Recurrent neural networks for spatiotemporal prediction of chaotic dynamics

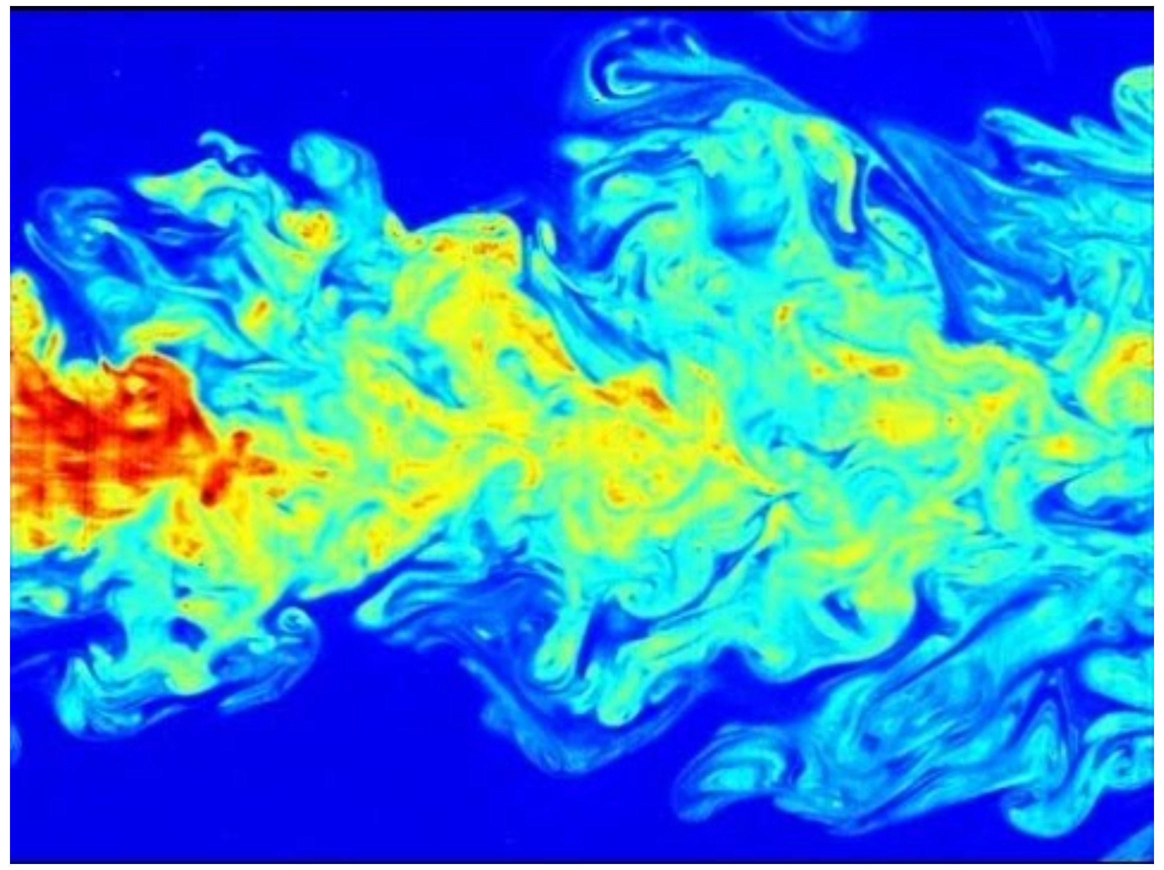
Vlachas Pantelis

Computational **S**cience and **E**ngineering **Lab**
ETH Zürich

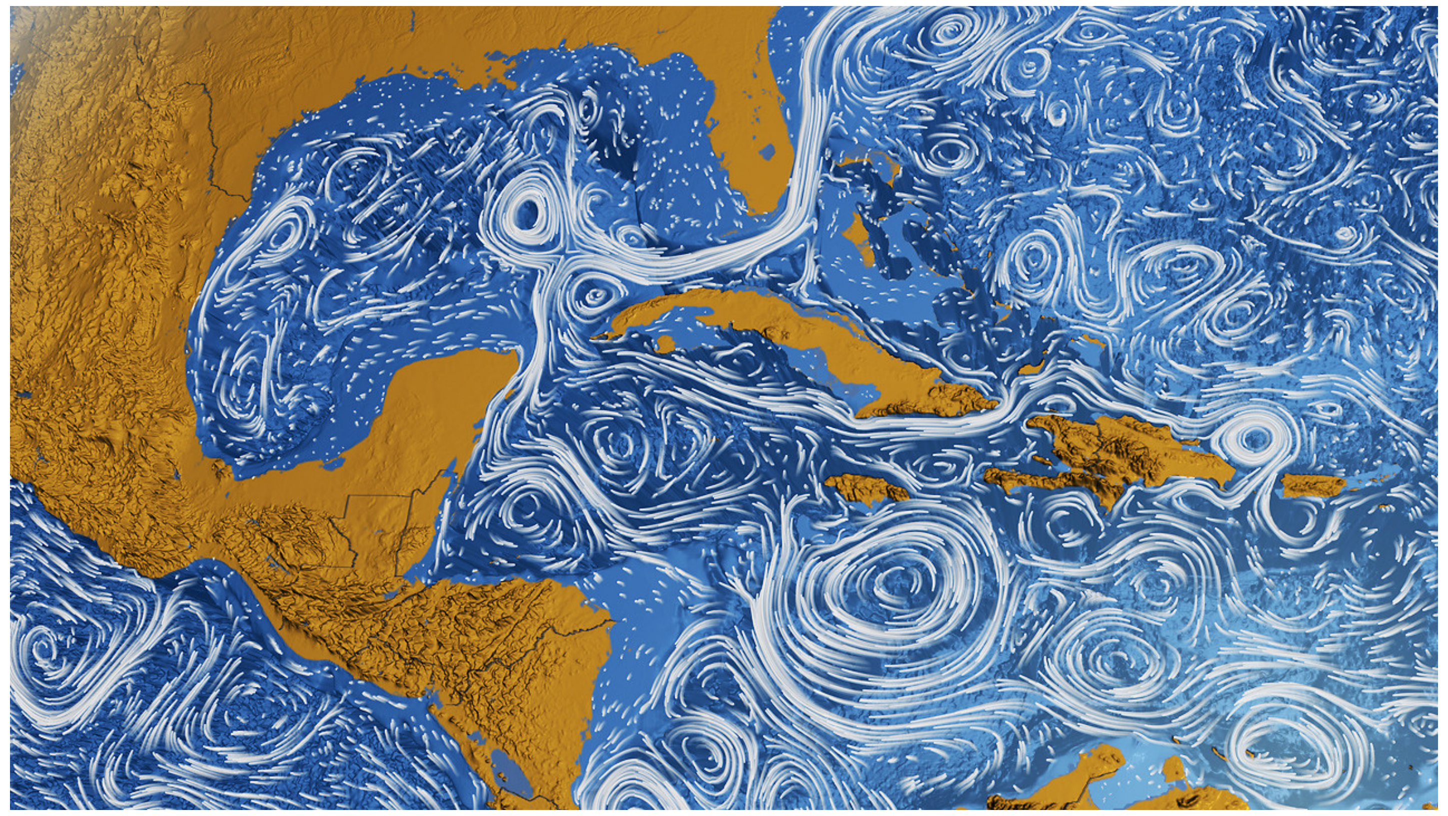
With:

Petros Koumoutsakos - ETH Zurich, **CSE L**ab
Wonmin Byeon - NVIDIA, Zhong Win Wan - MIT, Themistoklis Sapsis - MIT
Jaideep Pathak, Brian Hunt, Edward Ott - Univ. of Maryland

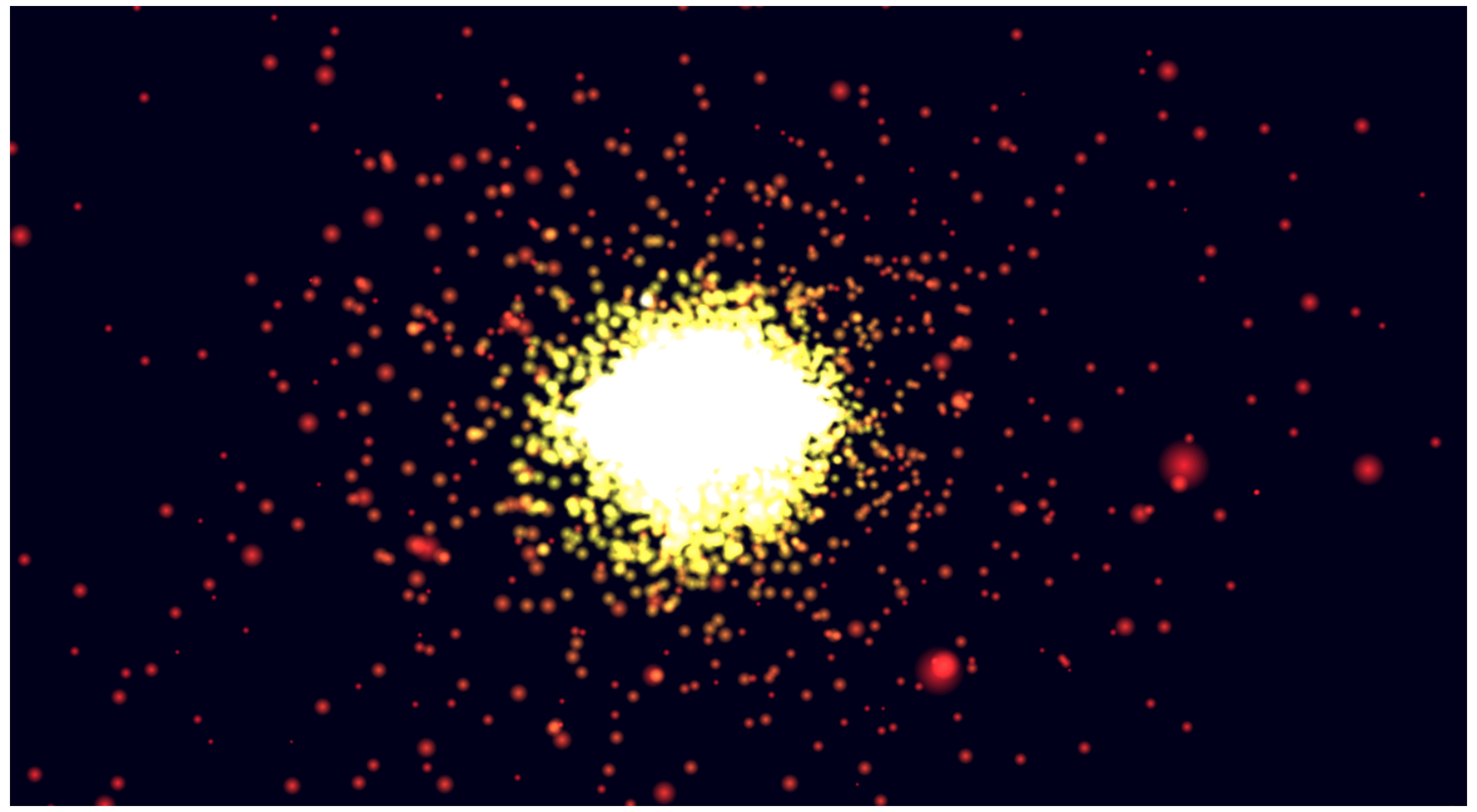
Motivation



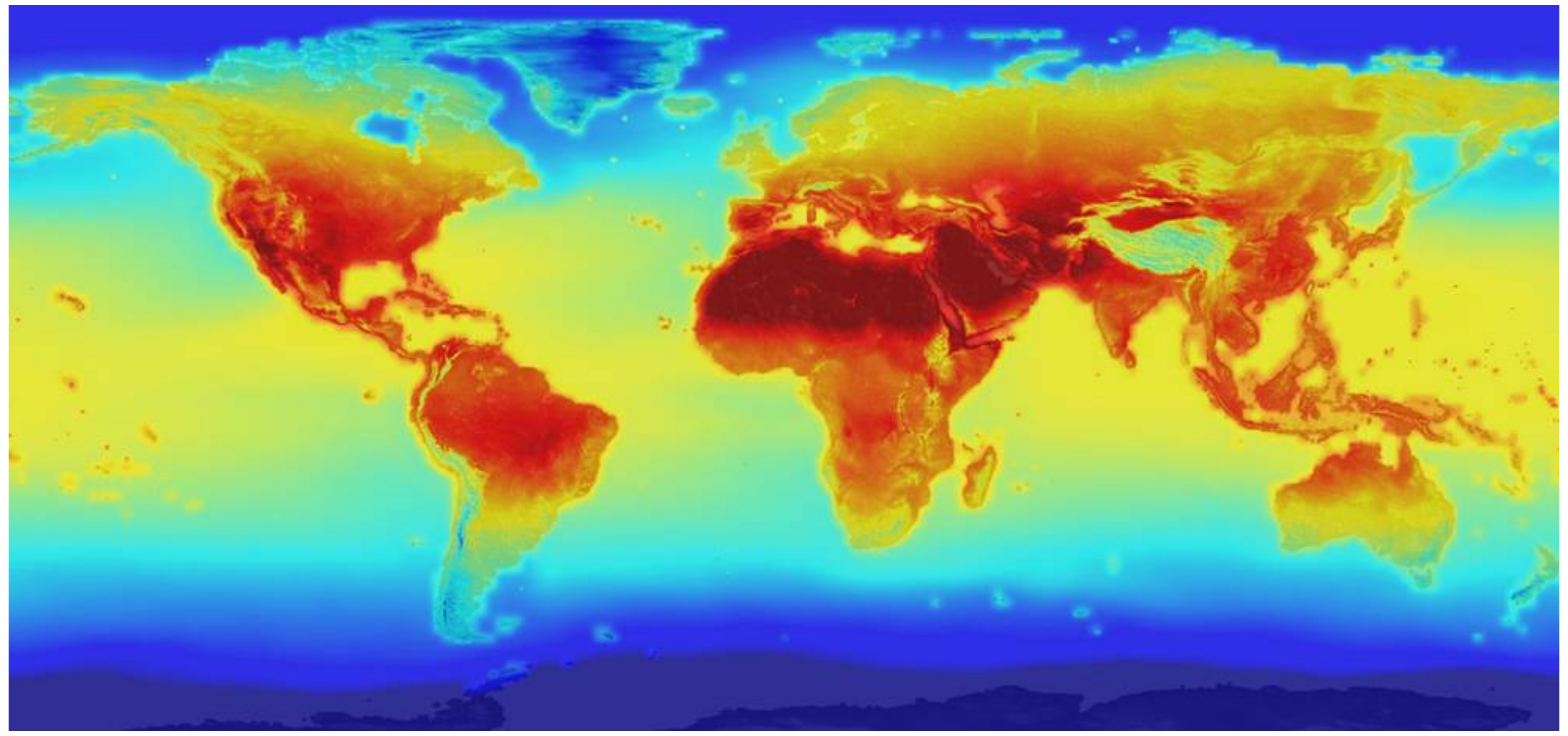
Turbulent phenomena



Ocean currents



N body problems in astrophysics

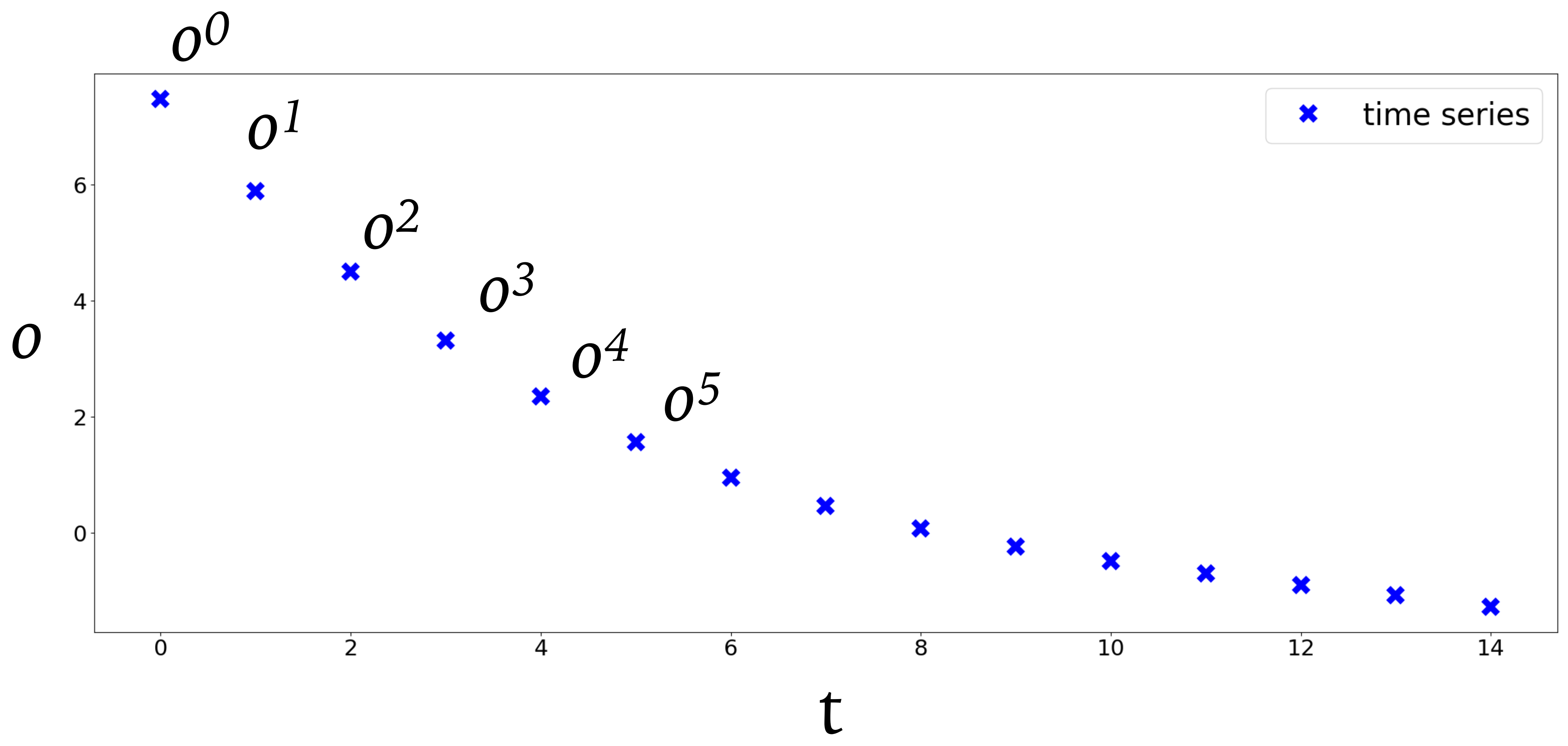


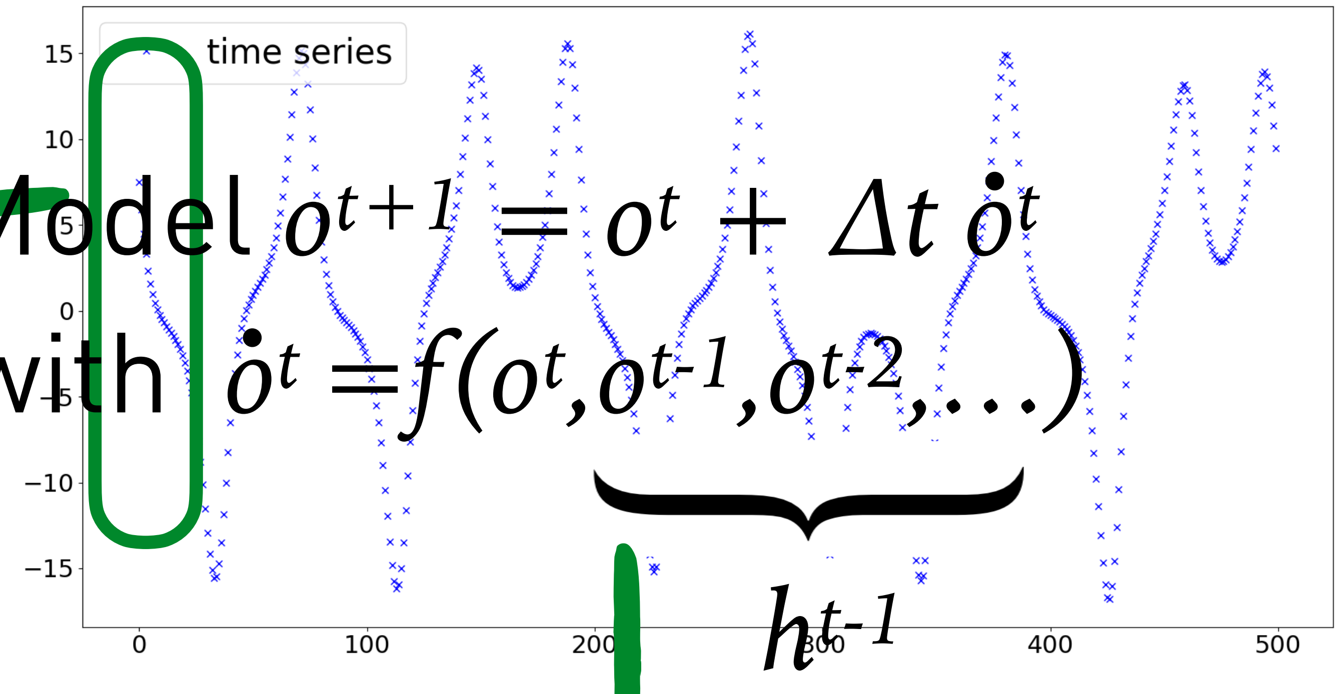
Climate

High-dimensional systems with chaotic behaviour are abound in nature and engineering

Forecasting of such systems is of great practical importance

Recurrent Neural Networks



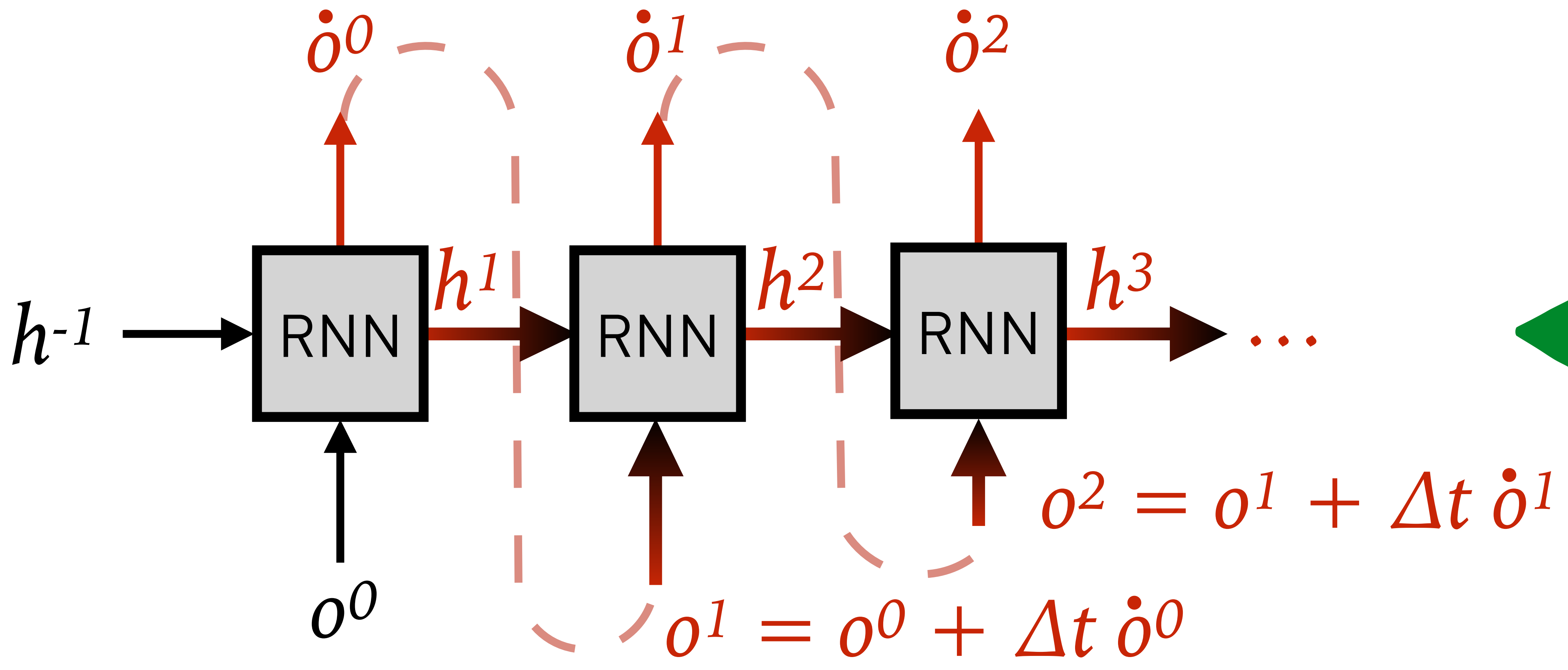


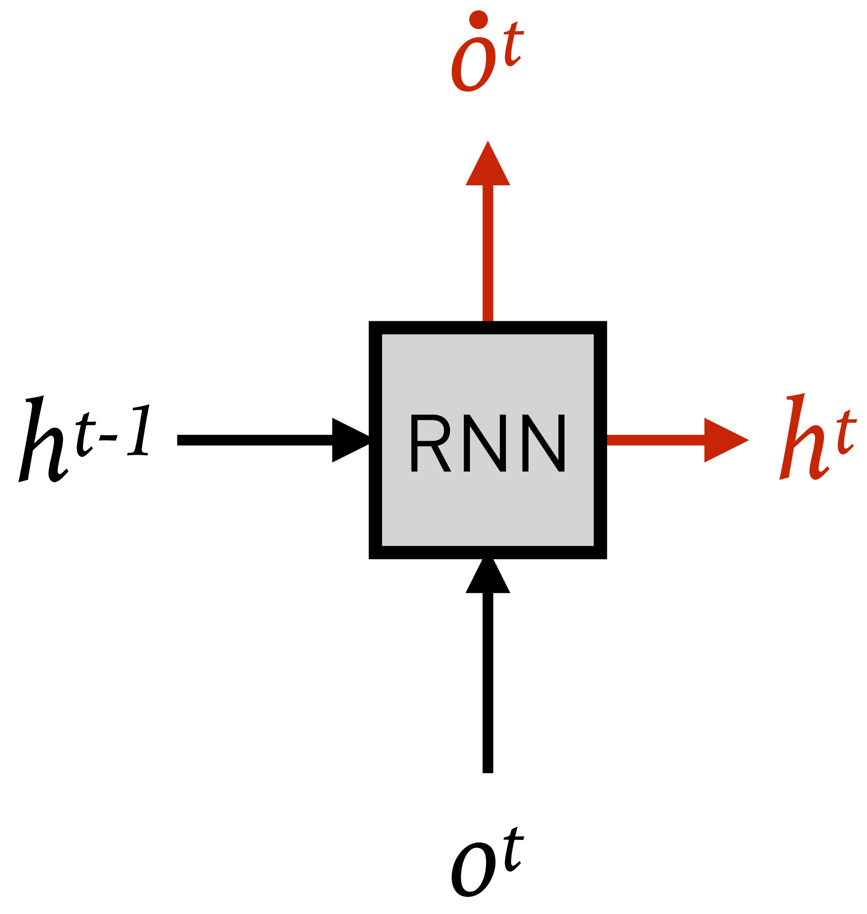
Model $o^{t+1} = o^t + \Delta t\, \dot{o}^t$

with $\dot{o}^t = f(o^t, o^{t-1}, o^{t-2}, \ldots)$

history

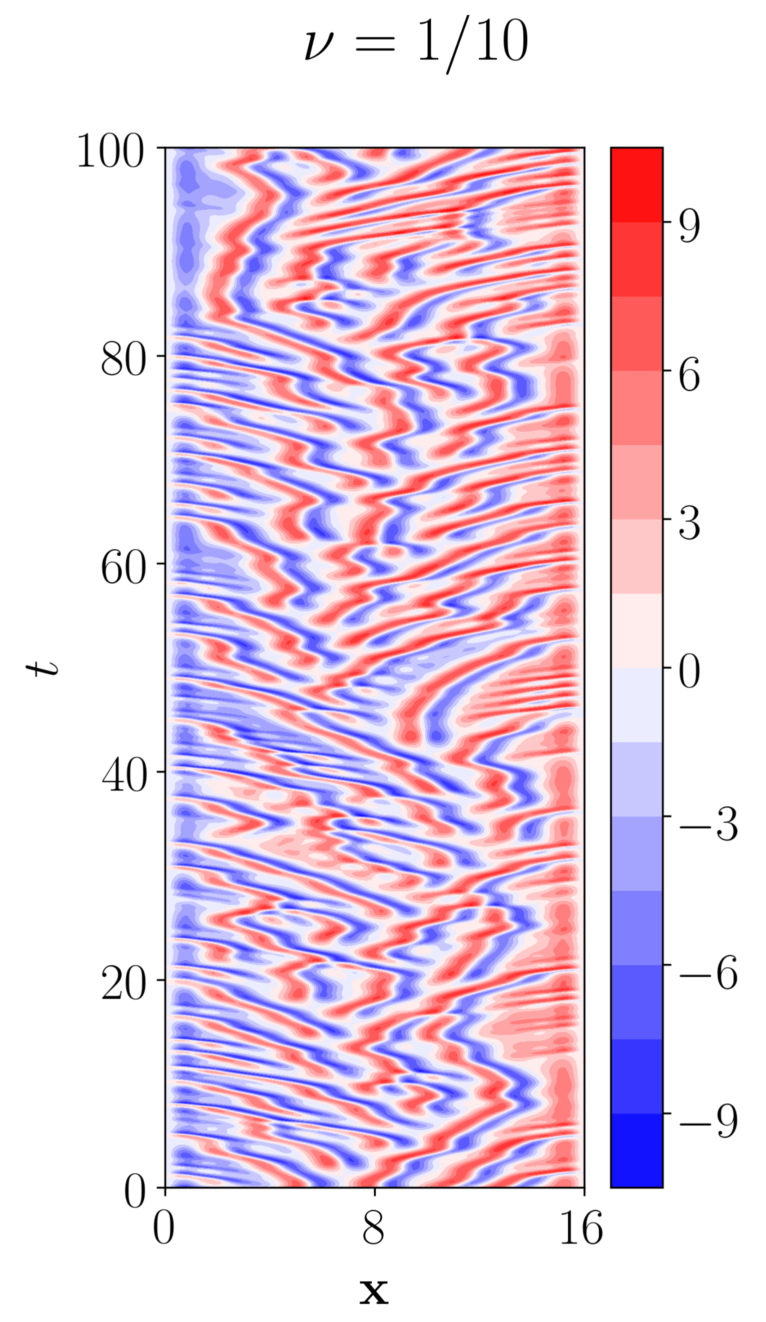
encoded in h^{t-1}

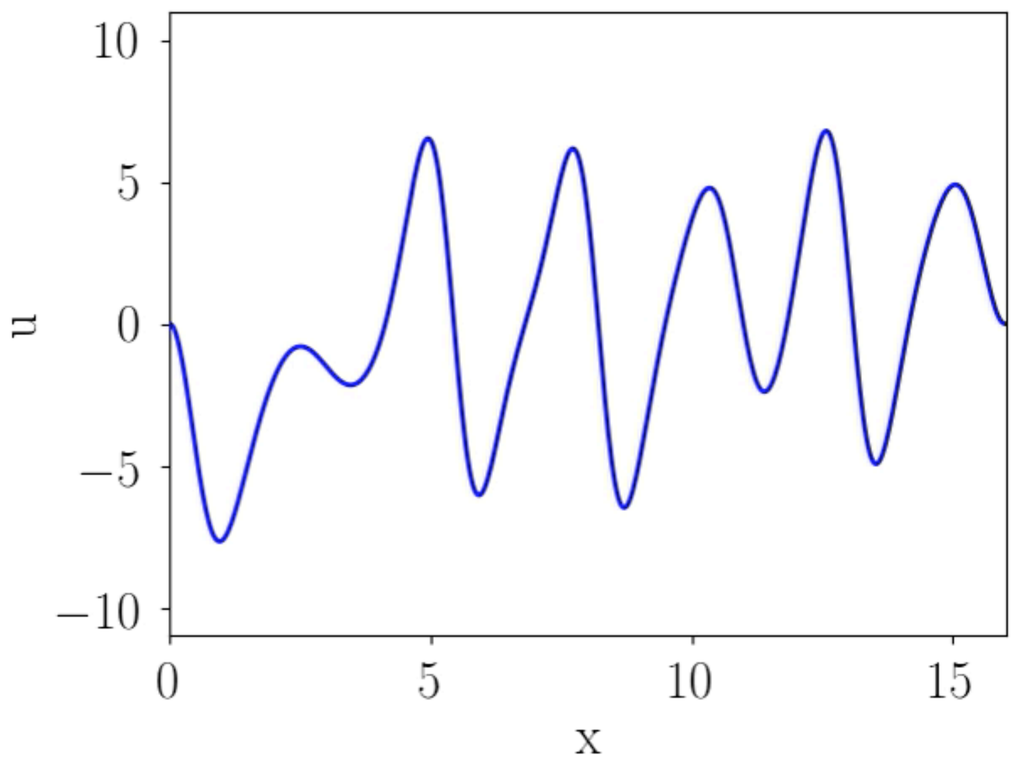




Training Data - Reducing the dimensionality **(observable)**

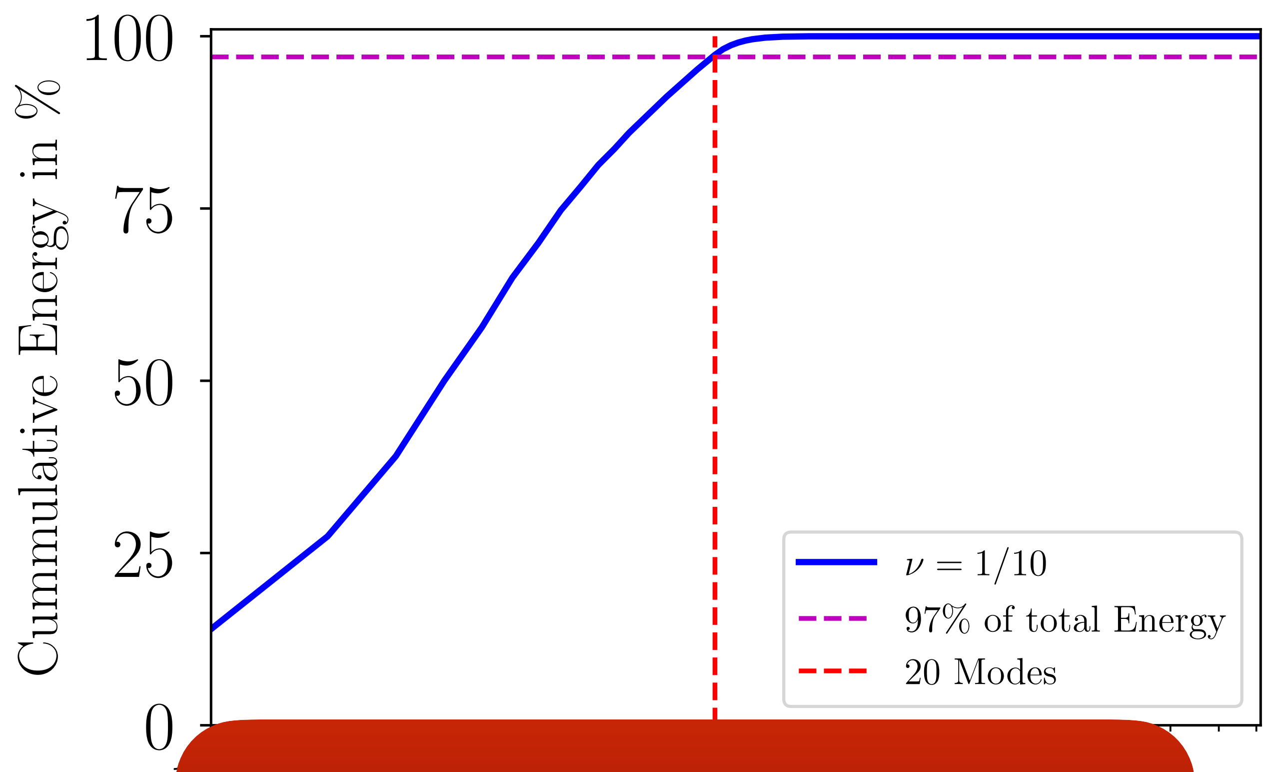
High dimensional





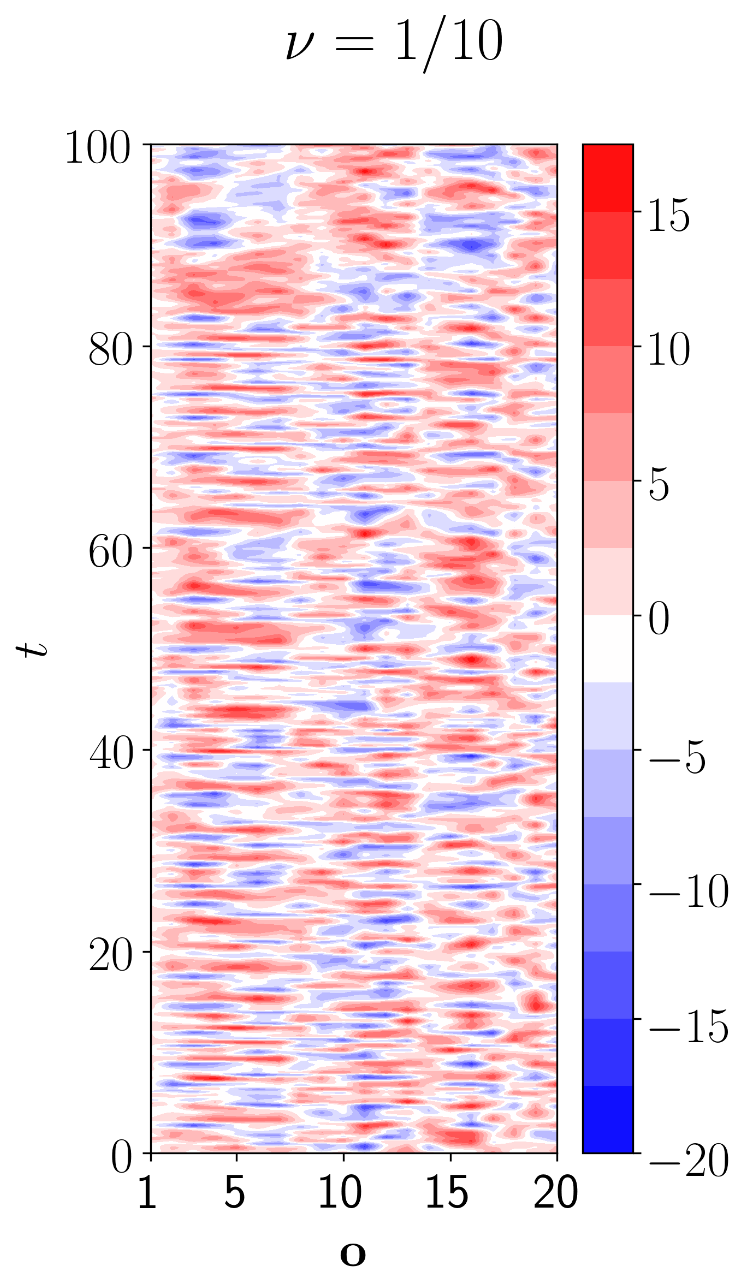
SVD

Singular
Value
Decomposition
(SVD)

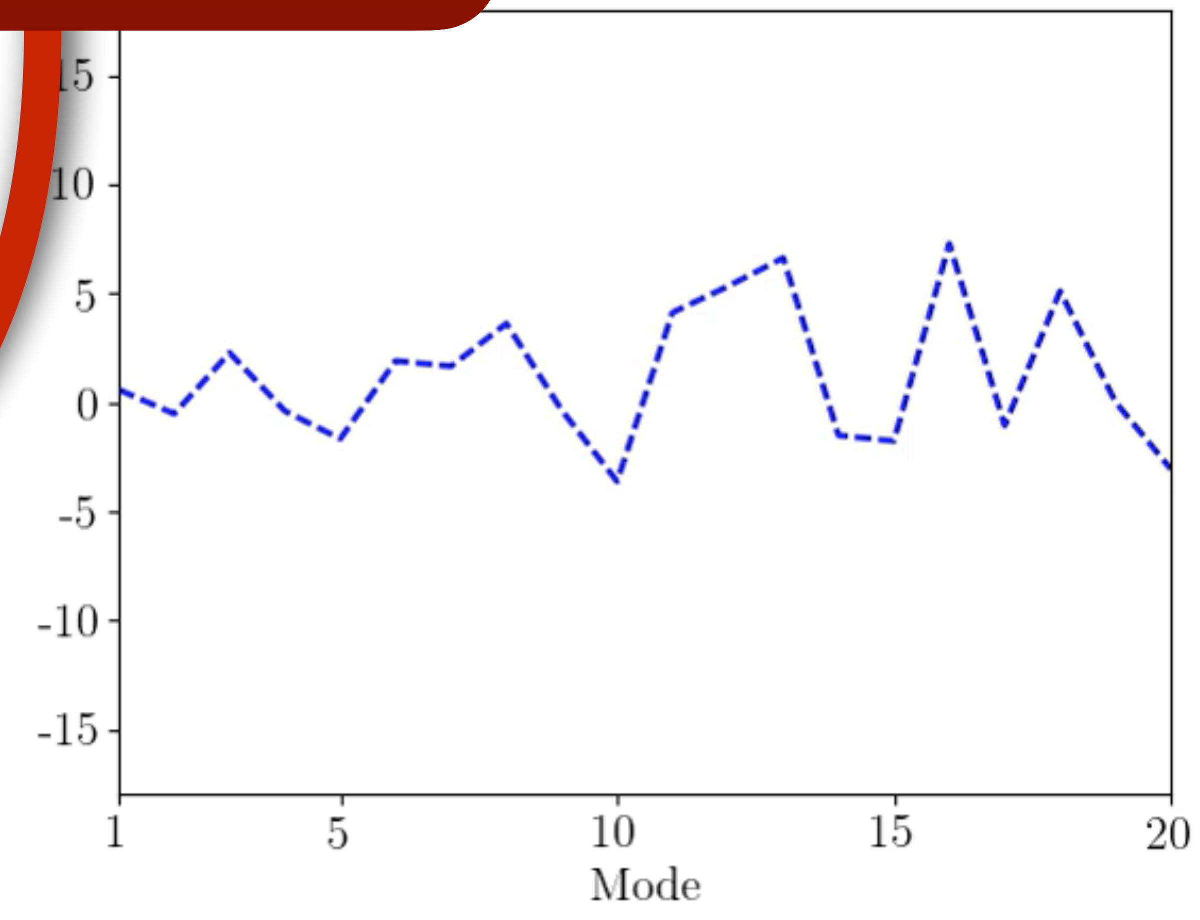


Throw away modes with low energy

Training Data!



20 Modes (observable)



Low dimensional state
(most energetic modes)

How to train these networks ?

Time-series: $\{\tilde{o}_1, \tilde{o}_2, \tilde{o}_3, \ldots, \tilde{o}_{11}, \tilde{o}_{12}\}$

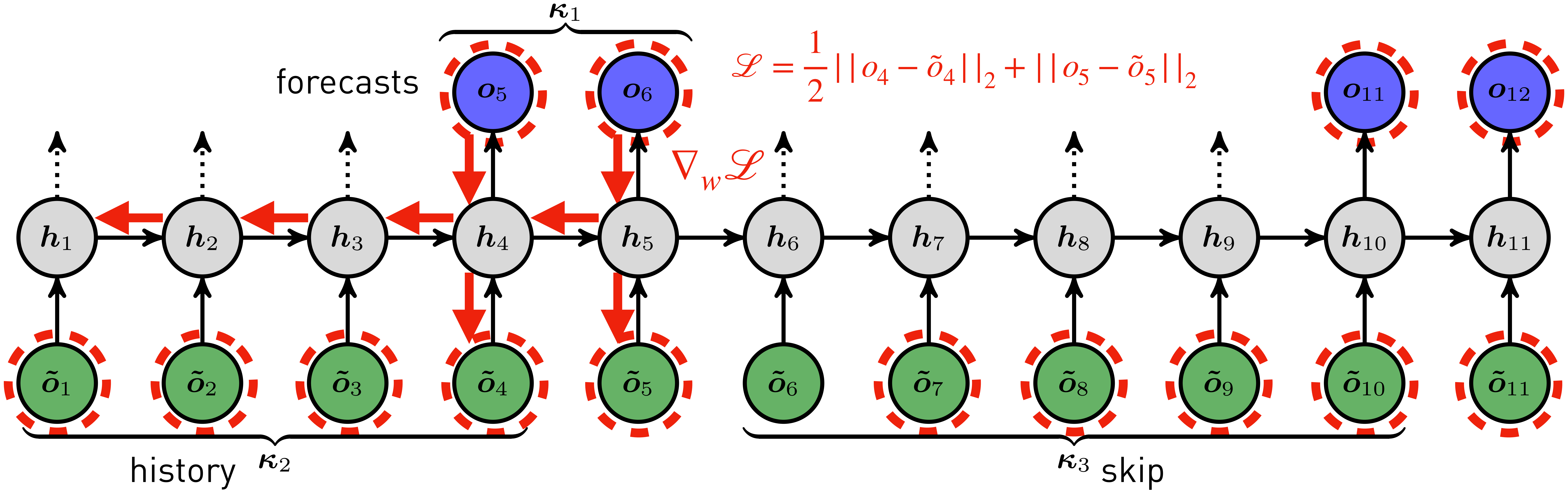
Target: $\{\tilde{\tilde{o}}_1, \tilde{\tilde{o}}_2, \tilde{\tilde{o}}_3, \ldots, \tilde{\tilde{o}}_{11}\}$ or $\{\tilde{o}_2, \tilde{o}_3, \tilde{o}_4, \ldots, \tilde{o}_{12}\}$

Algorithm: Back-propagation through time BPPT

Target batch: $[\tilde{o}_4, \tilde{o}_5]$

Target batch: $[\tilde{o}_{10}, \tilde{o}_{11}]$

$$\mathscr{L} = \frac{1}{2}||o_4 - \tilde{o}_4||_2 + ||o_5 - \tilde{o}_5||_2$$



Input batch: $[\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5]$

Input batch: $[\tilde{x}_7, \tilde{x}_8, \tilde{x}_9, \tilde{x}_{10}, \tilde{x}_{11}]$

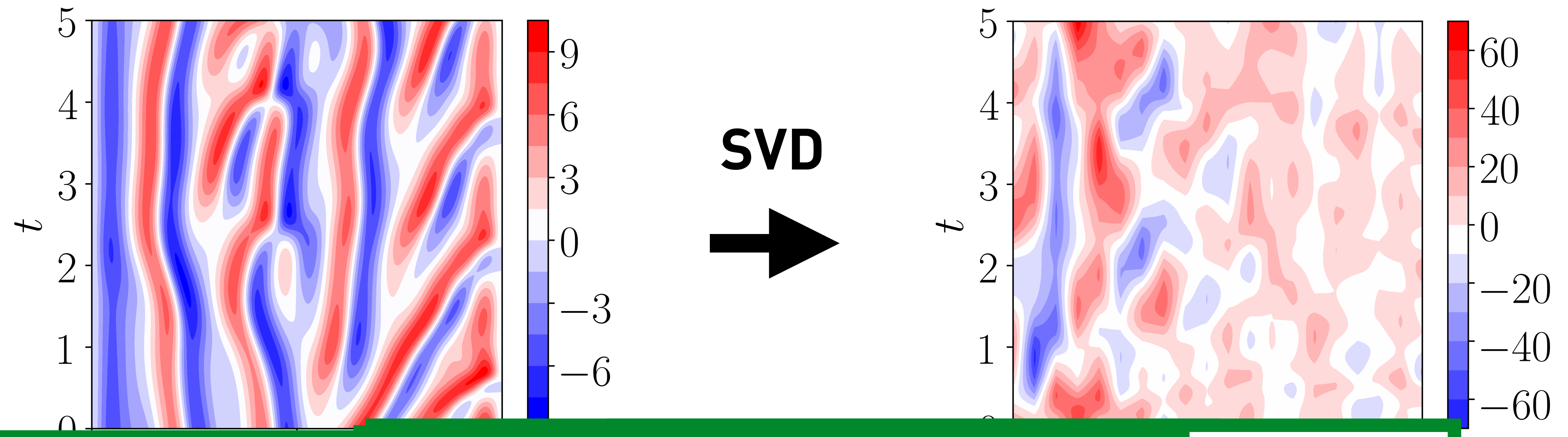
Forecasting on UNSEEN data - Iterative prediction in practice d_h=100, d=10

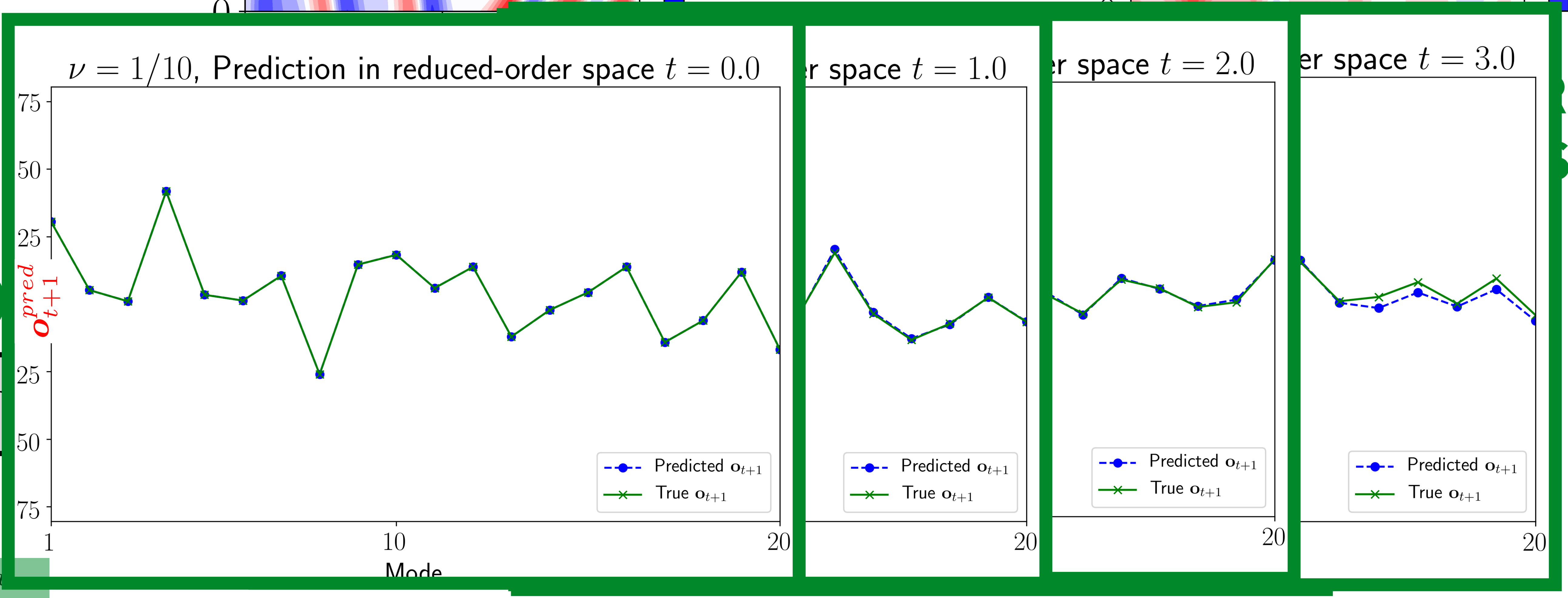
1 Train the model with BPTT*

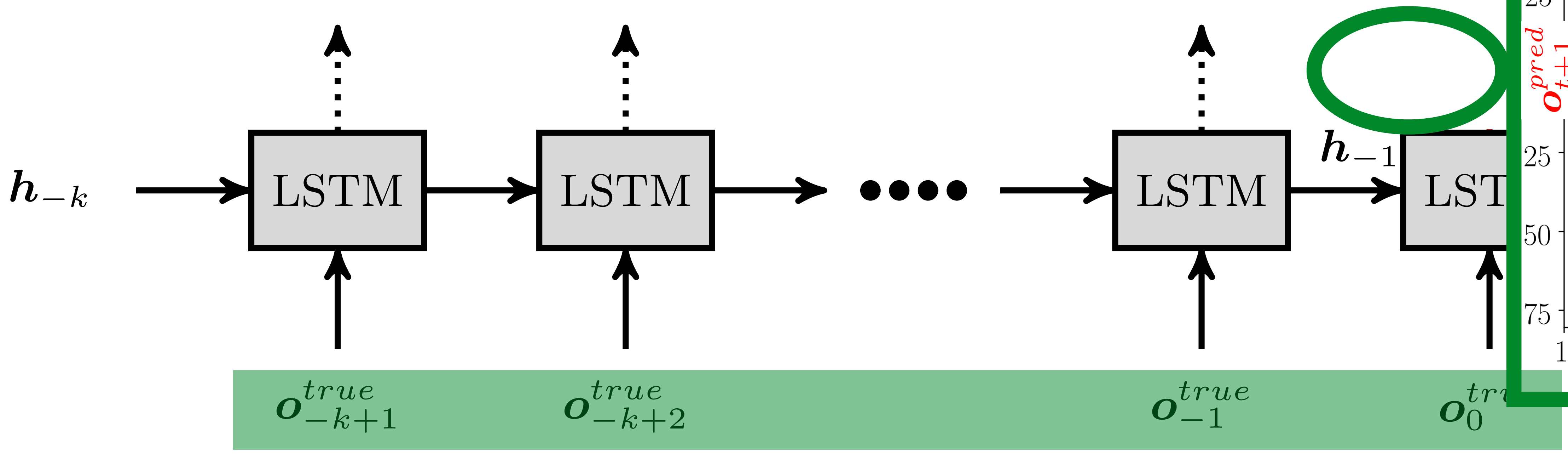
2 How to predict on Test (unseen) data -set ?

TEST case - STATE evolution

TRUE mode evolution







Short term history known

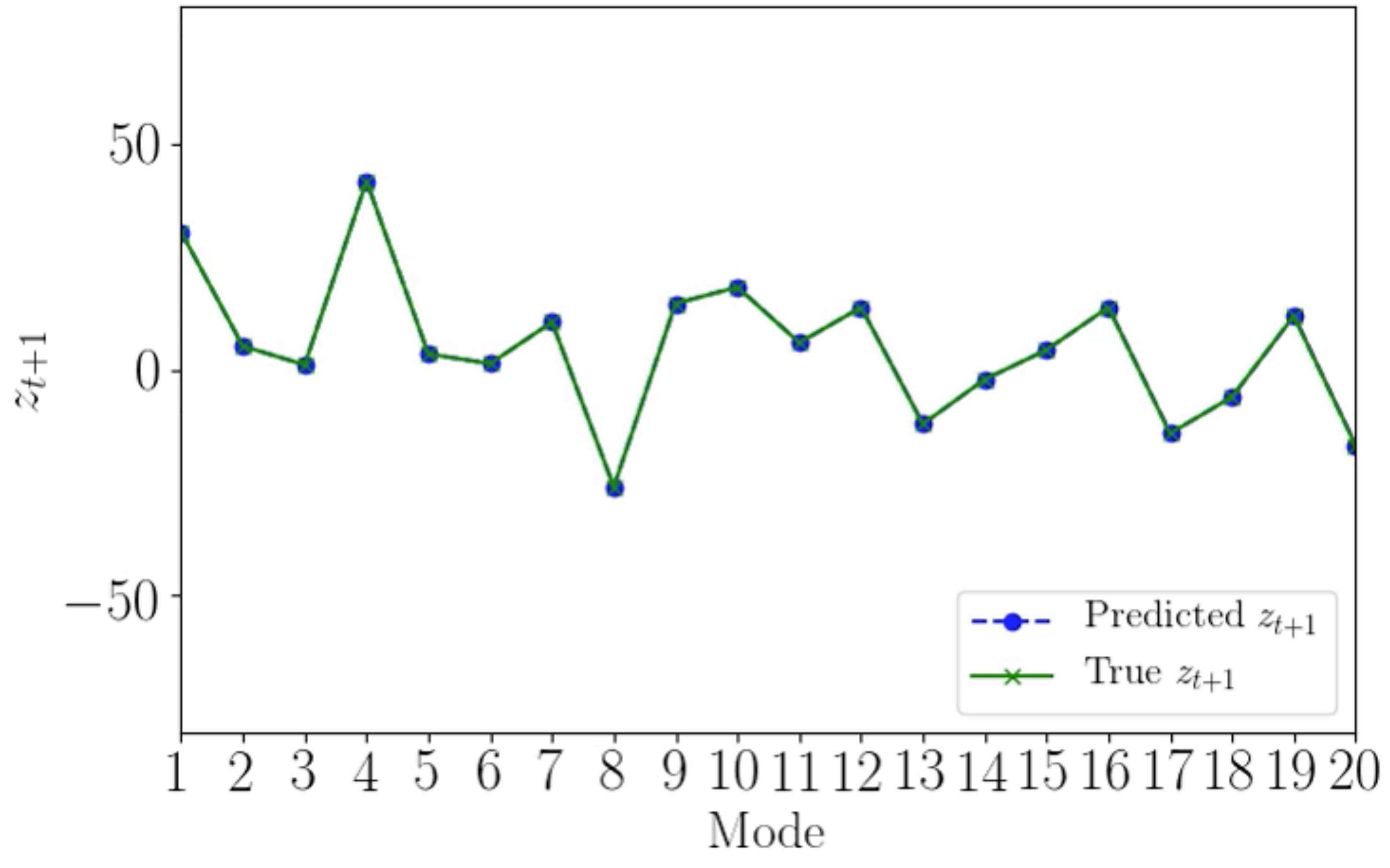
*Back-propagation through time BPPT

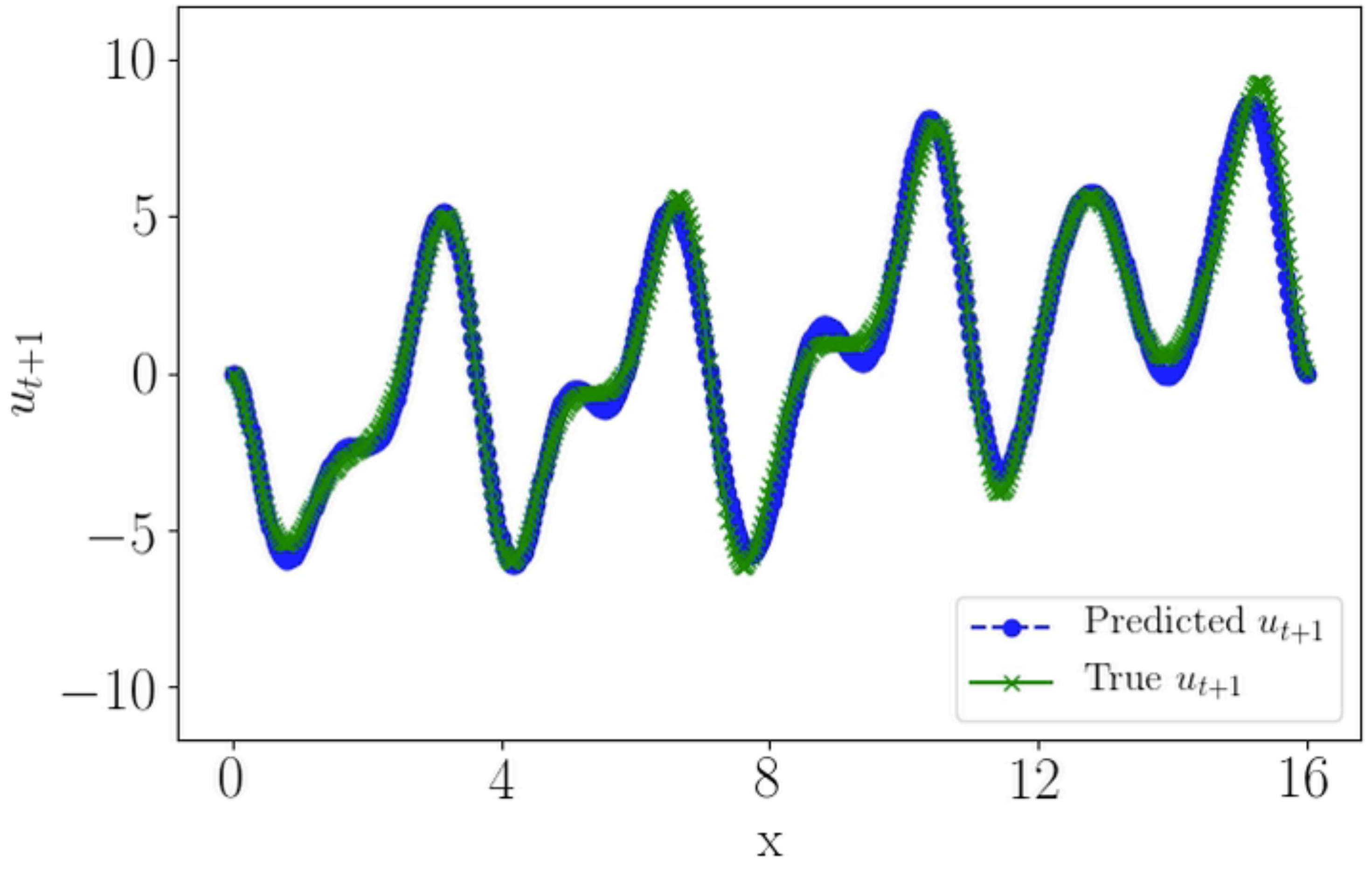
Forecasting - Iterative prediction in practice $d_h=100, d=10$

prediction in the **reduced** space

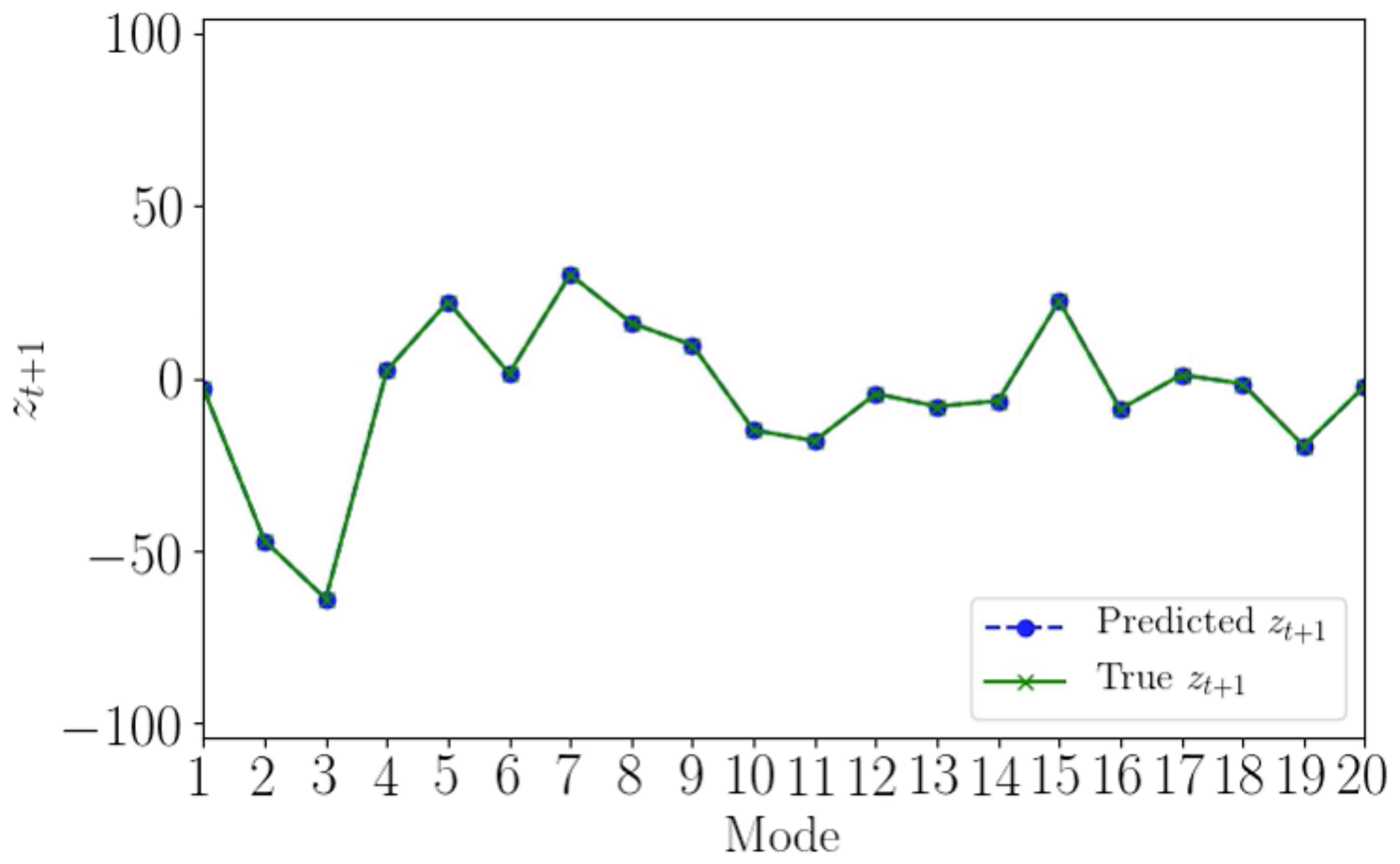
expanded prediction in the **original** space

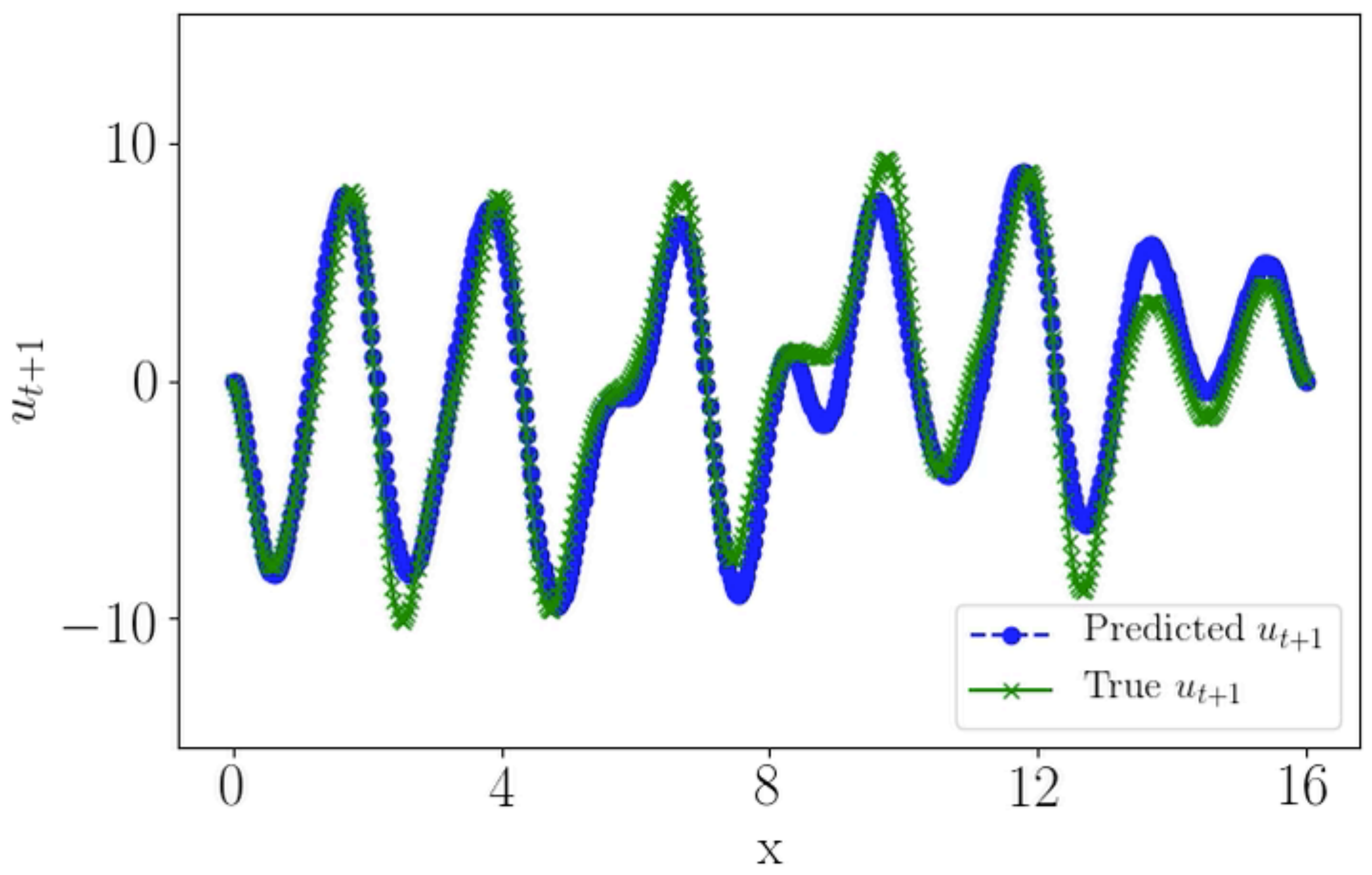
v = 1/10





v = 1/16





Problems ?

1 Iterative prediction causes **accumulation of prediction errors.** Especially in reduced order observables, it may cause **divergence.**

- Dynamics underrepresented in training data
- Scarce data in attractor boundaries
- Under-resolved high dimensional dynamics
- Models not generalising

Capturing Long-Term Behavior

PROBLEMS

- Iterative forecasting with LSTM suffers from **accumulation** of prediction **errors**
- There is no mechanism that guarantees that trajectories remain on the attractor

SOLUTION - MEAN STOCHASTIC MODEL (MSM)

- Ornstein-Uhlenbeck process - computationally cheap
- Converges in the **long-term** in **mean statistical behavior**
- Relies on **global attractor statistics**
- Efficient in highly chaotic systems

Hybrid LSTM - MSM approach: Use **MSM** in **regions underrepresented in the training** data or near attractor **boundaries**

Mean Stochastic Model

$$dz_t = c\, z_t\, dt + \zeta\, dW_t$$

parameters estimated from **data**

wiener process

$$\zeta = \sqrt{-2\, c\, \sigma_z}$$

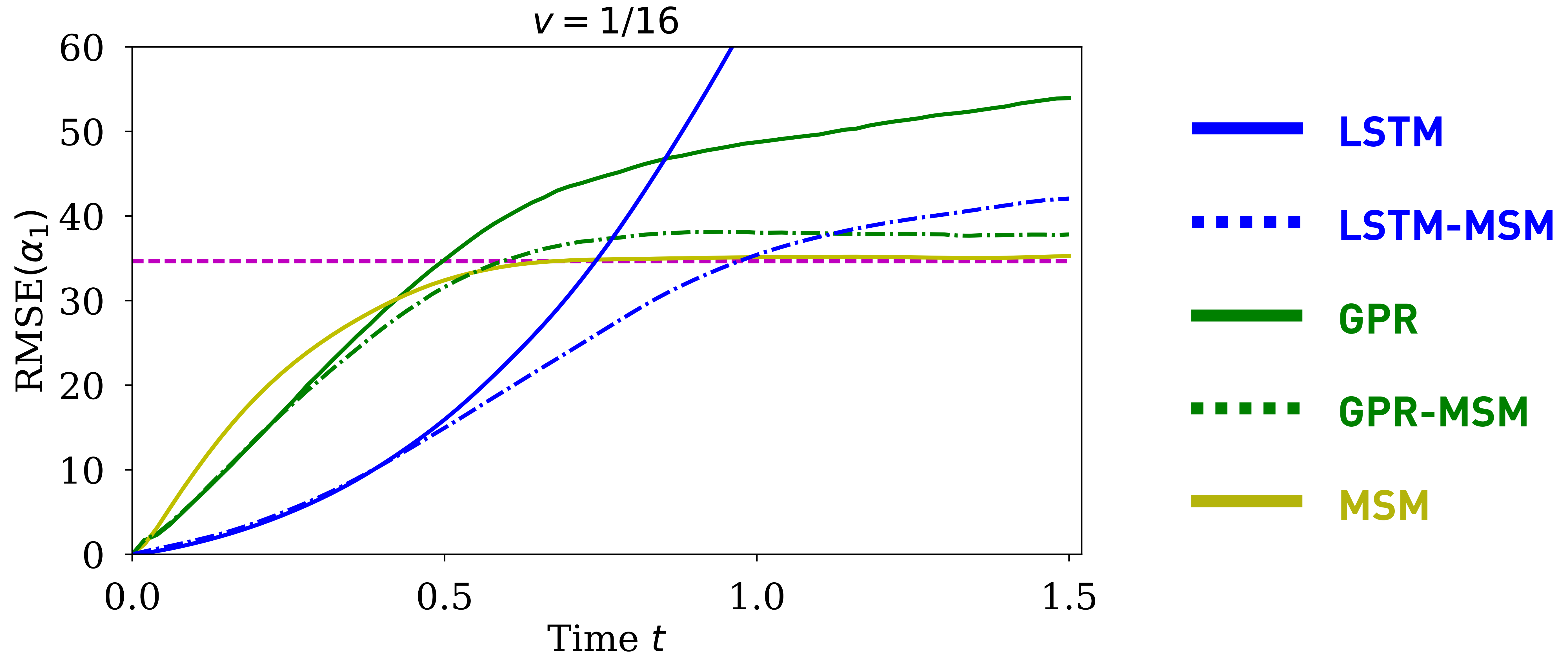
data standard deviation

$$c = \frac{1}{T}$$

decorrelation time

Results - Kuramoto-Sivashinsky - **Comparison with Gaussian Process Regression (GPR)**

V	Total number of initial conditions (**ic**)
k	Mode number
i	IC index
z_k^i	**True** state of mode k starting from **ic** i
$\widetilde{z}_k^i$	**Predicted** state of mode k starting from **ic** i



Root mean square error (RMSE)

$$RMSE(z_k) = \sqrt{1/V \sum_{i=1}^{V} \left(z_k^i - \widetilde{z}_k^i\right)^2}$$

Time evolution of the RMSE of **the most energetic PCA** mode over 1000 initial conditions

Problems ?

1 Iterative prediction causes **accumulation of prediction errors.** Especially in reduced order observables, it may cause **divergence.**

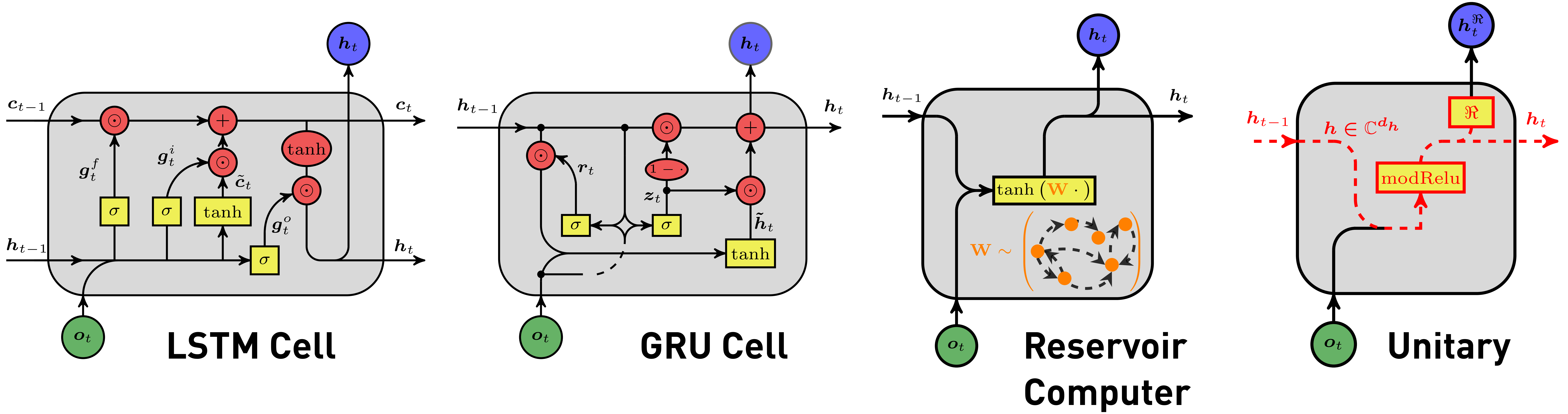
- Dynamics underrepresented in training data
- Scarce data in attractor boundaries
- Under-resolved high dimensional dynamics
- Insufficient training, low quality data, noisy

Solution? Hybrid LSTM - MSM approach

2 **Vanishing gradients problem during training:** As the gradient is back-propagated during training of the networks it may **vanish to zero or explode.**

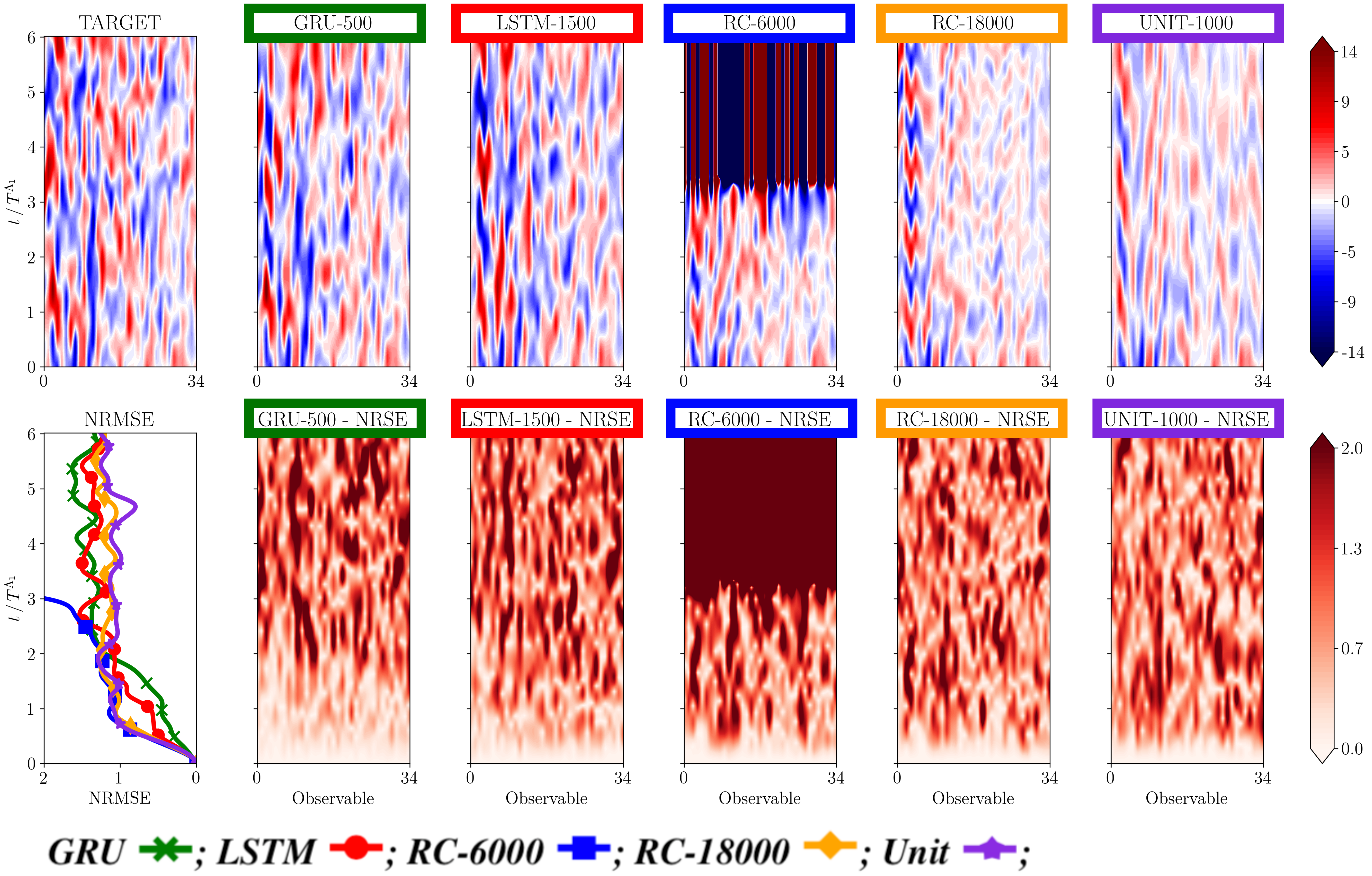
Solution? Sophisticated architectures

RNN Models

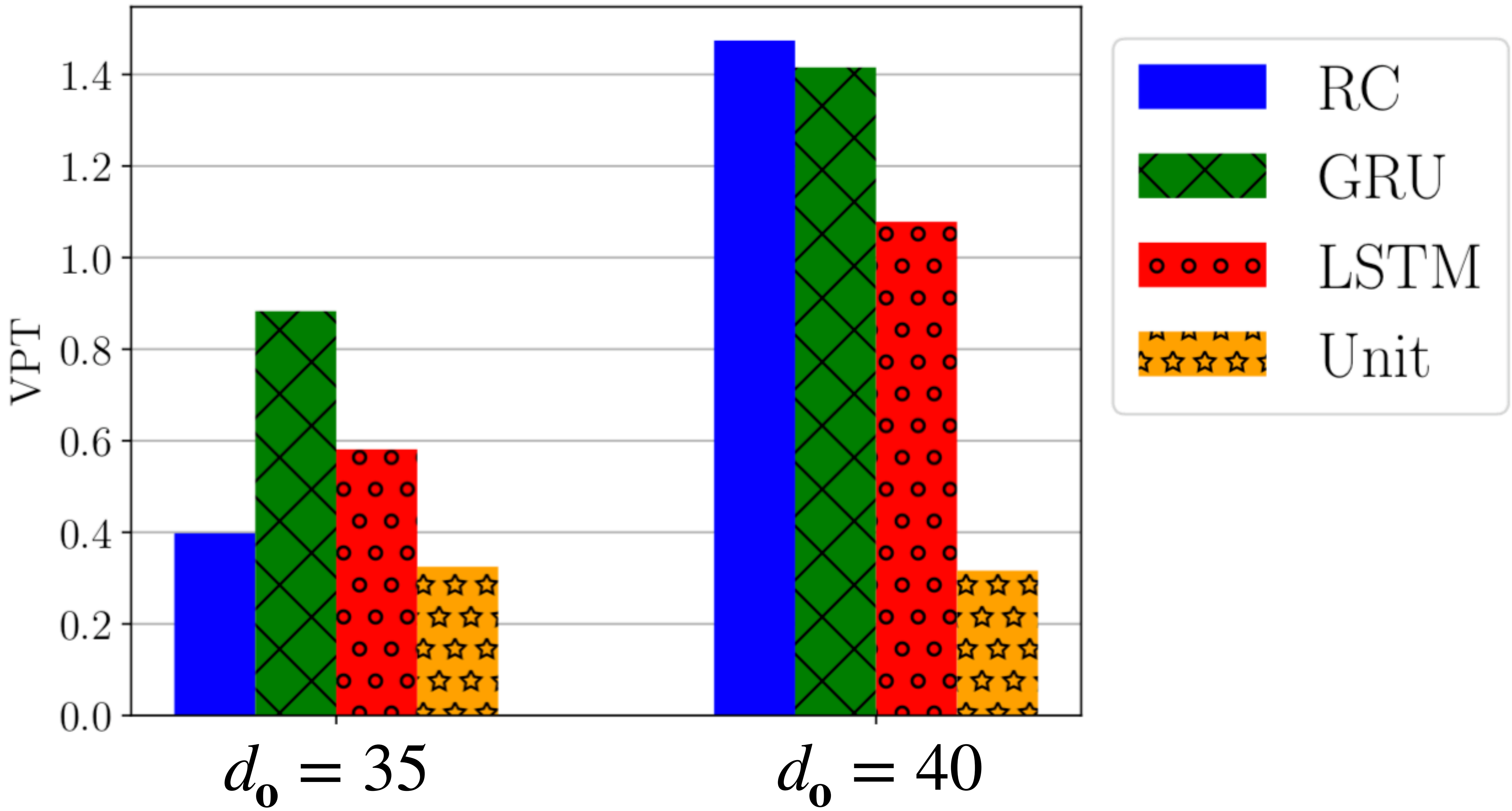


c_{t-1}
c_t
h_{t-1}
h_t
g_t^f
g_t^i
g_t^o
$\tilde{c}_t$
σ
tanh
o_t
LSTM Cell
r_t
z_t
$1-\cdot$
$\tilde{h}_t$
GRU Cell
tanh (W ·)
W ~
Reservoir Computer
$h \in \mathbb{C}^{d_h}$
$\Re$
modRelu
$h_t^{\Re}$
Unitary

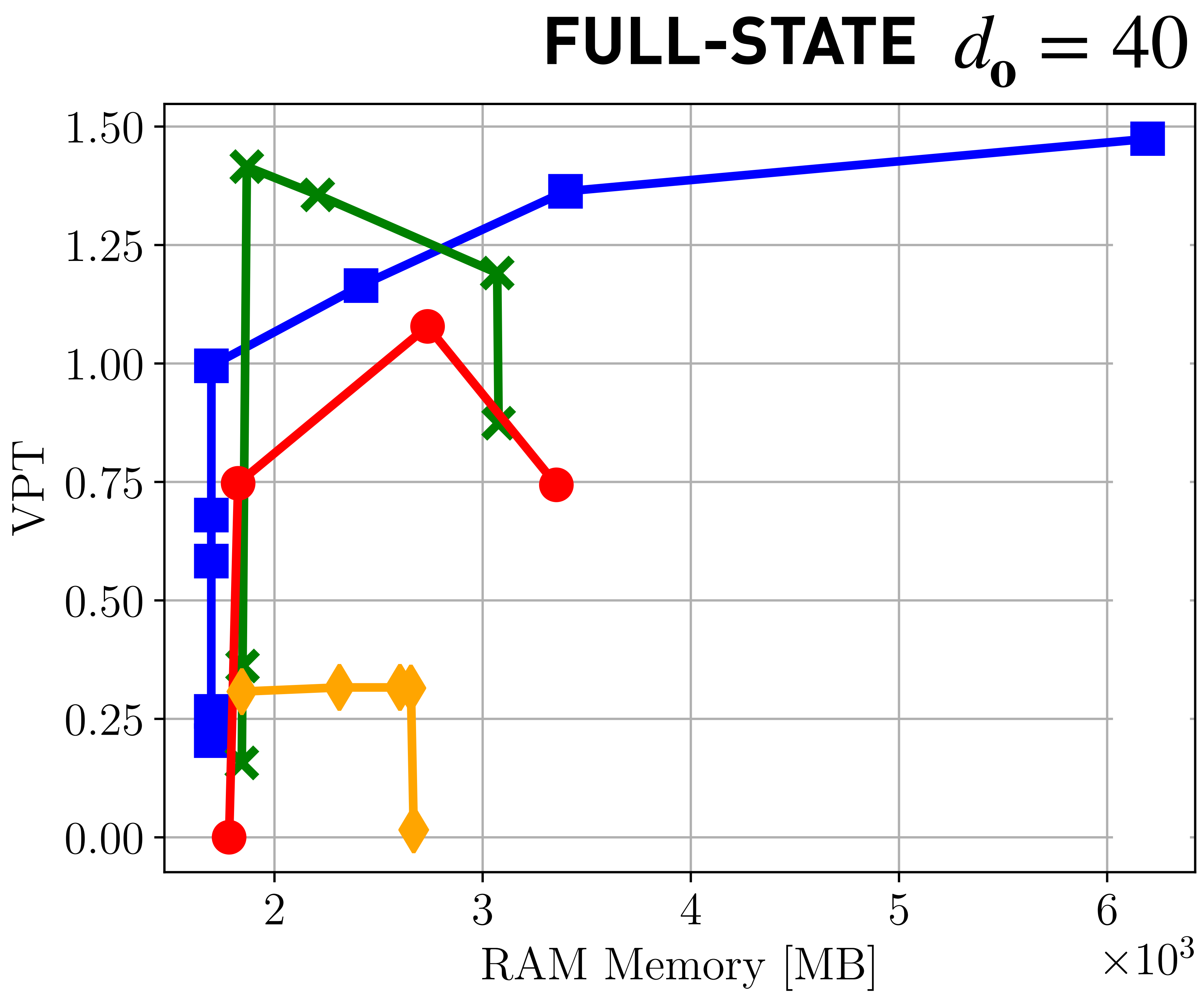
Lorenz-96 - 35/40 mode observable



Benchmarking - Lorenz-96 - 35/40 **SVD** mode observable

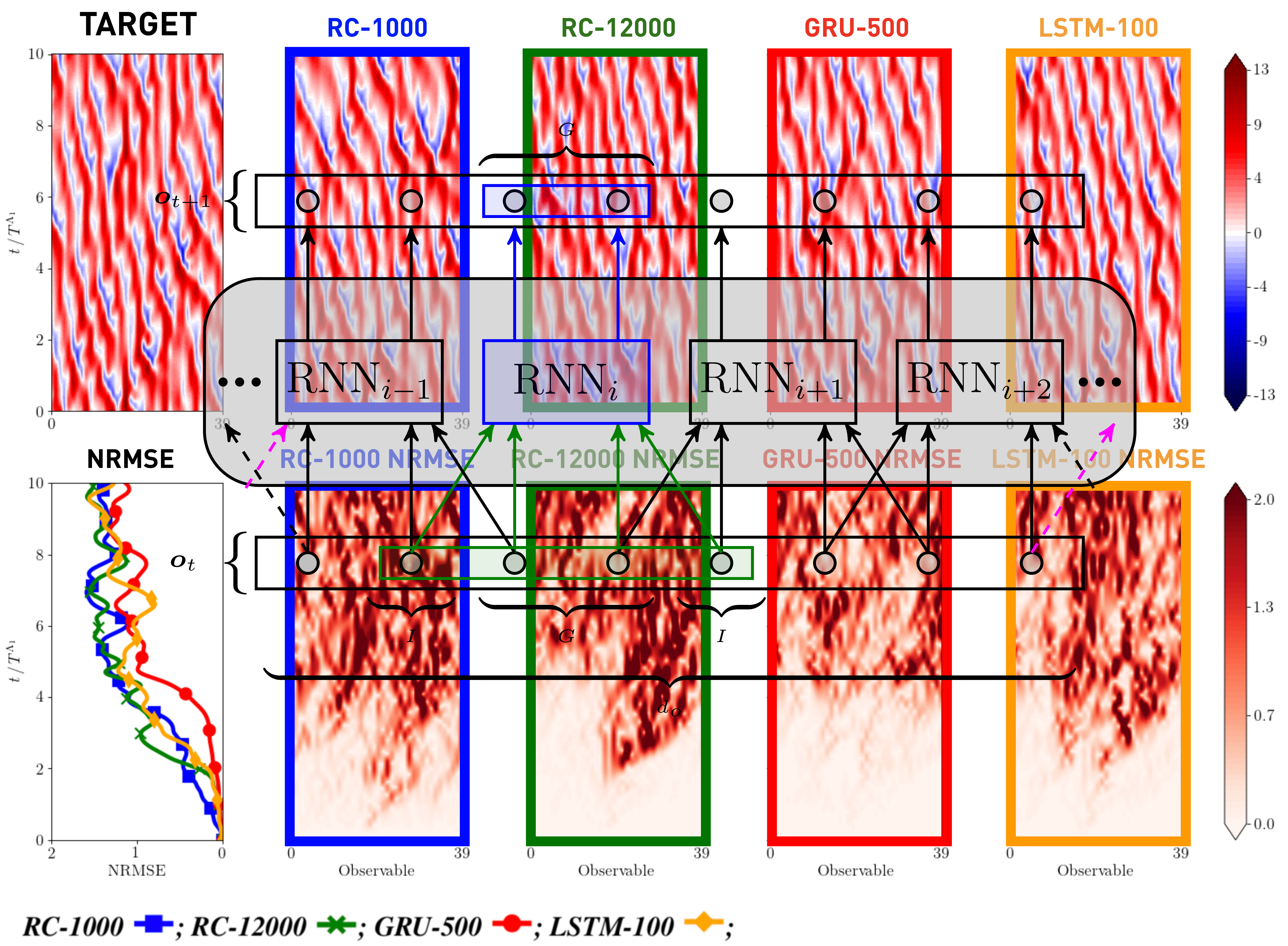


reduced-order **full-state**

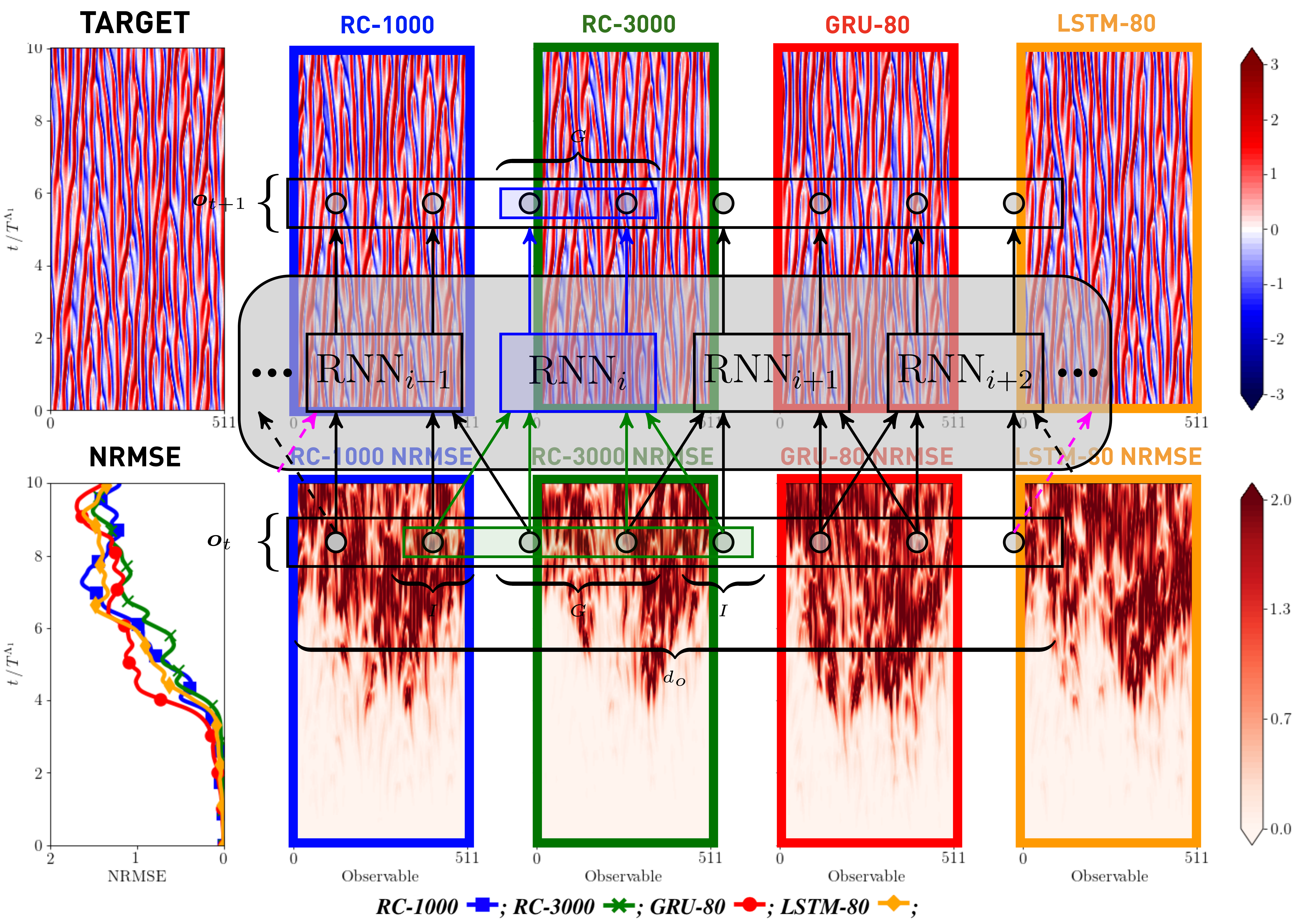


RC; *GRU*; *LSTM*; *Unit*

Lorenz-96 - full state information + parallelism

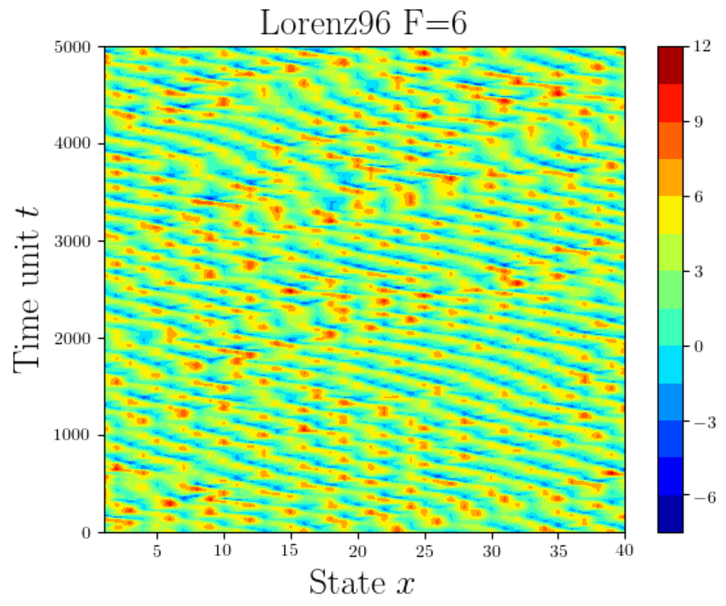


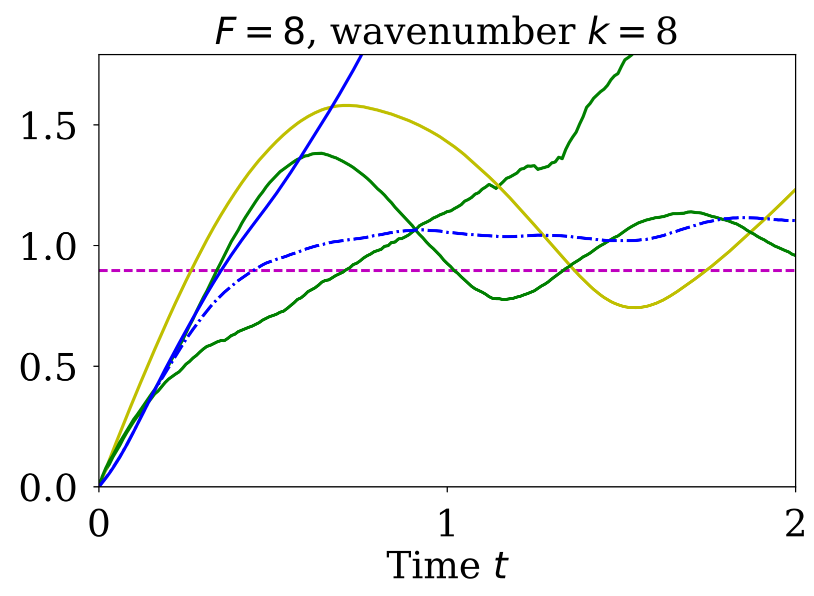
Kuramoto-Sivashinsky - Full state information + Parallelism



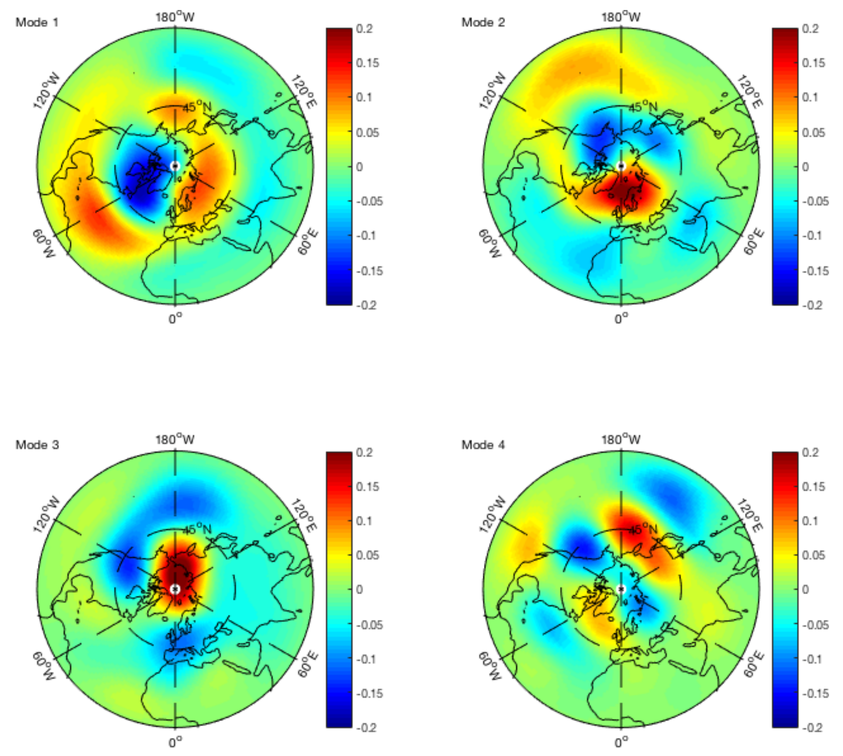
More applications...

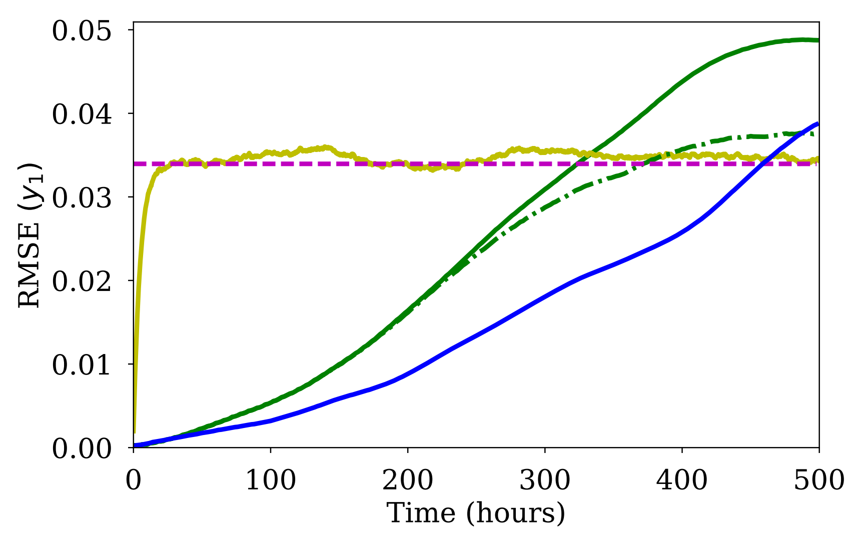
1. Lorenz 96



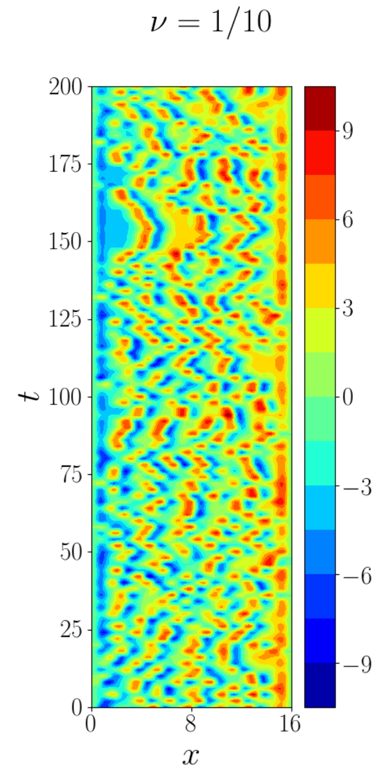


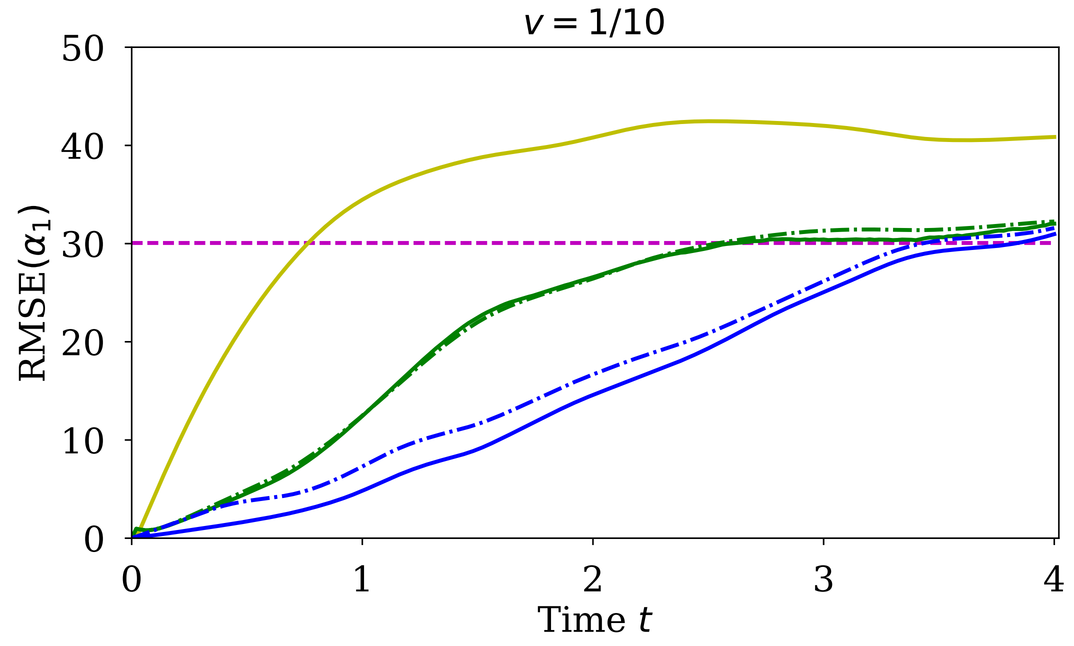
2. Barotropic Model





3. Kuramoto Sivashinsky

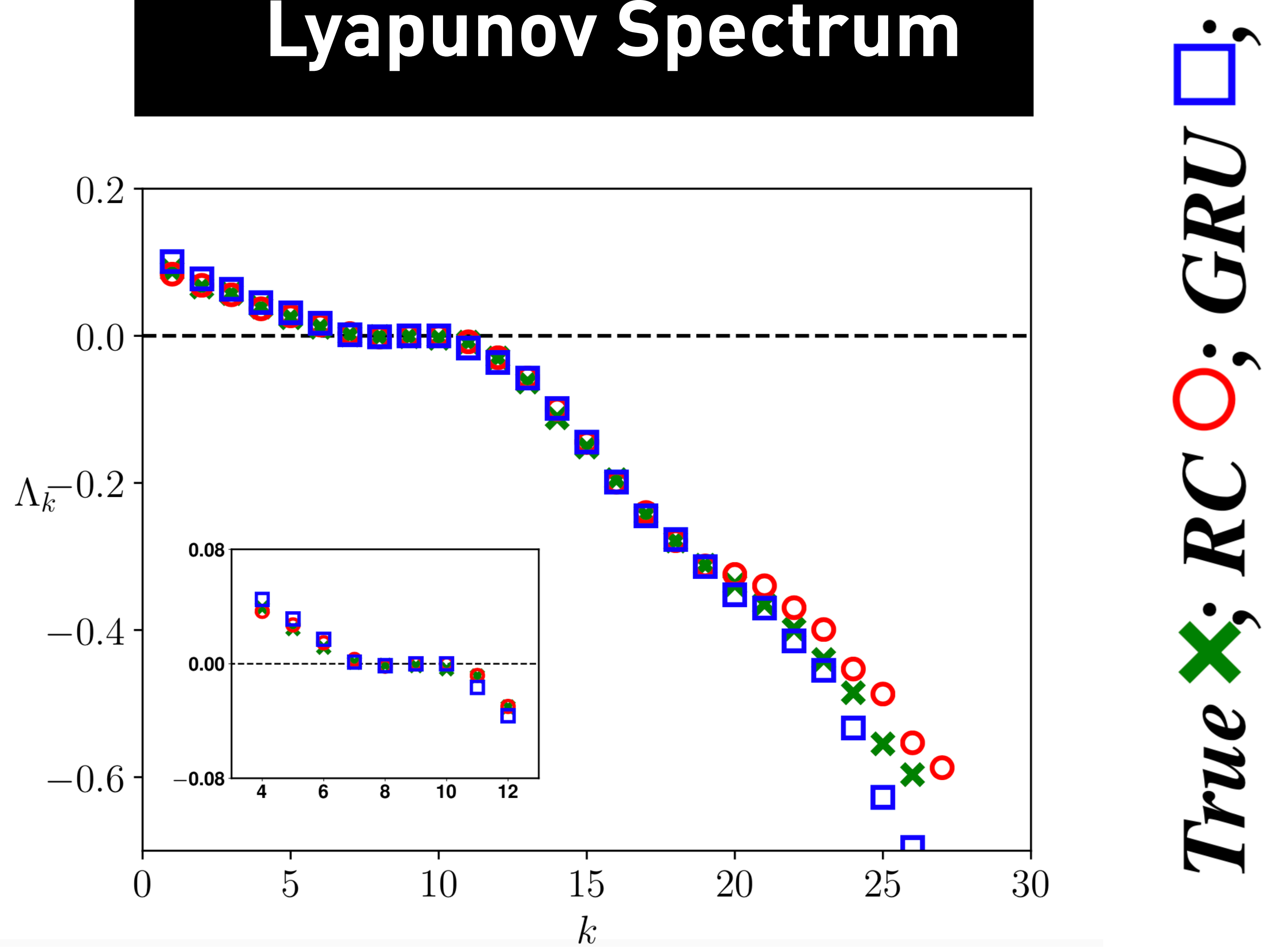




4. Deep Reinforcement Learning

$$Q^{\pi}(\mathbf{s},\alpha) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty}\gamma^{k} r_{t+k+1} \middle| s_t = s_1, s_{t-2} = s_2, \dots, \alpha_t = \alpha_1, \alpha_{t-1} = \alpha_2, \dots\right] \hat{=} Q^{w}(\mathbf{s},\alpha)$$

Lyapunov Spectrum



Summary & Future Outlook

- **LSTMs** utilized as **nonlinear** data-driven predictors of high dimensional chaotic dynamical systems
- Coupled with **Mean Stochastic Model (MSM)** to capture **long-term statistics**
- Their prediction **accuracy** was benchmarked against:
  - Gaussian Processes (GPR), Reservoir Computers (RC), GRUs, Unitary RNNs **ongoing work…**
  - **in** Kuramoto-Sivashinsky, Lorenz-96 system, Barotropic climate model
- Dynamical system surrogates - **Lyapunov spectrum** - employed in **Reinforcement Learning**
- Other open questions
  - Stochastic (bayesian) RNNs estimating uncertainty of forecasts ?
  - Training procedures ?