根据光学标测记录重建心脏电激动

C. D. Marcotte,[1 a)] [2] [3] [4], M. J. Hoffman、F. H. Fenton 和 E. M. Cherry[1] *Department of Computer Science Durham University Durham UK DH1 3LE*,,,,[2] *School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY, USA, 14623*[3] *School of Physics Georgia Institute of Technology Atlanta GA USA 30332*,,,,,[4] *School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA, 30332*

（日期：2023年9月6日）

重建组织未观测深度范围内的电激动模式，对于充分发挥计算模型在心脏医学中的潜力至关重要。我们利用犬心室心外膜和心内膜表面的心脏电激动实验性光学标测记录作为观测，指导局地集合变换卡尔曼滤波（LETKF）数据同化方案。我们展示了，纳入有关刺激方案的明确信息，能够小幅提升集合重建的可信度以及同化随时间推移的可靠性。同样，我们考察了在实验所得观测集背景下，向同化方案中加入随机建模的有效性。在观测和建模两个阶段，均通过观测的不确定性以及对同化集合中所用模型的设定来处理近似误差。我们发现，对观测进行扰动性修改，对状态重建的精度和稳健性影响不大，甚至可能造成不利影响。此外，我们发现，将观测中的额外信息纳入模型本身（在刺激和随机电流的情形下），相较于完全自治模型，虽然能略微提高重建精度，但也会使模型本身变得复杂，因而可能引入新类型的模型误差。明确建模信息的纳入对重建的影响微乎其微，甚至有负面作用，这表明有必要探索新的途径，优化应用于心脏电激动的数据同化方案。

关键词：数据同化、心脏动力学、实验数据

arXiv:2305.00009v3 [q-bio.QM] 5 Sep 2023

典型的心脏电生理实验使用光学相机记录兴奋数据，相机经过调校，可测量电压敏感染料的荧光，但只能观测组织表面的兴奋模式。理论和计算方面的证据表明，组织深处的兴奋模式细节对预测和控制十分重要，而表面可观测到的内容只是整体兴奋结构的一小部分。组织表面附近不可观测的结构可能使这些模式具有促心律失常特性；这些结构能够锚定动力学的组织特征，从而简化兴奋模式。由于传统实验手段无法探查组织内部，我们转而使用计算方法重建其内部，具体而言，就是采用数据同化。

I. 引言

心脏兴奋动力学实验只能让我们有限地窥见一个极其复杂的系统。微电极记录以牺牲空间分辨率为代价换取时间分辨率；心电图舍弃细胞细节，转而呈现器官层面的整体性；光学标测实验则聚焦于组织暴露表面附近的电兴奋，而组织的内部动力学可能复杂得多。数据同化（DA）有望解决后一类实验面临的难题：借助可靠的动力学模型和对实验状态足够精确的观测，数据同化技术旨在依据组织表面上的观测，重建不可观测的内部动力学。重要的是，数据同化技术能够稳健应对不确定性，因为在真实实验数据的处理中，真实状态和观测值的不确定性，以及模型预测的具体程度，始终都需要考虑。

数据同化与天气和气候预测领域共同发展，这些领域所用的工具通常经过专门设计并经过精细调校。集合方法尤其是应用领域中的新兴方法[1],，但由于它们与非线性动力学模型之间的关系较为简单，因此备受青睐。

a)通讯作 者：christopher.marcotte@durham.ac .uk

许多研究开发了新的数据同化方法，并将其与现有方法进行比较[2–7]；提出了提升数据同化结果性能[8,9] 和稳健性的新方法[10], ；制定了公平的评分规则，以严格界定何谓“改进”数据同化结果[11–14]；甚至还探讨了观测、同化滤波器与模型之间联系等细微问题[15–18]。

在三维组织中同化心脏激励模式方面，大量研究聚焦于这一问题的基础层面——数据分布、不确定性量化、模型误差以及对噪声的稳健性[19–21]——这些研究是在模型生成的合成观测背景下开展的，而非基于实验。此前的研究[22] 发现，局部集合变换卡尔曼滤波器（LETKF）重建结果的精度对某些参数不确定性较为敏感；此外，采用底层模型的随机形式可提高稳健性和精度。本研究考察重建方案在该领域典型实验数据上的效能：对组织进行周期性起搏，并通过光学标测捕获组织心外膜和心内膜表面的激励。

本文将使用一项此前在实验研究中发表的数据集[23–26]。该数据集对应于犬心室组织中受局部周期性电流刺激产生的激活模式。我们将探究使用自治模型和非自治模型重建这些实验观测的效果，并研究随机模型干预对重建精度的影响。最后，我们将考察在组织内部加入额外的合成观测，对同化稳健性以及实验动力学重建精度的影响。

II. 方法

本节概述动力学模型的数学细节、合理模型参数的确定、如何根据实验数据生成观测，以及最后用于同化观测、以重建实验动力学的局部集合变换卡尔曼滤波器（LETKF）。

A. 模型

我们使用 Fenton-Karma 三变量模型[27] 描述电生理学，其中各向异性扩散由纤维取向控制；该取向在x, y-平面内随深度旋转，其相对于 x轴的夹角记为 $\theta(z)$。演化方程为

$$\partial_t u - \nabla \cdot (\mathbf{D}(\mathbf{x})\nabla u) = I_{\text{stim}} - I_{fi} - I_{so} - I_{si}, \quad (1)$$
$$\partial_t v = \Theta(u_c - u)(1 - v)/\tau_v^-(u) - \Theta(u - u_c)v/\tau_v^+ \quad (2)$$
$$\partial_t w = \Theta(u_c - u)(1 - w)/\tau_w^- - \Theta(u - u_c)w/\tau_w^+, \quad (3)$$

结合无通量边界条件，$\hat{\mathbf{n}} \cdot \mathbf{D}(\mathbf{x})\nabla u = 0$。电流以及 v 的显式弛豫尺度定义如下：

$$I_{fi}(u, v) = -v\Theta(u - u_c)(1 - u)(u - u_c)/\tau_d, \quad (4)$$
$$I_{so}(u) = u\Theta(u_c - u)/\tau_o + \Theta(u - u_c)/\tau_r, \quad (5)$$
$$I_{si}(u, w) = -w(1 + \tanh(k(u - u_c^{si})))/(2\tau_{si}), \quad (6)$$
$$\tau_v^-(u) = \tau_{v1}^-\Theta(u - u_v) + \tau_{v2}^-\Theta(u_v - u), \quad (7)$$

除显式刺激电流 $I_{\text{stim}}(t,\ \mathbf{x})$外，该电流在自主动力学中恒为零，并在特定同化实验中指定（见第 III A 3 节）。组织中的扩散空间耦合采用有限差分格式实现，遵循原始模型描述[27]。组织的扩散耦合受纤维取向影响，该取向随组织深度旋转 d，$\theta(z) = \theta_0 + \delta\theta(-1/2 + z/d)$，其中 $0 \le z \le d$、θ_0 和 $\delta\theta$ 在本文中均有指定；对于（扩散）各向异性比$D_\parallel/D_\perp = 5$，其值为 $D_\parallel = 0.001$cm^2/ms和$D_\perp = 0.0002$ cm^2/，与心脏组织中测得的一些比值相比相对较小[28,29]。

对于此模型和实验数据集，我们采用参考文献 26 中定义的模型参数。这些参数与该模型用于产生具有物理意义的动力学时所采用的其他参数集截然不同；参见参考文献 27。不过，在本研究中，我们选择考察模型误差最小的情况，以便更好地确定改变模型和观测属性所产生的影响。

在所有实验中，我们对门控变量 v 和 w 采用 Rush-Larson 时间步进方案，并对跨膜电位 u[30]采用前向欧拉法。纳入随机效应时（参见第 III A 4 节），我们将跨膜电位 u 所用的前向欧拉法改为 Euler-Maruyama 法。本研究未在门控变量的动力学中纳入随机项，因为先前研究表明，对于该模型，仅使用随机电流[22], 也能达到类似的集合膨胀效果，同时使实现更加简单、高效。

B. 观测提取

实验数据包括荧光的光学记录（有关原始制备的详细信息，请参见参考文献 23 和 31），这些记录对应于跨膜电位，并经过平滑和去噪处理。观测值（$\mathbf{y}^o$）是由元组（$\mathbf{y}_i^o$）组成的向量，其中包含

对于每个观测值，均包含用于标示观测场变量的标签、区域中的坐标（$[x_i, y_i, z_i]$）、这些坐标处的观测场变量（例如，$u_i = u(x_i, y_i, z_i)$）以及观测值的不确定性（η_i）。观测场值还经过归一化，使得 $0 \leq u_i \leq 1$。由于观测值仅涵盖跨膜电位 u，因此我们可以交替使用 $\mathbf{y}_i^o$ 和 u_i 来表示观测值。记录数据涵盖了犬心室组织在单一位置施加周期性刺激后产生的响应激励，刺激周期或基本周期长度（BCLs）范围很大。

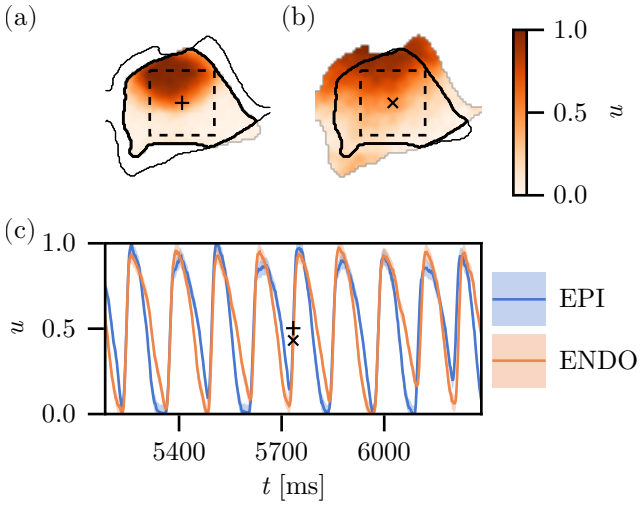

图 1. 用于重建实验的实验数据示例。(a) 心外膜快照和 (b) 心内膜快照，拍摄时间为 $t = 5734$ ms；(c) 标记为 $+$（心外膜）和 $\times$（心内膜）的位置处的记录时间轨迹。(a) 和 (b) 中的黑色细曲线表示对侧表面记录的边界，黑色粗曲线表示两者共同交集的边界。插图中的虚线正方形边长为 3 cm，与水平轴和垂直轴对齐，显示了为同化提取观测的选定区域。

本研究聚焦于一个刺激周期为 118 ± 2 ms 的数据集，因为如此短的刺激周期所产生的模式在空间和时间上最不规则，因此构成了最具挑战性的重建任务。图 1 展示了该数据集的记录示例，具体包括 (a) 心外膜和 (b) 心内膜表面记录，以及 (c) 组织表面中心点处记录的时间轨迹。图中还展示了一个用于生成同化实验观测集的 3 cm 方形插图，以及心外膜和心内膜表面的边界。记录中可以看到一个传播速度极快的波前，从图的左上方传播至右下方，几乎同时出现在两个表面上（参见图 1(c)）；该波前由施加于心内膜基底部的刺激电流引发[23]。

随着时间推移，这些数据呈现出典型的交替现象其中心脏组织兴奋以快于正常节律的速度起搏，即动作电位的持续时间和振幅发生变化。对于心外膜和心内膜数据集，我们通过在空间上构造三次插值函数、在时间上进行线性插值，并在对应于图 1(a、b) 中虚线方框所示嵌入区域的预先指定坐标处对该插值函数进行采样，从数据集中生成观测。对于在域内部生成的合成观测（参见第 III B 1 节），我们根据观测在表面投影位置处的数值，沿深度方向进行线性插值，以确定域内部合成观测的数值。这确保了信息能够从表面传播到中间深度。

每个$\mathbf{y}_i^o$ 同样被赋予一个不确定性 η_i，我们将其设为 $\eta = 0.05$，默认值为动作电位幅度范围的 5%。$\eta = 0.05$ 的值是根据该数据集在持续时间指数 τ 内，$u^o(t) - u^o(t - \tau)$分布的半峰全宽估算得出的。关于此计算的详细依据，请参见图 S2。对于插值到组织中间深度的合成观测（参见第 III B 1 节），我们赋予其较大的不确定性 $\eta_i = \eta = 0.5$，以考虑这些合成观测对同化状态的扰动效应，因为观测协方差的对角元素与其逆值 $R_{ii} = \eta_i^{-1}$ 相关；也就是说，除了一种高度不确定的提示（指出中平面在特定位置和时间处于激发或未激发状态）之外，它们不包含来自实验数据的新信息。对于对波前具体形式的不确定性进行编码的情况（参见第 III B 2 节），我们对观测不确定性进行类似的修正，使用观测场值 $\eta(\mathbf{y}_i^o)$ 的非线性函数；该函数的取值范围由 $[\underline{\eta}, \eta]$ 限定，并在$\mathbf{y}_i^o = 0.5$时达到峰值 $\eta = 0.5$。

C. 局部集合变换卡尔曼滤波

集合变换卡尔曼滤波（ETKF）通过寻找先验状态的最佳拟合线性组合，以二次意义匹配一组观测。局部集合变换卡尔曼滤波[32]根据状态的局部观测，对表示先验协方差和均值的集合进行变换，从而得到表示后验协方差和均值的集合。我们考虑经典卡尔曼滤波优化泛函[33]，

$$J(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}}^b)^\top (P^b)^{-1} (\mathbf{x} - \overline{\mathbf{x}}^b) + (\mathbf{y}^o - H(\mathbf{x}))^\top R^{-1} (\mathbf{y}^o - H(\mathbf{x})), \tag{8}$$

其中，规模为 M 的先验集合由集合偏差矩阵（$\mathbf{X}^b$）和均值（$\mathbf{x}^b$）组成，$\mathbf{x}^b = \mathbf{X}^b + \mathbf{x}^b[1, \ldots]^\top$；后验集合也可类似地分解为 $\mathbf{x}^a$。先验集合的协方差由矩阵 $P^b = (M - 1)^{-1} \mathbf{X}^b (\mathbf{X}^b)^\top$ 描述。物理状态的观测收集于

$\mathbf{y}^o \in \mathbb{R}^O$，而观测算子 $H: \mathbb{R}^N \to \mathbb{R}^O$ 将状态空间转换到观测空间，矩阵 $R \in \mathbb{R}^{O\times O}$ 描述观测的协方差。根据不确定性，观测算子将状态向量映射为观测，并为每个观测添加方差为 η 的加性噪声。式（8）的最小化解记为分析集合 $\mathbf{x}^a$，它根据数据 $\mathbf{y}^o$ 观测到的真实状态作出最佳近似，并由背景集合 $\mathbf{x}^b$ 构成。

ETKF 在背景集合状态张成的空间内执行分析更新。分析集合协方差表示为 $P^a = [(M-1)I/\rho + (\mathbf{Y}^b)^\top R^{-1}(\mathbf{Y}^b)]^{-1}$，其中 $\tilde{P}^a \in \mathbb{R}^{M\times M}$，分析集合均值为 $\mathbf{x}^a = \mathbf{x}^b + \mathbf{X}^b w^a$。标量因子 $\neg p \geq \vdash$ 是（均匀、恒定的）乘性协方差膨胀因子。均值权重由 $\mathbf{x}^a = \tilde{P}^a(\mathbf{Y}^b)^\top R^{-1}(\mathbf{y}^o - \mathbf{y}^b)$ 确定。分析集合均值偏差的权重同样用背景均值偏差的基底表示，即 $W^a = [(M-1)\,\tilde{P}^a]^{1/2}$，且 $\mathbf{x}^a = \overline{\mathbf{x}}^b + \mathbf{X}^b(w^a + W^a)$。这不同于误差子空间变换卡尔曼滤波（ESTKF）[2]，，后者在由 $\mathbf{X}^b$ 张成的 $(\overline{M}-1)$ 维子空间中执行变换。

LETKF 在背景集合空间中局部扩展了这一分析更新——这里的“局部”是指观测影响和协方差方面。引入观测局地化和协方差局地化后，便可对状态向量的每个元素应用前述分析；某个状态向量元素所纳入的观测，则由局地化过程决定。在本研究中，我们采用各向同性局地化权重：将背景协方差矩阵 P^b 乘以一个随距离衰减的五阶函数，该函数在距离达到 $2\sqrt{10/3}\,\sigma_o$ 时降为零。为与我们先前的工作保持一致，我们保留了 $\sigma_o = 6$ (21.9 个空间单位，即 ≈ 0.33 cm)。有关局地化过程实现的更多数学细节，请参阅文献 20、22 和 32。

最后，由于没有采用优化的乘性与加性膨胀方案，我们仅使用了乘性膨胀。LETKF 对分析集合偏差（$\mathbf{X}^a = \mathbf{x}^a - \mathbf{x}^a$）进行乘性膨胀，膨胀系数固定为 $\rho = 1.05$。尽管先前的结果表明加性膨胀有效[22]，，本研究并未采用加性膨胀；我们预期随机方法也能以类似方式影响集合，同时显著减少所需的磁盘空间。

III. 结果

我们报告了若干数据同化实验的结果。在所有实验中，我们使用一个由犬心室周期性刺激的光学标测记录数据推导而来的观测数据集，并采用固定的模型积分时间（亦即同化间隔） $T_a = 2$毫秒该时间根据数据记录频率确定。同化输出包括背景集合（$\mathbf{x}^b$）和分析集合（$\mathbf{x}^a$）、各自的均值 $\mathbf{x}^* = M^{-1}\sum_{m=1}^{M}\mathbf{x}^*_m$，其中$m = 1$、.....、 M表示集合中的索引，以及它们的集合离散度，

$$\mathrm{SPRD}^* = (M-1)^{-1}\left(\sum_{m=1}^{M}|\mathbf{x}^*_m - \overline{\mathbf{x}}^*|^2\right)^{1/2}, \quad (9)$$

即集合状态向量的方差。这里我们对集合采用向量组表示法，$\mathbf{x}^* = [\mathbf{x}^*_1, \ldots, \mathbf{x}^*_M]$；但也可以等价地在五索引数组$\mathbf{x}^* = [x^*_{ijklm}]$ 中明确空间（ijk）、变量（l）和集合（m）索引，其中索引（i, j, k, l）在线性向量形式中被线性化。

我们通过比较集合均值与观测$\mathbf{y}^o(t)$，利用重建状态的逐点误差报告重建的均方根误差（RMSE），

$$\mathrm{RMSE}^* = \left(N_{\mathrm{obs}}^{-1}\sum_{i=1}^{N_{\mathrm{obs}}}|\mathbf{y}^o_i - \overline{\mathbf{x}}^*_{\iota(i)}|^2\right)^{1/2}, \quad (10)$$

其中$\mathbf{y}^o_i(t)$是第 i个观测，索引 $\iota(i)$将相应的观测索引 i 映射到状态向量中对应的元素。因此，在我们的实验中，RMSE*的求和范围仅限于区域表面。SPRD*提供深度信息；通过对状态向量中对应长度和宽度的索引进行归约，我们可以得到每个状态变量在区域深度方向上的离散度度量，

$$\mathrm{SPRD}^*_{kl} = \sqrt{(n_x n_y)^{-1}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}(\mathrm{SPRD}^*_{ijkl})^2}, \quad (11)$$

其中，我们隐式地将 SPRD^* 从向量重塑为一个四维数组，并显式标出空间索引（i、 j、 k）和变量索引（l）。在实验中，我们针对背景集合$\mathbf{x}^b$ 和 u 个场元素，随时间计算该标量场 SPRD^b，并将结果报告为 $\mathrm{SPRD}^b_u(t$、 $z)$。

(10) 的定义将其评估范围限定在区域内存在观测的子集，因此无法说明内部重建的一致性。对于未知状态，通常使用适当的评分规则来报告重建一致性，例如集合的连续排序概率评分（CRPS）。该评分仅根据集合状态即可表述，无须（必然）诉诸观测或“真实”状态。[34] 我们采用调整后的 CRPS 的核表示[35]，

$$\mathrm{CRPS}^*(\mathbf{x}^b, \mathbf{y}) = \sum_{m=1}^{M}\frac{\|\mathbf{x}^b_m - \mathbf{y}\|}{M} - \sum_{m,m'=1}^{M}\frac{\|\mathbf{x}^b_m - \mathbf{x}^b_{m'}\|}{2M(M-1)}, \quad (12)$$

其中 $\mathbf{y} = \mathbf{y}^o$ 或 $\mathbf{y} = \mathbf{x}^a$ 分别表示 CRPS^o 或 CRPS^a。对于成员独立同分布的可交换集合，无论集合规模如何，该评分均具有理想性质 M。上文中的范数（$\|\cdots\|$）是对向量 $\mathbf{y}$ 的索引取 1-范数。当 $\mathbf{y}$ 指代观测 $\mathbf{y}^o$时，求和针对观测进行（因此受限于表面）；当 $\mathbf{y}$ 指代分析集合均值 $\mathbf{x}^a$ 时，求和则衡量整个背景集合相对于 LETKF 根据全部状态对观测所作最佳近似的重建一致性，$\mathbf{x}^a$。建议读者查阅补充材料，以进一步了解重建随时间演变的代表性动力学。

A. 模型变化对重建精度的影响

我们针对心脏激发模式的重建开展了多项实验。在最简单的情形下，我们将观测数据中的刺激模式信息纳入模型动力学，但不采用任何同化过程，以此计算重建过程的基准精度。此外，我们还考察了不包含刺激信息、但采用强同化过程的情形，以评估在重建精度方面，对观测数据中固有的非自治强迫进行建模的重要性。最后，我们将刺激强迫纳入模型动力学，并结合强同化过程；这种方法向同化提供了最多的信息，也清晰地体现了非自治模型信息对重建可靠性的重要性。

1. 自由运行

首先，我们要探究，与集合的平均过程以及集合状态与刺激节律的同步相比，同化过程在多大程度上促成了状态重建。对于重建问题，我们仅评估集合成员在刺激强迫下的非受控动力学（因为在没有刺激强迫时，所有集合成员状态都会回归静息状态）。为此，我们报告去除同化更新后的重建结果——等价地说，不允许观测影响任何状态变量——此时模型会受到周期性刺激，以匹配观测数据中观测到的刺激。

在自由运行设置中，我们不进行任何同化，具体做法是从分析计算中明确滤除所有观测，因此 LETKF 简化为一个缩放—置换算子，因为无法保证集合索引保持不变，并且 $\mathbf{x}^a \equiv \mathbf{x}^b$。自由运行 $\quad -$ 实验是区分长时间同化过程中 LETKF 和刺激模型对重建带来的改进的一种有效方法。

此外，为避免得到平凡解，我们在自由运行实验的动力学模型中加入了刺激电流强迫。因此，在此设置下，我们测试的是仅凭模型和刺激信息重现观测的准确性，而不借助同化方法。我们利用观测数据，为集合模型构建了一个显式的刺激电流。我们确定了该刺激电流的周期 T、时间偏移 t_0 和质心位置 $\mathbf{x}_0$，并假设其模型形式为：

$$I_{\mathrm{stim}}(t,\mathbf{x}) = I_0\,\Pi(t)\,\mathrm{e}^{-\kappa^2\left((x-x_0)^2+(y-y_0)^2\right)}, \qquad (13)$$

它通过进入演化方程，影响 $u(t,\mathbf{x})$ 的动力学；这些方程中还包含动力学模型已有的自主电流I_{so}、 I_{fi}和 I_{si} 。刺激电流由“顶帽”函数 $\Pi(t)$ 在时间上调制，该函数会在短暂的时间间隔 τ_{on} ($\Pi(t) = 1$) 内将刺激电流“开启”，其余时间则保持“关闭”， $I_{\mathrm{stim}} = 0$ ($\Pi(t) = 0$)。

本研究所用数据集对应的实验采用周期性双相刺激，在已同化区域之外的心内膜基底部施加。刺激电流模型，即方程 (13)，为单相且与深度无关，这与实验中的刺激电流并不相同。实际上，我们的刺激电流模型近似表示实验刺激对已同化区域内观测的影响；换言之，这是一个数据驱动模型，只能近似重现观测到的效应。由于我们只能在距离实验电流直接施加位置一定距离处观测其效应，因此实际上是通过组织观测对实验刺激电流进行滤波，从而构建刺激电流模型。整个研究中，我们设定 $I_0 = 0.1$ ms^{-1}，刺激持续时间设为 $\tau_{\mathrm{on}} = 10$ ms，并将刺激电流的空间尺度设为 $(\kappa/h_x)^2 = 5 \times 10^{-4}$，以匹配观测数据中的结果。相邻刺激之间的周期为 118 ms，同样与观测数据相符。我们进行了测试，以评估刺激电流参数对状态的影响；结果发现这些参数分为两类：(a) 亚临界（不具激发性）和 (b) 超临界（具激发性）刺激。在每一类中，对于物理上合理的取值，均未发现有意义的变化。一维情况下的这种临界激发现象已得到充分理解。[36–38]

在图 2 中，我们展示了使用 Barone 等人参数集进行自由运行实验时，CRPS$^{a,\,o}(t)$、RMSE$^b(t)$ 以及组织深度方向上的离散度 SPRD$^b_u(t,\ z)$。CRPS$^o(t)$ 和 CRPS$^a(t)$ 表明背景集合与观测之间存在显著差异，与分析集合均值相比也是如此。RMSE$^b(t)$ 表明，在没有同化的情况下，该集合无法准确重建表面动力学，这符合预期。

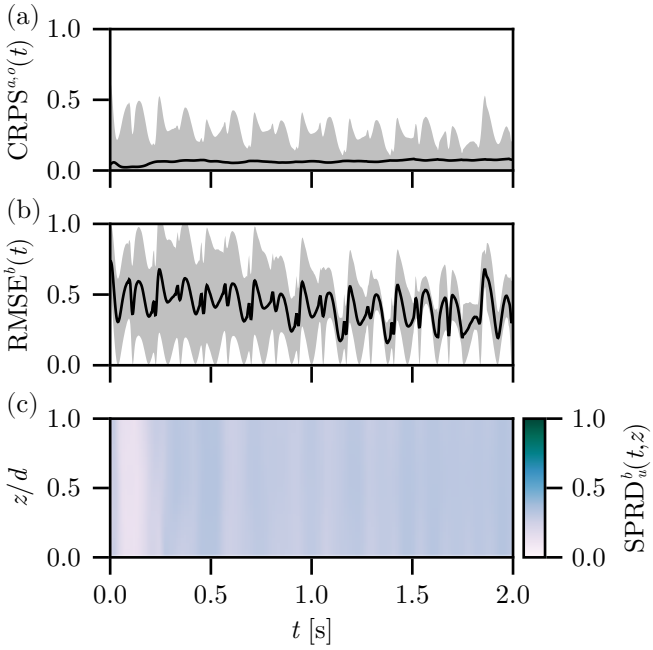

图 2. 自由运行实验结果，展示了使用 Barone *et al* 参数集进行重建时的 (a) CRPS$^a(t)$（线）和 CRPS$^o(t)$（带状区域）、(b) RMSE$^b(t)$（线）及正负一个标准差（带状区域），以及 (c) SPRD$^b(t,\ z)$（颜色）。

此外，RMSE$^b(t)$ 的变异性随时间推移而降低，因为刺激电流的驱动使集合状态的大尺度特征逐渐同步，同时有关集合初始化的信息也慢慢丢失。同样，SPRD$^b_u(t,\ z)$ 对 z 几乎保持不变，这表明当重建过程中不存在观测时，我们正确恢复了集合状态的 z 不变对称性。

这些结果可作为后续实验的比较基准。自由运行实验使用了未经修改的观测集和标准集合初始化，并采用刺激模型，且未进行任何同化。因此，集合均值并不对应于重建，因为没有观测数据提供引导；所以我们不展示平均场。

2. 自治模型

第一个实验采用自主、确定性的心脏兴奋模型作为集合成员，在同化这些观测时，考察模型参数导致的模型误差的重要性；这些参数可能有效地重现观测到的动作电位，也可能无法做到。与其针对每种刺激方案调整模型，不如利用卡尔曼滤波的性质来匹配刺激的周期性，而无需明确采用建模的刺激模式，这样要简单得多——这显然存在局限：我们舍弃了一个潜在的额外信息来源，并在同化过程中引入了延迟，因为集合只能在刺激已经出现在观测中之后 才作出反应。鉴于这一限制，本实验展示了使用 LETKF 同化框架对系统状态进行的最差情形、 «低信息量» 估计。我们将此实验称为«自主实验»。

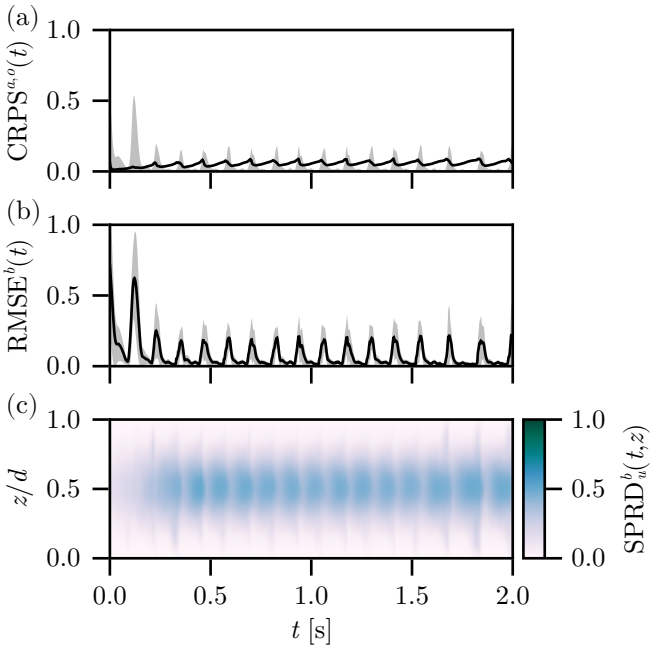

图 3. 自主实验结果，展示了使用 Barone *et al* 参数集进行重建时的 (a) CRPS$^a(t)$（线）和 CRPS$^o(t)$（带状区域）、(b) RMSE$^b(t)$（线）及正负一个标准差（带状区域），以及 (c) SPRD$^b(t,\ z)$（颜色）。

图 3 显示了采用 Fenton-Karma 模型和 Barone 等人的参数集进行重建时，随时间变化的 (a) 连续排序概率得分$^{a,\ o}(t)$、(b) 均方根误差$^b(t)$ 和 (c) 标准化预测半径离散度$^b_u(t,\ z)$ 指标。初始瞬态之后，同化能够很好地重建系统状态的表面，使其与观测相匹配。初始瞬态源于初始集合与观测之间的相位失配，从而导致较大的初始误差。正如预期，连续排序概率得分$^o(t)$与均方根误差$^b(t)$高度相关，而且连续排序概率得分$^a(t)$在较长时间尺度内始终较小，仅缓慢增大，并因底层激发的周期性而振荡。这种行为是该同化实验固有的滞后所致：下一次激发的信息需要传播经过整个局部集合变换卡尔曼滤波，才能对背景集合产生影响。均方根误差$^b(t)$的主要贡献对应于整个区域完全受到激发时出现的峰值；随着激发幅度消散，误差迅速回落。对这一贡献的一种解释是

在表面上，局部集合变换卡尔曼滤波难以处理波前的不确定性，以及记录到的波前形状与非线性模型为背景集合生成的波前形状之间的偏差（模型按构造会滞后于观测到的波前），但能够准确再现波后沿。事实上，在波前传播期间，局部集合变换卡尔曼滤波的创新量较大——它正在修正模型中较慢的传导速度，以匹配数据中观测到的兴奋模式；而在波后沿传播期间，创新量则相对较小。对于这一参数集，离散度初始较小，并迅速在内部增大，这如预期所示，代表了系统状态不确定性的主要贡献。随着激发信息从表面向区域内部传播，显然，仅凭表面观测，局部集合变换卡尔曼滤波无法约束解的内部。尽管如此，$z/d = 1/2$ 处的离散度最大值在 $t \approx 250$ 毫秒后趋于饱和，表明这一时间段反映了内部的真实不确定性。

在表面上，激发后，正如预期，下一次动作电位（AP）由同化驱动：同化在适当区域扰动状态，从而引发新的传播波，使其与观测中的波相匹配。此外，由于不同参数集产生的动作电位具有不同的波速（$c_{\parallel} \approx \sqrt{D_{\parallel}/\tau_d}$），同化会持续直接修正波前正前方的传播。当波速略慢时，这种修正是有效的，因为波前前方的区域预计是可激发的；因此，在波前前方施加微小扰动会产生新的锋面，并在下一次同化时刻之前与已有的滞后波前相连。当波速略快时，这种修正则无效，因为在波前上将 u 的值稍微降低，并不会显著阻碍传播——要做到这一点，需要更大的、可使波熄灭的扰动，该扰动与 I_{fi} 的中间不稳定零流线和静息态之间的间隔有关。这种有效控制的不对称性在表面上显而易见，这是因为表面存在观测；同样，这也意味着未受控制的波速误差会在内部传播。

图4展示了背景集合平均的 z切片，以此说明同化动力学的现象学特征

$\bar{\mathbf{x}}^b$ 在自主实验的三个样本时刻、固定深度处。观测中的一次新激发即将开始时，参见图4(a)，集合均值很好地近似了表面处激发的分布（$z/d = 0.0$，1。0），而内部切片并不能合理地近似贯穿深度的插值函数。相反，内部区域的不确定性过大，无法匹配波在区域内的传播；我们预期该波会在 z中与表面同步均匀传播。

此外，这些切片凸显了模型误差的一个重要成因：集合中锋面的出现时间晚于观测数据中的时间。这种

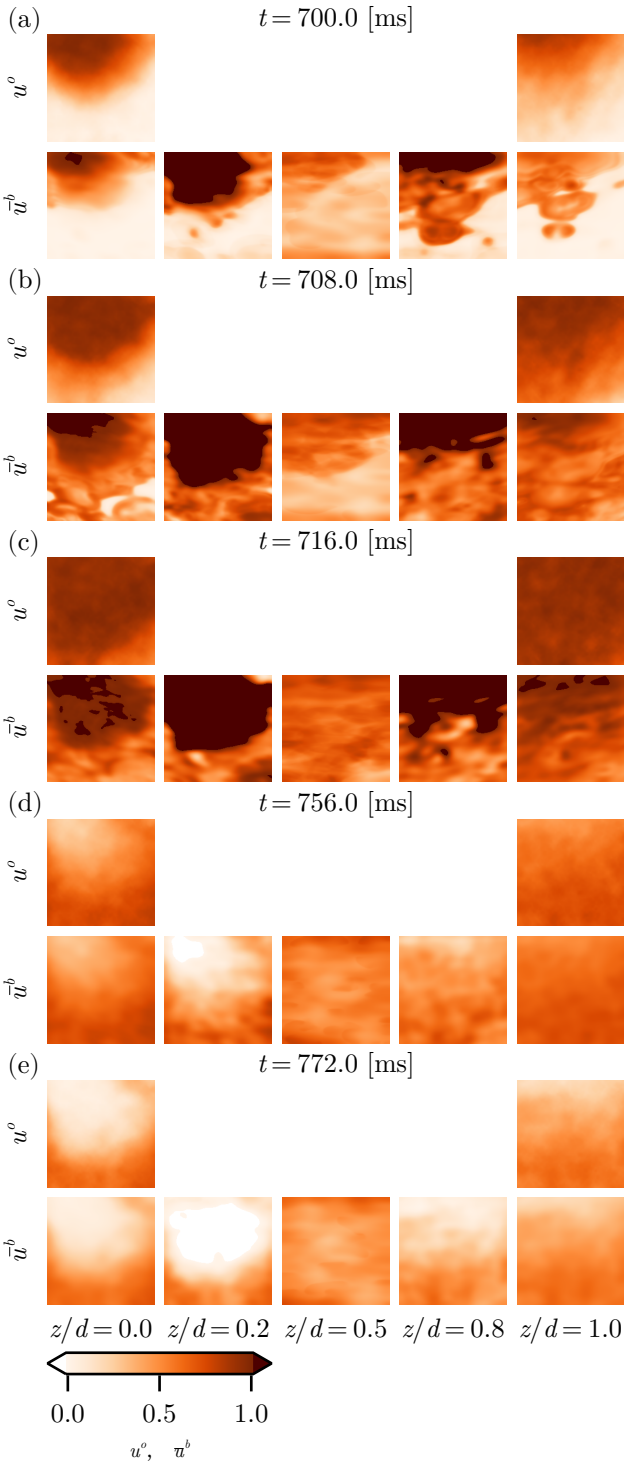

图4. 观测的 u场（u^o）和背景集合平均（$\bar{u}^b$，z/d = 0.0 0.2 0.5 0.8 1.0）在深度、、、、、以及时刻、、、、t = 700 708 716 756 772 ms处的切片，分别对应(a-e)。

行为是由三种现象相互作用造成的，这些现象与观测周期性的时序有关，并影响集合动力学的同步方式。

首先也是最重要的问题是，模型的传导速度比观测数据中的速度慢。尽管参数集经过人工调校以匹配数据[26]，，但若不大幅高估组织电导率，就无法充分匹配观测到的锋面速度（参见参考文献 26，其中的 $\sigma_{\parallel} = 9 \times D_{\parallel}$ 和 $\sigma_{\perp} = 15 \times D_{\perp}$）[39]。因此，集合中的激发波与观测之间的任何时间差，都需要 LETKF 持续进行显著校正。以往的计算机模拟研究发现，LETKF 能够有效处理 3D 模型中的多种系统性模型误差，包括纤维角度旋转、组织电导率以及模型的显式时间尺度。[20–22] 在波前前方，只要集合成员中已有成员在该前导区域处于激发状态，LETKF 就能轻易激发集合成员状态中静息组织的前沿，从而加快波前，使其与观测中的激发传播相匹配。由于传导速度较慢，这种情况很少发生，因为在 LETKF 的驱动下，集合成员会使其相位与观测同步。

模型误差的第二个来源，是自主实验模型中向区域内部传播时固有的滞后。对于波速为 $c_{\perp}$ 的横向波，内部动力学至少会滞后 $\tau \approx c_{\perp} d/2$，也就是表面信息传播到内部所需的时间。因此，即使表面能够正确估计观测中的激发分布，内部的分布仍主要由数个同化步骤之前的动力学决定。这些重建中的差异属于系统误差，均方根误差^{b}指标在很大程度上无法反映这些误差，因为它依赖于表面观测；但如果集合成员在内部存在差异，这些误差应会影响连续排序概率得分^a，同样也会影响标准化预测半径离散度$^b_u(t,\ z)$。

第三个来源是由于缺少刺激信息而导致的自主实验模型动力学滞后。在静息状态期间，下一个激发波在时刻 t被观测到；这些观测被用于局部集合变换卡尔曼滤波，其影响会体现在时刻 t 的分析集合状态中。这些分析集合状态构成动力学模型的新初始条件，而由初始观测引起的这种扰动，其影响会在时刻 $t+T_a$的下一个背景集合中显现，从而使背景集合与观测之间产生不小于一个同化间隔的滞后。如果分析集合状态未达到激发阈值，那么观测到的激发刺激与由此产生的波之间的滞后会更长；这种情况通常会发生，因为按定义，激发的初始观测较小，很可能低于阈值。

最后，我们必须指出观测和重建的若干特征，这些特征将在后续结果中发挥作用。观测从静息状态（$u(t,\ x,\ y,\ z) \approx 0$）平滑地过渡到完全激发状态（$u(t,\ x,\ y,\ z) \approx 1$），横跨波前，参见图4(a)。这种钝化的波前代表了一种当前模型无法实现的配置——该模型具有阈值，因此波前的定义十分尖锐；也就是说，$u(t,\ x,\ y,\ z)$ 很可能接近 0 或接近 1，而沿波前 $u(t,\ x,\ y,\ z)$ 接近 0。5 的概率极低。观测表明，在这个特定时刻，u^o 有显著概率接近 0。5。空间平滑数据与模型尖锐特征之间的这种矛盾，严重阻碍了可靠重建。

此外，尽管该模型的选定参数集是针对这一数据集手动调优的，其动力学仍往往会高估观测到的动作电位幅度（APA），参见图4(a-c)。这一现象首先是由于 LETKF 的分析步过度修正了集合状态。随后，由于模型积分区间相对较短（$T_a = 2$ ms），不足以使状态弛豫至慢流形，这一影响便持续存在。在自由运行实验中也可能观察到这一现象，这是因为刺激电流的频率相对较高（未显示）。值得注意的是，高频起搏刺激电流的影响与分析步的过冲效应并非简单相加。

3. 显式刺激电流建模

本实验考察在模型动力学中加入非自主、空间局部化的显式刺激电流所产生的影响。该刺激电流在空间和时间上均经过调节，以定性匹配实验数据集中观测到的刺激，并考察在采用这一显式建模方案时集合发生灾难性同步的可能性。所用刺激电流与自由运行结果中使用的相同（第 III A 1 节）。我们将此实验称为“刺激”。

模型的结果如图 5 所示，其中重建基本不受是否纳入刺激电流的影响。与自主实验的结果相比，RMSE$^b(t)$略有变化。在激发阶段，重建误差的峰值较小；在自主实验中，刺激电流会在观测-LETKF-模型路径所导致的滞后之前触发激发波。$t = 1.5$ s 后，刺激时序发生漂移，导致刺激提前；LETKF 会在表面上对此进行校正，从而有效限制时序误差。这些提前的刺激表现为 RMSE$^b(t)$相较于基线的小幅（$\lesssim 0.1$）上升，且不包括图 5(a) 中激发波造成的主要影响——表面误差的影响持续时间较短，也表明重建过程对刺激周期设定误差具有鲁棒性；鉴于这些误差与时序相关所造成的影响，这正是一个主要顾虑，

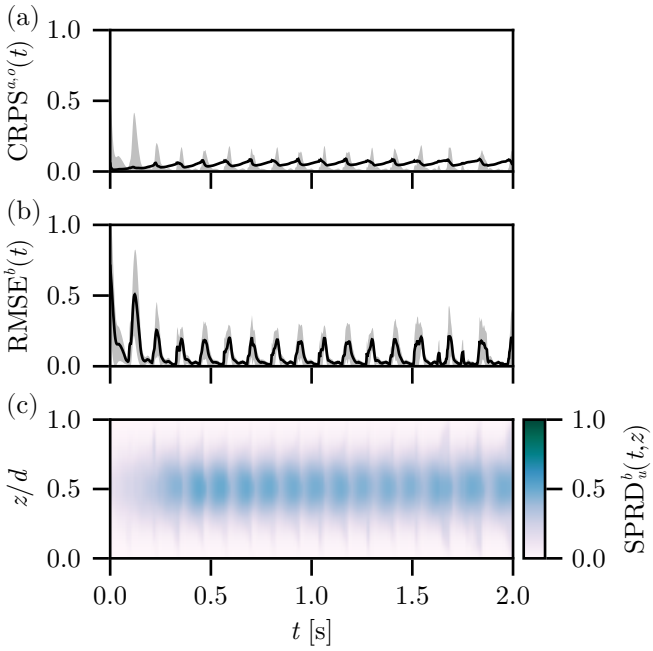

图 5. 刺激实验结果，展示了使用 Barone *et al* 参数集进行重建时的 (a) $\mathrm{CRPS}^a(t)$（曲线）和 $\mathrm{CRPS}^o(t)$（带状区域），(b) $\mathrm{RMSE}^b(t)$（曲线）及其正负一个标准差（带状区域），以及 (c) $\mathrm{SPRD}^b(t,\ z)$（颜色）。

自主实验重建中覆盖的分布。

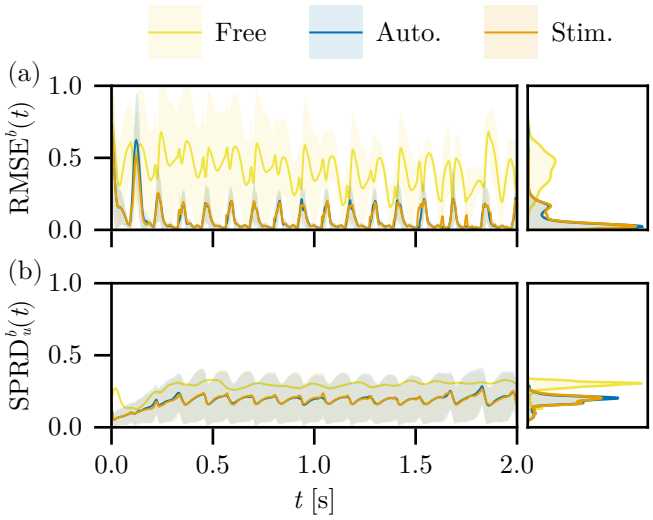

图 6. 自由运行、自主实验和刺激实验的 (a) $\mathrm{RMSE}^b(t)$ 和 (b) 深度平均 $\mathrm{SPRD}^b_u(t)$，以及随时间变化的（右上）表面误差分布和（右下）集合离散度。

图 6 展示了使用 Fenton-Karma 和 Barone 等参数集进行重建时，自主实验、刺激实验和自由运行实验随时间变化的 (a) $\mathrm{RMSE}^b(t)$（左），以及表面误差的边缘密度图（右）。其与自由运行实验相比，RMSE^b 有所改善，这一改善具有显著性，但自主实验和刺激实验的表面误差总体上及随时间变化都相近。刺激实验结果相较于自主模型在 $\mathrm{RMSE}^b(t)$ 上最显著的改进是初始瞬态更短：第一次激发（$t \approx 125$ ms）在表面上的振幅误差较低，因此集合能更快与真实刺激节律同步。RMSE^b 清楚地表明，纳入 LETKF 能够总体降低表面误差的期望值。自主实验和刺激实验的表面误差几乎完全相同——看来，将刺激信息纳入模型动力学，对长期表面重建误差影响甚微，除了上述细微优势之外。自由运行实验的深度平均 $\mathrm{SPRD}^b_u(t)$ 大于自主实验和刺激实验，反映出缺少 LETKF 时重建的置信度较低。事实上，自主实验和刺激实验结果的深度平均 $\mathrm{SPRD}^b_u(t)$ 几乎相同，但后者随时间变化的幅度略有偏大，见图 6(b)。

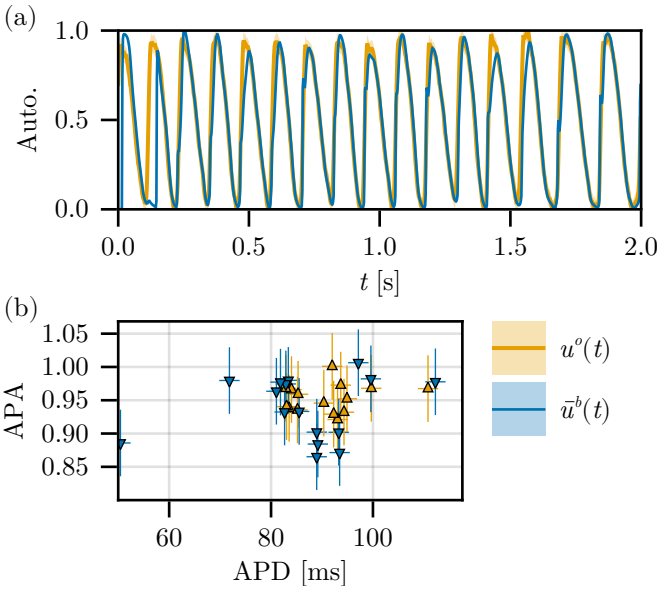

图 7. (a) 自主实验的时间轨迹及相关观测，采样位置为心外膜 ($z/d = 0$) 上的 $(x,\ y) = (0.825,\ 0.42)$ cm；(b) 对应的动作电位时程 (APD) 和动作电位幅度 (APA) ($u_{\mathrm{thr}} = 0.1$)。

除了场的重建误差外，还值得考虑由状态推导出的、与生理相关的量的误差，例如那些源自动作电位结构的量，包括动作电位时程（APD）和幅度（APA）。我们在图 7(a) 中报告了观测中随机选取位置处自主重建的时间轨迹。经过一个持续时间不足两个 BCL（{v1}ms）的

初始瞬态后，该点的重建效果很好，动作电位峰部出现了轻微误差，因此导致真实 APA 的估计不佳，参见图 7(b)。观测轨迹与重建轨迹的 APD 分布极为吻合，这符合预期，因为重建中的主要误差似乎出现在远离阈值（$u_{\mathrm{thr}} = 0.1$）的位置。补充材料中收录了所有由 LETKF 驱动的重建结果的类似计算，参见图 S1。

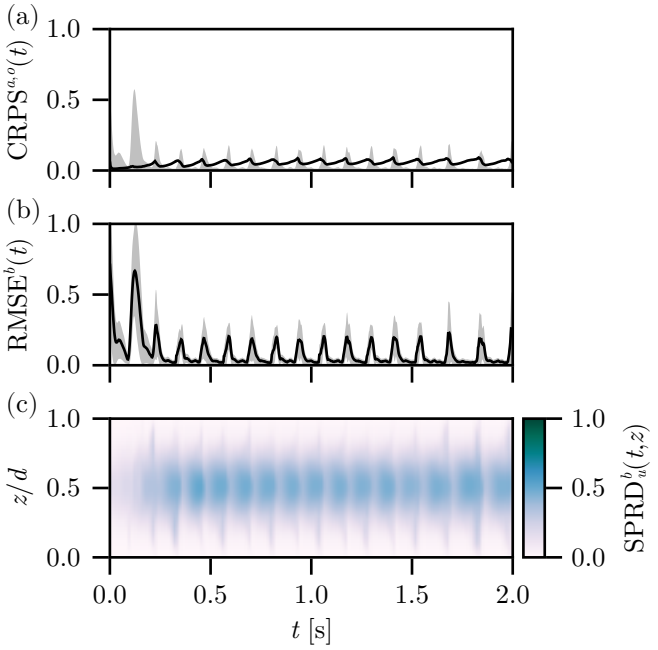

图 8. 随机实验结果，展示了使用 Barone *et al* 参数集进行重建的 (a) $\mathrm{CRPS}^{a}(t)$（线）和 $\mathrm{CRPS}^{o}(t)$（带），(b) $\mathrm{RMSE}^{b}(t)$（线）及正负一个标准差（带），以及 (c) $\mathrm{SPRD}^{b}(t,\ z)$（颜色）。

4. 随机电流模型效应

本实验将随机性引入模型，以研究在存在非控制模型误差的情况下，其对状态重建准确性的影响。我们此前介绍过这些随机模型效应，用于同化由模型生成的状态序列所产生的观测数据，并考虑了模型不确定性[22]。与确定性集合模型结合使用时，这些技术相比局部集合变换卡尔曼滤波提升了状态重建效果，有时优势显著；将其用于同化实验观测，或有助于在模型误差源于模型无法完美再现实验数据时恢复状态动力学。本文将向模型中引入随机电流（SDE-u）和随机选择的时间尺度参数（SMP-τ）。SDE-u 公式在（1）中电压变量 u 的动力学里引入新的随机电流 I_{sto}，使其发挥 I_{stim} 的作用，

$$I_{\mathrm{sto}} = \xi, \quad \xi \sim N(0, \sigma_u^2), \tag{14}$$

而 SMP-τ 会影响模型所有显式时间尺度的选择，

$$\tau_* \sim N(\overline{\tau}_*, (\overline{\tau}_* \sigma_p)^2), \tag{15}$$

其中 $\overline{\tau}_*$ 是基准参数值。我们将此实验称为“随机实验”。对于两种随机效应，我们采用固定的标准差（$\sigma_u = 0.04$ 和 $\sigma_p = 0.04$）。

图 8(a) 展示了采用 Fenton-Karma 模型 Barone 等人参数集时，使用基于电流的 SDE 和 SMP-τ 随机效应进行重建的表面误差均方根误差$^{b}(t)$。对于这些幅度相对较低的随机效应，我们发现表面重建误差仅发生了极小变化——与激发波前和波后沿对应的均方根误差$^{b}(t)$峰值被有效平滑，这符合均匀随机效应的预期。与刺激结果类似，均方根误差中的瞬态

$^{b}(t)$因加入随机效应而略有恶化；$t \approx 125$ ms 附近的激发出现了更大的峰值误差。

在先前的工作中，我们表明，随机扰动能够通过有效扩大给定集合规模下的离散度，使集合更好地逼近真实动力学的局部自由度。在本次实验中，我们没有发现这种效应。我们发现，加入随机电流会加快深度平均 $\mathrm{SPRD}^{b}_{u}(t)$，使初始相位内的平均离散度增长更快；随后，在长期重建过程中则转为保守效应，抑制离散度中的短暂变化。事实上，这种行为表明，在某些情况下，向模型中加入随机过程可能会增强集合的同步性，降低集合离散度，或抑制集合的完全失同步。

比较自主实验（图 3）、刺激实验（图 5）和随机实验（图 8）模型实验对应的表面误差 $\mathrm{RMSE}^{b}(t)$，我们发现，加入刺激电流或随机效应对整体重建误差的影响很小。对此，我们最多可以将其解读为：在动力学模型中加入随机效应或刺激电流，只会略微提高表面重建的准确性。或者，我们也可以认为，自主实验的结果实现了低于预期的重建误差，且置信度与这些信息更充分的模型方法相当。同样，由于观测数据和模型中刺激电流的时序缓慢漂移，我们可以确信，重建对于与刺激时序相关的模型误差具有稳健性。

B. 观测扰动对重建精度的影响

如果对模型的修改对状态动力学的重建只有细微影响，那么对实验观测的调整或许会产生更显著的影响。在这些实验中，我们对实验数据中的观测进行扰动，以更好地模拟记录数据所涉及物理过程的不确定性；或者依据数据表面呈现的时间结构，为极其稀疏的观测补充一些中间信息。同样，本研究也提供了一个机会，以确保我们的重建对于心脏研究人员可获得的这类数据具有稳健性。

1. 添加合成观测

在本实验中，我们探索生成中间深度的合成观测，并将其与表面观测一并用于同化过程。这些合成观测的作用是轻柔地约束组织内部的集合离散度；我们此前已发现，组织内部是该系统状态重建误差的主要来源之一[22]。内部合成观测通过对投影到心外膜表面和心内膜表面的观测值进行插值得到。合成观测被追加到真实观测之后，前者具有固定的逐点不确定性估计$\eta = 0.50$，后者具有固定的逐点不确定性估计 $\eta = 0.05$。中间深度观测的不同空间分布会显著影响重建的稳健性和精度。在与表面相同的位置对（x，y）网格进行采样的完整观测布置，会因分析步强烈收缩而导致集合立即坍缩（未展示）。我们采用相对稀疏的布置，将 64 个观测均匀分布在一个 8×8 网格中，这并未导致集合坍缩。

由于底层激励具有时空模式，我们加入了这些合成观测；波在表面上几乎均匀地穿过（x，y），并随时间推进。因此，我们将每个表面上的观测视为强耦合的，因而可以通过深度进行插值，即在各表面集合之间插值。我们报告了额外增加的 64 个观测（观测总数增加了 0.7%）对

重建保真度的影响。

图 9 展示了使用合成中间深度观测进行同化的结果。我们立即注意到，尽管 CRPS$^a(t)$ 未受到显著影响，但 CRPS$^o(t)$ 的峰值现在显著高于其他任何同化实验，参见图 9(a)。这一发现立即表明结果与观测存在显著偏差，但未能揭示这种差异

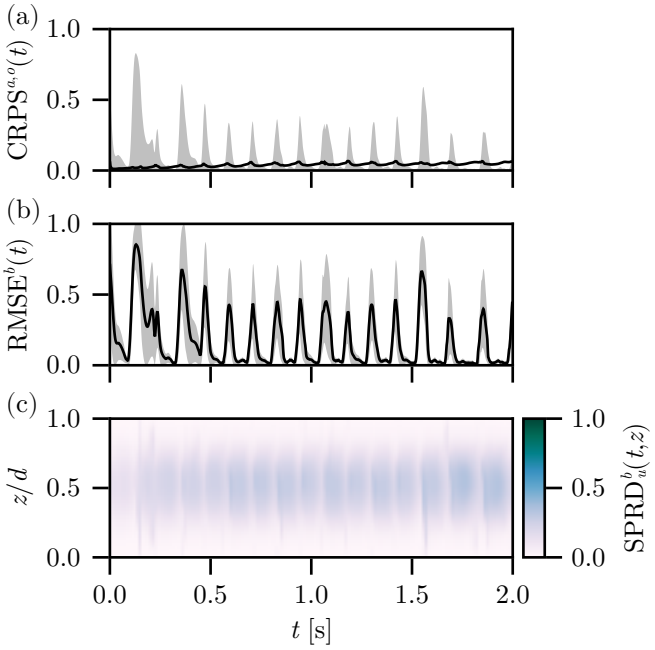

图 9. 合成观测实验结果，展示了使用 Barone *et al* 参数集进行重建时的 (a) 连续排序概率得分$^a(t)$（线）和连续排序概率得分$^o(t)$（带），(b) 均方根误差$^b(t)$（线）及其正负一个标准差（带），以及 (c) 标准化预测半径离散度$^b(t,\ z)$（颜色）。

是由内部、合成观测，还是“真实”的表面观测造成的。RMSE$^b(t)$ 可用于区分这些因素；它的峰值同样显著高于其他任何同化实验（参见图 9(b)），且为与其他重建实验保持一致，仅根据表面观测计算。值得注意的是，RMSE$^b(t)$ 可以分解为来自表面的贡献和来自内部合成观测的贡献；若进行这种分解，在插值假设正确且合理的前提下，便可直接估计内部重建误差。然而，我们没有严谨的方法来评估插值假设，因此无法据此断言内部重建是正确的。因此，我们得出结论：尽管合成观测的不确定性很高，将根据表面观测插值得到的内部合成观测纳入其中，仍会导致表面重建结果显著变差。

通过考察集合离散度，或许可以解释这一难题。尽管额外的中间深度观测非常有助于限制区域中间深度处重建的不确定性（参见图 9(c)），由于表面观测的信息能够影响解的内部层，这种影响并不总是理想的。通过约束内部区域，我们也限制了用于重建表面的状态离散度。这样一来，这个受到限制的状态集合可能无法充分涵盖观测到的表面动力学中存在的物理不稳定性，

从而为了不确定的观测（即内部观测），牺牲了对我们有信心的观测（即表面观测）的重建精度。与表面观测相比，对内部观测的不确定性和分布进行调节，以有针对性地约束集合，由此提供了进一步优化的空间，不过这超出了本研究的范围。

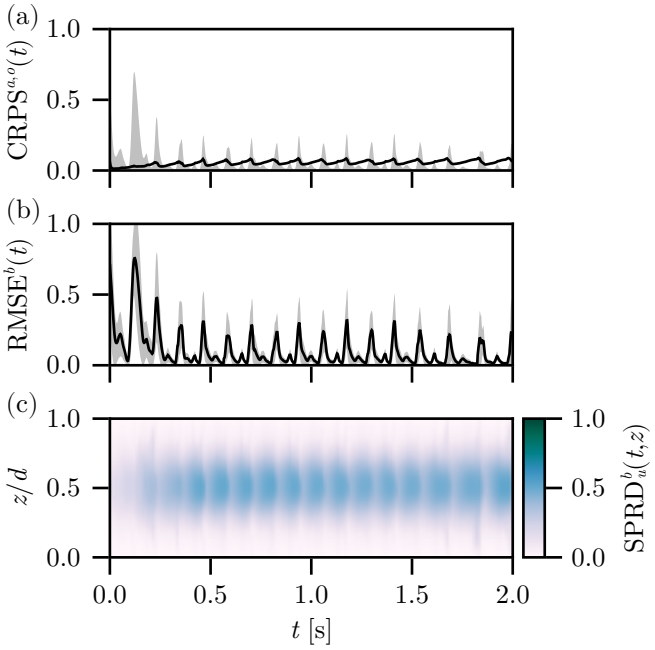

图 10。波前不确定性实验结果，展示了使用 Barone *et al* 参数集进行重建时的 (a) $\mathrm{CRPS}^{a}(t)$（曲线）和 $\mathrm{CRPS}^{o}(t)$（带状区域）、(b) $\mathrm{RMSE}^{b}(t)$ 加减一个标准差（带状区域），以及 (c) $\mathrm{SPRD}^{b}(t, z)$（颜色）。

2. 锋面观测不确定性建模

在本实验中，我们旨在根据对实验数据的物理与数值滤波过程以及模型动力学的认识，为观测数据赋予不确定性。为此，我们识别了 u 在静息态（$u \approx 0$）和激发态（$u \approx 1$）附近的慢动力学，并注意到，尽管模型产生了清晰的波前特征，观测数据却相对平滑。在实践中，这意味着观测不确定性取决于观测值 $\eta_i = \eta(\mathbf{y}_i^o)$，具体而言，取决于 $u < u < u$的值，其中 $u = 0.1$对应于标记静息状态边界的阈值，而 $u = 1.0 - u$标记完全激发态的边界。非线性不确定性函数采用缩放并截断的高斯分布形式，

$$\eta(\mathbf{y}_i^o) = \underline{\eta} + (\overline{\eta} - \underline{\eta}) \exp\left(-k^2 (u_i - (\overline{u} + \underline{u})/2.0)^2\right), \quad (16)$$

其中，$\underline{\eta} = \underline{u}/2 = 0.05$ 和 $\overline{\eta} = 0.50$ 分别是不确定性的下界和上界，而 k 的选取使得 $\eta(0.0) = \eta(1.0) = \eta$ 且 $\eta((u + u)/2) = \eta$。对波前不确定性最大值 η 最简单的解释是：沿波前的任意位置，由于观测数据经过平滑处理，无法确定集合成员在同一时刻的同一位置是否应处于完全激发态还是静息状态，即 $|u^b - \mathbf{y}_i^o| \sim U(-\eta, +\eta)$。原则上，这种方法会促使集合离散度沿波前增大，从而使集合均值能够很好地近似观测数据，同时各个集合成员仍保留清晰的波前特征。调整不确定性分布的具体形式，无论是采用不同的（η, η）组合，还是完全改用其他依据构建的非线性函数 $\eta(\mathbf{y}^o)$，均超出了本研究的范围；在本研究中，我们旨在扰动现有观测数据，以确定能否提高重建精度。

在图 10 中，我们报告了波前不确定性实验的影响。我们发现，额外的
观测数据中波前和波后沿的不确定性，导致波前末端附近的均方根误差$^{b}(t)$小幅但持续增加；同样，波后沿末端的表面误差也随之增大，并在通常与波前末端相关的一组峰值之间，产生一系列较小的均方根误差$^{b}(t)$峰值。由于我们正在促使集合成员在激发波的起始和末端附近产生偏差，因此这一发现并不意外——在激发与未激发状态的转换处，离散度会略微增加，并在误差本就较低的时空位置，对表面误差产生最显著的相对贡献。此外，由于观测数据在观测协方差逆加权作用下会促使集合成员趋于同步，^{b}，观测的不确定性越高，整体同步程度就应越低，集合离散度也越大。与自主实验相比，我们观察到标准化预测半径离散度$_{u}(^{b}$，$)$初期增长更快，但长期最大值并未发生显著变化。

在激发波的起始和末端附近促使集合成员产生偏差，因此这一发现并不意外——恰恰在激发与未激发状态的转换处，离散度会略微增加，并在误差本就较低的时空位置，对表面误差产生最显著的相对贡献。此外，由于观测数据在观测协方差逆加权作用下会促使集合成员趋于同步，R^{-1}，观测的不确定性越高，整体同步程度就应越低，集合离散度也越大。与自主实验相比，我们观察到标准化预测半径离散度$^{b}_{u}(t,\ z)$初期增长更快，但长期最大值并未发生显著变化。

在图 11 中，我们展示了自主实验、波前不确定性实验和合成观测实验的 $\mathrm{RMSE}^{b}(t)$ 和 $\mathrm{SPRD}^{b}_{u}(t)$。与自主实验相比，波前不确定性实验和合成观测实验总体上都会导致更大的表面重建误差（RMSE^{b}），但它们随时间变化的相对贡献极不均匀。自主实验的重建结果能够匹配观测数据的动作电位时程和持续时间指数，而合成观测实验的重建结果则会大幅偏离准确的重建结果，这种偏差持续时间往往超过观测数据的动作电位时程。同样，波前不确定性实验明确地以牺牲对波前精确位置的置信度为代价，换取更高的集合离散度所带来的稳健性，但这种稳健性在实践中始终未能体现；自主实验的

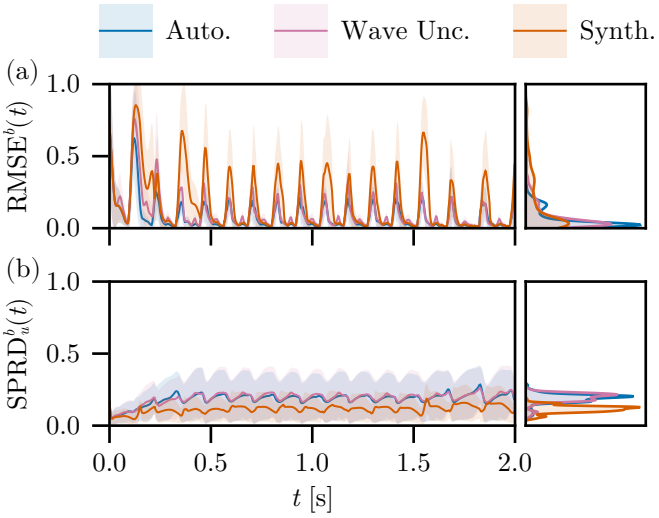

图 11. (a) 自主实验、波不确定性实验和合成观测实验的 RMSE$^b(t)$ 以及 (b) SPRD$^b_u(t)$，同时展示了随时间变化的（右上）表面误差分布和（右下）集合离散度。

集合离散度与波前不确定性实验的集合离散度相当。

IV. 讨论

在所有结果中，未观测且未受 LETKF 分析集合更新驱动的内部层，是重建误差的主要来源，这与此前使用模型运行生成表面观测所得的结果一致[22]。然而，此前的实验侧重于自主且（在大尺度上）由拓扑特征组织的动力学，即卷波的动力学围绕其丝状体组织；本研究则侧重于必然具有非自治性质且受外部刺激驱动的实验结果，因为这些条件普遍存在于心脏组织兴奋的实验研究中。在当前情形下，所有动力学均由观测及其对分析集合的 LETKF 更新所驱动；自由运行实验除外，该实验对应于不存在任何观测的情形，且动力学不受此类拓扑约束。

A. 模型误差与同化的相互作用

有几个因素会显著影响该数据集重建的准确性和稳定性。首先是模型误差，这主要由参数设定不当造成。例如，当不应期足够长时，集合变换卡尔曼滤波（LETKF）生成的后续激发会因模型的不应期特性而发展为传导阻滞。受阻的波前随后会妨碍准确重建下一个同化步骤中的状态。此外，在模型参数化中匹配波的传播速度也很重要：慢波依赖同化方案激发波前前方的组织，以匹配观测；而快波则依赖同化方案抑制模型中的波传播，以匹配观测到的波前。乍看之下，点燃问题与熄灭问题之间的不对称性竟会显著影响状态重建的可靠性，这一点令人意外。然而，这种效应与我们对可激发介质临界转变的理解相符：点燃问题仅取决于可激发介质的局部阈值响应，而熄灭问题则对激发的具体模式敏感，因此是一个非线性更强的过程。

模型的参数化也会影响集合状态的动力学范围，使分析可能将场值推到观测范围之外，即出现“过冲”。在当前问题中，Fenton-Karma 模型的动作电位仅被弱约束在与观测相同的区域内，即 $0 \leq u \leq 1$；而同化步骤对分析状态采用严格的边界 $0 \leq u^a \leq 1.5$。由于模型积分时间较短（$T_a = 2.0$ ms），且分析步发生过冲，状态无法足够迅速地松弛以补偿过冲，因而产生异常大的场值；这些场值也必须在下一次同化步骤中得到处理。这种模型误差与同化之间的反馈循环，为进一步研究同化滤波器提供了机会。

提高重建过程稳健性的一种可能方法是，除了状态向量之外，也将重建扩展到模型参数。这种方法在缺乏已知模型（参数集）的情况下，尤其适用于模型（参数）识别。我们选择将这一研究方向留待未来开展，一方面是因为已有文献针对该数据集和模型发表了模型参数集[25]，另一方面是因为我们此前已证明，随机模型膨胀技术善于处理模型误差（尤其是由参数化不当导致的误差）[22]。深入研究如何利用数据同化技术识别心脏电激发模型的模型参数，或可推动临床应用取得有益进展。

B. 观测与模型动力学的交互作用

观测处理也比乍看之下更为微妙。光学标测记录不仅记录组织最表层细胞发出的荧光，还会记录下方受激细胞层传递上来的荧光，因此实际上会模糊波前的位置。这种模糊表现为：波前不再尖锐，而是形成一个从静息态平滑过渡到兴奋态的前沿，宽度可达 1 cm。为提高数据的信噪比而进行的数值后处理——

通常采用时空卷积，这会进一步平滑波前的尖锐特征，使其变成临界激发细胞的分布。然而，由于我们在恒定起搏期间使用的实验数据已达到稳态，因此采用了叠加[40]，，以提高信噪比而不进一步进行空间平滑。阈值可激发模型无法重建这种临界激发细胞的分布——我们依靠同化方案，有效地将集合模型生成的陡峭波前解释为给定的平滑观测波前。我们考虑了状态观测$\mathbf{y}^b$的模糊效应：观测算子并非只对特定位置的场进行采样，而是通过固定宽度的方框滤波卷积计算场的局部平均。我们预计，卷积的长度尺度会与局域化过程的长度尺度相互竞争；同样，卷积滤波器也会与 LETKF 的滤波步骤相互竞争。我们预期，集合的信息含量总体上会减少，因为现在，l维$\mathbf{y}^b$的各个元素之间的相关性显著增强，从而在集合观测中引入实质上无法控制的误差，损害重建效果。我们发现，实际应用中，与本研究采用的基于采样的观测算子相比，差异很小——更多细节和结果请参见补充材料中的图 S3。

同样，心脏实验记录的观测不确定性估计也是一个重要课题，而我们在本研究中仅对其进行了浅显探讨。由于光学标测记录及其后处理涉及空间平均，这在本质上会平滑真实可激发模型中常见的陡峭波前，因此，我们可以像第 III B 2 节那样，根据观测值的近似大小，启发式地为观测数据指定非线性不确定性。由于数据中跨膜电位实际上呈现出三种“缓慢”状态——静息、轻度激发和完全激发——而且我们认为中间状态显然是测量方案造成的假象，因此可以估计处于该状态的细胞在下一观测时刻（等效地说，即下一个同化间隔）仍保持在同一状态的概率，并将这一比例因子解释为所观测跨膜电位的不确定性。对于 Fenton-Karma 模型，我们可以根据 u 和模型参数的取值范围，将这些观测到的缓慢状态映射到模型中，具体而言，$u_c < u_i(t) < u_c^{si}$ 表示轻度激发状态。此范围以外的观测不确定性应由观测的噪声底决定。由于 u_c 和 u_c^{si} 的真实值未定义，但可以假设其接近本研究中使用的值，因此需要将这一映射平滑地扩展至所有观测值；根据状态设定非线性观测不确定性的启发式方法，会降低对波前和波后沿观测的权重（即提高其不确定性）。要精确构建这些估计，需建立观测的统计模型，这超出了本研究的范围。我们基于启发式方法开展的初步实验表明，通过平滑地提高轻度激发区域内观测的不确定性，这一细微的不确定性调整可能会显著影响状态重建的准确性和稳健性。

最后，虽然我们针对重建任务重点研究了一个特定数据集，但也使用相关数据（BCL $= 234 \pm 2$ ms）和不同的模型参数集计算了重建结果。对于较慢的基本周期长度，结论相同：观测中的激发信息向集合状态向量传播较慢，需要 LETKF 持续进行校正，这表现为表面重构中出现与每次新激发相对应的尖峰误差。对于其他模型参数，周期性仍然匹配，但波形差异会导致重建误差显著增大。这些实验表明，我们的方法对数据集和模型设定中的特殊情况具有稳健性。更多详情请参见补充材料中的图 S4–S8。

C. 同化与观测不确定性的相互作用

LETKF 的规格包含若干数值参数，我们发现这些参数会影响同化程序的效果。第一个是名为“粗差”的浮点参数，以 σ_g 表示，用于控制某个观测与背景均值$\mathbf{x}^b$的偏离程度是否足以将其截断，其目的是让同化过程保持“轻触式”处理，并在观测值存在较大偏差时提高方案的稳定性。对于一组观测—不确定性配对 $(\mathbf{y}_i^o, \eta_i)$ 以及相应的背景均值场值 $u_i^b(t)$，若 $|\mathbf{y}_i^o(t) - u_i^b(t)| > \sigma_g\,\eta_i$，则 $(\mathbf{y}^o - H(\mathbf{x}^b))$ 中的相应元素会被覆盖为 0，也就是说，分析集合的更新不受该观测影响。由于 $0 \lesssim u^b \lesssim 1$ 的取值范围以及表面观测的不确定性下限 $\eta = 0.05$ 都是固定的，当 $\sigma_g \gtrsim 20$ 时，每个同化步骤都会使用所有观测。然而，我们发现，将该参数设为低于饱和值（例如 $\sigma_g = 10$）会显著影响结果，使表面重建误差的最大值比饱和情形（$\sigma_g = 20$）低近 25%。因此，本研究的所有实验均设定 $\sigma_g = 10$。程序日志清楚显示，在自主实验中，到第 50 个同化步骤（100 毫秒）时，被忽略的观测不到 5%，这表明我们只过滤了真正极端的值，而且只发生在同化初始迭代阶段。原则上，调节这一参数类似于膨胀：它衡量我们对观测和背景集合的相对信任程度

成员。该参数应取何值$a\ priori$尚不明确；由于它会影响重建精度，未来研究应优先对其影响进行表征。

此外，我们在整个研究中使用了 $\eta = 0.05$ ，其估计值依据观测值分布的半高全宽得出。虽然我们发现不确定性的具体取值会带来一定影响（例如，使用 $\eta = 0.025$和$\eta = 0.10$ 进行的实验总体上产生了均方根误差[b]较高的重建结果），但这些结果在定性上与本文所呈现的结果相似。根据数据集的统计特性严格估计观测不确定性超出了本研究的范围，但这或许会表明，通过更好地估计观测协方差矩阵，可以提高重建精度。

在本研究中，所有实验的状态变量分析更新均采用了“强耦合同化”（类比天气预报中域分解模型所使用的同一术语）。我们允许对状态 u 的观测影响v 和 w 场的分析更新。相比之下，“弱耦合同化”仅通过模型评估在状态变量之间传递信息。强耦合同化预计能够从相同的观测中提取比弱耦合同化更多的信息[41]。对于相对稀疏的观测模式，强耦合同化允许对状态进行更大幅度的修正，但也有可能过度驱动 v 和 w 的修正，使其偏离心脏激发的典型情况。将对 u变量观测所引起的分析更新限制在仅影响 u 变量，可以在分析中保留更多集合状态信息，从而降低同化效果，同时允许由于 v 和 w的非受控动力学而产生更大的集合离散度。在数值实验中，我们发现，在某些情景下，同化过程中可以将门控变量驱动至其静息态值；而在具有相近动力学的模型模拟中，门控变量从未达到其静息值。尚待探讨的问题包括：心脏模型相对较弱的耦合是否足以缓解强耦合同化可能暴露出的某些稳健性挑战，抑或弱耦合同化会不可接受地降低重建精度。在心脏背景下系统分析强耦合与弱耦合同化，将有助于厘清这一选择对稳健性和精度的影响，应成为未来研究的重点。

此外，当我们发现 LETKF 对门控变量的驱动过强，导致其取值在物理上不合理时，便将门控变量限制在其已明确知晓的极限值之间， $0 \le v^a,\ w^a \le 1$。这种限制可防止同化过程生成无法达到、且与真实动力学几乎无关的分析状态——这对模型而言，实际上相当于一种初始条件，因此会带来隐患。此外，我们还发现，在 u^a 中也会出现类似效应：部分集合成员会被驱动着对状态进行幅度极大且高度局部化的修正，导致集合离散度异常增大（即 $\mathrm{SPRD}_u^a \gtrsim 4.0$）；对于 u 的典型取值，这种情况本不应发生。为抑制这种增长，我们对 $-0.1 \le u^a \le 1.5$ 加以限制，这也带来了限制集合离散度的积极效果。此外，虽然对 u^a 的限制足以约束集合离散度，却不足以防止过冲，以致 $u^a > 1 \ge u^o$。不过，目前尚不清楚，这究竟只是为应对模型与局地集合变换卡尔曼滤波交互中的数值不稳定性，而在同化过程中加入的一项无害补充，还是反映了我们方法的局限性。对于模型初始条件的约束，其数学定义也不够严谨，因此无法断言这些限制在当前模型或其他模型中永远不需要调整。

观测局地化由两个空间参数和一个时间参数控制；后者设定为仅当前时刻影响同化，因为这符合临床应用需求，且计算效率较高。空间局地化尺度为各向同性，我们尚未研究不同局地化尺度或各向异性的选择。原则上，鉴于表面观测密度较高且噪声较低，从计算角度而言，让这些观测的影响延伸至组织深处，同时减少它们在表面上的重叠，或许更有利。然而，这种观点在物理上难以论证——事实上，它假定内部存在表面主导动力学，而这正是我们在本研究中探讨的现象学问题，只不过处于抽象层面。从物理角度看，每个观测的影响范围可能延伸至这样一个空间区域：在两次观测之间的时间内，该区域可能因激发信息的传播而被激发——也就是说，对于静息状态，$\sigma_{o,i} \propto c_i T_a \propto \sqrt{D_i}$ ——这会使其完全呈各向异性；实际上，与沿表面的影响相比，其向组织厚度方向的影响范围会更短，因为 $D_z = D_\perp \ll D_\parallel$的设计如此。相应的、结合纤维信息的局地化尺度之间具有与电导各向异性比的倒数相关的关系，$\sigma_{o,\parallel} = (D_\parallel / D_\perp)\sigma_{o,\perp}$，并随空间位置而变化。同样，这种缩放方式也为观测分布提供了依据——虽然我们只能获得电荧光的表面记录数据，但依据心脏兴奋特有的性质来考量局地化，或许能让心脏研究人员受益。

D. 在缺乏“真值”状态时的准确性测量

最后，我们关注状态重建结果的表面误差；在缺乏内部观测或“真实”状态的情况下，采用连续排序概率得分（CRPS）来衡量集合一致性。选择这一度量，是因为它具有关于集合规模的收敛性质且易于解释，而非因为它与心脏状态重建特别相关。其他研究采用了替代性的误差度量，例如

基于阈值的误差[21],，它需要参考未知真实状态、预先设定的阈值以及对模型动力学特性的了解。要构建一种公平的（按 CRPS 的意义而言）未知状态特征误差度量，需要建立一个描述状态本身动力学的统计模型，以便将重建状态的期望值与之比较。如上所述，这项工作也有助于基于物理原理估计观测不确定性。

为状态重建问题创建动力学统计模型，正是机器学习领域研究人员采用的方法，并已展现出令人鼓舞的结果。文献42研究了长短期记忆循环神经网络、自编码器和扩散人工神经网络（ANN）模型在简单可激发模型三维重建问题中的应用，同时改变模型可用的历史数据量和重建的层深度。该方法的核心结果令人鼓舞，尤其是在深度较浅（接近观测表面）且历史数据较长时。值得注意的是，考虑到与本研究相比，他们的观测数据较为精简（$l/N \approx T \cdot 8.3 \times 10^{-3}$，对于$1 \leq T \leq 32$），作者发现，尽管动力学模型和几何形态截然不同，其重建局限性与我们的方法相似。文献43采用基于 U-net 架构的人工神经网络，研究了 Aliev-Panfilov模型的重建问题，并像本研究一样考察了观测所含的信息量。具体而言，作者发现，同时观测可激发介质的顶层和底层，比仅观测单个表面能获得更可靠的重建结果；而对组织深度进行积分的“投影”观测效果更佳。在“层流”动力学状态下（孤立卷波的尺度与组织深度相当），他们使用 U-net 人工神经网络架构获得的结果，与此前采用不同心脏激发模型并结合 LETKF 和随机膨胀、针对类似动力学模式得到的计算机模拟结果相当。[22] 作者还发现，将深度信息编码进观测数据对其方法的成功至关重要，这限制了该方法在心室颤动中的应用。[43]

将统计模型（机器学习）与动力学模型（数据同化）方法结合起来进行重建，是一个活跃的研究方向。集合数据同化方法的一个主要实际局限是，存储需求随 $O(MN)$ 增长，而计算量（对于 LETKF[32]）随 $O(M^2N)$ 增长；以预训练的机器学习模型替代部分集合成员，或可降低这两项成本。同样，机器学习模型需要训练，且无法提供预测不确定性的估计；借助数据同化框架进行协同，可利用现有基础设施缓解后一个问题，而通过将动力学模型热切换为机器学习模型，则可能缓解前一个问题。

尤其是，我们发现，集合可能需要长达 250 ms 才能稳定下来，对该系统的真实不确定性作出良好近似；机器学习或可通过基于大量训练数据，更准确地估计初始状态协方差或集合状态初始条件，从而提供一条加速收敛的途径。

V. 结论

本研究使用真实物理数据开展了多项同化实验，并对模型和观测进行了变化。我们发现，与同化模型生成的观测相比，同化实验观测以进行心脏重建面临独特挑战。状态重建的主要难点包括：所研究动力学的类型，即受刺激的非自主实验动力学；参数设置不当导致的模型误差；同步输入（即刺激电流）的存在；以及实验观测不确定性的估计。

当模型生成的动力学类型与实验数据所描述的动力学存在根本差异时，我们需要依靠 LETKF 进行大幅修正。而这些大幅修正又会增加生成物理上不可实现状态的可能性。在心脏背景下，非物理行为可能表现为激发的传播速度远快于当前参数设置下模型所允许的速度。因此，模型能否再现波速，是再现状态动力学的一项重要特征。

本文仅简要涉及但十分重要的一个问题，是调整模型参数以解释数据特征。对于该数据集而言，尤其重要的是重现交替现象——包括激励持续时间（APD）和幅度（APA）的交替现象。虽然使用 Fenton-Karma 模型可以轻易重现 APD 交替现象，但 APA 交替现象并不是该模型的常见特征。可以设计对动力学或模型参数进行修改，以便在 Fenton-Karma 模型中重现 APA 交替现象，但这超出了本研究的范围。

模型误差是表面重建误差的主要组成部分，并且对集合在内部的离散度以及对内部重建的置信度具有显著影响。由于观测随时间呈现周期性而产生同步，CRPS^a 可能会对内部重建精度的置信度给出过于乐观的估计。还需要采用不依赖观测的其他方法来评估重建性能。

尽管我们的数据同化方案能够根据模型生成的观测有效地重建状态（这些观测与本研究所用的分布模式相同），但它难以处理本研究所用的实验数据

。我们推测，这一困难反映了我们此前的三维重建工作与当前数据集之间的根本差异，即观测到的动力学具有非自治性质。利用展现自主动力学（即螺旋波或卷波）的实验数据开展数据同化实验，将是对本研究有趣且富有启发性的拓展。

总之，我们发现，凭借表面上的实验观测，可以可靠地重建受驱动可激发态的表面动力学；但要可靠地重建内部动力学，还需要进一步了解我们的数据集及其与可激发模型动力学的相互作用。我们发现，虽然显式建模刺激电流或随机效应可能会略微改善重建效果，但集合变换卡尔曼滤波（LETKF）、观测和模型之间的耦合已经足够紧密，无法纠正该数据集重建中的主要误差来源。在表面上，这些误差主要源于模型对激发波前和波后的拟合较差。在内部，问题则在于信息无法从表面传播进来，导致内部成为不确定激发的源头。同样，我们也发现，将更多关于观测特征的建模见解纳入其中——无论是识别相干波模式并据此约束内部动力学的不确定性，还是对观测相关的不确定性进行建模——虽能带来一些细微变化，却也会产生新的挑战；在目前这一初始阶段，这些方法似乎并未优于低信息量方法。

VI. 补充材料与数据可用性

我们在在线补充材料中附上了两个视频文件。第一个视频 auto ensmeans.mp4 展示了自主模型实验中 $u^b(t,x,y,z)$ 和 $u^a(t,x,y,z)$ 在 z/d =0、0.2、0.4、0.6、0.8、1.0 时的动力学，并着重呈现同化 $-u^a - u^b$ 与观测u^o造成的差异。第二个视频 auto b.mp4 展示了自主模型实验中 $u^b(t,x,y,z)$、$v^b(t,x,y,z)$ 和 $w^b(t,x,y,z)$ 在 z/d =0、0.2、0.4、0.6、0.8、1.0 时随时间变化的动力学。

我们还附上了一份简短的补充材料，其中包含全文多处提及的若干实验结果。

支持本研究结论的数据可应合理请求向通讯作者获取。

致谢

我们感谢 Alessio Gizzi 提供本研究所用数据。M.J.H. 和 E.M.C. 获得了美国国家科学基金会（NSF）资助，项目编号为 CMMI-2011280 和 CMMI-1762803。F.H.F. 获得了美国国家科学基金会（NSF）资助，项目编号为 CMMI-1762553。C.D.M.、F.H.F. 和 E.M.C. 获得了美国国立卫生研究院（NIH）资助，项目编号为 1R01HL143450-01。本研究部分使用了由佐治亚理工学院先进计算环境合作伙伴计划（PACE）提供的研究基础设施资源和服务，佐治亚州亚特兰大，美国[44]。本研究还使用了杜伦大学 Hamilton高性能计算服务。本报告中的所有图表均使用 Makie.jl[45] 软件包生成。

[1]G. Evensen, 《使用蒙特卡洛方法预测误差统计量的非线性准地转模型序贯数据同化》, 《地球物理研究杂志：海洋》**99**, 10143–10162（1994）。[2]L. Nerger、T. Janji´c、J. Schr¨oter 和 W. Hiller, 《集合平方根卡尔曼滤波器的统一》, 《月刊天气评论》**140**, 2335–2345（2012）。[3]L. Nerger、T. Janji´c、J. Schr¨oter 和 W. Hiller, 《一种用于基于集合的卡尔曼滤波器的受限局地化方案》, 《英国皇家气象学会季刊》**138**, 802–812（2012）。[4]J. T¨odter、P. Kirchgessner、L. Nerger 和 B. Ahrens, 《高维数据同化中非线性集合变换滤波器的评估》, 《月刊天气评论》**144**, 409–427（2016）。[5]P. Kirchgessner、J. T¨odter、B. Ahrens 和 L. Nerger, 《非线性集合变换滤波器的平滑器扩展》, 《Tellus A辑：动力气象学与海洋学》**69**, 1–15（2017）。[6]S. Vetra-Carvalho、P. J. van Leeuwen、L. Nerger、A. Barth、M. U. Altaf、P. Brasseur、P. Kirchgessner 和 J.-M. Beckers, 《用于高维非高斯问题的先进随机数据同化方法》, 《Tellus A辑：动力气象学与海洋学》**70**, 1–43（2018）。[7]L. Nerger, 《采用混合非线性卡尔曼集合变换滤波器对非线性系统进行数据同化》, 《英国皇家气象学会季刊》**148**, 620–640（2022）。[8] E. Cosme、J.-M. Brankart、J. Verron、P. Brasseur 和 M. Krysta, 《用于高分辨率海洋数据同化的降秩平方根平滑器的实现》, 《海洋建模》**33**, 87–100（2010）。[9]U. K. Dash、S.-Y. Park、C. H. Song、J. Yu、K. Yumimoto 和 I. Uno, 《三种数据同化方法在提高PM2·5可预测性方面的性能比较：集合卡尔曼滤波器、集合平方根滤波器和三维变分方法》, 《环境污染》**322**, 121099（2023）。[10]V. Rao、A. Sandu、M. Ng 和 E. D. Nino-Ruiz, 《采用$L 1$范数和Huber范数的稳健数据同化》, 《SIAM科学计算杂志》**39**, B548–B570（2017）。[11]H. Hersbach, 《集合预测系统中连续排序概率评分的分解》, 《天气与预报》**15**, 559–570（2000）。[12]J. Br¨ocker, 《使用连续排序概率评分评估原始集合》, 《英国皇家气象学会季刊》**138**, 1611–1617（2012）。[13]J. Thorey、V. Mallet 和 P. Baudin, 《利用连续排序概率评分进行集合预报的在线学习》, 《英国皇家气象学会季刊》**143**, 521–529（2017）。[14]M. Leutbecher 和 T. Haiden, 《理解连续排序概率评分的变化

：采用齐次高斯近似》，《英国皇家气象学会季刊》**147**，425–442（2021）。[15]D. J. Lea、I. Mirouze、M. J. Martin、R. R. King、A. Hines、D. Walters 和 M. Thurlow，《基于英国气象局耦合大气–陆地–海洋–海冰模型的新型耦合数据同化系统评估》，《月刊天气评论》**143**，4678–4694（2015）。[16]M. Goodliff、T. Bruening、F. Schwichtenberg、X. Li、A. Lindenthal、I. Lorkowski 和 L. Nerger，《将温度同化到近岸海洋生物地球化学模型中：弱耦合与强耦合数据同化的评估》，《海洋动力学》**69**，1217–1237（2019）。[17] S. G. Penny、E. Bach、K. Bhargava、C.-C. Chang、C. Da、L. Sun 和 T. Yoshida，《多尺度介质中的强耦合数据同化：准地转耦合模型实验》，《地球系统建模进展》**11**，1803–1829（2019）。[18]Q. Tang、L. Mu、H. F. Goessling、T. Semmler 和 L. Nerger，《将海洋观测强耦合同化到海洋–大气模型中》，《地球物理研究快报》**48**（2021），10.1029/2021GL094941。[19]N. S. LaVigne、N. Holt、M. J. Hoffman 和 E. M. Cherry，《模型误差对利用数据同化重建心脏电波状态的影响》，《混沌：非线性科学跨学科期刊》**27**，093911（2017）。[20]M. J. Hoffman、N. S. LaVigne、S. T. Scorse、F. H. Fenton 和 E. M. Cherry，《利用数据同化重建三维折返性心脏电波动力学》，《混沌：非线性科学跨学科期刊》**26**，013107（2016）。[21] M. J. Hoffman 和 E. M. Cherry，《用于重建三维心脏电动力学的数据同化系统的敏感性》，《英国皇家学会哲学汇刊A辑》**378**，20190388（2020）。[22]C. D. Marcotte、F. H. Fenton、M. J. Hoffman 和 E. M. Cherry，《含噪声的稳健数据同化：在心脏动力学中的应用》，《混沌：非线性科学跨学科期刊》**31**，013118（2021）。[23]A. Gizzi、E. Cherry、R. F. Gilmour Jr、S. Luther、S. Filippi 和 F. H. Fenton，《起搏部位和刺激历史对交替变化动力学及心脏组织中复杂时空模式形成的影响》，《生理学前沿》**4**，71（2013）。[24]A. Loppini、A. Gizzi、C. Cherubini、E. M. Cherry、F. H. Fenton 和 S. Filippi，《时空相关性揭示心脏组织中的特征长度》，《物理评论E》**100**，020201（2019）。[25]A. Barone、M. G. Carlino、A. Gizzi、S. Perotto 和 A. Veneziani，《心脏电导率的高效估计：一种适当的广义分解方法》，《计算物理学杂志》**423**，109810（2020）。[26]A. Barone、A. Gizzi、F. Fenton、S. Filippi 和 A. Veneziani，《用于估计空间依赖性心脏电导率的变分数据同化方法的实验验证》，《计算机方法在应用力学与工程中的应用》**358**，112615（2020）。[27]F. Fenton 和 A. Karma，《具有纤维旋转的三维连续心肌中的涡旋动力学：丝状体不稳定性与颤动》，《混沌：非线性科学跨学科期刊》**8**，20–47（1998）。[28]M. S. Spach、W. T. Miller、D. B. Geselowitz、R. C. Barr、J. M. Kootsey 和 E. A. Johnson，《传播的不连续特性

常规犬心肌中的传播。细胞内电阻反复出现不连续性并影响膜电流的证据。”《循环研究》**48**, 39–54 (1981)。[29]I. Kotadia, J. Whitaker, C. Roney, S. Niederer, M. O’Neill, M. Bishop 和 M. Wright，“各向异性心脏传导”，《心律失常与电生理学评论》**9**, 202–210 (2020)。[30]M. Perego 和 A. Veneziani，“求解电生理膜方程的 Rush–Larsen 方法的高效推广”，《数值分析电子汇刊》**35**, 234 256 2009– ( )。[31]E. M. Cherry 和 F. H. Fenton，“模拟和实验心脏组织中螺旋波与卷波的可视化”，《新物理学杂志》**10**, 125016 (2008)。[32]B. R. Hunt, E. J. Kostelich 和 I. Szunyogh，“时空混沌的高效数据同化：局地集合变换卡尔曼滤波器”，《物理学D：非线性现象》**230**, 112–126 (2007)。[33]R. E. Kalman *et al.*，“线性滤波与预测问题的一种新方法 [j],”《基础工程学杂志》**82**, 35–45 (1960)。[34]T. Gneiting 和 A. E. Raftery，“严格适当的评分规则、预测与估计”，《美国统计学会杂志》**102**, 359–378 (2007)。[35]M. Leutbecher，“集合规模：低于无穷大时有多不理想？”《英国皇家气象学会季刊》**145**, 107–128 (2019), https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3387。[36] I. Idris 和 V. Biktashev，“激发波启动中的临界前沿”，《物理评论E》**76**, 021906 (2007)。[37]R. D. Simitev 和 V. N. Biktashev，“心房组织渐近模型中激发的传播与阻滞条件”，《生物物理学杂志》**90**, 2258–2269 (2006)。[38]C. D. Marcotte 和 V. N. Biktashev，“慢快可激发模型中的临界点燃预测”，《物理评论E》**101**, 042201 (2020)。[39]在我们的实验中，这些数值可使波传播更快，但也会强烈驱使集合状态进入静息状态，导致集合坍缩（虽无实际意义），从而显著恶化整体重建效果。即使将刺激电流纳入模型动力学，重建误差仍明显更大（未显示）。因此，本研究采用在生理学上更合理的组织电导率数值。[40]I. Uzelac 和 F. H. Fenton，“用于光学标测信号定量分析且无需滤波的稳健框架”，载于 *2015 Computing in Cardiology Conference (CinC)* 2015 (IEEE,) pp. 461–464。[41]T. C. Sluka, S. G. Penny, E. Kalnay 和 T. Miyoshi，“利用强耦合集合数据同化将大气观测同化至海洋”，Geophysical Research Let- ters **43**, 752–759 (2016)。[42]R. Stenger, S. Herzog, I. Kottlarz, B. R¨uchardt, S. Luther, F. W¨org¨otter 和 U. Parlitz，“基于表面数据重建混沌三维时空可激发介质模型的深部活动”，《混沌：非线性科学跨学科期刊》**33** (2023)。[43]J. Lebert, M. Mittal 和 J. Christoph，“利用深度神经网络从二维观测重建可激发介质中的三维卷波”，《物理评论E》**107**, 014221 (2023)。[44]PACE Partnership for an Advanced Computing Environment *(PACE)* 2017 ( )。

[45]S. Danisch 和 J. Krumbiegel，“Makie.jl：灵活的高性能 Julia 数据可视化”，《开放源代码软件杂志》**6**, 3349 (2021)。

arXiv:2305.00009v3 [q-bio.QM] 5 Sep 2023

I. 补充材料

A. 重建动作电位

此处给出了正文中若干数值实验的重建时间轨迹以及对应的观测轨迹。对于每个实验，轨迹位置均从观测坐标中随机选取。

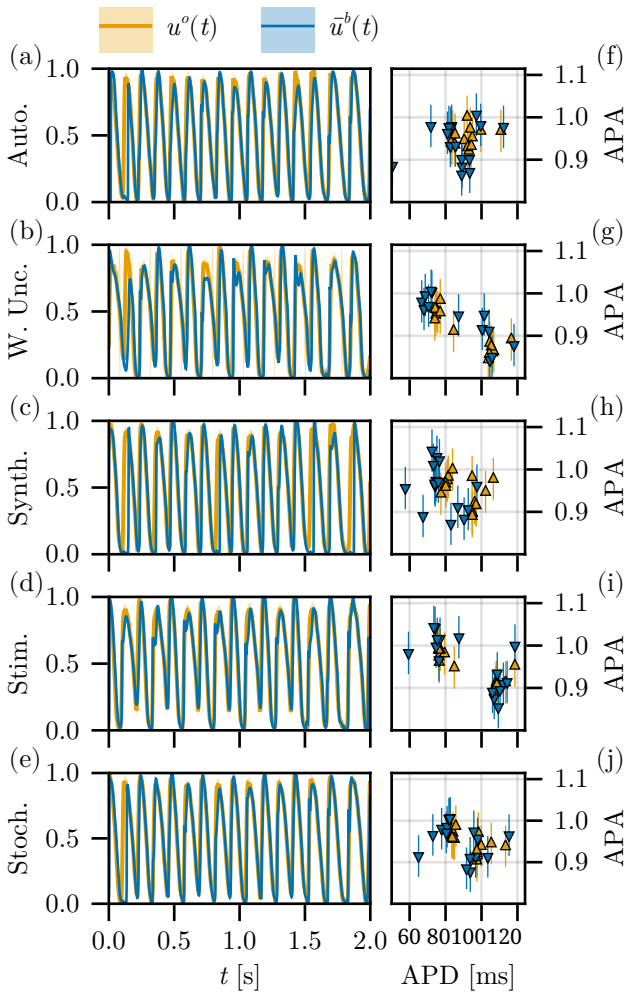

图 S1. 实验中随机选取位置处重建结果 $u^b(t)$ 与观测结果 $u^o(t)$ 的时间轨迹对（a–e），以及两种轨迹的动作电位时程和动作电位幅度分布（f–j）（$u_{\mathrm{thr}} = 0.2$）。

在所有实验中，经过少于两个周期长度（< 250 毫秒）的短暂过渡期后，重建都能可靠地恢复动作电位（AP）形状，参见图 S1(a-e)。动作电位形状的主要误差出现在峰部附近，重建结果的过冲和欠冲导致动作电位幅度（APA）的估计不佳。相比之下，由于动作电位时程（APD）的估计采用了一个低于峰部的阈值，因此观测结果与重建结果之间的估计值一致。用于生成图 S1(f-j) 的动作电位时程（APD）和动作电位幅度（APA）计算采用了一个阈值 $u_{\mathrm{thr}} = 0.2$ 以避免报告对于在重新激发前未达到静息状态的长波而言的动作电位时程（APD）+ 基本周期长度（BCL）。

B. 最小观测不确定性的估计

严格计算最小观测不确定性超出了本研究的范围。不过，我们根据数据集中延迟差值分布的半峰全宽估计了最小观测不确定性 η，其中延迟时间取同化间隔 $\tau = 2.0$ 毫秒。

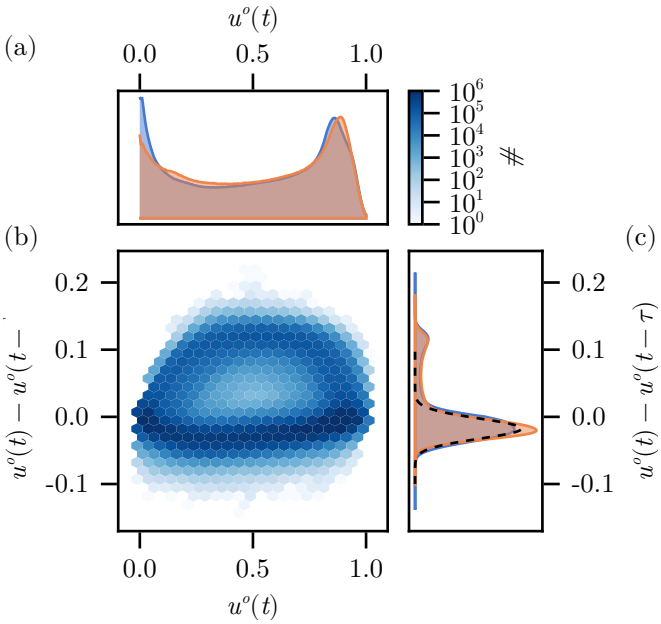

图 S2. (a) 数据集中场值的分布；(b) 数据集中数值与同化间隔对应的时间延迟差值之间的分布；(c) 数据集中与同化间隔相对应的时间延迟差值分布。观测的基线不确定性根据零附近出现的高斯分布峰值（黑色虚线曲线）的半峰全宽进行估计。

在图 S2(b) 中，我们展示了观测值与其同化间隔延迟差的分布，分辨率为 $(0.05, 0.05/3)$，分别对应 $(u^o(t)$ $u^o(t) - u^o(t-\tau))$ 坐标轴。从静息态 (0,0) 开始，动力学沿着表观环路顺时针演化。图 S2(a,c) 分别展示了 $u^o(t)$ 和 $u^o(t) - u^o(t-\tau)$ 的边缘分布。我们使用标准差为 $\sigma = \eta/(2\sqrt{2\ln 2})$ 的高斯分布，估算 (c) 中分布的半峰全宽；该高斯分布对应于 0 附近的峰值，其 $\eta \approx 0.05$ 如 $u^o(t) - u^o(t-\tau) \in [-0.1, 0.1]$ 所示，为黑色虚线。该数值同样使得在 $0 \le u^o \le 2\eta$ 之间识别静息观测更加可靠。

C. 观测算子局部平均

在本文主体中，观测算子$H(\mathbf{x}^b) = \mathbf{y}^b$ 是一个大型稀疏矩形矩阵的作用，用于将 N 维状态向量降至 l 维状态观测值 $l \ll N$。该算子严格地在特定位置对 u 场进行采样。与此同时，实验获得的观测数据经过了两次滤波：首先是由组织下方荧光的传输造成的滤波，其次是通过时空核进行后处理，以提高记录的信噪比。这会通过减弱离群信号的影响来平滑数据集，但也会使数据丧失系统中真正尖锐特征的分辨率。

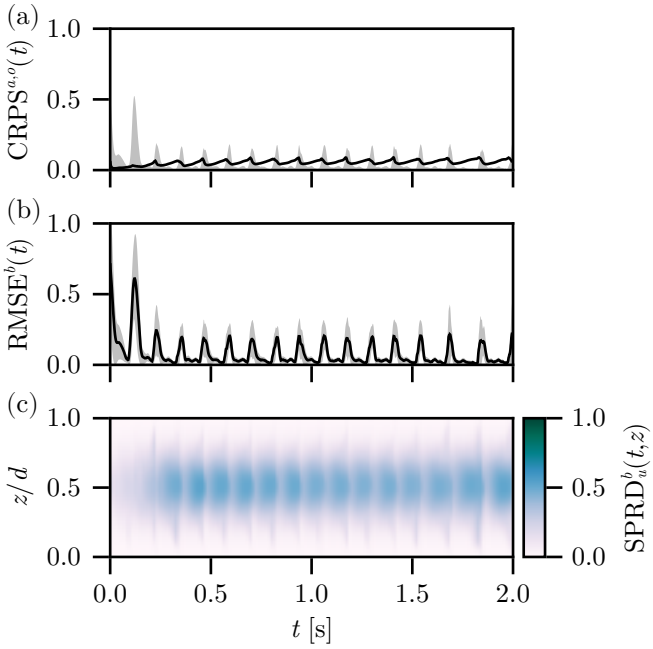

图 S3. 采用空间卷积观测算子的自主模型重建结果，报告 (a) CRPS$^{a,\ o}(t)$、(b) RMSE$^b(t)$，以及 (c) 沿深度方向的离散度 SPRD$^b_u(t,\ z)$。

为此，我们扩展了观测算子：不再由 $H(\mathbf{x}^b)$ 返回场在特定索引（位置） u^b_i处的值，而是计算并返回以该索引（位置）为中心、覆盖 $p \times p$ patch 的局部平均值。我们选择 $p = 5$，使得观测（在默认的表面分布中）总体上能够覆盖整个表面，但每个观测的平均范围都不包含其他观测位置。这等价于使用二维方框滤波器进行空间卷积。图 S3 展示了该滤波观测算子的结果，对应于自主重建问题（参见正文 III.A.2 节）。总体而言，结果与规范采样观测算子略有不同，均方根误差略有增加$^b(t)$，而集合离散度略有降低总体而言，在集合离散度中。

D. 刺激周期鲁棒性

除了我们的方法对模型参数的稳健性之外，我们还将重建应用于具有更长周期长度的其他数据集，这些数据集产生的动力学明显更慢，交替现象也不那么显著。然而，尽管动力学相对平缓，同化对错误动作电位形状的敏感性却显著增加。图 S4 展示了基本周期长度为 $= 234 \pm 2$ ms 的自主模型重建结果。该重建的连续排序概率得分$^a(t)$和最大均方根误差$^b(t)$均显著增大。

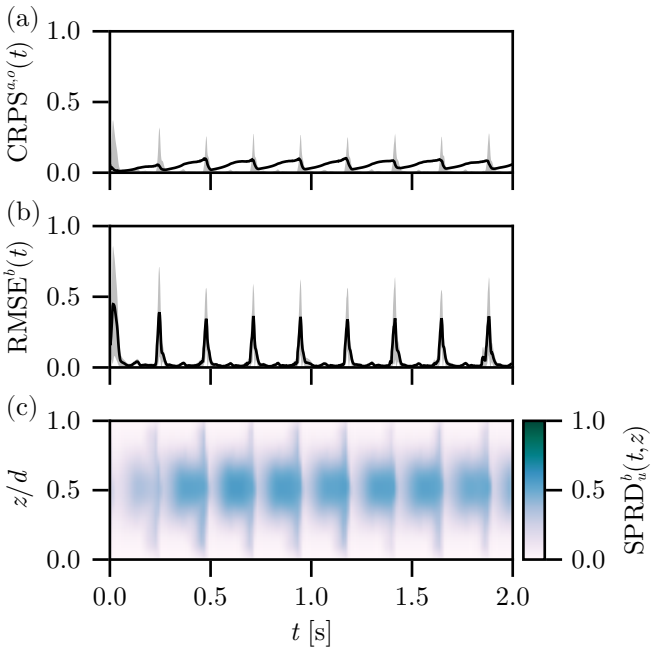

图 S4. 基本周期长度为 $=234 \pm 2$ ms 的数据集的重建结果，报告 (a) 连续排序概率得分$^{a,\ o}(t)$、(b) 均方根误差$^b(t)$，以及 (c) 沿深度方向的离散度标准化预测半径离散度$^b_u(t,\ z)$。

虽然均方根误差$^b(t)$中的偏差幅度较大，但持续时间很短（参见图 S4(b)），而且对于该数据集，动作电位形状得到了可靠的恢复（参见图 S5(a)）。由于模型能够在激励之间完全恢复，模型、观测与局部集合变换卡尔曼滤波之间的非线性反馈机制便不再起作用。正如较长基本周期长度下较简单的动力学所预期的那样，我们的方法无论在细节上还是总体上都表现出色。值得注意的是，图 S5(a) 中重建的动作电位形状往往呈现出明显的尖峰-穹顶结构。这种结构通常并非该模型所呈现的特征，却见于许多心脏兴奋离子模型。因此，重建得到的动作电位图示必须谨慎看待，因为这种结构的出现

很可能是模型过度激发后，局部集合变换卡尔曼滤波对集合状态进行校正所致，而非重建过程复现的数据所具有的统计特征。

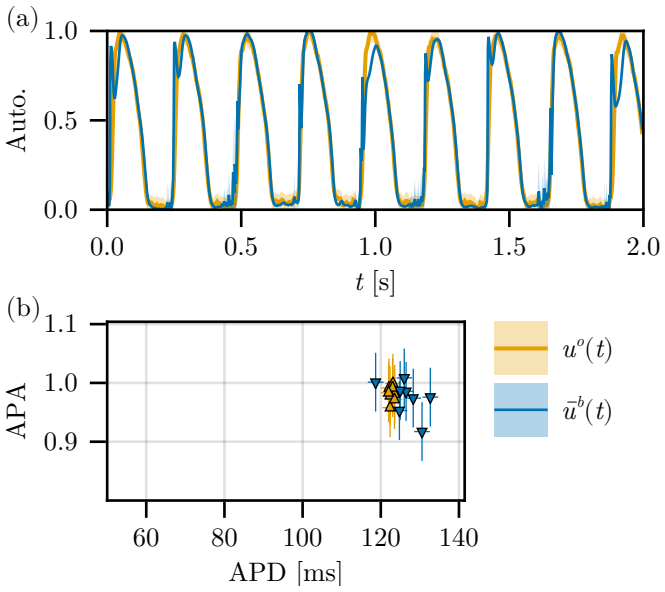

图 S5。（a）BCL 为 =234±2 ms 时的观测时间轨迹，以及使用自治模型进行的重建；（b）该序列中动作电位的动作电位时程和幅度分布（$u_{\mathrm{thr}} = 0.2$）。

这条轨迹的动作电位时程和幅度分布匹配良好，参见图 S5（b）；前者符合较短基本周期长度下的结果，而后者似乎是因为激发之间的周期较长，减弱了动作电位幅度交替的影响，因此更好地拟合了 Fenton-Karma 模型所能实现的动作电位形态。也就是说，尽管动力学完全处于 1:1 区域内，重建仍能更好地校正动作电位形状，因为整个集合必然具有相同的周期性和相位；而在交替现象强烈的区域，部分集合成员的节拍可能与其他成员不同，从而限制了校正动作电位形状时可用的自由度。

E. 模型参数敏感性

1. 电导率

在参考文献 1 中，模型参数是通过人工调优确定的，而组织的电导率则通过定制的数据同化方法进行估计。除估计电导率外，本文在重建问题中沿用这些模型参数；估计电导率的数值则不符合生理实际，因为它们用于补偿激发在组织中极快的传播速度。本文采用的数值为 $D_{\parallel} = \sigma_{\parallel}/9$和 $D_{\perp} = \sigma_{\perp}/15$，其中 $\sigma_{\parallel,\ \perp}$ 是文中报告的分别对应沿纤维方向和垂直于纤维方向的电导率估计值。

为考察稳健性，我们在重建问题中测试了不符合生理实际的高电导率（同时缩短了模型积分时间步长 $\delta t = 2 \times 10^{-4}$）。如正文所述，我们发现重建效果明显差于自由运行实验：由于没有周期性刺激电流，集合会回到静息状态，单靠 LETKF 驱动无法再次激发。

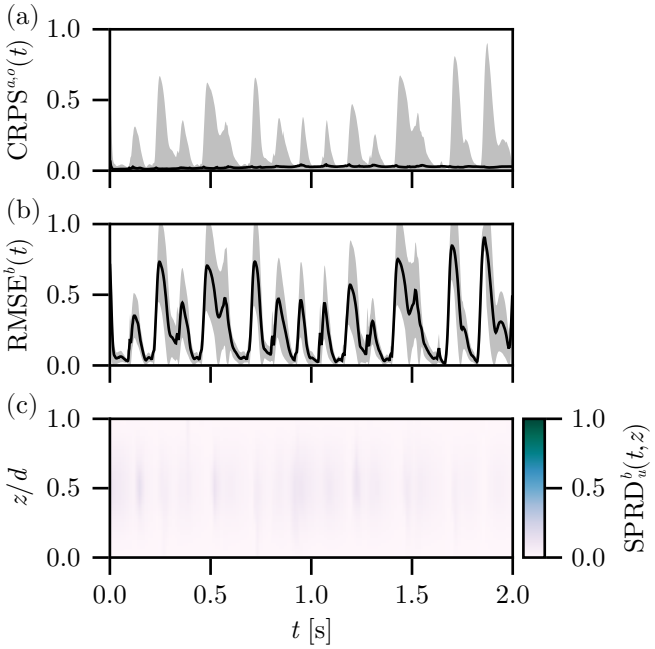

图 S6. 对受刺激的 Barone 模型进行重建，采用来自参考文献 1 的 $D_{\parallel,\ \perp} = \sigma_{\parallel,\perp}$ = (0.009,0.003) cm^2/ms 电导率，并展示 (a) 连续排序概率得分$^{a,\ o}(t)$、(b) 均方根误差$^b(t)$，以及 (c) 沿深度的离散度标准化预测半径离散度$^b_u(t,z)$。

结合刺激电流，我们能够匹配观测中激励的基线频率，并持续激活集合状态。图 S6 展示了采用不符合生理实际的超大电导率时，对受刺激的“快速”模型进行重建的结果。显然，这种重建在各项指标上的表现都更差——事实上，即使电导率如此之高，集合仍无法匹配观测中的波速，因为它持续与不应期组织相互作用。

2. 可激发模型参数

最后，我们转而讨论可激发模型参数本身。由于 Fenton-Karma 模型具有足够的灵活性，能够拟合大量可能的恢复曲线，因此我们使用了三个不同的参数集来比较重建对可激发模型参数的敏感性，汇总于表 I。Beeler-Reuter（BR）

且 Guinea Pig（GP）参数集改编自参考文献 2。

参数	BR	GP	Barone	单位
τ_{si}	30.0	22.0	127	[ms]
τ_{v1}^{-}	1250	333	38.2	[ms]
τ_{v2}^{-}	19.6	40.0	38.2	[ms]
τ_{v}^{+}	3.33	10.0	1.62	[ms]
τ_{w}^{-}	41.0	65.0	80	[ms]
τ_{w}^{+}	870	1000	1020	[毫秒]
τ_{d}	0.25	0.12	0.1724	[毫秒]
τ_{o}	12.5	12.5	12.5	[毫秒]
τ_{r}	33.0	25.0	130	[ms]
k	10.0	10.0	10	
u_c	0.13	0.13	0.13	
u_v	0.04	0.025	*	
u_c^{si}	0.85	0.85	0.85	

表 I. 用于重建实验数据的动力学模型中 Fenton-Karma 模型的参数值。

*注：由于 $\tau_{v1}^{-} \equiv \tau_{v2}^{-}$，参考文献 1 中未指定切换参数 u_v。

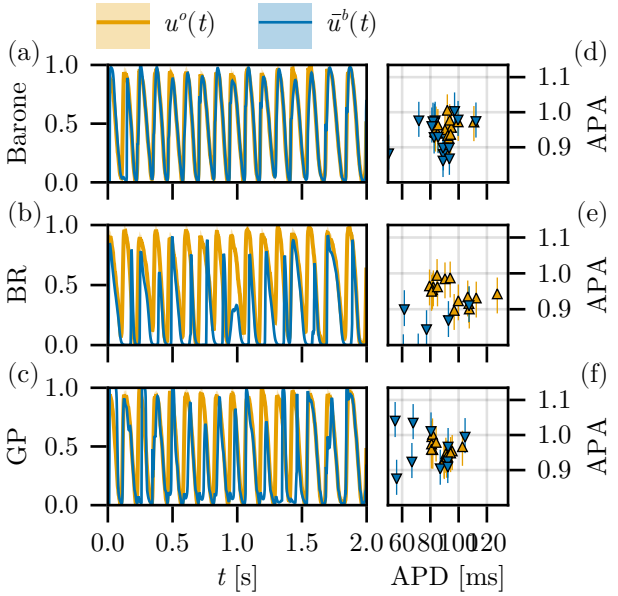

图 S7. Barone *et al.*、BR 和 GP 参数集在随机选取位置处的 (a-c) 重建时间轨迹 $u^b(t)$ 与观测 $u^o(t)$，以及 (d-f) 相应的动作电位时程和动作电位幅度分布 ($u_{\text{thr}} = 0.2$)。

在图 S7(a) 中，我们复现了使用 Barone 等人参数集时，在观测覆盖范围内随机选取位置的观测与重建结果的时间轨迹，并在图 S7(d) 中展示了两种信号的动作电位时程和动作电位幅度分布。BR 和 GP 模型参数主要影响动作电位形状，导致图 S8(b,c) 中较大的均方根误差$^b(t)$。使用 BR 和 GP 参数集进行重建时，动作电位形状无法可靠恢复，分别参见图 S7(b,c)。因此，BR 和 GP 模型在近似观测的动作电位时程和动作电位幅度分布方面表现较差，分别参见图 S5(e,f)。这表明，尽管每次新激发期间都存在局部表面重建误差，Barone 等 参数集仍能充分模拟潜在的真实动力学。

我们针对每个新的参数集计算了自主数据同化重建，以了解可激发模型参数对重建的影响，并在图 S8 中总结了这些结果：其中 (a) 重现了正文中始终采用 Barone等 参数集得到的基线自主重建结果，并考察了采用 (b) Beeler-Reuter 和 (c) Guinea Pig 参数集进行的重建；报告了连续排序概率得分^a, $^o(t)$、均方根误差$^b(t)$，以及沿深度的离散度 SPRD$_u^b(t,\ z)$。对于 (b) Beeler-Reuter（BR）和 (c) Guinea-Pig（GP）参数集，均方根误差$^b(t)$不仅显著增大（表现更差），可激性参数 τ_d 的变化还会导致动作电位形状和传导速度发生显著差异；这也改变了系综离散度沿深度的稳态分布。例如，在同化间隔内，BR 参数集产生的较慢波向内部传播得较少，导致不准确的确定性更高；而 GP 参数集产生的相对较快的波则将信息从表面传播到内部。

特别值得注意的是，BR 和 GP 参数集的 CRPS$^o(t)$ 发生了变化，参见图 S8(b,c)。在这两种情况下，与 Barone等参数集相比，背景集合与观测序列存在显著不一致；这也再次印证了我们此前从动作电位轨迹得出的结论：本研究使用的参数集能够充分刻画该数据集和模型的真实动力学。

F. 偏离自主模型重建

在图 S9 中，我们比较了随机模型和刺激模型的重建误差与重建离散度的偏差，并以自主实验的结果为基线。随机模型和刺激模型的重建结果与自主实验的结果都只有轻微偏差，表现为随时间振荡的改善（对于 ΔRMSE$^b(t) > 0$）或恶化（对于ΔRMSE$^b(t) < 0$）。重建误差对应于 $\Delta\text{RMSE}^b(t) = \text{RMSE}_*^b(t) - \text{RMSE}_{\text{Auto.}}^b(t)$，离散度也类似：$\Delta\text{SPRD}_u^b(t) = \text{SPRD}_{\text{u,*}}^b(t) - \text{SPRD}_{\text{Auto.,u}}^b(t)$，其中 $* \in \{\text{Stim.,Stoch.}\}$。此外，我们还可以比较背景集合平均与观测之间的误差，并将这一量表示为向量函数 $e_*(t) = u^b - u^o \in \mathbb{R}^l$，其中 l为观测维度。对于刺激模型和随机模型，误差在时间和空间上的重建误差变化期望值小于重建误差的变化幅度，

$$\text{mean}(|e_*(t) - e_{\text{Auto.}}(t)|) \ll \text{std}(|e_*(t)|),$$

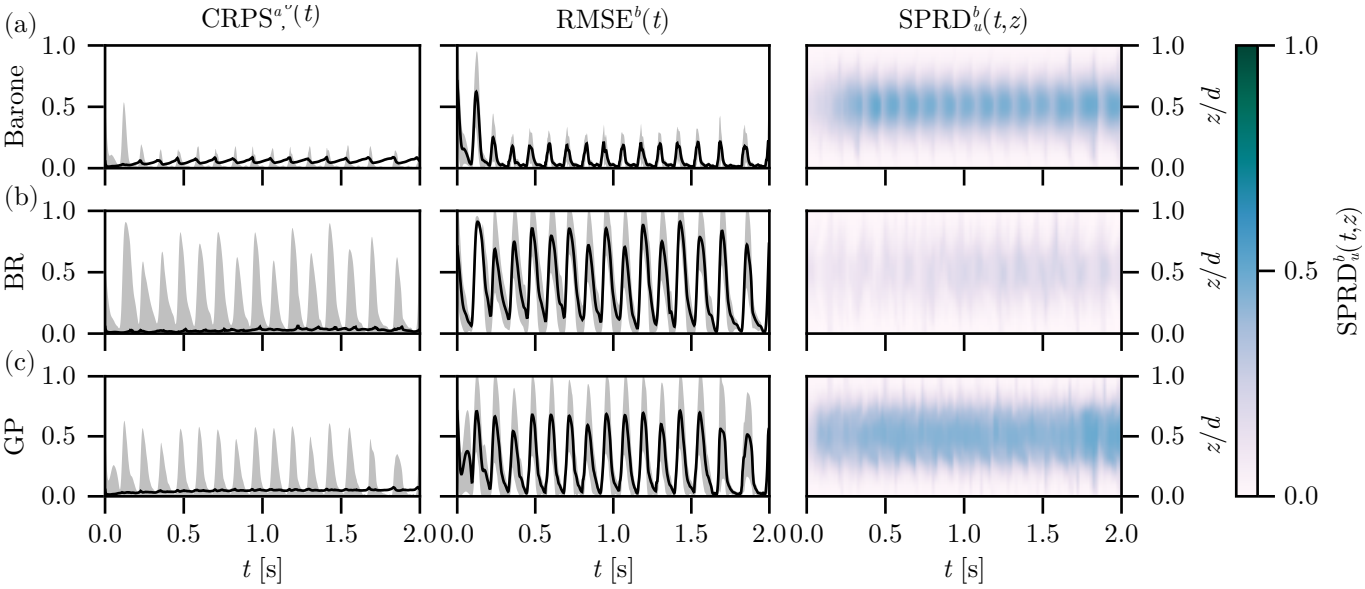

图 S8. 使用 (a) Barone 等人、(b) Beeler-Reuter 和 (c) Guinea Pig 参数集的自治模型重建结果。

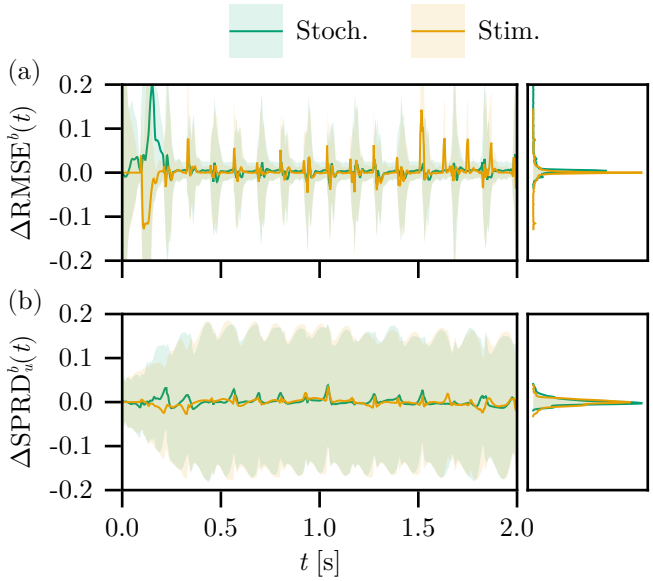

图 S9. 随机实验和刺激模型重建相对于自治模型重建的误差差异；展示 (a) Δ均方根误差$^{b}(t)$ 和 (b) Δ标准化预测半径离散度$^{b}_{u}(t)$，以及它们的边缘分布。

其中 mean 和 std 分别为标准均值函数和标准差函数，计算时覆盖全部 1000 个同化间隔以及 l 中所有 $e_{*}(t)$ 索引，这些索引用于枚举观测。在这些尺度下，图 S9(b) 清楚地显示，随机模型重建的集合离散度有所增加；其中 $\Delta\mathrm{SPRD}^{b}_{\mathrm{Stoch.},u}(t)$ 的峰值达到与刺激模型相近的幅度，但每个峰值在较长时间间隔内都明显非零。

[1]A. Barone、A. Gizzi、F. Fenton、S. Filippi 和 A. Veneziani，“用于估计空间依赖性心脏电导率的变分数据同化方法的实验验证”，《计算机辅助应用力学与工程方法》**358**，112615 (2020)。[2]F. Fenton 和 A. Karma，“纤维旋转的三维连续心肌中的涡旋动力学：丝状体不稳定性与颤动”，《混沌：非线性科学跨学科期刊》**8**，20–47 (1998)。