CalCAST Updates

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

Modeling Workgroup Quarterly Review

06/20/2023

Today's updates

1. Developing a reproducible workflow to download water quality data from EPA's Water Quality Portal for watershed model calibration

2. Testing watershed predictors of total flow in CalCAST

Developing a reproducible workflow to download water quality data from EPA's Water Quality Portal for watershed model calibration

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
    - Need for a transparent and streamlined process to submit monitoring data for use in watershed model calibration

Developing a reproducible workflow to download water quality data from EPA's Water Quality Portal for watershed model calibration

Advantages of using EPA's Water Quality Portal (WQP):

- Standardized, mature, and stable infrastructure for data submission, maintenance and retrieval already in place

- It already hosts over 380 million water quality records from over 900 federal, state and tribal agencies

- It brings together EPA's STORET, USGS' NWIS, and USDA's STEWARDS

- USGS has a convenient R package (dataRetrieval; DeCicco et al. 2022) that can be used to develop transparent and reproducible scripts to access and retrieve data from the WQP

Developing a reproducible workflow to download water quality data from EPA's Water Quality Portal for watershed model calibration

Main **objectives** of today's presentation:

- Begin evaluating feasibility of using EPA's WQP as our major tool to retrieve water quality data for watershed model calibration

- Perform an initial exploratory analysis to get an overview of how much and what type of data is available on the WQP

- Compare data retrieved from WQP to P6 calibration dataset

Example: finding monitoring stations with N data in the WQP

Step 1a: Find monitoring stations with water quality parameter(s) of interest

USGS data in the WQP portal can be searched using USGS Parameter Codes

Table 10-4: The aggregation rules used for creating the calibration dataset from water quality observations

Phase 6 observed data	USGS parameter name and code
Total Nitrogen	1. Total nitrogen [P600] 2. Total ammonia [P610] + total nitrate [P620] + total nitrite [P615] + total organic nitrogen [P605] 3. Total nitrite + nitrate [P630] + total kjeldahl [P625] 4. Total nitrate [P620] + total kjeldahl [P625] 5. Dissolved nitrogen [P602] + particulate nitrogen [P601] 6. Total nitrite + nitrate [P630]
Nitrate	1. Dissolved nitrite [P613] + total nitrate [P620] 2. Total nitrite + nitrate [P630] 3. Dissolved nitrite + nitrate [P631] 4. Total nitrate [P620] 5. Dissolved nitrate [P618]
Ammonia	1. Dissolved ammonia [P608] 2. Total ammonia [P610]

Example: finding monitoring stations with N data in the WQP

Step 1b: Find monitoring stations with water quality parameter(s) of interest

Non-USGS data within the WQP portal can be searched using a qualitative «CharacteristicName»

[1] "Ammonia"
[2] "Ammonia-nitrogen"
[3] "Ammonia-nitrogen"
[4] "Ammonia-nitrogen as N"
[5] "Ammonia and ammonium"
[6] "Ammonia as NH3"
[7] "Ammonium"
[8] "Ammonium-nitrogen"
[9] "Ammonium-Nitrogen"
[10] "Ammonium as N"
[11] "Ammonium as NH4"
[12] "Inorganic nitrogen"
[13] "Inorganic nitrogen (nitrate and nitrite)"
[14] "Inorganic nitrogen (nitrate and nitrite) as N"
[15] "Inorganic nitrogen (NO2, NO3, & NH3)"
[16] "Kjeldahl nitrogen"
[17] "Nitrate"
[18] "Nitrate-N"
[19] "Nitrate-Nitrite"
[20] "Nitrate-nitrogen"
[21] "Nitrate-Nitrogen"
[22] "Nitrate + Nitrite"
[23] "Nitrate as N"
[24] "Nitrite"
[25] "Nitrite-N"
[26] "Nitrite as N"
[27] "Nitrogen"
[28] "Nitrogen ion"
[29] "Nitrogen, mixed forms (NH3), (NH4), organic, (NO2) and (NO3)"
[30] "Nutrient-nitrogen"
[31] "Organic Nitrogen"
[32] "Particulate Organic Nitrogen and Particulate Nitrogen"
[33] "Total Kjeldahl nitrogen"
[34] "Total Kjeldahl nitrogen (Organic N & NH3)"
[35] "Total Nitrogen, mixed forms"
[36] "Total Nitrogen, mixed forms (NH3), (NH4), organic, (NO2) and (NO3)"
[37] "Total Particulate Nitrogen"

Example: finding monitoring stations with N data in the WQP

Step 2: Retain only monitoring stations within the Bay watershed

Monitoring stations in the WQP can be filtered geographically in a few different ways:

- Find all stations within a bounding box defined by lat/long
- Find all stations in HUC8 of interest
- Find all stations in state of interest

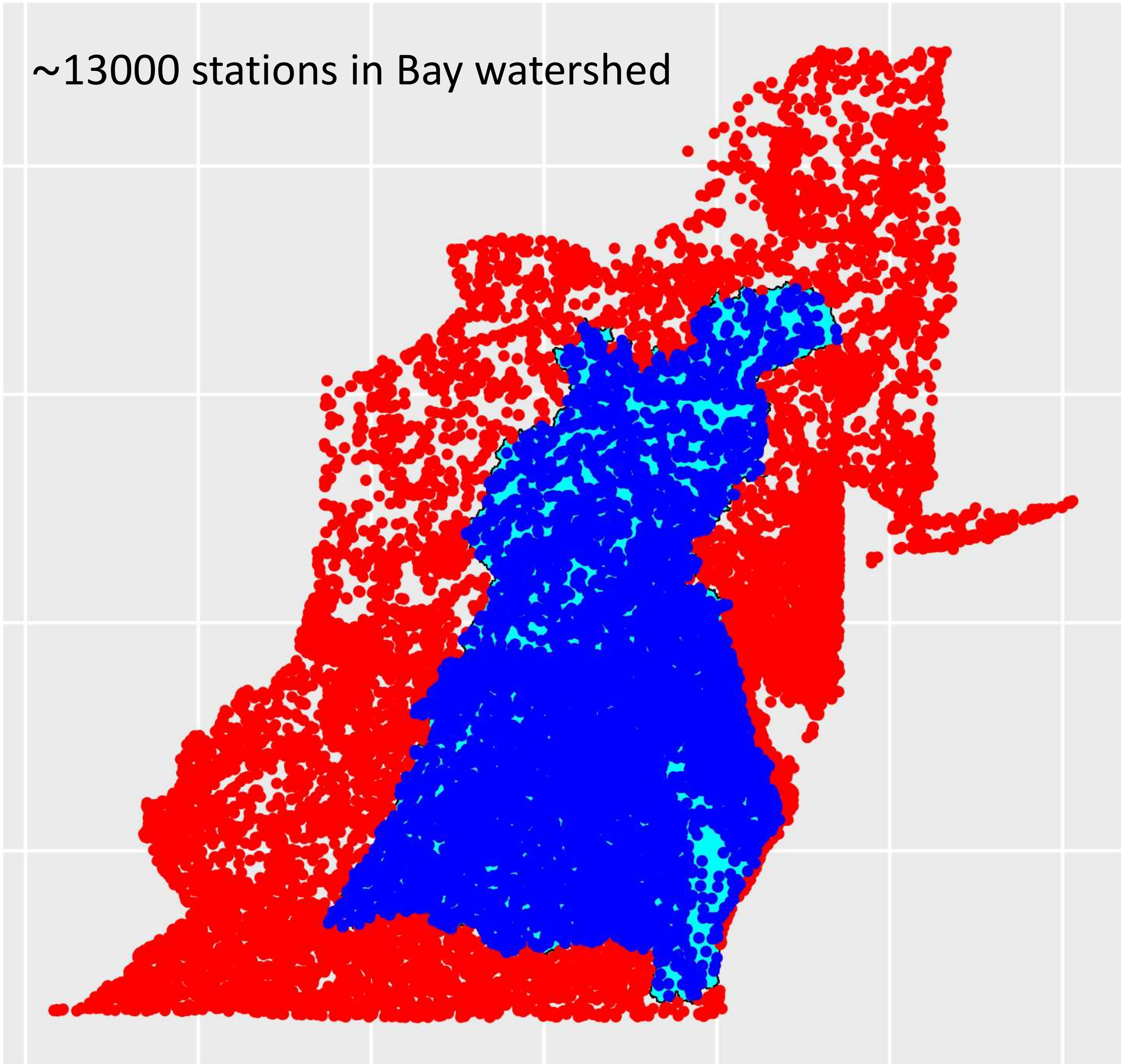
We are using all these methods to minimize the likelihood of missing stations due to errors in data input

Example: finding monitoring stations with N data in the WQP

Step 3: Retain only river/stream monitoring stations

"Channelized Stream"
"River/Stream"
"River/Stream Intermittent"
"River/Stream Perennial"
"Riverine Impoundment"
"Spring"
"Stream"
"Stream: Canal"
"Stream: Ditch"
"Stream: Tidal stream"

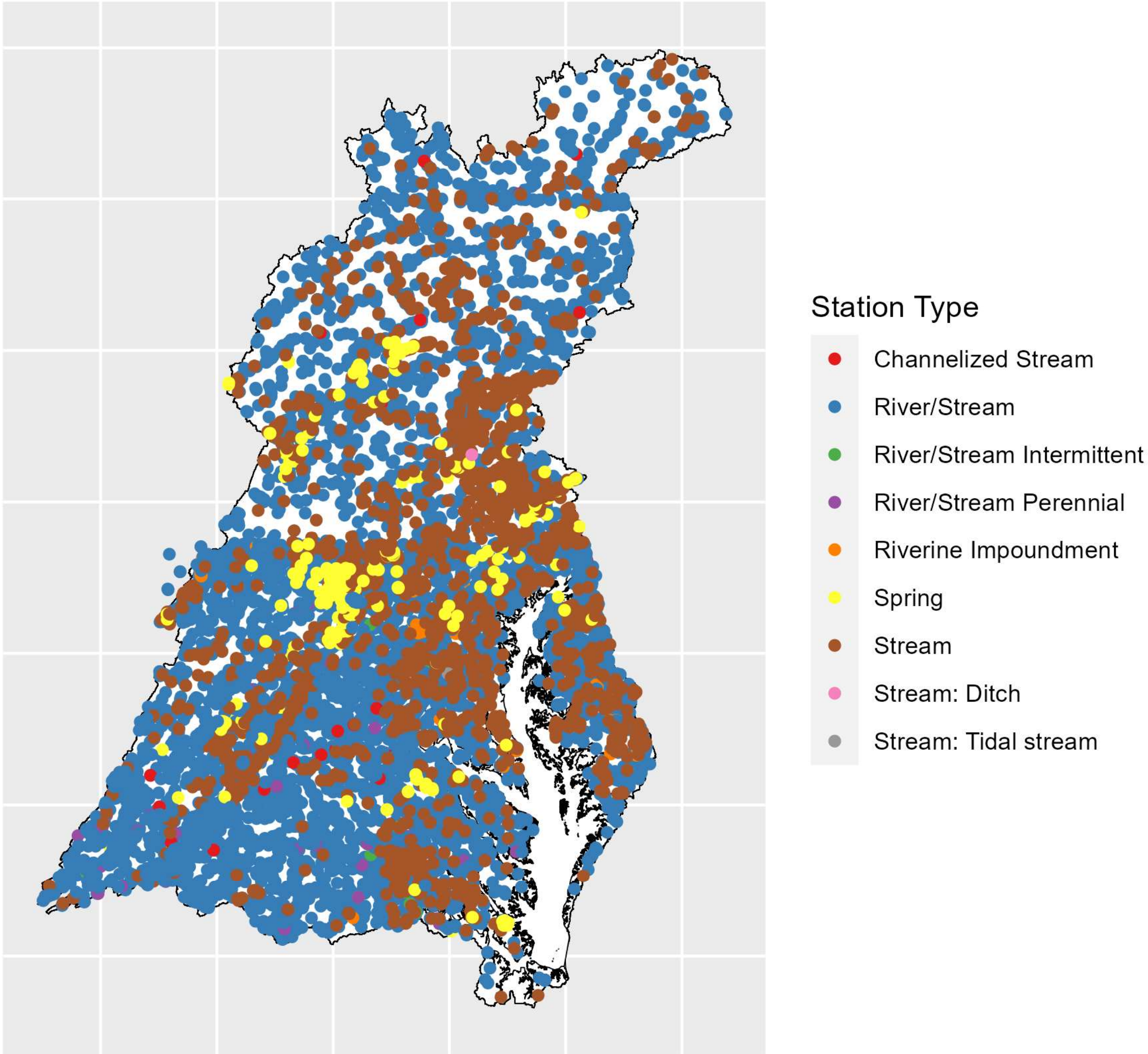
Example: finding monitoring stations with N data in the WQP



Example: finding monitoring stations with N data in the WQP

~11000 stations in Bay watershed after retaining only stations with data for 1984-present and not in the Bay

94% are classified as "Stream" or "River/Stream"

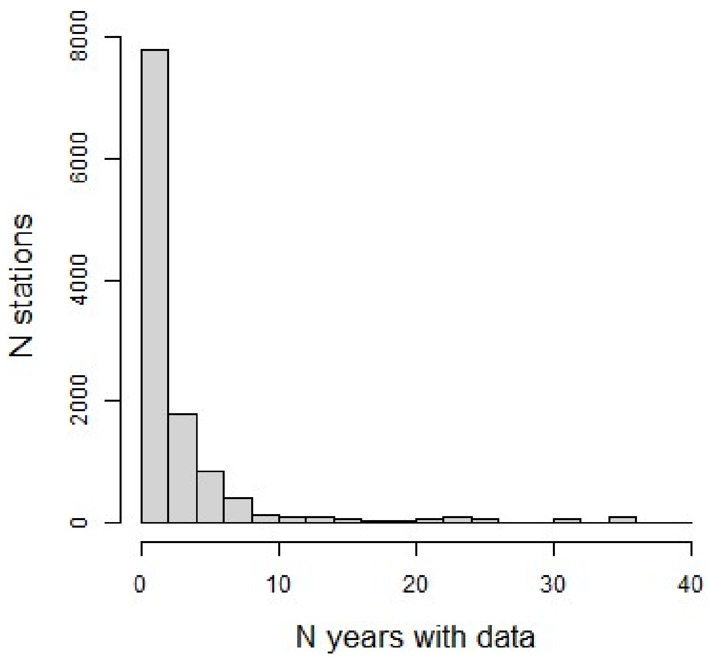


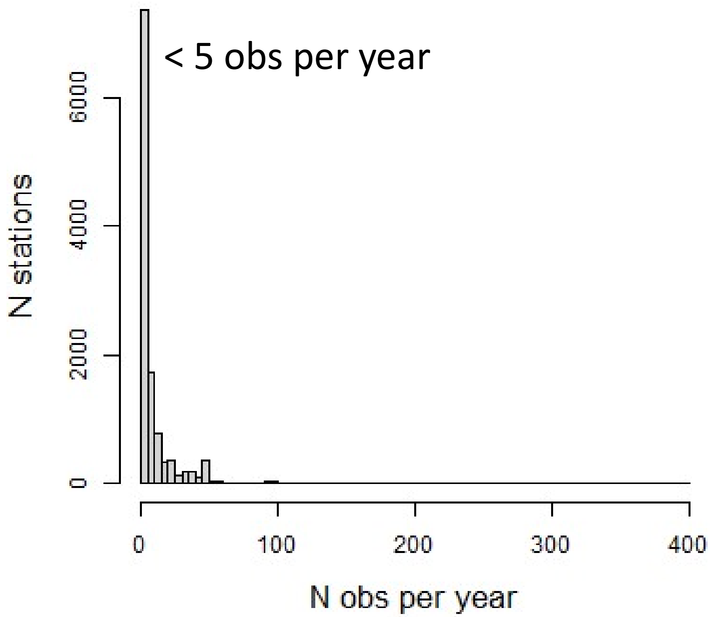
Example: finding monitoring stations with N data in the WQP

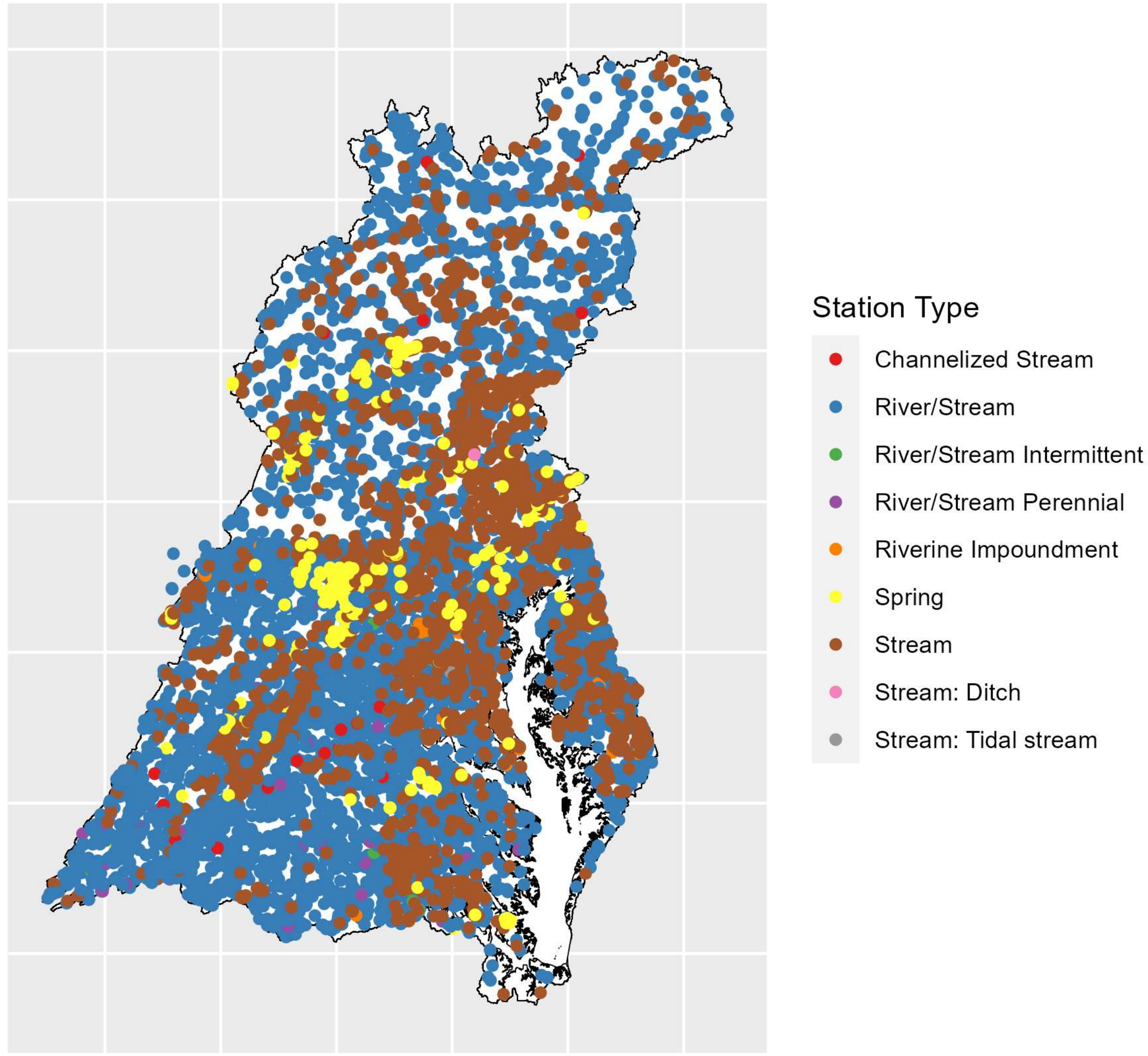
~ 30 reporting organizations

[1] "Antietam National Battlefield"
[2] "Chesapeake Bay Program (CBP)"
[3] "Delaware Department Of Natural Resources And Environmental Control"
[4] "District Of Columbia Department Of Energy And Environment-water Quality Division"
[5] "Environmental Assessment and Standard Program"
[6] "EPA National Aquatic Resource Survey Data"
[7] "EPA National Aquatic Resources Survey (NARS)"
[8] "Friends of Sligo Creek"
[9] "Izaak Walton League of America"
[10] "Keystone Watershed Montioring Network (Pennsylvania)"
[11] "Maryland Department of Natural Resources"
[12] "Maryland Dept. of the Environment In House Water Data"
[13] "Maryland Dept. of the Environment Toxics Data"
[14] "MDE Private Groups/Local Subdivision Data"
[15] "National Park Service Water Resources Division"
[16] "National Wildlife Refuge System, Fish and Wildlife Service"
[17] "New York State Dec Division Of Water"
[18] "NYS Dept. of EnCon, Division of Water"
[19] "PA DEPARTMENT OF ENVIRONMENTAL PROTECTION"
[20] "Potomac Appalachian Trail Club Monitoring - VA,MD"
[21] "Susquehanna River Basin Commission"
[22] "Susquehanna River Basin Commission (Pennsylvania)"
[23] "The Conservation Fund Freshwater Institute (Volunteer)"
[24] "TMDL Technical Development Program"
[25] "University of Virginia Environmental Sciences Department"
[26] "US Army Corps of Engineers, Baltimore District"
[27] "USGS Colorado Water Science Center"
[28] "USGS Maryland Water Science Center"
[29] "USGS New York Water Science Center"
[30] "USGS Pennsylvania Water Science Center"
[31] "USGS Virginia Water Science Center"
[32] "USGS West Virginia Water Science Center"
[33] "VIRGINIA DEPARTMENT OF ENVIRONMENTAL QUALITY"
[34] "Water Quality Protection and Restoration Program"
[35] "West Virginia Department of Environmental Protection Watershed Improvement Branch"
[36] "WV Div of Environmental Protection, Office of Water Resource"

Example: finding monitoring stations with N data in the WQP







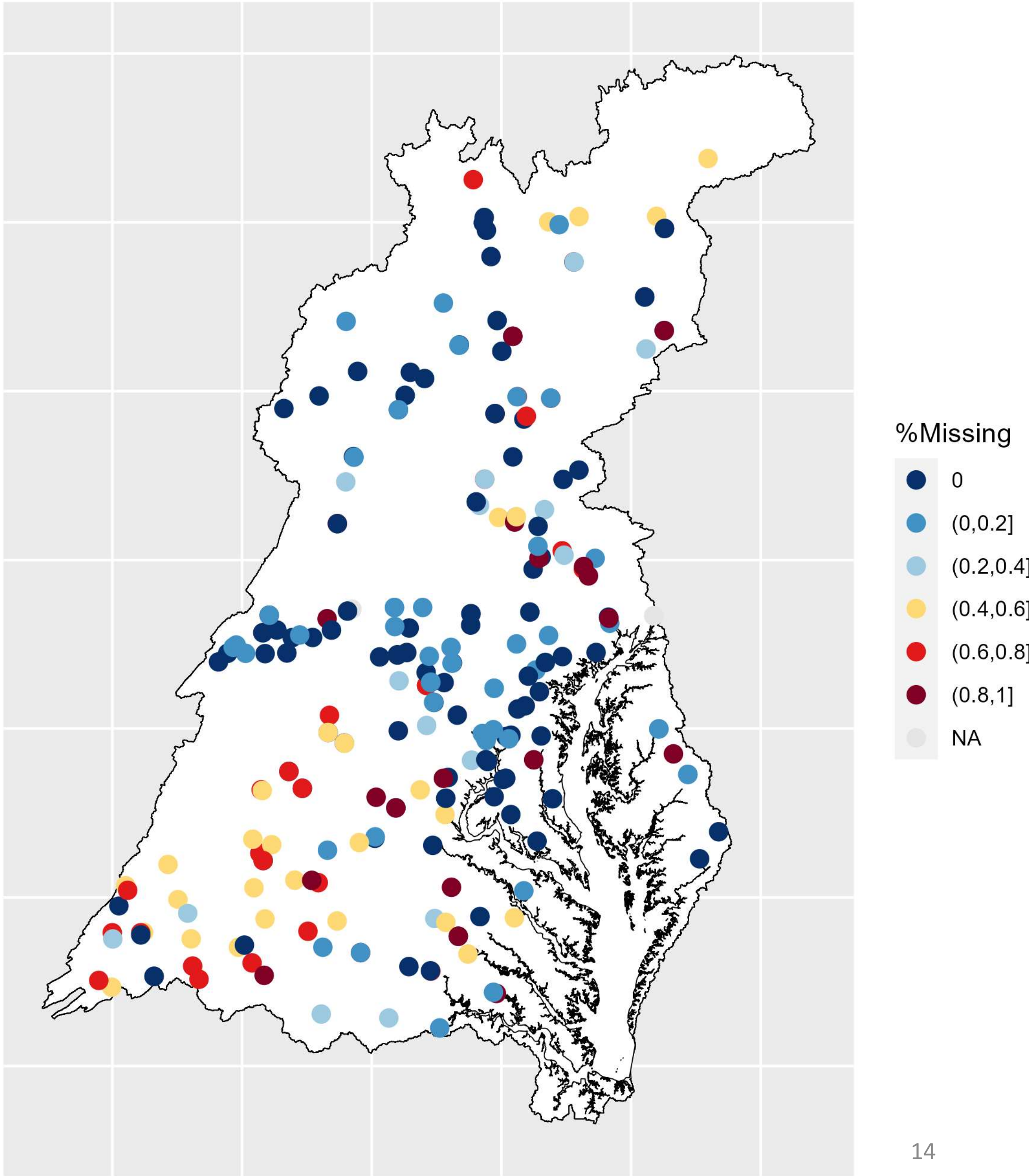
Example: finding monitoring stations with N data in the WQP

26 out of 274 P6 stations not found in the WQP (4 of those available from USGS NTN dataset)

Fraction of P6 sampling dates missing from the WQP

% P6 Dates Missing	% P6 Stations
0%	40%
(0,20%]	23%
(20%,40%]	7%
(40%,60%]	11%
(60%,80%]	10%
(80%,100%]	8%

P6 calibration stations



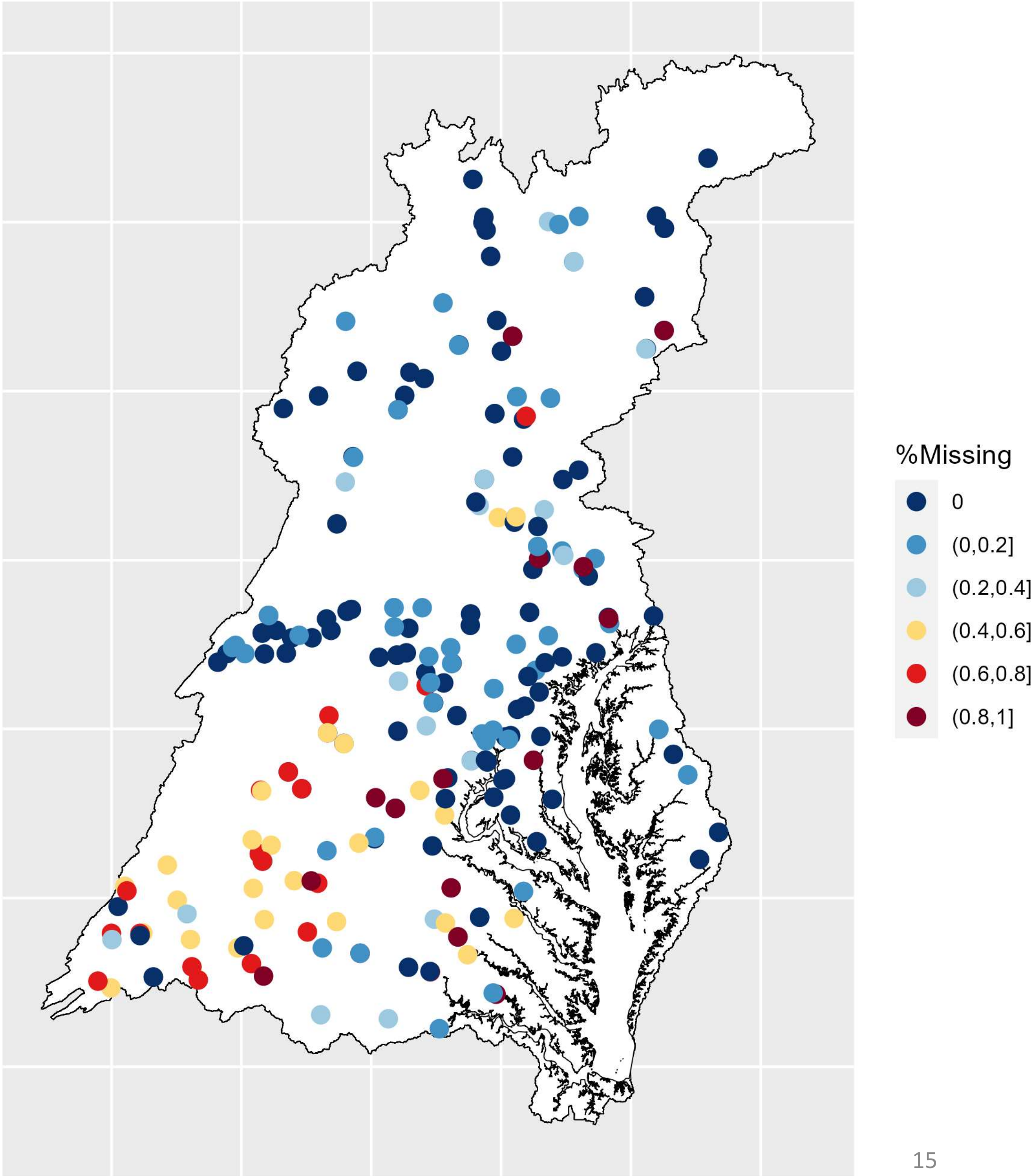
Example: finding monitoring stations with N data in the WQP

26 out of 274 P6 stations not found in the WQP (4 of those available from USGS NTN dataset)

Fraction of P6 sampling dates missing from the WQP + NTN

% P6 Dates Missing	% P6 Stations
0%	46%
(0,20%]	24%
(20%,40%]	7%
(40%,60%]	10%
(60%,80%]	7%
(80%,100%]	6%

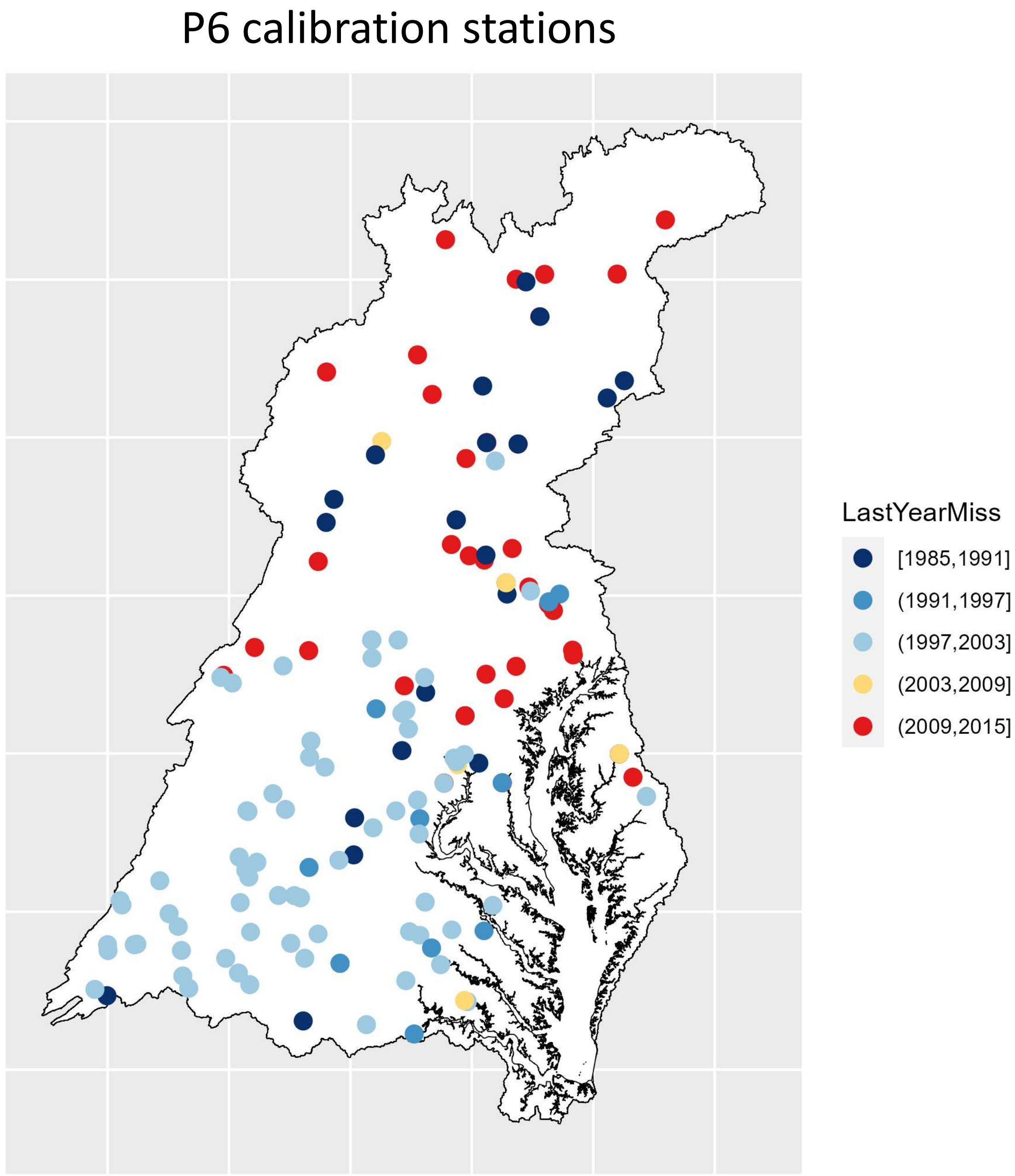
P6 calibration stations



Example: finding monitoring stations with N data in the WQP

Last year with missing sampling dates from WQP

At several stations, most missing data is in the past



Proposed next steps

Continue developing R code to build transparent, reproducible, and automated workflow to retrieve monitoring data from WQP

Work with states and monitoring agencies to make sure that we are as inclusive as possible:

- Provide monitoring organizations with list of stations that we retrieved from WQP so that they can help us fill in the gaps (i.e., stations either fully or partially missing from WQP)

- Maintain flexibility in how we set up the data acquisition process

- Are there better tools for this job that we should consider?

Today's updates

1. Developing a reproducible workflow to download water quality data from EPA's Water Quality Portal for watershed model calibration

2. Testing watershed predictors of total flow in CalCAST

What is CalCAST?

- Relatively parsimonious, spatially explicit, largely data-driven watershed modeling tool calibrated in a statistical framework

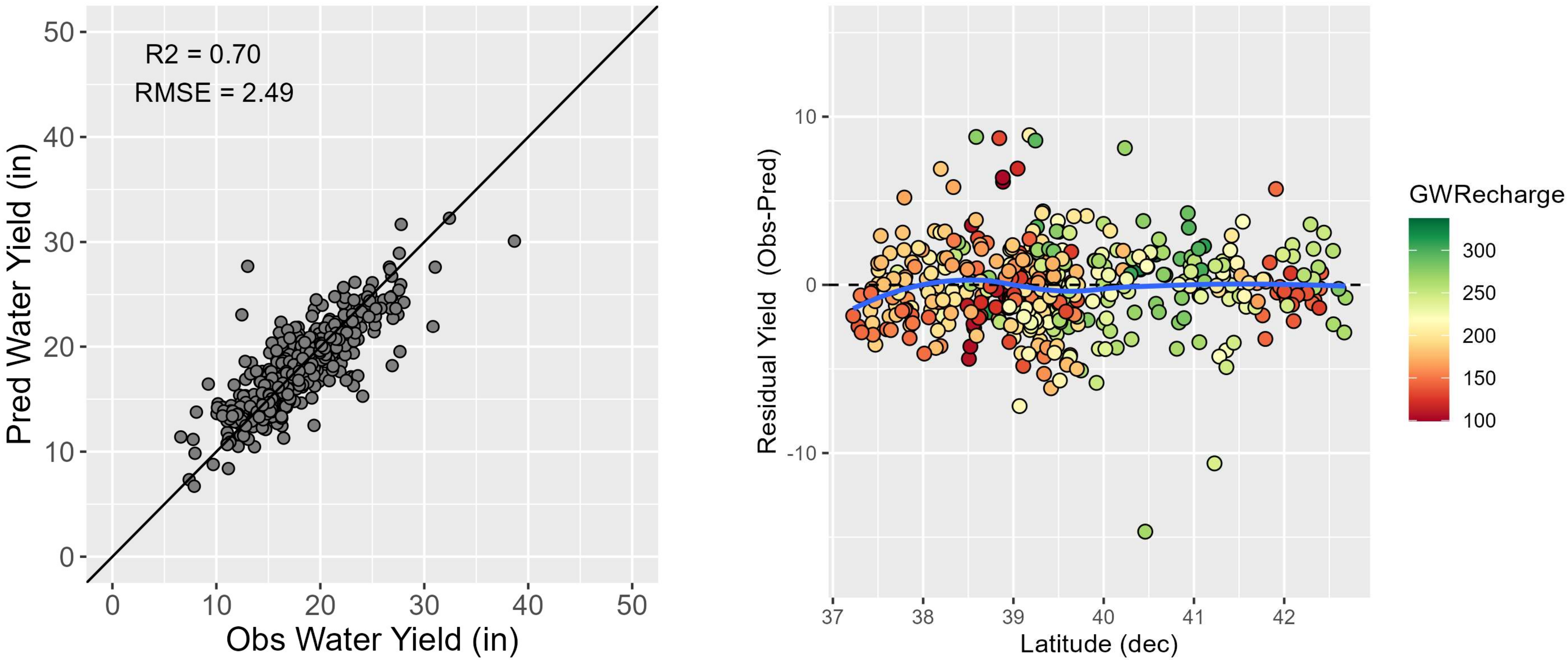
- Represents > 80,000 National Hydrography Dataset Plus (NHDPlus) catchments within the Bay watershed and leverages data from > 400 USGS monitoring stations for calibration

- Predicts long-term average and annual streamflow, %stormflow, sediment, and nutrients at NHDPlus catchments

Why CalCAST?

- Primarily used as spatial calibration tool
- Main purpose: probabilistically test hypotheses on factors related to spatial variation in contaminant loads and quantify parameters that describe such relationships
- Spatial parameters estimated by CalCAST will inform CAST and the dynamic model
- Incorporate data-driven line of evidence into modeling approach

Average Annual Water Yield
PRISM PPT and PRISM Priestley-Taylor PET



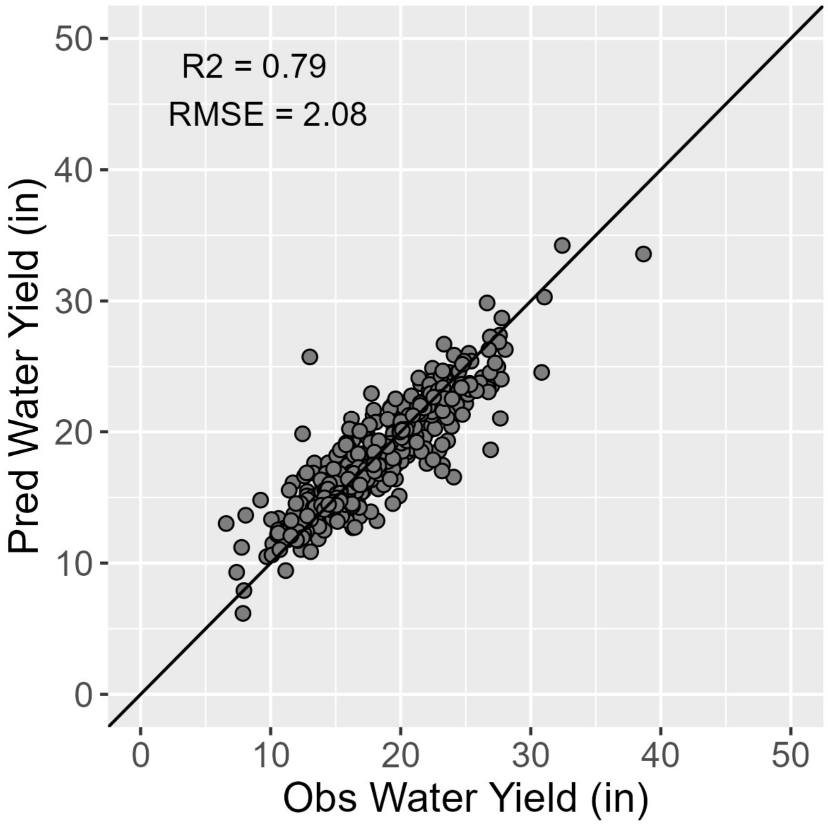
Priestley-Taylor = f(Temperature, Relative Humidity, Altitude, Latitude, Solar Radiation, day of year)

Average Annual Water Yield
PRISM PPT and PRISM Priestley-Taylor PET + watershed characteristics

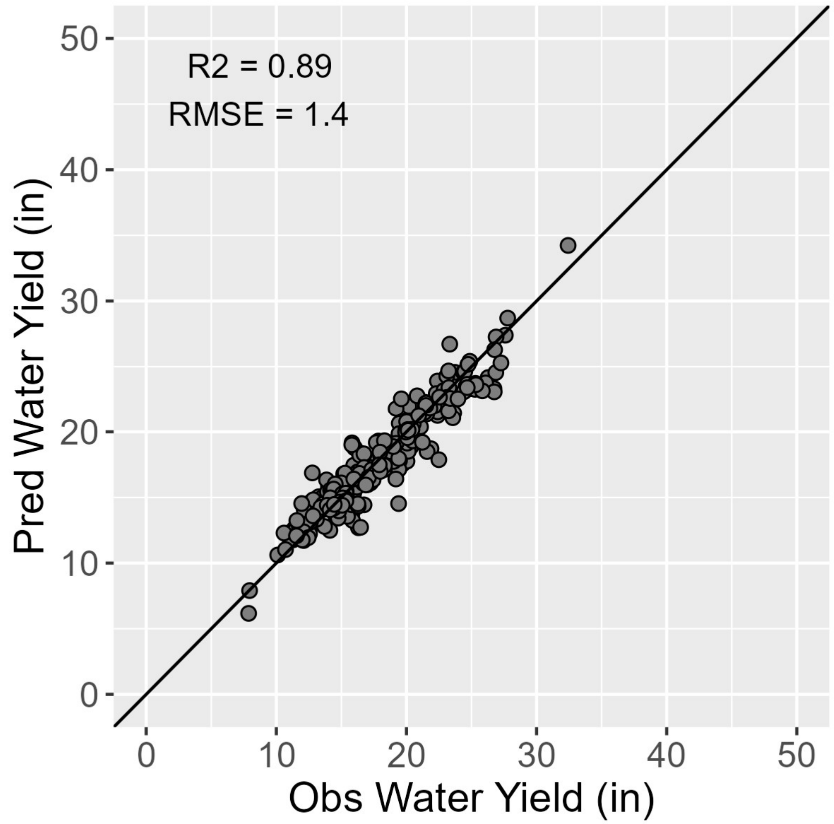
Variable	Coefficient mean	Coefficient 95% CI
Average Annual Temperature (degC)	-0.14	-0.18 - -0.10
Catchment elevation (m)	0.03	0.02 – 0.04
Hunt geological unit: Sandy and stony colluvium (%)	0.01	0.01 – 0.02
Anthracite Mine Land (%)	-0.19	-0.31 - -0.05
Hydrogeomorphic unit: Mesozoic Lowland (%)	0.12	0.01 - 0.22
Soller surficial geology: sandy glacial till sediments (%)	0.03	0.01 – 0.06

Average Annual Water Yield
PRISM PPT and PRISM Priestley-Taylor PET + watershed characteristics

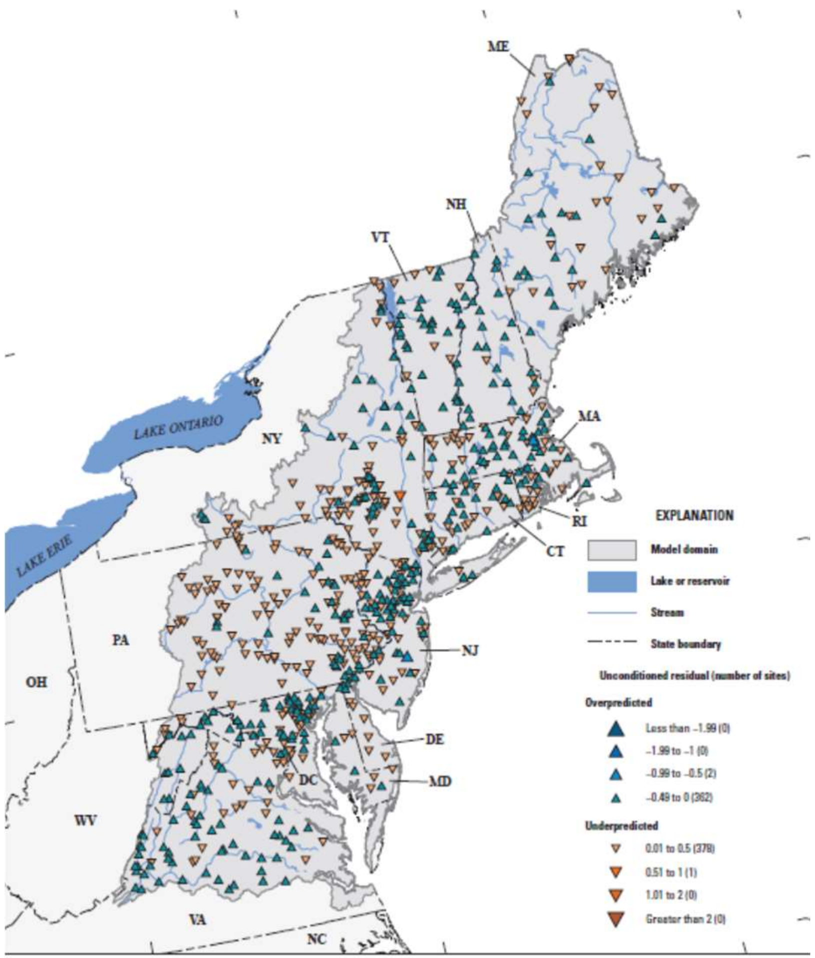
All CalCAST stations



P6 stations



SPARROW Northeast

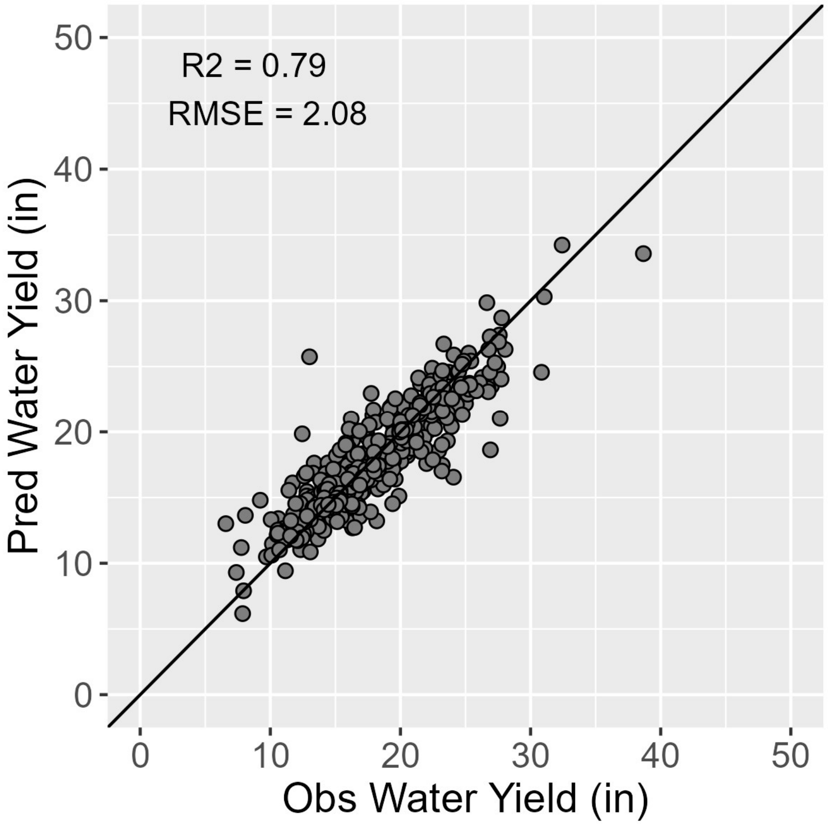


Variable	Variable unit	Coefficient unit	Model coefficient value[a]
		Source	
Mean-annual precipitation minus actual evapotranspiration, 2000–14	ft^3/s	Fraction, dimensionless	1.24
Wastewater point sources	ft^3/s	Fraction, dimensionless	0.998
Transfers from outside region	ft^3/s	Fraction, dimensionless	1
		Land-to-water delivery	
Ln(Mean-annual EVI, 2012)			–0.405
Ln(Mean-annual air temperature, 2000–14)	°C	$°C^{-1}$	–0.129
Ln(Mean-annual soil moisture, 2000–14)	m^3	m^{-3}	–0.0314
		Aquatic loss	
Evaporation from impoundments	Percent	Dimensionless	1.22
Withdrawals for public supply	Percent	Dimensionless	0.0528
Diversions coded in network	ft^3/s	Fraction, dimensionless	1
Spatial test	**Number**	**Correlation/ value**	***p*-value**
Tight clusters—pairs of nested sites within 5 km	16	–0.303	0.2536
		Model summary statistic	
Conditioned RMSE[c] in natural logarithm units		0.1265	
Conditioned RMSE[c], percent in real space units[d]		12.7	
Unconditioned RMSE[c] in natural logarithm units		0.1206	
Unconditioned RMSE[c], percent in real space units[d]		12.1	
Mean exponentiated weighted error		1.007	
R^2		0.9952	
Yield R^2		0.8357	
Number of sites		741	

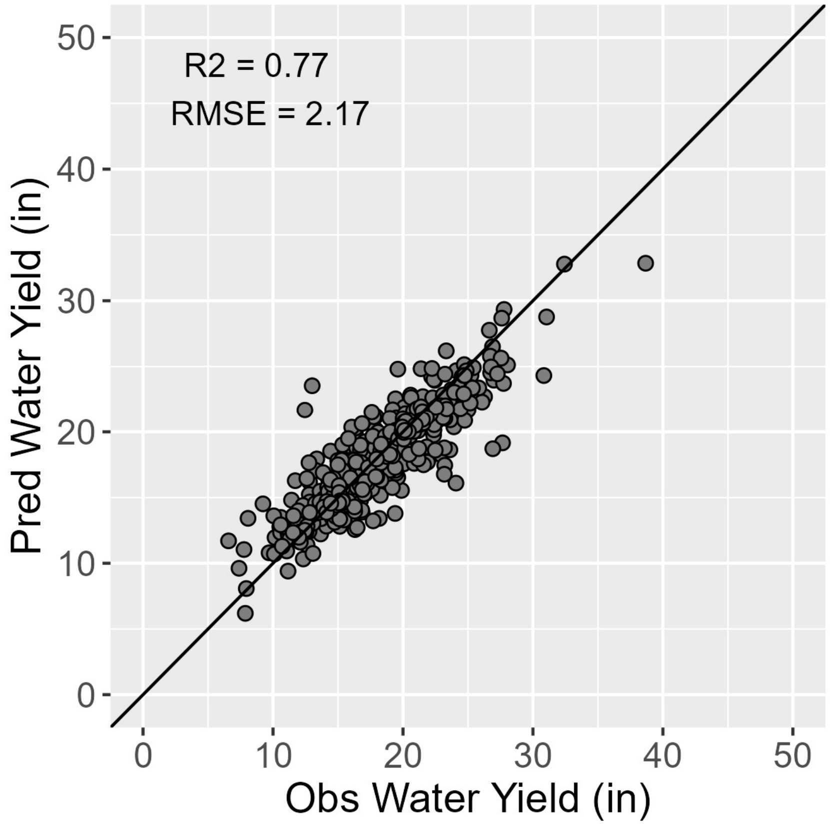
Ator et al. 2019

Average Annual Water Yield
PRISM PPT and PRISM Priestley-Taylor PET + watershed characteristics

Best performing model so far



Replacing Hunt geological unit with average annual soil moisture storage



Proposed next steps

Start exploring watershed predictors of nutrient and sediment loads

Continue making incremental progress on streamflow, also as new datasets become available (e.g., new hydrography and land use data from GIS/Land Use team)

Make sure we are «right for the right reasons»