arXiv:2304.06310v1 [cs.LG] 13 Apr 2023

序贯蒙特卡洛在虚拟流量计校准中的应用

安德斯·T. 桑德内斯*[1,2]，比亚内·格里姆斯塔德[1,3]，和奥德·科尔比约恩森[2]

[1]Solution Seeker AS，挪威奥斯陆[2]奥斯陆大学数学系，挪威奥斯陆[3]挪威科技大学工程控制系，挪威特隆赫姆

2023年4月14日

摘要

软测量仪因能够以较低成本、只需对油气资产进行少量干预，便可提供关键过程变量的估计值，正日益受到青睐。在石油和天然气生产中，虚拟流量计（VFM）是一种常用的软测量仪，旨在实时估计多相流量。虚拟流量计基于模型，而这些模型需要校准。校准高度依赖具体应用，这既是因为模型种类繁多，也因为可用测量值各不相同。最准确的校准方法是针对井测试仔细调节虚拟流量计参数，但这可能需要大量工作，而且并非所有油井都有频繁的油井测试数据可用。本文提出一种校准方法，该方法基于生产分离器提供的测量值，并假设观测到的流量应等于各个油井流量之和。这使我们能够持续联合校准虚拟流量计。该方法采用序贯蒙特卡洛（SMC）推断每口油井的调节因子和流体组成。我们使用合成数据和真实数据，在一个包含十口油井的案例中对该方法进行了测试。结果令人鼓舞，该方法无需依赖井测试便能给出合理的参数估计值。然而，我们也发现并讨论了一些挑战，尤其是与过程噪声以及如何处理数据质量变化有关的问题。

1 引言

第四次工业革命在数据采集、存储和可用性方面取得了令人瞩目的进展（Kagermann and Wahlster, 2022）。大量工业数据正以实时方式采集和存储。随着信息量不断增长，人工数据分析师越来越依赖机器智能系统来处理和筛选数据（Kagermann and Wahlster, 2022）。软测量仪是运行于这些数据流之上的预测模型，可为工程师和操作人员提供洞察与态势感知（Kadlec et al., 2009）。软测量仪是结合了配套软件的数学模型，因此安装和维护成本较低，因为无需在现场增添额外硬件。这些模型可以是数据驱动的、基于物理原理和过程知识的，也可以结合这两种方法。

在石油和天然气生产领域，数据基础设施已经相当完善，但实时数据的利用仍然有限（Lu et al., 2019）。虚拟流量计（VFM）是一项受到广泛研究的应用（Bikmukhametov and J¨aschke, 2020）。虚拟流量计使用数学模型和常见的测量值来估计天然气、石油和水的流量。文献中介绍了多种建模策略，包括物理模型（Sch¨uller et al., 2003）、机器学习模型（Sandnes et al., 2021）和混合模型（Hotvedt et al., 2022a）。虚拟流量计传统上基于系统中的一个或多个压降，例如节流阀两端或管道长距离区段上的压降。

虚拟流量计最常用于估算单口油井的流量。理想情况下，虚拟流量计能够同时估算天然气、石油和水的流量，但可用的测量值通常

*通讯作者：anders@solutionseeker.no

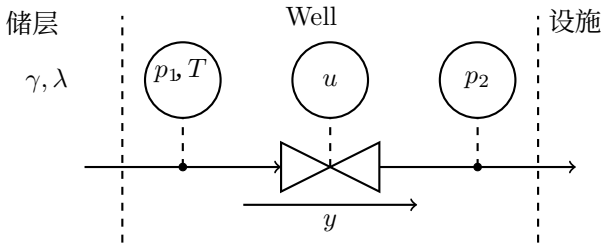


图1：油井拓扑结构。测量值包括节流阀开度 u、压力 p_1 和 p_2，以及温度 T。流经油井的流量 y 未知。流体来自储层，其中天然气、石油和水的混合比例未知。混合物由气体分数 γ 和油因子 λ 描述。

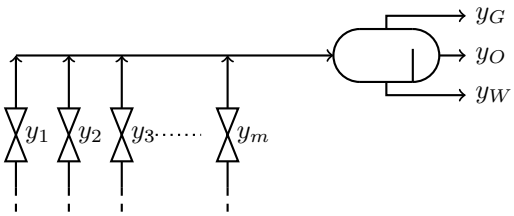


图2：共用分离器的油气资产，包含 m 口油井。各油井产出的流体汇入共用分离器，并在此测量流量。例如，各油井的天然气流量相加后等于分离器的天然气流量，$y_G = \sum_{j=1}^{m} y_{j,\,G}$，石油和水的流量也同样如此。

不足以实现这一点。因此，通常会假设流体组成部分已知或完全已知。这样便可将问题简化为仅推断总流量（Bikmukhametov 和 J¨aschke，2020）。例如，考虑图1所示的一口油井，其仪表配置可连续实时提供节流阀、压力和温度测量值。仅凭这四项测量值，要推断三种流体各自的流量会很困难。相反，可以估算天然气、石油和水的总和。这通常需要对混合物的组分作出假设。

无论采用何种测量和建模方法，都需要及时更新模型，以确保令人满意的性能。这是由于系统发生了变化，其中包括流体组成、流体性质和操作条件（Bikmukhametov 和 J¨aschke，2020）。已有多种方法被提出，而这些方法会根据模型类型和可用于校准的数据类型而显著不同。Hotvedt 等人（2022b）应用在线学习，基于 油井测试数据 更新混合模型和数据驱动模型。油井测试是在油井单独通过测试分离器生产时采集的数据，此时该油井的天然气、石油和水流量在物理分离后直接测量。基于井测试的校准当然非常精确。然而，由于开展井测试需要耗费大量时间，且多个油井共用同一测试设施，因此井测试的频率可能会被限制为每月仅进行几次（Monteiro 等人，2020）。

另一种校准策略是利用一系列持续可用的测量值（通常是压力和温度），并结合卡尔曼滤波器和粒子滤波器等状态空间方法（Adukwu et al., 2022）。这些方法不使用流量的直接测量值来调整模型。Lorentzen et al.（2003）利用集合卡尔曼滤波器，通过一系列高频压力测量值校准动态模型。He et al.（2020）应用粒子滤波器，校准同时包含数据驱动组件和机理模型组件的混合模型。

与虚拟流量计校准类似的问题是流量分配 （Kanshio，2020）。分配关注的是观测到的混合流量，它是多口油井产量的总和。任务是将观测到的流量分配给各个贡献者。考虑图2所示的油气资产，其中 m油井产出流体并汇入同一台分离器。生产分配就是确定各口油井对分离器测得流量的贡献。如果每口油井都配有虚拟流量计，则可利用这些流量计算分配因子（Kanshio，2020）。这样得到的分配结果符合系统的质量平衡（Badings 和 van Putten，2020）。

本文提出一种基于序贯蒙特卡洛（SMC）的虚拟流量计校准方法。我们尝试结合生产数据和井测试，估算多口油井的流体组成和调节因子。生产数据是所有油井经分离后测得的联合产量，如图2所示。生产数据会持续采集。我们的目标是在缺少频繁井测试的情况下校准虚拟流量计。

本文剩余部分的结构如下。第2节介绍序贯蒙特卡洛的必要背景。第3节详细描述该问题，包括需要推断的量以及可用数据。第4节使用真实数据和合成数据，对一项包含十口油井的油气资产开展实证研究。第5节总结研究结果并得出结论。

2 状态空间模型与序贯蒙特卡洛

SMC 是一类递归贝叶斯滤波方法（Chopin et al., 2020）。虽然它们通常适用于蒙特卡洛模拟，但在用于推断马尔可夫状态空间模型的潜变量时，效率尤其高。设 θ_t 为状态，$\mathcal{D}_t = \{（y_t,\ x_t）\}$ 为时间步 t 的观测值，其中 $t = 0,\ 1,\ \ldots,\ n$。马尔可夫状态空间模型由一个先验分布定义。

$$p(\theta_0), \tag{1}$$

一个转移分布

$$p(\theta_t|\theta_{t-1}), \tag{2}$$

对于 $t > 0$ 有效，以及观测值的似然

$$p(y_t|\theta_t, x_t). \tag{3}$$

所关注的滤波问题是后验分布 $p(\theta_t|\mathcal{D}_{0:t})$，其中符号 $\mathcal{D}_{0:t}$ 指代元素集合 $\{\mathcal{D}_0, \ldots \mathcal{D}_t\}$。将完整的状态序列 $\theta_{0:t}$ 纳入考虑后，滤波问题可以递归地写为

$$p(\theta_{0:t}|\mathcal{D}_{0:t}) = \frac{p(y_t|\theta_t, x_t)p(\theta_t|\theta_{t-1})}{p(\mathcal{D}_t|\mathcal{D}_{0:t-1})} p(\theta_{0:t-1}|\mathcal{D}_{0:t-1}), \tag{4}$$

当新数据到来时，可利用这一点高效地更新后验分布（Doucet 等，2001）。

如果公式（1）至（3）中的状态空间模型各组成部分均为线性且具有加性高斯噪声，则可以通过解析方法推导出滤波器的闭式解。这就是著名的卡尔曼滤波器 （Doucet 等，2001）。对于大多数问题，闭式解并不存在，因此必须对后验分布进行近似。蒙特卡洛方法通过从后验分布中生成样本，再利用这些样本近似感兴趣的量来实现这一点。

序贯蒙特卡洛是一种抽样方案，允许随着观测值的出现依次处理它们。它基于重要性抽样：从提议分布 中抽取样本，并根据真实的目标分布为样本赋权（Doucet 等，2001）。我们采用算法1中所述的序贯蒙特卡洛方法。有关该算法及其变体的深入介绍，请参阅 Chopin 等（2020）。

序贯蒙特卡洛的主要设计要素是构建提议分布。一种方便且常用的方法是直接从转移密度 $q(\theta_t|\theta_{t-1}, y_t, x_t) = p(\theta_t|\theta_{t-1})$ 中采样。这称为**自助滤波器** （Chopin 等，2020）。自助提议分布易于实现，但不考虑观测数据。如果似然信息量很大，这可能导致样本退化（Doucet 等，2001）。可以通过提高观测中的噪声水平来缓解这一问题，但这会引入一个额外的调节参数。引导提议分布试图生成能够考虑数据的样本，$q(\theta_t|\theta_{t-1}, y_t, x_t) \approx p(y_t|\theta_t, x_t)p(\theta_t|\theta_{t-1})$，以期提高样本效率。不过，这些提议分布可能较难设计。

本研究将使用有效样本量来衡量样本效率。它试图反映在考虑样本重要性的情况下，实际获得了多少粒子。时间 t 时的有效样本量计算为 $\mathrm{ESS}(W_t^{1:N}) = 1/(\sum_{i=1}^{N}(W_{(i)\,t})^2)$，其取值介于 1 和 N 之间（Chopin 等，2020）。我们将报告有效样本量与样本数量的比值。

算法1 引导粒子滤波器。上标表示粒子编号。所有操作都必须针对每个粒子执行，$i = 1, \ldots, N$。

$\theta_0^{(i)} \sim q(\theta_0 | y_0, \ x_0)$ ▷ 从提议先验分布中采样。

$w_0^{(i)} = \frac{p(y_0|\theta_0, x_0)p(\theta_0)}{q(\theta_0|y_0, \ x_0)}$ ▷ 计算权重。

$W_0^{(i)} = w_0^{(i)} / \sum_{k=1}^{N} w_0^{(k)}$ ▷ 归一化权重。

针对 $t = 1, 2, \cdots$ **执行**

$A_t^{1:N} \leftarrow$ 重采样$(W_{t-1}^{(i)})$ ▷ 每一步都进行重采样。

$\theta_t^{(i)} \sim q(\theta_t | \theta_{t-1}^{(A_t^t)}, y_t, x_t)$ ▷ 从转移提议分布中采样。

$w_t^{(i)} = \frac{p(y_t|\theta_t, x_t)p(\theta_t|\theta_{t-1}^{(A_t^t)})}{q(\theta_t|\theta_{t-1}^{(A_t^t)}, y_t, x_t)}$ ▷ 计算权重。

$W_t^{(i)} = w_t^{(i)} / \sum_{k=1}^{N} w_t^{(k)}$ ▷ 归一化权重。

结束循环

3 问题描述

我们考虑一组 m 油井，这些油井向共用分离器设施生产，如图2所示。这些油井的仪表配置如图1所示。对于油井 j 在时间 t，测量值收集于记为$x_{j,t} \in \mathbf{R}^4$的向量中。这些测量值包括节流阀开度、压力和温度。所有油井产出天然气、石油和水的混合物。来自一口油井的流量记为$y_{j,t}^{\top} = \begin{bmatrix} y_{j,t,G} & y_{j,t,O} & y_{j,t,W} \end{bmatrix}$。令$y_{j,t,T} = y_{j,t,G} + y_{j,t,O} + y_{j,t,W}$ 表示该油井的总流量，并令$\phi_{j,t} = y_{j,t}/y_{j,t,T}$ 表示流体组成。

我们考虑的虚拟流量计只能估算流经节流阀的总流量，并且需要了解流体组成。它们的形式为 $y_{j,t,T} = f_j(x_{j,t}, \phi_{j,t}) + v_{j,t}$，其中误差项为 $v_{j,t}$。我们的目标是推断虚拟流量计所需的未知组分 $\phi_{j,t}$。此外，虚拟流量计还需要通过调节因子进行校准，该因子用于缩放预测值。令 $\beta_{j,t} \in \mathbf{R}$ 为油井 j 在时间 t的调节因子。多相流量因此表示为

$$y_{j,t} = \beta_{j,t}\phi_{j,t}f_j(x_{j,t}, \phi_{j,t}) + e_{j,t}, \tag{5}$$

$$e_{j,t} \sim \mathcal{N}(0, \Sigma_{j,t}). \tag{6}$$

虚拟流量计本身是一个已知函数 f_j：$\mathbb{R}^4 \times [0,1]^3 \mapsto \mathbb{R}$，给定测量值和组分后即可自由求值。关于函数 f_j 的更多细节见附录A。

为进行推断，我们可以获取来自分离器的测量值。在分离器处测得的流量 y_t，是所有油井生产量的总和，

$$y_t = \sum_{j=1}^{m} y_{j,t} + \epsilon_t, \tag{7}$$

其中 $\epsilon_t \sim \mathcal{N}(0, \Sigma_{\epsilon,t})$。对于每口油井，我们都有测量值 $x_{j,t}$。完整的数据集表示为 $\mathcal{D}_{1:n} = \{(y_t, X_t)\}_{t=1}^n$，其中 $X_t = \{x_{1,t}, \ldots, x_{m,t}\}$。如果在某个时间步 t，除一口油井外的所有油井流量均为零，则该数据点为井测试。否则，该数据点为生产数据点。

目标是校准虚拟流量计。这意味着根据截至时间步 t 的数据，为所有油井提供调节因子 $\beta_{j,t}$ 和流体组成 $\phi_{j,t}$ 的估计值。组分向量 $\phi_{j,t}$有两个自由度，可以用气体分数进行参数化。

$$\gamma_{j,t} = \frac{y_{j,t,G}}{y_{j,t,G} + y_{j,t,O} + y_{j,t,W}} \tag{8}$$

以及油因子

$$\lambda_{j,t} = \frac{y_{j,t,O}}{y_{j,t,O} + y_{j,t,W}}. \tag{9}$$

其中 $\gamma_{j,t}$ 和 $\lambda_{j,t}$ 均取值于单位区间。这些量与组分的关系为

$$\phi_{j,t} = \begin{bmatrix} \gamma_{j,t} \\ (1-\gamma_{j,t})\lambda_{j,t} \\ (1-\gamma_{j,t})(1-\lambda_{j,t}) \end{bmatrix},$$

油井 j 在时间 t 的参数集记为 $\Theta_{j,t} = \{\beta_{j,t}, \gamma_{j,t}, \lambda_{j,t},\}$，而时间 t 的完整参数集为 $\Theta_t = \{\Theta_{1,t}, \ldots, \Theta_{m,t},\}$。

3.1 状态空间模型

滤波问题是从 $p(\Theta_t|\mathcal{D}_{1:t})$ 中采样，其中 $t = 1, \ldots, n$。要将序贯蒙特卡洛框架应用于此问题，我们需要指定先验、状态转移密度以及观测值的似然。

对于状态转移，假设组分和调节因子均为马尔可夫链，大部分时间保持不变，但有小概率发生变化。为对这些阶跃变化进行建模，引入了参数增广变量 $Z_t = \{z_{1,t}, \ldots, z_{m,t}\}$。油井 j 在时间 t 发生变化的情况表示为

$$z_{j,t} \sim \text{Bernoulli}(p_z). \tag{10}$$

每口油井使用一个变量，这意味着在任意给定时间，三个油井参数要么都可以变化，要么都不能变化。如果发生变化，则使用截断正态分布对转移进行建模。令 $\mathcal{N}_a^b$ 表示截断于区间 $[a, b]$ 内的正态分布。特殊情况 $\mathcal{N}_0^\infty$ 为半正态分布。调节因子的先验和转移方程为

$$\beta_{j,0} \sim \mathcal{N}_0^\infty(\mu_{\beta,0}, \sigma^2_{\beta,0}), \tag{11}$$

$$\beta_{j,t} = (1 - z_{j,t})\beta_{j,t-1} + z_{j,t}\tilde{\beta}_{j,t}, \tag{12}$$

$$\tilde{\beta}_{j,t} \sim \mathcal{N}_0^\infty(\beta_{j,t-1}, \sigma^2_\beta), \tag{13}$$

组成参数采用相同的结构，但截断在单位区间内，

$$\tilde{\gamma}_{j,t} \sim \mathcal{N}_0^1(\gamma_{j,t-1}, \sigma^2_\gamma), \tag{14}$$

$$\tilde{\lambda}_{j,t} \sim \mathcal{N}_0^1(\lambda_{j,t-1}, \sigma^2_\lambda). \tag{15}$$

方程（10）–（15）得出转移密度

$$p(\Theta_t, Z_t|\Theta_{t-1}, Z_{t-1}) = p(\Theta_t|\Theta_{t-1}, Z_t)p(Z_t). \tag{16}$$

似然项 $p(y_t|\Theta_t,\ X_t)$ 通过结合方程（7）和方程（5）构造，得到以下表达式

$$y_t = \sum_{j=1}^{m} [\beta_{j,t}\phi_{j,t}f_j(x_{j,t}, \phi_{j,t}) + e_{j,t}] + \epsilon_t. \tag{17}$$

似然不依赖于方程（16）中引入的增广变量，因此得到的似然可简单表示为

$$y_t \sim \mathcal{N}\left(\bar{y}_t, \Sigma_t\right), \tag{18}$$

$$\bar{y}_t = \sum_{j=1}^{m} \phi_{j,t}\beta_{j,t}f_j(x_{j,t}, \phi_{j,t}), \tag{19}$$

$$\Sigma_t = \Sigma_{\epsilon,t} + \sum_{j=1}^{m} \Sigma_{j,t}. \tag{20}$$

各个误差项 $\Sigma_{j,\ t}$ 均随时间变化。如果某口油井在时间 t 关闭，则其误差项不会对观测产生贡献。

3.2 数据

我们的数据集包含来自某油气资产的一年数据，该资产有十口油井。资产配有两台分离器，一台用于油井测试，另一台用于主生产线。我们主要使用生产数据进行推断，并使用油井测试数据进行性能评估。数据随时间的分布概况如图3所示，待估计的量汇总于表1。由表1可知，我们看到有些油井在整个时期内相当稳定，而另一些油井的油因子则持续变化。这些油井彼此也都相当相似。另请注意，有一口油井在数据集的前半段没有进行测试。

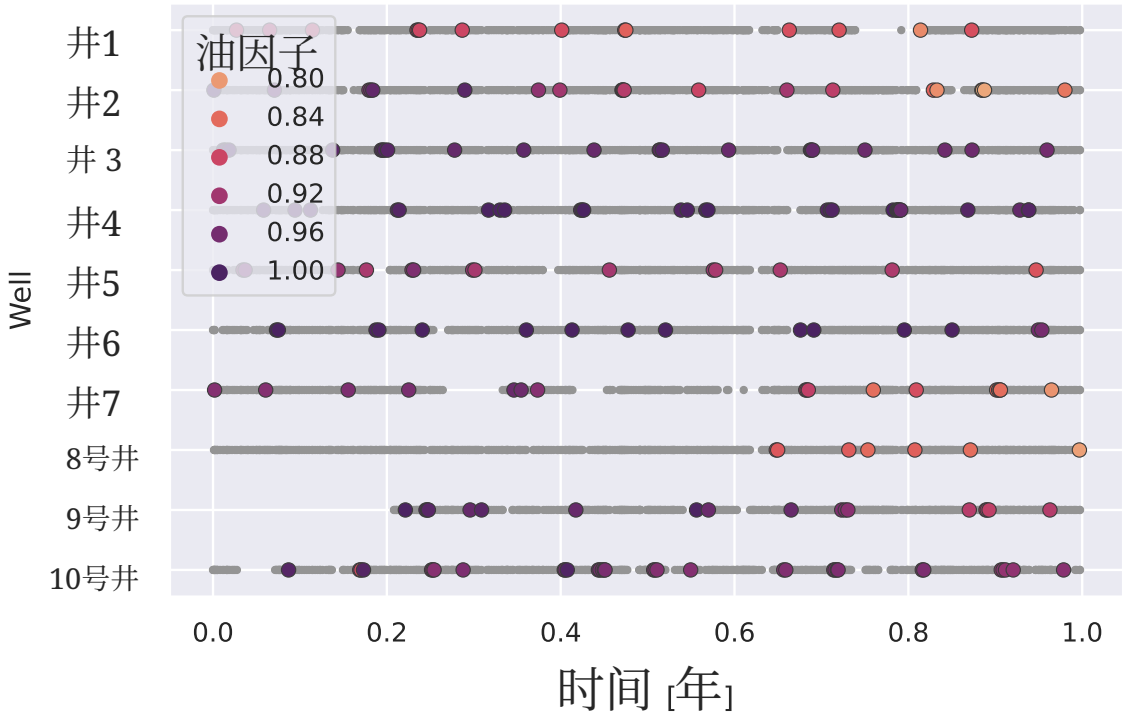


图3：数据概览。每条线对应一口油井。油井处于生产状态时，每个生产数据点均以灰色点标出。每个井测试均以彩色点标出，颜色表示该次测试计算得出的油因子。

表1：数据点汇总。列出了各油井所贡献的生产数据点数量、井测试数量，以及流体组成值的最小值和最大值。

	生产	井测试	最小值 γ_j	最大值 γ_j	min λ_j	最大 λ_j
井1	622	14	0.13	0.14	0.81	0.90
井2	664	21	0.14	0.16	0.78	1.00
井 3	643	22	0.14	0.17	0.96	1.00
井4	645	29	0.15	0.16	0.97	1.00
井5	635	14	0.08	0.15	0.86	0.95
井6	641	15	0.15	0.17	0.96	1.00
井7	565	17	0.13	0.29	0.81	0.97
8号井	666	7	0.05	0.14	0.79	0.87
9号井	492	18	0.14	0.15	0.89	1.00
10号井	536	33	0.14	0.16	0.87	0.99

4 结果与讨论

我们通过探索数据集的三种变体，研究使用序贯蒙特卡洛推断潜在状态空间变量的可行性。除了真实数据外，我们还探索了两个半合成 案例，以展示该方法在受控条件下的特性。这些案例以真实数据为基础构建。我们考虑了真实数据集的两种变体：一种是仅 使用生产数据进行推断，而井测试仅用于验证性能；另一种是同时 使用生产数据和油井测试数据进行推断。

验证采用根据井测试得出的数值作为目标值，并将前一个 时间步的样本均值作为估计值。气体分数和油因子的目标值根据方程8 和方程9 由井测试计算得出。调节因子的目标值通过对方程 5求解 $\beta_{j,t}$ 得到。对于真实数据，目标值取井测试。对于合成数据，目标值取底层真实值。所有结果可视化均使用样本均值，并以两种不同色调的彩色带分别覆盖第 25 至第 75 百分位数和第 5 至第 95 百分位数范围。

在所有案例中，我们都使用自助滤波器生成提议样本。根据第3.1节中的描述，这一过程很简单。在所有案例中，我们都使用 100 000 个样本运行算法。模型所用的具体参数见附录 B。

4.1 三口油井的构造案例

我们首先研究一个包含三口油井的构造案例。仅考虑前 50 个数据点，包括井测试。除时间步 $t = 10$ 时井2的油因子发生跳变外，所有状态参数均设为常数。特征 $x_{j,\ t}$ 取自真实数据，但流量是根据构造的状态值计算虚拟流量计的流量之和得到的。按照模型为分离器测量值添加噪声。这意味着该案例不存在建模误差，因为虚拟流量计和状态空间模型与系统完全匹配。所得总流量如图 4 所示。

对于这个构造案例，我们只对结果进行直观考察。由此得到的油因子估计值如图 5 所示。图 6 将图 4 和图 5 中突出显示区域的样本以散点图的形式呈现。

如图 4 所示，有效样本量整体上较为理想，平均值略低于 50%。不过，有些事件会导致有效样本量大幅下降。这种情况出现在$t = 10$ ，此时井2的油因子发生了显著变化。其他具有信息量的事件（例如井测试）也会导致有效样本量下降。此外，在最初的几个时间步中，样本量需要一些时间才能趋于稳定。我们可以在 $t = 25$ 的图 6 中清楚地看到样本多样性不足。不过，样本量随后会恢复为较为理想的分布，如 $t = 40$ 处所示。

油因子仅发生一次跳变的简单案例揭示了我们推断问题的一个重要方面。由于我们只能依据流量总和来推断隐变量，因此很难判断变化发生在哪口油井上。在本案例中，变化被错误地归因于井1。这种情况一直持续到 $t = 24$ 和 $t = 25$ 处发生了具有信息量的事件：井1停止生产并接入测试分离器。这为井1的真实状态提供了信息，进而也使其他油井的估计得以修正。值得注意的是，对于井2而言，井1在 $t = 24$ 处被移除 时提供的信息同样多；而对于井1，主要修正则发生在 $t = 25$ 处。序列剩余部分缺少进一步改进估计值所需的激励。这反映出这类数据中存在变化是多么重要，也表明油井关闭或启动时的时间步对于推断最有价值。

4.2 完整合成案例

此案例是一个合成案例，采用与第4.1节所述相同的策略，但纳入了全部十口油井，且参数的生成方式有所不同。我们介绍此案例的两种变体。它们的区别仅在于组成参数的生成方式。

在第一种案例中，组分取为真实井测试值的线性插值 ，该案例称为合成副本。所有油井的调节因子均为常数，且都等于一。结合真实测量值 $x_{j,\ t}$，以此评估虚拟流量计，从而生成响应变量

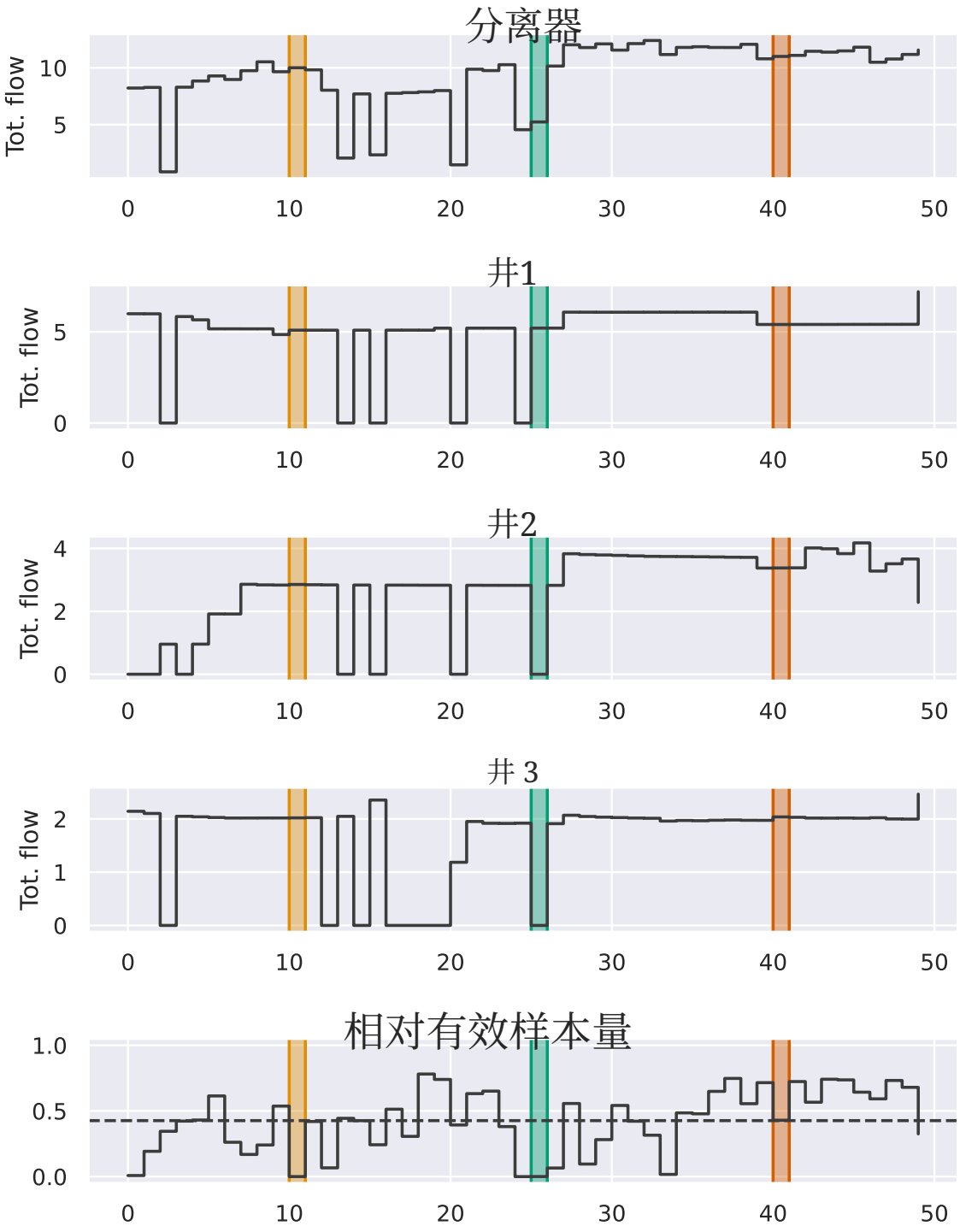


图 4：构造案例中包含 $m = 3$ 的总流量。顶部图显示分离器流量之和，该值为测量值。中间的三幅图显示三口油井的总流量。底部图显示序贯蒙特卡洛估计值的相对有效样本量。三个高亮时间步对应的油因子样本见图 6。

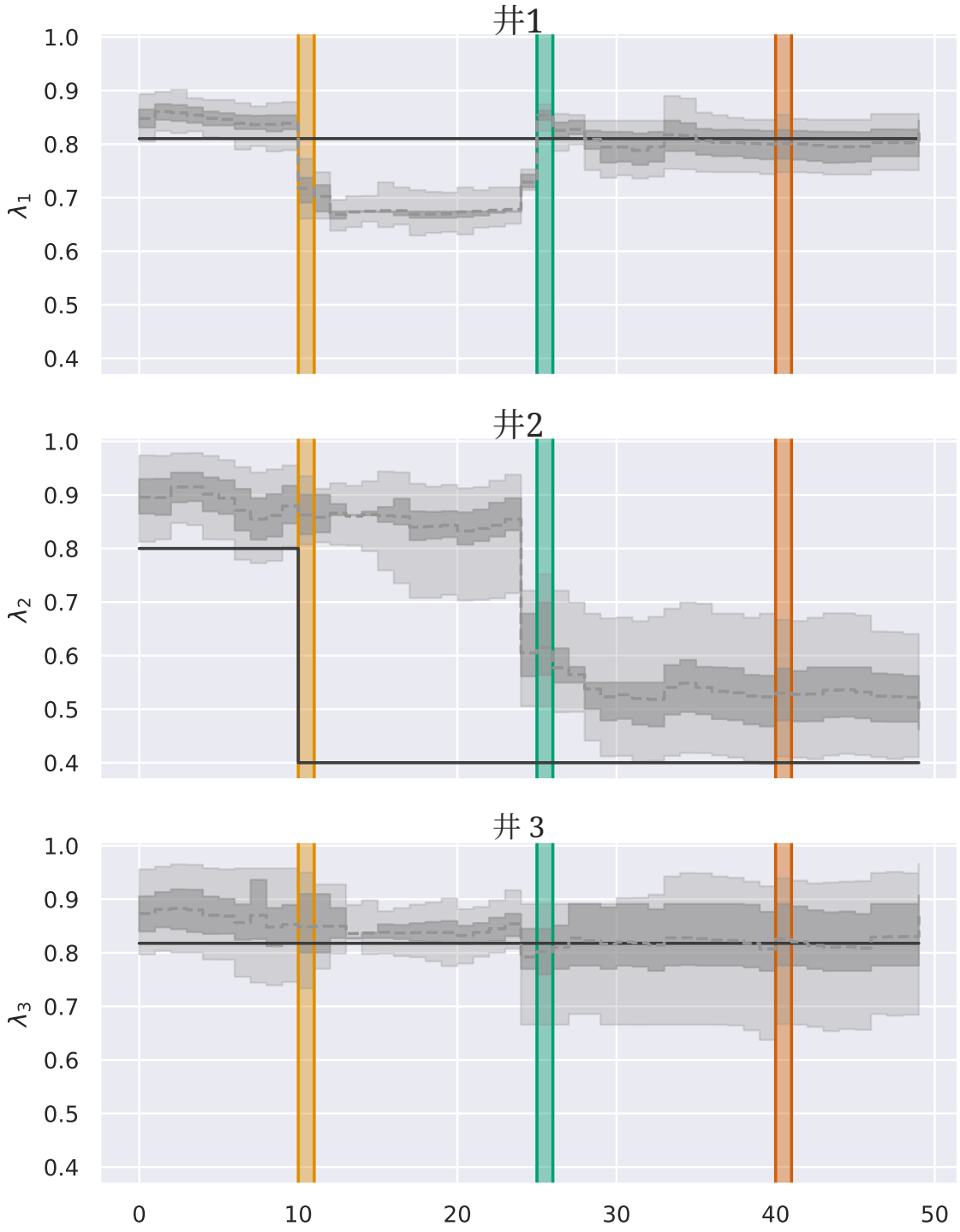


图 5：根据图 4 所示数据估算的油因子。图中以灰色显示各时间步的边缘样本分布，以黑色显示真实值。彩色区间将在图 5 中进行考察。

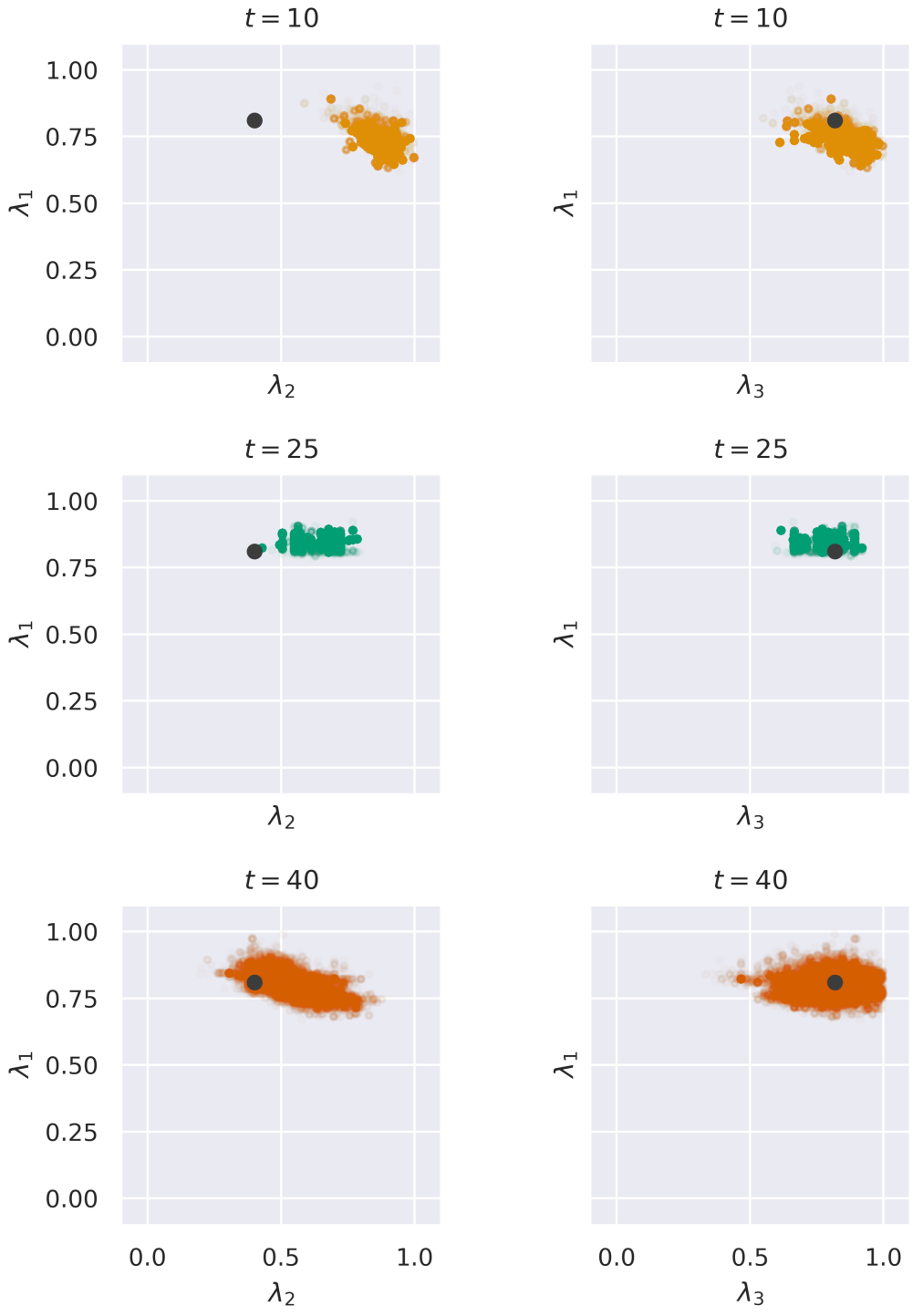


图 6：在图 5 中突出显示的时间步处获取的油因子样本散点图。彩色点表示样本，黑点表示真实数值。

y_t。如果虚拟流量计完美无误，那么这个案例便能反映真实案例。请注意，我们的状态空间模型假设状态按分段常数序列演变，而此处使用的生成组分则在井测试之间按分段线性趋势演变。参数的先验值设定为接近数据集开始时的分离器平均值。

第二个案例称为合成随机，与合成复制案例相同，但组分并非完全遵循井测试。每口油井的组分都被分配了一个位于单位区间内的随机起点。随后，组分随时间的变化被设定为对真实油井测试中观测到的增量进行随机打乱后的结果。提出第二个案例的原因是，真实数据中的组分都非常相似，因此资产平均值能够很好地估计各口油井的组分。

我们详细研究合成随机案例。井2和井7的边际估计分别见图7和图8。对于这两口油井，我们发现最初几天的估计值受到先验的显著影响；在此案例中，气体分数和油因子的先验值都接近0.4。最终，估计值会逐渐收敛到正确数值；能够获取井测试数据的推断自然会略快一些。将所有油井的估计误差汇总后，在图9中也能看到同样的效果。即使没有井测试数据，估计值也能够很好地追踪组分的变化。对于这两口油井，调节因子的估计值相当稳定，但两种情况下偶尔都会出现轻微漂移。

两个合成案例的平均性能汇总于表2。我们看到，与复制案例相比，井测试的纳入对随机案例的影响更大。在复制案例中，我们推测这是因为各组分都相当相似，因此先验信息的作用相当强。此外，仅将组分设为油气资产平均值，本身就能得到不错的估计。两个案例的性能相近，另一个原因是，如第4.1节所述，将一口油井从生产中移除所提供的信息可能相当重要。最后，我们注意到，两个数据集的平均样本量都可以接受。

4.3 真实数据案例

最后，我们在完整的真实数据集上进行推断。在这种情况下，虚拟流量计并非完全准确，且无法保证在测试分离器上观测到的数值能够直接对应于在生产分离器上看到的数值。这也意味着，用于噪声项的数值并不能描述真值。

表2汇总了平均性能及合成数据的结果，图 10则展示了误差随时间的变化。在这两种情况下，估计值都需要大约100天的数据，才能使样本趋于稳定并使误差稳定下来。与合成复制案例相比，真实数据下的误差显著更高。这是正常的，因为虚拟流量计是相当简单的模型，不太可能涵盖一年内出现的全部运行范围。这意味着调节因子会发挥重要作用，而且从一次测试到下一次测试之间可能发生较大变化。这也会导致相当高的误差。例如，调节因子的误差为0.1，意味着乘法校正因子存在10%的误差。

将井测试纳入数据集与否，观测到的误差几乎没有差异。合成复制案例中也观察到了这一点，可能的原因已在第4节4.2中讨论。

图11展示了8号井的估计值，该井首次井测试发生在约220天后。推断仅依据生产数据中的联合观测值，便得出了合理的参数值。可以注意到，根据井测试计算出的参数值与先验假设相当接近。然而，在最初的200天里，气体分数和油因子的估计值都取了不同的数值，之后才回归到时间序列后半段观测到的数值，这表明估计的准确性并非仅由先验决定。

真实数据案例中的噪声水平不像合成案例那样经过精确校准。这一影响见图12。井6的油因子接近1，但由于存在不确定性，样本均值偏离了边界。对于井6，这意味着无论收集多少数据点，油因子的样本均值都保持在0.90左右。我们可以调低噪声水平，但这也可能带来问题。例如，看看图11中的气体分数。在七次井测试中，有一次的气体分数明显低于其他测试。这可能是由于某种额外的过程扰动，而静态正态噪声模型难以捕捉这种扰动。结果便是出现了一种权衡：大多数井测试所获得的噪声水平都过低

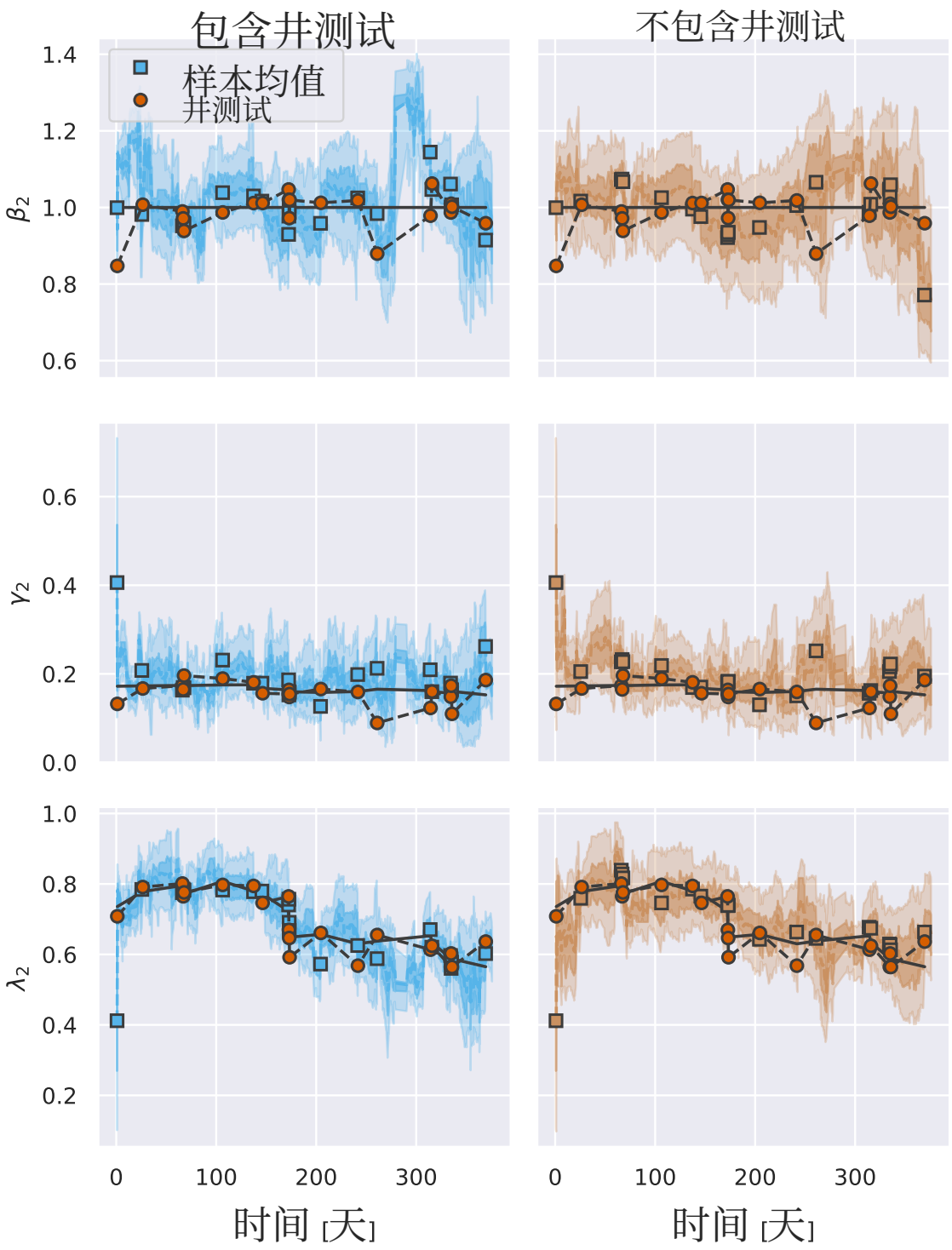


图7：合成随机数据中井2的边际估计。左列为数据集中包含井测试时的估计值，右列为不包含井测试时的估计值。各行依次表示调节因子、气体分数和油因子。橙色圆点表示根据井测试值计算出的参数。蓝色和棕色方块表示井测试之前 该时间步的样本均值。

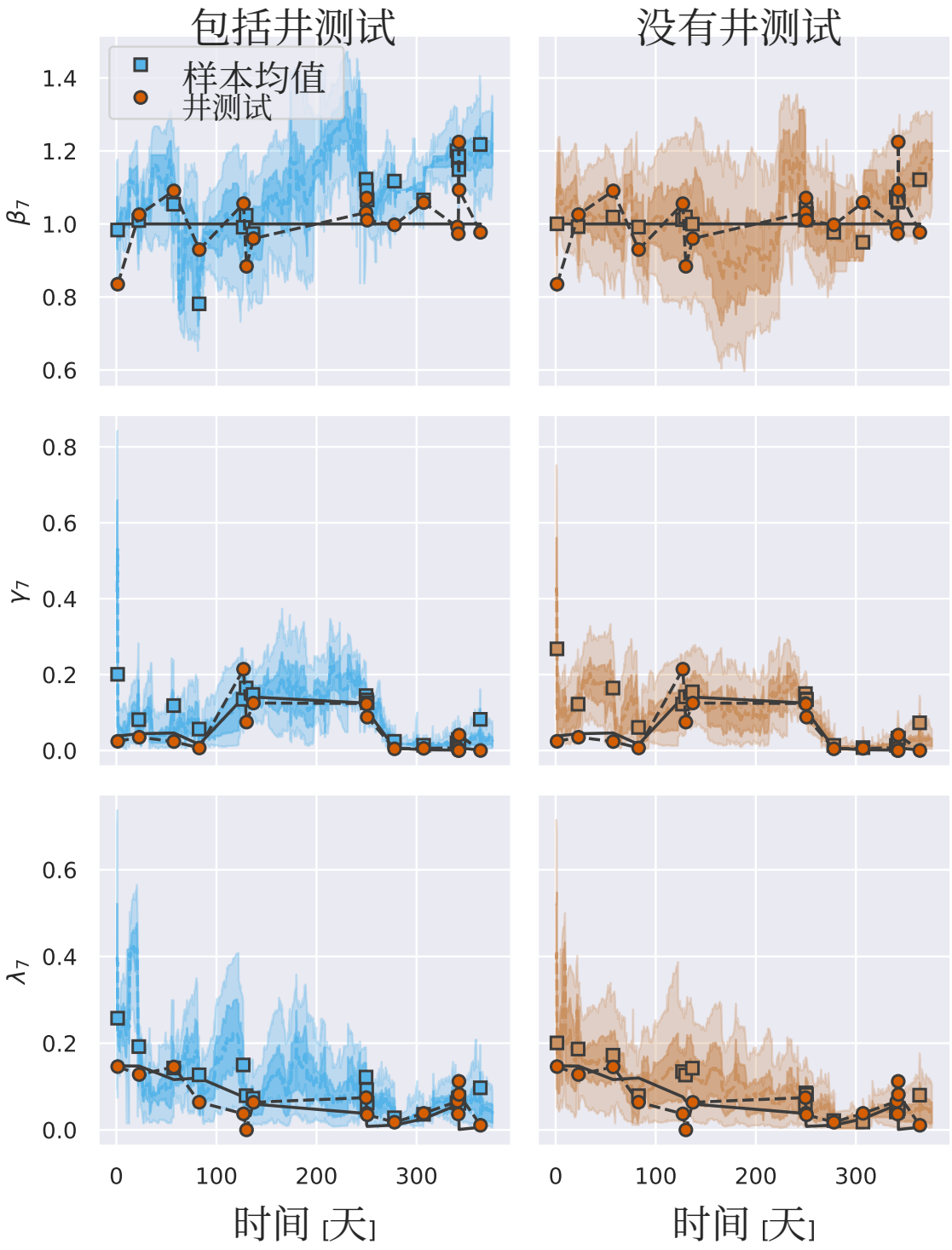


图8：合成随机案例中井7的边际估计。内容与图7相同。

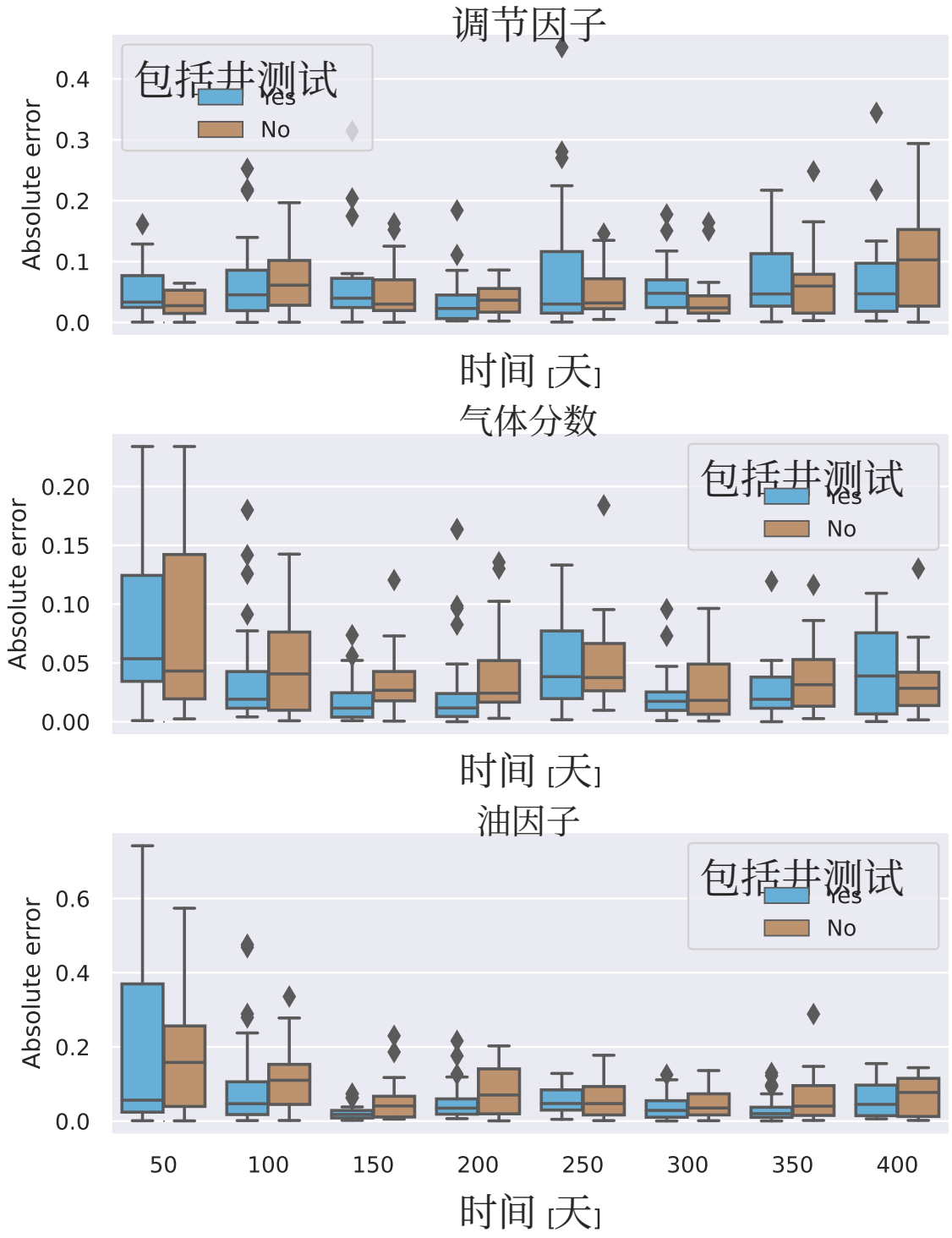


图9：合成随机案例中绝对误差的箱线图，按每50天划分为一个区间。

表2：真实数据、合成副本和合成随机三种情形下的性能汇总。所有情形均列出了纳入和未纳入井测试的数据集结果。所有情形下的绝对误差均为十口油井的平均值，这些油井共进行了190次井测试。

数据类型 井测试	Real Yes	Real No	合成副本 Yes	合成副本 No	合成随机 Yes	合成随机 No
调节因子	0.104	0.100	0.055	0.056	0.063	0.053
气体分数	0.031	0.031	0.023	0.025	0.036	0.043
油因子	0.055	0.052	0.034	0.038	0.066	0.081
相对有效样本量	0.598	0.531	0.358	0.385	0.396	0.417

影响，因为否则异常值会对重要性权重产生过大的影响。在实践中，可以评估每项测试并为其指定单独的噪声水平。这会导致先验具有很强的主观性，因此本文不采用这种做法。

5 结论

本文提出了一种基于序贯蒙特卡洛的虚拟流量计校准方法。该方法利用生产分离器的稳态数据，并结合质量平衡，同时更新十台虚拟流量计的调节因子和流体组成参数。通过真实数据和合成数据对该方法进行了研究，结果令人鼓舞。结果表明，纳入油井测试数据有助于提升效果；但即使缺少井测试数据，该方法也能取得不错的性能。不过，方法性能高度依赖生产数据中的激励程度，而这一程度因具体情况而异。由于可能存在异常值，很难提供一组固定参数来描述过程噪声和测量噪声。因此，井测试等信息量丰富的事件所发挥的作用会受到限制。实际应用中，建议采用更复杂的方法评估各个数据点的准确性。

虚拟流量计规格

本节详细说明第3.1节中介绍的模型和虚拟流量计函数。

我们的虚拟流量计采用 Sachdeva 等人（1986）的机理流动模型。所用的具体公式见 Hotvedt 等人（2022b）。$x_{j,\ t}$的测量分量分别记为： u 表示节流阀开度， p_1 表示节流阀上游压力， p_2 表示节流阀下游压力， T 表示温度。为方便起见，省略下标。该模型求得总质量流量的公式为

$$y_T = C_D A(u) \sqrt{2\rho_2^2 p_1 \left[\frac{\kappa}{\kappa - 1} \phi_G \left(\frac{1}{\rho_{G,1}} - \frac{p_r}{\rho_{G,2}} \right) + \left(\frac{\phi_O}{\rho_O} + \frac{\phi_W}{\rho_W} \right) (1 - p_r) \right]}. \tag{21}$$

该模型通过压力比项$p_r = \max(p_2/p_1,\ r_c)$区分临界流动和亚临界流动。由于 $p_2 \leq p_1$，该比值介于零和一之间，临界压力比 r_c通常接近 0.6。节流阀两侧的气体密度由下式给出

$$\rho_{G,1} = \frac{p_1 M}{ZRT_1},$$
$$\rho_{G,2} = \rho_{G,1} p_r^{1/\kappa},$$

其中， κ 为气体膨胀系数， Z 为压缩因子， M_G 为摩尔质量， R为理想气体常数。该公式旨在捕捉与气体可压缩性相关的影响。下游混合物密度采用加权调和平均值计算，

$$\frac{1}{\rho_2} = \frac{\phi_G}{\rho_G} + \frac{\phi_O}{\rho_O} + \frac{\phi_W}{\rho_W}.$$

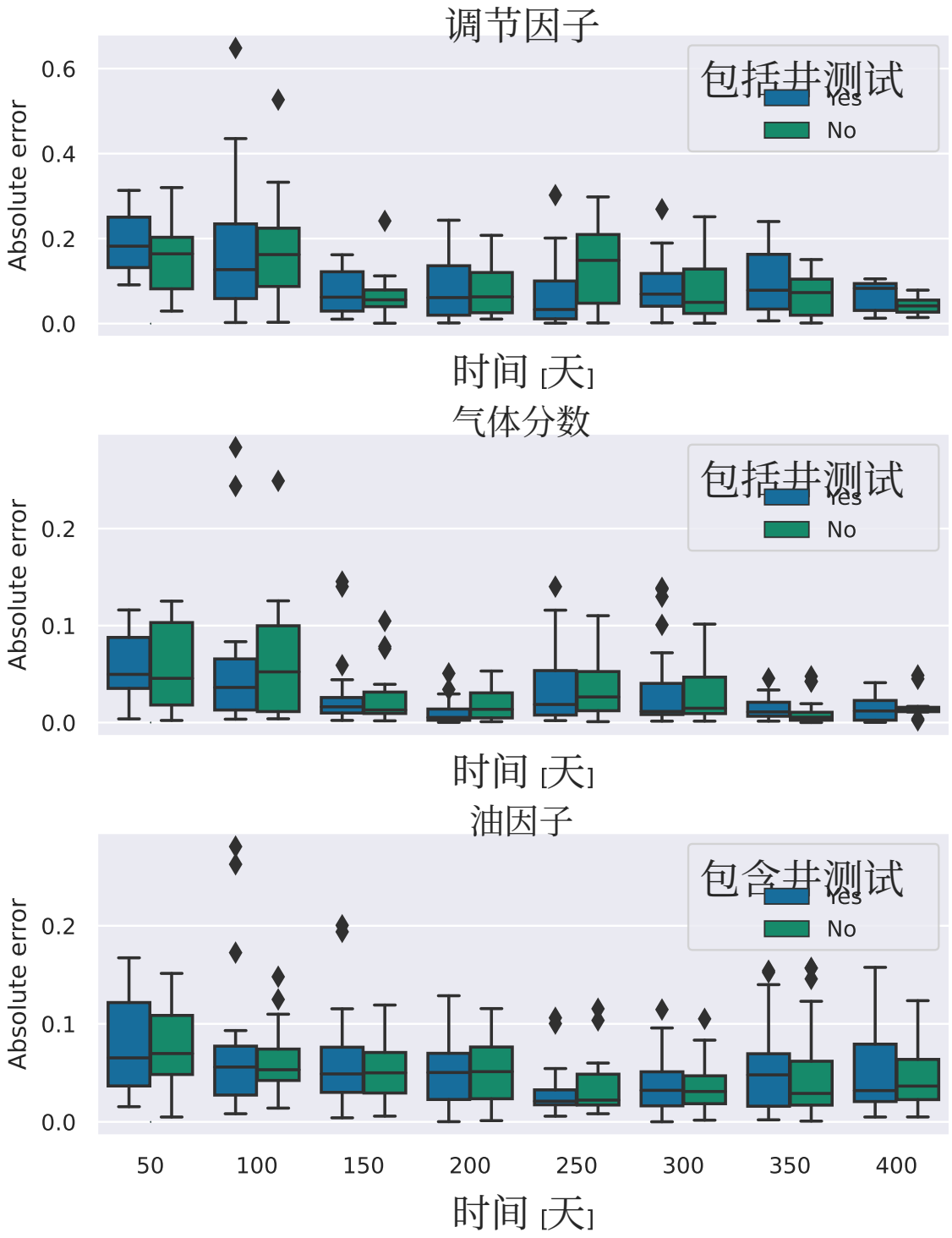


图 10：真实数据绝对误差的箱线图，按 50 天的区间分组。

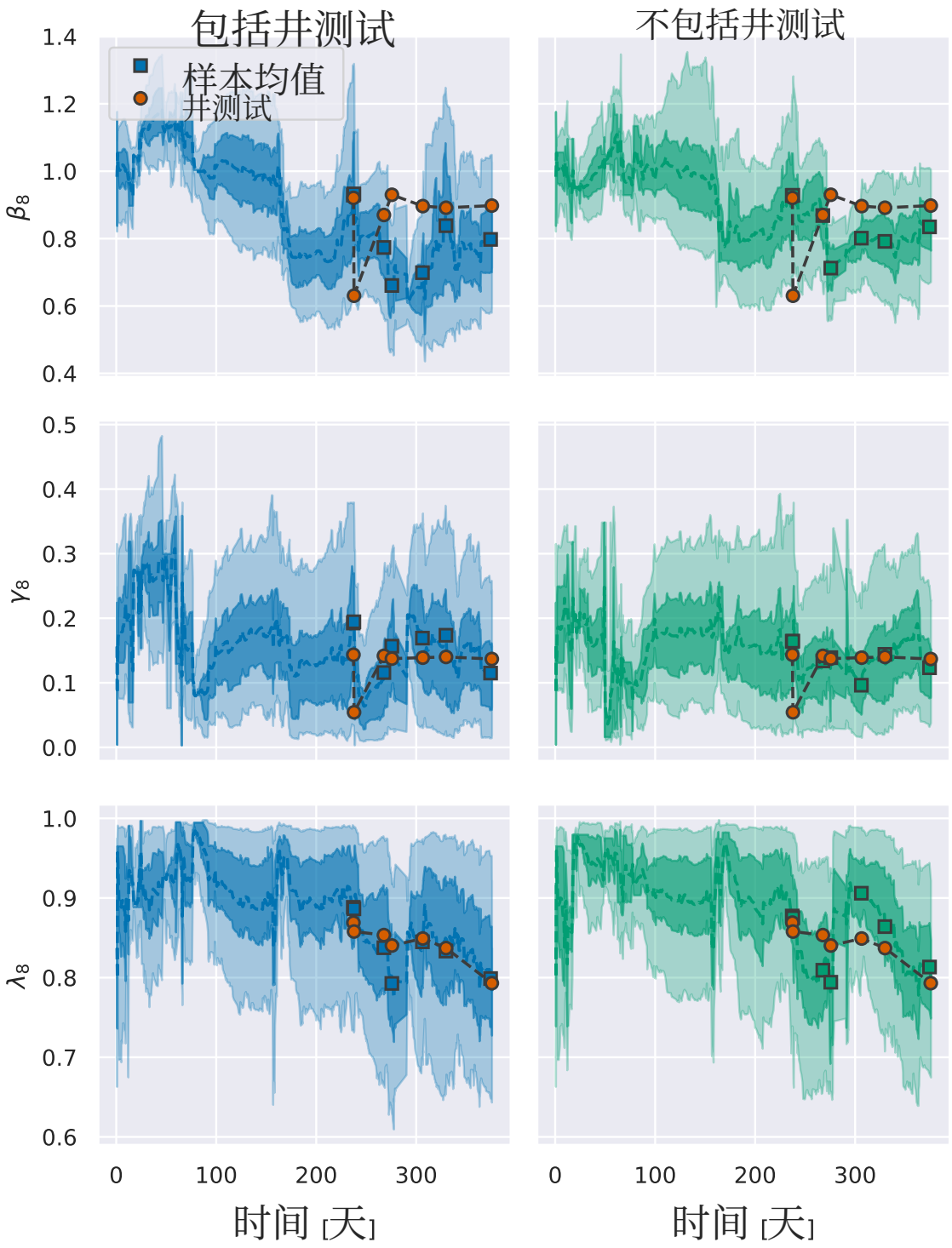


图11：真实数据集中的8号井边际估计。左列显示数据集中包含井测试时的估计值，右列显示不包含井测试时的估计值。各行分别为调节因子、气体分数和油因子。橙色点表示根据井测试值计算出的参数。蓝色和绿色方块表示井测试之前 时间步的样本均值。

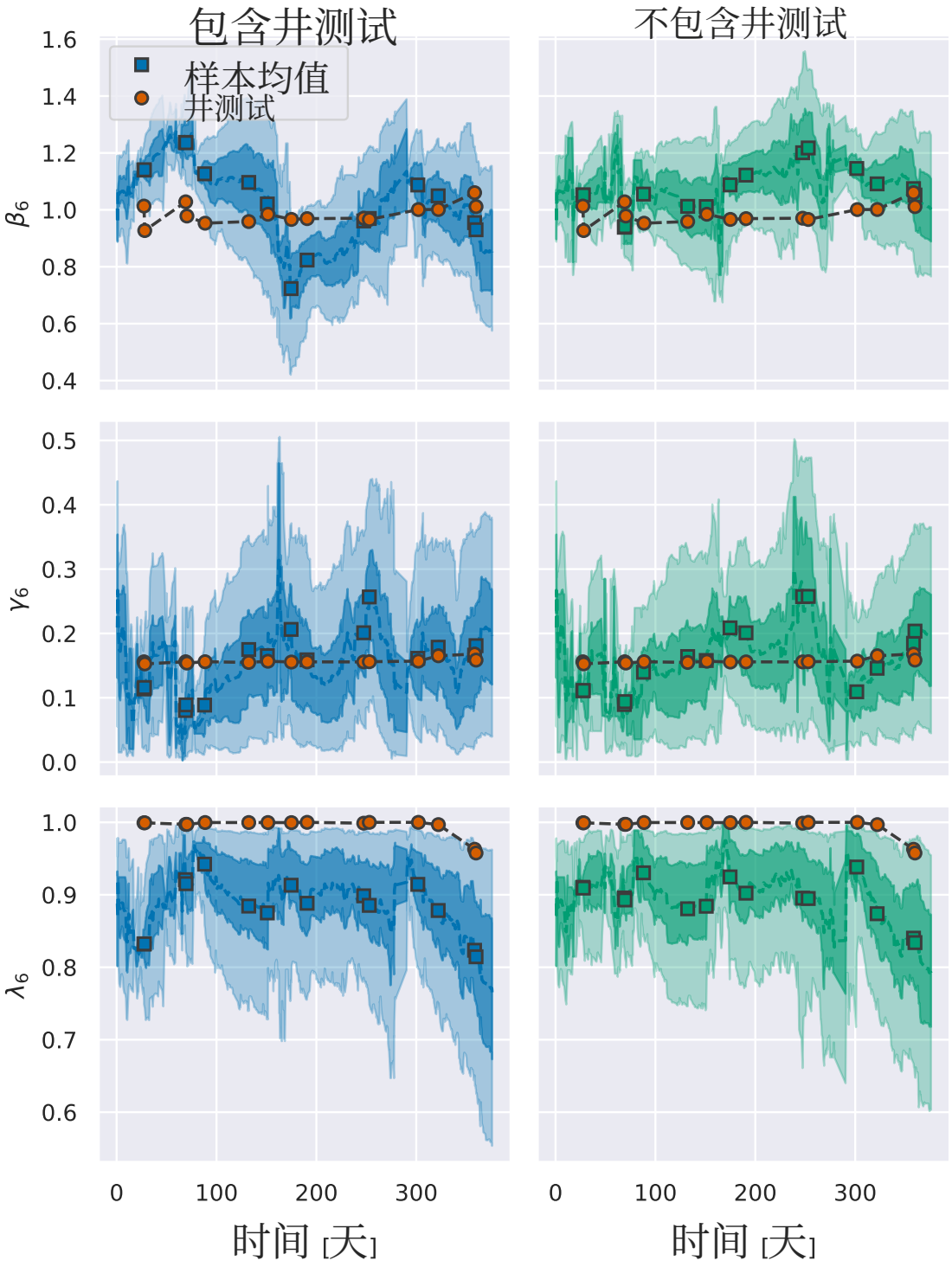


图12：真实数据集的井6边际估计。图中内容与图11相似。

表3：设定　状态空间模型所需的先验密度和转移密度的设置，详见第3.1节。参

数	Value
p_z	0.1
$\mu_{\beta,0}$	1.0
$\sigma_{\beta,0}$	0.1
$\sigma_{\beta,0}$	0.1
σ_{β}	0.05
$\sigma_{\gamma,0}$	0.1
σ_{γ}	0.05
$\sigma_{\lambda,0}$	0.1
σ_{λ}	0.05
σ_{λ}	0.05

节流阀的几何结构由流出系数 C_D 和面积函数 $A(u)$描述。有关此特定模型的更多详情，请参见 Hotvedt 等人 (2022b)。

B 参数设置

本节给出了第4节所有实验中状态空间模型所使用的参数。先验参数和转移参数列于表3中。对于真实数据案例，流量经过缩放，使得 $E[\beta_{j,t}] \approx 1$。对于组分，预期的先验数值大致以数据集开始时的生产分离器平均值为中心。

噪声模型包含三个组成部分：一个用于建模分离器测量值的共享元素，以及来自每口油井的两个组成部分。分离器测量噪声 $\epsilon_t \sim \mathcal{N}(0, \Sigma_{\epsilon,t})$ 被设定为与测量值成比例，

$$\Sigma_{\epsilon,t} = \sigma_\epsilon^2 \begin{bmatrix} y_{t,G} & 0 & 0 \\ 0 & y_{t,O} & 0 \\ 0 & 0 & y_{t,W} \end{bmatrix},$$

其中 $\sigma_\epsilon = 0.05$。油井组成部分 $e_{j,t} \sim \mathcal{N}(0, \Sigma_{j,t})$ 被分解为两个项

$$\Sigma_{j,t} = \sigma_e^2 I_3 + \sigma_f^2 (\beta_{j,t}\phi_{j,t})(\beta_{j,t}\phi_{j,t})^\top,$$

其中 $\sigma_e = 0.05$ 和 $\sigma_f = 0.05$。第一项是用于捕捉来自油井的过程扰动的静态元素，第二项是虚拟流量计总流量的预测误差。根据公式（5），该误差乘以调节因子和组分，从而得到由这些项的外积构成的协方差矩阵。

参考文献

Adukwu, O., Odloak, D., Saad, A. M., 和 Junior, F. K.（2022）。使用非线性滤波器估计气举油井的状态。传感器，22(13):1–16。

Badings, T. S. 和 van Putten, D. S.（2020）。多相分配系统中的数据验证与协调，用于纠正错误和检测重大错误。石油科学与工程杂志，195(6月):107567。

Bikmukhametov, T. 和 J¨aschke, J.（2020）。虚拟流量计量的第一性原理与机器学习：文献综述。石油科学与工程杂志，184(2019年9月):106487。

Chopin,N.,Papaspiliopoulos,O.，以及其他作者（2020）。序贯蒙特卡洛简介，第4卷。Springer。

Doucet, A., De Freitas, N., and Gordon, N. J.（2001）。实践中的序贯蒙特卡洛方法。Springer。He, L., Wen, K., Wu, C., Gong, J., and Ping, X.（2020）。基于粒子滤波器和 NARX 的混合方法，用于多产品管道中的实时流量估算。过程控制期刊，88:19–31。Hotvedt, M., Grimstad, B., Ljungquist, D., and Imsland, L.（2022a）。关于虚拟流量计量的灰箱建模。控制工程实践，118:104974。Hotvedt, M., Grimstad, B. A., and Imsland, L. S.（2022b）。通过被动学习应对虚拟流量计量应用中的非平稳性。专家系统与应用，210:118382。Kadlec, P., Gabrys, B., and Strandt, S.（2009）。流程工业中的数据驱动软测量。计算机与化学工程，33(4):795–814。Kagermann, H. and Wahlster, W.（2022）。工业 4.0 十年。科学，4(3)。Kanshio, S.（2020）。上游石油和天然气行业烃类分配方法综述。石油科学与工程杂志，184。Lorentzen, R. J., Nævdal, G., and Lage, A.（2003）。使用集合卡尔曼滤波器调整两相流模型中的参数。国际多相流期刊，29(8):1283–1309。Lu, H., Guo, L., Azimi, M., and Huang, K.（2019）。石油和天然气 4.0 时代：系统性综述与展望。工业计算机，111:68–90。Monteiro, D. D., Duque, M. M., Chaves, G. S., Filho, V. M. F., and Baioco, J. S.（2020）。利用数据分析量化生产测试不确定性对石油流量预测的影响。*IFPEnergies Nouvelles*，75:1–15。Sachdeva, R., Schmidt, Z., Brill, J. P., and Blais, R. M.（1986）。节流阀中的两相流。SPE 年度技术会议暨展览会。Sandnes, A. T., Grimstad, B., and Kolbjørnsen, O.（2021）。用于虚拟流量计量的多任务学习。知识库系统，232:107458。Sch¨uller, R. B., Solbakken, T., and Selmer-Olsen, S.（2003）。亚临界石油/天然气/水流动条件下节流阀多相流量模型的评估。*SPE* 生产与设施，18(3):170– 181。