

Tools to Tame the Continuous Data Beast: Examples from the Regional Monitoring Networks

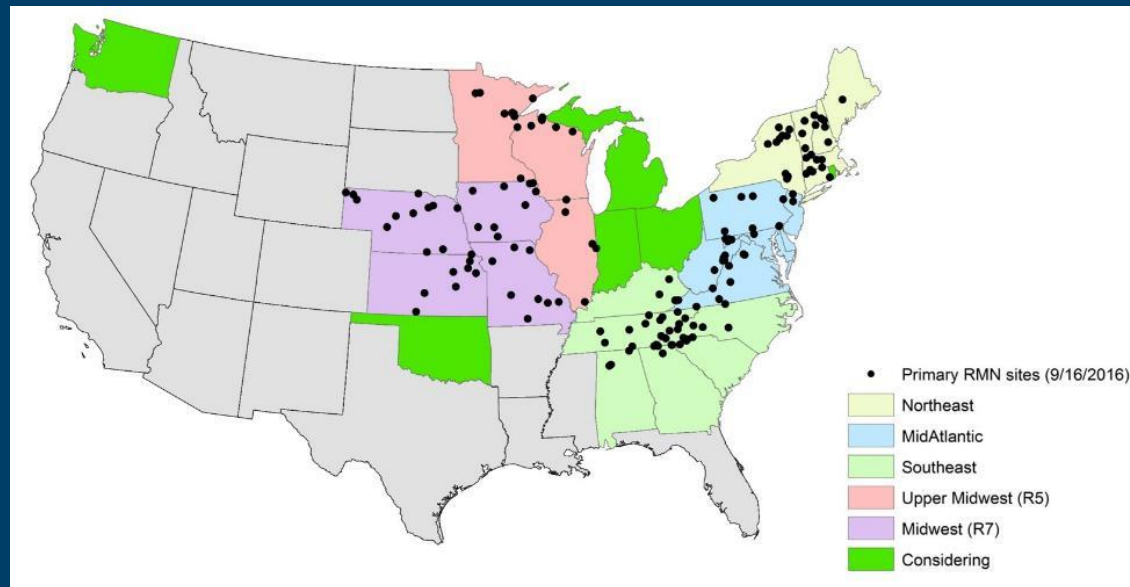
*Britta Bierwagen (USEPA), Jen Stamp, Erik W. Leppo,
Anna Hamilton (Tetra Tech)*



**2019 National Training Workshop,
Shepherdstown, WV
May 31, 2019**

Regional Monitoring Networks (RMNs)

- A volunteer, grassroots effort to document current conditions and detect long-term trends at a regional scale.
- Collect biological, thermal, hydrologic, water quality and habitat data one or more times a year, for 10 or more years, at a set of targeted sites, using regional protocols



Sampling efforts began in the Northeast in 2012, followed by the Southeast in 2013, the Mid-Atlantic in 2014 and the Midwest in 2016-2017.

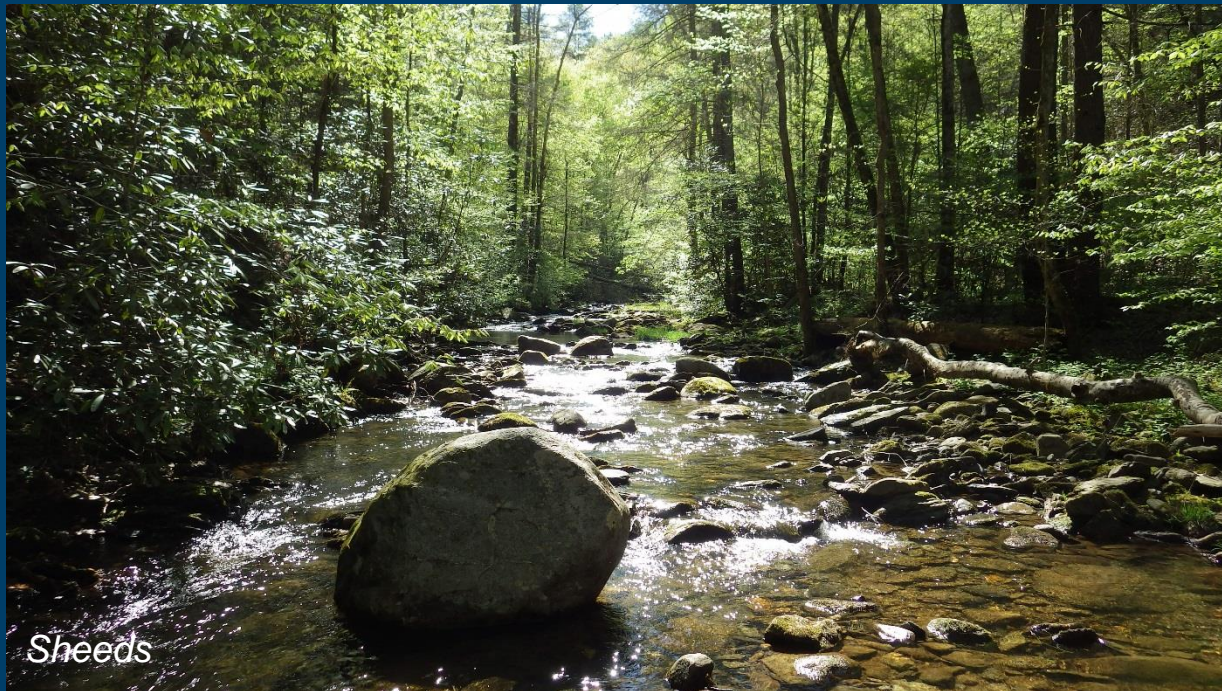
Why?

Baselines are changing.

We need to know how they are changing and how to respond

Data gaps

- Lack of **contemporaneous biological, thermal and hydrologic data**, especially in smaller, headwater, minimally disturbed sites
- This impedes identification and analyses of natural variability and long-term trends



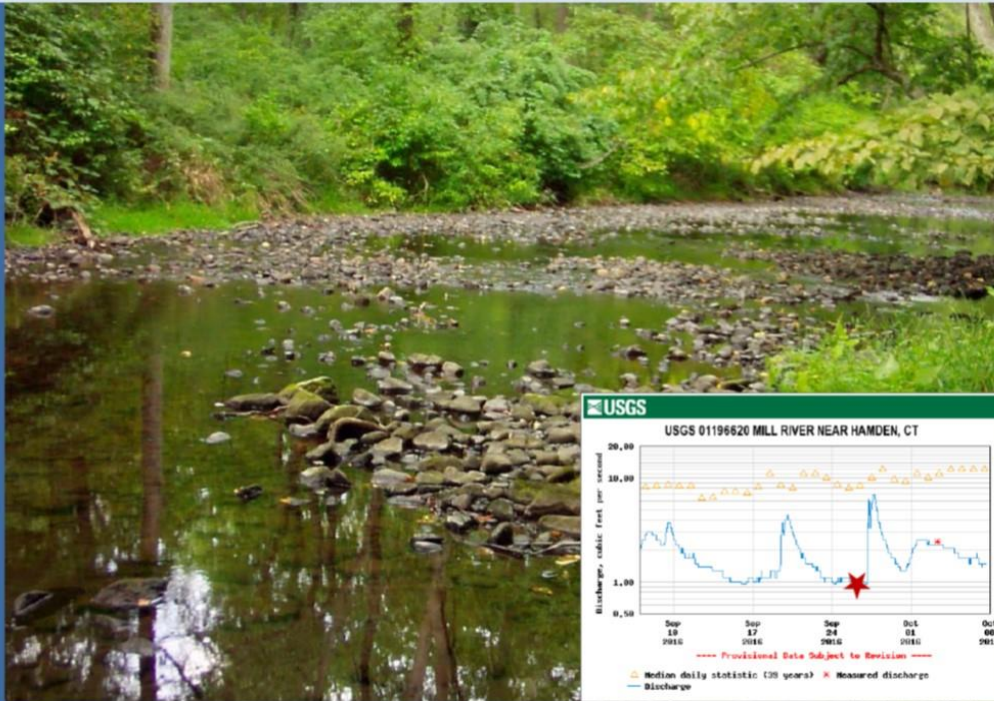
Data collection at RMNs

- **Biological indicators**
 - Benthic macroinvertebrates, optional fish and periphyton
- **Temperature**
 - Continuous water and air temperature
- **Hydrology**
 - Continuous water level data, converted to discharge
- **Habitat** (rapid visual habitat methods; quantitative optional)
- **Water chemistry**



'Photo-rating curve'

09/26/2016 – 0.97 cfs



MOULTRIE

19°C

MOULTRIECAM

26 SEP 2016 03:00 pm

Connecticut Department of Energy and Environmental Protection

Can be used to help identify ecologically important thresholds.

- Drying event – how much wetted habitat is lost at the lowest water levels?
- What water level corresponds with a bedload moving event?

Can also be used for -

- QC (take daily picture of staff gage)
- Documenting changes in riparian vegetation

Slide provided by Chris Bellucci

New approaches to measure flow

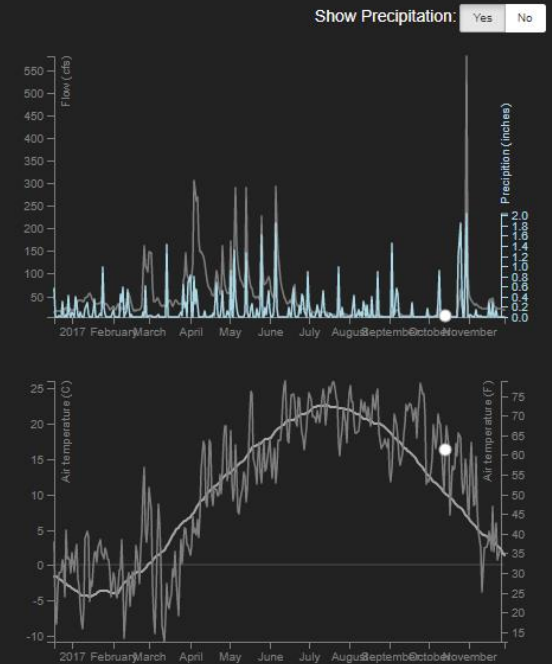
Collaboration to deploy time lapse cameras to take daily images

SHEDS: FLOW PICTURES EXPLORER



Cycle Stop

Show image dashes: Yes No



fpe.ecosheds.org

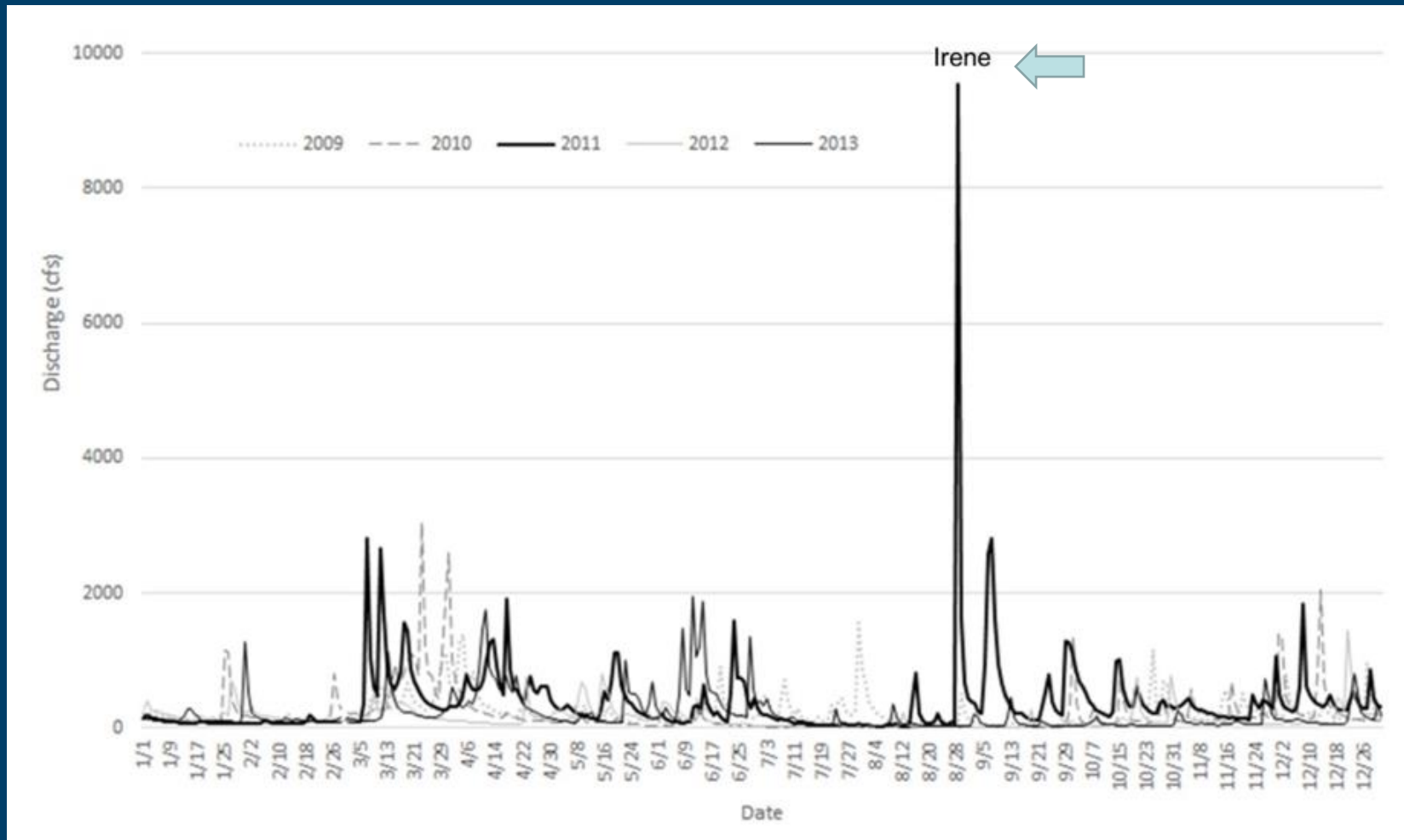
Why continuous thermal and hydrologic data?

Aquatic biota and watersheds are being exposed to **more frequent extreme weather events, warming temperatures and changing hydrologic patterns** (Wuebbles et al. 2017).

Even the most pristine, minimally disturbed sites will be affected by changing temperature and hydrology.



Extreme, episodic events



At some sites we've captured extreme events like high flows from Tropical Storm Irene in VT (note the magnitude & timing of this event!)

General data flow

Discrete data

- Secchi depth
- Water chemistry
- General assessment
- Biology

QA/QC

WQX/STORET
Water Quality Portal (EPA/USGS)

Query – 2 options:

1. WQ portal online interface
2. Data Discovery Tool (DDT)

Data source: RMN partners

QC/processing prior to upload

Upload to data storage system

Query the data system
& download data to your computer



Run data visualization and analysis tools

Continuous data

- Temperature
- Dissolved oxygen
- Water level

QA/QC

Data management/storage system
(each RMN partner needs its own system)

Request data from source

GENERIC QUALITY ASSURANCE PROJECT PLAN For

Monitoring Networks to Document Long-Term Conditions and
Detect Changes in High Quality Wadeable Streams

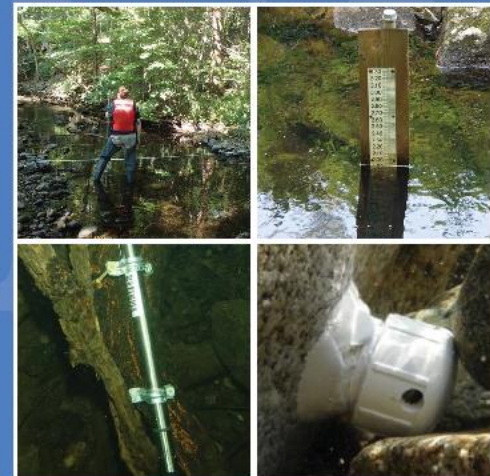
A collaboration among:

U.S. Environmental Protection Agency
U.S. Geological Survey
State and Tribal Agencies
River Basin Commissions
Universities

- Generic QAPP to encourage consistent quality and methods
- Report on best practices
 - Data collection methods for year-round deployments
 - Increases data comparability and quality across all participating entities



Best Practices for Continuous Monitoring of Temperature and Flow in Wadeable Streams



Quality Control guides Consistency

National Generic Quality Assurance Project Plan (QAPP) (February 2016)

Region Specific-Addendums

Many different aspects to collection of high quality continuous data:

- Proper equipment
- Accuracy checks
- Sensor configuration & placement
- Installation techniques
- Documentation
- Maintenance
- Data retrieval
- Data processing & storage

Parameter	Matrix	Accuracy	Precision
Temperature	Water and air	$\pm 0.5^{\circ}\text{C}$	$< 0.5^{\circ}\text{C}$
Water Level	Water	≤ 0.015 ft	Typically 0.2% of full scale of measurement
Discharge	Water	5%	5%
Minimum number of measurements for rating curve	Water	5	5

HOW'S THE
BIG DATA PROJECT
COMING ALONG,
HOSKINS?



© D.Fletcher for CloudTweaks.com

States & tribes collect the data, now what?

- Create free tools to support biomonitoring programs in working with continuous thermal and hydrologic data
- Make biological data preparation and metric calculation faster and easier
- Ensure that a certain (minimum) level of QC is performed
- Format data consistently to facilitate reporting and analysis
- Explore ways to evaluate biological, thermal and hydrologic data in combination without losing the richness of the continuous dataset



R-based tools

ContDataQC

<https://github.com/leppott/ContDataQC>

BioMonTools

<https://github.com/leppott/BioMonTools/>

Development was funded by EPA ORD

Written and maintained by Erik W. Leppo, Tetra Tech.

What does the ContDataQC R package do?

ContDataQC function	Description
FormatHOBOWare	Formats an exported file from HOBOWare for use with ContDataQC, <i>as long as the user follows our export instructions.</i>
GetGageData	Quick download of USGS data.
QCRaw	Generates QC reports. Data are run through four tests (gross, spike, rate of change and flat line). Values that fail the tests are flagged.
Aggregate	Merges files from the same site together. These can be files that have the same parameters but that cover different time periods, or files that contain different parameters for overlapping time periods. This function can also be used to subset files by date.
SummaryStats	Generates summary statistics and time series plots.
PeriodStats	Generates summary statistics and time series plots for the desired time period(s) preceding the biological sampling event
StreamThermal	Exports data in a format that can be run through the ThermalStats R package
IHA	Exports data in a format that can be run through the IHA R package
Flashiness Index	Calculates index for desired time period
CompSite	Creates Cumulative Distribution Function (CDF) plots that allow for comparisons of thermal regimes across sites

It is very important to QC your continuous data!



ContDataQC performs 4 tests

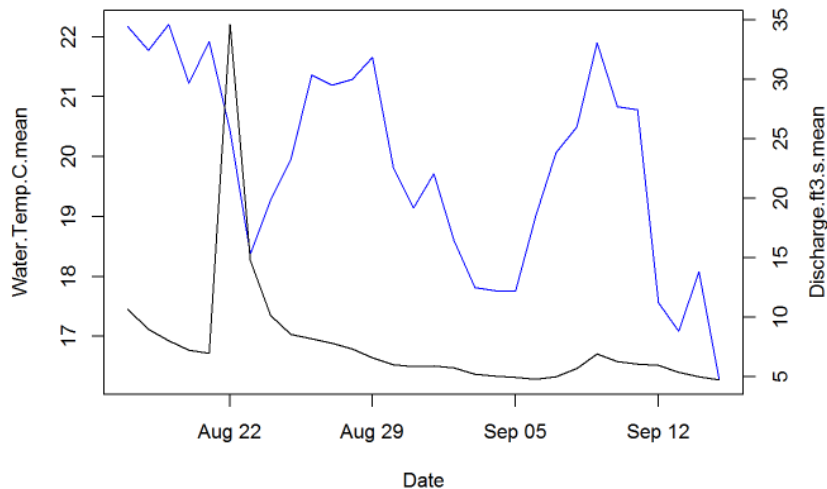
- **Unrealistic values** ('Gross range')
 - Entries are flagged if values are above or below upper and lower limits
- **Spikes**
 - Entries are flagged if adjacent points change by more than 'x' amount
- **Rate of change (RoC)**
 - Entries are flagged if the RoC exceeds a given threshold (e.g., ≥ 3 st dev within 25 hrs)
- **Flat line**
 - Entries are flagged if a certain # of consecutive measurements are within a certain amount of each other

Plus provides time series plots for visual checks

PeriodStats – time series plots

30 days

2016-09-15
(30d)

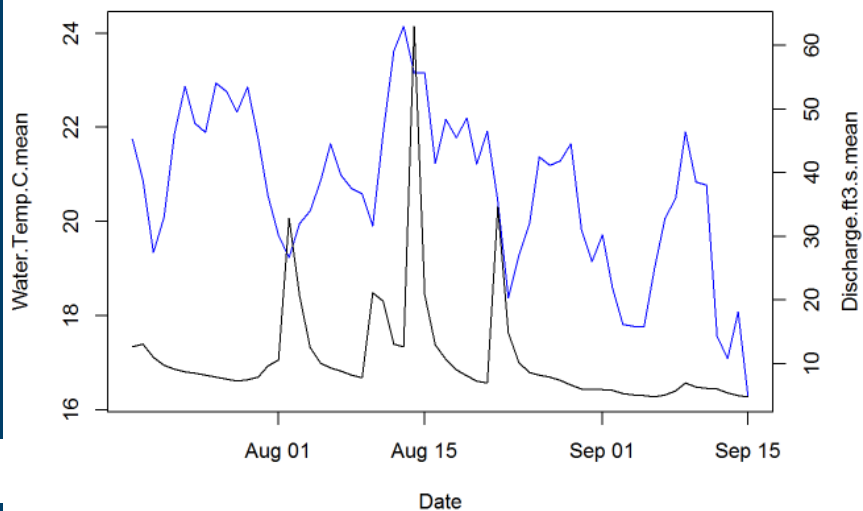


Were flow conditions stable prior to the biological sampling event?

Were temperatures higher than normal?

60 days

2016-09-15
(60d)



Were there any unusual episodic events? If so, characterize the event (magnitude, timing, duration and frequency).

○ Water.Temp.C
□ Discharge.ft3.s

What does the BioMonTools R package do?

BioMonTools function	Description	Bugs	Fish	Algae
Subsample (Rarify)	Subsample (rarify)	x	x	x
Mark excluded taxa	Mark excluded taxa	x	x	x
	Calculate metric values			
Metric values	Metric scores/IBIs	x		
Metric scores		x		
MapTaxaObs	Taxa distribution maps	x	x	x

BioMonTools

- Calculates hundreds of metrics, ranging from commonly-used EPT metrics to thermal and hydrologic indicators
- Having consistent sets of metrics facilitates regional analyses as well as data sharing across entities
- We are developing a master taxa list for the Eastern RMNs and evaluating attribute assignments (such as tolerance values and functional Feeding Group (FFG) assignments)
- Over time, as more data become available, we need to refine the lists of thermal and hydrologic indicator taxa and gain a better understanding of ecologically meaningful thresholds

R Shiny tools

Beta version performs the basic functions of the ContDataQC R package.

The user does not have to download R software onto their computer or work with R code.

Log onto this (temporary) website and select the functions from drop-down menus:

https://davidagibbs.shinyapps.io/rmn_continuous_data_active/

Regional Monitoring Networks (RMNs)

Framework



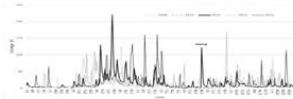
- [What are RMNs?](#)
- [Who can participate?](#)
- [Where do RMNs exist?](#)
- [How do I join?](#)
- [History](#)

Streams



- [Stream RMNs \(context/overview/science for decision making\)](#)
- [Protocols](#)
- [Site selection criteria](#)
- [Data management](#)
- [Data analysis](#)
- [Publications](#)

Tools



- [Continuous sensor data](#)
- [Biological data](#)
- [GIS-based site screening](#)

Lakes



- [Lake RMNs \(overview\)?](#)
- [Protocols](#)
- [Site selection criteria](#)
- [Data management](#)
- [Data analysis](#)
- [Publications](#)

Resources

- [Information for regional coordinators](#)
- [Existing state and tribal partners](#)
- [Thinking of joining?](#)
- [Case studies](#)

Wetlands



- [Wetland RMNs \(overview?\)](#)
- [Publications](#)

Related EPA research

- [Resiliency](#)
- [Vulnerability assessments](#)
- [Adaptation Design Tool](#)
- [EPA NARS](#)

- [Calendars/upcoming events](#)

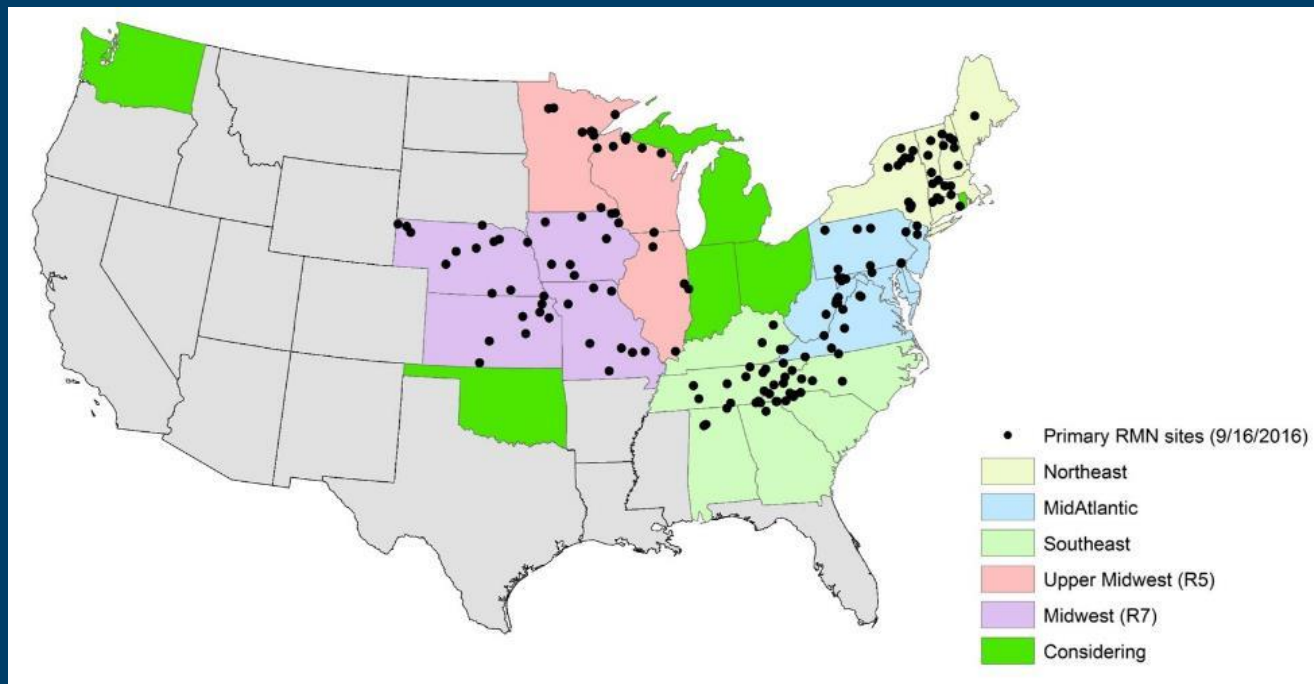
Other related research

- [National Reference Network](#)
- [NPS](#)
 - [Eastern Rivers and Mountains \(ERMN\)](#)
 - [Great Lakes Inventory & Monitoring Network](#)

Acknowledgements



All of our RMN partners!!

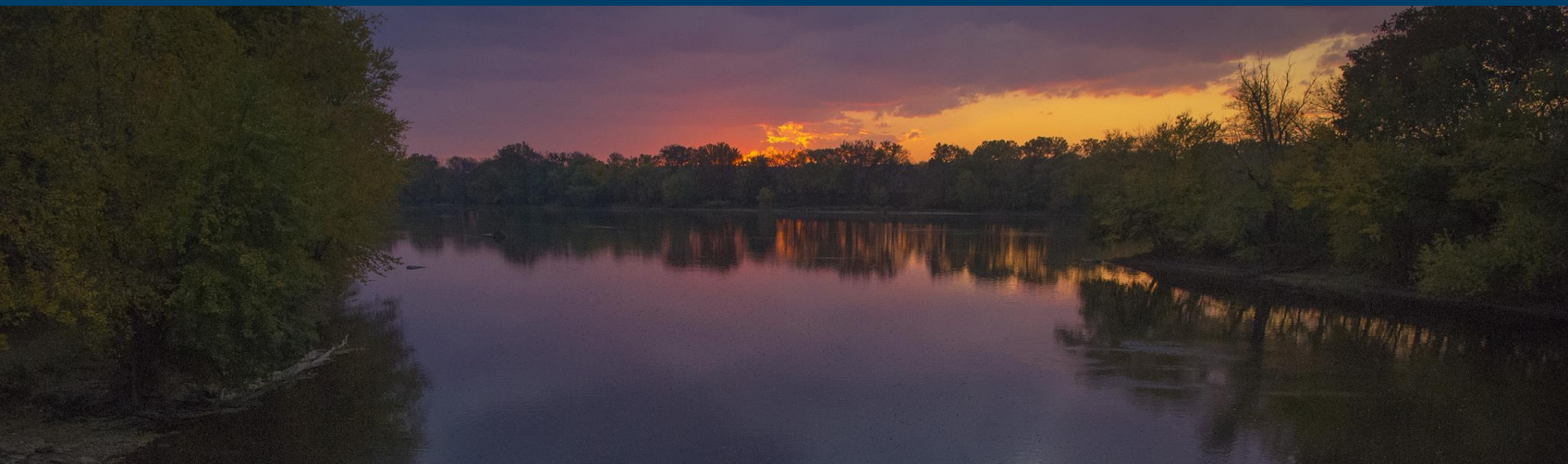


QUESTIONS? COMMENTS?

Jen Stamp (Jen.Stamp@tetratech.com)

Erik W. Leppo (Erik.Leppo@tetratech.com)

Britta Bierwagen (bierwagen.britta@epa.gov)



Example of using the PeriodStats function to evaluate antecedent conditions prior to the biological sampling event

- Were flow conditions stable prior to the biological sampling event (30, 60, 90+ days)?
- Were temperatures higher than normal?
- Were there any unusual episodic events?
 - If so, characterize the event (magnitude, timing, duration and frequency).
- Important to evaluate temperature and flow in combination

PeriodStats – summary statistics

- .csv file with **many statistics** (mean, max, min, stdev, percentiles for daily/monthly/seasonal/annual time periods)

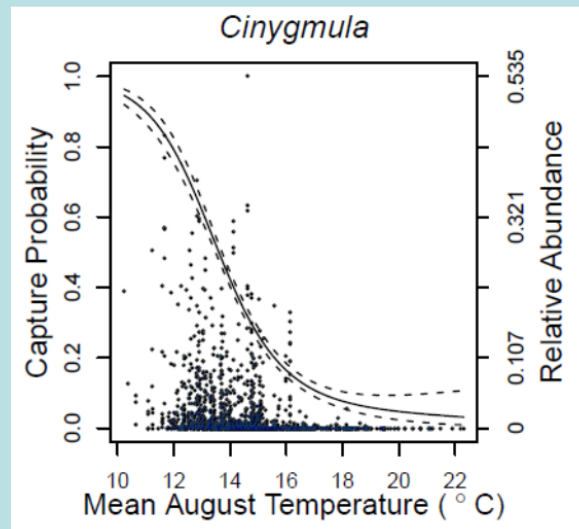
1	Date	Water.Temp.C.mean	Water.Temp.C.median	Water.Temp.C.min	Water.Temp.C.max	Water.Temp.C.range	Water.Temp.C.sd
2	2016-08-17	22.17395833	21.55	20.6	24.2	3.6	1.400018405
3	2016-08-18	21.77604167	22.05	19.1	24.4	5.3	1.922703403
4	2016-08-19	22.20416667	21.8	20.1	24.5	4.4	1.527289768
5	2016-08-20	21.22395833	21.65	18.1	24	5.9	2.139392181
6	2016-08-21	21.92395833	21.9	20.8	23.3	2.5	0.894279456
7	2016-08-22	20.43333333	20.15	18.5	22.5	4	1.105076008
8	2016-08-23	18.37083333	18.3	15.6	21.3	5.7	1.885061309
9	2016-08-24	19.26875	19.5	16.3	22.5	6.2	2.265963836

18 metrics

Need to continue to improve indicator taxa lists and our understanding of ecologically meaningful thresholds!

Modeled relationship between probability of occurrence of *Cinygmula* & modeled mean August stream temperature*

- **Points:** actual data of relative abundance
- **Curved lines:**
 - *Solid* – modeled capture probability based on the generalized additive model (GAM) fit
 - *Dotted* - estimated 90% confidence limits for the GAM model fit



- Based on BCG dataset (Puget Lowlands/Willamette Valley)
- Can help inform development of the thermal indicator taxa list
- Also considering results from other analyses from the region (Idaho, Oregon)

*averaged 1993-2011; based on *Isaak et al. 2015– NorWeST network*

RMN Data Usage

Most RMN sites are in
this time period ↓

1-5 years

- Establish current (“baseline”) conditions
 - Biology
 - Temperature
 - Hydrology
 - Water chemistry
 - Habitat
- Supplement Clean Water Act (CWA) programs
 - Refine lists of thermal, hydrologic and biological indicators
 - Inform criteria development or refinement (e.g., defining natural conditions)
- Refine classification (make comparisons across sites; evaluate differences and similarities)
- Evaluate variability (within-year (e.g., seasonal) and across years)
- Compile reference datasets that can be used to further biocriteria development

5-10 years

- Track indicators
- Explore relationships between water quality, climate drivers and other variables of interest

10+ years

- Detect and track trends in regional phenomena such as climate variability, atmospheric deposition and spreading of invasive species
- Track effectiveness of adaptation strategies

Document ecosystem response and recovery to extreme weather events (if they happen to occur)