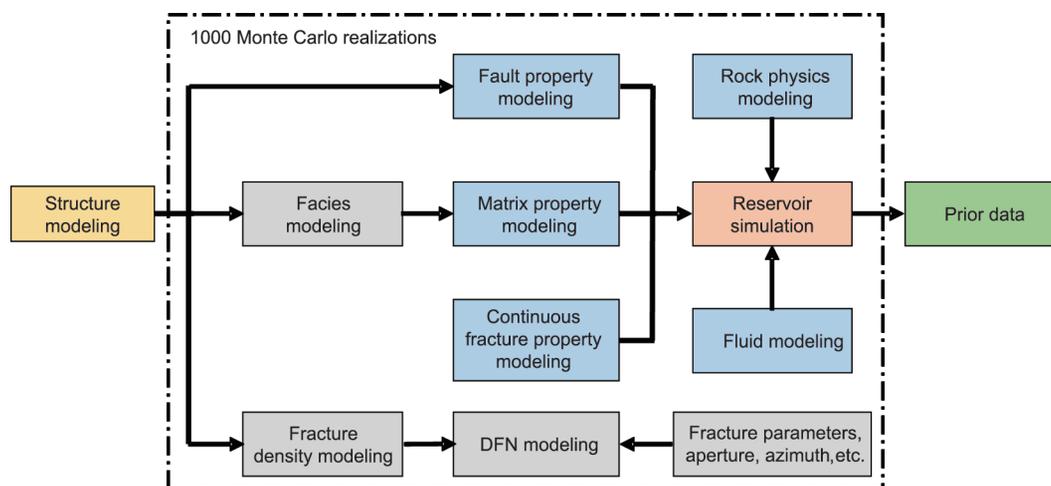


图1. 储层建模和储层模拟工作流程。

表1 不确定性参数及其分布

Model	Global parameters	Variable name	Prior uncertainty	Unit
Facies	Major variogram range	MajorVarioFacies	U (500, 3000)	m
	Ratio of minor to major variogram	MinorVarioFacies	U (0.3, 1.0)	m
	Vertical variogram	VertiVarioFacies	U (2, 10)	m
	Major variogram direction	VarioDirectFacies	U (-90, 90)	°
Petrophysical properties	Major variogram range	MajorVarioPoro	U (500, 3000)	m
	Ratio of minor to major variogram	MinorVarioPoro	U (0.3, 1.0)	m
	Vertical variogram	VertiVarioPoro	U (2, 10)	m
	Major variogram direction	VarioDirectPoro	U (-90, 90)	°
	Coefficient of seismic impedance to porosity	CoImpedancePoro	U (-1, 0.5)	—
	Matrix porosity multiplier	MtrxPoroMulti	U (0.5, 1.5)	—
	Fracture density	Major variogram range of fracture density	MajorVarioFrac	U (500,3000)
Ratio of minor to major variogram of fracture density		MinorVarioFrac	U (0.3, 1.0)	m
Vertical variogram of fracture density		VertiVarioFrac	U (2, 10)	m
Coefficient of seismic anisotropy to fracture density		CoSeisAzimuFD	U (-1, 1)	—
Major variogram direction of fracture density		VarioDirectFrac	U (-90, 90)	°
DFN	Fracture aperture	FracAperture	Log-U (-2.2, -1.3)	mm
	Direction variation of fracture azimuth	DirectVariFracAzimu	U (-90, 90)	°
	Fracture length distribution shape	FracLengthShape	U (2, 3)	m
	Maximum fracture length	MaxFracLength	U (100, 200)	m
	Minimum fracture length	MinFracLength	U (5 30)	m
	Minimum explicit fracture length	MinExplicitFracLength	U (100, MaxFracLength -10)	m
	Fracture concentration	FracConcentration	Log-U (2, 5)	—
Fluid	Oil-water contact	WOC	U (-1870, -1890)	m
	Oil density	OilDensity	U (920, 960)	kg·m ⁻³
	Bubble point pressure	BubPointPressure	U (80, 120)	MPa
	Gas-oil contact	GOC	U (-1570, -1400)	m
	Gas density	GasDensity	U (0.75, 0.8)	kg·m ⁻³
Rock physics	Maximum capillary pressure	MaxPc	Normal (0.1, 3.0)	MPa
	Water saturation at Cp = 0	SwPc0	U (0.6, Sorw)	—
	Initial water saturation	Swi	U (0, 0.4)	—
	Residual oil saturation	Sorw	U (0, 0.3)	—
	Rock compressibility	RockCompressibility	U (0.00004, 0.00700)	m
Fault property	Fault seal/open scenarios	FT1/FT2/.../FT46	Binary	—

Cp: capillary pressure; U: uniform; GOC: gas-oil contact.

裂方向对断层线的偏离是不确定的。接着用 Oda 方法[50] 将 DFN 模型升格为连续网格属性，以降低流体流动模拟的计算复杂性。

2.3.4. 岩石和流体性质的不确定性

裂缝使得获得岩石物理学测量和流体与岩石之间的相互作用变得困难。在这种情况下，对于储层而言，没有可用的直接测量数据，如岩石压缩性、相对渗透率曲线、初始水饱和度或毛细管压力。因此，我们参考先前的研究 [8,51]和类似的储层来定义岩石物理特性和不确定性范围。通过引入石油公司提供的参考（基准）值的范围，压力-

体积-温度（PVT）属性，包括泡点压力、油气密度和其他流体分布参数，如气-油接触（GOC）、油-水接触（WOC），被视为不确定性。

2.3.5. 所有先验不确定因素表

基于贝叶斯模型的不确定性量化需要有模型参数化的先验声明和每个参数的概率分布。在前面几节对不确定性来源的分析中，我们观察到很多方面的属性是不确定的，包括油层物理特性、流体储存、PVT、岩石物理以及不同尺度的裂缝所造成的空间异质性。一方面，我们试图在整个建模和模拟过程中考虑每一个可能的不确定参数，以避

免遗漏重要参数。另一方面，我们需要平衡模型的复杂性和计算需求，以解决多种不确定因素。表1中列出了不确定参数的详细名称和分布。

2.4. 静态建模、流动模拟和蒙特卡罗模拟

由于基于物理学的地下系统模型是高维和非线性的，因此需要进行蒙特卡罗模拟。先验模型中指定的每个变量都根据其概率分布进行抽样，通过蒙特卡罗模拟产生单一模型的实现。我们应用双开度双渗透性模拟作为正演模型，计算生产数据作为对不确定参数变化的反应。商业软件PETREL和ECLIPSE E100，分别被用来做静态建模和储层模拟。正演模型的功能可以形式化为：

$$d = g_d(m_{\text{matrix}}, m_{\text{frac}}, m_{\text{fault}}, m_{\text{fluid}}, m_{\text{rock}}) \quad (1)$$

式中， g_d 是一个确定性函数，将模型映射到数据变量 d ，即模拟生产数据； m_{matrix} 是一个基质模型； m_{frac} 是一个断裂模型； m_{fault} 是一个断层模型； m_{fluid} 是一个流体模型，包括流体接触和PVT； m_{rock} 是一个岩石物理模型，包括岩石压缩性、毛细管压力和相对渗透性。

地质建模和储层模拟的整个工作流程见图1。我们从表1的分布中对所有不确定的参数进行蒙特卡罗抽样；生

成1000个模型（附录A中的图S5）和模拟结果。

图2展示了模拟的生产数据与生产历史的比较。可以看出，对于一些生产曲线，先验似乎涵盖了观察结果，如P1、P3、P4和P5井的气体速率。P3井的BHP曲线则不是这样。P1的BHP和P2的气体速率有一定程度的覆盖，但在时间上有不同的趋势。这些结果表明，即使考虑到许多方面的不确定性，先验也是不够的。

在下一节中，我们开发了一套系统的方法来识别先验的问题，更重要的是，当先验被证伪时，我们该如何纠正它。

3. 方法

3.1. 贝叶斯证据学习中的证伪问题

一旦确立了正演模型，就有必要检查是否对定义的先验模型进行证伪。被证伪的先验意味着先验声明不具有涵盖观察的信息性。因此，后验计算将是失败的。在概率论方法中，通常试图证明先验模型是不正确的，而不是证明它是正确的（即拒绝“先验模型预测了数据”的无效假设）。如果先验模型不能被证伪，那么它所做的假设就得

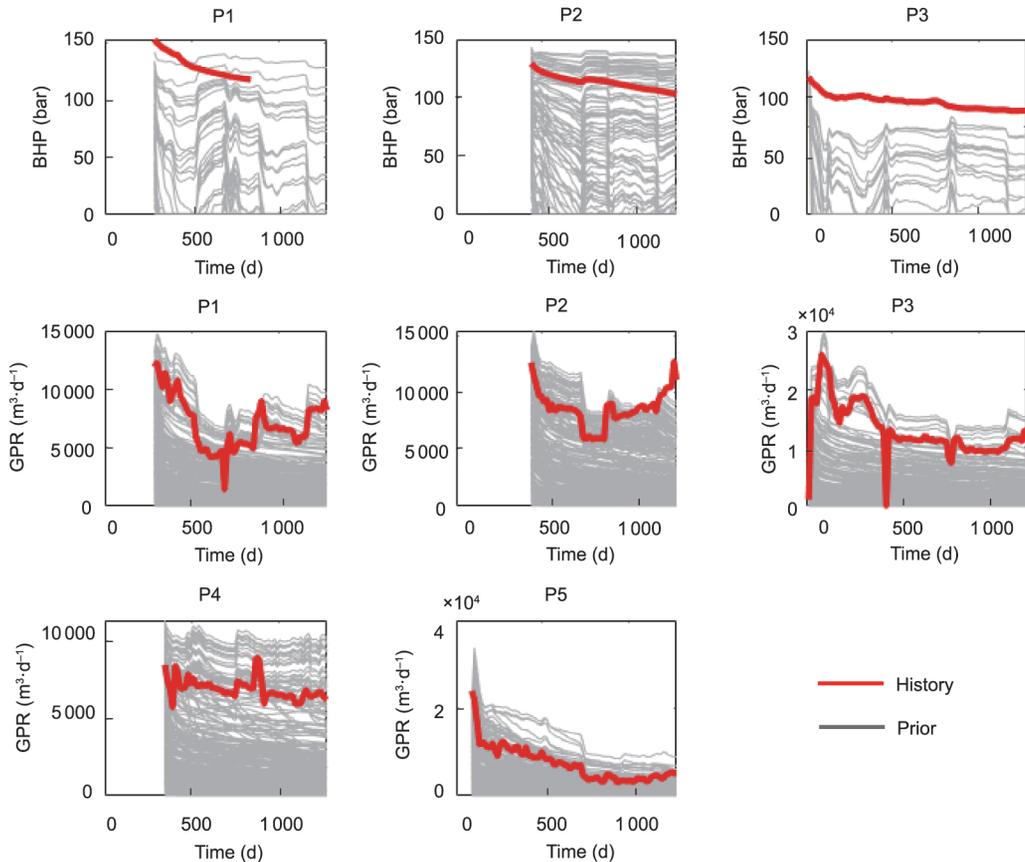


图2. 使用表1中定义的先验的模拟数据。

到了加强,但不一定被证明是正确的。证实过程可以通过验证以下内容来实现:①模型再现了系统中已知的物理变化,或者②模型可以预测观察到的数据[23]。后者不需要与历史数据(即观测数据, d_{obs})相匹配,而是表明在使用公式(1)生成的数据响应群体中,观测数据的概率不为零。这种证伪过程已经在以前的许多研究中使用[25,42,52]。开发先验模型通常是迭代的,也就是说,先验模型的第一次选择往往是被证伪的。根据先验模型被证伪的原因,先验模型可能需要通过增加①模型的复杂性、②参数的不确定性,或③两者都增加来调整。

当数据变量的维度较低时,通过视觉检查的方式就可以直接发现证伪行为。在本研究中,数据变量是高维的,这意味着需要一种系统的方法来检查。基于马氏距离的方法在证伪检测方面已经得到了关注。Yin等[24]使用稳健的马氏距离(RMD)方法来检测具有多高斯分布的先验。在现场案例的应用中,模拟的生产数据可能不是高斯的。Alfonza和Oliver[48]使用观察和模拟集合之间的马氏距离的近似值来诊断先验模型参数。他们的方法既可用于高斯和非高斯数据,也可用于有大量数据的大型模型。然而,他们的方法侧重于可接受多少个样本来改进先验模型。结果并没有给出先验模型是否被证伪的直接标准。单类支持向量机(SVM)是一种流行的离群点检测方法[41,53],在数据空间中围绕样本的最小体积超球被拟合。任何落在超球之外的观察都被归类为与先验不一致。使用单类SVM为检测高维和非线性决策边界中的异常值提供了一个强大的工具。我们采用单类SVM作为证伪的方法来诊断先验模型。

3.2. 全局敏感性分析

分析蒙特卡罗模拟的另一种方法是通过GSA,调查哪些模型参数(或其组合)影响响应(如模拟产量)。GSA方法是基于蒙特卡罗抽样,其中所有的参数是共同变化的[54]。在本研究中,我们使用基于距离的广义敏感性分析(DGSA)方法进行全局敏感性分析[55-56]。DGSA是一种区域化的GSA方法[57],它将样本的数量减少为少量的类。然后计算先验累积分布函数(CDF)和类条件CDF之间的距离来衡量敏感性。DGSA的一个优点是它支持所有类型的输入参数分布,如连续、离散和基于情景的分布。DGSA的另一个优点是它考虑了计算机模型可能的高维响应[41],这在地下系统中是很常见的。

我们采用DGSA从两个方面来协助先验的去伪存真。第一个方面是对当前先验模型参数的调整。DGSA通过降低模型的复杂性来过滤掉非敏感参数,从而简化了去伪存

真的过程。另一个方面是通过增加新的不确定参数来增加先验模型的复杂性。DGSA可以检验新模型参数的有效性。在应用中,我们将展示DGSA如何用于识别先验模型被证伪时的问题。

3.3. 用数据进行不确定性量化

如果先验没有被证伪,就可以进行适当的贝叶斯不确定性量化。完全贝叶斯方法估计在先验分布 $P(m)$ 下产生的函数 $P(d|m)$ 。然后通过以下公式计算模型的后验分布 $P(m|d)$ 。

$$P(m|d) = \frac{P(d|m)P(m)}{P(d)} \quad (2)$$

然而,在具有高维和不同分布类型的参数的实际案例中,评估似然的计算成本很高,或者说完全不可行[58]。这里,我们采用ABC方法来生成变量的后验分布[58-59]。ABC方法源于拒绝抽样。ABC通过比较模拟数据和观察数据来逼近似然函数。在ABC方法中,如果模型的模拟数据 $g_d(m)$ 与观测数据 d_{obs} 接近,则模型 m 被接受。因此,这种方法绕过了精确的似然计算[60]。随着数据维度的增加,产生与观测数据差距较小的被接受模型的概率也会下降。概括统计,即从数据中计算出的数值,以最简单的形式代表最大的信息量,被引入ABC方法来代替数据。如果模拟数据集的汇总统计量与观察数据集的汇总统计量的距离在 ε 以内,该方法就会接受样本。被接受的样本的分布是一个近似的后验,由以下公式给出:

$$f_\varepsilon(m|d) = f(m) P(d|m) I(d(g_d(m), d_{\text{obs}}) < \varepsilon) dd \quad (3)$$

式中, $I(d(g_d(m), d_{\text{obs}}) < \varepsilon)$ 是一个汇总统计量,定义了一个窗口大小 ε ,定义了接受区域。

必须强调的是,除非数据 d 足够多,否则近似误差不会随着 $\varepsilon \rightarrow 0$ 而消失[58,60]。另一方面,随着数据变量维度 d 的增加,接受率也会下降。这时我们应该增加 ε ,这将导致近似的质量降低。基于这些限制,我们必须:①产生尽可能多的模型 m ,并将模型 m 模拟到数据 d 上;②减少数据变量 d 的维度。

由于对有数百万个网格单元的实地案例进行DFN建模和模拟非常耗时,要产生ABC方法所要求的那么多个,甚至可能是几万个实际情况是不现实的。因此,需要一个代理模型,这将在下一节进行解释。

3.4. 代理模型

3.4.1. 基于树的回归

我们使用基于树的回归[61]的代理模型,以取代耗时

的静态和正演模拟。基于决策树的方法是基于输入空间的二进制分区，采用树状拓扑结构，模型 y 在分区区域采用常数。模型的输出是片状不连续的，可以表述为：

$$y = \sum_{m=1}^M c_m I(x \in R_m) \quad (4)$$

式中， c_m 是常数； R_m 是分区区域； $I(x)$ 是 M 个地区的离散指标函数的线性组合。回归树的优点是其对连续和分类输入变量都有效，这在地质问题中是常见的。然而，由于其片断性，回归树方法权衡了回归精度。其预测精度低于大多数其他方法[62]。

3.4.2. 集合树

“Boosting” [63]和“Bagging” [64]是两种通过树的集合来提高树状回归的预测性能的方法。Boosting提升法使用了较弱的估计器的结果，然后结合每个弱分类器的优势来提高预测性能[65–66]。Boosting提升法的树是按顺序生成的。在每个步骤中，用早期估计器训练树估计器，对数据进行简单的模型拟合，再对数据进行错误分析。然后增加有较多错误的样本的权重。最终的估计器是对估计器序列的加权组合。梯度Boosting提升法是Boosting提升法的延伸；它结合了梯度下降算法来优化可微损失函数。每一

代的新树都试图恢复实际值和预测值之间的损失[65]。

用树进行Bagging (bootstrap)集成[64]，从训练样本中随机选择几个数据子集进行替换。使用每个数据子集独立训练树状估计器。最终的估计器是通过用这些自举样本对不同树产生的所有估计器进行平均而产生的。随机森林[67]法是Bagging的一个扩展。除了选择一个随机的数据子集外，它还随机选择输入特征，而不是使用所有的特征来生长树。这种扩展可以处理缺失值，保持缺失数据的准确性，并且能很好地处理高维数据。

4. 先验证伪和不确定性量化——实例研究

4.1. 证伪识别

现在我们回到案例研究中，按照BEL框架的顺序进行。我们已经说明了先验的不确定性并进行了蒙特卡罗模拟；因此，下一步是证伪检测。我们将模拟的数据和观察到的数据映射到较低的维度上，然后在较低的维度空间上进行离群点检测，以进行证伪检测。这里，我们使用主成分分析（PCA）进行降维。图3显示了不同模拟数据的主成分（PC）得分图。

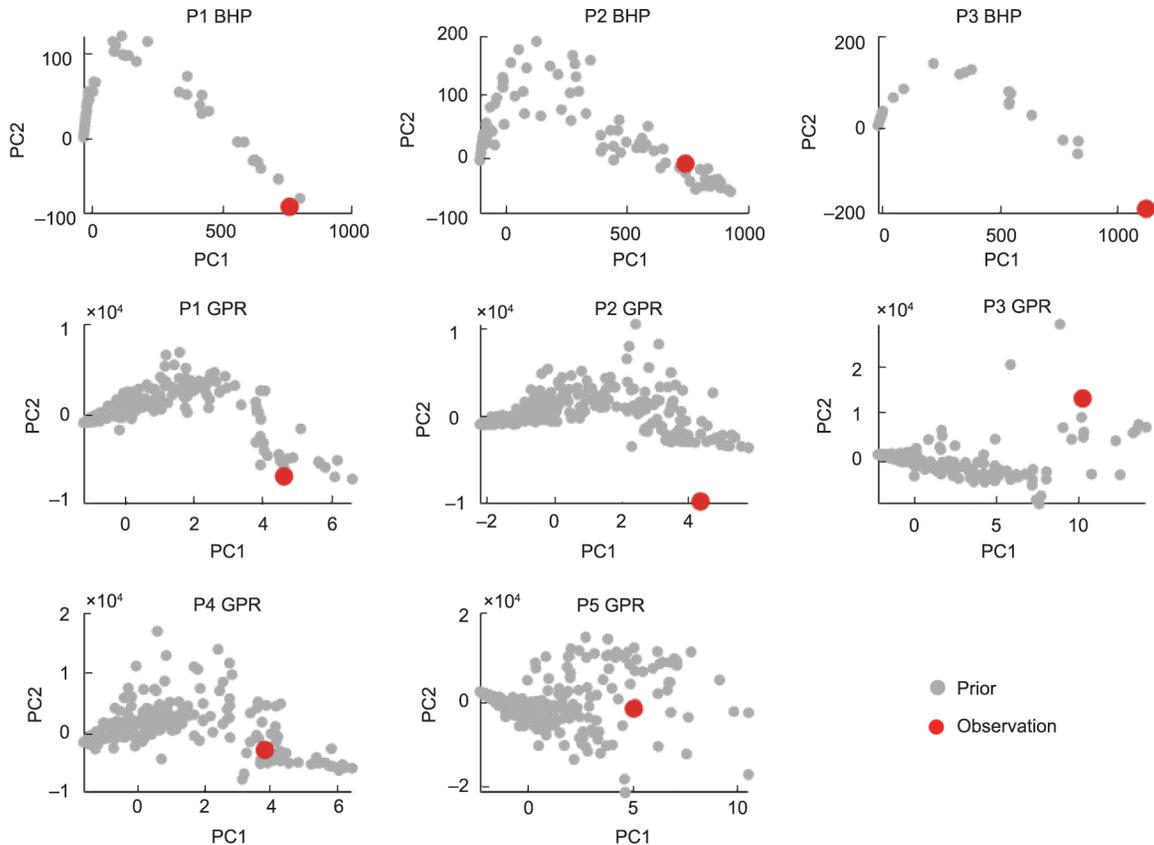


图3. 各生产井和数据类型的PCA图。灰点是模拟结果，红点是实际的观测结果。

如图3所示,一些观测数据(红点)离模拟数据(灰点)很远,如P3井的BHP和P2井的气率。有些观测数据接近模拟点,但覆盖面不够,如P1井的BHP、P1井的气率和P3井的气率。P2井的BHP似乎没有被证伪,因为观察点在PC1和PC2图的先验点内。然而,在更高的维度上,它被证伪。这可以从PC2和PC3图中看出(图4)。

因此,我们定义的先验被证伪,也就是说,它被证明是不正确的。我们需要一个系统的方法来确定先验的问题。任何修改都需要增加不确定性,增加复杂性,或进行两者的组合。简单的局部修改,如对物性进行乘法调整,都是非贝叶斯的[40]。相反,我们将把这个问题表述为一个假设检验。我们将对先验进行修改,将其作为一个假设陈述,然后试图将其作为一个假设拒绝。

4.2. 假设1:当前不确定参数的分布不当

为了确定先验的问题是什么,我们使用DGSA来进行敏感性分析。结果(图5)显示,裂缝开度是最敏感的参数。PCA图(图4)显示,离观测点较近的点更可能有较大的裂缝开度值。这两个特征表明,断裂开度的分布可能

太窄。因此,我们将裂缝开度分布的上限从-1.3提高到-0.2,而下限仍为-2.2。

在增加先验的断裂开度范围后,我们重新进行蒙特卡罗和储层模拟,生成先验数据(附录A中的图S6)。图6显示了假设1的先验数据(蓝点)和初始先验(灰点)的比较。可以看出,假设1的数据点对观察的覆盖面更大。这个证据表明,假设1没有错。然而,尽管数据对观察的覆盖面比初始先验更多,但先验仍然是被证伪的。如图6所示,P2井的气率和P1井的BHP的观察点仍然在模拟点之外。因此,我们应该提出新的假设来解决先验问题。

4.3. 假设2:增加模型的复杂性-动态传导率

如图2所示,P1井的BHP历史曲线在生产过程中持续下降。然而,模拟的曲线更平坦,这意味着在模拟中井眼的流体供应比现实中要多得多。一个潜在的原因是,渗透率是依赖于压力的,也就是说,渗透率是随着储层压力而动态变化的。一些研究[9-10]对裂缝储层的这种现象进行了探究。我们在以前的先验定义中忽略了这一机制。Chen等[10]报道了裂缝渗透率随储层压力下降而变化的幂

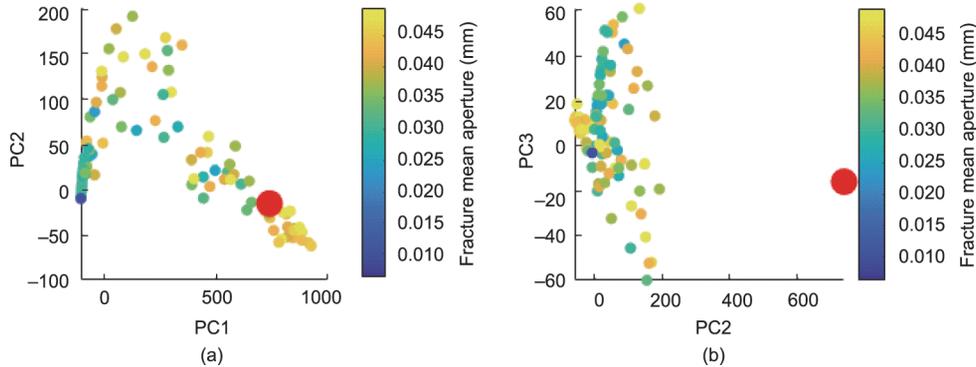


图4. P2BHP数据的PCA图。(a) PC1和PC2图;(b) PC2和PC3图。圆中圆点的颜色代表断裂开度值。红点是观测值的PC,而其他点是先验数据的PC。

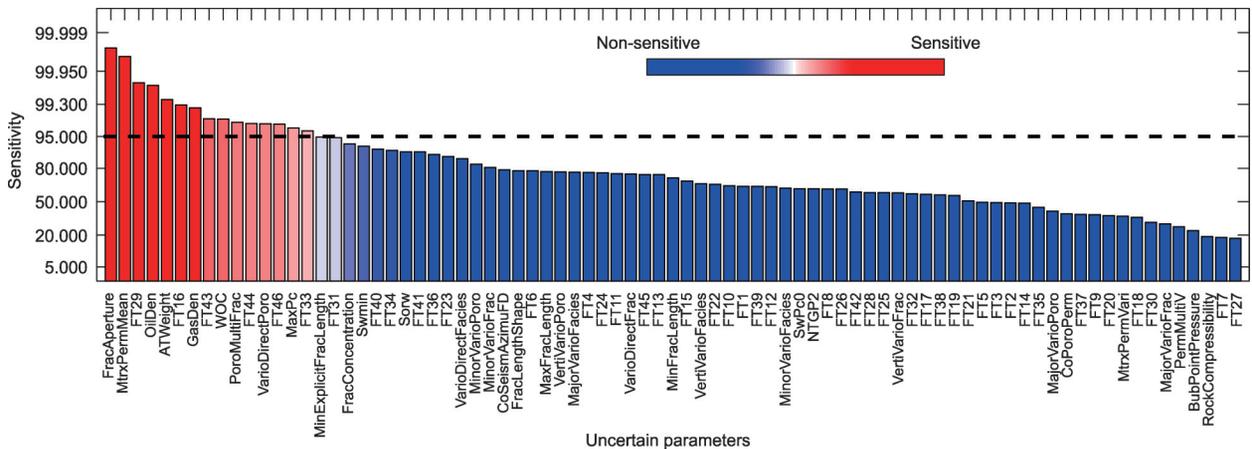


图5. P2井BHP数据的不确定参数的敏感度分析。

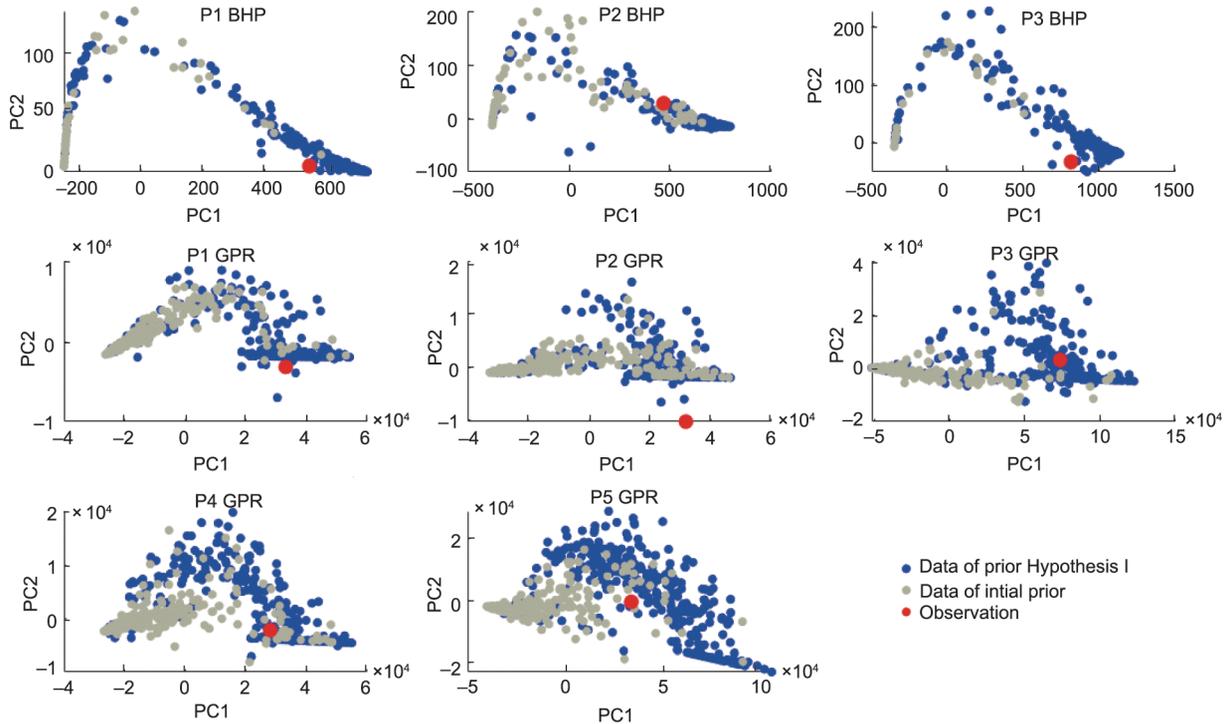


图6. 假设1的数据与初始先验的数据的比较。

律关系。该函数如下：

$$k = k_0 e^{3c_f \frac{v}{1-v} \alpha_b (p - p_0)} \quad (5)$$

式中， p_0 是参考压力； p 是储层压力； k_0 是 p_0 下的裂缝透率； c_f 是岩石压缩性； v 是泊松比； α_b 是Biot系数。

由于没有研究区域的泊松比和Biot系数的信息，我们定义了一个不确定参数DynamicTM（表2）来共同表示系数 $3c_f \frac{v}{1-v} \alpha_b$ 的值。在水库模拟中，我们使用动态透射率乘数（附录A中的图S7）来表示这一现象。

当动态透射率乘数被添加到模型中时，蒙特卡罗模拟结果在模拟曲线（附录A中的图S8）和PCA图（附录A中的图S9）上都有较大的观测覆盖率。DGSA的结果显示，动态透射率乘数对模拟数据很敏感（附录A中的图S10）。这些证据表明，该假设并没有错；应该在先验模型中考虑该假设。然而，仔细检查P2井的气率点（附录A中的图S9），发现观察结果仍在先验数据点之外。图S9中的生产曲线显示，P2井中天然气产量在2018年后有所增加，而模拟曲线则没有这一特征。P2井的BHP观测值在生产历史上有所下降。这说明我们建立的模型仍然不够复杂，不能反映生产历史中剩余的所有特征。造成这种现象的可能原因是压力供应不足，导致P1和P2井区域出现脱气现象。这是合理的，因为储层的高部位就在这个区域。

4.4. 假设3:增加模型的复杂性-减少生产者的压力供应

压力供应来自两个方面：底水和注入水。我们引入了两个额外的不确定参数，即断裂开度自上而下的递减程度[图7(a)]和底部区域的孔隙体积乘数[图7(b)]，来表示来自底部水的压力供应不确定性。在原始断层网络中[图7(c)]，可以看到生产井P1和P2与附近的注水井INJ1和INJ2没有被断层完全隔开。这意味着，尽管断层是完全密封的，但生产者和注水井之间仍然存在着一些联系。这可能是P1和P2井压力供应充足的一个原因。在此，我们定义了一个变量，代表我们是否使用修改后的断层网络。添加的参数列于表2，断层修改策略见图7(c)。

然后我们进行蒙特卡罗模拟，生成先验数据。模拟结果（附录A中的图S11）显示，P1和P2井的一些BHP曲线的表现与现场观测结果更加一致。更重要的是，P2井的一些产气速率曲线在2018年后开始增加，这在以前的先验数据中没有体现。PC1和PC2的先验数据点对观测数据有很好的覆盖（附录A中的图S12）。DGSA的结果（附录A中的图S13）也显示，假设3应该被包括在先验中。

至此，我们已经着手对低维的PC得分图进行了视觉检查。现在，我们包括了所有的PC维度，并使用一种系统的方法来证伪。我们采用一个单类SVM，即一个离群点检测方法，来进行证伪。其基本原理是，如果观测值相对于模拟观测值是一个离群点，那么先验就被证伪了。因

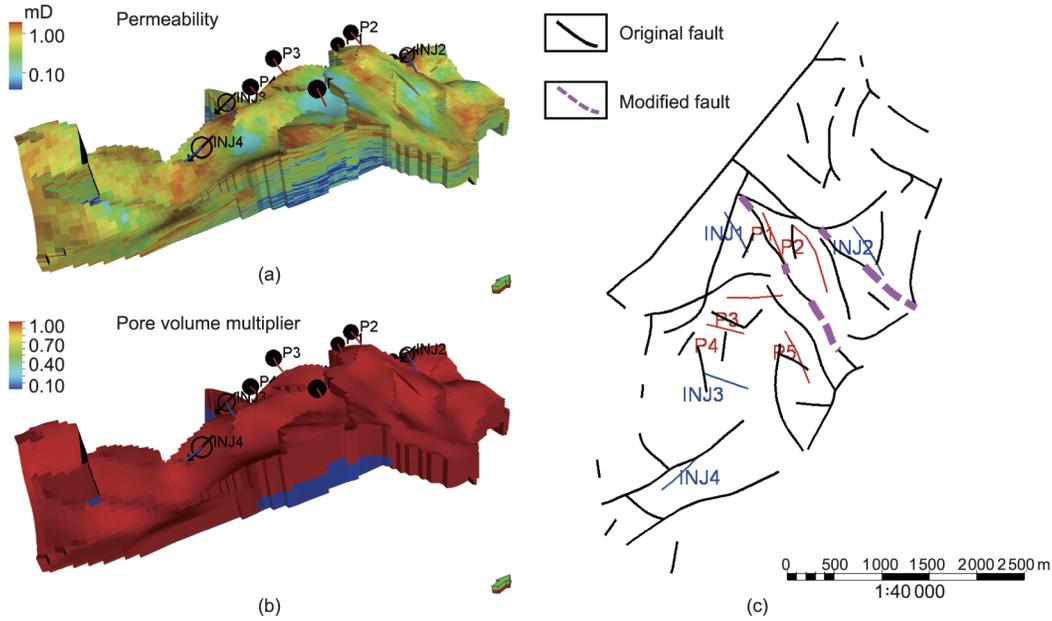


图7. 假设3中的模型修改。(a) 渗透率的一个实现，考虑到断裂开度从上到下的递减程度；(b) 区域3的孔隙体积乘数的一个实现；(c) 断层修改。INJ: 注入器。mD: 毫达西。

表2 假设2和假设3中先验的修改

	Parameter	Distribution	Description
Parameter added in hypothesis 2	DynamicTM	$U(0, 0.09)$	A larger value indicates that the transmissibility decreases more per draw-down of reservoir pressure
Parameter added in hypothesis 3	PermFracDRS	$U(-3, 0)$	Degree of decrease in fracture permeability from top to bottom
	FaultNetCase	Two geological scenarios	Whether or not the modified fault shown in Fig. 7 is used
	PoroVolMultiZone3	Two geological scenarios	Whether or not the pore volume multiplier is used for the bottom zone

为数据并不是高斯分布，所以我们采用单类SVM而不是基于马氏距离的方法[图8(a)]。

在单类SVM中，所有的模拟数据都被认为是训练样本，用于训练分类模型。然后，使用训练好的模型来计算观测分数。分数为负的观测值被认为是离群点，这意味着先验被证伪。图8(b)显示，当包含99%以上信息的10

个PC被包括在训练数据中时，观测得分仍然为正。这表明，先验没有被证伪。

4.5. 代理模型训练

贝叶斯定理允许我们通过计算后验分布来减少不确定性。在本研究中，我们采用ABC来计算后验分布，这需

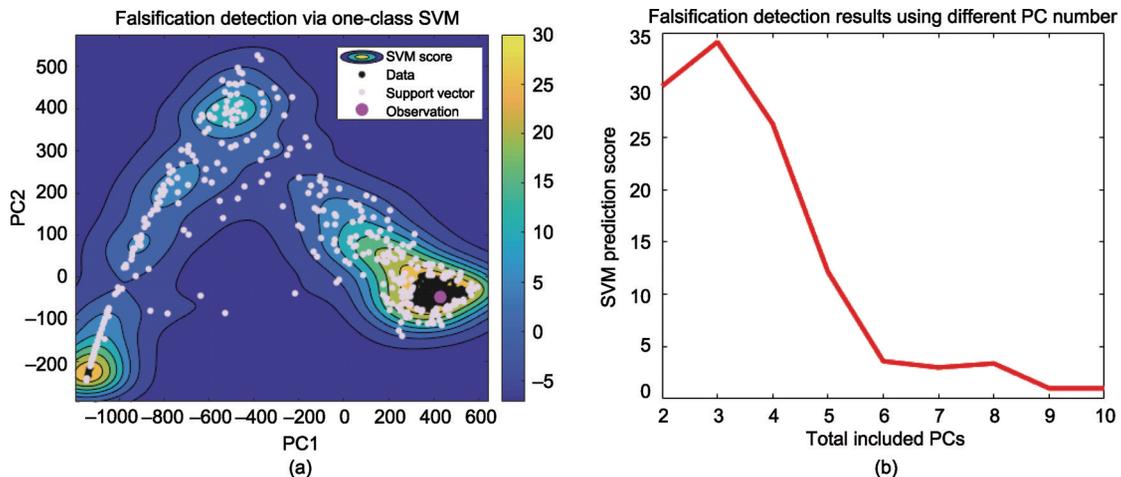


图8. 使用单类SVM的离群点检测。(a) 使用PC1和PC2的SVM结果；(b) 考虑到高维PC，使用SVM模型对观测结果的预测得分。

要足够的先验实现。因此，需要一个代理模型来建立参数和模拟结果之间的统计模型。在这里，我们采用了两种基于树的回归技术，即随机森林和梯度提升，来建立参数与模拟曲线和观测曲线距离之间的关系。

我们将所有的不确定参数作为回归模型的输入。目标输出（标签）是模拟与观测的不匹配；因此，我们的输出是模拟和观测值之间的距离。所有的数据曲线首先归一化到0和1之间，然后进行合并，以捕捉来自不同油井和不同数据类型的特征。

我们使用75%的样本作为训练数据集，25%作为测试集。图9显示，随机森林模型的回归性能和梯度提升模型的回归性能没有明显差异，相关系数分别为0.83 [图9 (a)]和0.85 [图9 (b)]。测试数据集在随机森林和梯度提升模型上的相关系数分别为0.72 [图9 (a)]和0.77 [图9 (b)]。梯度提升回归模型在训练和测试数据集上似乎都有更好的表现。然而，随机森林模型在距离值较小的样本上有更好的表现，这一点更重要，因为这些样本更接近于观察结果。图9 (c) 比较了随机森林模型和梯度提升模型的排名关系。可以看出，当考虑到离观察点较近的样本时，随机森林模型具有更好的性能。因此，我们选择随机森林模型作为代理模型。

4.6. 后验全局参数分布和数据

参数的后验分布是通过ABC方法计算的。首先，我们生成50 000个先验参数的实测值。然后通过随机森林模型计算预测的距离。在ABC程序中，我们使用值1.1作为选择后验样本的标准（附录A中的图S14）。图10比较了一些敏感参数（如裂缝开度、石油密度、P1和P2井区域的断层改造）的后验分布与它们的先验分布。敏感参数的不确定性有了明显的降低。

最后，为了验证是否能有效降低不确定性，我们使用更新的参数再次建立地质和模拟模型，然后进行实际的流量模拟。图11显示了使用更新后的参数的模拟曲线。结

果表明，整个工作流程是有效的，因为后验曲线更接近于观测曲线。为了进一步提高匹配度，可以拒绝那些被认为离生产历史太远的模拟算例。在匹配历史时应考虑到计算的复杂性。在本研究中，每个建模和模拟实现的平均时间约为30 min。因为我们只使用蒙特卡罗模拟而不是迭代匹配，所以可以使用并行计算来加快先验和后验模型的生成。

5. 讨论和结论

在一些实地案例研究中，先验的伪造一直是研究人员关注的问题[24,42,49]。在现场应用中，特别是对于我们知之甚少的案例，必须先验进行非伪造的说明。在本文的研究中，我们考虑了初始先验中的许多不确定性。即使如此，结果还是发现先验是被伪造的。许多关于历史匹配的论文都没有提到证伪的步骤。通常，实验做了特定修改但不会展示出来，只有研究到了成功地应用某种技术，才会被强调。然而，这通常不是工程师在实际案例中面临的最基本问题。

在去伪存真过程中，通过敏感性分析和降维，很容易发现当前不确定参数分布过窄的问题，如假设1中的断裂开度问题（第4节）。更具挑战性的部分是发现先验定义中被忽略的重要的不确定因素，而这需要大量的领域知识和经验。例如，在本研究中，我们认为初始先验中的断层位置和连接性是确定的。如果没有关于地震解释的领域知识，数据科学家或储层工程师会发现识别这种元素具有挑战性。另一种方法是与不同领域的专家进行深入讨论，以确定每个领域的不确定性，如地震解释、测井解释和生产管理。

考虑到建模和仿真的计算复杂性，我们生成了1000个实测数据，然后使用了一个代理模型。一个统计学上的代理模型总是会降低不确定性约简的性能。我们使用基于树

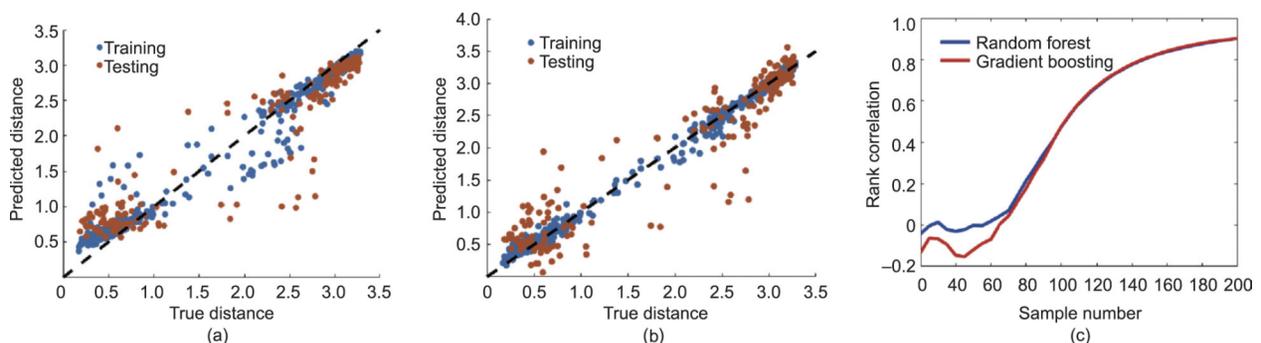


图9. 代理模型的性能。(a) 随机森林；(b) 梯度提升；(c) 随机森林和提升模型的等级相关性。(a)和(b)中蓝色的点代表训练数据集上的预测性能，而橙色的点代表测试数据集上的预测性能。

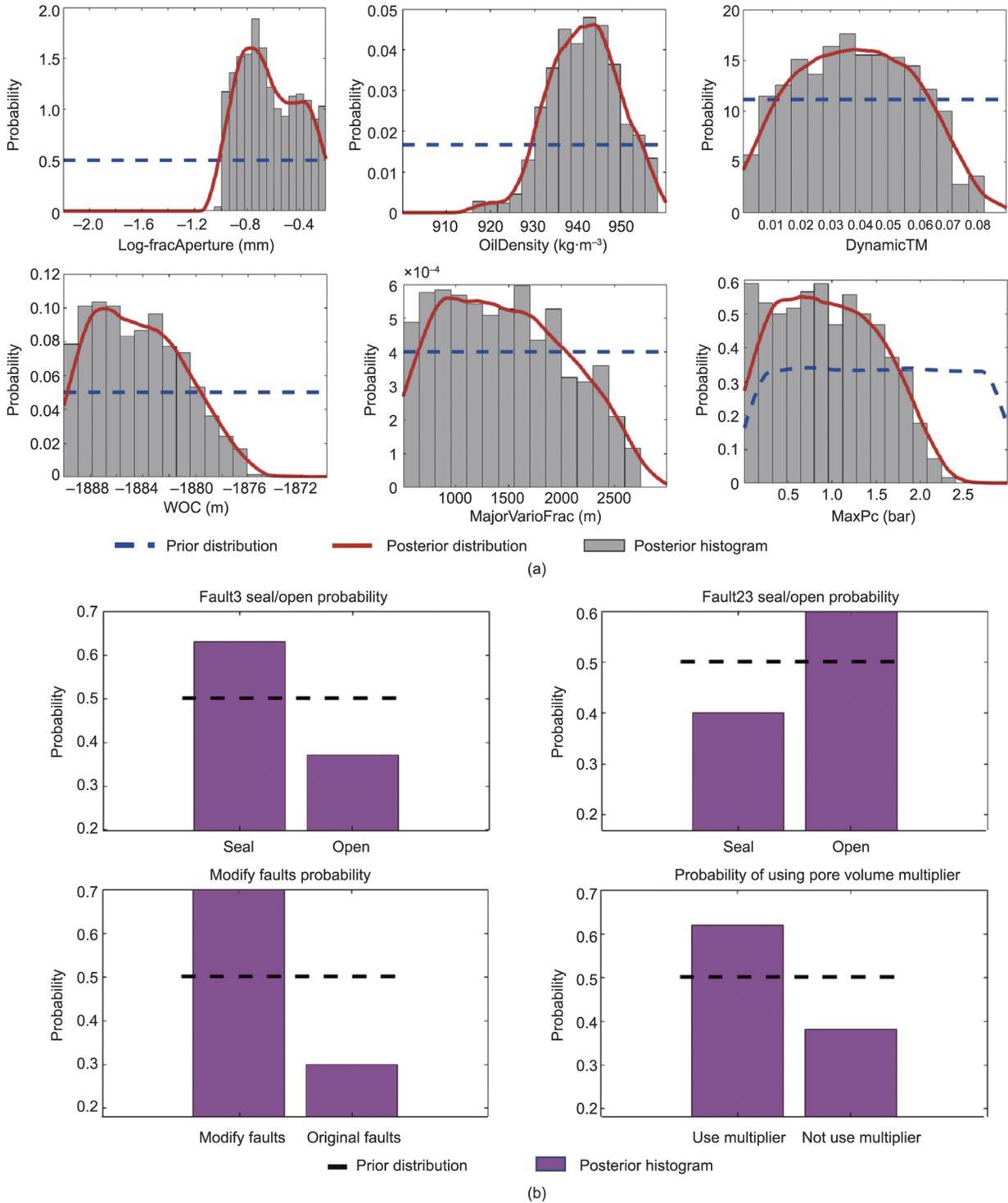


图10. 敏感连续参数 (a) 和分类敏感参数 (b) 的先验和后验分布。

的代理模型，因为一些输入参数是分类的。然而，基于树的回归模型牺牲了一些回归的准确性，因为它在每个节点中使用的是片状平均。如果地质模型和前向模拟时间可以容忍运行几万到几百万次实现，就不需要代理模型。

总之，我们提出了一种系统的方法来诊断真实裂缝储层案例的不确定性量化中先验的证伪问题。所提

出的方法整合了GSA、降维和蒙特卡罗抽样。我们证明了该方法能有效地识别先验的过小不确定性和模型复杂性的不足。更重要的是，该方法可以诊断出现场数据中存在的冲突，如本研究中的断层解释和生产历史，因为这种冲突经常发生在测量不同方面数据的现场地质研究中。

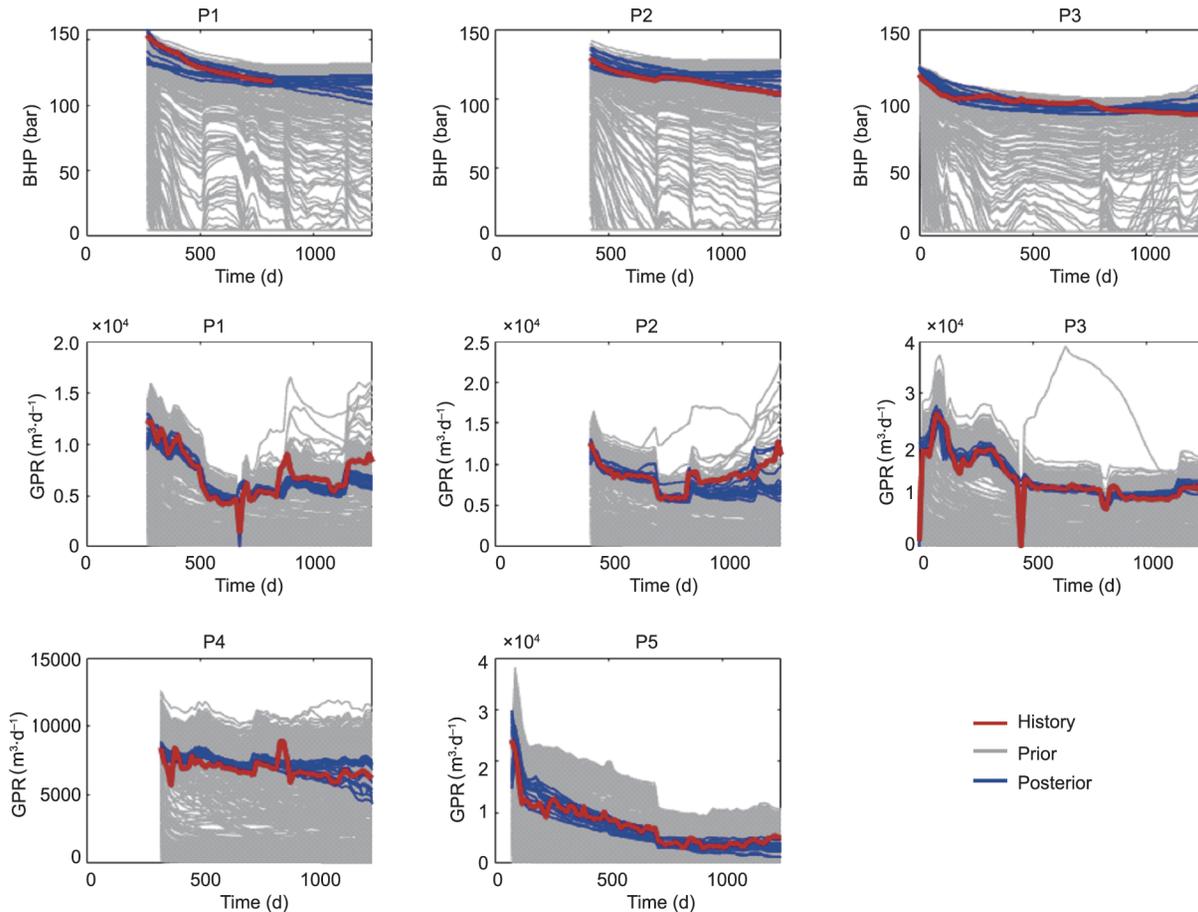


图 11. 使用后验参数的模拟数据曲线。

致谢

我们感谢国家留学基金管理委员会、中国地质大学(武汉)、特雷西能源科技有限公司和斯坦福地球资源预测中心对这项研究的支持。作者要特别感谢 Markus Zechner、Zhen Yin 和 Yizheng Wang 的技术讨论和支持。

Compliance with ethics guidelines

Junling Fang, Bin Gong, and Jef Caers declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2022.04.015>.

References

- [1] Bourbiaux B. Fractured reservoir simulation: a challenging and rewarding issue. *Oil Gas Sci Technol* 2010;65(2):227–38.
- [2] Azizmohammadi S, Matthai SK. Is the permeability of naturally fractured rocks scale dependent? *Water Resour Res* 2017;53(9):8041–63.
- [3] Nelson RA. Geological analysis of naturally fractured reservoirs. 2nd ed. Woburn: Gulf Publishing Company; 1985.
- [4] Turcott DL, Moores EM, Rundle JB. Super fracking. *Phys Today* 2014;67(8):34–9.
- [5] Bonnet E, Bour O, Odling NE, Davy P, Main I, Cowie P, et al. Scaling of fracture systems in geological media. *Rev Geophys* 2001;39(3):347–83.
- [6] Li Y, Shang Y, Yang P. Modeling fracture connectivity in naturally fractured reservoirs: a case study in the Yanchang formation. *Fuel* 2018;211:789–96.
- [7] Pieters DA, Graves RM. Fracture relative permeability: linear or non-linear function of saturation. In: *Proceedings of International Petroleum Conference and Exhibition of Mexico*; 1994 Oct 10–13; Veracruz, Mexico. 1994.
- [8] Schiozer D, Muñoz Mazo E. Modeling fracture relative permeability—what is the best option? In: *Proceedings of the 75th EAGE Conference & Exhibition incorporating SPE EUROPEC*; 2013 Jun 1013; London, UK. 2013.
- [9] Cho Y, Ozkan E, Apaydin O. Pressure-dependent natural-fracture permeability in shale and its effect on shale-gas well production. *SPE Reserv Eval Eng* 2012;16(02):216–28.
- [10] Chen D, Pan Z, Ye Z. Dependence of gas shale fracture permeability on effective stress and reservoir: model match and insights. *Fuel* 2015;139:383–92.
- [11] Mukerji T, Jorstad A, Avseth P, Markov G, Granli JR. Mapping lithofacies and pore-fluid probabilities in a North Sea reservoir: seismic inversions and statistical rock physics. *Geophysics* 2001;66(4):988–1001.
- [12] Pedersen SI, Randen T, Sonneland L, Steen O. Automatic fault extraction using artificial ants. *SEG Tech Prog Exp Abstr* 2002:512–5.
- [13] Leite EP, Vidal AC. 3D porosity prediction from seismic inversion and neural

- networks. *Comput Geosci* 2011;37(8):1174–80.
- [14] Xu K, Sun Z. A new fracture detection method based on full-azimuth anisotropic migration and shaping regularization. *J Geophys Eng* 2018;15(4):1624–37.
- [15] Guerriero V, Iannace A, Mazzoli S, Parente M, Vitale S, Giorgioni M. Quantifying uncertainties in multi-scale studies of fractured reservoir analogues: implemented statistical analysis of scan line data from carbonate rocks. *J Struct Geol* 2010;32(9):1271–8.
- [16] Wilson CE, Aydin A, Karimi-Fard M, Durlafsky LJ, Sagy A, Brodsky EE, et al. From outcrop to flow simulation: constructing discrete fracture models from a LIDAR survey. *AAPG Bull* 2011;95(11):1883–905.
- [17] Agada S, Geiger S, Elsheikh A, Oladyskhin S. Data-driven surrogates for rapid simulation and optimization of WAG injection in fractured carbonate reservoirs. *Petrol Geosci* 2016;23(2):270–83.
- [18] Williams JH, Johnson CD. Acoustic and optical borehole-wall imaging for fractured-rock aquifer studies. *Appl Geophys* 2004;55(1–2):151–9.
- [19] Kovac KM, Lutz SJ, Drakos PS, Byersdorfer J, Robertson-tait A. Borehole image analysis and geological interpretation of selected features in well DP 27-15 at desert peak, Nevada: pre-simulation evaluation of an enhanced geothermal system. In: *Proceedings of Thirty-Fourth Workshop on Geothermal Reservoir Engineering*; 2009 Feb 9–11; California, USA. 2009.
- [20] Caers J. History matching under training-image-based geological model constraints. *SPE J* 2003;8(3):218–26.
- [21] Oliver D, Chen Y. Recent progress on reservoir history matching: a review. *Comput Geosci* 2011;15(1):185–221.
- [22] Ghaedi M, Masihi M, Heinemann ZE, Ghazanfari MH. History matching of naturally fractured reservoirs based on the recovery curve method. *J Petrol Sci Eng* 2015;126:211–21.
- [23] Athens ND, Caers JK. A Monte Carlo-based framework for assessing the value of information and development risk in geothermal exploration. *Appl Energy* 2019;256:113932.
- [24] Yin Z, Strebelle S, Caers J. Automated Monte Carlo-based quantification and updating of geological uncertainty with borehole data (AutoBEL v1.0). *Geosci Model Dev* 2020;13(2):651–72.
- [25] Aydin O, Caers J. Quantifying structural uncertainty on fault networks using a marked point process within a Bayesian framework. *Tectonophysics* 2017;712–713:101–24.
- [26] Bonet-Cunha L, Olive DS, Rednar RA, Reynolds AC. A hybrid Markov chain Monte Carlo method for generating permeability fields conditioned to multiwell pressure data and prior information. *SPE J* 1998;3(3):261–71.
- [27] Tjelmeland H, Eidsvik J. On the use of local optimizations within Metropolis-Hastings updates. *J R Stat Soc B* 2004;66(2):411–27.
- [28] Romary T. Integrating production data under uncertainty by parallel interacting Markov chains on a reduced dimensional space. *Comput Geosci* 2009;13(1):103–22.
- [29] Wen X, Chen W. Real-time reservoir model updating using ensemble Kalman filter with confirming option. *SPE J* 2006;11(4):431–42.
- [30] Haugen V, Naevdal G, Natvik LJ, Evensen G, Berg AM, Flornes KM. History matching using the Ensemble Kalman Filter on a North Sea Field case. *SPE J* 2008;13(4):382–91.
- [31] Nejadi S, Trivedi JJ, Leung J. History matching and uncertainty quantification of discrete fracture network models in fractured reservoirs. *J Petrol Sci Eng* 2017;152:21–32.
- [32] Evensen G, van Leeuwen PJ. An ensemble Kalman smoother for nonlinear dynamics. *Mon Weather Rev* 2000;128(6):1852–67.
- [33] Chen Y, Oliver DS. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math Geosci* 2012;44(1):1–26.
- [34] Gu Y, Oliver DS. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE J* 2007;12(4):438–46.
- [35] Emerick AA, Reynolds AC. Ensemble smoother with multiple data assimilation. *Comput Geosci* 2013;55:3–15.
- [36] Satija A, Caers J. Direct forecasting of subsurface flow response from nonlinear dynamic data by linear least-squares in canonical functional principal component space. *Adv Water Resour* 2015;77:69–81.
- [37] Satija A, Scheidt C, Li L, Caers J. Direct forecasting of reservoir performance using production data without history matching. *Comput Geosci* 2017;21(2):315–33.
- [38] Sun W, Hui MH, Durlafsky LJ. Production forecasting and uncertainty quantification for naturally fractured reservoirs using a new data-space inversion procedure. *Computat Geosci* 2017;21(5–6):1443–58.
- [39] Sun W, Durlafsky LJ. A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. *Math Geosci* 2017;49(6):679–715.
- [40] Caers J. Bayesianism in the geosciences. In: Daya Sagar BS, Cheng Q, Agterberg F, editors. *Handbook of mathematical geosciences*. Chem: Springer; 2018.p. 527–66.
- [41] Scheidt C, Li L, Caers J. *Quantifying uncertainty in subsurface systems*. Hoboken: Wiley; 2018.
- [42] Hermans T, Nguyen F, Klepikova M, Dassargues A, Caers J. Uncertainty quantification of medium-term heat storage from short-term geophysical experiments using Bayesian evidential learning. *Water Resour Res* 2018;54(4):2931–48.
- [43] Slotte PA, Smorgrav E. Response surface methodology approach for history matching and uncertainty assessment of reservoir simulation models. In: *Proceedings of the 70th EAGE Conference & Exhibition*; 2008 Jun 9–12; Rome, Italy. 2008.
- [44] Castellini A, Gross H, Zhou Y, He J, Chen W. An iterative scheme to construct robust proxy models. In: *Proceedings of the 12th European Conference on the Mathematics of Oil Recovery*; 2010 Sep 6–9; Oxford, UK. 2010.
- [45] Friedmann F, Chawathe A, Larue DK. Assessing uncertainty in channelized reservoirs using experimental designs. *SPE Reservoir Eval* 2003;6(4):264–74.
- [46] Aulia A, Jeong D, Saaid IM, Kania D, Shuker MT, El-Khatib NA. A random forests-based sensitivity analysis framework for assisted history matching. *J Petrol Sci Eng* 2019;181:106237.
- [47] Brantson ET, Ju B, Omisore BO, Wu D, Selase AE, Liu N. Development of machine learning predictive models for history matching tight gas carbonate reservoir production profiles. *J Geophys Eng* 2018;15(5):2235–51.
- [48] Alfonzo M, Oliver DS. Evaluating prior predictions of production and seismic data. *Computat Geosci* 2019;23(6):1331–47.
- [49] Pradhan A, Mukerji T. Seismic Bayesian evidential learning: estimation and uncertainty quantification of sub-resolution reservoir properties. *Comput Geosci* 2019;24(3):1121–40.
- [50] Oda M. Permeability tensor for discontinuous rock masses. *Geotechnique* 1985;35(4):483–95.
- [51] Akin S. Estimation of fracture relative permeabilities from unsteady state core floods. *J Petrol Sci Eng* 2001;30(1):1–14.
- [52] Scheidt C, Jeong C, Mukerji T, Caers J. Probabilistic falsification of prior geologic uncertainty with seismic amplitude data: application to a turbidite reservoir case. *Geophysics* 2015;80(5):M89–M.
- [53] Noumir Z, Honeine P, Richard C. On simple one-class classification methods. In: *Proceedings of the IEEE International Symposium on Information Theory*; 2012 Jul 1–6; Cambridge, MA, USA. 2012.
- [54] Saltelli A, Ratto M, Andrews T, Campolongo F, Cariboni J, Gatelli D, et al. *Global Sensitivity Analysis*. New York: John Wiley & Sons, Ltd; 2008.
- [55] Fenwick D, Scheidt C, Caers J. Quantifying asymmetric parameter interactions in sensitivity analysis: application to reservoir modeling. *Math Geosci* 2014;46(4):493–511.
- [56] Park J, Yang G, Satija A, Scheidt C, Caers J. DGSA: a Matlab toolbox for distancebased generalized sensitivity analysis of geoscientific computer experiments. *Comput Geosci* 2016;97:15–29.
- [57] Spear RC, Hornberger GM. Eutrophication in peel inlet-II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Res* 1980;14(1):43–9.
- [58] Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. *Genetics* 2002;162(4):2025–35.
- [59] Sadegh M, Vrugt JA. Approximate Bayesian Computation using Markov Chain Monte Carlo simulation: DREAM (ABC). *Water Resour Res* 2014; 50(8):6767–87.
- [60] Barthelme S, Chopin N. ABC-EP: expectation propagation for likelihoodfree Bayesian computation. In: *Proceedings of the 28th International Conference on Machine Learning*. 2011 Jun 28–Jul 2; Washington, USA. 2011.
- [61] Trehan S, Carlberg KT, Durlafsky LJ. Error modeling for surrogates of dynamical systems using machine learning. *Int J Numer Methods Eng* 2017; 112(12):1801–27.
- [62] Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. 2nd Edition. Springer Series in Statistics; 2009.
- [63] Schapire RE. The strength of weak learnability. *Mach Learn* 1990; 5(2):197–227.
- [64] Breiman L. Bagging predictors. *Mach Learn* 1996;24(2):123–40.
- [65] Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2000;29(5):1189–232.
- [66] Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002;38(4):367–78.
- [67] Breiman L. Random forests. *Mach Learn* 2001;45(1):5–32.