Evaluating the prediction skill of correlative estuarine species distribution models trained with mechanistic model output

Dante M. L. Horemans
Virginia Institute of Marine Science

Marjorie A. M. Friedrichs, Pierre St-Laurent, Raleigh R. Hood, and Christopher W. Brown

20 November 2024,
Paris, France

Chesapeake Bay Environmental Forecast System (CBEFS)

www.vims.edu/cbefs

DATA PRODUCTS

- **Chesapeake Bay Environmental Forecast System**
  - Background
  - Hypoxia (Oxygen)
  - Dead Zone Size
  - Depth to Low Oxygen
  - Hypoxia Line Plots
  - Bay-wide Salinity
  - Bay-wide Temperature
  - Focused Salinity and Temperature Forecasts
  - Acidification Forecasts
  - Harmful Algal Blooms
  - Pathogens (Vibrio)
  - Sea Nettles
  - Waves
  - Contact Information and Requests
- **Dead Zone Forecasts**
- **Sea-Level Report Cards**
- **Tidewatch**

CBEFS

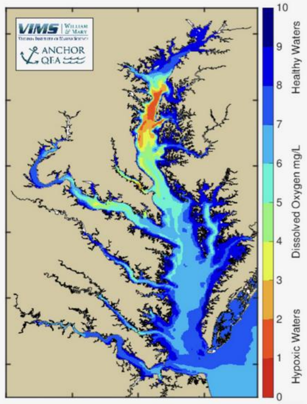
Chesapeake Bay Environmental Forecast System

Use our forecasts and "nowcasts" of temperature, salinity, dissolved oxygen, and other physical and chemical factors within the Chesapeake Bay to help monitor Bay health and plan your on-the-water activities. Based on observations and **computer models** developed by the Virginia Institute of Marine Science and partners, these tools accurately predict the current status of important environmental variables and how they are likely to change in the short-term.

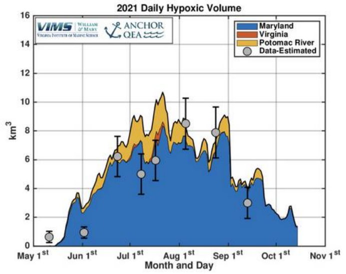
Our Chesapeake Bay Environmental Forecast System simulates 3 conditions for each selected variable:

1. **Nowcast**: present-day status of selected variable in Chesapeake Bay
2. **2-Day Forecast**: status of selected variable in the Bay 2 days from now, and
3. **Forecast Trend**: difference between nowcast and forecast (% change over 2 days)

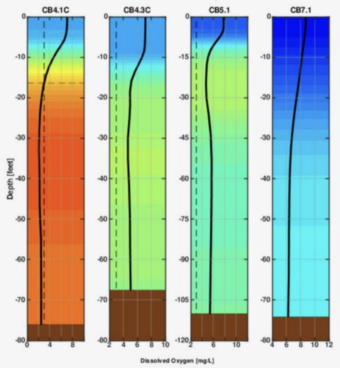
Click a selection below to access the specified simulation. Please see the **contact information page** for data requests and general contact information.



DISSOLVED OXYGEN (DO)



DEAD ZONE SIZE



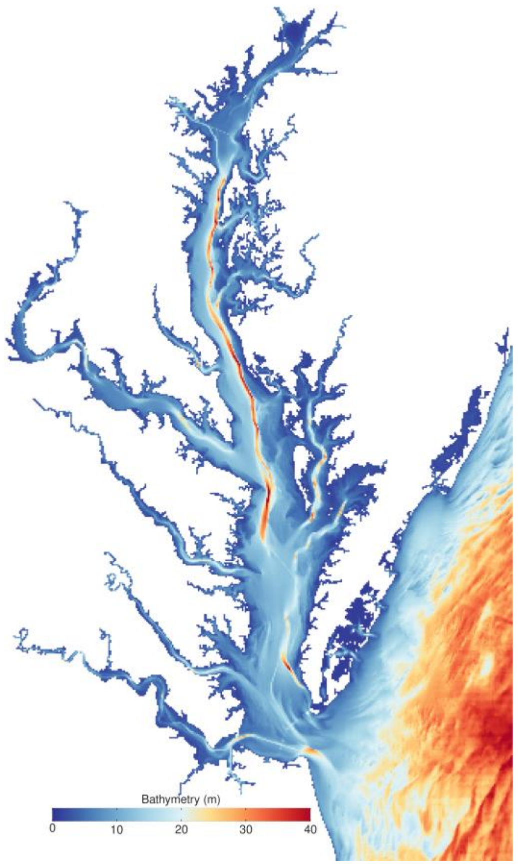
DEPTH TO LOW OXYGEN

Chesapeake Bay Environmental Forecast System (CBEFS)

- 3D hydrodynamic – biogeochemical ROMS-ECB
- Implementation of the Regional Ocean Modeling System (ROMS)
- 600 m x 600 m
- 20 vertical levels

Atmospheric inputs

Land inputs

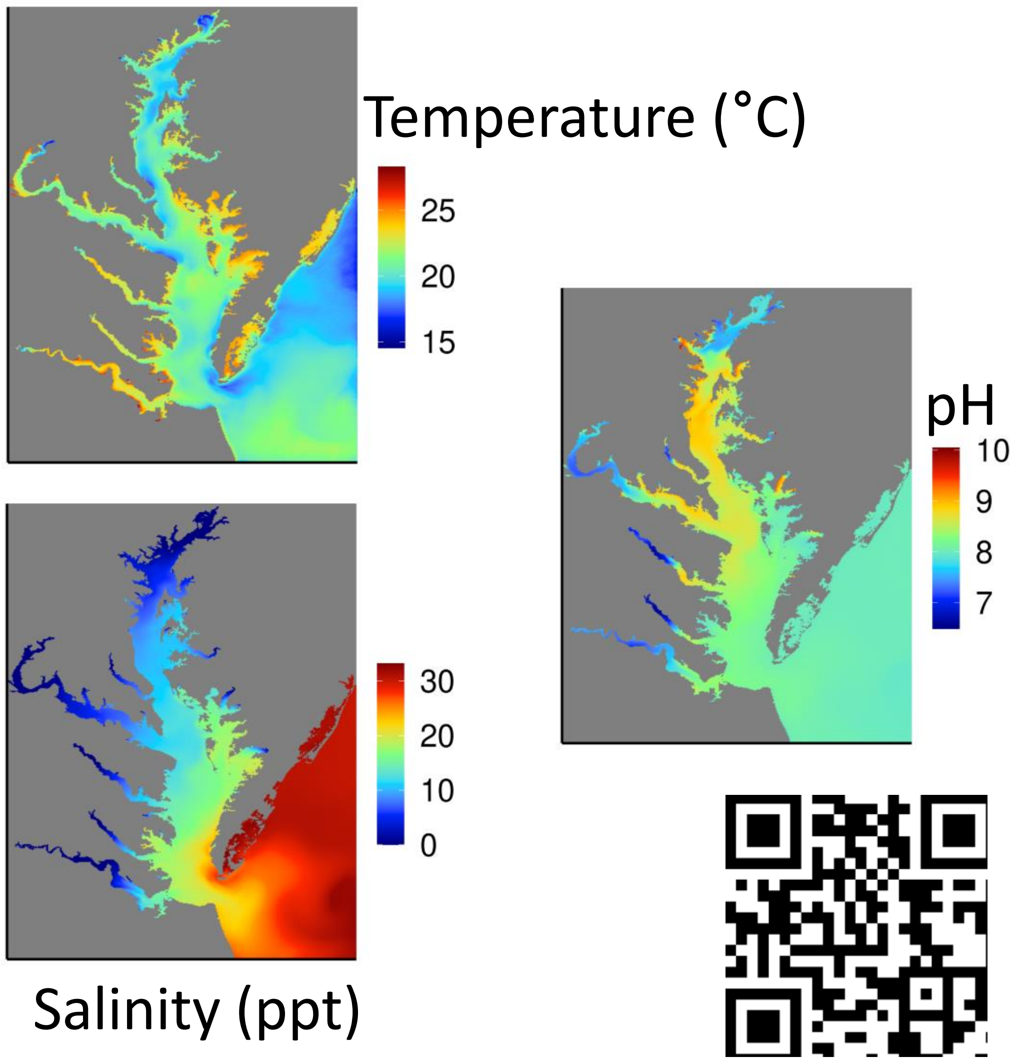


Coastal inputs

Bever et al., Env Mod & Software, 2021
St-Laurent et al., BG, 2020

Extend CBEFS with forecasts of harmful algal blooms

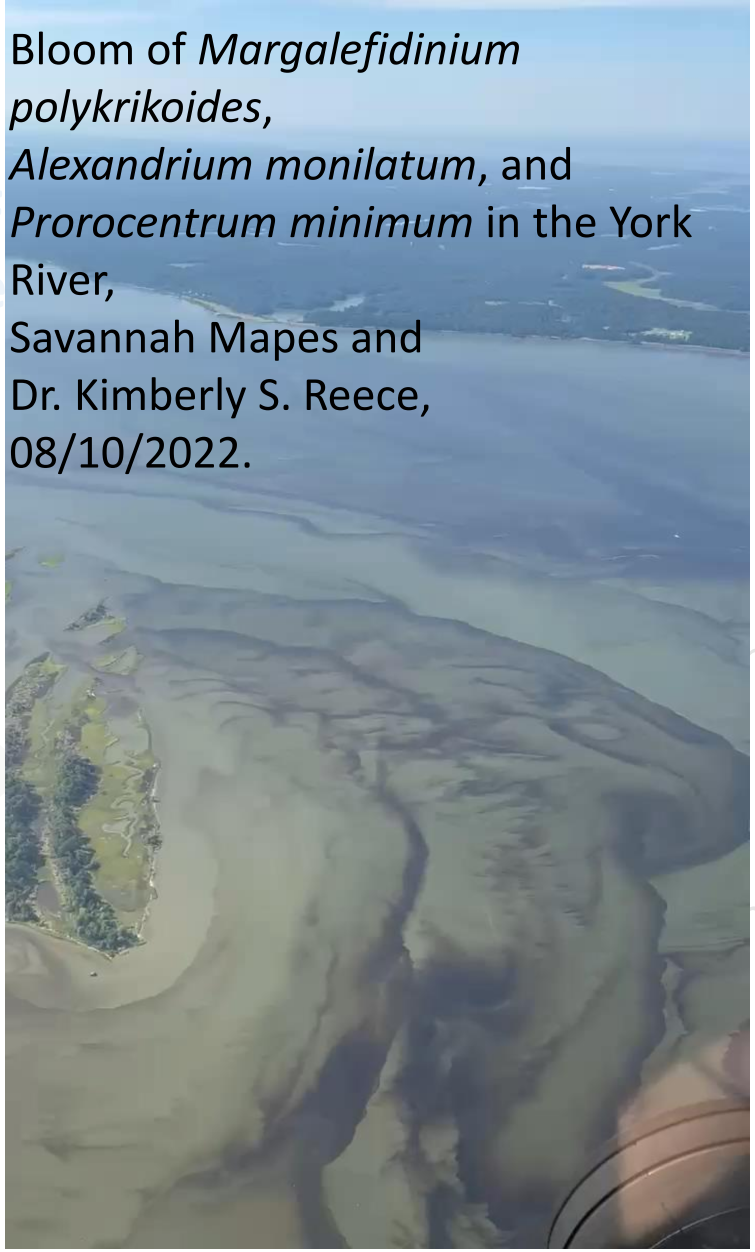
Existing model forecasts using a mechanistic model - Chesapeake Bay Environmental Forecasting System (CBEFS)



in

Statistical model

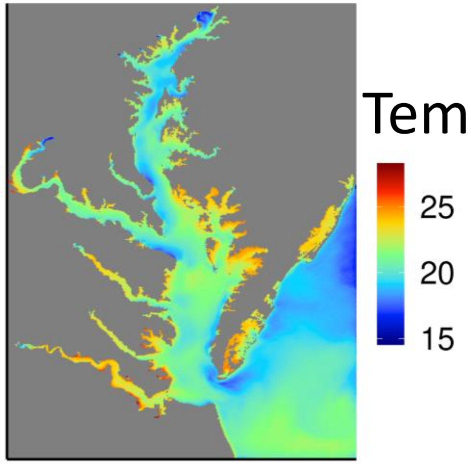
out


Bloom of *Margalefidinium polykrikoides*, *Alexandrium monilatum*, and *Prorocentrum minimum* in the York River, Savannah Mapes and Dr. Kimberly S. Reece, 08/10/2022.

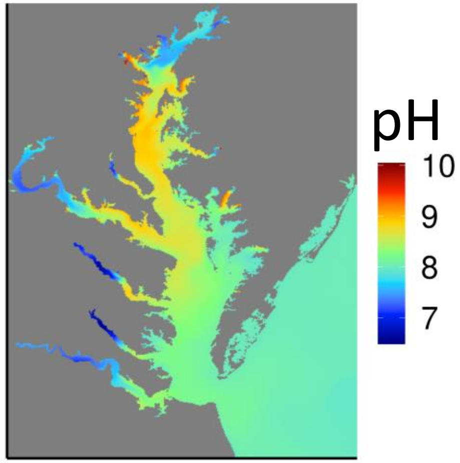
Extend CBEFS with forecasts of harmful algal blooms

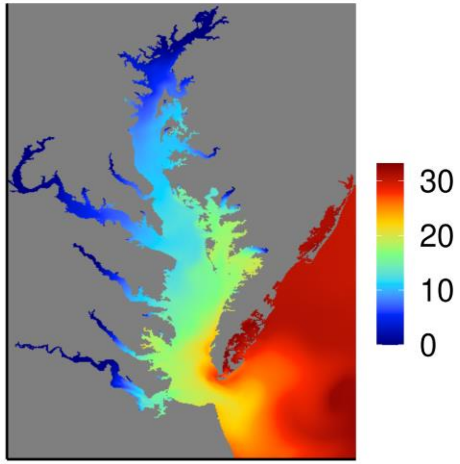
Existing model forecasts using a mechanistic model - Chesapeake Bay Environmental Forecasting System (CBEFS)

Temperature (°C)



pH

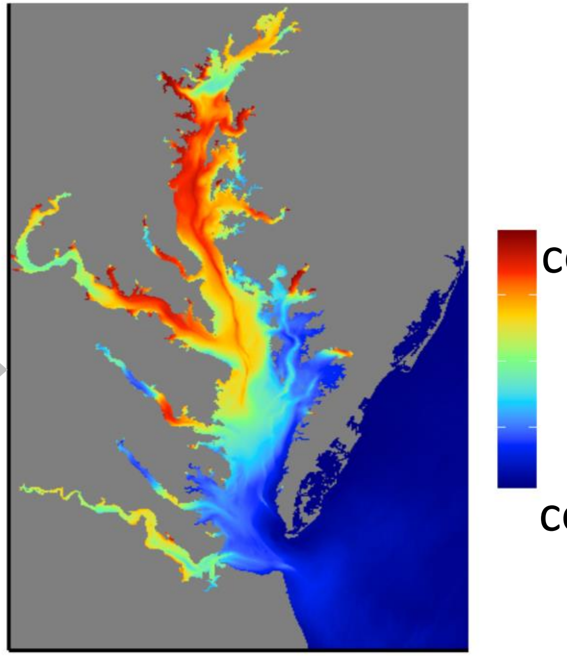




Salinity (ppt)

in → **Statistical model** → out

Forecast how suitable the environmental conditions are for a bloom

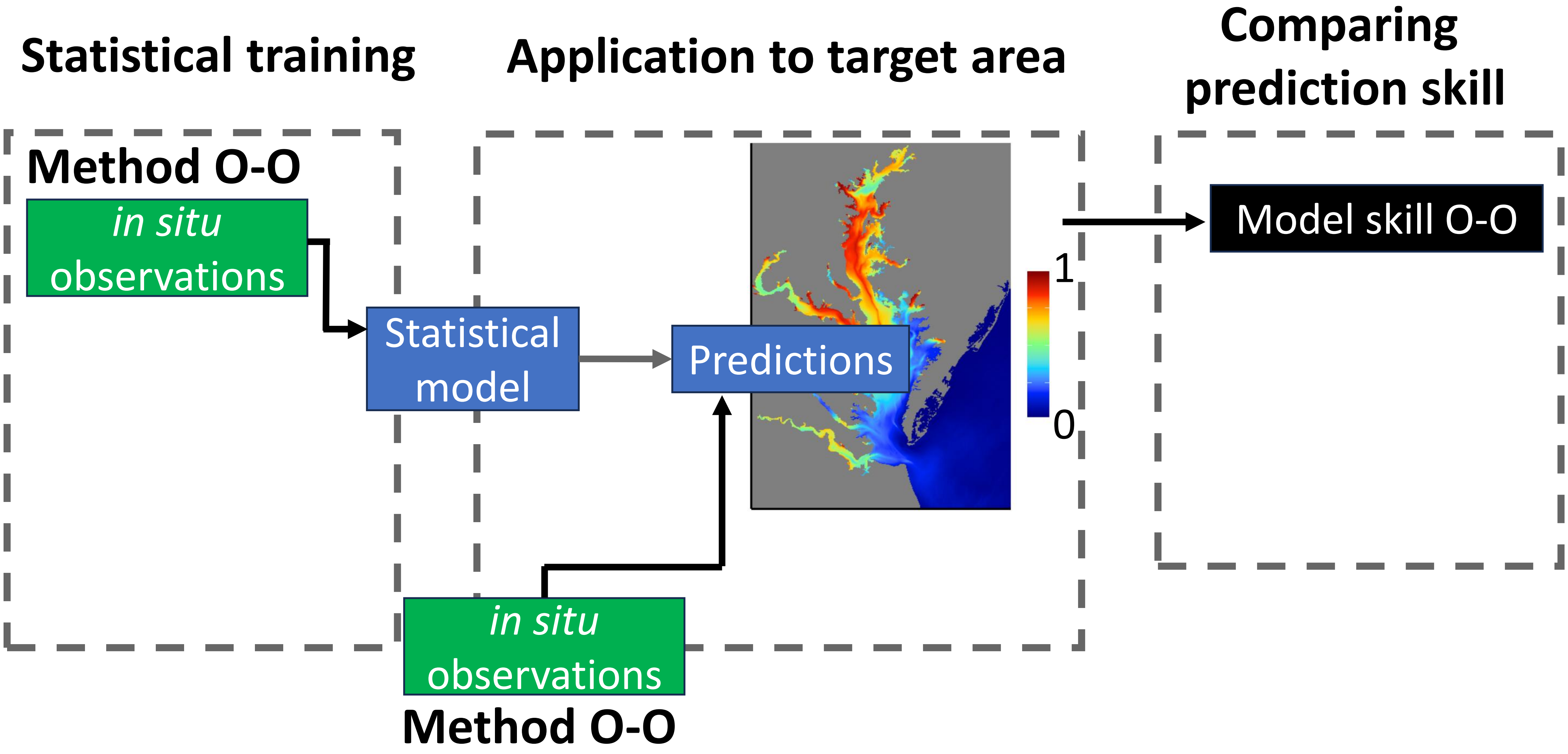


Research question

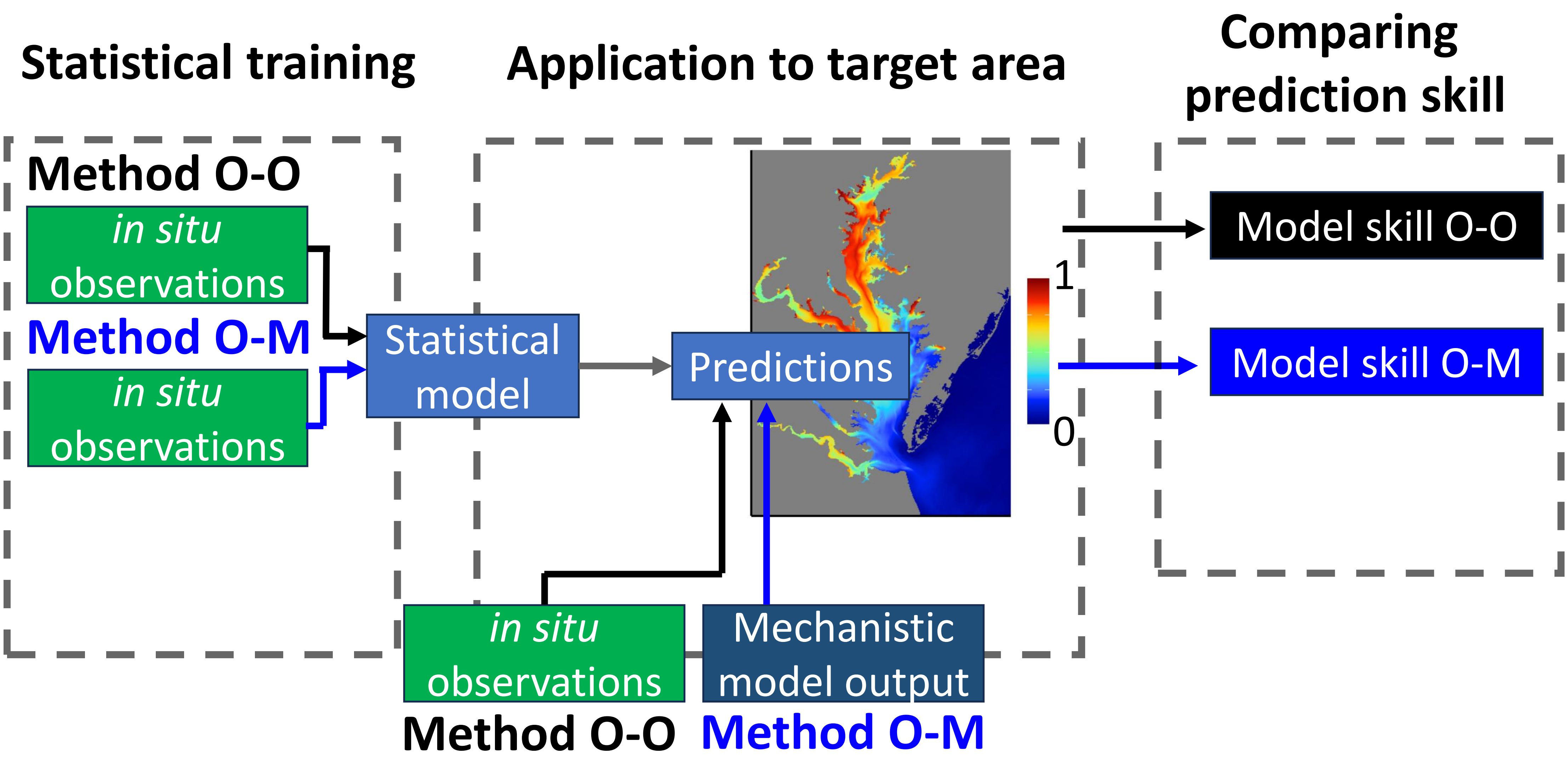
- We apply the statistical forecasting model using mechanistic model output (i.e., CBEFS forecasts)

- Should we also train the statistical model using mechanistic model output or can we train it using *in situ* observations?

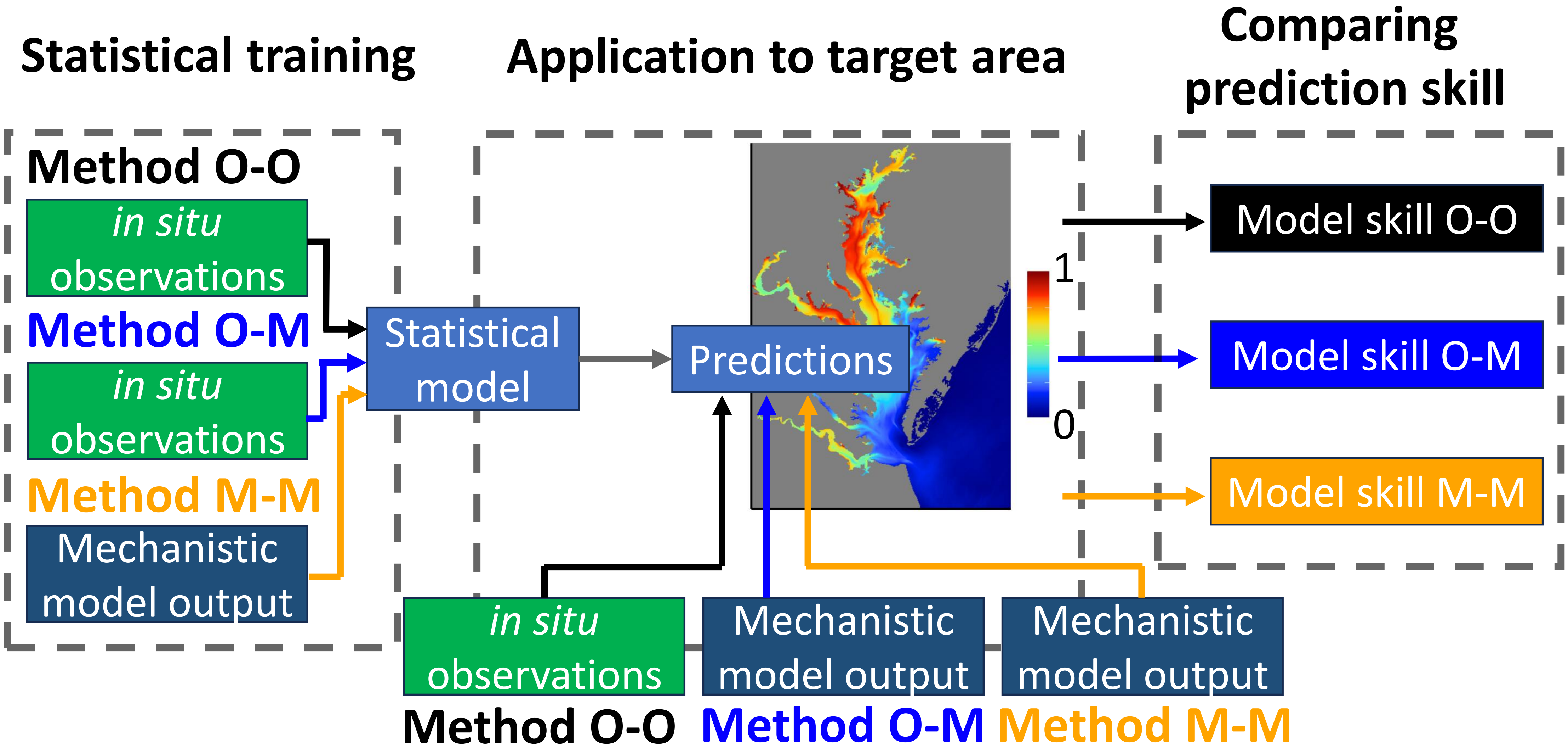
We compare three methods to train the statistical forecasting model



We compare three methods to train the statistical forecasting model



We compare three methods to train the statistical forecasting model



Methodology – environmental training information *in situ* observations

Chesapeake Bay Program
Science. Restoration. Partnership.

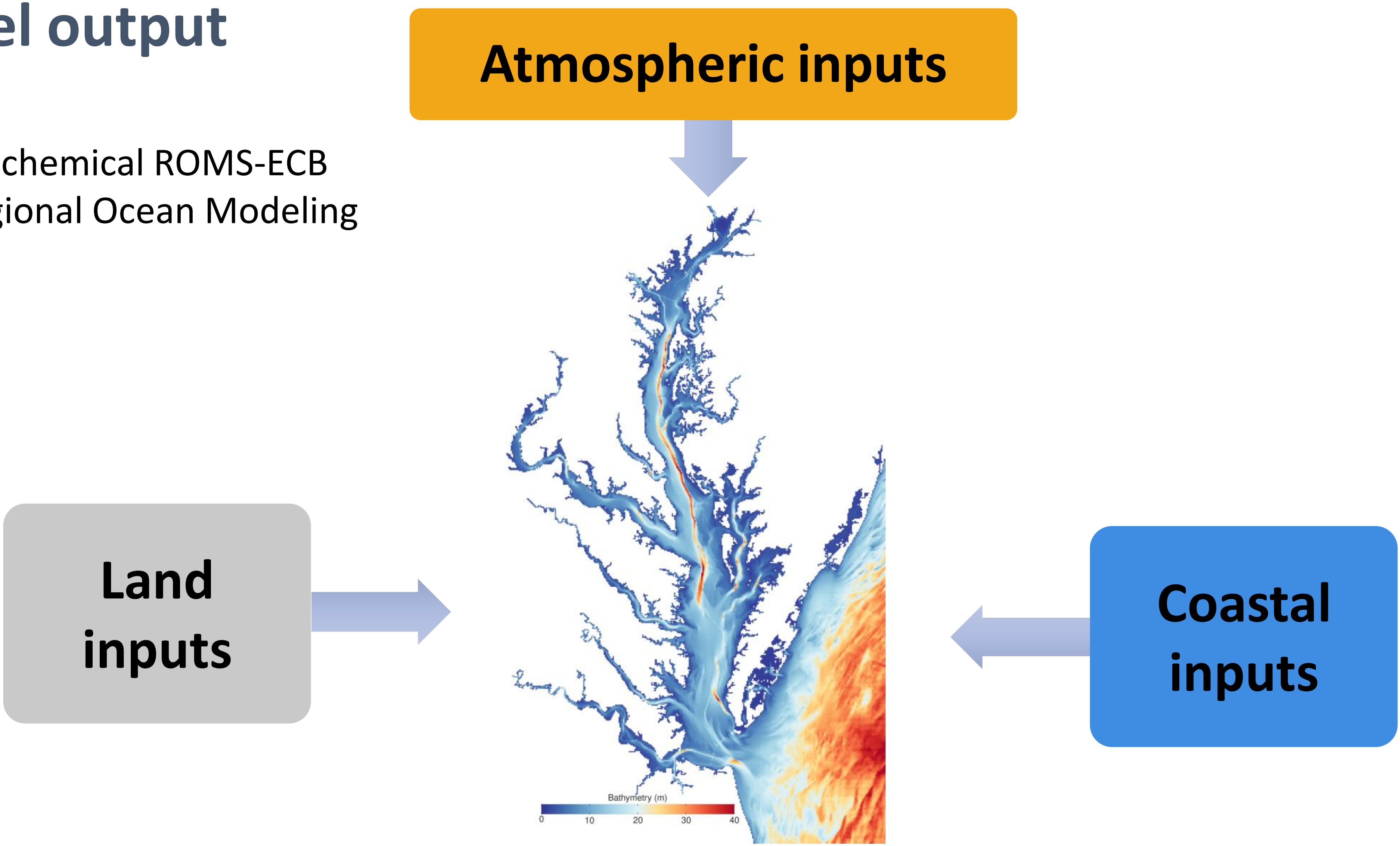
- Data provided by the Chesapeake Bay Program
- Use data from 1985-2020 (> 7,000 data points per taxon)
- At 42 stations covering both the main channel and tributaries

Environmental variable	Mean		Max		Standard deviation		Units
	in situ	model	*in situ*	model	*in situ*	model	
Water temperature (T)	17.2	17.5	31.0	30.5	7.9	7.4	°C
Salinity (S)	16.0	15.2	33.4	32.6	8.1	8.3	ppt
Vertical gradient of salinity (gradS)	0.38	0.40	1.99	1.83	0.33	0.30	ppt m^{-1}
Apparent oxygen utilization (AOU)	-0.81	-0.34	7.81	7.75	1.9	1.8	mg L^{-1}
pH	7.9	8.1	9.4	9.9	0.37	0.36	/
Dissolved inorganic nitrogen (DIN)	0.23	0.37	2.39	3.68	0.35	0.52	mg L^{-1}
Total organic nitrogen (TON)	0.46	0.40	1.70	1.02	0.20	0.14	mg L^{-1}
Solar irradiance at the water surface (swrad)†	188	188	251	251	55	55	W m^{-2}
Total water depth	16.3	16.3	31.0	31.0	6.7	6.7	m

† Derived from the ERA5 reanalysis Hersbach et al. (2020).

Methodology – environmental training information mechanistic model output

- 3D hydrodynamic – biogeochemical ROMS-ECB
- Implementation of the Regional Ocean Modeling System (ROMS)
- 600 m x 600 m
- 20 vertical levels



Bever et al., Env Mod & Software, 2021
St-Laurent et al., BG, 2020

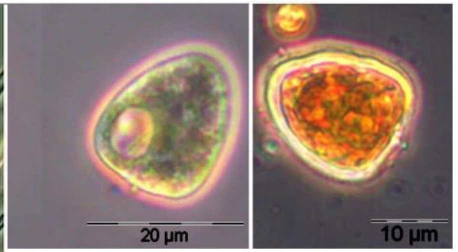
Methodology – *in situ* algal cell count data

- We focus on seven (mostly harmful) algal taxa
- We translate observed cell counts to binary bloom data using fixed cell count thresholds

Pseudo-nitzschia pungens

(photo by Regina Hansen)

Prorocentrum minimum



(photo by Susanne Busch)

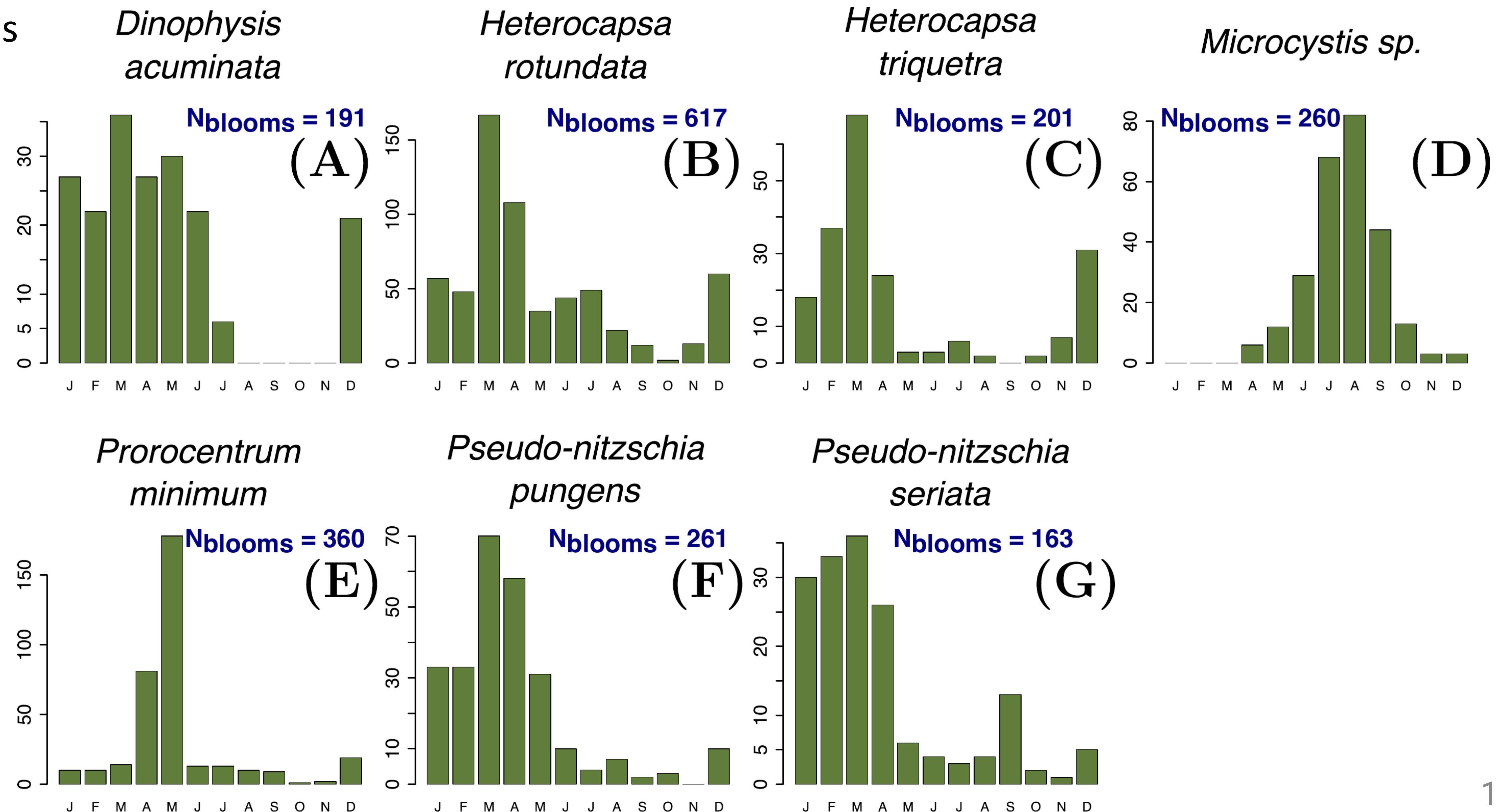
Dinophysis acuminata

(photo by Mats Kuylenstierna)

Taxon name	Number of blooms	Bloom threshold value	References
Dinophysis acuminata	191	0.4 cells mL^{-1}	Díaz et al. (2016)
Heterocapsa rotundata	617	1,000 cells mL^{-1}	Marshall and Egerton (2009) and Mulholland et al. (2018)
Heterocapsa triquetra (or *steinii*)	201	200 cells mL^{-1}	Baek et al. (2011) and Marshall and Egerton (2009)
Microcystis sp.	260	10,000 cells mL^{-1}	Marshall and Egerton (2009) and Ho et al. (2015)
Prorocentrum minimum (or *P.cordatum*)	360	1,000 cells mL^{-1}	Marshall and Egerton (2009), Pease et al. (2021), and Mulholland et al. (2018)
Pseudo-nitzschia pungens	261	500 cells mL^{-1}	Anderson et al. (2010)
Pseudo-nitzschia seriata	163	250 cells mL^{-1}	Anderson et al. (2010)

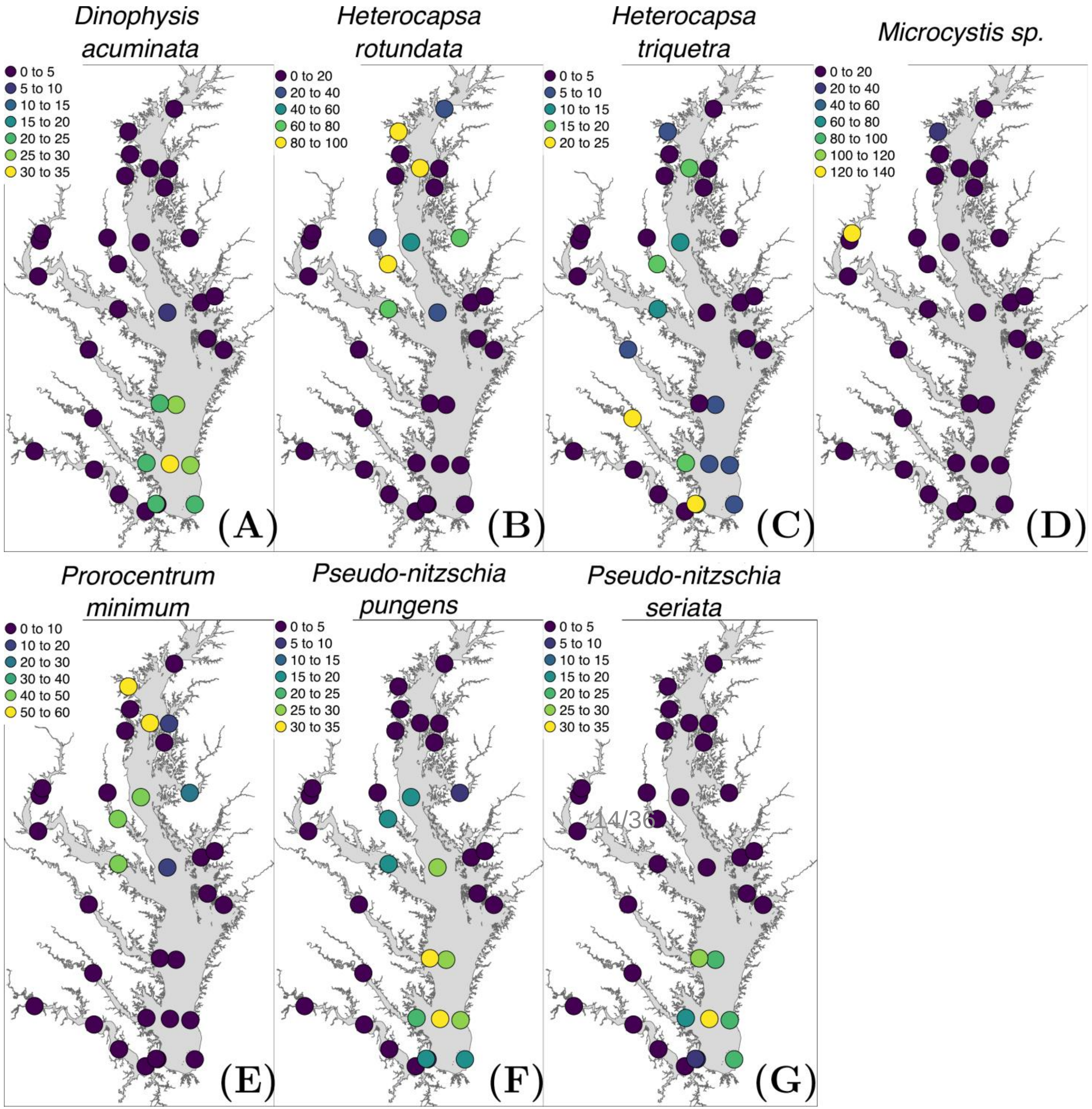
Methodology – *in situ* algal cell count data

Seven taxa exhibit a variety of habitat preferences, blooming in various seasons and regions



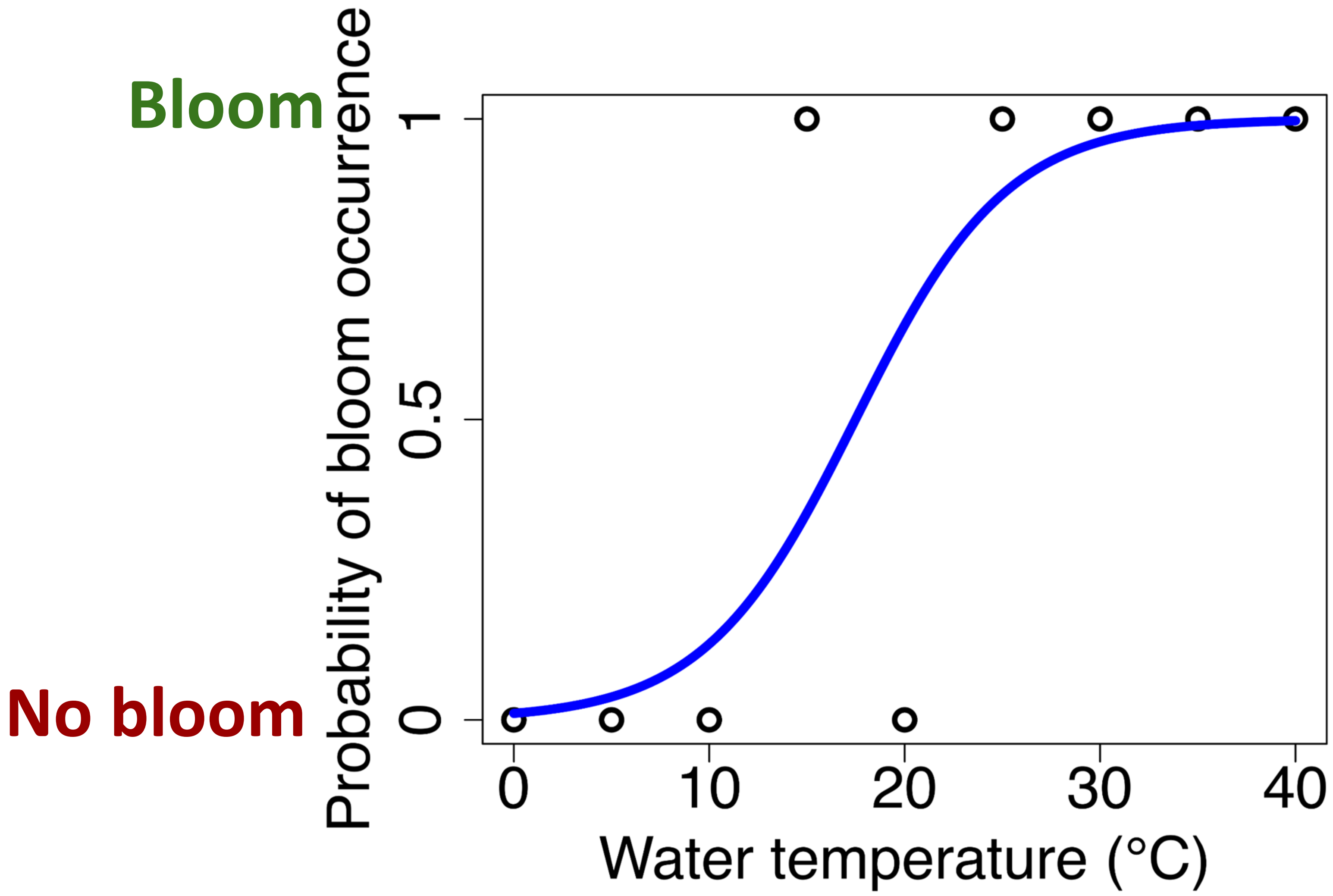
Methodology – *in situ* algal cell count data

Seven taxa exhibit a variety of habitat preferences, blooming in various seasons and regions

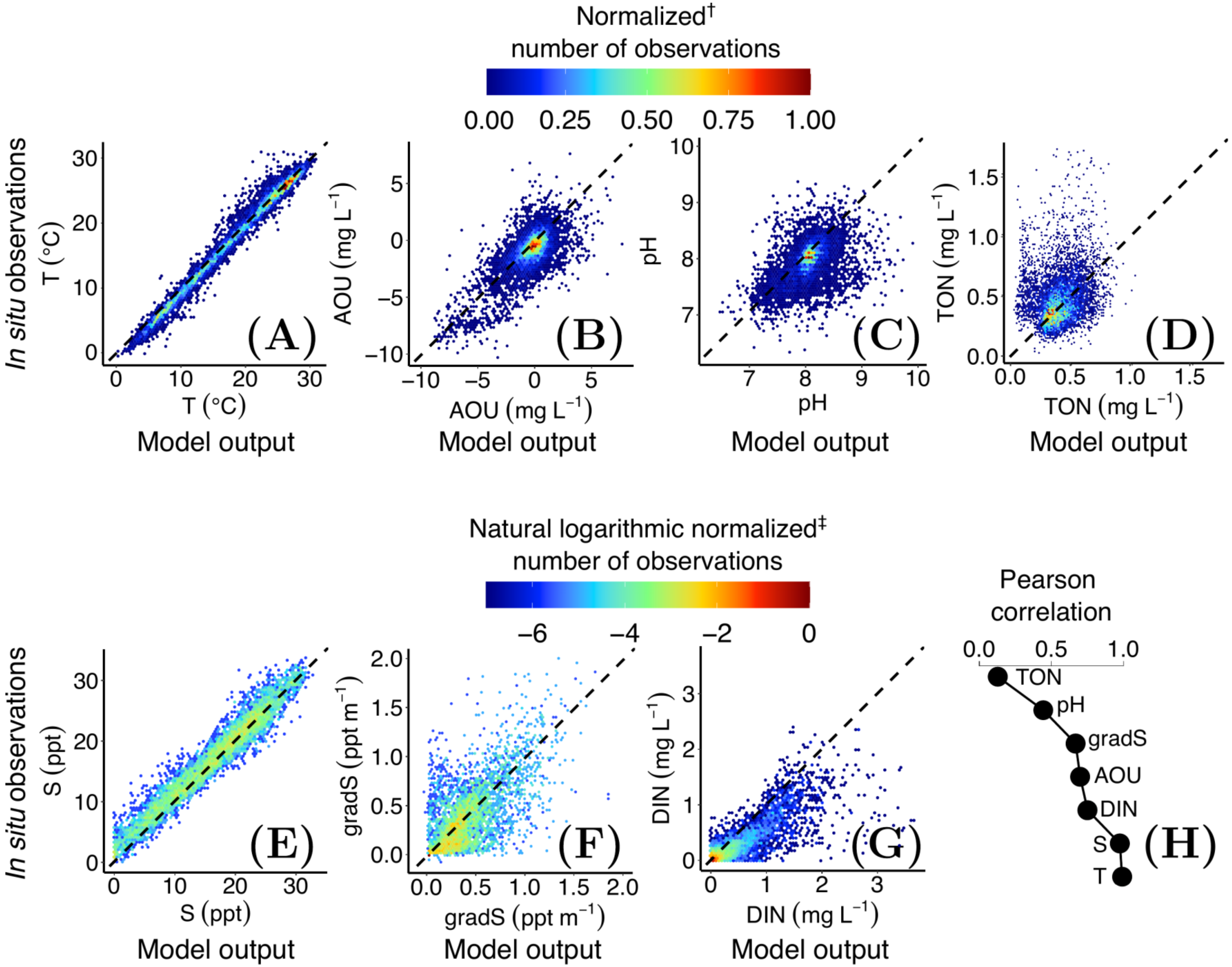


Statistical model: generalized linear model

- Similar to linear regression, the technique allows us to fit a curve to the data.
- Using this fitted curve, we can easily compute the probability of a bloom, given a set of environmental conditions.

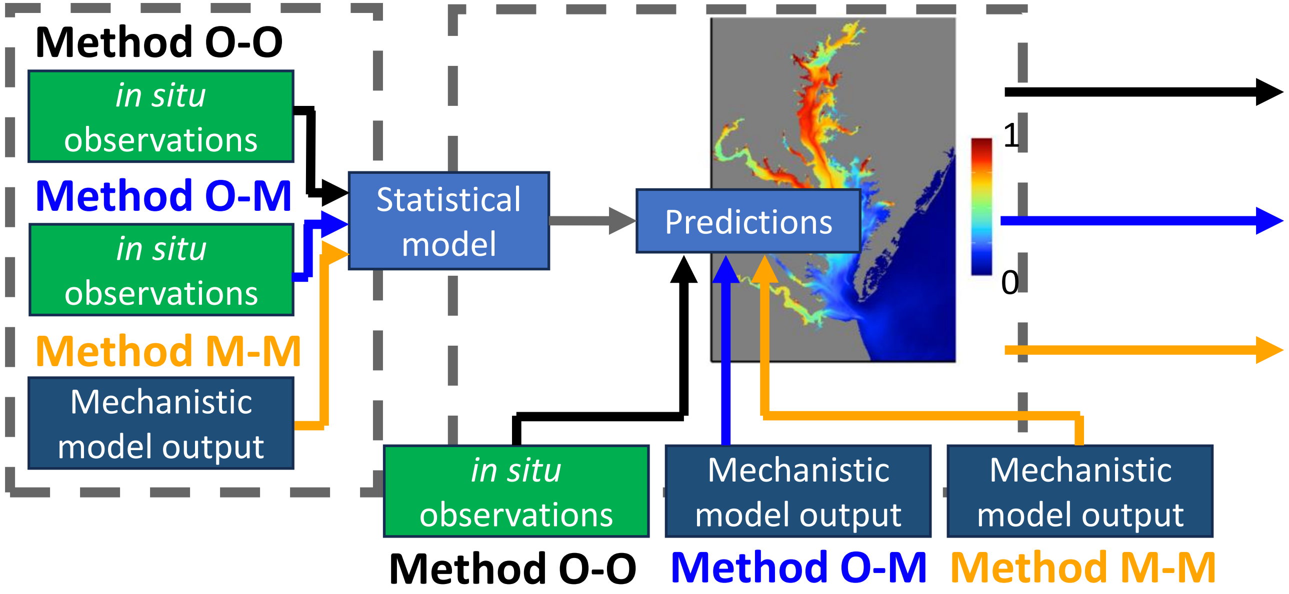


Comparing *in situ* observations and mechanistic model output

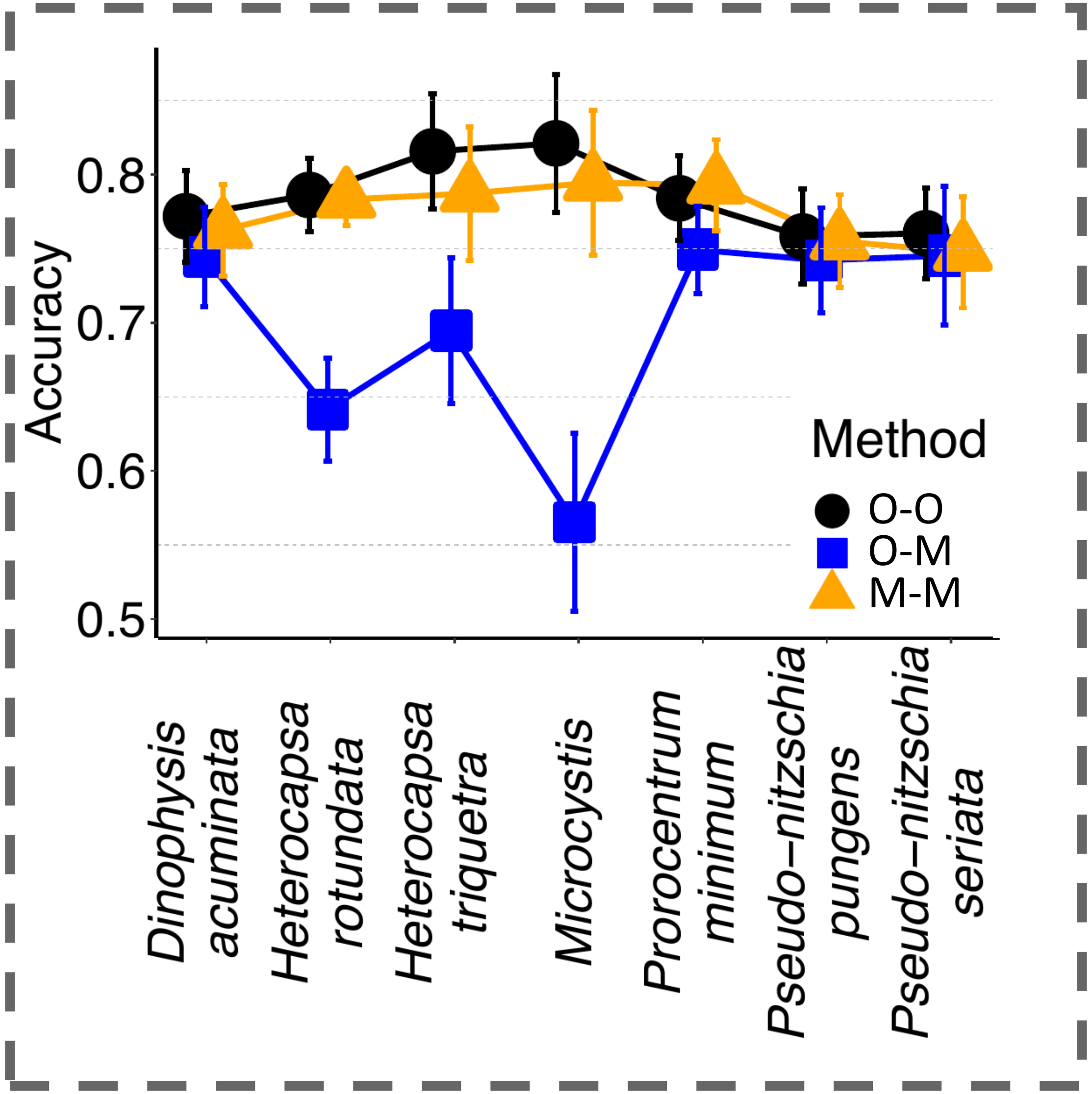


Both training and applying using mechanistic model output enhances model prediction skill

Statistical training **Application to target area**



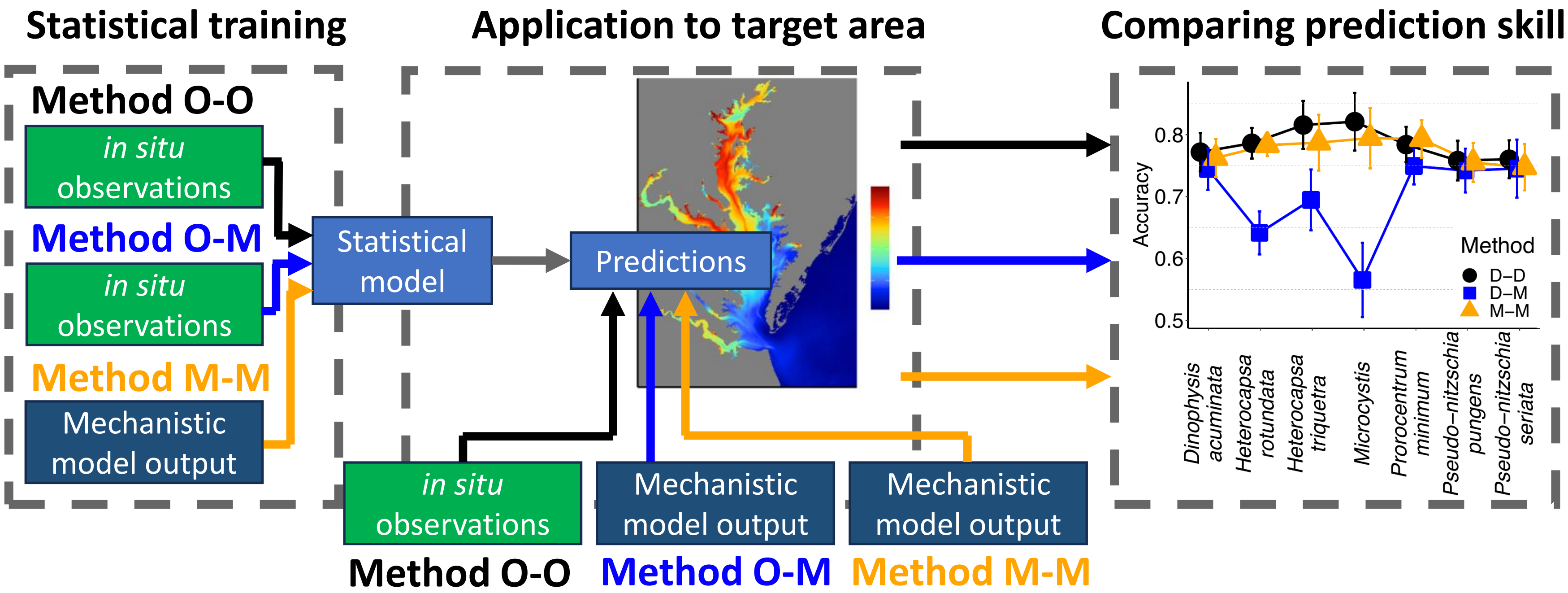
Comparing prediction skill



Take home messages

i) Statistical models trained using *in situ* observations are less accurate when applied to model output (Method O-M) than when applied to *in situ* observations (Method O-O)

ii) We can enhance the model prediction skill corresponding to Method O-M by both training and applying the statistical model using mechanistic model output (Method M-M)

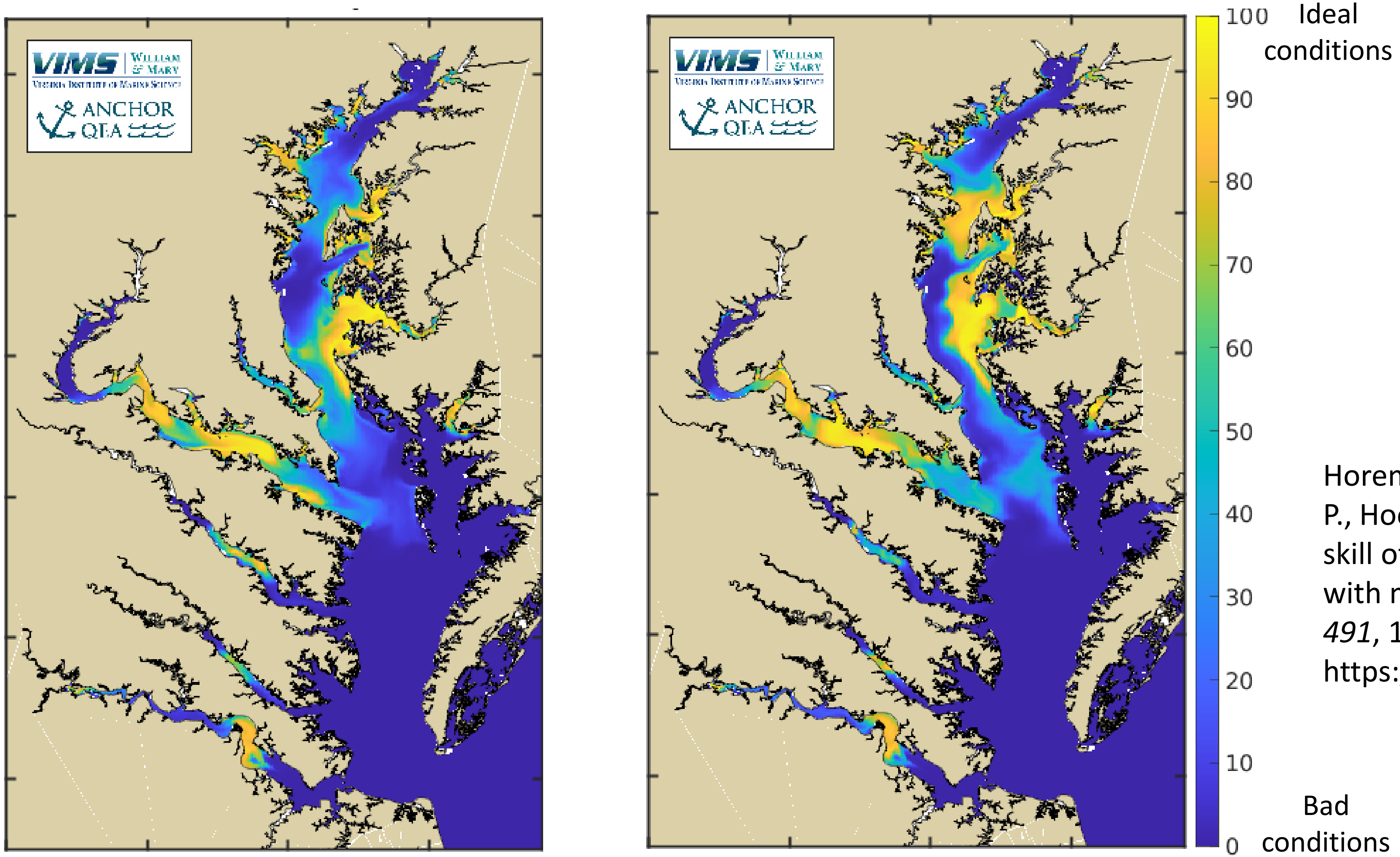


Implications of our results

Our results are used to extend CBEFS with forecasts of harmful algal blooms

Habitat suitability for *Prorocentrum minimum* blooms – Nowcast: April 14, 2023

Habitat suitability for *Prorocentrum minimum* blooms – Forecast: April 16, 2023



www.vims.edu/cbefs

Horemans, D. M. L., Friedrichs, M. A. M., St-Laurent, P., Hood, R. R., & Brown, C. W. (2024). Evaluating the skill of correlative species distribution models trained with mechanistic model output. *Ecological Modelling*, *491*, 110692. https://doi.org/10.1016/J.ECOLMODEL.2024.110692

Extra slides

Extra slides: model prediction skill