利用机器学习将机理模型转化为预测模型

Amit K. Chakraborty[1],Hao Wang[1,*]和Pouria Ramazi[2]

[1]阿尔伯塔大学数学与统计科学系及数学生态学与流行病学跨学科实验室（*ILMEE*），加拿大艾伯塔省[2]卡尔加里大学生物科学系及数学与统计系，加拿大艾伯塔省[*]通讯作者：*hao8@ualberta.ca*

arXiv:2602.04114v1 [cs.LG] 4 Feb 2026

摘要

复杂动力系统的方程未必能凭借专家知识加以识别，尤其是在底层机制未知的情况下。数据驱动发现方法通过利用由测量变量构建的函数库，从时间序列数据中推断控制方程，从而应对这一挑战。然而，这些方法通常假设系数为时不变系数，这限制了它们刻画不断演化的系统动力学的能力。为克服这一局限，我们允许部分参数随时间变化，直接从数据中学习其时间演化，并推断一个同时包含常数参数与时变参数的方程组。随后，我们通过预测时变参数并将这些预测结果代入学习得到的方程，将该框架转化为预测模型。该模型使用易感者–感染者–康复者、消费者–资源、温室气体浓度以及蓝细菌细胞计数等数据集进行了验证。通过动态适应时序变化，我们所提出的模型在学习时间序列时的平均绝对误差低于3%，在提前最多一个月的预测中低于6%。我们还将其预测性能与卷积神经网络–长短期记忆网络以及梯度提升机（GBM）进行了比较，结果表明我们的模型在大多数数据集上优于这些方法。我们的研究结果表明，将时变参数纳入微分方程的数据驱动发现，能够同时提高建模精度和预测性能。

关键词— 数据驱动动力系统、时变参数、稀疏回归、常微分方程、机理预测、有限时域预测误差界、机器学习

AMS 学科分类——37M05, 37N25, 65L20, 62M45, 92D30, 86A10

1 引言

机理模型提供了系统具有因果性且可解释的支配定律，但许多自然系统过于复杂，难以通过专家知识加以刻画 [1–3]。例如，甲烷和二氧化碳等温室气体的排放会受到气温、湿度、风速及其他气象变量的影响 [4, 5]。在流行病学中，疾病发病率受到环境因素、人口密度、社会行为和政策干预的共同影响 [6–9]。如此高维系统中的变量之间存在众多时变且未知的关系，因此几乎不可能用基于专家知识的方程加以描述 [10]。

一种颇具前景的替代方案是通过数据驱动发现来确定控制方程，通常将其表示为微分方程 [11–14]。与依赖预设假设的机理建模不同，数据驱动方法能够适应系统的内在结构，捕捉系统变量之间的线性和非线性交互作用 [13]。

非线性动力学稀疏识别（SINDy）是一种广泛用于数据驱动发现控制方程的技术 [12]。SINDy 及其变体已被用于揭示流体动力学 [12, 15–19], 、生物系统 [20–26], 、机械系统 [27–29], 以及气候和环境建模 [30–32]中的常微分方程和偏微分方程动力系统。该方法采用稀疏回归，识别描述系统动力学所需的最小候选函数集 [12]。SINDy 最初是为具有常系数的系统开发的，后来已扩展到处理时变参数 [14, 33]。对于动力学不断演化的系统，这类扩展不可或缺，因为固定参数无法准确表示随时间发生的变化 [8]。例如，流行病模型中的传播率和死亡率会随时间变化，可通过将模型与流行病数据进行校准来估计 [8, 9, 34]。

然而，对于包含许多变量的系统，究竟需要多少个时变参数才能准确捕捉数据，目前仍不明确。仅识别必要的时变参数，可以降低计算成本、防止过拟合并提高可解释性。

另一个挑战是将配有时变参数的方程组用于预测。通过模型拟合，可以从过去的数据中找出时变参数（本质上是一种变量）的时间序列。然而，预测需要在给定初始条件的情况下运行模型，这就需要参数未来的时间序列 [8]。该时间序列受外部作用力、变量之间的相互作用以及其他动态因素影响 [8, 9]。为应对这一挑战，人们在各个领域采用机器学习（ML）模型，利用历史时间序列和已知协变量来预测时变参数 [34]。在流行病学中，人们利用流行病模型，根据政策或天气预测疾病传播和死亡率 [8, 9, 35–37]。在气候科学中，人们通过预测时变参数来研究温室气体驱动因素 [38]。因此，一旦

系统采用时变参数建模后，便可通过机器学习模型，利用参数的过去值以及未来值已知的协变量进行预测 [34, 39]。

本研究旨在开发一种数据驱动框架，用于发现具有固定参数和时变参数、能够刻画演化中动力系统的常微分方程。我们通过分两个阶段应用 SINDy 算法来实现这一框架。第一阶段，将 SINDy 应用于完整时间序列，以识别在假设参数恒定的情况下，哪些变量会主动影响动力学及其相互作用。第二阶段，在较短的时间区间内重复分析，使根据验证性能选定的一部分参数能够随时间变化，同时将其余参数保持固定。此过程生成的常微分方程模型既保留了底层机理结构，又能适应关键过程随时间的变化。为实现预测，我们利用由影响时变参数的外部协变量驱动的机器学习模型，对识别出的时变参数进行预测。随后，将这些预测参数代入学习得到的常微分方程，从而运行模型并生成能够动态适应不断变化条件的预测。

我们将这一方法应用于两个著名的数学模型——易感者–感染者–康复者（SIR）模型 [40] 和消费者–资源（CR）模型 [41]——以及两个真实世界的数据集：艾伯塔省油砂尾矿池中的温室气体浓度动态，以及艾伯塔省各湖泊中的蓝藻细胞计数。我们利用气温、相对湿度、降水和风速等天气相关驱动因素，通过随机森林（RF）模型预测时变参数。为评估所提方法的预测性能，我们将其与两种广泛使用的数据驱动基线进行了比较：混合卷积神经网络-长短期记忆（CNN-LSTM）模型和梯度提升机（GBM）。

2 方法

2.1 问题表述

考虑自然数 $n \in \mathbb{N}$ 所对应的状态向量 $\mathbf{x}(t) \in \mathbb{R}^n$ 的演化，其由以下动力系统描述

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \tag{2.1}$$

其中，$\mathbf{u}(t) \in \mathbb{R}^q$, $q \in \mathbb{N}$ 表示输入变量，$\mathbf{f}: \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}^n$ 是一个未知函数。状态变量的测量值在 m 个时间点 $\mathbf{t} = [t_1, \ldots, t_m]$ 可用，

从而得到时间序列数据

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}(t_1)^\top \\ \mathbf{x}(t_2)^\top \\ \vdots \\ \mathbf{x}(t_m)^\top \end{bmatrix}, \tag{2.2}$$

其中， m 为测量值的数量，所得矩阵的维度为 $m \times n$。相应的时间导数数据为

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{x}}(t_1)^\top \\ \dot{\mathbf{x}}(t_2)^\top \\ \vdots \\ \dot{\mathbf{x}}(t_m)^\top \end{bmatrix}. \tag{2.3}$$

类似地，输入的时间序列数据为

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}(t_1)^\top \\ \mathbf{u}(t_2)^\top \\ \vdots \\ \mathbf{u}(t_m)^\top \end{bmatrix}.$$

用于估计**f(·)**的候选函数库由**X** 和**U** 构造而成：

$$\boldsymbol{\Theta}(\mathbf{X}, \mathbf{U}) = \left[\mathbf{1},\ \mathbf{X},\ \mathbf{U},\ (\mathbf{X} \odot \mathbf{U}),\ \mathbf{X}^2,\ \mathbf{U}^2,\ \ldots,\ (\mathbf{X}^{d-1} \odot \mathbf{U}),\ (\mathbf{X}^{d-2} \odot \mathbf{U}^2),\ \ldots,\ (\mathbf{X} \odot \mathbf{U}^{d-1}),\ \mathbf{X}^d,\ \mathbf{U}^d \right], \tag{2.4}$$

其中 $d \in \mathbb{N}$，候选函数库包含**x(t)**和**u(t)**中总次数不超过 d的所有单项式。除多项式项外，候选函数库还可以包含非线性基函数，例如 sin(·)、cos(·)、tan(·)、它们的幂以及其他用户定义的非线性函数。

根据方程 (2.1)，导数**X** 与候选函数库 **Θ(X,U)**之间的关系可表示为

$$\dot{\mathbf{X}} = \boldsymbol{\Theta}(\mathbf{X}, \mathbf{U})\, \boldsymbol{\Xi}, \tag{2.5}$$

其中

$$\boldsymbol{\Xi} = \begin{bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \ldots & \boldsymbol{\xi}_n \end{bmatrix}$$

是未知系数向量的矩阵 ξ_k。对于具有 n 个变量以及含 l 项的候选库 **Θ(·)**的系统，系数矩阵 **Ξ** 的维数为 $l \times n$。$\boldsymbol{\xi}_k$ 的非零元素表示对 k^{th} 状态的动力学有贡献的活跃候选函数。一旦

$\boldsymbol{\xi}_k$ 被确定， k^{th} 状态的非线性方程由下式给出

$$\dot{x}_k = \boldsymbol{\Theta}(\mathbf{x}, \mathbf{u})\, \boldsymbol{\xi}_k, \tag{2.6}$$

其中， x_k 是**x**的 k^{th} 分量，而 **Θ(x，u)**表示在**x** 和**u**处求值的候选函数向量。

示例 2.1.1. 考虑 $n = 2$ 个状态变量，即$\mathbf{x} = [x_1, x_2]^\top$，以及 $q = 2$ 个输入变量，即$\mathbf{u} = [u_1, u_2]^\top$，并设函数库的最高多项式次数为 $d = 2$。如果在时间点 $m = 2$测量这些变量，即$\mathbf{t} = [t_1, t_2]$，那么

$$\mathbf{X}^2 = \big[\, x_1^2(\mathbf{t}),\ x_1x_2(\mathbf{t}),\ x_2^2(\mathbf{t})\big], \qquad \mathbf{U}^2 = \big[\, u_1^2(\mathbf{t}),\ u_1u_2(\mathbf{t}),\ u_2^2(\mathbf{t})\big],$$

且候选库包含 $l = 15$ 项

$$\begin{aligned}\boldsymbol{\Theta}(\mathbf{X}, \mathbf{U}) = [&\mathbf{1},\ x_1(\mathbf{t}),\ x_2(\mathbf{t}),\ u_1(\mathbf{t}),\ u_2(\mathbf{t}),\ x_1x_2(\mathbf{t}),\ u_1u_2(\mathbf{t}),\ x_1u_1(\mathbf{t}),\ x_1u_2(\mathbf{t}),\ x_2u_1(\mathbf{t}),\ x_2u_2(\mathbf{t}),\\ &x_1^2(\mathbf{t}),\ x_2^2(\mathbf{t}),\ u_1^2(\mathbf{t}),\ u_2^2(\mathbf{t})].\end{aligned}$$

假设真实动力学为 $\dot{x}_1 = 0.6x_1 + 0.2x_2u_1$ 和 $\dot{x}_2 = -0.4x_2 + 0.1u_2^2$。方程(2.5) 旨在求得 $\boldsymbol{\Xi} \in \mathbb{R}^{15\times 2}$，其列为系数向量 $\boldsymbol{\xi}_k \in \mathbb{R}^{15\times 1}$。在此例中，

$$\boldsymbol{\xi}_1 = [\,0,\ 0.6,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0.2,\ 0,\ \ldots,\ 0\,]^\top, \quad \boldsymbol{\xi}_2 = [\,0,\ 0,\ -0.4,\ 0,\ \ldots,\ 0,\ 0.1\,]^\top,$$

其中，非零项标识函数库中的活跃项。

方程（2.5）可作为稀疏回归问题求解，以确定候选非线性项 **Θ**（**X**，**U**）的活跃系数 **Ξ** 。Brunton 等人（2016） [12]提出了一种非线性动力学稀疏识别（SINDy）算法。SINDy 的早期变体使用 L^1 正则化项进行回归，并在整个时间域内应用该项，以识别具有常系数的方程组 [12, 22]。因此，Li 等人（2019） [14]提出了 SINDy 的时变变体，用于识别动力系统的时变系数；Rudy 等人（2017年、2019年） [13, 33] 则采用岭回归，在损失函数中加入 L^2 正则化项。

2.2 发现时变动力学

按照 Li 等人（2019） [14]提出的框架，将结合了岭回归与 Rudy 等人（2017年、2019年） [13, 33], 所述稀疏性阈值化技术的序贯阈值岭回归（STRR），应用于发现方程 2.5 中的时变系数 **Ξ** 。

首先，根据先验知识选取了式 (2.1) 中的输入变量 **u**(*t*)，并收集了所有变量（**x**(*t*) 和 **u**(*t*)) 的时间序列数据（图 1a）。每个变量的时间

序列都在若干个时间点上进行滚动处理，以降低噪声。随后，对状态变量的导数 $\mathbf{X}$ 进行数值估算，并构建了候选库 $\boldsymbol{\Theta}(\mathbf{X},\mathbf{U})$（式 2.4）。

利用 STRR，在整个训练时间域内提取了式（2.5）中 $\boldsymbol{\xi}_k$ k^{th} 状态的活跃常系数（图 1b）。STRR 方法通过迭代式两步优化，在施加正则化的同时强制实现稀疏性，从而最小化误差。首先，通过最小化目标函数应用岭回归

$$\arg\min_{\boldsymbol{\xi_k}} \|\dot{\mathbf{X}}_k - \boldsymbol{\Theta}(\mathbf{X},\mathbf{U})\boldsymbol{\xi}_k\|_2^2 + \lambda\|\boldsymbol{\xi}_k\|_2^2, \tag{2.7}$$

其中，$\lambda\|\boldsymbol{\xi}_k\|_2^2$ 表示用于惩罚非零系数的正则化分量，由超参数 λ 控制，有助于避免过拟合。随后应用强制稀疏性的阈值处理步骤，将小于预定义阈值 τ 的系数设为零:

$$\xi_{i,k} = \begin{cases} 0 & \text{if } |\xi_{i,k}| < \tau \\ \xi_{i,k} & \text{otherwise} \end{cases} \tag{2.8}$$

其中$\xi_{i,\ k}$ 是 $\boldsymbol{\xi}_k$ 的第 i个条目。此步骤会移除较小、较不重要的特征。随后，使用相同的损失函数，对剩余的活跃特征（其系数未被置零） $\boldsymbol{\Theta}_{\text{active}}$（$\mathbf{X}$，$\mathbf{U}$）重新拟合岭回归

$$\arg\min_{\boldsymbol{\xi}_k} \|\dot{\mathbf{X}}_k - \boldsymbol{\Theta}_{\text{active}}(\mathbf{X},\mathbf{U})\boldsymbol{\xi}_k\|_2^2 + \lambda\|\boldsymbol{\xi}_k\|_2^2. \tag{2.9}$$

这种迭代细化确保模型持续适应最相关的特征。该过程不断重复，直到系数收敛，即连续两次迭代之间的系数变化小于指定容差 ϵ:

$$\|\boldsymbol{\xi}_k^{(t)} - \boldsymbol{\xi}_k^{(t-1)}\| < \epsilon, \tag{2.10}$$

其中，$\boldsymbol{\xi}_k^{(t)}$ 和 $\boldsymbol{\xi}_k^{(t-1)}$ 分别表示第 t 次和第 $t-1$次迭代时的系数。

计算了 k^{th} 状态下活跃候选项（$\boldsymbol{\Theta}_{\text{active}}(\mathbf{X},\ \mathbf{U})$)与目标变量 $\dot{\mathbf{X}}_k(\mathbf{t})$之间的相关性（图 1c）。随后，从候选项集合（$\boldsymbol{\Theta}_{\text{active}}(\mathbf{X},\ \mathbf{U})$）中找出与 $\mathbf{X}_k(\mathbf{t})$相关性最高的前 N 个活跃候选项$\boldsymbol{\Theta}_{\text{top N}}$。将 $\boldsymbol{\Theta}_{\text{top N}}$ 的系数选为时变系数（称为时变参数）（图 1c），其余系数则保持不变。此外，默认将偏置项设为时变，以反映不同时间区间之间的基线偏移。

为捕捉系统 (2.1) 的时变动力学，定义了窗口大小 w，并在整个训练期间保持不变。STRR ((2.7)-(2.10)) 被应用于每个时间窗口 $[t, t+w]$（图 1d），每个参数均使用其在前一个

时间区间中的取值进行初始化。这种序贯拟合策略使时变参数能够动态调整，同时将影响较小的参数保持不变，从而捕捉模型结构中的关键变化，并维持各时间区间之间的连续性。发现的 k^{th} 状态表达式可写为

$$\dot{x}_k(t) = \xi_{k,0}(t) + \sum_{z_i(t)\in\Theta_{\text{top N}}} \xi_{k,i}(t) z_i(t) + \sum_{z_j(t)\notin\Theta_{\text{top N}}} \bar{\xi}_{k,j} z_j(t), \tag{2.11}$$

其中，$\xi(t)$ 为时变参数，$\bar{\xi}$ 为常系数。此处，$z_i(t) \in \boldsymbol{\Theta}_{\text{top N}}$表示排名前 N的活跃候选项，而 $z_j(t) \in \boldsymbol{\Theta}_{\text{active}}(\mathbf{X}, \mathbf{U}) \setminus \boldsymbol{\Theta}_{\text{top N}}$ 表示其余采用固定系数的活跃项。

2.3 基于学习得到的动力学进行预测

为预测状态变量未来的动力学，本文利用了方程（2.11）中发现的表征，该表征同时包含固定系数和时变系数。其核心思路是在保持常系数（$\bar{\xi}$）不变的同时，预测时变参数 $\xi(t)$。因此，只需更新模型中随时间演化的分量，即可实现预测。

本文采用机器学习方法对时变参数的演化进行建模。在模型开发过程中，从 STRR 得到的时变参数 $\xi(t)$在每个长度为 w的窗口内均为分段常数。然而，在预测过程中，预测参数不受窗口内保持恒定的约束；机器学习模型会根据输入特征生成连续预测。使用相关预测变量训练模型，并将每个时变参数指定为目标输出（图 1e）。将预测的时变参数与固定系数及活跃候选项结合后，动力学模型（2.11）的右端项随之更新，并通过以下方式获得 k^{th} 状态变量的预测：

$$\dot{\hat{x}}_k(t) = \hat{\xi}_{k,0}(t) + \sum_{z_i(t)\in\Theta_{\text{top N}}} \hat{\xi}_{k,i}(t)\, z_i(t) + \sum_{z_j(t)\notin\Theta_{\text{top N}}} \bar{\xi}_{k,j}\, z_j(t), \tag{2.12}$$

其中，$\hat{\xi}(t)$ 是预测的时变参数，$\bar{\xi}$ 是常系数。这里，$z_i(t) \in \boldsymbol{\Theta}_{\text{top N}}$ 表示排名前 N 的活跃候选项，而 $z_j(t) \in \boldsymbol{\Theta}_{\text{active}}(\mathbf{X}, \mathbf{U}) \setminus \boldsymbol{\Theta}_{\text{top N}}$表示其余系数固定的活跃项。

3 案例研究

我们将所提出的方法（第 2 节）应用于四个数据集：两个由已知的常微分方程系统模拟生成，另外两个为实证数据集。模拟数据集分别由 1）易感者-

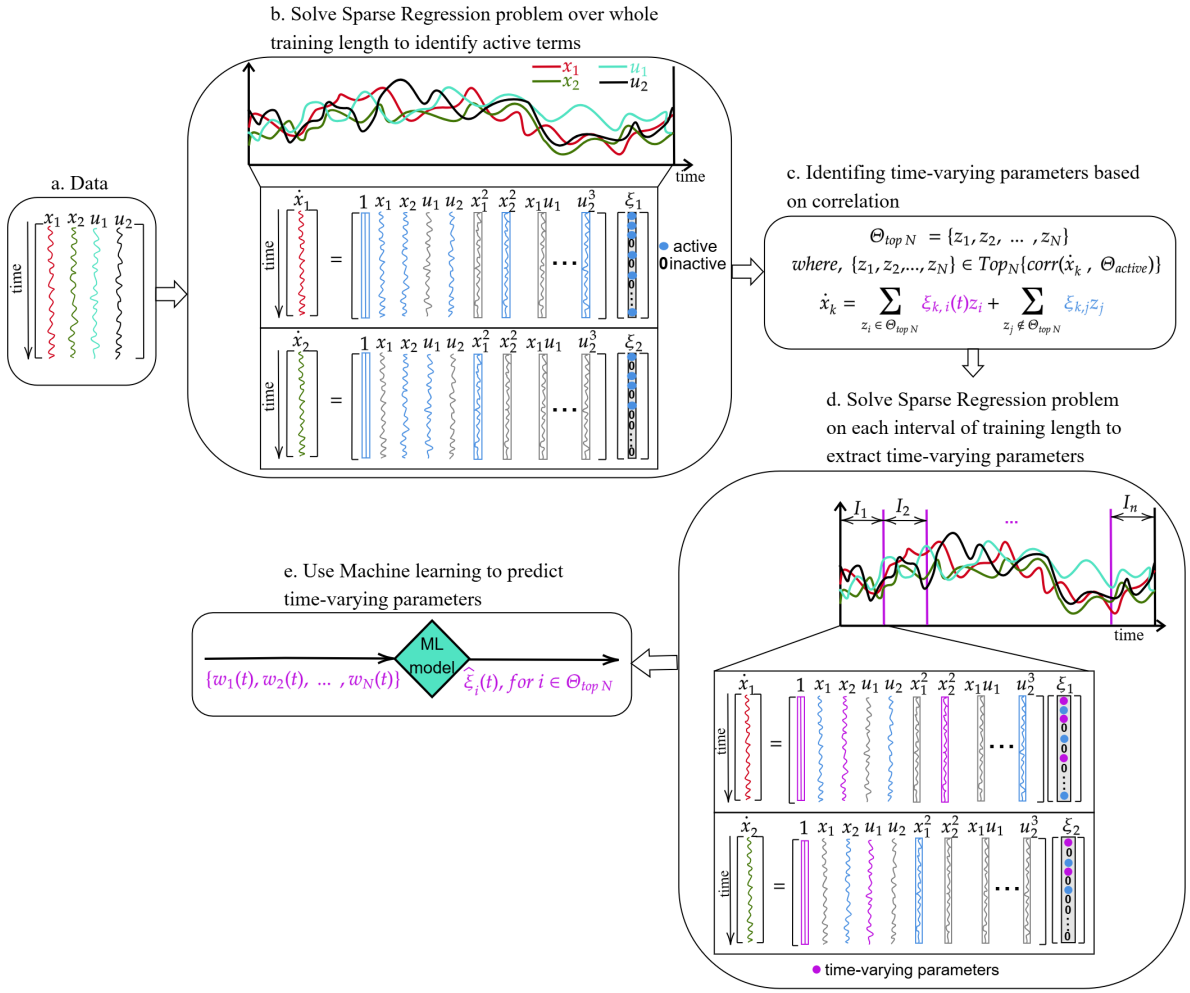


图 1：数据驱动发现控制方程并基于学习得到的系统进行预测的示意图。a）从自然系统或仿真系统中收集变量的时间序列数据。b）针对整个时间序列求解稀疏回归问题，以识别活跃项和常系数。灰色变量表示非活跃项，蓝色变量表示具有常系数的活跃项。方框中的变量表示生成项。c）根据活跃候选项与状态变量导数之间的相关性，选出排名前 N 的候选项。这些项的系数被视为时变参数。默认情况下，偏置项也被视为时变参数。d）将时间序列数据划分为多个时间区间，并在每个区间内进行稀疏回归，以确定排名前列的候选项所对应的时变参数。紫色表示排名靠前的候选项及其时变参数，蓝色表示具有常系数的项。e）采用机器学习模型预测时变参数。输入为相关预测变量，输出为时变参数。

感染者–康复者（SIR）模型和 2）消费者–资源（CR）模型生成。实证数据集包括：1）在艾伯塔省油砂尾矿池中测得的二氧化碳（CO_2）和甲烷（CH_4）浓度，以及 2）从艾伯塔省各地湖泊采集的蓝藻细胞计数。

3.1 SIR模型

SIR模型 [40]是大多数疾病传播模型的基础。在该模型中，种群被划分为三个舱室：易感者（S）、感染者（I）和康复-

者（R）。易感舱室代表尚未感染、但有可能通过接触感染者而罹患该疾病的个体。感染舱室由正在携带该疾病并能将其传播给他人的个体组成。最后，康复舱室包括已从该疾病中康复、并假定已获得免疫力、不再参与疾病传播的个体。这些舱室动态地相互作用，其间的转移由一组常微分方程控制，用以刻画各群体随时间的变化率。模型方程如下

$$\begin{cases} \frac{dS(t)}{dt} &= \Lambda - \beta(t)S(t)I(t) - \mu S(t), \\ \frac{dI(t)}{dt} &= \beta(t)S(t)I(t) - \alpha I(t) - \mu I(t), \\ \frac{dR(t)}{dt} &= \alpha I(t) - \mu R(t), \end{cases} \tag{3.1}$$

其中， Λ 是易感者群体的补充率， $\beta(t)$是随时间变化的疾病传播率， μ是自然死亡率， α 是疾病康复率。式（3.1）的前两个方程与康复者（R）无关，化简后的方程如下:

$$\begin{cases} \frac{dS(t)}{dt} &= \Lambda - \beta(t)S(t)I(t) - \mu S(t), \\ \frac{dI(t)}{dt} &= \beta(t)S(t)I(t) - \alpha I(t) - \mu I(t). \end{cases} \tag{3.2}$$

在 SIR模型（3.2）的方程中加入白噪声后，具有加性白噪声的 SIR模型形式为 [39]

$$\begin{cases} dS(t) &= \Lambda dt - \beta(t)S(t)I(t)dt - \mu S(t)dt + \sigma_1 dW_1(t), \\ dI(t) &= \beta(t)S(t)I(t)dt - \alpha I(t)dt - \mu I(t)dt + \sigma_2 dW_2(t), \end{cases} \tag{3.3}$$

其中， σ_i、 $i = 1$、2、是白噪声的强度， $W_i(t)$ $i = 1$、2、是相互独立的维纳过程。

我们使用方程（3.3）模拟了易感者和感染者群体在1500天内的时间序列。模型参数设为 $\Lambda = 100$、 $\alpha = 0.1$、 $\mu = 0.1$，初始条件为 $S_0 = 500$ 和 $I_0 = 7$。传播率 $\beta(t)$定义为正弦函数的组合，以捕捉不规则的季节性峰值，具体如下:

$$\beta(t) = A + A_1 \sin(\omega_1 t) + A_2 \sin(\omega_2 t + \pi/4) + A_3 \sin(\omega_3 t + \pi/6) \sin(0.005t), \tag{3.4}$$

其中 $A = 0.0003$, $A_1 = 0.00015$, $A_2 = 0.00012$, $A_3 = 0.00010$，角频率分别为 $\omega_1 = 2\pi/(365/2)$、 $\omega_2 = 2\pi/365$和 $\omega_3 = 2\pi/(365/2)$。每个时间步都会对 $\beta(t)$添加小幅随机波动，以考虑不规则的环境变异，并将数值截断为非负值。通过维纳过程引入加性随机性， $dW_i \sim \mathcal{N}(0, \sqrt{dt})$ $i = 1$, 2，并使用欧拉–丸山方法 [42] 进行数值积分，步长为 $dt = 0.01$。使用不同噪声强度重复模拟

$\sigma_1 = \sigma_2 = 0$，0。1，0。2，0。3，以及0。4，以考察随机性增强对模型预测结果的影响。

随后，利用 $S(t)$和 $I(t)$的模拟轨迹，按照第2节提出的方法，使用时变参数学习SIR模型的动力学。为发现控制方程，我们将系统表示为如下形式

$$\begin{cases} \frac{dS}{dt} = f(S, I), \\ \frac{dI}{dt} = f(S, I), \end{cases} \tag{3.5}$$

其中，时变参数按照第2节所述程序进行系统选择。

3.2 CR模型

CR模型 [41] 是用于研究两个种群之间相互作用的基本框架：资源种群（例如猎物或营养源）和消费者种群（例如捕食者或依赖该资源的生物）。该模型刻画了种群增长与相互作用的动力学，为理解生态系统提供了洞见。

在该模型中，资源种群（R）以与其规模成比例的速率增长，由增长率 r控制。 R 的增长受到消费者种群（C）消耗的制约，该消耗遵循饱和型功能反应。该反应由消耗率 a 和半饱和常数 b表征。同时，消费者种群（C）通过消耗资源而增长，但经历恒定的死亡率m。该模型的动力学由以下方程描述

$$\begin{cases} \frac{dR}{dt} = rR - \frac{aRC}{1+bR}, \\ \frac{dC}{dt} = \frac{aRC}{1+bR} - mC. \end{cases} \tag{3.6}$$

为了将随机波动纳入模型，我们在系统(3.6)的每个方程中都加入了噪声。受噪声影响的捕食者-猎物模型形式如下

$$\begin{cases} dR(t) = rR(t)dt - \dfrac{aR(t)C(t)}{1+bR(t)}dt + \sigma_1 dW_1, \\ dC(t) = \dfrac{aR(t)C(t)}{1+bR(t)}dt - mC(t)dt + \sigma_2 dW_2, \end{cases} \tag{3.7}$$

其中 $\sigma_i, i = 1,2$，表示白噪声的强度， $W_i(t)$ $i = 1$，2，表示相互独立的维纳过程。

我们使用欧拉–丸山方法模拟了CR模型的动力学，该方法非常适用于随机微分方程。模型参数设为：资源增长率 $r = 0.1$、消耗率 $a = 0.005$、半饱和常数 $b = 0.001$以及消费者死亡

率 $m = 0.025$。资源和消费者的初始种群数量分别设为 $R_0 = 20$ 和 $C_0 = 10$。通过从 $dW_i \sim \mathcal{N}(0, \sqrt{dt})$ $i = 1, 2$, 采样维纳过程，在两个方程中引入加性环境随机性，并针对一系列噪声强度 $\sigma_1 = \sigma_2 \in \{0,\ 0.5,\ 1.0,\ 1.5,\ 2.0\}$.重复进行模拟。系统模拟时长为1500个时间单位，时间步长为 $dt = 0.01$。每次欧拉–丸山更新后，均通过设置较小的正下限来避免种群数量出现负值，从而保持生物学上的合理性。

我们使用 C 和 R 的模拟时间序列，采用第2节所述的提出方法学习CR模型的动力学。在学习阶段，我们假设该系统可以表示为一般形式

$$\begin{cases} \frac{dR}{dt} = f(C, R), \\ \frac{dC}{dt} = f(C, R), \end{cases} \tag{3.8}$$

其中，时变参数按照第2节中的流程进行系统选取。

3.3 气体浓度

CO_2 和CH_4 是两种主要的温室气体，油砂尾矿区中存在大量此类气体。我们的目标是基于相关天气条件和稀释剂输入，学习一个能够捕捉艾伯塔省油砂尾矿池中CO_2 和 CH_4 浓度时间序列的方程组。此外，我们还利用所学习的方程，根据估计的天气条件和稀释剂输入预测未来浓度。

我们使用了两个不同的空气监测站数据集，其中包含CO_2 和CH_4 的浓度数据，以及伍德布法罗环境协会（WBEA） [43], 提供的一个附近气象塔数据集。这些站点位于加拿大艾伯塔省东北部的伍德布法罗区域。我们使用的空气监测站为AMS-1: Bertha Ganter–Fort McKay（纬度：57.189428，经度：-111.640583）和AMS-18: Stony Mountain（纬度：55.621408，经度：-111.172686）。我们使用的气象塔为JP104-Site 1004（纬度：57.11901，经度：-111.42542）。本研究所使用的参数及其来源见表1。空气监测站和气象塔数据可从 [43]获取。所收集数据的时间段为2019年1月至2024年12月。空气监测站按小时记录数据，其中存在少量缺失数据。我们删除了包含缺失测量值的行，并对记录数据取日平均值。

两个站点均测量 CO_2 和 CH_4 ，以及一些常见的气象变量（气温、相对湿度、风速、风向、降水和全球辐射）（表 1）。我们希望每个数据集中包含相同数量的气象变量（气温、相对湿度、风速、风向、全球辐射、气压、DEW、PAR、SR 和降水）。因此，其他气象变量（气压、DEW、PAR 和 SR）被视为在两个数据集中相同，它们

采集自附近的气象塔JP104（表 1）。

为纳入产生这些气体的碳氢化合物来源，我们使用了艾伯塔能源监管局（AER）报告中的稀释剂总损失数据 [44]。由于不清楚哪些监测站采集的是来自哪些具体尾矿池的气体，我们汇总了该区域所有报告中的采矿公司在2019年1月至2024年12月期间按月报告的稀释剂损失数据。随后，我们将总量平均分配到各个月份，以估算每日稀释剂损失。

表1：本案例研究中使用的变量和数据来源。

变量	**缩写**	**Unit**	**数据来源**
甲烷	CH_4	百万分率 (ppm)	AMS-1、AMS-18
二氧化碳	CO_2	百万分率 (ppm)	AMS-1、AMS-18
环境温度	AT	摄氏度（°C）	AMS-1，AMS-18
相对湿度	RH	百分比（%）	AMS-1，AMS-18
风速	WS	千米/小时$^{-1}$	AMS-1、AMS-18
风向	WD	度 $\in [0, 360)$	AMS-1、AMS-18
总辐射	GR	W m^{-2}	AMS-1 ，AMS-18
降水	PC	毫米（mm）	AMS-1，AMS-18
气压	BP	mB	JP104
露点	DEW	摄氏度（°C）	JP104
16米处的光合有效辐射	PAR	μ摩尔 米$^{-2}$ 秒$^{-1}$	JP104
太阳辐射	SR	W m^{-2}	JP104
稀释剂	DIL	m^3	AER

为学习方程组，我们从候选非线性函数的函数库 $\Theta(\mathbf{X}, \mathbf{U})$（方程（2.4））中选取了最高至二次项，用二次方程拟合每种气体的数据。我们将稀释剂损失数据视为 CO_2 和 CH_4, 的来源；由于气体浓度会随天气条件变化，因此将所有气象变量都纳入候选项，作为输入变量。此外，CH_4 可通过 CO_2的化学反应生成。因此，基于这些考虑，我们认为学习得到的方程组具有以下一般形式

$$\begin{cases} \frac{dCO_2}{dt} = f(DIL, \text{weather variables}), \\ \frac{dCH_4}{dt} = f(CO_2, DIL, \text{weather variables}), \end{cases} \tag{3.9}$$

其中，气象变量包括气温、相对湿度、风速、风向、全球辐射、气压、DEW、PAR、SR 和降水，详见表 1。

3.4 蓝细菌细胞计数

蓝藻水华常见于淡水生态系统，且可能具有很强的毒性。它们会产生有害毒素，污染水源，从而对水生生物、野生动物和人类健康构成重大风险。我们的目标是通过数据驱动发现蓝藻细胞计数的方程，并基于所发现的方程进行预测，依据包括

营养物质可利用性、气象条件、湖泊特征和流域数据。

该数据集包含加拿大艾伯塔省 78 个湖泊和水库的观测数据，时间跨度为 1986 年至 2017 [45]。这些湖泊和水库位于北纬 49.3° 至 58.8°、西经 110.0079° 至 119.21667° 之间 [46]。湖泊监测数据采集时间为 5 月至 10 月，其中包括通过艾伯塔湖泊监测计划收集的浮游植物数据。原始湖泊监测数据集包含多种参数，例如湖水中的化学物质浓度、塞氏透明度测量值、湖泊分层数据、蓝藻细胞计数等，由艾伯塔省环境与公园部（AEP）提供。我们使用了 Heggerud等人 [46], 处理后的数据集，可从 [45] 获取。我们对不同湖泊的数据按年份和月份取平均值。本案例研究使用的变量见表 2。

表 2：CB细胞计数所用变量及其缩写列表

变量	**描述**
CB_{cell}	监测日采样的蓝细菌细胞计数
P	监测日采样的磷浓度
N	监测日采样的氮浓度
SD	监测日湖泊的塞氏透明度
ST	监测日湖泊的分层状态
AT	从监测日起至2周的日平均气温 在下一个可用的CB样本之前
PC	从监测日起至2周的日平均降水量 在下一个可用样本之前
WS	从监测当天到下一个可用样本前 2 周的日平均风速 在下一个可用样本之前
SR	从监测当天到下一个可用样本前 2 周的日平均太阳辐射 在下一个可用样本之前
LD	湖泊深度
LE	湖面高程
PL	流域中牧草地的百分比
CL	流域中耕地的百分比
FT	流域中森林的百分比
WL	流域中湿地的百分比
LWAR	湖泊-流域面积比

我们使用第 2 节所述的方法，根据表 2 中列出的所有协变量，学习了一个用于 CB 细胞计数（CB_{cell}）的单一方程。所学方程的一般形式为

$$\frac{dCB_{cell}}{dt} = f(CB_{cell}, \text{湖泊监测}, \text{气象}, \text{湖泊特征}, \text{流域特征}), \tag{3.10}$$

其中，湖泊监测包括 P、N、SD 和 ST。气象包括 AT、PC、WS 和 SR。湖泊特征包括 LD 和 LE。流域特征包括 PL 、 CL 、 FT 、 WL 和 LWAR。CB_{cell} 保留在方程右侧，以表示新的细胞计数取决于现有的细胞计数。

3.5 模拟天气数据

通过时变参数学习系统后，我们使用机器学习模型根据天气条件预测这些参数。在本研究中，我们将温度、湿度、风速和降水这四种气象变量作为时变参数的预测变量。对于气体和 CB 数据集，这些气象变量在数据集中均可获取，并直接用于预测。相比之下，对于 SIR 和 CR 模型，我们模拟了相应的气象变量，以便以相同方式应用机器学习模型，从而根据这些模拟的环境条件预测时变参数。

我们使用 Ornstein-Uhlenbeck（OU）过程 [47] 来模拟气象变量的时间序列。OU 过程由随机微分方程（SDE）表示，通常用于对均值回复现象建模。该过程由以下方程描述

$$dX(t) = -\theta_X \left(X(t) - X_{\text{mean}}\right) dt + \sigma_X \, dW(t), \tag{3.11}$$

其中， $X(t)$ 表示时间 t 时的状态变量（例如温度）， X_{mean} 是过程围绕其波动的长期均值。参数 θ_X 表示均值回复速率，即偏离均值的情况被纠正的速度。项 σ_X 表示随机噪声的强度， $dW(t)$ 是维纳过程，用于引入随机性。

为了模拟 OU 过程的数据，我们使用 Euler-Maruyama 方法对连续时间方程进行离散化。离散化后的 SDE 表示为

$$X_{i+1} = X_i - \theta_X \left(X_i - X_{\text{mean}}\right) \Delta t + \sigma_X \sqrt{\Delta t} \, \xi_i,$$

w这里， Δt 是时间增量， ξ_i 是相互独立的标准正态随机变量 es.

天气模拟模型的参数值经过精心设定，使温度、湿度、风速和降水在模拟期间呈现与现实情况高度相似的动力学。我们采用了 $\Delta t = 0.01$ 的时间步长，足以细致捕捉总计 1500 天内天气条件的变化。初始温度为 20℃，初始湿度为 60%，初始风速为 10 m/s，初始降水量为 0.2 mm/day。各参数的长期均值分别设定为 $T_{\text{mean}} = 0$℃、 $H_{\text{mean}} = 50$%、 $W_{\text{mean}} = 5\,\text{m/s}$ 和 $P_{\text{mean}} = 0.1\text{mm/day}$。均值回复速率 ($\theta$) 分别设定为 $\theta_T = 0.1$、 $\theta_H = 0.05$、 $\theta_W = 0.2$ 和 $\theta_P = 0.1$。噪声强度分别设定为 $\sigma_T = 10$、 $\sigma_H = 5$、 $\sigma_W = 1$ 和 $\sigma_P = 0.05$，依次对应温度、湿度、风速和降水的波动。温度和湿度的噪声强度较高，反映了它们通常比风速和降水具有更大的变异性。此外，还为每个变量设定了边界，以避免出现不现实的数值。温度限制在 $[-20, 35]$℃ 范围内，湿度限制在 $[0, 100]$% 范围内，风速

至 $[0, 30]$m/s，降水量至 $[0, 100]$ mm/day。

3.6 机器学习模型、误差指标与交叉验证

首先，使用 30 点滚动窗口对每个变量的时间序列进行平滑处理，以减少高频噪声。为考虑不同变量之间的尺度差异，使用最小-最大缩放将所有序列归一化至 $[0, 1]$ 范围，

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}},$$

其中， X' 表示归一化后的变量， $X_{\min}$ 和 $X_{\max}$ 分别是 X的最小值和最大值。

作为案例研究，我们采用随机森林（RF）回归模型 [48]，根据四个气象协变量（气温、相对湿度、风速和降水）分别预测每个时变参数。RF 模型通过 sklearn.ensemble 中的 RandomForestRegressor 类 [49]实现。为确保集成稳定性，我们使用了 5000 棵树（n_estimators = 5000）。我们将最大树深度限制为 5（max_depth=5），并要求至少有 10 个样本才能分裂内部节点（min_samples_split=10），同时确保每个叶节点至少包含 5 个样本（min_samples_leaf=5），以控制模型复杂度。气象变量作为预测变量，每个时变参数则分别作为一个目标变量（图 1e）。

对于每个数据集，时间序列被划分为训练集、验证集和测试集。测试期间被分为五折，每折30天；CB数据集除外，其包含五个月度观测值。严格保留时间顺序。对于每个测试折，紧邻测试期间之前选取一个时长相同的验证折（图2）。使用方程（2.11）对训练数据拟合控制方程，并使用方程（2.12）生成预测。

使用扩展窗口时间序列交叉验证策略进行模型选择（图2）。窗口长度 w 以7天为增量，从7天变化至28天（对于CB数据集，则以5为增量，从5变化至25）。同时，时变参数的数量从零个（固定参数模型）逐步增加至全部活跃参数，并始终包含偏置项。从一组初始验证折（例如第1至第5折）开始，逐步扩展验证窗口（第1至第6折、第1至第7折，.....第1至第9折）（图2）。对于每种配置，使用平均绝对误差（MAE）评估性能，

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{y}_i - y_i \right|,$$

其中， y_i 和ˆy_i 分别表示观测值和预测值。选择使验证集平均绝对误差最小的配置，并用于预测后续测试

折。

为考察区间长度和时变参数数量的影响，我们在每个测试折上进行了模拟，并计算所得的平均绝对误差。在每一折中，我们将平均绝对误差最低的配置确定为最优配置，并与交叉验证选定模型进行比较。对于所有数据集，我们还将所提出的时变参数模型与固定参数基线进行比较；在后者中，整个时间序列由常系数表示（图 1b） [13]。两种情况下，正则化参数均设为 $\lambda = 0.01$，稀疏阈值设为 $\tau = 0.001$，最大迭代次数设为 10,000。

为进一步评估我们的框架，我们将其预测性能与两种标准机器学习基线进行了比较：混合卷积神经网络–长短期记忆网络模型 [50–52] 和梯度提升机（GBM） [53]。

对于 CNN–LSTM 模型，输入采用长度为 w 天的滑动窗口构建。对于每个目标变量，预测变量由相同 w 天区间内的外生协变量构成，而响应则由时间上对齐的相应 w 天目标状态序列构成。对于 SIR 和捕食者 - 猎物模型数据集，协变量包括气温（ AT ）、风速（WS）、降水（PC）和相对湿度（RH）。对于气体浓度和 CB 数据集，则使用表 1 和表 2 中列出的所有气象变量。

CNN–LSTM 架构结合了用于特征提取的卷积层 [52]与循环记忆单元 [51], ，并在 TensorFlow/Keras [54, 55] 中实现。该架构由一个一维因果卷积层组成，Conv1D（滤波器 $=F$，卷积核大小 $= K$，激活函数 $=$ ReLU），后接池化大小为 2 的最大池化层、一个包含 U 个隐藏单元的 LSTM 层、一个丢弃率为 d的丢弃层，以及两个全连接层：一个使用 ReLU 激活函数且包含 D 个单元的稠密层，以及一个最终的线性输出层，生成 w 个输出（在我们的实验中为 $w = 30$）。通过 KerasTuner [56] 随机搜索选择超参数（F、 K、U、 D、d）和学习率，以最小化验证集均方误差（MSE）。训练采用 Adam优化器 [57]并结合提前停止 [58]（耐心值为 10 个周期，并恢复表现最佳的权重），使用未进行随机打乱的时间序列交叉验证方案。在每一折中，将训练数据中与测试长度相匹配的最新数据块留作验证集。

作为额外的基线，我们实现了一个GBM模型，其所用预测变量集合与卷积神经网络–长短期记忆网络设置相同。GBM的输入由每个时间步的外生协变量组成，响应则为相应的目标变量值。该模型使用1000棵提升树、0.01的学习率、30的最大树深度以及0.9的子采样比例进行训练，以减轻过拟合。每个终端节点被约束为至少包含10个样本。采用了平方误差损失函数，对应于高斯回归框架。

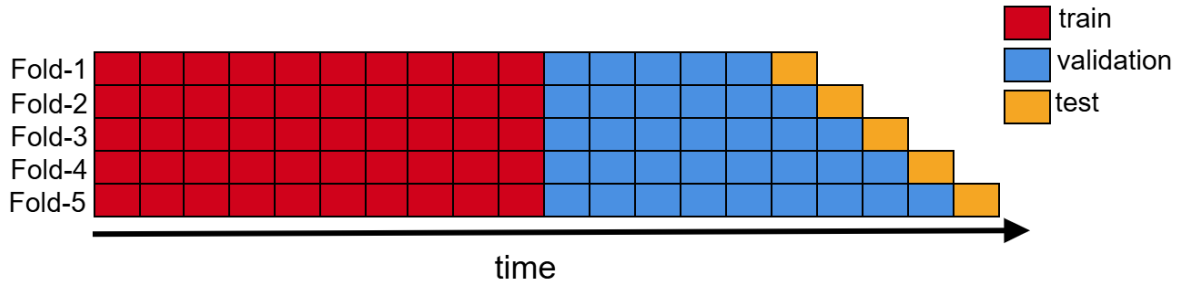


图2：本研究所采用的依序、不重叠的训练集、测试集和验证方案示意图。训练期间（红色）覆盖了时间序列的大部分。验证窗口块（蓝色）的长度与测试窗口相同，并紧邻测试期间（黄色）之前。对于每一折测试，仅向前扩展验证窗口，而训练集保持固定，从而产生多个验证折，同时不扩展训练数据，也不使其重叠。

4 结果

4.1 理论保证

在本小节中，我们为我们框架的三个关键分量建立理论保证：（i）从学习得到的方程和预测的时变参数到预测状态的误差传播，（ii）学习得到的方程与预测方程之分裂结构的误差界，以及（iii）时变参数化相对于建立在同一函数库上的任何固定参数模型具有严格更紧的有限时域预测误差界的条件。

我们的分析基于经典常微分方程稳定性理论和稀疏方程发现的结果 [12, 59, 60]。在此基础上，我们推导了时变参数模型的有限时域预测误差界，并量化了机器学习预测参数的误差向状态轨迹传播的过程。

4.1.1 有限时域预测误差界

现在，我们量化已学习控制方程中的误差以及预测时变参数中的误差如何传播为预测轨迹中的误差。设 $\mathbf{f}$： $[0, H] \times \mathbb{R}^n \to \mathbb{R}^n$表示真实系统的（未知）右端项，

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t)), \qquad t \in [0, H],$$

其中， $H > 0$ 是有限时间范围。我们假设$\mathbf{f}(\cdot)$可以在紧区域上由基于候选库 $\Theta(\cdot)$构造的表示进行一致逼近。

假设4.1.1 (Lipschitz动力学和函数库近似 [12, 59, 60])。 设 $D \subset \mathbb{R}^n$是一个包含所关注轨迹的紧集。假设：

(a) (*Li* 施茨动力学）对于每个 $t \in [0, H]$, 映射 $x \mapsto f(t, x)$ 满足全局 *Lipschitz* 连续性，*n*

D 其常数为 $L > 0$:

$$\|f(t,x) - f(t,y)\| \le L\|x - y\| \quad \textit{for all } x, y \in D.$$

(b)（库近似）存在系数函数 $\xi^\star$： $[0,H] \to \mathbb{R}^p$ 和库向量 Θ： $D \times [0,H] \to \mathbb{R}^p$，使得

$$\big\|f(t,x) - \Theta(x,u(t))\,\xi^\star(t)\big\| \le \varepsilon_{\text{lib}} \quad \textit{for all } (t,x) \in [0,H] \times D,$$

对于某个 $\varepsilon_{\text{lib}} \ge 0$。

(c)（有界库）存在 $B_\Theta > 0$，使得 $\|\Theta（x, u（t））\|_{\text{op}} \le B_\Theta$ 对于所有 $（t, x）\in [0,H] \times D$均成立，其中 $\|\cdot\|_{\text{op}}$ 表示谱范数。

ˆ 我们的预测流程通过在滑动窗口上使用 STRR，随后进行机器学习模型预测，生成估计的时变系数 $\xi(t)$，并采用以下模型

$$\hat{f}(t,x) := \Theta(x,u(t))\,\hat{\xi}(t).$$

我们假设机器学习模型预测在整个预测时域内，以一致有界误差近似理想的系数函数 $\xi^\star(t)$。

假设 4.1.2（系数预测精度）。 存在 $\varepsilon_{\text{par}} \ge 0$使得

$$\sup_{t\in[0,H]} \big\|\hat{\xi}(t) - \xi^\star(t)\big\| \;\le\; \varepsilon_{\text{par}}.$$

在假设 4.1.1-4.1.2 下，真实向量场 f 与学习预测向量场 $\hat{f}$ 之间的差异一致有界:

$$\big\|f(t,x) - \hat{f}(t,x)\big\| \le \varepsilon_{\text{lib}} + B_\Theta\varepsilon_{\text{par}} =: \delta \quad \text{for all } (t,x) \in [0,H] \times D.$$

现在，我们来量化这一误差界对状态轨迹的影响。

定理 *4.1.1*（有限时域预测误差界）。假设假设 *4.1.1* 和 *4.1.2* 成立，且令 $\delta = \varepsilon_{\text{lib}} + B_\Theta\varepsilon_{\text{par}}$。令 $x:\,[0,H] \to D$ 为满足真实系统 $\dot{x}(t) = f(t, x(t))$ 的解，其初始条件为 $x(0) = x_0$,；再令 $\hat{x}:\,[0,H] \to D$ 为满足学习得到的系统 $\dot{\hat{x}}(t) = f(t, \hat{x}(t))$ 的解，且具有相同的初始条件 $\hat{x}(0) = x_0$。那么，对所有 $t \in [0,H]$,

$$\|x(t) - \hat{x}(t)\| \;\le\; \frac{\delta}{L}\big(e^{Lt} - 1\big) \;\le\; \frac{\delta}{L}\big(e^{LH} - 1\big).$$

证明。 定义误差 $e(t) := x(t) - \hat{x}(t)$。求导得

$$\dot{e}(t) = f(t,x(t)) - \hat{f}(t,\hat{x}(t)) = \big(f(t,x(t)) - f(t,\hat{x}(t))\big) + \big(f(t,\hat{x}(t)) - \hat{f}(t,\hat{x}(t))\big).$$

取范数并应用假设 4.1.1(a)以及 δ 的定义，可得

$$\|\dot{e}(t)\| \le L\|e(t)\| + \delta.$$

令 $y(t) := \|e(t)\|$。则 $y(0) = 0$并且

$$\dot{y}(t) \le Ly(t) + \delta \quad \text{for a.e. } t \in [0, H].$$

根据格朗沃尔不等式的微分形式 [59, 61],

$$y(t) \le \frac{\delta}{L}\big(e^{Lt} - 1\big) \quad \text{for all } t \in [0, H],$$

由此证明了所述界。 □

定理 4.1.1 是一个标准的常微分方程稳定性结果，针对我们基于函数库的表示进行了专门化。它表明，我们流程中的所有误差来源都通过单一量 $\delta = \varepsilon_{\text{lib}} + B_\Theta \varepsilon_{\text{par}}$ 影响状态预测，该量综合了函数库表示的近似误差和时变参数预测的误差。

4.1.2 学习方程与预测方程分裂结构的界

现在，我们将基线界专门应用于本文的学习得到的预测模型。作为参考，我们重述学习得到的动力学（2.11）的分裂形式及其对应的预测形式（2.12）。令

$$\dot{\mathbf{x}}(t) = \mathbf{f}_{\text{tv}}(t, \mathbf{x}(t)) := \Theta_{\text{top}N}(x(t), u(t))\, \xi_{\text{tv}}(t) + \Theta_{\text{fix}}(x(t), u(t))\, \bar{\xi}, \tag{4.1}$$

其中 $\Theta_{\text{fix}} := \Theta_{\text{active}} \setminus \Theta_{\text{top}N}$ 且 $\bar{\xi}$ 表示非前n 个活跃项的固定系数向量。预测模型为

$$\dot{\hat{\mathbf{x}}}(t) = \hat{\mathbf{f}}_{\text{tv}}(t, \hat{\mathbf{x}}(t)) := \Theta_{\text{top}N}(\hat{x}(t), u(t))\, \hat{\xi}_{\text{tv}}(t) + \Theta_{\text{fix}}(\hat{x}(t), u(t))\, \bar{\xi}. \tag{4.2}$$

假设 4.1.3 （前n 函数库块的一致有界性）。 存在 $B_{\text{top}N} > 0$使得

$$\sup_{t\in[0,H]} \sup_{x\in D} \|\Theta_{\text{top}N}(x, u(t))\|_{\text{op}} \le B_{\text{top}N}.$$

假设 4.1.4 （时变系数预测的一致精度）。 存在 $\varepsilon_{\text{top}N} \ge 0$ 使得

$$\sup_{t\in[0,H]} \|\hat{\xi}_{\text{tv}}(t) - \xi_{\text{tv}}(t)\| \le \varepsilon_{\text{top}N}.$$

定理 4.1.2 （分割模型预测界：仅 $\Theta_{\text{top}N}$ 预测误差参与）。 假设对于 f_{tv}， 式 (4.1) 满足假设 *4.1.1a*，且假设 *4.1.3–4.1.4* 成立。令 $\dot{\mathbf{x}}$ 为式 (4.1) 的解，并

$\dot{\mathbf{x}}$ 为式 (4.2) 的解，且具有相同的初始条件 $\mathbf{x}(0) = \hat{\mathbf{x}}(0)$。若 $x(t)$、$\hat{x}(t) \in D$ 对所有 $t \in [0, H]$ 均成立，则对所有 $t \in [0, H]$,

$$\|x(t) - \hat{x}(t)\| \le \frac{B_{\mathrm{top}N}\varepsilon_{\mathrm{top}N}}{L}\big(e^{Lt} - 1\big) \le \frac{B_{\mathrm{top}N}\varepsilon_{\mathrm{top}N}}{L}\big(e^{LH} - 1\big).$$

证明。 对于任意 $(t, x) \in [0, H] \times D$，固定块恰好抵消：

$$f_{\mathrm{tv}}(t, x) - \hat{f}_{\mathrm{tv}}(t, x) = \Theta_{\mathrm{top}N}(x, u(t))\big(\xi_{\mathrm{tv}}(t) - \hat{\xi}_{\mathrm{tv}}(t)\big).$$

因此，根据假设 4.1.3–4.1.4,

$$\sup_{(t,x)\in[0,H]\times D} \|f_{\mathrm{tv}}(t, x) - \hat{f}_{\mathrm{tv}}(t, x)\| \le B_{\mathrm{top}N}\varepsilon_{\mathrm{top}N}.$$

应用定理 4.1.1，并令 $\delta = B_{\mathrm{top}N}\varepsilon_{\mathrm{top}N}$，即可得到所述误差界。 □

注释 4.1.1。 定理 *4.1.2* 适用于 (4.1)–(4.2) 的分裂结构：固定系数-¯向量 ξ 不会出现在误差界的强迫项中。因此，提高所选 $\Theta_{\mathrm{top}N}$ 项的 ξ_{tv} 预测准确度，可直接收紧有限时域状态预测界；而固定块仅通过所得向量场的利普希茨常数 L 间接影响该界。

4.1.3 时变参数优于固定参数的情形

最后，我们将时变参数化与固定参数模型进行比较。令$\mathcal{C}_{\mathrm{const}}$ 表示常系数向量的类别

$$\mathcal{C}_{\mathrm{const}} := \{\xi : [0, H] \to \mathbb{R}^p : \ \xi(t) \equiv c\},$$

并令 $\mathcal{C}^{(m)}{}_{\mathrm{tv}}$ 表示至多包含 m 个子时间区间的分段常值系数函数类：

$$\mathcal{C}^{(m)}_{\mathrm{tv}} := \Big\{\xi : [0, H] \to \mathbb{R}^p : \ \xi(t) = c_r \text{ for } t \in I_r, \ r = 1, \dots, m\Big\},$$

其中 $\{I_r\}_{r=1}^m$ 是将 $[0, H]$ 分成 m 个时间区间的一个划分。我们量化这些函数类对理想系数轨迹 $\xi^\star(t)$的逼近程度。

定义 4.1.1 （最佳一致系数逼近误差）。 设

$$E_{\mathrm{const}} := \inf_{\xi\in\mathcal{C}_{\mathrm{const}}} \sup_{t\in[0,H]} \|\xi^\star(t) - \xi(t)\|, \qquad E^{(m)}_{\mathrm{tv}} := \inf_{\xi\in\mathcal{C}^{(m)}_{\mathrm{tv}}} \sup_{t\in[0,H]} \|\xi^\star(t) - \xi(t)\|.$$

显然， $\mathcal{C}_{\mathrm{const}} \subset \mathcal{C}^{(m)}{}_{\mathrm{tv}}$，因此对每个 m，都有 $E^{(m)}{}_{\mathrm{tv}} \le E_{\mathrm{const}}$。下一个引理表明，在满足温和正则性条件时，只要 $\xi^\star(t)$ 确实随时间变

化，当 足够大时，不等式便是严格的。这是这样一个事实的构造性实例：阶梯函数在上确界范数 [62, 意义下稠密于 $C([0,H])$。第 2] 章。

假设4.1.5（真实系数的正则性与非平稳性）。 $\xi^\star$（t）的每个分量在 $[0,H]$上连续，且 $\xi^\star$ 在 $[0,H]$ 上不是常数（即存在 t_1、 t_2，使得$\xi^\star$（t_1） $\neq \xi^\star$（t_2））。

引理4.1.1（阶梯函数逼近优于常数逼近）。假设假设*4.1.5*成立。则 $E_{\text{const}} > 0$，并且对每一个 $0 < \varepsilon < E_{\text{const}}$，存在整数 $m_\varepsilon \geq 1$，使得对所有 $m \geq m_\varepsilon$，

$$E_{\text{tv}}^{(m)} \leq \varepsilon < E_{\text{const}}.$$

特别地，对所有充分大的 m，我们有 $E^{(m)}{}_{\text{tv}} < E_{\text{const}}$。

证明。由于 $\xi^\star$ 在紧区间 $[0,H]$上连续，故它是一致连续的。因此，对任意 $\varepsilon > 0$，存在 $\delta > 0$，使得只要 $|t-s| < \delta$，就有 $\|\xi^\star(t) - \xi^\star(s)\| < \varepsilon$。

给定这样的 δ，选取足够大的 m_ε，使得将 $[0,H]$ 划分为 m_ε 个子区间时，每个时间区间的长度至多为 δ。定义一个分段常数函数 $\tilde{\xi} \in \mathcal{C}^{(m_\varepsilon)}{}_{\text{tv}}$，方法是令$\xi(t)$等于 $\xi^\star$ 在每个子区间中某一参考点处的取值。由构造可知，

$$\sup_{t\in[0,H]} \|\xi^\star(t) - \tilde{\xi}(t)\| \leq \varepsilon,$$

因此 $E^{(m_\varepsilon)}{}_{\text{tv}} \leq \varepsilon$，且同一界对所有 $m \geq m_\varepsilon$成立。

为证明 $E_{\text{const}} > 0$，注意到根据假设4.1.5，存在 t_1, t_2，使得 $\xi^\star(t_1) \neq \xi^\star(t_2)$。令

$$V := \|\xi^\star(t_1) - \xi^\star(t_2)\| > 0.$$

对任意常向量 $c \in \mathbb{R}^p$，

$$V = \|\xi^\star(t_1) - \xi^\star(t_2)\| \leq \|\xi^\star(t_1) - c\| + \|\xi^\star(t_2) - c\| \leq 2 \sup_{t\in[0,H]} \|\xi^\star(t) - c\|,$$

因此对每个 c，$\sup_{t\in[0,H]} \|\xi^\star(t) - c\| \geq V/2$。对 c取下确界即得 $E_{\text{const}} \geq V/2 > 0$。再任取 $0 < \varepsilon < E_{\text{const}}$，即可得到所述结论。 □

现在，我们结合引理4.1.1与定理 4.1.1，得出一个充分条件：在该条件下，时变参数化的有限时域预测误差界严格小于任何固定参数模型的有限时域预测误差界。

定理 4.1.3（当时变参数优于固定参数时）。 假设条件 *4.1.1*、*4.1.2* 和 *4.1.5* 成立。考虑基于同一函数库Θ构建的两个理想化模型：

- 一个系数轨迹为$\xi^\star{}_{\text{const}} \in \mathcal{C}_{\text{const}}$ 且误差为E_{const}的固定参数模型，

• 一个具有 m 时间区间和系数轨迹$\xi^{\star}{}_{\mathrm{tv}} \in \mathcal{C}_{(m)\ \mathrm{tv}}$ 且误差为 $E_{(m)\ \ \mathrm{tv}}$ 的时变模型。

令$\hat{}\mathbf{x}^{(m)}{}_{\mathrm{tv}}(t)$ 表示时变模型的轨迹

$$\dot{\hat{\mathbf{x}}}(t) = \boldsymbol{\Theta}(\hat{\mathbf{x}}(t), \mathbf{u}(t))\, \boldsymbol{\xi}^{\star}_{\mathrm{tv}}(t), \qquad \hat{\mathbf{x}}^{(m)}_{\mathrm{tv}}(0) = \mathbf{x}_0,$$

其中$\boldsymbol{\xi}^{\star}{}_{\mathrm{tv}} \in \mathcal{C}^{(m)}{}_{\mathrm{tv}}$。对于这两个模型，定义

$$\delta_{\mathrm{const}} := \varepsilon_{\mathrm{lib}} + B_{\Theta} E_{\mathrm{const}}, \qquad \delta^{(m)}_{\mathrm{tv}} := \varepsilon_{\mathrm{lib}} + B_{\Theta} E^{(m)}_{\mathrm{tv}}.$$

则对于任意 $H > 0$以及任意满足 $E^{(m)}{}_{\mathrm{tv}} < E_{\mathrm{const}}$的 $m \geq m_{\varepsilon}$，时变模型的有限时域预测误差界满足

$$\sup_{t\in[0,H]} \|x(t) - \hat{x}^{(m)}_{\mathrm{tv}}(t)\| \ \leq\ \frac{\delta^{(m)}_{\mathrm{tv}}}{L}\big(e^{LH} - 1\big) < \frac{\delta_{\mathrm{const}}}{L}\big(e^{LH} - 1\big) \ \geq\ \sup_{t\in[0,H]} \|x(t) - \hat{x}_{\mathrm{const}}(t)\|.$$

特别地，对于所有足够大的 m，时变参数化所具有的最坏情形有限时域预测误差界严格小于基于同一函数库的任何固定参数模型。

证明。 根据引理4.1.1，对于所有足够大的 m ，存在一条时变系数轨迹 $\xi^{\star}_{\mathrm{tv}} \in \mathcal{C}^{(m)}{}_{\mathrm{tv}}$，满足 $E^{(m)}{}_{\mathrm{tv}} < E_{\mathrm{const}}$。对于 m 的这一选择，我们有 $\delta^{(m)}{}_{\mathrm{tv}} < \delta_{\mathrm{const}}$。将定理 4.1.1分别应用于固定参数模型和时变模型，可得

$$\sup_{t\in[0,H]} \|x(t) - \hat{x}_{\mathrm{const}}(t)\| \leq \frac{\delta_{\mathrm{const}}}{L}(e^{LH} - 1), \qquad \sup_{t\in[0,H]} \|x(t) - \hat{x}^{(m)}_{\mathrm{tv}}(t)\| \leq \frac{\delta^{(m)}_{\mathrm{tv}}}{L}(e^{LH} - 1),$$

并且严格不等式 $\delta^{(m)}{}_{\text{时变}} < \delta_{\mathrm{const}}$ 给出了所声称的界的排序。 □

定理4.1.3将如下直觉形式化：当真实参数随时间呈非平稳时，足够丰富的时变参数化能够比任何固定参数模型更准确地表示系数轨迹，因而对有限时域预测误差给出严格更紧的最坏情形界。在实践中，我们的STRR+随机森林流程会在窗口长度和时变参数数量上进行搜索，从而对上文所识别的理论上有利的情形给出数据驱动的近似。

4.2 实验结果

4.2.1 学习并复现系统动力学

在SIR数据集中，时变参数模型展现出强大的动力学捕捉能力，易感者（S）种群的平均绝对误差达到0.9%，以及1.7%

对感染者（I）种群而言（图3a）。相比之下，固定参数模型产生的误差显著更高——S 为29.7%， I 为29.9%（图3a）。在所有噪声水平下，时变模型始终将平均绝对误差的平均值保持在2%以下，而固定参数模型的准确度则差得多，平均绝对误差介于28%至33%之间（补充信息附录图6）。

对于CR数据集，时变参数模型有效地再现了时间序列动力学，消费者（C）变量的平均绝对误差为1.9%，资源（R）变量的平均绝对误差为2.5%（图3b）。相比之下，固定参数模型的表现明显较差， C 的平均绝对误差为21.5%， R 的平均绝对误差为31.1%（图3b）。在不同噪声水平下进行考察时，时变模型的平均绝对误差稳定在1.6–2.7%之间，而固定参数模型则在11–38%之间波动（补充信息附录图6）。

在气体数据集中，时变参数方法同样表现出优异性能，成功捕捉了CO_2 和CH_4 的时间变化特征，其平均绝对误差分别为1.4%和1.9%（图3c）。相反，固定参数模型的平均绝对误差明显更高，CO_2 为16.9%，CH_4 为9.1%（图3c）。在多个站点数据集上取平均后，时变模型的总体平均绝对误差保持在3%以下，而固定参数模型则介于11–17%之间（补充信息附录图6）。

在CB数据集中，两种模型的表现相当，时变模型和固定参数模型的平均绝对误差几乎相同，分别为3.51%和3.53%（图3d）。

样例图显示，在我们的案例研究中，除CH_4和CB外，固定参数模型通常无法捕捉大多数变量时间序列中不断变化的趋势（图10）。该模型在仅出现单个峰值的CB数据集上表现良好，在具有多个峰值的CH_4, 上表现尚可（补充信息附录图10）。相比之下，时变参数模型能够准确捕捉所有变量的时间动态（补充信息附录图10）。

4.2.2 预测系统动态

对于SIR数据集，时变参数模型对易感者（S）和感染者（I）种群均展现出较强的预测能力，平均绝对误差分别为10.0%和9.9%（图4a）。相比之下，固定参数模型的误差明显更高， S 的平均绝对误差为13.8%， I 为14.0%（图4a）。在所有噪声水平下，时变模型的平均绝对误差始终保持在9%至13%之间，而固定参数模型的表现则有所下降，误差介于12%至15%之间（补充信息附录图7）。

纯数据驱动基线的预测表现明显较差。CNN–LSTM模型对 S 和 I的平均绝对误差均超过21%，而梯度提升机取得的误差略

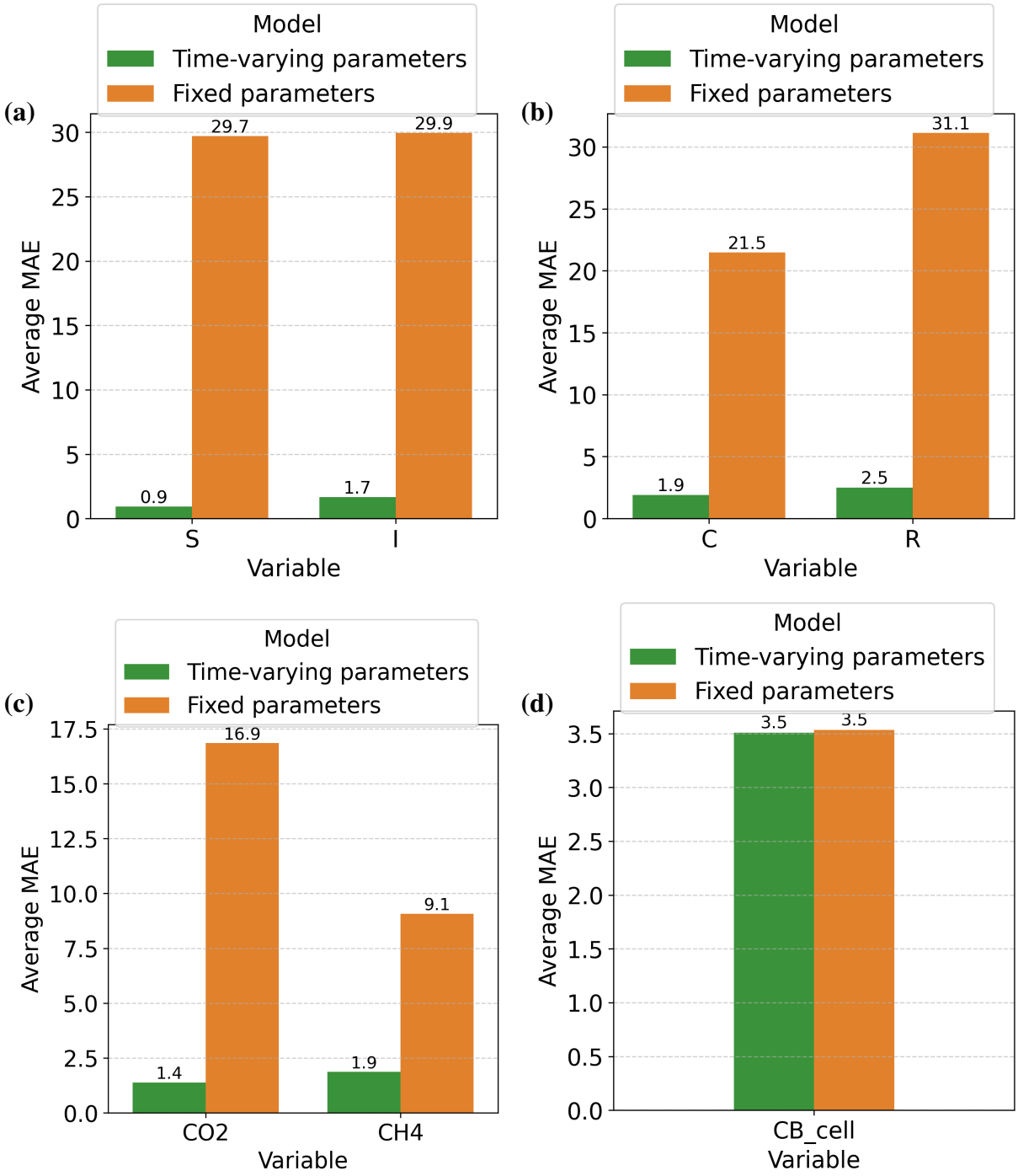


图 3： (a) SIR数据集、(b) CR数据集、(c) 气体数据集和 (d) CB数据集的平均学习MAE。绿色柱表示时变参数模型，橙色柱表示固定参数模型。针对每个数据集，在不同噪声水平或不同监测站下分别计算各变量的MAE，然后取平均值。

低一些，但两个变量的误差仍高达约21%（图4a）。随着噪声水平不断提高，这一性能差距依然存在：与机理模型相比，CNN–LSTM和梯度提升机的误差急剧上升（补充信息附录图7）。

在 CR数据集中，时变参数模型准确预测了消费者（C）和资源（R）变量的动力学，平均绝对误差分别为 3.9% 和

5.4%（图4b）。固定参数模型的准确度相当，C 的平均绝对误差为 3.9%，R 的平均绝对误差为 5.1%（图4b）。在不同噪声条件下，两种机理模型的预测误差均保持相对较低，通常介于 2% 到 7% 之间（补充信息附录图7）。

相比之下，卷积神经网络–长短期记忆网络和梯度提升机基线的误差要高得多。卷积神经网络–长短期记忆网络在 C 和 R上的平均绝对误差分别约为18.5%和18.8%，而梯度提升机的平均绝对误差分别约为18.0%和15.9%（图4b）。在所有噪声水平下，卷积神经网络–长短期记忆网络和梯度提升机的误差仍约为机理模型的三至五倍（补充信息附录图7）。

在气体数据集中，时变模型有效地预测了CO_2 和CH_4, 的时间变化特征，平均绝对误差分别为10.4%和12.0%（图4c）。相比之下，固定参数模型对CH_4, 的预测误差较高，平均绝对误差为24.5%（图4c）。在多个站点取平均后，时变方法的总体平均绝对误差维持在约11%，而固定参数模型的平均绝对误差则介于6%至27%之间（补充信息附录图7）。

卷积神经网络–长短期记忆网络和GBM模型的表现明显较差，卷积神经网络–长短期记忆网络的平均绝对误差在CH_4 中超过25%，在CO_2, 中超过35%；GBM的误差则分别高于16%和34%（图4c）。在不同站点之间，时变机理模型的表现仍相对稳定，而卷积神经网络–长短期记忆网络和GBM的误差在23%至34%之间波动（补充信息附录图7）。

对于CB数据集，时变参数模型和固定参数模型的预测性能相当，平均绝对误差分别为3.0%和3.3%（图4d）。卷积神经网络–长短期记忆网络基线的平均绝对误差约为2.2%，优于机理模型；而GBM的平均绝对误差约为10.8%。

4.2.3 超越交叉验证的最优预测

扩展窗口交叉验证方法并不总能确定时变参数模型最有效的一组超参数，即区间长度和时变参数数量。对于每个数据集，都存在能够产生更优预测性能的不同配置（补充信息附录表3、4、5、6）。

在 SIR数据集中，最优配置使感染者（I）种群的平均绝对误差达到 2.5%，易感者（S）种群的平均绝对误差达到 2.2%（图5a）——与扩展窗口交叉验证设置中的超参数相比，提升了约 8%。在 CR数据集中，该模型对消费者（C）和资源（R）动力学的平均绝对误差分别为 1.3% 和 2.7%（图5b），提升幅度约为 2%。

同样，对于气体数据集，最优配置在 CH_4 和 CO_2 上分别取得了 5.7% 和 2.6% 的平均绝对误差（图5c），与扩展窗口交叉验证方法相比，提升了约 5%。最后，在 CB数据集中，表现最佳的配置取得了 0.5% 的平均绝对误差（图5d），这表明预测性能提高了 2.5%

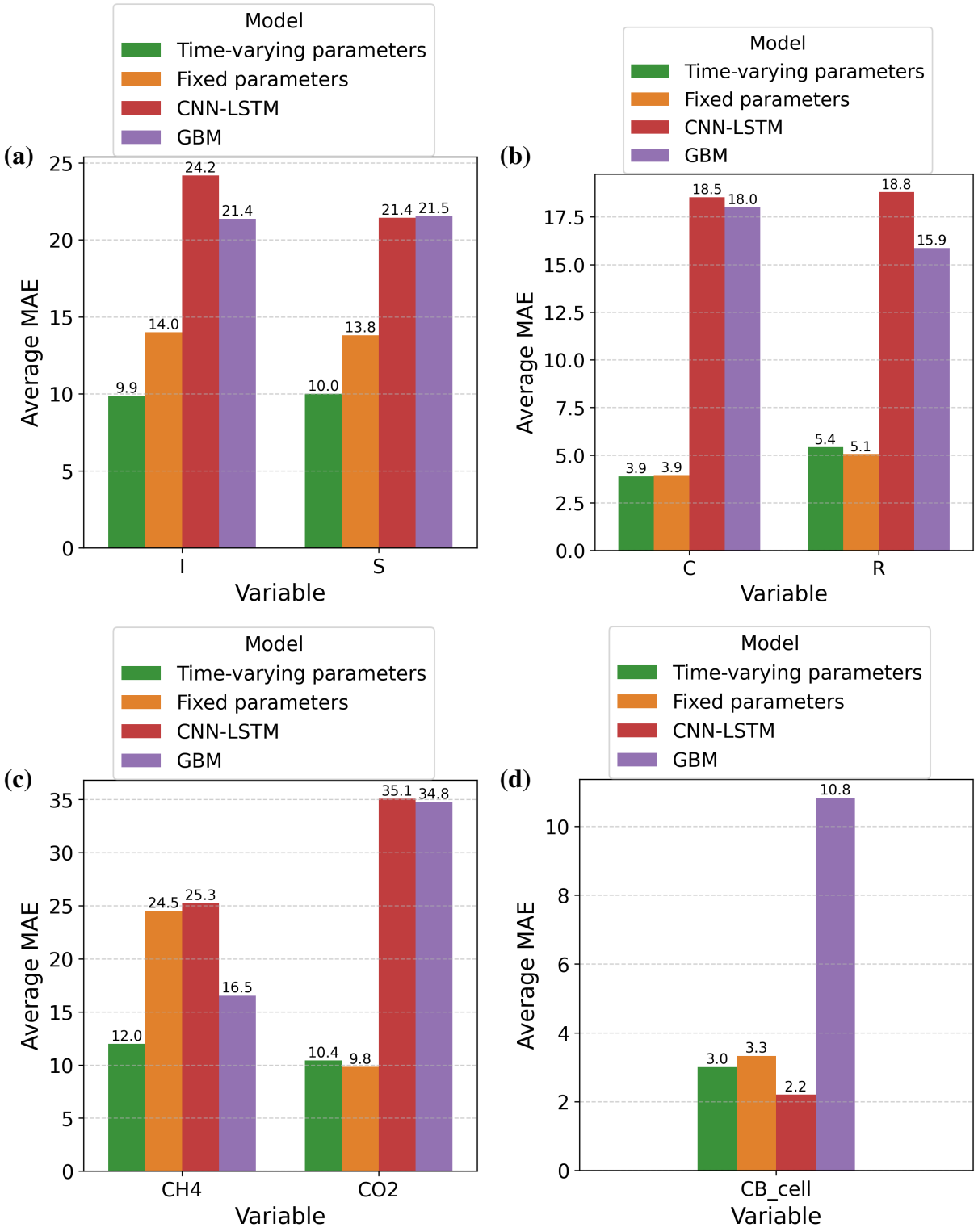


图4：使用扩展窗口交叉验证方法得到的(a) SIR数据集、(b) CR数据集、(c) 气体数据集和(d) CB数据集的平均预测平均绝对误差。绿色条表示时变参数模型，橙色条表示固定参数模型，红色条表示CNN-LSTM模型，紫色条表示GBM模型。平均绝对误差先针对每个数据集内各变量在不同噪声水平或监测站上计算，然后取平均。

相对于扩展窗口交叉验证方法的准确度。

Colle 总体而言，这些结果表明，尽管扩展窗口交叉验证框架 -

该方法提供了稳健的基线；微调区间长度和时变参数的数量，可以显著提高不同动力系统的预测精度。

示例预测图表明，在各案例研究中，采用最优配置的时变参数模型始终能够跟踪所有变量的观测轨迹，从而显著提升预测性能（SI 附录图 11）。相比之下，固定参数模型生成的预测仅能大致匹配 R 和 CH_4，的时间序列，且无法捕捉其余变量的演化动力学（SI 附录图 11）。

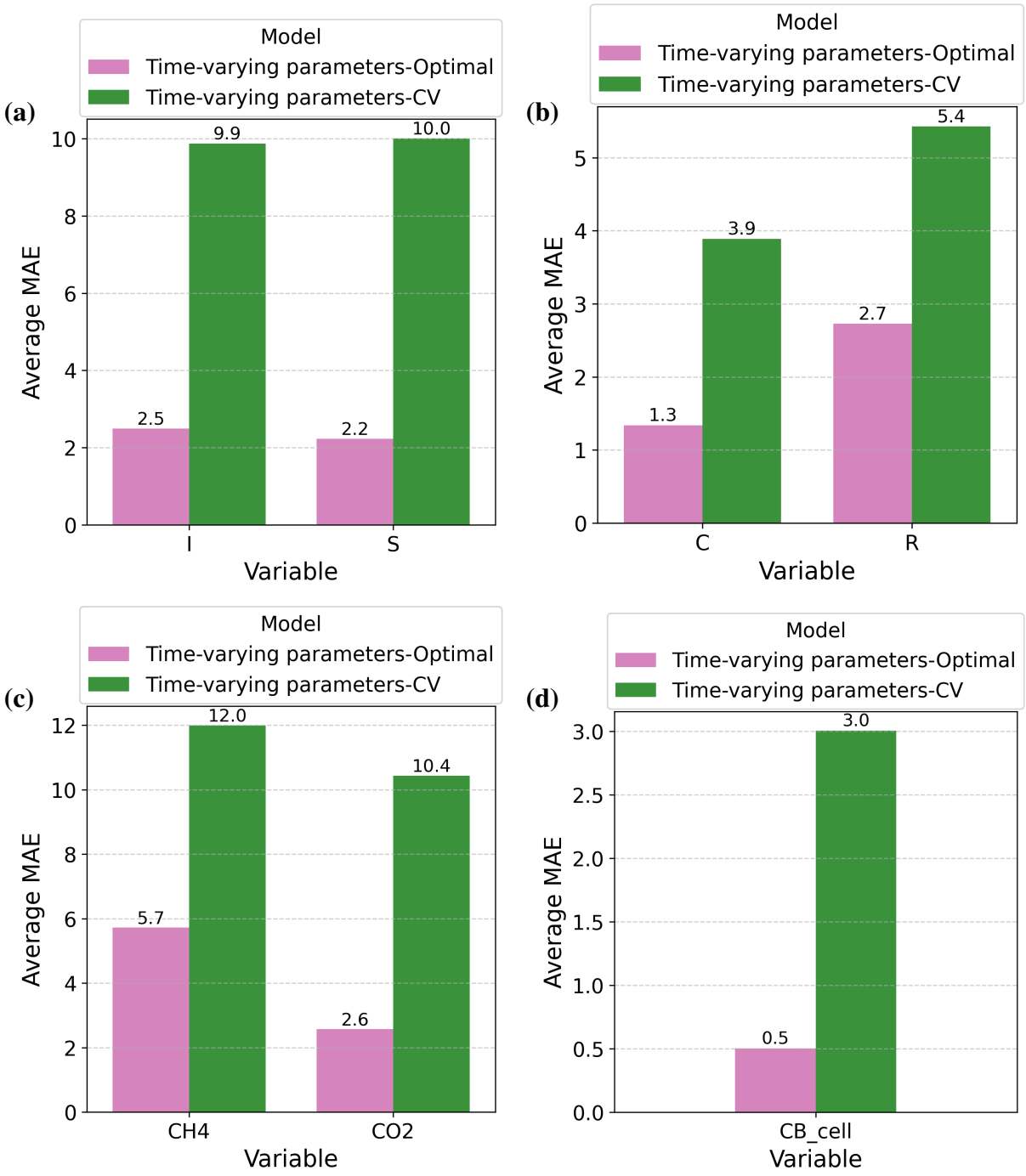


图 5：基于最优配置和交叉验证的时变参数模型的平均预测平均绝对误差，分别对应 (a) SIR数据集、(b) CR数据集、(c) 气体数据集和 (d) CB数据集。绿色柱表示基于交叉验证（CV）方法的模型误差，粉色柱表示模型在最优配置下的预测误差。对于每个数据集，分别计算各变量在不同噪声水平或监测站上的平均绝对误差，并取其平均值。

5 讨论

我们提出了一种数据驱动框架，将稀疏方程发现与机器学习得到的参数演化相结合，从而实现对动力系统的自适应学习与预测。结果表明，允许部分系数随时间变化，既能提升学习能力，也能提高——

高保真度和短期预测准确性，尤其适用于流行病传播和环境气体排放等非平稳过程。与固定系数稀疏回归相比，所提出的模型在大多数数据集上的预测平均绝对误差降低了 2–7%，且训练误差保持在 3% 以下，这表明时序参数化能够在不牺牲可解释性的前提下增强模型的表达能力。

传统机器学习模型广泛应用于流行病学和环境预测[63–65]。虽然这些模型通常能够实现较强的点预测能力，但其黑箱结构限制了机制层面的解释 [14,31]。在我们的实验中，两种常见的数据驱动预测器——LSTM-CNN 和梯度提升机——产生的预测平均绝对误差显著更高，且对噪声更加敏感，这表明当系统机制发生变化或外部驱动因素仅能提供部分信息时，仅依靠这类模型可能难以实现良好的泛化。我们不应将这一结果视为机器学习本身的弱点；相反，它凸显了归纳结构的重要性：将机器学习预测的变化系数嵌入已发现的微分方程中，能够同时提高可解释性和预测稳定性 [8, 9]。我们的框架通过将稀疏方程学习器与随机森林预测的时变参数相结合，使自适应预测成为可能，同时保留动力学的透明性。这种整合对于不断演变的自然系统尤其有价值，包括流感传播、气候驱动的生态相互作用、大气气体排放和蓝藻动力学；在这些系统中，固定系数模型或不受约束的机器学习预测器可能无法捕捉瞬态或非平稳响应 [4,66–68]。

模拟的 SIR 和捕食者-猎物模型系统为内在非平稳与近似平稳的参数状态提供了受控比较。尽管我们的框架准确恢复了两个系统的动力学，但时变系数带来的预测优势在 SIR 系统中显著更为突出。这一结果符合理论预期：由近乎恒定的相互作用强度驱动的捕食者-猎物模型轨迹，在较短的预测时域内可用固定系数合理近似；而 SIR 的演化则很大程度上受非平稳驱动因素支配，其中最关键的是随时间变化的传播率。因此，固定系数稀疏回归基线在 SIR 系统中表现出较大的预测误差，进一步凸显了时间参数化的重要性 [8, 34]。在季节性驱动和外部扰动下的流行病学与生态学模型中，也有研究报告了时变模型的类似优势 [69, 70]。

在实证数据集中，时变模型在学习和预测 CO_2 和 CH_4方面都取得了显著更低的误差。固定系数稀疏模型在预测 CH_4, 时尤其表现不佳，而 LSTM-CNN 和梯度提升机基线在两者上都表现不佳，这反映了季节周期和排放波动对大气气体动力学的影响。由于环境系统会持续变化，要捕捉其行为，就需要采用时变表示 [71]。在 CB数据集中，两种模型在学习阶段的表现相近，这表明某些系统可以用固定参数进行近似；然而，在预测方面，时变模型始终表现更佳。总的来说，这些结果支持一个积极的结论：融入时变参数的方程学习器

相比固定系数稀疏回归或独立的机器学习预测器，能够产生对噪声更稳健且更准确的有限时域预测

在无噪声条件下，两种稀疏模型都取得了较低的预测平均绝对误差，证实当动力学近乎平稳且数据干净时，固定系数表示足以胜任。然而，在噪声和瞬态动力学状态下，模型表现出现了差异。在 SIR 和 CR 学习区间内，时变形式均保持稳定，而固定系数模型在 CR 折中的方差较大，导致学习误差显著升高。有趣的是，在某些 CR 噪声水平下，固定系数模型的学习平均绝对误差较高，但短时域预测误差较低，这是因为其系数收敛到长期平均值，从而平滑了局部波动。尽管这种平均化在较短的预测窗口内看起来颇具竞争力，却会掩盖导数层面的动力学，并限制特定状态下的泛化。相比之下，在较短区间内学习系数能够提升动力学表达能力并降低跨折误差，但如果预测时域与系数演化的时间尺度不匹配，预测准确度仍可能下降。

在我们的框架中，预测准确度取决于时变参数的预测能力；本研究使用单个随机森林模型来估计这些参数。尽管随机森林能够有效捕捉非线性特征—参数关系，但对于动力学不断演化的预测模型，采用具有互补归纳偏置或更强时间适应能力的预测变量或许会带来益处 [72, 73]。神经网络、贝叶斯网络和混合学习器在建模复杂、含噪且瞬态的特征—响应相互作用方面展现出良好前景 [8, 9]。因此，未来研究应探索多基线或基于集成的参数预测器，以提升泛化能力并减少噪声下的传播漂移，尤其是在快速演化或受季节性强迫的系统中 [39, 74]。

尽管扩展窗口交叉验证是超参数选择的标准方法，但它可能无法确定最佳区间长度或最有效的时变参数数量，因为它优化的是平均折性能，而非特定片段的预测能力。逐折评估揭示了一些始终优于交叉验证选定模型的配置——其中最显著的是 SIR 系统：通过优化区间对齐， S 和 I的预测平均绝对误差均降低了约 8%；此外，CR、气体和 CB 数据集中也观察到 2%–5% 的提升。我们还发现，存在噪声时，交叉验证误差会更急剧地上升。这是基于平均的验证已知的局限，可能会掩盖局部最优配置 [75, 76]。这些发现促使未来进一步研究预测感知验证目标、噪声加权模型选择，以及针对瞬态或外部强迫动力学的自适应区间调优。

候选项通过领域知识进行初始化，以确保函数库定义合理且具有可解释性。对于仿真系统（SIR 和捕食者-猎物模型），其控制相互作用已知，因此直接将状态变量用作候选函数。对于实证数据集（气体和蓝细菌，CB），函数库则扩展至纳入测量的源变量和预计会调节动力学的天气协变量。尽管这种设计有助于取得优异的性

能，但通常还可以在不降低预报技巧的情况下进一步缩小库规模。因果选择策略——例如，在贝叶斯网络 [46] 中将预测变量限制为马尔可夫毯中的变量——能够在保持短时程精度的同时大幅降低维度。由于学习和预测演变中的系数所需的成本随活跃项数量增加，缩减函数库也能限制时变参数的数量，从而提高计算效率和稳健性。我们的框架能够轻松适应针对特定数据集的函数库调整，未来还可整合自动化的因果选择或基于相关的项选择，以进一步提升可扩展性和泛化能力。

尽管该框架在学习和预测准确度方面表现出色，分析也发现了若干改进机会。首先，扩展窗口交叉验证虽然是标准做法，但它优先考虑平均折性能，而非特定预测时距的预报技巧，因此可能无法找到区间长度和时变参数数量的局部最优选择。其次，自回归机器学习预报器在较长预测时距下会出现误差累积，尤其是在存在噪声或机制快速转变时；这一挑战可通过动力学约束和噪声感知验证目标来缓解。第三，参数预测依赖外生驱动因子（此处为四个天气协变量）；这些输入的变异有限或存在测量误差，可能降低系数预测的可靠性，因此未来可研究自适应或扩展的驱动因子集。第四，时变参数的选择依据最高相关性（偏置项始终视为时变参数），这可能引入选择偏差；当相关性是短暂的、含有噪声或无法反映真实因果影响时，这种方法也可能失效。最后，候选库根据系统先验知识进行初始化，这确保了可解释性，但可能排除潜在或未观测到的驱动因子；未来的扩展应研究自动化的因果或基于相关性的函数库缩减与选择方法，以提升可扩展性和泛化能力。总之，这些考量凸显了谨慎设计模型以及未来改进方法的重要方向。

数据可用性

本研究使用的所有数据和代码均可在 Zenodo 存储库 [77] 中获取。

资助

Hao Wang的研究部分得到加拿大自然科学与工程研究委员会的资助（个人发现资助RGPIN-2020-03911和发现加速补充奖RGPAS-2020-00090），以及加拿大研究讲席计划的支持（一级加拿大研究讲席奖）。Pouria Ramazi也感谢NSERC发现资助提供的支持。

致谢

我们感谢ILMEE成员提供宝贵的反馈。我们使用ChatGPT [78]对本文的拼写、语法和整体文风进行了润色。计算资源由加拿大数字研究联盟（前身为Compute Canada）提供。

利益冲突

我们声明不存在任何利益冲突。

参考文献

[1] Ruth E Baker、Jose-Maria Pena、Jayaratnam Jayamohan 和 Antoine Jérusalem。机理模型与机器学习：一场值得为生物学界而战的较量？《生物学快报》，14(5):20170660，2018年。

[2] Yunduo Lan、Sung-Young Shin 和 Lan K Nguyen。从浅层到深层：机器学习与机理模型在癌症研究中的整合演进。《系统生物学最新观点》，40:100541，2025年。

[3] Martin-I Trappe 和 Ryan A Chisholm。跨尺度生态学的密度泛函理论。《自然通讯》，14(1):1089，2023年。

[4] Christina C Small、Sunny Cho、Zaher Hashisho 和 Ania C Ulrich。油砂尾矿池的排放：尾矿池参数与排放估算综述。《石油科学与工程杂志》，127:490–501，2015年。

[5] Sivakiruthika Natchimuthu、Balathandayuthabani PanneerSelvam 和 DavidBastviken。气象变量对浅水池塘甲烷和二氧化碳通量的影响。《生物地球化学》，119:403–413，2014年。

[6] Anice C Lowen、Samira Mubareka、John Steel 和 Peter Palese。流感病毒的传播取决于相对湿度和温度。《公共科学图书馆·病原体》，3(10):e151，2007年。

[7] Jeffrey Shaman 和 Melvin Kohn。绝对湿度调节流感病毒存活、传播和季节性。《美国国家科学院院刊》，106(9)：3243–3248，2009。

[8] Amit K Chakraborty、Hao Wang 和 Pouria Ramazi。从政策到预测：在机器学习与疾病模型集成框架中评估预测准确度。《计算生物学杂志》，31(11):1104–1117，2024年。

[9] Xiunan Wang、Hao Wang、Pouria Ramazi、Kyeongah Nah 和 Mark Lewis。从政策到预测：在疫苗接种不完善的情况下预测新冠肺炎动力学。《数学生物学通报》，84(9):90，2022年。[10] Julio R Banga 和 Alejandro F Villaverde。生物系统的机理动力学建模：未来之路。《系统生物学最新观点》，42:100553，2025年。[11] Wenxiang Song、Shijie Jiang、Gustau Camps-Valls、Mathew Williams、Lu Zhang、Markus Reichstein、Harry Vereecken、Leilei He、Xiaolong Hu 和 Liangsheng Shi。迈向地球科学中控制方程的数据驱动发现。《地球与环境通讯》，5(1):589，2024年。[12] Steven L Brunton、Joshua L Proctor 和 J Nathan Kutz。通过非线性动力系统稀疏识别，从数据中发现控制方程。《美国国家科学院院刊》，113(15):3932–3937，2016年。[13] Samuel H Rudy、Steven L Brunton、Joshua L Proctor 和 J Nathan Kutz。偏微分方程的数据驱动发现。《科学进展》，3(4):e1602614，2017年。[14] Shanwu Li、Eurika Kaiser、Shujin Laima、Hui Li、Steven L Brunton 和 J Nathan Kutz。通过非线性动力系统稀疏识别，发现原型桥梁的时变空气动力学特性。《物理评论*E*》，100(2):022220，2019年。[15] Kai Fukami、Takaaki Murata、Kai Zhang 和 Koji Fukagata。利用低维化流动表征进行非线性动力学稀疏识别。《流体力学杂志》，926:A10，2021。[16] Jared L Callaham、Steven L Brunton 和 Jean-Christophe Loiseau。非线性相关性在降阶建模中的作用。《流体力学杂志》，938:A1，2022年。[17] Jared L Callaham、Georgios Rigas、Jean-Christophe Loiseau 和 Steven L Brunton。湍流尾流中对称性破缺的经验平均场模型。《科学进展》，8(19):eabm4786，2022年。[18] Jean-Christophe Loiseau。环形热虹吸管中混沌热对流的数据驱动建模。《理论与计算流体动力学》，34(4):339–365，2020年。[19] Jean-Christophe Loiseau 和 Steven L Brunton。约束稀疏伽辽金回归。《流体力学杂志》，838:42–67，2018年。[20] Bartosz Prokop 和 Lendert Gelens。利用 SINDy 从生物数据构建振荡器模型。《*iScience*》，27(4)，2024年。

[21] Niall M Mangan、Steven L Brunton、Joshua L Proctor 和 J Nathan Kutz。通过非线性动力学稀疏识别推断生物网络。*IEEE* 分子、生物与多尺度通信汇刊，2(1):52–63，2016年。

[22] Eurika Kaiser、J Nathan Kutz 和 Steven L Brunton。在低数据极限下，为模型预测控制进行非线性动力学稀疏识别。英国皇家学会学报*A*辑，474(2219):20180335，2018年。

[23] Xiaojun Wu、MeiLu McDermott 和 Adam L MacLean。面向含噪生物系统的数据驱动模型发现与模型选择。*bioRxiv*，第2024–10页，2024年。

[24] Bartosz Prokop 和 Lendert Gelens。使用 SINDy 对振荡器模型进行数据驱动发现：迈向其在生物学实验数据中的应用。*bioRxiv*，第2023–08页，2023年。

[25] Antoine Sandoz、Verena Ducret、Georg A Gottwald、Gilles Vilmart 和 Karl Perron。将 SINDy 用于时滞微分方程：应用于建立细菌锌响应模型。英国皇家学会学报 *A* 辑，479(2269):20220556，2023年。

[26] Siddharth Prabhu、Srinivas Rangarajan 和 Mayuresh Kothare。利用数据驱动方法发现用于模型预测控制的心血管系统稀疏动力学模型。生物学与医学中的计算机，166:107513，2023年。

[27] Gregor Thiele、Arne Fey、David Sommer 和 Jörg Krüger。使用 SINDy 对滞后控制泵系统进行系统辨识。载于*2020*年 第*24*届国际系统理论、控制与计算会议（*ICSTCC*），第457–464页。IEEE，2020年。

[28] Paul Wulff、Nils Gräbner 和 Utz von Wagner。通过 SINDy 对鼓式制动器尖叫进行最简模型辨识。应用力学档案，94(10):3101–3117，2024年。

[29] Hao Lee、Ruoning Ren、Yifei Qian 和 Jacob Rosen。利用 SINDy 和时变模型预测控制降低可穿戴气动阀系统的能耗。*IEEE/ASME* 机电一体化学报，2024年。

[30] Lin Guo、Xiaokai Yang、Zhonghua Zheng、Nicole Riemer 和 Christopher W Tessum。利用集成 SINDy 对气相大气化学降阶建模中的不确定性进行量化。*arXiv* 预印本 *arXiv:2407.09757*，2024年。

[31] Javier Rubio-Herrero、Carlos Ortiz Marrero 和 Wai-Tong Louis Fan。大气数据建模与动力学识别：空气污染物的时序数据驱动建模。《清洁生产杂志》，333:129863，2022年。

[32] Xiaokai Yang、Lin Guo、Zhonghua Zheng、Nicole Riemer 和 Christopher W Tessum。利用非线性动力学稀疏识别进行大气化学代理建模。地球物理研究杂志：机器学习与计算，1(2):e2024JH000132，2024年。[33] Samuel Rudy、Alessandro Alla、Steven L Brunton 和 J Nathan Kutz。参数化偏微分方程的数据驱动识别。*SIAM*应用动力系统杂志，18(2):643–660，2019年。

[34] Xiunan Wang 和 Hao Wang。离散逆方法：从可获取的感染数据中提取疾病传播率。*SIAM*应用数学杂志，84(3):S336– S361，2023年。

[35] Arthur Bousquet、William H Conrad、Said Omer Sadat、Nelli Vardanyan 和 Youngjoon Hong。利用 SIRD 模型的时变参数进行深度学习预测新冠肺炎。《*Scientific Reports*》，12(1):3030，2022年。[36] Jie Long、AQM Khaliq 和 Khaled M Furati。新冠肺炎模型时变参数的识别与预测：一种数据驱动的深度学习方法。国际计算数学杂志，98(8):1617–1632，2021。[37] Juping Ji、Shohel Ahmed 和 Hao Wang。一种研究和预测美国气候敏感型诺如病毒感染的混合方法。理论生物学杂志，598:112007，2025年。

[38] Pathairat Pastpipatkul与PanichaSubsai。泰国温室气体排放与其决定因素之间的时变系数及预测效应分析。载于最优传输在经济学及相关主题中的应用，第591–604页。施普林格，2024年。

[39] Amit K Chakraborty、Shan Gao、Reza Miry、Pouria Ramazi、Russell Greiner、Mark A Lewis 和 Hao Wang。基于具有不同噪声的随机疾病传播模型训练的早期预警指标。《*Journal of the Royal Society Interface*》，21(217):20240199，2024年。[40] William Ogilvy Kermack 和 Anderson G McKendrick。对流行病数学理论的贡献。《伦敦皇家学会会刊：*A*辑，数学与物理学论文集》，115(772):700–721，1927。[41] Michael L Rosenzweig 和 Robert H MacArthur。捕食者-猎物相互作用的图示及稳定性条件。美国博物学家，97(895):209–223，1963。

[42] Peter E. Kloeden 和 Eckhard Platen。随机微分方程的数值解法，《随机建模与应用概率》第23卷。施普林格，柏林、海德堡，1992年。[43] 网络地图与站点数据——伍德布法罗环境协会——wbea.org。https://wbea.org/data/network-map-station-data/. [访问日期：06-11-2024]。[44] 艾伯塔能源监管局。艾伯塔可开采油砂厂统计数据。https://www.aer.ca/data-and-performance-reports/statistical-reports/ st39. [访问日期：01-23-2025]。[45] Christopher M Heggerud。cheggerud/cb-prediction：Nov_10_2023。https://doi. org/10.5281/zenodo.10109225，2023年。 [访问日期：24-01-2024]。[46] Christopher M Heggerud、Jingjing Xu、Hao Wang、Mark A Lewis、Ron W Zurawell、Charlie JG Loewen、Rolf D Vinebrooke 和 Pouria Ramazi。利用不完整的及时数据预测湖泊中即将发生的蓝藻水华。《水资源研究》，60(2):e2023WR035540，2024年。[47] George E. Uhlenbeck 和 Leonard S. Ornstein。关于布朗运动理论。《物理评论》，36(5):823–841，1930年。[48] Leo Breiman。随机森林。机器学习，45:5–32，2001年。[49] Fabian Pedregosa 等。Scikit-learn：用于 Python 的机器学习。《*Journal of Machine Learning Research*》，12:2825–2830，2011年。[50] Xingjian Shi 等。卷积长短期记忆网络：一种用于降水临近预报的机器学习方法。《神经信息处理系统进展》，28，2015年。[51] Sepp Hochreiter 和 Jürgen Schmidhuber。长短期记忆。《神经计算》，9:1735–1780，1997年。[52] Shaojie Bai、J Zico Kolter 和 Vladlen Koltun。用于序列建模的通用卷积网络和循环网络的实证评估。*arXiv*预印本 *arXiv:1803.01271*，2018年。[53] Jerome H Friedman。贪心函数逼近：梯度提升机。《统计年鉴》，29:1189–1232，2001年。[54] Martín Abadi 等。TensorFlow：一种用于大规模机器学习的系统。第*12*届*USENIX*操作系统设计与实现研讨会论文集，第265–283页，2016年。

[55] François Chollet。Keras。*GitHub* 仓库，2015年。https://keras.io.[56] Tom O’Malley 等。Keras 调参器，2019年。https://keras.io/keras_tuner.[57] Diederik P Kingma 和 Jimmy Ba。Adam：一种随机优化方法。*arXiv* 预印本 *arXiv:1412.6980*，2014年。[58] Lutz Prechelt。提前停止——但何时停止？《神经网络：行业诀窍》，第55–69页，1998年。[59] Gerald Teschl。常微分方程与动力系统。美国数学学会，2012年。[60] Zeyu Zhang 和 Hayden Schaeffer。论 SINDy 算法的收敛性。《多尺度建模与仿真》，17(3):948–972，2019年。[61] Thomas H. Gronwall。关于微分方程组解对参数的导数的注记。《数学年刊》，20(4):292–296，1919年。[62] Walter Rudin。实分析与复分析。麦格劳-希尔，第3版，1987年。[63] Abderrachid Hamrani、Abdolhamid Akbarzadeh 和 Chandra A Madramootoo。用于预测农业土壤温室气体排放的机器学习。《整体环境科学》，741:140338，2020年。[64] Pi Guo、Tao Liu、Qin Zhang、Li Wang、Jianpeng Xiao、Qingying Zhang、Ganfeng Luo、Zhihao Li、Jianfeng He、Yonghui Zhang 等。利用机器学习构建登革热预测模型：中国案例研究。《*PLOS* 被忽视的热带病》，11(10):e0005973，2017年。[65] Esha Saha、Oscar Wang、Amit K Chakraborty、Pablo Venegas Garcia、Russell Milne 和 Hao Wang。用于阿尔伯塔尾矿池甲烷监测的基于弥散的循环神经网络模型。《环境管理杂志》，395:127748，2025年。[66] Anice C Lowen 和 John Steel。湿度和温度在塑造流感季节性中的作用。《病毒学杂志》，88(14):7692–7695，2014年。[67] Angela N Laws。气候变化对捕食者—猎物相互作用的影响。《昆虫科学前沿》，23:28–34，2017年。[68] Shih-Hsien Tseng、Pei Han Feng 和 Thi Ha Trang Duong。应用数据科学与特征选择技术预测台湾的二氧化碳排放。《随机环境研究与风险评估》，第1–23页，2025年。

[69] Adam Spannaus、Theodore Papamarkou、Samantha Erwin 和 J Blair Christian。推断新冠肺炎的传播：时变报告率在流行病学建模中的作用。《*Scientific Reports*》，12(1):10761，2022年。[70] Xiaodong Song、Brett A Bryan、Auro C Almeida、Keryn I Paul、Gang Zhao 和 Yin Ren。基于过程的生态模型的时变敏感性。《*Ecological Modelling*》，265:114–123，2013年。[71] Melkior Ornik 和 Ufuk Topcu。使用最大似然估计对时变马尔可夫决策过程进行学习与规划。《*Journal of Machine Learning Research*》，22(35):1–40，2021。[72] Nesreen K Ahmed、Amir F Atiya、Neamat El Gayar 和 Hisham El-Shishiny。机器学习模型用于时间序列预测的实证比较。《*Econometric Reviews*》，29(5-6):594–621，2010年。

[73] Bushra Majeed Muter 和 Ayat Jasim Mohammed。人工智能的未来：评估深度学习和机器学习的优势与局限。《*Wisdom Journal For Studies & Research*》，4(06):348–374，2024年。

[74] Amit K Chakraborty、Reza Miry、Russell Greiner、Mark A Lewis、Hao Wang、Tianyu Guan 和 Pouria Ramazi。用于疾病暴发预测的深度学习：并行 LSTM-CNN 模型。《*Journal of the Royal Society Interface*》，22(229):20250046，2025年。[75] Stephen Bates、Trevor Hastie 和 Robert Tibshirani。交叉验证：它估计什么，又估计得有多好？《*Journal of the American Statistical Association*》，119(546):1434–1445，2024年。[76] Christoph Bergmeir、Rob J Hyndman 和 Bonsoo Koo。关于交叉验证在评估自回归时间序列预测时有效性的说明。《*Computational Statistics & Data Analysis*》，120:70–83，2018年。[77] Amit K. Chakraborty。数据与代码：运用机器学习将机理模型转化为预测器，2026年2月。[78] OpenAI。ChatGPT，2023年。用于校对辅助的大语言模型。

补充信息

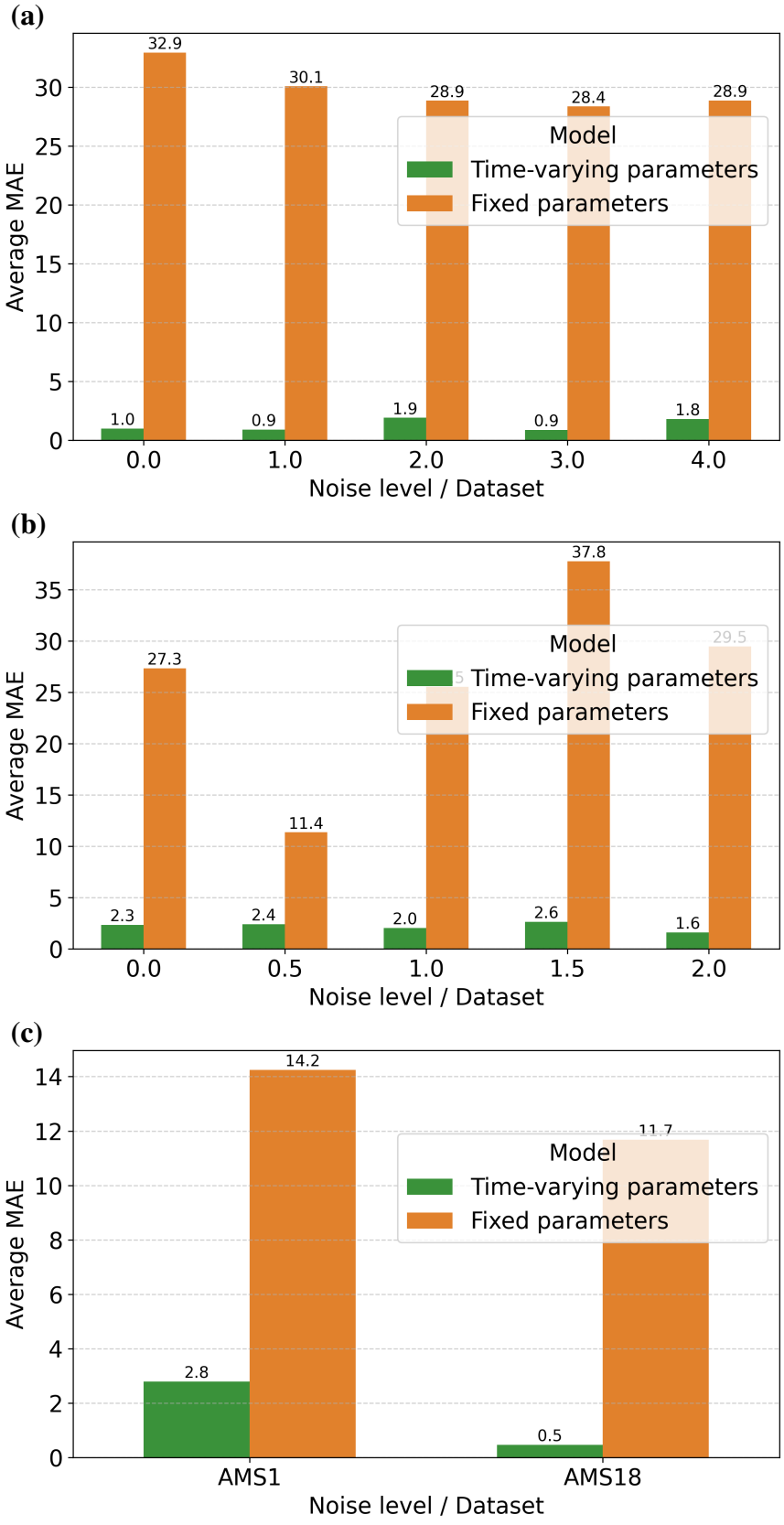


图6: （a）SIR数据集中不同噪声下的平均学习MAE，（b）CR数据集中不同噪声下的平均学习MAE，以及（c）气体数据集中不同噪声下的平均学习MAE

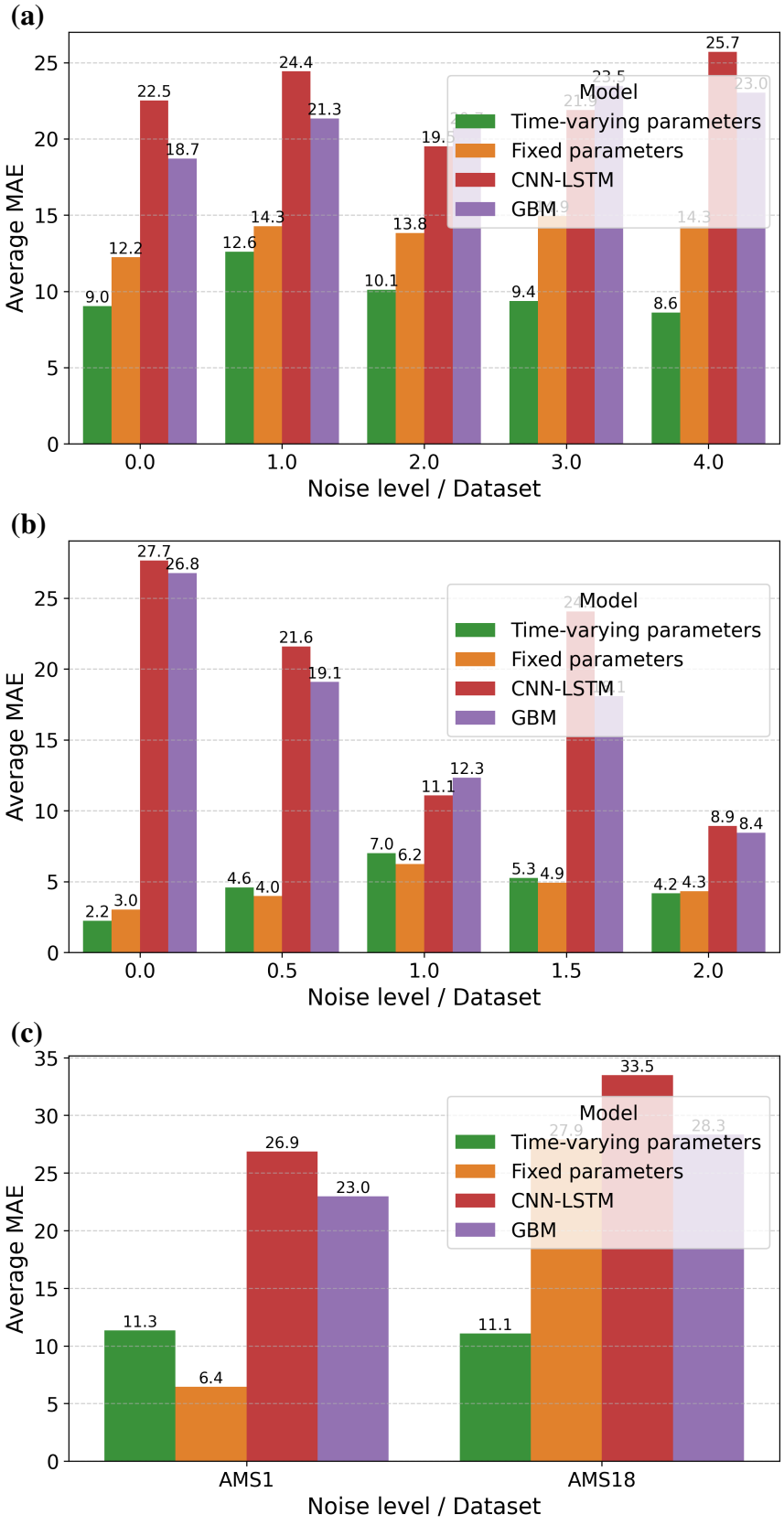


图7：A　不同噪声条件下的平均预测平均绝对误差：(a) SIR数据集，(b) CR数据集，以及(c)气体数据集　t.

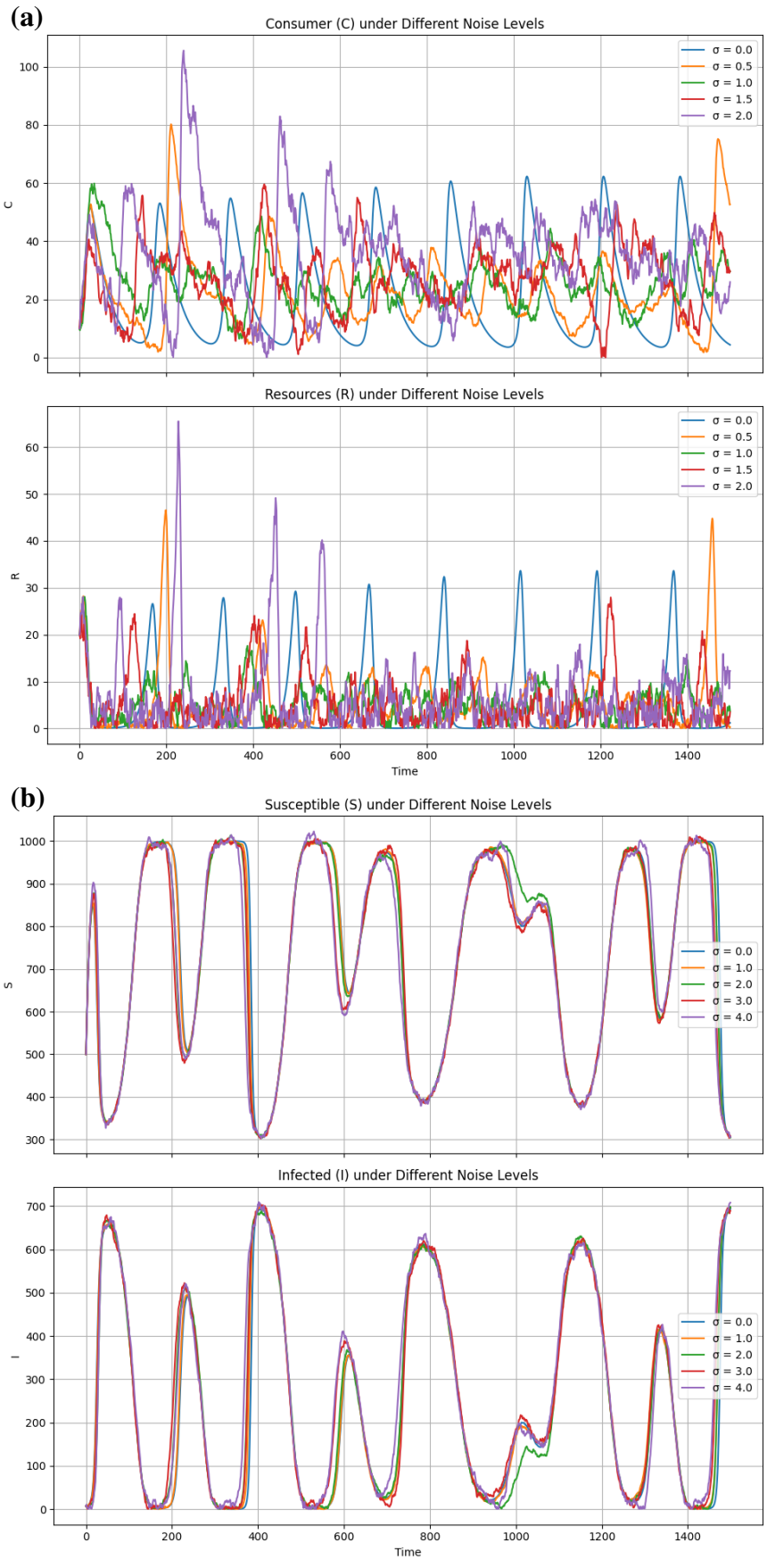


图8：a）捕食者-猎物模型和 b）SIR模型的模拟。

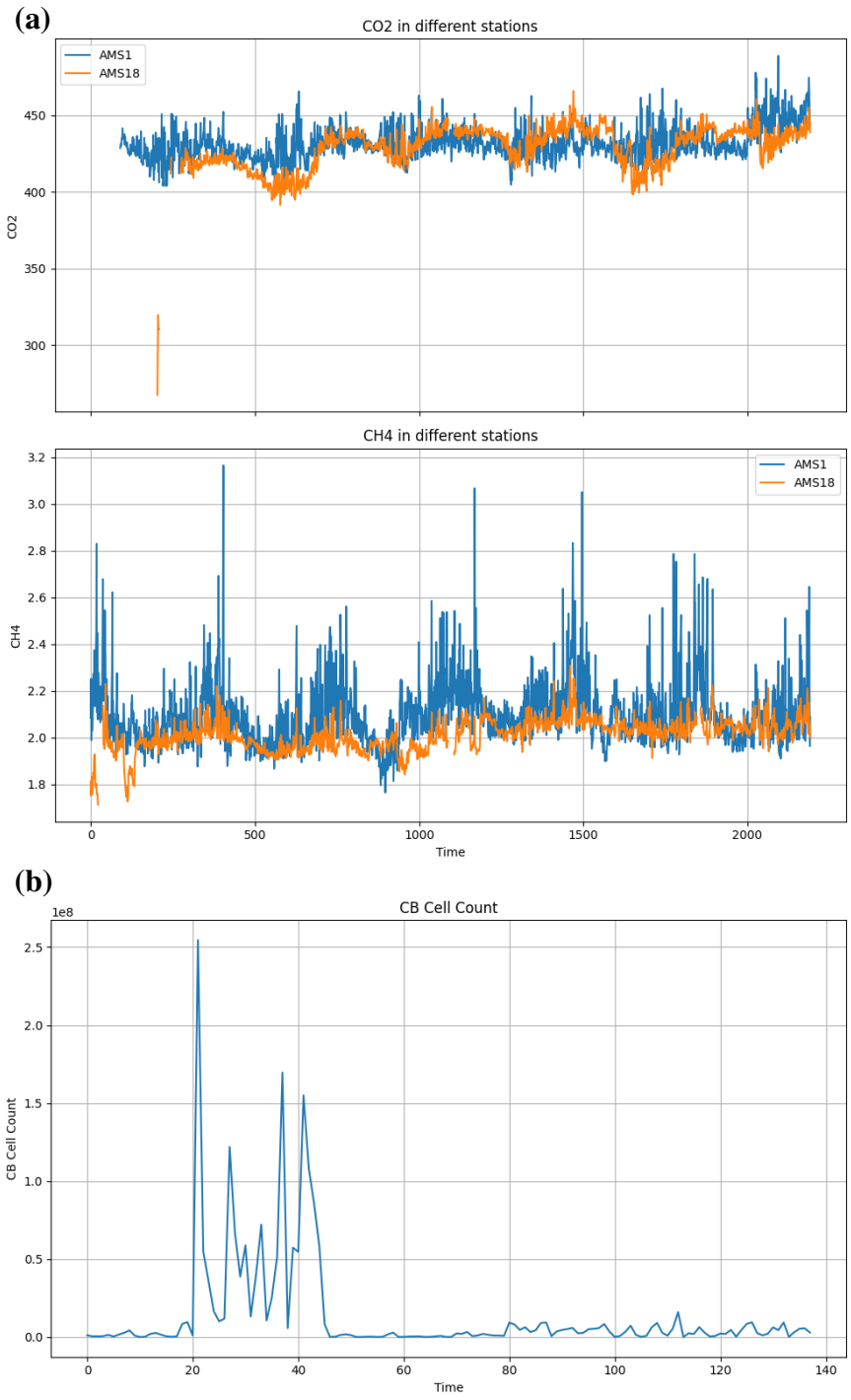


图9：a) 气体数据集和 b) CB数据集的时间序列。

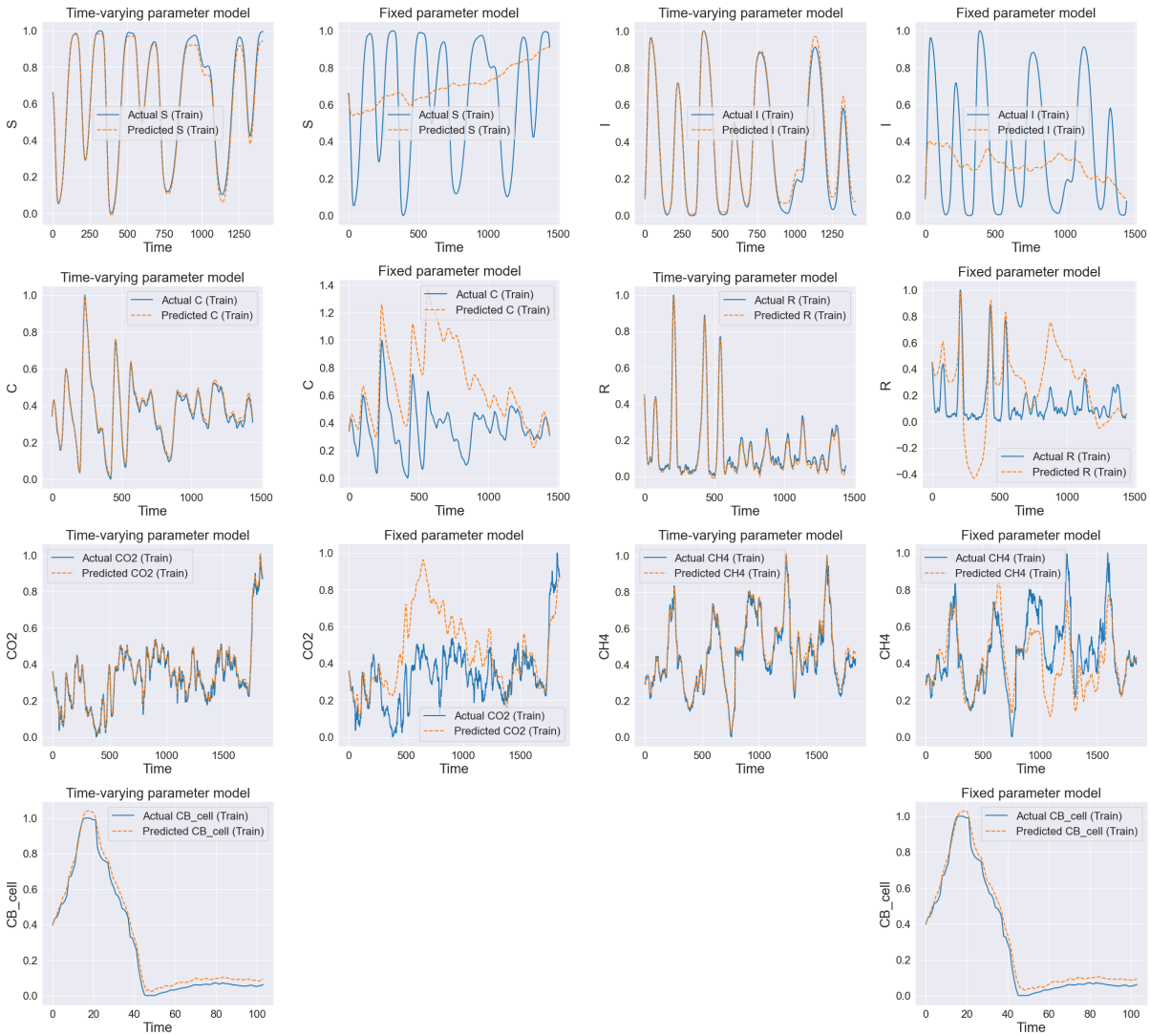


图10：时变参数模型与固定参数模型在四个数据集上的学习对比。第一列和第三列展示时变参数模型的学习结果，第二列和第四列展示固定参数模型的学习结果。

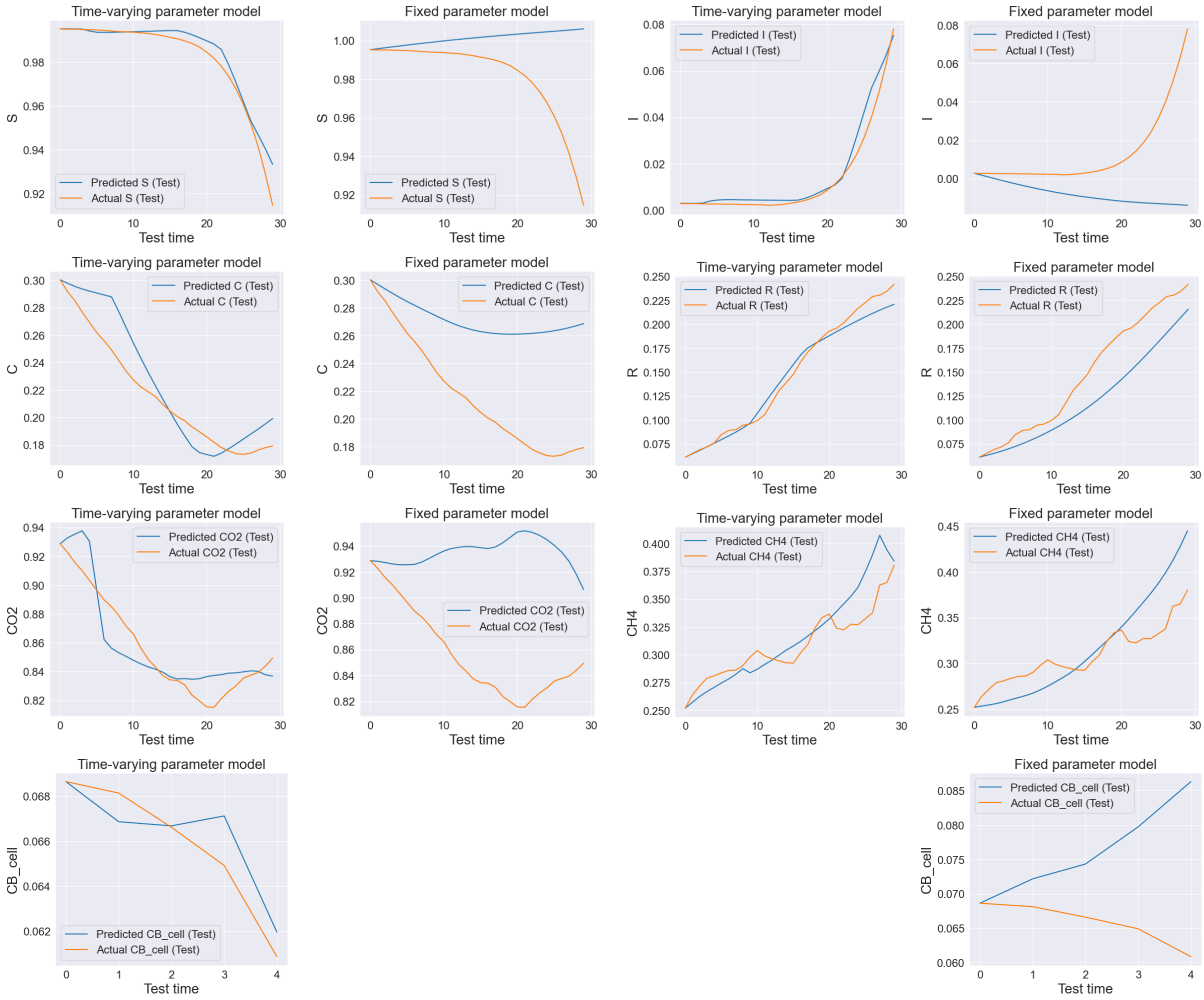


图11：时变参数模型与固定参数模型在四个数据集上的预测比较。第一列和第三列是时变参数模型做出的预测结果，而第二列和第四列是固定参数模型做出的预测结果。

表3：使用扩展窗口交叉验证（CV）方法时，CR模型模拟中每种噪声设置的最优配置（OC）在不同测试折上的时变参数数量、区间长度和平均绝对误差。

数据集	噪声	变量	Fold	CV: 区间长度	CV: Num. 时间的变化 时变参数	CV: MAE test	OC: 区间长度	OC: Num. 时间的变化的参数。	OC: MAE test
CR	0.0	C	1	28	1	1	21	2	0.29
			2	28	1	2.02	14	2	0.59
			3	28	1	0.93	7	5	0.46
			4	28	1	0.1	7	1	0.07
			5	28	1	0.21	21	1	0.08
		R	1	28	1	5.22	28	2	1.72
			2	28	1	3.79	21	2	2.16
			3	28	1	6	14	4	1.07
			4	28	1	1.56	7	1	0.16
			5	28	1	1.62	14	1	0.16
	0.5	C	1	21	1	2.75	28	1	2.42
			2	21	1	5.54	14	2	3.86
			3	21	1	1.39	21	6	0.72
			4	21	1	2.97	28	1	2.77
			5	14	1	4.06	28	2	2.85
		R	1	21	1	1.55	28	1	1.45
			2	21	1	5.53	14	4	4.01
			3	21	1	1.85	28	4	0.72
			4	21	1	15.28	7	2	8.01
			5	14	1	4.91	7	2	3.63
	1.0	C	1	28	2	3.1	28	6	1.55
			2	28	2	4.48	21	5	2.9
			3	28	2	2.38	21	2	1.54
			4	28	2	1.06	14	6	0.55
			5	28	2	19.81	14	4	0.95
		R	1	28	2	2.77	28	1	2.69
			2	28	2	16.36	21	6	10.66
			3	28	2	2.3	28	2	2.3
			4	28	2	3.5	28	6	1.23
			5	28	2	14.45	21	4	3.06
	1.5	C	1	14	1	3.94	7	5	1.18
			2	14	1	3.46	21	3	1.04
			3	14	1	9.69	7	6	1.45
			4	14	1	6.17	28	2	2.69
			5	28	1	5.04	14	3	1.08
		R	1	14	1	2.84	21	4	1.54
			2	14	1	4.03	14	6	1.62
			3	14	1	5.8	21	6	5.39
			4	14	1	8.27	14	6	3.62
			5	28	1	3.35	28	3	1.21
	2.0	C	1	14	2	1.34	21	1	0.94
			2	14	2	4.72	7	5	0.62
			3	14	2	3.87	28	2	0.59
			4	14	2	1.42	7	1	0.83
			5	14	2	5.71	14	3	1.45
		R	1	14	2	6.57	21	6	2.97
			2	14	2	2.86	14	5	1.18
			3	14	2	10.22	14	3	6.43
			4	14	2	1.07	28	5	0.43
			5	14	2	4	14	3	0.74

表 4：使用扩展窗口交叉验证（CV）方法，针对 SIR 模型模拟中每种噪声设置的最优配置（OC），不同测试折的时变参数数量、区间长度和平均绝对误差（MAE）。

数据集	噪声	变量	Fold	CV: 区间长度	CV: Num. 随时间变化时变参数。	CV: MAE test	OC: 区间长度	OC: Num. 时间变化的参数。	OC: MAE test
SIR	0.0	I	1	14	6	7.81	28	5	3.42
			2	14	6	16.78	28	6	1.67
			3	28	6	11.43	28	1	1.78
			4	28	6	0.66	28	6	0.66
			5	28	6	6.49	28	6	6.49
		S	1	14	6	8.66	28	6	3.61
			2	14	6	17.54	28	6	1.26
			3	28	6	11	28	1	1.8
			4	28	6	1.41	7	1	0.74
			5	28	6	8.47	21	4	4.03
	1.0	I	1	21	6	7.83	28	5	3.2
			2	21	6	10.75	28	6	1.57
			3	28	6	13.98	14	6	2.24
			4	21	6	25.82	21	1	0.48
			5	28	6	5.08	14	5	1.08
		S	1	21	6	8.08	14	2	1.75
			2	21	6	11.88	28	6	0.73
			3	28	6	13.2	28	1	2.94
			4	21	6	25.76	21	1	0.47
			5	28	6	3.69	21	6	3.62
	2.0	I	1	21	6	11.43	14	3	4.16
			2	14	3	17.39	28	5	4.06
			3	14	3	9.05	28	1	1.28
			4	14	3	2.45	7	6	0.26
			5	14	3	9.33	21	5	1.42
		S	1	21	6	10.75	7	2	3.16
			2	14	3	13.36	21	5	3.11
			3	14	3	2.02	28	2	0.79
			4	14	3	14.36	7	6	0.31
			5	14	3	10.89	21	6	4.29
	3.0	I	1	21	6	12.46	14	4	2.87
			2	21	6	12.66	28	4	1.13
			3	28	6	7.75	28	1	1.31
			4	28	6	2.62	28	5	1.13
			5	28	6	8.35	7	6	4.47
		S	1	21	6	10.3	28	4	1.74
			2	21	6	12.48	14	4	0.57
			3	28	6	8	28	1	1.88
			4	28	6	2.48	14	6	0.7
			5	28	6	16.5	28	4	5.17
	4.0	I	1	14	1	10.01	14	3	3.05
			2	21	6	9.14	28	5	8.06
			3	21	6	6.09	28	1	2.71
			4	21	6	6.67	28	2	1.32
			5	21	6	14.84	14	6	2.46
		S	1	14	1	10.13	21	2	2.48
			2	21	6	5.88	21	3	1.19
			3	21	6	3.8	21	5	3.07
			4	21	6	6.15	21	5	2.38
			5	21	6	13.35	14	5	3.97

表 5：使用扩展窗口交叉验证（CV）方法和气体数据集中各站点的最优配置（OC），在不同测试折上的时变参数数量、区间长度和平均绝对误差。

数据集	站点	变量	Fold	CV: 区间长度	CV: Num. 时间变化时变参数	CV: MAE test	OC: 区间长度	OC: Num. 时间变化的参数。	OC: MAE test
气体	AMS1	CH_4	1	7	3	5.68	21	4	1.38
			2	7	3	37	14	3	1.52
			3	14	3	5.07	14	15	1.13
			4	14	3	10.58	7	36	1.02
			5	14	3	4.64	14	81	1.27
		CO_2	1	7	3	12.56	7	28	3.02
			2	7	3	10.91	14	8	2.61
			3	14	3	7.99	7	3	2.48
			4	14	3	6.47	14	5	2.89
			5	14	3	12.54	21	23	2.46
	AMS18	CH_4	1	7	96	24.6	7	81	23.78
			2	7	99	4.94	7	5	4.51
			3	7	99	15.28	14	95	14.08
			4	7	99	6.56	28	2	4.61
			5	7	99	5.57	7	75	4
		CO_2	1	7	96	11.16	7	3	2.65
			2	7	99	10.9	7	7	1.63
			3	7	99	5.75	14	12	3.43
			4	7	99	12.2	7	6	3.05
			5	7	99	13.86	7	40	1.47

表 6：使用扩展窗口交叉验证（CV）方法和每个站点的最优配置（OC），蓝细菌数据集中不同测试折的时变参数数量、区间长度和平均绝对误差。

数据集	区域	变量	Fold	CV: 区间长度	CV: Num. 时间变化参数。	CV: MAE test	OC: 区间长度	OC: Num. 的时间变化参数。	OC: MAE test
CB	AB	CB_{cell}	1	20	103	2.26	15	82	0.07
			2	15	55	0.91	25	65	0.29
			3	25	64	5.46	25	106	1.74
			4	25	64	3.5	15	58	0.26
			5	25	60	2.88	10	69	0.15