

# EGRET: Exploration and Graphics for RivEr Trends

An R-package for the analysis of long-term changes in river-water quality and streamflow, including the water quality method “Weighted Regressions on Time, Discharge, and Season” (WRTDS)

Developed by  
Bob Hirsch &  
Laura De Cicco:  
**In Beta testing**



# The WRTDS analysis is a method that has been described in peer reviewed publications

Links to these papers, plus the R-packages, draft manual and this presentation are on the EGRET web site

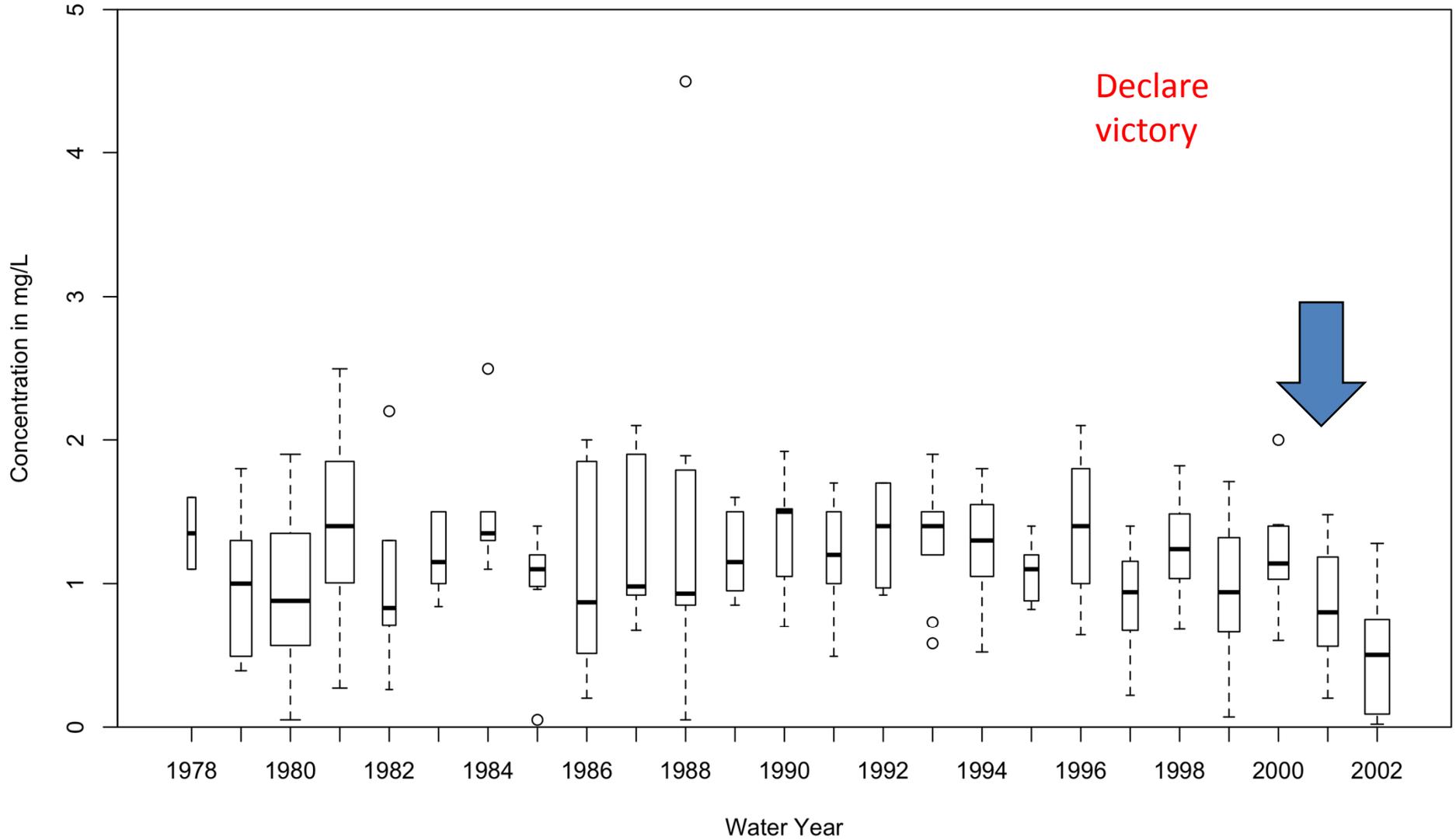
<https://github.com/USGS-CIDA/WRTDS/blob/master/README.md>

**The driving philosophy for  
WRTDS and EGRET:**

**“The only way to figure out what is  
happening to our planet is to  
measure it,  
and this means tracking changes  
decade after decade  
and poring over the records.”**

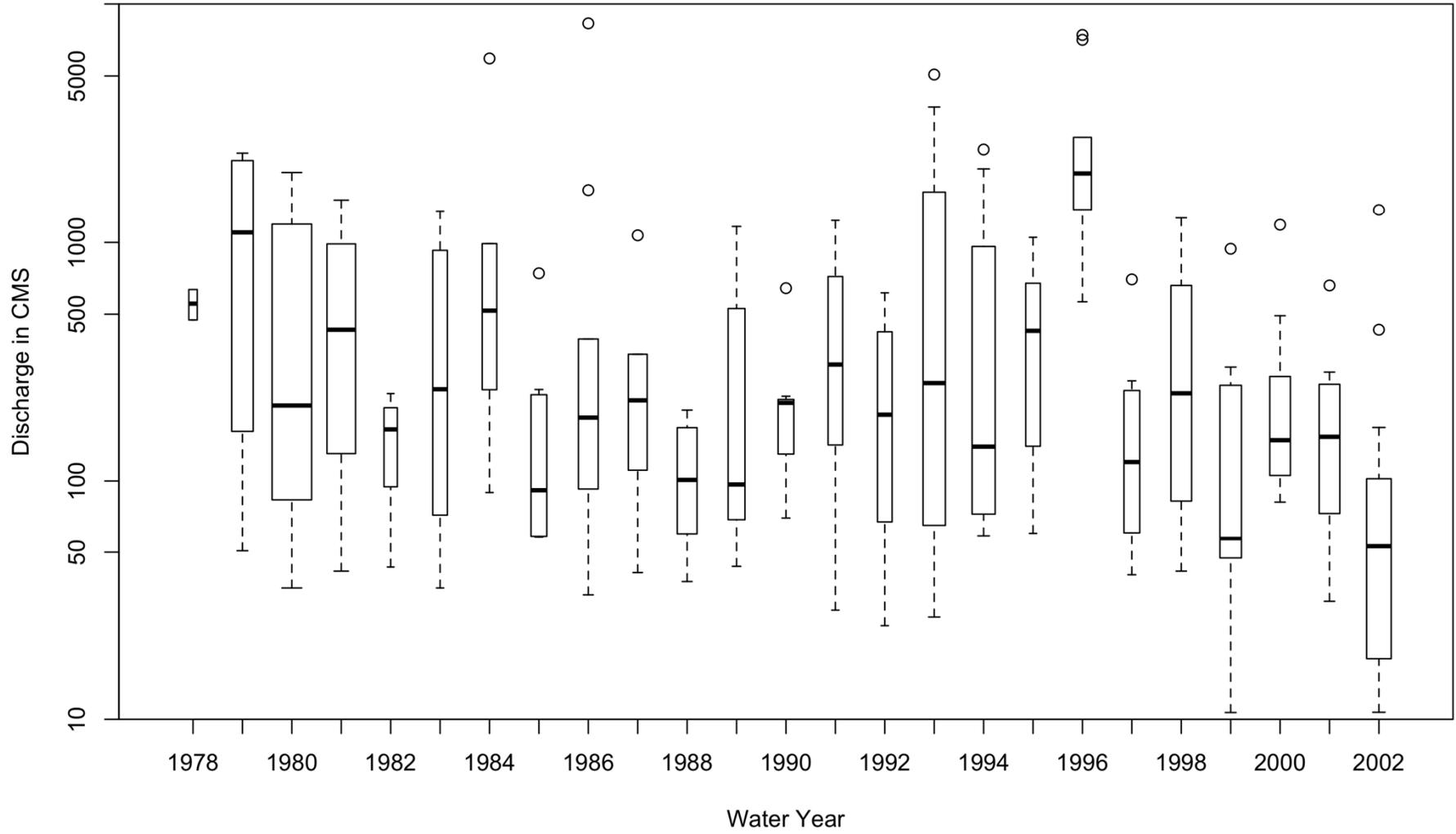
*(Ralph Keeling, Science Magazine, 2008)*

Potomac River at Chain Bridge, Washington DC  
Box plot of sample values by Water Year  
Nitrate as N

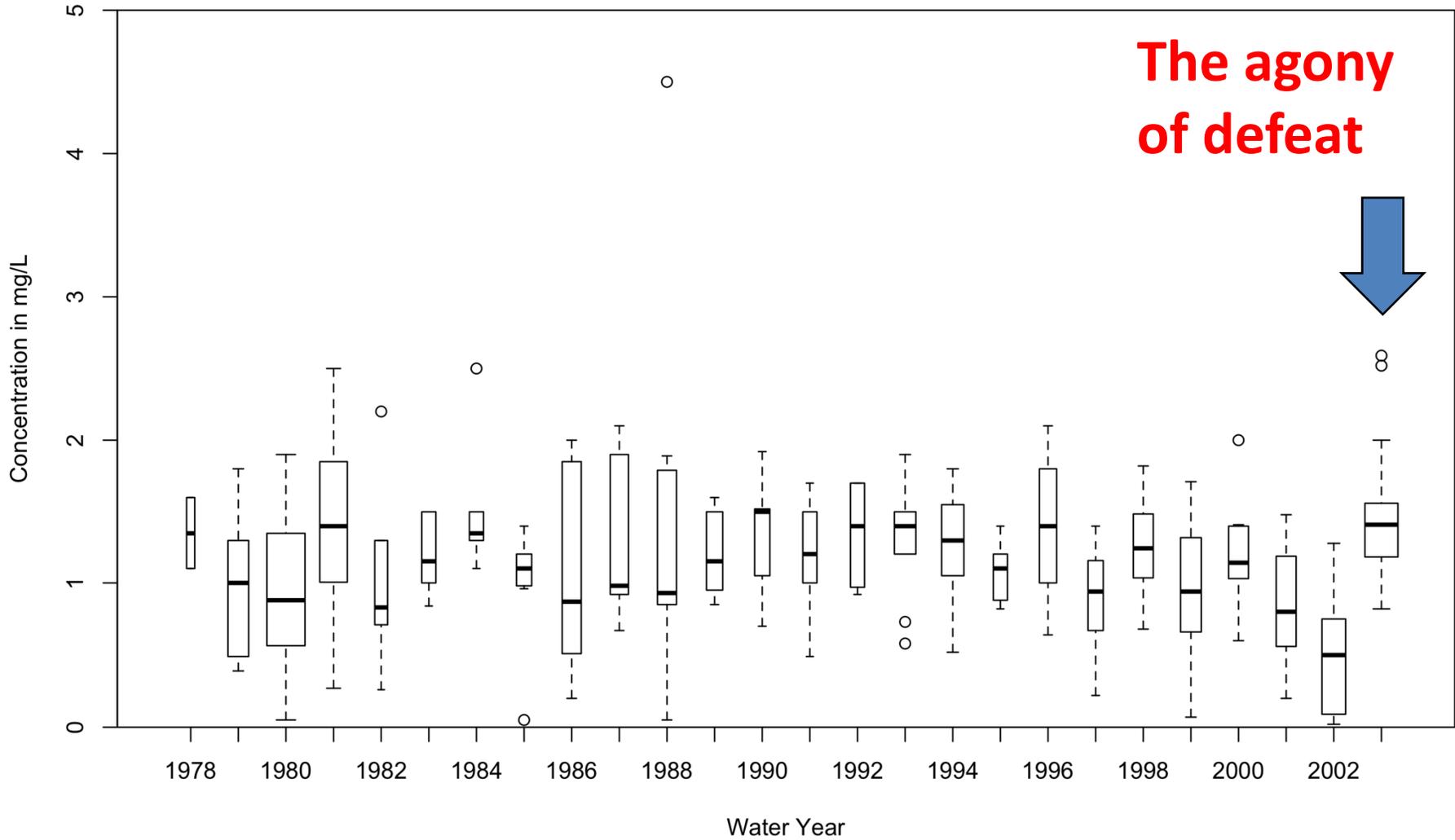


Data through September 2002

Potomac River at Chain Bridge, Washington DC  
Boxplot of Discharge on Sampling Date by Water Year

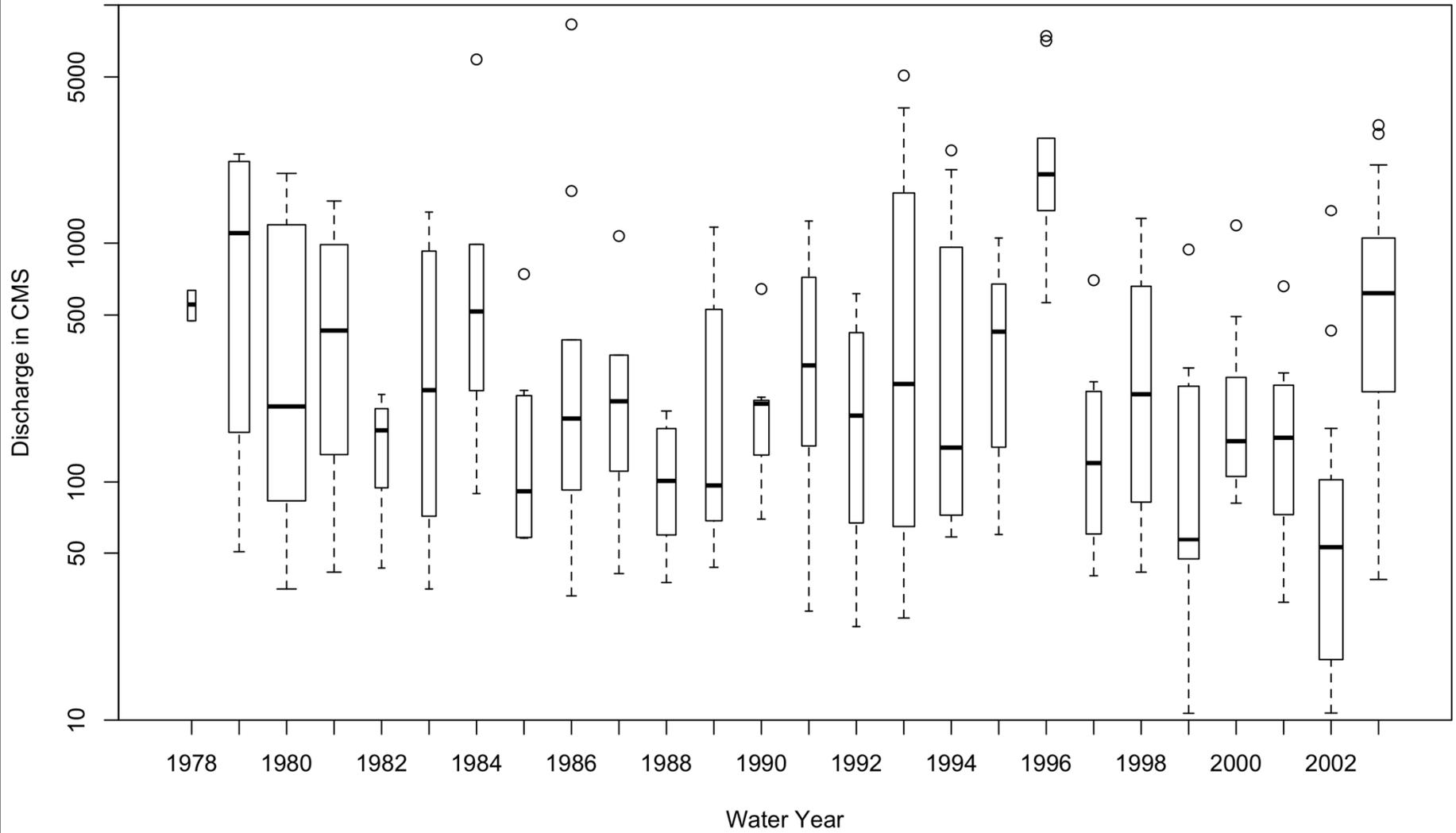


Potomac River at Chain Bridge, Washington DC  
Box plot of sample values by Water Year  
Nitrate as N

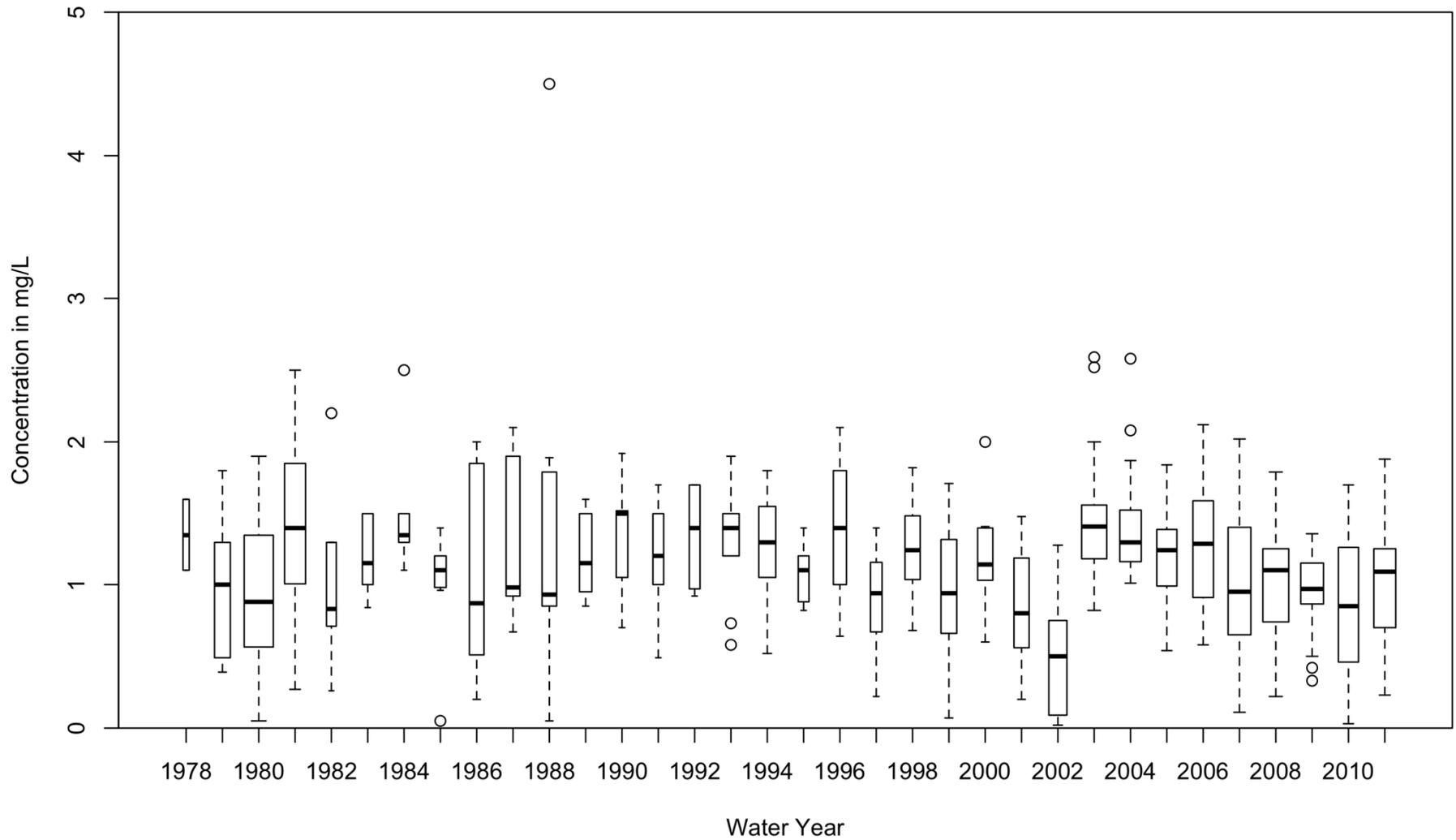


Data through September 2003

Potomac River at Chain Bridge, Washington DC  
Boxplot of Discharge on Sampling Date by Water Year



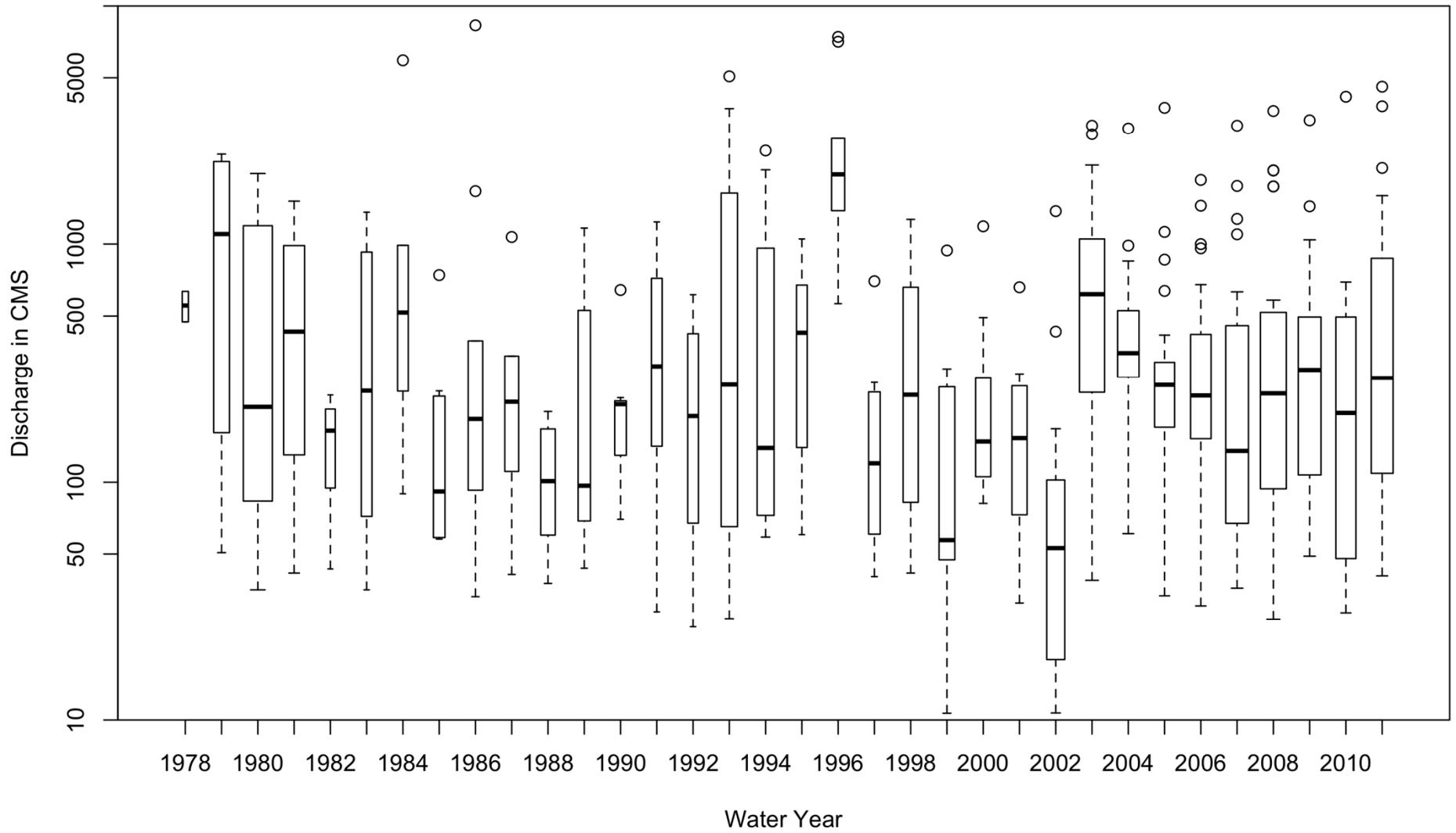
Potomac River at Chain Bridge, Washington DC  
Box plot of sample values by Water Year  
Nitrate as N



Data through September 2011



Potomac River at Chain Bridge, Washington DC  
Boxplot of Discharge on Sampling Date by Water Year



# What needs motivated WRTDS?

- Let the data inform the statistical model. No mathematical straight-jacket!!
- Estimate both concentration & flux.
- Estimate the actual history but also a flow-normalized history.
- Try to resolve a serious flux-bias problem.
- Be quantitative but also exploratory.
- Focus on description and understanding not on statistical significance. Like economic and social time series.

# EGRET overview

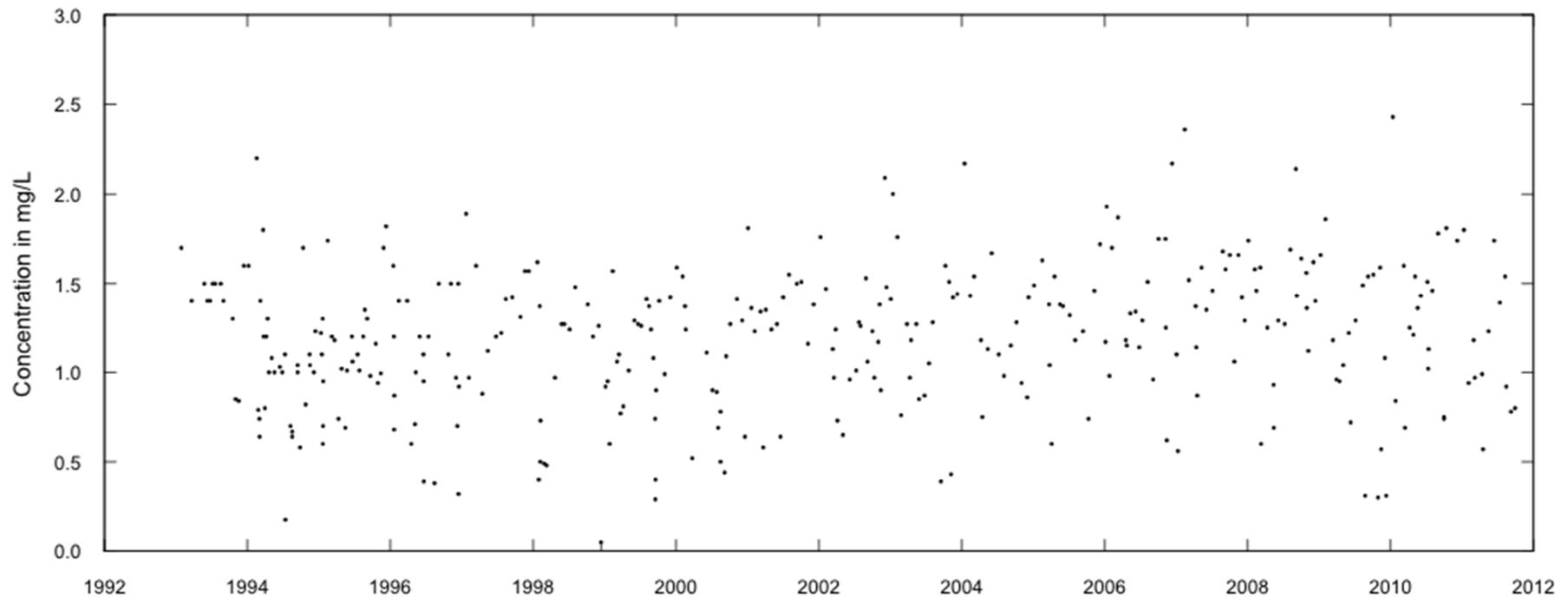
- Ingest water quality sample data, daily streamflow data, and meta data from **USGS Web services** or from **user-supplied files**
- Sub-systems: **WRTDS** for river water quality data analysis and **flowHistory** for analysis of streamflow alone
- Goal: Exploration of the data to describe the evolving hydrologic system. Produce: graphs, summary statistics, understanding, and hypotheses.

# What's the underlying idea of WRTDS?

Water quality is influenced by many factors, we need tools to sort out the factors

What's going on here?

Choptank River near Greensboro, MD , Nitrate plus Nitrite, filtered, as N



# What's the underlying idea of WRTDS?

Use the data and a simple, **highly-flexible** smoothing model, to compute **for every day** in the study period, an estimate of **concentration** and an estimate of **flux**.

The smoothing model expresses the behavior as a combination of influences:

- 1) **Time trend**
- 2) **Discharge**
- 3) **Seasonal cycle**
- 4) **Random component**

# Locally Weighted Regression

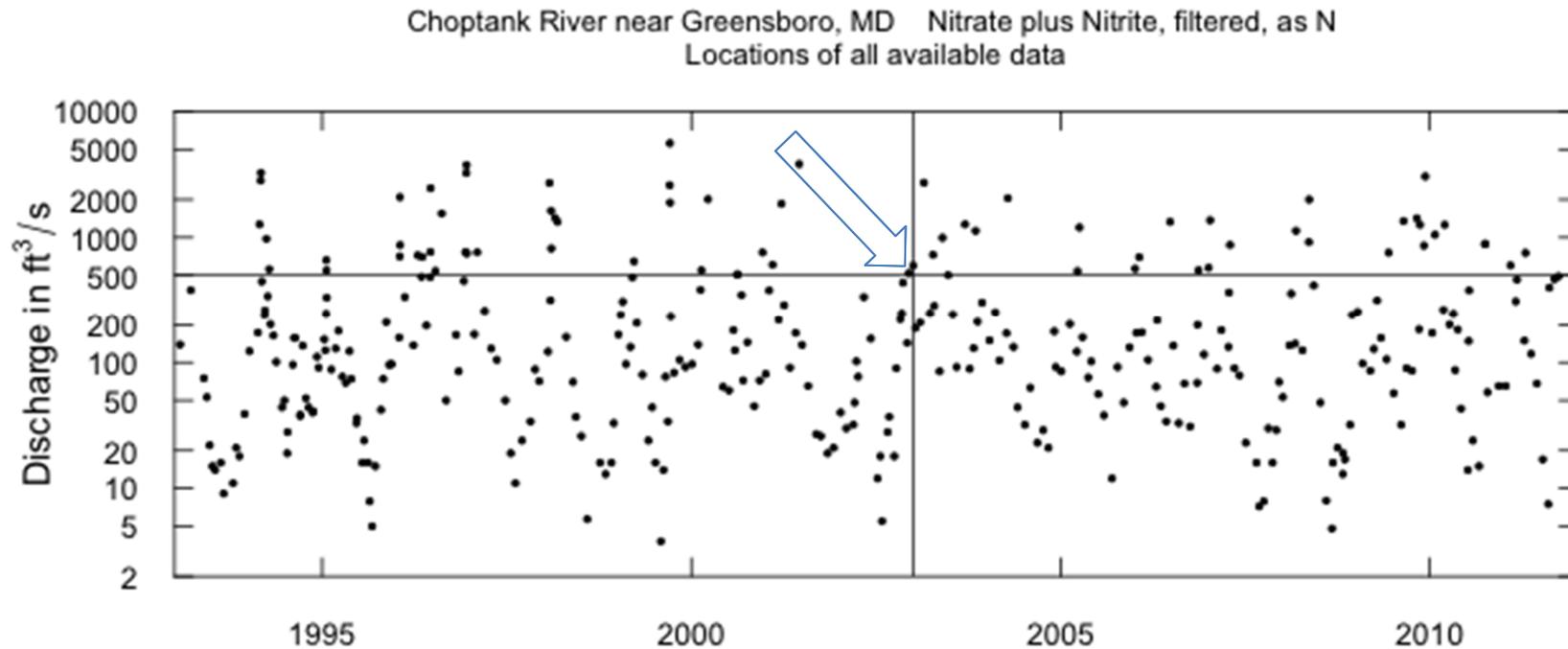
For any location in time - discharge space  
( $t$  and  $Q$ ) we assume that concentration ( $c$ ) follows  
this model

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

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

Use weighted regression at many points in that space.  
The weight on each sample is determined by its  
“relevance” to that particular point in the space.

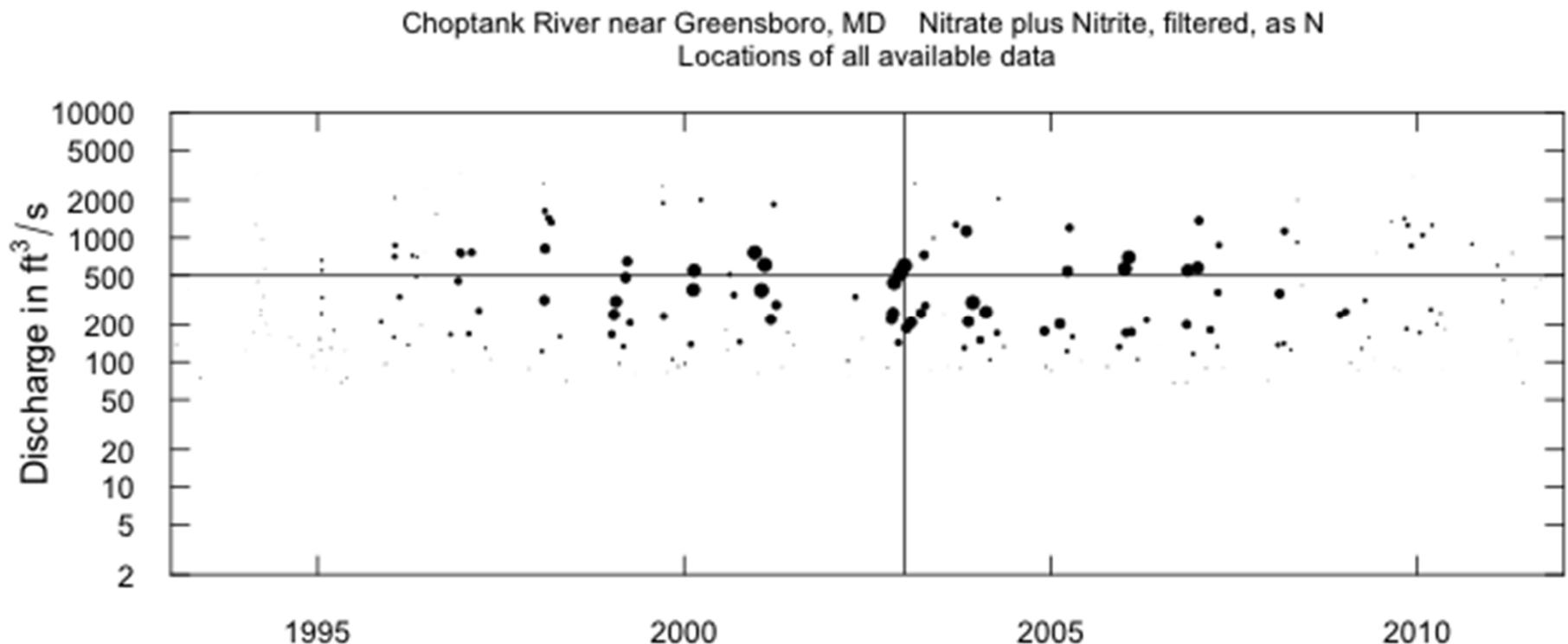
Every dot is a data point from 1993 to 2012  
Let's say we want to use the data to estimate  
the expected value of concentration for  
January 1, 2003 at  $Q=500$  cfs



## How do we set the weights for the regression?

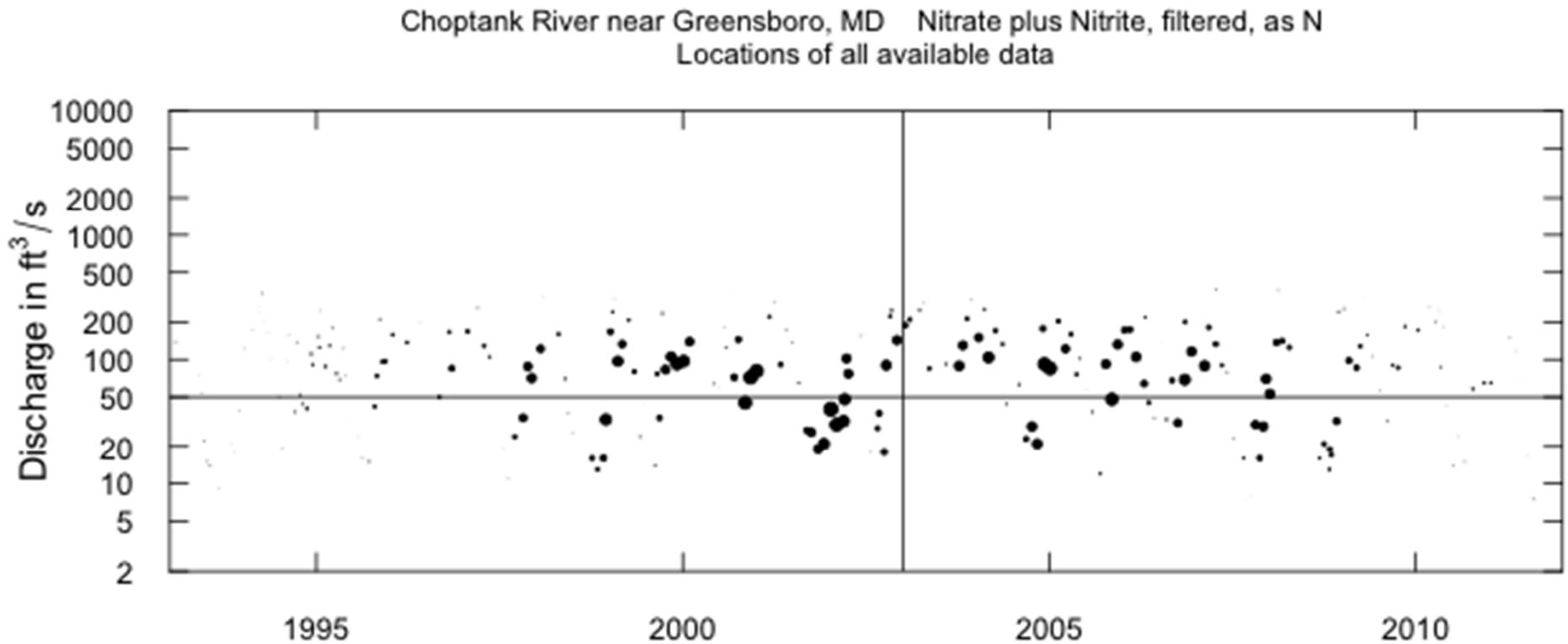
These are the same points we just saw, but the radius of the dot is proportional to weight assigned to that point for purposes of estimating concentration for January 1, 2003 at Q=500 cfs

The weight depends on distance in: time, log discharge, and season from January 1, 2003 at Q = 500 cfs



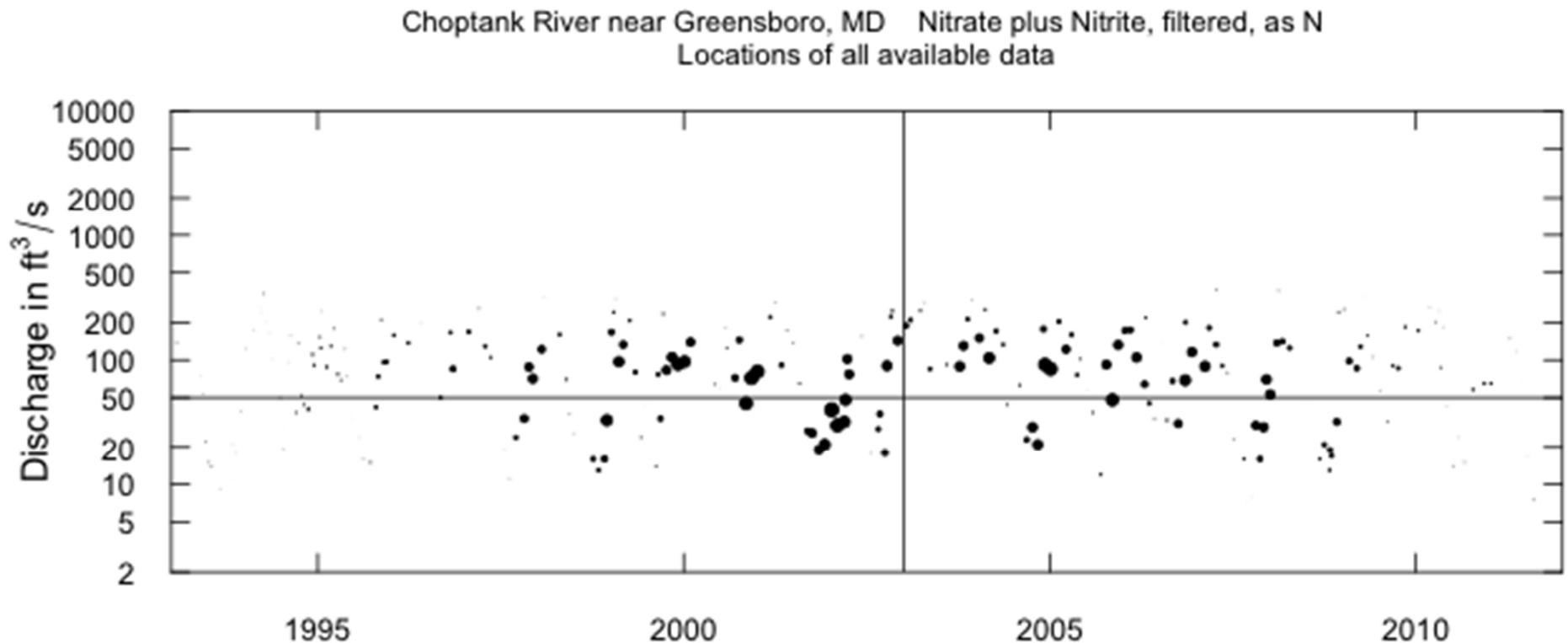


What if we wanted to make an estimate for  
January 1, 2003 but for  $Q = 50$  cfs  
Redo the weights for distance from that point



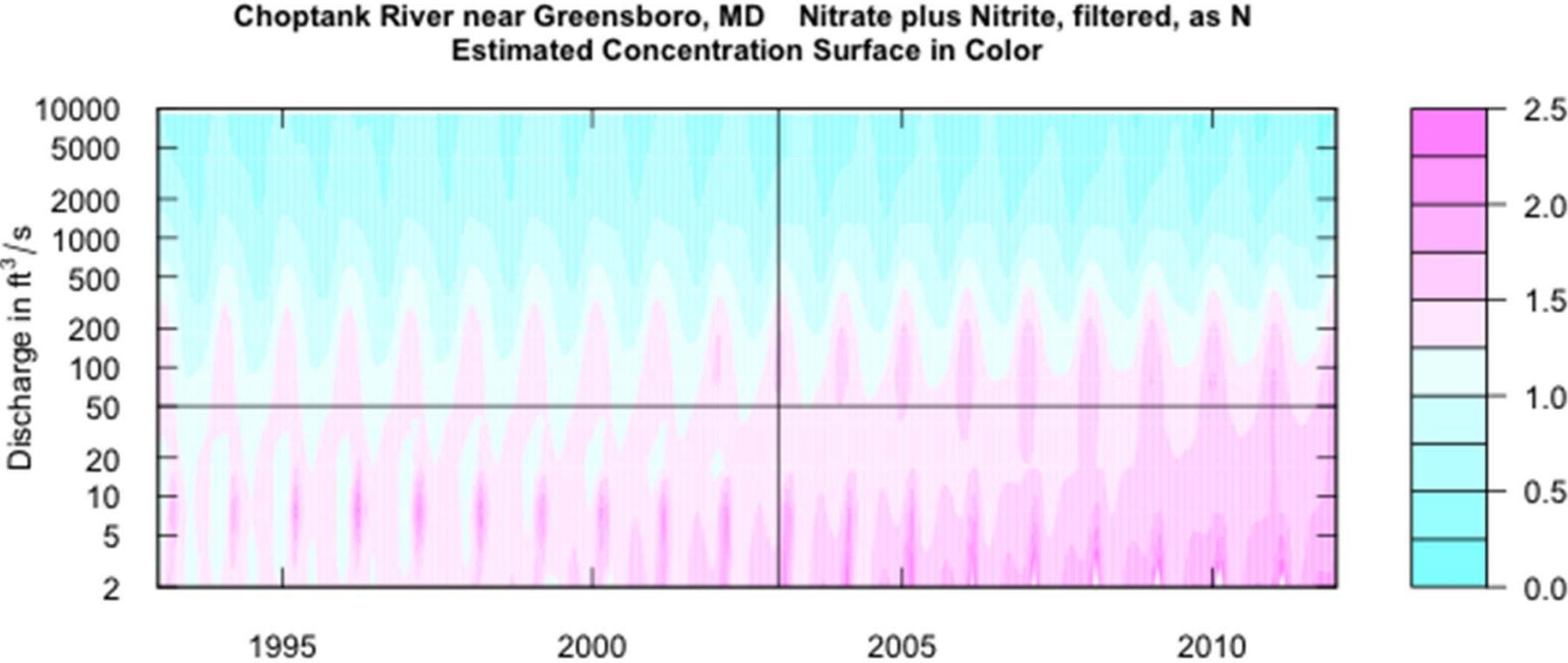
To organize the work, let's make estimates for a fine mesh of points in this space.

Over this space we will do it at 14 Q values and 177 time values, for a grid of 2,478 points.

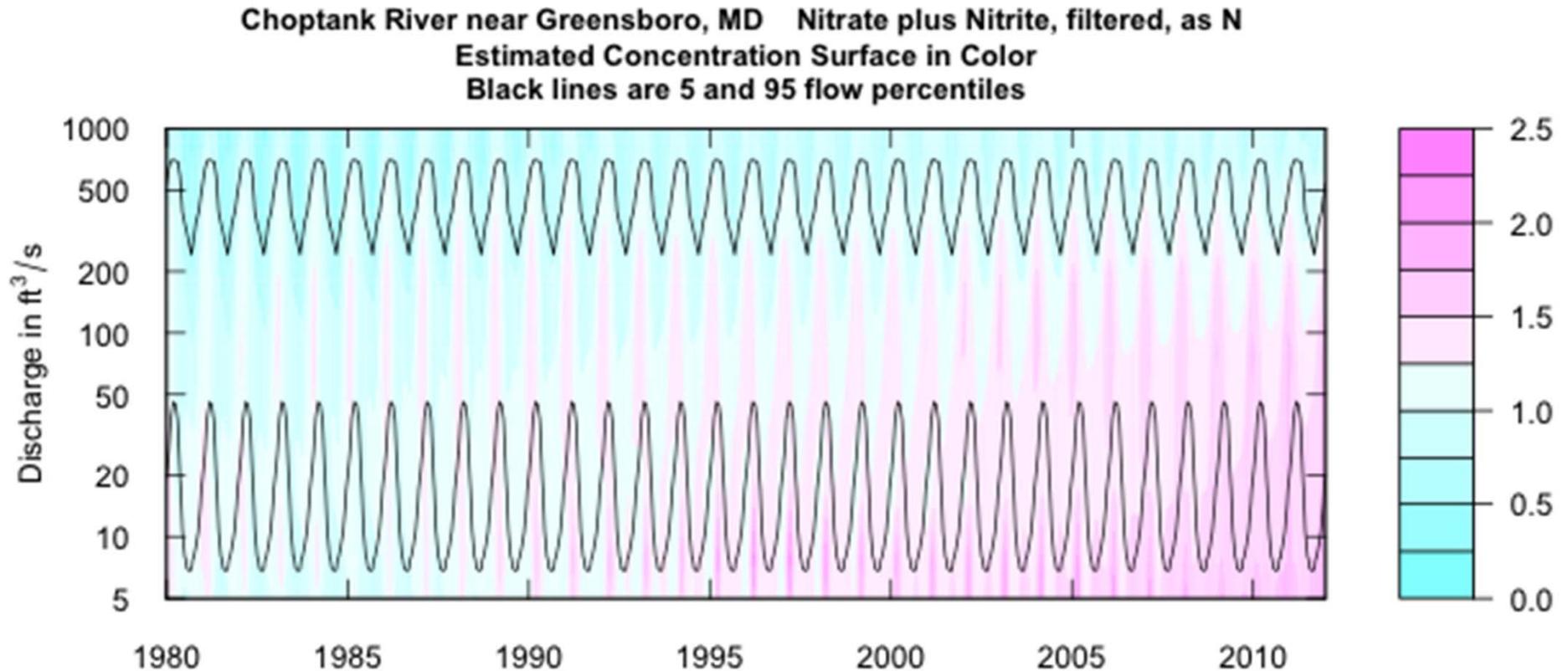


Here is the “surface” computed:

It is the estimates of the Expected value of Concentration as a function of time and discharge

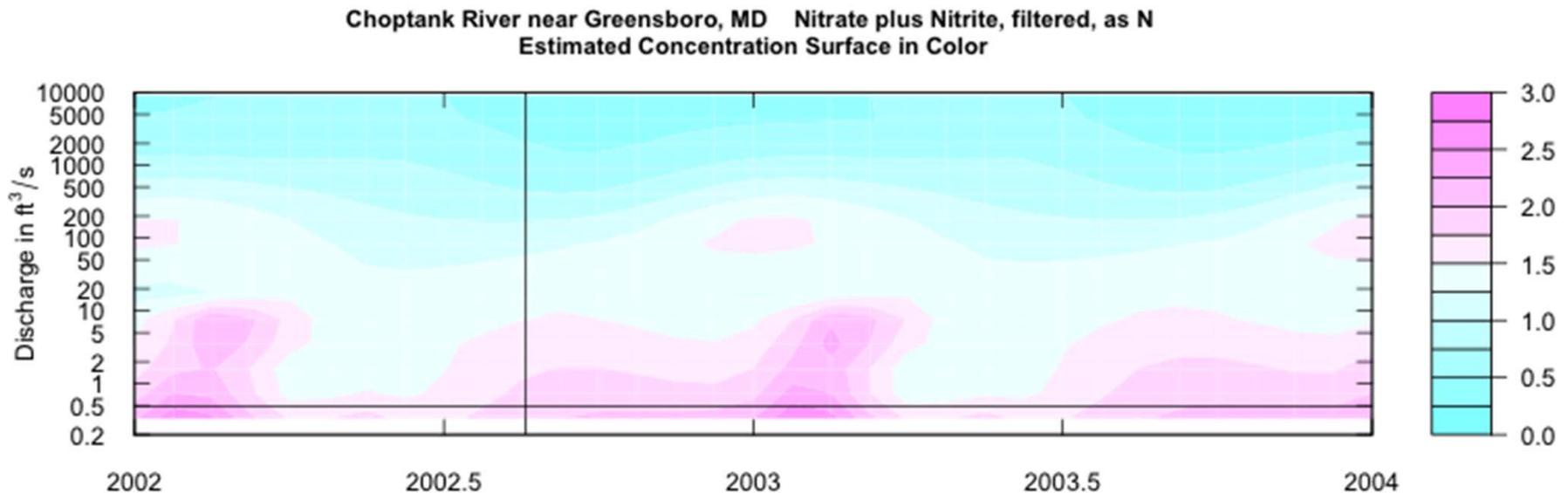


Here is the whole surface: 513 time values by 14 discharge values, for a total of 7,168 points



**Added the black lines to illustrate the 5<sup>th</sup> and 95<sup>th</sup> percentiles on the seasonal flow duration curve**

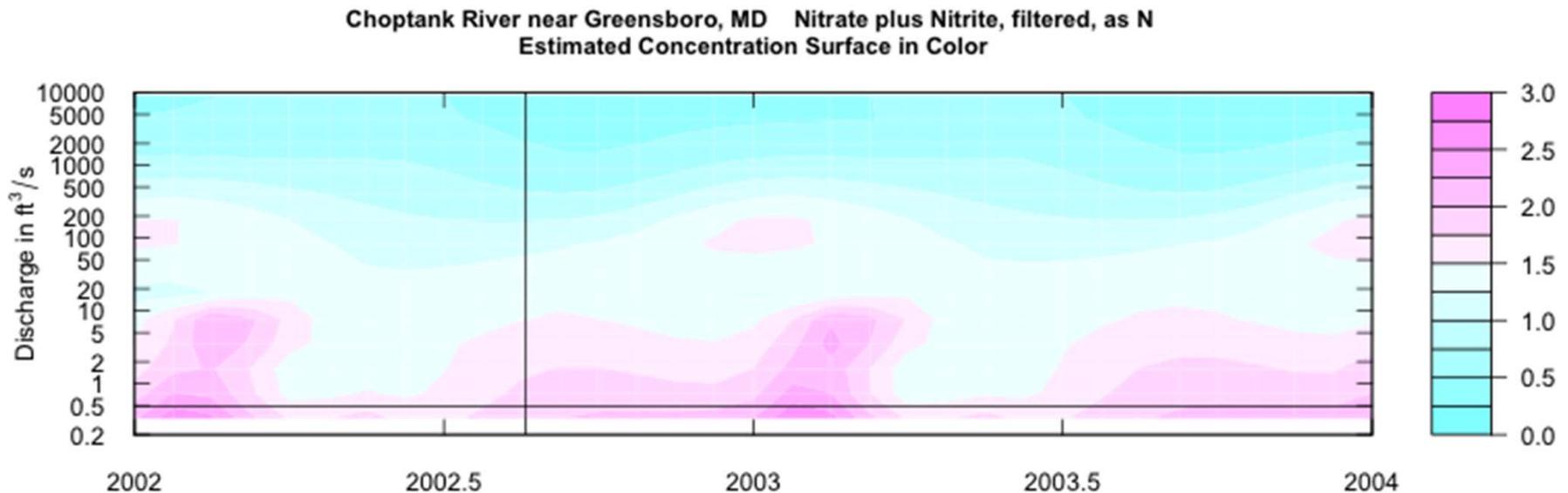
Look at a small part of this contour plot  
Let's look at August 20, 2002 (a serious drought), discharge was 0.49 cfs



Interpolation off this surface gives me an estimate of **concentration** for that day: 1.86 mg/L

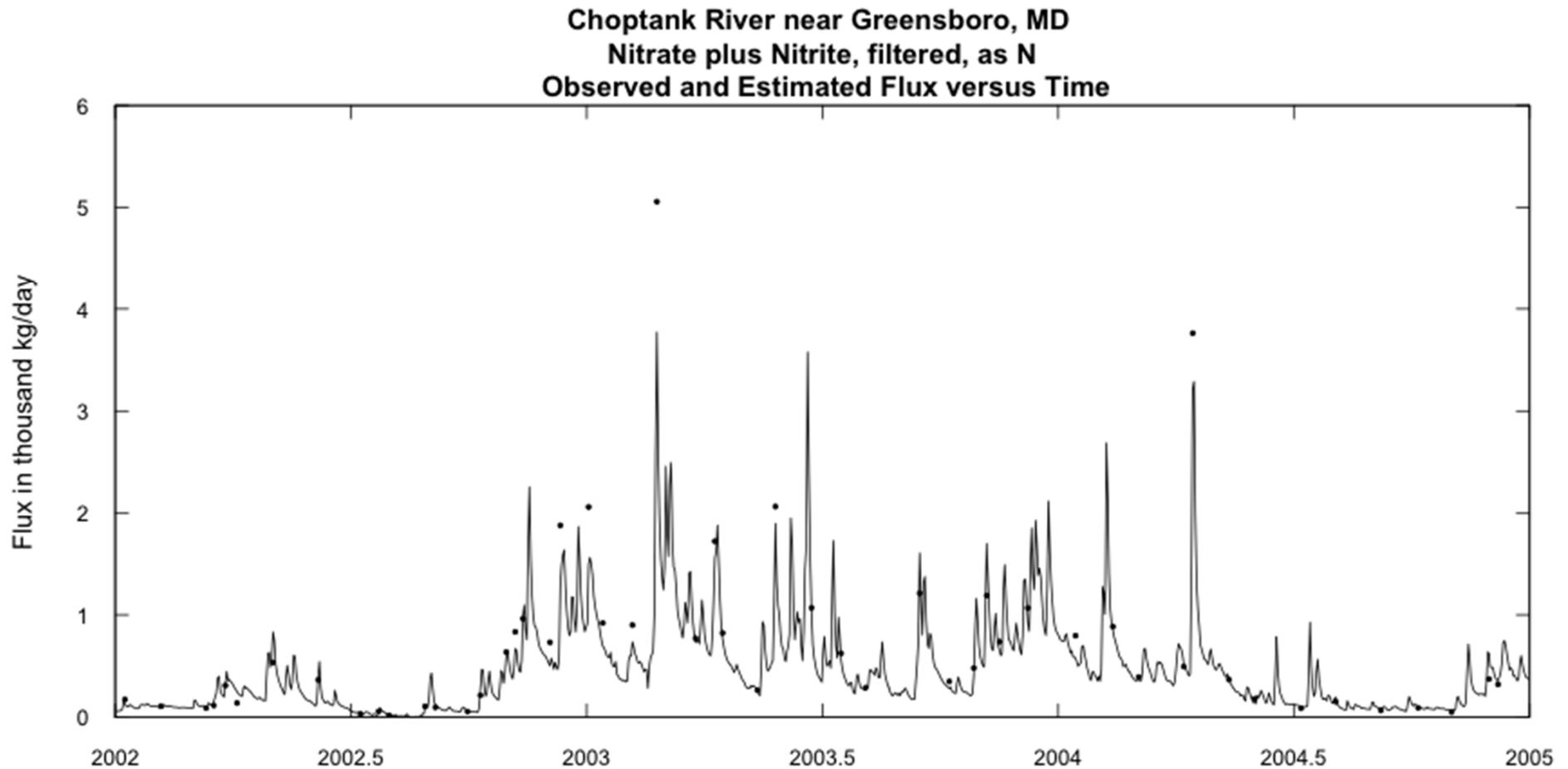
For **flux** we compute:  $E(C) * Q$  to get 2.24 kg/day

EGRET does this process for each of the 11,718 days in the period of record.

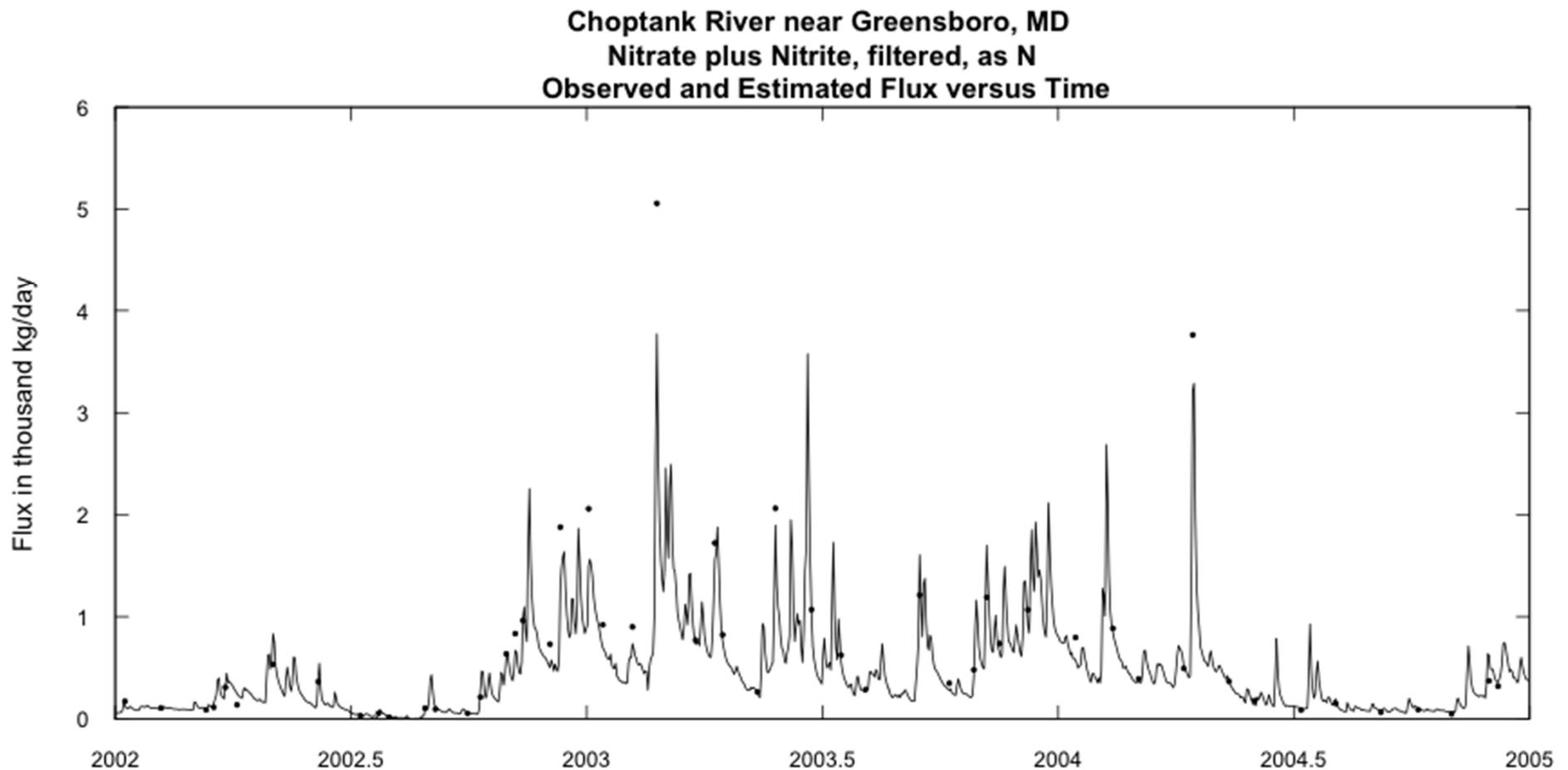


It computes and stores:  
11,718 estimated daily concentrations  
11,718 estimated daily fluxes

We can look at these flux estimates and compare them to the actual fluxes for the sampled days



**Most of the difference between years is driven by the natural random variation in streamflow. Water years 2003 and 2004 were very wet.**





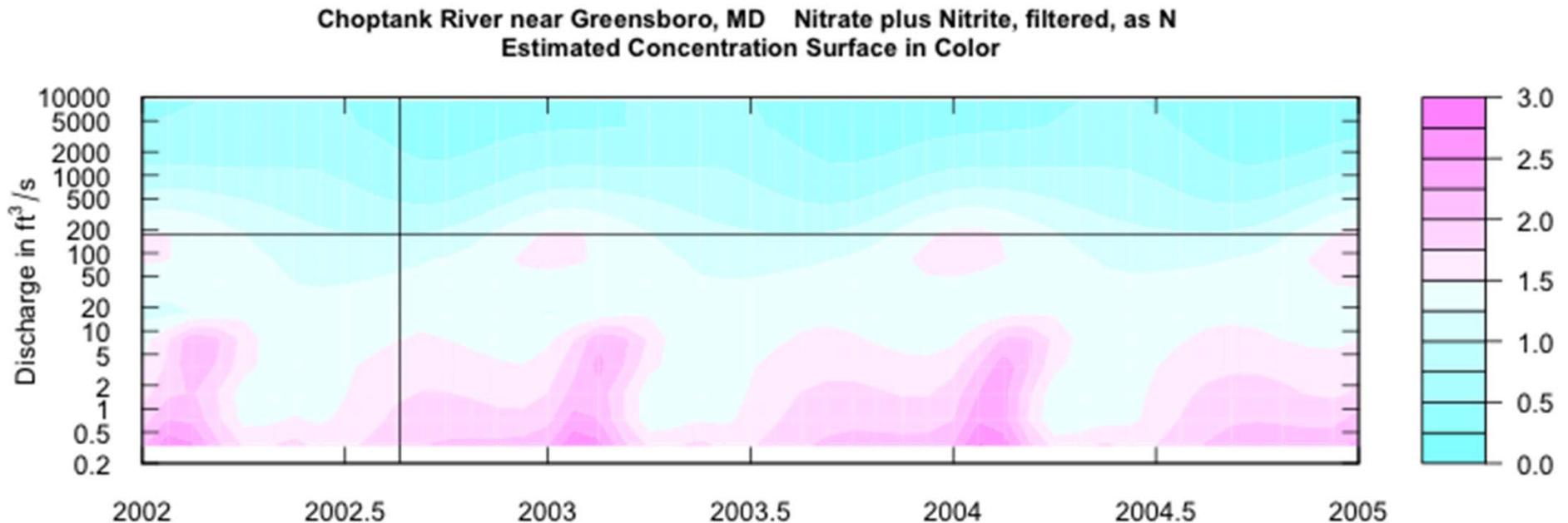
**The estimated concentrations and fluxes are important, particularly if one wants to understand the estuary's response to inputs.**

**But, if we really want to understand “progress” or “lack of progress” in improving water quality in this watershed, we need to remove the effect of these random year-to-year variations driven by streamflow.**

**We do this through “flow-normalization”**

**How does that work?**

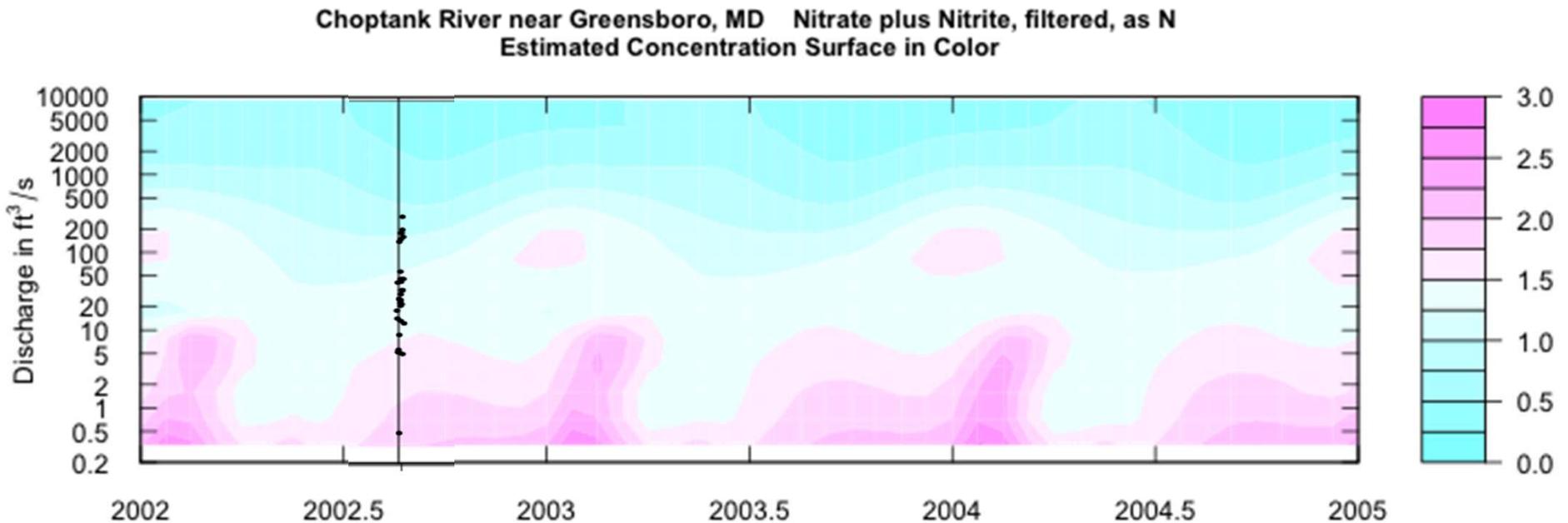
We can say that it is just as likely that Q could have been 172 cfs that day (it happens to be the Q from August 20, 2003) as the 0.49 cfs that happened August 20, 2002



The WRTDS estimate of concentration for that discharge, but based on the “behavior of the system” on August 20, 2002, is 1.0 mg/L

The flux for that discharge and that concentration comes to 431 kg/day

Let's think of every Aug. 20 Q value we have seen as equally likely to have happened on Aug. 20, 2002 and estimate the concentration and flux we would have had on Aug. 20, 2002 with that Q



The mean of those concentrations is 1.31 mg/L, we call this the “flow-normalized concentration” for Aug. 20, 2002

The mean of the flux values is 164.4 kg/day, we call this the “flow-normalized flux” for Aug. 20, 2002

**Flow-normalization allows us to remove the influence of the year to year variation in discharge (but not the natural seasonal variation in discharge)**

**The program computes, for every day:**

- The estimated concentration**
- The flow-normalized concentration**
- The estimated flux (flow \* concentration)**
- The flow-normalized flux (by integrating flux over the frequency distribution of flow)**

Now we have, for August 20, 2002

Estimated Concentration = 1.86 mg/L

Flow Normalized Concentration = 1.31 mg/L

Estimated Flux = 2.24 kg/day

Flow Normalized Flux = 164.4 kg/day

Now: do that process again and again for all  
11,718 days in the record and store each of  
those four numbers for each day

It is assumed here that streamflow is stationary over the period of interest.

If I had a strong reason to believe that the past 34 years of August 20 streamflows is not a reasonable sample of the probability distribution of streamflows I can expect to see on this year's August 20, then I should not use flow normalization.

Major new water controls (dams built or removed), or major changes in flow due to changes in water withdrawals, changes in impervious surface area, or groundwater drawdown, would require a modified approach.

I do not see climate change as sufficiently large at this time to invalidate this approach.

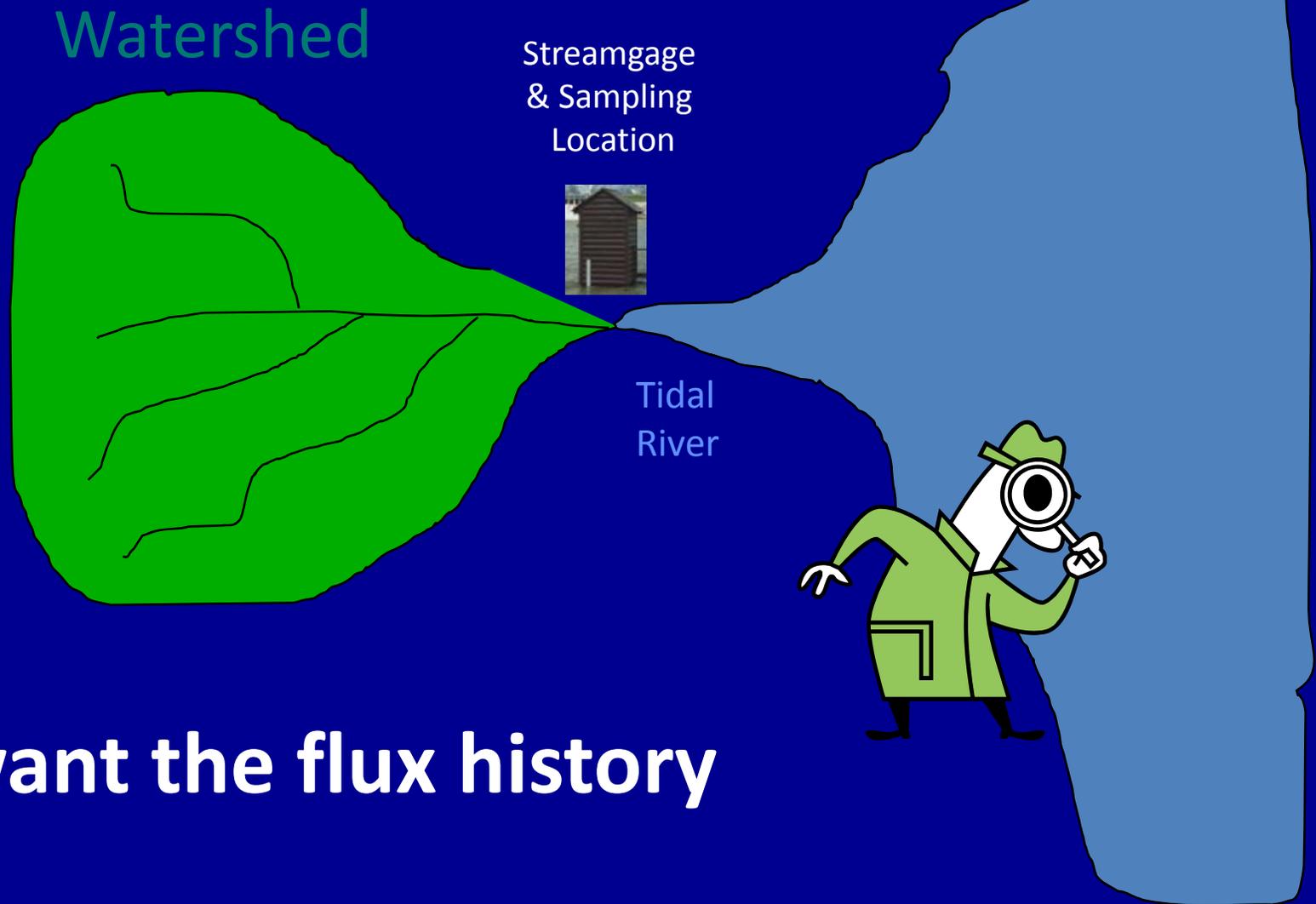
**Why all this complexity?**

**Different products for different  
purposes**

**Concentration vs Flux**

**History vs Flow-normalized history**

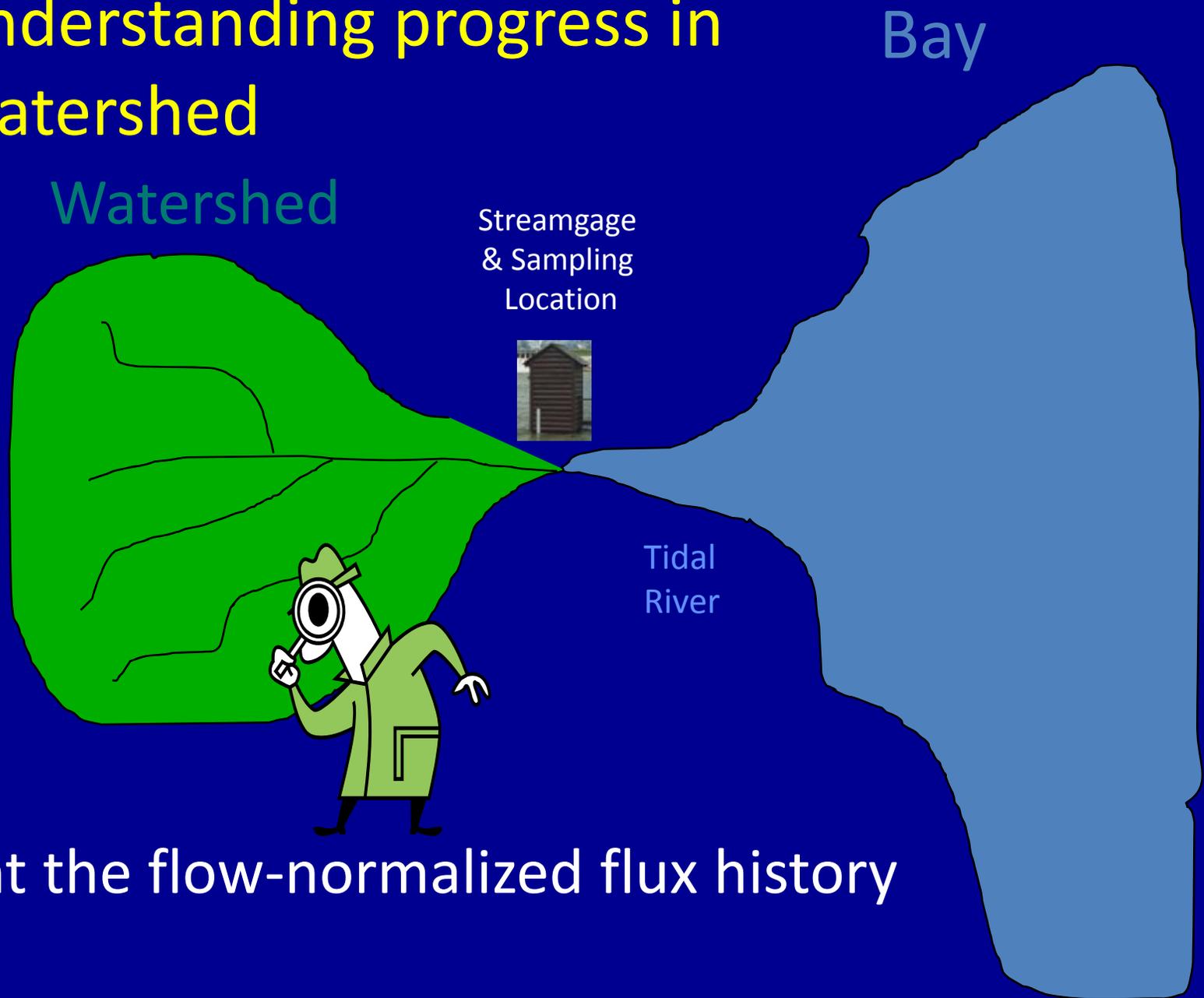
# For understanding impact on the Bay ecosystem



**We want the flux history**

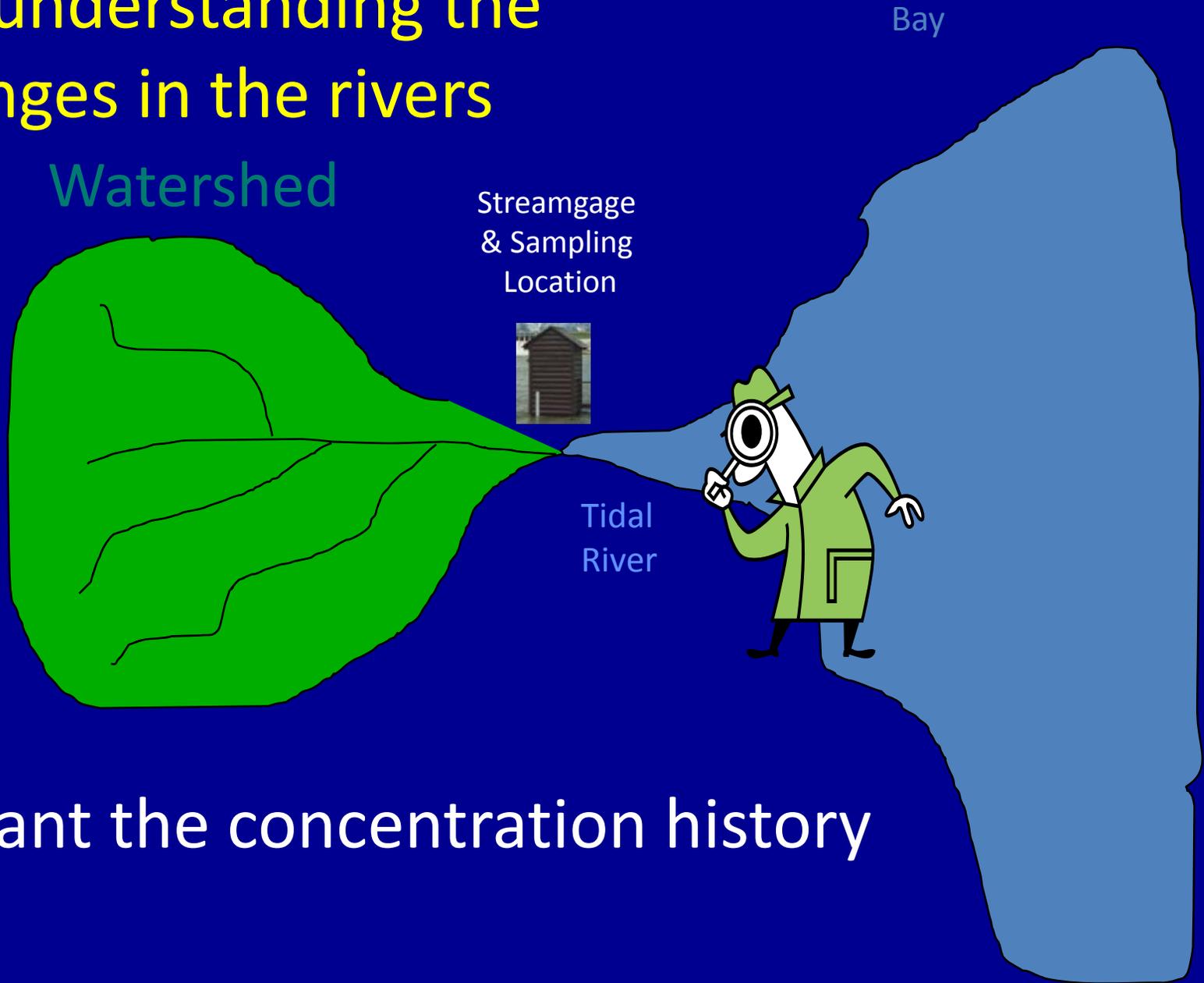


# For understanding progress in the watershed



We want the flow-normalized flux history

# For understanding the changes in the rivers



We want the concentration history

# Let's run an analysis

- Read in data
- Look at the data
- Run the model in cross-validation mode and check quality of the fit
- Compute the “surface” of the expected value of concentration as a function of time, discharge and season
- Use that surface to produce the 4 daily time series:
- Reports those results as graphs, tables, and change measures
- Explore the data and the model

<https://github.com/USGS-CIDA/WRTDS/blob/master/README.md>

- At the downloads page you will see something like this: this:

### Download Packages

---

 [EGRET manual.doc](#) — Latest EGRET Manual  
7.0MB · Uploaded 4 days ago

 [EGRET\\_1.1.1.tar.gz](#) — Latest EGRET package  
511KB · Uploaded 7 days ago

 [dataRetrieval\\_1.0.tar.gz](#) — Latest dataRetrieval package  
376KB · Uploaded 20 days ago

---

Download the tar.gz files them and then open R

```

> library(dataRetrieval)
> library(EGRET)
> sta<-"01491000"
> param<-"00631"
> StartDate<-"1979-09-01"
> EndDate<-"2011-09-30"
> Sample<-getSampleData(sta,param,StartDate,EndDate)
> summary(Sample)

```

Date	ConcLow	ConcHigh	Uncen	ConcAve	Julian
Min. :1979-09-25	Min. :0.176	Min. :0.050	Min. :0.0000	Min. :0.025	Min. :47383
1st Qu.:1988-12-30	1st Qu.:0.880	1st Qu.:0.880	1st Qu.:1.0000	1st Qu.:0.880	1st Qu.:50768
Median :1994-03-06	Median :1.100	Median :1.100	Median :1.0000	Median :1.100	Median :52659
Mean :1995-05-05	Mean :1.125	Mean :1.123	Mean :0.9985	Mean :1.123	Mean :53085
3rd Qu.:2002-04-12	3rd Qu.:1.400	3rd Qu.:1.400	3rd Qu.:1.0000	3rd Qu.:1.400	3rd Qu.:55618
Max. :2011-09-29	Max. :2.430	Max. :2.430	Max. :1.0000	Max. :2.430	Max. :59075
	NA's :1.000				

Month	Day	DecYear	MonthSeq	SinDY	CosDY
Min. : 1.00	Min. : 2.00	Min. :1980	Min. :1557	Min. :-1.0000	Min. :-0.999963
1st Qu.: 3.00	1st Qu.: 81.75	1st Qu.:1989	1st Qu.:1669	1st Qu.: -0.6306	1st Qu.: -0.672949
Median : 6.00	Median :155.50	Median :1994	Median :1731	Median : 0.1961	Median : -0.017166
Mean : 6.06	Mean :168.06	Mean :1995	Mean :1745	Mean : 0.0875	Mean : 0.004315
3rd Qu.: 9.00	3rd Qu.:255.25	3rd Qu.:2002	3rd Qu.:1828	3rd Qu.: 0.7841	3rd Qu.: 0.710135
Max. :12.00	Max. :363.00	Max. :2012	Max. :1941	Max. : 0.9999	Max. : 0.999668

```

> length(Sample$Date)
[1] 652
>

```

# Treatment of "less than values"

In the uncensored case, let's say concentration is 1.0

Then ConcLow = 1.0 and ConcHigh = 1.0

In the usual type of censored case, say concentration is reported as <1.0

Then ConcLow = NA and ConcHigh = 1.0

Because of censoring we use weighted "survival regression" (the function `survreg` in R) in place of ordinary weighted regression.

It views every data point as an interval:

In the first case the interval is (1.0 to 1.0)

In the second case the interval is (0.0 to 1.0)

# Treatment of "less than values"

There is one more kind of case. The analyte of interest is the sum of two or more measured analytes.

Here is a real example for Total Nitrogen in the Susquehanna River, Maryland, April 27, 1988.

The rule is: Compute Total N as Ammonia plus organic N, unfiltered + Nitrate plus nitrite, filtered

They were reported as <0.2 and 0.9 mg/L respectively.

Then ConcLow = 0.9 and ConcHigh = 1.1

Conventional left-censored approach calls this (0, 1.1)

WRTDS calls this (0.9 to 1.1)

```
> Daily<-getDVData(sta,"00060",StartDate,EndDate)
```

There are 11718 data points, and 11718 days.

There are 0 zero flow days

If there are any zero discharge days, all days had 0 cfs added to the discharge value.

```
> summary(Daily)
```

Date	Julian	Month	Day	DecYear
Min. :1979-09-01	Min. :47359	Min. : 1.000	Min. : 1.0	Min. :1980
1st Qu.:1987-09-08	1st Qu.:50288	1st Qu.: 4.000	1st Qu.: 92.0	1st Qu.:1988
Median :1995-09-15	Median :53218	Median : 7.000	Median :184.0	Median :1996
Mean :1995-09-15	Mean :53218	Mean : 6.529	Mean :183.3	Mean :1996
3rd Qu.:2003-09-22	3rd Qu.:56147	3rd Qu.:10.000	3rd Qu.:274.0	3rd Qu.:2004
Max. :2011-09-30	Max. :59076	Max. :12.000	Max. :366.0	Max. :2012

MonthSeq	Qualifier	i	LogQ	Q7	Q30
Min. :1557	Length:11718	Min. : 1	Min. :-4.61412	Min. : 0.01808	Min. : 0.09606
1st Qu.:1653	Class :character	1st Qu.: 2930	1st Qu.: -0.06779	1st Qu.: 0.99109	1st Qu.: 1.17609
Median :1749	Mode :character	Median : 5860	Median : 0.87835	Median : 2.55661	Median : 2.87133
Mean :1749		Mean : 5860	Mean : 0.76602	Mean : 4.08154	Mean : 4.08059
3rd Qu.:1845		3rd Qu.: 8789	3rd Qu.: 1.52945	3rd Qu.: 4.91095	3rd Qu.: 5.68036
Max. :1941		Max. :11718	Max. : 5.50678	Max. :84.00395	Max. :25.47478
				NA's : 6.00000	NA's :29.00000

```
> length(Daily$Q)
```

```
[1] 11718
```



## Reasons for the meta-data:

- Knowing what data you have and where they came from
- Putting labels on figures and tables
- Putting a name on your saved workspaces, using abbreviations

**> INFO<-getMetaData(sta,param)**

Your site for streamflow data is 01491000 .

Your site name is CHOPTANK RIVER NEAR GREENSBORO, MD ,but you can modify this to a short name in a style you prefer.

This name will be used to label graphs and tables.

If you want the program to use the name given above, just do a carriage return, otherwise enter the preferred short name(no quotes):

**Choptank River near Greensboro, MD**

The latitude and longitude of the site are: 38.99719 , -  
75.78581 (degrees north and west).

The drainage area at this site is 113 square miles which is  
being stored as 292.6687 square kilometers.

It is helpful to set up a station abbreviation when doing  
multi-site studies, enter a unique id (three or four  
characters should work).

It is case sensitive. Even if you don't feel you need an  
abbreviation for your site you need to enter something (no  
quotes):

**Chop**

Your water quality data are for parameter number 00631 which has the name: ' Nitrate plus nitrite, water, filtered, milligrams per liter as nitrogen '.

Typically you will want a shorter name to be used in graphs and tables. The suggested short name is: ' Nitrate-nitrite '.

If you would like to change the short name, enter it here, otherwise just hit enter (no quotes):

**Nitrate, filtered, as N**

The units for the water quality data are: mg/l as N .

It is helpful to set up a constituent abbreviation when doing multi-constituent studies, enter a unique id (three or four characters should work something like tn or tp or NO3).

It is case sensitive. Even if you don't feel you need an abbreviation you need to enter something (no quotes):

**no3**

# This command works even if all of your data came from your own spreadsheet

**Nothing in the system will work if you don't have, at least:**

- **A site name**
- **A parameter name**
- **A site abbreviation**
- **A parameter abbreviation**

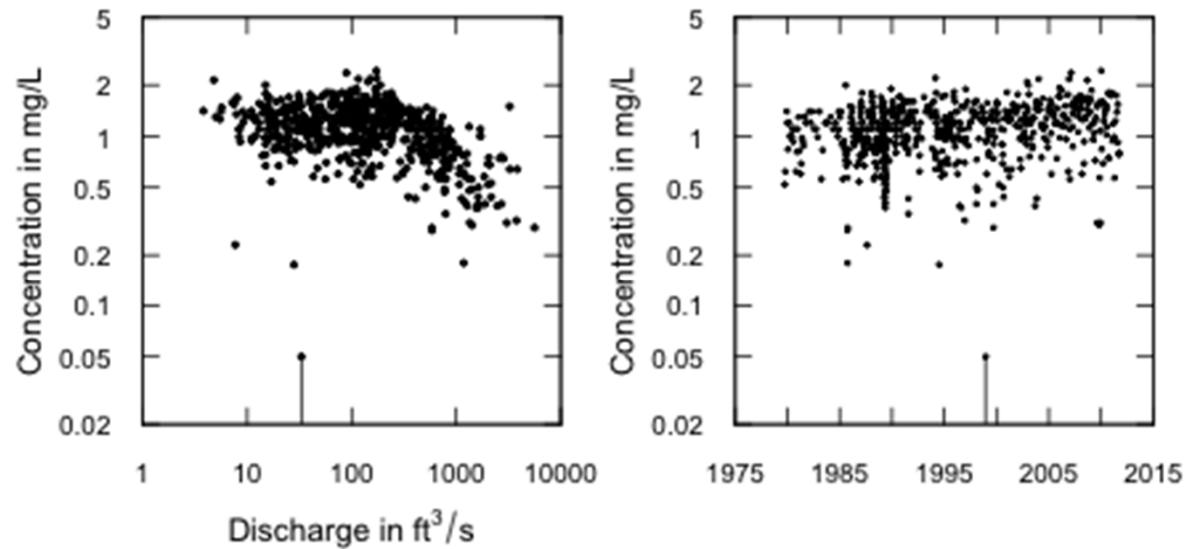
# Two more commands before we can start our analysis of the data

**> Sample<-mergeReport()**

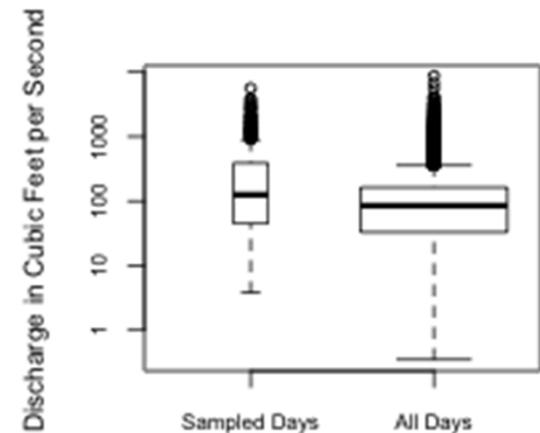
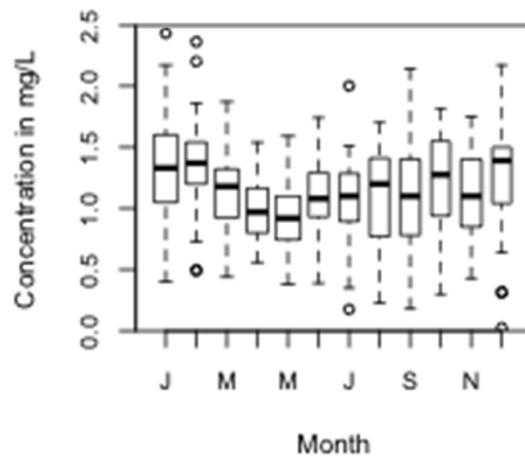
```
Discharge Record is 11718 days long, which is 32 years
First day of the discharge record is 1979-09-01 and last day is 2011-09-30
The water quality record has 652 samples
The first sample is from 1979-09-25 and the last sample is from 2011-09-29
Discharge: Minimum, mean and maximum 0.00991 4.08 246
Concentration: Minimum, mean and maximum 0.05 1.1 2.4
Percentage of the sample values that are censored is 0.15 %
```

Let's look at  
the data  
before we  
proceed, the  
function is:

Choptank River near Greensboro, MD  
Nitrate, filtered, as N



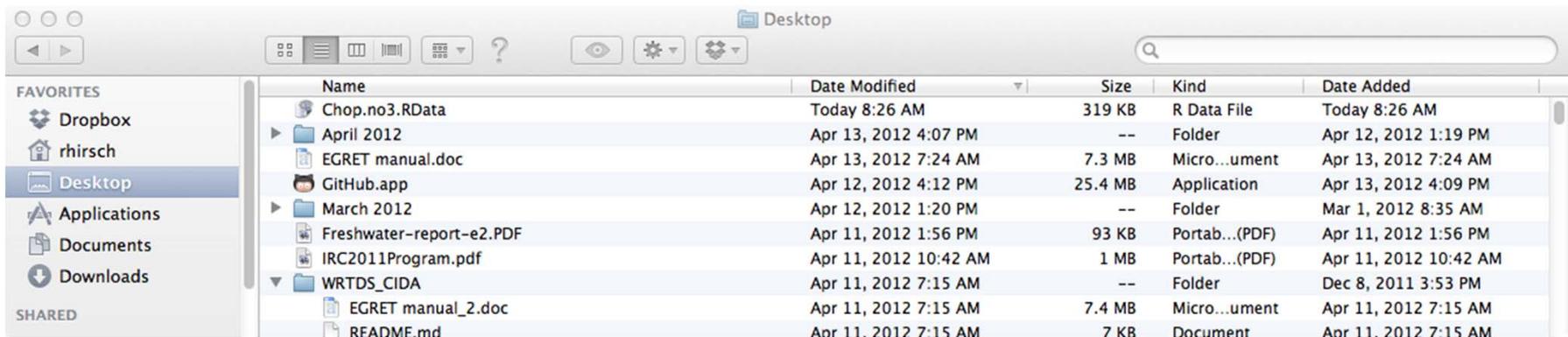
> multiPlotDataOverview(qUnit=1)





# We've gone to all this effort, let's save our work

- > `savePath<-" /Users/rhirsch/Desktop/"`
- > `saveResults(savePath)`



Chop.no3.RData

## Save it over and over as you proceed and add results

# **We now have 3 data frames**

- Sample (652 rows, 14 columns)**
- Daily (11,718 rows, 12 columns)**
- INFO (1 row, 42 columns)**

## **> modelEstimation()**

- **Jack-knife cross-validation of model**
- **Sets up the grid in  $\ln(Q)$  and time**
- **Estimates surfaces for  $\log(C)$ , standard error and Concentration**
- **Estimates the daily values from these surfaces (Conc, Flux, FNConc, FNFlux)**
- **Creates a summary of monthly results**

Various options are possible here, for window widths and minimum data requirements

# Setting up the “Period of Analysis”

- Could be water year
- Could be calendar year
- Could be April-May-June
- Could be Dec-Jan-Feb-Mar
- Could be only May

**paStart = calendar month that starts the Period**

**paLong = length of Period, in months**

# Setting up the “Period of Analysis”

	<b>paStart</b>	<b>paLong</b>
• Could be water year	<b>10</b>	<b>12</b>
• Could be calendar year	<b>1</b>	<b>12</b>
• Could be April-May-June	<b>4</b>	<b>3</b>
• Could be Dec-Jan-Feb-Mar	<b>12</b>	<b>4</b>
• Could be only May	<b>5</b>	<b>1</b>

**paStart = calendar month that starts the Period**

**paLong = length of Period, in months**

# Now what is in Daily?

> summary(Daily)

Date	Q	Julian	Month	Day	DecYear	MonthSeq
Min. :1979-09-01	Min. :9.911e-03	Min. :47359	Min. : 1.000	Min. : 1.0	Min. :1980	Min. :1557
1st Qu.:1987-09-08	1st Qu.:9.345e-01	1st Qu.:50288	1st Qu.: 4.000	1st Qu.: 92.0	1st Qu.:1988	1st Qu.:1653
Median :1995-09-15	Median :2.407e+00	Median :53218	Median : 7.000	Median :184.0	Median :1996	Median :1749
Mean :1995-09-15	Mean :4.082e+00	Mean :53218	Mean : 6.529	Mean :183.3	Mean :1996	Mean :1749
3rd Qu.:2003-09-22	3rd Qu.:4.616e+00	3rd Qu.:56147	3rd Qu.:10.000	3rd Qu.:274.0	3rd Qu.:2004	3rd Qu.:1845
Max. :2011-09-30	Max. :2.464e+02	Max. :59076	Max. :12.000	Max. :366.0	Max. :2012	Max. :1941

Qualifier	i	LogQ	Q7	Q30	Leap
Length:11718	Min. : 1	Min. :-4.61412	Min. : 0.01808	Min. : 0.09606	Min. : 1.0
Class :character	1st Qu.: 2930	1st Qu.: -0.06779	1st Qu.: 0.99109	1st Qu.: 1.17609	1st Qu.: 92.0
Mode :character	Median : 5860	Median : 0.87835	Median : 2.55661	Median : 2.87133	Median :184.0
	Mean : 5860	Mean : 0.76602	Mean : 4.08154	Mean : 4.08059	Mean :183.3
	3rd Qu.: 8789	3rd Qu.: 1.52945	3rd Qu.: 4.91095	3rd Qu.: 5.68036	3rd Qu.:274.0
	Max. :11718	Max. : 5.50678	Max. :84.00395	Max. :25.47478	Max. :365.0
			NA's : 6.00000	NA's :29.00000	

yHat	SE	ConcDay	FluxDay	FNConc	FNFlux
Min. :-1.534999	Min. :0.1347	Min. :0.230	Min. : 1.633	Min. :0.8208	Min. : 77.3
1st Qu.: -0.006091	1st Qu.:0.2186	1st Qu.:1.033	1st Qu.: 98.053	1st Qu.:1.0524	1st Qu.:168.3
Median : 0.130494	Median :0.2504	Median :1.195	Median : 248.578	Median :1.2042	Median :317.3
Mean : 0.117362	Mean :0.2678	Mean :1.194	Mean : 365.030	Mean :1.1975	Mean :361.5
3rd Qu.: 0.258278	3rd Qu.:0.3002	3rd Qu.:1.353	3rd Qu.: 482.075	3rd Qu.:1.3286	3rd Qu.:536.0
Max. : 0.659426	Max. :0.6129	Max. :1.962	Max. :5705.826	Max. :1.7017	Max. :928.3

# Compute AnnualResults

```
> AnnualResults<-setupYears(paLong=12,paStart=1)
```

```
> AnnualResults<-setupYears()
```

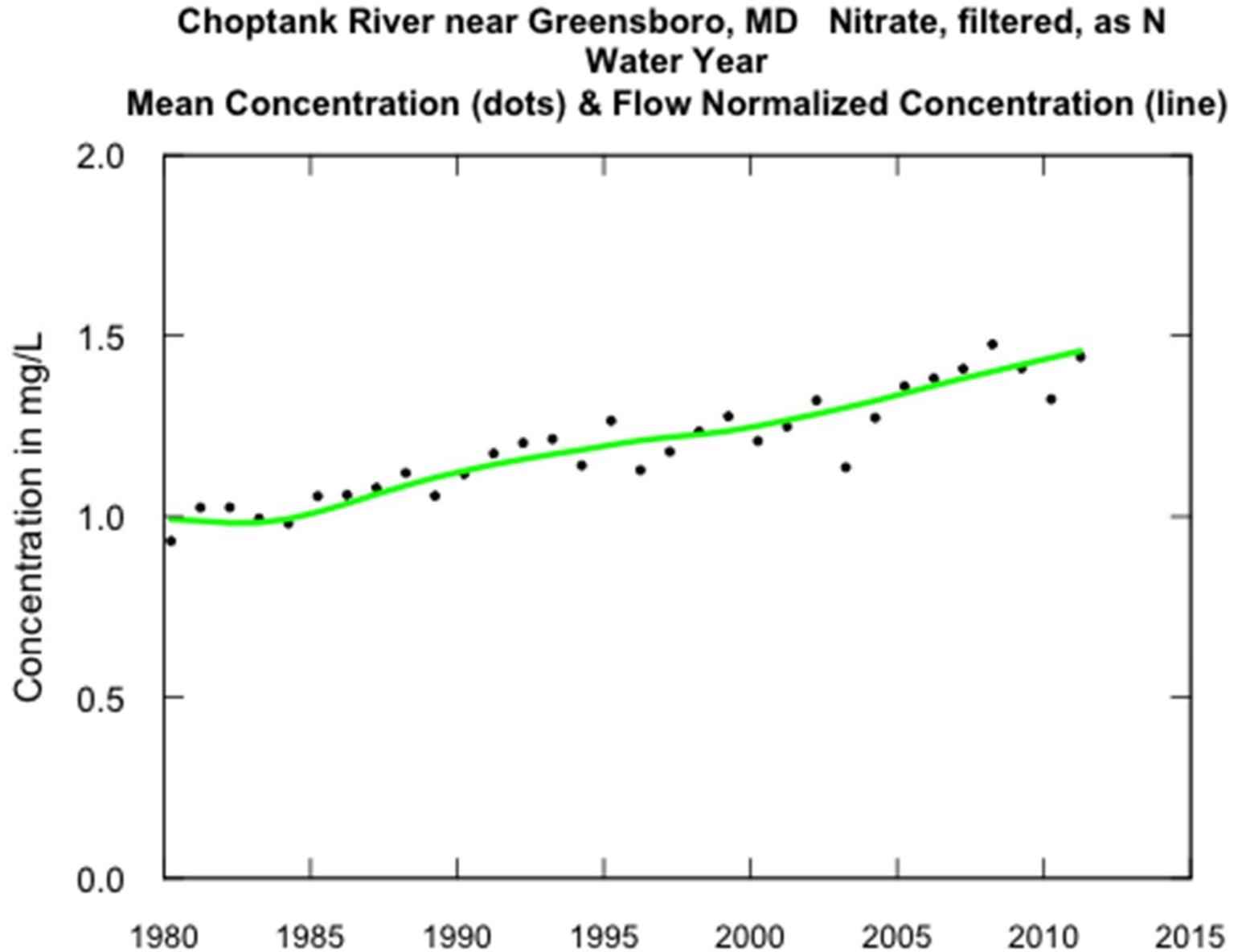
## **We now have 5 data frames**

- Sample (652 rows, 17 columns)**
- Daily (11,718 rows, 19 columns)**
- INFO (1 row, 53 columns)**
- MonthlyResults (385 rows, 7 columns)**
- AnnualResults (32 rows, 8 columns)**

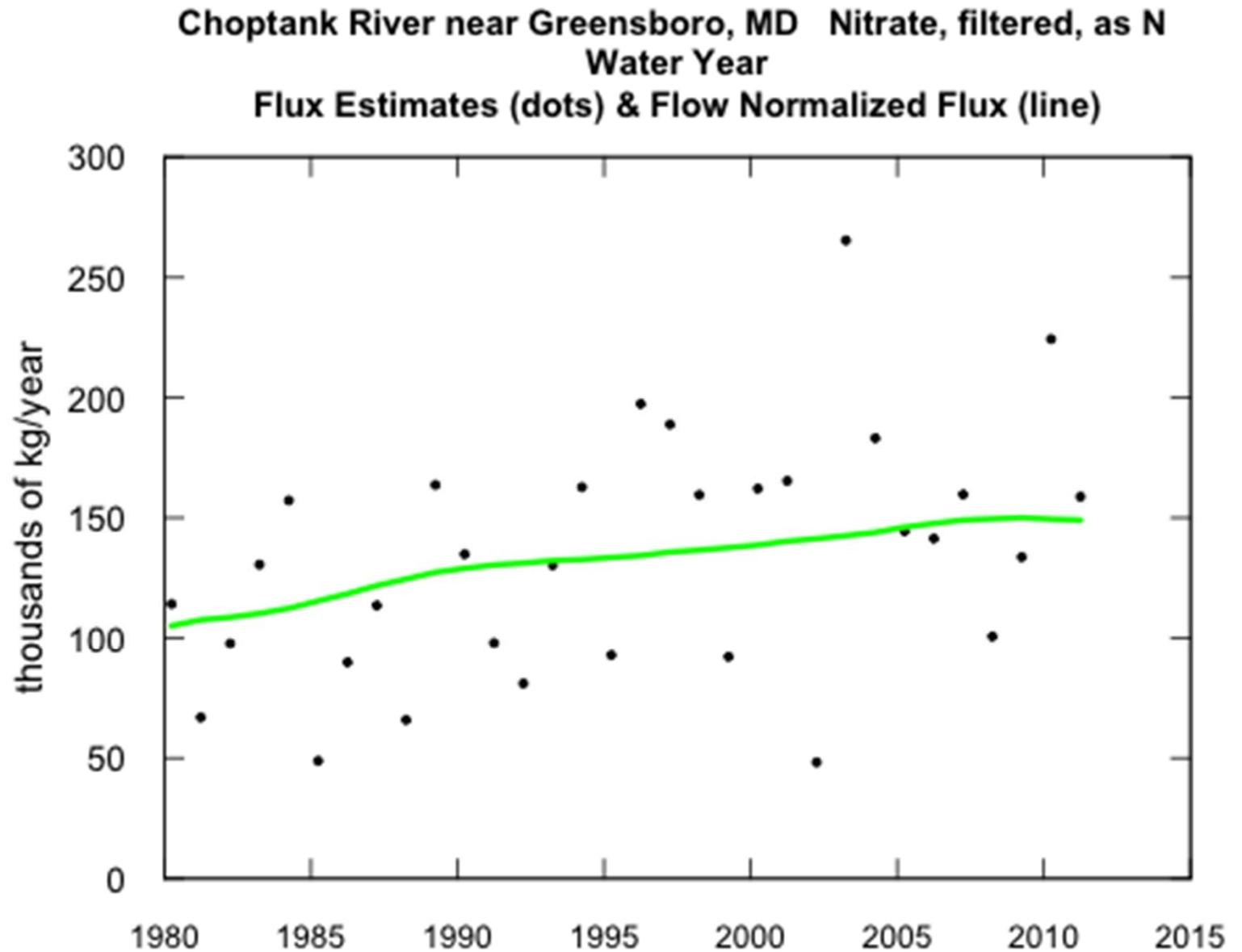
**And a matrix *surfaces* 14\*529\*3**



> plotConcHist(1980,2012)



```
> plotFluxHist(1980,2012,fluxUnit=8)
```



```
> tableResults(qUnit=1,fluxUnit=5)
```

Choptank River near Greensboro, MD

Nitrate, filtered, as N

Water Year

Year	Discharge cfs	Conc	FN_Conc mg/L	Flux tons/yr	FN_Flux
1980	150.2	0.932	0.992	126.0	116
1981	78.3	1.025	0.988	73.7	119
1982	107.6	1.025	0.982	107.9	120
1983	176.1	0.995	0.983	143.9	122
1984	201.9	0.981	0.994	173.3	124
1985	53.6	1.056	1.012	53.8	127
1986	92.8	1.060	1.036	99.0	131
1987	119.1	1.079	1.062	125.2	134
1988	66.0	1.120	1.086	72.6	137
1989	198.2	1.057	1.108	180.4	140
1990	141.5	1.118	1.126	148.7	142
1991	97.0	1.174	1.144	108.1	144
1992	77.2	1.204	1.158	89.2	145
.					
.					
.					
.					
2007	151.2	1.408	1.382	176.0	164
2008	90.5	1.476	1.401	111.1	165
2009	130.0	1.410	1.420	147.4	165
2010	254.0	1.324	1.438	247.1	165
2011	185.2	1.441	1.458	175.0	164

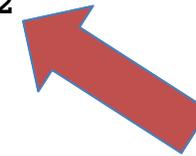
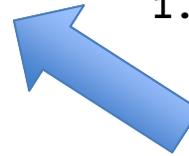
```
> tableChange(fluxUnit=5,yearPoints=c(1980,1995,2011))
```

```
> tableChange(fluxUnit=5,yearPoints=c(1980,1995,2011))
```

Choptank River near Greensboro, MD  
Nitrate, filtered, as N  
Water Year

Concentration trends						
time span			change	slope	change	slope
			mg/L	mg/L/yr	%	%/yr
1980	to	1995	0.21	0.014	21	1.4
1980	to	2011	0.47	0.015	47	1.5
1995	to	2011	0.26	0.016	22	1.4

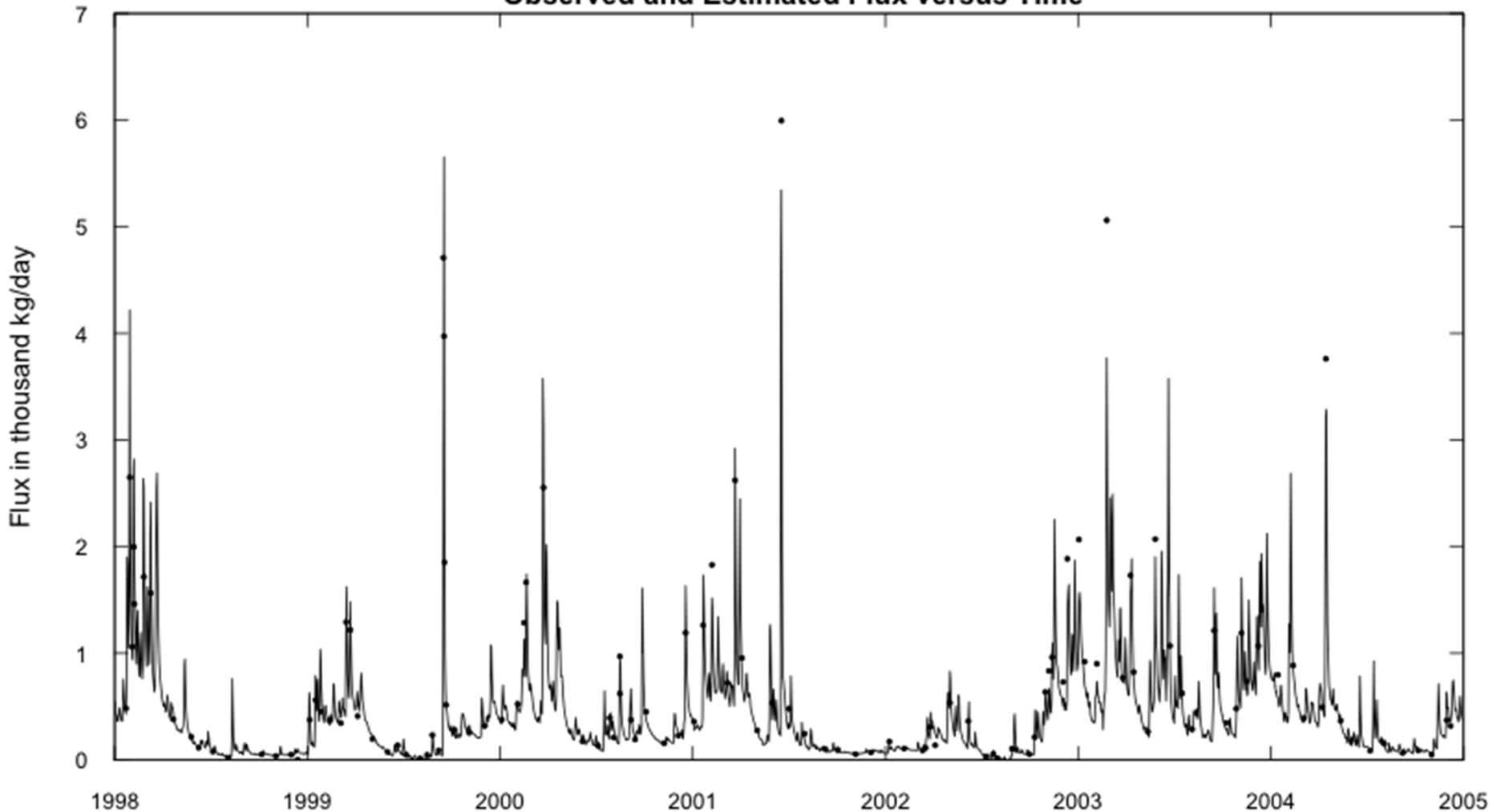
Flux Trends						
time span			change	slope	change	slope
			tons/yr	tons/yr/yr	%	%/yr
1980	to	1995	31	2.1	27	1.8
1980	to	2011	48	1.6	42	1.3
1995	to	2011	17	1.1	12	0.73



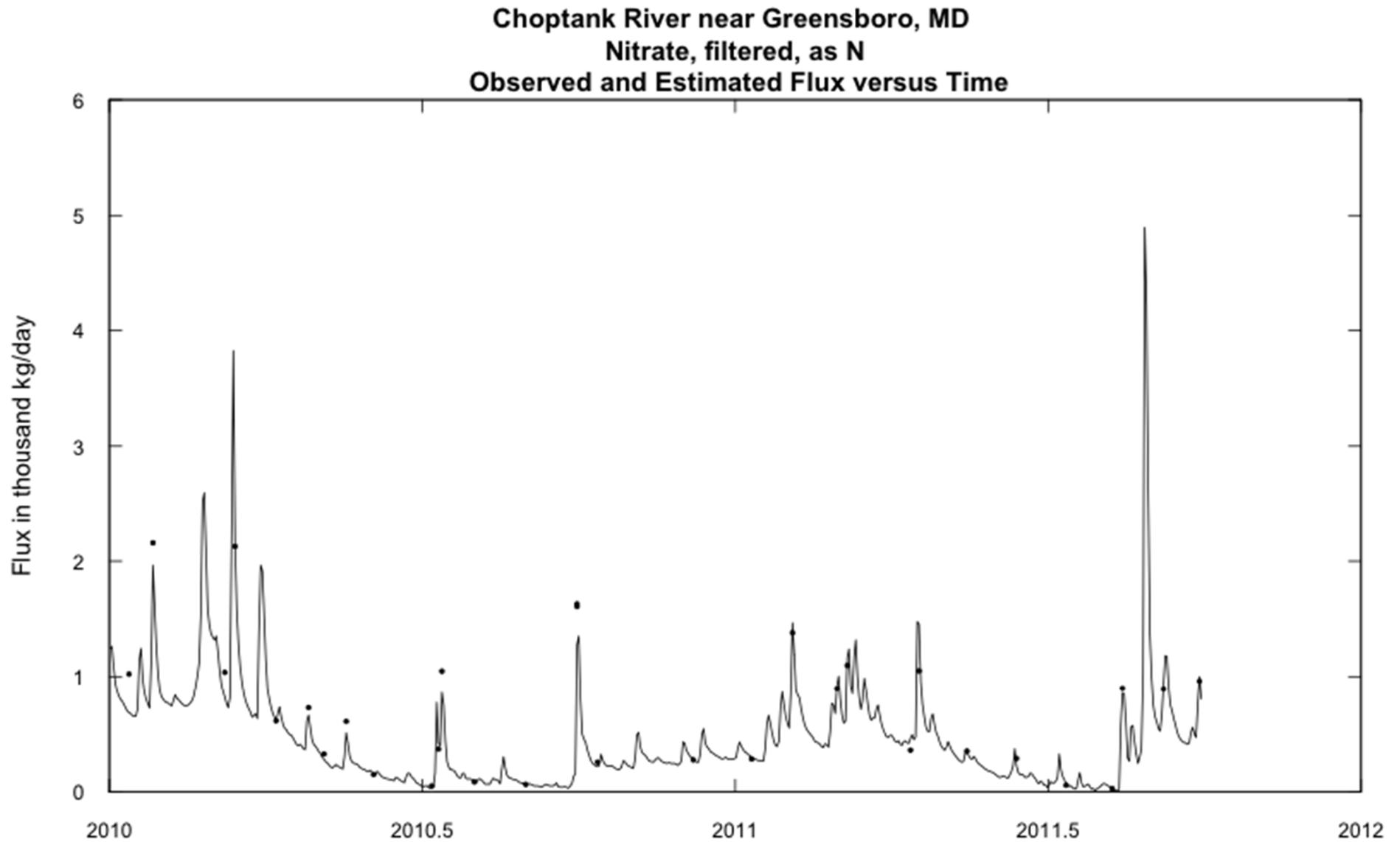
# Looking at the WRTDS model - Diagnostics

> plotFluxTimeDaily(1998,2005)

Choptank River near Greensboro, MD  
Nitrate, filtered, as N  
Observed and Estimated Flux versus Time



> plotFluxTimeDaily(2010,2011.75)



## More Diagnostics – Looking at quality of fit and flux bias

EGRET computes a flux bias statistic which is computed from the Sample data frame, using the Jack-Knife estimates of concentration. The statistic is:

$$\frac{\text{Mean}(\text{Sampled day flux estimates}) - \text{Mean}(\text{Sampled day flux values})}{\text{Mean}(\text{Sampled day flux values})}$$

For example a value of 0.25 would mean that the estimates are, on average, 25% too high.

Because of censoring we don't know the Sampled day flux values exactly, but we can put bounds on them.

It is rare that the censoring matters to this statistic.



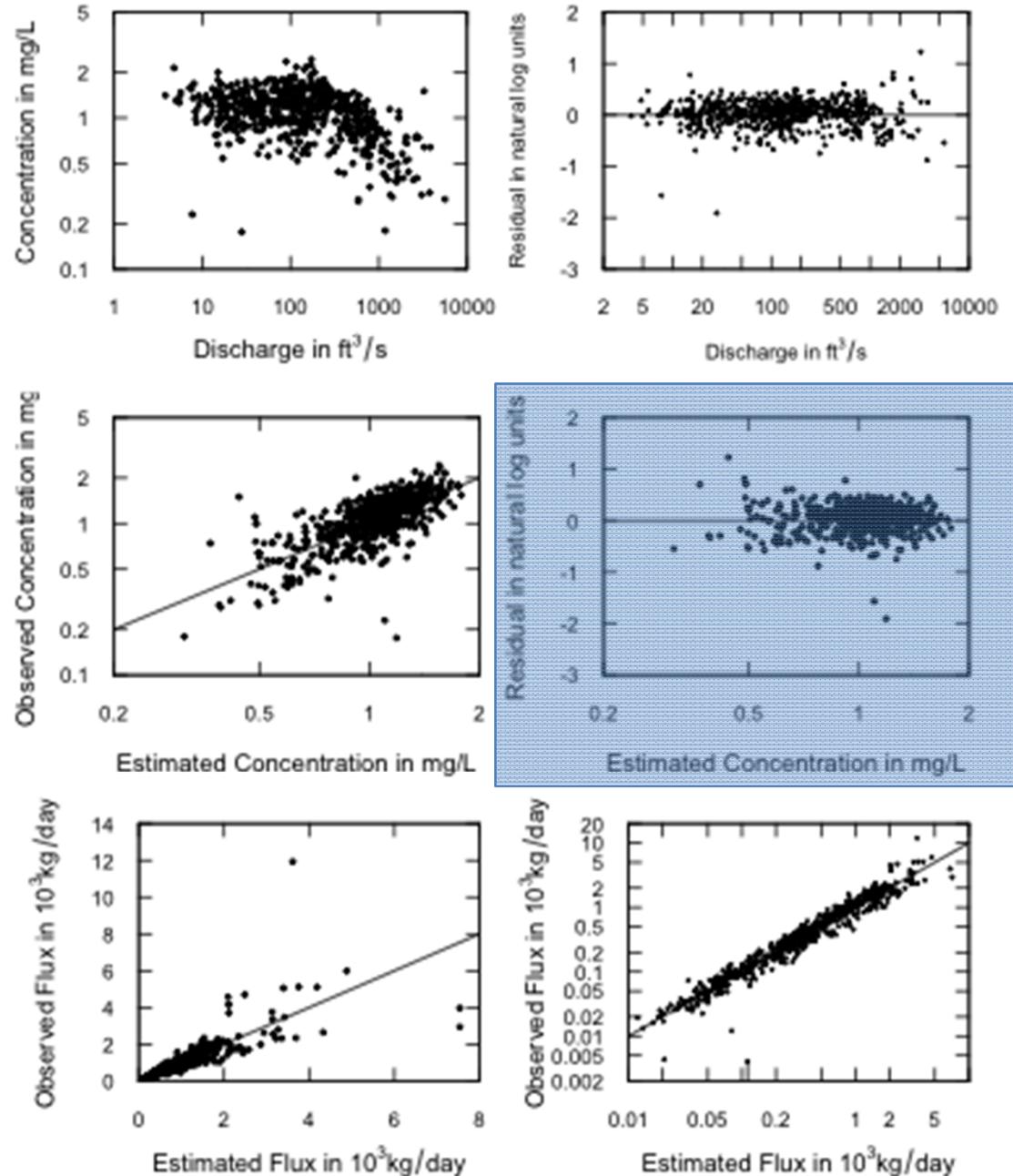
>  
fluxBiasMulti(qUnit=1,fluxUnit=4)

For this data set  
flux bias statistic  
looks fine.

Average estimates  
are about 2.5%  
below average  
observed.

This is very  
acceptable.

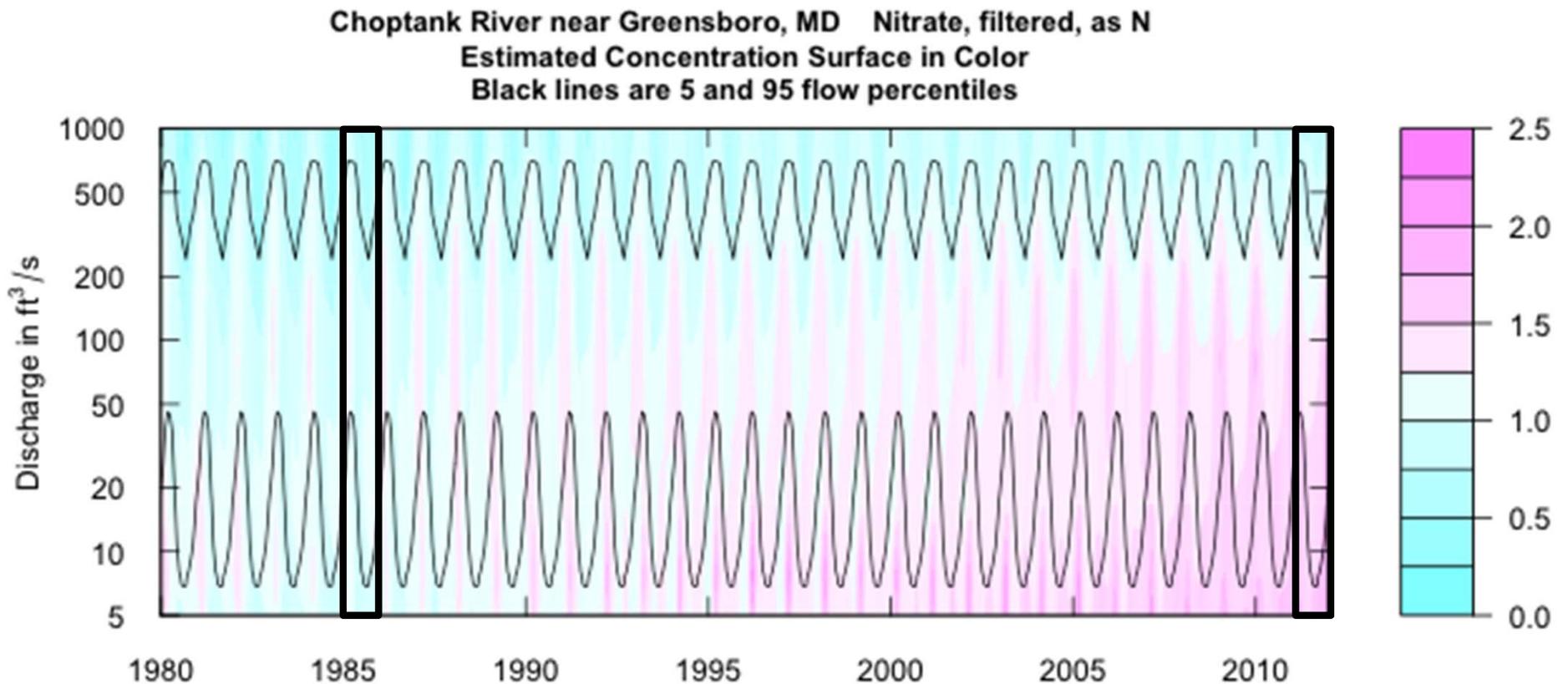
### Choptank River near Greensboro, MD Nitrate, filtered, as N Flux Bias Statistic -0.02492 ( -0.02492 , -0.02492 )



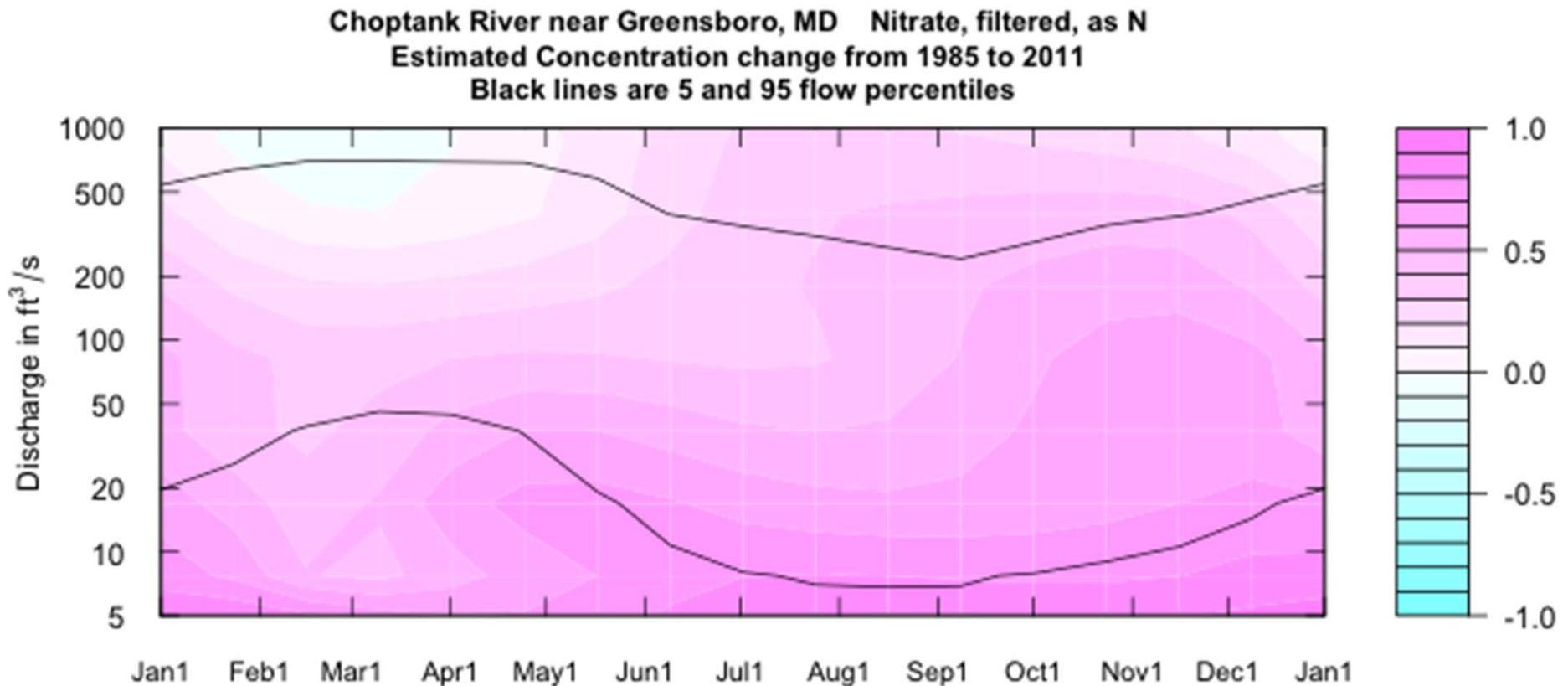
So, what does the model “look like”?

How has the system’s behavior changed over time?

```
> plotContours(1980,2012,5,1000,qUnit=1,contourLevels=seq(0,2.5,0.25))
```



> plotDiffContours(1985,2011,5,1000,qUnit=1,maxDiff=1.0)

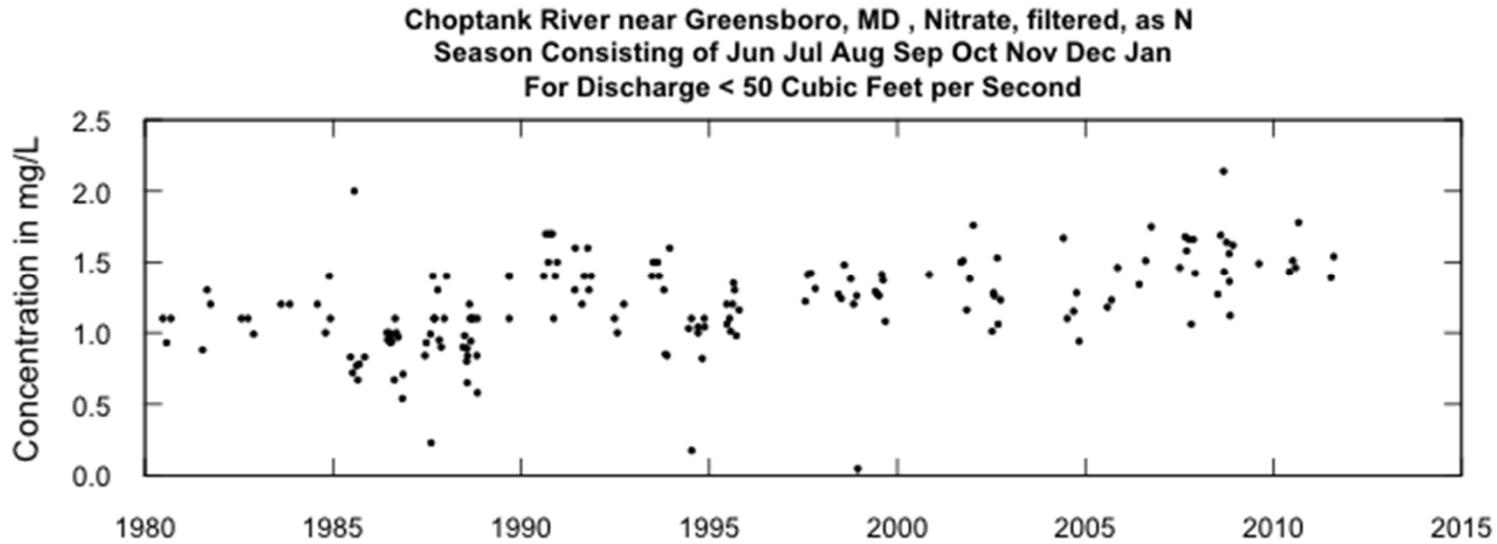


Increases at almost all seasons and flows, greatest increases are at lower flows in the months June through January

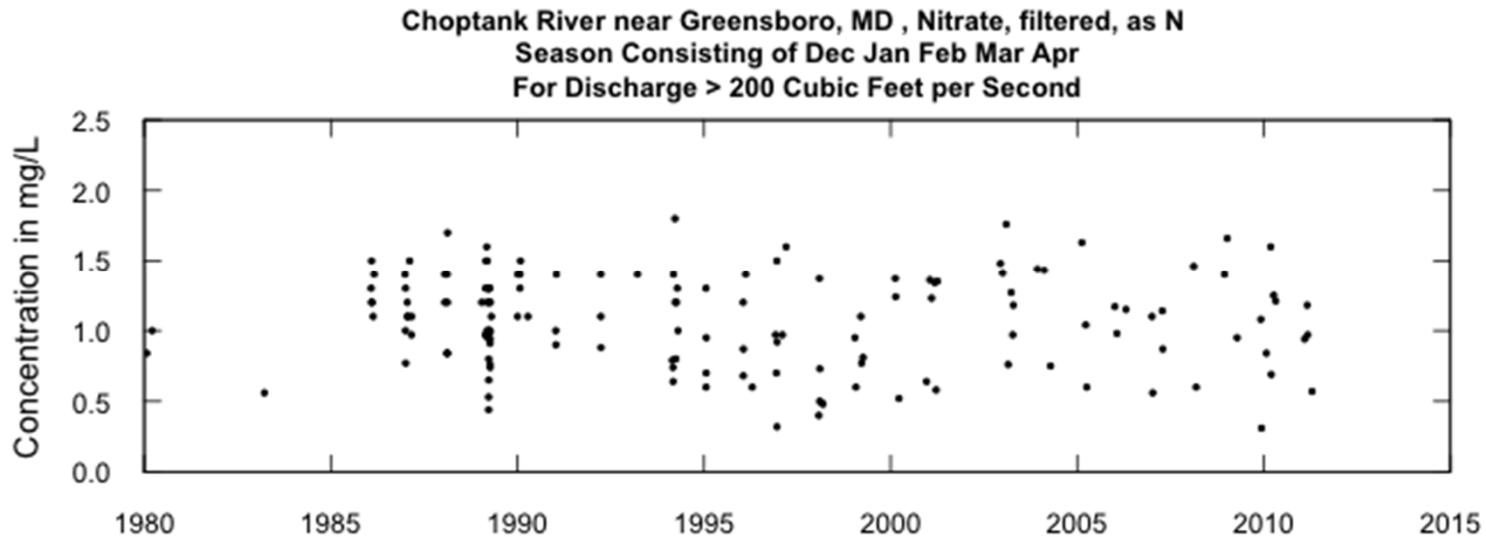
Very small improvement at very high flows in January through April

Now, can we look at *the data* without all the “statistical tricks?”

> plotConcTime(qUnit=1,qUpper=50,paLong=8,paStart=6,concMax=2.5)



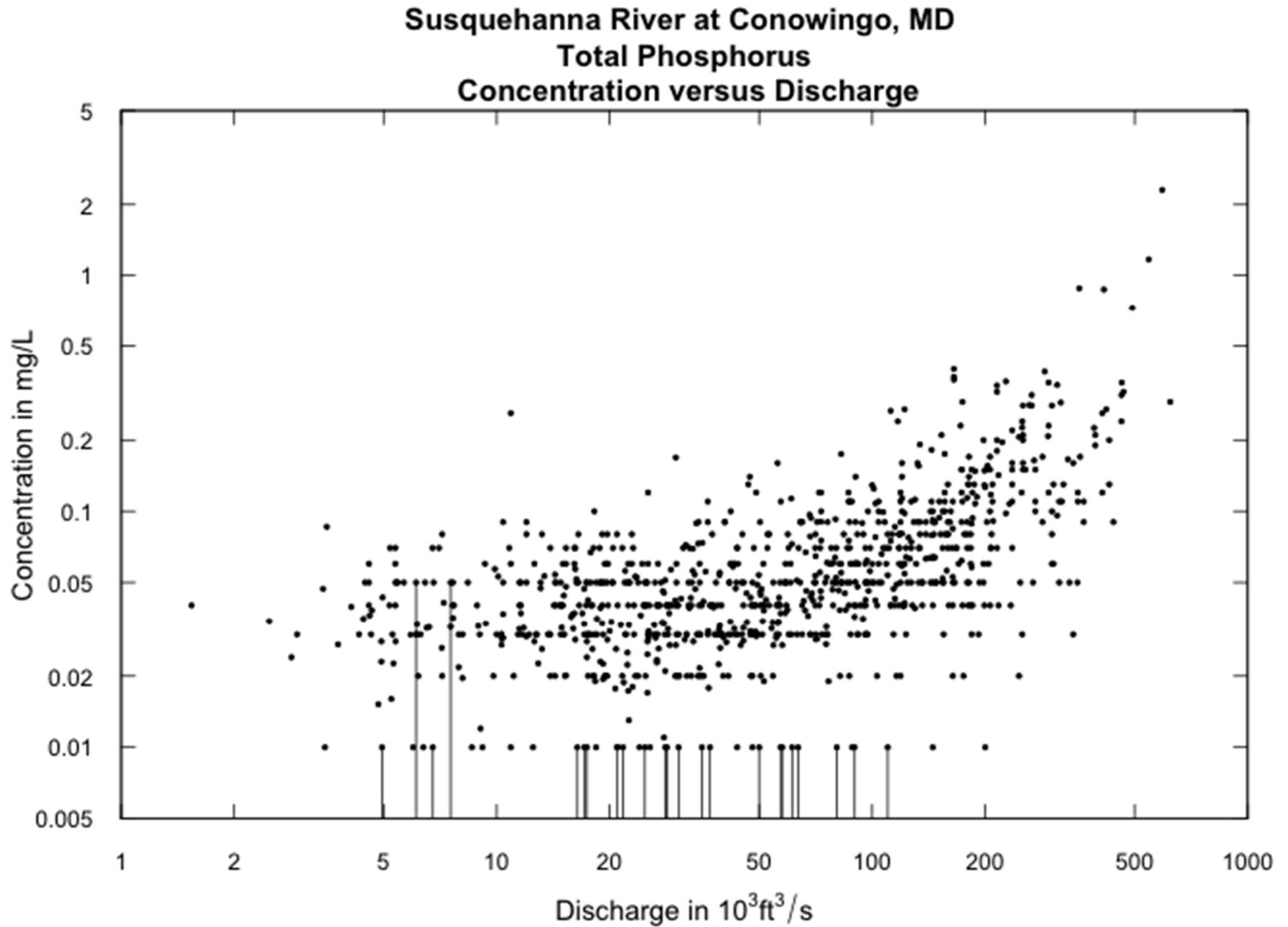
> plotConcTime(qUnit=1,qLower=200,paLong=5,paStart=12,concMax=2.5)



Look at a different data set:  
Total Phosphorous  
Susquehanna River at Conowingo, MD

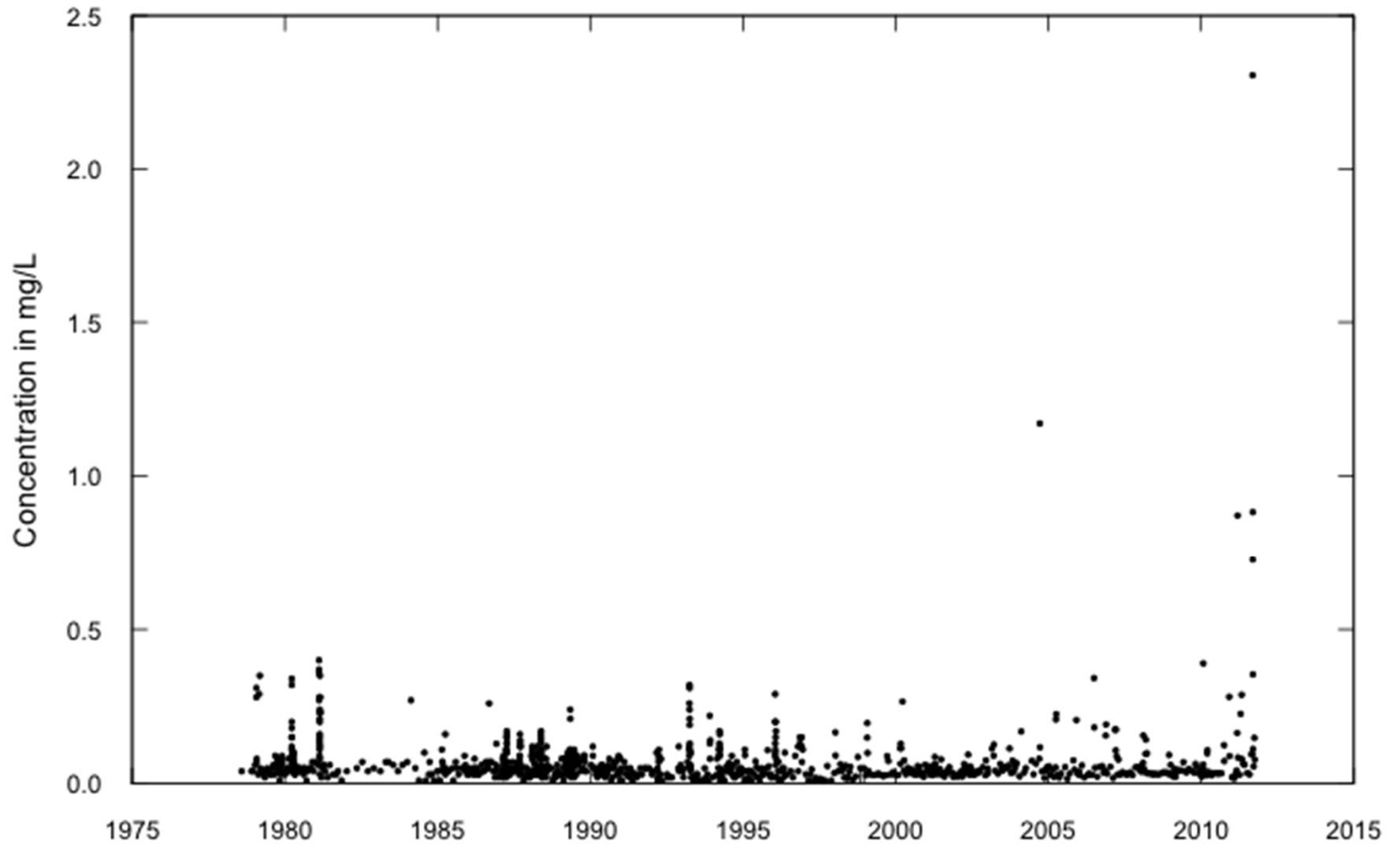
- Single largest river input to Chesapeake Bay
- Sampling at Conowingo Dam, reservoir capacity getting nearly full of sediment

> plotLogConcQ(concMin=0.005,qUnit=3)

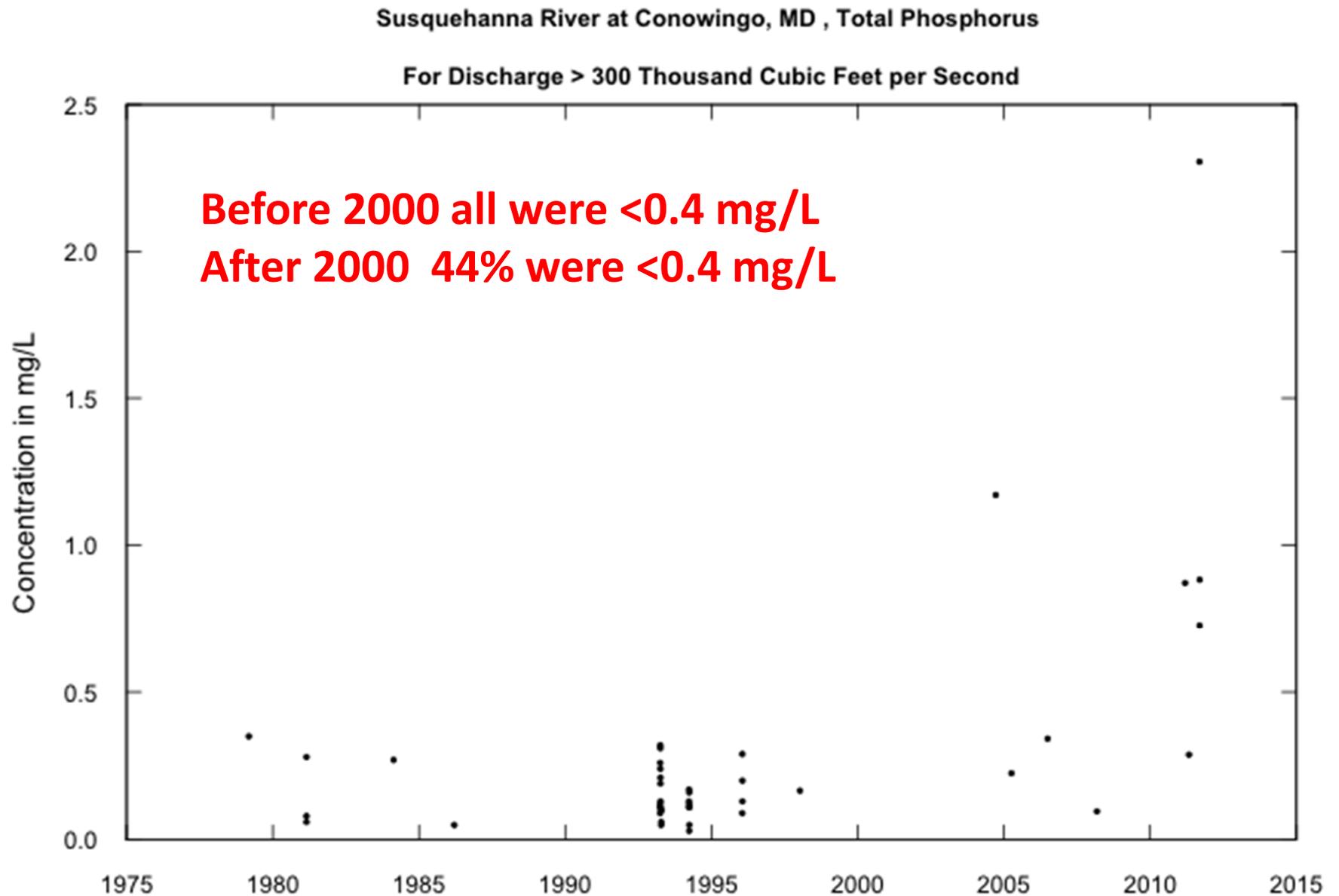


> plotConcTime()

Susquehanna River at Conowingo, MD , Total Phosphorus

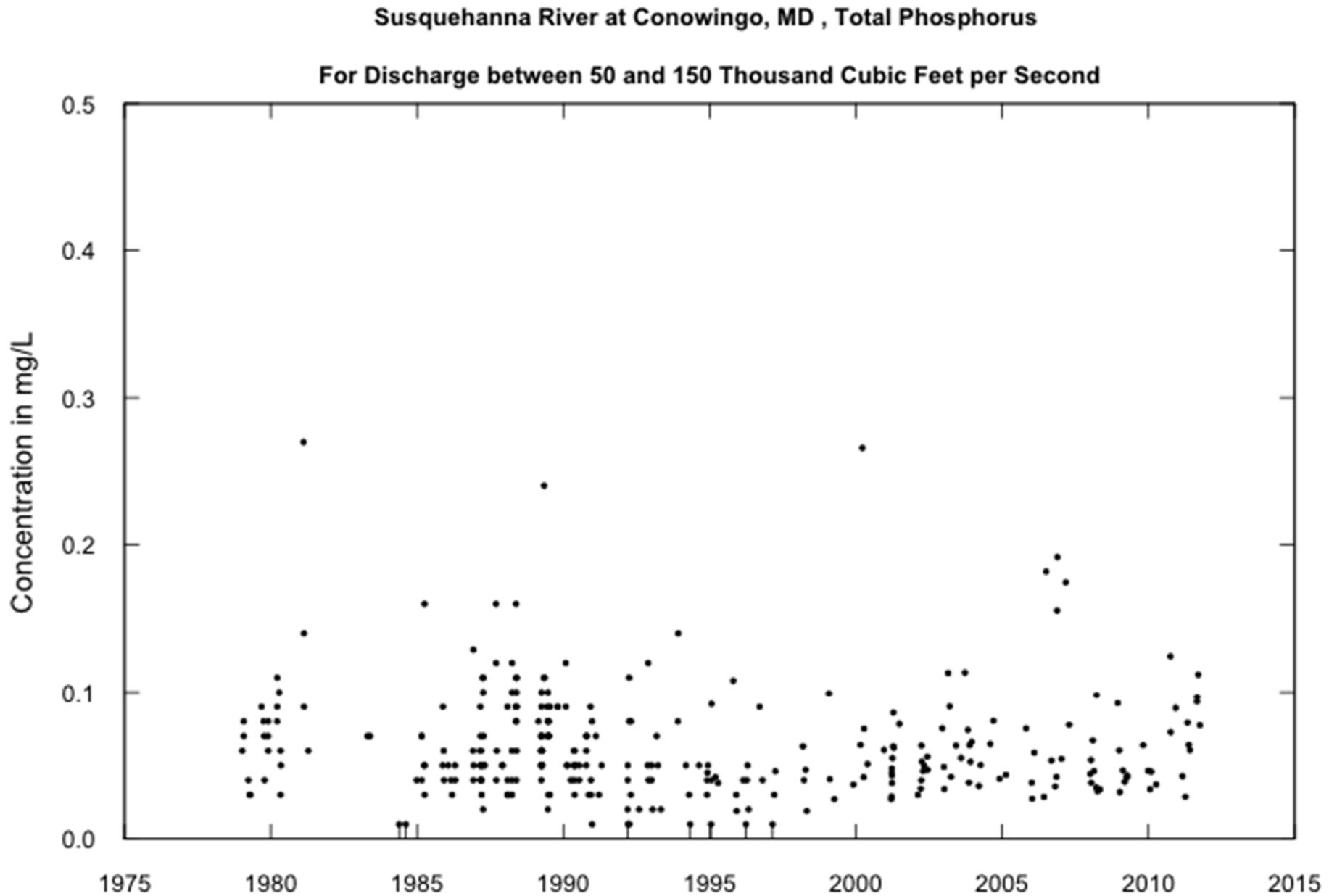


```
> plotConcTime(concMax=2.5,qLower=300,qUnit=3)
```

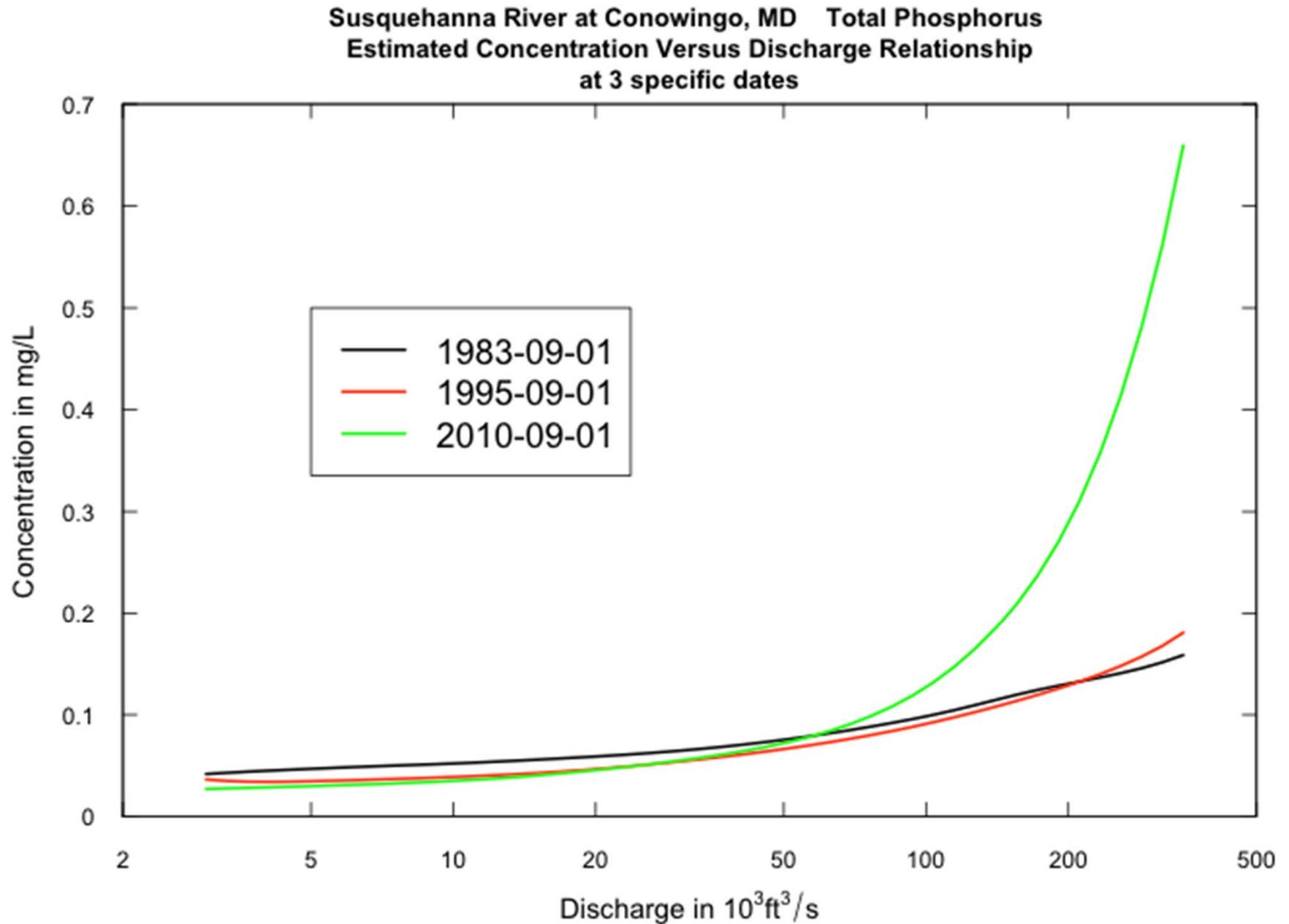




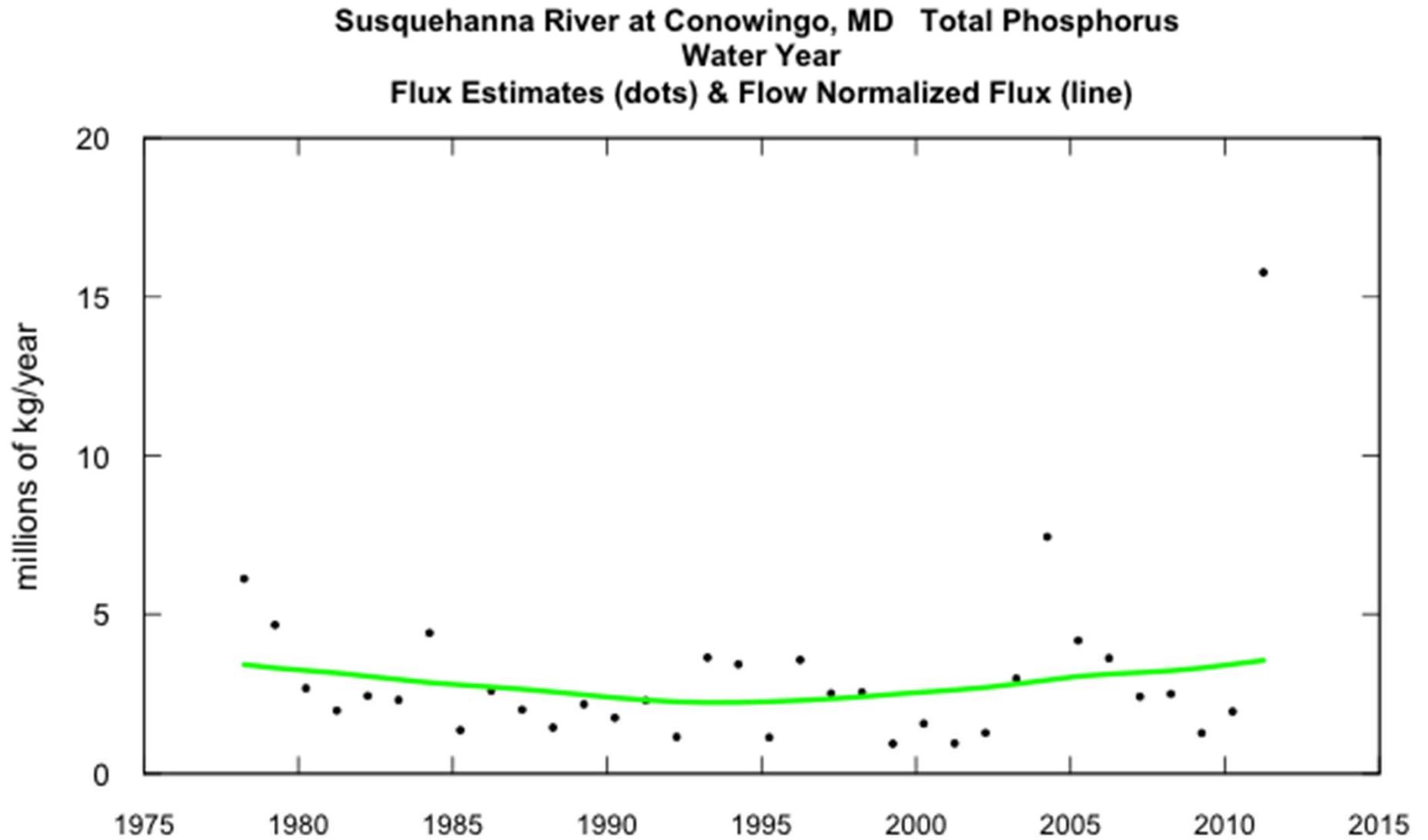
```
> plotConcTime(concMax=0.5,qLower=50,qUpper=150,qUnit=3)
```



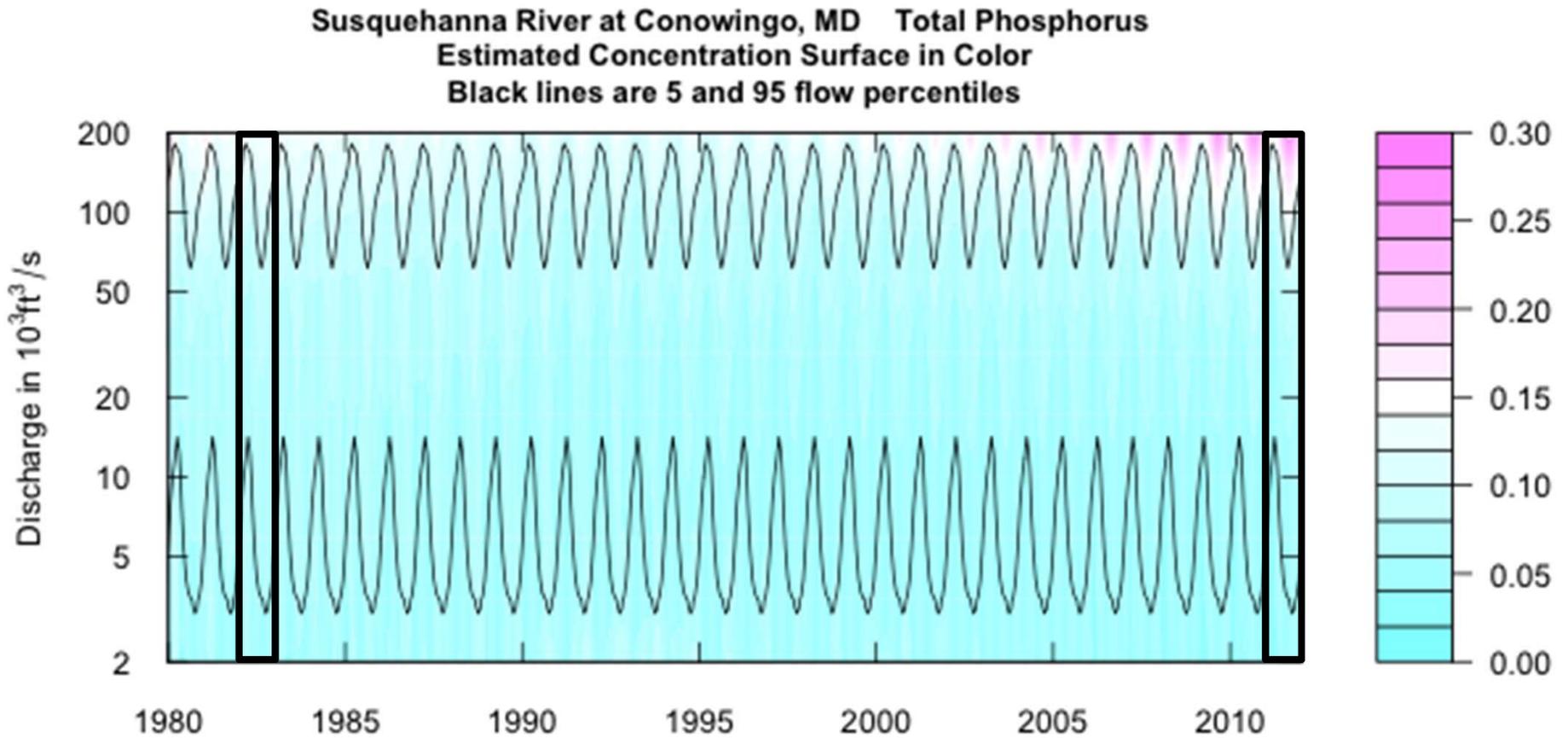
```
> plotConcQSmooth("1983-09-01","1995-09-01","2010-09-01",qLow=3,qHigh=350,  
qUnit=3,legendLeft=5,legendTop=0.5)
```



> plotFluxHist(1975,2012)

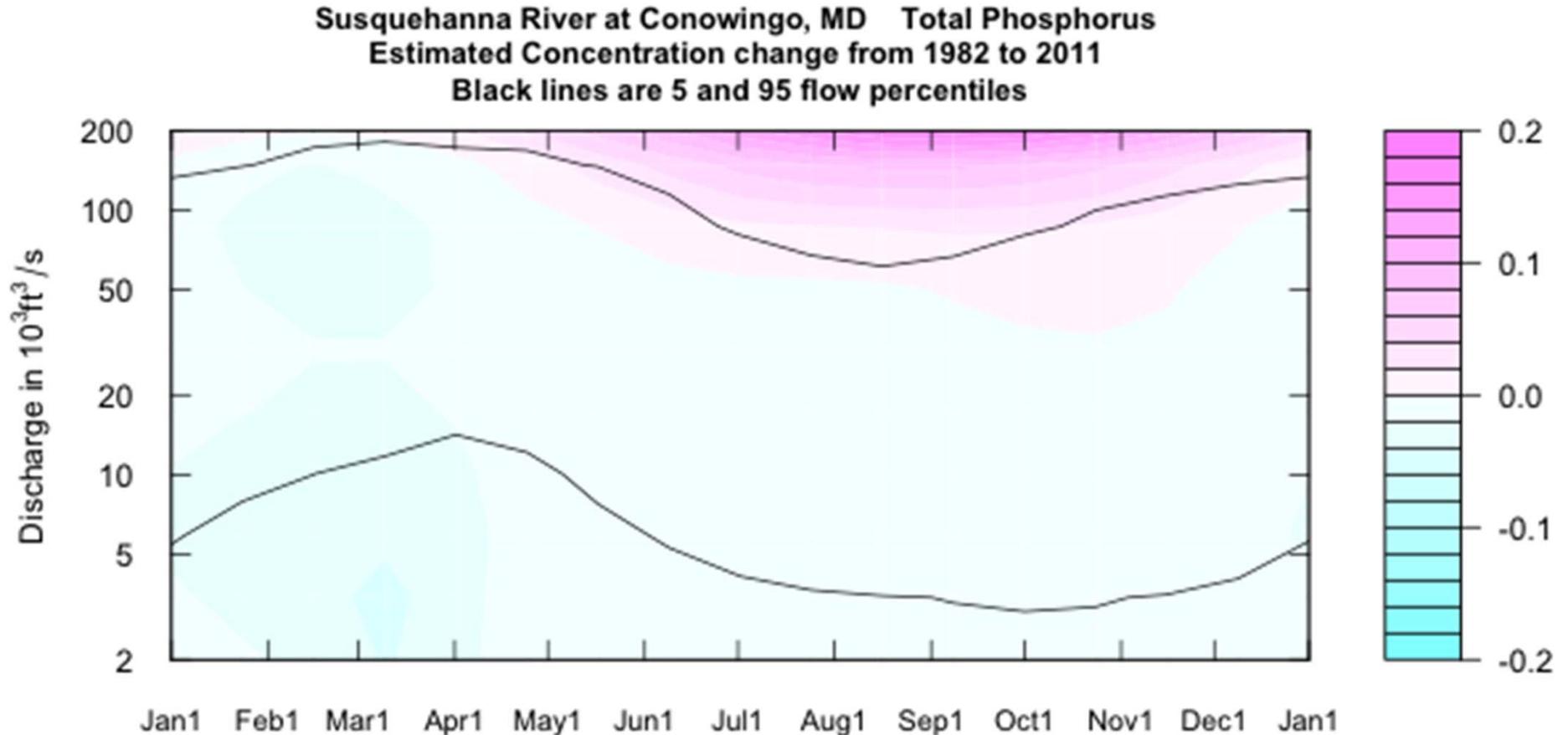


```
> plotContours(1980,2012,2,200,qUnit=3,contourLevels=seq(0,0.3,0.02))
```

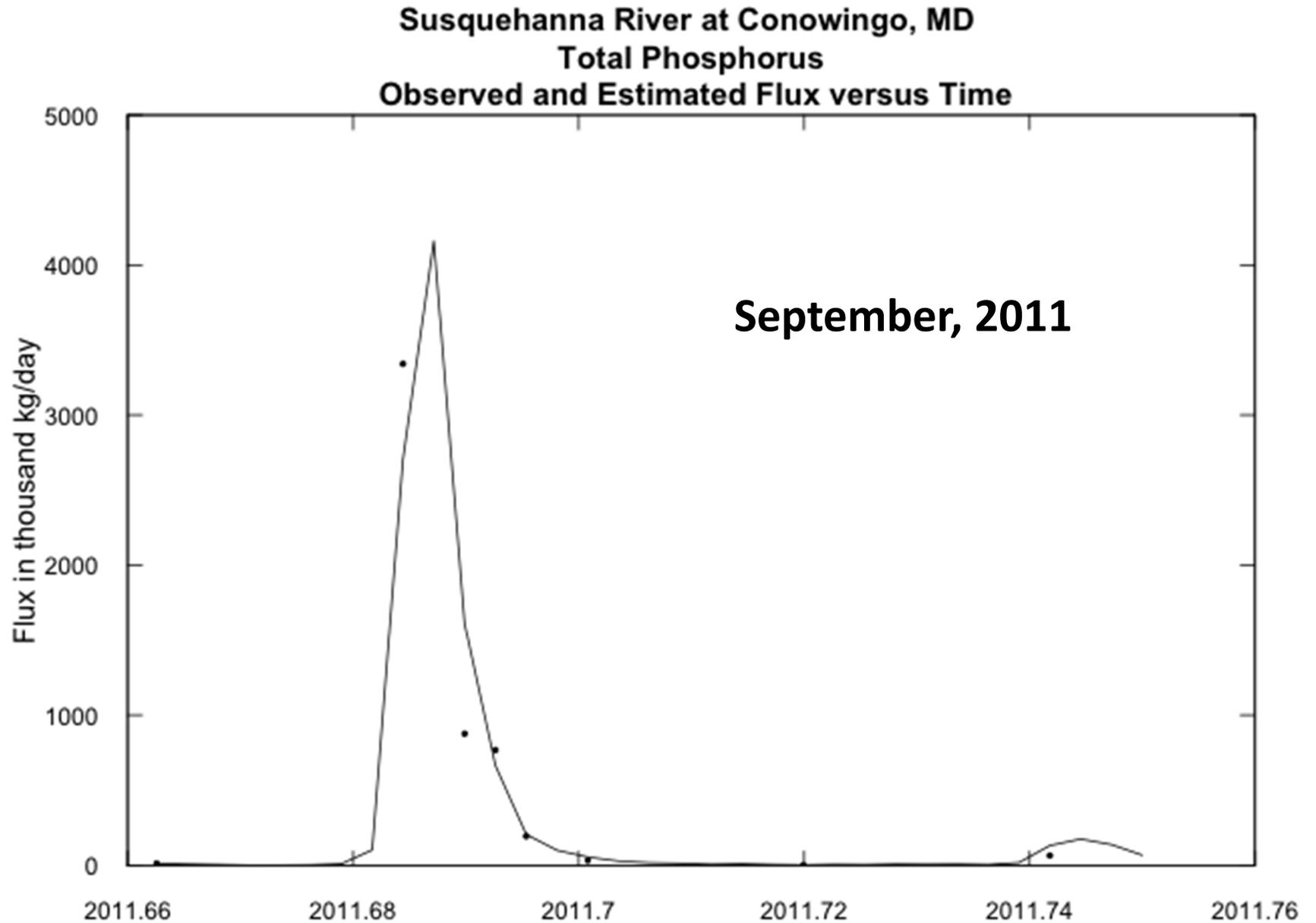


> plotDiffContours(1982,2011,2,200,qUnit=3,maxDiff=0.2)

- Over most flows and most seasons there has been little change between 1982 and 2011
- But, for discharge above about 100,000 cfs there are substantial increases, especially in the tropical storm season.
- Slight improvement indicated in winter at most flows.



> plotFluxTimeDaily(2011.66,2011.75)



**Because we have the Daily data frame, with an estimated flux for every day of the record, we can look at it to calculate things like:**

- Tropical Storm Lee flux, about 10,600 tons**
- Water Year 2011 flux, about 17,400 tons**
- Average flux over past decade about 4,800 tons/year**
- Average flux over entire 34 year record about 3,300 tons/year**

### **Tropical Storm Lee carried**

- 12% of 2011 streamflow and 61% of 2011 Total Phosphorus**
- 1.8% of the past decade's streamflow and 22% of the Total Phosphorus**
- 0.6% of the streamflow of the 34 years and 9% of the Total Phosphorus**

```
> tableChange(fluxUnit=8,yearPoints=c(1980,1995,2011))
```

Susquehanna River at Conowingo, MD  
Total Phosphorus  
Water Year

Concentration trends

time span			change	slope	change	slope
			mg/L	mg/L/yr	%	%/yr
1980	to	1995	-0.014	-0.00095	-24	-1.6
1980	to	2011	-0.0098	-0.00031	-17	-0.54
1995	to	2011	0.0046	0.00028	10	0.64

Flux Trends

time span			change	slope	change	slope
			10 <sup>3</sup> kg/yr	10 <sup>3</sup> kg/yr /yr	%	%/yr
1980	to	1995	-977	-65	-30	-2
1980	to	2011	314	10	9.7	0.31
1995	to	2011	1291	81	57	3.6





```
> AnnualResults<-setupYears(paLong=3,paStart=4)
> tableChange(fluxUnit=8,yearPoints=c(1980,1995,2011))
```

Susquehanna River at Conowingo, MD  
 Total Phosphorus



Concentration trends

time span			change mg/L	slope mg/L/yr	change %	slope %/yr
1980	to	1995	-0.016	-0.001	-27	-1.8
1980	to	2011	-0.0045	-0.00015	-7.9	-0.26
1995	to	2011	0.011	0.00069	26	1.7

Flux Trends

time span			change 10 <sup>3</sup> kg/yr	slope 10 <sup>3</sup> kg/yr /yr	change %	slope %/yr
1980	to	1995	-972	-65	-27	-1.8
1980	to	2011	529	17	15	0.47
1995	to	2011	1501	94	57	3.6



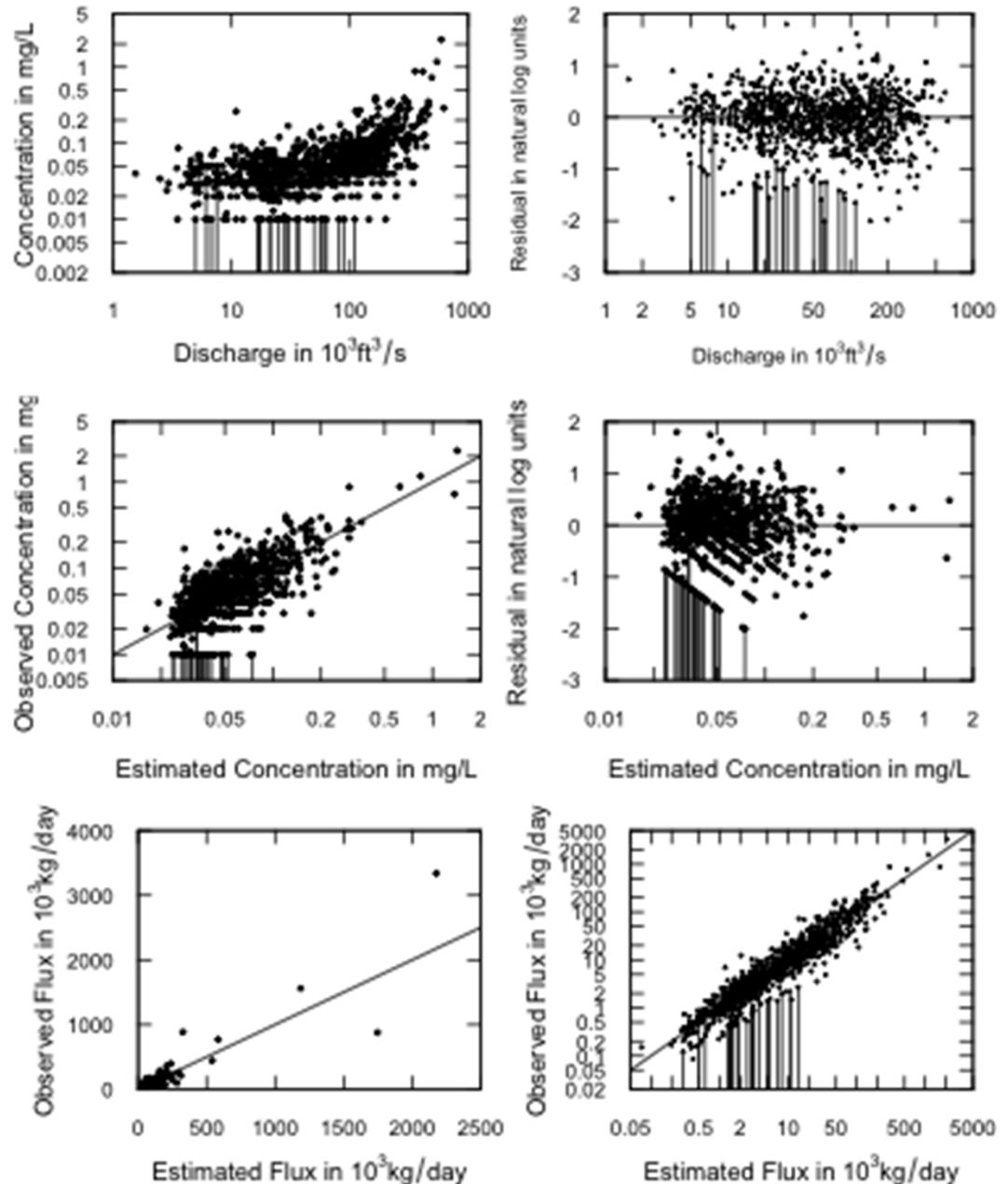
>  
fluxBiasMulti(qUnit=3,  
fluxUnit=4,  
moreTitle="WRTDS")

A short aside about the “flux bias” problem

Models such as WRTDS or LOADEST are fit in log space but used in “real” space.

None of them should be used without checking that the model behavior is reasonably close to actual behavior.

Susquehanna River at Conowingo, MD Total Phosphorus  
Flux Bias Statistic -0.004883 ( -0.005287 , -0.00448 ) WRTDS

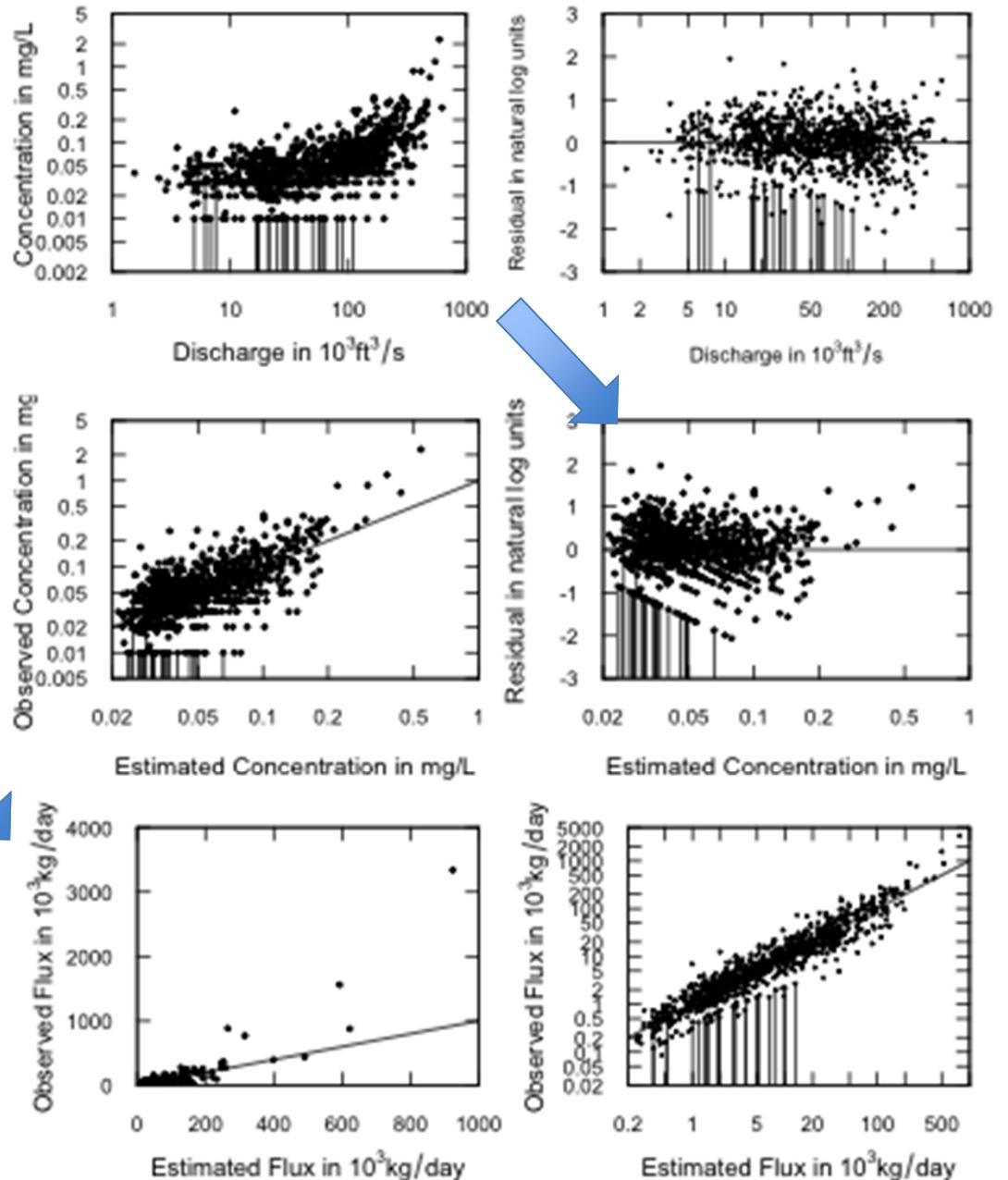


Average flux bias for Loadest on sampled days is about -16%

Notice the U shape of residuals versus predicted

Notice the severe under-prediction of flux on the highest flux days

Susquehanna River at Conowingo, MD Total Phosphorus Flux Bias Statistic -0.1586 ( -0.1591 , -0.1582 )

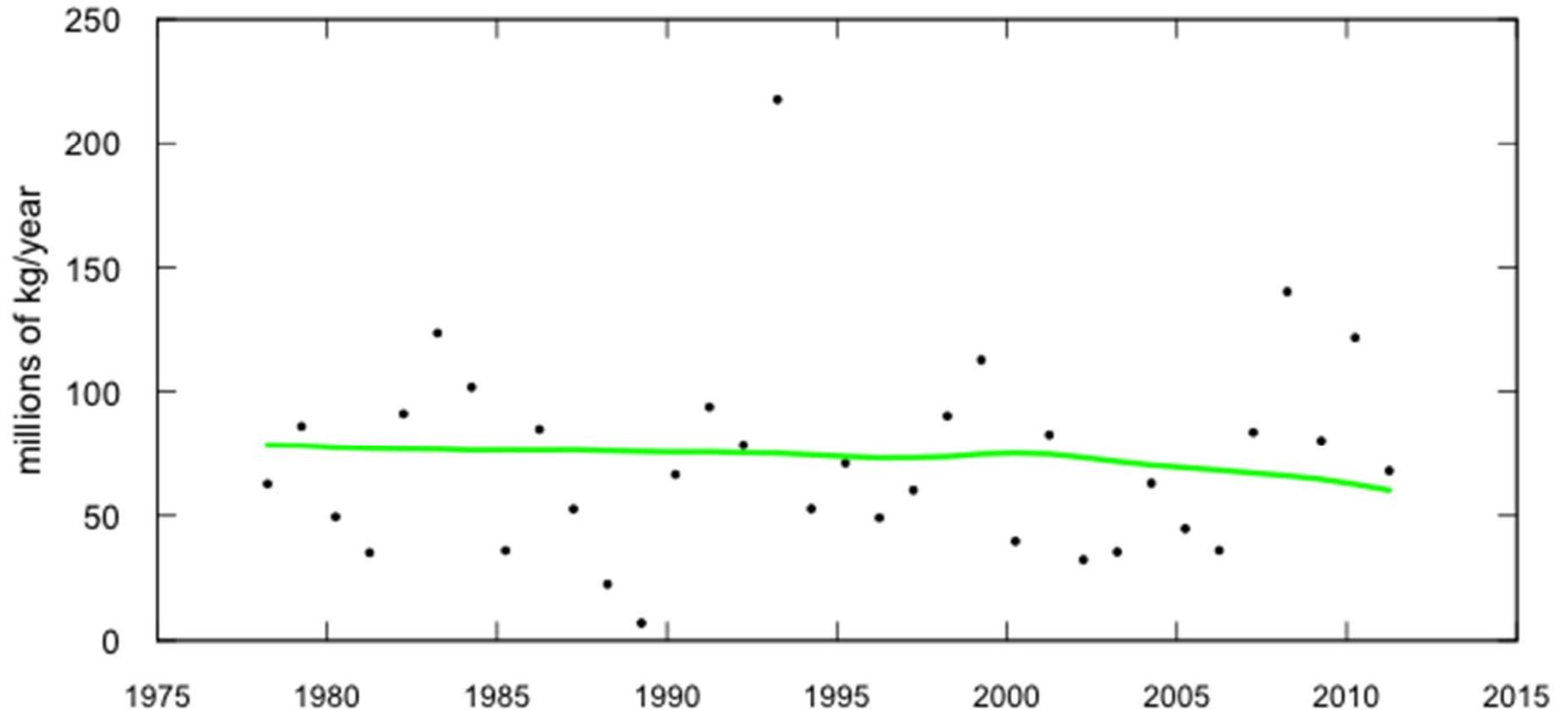


**Let's move to the corn belt:**

**Iowa River at Wapello, IA**

**Total Nitrogen**

Iowa River at Wapello, IA Total Nitrogen  
Water Year  
Flux Estimates (dots) & Flow Normalized Flux (line)



**Pretty good news, Flow Normalized Flux**  
**1980 – 1993 decreasing about 0.2%/year**  
**1993 – 2011 decreasing about 1.1%/year**

> flowDuration()

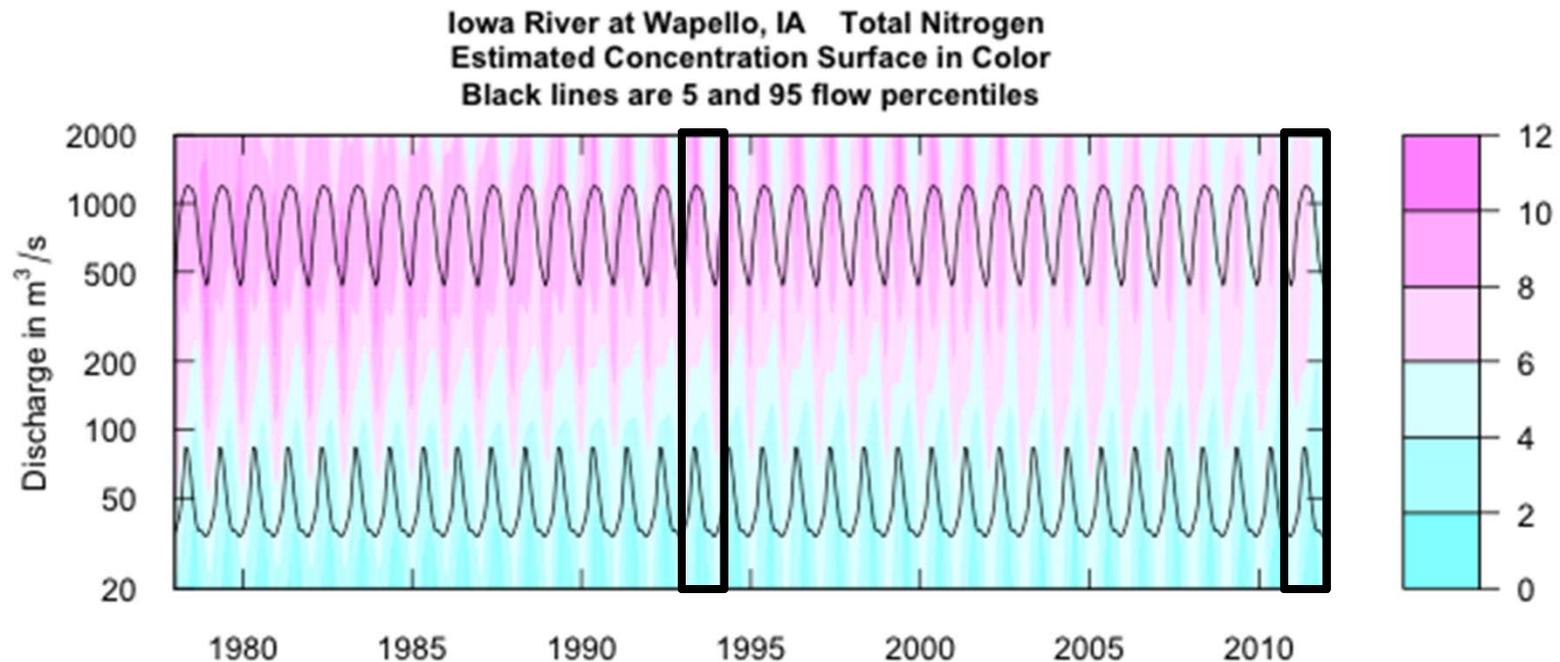
Flow Duration for Iowa River at Wapello, IA

Flow duration is based on full year

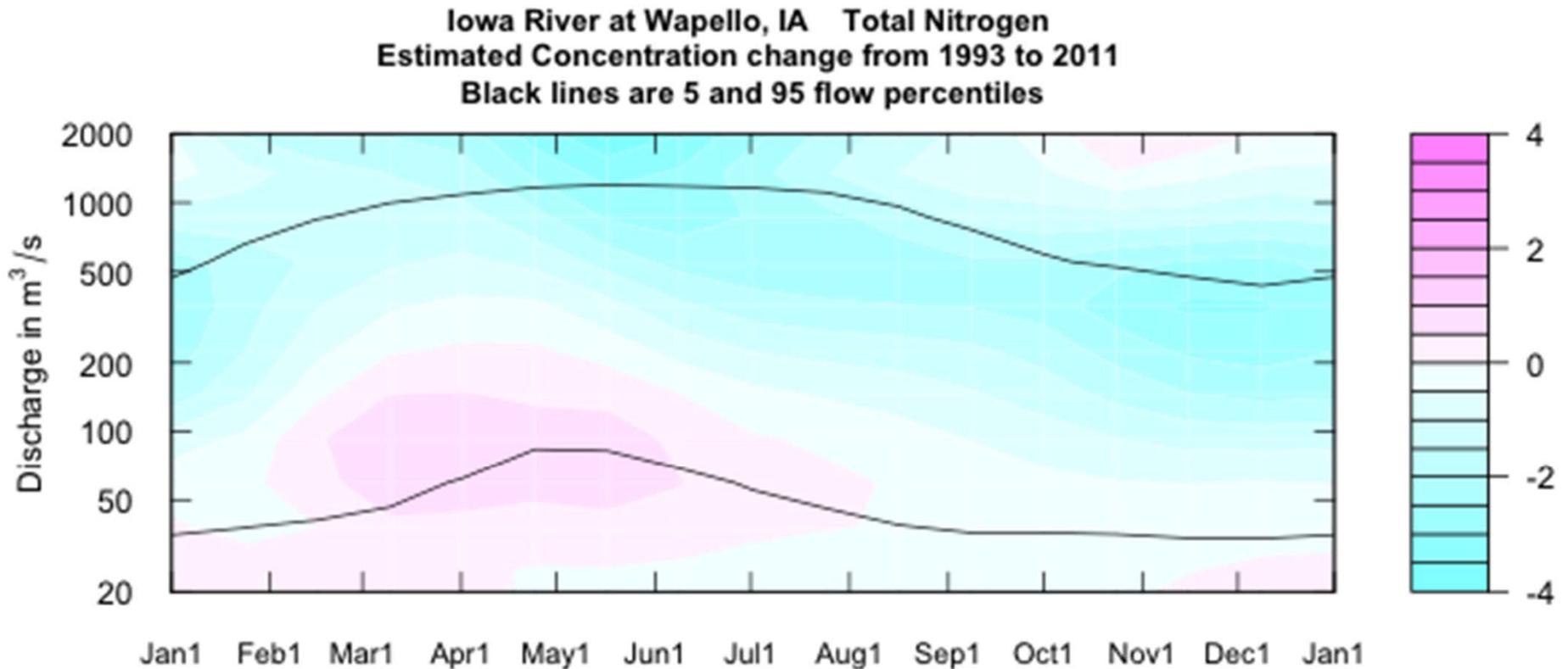
Discharge units are Cubic Meters per Second

min	5%	10%	25%	50%	75%	90%	95%	max
16.4	44.5	68.0	104.5	194.0	390.8	690.9	900.5	4870.5

> plotContours(1978,2012,20,2000,contourLevels=seq(0,12,2))



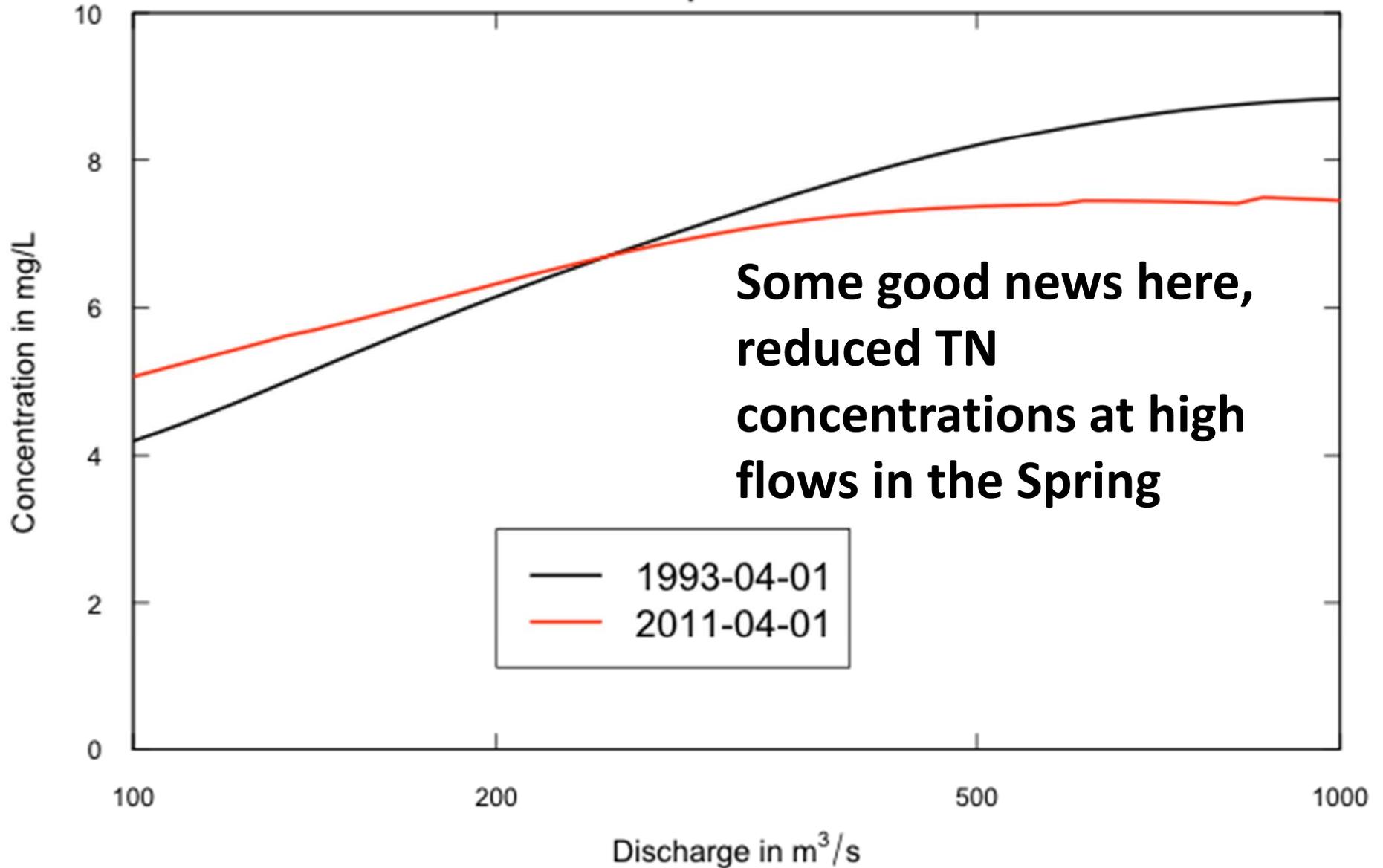
> plotDiffContours(1993,2011,20,2000,maxDiff=4)



**Concentrations decreasing at high discharge throughout the year**  
**Biggest declines around May and June**  
**Some increase at low flows from February through June**

```
> plotConcQSmooth("1993-04-01","2011-04-01",NA,qLow=100,qHigh=1000)
```

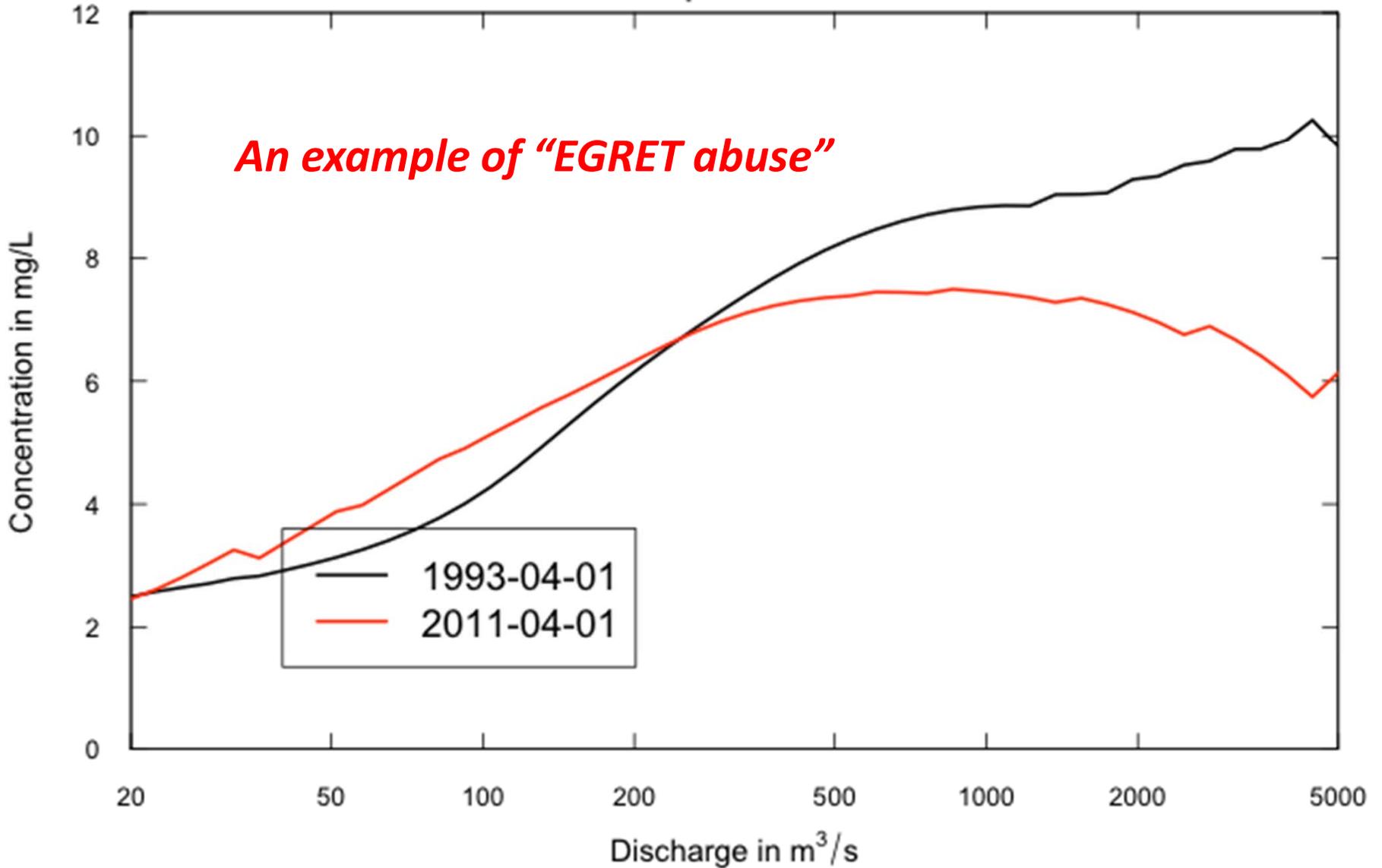
Iowa River at Wapello, IA Total Nitrogen  
Estimated Concentration Versus Discharge Relationship  
at 2 specific dates



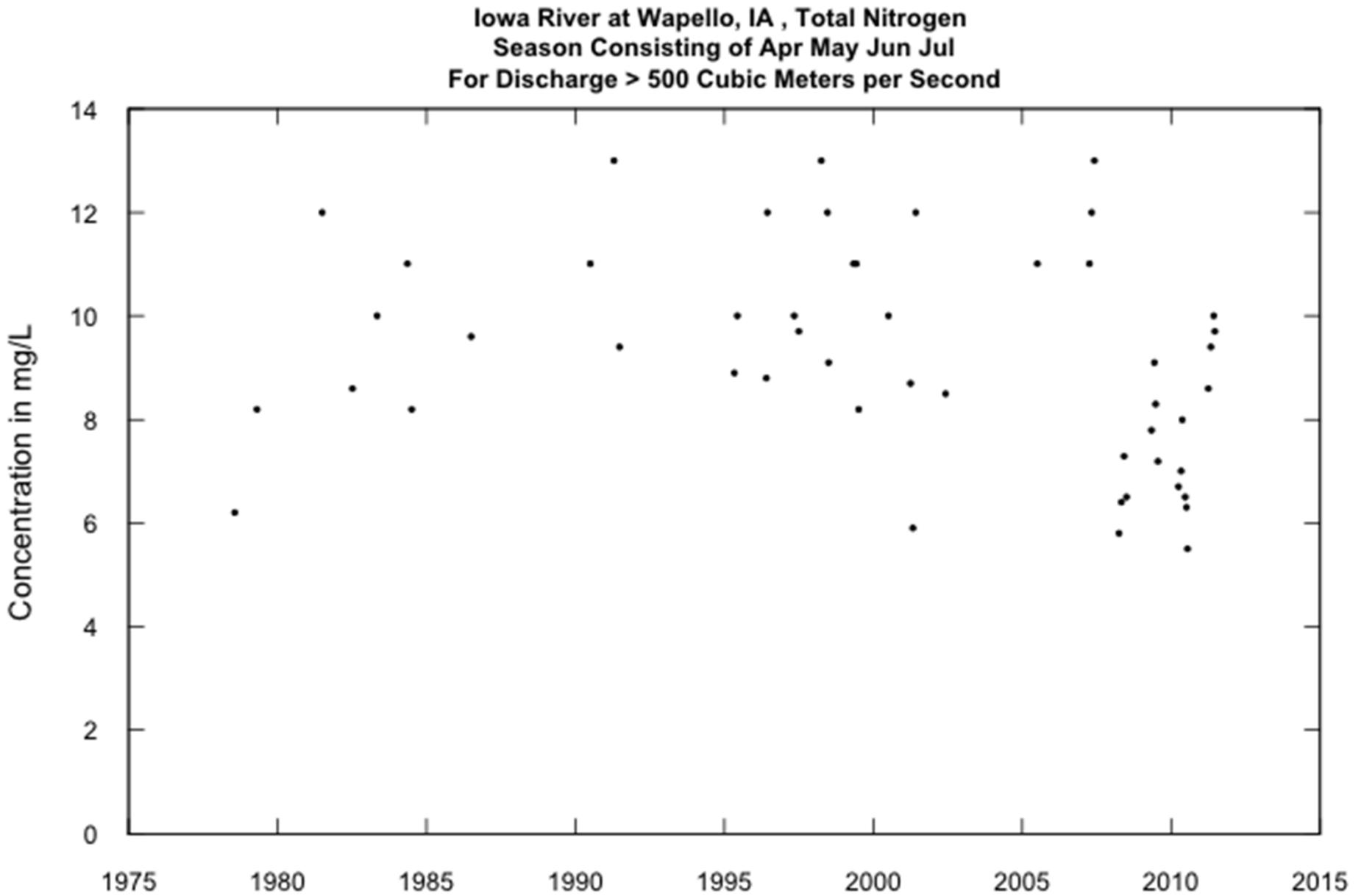


```
> plotConcQSmooth("1993-04-01","2011-04-01",NA,qLow=20,qHigh=5000)
```

Iowa River at Wapello, IA Total Nitrogen  
Estimated Concentration Versus Discharge Relationship  
at 2 specific dates



```
> plotConcTime(qLower=500,paLong=4,paStart=4)
```



# Moving to the western edge of the corn belt – the Missouri River at Hermann, MO

Missouri River at Hermann, MO

Nitrate as N

Water Year

## Concentration trends

time span	change mg/L	slope mg/L/yr	change %	slope %/yr
1980 to 2000	0.24	0.012	24	1.2
1980 to 2011	0.74	0.024	75	2.4
2000 to 2011	0.5	0.045	40	3.7

## Flux Trends

time span	change 10 <sup>6</sup> kg/yr	slope 10 <sup>6</sup> kg/yr /yr	change %	slope %/yr
1980 to 2000	31	1.5	36	1.8
1980 to 2011	48	1.6	56	1.8
2000 to 2011	17	1.6	15	1.3

```
> tableChange(fluxUnit=9,yearPoints=c(1980,2000,2011))
```

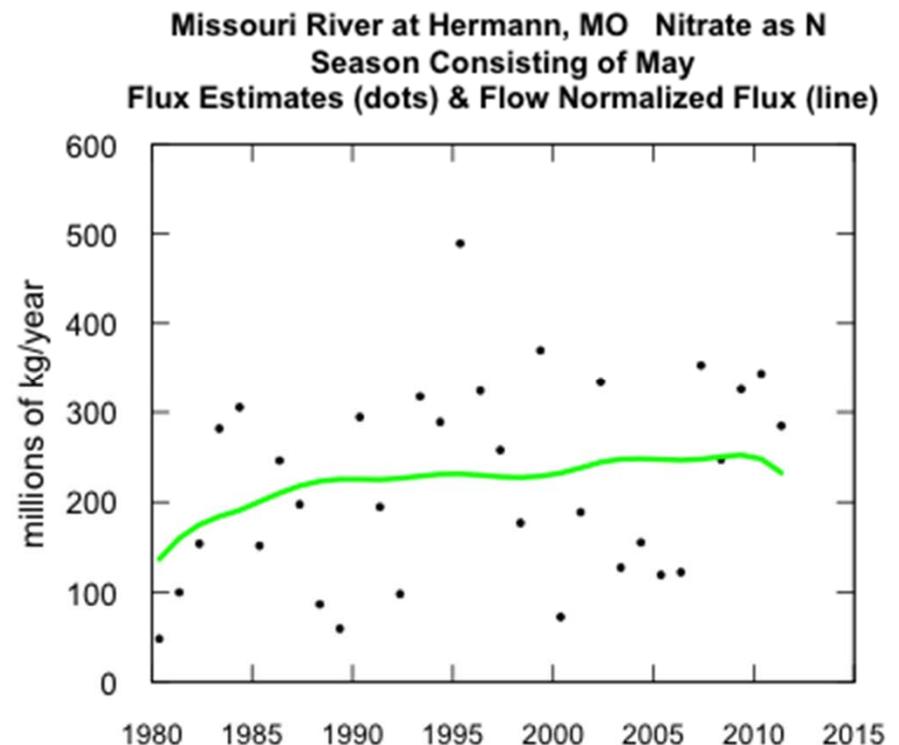
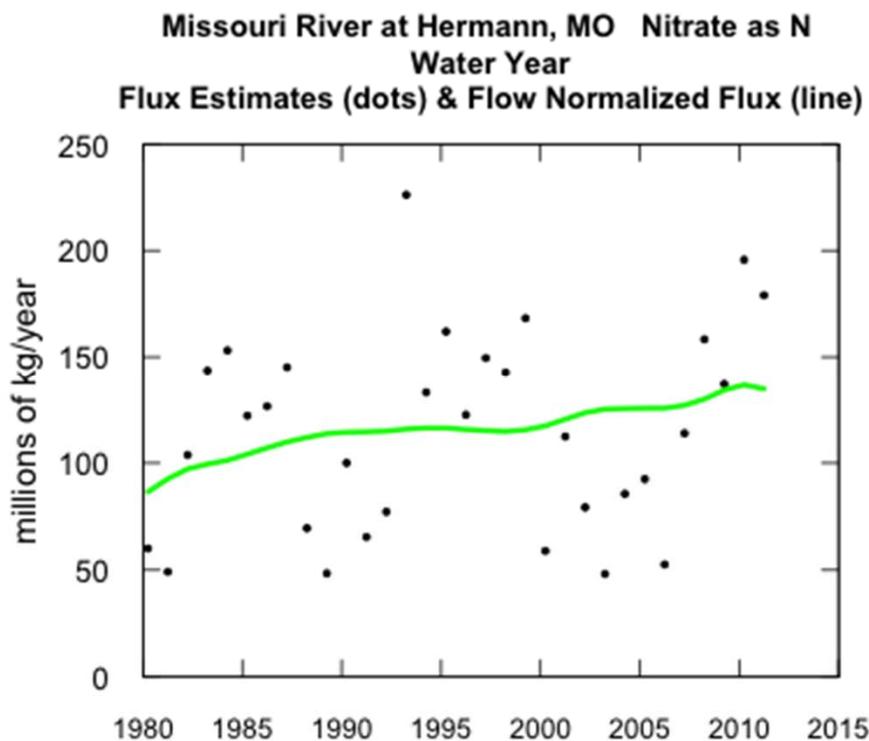
# Gulf Hypoxia concerns are focused on May flux (rather than annual). What's been the trend in May flux.

> AnnualResults<-setupYears(paLong=12,paStart=10)

> plotFluxHist(1980,2012)

> AnnualResults<-setupYears(paLong=1,paStart=5)

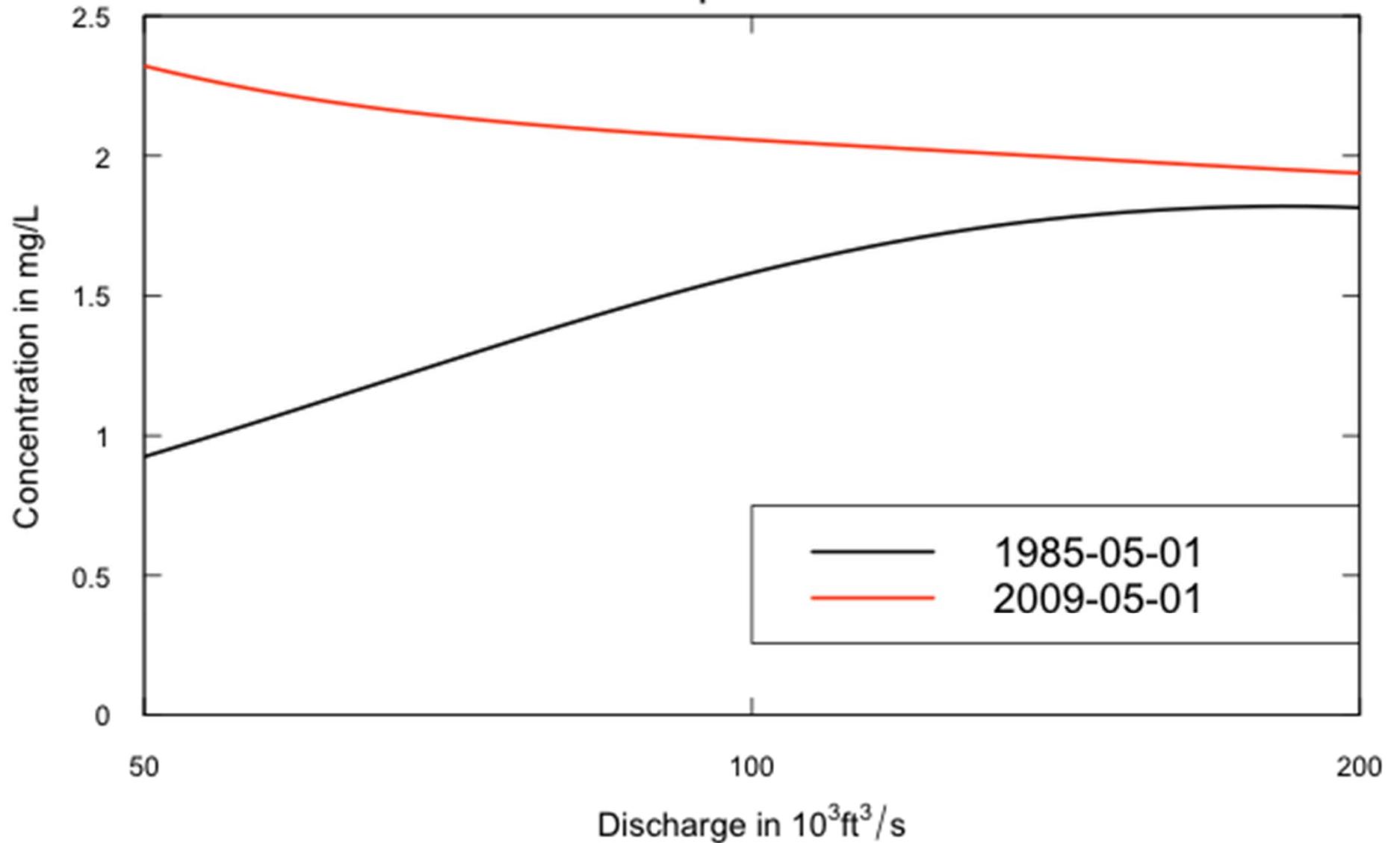
> plotFluxHist(1980,2012)



Annual flux keeps rising, but May flux seems flat, suggesting that the flux is getting more “spread out” over the course of the year. *The role of groundwater?*

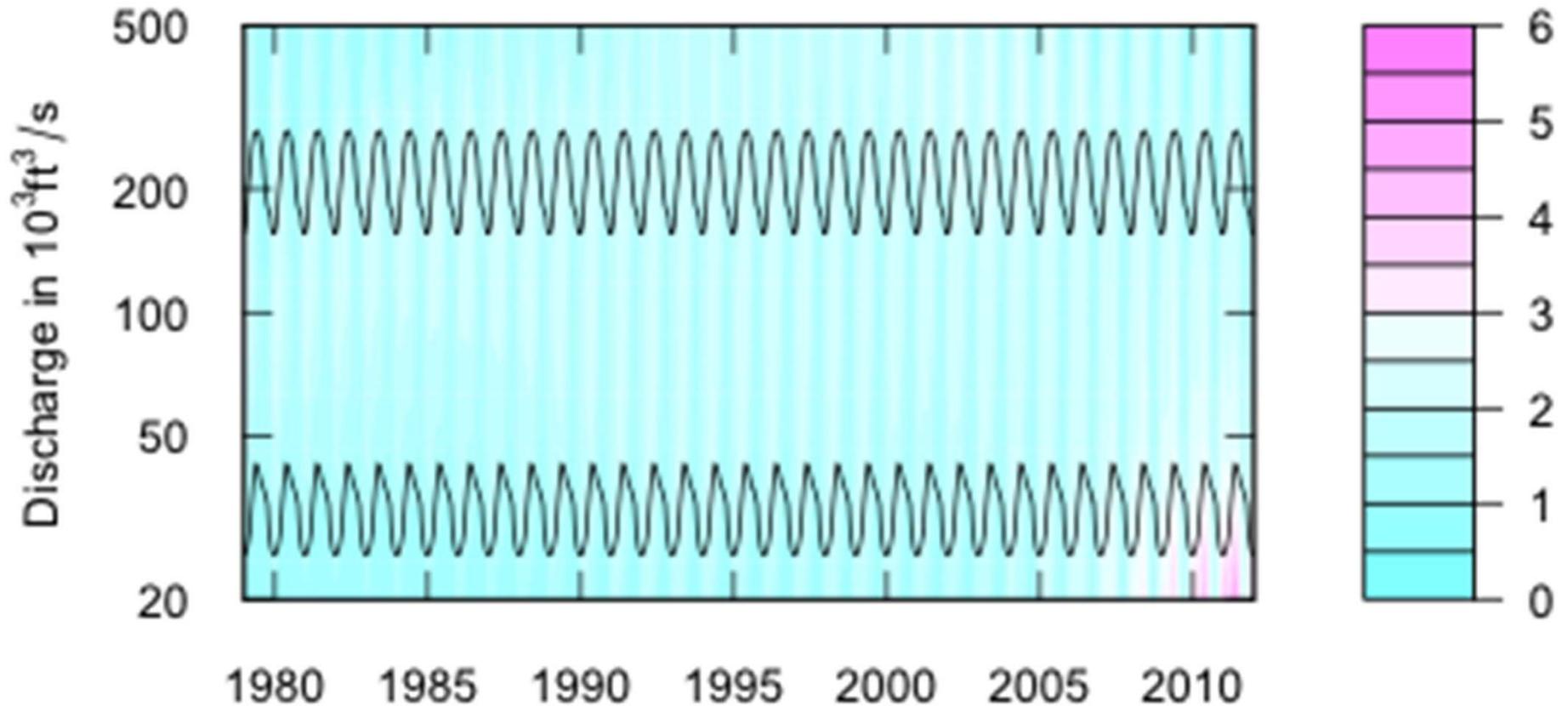
```
> plotConcQSmooth("1985-05-01","2009-05-01",NA,qLow=50,qHigh=200,qUnit=3)
```

Missouri River at Hermann, MO Nitrate as N  
Estimated Concentration Versus Discharge Relationship  
at 2 specific dates



```
> plotContours(1979,2012,20,500,qUnit=3,contourLevels=seq(0,6,0.5))
```

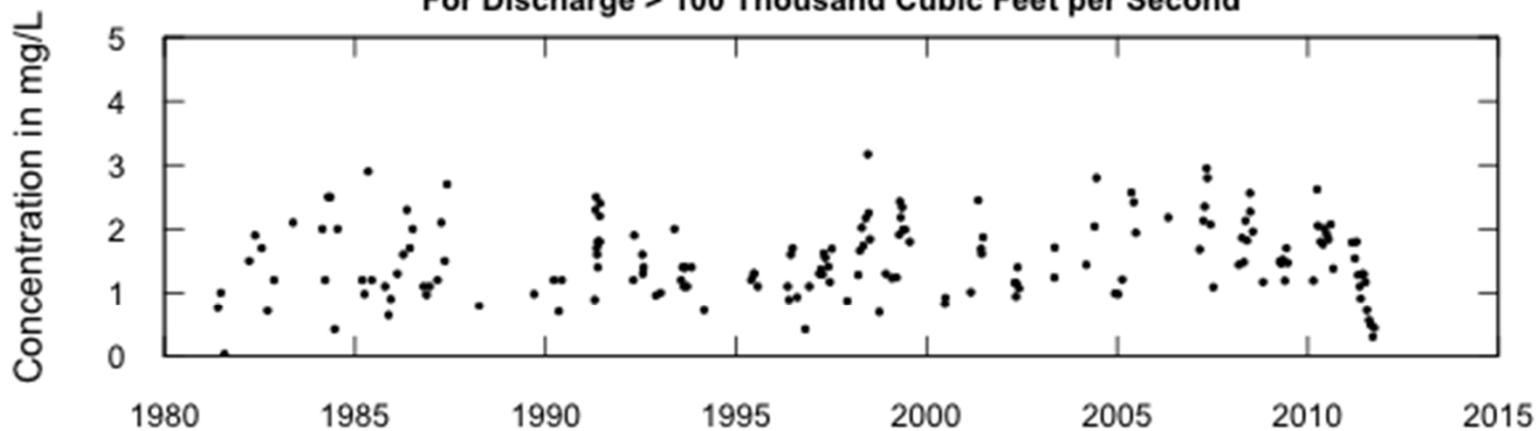
**Missouri River at Hermann, MO Nitrate as N  
Estimated Concentration Surface in Color  
Black lines are 5 and 95 flow percentiles**





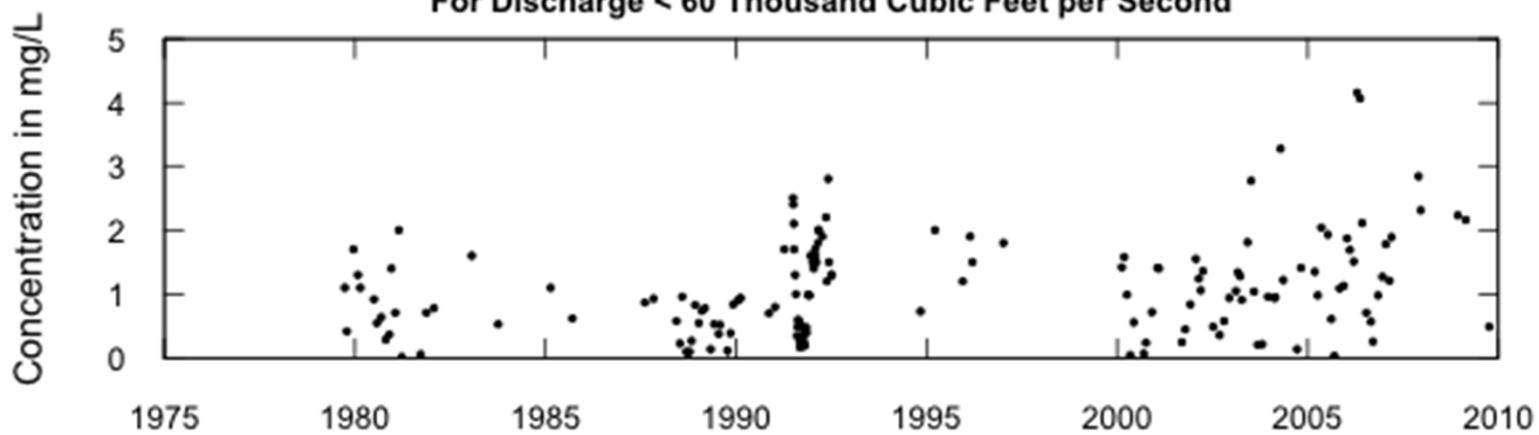
### Missouri River at Hermann, MO , Nitrate as N

For Discharge > 100 Thousand Cubic Feet per Second



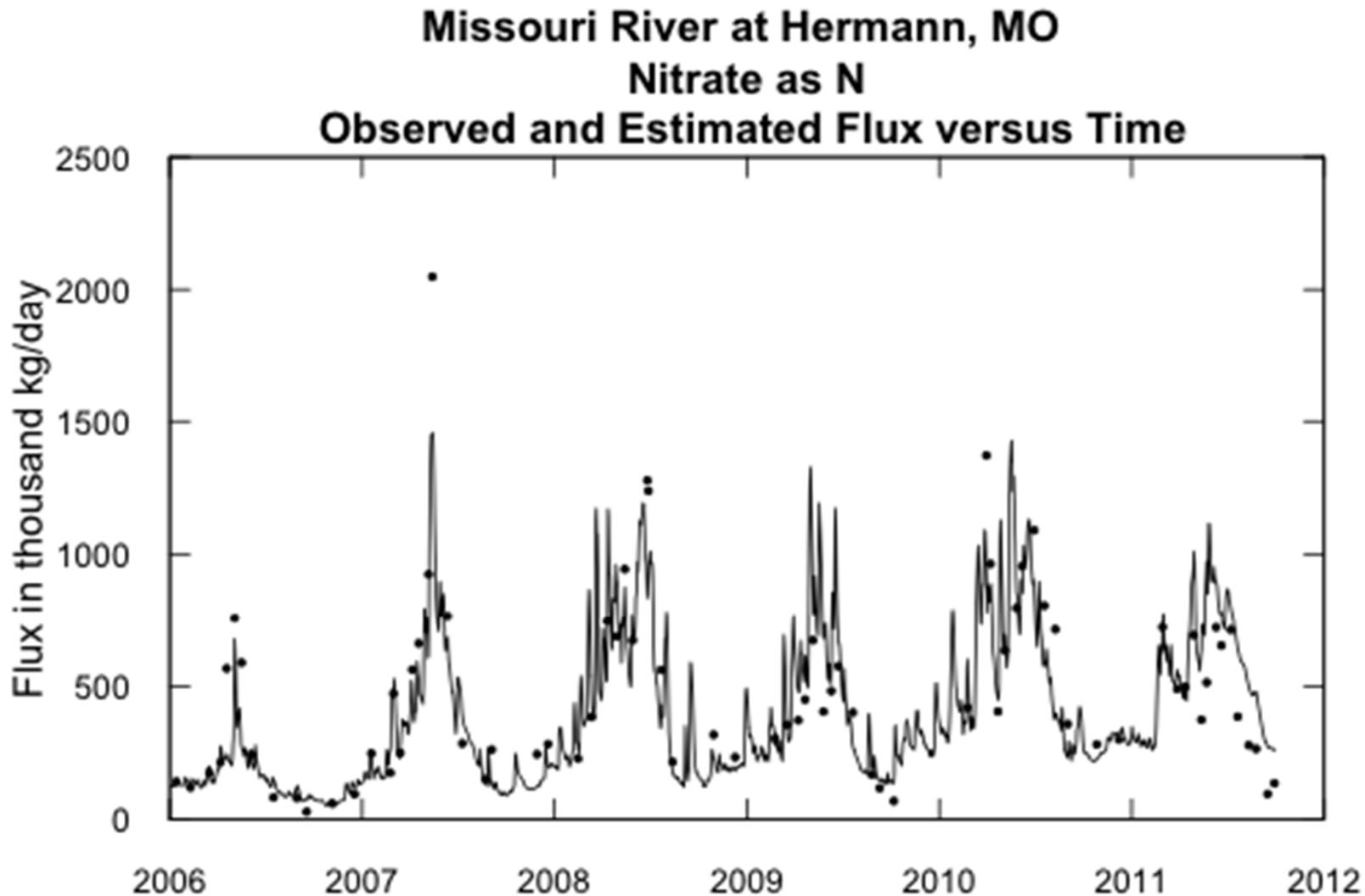
### Missouri River at Hermann, MO , Nitrate as N

For Discharge < 60 Thousand Cubic Feet per Second





Sometimes what doesn't work well is highly informative and suggests the need for a more complex model. *> plotFluxTimeDaily(2006,2012)*



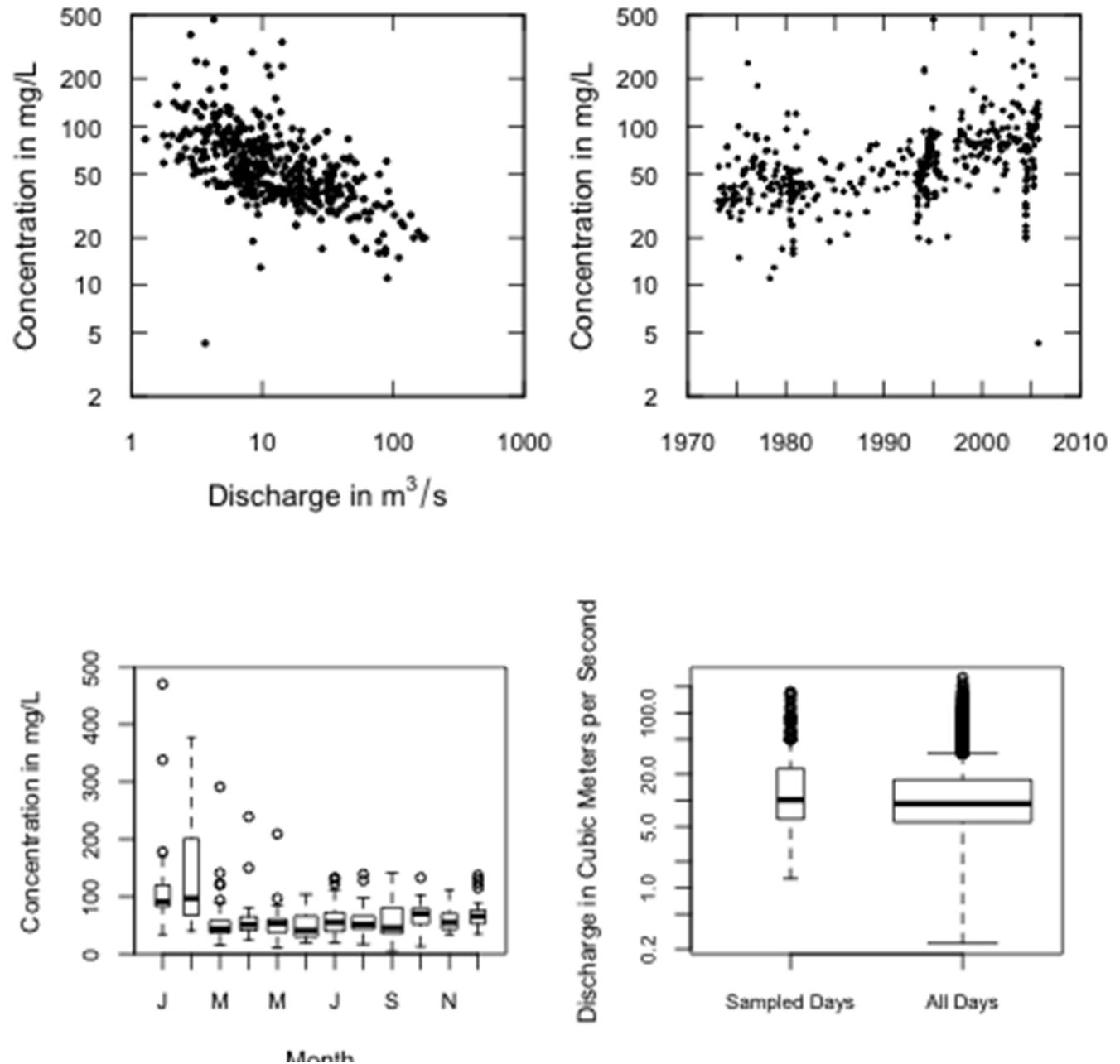
What was special about 2011 for the Missouri River?

## One last example

It doesn't take a rocket scientist to see that chloride decreases with flow, has increased with time, and is highest in the winter

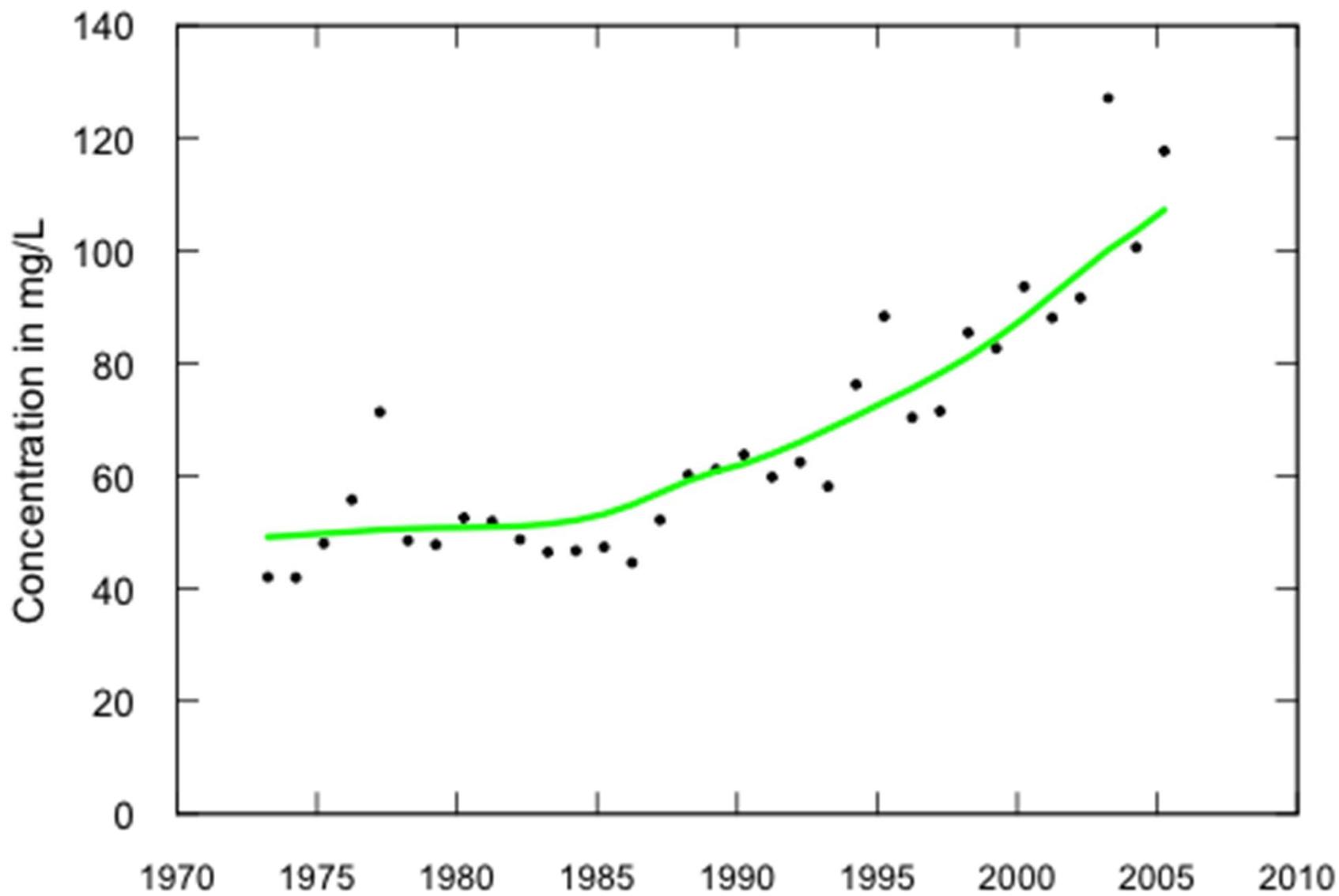
But is there more to the story?

### Milwaukee River at Milwaukee, WI Chloride



Milwaukee River at Milwaukee, WI Chloride  
Water Year

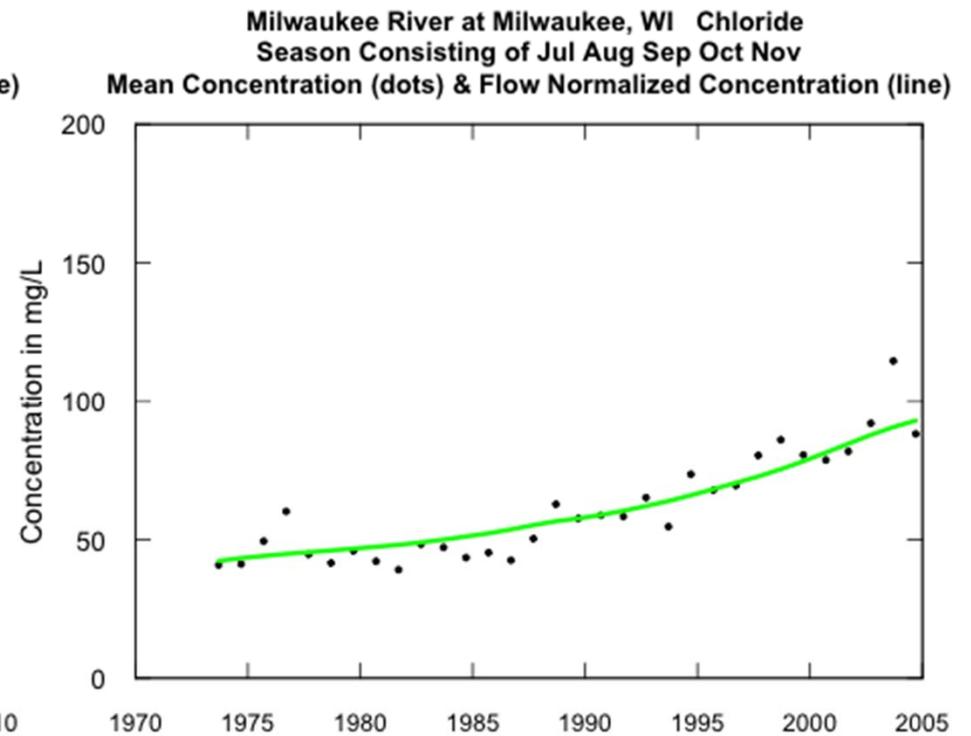
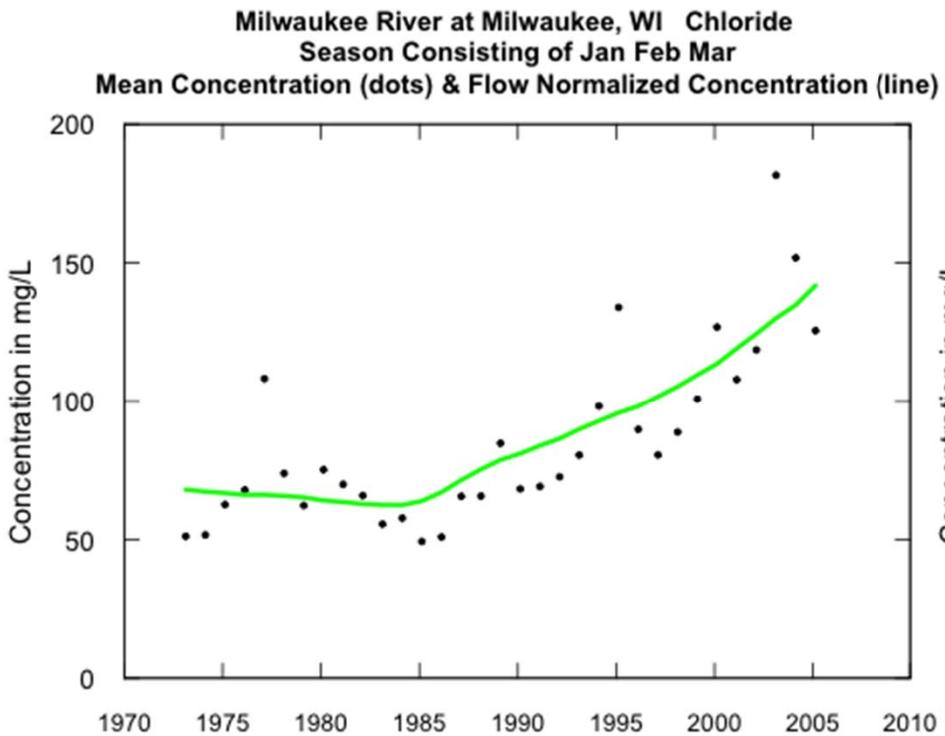
Mean Concentration (dots) & Flow Normalized Concentration (line)



## The changes are not confined to the winter months

```
> AnnualResults<-setupYears(paLong=3,paStart=1)  
> plotConcHist(1970,2010)
```

```
> AnnualResults<-setupYears(paLong=5,paStart=7)  
> plotConcHist(1970,2010,concMax=200)
```



```
> flowDuration()
```

Flow Duration for Milwaukee River at Milwaukee, WI

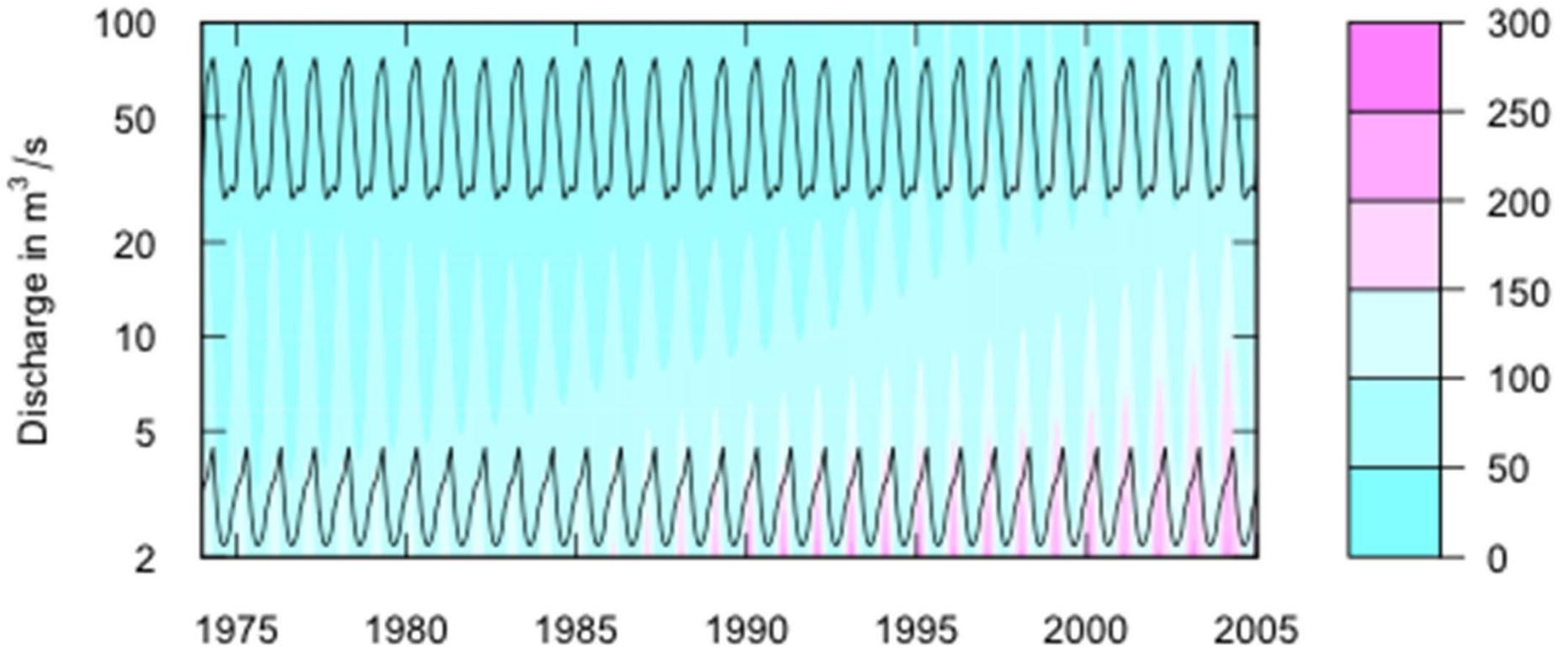
Flow duration is based on full year

Discharge units are Cubic Meters per Second

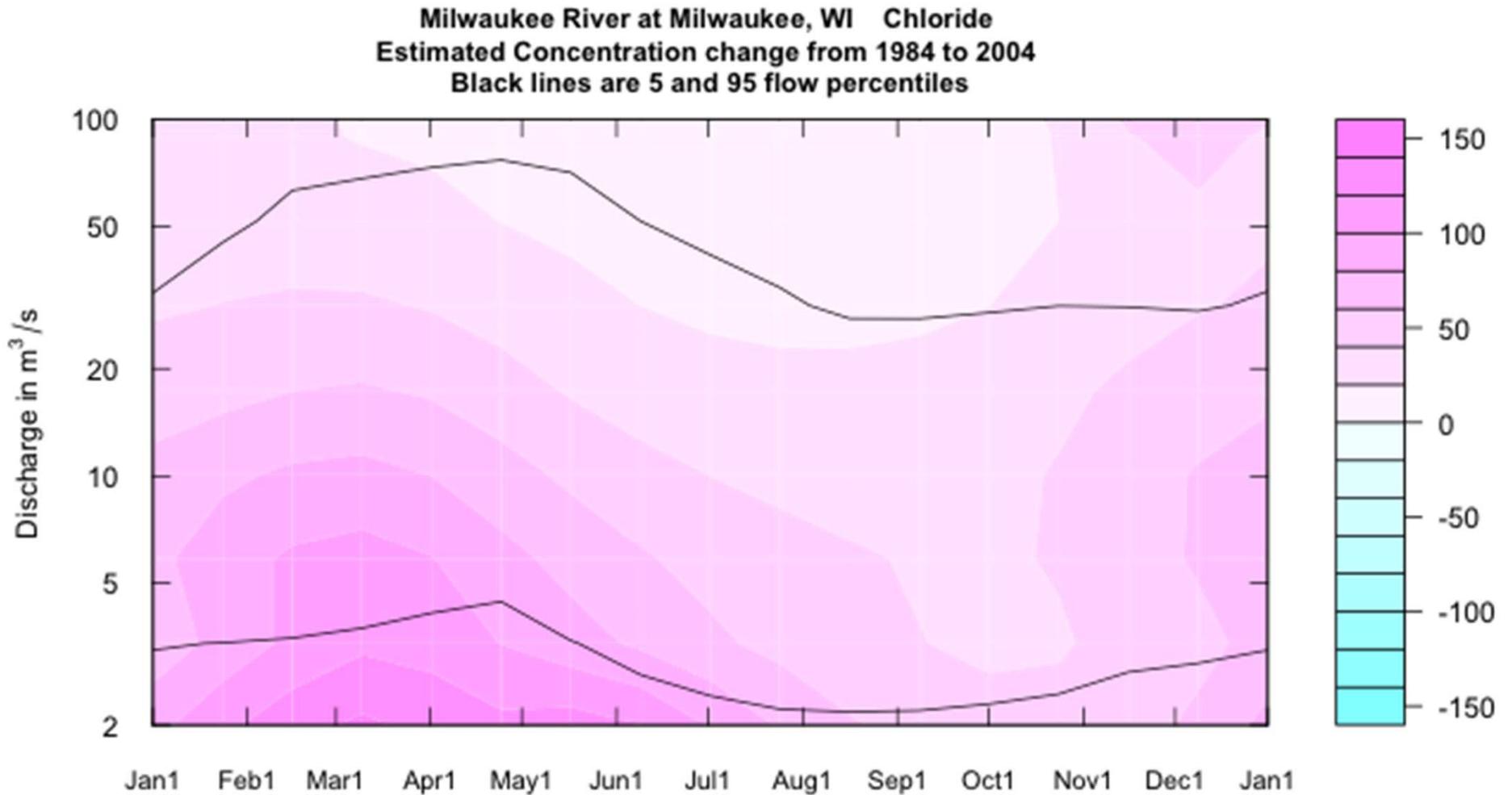
min	5%	10%	25%	50%	75%	90%	95%	max
0.232	2.860	3.681	5.663	9.175	17.188	32.564	47.572	254.002

```
> plotContours(1974,2005,2,100,contourLevels=seq(0,300,50))
```

**Milwaukee River at Milwaukee, WI Chloride  
Estimated Concentration Surface in Color  
Black lines are 5 and 95 flow percentiles**



> plotDiffContours(1984,2004,2,100,maxDiff=150)

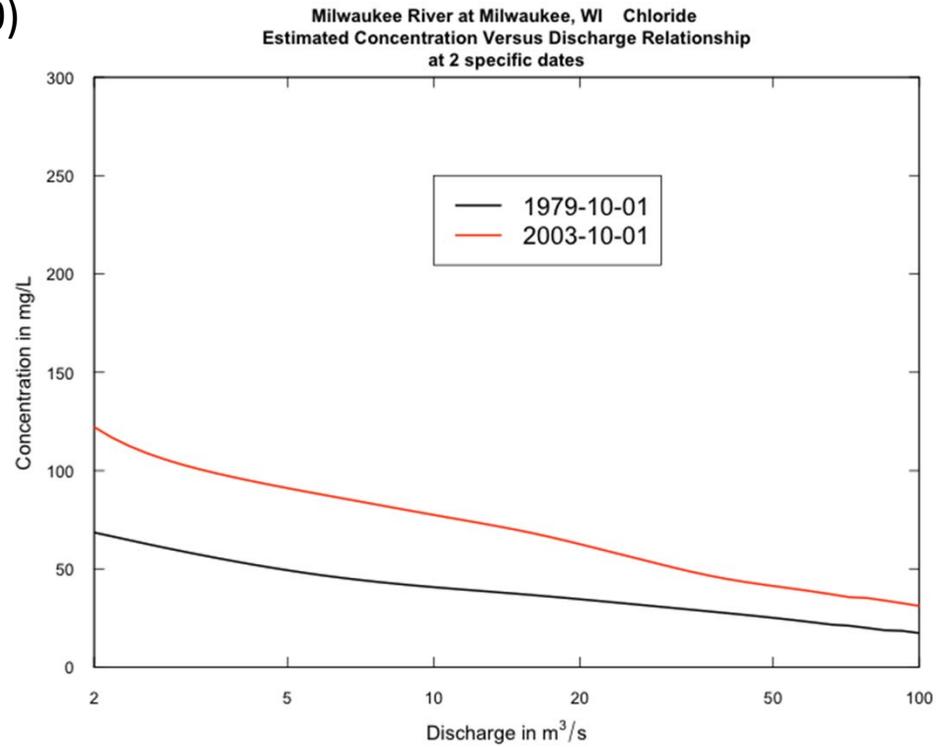
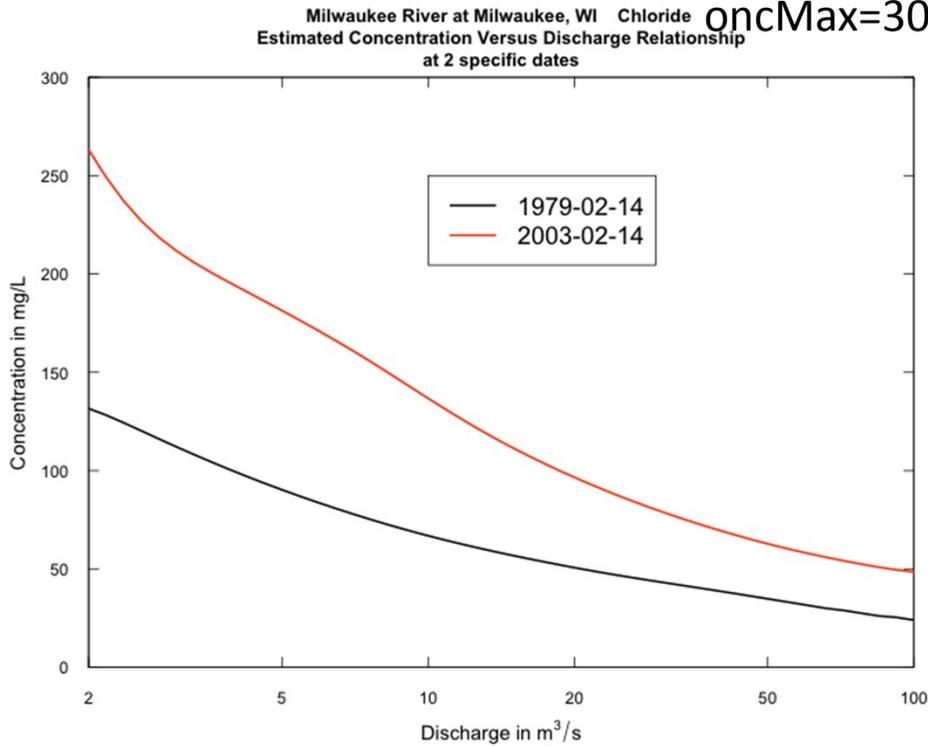


**Increases at all flows and all seasons**

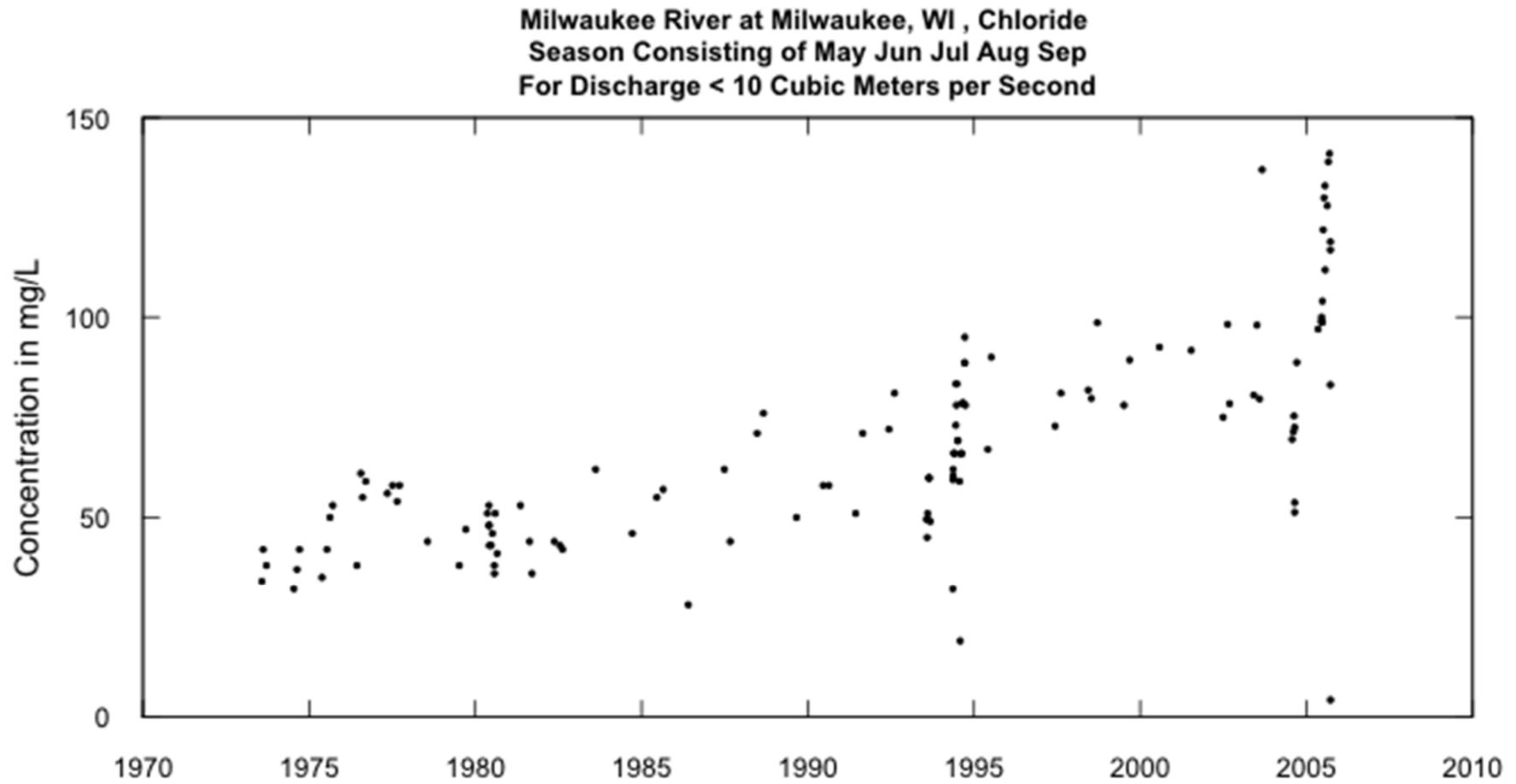
**The largest changes are at winter low to moderate flow conditions**

```
plotConcQSmooth("1979-02-14","2003-02-14",NA,qLow=2,qHigh=100,legendLeft=10,legendTop=250)
```

```
plotConcQSmooth("1979-10-01","2003-10-01",NA,qLow=2,qHigh=100,legendLeft=10,legendTop=250,concMax=300)
```



> plotConcTime(qUpper=10,paLong=5,paStart=5)





## EGRET functionality

- Graphs and tables are self-labeling and suitable for presentation or publication.
- Reporting units selected by user.
- Works interactively or in batch.
- Structures the data and results so it is easy to go back and ask further questions.
- Data structure opens up options for many other kinds of analysis using a wide-range of functions that are part of R.
- Data frames can be easily shared among users.

# EGRET philosophy

- Get the data easily and organize it for analysis.
- Don't **only** drive to get numbers or significance levels.
- Drive to understanding, hypotheses, & descriptions of our changing world.



# EGRET philosophy

There are changes all around us, now  
describe them to help guide how we  
manage our water resources!



**We welcome your questions and feedback on: *the methods, the outputs, the manual***

The R-packages, draft manual and this presentation are on the EGRET web site:

<https://github.com/USGS-CIDA/WRTDS/blob/master/README.md>

Send your questions and feedback to:  
[egret\\_comments@usgs.gov](mailto:egret_comments@usgs.gov)

