EPA
United States
Environmental Protection
Agency

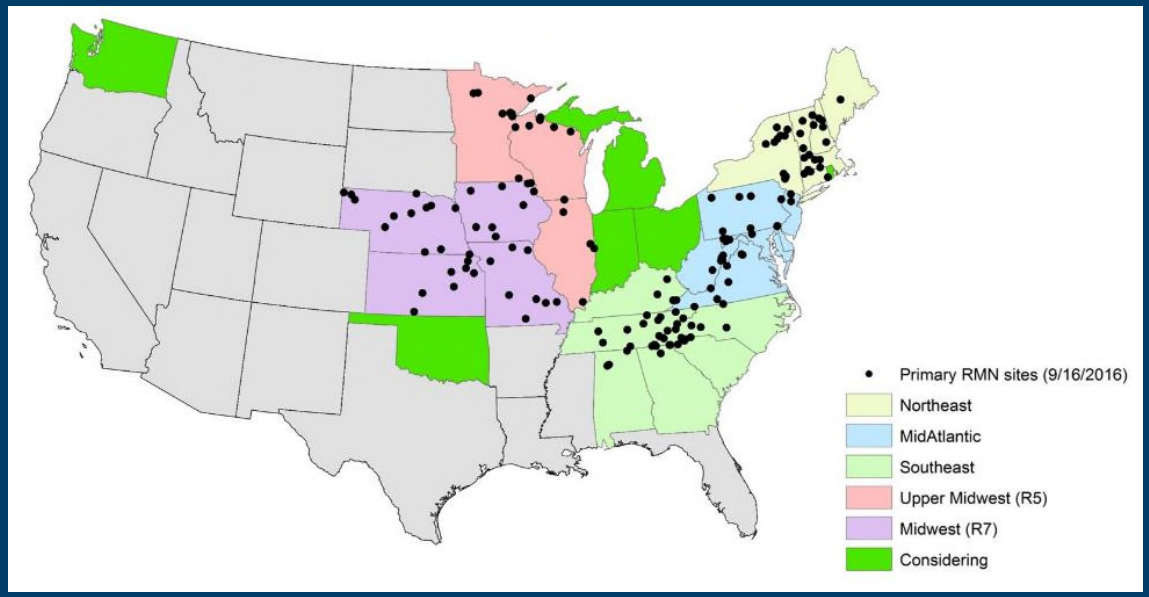
Analyzing stream macroinvertebrate data in combination with continuous thermal and hydrologic data

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

NWQMC conference, Denver, CO
March 28, 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

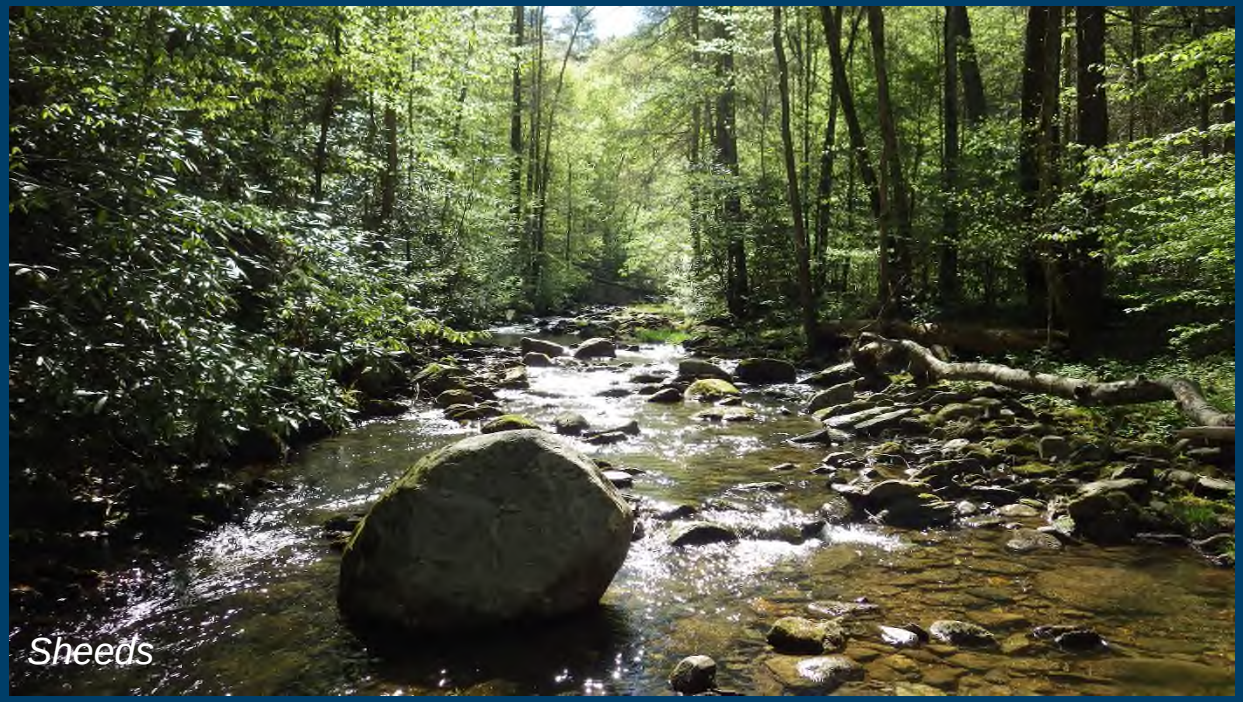


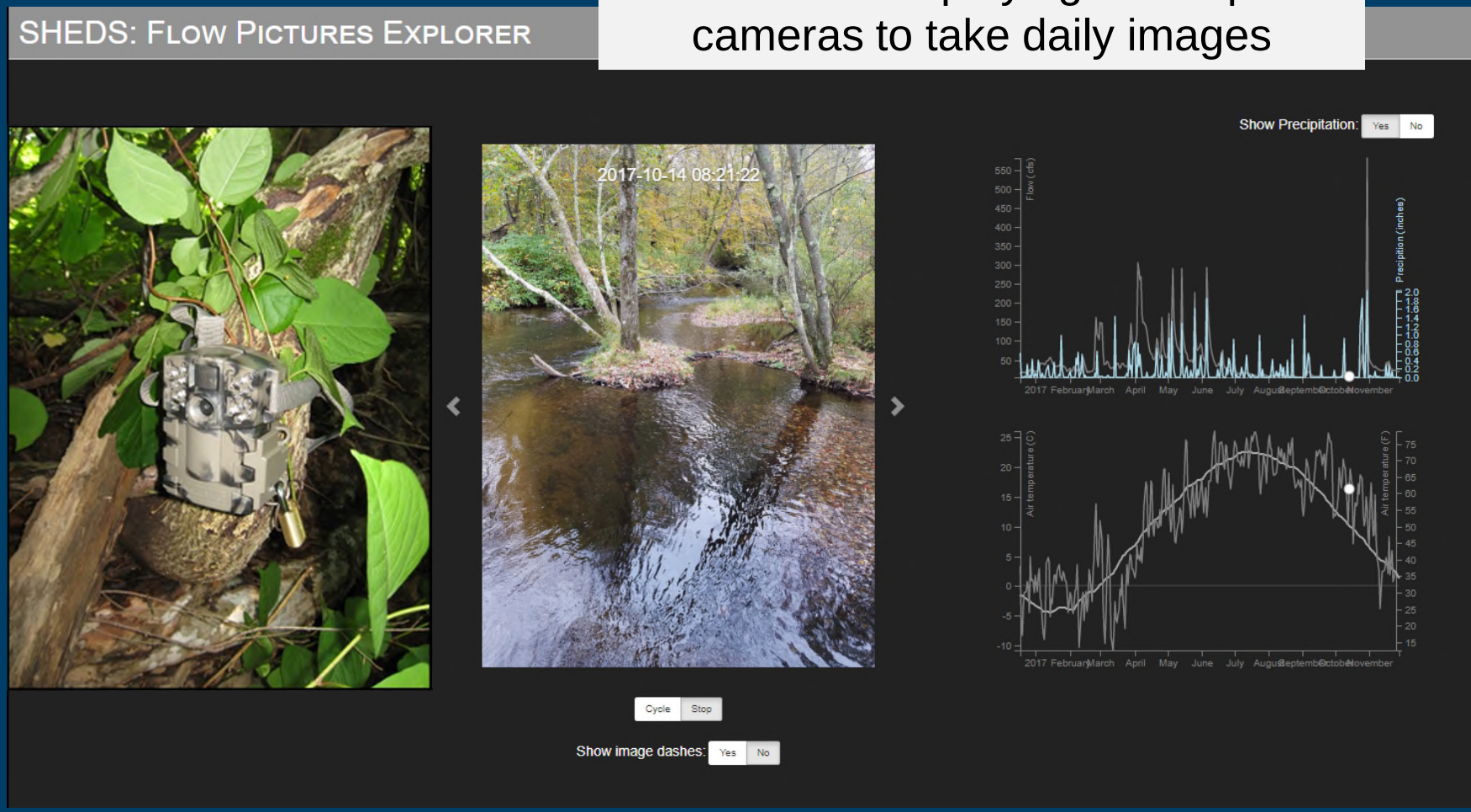
Photo provided by TN DEC

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**

Data collection at RMNs

NEW! Also deploying time lapse cameras to take daily images



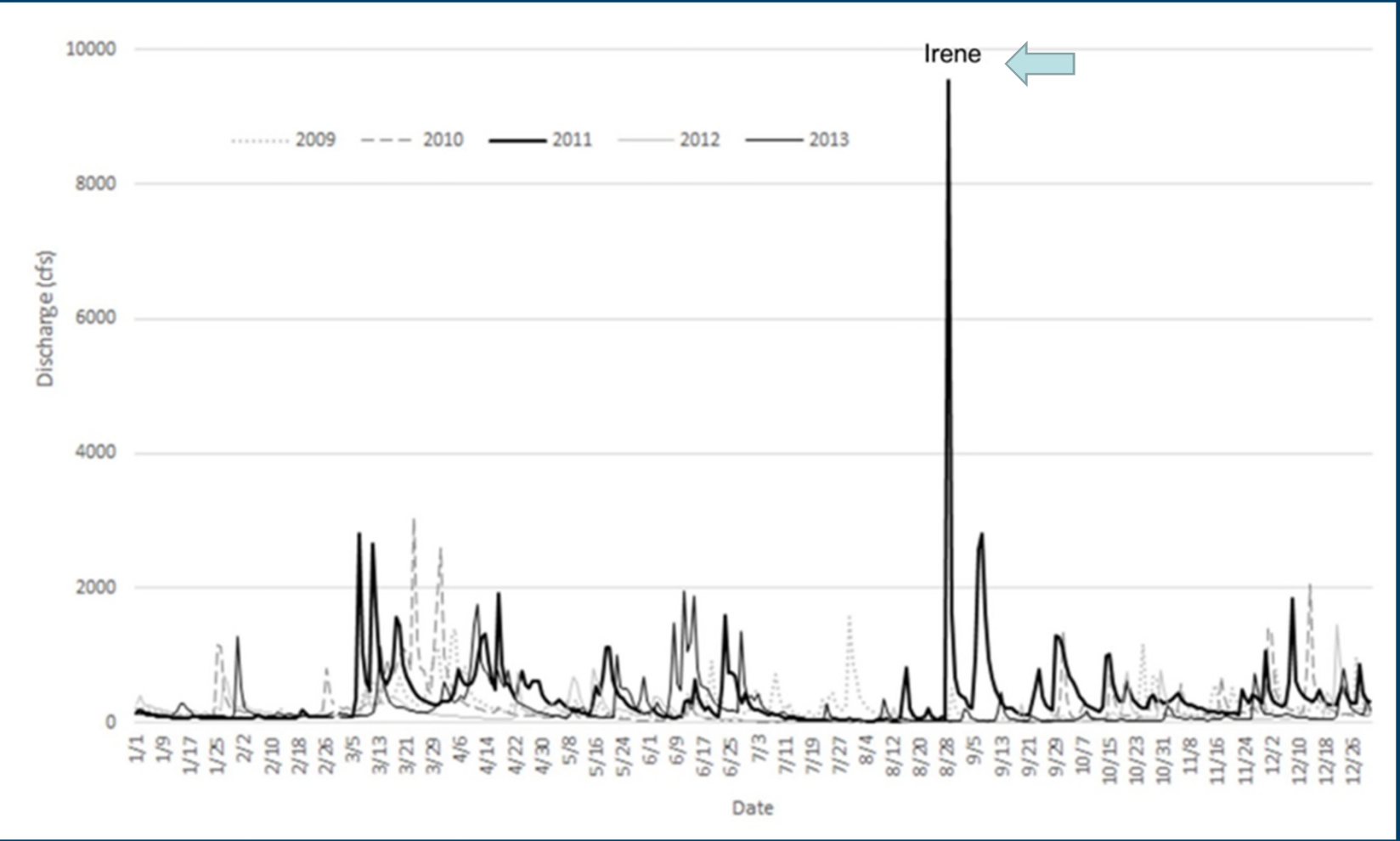
fpe.ecosheds.org

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!)

Changing thermal and hydrologic conditions effect biomonitoring programs

High flows from Irene affected density scores in 2011

VERMONT DEPARTMENT OF ENVIRONMENTAL CONSERVATION
WATERSHED MANAGEMENT DIVISION

Macroinvertebrate Site Summary - River/Stream

Green River

Located about 1.0mi below Green River Village, off River Rd.
Guilford, VT (42.7588, -72.66223)
Stream Type: Medium High Gradient

Macroinvertebrate Community Metrics

Macroinvertebrate Community Assessments are based primarily on eight metrics of the Macroinvertebrate community. These include metrics of abundance, species richness, and indexes of Sensitive to tolerant species

Date	Density	Richness	EPT Richness	PMA-O	B.I.	Oligo.	EPT/EPT + Chiro	PPCS-F	Community Assessment
10-24-1991	1379	48.0	26.5	77.3	3.35	0.00	0.70	0.59	Excellent
10-6-1992	2136	50.0	29.0	84.1	3.12	0.38	0.79	0.61	Excellent
10-14-1993	1568	37.0	23.5	82.0	3.22	0.00	0.83	0.58	Excellent
10-6-2011	203	41.0	28.5	85.6	2.53	0.00	0.98	0.58	Fair
9-26-2012	4632	58.0	36.5	78.7	3.22	0.14	0.95	0.51	Excellent
10-11-2013	2025	39.0	24.0	59.5	4.23	0.00	0.91	0.27	Fair
9-30-2014	2384	45.5	28.5	85.2	2.65	0.61	0.85	0.56	Excellent
9-8-2015	1996	60.0	34.0	78.0	3.30	0.00	0.90	0.56	Excellent
9-15-2016	1804	58.5	32.0	70.5	3.69	0.00	0.93	0.52	Excellent - Very Good
Scoring Guideline for a MHG Stream of Water Quality Class B(2)									
	≥ 300	≥ 30	≥ 18	≥ 45	≤ 5	≤ 12	≥ 0.45	≥ 0.4	Full Support
	≥ 250	≥ 28	≥ 16	≥ 40	≤ 5.15	≤ 14.5	≥ 0.43	≥ 0.35	Indeterminate
	< 250	< 28	< 16	< 40	> 5.15	> 14.5	< 0.43	< 0.35	Non-Support

Provided by VT DEC

Our goals

- Create free tools to make it easier for biomonitoring programs to work 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
- Explore ways to evaluate biological, thermal and hydrologic data in combination without losing the richness of the continuous dataset
- Make the tools accessible (“bioassessment toolbox”)

R-based tools

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

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

Development was funded by EPA ORD (lead: Britta Bierwagen).

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

What does the ContDataQC R package do?

ContDataQC function	Description
FormatHOBO	Formats an exported file from HOBOWare for use with ContDataQC, *as long as the user follows our export instructions.*
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

ContDataQC R package

Parameters

- **Temperature (water and air)**
- **Water level**
- **Discharge**
- DO
- pH
- Conductivity
- Turbidity
- Chlorophyll-a

Have experience from the stream RMNs

Unsure of what thresholds are appropriate for these parameters
(if you have experience with these parameters, we'd love to get your feedback)

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’ function in ContDataQC

```
#Command lines
# function inputs
myDate <- "2016-09-15"
myDate.Format <- "%Y-%m-%d"
myPeriod.N <- c(30, 60, 90, 120, 1)
myPeriod.Units <- c("d", "d", "d", "d", "y")
myFile <- "DATA_Green_Gage_20150101_20161230.csv"
myDir.import <- getwd()
myDir.export <- getwd()
myParam.Name <- c("Water.Temp.C", "Discharge.ft3.s")
myDateTime.Name <- "Date.Time"
myDateTime.Format <- "%Y-%m-%d %H:%M:%S"
myThreshold <-
myConfig <- ""
myReport.format <- "html"

PeriodStats(myDate
            , myDate.Format
            , myPeriod.N
            , myPeriod.Units
            , myFile
            , myDir.import
            , myDir.export
            , myParam.Name
            , myDateTime.Name
            , myDateTime.Format
            , myThreshold
            , myConfig
            , myReport.format)
```

Entries

Biological sampling event
- 2016-09-15

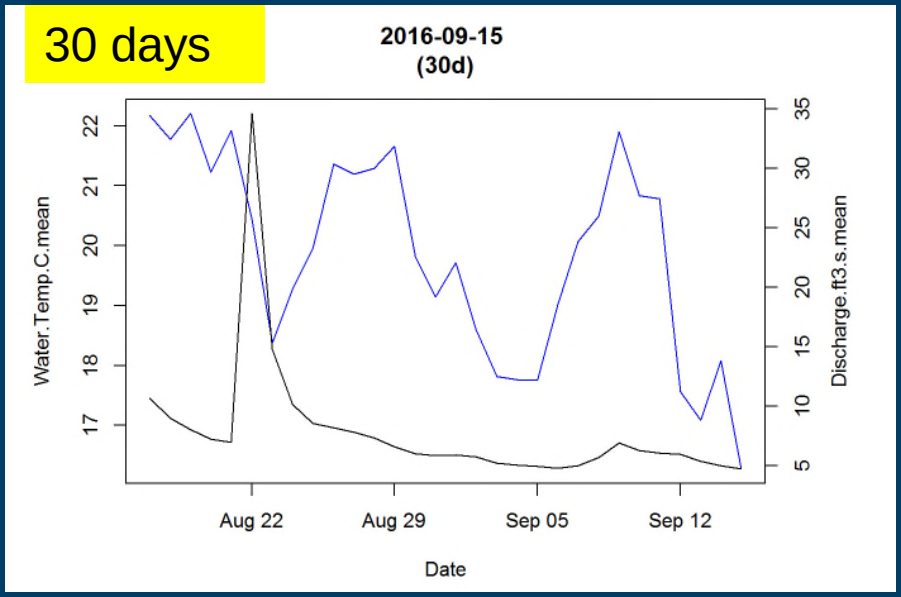
Desired time period(s)
- 30, 60, 90, 120 days, 1 year

Parameter
- Temperature & discharge

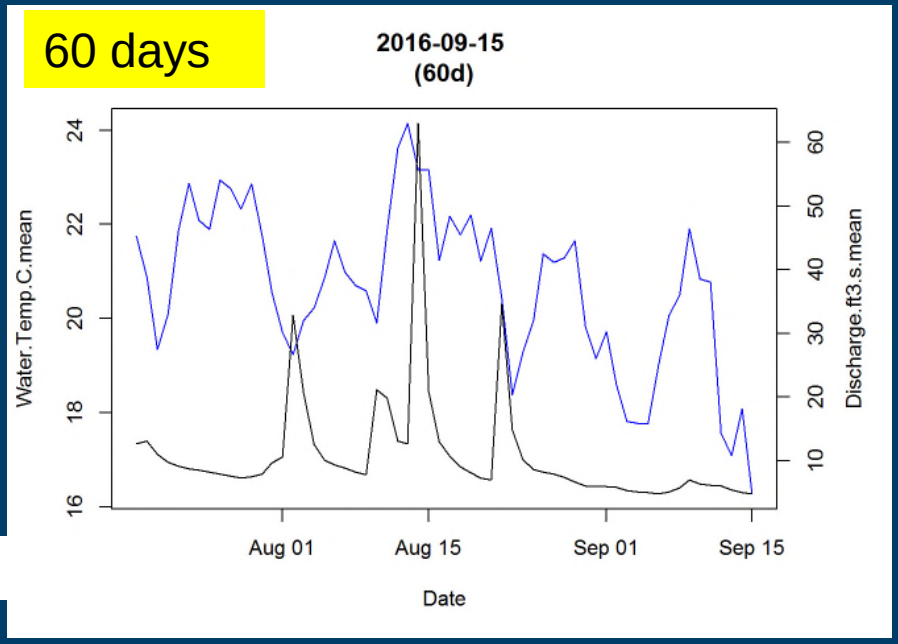
Threshold (if desired)

PeriodStats – time series plots

30 days



60 days



Were flow conditions stable prior to the biological sampling event?

Were temperatures higher than normal?

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

Water.Temp.C
Discharge.ft3.s

PeriodStats – summary statistics

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

	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

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
Metric values	Calculate metric values	x		
Metric scores	Metric scores/IBIs	x		
MapTaxaObs	Taxa distribution maps	x	x	x

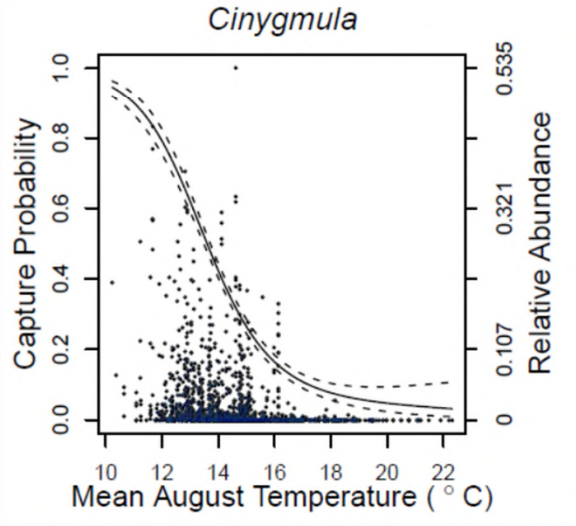
BioMonTools

- 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

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)

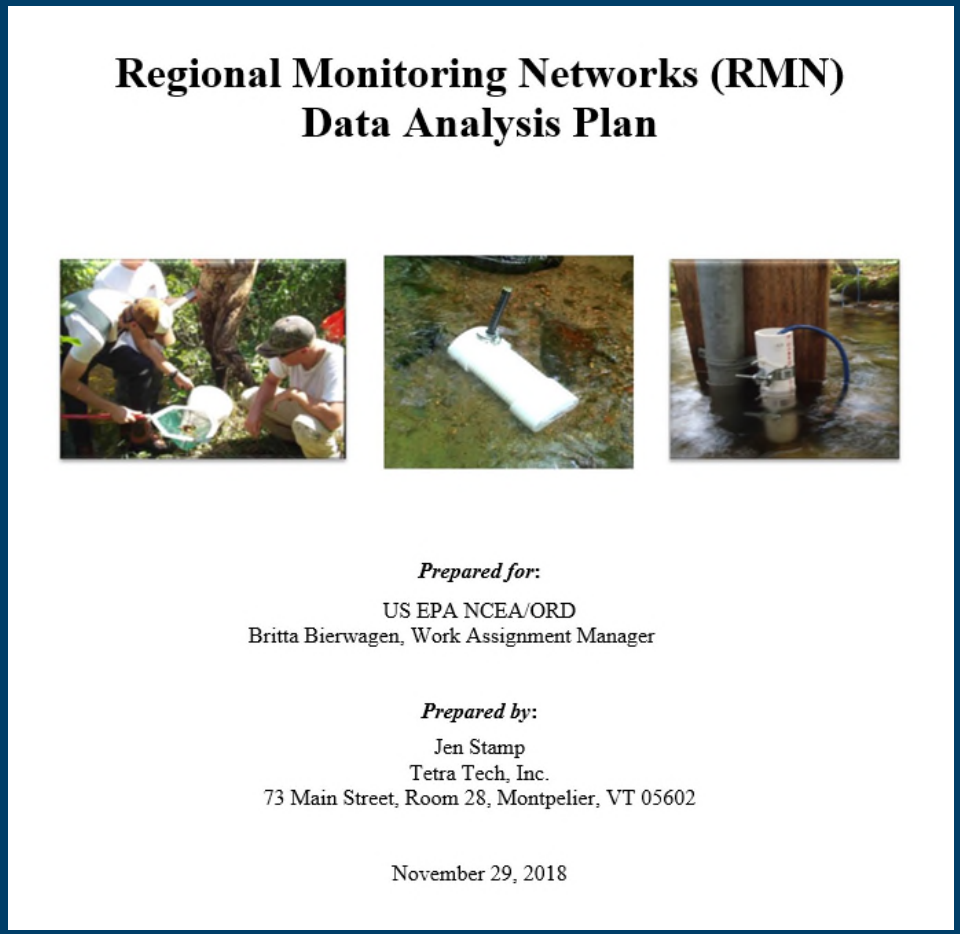
*averaged 1993-2011; based on *Isaak et al. 2015– NorWeST network*

Analyses

- Many possibilities
  - Correlation and regression analyses
  - Analysis of variance (ANOVA)
  - Non-metric multidimensional scaling (NMDS) ordinations
  - Visualization
    - Time series plots
    - Scatterplots
    - Box plots
    - Cumulative distribution function (CDF) plots

RMN Data Analysis Plan

- Guidance on
  - Characterizing current conditions and similarities/differences across sites
  - Trend analyses
  - Analyzing relationships between different types of data
  - Opportunistic evaluation of extreme events



Regional Monitoring Networks (RMN)
Data Analysis Plan

Seeking feedback

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

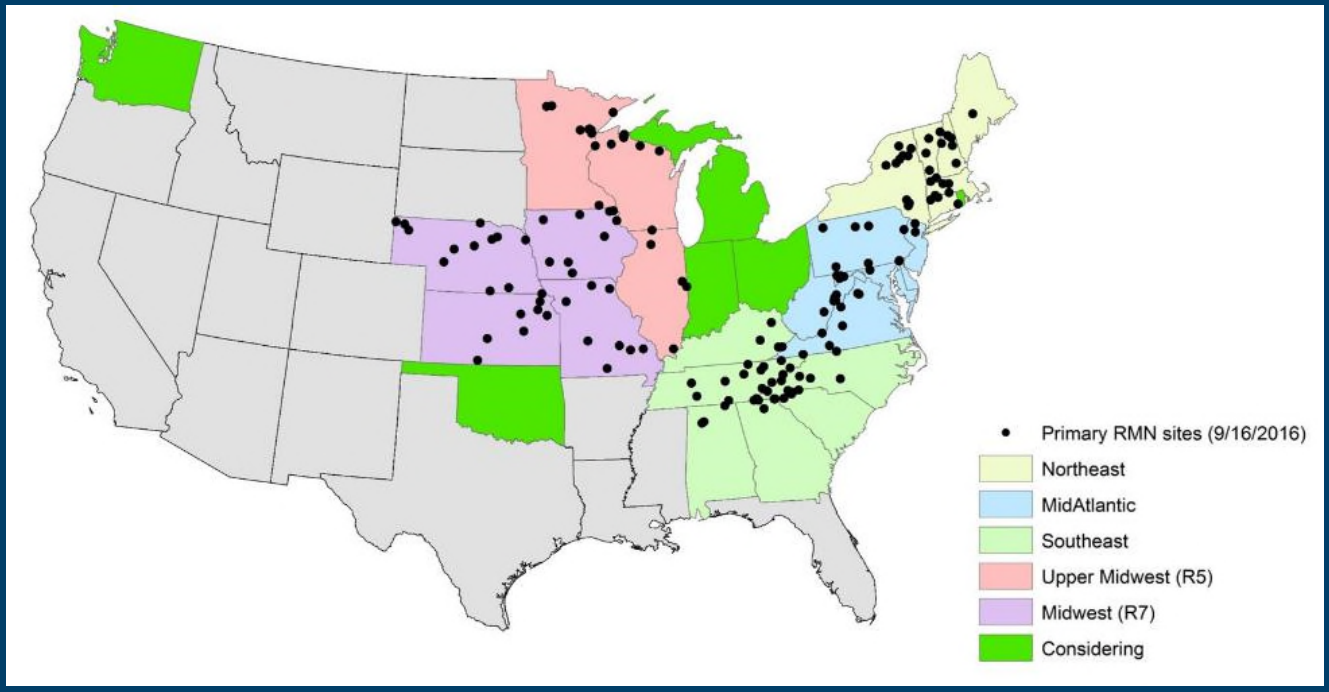
BioMonTools https://github.com/leppott/BioMonTools/

- Running into problems?
  - Please email us screenshots, copies of the specific error messages and input files
  - On GitHub, enterprising R users can log issues
- Can you think of ‘wish list’ items (improvements that would be ‘nice to have’)?

Are you aware of other similar efforts that we can learn from?

Acknowledgements

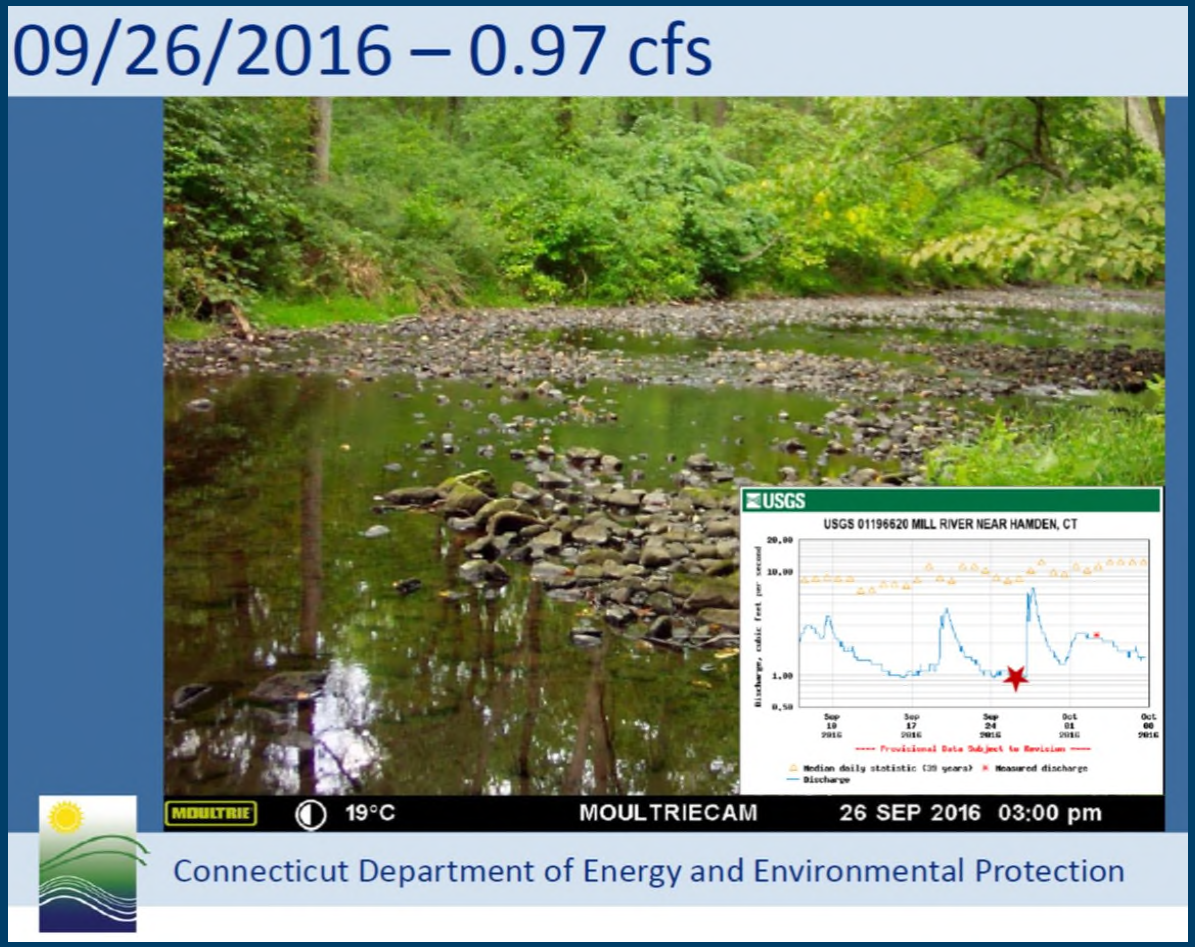
All of our RMN partners!!



QUESTIONS? COMMENTS?

Jen Stamp (Jen.Stamp@tetratech.com)
Erik W. Leppo (Erik.Leppo@tetratech.com)
Britta Bierwagen (bierwagen.britta@epa.gov)

‘Photo-rating curve’



Slide provided by Chris Bellucci

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