# Hack The Bay

Challenge 2: Data Gaps

# AUTHORS

Richard Latham
Data Scientist Booz Allen

Lu Sevier
Data Scientist Booz Allen

# Agenda

- **Overview:** Growth of citizen science data
- **Solution:** Visualizing how CMC data compliments CBP data
- **Next Steps:** Analyzing the outcomes of the CMC prioritization report published in 2017

# Overview: Data collection in the Chesapeake Bay watershed has exploded in the last 10 years

- The combination data from the Chesapeake Bay Program (CBP) and the Chesapeake Bay Monitoring Cooperative (CMC) represented **1,796,495** measurements from 2010-2020

- Since 2010, CMC data has increased from 4K points a year to 63K points a year. In 2019, CMC took measurements across **278 distinct HUCs**, across 6 states and the District of Columbia.

- Since 2010, CBP data has stayed relatively stable, changing from 147K points a year to 144K points a year. In 2019, CBP took measurements across **238 distinct HUCs**, across 6 states and the District of Columbia.

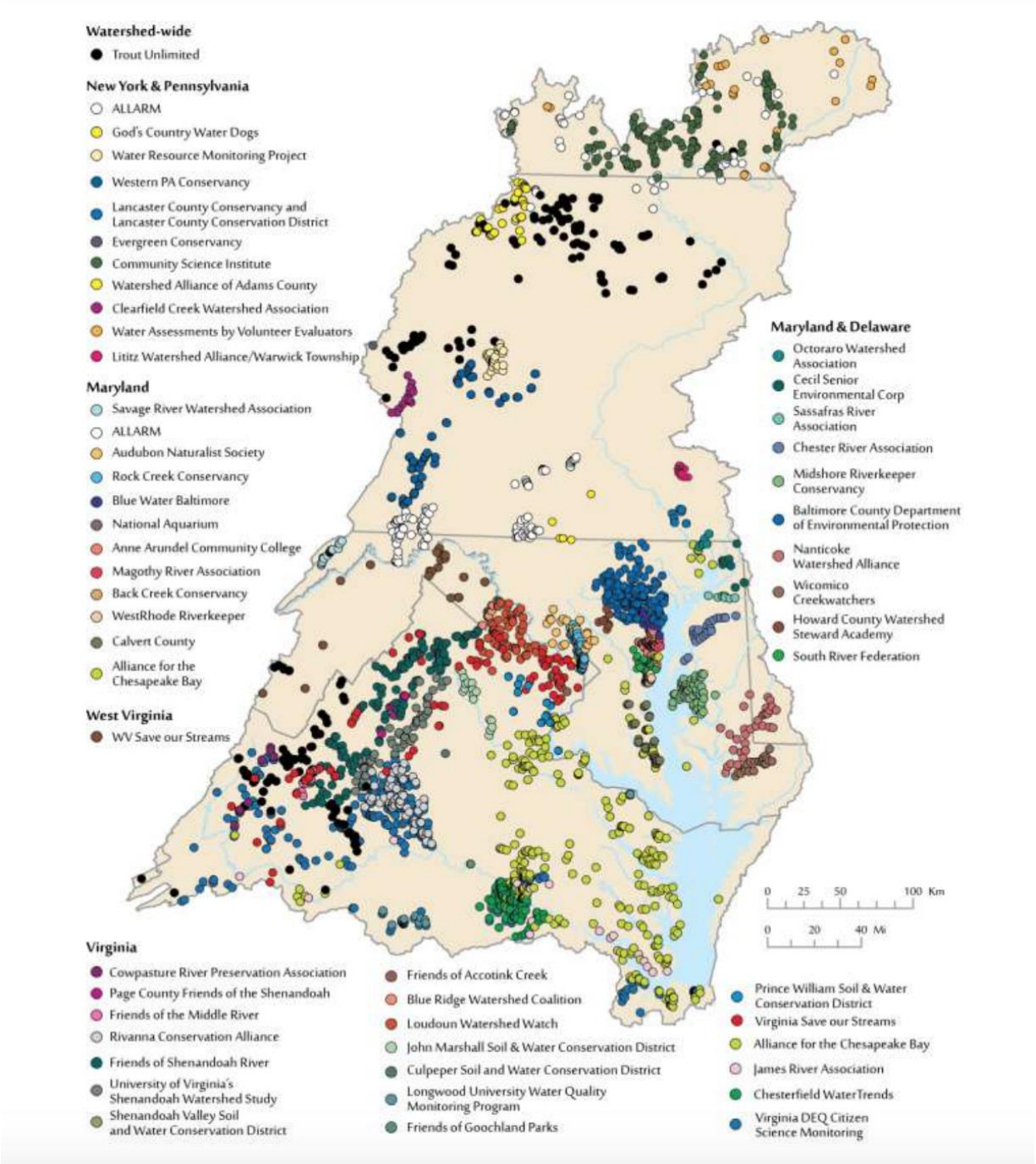
Comparing Water Quality Data Across Databases

| | CMC | | CBP | | All | |
|---|---|---|---|---|---|---|
| Year | Distinct HUCs | Total Points | Distinct HUCs | Total Points | Distinct HUCs | Total Points |
| 2010 | 30 | 4,289 | 202 | 146,804 | 231 | 151,093 |
| 2011 | 67 | 6,561 | 251 | 162,967 | 308 | 169,528 |
| 2012 | 143 | 33,899 | 269 | 176,002 | 398 | 209,901 |
| 2013 | 162 | 14,249 | 269 | 167,935 | 417 | 182,184 |
| 2014 | 189 | 14,129 | 262 | 156,070 | 436 | 170,199 |
| 2015 | 199 | 13,294 | 257 | 156,379 | 443 | 169,673 |
| 2016 | 186 | 13,957 | 246 | 152,746 | 418 | 166,703 |
| 2017 | 234 | 22,220 | 246 | 149,062 | 462 | 171,282 |
| 2018 | 215 | 34,981 | 242 | 146,424 | 427 | 181,405 |
| 2019 | 278 | 63,437 | 238 | 144,301 | 459 | 207,738 |
| 2020 | 163 | 6,448 | 83 | 10,341 | 227 | 16,789 |

Data as of 7/2020; the year 2020 has reduced data points due to partial year data and COVID-19.

This citizen scientist call to action has increased data collection and united monitoring groups across the watershed.

In 2019, CMCs data represented **30.5%** of all water quality data collected in a year.

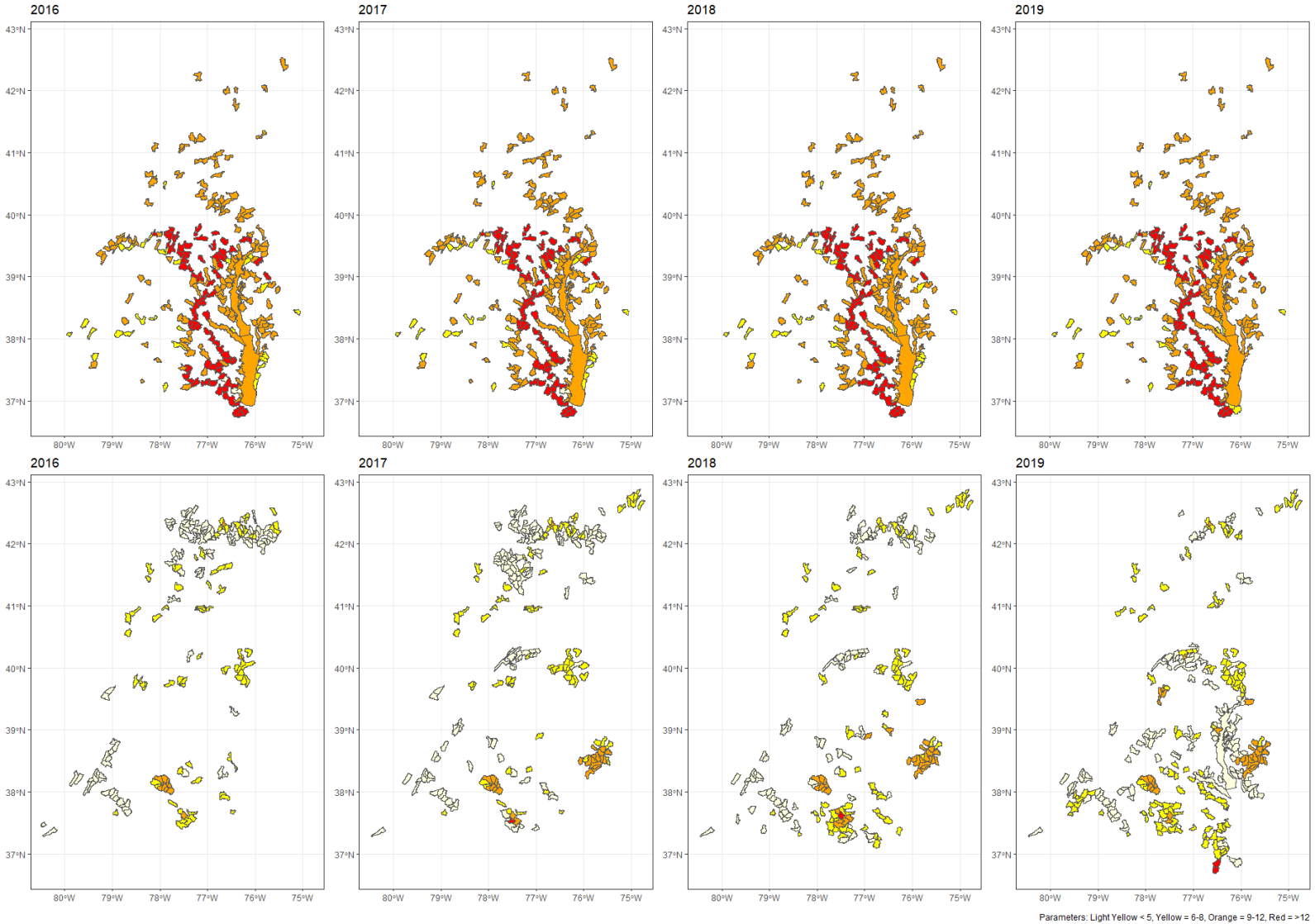


# Visualizing how CMC compliments CBP data

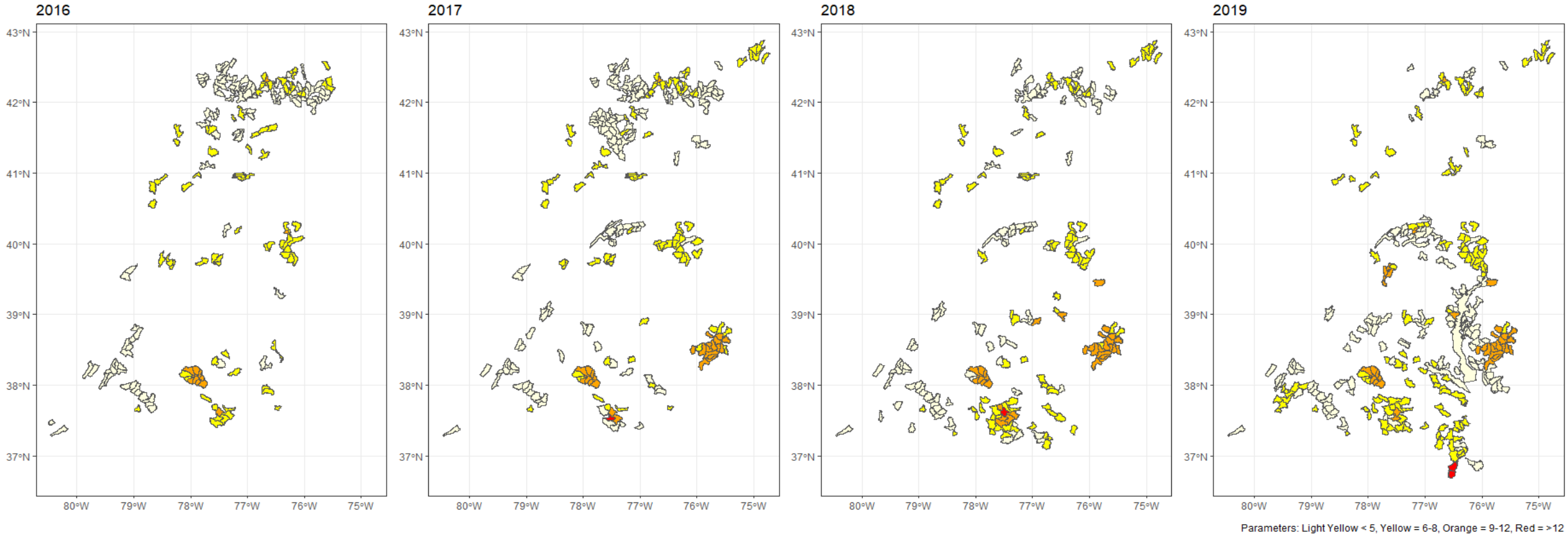
# Visualization: CBP has a large diversity of data parameters across 4 years compared to CMC

CBP

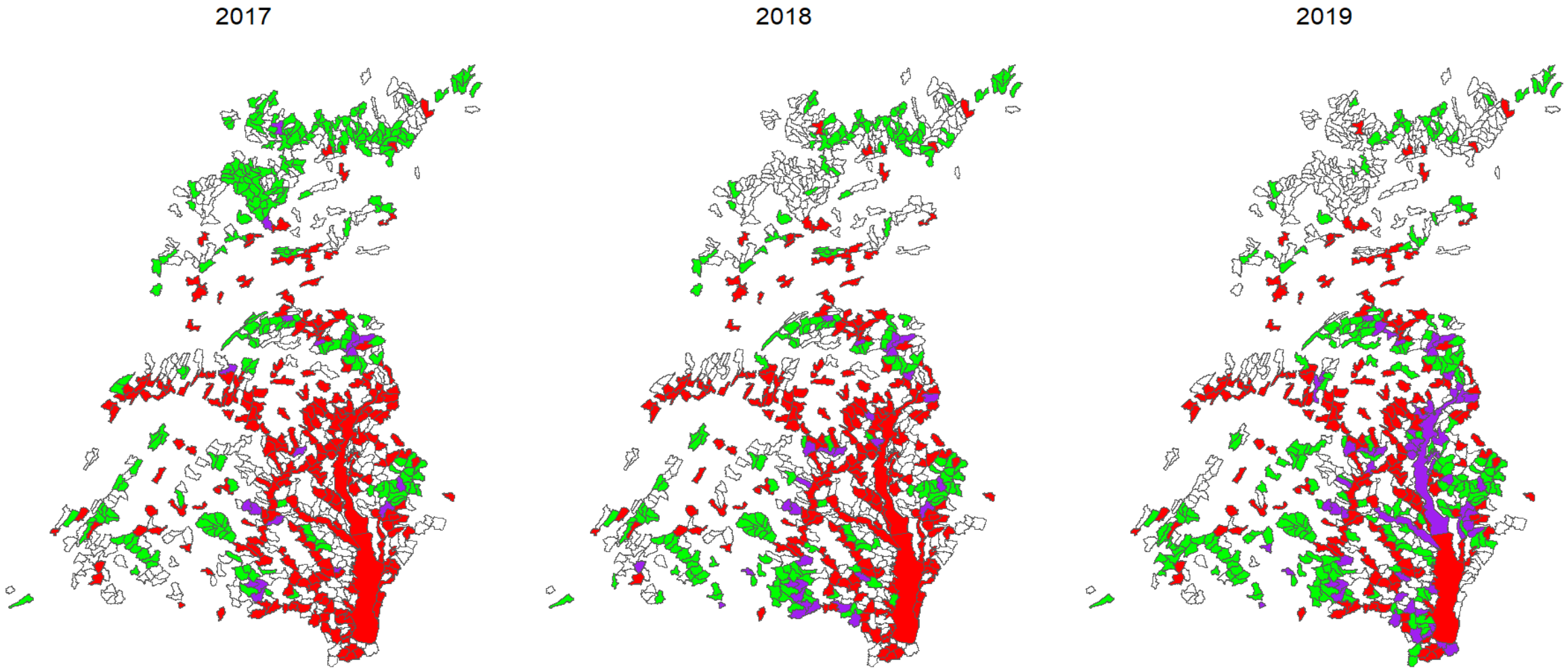
CMC



# Visualization: The diversity of data parameters of CMC has increased in the past 4 years



# Visualization: CMC data continues to measure water quality in areas that the CBP does not reach



# Does CMC data reflect the goals outlined in the 2017 Prioritization report?

# CMC 2017 Prioritization report

The Chesapeake Bay Monitoring Cooperative published a report in 2017 outlining how volunteer and nontraditional monitoring can help fill data gaps in the Chesapeake Bay watershed. The report not only identifies basic monitoring objectives in each state but outlines  preliminary plans of action for how local monitoring groups could support state agency environmental measurement efforts.

**Table 2. Priority objectives for volunteer and nontraditional data use identified by environmental agencies in the Chesapeake Bay watershed**

| Monitoring Data Use/Need | State/Jurisdiction |
|---|---|
| Fill data gaps for Clean Water Act 305(b)/303(d) assessments | DE, DC, MD, NY, VA |
| Monitor restoration progress and conservation effectiveness or conduct impact assessment (pre- and post-implementation) | DE, DC, MD, NY, PA, WV |
| Stormwater MS4 monitoring | PA |
| Collect longitudinal data and monitor trends over time | MD, NY |
| Establish baseline water quality data (tidal) | DE, DC, MD, VA |
| Establish baseline water quality data (nontidal) | DE, DC, MD, PA, VA |
| Monitor the impacts of road salt (conductivity) | MD |
| Marcellus Shale natural gas and acid mine drainage mitigation | MD, PA |
| Determine if a TMDL is needed | MD |
| Higher frequency monitoring of impaired waters | DC, NY, VA |
| Promote stewardship and provide opportunities for community outreach and engagement | DE, DC, MD, NY, PA, VA, WV |
| Identify areas of high nutrient and sediment loading | NY, WV |
| Monitor  areas undergoing change (i.e. development) | DC, WV |
| Climate change resiliency | NY, PA, VA, WV |
| Monitor the impacts of pipelines | PA, VA, WV |
| Monitor areas with high concentrations of agriculture | PA |
| Monitor presence of aquatic invasive species | MD |

CMC Prioritization Report, March 2017 - https://www.chesapeakemonitoringcoop.org/wp-content/uploads/2018/08/Prioritization-Report_Final.pdf

# Example: District of Columbia (2017 – 2019)

Department of Energy and Environment (DOEE) is interested in baseline data, such as **dissolved oxygen, water temperature, conductivity, pH, and benthic macroinvertebrates**. They identified five small watersheds to collect more measurements.

1. Pope Branch
2. Nash Run
3. Hickey Run
4. Watts Branch
5. Ft. Dupont

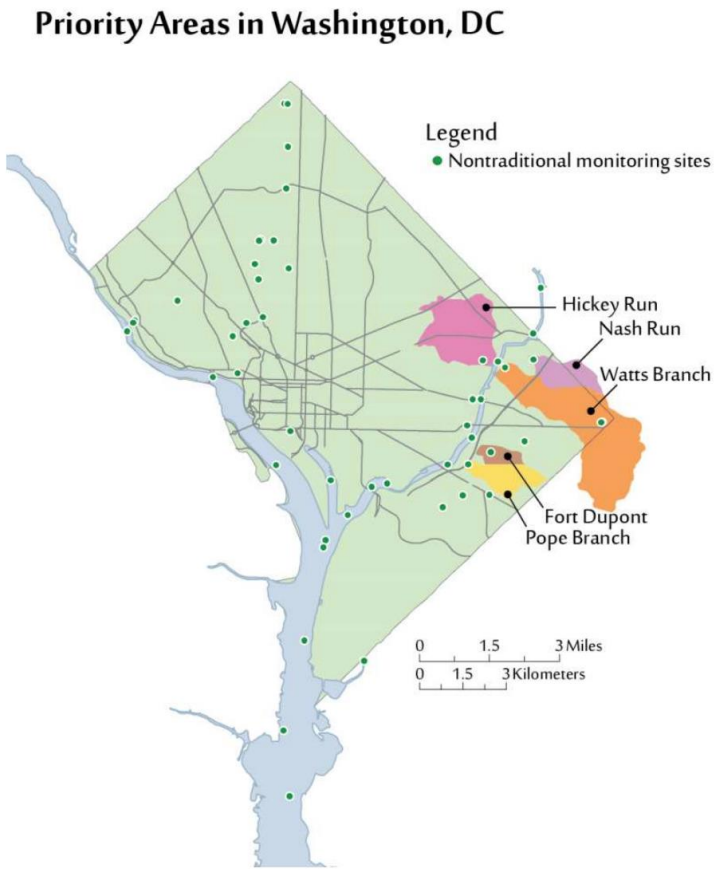



*Figure 5. DOEE identified five watersheds in the District of Columbia where DOEE is seeking more water quality and benthic macroinvertebrate monitoring data. The CMC will investigate volunteer and nontraditional monitoring occurring within those five watersheds.*

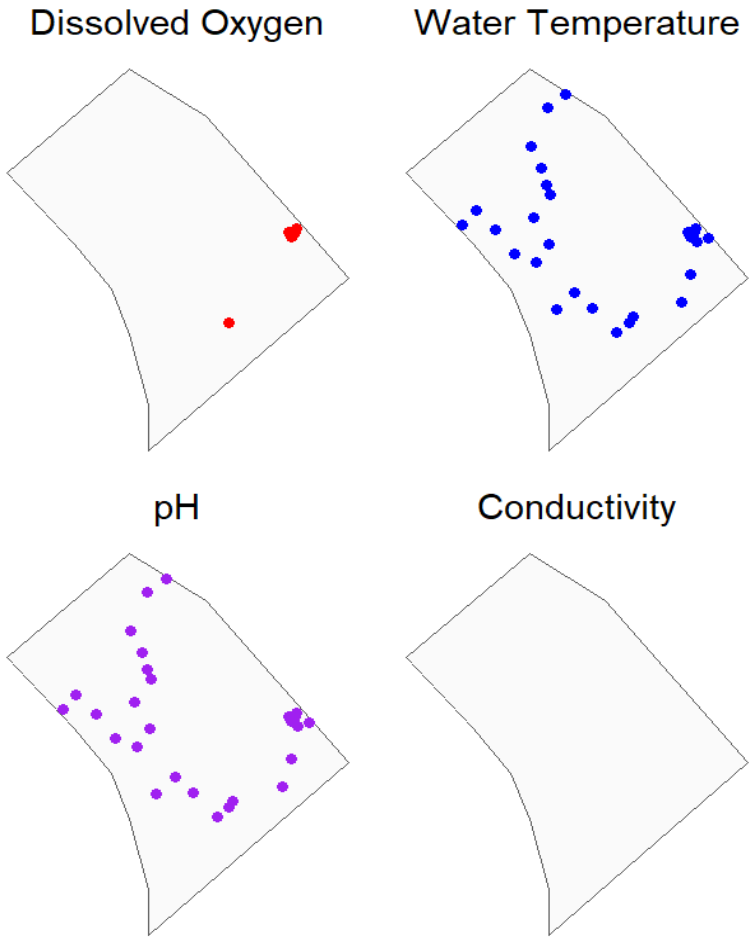
# Example: District of Columbia (2017 – 2019)

Since the publishing of the CMC prioritization report, dissolved oxygen, water temperature, pH and water quality measurements have increased in the DMV in line with the requests from the DOEE. Note that the team did not find any measurements for conductivity.

Growth across parameters in DC

Date range: 2017 - 2020

| Year | Conductivity | Dissolved Oxygen | pH | Water Temperature |
|---|---|---|---|---|
| 2017 | 0 | 2 | 1 | 1 |
| 2018 | 0 | 215 | 134 | 134 |
| 2019 | 0 | 232 | 557 | 558 |
| 2020 | 0 | 0 | 206 | 205 |




Data from 2017-2020 from CMC database.

# Example: District of Columbia (2017 – 2019)

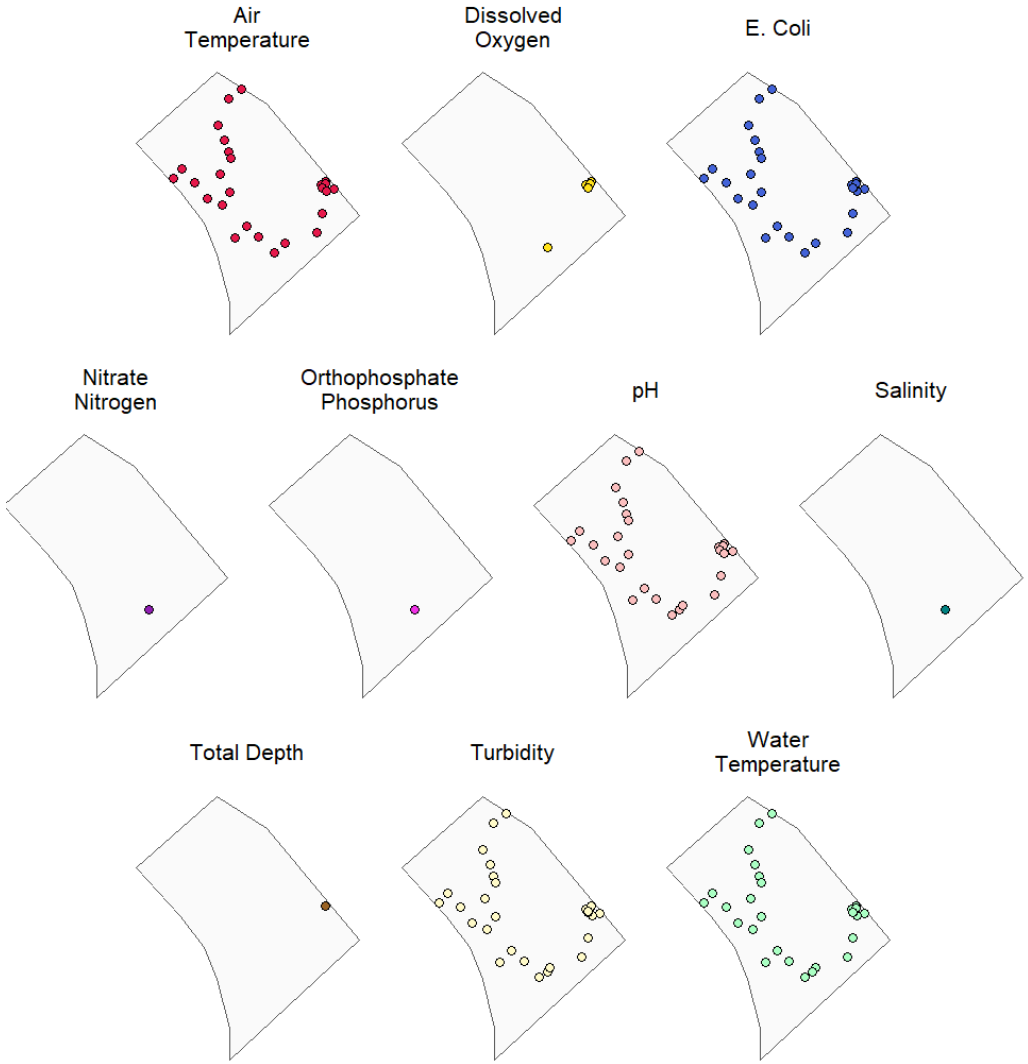
In these three years, DC has also had an increase in the following variables **Air Temperature and Turbidity** which may be of help to the DOEE and other environmental measurement efforts.

Growth in all parameters in DC (CMC)

Date range: 2017 - 2020

| Year | Air Temperature | Dissolved Oxygen | E. Coli | Nitrate Nitrogen | Orthophosphate Phosphorus | pH | Salinity | Total Depth | Turbidity | Water Temperature |
|---|---|---|---|---|---|---|---|---|---|---|
| 2017 | 0 | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| 2018 | 127 | 215 | 120 | 5 | 4 | 134 | 4 | 0 | 127 | 134 |
| 2019 | 563 | 232 | 551 | 0 | 0 | 557 | 0 | 1 | 532 | 558 |
| 2020 | 207 | 0 | 207 | 0 | 0 | 206 | 0 | 0 | 207 | 205 |

Data as of 7/2020. The year 2020 has reduced data points due to partial year data and COVID-19.



# Next Steps: Prioritization

# Recommendations

**Replicating this simple data gaps analysis can provide quick answers to the location of data gaps in the watershed and tangible actions. Possible next steps:**

- Complete a data gap analysis for all listed states the CMC 2017 Prioritization report.
- Share the successes. Where has the data met the goals defined in the report? How has data also increased in this region? Thank the monitoring groups responsible for this great work.
- Collect feedback from the agencies who requested data to learn about their experience using the data. How did they use the data? Was it at the tier that they were looking for?
- Identify data gaps. What data is missing from these regions based on the plan of action defined in the report?
- Scale the analysis by creating a dashboard or yearly report to compare data to outcomes of the 2017 Prioritization Report.