Discovering Dynamical Parameters by Interpreting Echo State Networks

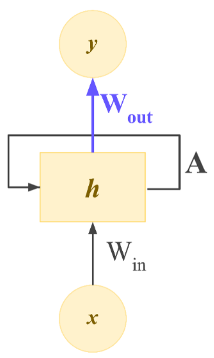
Oreoluwa Alao, Peter Y. Lu, Marin Soljačić

MIT Physics

NeurIPS 2021
AI for Science Workshop

Background: Echo State Networks

ESN architecture:
- Randomly initialized RNN-like architecture
- Sparse reservoir **A**, input weights $\mathbf{W}_{\text{in}}$ both fixed
- Only output/readout layer $\mathbf{W}_{\text{out}}$ is trained (via *linear regression*)



Strengths:
- High prediction performance
- Efficient and stable learning of chaotic dynamics
- Computationally inexpensive

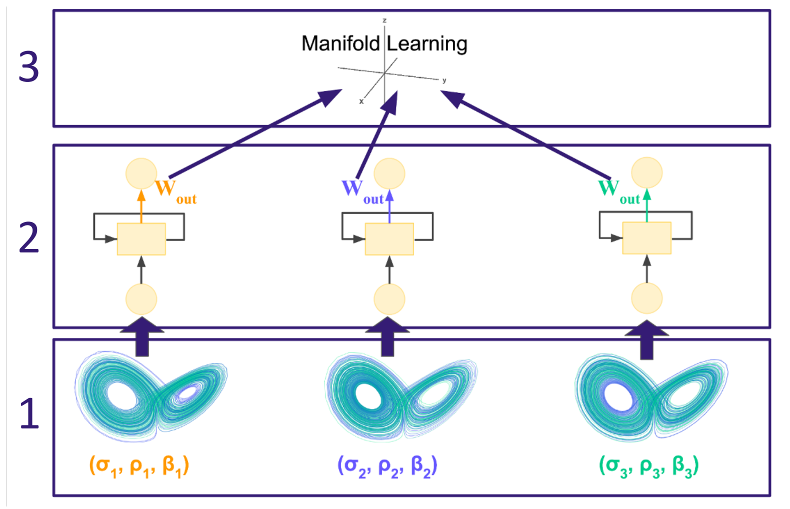
ESN applications in forecasting:
- Improve training stability on chaotic data
- Amplify forecasting performance

New application and approach: *Interpret ESNs to discover varying dynamical parameters in trajectory data.*

Question: *Are dynamical parameters embedded in the readout layer?*

Interpreting ESNs Using Manifold Learning

1. **Collect trajectory data from dynamical system** with uncontrolled or varying dynamical parameters. We use data from the Lorenz system.
2. **Train an ESN on each trajectory**, where each ESN shares the same reservoir so that corresponding readout weights are comparable.
3. **Perform manifold learning** to identify components of the readout layers that correspond to dynamical parameters which differ between trajectories in the data.

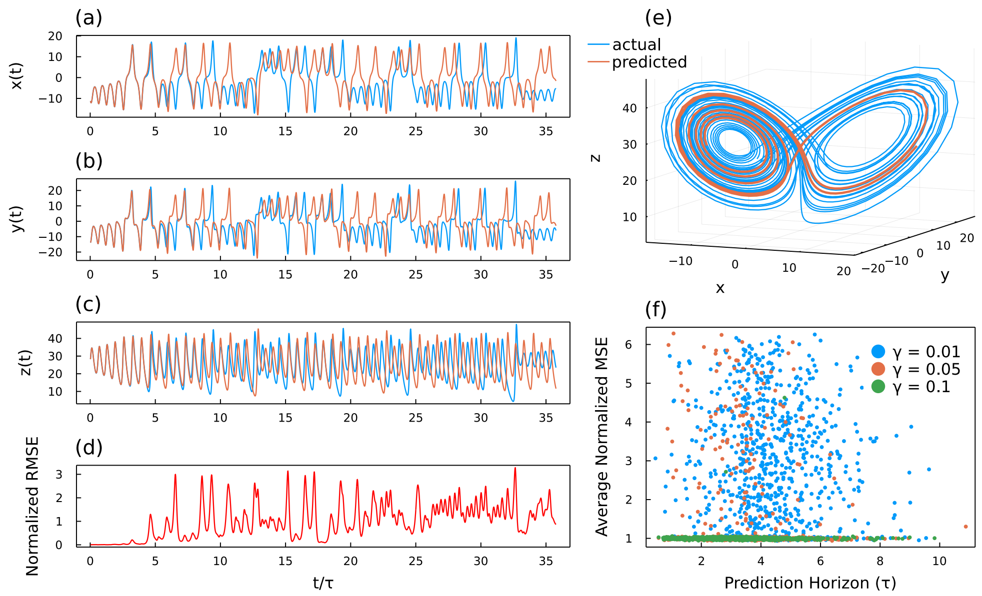


Optimizing Prediction Performance

Lorenz system with 3 varying dynamical parameters σ, ρ, β:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y-x), \quad \frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y, \quad \frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z$$

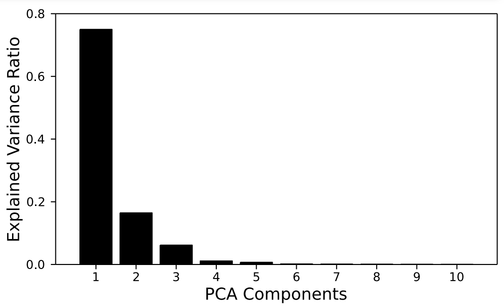
Accurate ESN models of trajectory data necessary for extracting high quality representations of the dynamics from ESN readout layers:
- Prediction performance **(a–e)**:
  Good short-term forecasting (~4 Lyapunov times τ)
- Effect of ridge regression regularization **(f)**:
  Higher regularization → improved long-term stability



Applying Manifold Learning: PCA

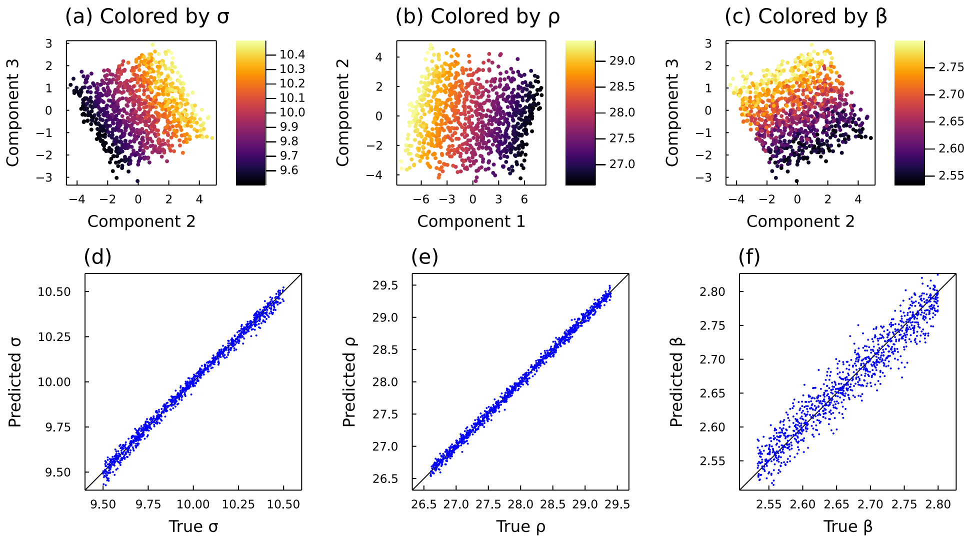
1. **Apply PCA** to flattened ESN readout layer weights.
2. **Identify relevant PCA components** using explained variance ratio.

(total variance threshold: 0.95)



Extracting Dynamical Parameters

PCA components form embedding of dynamical parameters:
- PCA projections **(a–c)**: Relevant components correspond to the varying dynamical parameters σ, ρ, β.
- Fitted parameters vs. ground truth **(d–f)**: Linear fits of the extracted embedding to σ, ρ, β with R^2 = 0.993, 0.996, 0.925, respectively.



Conclusions

We have demonstrated on the Lorenz system that…
- ESN readout layers contain encoding of the dynamical parameters.
- System parameters can be extracted through manifold learning.

Potential applications for ESNs in interpretable modeling:
- Provides an alternative, computationally inexpensive method for unsupervised identification of dynamics.
- Easily applied to any existing ESN prediction models.
- Acts as a generic dynamics encoder for nonlinear dynamics.

Future work:
- Test on high-dimensional data, e.g. from PDEs.
- Adapt for parameter extraction with only partial state information.
- Develop better theoretical understanding of how ESN dynamics are encoded in the readout layer.

We would like to acknowledge useful discussions with Rumen Dangovski, Samuel Kim, Andrew Ma, and Charlotte Loh, and Dr. John Rickert. This work is supported by the U.S. Department of Defense, the National Science Foundation through the Institute for Artificial Intelligence and Fundamental Interactions (IAIFI), and the Center for Excellence in Education (CEE) through the Research Science Institute (RSI) summer program.