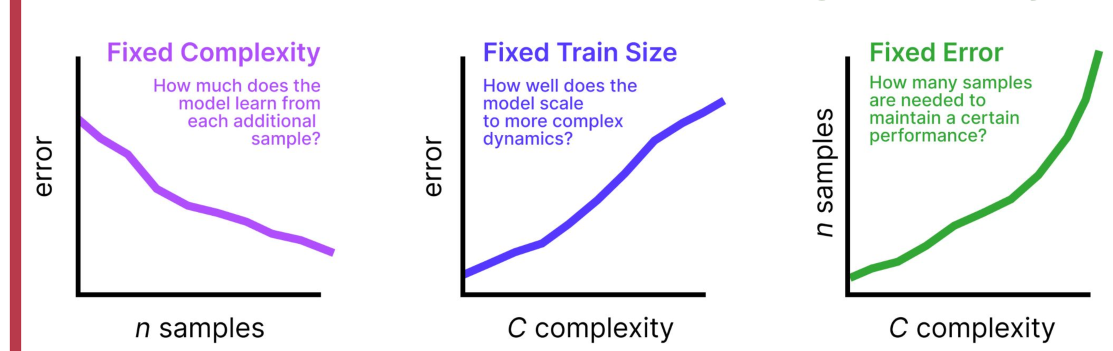
Modeling Chaos: Advancing Benchmarking for Data-Driven Dynamical System Identification

Lauren McLane, Brains in Silicon, Boahen's Lab

Introduction:

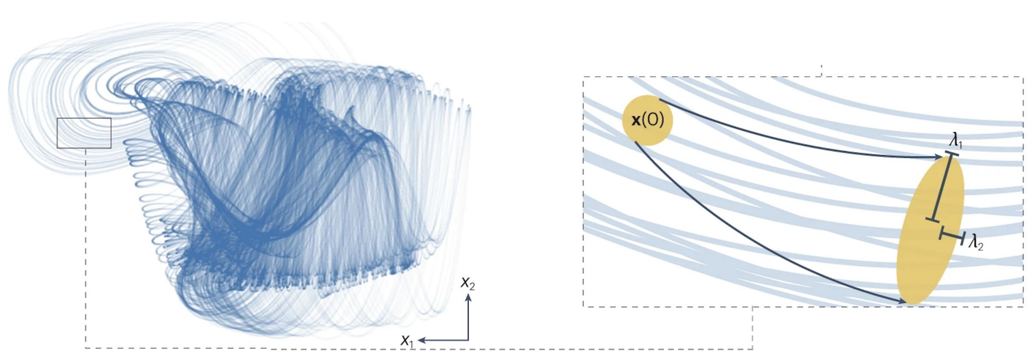
Dynamical systems model evolving processes all around us. While some dynamical systems can be derived from known mathematical equations, like a swinging pendulum, others are too complex or unknown to express analytically, such as the behavior of the stock market. To address this, data-driven approaches use machine learning to extract patterns directly from observed data. However, progress in this field is hindered by the lack of standardized benchmarks.

DynaDojo aspires to be a comprehensive benchmarking platform for dynamical system identification, evaluating algorithms across three key dimensions:



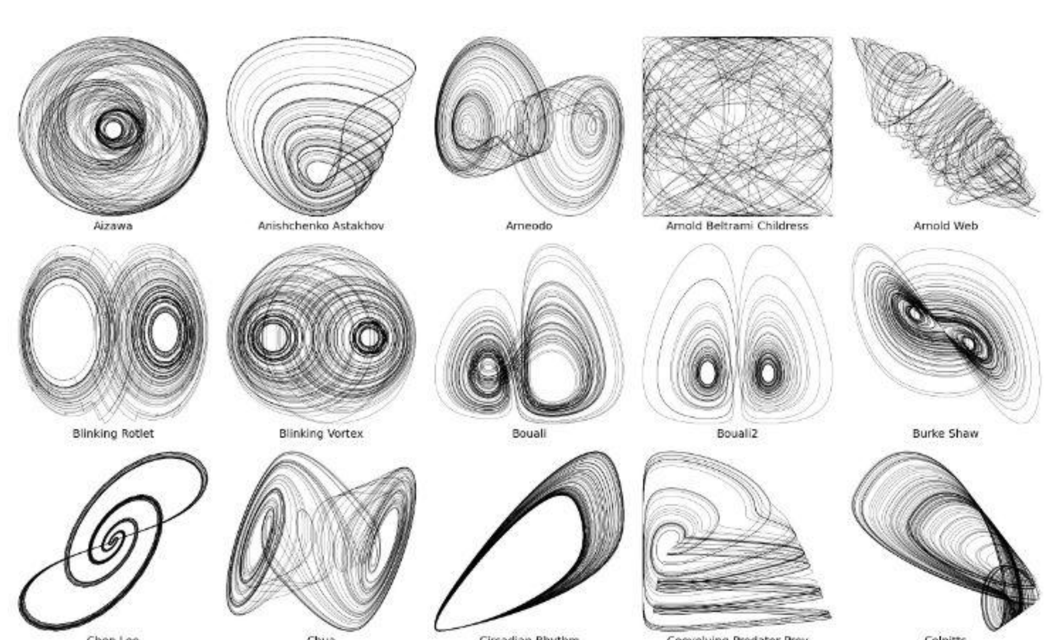
Chaotic Systems

Chaotic systems produce complex and unpredictable behavior that challenge modern forecasting methods, while retaining deterministic properties that make them controllable and interpretable as benchmarks.



The divergence of two trajectories in a chaotic system, described by Lyapunov exponents (Gilpin, 2024).

Drawing from an open-source library of chaotic systems, over 130 new continuous chaotic systems were integrated into DynaDojo's library of 20 dynamical systems. Each system can be simulated with customizable length, granularity, and number of trajectories.



(Gilpin, 2024)

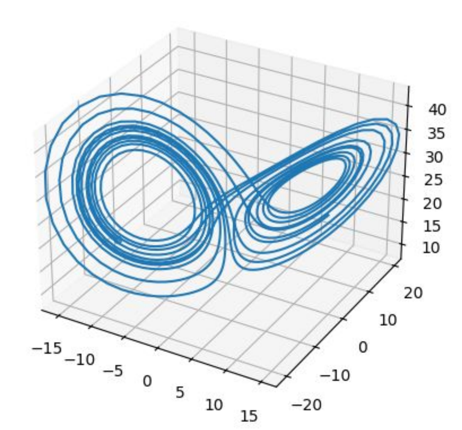
Static Datasets

Ultimately, we are interested in applying forecasting algorithms to real datasets. By, treating a static dataset as the instance of a system with the most trajectories, longest timesteps, and finest granularity, we can apply scaling benchmarks on real data.

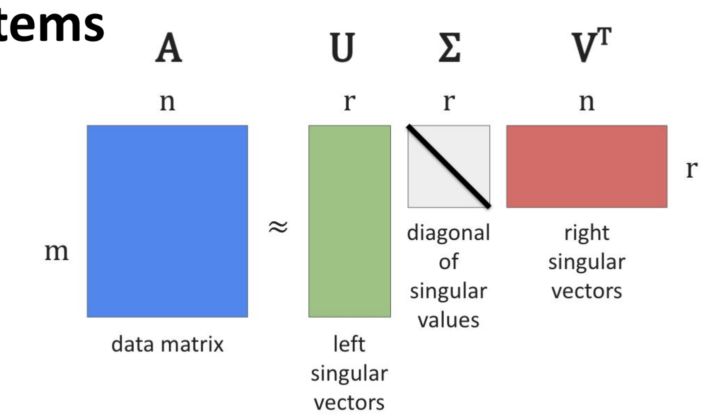
Out-of-Distribution (OOD) Generalization

Out-of-distribution (OOD) generalization measures how well an algorithm performs on new, unseen scenarios that differ from the data it was trained on. To study OOD generalization for the newly integrated systems, we develop methods for OOD generation and categorization.

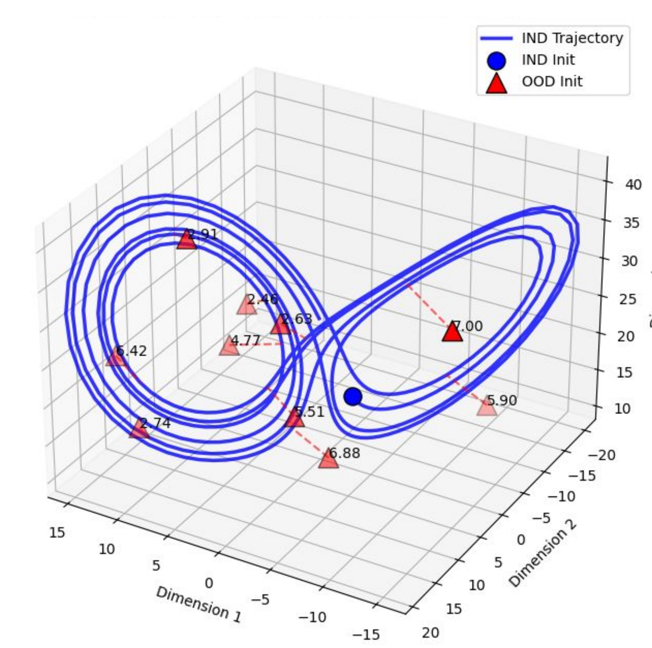
OOD Generation for Chaotic Systems



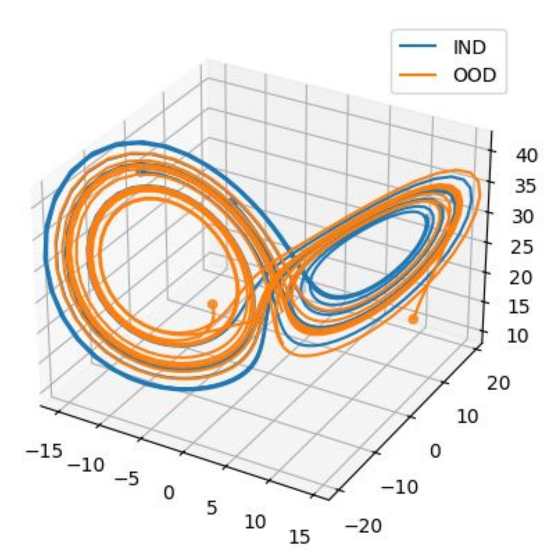
Generate a single sample trajectory of 1000 timesteps.



Perform singular value decomposition to determine the direction(s) of least variance, given by the right singular vectors corresponding to the smallest singular values.

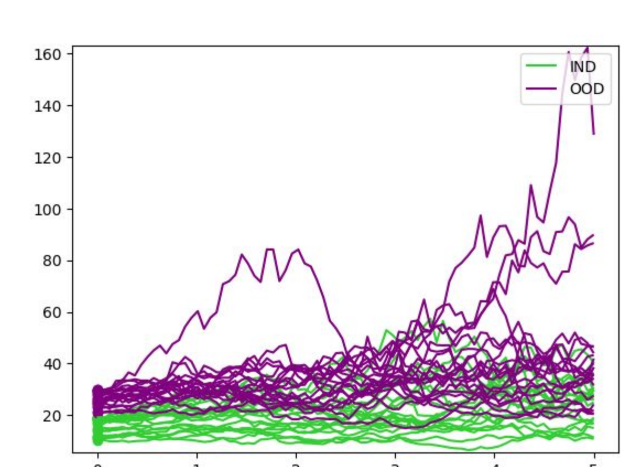


Select random points in the sample trajectory. Perturb them in the direction(s) of least variance.

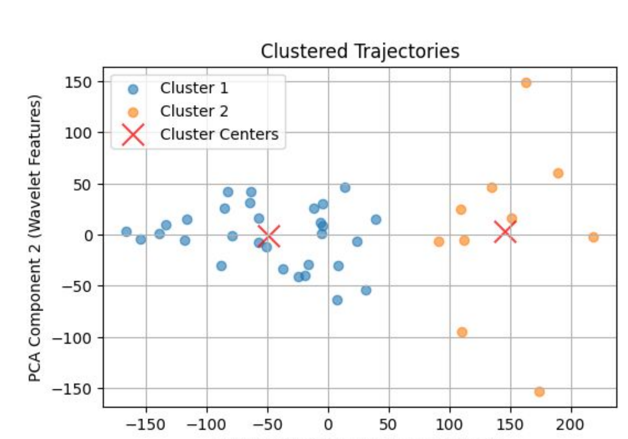


Plot full in-distribution (IND) and OOD trajectories.

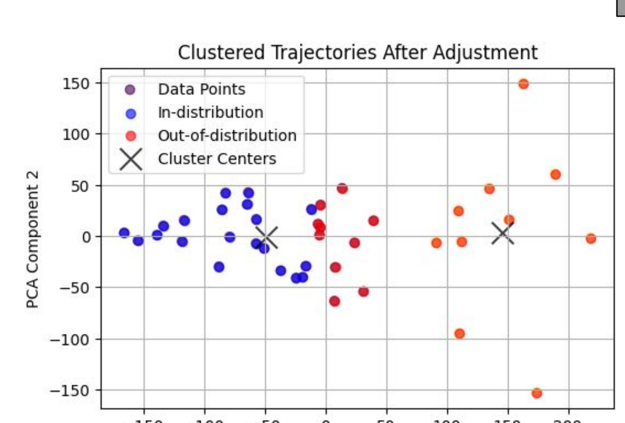
OOD Categorization for Static Datasets



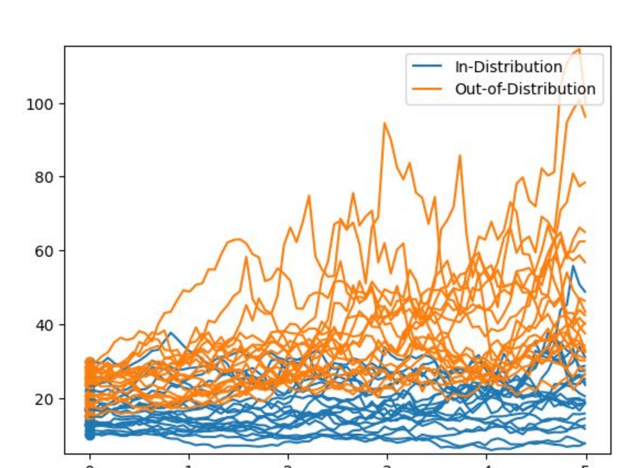
Dynadojo defined IND and OOD trajectories for Black-Scholes-Barenblatt PDE System



Extract Haar wavelet features and cluster with K-means. Let the larger cluster be IND points.



Move trajectories from one cluster to the other until desired ratio is achieved (in this case 1:1).



Plot newly defined IND and OOD trajectories.

Tuning System Complexity

To study how algorithms handle systems of increasing complexity, we must understand how complexity changes as we tweak a system's parameters. We seek an algorithm that predicts the complexity of the system across the entire parameter space with as few samples as possible.

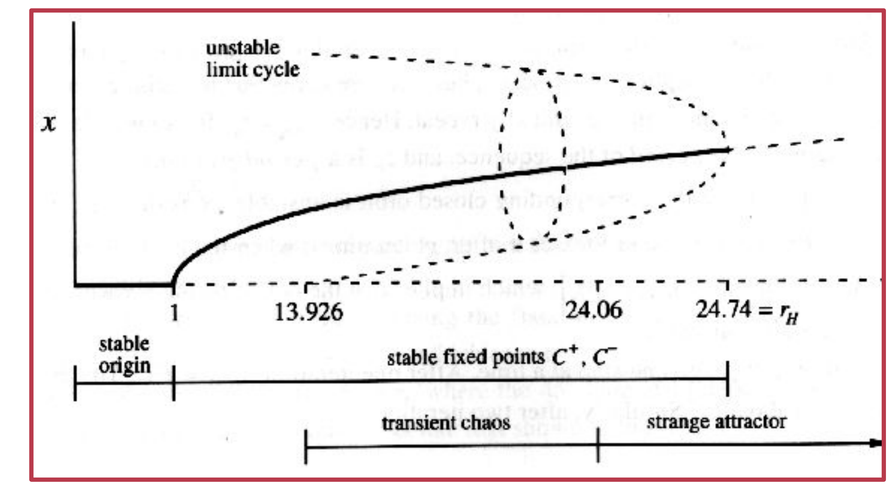
Complexity Measures:

- ***The Lyapunov exponent*** tells us the rate of divergence of nearby trajectories.
- ***The correlation dimension (CD)*** estimates the fractal dimension of the attractor.
- ***Multiscale entropy (MSE)*** calculates the unpredictability or irregularity of a system over multiple time scales.

Toy Implementation: Lorenz System and Correlation Dimension

$$\frac{dx}{dt} = \sigma(y - x),$$
$$\frac{dy}{dt} = x(\rho - z) - y,$$
$$\frac{dz}{dt} = xy - \beta z.$$

Let σ=10, β=8/3, vary ρ (or r)

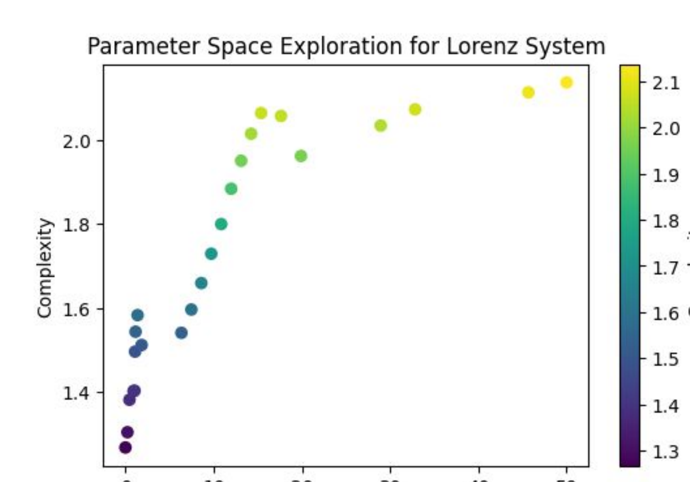


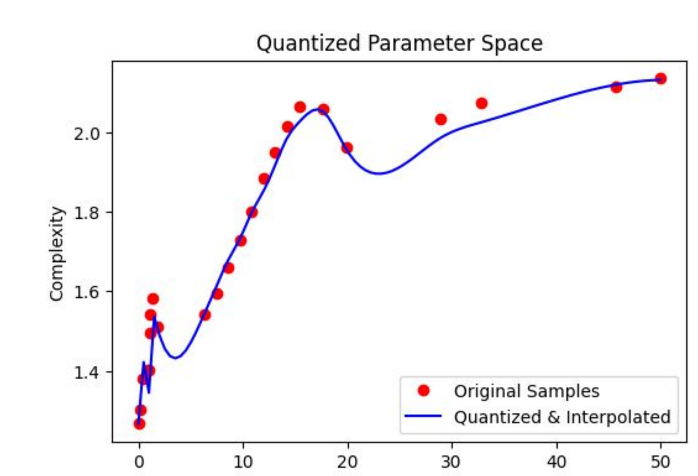
Bifurcations give us intuition for when a particular measure of complexity might change.

We anticipate the correlation dimension to increase during the homoclinic bifurcation (r=13.926), when the fractal structure of the Lorenz attractor begins to emerge.

Adaptive Sampling & Quantization Algorithm

- Generate initial samples
- Compute complexity for each sample.
- Detect significant jumps in complexity and refine sampling
- Iterate until improvement is minimal.
- After sampling, quantize the function using wavelets and interpolate the quantized complexities.

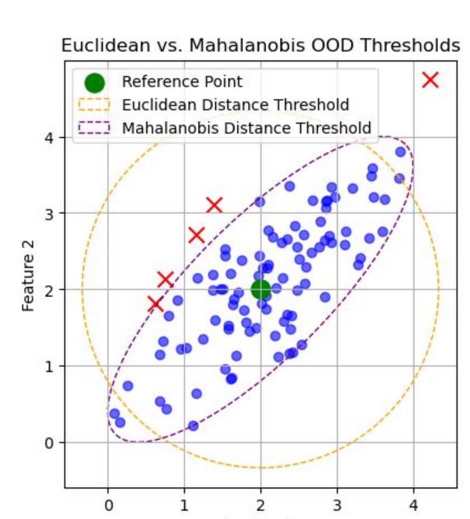




Evaluations & Next Steps

OOD Generation

- Criteria are needed to assess whether generated OOD points meaningfully deviate from in-distribution data.
- **Next:** Use methods like Mahalanobis distance, which accounts for the data's covariance, to evaluate OOD points.



Lorenz Adaptive Sampling and Quantization

- The adaptive algorithm sampled 24 points in total and 4 points in the 0 to 1 range, where complexity rises sharply.
- A grid search over the same range would require 200 samples to achieve a comparable understanding of system behavior between 0 and 1.
- Potential Flaws: May miss subtle complexity transitions in higher dimensions or gradual changes.
- **Next:** Use methods like active learning or Bayesian optimization for global exploration in the high-dimensional spaces, before switching to a refinement approach.

Acknowledgements

I want to thank Max Kanwal for taking me on as a research student and mentoring me this summer. Not only have I learned so much about this fascinating field of ML, but I have also learned how to approach research problems thoughtfully and rigorously. Thank you also to Victor, Vincent, and Hadil for being kind and dedicated labmates.

References

Bhamidipaty, L. M., Bruzzese, T., Tran, C., Ratl Mrad, R., & Kanwal, M. S. (2023). *DynaDojo: An extensible platform for benchmarking scaling in dynamical system identification*. *Advances in Neural Information Processing Systems, 36*, 15519–15530.

Gilpin, W. (2021). *Chaos as an interpretable benchmark for forecasting and data-driven modeling*. NeurIPS Datasets and Benchmarks Track.

Gilpin, W. (2024). Williamgilpin/dysts [Jupyter Notebook].

Gilpin, W. (2024). *Generative learning for nonlinear dynamics*. *Nature Reviews Physics, 6*(3), 194–206.

Strogatz, S. H. (2015). *Nonlinear dynamics and chaos* (2nd ed.). Westview Press.