

# BIG DATAbases for Streams



**Part 1a: Designed databases**



# NAWQA

Stream Internet Workshop  
Boise, ID

*Daren Carlisle, Ecological Studies Coordinator,  
National Water-Quality Assessment Program,  
US Geological Survey*

**Website: <http://water.usgs.gov/nawqa/>**

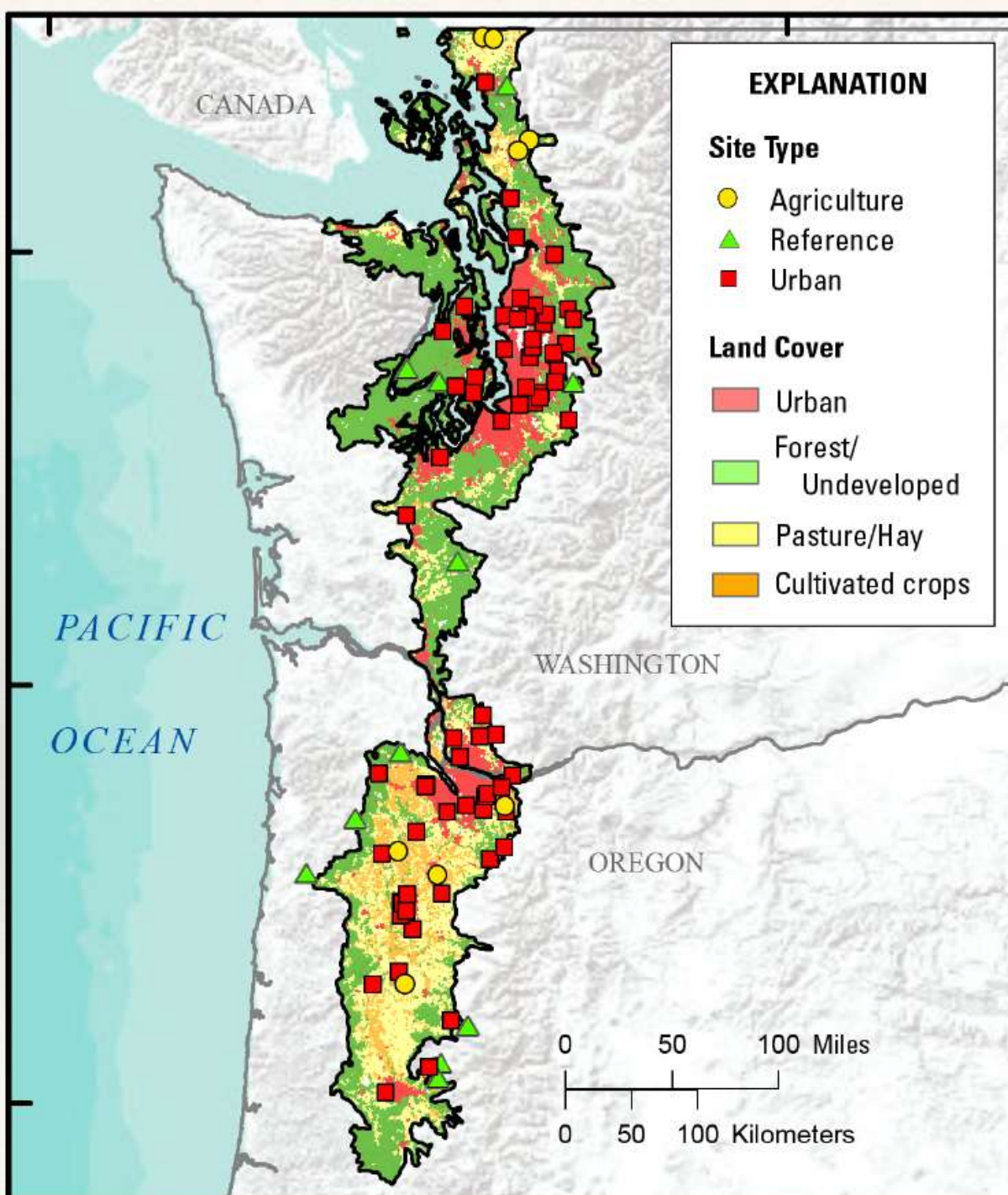
---

# NAWQA Regional Assessments

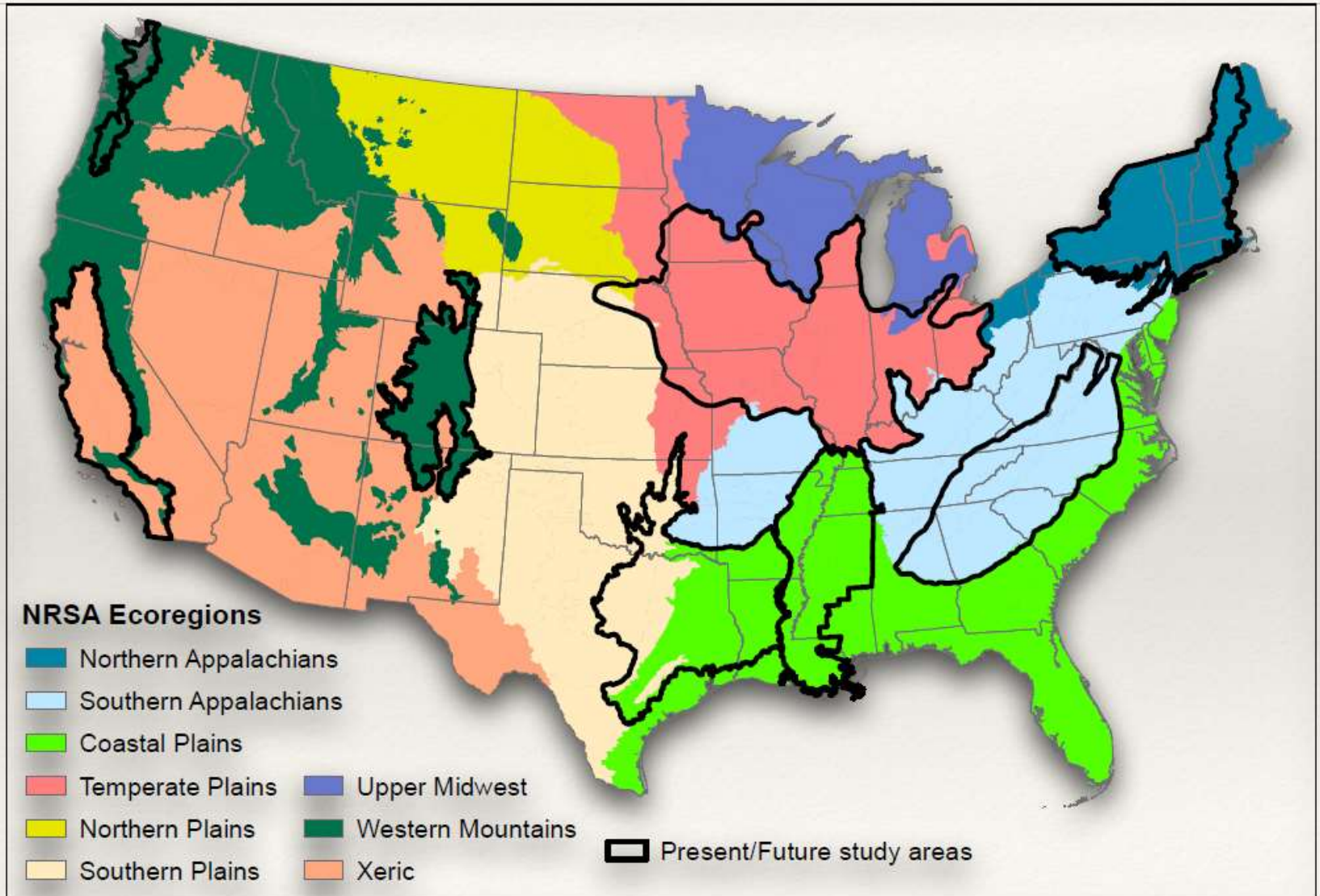
---

- ❖ GOAL: Assess stream health and the potential stressors that affect it
- ❖ 90-130 stream sites
- ❖ Gradient of anthropogenic stressors
- ❖ Characterize chemical & physical stressors for 10 weeks
- ❖ Characterize physical & biological conditions during week 10 of chemical sampling

# Pacific Northwest Stream-Quality Assessment



# Regional Assessments



## Water: Monitoring & Assessment

[Water Home](#)

[Drinking Water](#)

[Education & Training](#)

[Grants & Funding](#)

[Laws & Regulations](#)

[Our Waters](#)

[Drinking Water](#)

[Ground Water](#)

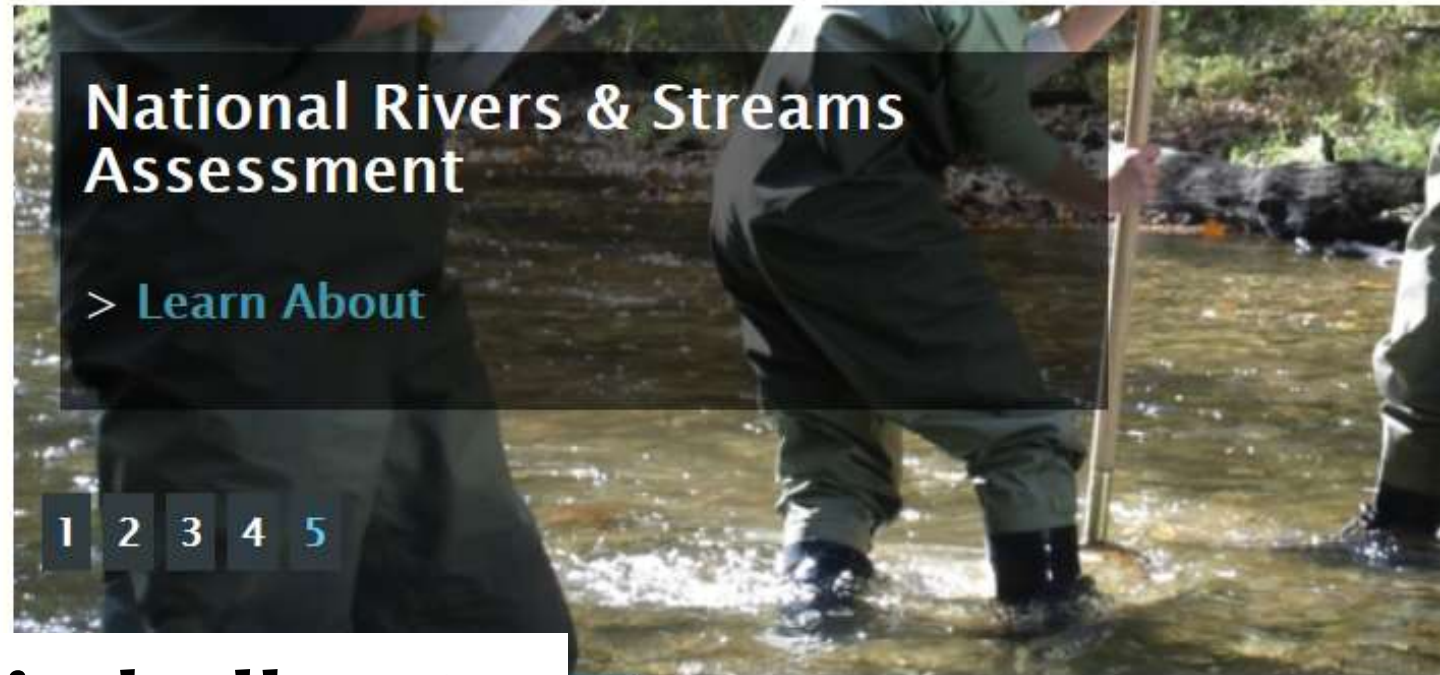
[Lakes](#)

[Oceans, Coasts,  
Estuaries & Beaches](#)

[Rivers & Streams](#)

You are here: [Water](#) » [Our Waters](#) » [Watersheds](#) » [Monitoring & Assessment](#) » [National Aquatic Resource Surveys](#)

# National Aquatic Resource Surveys



## National Rivers & Streams Assessment

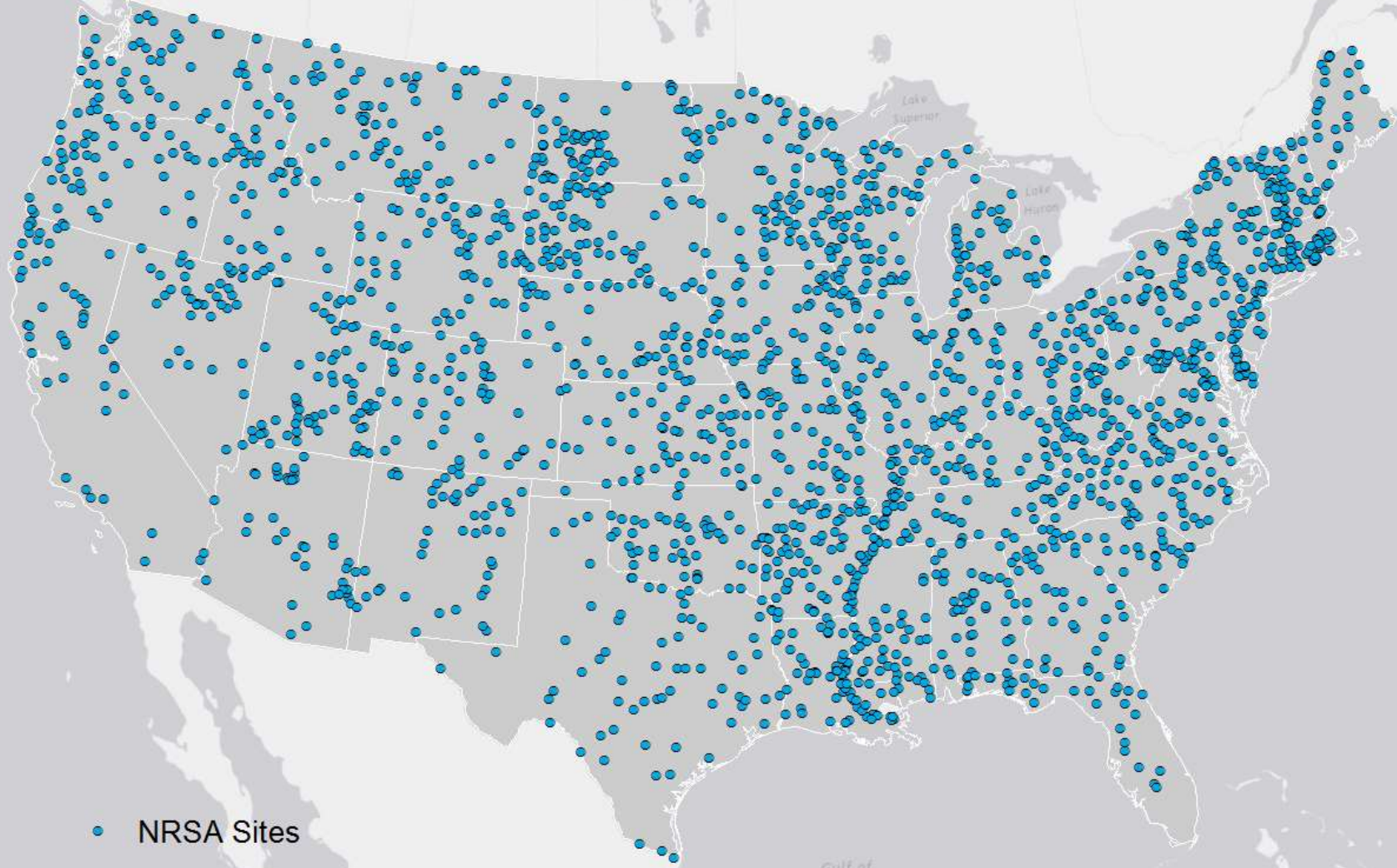
[> Learn About](#)

[1](#) [2](#) [3](#) [4](#) [5](#)

ands

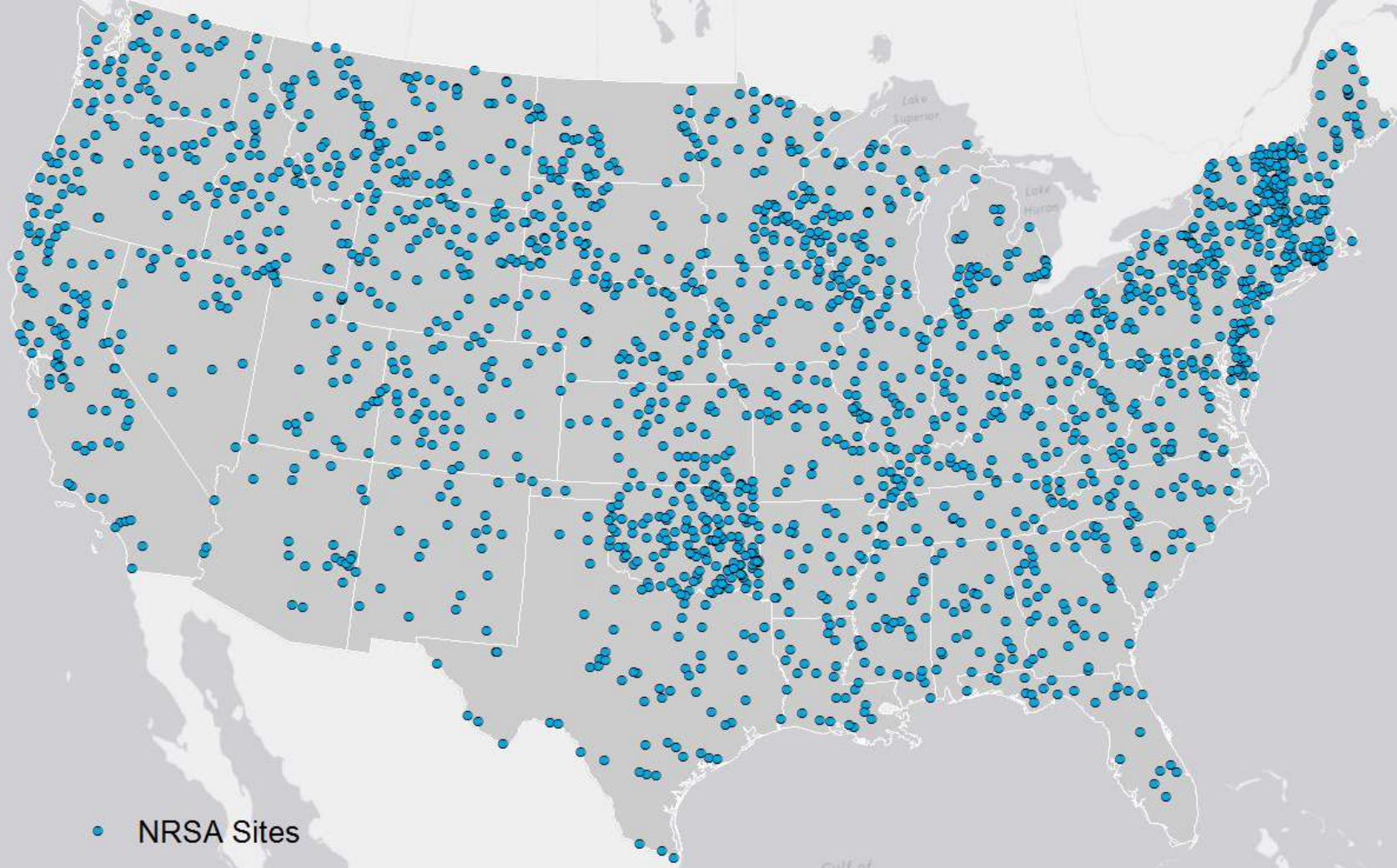
**Richard Mitchell - EPA**

# NRSA 2008/2009 Sample Locations



• NRSA Sites

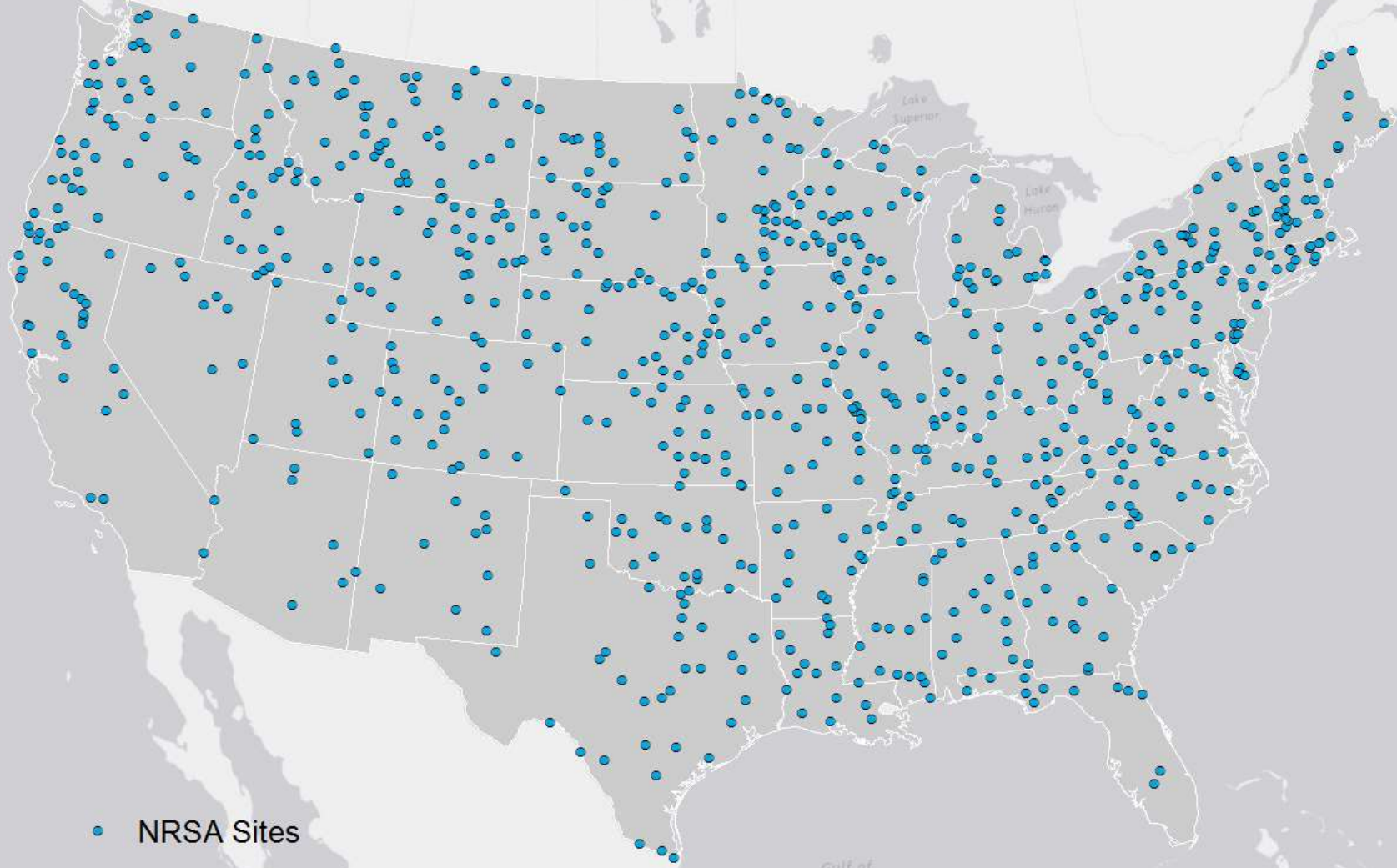
# NRSA 2013/2014 Sample Locations



• NRSA Sites



# NRSA Between Survey Resample Sites



• NRSA Sites

# BugLab & WMC

The screenshot shows the top portion of the BugLab website. On the left is the BugLab logo, a black silhouette of a dragonfly inside a water drop. In the center is a landscape photograph of a river flowing through a forested valley. On the right is the Utah State University logo. Below the header is a dark navigation bar with white text for 'HOME', 'SAMPLE PROCESSING', 'MONITORING RESOURCES', 'PROJECTS', 'PHOTOS', 'MEET THE STAFF', 'LINKS', and 'CONTACT US'. The main content area begins with the heading 'Welcome to the BugLab' and a section titled 'Our Mission' which contains a paragraph of text describing the center's cooperative work with Utah State University and the U.S. Bureau of Land Management.

This screenshot displays the Western Center for Monitoring & Assessment of Freshwater Ecosystems website. On the left is a circular logo with a map of the western United States in various colors. The main heading reads 'Western Center for Monitoring & Assessment of Freshwater Ecosystems'. Below this is a blue navigation bar with white text for 'Home', 'Bioassessments', 'Predictive Models', 'Data', 'Publications', 'Projects', and 'Workshop'. At the bottom of the screenshot, the text 'WMC / Bug Lab Database Queries' is visible.



## Scott Miller, Chuck Hawkins – USU/BLM

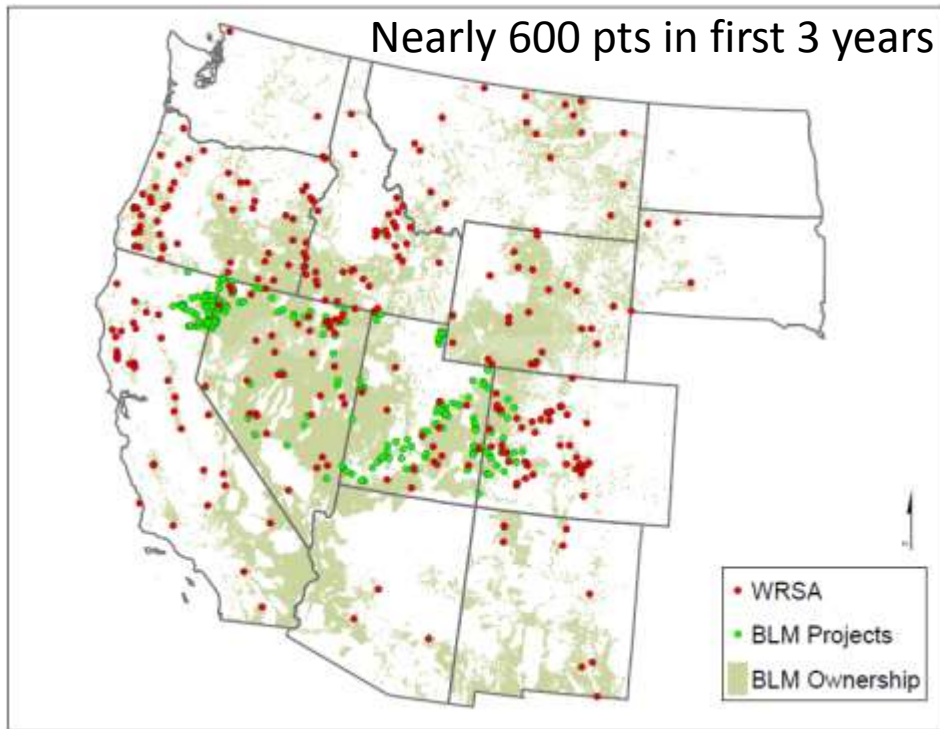
BugLab Website: <http://www.usu.edu/buglab/>

WMC Website: <http://www.cnr.usu.edu/wmc/>

# NAMF (National Aquatic Monitoring Framework)

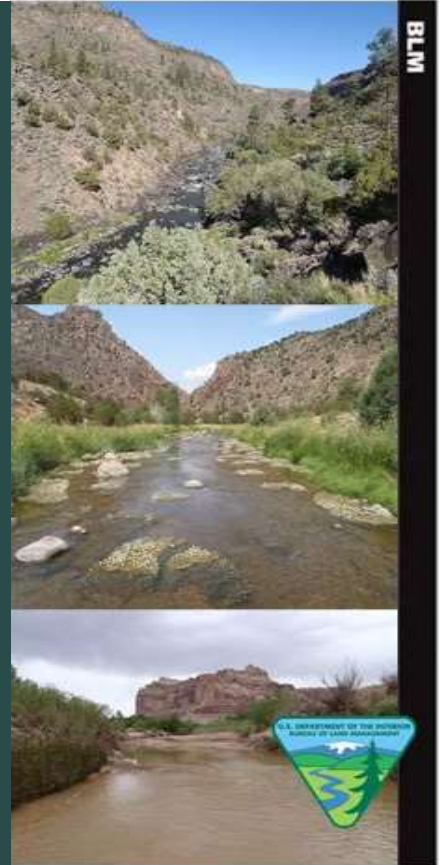
Lotic ecosystem monitoring (started 2013)

- Standardized methods for chemical, physical and biological indicators
- Statistically valid sample designs
- Collaborations with EPA, USFS, States



AIM National Aquatic Monitoring Framework: Introducing the Framework and Indicators for Lotic Systems

Technical Reference  
1735-1



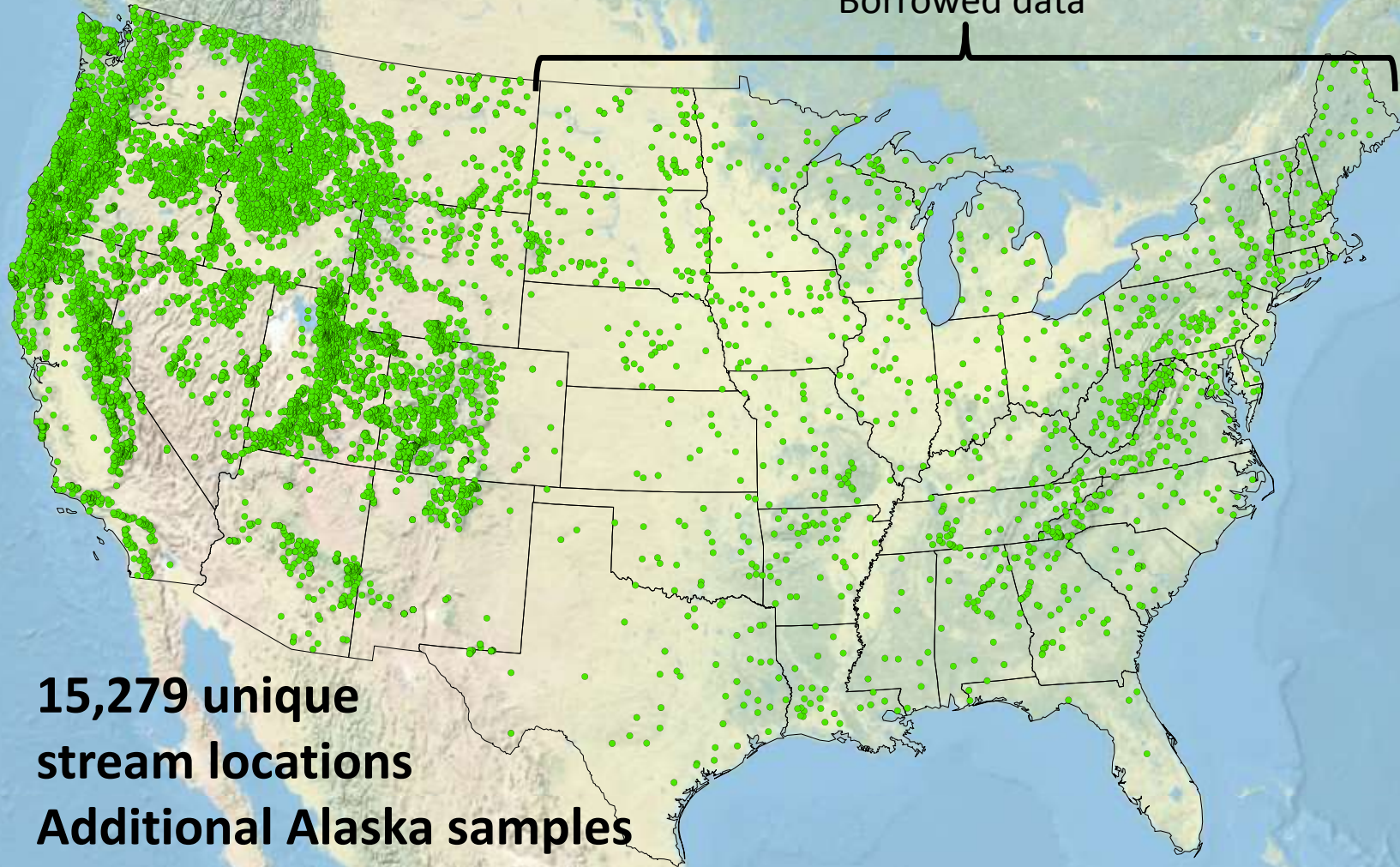
**Scott Miller (BLM/USU)**

**Website:** <http://www.usu.edu/buglab/Projects/CurrentProjects/>

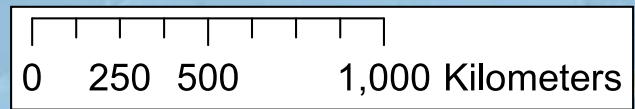
# WMC/BugLab Invertebrate Spatial Database

1.5 million records for 5,473 taxa from 49,450 samples

Borrowed data



15,279 unique  
stream locations  
Additional Alaska samples





Western Center for Monitoring & Assessment of Freshwater Ecosystems



BLM/USU National Aquatic Monitoring Center



BUGLAB

## MAPIT - a Mapping Application for Freshwater Invertebrate Taxa

Type in a taxon's name

Site Clusters  Pins

[Read Me First!](#)

Results 1 - 2000 of 4468

### ProjectName

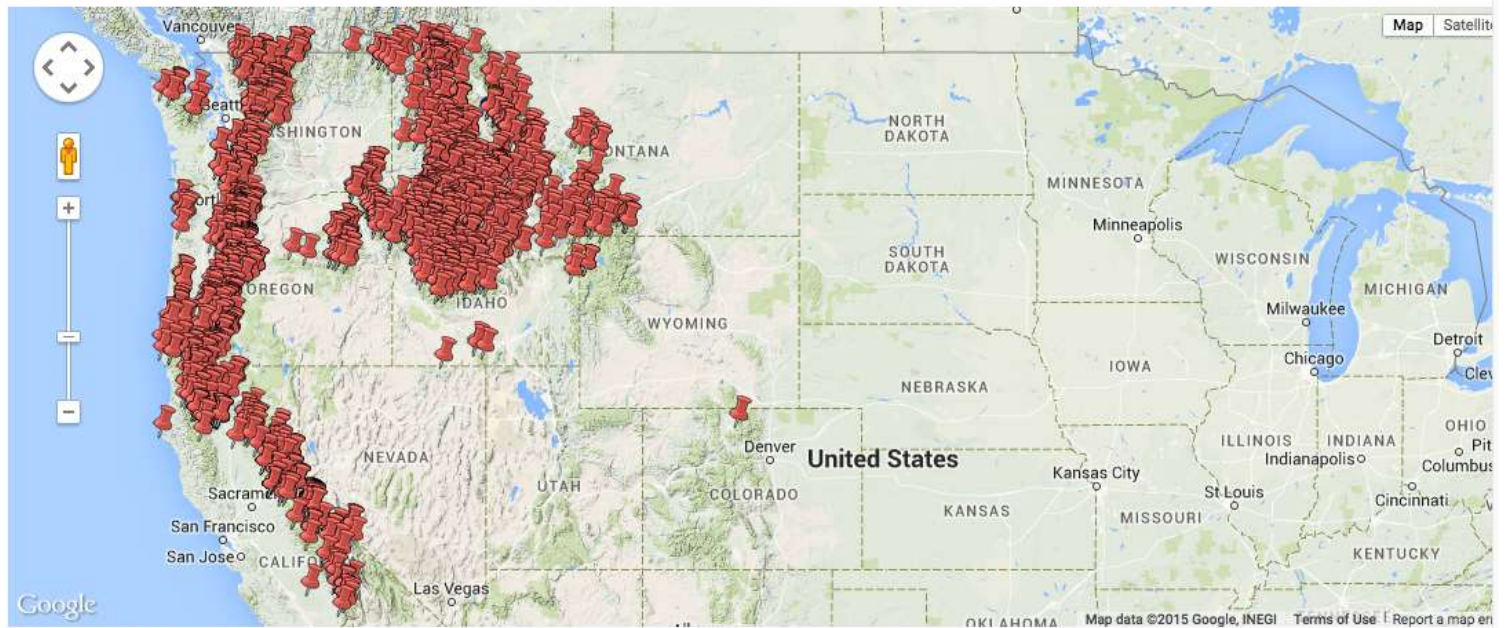
- + [AREMP \(422\)](#)
- + [CA-R5 \(91\)](#)
- + [EMAP-WEST \(304\)](#)
- + [HAWKINS-CWE \(139\)](#)
- + [NAOWA-REF \(41\)](#)
- + [PIBO \(1088\)](#)
- + [R6-USU-EAST \(10\)](#)
- + [R6-USU-WEST \(286\)](#)
- + [USU-STAR-I \(38\)](#)
- + [USU-STAR-II \(8\)](#)
- + [USU/BLM-BUGLAB \(2041\)](#)

### State

- + [CA \(1370\)](#)
- + [CO \(2\)](#)
- + [ID \(1034\)](#)
- + [MT \(595\)](#)
- + [NV \(17\)](#)
- + [OR \(998\)](#)
- + [UT \(1\)](#)
- + [WA \(430\)](#)
- + [WY \(21\)](#)

### SampleDate

- + [1981 \(1\)](#)
- + [1984 \(4\)](#)
- + [1987 \(4\)](#)
- + [1988 \(49\)](#)
- + [1989 \(82\)](#)
- + [1990 \(4\)](#)
- + [1991 \(8\)](#)
- + [1992 \(11\)](#)
- + [1993 \(19\)](#)
- + [1994 \(63\)](#)
- + [1995 \(33\)](#)
- + [1996 \(45\)](#)
- + [1997 \(75\)](#)
- + [1998 \(308\)](#)
- + [1999 \(226\)](#)
- + [2000 \(270\)](#)
- + [2001 \(249\)](#)
- [More ...](#)

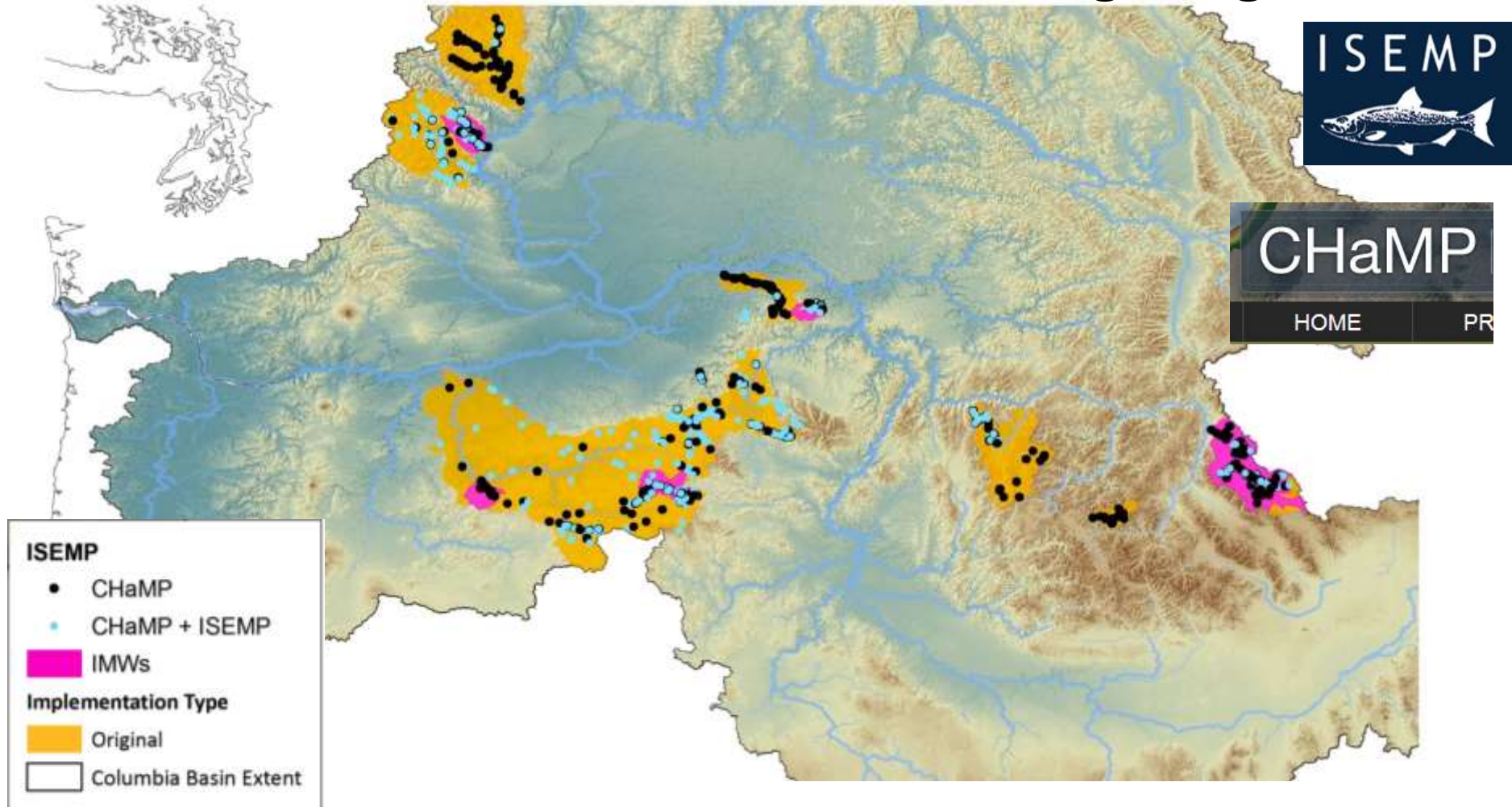


[Download all data](#)

Results Page: [1](#) [2](#) [3](#) [Next](#)

SampleId	Station	ProjectName	Method	Habitat	Area	LabSplit	Qualitative	Mesh	SampleDate
148461	CSNM_11	USU/BLM-BUGLAB	Surber net	Targeted Riffle	0.74	7.81	N		2011/08/29 00:00:00
151068	2013CAEIN1003	USU/BLM-BUGLAB	Surber net	Targeted Riffle	0.74	62.5	N		2013/06/25 00:00:00
151090	2013CAIND1008	USU/BLM-BUGLAB	Surber net	Targeted Riffle	0.74	100	N		2013/05/15 00:00:00
151098	2013CALIN1007	USU/BLM-BUGLAB	Surber net	Targeted Riffle	0.74	100	N		2013/07/03 00:00:00
151099	WCAP99-R037	USU/BLM-BUGLAB	Surber net	Targeted Riffle	0.74	100	N		2013/07/03 00:00:00
151100	2013CALIN1009	USU/BLM-BUGLAB	Surber net	Targeted Riffle	0.74	100	N		2013/07/03 00:00:00

# Columbia Habitat Monitoring Program & Integrated Status and Effectiveness Monitoring Program



Website: <https://www.champmonitoring.org/>

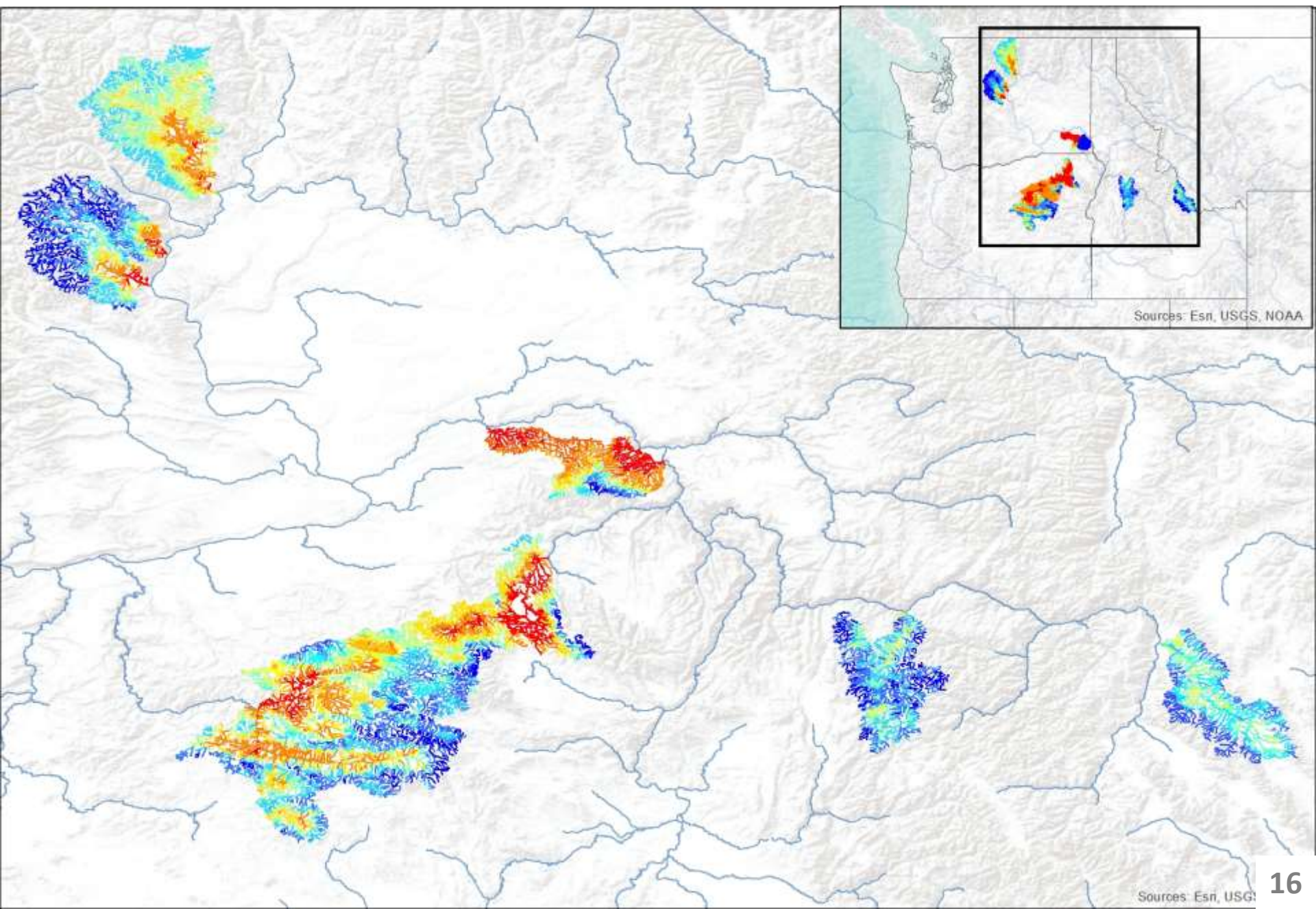
**Chris Jordan - NOAA**

Website: <http://isemp.org/>

# CHaMP and ISEMP

- CHaMP samples stream habitat conditions at ~500 unique locations in PNW, paired with ISEMP sampling stream fish (salmonids) at 250 CHaMP sample sites annually.
- Goals
  - Developing fish-habitat relationships across all CHaMP sampling sites
  - Developing generalized fish-habitat relationships across all CRB salmonid domain
  - Distributing metric data from sampling sites
- Need help thinking about
  - Distributing derived metrics
  - Distributing network-scale data products

# Spatially and Temporally Continuous Predictions of Stream Temperature



Sources: Esri, USGS, NOAA



Wenatchee



Asotin




Tucannon



Lemhi

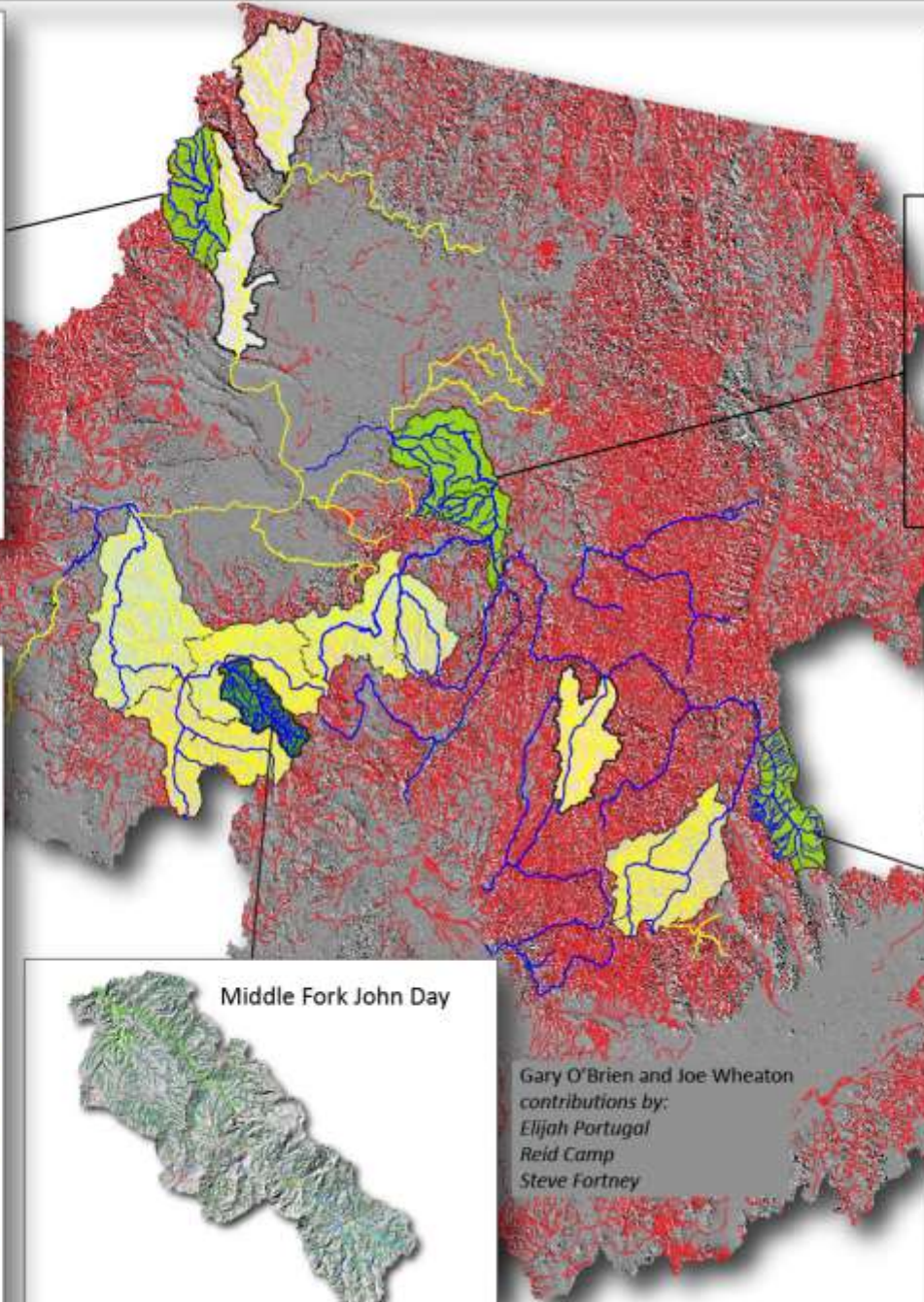


**CHaMP Basins**

- river style study completed 
- river style study in draft stage 
- river style study begun, data available 
- river style study planned 

**River Styles**

- river styles delineated for perennial streams 
- Chinook and steelhead stream network, river styles not yet completed 
- perennial streams of the Columbia River Basin slated for automated river styles delineation 



Middle Fork John Day



Gary O'Brien and Joe Wheaton  
contributions by:  
Elijah Portugal  
Reid Camp  
Steve Fortney

[National Oceanic and Atmospheric Administration](#) | [Watershed: Upper Grande...](#) | [https://www.champmonitoring.org/Watershed/Details/12#tab-visits](#)

[NWFS](#) | [Inside NWFS](#) | [ISEMP](#) | [Google Maps](#) | [Doodle](#) | [SDM\\_WebApps](#) | [Calendar](#) | [CHAAMPMonitoring](#) | [Dropbox](#) | [skcd](#) | [AmeriGas](#) | [Weather](#) | [TripCheck](#) | [NWFS Weekly](#)

[Home](#) > [Watersheds](#) > [Upper Grande Ronde](#) > [Details](#)

## Watershed: Upper Grande Ronde (ID: 12)

[Reset Map](#) | [Hide Map](#)

[Overview](#) | [Study Design](#) | [Field Support](#) | **Visits** | [Measurements](#) | [Metrics](#) | [Status](#)

Year: 2014

### Visits Tab

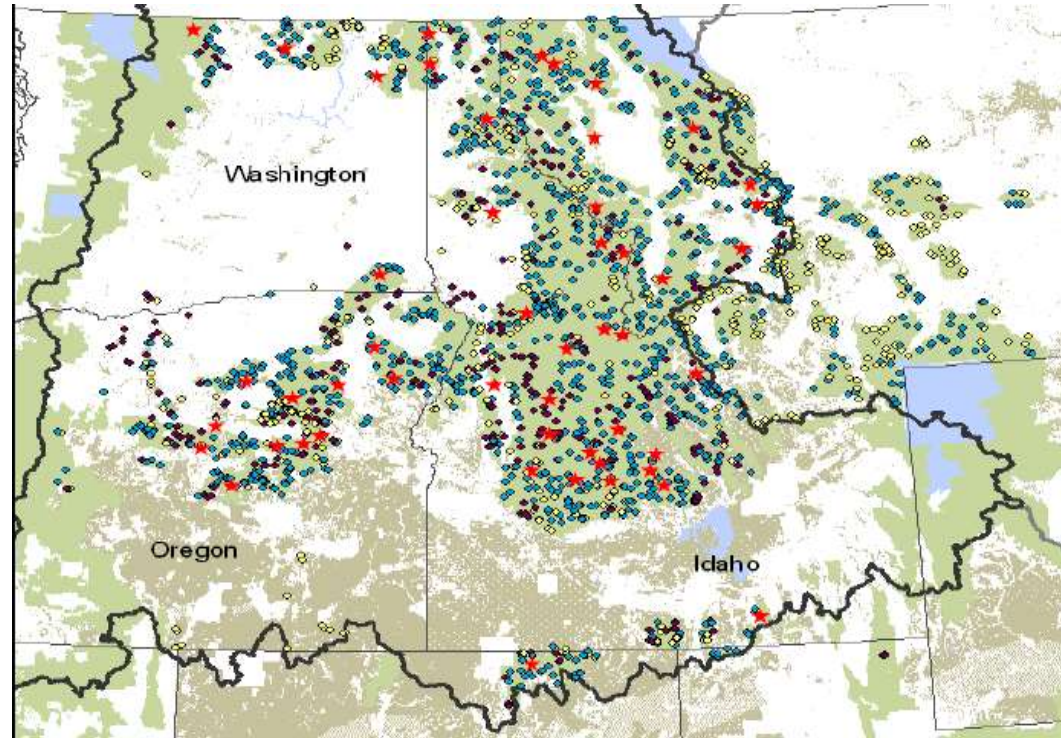
Below are the site visits where monitoring data has been received and loaded to the database in preparation for the quality assurance (QA) process. [show more](#)

Currently viewing 75 of 75 visits [Download Data](#)

Rel#	Site ID	Sample Date	Organization	Use Order	Protocol	Hitch Date	Hitch	Crew	Visit Phase	Visit Status	Visit ID	Panel	Category	Stream Name	Edit Purpose of Visit	CHAAMP Core
	CBW05583-813862	06/30/2014	Oregon Department	1	CHAAMP 2014	06/15/2014	COFW Small Sites	Chris Horn Crew	Data Approved	Released to Public	2272	Annual	Grande Ronde Stee	Peel Creek		Yes
	CBW05583-880282	07/15/2014	Oregon Department	10	CHAAMP 2014	07/01/2014	COFW Big Sites	Chris Horn Crew	Data Approved	Released to Public	2142	Rotating Panel 1	Catherine Creek Chi	Catherine Creek		Yes
	CBW05583-882988	07/15/2014	Oregon Department	1	CHAAMP 2014	05/09/2014	COFW Medium Streams	Chris Horn Crew	Quality Assurance	In Q/A	2136	Rotating Panel 1	Grande Ronde Stee	Fly Creek		Yes
	CBW05583-885642	06/17/2014	Oregon Department	1	CHAAMP 2014	06/15/2014	COFW Small Sites	Chris Horn Crew	Data Approved	Released to Public	2123	Annual	Grande Ronde Stee	McCoy Creek		Yes
	CBW05583-108518	06/18/2014	Columbia River Inter	10	CHAAMP 2014	06/18/2014	CRITFC Upper Grande Ronde	Monica Crew	Data Approved	Released to Public	2014	Rotating Panel 1	Upper Grande Ronde	Linber Jim Creek		Yes
	CBW05583-142496	07/14/2014	Oregon Department	2	CHAAMP 2014	06/26/2014	COFW Medium Streams	Chris Horn Crew	Data Approved	Released to Public	2137	Annual	Grande Ronde Stee	Clark Creek		Yes
	CBW05583-148976	07/09/2014	Columbia River Inter	8	CHAAMP 2014	06/18/2014	CRITFC Upper Grande Ronde	Monica Crew	Data Approved	Released to Public	2015	Rotating Panel 1	Upper Grande Ronde	Grande Ronde River		Yes
	CBW05583-138815	07/09/2014	Oregon Department	2	CHAAMP 2014	06/15/2014	COFW Small Sites	Chris Horn Crew	Data Approved	Released to Public	2188	Annual	Grande Ronde Stee	Gordon Creek		Yes
	CBW05583-130554	06/21/2014	Columbia River Inter	2	CHAAMP 2014	06/18/2014	CRITFC Upper Grande Ronde	Monica Crew	Data Approved	Released to Public	2016	Rotating Panel 1	Upper Grande Ronde	Sheep Creek		Yes
	CBW05583-138868	09/18/2014	Columbia River Inter	6	CHAAMP 2014	06/18/2014	CRITFC Catherine Creek 2014	Monica Crew	Data Approved	Released to Public	2064	Rotating Panel 1	Catherine Creek Chi	North Fork Catharhu		Yes
	CBW05583-150018	09/22/2014	Columbia River Inter	8	CHAAMP 2014	06/18/2014	CRITFC Catherine Creek 2014	Monica Crew	Data Approved	Released to Public	2005	Rotating Panel 1	Catherine Creek Chi	Little Catherine Cree		Yes
	CBW05583-149584	06/30/2014	Oregon Department	3	CHAAMP 2014	06/15/2014	COFW Small Sites	Chris Horn Crew	Data Approved	Released to Public	2124	Rotating Panel 1	Grande Ronde Stee	Dark Canyon Creek		Yes

# PIBO (PacFish/InFish Biological Opinion)

## AREMP (Aquatic & Riparian Effectiveness Monitoring Program)



Rotating panel designs

250 sites/year

1,500 sites total

- Habitat
- Macroinvertebrates
- Stream temperature

Jeff Kershner, Brett Roper, Stephanie Miller (USFS)

PIBO Website: <http://www.fs.fed.us/biology/fishecology/emp/index.html>

AREMP Web: <http://reo.gov/monitoring/reports/watershed/aremp/Welcome.htm>

# BIG DATAbases for Streams



**Part 1b: Aggregated databases**

# MARIS (Multistate Aquatic Resources Information System)



**MARIS**  
Multistate Aquatic Resources Information System

HOME ABOUT DOWNLOAD WEB SERVICES SEARCH BROWSE MAP CONTRIBUTE CONTACT

The Multistate Aquatic Resources Information System (MARIS) is an online resource that contains over one million fish sampling and water quality records for over one thousand fish species compiled from state fish and wildlife agency and other sampling programs. *(Learn More)*

UNITED STATES MEXICO

bing

MARIS Status map - click on a state for more information

Legend:  
■ Data Available  
■ Data Available (no lat/long)  
■ Data Coming Soon

USDA NRCS  
U.S. Department of Agriculture  
Natural Resources Conservation Service

**Download**  
Download the entire MARIS database or selected states in various formats.

**Browse Map**  
View a clickable map of every lake and stream surveyed by the MARIS partner agencies.

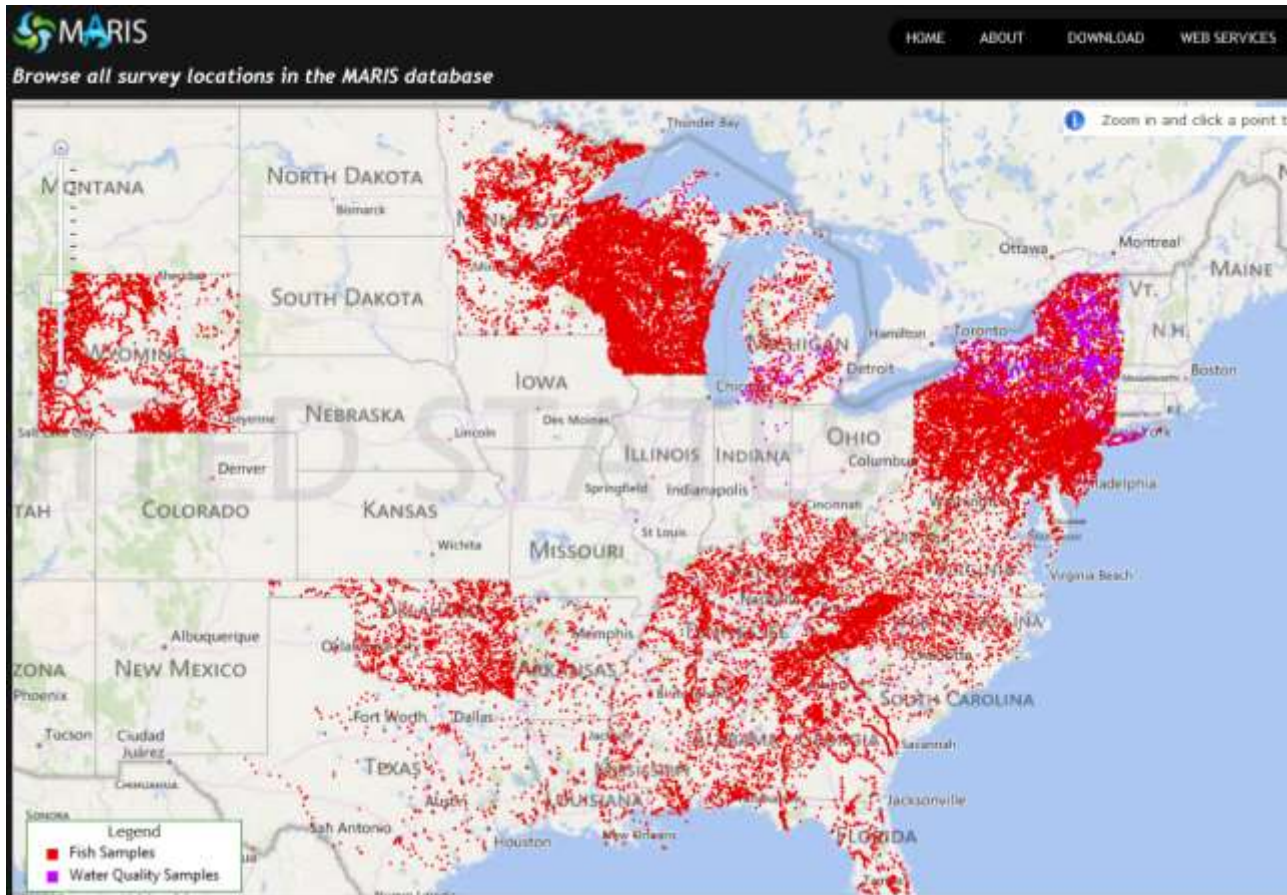
**Search Data**  
Search the MARIS database using several parameters and view the results.

**Andy Loftus (Natural Resources Consultant)**

Website: <http://www.marisdata.org/>

# MARIS (Multistate Aquatic Resources Information System)

“contains over one million fish sampling and water quality records for over one thousand fish species compiled from state fish and wildlife agency and other sampling programs.”



Andy Loftus (Natural Resources Consultant)

Website: <http://www.marisdata.org/>

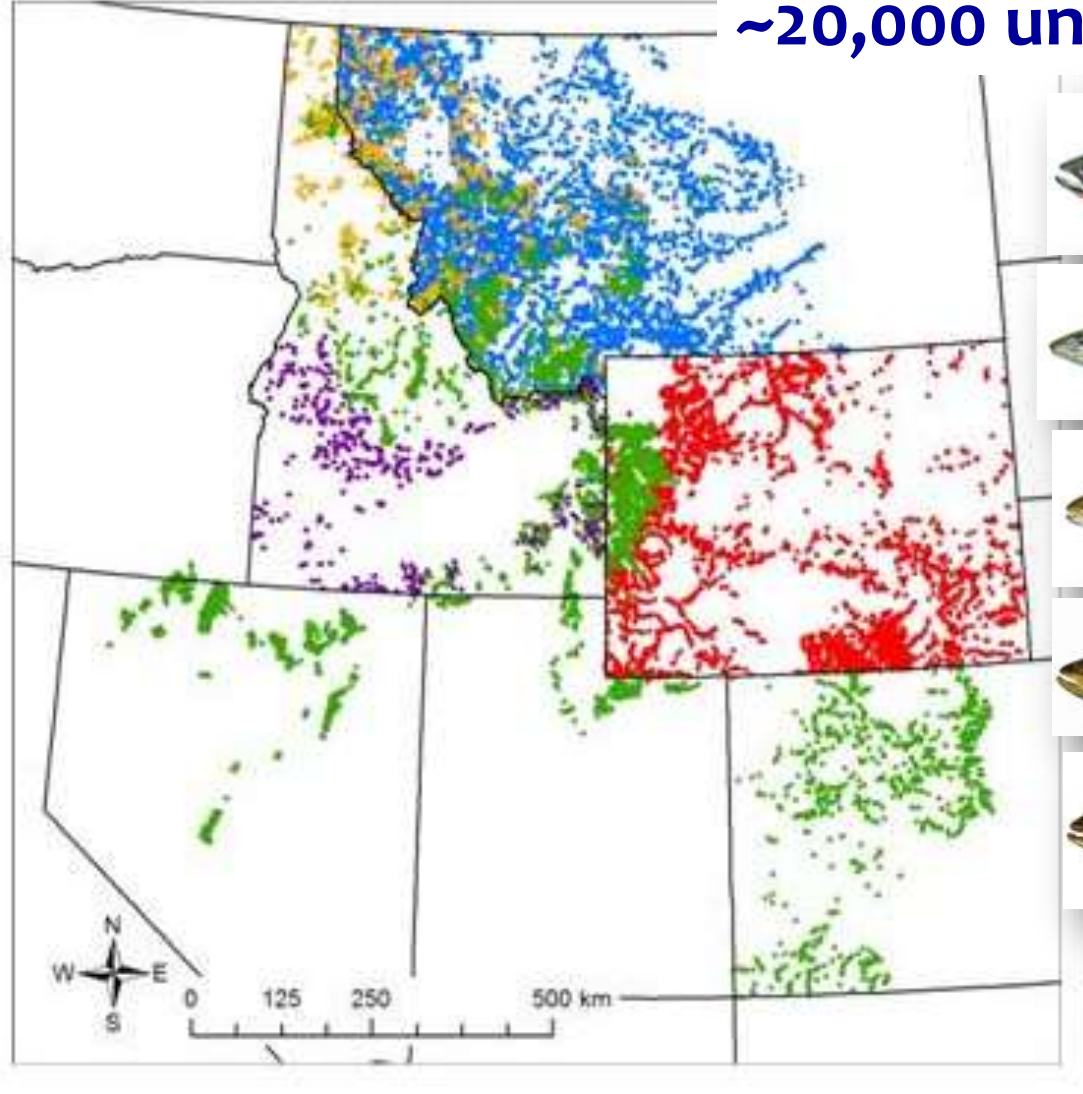
# Western Trout Interagency Database Compilation



Montana Fish,  
Wildlife & Parks



~20,000 unique sites

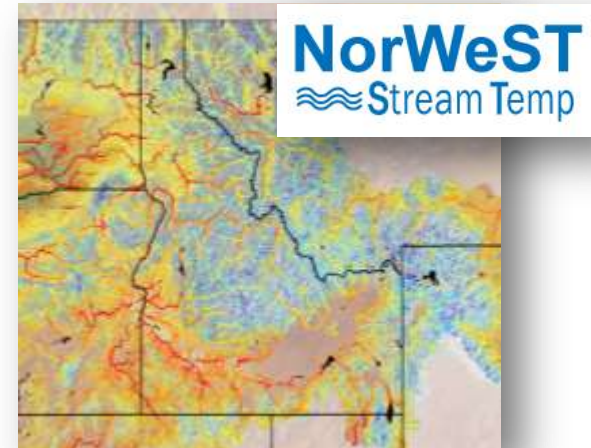
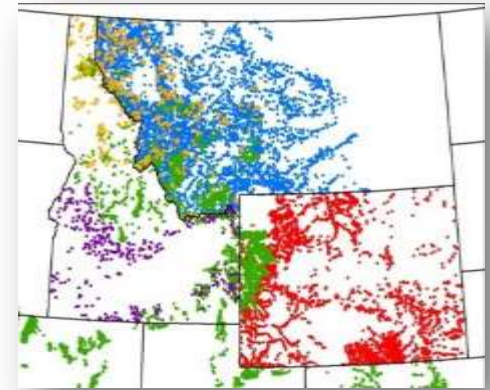
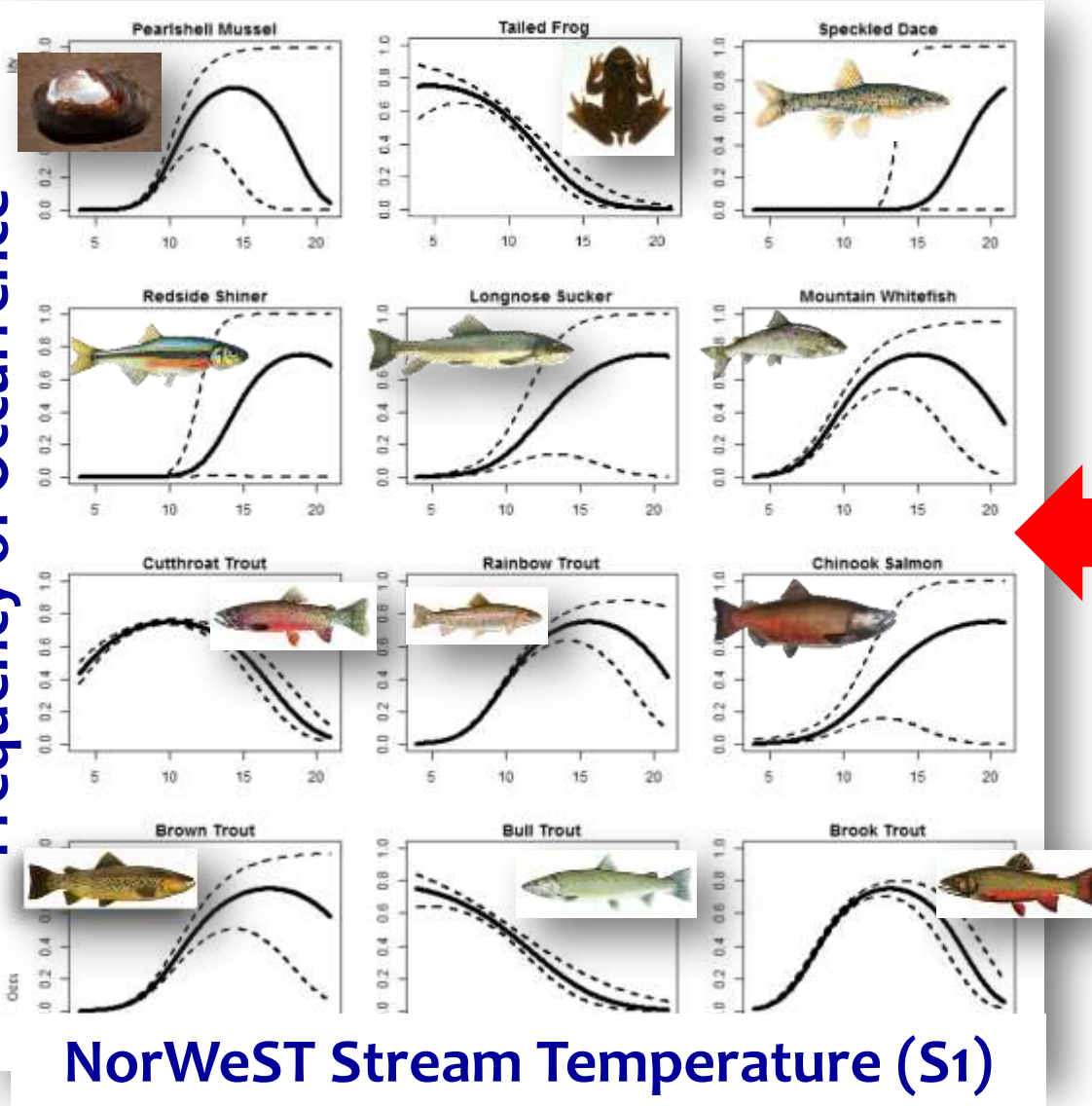


Seth Wenger

(Wenger et al. 2011a; 2011b; In preparation)

# Thermal Criteria in Batch Mode... **BIG FISH**

Frequency of Occurrence

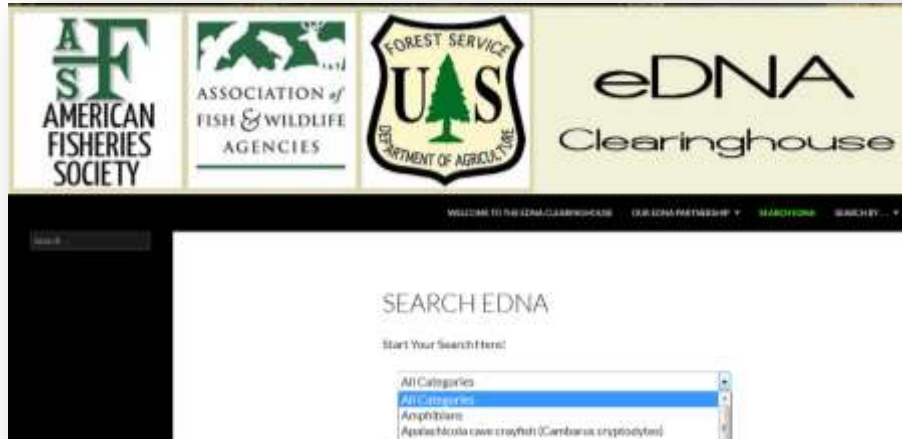


Wenger et al. *In Preparation*. Description of realized thermal niches using massive biological and temperature databases.

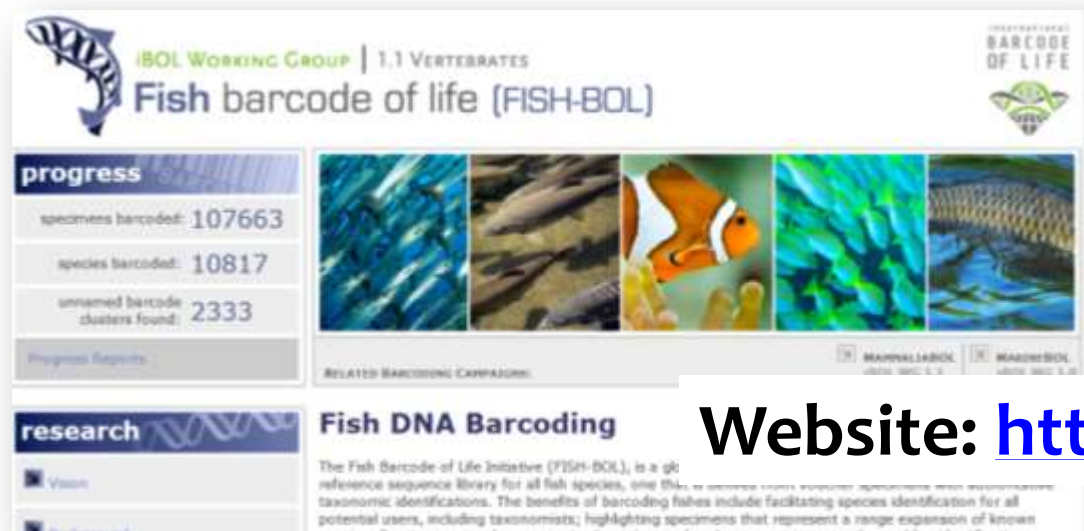


# Genomics Frontier

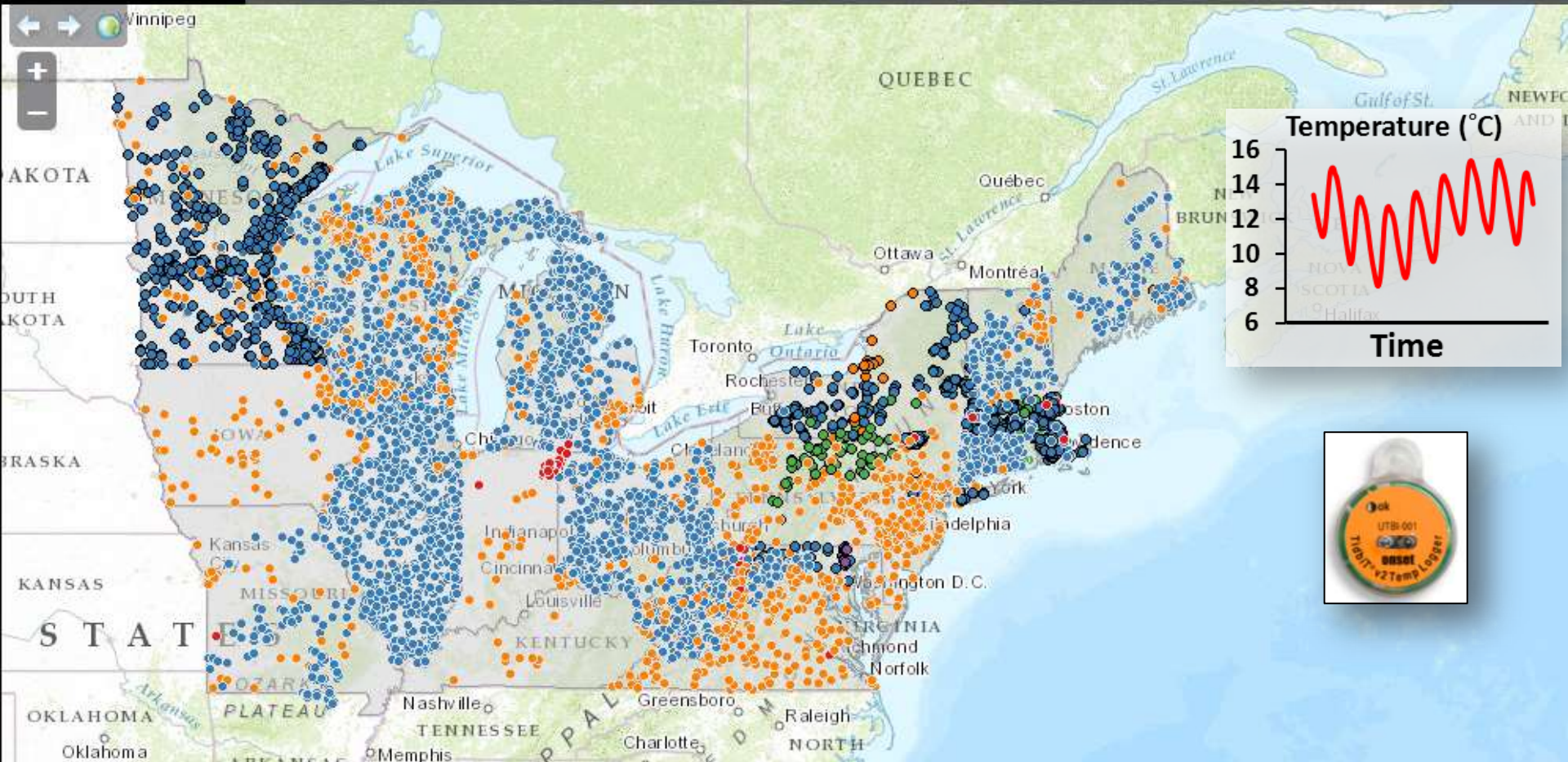
## eDNA, DNA Barcoding, etc.



Website: <http://edna.fisheries.org/>



Website: <http://www.fishbol.org/>



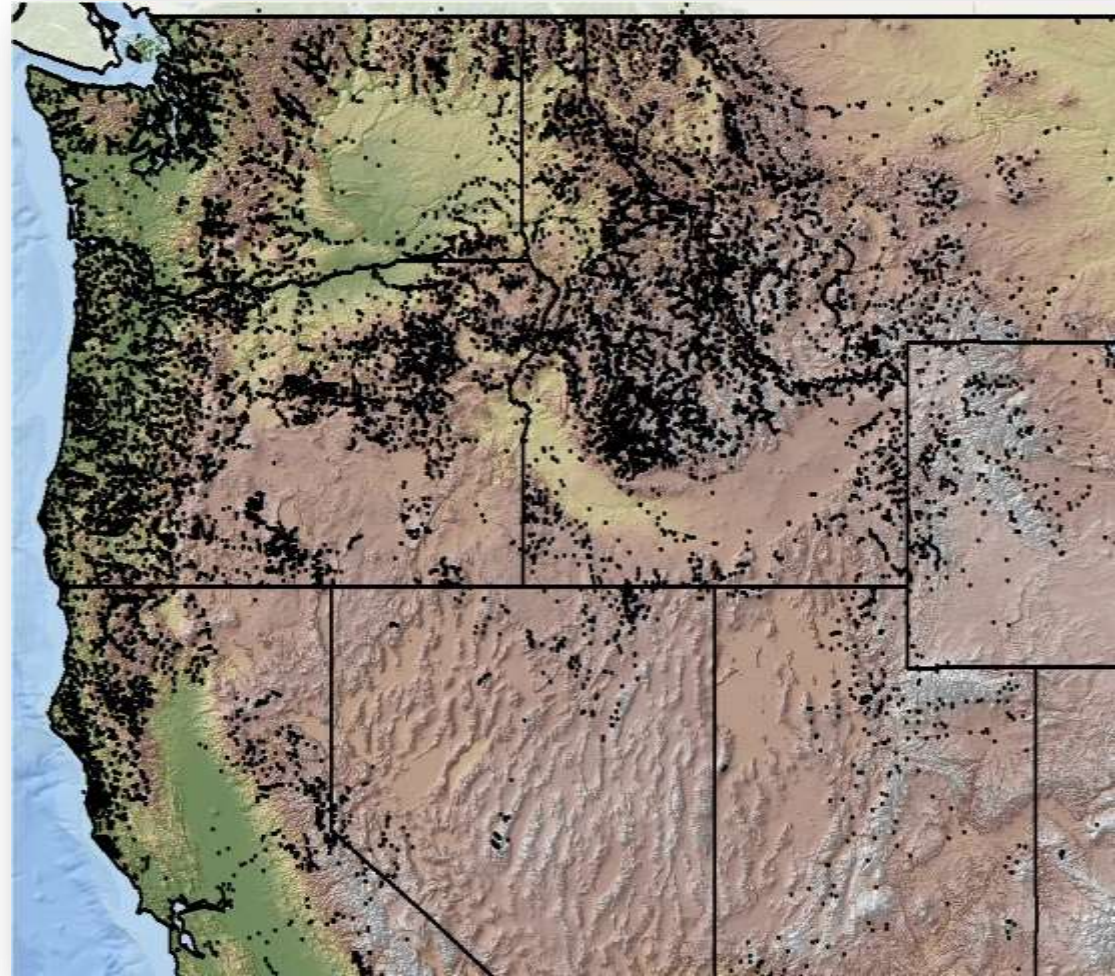
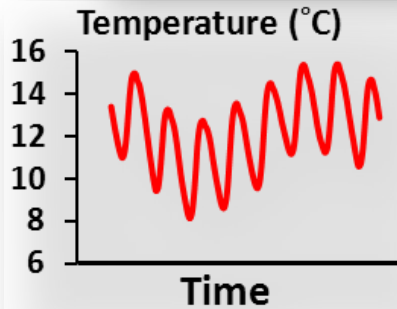
**Dana Infante**

Website: <http://wim.usgs.gov/NorEaST/>

# NorWeST

Stream Temp

>50,000,000 hourly records  
>15,000 unique stream sites  
>80 resource agencies

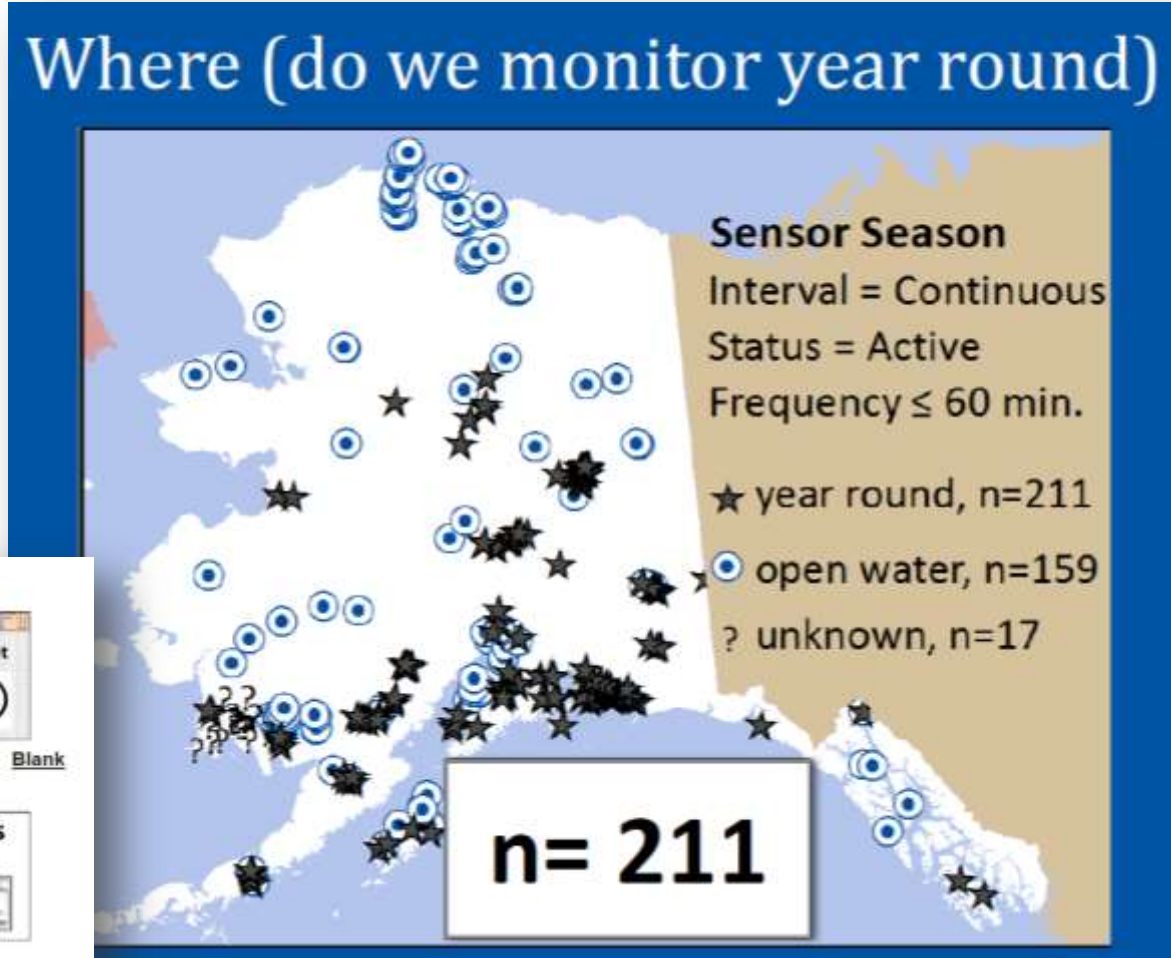
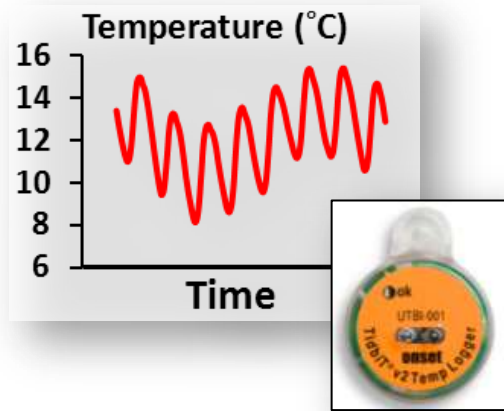


**Dan Isaak**

Website:

<http://www.fs.fed.us/rm/boise/AWAE/projects/NorWeST.html>

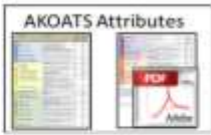
# AKOATS (Alaska Online Aquatic Temperature Sites)



## AKOATS Products



Temperature Monitor Web Map  
Data Table



Metadata Attributes



Blank



AKOATS in ArcMap  
Instructions: Enter New Data



AKOATS Report



Joel Reynolds – FWS, Western Alaska LCC



# ALASKA'S WATER

---

> 3,000,000  
Mapped Lakes and Ponds

---

> 847,000  
Length of Mapped Streams and Rivers in miles

---

47,000  
Length of Mapped Coastline in miles

---

1,980  
Length of the longest river, the Yukon, in miles

---

616  
Named Glaciers

---

40%  
Share of the Nation's Surface Water

**Lucinda - Lakes?**

**Dan Magoulick – SE additions?**

# Update: Open Water Data Initiative

Stream Internet Workshop

April 23, 2015

Al Rea, USGS National Geospatial Program

Ed Clark, NOAA National Weather Service

Subcommittee on Spatial Water Data

## ***Open Water Data Initiative***

Water Data Catalog	Water Data As a Service	Enriching Water Data	Water Data and Tools MarketPlace
Find Source Data	Consensus standards	River routing	Community exercise of tools & data
Create water & climate themes	Water Map Themes	Coupling with models	Data usage tracking
Recruit/engage partners	High performance data delivery	Grounded to geofabric	Community-built extensions



***Technical: National Water Data Infrastructure***

***Social: Open Water Web***



# OWDI Activities To Date

- **FGDC** Steering Committee (6/26) and **ACWI** (8/19) voted unanimously to revitalize and charge the Subcommittee on Spatial Water Data to scope and design a national Open Water Data Infrastructure
- **AWRA**
  - National Meeting (Nov 3-6, 2014) special track on OWDI
  - JAWRA special collection OWDI/NFIE (source: <http://www.awra.org/jawra/jawra-owdi-call.html>)
  - National Meeting, Denver (Nov 16-19, 2015) proposed special track
- **Subcommittee on Spatial Water Data (SSWD)**
  - Monthly meetings since August
  - 40+ regular attendees, mailing list of 80+
  - > 30 organizations represented
  - Applying “Lean Startup” principles
  - Three initial use cases
  - Four data work groups
  - Technology/standards work group

# OWDI Use-Case Working Groups



## Work Group 1: National Flood Interoperability Experiment

Identify flood data including stream-flow observations, forecasts and impacts  
Developing *Hydrofabric*\* v 0.1 and exploring data conflation

\*Supported by 3 sub-teams



## Work Group 2: Drought Decision Support System

Identify water resources data including natural flow, reservoir storage and drought impacts  
Explore visualization of drought in Lower Colorado



## Work Group 3: Spill Response Tool

Identify water quality data including potential points sources and impacts  
Exploring requirements for new/additional data (e.g. velocity forecasts and reservoir residence times)

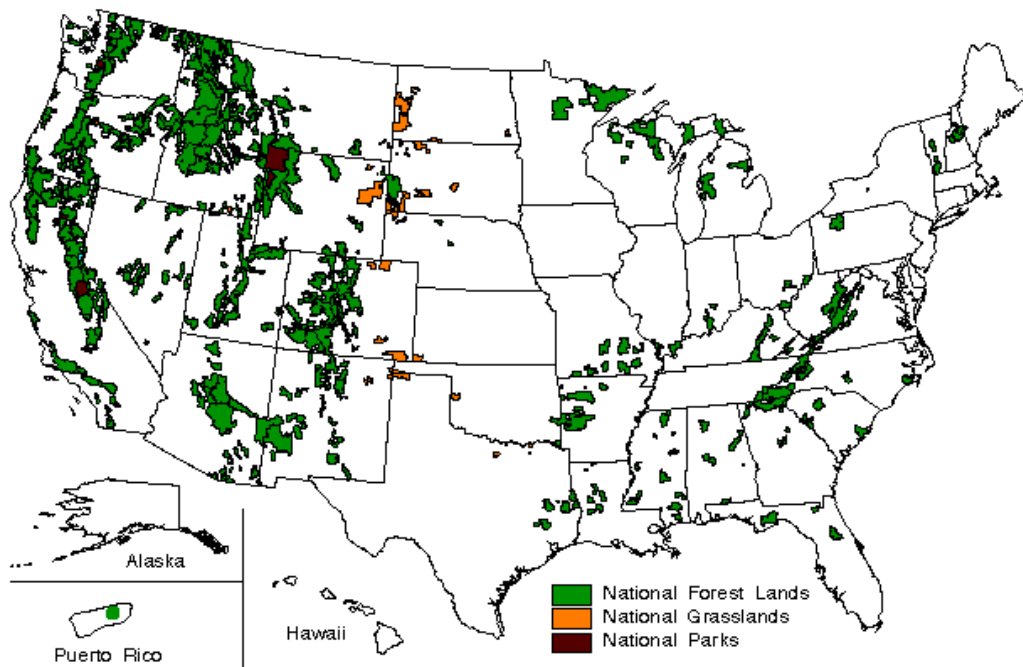
# Hydro Event Services

- Hydro Event Data Catalog (HEDC)
  - Open – federated catalog
  - Web-service based
  - Starting with USGS ScienceBase platform for ESRI services, USGS CIDA group will host open (GeoServer) services
  - Intent is to complement, not replace WATERS web services
  - Iterative approach, will seek feedback



# Aquatic Surveys Module in Natural Resource Monitor

## Corporate database for National Forests & Grasslands



## Survey Type Modules

- ❖ Stream temperature
- ❖ Fish, amphibian
- ❖ Reach habitat
- ❖ Passage conditions

A screenshot of the Natural Resource Monitor software interface. The interface shows a list of survey data on the left and a detailed data table on the right. The table has columns for 'Station Name', 'Date', 'Start Date', 'End Date', and 'Status'. The data rows show various survey stations and their corresponding dates and statuses.

**Callie McConnell,  
Brian Sanborn**

Website: <http://www.fs.fed.us/nrm/index.shtml>

# BIG DATAbases for Streams



## Part 2: Environmental predictors

## A Hierarchical Spatial Framework and Database for the National River Fish Habitat Condition Assessment

Lizhu Wang

Institute for Fisheries Research, Michigan Department of Natural Resources and University of Michigan, 212 Museums Annex, Ann Arbor, MI 48109. E-mail: wangl@michigan.gov

Dana Infante

Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824

Peter Esselman

Institute for Fisheries Research, Michigan Department of Natural Resources and University of Michigan, 212 Museums Annex, Ann Arbor, MI 48109; and Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824

Arthur Cooper

Institute for Fisheries Research, Michigan Department of Natural Resources and University of Michigan, 212 Museums Annex, Ann Arbor, MI 48109; and Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824

Dayong Wu

Department of Life Science, Hengshui University, Hengshui, Hebei, China 053000

William Taylor

Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824

Doug Beard

National Climate Change and Wildlife Science Center, U.S. Geological Survey, 12201 Sunrise Valley Drive, Reston, VA 20192

Gary Whelan

Fisheries Division, Michigan Department of Natural Resources, 530 West Allegan, Lansing, MI 48909

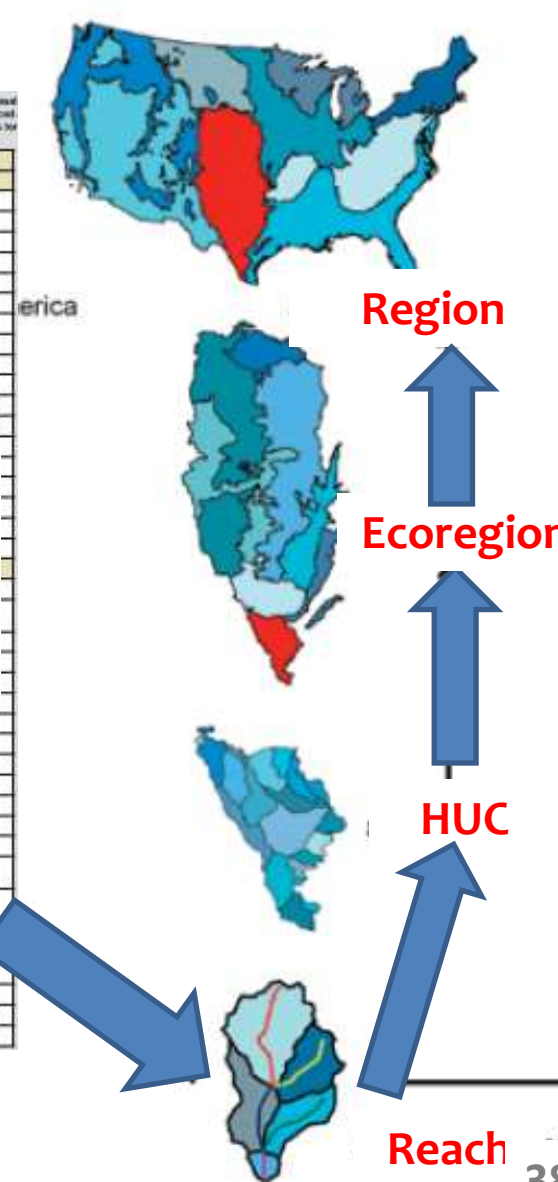
Andrea Ostroff

National Climate Change and Wildlife Science Center, U.S. Geological Survey, 12201 Sunrise Valley Drive, Reston, VA 20192

TABLE 1. Summary of natural and human disturbance data sources that were attributed to each of the river reaches and included in our database. NCD = National Land Cover Database; NHEPles = National Hydrography Dataset Plus; STATSO = State Soil Geographic Database; TRSER = Topologically Integrated Geographic Encoding and Referencing System; SPMR0W = Spatially Referenced Regression on Watershed Attributes; HUC = Hydrologic Unit Code; Date = year or range of years for each data source.

Description	Source	Resolution	Date
Natural variables			
Land cover—Deciduous forest	NLCD ( <a href="http://www.arlc.gov/">http://www.arlc.gov/</a> )	30 m	2001
Land cover—Evergreen forest	NLCD ( <a href="http://www.arlc.gov/">http://www.arlc.gov/</a> )	30 m	2001
Land cover—Mixed forest	NLCD ( <a href="http://www.arlc.gov/">http://www.arlc.gov/</a> )	30 m	2001
Land cover—Open water	NLCD ( <a href="http://www.arlc.gov/">http://www.arlc.gov/</a> )	30 m	2001
Land cover—Shrub/scrub	NLCD ( <a href="http://www.arlc.gov/">http://www.arlc.gov/</a> )	30 m	2001
Land cover—Grassland/forbaceous	NLCD ( <a href="http://www.arlc.gov/">http://www.arlc.gov/</a> )	30 m	2001
Land cover—Woody wetlands	NLCD ( <a href="http://www.arlc.gov/">http://www.arlc.gov/</a> )	30 m	2001
Land cover—Open wetlands	NLCD ( <a href="http://www.arlc.gov/">http://www.arlc.gov/</a> )	30 m	2001
Local catchment area	NHDPlus ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	1:100,000	2007
Network catchment area	NHDPlus ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	1:100,000	2007
Mean annual air temperature	NHDPlus ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	4 km	1961–2000
Mean annual precipitation	NHDPlus ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	4 km	1961–2000
Human variables			
Imperviousness	NLCD ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	30 m	2001
Land use—Pasture/hay	NLCD ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	30 m	2001
Land use—Cultivated crops	NLCD ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	30 m	2001
Land use—Open space urban	NLCD ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	30 m	2001
Land use—Low intensity urban	NLCD ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	30 m	2001
Land use—Medium intensity urban	NLCD ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	30 m	2001
Land use—High intensity urban	NLCD ( <a href="http://www.bartov-systems.com/nhdplus/">http://www.bartov-systems.com/nhdplus/</a> )	30 m	2001
Mining density	USGS Active Mines ( <a href="http://nats.usgs.gov/natprod/">http://nats.usgs.gov/natprod/</a> )	Point data	2003
National Pollutant Discharge Classification System Density	EPA Geospatial Data ( <a href="http://www.epa.gov/nwtr/gis_data.html">http://www.epa.gov/nwtr/gis_data.html</a> )	Point data	2007
Road crossing density	Census 2000 TIGER Roads ( <a href="http://www.census.gov/data/download/census2000-tiger-line.html">http://www.census.gov/data/download/census2000-tiger-line.html</a> )	1:100,000	2000
Road length density	Census 2000 TIGER Roads ( <a href="http://www.census.gov/data/download/census2000-tiger-line.html">http://www.census.gov/data/download/census2000-tiger-line.html</a> )	1:100,000	2000
Superfund National Priority List Density	EPA Geospatial Data ( <a href="http://www.epa.gov/nwtr/gis_data.html">http://www.epa.gov/nwtr/gis_data.html</a> )	Point data	2007
Toxic Release Inventory density	EPA Geospatial Data ( <a href="http://www.epa.gov/nwtr/gis_data.html">http://www.epa.gov/nwtr/gis_data.html</a> )	Point data	2007
Total phosphorus yield	USGS SPMR0W ( <a href="http://nats.usgs.gov/nwtr/spmr0w/rr92/outputs.html">http://nats.usgs.gov/nwtr/spmr0w/rr92/outputs.html</a> )	8-Digit HUC	1974–1989
Total nitrogen yield	USGS SPMR0W ( <a href="http://nats.usgs.gov/nwtr/spmr0w/rr92/outputs.html">http://nats.usgs.gov/nwtr/spmr0w/rr92/outputs.html</a> )	8-Digit HUC	1974–1989

Many databases in one



Dana Infante, Gary Whelan

# StreamCat: Watershed Variables for Predicting Aquatic Condition and Watershed Integrity

## Phase 1 Indicators



## Phase 2 Indicators

Examples of possible Phase 2 indicators:

- Topographic Wetness Index
- Mean summer temperature/precip
- Functional Process Zone (e.g., valley floor width, channel belt width and sinuosity)
- Pesticide applications rates
- Agricultural fertilizer application rates
- Recent forest loss

Tony Olsen (EPA)

# StreamCat: Watershed Variables for Predicting Aquatic Condition and Watershed Integrity

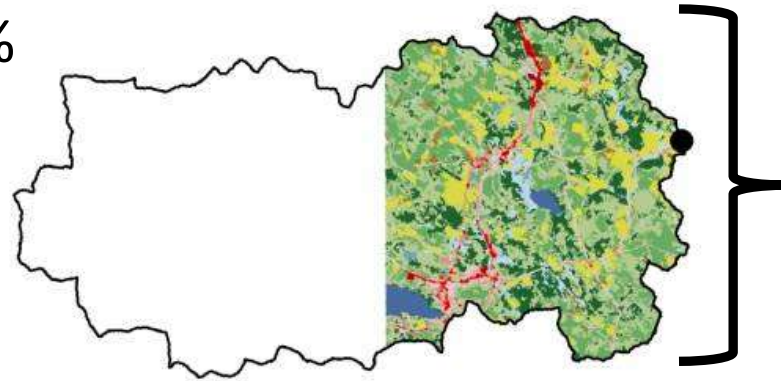
## Quality Assurance

Landscape layers (LL)  
*(CONUS-wide layers of climate, geology, soils, land cover/land use, and others)*



*Problem* – Most Landscape Layers do not cross international boundaries or have missing values in some locations

*Solution* – Calculate the % completeness of each Landscape Layer within each watershed



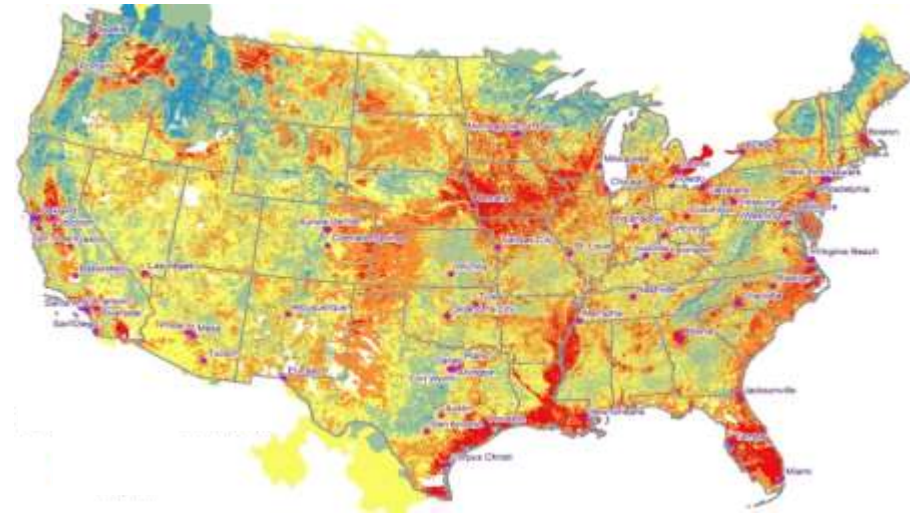
48% complete



# StreamCat: Watershed Variables for Predicting Aquatic Condition and Watershed Integrity

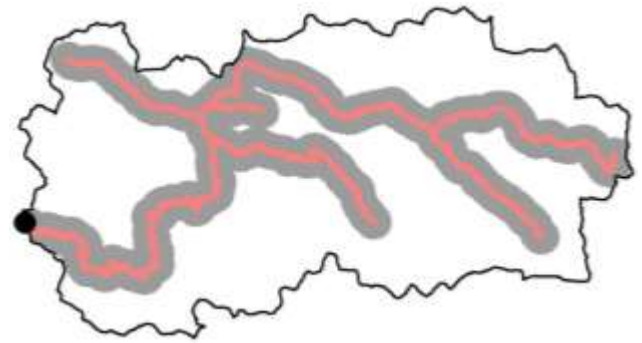
## Data Assembly Products

- 30 landscape layers with QA documentation
- 84 initial watershed metrics for:
  - 2.7 million NHD watersheds
  - 1,883 NRSA watersheds
- Data dictionary for 84 watershed metrics
- Python and R code for each step in metric calculation process
- 7 page description of methods



### Multiple Versions of Some Variables

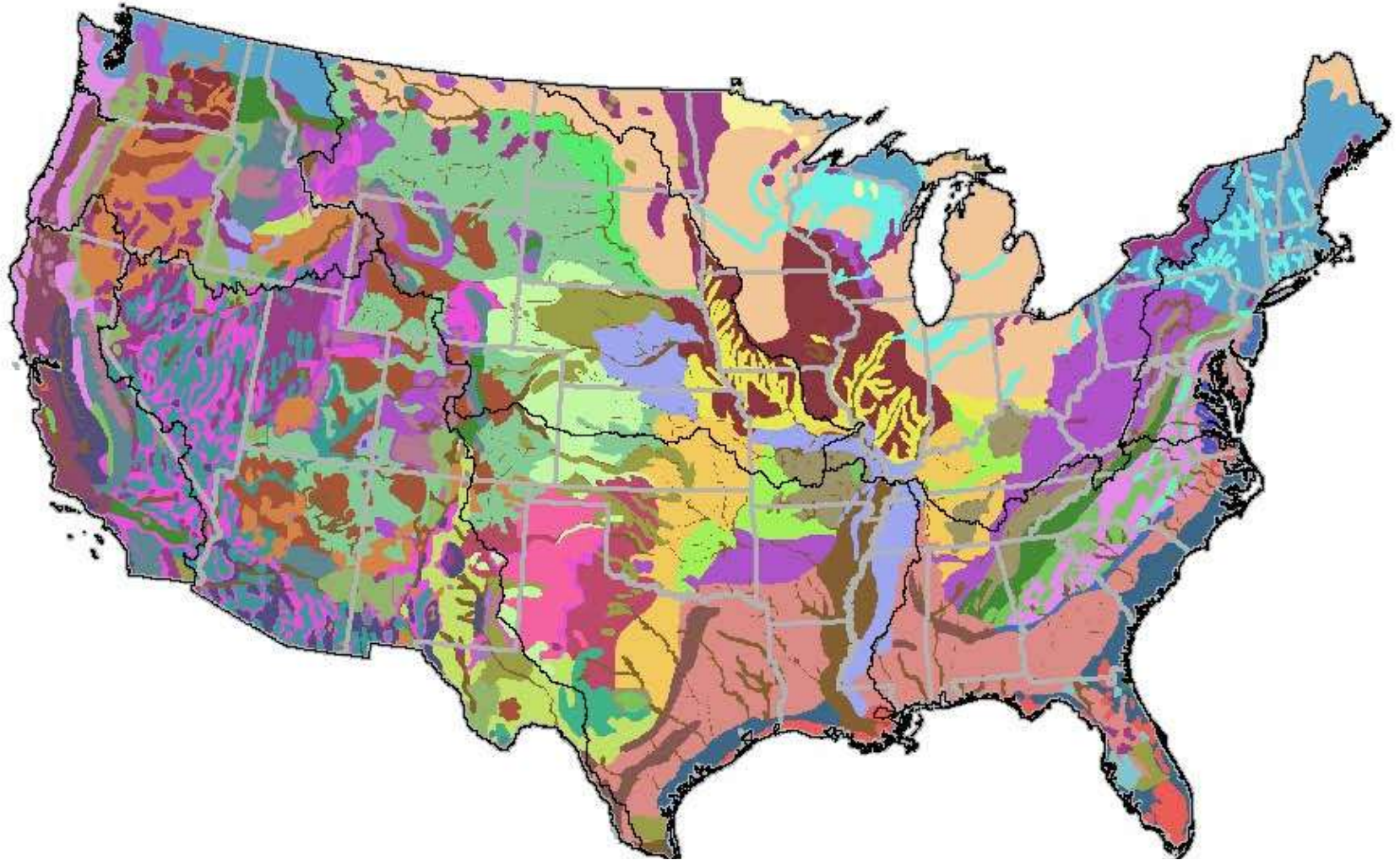
- Watershed-level
- Within 100-m buffer
- Within 600-m buffer



**Tony Olsen (EPA)**

# Not very useful for ecologists

## Surficial Geology maps

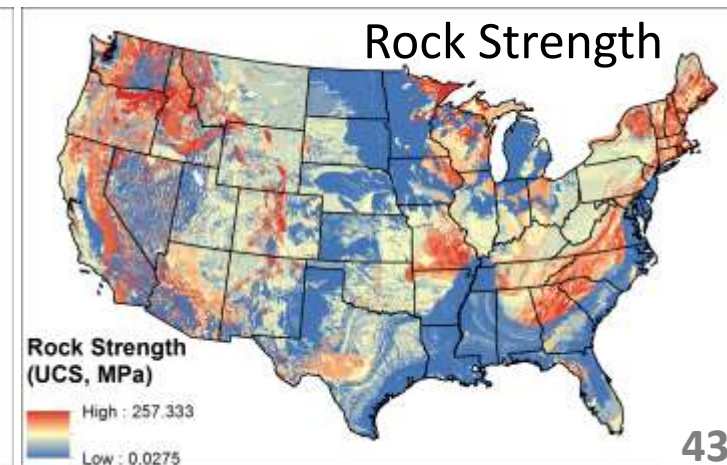
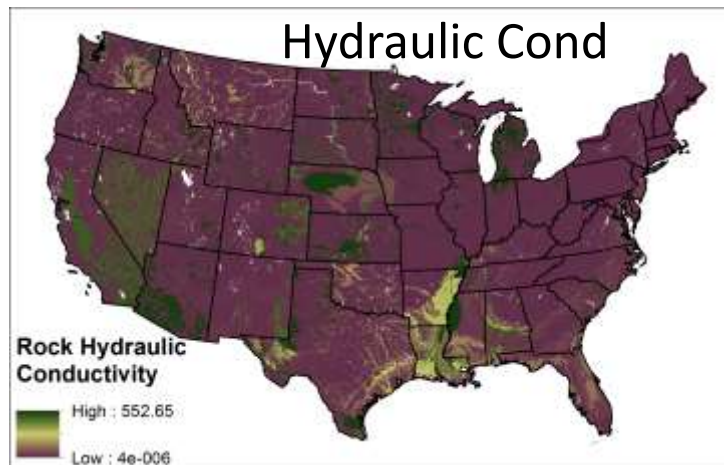
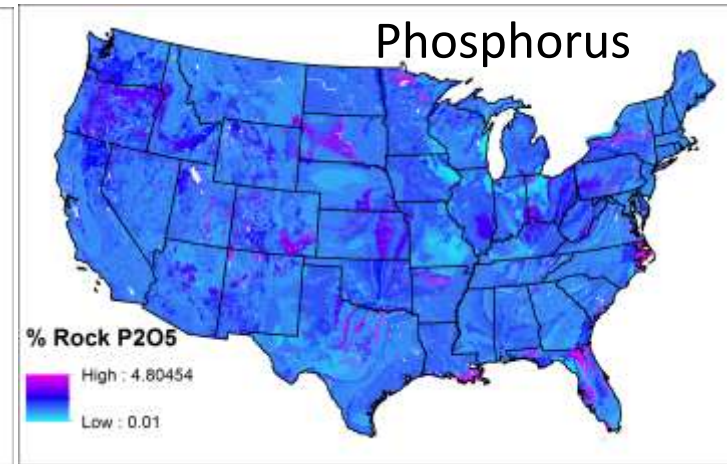
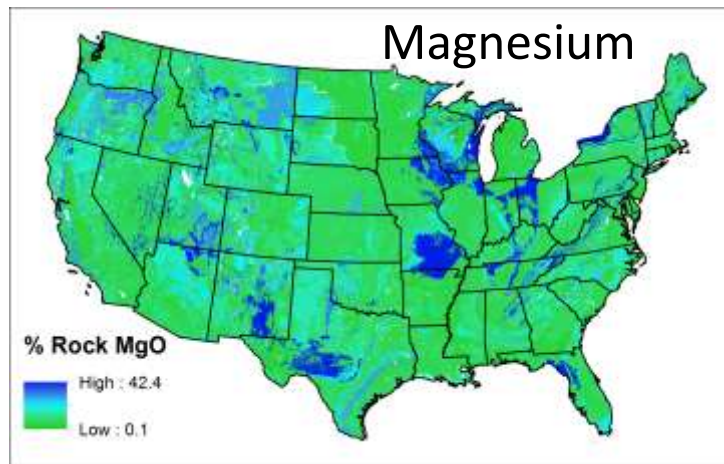
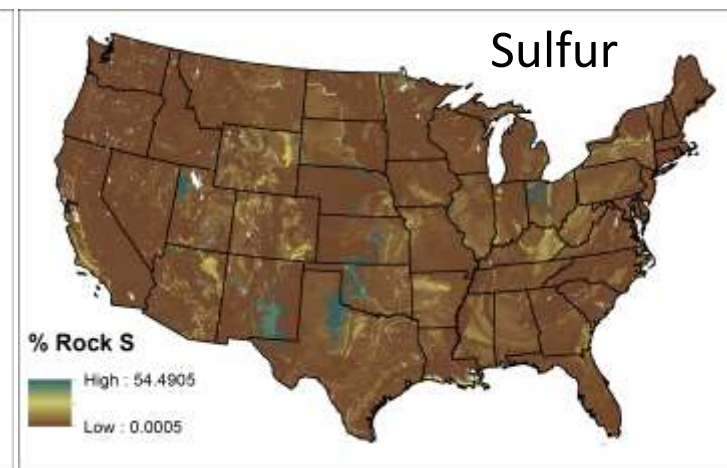
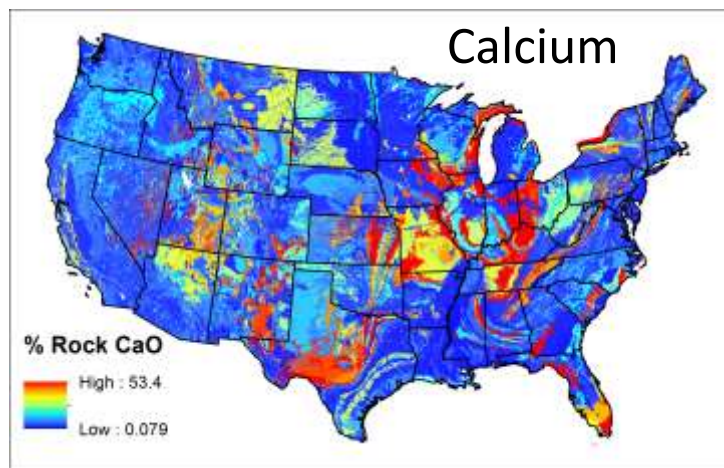


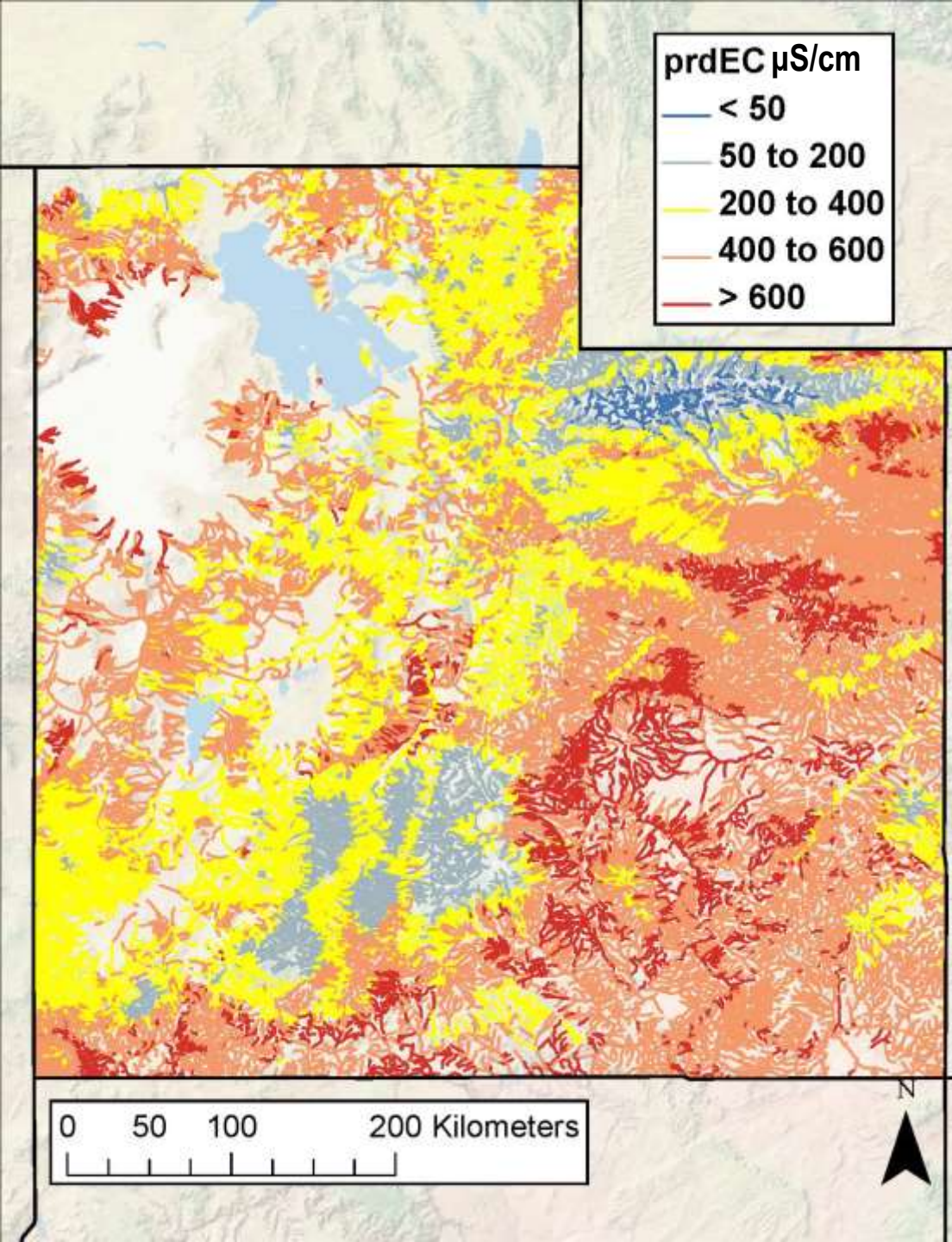
**John Olson/Chuck Hawkins (USU)**

# John Olson translations

Derived from Integrated Geologic Maps Databases and other sources

[www.sciencebase.gov](http://www.sciencebase.gov)





Use derived geology  
to model water  
chemistry

Predicted spatial  
variation in natural  
base flow salinity  
( $\mu\text{S}/\text{cm}$ ) in Utah  
streams

John Olson's  
nationwide model

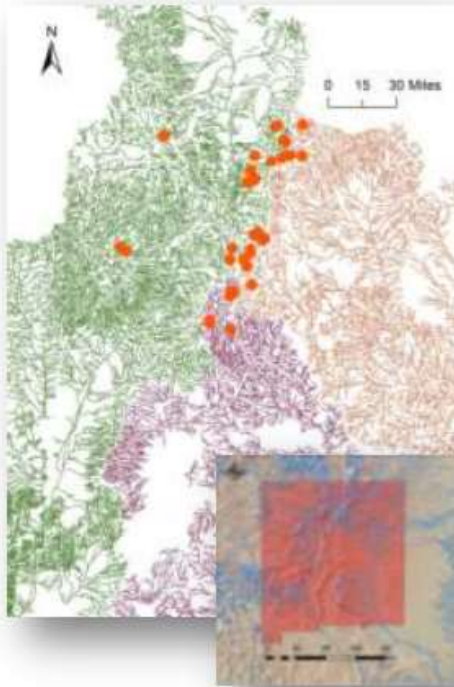
# Some new spatial analyses



# POPULATION GENETICS & SSN

David Cowley, Sabela Lois, Erin Peterson, Dan Isaak et al.





Canadian, 7 streams, N=160  
 Rio Grande, 26 streams N=757  
 Pecos, 9 streams, N=254

Georeference Genetic Samples

GIS: spread individual genotypes @ 50 m

Create network outlets @

Format Genetic Data (0, 1, 2)

Sample	allele 1	allele 2
1 (heterozy.)	1	1
2 (homozy.)	0	2

PCA on allele-count data

PC scores for each individual

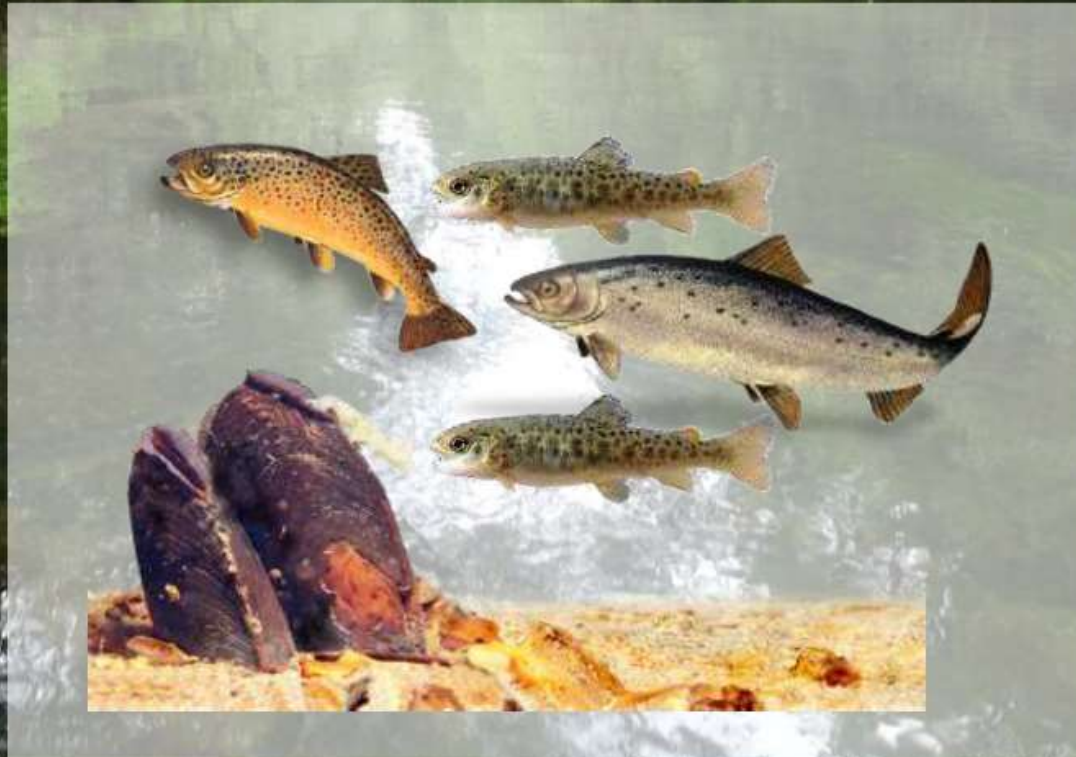
Separate SSN Analyses on "Important" PCs

Select Weighting & Fixed Effects

Simulation of Genetics on SSNs

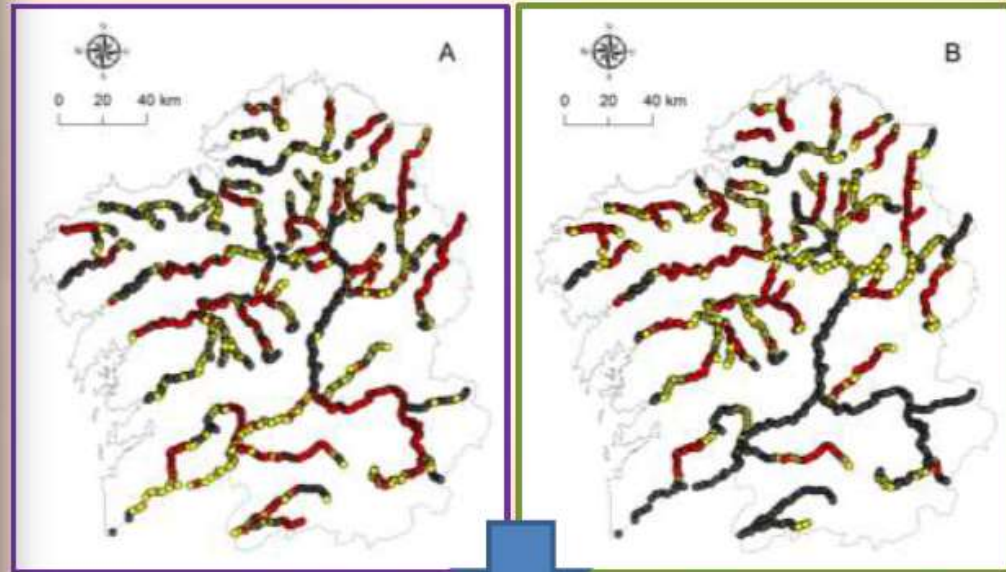
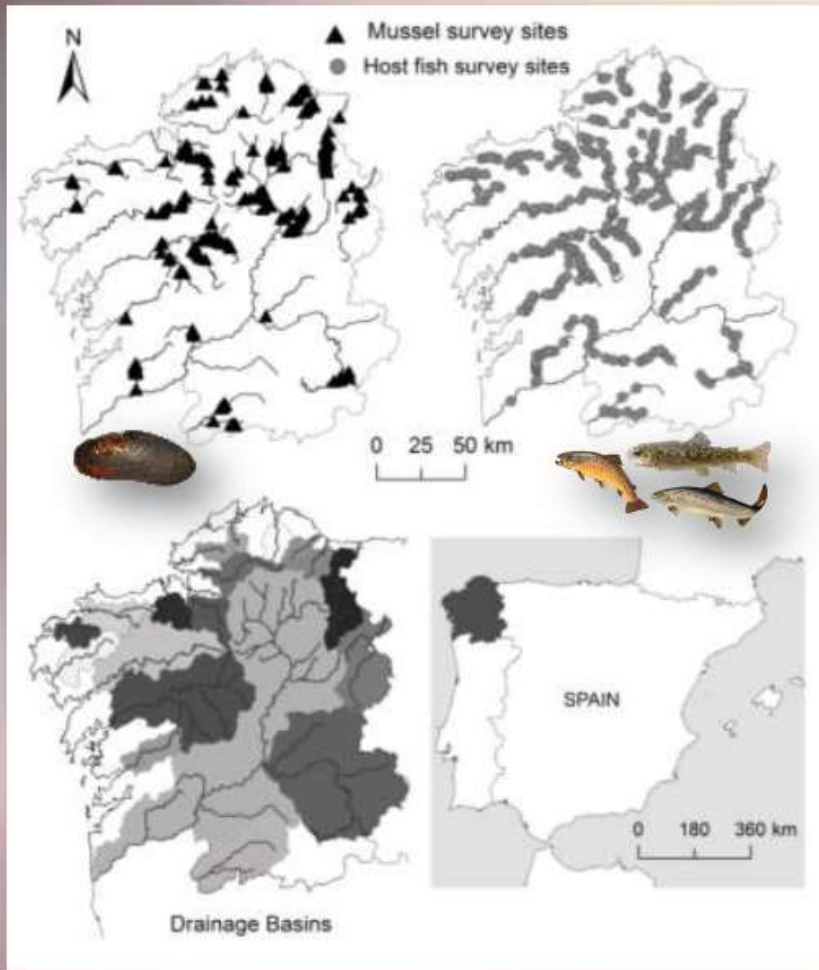
# Multi-Species Conservation Using SSNs

Sabela Lois & David Cowley

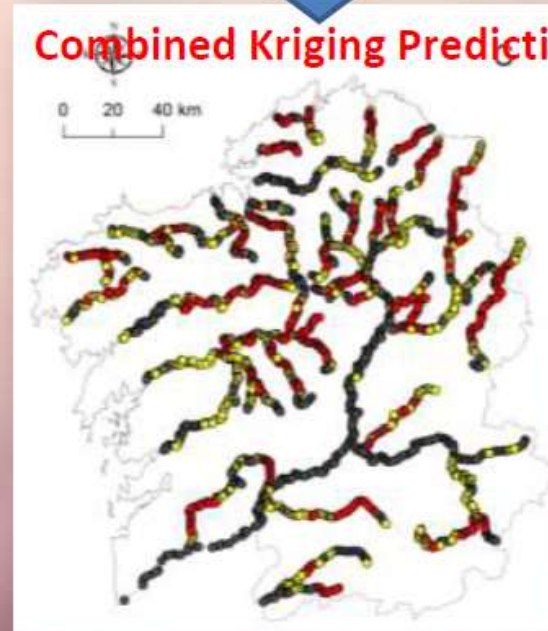




# Host-Parasite Conservation



## Combined Kriging Predictions



Conservation

Restoration

Recovery

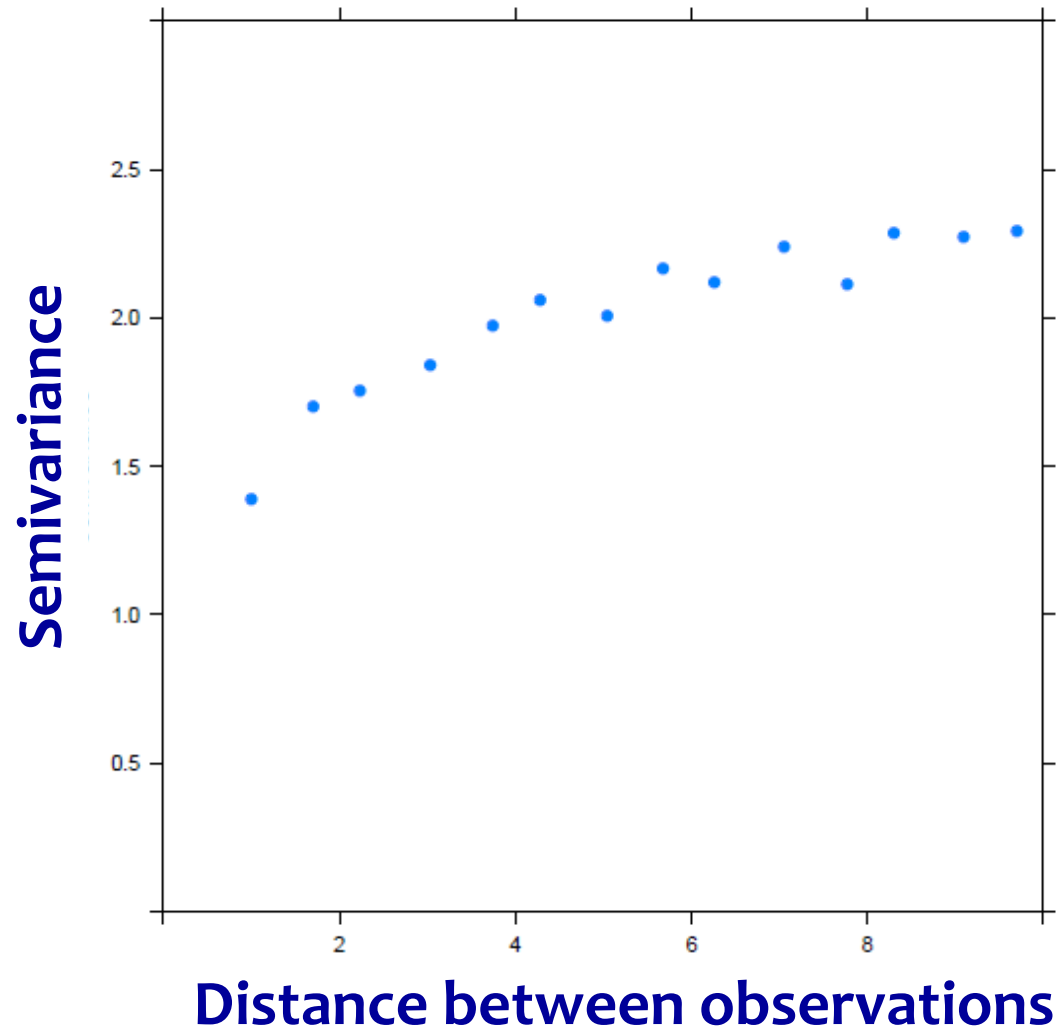
CLIMATE  
HOST FISH  
GEOLOGY  
LANDFORM  
LAND USE

+

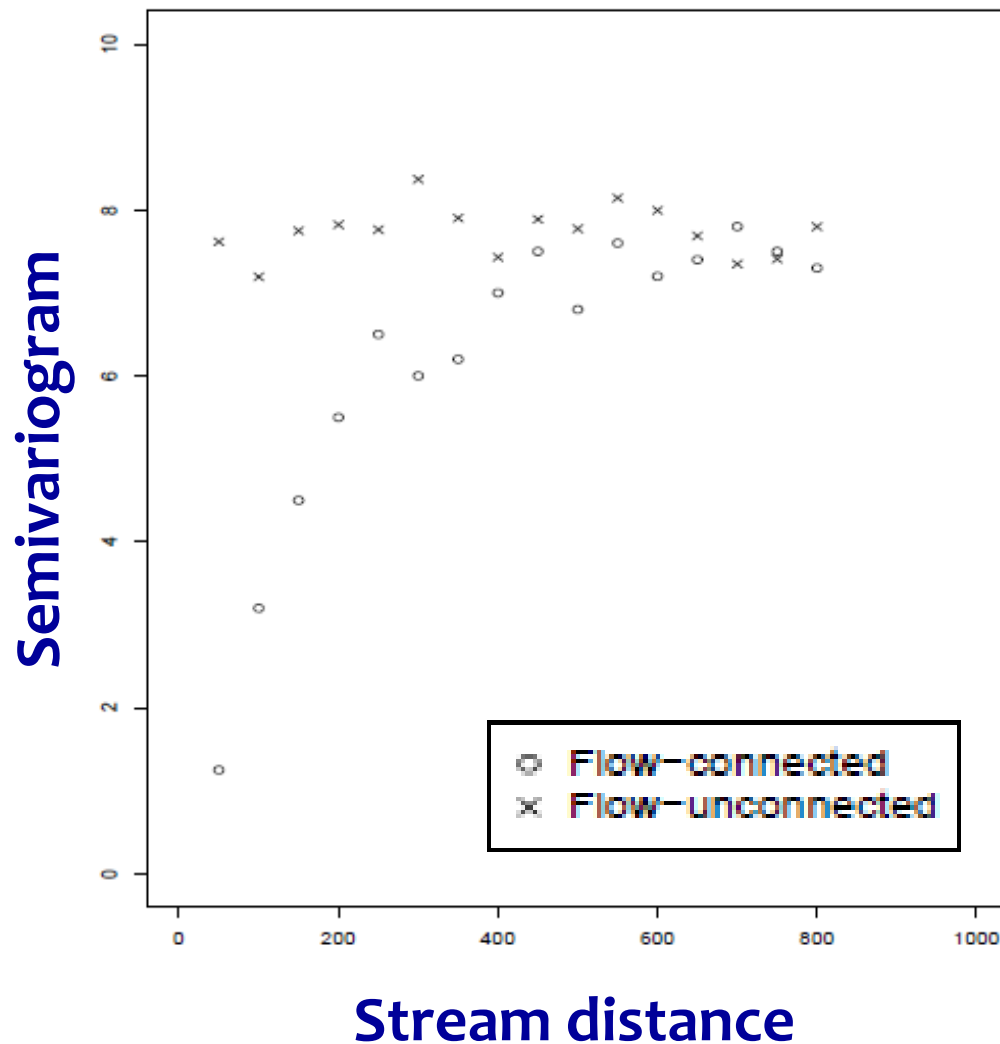
SSN with  
Tail-up &  
Tail-down

# Characterizing spatial dependence on streams

A “typical” empirical semivariogram (in Euclidean space):



A Torgegram (empirical stream-network semivariogram) consists of separate semivariograms for flow-connected and flow-unconnected sites, and may also account for flow volume:



- The Torgegram has tremendous diagnostic value, and for very large stream datasets may be the only estimate of spatial dependence that is feasible to compute
- I'm developing formal hypothesis testing methods which use the Torgegram to determine whether the best model for a given stream dataset is tail-up, tail-down, or a hybrid
- I'm also developing methods for testing for (spatial) stationarity based on comparisons of watershed-specific Torgegrams
- Jay and I plan to extend the Torgegram to space-time and multivariate data
- Reference for our working paper:  
Zimmerman, D.L. and Ver Hoef, J. (2015). The Torgegram for fluvial variography: Characterizing spatial dependence on stream networks. In progress.

# Introduction

- **Associate Professor (CSU)**

Fish, Wildlife, and Conservation  
Biology

Statistics

- **Assistant Unit Leader (USGS)**

Colorado Cooperative Fish and  
Wildlife Research Unit



**Mevin Hooten**

Fish, Wildlife, & Conservation Biology and Statistics  
Colorado State University  
Colorado Cooperative Fish and Wildlife Research Unit  
U.S. Geological Survey

# Spatial Occupancy Models

Ecology 2014, 95, 181-192  
© 2013 by the Biological Society of America

## Spatial occupancy models for large data sets

David E. Hooten,<sup>1,2</sup> Paul H. Cox,<sup>3</sup> Mevin Hooten,<sup>1,2,4</sup> James C. Ray,<sup>5</sup> and Bruce A. Pusek<sup>6</sup>

<sup>1</sup>National Wildlife Research Laboratory, USDA, 3601 Sand Point Way SE, Seattle, Washington 98112 USA  
<sup>2</sup>U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University,  
Fort Collins, Colorado 80523 USA

<sup>3</sup>Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado 80523 USA  
<sup>4</sup>Department of Statistics, Colorado State University, Fort Collins, Colorado 80523 USA

<sup>5</sup>Wildlife Conservation Society Canada, Toronto, Ontario M5S 2T7 Canada

<sup>6</sup>Wildlife Research and Development Division (Ottawa Ministry of Natural Resources, Peterborough, Ontario E3V 5M7 Canada)

**Abstract.** Since its development, occupancy modeling has become a popular and useful tool for ecologists wanting to learn about the dynamics of species occurrence over time and space. Such models require presence-absence data to be collected at spatially indexed survey units. However, only recently have researchers recognized the need to account for spatially induced overdispersion by explicitly accounting for spatial autocorrelation in occupancy probability. Previous efforts to incorporate such autocorrelations have largely focused on logit-normal formulations for occupancy, with spatial autocorrelations induced by a random effect within a hierarchical modeling framework. Although useful, computational time generally limits such an approach to relatively small data sets, and there are often problems with algorithm instability, yielding uninterpretable results. Further, recent research has revealed a hidden form of multicollinearity in such applications, which may lead to parameter bias if not explicitly addressed. Combining several techniques, we present a unified hierarchical spatial occupancy model specification that is particularly effective over large spatial extents. This approach employs a product mixture framework for occupancy and can easily accommodate a reduced-dimensional spatial process to resolve issues with multicollinearity and spatial overfitting while improving algorithm convergence. Using open-source software, we demonstrate this new model specification using a case study involving occupancy of cutthroat trout (*Oncorhynchus clauseni*) over a set of 1000 survey units spanning a large contiguous region (100 000 km<sup>2</sup>) in northern Ontario, Canada. Overall, the combination of a more efficient specification and open-source software allows for a fast and stable implementation of spatial occupancy models for large data sets.

**Key words:** fish, cutthroat trout, conditionally autoregressive, occupancy model, profit regression, weighted likelihood, reduced-rank, spatial regression.

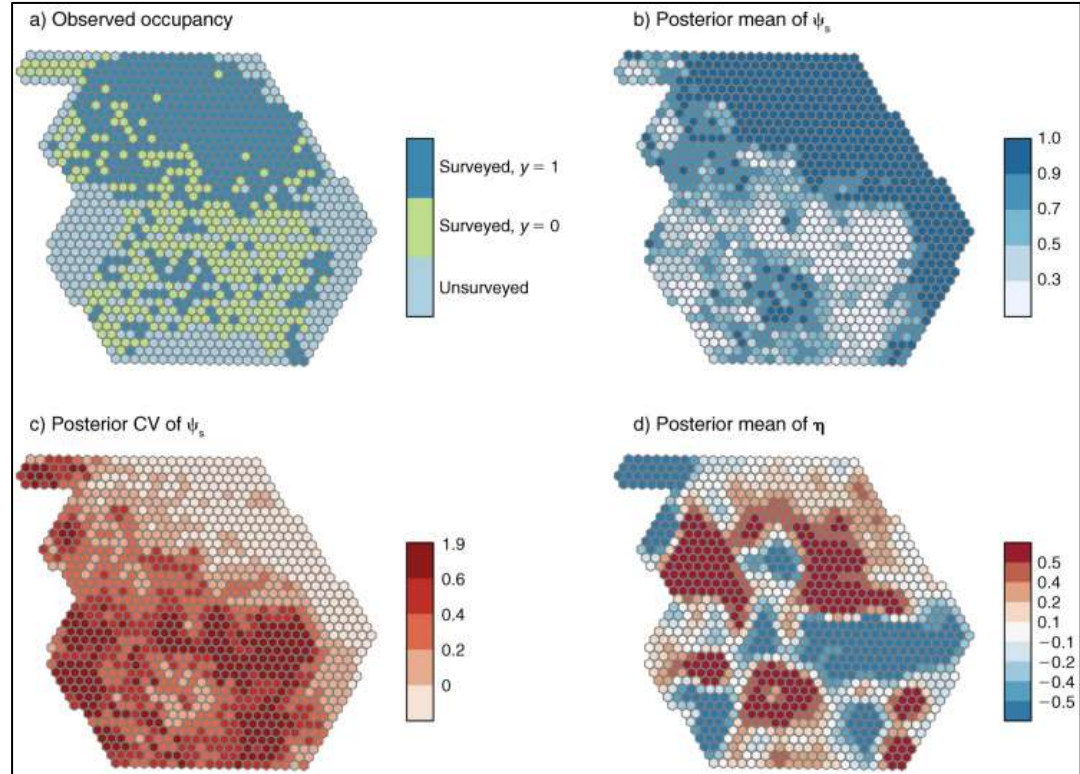
### INTRODUCTION

Since the seminal work by MacKenzie et al. (2002), there has been an explosion in ecological studies designed to estimate occupancy and related parameters (see MacKenzie et al. (2005) and Long et al. (2008) for reviews). Occupancy, defined here as the probability that a local taxa occurs (in some meaningful sense) in a survey unit, is commonly used for population monitoring and for assessing whether hypothesized covariates influence species presence or absence. The appeal of occupancy studies is undeniable, as inferential about population-level processes can be made without physically capturing and marking animals or otherwise assessing the population. Well-designed occupancy studies are particularly advantageous for assessing spatial distributions of wide-ranging, elusive species or

the landscape scale, where more intensive sampling designs are costly and inefficient (e.g., Magnus et al. 2007; Kersch et al. 2011).

Typical occupancy study designs involve identification of a number of survey units or habitat patches within a larger study area. Each unit is visited by one or more trained observers, with a subset of units being visited more than once. The rationale for visiting a unit more than once is to help account for false negatives. If the species is truly absent, the probability of detecting the species is usually assumed to be zero (with the exception of a few studies, e.g., Royle and Link (2005), Hartz et al. (2011)). Observed absences are thus a mixture of true absences and nondetections, sampling a subset of units multiple times provides the information necessary to discriminate between the two. Although MacKenzie et al. (2002) is often cited as the initial paper on occupancy modeling, Hooten et al. (2008) provides an earlier description of presence in the face of uncertain detection. Moreover, the model Hooten et al. (2008) proposed was also spatially explicit,

Reports



Mevin Hooten

Fish, Wildlife, & Conservation Biology and Statistics  
Colorado State University  
Colorado Cooperative Fish and Wildlife Research Unit  
U.S. Geological Survey

# Colorado Plains Fish

## Methods in Ecology and Evolution

Methods in Ecology and Evolution 2013, 4, 99–108

doi: 10.1111/1365-2113.12206

### Accounting for imperfect detection in Hill numbers for biodiversity studies

Kristin M. Broms<sup>1\*</sup>, Mevin B. Hooten<sup>1,2,3</sup> and Ryan M. Fitzpatrick<sup>4</sup>

<sup>1</sup>Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA; <sup>2</sup>U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Unit, Fort Collins, CO 80523, USA; <sup>3</sup>Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA; and <sup>4</sup>Aquatic Wildlife Research Group, Colorado Parks and Wildlife, Fort Collins, CO 80526, USA

#### Summary

- Hill numbers rely on biodiversity metrics by averaging several axes in an ordination. For example, species richness, Shannon's diversity index and the Gini-Simpson index are a few of the more used diversity measures, and they can be expressed as Hill numbers. Traditionally, Hill numbers have been calculated from relative abundance data, but the expression has been modified to use incidence data as well. We demonstrate an approach for estimating Hill numbers using an occupancy modelling framework that accounts for imperfect detection.
- We apply the Hill numbers formula to our occupancy probabilities as opposed to the incidence probabilities that have been used previously and re-calculate its counterparts from the modified species richness. After incorporating the occupancy-based Hill numbers, we demonstrate the difference between them and the incidence-based Hill numbers previously used through a simulation study and two applications.
- In the simulation study and the two examples using real data, the occupancy-based Hill numbers were larger than the incidence-based Hill numbers, although species richness was estimated similarly using both methods.
- The occupancy-based Hill number estimates are always at their asymptotic values (i.e. as if an infinite number of samples have been taken for the study region), therefore making it easy to compare biodiversity between different assemblages. In addition, the Hill numbers are compared as default quantities within a Bayesian hierarchical model, allowing for straightforward inference.

**Key-words:** Bayesian methods, Gini-Simpson index, incidence matrix, multi-species occupancy model, Shannon's entropy, species richness

#### Introduction

Biodiversity is one of the most important concepts in the study of ecology and is commonly measured by species richness, the Gini-Simpson index and Shannon's entropy (Lande 1996; Jost 2006; Mac 2007; Gotelli & Chao 2014; Chao et al. 2014). Multiple measurements of biodiversity are valuable because no technique does not measure the presence of all species. To adapt the example from Gotelli & Chao (2013), suppose two communities both contain exactly five species. The first community has one species comprising 0.90 of the total number of individuals, with the other species each comprising 0.05 of the assemblage. In the second community, each species comprises 0.20 of the population. Arguably, the second community should be seen as more diverse, but the species richness estimator is not robust enough to distinguish the two communities.

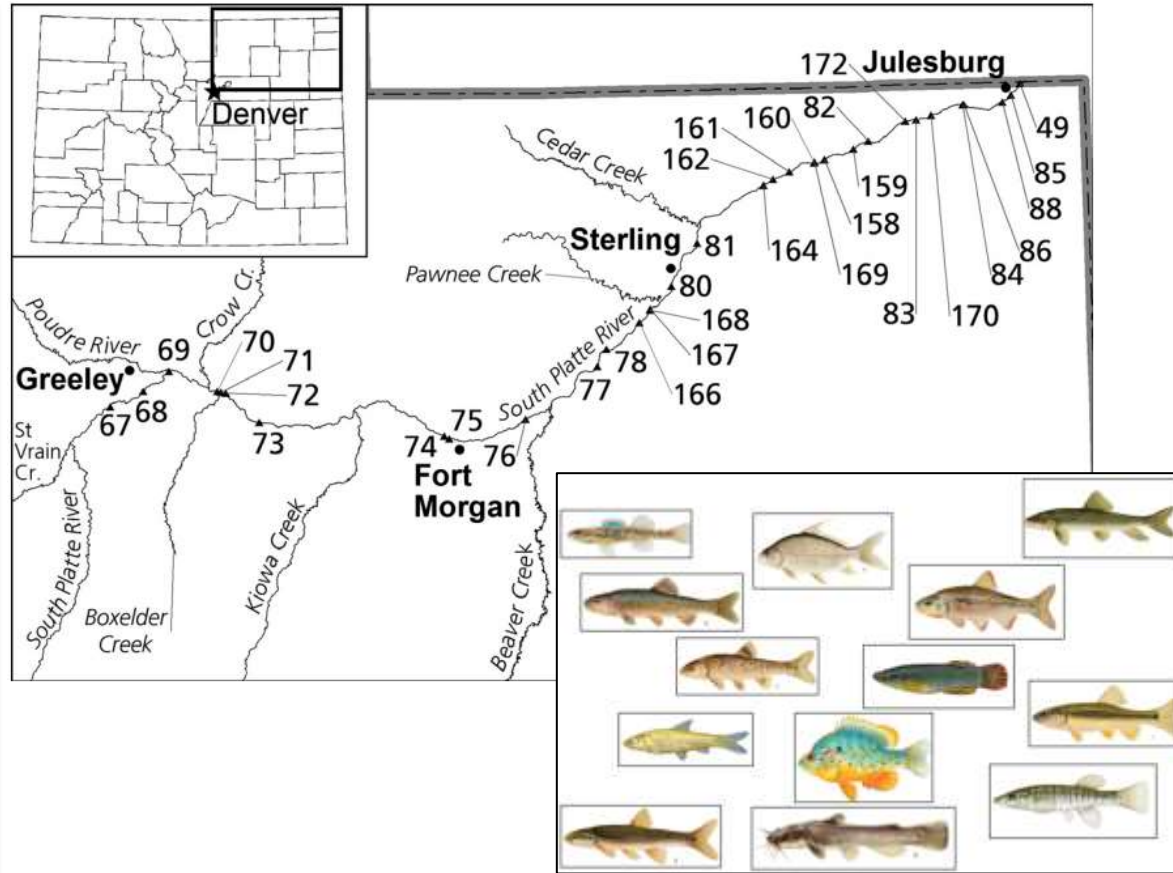
Therefore, other measurements of biodiversity are also used; both the Shannon entropy and the Gini-Simpson index take relative abundances of each species into account. The Shannon entropy 'quantifies the uncertainty in the species identity of a randomly chosen individual in the assemblage' (Gotelli &

Chao 2013). It is also called the Shannon's diversity index or the Shannon-Wiener index (Jost 2006). The Gini-Simpson index 'measures the probability that two randomly chosen individuals (selected with replacement) belong to two different species' (Gotelli & Chao 2013). Variations of the Gini-Simpson index include the Simpson concentration, the inverse Simpson concentration, the second-order Rényi entropy or the Hill number (Gotelli & Chao 2013).

Hill numbers conveniently summarize all three types of biodiversity using a single expression, providing a unification (Hill 1975; Chao et al. 2014; Chao, Jost & Chao 2016) and a framework from which to derive alpha and beta diversities (Jost 2007). We illustrate the Hill number formula and its exact relationship to Shannon entropy and the Gini-Simpson index in the 'Implementation' section.

Traditionally, biodiversity measurements are functions of the relative abundances of each species in an assemblage, as obtained from the sampling design. However, Hill numbers have also been calculated using presence-absence data (Caldwell & Colwell 1989; Colwell, Mao & Chang 2004; Colwell et al. 2012; Gotelli & Chao 2013; Chao et al. 2014). The presence-absence data are less informative than species counts, but they can be easier to collect, they may allow for comparisons

\*Correspondence author. Email: kristin.broms@colostate.edu



Mevin Hooten

Fish, Wildlife, & Conservation Biology and Statistics  
Colorado State University  
Colorado Cooperative Fish and Wildlife Research Unit  
U.S. Geological Survey

# Landscape Genetics

## Circuit Theory and Model-Based Inference for Landscape Connectivity

Ephraim M. Hooten and Mewin B. Hooten

Circuit theory has been extensively used in the field of ecology, where it is often applied to study functional connectivity. The landscape is typically represented by a network of nodes and resistors, with the resistance between nodes a function of landscape characteristics. The effective distance between two locations on a landscape is represented by the resistance distance between the nodes in the network. Circuit theory has been applied to many other scientific fields for exploratory analyses, but parameter estimates for circuits are not common in the scientific literature. To model circuits explicitly, we demonstrate a link between Gaussian Markov random fields and contemporary circuit theory using a maximum likelihood framework to estimate resistance distance. This provides a parsimonious model for second order observations from such a network. In the landscape ecology setting, the proposed model provides a simple framework whose estimates can be obtained for efforts that landscape features have on functional connectivity. We illustrate the approach through a landscape genetics study testing gene flow in alpine thistles (*Leucophaea nuttallii*) in the underlying landscape.

**KEY WORDS:** Conditional autoregressive models, Gene flow, Landscape genetics, Nonsatiable spatial correlation

### 1. INTRODUCTION

Circuit theory has been successfully used to study connectivity in a wide range of fields, including molecular chemistry (Zhu and Klein 1996; Klein et al. 2004), collaborative recommendation (Dumas et al. 2007), communications network analysis (Tijssen and Leon-Garcia 2010, 2011), social network analysis (Katz and Leskovec 2009; and Bhaskar 2009), and random walks on a graph (Chandra et al. 1996; Mikami 2011). Circuit theory has also seen extensive recent use in landscape ecology, where it has been theoretically linked to animal movements and gene flow in heterogeneous landscapes (McRae 2004; Cushman et al. 2009; McRae and Beier 2007; McRae et al. 2008; Urban et al. 2009; Cushman and Landguth 2010; Dyer, Nason, and Gurrick 2010; Lookingbill et al. 2010; Owen-Smith, Fryxell, and Merrill 2010; Rayfield, Fortin, and Fall 2010; Saura and Rubio 2010). In these latter cases, the landscape is specified as a raster grid with connectivity between grid cells determined by landscape characteristics and modeled based on circuit theory (Figure 1). Circuit theory provides a flexible framework for modeling nonstationary connectivity, and shows promise for predicting effects of landscape and environmental change on connectivity (e.g., Suter et al. 2007; Spear et al. 2010).

A key challenge in modeling landscape connectivity using circuit theory is to estimate the relative resistance values of various landscape characteristics (e.g., Spear et al. 2010). In applications of circuit theory other than landscape ecology, resistance values are typically known, or all resistors in the circuit are assumed to have equal resistance. In those cases, the focus is typically on exploratory analysis of the connectivity implied by viewing the system as a circuit, rather than on estimating resistance values based on observations. In contrast, the goal in landscape ecological applications of circuit theory is to understand the impact that

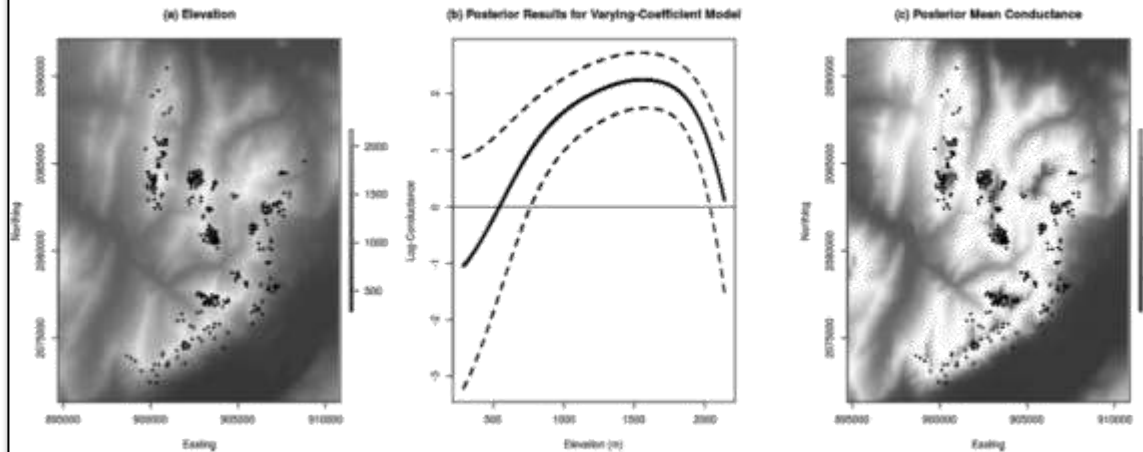
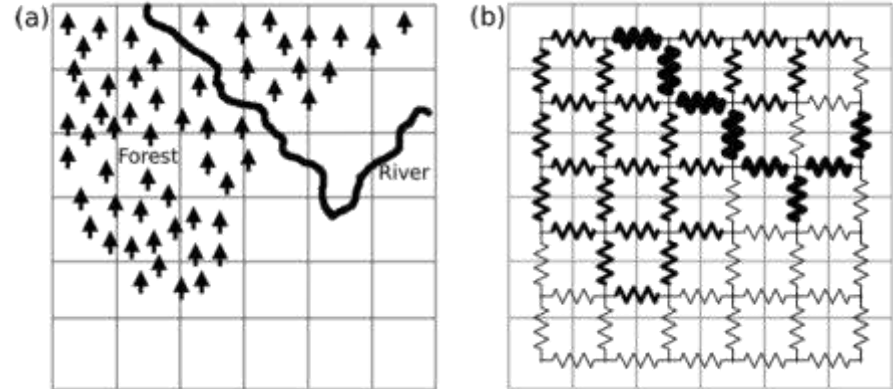
different landscape characteristics have on connectivity. Observations are typically second order, and come in the form of an observed pairwise distance matrix representing the process under study (e.g., spatial gene flow in landscape genetics). The most common approach used to estimate resistance values for different landscape characteristics is to choose between a set of prospectively candidate resistance values for each landscape covariate hypothesized to have an effect on connectivity (e.g., Cushman et al. 2009; Cushman, McKelvey, and Schwartz 2009). Each set of candidate resistance values is used to create a hypothesized resistance distance matrix between the observed spatial locations in the study, where the resistance distance is computed based on circuit theory. The correlations between each of the hypothesized distance matrices and the observed distance matrix are computed, and the set of candidate resistance values that results in the highest correlation to the observed distance matrix is chosen (e.g., Cushman et al. 2009; Cushman, McKelvey, and Schwartz 2009; Wang, Saura, and Bradley Shaffer 2009; Sisk et al. 2010), with significance assessed through Mantel permutation tests (e.g., Legendre and Fortin 2010).

One major drawback of this approach is that there is no obvious way to assess the uncertainty in the parameter estimates for the resistance values of the landscape covariates. This is a critical point, as the results of spatial connectivity studies are being used to influence policy decisions (e.g., Theobald, Cooks, and Norman 2011) and predict the results of landscape change over time (e.g., Spear et al. 2010).

Our goal is to put the estimation of resistance values from observed genetic distance matrices within a model-based framework. Recent work by McLaughlin (2009) shows that observed squared-Euclidean distance matrices can be modeled using the generalized (or isotropic) Wishart distribution with a spatial covariance matrix as a parameter. However, it is not immediately obvious how to parameterize a covariance matrix in a way that models connectivity based on circuit theory. As circuits are based on a graph or network of nodes, it seems natural to

Ephraim M. Hooten is Ph.D. candidate in the Department of Statistics, Colorado State University, Fort Collins, CO 80523. E-mail: hooten@stat.colostate.edu. Mewin B. Hooten is Assistant Unit Leader at the U.S. Geological Survey Colorado Cooperative Fish and Wildlife Research Unit and Assistant Professor in Fish, Wildlife, and Conservation Biology and Statistics at Colorado State University, Fort Collins, CO 80523. E-mail: mhooten@colorado.gov. The use of trade names or products does not constitute endorsement by the U.S. government.

In the Public Domain  
Journal of the American Statistical Association  
March 2013, Vol. 108, No. 501, Applications & Case Studies  
DOI: 10.1080/16214997.2012.728487



# Mewin Hooten

Fish, Wildlife, & Conservation Biology and Statistics  
Colorado State University  
Colorado Cooperative Fish and Wildlife Research Unit  
U.S. Geological Survey



# NFWF (National Fish & Wildlife Foundation)



# NFWF

Who We Are    What We Do    Partnerships    Results

Home » Themes » Freshwater



search



FRESHWATER



FORESTS AND GRASSLANDS



Eastern Brooke Trout | Credit: USFWS

## Freshwater

Freshwater ecosystems are among the most threatened

### CONSERVATION PROGRAMS

- » Acres for America
- » Alaska Fish and Wildlife Fund

### FRESHWATER FEATURE STORY



Penobscot Indian elder Butch Phillipps describes how this project will restore 1000 miles of habitat for fish and wildlife.

Watch the video ▶

### FRESHWATER ANNOUNCEMENTS

3/20/2015  
Unassessed Waters Initiative 2015 RFP Has Been Released!

# Dave Lawrence

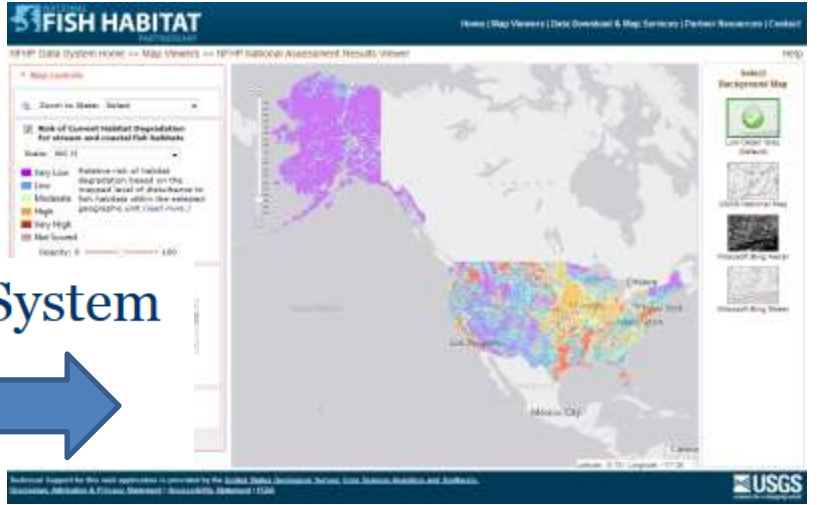
Website: <http://www.nfwf.org/>

# NFHP (National Fish Habitat Partnership)



Reservoir FHP featured on Bass Pro Shops Outdoor World Radio  
 Jeff Boxrucker, coordinator for the Reservoir Fisheries Habitat Partnership recently did an interview with Rural Radio of Sirius XM's Rural Radio (Channel 80) during Bass Pro

Join the effort!  
 - become a partner coalition member today  
 Create an account to receive Fish Habitat News and join the growing Fish Habitat Partner Coalition!



## National Fish Habitat Partnership Data System

The National Fish Habitat Partnership (NFHP) Data System supports coordinated efforts of scientific assessment and data exchange among the partners and



Gary Whelan Website: <http://www.fishhabitat.org/>

**Anything else?**