arXiv:2601.00172v1 [cs.LG] 1 Jan 2026

用于高效高维时空预测的序列储备池计算 *

Ata Akbari Asanjan,FilipWudarskiUSRA 高级计算机科学研究所（RIACS） 加利福尼亚州 {aakbariasanjan, fwudarski}@usra.edu

Daniel O'Connor,Shaun Geaney,Elena Strbac渣打银行, Basinghall大道1号，伦敦，英国 {daniel.oconnor, shaun.geaney, elena.strbac}@sc.com

P.Aaron LottUSRA 高级计算机科学研究所（RIACS） 加利福尼亚州 plott@usra.edu

Davide VenturelliUSRA 高级计算机科学研究所（RIACS） 加利福尼亚州 dventurelli@usra.edu

摘要

对于循环神经网络（RNN）和长短期记忆网络（LSTM）模型而言，高维时空系统的预测在计算上仍然面临挑战，这是因为它们采用基于梯度的训练方法，且存在内存瓶颈。储备池计算（RC）通过固定循环层和凸读出层优化取代反向传播，缓解了这些挑战；然而，传统储备池计算架构在输入维度增大时仍难以有效扩展。我们提出了一种序贯储备池计算（Sequential RC）架构，将一个大型储层分解为一系列相互连接的小型储层。这种设计在保留长期时间依赖关系的同时，降低了内存和计算开销。通过对低维混沌系统（Lorenz63）和高维物理模拟（二维涡量方程和浅水方程）进行实验，序贯储层计算与长短期记忆网络和标准循环神经网络基线模型相比，实现了长 15%–25% 的有效预测时域、低 20%–30% 的误差指标（结构相似性指数（SSIM）、均方根误差（RMSE）），训练成本最高降低了三个数量级。结果表明，序贯储层计算在保持传统储备池计算简洁高效的同时，对于高维动力系统具有更优的可扩展性。这种方法为科学和工程应用中的实时、高能效预测提供了一条切实可行的路径。

关键词： 储备池计算 · 时间序列预测 · 序贯架构 · 高维数据

*引用：**Akbari Asanjan, A., Wudarski,F., O'Connor, D., Geaney, S., Strbac, E.,Lott, P. A.,& Venturelli, D.《用于高效高维时空预测的序列储备池计算》，arXiv 预印本，arXiv:2301.00001**。

1 引言

近年来，机器学习技术的迅速发展推动科学研究领域发生了显著变革。这些进展为我们开启了一个充满可能性的新时代，使我们能够理解和预测广泛研究领域中的复杂现象 [1, 2, 3]。值得注意的是，提升这些前沿模型性能的主流趋势，是不断增加其架构的复杂度。这通常意味着参数数量不断增加，而参数是影响模型捕捉数据中复杂模式和关系能力的关键因素 [4, 5, 2, 1]。先进架构在处理复杂数据集、提取细微模式方面表现出色，而这些任务曾令传统方法难以应对。然而，架构日益复杂也带来了一个值得关注的后果：对计算资源的需求增加，尤其是内存和处理能力方面 [6]。因此，要确保科学界能够在各个学科领域充分发挥机器学习进步的潜力，关键在于兼顾对复杂架构的追求与计算效率。

时间序列分析领域架构日益复杂的一个显著例证，是模型从 Elman型循环神经网络（RNN） [7]发展到长短期记忆网络（LSTM） [2] 和门控循环单元（GRU） [8]等门控结构的演进。这种循环模型架构的改进，已被证明有助于提升模型理解较长序列的能力，并缓解与次优学习方案相关的训练难题，例如梯度消失和梯度爆炸 [9, 10, 7]。
门控循环模型通过引入门控单元来调节流入循环模型核心单元的信息流，有效解决了学习问题。这一机制能够保留关键信息，同时过滤掉无关细节 [2, 7]。然而，这种改进的代价是参数数量大幅增加——每个循环层中的参数约为原来的三到四倍 [11]。尽管存在这种权衡，门控循环模型克服固有训练困难的能力有所提升，因此已成为时间序列分析中的关键工具，也体现了一个持续发展的趋势：在提升模型复杂度的同时，兼顾计算效率和训练效果等实际因素。

储备池计算（RC）近年来再度兴起，凭借训练高效、预测准确等独特优势，重新受到关注。储备池计算的起源可追溯至开创性工作 [12] ，其提出了这一概念。这项先驱性工作引入了由随机且固定的脉冲神经元构成的网络，并将其配置为动态突触，同时搭配采用基于相关性的训练方法的可训练输出层。这项工作的启发为两种主要储备池计算框架的发展奠定了基础：Jaeger 的回声状态网络 [13] 和 Maas 的液态状态机 [14]。

在回声状态网络（ESN）这一储备池计算中最广为人知的架构中，作者阐述了回声状态性质，并指出模型的状态最终应仅取决于驱动输入信号。换言之，状态会成为输入的回声，而初始条件的影响应随时间推移逐渐减弱 [13]。这一性质通过输入层与储层构成的结构实现；储层由循环连接的神经元组成，通常采用随机且稀疏的连接方式。储层充当动态记忆系统，善于捕捉和处理输入数据中的时间依赖关系。ESN 架构（下文称为储备池计算）在输入层和储层中保留固定参数，同时训练最终的线性层，即“读出层”，并采用带 L2 范数正则化的最小二乘误差进行训练 [13]。

众多研究工作深入探索了储备池计算（RCs）的多种应用，尤其关注其在物理系统中的实用性。在这一领域，研究重点主要集中于利用储备池计算预测混沌系统的行为，例如多项研究对洛伦兹系统进行了考察 [15, 16, 17]。此外，大量研究致力于利用储备池计算预测时空模式。其中一些研究拓展了储备池计算的应用范围，将预测任务从低维扩展至半高维。例如，有研究应对了具有空间相关输入变量的数据集所带来的挑战，如库拉莫托-西瓦辛斯基数据集；在该数据集中，复杂关系存在于单一维度中 [18, 19, 20]。与此同时，另一些研究进一步探索了完整的二维空间数据集 [21], , 从而深化了对储备池计算处理复杂空间关系能力的理解。这些研究的总体发现有助于细致认识储备池计算的多功能性与适用性：它不仅能够预测混沌系统，还能应对不同维度下时空预测任务所带来的挑战。尽管储备池计算具有良好的有效性，研究人员仍遇到了一个阻碍其在高维数据集中得到全面应用的重大瓶颈，即在此类复杂问题上有效训练储备池计算所必需的指数级内存需求 [15]。这一挑战以及更高的计算需求凸显了进一步解决内存相关限制的必要性，以释放储备池计算处理高维数据集的全部潜力。

在传统深度学习发展的连续进程中，储备池计算不断涌现出多种变体，利用分层架构来揭示多层次时间特征 [22, 23], ，通过多个读出层实现更为复杂的状态到预测映射 [24, 25] ，或串联多个输入 - 储层 - 读出模块 [26]。多项研究 [23, 22, 27] 表明，在深层储层计算架构中增加带有非线性激活的储层层数，能够捕捉跨越不同时间尺度的依赖关系。据报道，深层架构能够增强输入时间序列数据进行神经编码时形成的内部表示，从而丰富最终模型的深度与复杂性。

尽管序贯储层计算已被证明能够有效捕捉各种时间信息，但尚无研究探讨该架构在处理时空模式方面的优势；在这类任务中，模型需要学习多层次的空间和时间特征。鉴于序贯储层计算的应用，并认识到对更精细时空能力的需求日益增长，我们研究了这一问题，同时考察深层储层计算（下文称为序贯储层计算）模型的硬件效率，并将其与原始储层计算模型进行比较。
我们在两种不同的低维和高维数据设置下，对基线序贯储层计算和所提出的序贯储层计算进行了实验。我们研究了一种名为 Lorenz63 的低维数据，它是一组用于表示简化大气对流模型的微分方程。Lorenz63 数据属于混沌数据集，小小的当前状态误差就可能导致近期未来状态出现显著误差 [28]。为评估模型在高维基准数据上的表现，我们采用了全面的方法，使用了两种不同的纳维-斯托克斯模拟。具体而言，这些模拟聚焦于涡量动力学和浅水方程，两者均在二维配置中实现。这些基准数据集使我们得以全面探索和分析基于物理的行为，并在复杂的多维场景中对其性能进行稳健且深入的评估。

本文后续各节安排如下：第 2 节介绍基线模型和所提出模型的方法。第 3 节介绍用于低维和高维基准测试的数据集。第 4 节展示实验结果并进行分析，阐明模型架构在各种设置下的优势与不足。最后，第 5 节总结本文的研究发现。

2 方法

本节介绍本文采用的方法。首先介绍两种循环神经网络基线模型（Elman型和长短期记忆网络），随后介绍所提出的方法——储备池计算，并说明具体实现细节。

2.1 循环神经网络

Elman型循环神经网络（RNNs）通常称为普通循环神经网络，是一类独特的神经网络架构，擅长对序列数据建模。正如 [29], 所介绍的，Elman型循环神经网络因在隐藏层中设置了上下文单元而脱颖而出，使网络能够学习并保留时间依赖信息[10]。Elman型循环神经网络的核心是其隐藏层，其中包含循环连接。这些连接使网络能够维持内部状态或内存，从而捕捉过去时间步的信息。隐藏层对于建模序列数据中的短期依赖至关重要。Elman型循环神经网络的一个显著特征是包含上下文或先前状态。该层与隐藏层一一连接，在网络架构中形成内部循环。上下文层充当时间依赖信息的记忆机制，使网络能够学习并在更长序列中保留时间模式。然而，需要注意的是，传统的普通循环神经网络在处理长期依赖时可能会遇到困难，这通常是由于训练过程中出现梯度消失或爆炸问题 [10, 2]。这是因为循环反向传播长循环中的非线性激活梯度会压缩或放大权重。从数学上讲，给定输入数据（x_t），其中 $t \in [0, T]$，以及上下文状态 $h_{t-\Delta t}$，模型按如下方式计算时间 t处的隐藏状态:

$$h_t = g(W_x x_t + W_h h_{t-\Delta t} + b_h) \tag{1}$$

其中 W_x 和 W_h 分别是输入与隐藏状态的权重矩阵。 b_h 是偏置项， g 是非线性激活函数，通常为双曲正切（tanh）或整流线性单元（ReLU）。

2.2 长短期记忆网络

Elman型循环神经网络（RNNs）具有一个已被充分记载的挑战，即梯度消失或爆炸问题。该问题出现在时间反向传播（BPTT）过程中，具体是由于网络架构中涉及多次递归，并经过双曲正切和Sigmoid函数等非线性激活函数。因此，某些梯度会变得过小（消失）或过大（爆炸），从而妨碍有效训练。为应对与Elman型循环神经网络相关的这些挑战，已发展出长短期记忆网络（LSTM）和门控循环单元（GRU）等先进架构。这些架构包含专门机制，包括门控单元，以在训练期间有效管理信息流。通过缓解梯度消失和爆炸问题，这些先进模型增强了网络在序列数据中捕获并保留长程依赖关系的能力。尤其是长短期记忆网络，由三个门控单元组成：输入门（i_t）、输出门（o_t），以及后来增加的称为遗忘门（f_t）的单元 [2]。这些门控单元以类似于Elman型循环神经网络的方式，各自独立地更新相应状态（式2a、2b、2c），然后输入门和遗忘门共享信息以更新细胞状态（c_t）。细胞状态作为贯穿整个序列的长期记忆单元，使长短期记忆网络能够在较长时间内捕获并记住信息（式2d）。在长期记忆之后，代表短期记忆的隐藏状态被更新。它同时受当前输入和先前隐藏状态的影响。隐藏状态（h_t）用于进行预测，并基于细胞状态和输出门进行更新（式2e）。

$$i_t = \sigma(W_{ii}x_t + W_{hi}h_{t-\Delta t} + b_i) \tag{2a}$$
$$o_t = \sigma(W_{io}x_t + W_{ho}h_{t-\Delta t} + b_o) \tag{2b}$$
$$f_t = \sigma(W_{if}x_t + W_{hf}h_{t-\Delta t} + b_f) \tag{2c}$$
$$c_t = f_t \odot c_{t-\Delta t} + i_t \odot \tanh(W_{ic}x_t + W_{hc}h_{t-\Delta t} + b_c) \tag{2d}$$
$$h_t = o_t \odot \tanh(c_t) \tag{2e}$$

其中，i_t、o_t和 f_t 分别表示输入门、输出门和遗忘门。σ 是Sigmoid激活函数，$\odot$ 表示逐元素乘法。不同组件的权重记为 $W_{[i/h][gate-first-letter]}$，其中 i 和 h 分别表示输入层和隐藏层。例如，遗忘门隐藏层的权重记为 W_{if}，而细胞状态输入层的权重记为 W_{ic}。

这些方程共同定义了长短期记忆网络架构，阐述了网络中的输入门、遗忘门和输出门，以及细胞状态和隐藏状态如何随时间步更新并相互作用，从而防止系统中的梯度下溢或溢出，并保持对数据更长时间范围的理解能力。

2.3 储备池计算

储备池计算（RC）最初是在回声状态网络（ESN） [13] 和液态状态机（LSM） [14] 框架下提出的，现已重新兴起，成为一种强大而高效的计算框架，专为时间依赖序列的建模与预测而设计 [30, 31]。遵循回声状态网络模型架构的储备池计算由两个随机初始化的组件构成，即输入层和储层；它们是固定的，在训练阶段不会进行训练。输入层 $W_{in} \in R^{D\times N}$ 负责将输入数据随机变换为潜在表示，随后由储层接收，并通过状态计算提取动态时间特征。储层是一个随机方阵 $W_{res} \in R^{N\times N}$，用作储层图的邻接矩阵。储层图遵循Erd˝os–Rényi图结构，其中每条图边都以一定概率存在， [32] 以确保最佳连通性。实验 [33, 34] 表明，除图结构外，谱半径约等于1的储层能够保持回声状态属性，并展现出最佳性能 [33]。在数学上，给定时间 t 的数据，记为 x_t，前一状态记为 $r_{t-\Delta t}$，输入层和储层的权重分别记为 W_x 和 W_r，即可得到模型的状态：

$$r_t = (1-\alpha)r_{t-\Delta t} + \alpha g\left(W_r r_{t-\Delta t} + W_x x_t\right) \tag{3}$$

其中，α 是泄漏率，用于控制新信息纳入系统的速率；g 表示用作输入层激活函数的非线性函数。

这种架构（图 1）使模型能够捕捉复杂的状态模式，从而实现有效且高效的预测。模型的状态以自回归方式计算，随后可通过一个称为“读出层”的输出层映射为输出数据。读出层本质上是一个多元线性回归层，将状态向量（r_t）映射为输出数据（y_t）。为提高泛化能力并防止过拟合，

读出层通常采用蒂霍诺夫正则化，通常称为岭回归。该正则化的目标是最小化以下代价函数：

$$\hat{W}_o = \arg\min_{W_o} \left\{ \|y - RW_o\|_2^2 + \beta \|W_o\|_2^2 \right\} \tag{4}$$

ˆ其中， W_o 和 W_o 分别表示读出层参数及其优化后的版本。在该方程中， y表示目标向量， R 表示状态向量 $[1, x_t, r_t]$， β为正则化参数。方程（4）的向量化解为：

$$\hat{W}_o = (R^T R + \beta I)^{-1} R^T y \tag{5}$$

此处， I 表示单位矩阵。方程（3）和（5）所述过程的结合，构成了储层计算学习和预测动态时间演化的基础。

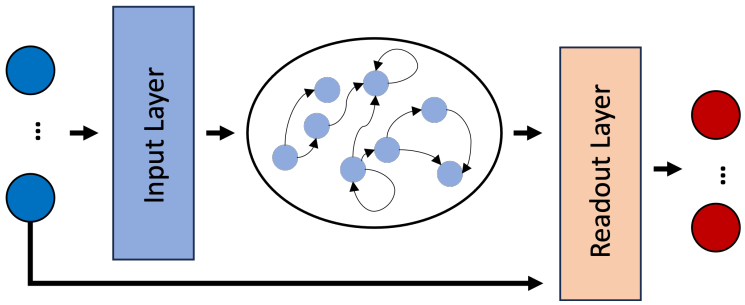

图 1：储备池计算模型架构示意图。左侧蓝色节点和右侧红色节点分别表示输入变量和目标变量。

2.4 序贯储备池计算

所提出的序列储备池计算模型（图 2）旨在通过引入多个顺序连接的储备池层，增强传统储备池计算的能力。在该模型中，每个储备池层处理后续层的信息，并将各储备池的输出拼接后馈入读出层，以进行最终预测（图 2）。储备池层的序贯排列使模型能够捕获复杂的时间依赖关系，并在显著减少储备池参数数量的同时提升整体性能；储备池参数数量过多可能成为扩展至高维数据的瓶颈。用数学形式表示，序列储备池计算由一个参数记为 W_x 的输入层以及多个依次连接的储备池层组成，分别记为 W_r^1、 W_r^2、.....、 W_r^n。每个储备池层 W_r^i都有其自身的内部动力学和权重。储层 W_r^i的动力学可由以下方程描述：

$$r_t^i = \begin{cases} (1-\alpha) r_{t-\Delta t}^i + \alpha g\left(W_r^i r_{t-\Delta t}^i + W_x x_t\right), & \text{for } i = 1 \\ (1-\alpha) r_{t-\Delta t}^i + \alpha g\left(W_r^i r_{t-\Delta t}^i + W_r^{i-1} r_t^{i-1}\right), & \text{for } i \neq 1 \end{cases} \tag{6}$$

其中， i 表示储层层的顺序 $i = 1, \ldots, n$， n 表示储层层的总数。

与储层计算的记法类似，将每个储层层的输出（记为$\mathbf{r}_t^i$）拼接起来，作为读出层的输入。拼接后的输出表示为$\mathbf{r}_t^{\text{concat}}$：

$$R_t = \mathbf{r}_t^{\text{concat}} = [1, x_t, r_t^1, r_t^2, ..., r_t^n] \tag{7a}$$

$$\hat{W}_o = (R^T R + \beta I)^{-1} R^T y \tag{7b}$$

序贯储层计算的读出层参数采用与储层计算类似的方法计算，如公式 7b 所示。

3 数据集

本节将介绍实验研究中使用的数据集，详细说明其各自的属性，并阐述这些实验所采用的具体超参数。

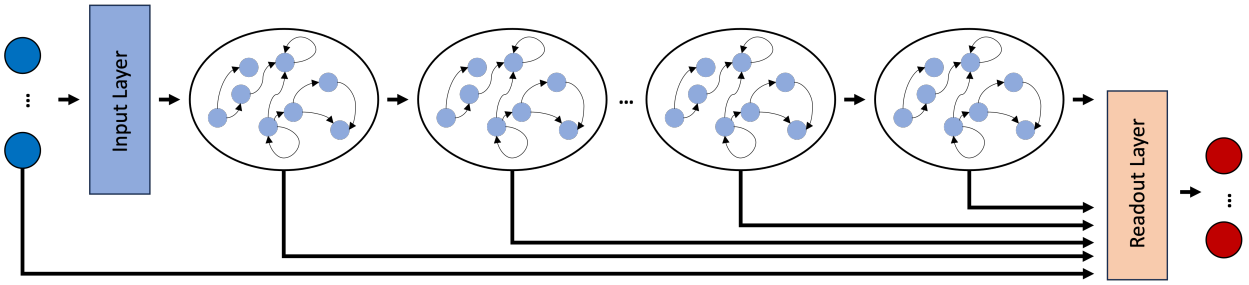

图 2：储备池计算模型架构示意图。左侧的蓝色节点和右侧的红色节点分别表示输入变量和目标变量。每个储层的输出与输入数据一起送入读出层。

3.1 Lorenz63

Lorenz63 系统是一组常微分方程，是混沌动力系统的经典示例。该系统由 Edward N. Lorenz 于 1963 [28] 提出，是一个简化的大气对流模型。该系统由三个相互耦合的非线性微分方程组成，用于描述三个状态变量的演化。Lorenz63 方程如下：

$$\frac{dx}{dt} = \sigma(y - x) \tag{8a}$$

$$\frac{dy}{dt} = x(\rho - z) - y \tag{8b}$$

$$\frac{dz}{dt} = xy - \beta z \tag{8c}$$

其中，x y和 z 是状态变量，σ、ρ和 β 是控制系统行为的参数。Lorenz63 系统以对初始条件的敏感性著称，这会导致其相空间中出现混沌轨迹。由于其结构简单而动力学丰富，Lorenz63 系统已成为研究混沌、吸引子以及不同科学领域（包括大气科学、物理学和数学）中动力系统行为的标准测试平台。在我们的实验中，我们利用 Lorenz63数据集，探索并分析混沌系统固有的复杂动力学，将其作为低维基准数据集。我们分别使用 17.67715816276679、12.931379185960404 和 43.91404334248268 作为 σ、ρ 和 β 的初始值，以初始化动力学。

3.2 涡量

我们研究了二维纳维–斯托克斯方程，该方程以涡度形式描述环面区域内不可压缩流体的黏性行为动力学。控制方程如下：

$$\partial_t w(x,t) + u(x,t) \cdot \nabla w(x,t) = \nu \Delta w(x,t) + f(x), \quad x \in (0,1)^2, \quad t \in (0,T], \tag{9a}$$

$$\nabla \cdot u(x,t) = 0, \quad x \in (0,1)^2, \quad t \in [0,T], \tag{9b}$$

$$w(x,0) = w_0(x), \quad x \in (0,1)^2 \tag{9c}$$

在方程 9a–9c 中，$w(x,t)$、$u(x,t)$、ν以及 $f(x)$ 分别表示涡量随时间的演化、流体速度、黏度和外力。速度由涡量导出，整个系统受不可压缩条件（方程 9b）约束。初始条件 $W_0(x)$ 根据 $w_0\,\mathcal{N}(0, 7^{3/2}(-\Delta + 49I)^{-2.5}$ 生成，并采用周期边界条件 [35]。我们使用固定外力 $f(x) = (sin(2\pi(x_1 + x_2)) + cos(2\pi(x_1 + x_2)))/10$。有关求解器规格和假设的更多信息，请参阅 [35]。在实验中，我们使用固定的雷诺数 500，该数值影响方程 9a 中的黏度（ν）。为确保训练和测试的一致性，我们将模拟网格的分辨率固定为64×64，因为一些现有方法对分辨率变化较为敏感。

3.3 浅水方程

浅水方程（SWEs）是海洋学、水文学及相关领域中的一组基本方程。它们描述薄层中自由表面流的动力学，其中水深远小于其

水平长度尺度。在二维情况下求解浅水方程，这是一种能够有效捕捉流体深度显著小于其水平延展范围时流体运动动力学的典型系统。该模型的控制方程同时包含线性化的动量方程和以非线性方式求解的连续性方程。对动量方程进行线性化，可以对流体运动进行易于处理的表示，从而有利于波传播及其相互作用的数值模拟；而对连续性方程进行非线性求解，则确保系统内质量守恒。该模型以一个显著的高斯凸起作为初始配置，充当水面的初始扰动，进而产生从扰动处向外传播的波。随后，这些波与模型边界发生相互作用，并遵循无流条件。

该模型采用线性动量方程，同时求解非线性形式的连续性方程。脚本具有灵活性，可选择启用或停用各种项。在最完整的配置下，模型由以下方程组描述：

$$\frac{\partial u}{\partial t} - fv = -g\frac{\partial \eta}{\partial x} + \frac{\tau_x}{\rho_0 H} - \kappa u \tag{10a}$$

$$\frac{\partial v}{\partial t} + fu = -g\frac{\partial \eta}{\partial y} + \frac{\tau_y}{\rho_0 H} - \kappa v \tag{10b}$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial (\eta + H)u}{\partial x} + \frac{\partial (\eta + H)v}{\partial y} = \sigma - w \tag{10c}$$

其中， $f = f_0 + \beta y$ 表示随纬度变化的完整科里奥利参数。动量方程采用常规的时间前向、空间中心格式。然而，处理复杂的科里奥利项需要分两步进行：先预测 x 和 y 方向的水平速度分量（u 和 v），再对其进行修正，以纳入科里奥利效应。连续性方程对时间导数采用前向差分，对非线性项采用迎风格式。

模型的稳定性分析基于 Courant-Friedrichs-Lewy（CFL）条件，确保时间步长（dt）受限于 $dt \leq \min(dx, dy)/\sqrt{gH}$。此外，考虑科里奥利效应时，稳定性还要求 $\alpha \ll 1$；其中， dx 和 dy 分别表示 x 和 y 方向的网格间距， g 表示重力加速度， H 表示流体的静水深度。 η、 g、 ρ_0、 κ、 σ、 w、 τ_x 和 τ_y 分别表示表面高度、重力、流体密度、速度系数、附加源汇项、垂直速度，以及 x 和 y 方向的外部强迫项。

4 结果与讨论

4.1 指标

4.1.1 有效预测时间

我们采用有效预测时间（VPT） [19, 20, 17] ，直接追踪低维时间序列数据（即 Lorenz63）中的预测范围。VPT 定义为时间点 t： 模拟预测与真实值之间的偏差超过基于均方根误差（RMSE）预先设定的阈值：

$$\text{RMSE}(t) \geq \varepsilon \tag{11a}$$

$$\text{RMSE}(t) = \sqrt{\frac{1}{D}\sum_{i=1}^{D}\left(\frac{\hat{y}_i(t) - y_i(t)}{\sigma_i}\right)^2} \tag{11b}$$

预测中的第 i 个分量在时间 t 处的值（$\hat{y}_i(t)$）及其对应的真实值（$y_i(t)$）记作相应符号。本研究选取的阈值（ε）设为 0.3。在均方根误差指标中， σ_i 表示真实数据标准差的第 i 个分量，用于归一化。符号 D 表示所考虑问题的维数。

4.1.2 回归映射

回归映射是研究动力系统时使用的一个概念，由洛伦兹 [28]提出，用于分析模拟的稳定性，尤其是通过观察洛伦兹63的动力学。洛伦兹观察到，变量 $z(t)$ 的局部最大值

大约出现在洛伦兹63蝴蝶形轨迹中的一个回路（完整一圈）完成时。分析 $z(t)$ 中连续出现的局部最大值，可以深入了解预测与真实值的整体吻合程度。按照洛伦兹的方法，我们计算了 $z(t)$ 在其邻域内的最大值，然后绘制 $z_i(t)$ 与 $z_{i+1}(t)$ 的关系图，以生成回归映射曲线。

4.1.3 峰值信噪比

峰值信噪比（PSNR）是图像和视频处理领域广泛使用的一种质量评估指标，用于定量衡量原始信号与失真信号之间的保真度。从数学上看，它表示为最大可能强度值的平方与原始信号和失真信号之间均方误差之比的对数。具体而言，在图像比较中，PSNR 表示为 $PSNR = 10 \cdot \log_{10}\left(\frac{\text{最大值}^2}{MSE}\right)$，其中，$\max$ 表示最大可能的像素强度值，MSE 表示均方误差。PSNR 值越高，失真越小，因此重建信号的质量也越高。尽管 PSNR 应用广泛，但需要注意的是，由于它依赖均方误差，且对某些感知细节不敏感，因此可能无法与人们感知到的图像质量完全一致。尽管如此，在各种图像和视频处理应用中，它仍是评估重建信号保真度的重要定量工具。

4.1.4 结构相似性指数

结构相似性指数（SSIM）是一种图像处理领域中广泛使用的综合指标，用于量化原始图像（X）与失真图像（Y）之间的相似程度。与均方根误差或峰值信噪比等传统指标不同，SSIM 不仅考虑亮度差异，还纳入了人类视觉系统可感知的结构模式、纹理和对比度信息。SSIM 指数（$SSIM(X,\ Y)$）通过评估三个分量计算得出：亮度（$l(X,\ Y)$）、对比度（$c(X,\ Y)$）和结构（$s(X,\ Y)$）。这些分量通过以下公式组合：

$$SSIM(X,Y) = \frac{[l(X,Y) \cdot c(X,Y) \cdot s(X,Y)]^{\alpha}}{[l(X,Y)'' \cdot c(X,Y)'' \cdot s(X,Y)'']^{\beta}}. \tag{12}$$

其中，α、β和 γ是用于稳定除法运算的常数；$l(X,\ Y)$、$c(X,\ Y)$和 $s(X,\ Y)$分别表示局部均值、标准差和互协方差函数。SSIM 指数的取值范围为 -1 到 1，其中 1 表示完全相似。当人类感知起关键作用时，该指标尤其有优势，因为它与人们感知到的图像质量更加一致。它已广泛应用于图像和视频压缩算法、图像复原技术的评估，以及其他需要保留结构信息以确保视觉保真度的应用领域。

4.2 低维结果

本节研究聚焦于低维数据，具体而言，以洛伦兹63系统作为实验基础。我们的目标是评估基线模型的性能，并将其与储层计算和序列储备池计算模型进行比较。为深入探究训练数据对预测能力的影响，并了解参数较少的模型（如储层计算和序列储备池计算）可能带来的优势，我们针对三种不同的训练数据长度开展实验，分别使用 2,000、5,000 和 10,000 个样本训练模型，从而考察训练数据集规模变化所产生的影响。此外，我们还通过考察这些模型的可训练参数和固定参数数量、每秒浮点运算次数（浮点运算次数）、内存使用量及训练时间，进一步评估其效率，为了解其计算效率提供有价值的见解。通过这一多维度分析，我们旨在加深对低维系统中训练数据、模型复杂度与预测效率之间关系的理解。

在第一个实验中，我们训练了多对一架构的循环神经网络和长短期记忆网络模型，其结构包括一个全连接输入层、一个循环神经网络/长短期记忆网络层和输出层，以缩小基线模型与所提出的储层计算和序贯储层计算之间的架构差异。选择多对一架构，是为了为循环神经网络/长短期记忆网络预测提供相关上下文，这与储层计算/序贯储层计算在预测中所起的上下文作用类似。我们针对历史窗口长度对模型性能的影响进行了超参数搜索（见附录A）。超参数搜索结果表明，延长历史窗口可以提升模型性能，但会增加计算负担。根据我们在A部分图10中的观察，我们选择了长度为49的历史窗口，并使用 RMSProp [36] 和1e-4的学习率对两个模型进行优化。循环神经网络、长短期记忆网络和储层计算的结果基于大小为256的循环层/储层；对于序贯储层计算，我们使用了8个大小为32的储备池层。图3展示了基于2,000个训练样本的模型性能（有关5,000和10,000个训练样本的结果，请参阅附录2）。

结果显示，在自回归预测方面，循环神经网络和长短期记忆网络模型的表现有限；而储层计算和序贯储层计算在准确预测的持续时间方面表现显著。值得注意的是，尽管序贯储层计算的架构限制比储层计算更多，它仍能在更长时间内保持预测性能。具体而言，序贯储层计算的可训练参数数量相同，而固定参数数量不到其14%（表2）。

在图3中，从上到下各行分别表示 RNN、长短期记忆网络、储层计算和序列储层计算模型；从左到右各列分别表示 $x(t)$ 、 $y(t)$ 和 $z(t)$ 变量。RNN 模型（蓝色曲线）的预测能力较弱，在预测时域开始不久便偏离真实值（绿色曲线），随后开始振荡。与 RNN 相比，长短期记忆网络（第二行的青色曲线）在更长时间内与真实值保持一致，但从 1.73 VPT 起开始偏离真实值。在训练样本量较小的情况下，储层计算和序列储层计算都能比基于反向传播的模型维持更长时间的预测能力。序列储层计算的预测能力可达 8.65 VPT，高于储层计算的 7.19 VPT。随着训练样本数量从 2,000 增至 5,000 和 10,000，除长短期记忆网络外，各模型呈现出相似的性能；对于长短期记忆网络而言，训练样本数量从 2,000 增至 5,000 后，VPT 约翻了一倍。基于反向传播的方法表现欠佳，与学习混沌行为时的梯度爆炸有关 [9]。该论文提出了应对在混沌数据集上训练循环模型所面临挑战的策略，但这一主题超出了我们的研究范围。此外，对于 2,000 个训练样本，储层计算和序列储层计算的性能结果均来自谱半径为 1.1、泄漏率为 0.7 且无稀疏性的配置（模型规格见附录B）。表1展示了四种模型在不同样本量下的 VPT 性能。值得注意的是，在所有训练样本量下，序列储层计算的性能始终与储层计算相当或更优。

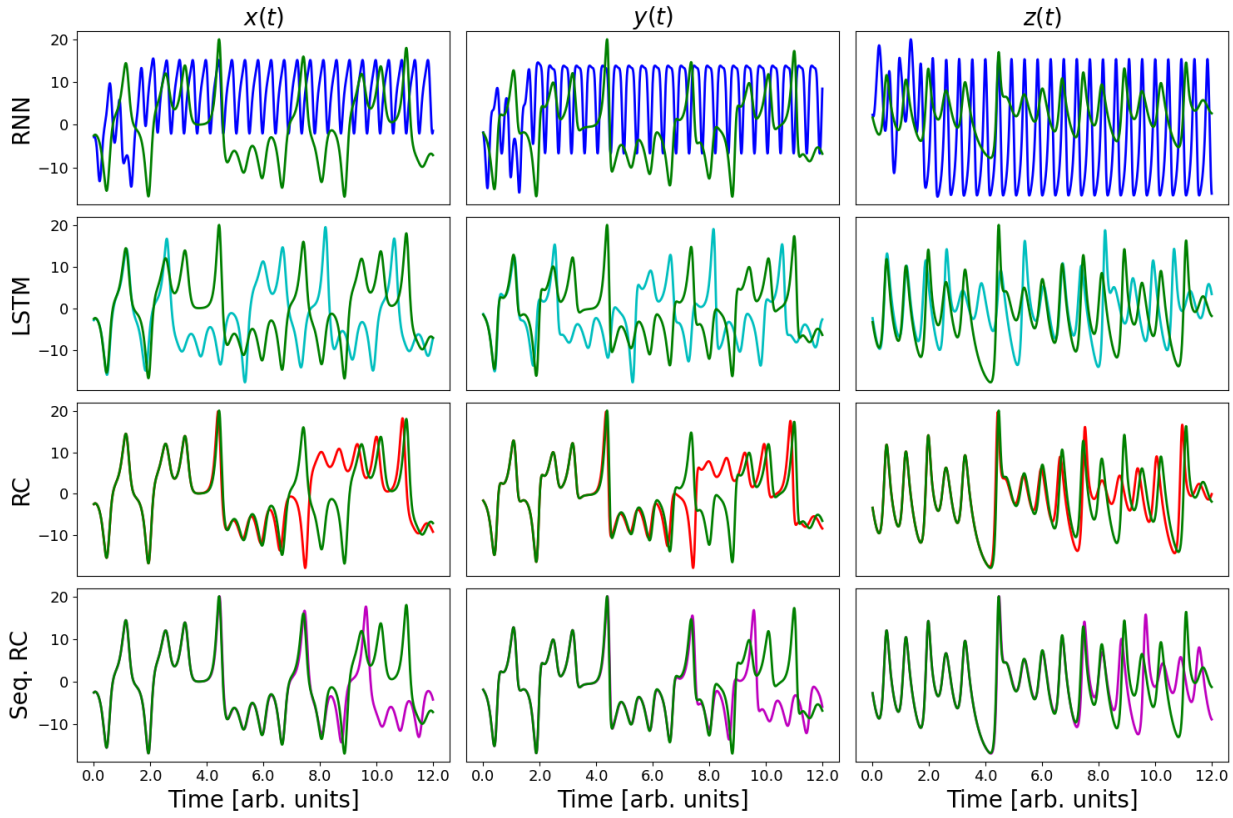

图3：基于 2,000 个训练样本的预测。每个子图中的 Y 轴表示相应变量的取值，X 轴表示预测的提前时间，时间单位任意。

此外，图4表明，除了循环神经网络外，所有模型都能够在不同训练样本量下学习混沌动力学。

我们绘制了第 4.1.2 节所述的回归映射，以观察长程动力学的稳定性。结果显示，对于任何训练样本量，真实值与循环神经网络之间都没有收敛。长短期记忆网络的收敛效果更好，但其回归映射与真实值并未完全对齐。值得注意的是，长短期记忆网络回归映射的收敛性

训练样本	RNN	LSTM	RC	序列储层计算
2,000	0.10	1.73	7.19	8.65
5,000	0.24	3.11	8.58	8.57
10,000	0.17	3.17	7.96	8.67

表1：Lorenz63实验中，不同数据集规模下各模型的时间预测有效期（VPT），单位为任意时间单位。

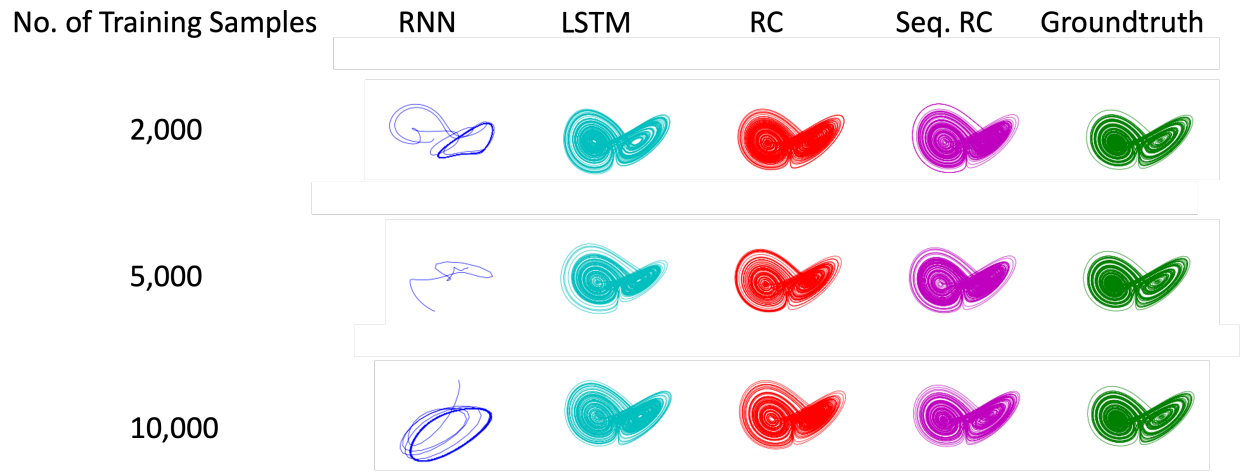

图4：RNN（最左列，蓝色）、LSTM（左起第二列，青色）、RC（左起第三列，红色）、序贯RC（左起第四列，洋红色）与真实值（最右列，绿色）的Lorenz吸引子对比。Lorenz63吸引子展示了系统的动态行为，以及轨迹模式和稳定性的差异。无论训练样本量是多少，RNN都无法跟随轨迹。LSTM无法学习该轨迹，但在使用5,000和10,000个样本时能够捕捉轨迹。RC和序贯RC均能在所有训练样本量下预测轨迹。

从 2,000 个训练样本增加到 5,000 个时显著提升。储层计算和序列储层计算在所有训练样本量下，其回归映射都与真实值高度对齐。

本研究所采用模型的计算和内存需求，对于理解其效率和可扩展性至关重要。尽管已知RC和序贯储层计算中的GPU加速运算有限，为确保评估过程尽可能公平，本文讨论的所有模型均在单个图形处理器（GPU）上训练。报告的内存指标特指GPU内存使用量。在内存需求方面，GPU内存利用率是一个关键因素。内存需求受模型架构、数据集规模和任务复杂度的影响。训练过程中，GPU内存对于存储中间结果、梯度和参数至关重要。了解并监控GPU内存使用量，对于确保所选GPU具备足够容量以应对计算负载至关重要。

为确保公平，本研究报告的所有实验均在一台 NVIDIA GeForce RTX 3090 和 4 颗 Intel(R) Xeon(R) E5-1660 v4 @ 3.20GHz CPU 上进行。表2中报告的内存指标仅涉及 GPU内存使用量，可用于了解各模型的资源需求，并帮助评估其在实际应用中的可行性。表2展示了各模型的模型参数、推理计算、内存负担和训练时间。结果表明，与循环神经网络、长短期记忆网络和储层计算相比，序贯储层计算的运算次数显著更少，分别减少了 724.5 倍、2885 倍和 7.3 倍。类似地，与循环神经网络、长短期记忆网络和储层计算相比，序贯储层计算的内存效率更高，内存使用量分别减少了 111.8 倍、471.7 倍和 1.5 倍。在训练时间方面，由于计算过程具有序贯性质，且存在相应的延迟，序贯储层计算的训练时间未能优于储层计算。不过，它仍具有可比性和优势：随着储层计算的内存消耗呈指数级增长，序贯储层计算的内存需求则明显较低。

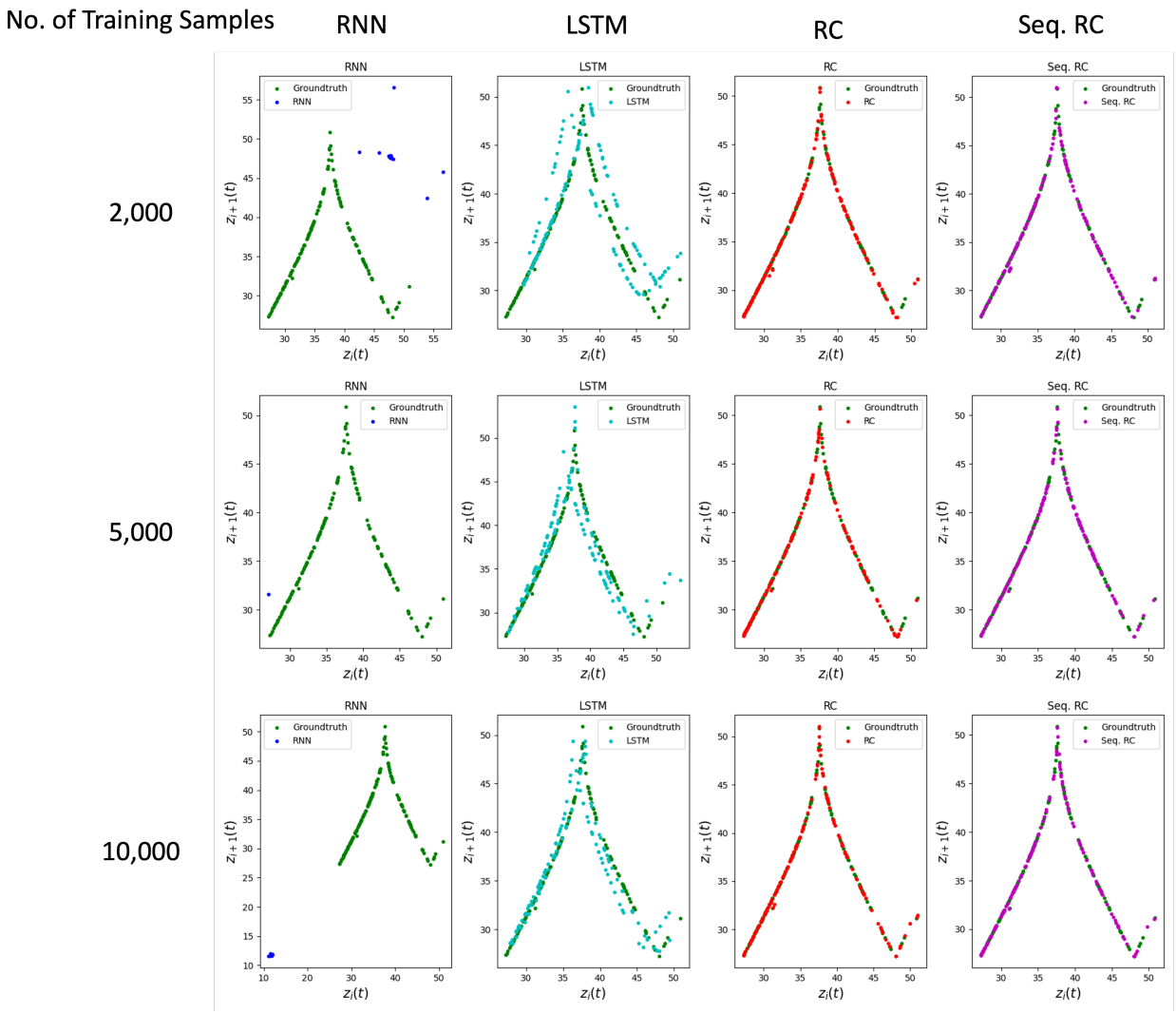

图 5：Lorenz63 系统 z 维的回归映射。图中显示，储层计算和序列储层计算模型与真实值高度吻合，在各种训练样本量下均表现稳健。相比之下，循环神经网络出现了发散，而长短期记忆网络虽然与真实值高度吻合，却未能完全捕捉动力学特征。该比较凸显了储层计算和序列储层计算在保持稳定性并捕捉 Lorenz63 动力学特征方面的有效性。

指标	训练样本数	RNN	LSTM	RC	序列储层计算
可训练参数	-	133,379	528,131	780	780
固定参数	-	0	0	66,304	8,288
浮点运算次数（百万次浮点运算）	-	6.5927	26.2535	0.0673	0.0091
内存消耗（MB）	-	1.0176	4.2930	0.0138	0.0069
训练时间（秒）	2,000	508.0	576.0	0.4	1.4
	5,000	1290.0	1452.0	1.1	3.9
	10,000	2588.0	2932.0	2.2	7.7

表2：该表展示了模型统计数据，包括可训练参数和固定参数的数量、单次前向运行的百万次浮点运算（MFLOPS）及 GPU 内存消耗。此外，还使用 NVIDIA GeForce RTX 3090 和 4 个 Intel(R)Xeon(R)E5-1660 v4 @ 3.20GHz CPU，以秒为单位计算了 Lorenz63实验中不同数据集规模下各模型的训练时间。

4.3 高维结果

在本节研究中，我们将注意力转向高维数据集，具体而言是第 3 节介绍的涡量数据集和 SWE 数据集。与对低维数据的探究相似，我们旨在分析基线模型的性能，并在这一高维背景下将其与储层计算模型和序列储层计算模型进行比较。随着我们深入研究更高维的非混沌数据集，我们希望揭示参数较少的模型（如储层计算和序列储层计算）所带来的潜在优势。我们的实验涵盖了不同长度的训练数据，包括 2,000 个和 5,000 个样本，使我们能够研究训练数据集规模对模型性能的影响。此外，与低维小节类似，我们还通过评估这些模型的可训练参数和固定参数数量、浮点运算次数、内存使用量及训练时间，进一步考察其效率。RNN、LSTM 和储层计算均采用与 Lorenz63 模型相似的架构，每个模型都包含一个大小为 512 的隐藏层。对于序列储层计算，我们采用了 8 个储备池层，每层大小为 64。RNN 和 LSTM 均使用 RMSProp 进行训练，学习率为 10^{-4}。对于储层计算和序列储层计算，我们使用 1.1 的谱半径、0.7 的泄漏率，且不设置稀疏性。我们首先考察基线模型和所提出的序贯储层计算在涡量数据集上的可视化结果（图6）。在图6中，我们展示了训练样本量为 2,000 时，涡量数据集的模型预测可视化分析。每一行对应一种特定模型（循环神经网络、长短期记忆网络、储层计算、序贯储层计算和真实值），各列则表示连续的预测时间步，范围从 $t+1$ 到 $t+60$。子图以较小的 Δt 间隔缓慢推进，从而细致呈现各模型的预测能力。通过可视化对比可以发现，循环神经网络从最初几个预测时间步开始便出现偏离，表明其有效捕捉系统动力学的能力有限。对于长短期记忆网络，偏离现象在 $t+30$左右开始显现，凸显了其难以长期保持预测准确性的挑战。储层计算模型虽然最初表现良好，但到 $t+50$时，预测中的精细结构开始逐渐消失。序列储备池计算模型则表现突出，即使在较晚的时间步也能保持预测效果。子图的演变展现了该模型在捕捉涡量数据集复杂模式方面的稳定性和有效性。

为补充视觉分析，我们使用结构相似性指数、峰值信噪比和均方根误差进行了定量评估，如图 7 所示。在这些图中，Y 轴表示指标值，X 轴表示预测提前时间，起始值为 1。如图 7 所示，循环神经网络在预测时域的早期阶段，在所有指标上都持续偏离真实值，体现出其预测能力有限。长短期记忆网络模型在所有指标上偏离真实值的速度较慢，但其预测性能仍在提前时间的早期阶段有所下降。储层计算和序贯储层计算在初期的表现较为接近，表明它们在预测提前时间的早期阶段具有相当的预测能力。然而，在约 50 个时间步之后，储层计算模型的结构相似性指数和峰值信噪比显著下降，同时均方根误差骤然上升，表明其预测能力有所减弱。相比之下，序贯储层计算直到第 94 个时间步仍保持了可接受的预测能力，展现出其稳健性，并能在更长的提前时间内维持性能。这一定量评估补充了我们的视觉分析结果，从不同角度呈现了各模型在不同提前时间下的预测能力。

同样，对于 SWE 数据集，我们通过图 8 考察了基线模型和所提出的序贯储层计算在视觉上的结果。在这项视觉分析中，我们重点考察了训练样本量为 2,000 时 SWE 数据集的模型预测。图中每一行对应一个不同的模型，包括循环神经网络、长短期记忆网络、储层计算、序贯储层计算以及真实值。各列表示连续的预测时间步，并分为三个区块：顶部区块为 $t+1$ 至 $t+20$，中间区块为 $t+21$ 至 $t+40$，底部区块为 $t+41$ 至 $t+60$。图中的子图展示了 SWE 动力学，并保持了第 3 节所述的边界限制。通过目视检查，可以发现与涡量数据集中的观察结果相似的明显模式。值得注意的是，循环神经网络模型很早便偏离真实值，表明其难以有效捕捉不断演变的动力学，这一点在时间步 5 左右尤为明显。长短期记忆网络模型虽然表现出较为持久的预测能力，但在 $t+20$ 左右开始出现偏离，表明其难以在较长预测时域内维持准确性。储层计算和序贯储层计算模型的表现相近，能够在更长的预测范围内保持动力学特征，但在 $t+48$ 左右开始出现性能下降的迹象。这项全面的视觉分析揭示了各模型在预测 SWE 动力学方面的优势与局限。

与我们对涡量的处理方法一致，我们在视觉检查的基础上，对 SWE 数据集进行了定量评估，采用的指标包括结构相似性指数、峰值信噪比和均方根误差，如图 9 所示。在这些图示中，Y 轴表示指标值，而X 轴表示预测提前时间，从 1 开始。如图 7 所示，在预测时域的早期，循环神经网络在所有指标上都持续偏离真实值，凸显了其有限的预测能力。长短期记忆网络模型在所有指标上的表现均逐渐偏离真实情况，但在提前时间较早时，其预测性能仍有所下降。储层计算和序列储备池计算模型的表现相近，在整个预测提前时间内展现出相当的预测能力。值得注意的是，储层计算和序列储备池计算的结构相似性指数和峰值信噪比均有所下降，并低于 0.6



图6：该图以可视化方式展示了各个模型的预测结果，按行排列（循环神经网络、长短期记忆网络、储备池计算、序贯储层计算和真实值），按列展示预测时间步（$t+1$ 至 $t+60$）。在每个区块中从上到下，子图展示了逐步递增的预测时间间隔。尽管由于 Δt 较小，预测进程看似缓慢，但仍能观察到关键模式。循环神经网络从最初几个时间步起便出现偏离；长短期记忆网络在约 $t+30$ 时开始偏离；储备池计算到 $t+50$ 时开始丢失精细结构；而序贯储层计算即使在较后的时间步仍能保持预测结果。该图展现了各模型在涡量数据预测方面能力的差异。

结构相似性指数阈值，分别出现在第 68 和第 69 个时间步。储层计算和序列储备池计算在均方根误差上的接近趋势也类似；与循环神经网络相比，长短期记忆网络的均方根误差相对更优，但不及储层计算和序列储备池计算。
表3中详述的结果，为我们了解不同神经网络架构（循环神经网络、长短期记忆网络、储备池计算和序贯储层计算）在各种条件下的性能与计算特性提供了宝贵见解。通过考察可训练参数、固定参数、百万次浮点运算（MFLOPS）、GPU内存消耗和训练时间等一系列指标，可以全面了解各模型的能力与效率。可训练参数的数量可作为模型复杂度的指标，体现出长短期记忆网络的预期容量高于循环神经网络。值得注意的是，储备池计算和序贯储层计算的可训练参数数量都相当多，这是因为它们采用最小二乘回归模型进行训练。这种方法显著缩短了训练时间，这一点也反映在训练时间指标上。在固定参数方面，储备池计算的固定参数数量是序贯储层计算的18倍以上。就以浮点运算次数衡量的计算复杂度而言，序贯储层计算的浮点运算次数在各模型中最低，表明其推理时的计算负担较低。各模型的内存使用模式有所不同：循环神经网络和长短期记忆网络所需的内存少于储备池计算和序贯储层计算。对于储备池计算，这是由于其储层分配规模庞大，并且还包含输入层和读出层。相比之下，序贯储层计算将大部分内存分配给读出层和储层序列。不过，由于储层层被拆分，其内存占用更少。就训练时间而言，观察到的平均值一致

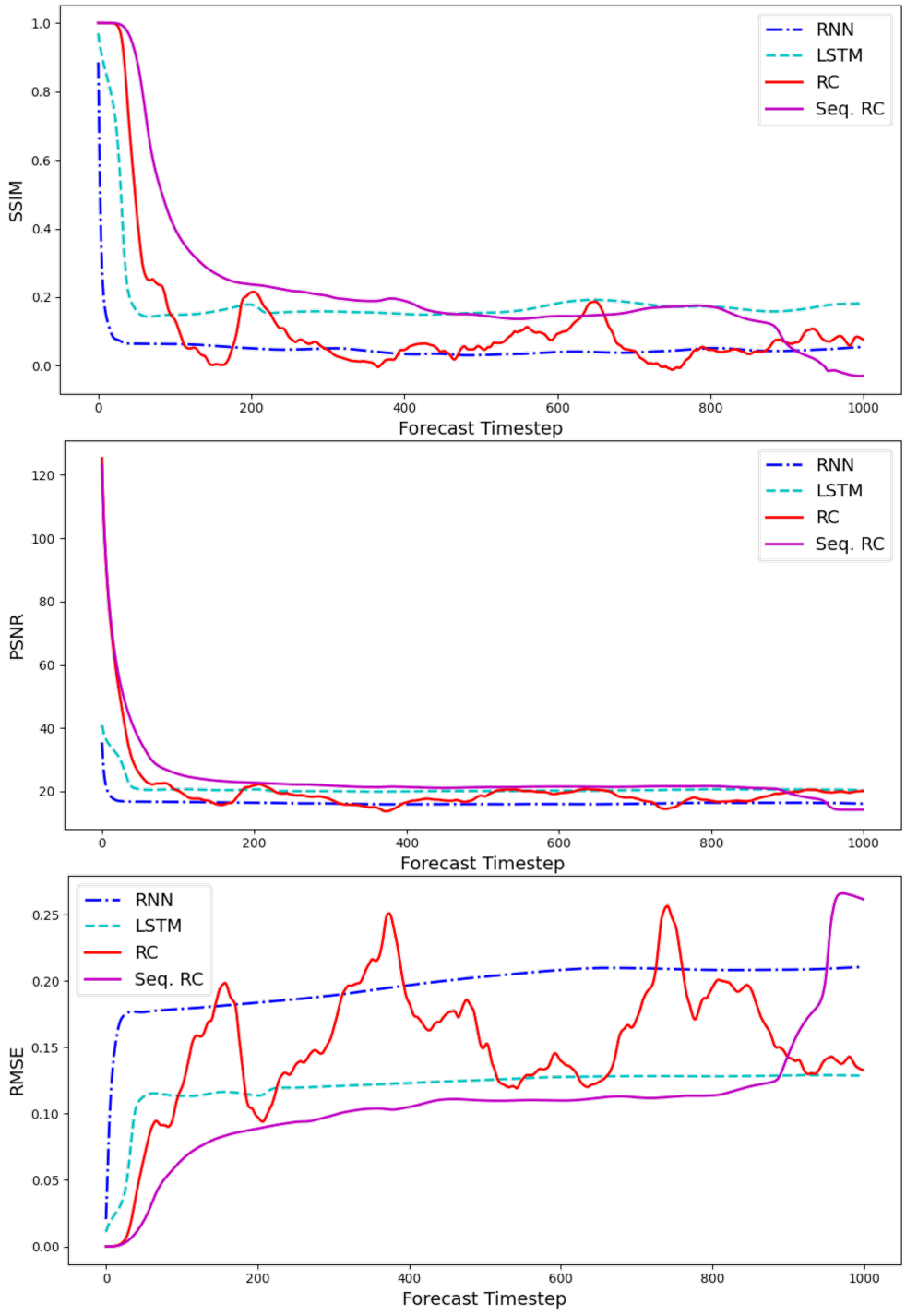

图 7：预测模型的定量评估：绘制结构相似性指数、峰值信噪比和均方根误差指标随预测提前时间的变化。Y 轴表示指标值，X 轴表示从 1 开始的预测提前时间。值得注意的是，循环神经网络较早偏离真实值，表明其预测能力有限。长短期记忆网络的性能退化较慢，但在较早的预测提前时间内准确率有所下降。储备池计算和序贯储层计算的初始表现相近，但约 50 个时间步后，储备池计算的结构相似性指数和峰值信噪比下降，同时均方根误差上升。相比之下，序贯储层计算在第 94 个时间步之前始终保持尚可的预测能力，彰显了其在较长预测提前时间下的稳健性和持续性能。

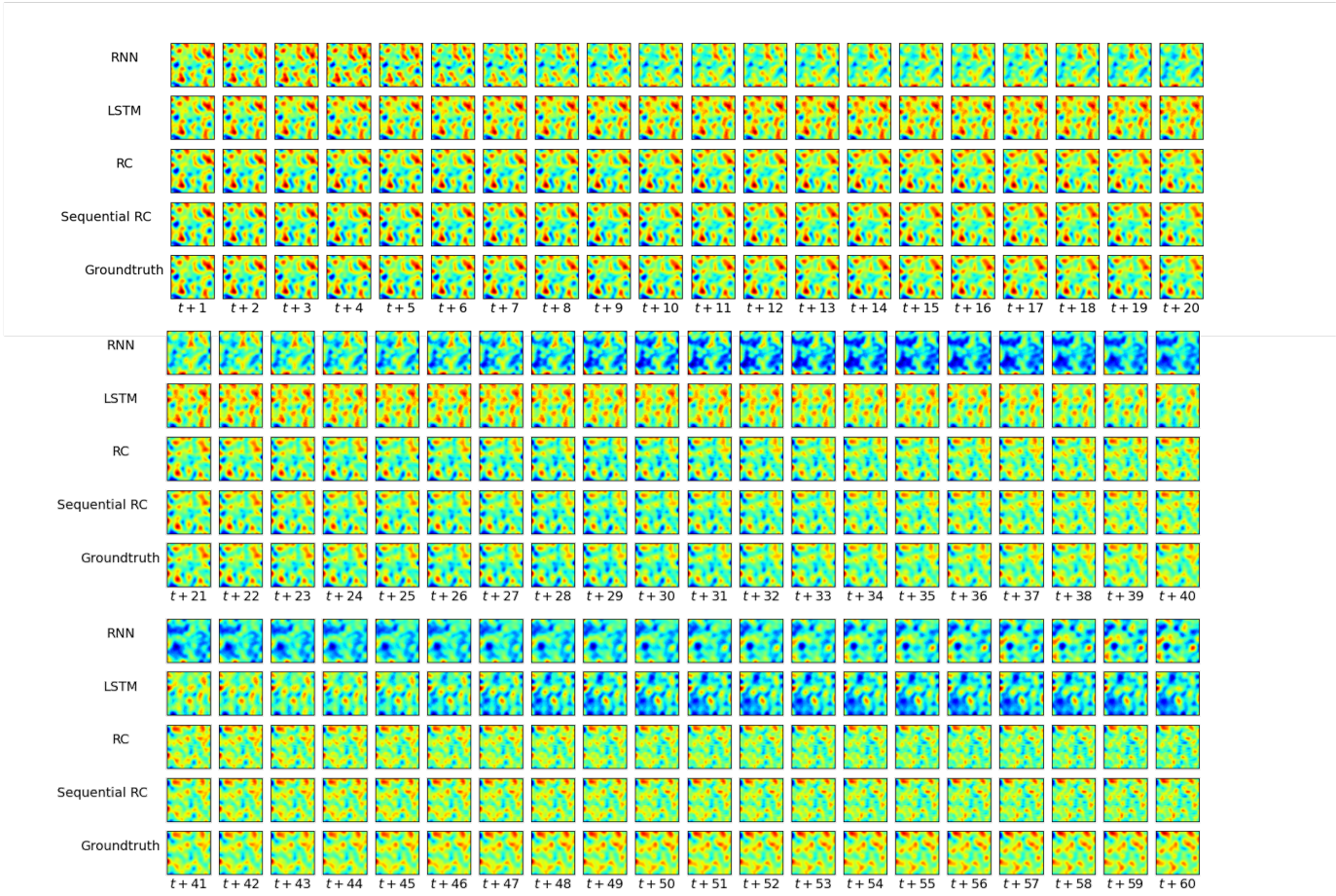

图8：训练样本量为 2,000 时，SWE 数据集上模型预测的可视化分析。每一行对应一种模型，各列表示连续的预测时间步。各子图展示了边界受限条件下的 SWE 动力学，其模式与涡度数据集相似。值得注意的是，RNN 很早便偏离真实值，而长短期记忆网络模型在约 $t + 20$ 之前持续保持良好表现，随后出现偏离。RC 和序贯储层计算模型在更长的预测范围内保留了动力学特征，约在 $t + 48$ 开始性能下降。

涡量数据集与 SWE 数据集之间的比较。对于训练样本量为 2,000 和 5,000 的数据集，RC 的运行速度最快。序贯储层计算紧随 RC 之后，训练时间仅多出几秒。

5 结论

本研究提出了序贯储备池计算（Sequential RC）——一种面向时空预测的标准储备池架构的分层、内存高效扩展。通过依次连接多个小型储层，序贯储备池计算无需承担基于反向传播的循环模型所需的计算开销，即可捕获多尺度时间特征。在低维和高维混沌系统中，序贯储备池计算的表现始终优于 RNN、LSTM 和传统 RC 模型。它将有效预测时域延长了至多 25%，提升了长提前时间下 SSIM 和 PSNR 的稳定性，并将浮点运算次数和内存需求降低了一个数量级（相较于 LSTM）。这些提升并未增加可训练参数的数量；与标准 RC 相比，延迟也仅略有增加。

在低维领域中，我们观察到，与循环神经网络、长短期记忆网络及传统储层计算模型相比，序贯储层计算显著减少了运算次数，并展现出良好的内存效率。我们针对三种不同的训练数据长度（2,000、5,000 和 10,000 个样本）进行了实验，使我们能够

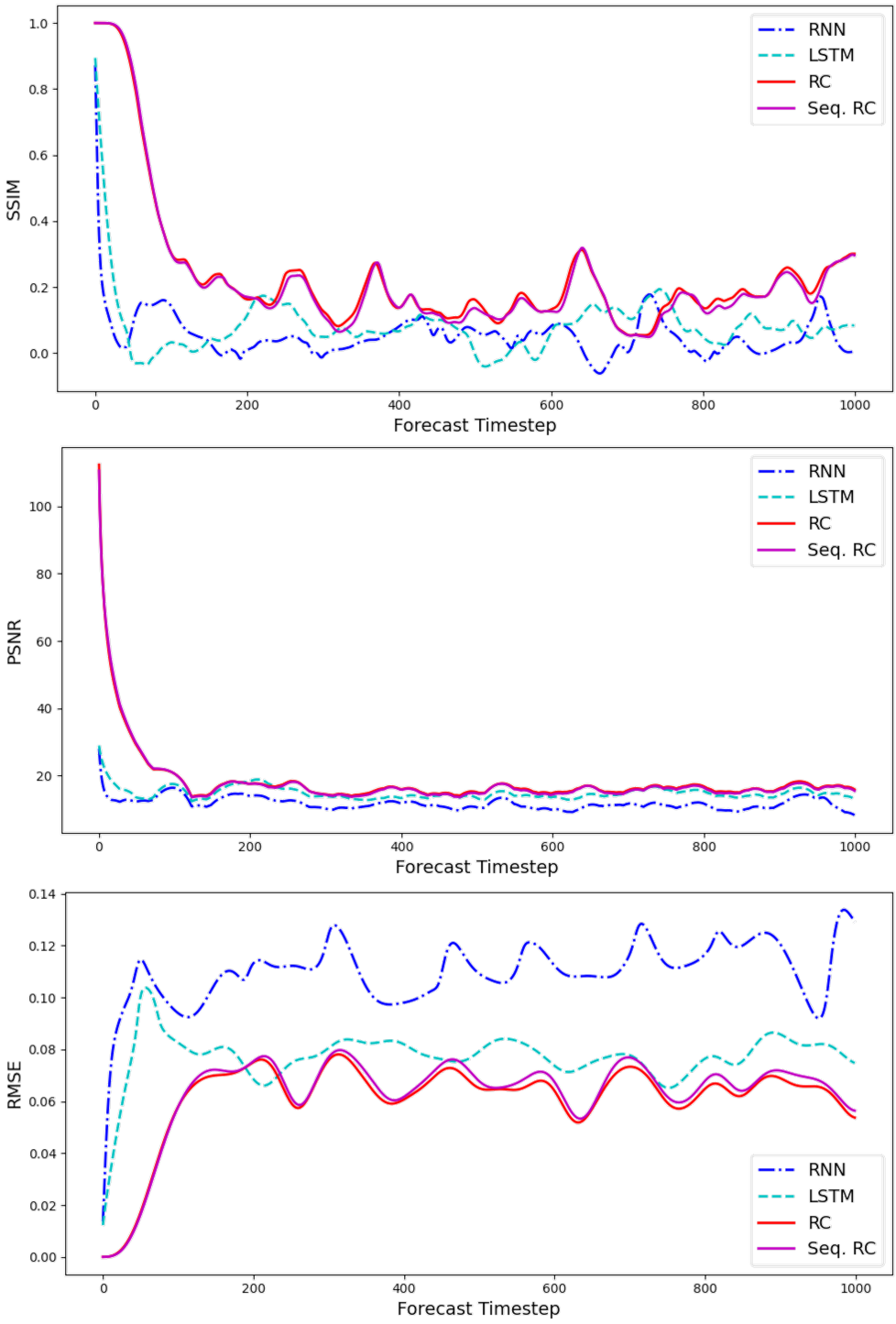

图 9：图 9 展示了使用结构相似性指数、峰值信噪比和均方根误差指标对 SWE 数据集进行的定量评估。Y 轴表示指标值，X 轴表示预测提前时间，从 1 开始。与图 7 对比可见，循环神经网络始终较早偏离真实值，体现出其预测能力有限。长短期记忆网络模型的指标下降较慢，而储层计算和序列储层计算模型在整个预测提前时间内表现出相当的预测能力。储层计算和序列储层计算模型的结构相似性指数与峰值信噪比均有所下降，并分别在时间步 68 和 69 低于 0.6 的结构相似性指数阈值。均方根误差也呈现类似趋势：长短期记忆网络的表现相对优于循环神经网络，但逊于储层计算和序列储层计算模型。

指标	训练样本	RNN	LSTM	RC	序列储层计算
可训练参数（M）	–	10.49	16.79	20.97	18.87
固定参数（M）	–	0.00	0.00	5.24	0.29
浮点运算次数（百万次浮点运算）	–	35.65	67.10	26.21	19.17
内存（MB）	–	80.05	128.10	192.03	146.03
训练时间（秒）	2,000	716.0	852.0	1.3	2.3
	5,000	1,568.0	1,744.0	2.4	4.9

表3：模型统计数据，包括可训练参数和固定参数（以百万计）、百万次浮点运算（MFLOPS）、单次前向运行的GPU内存消耗，以及涡量和浅水方程实验在不同数据集规模下的训练时间。所有实验均在配备4颗 Intel(R) Xeon(R) E5-1660 v4 @ 3.20GHz CPU 的 NVIDIA GeForce RTX 3090 上进行。

使我们能够辨析不同训练数据集规模对预测能力的影响。此外，我们从参数、每秒浮点运算次数（浮点运算次数）和训练时间等方面评估模型效率，从而深入了解这些模型在此情境下的计算效率。

转向高维数据集，具体而言是涡量和浅水方程数据集时，我们的分析沿用了低维探索的平行结构。我们的目标是了解上述模型在更复杂的数据环境中的表现。通过仔细考察训练数据对预测能力和模型效率的影响，我们试图揭示计算性能中的规律与差异。

结果表明，序贯储层计算是一种颇具前景的框架，可用于实时高维动力学预测，并在机器学习与物理建模之间架起桥梁。未来的工作将探索自适应或物理信息储备池连接、将方法扩展至湍流和多模态地球物理数据集、在神经形态或FPGA系统上进行硬件实现，并有望将该架构移植到量子混合实现中 [37, 38, 39]。

A 附录A：基于反向传播的模型的历史窗口

在图10中，我们研究了历史窗口对模型性能和模型内存消耗的影响。我们分析了不同历史窗口对模型性能的影响，重点考察循环神经网络和长短期记忆网络架构。左侧子图展示了不同历史窗口下的验证损失趋势，结果显示，对于这两种模型，输入时间步为50时损失最低。这表明存在一个有利于预测准确性的最佳历史上下文，我们据此报告主要结果。图10右侧子图深入分析了 GPU 内存消耗，单位为兆字节，其中包括模型本身以及每个历史窗口对应的中间张量。该分析揭示了不同历史窗口带来的计算需求。研究结果为选择合适的历史窗口大小提供了有价值的信息，有助于权衡模型性能与计算效率。从图中可以看出，随着历史窗口大小增加，长短期记忆网络的 GPU 需求增长明显快于循环神经网络。

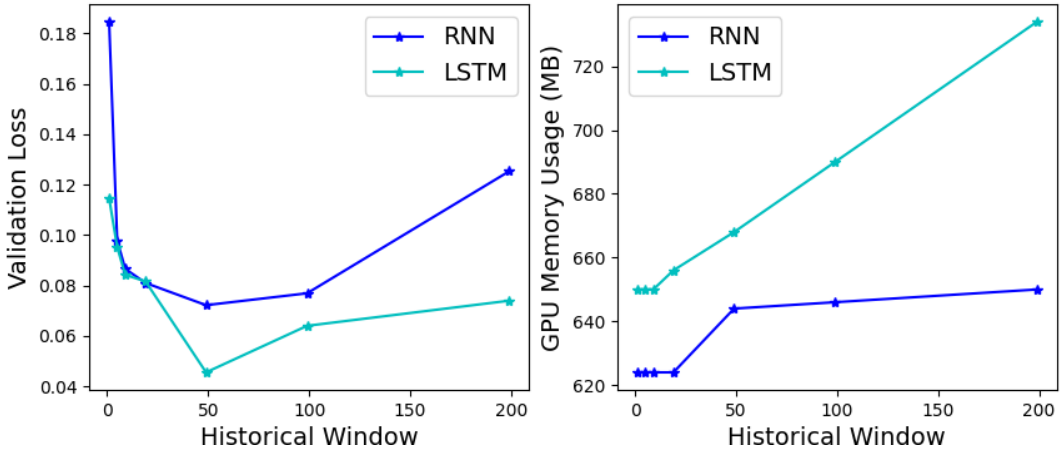

图10：左：循环神经网络（蓝色）和长短期记忆网络（青色）在不同历史窗口大小下报告的验证损失。值得注意的是，循环神经网络和长短期记忆网络模型在使用包含50个输入时间步的历史窗口时，损失均达到最小值。右：两种模型及其中间张量在不同历史窗口大小下的GPU内存消耗（以兆字节为单位）。

B 附录B：建模超参数

本节在表4中列出不同模拟所用模型的全部超参数。

超参数	描述	数据集	模型	值
模型精度	模型参数精度	All	All	双精度浮点数
数据精度	数据集精度	All	All	双精度浮点数
隐藏维度	隐藏维度数量-数	Lorenz63	RNN, LSTM, RC	256
隐藏维度	隐藏维度数量数	涡量，浅水方程	循环神经网络、长短期记忆网络、储层计算	512
序列层	储层层数	All	序列储层计算	8
层大小	每个储层层的大小	Lorenz63	序列储层计算	32
层大小	每个储层层的大小	涡量，浅水方程	序列储层计算	64
学习率	优化器学习率	All	循环神经网络、长短期记忆网络	1e-4
优化器	训练优化器	All	循环神经网络、长短期记忆网络	RMSProp
谱半径	储备池谱半径	All	储层计算，序列储层计算	1.1
泄漏率	储备池泄漏率	All	储层计算，序列储层计算	0.7
稀疏性	储备池稀疏性	All	储层计算、序列储层计算	0.0
历史窗口	输入序列长度	All	All	49

表 4：所有实验中使用的超参数列表，包括其描述、适用的数据集、关联模型及取值。

致谢

USRA 作者感谢渣打银行提供的支持。本研究使用了通过加利福尼亚大学圣迭戈分校的 USRA 高级计算机科学研究所（RIACS）和国家研究平台（NRP）提供的资源。NRP 的开发和部分支持资金来自美国国家科学基金会的资助项目，编号为 1730158、1540112、1541349、1826967、2112167、2100237 和 2120019，此外还获得了社区合作伙伴提供的额外资金。

参考文献

[1] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。Attention is all you need。神经信息处理系统进展，30，2017年。[2] Felix A Gers、Jürgen Schmidhuber 和 Fred Cummins。学习遗忘：使用长短期记忆网络进行持续预测。神经计算，12(10):2451–2471，2000年。[3] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。用于图像识别的深度残差学习。载于 IEEE 计算机视觉与模式识别会议论文集，第770–778页，2016年。[4] Steve Lawrence、C Lee Giles 和 Ah Chung Tsoi。多大的神经网络才能实现最佳泛化？反向传播的收敛特性。技术报告，1998年。[5] Anna Golubeva、Behnam Neyshabur 和 Guy Gur-Ari。在参数数量相同的情况下，网络越宽越好吗？arXiv预印本 arXiv:2010.14495，2020年。[6] Radosvet Desislavov、Fernando Martínez-Plumed 和 José Hernández-Orallo。人工智能推理能耗趋势：超越深度学习的性能与参数定律。Sustainable Computing: Informatics and Systems，38:100857，2023年。[7] Junyoung Chung、Caglar Gulcehre、KyungHyun Cho 和 Yoshua Bengio。门控循环神经网络在序列建模中的实证评估。arXiv预印本 arXiv:1412.3555，2014年。[8] Kyunghyun Cho、Bart Van Merriënboer、Dzmitry Bahdanau 和 Yoshua Bengio。论神经机器翻译的特性：编码器-解码器方法。arXiv预印本 arXiv:1409.1259，2014年。[9] Jonas Mikhaeil、Zahra Monfared 和 Daniel Durstewitz。使用循环神经网络学习混沌动力学的难点。神经信息处理系统进展，35:11297–11312，2022年。[10] Ata Akbari Asanjan、Tiantian Yang、Kuolin Hsu、Soroosh Sorooshian、Junqiang Lin 和 Qidong Peng。基于 PERSIANN系统和长短期记忆循环神经网络的短期降水预报。《地球物理研究杂志：大气》，123(22):12–543，2018年。[11] Hojjat Salehinejad、Sharan Sankar、Joseph Barfett、Errol Colak 和 Shahrokh Valaee。循环神经网络的最新进展。arXiv预印本 arXiv:1801.01078，2017年。[12] Dean V Buonomano 和 Michael M Merzenich。神经网络将时间信息转换为空间编码的过程及其现实特性。Science，267(5200):1028–1030，1995年。[13] Herbert 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 GMD 技术报告，148(34):13，2001年。[14] Wolfgang Maass、Thomas Natschläger 和 Henry Markram。无需稳定状态的实时计算：一种基于扰动的神经计算新框架。神经计算，14(11):2531–2560，2002年。[15] Ashesh Chattopadhyay、Pedram Hassanzadeh 和 Devika Subramanian。使用机器学习方法预测多尺度 Lorenz 96 混沌系统：储备池计算、人工神经网络和长短期记忆网络。《地球物理非线性过程》，27(3):373–389，2020年。[16] Md Raf E Ul Shougat、XiaoFu Li、Tushar Mollik 和 Edmon Perkins。使用 Lorenz 系统进行储备池计算。载于国际设计工程技术会议与计算机和信息工程会议，2021年。[17] Jason A Platt、Stephen G Penny、Timothy A Smith、Tse-Chun Chen 和 Henry DI Abarbanel。针对复杂时空动力学预测的储备池计算系统性探索。Neural Networks，153:530–552，2022年。[18] Mushegh Rafayelyan、Jonathan Dong、Yongqi Tan、Florent Krzakala 和 Sylvain Gigan。用于预测时空混沌系统的大规模光学储备池计算。Physical Review X，10(4):041037，2020年。[19] Pantelis R Vlachas、Jaideep Pathak、Brian R Hunt、Themistoklis P Sapsis、Michelle Girvan、Edward Ott 和 Petros Koumoutsakos。使用循环神经网络预测时空混沌动力学：储备池计算与反向传播算法的比较研究。arXiv预印本 arXiv:1910.05266，42，2019年。[20] Pantelis R Vlachas、Jaideep Pathak、Brian R Hunt、Themistoklis P Sapsis、Michelle Girvan、Edward Ott 和 Petros Koumoutsakos。用于预测复杂时空动力学的循环神经网络中的反向传播算法与储备池计算。Neural Networks，126:191–217，2020年。[21] Kevin Yao、Eric Forgoston 和 Philip Yecko。使用储备池计算学习海洋环流模型。Physics of Fluids，34(11)，2022年。

[22] Claudio Gallicchio 和 Alessio Micheli。深度储备池计算网络的回声状态属性。*CognitiveComputation*，9:337–350，2017年。[23] Claudio Gallicchio、Alessio Micheli 和 Luca Pedrelli。深度储备池计算：关键实验分析。*Neurocomputing*，268:87–99，2017年。[24] Razvan Pascanu、Caglar Gulcehre、Kyunghyun Cho 和 Yoshua Bengio。如何构建深度循环神经网络。*arXiv*预印本 *arXiv:1312.6026*，2013年。[25] Claudio Gallicchio 和 Alessio Micheli。深度储备池计算。储备池计算：理论、物理实现与应用，第77–95页，2021年。[26] Stefano Nichele 和 Andreas Molund。使用元胞自动机进行深度储备池计算。*arXiv*预印本 *arXiv:1703.02806*，2017年。[27] Mirko Goldmann、Felix Köster、Kathy Lüdge 和 Serhiy Yanchuk。深度时延储备池计算：动力学与记忆容量。*Chaos: An InterdisciplinaryJournal of Nonlinear Science*，30(9)，2020年。[28] Edward N Lorenz。确定性非周期流。大气科学杂志，20(2):130–141，1963年。[29] Jeffrey L Elman。发现时间中的结构。认知科学，14(2):179–211，1990年。[30] Benjamin Schrauwen、David Verstraeten 和 Jan Van Campenhout。储备池计算综述：理论、应用与实现。载于第*15*届欧洲人工神经网络研讨会论文集，第*471–482*页，*2007*年，第471–482页，2007年。[31] David Verstraeten、Benjamin Schrauwen、Michiel d'Haene 和 Dirk Stroobandt。储备池计算方法的实验性统一。神经网络，20(3):391–403，2007年。[32] Xinyu Han、Yi Zhao 和 Michael Small。储备池计算的更紧泛化界。*Chaos: AnInterdisciplinary Journal of Nonlinear Science*，32(4)，2022年。[33] Jaideep Pathak、Zhixin Lu、Brian R Hunt、Michelle Girvan 和 Edward Ott。利用机器学习从数据中重现混沌吸引子并计算李雅普诺夫指数。*Chaos: An Interdisciplinary Journal ofNonlinearScience*，27(12)，2017年。[34] Jaideep Pathak、Brian Hunt、Michelle Girvan、Zhixin Lu 和 Edward Ott。基于数据的大型时空混沌系统无模型预测：一种储备池计算方法。*Physical Review Letters*，120(2):024102，2018年。[35] Zongyi Li、Nikola Kovachki、Kamyar Azizzadenesheli、Burigede Liu、Kaushik Bhattacharya、Andrew Stuart 和 Anima Anandkumar。用于参数化偏微分方程的傅里叶神经算子。*arXiv*预印本 *arXiv:2010.08895*，2020年。[36] Tijmen Tieleman 和 Geoffrey Hinton。第6.5讲——RMSProp，Coursera：机器学习神经网络。多伦多大学，技术报告，6，2012年。[37] Casper Gyurik、Filip Wudarski、Evan Philip、Antonio Sannia、Hossein Sadeghi、Oleksandr Kyriienko、Davide Venturelli 和 Antonio A Gentile。从量子特征映射到量子储备池计算：展望与应用。*arXiv*预印本 *arXiv:2510.01797*，2025年。[38] Filip Wudarski、Daniel OConnor、Shaun Geaney、Ata Akbari Asanjan、Max Wilson、Elena Strbac、P Aaron Lott 和 Davide Venturelli。用于混沌系统模拟的混合量子-经典储备池计算。*arXiv*预印本 *arXiv:2311.14105*，2023年。[39] Osama Ahmed、Felix Tennie 和 Luca Magri。用于预测混沌动力学的稳健量子储备池计算机：广义同步与稳定性。*arXiv*预印本 *arXiv:2506.22335*，2025年。