CalCAST Updates

Isabella Bertani, Gopal Bhatt, Lewis Linker, and the Modeling Team

Modeling Workgroup Quarterly Review

10/17/2023

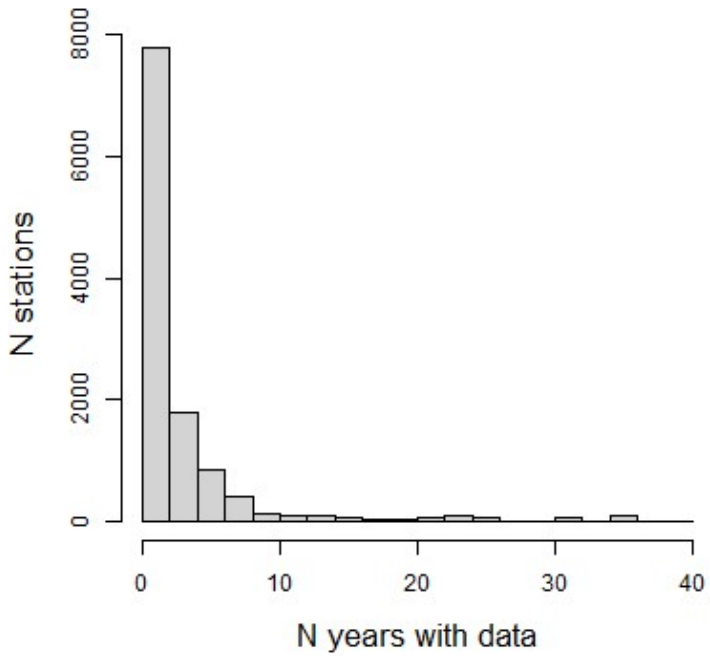
Developing a reproducible workflow to download water quality data from EPA's Water Quality Portal (WQP)

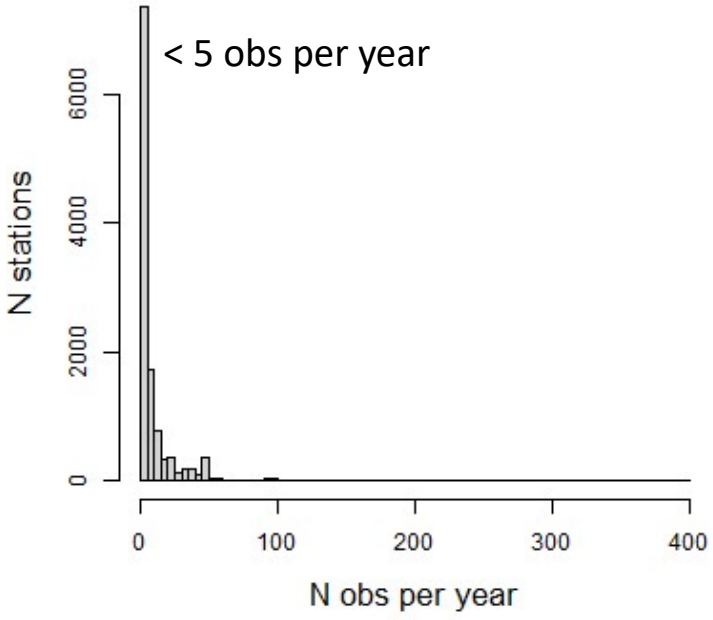
JUNE QUARTERLY RECAP

Motivation

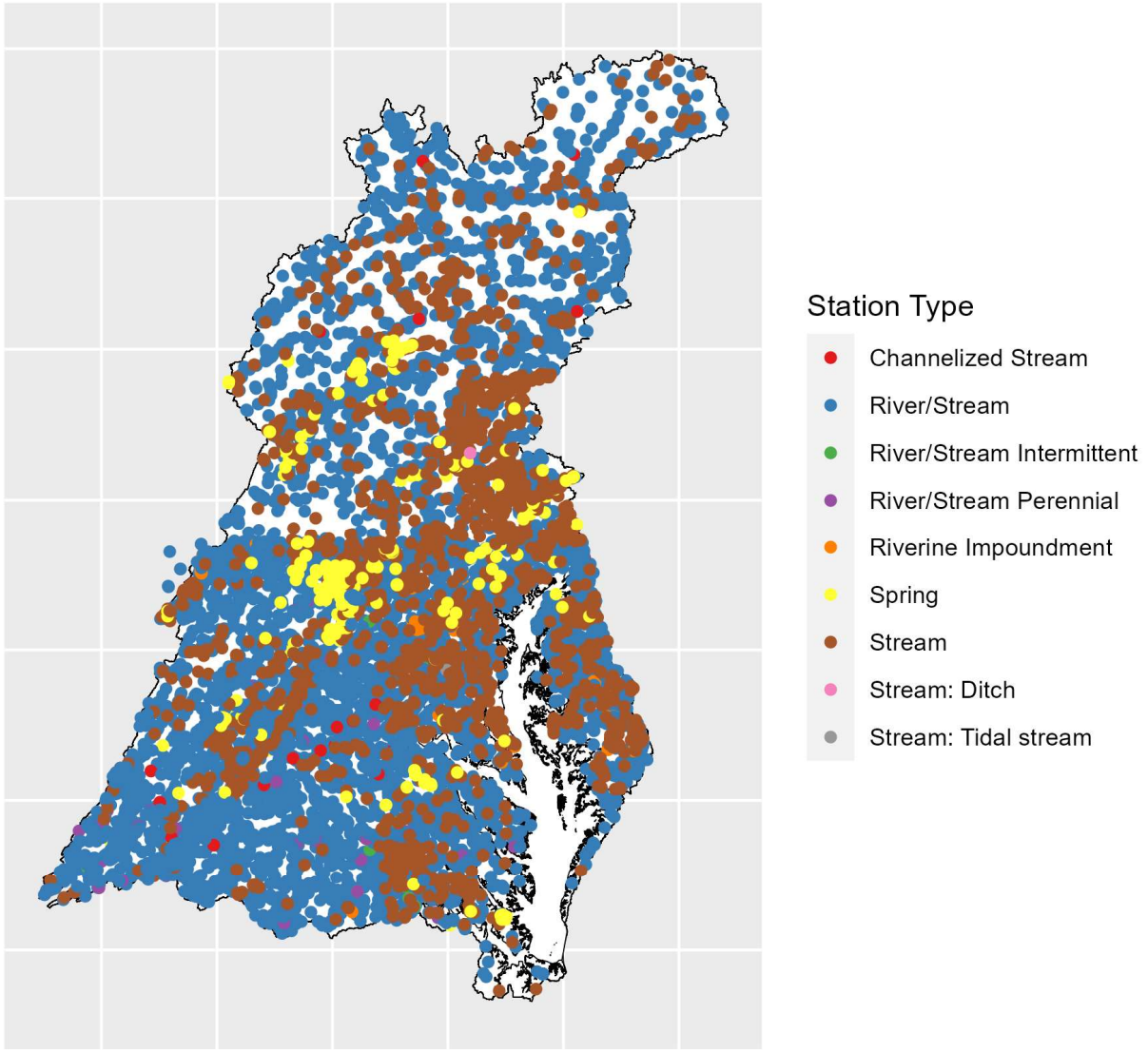
- We need to update the watershed model calibration dataset for P7
- The P6 calibration dataset was obtained through an ad-hoc process that is not easily reproducible
- We would like to develop an automated workflow that allows us to regularly update and expand our calibration dataset without major efforts
- March 2023: STAC Workshop "Using Local Monitoring Results to Inform CBP's Watershed model":
  - It is very important to local monitoring agencies that their stations are used to calibrate the CBP watershed model if possible
  - Need for a transparent and streamlined process to submit monitoring data for use in watershed model calibration

Monitoring stations with N data in the WQP





~11000 stations in Bay watershed



Three purposes for WQP dataset

Each purpose has different data requirements and different criteria for station/data inclusion

- **Dynamic watershed model** calibration and verification
    - Raw constituent concentrations (for calibration)
    - Load estimation (for verification)

- **CalCAST** calibration
    - Load estimation

- **Other forthcoming applications** (e.g., Machine Learning RFP)
    - Not sure yet, retain as much data as possible (while flagging/fixing issues)

Workflow

1. Raw WQP data processing/cleanup
2. Raw USGS streamflow data processing/cleanup
3. WQP station – USGS streamflow gage matching for load estimation
4. Application of screening criteria for load estimation

References we heavily relied on

1. **Oelsner et al., 2017** - **Water-quality trends** in the **Nation's rivers and streams**, 1972–2012—**Data preparation**, **statistical methods**, and trend results (ver. 2.0, October 2017): *U.S. Geological Survey Scientific Investigations Report 2017–5006*.

2. **Shoda et al., 2022** - **Water-quality trends** in the **Delaware River Basin** calculated using multisource data and two methods for trend periods ending in 2018: *U.S. Geological Survey Scientific Investigations Report 2022–5097*.

3. **Saad et al., 2019** - Estimates of long-term mean daily streamflow and annual nutrient and suspended-sediment loads considered for use in regional **SPARROW** models of the conterminous United States, 2012 base year: *U.S. Geological Survey Scientific Investigations Report 2019–5069*.

4. **Hirsch and DeCicco, 2015** - User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval—**R packages for hydrologic data** (version 2.0, February 2015): *U.S. Geological Survey Techniques and Methods book 4, chap. A10*.

5. **CBP's Nontidal Network**

Workflow

1. Raw WQP data processing/cleanup
2. Raw USGS streamflow data processing/cleanup
3. WQP station – USGS streamflow gage matching for load estimation
4. Application of screening criteria for load estimation

Raw WQP data processing/cleanup

- Flag/remove samples in wrong **media** (e.g., «biological tissue» or «bed sediment»)
- Standardize/flag **sample fraction** (e.g., «dissolved» vs. «total» vs. «suspended»)
- Standardize/flag **sample type** (e.g., «lab replicate», «reagent blank», «field obs»)
- Standardize/flag **censoring** information (e.g., «Present Below Quantification Limit»)
- Standardize/flag **qualifying issues** that affect results (e.g., «No Result, lab accident»)
- Standardize **parameter names** and assign appropriate USGSPCode
- Standardize/fix **units of measure**
- Resolve **duplicate** samples
- Flag «**suspicious»** values
- Aggregate constituents to obtain **surrogate TN** measurements

Workflow

1. Raw WQP data processing/cleanup
2. Raw USGS streamflow data processing/cleanup
3. WQP Station – USGS Streamflow gage matching for load estimation
4. Application of screening criteria for load estimation

USGS streamflow data processing/cleanup

1. Retrieved daily flow data for 1984-present at >500 USGS streamflow gages originally assessed for inclusion in DM/CalCAST calibration dataset

Calculation of loads requires continuous record of daily flow (no NAs, zeros, or negatives). Following Oelsner et al. 2017 and Shoda et al. 2022:

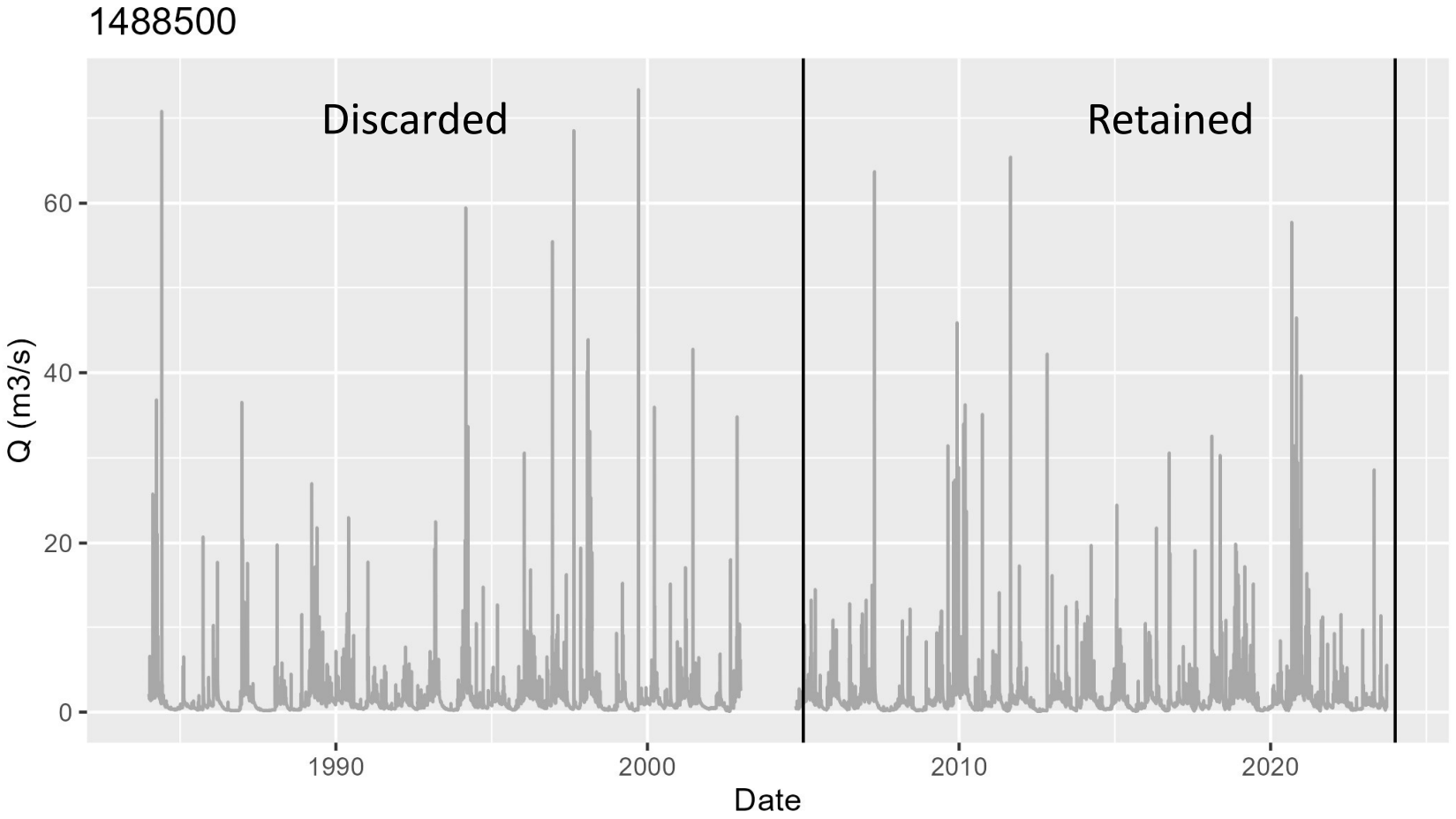
2. Each streamflow record was required to meet the following criteria:
   - No more than 30 days with **missing streamflow** per year
   - No more than 3 consecutive days with **missing streamflow** per year
   - No days with **negative streamflow**
   - No more than 30 days with **zero streamflow** per year

3. Added small constant to zero streamflow values

4. Filled in missing streamflow using the R function “fillMiss” from the R package “waterData” (Ryberg and Vecchia, 2012)

USGS streamflow data processing/cleanup

In some cases, a streamflow record failed to meet all criteria; whenever possible, we deleted parts of the period of record that did not meet criteria and retained the longest period of record that met criteria.



Workflow

1. Raw WQP data processing/cleanup
2. Raw USGS streamflow data processing/cleanup
3. WQP Station – USGS Streamflow gage matching for load estimation
4. Application of screening criteria for load estimation

WQP Station – USGS Streamflow gage matching

Calculation of loads requires to match streamflow gages and water quality stations:

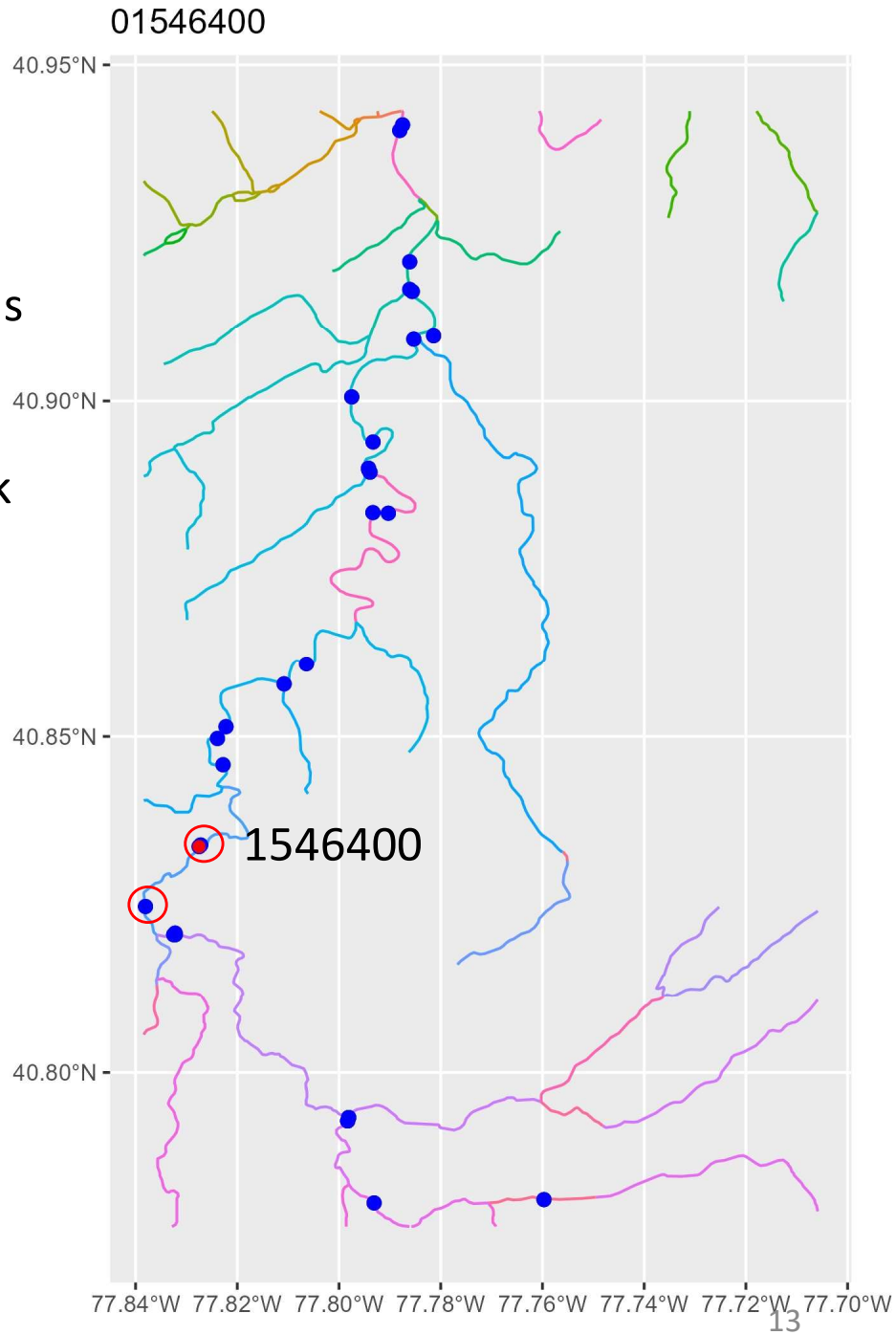
1. Assign NHDPlus Flowline (ComID) to WQP stations and some USGS streamflow gages (most USGS streamflow gages already have ComID information) – Combination of NLDI and GIS work

2. For each USGS streamflow gage, identify WQP stations on same LevelPathID (Oelsner et al., 2017)

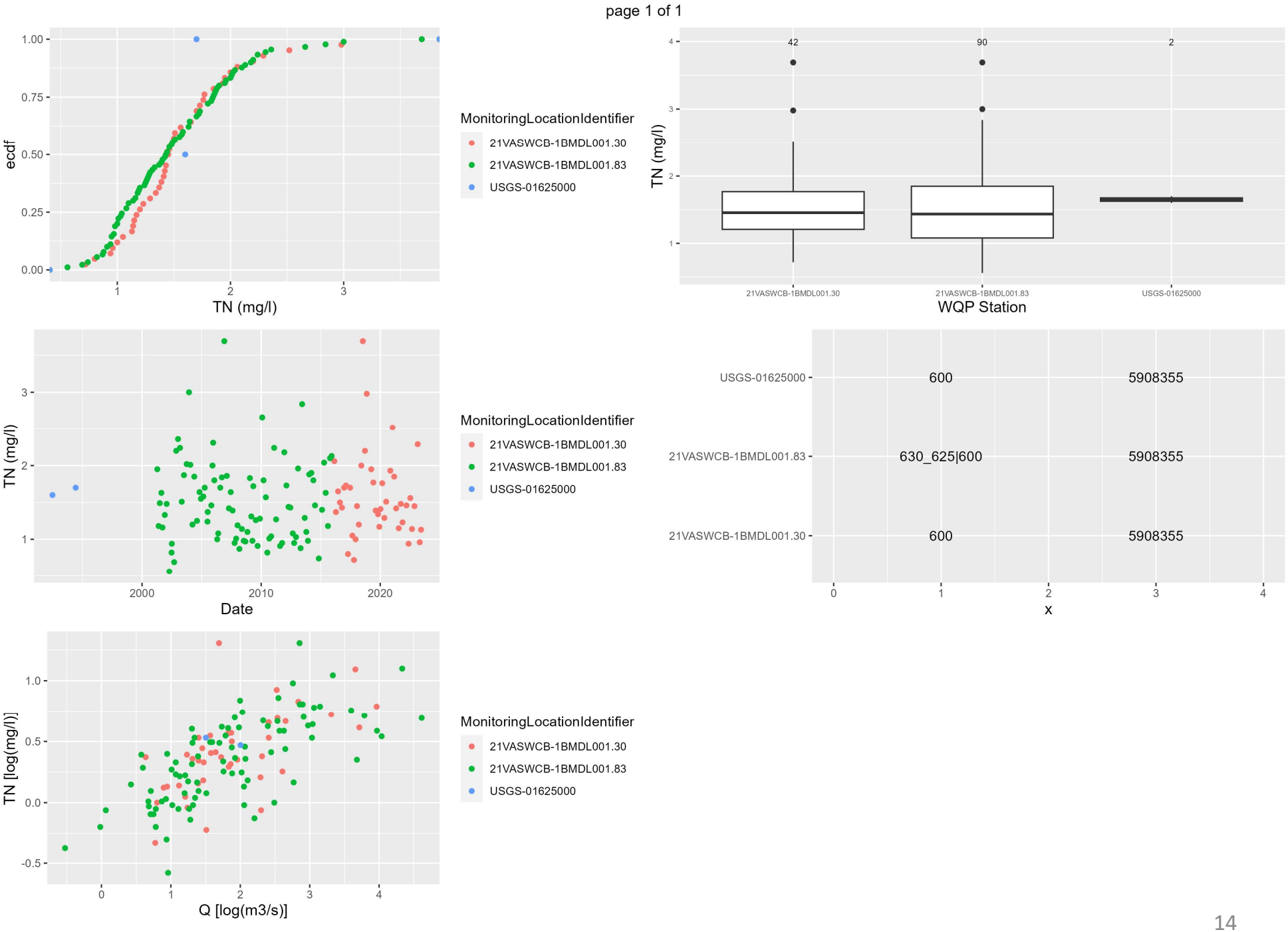
3. Retain only WQP stations within ~2 km (linear distance) of each streamflow gage and without intervening tributaries

4. Double check that we obtain same matches as NTN sites – Verified

5. Visually assess the combined water-quality data

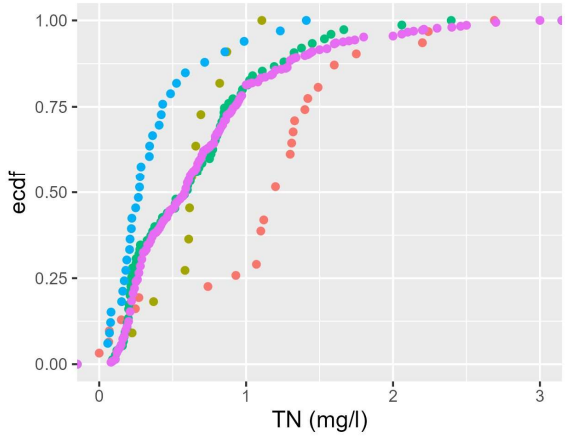


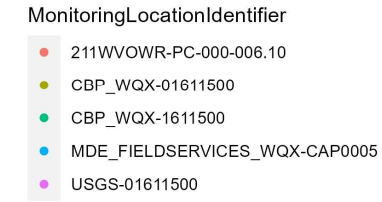
WQP Station – USGS Streamflow gage matching

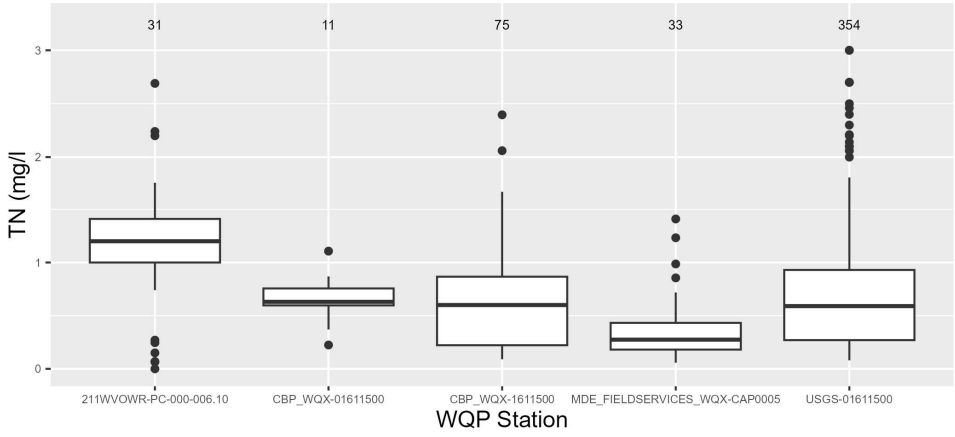


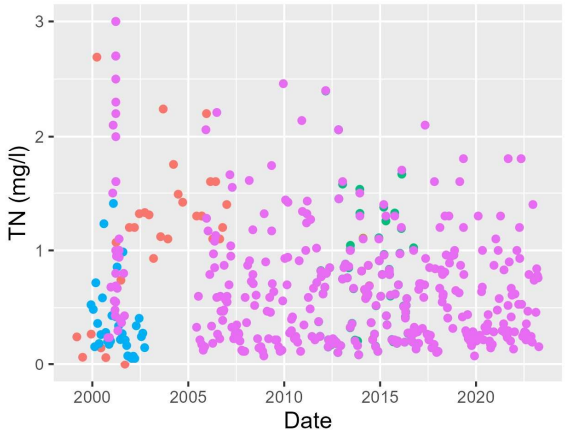
WQP Station – USGS Streamflow gage matching

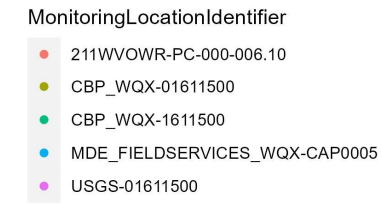
page 1 of 1

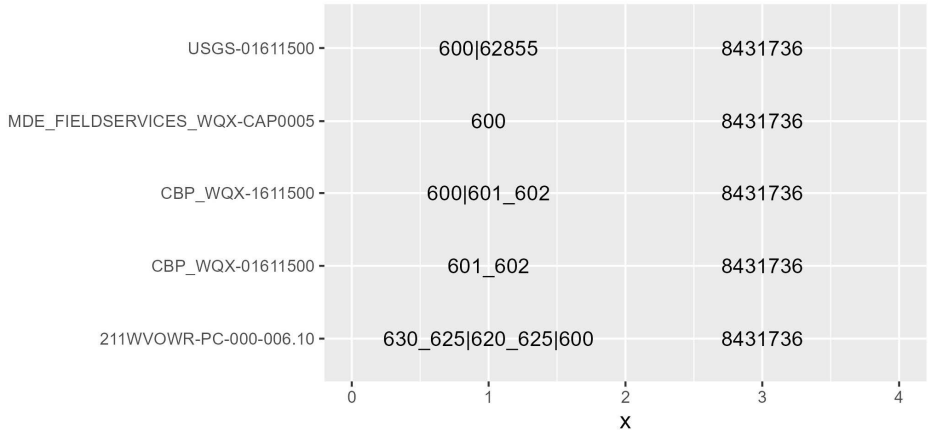


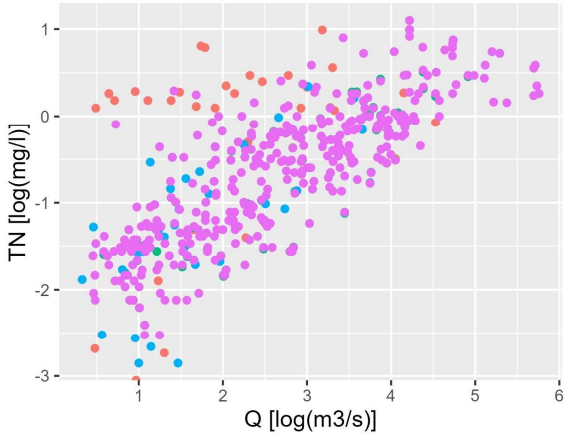


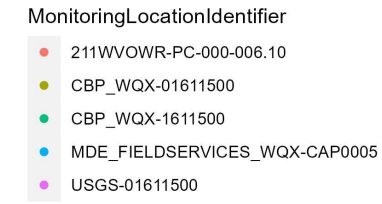


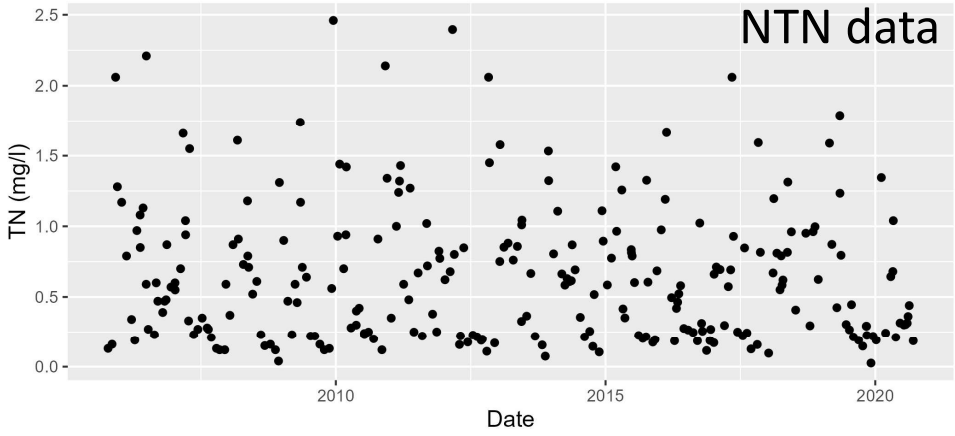












Workflow

1. Raw WQP data processing/cleanup
2. Raw USGS streamflow data processing/cleanup
3. WQP Station – USGS Streamflow gage matching for load estimation
4. Application of screening criteria for load estimation

Application of screening criteria for load estimation

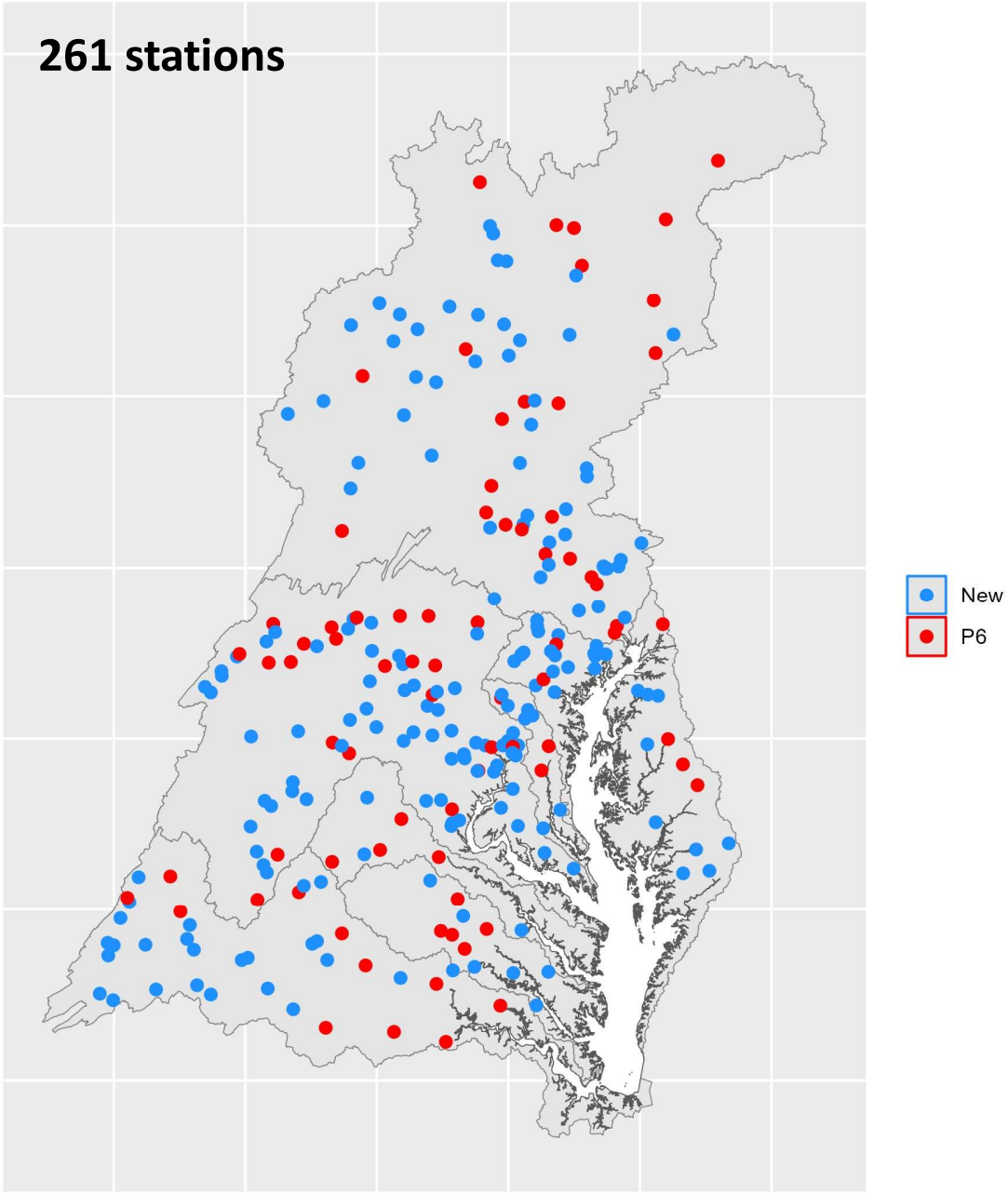
Screening criteria to estimate loads using **FluxMaster** (Saad et al., 2019/SPARROW):

a) At least 2 years of overlap between water quality and streamflow data
b) Minimum of 5 years of daily streamflow (originally 10 years)
c) Minimum of 3 years of water quality data
d) At least 24 samples
e) At least 3 samples in each season

Application of screening criteria for load estimation

Screening criteria to estimate loads using **FluxMaster** (Saad et al., 2019/SPARROW):

a) At least 2 years of overlap between water quality and streamflow data
b) Minimum of 5 years of daily streamflow (originally 10 years)
c) Minimum of 3 years of water quality data
d) At least 24 samples
e) At least 3 samples in each season



Application of screening criteria for load estimation

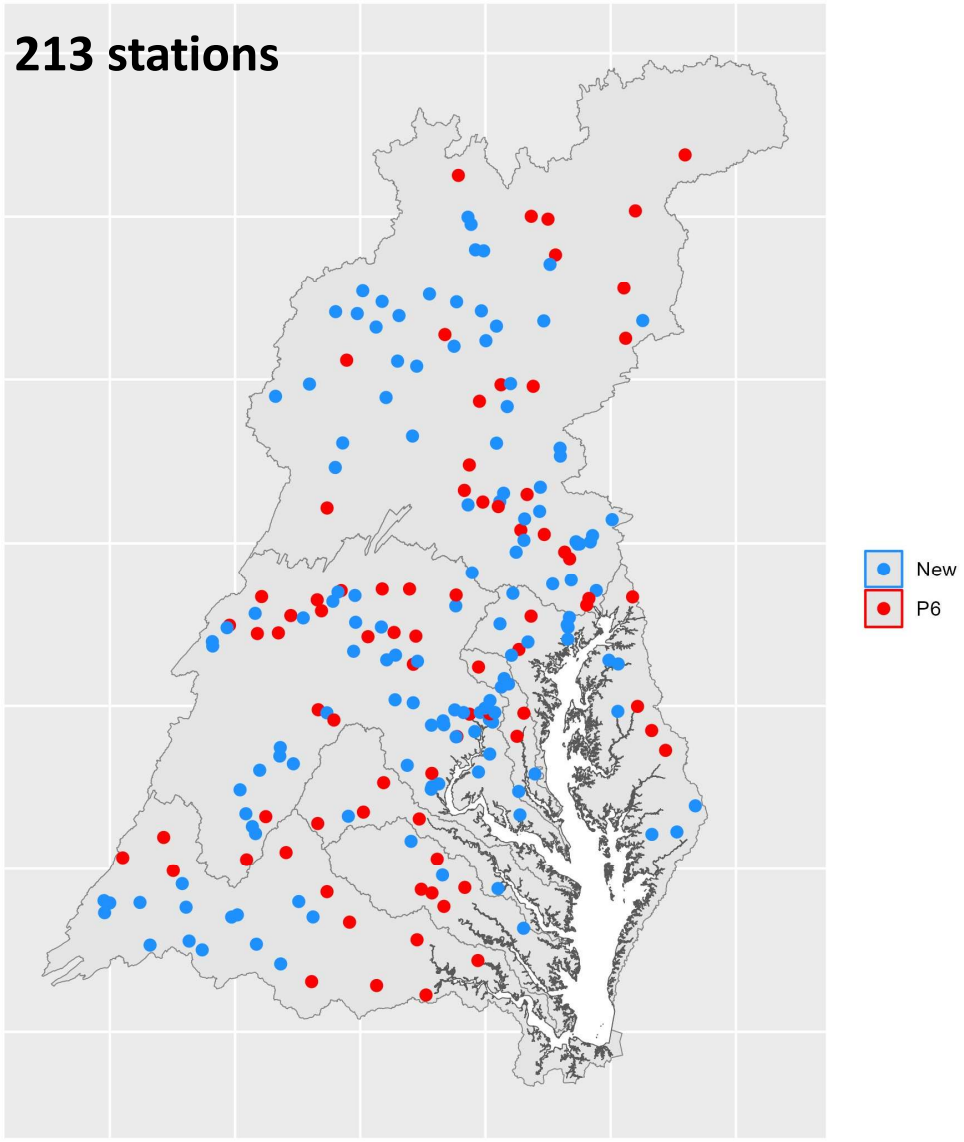
Screening criteria to estimate loads using **WRTDS** (Oelsner et al., 2017; Shoda et al., 2022; Hirsch and DeCicco, 2015):

a) At least 3 years of water quality (WQ) data
b) If one or more gaps of >2 years in WQ data is present, retain longest period of record without such gaps (gaps <3 years are allowed)
c) Coverage of seasons: 70% of years must have at least quarterly samples (relaxed criterion to retain stations with > 60 samples)
d) At least 3 consecutive years with quarterly WQ samples (relaxed criterion to retain stations with > 60 samples)
e) Coverage of streamflow: At least half of the decades must have at least 14% high-flow samples, or 20 high-flow samples, and the other decades must have at least 10% high-flow samples
f) Minimum 5 years of daily streamflow (originally 10 years)

Application of screening criteria for load estimation

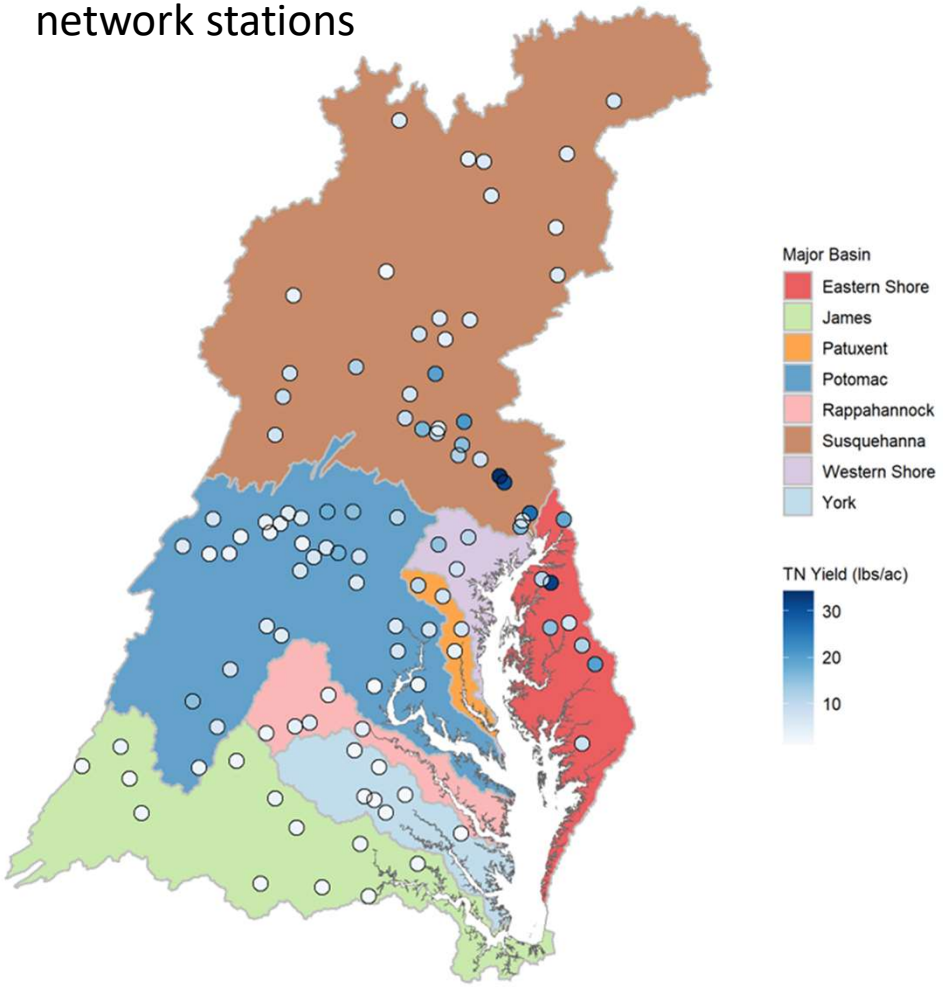
Screening criteria to estimate loads using **WRTDS** (Oelsner et al., 2017; Shoda et al., 2022; Hirsch and DeCicco, 2015):

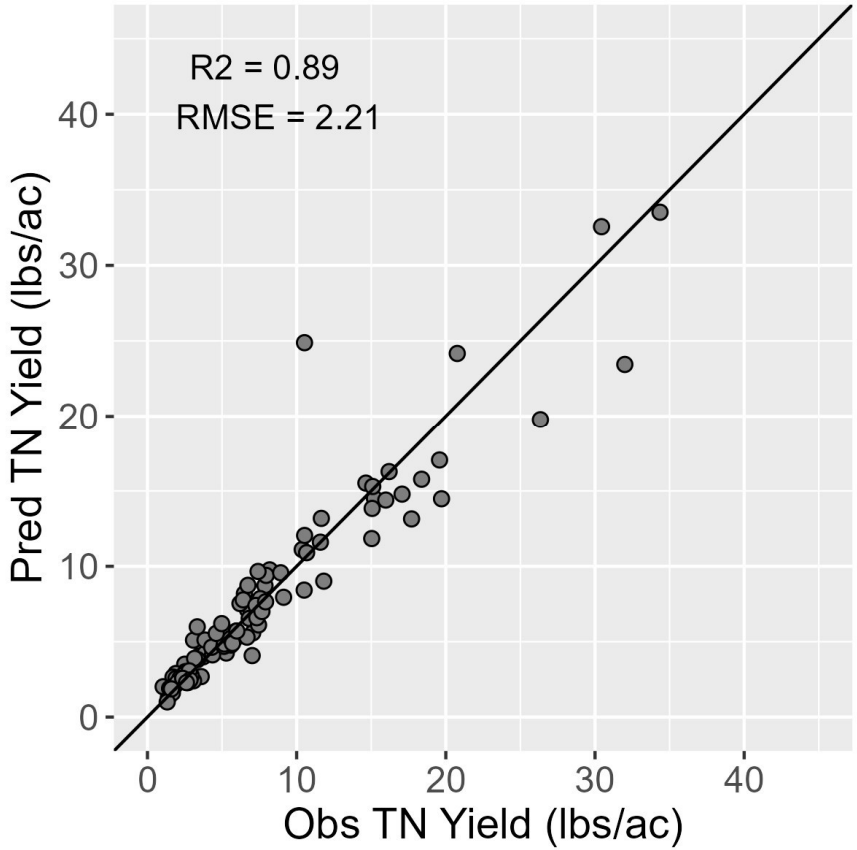
a) At least 3 years of water quality (WQ) data
b) If one or more gaps of >2 years in WQ data is present, retain longest period of record without such gaps (gaps <3 years are allowed)
c) Coverage of seasons: 70% of years must have at least quarterly samples (relaxed criterion to retain stations with > 60 samples)
d) At least 3 consecutive years with quarterly WQ samples (relaxed criterion to retain stations with > 60 samples)
e) Coverage of streamflow: At least half of the decades must have at least 14% high-flow samples, or 20 high-flow samples, and the other decades must have at least 10% high-flow samples
f) Minimum 5 years of daily streamflow (originally 10 years)



CalCAST average annual TN predictions at current calibration stations

- **Calibration target**: WRTDS flow-normalized TN load at **123** non-tidal network stations





Land to water factors

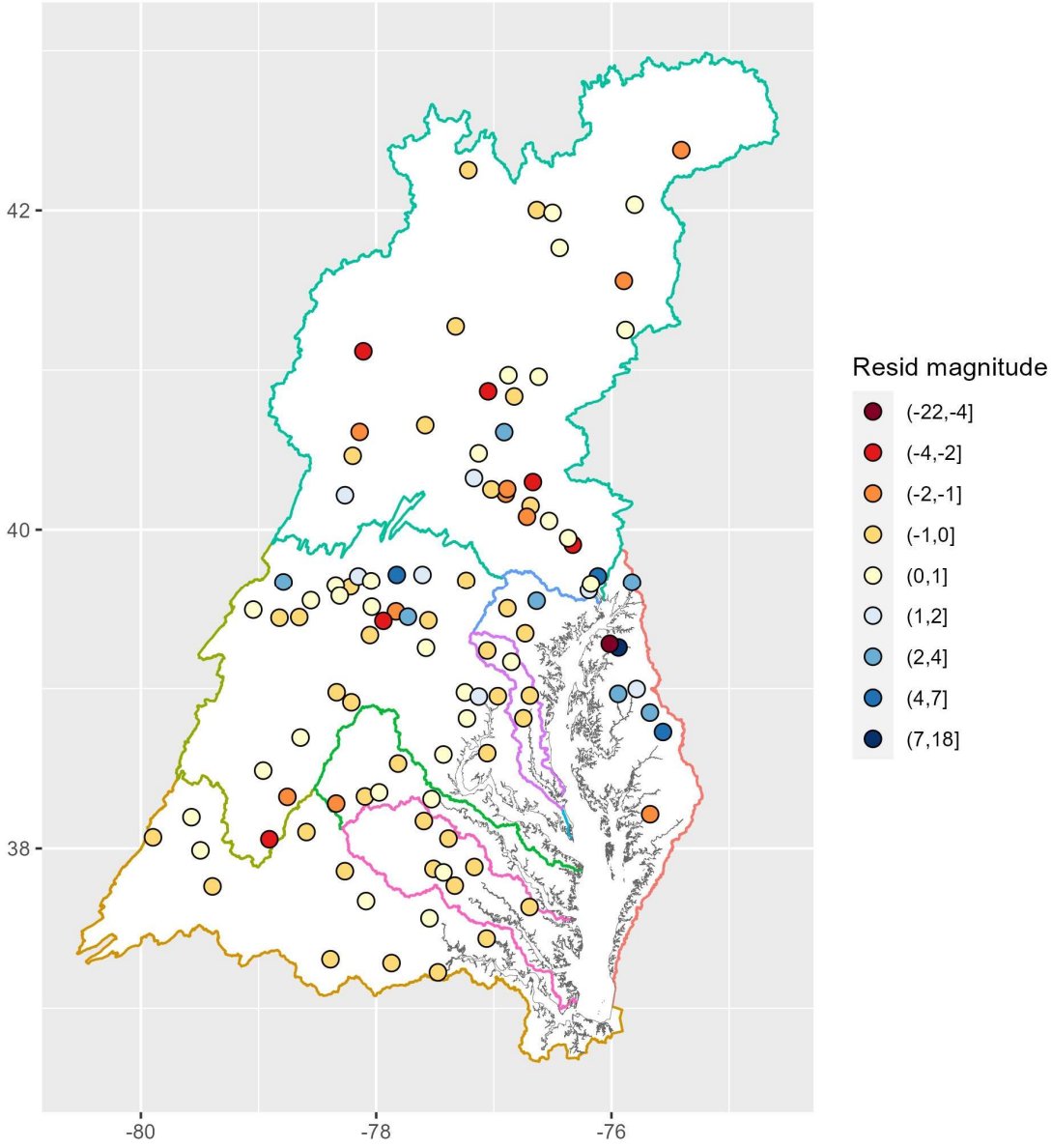
Groundwater recharge (mm)

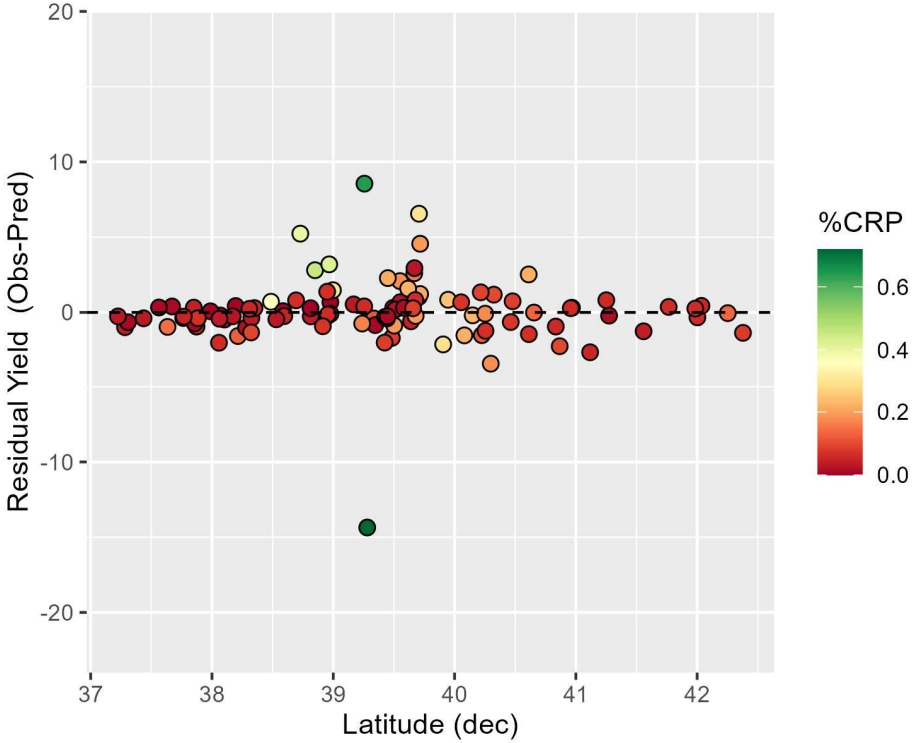
Hydrogeomorphic unit: Coastal Plain Upland (%)

Hydrogeomorphic unit: Piedmont Carbonate (%)

CalCAST average annual TN predictions at current calibration stations

Map of residual yield (Obs – Pred)





Next Steps

- Finalize data processing workflow
- Calculate constituent loads and compare to P6/NTN/SPARROW wherever possible
- Ask Monitoring Agencies to provide feedback and fill in data gaps
- Update CalCAST and DM calibration datasets
- Re-do data pull and processing in 2025