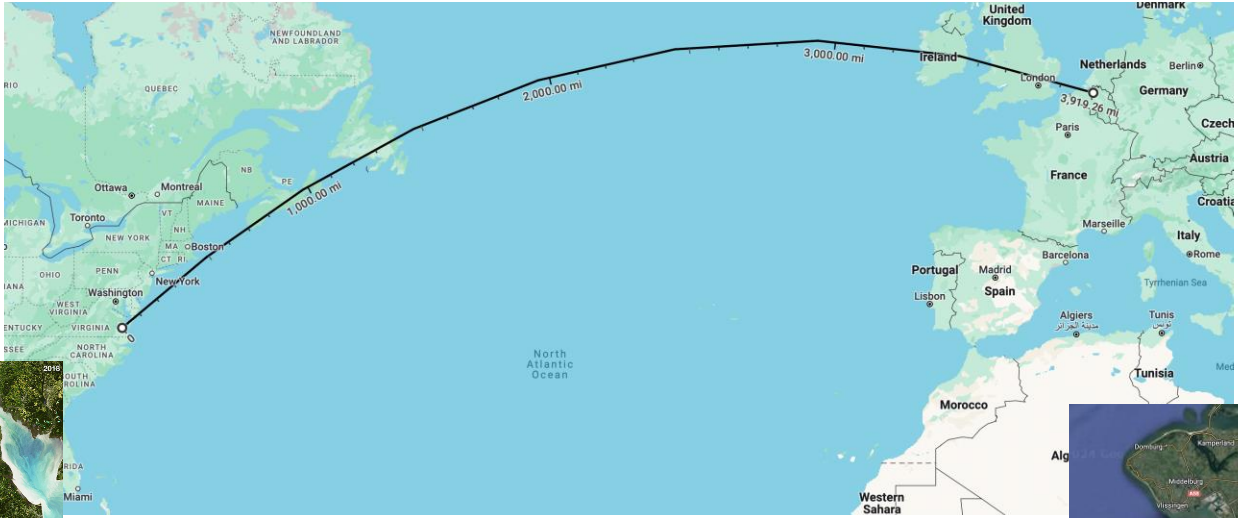
Real-time forecasts of Chesapeake Bay harmful algal blooms using empirical habitat suitability models

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

3 April 2024,
Gloucester, VA

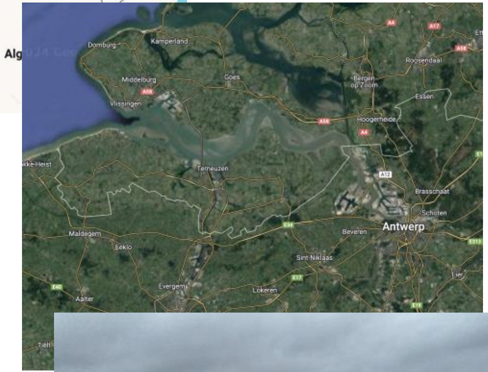
VIMS | William & Mary
Virginia Institute of Marine Science
University of Maryland
Center for Environmental Science

Biography



- Chesapeake Bay
- Statistical modeling
- Forecasting harmful algal blooms

- Scheldt estuary
- Mechanistic and idealized modeling
- Sediment-phytoplankton interactions



Estuary	SPM (kg m^{-3})	Bathymetry and geometry	Discharge (m^2 s^{-1})	Tidal range (m)
Chesapeake Bay	$\mathcal{O}(0.01)$	deep, and non-convergent	$\mathcal{O}(2000)$	$\mathcal{O}(0.1)$
Scheldt	$\mathcal{O}(0.1)$	shallow and funnel-shaped	$\mathcal{O}(100)$	$\mathcal{O}(1)$

Harmful algal blooms

Harmful algal blooms (HABs) are algal blooms that negatively impact other organisms in natural waters by excreting toxins or severely lowering the oxygen levels.

ODU Professors Receive $3M NOAA Grant to Track Harmful Algal Blooms

Virginia Waters Troubled by Multiple Harmful Algal Blooms

WYDAILY
WILLIAMSBURG YORKTOWN DAILY

By **Charlie Paullin, Virginia Mercury** September 4, 2022 0

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

Harmful algal blooms

Harmful algal blooms (HABs) are algal blooms that negatively impact other organisms in natural waters by excreting toxins or severely lowering the oxygen levels.

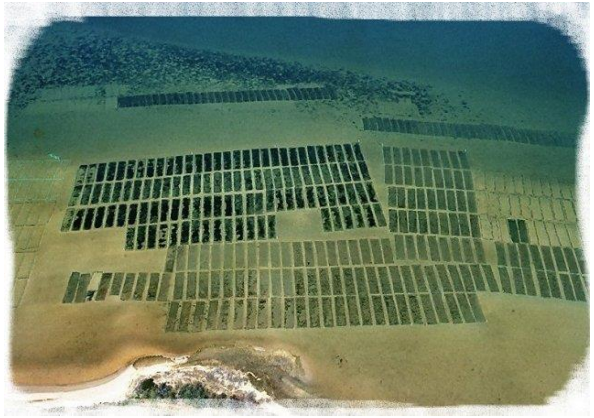
Fish mortality

U.S. National Office for Harmful Algal Blooms

Beach closures

NJ Advance Media, 06/27/2019

Economic impact (e.g., aquaculture)

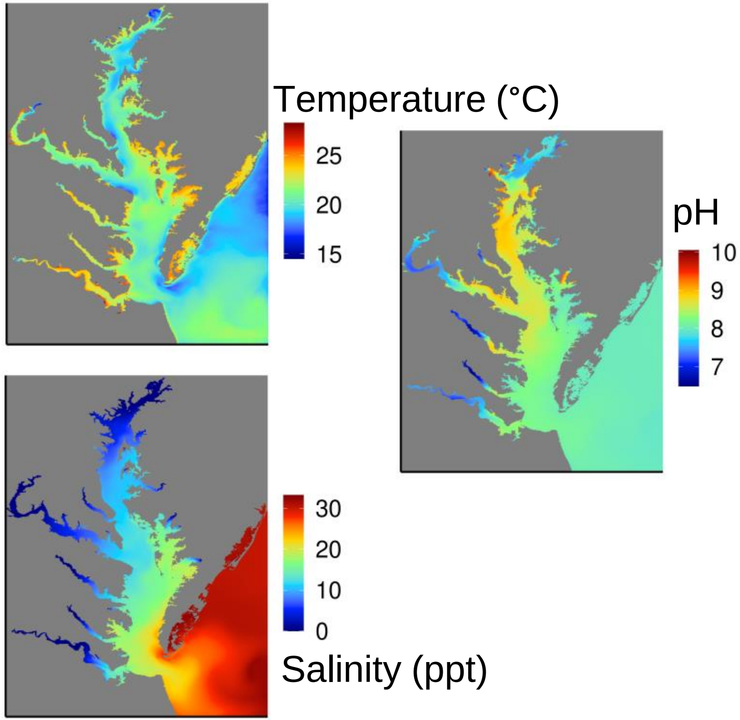


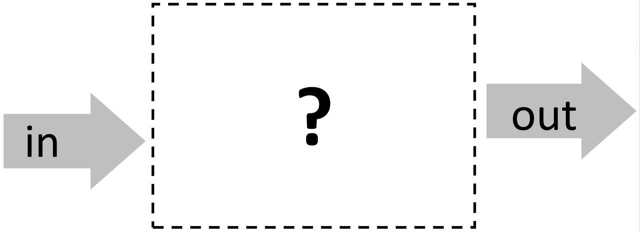
Virginia Marine Resources Commission

It is therefore important to forecast HABs and get insight into the mechanisms driving these blooms

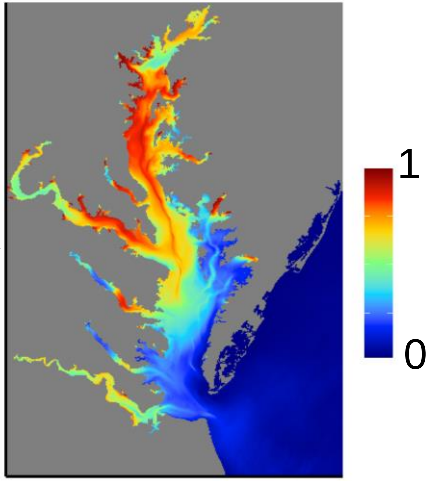
Objective – forecast harmful algal blooms (HABs), add HABs to our forecasting tool (CBEFS), and organize workshops with stakeholders

Existing model forecasts using a mechanistic model (ChesROMS-ECB)





Forecast probability of HABs



www.vims.edu/cbefs

Chesapeake Bay Environmental Forecast System (CBEFS)

ChesROMS-ECB

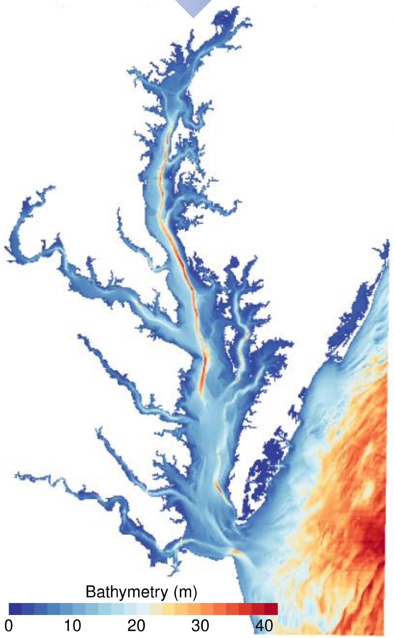
Estuarine model framework

- ~ 600 m x 600 m
- 20 vertical levels
- Hydrodynamics, tides, etc.
- BGC cycles: C, N, etc.

Atmospheric inputs

NOAA atm. forcing

- Winds
- Solar radiation
- Temperature
- Precipitation



Land inputs

Terrestrial inputs from watershed models and USGS data

Coastal inputs

Observed water levels; climatologies of NOAA data

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

Marjy Friedrichs
VIMS

Aaron Bever
Anchor QEA

Pierre St-Laurent
VIMS

Raleigh Hood
UMCES

Chesapeake Bay Environmental Forecast System (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

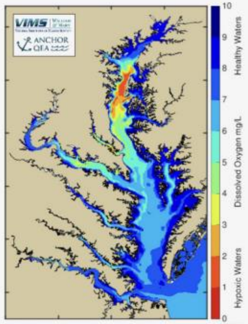
Chesapeake Bay Environmental Forecast System

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.

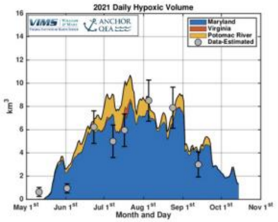
Our Chesapeake Bay Environmental Forecast System simulates 3 conditions for each selected variable:

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)

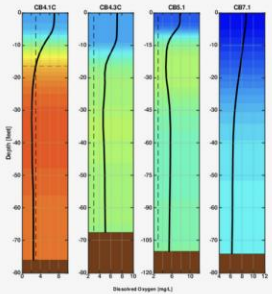
Click a selection below to access the specified simulation. Please see the **contact information page** for data requests and general contact information.



DISSOLVED OXYGEN (DO)



DEAD ZONE SIZE

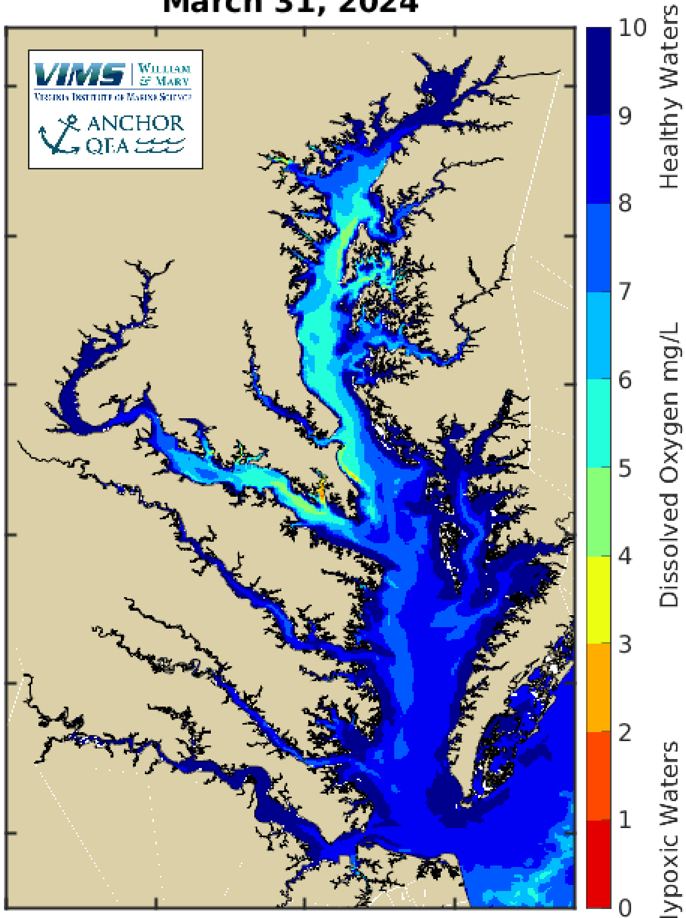


DEPTH TO LOW OXYGEN

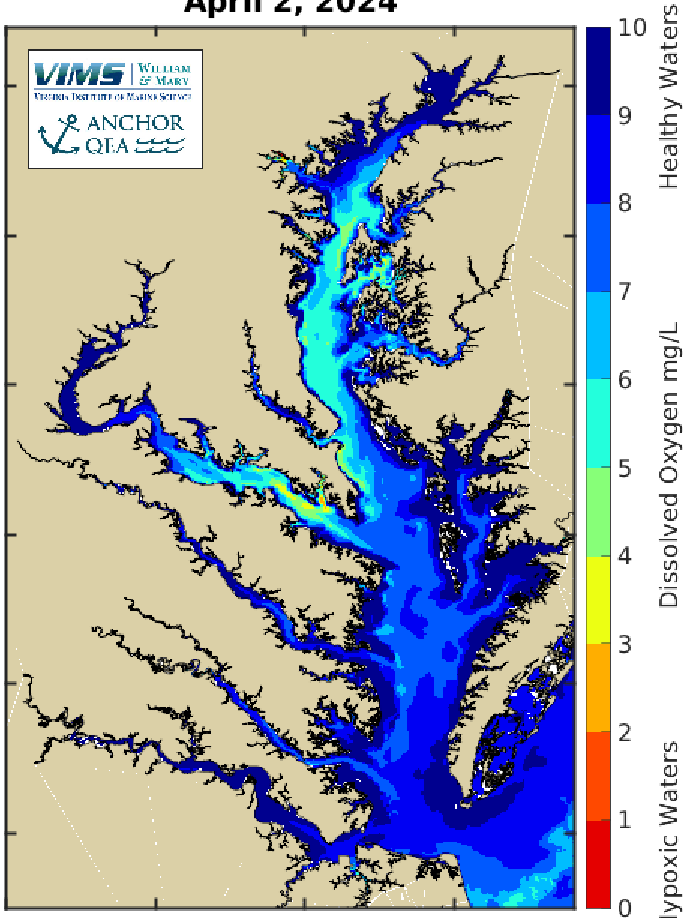
www.vims.edu/cbefs

Chesapeake Bay Environmental Forecast System (CBEFS)

Bottom Oxygen: Nowcast
March 31, 2024



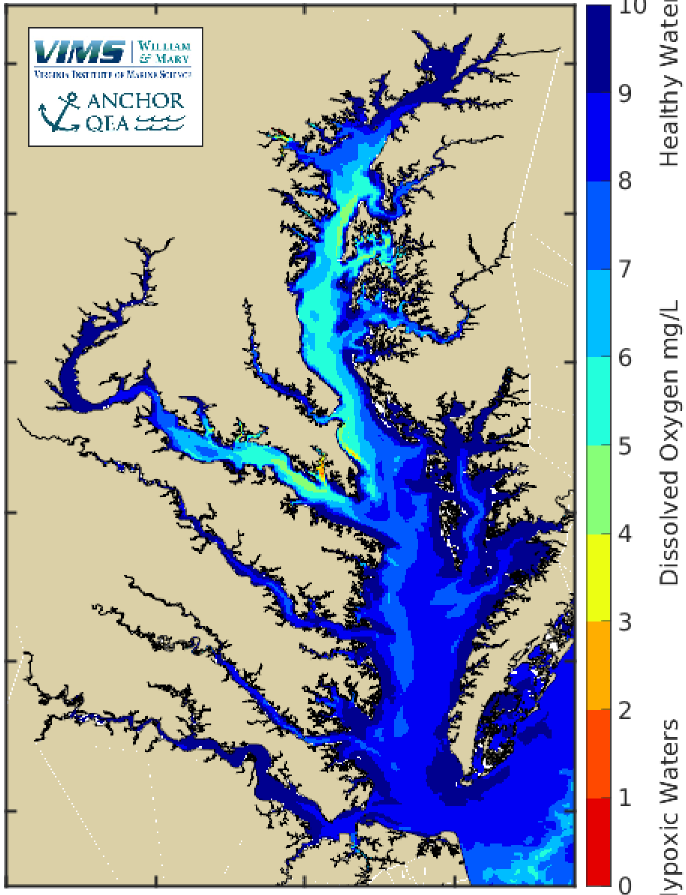
Bottom Oxygen: Forecast
April 2, 2024



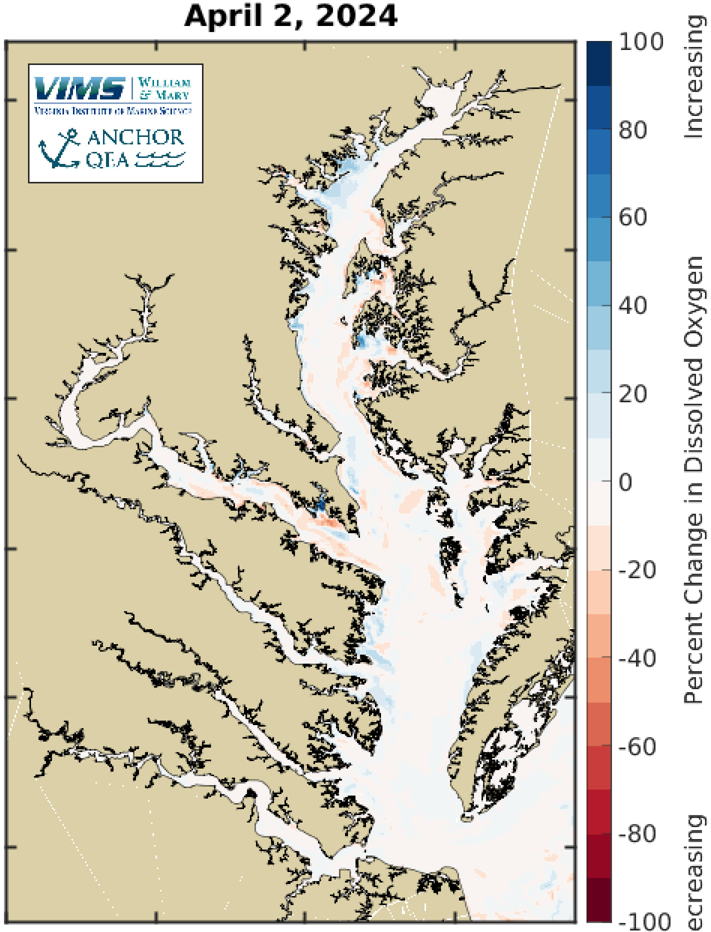
www.vims.edu/cbefs

Chesapeake Bay Environmental Forecast System (CBEFS)

Bottom Oxygen: Nowcast
March 31, 2024



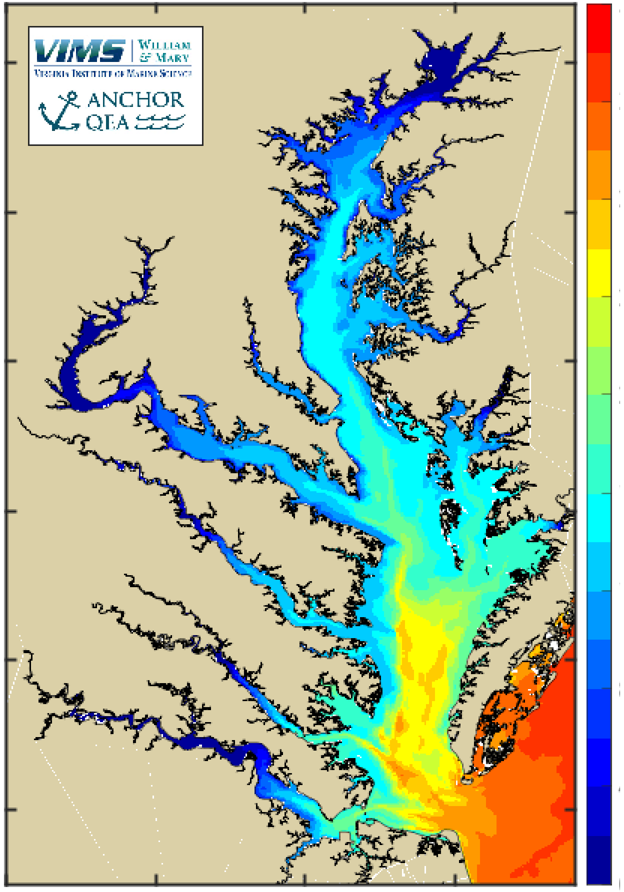
Bottom Oxygen: Forecast Trend
April 2, 2024



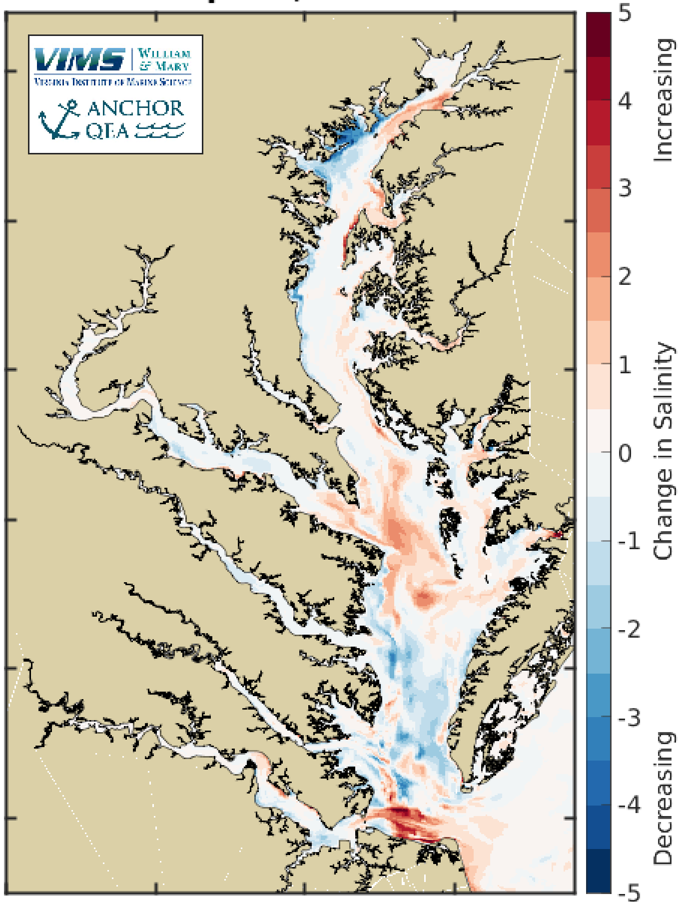
www.vims.edu/cbefs

Chesapeake Bay Environmental Forecast System (CBEFS)

Bottom Salinity: Nowcast
March 31, 2024



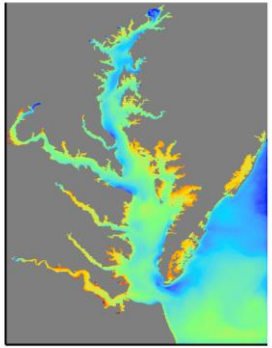
Bottom Salinity: Forecast Trend
April 2, 2024



www.vims.edu/cbefs

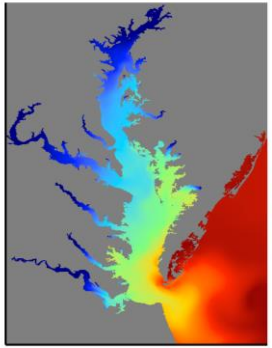
Main objective – forecast harmful algal blooms (HABs)

Existing model forecasts



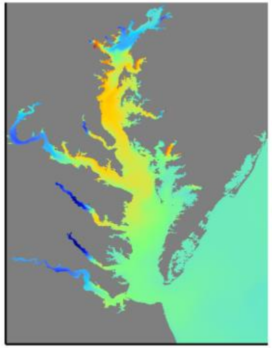
Temperature (°C)

25
20
15



30
20
10
0

Salinity (ppt)



pH

10
9
8
7

www.vims.edu/cbefs

Chesapeake Bay Program
Science. Restoration. Partnership.

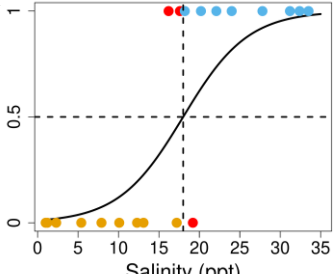
training information

in

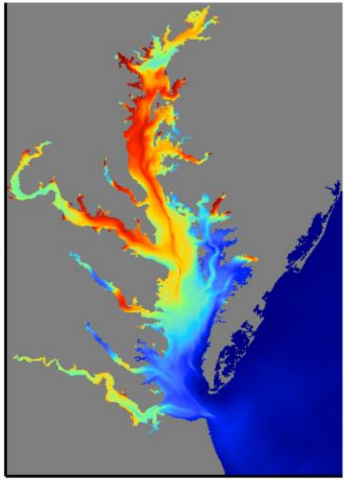
Empirical habitat model

out

Statistical model type



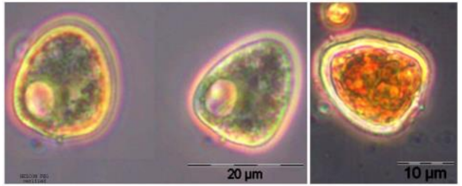
Forecast probability of HABs



1
0

Challenges – forecast harmful algal blooms (HABs)

1. How does model prediction skill change in time and space depending on **model type/complexity**?



(photo by Susanne Busch)


ORIGINAL RESEARCH article

Front. Mar. Sci., 21 March 2023
Sec. Coastal Ocean Processes
Volume 10 - 2023 | https://doi.org/10.3389/fmars.2023.1127649

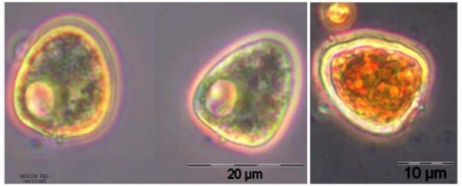
This article is part of the Research Topic
Eutrophication, Algal Bloom, Hypoxia and Ocean Acidification in Large River Estuaries, Volume II
View all 8 Articles >

Forecasting *Prorocentrum minimum* blooms in the Chesapeake Bay using empirical habitat models

Dante M. L. Horemans[1*] Marjorie A. M. Friedrichs[1] Pierre St-Laurent[1] Raleigh R. Hood[2] Christopher W. Brown[3]

Challenges – forecast harmful algal blooms (HABs)

1. How does model prediction skill change in time and space depending on **model type/complexity**?



(photo by Susanne Busch)

ORIGINAL RESEARCH article

Front. Mar. Sci., 21 March 2023
Sec. Coastal Ocean Processes
Volume 10 - 2023 | https://doi.org/10.3389/fmars.2023.1127649

This article is part of the Research Topic
Eutrophication, Algal Bloom, Hypoxia and Ocean Acidification in Large River Estuaries, Volume II
View all 8 Articles

Forecasting *Prorocentrum minimum* blooms in the Chesapeake Bay using empirical habitat models

Dante M. L. Horemans[1*] Marjorie A. M. Friedrichs[1] Pierre St-Laurent[1] Raleigh R. Hood[2]
Christopher W. Brown[3]

2. How does model prediction skill change depending on **training information**?

Should we use *in situ* observations to train our statistical model?

ELSEVIER
Ecological Modelling
Volume 491, May 2024, 110692

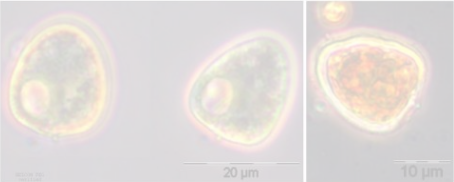
Review

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

Dante M.L. Horemans [a], Marjorie A.M. Friedrichs [a], Pierre St-Laurent [a],
Raleigh R. Hood [b], Christopher W. Brown [c d]

Challenges – forecast harmful algal blooms (HABs)

1. How does model prediction skill change in time and space depending on **model type/complexity**?



(photo by Susanne Busch)


ORIGINAL RESEARCH article

Front. Mar. Sci., 21 March 2023
Sec. Coastal Ocean Processes
Volume 10 - 2023 | https://doi.org/10.3389/fmars.2023.1127649

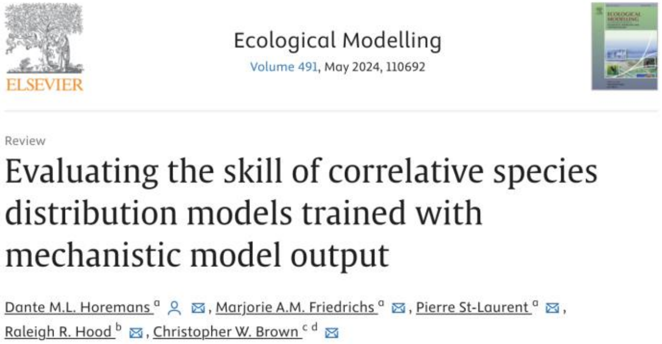
This article is part of the Research Topic
Eutrophication, Algal Bloom, Hypoxia and Ocean Acidification in Large River Estuaries, Volume II
View all 8 Articles >

Forecasting *Prorocentrum minimum* blooms in the Chesapeake Bay using empirical habitat models

Dante M. L. Horemans[1*] Marjorie A. M. Friedrichs[1] Pierre St-Laurent[1] Raleigh R. Hood[2] Christopher W. Brown[3]

2. How does model prediction skill change depending on **training information**?

Should we use *in situ* observations to train our statistical model?


ELSEVIER

Ecological Modelling
Volume 491, May 2024, 110692

Review

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

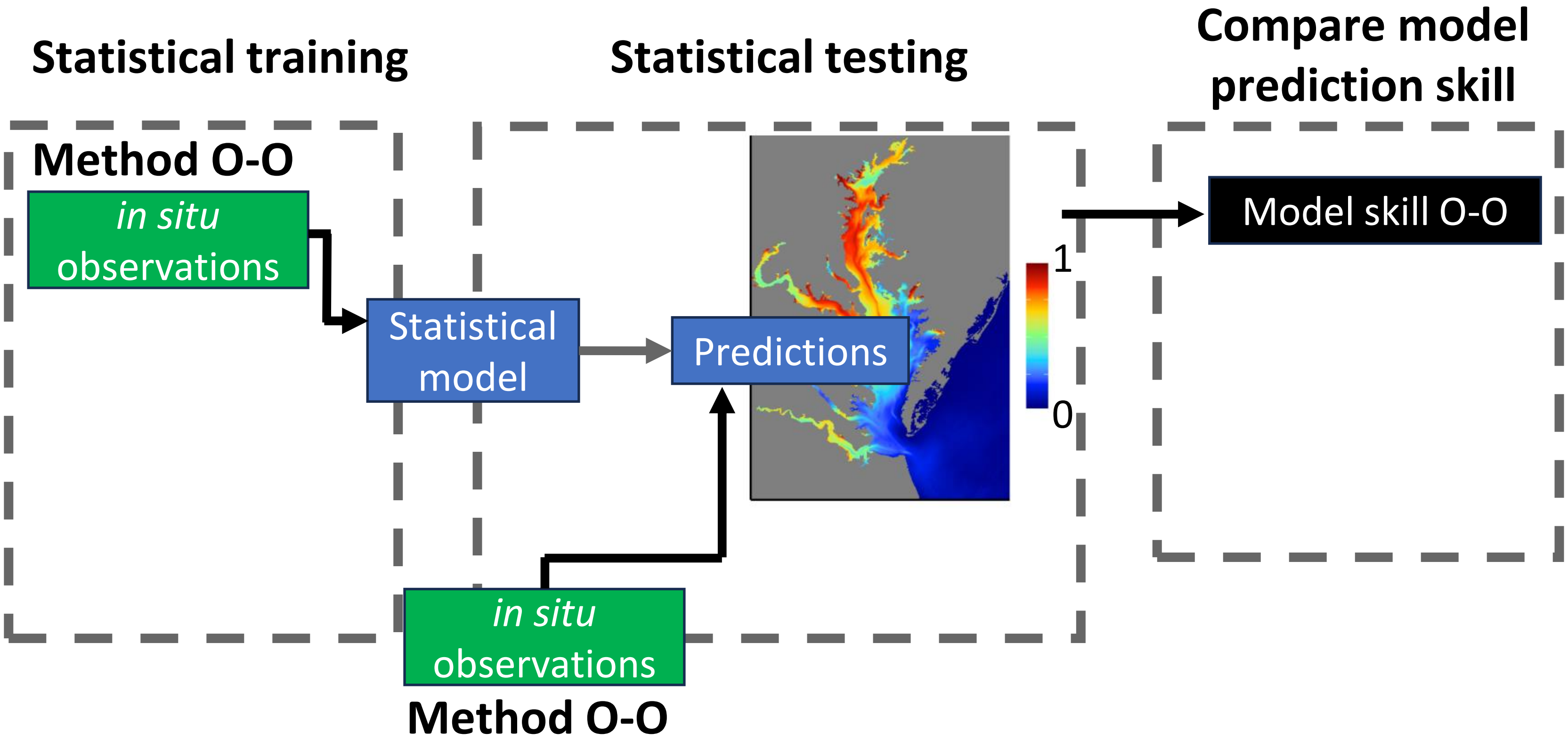
Dante M.L. Horemans[a], Marjorie A.M. Friedrichs[a], Pierre St-Laurent[a], Raleigh R. Hood[b], Christopher W. Brown[c d]

Research question – should we use *in situ* observation to train our statistical model?

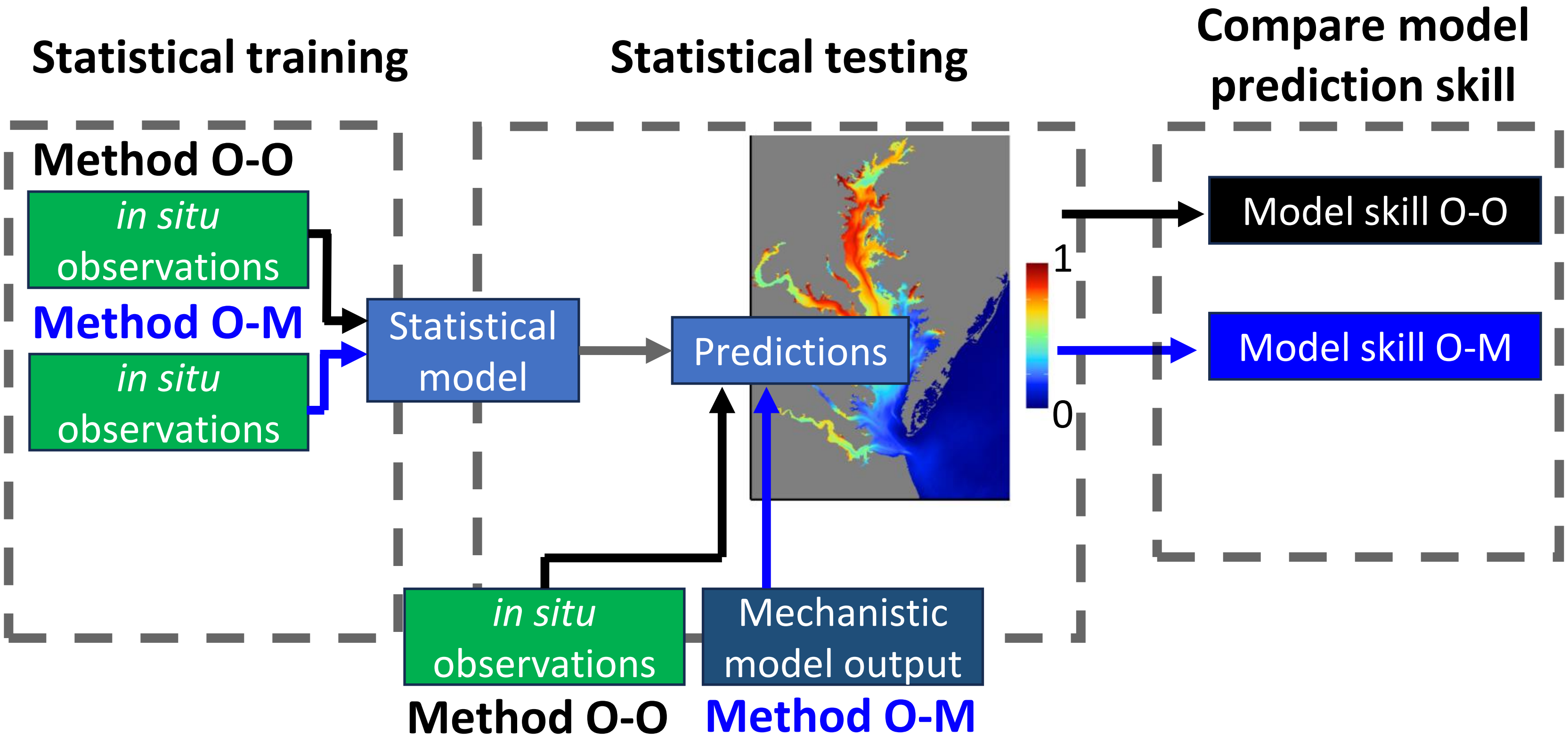
- We apply the statistical forecasting model using mechanistic model output (i.e., CBEFS forecasts)

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

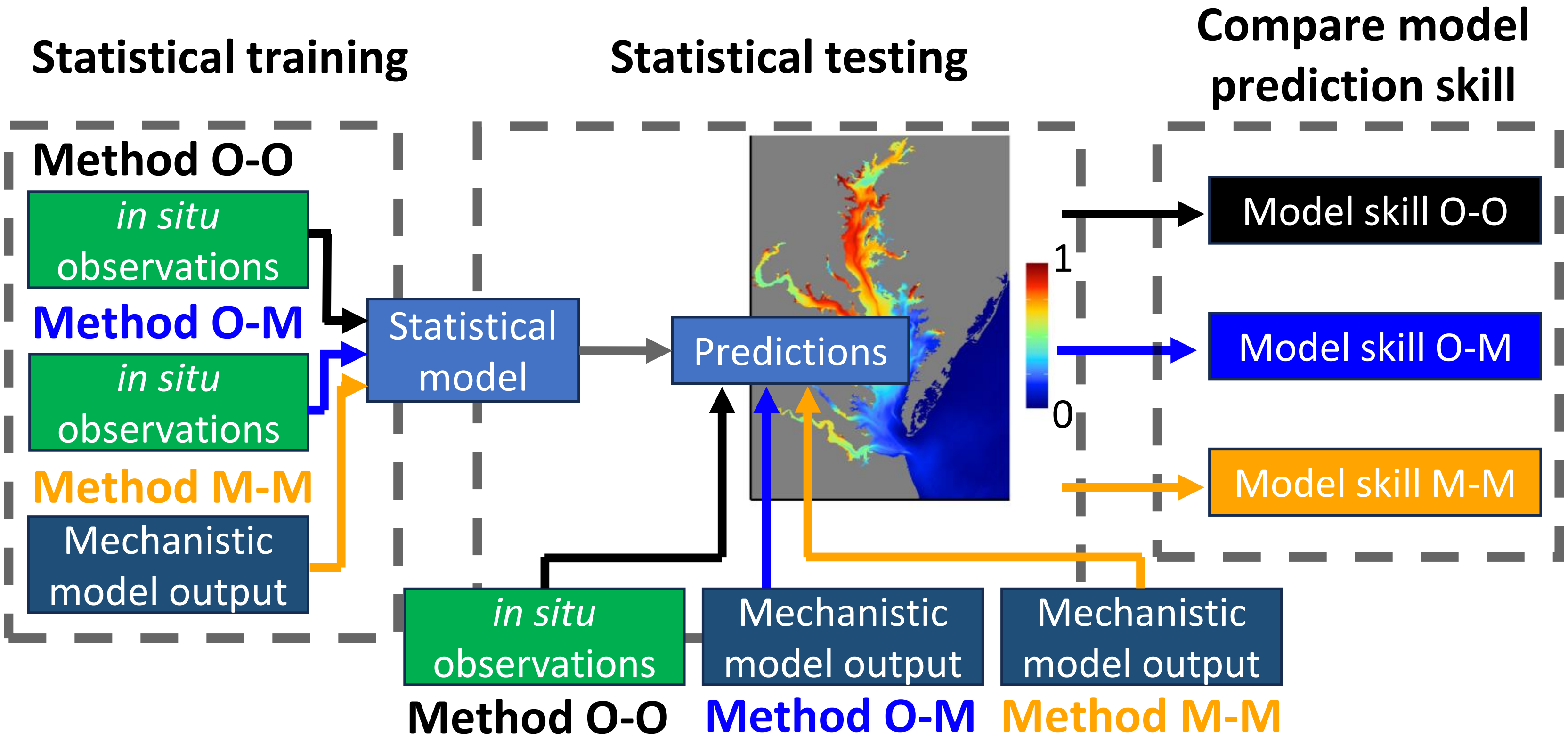
We compare three methods to train the statistical forecasting model



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

- 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

Chesapeake Bay Program
Science. Restoration. Partnership.

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

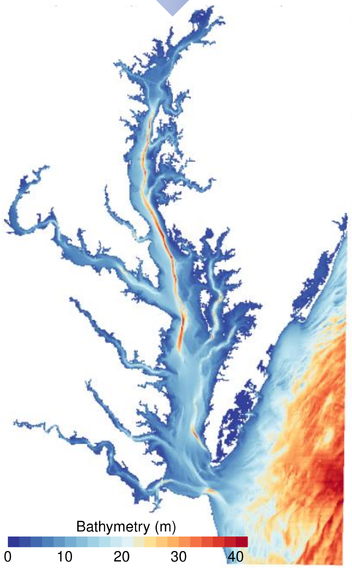
Chesapeake Bay Environmental Forecast System (CBEFS)

ChesROMS-ECB

Estuarine model framework

- ~ 600 m x 600 m
- 20 vertical levels
- Hydrodynamics, tides, etc.
- BGC cycles: C, N, etc.

Atmospheric inputs



NOAA atm. forcing

- Winds
- Solar radiation
- Temperature
- Precipitation

Land inputs

Terrestrial inputs from watershed models and USGS data

Coastal inputs

Observed water levels; climatologies of NOAA data

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

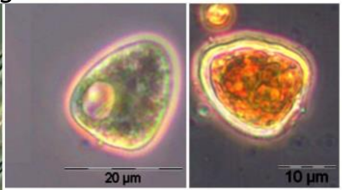
Methodology – *in situ* algal cell count data

- 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

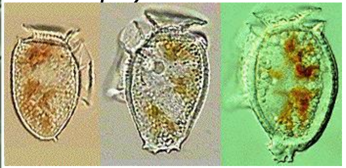
(photo by Regina Hansen)

Prorocentrum minimum



(photo by Susanne Busch)

Dinophysis acuminata

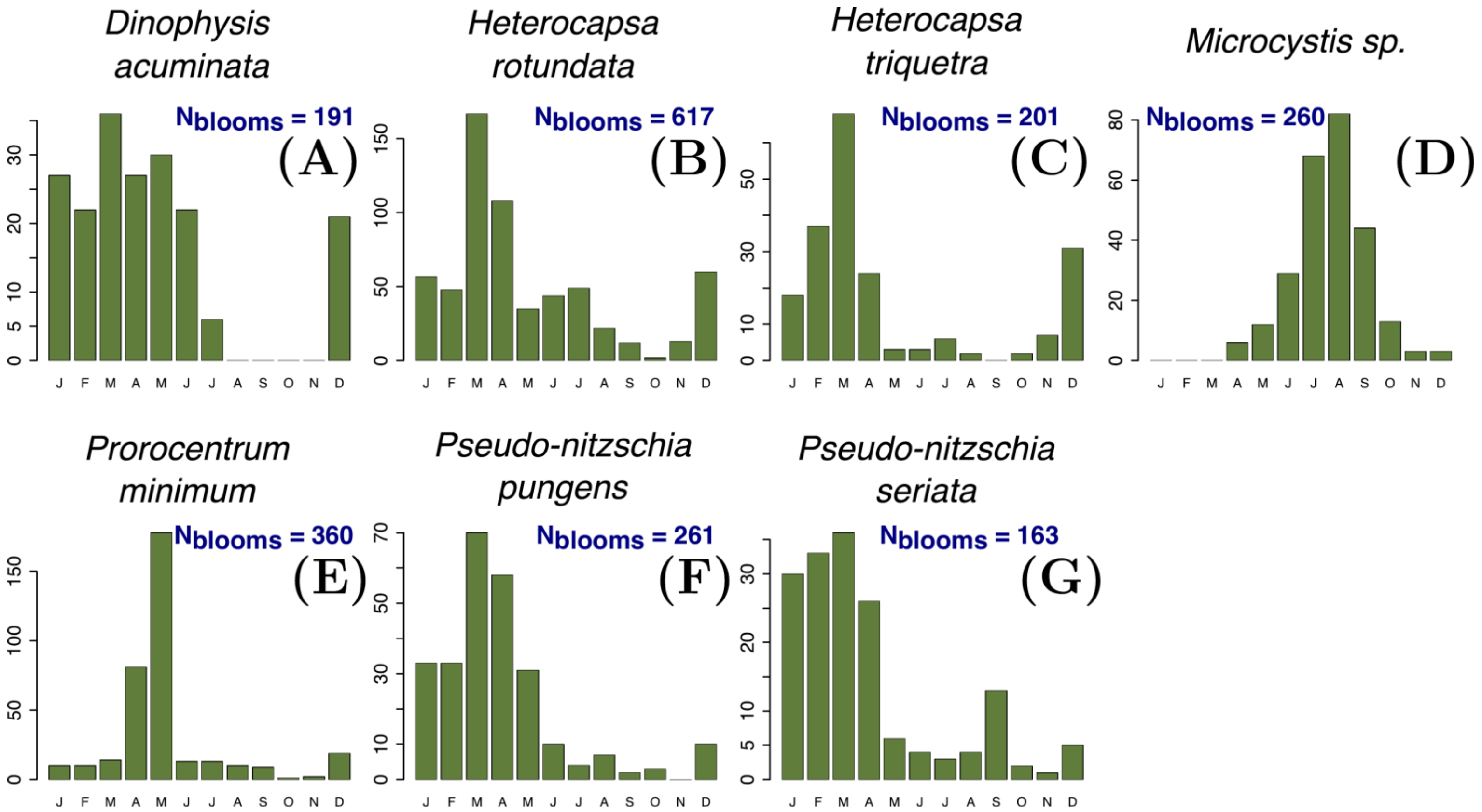


(photo by Mats Kuylenstierna)

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)

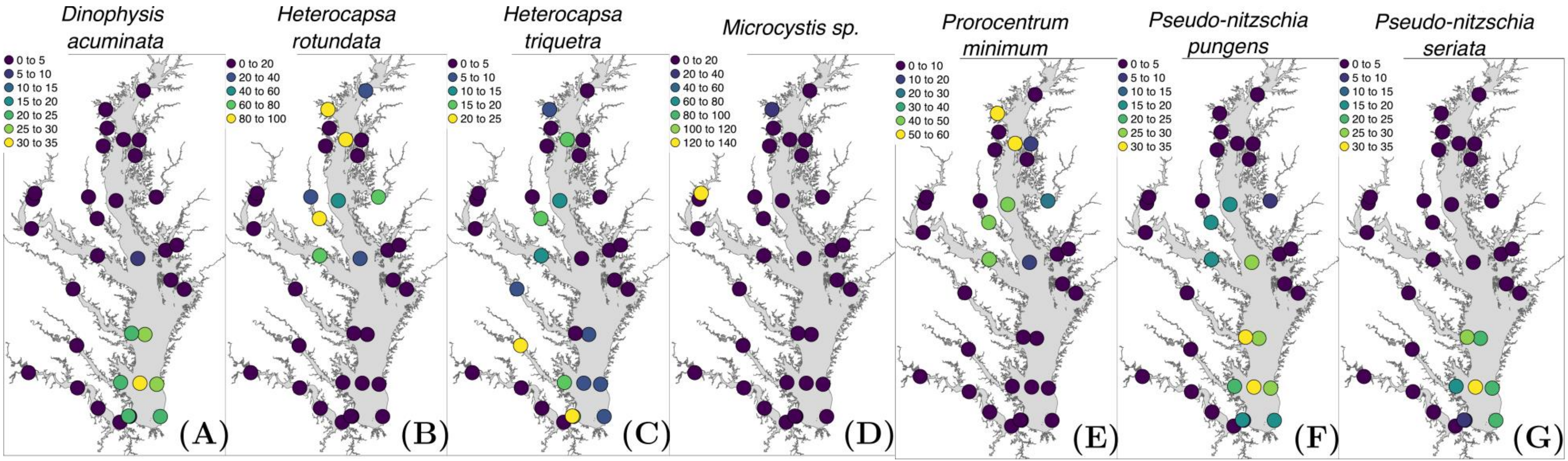
Methodology – *in situ* algal cell count data

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

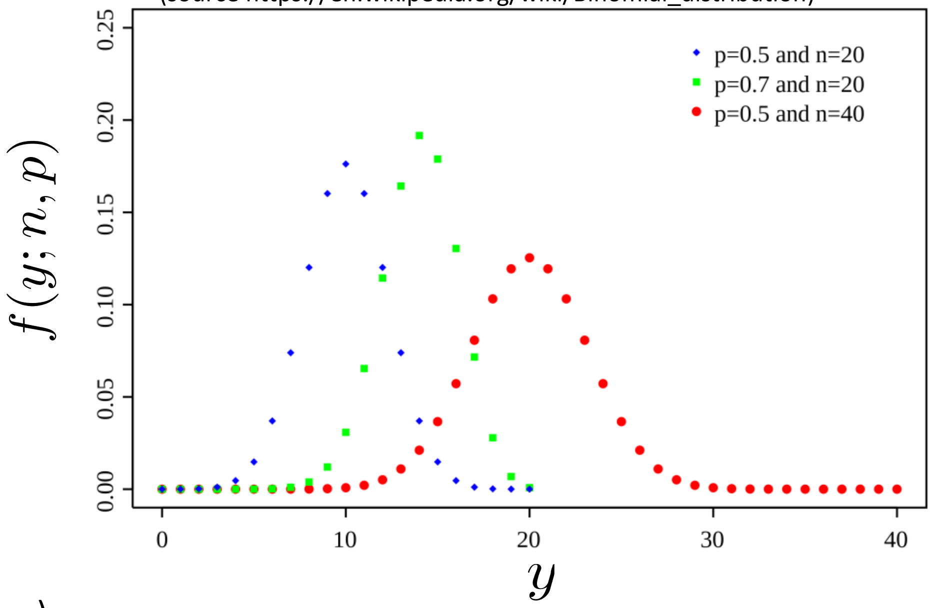


Methodology – statistical model type
Generalized linear models

1. We assume that the binary bloom data follows a binomial distribution

$$f(y; n, p) = \binom{n}{y} p^y (1-p)^{n-y},$$

(source https://en.wikipedia.org/wiki/Binomial_distribution)



p Probability bloom

n total number of trials

$y \in \{0, 1, \ldots, n\}$ number of successes (i.e., blooms)

$\binom{n}{y} = n!/[y!(n-y)!]$ binomial coefficient

Methodology – statistical model type
Generalized linear models

2. Link probability of bloom *p* to environmental conditions

$$f(y;n,p) = \binom{n}{y} p^y (1-p)^{n-y},$$

Apply identity operator exp(ln())

$$= \binom{n}{y} \exp[\eta y - n \ln(1+e^{\eta})],$$

with $\eta = \ln \frac{p}{1-p}$

p Probability bloom

n total number of trials

$y \in \{0, 1, \ldots, n\}$ number of successes (i.e., blooms)

$\binom{n}{y} = n!/[y!(n-y)!]$ binomial coefficient

Methodology – statistical model type
Generalized linear models

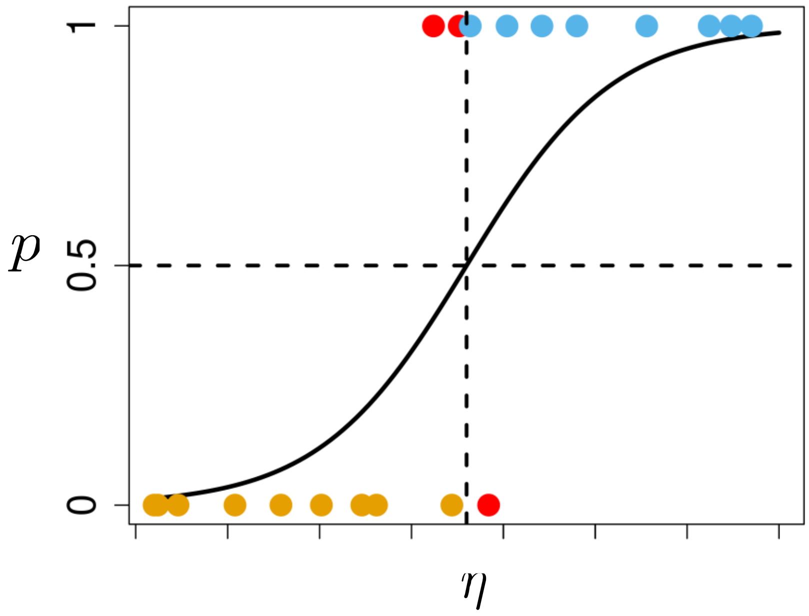
2. Link probability of bloom *p* to environmental conditions

$$f(y;n,p) = \binom{n}{y} p^y (1-p)^{n-y},$$

Apply identity operator exp(ln())

$$= \binom{n}{y} \exp[\eta y - n \ln(1+e^\eta)],$$

with $\eta = \ln \frac{p}{1-p}$



p Probability bloom

n total number of trials

$y \in \{0, 1, \ldots, n\}$ number of successes (i.e., blooms)

$\binom{n}{y} = n!/[y!(n-y)!]$ binomial coefficient

Methodology – statistical model type
Generalized linear models

2. Link probability of bloom p to environmental conditions

$$f(y;n,p)=\binom{n}{y}p^{y}(1-p)^{n-y},$$

Apply identity operator exp(ln())

$$=\binom{n}{y}\exp[\eta y-n\ln(1+e^{\eta})],$$

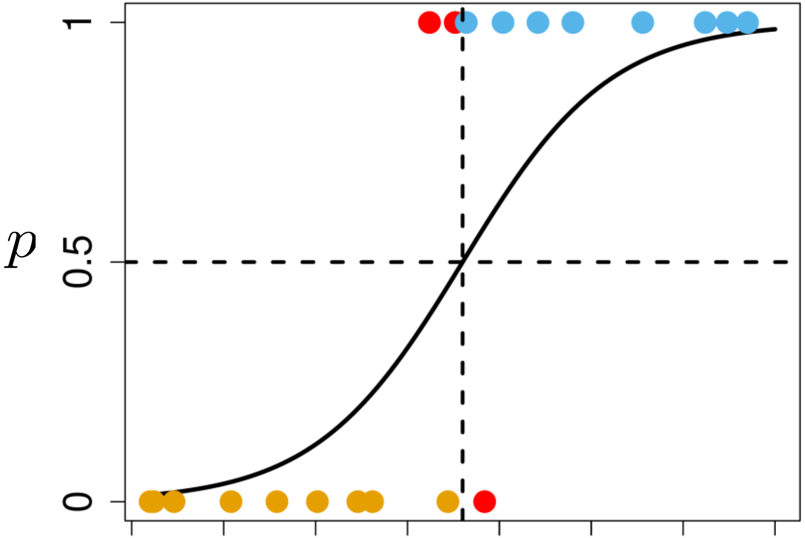
with $\eta=\ln\frac{p}{1-p}$

p Probability bloom

n total number of trials

$y\in\{0,1,\ldots,n\}$ number of successes (i.e., blooms)

$\binom{n}{y}=n!/[y!(n-y)!]$ binomial coefficient



$$\eta=\beta_0+\sum_{i=1}^{k}\beta_i x_i.$$

x_i environmental variables

β_i linear coefficients

Methodology – how to choose the environmental predictors?

- To avoid overfitting, we choose up to five environmental variables x_i

$$\mathcal{C} = \sum_{D=1}^{5} \frac{N^{\mathrm{var}}!}{D!(N^{\mathrm{var}}-D)!} = 381,$$

$N^{\mathrm{var}} = 9$ total number of environmental variables to choose from

D total number of variables selected

Methodology – how to choose the environmental predictors?

- To avoid overfitting, we choose up to five environmental variables x_i

$$\mathcal{C} = \sum_{D=1}^{5} \frac{N^{\mathrm{var}}!}{D!(N^{\mathrm{var}}-D)!} = 381,$$

$N^{\mathrm{var}} = 9$ total number of environmental variables to choose from

D total number of variables selected

- Next, we try all variable combinations and select the optimal set based on the accuracy forecasting a bloom/no-bloom and the Akaike Information Criterion (AIC):

$$\mathrm{AIC} = 2k' - 2\ln(\mathcal{L}),$$

k' number of model parameters

$\mathcal{L}$ likelihood function of the model

Methodology – how to choose the environmental predictors?

- To avoid overfitting, we choose up to five environmental variables x_i

$$\mathcal{C} = \sum_{D=1}^{5} \frac{N^{\mathrm{var}}!}{D!(N^{\mathrm{var}}-D)!} = 381,$$

$N^{\mathrm{var}} = 9$ total number of environmental variables to choose from

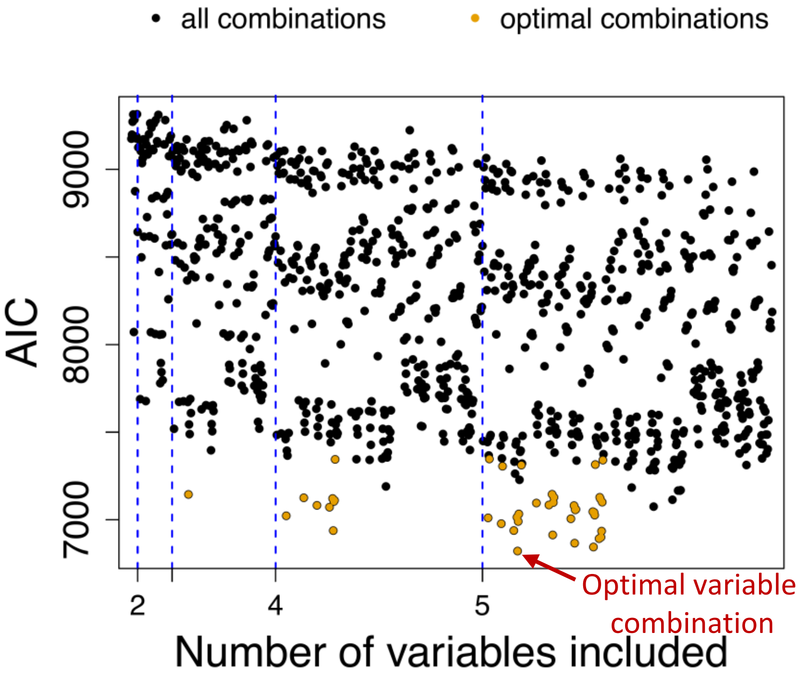
D total number of variables selected

- Next, we try all variable combinations and select the optimal set based on the accuracy forecasting a bloom/no-bloom and the Akaike Information Criterion (AIC):

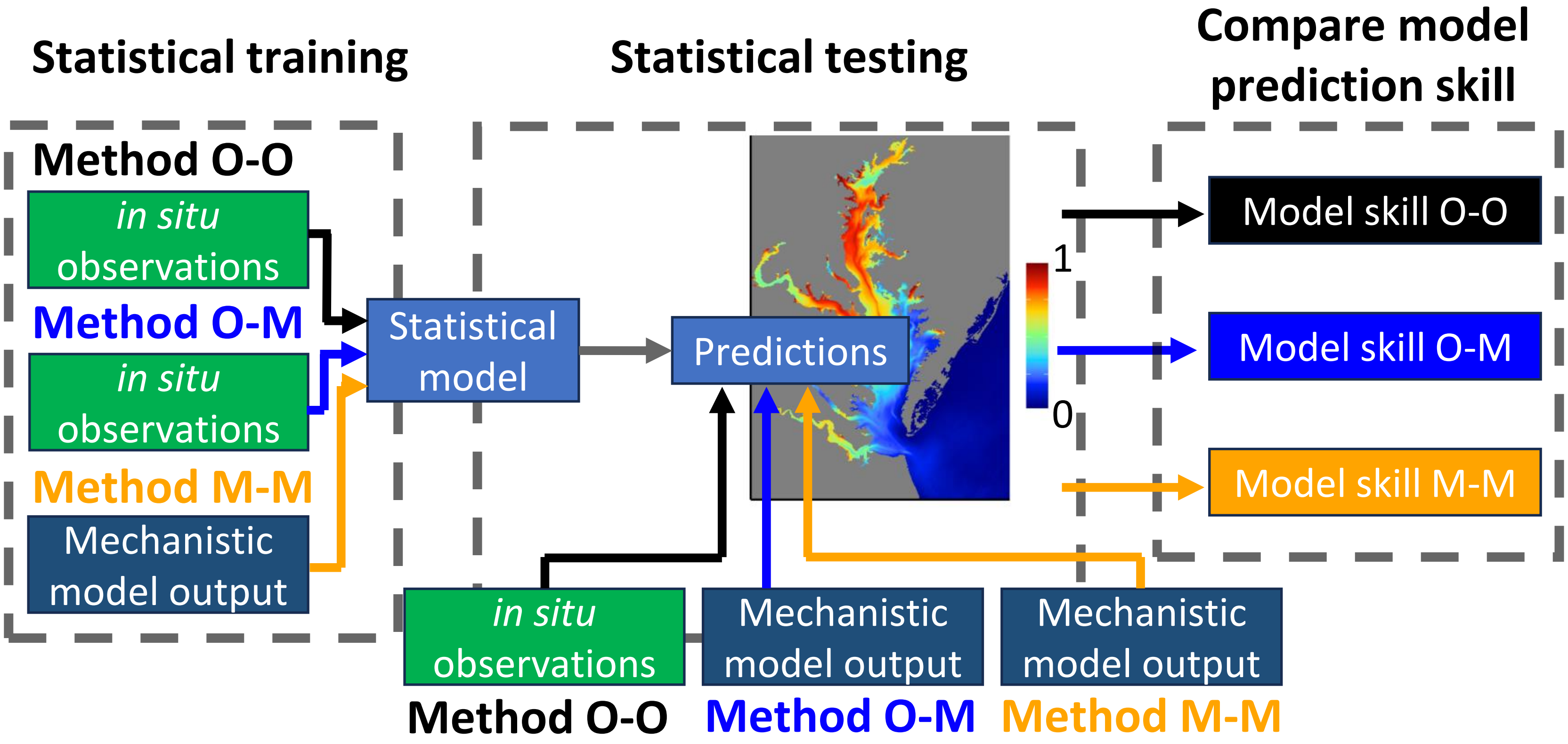
$$\mathrm{AIC} = 2k' - 2\ln(\mathcal{L}),$$

k' number of model parameters

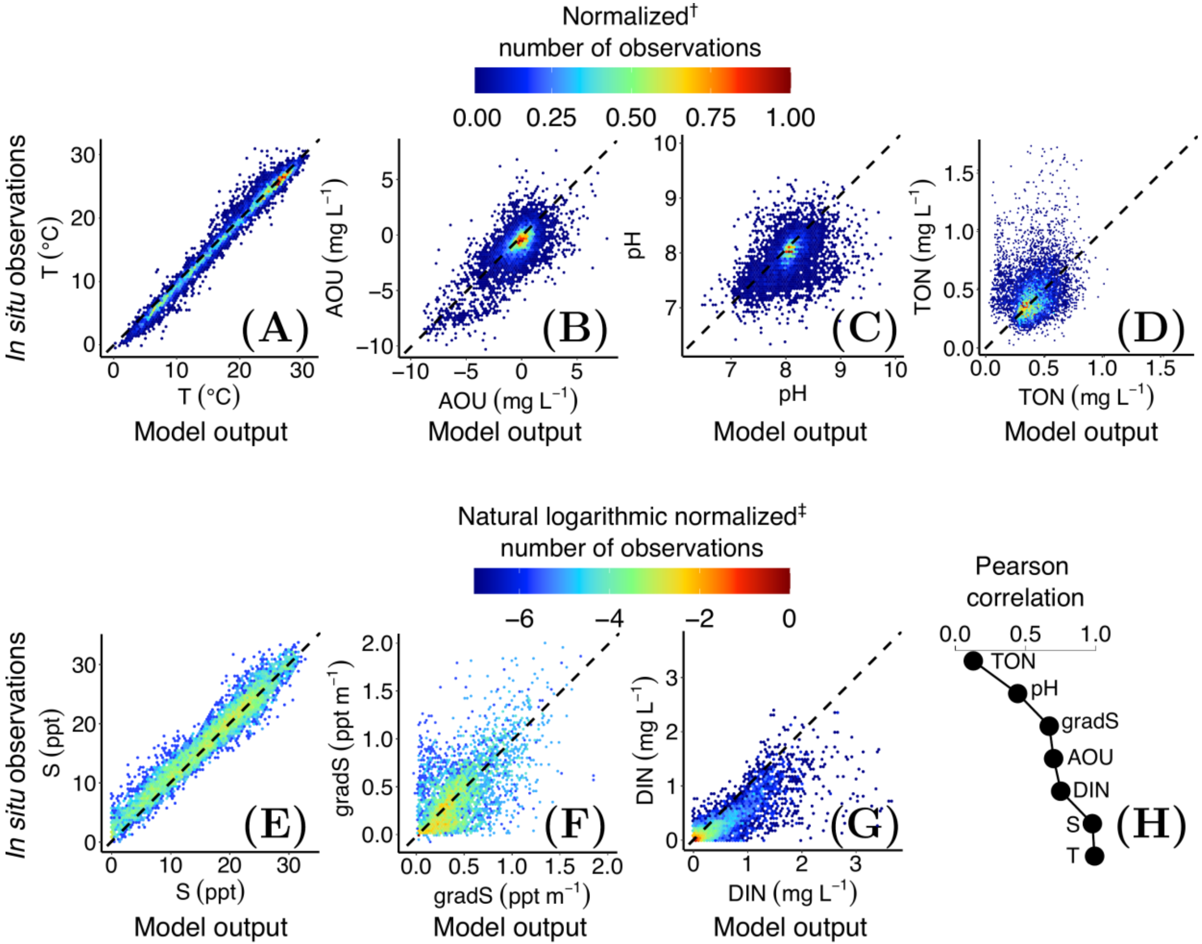
$\mathcal{L}$ likelihood function of the model



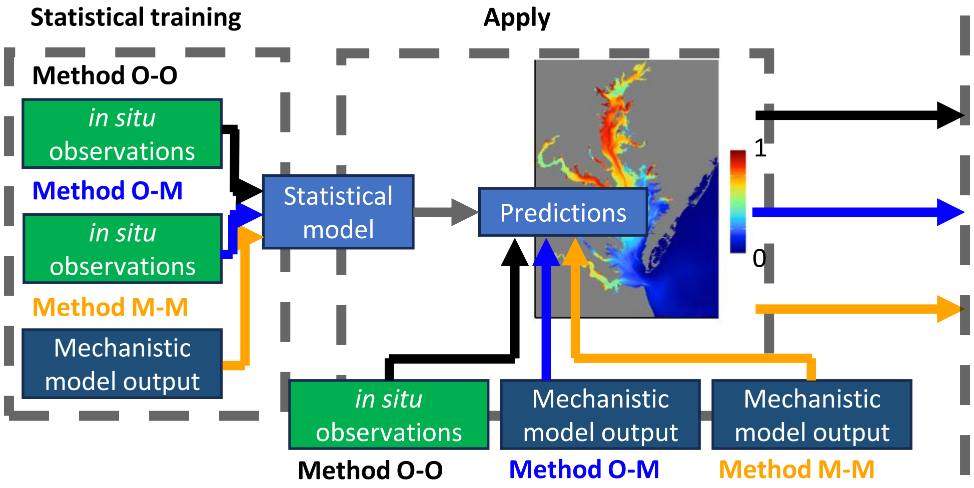
We compare three methods to train the statistical forecasting model



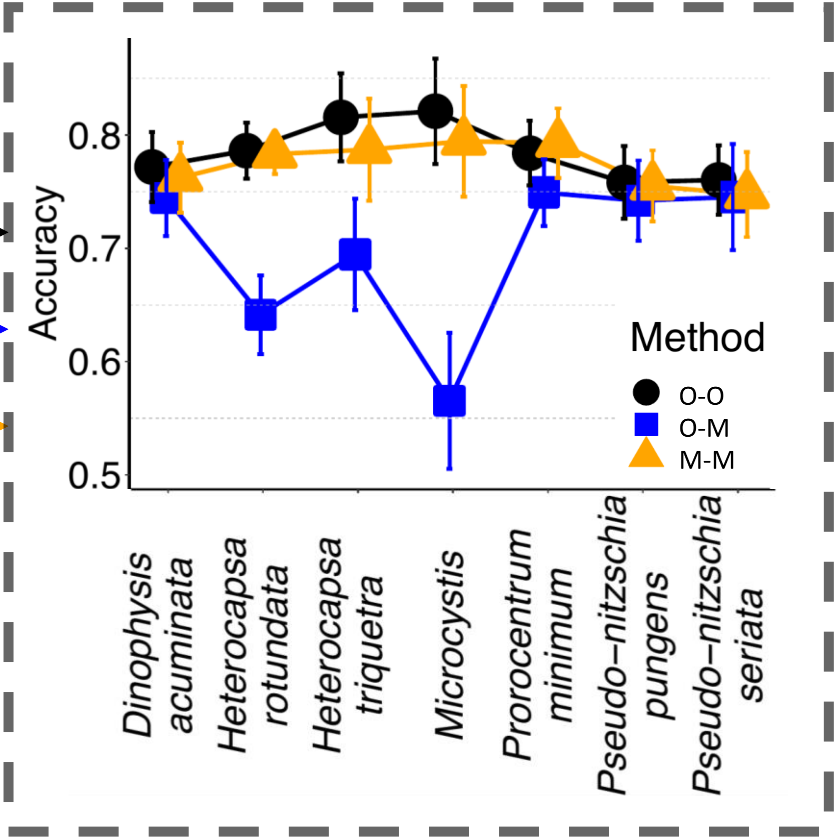
Comparing *in situ* observations and mechanistic model output



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



Test prediction skill

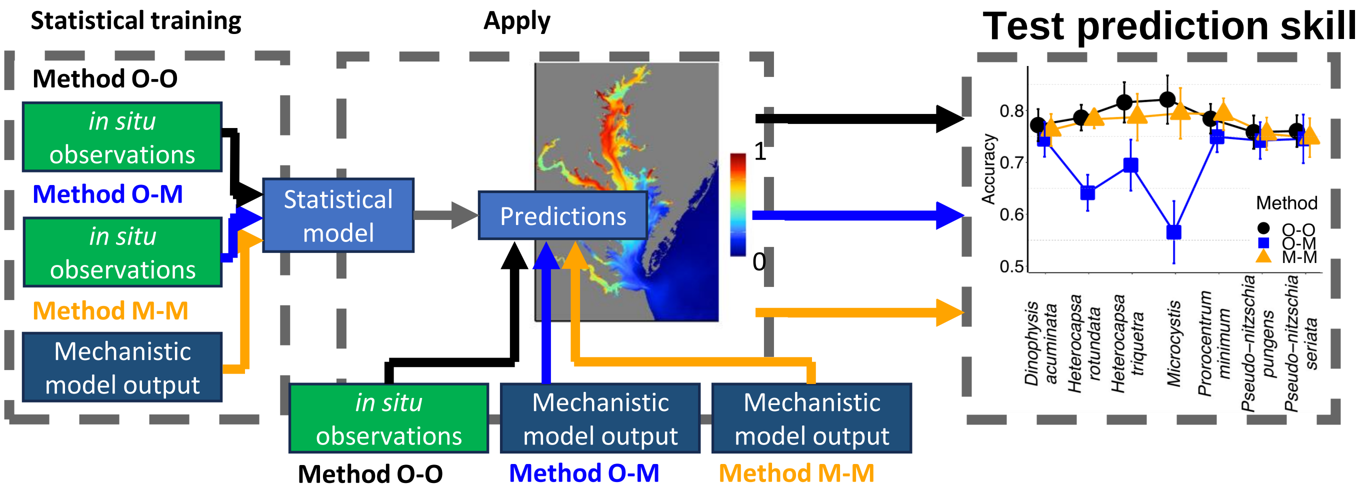


Conclusions

Should we use *in situ* observations to train our statistical model?

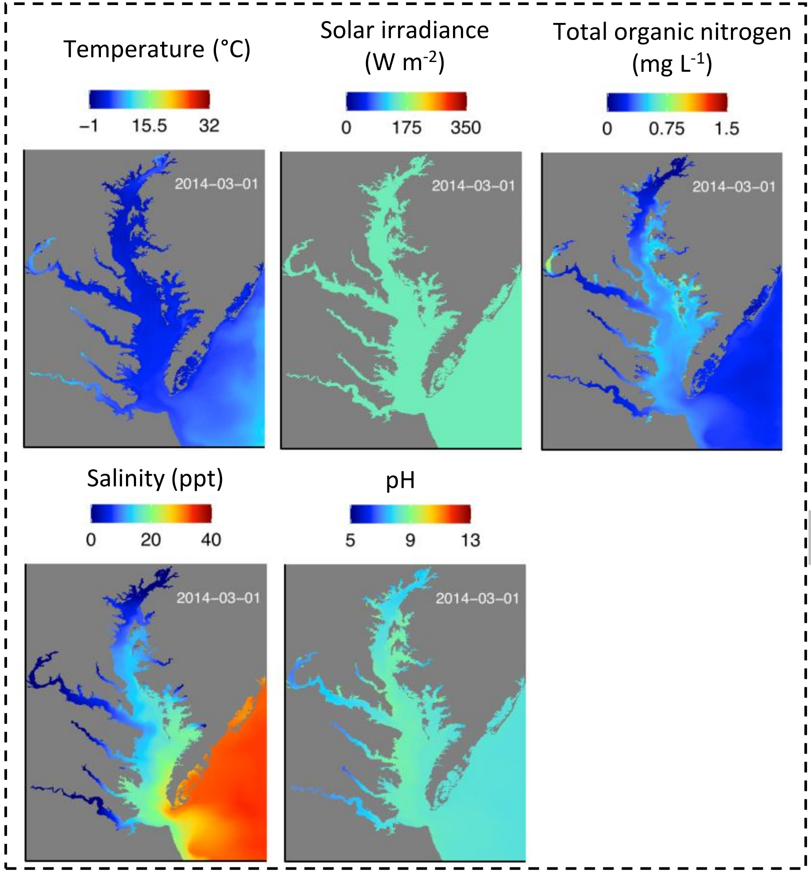
i) Statistical models trained using *in situ* observations are less accurate when applied to model output (Method O-M)

ii) Training and applying the statistical model using mechanistic model output (Method M-M) enhances model prediction skill



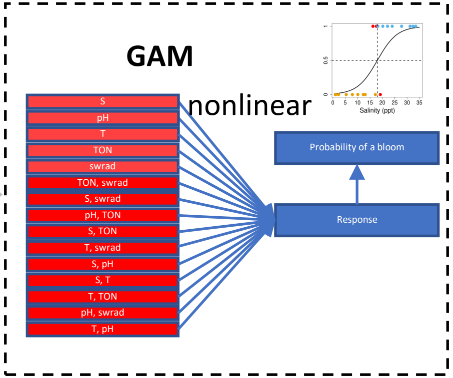
Adding forecasts of *Prorocentrum minimum* to CBEFS

Existing model forecasts



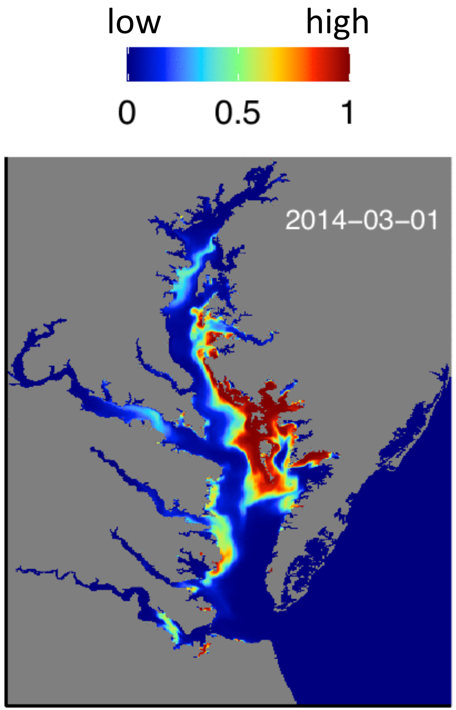
in

Empirical habitat model



out

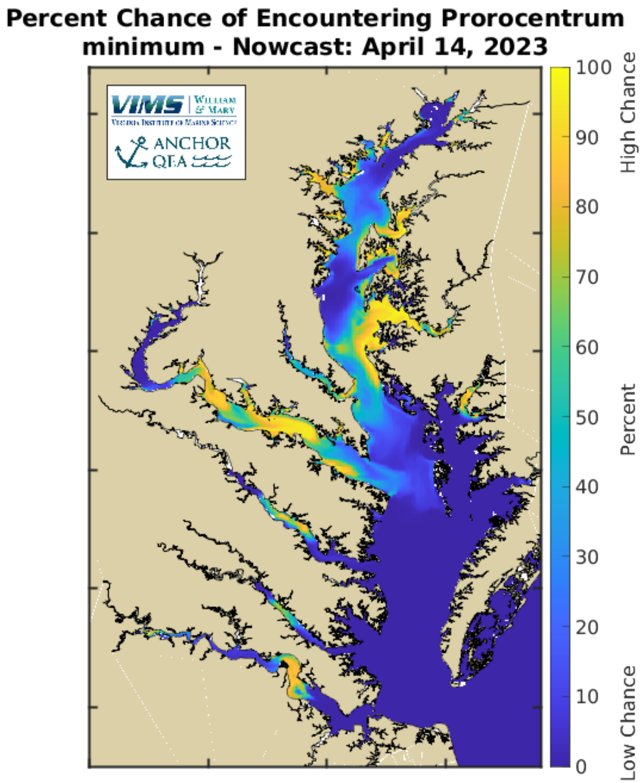
Probability of a *Prorocentrum minimum* bloom

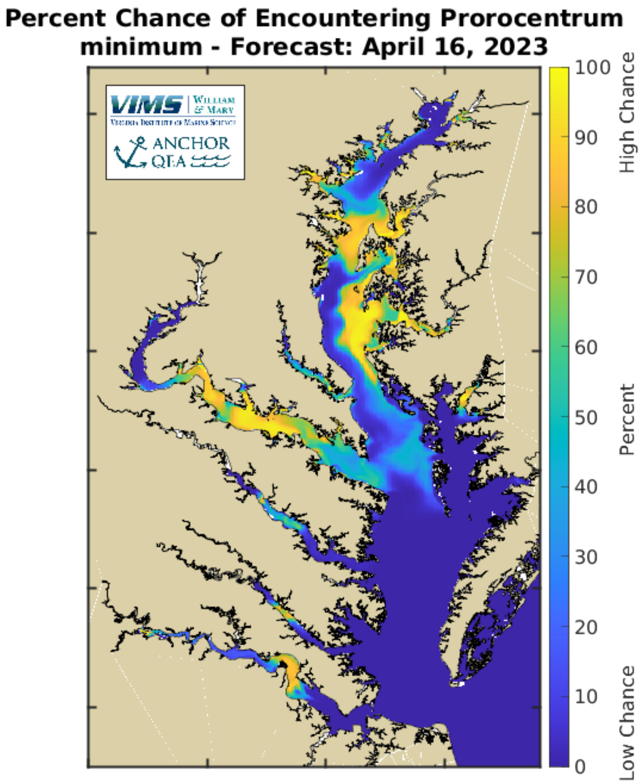


www.vims.edu/cbefs

Adding forecasts of *Prorocentrum minimum* to CBEFS

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



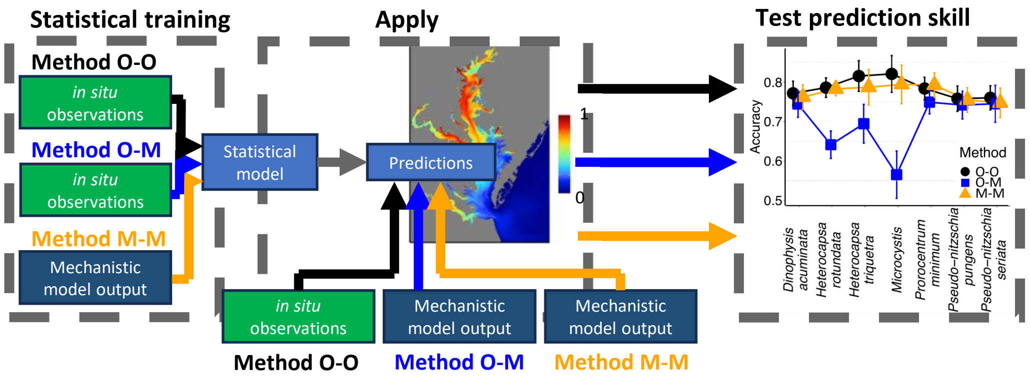


www.vims.edu/cbefs

Take home messages

Type of training information affects the construction of statistical habitat suitability models

in situ observations versus mechanistic model output



Please visit us at
www.vims.edu/cbefs

dmlhoremans@vims.edu