SPATIAL-TEMPORAL
ASPECTS OF
WATER QUALITY

N. Scott Urquhart
F. Jay Breidt
Mark J. Delorey
Department of Statistics
Colorado State University
Fort Collins, CO 80523-1877

STARMAP FUNDING

- **EPA Funded Program**
- **ROUTINE (REQUIRED) DISCLAIMER:**
  - **The work reported here today was developed under the STAR Research Assistance Agreement CR-829095 awarded by the U.S. Environmental Protection Agency (EPA) to Colorado State University. This presentation has not been formally reviewed by EPA. The views expressed here are solely those of presenters and the STARMAP, the Program they represent. EPA does not endorse any products or commercial services mentioned in this presentation.**

This research is funded by
U.S.EPA – Science To Achieve Results (STAR) Program
Cooperative Agreement # CR - 829095

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

REALITY = TWO TALKS

- **First part – (45 minutes or so)**
  - **Urquhart**
  - **An example of an important context**
  - **Questions in need of a solution**
  - **+ a bit about STARMAP = Space-Time Aquatic Resources Monitoring and Analysis Program at CSU**

- **Second part (15 minutes or so)**
  - **Breidt and Delorey**
  - **Beginning of a solution for one of the problems**
  - **+ a bit about PRIMES at CSU**

WHAT IS SIMILAR/DIFFERENT ABOUT AQUATIC SYSTEMS?

- **Similar to yesterday morning's presentations:**
  - **Highly multivariate**
  - **Stationary – sometimes, but probably not often**
  - **Sometimes spatially & temporally smooth – storms!!!**
- **Different from yesterday morning's presentations:**
  - **Very different time scale – years, not days**
  - **Data sparse – compared to Cressie's**
  - **Spatially isolated data points**
  - **Frequently spatially one dimension in two-space**
  - **Most aquatic responses are not (currently) sensible from remote platforms**
  - **Berliner: N >> n. Aquatic systems N > n or even N < n**

WHY TALK ABOUT AN EPA REPORT HERE?

- **Many of the presentations here have dealt with SOLUTIONS**
- **My objective today is to expose you to an important time/space problem containing features needing solution**
- **Distinctive features:**
  - **Extensive data set for the type of problem**
  - **Spatially extensive situation**
  - **Data: from probability surveys and convenience sites**
  - **Response = trend, not response size**
  - **Primary summary is estimated cumulative distribution function (cdf)**

A BIT OF HISTORY

- **Initially (late 1800s) electricity was delivered as direct current**
  - **"Generation" facilities had to be close to user**
- **Switch to alternating current occurred during early 1900s, but generating facilities already were in cities**
  - **WW II led to great industrialization, expansion of power generation from coal, and LOTS of air pollution**
  - **This was regarded as a local problem, regulated by cities and counties**
    - **Power generation was exported from cities to coal fields**
    - **Electricity delivered by massive transmission lines**

COMPETING FORCES

- State Public Utility Commissions leaned on utilities to keep prices down (1950s - 1960s)
- Power plants and their pollution got exported hundreds of miles from users
- Ex: Los Angeles and four corners generation
- To avoid local pollution, high smoke stacks pushed smoke plumes up hundreds of feet
  - Smoke plumes traveled great distances
  - This led to the Clean Air Act of 1977
  - It mainly regulated particulate emissions
    - Began working on auto emissions

IN THE 1980s

- **Importance of other emissions was recognized:**
  - **Ozone**
  - **Precursors of "acid rain"**
    - **Sulfur dioxide (SO_2 & SO_X) + H_2O ===> sulfuric acid**
    - **Nitric oxide (NO_2 & NO_X) + H_2O ===> nitric acid**
  - **Health effects of "invisible" emissions documented**
- **EPA conducted probability surveys (one-time) of streams and lakes to identify acid sensitive areas**
  - **Predecessor of Environmental Monitoring and Assessment Program (EMAP)**
- **Above led to the 1990 amendments to the Clean Air Act → Report due in 2002**

RESPONSE OF SURFACE WATER CHEMISTRY TO THE CLEAN AIR ACT AMENDMENTS OF 1990

{an EPA Report to Congress}

by

John Stoddard, Jeffrey Kahl, Frank Deviney, David DeWalle, Charles Driscoll, Alan Herlihy, James Kellogg, Peter Murdoch, James Webb, and Katherine Webster

INTERNET ADDRESS:

www.epa.gov/ordntrnt/ORD/htm/CAAA-2002-report-2col-rev-4.pdf/

Rest of citation: Environmental Protection Agency,

EPA 620/R-03/001, Research Triangle Park, NC 27711

BACKGROUND

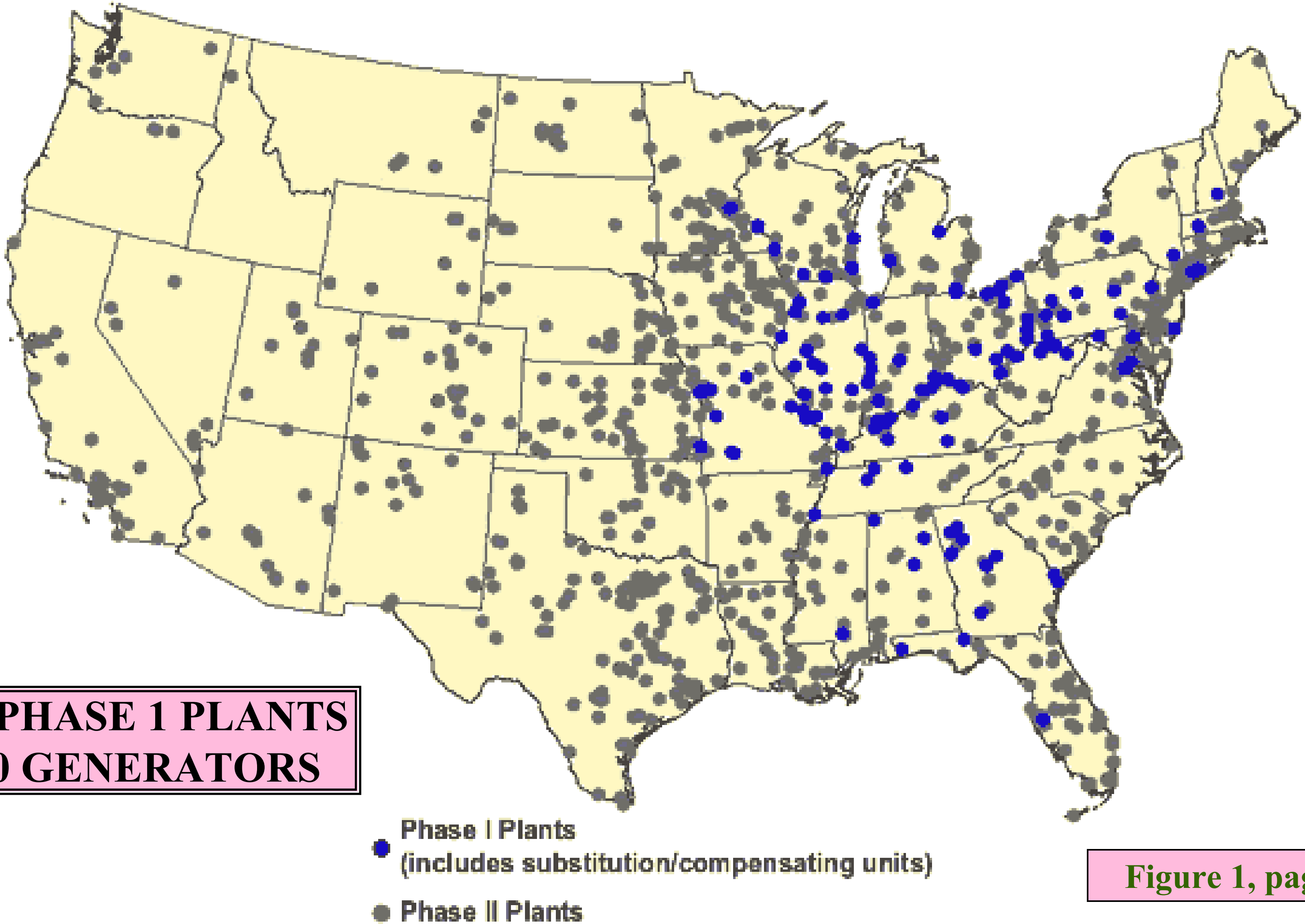
- Congress enacted Amendments to the Clear Air Act of 1977 in 1990.
  - SO_2 was the major atmospheric pollutant contributing to "acid rain."
  - 1990 output of SO_2 was about 20 million tons/year.
  - 110 power plants were required to reduce their SO_2 output by 10 million tons (50% of total) by 1995.
  - About 2,000 power plants were required to reduce their output of SO_2 by more than an additional 50% by 2000.
  - Substantial penalties for noncompliance = $1/# of SO_2 output.

BACKGROUND
CONTINUED

- **Congress enacted Amendments to the Clear Air Act of 1977 in 1990.**
  - **Section 901. CLEAN AIR RESEARCH**
    - **(j) specified a biennial report to Congress**
      - **Actual and projected emissions and acid deposition trends;**
      - **Average ambient concentrations of acid deposition precursors and their transformation products;**
      - **The status of ecosystems (including forests and surface waters), materials, and visibility affected by acid deposition;**
      - **The causes and effects of such deposition, including changes in surface water quality and forest and soil conditions;**
      - **The occurrence and effects of episodic acidification, particularly with respect to high elevation watersheds; and**
      - ***The confidence level associated with each conclusion to aid policymakers in use of the information.***

AFFECTED SOURCES

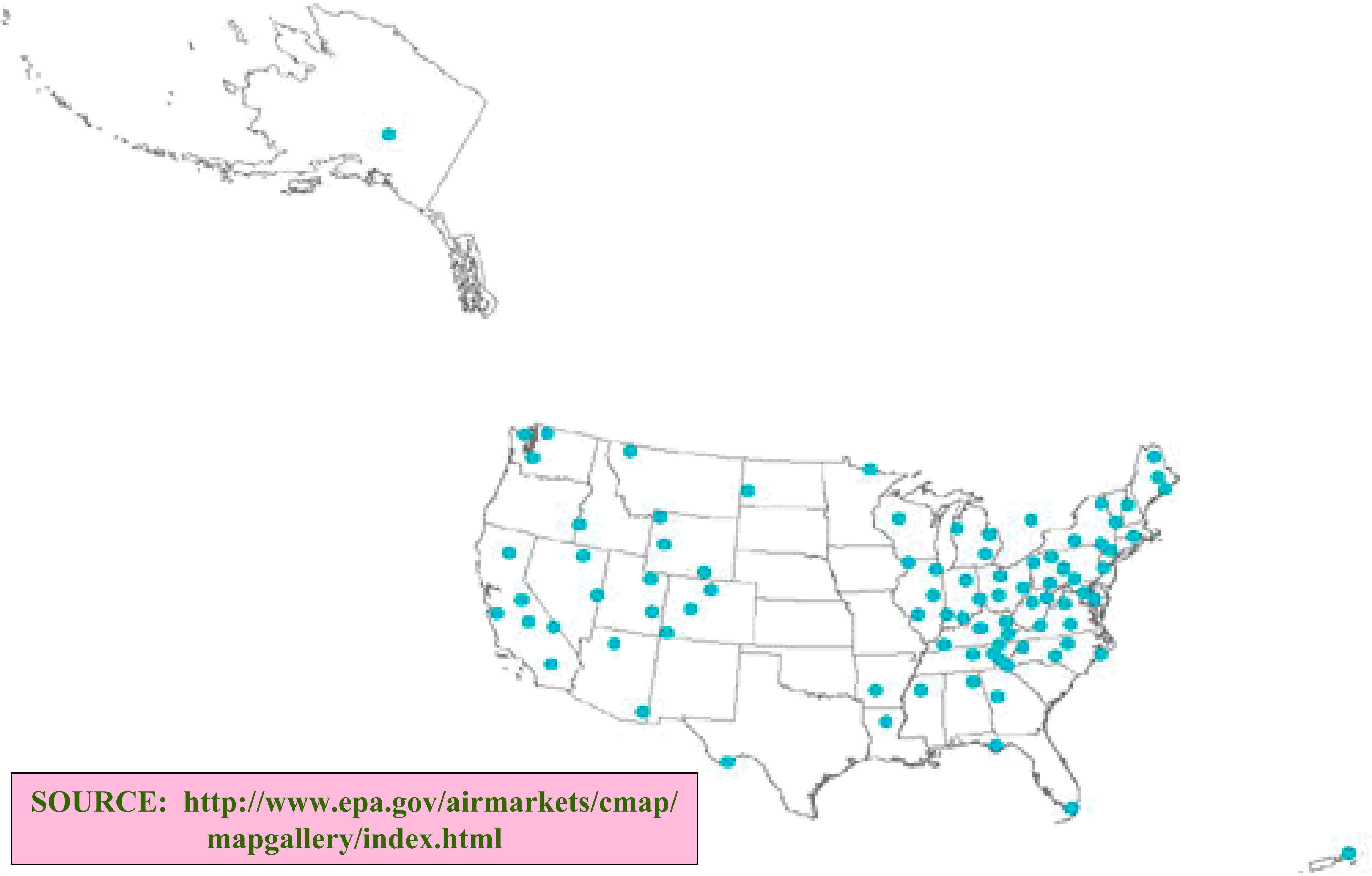
by CAAA, 1990



110 PHASE 1 PLANTS
220 GENERATORS

Figure 1, page 3

Clean Air Status and Trends Network CASTNet



SOURCE: http://www.epa.gov/airmarkets/cmap/mapgallery/index.html

ACID SENSITIVE REGIONS
of the
NORTHERN and EASTERN
UNITED STATES

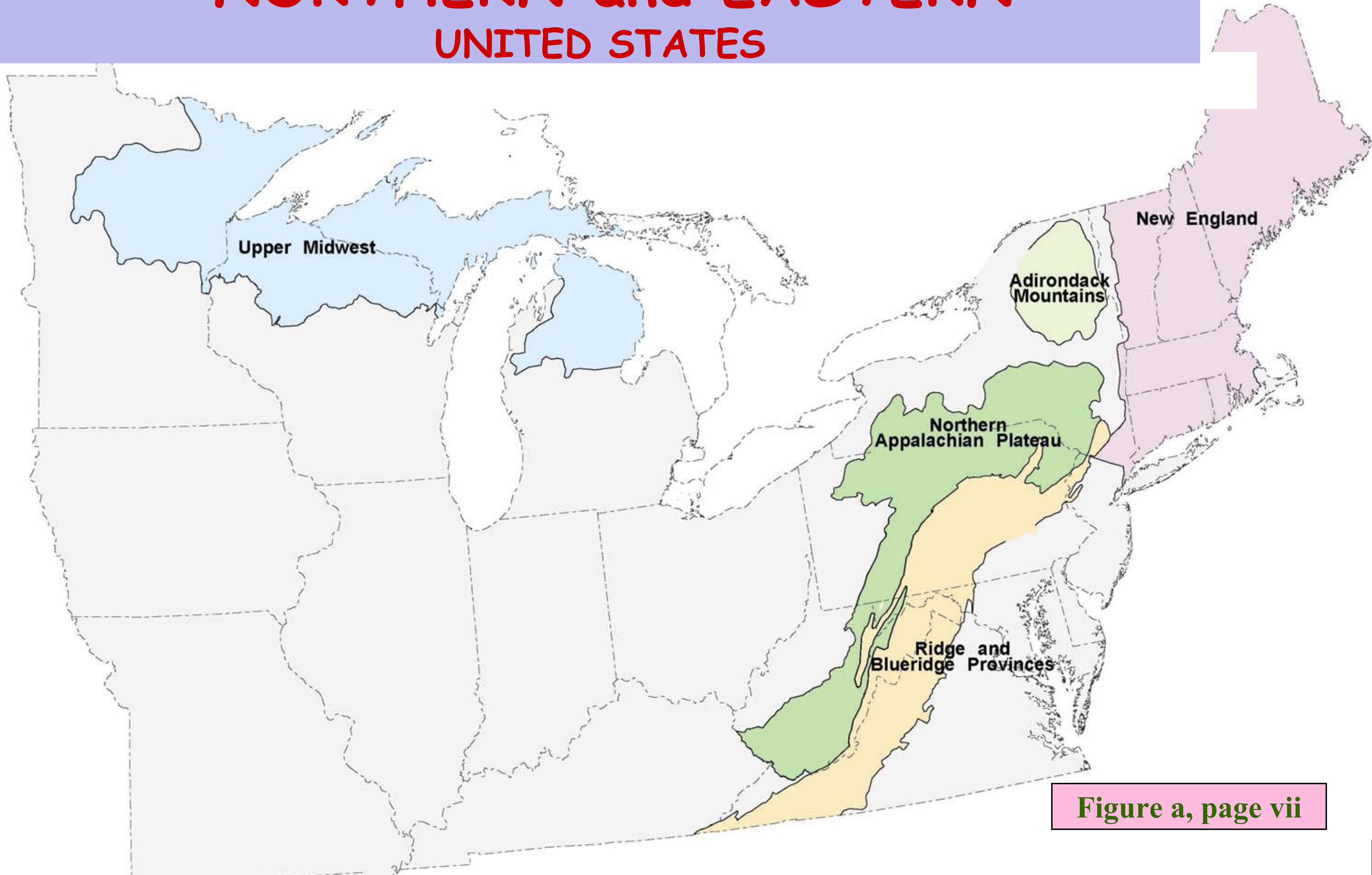


Figure a, page vii

SULFATE EMISSIONS BEFORE & AFTER CAAA, 1990

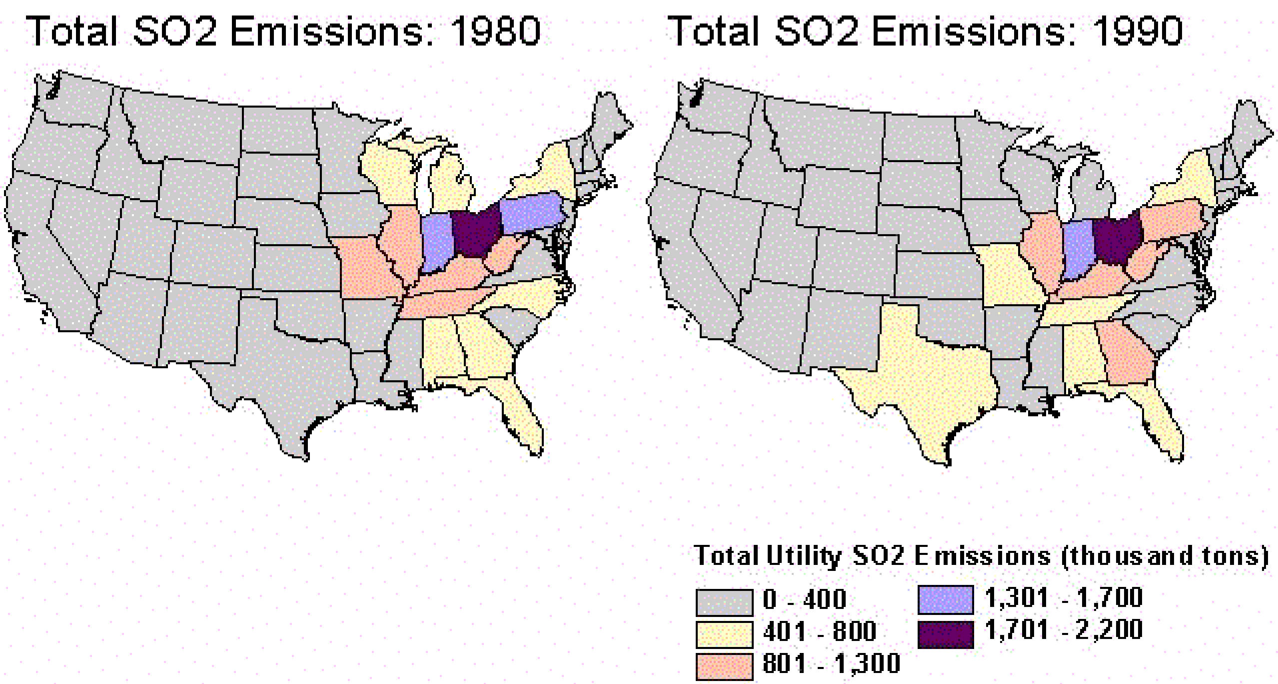
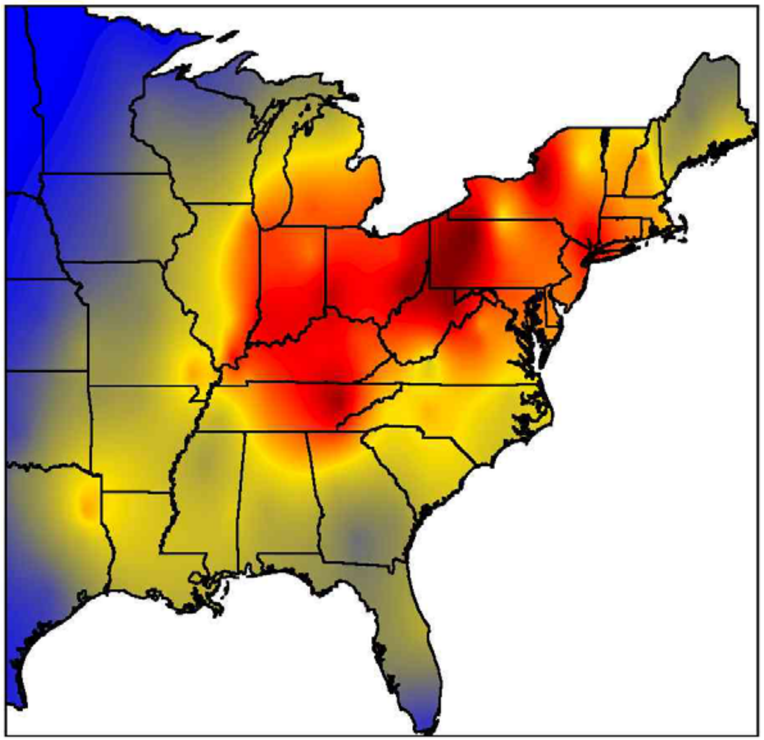


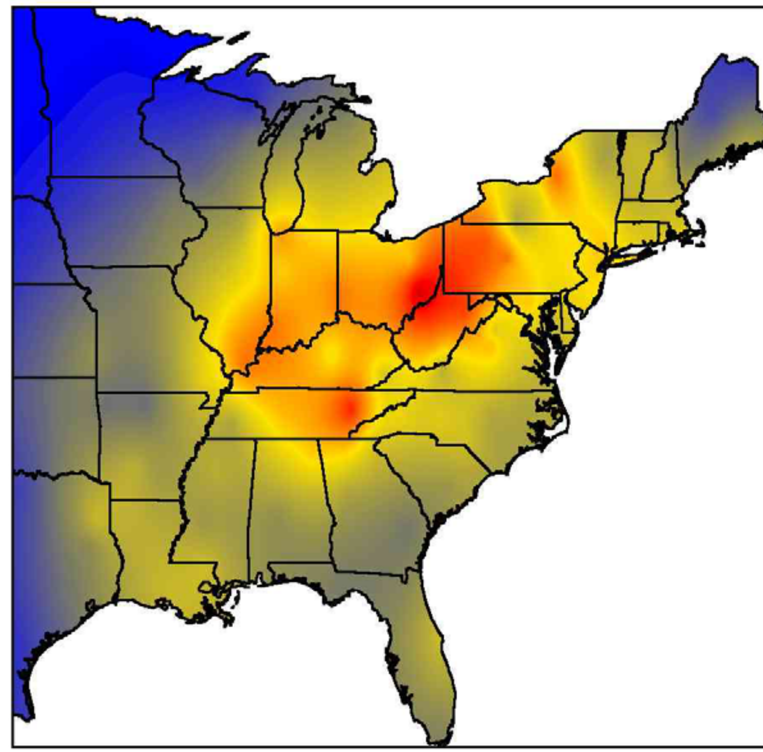
Figure 5, page 21, also http://www.epa.gov/airmarkets/cmap/mapgallery/index.html

WET DEPOSITION OF SULFATE

1989 - 1991



1995 - 1998



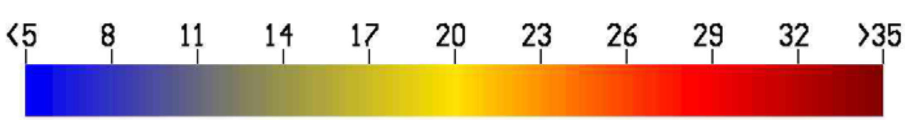


Figure 6, page 21, also http://www.epa.gov/airmarkets/cmap/mapgallery/index.html

NITRATE EMISSIONS BEFORE & AFTER CAAA, 1990

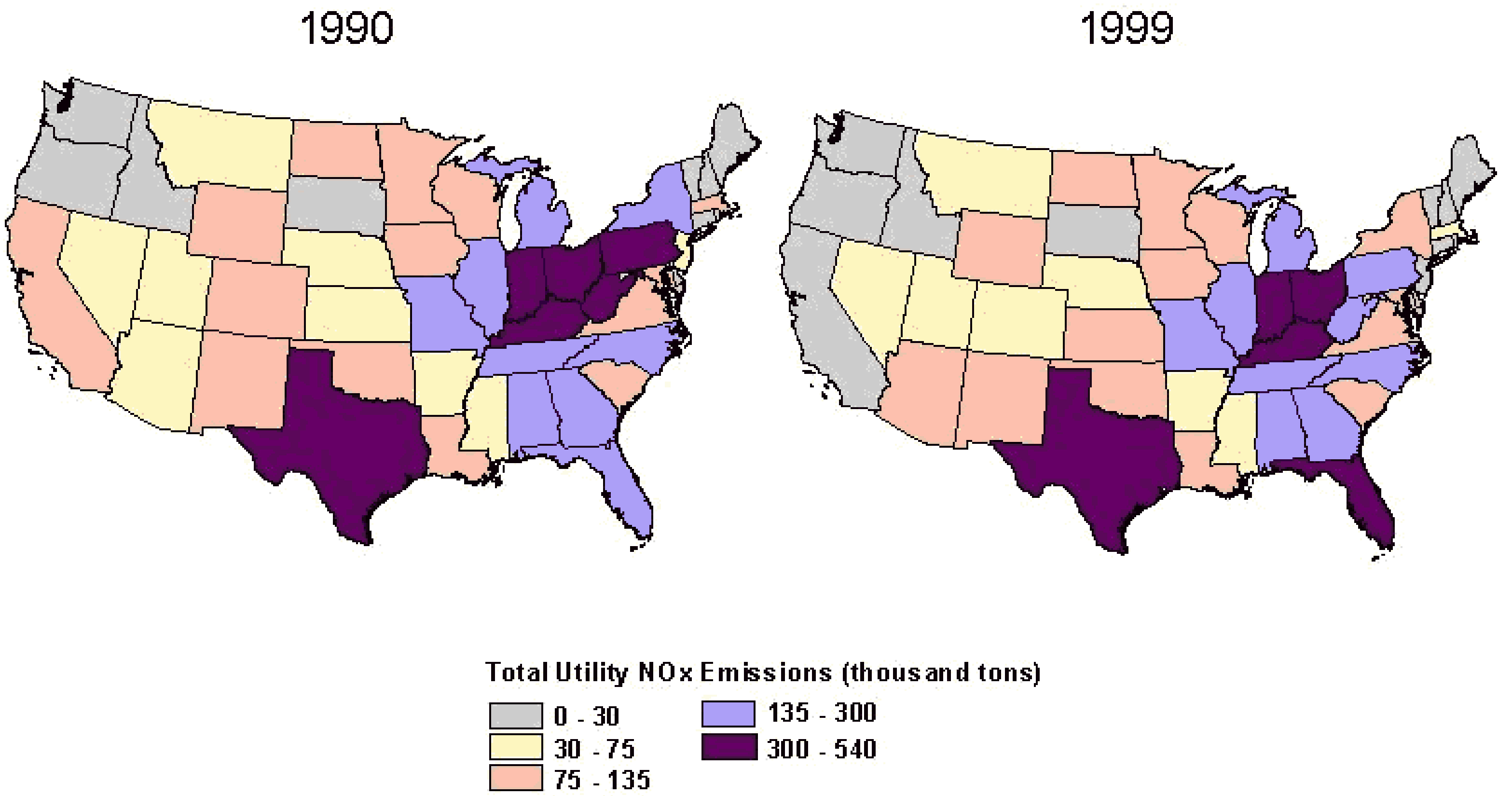
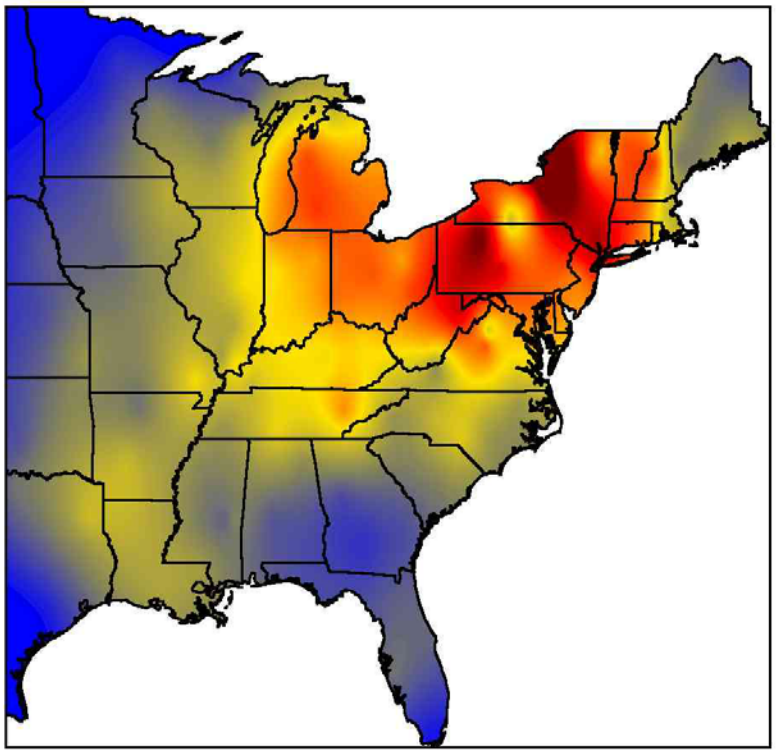


Figure 8, page 24, also http://www.epa.gov/airmarkets/cmap/mapgallery/index.html

WET DEPOSITION OF NITRATE

1989 - 1991



1995 - 1998

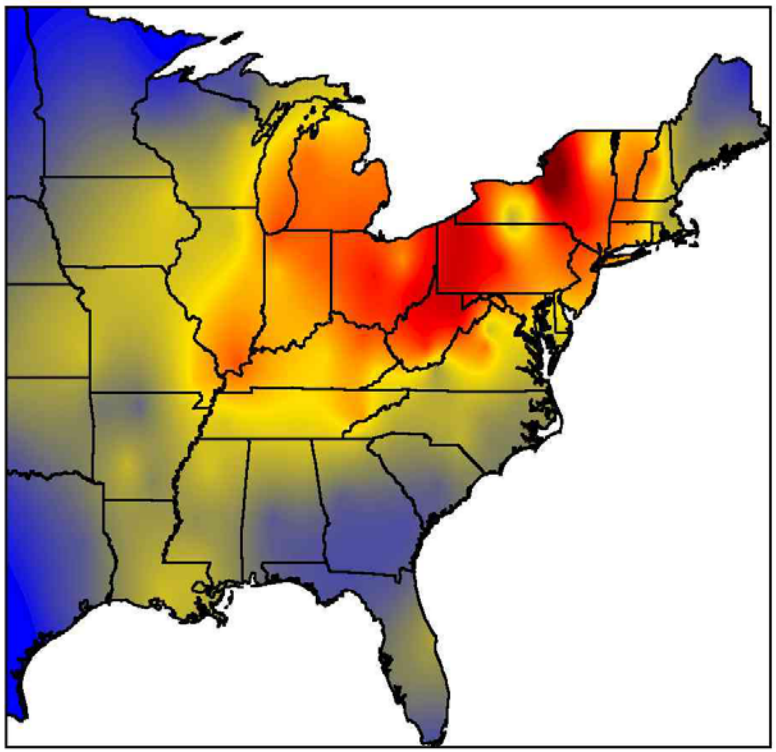


Figure 9, page 25, also http://www.epa.gov/airmarkets/cmap/mapgallery/index.html

SOURCE OF DEPOSITION MAPS

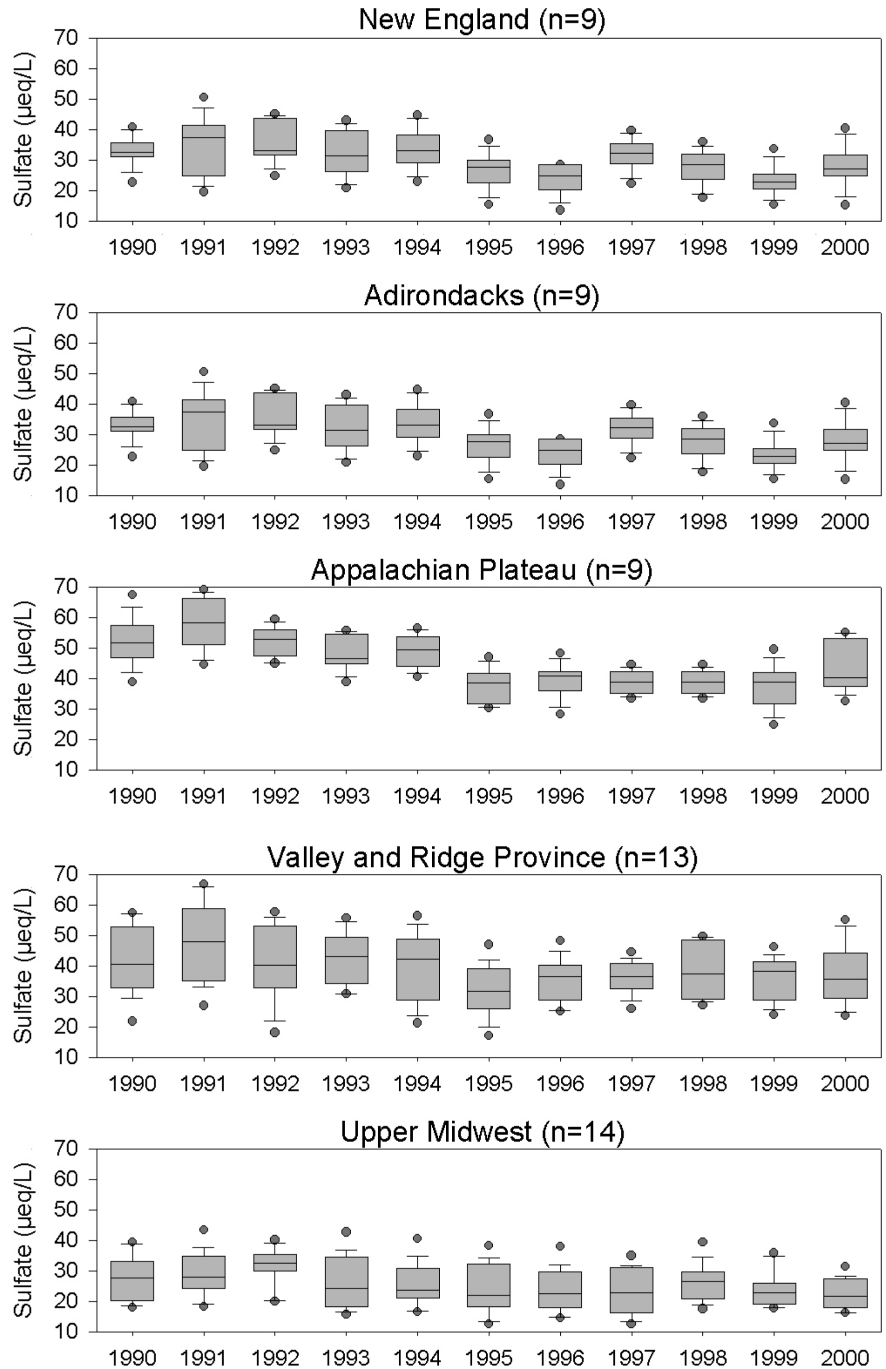
- **http://www.epa.gov/airmarkets/cmap/mapgallery/index.html**
- **The deposition maps were calculated in what seems to be a very statistically naïve way: "Multiquadric Equations."**
  - for data points (x_i, y_i, z_i), solve the equations

$$\sum_{i=1}^{n} c_j [(x_j - x_i)^2 + (y_j - y_i)^2]^{1 \ or \ 0.5} = z_i$$

Then evaluate the result over a dense grid, and map it.

SULFATE DEPOSITION IN THE AREAS OF INTEREST

(Figure 7, page 23)



Note decrease through 1995/1996. Utilities realize they had overshot the requirements of the CAAA for 1995!

ANALYSIS APPROACH

- **Deal with time and "trends" by**
  - **Fitting lines to ANC vs date**
  - **Trends, as conceived here, would be detectable as linear trend (without implying all trend is linear)**
- **Summarize with estimated cumulative distribution functions (cdf s)**
  - **Allows for incorporation of variable probability in the estimation process**

SULFATE CONCENTRATION TRENDS IN WET DEPOSITION
(NADP/NTN SITES 1990 - 2000)

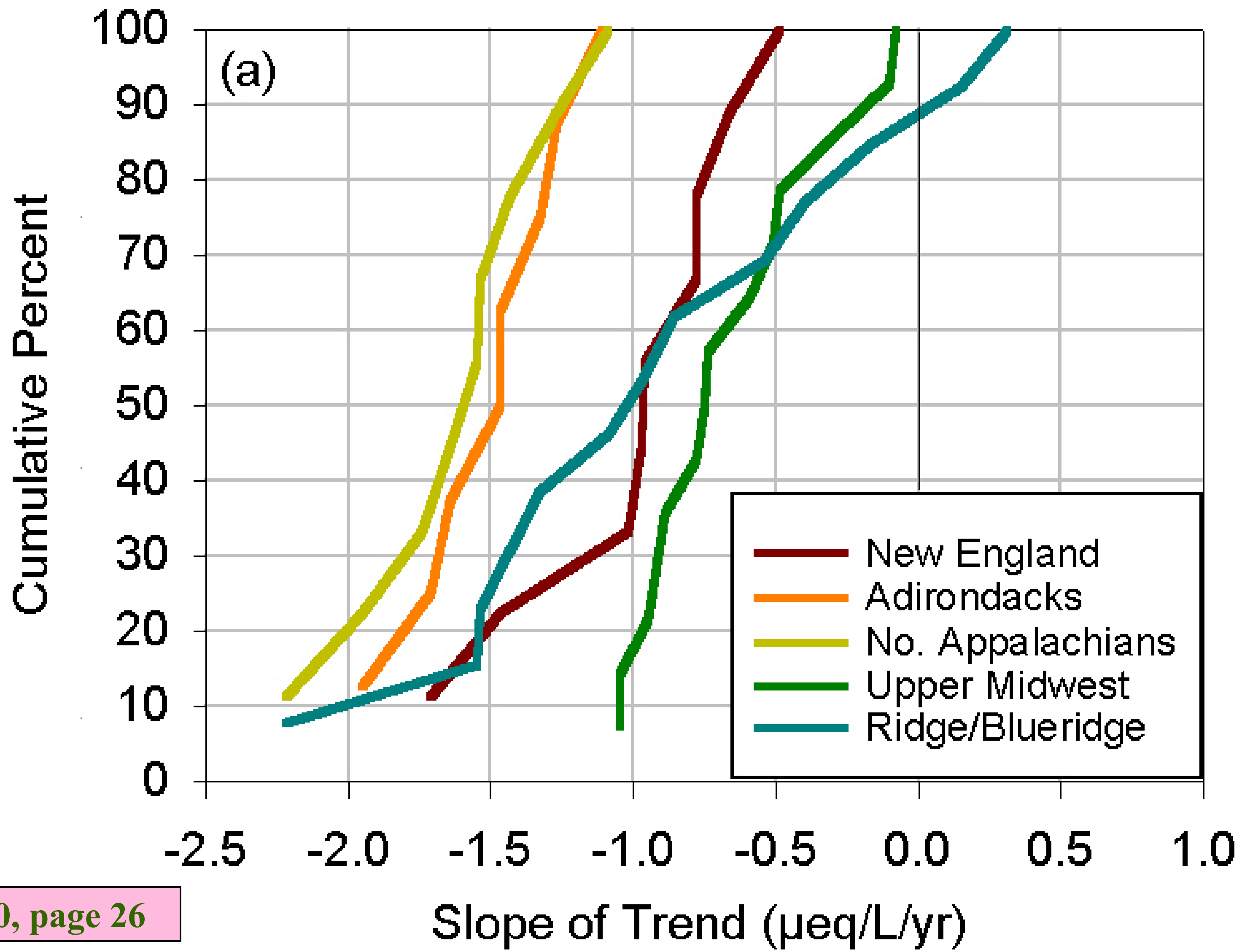


Figure 10, page 26

PRECEDING SHOWED

- **Emissions &**
- **Deposition**
- **Of**
  - **Sulfate &**
  - **Nitrate**
- **But what about their effects on surface water?**

Table 1: SOURCES of DATA and SAMPLE SIZES

Sources of Data	No. of sites	Size of Population	Percent acidic 1980s
Statistical Surveys			
New England Lakes	30	4,327 lakes	5%
Adirondack Lakes	43	1,290 lakes	14%
Appalachian Plateau Streams	31	72,000 stream miles	6%
Sensitive Surface Waters			
New England Lakes	24	N.A.	5%
Adirondack Lakes	48	N.A.	14%
Northern Appalachian Streams	9	N.A.	6%
Upper Midwest Lakes	38	N.A.	3%
Ridge/Blue Ridge Streams	69	N.A.	5%

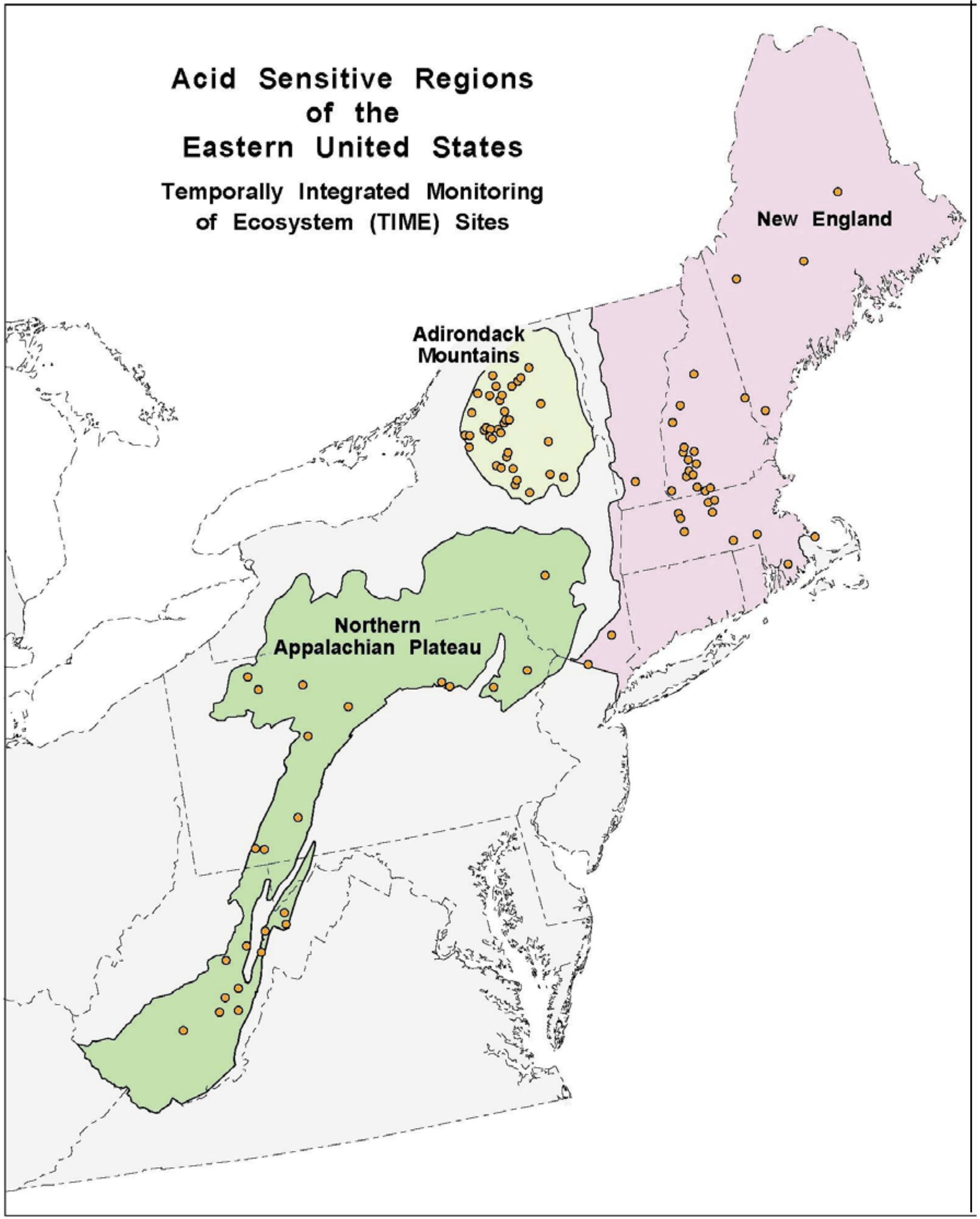


Figure 3, page 7

SELECTION OF TIME SITES

- **Variability probability sample of known population of lakes**
  - **Probability increased with lake size**
- **Variability density sample of known population of streams (continuous sampling model)**
  - **Probability increased with Strahler order of stream**
    - **Strahler order captures how far down in a stream network a particular stream segment is**
    - **Sampling density varied similar to that of lakes**

Table 1: SOURCES of DATA and SAMPLE SIZES

Sources of Data	No. of sites	Size of Population	Percent acidic 1980s
Statistical Surveys			
New England Lakes	30	4,327 lakes	5%
Adirondack Lakes	43	1,290 lakes	14%
Appalachian Plateau Streams	31	72,000 stream miles	6%
Sensitive Surface Waters			
New England Lakes	24	N.A.	5%
Adirondack Lakes	48	N.A.	14%
Northern Appalachian Streams	9	N.A.	6%
Upper Midwest Lakes	38	N.A.	3%
Ridge/Blue Ridge Streams	69	N.A.	5%

ACID SENSITIVE REGIONS OF THE NORTHERN and EASTERN UNITED STATES SHOWING the LONG-TERM MONITORING (LTM) SITES

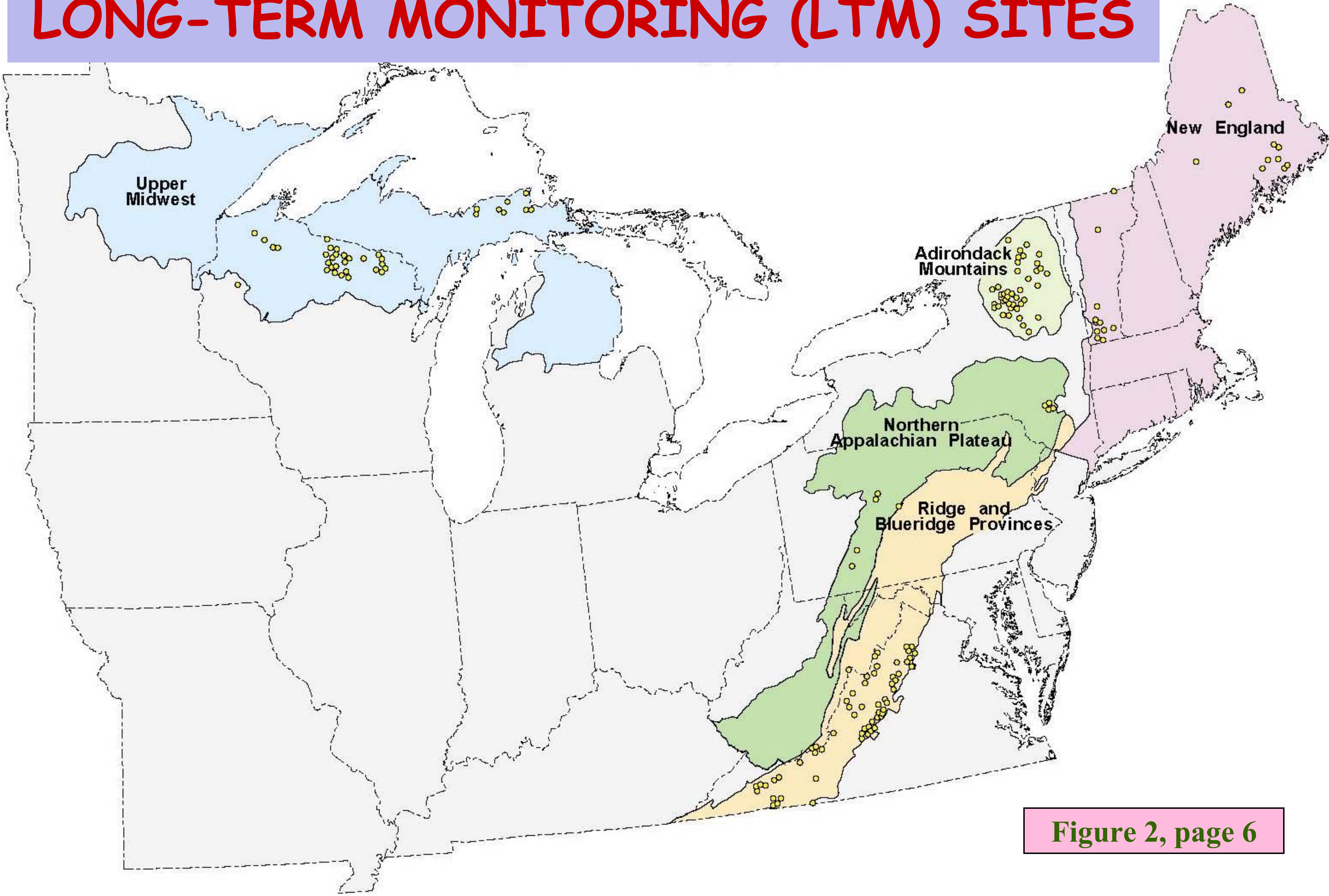


Figure 2, page 6

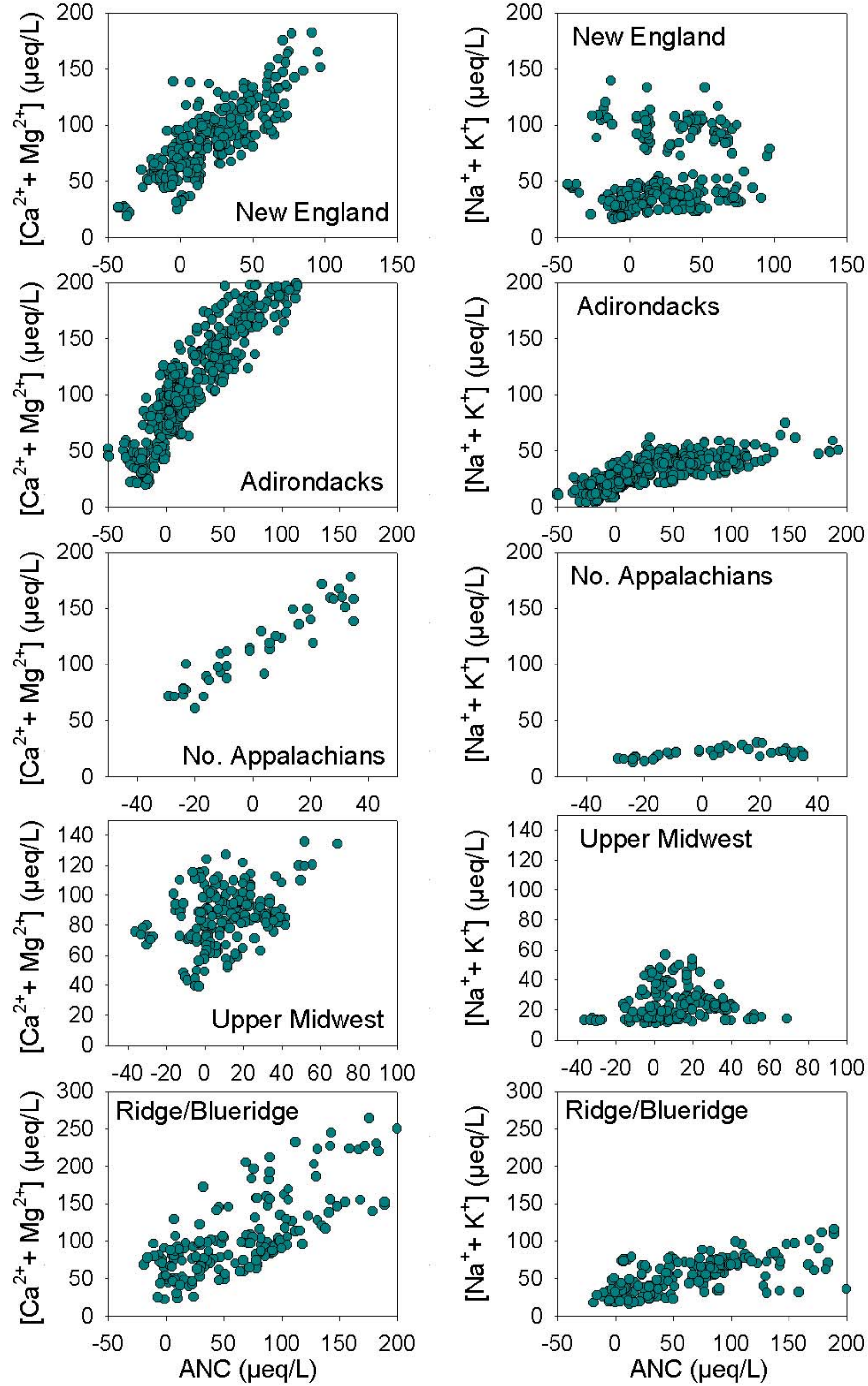
CHOICE OF LONG TERM MONITORING SITES

- **Strictly convenience collection**
  - **Some funded by EPA, others perhaps by state or non-profit organizations**
  - **EPA funded**
    - **Someone at a nearby university would offer to collect the required water samples according to EPA QA/QA**
      - Major factor = ease of access for boat

RELATION OF INTERESTING ANALYTES TO ACID NEUTRALIZING CAPACITY

Figure 13, page 31

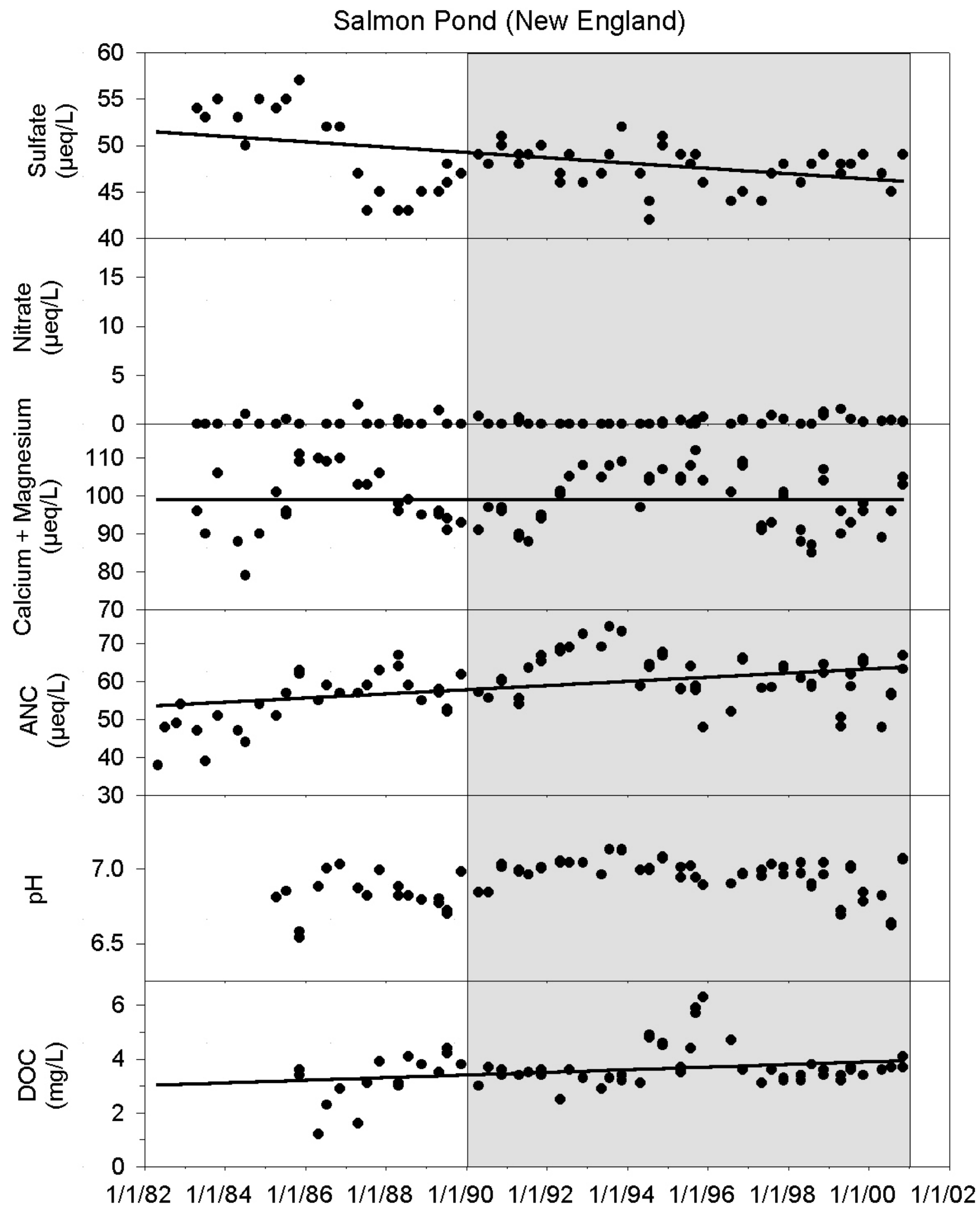
- **Focus on Acid Neutralizing Capacity (ANC) as the major response**
- **Consider the ANC response across regions**



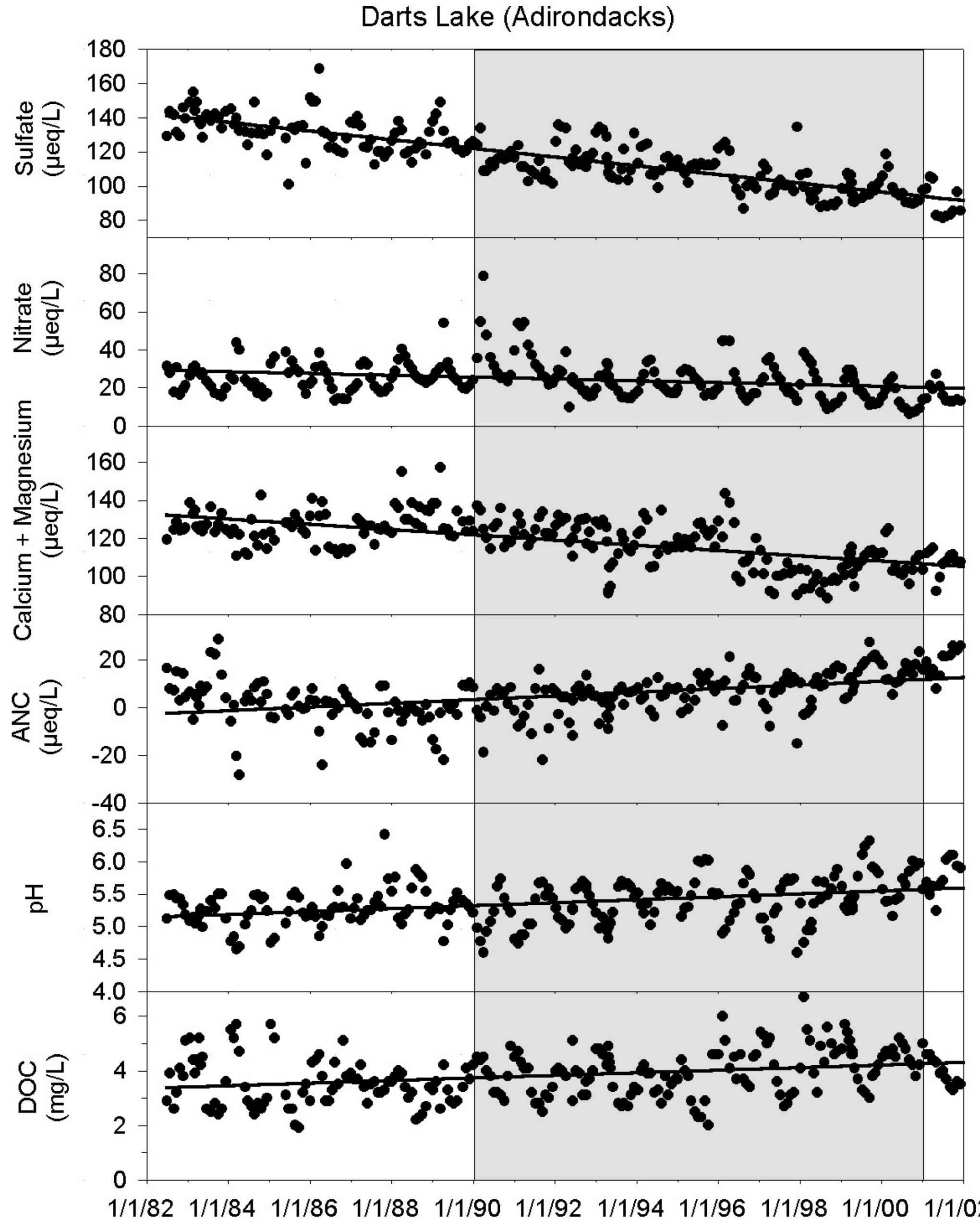
TRENDS IN ANALYTES AT SALMON POND (LTM SITE) (NEW ENGLAND)

Figure 14, page 34

ANC

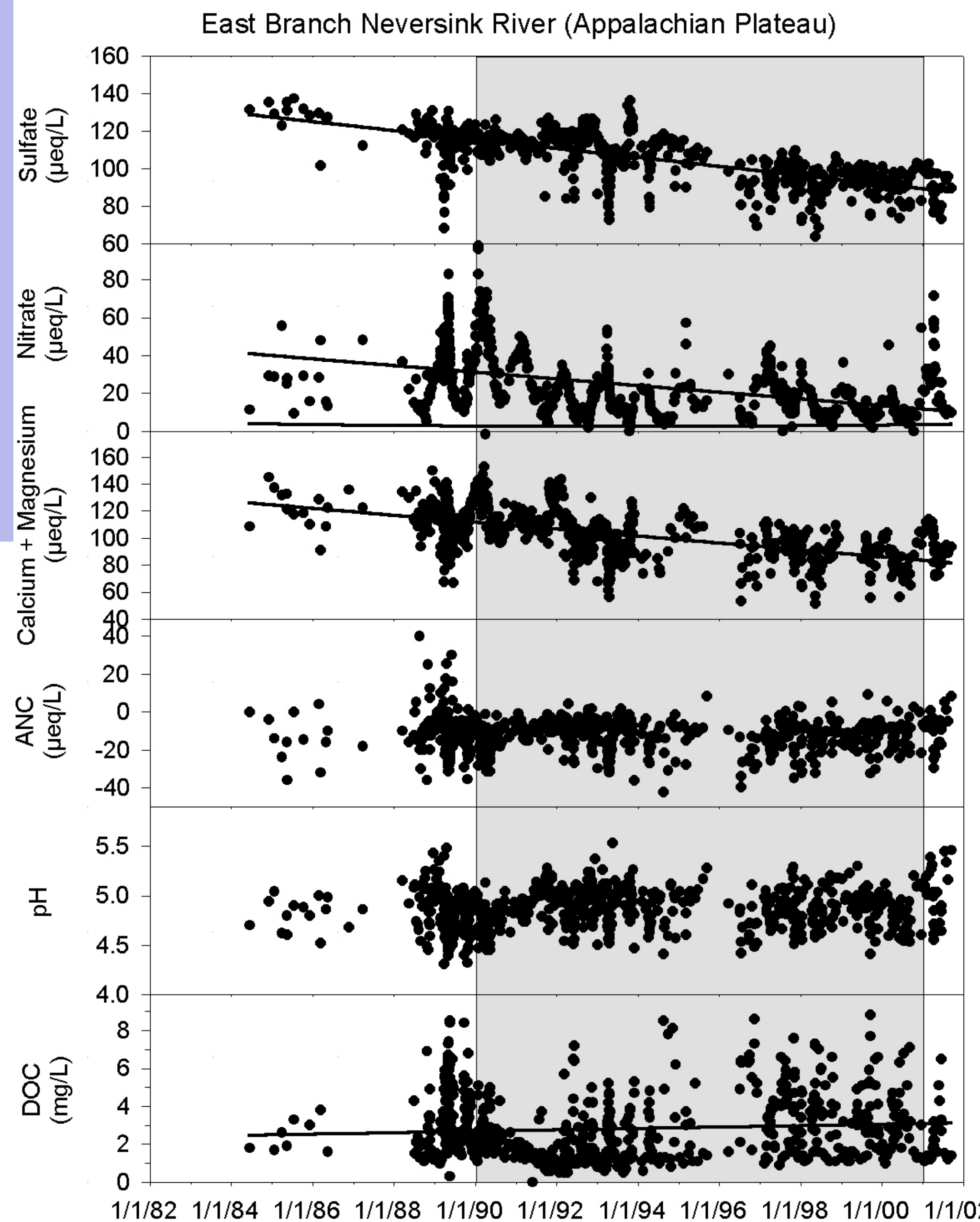


TRENDS IN ANALYTES AT DART LAKE
(LTM SITE) (ADIRONDACKS)
Figure 15, page 35



TRENDS IN ANALYTES AT NEVERSINK RIVER
(LTM SITE)
(APPALACHIAN PLATEAU)
Figure 16, page 36

ANC

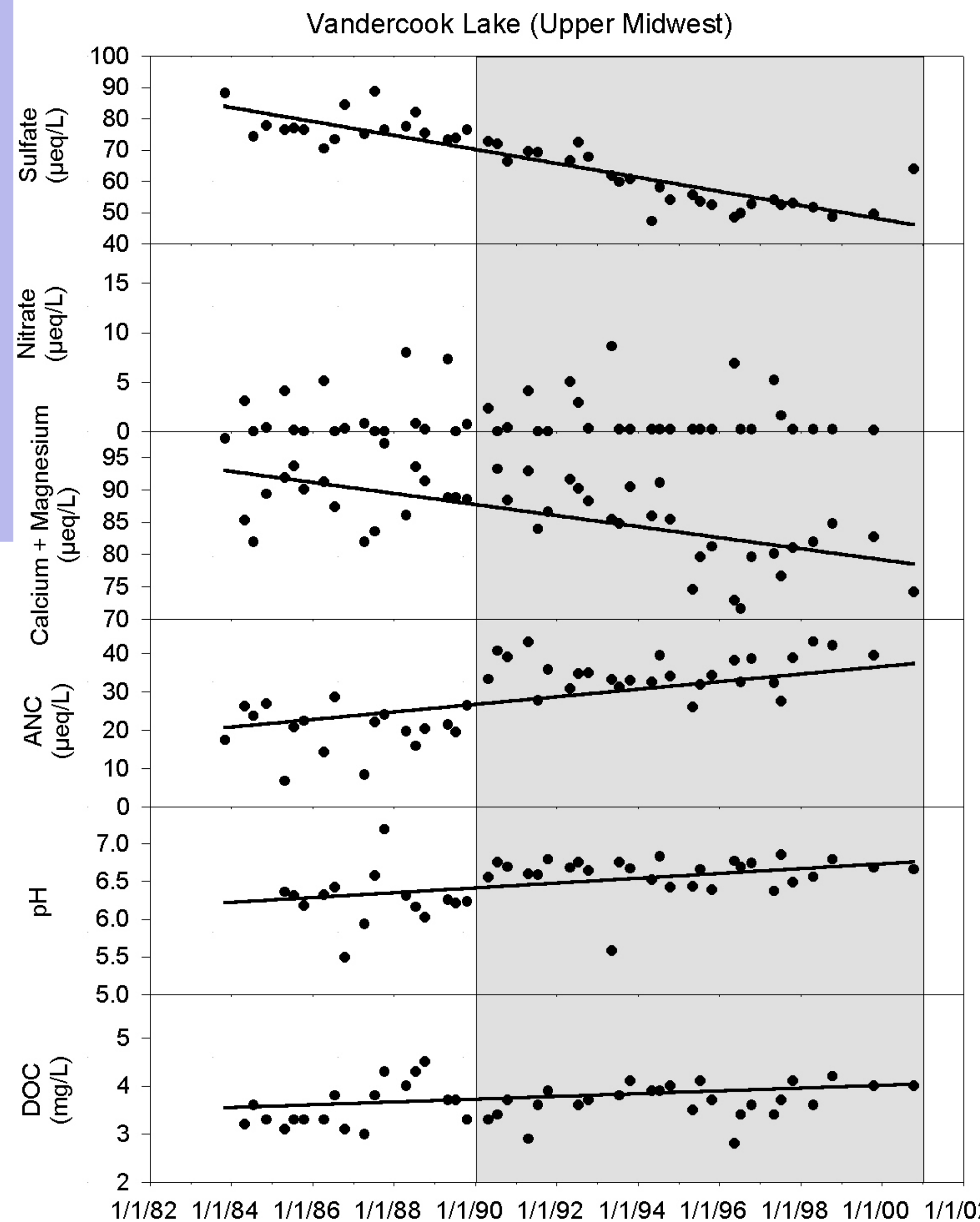


TRENDS IN ANALYTES AT VANDEROOK LAKE (LTM SITE)

(UPPER MIDWEST)

Figure 17, page 37

ANC



TRENDS IN ANALYTES AT STAUNTON RIVER

(LTM SITE)

(RIDGE/BLUE RIDGE REGION)

Figure 18, page 38

ANC

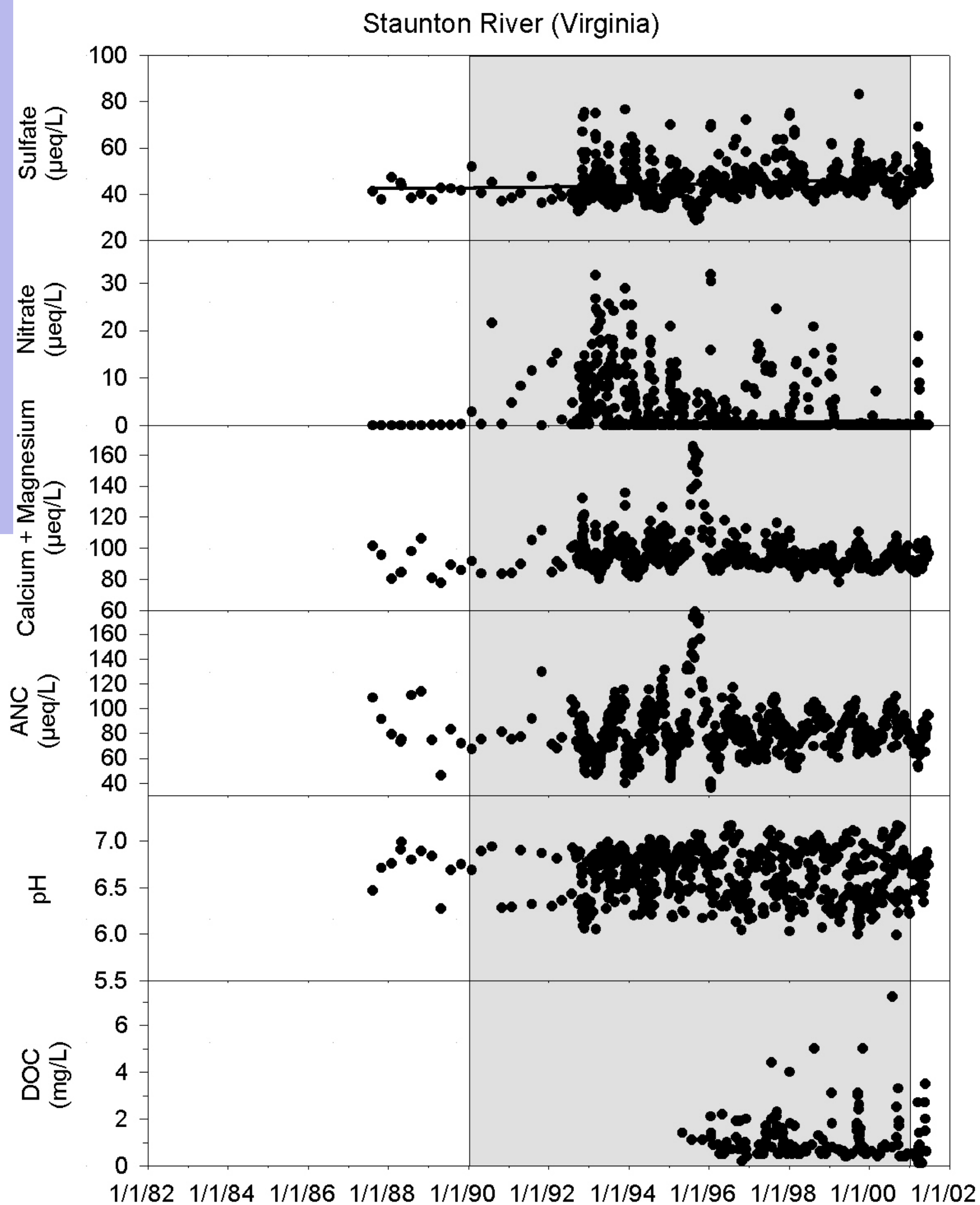
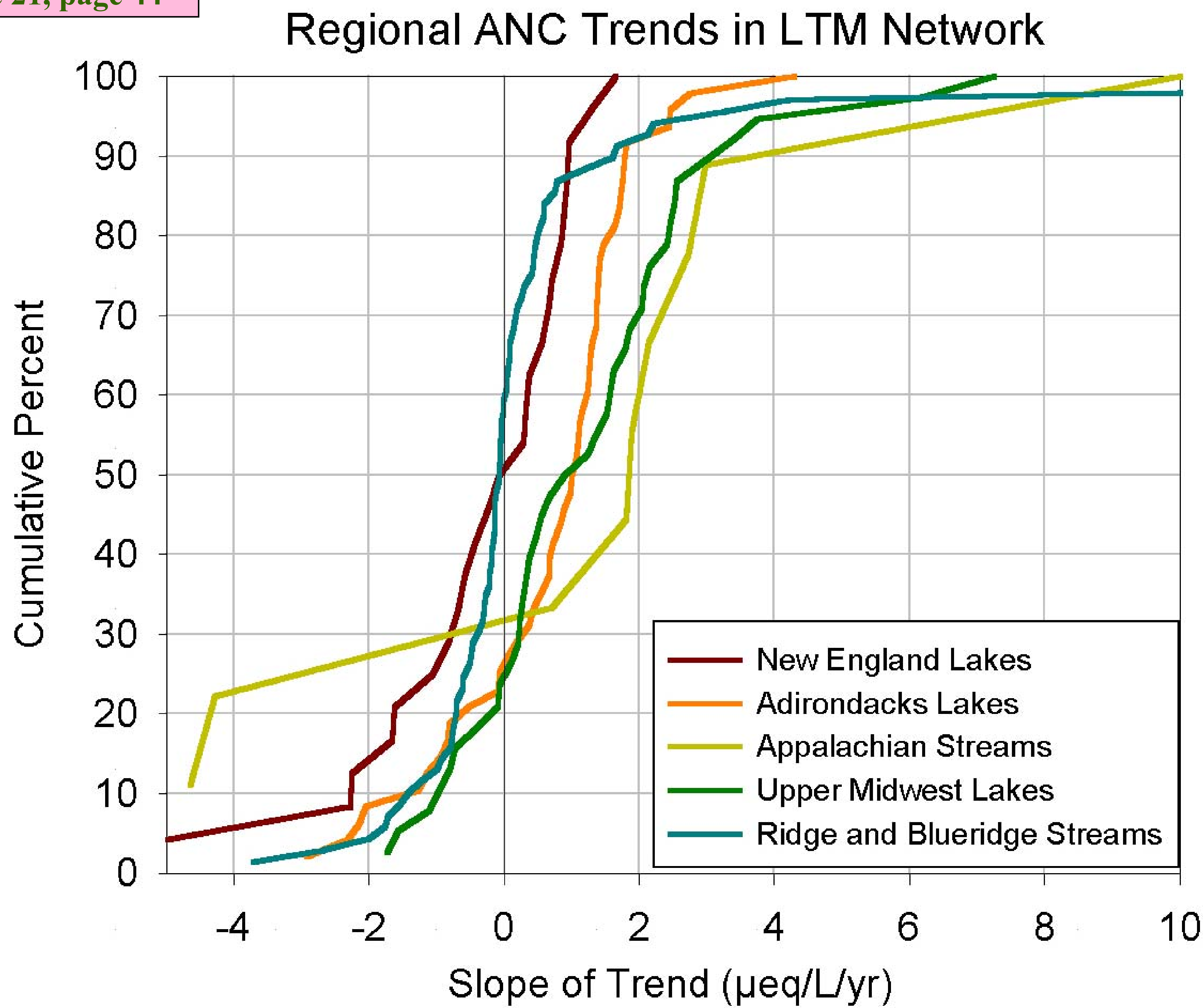
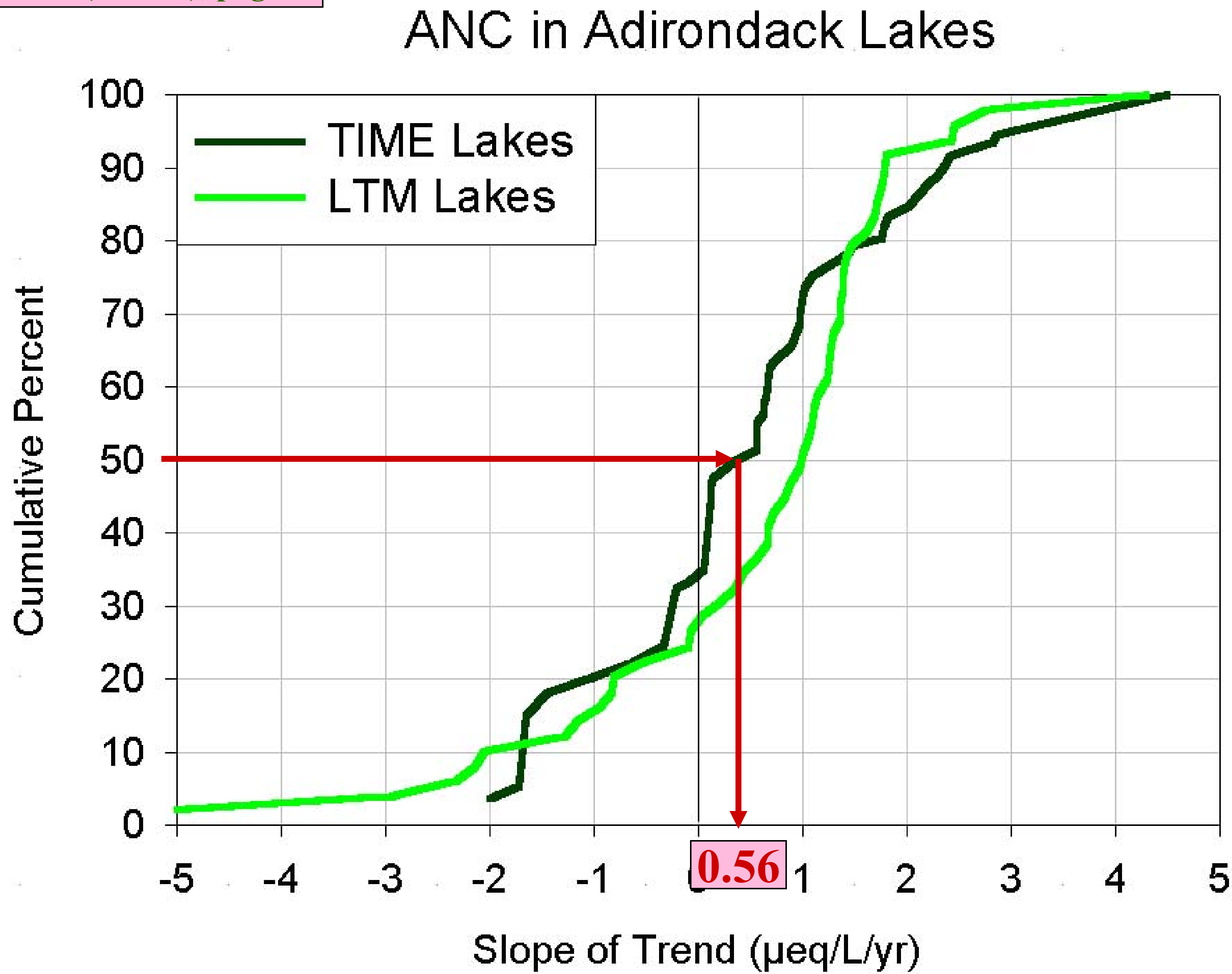


Figure 21, page 44




ANC in Adirondack Lakes
TIME Lakes
LTM Lakes
Cumulative Percent
0
10
20
30
40
50
60
70
80
90
100
-5
-4
-3
-2
-1
0
1
2
3
4
5
0.56
Slope of Trend (µeq/L/yr)

ANC in New England Lakes

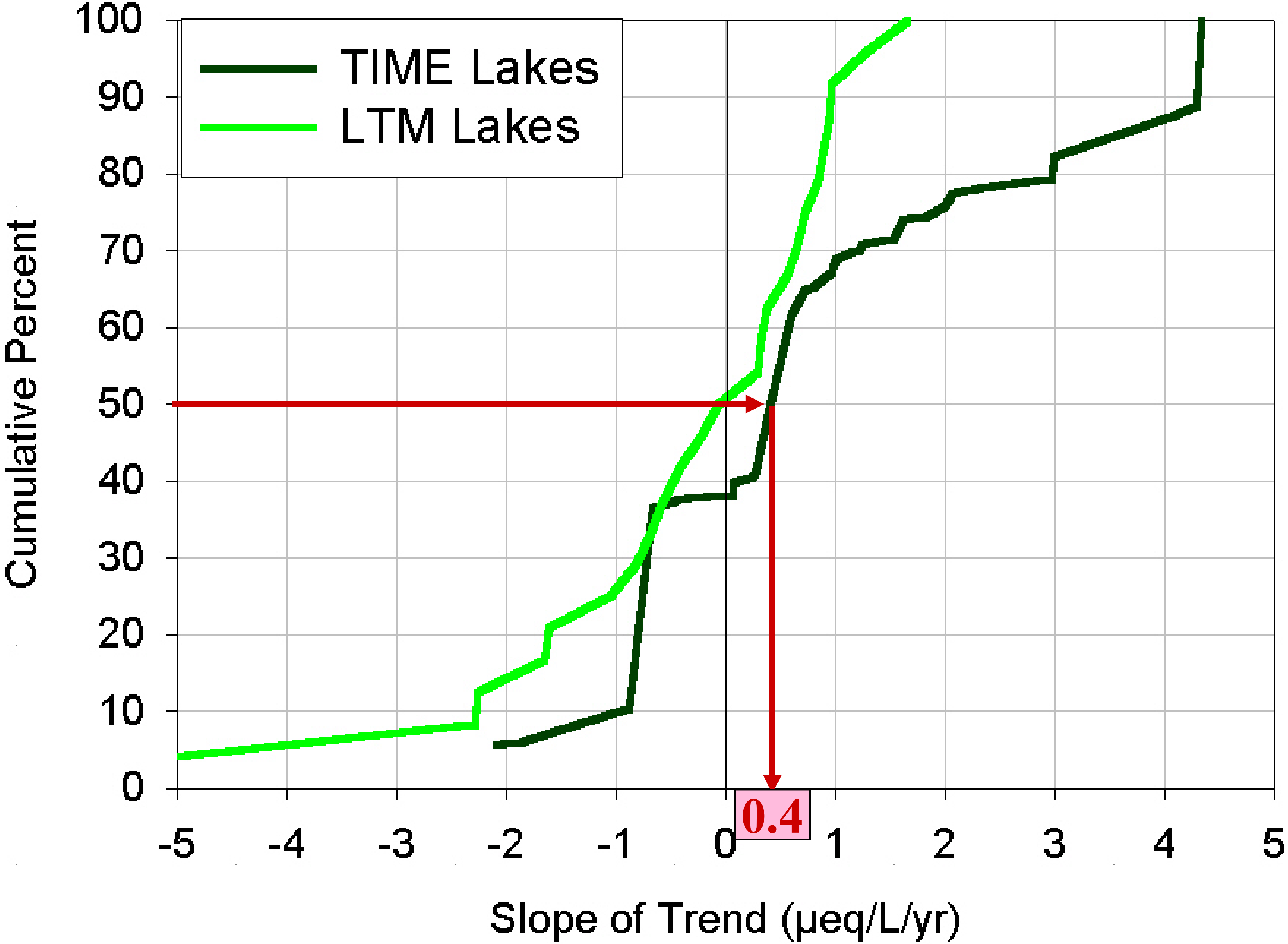
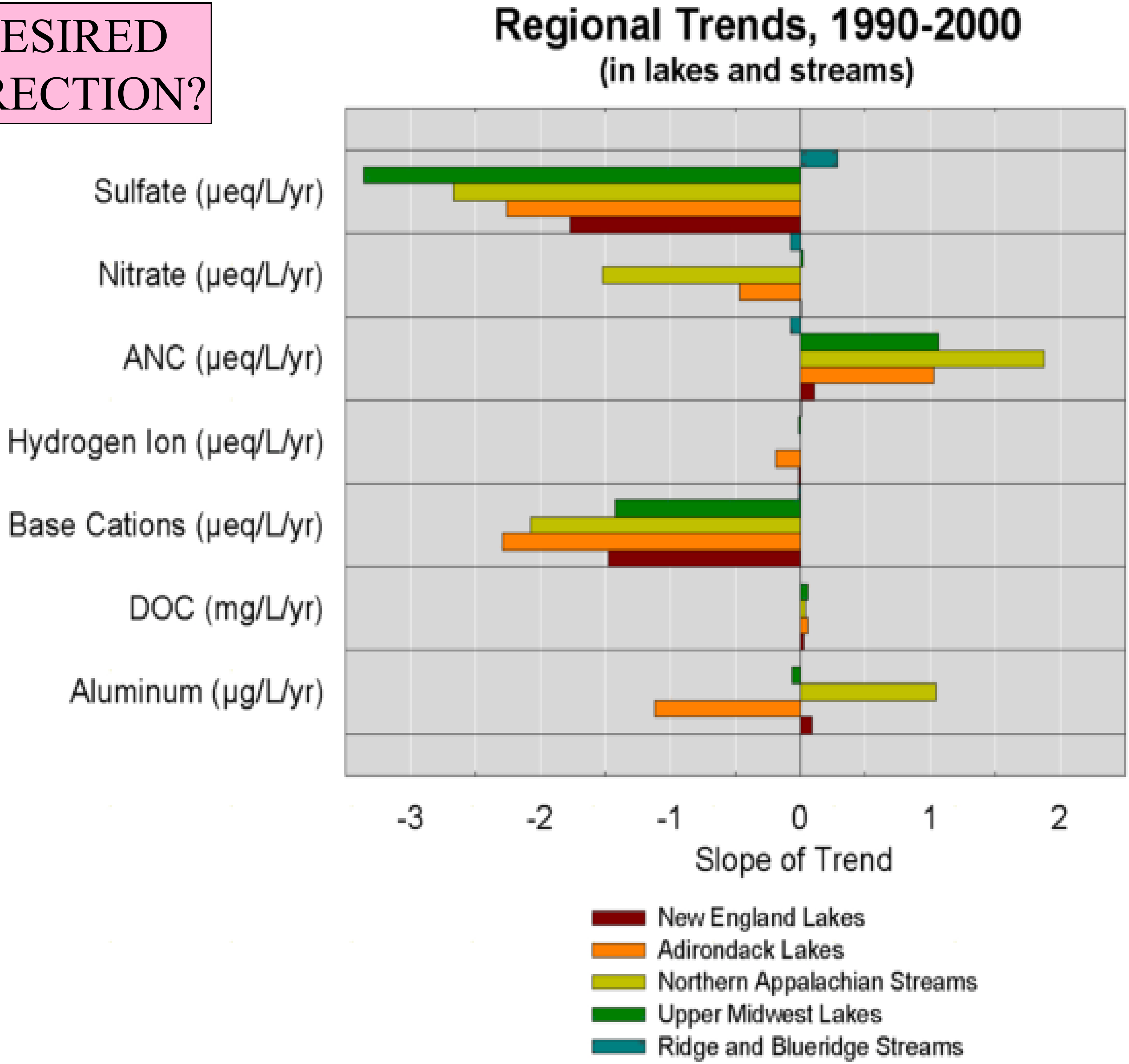


Table 7. Regional trend results for populations of sites in acid sensitive regions. Results from TIME probability sites are extrapolated to regional target populations.

Region	SO_4^{2-}	NO_3^-	Base Cations	Gran ANC	Hydrogen	DOC	Alumin
Adirondack Lakes	-2.10**	+0.01^{ns}	-1.22*	+0.56*	-0.09^{ns}	+0.09*	+0.66^{ns}
New England Lakes	-1.88**	+0.02^{ns}	-1.57**	+0.40*	+0.01^{ns}	+0.08*	-1.94^{ns}
Appalachian Streams	-0.64*	+0.04^{ns}	-0.32^{ns}	+0.34*	-0.01^{ns}	+0.01^{ns}	+0.14^{ns}

DESIRED DIRECTION?

+
+
+
+
-
-
+



IMPORTANT FEATURES OF THIS SET OF DATA & ANALYSES

- **This is the most extensive set of data that exists to look at acidification of surface waters in the WORLD**
  - **A few Scandinavian surveys have similar coverage, but are much less spatially extensive.**
- **Very important feature:**
  - **Probability samples of known populations, and**
  - **Long term data at convenience collection of sites.**
  - **Need to combine results from such data sources**
- **Response is TREND, not response size**

IMPORTANT FEATURES OF THIS SET OF DATA/ ANALYSES

continued

- **Trends are summarized in estimated cumulative distribution functions (cdf).**
  - **Measurement error spreads observed cdf out from its underlying true value**
    - **"Deconvolution" has been attempted for such problems, but**
      - assuming each observation has the same "error" distribution
    - **Here, slopes at different sites have different variances, even if "measurement error" has the same distribution, because and different series are observed at different times.**

$$\operatorname{var}(\hat{\beta}) = \frac{\sigma^2}{\sum (t_i - \bar{t})^2}$$

IMPORTANT FEATURES OF THIS SET OF DATA/ ANALYSES - NOT DONE

- No Spatial analysis – relation to objectives?
- Temporal analysis limited to linear regression
- No spatial-temporal analyses
- No combination of
  - Probability-selected sites with
  - Convenience-selected sites (Convenience = purosefully selected)
- No recognition of unequal var(slopes)
- Why not?
  - Our tools range from handcrafted solutions requiring substantial knowledge to apply them, to
  - Non-existent tools

DESIGN BASED *vs* MODEL BASED

- Spatial statistics is "model based" in that all inferences are made through the model, so validity of results rests on the model used.
- RISK: Biased data going into a model-based analysis will produce a biased analysis.
- Legitimate concern?
  - For environmental data, I think so.
  - I have several examples showing such problems!
    - Talk to me if you want to see the one I have along.
  - How do we address this in model based analyses?

STARMAP IS ?

- Space-Time Aquatic Resources Modeling and Analysis Program
- EPA funded
- Companion Program at Oregon State University
  - Focused on design-based perspectives
- Associated Center at the University of Chicago
  - Michael Stein is director of that Center

SUMMARY OF BACKGROUND for STARMAP

- Probability-based surveys of aquatic resources have a role and will be implemented
- Important associated questions
  - How should we combine
    - Probability survey data with
    - Data from purposefully picked sites?
  - How can we incorporate remotely sensed information (satellite) with ground data?
  - Role of landscape data (GIS) is?
  - How can we make accurate predictions of water quality at unvisited sites, using all of above?

STARMAP'S MAJOR OBJECTIVES

- To advance the science of statistics to address questions such as
  - Spatial and temporal modeling relevant to aquatic monitoring
  - Adapt Bayesian methods to needs of aquatic monitoring
  - Develop allied small area estimation methods
  - Integrate the above with techniques of hierarchical survey design and allied techniques
- To develop and extend the expertise on design and analysis to the states and tribes
- To train future generations of environmental statisticians

STARMAP'S VISION

- **PERSPECTIVE:**
  - **A searching analysis of a real, moderately complex, data set almost always generates questions whose answer calls for an extension of existing statistical theory or methodology.**

STARMAP PROJECTS

- COMBINING ENVIRONMENTAL DATA SETS - JENNIFER HOETING
- LOCAL ESTIMATION - JAY BREIDT
- INDICATOR DEVELOPMENT - DAVE THEOBALD (CSU'S Natural Resources Ecology Lab)
- OUTREACH - SCOTT URQUHART

THANK YOU FOR YOUR ATTENTION

QUESTIONS and/or COMMENTS ARE WELCOME