USGS
science for a changing world

A new approach to analysis of large and long-duration surface water quality data sets

Robert M. Hirsch,
Research Hydrologist, USGS

April 27, 2010

U.S. Department of the Interior
U.S. Geological Survey

Motivations

- Larger and longer data sets
- Desire to know: How successful are our control efforts?
- Desire to **understand** and report
- Make use of new statistical ideas about exploratory data analysis

USGS

Characteristics of new method

- Focus on change description not hypothesis testing
- Concentration versus discharge relationship must be free to change shape over time
- Seasonal patterns must be free to change shape over time

USGS

Characteristics of new method

- Trend pattern is free form: not linear or quadratic – new data shouldn't influence old estimates
- Report results for concentration and for flux (from one unified approach)

USGS

Characteristics of new method

- Report the actual history & results that are free of the particular flow history that took place
- Method should compute trends **and** provide diagnostic information

USGS

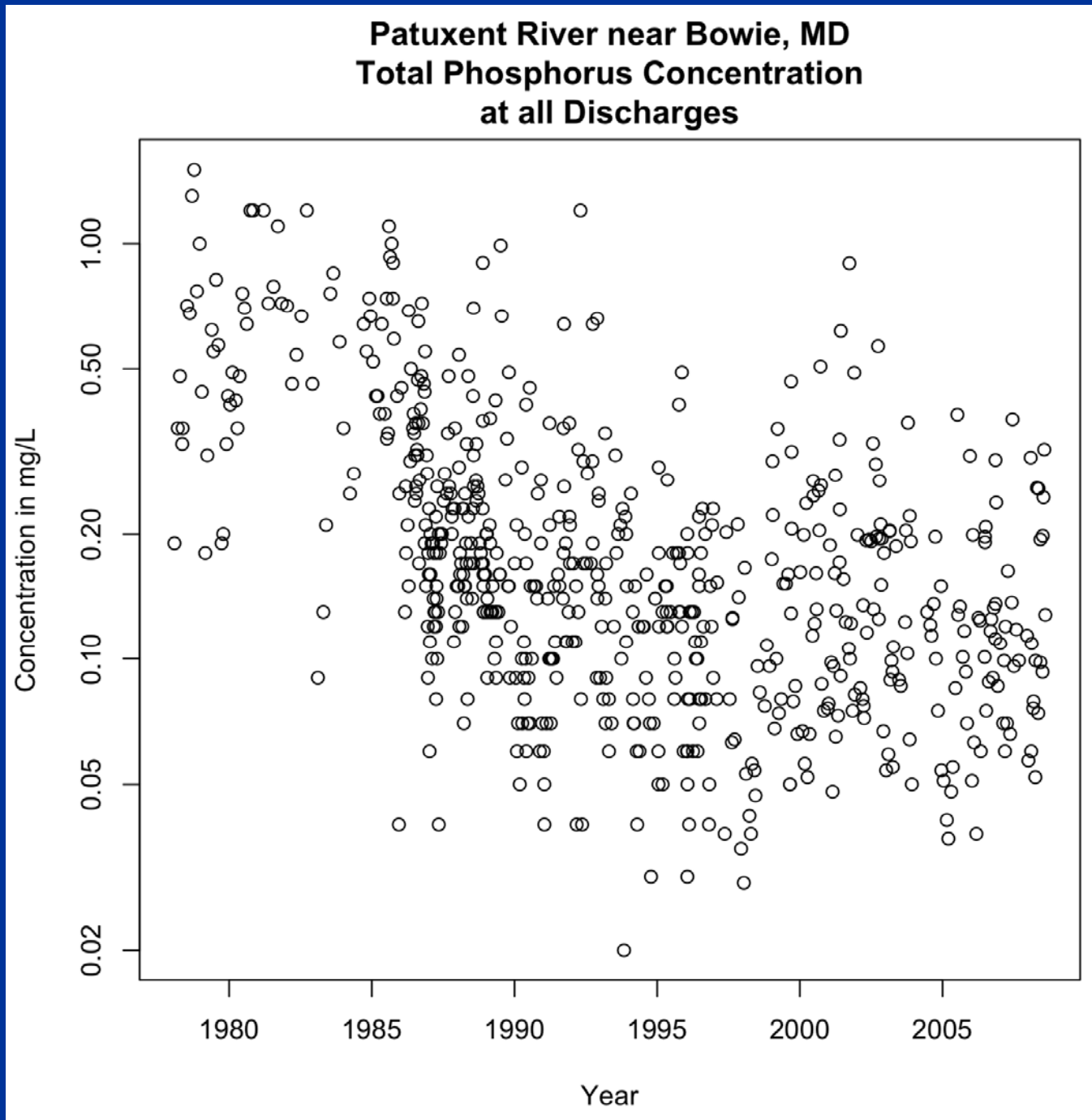
Lets look at a data set: Patuxtent River, Maryland Total Phosphorus

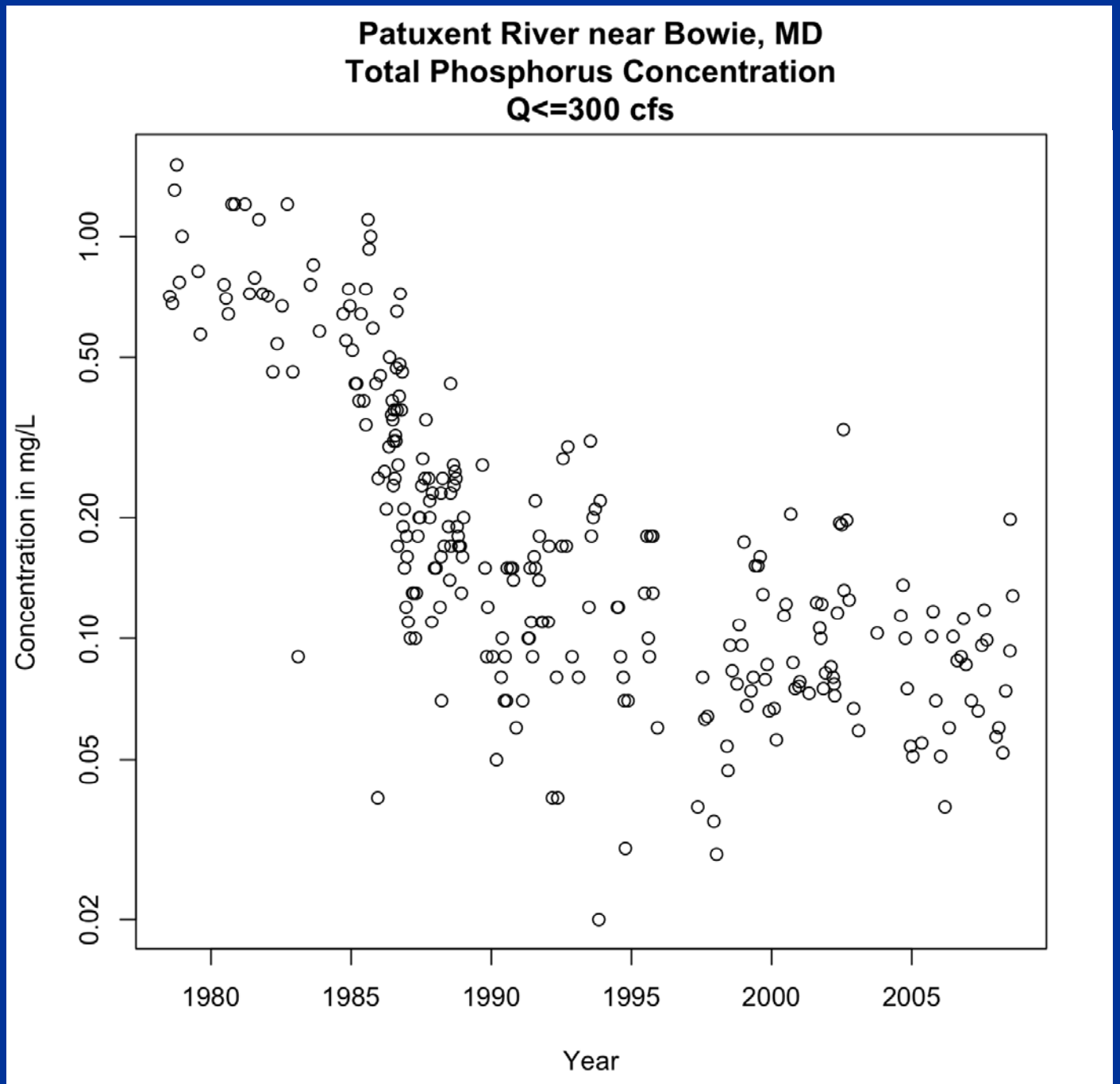
1978-2008

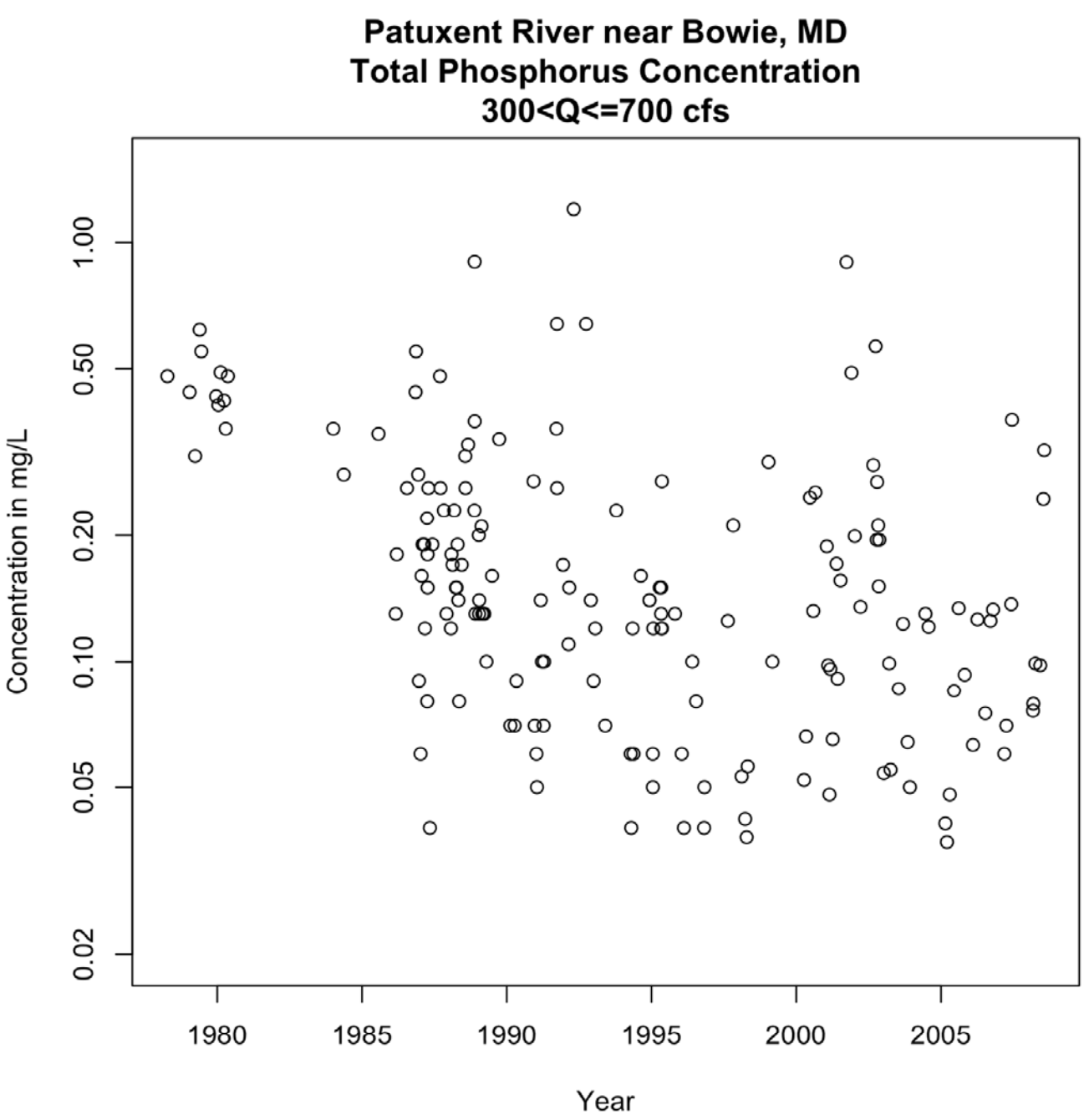
737 Observations

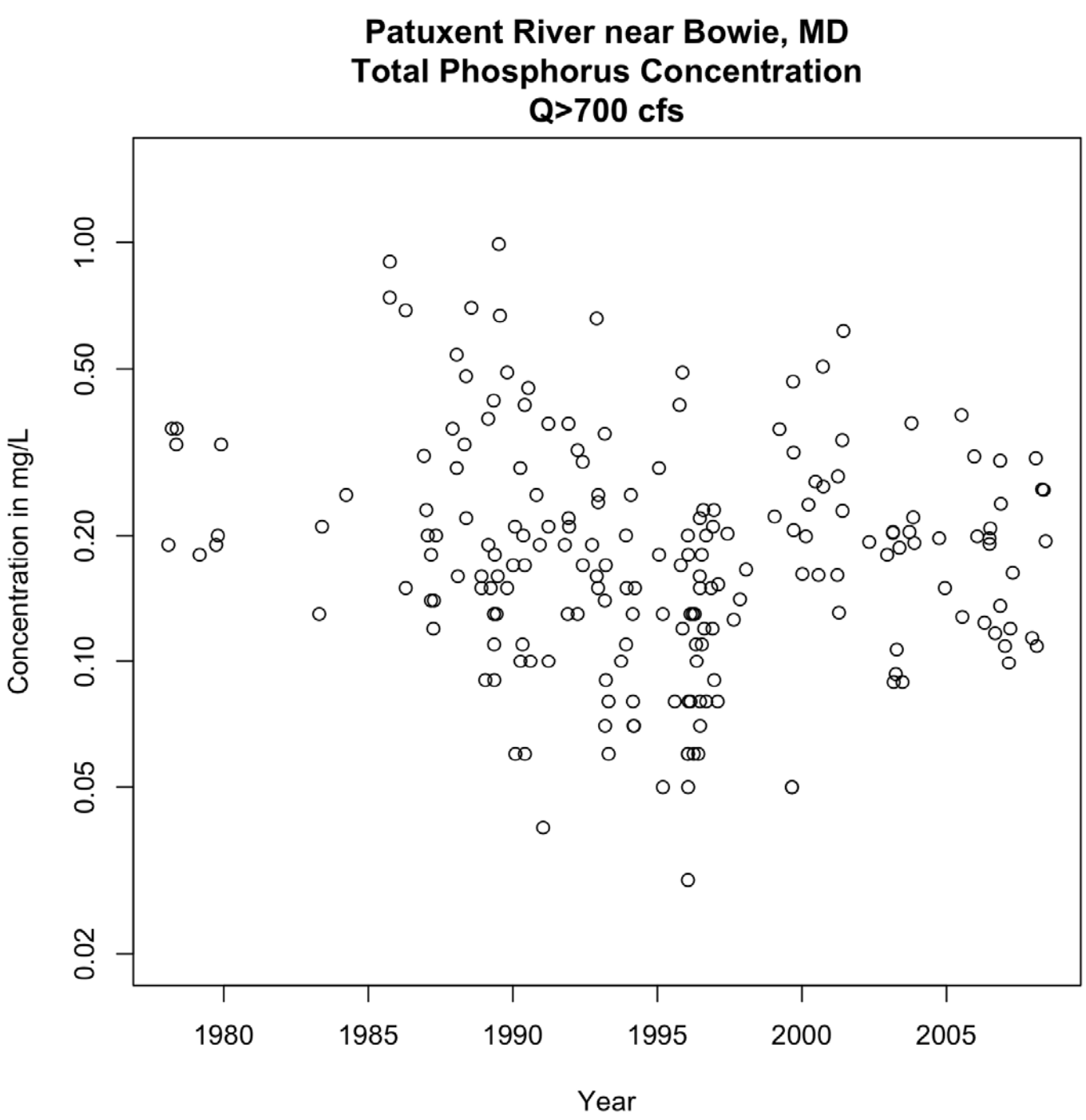
Huge reductions in point sources

USGS


Patuxent River near Bowie, MD
Total Phosphorus Concentration
at all Discharges
Concentration in mg/L
0.02
0.05
0.10
0.20
0.50
1.00
1980
1985
1990
1995
2000
2005
Year


Patuxent River near Bowie, MD
Total Phosphorus Concentration
Q<=300 cfs
Concentration in mg/L
0.02
0.05
0.10
0.20
0.50
1.00
1980
1985
1990
1995
2000
2005
Year


Patuxent River near Bowie, MD
Total Phosphorus Concentration
300<Q<=700 cfs
Concentration in mg/L
0.02
0.05
0.10
0.20
0.50
1.00
1980
1985
1990
1995
2000
2005
Year


Patuxent River near Bowie, MD
Total Phosphorus Concentration
Q>700 cfs
Concentration in mg/L
0.02
0.05
0.10
0.20
0.50
1.00
1980
1985
1990
1995
2000
2005
Year

What do we know?

- C vs Q relationship changes over time
- Trend is large in early years, rather flat in later years
- We would like to know consequences for concentration and also for flux
- We need a flexible model of this behavior

USGS

The specific method being implemented is called:

Weighted Regressions on Trend, Discharge, and Season (WRTDS)

It is a "smoother" that builds a flexible statistical model to estimate concentration for every day in the period of record.

USGS

It decomposes the record into

- Seasonal component
- Discharge component
- Trend component
- Random component

Then makes estimates of concentration and flux for every day of the period of record

USGS

Locally Weighted Regression

For any location in time - discharge space (*t* and *Q*) we assume that concentration (*c*) follows this model

$$\ln(c) = \beta_0 + \beta_1 \bullet t + \beta_2 \bullet \ln(Q) + \beta_3 \bullet \sin(2\pi t) + \beta_4 \cos(2\pi t) + \varepsilon$$

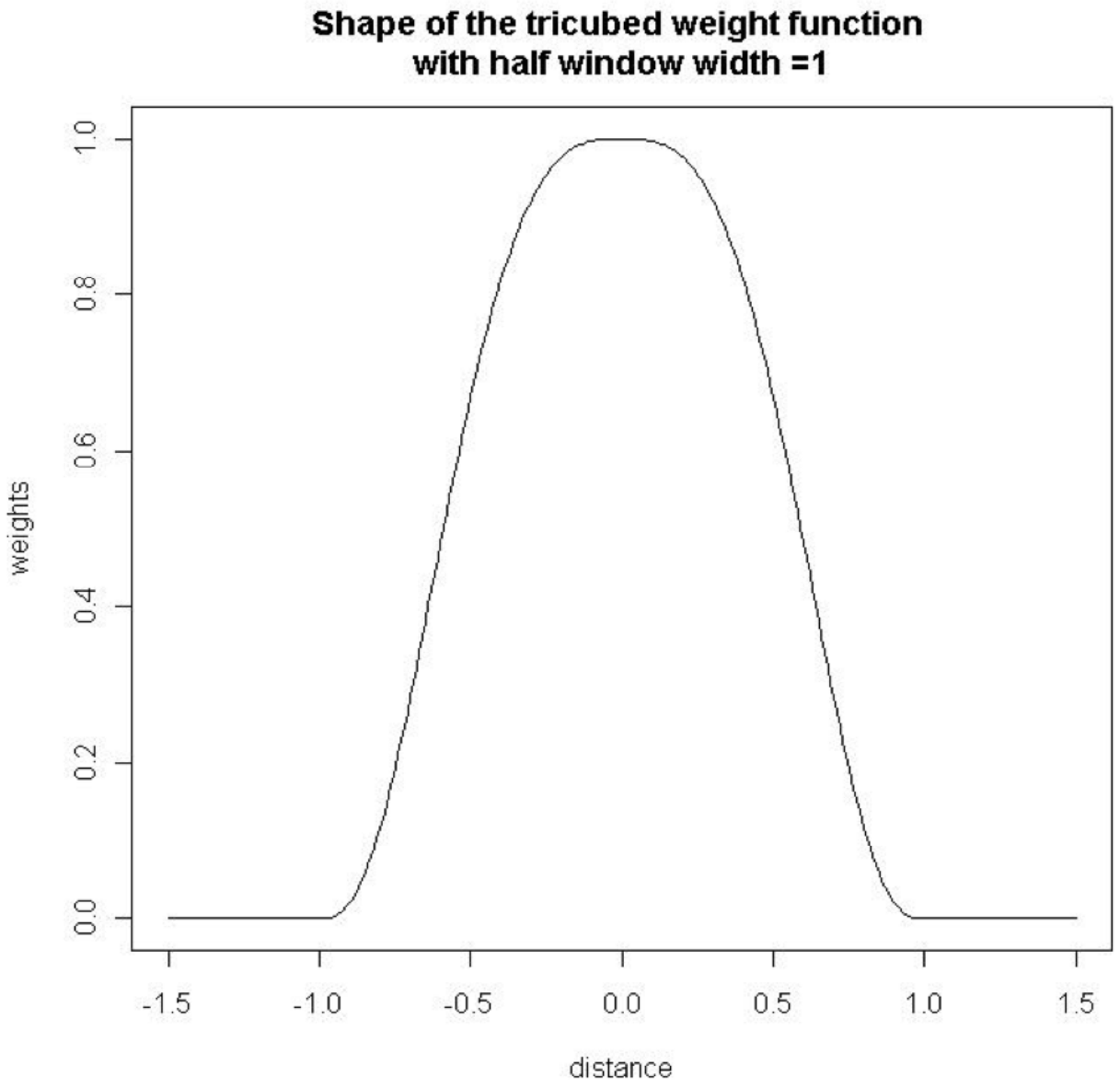
But the coefficients should be smoothly changing as we move through the space

We estimate the coefficients at any location in this space using a weighted regression model, where the weights are based on "relevance" to the location of interest

Measures of relevance

- Time in years
- Time in seasons
- Discharge (as ln(Q))

Use Tukey's tri-cubed weight function



Set up the 3 windows

- **Time: half-window width is 10 years**
- **Seasons: half-window width is 0.5 years**
- **Discharge: half-window width is 2 natural log units (a range of 0.13 to 7.4 times the discharge of interest)**

USGS

What's the process?

- For any "place" in *Q* vs *t* space, look at every one of the 737 observations
- Compute a time weight, a season weight, and a discharge weight
- The overall weight is the product of these 3 weights

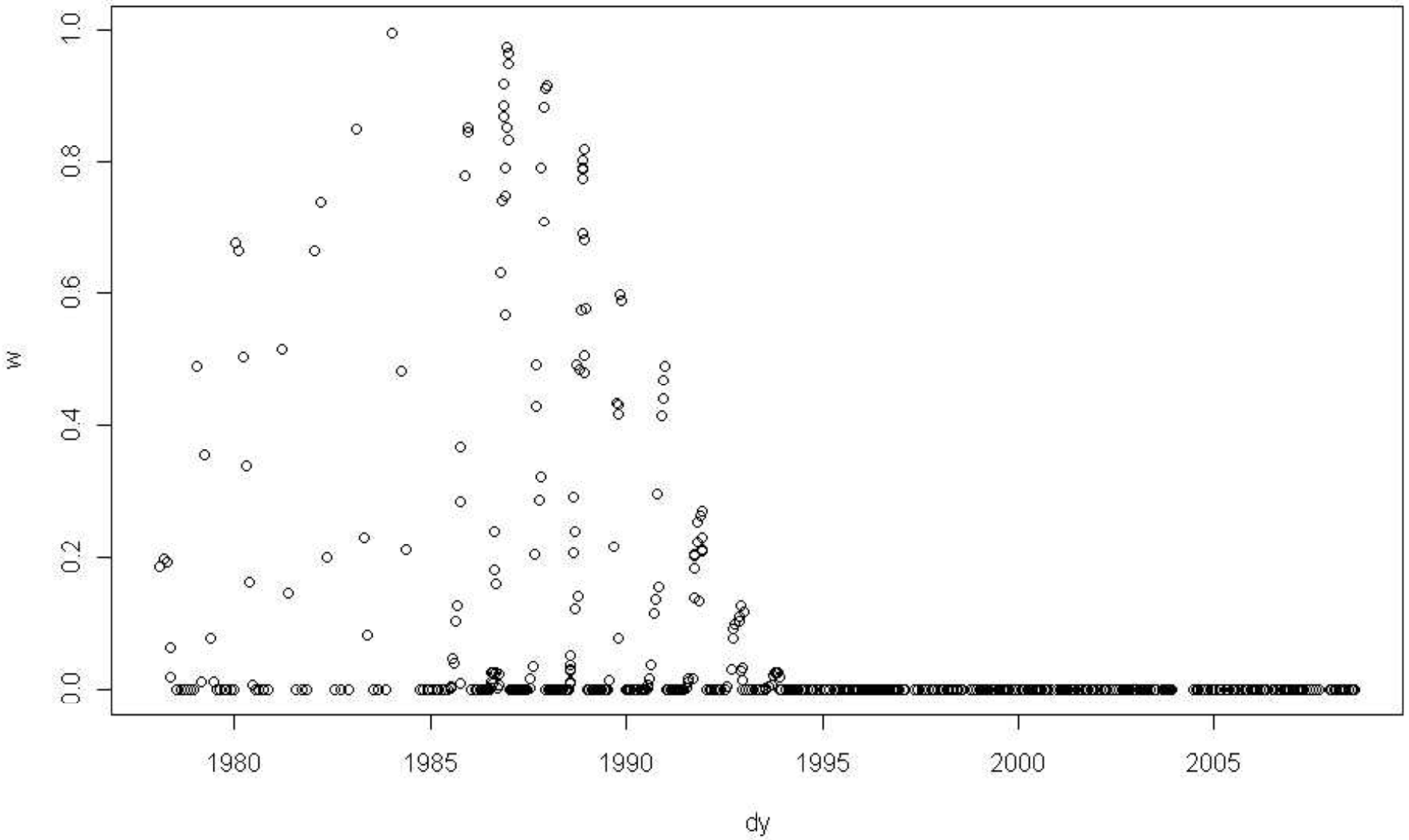
USGS

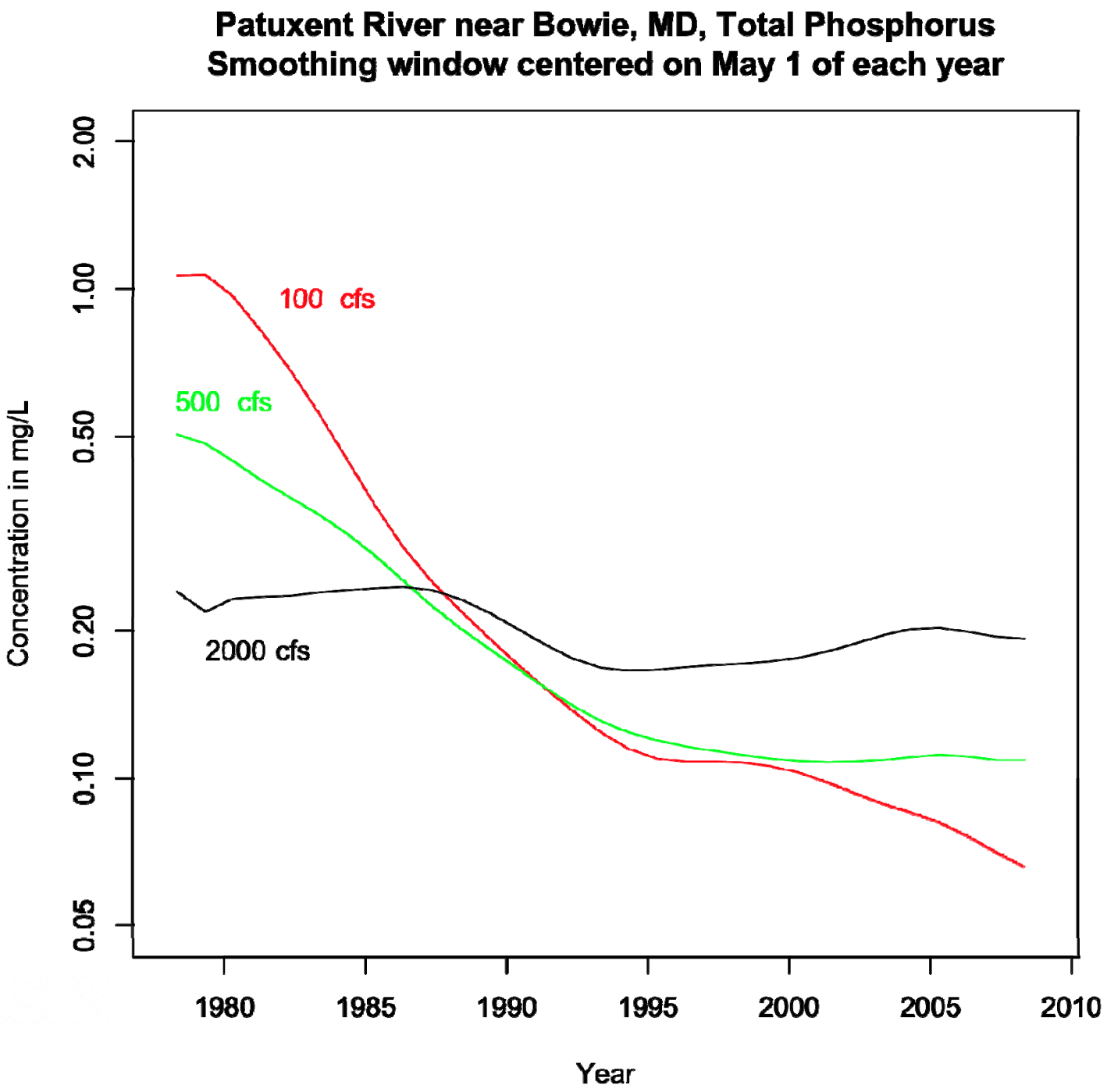
What's the process?

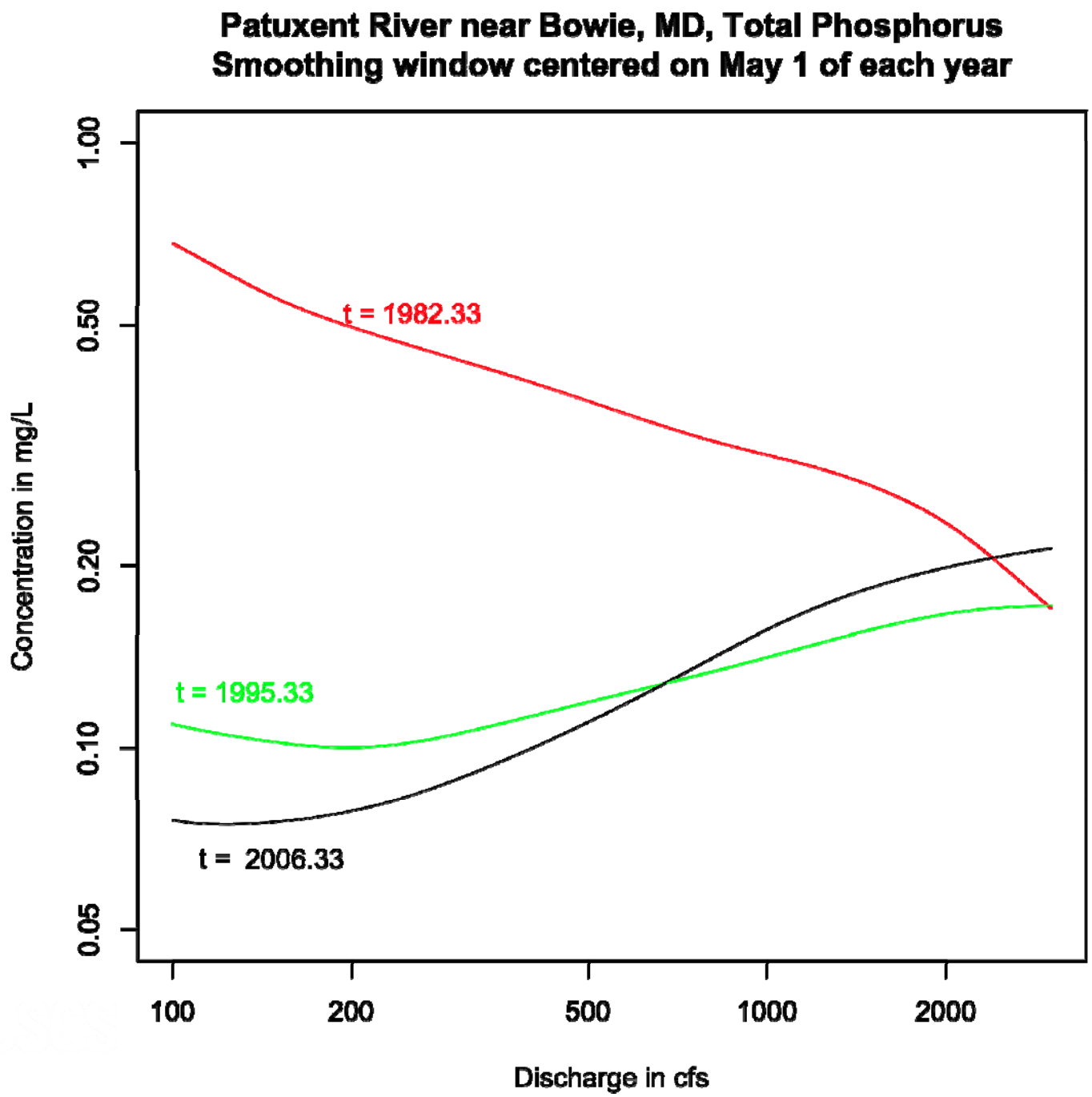
- Now run a weighted regression
- Transform the results from ln(*c*) to *c* adding in a bias correction (based on the weighted residuals)
- Use it to estimate concentration for any combination of discharge and time

USGS

An example: weights for 1/1/1985, Q of 372 cfs, the long term median




Patuxent River near Bowie, MD, Total Phosphorus
Smoothing window centered on May 1 of each year
Concentration in mg/L
2.00
1.00
0.50
0.20
0.10
0.05
100 cfs
500 cfs
2000 cfs
1980
1985
1990
1995
2000
2005
2010
Year

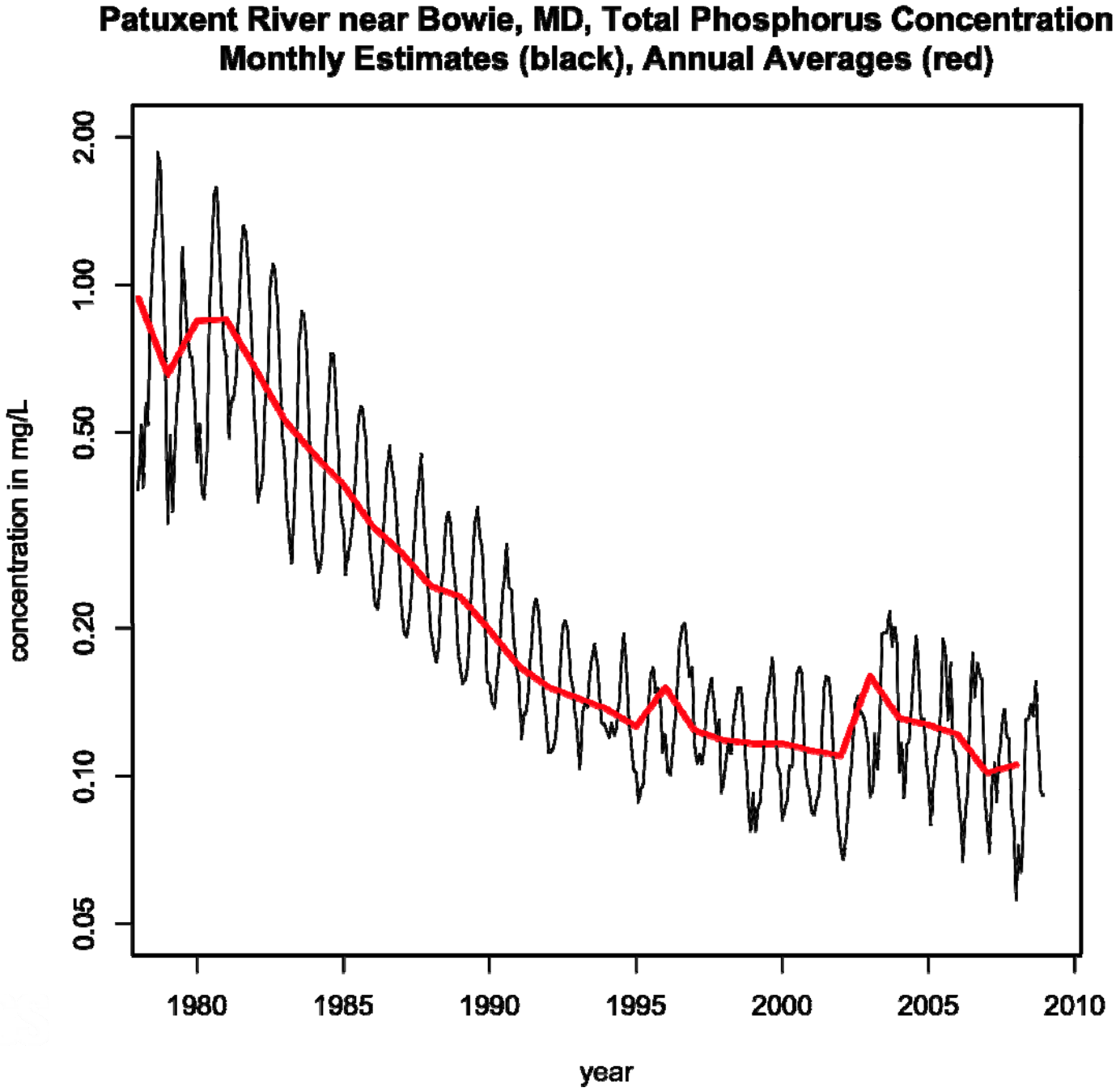

Patuxent River near Bowie, MD, Total Phosphorus
Smoothing window centered on May 1 of each year
Concentration in mg/L
1.00
0.50
0.20
0.10
0.05
t = 1982.33
t = 1995.33
t = 2006.33
100
200
500
1000
2000
Discharge in cfs

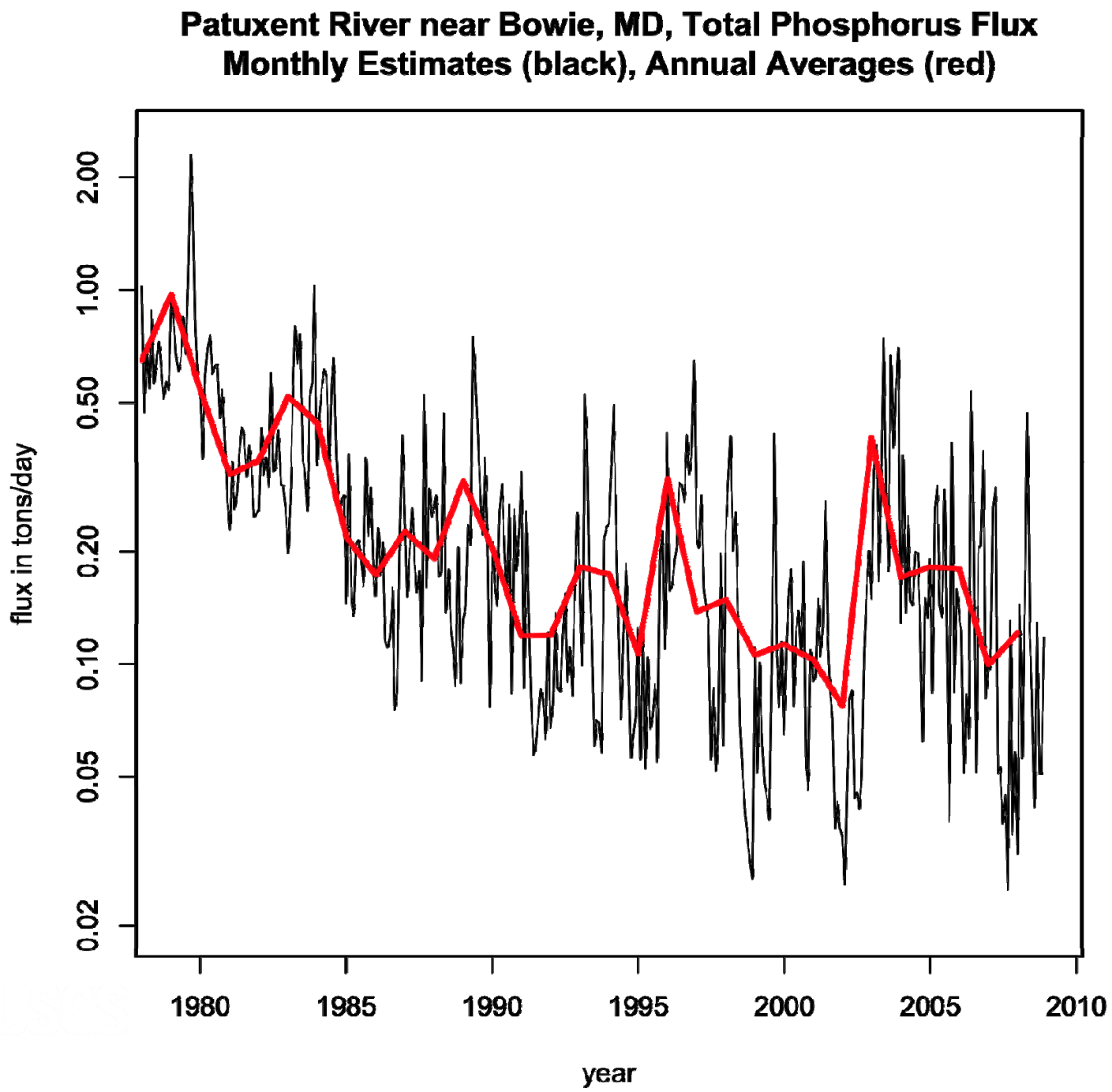
Now create an estimated concentration history

Using daily value Q from every day from Water Years 1978 – 2008

Create an estimated flux history the same way, flux = c x Q x 0.0027

USGS


Patuxent River near Bowie, MD, Total Phosphorus Concentration
Monthly Estimates (black), Annual Averages (red)
concentration in mg/L
0.05
0.10
0.20
0.50
1.00
2.00
1980
1985
1990
1995
2000
2005
2010
year


Patuxent River near Bowie, MD, Total Phosphorus Flux
Monthly Estimates (black), Annual Averages (red)
flux in tons/day
0.02
0.05
0.10
0.20
0.50
1.00
2.00
1980
1985
1990
1995
2000
2005
2010
year

These are best estimates of what actually happened

- For those focused on actual conditions in the downstream water body (bay, reservoir, lake) these results are ideal “drivers” for water quality and ecological models of that water body.
- But, we also want to talk about “performance” of the watershed

USGS

For “performance” the actual flow history is “noise”

- The drought of 1999-2002 could have happened any time,
- The high flows of 2003 could have happened any time
- These random events shouldn’t determine success
- So, lets “randomize” the flows and run them through our model

USGS

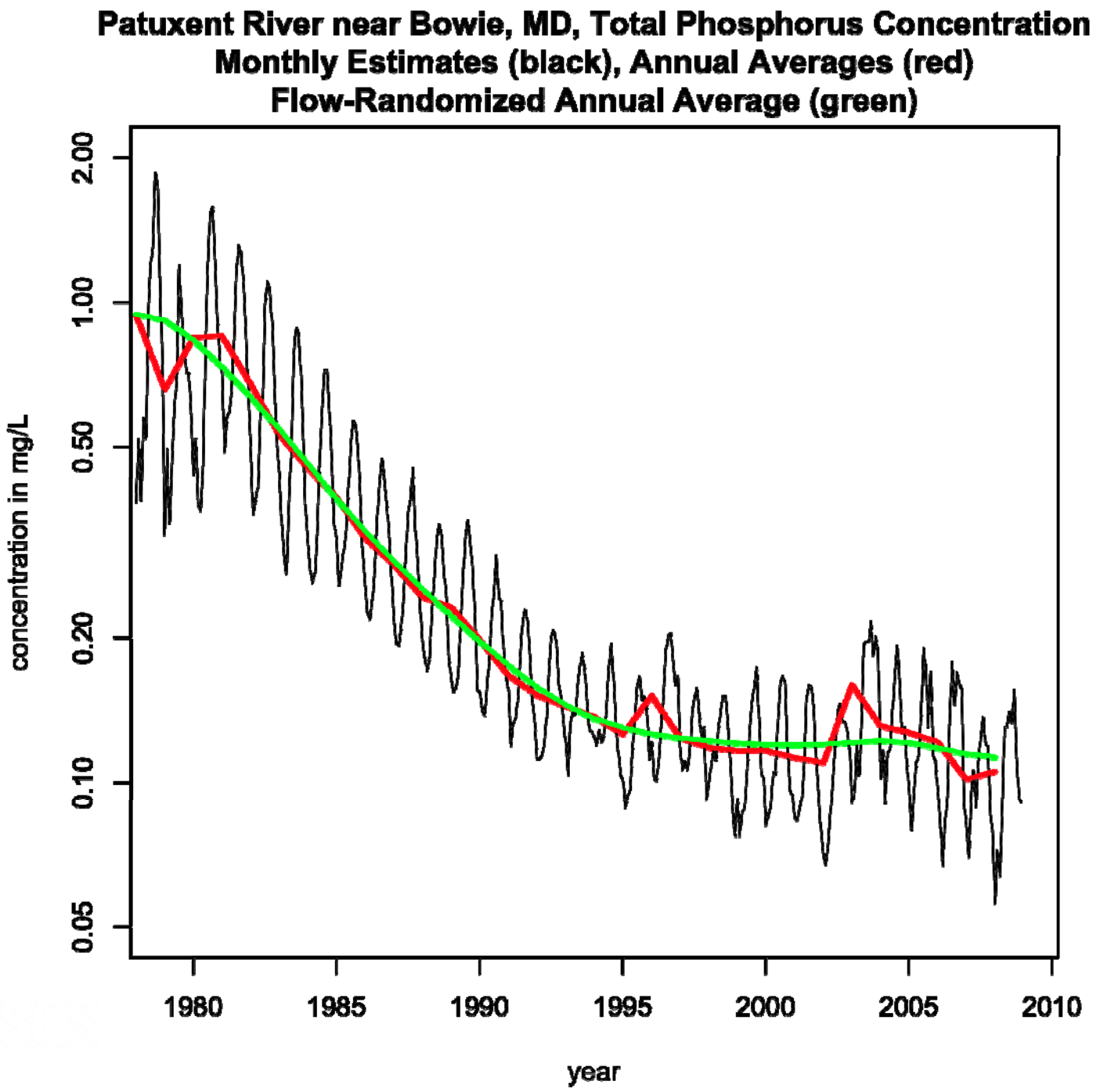
To generate a flow-random mean concentration for, let's say, April 20, 2003 we do the following:

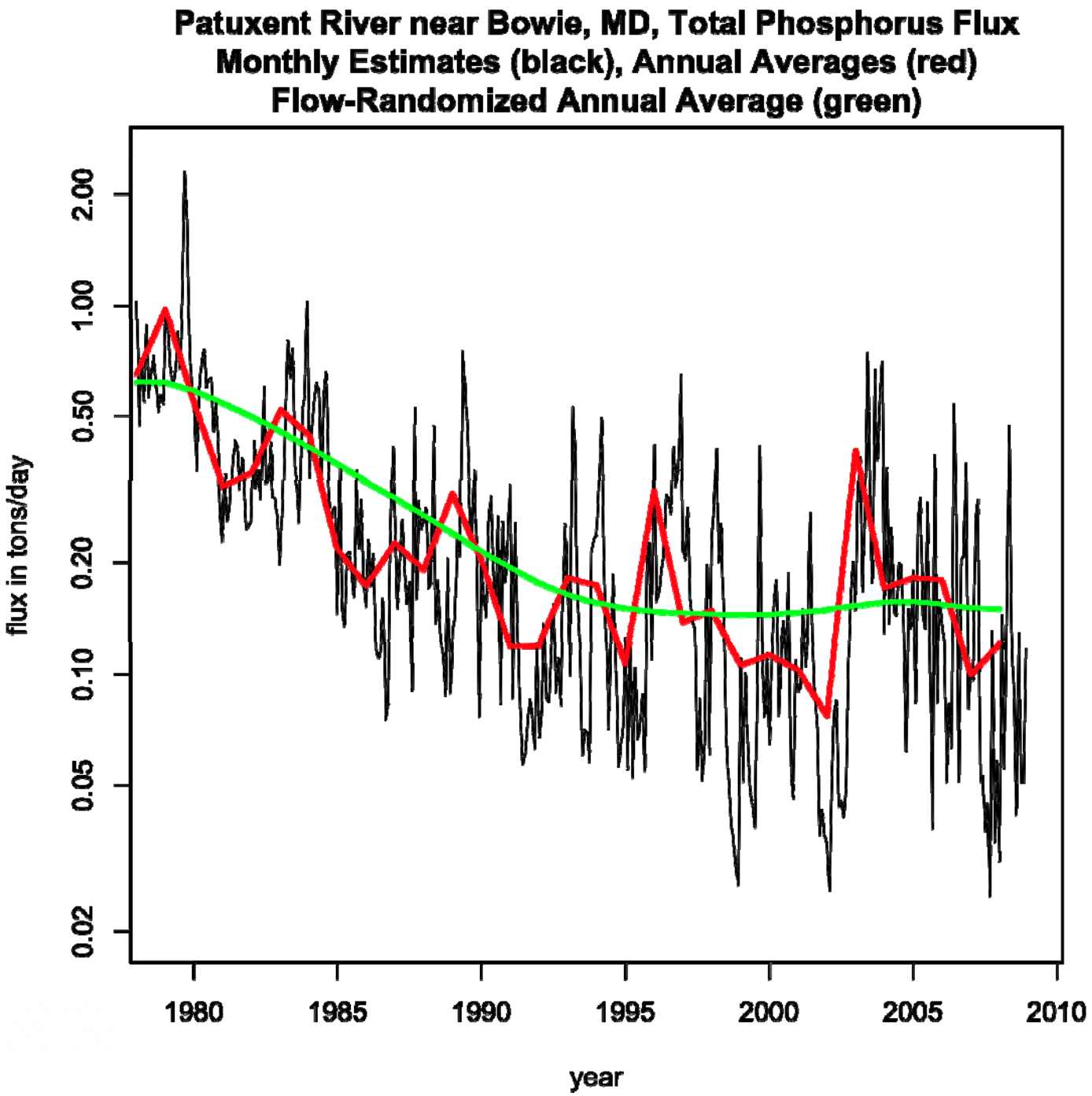
- Set t = 2003.3 and take each of the 31 different Q values for April 20 of each year from 1978-2008 and use the weighted regression method to estimate the concentration for each one.
- Take the average of these and that's the flow-randomized estimated concentration for April 20, 2003.

USGS

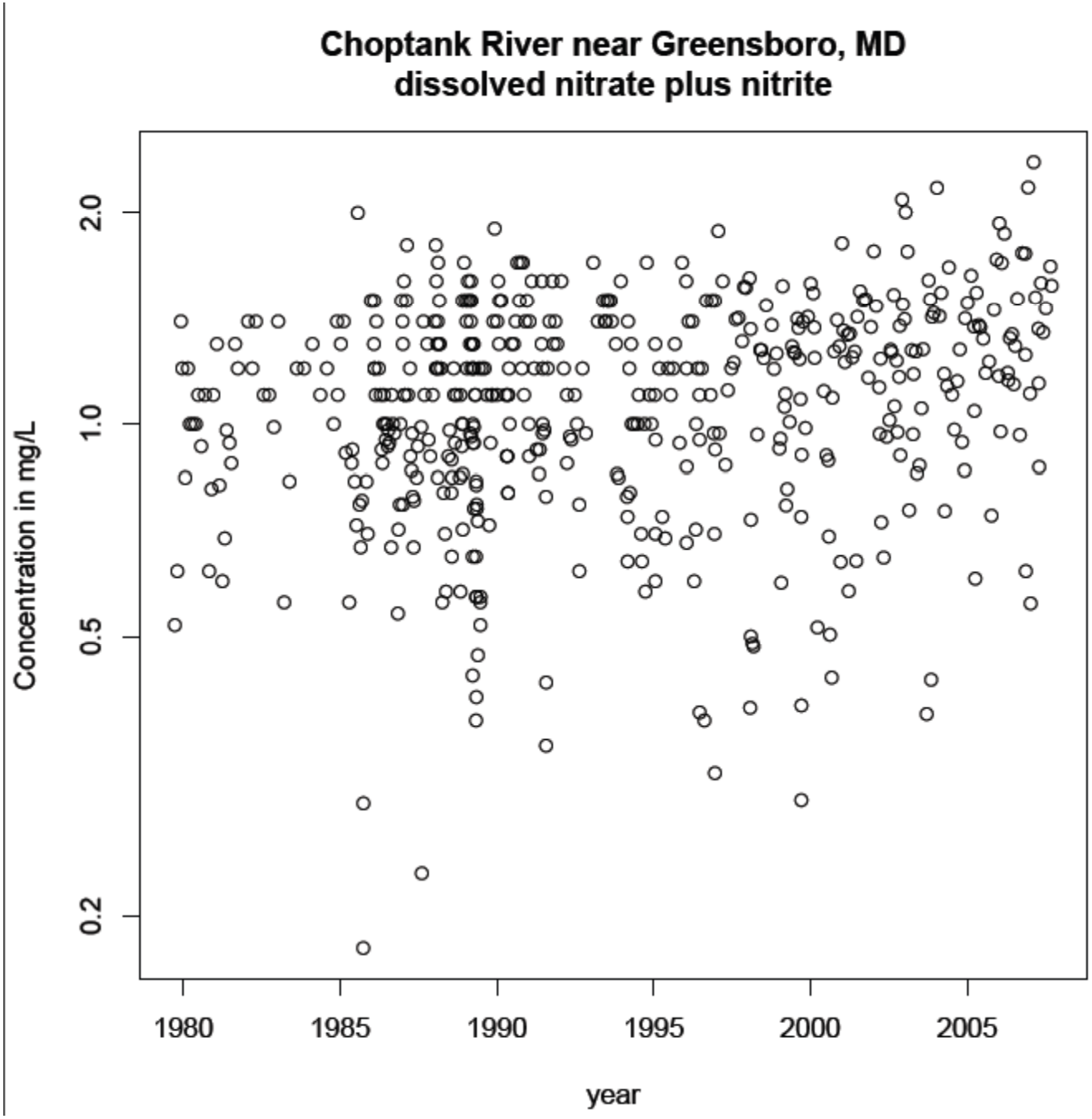
- Do that whole process again for April 21, 2003, and do it for every single day for every year in the record (in this case 11,323 days)
- We can do the same thing for flux and compute a “flow randomized average flux” for each day.
- We can then take monthly or annual averages of these flow randomized concentrations and fluxes

USGS

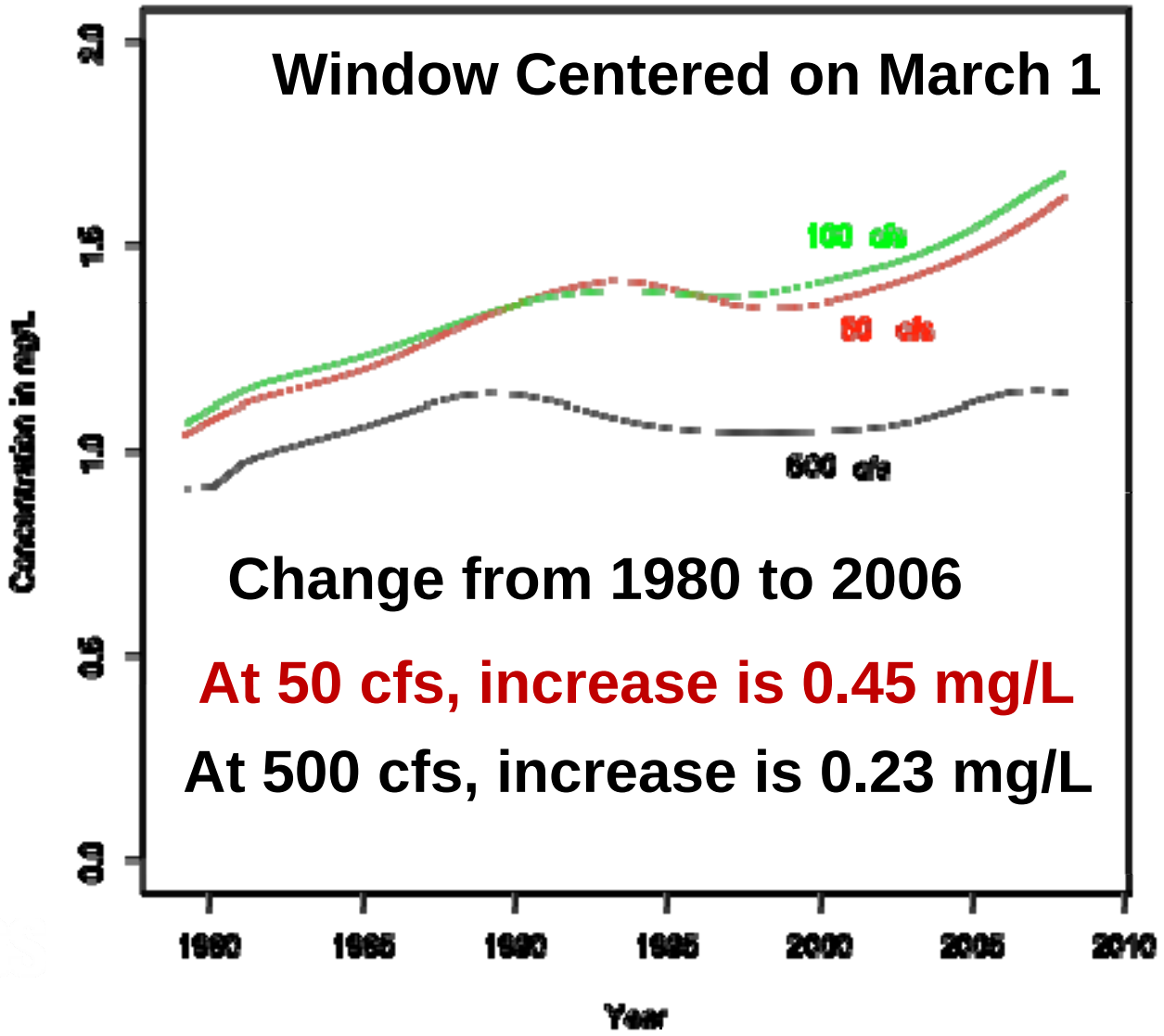

Patuxent River near Bowie, MD, Total Phosphorus Concentration
Monthly Estimates (black), Annual Averages (red)
Flow-Randomized Annual Average (green)
concentration in mg/L
2.00
1.00
0.50
0.20
0.10
0.05
1980
1985
1990
1995
2000
2005
2010
year


Patuxent River near Bowie, MD, Total Phosphorus Flux
Monthly Estimates (black), Annual Averages (red)
Flow-Randomized Annual Average (green)
flux in tons/day
0.02
0.05
0.10
0.20
0.50
1.00
2.00
1980
1985
1990
1995
2000
2005
2010
year

A quick look at a ground-water dominated stream Maryland Coastal Plain



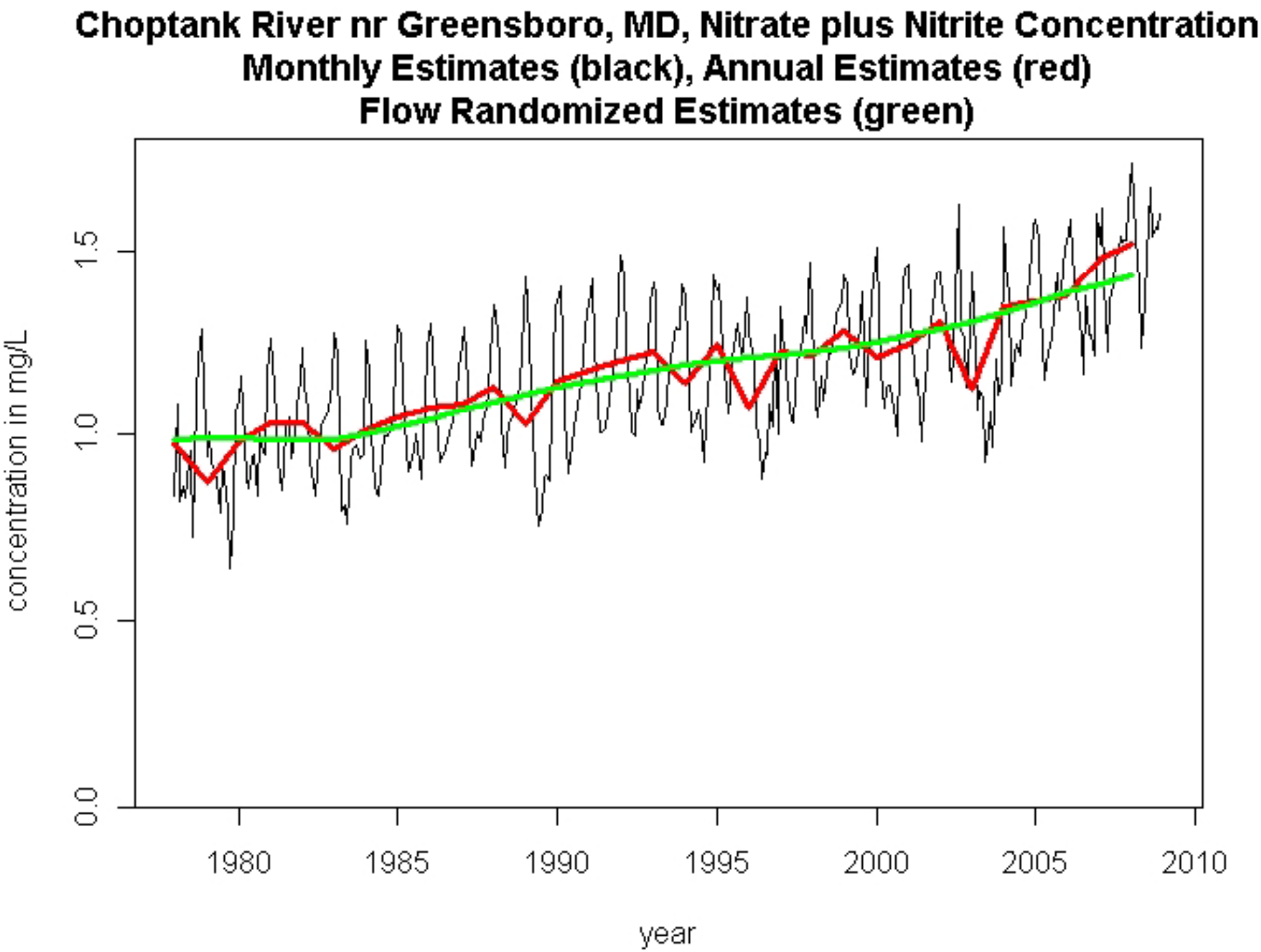
Choptank River nr Greensboro MD, Dissolved Nitrate plus Nitrite

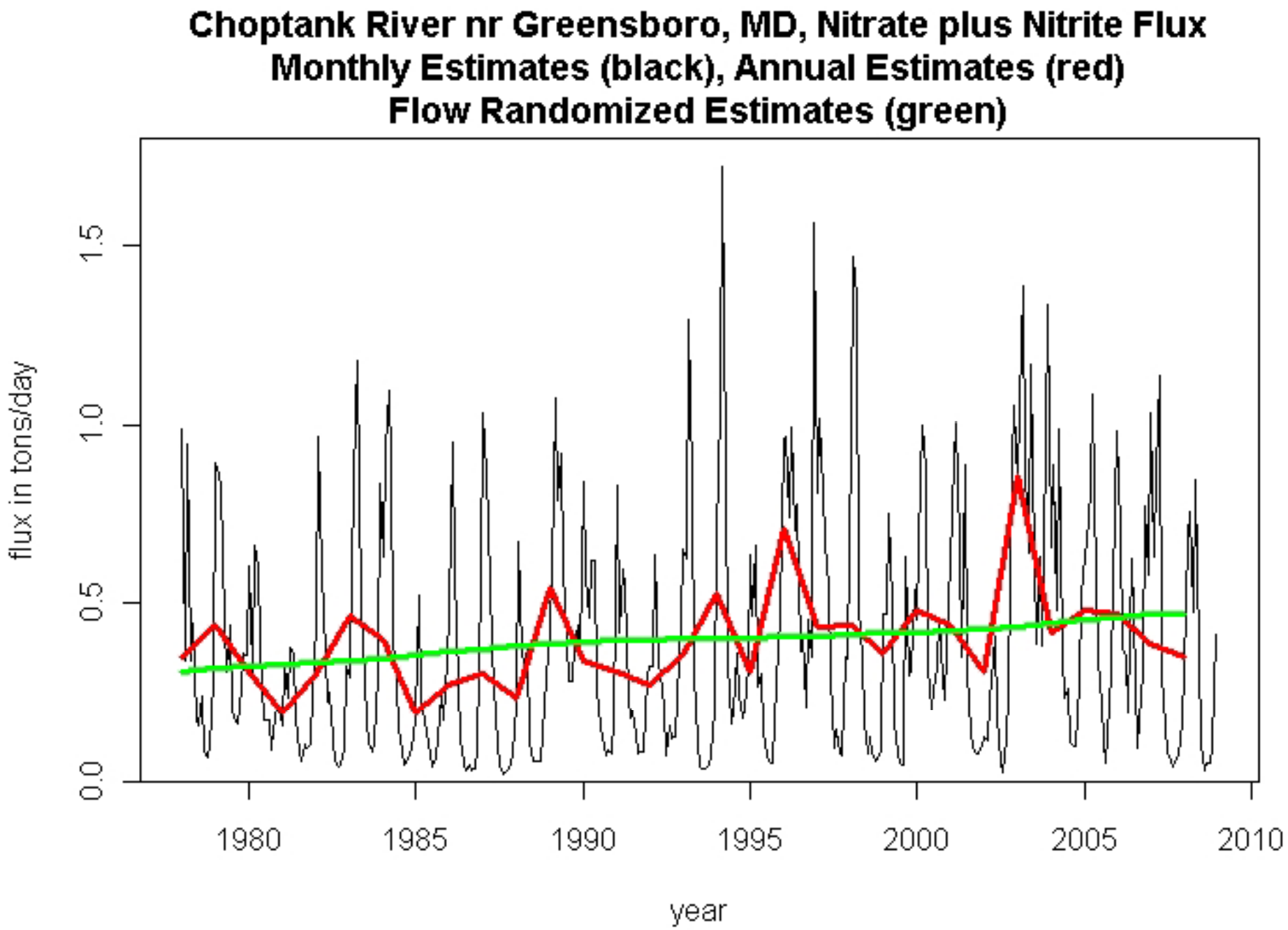


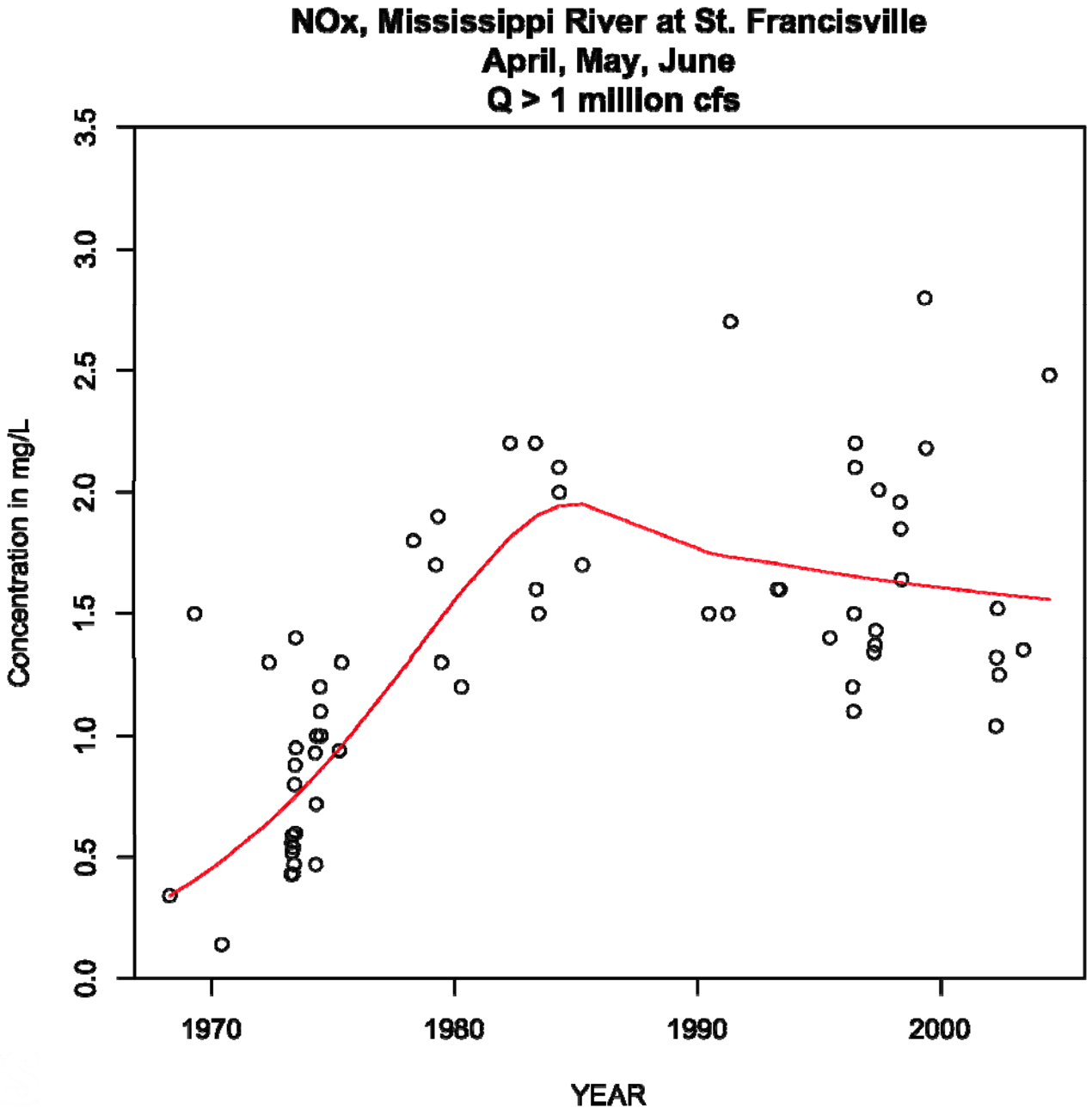
Change from 1980 to 2006

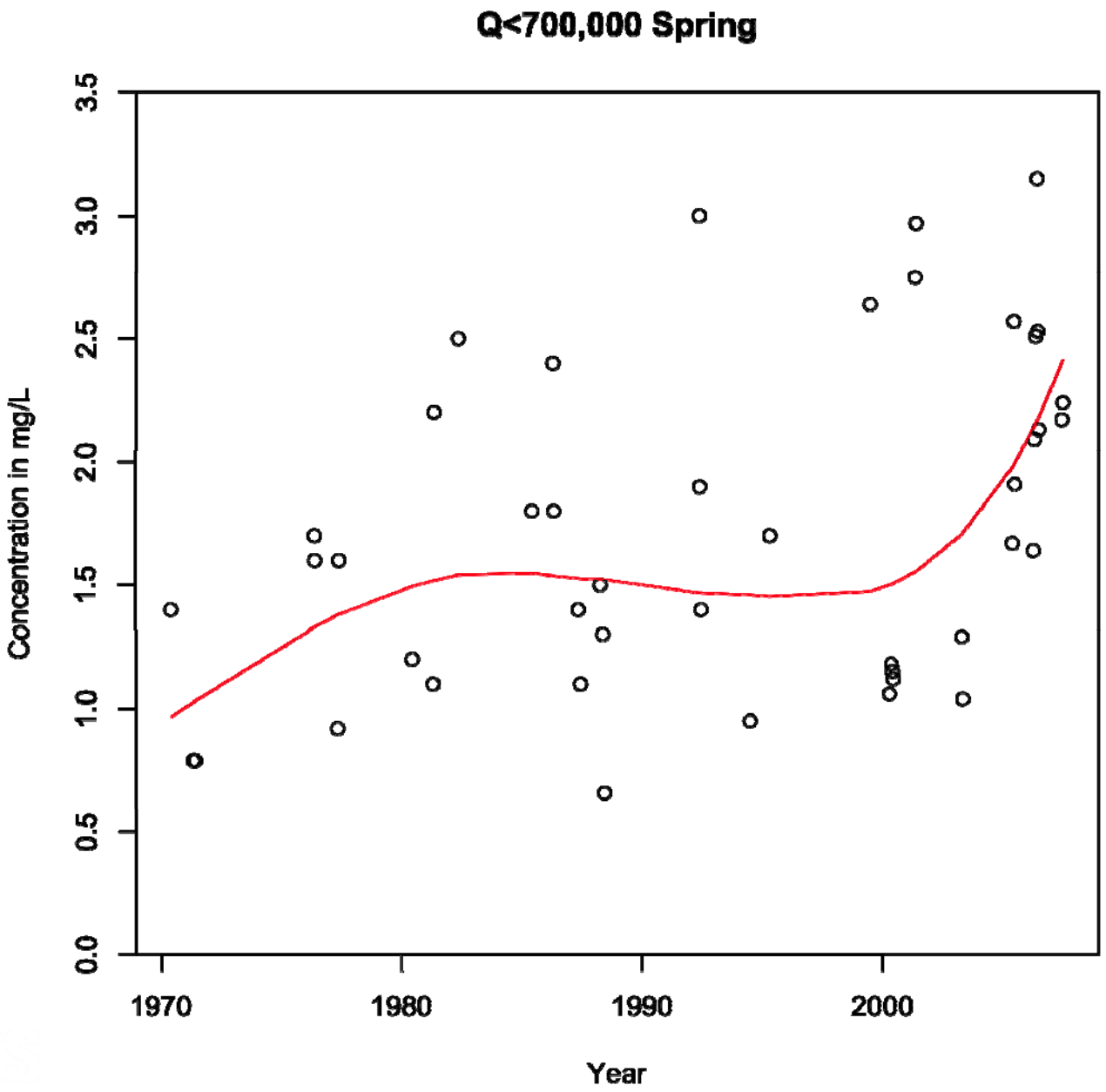
At 50 cfs, increase is 0.45 mg/L

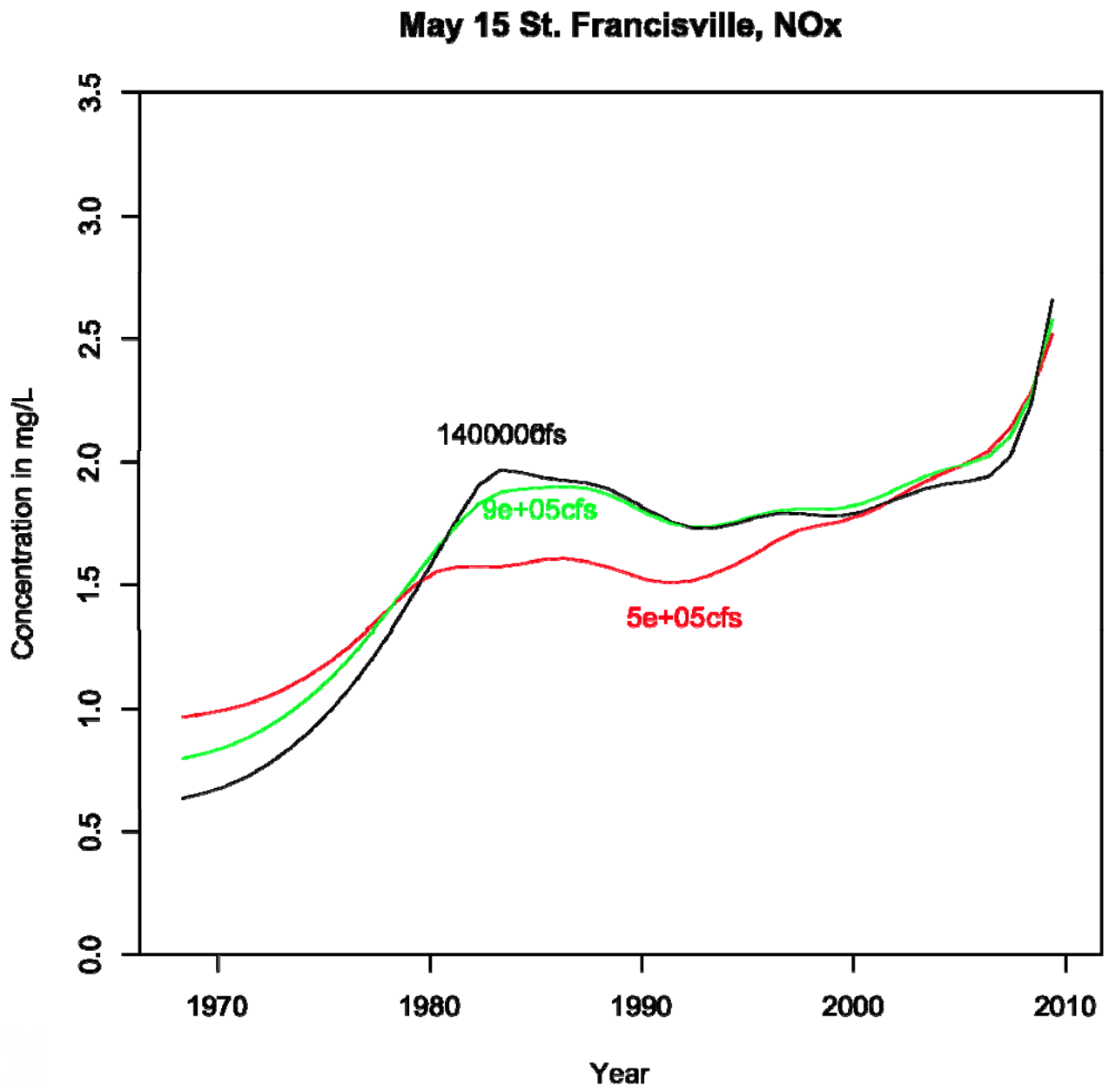
At 500 cfs, increase is 0.23 mg/L

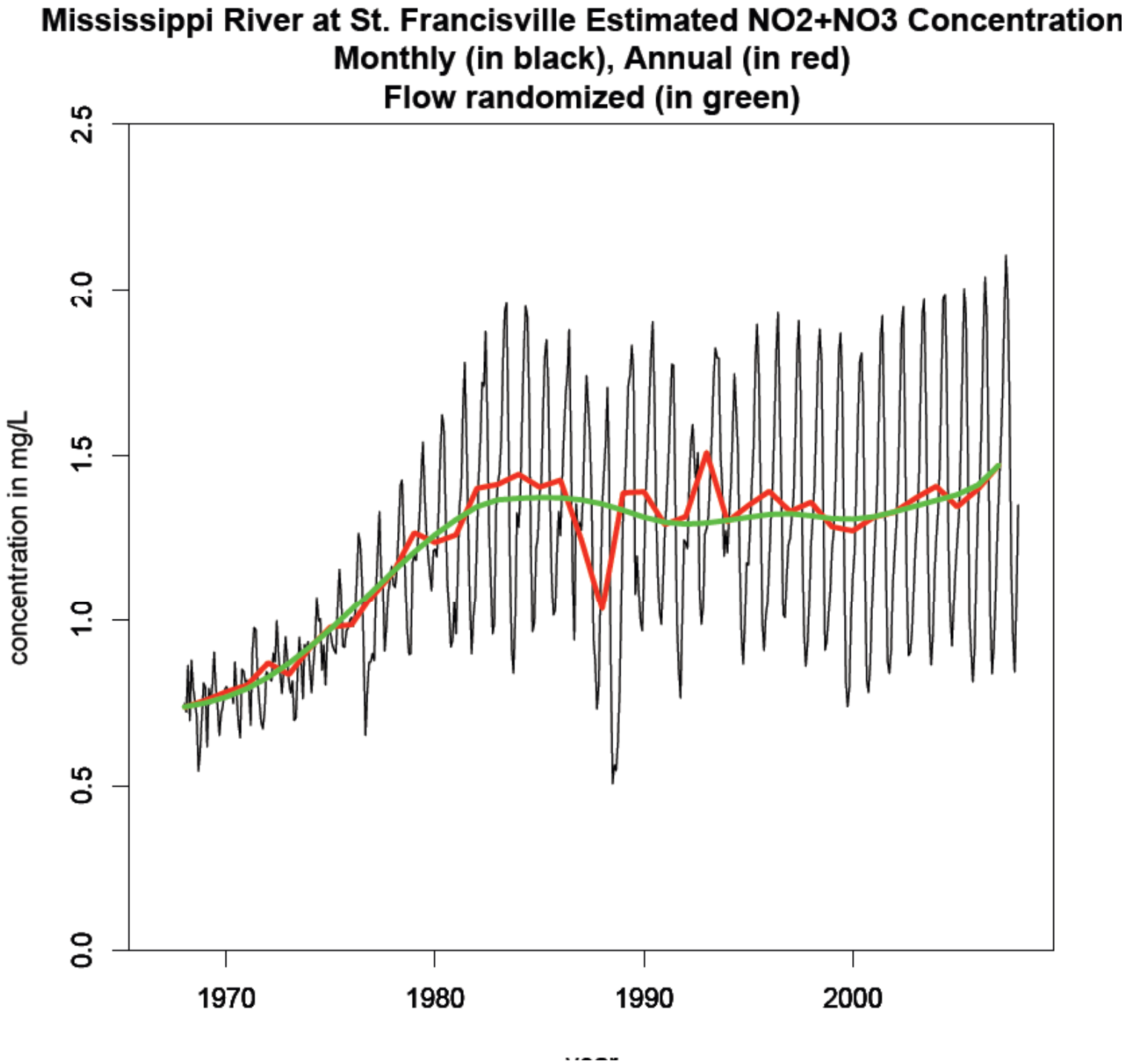

Choptank River nr Greensboro, MD, Nitrate plus Nitrite Concentration
Monthly Estimates (black), Annual Estimates (red)
Flow Randomized Estimates (green)
concentration in mg/L
0.0
0.5
1.0
1.5
1980
1985
1990
1995
2000
2005
2010
year

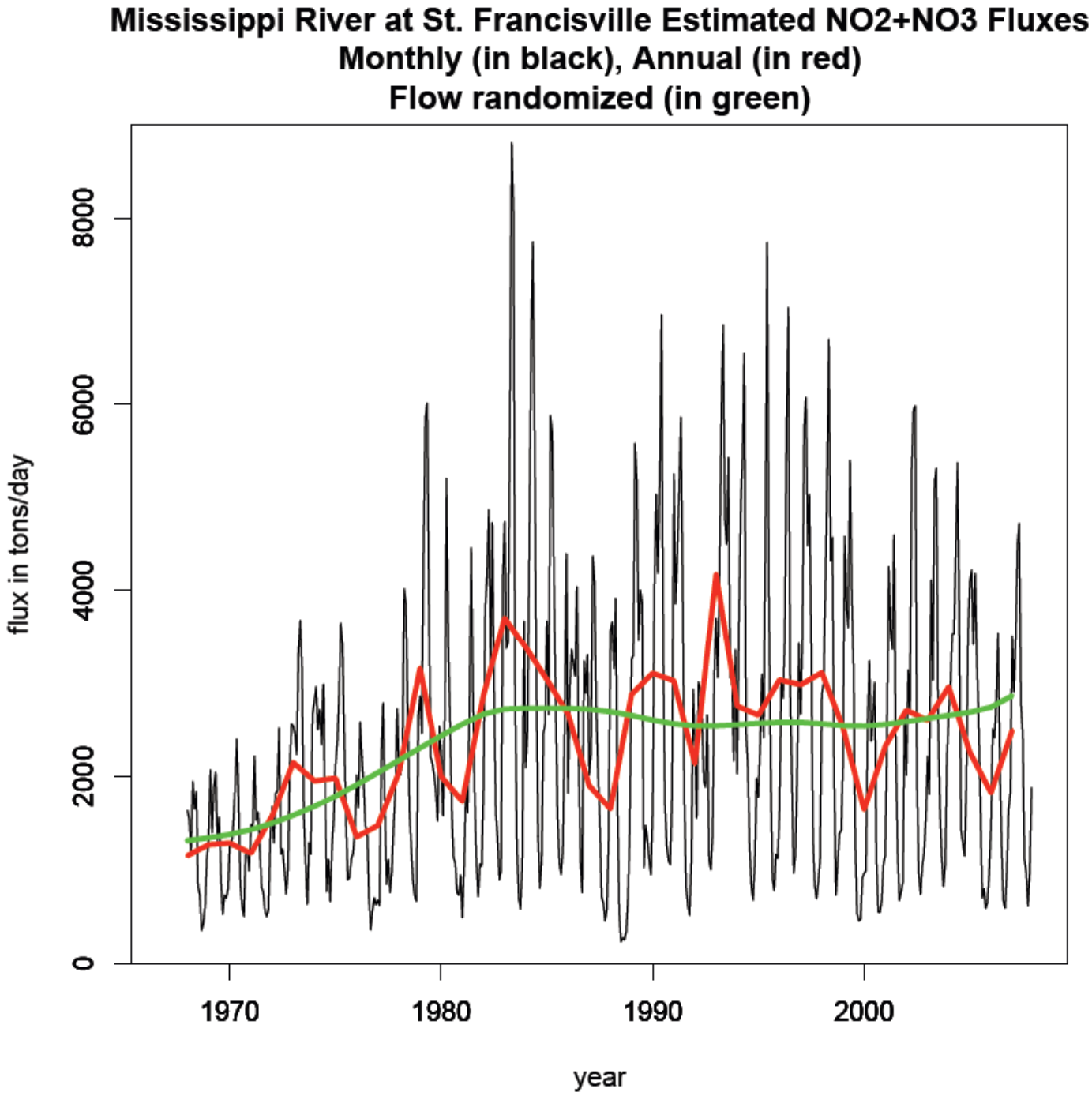

Choptank River nr Greensboro, MD, Nitrate plus Nitrite Flux
Monthly Estimates (black), Annual Estimates (red)
Flow Randomized Estimates (green)
flux in tons/day
0.0
0.5
1.0
1.5
1980
1985
1990
1995
2000
2005
2010
year


NOx, Mississippi River at St. Francisville
April, May, June
Q > 1 million cfs
Concentration in mg/L
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1970
1980
1990
2000
YEAR


Q<700,000 Spring
Concentration in mg/L
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1970
1980
1990
2000
Year


May 15 St. Francisville, NOx
Concentration in mg/L
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1970
1980
1990
2000
2010
Year
1400000fs
9e+05cfs
5e+05cfs


Mississippi River at St. Francisville Estimated NO2+NO3 Concentration
Monthly (in black), Annual (in red)
Flow randomized (in green)
concentration in mg/L
0.0
0.5
1.0
1.5
2.0
2.5
1970
1980
1990
2000


Mississippi River at St. Francisville Estimated NO2+NO3 Fluxes
Monthly (in black), Annual (in red)
Flow randomized (in green)
flux in tons/day
0
2000
4000
6000
8000
1970
1980
1990
2000
year

Some topics not addressed

- Censored Data
- Importance of flow history
- Trends in streamflow
- Estimation of frequency of exceedence of threshold values
- Analysis of multiple variables
- Uncertainty analysis

USGS

Final thoughts about WRTDS

- Diagnostic regarding source
- Show variability **and** the "signal"
- Concentration **and** flux history
- No revision of the distant past
- Stationarity is dead – get over it!
- Change is likely to be gradual

USGS