arXiv:2303.06951v1 [q-bio.SC] 13 Mar 2023

用于合成细胞内生物分子时空动力学的生成模型

Lisa Balsollier[1,2,3],Fr´ed´eric Lavancier[1], Jean Salamero[2,3], 和 Charles Kervrann[2,3]

[1]LMJL，法国国家科学研究中心（CNRS）UMR 6629、南特大学，F-44322 南特，法国[2]法国国家科学研究中心（CNRS）雷恩大学 Inria 中心 SERPICO 项目团队，F-35042 雷恩 Cedex，法国[3]居里研究所、法国国家科学研究中心（CNRS）UMR 144、巴黎文理研究大学、索邦大学，F-75005 巴黎，法国

摘要

生物成像中的时空动力学生成器已成为构建真实值数据集的重要工具，可用于评估图像处理算法（如生物分子检测器和跟踪器），也可用于生成深度学习算法的训练数据集。本文采用一种称为生-死-移动（BDM）点过程的随机模型，生成细胞内生物分子的联合动力学。这种方法非常灵活，能够对运动中的粒子系统进行建模；粒子之间可能相互作用，并且每个粒子都可能从一种运动状态（如布朗运动）切换到另一种状态（如定向运动）。此外，新的轨迹会随时间出现，并在经过一定寿命后消亡；这些特征均可能取决于所有现存粒子的当前空间构型。我们将介绍如何设定 BDM模型的各项特征，并提供许多与生物成像应用相关的实例。最后，我们基于真实的荧光显微镜数据集，对模型进行校准，以模拟质膜附近 Langerin 和 Rab11 蛋白的联合动力学。我们展示了生成的合成序列具有与真实显微镜图像序列中观察到的特征相似的特征。

关键词： 模拟与图像合成、细胞内动力学与分子运动、荧光显微镜、出生-死亡-移动过程、空间统计。

1 引言

基础生物学的长期目标之一，是在单细胞水平上揭示分子复合物内相互作用分子的时空动力学协调与组织方式。这包括对细胞内动力学的表征，这对于更好地理解膜运输等基本机制至关重要。为此，人们开发了专门的图像分析方法，用于处理通过荧光显微镜采集的、具有挑战性的二维和三维图像时间序列，例如 [13]。

在此背景下，数学模型和生物物理模型对于解读生物分子的运输流至关重要。它们构成了大多数粒子追踪程序中的关键先验运动模型 [8; 24], ，而且是开展模拟所必需的，例如可用于评估图像分析算法的性能，以及训练监督式深度神经网络等复杂模型 [16; 1]。近年来，人们提出了针对单个粒子轨迹的随机模型，可描述细胞内速度可变的布朗运动、受限运动或定向运动 [4; 5; 10; 6; 23; 7; 16], ，并且还可能模拟粒子沿细胞骨架网络运动的情形 [14; 12]。然而，这些建模方法难以体现粒子的集体运动和整体生物分子运输。理想情况下，后者应纳入生物分子之间可能发生的相互作用，尤其是不同类型蛋白质之间的相互作用，从而产生多种应用中观察到的共定位现象 [2; 9; 15; 19]。此外，它还应能够生成每条轨迹的寿命、空间分布，以及其在模拟图像序列中的起始时间。

在本文中，我们提出利用[18], 中介绍的一种定制随机模型，该模型能够生成上述特征。这种所谓的出生-死亡-移动（BDM）空间点过程，是一种灵活的模型，用于描述粒子系统（此处为一组生物分子）的动力学：粒子随时间移动，同时细胞中可能出现一些新粒子，也可能有一些现存粒子消失。我们进一步允许为每个粒子赋予一个给定的标签，例如不同种类的带标签蛋白质和/或不同的运动状态；这些标记还可能随时间改变，例如粒子可能从一种状态（如布朗运动）切换到另一种状态（如亚扩散）。文献中有时将这种标记切换称为“突变”，但为避免与真正的生物突变混淆，我们在此更倾向于使用“转化”一词。此外，轨迹可以由任意连续马尔可夫扩散模型驱动，这涵盖了文献中此前考虑过的所有单条轨迹模型；还可以引入一些相互作用，从而生成共定位现象。出生强度决定了等待下一个新粒子出现的时间，它可能取决于粒子的当前构型；死亡强度也同样如此。例如，我们可以设定细胞中的生物分子越多，死亡强度越高，从而导致粒子迅速消失。还可以考虑一些空间效应

，以便在细胞的某些区域形成不同的运动状态，或促使新粒子在某些空间区域出现，例如在共定位作用下出现在现存粒子附近。

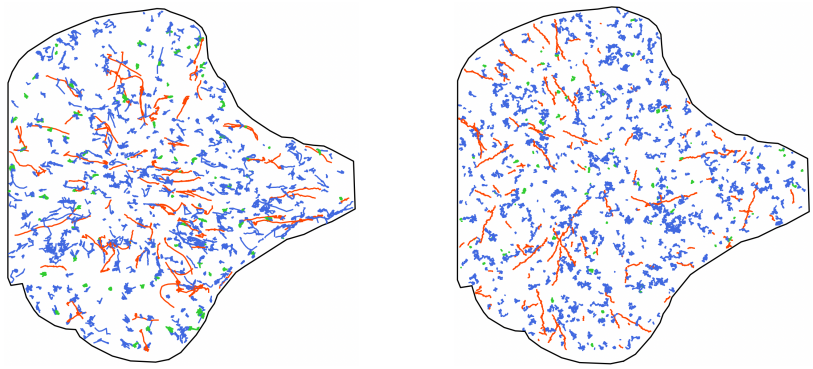


图 1：左：在 Langerin 蛋白的真实图像序列中检测并跟踪的所有轨迹，按估计的运动状态着色（布朗运动为蓝色，超扩散为红色，亚扩散为绿色）。右：由我们的随机模型生成的合成序列结果。

本文其余部分安排如下。在第2节中，我们给出随机过程的精确定义，并特别列出完整定义模型所需的所有要素。第2.2节介绍一种迭代构造，阐明动力学的演进方式；附录中则对一种高效的模拟算法作出正式说明，并在线提供该算法。第2.3节中，我们给出模型设定的诸多示例，我们认为这些设定与许多真实生物分子的动力学相关。第3节中，我们重点考察参与膜运输的 Langerin 和 Rab11 蛋白的联合动力学，以展示生成器的潜力。我们的方法首先检查真实数据集，以便审慎地选择并校准 BDM 模型的各项参数。该数据集由一组通过三维多角度全内反射荧光（total internal reflection fluorescence）显微镜技术采集的图像构成 [3], ，描绘了细胞质膜附近 Langerin 和 Rab11 蛋白的位置，具体而言，范围为 z轴方向上的 $1\ \mu$m。经过一些后处理后，这组图像呈现出两类蛋白质各自的一组轨迹；这些轨迹遵循不同的运动状态，在细胞内按特定方式进行空间分布，并在序列的不同时段出现，这与 BDM 过程的动力学完全吻合。图1最左侧的图展示了 Langerin 通道中观测到的轨迹。我们根据该数据集计算出一组描述符，用于校准随机过程的参数，同时也用于构建评估合成序列的基准特征。最后，在第3.2节中，我们生成若干

模拟序列，并展示它们具有与真实序列中观测到的特征相当的特征。图1最右侧的图展示了一个生成轨迹的示例。

补充材料包括用于模拟的 Python 代码、原始数据以及一些额外的模拟序列，可在我们的在线 GitHub代码库 `https://github.com/balsollier-lisa/BDM-generator-for-bioimaging` 中获取。

2 数学模型

2.1 启发式描述与符号

为了模拟生物分子的动力学，我们考虑一种带有突变的多类型出生-死亡-移动过程，记作 $(\mathbf{X}_t)_{t\geq 0}$。该过程是对出生-死亡-移动过程的推广，相关过程最初见于 [18]。以下为避免误解，我们采用“转化”一词，而非“突变”，具体解释见第 1 节。本节描述 $(\mathbf{X}_t)_{t\geq 0}$ 的时空动力学，并介绍一些符号。

在每个时刻 t，$\mathbf{X}_t$ 是位于 $\mathbb{R}^d$ 的有界集合 Λ 中的一组粒子。每个粒子都被赋予一个标记，用来表示某种特征。我们用 $\mathbb{M}$ 表示所有可能标记的集合。随着时间推移，粒子会移动（移动方式可能取决于其关联的标记以及粒子之间的相互作用），并可能发生三种突变，我们称之为“跳跃”：

1. 一次“出生”：一个新粒子可能出现，并被赋予一个标记；
2. 一次“死亡”：一个现有粒子可能消失；
3. 一次“转化”：一个现有粒子的标记可能发生变化。

示例：在第 3 节讨论的生物学应用中，Λ表示维度为 $d = 2$ 或 $d = 3$ 的细胞。我们观察到这个细胞内有两类粒子，分别与 Langerin 和 Rab11 蛋白相关联；每个粒子都可能按照三种不同的状态运动：布朗运动、超扩散或亚扩散。因此，在这个例子中，每个粒子都从 6 种可能性中进行标记：它与 Langerin (L) 或 Rab11 (R) 相关联，且取决于其运动状态（1 至 3），从而得到 $\mathbb{M} = \{(L,\ 1),\ (L,\ 2),\ (L,\ 3),\ (R,\ 1),\ (R,\ 2),\ (R,\ 3)\}$。随着时间推移，每个粒子都根据自身的运动状态独立于其他粒子运动；最终可能出现一个新粒子，已有粒子可能消失，而某些粒子的运动状态也可能改变。

我们用 $n(\mathbf{X}_t)$ 表示时刻 t 的粒子数。每个粒子 $x_i \in \mathbf{X}_t$，其中 $i = 1, \ldots\ldots, n(\mathbf{X}_t)$，可分解为 $x_i = (z_i, m_i)$，其中 $z_i \in \Lambda$ 表示其位置，而 $m_i \in \mathbb{M}$ 表示其标记。因此，我们有 $\mathbf{X}_t \in (\Lambda \times \mathbb{M})^{n(\mathbf{X}_t)}$。（严格来说，$\mathbf{X}_t$ 中粒子的排列顺序并不重要，因为对粒子进行任何排列都会得到相同的粒子集合。本文为避免这一细节，采用相同的记法，仿佛 $\mathbf{X}_t$ 是一个粒子向量，尽管它实际上是一个粒子集合。）由于粒子数随时间变化，随机过程 $(\mathbf{X}_t)_{t\geq 0}$ 取值于空间

$$E = \bigcup_{n\in\mathbb{N}} (\Lambda \times \mathbb{M})^n.$$

为强调 $\mathbf{X}_t$ 并非一个简单的值，而是编码了粒子系统的位置和标记这一事实，我们称该系统在时刻 t 处于构型 $\mathbf{X}_t$。

要完整描述 $(\mathbf{X}_t)_{t\geq 0}$ 的动力学，我们需要以下要素 ：

1. 一个方程组 $(Move)$，用于规定 $\mathbf{X}_t$ 中的每个粒子在两次跳跃之间连续运动的方式。我们通常考虑一组作用于每个粒子位置的随机微分方程，这些方程可以依赖于与粒子关联的标记，也可以与其他粒子相互作用；

2. 三个从 E 到 $[0, \infty)$ 的连续有界函数 β、δ 和 τ，分别称为出生、死亡和转化强度函数，用于控制新粒子出生、死亡和转化之前的等待时间。在每个时刻 t，我们可以将 $\beta(\mathbf{X}_t)dt$ 解释为：给定粒子系统处于构型 $\mathbf{X}_t$，在区间 $[t, t+dt]$ 内发生出生的概率；对于 δ 和 τ，也可以作类似解释。

3. T表示每次跳变如何发生的三个转移概率函数：

- $p^{\beta}((z, m)|\mathbf{x})$：在已知发生出生，且出生时粒子系统处于构型 $\mathbf{x}$ 的情况下，出生发生在位置 z 且标记为 m 的概率密度函数；

- $p^{\delta}(x_i|\mathbf{x})$，对于 $x_i \in \mathbf{x}$：在已知发生死亡，且死亡时粒子系统处于构型 $\mathbf{x}$ 的情况下，死亡涉及 $\mathbf{x}$ 中的粒子 x_i 的概率；

- $p^{\tau}((z_i, m_i)\ m|\mathbf{x})$，对于 $(z_i, m_i) \in \mathbf{x}$：粒子 $x_i = (z_i, m_i)$ 在 $\mathbf{x}$ 中改变其标记的概率，以及该转化导致其获得新标记 $m \neq m_i$ 的概率，前提是发生了转化，且粒子系统在转化时刻处于构型 $\mathbf{x}$。

我们在第2.3节中提供了一些示例，说明如何选择这些特征。最后，我们将用 T_1，T_2、.....表示该过程的跳跃时刻，并约定 $T_0 = 0$。

2.2 算法构造

假设已知上一节介绍的过程（$\mathbf{X}_t$）$_{t\geq 0}$ 的特征，即方程组（$Move$）、强度函数 β, δ, τ以及转移概率函数 $p^\beta, p^\delta, p^\tau$。那么，从时间 $T_0 = 0$时的初始构型 $\mathbf{X}_0$出发，我们按如下方式在时间区间 $[0, T]$ 内迭代构造该过程。这里令 $\alpha = \beta + \delta + \tau$ 表示总跳跃强度。

1. 给定初始条件 $\mathbf{X}_0$，在区间 $[0, T - T_0]$内生成 n条（$\mathbf{X}_0$）连续轨迹，作为方程组（$Move$）的解。将这些轨迹记为（$\mathbf{Y}_t$）$_{t\in[0,T-T_0]}$ 。

2. 通过抛硬币，检验跳跃时间 T_1 是发生在 T 之后（其概率为 p），还是发生在 T之前（其概率为 $1 - p$），其中

$$p = \exp\left(-\int_0^{T-T_0} \alpha(\mathbf{Y}_u)du\right).$$

- 若 $T_1 > T$，则对于所有 $t \in [T_0, T]$，都有 $\mathbf{X}_t = \mathbf{Y}_{t-T_0}$，至此模拟完成。
- 否则，我们继续生成 T_1，在 $[T_0, T]$ 中生成，并按如下方式生成相应的跳跃。

3. 根据概率分布，在满足 $T_1 < T$ 的条件下生成 $T_1 - T_0$

$$\mathbb{P}(T_1 - T_0 < t | T_1 < T) = \frac{1}{1-p}\left(1 - \exp\left(-\int_0^t \alpha(\mathbf{Y}_u)du\right)\right), \quad 0 < t < T - T_0.$$

截至时间 T_1 的过程由生成的轨迹给出，即

$$(\mathbf{X}_t)_{T_0\leq t<T_1} = (\mathbf{Y}_{t-T_0})_{T_0\leq t<T_1}.$$

4. 抽取在 T_1 发生哪种跳跃（我们用 $\mathbf{X}^{T_1^-}$ 表示跳跃发生前一刻过程的构型；根据 $(\mathbf{Y}_t)$ 的连续性，该构型为 $\mathbf{Y}^{T_1-T_0}$ ）：

- 这是一次出生，其概率为 $\beta(\mathbf{X}^{T_1^-})/\alpha(\mathbf{X}^{T_1^-})$;
- 这是一次死亡，其概率为 $\delta(\mathbf{X}^{T_1^-})/\alpha(\mathbf{X}^{T_1^-})$;
- 这是一次概率为 $\tau(\mathbf{X}^{T_1^-})/\alpha(\mathbf{X}^{T_1^-})$ 的转化。

5. 按如下方式生成在 $t = T_1$ 处的跳跃，以得到 $\mathbf{X}_{T_1}$：

- 如果这是一次出生，则根据概率密度函数 $p^{\beta}((z,\ m)|\mathbf{X}_{T_1^-})$ 生成新粒子 $x = (z,\ m)$。然后设置 $\mathbf{X}_{T_1} = \mathbf{X}_{T_1^-} \cup (z,\ m)$；
- 如果这是一次死亡，则根据概率 $p^{\delta}(x_i|\mathbf{X}_{T_1^-})$ 抽取要删除的粒子 $x_i \in \mathbf{X}_{T_1^-}$，其中 $i = 1,\ \ldots\ldots,\ n(\mathbf{X}_{T_1^-})$。然后设置 $\mathbf{X}_{T_1} = \mathbf{X}_{T_1^-} \setminus x_i$；
- 如果这是一次转化，则抽取要发生转化的粒子 $x_i = (z_i,\ m_i) \in \mathbf{X}_{T_1^-}$，并根据 $p^{\tau}((z_i,\ m_i)\ m|\mathbf{X}_{T_1^-})$ 生成其转化，其中$i = 1,\ \ldots\ldots,\ n(\mathbf{X}_{T_1^-})$。然后设置 $\mathbf{X}_{T_1} = (\mathbf{X}_{T_1^-} \setminus (z_i,\ m_i)) \cup (z_i,\ m)$。

6. 返回步骤 1，依次使用 $T_0 \leftarrow T_1$ $\mathbf{X}_0 \leftarrow \mathbf{X}_{T_1}$，以生成从 $\mathbf{X}_{T_1}$ 开始的新轨迹以及下一次跳跃时间 T_2，依此类推。

在上述构造的第一步中，轨迹会生成至终止时间 T。然而，下一次跳跃很可能会在远早于 T 的时刻发生，因此只需在较短的时间区间内生成这些轨迹，这样计算效率也更高。附录中给出了一个模拟$\mathbf{X}_t$（针对 $t \in [0, T]$）的正式算法，该算法遵循上述构造并纳入了这一思路。此算法已使用 Python 实现，可在我们的 GitHub代码库中获取。

从理论角度来看，请注意，步骤 3 中跳跃间等待时间概率分布的特定指数形式，是得出上一节所阐释的 β、δ 和 τ 的解释所必需的。这种指数形式还意味着，$(\mathbf{X}_t)_{t\geq 0}$ 是一个马尔可夫过程，也就是说，其未来动力学仅取决于其当前构型。关于这些理论方面的更多细节，请参见 [18] 。

2.3 模型的示例性规范

2.3.1 跳跃间运动

回想一下，在跳跃间隔内，过程（$\mathbf{X}_t$）具有恒定的基数 $n = n$（$\mathbf{X}_t$），且其所有粒子的标记保持不变。我们用（$\mathbf{Y}_t$）表示由 n 个此类粒子组成的系统（$z_{i,t}, m_i$），其中 $i = 1, \ldots, n$，$z_{i,t} \in \Lambda$ 表示时刻 t 第 i 个粒子的位置，而 $m_i \in \mathbb{M}$ 是其恒定标记，即

$$\mathbf{Y}_t = ((z_{1,t}, m_1), \cdots, (z_{n,t}, m_n)).$$

与上一节的构造一致，$(\mathbf{X}_t)$ 中每个粒子的跳跃间轨迹将在此期间与 $(\mathbf{Y}_t)$ 的 n 轨迹重合。

作为一个一般性示例，我们假设 $(\mathbf{Y}_t)$ 遵循以下随机微分方程组，并从构型 $\mathbf{y}_0 \in (\Lambda \times \mathbb{M})^n$ 处的 $t = 0$ 开始，

$$(Move): \quad \begin{cases} dz_{i,t} = v_i(t, \mathbf{Y}_t)dt + \sigma_i(t, \mathbf{Y}_t)dB_{i,t}, \quad t \geq 0, \quad i = 1, \ldots, n, \\ \mathbf{Y}_0 = \mathbf{y}_0, \end{cases}$$

其中，漂移函数 v_i 的取值属于 $\mathbb{R}^d$，扩散 σ_i 为非负函数，而 $(B_{i,t})$、$i = 1, \ldots, n$ 是 n 相互独立的标准布朗运动，取值于 $\mathbb{R}^d$。这里，$\mathbf{y}_0$、v_i 和 σ_i 是待选的自由参数。

要确保 $(Move)$ 的解存在且唯一，需要对漂移函数和扩散函数施加一些条件。例如，若这些函数满足利普希茨条件 [20], ，则可保证这一点；以下示例均满足该条件。此外，由于假设每个粒子都在 $\mathbb{R}^d$的有界集合 Λ 中运动，实际应用中需要确保 $(Move)$ 的轨迹始终留在 Λ内。可以通过使轨迹在 Λ的边界处反射来实现这一点。

一般而言，$(Move)$允许每个粒子的运动取决于其标记，也取决于其他粒子（属于 $\mathbf{Y}_t$）的位置和标记。下面将详细介绍几个可能适用于生物学应用的实例。

例 1*(布朗运动)*：如果 $v_i(t, \mathbf{Y}_t) = 0$ 且 $\sigma_i(t, \mathbf{Y}_t) = \sigma$ (对于 $\sigma > 0$) 为常数，那么每个粒子都遵循具有相同扩散系数 σ 的布朗运动，且彼此独立。

例 2*(空间变化的扩散系数)*：如果 $v_i(t, \mathbf{Y}_t) = 0$ 且 $\sigma_i(t, \mathbf{Y}_t) = \sigma_{m_i}(z_{i,\,t})$，其中 σ_{m_i} 是定义在 Λ 上的正函数，那么每个粒子都遵循相互独立的扩散运动，其扩散系数取决于相应的标记，并且可能随空间变化。例如，假设 $\Lambda = \Lambda_1 \cup \Lambda_2$，其中 $\Lambda_1 \cap \Lambda_2 = \varnothing$，且对于 $m \in \mathbb{M}$,

$$\sigma_m(z) = \begin{cases} \sigma_{1,m} & \text{if} \quad z \in \Lambda_1, \\ \sigma_{2,m} & \text{if} \quad z \in \Lambda_2, \end{cases}$$

其中 $\sigma_{1,\,m} > 0$，$\sigma_{2,\,m} > 0$。那么，标记为 m 的每个粒子在 Λ_1 局部遵循扩散系数为 $\sigma_{1,\,m}$ 的布朗运动，并在 Λ_2 局部遵循扩散系数为 $\sigma_{2,\,m}$ 的布朗运动。需要注意的是，这样的 σ_m 并非利普希茨函数，因此需要足够光滑才能符合理论设定。这可以通过对 σ_m 与光滑截断函数进行卷积来实现。

例 3（超扩散和亚扩散运动）：若 $v_i(t, \mathbf{Y}_t) = v_{m_i}(z_{i,\,t})$ 且 $\sigma_i(t, \mathbf{Y}_t) = \sigma_{m_i}$，其中 v_{m_i} 定义在 Λ 和 $\sigma_{m_i} > 0$ 上，则每个粒子彼此独立地演化，其漂移和扩散系数取决于其标记。此例包含 [6] 中讨论的超扩散情形，其中 $v_{m_i}(z) = v_{m_i}$ 是恒定漂移。

其中还包括 [6], 中讨论的奥恩斯坦-乌伦贝克动力学：当 $v_{m_i}(z_{i,\ t}) = -\lambda_{m_i}(z_{i,\ t} - z_{i,\ 0})$ 时，$\lambda_{m_i} > 0$ 可解释为指向初始位置 $z_{i,\ 0}$的吸引力，从而形成一条受限轨迹。

示例 4（相互作用粒子）：本例展示如何通过朗之万动力学，在粒子之间纳入相互作用。为此，对于m、$m' \in \mathbb{M}$，我们引入统计物理学中所考虑的成对相互作用函数 $\Phi_{m,\ m'}$：对于$r > 0$、$\Phi_{m,\ m'}(r)$表示标记为 m 的粒子与标记为 m' 的粒子在相距 r 时的成对相互作用。若$\Phi_{m,\ m'}(r) = 0$，则不存在相互作用；若$\Phi_{m,\ m'}(r) > 0$，则两个相距 r的粒子之间存在抑制；若$\Phi_{m,\ m'}(r) < 0$，则存在吸引。[17]中给出了抑制性相互作用函数的示例。与这些相互作用相关的（过阻尼）朗之万动力学可写为 $(Move)$，其中$\sigma_i(t, \mathbf{Y}_t) = \sigma_{m_i}$，$\sigma_{m_i} > 0$，并且

$$v_i(t, \mathbf{Y}_t) = -\sum_{j \neq i} \nabla \Phi_{m_i, m_j}(\|z_{i,t} - z_{j,t}\|),$$

其中，∇ 表示梯度算子。因此，每个粒子都会沿着使其与其他粒子之间的两两相互作用函数值降低的方向运动。

示例 5（共定位粒子）：假设某些粒子（例如带有标记 m 的粒子）被认为与带有标记 $\tilde{m}$ 的粒子共定位。这意味着我们预期前者位于后者附近，并大致遵循相同的运动。具体而言，若要使带有标记 m 的粒子 i 与带有标记$\tilde{m}$ 的粒子 j 共定位，只需定义 $z_{i,\ t} = z_{j,\ t} + \sigma_i B_{i,\ t}$, $t > 0$，其中 $B_{i,\ t}$ 是 $\mathbb{R}^d$ 中的标准布朗运动，表示轨迹 i 相对于轨迹 j 的标准差，而$\sigma_i > 0$ 量化了该标准差的强度。此处，$z_{j,\ t}$ 可按前述示例中的方式定义，例如定义为带有标记 $\tilde{m}$ 的粒子的典型轨迹。

2.3.2 强度函数

回顾一下，强度函数 β、δ 和 τ 分别决定下一次出生、死亡和转化的等待时间。直观地说，已知粒子处于构型 $\mathbf{X}_t$ 时，出生事件在时间区间 $[t, t + dt]$ 内发生的概率为 $\beta(\mathbf{X}_t)\ dt$，δ 和 τ也同理。因此，这些概率可能会随时间根据粒子构型而变化。例如，由于竞争，粒子数量很多或某个区域内粒子浓度较高时，死亡事件就更有可能发生。下面给出一些自然的示例。对于每个示例，β、δ 或 τ 中的任一项都可以用类似方式设定，即使我们只聚焦于其中一项。

示例6（常数强度）：最简单的情况是强度函数为常数，例如 $\beta(\mathbf{X}_t) = \beta$，其中 $\beta > 0$。此时，出生以恒定速率发生，因此我们可以预期，在时间区间 $[s,t]$内平均会出现 $\beta \times (s-t)$ 个新粒子。

例 7（取决于基数的强度）：如果 $\delta(\mathbf{X}_t) = \delta n(\mathbf{X}_t)$，且 $\delta > 0$，那么粒子越多，我们观察到的死亡就越多。当每个粒子都被认为具有恒定的死亡率 δ 时，这是一种自然情形，因此时刻 t 粒子系统的总死亡强度就是各粒子死亡强度之和，即 $\delta n(\mathbf{X}_t)$。

例 8（空间变化强度）：假设粒子的标记（例如其运动状态）在 Λ 的某些区域比其他区域更容易发生变化，那么变换强度 τ 可以反映这种依赖关系。例如，令 $\Lambda = \Lambda_1 \cup \Lambda_2$，其中$\Lambda_1 \cap \Lambda_2 = \varnothing$，并定义 $\tau(\mathbf{X}_t) = \tau_1 n_1(\mathbf{X}_t) + \tau_2 n_2(\mathbf{X}_t)$，其中 $\tau_1 < \tau_2$，且 $n_1(\mathbf{X}_t)$（分别为$n_2(\mathbf{X}_t)$）表示 Λ_1（分别为 Λ_2）中的粒子数。那么，对于给定的基数$n(\mathbf{X}_t)$，Λ_2 中的粒子占比越高，发生的转化就越多。请注意，为了严谨起见，我们应考虑 τ 的连续版本，这可以通过与光滑截断函数进行卷积来实现。

例9（由共定位引起的转化）：假设某些 m-粒子（即带有标记 m的粒子）可以与某些 $\tilde{m}$-粒子共定位。此外，假设这些粒子被赋予第二标记，用以编码其运动状态（例如扩散、亚扩散或超扩散）。最终，在粒子的动力学过程中，一个未共定位的 m-粒子可能与一个 $\tilde{m}$-粒子共定位，也就是说，它与一个 $\tilde{m}$-粒子的距离变为 D，其中 D 是指定的共定位距离。如果发生这种情况，我们可能会预期 m-粒子的运动状态会变得与 $\tilde{m}$-粒子相似，因此必须发生转化。令$n_{D,m,\tilde{m}}(\mathbf{X}_t)$表示标记为 m 和 $\tilde{m}$且运动状态不同的粒子对中，距离小于 D的粒子对数。于是，对于某个 $\tau > 0$，我们可以定义$\tau(\mathbf{X}_t) = \tau\, n_{D,m,\tilde{m}}(\mathbf{X}_t)$，使得在上述情况下（此处为运动状态的）转化很可能发生。注意，如果 τ 很大，那么一旦$n_{D,m,\tilde{m}}(\mathbf{X}_t) = 1$，这种转化就会很快发生，因此$n_{D,m,\tilde{m}}(\mathbf{X}_t) > 1$不太可能被观测到。这里同样可以通过卷积引入 τ 的光滑版本，以确保其连续性。

2.3.3 转移概率函数

下面依次详细介绍三种可能的转变：出生、死亡和转化。

对于出生，请记住，$p^{\beta}((z,\ m)|\mathbf{x})$ 表示在粒子系统处于构型 $\mathbf{x}$ 的条件下，粒子出现在位置 z 并带有标记 m 的概率密度函数（pdf）。设定该概率有两种方法：

1. 首先根据某个概率$p^\beta(m|\mathbf{x})$抽取新粒子的标记 m ，然后根据某个概率密度函数 $p^\beta(z|\mathbf{x}\ m)$，在给定其标记的条件下抽取新粒子的位置。这会得到分解式 $p^\beta((z,\ m)|\mathbf{x}) = p^\beta(m|\mathbf{x})p^\beta(z|\mathbf{x}\ m)$。

2. 首先，根据某个概率密度函数 $p^\beta(z|\mathbf{x})$ 生成新粒子的位置，然后根据给定位置，以概率 $p^\beta(m|\mathbf{x}\ z)$ 生成其标记 m 。由此可得分解式 $p^\beta((z,m)|\mathbf{x}) = p^\beta(z|\mathbf{x})p^\beta(m|\mathbf{x}\ z)$。

例10*(均匀出生)*：这是一个简单示例，其中出生不依赖于环境，在空间中均匀分布，且标记按照某些指定概率 p_m 抽取，其中 $\sum_{m\in\mathbb{M}} p_m = 1$。上述两种方法因此等价于 $p^\beta(m|\mathbf{x}) = p^\beta(m|\mathbf{x}\ z) = p_m$ 以及 $p^\beta(z|\mathbf{x}\ m) = p^\beta(z|\mathbf{x}) = 1/|\Lambda|$ 当 $z \in \Lambda$ 时。

例11*(共定位出生)*：这里我们采用上述第一种方法。首先，像前一个示例一样，通过设定 $p^\beta(m|\mathbf{x}) = p_m$ ，使标记的抽取独立于环境，其中 $\sum_{m\in\mathbb{M}} p_m = 1$。其次，为了生成一个新 m-粒子的位置，并使其与 $\tilde{m}$-粒子共定位，我们可以使用各向同性正态分布混合，其中心位于每个 $\tilde{m}$-粒子处，标准差为 $\sigma > 0$。记 $\tilde{n}(\mathbf{x})$ 为 $\mathbf{x}$ 中 $\tilde{m}$-粒子的数量， $\tilde{z}_i$ 为它们的位置（$i = 1, \ldots, \tilde{n}(\mathbf{x})$），则这意味着

$$p^\beta(z|\mathbf{x}, m) = \frac{1}{\tilde{n}(\mathbf{x})} \sum_{i=1}^{\tilde{n}(\mathbf{x})} \frac{1}{(\sigma\sqrt{2\pi})^d} \exp\left(-\frac{\|z - \tilde{z}_i\|}{2\sigma^2}\right). \tag{2.1}$$

需要注意的是，严格来说， $p^\beta(z|\mathbf{x}\ m)$ 应限制在 Λ 内，并进行适当的归一化，否则可能会在 Λ 之外生成粒子。我们省略这些细节。

例 12（空间依赖的新标记）：我们可以采用第二种方法，先为新粒子生成一个均匀分布的位置，即 $p^\beta(z|\mathbf{x}) = 1/|\Lambda|$ ，其中 $z \in \Lambda$；然后根据生成的位置抽取标记。例如，令$\Lambda = \Lambda_1 \cup \Lambda_2$，其中 $\Lambda_1 \cap \Lambda_2 = \varnothing$，并设

$$p^\beta(m|\mathbf{x}, z) = \begin{cases} p_{1,m} & \text{if} \quad z \in \Lambda_1, \\ p_{2,m} & \text{if} \quad z \in \Lambda_2, \end{cases}$$

其中 $\sum_{m\in\mathbb{M}} p_{1,m} = \sum_{m\in\mathbb{M}} p_{2,m} = 1$。那么，标记的分布可能因位置而异。

现在我们聚焦于死亡转移，即概率 $p^\delta(x_i|\mathbf{x})$，对于 $x_i \in \mathbf{x}$，当发生死亡时，粒子 x_i in $\mathbf{x}$ 消失。

示例 13（均匀死亡）：最简单的情况是，死亡均匀地发生在现有粒子中，即 $p^{\delta}(x_i|\mathbf{x}) = 1/n(\mathbf{x})$，其中 $i = 1, \ldots\ldots, n(\mathbf{x})$。

示例 14（竞争导致的死亡）：我们可以设想，由于竞争，如果某个粒子周围的邻居过多，它就更有可能消失。令 $n_D(x_i)$ 表示 x_i 距离 $D > 0$以内的邻近粒子数量。为了考虑距离 $D > 0$ 范围内的竞争，我们可以定义 $p^{\delta}(x_i|\mathbf{x}) = n_D(x_i)/\sum_{j=1}^{n(\mathbf{x})} n_D(x_j)$。类似地，如果有相关需要，我们也可以只计算具有特定标记的邻居数量。

最后，我们关注 $p^{\tau}((z_i, m_i)\, m|\mathbf{x})$，其中 $(z_i, m_i) \in \mathbf{x}$，它表示当发生转化时，位于 $\mathbf{x}$ 的粒子 $x_i = (z_i, m_i)$ 将其标记从 m_i 改为 $m \neq m_i$ 的概率。与出生转移概率类似，自然可以将此概率分解为

$$p^{\tau}((z_i, m_i), m|\mathbf{x}) = p^{\tau}((z_i, m_i)|\mathbf{x}) p^{\tau}(m|\mathbf{x}, (z_i, m_i)),$$

其中，$p^{\tau}((z_i, m_i)|\mathbf{x})$ 表示在构型 $\mathbf{x}$ 中选择粒子 (z_i, m_i) 以改变其标记的概率，而 $p^{\tau}(m|\mathbf{x}, (z_i, m_i))$ 表示在已知转化后的粒子位于 z_i 且标记为 m_i 的条件下，选择新标记 m 的概率。

例15（独立于环境的变换）：一种典型情况是，从现有粒子中均匀抽取待转化的粒子，即$p^{\tau}((z_i, m_i)|\mathbf{x}) = 1/n(\mathbf{x})$，并依据一个元素为 $p_{m,\, m'} \geq 0, m, m' \in \mathbb{M}$ 的转移矩阵，独立于环境进行转化，这些元素表示从标记 m 转化为标记 m'的概率。这里，对于所有 $m \in \mathbb{M}$，我们假设 $p_{m,\, m} = 0$，以确保这是真正的转化；当然还有 $\sum_{m' \in \mathbb{M}} p_{m,\, m'} = 1$。采用这种形式，我们便有 $p^{\tau}(m|\mathbf{x}, (z_i, m_i)) = p_{m_i,\, m}$。

例16（空间依赖变换）：要使前一个例子具有空间依赖性，引入 $p_m(z)$，它是关于 Λ 的概率密度函数，表示在 Λ 中的位置上，标记为 m 的粒子发生转化的可能性。然后我们可以设定

$$p^{\tau}((z_i, m_i)|\mathbf{x}) = \frac{n_i(\mathbf{x})}{n(\mathbf{x})} \frac{p_{m_i}(z_i)}{\sum_{j=1}^{n_i(\mathbf{x})} p_{m_i}(z_j^{(m_i)})},$$

其中 $n_i(\mathbf{x})$ 表示标记为 m_i 且位置为 $\mathbf{x}$ 的粒子数量，$z^{(m_i)}{}_j, j = 1, \ldots, n_i(\mathbf{x})$。在此表达式中，$n_i(\mathbf{x})/n(\mathbf{x})$ 是一个权重，用于计入标记 m_i 在 $\mathbf{x}$ 处的占比；分母中的求和则用于归一化，使概率之和为 1。注意，如果 $p_m(z)$ 是 Λ上的均匀概率密度函数，那么我们就得到均匀分布 $p^{\tau}((z_i, m_i)|\mathbf{x}) = 1/n(\mathbf{x})$。此外，按照上述方式选定粒子后，我们可以应用如下依赖于空间的转化。令 $\Lambda = \Lambda_1 \cup \Lambda_2$ 满足

$\Lambda_1 \cap \Lambda_2 = \varnothing$ 并令两个不同的转移矩阵，其对应元素分别为 $p_{m}^{(1)}{}_{,m'}$ 和$p_{m}^{(2)}{}_{,m'}$，其中 $m, m' \in \mathbb{M}$。那么我们可以设定

$$p^{\tau}(m|\mathbf{x},(z_i,m_i)) = \begin{cases} p^{(1)}_{m_i,m} & \text{if} \quad z_i \in \Lambda_1, \\ p^{(2)}_{m_i,m} & \text{if} \quad z_i \in \Lambda_2. \end{cases}$$

因此，无论所选要转化的粒子位于 Λ_1 还是 Λ_2，转化都不会遵循相同的分布。

例17（由共定位引起的转化）：假设我们处于与例9相同的情形，其中 m-粒子可以与 $\tilde{m}$-粒子共定位。与该例一样，我们假设当 $n_{D,\ m,\ \tilde{m}}(\mathbf{X}_t) \geq 1$ 时会发生转化，其中 $n_{D,\ m,\ \tilde{m}}(\mathbf{X}_t)$ 表示标记为 m 和 $\tilde{m}$、运动状态不同且彼此距离小于 D 的粒子对数目。于是，发生转化时，我们可以从所有与运动状态不同的 $\tilde{m}$-粒子距离小于 D 的 m-粒子中，均匀选择一个 m-粒子进行转化。随后，转化会使所选 m-粒子的运动状态变得与其最近的 $\tilde{m}$-粒子的运动状态相似。

3 应用于 Langerin/Rab11 蛋白的联合动力学

3.1 数据集描述

我们所考虑的数据集来自一项观测：研究人员采用三维多角度全内反射荧光（全内反射荧光）显微镜技术，观察 RPE1 活细胞中 m- Cherry Langerin 和 GFP Rab11 蛋白的细胞内运输 [3], ，具体而言，是将其沿 z轴投影到靠近质膜的二维平面上。由此得到一段包含 1199 帧的二维图像序列，每帧时长为 140 ms，显示了两种蛋白质的同步位置。图2顶部的两幅图分别展示了原始序列的第一帧中 Langerin 荧光通道和 Rab11 荧光通道的图像，这些图像由双视野光学装置同步记录。请注意，细胞贴附在纤连蛋白微图案上，该图案会限制细胞骨架元素等细胞内组分的分布，并使细胞形状具有可重复性，这解释了细胞呈现的“伞状”形态。这些原始序列按照 [22进行后处理； 21], 随后，每个亮点均以一个点表示，并应用 U-track算法 [11] 估算粒子轨迹。图2底部的图像分别展示了 Langerin 通道和 Rab11 通道得到的轨迹。这些轨迹已

随后采用 [6] 中提出的方法进行进一步分析，将其分为三种扩散状态：布朗运动、超扩散和亚扩散，在图2中分别以蓝色、红色和绿色表示。

更具体地说，在分析所有轨迹时，我们根据其所处状态，按照 [6]，对每条轨迹拟合三个参数模型：

1. 对于布朗运动状态（蓝色）：布朗运动；2. 对于超扩散状态（红色）：带恒定漂移的布朗运动；3. 对于亚扩散状态（绿色）：奥恩斯坦-乌伦贝克过程。

每条轨迹都有各自的参数（见示例 1 和 3），通过最大似然法估计。此外，一些轨迹可能会从一种状态转变为另一种状态，这在 BDM模型中对应一种“转化”，将在下一节中具体说明。

图 3 总结了 Langerin序列中所得轨迹的不同特征（Rab11序列也分析了相同的特征，但此处不作详细介绍）。底部的直方图按状态显示了所有轨迹的持续时间（以帧为单位）。可以观察到，与亚扩散轨迹（绿色）相比，布朗运动轨迹（蓝色）和超扩散轨迹（红色）的平均寿命都相当短。右上方的箱线图按状态显示了每帧粒子数量的分布：布朗运动占多数，其次是亚扩散运动，超扩散运动则占少数。最后，左上方的圆形直方图旨在展示超扩散轨迹（红色）的漂移向量方向。具体而言，在此图中，我们记录了漂移角相对于指向细胞中心方向的偏转角（以度为单位）。例如，若漂移朝向细胞中心，则该偏转角为 0°；若漂移方向相反，则为 180°。从图中可以看出，大多数偏转角约为 0° 或 180°，这意味着红色轨迹主要沿径向朝细胞中心移动（或从细胞中心出发）。

上述描述符有助于在下一节中校准BDM模型的参数，也将作为评估我们模拟质量的基准。然而，重要的是要牢记，它们存在一些由不完善的跟踪算法导致的近似和误差。具体而言，数据中的任何轨迹都不会短于10帧，这是我们的跟踪方法所能检测到的最短长度。从图2底部的图中还可以清楚地看到，一些有方向性的（超扩散）轨迹被错误地标记为蓝色，这可以用 [6]的多重检验程序来解释，该程序旨在尽量减少假阳性（即不良

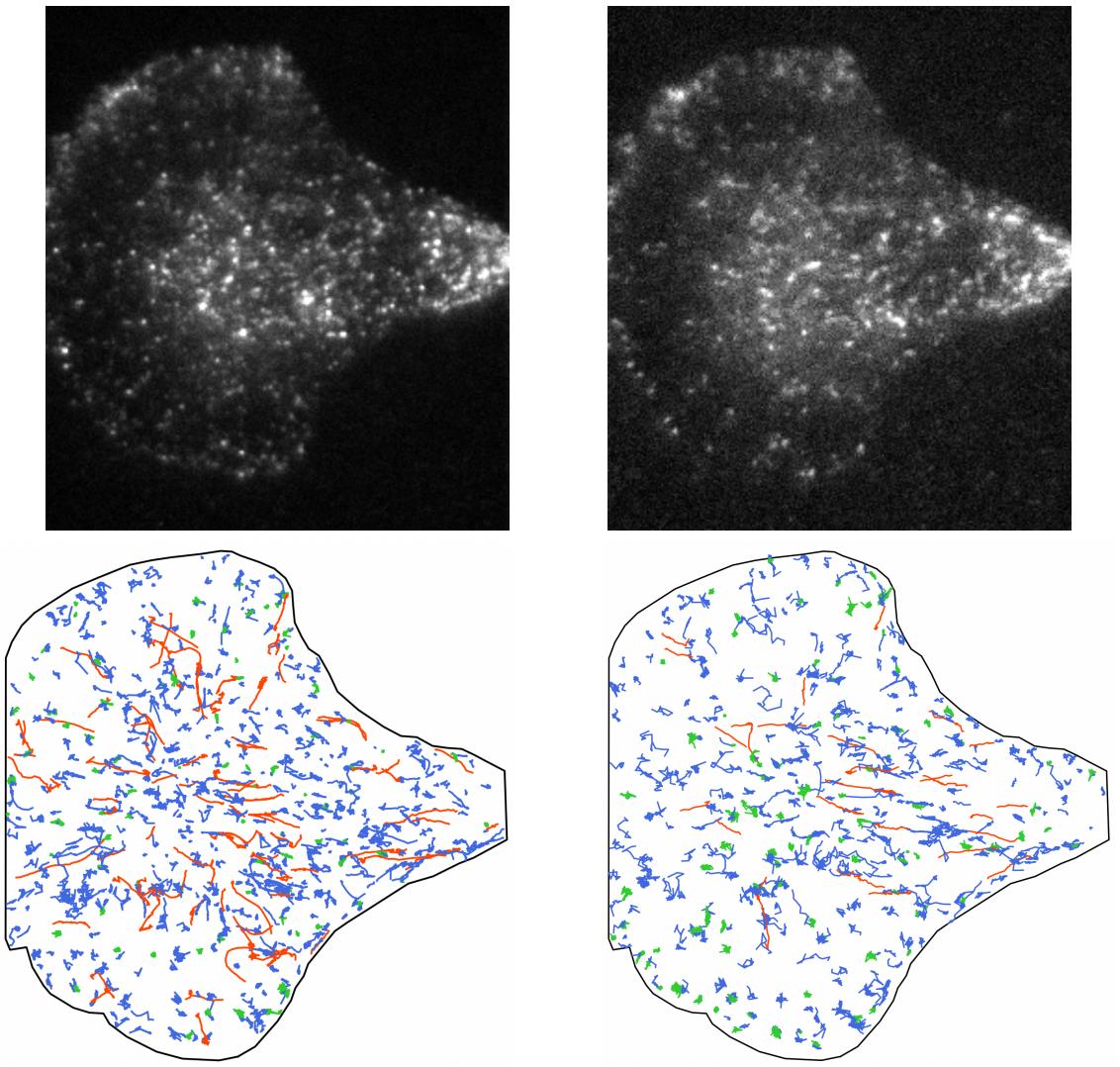


图2：上：原始序列的第一帧，亮点显示了 Langerin 蛋白（左侧）和 Rab11 蛋白（右侧）的位置。下：在 Langerin 蛋白（左侧）和 Rab11 蛋白（右侧）的序列中检测并追踪到的所有轨迹，按估计的运动状态着色（布朗运动为蓝色，超扩散为红色，亚扩散为绿色）。

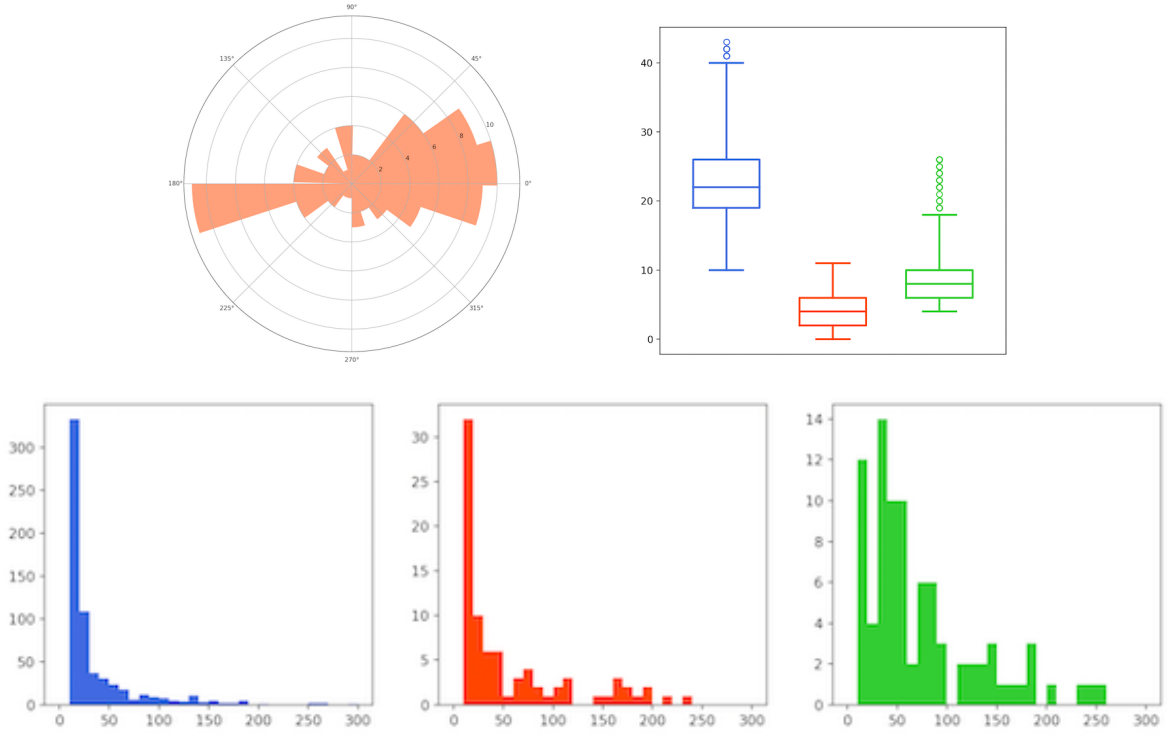


图 3：真实数据序列中 Langerin 轨迹的描述符。左上：超扩散轨迹的漂移相对于指向细胞中心方向的偏转角圆形直方图。右上：按运动状态分类的每帧轨迹数量箱线图（蓝色：布朗运动，红色：超扩散，绿色：亚扩散）。下方：按运动状态分类的各条轨迹寿命（以帧为单位）直方图（颜色标记同上）。

绿色轨迹或误判为红色的轨迹），而可能会导致过多的假阴性（即误判为蓝色的轨迹）。

关于轨迹的出生和死亡，我们在表 1 中按蛋白质类型和运动状态汇总了真实数据集中的观察总数。Langerin 蛋白的状态变换次数则列于表2。对于 Rab11 蛋白，在整个序列中仅观察到一次从布朗运动向亚扩散运动的状态转换。

为详细研究这些跳跃动力学，我们借鉴了 [18]，中针对同一数据集开展的研究。该研究得出结论：对于每种蛋白质和运动状态，出生强度都是恒定的，如示例6所示；而死亡强度

		布朗运动	超扩散	亚扩散	总计	
出生	Langerin	603	78	66	747	1248
	Rab11	393	24	84	501	
死亡	Langerin	602	77	89	768	1282
	Rab11	395	26	93	514	

表 1：根据蛋白质类型和运动状态，真实数据集序列中观测到的轨迹出生和死亡总数。

从 \到	布朗运动	超扩散	亚扩散
布朗运动	0	0	9
超扩散	0	0	1
亚扩散	5	0	0

表2：数　　真实数据中的 Langerin 轨迹的状态转化观测数量

与现存粒子数量成正比，如示例7所示。鉴于观察到的运动状态转变数量较少，其强度也可视为恒定。关于转移概率函数，死亡均匀地发生在所有现存粒子中，如示例13所示。至于出生转移，对于 Rab11 蛋白，没有理由不采用细胞内的均匀分布作为密度（示例10）。但由于存在共定位（该数据集中的观察结果见 [19]），Langerin 阳性结构的出生密度可近似为两种分布的混合：93% 的 Langerin 出生服从均匀分布，7% 的 Langerin 出生服从围绕现存 Rab11 囊泡的共定位密度，如示例11所示。这些比例以及其他参数均通过最大似然估计得出。

3.2 合成序列模拟

3.2.1 模型参数设置

基于上一节的数据分析，我们现在可以明确第 2 节中介绍的带转化的 BDM 过程的所有特征，以模拟细胞内 Langerin/Rab11 蛋白的联合动力学。为了建立

与理论符号的联系，感兴趣区域 Λ 表示维度为 $d = 2$的细胞。 Λ 中的每个粒子都将用集合 $\mathbb{M} = \{(L, 1), (L, 2), (L, 3), (R, 1), (R, 2), (R, 3)\}$中的一个标记，其中 L 代表 Langerin 蛋白， R 代表 Rab11 蛋白，数字 1、2 或 3 分别表示粒子的运动状态：布朗运动、超扩散或亚扩散。

关于每条轨迹的运动，它遵循其标记所指示的状态，并与上一节详述的真实数据集中的观测轨迹一致，另见示例 1 和 3：

- 对于布朗运动，我们根据从真实数据集中同类型蛋白质（L 或 R）的布朗运动中估算出的扩散系数经验分布，抽取扩散系数；

- 对于超扩散运动，我们生成具有恒定漂移的布朗运动，并采用相同的策略选择扩散系数；漂移向量的选择方式如下：默认情况下，其方向朝向细胞中心，但该方向会受到一个偏差的影响，该偏差从图 3 左上方圆形直方图所示的经验分布中抽取。此外，其范数从真实数据集中观测到的漂移范数经验分布中抽取。同样，Langerin 和 Rab11 蛋白各自使用不同的参数集；

- 对于亚扩散运动，我们为所有粒子生成扩散系数为 0.4 的奥恩斯坦-乌伦贝克过程（该系数取自真实数据集的平均值），并为 Langerin 蛋白设置参数 $\lambda = 5.96$ ，为 Rab11 蛋白设置参数 $\lambda = 7.41$ 。

这些数值的单位是像素，在我们的图像中，一个像素等于 $160 \times 160\,\mathrm{nm}^2$ 。

关于强度函数，根据真实数据分析的结论，我们将出生强度和变换强度设为常数。与表 1 一致，总出生强度可估计为 $\beta(\mathbf{x}) = 1248/167.86 = 7.43$，无论粒子构型 $\mathbf{x}$ 如何，因为 1248 是观测到的出生总数，167.86 是序列的总时长（秒）。同样，由于在真实序列中观测到 16 次变换，我们设 $\tau(\mathbf{x}) = 16/167.86 = 0.095$ 。对于死亡强度，针对每种标记 $m \in \mathbb{M}$，我们令其与粒子数成正比，即 $\delta_m n_m(\mathbf{x})$，其中 $n_m(\mathbf{x})$ 是构型 $\mathbf{x}$ 中具有标记 m 的粒子数，而 δ_m 根据真实数据集估计如下： $\delta_m = 0.17$ if $m = (L, 1)$, $\delta_m = 0.14$ if $m = (L, 2)$, $\delta_m = 0.07$ if $m = (L, 3)$, $\delta_m = 0.21$ if $m = (R, 1)$, $\delta_m = 0.25$ if $m = (R, 2)$, 以及 $\delta_m = 0.08$ if $m = (R, 3)$。因此，粒子构型 $\mathbf{x}$ 的总死亡强度为 $\delta(\mathbf{x}) = \sum_{m \in \mathbb{M}} \delta_m n_m(\mathbf{x})$。

最后，我们将转移概率函数设定如下。对于死亡转移，在构型 $\mathbf{x}$ 中杀死粒子 $x_i = (z_i, m_i)$ 的概率设为

$$p^{\delta}((z_i, m_i)|\mathbf{x}) = \frac{\delta_{m_i}}{\delta(\mathbf{x})},$$

这意味着我们首先以概率 $\delta_m n_m(\mathbf{x})/\delta(\mathbf{x})$ 抽取标记 m ，然后从所有标记为 m 的现有粒子中均匀选取一个。对于变换转移，我们首先按照真实序列中观察到的转化率，以 $15/16$ 的概率选择 Langerin，以 $1/16$ 的概率选择 Rab11，确定要转化的蛋白质类型；其次，从该类型的所有现有粒子中均匀选取一个；最后，如例15所示，根据以下转移矩阵进行状态变换（行表示变换前的状态，列表示变换后的状态）：

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1/4 & 3/4 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{pmatrix} \end{array}.$$

该矩阵与表2中关于 Langerin 蛋白的内容一致；我们还加入了一些从状态1到状态2以及从状态2到状态1的可能转移，我们认为这些转移很可能发生，尽管在真实序列中观察到的转化（相当罕见）并未出现这些转移。Rab11 蛋白也采用了相同的转移矩阵，因为真实序列中观察到的转化数量不足（仅一次），无法据此设计更精细的选择方案。

接下来需要设定出生转移概率。首先，根据表 1，我们以概率 $747/1248$ 选择生成 Langerin 蛋白，以概率 $501/1248$ 选择生成 Rab11 蛋白。如果选中的类型为 Rab11，则在细胞中均匀生成该蛋白，其状态为 1 的概率为 0.784，状态为 2 的概率为 0.048，状态为 3 的概率为 0.168，这对应于真实 Rab11 序列中各状态出生数占所有出生数的相对比例。如果选中的类型为 Langerin，则以概率 $p = 0.07$抛硬币决定是否共定位。如果没有共定位，则在细胞中均匀生成新的 Langerin 蛋白，其状态为 1 的概率为 0.807，状态为 2 的概率为 0.104，状态为 3 的概率为 0.088（观测到的出生相对比例）。如果发生共定位，则根据例11中的密度（2.1），在现有 Rab11 蛋白周围生成新的 Langerin 蛋白，其中通过最大似然估计 $\sigma = 1.1$。在这种情况下，新生成的 Langerin 蛋白的状态及其超扩散运动的漂移向量与其共定位的 Rab11 蛋白相似。

3.2.2 对生成的模拟结果进行分析

我们按照上一节的模型生成了 100 个序列，其时长与第 3.1 节真实序列的时长相同，即 167.86 秒，共 1199 帧。图 4 和图 5 展示了来自两个模拟序列的 Langerin 轨迹的一些描述符，可与图 2 和图 3 中真实数据的类似结果进行比较。其他模拟序列的结果可在我们的 GitHub代码库中查看。我们还在表 3 中汇总了 100 个模拟序列的平均出生数和死亡数，可与表 1 进行比较。图形和定量结果均表明，我们的生成器能够生成联合动力学，其特征与真实数据序列中观察到的特征相当。

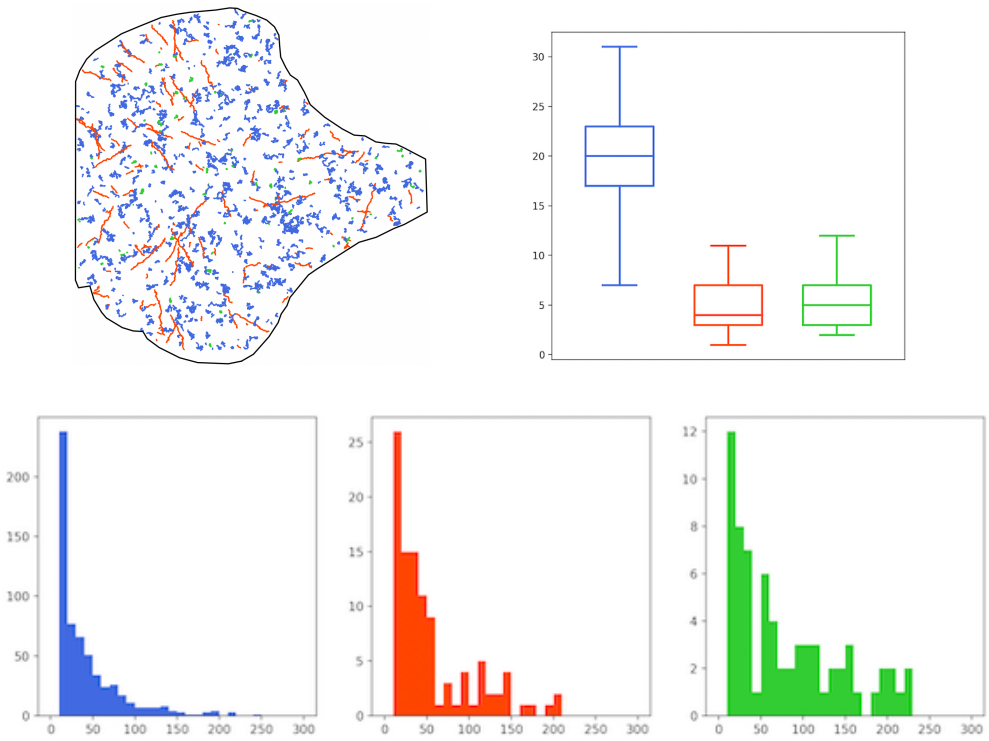


图 4：首个模拟序列中 Langerin 轨迹的描述符。左上：轨迹集合，按其运动状态着色（蓝色：布朗运动，红色：超扩散，绿色：亚扩散）。右上：每帧的轨迹数量箱线图，按其状态分类（颜色标记相同）。下方：各轨迹寿命（以帧为单位）的直方图，按其状态分类（颜色标记相同）。

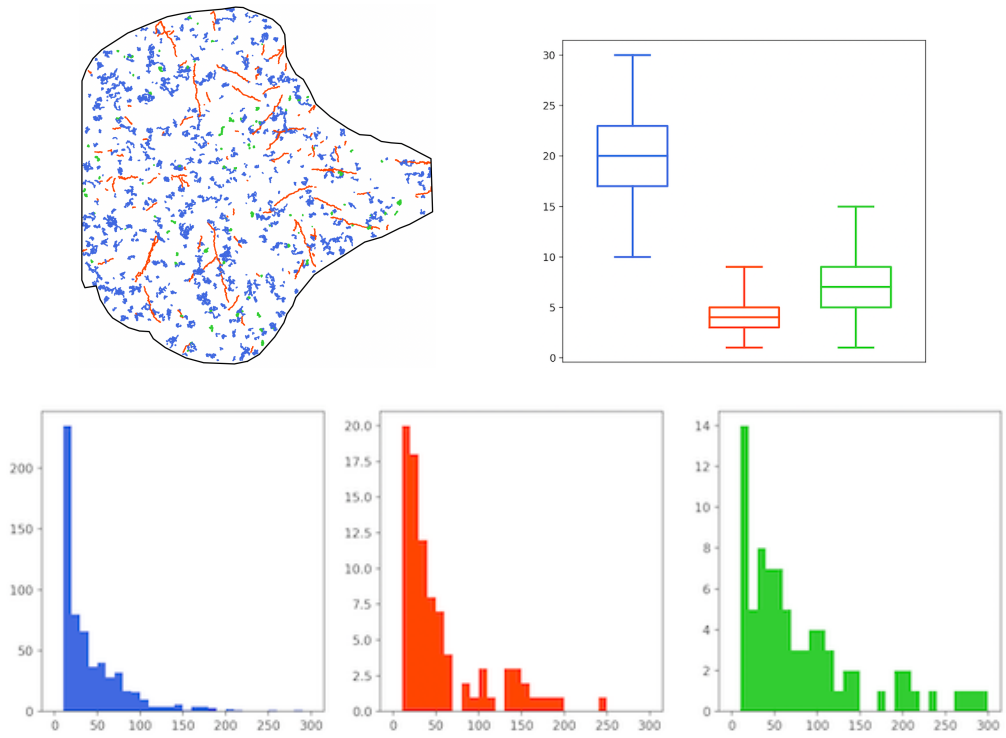


图 5：描述　　第二个模拟序列的 Langerin 轨迹，与图 4 中相同。

4 结论与展望

我们采用了一种原创随机模型，即带有转化的多类型出生-死亡-移动过程，以生成细胞内生物分子动力学的逼真图像序列。该随机过程不仅对粒子的个体轨迹进行建模，还能够生成每条轨迹随时间的出现（即出生）、消失（即死亡）和状态切换（即转化）；这些事件的发生可能取决于现有粒子的构型（例如粒子数量）。该模型非常灵活，由三组参数定义：1）描述轨迹集合的系统（通常是一组随机微分方程）；2）决定新粒子出现、粒子消失或发生状态切换之前等待时间的强度函数；3）转移概率函数，用于确定出生时新粒子出现的位置、死亡时消失的是哪个粒子，以及转化时哪个粒子发生状态切换（以及如何切换）。我们详细介绍了这种模型设定的许多示例。我们通过生成

		布朗运动	超扩散	亚扩散	总计	
出生	Langerin	581	85	76	742	1239
	Rab11	391	22	83	496	
死亡	Langerin	584	86	77	747	1248
	Rab11	397	23	80	500	

表 3：每个序列中轨迹出生和死亡总数的平均值，基于 100 个模拟序列。

细胞内 Langerin/Rab11蛋白质联合动力学的逼真图像序列，展示了这种方法的有效性；为精确校准模型，我们还依据初步的数据分析进行了验证。

由于模型具有很高的灵活性，选择模型特征和参数是一个重要步骤。为了改进我们的方法，基于大量图像序列开展更深入的实证研究，或许有助于稳健地校准模型。参数拟合完成后，生成一个图像序列的速度相当快：在普通笔记本电脑上，生成 2000 帧约需一分钟，每帧包含约 70 条相互作用的轨迹。为了生成更加逼真的图像序列，我们可以考虑使用点扩散函数对生成的粒子系统进行模糊处理，并添加一些噪声和背景。此外，还可以从真实图像序列和合成图像序列中提取更多特征，以加强对生成器的质量评估。

附录：模拟算法

在本附录中，我们按照第2.2节的构造，提供一种模拟带突变（或转化）的出生-死亡-移动过程的正式算法。这是对 [18] 中介绍的出生-死亡-移动过程（不含转化）算法的改进。其思路是先在较短的时间长度 Δ 内生成跳跃间移动，然后检验这段时间内是否发生了跳跃（在下文的算法1中，发生跳跃的概率为 p）。如果发生了跳跃，我们就生成跳跃时间和跳跃；否则，我们继续在后续的时间长度 Δ 内模拟跳跃间移动，检验是否发生了跳跃，以此类推。无论 $\Delta > 0$ 取何值，该算法都有效，但一种高效的做法是为 Δ 设定一个较小的值。其实现可在我们的 GitHub代码库中获取。

我们按照第 2 节中的设定。$2\ \alpha = \beta + \delta + \tau$和 $T_0 = 0$。在算法 2 中，我们用 $\mathbf{X}^{T_j^-}$ 表示跳跃时间 T_j之前的构型。为了运行算法 1 和算法 2，我们需要

以下输入：

- $T > 0$：模拟的终止时间；
- $\mathbf{X}_0 \in E$：粒子的初始构型；
- $\Delta > 0$：分段模拟的时间长度；
- β, δ, τ：出生、死亡和转化的强度函数；
- p^β, p^δ, p^τ：出生、死亡和转化的转移概率；
- $Algo.Move(\mathbf{y}_0, \Delta)$：算法，对于$\mathbf{y}_0 \in E$和 $\Delta > 0$, $n(\mathbf{y}_0)$，返回在 $[0, \Delta]$ 上的轨迹，这些轨迹遵循随机微分方程组（$Move$），初始构型为$\mathbf{y}_0$。

参考文献

[1] Badoual, A., Arizono, M., Denizot, A., Ducros, M., Berry, H., Nägerl, U. V., and Kervrann, C. 晶格光片显微图像中的星形胶质细胞钙动力学模拟。载于《2021年 IEEE 第18届国际生物医学成像研讨会（ISBI）》（2021），IEEE，第135–139页。[2] Bolte, S., and Cordelieres, F. 光学显微镜中亚细胞共定位分析指南。《显微镜技术杂志》224（2006），213–232。[3] Boulanger, J., Gueudry, C., Munch, D., Cinquin, B., Paul-Gilloteaux,P., Bardin, S., Guérin, C., Senger, F., Blanchoin, L., and Salamero, J.通过入射角扫描和方位角平均实现快速高分辨率三维全内反射荧光显微镜。《美国国家科学院院刊》111，48（2014年），17164–17169。[4] Boulanger, J., Kervrann, C., and Bouthemy, P. 用于视频显微和荧光图像中细胞内动力学与运输的模拟与估计框架。《医学图像分析》13，1（2009），132–142。[5] Bressloff, P. C., and Newby, J. M. 细胞内运输的随机模型。《现代物理评论》85，1（2013），135。[6] Briane, V., Kervrann, C., and Vimond, M. 活细胞中粒子轨迹的统计分析。《物理评论 E》97，（6-1）：062121（2018）。

算法1: 在时间区间 $[0,T]$ 上进行模拟

设置 $t=0$ 和 $j=0$;
当 $t<T$ **时执行**
  设置 $k=1$;
  当 $k\neq 0$ **时执行**
    设置 $\Delta_j =$ 最小值$(\Delta, T-T_j-(k-1)\Delta)$;
    生成 $(\mathbf{Y}_s)_{0\le s\le \Delta_j}$ 通过 $Algo.Move(\mathbf{X}_{T_j+(k-1)\Delta},\ \Delta_j)$;
    设 $p=\exp\left(-\int_0^{\Delta_j}\alpha(\mathbf{Y}_u)du\right)$;
    生成 $U\sim U([0,1])$;
    如果 $U\le p$ **则**
      设置 $\mathbf{X}_s=\mathbf{Y}_{s-T_j-(k-1)\Delta}$ 对于 $s\in[T_j+(k-1)\Delta,\ T_j+(k-1)\Delta+\Delta_j]$;
      如果 $\Delta_j=T-T_j-(k-1)\Delta$ **则**
        $k=0$;
        $t\leftarrow T$;
      else
        $k\leftarrow k+1$;
      end
    else
      生成等待时间 τ , 根据分布

$$\mathbb{P}(\tau<s|(k-1)\Delta<\tau<(k-1)\Delta+\Delta_j)=\frac{1-\exp\left(-\int_0^{s-(k-1)\Delta}\alpha(Y_u)du\right)}{(1-p)},\quad \forall s\in[(k-1)\Delta,(k-1)\Delta+\Delta_j]\ ;$$

      设置 $T_{j+1}=T_j+\tau$ 并 $\mathbf{X}_s=\mathbf{Y}_{s-T_j-(k-1)\Delta}$ 对于 $s\in[T_j+(k-1)\Delta,\ T_{j+1}[$;
      通过算法 2 生成 $\mathbf{X}_{T_{j+1}}$, 给定 $\mathbf{X}_{T^-_{j+1}}=\mathbf{Y}_{T_{j+1}-T_j-(k-1)\Delta}$;.
      $k=0$;
      $t\leftarrow T_{j+1}$;
      $j\leftarrow j+1$;
    end
  end
end

Algorithm 2: Simulation of the jump at $t = T_{j+1}$ given $\mathbf{X}_{T_{j+1}^-}$

generate $U \sim U([0,1])$;
if $U \leq \beta(\mathbf{X}_{T_{j+1}^-})/\alpha(\mathbf{X}_{T_{j+1}^-})$ **then**
  generate a new particle (z,m) according to $p^{\beta}((z,m)|\mathbf{X}_{T_{j+1}^-})$;
  set $\mathbf{X}_{T_{j+1}} = \mathbf{X}_{T_{j+1}^-} \cup (z,m)$;
else
  if $U \leq (\beta+\delta)(\mathbf{X}_{T_{j+1}^-})/\alpha(\mathbf{X}_{T_{j+1}^-})$ **then**
    sample $x_i \in \mathbf{X}_{T_{j+1}^-}$ according to $p^{\delta}(x_i|\mathbf{X}_{T_{j+1}^-})$;
    set $\mathbf{X}_{T_{j+1}} = \mathbf{X}_{T_{j+1}^-} \setminus (z,m)$;
  else
    sample $x_i = (z_i, m_i) \in \mathbf{X}_{T_{j+1}^-}$ and $m \in \mathbb{M}$ according to $p^{\tau}((z_i,m_i),m|\mathbf{X}_{T_{j+1}^-})$;
    set $\mathbf{X}_{T_{j+1}} = (\mathbf{X}_{T_{j+1}^-} \setminus (z_i,m_i)) \cup (z_i,m)$;
  end
end

[7] Briane, V., Vimond, M., and Kervrann, C. 细胞内动力学分析中扩散模型的综述。*Briefings in Bioinformatics*（2019年）。bbz052。

[8] Chenouard, N., Smal, I., De Chaumont, F., Maška, M., Sbalzarini, I. F., Gong, Y., Cardinale, J., Carthel, C., Coraluppi, S., Winter, M., et al.粒子追踪方法的客观比较。*Nature Methods 11*，3（2014年），281–289。

[9] Costes, S., Daelemans, D., Cho, E., Dobbin, Z., Pavlakis, G., and Lock-ett, S. 活细胞中蛋白质-蛋白质共定位的自动定量测量。*Biophysical Journal 86*，6（2004年），3993–4003。

[10] Hozé, N., and Holcman, D. 细胞生物学中超分辨率随机单粒子轨迹大型集合的统计方法。*Annual Review ofStatistics and Its Application 4*（2017年），189–223。

[11] Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S.,Schmid, S. L., and Danuser, G. 活细胞时间序列成像中的稳健单粒子追踪。*Nature Methods 5*,8 (2008年),695.

[12] Kervrann, C. 生物分子运输与基于网络断层成像的模拟。载于《生物医学图像合成与仿真》。爱思唯尔，2022年，第543–569页。[13] Kervrann, C., Sorzano, C. Ó. S., Acton, S. T., Olivo-Marin, J.-C., andUnser, M. 荧光显微镜与电子显微镜图像处理和分析方法精选导览。IEEE信号处理选题期刊 10, 1 (2015年), 6–30。[14] Lagache, T., Dauty, E., and Holcman, D. 病毒和质粒在细胞中的运输定量分析。《物理评论E》79, 1 (2009年), 011921。[15] Lagache, T., Sauvonnet, N., Danglot, L., and Olivo-Marin, J.-C. 生物成像中分子共定位的统计分析。《细胞计量学A辑》87, 6 (2015年), 568–579。[16] Lagardère, M., Chamma, I., Bouilhol, E., Nikolski, M., and Thoumine,O. Fluosim：用于荧光活细胞成像和膜蛋白超分辨成像的单分子动力学模拟器。《科学报告》10, 1 (2020年), 1–14。[17] Lavancier, F., Le Guével, R., and Manent, E. 跳跃-移动过程的费勒性质与遍历性质及其在相互作用粒子系统中的应用。arXiv预印本 arXiv:2204.02851 (2022年)。[18] Lavancier, F., and Le Guével, R. 空间出生–死亡–移动过程：基本性质及其强度函数估计。《英国皇家统计学会会刊：B辑（统计方法学）》83, 4 (2021年), 798–825。[19] Lavancier, F., Pécot, T., Zengzhen, L., and Kervrann, C. 检验两个随机集之间的独立性，以分析生物成像中的共定位。《生物统计学》76, 1 (2020年), 36–46。[20] Øksendal, B. 随机微分方程：简介与应用。施普林格科学与商业媒体，2013年。[21] Pécot, T., Bouthemy, P., Boulanger, J., Chessel, A., Bardin, S., Salamero, J., and Kervrann, C. 利用条件随机场进行活细胞成像中的背景荧光估计与囊泡分割。IEEE图像处理汇刊 24, 2 (2014年), 667–680。[22] Pécot, T., Kervrann, C., Bardin, S., Goud, B., and Salamero, J. 用于荧光生物成像事件检测的基于图块的马尔可夫模型。载于《国际

医学图像计算与计算机辅助干预会议 (2008年)，施普林格，第95–103页。

[23] Pécot, T., Zengzhen, L., Boulanger, J., Salamero, J., and Kervrann, C.一种定量方法，用于分析荧光显微镜下细胞内3D事件的时空分布。*eLife 7* (2018), e32311。[24] Roudot, P., Ding, L., Jaqaman, K., Kervrann, C., and Danuser, G.用于追踪密集环境中异质粒子运动的分段平稳运动建模与迭代平滑。*IEEE*图像处理汇刊 *26*，11（2017年），5395–5410。