# Machine Learning approaches in Ecological Forecasting

Ethan Deyle

erd.carrd.co/

BU Marine Program

UC San Diego | SCRIPPS INSTITUTION OF OCEANOGRAPHY

NSF

SERDP DOD · EPA · DOE

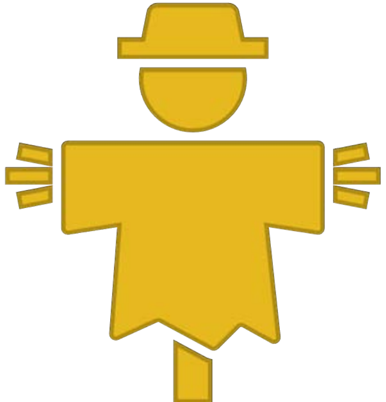
# Outline

- Myths and Merits.
- Definitions and history.
- Application to red tide forecasting.
  - Time-scales.
  - Limitations (e.g. bias)
- Application to hypoxia.
  - Non-stationarity.
  - Actionable and predictive insight into mechanism.
  - Hybrid approaches: incorporating first-principles understanding when at hand; structural agnosticism.
- Closing Thoughts

# Myths about Machine Learning

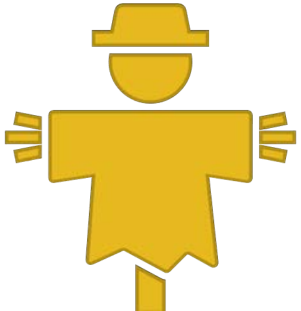
## Myths

- Machine learning is new and untested.
- Machine learning approaches are a black box.
- Machine learning approaches throw away our first-principles understanding of systems.
- Machine learning approaches are complicated.



## Merits

- Machine learning is not new to ecological forecasting and has a 30-year track-record.
- Can yield actionable and predictive insight into mechanism
- Can be part of forecasting non-stationary and non-equilibrium futures.
- Can be minimally assumptive and surprisingly unsophisticated!



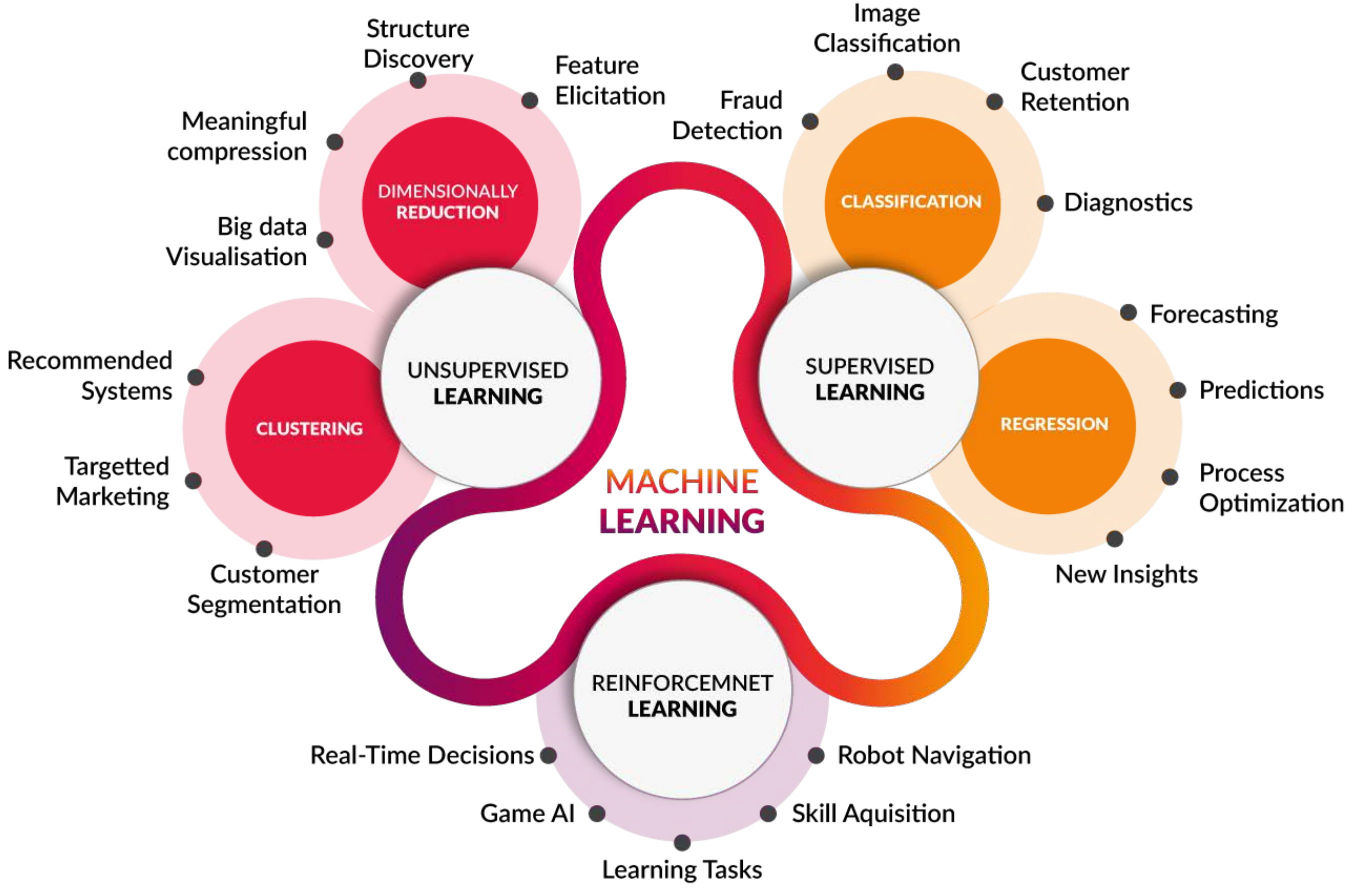
Machine Learning is new and untested.

Machine learning has a 30-year track-record in ecological forecasting (including marine and coastal systems!).

…Depending on what you mean by machine learning.

# Definitions and History

# What do we we mean by “machine learning”?




“Use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.”
“Through the use of statistical methods, algorithms are trained to make classifications or predictions.”

– IBM.com

Supervised: labeled *inputs* and *outputs*.

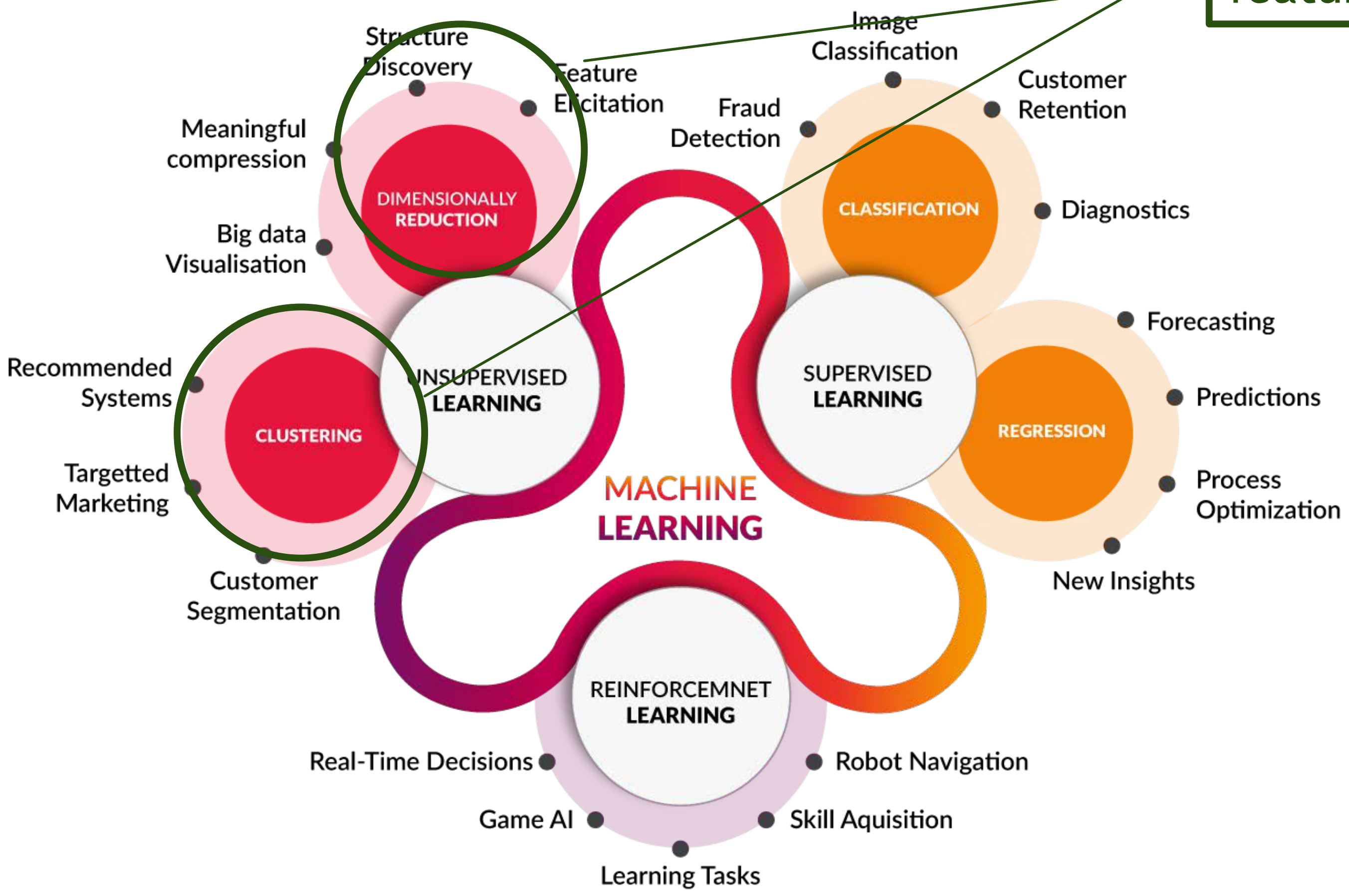
- *Predictors vs. predictees*
- Categories

Unsupervised: eliminates human intervention.

http://www.cognub.com/index.php/cognitive-platform/

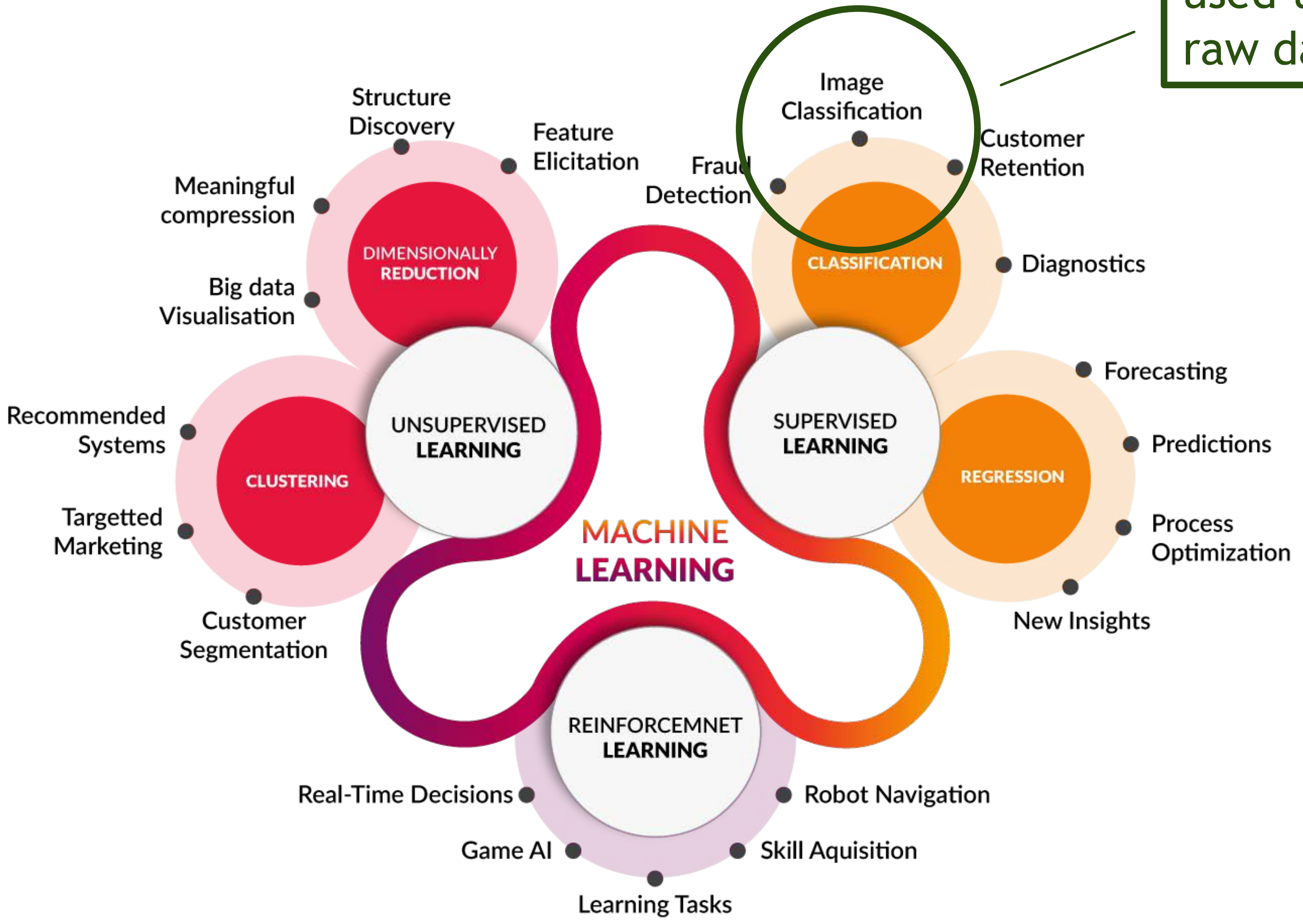
# What do we we mean by “machine learning”?

Gulf-stream meander: Unsupervised self-organizing map (ANN) to do feature selection and clustering.




http://www.cognub.com/index.php/cognitive-platform/

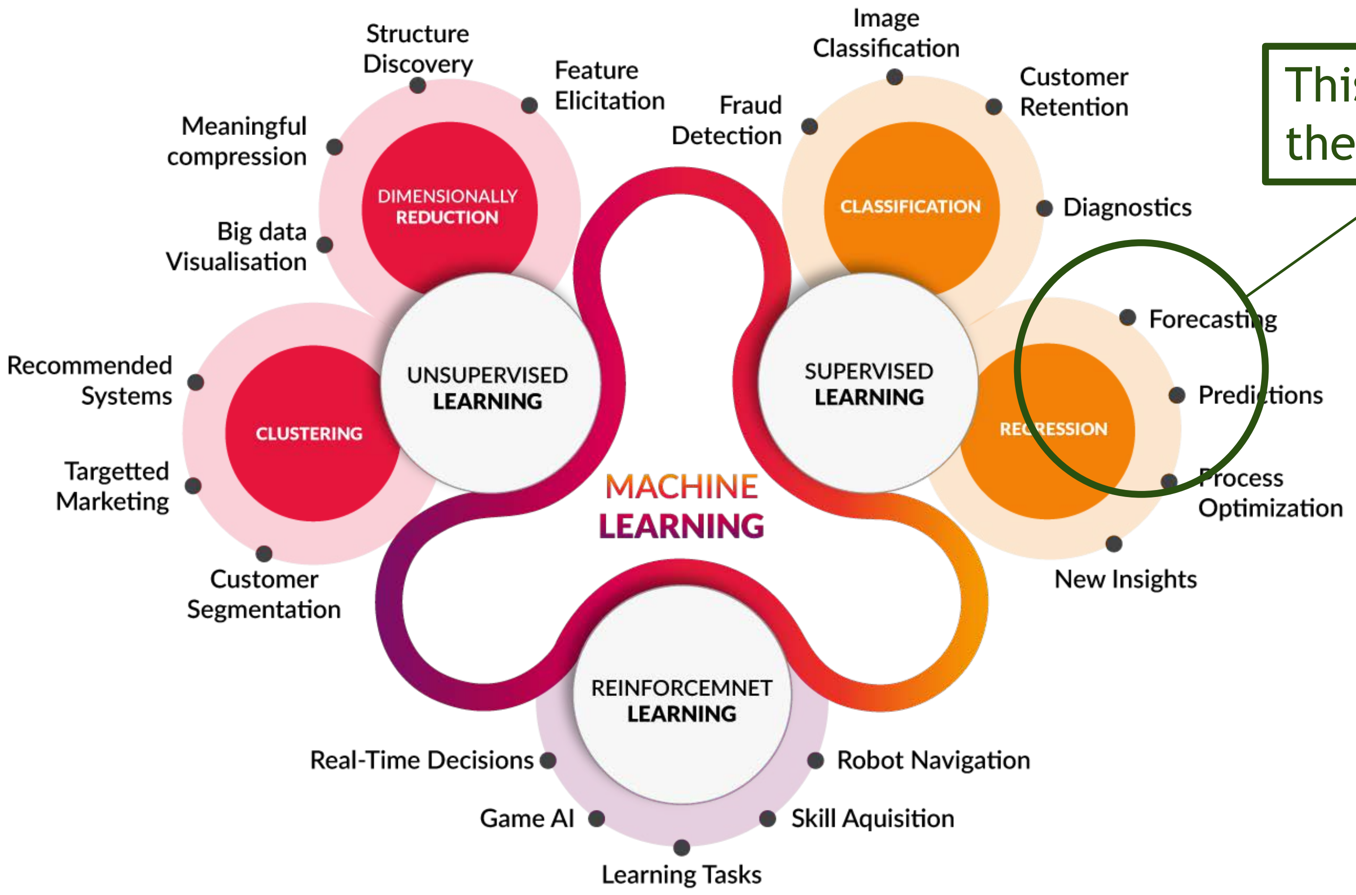
# What do we we mean by “machine learning”?




Cytobot: Image classification used to support the pipeline from raw data to models.

http://www.cognub.com/index.php/cognitive-platform/

# What do we we mean by "machine learning"?




This talk: Machine learning to do the actual forecast.

http://www.cognub.com/index.php/cognitive-platform/

# Sugihara & May 1990. Sugihara 1994.

- If ecological measurements are random variations within an equilibrium system, they shouldn’t be predictable!
- W.E. Allens diatom counts from the end of Scripps Pier (1929-1939).
- Simplex projection: nearest neighbor forecasting.
- S-map: locally weighted linear regression (kernel regression).

Published: 19 April 1990

**Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series**

George Sugihara & Robert M. May

*Nature* **344**, 734–741(1990) | Cite this article



Article

**Nonlinear forecasting for the classification of natural time series**

George Sugihara

Published: 15 September 1994 | https://doi.org/10.1098/rsta.1994.0106

# Sugihara & May 1990.

- Simplex projection: knn (nearest neighbor) forecasting with a single parameter, *E*.
- Predict first-differences in weekly counts (remove persistence).
- Autoregressive linear predictor **ρ** = 0.13
- Significant forecast skill to 2-weeks.

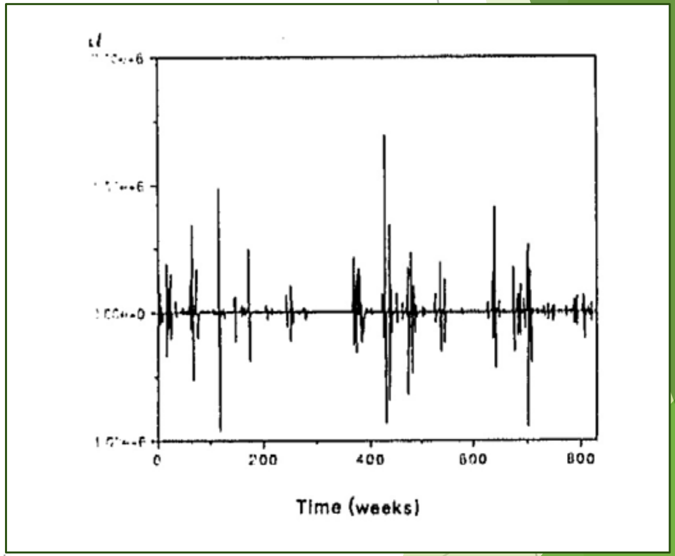
Published: 19 April 1990

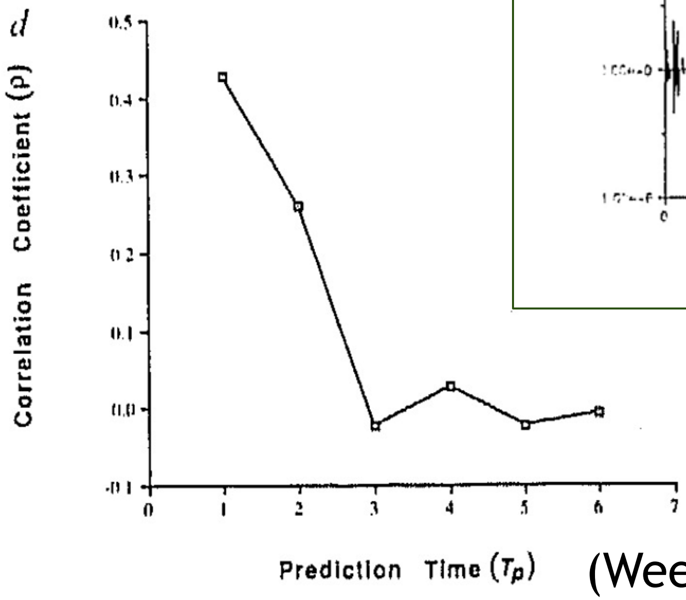
**Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series**

George Sugihara & Robert M. May

*Nature* **344**, 734–741(1990) | Cite this article

**1976** Accesses | **1151** Citations | **11** Altmetric | Metrics




(Weeks)

# Sugihara 1994.

- Sugihara 1994: S-map.
- 1 additional tunable parameter.
- Explicitly compare multivariate linear predictor to a nonlinear predictor.

$$\hat{Y}_t = \sum_{j=0}^{m} \boldsymbol{C}_t(j)\boldsymbol{X}_t(j).$$

$$\boldsymbol{B} = \boldsymbol{A}\boldsymbol{C},$$

$$B_i = w(||X_i - X_t||)Y_i, \quad A_{ij} = w(||X_i - X_t||)\boldsymbol{X}_i(j)$$

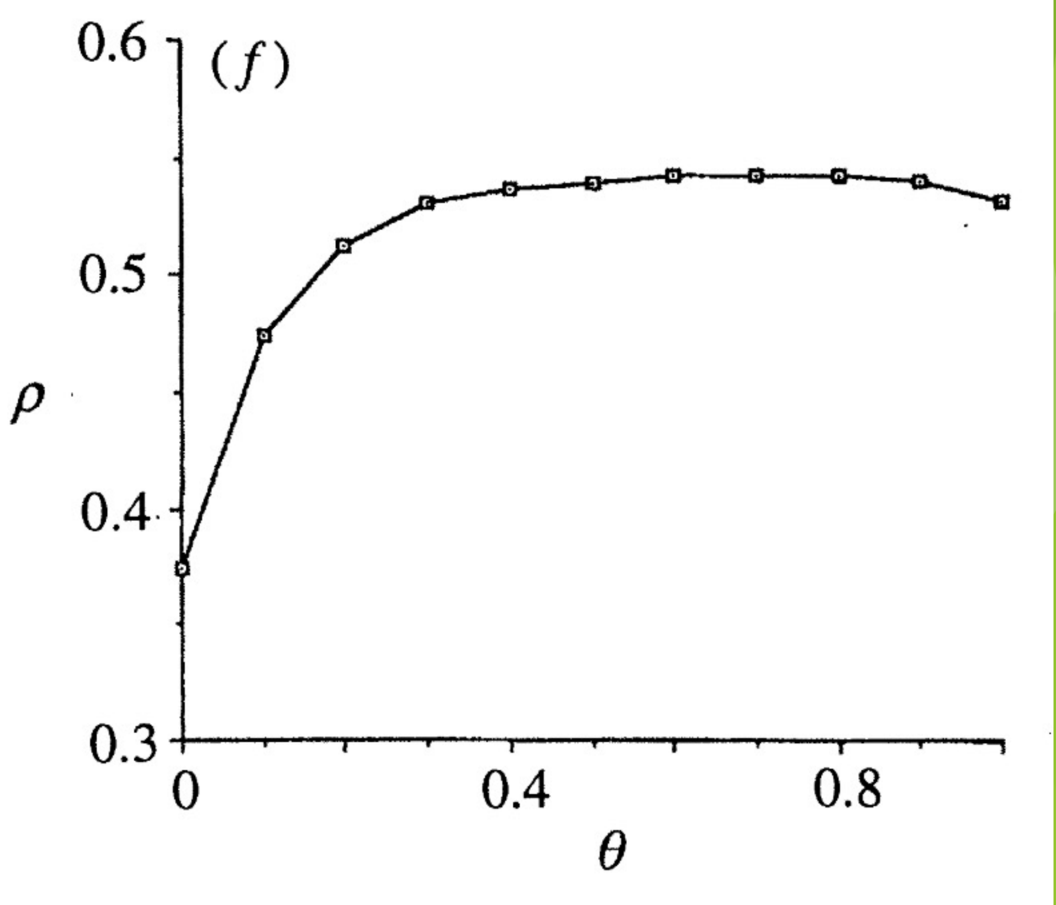
$$w(d) = \mathrm{e}^{-\theta d/\bar{d}},$$

Article

**Nonlinear forecasting for the classification of natural time series**

George Sugihara

Published: **15 September 1994** | https://doi.org/10.1098/rsta.1994.0106



# Think of “supervised Machine learning for forecasting” as universal function approximation.

- “Machine-learning”
- “Non-parametric”
- “Model-free”
- “Non-structural”
- “Empirical model”

*Ecological Applications*, 27(7), 2018, pp. 2048–2060


Prediction in ecology: a first-principles framework

Michael C. Dietze[1]

*Department of Earth and Environment, Boston University, 685 Commonwealth Avenue, Room 130, Boston, Massachusetts 02215 USA*

$$Y_{t+1} = f(Y_t, X_t | \bar{\theta} + \alpha) + \varepsilon_t \quad (1)$$

Universal function approximators

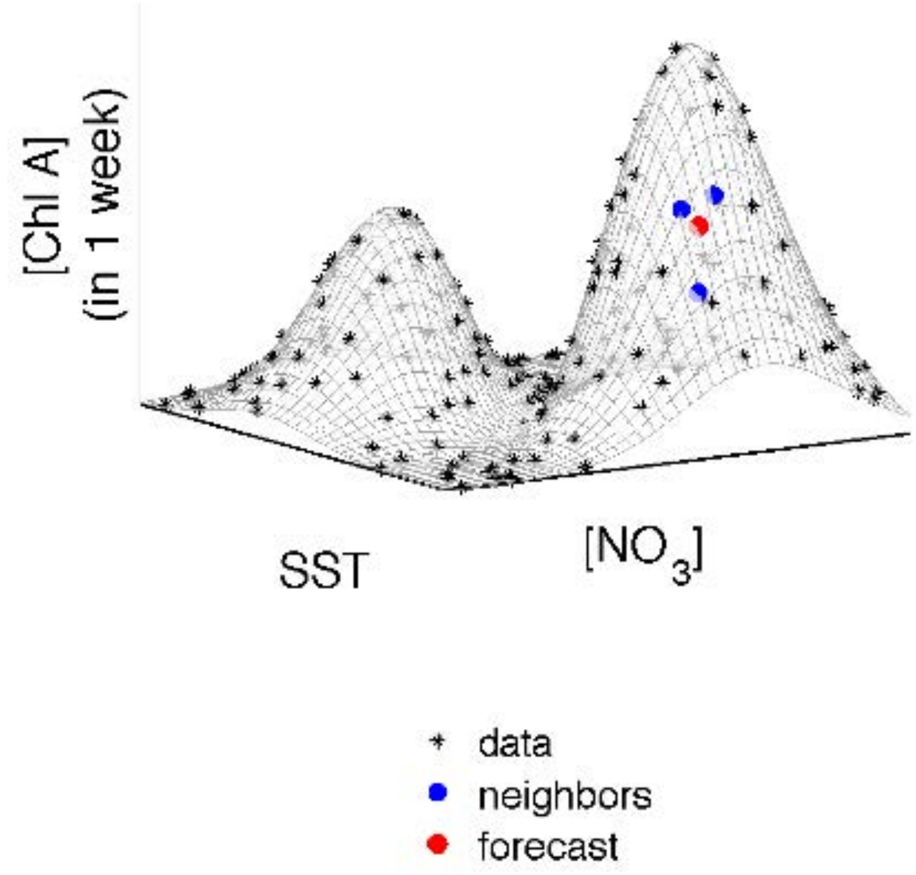
# Think of "supervised Machine learning for forecasting" as universal function approximation.

$$Y_{t+1} = f(Y_t, X_t | \bar{\theta} + \alpha) + \varepsilon_t$$

- Nearest neighbor forecasting.
- Generalizations of regression.
  - Local linear regression/kernel regression.
  - Dynamic linear models.
- Artificial neural networks.
- Gaussian processes.
- Random forest.

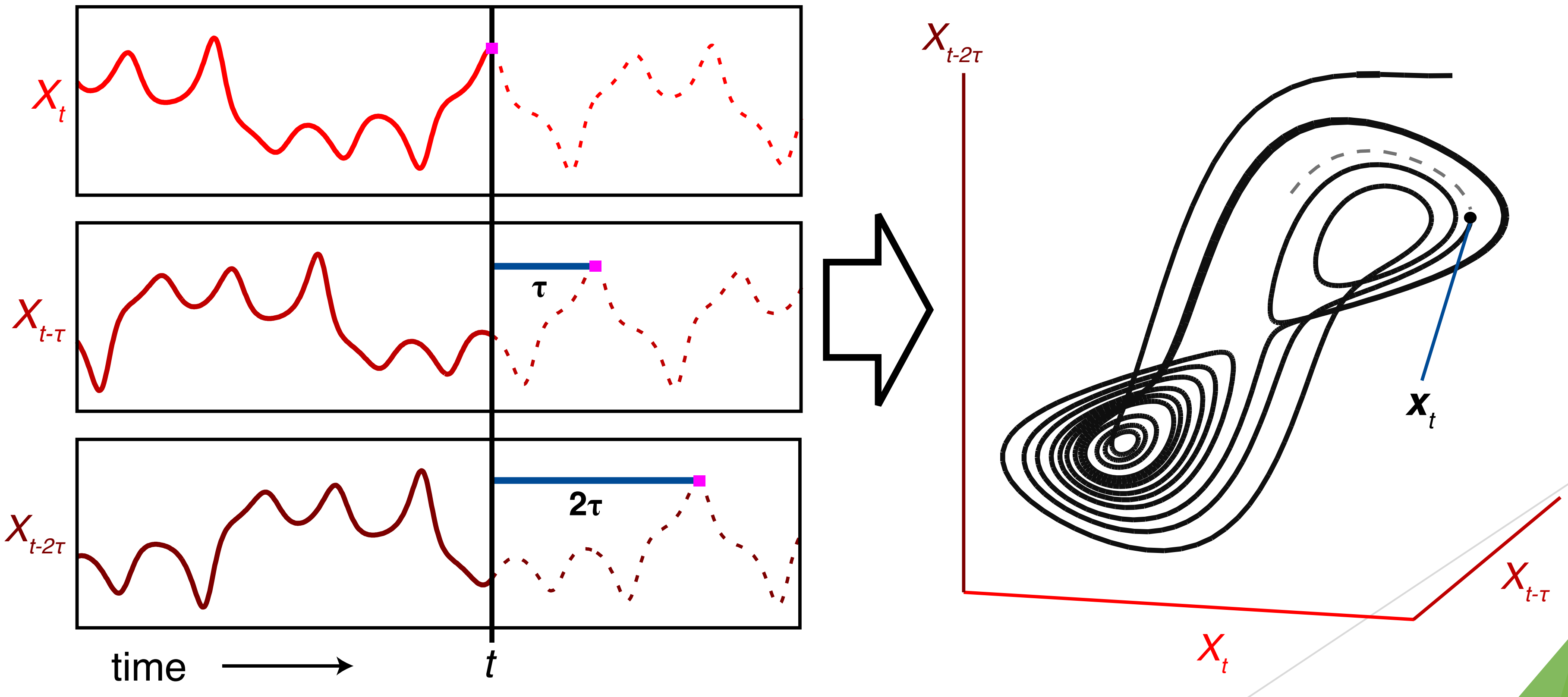
- All of these can incorporate endogenous and exogenous variables.
- Some treat observation error implicitly (averaging), some can treat it explicitly.
- Process uncertainty looks quite different when there are no parameters.
- "Tunability"/ "interpretability" can vary widely.

# Tools look a lot like non-parametric surface fits but often we want *dynamics* not a *response surface*.



# Tools look a lot like non-parametric surface fits but often we want *dynamics* not a *response surface*.

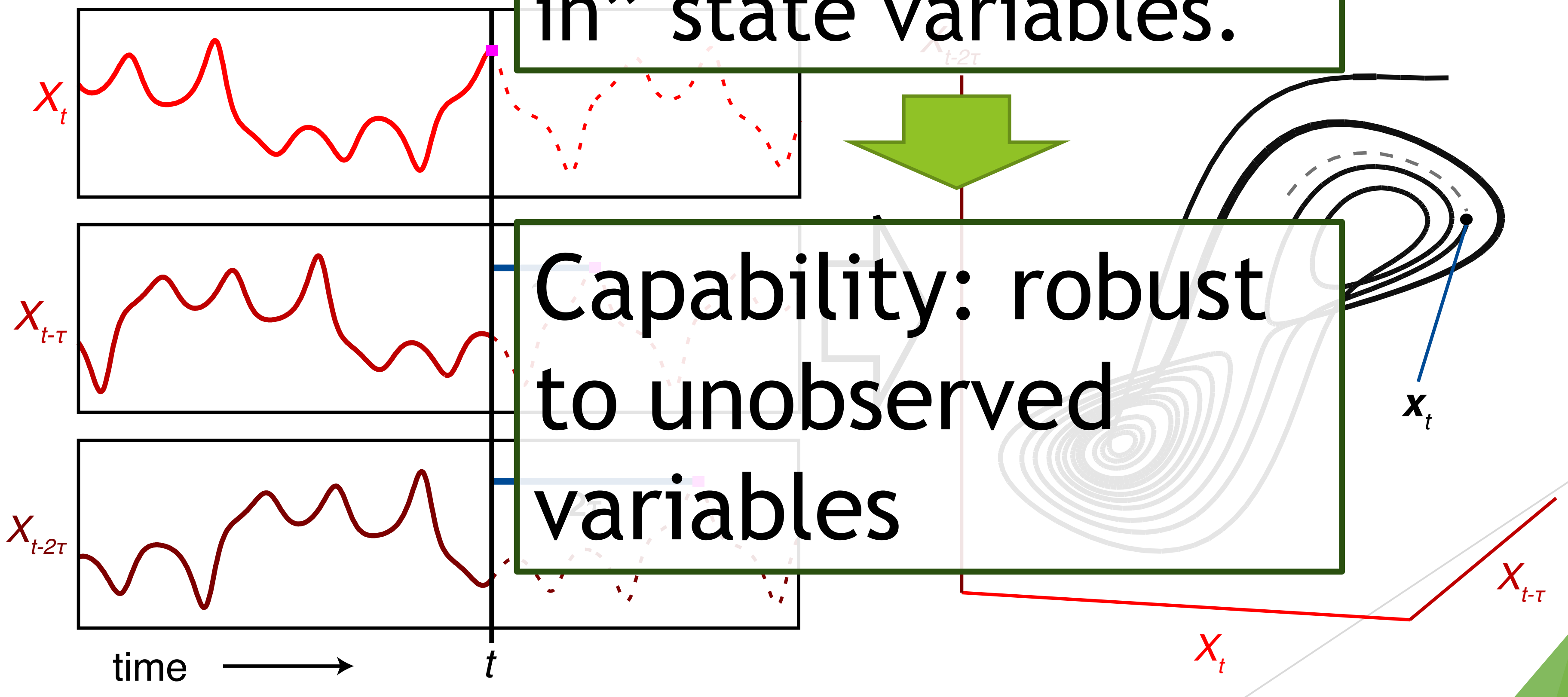
"Takens Theorem" — observing change-over-time gives a window into the coupled dynamics of the system even when there are unobserved state variables.



# Tools look a lot like non-parametric surface fits but often w... *response surfa...*

"Takens Theorem" — observing change over time gives a window into the coupled dynamics of the system even when there are unobserved state variables.

**Time-lag values of measured variables can be used as "fill-in" state variables.**




**Capability: robust to unobserved variables**

# ML lets us sidestep model misspecification **aka** the “other” error **aka** structural uncertainty.

ICES Journal of Marine Science

ICES CIEM

ICES Journal of Marine Science (2018), 75(3), 903–911. doi:10.1093/icesjms/fsx202

**Review Article**

**Assessing causal links in fish stock–recruitment relationships**

Maud Pierre[1]*, Tristan Rouyer[2], Sylvain Bonhommeau[3], and J. M. Fromentin[2]

Received: 21 November 2017 | Revised: 9 May 2018 | Accepted: 23 May 2018

DOI: 10.1111/faf.12304

ORIGINAL ARTICLE

WILEY FISH and FISHERIES

**Nonlinear dynamics and noise in fisheries recruitment: A global meta-analysis**

Stephan B. Munch[1,2] | Alfredo Giron-Nava[3] | George Sugihara[3]

Article | Open Access | Published: 24 April 2020

**Circularity in fisheries data weakens real world prediction**

Alfredo Giron-Nava, Stephan B. Munch, Andrew F. Johnson, Ethan Deyle, Chase C. James, Erik Saberski, Gerald M. Pao, Octavio Aburto-Oropeza & George Sugihara

*Scientific Reports* **10**, Article number: 6977 (2020) | Cite this article



All use variations of empirical dynamic modeling (ML) to investigate recruitment prediction and S-R relationship.

Across these cases, evidence that blaming poor fits on observational and processed uncertainty can mask a deeper problem.

“What we know for sure that just ain’t so”.

# ML lets us sidestep model misspecification **aka** the "other" error **aka** structural uncertainty.

ICES Journal of Marine Science

ICES Journal of Marine Science (2018), 75(3), 903–911. doi:10.1093/icesjms/fsx202

Review Article

Assessing causal links in fish stock–re

Maud Pierre[1]*, Tristan Rouyer[2], Sylvain Bonhommeau[3],

Received: 21 November 2017 | Revised: 9 May 2018 | Accepted: 23 May 2018

DOI: 10.1111/faf.12304

ORIGINAL ARTICLE

WILEY FISH and FISHERIES

e in fisheries recruitment: A global

Nava[3] | George Sugihara[3]

Article | Open Access | Published: 24 April 2020

Circularity in fisheries dat

world prediction

Alfredo Giron-Nava, Stephan B. Munch, Andrew F. Johnson, Ethan Deyle, Chase C. James, Erik Saberski, Gerald M. Pao, Octavio Aburto-Oropeza & George Sugihara

*Scientific Reports* **10**, Article number: 6977 (2020) | Cite this article



Capability: sidestep or capture structural uncertainty.

empirical dynamic

stigate recruitment

elationship.

Across these cases, evidence that blaming poor fits on observational and processed uncertainty can mask a deeper problem.

"What we know for sure that just ain't so".

# Application to “red tide” forecasting.

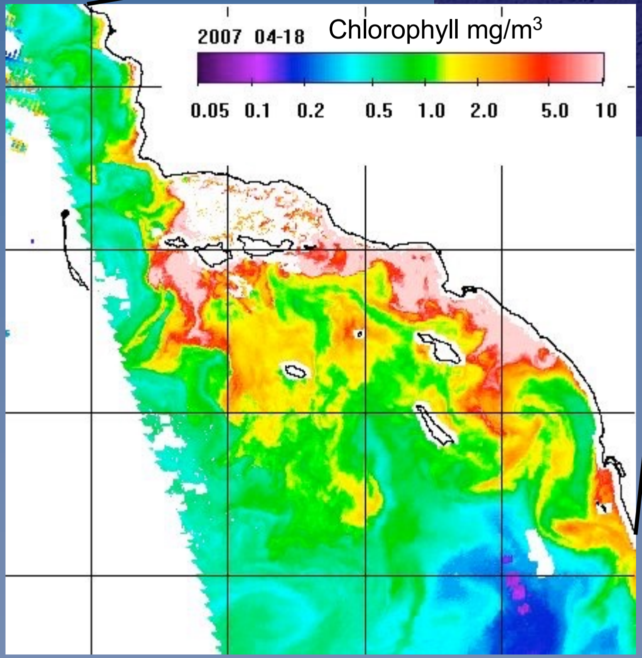
Talking about time-scales; touching on limitations

John A. McGowan, Hao Ye, Melissa L. Carter, Charles T. Perretti, Kerri D. Seger, Alain de Verneil, George Sugihara

Drew Lucas, Art Miller, Steve Munch, Enrique Curchitser

# Coastal Algal Blooms in Southern California

Southern California Bight

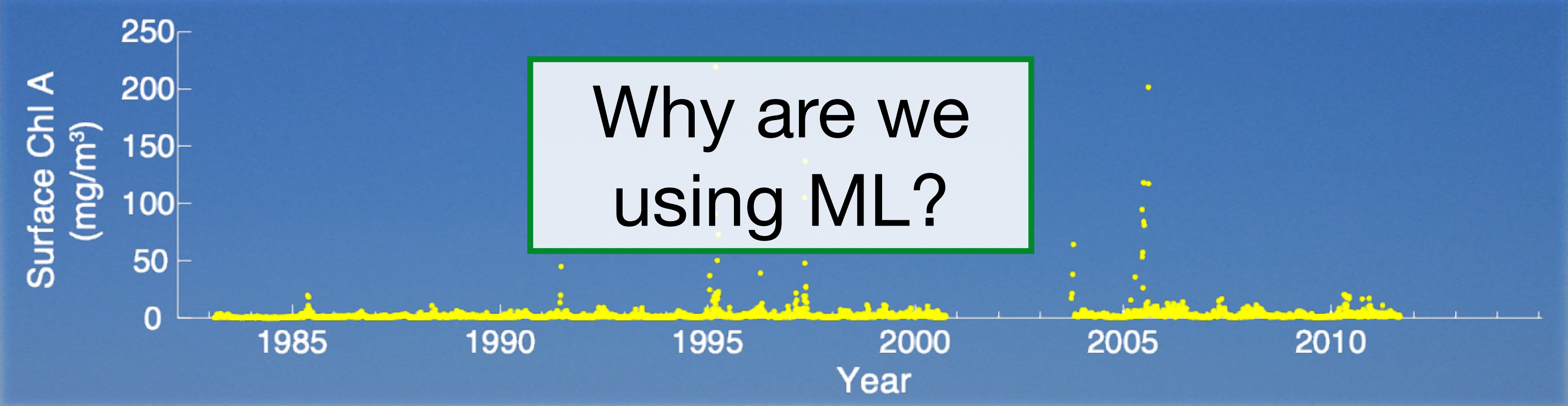



Scripps Pier, La Jolla

(1084 ft.)

McGowan et al. 2017 *Ecology* [*10.1002/ecy.1804*]

# Red Tides/Coastal Algal Blooms in Southern California




- Red tide observations in La Jolla date back to Allen (1917-1945).
- Systematic data collection beginning in 1983 (n=2595). [Chlorophyll blooms].

# Red Tides/Coastal Algal Blooms in Southern California


Why are we using ML?

Uncertain drivers and mechanisms.

Hypothesis poorly supported in traditional linear statistical analysis.

- Red tide observations in La Jolla date back to Allen (1917-1945).
- Systematic data collection beginning in 1983 (n=2595). [Chlorophyll blooms].

# Red Tides/Coastal Algal Blooms in Southern California


Why are we using ML?

Uncertain drivers and mechanisms.

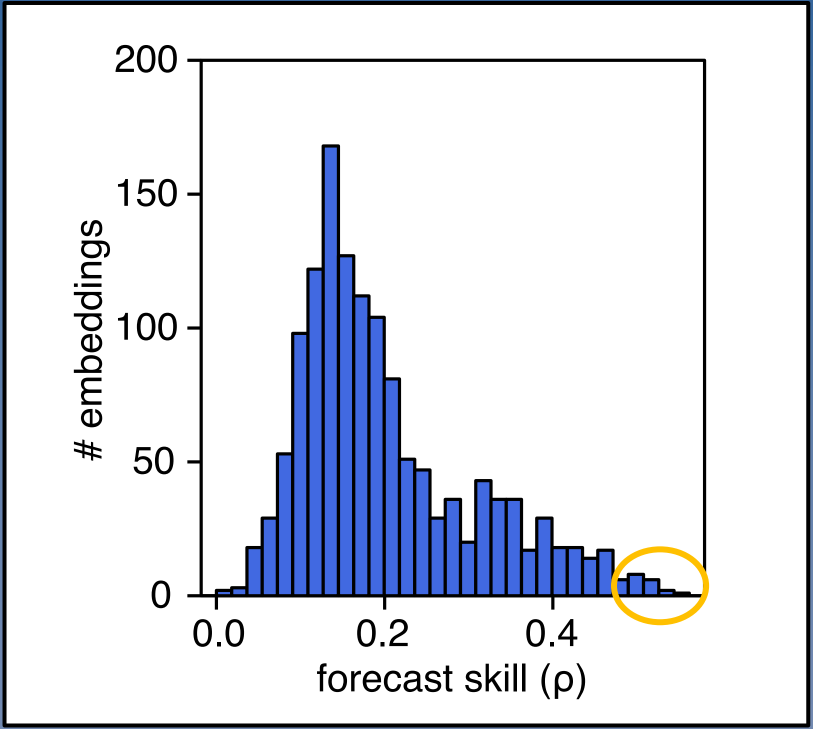
Hypothesis poorly supported in traditional linear statistical analysis.

- Red tide observations in La Jolla date back to Allen (1917-1945).
- Systematic data collection beginning in 1983 (n=2595). [Chlorophyll blooms].

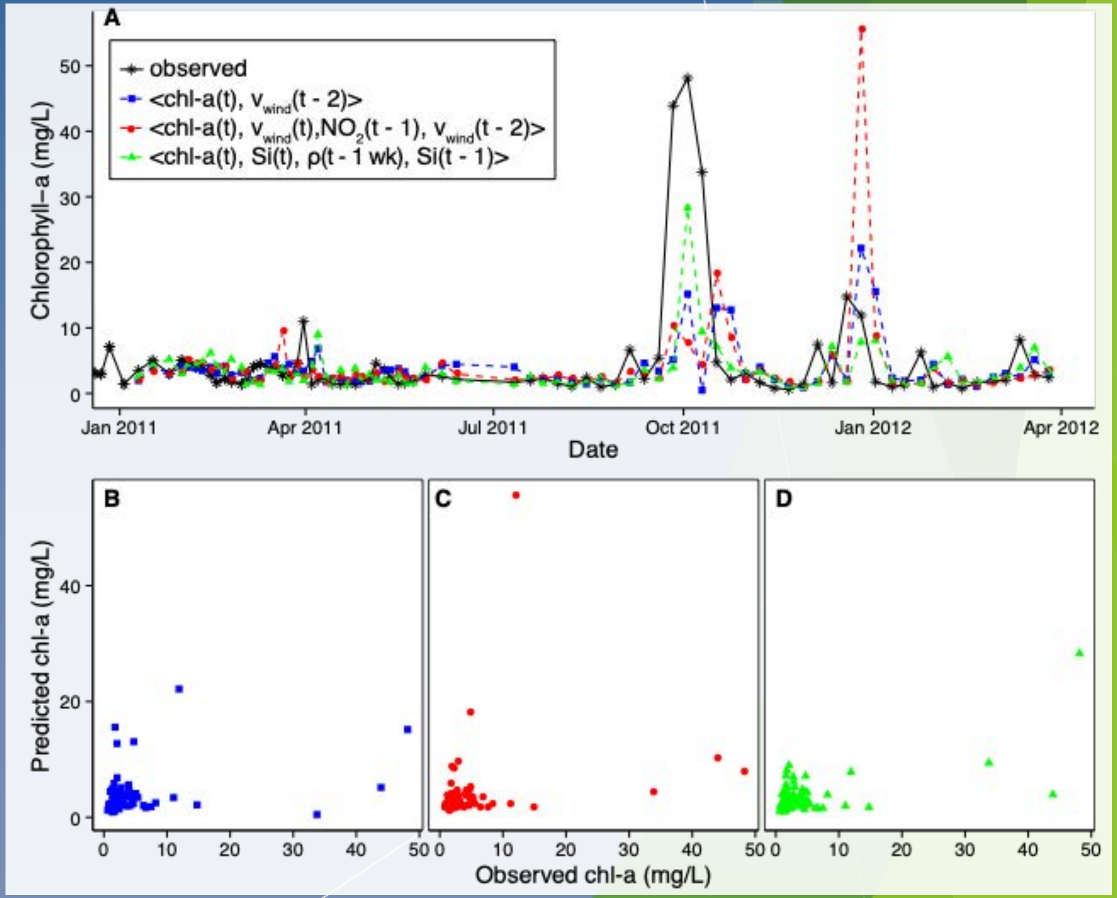
Are they predictable at all?

# McGowan et al. 2017: Short-term Forecasting

In-sample
variable selection




Out-of-sample
ultimate test



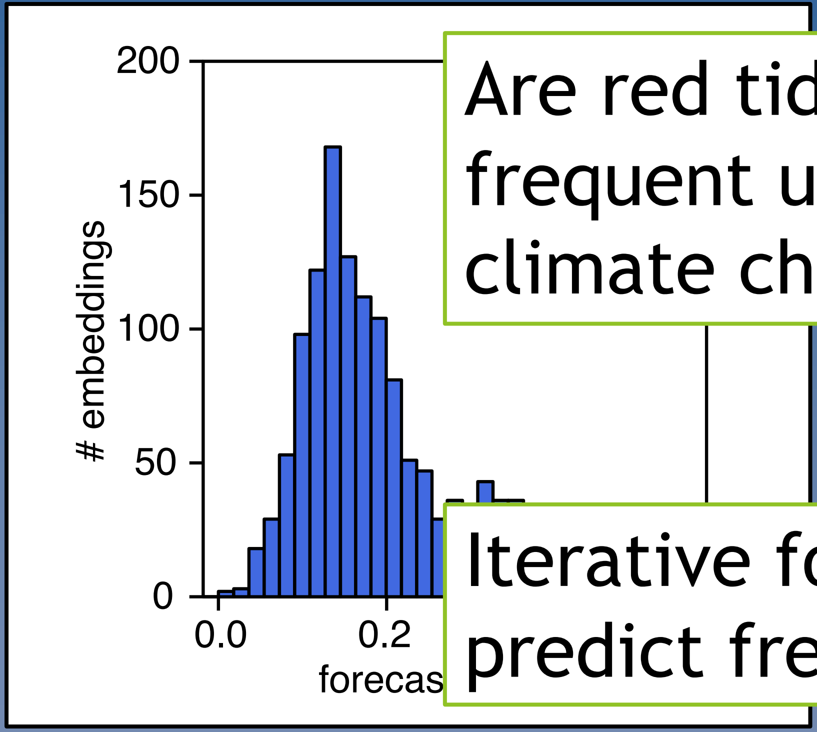

Blooms are stochastic chaos: deterministic dynamics forced by high-dimensional physics.

Multi-model suitability: Some proximal drivers not measured, need model averaging.

# McGowan et al. 2017: Short-term Forecasting

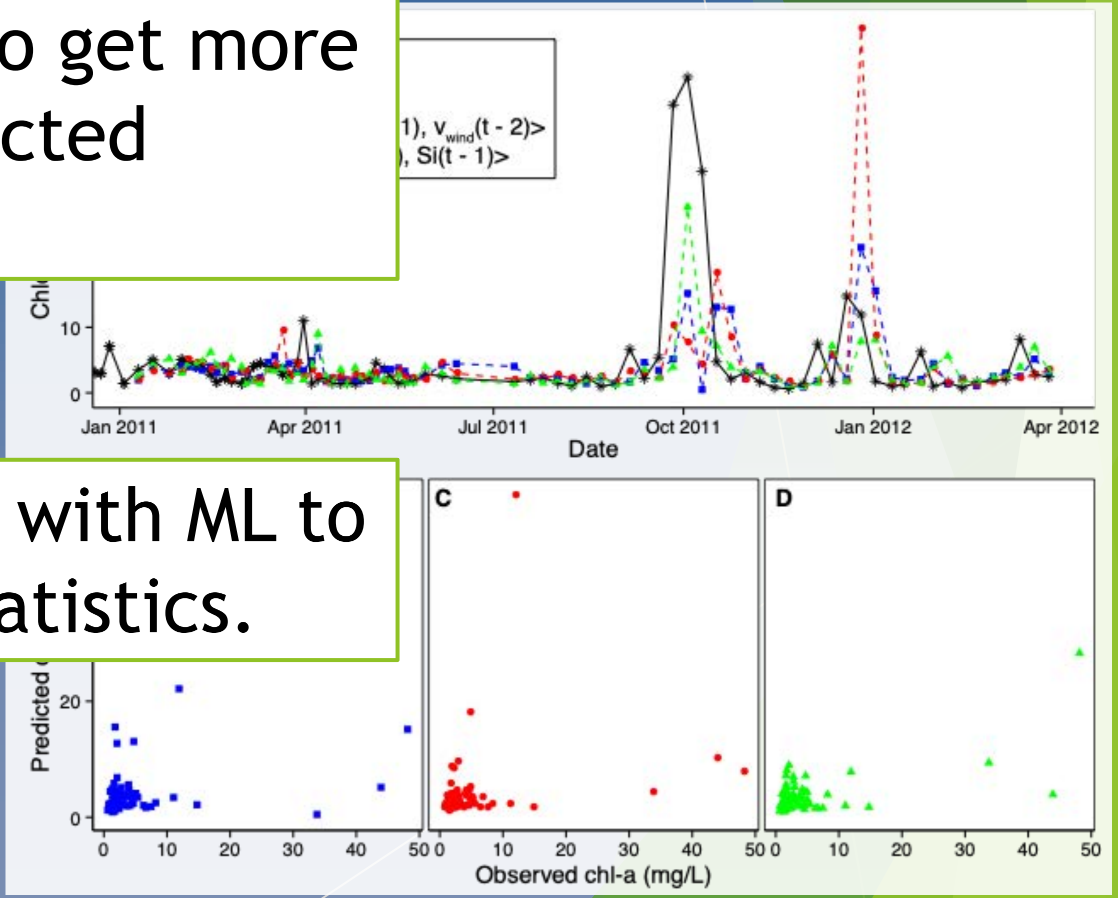
In-sample
variable selection

Out-of-sample
ultimate test




Are red tides going to get more frequent under expected climate change?

Iterative forecasting with ML to predict frequency statistics.

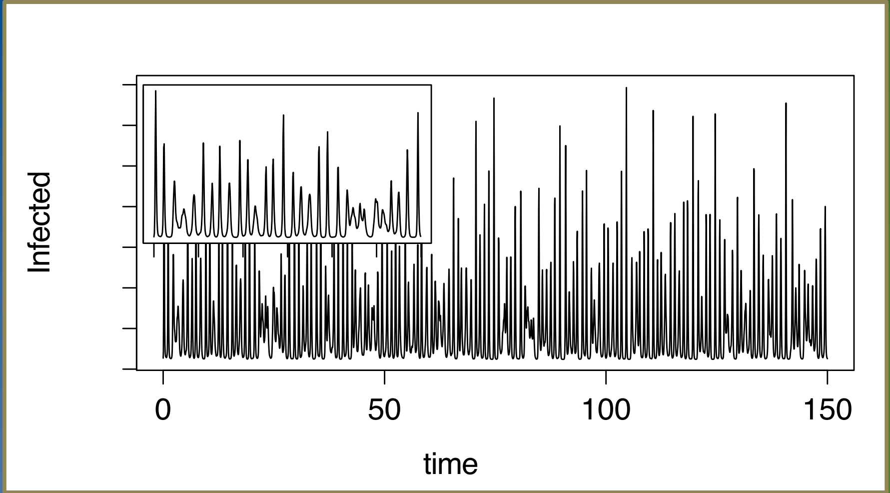


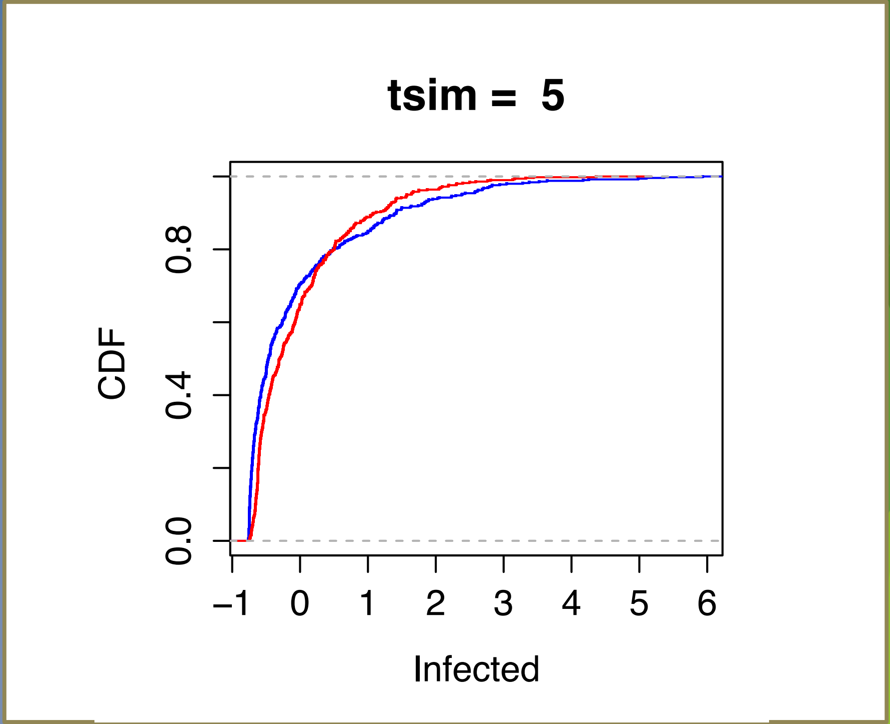

Blooms are **stochastic chaos**: deterministic dynamics forced by high-dimensional physics.

Multi-model suitability: Some proximal drivers not measured, need model averaging.

# Iterative forecasting with short-term predictor exposes challenge:

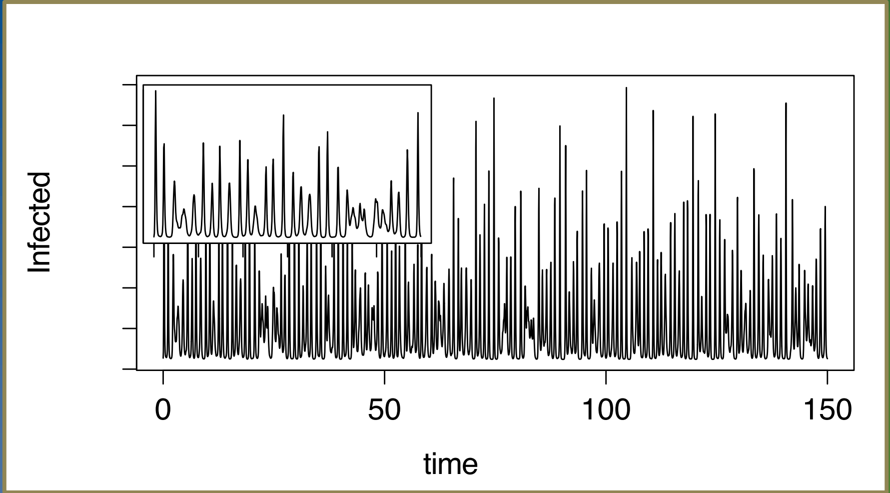
- Considering systems with sharp Lyapunov horizons, but underlying behavior driven by climate.
- Cumulative distributions: toy model realization (blue) versus the iterated ML forecast (red).
- Under short time horizons, the distribution is quite accurately recovered, but…

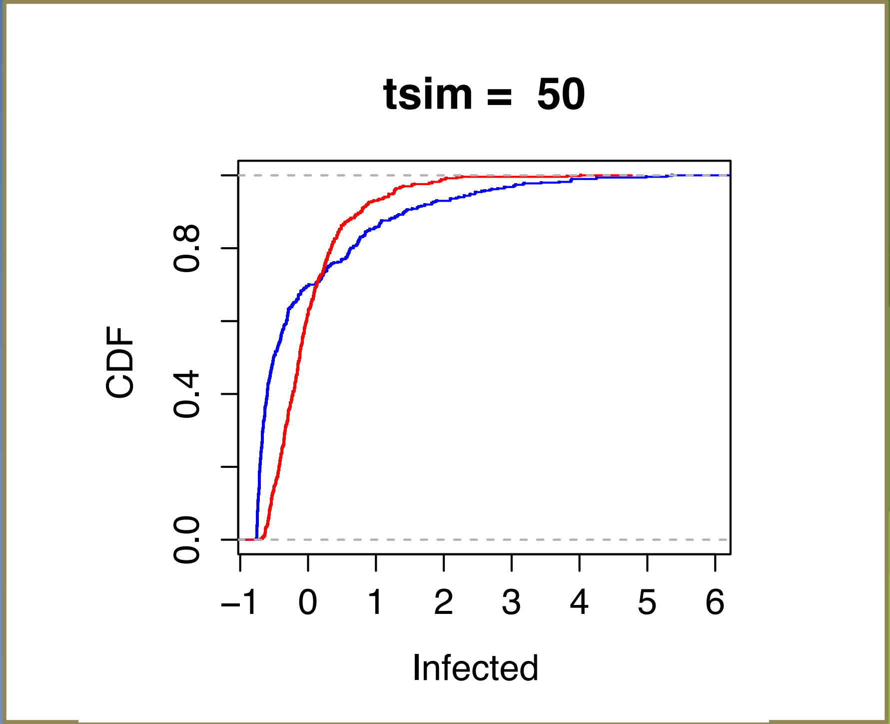


# Iterative forecasting with short-term predictor exposes challenge:

- Under longer time horizons, bias towards the median becomes more evident, and the frequency of large events is under-estimated.
- At 50 time-steps, ML fails to simulate any outbreaks over 2 (normalized units) despite these occurring roughly 10% of the time in the true system.
- Add stochasticity based on uncertainty. [Turned out to be easiest just to subsample].

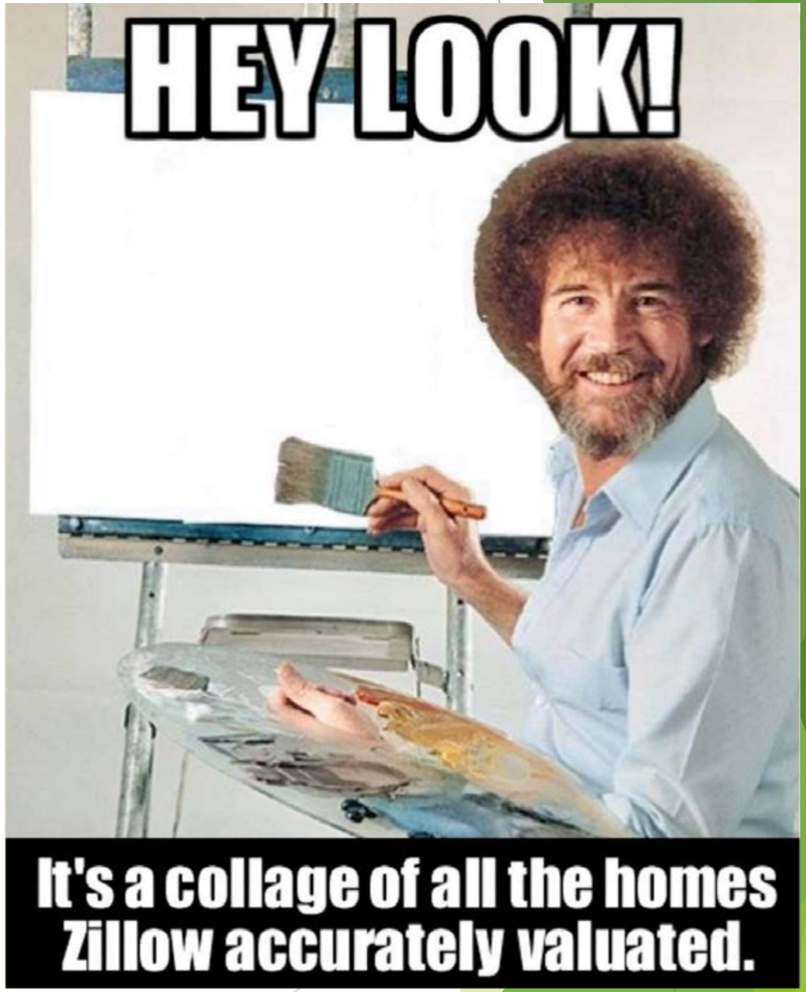


# Challenge: Bias and Treating ML Methods as a Black-Box.

- "PROPHET" and the collapse of Zillow.
  - Like parametric models, machine learning approaches have tunable knobs.
  - Can be very opaque what their effect is.
  - Forecast bias.

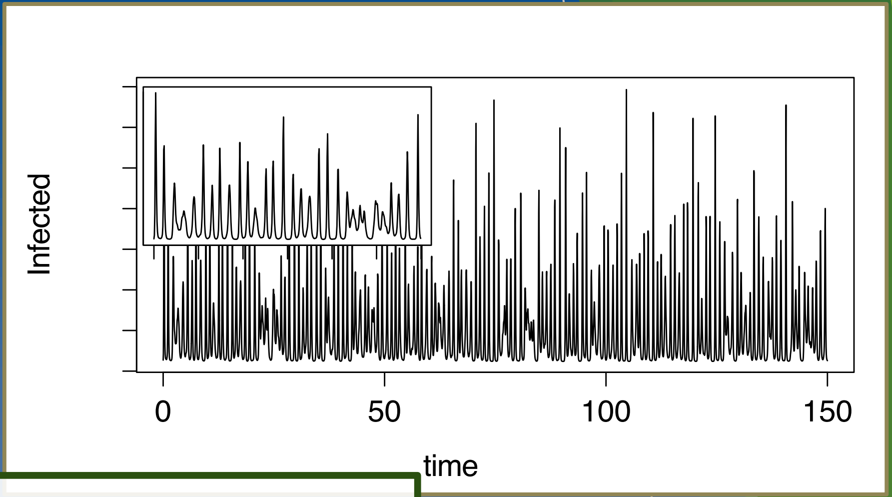
https://towardsdatascience.com/in-defense-of-zillows-besieged-data-scientists-e4c4f1cece3c



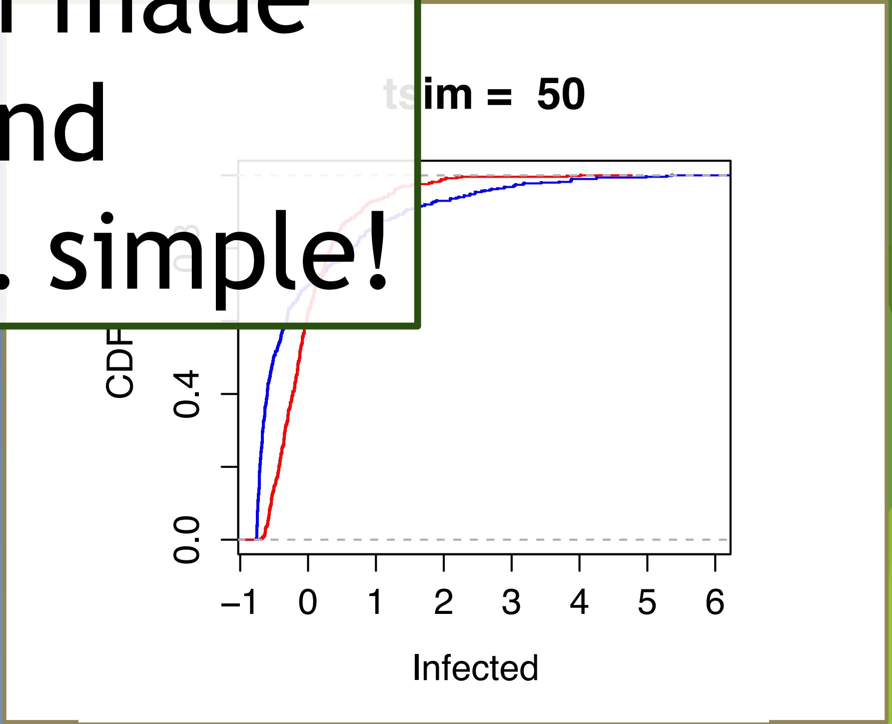

https://lightersideofrealestate.com/

# Iterative forecasting with short-term predictor exposes challenge:

- Under longer time horizons, bias towards the median becomes more evident, and the frequency of large events is under-estimated.
- At 50 time-steps, ML fails to simulate any outbreaks over 2 (normalized units) despite these occurring roughly 10% of the time in the true system.
- Add stochasticity based on uncertainty. [Turned out to be easiest just to subsample].




**Simple approach made understanding and correcting bias… simple!**



# Extrapolate EDM short-term forecasting to Non-Analogue Futures

## (1) Direct forcing of EDM model

ROMs predictions of coastal environment → Multivariate EDM → Red Tides under climate change

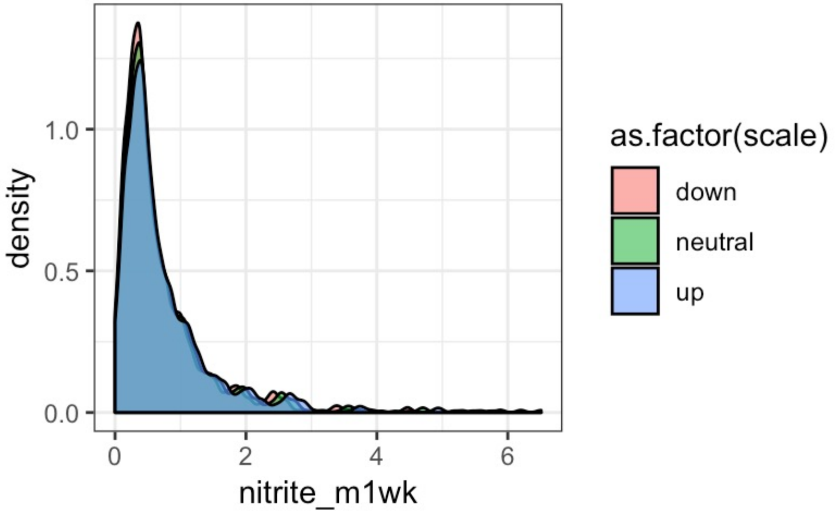
## (2) EDM Scenario Exploration (climate sensitivity analysis)

Qualitative predictions of coastal environment.

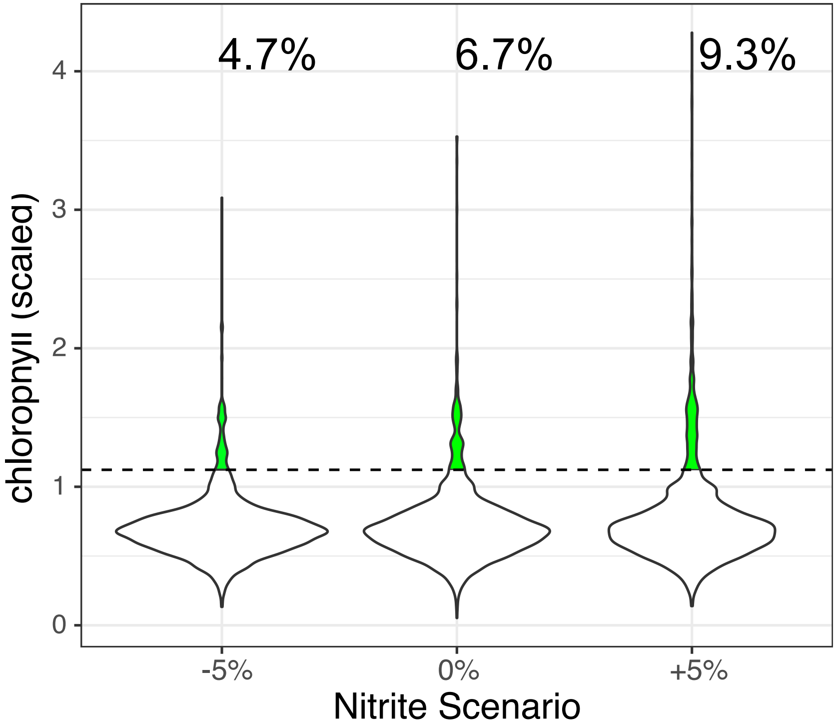
Simple changes in variable -5% +5% → Multivariate EDM → Red Tides under climate change

# Extrapolate EDM short-term forecasting to Non-Analogue Futures

Climate sensitivity, change in nitrite.

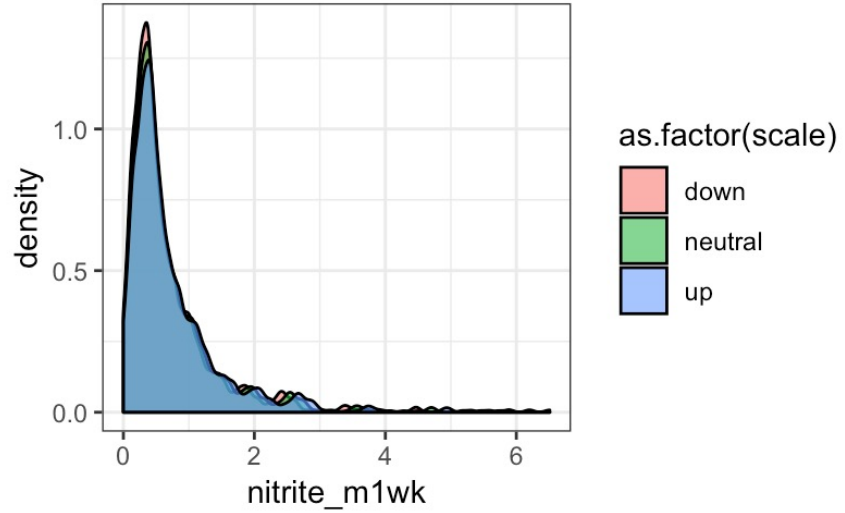



EDM predicts bloom frequency will increase/decrease by 50% with a 5% increase/decrease in nitrite from current levels.

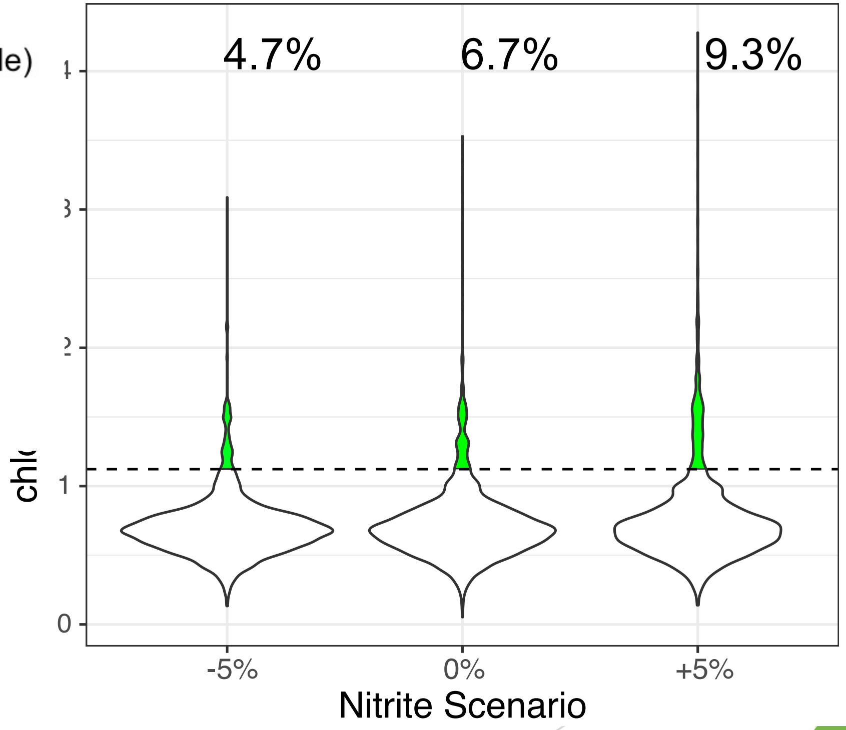


# Extrapolate EDM short-term forecasting to Non-Analogue Futures

Climate sensitivity, change in nitrite.

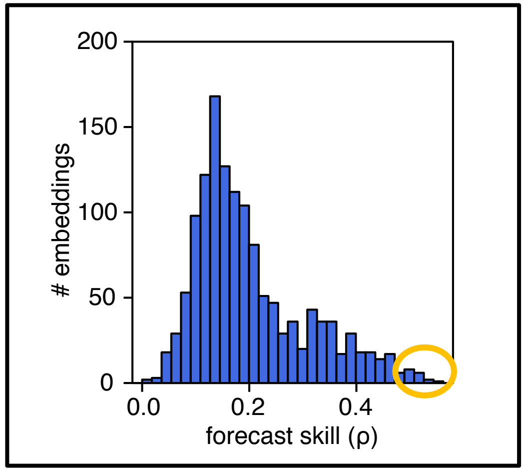


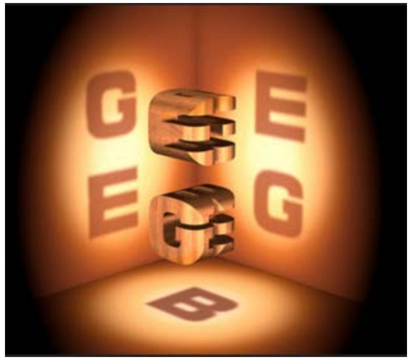

EDM predicts bloom frequency will increase/decrease by 50% with a 5% increase/decrease in nitrite from current levels.

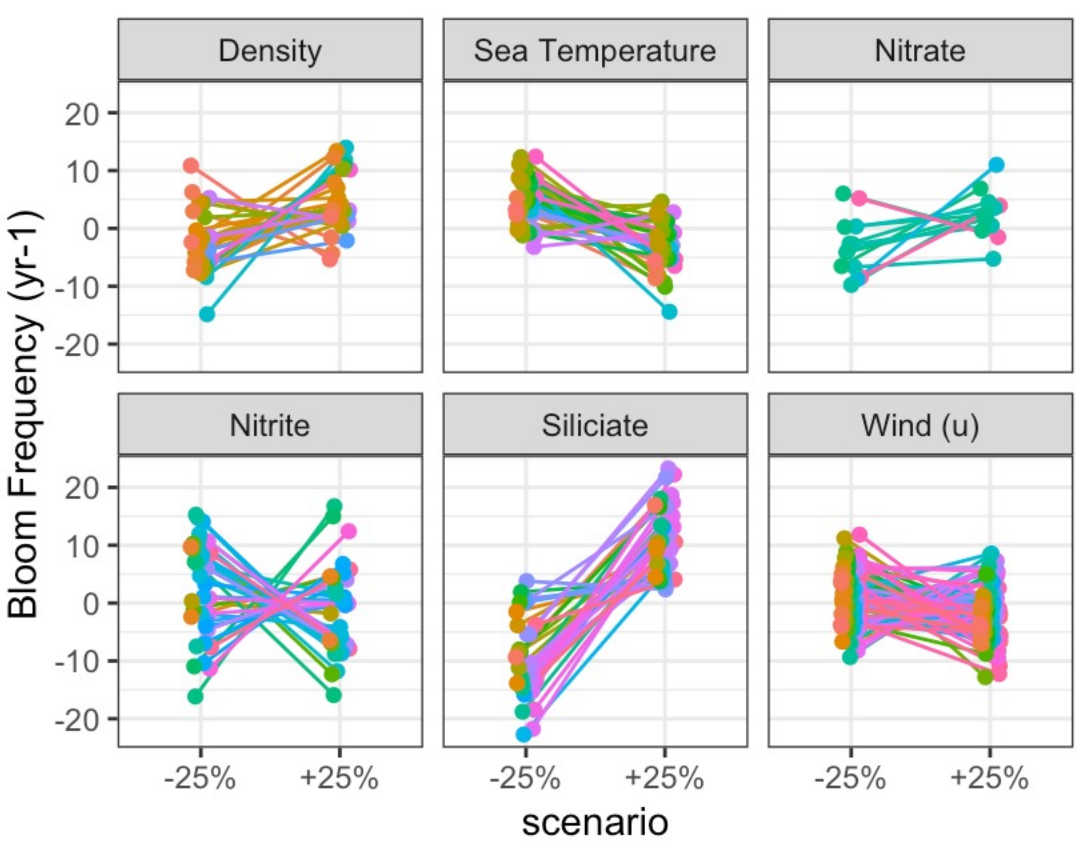


# Extrapolate EDM short-term forecasting to Non-Analogue Futures

Multi-model predictions of climate sensitivity





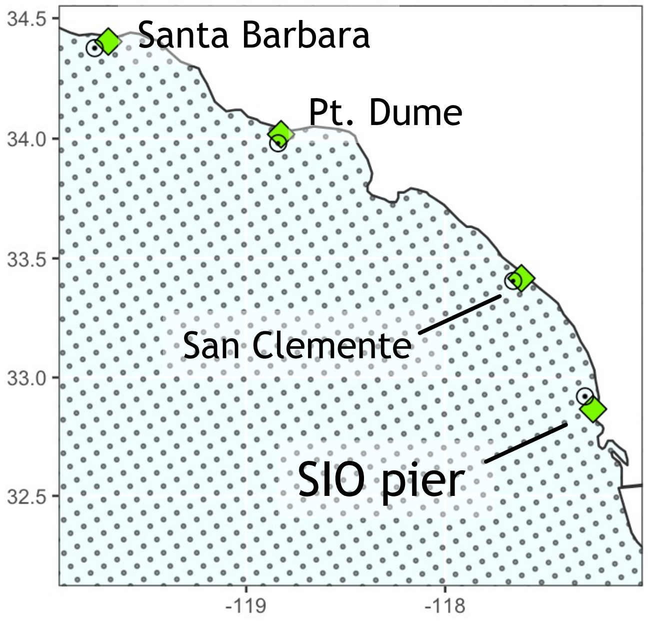
# Extrapolate EDM short-term forecasting to Non-Analogue Futures

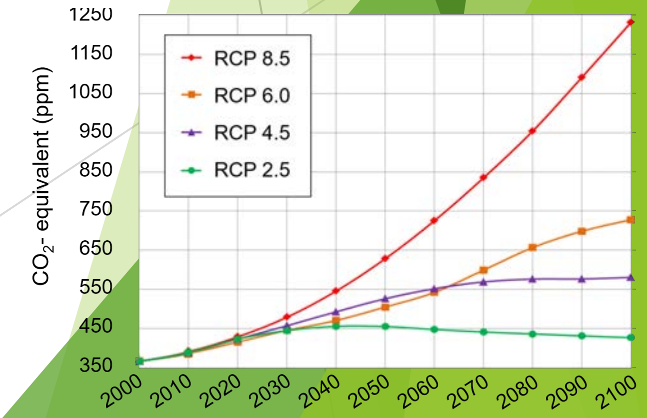
## ROMs Forcing

Curchitser, E. N.; Dussin, R.; Stock, C. A.

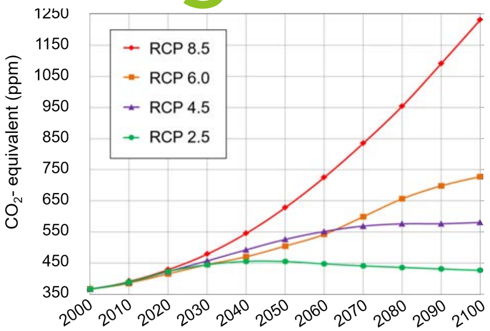
Regional Ocean Modeling System (ROMS) + NOAA/GFDL's Carbon, Ocean Biogeochemistry and Lower Trophics (COBALT) biogeochemical model

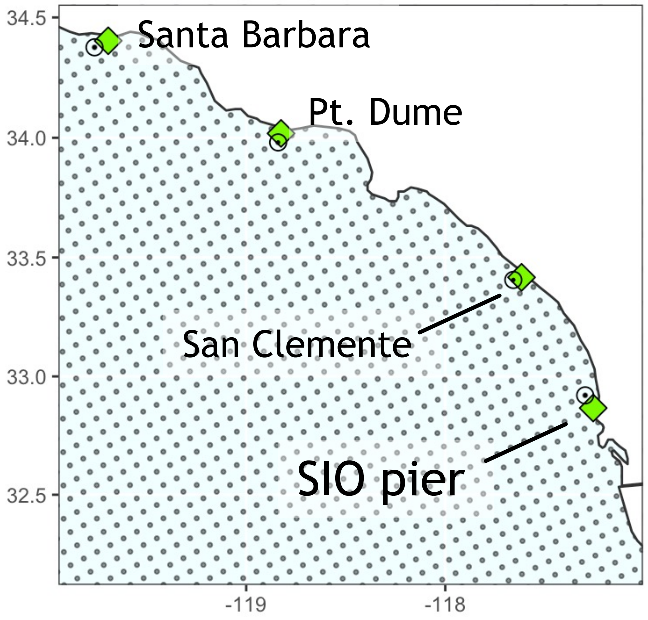
Forced by GFDL ESM2M RCP8.5 future projection.

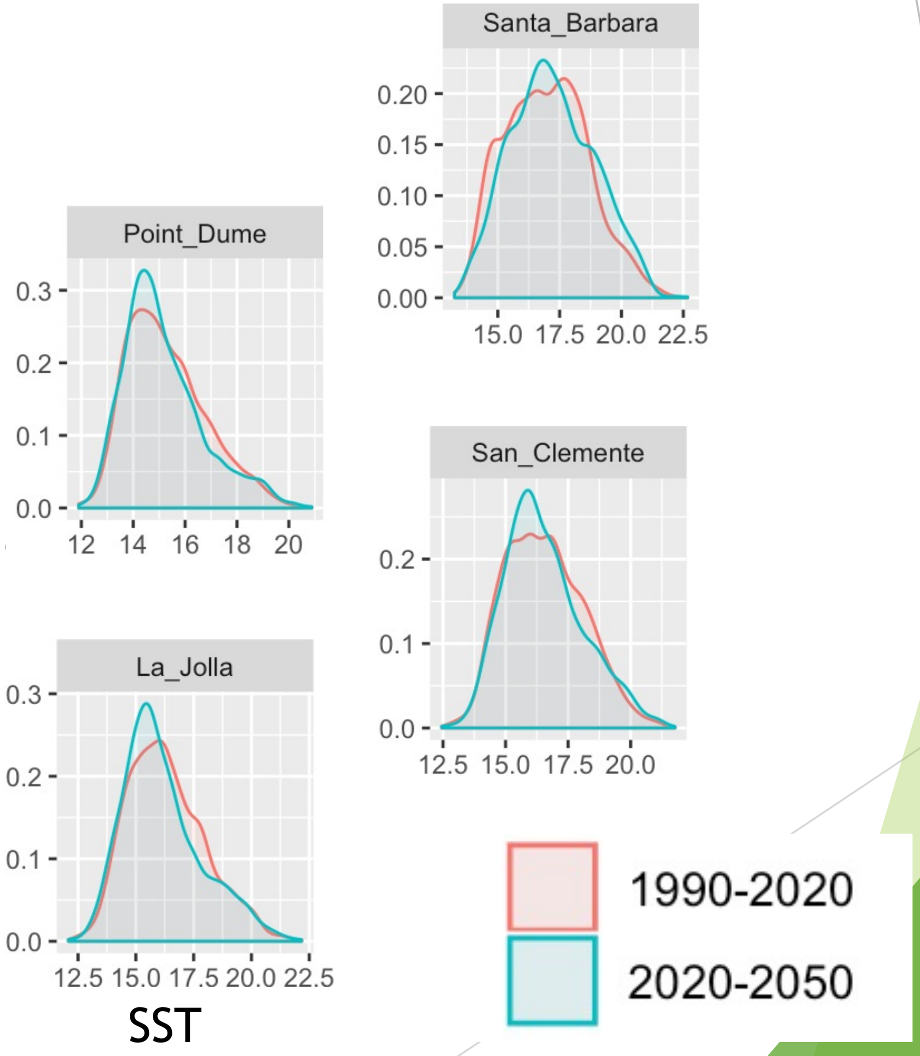


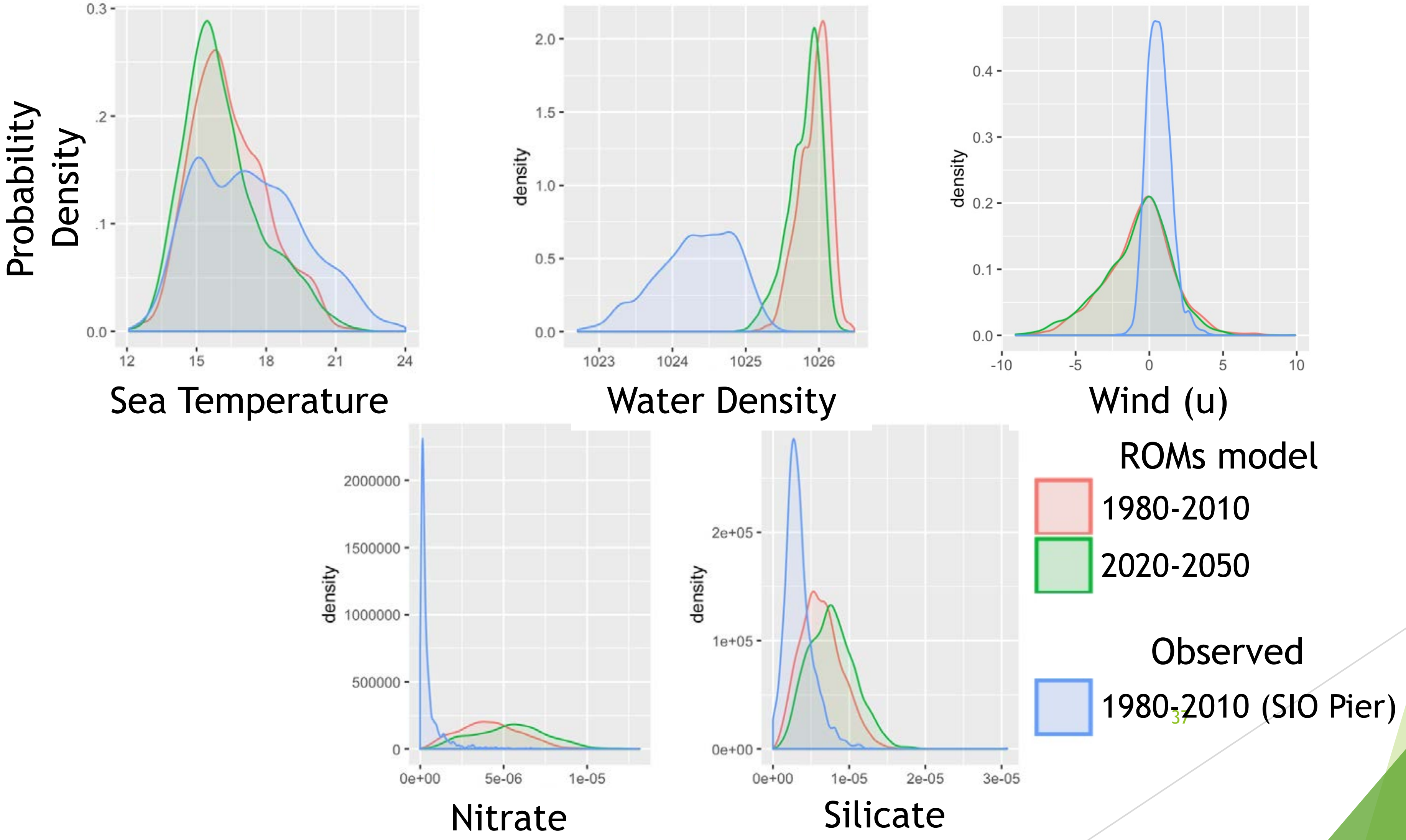
# Extrapolate EDM short-term forecasting to Non-Analogue Futures



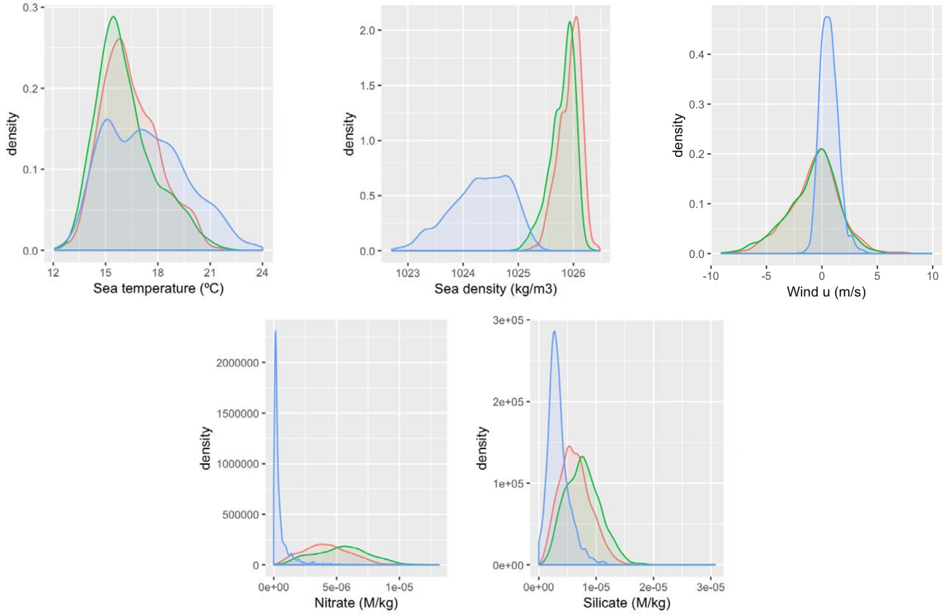



# However, the EDM models were built on near-shore observations and the ROMs model is not.



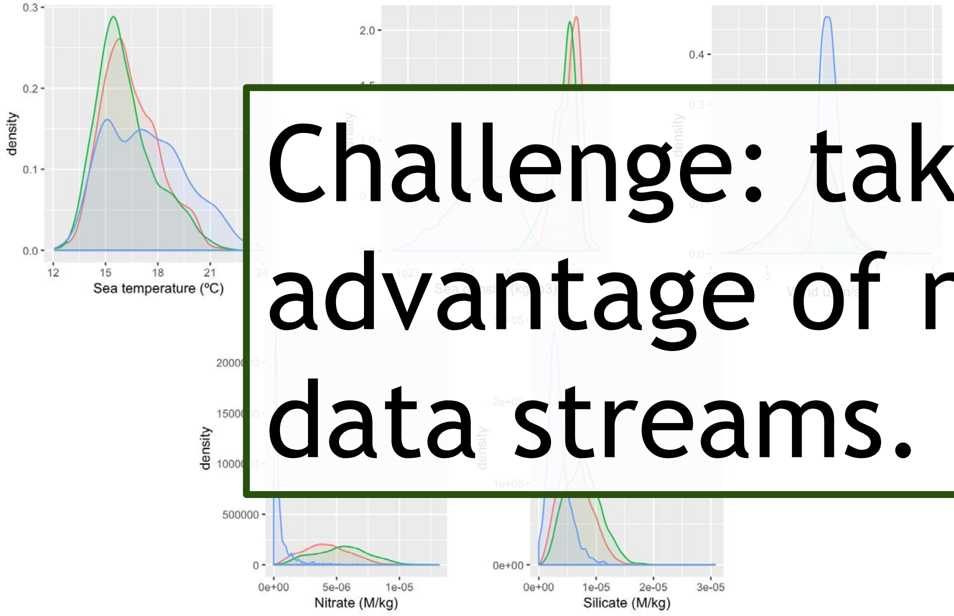
# If you are willing to assume that the direction of change in the ROM is the same as near-shore environment…




| | |
|---|---|
| SST | ↓ |
| Density | ↓ |
| u-Wind | ? |
| Silicate | ↑ |
| Nitrate | ↑ |

# …our best prediction is that there will be an increase in red tide frequency along the San Diego coast moving towards 2050.

If you are willing to assume that the direction of change in the ROM is the same as near-shore environment…




Challenge: taking advantage of new data streams.

| | |
|---|---|
| SST | ↓ |
| Density | ↓ |
| u-Wind | ? |
| Silicate | ↑ |
| Nitrate | ↑ |

…our best prediction is that there will be an increase in red tide frequency along the San Diego coast moving towards 2050.

# Application to hypoxia in Lake Geneva.

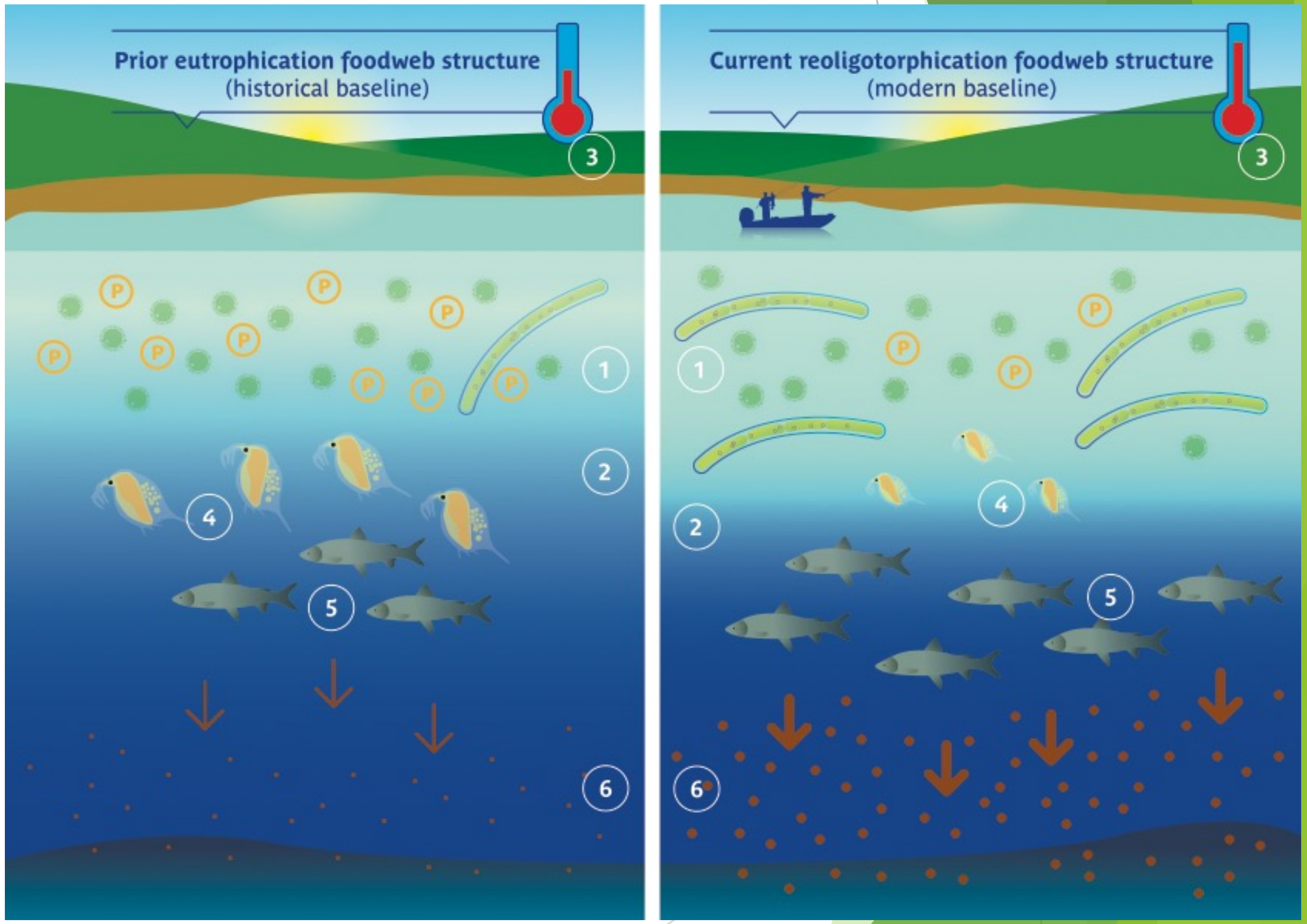
Overcoming challenge of interpretability; hybrid approaches to leverage first-principles understanding.

Damien Bouffard, Victor Frossard, Robert Schwefel,
Johnb Mellack, George Sugihara.

Tom Lorimer.

# Hybrid Approaches to Hypoxia: Lake Geneva Study

- Fixed, parameterized rates really most immediately appropriate to capture a fixed community / food-web.
- These are changing! *Non-stationary world.*
- Low-dimensional nonlinear regression (aka simple supervised machine learning) can represent changing interactions between variables (rates).

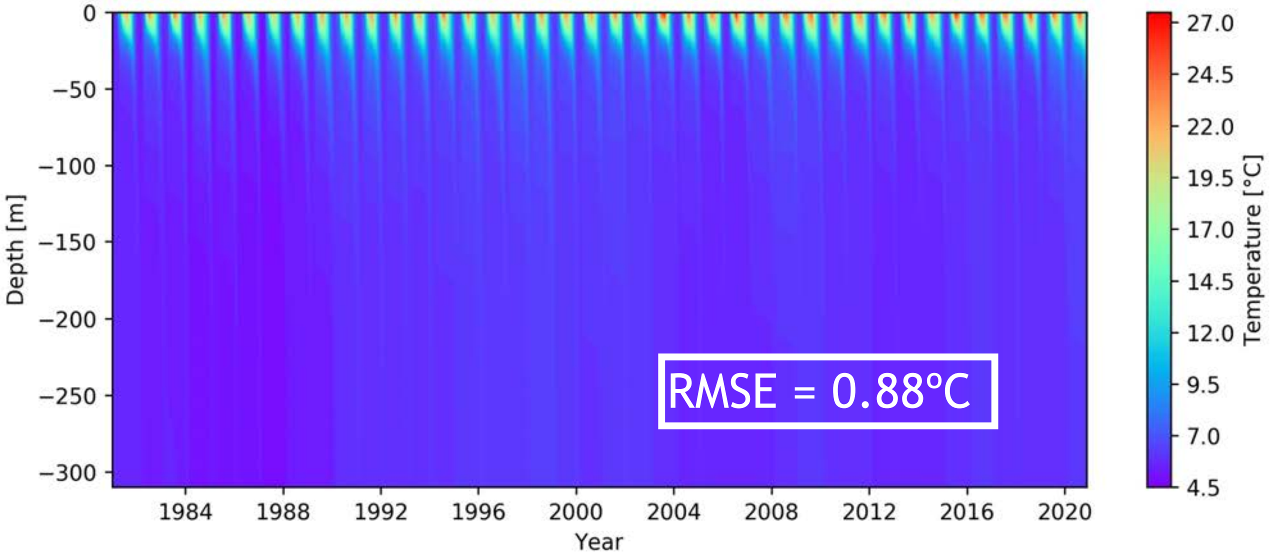



- *Deyle et al. PNAS in press*

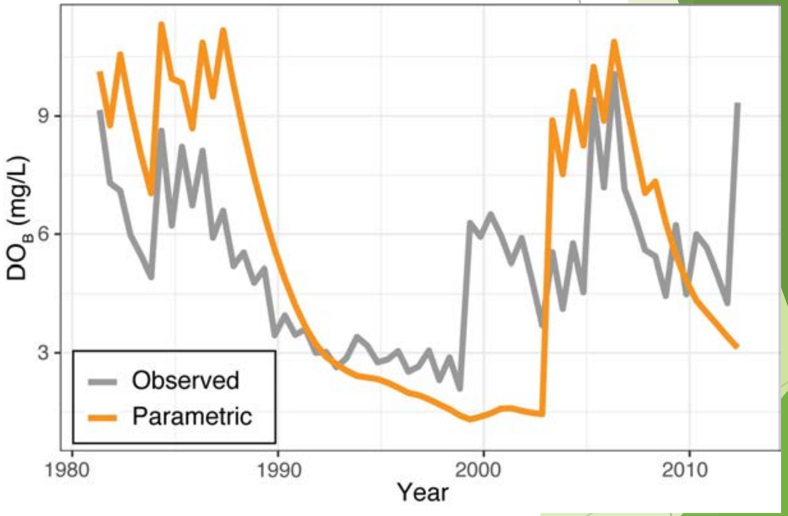
# Parametric modeling of stratification has shown predictive success, but extending the framework to water-quality has been harder.

Weather
Inflow
Bathymetry → 1D Physical State

1D Physical State
Chlorophyll → DO




Simstrat (v2.0) predictions of thermal structure
[simstrat.eawag.ch/LakeGeneva](simstrat.eawag.ch/LakeGeneva)

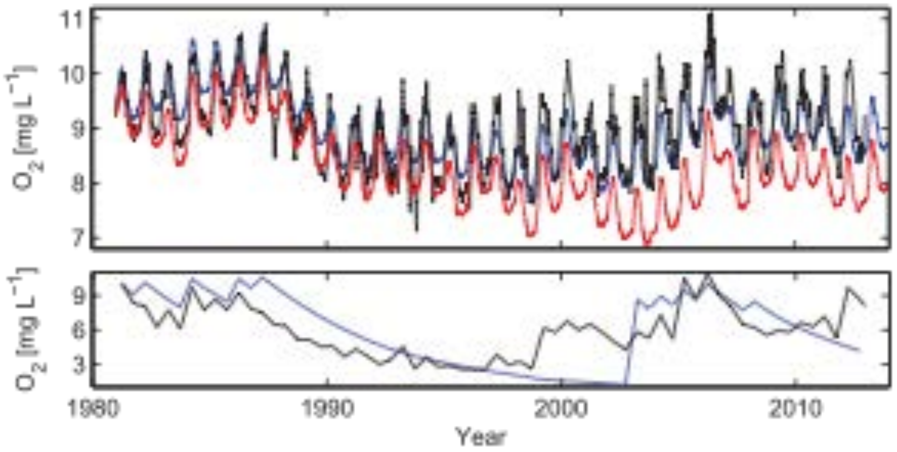



Annual predictions of $DO_B$ from coupled parametric model (Schwefel et al. 2016)
[doi.org/10.1002/2016WR019194](doi.org/10.1002/2016WR019194)

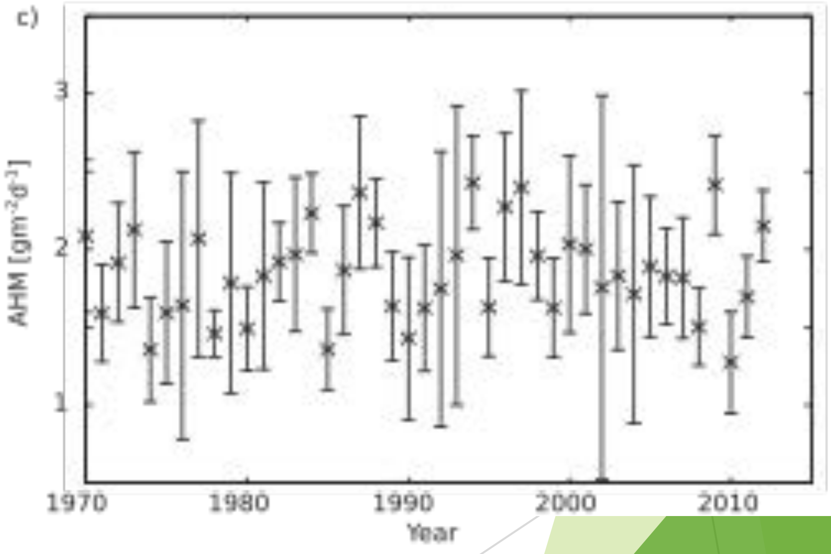
# Parametric description faces a trade-off between over-simplification and over-fitting that presents a major obstacle for management.

Only included effect of phosphorous indirectly through observed chlorophyll

$$\frac{dC_{ul}}{dt} = c_1\ k[C_{sat}(T,t) - C_{ul}] + c_2 Min(t)\ 1.036^{(T-20^\circ C)}\ P_{max}\ Chl(t) - c_3\ VOD(z,t)$$

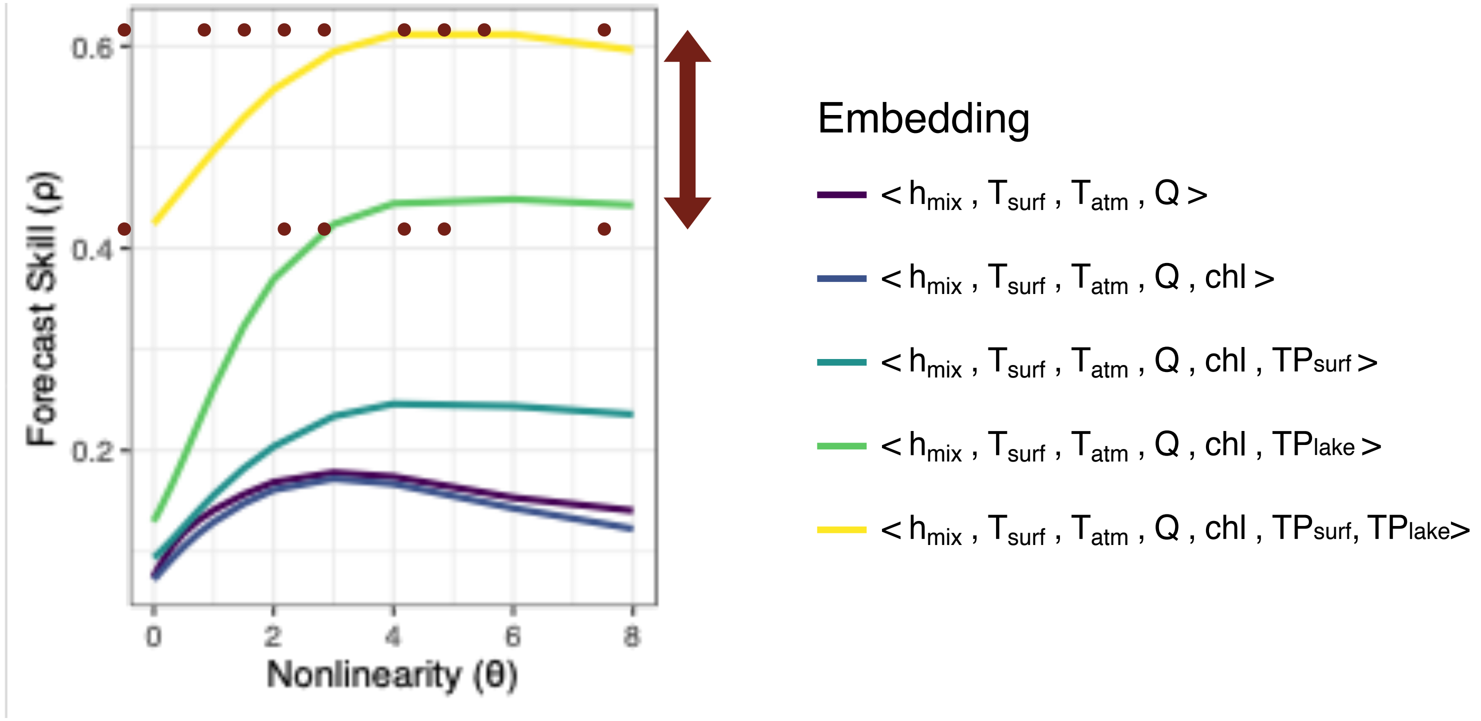



Observed rates of oxygen depletion vary substantially in and between years




(Schwefel et al. 2016)
doi.org/10.1002/2016WR019194

# One-step forecasts of $DO_B$ show potential to capture emergent dynamics of BGC with Empirical Dynamics

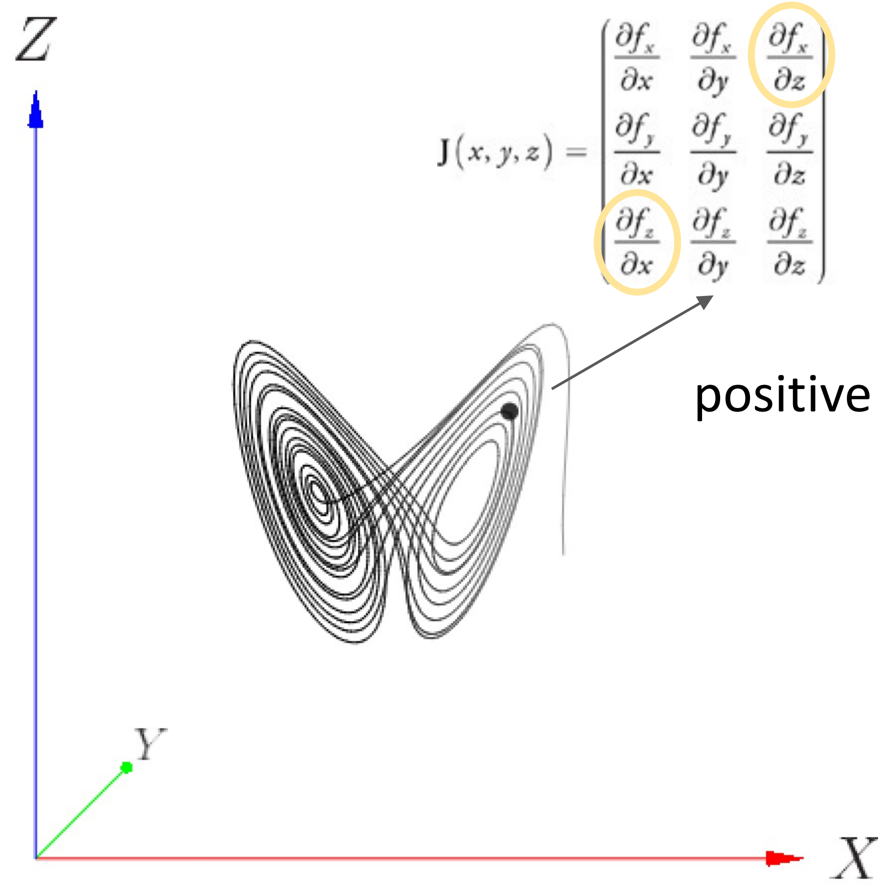


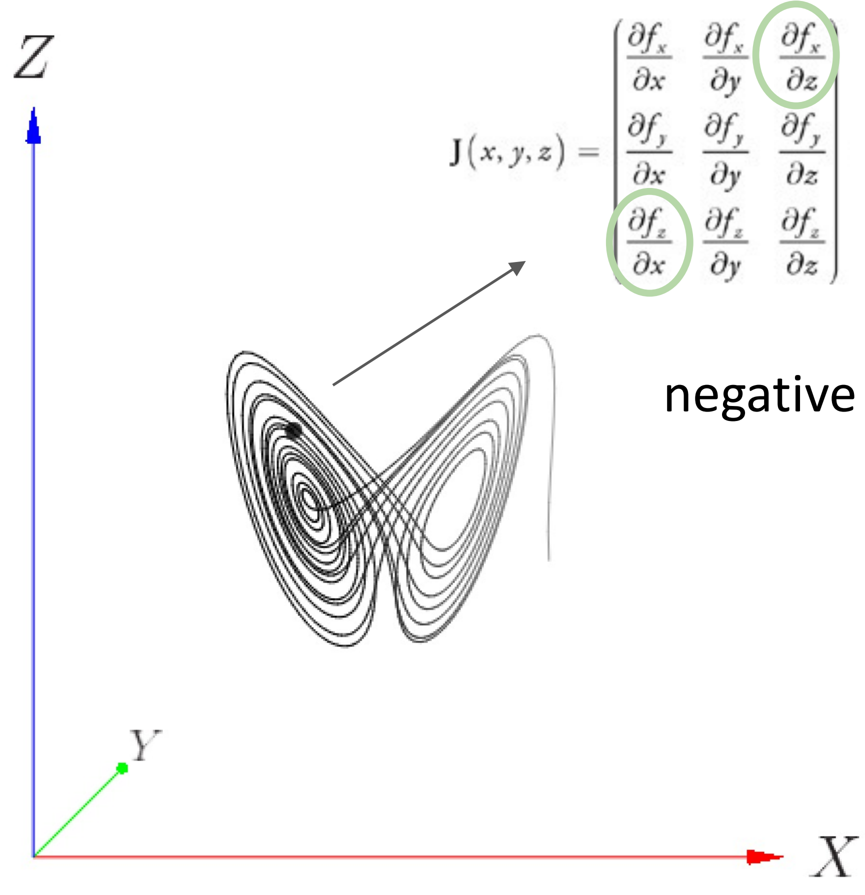

Nonlinearly tuned S-map models forecast substantially better than vector auto-regression

Incorporating biogeochemical variables leads to improved forecasts

rEDM package available at (github.com/SugiharaLab/rEDM)

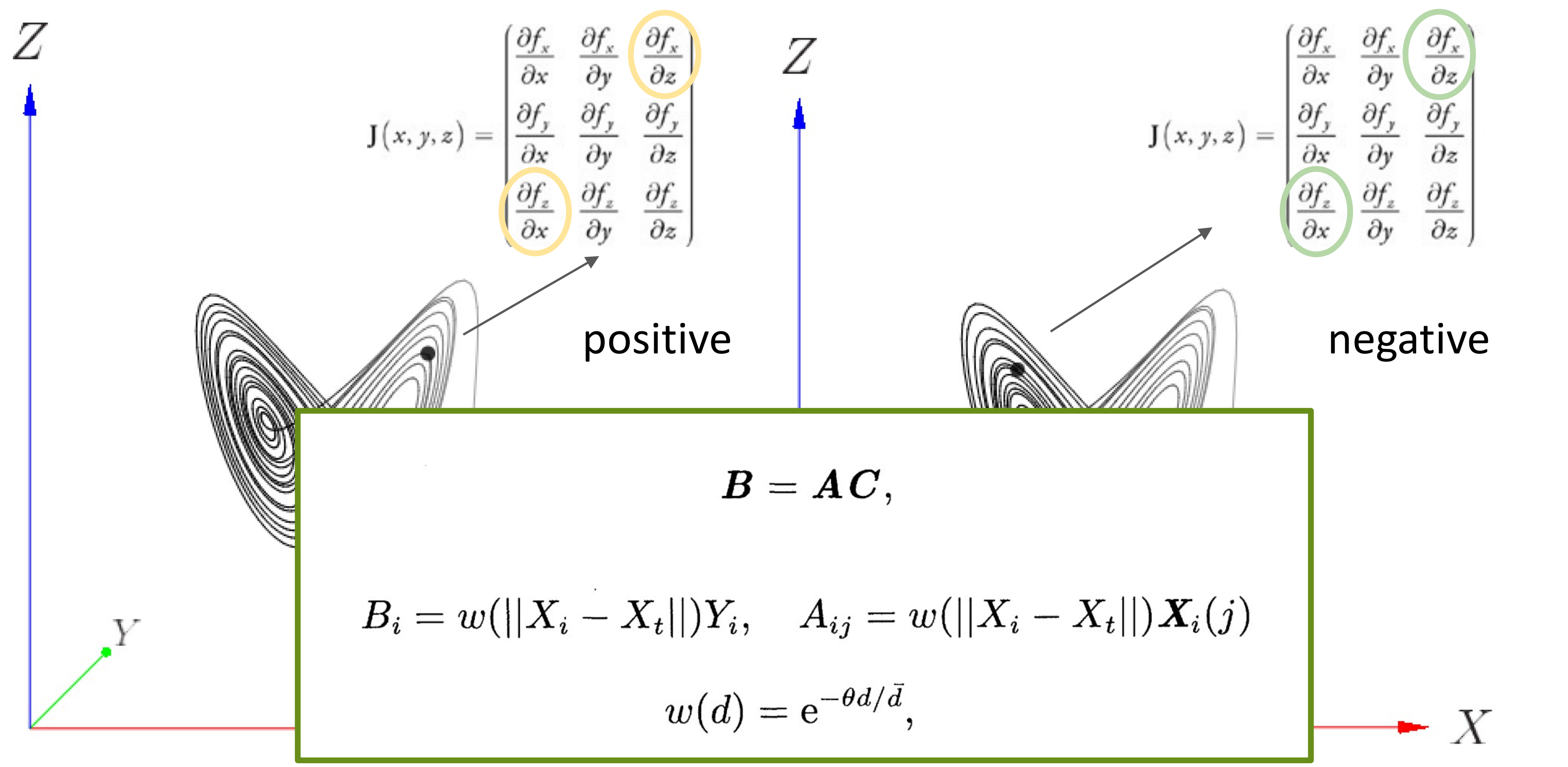
# Interpretability: extracting rates and interaction coefficients.




Deyle et al. 2016 *Proc Roy Soc B*

# Interpretability: extracting rates and interaction coefficients.




$$\boldsymbol{B} = \boldsymbol{A}\boldsymbol{C},$$

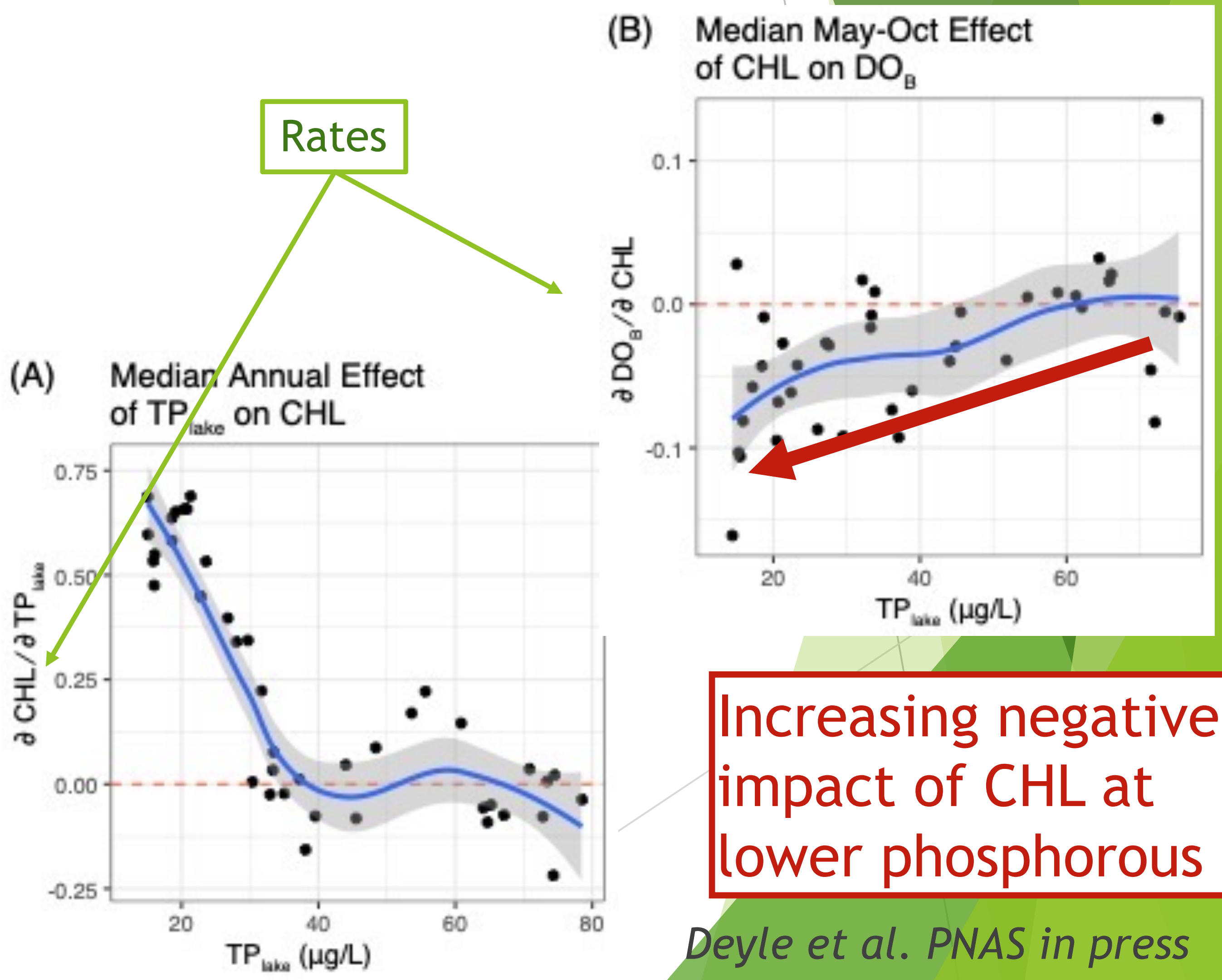
$$B_i = w(||X_i - X_t||)Y_i, \quad A_{ij} = w(||X_i - X_t||)\boldsymbol{X}_i(j)$$

$$w(d) = \mathrm{e}^{-\theta d/\bar{d}},$$

Deyle et al. 2016 *Proc Roy Soc B*

# Hybrid Approaches to Hypoxia: Lake Geneva Study

- Fixed, parameterized rates really most immediately appropriate to capture a fixed community / food-web.
- These are changing! *Non-stationary world.*
- Low-dimensional nonlinear regression (aka simple supervised machine learning) can represent changing interactions between variables.

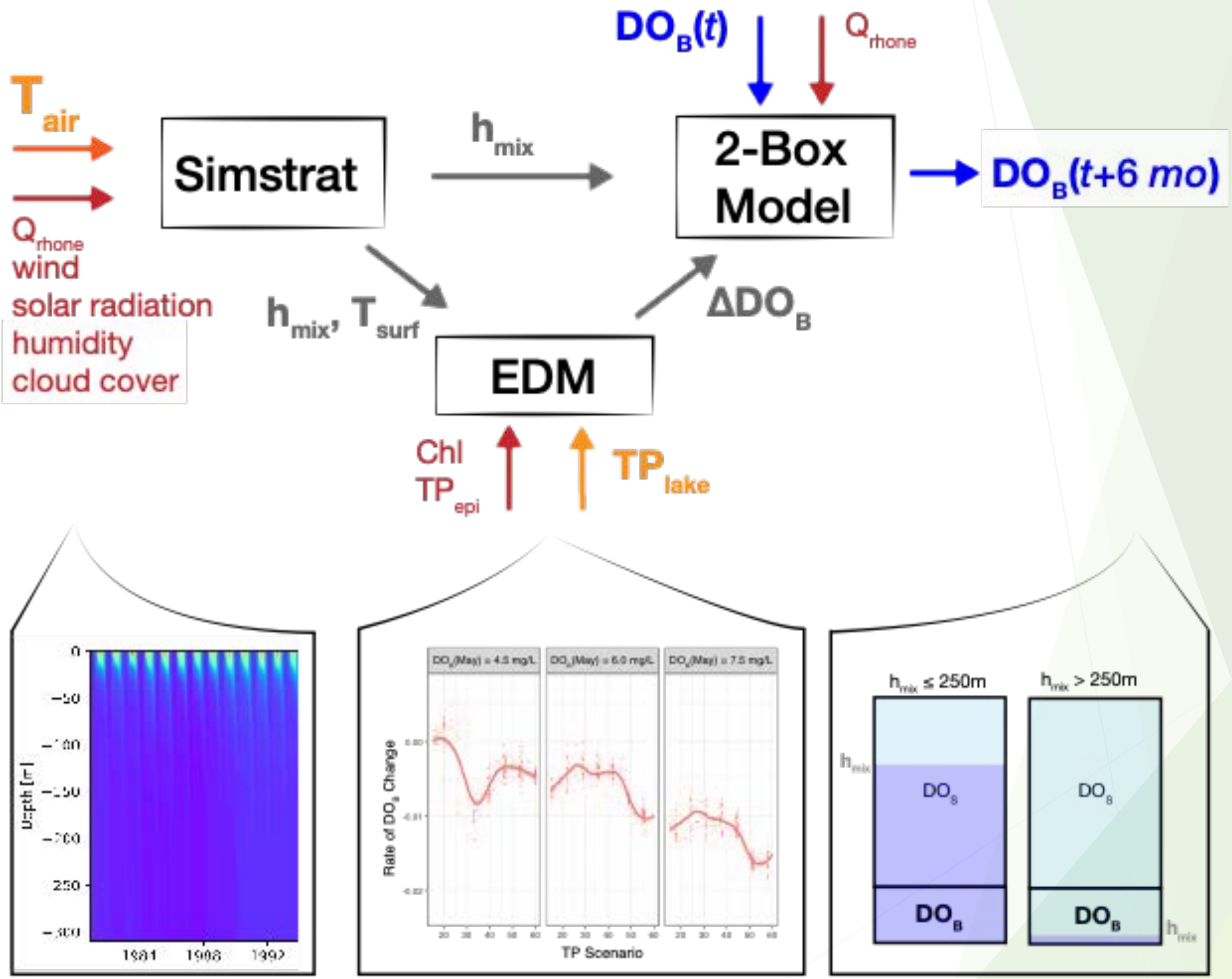



Increasing negative impact of CHL at lower phosphorous

*Deyle et al. PNAS in press*

# First-princples + empirical

2 box model but instead of parameterized equations for ecosystem processes (e.g. primary production and respiration), use empirical dynamic models.




*Deyle et al. PNAS in press*

# Closing Thoughts

# There are many flexible approaches in machine learning to make forecasts but practical considerations can lead to strong preferences.

- Kernel Regression:
  - Have to “carry the data around” to use the empirical model.
- Random Forest
  - Relatively lightweight specification, can just carry around a sparse matrix and a few other things.

# There are many flexible approaches in machine learning to make forecasts but underlying assumptions can limit interpretability of some versus others.

- Kernel Regression:
  - Dynamic interactions can be read nearly straight out of the model.
- Random Forest
  - Dynamic interactions are 0 almost everywhere and undefined on a sparse, finite set.

# There are many flexible approaches in machine learning to make forecasts but <u>accessibility remains a major challenge that can override any other consideration.</u>

- Kernel Regression:
  - Can characterize forecast uncertainty in a frequentist
  - Full version controlled, documented R, Python, and C++ packages (Hao Ye, Joseph Park) with training materials.

- Gaussian Processes:
  - Can use a fully Bayesian framework for uncertainty propagation.
  - Currently available packages poorly matched to ecological forecasting use-cases.

github.com/SugiharaLab

# Thank you!

## Capabilities

- Sidestep structural uncertainty.
- Robust to unobserved variables.
- Practically cope with shifts in ecology, rather than trying to approximate systems as fixed and unchanging.
- Can be part of forecasting non-stationary and non-equilibrium futures.

## Challenges

- Interoperability of modeling frameworks.
    - Both the parametric hydrodynamic model and our “EDM” package went through major version turn-over.
- Using as “black boxes” can obscure bias.
- Knobs & Tuning can be hiding (what data do you put in?).