# **IHCV:** 非线性动力系统中隐藏的时变控制变量的发现

胡安·穆尼奥斯[1], 苏巴什·巴尔萨米[1], 胡安·P. 贝尔纳尔-塔马约[1], 阿里·巴鲁拜德[1], 阿尔贝托·马伊略·鲁伊斯·德·因凡特[1], 文森佐·拉加尼[1,4,5], 大卫·戈麦斯-卡夫雷罗[1,3], 纳尔西斯·A. 基亚尼[6,7], 和耶斯佩尔·泰格纳[1,2,7,8,*]

[1] 沙特阿拉伯图瓦勒23955-6900，阿卜杜拉国王科技大学（KAUST）生物与环境科学

[2] 计算机、电气与数学科学及工程学部，阿卜杜拉国王科技大学（KAUST），图瓦尔

[3] 转化生物信息学部，纳瓦拉生物医学研究中心（Navarrabiomed），纳瓦拉医院综合体（CHN），纳瓦拉公立大学（UPNA），纳瓦拉健康研究所（IdiSNA），潘普洛纳，西班牙。

[4] SDAIA-阿卜杜拉国王科技大学数据科学与人工智能卓越中心，图瓦尔 23952，沙特阿拉伯

[5] 伊利亚州立大学化学生物学研究所，第比利斯 0162，格鲁吉亚

[6] 算法动态实验室，卡罗琳斯卡学院肿瘤学与病理学系，斯德哥尔摩，瑞典。

[7] 计算医学组，卡罗琳斯卡医学院医学系，分子医学中心，卡罗琳斯卡大学医院，L8:05，SE-171 76，斯德哥尔摩，瑞典

[8] 生命科学实验室，Tomtebodavagen 23A，SE-17165，索尔纳，瑞典

[*] 通信地址：jesper.tegner@kaust.edu.sa

## 摘要

从数据中发现非线性动力学模型是科学的核心。近期进展依赖于利用丰富的候选函数库对可观测量进行稀疏回归。然而，对支配不同动力学状态之间切换的隐藏、不可观测控制变量进行建模，仍然颇具挑战。

本文提出了一种数据高效的无导数方法——IHCV，用于识别隐藏控制变量。首先，通过使用众所周知的分岔模型（鞍结分岔、跨临界分岔、叉式分岔、Hopf分岔）对IHCV方法进行基准测试，评估IHCV在噪声下的性能和稳健性。接下来，我们展示了模型中发现隐藏的驱动变量。最后，我们使用toggle switch模型、遗传抑制子振荡器和Waddington景观模型对IHCV进行演示，说明IHCV可以推广到仅提供部分观测数据的情况。

我们的原理验证表明，利用正规形有助于以数据高效且可扩展的方式，发现控制不同动力学状态与非线性模型之间转变的隐变量。

科学与机器学习界正日益联合起来，利用数据驱动技术加速并推动科学发现[1]。这种追求超越了数据分析和模式检测。在科学领域，我们需要回答关于如何控制和引导系统在不同动力学状态之间转换的假设性问题 [2,3]。

总体而言，迄今为止的大多数研究都依赖强假设（这些假设可能并不现实），以便根据观测结果识别模型[4,5]。这些研究包括线性建模，以及假设采用非零系数较少的稀疏线性模型；然而，自然现象的动力学种类繁多，因此需要使用非线性微分方程模型[6]。此外，若干方法假设可以获得大量观测训练数据、实验数据或合成数据，而且通常还需要扰动数据[7,8]。然而，在研究自然系统时，无论是否施加扰动，我们都很少能找到如此丰富的观测数据。相反，人们采用函数库，并结合稀疏性或简约性等约束，以应对解码非线性系统这一挑战[5,9]。然而，这些方法要么随着基函数数量增加而难以扩展，要么要求作出强假设，即仅使用少数几项的组合来表示数据。

一个非线性动力系统可以轻易展现复杂的切换动力学，在不动点、振荡或混沌动力学之间交替[10–12]。然而，发现支配如此丰富的动力学特征的核心控制机制仍然具有挑战性。此外，我们不能假定能够获取所有相关变量来进行此类识别。例如，数据通常因测量的稀疏性而不完整，我们甚至可能不知道属于所关注物理系统的未知变量。这一表述在地球科学/气候建模、神经科学、基因组学、干细胞生物学和发育生物学[10](图1A)中具有广泛应用。在这些领域，我们通常无法获得由第一性原理构建的基础模型。因此，模型发现问题与这一挑战内在相关

以及存在 此类隐变量 [13,14]。在此，我们开发了一种方法 IHCV，用于从非线性动力系统的时间观测中识别隐藏控制变量，而无需扰动数据。我们的核心思想是使用动力系统理论中的正规形作为最小模型构建模块 [15]。因此，我们将不同动力学状态之间的转变建模为系统动力学根据未观测控制变量而演化的过程。

受动力系统理论[15]的启发，我们使用正规形作为基本的模型构建模块[16,17]，来解决复杂动力系统及其在不同状态之间转变的建模问题。为了评估这一概念的可行性，我们使用叉式分岔、跨临界分岔、鞍结点和霍普夫分岔作为原理验证测试模型（图1B）。举例来说，假设我们观测并测量一个且仅一个变量（x），并发现它在某个时间点开始随时间增长（图1C）。因此，存在另一个我们无法观测的过程，驱动了x的增加，或者等价地说，使x的不动点值发生了偏移。我们使用鞍结正规形$\dot{x} = -\alpha + x^2$对观测结果进行建模。重要的是，参数α 是隐藏的，并且是时间的函数（α(t)）。我们的任务是根据观测到的数据（x）重构隐藏变量α 的时间动力学。为了解决这一问题，我们近似认为系统接近一个稳定流形。这意味着控制变量的动力学足够缓慢，从而由此诱导的x的偏移使系统能够在其稳定流形附近演化。因此，这将问题简化为在观测时间区间内寻找α 的多项式表示。在这里，我们对α 进行建模，使得$\alpha(t) = 在 + at^2 + bt^3$ ct，并求解权重a、b和c。这使我们能够重构alpha，足以解释所观测到的x的动力学（图1C）。接下来，我们询问这一原理验证是否能够推广到更广泛的模型类别。为此，我们使用正规形、经典动力系统等模型对我们的方法进行了评估和基准测试，

神经系统，以及种群生物学、合成生物学、干细胞生物学和发育生物学中的模型（图 1A）。

首先，我们通过考察不同动力学状态之间的转变来评估 IHCV，例如标准形模型中两个不同不动点之间的转变。我们生成了观测时间序列数据，其中低稳态在一段时间后切换至高稳态水平（图 2a，左）。IHCV方法利用叉形模型（图 2a，中）求解驱动这一跃迁的隐藏驱动变量的参数化表示（图 2a，右）。使用鞍结模型则求解了不动点增长的情形（图 2b）。接下来，我们探讨了振荡失稳、系统中出现不动点的情形（图 2c，左）。利用 Hopf标准形模型（图 2c，中），我们的方法能够求解隐藏驱动变量的参数化表示（图 2c，右）。同样，我们也能求解跨临界分岔和叉形分岔对应的情形（补充材料 1）。

动力系统文献已广泛研究 Lorenz 和 vanderPol模型[15,18]。在这里，我们采取了不同的路径，探讨能否从观测数据中识别出一个隐藏变量，该变量在这些系统中以时变方式改变各自振荡行为的频率和振幅。对于Lorenz吸引子，我们生成了振幅和频率随时间平滑增加的数据（图 2d，左）。随着时间推移，我们减慢了 van der Pol 的弛豫振荡器（图2e，左）。在这两种情况下，结合观测数据，使用相应的模型（图2d、e，中）我们都能够识别出隐藏变量的时变动力学（图 2d、e，右）。

Hodgkin-Huxley 模型是用于对生物神经网络中的节点动力学进行建模的经典模型 [19]。在这里，我们将外部电流注入用作隐藏控制变量，

并且我们能够从观测数据（图 2f，左）中重构电流注入的隐藏动力学（图 2f，右）。

接下来，我们使用经典的二维约化 Fitzhugh-Nagumo [19] 神经（FHN）模型（图 2f）验证了这些观察结果（补充材料2）。在此，IHCV方法面临的挑战是：发现使FHN模型中的不动点失稳的隐藏驱动，随后导致松弛振荡出现，并返回到另一个增长的不动点。最后，我们还使用由用于种群动力学的Lotka-Volterra捕食者-猎物模型 [18] 产生的不同动力学状态的数据，进一步验证了IHCV方法（补充材料3）。

在所有这些概念验证演示中，我们都恢复了隐藏变量的时间动力学。接下来，我们询问 IHCV 能否推广到更具挑战性的情形：无法获取特定领域的模型方程。这一问题与合成生物学、干细胞生物学和发育生物学相关（图 1A）。首先，我们使用遗传学拨动开关模型[20],，该模型描述了一个耦合双基因回路中的两种稳定基因表达状态。这个经过工程设计的合成生物学基因回路模拟了这样一种实验：实验诱导使系统从一种稳定状态切换到另一种稳定状态。作为 IHCV 的输入（图 2h，左），我们使用一个较高的稳态，该状态经过一段时间后切换到低态。即使无法获取模型方程，IHCV 也成功重建了隐藏驱动变量的定性动力学（图 2h，右）。值得注意的是，虽然我们不假设能够获取“真实”的模型方程，但拟合的均方误差并不为零。尽管如此，IHCV 仍捕捉到了隐藏变量的单调递增趋势。合成的阻遏振荡器回路[21] 由三个相互连接、能够产生振荡系统的基因组成。我们驱动该回路从一个不动点进入频率不断增加的振荡模式。仅利用这些观测数据（图 2i，左），在无法获取模型方程的情况下，IHCV 识别出了隐藏控制变量的定性动力学（图 2i，右）。

细胞重编程是干细胞生物学领域的一项重要发现[22]。干细胞具有多能性，可以转化为其他不同类型的细胞。从这个意义上说，细胞处于发育景观中的不同位置，这种景观在技术上称为 Waddington 景观[23]。这类景观也被用于解释不同细胞类型和器官的发育过程，最终形成一个生物体。多种因素（其中一部分尚不明确）共同促使细胞沿着这样的 Waddington 景观逐步发展[10,24]。在这里，我们探讨该方法能否通过观察系统从一种稳定状态转变为两种新的稳定状态（图 2h，左），揭示这样的隐藏驱动因素。使用超临界叉式分岔模型（图 2j，中），IHCV 恢复了隐藏驱动变量的时间动力学（图 2j，右）。

我们在此开发了一种数据高效的无导数模型发现技术，用于识别具有未知动力学的隐藏控制变量，并在不同动力学状态之间引导系统。我们使用正规形作为通用且可扩展的模型构建模块，以捕捉慢流形附近的样本观测数据。动力学系统的建模问题被转化为选择在稳定流形上具有相似轨迹的分岔正规形。此类稳定流形可利用分岔参数的一种便利动力学进行变换。换言之，仅通过调控动态分岔参数，我们就能基于单一系统生成丰富的动力学。该技术适用于众多模型，从作为原理验证的正规形（包括Lorenz模型和van der Pol模型），到种群生物学、神经科学、遗传学和干细胞生物学中的动力学系统模型。未来工作包括将这一原理验证扩展到更大规模的系统，以实现大规模机制建模[25]。

图注

图1：IHCV方法：a）适用范围包括气候建模、神经科学、基因组学和干细胞生物学。b）隐藏驱动变量（蓝线）在不同动力学状态之间切换系统动力学，分别由叉式分岔系统（上方红线）和Hopf分岔系统（下方红线）示意。c）使用鞍结分岔对IHCV方法的说明。上图：隐藏驱动变量α （x轴）控制观测状态变量（y轴）。下图：采集观测状态变量随时间变化的数据（左）足以通过多项式回归恢复α 的动力学（右），求解权重a、b和c，使得$\alpha(t) = at + bt^2 + ct^3$。

图2. 从观测中学习隐藏控制。定性动力学（第一列），第二列显示原始观测数据点。生成模型（第三列）预测隐藏变量的动力学（第四列）。蓝色虚线标示学习得到的动力学，红线则表示实际动力学。前三个模型（a-c）基于正规形理论，包括叉式分岔、鞍结分岔和霍普夫分岔。洛伦兹与范德波尔物理模型（d-e）。其余两个来自神经科学的模型（f-g）为菲茨休-南云模型和霍奇金-赫胥黎模型。最后，在最后三个模型（h-j）中，当我们无法获得基本方程时，利用分岔正规形库预测模型的定性动力学。尽管h-j在第4列中实际值与学习值之间存在差异，正规形仍能充分刻画系统的动力学。

## 材料与方法

方法已上传至单独的文件中。此外，数据集、模型和代码已共享在专用的 GitHub 页面 https://github.com/munozdjp/IHCV，其中包括已演算的数据示例。

## 数据可用性

所有数据和软件均可公开访问，具体信息详见“材料与方法”部分。

## 资助

阿卜杜拉国王科技大学为本研究提供了支持。

## 利益冲突

作者声明不存在利益冲突。

## 参考文献

1. Karniadakis, G. E. 等. 物理信息机器学习。《自然综述：物理学》3, 422–440 (2021). 2. Li, Z., Ji, J. 与 Zhang, Y. 从开普勒到牛顿：用于科学发现的可解释人工智能。载于《国际机器学习会议》(2022). 3. Sapoval, N. 等. 深度学习在生物科学各领域的应用：当前进展与开放挑战。《自然通讯》13, (2022). 4. Schmidt, M. 与 Lipson, H. 从实验数据中提炼自由形式的自然定律。《科学》(1979) 324, 81–85 (2009). 5. Brunton, S. L., Proctor, J. L., Kutz, J. N. 与 Bialek, W. 通过非线性动力系统的稀疏辨识，从数据中发现控制方程。《美国国家科学院院刊》113, 3932–3937 (2016). 6. Tegnér, J., Yeung, M. K. S., Hasty, J. 与 Collins, J. J. 基因网络的逆向工程：整合遗传扰动与动力学建模。《美国国家科学院院刊》5月13日, 5944 5949 2003– ( ). 7. Tegnér, J. 与 Björkegren, J. 通过扰动揭示基因网络。《遗传学趋势》卷 23 34–41 预印本，网址：https://doi.org/10.1016/j.tig.2006.11.003 (2007). 8. Peña, J. M., Björkegren, J. 与 Tegnér, J. 在忠实性假设下，对马尔可夫边界进行可扩展、高效且正确的学习。载于《符号与定量不确定性推理方法》。ECSQARU

2005. 《计算机科学讲义》(), 卷 3571. 施普林格，柏林，海德堡。 https://doi.org/10.1007/*11518655_13* (编者. 载于：Godo, L. (编者) 不确定性的符号与定量方法。E. 2005.) (2005). 9. Gustafsson, M., Hörnquist, M., Lundström, J., Björkegren, J. 与 Tegnér, J. 使用最小绝对收缩和选择算子及非线性基函数对基因网络进行逆向工程。《纽约科学院年刊》1158, 265–275 (2009). 10. Moris, N., Pina, C. 与 Arias, A. M. 表观遗传景观中的过渡态与细胞命运决定。《自然综述：遗传学》卷 17 693–703 预印本，网址：https://doi.org/10.1038/nrg.2016.98 (2016). 11. Zhou, J. X. 与 Huang, S. 理解细胞命运分支点处的基因电路，以实现理性的细胞重编程。《遗传学趋势》卷 27 55–62 预印本，网址：https://doi.org/10.1016/j.tig.2010.11.002 (2011). 12. Teschendorff, A. E. 与 Feinberg, A. P. 统计力学与单细胞生物学相遇。《自然综述：遗传学》(2021) doi:10.1038/s41576-021-00341-z. 13. Battistin, C., Dunn, B. 与 Roudi, Y. 与未知因素共学：在存在隐变量的情况下分析生物学数据。《系统生物学当前观点》1, 122–128 (2017). 14. Agrawal, R., Squires, C., Prasad, N. 与 Uhler, C. DeCAMFounder：存在隐变量时的非线性因果发现。https://arxiv.org/abs/2102.07921 (2021). 15. Wiggins, S. 《应用非线性动力系统与混沌导论》。(2003). 16. Yair, O., Talmon, R., Coifman, R. R. 与 Kevrekidis, I. G. 通过从数据中学习信息充分的观测几何来重构正规形。《美国国家科学院院刊》114, E7865–E7874 (2017). 17. Kalia, M., Brunton, S. L., Meijer, H. G. E., Brune, C. 与 Kutz, J. N. 学习正规形自编码器，以数据驱动方式发现通用的、依赖参数的控制方程。https://arxiv.org/abs/2106.05102 (2021). 18. Strogatz, S. 《非线性动力学与混沌》。(CRC出版社, 2015). 19. Catterall, W. A., Raman, I. M., Robinson, H. P. C., Sejnowski, T. J. 与 Paulsen, O. Hodgkin-Huxley的传承：从通道到电路。《神经科学杂志》32, 14064 14073 2012– ( ). 20. Gardner, T. S., Cantor, C. R. 与 Collins, J. J. 在大肠杆菌中构建遗传学拨动开关。《自然》卷 403 www.nature.com (2000). 21. Elowiotz 与 Leibler. 转录调节因子构成的合成振荡网络。《自然》(2000). 22. Takahashi, K. 与 Yamanaka, S. 通过确定因子从小鼠胚胎及成年成纤维细胞培养物中诱导多能干细胞。《细胞》126, 663–676 (2006). 23. Ladewig, J., Koch, P. 与 Brüstle, O. 拉平Waddington景观：直接编程的兴起与细胞命运层级的消失。《自然综述：分子细胞生物学》卷 14 225–236 预印本，网址：https://doi.org/10.1038/nrm3543 (2013). 24. Ferrell, J. E. 双稳态、分岔与Waddington表观遗传景观。《当代生物学》卷 22 预印本，网址：https://doi.org/10.1016/j.cub.2012.03.045 (2012). 25. Tegnér, J. N. 等. 计算疾病建模——事实还是虚构？《BMC系统生物学》3, (2009).

a)
Climate
Neuroscience
Genomics
Caenorhabditis elegans
Stem Cell Biology
Lorentz Atractor
Neural Oscilattor
FitzHugh Nagumo
Hudgking Huxley
Cellular level
Gene expression
Cell growth
Genetic Toggle Switch
Genetic Oscillator
Developmental Biology
Waddington Landscape
c)
Saddle Node
State Variable
Alpha
图1
b)
Hidden Variable
State Variable
Time
Pitchfork
State Variable
Alpha
Hopf
State Variable
Alpha
Saddle Node $\dot{x} = -\alpha(t) + x^2$
$x$
Observation x
Stable Manifold
Time
Reconstruct $\alpha(t)$
Assumption: $\dot{x} \approx 0$
$0 \approx -\alpha(t) + x^2$
$\alpha(t) \approx x^2$
Hidden variable discovery
$\alpha(t)$
Prediction
Ground Truth
$\alpha(t) = at + bt^2 + ct^3$
Time

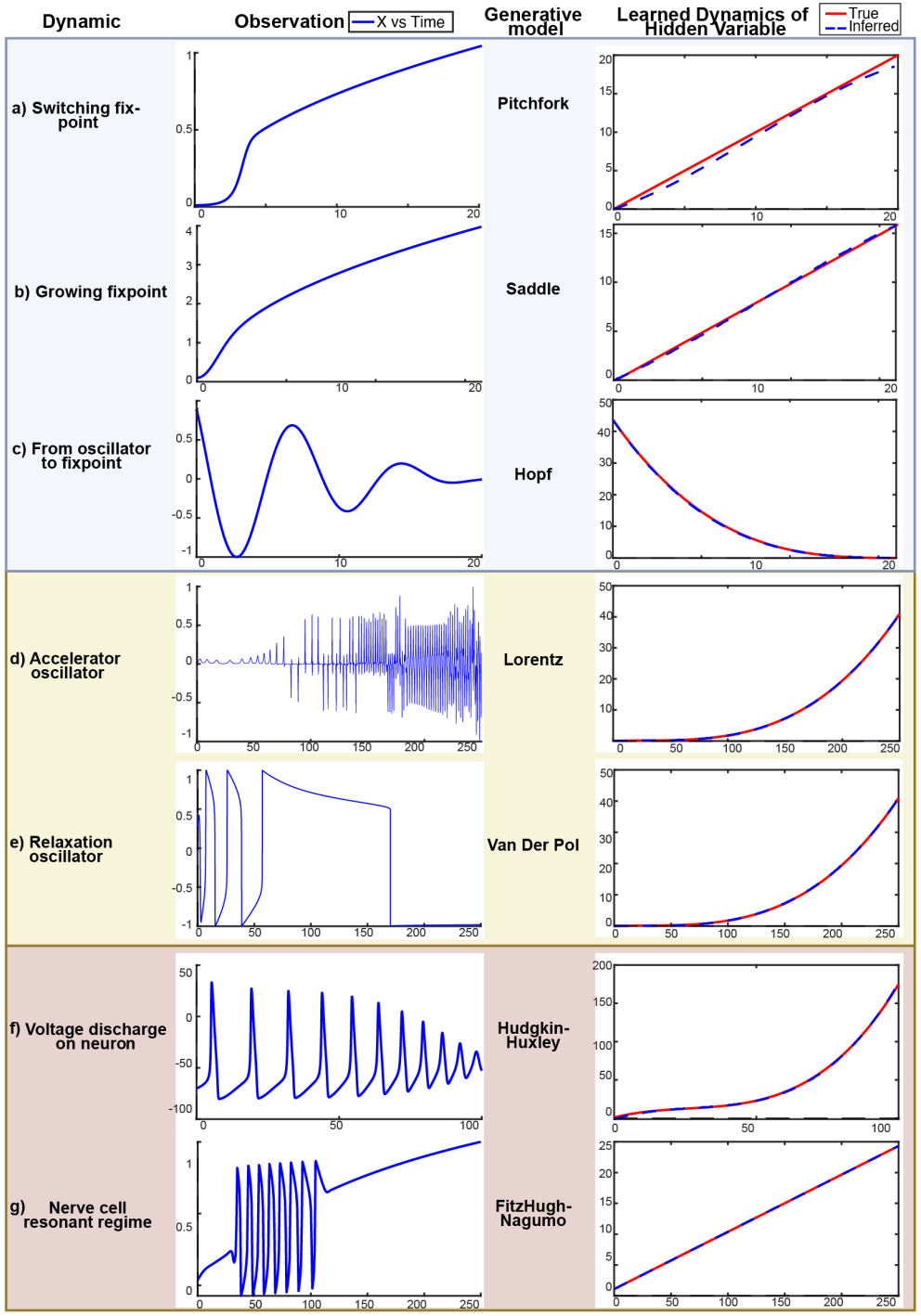
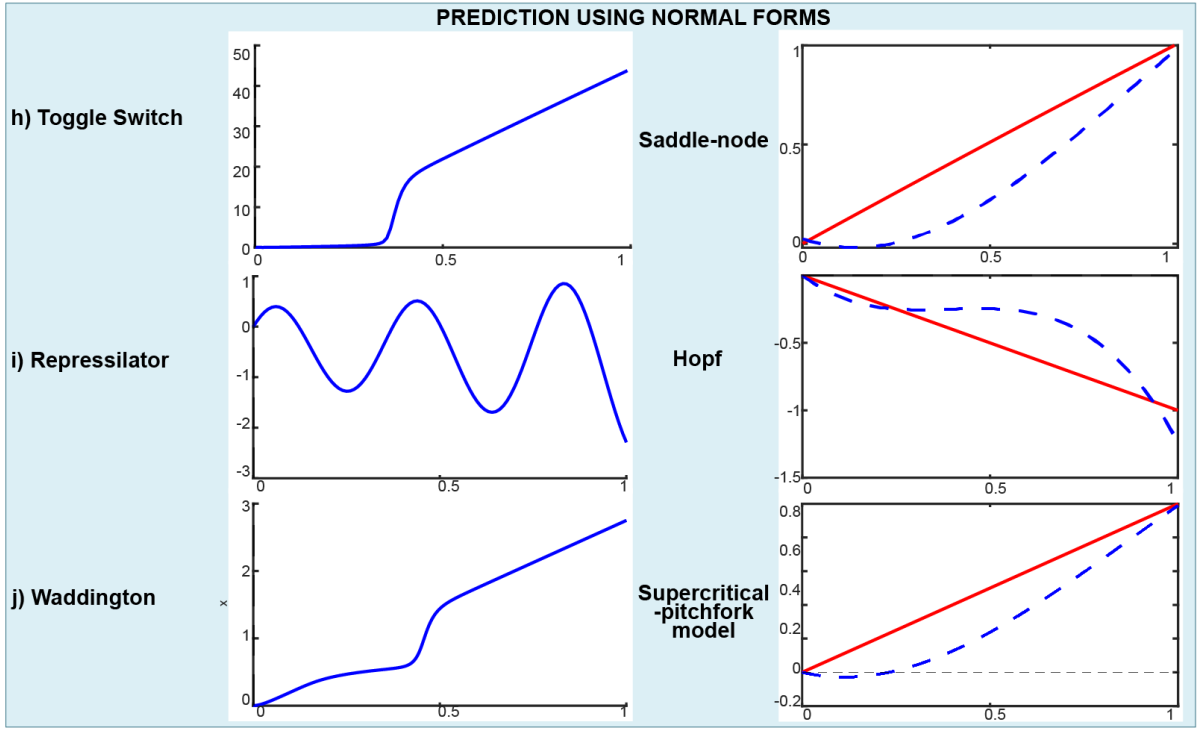

Dynamic
Observation
X vs Time
Generative model
Learned Dynamics of Hidden Variable
True
Inferred
a) Switching fix-point
b) Growing fixpoint
c) From oscillator to fixpoint
d) Accelerator oscillator
e) Relaxation oscillator
f) Voltage discharge on neuron
g) Nerve cell resonant regime
Pitchfork
Saddle
Hopf
Lorentz
Van Der Pol
Hudgkin-Huxley
FitzHugh-Nagumo

图2


PREDICTION USING NORMAL FORMS
h) Toggle Switch
i) Repressilator
j) Waddington
Saddle-node
Hopf
Supercritical -pitchfork model