Leveraging Big Data and Citizen Science to Understand Sub-Continental Scale Ecological Patterns

Noah R. Lottig
University of Wisconsin Center for Limnology

CSILIMNOLOGY

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Center for Limnology
University of Wisconsin

LTER

Roadmap

1. Approach to addressing sub-continental scale research questions

2. Regional and sub-continental water clarity patterns
   - Regional patterns (citizen data)
   - Data driven approach

3. Potential for future contributions and citizen driven research


nature
The Economist
Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
POPULAR
COMPUTERWORLD
IT Workers With Heart
When Machines Make Decisions
After Jobs: The Enterprise
WHY BIG DATA
IS A BIG DEAL
COVER STORY
BI & ANALYTICS
A new group of data mining technologies promises to forever change the way we sift through our vast stores of data.
Science
11 February 2011
scientific
information
climate
science
research
analysis
visualization
researchers
new
many
access
human
knowledge
data
AAAS
significance
statistics making sense
BIG DATA
SPECIAL ISSUE
Harvard Business Review
OCTOBER 2012
The True Measures Of Success
10 Rules for Managing Global Innovation
What Ever Happened To Accountability?
GETTING CONTROL OF BIG DATA
How vast new streams of information are changing the art of management

CONTEXT NUMBERS

ScienceNews
MAGAZINE OF THE SOCIETY FOR SCIENCE & THE PUBLIC

Why Big Data is

BY

FINANCIAL TIMES

Big

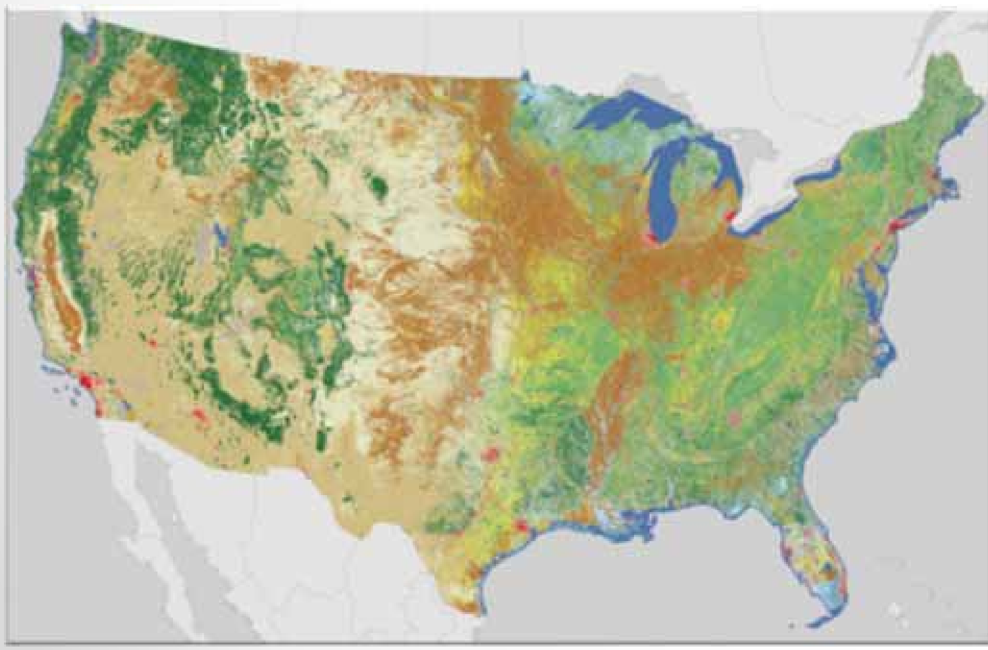
By Tim

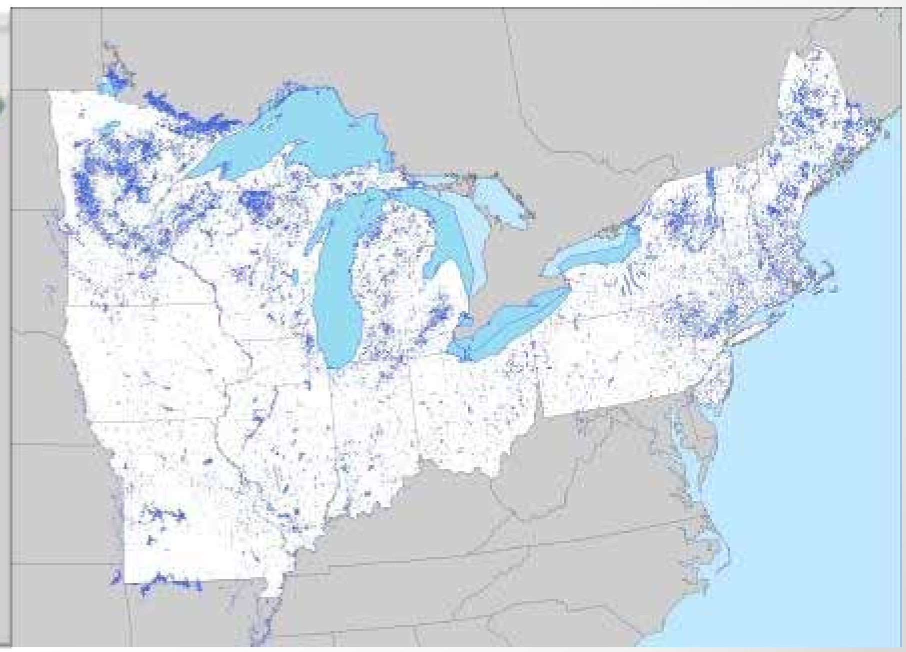
The New York Times | http://nyti.ms/1kgErs2

Eight (No, Nine!) Problems With Big Data

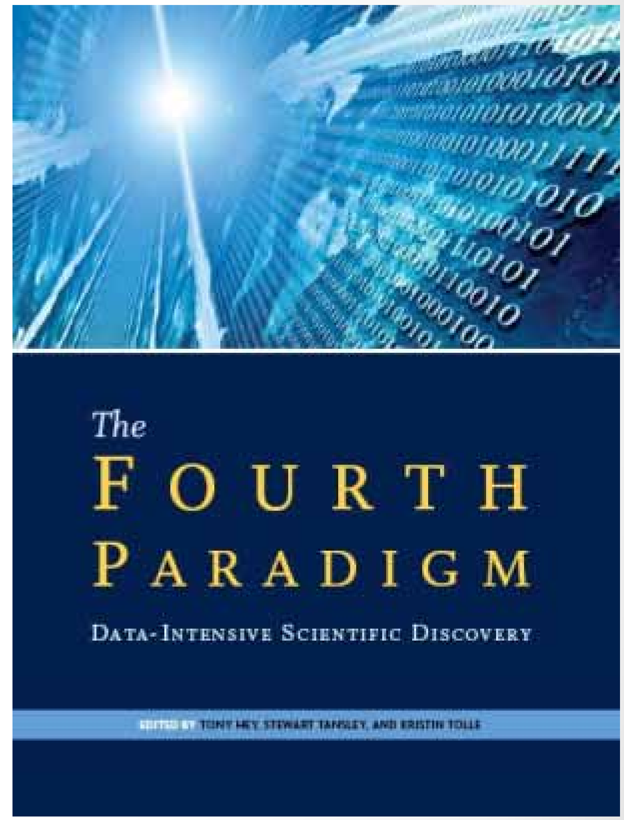
By GARY MARCUS and ERNEST DAVIS APRIL 6, 2014

So why use Big Data?





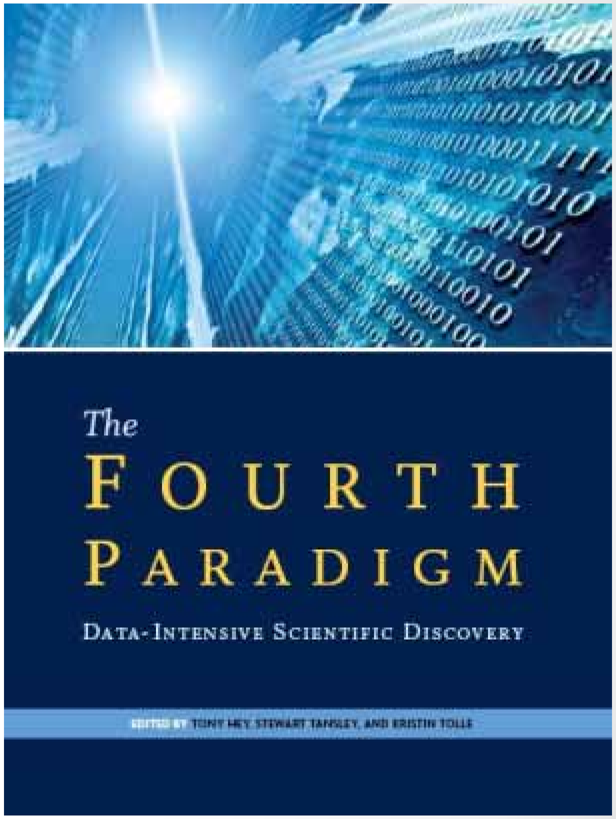
“Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets. The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud-computing technologies.”



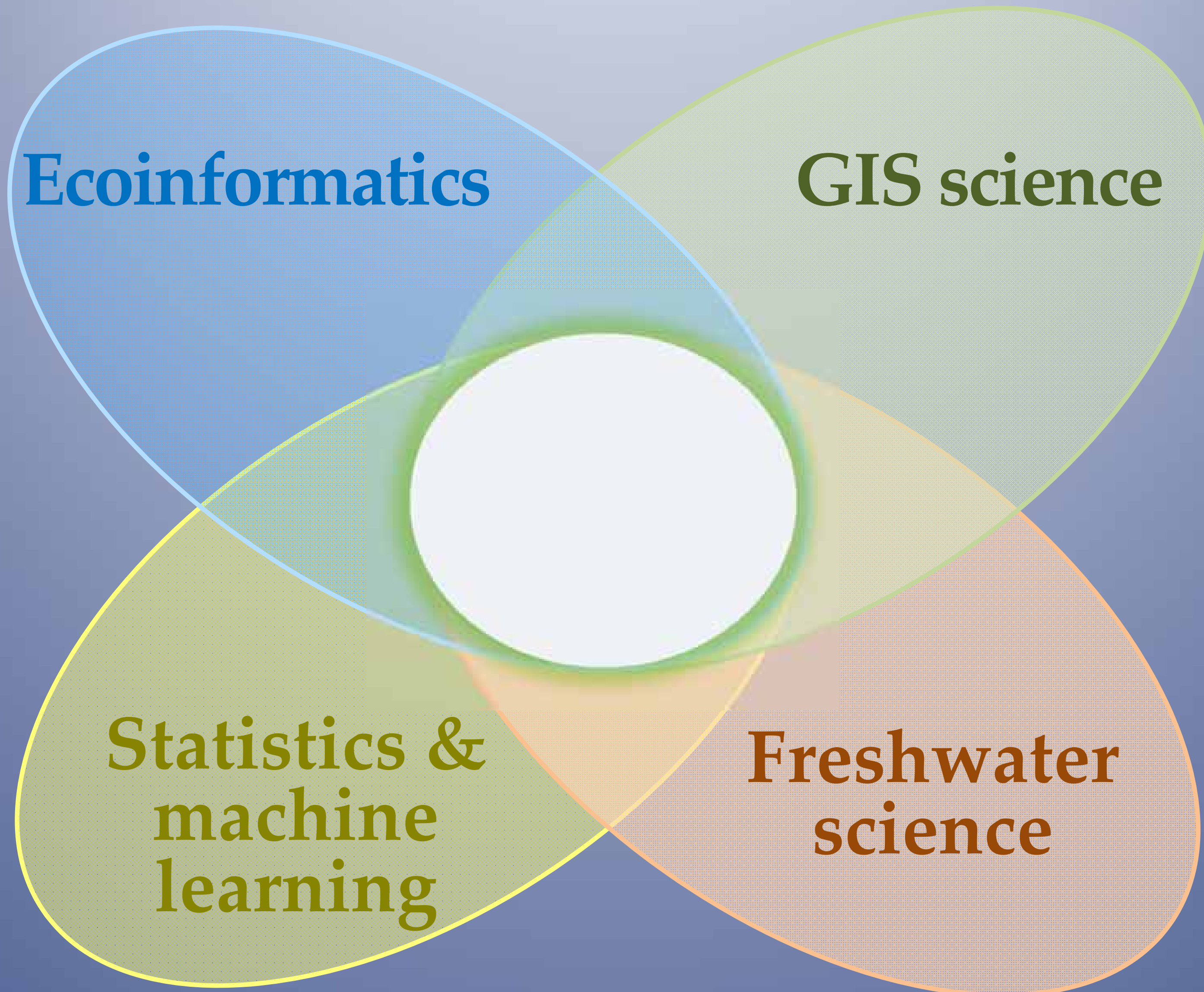
Data intensive approaches

~~**Big data approaches**~~ can be used to study freshwaters at broad scales if 3 conditions are met:

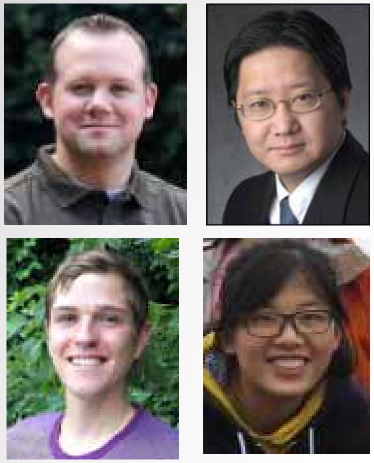
1) Interdisciplinary team science approach
2) Conceptual foundation
3) Robust quantitative methods



Approach: *Interdisciplinary team science*



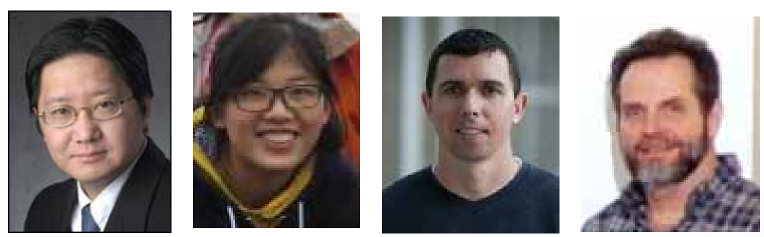
Ecoinformatics



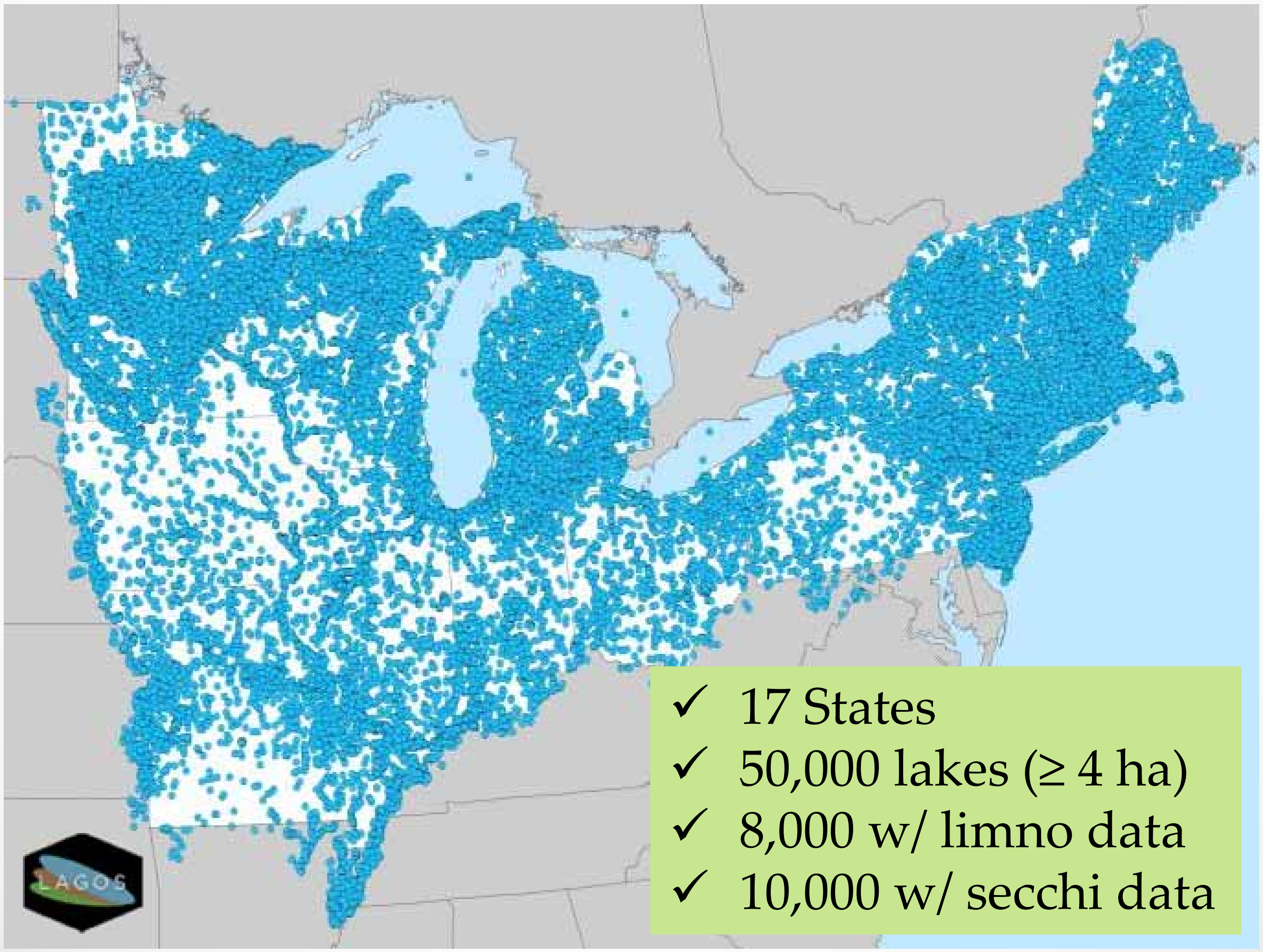
FW science

GIS science

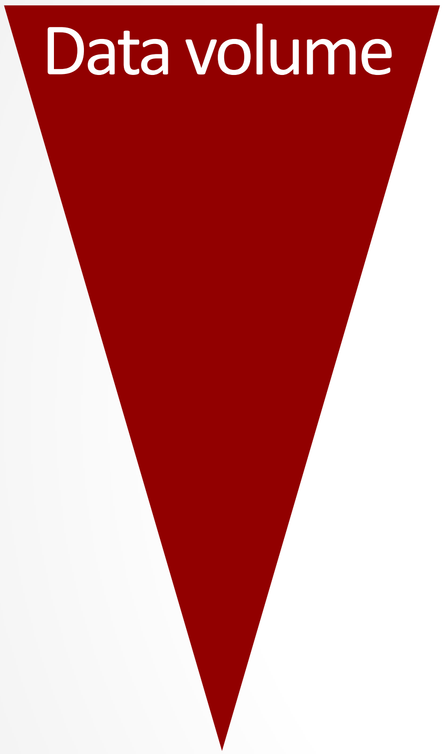
Statistics & machine learning



LAke GeOSpatial temporal database



Limnological data sources *(~90 total)*



State & tribal agencies

Citizen monitoring programs

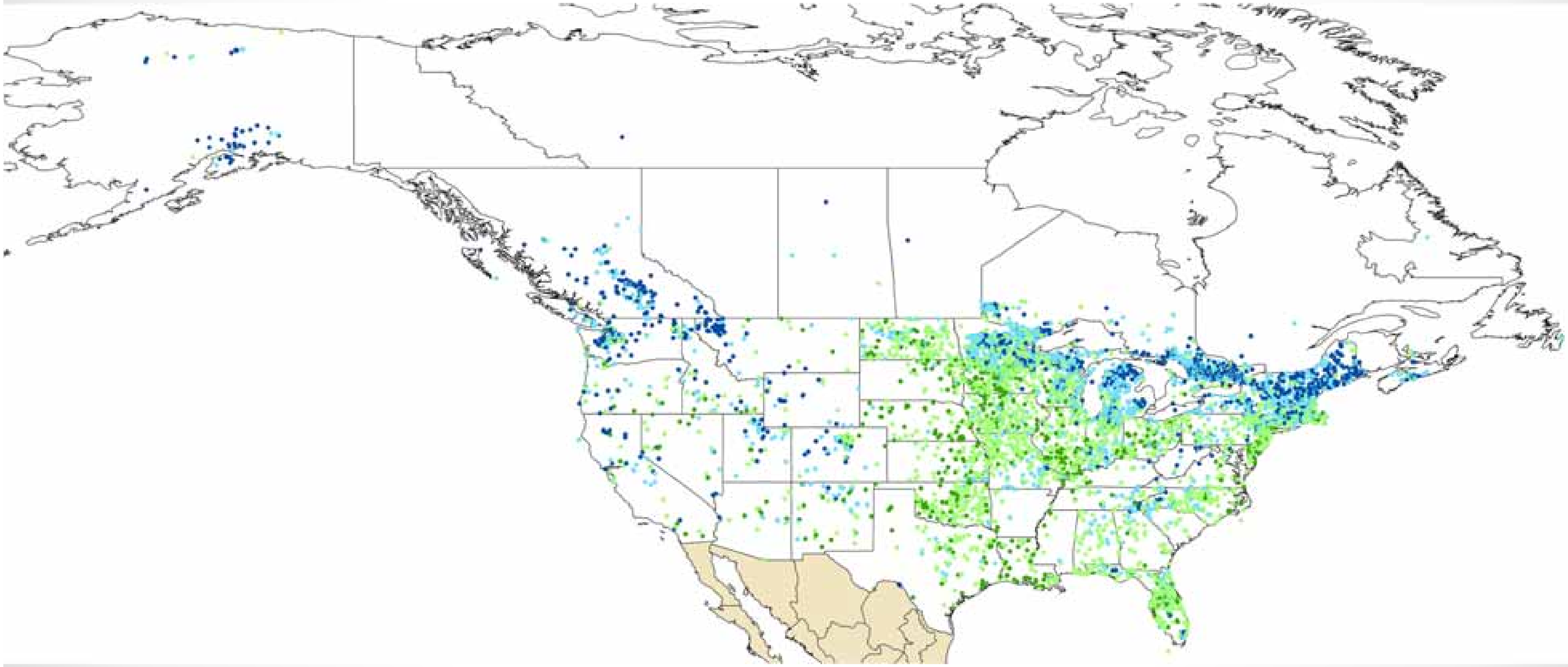
Federal agencies

University researchers

What are the long-term patterns of water clarity

Pietro Angelo Secchi

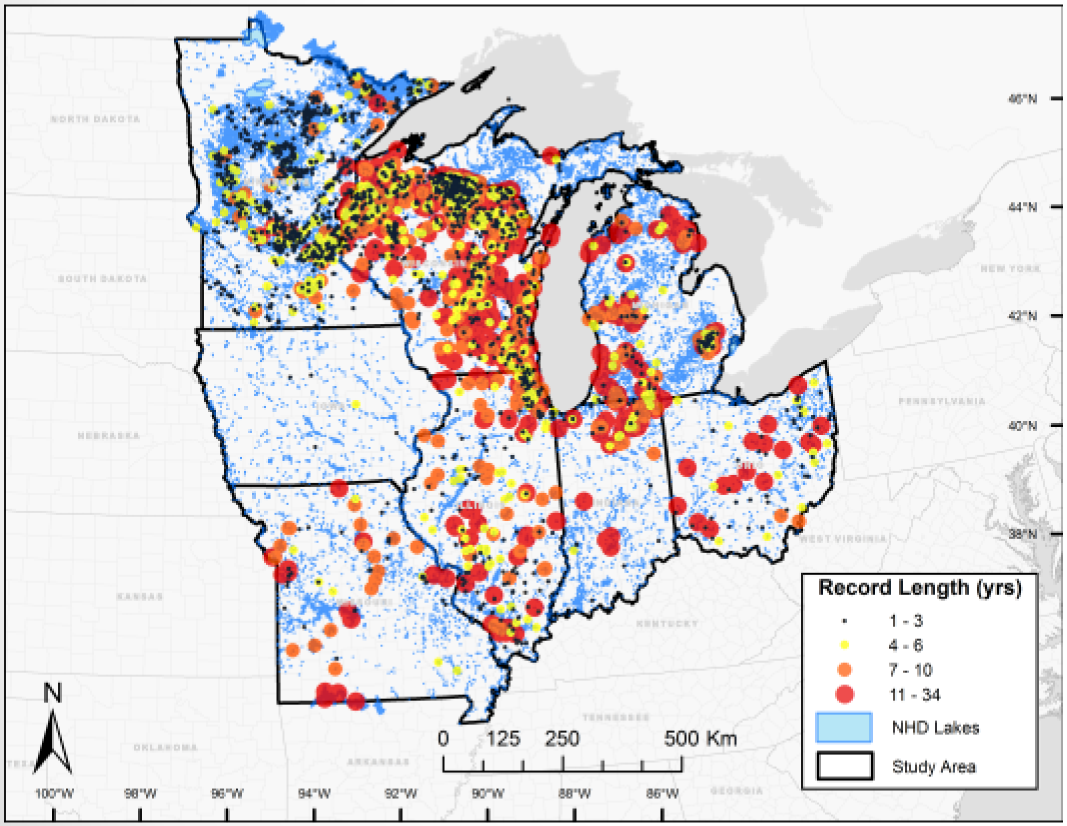
Citizen Science!



The Secchi Dip-In Project

13k lakes/41k secchi observations

Long-term lake clarity trends



- Citizen Data!
- 8 States
- 239,741 observations
- 21,020 annual values
- 3,251 different lakes

Lottig et al. PLoS ONE (2014)

Research Questions

- What are the long-term patterns in water clarity across a broad spatial extent
- How does spatial location, size, and monitoring time period influence long-term patterns

Bayesian Hierarchical Linear Modeling

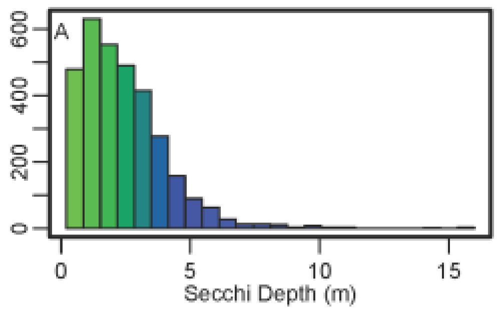
$$y_i \sim N\left(\alpha_{j(i)} + \beta_{j(i)} x_i, \sigma_j^2\right), \text{for } i = 1, \ldots n$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right), \text{for } j = 1, \ldots \text{J}$$

$$log(\sigma_j) \sim N\left(\mu_\sigma, \omega_\sigma^2\right), \text{for } j = 1, \ldots \text{J}$$

Lottig et al. PLoS ONE (2014)

Long-term lake clarity trends

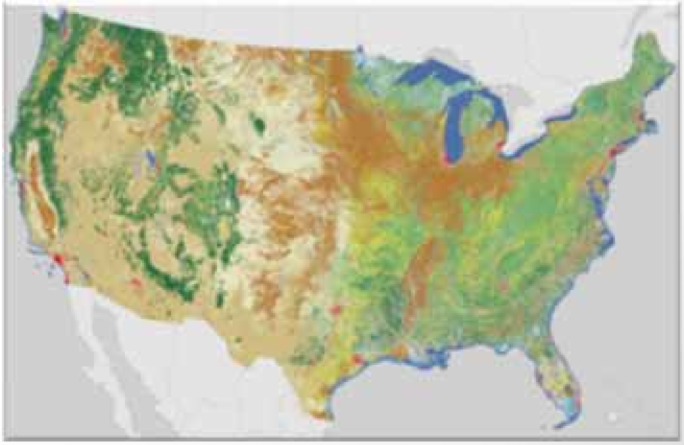


- Average Secchi (2.4m)
- Low inter-annual variability (30%)
- 1% yr^{-1} increase in Secchi depth
- Few lakes with significant long-term trends
  - 7% increasing
  - 4% decreasing

Lottig et al. PLoS ONE (2014)

Long-term lake clarity trends – Spatial Location

1. Latitudinal gradients in Secchi depth
2. Latitudinal gradients in long-term trends
3. Latitudinal gradients in inter-annual variability



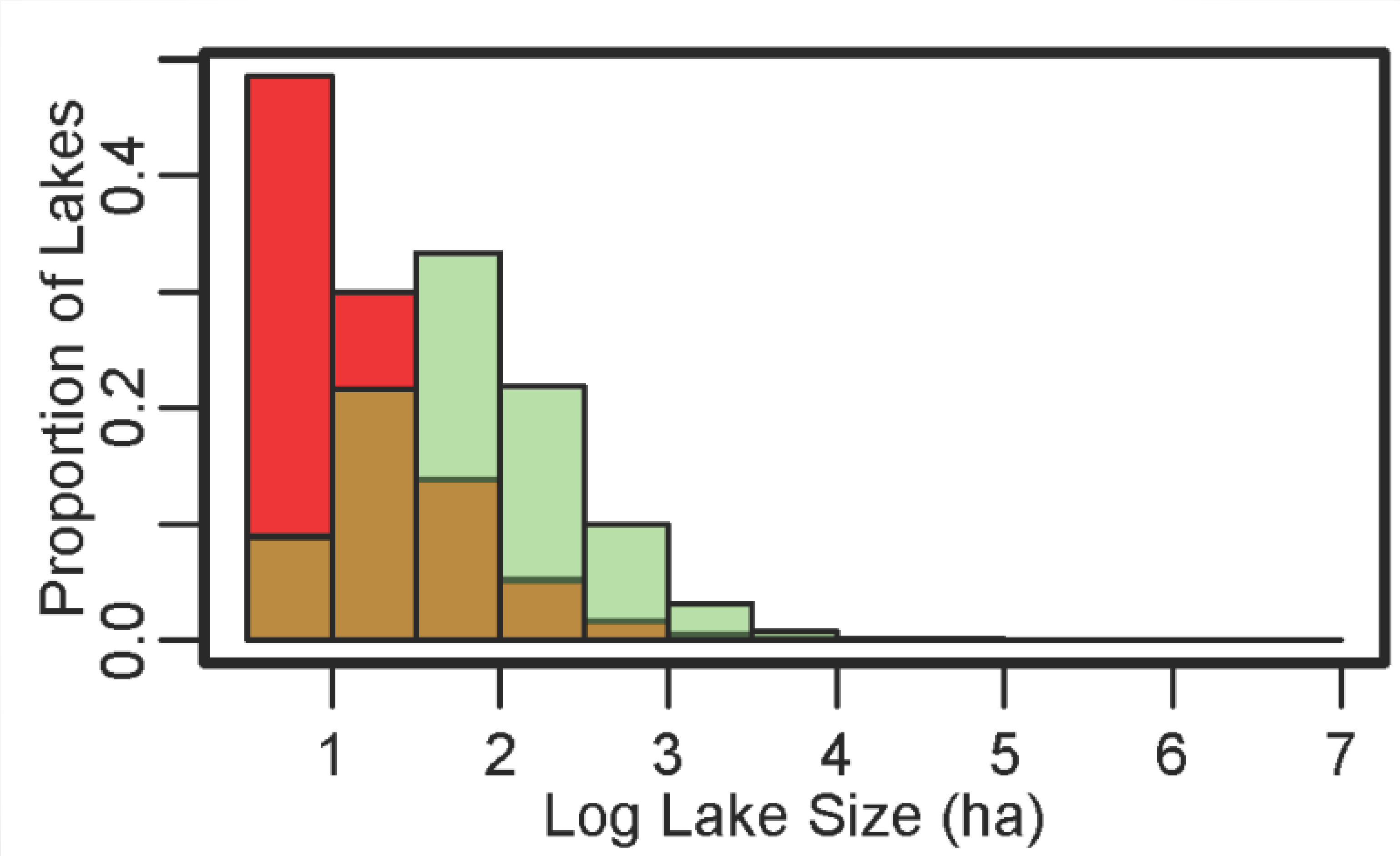
Lottig et al. PLoS ONE (2014)

Long-term lake clarity trends – Time Period

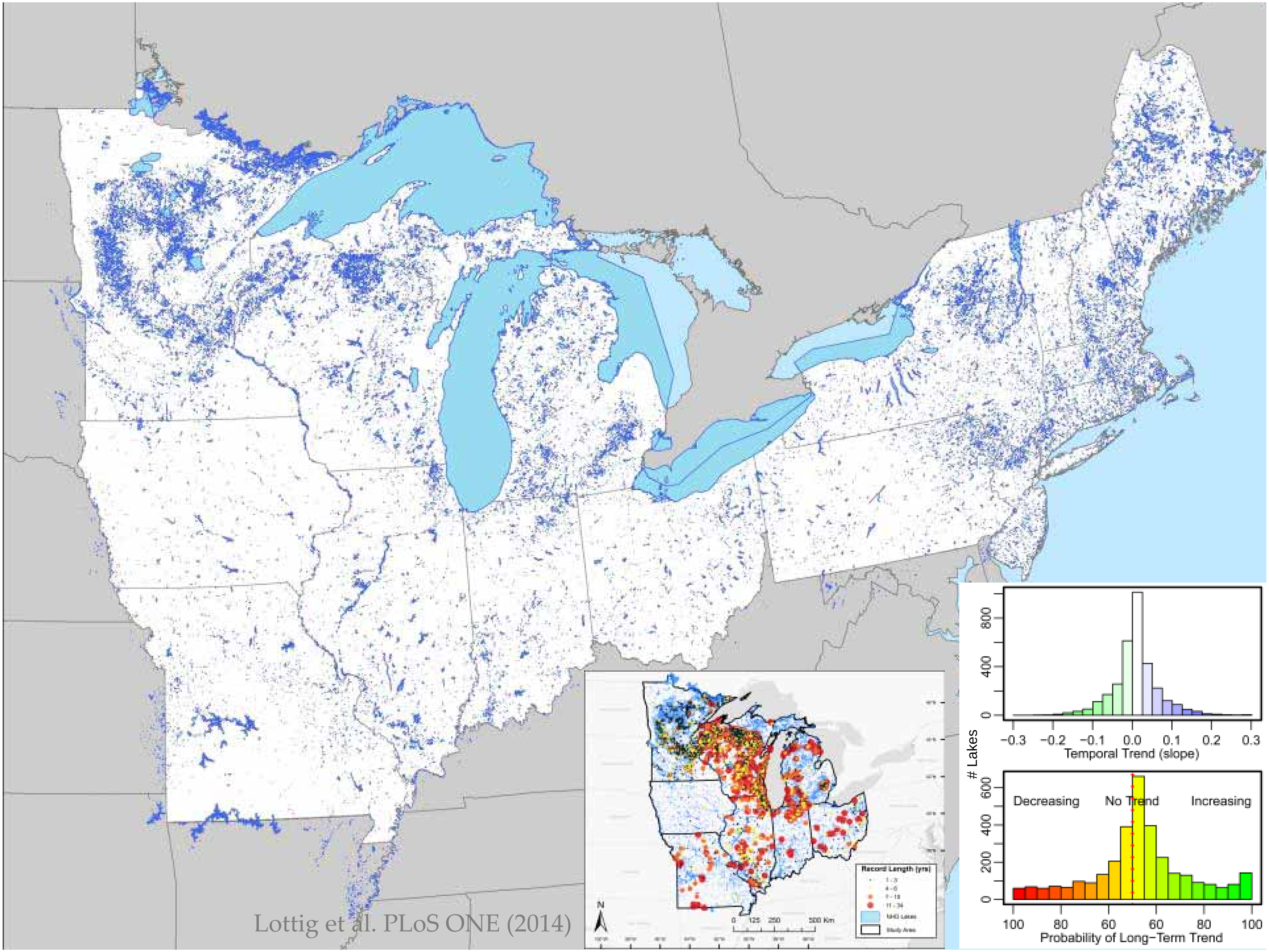
1. Average Secchi hasn't changed
2. Recent lakes have declining trends
3. Significant reductions in inter-annual variability

Lottig et al. PLoS ONE (2014)

Biases in the lakes monitored

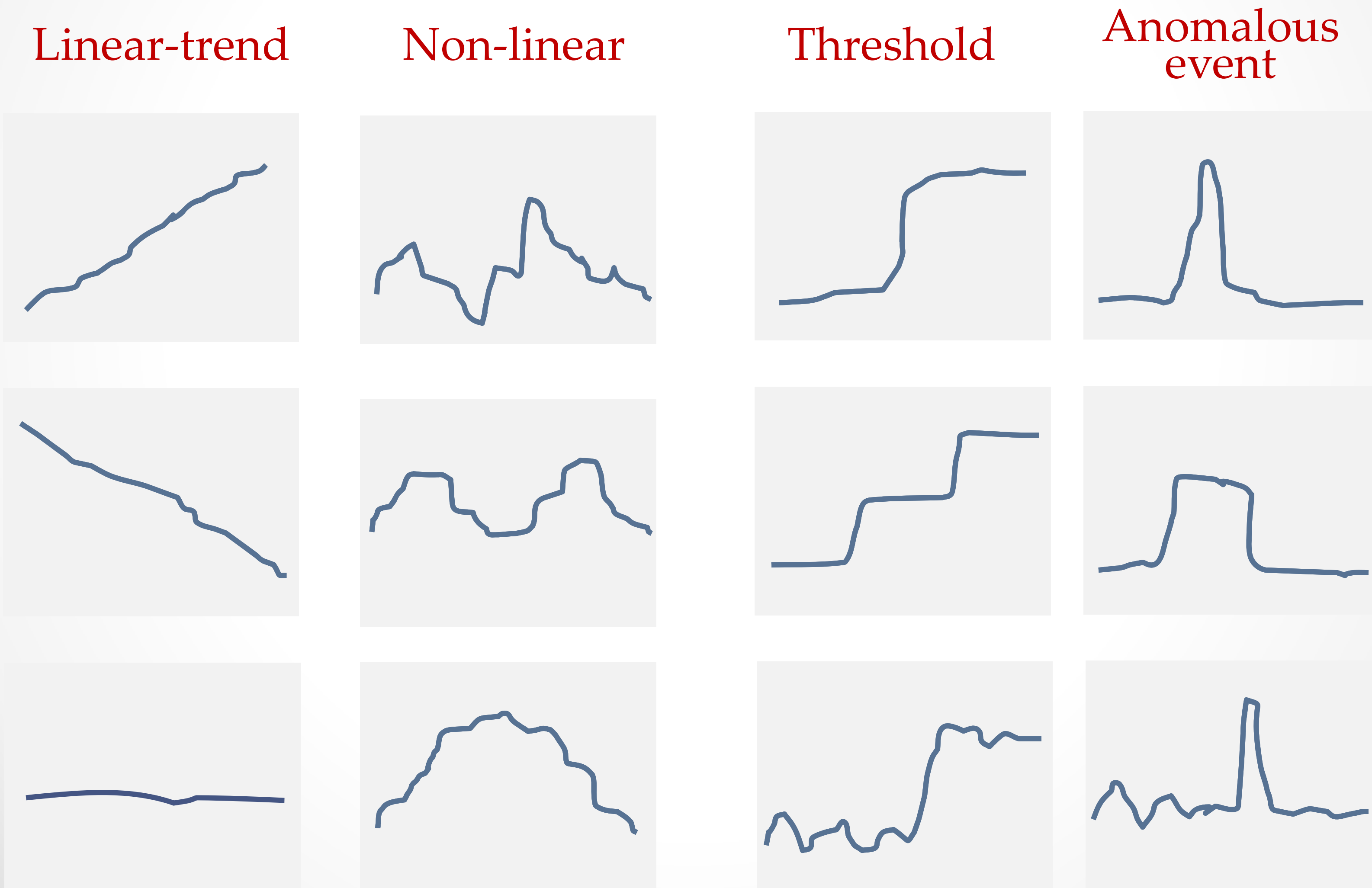


Lottig et al. PLoS ONE (2014)



Lottig et al. PLoS ONE (2014)
Record Length (yrs)
NHD Lakes
Study Area
0 125 250 500 Km
N
Lakes
Temporal Trend (slope)
-0.3 -0.2 -0.1 0.0 0.1 0.2 0.3
0 400 800
Decreasing
No Trend
Increasing
Probability of Long-Term Trend
100 80 60 60 80 100
0 200 400 600

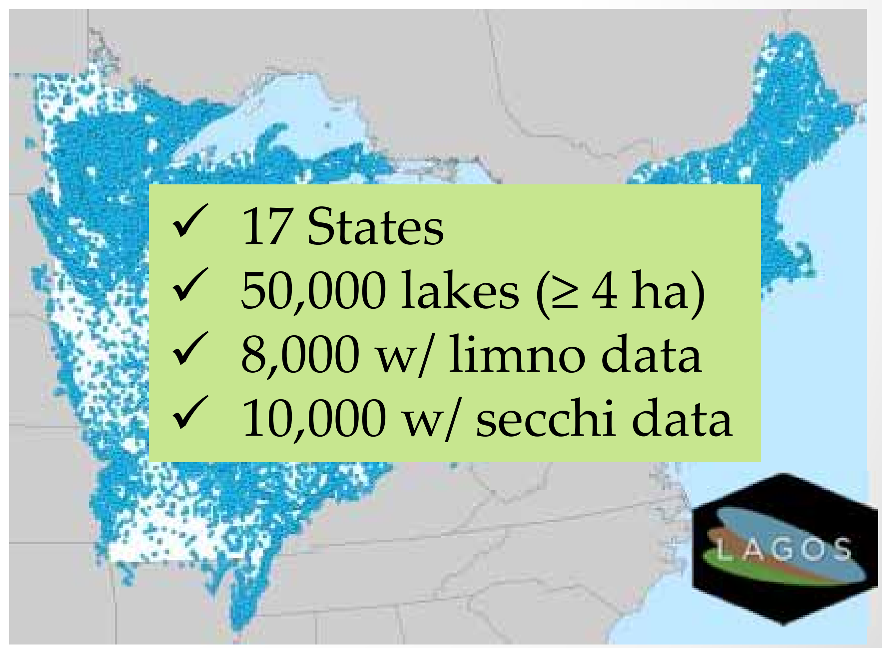
What are patterns of ecological change?



Research Questions:

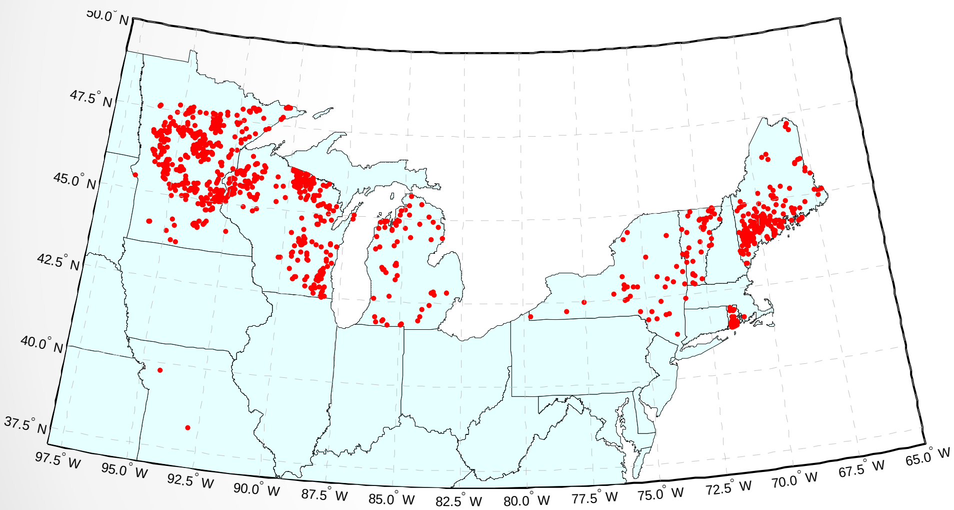
1. What are ecological patterns across large spatial scales?
2. Do clusters of ecosystems exhibit similar patterns?
3. Can we identify factors influencing clusters
   - Spatial
   - Local
   - Regional

LAke **G**e**OS**patial temporal database



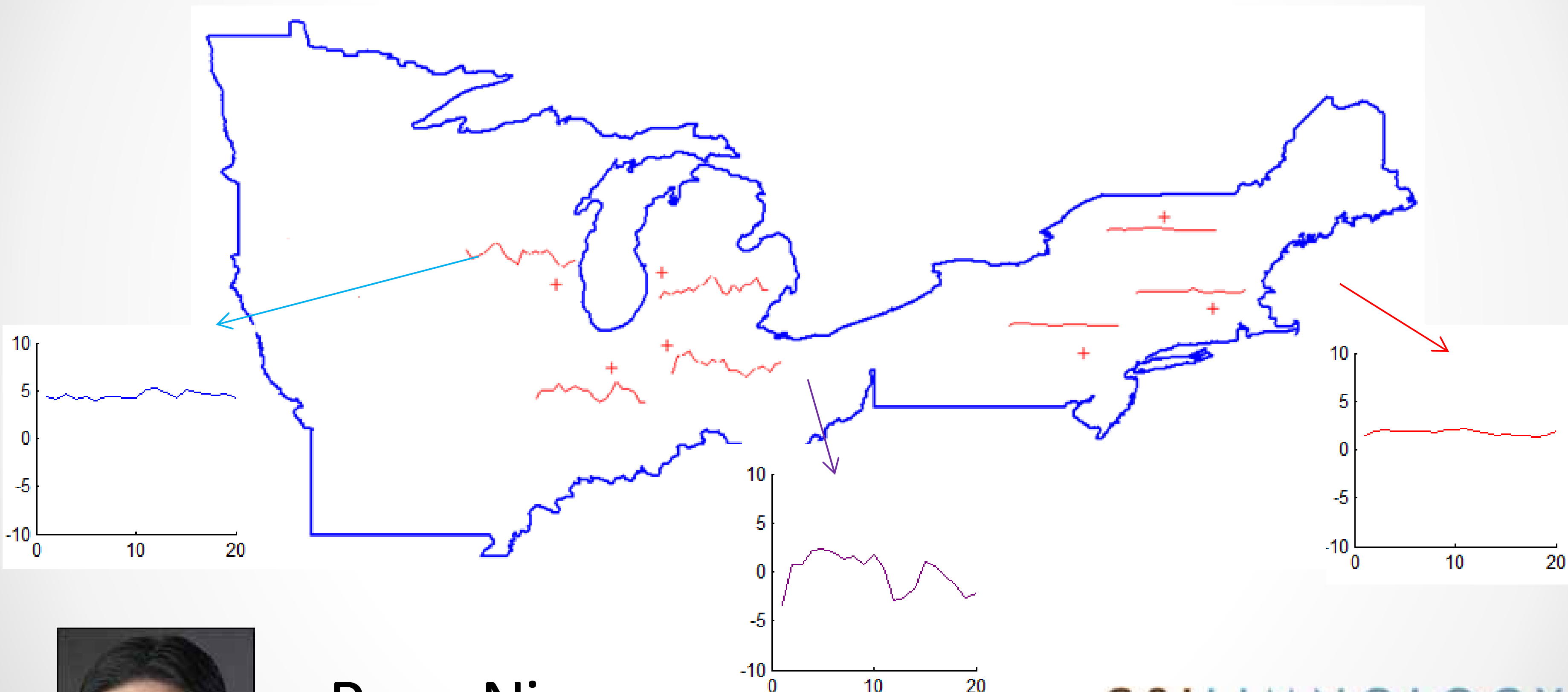
ver 1.032.0

Long-term Secchi data



- 18+ years of data
- 1987 – 2012
  - o ≥ 75% overlap among time-series
- Annual median value
  - o 15 June – 15 September
- 1007 lakes / 8 states

Approach (Q1 & Q2)
Patterns and clusters

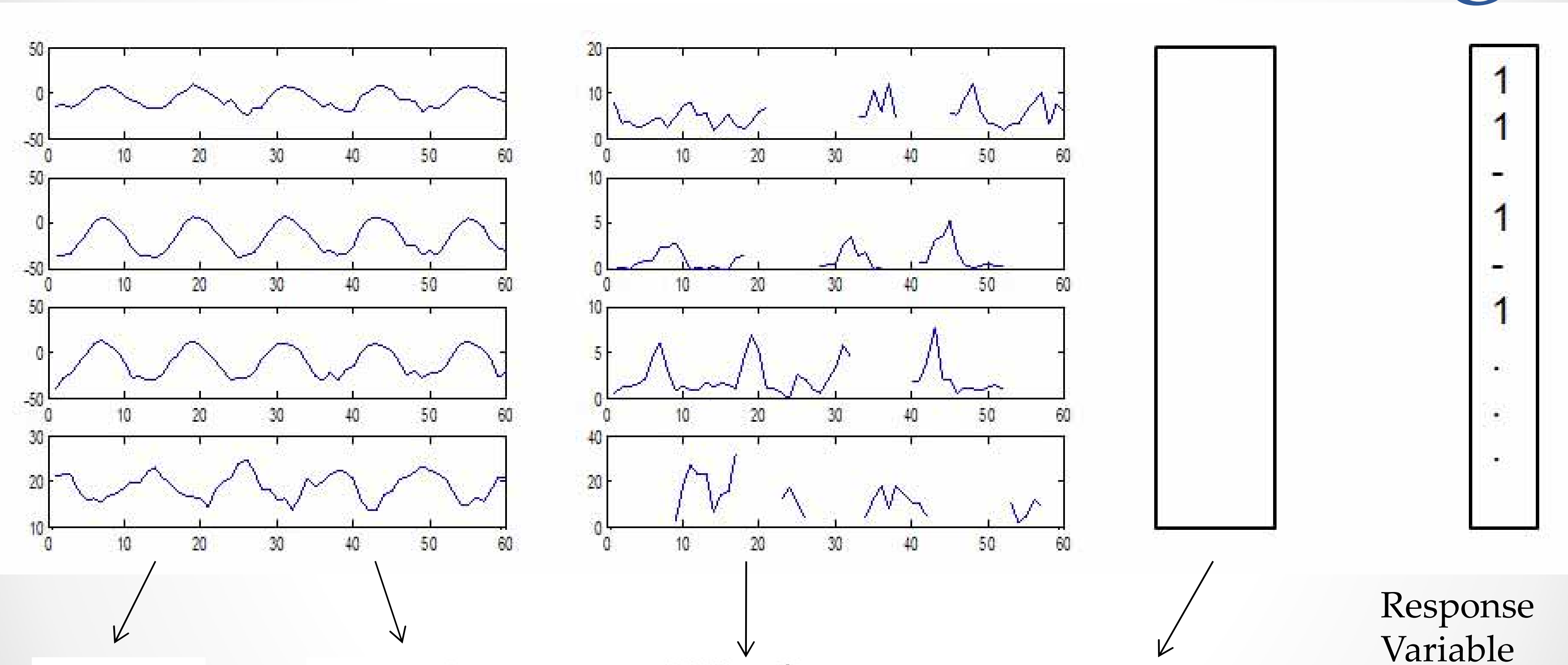


Pang-Ning
Tan

CSILIMNOLOGY

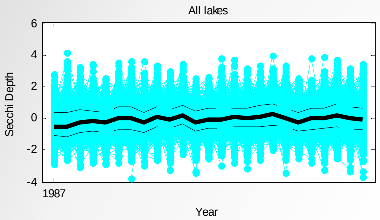
Approach (cont.)

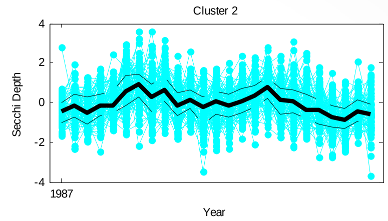
Kernel K-means clustering

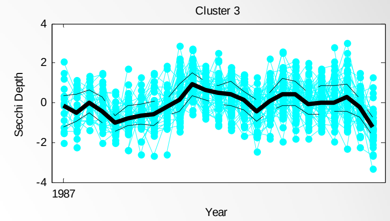


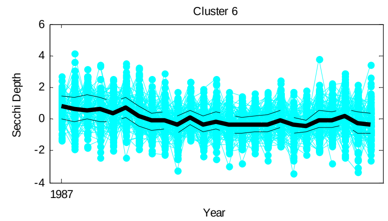
$K_1(x, x')$ $K_4(x, x')$ $K_2(x, x')$ $K_3(x, x')$

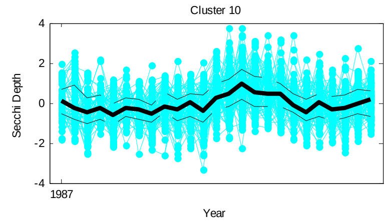
$$K(x, x') = \sum_{i=1}^{4} \beta_i K_i(x, x') \Longrightarrow f(x) = \sum_{i=1}^{N} \alpha_i y_i k(x, x_i)$$

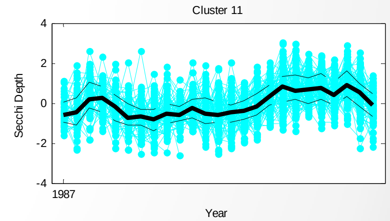

All lakes
Secchi Depth
6
4
2
0
-2
-4
1987
Year

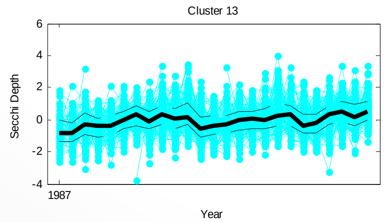




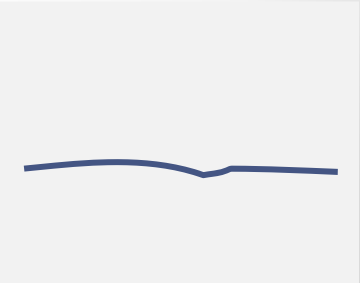




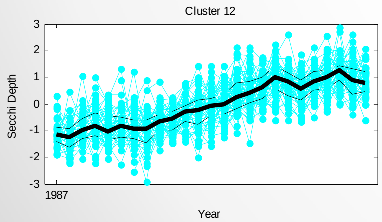


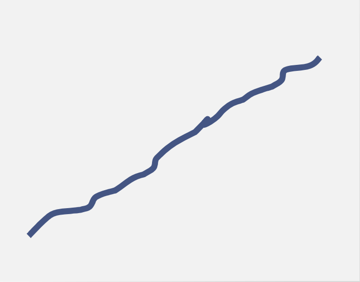


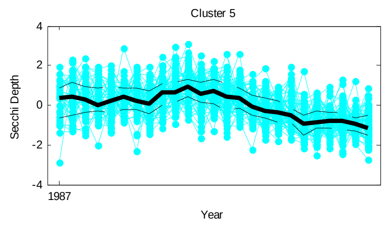
Linear-trend



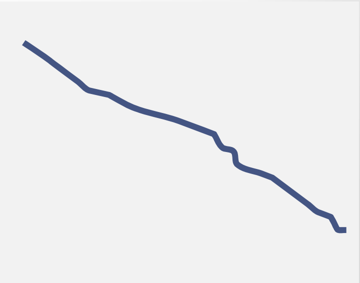
Linear-trend

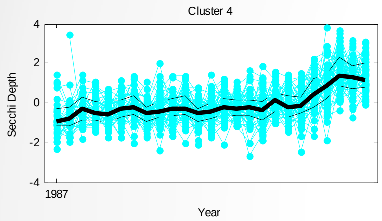




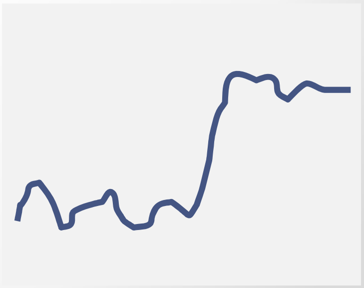


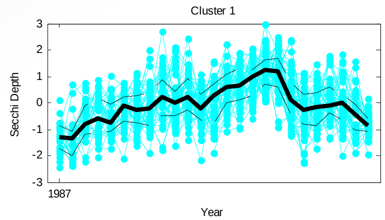
Linear-trend

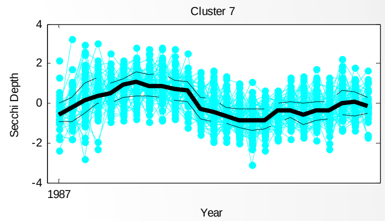


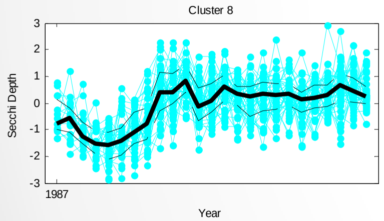


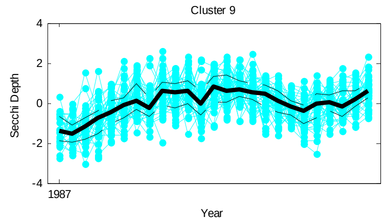
Threshold



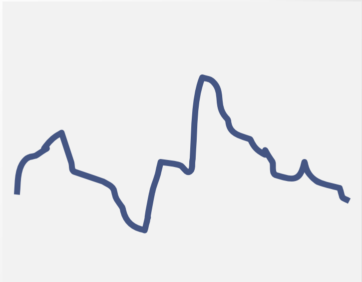

Cluster 1
Secchi Depth
Year
1987

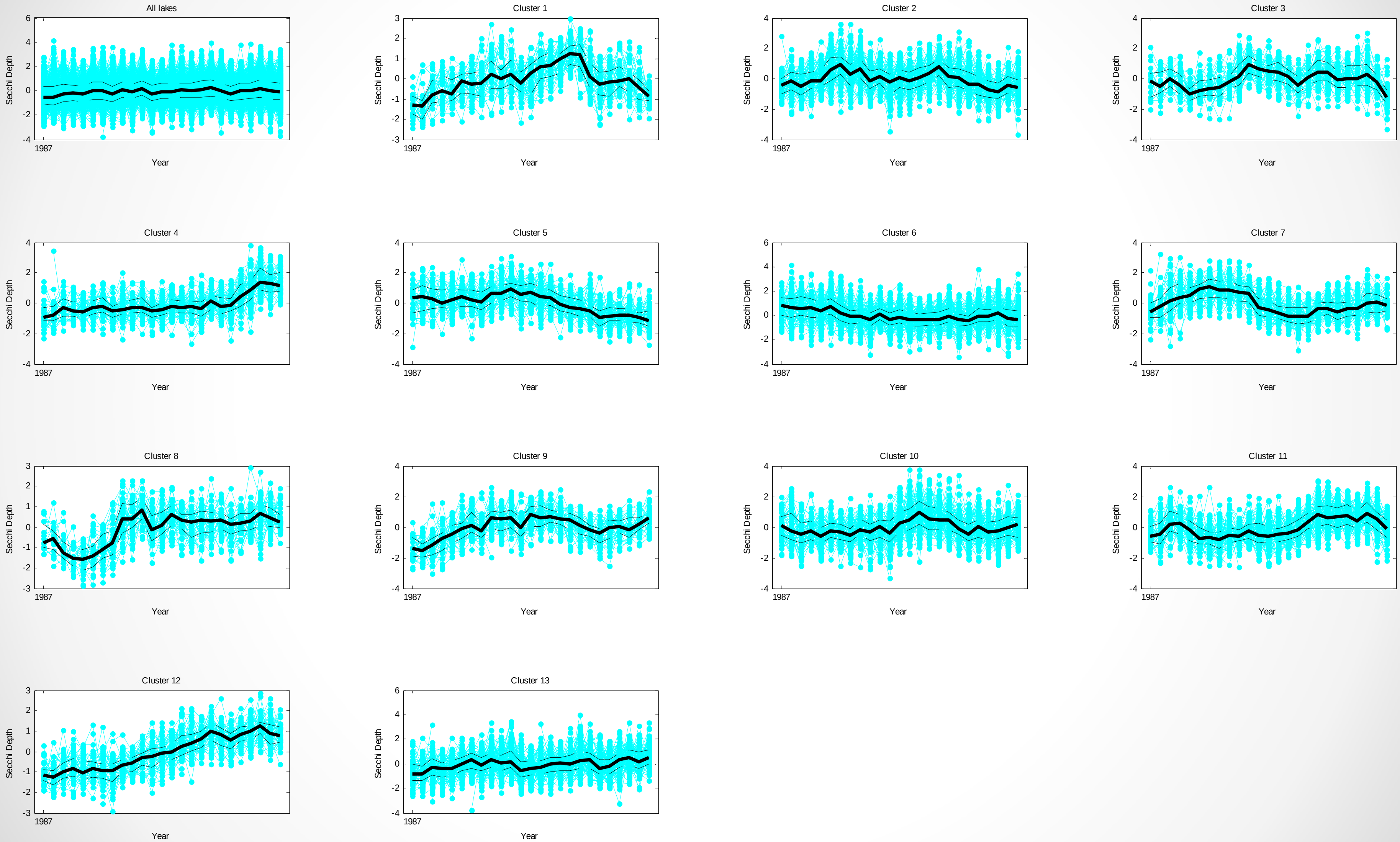

Cluster 7
Secchi Depth
Year
1987


Cluster 8
Secchi Depth
Year
1987


Cluster 9
Secchi Depth
Year
1987

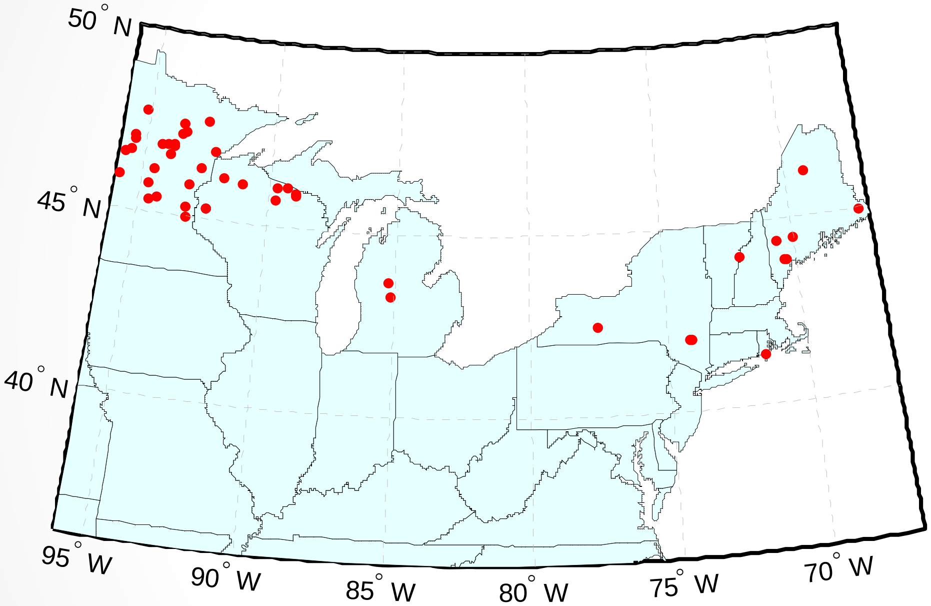
Non-linear



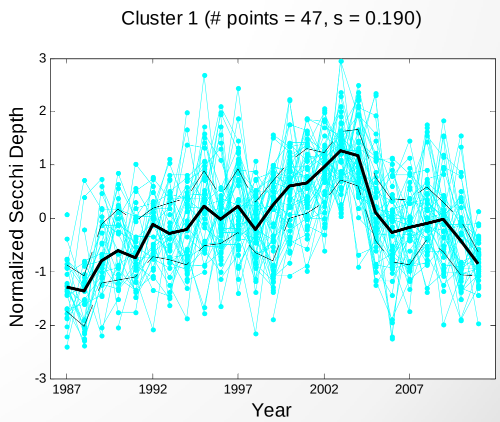


All lakes
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
Cluster 13
Secchi Depth
Year
1987

Cluster 1

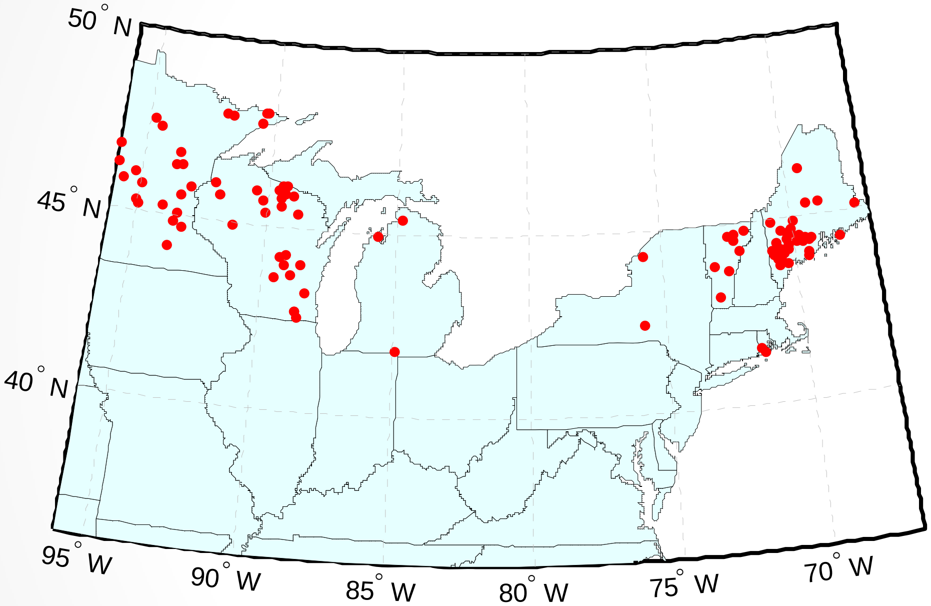


50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W

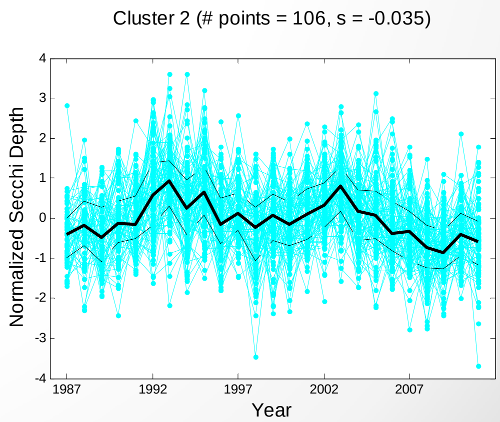


Cluster 1 (# points = 47, s = 0.190)
Normalized Secchi Depth
Year
3
2
1
0
-1
-2
-3
1987
1992
1997
2002
2007

Cluster 2

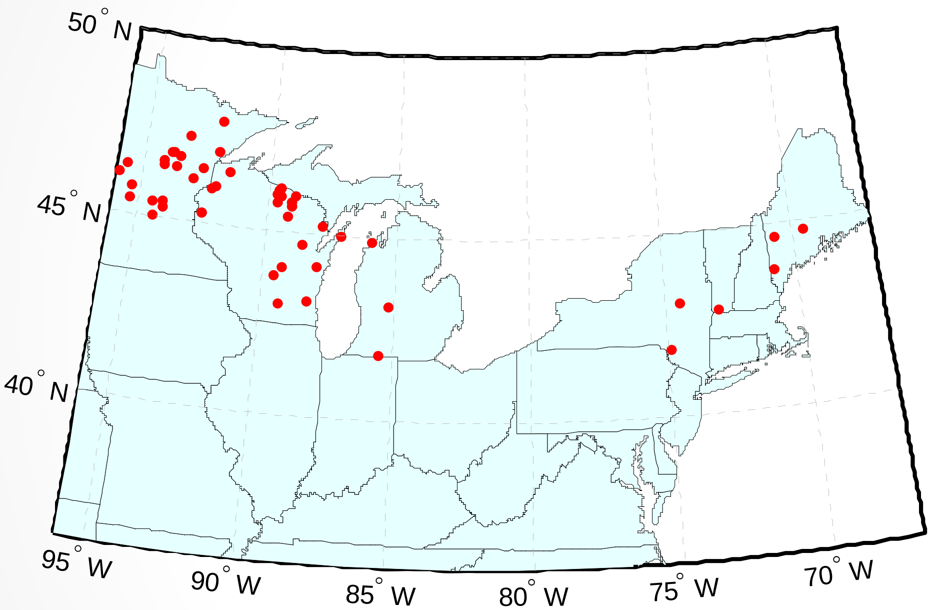


50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W

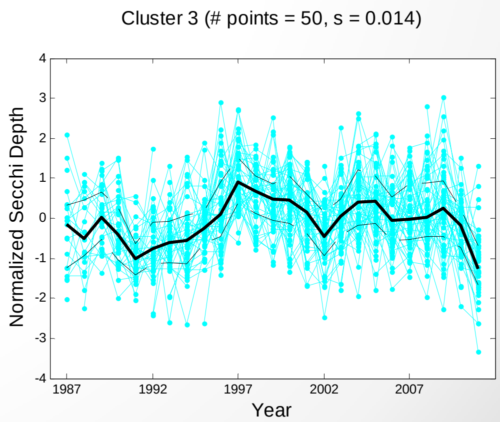


Cluster 2 (# points = 106, s = -0.035)
Normalized Secchi Depth
Year
1987
1992
1997
2002
2007
4
3
2
1
0
-1
-2
-3
-4

Cluster 3

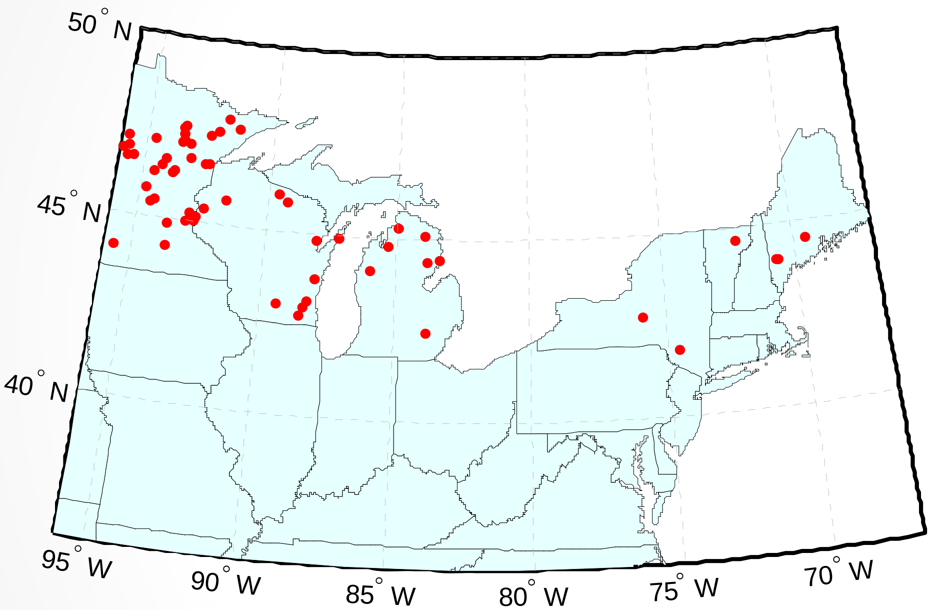


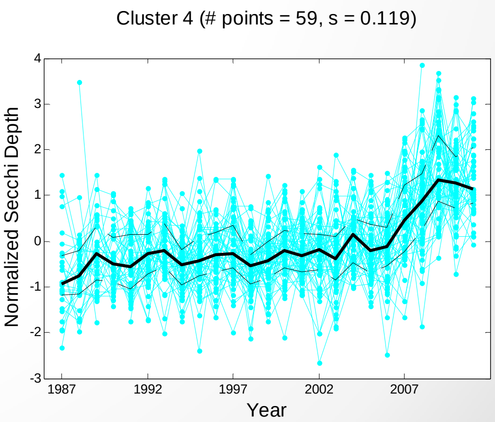
50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W



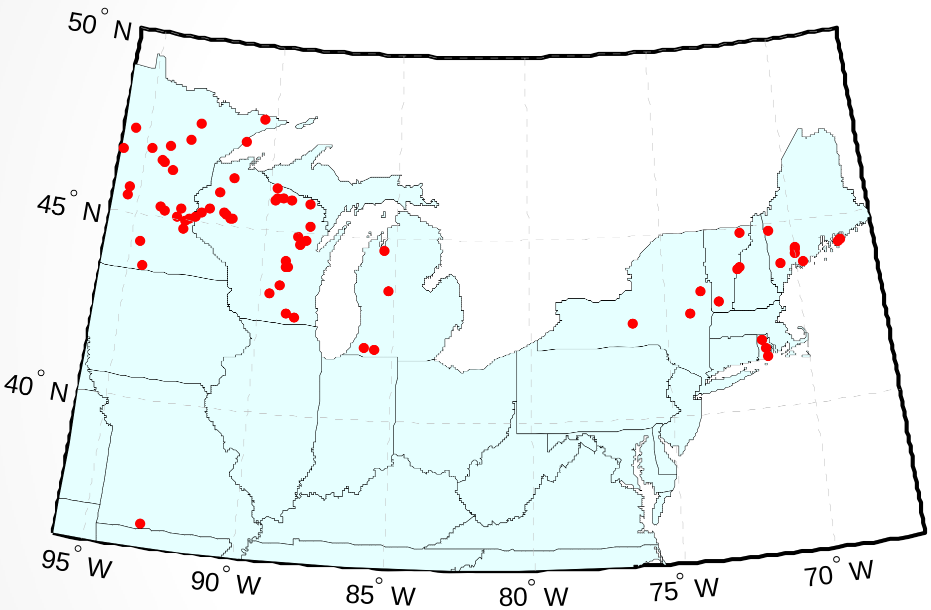
Cluster 3 (# points = 50, s = 0.014)
Normalized Secchi Depth
Year
1987
1992
1997
2002
2007

Cluster 4

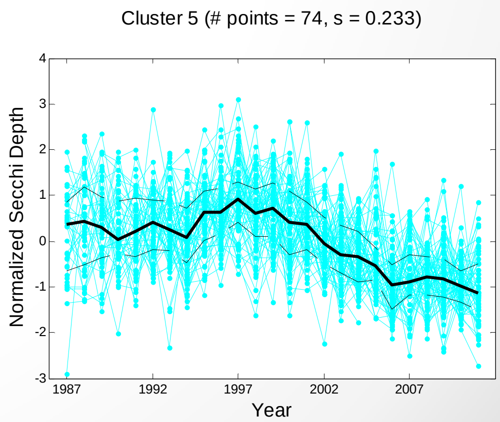

50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W


Cluster 4 (# points = 59, s = 0.119)
Normalized Secchi Depth
Year
1987
1992
1997
2002
2007
4
3
2
1
0
-1
-2
-3

Cluster 5

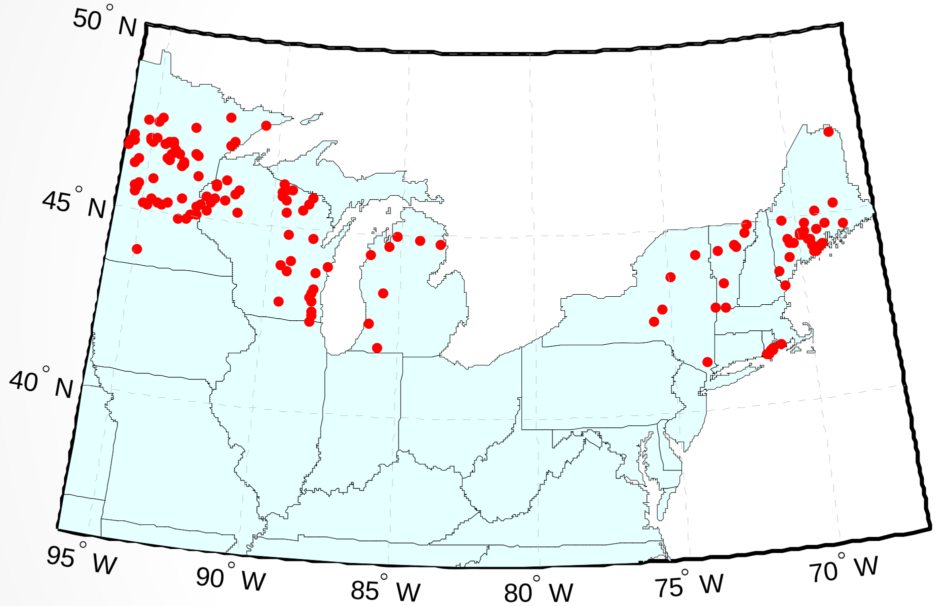


50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W

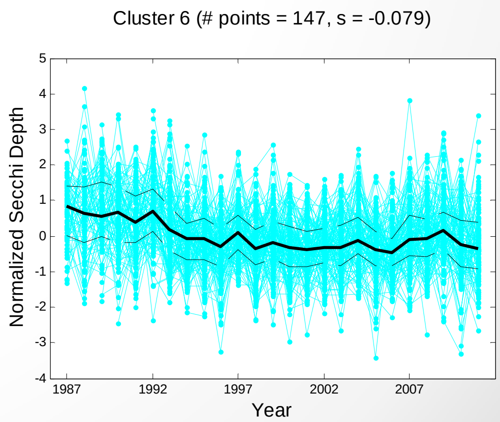


Cluster 5 (# points = 74, s = 0.233)
Normalized Secchi Depth
Year
1987
1992
1997
2002
2007
4
3
2
1
0
-1
-2
-3

Cluster 6

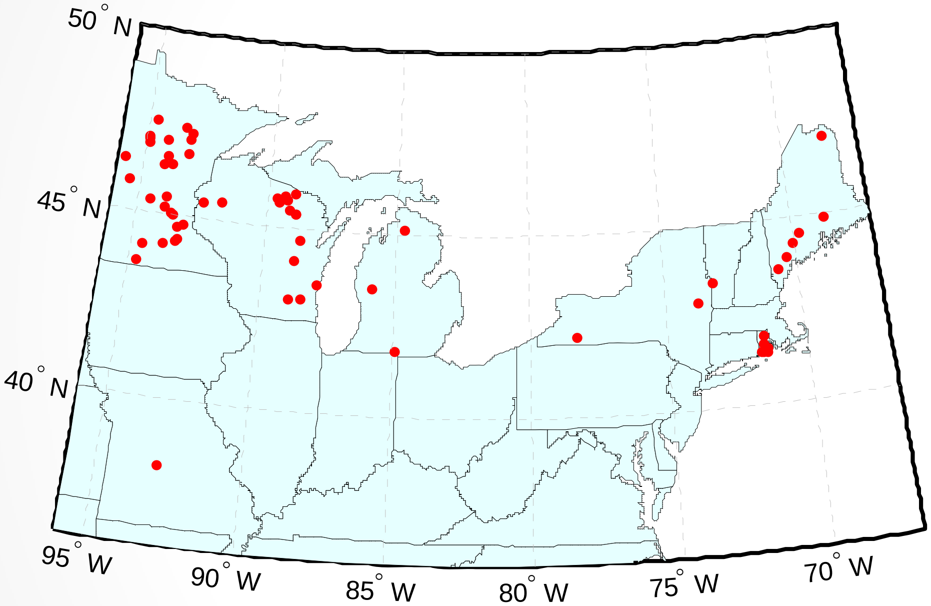


50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W

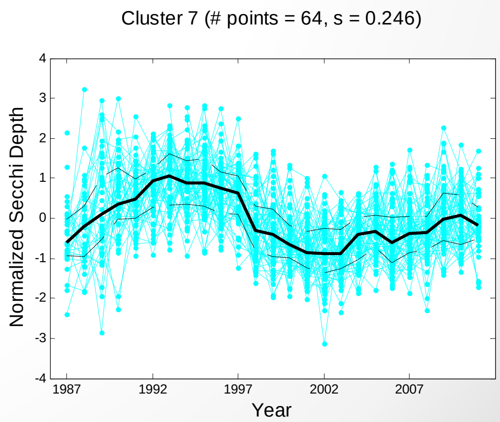


Cluster 6 (# points = 147, s = -0.079)
Normalized Secchi Depth
Year
1987
1992
1997
2002
2007

Cluster 7

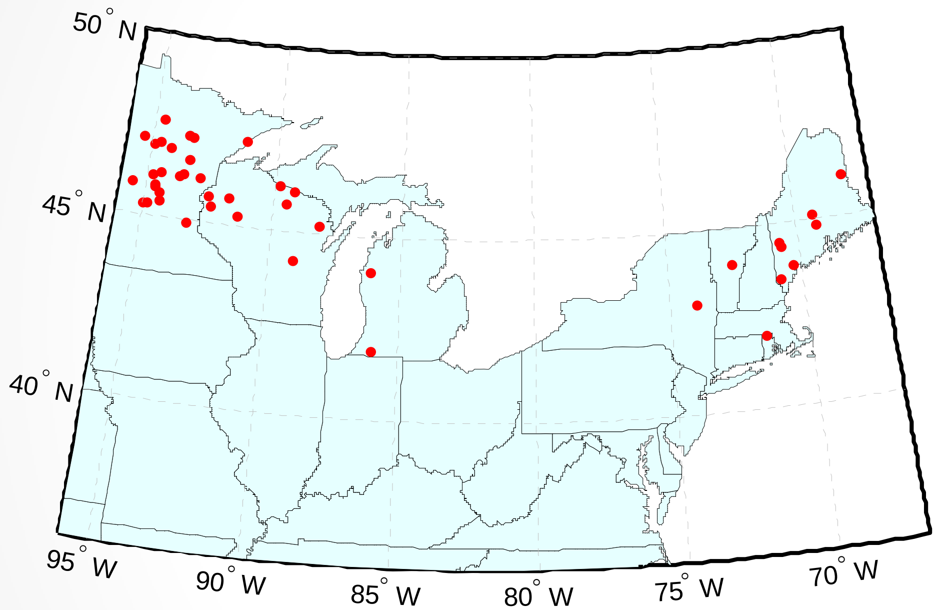


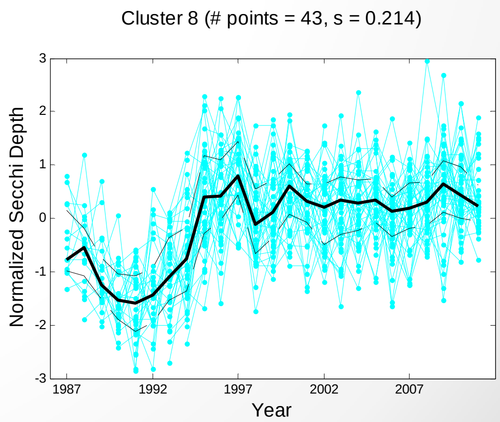
50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W



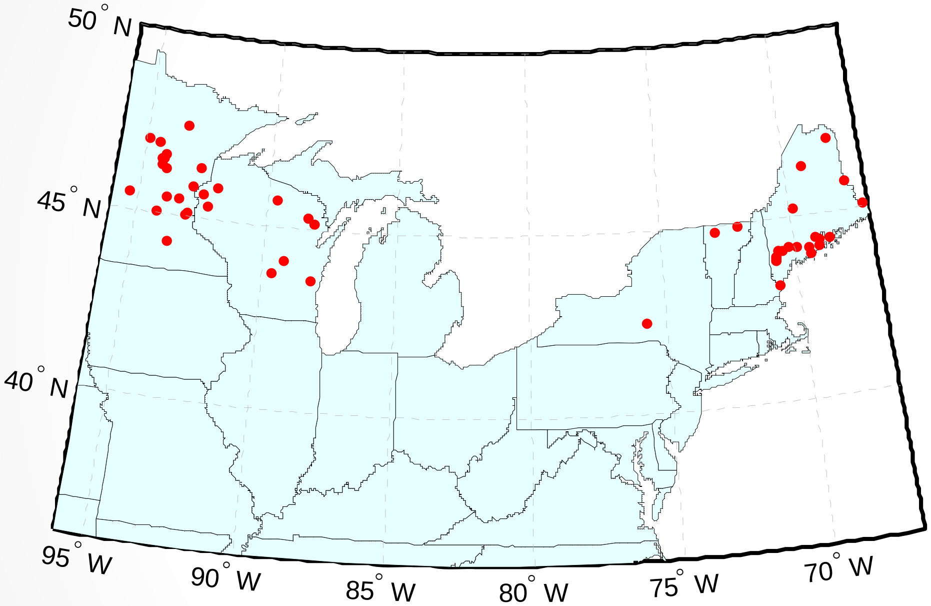
Cluster 7 (# points = 64, s = 0.246)
Normalized Secchi Depth
Year
4
3
2
1
0
-1
-2
-3
-4
1987
1992
1997
2002
2007

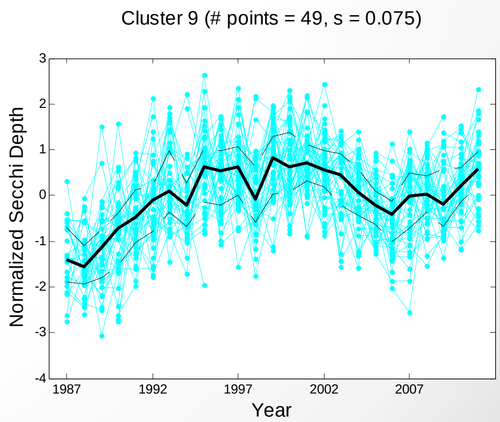
Cluster 8


50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W

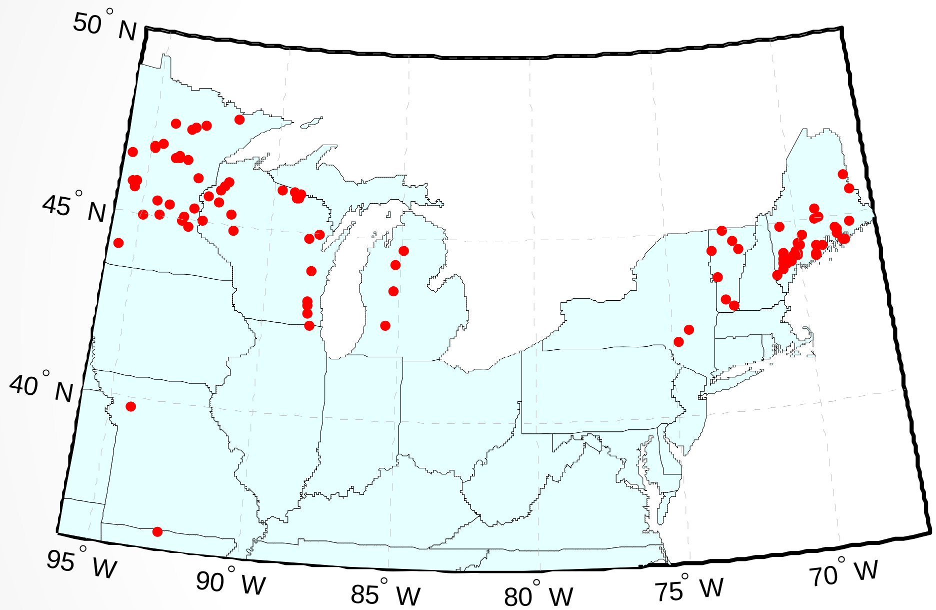

Cluster 8 (# points = 43, s = 0.214)
Normalized Secchi Depth
Year
1987
1992
1997
2002
2007
3
2
1
0
-1
-2
-3

Cluster 9

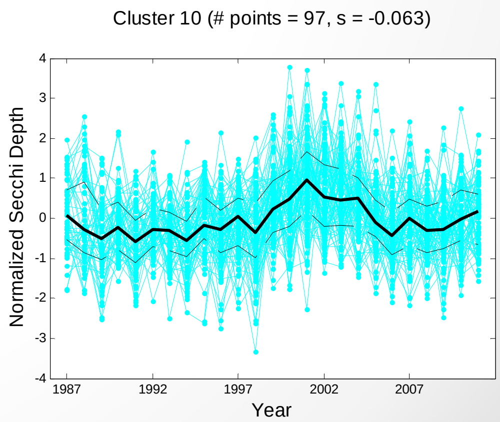

50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W


Cluster 9 (# points = 49, s = 0.075)
Normalized Secchi Depth
Year
3
2
1
0
-1
-2
-3
-4
1987
1992
1997
2002
2007

Cluster 10

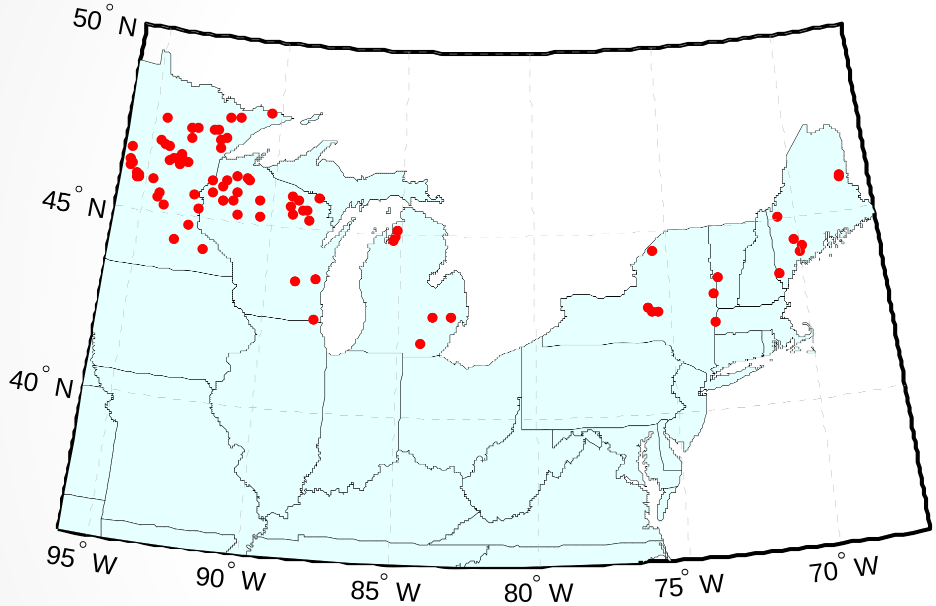


50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W

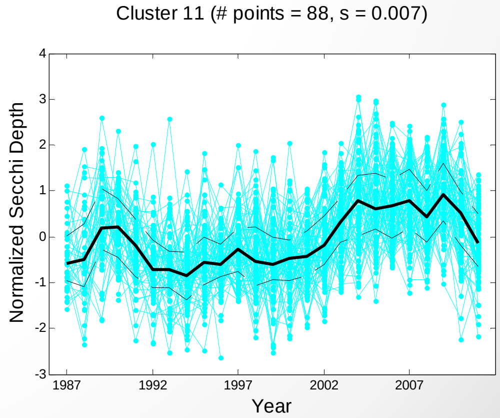


Cluster 10 (# points = 97, s = -0.063)
Normalized Secchi Depth
Year
1987
1992
1997
2002
2007
4
3
2
1
0
-1
-2
-3
-4

Cluster 11

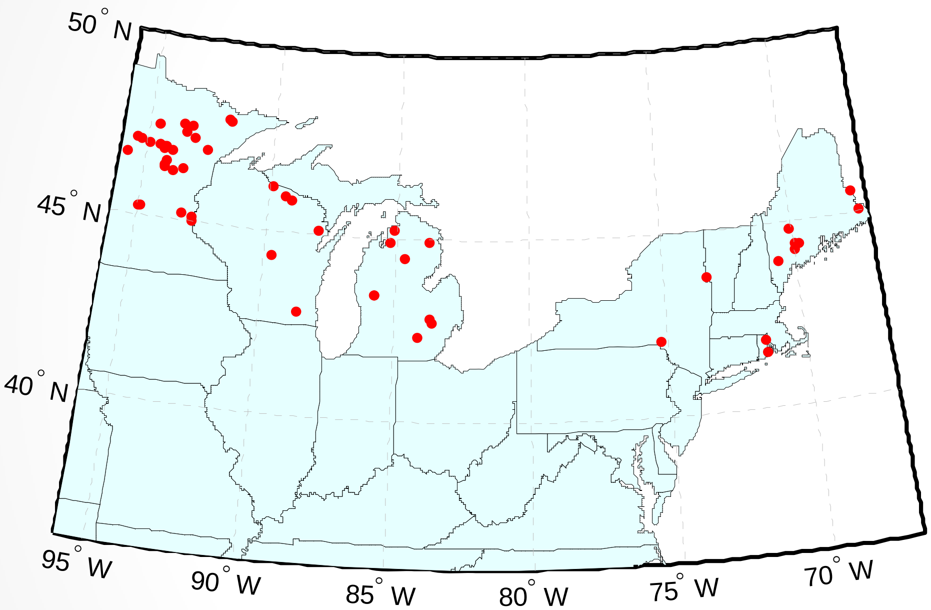


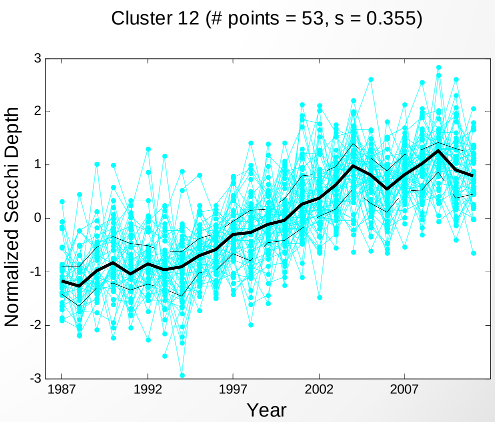
50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W



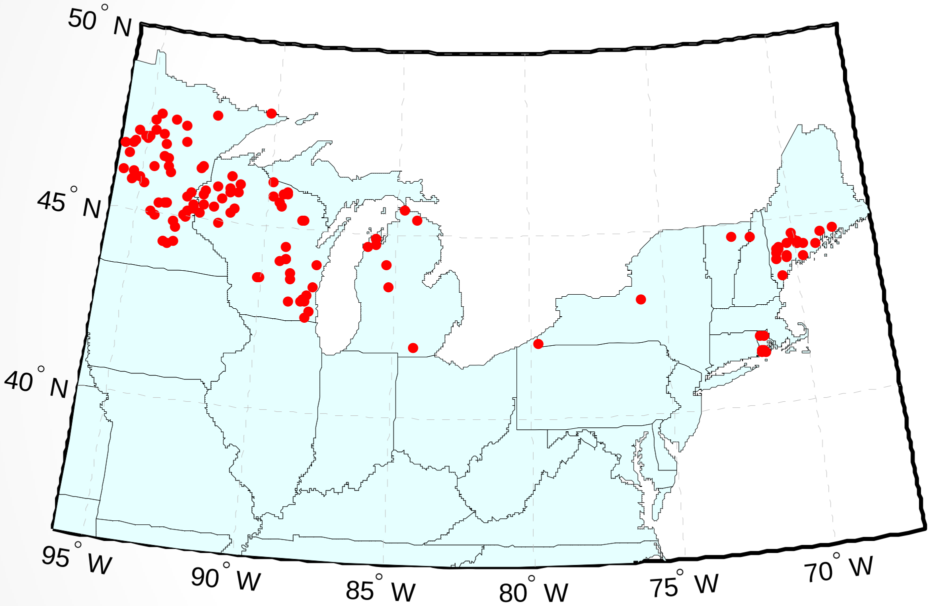
Cluster 11 (# points = 88, s = 0.007)
Normalized Secchi Depth
Year
4
3
2
1
0
-1
-2
-3
1987
1992
1997
2002
2007

Cluster 12

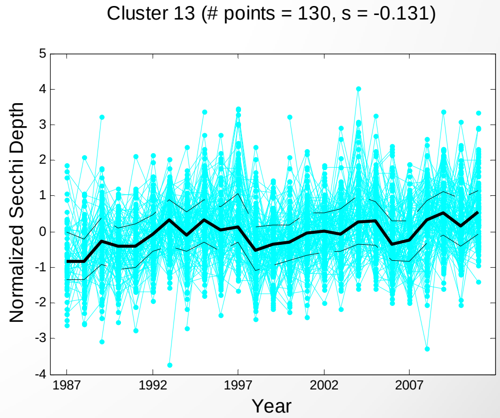

50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W


Cluster 12 (# points = 53, s = 0.355)
Normalized Secchi Depth
Year
3
2
1
0
-1
-2
-3
1987
1992
1997
2002
2007

Cluster 13



50° N
45° N
40° N
95° W
90° W
85° W
80° W
75° W
70° W



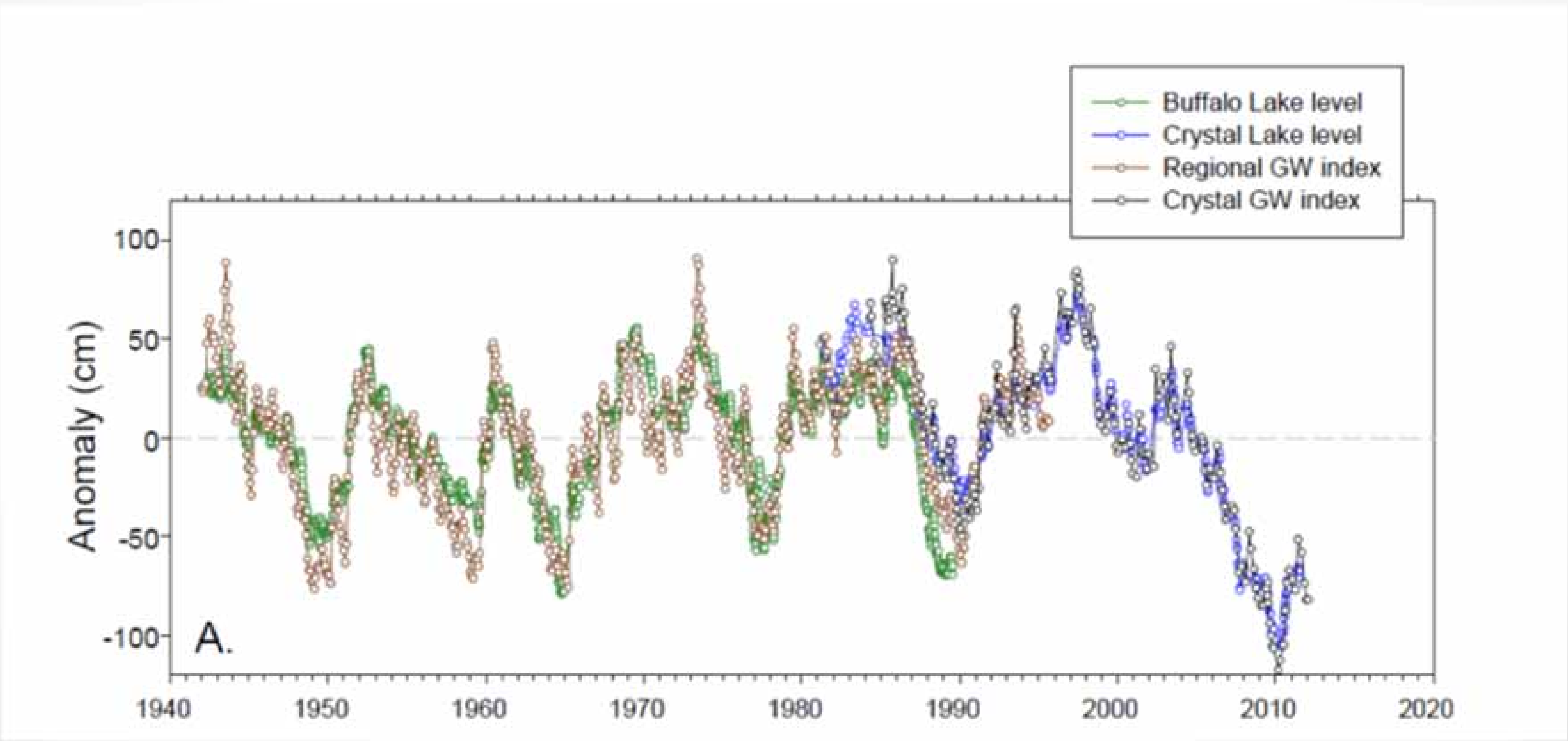
Cluster 13 (# points = 130, s = -0.131)
Normalized Secchi Depth
Year
1987
1992
1997
2002
2007
5
4
3
2
1
0
-1
-2
-3
-4

Conclusions

- Simple models may not be adequate for perceiving of long-term patterns in the age of Big Data
- Data driven perception of long-term change
- Diverse set of common water clarity dynamics

Opportunities…

- Citizens data is a critical source of information

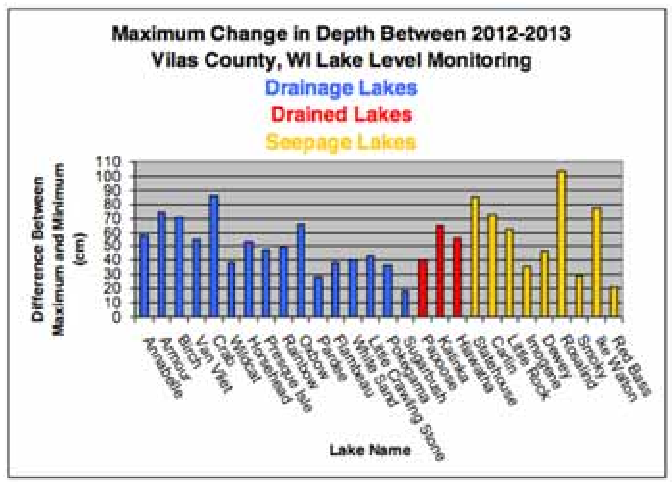


Watras et al. GRL (2014)

North Lakeland Discovery Center

WISCONSIN DNR
Department of Natural Resources

Citizen-based lake level monitoring



lakechange.org

discoverycenter.net

Poster Session