时间序列数据上的可解释系统辨识与长期预测

朱霞，　　DuxinChen*、韦文佳、朱霞和Wenwu Yu*，*IEEE*高级会员

摘要—过去几十年来，随着深度学习方法不断取得突破，时间序列预测受到了广泛关注。然而，大多数基于神经网络的方法缺乏可解释性，无法提取目标物理系统的隐藏机制。为克服这些缺陷，本文提出了一种无需任何先验知识的可解释稀疏系统辨识方法。该方法采用傅里叶变换减少字典矩阵中的无关项，而不是像大多数系统辨识方法那样不加区分地使用多项式函数。该方法能够实现可解释的系统表示，并显著降低计算成本。通过采用 l_1 范数对参数矩阵进行正则化，可以获得系统模型的稀疏描述。此外，本文采用包括水利数据、全球气温数据和金融数据在内的三个数据集，对所提方法的性能进行测试。尽管没有关于物理背景的先验知识，实验结果表明，与广泛使用的基线数据驱动方法相比，无论原始数据中存在噪声还是不完整性，我们的方法都能更准确地实现长期预测。本研究可为时间序列预测研究提供一些启示，并表明白盒系统辨识方法能够提取那些容易被忽视但内在存在的周期性特征，并在长期预测任务中击败基于神经网络的黑盒方法。

索引词—　　时间序列预测、长期预测、稀疏辨识、基于模型的方法、机器学习

arXiv:2303.01193v1 [cs.LG] 2 Mar 2023

1 引言

ACCURATE 时间序列预测在大数据时代至关重要，广泛应用于现代工业的诸多领域，例如能源消耗、天气变化、疾病传播、交通流量和经济演变预测 [1]–[5]。它吸引了数学、计算机科学及其他相关工程领域众多科学家的广泛关注，并取得了大量成果，展现出令人瞩目的性能。然而，由于人们尚未深入理解目标物理系统中的隐藏机制，实现长期（例如超过一千步）且稳定的预测仍是一项棘手但意义重大的挑战 [6]。此外，现有方法大多基于深度学习框架提出，难以满足大规模工业系统对机制可解释性和可靠性的要求。这促使我们进一步以白盒形式探究系统演化机制，并摆脱对黑盒神经网络模块的依赖。

针对时间序列预测，近年来为解决这一问题逐渐形成了两种主流方法。一种源于数据驱动方法，例如神经网络（NN）和统计学习。这类方法无需显式的机理模型，即可实现端到端预测。早期的数据驱动方法包括反向传播网络（BP 网络） [7] 和自回归差分移动平均模型（ARIMA） [8], ，它们在具有简单规则的小规模系统预测中表现良好。然而，在大多数具有复杂非线性和随机性的系统中，这两种模型都无法保证长期预测的准确率。为了提升长期预测能力，人们提出了各种循环神经网络（RNN）。应用广泛的两种 RNN 是长短期记忆网络（LSTM） [9] 和门控循环单元（GRU） [10],，它们在处理和预测时间序列中具有长期交互的重要事件方面表现出优势 [11], [12]。但 RNN 及其变体模型的参数远多于典型的神经网络，因此模型训练时间更长，且过拟合风险更高。为解决这一问题，随后提出了时间卷积网络（TCN） [3]。TCN 使用因果卷积对时间因果关系进行建模，并允许并行计算输出，因此与 RNN 相比能够提升网络性能。遗憾的是，使用 TCN 进行长期预测需要多次迭代进行多步（或单步）预测，这会导致误差累积，从而削弱其捕捉长期机制的能力 [13]。近期，随着仅基于自注意力机制的 Transformer [14]问世，许多 Transformer 模型的变体 [1], [4], [15] 相继被用于长期预测，无需多次迭代。其中，Autoformer [1] 在准确率和计算速度方面均取得了最先进的结果。然而，Autoformer 需要充足的数据进行训练，而且在保证准确率的前提下，预测长度最多只能达到数百个时间点。尽管这些数据驱动方法在特定预测任务中表现出色，并已广泛应用于实际系统，但在提取物理机制方面仍存在可解释性和稳定性问题。这是因为它们均采用黑箱神经网络框架，缺乏推断系统动力学的能力。

* 通讯作者 *Xiaoyi Liu*、*Duxin Chen* 和 *Wenwu Yu* 就职于江苏省网络化群体智能重点实验室、东南大学数学学院，中国南京 *210096*。（电子邮箱：*xiaoyiliu@seu.edu.cn*、*chendx@seu.edu.cn*、*wwyu@seu.edu.cn*）

另一种主流途径基于系统辨识。考虑到大多数工程系统在真实场景中都遵循潜在的物理机制 [16]–[19]，，原问题可以转化为时间序列的稀疏辨识。正交匹配追踪算法（OMP） [20] 是最早用于解决这一问题的算法之一。该方法采用贪心思想，能够快速获得稀疏模型，但在大多数情况下无法得到最优解。为改进这一方法，人们分别提出了 LASSO [21] 和稀疏贝叶斯方法 [22]，它们能够在先验知识和假设的基础上得到最优解。然而，这类先验假设在现实中可能难以满足，而且在大多数目标系统中，先验知识十分有限，甚至无法获得。近期，基于在含噪复杂系统中发现支配性势方程的非线性动力学稀疏辨识（SINDy） [23]和混合动力系统辨识（HYDE） [24]在自动驾驶车辆、流体对流、医疗应用等真实场景中表现良好。该领域很快涌现出许多相关模型 [25]–[28]。然而，当前稀疏辨识方法不可避免的瓶颈在于，难以兼顾基函数的数量与计算复杂度。过度关注前者会导致计算时间难以承受 [29]，，甚至比训练大规模神经网络耗时更长。相反，过度关注后者则会导致辨识结果不准确 [23]。这种两难困境阻碍了稀疏辨识方法对时间序列进行长期预测。

因此，为了提高时间序列预测的准确率和计算速度，尤其是在长期预测任务中，本研究提出了一种采用自适应基函数的新型稀疏系统识别方法（SIABF），该方法包含三个实施步骤。第一步，给定先验基函数库并评估选择准则。考虑到绝大多数时间序列可能具有周期性 [31]，，首先使用离散傅里叶变换（DFT） [32] 将时域转换为频域。根据 DFT 频谱的振幅，按频率秩降序排序，得到能够反映内在潜在周期性强度的重排序图。利用时频转换公式 [33]，在重排序图中排名靠前的频率对应的先验周期将用于构建先验傅里叶基函数。第二步，在进行预测之前，提供一个用于评估先验基质量的评价指标。我们将准周期指数 $I^{(10)}$ 定义为排序图中前十个最快下降斜率与最大振幅之比，该指数反映了整个时间序列的周期复杂度。$I^{(10)}$ 的值可为先验基函数选择的适用性提供有效参考。第三步，利用稀疏辨识进行长期预测。假设先验基函数主要由一系列正弦和余弦函数构成，其频率为第一步评估得到的排名靠前的频率。通过采用 l_1 范数对参数矩阵 [21], [22], 进行正则化，即可实现时间序列稀疏辨识。利用识别结果可对时间序列进行长期预测。

为进一步说明现有基于模型的方法和数据驱动方法的瓶颈，我们开展了一项对照实验。构造了以下两类时间序列，并分别使用基于模型的方法和数据驱动方法对其进行预测。稀疏辨识方法无法提取图1.(a)中的机制，因为生成规则并未包含在基函数中；而门控循环单元则能很好地进行预测，这是由于时间序列 [30]的生成规则较为简单。相比之下，在拟合图1.(b)中的不规则曲线时，神经网络只能进行短期准确预测。但由于所假设的

基函数包含部分生成规则时，SINDy 可获得更好的预测结果 [23]。

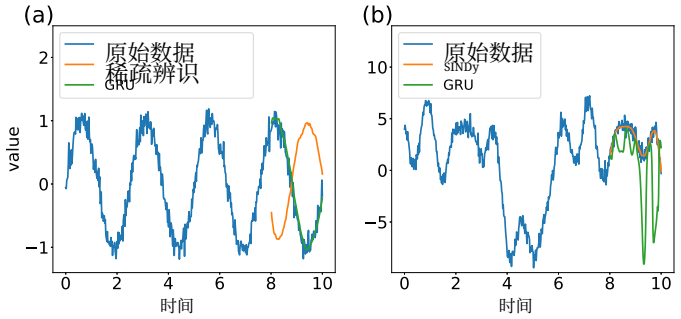


图1。两种预测方法的示例。(a) 生成规则：$x = \quad + \epsilon$ $x = 5 \quad t + \sin(0.5t)$ $3 \quad 2t - 2.3 \quad 5t + 1.2 \quad 7t + \epsilon$；(b) 生成规则：$\sin() \cos() \sin() \cos()$。识别函数库假设基函数为 $\sin(t)$、$\cos(t)$、$\sin(2t)$、$\cos(2t)$、.....、$\sin(100t)$、$\cos(100t)$。在这两种情况下，训练时长和预测时长分别为 8 秒和 2 秒。

与现有的基于模型的方法和数据驱动方法相比，SIABF 具有以下优势。通过离散傅里叶变换获得的先验周期不仅提供了基函数库中最有可能的候选项，还赋予整个模型频域中的可解释性，这是基于神经网络的方法无法提供的 [34]。与其他基于模型的方法相比，SIABF 大幅降低了基函数矩阵的维度 [23], [24], [28]。此外，与大多数广泛使用的知名数据驱动方法相比，它能够利用隐藏于任何物理系统中的底层主导频率，实现长期预测 [1], [10], [11]。

本文的主要贡献如下：

1. 本文提出了一种新颖的可解释稀疏系统辨识方法，通过变换到频域来捕捉关键周期性因素。所获得并用于基函数的底层频率特征，清晰地描述了物理系统的演化机制。2. 在时间序列长期预测任务中，所提出的 SIABF 模型以更少的计算量和训练时间，实现了高精度且稳定的预测结果。3. 通过在温度、水利和金融数据集上进行测试，并与其他广受认可的最新基于模型和数据驱动的基线方法进行全面比较，取得了最先进的结果。明显的优势已经

在对比实验中观察到。

本文其余部分安排如下。第 2 节介绍时间序列预测中稀疏识别的问题及预备知识。第 3 节详细介绍所提出的 SIABF 模型。第 4 节通过数值实验展示与其他基线方法的比较。第 5 节总结全文。

2 问题与预备知识

本节介绍稀疏识别问题的描述，并给出一些预备知识，以便更清晰地阐述后续的方法论。

2.1 问题描述

传统的时间序列预测稀疏识别问题可描述如下 [23]:

$$\mathbf{X}_k = \boldsymbol{\Theta}(\mathbf{t}, \widetilde{\mathbf{X}})\boldsymbol{\Xi}_k + \mathcal{N}, k = 1, 2, ..., m, \tag{1}$$

其中，$\mathbf{X}_k \in \mathbf{R}^n$表示具有 n 个预测步长的系统第 k个 m 输出的状态。 $\boldsymbol{\Theta}$ 是一个 p维向量值函数，表示稀疏辨识中的基函数库。$\mathbf{t} \in \mathbf{R}^n$ 表示预测时间，$\widetilde{\mathbf{X}} \in \mathbf{R}^{(m\times l)\times n}$ 表示 l阶动态方程中预测数据之前所有系统的第1、2、.....、 l步数据。特别地，对于0阶动态方程（即代数方程），有$\mathbf{X} \equiv 0$。 $\mathcal{N}$ 是模型的容许误差项，且 $\mathcal{N} \sim N\left(0, \sigma^2\mathbf{I}\right)$。 $\boldsymbol{\Xi}_k$ 是稀疏辨识的核心——稀疏系数矩阵。目标是获得模型中的稀疏矩阵 $\boldsymbol{\Xi}_k$，即,

$$\min \left\|\boldsymbol{\Xi}_k\right\|_0 = \sum_{i=1}^{(ml+1)p} \left(\xi_i^0\right)_k, k = 1, 2, ..., m. \tag{2}$$

这里， $0^0 = 0$。利用稀疏辨识进行时间序列预测的基本思想是：使用样本数据中的$\mathbf{X}_k$,$\mathbf{X}$ $\boldsymbol{\Theta}$ 训练 $\boldsymbol{\Xi}_k$，使其符合公式（2），并且$\mathbf{X}_k$,$\mathbf{X}$ $\boldsymbol{\Theta}$满足公式（1）。

2.2 系统辨识的预备知识

不失一般性，对于时间序列的稀疏辨识，需要解决两个基本问题。一个是基函数库的选择 $\boldsymbol{\Theta}$，另一个是稀疏解 $\boldsymbol{\Xi}_k$ 在方程 Eq. (1) 的样本内数据上的适用性。

基函数的选择是稀疏辨识中的一个棘手问题。现有的大多数方法都将所有基本初等函数纳入基函数库矩阵 [23], [28], , 包括多项式基、三角函数基、对数基、指数基、常数基等。简而言之， $\boldsymbol{\Theta}(\mathbf{.})$ 可写为

$$\boldsymbol{\Theta}(\mathbf{.}) = \begin{bmatrix} | & | & | & | & | \\ \sin / \cos(c_1\mathbf{.}) & \mathbf{.}^{c_2} & e^{(\mathbf{.})} & \ln(|\mathbf{.}|) & \mathbf{1} \\ | & | & | & | & | \end{bmatrix}, \tag{3}$$

其中， c_1， c_2 为实数，$(\mathbf{.})$ 表示$(\mathbf{t}$，$\mathbf{X}j)$。在基函数库中，由于不同指数或对数的函数之间存在线性相关性，因此只需要一种指数基或对数基。在这种情况下， $\boldsymbol{\Theta}(\mathbf{.})$ 的主体由多项式基和三角函数基构成。考虑到式（3）中的 c_1， c_2 可以是任意实数，一种解决方案是根据数据选择一种基，并遍历 c_1 或 c_2，以获得尽可能多的自然数。其理论依据是魏尔斯特拉斯两个逼近定理 [35]。如果知道动力学方程右侧项可能的组成，这似乎是个不错的方法。许多研究利用对潜在控制方程的先验知识，取得了显著成果 [24]–[26] 。然而，无法事先得知右侧项的形式。由于缺乏先验知识，即使使用大量基函数拟合最简单的曲线，也可能得到较差的预测结果（见图 1.(a)）。与此同时，过多地选取基函数不仅会带来过拟合的风险，还会造成巨大的计算压力，这被称为维数灾难 [36]。在这种情况下，亟需提出一种寻找基函数库的新方法。

稀疏解 $\boldsymbol{\Xi}_k$ 的适用性也是一个难题。由于原问题是NP困难[37] 问题，现有的大多数解法都是近似解，或是在某些附加条件下的最优解。目前，稀疏求解 $\boldsymbol{\Xi}_k$ 主要有三种方法。第一种思路是通过贪心思想求解。OMP[20], 、松弛贪心算法（RGA） [38] 和 SINDy [23]都是这类思路的代表。它们在每次迭代中选择局部最优解，从而使原来的 NP困难问题能够在多项式时间内求解。以 SINDy [23] 为例，在第 j 次迭代中求解 $\boldsymbol{\Xi}_k$ 时，迭代过程如下:

$$\arg\min_{\boldsymbol{\Xi}_k} \left\| \mathbf{X}_k^{(j)} - \sum_{i\in\Gamma^{(j)}} \theta_i(\mathbf{.}) \left(\xi_i^{(j)}\right)_k \right\|^2, \tag{4}$$

其中， $\theta_i(\mathbf{.})$ 是 $\boldsymbol{\Theta}(\mathbf{t}$ $\widetilde{\mathbf{X}})$ 的第 i 列，而 $\Gamma^{(0)}$ 和 $\Gamma^{(j+1)}$ 定义为

$$\begin{aligned} \Gamma^{(0)} &= \{1, 2, \cdots, (ml+1)p\}, \\ \Gamma^{(j+1)} &\leftarrow \Gamma^{(j)} \bigcap \left\{ i \left| |\xi_i^{(j)}| \le \varepsilon \right. \right\}^c, \end{aligned}$$

其中，阈值 ε 是可调整的超参数。该方法通过使用式 (4) 中的最小二乘法，可以快速找到一个稀疏解$\boldsymbol{\Xi}_k$ ，但显然这并不是模型中最稀疏的情况。

第二种思路是将式 (2) 中的 l_0 范数替换为$l_q(q \ge 1)$ 范数，从而将原问题转化为凸优化问题。在这种情况下，可以在多项式时间内求解。原问题等价于以下优化问题，可描述为

$$\arg\min_{\boldsymbol{\Xi}_k} \left\| X_k - \boldsymbol{\Theta}(\mathbf{t}, \widetilde{\mathbf{X}})\boldsymbol{\Xi}_k \right\|^2 + \lambda \left\|\boldsymbol{\Xi}_k\right\|_{l_q}^q, \tag{5}$$

其中， λ 是可调节的超参数。原始问题指的是 $q = 1$下的 LASSO 和 $q = 2$下的岭回归。坐标下降法 [21]和最小角回归 [39] 都是该方法的成熟解法。在 l_p 范数中，使用 l_1 范数可以获得最佳的稀疏效果。然而，尽管得到的结果是稀疏的，但它并不是原始问题的解。

在不严格要求 $\boldsymbol{\Xi}_k$ 具有稀疏性的情况下，通过这种方法得到的$\boldsymbol{\Xi}_k$ 是原始问题的一个良好替代方案。

第三种思路是为原问题的系数赋予先验概率，并使用贝叶斯统计学习方法求解问题。可以假设各个 ξ_i 相互独立，且均服从高斯分布 $N(0,\ \sigma_i^2)$，即，

$$p(\xi_i) = \int N(\xi_i; 0, \sigma_k^2) p(\sigma_k^2) \mathrm{d}\sigma_k^2,$$

其中 $p(\sigma_k^2) \propto 1/\sigma_k^2$ （正态-杰弗里斯先验）。根据贝叶斯公式，化简后可得到 $\boldsymbol{\Xi}_k$ 的后验概率，即，

$$\begin{aligned} p(\boldsymbol{\Xi}_k) \propto & -\boldsymbol{\Xi}_k^\top H^\top H \boldsymbol{\Xi}_k \\ & - 2\boldsymbol{\Xi}_k H^\top \mathbf{X}_k - \boldsymbol{\Xi}_k^\top \Upsilon(\sigma_\Xi^2) \boldsymbol{\Xi}_k, \end{aligned}$$

其中 H是 $\Theta(\boldsymbol{.})$的核函数，而 $\sigma_\Xi^2 = \mathrm{diag}(\sigma_1^{-2},\ \cdots,\ \sigma^{-2}{}_{(ml+1)p})$。对后验概率 $p(\boldsymbol{\Xi}_k)$进行对数最大似然估计，可得到最终的优化形式，即，

$$\begin{aligned} \arg\min_{\boldsymbol{\Xi}_k} & \sum_{i=1}^{n} \frac{1}{2\sigma_i^2} \left| \mathbf{X}_k^{(i)} - \boldsymbol{\Theta}(\mathbf{t}^{(i)}, \widetilde{\mathbf{X}}^{(i)}) \boldsymbol{\Xi}_k^{(i)} \right| \\ & + \frac{n}{2} \sum_{i=1}^{(ml+1)p} \log \sum_{j=1}^{n} \left(\xi_i^{(j)} \right)_k, \end{aligned} \tag{6}$$

其中，上标 i （或 j）表示时间序列中第i个（或第 j个）时刻的取值。式（6）是一个可求解的凸优化问题。这里，先验概率不必是高斯先验，其他类型的先验概率也在迅速涌现 [40]–[42]。这三种思路，即式（4）—（6），极大地丰富了稀疏辨识的求解方法。

然而，沿着这一研究思路，自然而然会产生一个问题：能否选择基于模型的方法，而非数据驱动方法，来解决实际预测问题？换言之，在求解 $\boldsymbol{\Xi}_k$之前，我们能否预先判断稀疏辨识方法的有效性？现有方法并未提供评价模型有效性的指标，因此我们只能不断尝试各种方法，包括基于模型的方法和数据驱动方法，这既费力又耗时。

在下一节中，我们将为基函数给出一种独特且通用的先验知识，并提出 SIABF 的评价指标。这将在一定程度上缓解上述两个问题。

3 基于自适应基函数的稀疏辨识（SIABF）

本节将介绍核心方法论。SIABF 分为三个步骤。第一步用于在时间序列中寻找自适应基。第二步用于判别基的有效性。第三步用于计算稀疏矩阵 $\boldsymbol{\Xi}_k$ 并验证模型的鲁棒性。SIABF 的实施过程如图 2 所示。借助 SIABF，可以有效解决第 2 节中提到的当前基于模型的方法所存在的问题。

3.1 方法论

本部分将提出一种选择自适应基函数的方法，该方法不同于依赖人工经验的传统基于模型的方法。在此情况下，还将给出一种用于验证自适应基函数有效性的验证方法。

3.1.1 步骤 *I*：自适应周期

首先，需要确定在式（3）中选择什么基函数进行识别。注意，大多数实际数据都包含周期性、趋势性和随机性 [31]。因此，应设计有针对性的先验基函数。考虑到实际样本数据中往往存在突变点，而且高阶多项式识别的尾部会产生较大的龙格现象 [43]，多项式拟合并不适用。此外，由于多项式函数在无穷远处的取值趋于无穷，较远的预测点可能会产生较大误差。因此，识别用基函数库将主要采用式（3）中的三角基。

对于某个长度为 N 的物理系统的时间序列$\mathbf{X}$ ，为了获取隐藏在数据中的潜在周期信息，采用模值为 A_w、频率为 f_w 的 DFT(**X**)，

$$A_w = \left| \sum_{s=0}^{N-1} X_s \cdot e^{-\frac{j2\pi ws}{N}} \right|, w = 0, 1, \cdots, \left[\frac{N-1}{2} \right], \tag{7}$$

$$f_w = \frac{w}{N \cdot d},\ w = 0, 1, \cdots, \left[\frac{N-1}{2} \right], \tag{8}$$

其中，d 表示采样频率，X_s 指**X** 中第 s 个分量的值，j 是虚数单位。由于 DFT 具有共轭对称性，只需考虑正频率，这意味着 w 只需遍历数据长度的一半。根据 DFT 的性质，A_w 的值越大，相应频率 f_w 对时间序列数据的影响就越大。因此，我们将 A_w 从大到小排序，选取前 Q 个，并找出对应的自变量 w_q^*，即

$$w_q^* = \arg\max_{w}{}^{(q)} A_w, q = 1, 2, \cdots, Q,$$

其中，$\max^{(q)}$ 表示 A_w 中第 q 大的数。这样便可以找到时间序列的高频率 $f_{w_q^*}$。利用频率-周期转换公式 $T_{(q)} = 1/f_{w_q^*}$，即可得到时间序列中最可能存在的周期。这些周期记为 $T_{(1)} T^{(2)}, \cdots, T^{(Q)}$，并称为自适应周期。将这些 $T_{(q)}$ 代入识别过程的基函数库，其中，式（3）中的每个 $c_1^{(i)}$ 均可写为

$$c_1^{(q)} = \frac{2\pi}{T^{(q)}},\ q = 1, 2, \cdots, Q. \tag{9}$$

值得注意的是，时间序列的变化基本上可以用这些 Q 周期来表示。因此，可以通过在 Q 周期中选择对应振幅的最大值，描述该时间序列的主要变化周期。与人为选择自适应基函数或仅使用受Taylor启发的多项式相比，这能进一步提升模型的可解释性

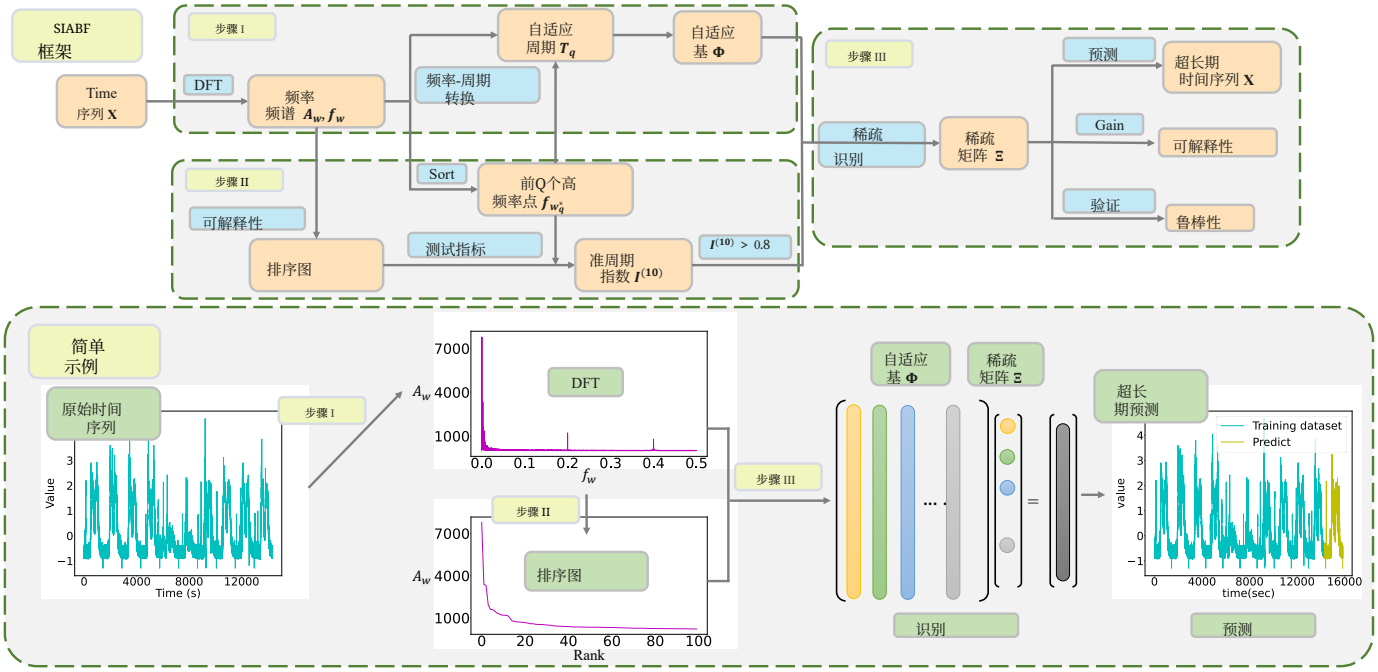


图 2. SIABF 框架及一个简单示例。

展开。此外，供选择的先前周期数（即 Q的大小）可能不必很多。与传统的自然数遍历方案[23], [24], 相比，我们的模型将大幅降低计算复杂度。这使模型更适合解决大规模长期预测问题。

这样，稀疏辨识中的基函数便从式 (7) 和式 (8) 中自适应地选出。值得注意的是，自适应基函数的主要部分仅通过样本时间序列选取，这种选取具有自适应性，并包含容易被忽略的关键周期性特征。由此解决了上述稀疏辨识中的第一个问题，即如何选择自适应基函数。

3.1.2 步骤 *II*：准周期指数

通常，复杂系统中蕴含着特定的物理机制，但这些机制无法被完全揭示和描述。换言之，并非所有先验知识（包括通过离散傅里叶变换获得的周期）都能纳入用于预测的识别模型。本文将在辨识与预测之前引入指数 $I^{(10)}$，用于评估基函数选择的质量。该指数的意义在于避免浪费时间训练不必要的模型。此外，该指数还有助于区分基于模型的方法与数据驱动模型之间的界限。

原始时间序列需要尽可能用少量自适应基函数来表示。其优势之一是，少量基函数识别发生过拟合的概率非常低。相反，如果需要大量自适应周期才能构成基函数，那么即使能够还原原始数据，也不会用于预测，因为发生过拟合的概率很高。

为了量化所提方法在自适应周期选择方面的有效性，我们将对离散傅里叶变换中的振幅进行排序，以得到排序图（图 3）。根据离散傅里叶变换的性质，排序图中的纵坐标表示相应频率处的振幅。如果图中排名靠前的振幅下降得足够快，也就是说，大多数离散傅里叶变换结果的模趋近于零，那么可以得出结论：影响时间序列的周期特征仅限于振幅较大的前几个频率。因此，只需使用少量先验周期就足以描述时间序列。

相应地，提出了准周期指数 $I^{(10)}$，其定义为任意由 Q 在离散傅里叶变换中具有最大振幅的先验基函数形成的排序图中前十个斜率 $k_{\max}^{(i)}$ 之和的比值，即：

$$I^{(10)} = \sum_{i=1}^{10} k_{\max}^{(i)} \Big/ \max A_w.$$

准周期指数 $I^{(10)}$ 表示对整个时间序列贡献最大的十个频率所占的比重。我们可以根据 $I^{(10)}$ 的值，凭经验判断得到的自适应周期能否进一步用于稀疏辨识。根据目前的实验情况，通常，如果 $I^{(10)} \geq 0.8$ （图 3 示例 1），其长期预测效果会远优于基于神经网络的方法；但如果 $I^{(10)} \leq 0.5$ （图 3 示例 2），这种系统辨识方法可能无法奏效。当 $I^{(10)}$ 处于中间值（$0.5 < I^{(10)} < 0.8$）时，与数据驱动方法相比，它有利也有弊。请注意，$I^{(10)}$ 的值越大，这种方法在长期预测中优于数据驱动方法的可能性就越大。该指标或可为区分基于模型的方法与数据驱动方法提供有效判据。

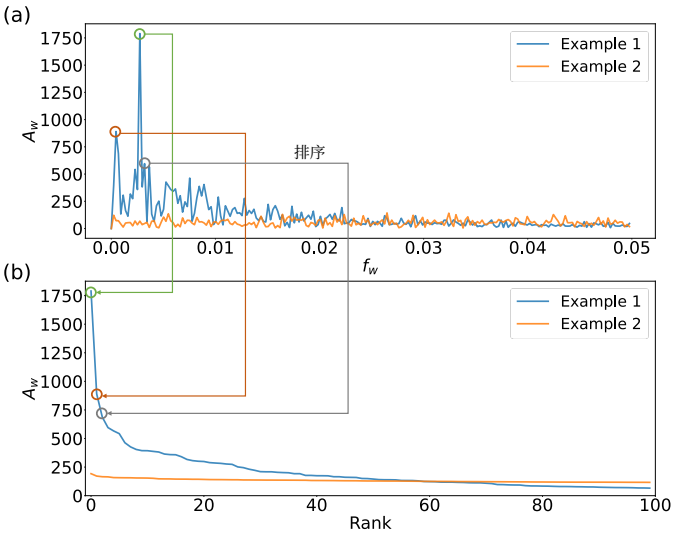


图3. 离散傅里叶变换（a）和排序图（ b).

这样一来，就可以在进行稀疏辨识之前判断该方法是否适用。上述稀疏辨识的第二个问题也就得到了解决。

3.1.3 步骤 *III*：稀疏辨识与鲁棒性测试

在此步骤中，利用步骤 I-II 中获得的先验基，可通过任何传统识别方法对 $\boldsymbol{\Xi}_k$ 进行稀疏识别。

与式（4）-（6）类似，使用 l_1 范数和先验基函数识别稀疏矩阵 $\boldsymbol{\Xi}_k$，即:

$$\arg\min_{\boldsymbol{\Xi}_k} \int_0^T \left(\left\| \mathbf{X}_k - \boldsymbol{\Theta}(\mathbf{t}, \tilde{\mathbf{X}}) \boldsymbol{\Xi}_k \right\|^2 \right) \mathrm{d}\mathbf{t} + \lambda \|\boldsymbol{\Xi}_k\|_1, \quad (10)$$

其中，T 是识别时长，λ 是一个超参数，用于调节 $\boldsymbol{\Xi}_k$的稀疏性。 λ的值可通过交叉验证法 [21], [28]确定。公式（10）的求解过程是一个凸优化问题，可通过坐标下降法 [21] 或最小角回归 [39]快速获得稀疏解。

$\boldsymbol{\Xi}_k$ 的结果捕捉了时间序列中真正有效的自适应基函数。这些识别出的基函数及其系数将共同用于长期预测。每一步可表示为:

$$\mathbf{X}_k^{(t+1)} = \boldsymbol{\Theta}(\mathbf{t}+1, \tilde{\mathbf{X}}) \boldsymbol{\Xi}_k, k = 1, 2, \cdots, m. \quad (11)$$

此外，对于代数方程，公式（11）无需迭代。在这种情况下，它将具有极强的长期预测能力，且不会产生误差累积。

此外，还需要测试 SIABF 的鲁棒性。进行如下实验：对照组使用完整的训练数据，通过 SIABF 进行预测。实验组则删除 5% 的训练数据。对于删除的数据，采用插值法填补缺失值，以使训练数据的维度与对照组保持一致。其次，对训练数据施加随机扰动，该扰动不一定服从正态分布，而是完全随机的。在这种情况下，训练数据将与对照组不同。分别对实验组和对照组进行实验，并比较预测指标。如果实验组的预测效果并未显著弱于对照组，则表明 SIABF 确实具有鲁棒性。

3.2 实现与算法

将自适应周期和准周期指数$I^{(10)}$ 纳入识别方法，使用 l_1 范数作为惩罚项，并考虑时间序列中的噪声，即可得到完整的 SIABF 模型流程。SIABF 能够自适应地构建基函数库，并实现自我评估。

如图2所示，整个识别过程可分为两个阶段和三个步骤。第一步和第二步通过验证获得自适应基。第三步求解稀疏识别问题并完成预测。回顾整个模型，所提出的方法总结于算法1中。

算法1： 使用自适应基函数的稀疏辨识（SIABF）

输入： 时间序列 $\mathbf{X}_k$，$k = 0$、1，$\cdots$，m。

输出： 稀疏矩阵 $\boldsymbol{\Xi}_k$ 和长期预测时间序列 $\hat{\mathbf{X}}_k, k = 0, 1, \cdots, m.$

```
/* 步骤 I 和 II： */
初始化和标准化。
对于 k = 0, 1, ··· , m do
    A_w^k, f_w^k = 离散傅里叶变换(X_k)。
    从 A_w^k中选出最大的 q，并获得与振幅对应的频率。
    计算先验基函数 Θ(.)，使用频率-周期转换和公式（9）。
    if I_k^(10) > 0.8然后
        对.....做出决策时具有高度信心
        准确的长期预测具有高度信心。
    else
        需要进行方法比较。
/* 步骤 III： */
使用公式（10）计算系数矩阵 Ξ_k 。
使用公式（11）求解长期预测。
测试 SIABF 的鲁棒性。
```

4 实验

4.1 实验设置

4.1.1 数据集

为测试所提出模型 SIABF 的性能和稳定性，使用了三个时间序列数据集，包括水利、全球温度和金融数据集[44]–[46]。

水利数据集（数据集 1），来源于美国国家海洋和大气管理局国家水文模型回溯数据集，记录了一条河流水利数据的均值、最大值和标准差

，采样间隔为 1 小时。记录时长长达数十年，记录点超过一万个。

全球温度数据集（数据集2）来自 NCEP-NCAR 再分析资料，记录了自1950年以来近70年的全球温度数据。记录的空间分辨率为$1.875° \times 1.875°$。时空数据点数量超过50万。

金融数据集（数据集3）记录了中国22个主要城市的金融数据。该数据集包含连续九天的记录，采样间隔为1小时。数据总量超过10万。

4.1.2 基线方法

为了展示 SIABF 的有效性，选取了几种常用且较新的时间序列预测模型。根据基于模型的方法和数据驱动方法，基线方法分为两类。对比的基于模型的方法包括：

- **ARIMA** [8] 自回归积分滑动平均模型。通常记作 ARIMA(p, d, q)，其中，p 为自回归项的数量，q 为移动平均项的数量，d 为使序列平稳所需的差分次数（阶数）。

- **SBL** [22] 稀疏贝叶斯学习方法。一种常见且成熟的稀疏辨识方法，其假设系数的先验分布服从均值为零的正态分布。
- **SINDy** [23] 一种通过阈值最小二乘法发现控制方程的方法。该方法利用贪心思想，可以快速找到系数矩阵的稀疏解。

所比较的数据驱动方法包括：

- **LSTM** [9] 一种常用的用于时间序列预测的循环神经网络，专门用于解决一般循环神经网络中的长期依赖问题。

- **GRU** [10] 一种针对长短期记忆网络改进的循环神经网络。该网络只需更少的训练参数，便可获得与长短期记忆网络接近的预测结果。

- **ResNet** [47] 2016 ImageNet冠军网络。此后，该网络已应用于包括时间序列在内的各种场景。

- **Autoformer** [1] 目前用于长期预测的最先进网络。在多个开放数据集的预测中取得了最佳结果。

4.1.3 评估方法

为确保公平比较，所有实验均在同一台机器上进行，且所有数据都经过归一化处理。预测结果从时间、准确率和可解释性三个角度进行评判。在准确率方面，采用四个常用指标来评估其性能，包括 RMSE（均方根误差）、MAE（平均绝对误差）、R^2（R 平方）和 MAPE（平均绝对百分比误差）。为避免接近零点对 MAPE 的影响，与传统的 MAPE 定义不同，采用以下形式

$$\mathrm{MAPE} = \mathrm{Median}\left\{\frac{|y_1 - \hat{y}_1|}{|y_1| + \varepsilon}, \frac{|y_2 - \hat{y}_2|}{|y_2| + \varepsilon}, \cdots, \frac{|y_n - \hat{y}_n|}{|y_n| + \varepsilon}\right\},$$

其中，$\hat{y}$ 为预测值，ε 为较小的正数，以防分母为零。每种方法运行十次，以获得平均结果。

4.1.4 参数设置

我们已搜索并测试了适用于验证数据的最优参数，并将其作为所有基线方法的初始输入。经过重复测试，基于模型的方法的超参数设置如下：LASSO 中为 λ = 5e-4，SBL 中为 $\alpha_1 = \lambda_1$ =1e-8、$\alpha_2 = \lambda_2$ = 5e-5，SINDy 中为 ϵ = 1e-2。b, p, q在 ARIMA 中自动选择。数据驱动方法中的部分超参数设置如下：批量大小为 = 50，学习率为 = 1e-3，网络训练轮数均为 $\geq$ 200。

4.2 性能比较

性能比较结果如表 I 所示。对于每个数据集，考虑了两种预测类型：使用完整训练数据和不完整训练数据进行长期预测。对于每种类型，共测试了包括所提出的 SIABF 在内的八种方法，并采用五项关于准确率和耗时的指标进行评估。从表 I 的结果中可以得出以下观察结论。

首先，SIABF 在完整和不完整的训练数据上均取得了一致的最先进结果。与数据驱动方法相比，SIABF 在水利数据上的均方根误差降低了 40%（0.491→0.814），在温度数据上降低了 26%（0.430→0.585），在金融数据上降低了 38%（0.336→0.540）。总体而言，与基线数据驱动方法相比，SIABF 的均方根误差平均降低了 33%。

其次，在所有预测方法中，只有 SIABF 的 R^2 始终大于 0.5。在大多数方法中，长期预测的 R^2 小于零。这清楚地表明，目前大多数方法都无法准确预测未来数据的趋势，也就是说，它们无法准确预测数据的峰值和谷值。然而，SI ABF 能够准确把握数据的长期趋势，这可归因于两个原因：1）SIABF 能够准确计算数据中潜在的周期，尽管事先并不知道数据的先验知识。2）SIABF 能够稀疏地选择潜在的先验周期，以确保训练集不会出现过拟合现象。

第三，SIABF 与数据驱动方法相比，可节省大量训练时间。在本文中，尽管所有数据驱动方法都是在高配置的图形处理器上训练的，但每次实验的训练时间仍长达数分钟。相比之下，SIABF 可在一秒内完成长期预测，同时保持预测准确率，速度比基于神经网络的方法快数百倍甚至上千倍。

最后，SIABF 对不完整数据和含噪数据具有鲁棒性。在第二类实验中，随机删除了 5% 的训练数据，并向所有训练数据添加了非规则噪声。然而，显然 SIABF 不受训练数据不完整或噪声的影响。表 I 中几乎所有指标的变化都小于基线方法。

总而言之，在时间序列长期预测实验中，SIABF 在精度、速度和鲁棒性方面优于当前方法。

表 I 三个数据集上不同模型的性能。所有实验均考虑长期预测。三个数据集的预测长度分别为1460、1220和1440个点。

预测类型		完整训练数据集				不完整且含噪的训练数据集				耗时
指标		RMSE	MAE	R^2	MAPE (%)	RMSE	MAE	R^2	MAPE (%)	时间（秒）
Data -组 1	ARIMA	1.052	0.859	-1.37E5	35.011	1.052	0.861	-7.89E5	34.517	30.874
	SBL	0.791	0.548	-0.833	23.153	0.789	0.547	-0.831	23.148	0.042
	SINDy	0.864	0.643	0.046	38.186	0.864	0.643	0.050	38.323	**0.022**
	LSTM	1.058	0.876	-9.22E3	33.054	1.586	1.271	-2.009	72.813	718.962
	GRU	1.330	1.113	-4.459	32.536	1.427	1.078	-1.736	34.386	650.110
	残差网络	1.087	0.800	-3.243	44.520	1.097	0.806	-2.183	48.667	71.883
	Autoformer	0.814	0.613	-1.382	25.668	0.903	0.661	-1.817	26.782	92.802
	SIABF	**0.491**	**0.393**	**0.777**	**19.222**	**0.494**	**0.392**	**0.772**	**19.749**	0.247
Data -第 2 组	ARIMA	1.067	0.928	-600.733	0.401	1.063	0.925	-513.806	0.401	56.346
	SBL	0.575	0.476	0.590	0.206	0.572	0.473	0.593	0.205	**0.003**
	SINDy	0.599	0.491	0.562	0.213	0.595	0.489	0.565	0.212	0.008
	LSTM	0.602	0.474	0.604	0.205	0.709	0.567	0.444	0.245	482.761
	GRU	0.683	0.553	0.509	0.239	0.666	0.537	0.508	0.232	337.221
	残差网络	0.585	0.471	0.532	0.204	0.529	0.420	0.646	0.182	38.537
	Autoformer	0.766	0.629	0.005	0.271	0.727	0.589	-0.148	0.254	183.429
	SIABF	**0.430**	**0.334**	**0.790**	**0.144**	**0.421**	**0.328**	**0.794**	**0.142**	0.033
Data -集 3	ARIMA	0.999	0.747	-16.372	123.700	1.134	0.822	-39.582	257.555	176.488
	SBL	0.901	0.822	0.070	64.558	0.898	0.819	0.036	64.522	**0.003**
	SINDy	1.207	1.124	-10.855	84.089	1.202	1.120	-10.858	83.714	0.008
	LSTM	0.638	0.447	0.541	46.816	0.610	0.426	0.416	80.708	1495.221
	GRU	0.540	0.388	0.328	50.072	0.595	0.435	0.240	60.308	1025.846
	残差网络	0.561	0.410	0.151	63.778	0.647	0.484	0.118	101.053	124.194
	Autoformer	1.065	0.909	-3.273	174.858	1.147	0.979	-3.112	144.069	141.836
	SIABF	**0.336**	**0.240**	**0.863**	**41.269**	**0.345**	**0.250**	**0.855**	**49.225**	0.032

4.3 可解释性

与基于神经网络的数据驱动方法相比，基于模型的方法还有一个优势，即具有良好的可解释性。本文以水利数据集为例，展示 SIABF 的可解释性。

4.3.1 周期意味着先验信息

与大多数基于模型的方法不同，例如 [23], [24], [28],在我们的实验中，数据集可能存在的自适应基函数是未知的。然而，一些可解释的信息可以通过周期先验方法提供。水利数据集周期秩排名前十的排序图如图 4（右）所示。

通过 SIABF 可以找到数据中最有可能存在的周期。仔细检查这些获得的先验周期后，会惊讶地发现，与神经网络相比，它们具有很强的可解释性。在排序图中，排名前三的先验周期分别为 365、2920 和 1460，分别对应一年、八年和四年的时长。这意味着水利领域存在一年、四年和八年的周期性变化。相关专业研究也证实了这一观察结果 [48]。同样，其他周期长度也可以被 16 年整除。还可以通过观测到的振幅来分析这些先验周期对水利数据的影响。

通过这个例子可以清楚地看出，即使数据背后的潜在函数未知，SIABF 也能提供一些先验知识。这表明 SIABF 能够高效地选择基函数，也说明 SIABF 中的参数比神经网络中的参数更具可解释性。

4.3.2 准周期指数的作用

此外，以水利数据集为例，计算准周期指数 $I^{(10)}$。当 Q =50 时，准周期指数的计算过程如图 4（左）所示。

根据公式（3），该数据集中的 $I^{(10)}$ 为 $0.830 > 0.8$。这表明，由水利数据构成的时间序列可以用更少的先验周期表示，尽管直观上并不明显（图 5）。在这种情况下，基于模型的方法基本上优于大多数数据驱动方法。表 I 也验证了这一点。在此情形下，SIABF 提供了一个重要的判据，用于区分基于模型的方法和数据驱动方法。

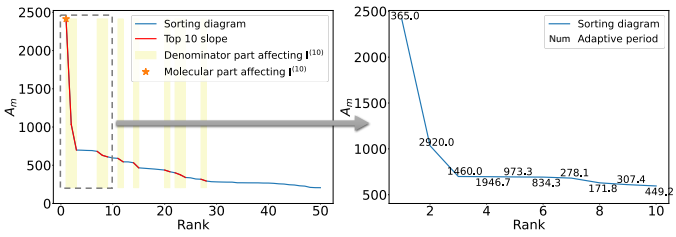


图 4. 用于计算 $I^{(10)}$ 的排序图（左）以及排名前十的水箱自适应周期可视化图（右）。

5 结论

本文提出一种新颖的基于模型的方法 SIABF，用于在没有先验知识的情况下对时间序列进行长期预测。通过简单地实施离散傅里叶变换得到的自适应周期，以极具说服力的形式展现了

底层基函数。所提出的准周期指数 $I^{(10)}$，也能在采用基于模型的白盒方法或数据驱动的黑盒方法之前，为量化预测效果提供较高的置信度。当准周期指数相对较大时（通常为 ≥ 0.8）SIABF 很有可能获得出色的预测性能。

此外，SIABF 提供了一个可解释、高效且有效的辨识与预测框架，且实现过程简单。值得注意的是，本研究表明，任何数据中的潜在周期性都不容忽视，因为它在量化复杂系统的物理机制方面起着至关重要的作用。未来，该框架将考虑周期性之外更多的内在特征，并将扩展 SIABF，使其能够识别和预测网络化系统。

附录 A 可视化与代码

以水利数据为例，不同模型长期预测的可视化结果如图 5 所示。这里，训练数据长度为 16 年（数据点 = 365*16），而预测长度为 4 年（数据点 = 365*4）。

此外，我们的代码可在 [46]获取。

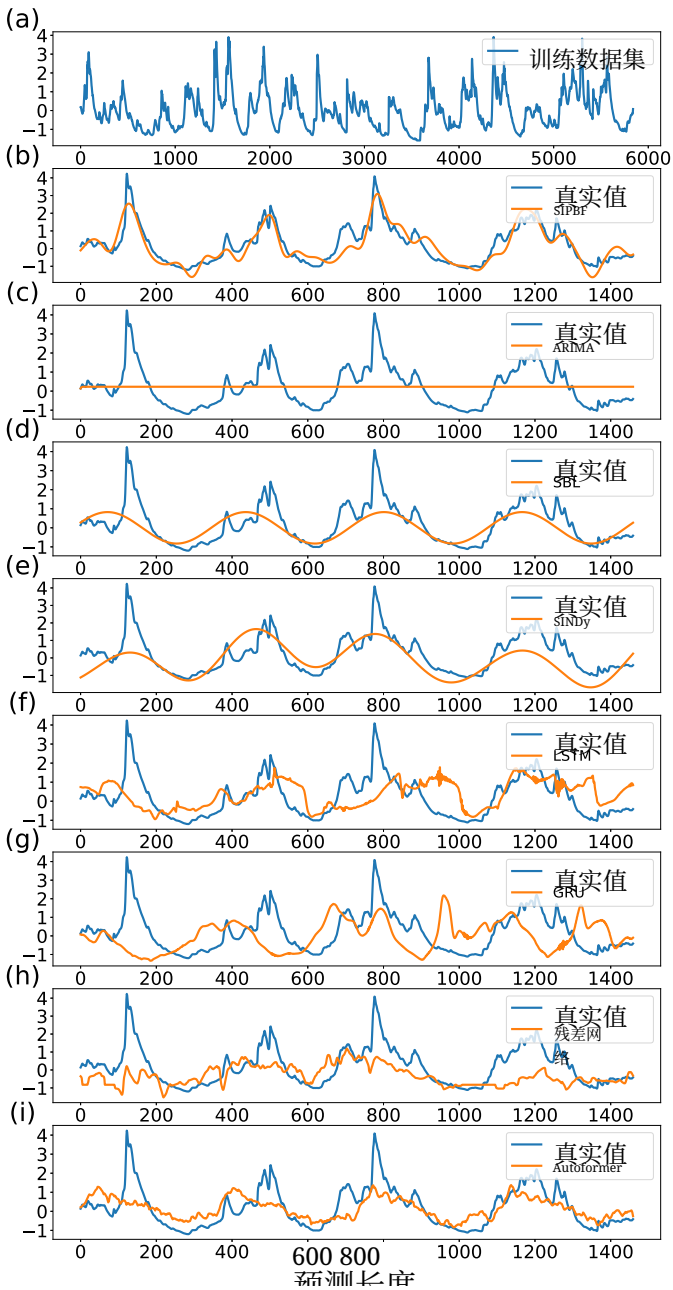


图 5. 不同方法在水利数据长期预测的可视化中的应用。(a) 为训练数据集，(b)–(i) 分别展示了 SIABF、ARIMA、SBL、SINDy、长短期记忆网络、门控循环单元、残差网络和 Autoformer 的测试数据集可视化结果。

参考文献

[1] H. Wu、J. Xu、J. Wang 和 M. Long，“Autoformer：利用自相关的分解式 Transformer 进行长期序列预测”，神经信息处理系统进展，第 34 卷，第 22 419–22 430 页，2021 年。[2] Z. Zhou、Y. Wang、X. Xie、L. Chen 和 C. Zhu，“Foresee 城市稀疏交通事故：时空多粒度视角”，*IEEE* 知识与数据工程汇刊，2020 年。[3] A. Borovykh、S. Bohte 和 C. W. Oosterlee，“利用卷积神经网络进行条件时间序列预测”，*arXiv* 预印本 *arXiv:1703.04691*，2017 年。[4] H. Zhou、S. Zhang、J. Peng、S. Zhang、J. Li、H. Xiong 和 W. Zhang，“Informer：超越高效 Transformer 的长序列时间序列预测”，载于*AAAI* 人工智能会议论文集，第 35 卷，第 12 期，2021 年，第 11 106–11 115 页。[5] J. Deng、X. Chen、R. Jiang、X. Song 和 I. W. Tsang，“用于多变量时间序列预测的多视角多任务学习框架”，*IEEE* 知识与数据工程汇刊，2022 年。[6] X. Xu、L. Zhang、Q. Kong、C. Gui 和 X. Zhang，“用于长期预测的增强历史平均法”，载于

*2022*年第*2*届计算机、控制与机器人国际会议（*ICCCR*）. IEEE，2022，第115–119页。[7] D. E. Rumelhart、G. E. Hinton 和 R. J. Williams，“通过反向传播误差学习表征”，《自然》，第323卷，第6088期，第533–536页，1986年。[8] R. H. Shumway 和 D. S. Stoffer，“ARIMA模型”，载于《时间序列分析及其应用》。施普林格，2017，第75–163页。[9] S. Hochreiter 和 J. Schmidhuber，“长短期记忆网络”，《神经计算》，第9卷，第8期，第1735–1780页，1997年。[10] J. Chung、C. Gulcehre、K. Cho 和 Y. 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。[11] G. Lai、W.-C. Chang、Y. Yang 和 H. Liu，“利用深度神经网络对长短期时间模式建模”，载于第*41*届 *ACM SIGIR* 信息检索研究与发展国际会议，2018，第95–104页。[12] D. Salinas、V. Flunkert、J. Gasthaus 和 T. Januschowski，“DeepAR：利用自回归循环网络进行概率预测”，《国际预测学杂志》，第36卷，1191，2020年第3期，第1181页。

[13] B. Alhnaity、S. Kollias、G. Leontidis、S. Jiang、B. Schamp 和 S. Pearson，《一种基于自编码器和小波、结合注意力机制的深度神经网络，用于植物生长多步预测》，《信息科学》，第560卷，第35–50页，2021年。[14] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、Ł. Kaiser 和 I. Polosukhin，《Attention is all you need》，神经信息处理系统进展，第30卷，2017年。[15] N. Kitaev、Ł. Kaiser 和 A. Levskaya，《Reformer：高效的Transformer》，*arXiv*预印本 *arXiv:2001.04451*，2020年。[16] Y. Gu、J. Jin 和 S. Mei，《l_0 用于稀疏系统辨识的范数约束LMS算法》，《*IEEE*信号处理快报》，第16卷，第9期，第774–777页，2009年。[17] R. Wang、C. Zhang、J. Bian、Z. Wang、F. Nie 和 X. Li，《用于无监督特征选择的稀疏灵活投影》，*IEEE* 知识与数据工程汇刊，2022年。[18] X. Ma、Y. Wang、X. Chu、L. Ma、W. Tang、J. Zhao、Y. Yuan 和

G. Wang，《通过医学特征的相关稀疏先验进行患者健康表征学习》，IEEE 知识与数据工程汇刊，2022年。[19] Y. Ye 和 S. Ji，《稀疏图注意力网络》，IEEE 知识与数据工程汇刊，2021年。[20] Y. C. Pati、R. Rezaiifar 和 P. S. Krishnaprasad，《正交匹配追踪：递归函数逼近及其在小波分解中的应用》，载于第27届阿西洛马信号、系统与计算机会议论文集。IEEE，1993年，第40–44页。[21] R. Tibshirani，《基于套索的回归收缩与变量选择》，英国皇家统计学会会刊 B辑：统计方法论，第73卷，第3期，2011年。[22] M. E. Tipping，《稀疏贝叶斯学习与相关向量机》，机器学习研究期刊，第1卷，6月号，第211–244页，2001年。[23] S. L. Brunton、J. L. Proctor 和 J. N. Kutz，《通过非线性动力系统稀疏辨识，从数据中发现控制方程》，美国国家科学院院刊，第113卷，第15期，第3932–3937页，2016年。[24] Y. Yuan、X. Tang、W. Zhou、W. Pan、X. Li、H.-T. Zhang、H. Ding 和 J. Goncalves，《信息物理系统的数据驱动发现》，自然-通讯，第10卷，第1期，第1–9页，2019年。[25] S. L. Brunton、J. L. Proctor 和 J. N. Kutz，《带控制的非线性动力学稀疏辨识（SINDyC）》，IFAC-PapersOnLine，第49卷，第18期，第710–715页，2016年。[26] N. M. Mangan、S. L. Brunton、J. L. Proctor 和 J. N. Kutz，《通过非线性动力学稀疏辨识推断生物网络》，IEEE 分子、生物与多尺度通信汇刊，第2卷，第1期，第52–63页，2016年。[27] M. Quade、M. Abel、J. Nathan Kutz 和 S. L. Brunton，《用于快速模型恢复的非线性动力学稀疏辨识》，《混沌：非线性科学跨学科期刊》，第28卷，第6期，文章号063116，2018年。[28] T.-T. Gao 和 G. Yan，《从不完整且含噪的数据中自主推断复杂网络动力学》，自然-计算科学，第2卷，第3期，第160–168页，2022年。[29] D. Feller，《数据库在支持计算化学计算中的作用》，计算化学期刊，第17卷，第13期，第1571–1586页，1996年。[30] A. A. Bataineh 和 D. Kaur，《使用住房数据集对人工神经网络中不同曲线拟合算法的比较研究》，载于 NAECON 2018—IEEE 全国航空航天与电子会议，2018年，第174–178页。[31] M. G. Elfeky、W. G. Aref 和 A. K. Elmagarmid，《时间序列数据库中的周期性检测》，IEEE 知识与数据工程汇刊，第17卷，第7期，第875–887页，2005年。[32] T. Lu、Y.-F. Chen、B. Hechtman、T. Wang 和 J. Anderson，《张量处理单元上的大规模离散傅里叶变换》，IEEE Access，第9卷，第93 422–93 432页，2021年。[33] Y. Wang，《地震信号的时频分析》。约翰·威立父子公司，2022年。[34] R. Agarwal、L. Melnick、N. Frosst、X. Zhang、B. Lengerich、R. Caruana 和 G. E. Hinton，《神经加性模型：利用神经网络实现可解释机器学习》，神经信息处理系统进展，第34卷，第4699–4711页，2021年。[35] D. P´erez 和 Y. Quintana，《魏尔斯特拉斯逼近定理综述》，arXiv预印本 math/0611038，2006年。[36] Y. Mao 和 F. Ding，《一种用于Hammerstein系统的新型参数分离辨识算法》，应用数学快报，第60卷，第21–27页，2016年。[37] A. Hernandez、F. Ruiz、A. Desideri、C. Ionescu、S. Quoilin、V. Lemort 和 R. De Keyser，《利用稀疏多项式模型对有机朗肯循环系统进行非线性辨识与控制》，载于2016年 IEEE 控制应用会议（CCA）。IEEE，2016年，第1012–1017页。[38] C. Novara，《非线性稀疏函数的辨识与参数集合成员最优性分析》，IEEE 自动控制汇刊，第57卷，第12期，第3236–3241页，2012年。[39] B. Efron、T. Hastie、I. Johnstone 和 R. Tibshirani，《最小角回归》，统计年鉴，第32卷，第2期，第407–499页，2004年。[40] J. Jin、Y. Yuan、W. Pan、C. Tomlin、A. A. Webb 和 J. Gonc.alves，《使用稀疏贝叶斯学习辨识非线性稀疏网络》，载于2017年 IEEE 第56届决策与控制年会（CDC）。IEEE，2017年，第6481–6486页。[41] W. R. Jacobs、T. Baldacchino、T. Dodd 和 S. R. Anderson，《使用变分推断进行稀疏贝叶斯非线性系统辨识》

ence," *IEEE* 自动控制汇刊，第63卷，第12期，第4172–4187页，2018年。[42] A. Shahmansoori，"通过动态滤波对时变大规模 MIMO 信道进行稀疏贝叶斯多任务学习"，*IEEE* 无线通信快报，第9卷，第6期，第871–874页，2020年。[43] Y. Wu 和 M. Zhu，"从惯性测量中进行姿态重构：缓解动态应用中的龙格效应"，*IEEE* 航空航天与电子系统汇刊，2021年。[44] "美国国家海洋和大气管理局国家水文模型美国本土回溯数据集"，https: //registry.opendata.aws/nwm-archive/。[45] "NCEP-NCAR 再分析"，https://psl.noaa.gov/data/gridded/data . ncep.reanalysis.html。[46] https://github.com/cccs-data/SIABF-and-Finance-data.git。[47] K. He、X. Zhang、S. Ren 和 J. 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，2016年，第770–778页。[48] P. Hellard、M. Avalos、F. Guimaraes、J.-F. Toussaint 和 D. B. Pyne，"精英游泳运动员4年期间训练相关的常见疾病风险。"《运动与锻炼医学与科学》，第47卷，第4期，第698–707页，2015年。