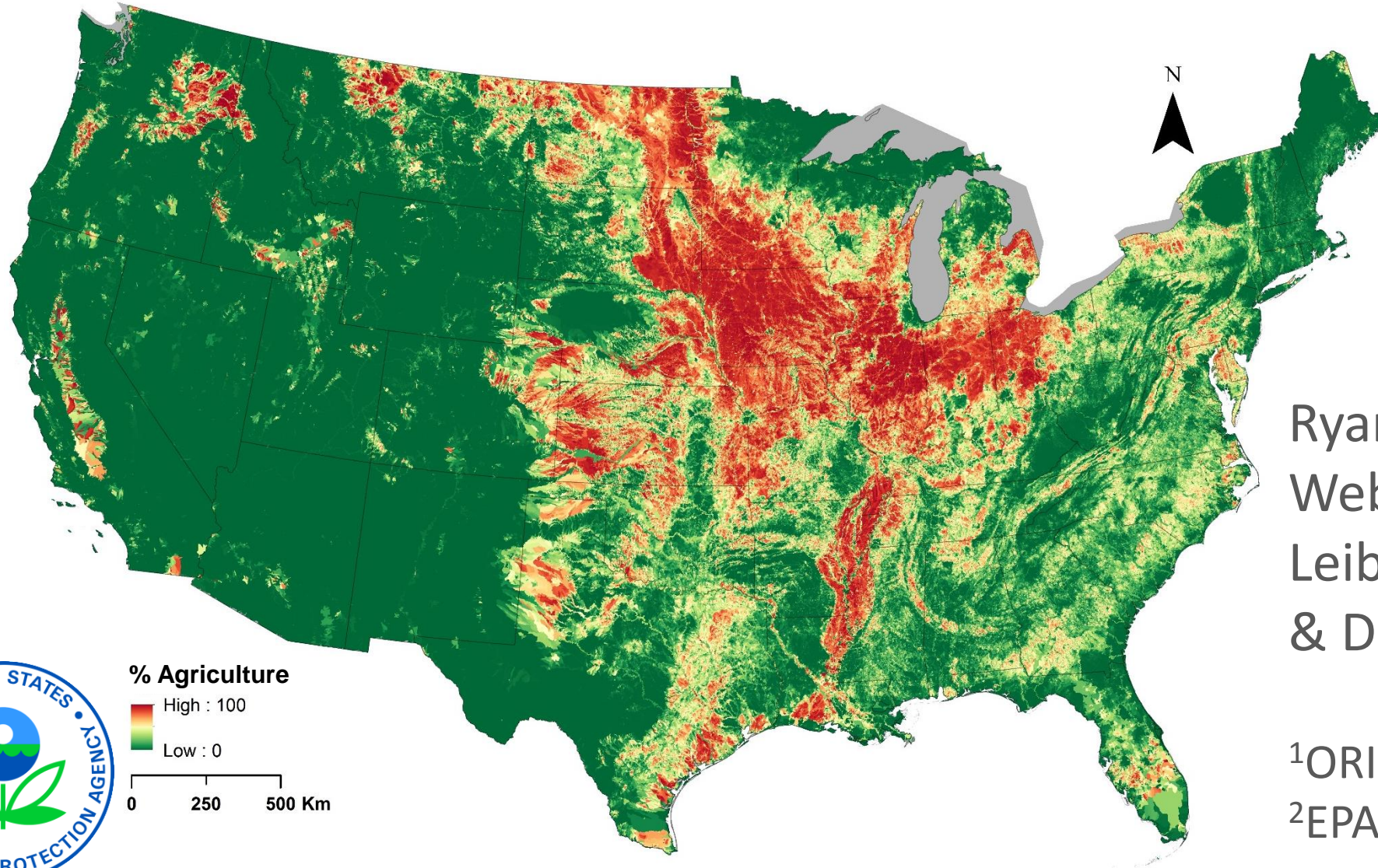


# The Stream-Catchment (StreamCat) Dataset:

A database of watershed metrics for the conterminous US



Ryan Hill<sup>1</sup>, Marc  
Weber<sup>2</sup>, Scott  
Leibowitz<sup>2</sup>, Tony Olsen<sup>2</sup>  
& Darren Thornbrugh<sup>1</sup>

<sup>1</sup>ORISE Postdoc

<sup>2</sup>EPA/ORD/NHEERL/WED



**% Agriculture**

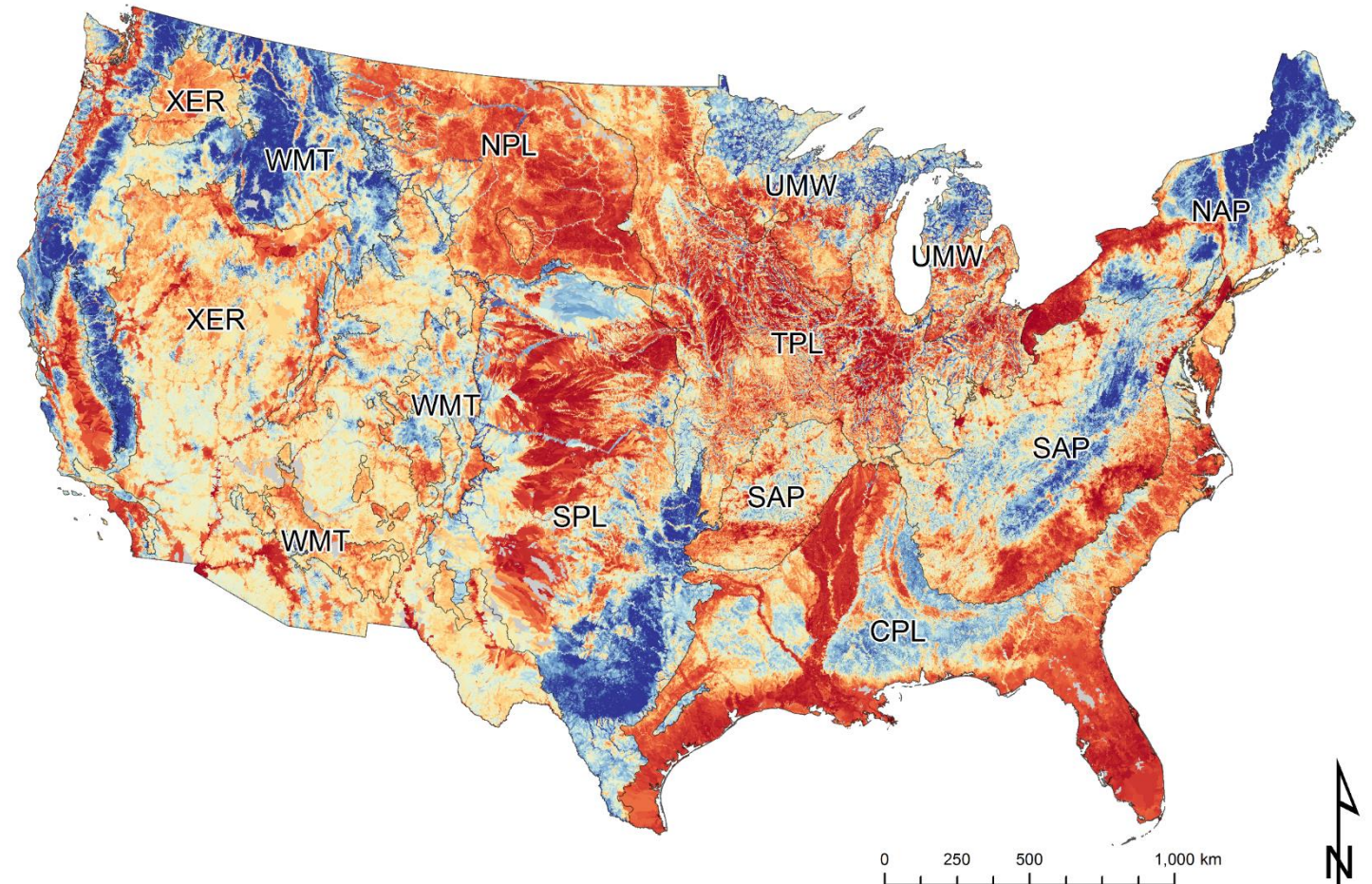
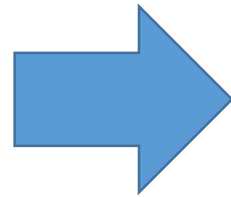
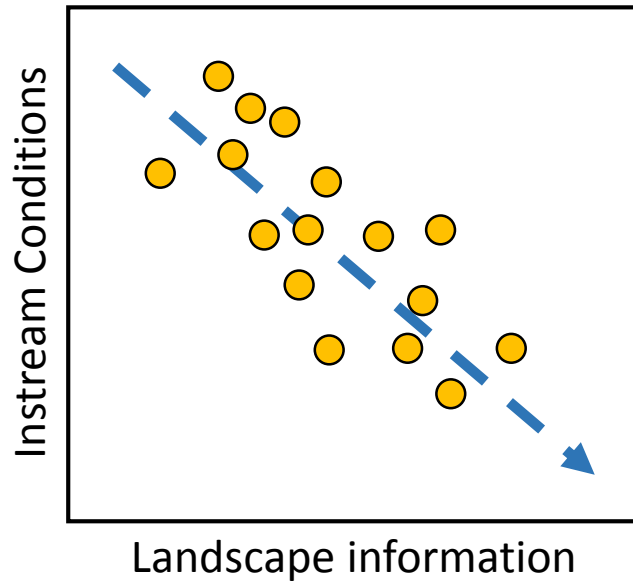
High : 100

Low : 0

0 250 500 Km

# StreamCat – Impetus

- Want to use models of instream conditions to predict and map these conditions nationally



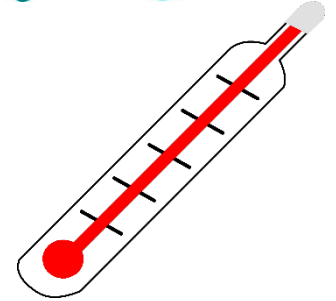
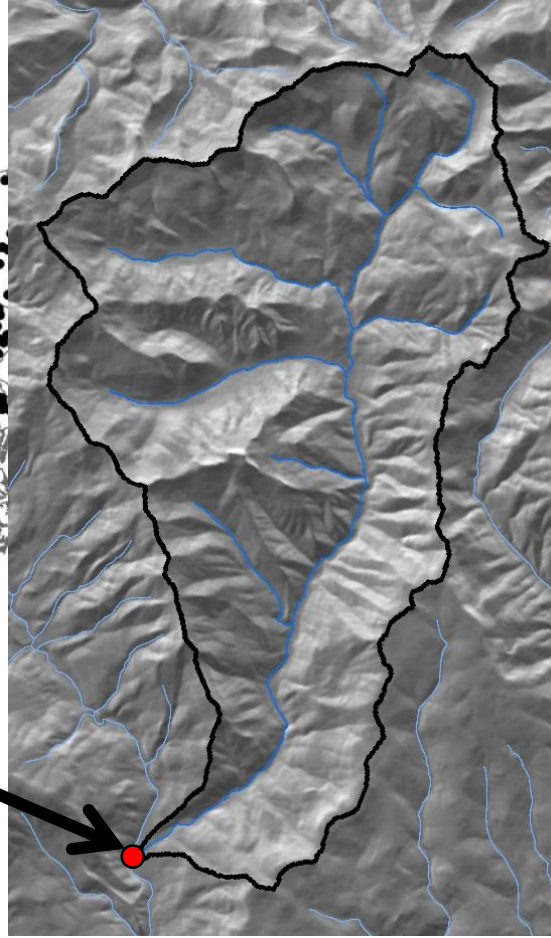
# StreamCat – Impetus

- US EPA National Rivers and Streams Assessment (NRSA) - >2,000 sample sites



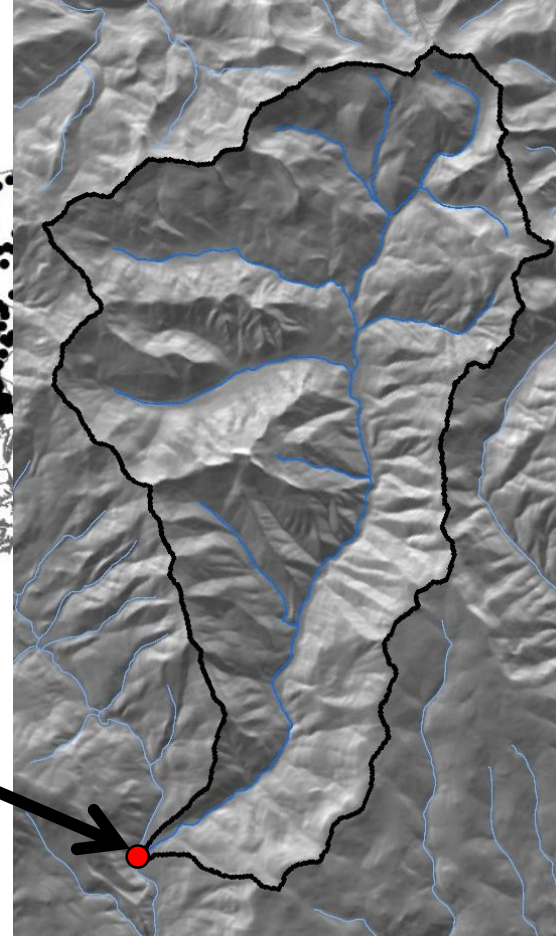
# StreamCat – Impetus

- Need information about landscapes that are upstream of specific sites...



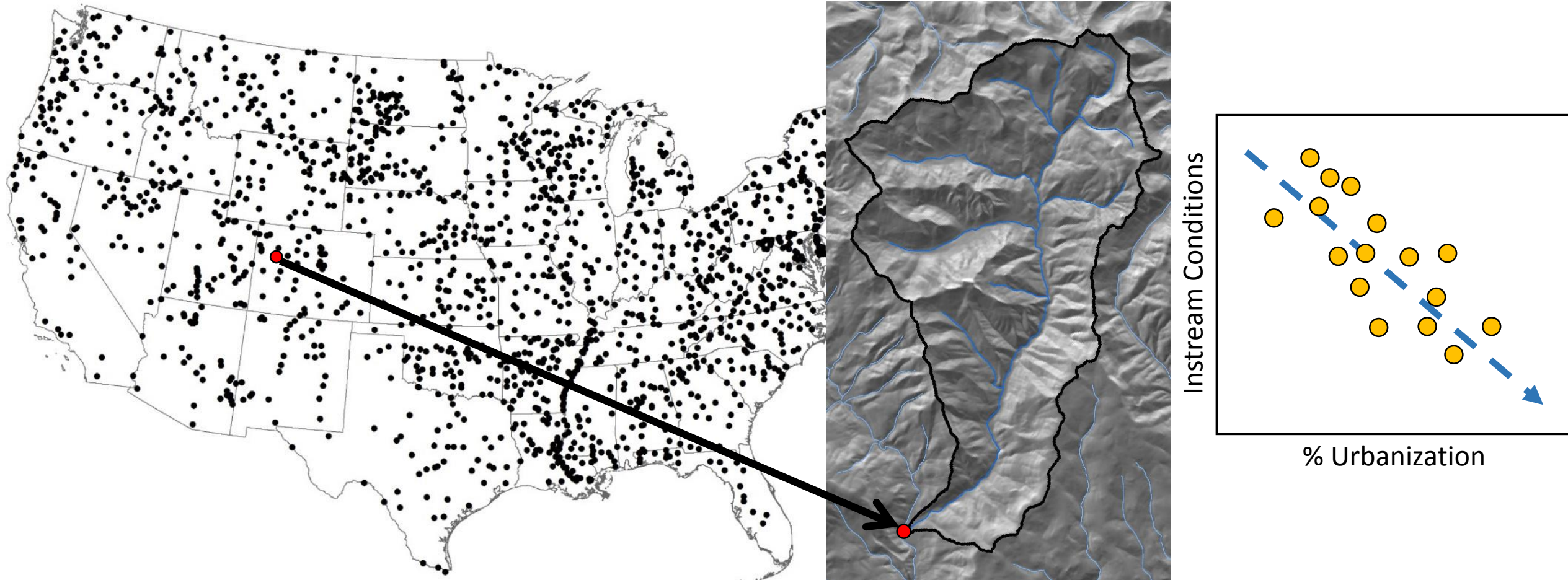
# StreamCat – Impetus

- Need information about landscapes that are upstream of specific sites...



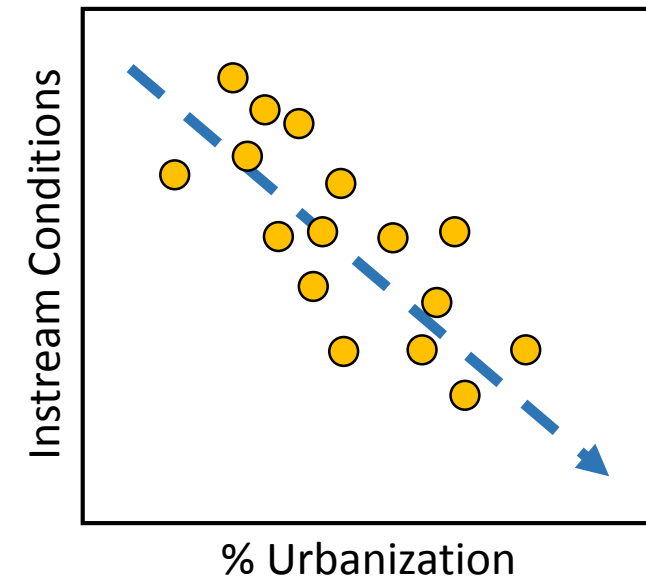
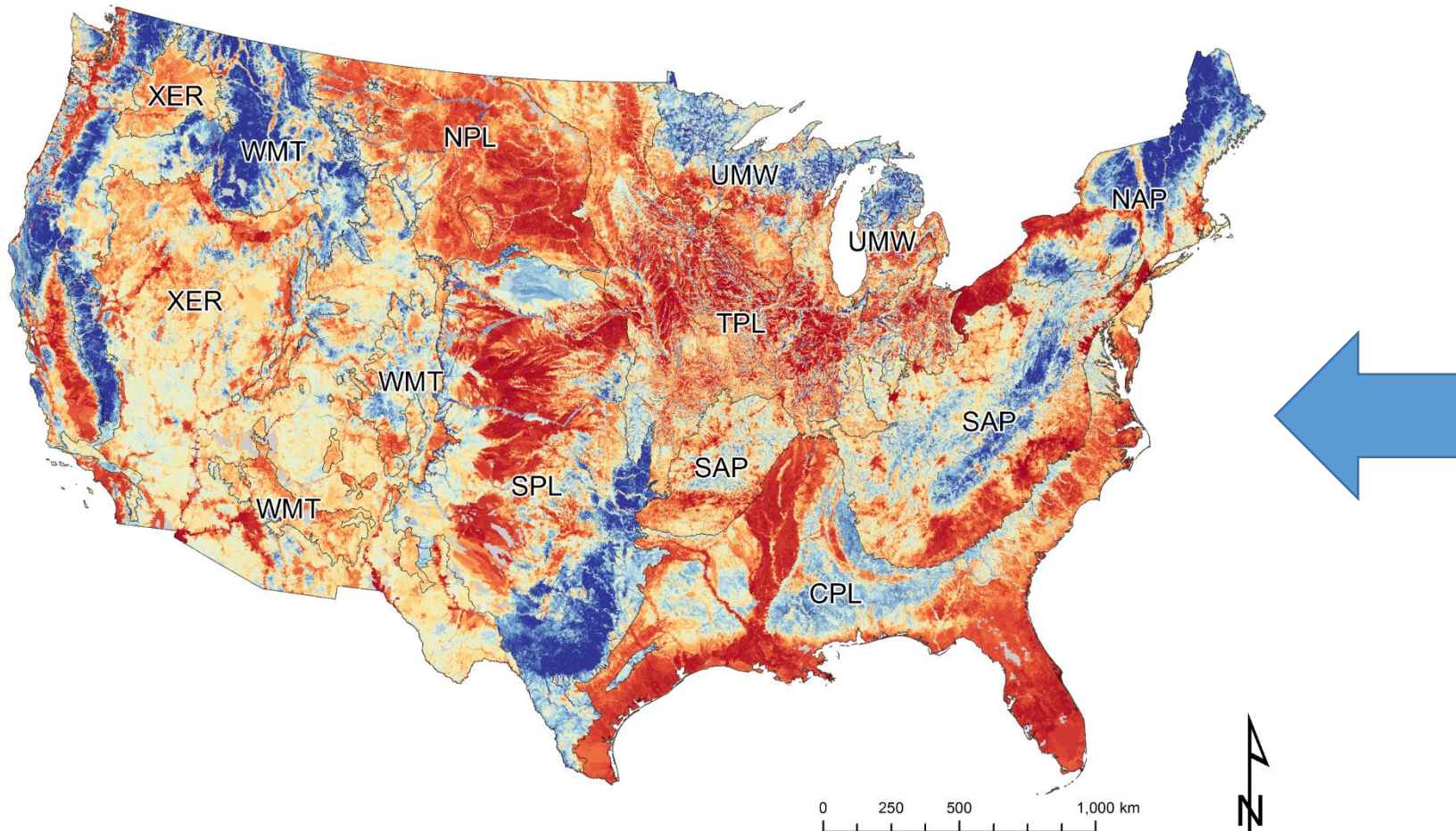
# StreamCat – Impetus

- ... to build empirical relationships between upstream features and instream conditions



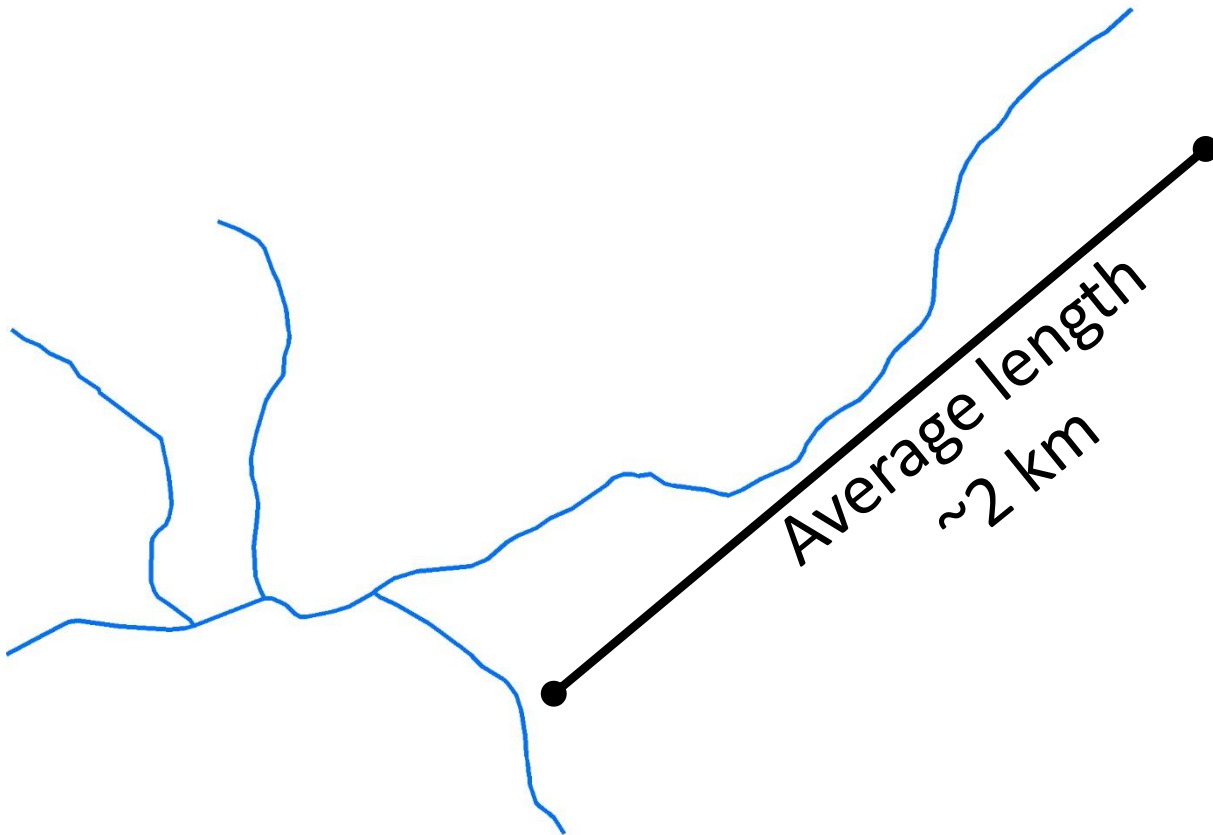
# StreamCat – Impetus

- Apply empirical models back to landscape to produce national maps of instream conditions



# NHDPlusV2 Framework

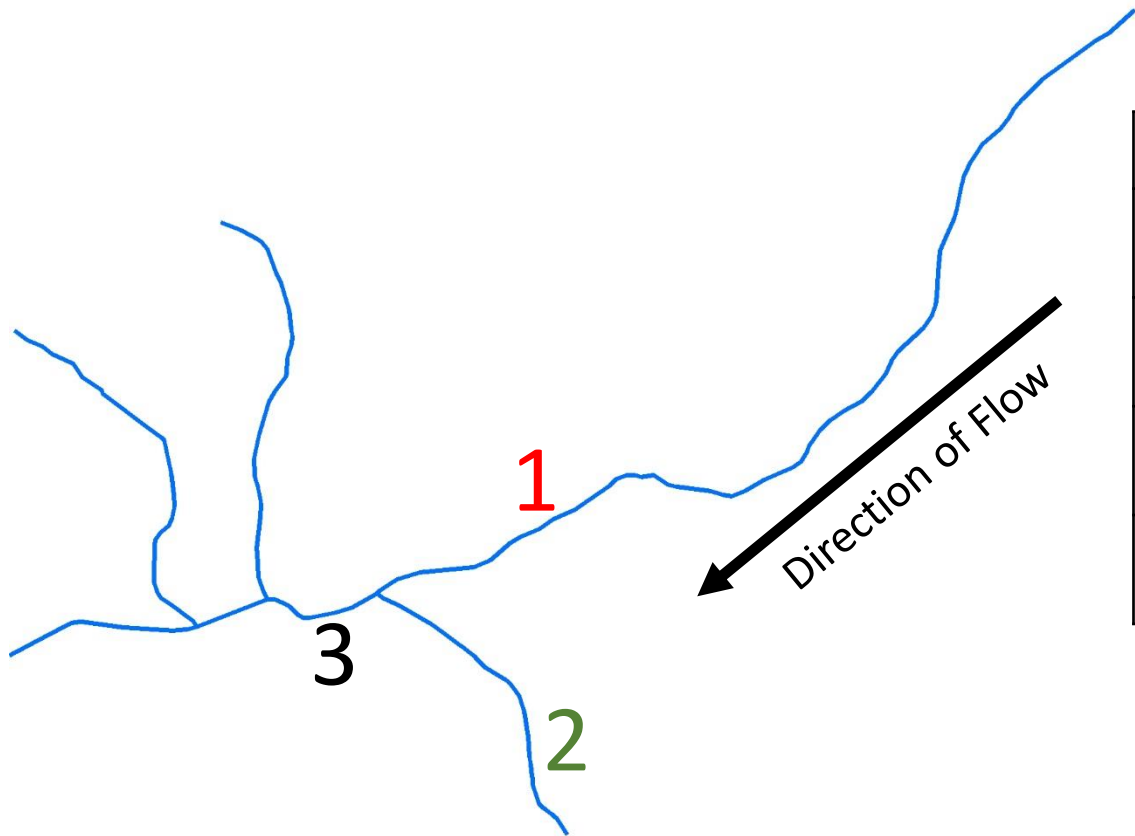
- National Hydrography Dataset Plus v.2 (NHDPlusV2)
- Line network of ~2.65 million streams in CONUS





# NHDPlusV2 Framework

- Topology (from-to relationships) documented for each line in NHDPlusV2

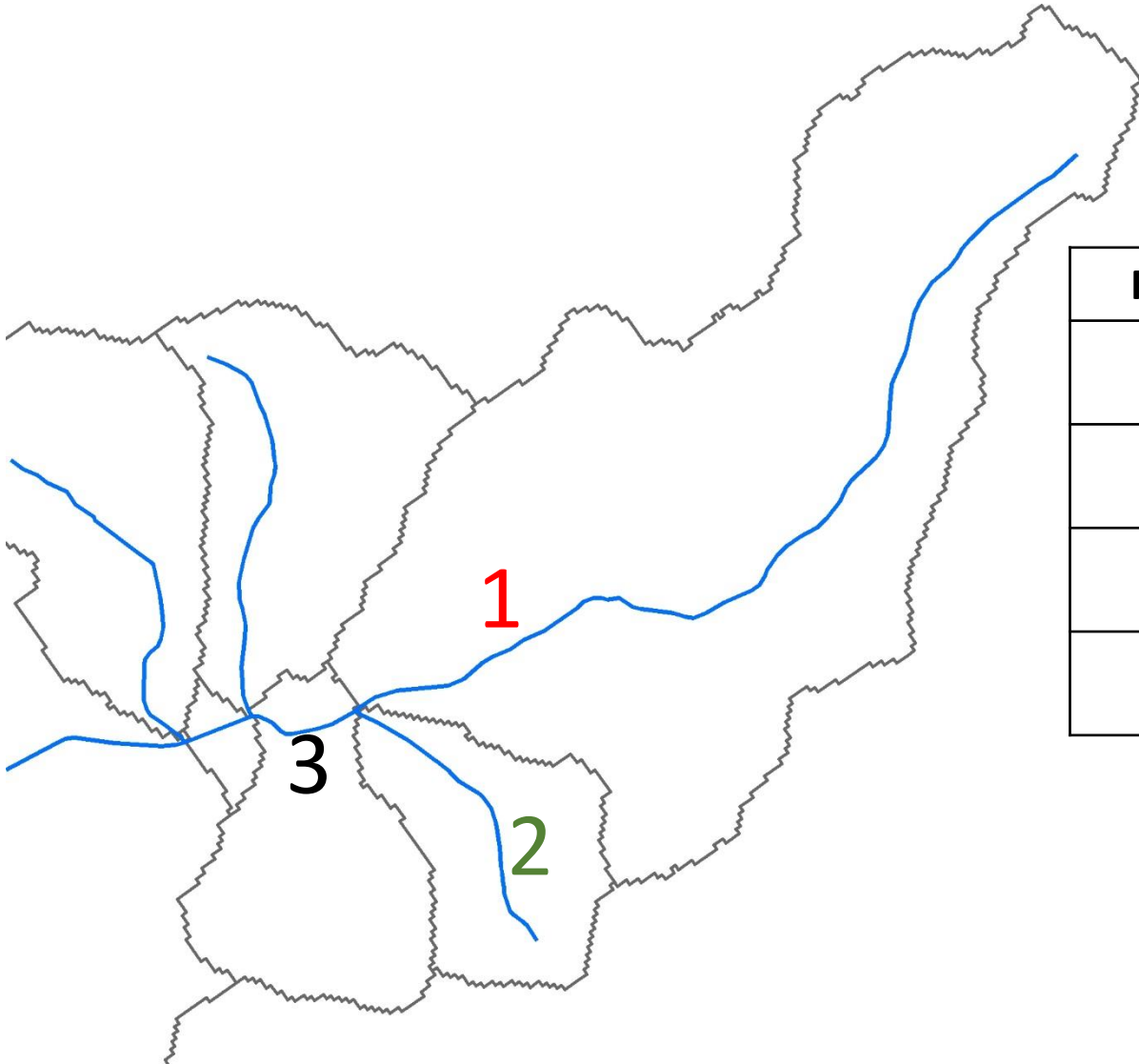


FROM	TO
-	1
-	2
1	3
2	3

·  
·  
·

# NHDPlusV2 Framework

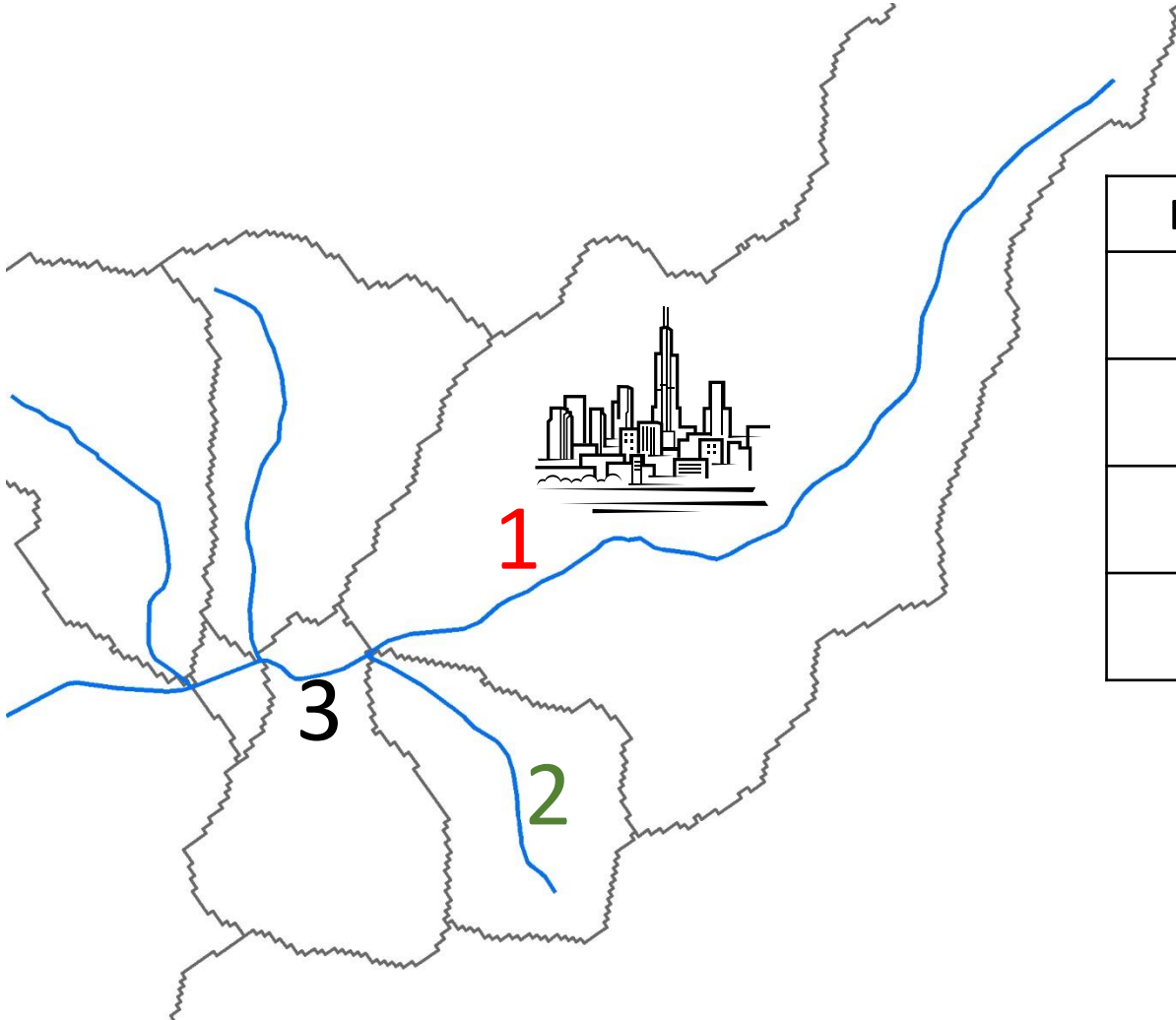
- Contains local catchments for each stream segment



FROM	TO
-	1
-	2
1	3
2	3
.	.
.	.
.	.

# NHDPlusV2 Framework

- Provides ability to overlay and accumulate information from upstream to downstream



FROM	TO
-	1
-	2
1	3
2	3
.	.
.	.
.	.

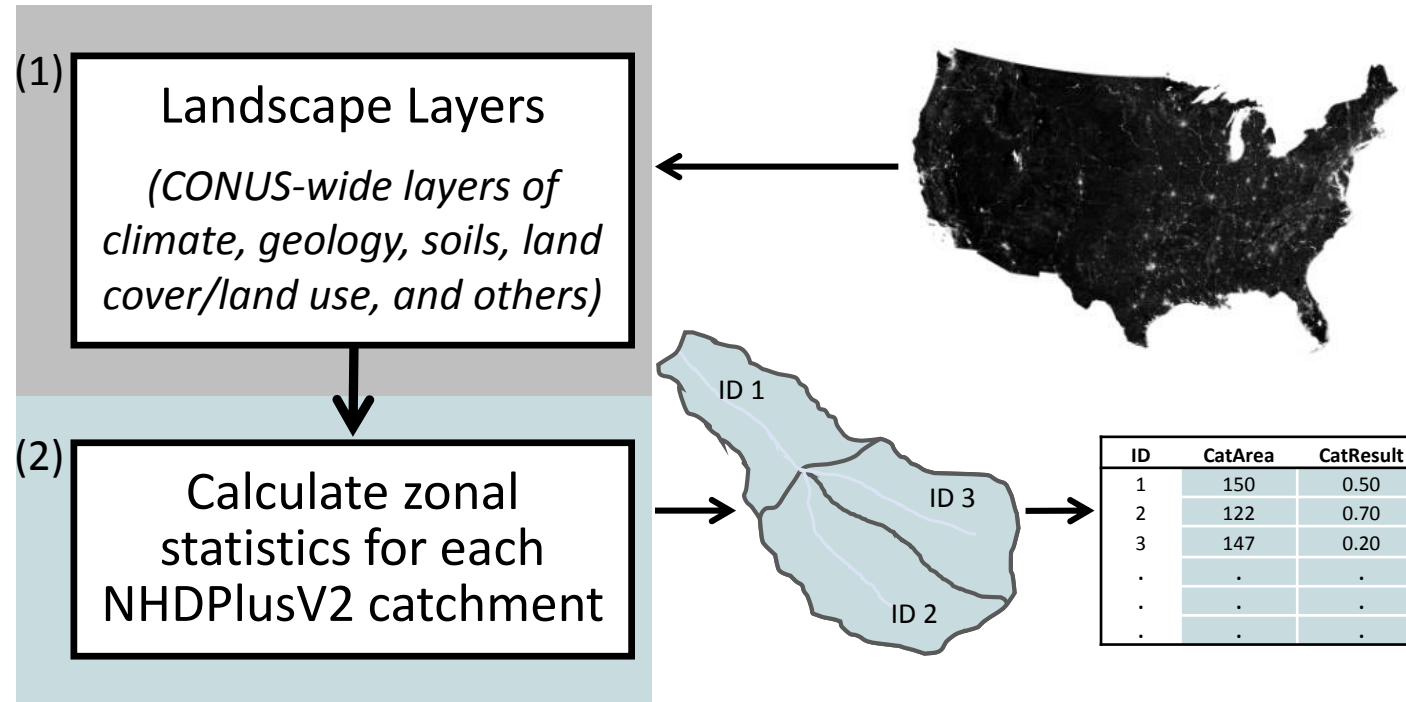
# StreamCat Approach –

(1)

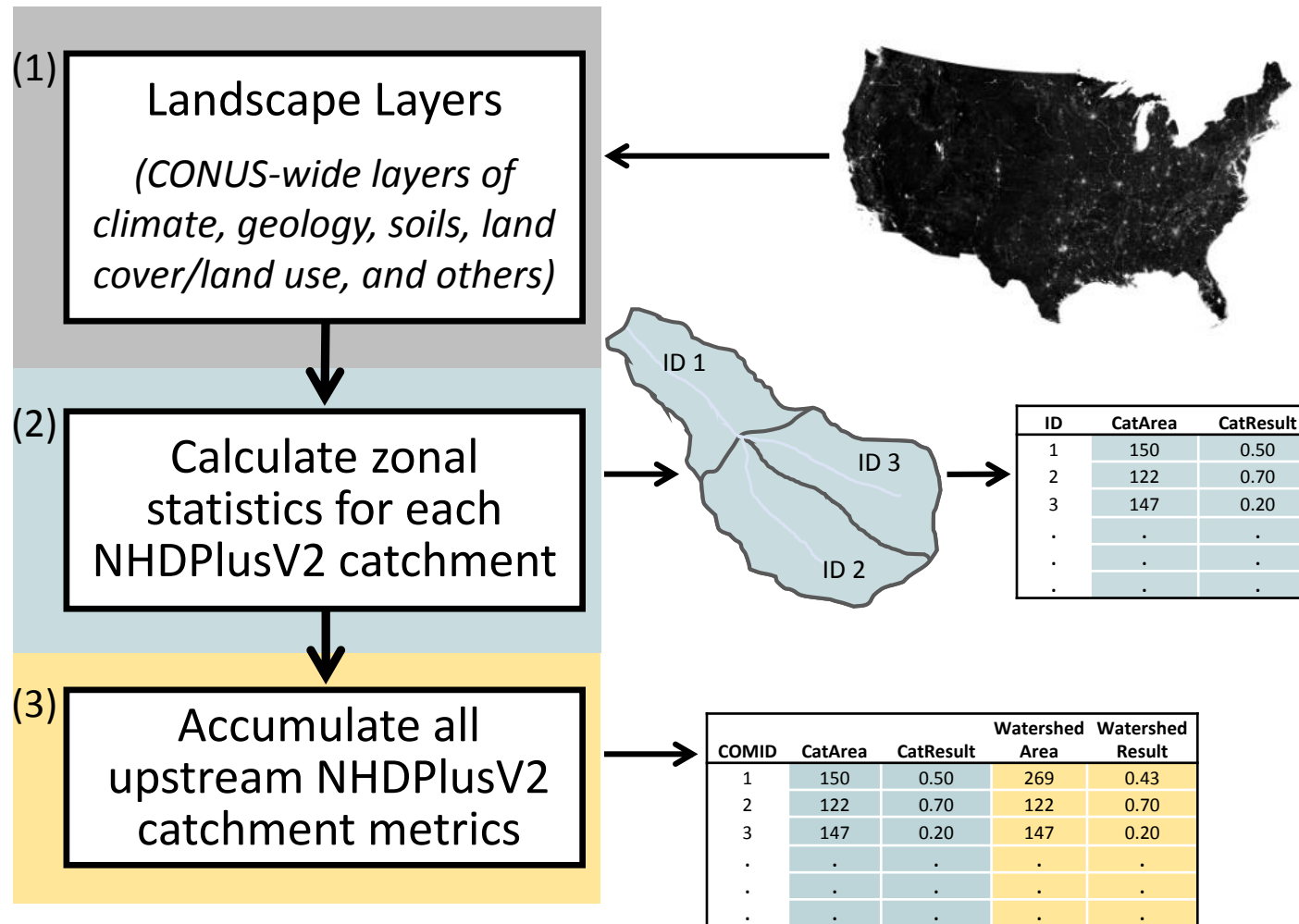
Landscape Layers  
*(CONUS-wide layers of  
climate, geology, soils, land  
cover/land use, and others)*



# StreamCat Approach –

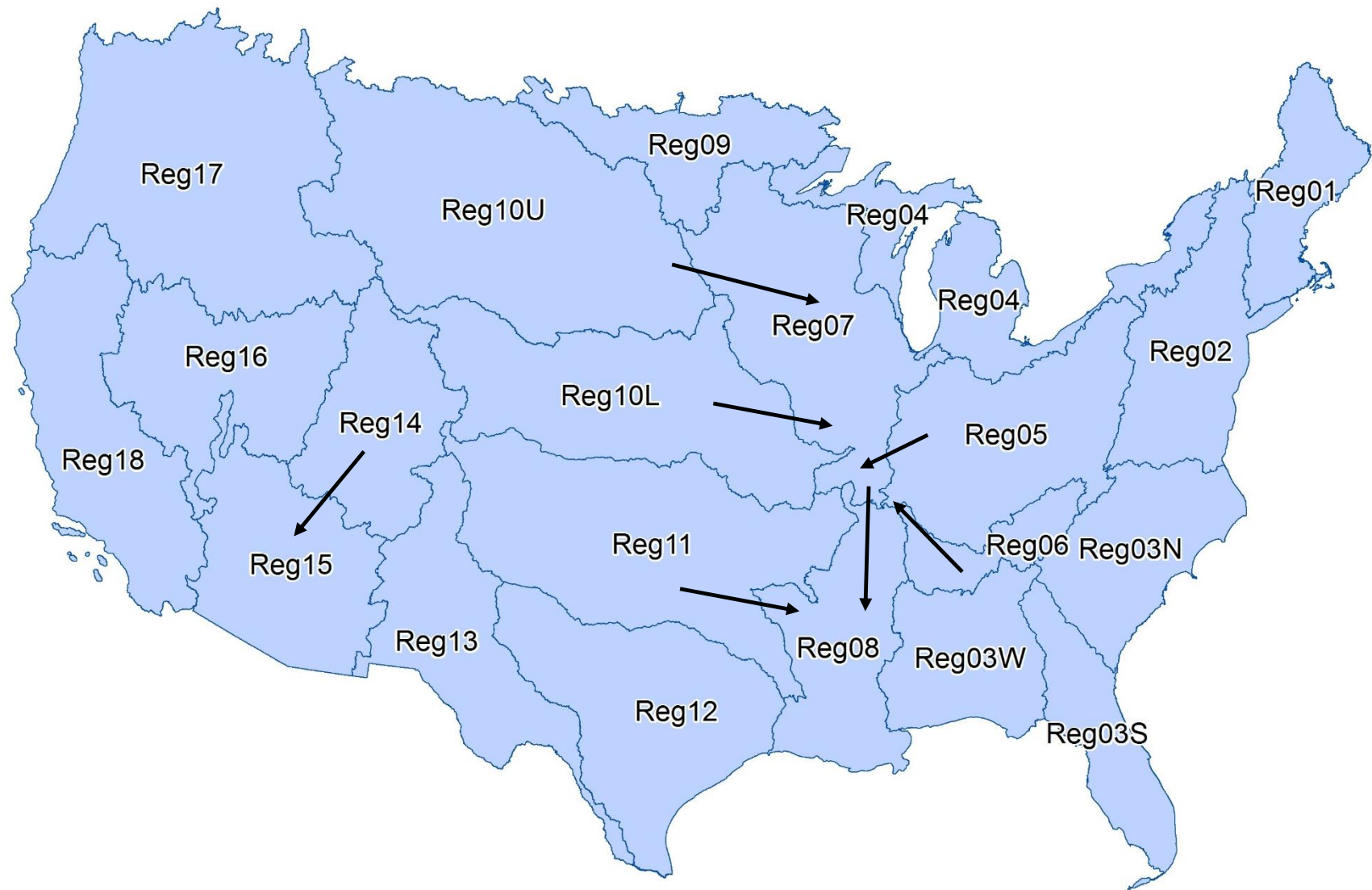


# StreamCat Approach –



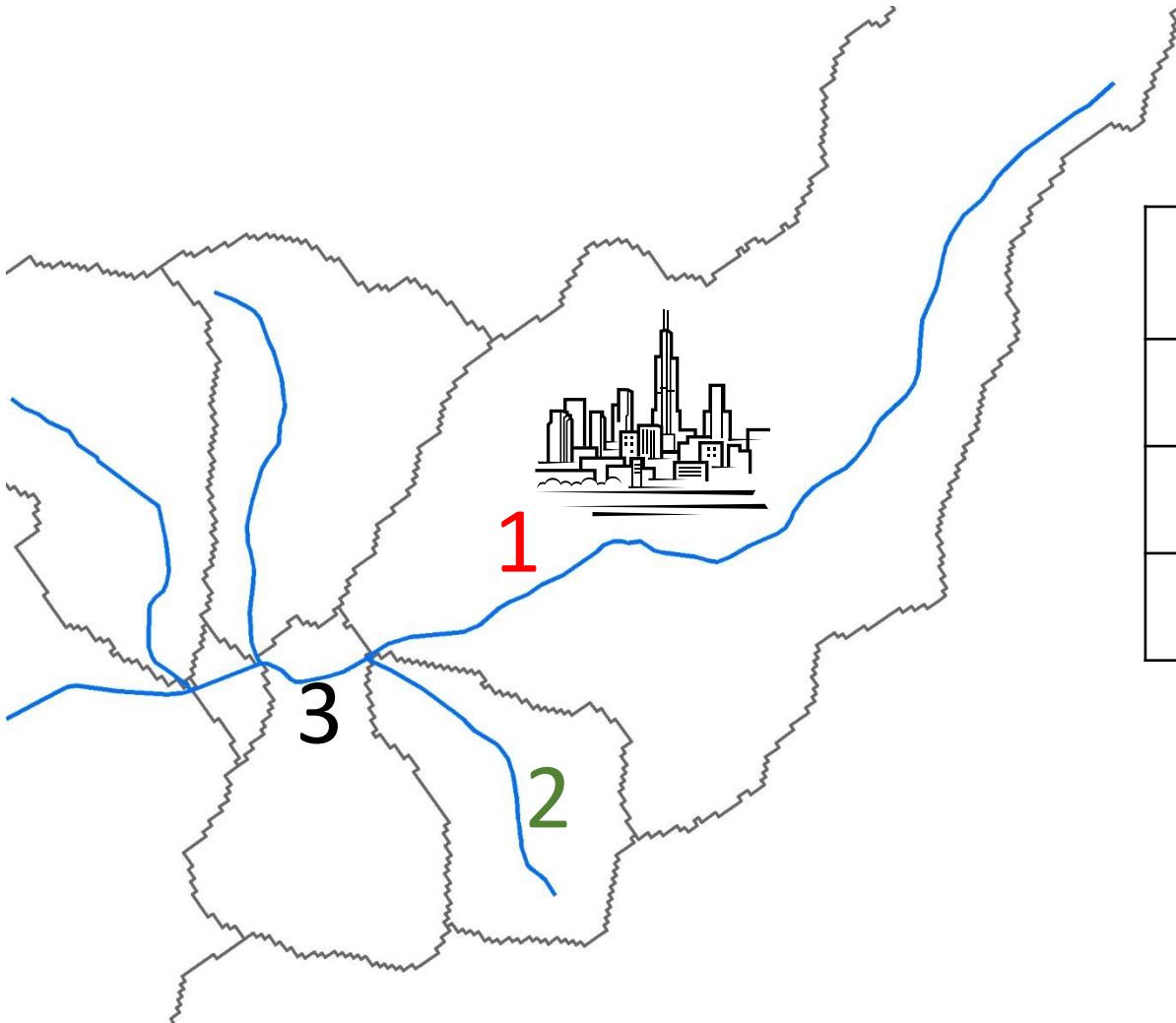
# StreamCat – Approach

- Connect HydroRegions



# StreamCat Dataset –

- Local catchment and full upstream watershed summaries for 2.65 million NHDPlusV2 stream segments in CONUS



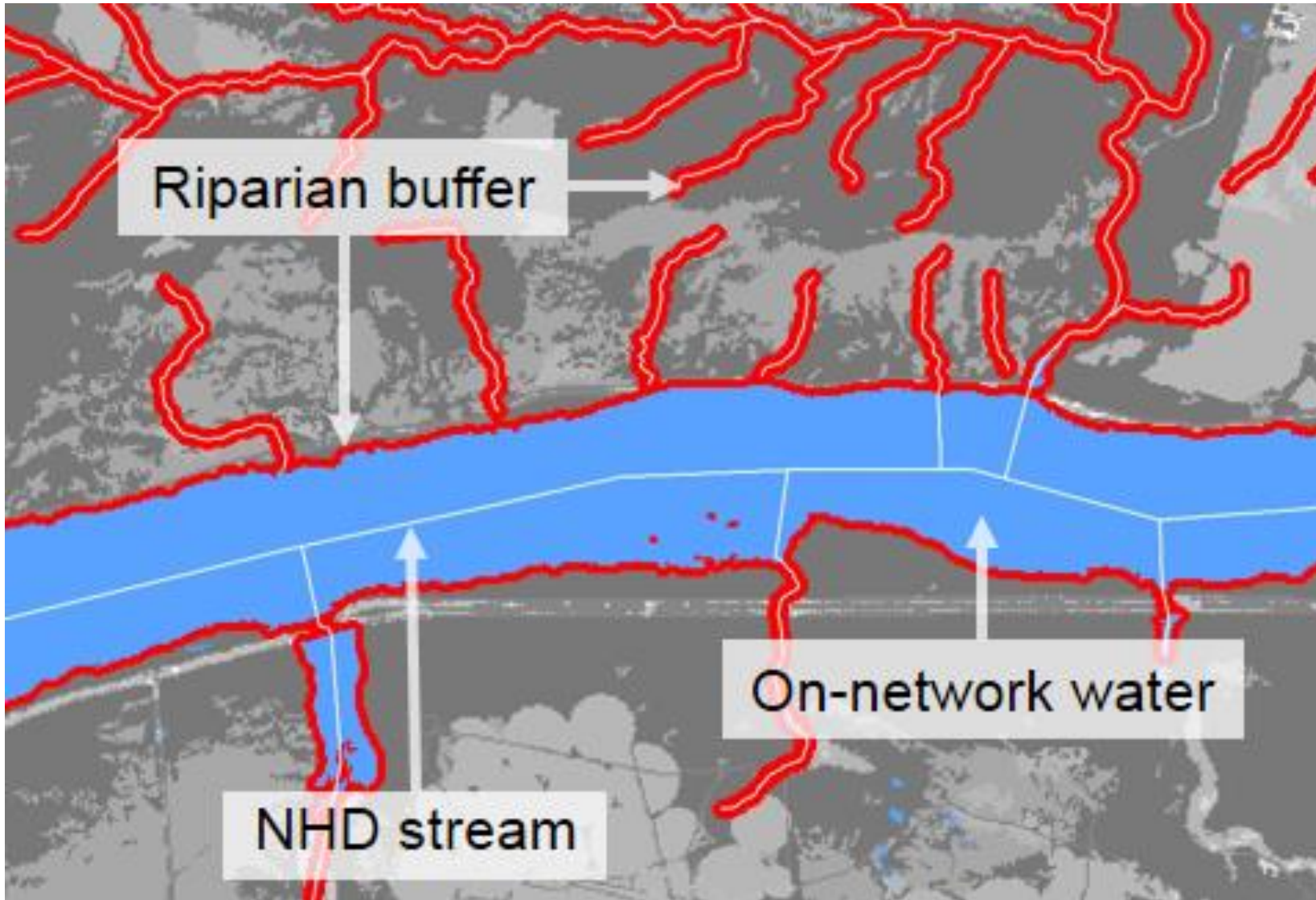
Stream ID	% Urban in local catchment	% Urban in full watershed
1	10.0	10.0
2	0	0
3	0	6.0

. . .  
. . .  
. . .



# StreamCat Dataset –

- Riparian buffers for some metrics



# StreamCat Dataset –

- *Transparency*
  - All processes implemented in R and Python (open source and ArcPy)
  - All code online – FTP site and now GitHub
  - Compliant metadata for each layer
- *Quality Assurance*
  - Standardized QA protocol for layers
  - All quality assurance documented online with R Markdown to show code and results

# StreamCat Dataset –



**JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION**  
AMERICAN WATER RESOURCES ASSOCIATION

**JAWRA**

## **THE STREAM-CATCHMENT (STREAMCAT) DATASET: A DATABASE OF WATERSHED METRICS FOR THE CONTERMINOUS UNITED STATES<sup>1</sup>**

*Ryan A. Hill, Marc H. Weber, Scott G. Leibowitz, Anthony R. Olsen, and Darren J. Thornbrugh<sup>2</sup>*

**ABSTRACT:** We developed an extensive database of landscape metrics for ~2.65 million stream segments, and their associated catchments, within the conterminous United States (U.S.): The Stream-Catchment (StreamCat) Dataset. These data are publically available (<http://www2.epa.gov/national-aquatic-resource-surveys/streamcat>) and greatly reduce the specialized geospatial expertise needed by researchers and managers to acquire landscape information for both catchments (i.e., the nearby landscape flowing directly into streams) and full upstream watersheds of specific stream reaches. When combined with an existing geospatial framework of the

# StreamCat Dataset –

- Data are publically available through EPA’s NARS website


## National Aquatic Resource Surveys Contact Us Share

**You are here:** [EPA Home](#) » [National Aquatic Resource Surveys](#) » [StreamCat](#)

### StreamCat

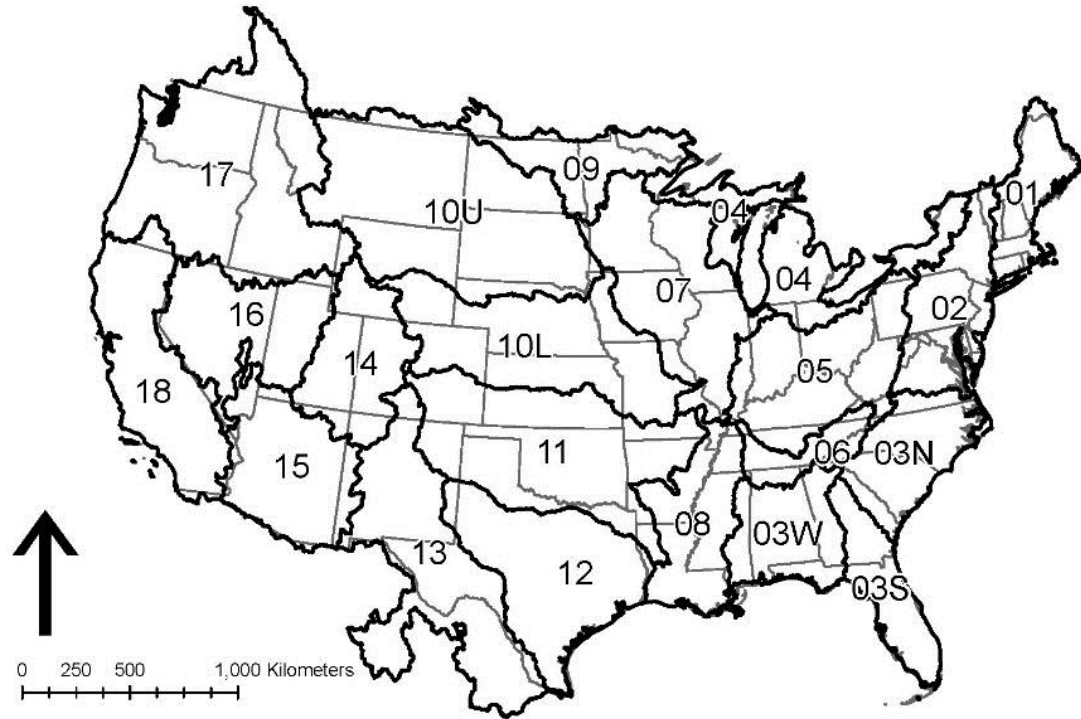
EPA's Office of Research and Development (ORD) has developed the Stream-Catchment (StreamCat) dataset, an extensive collection of landscape metrics for 2.6 million streams and associated catchments within the conterminous U.S. StreamCat includes both natural and human-related landscape features. The data are summarized both for individual stream catchments and for cumulative upstream watersheds, based on the National Hydrography Dataset Plus Version 2 geospatial framework.

StreamCat data are being utilized to develop national maps of aquatic condition and watershed integrity, and can be used to model and predict reference condition for the [National Rivers and Streams Assessment](#) (NRSA). The data will also be useful to states that are conducting similar assessments. StreamCat data, which are available to the public for [download](#), provide an important tool for stream researchers and managers to understand and characterize the Nation's rivers and streams.

A map of the United States showing stream catchment areas. The map is color-coded, with blue representing water bodies and other colors representing different landscape metrics. The map covers the entire United States, including Alaska and Hawaii.

# StreamCat Dataset –

- Data served as flat, comma-delimited text files
- Served by State or NHDPlus HydroRegion



- Unique ID in StreamCat direct match to unique ID in NHDPlusV2

# StreamCat Dataset –

- Welcome page, ReadMe page, and Updates page with extensive information on dataset

## StreamCat Dataset - Welcome

The StreamCat Dataset provides summaries of natural and anthropogenic landscape features for ~2.65 million streams, and their associated catchments, within the conterminous USA.

Similar data (LakeCat) on more than 356,000 USA lakes will be available within the next year.

## Updates & Corrections

Be sure to check [here](#) for recent updates and corrections to the StreamCat Dataset

## StreamCat Documentation

- A detailed description of the StreamCat Dataset and its development is now available in the [Journal of the American Water Resources Association](#).
- **StreamCat can be cited as:** Hill, Ryan A., Marc H. Weber, Scott G. Leibowitz, Anthony R. Olsen, and Darren J. Thornbrugh, 2016. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. Journal of the American Water Resources Association (JAWRA) 1-9. DOI: 10.1111/1752-1688.12372.
- The [Read Me](#) file provides additional detail and important information on how to access and use the StreamCat Dataset.
- The [Variable List](#) provides a quick reference of available watershed metrics.
- The [Data Dictionary](#) provides a comprehensive description of each watershed metric and how they were derived.
- Information on the steps we followed to acquire geospatial layers to develop the StreamCat Dataset, any manipulations done to those data, and quality assurance steps can be found in [Data Processing and Quality Assurance](#).
- The [StreamCat GitHub](#) page has the latest working versions of StreamCat code and a brief description of each script we used to develop the StreamCat Dataset, including examples of tables used to control the processing of StreamCat with .

<ftp://newftp.epa.gov/EPADataCommons/ORD/NHDPlusLandscapeAttributes/StreamCat/WelcomePage.html>

# StreamCat Dataset –

- Extensive documentation of 260 metrics (*more coming*)

	Variable Name	Description	Data Location
1	CatAreaSqKm	Area of local NHDPlus catchment (square km)	All tables
2	WsAreaSqKm	Watershed area (square km) at NHDPlus stream segment outlet, i.e., at the most downstream location of the vector line segment	All tables
3	CatAreaSqKmRp100	Area of local NHDPlus catchments (square km) within a 100-m buffer of NHD streams.	All RipBuf100 tables
4	WsAreaSqKmRp100	Watershed area (square km) within a 100-m buffer of NHD streams	All RipBuf100 tables

• • •  
• • •  
• • •

256	PctShrb2011WsRp100	% of watershed area classified as shrub/scrub land cover (NLCD 2011 class 52) within a 100-m buffer of NHD streams	NLCD2011RipBuf100_<RegionID>.csv, NLCD2011RipBuf100_<StateID>.csv
257	PctGrs2011Cat	% of catchment area classified as grassland/herbaceous land cover (NLCD 2011 class 71)	NLCD2011_<RegionID>.csv, NLCD2011_<StateID>.csv
258	PctGrs2011Ws	% of watershed area classified as grassland/herbaceous land cover (NLCD 2011 class 71)	NLCD2011_<RegionID>.csv, NLCD2011_<StateID>.csv
259	PctGrs2011CatRp100	% of catchment area classified as grassland/herbaceous land cover (NLCD 2011 class 71) within a 100-m buffer of NHD streams	NLCD2011RipBuf100_<RegionID>.csv, NLCD2011RipBuf100_<StateID>.csv
260	PctGrs2011WsRp100	% of watershed area classified as grassland/herbaceous land cover (NLCD 2011 class 71) within a 100-m buffer of NHD streams	NLCD2011RipBuf100_<RegionID>.csv, NLCD2011RipBuf100_<StateID>.csv

# StreamCat Dataset –

- Code and detailed description of running code published on GitHub

USEPA / StreamCat

Watch 12 Star 4 Fork 2

Code Issues 0 Pull requests 0 Wiki Pulse Graphs

No description or website provided.

84 commits 1 branch 0 releases 4 contributors

Branch: master New pull request Create new file Upload files Find file Clone or download

debbotr update makeFinalTables fro dams and NABD Latest commit 70164ae 5 days ago

File	Description	Time
.gitignore	handle Dens title with summarized metrics	5 days ago
ControlTable_StreamCat.csv	Added table and modified ReadMe	4 months ago
FieldCalcTable.csv	Adding control tables	4 months ago
FinalTablesStates.py	Folding in Rick's commits and adding FinalTablesStates.py	3 months ago
InterVPU.csv	add InterVPU table	3 months ago
Lithology_lookup.csv	Adding control tables	4 months ago
MakeFinalTables.py	handle Dens title with summarized metrics	5 days ago
NLCD2008_lookup.csv	Adding control tables	4 months ago
README.md	Add explicit instructions to run StreamCat.py from conda shell	18 days ago
RasterControlTable.csv	Adding control tables	4 months ago
ReclassTable.csv	Adding control tables	4 months ago
RiparianBuffer.py	Updates mainly to StreamCat_PreProcessing and RiparianBuffer scripts	4 months ago
StreamCat.Rproj	Added changes to pre-processing	3 months ago
StreamCat.py	fix numpy vector creation	22 days ago
StreamCat_PreProcessing.py	format fix, test push to repo	3 months ago
StreamCat_functions.py	minor commit	19 days ago
streamcat.txt	Add info on installing streamcat environment to readme and add stream...	18 days ago

## StreamCat

### Description:

The StreamCat Dataset (<http://www2.epa.gov/national-aquatic-resource-surveys/streamcat>) provides summaries of natural and anthropogenic landscape features for ~2.65 million streams, and their associated catchments, within the conterminous USA. This repo contains code used in StreamCat to process a suite of landscape rasters to watersheds for streams and their associated catchments (local reach contributing area) within the conterminous USA using the [NHDPlus Version 2](#) as the geospatial framework.

### Necessary Python Packages and Installation Tips

The scripts for StreamCat rely on several python modules a user will need to install such as numpy, pandas, gdal, fiona, rasterio, geopandas, shapely, pysal, and ArcPy with an ESRI license (minimal steps still using ArcPy). We highly recommend using a scientific python distribution such as [Anaconda](#) or [Enthought Canopy](#). We used the conda package manager to install necessary python modules. Our essential packages and versions used are listed below (Windows 64 and Python 2.7.11):

Package	Version
fiona	1.6.3
gdal	1.11.4
geopandas	0.2.0.dev
geos	3.4.2
libgdal	2.0.0
numpy	1.10.1
pandas	0.18.1
pyproj	1.9.5
pysal	1.10.0
rasterio	0.34.0
shapely	1.5.15



# StreamCat Dataset –

- Metadata in readable format for all metrics

The StreamCat Dataset: Accumulated Attributes for NHDPlusV2 (Version 2.1) Catchments for the Conterminous United States: State Soil Geographic Database (STATSGO)	
FGDC Metadata Show Definitions	
Description   Spatial   Data Structure   Data Quality   Data Source   Data Distribution   Metadata	
+ Resource Description	
<b>Citation</b>	
<i><b>Title:</b></i> The StreamCat Dataset: Accumulated Attributes for NHDPlusV2 (Version 2.1) Catchments for the Conterminous United States: State Soil Geographic Database (STATSGO) <i><b>Originators:</b></i> US Environmental Protection Agency <i><b>Publisher:</b></i> US EPA Office of Research & Development (ORD) - National Health and Environmental Effects Research Laboratory (NHEERL) <i><b>Publication place:</b></i> Corvallis, OR <i><b>Publication date:</b></i> 2015 <i><b>Data location:</b></i> <a href="ftp://newftp.epa.gov/EPADataCommons/ORD/NHDPlusLandscapeAttributes/StreamCat/WelcomePage.html">ftp://newftp.epa.gov/EPADataCommons/ORD/NHDPlusLandscapeAttributes/StreamCat/WelcomePage.html</a> , <a href="http://www.epa.gov/national-aquatic-resource-surveys/streamcat">http://www.epa.gov/national-aquatic-resource-surveys/streamcat</a>	
<b>Description</b>	
<b>Time Period of Data</b>	
<b>Status</b>	
<b>Key Words</b>	
<b>Data Access Constraints</b>	
<b>Data Security Information</b>	
+ Spatial Reference Information	
<b>Horizontal Coordinate System</b>	
<b>Spatial Domain</b>	
+ Data Structure and Attribute Information	
<b>Overview</b>	
<b>Attributes of STATSGO_Set1</b>	
<b>Attributes of STATSGO_Set2</b>	
+ Data Quality and Accuracy Information	
<b>General</b>	
<b>Positional Accuracy</b>	
+ Data Source and Process Information	
<b>Data Sources</b>	
<b>Process Steps</b>	
+ Data Distribution Information	
<b>General</b>	
<b>Distribution Point of Contact</b>	

# StreamCat Dataset –

- Documented QA processing steps

## Data Acquisition, Processing, & Quality Assurance

*Ryan Hill and Marc Weber*

*Originated: Wednesday, May 06, 2015*

- [Introduction](#)
- [Processing and QA Steps](#)
  - [1. Landscape Layer Acquisition](#)
  - [2. Layer Checks](#)
  - [3. Layer Manipulation](#)
  - [4. Zonal Statistics Geoprocessing](#)
  - [5. Catchment Results Cleanup](#)
  - [6. Metric Accumulation & QA](#)
  - [7. Final Tables & QA](#)

### Introduction

This document describes the main steps that we followed to develop the StreamCat Dataset, including quality assurance (QA) procedures.

In addition to this document, data acquisition, modification, and QA were tracked in an [spreadsheet](#). Processing steps and data sources are further documented in individual [metadata files](#). Additional details on code and associated output are contained in the [Code ReadMe](#).

### Processing and QA Steps

#### 1. Landscape Layer Acquisition

The development of the StreamCat Dataset is a part of a multi-phased project. For the first phase of this project, we identified potential landscape metrics through a literature review. Specifically, we used several landscape metrics identified in [Carlisle et al. 2009](#), [Falcone et al. 2010](#), and [Wang et al. 2011](#). Data were downloaded from online sources where available. Some layers (e.g., STATSGO Soils layers) were obtained directly from James Falcone of the USGS.

For the second phase of this project, we conducted an extensive search for publically available, conterminous USA-wide landscape layers that we hypothesized could improve our representation of natural and anthropogenic watershed features. These layers included climate, groundwater usage, forest cover change (yrs. 2001-2010), atmospheric N deposition, and fish passage barriers. These data will be added to this FTP site as they become available.

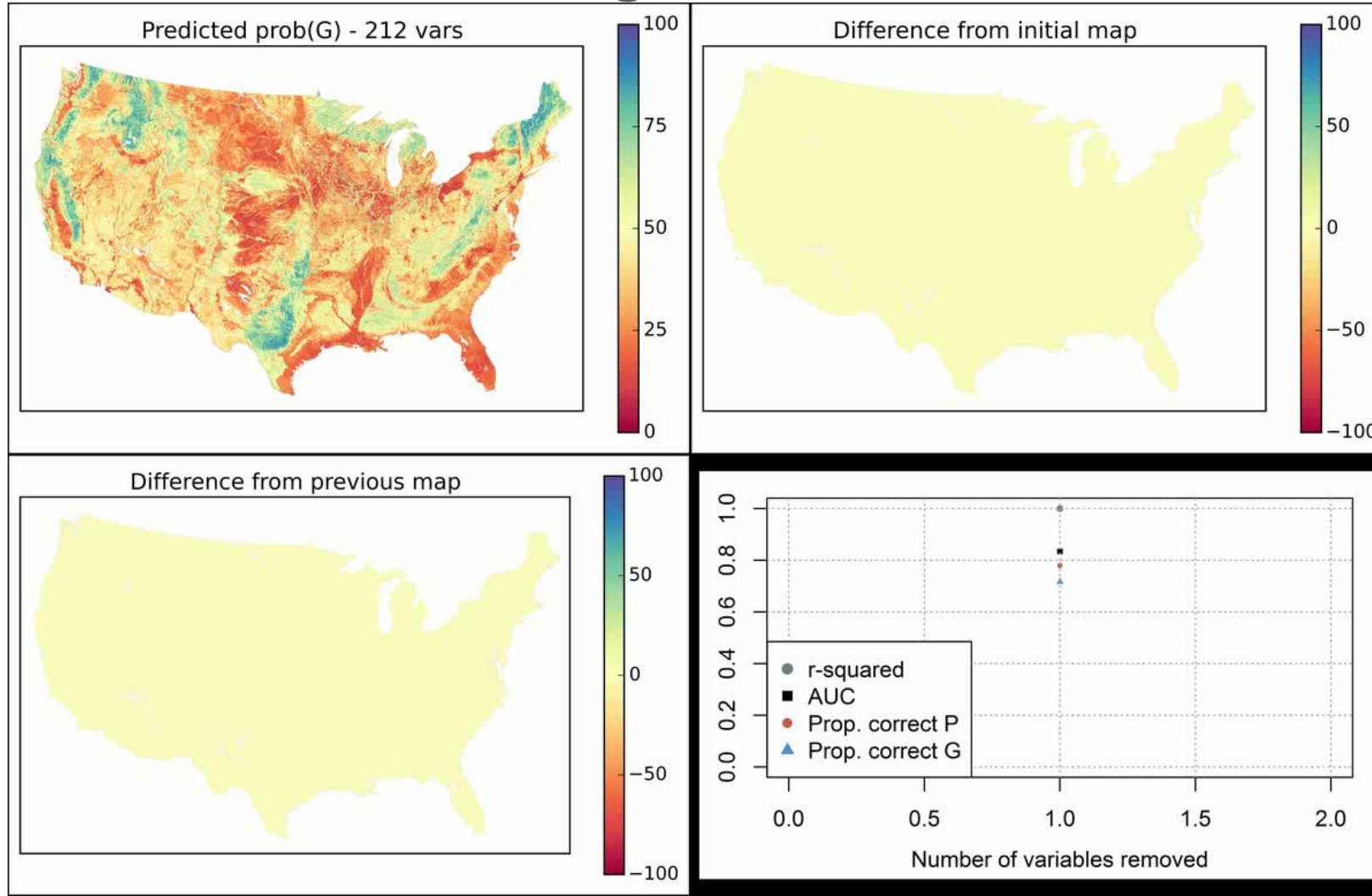
# StreamCat Dataset – *So what?*

- Provides access to 260 landscape metrics for 2.65 million streams in conterminous US
- Minimal GIS expertise to link field samples to watershed data



# StreamCat Dataset – *So what?*

- Can rapidly apply analyses to new, unsampled locations
- Helps understand and diagnose model behaviors



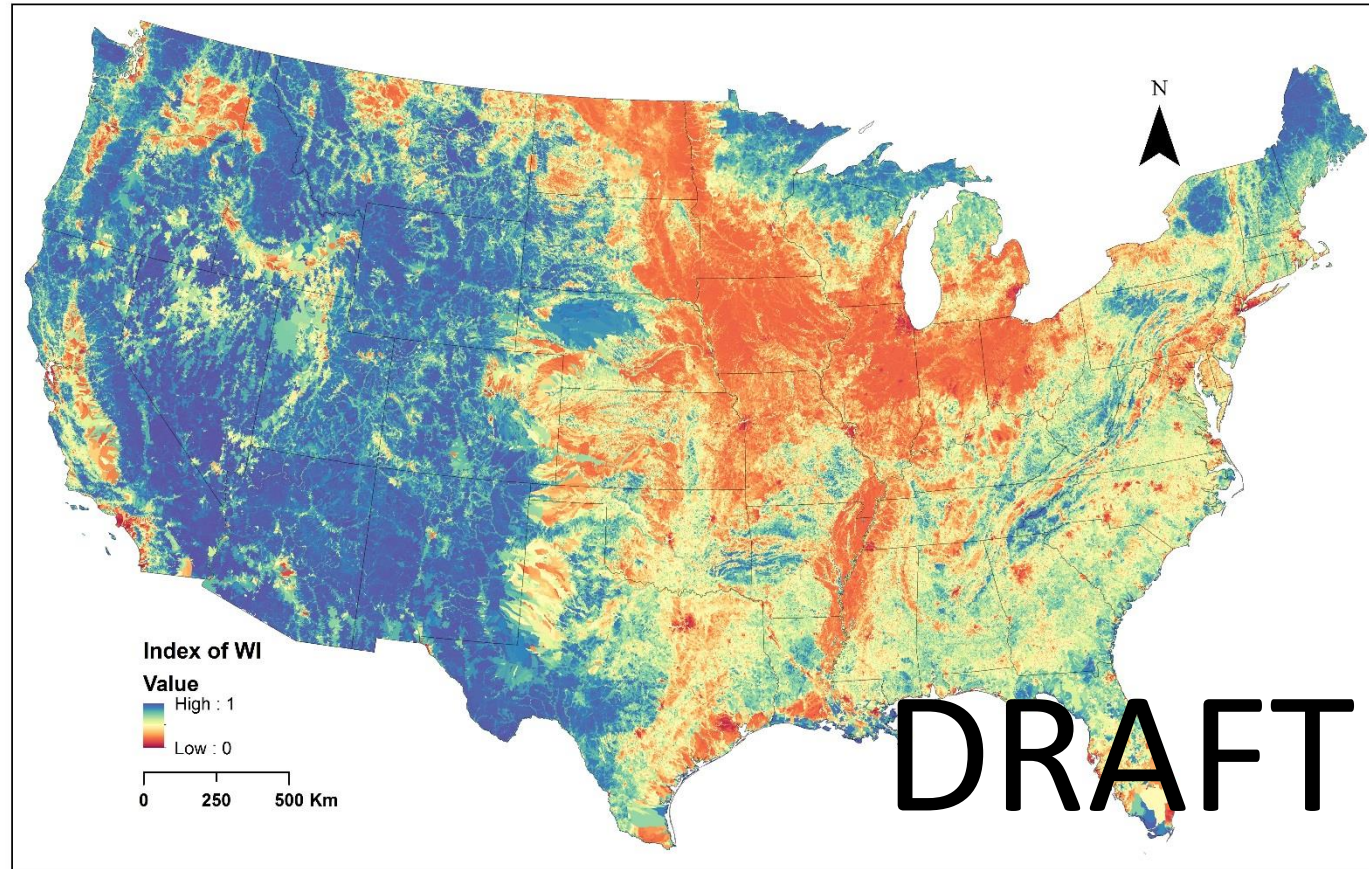
# StreamCat – Applications

- Applying ‘Watershed Integrity’ index to all NHDPlusV2 streams

River Research and Applications (2015)

A WATERSHED INTEGRITY DEFINITION AND ASSESSMENT APPROACH TO  
SUPPORT STRATEGIC MANAGEMENT OF WATERSHEDS

J. E. FLOTEMERSCH<sup>a\*</sup>, S. G. LEIBOWITZ<sup>b</sup>, R. A. HILL<sup>c</sup>, J. L. STODDARD<sup>b</sup>, M. C. THOMS<sup>d</sup> AND R. E. THARME<sup>e†</sup>



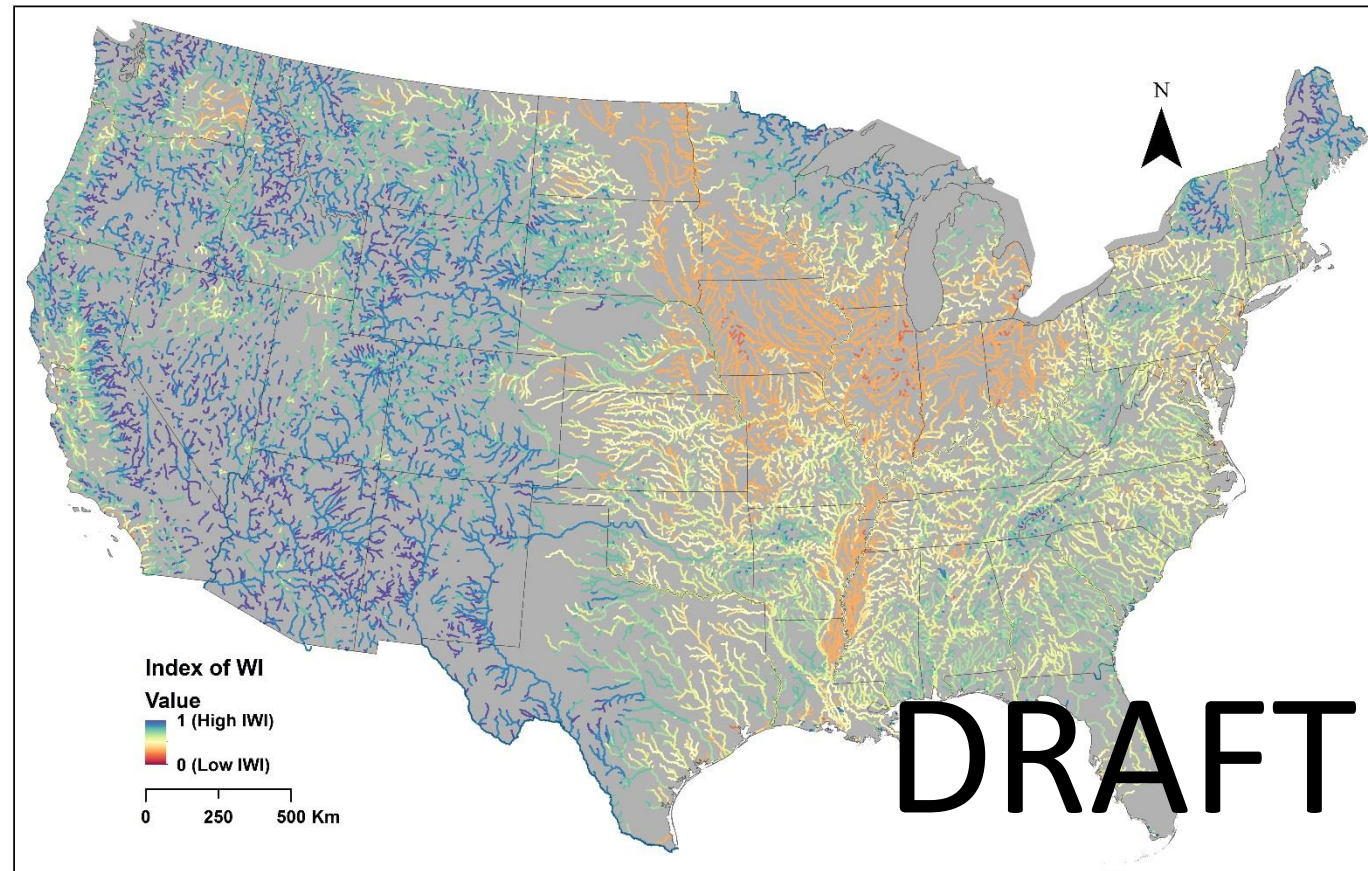
# StreamCat – Applications

- Applying ‘Watershed Integrity’ index to all NHDPlusV2 streams

River Research and Applications (2015)

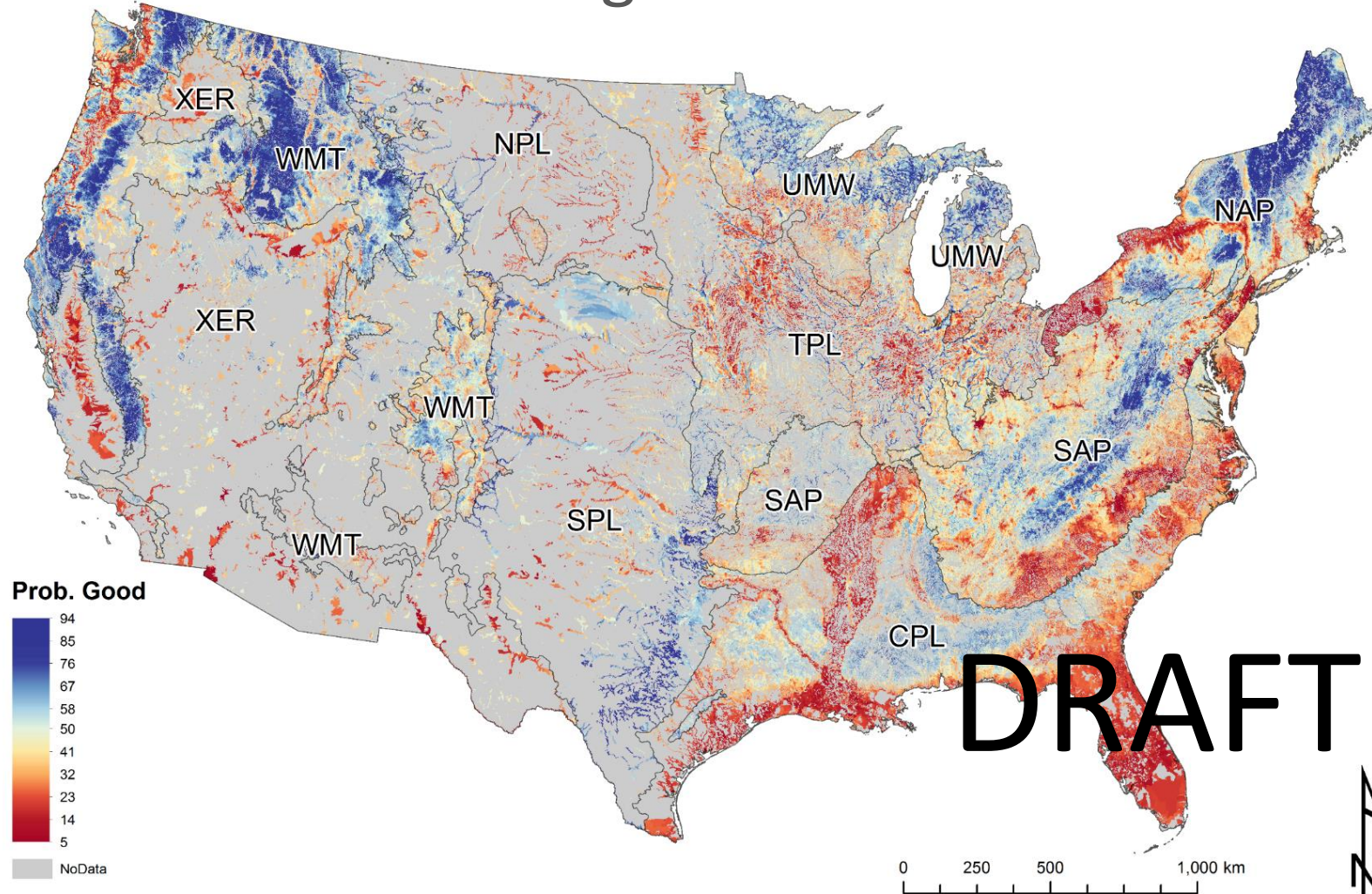
A WATERSHED INTEGRITY DEFINITION AND ASSESSMENT APPROACH TO  
SUPPORT STRATEGIC MANAGEMENT OF WATERSHEDS

J. E. FLOTEMERSCH<sup>a\*</sup>, S. G. LEIBOWITZ<sup>b</sup>, R. A. HILL<sup>c</sup>, J. L. STODDARD<sup>b</sup>, M. C. THOMS<sup>d</sup> AND R. E. THARME<sup>e†</sup>



# StreamCat – Applications

- Linking StreamCat to EPA's National Rivers and Streams Assessment to model biological condition



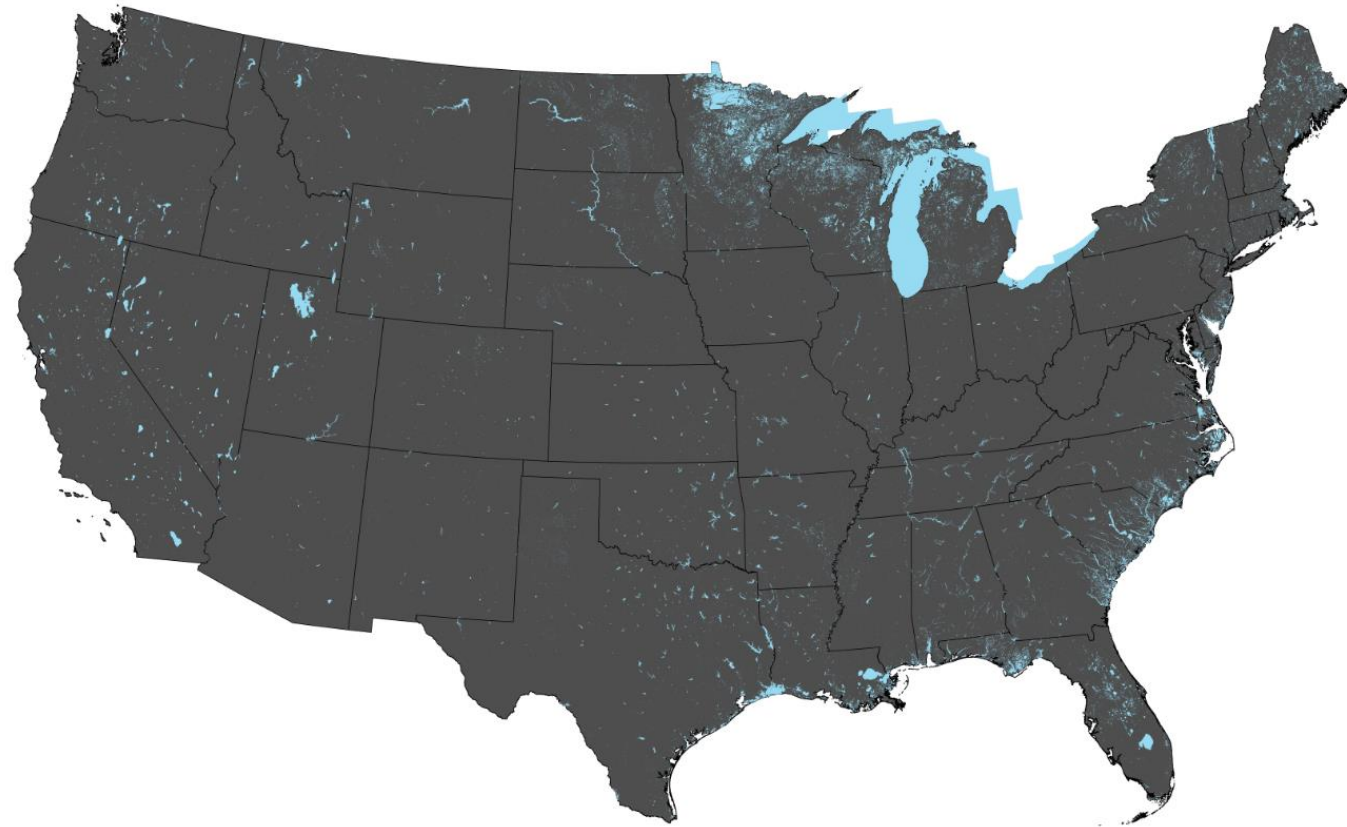
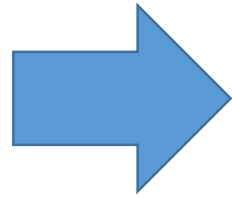
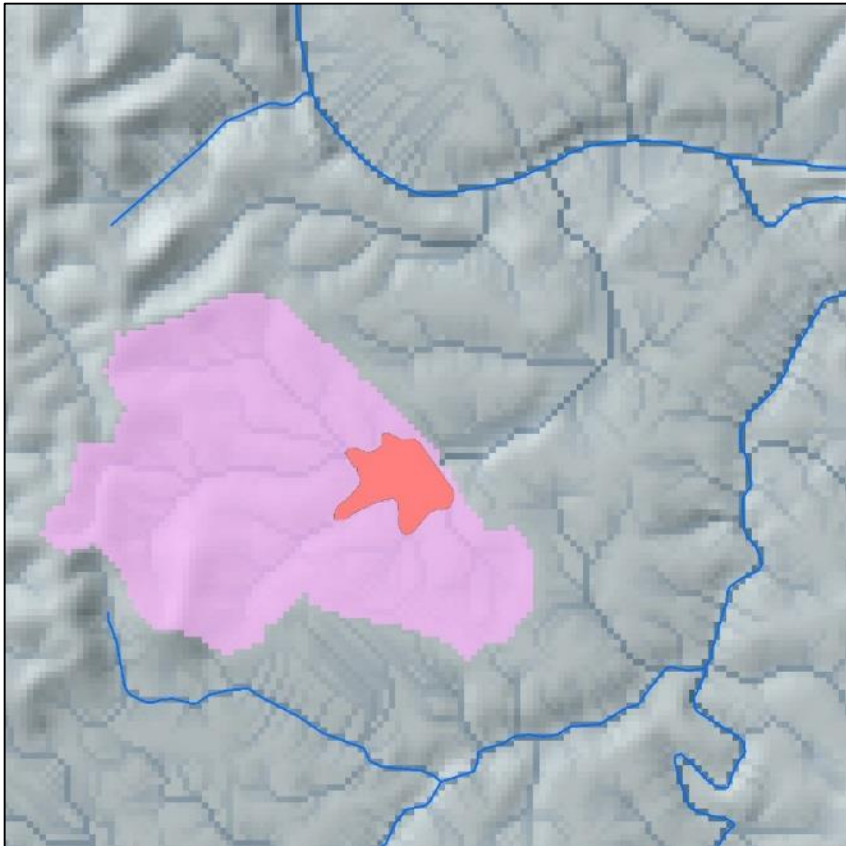
# StreamCat Dataset – *Other applications*

1. Online Watershed Characterization Tool
2. Update/revise EPA's Healthy Watershed Programs Online Tool
3. Refining aquatic species distribution maps in California
4. Modeling and mapping of invasive species in the Pacific Northwest



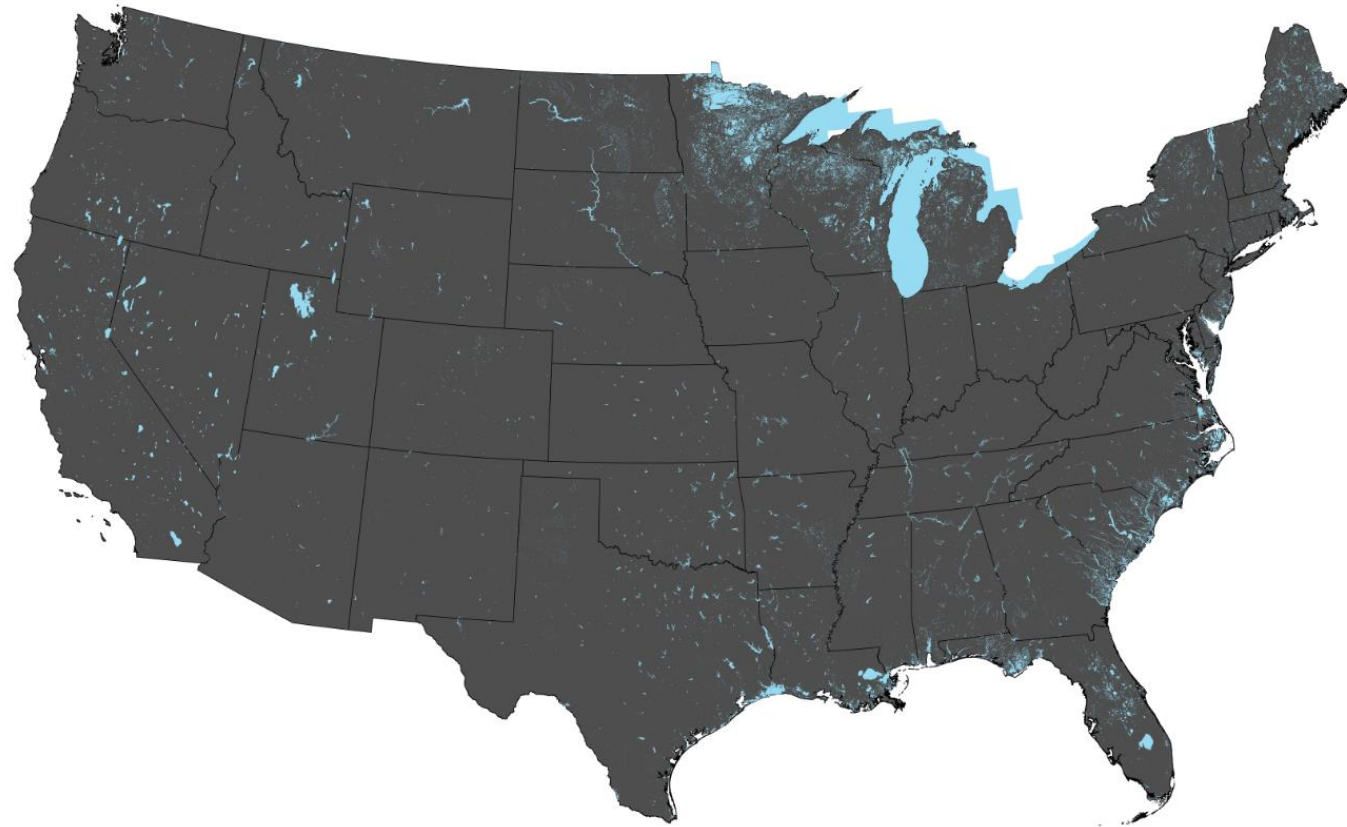
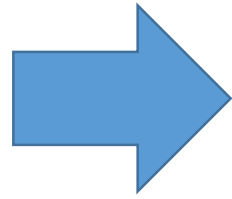
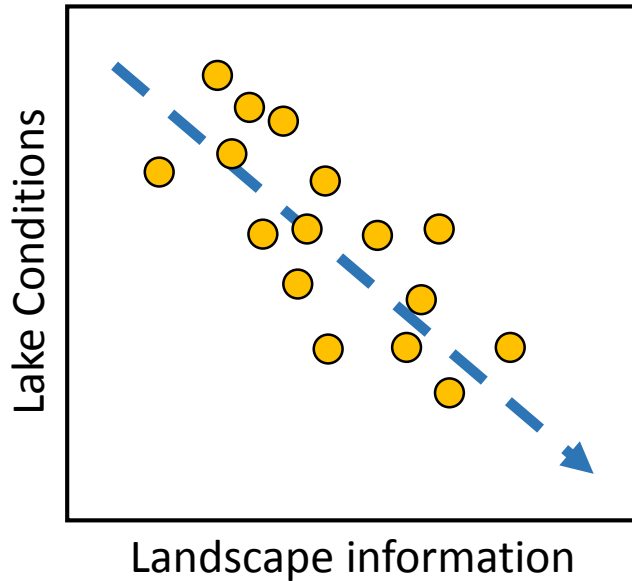
# What's next?

- LakeCat Dataset – Parallels StreamCat for >356,000 lakes



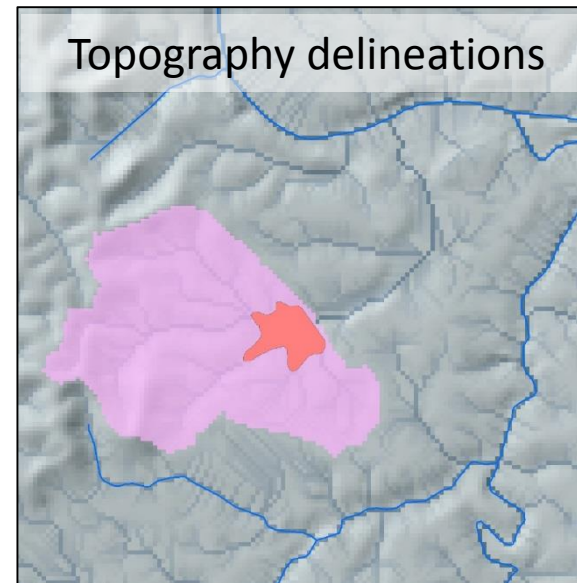
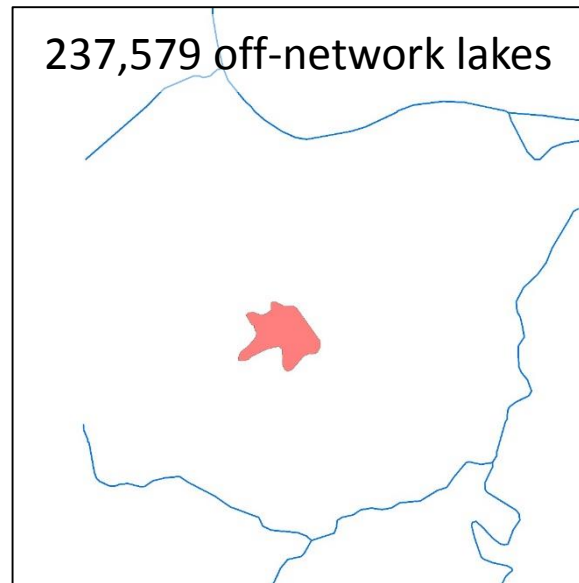
# What's next?

- LakeCat Dataset – Parallels StreamCat for >356,000 lakes



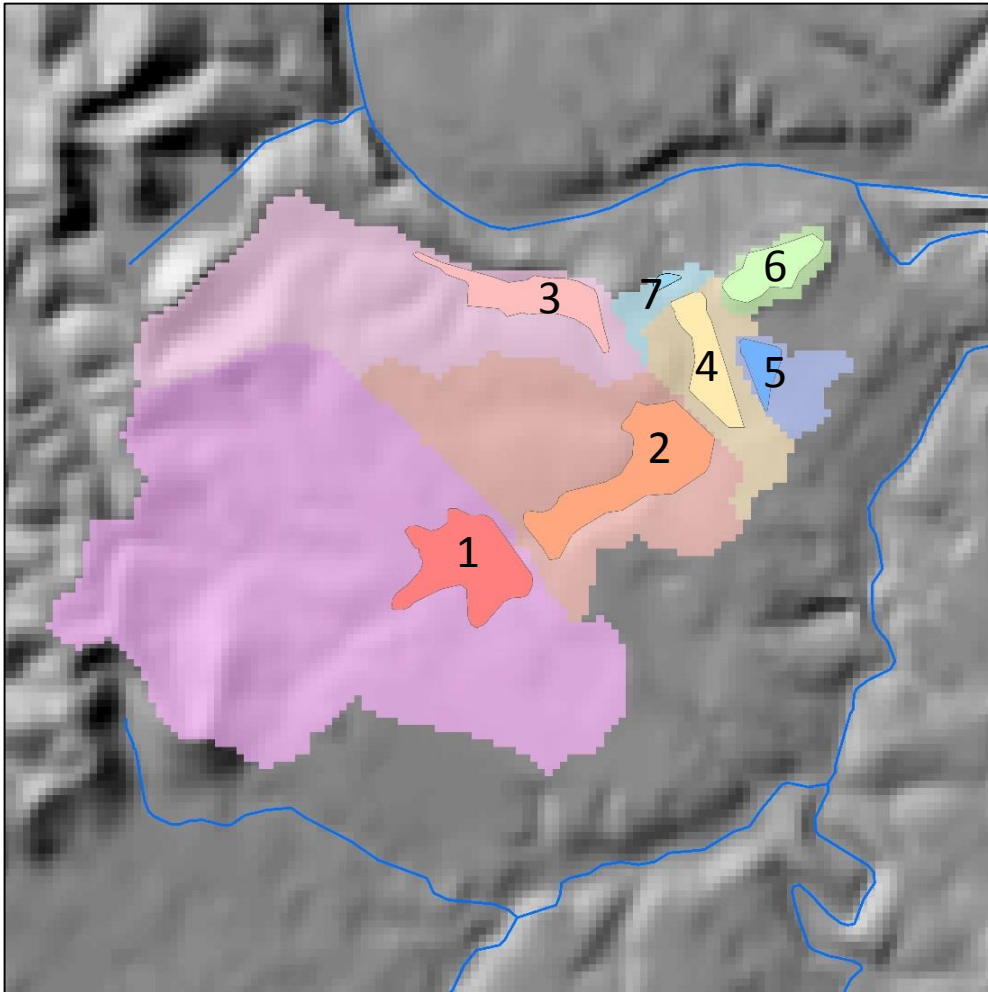
# What's next? –

- *LakeCat Dataset – 356K lakes*
  - *1/3 on NHDPlusV2 stream network – can pull directly from StreamCat*
  - *2/3 off-network lakes needed delineations*



# LakeCat –

- *Nested lakes* - created topological relationships for 237K off-network lakes



FROM	TO
-	1
-	3
1	2
.	.
.	.
3	7
4	7

# Questions?

