Reconstructing Chaotic Dynamics with Delay Embedding and Gaussian Process Regression: Applications to Ecological Biomass Data

Joey Park[1], Mentor: Prof. Sarah Day[2]
[1]Thomas Jefferson High School for Science and Technology | [2]Department of Mathematics, William & Mary
joyful.joeypark@gmail.com | sldayx@wm.edu

Introduction & Objective

Chaotic dynamics appear in many natural systems, including ecological populations, where small perturbations cause complex long-term behavior. To recover such dynamics from limited data, we combine **Takens' delay embedding** and **Gaussian Process Regression (GPR)**. Delay embedding reconstructs the hidden state space from a single time series using

$$y_t = [x_t,\, x_{t-\tau}, \ldots,\, x_{t-(Q-1)\tau}],$$

where Q is the embedding dimension (number of delayed coordinates) and τ the time lag between successive coordinates, unfolding the system's trajectory into a higher-dimensional space that reveals the attractor's geometry. GPR then learns the mapping $y_t \rightarrow x_{t+1}$

$$x_{t+1} \sim GP(m(yt), k(yt, yt')),$$

where m and k are the mean and kernel functions that capture smooth, nonlinear relationships in the reconstructed space. We test the method on logistic and Hénon maps, then apply it to Salvinia molesta biomass to assess its ability to recover meaningful ecological dynamics.

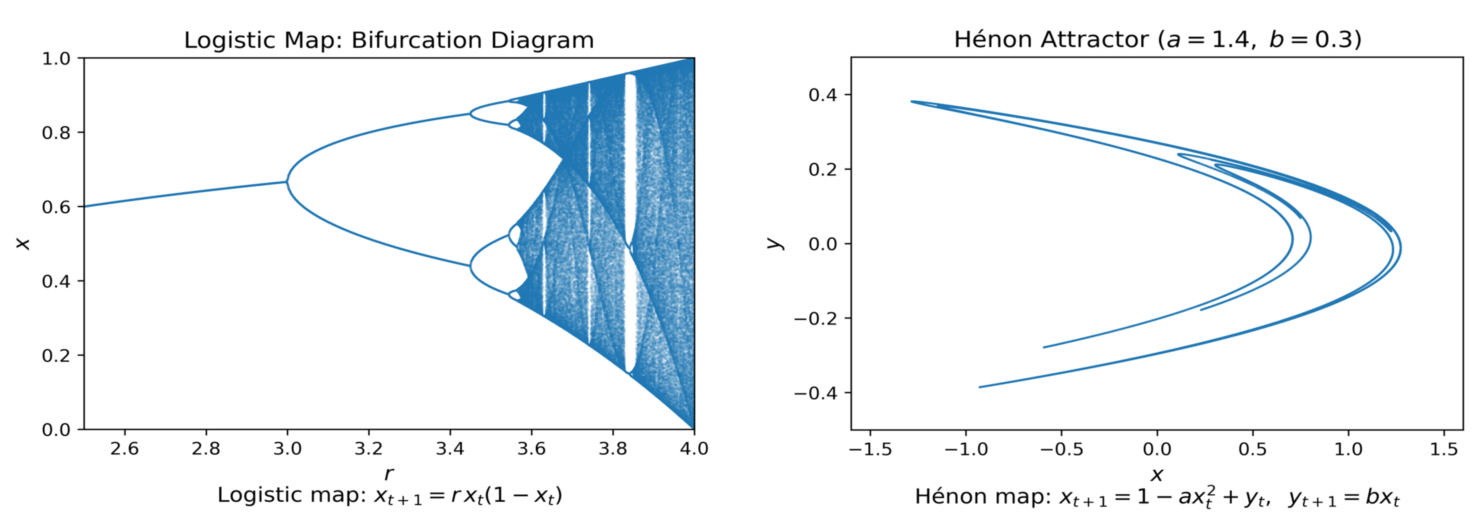


Figure 1. Representative chaotic systems used for testing: the logistic map (left) and the Hénon attractor (right).

Method

For the logistic map, we applied delay embedding + GPR to test whether the model can reproduce the long-term dynamics. We then varied noise levels and the GPR hyperparameter (α) to assess how data quality and regularization influence reconstruction accuracy.

For the Hénon map, we tested several measurement functions $h(x,y)$, generating scalar time series $s_t = h(x_t, y_t)$, and selected the one that best preserved the attractor's shape. We also varied the delay (τ), sample size, and noise to examine how embedding choices and data quality influence the reconstructed geometry within the same framework.

For the ecological data, we applied the same framework to Salvinia molesta biomass time series to test whether it can recover meaningful nonlinear patterns from limited and noisy real-world observations.

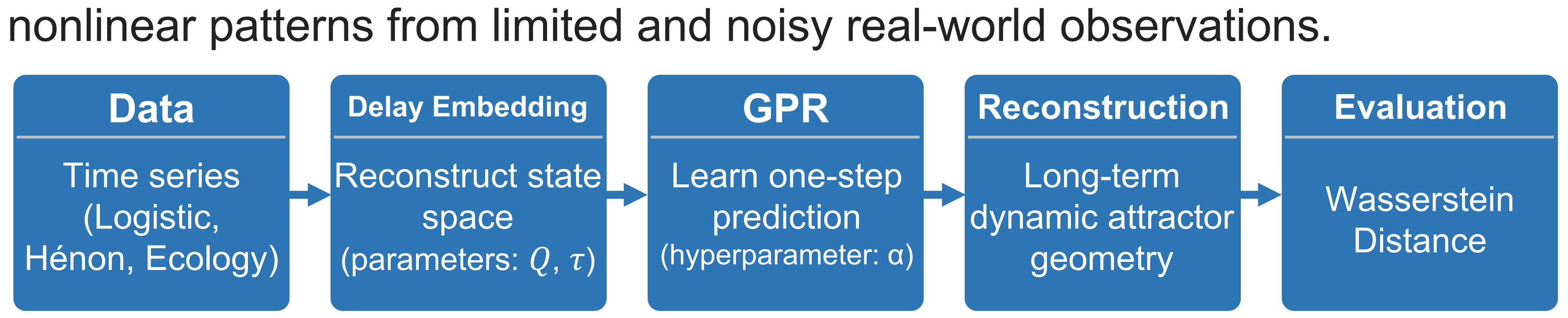
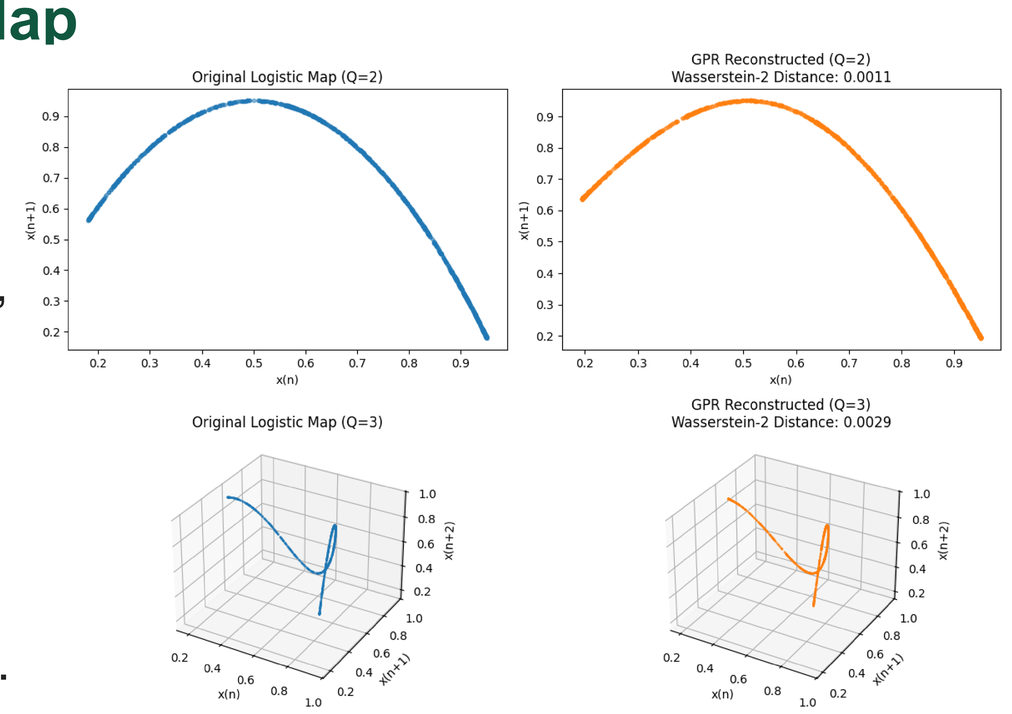
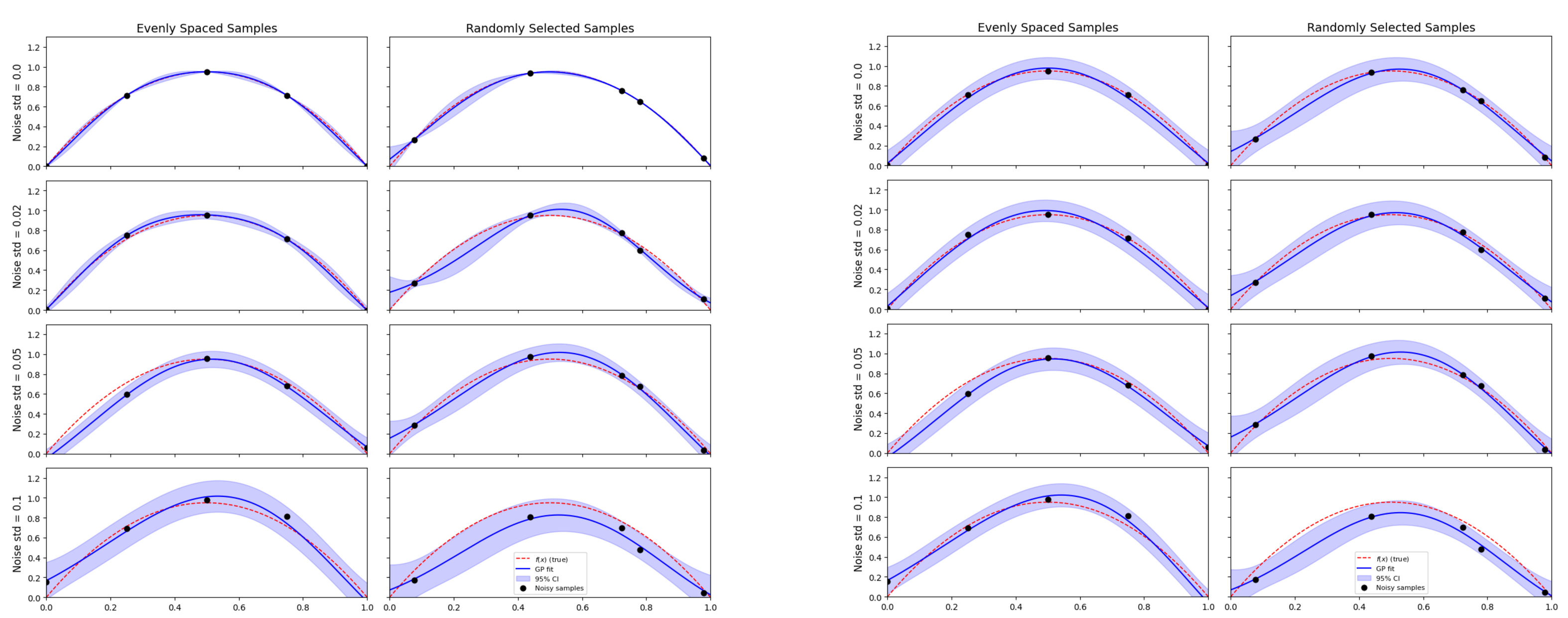


Figure 2. Workflow of the delay embedding and Gaussian Process Regression (GPR) framework used to reconstruct and evaluate system dynamics.

Result – Logistic Map

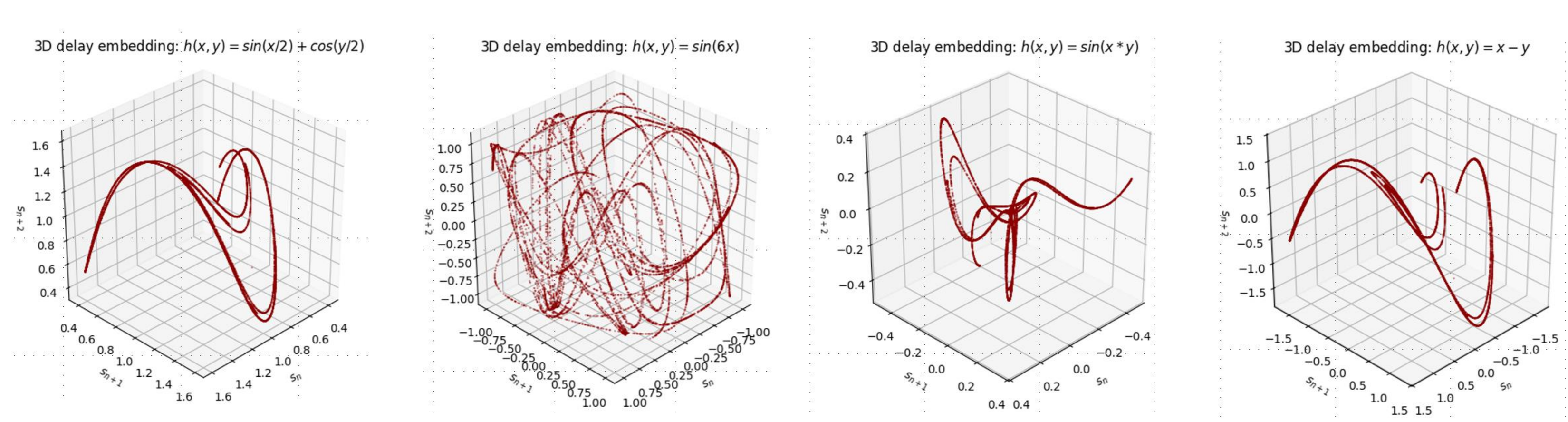
(a) Using a time series of length 10, we applied delay embedding in two and three dimensions (Q=2,3) with τ=1, generating 1,000 predicted points with GPR. The reconstructed trajectories closely matched the original logistic map, showing very small Wasserstein distances.



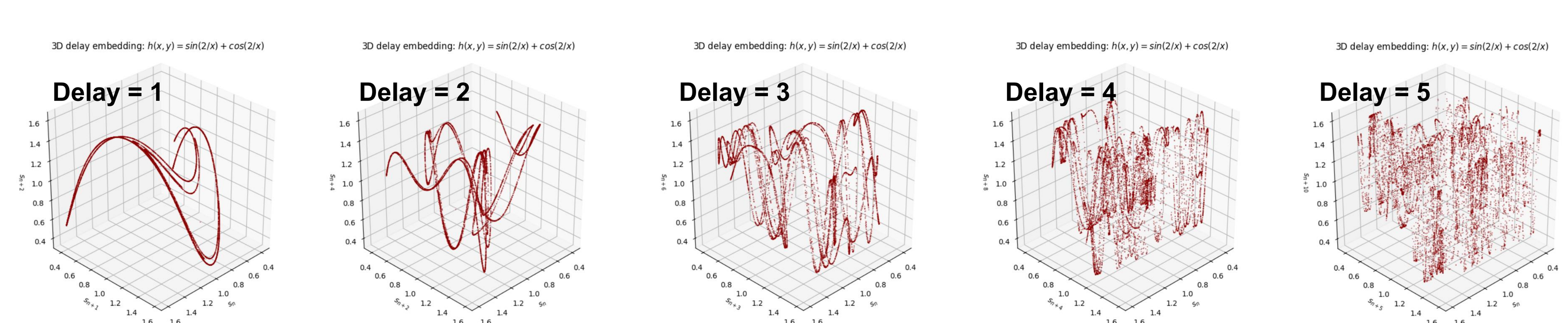


(b) Gaussian Process Regression (GPR) on the logistic map using five training points. Both panels use the same noise levels ($\sigma \in \{0, 0.02, 0.05, 0.1\}$). The left column adjusts the GPR noise hyperparameter $\alpha = \sigma^2$, which fits the data closely but overfits when both σ and α are small. The right column fixes α=0.005, producing smoother and more consistent fits across all noise levels, preserving the global shape while reducing overfitting.

Result – Hénon Map

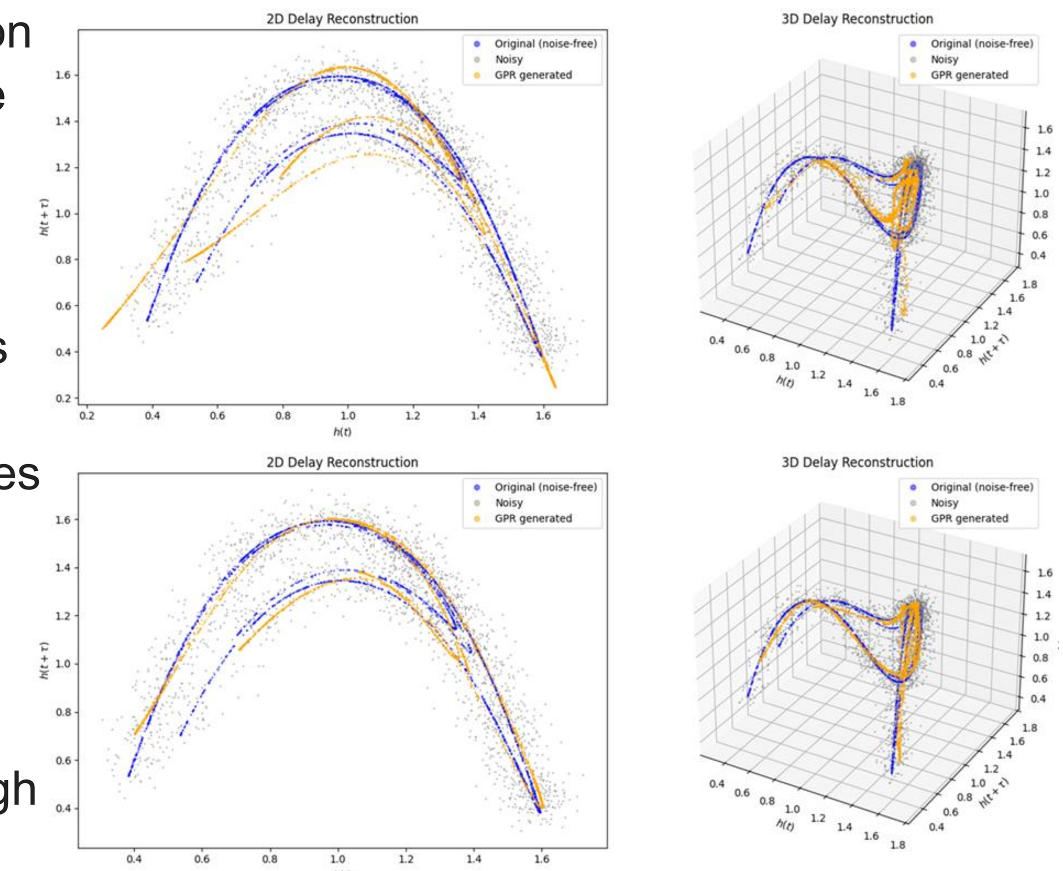


(a) 2D and 3D delay embeddings of the Hénon map using four measurement functions with τ=1. From top to bottom: $h(x,y) = \sin(x/2)+\cos(y/2)$, $\sin(6x)$, $\sin(xy)$, $x-y$

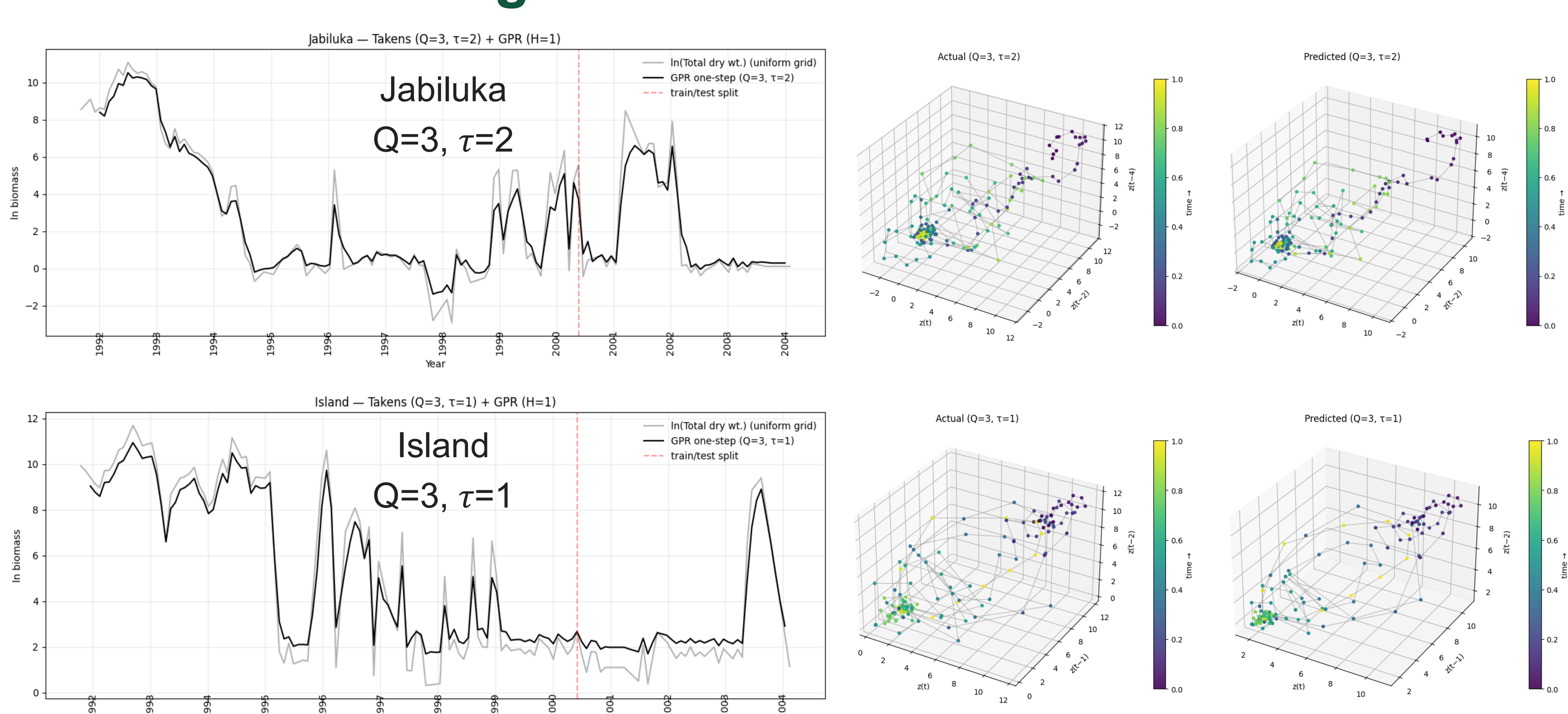


(b) Using $h(x,y) = \sin(x/2)+\cos(y/2)$, and embedding dimension Q = 3, small delays (τ = 1, 2) preserve the attractor's shape, while larger delays distort the geometry.

(c) Effect of sample size on GPR reconstruction of the Hénon map (σ=0.05, α=0.005). The top panels use time series of length 40, and the bottom panels use time series of length 200. Wasserstein distances were 0.3633 (2D) and 0.4027 (3D) for length-40 series, decreasing to 0.2642 and 0.2807 with length-200 series. Although reconstruction accuracy improves with longer datasets, even a length-40 series reproduces the original attractor's trajectory with high visual similarity.



Result – Ecological Data



Applying the framework to Salvinia molesta biomass at the Jabiluka and Island sites. Using 70% of each series for training, the models reconstructed nonlinear biomass dynamics with Wasserstein distances of 3.41 (Q=3, τ=2) at Jabiluka and 7.40 (Q=3, τ=1) at Island. Despite noise and limited samples, the reconstructions captured key nonlinear patterns and overall trends, demonstrating that the framework generalizes from synthetic systems to real ecological data.

Discussion & Conclusion

Across both the logistic and Hénon maps, the delay embedding + GPR framework reconstructed chaotic dynamics well, even with limited data. Applied to Salvinia molesta biomass, the method captured nonlinear patterns despite noise, suggesting potential for ecological modeling. Future work should explore a broader range of embedding dimensions Q and delays τ to refine reconstruction and improve model reliability.

Acknowledgement

I sincerely thank **Professor Sarah Day** for her guidance and mentorship throughout this research. I also appreciate Dr. Anthony Ives, Dr. Bill Kalies, Alethea Callahan, and Anaiel Dayrit for their helpful discussions.