# 结合时间尺度分离的可解释降阶建模

Sebastian Kaltenbach, [1, a)] Phaedon-Stelios Koutsourelakis, [2] 和 Petros Koumoutsakos[3]
1) 苏黎世联邦理工学院 *CSE-Lab* 和哈佛大学工程与应用科学学院
2) 慕尼黑工业大学数据驱动材料建模讲席
3) 哈佛大学工程与应用科学学院

（日期：2023年3月7日）

arXiv:2303.02189v1 [stat.ML] 3 Mar 2023

高维偏微分方程（PDE）在计算物理和工程学中十分常见。然而，求解这类偏微分方程的计算成本可能很高，因此模型降阶至关重要。我们提出一种数据驱动方案，可自动识别涉及的时间尺度，并在向前推演时间以及面对训练数据中未包含的不同初始条件时，生成稳定的预测。为此，我们将非线性自编码器架构与复空间中描述潜在动力学的时间连续模型相结合。该方法可以直接纳入稀疏且非规则采样的训练数据。所学习到的潜在动力学具有可解释性，并揭示了其中涉及的不同时间尺度。我们表明，该数据驱动方案能够自动学习独立过程，并沿系统矩阵的特征向量对线性常微分方程组进行分解。除此之外，我们还展示了所提出框架在隐马尔可夫模型和（离散化的）Kuramoto-Shivashinsky（KS）方程中的适用性。此外，我们还提出了一个概率版本，能够捕捉预测不确定性，并进一步提升确定性框架的结果。

a) 电子邮箱：skaltenbach@seas.harvard.edu

## 一、引言

对海洋动力学和流行病等临界现象进行高保真模拟，已成为决策制定的重要依据。这些模拟基于符合物理规律的偏微分方程，用以描述跨越多个时空尺度的系统动力学，而求解这些方程需要进行繁复的计算（Givon、Kupferman 和 Stuart，2004）。近年来，人们日益关注数据驱动模型的开发，这类模型既能加速求解这些偏微分方程（Alber等，2019），也能揭示控制长期演化的重要低维特征。在多数情况下，数据驱动的降阶模型缺乏可解释性。具体而言，尽管基于神经网络的模型具有良好的预测能力（Chen 等，2018；Li 等，2020；Vlachas 等，2022），但它们对系统动力学提供的仍是黑箱描述。一种可能的补救方法是对已学习的神经网络表示进行符号回归（Cranmer 等，2020），但这种两步流程会增加计算成本。SINDy（Brunton、Proctor 和 Kutz，2016）等多种框架能够学习可解释的动力学，但它们依赖于预先提供的低维描述符和时间导数；对于模拟数据和实验数据而言，时间导数都可能包含很大噪声。其他框架则针对特定问题设计，例如分子动力学（No´e，2020）。本文提出一种框架，只需将观测量的数值（而非其导数）作为训练数据，即可识别可解释的潜在动力学。采用可解释的潜在动力学，能够确保降阶模型保留重要性质的守恒（Kaltenbach and Koutsourelakis，2020，2021）。本方法与基于Koopman算子的方法（Koopman，1931；Budi˘si´c、Mohr 和 Mezi´c，2012；Klus 等，2018）及扩展动态模态分解（eDMD）（Williams、Kevrekidis 和 Rowley，2015）相关，但采用连续的复值潜在空间动力学，且每个潜在维度仅需一个标量变量即可描述潜在空间动力学。因此，我们无需对Koopman矩阵施加任何参数化约束（Pan 和 Duraisamy，2020）。此外，时间连续形式还允许纳入稀疏且非规则采样的训练数据，并在训练阶段结束后快速生成预测。采用复值潜在空间还可纳入谐波效应，并减少所需的潜变量数量。线性和非线性自编码器用于……

p 观测到的 ，high 维时间-

序列映射到低维潜在表示（Kingma and Welling, 2013；Champion等，2019），并通过优化组合损失函数，同时识别自编码器和潜在动力学。因此，降维与发现降阶动力学这两项任务得以统一（Horenko 等，2008；Horenko，2008），而其他框架则将这两部分分开处理（Vlachas 等，2022）。除了使用基于自编码器的架构来识别潜在空间外，也可以采用基于投影的方法（Snyder 等，2022）。

我们还提出了该算法的概率版本（Kaltenbach and Koutsourelakis，2021），其中采用了概率慢特征分析（Turner and Sahani，2007；Zafeiriou 等，2015）。这使得潜在表示不仅具有时间连续性，还能够量化预测不确定性，并将动力学按相关尺度进行层次分解，同时促进发现控制系统在较长时间范围内演化的慢过程。

本文其余部分的结构如下：第二节介绍方法框架及算法细节，并特别关注所推断低维动力学的可解释性。第三节展示三个数值示例，即一个线性常微分方程组、一个隐马尔可夫模型和离散化的 KS 方程。随后，第四节介绍该框架的概率扩展，并将其应用于 KS 方程。最后，我们总结全文，并简要讨论可能的后续研究方向。

## II. 方法论

我们介绍本研究所采用的自编码器，随后介绍可解释的潜在空间动力学，并讨论训练过程。我们考虑来自高维时间序列的数据 $\boldsymbol{x}_n \in \mathbb{R}^f$，其中包含 $n = 1, \ldots\ldots, T$。需要指出的是，不同状态之间的时间间隔不必均匀分布。

具有时间尺度分离的可解释降阶建模

**A. 自编码器**

该方法的一个核心假设是，每个高维状态 $\boldsymbol{x}_n$ 都可以压缩为具有 $c << f$ 的低维表示 $\boldsymbol{z}_n \in \mathbb{C}^c$ 。我们通过一个由参数化编码器和解码器组成的自编码器来确定这种低维表示。编码器将高维表示映射到潜在空间，如下所示：

$$\boldsymbol{z}_n = \boldsymbol{E}_{\boldsymbol{\theta}}(\boldsymbol{x}_n) \tag{1}$$

潜在空间是复数值的。解码器根据潜变量重建高维表示，如下所示：

$$\boldsymbol{x}_n = \boldsymbol{D}_{\boldsymbol{\theta}}(\boldsymbol{z}_n) \tag{2}$$

我们用 $\boldsymbol{\theta}$ 表示编码器和解码器的参数。如第二节 IIC 后文所述，两组参数在训练过程中同时优化，因此无需对它们加以区分。

**B. 可解释潜在空间动力学**

我们在潜在空间中采用传播算子，以捕捉系统的降阶动力学。与其他时间扩展变分自编码器框架不同，我们的表示采用复数值潜变量。此外，各潜变量彼此独立处理。后一特性使我们能够获得可解释的潜在动力学；同时，由于所需参数数量较少，该模型尤其适合在小数据情形下训练。这与 LSTM（Hochreiter 和 Schmidhuber，1997）等时间传播算子形成对比。

对于潜变量 z 的每个维度 $i$ ，我们在复平面中使用以下连续常微分方程：

$$\frac{dz_i}{dt} = -\lambda_i z_i \tag{3}$$

通过求解该常微分方程，我们可以定义该算子：

$$\boldsymbol{z}_{n+\Delta t_n} = \boldsymbol{P}_{\lambda}(\boldsymbol{z}_n, \Delta t_n) = \boldsymbol{z}_n \odot \exp(-\boldsymbol{\lambda}\Delta t_n) \tag{4}$$

其中， $\boldsymbol{\lambda}$ 是一个包含所有单独 $\lambda$的向量， $\Delta t_n$ 表示潜在状态之间的时间步长。符号 $\odot$ 用于表示逐分量乘法。需要指出的是，潜变量以及控制时间演化的参数都是复数，它们在描述系统动力学中的作用类似于特征值。实部与增长和衰减相关，而虚部则表示周期分量。这种方法与基于Koopman算子的方法（Koopman，1931；Mezi´c，2013；Rowley 等，2009）以及扩展动态模态分解（Williams、Kevrekidis 和 Rowley，2015）具有相似之处。与前述方法不同，我们在潜在空间中采用连续形式，从而能够纳入稀疏且非规则采样训练数据，并直接利用潜在空间中的复数。

### C. 训练与预测

我们优化一个损失函数，该函数结合了重构损失以及与我们在潜在空间中学习得到的传播算子误差相关的损失：

$$L = \frac{1}{T}\sum_{n=1}^{T} |\boldsymbol{x}_n - \boldsymbol{D}_{\boldsymbol{\theta}}(\boldsymbol{E}_{\boldsymbol{\theta}}(\boldsymbol{x}_n))|^2 + \frac{1}{T-1}\sum_{n=1}^{T} |\boldsymbol{E}_{\boldsymbol{\theta}}(\boldsymbol{x}_{n+1}) - \boldsymbol{P}_{\lambda}(\boldsymbol{E}_{\boldsymbol{\theta}}(\boldsymbol{x}_n), \Delta t_n)|^2 \qquad (5)$$

需要指出的是，我们可以通过仅对 $N$ 可用训练数据的一个子集进行求和，直接采用小批量训练。对于未见状态的新预测，我们使用编码器生成潜在表示，然后由学习得到的传播算子将其沿时间推进。在指定的时间步，我们使用解码器重构高维解。

## III. 数值示例

我们将算法应用于三个系统。首先，我们展示该算法能够精确复现线性常微分方程的解，并识别其特征值。随后，我们将该框架应用于由复杂潜在动力学生成的高维过程，并成功识别出其动力学。最后，我们将算法应用于 Kuramoto–Sivashinsky（KS）方程。

具有时间尺度分离的可解释降阶建模

**A. 线性常微分方程**

我们考虑关于 $\boldsymbol{x} = \begin{pmatrix} y_1 & y_2 \end{pmatrix}$ 的二维常微分方程系统：

$$\begin{pmatrix} \dot{y_1} \\ \dot{y_2} \end{pmatrix} = \begin{pmatrix} -5 & 2 \\ 2 & -5 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{6}$$

初始条件为 $y_1 = 10$ 和 $y_2(0) = -3$。使用龙格-库塔求解器求解常微分方程，生成 $t_{max}$=2 的训练数据，时间步长为 $\Delta t = 2.5 \cdot 10^{-4}$；随后从所得轨迹中采样 40 条时间序列，每条包含 150 个数据点。因此，训练数据由以下部分组成：

- 40条时间序列
- ，每条时间序列包含 $\boldsymbol{x}$ 的 150 个观测值，时间步长各不相同 $\Delta t$

基于所得训练数据，我们采用线性编码器和解码器结构以及两个潜变量 $\boldsymbol{z}$运行算法。使用 Adam 算法（ Kingma 和 Ba，2014 年）优化损失函数。

对于线性常微分方程，我们可以解析地计算其中涉及的特征值，并将其与算法识别出的参数 $\lambda$ 进行比较。从图1中可以看出，算法成功恢复了正确的值，即给定线性常微分方程的特征值 7 和 3。该系统没有周期分量，两个虚部也如预期趋于零，而实部则收敛到参考值。

此外，我们还能够识别出潜变量$z$ 与训练数据之间的线性映射矩阵。该矩阵由特征向量 (1,1) 和 (1,-1) 的倍数组成，因此对应于正确的解。选择这个示例是为了说明，该算法能够快速识别线性常微分方程的精确解，并将其表示为线性无关分量。

**B. 隐藏多尺度动力学**

我们考虑由底层二维复值过程生成的八维合成时间序列数据。具体而言，数据点 $\boldsymbol{x}$ 是通过先求解两个复值过程 $p_1$ 和 $p_2$ 的时间演化，再

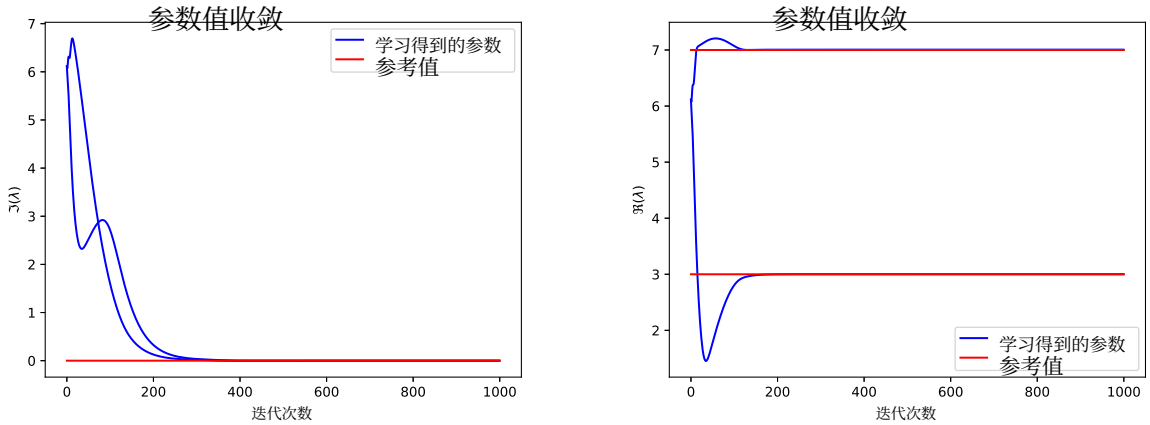



图1. 线性常微分方程测试案例中 $\boldsymbol{\lambda}$ 的收敛情况

使用随机采样得到的线性映射 $\boldsymbol{W}$ 映射到八维空间而生成的。用于生成数据的两个过程之一被设定为远慢于另一个过程，并且两个过程都具有周期分量。

$$\frac{dp_1}{dt} = (-0.1 + 1i)p_1 \tag{7}$$

$$\frac{dp_2}{dt} = (-0.9 + 1.5i)p_2 \tag{8}$$

$$\boldsymbol{x} = \boldsymbol{W} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \tag{9}$$

我们将 40 条时间序列作为训练数据，每条序列包含 150 个数据点。这些序列通过对所述过程进行模拟（最长模拟时间为 $t = 15$ 秒）获得，随后从所得数据点中进行采样。因此，训练数据由以下内容组成：

- 40条时间序列
- 其中每个包含 150 个 $\boldsymbol{x}$ 的观测值，时间步长为均匀的 $\Delta t = 0.0025$

得到的自编码器由一个线性层构成，该层同时用于解码器和编码器。

模型使用 Adam优化器训练了 5000 次迭代，学习率为$10^{-3}$。参数 $\lambda_1$ 和 $\lambda_2$ 的收敛结果见图2。我们注意到，衰减较慢、因而对系统长期演化影响更大的过程，其收敛速率高于衰减较快的过程。

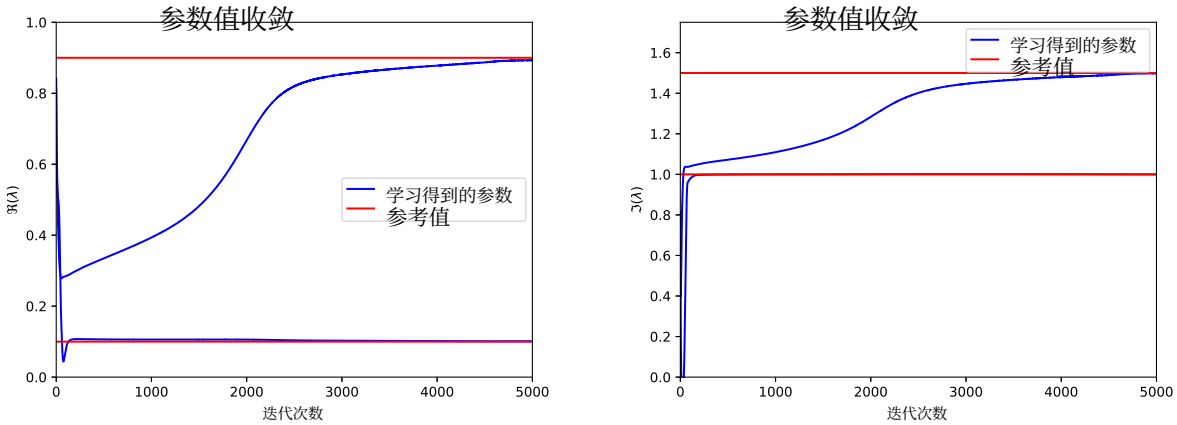



图 2. 潜在动力学参数实部的收敛情况。二者都在收敛至参考值。值得注意的是，较慢过程的收敛速度明显快于较快过程。

利用得到的参数 $\lambda$ 以及训练好的自编码器，我们基于训练所用的最后一个时间步计算预测，即应用编码器获得潜在表示，然后利用潜在动力学推进潜在表示随时间的演化。随后，我们使用解码器重建完整的高维系统。结果见图3，预测与参考数据吻合良好。

该示例表明，我们的模型能够成功执行降维，并且还表明潜在过程之间的收敛速率可能不同。后者与模型训练相关，因为为了获得准确的预测，所有潜在过程及其动力学都应达到收敛。

### C. Kuramoto-Sivashinsky

最后，我们将所提出的算法应用于 KS方程，旨在为解 $u(y,\ t)$ 建立一个降阶模型：

$$\frac{\partial u}{\partial t} = -\mu\frac{\partial^4 u}{\partial y^4} - \frac{\partial^2 u}{\partial y^2} - u\frac{\partial u}{\partial y} \tag{10}$$

我们采用了周期边界条件，$\mu = 1$ 并设置了区域大小 $y \in [0, 22]$。对于这一域大小，KS 方程表现出一个结构稳定的混沌吸引子，正如.....中所讨论的

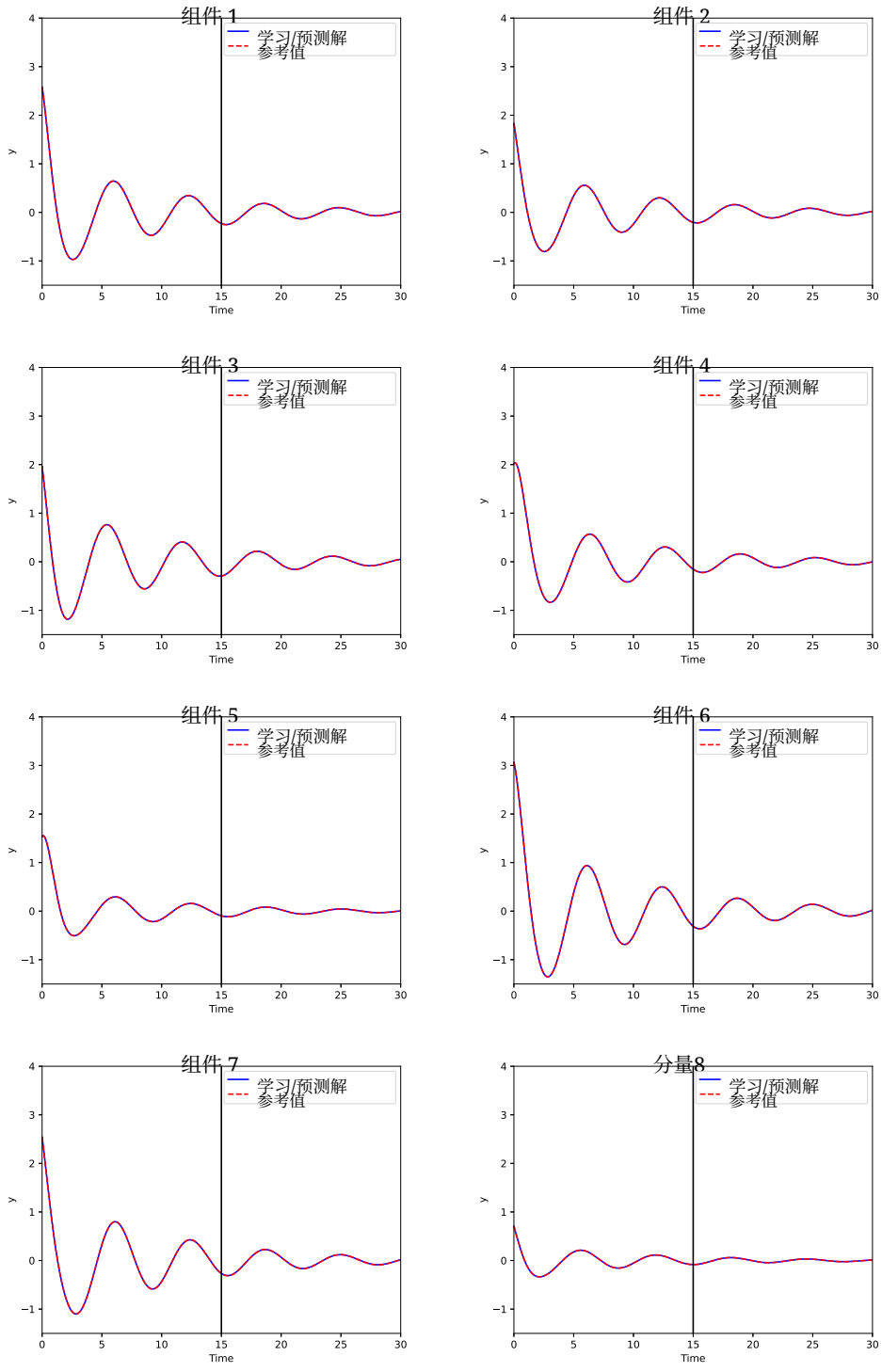



图3。八维系统所有分量的重构与预测演化。黑线将提供训练数据的区域与没有训练数据的区域分隔开。

Robinson (1994)；Kassam 和 Trefethen (2005)。该方程在空间上以$_{22\ 64}$ 的离散步长进行离散化，从而得到维度为 64 的状态向量 $\boldsymbol{x}$ 以及一个由耦合常微分方程构成的非线性系统。使用刚性四阶求解器对该系统求解（Kassam 和 Trefethen，2005），并将解在前 250 秒内的 1000 个数据点划分为 40 个时间序列，每个序列包含 25 个点，用作训练数据。因此，训练数据由以下内容组成：

- 40时间序列
- w每个序列包含 $\boldsymbol{x}$ 的 25 个观测值，时间步长均匀为 $\Delta t = 0.25$

我们采用了一个非线性编码器和一个解码器，每个都包含四个全连接层，并使用 ReLU 激活函数；此外，在全连接层之间还使用了 Dropout 层（Srivastava 等，2014）。我们使用 Adam 和 $5 \cdot 10^4$ 的学习率训练模型 200000 次迭代，并假设潜在空间为五维。我们得到了图 4 中的 $\lambda$。其中四个潜变量的 $\lambda$接近于零，因此具有缓慢的时间动力学，负责长期演化；另一个潜变量则迅速衰减。

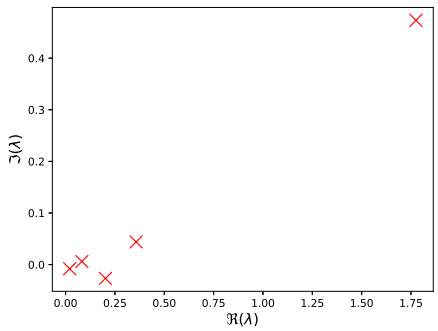



图 4. KS 方程的求得参数 $\lambda$

基于所获得的参数，我们针对训练数据中未包含的未知初始条件进行预测。尽管仅使用了非常有限的训练数据，我们仍能够根据预测结果重构出正确的相空间。相空间的结果如图 5 所示。尽管时间动力学中的小尺度波动未能得到很好的捕捉，但该模型识别出了正确的流形，与参考解相比具有较高的准确性。所有相空间均通过对数据或预测结果使用有限差分算子获得。这些结果与

Vlachas 等（2022年）的结果一致，其基于长短期记忆网络的时间动态模型同样能够找到正确的相空间，但无法在长期预测中跟踪实际动力学。我们的模型无法考虑时间演化中的噪声，因此处理混沌的小尺度波动颇具挑战性。我们认为，算法的概率版本可能会在此方面更具优势。

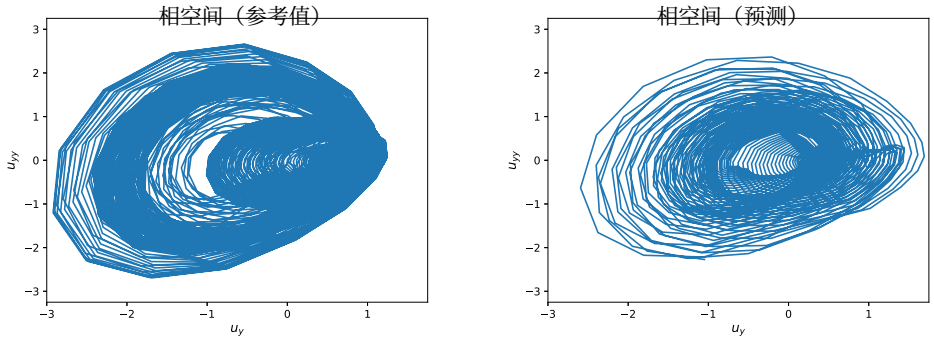



图 5. 参考解的相空间（左）与预测结果的相空间之间的比较

## IV. 概率扩展

本节针对前文讨论的确定性模型，给出一个完整的概率公式。我们将自编码器替换为变分自编码器，并将潜在空间中的常微分方程替换为随机微分方程。我们优化的损失函数是证据下界（ELBO）。

### A. 模型结构

我们假设概率模型满足以下关系：潜在空间的每个维度 $i$ 均遵循 Ornstein-Uhlenbeck（OU）过程，潜在空间中还有一个维纳过程 $W_t$：

$$dz_{t,i} = -\lambda_i z_{t,i}\, dt + \sigma_i dW_{t,i} \quad \lambda_i \in \mathbb{C},\ i = 1, 2, \ldots, c \tag{11}$$

$$\boldsymbol{x}_n = \boldsymbol{G}(\boldsymbol{z}_n,\ \boldsymbol{\theta}) \tag{12}$$

我们再次假设潜变量 $\boldsymbol{z}_t$ 为复数值且先验独立。选择复变量是因为其演化包含谐波分量，这种分量在许多物理系统中都能观测到。我们假设初始条件为 $z_{0,\ i} \sim \mathcal{CN}(0,\ \sigma^2_{0,\ i})$。因此，与模型潜在空间动力学相关的参数总数为$\{\sigma^2_{0,i}, \sigma^2_i, \lambda_i\}_{i=1}^c$，记为 $\boldsymbol{\theta}$；此外还有负责解码器映射 $\boldsymbol{G}$ 的所有参数（见下一节）。这些参数以及状态变量 $\boldsymbol{z}_t$都需要根据数据 $\boldsymbol{x}_t$ 进行推断。基于概率慢特征分析（SFA）（Turner and Sahani，2007年；Zafeiriou 等，2015年），我们设定 $\sigma^2_i = 2$；$\Re(\lambda_j)$ 和 $\sigma^2_{0,\ i} = 1$。因此，先验上，潜在动力学是平稳的。附录A中给出了这一选择的推导和依据。因此，唯一的独立参数是 $\lambda_i$，其虚部可以表示潜在动力学中的周期效应。

### B. 变分自编码器

我们采用变分自编码器来表征从低维表示 $\boldsymbol{z}_n$ 到高维系统 $\boldsymbol{x}_n$的概率映射。具体而言，我们

具有时间尺度分离的可解释降阶建模

采用概率解码器

$$\boldsymbol{G}(\boldsymbol{z}_n, \boldsymbol{\theta}) = p(\boldsymbol{x}_n \mid \boldsymbol{z}_n, \boldsymbol{\theta}) \quad \text{and} \tag{13}$$

编码器用于根据给定的数据推断状态变量 $\boldsymbol{z}$，其定义见推断与学习部分。

## C. 推断与学习

给定这些概率关系，我们的目标是推断潜变量 $\boldsymbol{z}_{0:T}$ 以及所有模型参数 $\boldsymbol{\theta}$。我们采用一种混合贝叶斯方法，通过摊销变分推断近似状态变量的后验分布，并计算 $\boldsymbol{\theta}$ 的最大后验（MAP）点估计。

对每个数据序列 $\boldsymbol{x}_{0:T}$ 应用贝叶斯定理，可得以下后验分布：

$$p(\boldsymbol{z}_{0:T}, \boldsymbol{\theta} | \boldsymbol{x}_{0:T}) = \frac{p(\boldsymbol{x}_{0:T}|, \boldsymbol{z}_{0:T}, \boldsymbol{\theta})\ p(\boldsymbol{z}_{0:T}, \boldsymbol{\theta})}{p(\boldsymbol{x}_{0:T})} \tag{14}$$

$$= \frac{p(\boldsymbol{X}_{0:T} \mid \boldsymbol{z}_{0:T}, \boldsymbol{\theta})\ p(\boldsymbol{z}_{0:T} \mid \boldsymbol{\theta})\ p(\boldsymbol{\theta})}{p(\boldsymbol{x}_{0:T})} \tag{15}$$

其中，$p(\boldsymbol{\theta})$表示模型参数的先验分布。在变分推断的背景下，我们对近似后验分布采用以下因子分解

$$q_{\boldsymbol{\phi}}(\boldsymbol{z}_{0:T}) = \prod_{n=1}^{T} \prod_{i=0}^{c} q_{\boldsymbol{\phi}}(z_{n,i} \mid \boldsymbol{x}_n) = \prod_{n=1}^{T} \prod_{i=0}^{c} \mathcal{N}(z_{n,i} | \mu_{n,i}(\boldsymbol{x}_n), \sigma_{n,i}(\boldsymbol{x}_n)) \tag{16}$$

也就是说，我们仅根据给定的数据点推断每个状态变量的均值 $\mu$ 和方差 $\sigma$。用于推断的这种条件密度，是上一节定义的概率解码器的编码器对应项。

很容易证明，最优参数值可通过最大化附录 B 中推导出的证据下界（ELBO） $\mathcal{F}(q_{\boldsymbol{\phi}}(\boldsymbol{z}_{0:T})\ \boldsymbol{\theta})$得到。我们利用重参数化技巧（Kingma and Welling，2013）计算 ELBO 关于 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的梯度的蒙特卡罗估计，并使用 ADAM 算法（ Kingma 和 Ba，2014 ）进行随机优化。

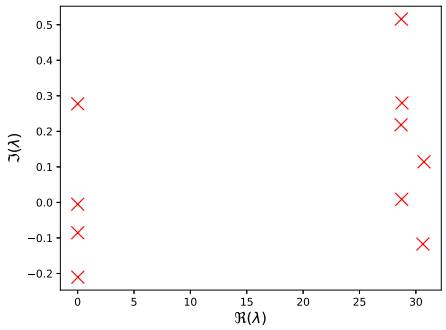



图 6. 使用概率框架学习的 KS 方程系统潜在动力学的 $\lambda$ ：与确定性框架类似，4 个分量是慢变量，负责长期演化

**D. 概率扩展的结果**

我们将概率版本应用于 KS 方程。我们采用了与确定性方法相同的设置，但考虑了最多 10 个复数潜变量。得到的$\lambda$’见图 6。

概率模型使我们能够量化预测中的不确定性。图 7 展示了针对一个未见过的初始条件，在多个时间步长下的预测及相应的不确定性界限。由于 KS 方程的混沌特性以及训练数据量较少，我们模型中的底层线性动力学只能在有限的时间范围内捕捉完整动力学。幸运的是，得益于概率方法，该模型能够捕捉混沌波动，且不确定性界限会逐渐变宽。

我们还根据模型得到的预测结果计算了 KS 方程的相空间表示，并将其与参考解进行比较。概率模型比确定性模型更准确地识别出正确的流形。由于一些小尺度波动被视为噪声，得到的流形在原点附近更加集中，且所得数值略小于参考流形，尽管两者的形状非常相似。

## 五、结论

我们提出了确定性和概率性两种模型阶数约化框架，在潜在空间中采用具有可解释性的线性动力学。与潜变量时间演化相关的参数直接关联系统动力学，并表征增长、衰减以及周期分量。我们将这些模型应用于三个不同的系统；概率版本还能够识别预测不确定性。

目前，其中的可解释动力学是线性的，这对我们的模型构成了限制。因此，我们计划扩展该框架，使其在保留可解释性的同时，也能够表示非线性潜在动力学。

## 数据可用性

支持本研究结论的数据可向通讯作者合理提出请求后获取。

## 利益声明

作者声明不存在任何利益冲突。

## 致谢

我们感谢欧洲高性能计算联合事业（EuroHPC）DComEX 项目（956201-H2020-JTI-EuroHPC-2019-1）资助。

**附录A：附录A**

在本附录中，我们推导转移概率密度以及先验平稳性的必要条件。对于如下形式的多元奥恩斯坦-乌伦贝克过程（OU）：

$$d\boldsymbol{z}_t = -\boldsymbol{A}\boldsymbol{z}_t dt + \boldsymbol{B}d\boldsymbol{W}_t \tag{A1}$$

其中 $\boldsymbol{A}$， $\boldsymbol{B}$ 是矩阵， $\boldsymbol{W}_t$ 是多元维纳过程，我们有：

$$\mathbb{E}[\boldsymbol{z}_t] = \boldsymbol{\mu}_t = e^{-\boldsymbol{A}t}\boldsymbol{\mu}_0 \tag{A2}$$

其中 $\mathbb{E}[\boldsymbol{z}_0] = \boldsymbol{\mu}_0$。平稳协方差矩阵 $\boldsymbol{\Sigma}_\infty$ 满足李雅普诺夫方程（Gardiner 等，1985）：

$$\boldsymbol{A}\boldsymbol{\Sigma}_\infty + \boldsymbol{\Sigma}_\infty\boldsymbol{A}^T = \boldsymbol{B}\boldsymbol{B}^T \tag{A3}$$

而在二维情况下，其表达式为（Gardiner 等，1985）：

$$\boldsymbol{\Sigma}_\infty = \frac{det(\boldsymbol{A})\boldsymbol{B}\boldsymbol{B}^T + (\boldsymbol{A} - tr(\boldsymbol{A})\boldsymbol{I})\boldsymbol{B}\boldsymbol{B}^T(\boldsymbol{A} - tr(\boldsymbol{A})\boldsymbol{I})^T}{2tr(\boldsymbol{A})det(\boldsymbol{A})} \tag{A4}$$

最后，平稳自相关由下式给出（Gardiner等人，1985）：

$$\mathbb{E}[\boldsymbol{z}_{t+\Delta t}\boldsymbol{z}_t^T] = e^{-\boldsymbol{A}\Delta t}\boldsymbol{\Sigma}_\infty \tag{A5}$$

我们考虑如下演化的复OU过程 $\boldsymbol{z}_t = z_{R,\,t} + iz_{I,\,t}$：

$$d\boldsymbol{z}_t = -\boldsymbol{\lambda}\boldsymbol{z}_t\, dt + \frac{\sigma}{\sqrt{2}}d\boldsymbol{W}_t \tag{A6}$$

其中 $\boldsymbol{\lambda} = \lambda_R + i\lambda_I$ 和 $d\boldsymbol{W}_t = dW_{R,\,t} + idW_{I,\,t}$ 满足 $W_{R,\,t}$， $W_{I,\,t}$ 是相互独立的维纳过程。我们进一步假设，在 $t = 0$ $\boldsymbol{z}_0 \sim \mathcal{CN}(0,\ \sigma_0^2)$ 处，即 $z_{R,\,0}$， $z_{I,\,0}$ 是相互独立的均值为零且方差为 $\sigma_0^2/2$ 的正态变量。若将 $\boldsymbol{z}_t$ 的实部和虚部排列成向量 $\boldsymbol{k}_t = \begin{bmatrix} z_{R,\,t} \\ z_{I,\,t} \end{bmatrix}$，则

这个二维向量的演化由以下 OU 过程决定：

- $\boldsymbol{A} = \begin{bmatrix} \lambda_R & -\lambda_I \\ \lambda_I & \lambda_R \end{bmatrix} \longrightarrow tr(\boldsymbol{A}) = 2\lambda_R,\quad det(\boldsymbol{A}) = |\boldsymbol{\lambda}|^2$
- $\boldsymbol{B} = \frac{\sigma}{\sqrt{2}}\boldsymbol{I} \longrightarrow \boldsymbol{B}\boldsymbol{B}^T = \frac{\sigma^2}{2}\boldsymbol{I}$

具有时间尺度分离的可解释降阶建模

根据.....可知:

$$\boldsymbol{\mu}_t = \mathbf{0}. \tag{A7}$$

根据.....可知:

$$\boldsymbol{\Sigma}_\infty = \frac{\sigma^2}{4\lambda_R}\boldsymbol{I} \tag{A8}$$

它满足.....。根据.....的平稳自相关特性以及高斯向量的性质，我们可以推导出平稳转移概率密度 $p(\boldsymbol{k}_{t+\Delta t}|\boldsymbol{k}_t)$，它同样服从高斯分布，即:

$$\boldsymbol{k}_{t+\Delta t}|\boldsymbol{k}_t \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_{\Delta t}, \bar{\boldsymbol{\Sigma}}_{\Delta t}) \tag{A9}$$

其中:

$$\bar{\boldsymbol{\Sigma}}_{\Delta t} = \boldsymbol{\Sigma}_\infty - e^{-\boldsymbol{A}\Delta t}\boldsymbol{\Sigma}_\infty\boldsymbol{\Sigma}_\infty^{-1}\boldsymbol{\Sigma}_\infty e^{-\boldsymbol{A}^T\Delta t} = \frac{\sigma^2}{4\lambda_R}(1 - e^{-2\lambda_R\Delta t})\boldsymbol{I} \tag{A10}$$

and:

$$\bar{\boldsymbol{\mu}}_{\Delta t} = \mathbf{0} + e^{-\boldsymbol{A}\Delta t}\boldsymbol{\Sigma}_\infty\boldsymbol{\Sigma}_\infty^{-1}(\boldsymbol{k}_t - \mathbf{0}) = e^{-\boldsymbol{A}\Delta t}\boldsymbol{k}_t = e^{-\lambda_R\Delta t}\begin{bmatrix}\cos(\lambda_I\Delta t) & \sin(\lambda_I\Delta t)\\ & \cos(\lambda_I\Delta t)\end{bmatrix}\boldsymbol{k}_t \tag{A11}$$

因此，要获得具有协方差 $\boldsymbol{\Sigma}_\infty = \frac{1}{2}\boldsymbol{I}$ 的平稳过程 $\boldsymbol{x}_t$，必须采用以下参数取值:

- $\sigma^2 = 2\lambda_R$
- $\sigma_0^2 = 1$

**附录 B: 附录 B**

本节详细推导证据下界（ELBO），该下界在训练过程中作为确定参数 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$的目标函数。具体而言，对于给定的 $N$ 时间序列 $\boldsymbol{x}$:

$$\begin{aligned}
&\log p(\boldsymbol{x}_{0:T}^{(1:N)}|\boldsymbol{\theta})\\
&= \log \int p(\boldsymbol{x}_{0:T}^{(1:N)}, \boldsymbol{z}_{0:T}^{(1:N)}, \boldsymbol{\theta})\ \ \boldsymbol{z}_{0:T}^{(1:N)}\\
&= \log \int \frac{p(\boldsymbol{x}_{0:T}^{(1:N)}|\boldsymbol{z}_{0:T}^{(1:N)}, \boldsymbol{\theta})p(\boldsymbol{z}_{0:T}^{(1:N)}, \boldsymbol{\theta})}{q_{\boldsymbol{\phi}}(\boldsymbol{z}_{0:T}^{(1:N)})} q_{\boldsymbol{\phi}}(\boldsymbol{z}_{0:T}^{(1:N)})\ d\boldsymbol{z}_{0:T}^{(1:N)}\\
&\geq \int \log \frac{p(\boldsymbol{x}_{0:T}^{(1:N)}|\boldsymbol{z}_{0:T}^{(1:N)}, \boldsymbol{\theta})p(\boldsymbol{z}_{0:T}^{(1:N)}, \boldsymbol{\theta})}{q_{\boldsymbol{\phi}}(\boldsymbol{z}_{0:T}^{(1:N)})} q_{\boldsymbol{\phi}}(\boldsymbol{z}_{0:T}^{(1:N)})\ d\boldsymbol{z}_{0:T}^{(1:N)}\\
&= \mathcal{F}(q_{\boldsymbol{\phi}}(\boldsymbol{z}_{0:T}^{(1:N)}), \boldsymbol{\theta})
\end{aligned} \tag{B1}$$

**参考文献**

Alber, M., Buganza Tepole, A., Cannon, W. R., De, S., Dura-Bernal, S., Garikipati, K., Karniadakis, G., Lytton, W. W., Perdikaris, P., Petzold, L., 等，“整合机器学习与多尺度建模——生物学、生物医学和行为科学领域的观点、挑战与机遇”，NPJ数字医学**2**, 115 (2019). Brunton, S. L., Proctor, J. L., and Kutz, J. N., “通过稀疏辨识非线性动力系统，从数据中发现控制方程”，美国国家科学院院刊 **113**, 3932–3937 (2016). Budiˇsi´c, M., Mohr, R., and Mezi´c, I., “应用Koopman理论”，《混沌：非线性科学跨学科期刊》**22**, 047510 (2012). Champion, K., Lusch, B., Kutz, J. N., and Brunton, S. L., “数据驱动地发现坐标与控制方程”，美国国家科学院院刊**116**, 22445–22451 (2019). Chen, T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K., “神经常微分方程”，载于神经信息处理系统进展 (2018) pp. 6571–6583. Cranmer, M., Sanchez Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., and Ho, S., “利用归纳偏置，通过深度学习发现符号模型”，神经信息处理系统进展**33**, 17429–17442 (2020). Gardiner, C. W.等，《随机方法手册》，第3卷（施普林格·柏林，1985）。Givon, D., Kupferman, R., and Stuart, A., “提取宏观动力学：模型问题与算法”，《非线性》(2004). Hochreiter, S. and Schmidhuber, J., “长短期记忆”，《神经计算》**9**, 1735–1780 (1997). Horenko, I., “关于同时进行基于数据的降维与隐含相位识别”，《大气科学杂志》, 1941–1954 (2008), 出版方：美国气象学会，栏目：《大气科学杂志》。Horenko, I., Klein, R., Dolaptchiev, S., and Sch¨utte, C., “自动生成降阶随机天气模型 I：时间序列分析中的同步维度与模型降阶”，《多尺度建模与仿真》**6**, 1125–1145 (2008), 出版方：工业与应用数学学会。

采用时间尺度分离的可解释降阶建模 Kaltenbach, S. and Koutsourelakis, P.-S., “将物理约束纳入用于动力系统粗粒化的深度概率机器学习框架”，计算物理学杂志 **419**, 109673 (2020). Kaltenbach, S. and Koutsourelakis, P.-S., “具备物理感知能力且保证稳定性的概率模型降阶”，arXiv预印本 arXiv:2101.05834 (2021). Kassam, A.-K. and Trefethen, L. N., “刚性偏微分方程的四阶时间步进”，SIAM科学计算杂志 **26**, 1214–1233 (2005). Kingma, D. P. and Ba, J., “Adam：一种随机优化方法”，arXiv预印本 arXiv:1412.6980 (2014). Kingma, D. P. and Welling, M., “自编码变分贝叶斯”，arXiv预印本 arXiv:1312.6114 (2013). Klus, S., N¨uske, F., Koltai, P., Wu, H., Kevrekidis, I., Sch¨utte, C., and No´e, F., “数据驱动的模型降阶与传递算子近似”，非线性科学杂志 **28**, 985–1010 (2018). Koopman, B. O., “希尔伯特空间中的哈密顿系统与变换”，美国国家科学院院刊**17**, 315–318 (1931). Li, X., Wong, T.-K. L., Chen, R. T., and Duvenaud, D., “随机微分方程的可扩展梯度”，arXiv预印本 arXiv:2001.01328 (2020). Mezi´c, I., “通过Koopman算子的谱性质分析流体流动”，流体力学年度评论**45**, 357–378 (2013). No´e, F., “长时间尺度上的分子动力学机器学习”，载于机器学习与量子物理学，物理学讲义，K. T. Sch¨utt, S. Chmiela, O. A. von Lilienfeld, A. Tkatchenko, K. Tsuda, and K.-R. M¨uller 编（Springer国际出版公司，Cham，2020）第331–372页。 Pan, S. and Duraisamy, K., “以物理信息概率学习对非线性动力学进行线性嵌入并保证稳定性”，SIAM应用动力系统杂志 **19**, 480–509 (2020). Robinson, J. C., “Kuramoto-Sivashinsky方程的惯性流形”，物理学快报A **184**, 190–193 (1994). Rowley, C. W., Mezi´c, I., Bagheri, S., Schlatter, P., and Henningson, D. S., “非线性流动的谱分析”，流体力学杂志**641**, 115–127 (2009).

采用时间尺度分离的可解释降阶建模 Snyder, W., Mou, C., Liu, H., San, O., De Vita, R., and Iliescu, T., “降阶模型闭合：简要教程”，arXiv:2202.14017 [physics] (2022), arXiv: 2202.14017. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., “Dropout：一种防止神经网络过拟合的简单方法”，机器学习研究杂志**15**, 1929–1958 (2014). Turner, R. and Sahani, M., “慢特征分析的最大似然解释”，《神经计算》**19**, 1022–1038 (2007). Vlachas, P. R., Arampatzis, G., Uhler, C., and Koumoutsakos, P., “通过学习有效动力学模拟复杂系统的多尺度行为”，自然·机器智能**4**, 359–366 (2022). Williams, M. O., Kevrekidis, I. G., and Rowley, C. W., “Koopman算子的数据驱动近似：扩展动态模态分解”，非线性科学杂志**25**, 1307–1346 (2015). Zafeiriou, L., Nicolaou, M. A., Zafeiriou, S., Nikitidis, S., and Pantic, M., “用于行为分析的概率慢特征”，IEEE神经网络与学习系统汇刊**27**, 1034–1048 (2015).

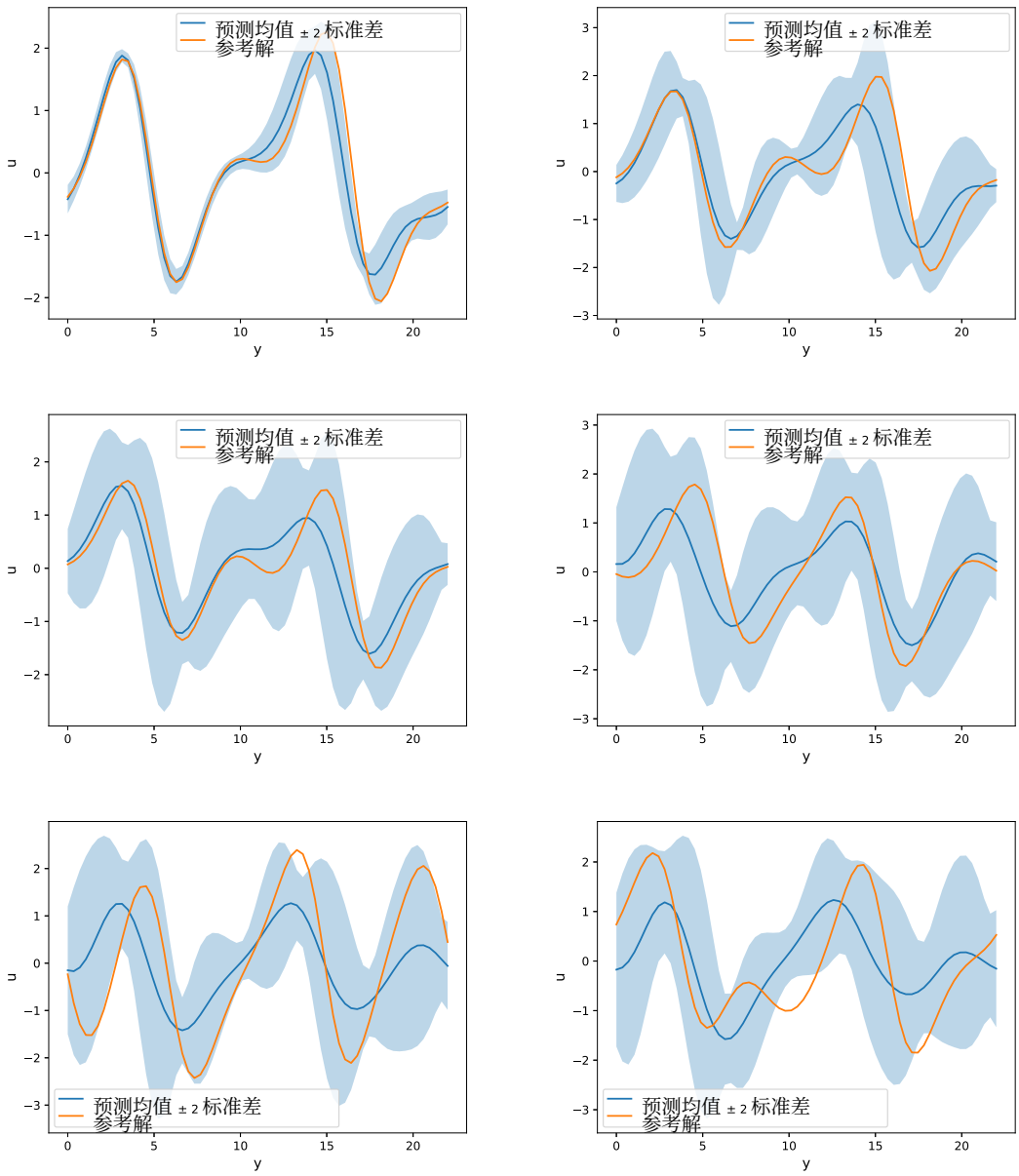



图 7. 针对新的初始条件，对预测结果与参考解进行比较，时间分别为$t = 1.25$，3。75， 7。5，12。5， 20， 30（从左到右、从上到下）。我们注意到，预测时间越长，不确定性界限越大。尽管 KS 方程具有混沌特性，但预测后验均值仍接近参考解 $t \leq 12.5$

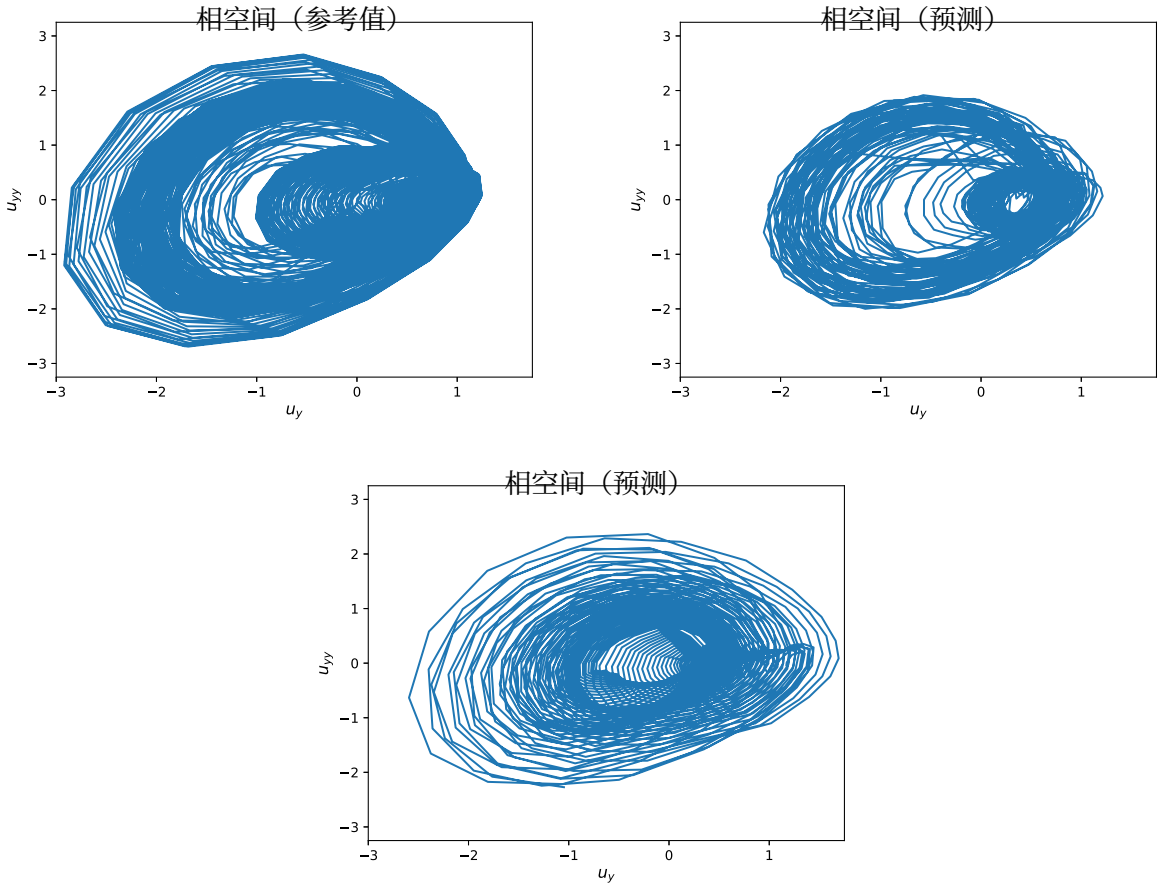



图 8. 参考解的相空间（左上）、概率模型预测的后验均值（右上）与确定性模型预测的相空间（下）的比较