Harmful Algal Blooms Forecasting

Athulith Paraselli, Ciro Zhang,

Esther Chung, Nian-Nian Wang

Contents

Background: What is an algal bloom?

- Rapid growth of algae in coastal waters
- Can discolor water, reduce oxygen, and release toxins
- Off the shores of Southern California, algae periodically blooms out of control

Predicting blooms even one week in advance helps:

- Public health officials respond to health concerns
- Coastal facilities shut down before equipment is damaged

Most common algae species in SoCal: Lingulodinium polyedrum

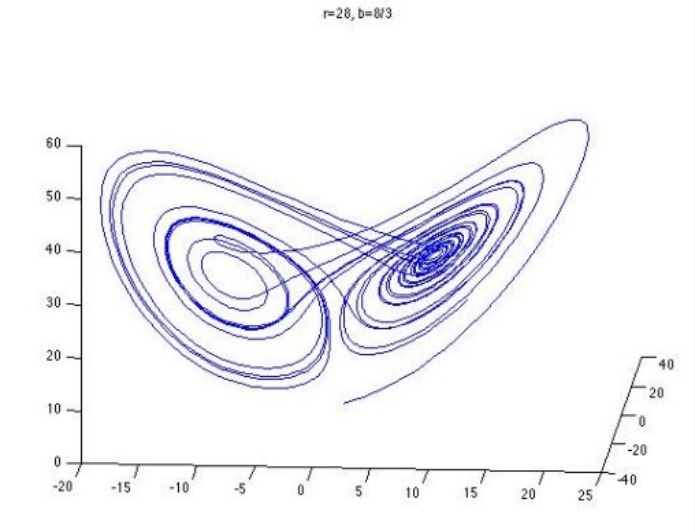
- 50% of algae samples at Scripps are Lingulodinium

When disturbed, this algae releases bioluminescent light

Predicting Lingulodinium blooms = Predicting when bioluminescent waves will be visible

Motivation

- Blooms are driven by **many interacting factors** (nutrients, salinity, wind, plankton communities)
- The system is **nonlinear and chaotic** — small changes can cause large effects
- Traditional statistical models often fail to capture this complexity
- Data is **noisy and limited**, especially for species-specific events like Lingulodinium
- Forecasting requires balancing **mechanistic understanding** with **predictive modeling**



Dataset: Carter dataset

- **Acquired from:** Prof. Melissa Carter, Scripps Institution of Oceanography
- **Time span:** June 2008 – December 2024 (16.5-year span)
- **Frequency:** Weekly
- **Size:** 823 observations

Columns include:

- Avg Chlorophyll
- Chloro 1, Chloro 2
- Salinity (Bottom, Surface, Average)
- Density (Bottom, Surface, Average)
- Temperature (Bottom, Surface, Average)
- Nitrate, Nitrite, Phosphate, Silicate
- Wind speed
- Lingulodinium (cell counts)
- Phaeo

Methods

01

Convergent Cross Mapping (CCM)

- Identifies causal relationships among environmental variables
- Filters out irrelevant predictors to focus on true drivers of blooms

02

Empirical Dynamic Modeling (EDM)

- Captures nonlinear ecological dynamics from time series data
- Builds ensembles of local models to represent system behavior

03

Temporal Convolution + Attention (Trust Gate)

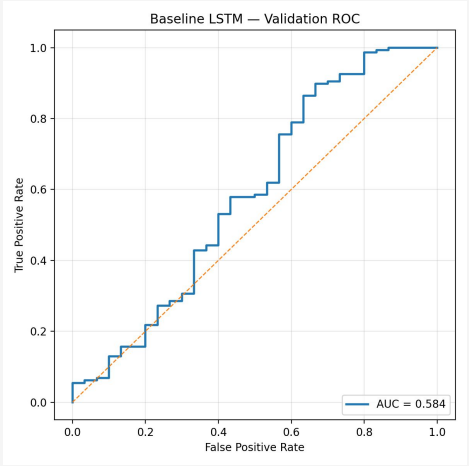
- Trust gate / attention mechanism highlights the most predictive features
- Convolutional layers extract short-term patterns in EDM embeddings

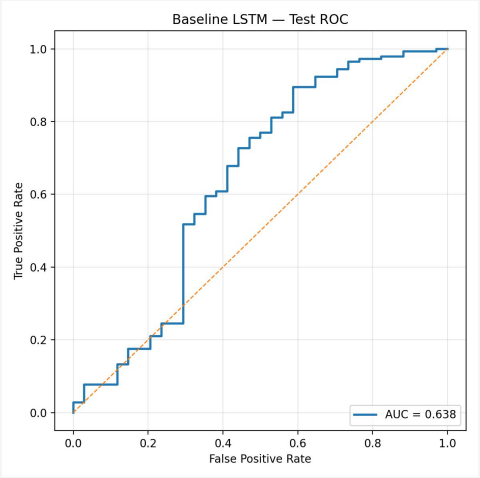
04

Long Short-Term Memory (LSTM)

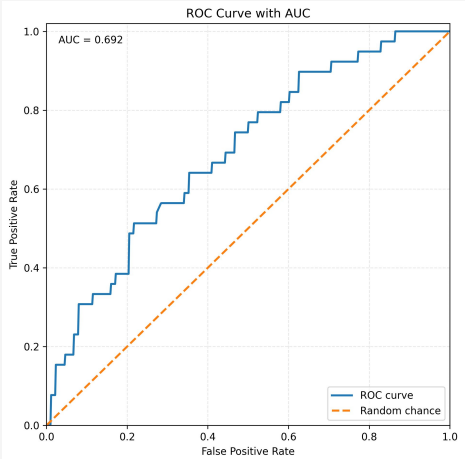
- Models long-term temporal dependencies and chaotic dynamics
- Produces bloom vs no bloom forecasts (bloom = above 95th percentile)

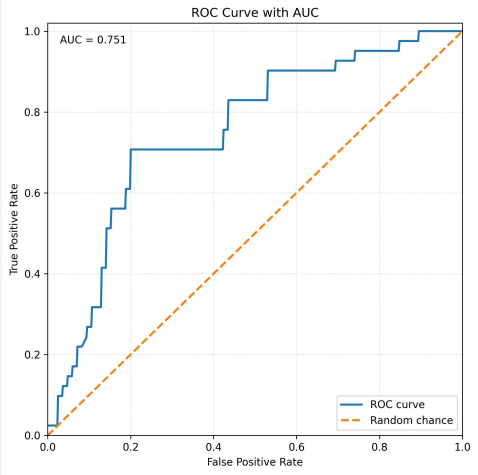
Raw LSTM Results





EDM-LSTM Results





Results: Pure-LSTM vs Hybrid EDM-LSTMs