嵌入时延的选择：嵌入技术概述及一种使用持久同调的新方法。

Eugene Tan, [1,a)] Shannon Algar[1]、Debora Corrˆea[1,2]、Michael Small[1,3]、Thomas Stemler[1,3,4] 和 David Walker[1]

[1)]*Complex Systems Group, Department of Mathematics and Statistics, The University of Western Australia,Crawley, Western Australia 6009, Australia*

[2)]*Forrest Research Foundation, The University of Western Australia, Crawley, Western Australia 6009,Australia*

[3)]*ARC Centre for Transforming Maintenance Through Data Science, The University of Western Australia, Crawley, Western Australia 6009, Australia*

[4)]*Mineral Resources, CSIRO, Kensington, Western Australia 6151, Australia*

（日期：2023年2月8日）

时延嵌入方法是时间序列分析与预测领域中的常用工具。然而，嵌入参数的选择可能会对最终分析产生重大影响。因此，人们提出了大量方法，用于优化嵌入时滞等参数的选择。本文旨在为初次接触该主题的读者全面概述嵌入理论的基础知识。我们概述了在均匀延迟嵌入和非均匀延迟嵌入情况下选择嵌入时滞的一系列现有方法。针对现有非均匀滞后选择方法在动力学可解释性方面的不足，我们提出了一种替代方法来选择嵌入滞后，该方法结合了动力学论证和拓扑论证。所提出的方法——持久链上的显著时间（SToPS）——利用持久同调构建特征时间谱，以量化每个时间滞后的相对动力学显著性。我们在周期性、混沌和快慢时间序列上测试了该方法，发现其表现与现有自动非均匀嵌入方法相近。此外，在使用 SToPS 构建的嵌入上训练的 n 步预测器，在预测快慢时间序列时表现优于其他嵌入方法。

arXiv:2302.03447v1 [math.DS] 7 Feb 2023

嵌入方法常用于分析无法完整或直接观测到系统完整状态的时间序列。然而，大多数嵌入方法都需要仔细选择参数，才能忠实地重构系统动力学。一类常见的嵌入方法——时间延迟嵌入——需要仔细选择嵌入滞后。本文概述了嵌入理论，并汇集了现有方法和原则，以指导嵌入滞后的选择。最后，我们阐述了采用非均匀嵌入的理由，并提出一种基于持久同调的新方法 SToPS，用于指导多个嵌入滞后的选择。

I. 嵌入的依据

自惠特尼[1] 和 Takens[2], 提出具有重要意义的结果以来，数学嵌入的思想几乎渗透到非线性动力学文献的所有领域。随后，Sauer、Yorke 和 Casdagli[3]在开创性论文《嵌入学》中对相关定理进行了形式化、系统整理和讨论。这为多种嵌入技术的发展铺平了道路，例如导数法[4]，时间延迟嵌入[2,5] 和主成分分析嵌入[6] 等；这些技术后来被应用于广泛的研究领域。如今，嵌入方法仍是非线性时间序列分析研究中不可或缺的工具。

时间序列 $x(t) \in \mathbb{R}^m$ 通常可视为数据生成过程的产物：该过程通过测量函数 $h : \mathbb{R}^n \to \mathbb{R}^m$，对具有状态 $\vec{s}(t) \in \mathbb{R}^n$ 的某个动力过程进行连续观测，尽管观测并不总是规则的。通常， $m < n$ 无法观测到底层动力过程的完整状态。为便于说明，本文将考虑最简单的情况，即 $x(t)$为标量（即 $m = 1$）。

时间序列分析的主要目标通常分为两大类：系统辨识或分类，以及预测。前者旨在刻画并理解底层动力系统的动力学特性和运行机制。研究范围可以从李雅普诺夫谱[7,8] 和关联维数[9–11] 等系统不变量，延伸至分岔分析[11,12]。后者即时间序列预测，目标则更具实用性，正如其名称所示：“给定一段历史$x(t_0)...x(t)$，找出对 $x(t+\tau)$ 的最佳预测估计”。值得说明的是，这两个领域并非互斥，且往往相辅相成。不过，本文将聚焦于后者，即时间序列预测问题。

尽管目标简单，但在实践中完成时间序列预测任务仍面临诸多挑战。

[a)]通讯作者，eugene.tan@uwa.edu.au

除了对简单玩具模型的研究之外，许多有趣的系统还表现出高维特性和混沌行为[13–15]。对于时滞系统，系统动力学的维数甚至可能不是有限的。无法获取完整系统状态也进一步增加了挑战。因此，通常会借助测量函数 h 来重新表述时间序列预测问题： $\mathbb{R}^n \to \mathbb{R}^m$ 其中 $m < n$。这样，时间序列预测就可以改写为：“给定一段观测历史 $h(x(t_0))$ ..., $h(x(t))$，预测未来状态 $x(t+\tau)$”。我们可以用概率论来表述这一问题，目标是计算条件概率 $P(x(t+\tau)|h(x(t_0))$ ..., $h(x(t)))$。

鉴于观测限制，任何时间序列预测器都应尽可能提取并利用观测时间序列 $x(t)$中包含的信息。这正是嵌入的价值所在，也就是通过维度增广来实现这一点。幸运的是，在精度足够且无噪声的假设下，Takens嵌入定理保证：由标量观测时间序列构造、在维度足够高的空间中定义动力学 Φ 的时间延迟嵌入 $\mathbb{R}^d$ 通常与底层动力系统的完整状态空间动力学[2]微分同胚。从本质上说，仅利用观测系统变量的一部分，就可以重构完整状态动力学的一个代理像。通常的做法是，将观测时间序列嵌入到一个维度足够高、包含“完整”状态动力学的空间中，使用偏好的建模工具（例如神经网络[16,17], 储备池计算[18–20] 支持向量机[21,22] 等）学习状态空间动力学或向量场，然后向前预测，并在新的嵌入空间中完成所有必要的计算。

尽管嵌入定理具有优雅性和实用性，但其对无限数据精度、无限数据长度以及无噪声信号的依赖，在实践中带来了问题。嵌入的有效性在很大程度上取决于用于扩充数据的嵌入参数[23,24]。例如，在时间延迟嵌入中，时间滞后[25] 和嵌入维数[3,26] 的选择会对所得重构的质量产生深远影响。

因此，我们可以将嵌入的主要挑战概括为以下问题：“我们如何选择良好的嵌入参数？”这将是我们讨论的重点。人们提出了许多方法来解决这一问题，其范围从纯粹的统计和动力学论据（如互信息[11,25]和连续性统计量[27], ）到更为纯粹的拓扑学论据（如畸变[28], 、噪声放大[28,29] 和填充因子[30]）。在许多情况下，这些方法中的每一种只对特定类型的系统表现良好，而对其他系统则不然。

本文有三个主要目标。首先，简要概述选择合适嵌入参数所面临的挑战。其次，汇总并比较动力学与拓扑之间各种广受欢迎的方法，这些方法旨在解决嵌入参数选择问题。本文将重点关注优化时间延迟嵌入这一具体情形。最后，介绍一种基于蓬勃发展的持久同调领域的新方法——显著性评分；该方法试图在选择嵌入参数时兼顾动力学和拓扑论证。

本文结构如下。我们首先在第II节概述嵌入理论和若干常见的嵌入方法。随后在第III节简要讨论嵌入参数选择的指导原则。第IV至V节介绍了几种适用于均匀和非均匀延迟嵌入的嵌入参数选择方法。最后，我们在第VIII节介绍了SToPS，这是一种基于持久同调的嵌入参数选择方法，也是我们对嵌入参数选择问题的贡献。

II. 嵌入方法

有多种嵌入方法可用于状态空间重构。不过，它们的目标都大致相同：将观测到的时间序列扩展到足够高的维度，以便描述潜在动力学。为完整起见，我们简要介绍嵌入及三种常见的嵌入方法：时间延迟嵌入、导数嵌入和全局主值嵌入。不过，本文的重点是时间延迟嵌入参数的选择。关于其他嵌入方法的更深入讨论，可参见其他文献[28]。

A. 嵌入理论

对于一个给定的动力系统，其状态为 $\vec{s}(t)$，在状态空间 $S \subseteq \mathbb{R}^d$ 上的动力学由演化算子 f 描述，满足：

$$\vec{s}(t+T) = f_T(\vec{s}(t)), \tag{1}$$

我们可以定义一个测量函数 h: $\mathbb{R}^d \to \mathbb{R}$ ，它模拟观测系统并提取信息的过程，在给定时间步评估这些信息，从而生成时间序列：

$$x(t) = h(\vec{s}(t)). \tag{2}$$

嵌入可以定义为一种变换 Ψ:$\mathbb{R} \to \mathbb{R}^m$ ，它利用某个时间窗口 $[t_1, t_2]$ 内观测到的值来扩展时间序列的维度。例如，均匀延迟嵌入定义如下：

$$\Psi(x(t)) = \vec{x}(t) = (x(t), x(t-\tau), ...), \tag{3}$$

其中 $\vec{x}(t)$是嵌入向量，其动力学定义在重构状态空间 $X \subseteq \mathbb{R}^m$ 中，并采用经过变换的演化算子 F。如果 Ψ 是有效嵌入，

Takens嵌入定理保证，一般而言，存在一个微分同胚 Φ: $S \to X$，它保留该系统的动力学，使得，

$$F = \Phi \circ f \circ \Phi^{-1}. \tag{4}$$

沿重构状态空间学习动力学 F，等价于学习真实系统的动力学 f （见图 1）。因此，利用嵌入进行时间序列预测的任务，简化为学习演化算子 F，其中

$$x(t+T) = \Psi^{-1} \circ F_T \circ \Psi(x(t)). \tag{5}$$

B. 时间延迟嵌入

时间延迟嵌入最早由 Packard等人[4], 提出，它通过构造延迟向量 $\vec{x}$（t），将标量时间序列x（t）扩展到更高维度，表示为

$$\vec{x}(t) = (x(t), x(t-\tau), ..., x(t-(m-1)\tau)), \tag{6}$$

需要选择的嵌入参数是延迟间隔 τ 和嵌入维数 m。根据 Takens 定理的保证，只要 m 足够大且测量值具有无限精度， τ 的任意取值都能产生有效嵌入。然而，这在实践中无法实现，而且对延迟间隔和嵌入维数的不同选择可能会产生不同的结果。III A 节对此作了进一步讨论。

我们还要指出，选择理想的延迟间隔 τ和嵌入维数 m 并非时间延迟嵌入所独有。在其他时间序列分析方法中，例如排列熵[31] 和序数划分网络[32,33]，选择 τ 和 m 也是关键决策。在这两种情况下，都会构造一个延迟向量，并根据各分量的大小顺序对其进行编码表示。随后，可以将时间序列视为不同编码状态之间的转换，并用于进一步分析。

许多自动化时间序列预测方法，例如循环神经网络[16,17] 和储备池计算[18–20, 34] 也可能与延迟嵌入有关。在这两种情况下，输入时间序列都会输入到一个具有某种记忆能力的动力网络中。这种记忆对过去状态的前向传播会影响未来状态，实际上等同于一种延迟间隔较小 τ 、嵌入维数较大的时间延迟嵌入m。

C. 导数嵌入

导数嵌入方法利用逐阶递增的观测时间序列的时间导数[4]来重构嵌入向量。其表达式为:

$$\vec{x}(t) = \left(x(t), \frac{dx(t)}{dt}, ..., \frac{d^m x(t)}{dt^m}\right) \tag{7}$$

导数通过数值近似求得。如果能够准确计算所需的导数，那么对于足够大的 m ，导数嵌入方法就是有效的嵌入方法。

D. 积分-微分嵌入

导数嵌入方法的一个缺点是需要根据数据计算数值导数。虽然这对于一阶导数或许可以接受，但随着信号噪声比往往受到负面影响，对后续高阶导数的近似通常并不准确。即使对于非常干净的数据集，情况也是如此。

导数嵌入的一种替代方法是积分-微分嵌入[35]。该方法将二阶导数替换为积分，从而避免计算连续的高阶导数。由此可构造出以下嵌入:

$$\vec{x}(t) = \left(\int_{-\infty}^{t} x(t) - \langle x(t) \rangle_t \, dt, x(t), \frac{dx(t)}{dt}\right), \tag{8}$$

其中，先将第一个分量设为零均值，再进行积分。采用一阶积分和数值导数，使第一和第三个嵌入分量的信噪比分别仅下降一个数量级。相比之下，在导数嵌入方法中，随着数值导数阶数的增加，每次连续求导都会使信噪比进一步降低。不过，对于更高维的嵌入，积分-微分嵌入方法也会受到与纯导数方法相同的噪声效应影响。这限制了该方法在系统动力学被认为具有高维特性的系统中的适用性。

E. 全局主值嵌入

主值嵌入方法由 Broomhead 和 King[6] 提出，是一种对 Takens 定理所采用的时间延迟嵌入方法进行改进的替代方案。该方法借鉴了主成分分析[36] 的思想，旨在为维数足够高的时间延迟嵌入寻找理想的旋转。给定一个长度为 N_T 的时间序列 $x(t)$，以及一个长度为 M的滑动窗口，我们可以构造一组

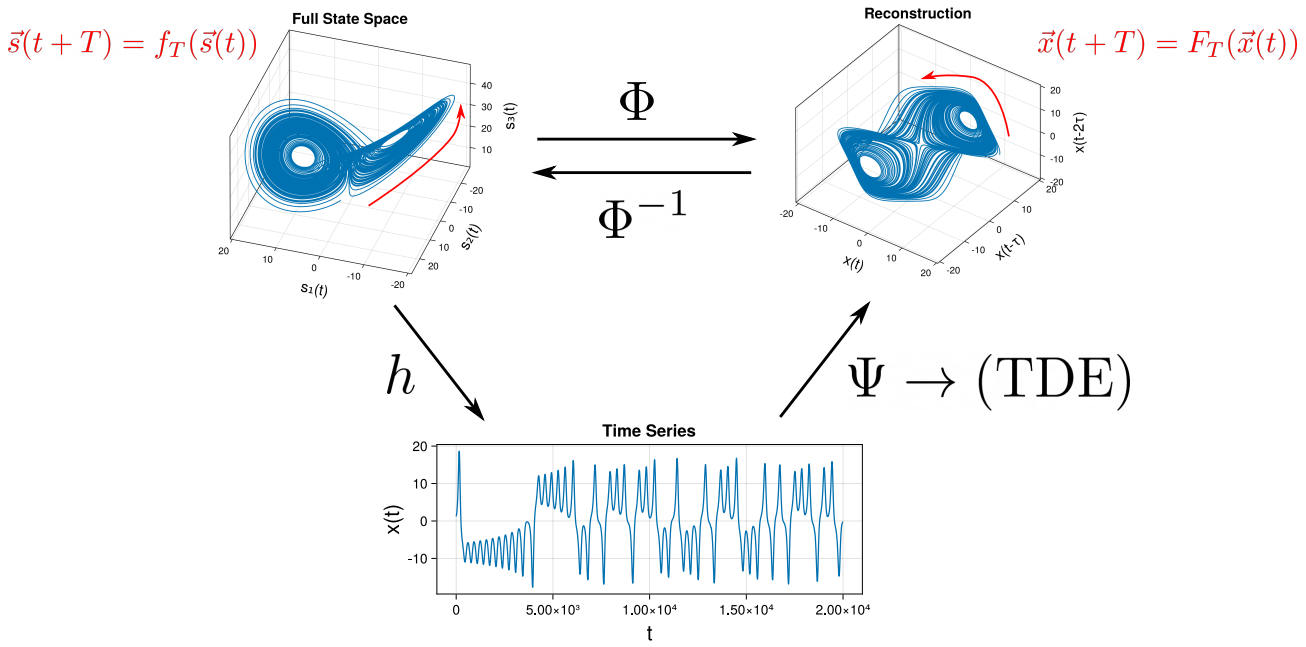


图 1. 嵌入过程及其组成部分之间关系的示意图 s.

$N = N_T - (M-1)$ 个延迟向量，

$$\mathbf{X} = N^{-1/2} \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_N \end{bmatrix} \tag{9}$$

其中，$\vec{x}_i$ 是以时间序列中的 i^{th}值作为第一个分量构造的延迟向量，

$$\vec{x}_i = (x(t_i), x(t_{i-1}), ..., x(t_{i-(M-1)})). \tag{10}$$

可由$\mathbf{X}$计算得到一个 $m \times m$ 协方差矩阵 C。该矩阵的元素 C_{ij} 可简单表示为：

$$C_{ij} = \langle x(t)\, x(t + (i-j)) \rangle_t \tag{11}$$

其中，$\langle ... \rangle_t$ 表示时间平均值。然后，通过计算 C 相应的特征值和特征向量，得到其主成分。选取与所需嵌入维数相对应的前 d 个主成分后，可利用特征向量矩阵计算$\mathbf{X}$ 的投影，从而得到最终的嵌入坐标。读者可参阅其他资料[6,28] 了解更多详情。

主成分值嵌入的核心目标，是将高维延迟嵌入（通常通过取大量滞后分量获得）提炼并简化到一个低维子空间中。有人认为，其余子空间对应的是动力学变化和重要性都很小的分量方向。这种方法的一种应用，是尝试简化最优嵌入维数的选择[6]，，其中理想的嵌入维数 m 对应于显著高于某个“噪声底限”的奇异值数量。然而，这种方法受到了一些批评[37,38]。其中最主要的批评认为，噪声底限开始形成平台，可能归因于数据的精度和噪声强度，而非相应特征向量方向的重要性。

在嵌入的一般研究背景下，Paluˇs 和 Dvˇor´ak 使用前 k 个主特征向量方向[38]，检验了降维后延迟嵌入的质量。作者指出，降维嵌入对关联维数等动力学不变量的估计，会随时间延迟和分量数量的不同而变化。他们认为，协方差矩阵的主成分仅适用于线性相关。因此，尽管嵌入中的分量在线性意义上可能彼此独立，但它们仍可能存在非线性依赖。相反，截断嵌入维数可能导致重要的非线性分量被排除。

在给定主分量方向中纳入较大的时间滞后，对于混沌系统而言可能也没有多大意义，因为在时间上相近的观测值会随时间呈指数级去相关。从选择时间滞后的角度来看，每个主成分几乎无一例外地都会包含来自所有 M 可能的滞后分量的贡献。除去降维这一论据[38]（对其解释必须谨慎），全局主值嵌入与具有较大嵌入维数 m的一般时间延迟嵌入并没有显著区别。

III. 嵌入考量

A. 嵌入质量

如前所述，Takens 的理论保证在有限精度和噪声[23,25] 存在时会失效，从而产生了“最优”嵌入参数这一概念。此类“最优”参数集的存在意味着，并非所有嵌入都具有相同的质量。然而，这需要有一种度量，用于比较嵌入质量。Casdagli[28]和 Potapov[39]对量化嵌入质量进行了深入研究。

嵌入质量的定义差异很大，例如基于信息论论证[25]，预测任务[28] 和吸引子拓扑[30,40]的定义。值得注意的是，一组参数的“最优性”取决于嵌入的用途。因此，尽管基于不同嵌入质量概念的方法可能会得出相似结果，但结果之间很可能始终存在分歧[11,23,28]。因此，除非数据本身存在支持这一判断的动力学和拓扑原因，否则最好不要声称某一组参数更为有利。不过，本节将重点介绍定义嵌入质量时通常考虑的一般因素。

嵌入质量的不同概念可归纳为两大类或论证：基于预测的论证和拓扑论证。基于预测的嵌入质量概念可视为受到嵌入在时间序列预测中的应用启发。从根本上说，良好的嵌入应能实现更准确的预测[28,39]。

在时间序列预测任务中，未知的测量函数 h以及含噪数据会给推断真实系统状态 $s(t)$带来一定程度的不确定性。Casdagli 指出，理想的嵌入应尽可能降低根据重构状态空间中的位置 $\vec{x}(t)$推断真实状态 $\vec{s}(t)$时的不确定性。本质上，在存在噪声的情况下，对构造出的延迟嵌入应用逆变换 Φ^{-1}时，真实系统状态应几乎不存在歧义，前提是能够完全了解 $\vec{s}(t)$[28]。从理论上说，这种抗噪性和低歧义性有利于时间序列预测，也是用于量化嵌入质量的噪声放大和畸变概念的基础。在质量较差的嵌入中，例如吸引子流形呈层状、彼此间隔很小的情况，噪声扰动跨越流形层所产生的影响会导致真实状态 $s(t)$存在显著的不确定性，从而使时间序列预测变得困难。

用于选择嵌入参数的 信息论论证（例如自相关、最小互信息、连续性统计）也与预测思想密切相关。这些方法通常试图最大限度地增加每个延迟维度所包含的新信息，以期提供更多关于真实系统状态$\vec{s}(t)$的信息，并有助于时间序列预测。

定义嵌入质量的另一大类方法基于拓扑和几何论证。其中许多方法着重研究吸引子结构以及流形在其环境状态空间中的分布。本质上，从拓扑和几何角度看，优质嵌入应在其环境空间中分布均匀且充分展开[11,30,40]。这一质量概念与 Casdagli 关于噪声放大的论证有相似之处。基于几何论证的方法包括填充因子[30] 和偏离对角线距离[40]等统计量。最终，上述许多用于确定时间延迟嵌入理想时滞和嵌入维数的考量，都可以用无关性和冗余[28,41]这两个概念来概括。

B. 无关性和冗余

选择时间滞后 τ 和嵌入维数m 是构建时间延迟嵌入时面临的主要挑战。嵌入时滞与嵌入维数的相对重要性尚不明确。此外，也有研究提出，这些嵌入参数可能并非相互独立。相反，有人提出，应将称为嵌入窗口的量 $\tau_w = m\tau$ 视为更重要的优化参数[28,41]。然而，当嵌入维数足够大时，可以认为，选择 τ_w 或许可以简化为恰当地选择 τ。

嵌入窗口 τ_w （进而也包括嵌入时滞 τ）的选择可以概括为一个原则：既不能太短（冗余），也不能太长（无关性）。这一解释适用于混沌或非周期信号。然而，周期信号可以通过较大的时滞 τ 成功嵌入，其中有效时滞 τ^* 与周期 T相关，

$$\tau = nT + \tau^*, \quad n \in \mathbb{N}. \tag{12}$$

冗余度较高的嵌入会产生彼此大致平行、呈层状分布的轨迹（例如靠近对角线）。存在足够噪声时，各层之间清晰的分隔会受到影响。因此，在重构的状态空间中，给定某个含噪观测 $\vec{x}$（t），对真实系统状态 $\vec{s}$（t）的不确定性会更大。

同样，具有高度无关性的嵌入包含与真实状态高度不相关的分量[28]。这也不理想，因为它可能会在重构流形中的轨迹之间引入不必要的交叉。这正是选择嵌入窗口 τ_w 时面临的金发姑娘问题：窗口既不能太大，也不能太小。

IV. 均匀延迟嵌入

最简单的延迟嵌入形式是均匀延迟，即为 τ 和m 选取单一的常数值。在这种情况下，嵌入向量按照均匀递增的时间延迟进行选取，如公式13所示，

$$\vec{x}(t) = (x(t), x(t-\tau), ..., x(t-(m-1)\tau)). \tag{13}$$

由于τ 和 m的选择优先级存在争议，人们提出了多种方法来同时估计这两个参数。一些方法包括 Gao and Zheng 的特征长度[42,43]，[23] 和 Schuster 的摆动乘积[44]。第 V 节概述了这些方法。

其他常见方法则试图通过假设 τ和 m 相互独立来简化问题，并选择分别估计这两个值。通常，首先使用多种度量方法之一确定嵌入 τ。选定后，构造维度逐渐增加的均匀延迟向量，并使用 Grassberger-Procaccia[10] 算法或伪近邻[26,45,46] 等算法进行检验，直至达到收敛。达到稳定时延迟向量的长度用于确定嵌入维数 m。

用于确定嵌入时滞 τ 的方法从简单的启发式方法到更复杂的统计论证不等。一种常见的启发式方法是将嵌入时滞选为信号周期（或混沌信号的准周期）的四分之一。深入探讨统计依据更充分的论证时，偶尔会使用自相关[11]及其非线性推广——互信息[47]、连续性统计量[27] 和 L统计量[29]，以确定 τ的合适取值。第 VI 至 VII 节对这些方法进行了全面概述和进一步讨论。

V. 均匀嵌入参数的同步优化

与许多需要独立选择 τ 和 m 的现有方法不同，已有若干嵌入方法提出使用单一指标同时估计这两个值。该指标通常在多个时滞上计算，并针对逐渐增大的 m重复计算。首先根据某种准则选择 m 的取值，然后再选择 τ。Celluci[23]详细概述了这些方法的实施步骤。

A. Gao and Zheng 方法与特征长度

Gao and Zheng 提出的嵌入方法基于假近邻的出现情况[42,43]。伪近邻可归因于冗余（展开不足）或无关性（吸引子中的虚假交叉）。Gao and Zheng 提出的方法基于这样一种观点：在理想的嵌入中，应尽可能减小分离距离和伪近邻的比例。

考虑嵌入空间中的一对点 $\vec{x}_i$，$\vec{x}_j$ 以及它们在未来 k 步后的演变 $\vec{x}_{i+k}$，$\vec{x}_{j+k}$。在时间延迟嵌入中，随着吸引子的展开，伪近邻点往往比真正的邻居点分离得更快。因此，对于伪近邻点对，其距离之比 $|\vec{x}_{i+k} - \vec{x}_{j+k}|/|\vec{x}_i - \vec{x}_j|$ 会更大；而对于真正的邻居点，该比值约等于 1。随后，Gao and Zheng 提出以下指标 Λ 来优化嵌入参数：

$$\Lambda(k, m, \tau) = \frac{1}{N_{ref}} \sum_{i,j} \ln \frac{|\vec{x}_{i+k} - \vec{x}_{j+k}|}{|\vec{x}_i - \vec{x}_j|}, \tag{14}$$

其中，N_{ref} 表示随机抽取的点对数量，距离比值在这些点对上取平均。点对的选取还受到若干其他限制 $\vec{x}_i$，$\vec{x}_j$。首先，这些点的初始间距应满足 $|\vec{x}_i - \vec{x}_j| \leq r$ 其中 r 是选定的一个较小阈值，即点的初始间距应足够小，使得计算间距增长具有意义。其次，所选点对的 Theiler 窗口不能相交$|i-j| > l_{\text{Theiler}}$，其中 $l_{\text{Theiler}} \in \mathbb{N}^+$。这样做是为了避免同一局部轨迹上的点之间产生不需要的相关性[48,49]。最后，常数 k 不能太大，且应根据系统动力学的自然时间尺度来选取。

为确定合适的嵌入参数，针对不断增大的嵌入维数 m，计算$\Lambda(\tau)$的曲线。选择曲线 $\Lambda(\tau)$中下降幅度最大的 m 值作为嵌入维数。随后，将嵌入时滞 τ 选为 $\Lambda(\tau)$的第一个极小值。

B. 特征长度

特征长度方法是 Gao and Zheng[23] 方法的扩展，旨在解决演化时间 k 的选择问题。它不再任意选择 k，而是计算一个特征长度 $J(m, \tau)$，用于描述系统吸引子的自然空间尺度，

$$J(m, \tau) = \langle |\vec{x}_i - \vec{x}_j| \rangle, \tag{15}$$

其中 $\langle ... \rangle$ 表示对吸引子采样点对的平均值。随后利用特征长度计算分离时间 $T_J(\vec{x}_i, \vec{x}_j)$，其定义为最近邻点对分离到特征长度 $J(m, \tau)$ 的某个比例所需的时间。对于真实的邻居点，随着嵌入维数增加，T_J 将收敛到与系统李雅普诺夫指数相关的某个值

，而错误的最近邻将导致较小的值 T_J ，因为轨迹会迅速分离。用于确定嵌入参数的新度量如下：

$$C(m,\tau) = \frac{1}{N_{ref}} \sum_{i,j} T_J(\vec{x_i}, \vec{x_j}), \tag{16}$$

其中， N_{ref} 是近邻采样点对的数量。使$C(m,\ \tau)$达到最大值的 m 和 τ 取值，分别选作嵌入维数和时滞。

C. 摆动乘积

摆动乘积[44] 与 Gao and Zheng 的方法以及特征长度相似，因为它们都基于最近邻的概念。作者提出，良好的嵌入应保留给定参考点附近邻居点的顺序对应关系（即，应保留按与某个参考点 $\vec{x}_i$ 的距离排序后的邻居点顺序）。具体做法是比较点 $\vec{x}_i$ 在给定嵌入 Φ_k中的 N个近邻的顺序，其有序序列为：

$$X_{\Phi_k} = \{\vec{x}_{i,1}, ..., \vec{x}_{i,N}\}_{\Phi_k}, \tag{17}$$

以及将其投影到下一个阶次的嵌入 Φ_{k+1}（通过增大 m 或 τ）后所得的序列：

$$Z_{\Phi_{k+1}} = \{\vec{z}_{i,1}, ..., \vec{z}_{i,N}\}_{\Phi_{k+1}}. \tag{18}$$

这里，$\vec{x}_{i,\ n}$ 对应于参考点 $\vec{x}_i$的第 n^{th} 个最近邻i^{th} 。投影 $\vec{z}_{i,\ n}$ 对应于同一个邻居数据点$\vec{x}_{i,\ n}$ ，其位置根据下一个阶次的嵌入 Φ_{k+1}重新计算。

同样，也可以通过降低 m 或 τ，将其投影到阶数较低的嵌入中，从而得到一组新的有序点，

$$V_{\Phi_{k-1}} = \{\vec{v}_{i,1}, ..., \vec{v}_{i,N}\}_{\Phi_{k-1}}. \tag{19}$$

理想情况下，良好的嵌入应在这些有序序列之间保持一一对应关系。这样，以下比值将等于 1，

$$\frac{|\vec{x}_i - \vec{z}_{i,n}|}{|\vec{x}_i - \vec{x}_{i,n}|}, \text{ and } \frac{|\vec{x}_i - \vec{x}_{i,n}|}{|\vec{x}_i - \vec{v}_{i,n}|}. \tag{20}$$

Schuster 和 Liebert 提出的方法将上述两个比率的乘积作为以下度量，

$$W_i(m\tau) = \prod_{n=1}^{N} \left\{ \left(\frac{|\vec{x}_i - \vec{z}_{i,n}|}{|\vec{x}_i - \vec{x}_{i,n}|} \right) \cdot \left(\frac{|\vec{x}_i - \vec{x}_{i,n}|}{|\vec{x}_i - \vec{v}_{i,n}|} \right) \right\}. \tag{21}$$

需要优化的度量定义为在 N_{ref} 个随机采样的参考点上取平均值，

$$W(m,\tau) = \ln \left(\frac{1}{N_{ref}} \sum_{i=1}^{N_{ref}} W_i(m,\tau) \right) \tag{22}$$

其中， m 被选为使 W 达到极限行为的维度，而 τ 对应于所得曲线的第一个极小值。

VI. 均匀嵌入的选择 AG

与第 V 节中利用单一指标来优化嵌入时滞和维度选择的嵌入方法不同，目前最常用的方法仍是依据不同指标，分别选择嵌入时滞 τ 和维度 m。本文将重点讨论用于选择时间延迟嵌入时滞的方法。随后可使用该时滞构造任意维度的延迟嵌入。

对于较简单的均匀延迟嵌入，需要选择 τ 的一个单一值，并以其整数倍构造所需的延迟向量。用于指导延迟间隔 τ 选择的方法，反映了 III A 节所述关于嵌入质量的两种主要观点。本节将各种优化嵌入时滞的方法分为两类：第一类基于动力学和信息论论证，第二类基于拓扑论证。我们还将在第 VII 节讨论非均匀嵌入。

基于动力学方法的方法可以理解为关注数据生成过程背后的机制，以及测量值之间的统计关系。信息论方法和统计方法也属于这一类。我们将在这一类中介绍的方法包括自相关、最小互信息和四分之一周期。这与第二类方法形成对比，后者的方法更偏向拓扑学。例如，填充因子[30] 和噪声放大[28,29]。

关于嵌入维数的简要说明：有几种方法可用于确定所需的嵌入维数，例如伪近邻（FNN）[26,45] 以及用于估算关联和与关联维数的 Grassberger-Procaccia算法[10]。其他类似于关联维数的不变量，例如 Kaplan-Yorke维数[50] 和盒计数维数[51] ，也常用于确定嵌入维数。这些结果通常与惠特尼定理 d结合使用；该定理指出，任何 d 维流形（例如吸引子）都可以嵌入至少 $m > 2d + 1$ 维的空间中[1]。然而，有人指出，直接应用该定理存在缺陷，因为惠特尼定理仅针对整数维数 d 得到了证明，而对于大多数

感兴趣的系统（例如表现出分形和混沌行为的系统）而言，维数通常并非整数。不过，我们注意到，Sauer 等人[11,52]将惠特尼定理推广到盒计数维数（$m > 2D_F$），使该不等式适用于一般情形。

A. 动力学方法

1. 自相关和最小互信息

许多常被分析的动力系统往往表现出混沌行为，其中相近的轨迹会迅速发散，并很快变得不相关。自相关方法基于这样一种思想：重构中的每个分量都应尽可能包含关于真实状态的新信息。有人提出，延迟向量中的各分量应力求最小化相关性[11]。这与最小化重构的冗余具有类似效果。

对于延迟嵌入，取自相关函数第一个极小值所对应的时滞作为嵌入时间延迟。另一种方法是采用自相关信号衰减至 $1/e$所需的时间[11]。此外，还有一种基于平均局部自协方差第一个根节点的变体方法，被提出作为最小自相关方法的一种稳健替代方案[53]。

自相关的一个缺点是无法处理时间序列中的非平稳性（例如相位、频率和幅值的漂移）。此外，其应用仅限于线性信号[54]。除最简单的情况外，动力系统都会表现出一定程度的非平稳和非线性行为。Fraser 和 Swinney[25] 提出，应使用系统之间的互信息来替代自相关。在他们的原始论文中，Fraser 和 Swinney 首先给出了几何解释，以补充关于互信息的理论论证。具体而言，考虑一组点，其中某一分量的值 x 位于某个固定窗口内。从这组点出发，跟踪它们在未来τ 步后的位置，并计算同一组点在同一分量中的值 p_τ（x）的分布。若取值为 τ 时得到更宽的分布p_τ（x），则应对应于一个较好的时滞，同时也对应于较小的互信息值。

互信息 $I(\tau)$ 可视为自相关函数的非线性对应形式，

$$I(\tau) = \int P(x(t), x(t+\tau)) \log_2 \left(\frac{P(x(t), x(t+\tau))}{P(x(t))P(x(t)+\tau)} \right) dt,$$

其中， $P(x(t))$ 表示在任意给定时刻观测到状态 $x(t)$ 的概率，而 $P(x(t)\ x(t+\tau))$ 则是以相同方式定义的、针对时刻 t 和未来时刻$t+\tau$的联合概率。借鉴信息论，互信息 $I(\tau)$ 旨在量化某一时刻的观测中所包含的关于未来状态的信息量即关于时刻 $t+\tau$的未来状态、包含在时刻 t的观测中的信息。对于给定时滞 τ，较高的互信息意味着状态之间具有较强的相关性，并会导致延迟重构中的冗余度更高。

最小互信息和自相关的优势在于，它们能够以相对简单、快速的计算为时滞提供合理估计。然而，对于给定的互信息曲线 $I(\tau)$[55]，并不能保证一定存在明确的极小值。此外，计算互信息需要对概率密度函数进行数值估计 $P(x(t))$ 和 $P(x(t)\ x(t+\tau))$，因此还需考虑直方图箱宽和数据长度要求的最优设置[56,57]。 人们提出了许多能够更有效估计互信息的替代方法，包括采用自适应分箱[25,58], 核密度估计器[59] 和 k- 最近邻[56]。

2. 四分之一周期

一种常用的启发式方法是，将嵌入时滞 τ 设为信号[60]中最主要周期的四分之一。这种方法能够在嵌入过程中编码系统动力学的自然时间尺度。这一启发式方法源自将正弦波嵌入二维空间的问题 $x(t) = \sin(\omega t)$。在这种情况下，$\tau = \frac{2\pi}{4\omega}$ 会产生最接近圆形的二维延迟嵌入；取其他值则会形成椭圆形轨迹。然而，这种启发式方法不能直接用于信号非周期性的混沌系统。相反，需要估计某种形式的伪周期，这也是本文所提出方法的重点。

B. 拓扑方法

1. 填充因子

填充因子方法最初由 Buzug 等人提出[30] ，是一种完全基于几何的评估给定嵌入质量的方法。该方法假设，理想的嵌入应能够展开吸引子，并最大限度地增大各条轨迹之间的间隔。作者认为，这种嵌入能够最优地利用环境空间，并减少重构状态空间中不同点所对应的系统真实状态的不确定性。

填充因子的计算方法是：首先从数据的 m 维延迟嵌入中采样 $m+1$ 个随机点。然后从这组点中选取一个参考点 $\vec{x}_r$ ，并计算相应的相对距离向量，

$$\vec{d_i}(\tau) = \begin{bmatrix} x_i(t) - x_r(t) \\ x_i(t-\tau) - x_r(t-\tau) \\ \vdots \\ x_i(t-(m-1)\tau) - x_r(t-(m-1)\tau) \end{bmatrix}. \quad (23)$$

相应的 $m \times m$ 矩阵可以表示为

$$M(\tau) = (\vec{d_1}, \vec{d_2}, ..., \vec{d_m}), \quad (24)$$

由此得到的平行六面体的体积可通过计算 M 的行列式得到,

$$V(\tau) = \det(M(\tau)). \quad (25)$$

填充因子的最终表达式为：计算一组随机采样的平行六面体 $V_i(\tau)$ 的平均体积，再除以采样数据点的取值范围,

$$f = \log\left(\frac{\frac{1}{N}\sum_{i=1}^{N} V_i(\tau)}{(\max_k x(t_k) - \min_k x(t_k))^m}\right). \quad (26)$$

作者建议选择能使区间 $\tau \in (0, T_c/2)$ 内的填充因子 f 达到最大值的 τ，其中 T_c 为特征复现时间。T_c 的值为:

$$T_c = \frac{1}{\omega_c}, \quad (27)$$

其中 ω_c 是时间序列功率谱中占主导地位的频率。

2. 噪声放大

噪声放大是 Casdagli 提出的一种度量，用于尝试量化嵌入 [28,39] 的质量。其依据是，良好的嵌入应有助于进行预测。此外，即使存在噪声，良好的嵌入也应仍能取得相对良好的表现。对于给定的嵌入 $\vec{x}(t)$，噪声放大是根据系统在存在噪声的情况下向未来 T 步的可预测性来定义的。通常，其表达式为:

$$\sigma(T, \vec{x}) = \lim_{\epsilon \to 0} \sigma_\epsilon(T, \vec{x}), \quad (28)$$

其中

$$\sigma_\epsilon(T, \vec{x}) = \frac{1}{\epsilon}\sqrt{\mathrm{Var}[x(T)|B_\epsilon(\vec{x})]}. \quad (29)$$

这里，$\mathrm{Var}[x(T)|B_\epsilon(\vec{x})]$ 表示在嵌入空间 $\mathbb{R}^m$ 中，从初始条件 $\vec{x}$ 出发，对未来 T 步进行预测所得的条件方差。该嵌入空间受到额外小幅观测噪声 ϵ的污染。在这种情况下，假设预测不存在模型误差。可以通过在嵌入 $\mathbb{R}^m$ 中选择近邻作为噪声初始条件的替代[29]来满足这一条件。

最后，对时间序列中采样的一组参考点上的噪声放大量 $\sigma(T, \vec{x})$ 求平均，以计算噪声放大值 σ。噪声放大程度较高的嵌入意味着，嵌入空间中的近邻 $\mathbb{R}^m$ 往往会对应于迅速发散的未来轨迹，因为它们并非真实流形 $\mathcal{M}$ 状态空间中的真实邻居点。因此，重构空间中的微小扰动 $\mathbb{R}^m$ 会导致系统真实状态出现较大的不确定性，从而使噪声的影响大幅放大。

3. L统计量

噪声放大度量的一项缺陷是必须定义 T，即计算噪声放大的预测时域。Uzal[29] 通过修改噪声放大的定义解决了这一问题，新的定义如下:

$$\sigma_\epsilon^2(\vec{x}) = \frac{1}{T_M}\int_0^{T_M} \sigma_\epsilon^2(T, \vec{x})\, dT. \quad (30)$$

该定义针对一系列预测时域计算噪声放大，预测时域的上限为 T_M；研究发现，当 T_M 足够大时，该定义具有较好的稳健性。

用于计算 σ 的算法依赖于将参考点 $\vec{x}_i$ 的 k 个最近邻作为代理。根据点的分布，这会导致每个点的有效噪声水平 ϵ 大小各异。因此，Uzal 提出了一个归一化常数 α_k 来适应这种变化，其定义如下:

$$\alpha_k^2 = \left[\sum_i \epsilon_k^{-2}(\vec{x}_i)\right]^{-1} \quad (31)$$

结合这两种思路，作者提出，噪声放大 σ 衡量某种冗余程度，而 α_k 衡量某种无关性。随后，他们将 L 统计量描述为一个代价函数，用于同时最小化这两个值,

$$L = \log(\sigma\alpha_k) = \log(\sigma) + \log(\alpha_k) \quad (32)$$

VII. 非均匀嵌入

均匀延迟嵌入广受欢迎，得益于其易于实现和优化。直接应用时，均匀延迟只需选择两个参数： τ 和 m。然而，这种方法虽然方便，却牺牲了灵活性并存在局限，尤其是在分析具有多个不同时间尺度动力学的系统时[60,61]。

首先，采用单一延迟会限制重构突出多个迥异时间尺度上的特征的能力[27]。例如，对于特征时间尺度分别为 τ_1 和 τ_2 、且$\tau_1/\tau_2 \gg 1$ 的快慢系统，选择 τ_1 （即慢动力学）作为嵌入时滞，可能会限制重构充分展开与快速动力学相对应的吸引子拓扑的能力。时间尺度为τ_2 的动力学（即快速动力学）会在重构状态空间中表现为噪声波动。

其次，能够满足要求的均匀延迟嵌入重构不一定是最优的。这里必须澄清，最优的定义预设了某种标准或质量概念。Casdagli 指出，嵌入的质量（定义为重构在含噪数据上进行预测时的稳健性）可能会在吸引子的不同区域局部变化[28]。Uzal 在扩展 Casdagli 的噪声放大和失真方法时也强调了这一现象[29]。此外，我们还应考虑到，李雅普诺夫指数等不变测度也会发生局部变化[62,63]。因此，选择单一嵌入时滞意味着这些变化都可能被平均化。

非均匀延迟嵌入被提出作为均匀嵌入的自然延伸，旨在解决后者的一些局限性。非均匀延迟嵌入需要选择多个延迟滞后 $\{\tau_1, \tau_2, ...\tau_{m-1}\}$ ，以构造延迟向量，

$$\vec{x}(t) = (x(t), x(t-\tau_1), ..., x(t-\tau_{m-1})). \tag{33}$$

延迟滞后的选择是一个组合复杂度极高的问题，且其复杂度会随着嵌入维数的增加而增长。为构建非均匀延迟嵌入而提出的方法，通常会迭代选择时间滞后，逐步构建延迟向量，直至达到所需的嵌入维数。本节概述为解决这一问题并实现其自动化而提出的各种方法，包括连续性统计量[27], PECUZAL[64], 投影导数最大化（MDOP）[40], 降阶自回归模型[60,61], 以及蚁群优化等搜索优化算法[65] 和蒙特卡洛决策树搜索（MTCDS）[66]。

A. Garcia 和 Almeida

Garcia 和 Alemeida[47] 较早提出了一种选择非均匀延迟的方法。他们将 Kennel 和 Hegger 提出的最近邻方法加以改进，并用于选择时间延迟。他们的方法还会在多个嵌入周期中，利用所提出的 N-统计量递归地选择滞后。每个周期结束时，都会使用假近邻算法评估新构建嵌入的质量。重复这一过程，直至伪近邻统计量 F 降至临界阈值以下。

对于首个时间延迟 τ_1 的选择，首先针对待测试的某个候选时间滞后 τ^* 进行二维延迟嵌入 $\vec{x}$（t），

$$\vec{x}(t) = (x(t), x(t-\tau^*)). \tag{34}$$

确定嵌入重构中每个点 $\vec{x}$（t_i）的最近邻 $\vec{x}$（t_j）。应选择在时间上不接近的邻居点（即相对于某个 Theiler 窗口）[48,49]。这样可以确保它们在空间上的接近并非完全由时间上的接近造成。随后，计算任意给定两个点之间的两个欧几里得距离 d_{1,τ^*}（t_i）d_{2,τ^*}（t_i），如下所示，

$$d_{1,\tau^*}(t_i) = ||\vec{x}(t_i) - \vec{x}(t_j)|| \tag{35}$$
$$d_{2,\tau^*}(t_i) = ||\vec{x}(t_i + \delta t) - \vec{x}(t_j + \delta t)||, \tag{36}$$

其中 δt 是数据的采样时间。简而言之$d_{1,\ \tau^*}$ 是重构状态空间中最近邻点对之间的空间间隔，而 $d_{2,\ \tau^*}$ 是向前一个时间步后得到的间隔。N-统计量取距离比 $d_{2,\ \tau^*}/d_{2,\ \tau^*} > 10$ 满足条件的点所占的比例，

$$N(\tau^*) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left(\frac{d_{2,\tau^*}(t_i)}{d_{1,\tau^*}(t_i)} > 10\right), \tag{37}$$

其中 N 是时间序列的长度，而 $\mathbb{1}$ 是示性函数。作者根据 Kennel 等人[46]的数值计算，凭经验选取阈值 10。 N (τ^*) 的第一个极小值对应的时间滞后被视为嵌入时滞。

对于非均匀延迟嵌入，每个后续嵌入周期中额外时滞的选择采用类似的过程。然而，用于计算最近邻和成对距离的重构空间，是根据先前选定的时滞有条件地计算的。因此，在非均匀嵌入过程中选择 m^{th} 嵌入时滞，需要使用包含已选定 $m-1$ 个时滞和新候选时滞 m^{th} 的嵌入来计算邻居点和距离d_{2,τ^*},d_{2,τ^*}

已选定的时滞以及新的候选时滞 τ^*,

$$\vec{x}(t) = (x(t), x(t-\tau_1), ..., x(t-\tau_{m-1}), x(t-\tau^*)). \quad (38)$$

B. 连续性统计量

连续性统计量最初由 Pecora 等人提出，作为一种基于向量坐标函数独立性这一思想，按步骤构造非均匀延迟向量的方法[27]。Takens 和 Sauer 都讨论了嵌入重构要求向量坐标彼此独立这一条件[3,67]。Pecora 等人提出，通过检验延迟向量各分量之间的函数依赖，来评估嵌入的质量。向量坐标之间存在函数依赖意味着,

$$x(t-\tau_m) = F(x(t), x(t-\tau_1), ..., x(t-\tau_{m-1})), \quad (39)$$

其中 F 是某个任意函数。构造非均匀延迟嵌入，需要迭代地构建一组时间滞后 $\vec{\tau} = \{\tau_1, ...\tau_{m-1}\}$，以尽可能降低各分量之间存在函数依赖的可能性。每次迭代中，都会检验一个候选时滞 τ_i 与现有滞后分量之间是否存在函数依赖，这些分量对应于 $\vec{\tau}$。如果不存在显著的函数依赖，则可以将 τ_i 加入滞后集合。为检验公式 39 是否成立，作者假设 F 是光滑的，并利用连续性这一性质来量化函数依赖。

考虑由时滞 $\tau = \{\tau_1, ...\tau_m\}$ 构造的一个已有的 m 维嵌入$\vec{x}_m(t) \in \mathbb{R}^m$，以及一个待检验的潜在新嵌入时滞 τ_{m+1}。为检验新时滞的函数依赖，选择嵌入空间中的一个参考点 $\vec{x}_m(t_0)$。如果存在平滑的函数依赖，那么连续性条件指出，重构空间中靠近参考点 ($||\vec{x}_{m,i} - \vec{x}_m(t_0)|| < \delta$) 的点$\vec{x}_{m,i}$ 应当具有彼此也相近的滞后分量 $\mathbb{R}^m$ $m+1^{th}$ ($|x_i(t-\tau_{m+1}) - x(t_0-\tau_{m+1})| < \epsilon$) （见图 2）。

可以计算滞后分量位于参考点的滞后分量 ϵ 范围内的点$\vec{x}_{m,i}$ 所占的比例 p。随后，将该比例与一个零假设进行比较；该假设认为这两组之间的对应关系纯属偶然。p 的值较大，表明 m 维重构与新的滞后分量 τ_{m+1} 之间存在较强关系。Pecora 等人建议使用临界值为 $p^* = 0.5$ 的二项分布，以判断对于选定的 ϵ 是否存在函数依赖；这种方法简单且对噪声具有稳健性[27]。

对于给定的待检验 τ_{m+1} 和一个点样本，使用递减的 ϵ值进行连续性检验，直至无法拒绝零假设。拒绝零假设所需的最小值为

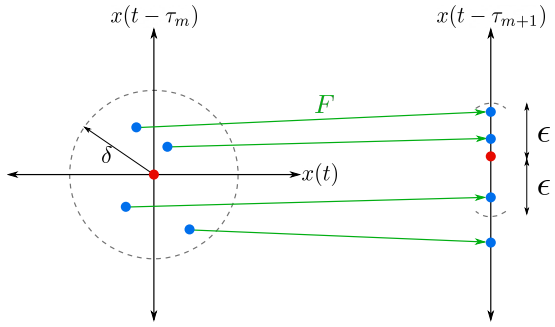


图 2. 对所提出的函数依赖 F 的示意，该依赖通过连续性统计量进行检验。参考点（红色）在半径为 δ的球内有 4 个邻居点（蓝色）。在映射 F下，这些邻居点中只有 3 个与参考点的像之间的距离小于 ϵ。值p表示嵌入空间中的邻居点（左侧）同时也是潜在新滞后分量中的邻居点（右侧）的比例。调整 ϵ 的值，直到 p 不足以拒绝零假设：嵌入空间中的邻居点（左侧）同时也是新滞后分量中的邻居点（右侧）的概率服从二项分布，参数为 $p^* = 0.5$。

ϵ^*。对从数据中抽取的一组参考点取该值的平均值，以计算连续性统计量 $\langle\epsilon^*\rangle(\tau_{m+1})$。

在为现有的一组时滞 $\vec{\tau} = \{\tau_1, ..., \tau_m\}$选择候选时滞 τ_{m+1}的每次迭代中，都会计算连续性统计量曲线 $\langle\epsilon^*\rangle(\tau_{m+1})$。新时滞 τ_{m+1} 取为连续性统计量曲线的相对极大值所对应的时滞。重复此过程，直至达到所需的嵌入维数（根据惠特尼定理）。Pecora 等人还提出了一种欠采样统计量，可用作迭代选择时间延迟的终止准则。详情请参阅原始论文。

将该方法应用于二维环面，得到的嵌入时滞和嵌入维数符合理论预期[27]。所得重构吸引子在视觉上也被认为是最优的。将该方法应用于 Lorenz 混沌时间序列时，也得到了类似结果。然而，所得重构在视觉上似乎存在过度折叠。

C. PECUZAL

对连续性统计量方法的一项批评是，在每次嵌入迭代中选择最优时滞 τ 存在歧义[64]。在 Pecora 等人的原始论文中，“相对极大值”的定义并不明确，也没有客观标准来从多个候选局部极大值中选出最佳时滞[64]。此外，该方法也未考虑选择不同距离 δ 来定义重构中的邻近点所产生的影响。最后，最初被提出作为嵌入算法终止条件的欠采样统计量计算开销很大，而且无法指

明应从候选时滞中选择哪一个。Kraemer 等人[64]提供了更详细的批评。

Kraemer 等人提出，可将连续性统计方法与 Uzal 的L统计量[29] 结合起来，从而提供一种全自动构建非均匀嵌入延迟的方法。在他们的论文中，他们给出了一套工作流程，利用连续性统计量对多个时滞进行粗略搜索，并确定一小组候选滞后。这些候选滞后通常对应于连续性统计量曲线 $\langle\epsilon^*\rangle$的各个局部极大值（τ_{m+1}）。

随后，将 L统计量用作评估标准，以选择每个嵌入周期中应采用哪些候选滞后。这解决了使用连续性统计时选择滞后所存在的歧义问题。在每个嵌入周期中，选择使新的扩展延迟嵌入带来最大 L统计量降幅的候选滞后。L统计量还为嵌入算法提供了终止条件。当从候选滞后中无法实现 L统计量的进一步降低时，嵌入周期便结束，即 $\Delta L > 0$ 在连续的嵌入周期之间。

D. 投影最大导数（MDOP）

投影最大导数（MDOP）方法最初由 Nichkawde 提出，是一种几何方法，可替代互信息和连续性统计量等统计学与信息论方法[40]。MDOP 根据以下标准优化嵌入：重构吸引子应尽可能展开，同时最大限度减少延迟分量中的冗余。与大多数非均匀嵌入方法类似，MDOP 通过嵌入周期递归构建延迟向量。每个周期都会找出一个新的时间滞后，使重构状态空间中各点的方向导数 $\phi'_d(\tau)$ 达到最大。

与 Pecora 在连续性统计量中采用的方法类似，MDOP 首先以每个新的候选时滞与现有时间延迟重构之间存在函数依赖为判据（见公式 40）。然而，与连续性统计量方法不同，Nichkawde 建议使用函数依赖的方向导数 F （见公式 39），以量化重构吸引子的嵌入与展开中的冗余程度。该方向导数定义如下：

$$F'_m = \left|\frac{\Delta F_m}{\Delta x_m}\right| \tag{40}$$

其中 Δx_m 对应于重构空间中一对相邻点 $\vec{x}(t_i)$ $\vec{x}(t_j)$ 之间的空间距离，该空间具有 $m-1$ 维度，

$$\Delta x_{m,ij} = ||\vec{x}(t_i) - \vec{x}(t_j)|| \tag{41}$$

$$= \sqrt{\sum_{k=1}^{m-1}[x(t_i - \tau_k) - x(t_j - \tau_k)]^2} \tag{42}$$

$$\tau_0 = 0. \tag{43}$$

所选的点对还应满足：其空间上的接近并非源于它们在时间上几乎相邻[48,49]。只需设置一个 Theiler 窗口 l_{Theiler}，其中 $|i-j| > l_{\text{Theiler}}$，即可轻松实现这一点。

测试是否纳入新的时间滞后 τ_m ，需要评估预期新滞后分量 $\Delta F_{m,\ ij}$ 的空间变化，其计算方法见公式 44，

$$\Delta F_{m,ij} = |x(t_i - \tau_m) - x(t_j - \tau_m)|. \tag{44}$$

该量用于评估重构吸引子上一个小区域的方向导数，

$$F'_{m,ij} = \left|\frac{\Delta F_{m,ij}}{\Delta x_{m,ij}}\right|. \tag{45}$$

针对每个预期的新时间滞后 τ_m 计算方向导数，并对整个重构吸引子中随机采样的近邻点对取平均，

$$\beta_d(\tau_m) = \langle F'_{m_{i,j}}\rangle_{i,j}, \tag{46}$$

其中 $\langle\ldots\rangle_{i,\ j}$ 表示所有采样点对的几何平均数。作者提出使用几何平均数，因为它在存在离群值时具有稳健性。在每个递归嵌入周期中，选择使方向导数 $\beta_d(\tau_m)$ 最大的时滞 τ_m，用于下一周期的重构。重复这一过程，直至达到所需的嵌入维数 m； m 则通过多种嵌入维数估计方法（如伪近邻法等）确定。

E. 简化的自回归模型

Judd 和 Mees 提出了用于非均匀嵌入的简化自回归模型，[60,68] 将其作为一种根据某个信息准则构建理想模型的方法。该方法涉及构建一个伪线性自回归预测模型，以所有 k 可能的滞后分量作为输入或基函数，

$$x(t+1) = a_0 + \sum_{i=1}^{k} a_i x(t - \tau_i) + \epsilon(t), \tag{47}$$

其中，$\lambda = (a_0, \ldots\ldots, a_k)^T$是各输入对应的待定系数，$\epsilon_t$ 是模型预测误差。对于长度为 N的时间序列，构建矩阵 V_τ，其中每一行都包含某一时刻 t的滞后向量，

$$V_\tau = \begin{pmatrix} x(\tau_k+1) & x(\tau_k) & \ldots & x(1) \\ x(\tau_k+2) & x(\tau_k+1) & \ldots & x(2) \\ \vdots & \vdots & \ddots & \vdots \\ x(N) & x(N-1) & \ldots & x(N-\tau_k) \end{pmatrix}. \quad (48)$$

矩阵 V_τ的维度为（$N-k$）$\times k$，其定义基于一组时滞 $B = \tau_1$、$\ldots\ldots \tau_k$和 $\tau_1 < \tau_2 < \ldots < \tau_k$。

系数的估计值通过最小二乘回归计算得到，

$$\hat{\lambda} = (V_\tau^T V_\tau)^{-1} V_\tau^T \xi, \quad (49)$$

其中 ξ =（x（τ_k+1） x（τ_k+2）$\ldots\ldots$， x（N））。因此，利用所有可能的滞后分量集合 B得到的模型误差可计算为 $e_\tau = V_\tau \lambda_\tau$。

为了将滞后分量的数量缩减到较小的集合，Judd 和 Mees 提出了最小描述长度方法。最小描述长度原理是将奥卡姆剃刀应用于模型选择。该原理认为，对于给定的时间序列预测任务，最佳模型是能够以最简洁的方式描述数据的模型。对于模型选择，这要求在模型准确度和模型复杂度（即模型描述长度）之间取得平衡。模型描述长度 L 可近似表示为，

$$\begin{aligned} L &= L_\epsilon + L_a \\ &= N \log(\bar{\epsilon}) + L_\lambda, \end{aligned}$$

其中，L_ϵ 是模型误差的描述长度，它是时间序列长度 N和均方预测误差 $\bar{\epsilon}$的函数；L_λ 是模型参数的描述长度。

减少滞后分量数量的算法如下：

1. 创建一个空集合 B' 用于存放选定的滞后，并创建一个空集合 B用于存放候选滞后

2. 针对选定的滞后集合 B'，将预测误差定义为 $e_{B'} = \xi - V_{B'}^T \lambda_{B'}$，其中 V_B^T和 $\lambda_{B'}$ 是针对较小的滞后集合 B'定义的。如果 B' 为空，则 $e_{B'} = \xi$。

3. 计算向量 $\mu = V_\tau^T e_{B'}$，并找出与最显著时滞分量对应的最大幅值元素的索引 p。添加此时滞 $\{B' \cup \tau_p\} \to B'$4. 针对新的已选时滞集合 B' 重新计算 μ，并确认最不显著的时滞分量是最近添加的时滞 τ_p。否则，返回步骤 2。

5. 评估模型描述长度 L_B 和 $L_{B'}$、时滞集合 B 和 B'。如果 $L_{B'} < L_B$，返回步骤 2。否则，结束算法并返回已选时滞集合 $B' \to B$。

Small 和 Tse[69,70]实现了最优嵌入时滞和窗口的最小描述长度准则。Hirata 等人[61] 提出了该方法的扩展，其中使用归一化最大似然 $\mathcal{L}$ 取代最小描述长度 L进行模型选择[71,72]。Hirata 等人还提出了对 Judd 和 Mees 提出的上述算法（交叉验证）的一种变体，该变体使用径向基模型，而非伪线性模型。

F. 搜索优化算法

许多常见的非均匀嵌入方法会在每个嵌入周期中执行一次优化步骤。相比之下，搜索优化算法会在可能的时滞构成的状态空间中进行搜索，以确定理想的嵌入滞后组合，而不一定选择最先遇到的局部最优解。这类方法有两个例子：蚁群优化（ACO）方法[65] 和蒙特卡洛决策树搜索（MTCDS）[66]。

蚁群优化（ACO）是一种群体智能方法，由 Dorigo[73,74] 首次提出，其灵感源自蚁群的觅食行为。与其他群体优化方法类似，ACO 会初始化多个智能体（“蚂蚁”），让它们同时对解空间进行初步搜索。根据目标函数评估每个尝试解的质量，并为相应的搜索路径分配“信息素”。信息素会随时间累积和消退，从而影响算法后续迭代的搜索方向，这与现实世界中蚁群寻找最优路径的行为相似。

应用于非均匀嵌入的蚁群优化方法（ACO-NE）在这一框架的基础上进行了几方面改进。(i) 使用基于多种嵌入质量指标（平均邻域距离、最小假近邻数、最小描述长度[69,70]）的目标函数来优化参数；(ii) 将启发式方法融入算法，以加快收敛。感兴趣的读者可参阅原论文，了解更多详情[65]。

Kraemer 等人提出的另一种搜索优化算法 MTCDS[66]，将非均匀时滞选择问题重新表述为决策树搜索。每个嵌入周期都表示为一组从根节点（原始时间序列）延伸出的叶节点或节点，其中每个叶节点代表一个特定候选嵌入时滞的选择。该算法使用蒙特卡洛方法对树进行随机采样，并针对给定的目标函数识别多个局部最优解，然后通过反向传播

来确定每一步中滞后的最佳选择。

VIII. 持久条带与特征时间

本文关注的是为延迟嵌入，尤其是非均匀延迟嵌入，选择时间延迟的问题。许多旨在优化均匀和非均匀情形下嵌入延迟的方法，大致可分为两类：动力学方法（如互信息、连续性等）和拓扑方法（如填充因子、MDOP）。除 PECUZAL 自动化嵌入框架外，非均匀延迟嵌入策略在定义良好嵌入时，都只关注这两个方面中的一个。

非均匀嵌入的另一个缺点是，它们并不总能全面呈现各个延迟的相对显著性。在通过迭代方式构造延迟向量时，每个候选的新时间延迟都必须根据最新更新的嵌入重新评估。因此，后续每个延迟的显著性都取决于此前选定的延迟序列。这对于 MDOP 和 PECUZAL 等全自动算法尤其是个缺点，因为这类算法会自动选择理想的嵌入时滞，却很少或完全不参考它们对嵌入质量的相对影响。此外，不同方法的结果往往缺乏一致性。这可归因于优化所依据的嵌入质量概念各不相同。这些方法通常无法从动力学角度解释每个时间滞后的显著性，以及其改善给定嵌入的能力。

鉴于此，我们认为，优秀的非均匀嵌入方法应具备两个主要特点。首先，嵌入判据在选择嵌入时滞时，应同时利用动力学特征（不相关性、周期性、坐标独立性）和拓扑特征（吸引子展开）。其次，每个选定的嵌入时滞都应具有可解释的显著性。

作为本文的贡献，我们提出一种新方法——“持久链上的显著时刻”（STOPS），利用持久同调和复现分析中的技术来识别非均匀嵌入的时间滞后。我们提出了基于时间序列周期性的信号特征时间尺度谱这一概念，并展示了如何利用它识别理想的时间延迟。我们还独立处理多个时间延迟的选择，这与大多数非均匀延迟嵌入方法中常见的迭代式延迟向量构造方法形成对比。我们在文献中的一组标准周期性和混沌时间序列上展示了该方法的性能。此外，我们还探讨了该方法在包含快慢动力学的实验神经元数据上的表现并表明 STOPS 能够灵敏地识别可解释的时间延迟。

A. 持久同调简介

近年来，持久同调在动力系统等领域尤其受到关注[75–79]。我们还注意到，近期也有研究尝试利用持久同调自动选择延迟嵌入参数[80]。本质上，持久同调旨在量化并追踪对象（网络、点云数据等）的拓扑性质[76]在距离度量不断增大时的演变。逐步增大距离的过程称为“加厚”。在加厚过程中，若某些拓扑特征在较大的距离区间内持续存在，通常认为它们具有显著性。为简便起见，我们将坐标或不断增大的距离 ϵ视为与不断增加的时间相对应。相反，短暂的拓扑变化通常被视为噪声。

我们将加厚过程描述如下。考虑一个按圆形排列的点云 $\mathcal{A} = \{\boldsymbol{x}_i\}$ 。在每个点处放置一个半径为 ϵ 、以该点为中心的开球 $B_\epsilon(\boldsymbol{x}_i)$，并将所有开球的并集作为研究对象，

$$\mathcal{B}(\epsilon) = \bigcup_{i=1}^{N} B_\epsilon(\boldsymbol{x}_i), \tag{50}$$

我们感兴趣的是计算集合 $\mathcal{B}(\epsilon)$ 的同调，其复形由 $K(\mathcal{B}(\epsilon))$ 给出。随着 ϵ 增大，开球的尺寸也随之增大，形成如下所示的过滤，

$$K_0 \subseteq K_1 \subseteq ... \subseteq K_n, \tag{51}$$

其中 $K_n = K(\mathcal{B}(\epsilon_n))$ 且 $\epsilon_0 \leq \epsilon_1 \leq ... \leq \epsilon_n$。更准确地说，每个集合 K_n 表示一组单纯复形，后续的 ϵ_n 产生一个称为过滤的嵌套序列。加厚过程的计算已有详尽记载，主要采用两种算法：Vietoris-Rips[81] 和 Čech[82], ，我们将使用前者。

在这种情况下，我们关注的是集合的同调如何随着 ϵ 增加而变化。简单来说，持久同调旨在枚举并追踪集合[83]中n维“孔洞”的数量。具体而言， H_0、 H_1 和H_2 分别对应低维同调的情况。0、1 和 2 维孔洞分别对应不相交分支、圈和空腔。单纯形（三角形、四面体等）被视为实心组成部分，因此不属于孔洞。

在低维奇异吸引子的背景下，持久同调利用来自嵌入或其他方式的空间数据，追踪吸引子中同调特征（通常为 H_1 孔洞）的持久性或生存期（死亡时间减去诞生时间）。所有与特征的诞生和持久性相关的信息

都可以很容易地表示在持久性图中。在持久性图中，每个同调特征（即 H_0 个不相交分支、H_1 个孔洞）都由一个绘制的点表示，其坐标为（ϵ_b，ϵ_d），分别对应其诞生时间和死亡时间。

追踪同调特征的诞生时间和死亡时间有两个重要作用。首先，它能追踪数据中孔洞的位置。在分析相空间轨迹时，这些孔洞可能对应于短期伪周期或时间序列中的转折点（见图 5）。通过追踪 ϵ 以及代表性圈，还可以识别孔洞的边界点。其次，同调特征的生命周期有助于估算该特征的相对大小；根据数据类型，这一大小可能与其显著性有关，也可能无关。所有这些特征都可以用持久图表示（见图 3），其中距离对角线越远的点代表越持久的同调特征。

B. 特征时间尺度

为非均匀嵌入选择滞后的一种方法，是选择与系统动力学自然时间尺度相关的数值。选择远小于或远大于该尺度的滞后，逻辑上分别对应于高度冗余和无关的情况。考虑最简单的周期信号情形。可以认为，理想的嵌入需要一个与其主要动力学时间尺度（即周期性）相关的延迟。为此，我们在特征时间 τ 的定义中采用四分之一周期启发式规则，

$$\tau = \frac{T}{4}.$$

将这一概念自然地拓展到更复杂的动力学中，可以从自然频率集合中选取时滞，

$$\tau_i = \frac{2\pi}{4\omega_i}.$$

由于在各个频率 ω_i处评估 τ_i ，而这些频率的幅值可能相差很大，因此有可能捕捉具有多重时间尺度的系统动力学，例如快慢动力系统。将嵌入时滞与自然频率和周期性联系起来，也使 τ 的选择具有一定的可解释性，并且与动力学直接相关。尽管这相比典型的非均匀嵌入方法具有潜在优势，但它要求能够从任意给定的信号中准确测量自然频率（或等效量）。

如果所分析的信号相对平滑，且易于分解为正弦波，那么傅里叶变换将是解决这一问题的绝佳方法。然而，在分析类不连续信号时，这种方法很快就会失效，例如神经元电压，其时间序列的特点是爆发期与静息动力学期交替出现。傅里叶变换还要求时间信号在统计特性上保持平稳。对于混沌时间序列而言，由于信号的相位和周期会随时间变化，因此无法满足这一条件。例如，洛伦兹和 Chua 等混沌信号的频率功率谱会呈现指数衰减的形状。分析实验时间序列时也可能遇到类似情况，因为相位和/或周期可能发生漂移和振荡。

检测周期性的另一种方法，是采用复现分析和不稳定周期轨道检测中使用的复现距离概念[84]，

$$d(\tau) = \frac{1}{N}\sum_{i=1}^{N}\left\| \begin{pmatrix} x(t_i) \\ x(t_i-\tau) \end{pmatrix} - \begin{pmatrix} x(t_i+4\tau) \\ x(t_i+3\tau) \end{pmatrix} \right\|, \quad (52)$$

其中，如果复现距离$d(\tau)$低于某个阈值 d^*，则 τ 为特征时间，

$$d(\tau) < d^*. \quad (53)$$

复现距离跟踪状态空间中某一点与其未来轨迹之间的位移。对于周期轨道，若正确选择了 τ，那么当周期轨道返回到接近其初始位置时， $d(\tau)$将出现局部极小值。这种特征时间检测方法 τ 并不适用，因为它无法区分 $T = \tau/2$ 的情况（此时得到的嵌入轨迹会沿对角线聚集）。

我们提出了一种从时间序列中识别特征时间并评估其显著性的方法。特征时间的识别通过对“数据片段”（二维延迟嵌入时间序列中连续的短窗口）进行采样，并计算其持久同调来完成。利用持久链计算得到的代表循环同调，为每个识别出的特征时间赋予显著性分数，从而构建特征时间谱。随后可利用该谱为非均匀时间延迟嵌入选择滞后。这个组合框架称为持久条带显著时间（SToPS）。

IX. 持久条带显著时间（SToPS）

A. 持久条带

首先要解决的挑战是识别所有周期为 T 的伪周期行为。对于特征时间 τ，四分之一周期启发式认为

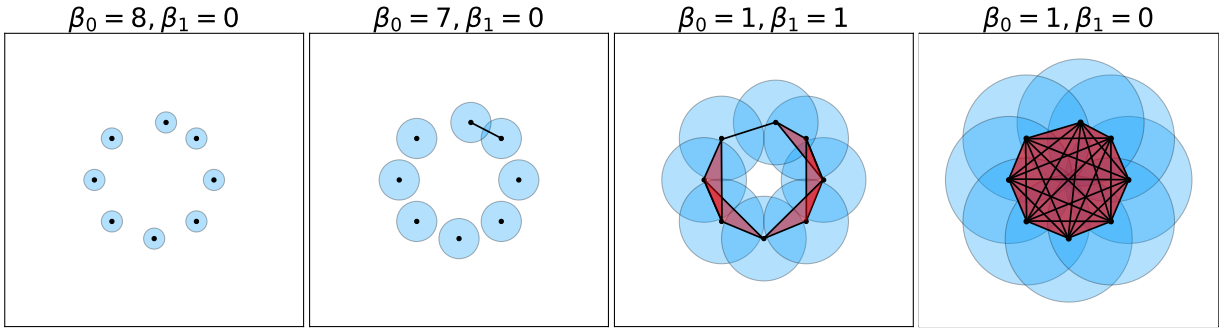


图 3. 带有贝蒂数的过滤过程示意图 β_0，β_1 分别表示同调中 H_0（不相交分支）和 H_1（孔洞）的数量。过滤过程使用半径逐渐增大的开球 ϵ。起初，当 ϵ较小时，构造出的球形成八个互不相交的分支，且没有孔洞（$\beta_0 = 8$，$\beta_1 = 0$）。随着 ϵ不断增大，两个球最终相交，导致两个分支合并。随后，ϵ继续增大，直至所有分支合并为一个分支，中心留下一个未填充的区域（$\beta_0 = 1$，$\beta_1 = 1$）。最终，ϵ继续增大，直至中心的孔洞被填满（以红色阴影标出），形成一个没有孔洞的实心分支（$\beta_0 = 1$，$\beta_1 = 0$）。

相应的二维嵌入（坐标为 $(x(t)\ x(t-\tau))$）将形成一个在重构空间中近似达到最大凸度的周期轨道。需要注意的是，无法保证该轨道的精确形状为圆形，而且其形状会因信号而异。

为检验信号 $x(t)$ 是否具有特征时间τ 下的周期行为，构造该时间序列的二维延迟嵌入，其中包含一个滞后分量，

$$\vec{x}(t) = (x(t), x(t-\tau)). \tag{54}$$

随后从时间序列中采样一组长度为 N 的 $l = 4\tau$ 数据片段，其初始位置随机选取 t_i，其中每个数据片段定义为：

$$\vec{x}_i(t) = (x(t), x(t-\tau)), \quad t \in [t_i, t_i + 4\tau]. \tag{55}$$

我们认为，根据四分之一周期启发式，这种长度的数据片段应足以检测二维嵌入中的环结构。可以计算每个数据片段的持续同调，以检测周期为 l的轨道是否存在。对于每个数据片段$\vec{x}_i$（t），从所得持久性图 PD中提取最大持久性。如果相应图中的最大持久性超过临界值 ρ，则认为该样本数据片段包含一个长度为 l 的轨道。ρ 取为相空间中两个连续观测点之间的平均距离。因此，可以将一个用于量化特征时间 τ 显著性的朴素值 P_i 定义为：

$$P_i = \begin{cases} \text{maxpers}(PD_i), & \text{maxpers}(PD_i) > \rho \\ 0 & \text{maxpers}(PD_i) < \rho \end{cases}. \tag{56}$$

通过对每个特征时间 τ 下所有非零得分 P_i 取平均，可以利用该值计算总体最大持续谱 $P(\tau)$，

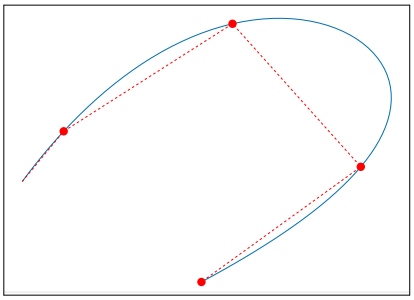
$$P(\tau) = \langle P_i \rangle_{P_i \neq 0}. \tag{57}$$

我们还施加了一项额外约束：用于重建轨道中孔洞边界（H_1 同调）的点数至少应为 N_{hole}。这样可以避免纳入伪孔洞：少量点可能会让人误以为存在孔洞，但嵌入弦长度不足以闭合轨道（图4）。基于至少需要 8 个点才能识别出小滞后 $\tau = 2$ 对应的孔洞，且不排除较大滞后的理由，我们为分析选取了 $N_{hole} = 8$ 的最小值。事实证明，该值在我们的分析中效果良好。

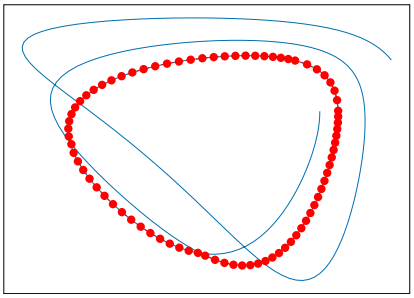
最大持久性谱提供了一种简单的方法，用于量化 τ 在多大程度上对应于动力学显著时间尺度（即周期）。然而，P（τ）由于依赖于每个同调特征的生存期，因此偏向较大的特征。因此，在较小时间尺度上，空间范围较小但具有动力学显著性的特征未能得到充分体现。我们还注意到，P（τ）无法区分病态低效嵌入，例如环并非最大程度地呈圆形，或带有拖尾的嵌入（见图5）。

正如我们从 III A 节所见，良好的嵌入应尽可能充分展开，以最大限度地利用重构状态空间，同时能够抵御噪声放大的影响。因此，每个特征时间 τ 的重要性应根据相应线段在时滞为 τ 的二维延迟嵌入中展开成环结构的程度来加权。

据此，我们提出显著性评分 $S(\tau)$，用以衡量每个特征时间 τ 的动力学显著性，同时考虑采样线段展开的质量。根据良好嵌入的拓扑学概念，$S_i(\tau)$ 的显著性评分



(a) 滞后过小，导致
链长度过短，无法完全包围
从而检测出伪
孔洞，其边界点数量极少
。



(b) 即使在较大时滞下也存在孔洞
这是由于相似轨道上的轨迹彼此重叠，
较长的线段无法
有效利用所有可用的点。

图4. SToPS 方法的病态问题。

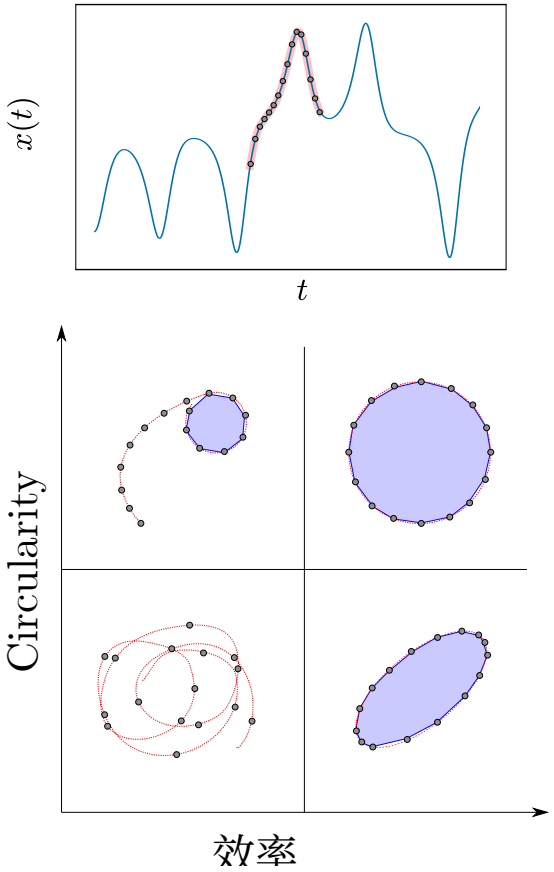


图 5. 随机抽取的时间序列中的一条“线段”（阴影红色）及不同时间尺度下嵌入的后续病态情形示意图 τ。每种情形都对应于根据圆度和效率概念判定的非最优嵌入。显著的特征时间应力求同时最大化这两项指标（右上象限）。蓝色阴影区域对应于由嵌入条带的代表性同调计算得到的孔洞同调生成元所构成的多边形。

定义为该 i^{th} 采样线段的显著性评分，

$$S_i(\tau) = \alpha_i(\tau)\gamma_i(\tau), \tag{58}$$

其中 $\alpha(\tau)$ 和 $\gamma(\tau)$ 是两个独立的度量，分别称为圆度和效率。该评分也不会偏向较大的同调特征，因此能够检测小型和大型的伪周期动力学。

圆度 $\alpha(\tau)$ 用于量化嵌入空间中持久条带展开的质量。能够产生圆形环的嵌入意味着所选时滞 τ能最大程度地展开动力学（即 τ 可能对应于特征时间）。因此，与椭圆及其他具有离心率的形状相比，更圆或更规则的环具有更高的取值。这也使圆度能够作为冗余度的度量，较低的 α 对应较高的冗余度。

为了计算 α_i 给定采样条带的边界点 $\vec{b}_i$ ，通过检查代表循环来确定与最持久同调特征的诞生相对应的点。使用主成分分析（PCA）确定嵌入孔洞的长轴和短轴。由于每个数据片段都用于评估二维延迟嵌入中的单个时滞，因此第一和第二主特征值大致对应于嵌入孔洞点的外接椭圆的相对大小。因此，我们将圆度定义为第一和第二主特征值之比，并对所有数据片段取平均

$$\alpha(\tau) = \langle \alpha_i \rangle_{P_i \neq 0} = \left\langle \frac{\lambda_{2,i}}{\lambda_{1,i}} \right\rangle_{P_i \neq 0}, \tag{59}$$

其中 $\alpha(\tau) \in (0, 1]$。由于 $\alpha \to 1$，嵌入孔洞呈现出更均匀的圆形。

效率 $\gamma(\tau)$ 定义为两个面积之比，

$$\gamma(\tau) = \frac{A_h}{A_{pc}}, \tag{60}$$

其中 A_h 是由有序边界点集合 $\vec{b}_i$ 所围成的孔洞面积，而 A_{pc} 是包含线段中所有点的最小凸多边形的面积 $\vec{x}_i(t)$。孔洞面积 A_h 可用鞋带公式直接计算[85]。后者的面积

A_{pc} 可通过 Graham 扫描算法类似地计算：首先确定最小的边界凸多边形[86]。与 $\alpha(\tau$类似，效率得分也受 $\gamma(\tau)$ $\in [0,1]$的约束。效率用于衡量嵌入弦相对于孔洞对环境空间的利用程度。它可用于检测这样的情况：检测到的孔洞呈圆形，但没有利用线段的全部长度（见图 5）。

X. 测试方法

STOPS 在三种不同类型的时间序列上进行了测试，涵盖周期性、混沌和快慢动力学。

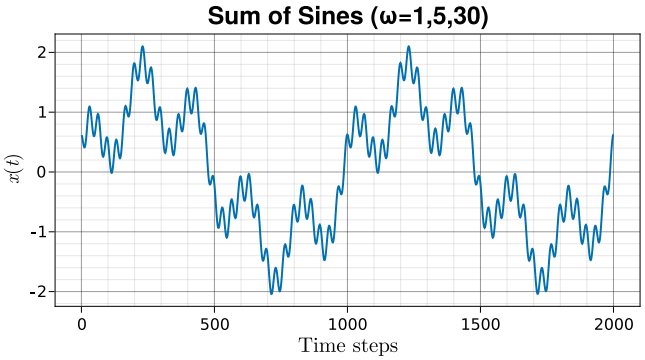
首先，使用正弦和信号（见图6），其步长为 $dt = 0.001$，以模拟周期时间序列的情况，

$$x(t) = \sum_{i=1}^{N} A_i \sin(2\pi\omega_i t + \pi\phi_i), \tag{61}$$

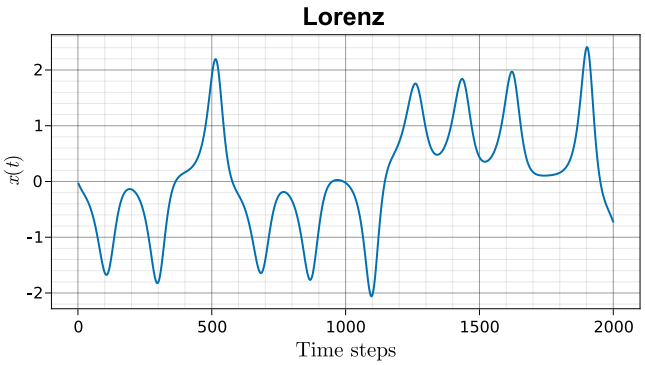
其中 $N = 3$ 是叠加正弦波的数量。选择相位 $\phi = \{0, 0.25, 0.75\}$ 和幅值 $A = \{1, 0.5, 0.2\}$，以确保动力学的空间尺度彼此不同（见图6）。

分析的第二个时间序列是经典洛伦兹混沌时间序列的 x分量，用于表示混沌时间序列的情况。该时间序列采用 4th 阶龙格-库塔格式进行数值积分，积分时间步长为0.0004，并下采样至有效时间步长 $dt = 0.004$，共25000步。

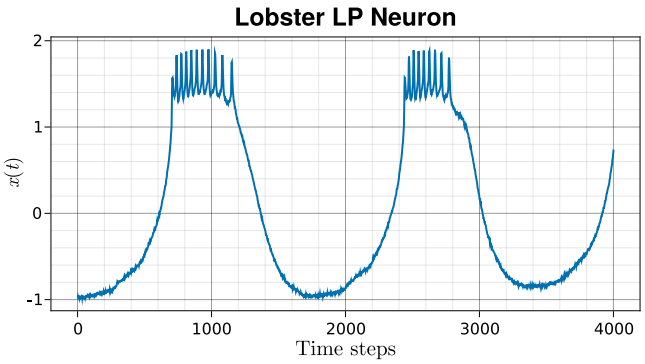
第三个时间序列由测自龙虾口胃神经节（STG）外侧幽门（LP）神经元的实验数据构成。该时间序列代表第三类动力学，即具有两种不同空间和时间尺度的快慢动力学。Abarbanel[87] 最初分析了该时间序列，其中包含两种特征性动力学：对应神经元尖峰动力学的小幅值、短时间尺度振荡，以及对应神经元爆发的长时间尺度周期性行为（见图6）。此外，爆发动力学的相位也会随时间略微变化。这会导致二维嵌入中预期环的空间位置逐渐偏移。因此，长度为 T 的过长数据片段会受到惩罚，因为这些环最终会被填满。



(a) 由包含 $\omega = \{1, 5, 30\}$、$\phi\{0, 0.25, 0.75\}$ 和 $A_i = \{1, 0.5, 0.2\}$ 的 3 个正弦项之和构造的人工时间序列。



(b) 混沌洛伦兹时间序列的第一个分量（x（t））。



(c) 展示伪周期性的龙虾 LP神经元时间序列在两个尺度上的动力学。快速（尖峰神经元）和缓慢（爆发神经元）。

图6。测试的三类时间序列：正弦和（周期性）、洛伦兹（混沌）、龙虾 LP 神经元（多个差异显著的时间尺度）。

在应用嵌入算法之前，先对所有三个输入时间序列进行归一化，使其具有零均值和单位标准差。对正弦和以及数字化的龙虾 LP 神经元实验数据添加少量加性噪声$\xi \sim N(0,\ 0.001_2)$，以确保数据值不会完全重合。在所有情况下，计算嵌入滞后时，输入数据均限制为 25000 个点，以确保计算时间足够短。这是因为采用了k-均值++ 随机采样器[88]，，其计算量会随点数增加，以选择数据片段滞后-均匀遍历重构状态空间的阳离子。

根据时间序列的类型，测试了不同范围的特征时间，以计算效率、圆度和显著性评分曲线。每个 τ 由采样的 250 个长度为 $T = 4\tau$ 的数据片段组成。由于 Vietoris-Rips 过滤的计算扩展性较差，当 T 较大时，较长的数据片段会按 k的因子进行下采样，

$$k = \left\lfloor \frac{4\tau}{N_p} \right\rfloor, \tag{62}$$

其中 $N_p = 250$ 是允许的最大数据片段长度的近似尺度。选择该数值是为了适应计算 Vietoris-Rips 过滤及其相关持久同调时较低的计算效率。 用于 ρ^* 定义将同调特征归类为显著特征所需的最短寿命的阈值，被设定为时间上相邻点之间距离的平均值，

$$\rho^* = \langle ||\vec{x}(t_i) - \vec{x}(t_{i+1})|| \rangle_{i.k}, \tag{63}$$

其中 $\langle ... \rangle_{i,\ k}$ 是二维延迟嵌入线段中所有点的平均值 $\vec{x}$（t）以因子k进行子采样。圆度、效率和显著性得分曲线由 SToPS 计算得出。同时计算了平均值 S_μ（τ）和标准差 S_σ（τ）曲线，并将其与一组选定的嵌入延迟优化指标进行比较。在我们的分析中，每条曲线的峰值均通过观察选取。不过，也可以通过使用搜索算法识别显著性评分曲线中的所有局部极大值，来实现峰值的自动识别。

将计算得到的显著性评分与用于为均匀延迟嵌入选择 τ 的互信息进行了比较。还使用了两种自动化方法 MDOP[40] 和 PECUZAL[64], ，作为非均匀延迟嵌入的比较基准。这些计算使用 ‘DynamicalSystems.jl’ 软件包[89] 提供的实现完成。

除了比较各方法输出的滞后值之外，我们还关注不同的非均匀嵌入会如何影响预测任务的表现。对于每种非均匀嵌入方法，均取前两个主导时滞来构建三维延迟嵌入。对于 SToPS，这是通过目视检查 $S_\mu(\tau)$ $S_\sigma(\tau)$ 曲线完成的。PECUZAL 和 MDOP 的时间滞后则分别取自各自迭代过程检测到的前两个时间滞后。

所得延迟嵌入用于训练一个简单的四层前馈神经网络，该网络由两个隐藏层、一个输入层和一个输出层组成。每个隐藏层包含 128 个节点，并采用 ReLU激活函数，整体网络架构为 3:128:128:3。

该神经网络经过训练后可输出一步预测，

$$\vec{x}(t + \delta t) = \mathrm{NN}(\vec{x}(t)). \tag{64}$$

学习率设为 0.001，批量大小为 512，训练 30 轮次。每种情况下，仅使用时间序列数据的前半部分进行训练，后半部分用于验证。验证时通过计算 n 步自由运行预测来进行。这一过程

通过提供初始条件并将神经网络的输出反馈 n 次，得到最终预测（见公式 65），然后据此计算预测误差。

$$\vec{x}(t + n\delta t) = \mathrm{NN}^{(n)}(\vec{x}(t)). \tag{65}$$

XI. 结果与讨论

A. 显著时间

1. 周期动力学 - 正弦波之和

周期性正弦和时间序列代表了具有多重时间尺度的周期动力学情况。分量频率选择为 $\omega = \{1, 5, 30\}$，对应于滞后τ =处的 3 个不同特征时间（250、50、8）。对于基线互信息 $I(\tau)$ 度量而言，正弦和时间序列是一项重大挑战（见图7），其中极小值只能识别出数据中最高频率的周期性。最大持久性 $P(\tau)$ 在从该时间序列中识别任何显著时间滞后方面同样没有发挥作用。

对于自动非均匀嵌入方法，PECUZAL 在 $\tau = 341$ 处发现了一个时间滞后。MDOP 在 $\tau =$ 处返回了四个时间滞后（241、42、271、180），其中两个接近预期的时间滞后 $\tau =$（250、50、8）。两种方法检测到的其余时间滞后均与预期的特征时间没有明显关系。

相比之下，所提出的显著性评分$S_\mu(\tau)$ 和 $S_\sigma(\tau)$ 均在预期时间滞后附近出现峰值，对应于分量周期约四分之一处（即 $T = 1, 1/5, 1/30$）（见图7）。然而，与 $\omega = 5$分量对应的峰值（即 $\tau = 45$）并不明显，反而在相近的时间滞后处出现了两个峰值。这种异常可能是因为第二个频率的空间和时间尺度与较大的时间尺度在动力学上区分得不够明显。在计算二维嵌入时，这可能导致持久条带在特征时间处形成螺旋，而非圆形孔洞。我们注意到，圆度 $\alpha(\tau)$ 与效率 $\gamma(\tau)$ 得出的曲线截然不同，后者会显著影响最终曲线的形状 $S_\mu(\tau)$。附录中提供了 MDOP、PECUZAL 和 SToPS 的相空间重构对比。

2. 混沌动力学——洛伦兹

对于洛伦兹系统，互信息的极小值给出的滞后为 $\tau \approx$ (40，150)。这一结果与最大持久性 $P(\tau)$ 曲线中取得的极大值非常吻合。在 PECUZAL 检测出的两个时滞中 $\tau =$ (46，24)，

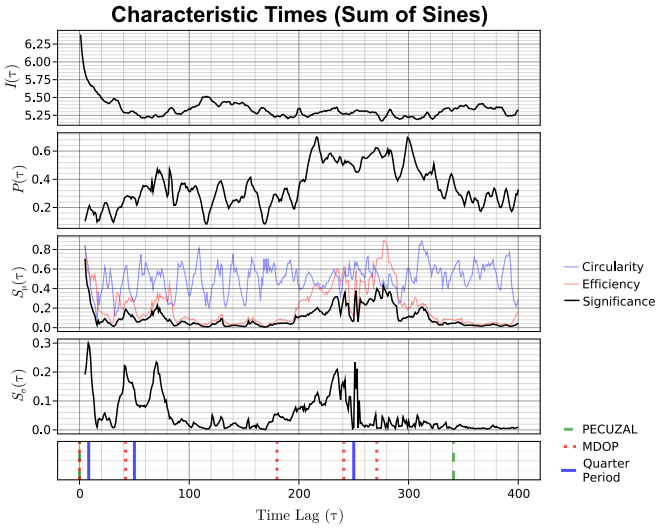


图7. 正弦和时间序列（分量频率为 $\omega = (1, 5, 30)$ 的嵌入延迟估计方法比较。从上到下：互信息、最大持久性、显著性评分（黑色），以及圆度（蓝色）和效率（红色），显著性评分的标准差。 基于分量频率周期四分之一得出的理论特征时间，以蓝色实线表示。 PECUZAL 和 MDOP 算出的非均匀延迟比较，分别以绿色和红色竖线表示。

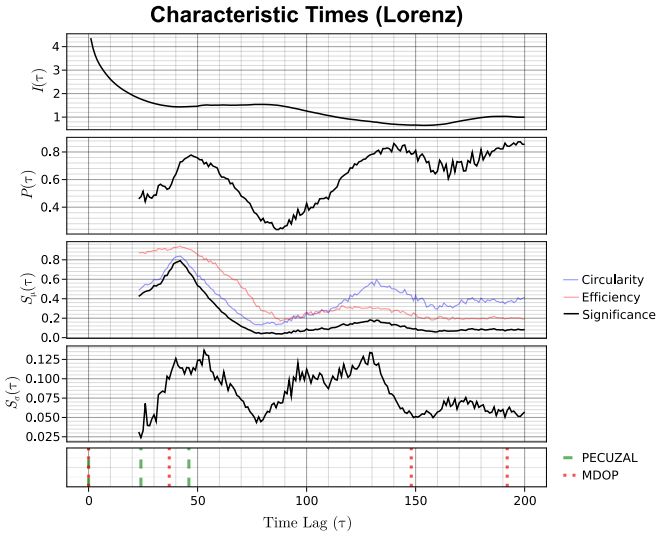


图8. 混沌洛伦兹时间序列嵌入延迟估计方法的比较。从上到下：互信息、最大持久性、显著性评分（黑色），圆度（蓝色）和效率（红色），显著性评分的标准差。PECUZAL 和 MDOP 计算出的非均匀延迟分别以绿色和红色竖线表示。

其中一个与互信息曲线的一个极小值相近。同样，MDOP 得出了 3 个不同的时滞$\tau = (148, 192, 37)$，其中一个与互信息的极小值大致吻合。

使用 STOPS 计算的显著性度量在 $S_\mu(\tau)$ $S_\sigma(\tau)$ 范围内产生了两个不同的极大值，对应的滞后为 $\tau = (41, 130)$。第一个滞后与 PECUZAL 和 MDOP 计算出的滞后相近。然而，第二个滞后 $\tau = 130$ 会导致时间序列过度嵌入，并且与 PECUZAL 或 MDOP 输出的任何滞后都没有直接对应关系。仔细检查 $\tau = 120$处嵌入的时间序列，可以看到洛伦兹吸引子的叶瓣在周期的 $3/4$处重新出现，其边界由多个密集的环构成。虽然这可能会在叶瓣附近的持久同调中形成定义清晰的孔洞，但并未有效利用采样线段中的所有点（即 $4\tau = 520$ 个点）（见图4），因此不应将其归类为该时间序列的特征时间滞后；这一点也反映在显著性评分大幅降低上。使用这一时间滞后会导致时间序列过度嵌入。不过，仔细检查采样线段后可以发现， $S_\mu(\tau)$ 的增加部分归因于吸引子鞍点附近轨迹扩散所形成的孔洞。

互信息的连续极小值与非均匀嵌入方法检测到的时间滞后之间，似乎也存在对应关系。这表明，互信息 $I(\tau)$ 或许有助于指导非均匀嵌入的滞后选择。然而，一个难点在于判断极小值对应的时滞是否处于可接受的嵌入窗口 $m\tau$ 内，从而避免无关性问题。

3. 快慢动力学 - LobsterLP神经元

通过观察龙虾神经元数据（见图6），可以推断出两个主导时间尺度，其预期滞后约为 $\tau = (12, 400)$。这些时间尺度的时间变化可能归因于观测噪声或潜在的混沌动力学。研究发现，互信息和最大持久性这两种均匀嵌入度量在识别滞后方面效果不佳。不过，当接近预期特征滞后时，最大持久性 $P(\tau)$ 开始迅速上升$\tau = 400$。这并不意外，因为在这些滞后下进行二维嵌入，开始展开较大时间尺度动力学中的大轨道；这些动力学对应于神经元的慢爆发阶段（见图10）。

对于非均匀嵌入，PECUZAL 和 MDOP 都得出了大量可能的嵌入滞后。然而，这两种方法都未能成功找回快速动力学对应的滞后（见图9）。此外，除了对应慢速动力学的滞后 $\tau \approx 500$ 之外，PECUZAL 和 MDOP 还产生了多个额外的时间滞后，而这些滞后与数据时间尺度之间没有明显可解释的关系。

使用 STOPS 得出的显著性评分能够检索出数据中存在的两个主要时间滞后，约为 $\tau = (5, 500)$。较大的时间滞后的位置似乎略有分歧——

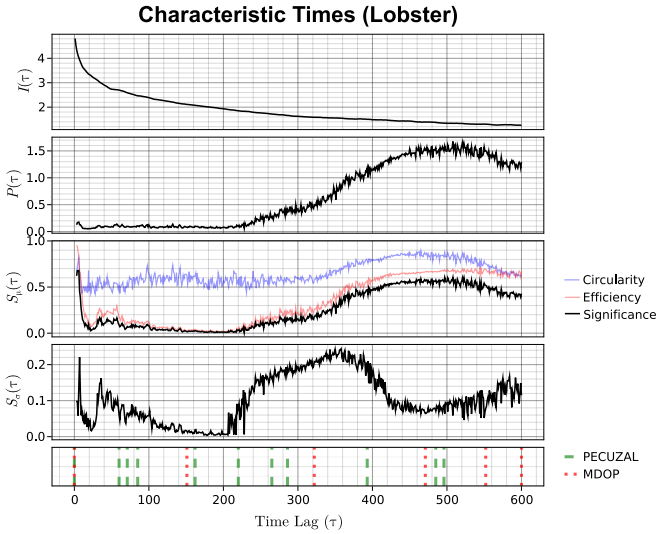


图9. 混沌洛伦兹时间序列嵌入延迟估计方法比较。从上到下：互信息、最大持久性、显著性评分（黑色），圆度（蓝色）和效率（红色），显著性评分的标准差。PECUZAL 和 MDOP 计算出的非均匀延迟分别以绿色和红色竖线表示。

不同方法对较大时间滞后的位置判断$S_\mu(\tau)$ 和 $S_\sigma(\tau)$ 得出的时间滞后略有不同。此外，$S_\mu(\tau)$ 在约$\tau = (40, 60)$ 处也显示出一个小峰值。进一步检查采样线段的代表性同调后发现，这是快速尖峰动力学产生多个重叠轨道所致。不过，$\tau = 5$能够捕捉到这一特征，而 $\tau = (40, 60)$ 处的峰值则是过度嵌入的时间序列落在相似轨道上的伪影（见图4）。

根据这些结果，只有 SToPS 能够检测到数据中的两个主导时间尺度。PECUZAL 和 MDOP 的滞后取决于此前检测到的滞后的选择，这是由算法采用的迭代方法所致。相比之下，SToPS 会生成单一的特征时间谱 $S(\tau)$，据此可以独立评估并选择每个潜在滞后的显著性。我们在图10中直观比较这三种方法得到的三维延迟嵌入。正弦和数据与洛伦兹数据的类似比较见附录。对于 PECUZAL 和 MDOP，我们分别按检测顺序选取前两个非零延迟，位置为 $\tau = (265, 496)$ 和 $\tau = (322, 552)$。SToPS 的滞后通过目视从$S_\mu(\tau)$和 $S_\sigma(\tau)$中选出，滞后时间为 $\tau = (5, 500)$。从重构状态空间的投影中，我们发现 SToPS 将不同投影与不同时间尺度的动力学联系起来。这使得展开后的结果在视觉上更易于解读。相比之下，PECUZAL 需要大量维度才能在视觉上检测到所有动力学成分。限制维度数量会导致快速动力学被遮蔽，而慢速动力学得以保留。

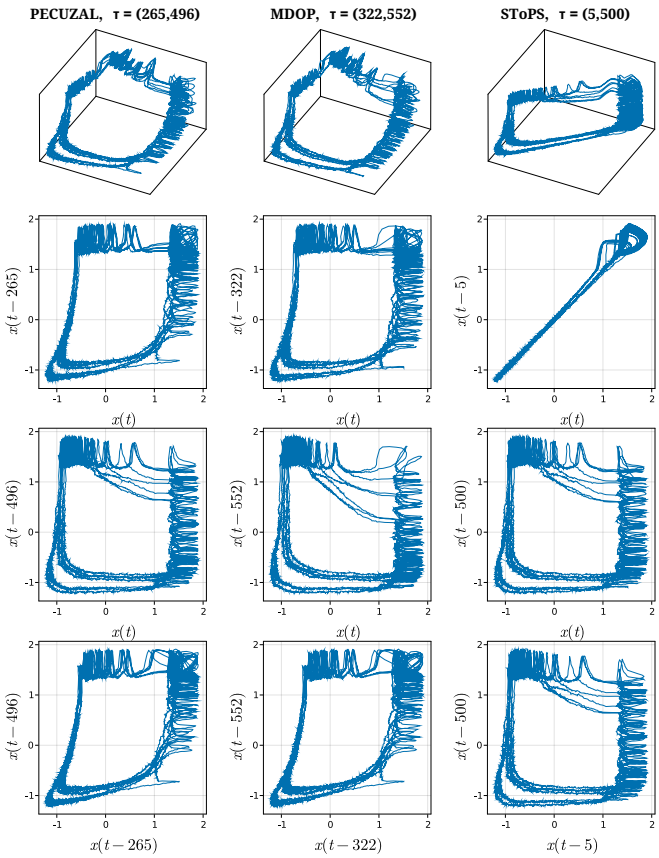


图10. PECUZAL、MDOP 和 SToPS 的三维时间延迟嵌入及其对应二维投影的比较。PECUZAL 和 MDOP 的时间延迟取自检测到的前两个延迟。SToPS 的延迟根据 $S_\mu(\tau)$ 目测选取。SToPS 能捕捉尖峰神经元快速的小尺度振荡，而 PECUZAL 和 MDOP 则不能。

B. 自由运行预测

使用不同非均匀嵌入构建的模型，计算了非周期性洛伦兹系统和龙虾 LP 神经元时间序列的自由运行预测误差。每种情况下，均使用 PECUZAL 和 MDOP 检测到的前两个时滞构建三维延迟嵌入。对于 SToPS，则选取 $S_\mu(\tau)$ 和 $S_\sigma(\tau)$ 曲线中目测最显著的前两个极大值。对于洛伦兹系统，由于仅存在一个相关时间滞后（图4和图8）（即 $\tau = 41$），我们采用均匀嵌入方案，其中第二个时滞是第一个时滞的倍数。

洛伦兹系统的预测误差是通过25 步自由运行预测计算得出的，该预测使用训练好的前馈神经网络（见第 X 节）。对于龙虾 LP 神经元，训练使用了降采样数据集，其中每隔三个点取一个点。进行子采样是为了缩短训练时间。随后，使用降采样数据对 10 步自由运行预测进行了评估（即相当于 30 步自由运行预测）。较少的自由运行

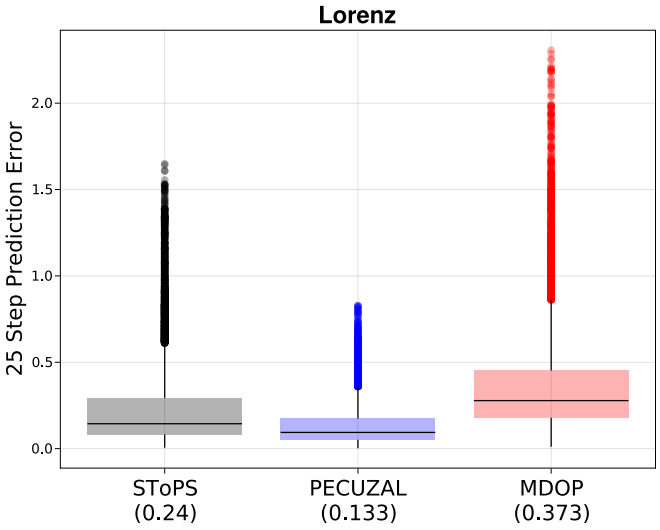


图 11. 使用在三维延迟嵌入上训练、用于一步预测的神经网络，对洛伦兹时间序列进行 25000 步自由运行预测时，25 步预测误差的分布。

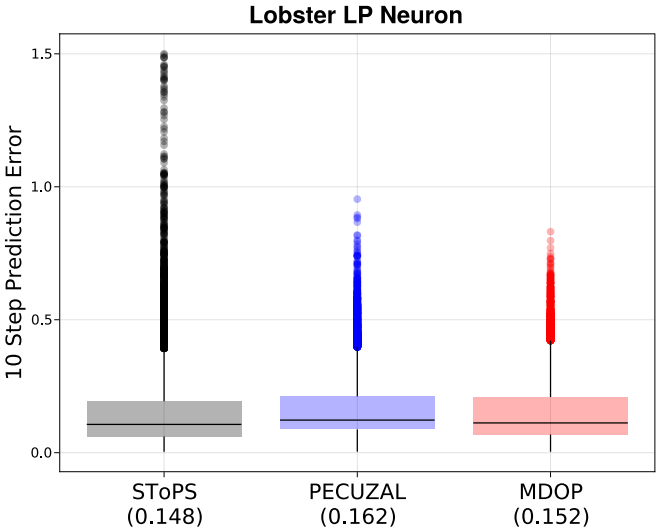


图 12. 对整个下采样数据集（约 200000 步）中的龙虾 LP 神经元时间序列进行 10 步自由运行预测的误差分布。所用神经网络基于三维延迟嵌入训练，用于一步预测。

预测步数是为了在时间序列快速神经元放电区间内更准确地评估预测性能。 在这两种情况下，前半部分数据用于训练，后半部分用于测试和评估。附录中还提供了使用非子采样数据进行自由运行预测的补充分析。

每次预测的误差计算为向前预测 n 步得到的延迟向量与实际值之间的误差幅度，

$$E(t) = ||\mathrm{NN}^{(n)}(\vec{x}(t)) - \vec{x}(t+n)||. \tag{66}$$

图 11 和图 12 给出了两种时间序列采用三种不同嵌入方法时 $E(t)$ 的分布。 对于洛伦兹时间序列，STOPS（持久链）的平均误差介于 PECUZAL 和 MDOP 之间。在实验数据（龙虾 LP 神经元）中，STOPS 的误差较低，优于另外两种方法。这些结果也反映在两种情况下相应的中位数中。洛伦兹时间序列的预测误差中位数为 0.106（STOPS）、0.122（PECUZAL）和 0.112（MDOP）。龙虾 LP 神经元的预测误差中位数为 0.106（STOPS）、0.122（PECUZAL）和 0.112（MDOP）。此外，与 PECUZAL 和 MDOP 相比，STOPS 在龙虾 LP 神经元的自由运行预测中呈现出尾部更重的误差分布。尽管预测误差可能更低，但我们指出，这种改进并不显著，也不应成为 STOPS 的目标优势。相反，我们认为，STOPS 的主要优势在于它提供的时滞能够结合时间序列中观测到的动力学进行解释。

使用 STOPS 的一个优势是，在时间序列的嵌入过程中，有意纳入快速和缓慢的时间尺度，因此，预计基于这种嵌入训练的模型能够更好地解析快速动力学；如果选择较大的延迟，这些快速动力学可能会被遗漏。图 13 验证了这一点：STOPS 生成了 10 步自由运行预测，能够更好地重现神经元时间序列中典型的小尺度快速脉冲动力学。相比之下，基于 PECUZAL 和 MDOP 的相同数量滞后训练的模型，其自由运行预测无法捕捉到尖峰动力学中的同等细节。我们注意到，当使用未经子采样的完整数据集训练预测模型时，这一优势并不那么明显。不过，在这种情况下，STOPS 的预测误差略低。

XII. 噪声效应

持久同调常被提及的一项优点是其对噪声的鲁棒性[90,91]。由于追踪的只是几何特征（即孔洞），只要噪声幅值没有破坏嵌入时间序列的底层结构，计算得到的同调就应当保持稳定。为了检验该性质在我们的方法中的表现，我们针对不同的信噪比，使用洛伦兹数据集重复分析，并观察由此得到的显著性评分曲线 $S_\mu(\tau)$ 的变化。在将输入的洛伦兹时间序列归一化为零均值和单位标准差后，加入了五种不同噪声水平的加性高斯噪声，其信噪比$SNR =$分别为(1000，500，100，50)。结果见图 14。我们发现，各条曲线都相对稳定，且

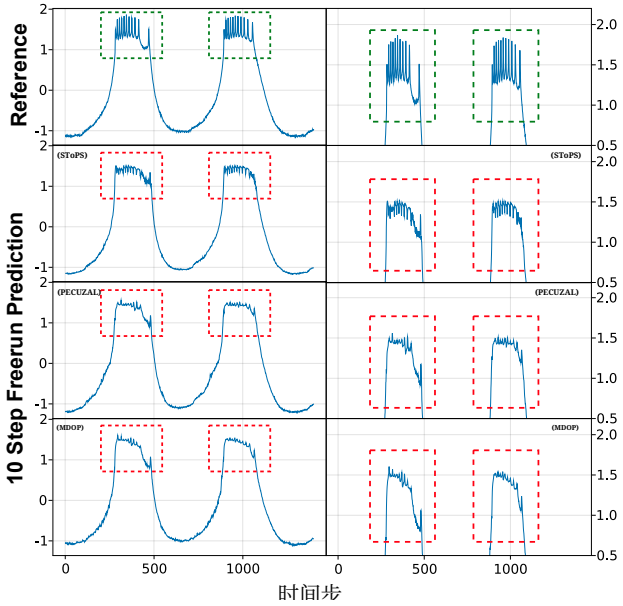


图13。比较了采用各种嵌入方法对龙虾 LP 神经元时间序列进行的 10 步自由运行预测。持久股线方法似乎更善于检测并复现快速的神经元尖峰动力学。预测的幅值已归一化为零均值和单位标准差。

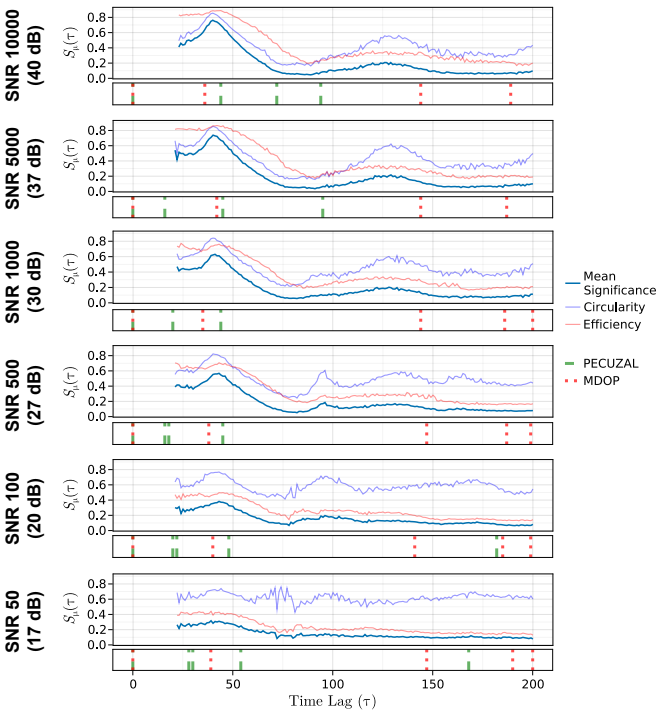


图14。比较了显著性评分 $S_\mu(\tau)$ 与不同加性噪声水平下 PECUZAL 和 MDOP 的时滞。SToPS 显著性评分相对稳定，在所有测试噪声水平下峰值位置均相同。低噪声水平下，PECUZAL 和 MDOP 保持稳定，但随着噪声水平升高，两者逐渐发生偏移。

随着噪声水平升高，结果仍具有鲁棒性，观察到的显著滞后几乎没有变化，甚至完全没有变化。在噪声水平较高时（> 20 dB），伪造的滞后伪影 $\tau = 120$ 消失了。这很可能是因为噪声影响了沿相似轨道运行的多条轨迹，从而破坏了底层同调结构。

同样，PECUZAL 和 MDOP 在低噪声水平（< 37 dB）下得出了相似的时滞。然而，噪声水平升高会导致这两种方法逐渐发生偏移。当噪声水平高于 30 dB 时，PECUZAL 得出的时滞在小τ 情况下也存在显著差异。相比之下，MDOP 受噪声影响相对较小，偏移效应也相对较弱。此外，只有 MDOP 能够对信噪比低于 17 dB 的信号给出时滞预测。SToPS 和 PECUZAL 均无法得出任何结果。

XIII. 计算复杂度

持久同调计算并不节省计算资源，且仍受计算扩展性较差的困扰。对于点数较少的数据集，持久同调计算相对较快；但即使数据集规模适中，计算时间也会迅速增长。这给 SToPS 方法带来了挑战，因为它需要对大量时间序列中的多个采样线段计算持久同调。根据所提出的算法，时间复杂度近似为：

$$\mathcal{O}_{SToPS} \approx \mathcal{O}(N_\tau N_s \phi_{PH}(N_p)),$$

其中，N_τ, N_s 和 N_p 分别表示测试的滞后数、采样线段数和线段长度，$\phi_{PH}(N_p)$表示持久同调计算的时间复杂度。该值取决于持久同调算法的实现方式以及所用过滤的类型（例如 Čech、Rips、Delaunay 等）。Vietoris-Rips 过滤的单纯复形大小为 K，其随 $2^{\mathcal{O}(N_p)}$ 呈指数级增长[92]。Vietoris-Rips 过滤的计算可分为两个阶段。不过，计算其复杂度界限并非易事[93]。值得注意的是，目前仍有工作致力于改进和优化持久同调算法，并已显著提升其性能[92,94]。

表 I 对 SToPS、PECUZAL 和 MDOP 的近似运行时间进行了比较。为确保结果具有可比性，计算在配备 16GB 随机存取存储器的 Ryzen 7 4800HS 上进行，且仅使用一个线程。尽管 SToPS 的计算时间远长于其他方法，但值得注意的是，SToPS 可以并行考察多个滞后，因为每个潜在时间滞后的显著性都是独立评估的。这与 PECUZAL 和 MDOP 不同，后两者采用迭代嵌入周期方法

表 I. 在洛伦兹和龙虾 LP 神经元时间序列（25000 个时间步）上，PECUZAL、MDOP 和 SToPS 在候选滞后的最大数量不同的情况下所需的计算时间（即 $\tau \in [1, \tau_{max}]$）。计算时间以秒为单位。列出了 PECUZAL 和 MDOP 计算得到的滞后。由于 SToPS 的结果以显著性评分曲线的形式给出，因此没有计算得到的滞后

时间序列	τ_{max}	PECUZAL	MDOP	SToPS
洛伦兹	50	6.0	0.19	148
		(45,23)	(42,23)	
	100	10.2	0.23	982
		(45,23)	(100,62)	
	150	15.8	0.28	1689
		(45,23)	(147,40)	
	200	21.2	0.42	2396
		(45,23)	(147,189,40)	
龙虾	50	7.1	0.39	95.9
		(50,24)	(50,31,11,44)	
	100	46.9	0.48	881
		(100,63,81,11,35)	(100,51,19,81)	
	300	112	0.82	3376
		(300,251,114)	(300,149,55,245)	
	600	2705	1.62	7266
		(515,304,132,505,32,20,253,394,81,232,55)	(544,274,599,114,423)	

并且每个新的时滞都是根据此前选定的时滞有条件地选择的。不过，Kr¨amer 等人提出了一种基于蒙特卡罗树搜索的新方法，试图解决这一问题[66]。独立评估嵌入滞后的灵活性，可能会显著提升计算速度，因为多个滞后的计算可以分配到多个线程上。我们当前的算法实现尚未支持这一功能。

根据先前已选滞后条件性地选择新滞后，也意味着 PECUZAL 和 MDOP 的结果对最大允许时滞 τ_{max}的变化并不稳健。潜在时滞范围的变化会影响后续滞后的选择顺序，因为如果引入一个能更好地优化目标统计量的新候选时滞，选择顺序就可能改变。对于复杂数据，最大时滞的变化可能导致结果出现很大差异。数据中的噪声也会导致滞后不同，如图14中发现的滞后漂移所示。研究发现，龙虾 LP 神经元时间序列的计算滞后甚至会因同分布噪声具有不同实现的试验而有所变化。

我们还注意到，这些嵌入算法的计算时间可能因输入时间序列的复杂程度而异。例如，PECUZAL 的连续嵌入圈通常会使计算时间逐渐增加。对于复杂或含噪的时间序列，PECUZAL 可能会产生多个显著性各异的时滞。从龙虾数据中最大时滞为 300 和 600 时估计时滞数量的跃升，可以看出这一点。相比之下，SToPS 算法会评估每个得分的显著性，其计算量应与时滞数量成线性比例增长。小型时滞 τ 是一个例外，此时时间片段过短，且没有可追踪的孔洞同调。

XIV. 结论

本文的目标之一是概述嵌入基础，并回顾现有的嵌入参数优化方法。我们首先在第一节中简要介绍嵌入定理的基本概念及其在时间序列分析中的应用。我们还为初学者或尚不熟悉该领域的读者提供了一个简单案例，说明嵌入在时间序列预测任务中的用法。本文聚焦于如何确定合适的嵌入参数，具体而言，是非均匀嵌入中嵌入滞后的选择。第三节概述了选择嵌入参数时需要考虑的因素，随后第四至第六节将全面回顾各种均匀延迟嵌入方法。

我们认为，非均匀嵌入方法在重构快慢动力系统时具有更大的灵活性。基于此，第VII节概述了现有的自动化方法，这些方法试图自动选择非均匀嵌入滞后。然而，尽管许多自动非均匀嵌入方法都能可靠地返回一组滞后，但它们未必彼此一致，也未必能为其选择提供令人满意的动力学解释。此外，由于选择延迟时采用了迭代过程，每个后续延迟的选择都取决于先前的选择。

我们在第 VIII 节和第 IX A 节提出了一种选择非均匀嵌入滞后的新方法——SToPS，旨在生成具有更强动力学解释力且彼此独立选取的滞后。SToPS 利用持久同调检测时间序列样本窗口的二维延迟嵌入所形成的环，我们称之为“持久链”。这一过程在样本窗口的多种不同长度上进行，以生成一个特征时间

谱 $S_\mu(\tau)$，其中显著性评分较大的值对应于动力学上显著的时间尺度（即它们与时间序列的某种周期性概念相关）。每个持久股环的结构由两个量来表征：圆度和效率；二者结合后得到显著性评分 $S_i(\tau)$。时间滞后τ 的选择基于显著性评分 $S_\mu(\tau)$ $S_\sigma(\tau)$的均值和标准差曲线。

我们在三类不同的时间序列上测试了 SToPS 嵌入方法：周期性（正弦和）、混沌（洛伦兹）和快慢（龙虾 LP 神经元）。在所有情况下，SToPS 都能检测到具有动力学可解释性的时间尺度，而这些时间尺度在其他参考非均匀嵌入方法 PECUZAL 和 MDOP 中均未体现。此外，在存在快慢动力学的龙虾 LP 神经元中，SToPS 在识别主导时间尺度方面优于 PECUZAL 和 MDOP。

我们还测试了不同嵌入方法对时间序列预测性能的影响。我们使用嵌入后的时间序列训练单步神经网络预测器。结果发现，对于洛伦兹和龙虾 LP 神经元时间序列，各种嵌入方法所得模型的表现相近。然而，使用 SToPS 嵌入滞后训练的模型，对龙虾 LP 神经元时间序列进行自由运行预测时，能够比参考嵌入方法更好地复现快速神经元尖峰动力学。我们还简要讨论并分析了 SToPS 的计算效率，以及其对含噪输入数据的鲁棒性。

总体而言，尽管 SToPS 的性能仅略优于现有的非均匀嵌入方法 PECUZAL 和 MDOP，但我们认为，与这些方法相比，SToPS 提供的时滞在动力学上更具可解释性。对各个时间滞后的评估还使该方法能够应用于多变量时间序列：将每个分量视为独立的标量时间序列，并识别其各自的时间滞后。随后可利用这些时间滞后构建一个使用时间序列所有分量的延迟向量。在这一背景下，SToPS 的性能尚未经过测试，这也为进一步研究提供了方向。此外，通过特征时间谱独立选择时滞，相较于现有的用于自动非均匀延迟嵌入的迭代方法，能够更清晰地展示每个时间滞后的相对重要性；在后者中，时滞的明确集合是预先给定的。

然而，追求动力学上更具可解释性的延迟，会使特征时间谱的解读带有一定主观性。尽管如此，我们认为，着重选择在动力学上相关且可解释的延迟，是一种更有意义的建模方法，有助于构建更贴近观测到的系统动力学的模型。这种优势在具有多个迥异时间和空间尺度的系统中尤为明显，龙虾 LP 神经元的动力学便是一个例子。因此，我们建议，在未来发展时延嵌入方法时，应将动力学相关性和可解释性作为重要的额外考量。

致谢

D.C.C. 获澳大利亚研究委员会资助，该资助通过由澳大利亚政府出资的通过数据科学转型维护中心（资助编号 IC180100030）提供。D.W. 和 M.S. 获澳大利亚研究委员会资助（资助编号 DP200102961）。S.D.A 获 Forrest 基金会资助。E.T. 获 Robert & Maude Gledden 基金会和 A.F. Pillow 应用数学信托基金资助。我们还要感谢 H. Abarbanel 及其团队为我们提供实验性 LP 神经元数据的访问权限。

数据可用性声明

本文引用的所有相关数据均可公开获取。本文提供了人工数据的生成方法。支持本文研究结果的其他实验数据，可向通讯作者合理索取。包含 SToPS 分析方法的 Julia 脚本也可在以下网址获取：
https://github.com/eugenetkj98/SToPS_Public

参考文献

[1]H. Whitney, 《可微流形》, 《数学年刊》, 645–680（1936）。[2]F. Takens, 《湍流中的奇异吸引子检测》, 载于 *Dynamical systems and turbulence, Warwick 1980*（施普普林格, 1981）, 第366–381页。[3]T. Sauer、J. A. Yorke和M. Casdagli, 《嵌入学》, 《统计物理学杂志》**65**, 579–616（1991）。[4]N. H. Packard、J. P. Crutchfield、J. D. Farmer和R. S. Shaw, 《从时间序列中得到的几何结构》, 《物理评论快报》**45**, 712（1980）。[5]H. D. Abarbanel、T. Carroll、L. Pecora、J. Sidorowich和L. S. Tsimring, 《时间延迟嵌入中的物理变量预测》, 《物理评论 E》**49**, 1840（1994）。[6]D. S. Broomhead和G. P. King, 《从实验数据中提取定性动力学》, 《物理学 D：非线性现象》**20**, 217–236（1986）。[7]A. M. Lyapunov, 《运动稳定性的一般问题》, 《国际控制杂志》**55**, 531–534（1992）。[8]A. Wolf、J. B. Swift、H. L. Swinney和J. A. Vastano, 《从时间序列中确定李雅普诺夫指数》, 《物理学 D：非线性现象》**16**, 285–317（1985）。[9]P. Grassberger和I. Procaccia, 《奇异吸引子的表征》, 《物理评论快报》**50**, 346（1983）。[10]P. Grassberger和I. Procaccia, 《测量奇异吸引子的奇异性》, 载于 *The theory of chaotic attractors*（施普林格, 2004）, 第170–189页。

[11]H. Kantz和T. Schreiber, *Nonlinear time series analysis*, 第7卷（剑桥大学出版社，2004）。[12]J. Guckenheimer和P. Holmes, *Nonlinear oscillations dynamical systems and bifurcations of vector fields* 42 第（施普林格科学与商业媒体公司，2013）卷。[13]N. A. Gershenfeld和N. Gershenfeld, *The nature of mathematical modeling* 1999

（剑桥大学出版社，）。[14]R. Govindan、K. Narayanan和M. Gopinathan, 《心电图中确定性混沌的证据：替代数据与可预测性分析》, Chaos: An Interdisciplinary Journal of Nonlinear Science 8，495–502（1998）。[15]A. S. Weigend和N. A. Gershenfeld, 《圣菲研究所时间序列预测竞赛结果》, 载于 *IEEE International Conference on Neural Networks* 1993

（IEEE, ）, 第1786–1793页。[16]M. Sangiorgio和F. Dercole, 《用于混沌时间序列多步预测的LSTM神经网络的鲁棒性》, 《混沌、孤子与分形》139，110045（2020）。[17]J. S. Zhang和X. C. Xiao, 《使用循环神经网络预测混沌时间序列》, 《中国物理快报》17，88–90（2000）。[18]A. Haluszczynski和C. R¨ath, 《优良与不良预测：通过储备池计算评估并改进混沌吸引子的复现》, Chaos: An Interdisciplinary Journal of Nonlinear Science 29，103143（2019）。[19]H. Jaeger, 《用于分析和训练循环神经网络的“回声状态”方法——附勘误说明》, 德国波恩：德国国家信息技术研究中心 Gesellschaft f¨ur Mathematik und Datenverarbeitung mbH（GMD）技术报告148，13（2001）。[20]J. Pathak、Z. Lu、B. R. Hunt、M. Girvan和E. Ott, 《使用机器学习复现混沌吸引子并从数据中计算李雅普诺夫指数》, Chaos: An Interdisciplinary Journal of Nonlinear Science 27，121102（2017）。[21]K.-R. M¨uller、A. J. Smola、G. R¨atsch、B. Sch¨olkopf、J. Kohlmor-gen和V. Vapnik, 《使用支持向量机预测时间序列》, 载于 *International Conference on Artificial Neural Networks* 1997 999 1004

（施普林格，）, 第–页。[22]N. I. Sapankevych和R. Sankar, 《使用支持向量机进行时间序列预测：综述》, 《IEEE计算智能杂志》4，24–38（2009）。[23]C. J. Cellucci、A. M. Albano和P. Rapp, 《嵌入方法的比较研究》, 《物理评论 E》67，066210（2003）。[24]J. Theiler、S. Eubank、A. Longtin、B. Galdrikian和J. D. Farmer, 《时间序列非线性测试：替代数据方法》, 《物理学 D：非线性现象》58，77–94（1992）。[25]A. M. Fraser和H. L. Swinney, 《基于互信息的奇异吸引子独立坐标》, 《物理评论 A》33，1134（1986）。[26]M. B. Kennel和H. D. Abarbanel, 《伪近邻与伪线段：一种可靠的最小嵌入维数算法》, 《物理评论 E》66，026209（2002）。[27]L. M. Pecora、L. Moniz、J. Nichols和T. L. Carroll, 《吸引子重构的统一方法》, Chaos: An Interdisciplinary Journal of Nonlinear Science 17，013110（2007）。[28]M. Casdagli、S. Eubank、J. D. Farmer和J. Gibson, 《存在噪声时的状态空间重构》, 《物理学 D：非线性现象》51，52–98（1991）。[29]L. C. Uzal、G. L. Grinblat和P. F. Verdes, 《动力系统的最优重构：噪声放大方法》, 《物理评论 E》84，016223（2011）。[30]T. Buzug和G. Pfister, 《用于延迟时间坐标的最优嵌入参数计算算法比较》, 《物理学 D：非线性现象》58，127–137（1992）。[31]C. Bandt和B. Pompe, 《排列熵：时间序列的自然复杂度度量》, 《物理评论快报》88，174102 2002（）。[32]M. McCullough、M. Small、T. Stemler和H. H.-C. Iu, 《用于捕捉动力学的时间滞后序数分区网络》, 连续动力系统”, 《混沌：跨学科非线性科学杂志》25，053101（2015）。[33]J. Zhang、J. Zhou、M. Tang、H. Guo、M. Small 和 Y. Zou, “从多变量时间序列构建序数分区转移网络”, 《科学报告》7，1–13（2017）。[34]B. Schrauwen、D. Verstraeten 和 J. Van Campenhout, “储备池计算概述：理论、应用与实现”, 载于 *Proceedings of the 15th European Symposium on Artificial Neural Networks. p. 471-482 2007* 2007 471 482 ( )

第 – 页。[35]R. Gilmore, “ 混沌动力系统的拓扑分析 ”, 《现代物理评论》70，1455（1998）。[36]J. Lever、M. Krzywinski 和 N. Altman, “重要观点：主成分分析”, 《自然·方法》14，641–643（2017）。[37]A. I. Mees、P. Rapp 和 L. Jennings, “奇异值分解与嵌入维数”, 《物理评论 A》36，340（1987）。[38]M. Palu˘s 和 I. Dvo˘r´ak, “吸引子重构中的奇异值分解：陷阱与注意事项”, 《物理学 D：非线性现象》55，221–234（1992）。[39]A. Potapov, “混沌吸引子重构的畸变”, 《物理学 D：非线性现象》101，207–226（1997）。[40]C. Nichkawde, “利用投影流形上的导数进行最优状态空间重构”, 《物理评论 E》87，022905（2013）。[41]J. F. Gibson、J. D. Farmer、M. Casdagli 和 S. Eubank, “实用状态空间重构的解析方法”, 《物理学 D：非线性现象》57，1–30（1992）。[42]J. Gao 和 Z. Zheng, “局部指数发散图与混沌时间序列的最优嵌入”, 《物理快报 A》181，153–158（1993）。[43]J. Gao 和 Z. Zheng, “确定性混沌的直接动力学检验”, 《欧洲物理快报》（EPL）25，485–490（1994）。[44]W. Liebert、K. Pawelzik 和 H. Schuster, “基于拓扑考量的混沌吸引子最优嵌入”, 《欧洲物理快报》（EPL 14，521（1991）。[45]R. Hegger 和 H. Kantz, “用于检测时间序列数据中确定性的改进假最近邻法”, 《物理评论 E》60，4970–4973（1999）。[46]M. B. Kennel、R. Brown 和 H. D. I. Abarbanel, “利用几何构造确定相空间重构的嵌入维数”, 《物理评论 A》45，3403–3411（1992）。[47]S. P. Garcia 和 J. S. Almeida, “采用不同时间延迟的最近邻嵌入”, 《物理评论 E》71，037204（2005）。[48]J. Theiler, “应用于有限时间序列数据的相关算法产生的伪维数”, 《物理评论 A》34，2427–2432（1986）。[49] J. Theiler, “估计分形维数”, 《美国光学学会杂志 A》7，1055–1073（1990）。[50]J. L. Kaplan 和 J. A. Yorke, “多维差分方程的混沌行为”, 载于 *Functional differential equations and approximation of fixed points* 1979 204 227

（施普林格，）第 – 页。[51]J. D. Farmer、E. Ott 和 J. A. Yorke, “混沌吸引子的维数”, 《物理学 D：非线性现象》7，153–180（1983）。[52] E. Ott、T. Sauer 和 J. A. Yorke, 《应对混沌：混沌数据分析与混沌系统的利用》, 威利非线性科学丛书（1994）。[53]J. M. Moore、D. M. Walker 和 G. Yan, “平均局部自协方差为基于高频采样类流数据的重构提供稳健而灵活的延迟选择”, 《物理评论 E》101，012214（2020）。[54]H. D. Abarbanel、R. Brown、J. J. Sidorowich 和 L. S. Tsimring, “物理系统中观测到的混沌数据分析”, 《现代物理评论》65，1331（1993）。[55]S. Wallot 和 D. Mønster, “在 MATLAB 中计算平均互信息（AMI）和假最近邻（FNN）, 以估计多维时间序列的嵌入参数”, 《心理学前沿》9，1679（2018）。

[56]A. Kraskov、H. St¨ogbauer 和 P. Grassberger，“互信息估计”，《物理评论 E》69，066138（2004）。[57]A. Papana 和 D. Kugiumtzis，“时间序列互信息估计器的评估”，《国际分岔与混沌杂志》19，4197–4215（2009）。[58]G. A. Darbellay 和 I. Vajda，“通过对观测空间进行自适应划分来估计信息”，《IEEE 信息论汇刊》45，1315–1321（1999）。[59]Y.-I. Moon、B. Rajagopalan 和 U. Lall，“使用核密度估计器估计互信息”，《物理评论 E》52，2318（1995）。[60] K. Judd 和 A. Mees，“将嵌入视为建模问题”，《物理学 D：非线性现象》120，273–286（1998）。[61]Y. Hirata、H. Suzuki 和 K. Aihara，“利用可变延迟从多变量数据重构状态空间”，《物理评论 E》74，026202（2006）。[62]H. D. Abarbanel、R. Brown 和 M. B. Kennel，“奇异吸引子上的李雅普诺夫指数变化”，《非线性科学杂志》1，175–199（1991）。[63]S. D. Algar、T. Stemler 和 B. De Saedeleer，“噪声诱发的同步模式间跃迁动力学”，《国际分岔与混沌杂志》25，1530034（2015）。[64]K. H. Kr¨amer、G. Datseris、J. Kurths、I. Z. Kiss、J. L. Ocampo-Espindola 和 N. Marwan，“吸引子重构的统一自动化方法”，《新物理学杂志》23，033017（2021）。[65]M. Shen、W.-N. Chen、J. Zhang、H. S.-H. Chung 和 O. Kaynak，“利用蚁群优化选择混沌时间序列非均匀嵌入的最优参数”，《IEEE 控制论汇刊》43，790–802（2013）。[66]K. H. Kr¨amer、M. Gelbrecht、I. Pavithran、R. Sujith 和 N. Marwan，“通过蒙特卡洛决策树搜索实现最优状态空间重构”，《非线性动力学》108，1525–1545（2022）。[67]D. A. Rand 和 L.-S. Young，*Dynamical Systems and Turbulence Warwick 1980: Proceedings of a Symposium Held at the*，*University of Warwick 1979/80* 898 2006

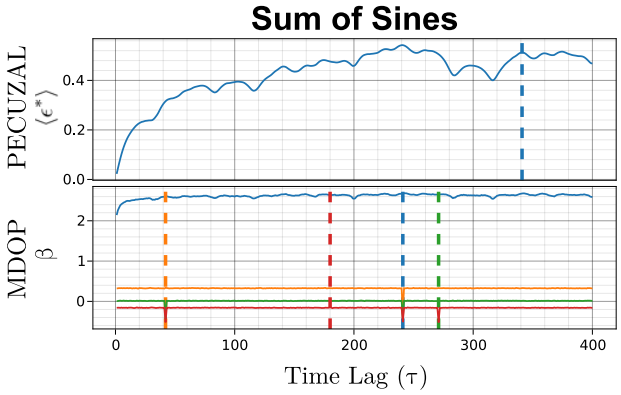
第 卷（施普林格，）。[68]M. Small 和 K. Judd，“利用线性建模技术检测实验数据中的周期性”，《物理评论 E》59，359 376 1999–（）。[69] M. Small 和 C. K. Tse，“时间序列建模中嵌入参数的最优选择”，载于*European Conference on Circuits Theory and Design (European Circuit Society and the Institute of Electrical and Electronic Engineers* 2003

（）。[70]M. Small 和 C. K. Tse，“最优嵌入参数：一种建模范式”，《物理学 D：非线性现象》194，283 296 2004–（）。[71]J. Rissanen，“MDL去噪”，《IEEE 信息论汇刊》46，2537–2543（2000）。[72]T. Nakamura、K. Judd、A. I. Mees 和 M. Small，“模型选择信息准则的比较研究”，《国际分岔与混沌杂志》16，2153–2175（2006）。[73] M. Dorigo、V. Maniezzo 和 A. Colorni，“蚁群系统：协作智能体群体的优化方法”，《IEEE 系统、人和控制论汇刊 B辑（控制论）》26，29 41 1996–（）。[74]M. Dorigo 和 L. M. Gambardella，“蚁群系统：一种解决旅行商问题的协作学习方法”，《IEEE 进化计算汇刊》1，53–66（1997）。[75]H. Adams、M. Aminian、E. Farnell、M. Kirby、J. Mirth、R. Neville、C. Peterson 和 C. Shonkwiler，“基于持久同调的测度分形维数”，载于 *Topological Data Analysis* 2020

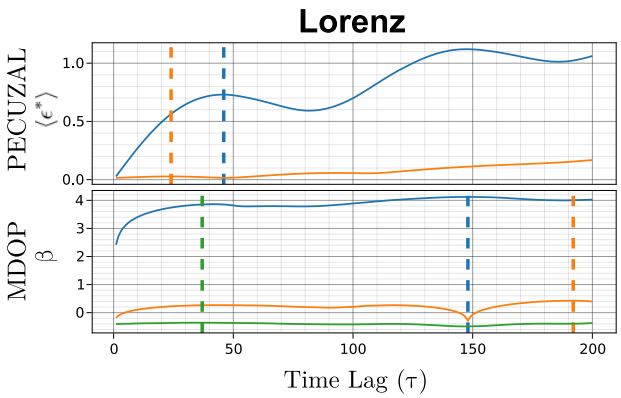
（施普林格，）。[76]J. Jaquette 和 B. Schweinhart，“利用持久同调估计分形维数：一项比较研究”，《非线性科学与数值模拟通讯》84，105163（2020）。[77]F. A. Khasawneh 和 E. Munch，“利用持久同调进行车削颤振检测”，《机械系统与信号处理》70，527–541（2016）。[78]A. Myers、E. Munch 和 F. A. Khasawneh，“用于动态状态检测的复杂网络持久同调”，《物理评论 E》**100**，22314（2019），1904.07403。[79]E. Tan、D. Corrˆea、T. Stemler 和 M. Small，“为模型评分：使用持久同调评估模型的动力学学习”，Chaos: An Interdisciplinary Journal of Nonlinear Science **31**，123109（2021）。[80]A. D. Myers 和 F. A. Khasawneh，“使用拓扑数据分析选择排列熵的延迟参数”，（2022），arXiv:1905.04329。[81]L. Vietoris，“¨论紧致空间的高阶连通性及一类保持连通性的映射”，Mathematische Annalen **97**，454–472（1927）。[82]A. Hatcher，*Algebraic Topology*（剑桥大学出版社，2002）。[83]S. Maleti´c、Y. Zhao 和 M. Rajkovi´c，“动力系统的持久拓扑特征”，Chaos: An Interdisciplinary Journal of Nonlinear Science **26**，053105（2016）。[84]N. Marwan、M. C. Romano、M. Thiel 和 J. Kurths，“用于复杂系统分析的递归图”，Physics Reports **438**，237–329（2007）。[85]B. Braden，“测量员面积公式”，The College Mathematics Journal **17**，326–337（1986）。[86]R. L. Graham，“确定有限平面点集凸包的高效算法”，Information Processing Letters **1**，132–133（1972）。[87]H. D. Abarbanel、R. Huerta、M. I. Rabinovich、N. F. Rulkov、P. F. Rowat 和 A. I. Selverston，“突触耦合混沌模型神经元的同步活动”，Neural Computation **8**，1567 1602 1996–（）。[88] D. Arthur 和 S. Vassilvitskii，“k-means++：精心选择种子选取的优势”，技术报告（斯坦福，2006）。[89]G. Datseris，“Dynamicalsystems.jl：用于混沌和非线性动力学的 Julia 软件库”，Journal of Open Source Software **3**，598（2018）。[90]H. Adams、T. Emerson、M. Kirby、R. Neville、C. Peterson、P. Shipman、S. Chepushtanova、E. Hanson、F. Motta 和 L. Ziegelmeier，“持久图像：持久同调的稳定向量表示”，Journal of Machine Learning Research **18** （2017）。[91]R. Turke˘s、J. Nys、T. Verdonck 和 S. Latr´e，“灰度图像上持久同调的噪声鲁棒性：跨过滤和特征的分析”，PLOS One **16**，e0257215（2021）。[92]N. Otter、M. A. Porter、U. Tillmann、P. Grindrod 和 H. A. Harrington，“持久同调计算路线图”，EPJ Data Science **6**，1–38（2017）。[93]A. Zomorodian，“Vietoris-Rips复形的快速构造”，Computers & Graphics **34**，263–271（2010）。[94]U. Bauer，“Ripser：高效计算 Vietoris–Rips 持久条形码”，Journal of Applied and Computational Topology **5**，391–423（2021）。

附录A：非均匀嵌入剖面

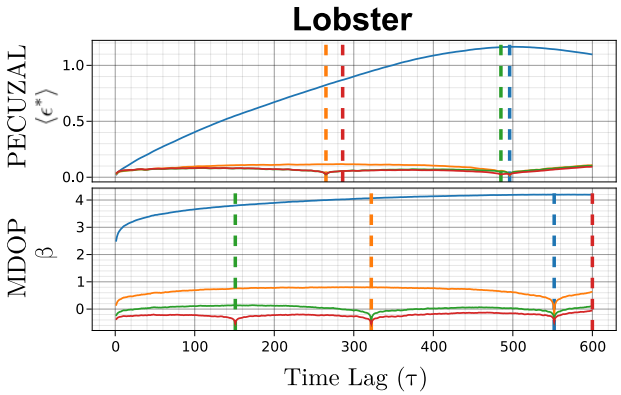
图15给出了用于选择自动非均匀嵌入算法 PECUZAL 和 MDOP 的嵌入滞后的计算统计量曲线集合。相关统计量分别是 PECUZAL 的连续性统计量 $\langle \epsilon^* \rangle$ 和 MDOP 的 β 统计量。对于 PECUZAL，与 $\langle \epsilon^* \rangle$ 中的局部极大值对应的 τ 滞后被用作候选时滞。在每个嵌入周期中，使 L-统计量下降幅度最大的时滞被选为最终嵌入时滞。MDOP 采用类似的过程，但选择 β的全局极大值。基于伪近邻（FNN）统计量的终止准则用于在



(a) 由 $\omega = \{1, 5, 30\}$、$\phi = \{0, 0.25, 0.75\}$ $A = \{1, 0.5, 0.2\}$ 和 . 构成的 3 个正弦项之和。



(b) 洛伦兹系统的第一个分量。



(c) 龙虾 LP 神经元时间序列

图15。自动非均匀嵌入方法比较中所计算统计量的嵌入剖面。

连接处。

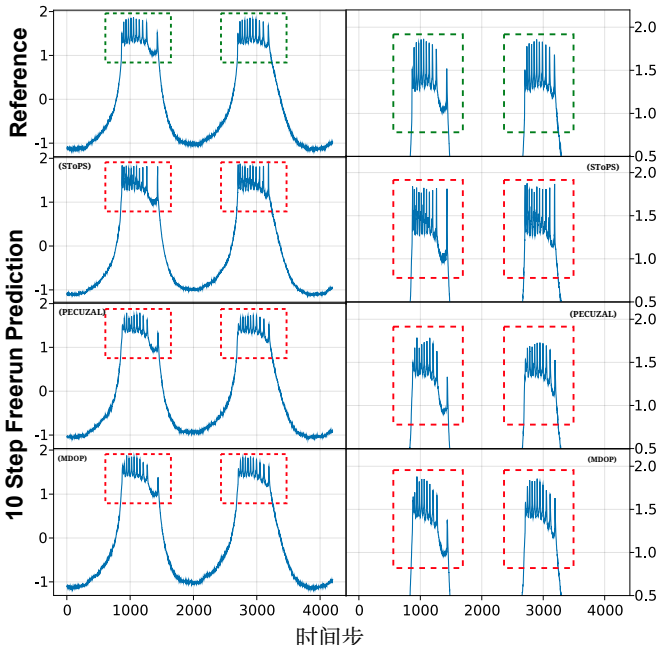
附录B：非子采样数据的自由运行预测

本节介绍了在完整的龙虾 LP 神经元数据集上训练的模型进行的额外自由运行预测测试结果，训练过程中未进行任何子采样。从理论上讲，这是一项更容易的任务，因为模型获得了更多的数据，并且每一步的预测幅值更小。测试了 10 步（图16）和 30 步（图17）的预测时域。后者的预测时域等于在经过子采样的数据上训练的模型的 10 步预测时域。在 10 步非子采样情况下，STOPS 的预测误差低于 PECUZAL 和 MDOP。所有方法对快速脉冲动力学的复现相对相似。在 30 步预测情况下也得到了类似结果。然而，与 10 步预测情况相比，快速脉冲动力学的复现效果较差，这可能是由于连续预测值中误差的累积所致。

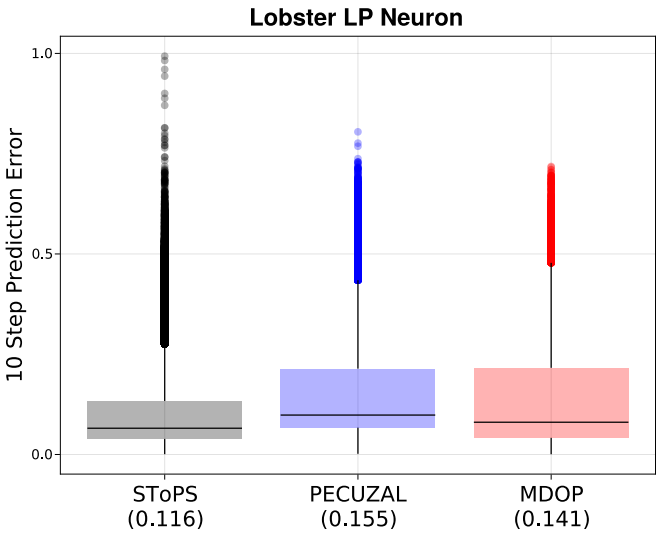
附录C：相空间重构比较

图18和图19给出了正弦和与洛伦兹时间序列的各种重构吸引子比较，并列出了检测到的前两个时滞。对于周期性正弦和，PECUZAL仅得到一个时滞。对于洛伦兹时间序列，SToPs仅在 $\tau = 41$ 处得到一个峰值，随后使用该峰值进行均匀嵌入。PECUZAL和SToPS均具有相似的时滞 $\tau \approx 40$。MDOP得到的时滞会导致过度嵌入。

图20、图21和图22分别给出了正弦和、洛伦兹时间序列以及龙虾 LP 神经元时间序列对应的二维和三维嵌入的复现图。所有嵌入方法都能够保留正弦和数据集的部分周期结构。然而，MDO和SToPS的复现图显示出比PECUZAL更多的小尺度结构。对于洛伦兹时间序列，PECUZAL和SToPS得到的复现图相似。相比之下，MDOP的复现图丢失了一些细节，这可能是由于较大的嵌入滞后导致吸引子过度折叠。在龙虾 LP 神经元时间序列中，使用SToPS选择较小的时滞，在复现图的若干对角线区域显示出预期的周期行为。在PECUZAL和MDOP嵌入中，这一现象并不明显，因为检测到的周期行为被缓慢的神经元动力学所掩盖。

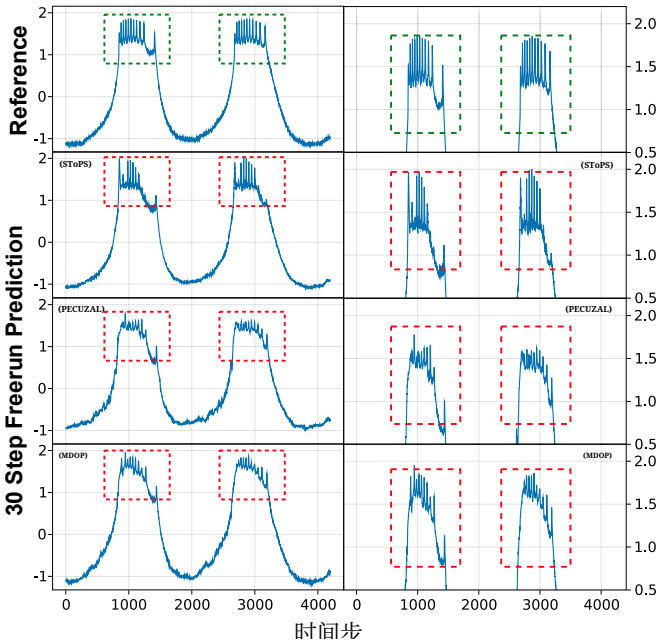


(a) 自由运行预测轨迹的比较。

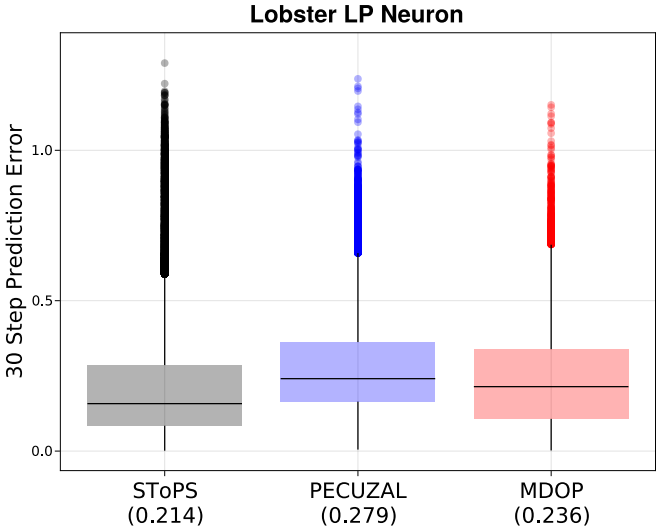


(b) 10步平均预测误差的分布

图16. 在非子采样数据上训练的神经网络模型的10步自由运行预测结果。



（a）自由运行预测轨迹的比较。



（b）30步平均预测误差的分布。

图17。在非子采样数据上训练的神经网络模型的30步自由运行预测结果。

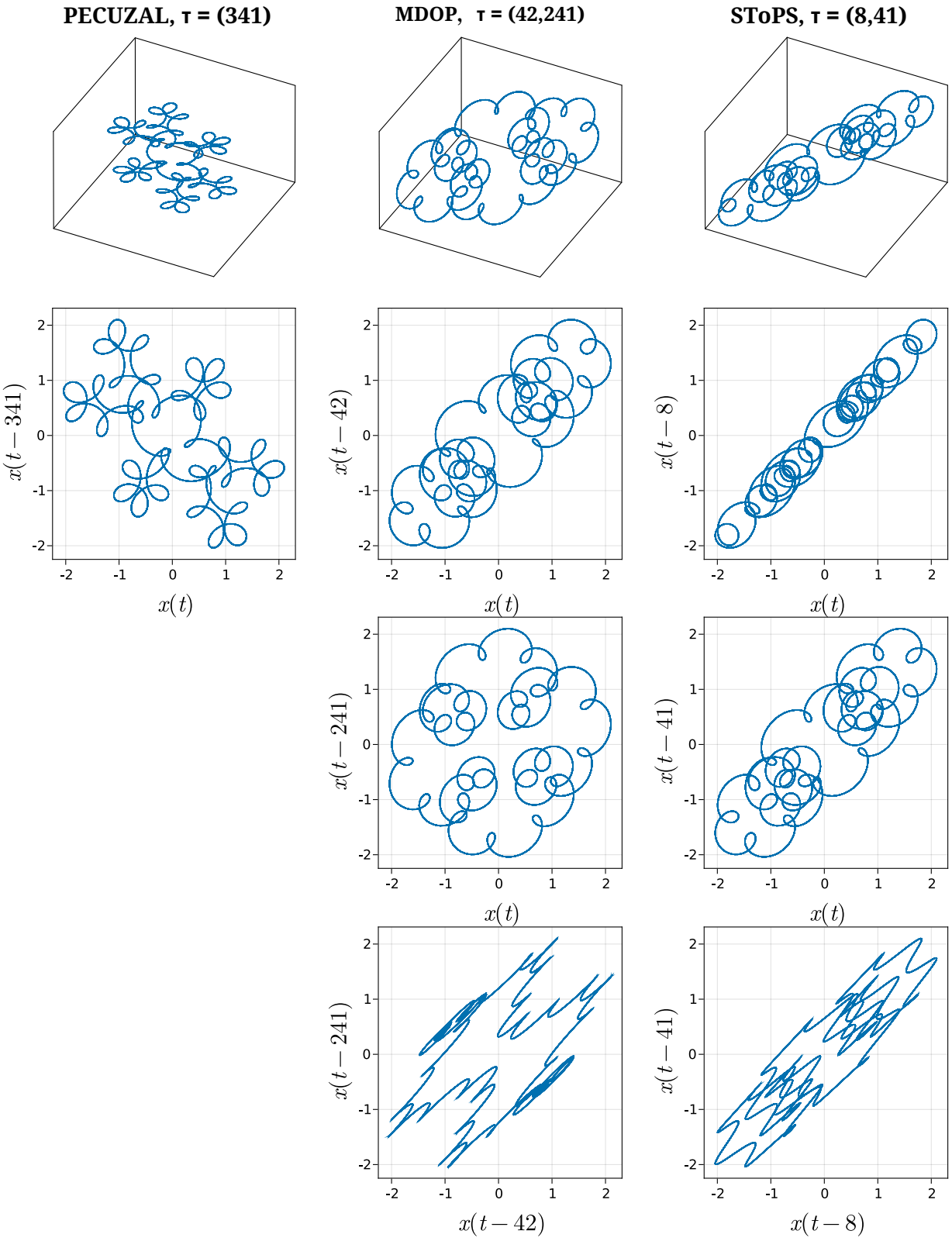


图18　。正弦和（$\omega = 1,\ 5,\ 30$）使用 PECUZAL、MDOP 和 SToP 进行的相空间重构　S.

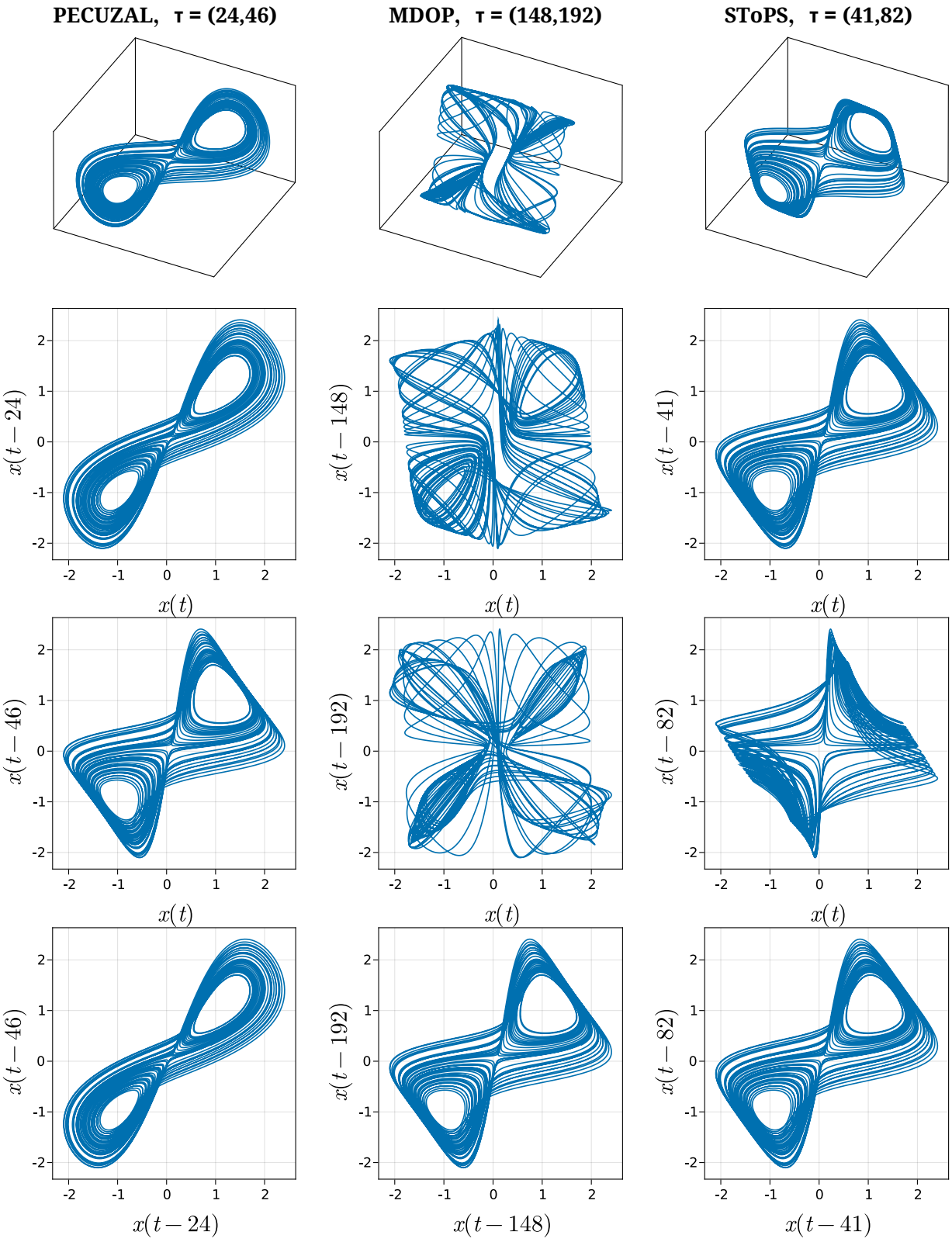


图 19. 洛伦兹相空间重构：使用 PECUZAL、MDOP 和 STOPS。

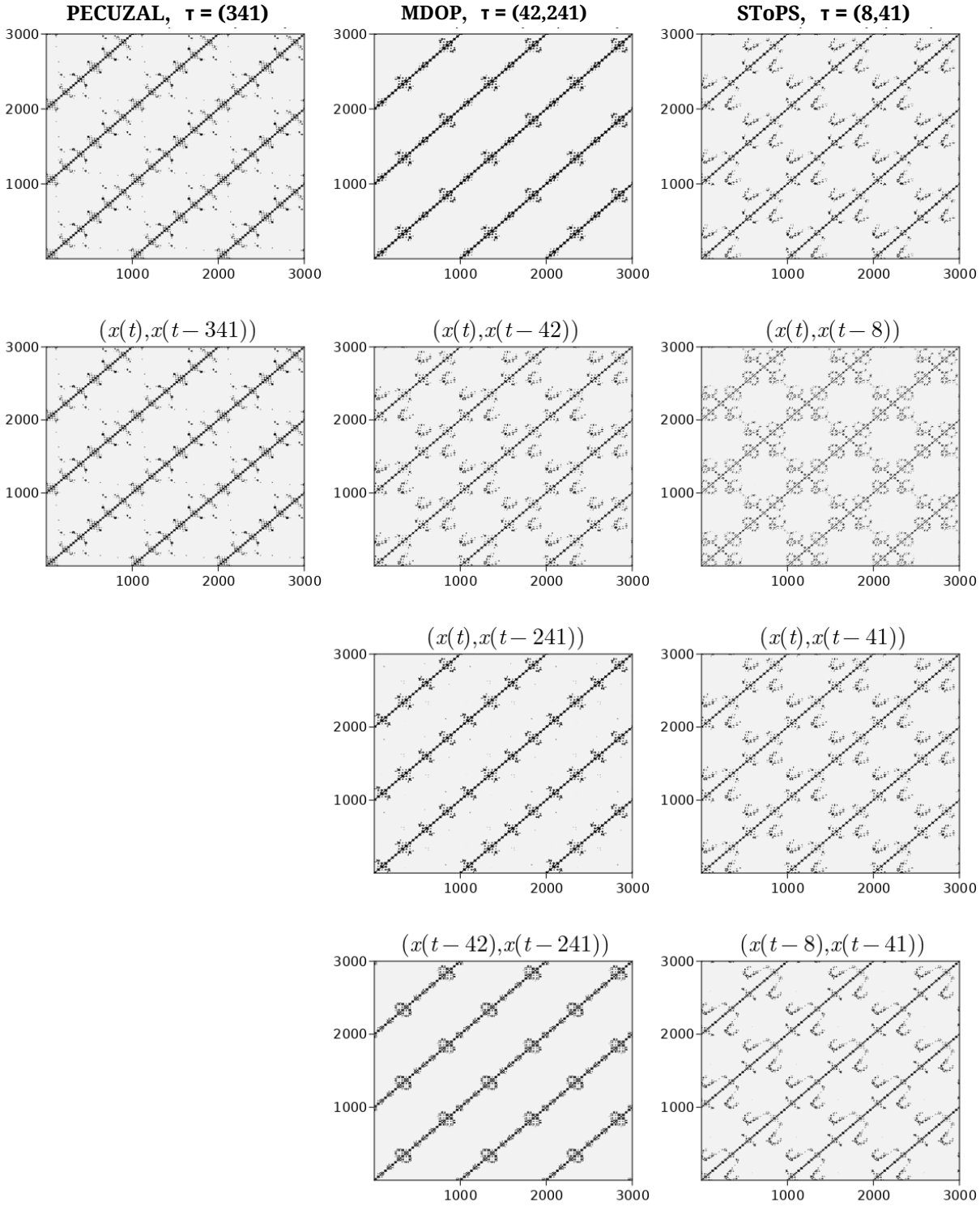


图 20。正弦和（$\omega = 1$，5，30）二维和三维嵌入的复现图。

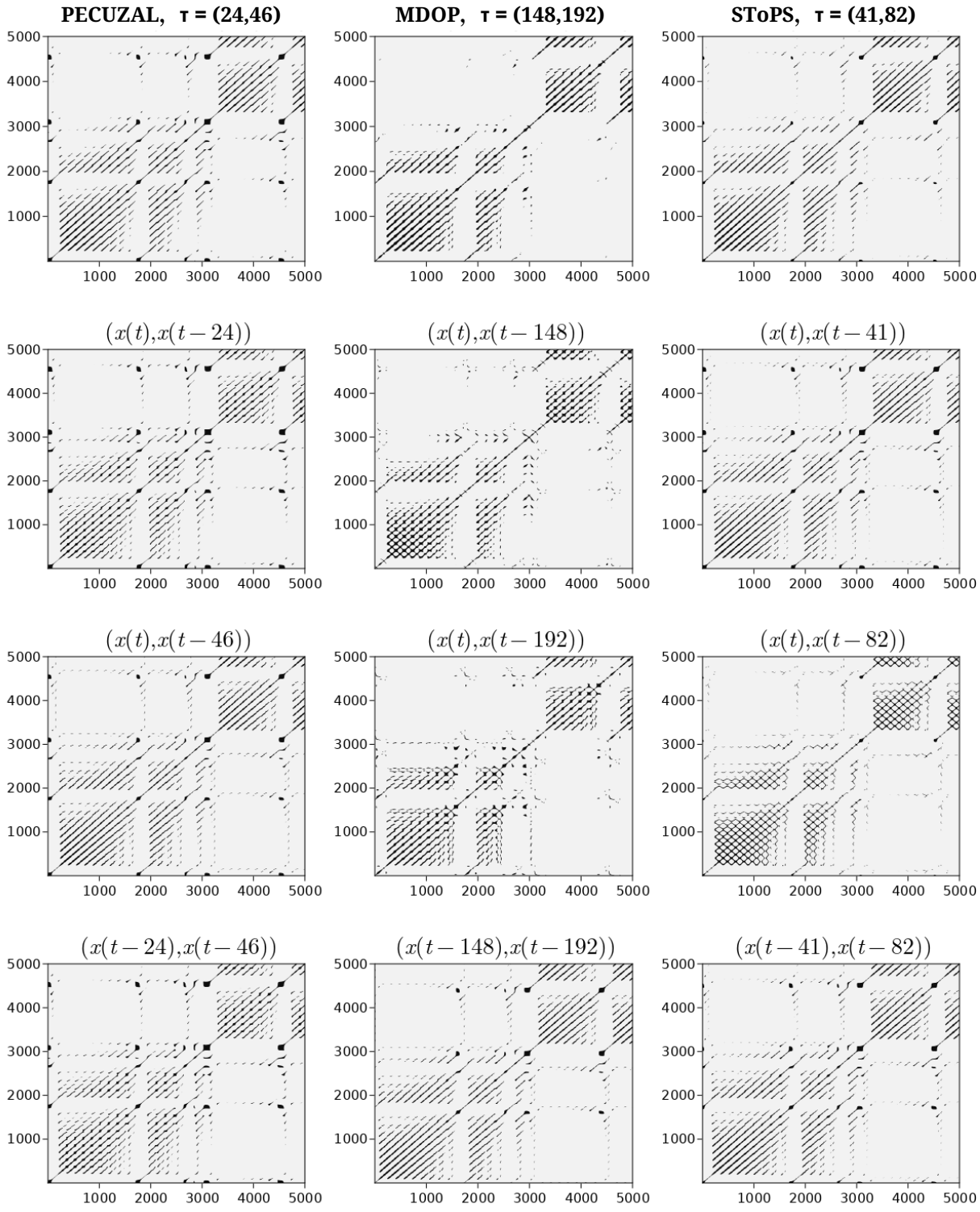


图21。 二维和三维嵌入的洛伦兹递归图。

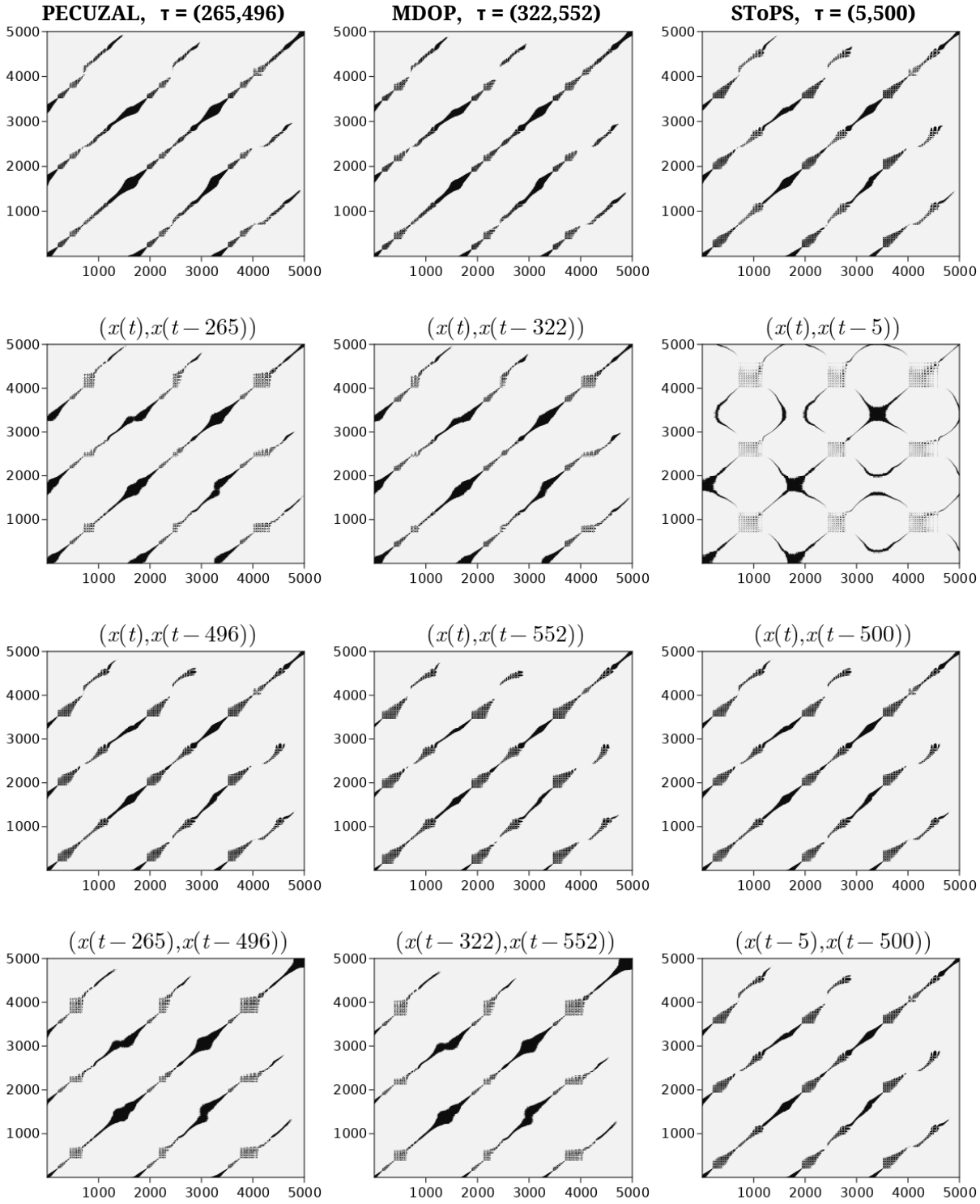


图 22. 龙虾 LP 神经元二维和三维嵌入的复现图。