# Application of CART to Predict Nutrient Limitation Patterns for the 2017 Bay Model Output

Qian Zhang, Richard Tian, Lew Linker

Modeling Workgroup Conference Call
September 10, 2020

Background

- Chesapeake Bay has well-documented seasonal and spatial patterns in nutrient limitation to algal growth (Kemp et al., 2005).

- These patterns were determined using bioassays collected from the 1992-2002 (Fisher et al., 2002, 2005).

- These patterns were used in the calibration of the 2017 Chesapeake Bay estuarine model.

Kemp et al. (2005)
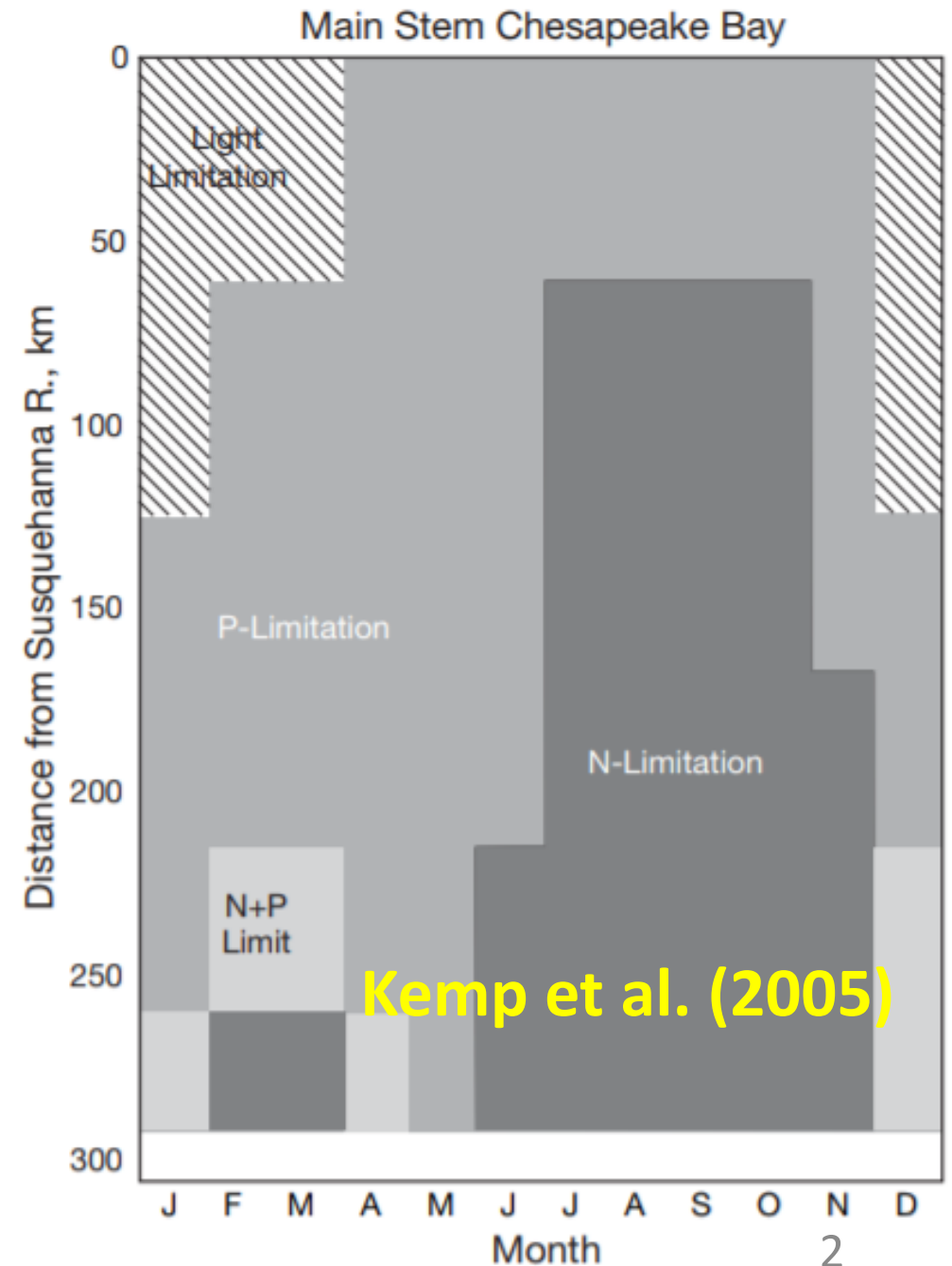
Recent Research*

1. Developed an <u>empirical approach</u> to relate tidal monitoring data to bioassay-based nutrient limitation ("truth") in the concurrent period of 1992-2002,

2. Applied the selected approach to tidal monitoring data in more recent periods to predict nutrient limitation and explore potential changes in limitation in response to altered nutrient loading ("Goal 2").

* **Zhang**, Q., T. R. **Fisher**, E. M. **Trentacoste**, C. **Buchanan**, A. B. **Gustafson**, R. **Karrh**, R. R. **Murphy**, J. **Keisman**, C. **Wu**, R. **Tian**, J. M. **Testa** and P. J. **Tango**, 2020. Nutrient limitation of phytoplankton in Chesapeake Bay: Development of an empirical approach for water-quality management. *Water Research,* 116407, doi: 10.1016/j.watres.2020.116407.

Study Sites & Data

- Bioassay-based Limitation Classes for 1992-2002
  - ❑ 6 stations x 12 months

- Tidal WQ Monitoring Data in 1990-2018 (21 Stations)
  - ❑ Chesapeake Bay Program Data Hub (> 3,000,000 values)
 - ❑ Aggregated 1992-2002 data to the same size as bioassay classes (6 stations x 12 months)

Main Stem Chesapeake Bay

"truth"
**72 classes**
6 stations x 12 months
(1992-2002)

Kemp et al. (2005)

4
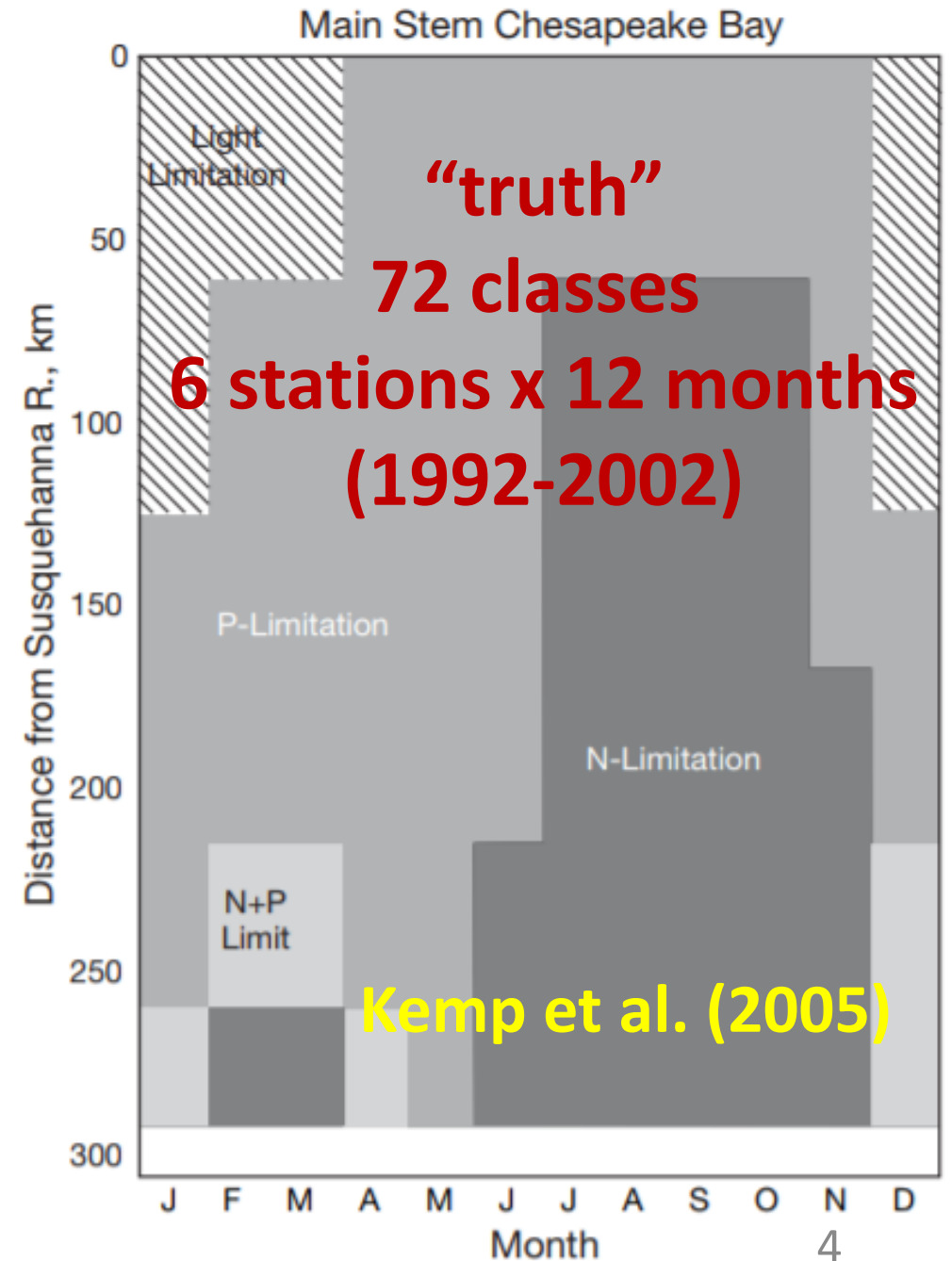
Analysis of Mainstem Limitation (1992-2002)

Target: 72 bioassay-based limitation classes

Approach	Variables	Decision Rules	Classification Rate
A1. Probability-based approach	DIN and DIP concentrations	N (low DIN ≥ 50%) P (low DIP ≥ 50%) NP (both ≥ 40%) NoR (else)	43% (31 matches / 72)
A2. Nutrient index-based approach	DIN and DIP indices (based on concentrations)	N (N-index ≥ 0.5) P (P-index ≥ 0.5) NP (both ≥ 0.4) NoR (else)	57% (41 matches / 72)
A3. ★ Classification and Regression Trees (CART)	DIN, DIP, + more (e.g., WTEMP, N:P ratio, CHLA, Secchi, Salinity)	Data-driven (through CART)	89% (64/72; LOOCV); 99% (71/72; Full Data)

Mainstem Chesapeake Bay (1992–2002) Bioassay Data
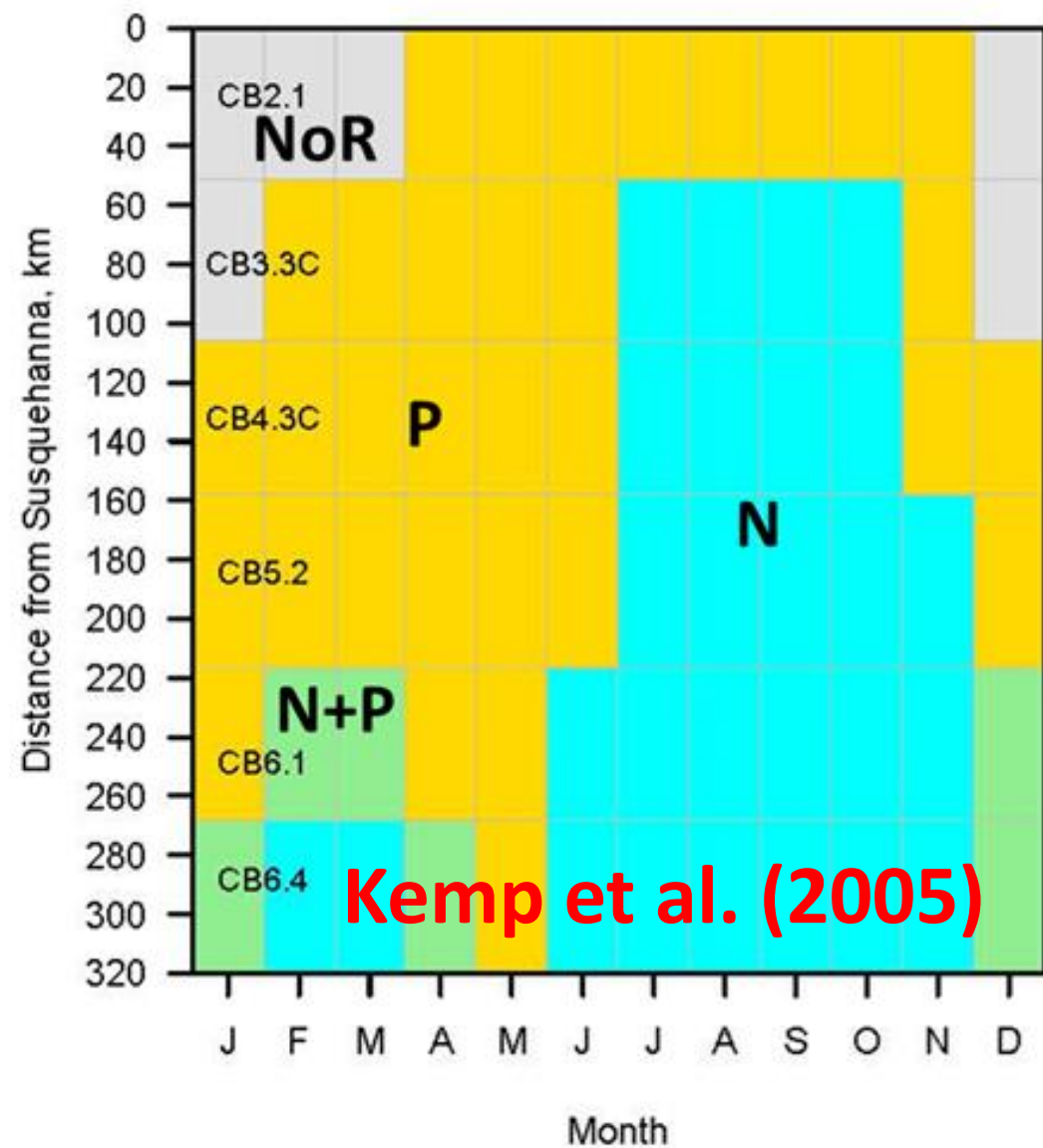
Mainstem Chesapeake Bay (1992–2002) Monitoring Data (CART – Full Data)

Kemp et al. (2005)
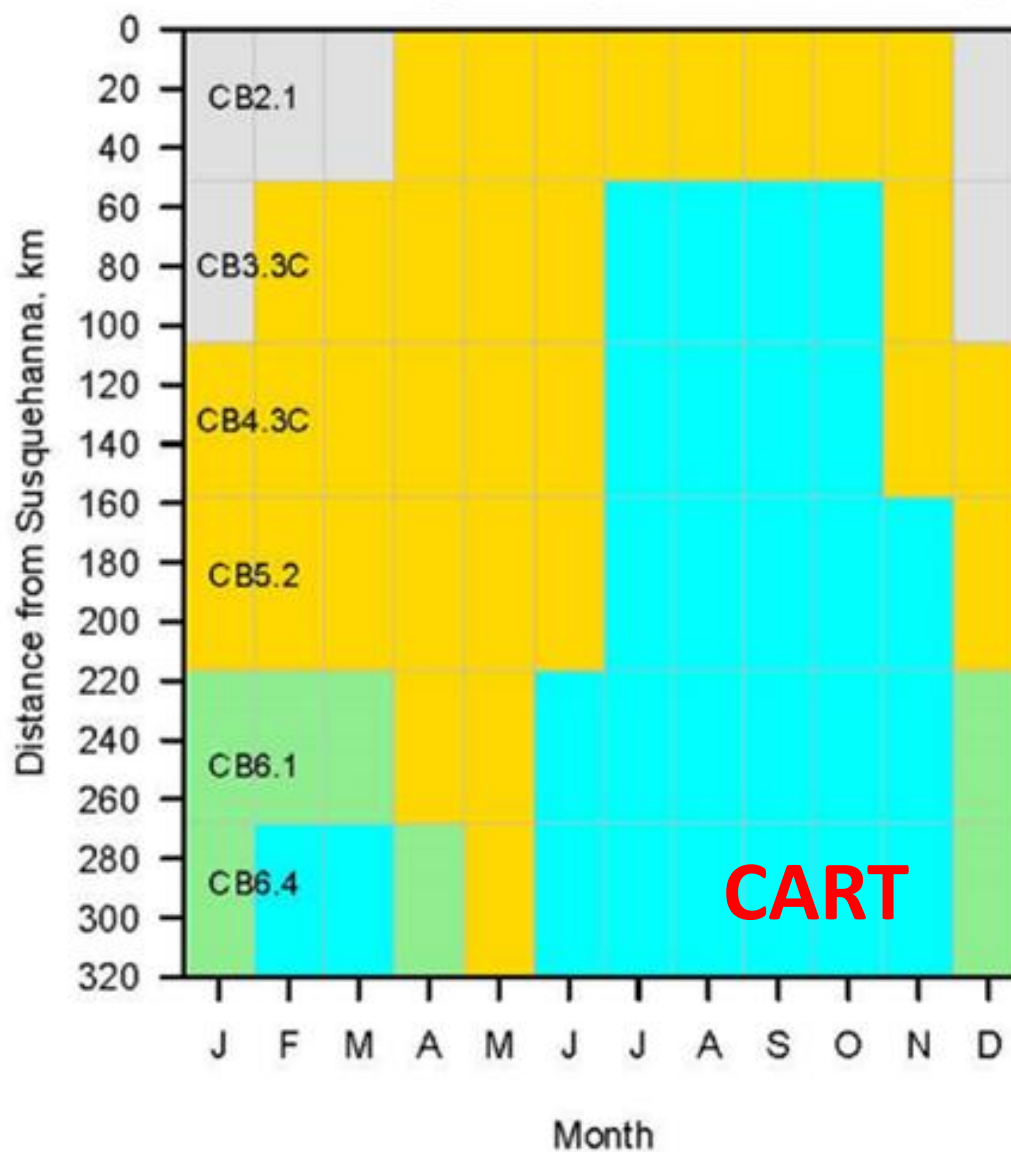
CART

Applications of CART to the 2017 WQSTM

1. Run CART on monitoring data and modeling data to determine and compare nutrient limitation for the 1991-2000 period:

    ❖ monitoring data in 1991-2000 (2017 WQSTM calibration and base years).

    ❖ WQSTM model data for days matched with the monitoring data in 1991-2000.

    ❖ WQSTM model data for all days in 1991-2000.

2. Run CART for key scenarios (high nutrient loads to low) for the 1991-2000 period:

    ❖ No Action

    ❖ WIP3

    ❖ E3

    ❖ All forest

    ❖ Progress runs

Two Considerations

1. CART model form:

   Due to data availability, we focused on the use of CART model 4, which uses the following features (explanatory variables): WTEMP, TNTP.ratio, indexN, and DIP.

2. Data matching:

   Output data of these variables from the Bay Model are available at different vertical depths, so we attempted a few different ways of matching these data with the tidal monitoring data in terms of vertical depths. Results are very insensitive to this treatment.

Monitoring vs. Modeling Data (1991-2000)

Mainstem Chesapeake Bay (1991–2000)
Monitoring Data (CART; 21 Stations)
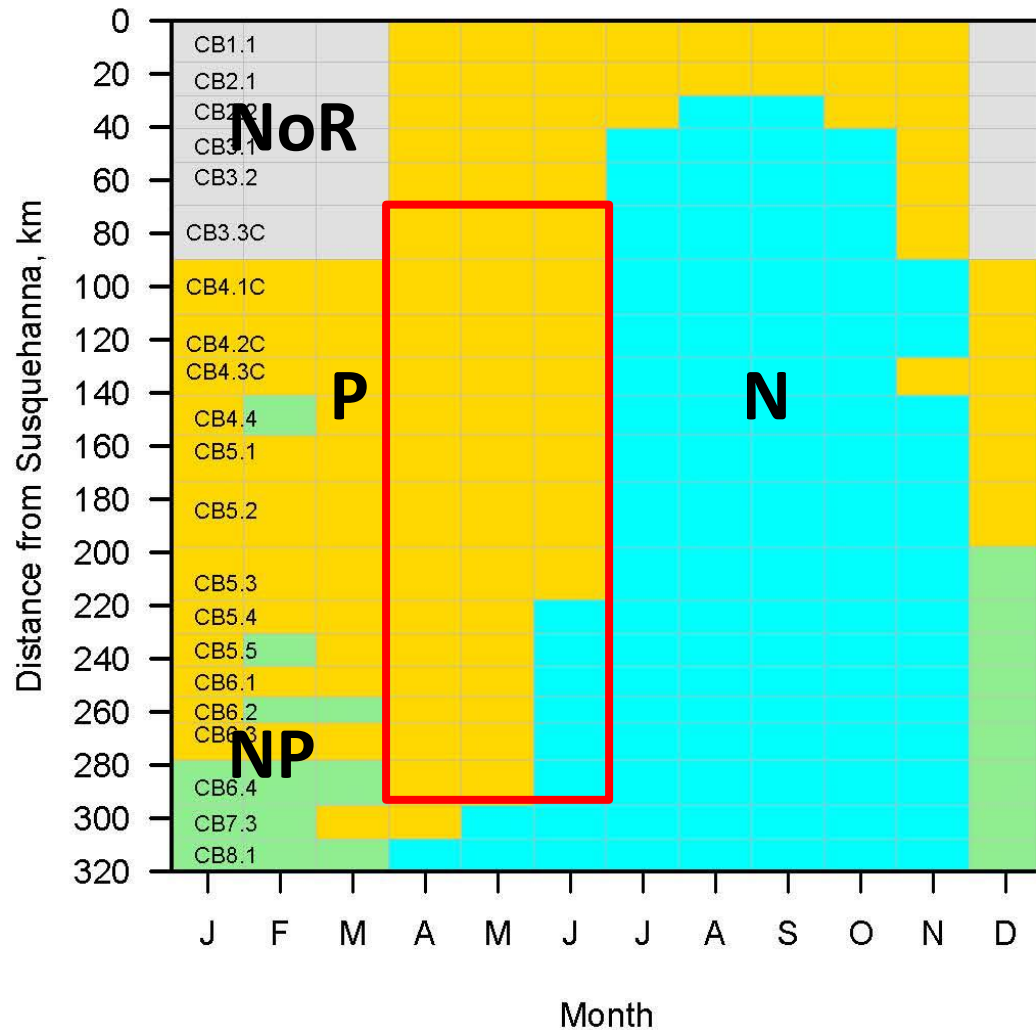
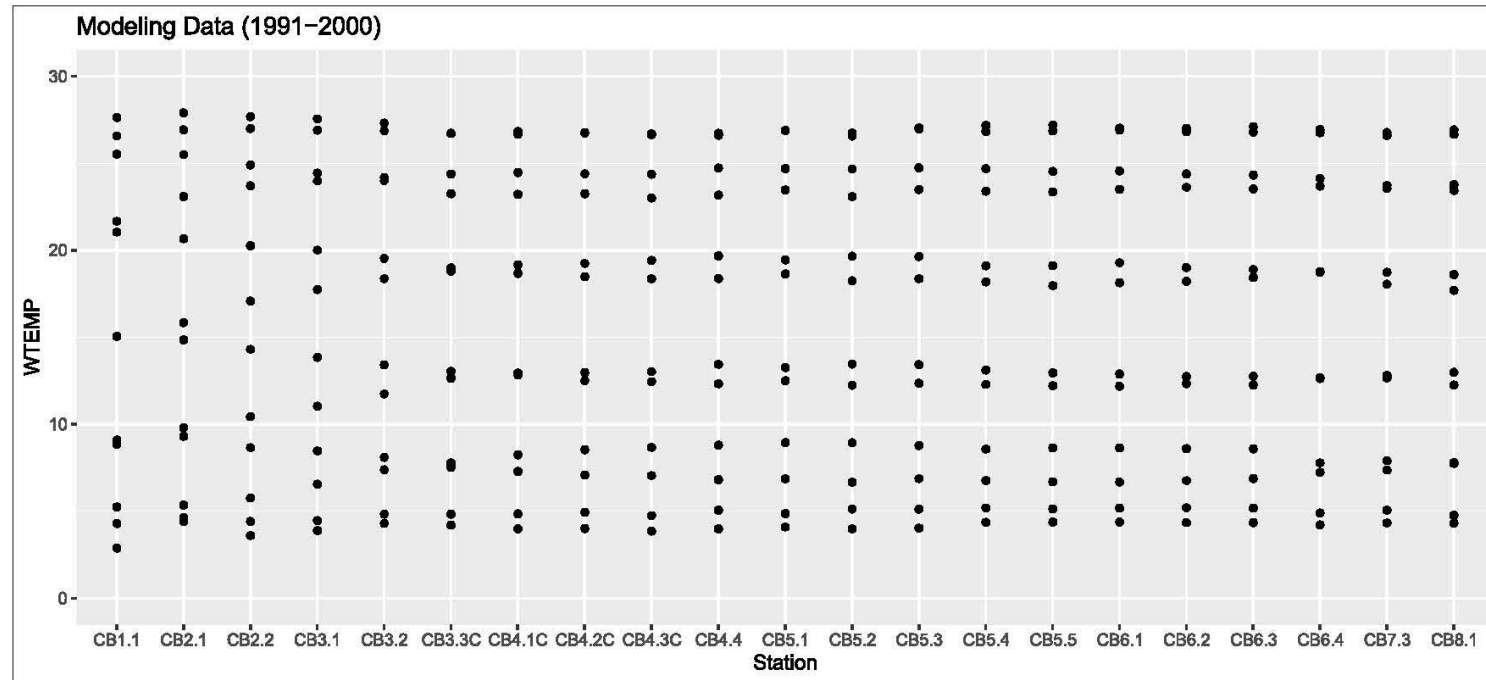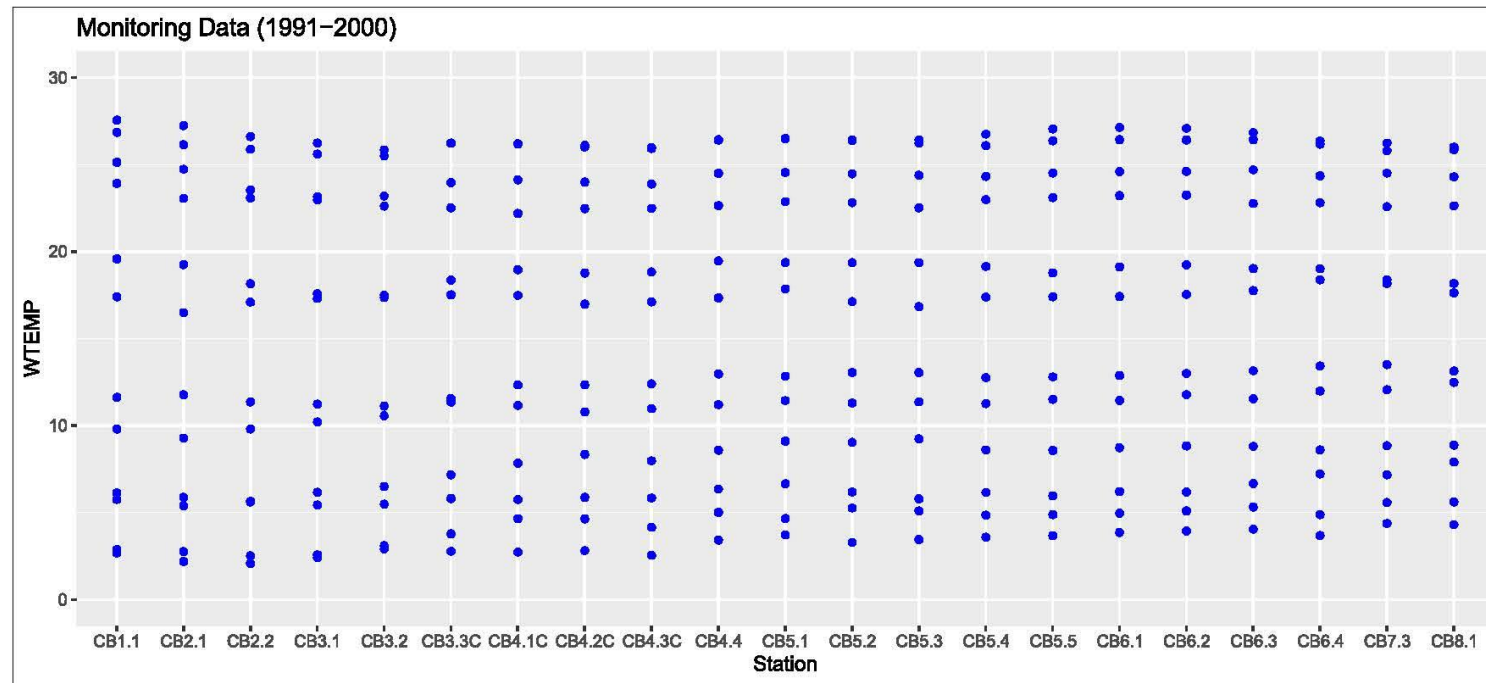Mainstem Chesapeake Bay (1991–2000)
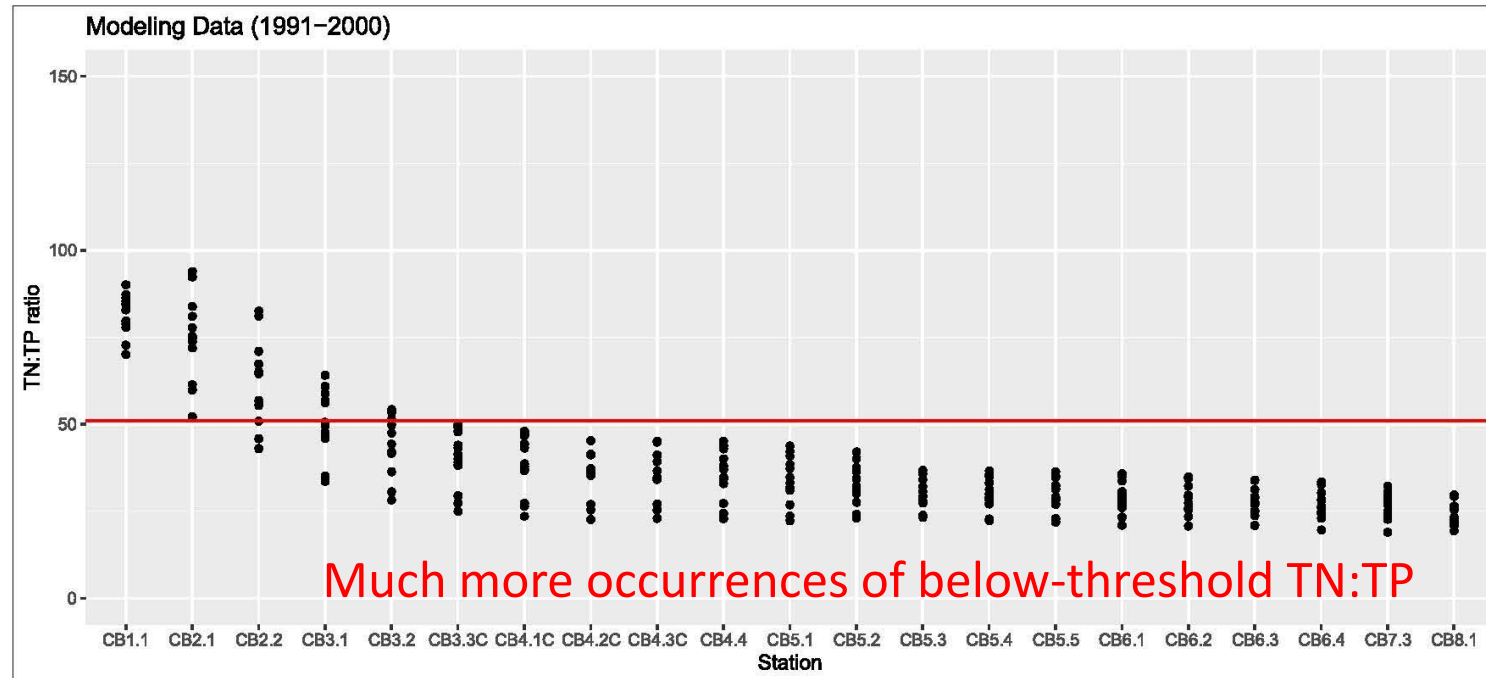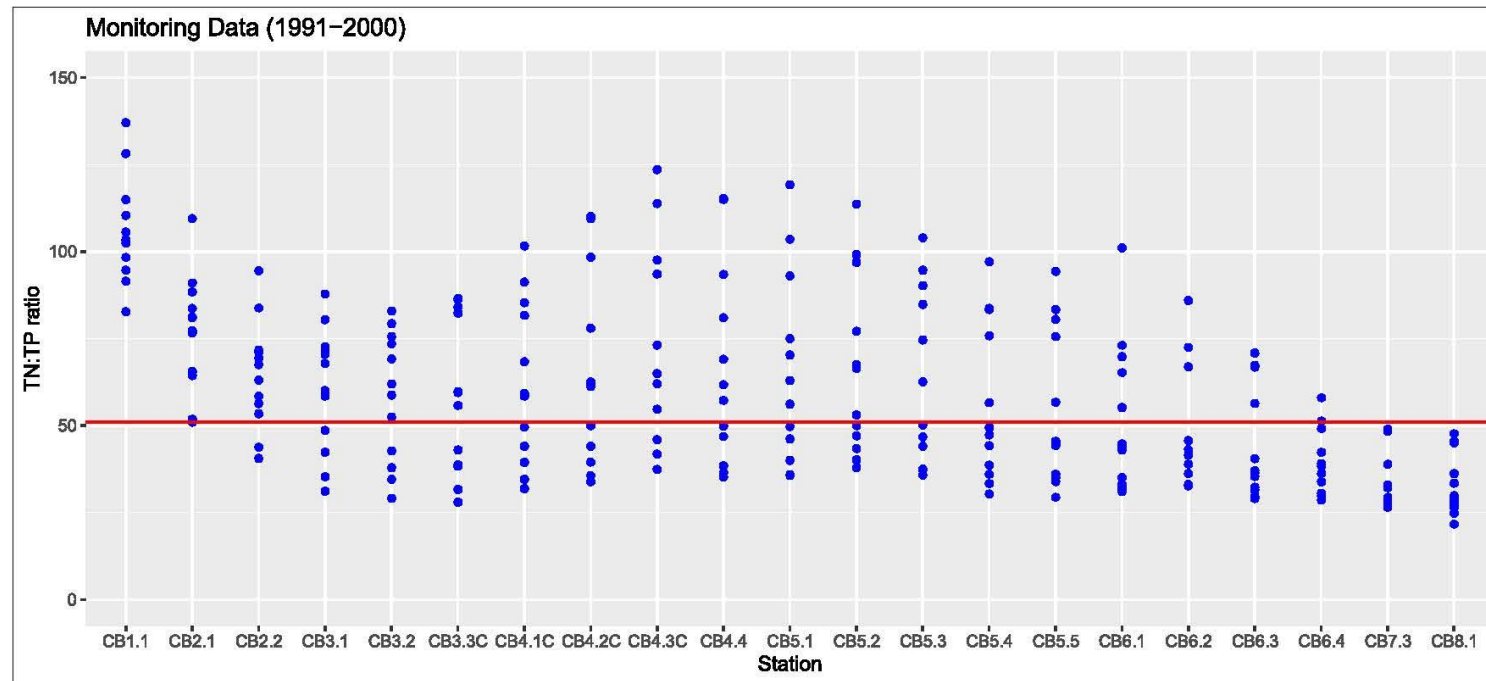Modeling Data (CART; 21 Stations)

CART Tree 4

WTEMP

TN:TP

Monitoring Data (1991-2000)

Modeling Data (1991-2000)

Much more occurrences of below-threshold TN:TP

12

DIP

Monitoring Data (1991−2000)

Modeling Data (1991−2000)

Much more occurrences of above-threshold DIP
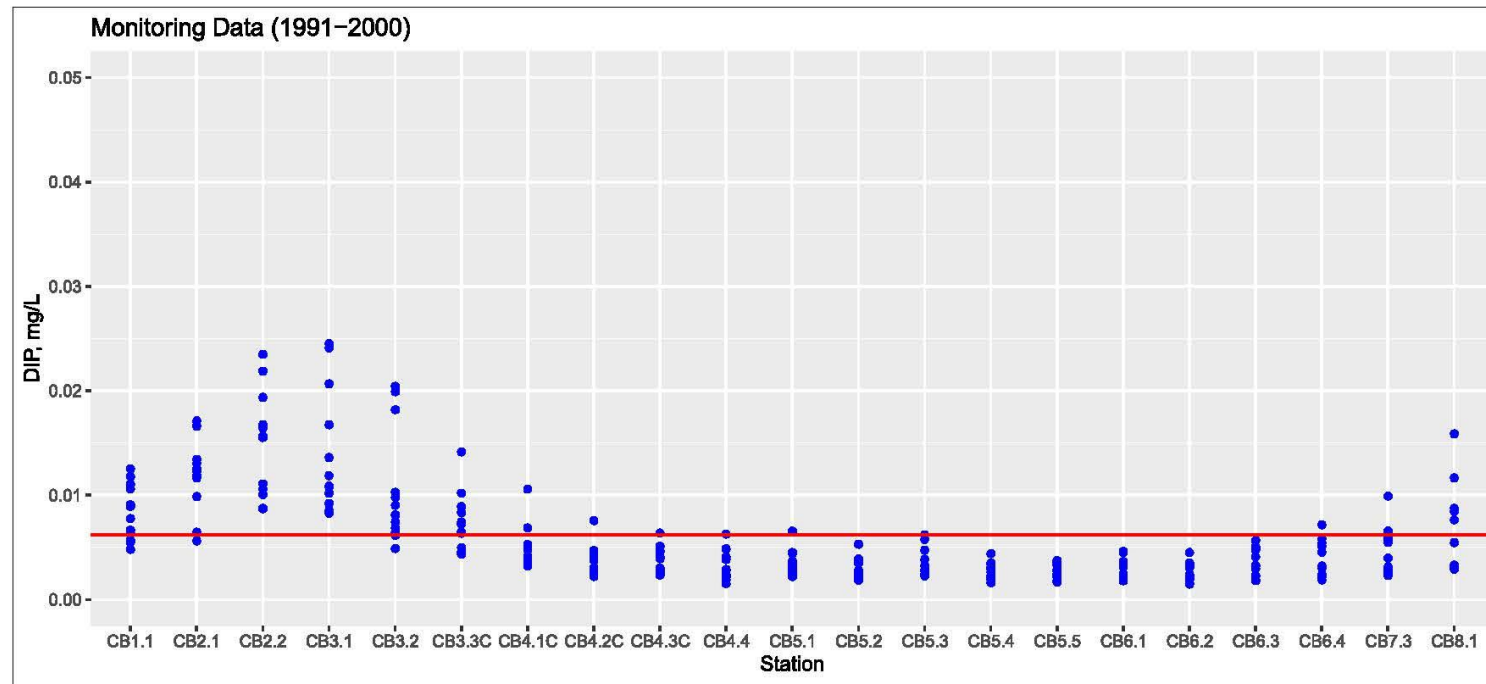
13

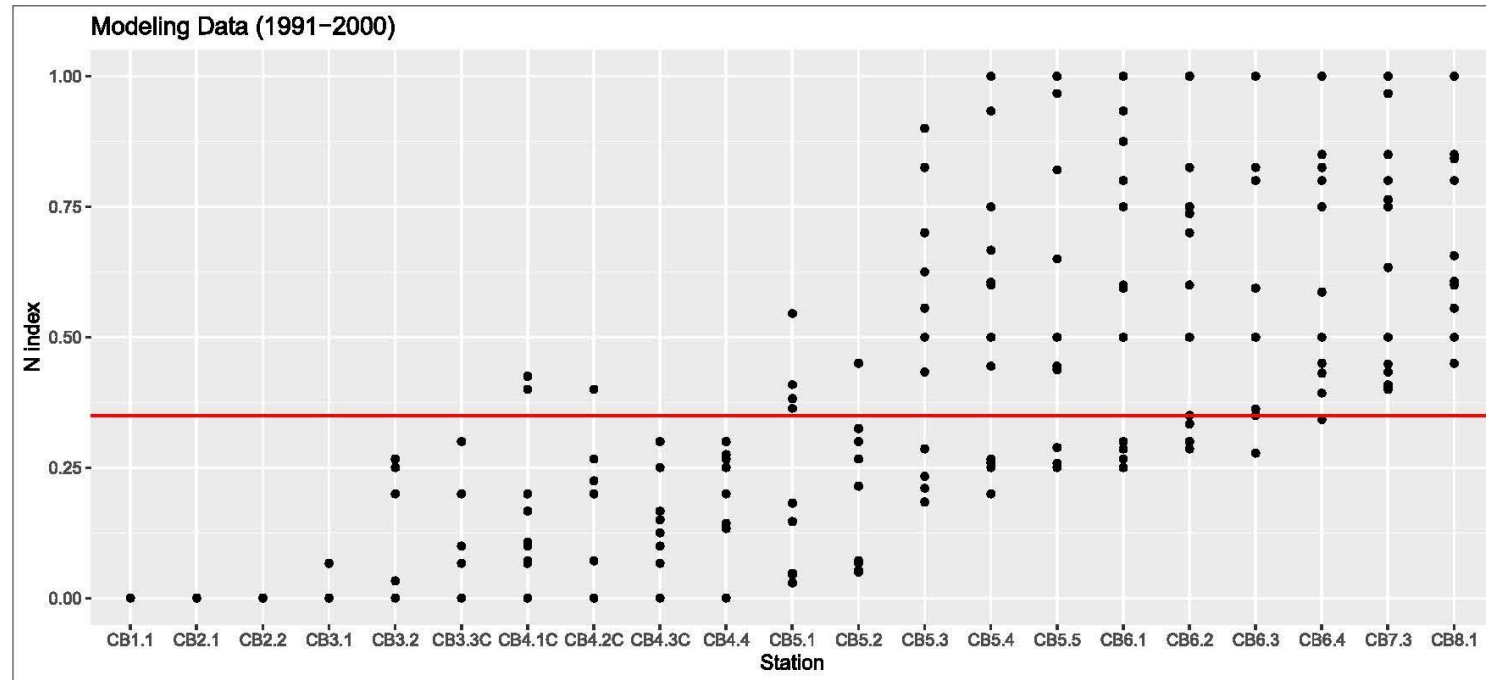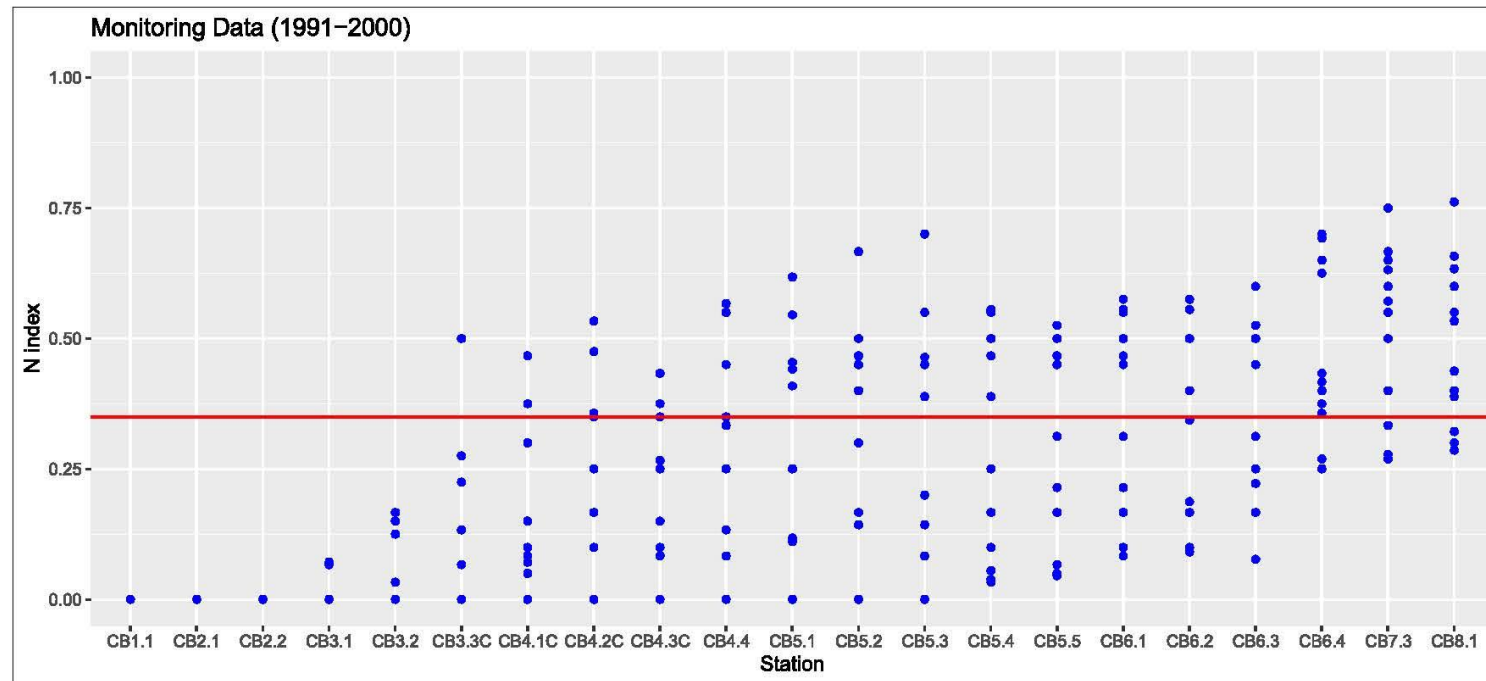# N index

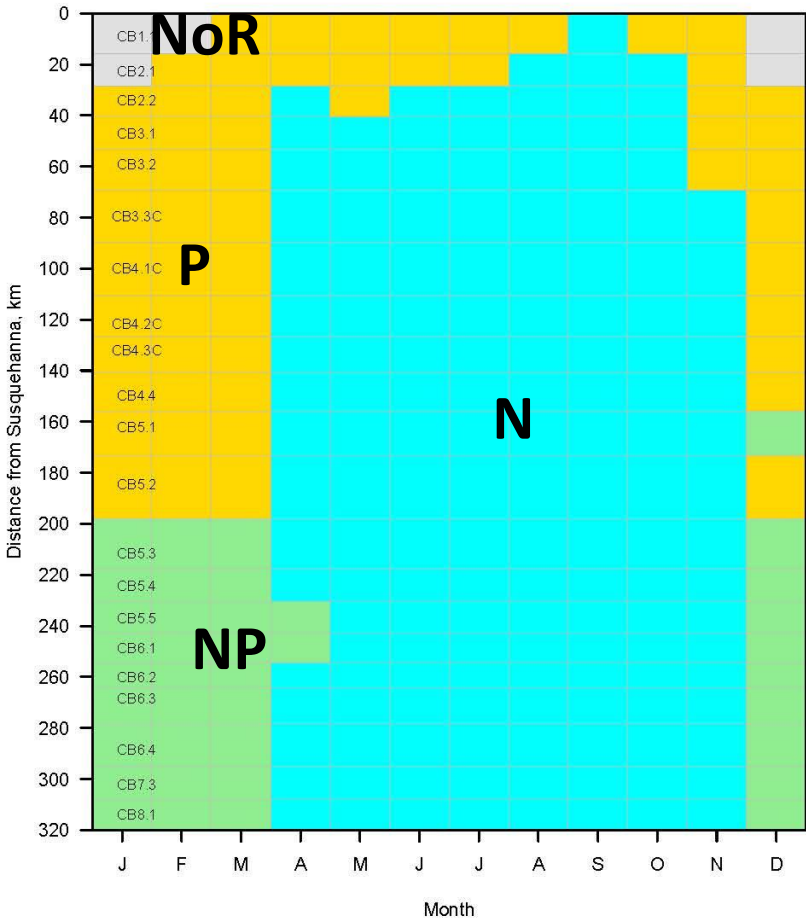Monitoring Data (1991–2000)

Modeling Data (1991–2000)

14
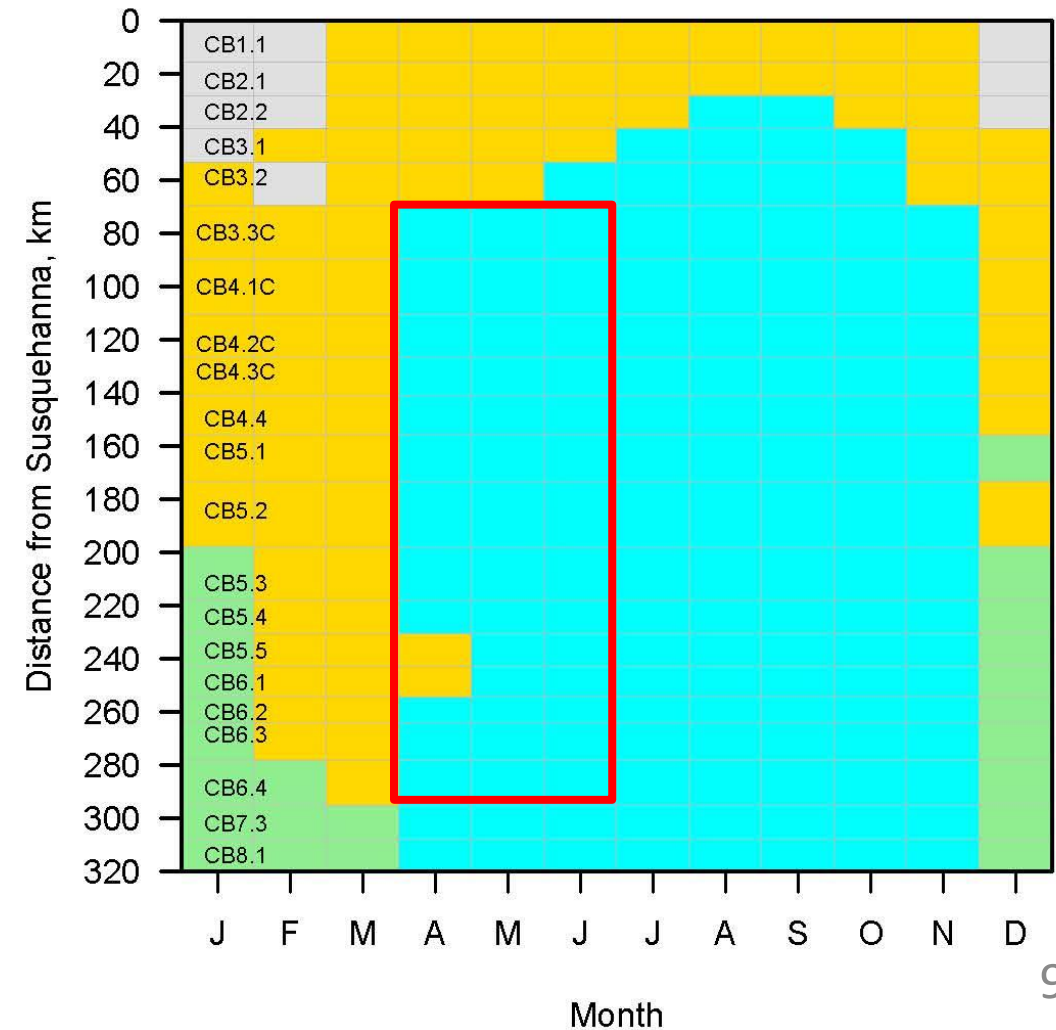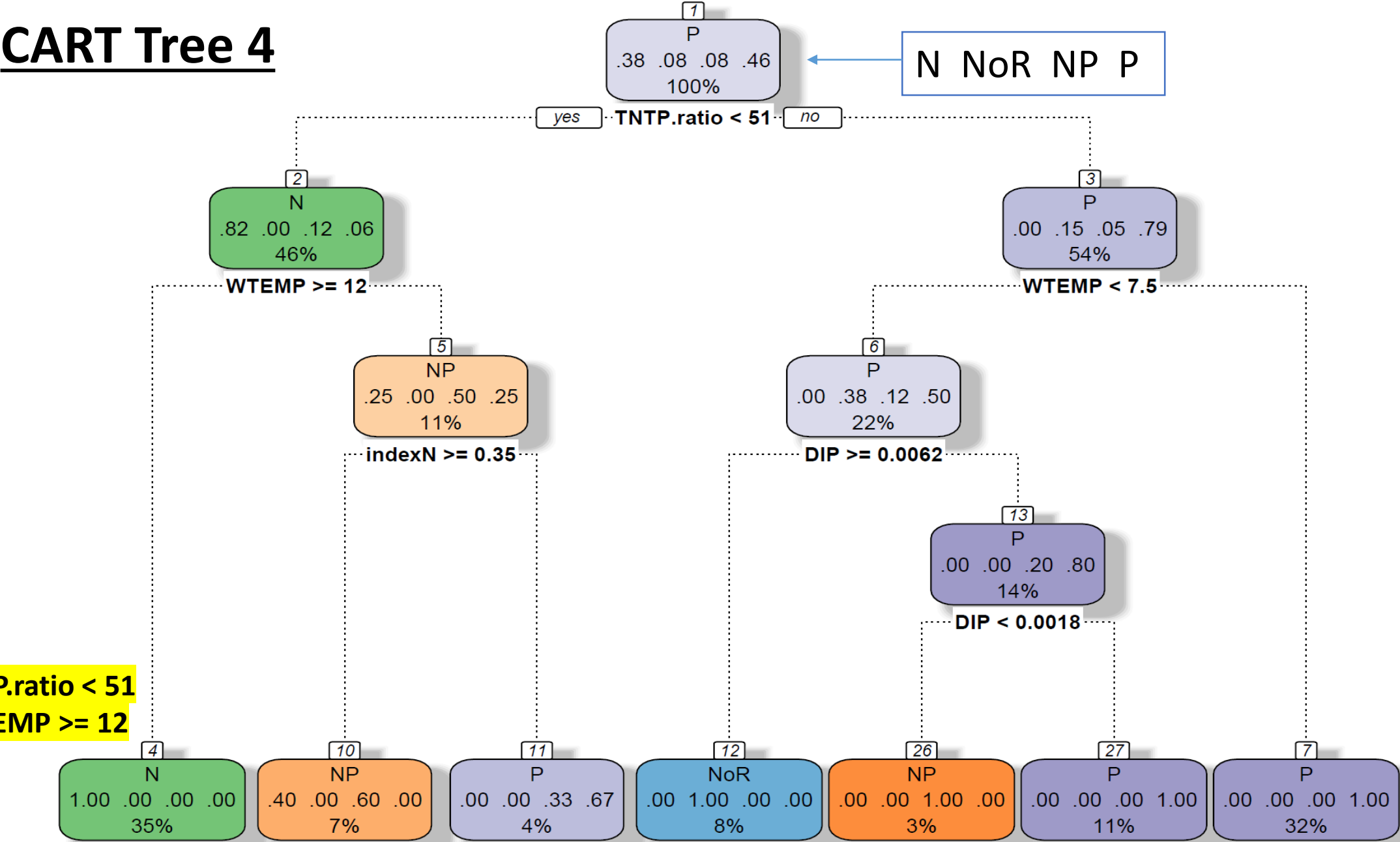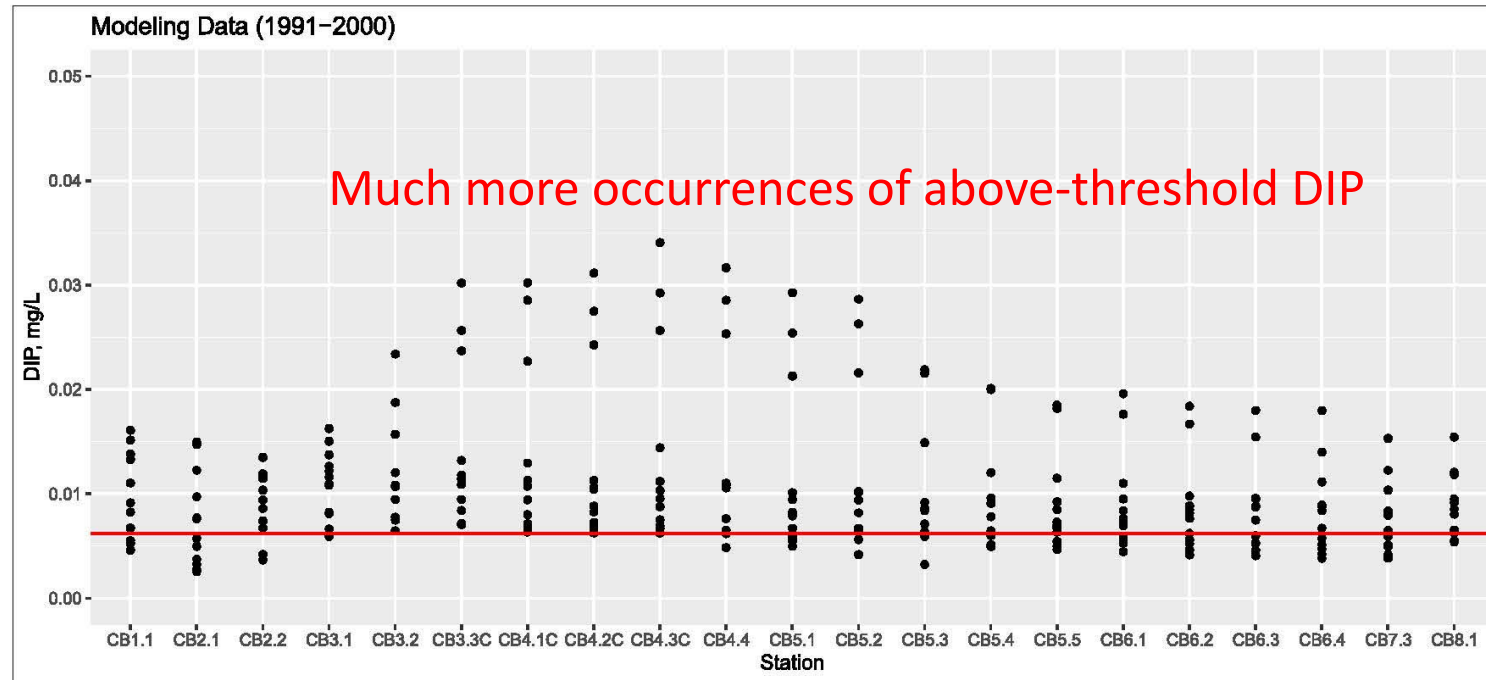
Potential Applications to the 2017 WQSTM

1. Run CART on monitoring data and modeling data to determine and compare nutrient limitation:

   ❖ monitoring data in 1991-2000 (2017 WQSTM calibration and base years).

   ❖ WQSTM model data for days matched with the monitoring data in 1991-2000.

   ❖ WQSTM model data for all days in 1991-2000.

2. Run CART for key scenarios (high nutrient loads to low) for 1991-2000:
   ❖ No Action
   ❖ WIP3
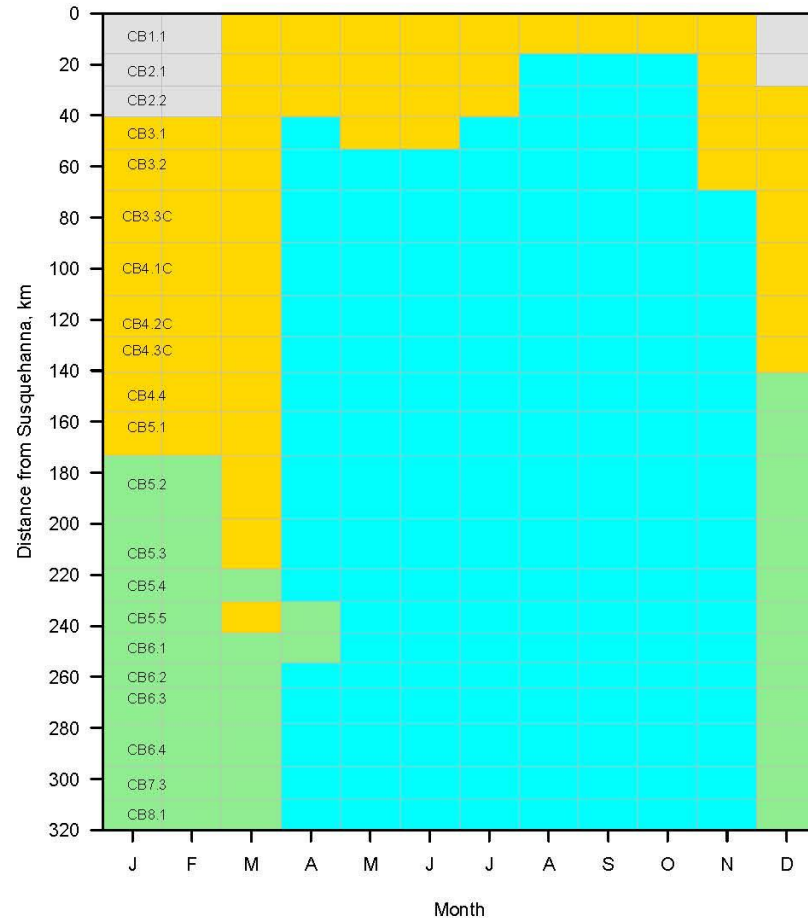   ❖ E3
   ❖ 1985 Progress
   ❖ All forest

No Action

Mainstem Chesapeake Bay (1991–2000)
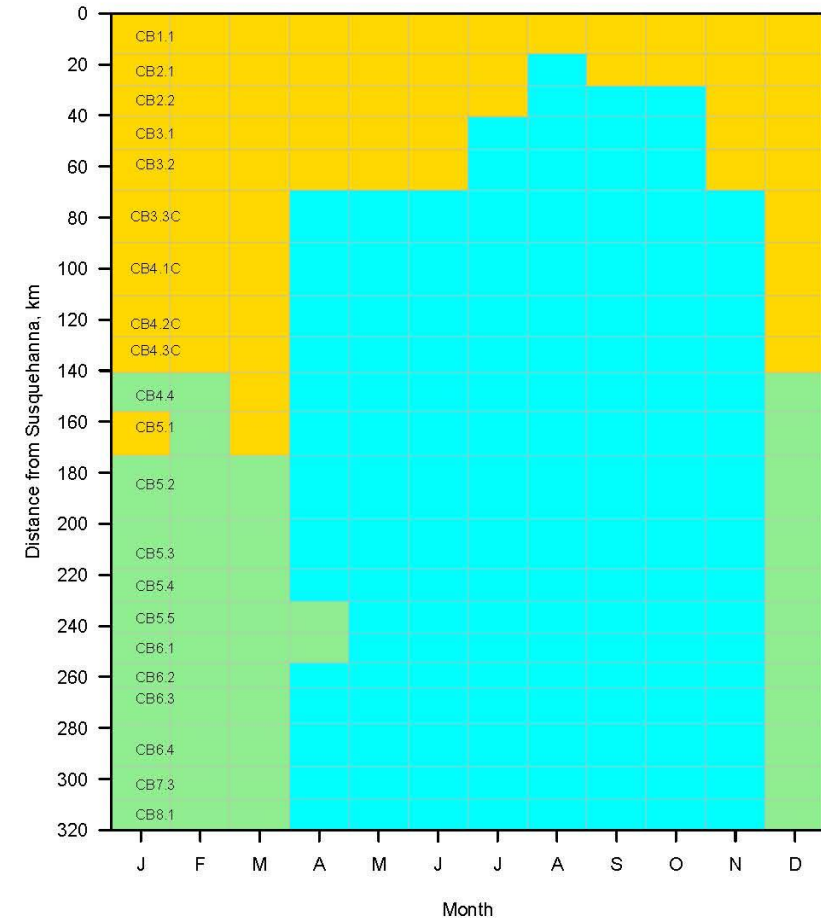**Modeling NoAct Data (CART; 21 Stations)**

WIP3

Mainstem Chesapeake Bay (1991–2000)
Modeling WIP3 Data (CART; 21 Stations)

E3

Mainstem Chesapeake Bay (1991–2000)
Modeling E3 Data (CART; 21 Stations)

No NoR
More NP

Preliminary; subject to change.

Next steps on analyzing the modeling data
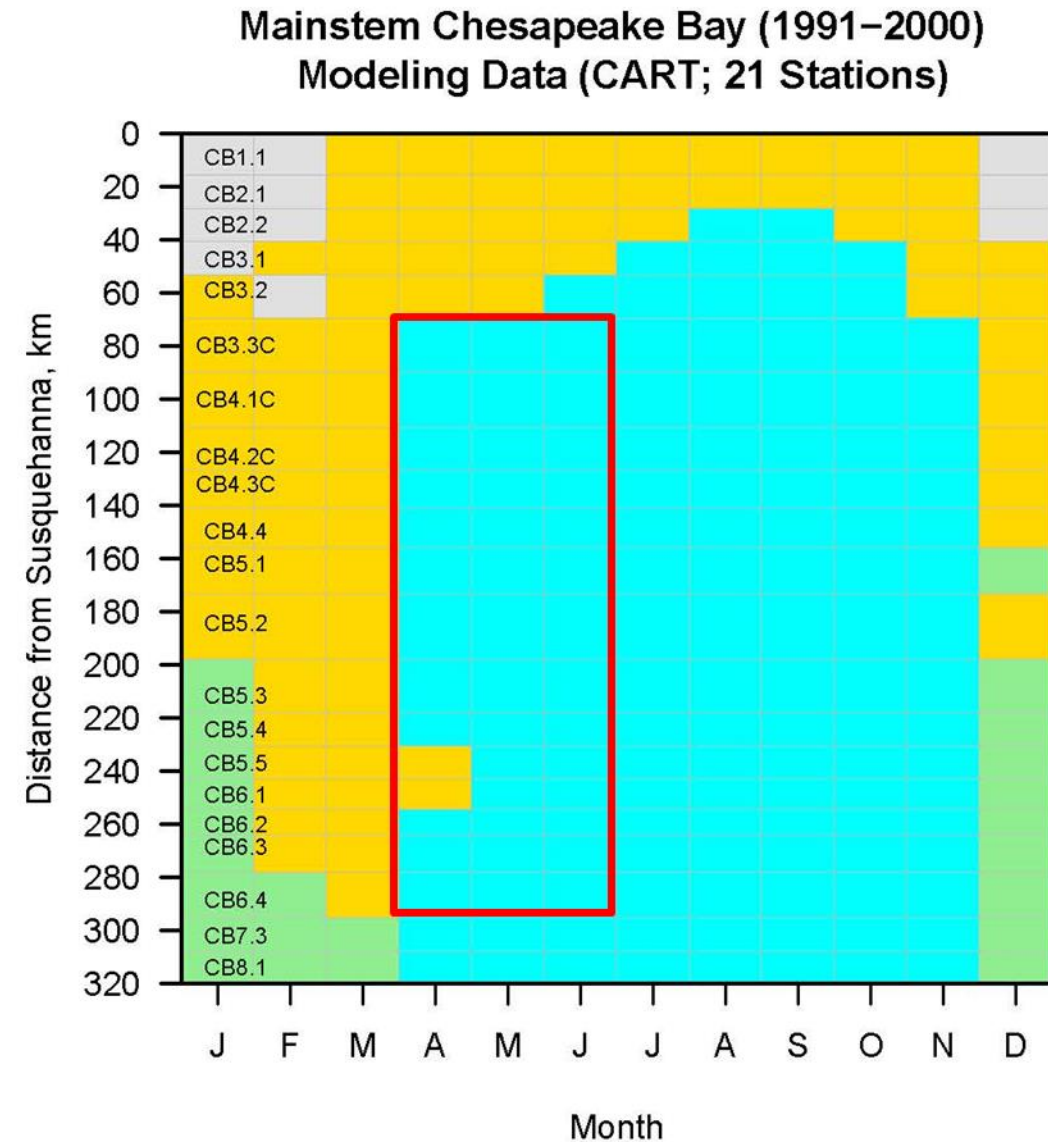

Mainstem Chesapeake Bay (1991–2000)
Monitoring Data (CART; 21 Stations)
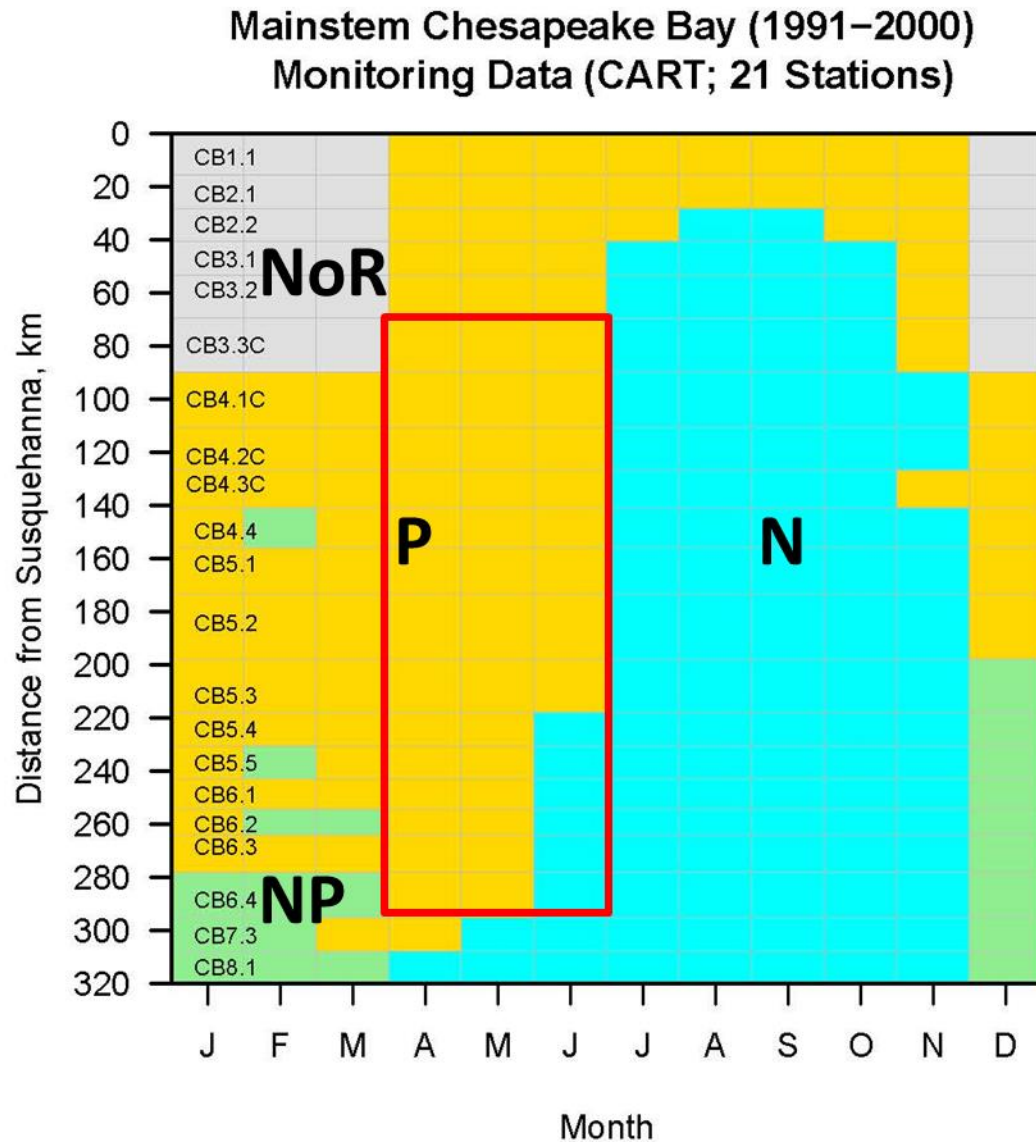
Mainstem Chesapeake Bay (1991–2000)
Modeling Data (CART; 21 Stations)

17

Next steps on analyzing the modeling data

- What are the likelihoods of the predicted nutrient limitation classes?

- Why were TN:TP ratios so different between the monitoring & modeling data?

- If DIN:DIP is used instead of TN:TP, would the observed differences remain?

- How would the limitation patterns of the modeling data vary if only the S layer is considered?

- How would the limitation patterns of the modeling data vary if only a smaller period is considered?

- How would the limitation patterns of the modeling data compare between 1990s and 2010s?

Remarks

- With limited or no bioassays, CART provides complementary information on nutrient limitation based on WQ monitoring data, greatly expanding the spatial & temporal extent of assessments to guide management.
  - An underlying assumption of CART is the stationarity in the derived model relationships (including decision rules).
  - The CART predictions should be viewed as the overall behavior for the specified periods.
- New bioassay experiments can be particularly useful for validating and updating the CART models.

#BayStrong