Extracting Environmental and Climate Signals in Coastal Ecosystem Data with Two-Stage Multivariate Dynamic Linear Models

Presenter: Jacob P. Strock,

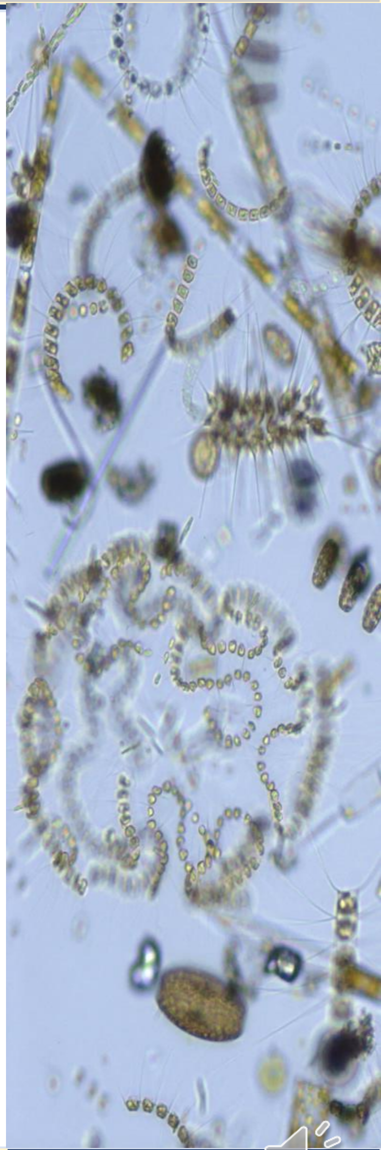
Dr. Gavino Puggioni,

Dr. Susanne Menden-Deuer

JSM

August 11, 2021

Why Phytoplankton?



What are phytoplankton?

- Microscopic marine algae

Why study phytoplankton?

- Production of 49-60 Gt C yr^{-1} (Carr et al. 2006)
- Drive global biogeochemical cycling (e.g. carbon; Falkowski 1994)
- Form the basis of production in marine food webs (Steinberg and Landry 2017)

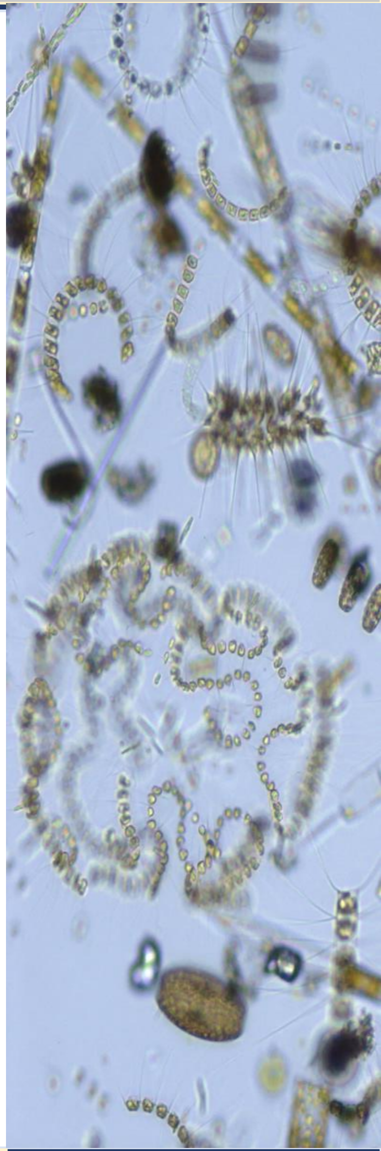
Why is phytoplankton size important?

Phytoplankton size affects:

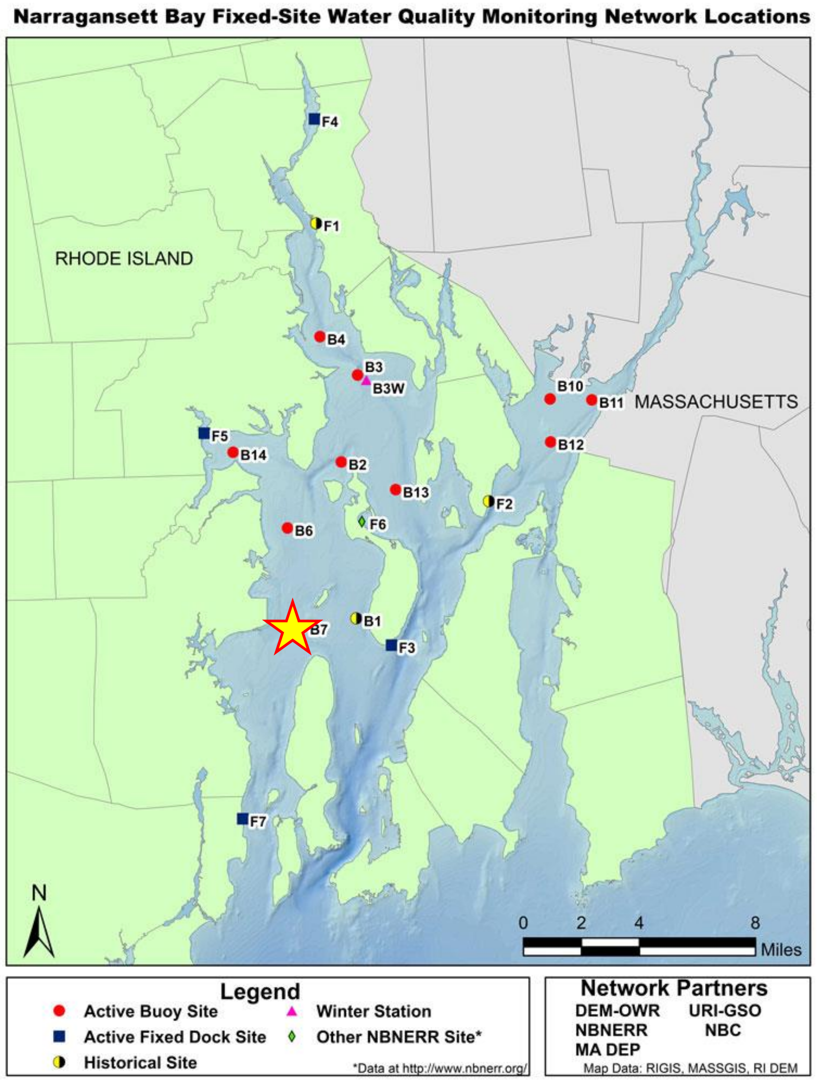
- Metabolic rate (López-Urrutia et al. 2011)
- Algae bloom formation (Irigoien et al. 2005)
- Food chain length (Sprules and Munawar 1986)

Phytoplankton size affected by:

- Cell size scales inversely with temperature (Atkinson et al. 2003)
- Cell size directly imposes a physical constraint on the potential rate of nutrient supply (e.g. Mei et al. 2009)



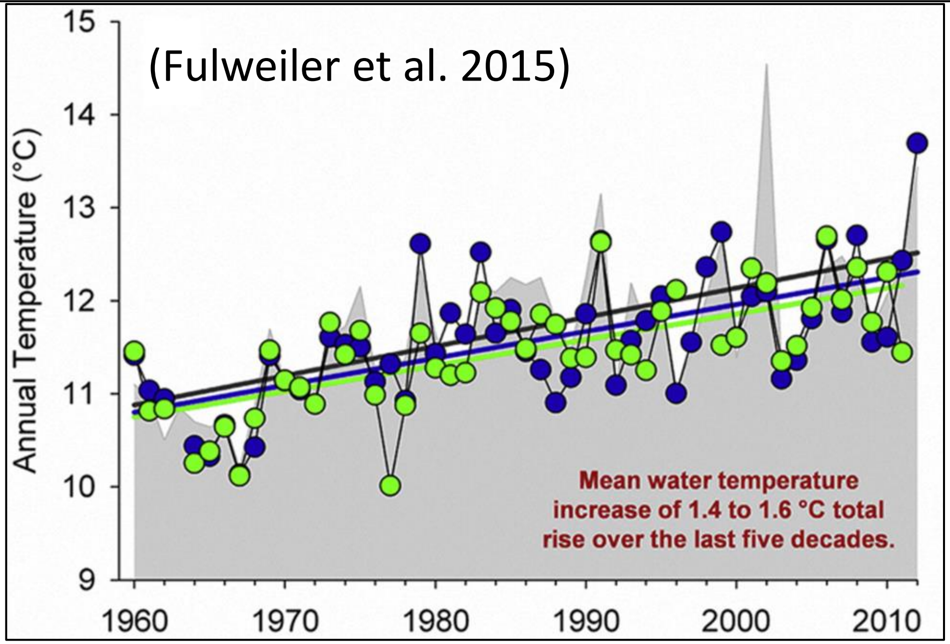
Could ecosystem changes be changing cell size locally?



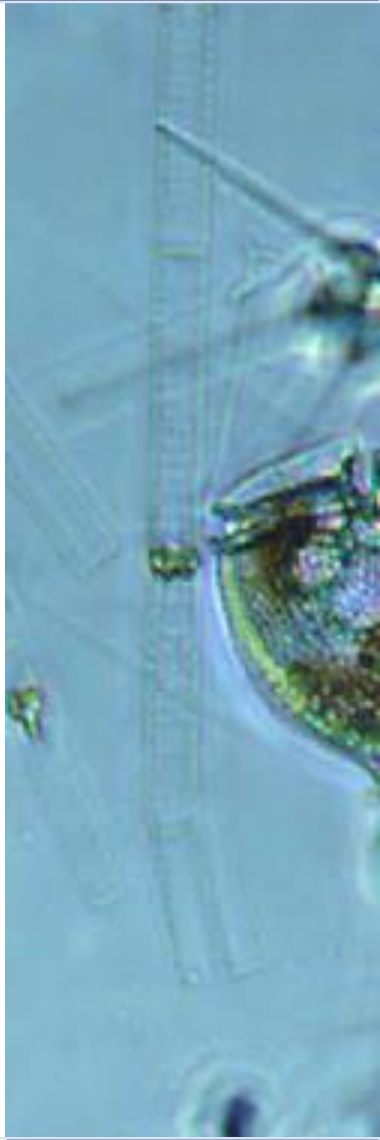
http://www.dem.ri.gov/programs/emergencyresponse/bart/stations.php

In Narragansett Bay, Rhode Island, USA:

- **Temperature** has shown an increase of 1.4-1.6°C by a linear regression of mean annual levels (Fulweiler et al. 2015)
- **Nitrogen** pollution in wastewater was required by RI Department of Environmental Management to be cut 50% from 2005 to 2012 (RI DEM 2005)



Research Objective

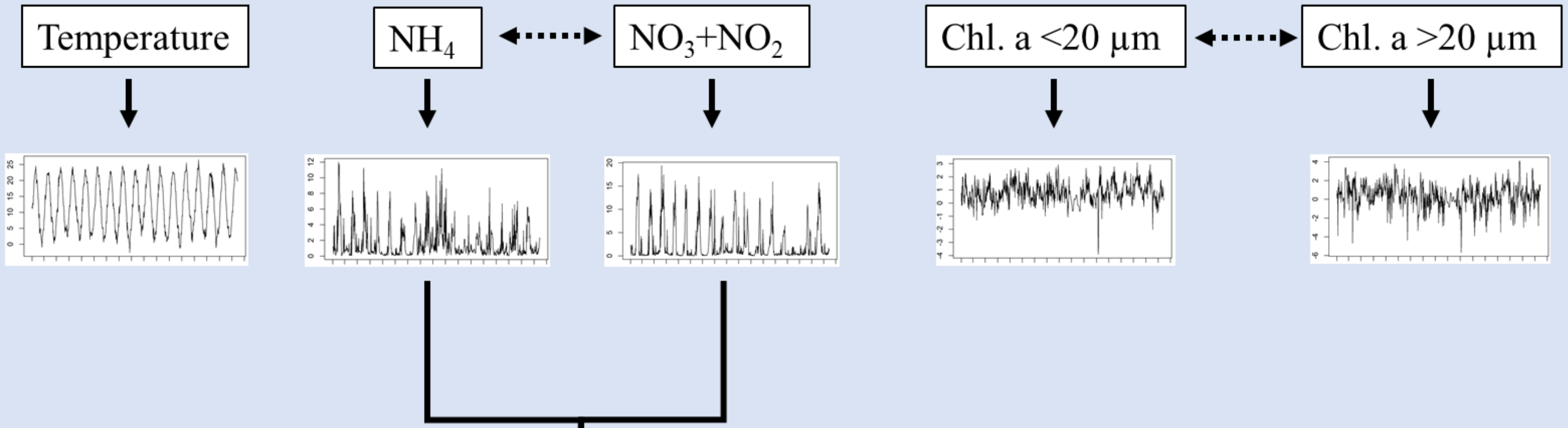


Use Bayesian Dynamic Linear Models to:

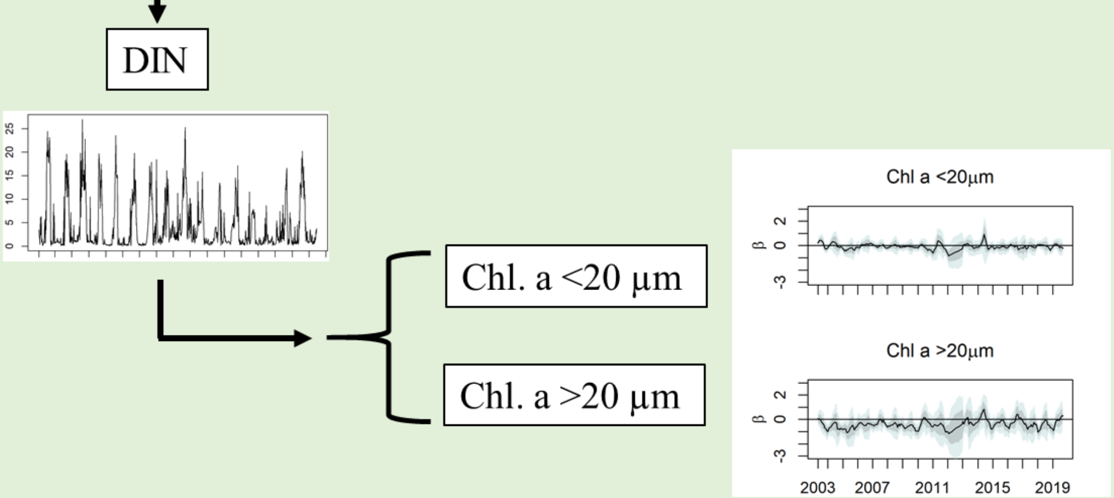
1. Characterize long-term changes in Narragansett Bay (2003-2019):
   - Temperature
   - DIN (Dissolved inorganic nitrogen)
   - Size Fractionated Chlorophyll
     (< 20µm, >20µm, proxy for phytoplankton biomass)
2. Impute missing data
3. Test the influence of DIN on phytoplankton size structure through dynamic linear regression

Outline

Stage 1 Models



Stage 2 Model



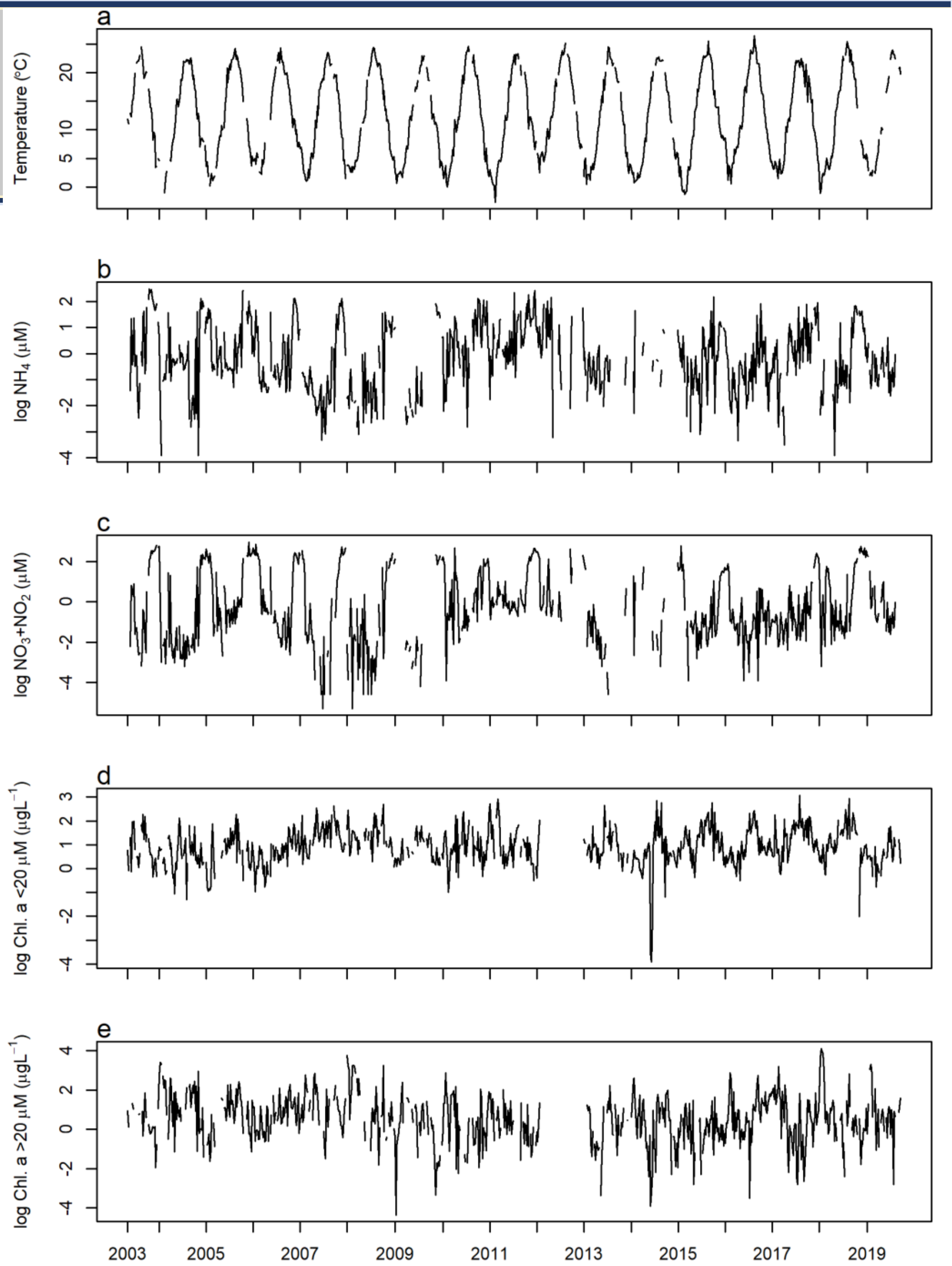
Purpose

- Missing data imputation
- Exploration
- Inference:
    - Long-term change
    - Seasonal change

- Describe the relationship of DIN and size structure

Data Description

- Narragansett Bay Long-term Plankton Time Series
  - Source: https://web.uri.edu/gso/research/plankton/
- Time-period:
  - April 29, 2003- September 17, 2019
  - Weekly resolution
- Variables:
  - water temperature (°C)
  - NH_4
  - NO_3+NO_2 (µm)
  - Chlorophyll a < 20 µm (µg L^{-1})
  - Chlorophyll a > 20 µm (µg L^{-1})
- Missingness:
  - MCAR
  - Missingness lengths 1-48 observations



Dynamic Linear Models

$$
\begin{tikzcd}
\boxed{\theta_1} \arrow[r] \arrow[d] & \boxed{\theta_2} \arrow[r] \arrow[d] & \cdots \arrow[r] & \boxed{\theta_{t-1}} \arrow[r] \arrow[d] & \boxed{\theta_t} \arrow[r] \arrow[d] & \boxed{\theta_{t+1}} \arrow[d] \\
Y_1 & Y_2 & & Y_{t-1} & Y_t & Y_{t+1}
\end{tikzcd}
$$

Dynamic Linear Models

$$\begin{tikzcd}
\theta_1 \arrow[r] \arrow[d] & \theta_2 \arrow[r] \arrow[d] & \cdots \arrow[r] & \theta_{t-1} \arrow[r] \arrow[d] & \theta_t \arrow[r] \arrow[d] & \theta_{t+1} \arrow[d] \\
Y_1 & Y_2 & & Y_{t-1} & Y_t & Y_{t+1}
\end{tikzcd}$$

$$Y_t = \boldsymbol{F}_t\boldsymbol{\theta}_t + v_t, \qquad v_t \sim N(0, \boldsymbol{V}) \quad \text{Observation equation}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + w_t, \qquad w_t \sim N(0, \boldsymbol{W}) \quad \text{State equation}$$

$$\left.\begin{array}{l} \boldsymbol{\theta}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0) \\ \boldsymbol{V} \sim IG(a_v, b_v) \\ \boldsymbol{W} \sim IW(a_w, b_w) \end{array}\right\} \text{Priors}$$

Variable	Definition
Y_t	Datum/Data
F_t	Observational matrix
G_t	Evolutional Matrix
θ_t	Unobserved state
V_t	Observational variance/covariance
W_t	Evolutional variance/covariance

Dynamic Linear Models

$$
\begin{tikzcd}
\boxed{\theta_1} \arrow[r] \arrow[d] & \boxed{\theta_2} \arrow[r] \arrow[d] & \cdots \arrow[r] & \boxed{\theta_{t-1}} \arrow[r] \arrow[d] & \boxed{\theta_t} \arrow[r] \arrow[d] & \boxed{\theta_{t+1}} \arrow[d] \\
Y_1 & Y_2 & & Y_{t-1} & Y_t & Y_{t+1}
\end{tikzcd}
$$

$$Y_t = \boldsymbol{F}_t \boldsymbol{\theta}_t + v_t, \qquad v_t \sim N(0, \boldsymbol{V}) \quad \text{Observation equation}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + w_t, \qquad w_t \sim N(0, \boldsymbol{W}) \quad \text{State equation}$$

$$\left.\begin{aligned} &\boldsymbol{\theta}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0) \\ &\boldsymbol{V} \sim IG(a_v, b_v) \\ &\boldsymbol{W} \sim IW(a_w, b_w) \end{aligned}\right\} \text{Priors}$$

Variable	Definition
Y_t	Datum/Data
F_t	Observational matrix
G_t	Evolutional Matrix
θ_t	Unobserved state
V_t	Observational variance/covariance
W_t	Evolutional variance/covariance

Why?

- Flexible structure
- Components are additive
  - long-term trend, season, regression components separately
- Any parameter and component can time vary
  - Important when we are hypothesizing changing ecosystem function
- Interpolation via inference
- Quantify uncertainty in missingness and states

Stage 1: Models, Environment

Stage	Model	Response Variable(s)	θ Components	V Specification	W Specification
1	1	Temperature	Dynamic, Intercept (μ)	Static, *IG* prior	Static, *IW* prior
			Dynamic, Season (S_i), i=1		
	2	log NH_4, log NO_3 + NO_2	Dynamic, Intercept (μ)	Static, *IW* prior	Static, *IW* prior
			Dynamic, Season (S_i), i=1,...,6		
	3	log Chl. <20, log Chl. >20	Dynamic, Intercept (μ)	Static, *IW* prior	Static, *IW* prior
			Dynamic, Season (S_i), i=1,...,6		

Stage 1

$$\begin{cases} \boldsymbol{Y}_t^Q = F_t^Q \boldsymbol{\theta}_t^Q + v_t^Q, & v_t^Q \sim \mathcal{N}_r(\mathbf{0}, \boldsymbol{V}_t^Q) \\ \boldsymbol{\theta}_t^Q = \boldsymbol{G}_t^Q \boldsymbol{\theta}_{t-1}^Q + w_t^Q, & w_t^Q \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{W}_t^Q) \end{cases} \quad (2)$$

Stage 2

$$\begin{cases} \boldsymbol{Y}_t^Z = F_t^Z \boldsymbol{\theta}_t^Z + v_t^Z, & v_t^Z \sim \mathcal{N}_z(\mathbf{0}, \boldsymbol{V}_t^Z) \\ \boldsymbol{\theta}_t^Z = \boldsymbol{G}_t^Z \boldsymbol{\theta}_{t-1}^Z + w_t^Z, & w_t^Z \sim \mathcal{N}_{rw}(\mathbf{0}, \boldsymbol{W}_t^Z \otimes \boldsymbol{V}_t^Z) \\ F_{X,t}^Z \sim P(g((\Psi_{Stage1}^X))) \end{cases} \quad (3)$$

$$\Psi_0 \sim \Pi_{k=1}^K P(\Psi_{0,k})$$

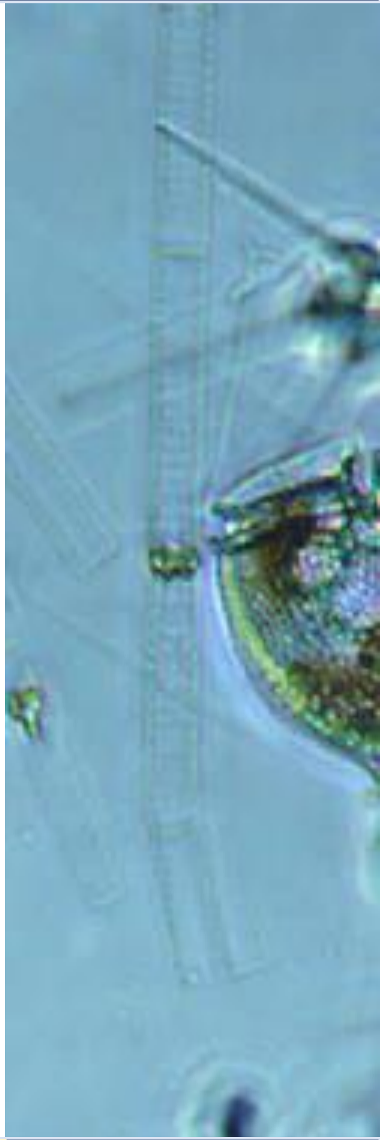
$$\Psi_0 = \{\theta_0^Z, \theta_0^Q, V_0^Z, V_0^Q, W_0^Z, W_0^Q\}$$

Stage 1: Models, Environment

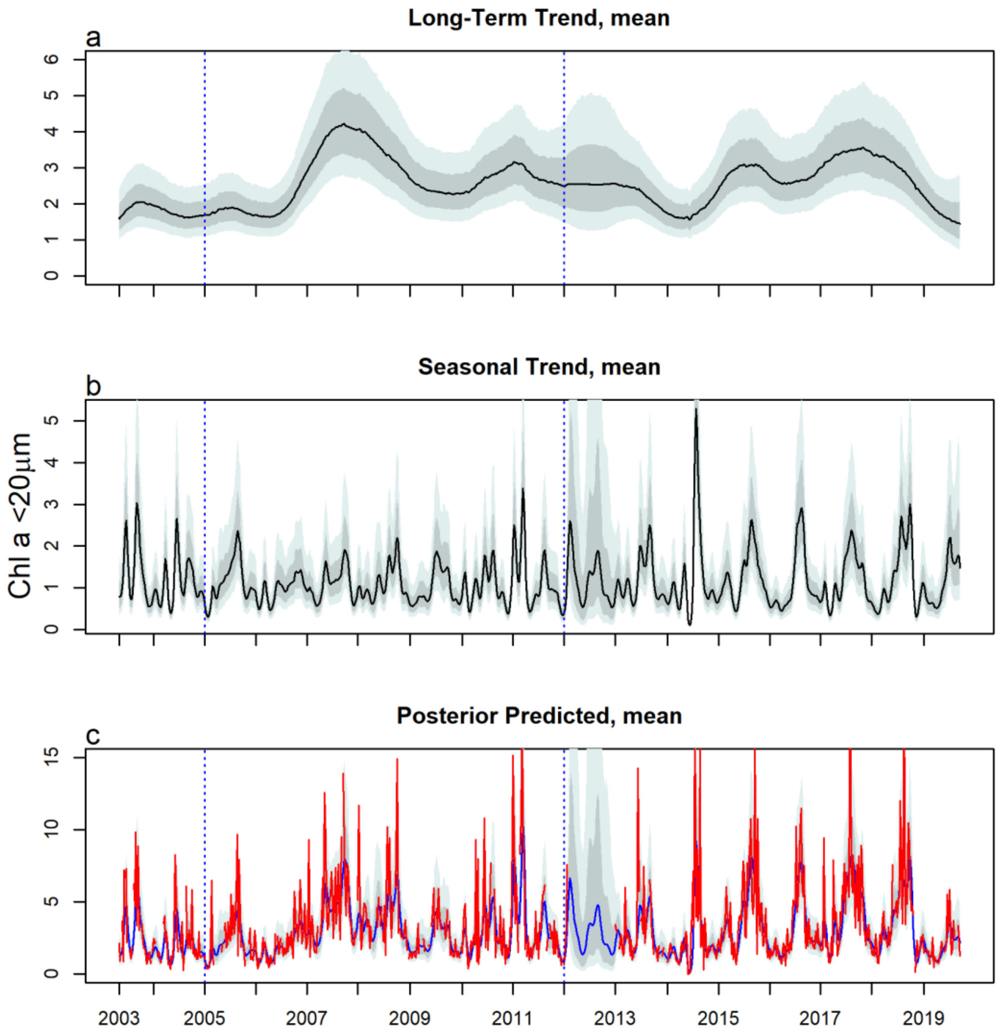
Inference and Sampling:

- Non-conjugate (no closed form of posterior distribution)
- Markov Chain Monte Carlo (MCMC) Simulation of Posteriors

- Gibbs Sample each conditional posterior distribution
  - $\theta_{1:T} | Y_{1:T}, V, W$ Kalman Filtering and Smoothing
  - $V | Y_{1:T}, W, \theta_{1:T}$ $\sim \; Inverse - Gamma \; (\nu_0 + T, S_0 + RSS_y)$
  - $W | Y_{1:T}, V, \theta_{1:T}$ $\sim \; Inverse - Wishart \; (\nu_0 + T, S_0 + RSS_\theta)$



Decomposing seasonal and long-term trends



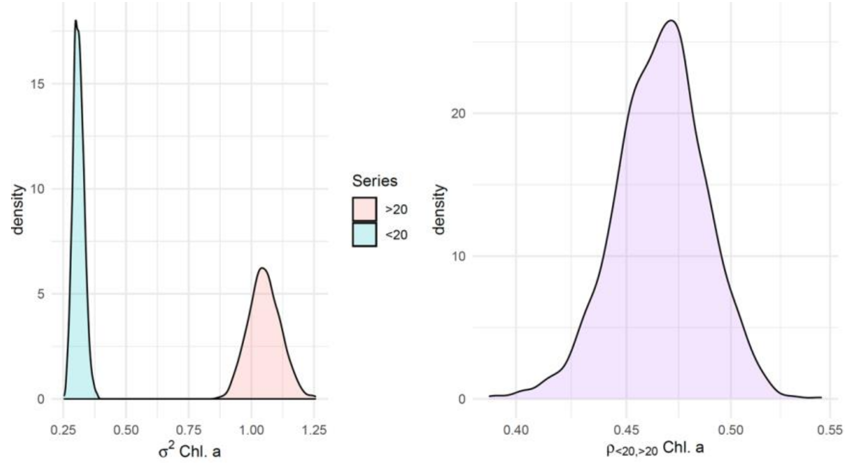
Stage 1: Models, Environment

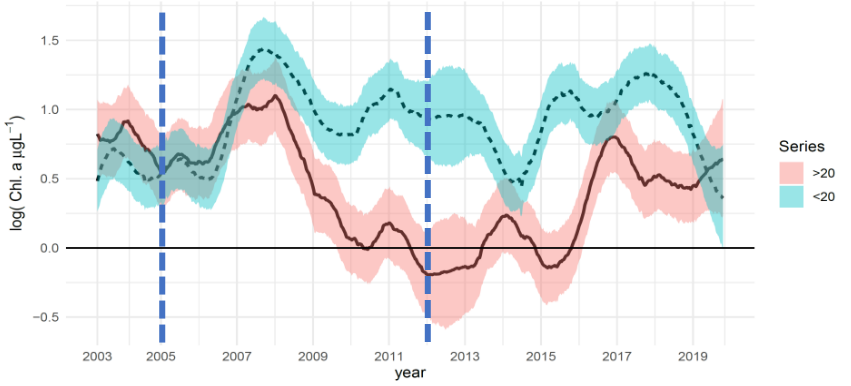
Chl. a < 20 µm

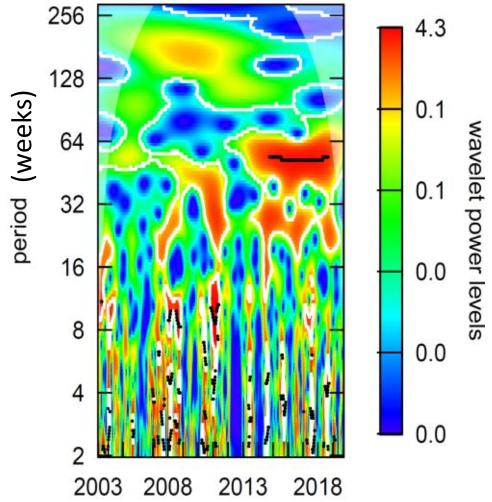
- Less variable
- Seasonal peak in autumn
- Increase in mean levels
- Intensifying seasonality

Chl. a > 20 µm

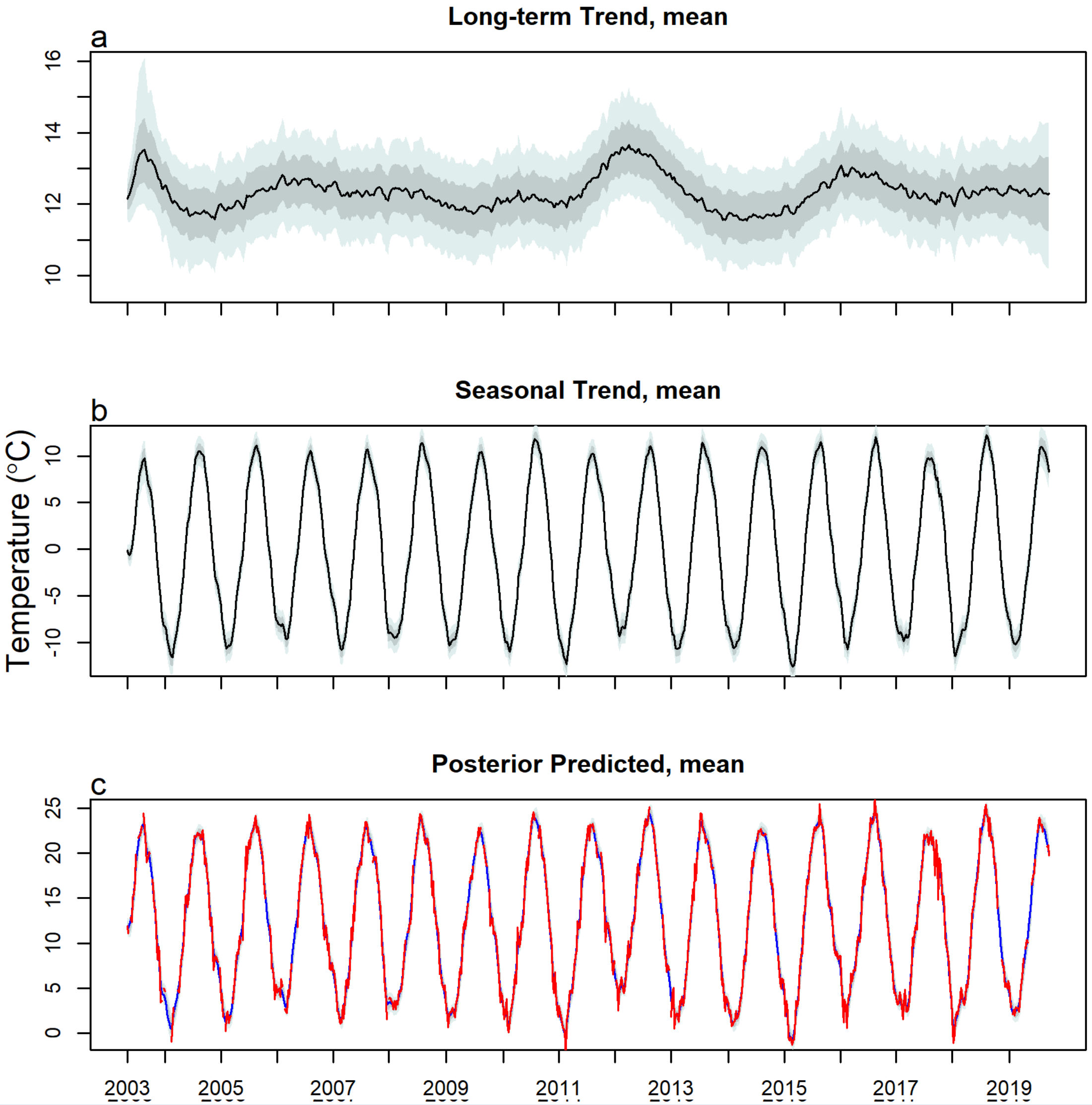
- More variable
- Seasonal peak in spring
- Decrease in mean levels
- Seasonal intensity inversely correlated with mean levels

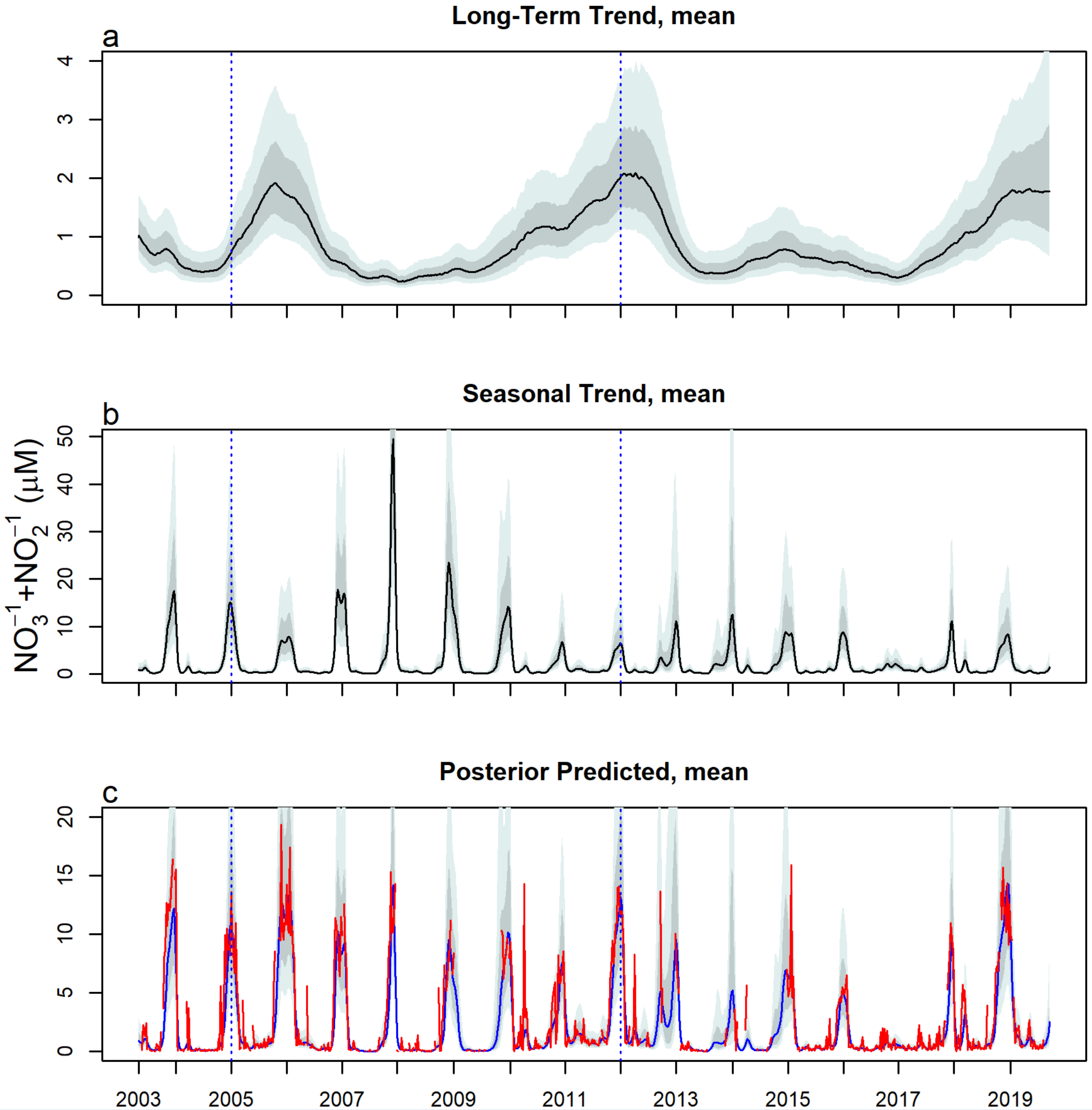






Temperature & Nitrogen Species





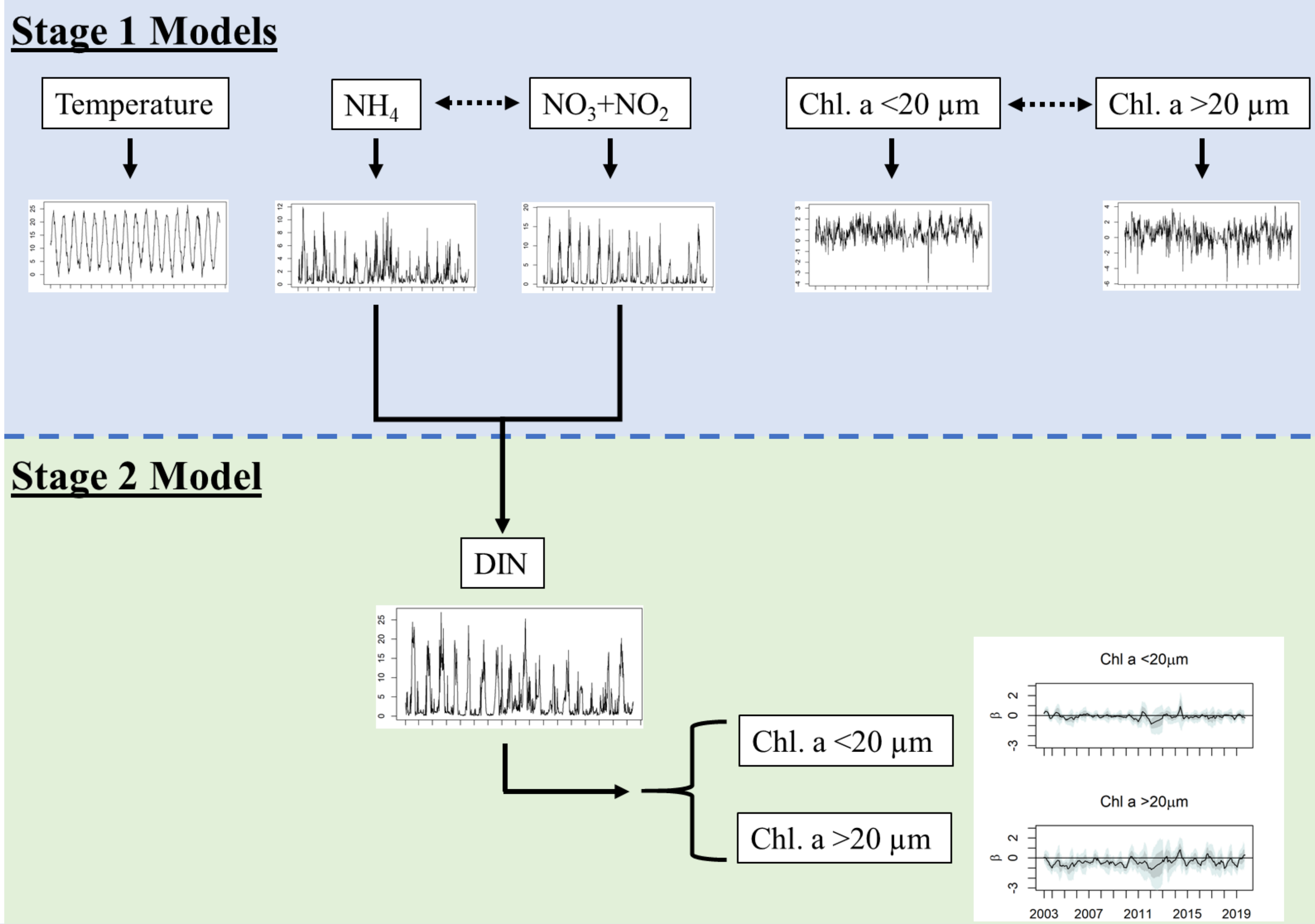
Outline

Stage 2 Model: Multivariate Dynamic Linear Regression

Stage 2 Model: Regression

Stage	**Model**	**Response Variable(s)**	***θ* Components**	***V* Specification**	***W* Specification**
2	4	log Chl. <20, log Chl. >20	Dynamic, Intercept (μ) Dynamic, Regression on DIN Static, Season (S_i), i=1	Static, *IW* prior	Dynamic, Fixed Discount Factor

Outline

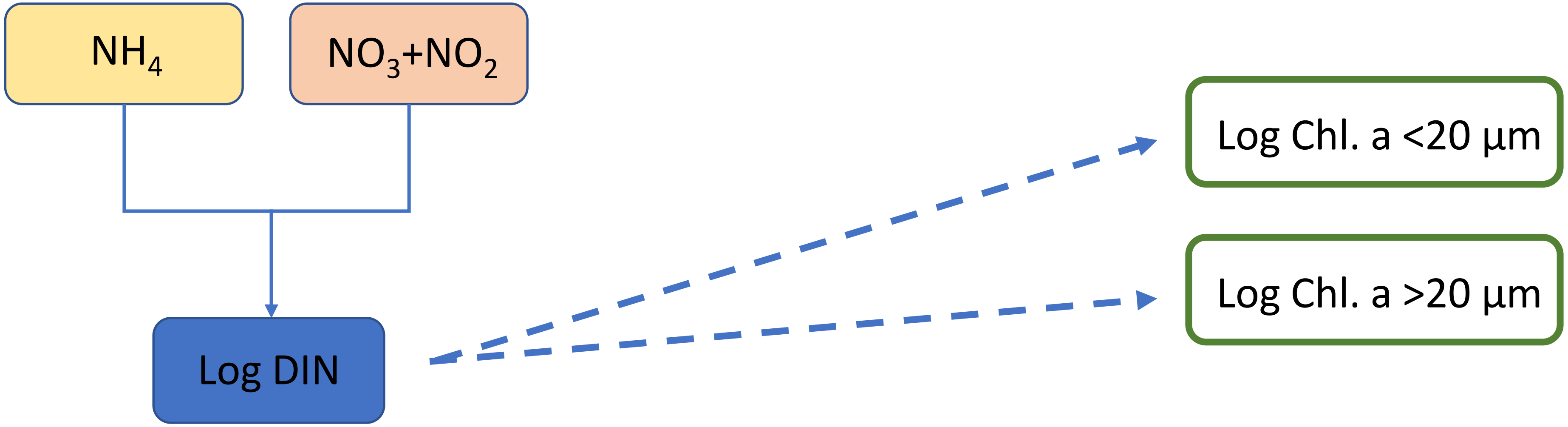


Purpose

- Missing data imputation
- Exploration
- Inference:
  - Long-term change
  - Seasonal change

- Describe the relationship of DIN and size structure

DIN as a regressor



$$\log DIN = \log(\exp(F_{1:T}\theta_{NH_4,1:T}) + \exp(F_{1:T}\theta_{NO_{3,2},1:T}))$$

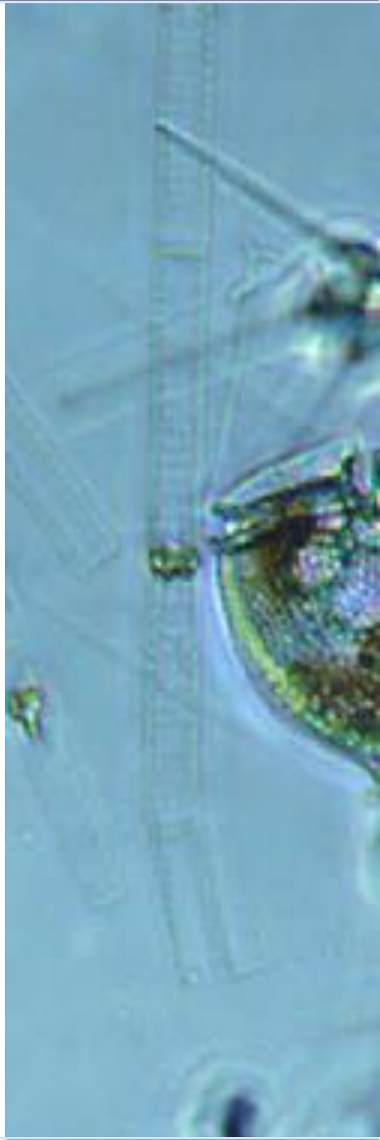
$$\theta_{NH_4,1:T}|Y_{1:T} \sim \int P(\theta_{NH_4,1:T}|Y_{1:T},\cdot)P(\cdot\,|Y_{1:T})d\,\cdot$$

$$\theta_{NO_{3,2},1:T}|Y_{1:T} \sim \int P\left(\theta_{NO_{3,2}\ 1:T}\middle|Y_{1:T},\cdot\right)P(\cdot\,|Y_{1:T})d\,\cdot$$

$$\int \underbrace{p(\Phi|DIN,\cdot)}_{\text{sample}}\underbrace{p(DIN)}_{\text{sample}}dDIN = p(\Phi|\cdot)$$

MCMC

Stage 2 Model: Regression



Inference:

- Non-conjugate (no closed form of posterior distribution)
- Markov Chain Monte Carlo (MCMC) Simulation of Posteriors
- Fixed δ levels, run in parallel in separate models
- Sample each posterior distribution
  - $DIN_{1:T} | \cdot$ From Stage 1 Model
  - $\theta_{1:T} | Y_{1:T}, V, W_t$ Kalman Filtering and Smoothing
    - W_t Discount Specification
  - $V | Y_{1:T}, W_t$ $\sim \ Inverse - Wishart \ (\nu_0 + T, S_0 + RSS_\theta)$

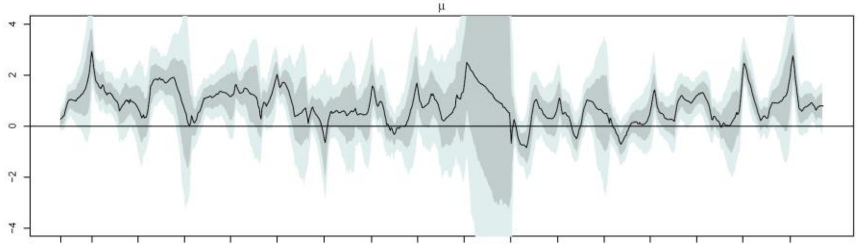
Evolutional Covariance specification:

$$W_t = \frac{1-\delta_i}{\delta_i} P_{i,t},$$
$$P_{i,t} = G_t C_{t-1} G_t'$$

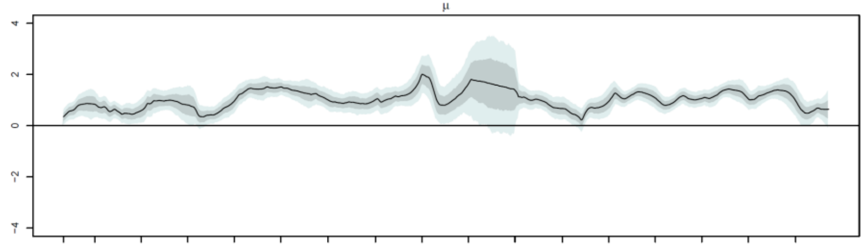
- Discount factor δ_i, describes the loss of information between time-steps
  - $\delta_i = 1$ is equivalent to a static model (no evolutional change in the states)
  - Commonly specified between 0.85 and 0.999
- Practical discounting constrains information loss to a linear rather than exponential rate during longer period missingness

$$R_t(k) = \frac{G^k C_t G'^k}{\delta^k}$$
$$if\ k > 1, \qquad R_t = G^{k-1} C_{t+1} G'^{\,k-1}$$

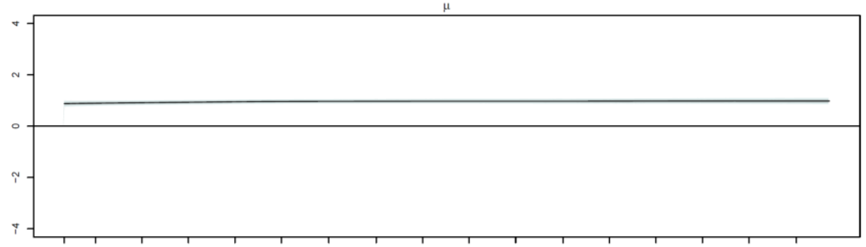
$\delta = 0.8$



$\delta = 0.9$

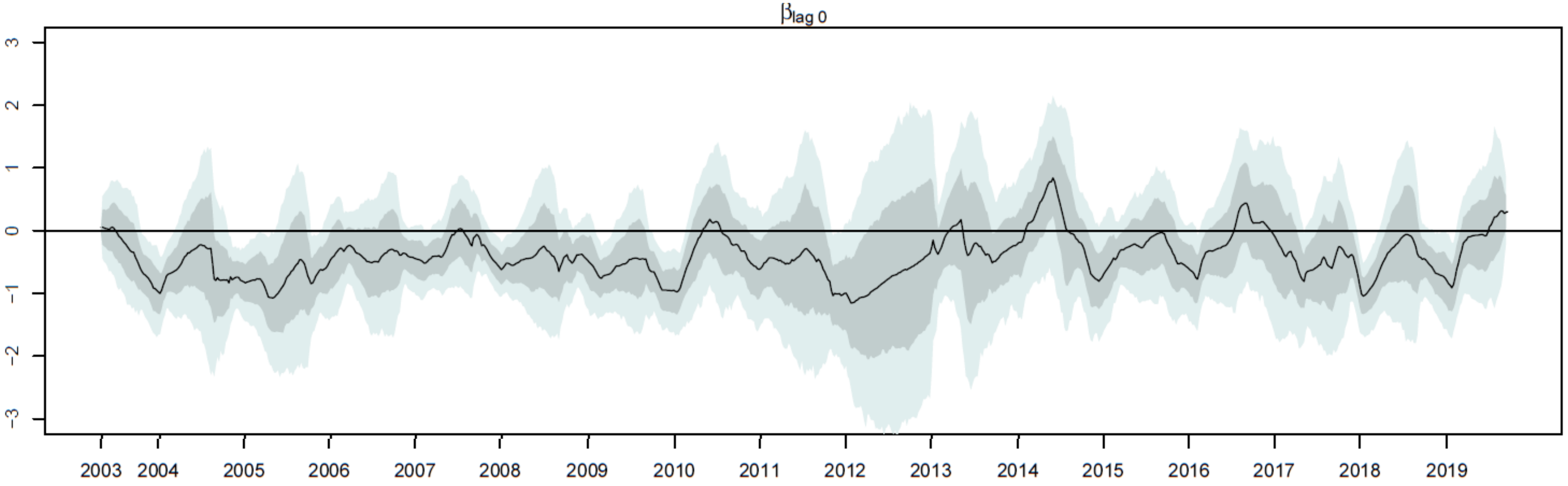


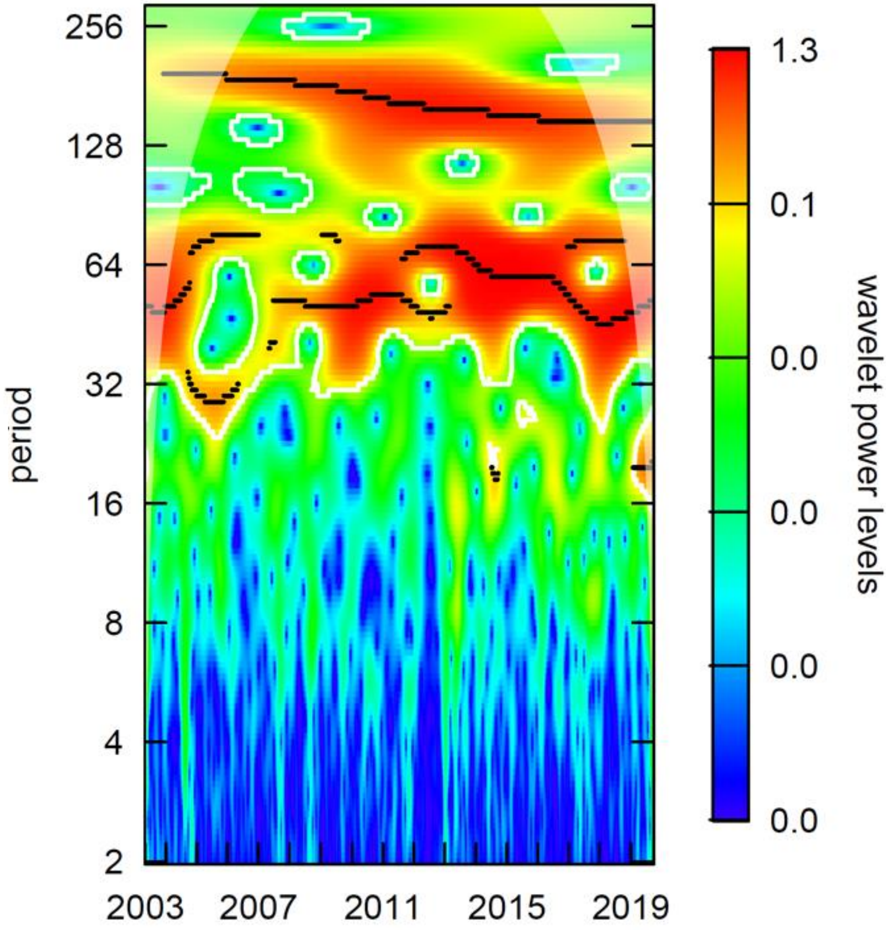
$\delta = 0.999$



Chl. a > 20 µm

- DIN significantly associated with Chl. a > 20
- Chl. a > 20 µm negatively associated with DIN, suggesting that DIN is shaped by the Chl. a > 20 µm community, but not <20 µm
- No long-term-prediction of decline by ambient DIN

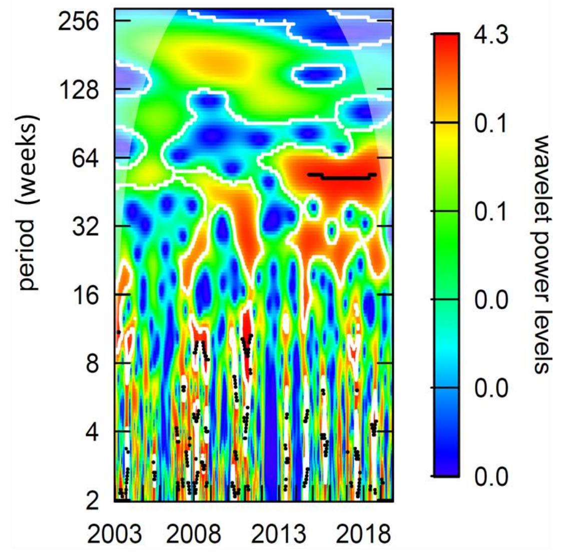


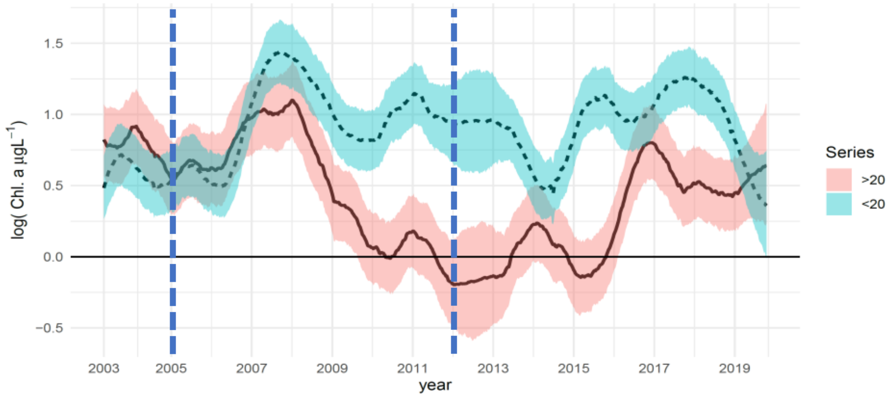


Conclusions

__Biotic Change__

- Narragansett Bay has switched from being dominated by organisms >20 µm to those <20 µm
- Seasonal variation has amplified in phytoplankton <20 µm
- The seasonally significant associations between large phytoplankton and DIN suggest large phytoplankton drive bay dynamics during winter periods

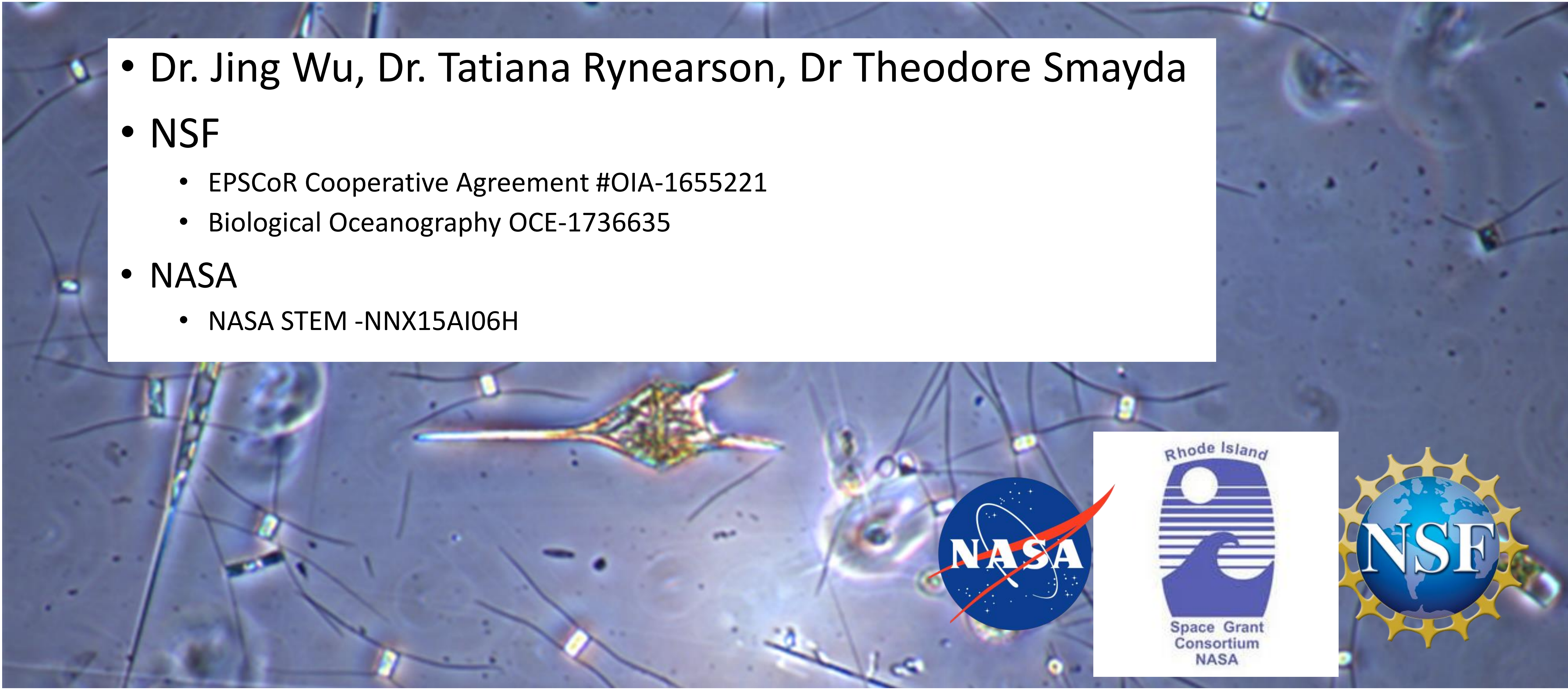




Conclusions

- The multi-stage DLM provides a convenient structure for multivariate time-series analysis
- Useful when :
  - Modeling may have multiple goals such as describing long-term patterns and regression on these features
  - Missing data or posterior parameters of time series may be useful in other models
- Advantage over a joint model due to simplicity, and computational savings during model development :
  - Posteriors of model stages can be saved and sampled, and thereby do not need to be re-estimated if later stages need to be adjusted

Acknowledgements



- Dr. Jing Wu, Dr. Tatiana Rynearson, Dr Theodore Smayda
- NSF
  - EPSCoR Cooperative Agreement #OIA-1655221
  - Biological Oceanography OCE-1736635
- NASA
  - NASA STEM -NNX15AI06H

Literature Cited

Atkinson, David, Benjamin J Ciotti, and David J S Montagnes. 2003. “Protists Decrease in Size Linearly with Temperature: Ca. 2.5% °C.” https://doi.org/10.1098/rspb.2003.2538.

Falkowski, Paul G. 1994. “The Role of Phytoplankton Photosynthesis in Global Biogeochemical Cycles.” Photosynthesis Research 39 (3): 235–58. https://doi.org/10.1007/BF00014586.

Finkel, Zoe V. 2007. “Does Phytoplankton Cell Size Matter? The Evolution of Modern Marine Food Webs.” In Evolution of Primary Producers in the Sea, 333–50. Academic Press.

Fulweiler, R. W., A. J. Oczkowski, K. M. Miller, C. A. Oviatt, and M. E.Q. Pilson. 2015. “Whole Truths vs. Half Truths - And a Search for Clarity in Long-Term Water Temperature Records.” Estuarine, Coastal and Shelf Science 157 (May): A1–6. https://doi.org/10.1016/j.ecss.2015.01.021.

Masson-Delmotte, V., Zhai, P., Pörtner, H. O., Roberts, D., Skea, J., Shukla, P. R., ... & Connors, S. (2018). Global warming of 1.5 C. An IPCC Special Report on the impacts of global warming of, 1.

Irigoien, X., K.J. Flynn, and R.P. Harris. 2005. “Phytoplankton Blooms: A ‘Loophole’ in Microzooplankton Grazing Impact? | Journal of Plankton Research | Oxford Academic.” Journal of Plankton Research 27 (4): 313–21. https://academic.oup.com/plankt/article/27/4/313/1508420.

Keller, Aimee A., Grace Klein-MacPhee, and Jeanne St Onge Burns. "Abundance and distribution of ichthyoplankton in Narragansett Bay, Rhode Island, 1989–1990." Estuaries 22, no. 1 (1999): 149-163.

López-Urrutia, Angel, Elena San Martin, Roger P Harris, and Xabier Irigoien. 2011. “Scaling the Metabolic Balance of the Oceans.” Proceedings of the National Academy of Sciences, 103. www.pnas.orgcgidoi10.1073pnas.0601137103.

Masson-Delmotte, Valérie, Panmao Zhai, Hans-Otto Pörtner, Debra Roberts, Jim Skea, Priyadarshi R. Shukla, Anna Pirani et al. "Global warming of 1.5 C." An IPCC Special Report on the impacts of global warming of 1 (2018).

Mei, Zhi-Ping, Zoe V Finkel, and Andrew J Irwin. 2009. “Light and Nutrient Availability Affect the Size-Scaling of Growth in Phytoplankton.” Journal of Theoretical Biology 259: 582–88. https://doi.org/10.1016/j.jtbi.2009.04.018.

Nixon, Scott W. "Prehistoric nutrient inputs and productivity in Narragansett Bay." Estuaries 20, no. 2 (1997): 253-261.

Nixon, Scott W., Robinson W. Fulweiler, Betty A. Buckley, Stephen L. Granger, Barbara L. Nowicki, and Kelly M. Henry. 2009. “The Impact of Changing Climate on Phenology, Productivity, and Benthic-Pelagic Coupling in Narragansett Bay.” Estuarine, Coastal and Shelf Science 82 (1): 1–18. https://doi.org/10.1016/j.ecss.2008.12.016.

Oviatt, Candace, Leslie Smith, Jason Krumholz, Catherine Coupland, Heather Stoffel, Aimee Keller, M. Conor McManus, and Laura Reed. "Managed nutrient reduction impacts on nutrient concentrations, water clarity, primary production, and hypoxia in a north temperate estuary." Estuarine, Coastal and Shelf Science 199 (2017): 25-34.

Rhode Island Department of Environmental Management (RIDEM). "Plan for Managing Nutrient Loadings to Rhode Island Waters." (2005).

Sprules, W. Gary, and M. Munawar. "Plankton size spectra in relation to ecosystem productivity, size, and perturbation." Canadian Journal of Fisheries and Aquatic Sciences 43, no. 9 (1986): 1789-1794.

Steinberg, Deborah K., and Michael R. Landry. 2017. “Zooplankton and the Ocean Carbon Cycle.” Annual Review of Marine Science 9 (1): 413–44. https://doi.org/10.1146/annurev-marine-010814-015924.