New Tools for Water Quality Data Access and Trend Analysis

An overview of the USGS R Packages:

dataRetrieval and EGRET

Robert M. Hirsch & Jessica Thompson
USGS
1 May 2014

Outline

- Motivations for the package
- The WRTDS concept and examples of results
- Overview of dataRetrieval
- How EGRET works, doing WRTDS analysis and producing graphs and tables

USGS

From Ralph Keeling

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.

Keeling, 2008, Recording Earth's vital signs, Science, p1771-1772

USGS

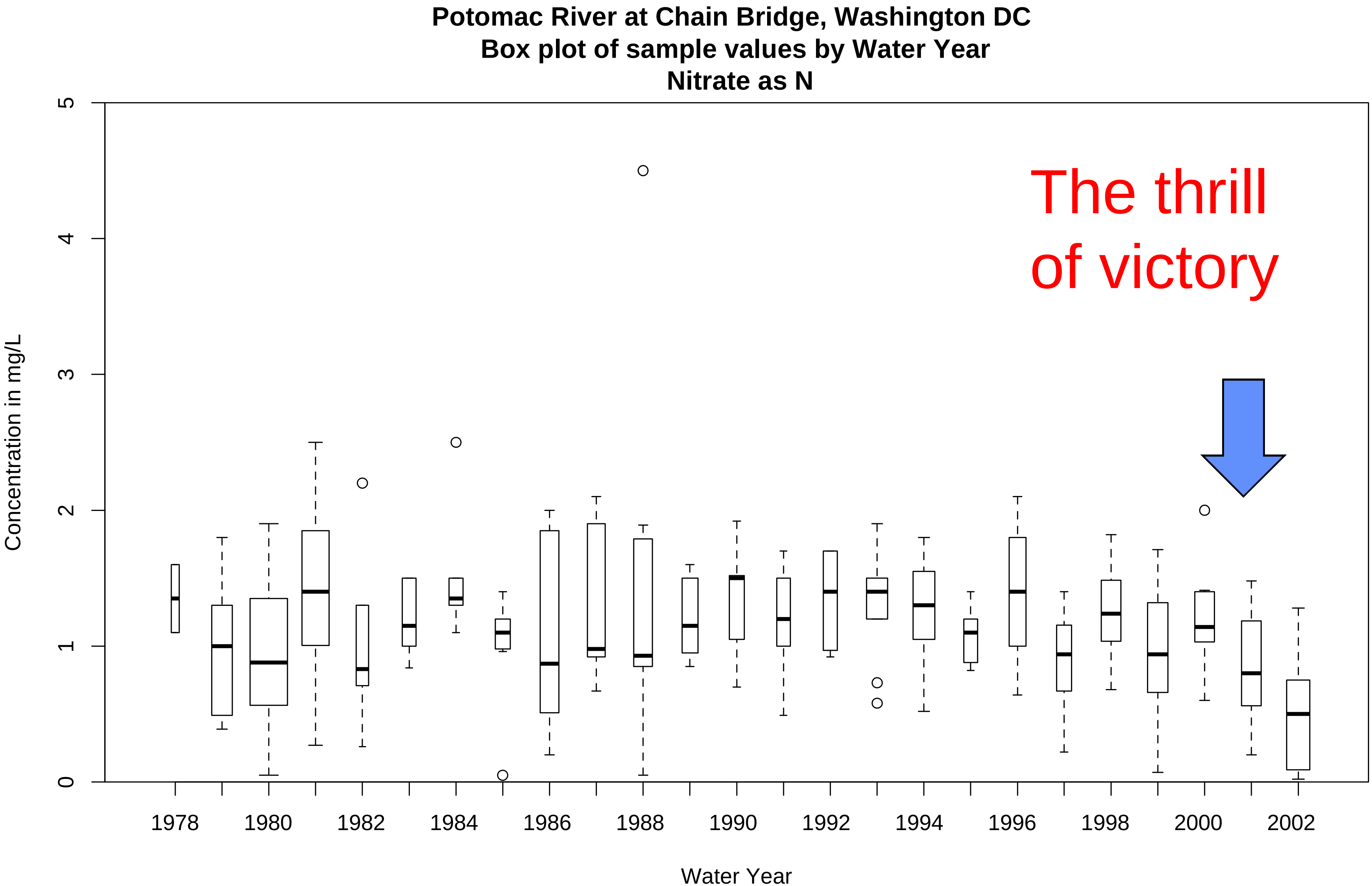
The challenge

How do we come to understand what is happening to water quality in large watersheds.

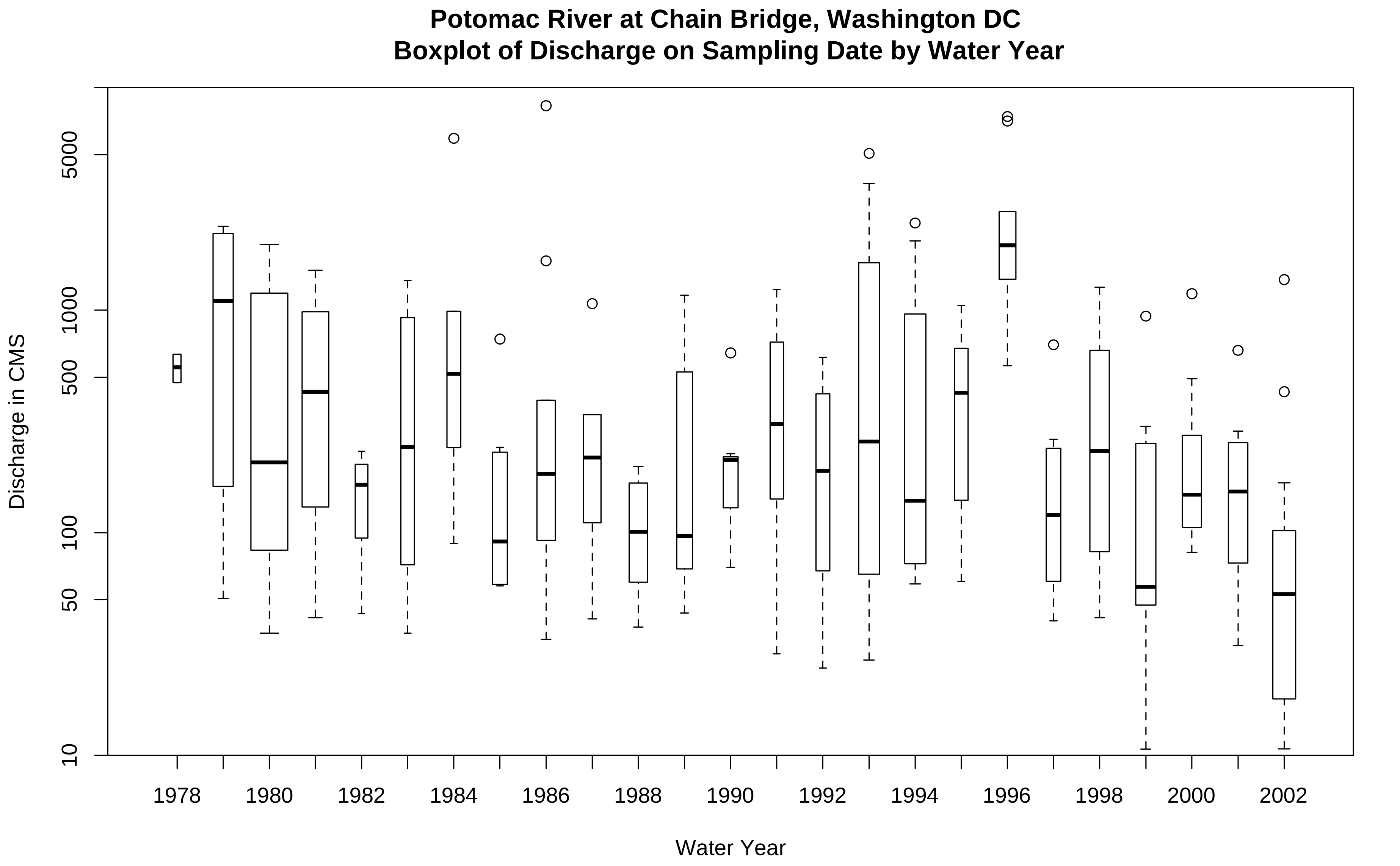
Is it getting better or worse?

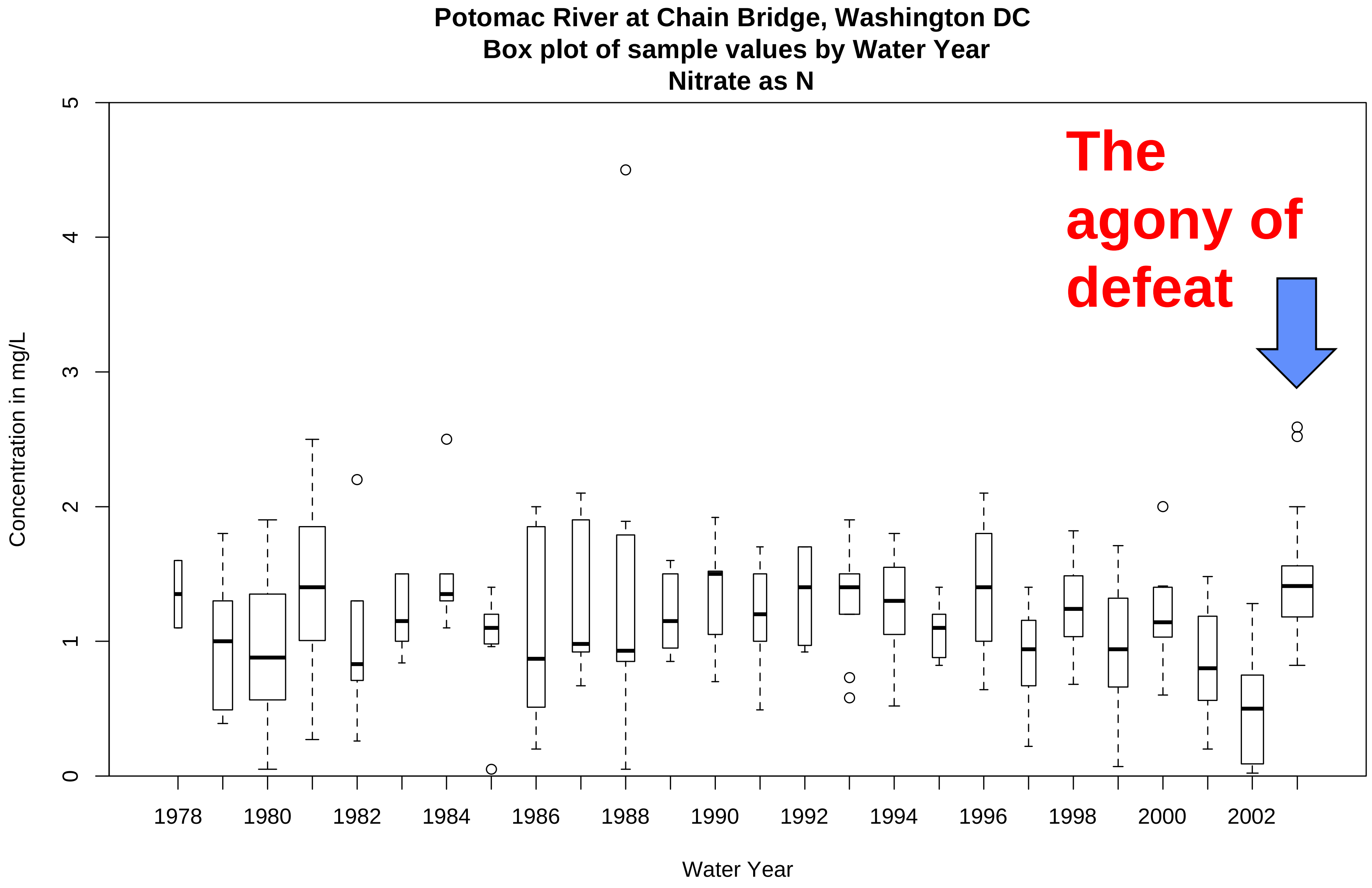
Can we develop ideas of causative factors and changes in processes?

Can these be used to guide management choices?

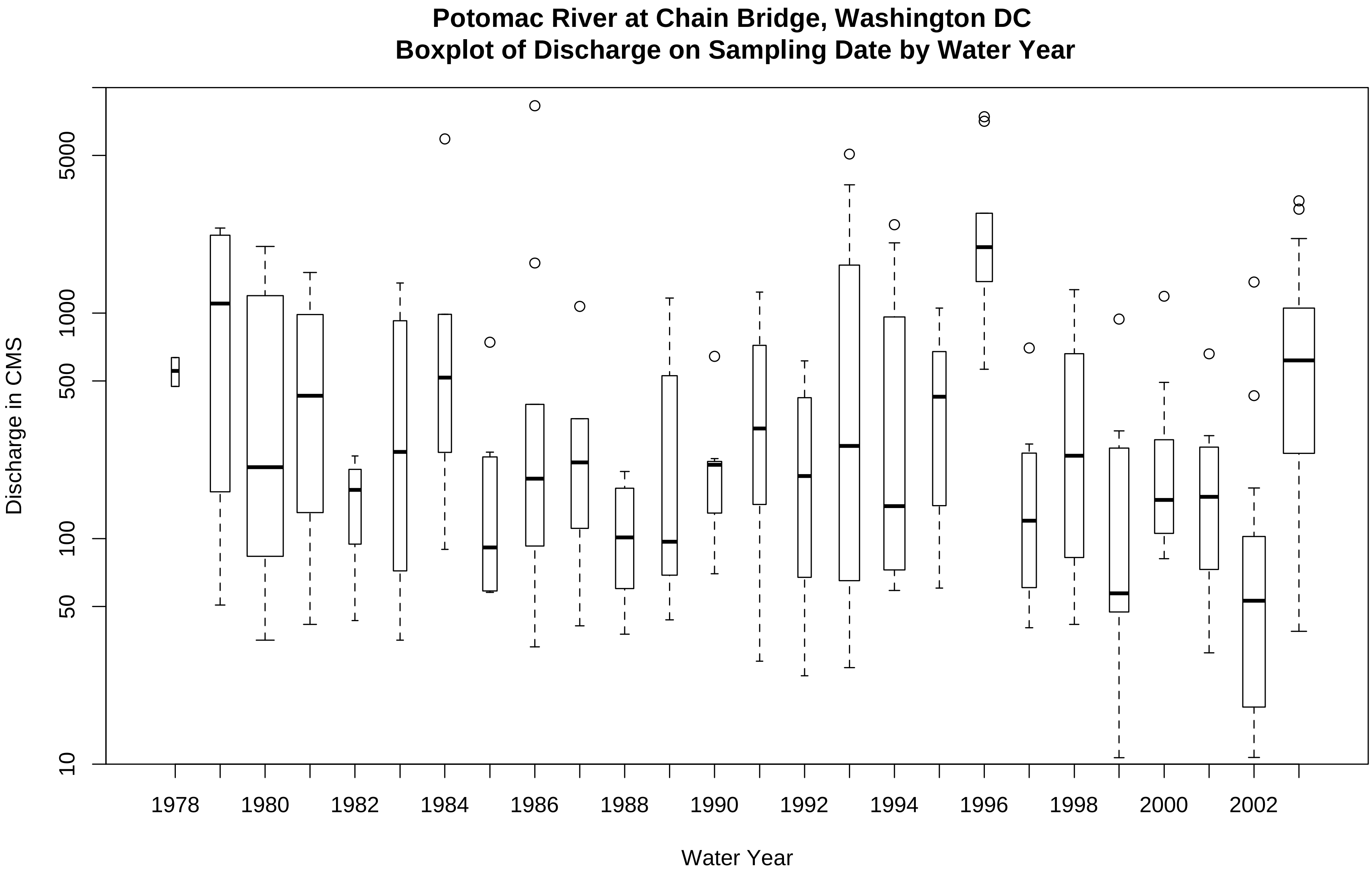


Data through September 2002


Potomac River at Chain Bridge, Washington DC
Boxplot of Discharge on Sampling Date by Water Year
Discharge in CMS
10
50
100
500
1000
5000
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
Water Year



Data through September 2003



Potomac River at Chain Bridge, Washington DC
Boxplot of Discharge on Sampling Date by Water Year
Discharge in CMS
10
50
100
500
1000
5000
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
Water Year

Motivations for the method

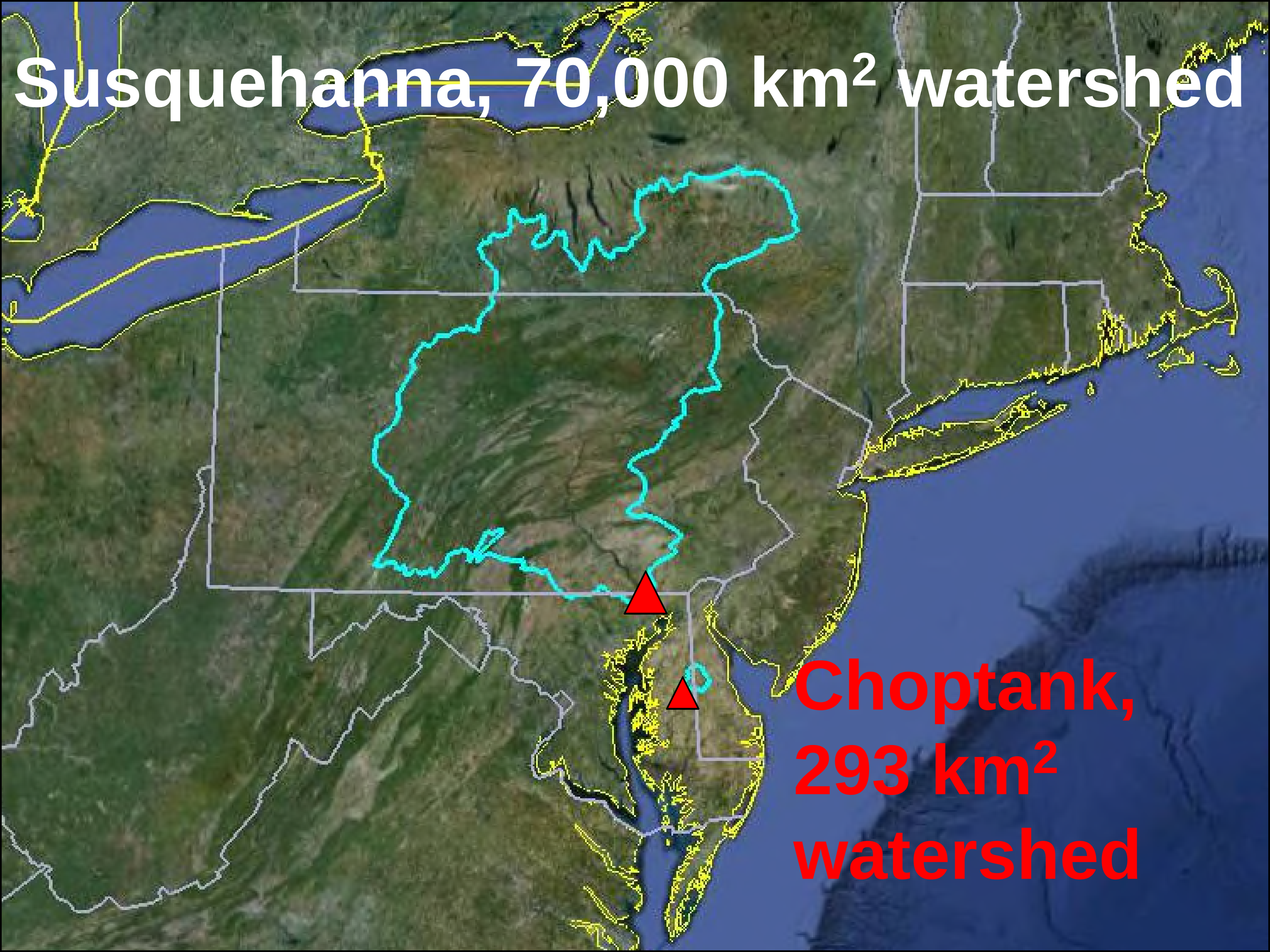
- Describe the evolving behavior of the watershed. No mathematical straight-jacket!!
- Estimate both concentration & flux (averages as well as trends).
- Estimate the actual history but also a flow-normalized history.
- Resolve a serious bias in flux estimates.
- Be quantitative but also exploratory.

USGS

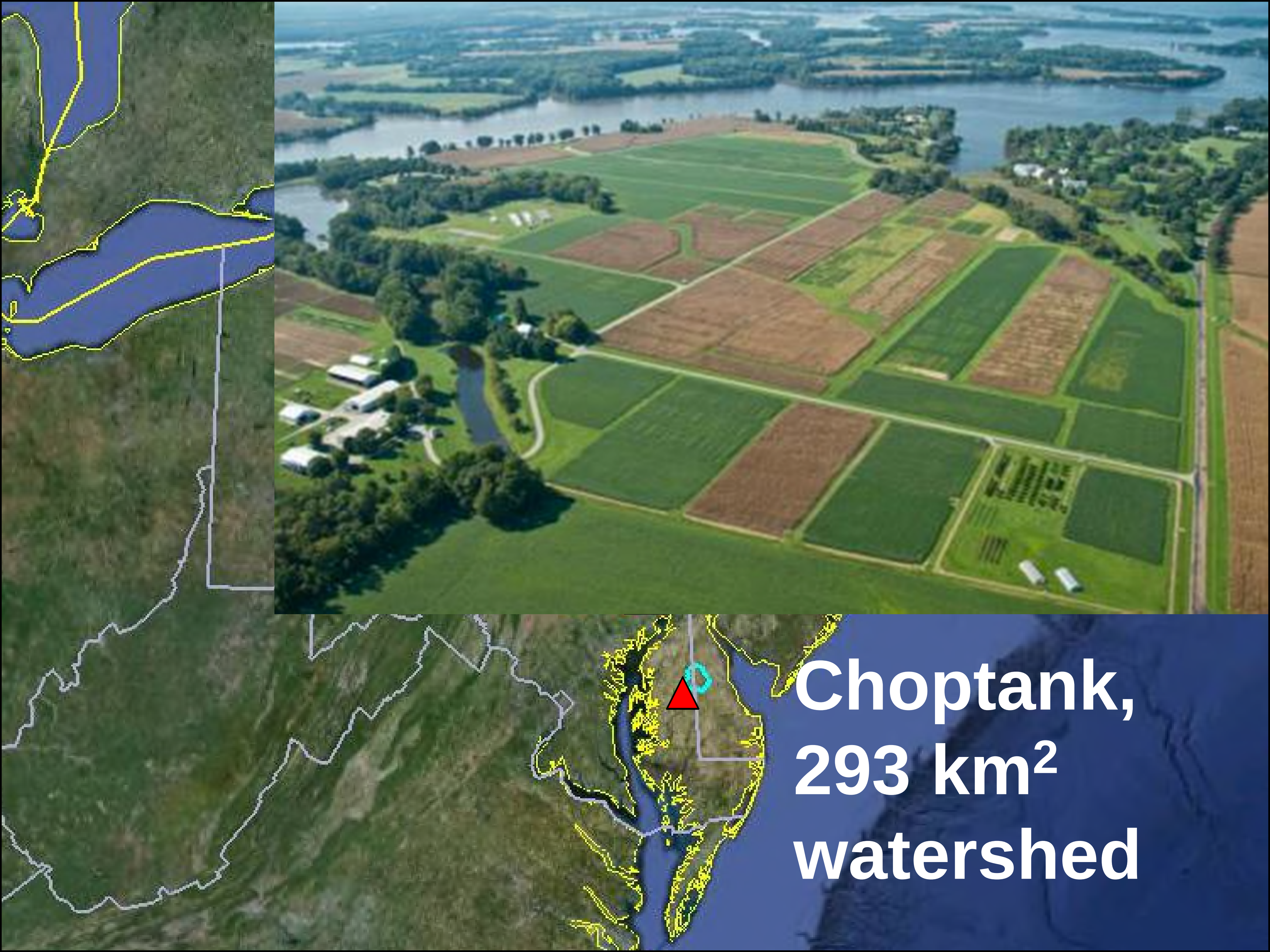
Data requirements

- Low intra-day variability (not flashy)
- Requires a complete daily discharge record
- Intended for >200 samples, but has been used for some purposes with as few as 60 samples
- Water quality samples cover most of the discharge range
- For trend studies: 20+ years
- For average flux computations: 5 – 10 years.

USGS

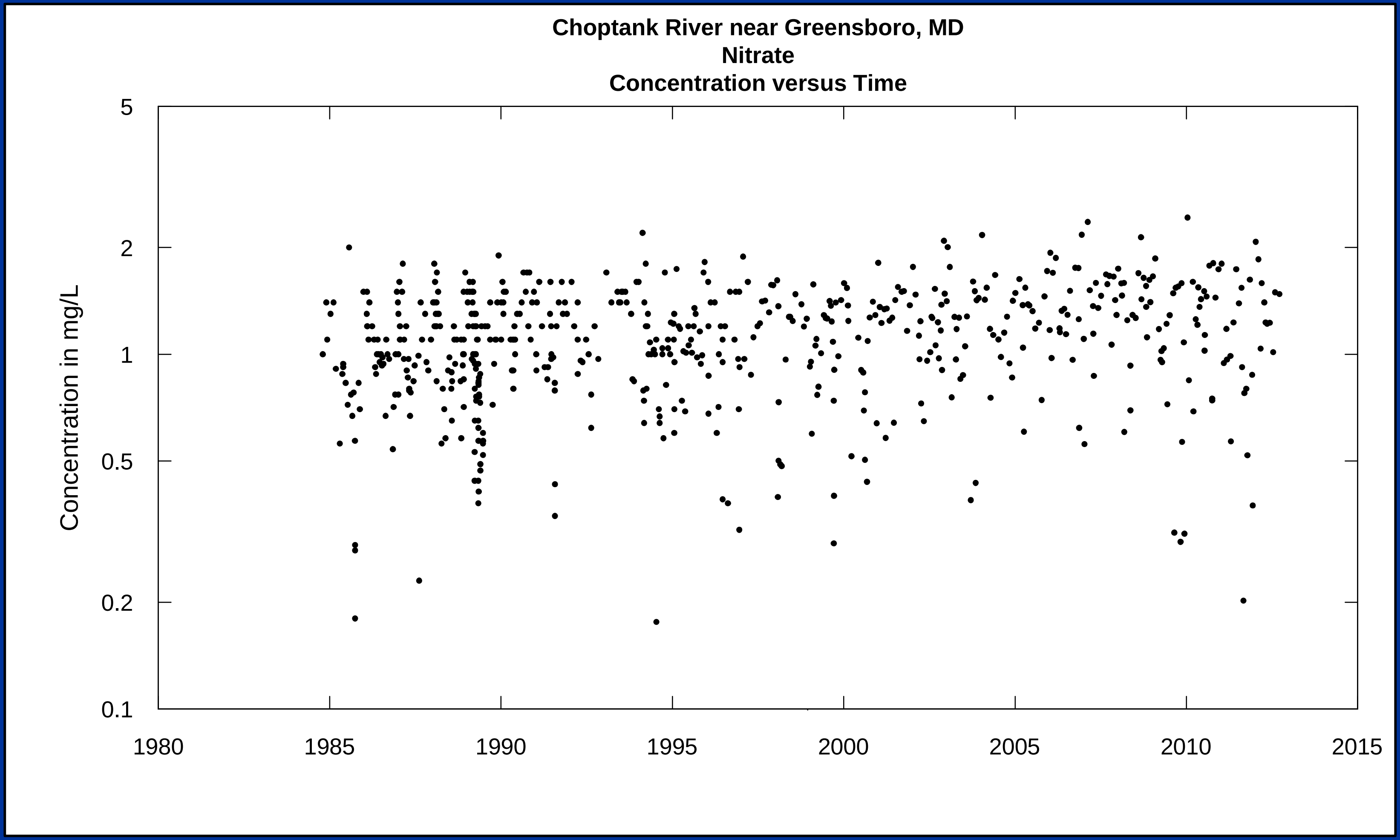


Susquehanna, 70,000 km² watershed
Choptank, 293 km² watershed


Choptank,
293 km²
watershed

“Data without models are chaos, but models without data are fantasy”

Nesbit, Dlugokencky and Bousquet, Science, 31 January 2014, pp. 493-495



USGS

Use the data and a simple, highly-flexible smoothing model to decompose the data into 4 components.

1) Discharge related component

2) Seasonal component

3) Time trend

4)Random component

Weighted Regressions on Time, Discharge and Season (WRTDS)

USGS

Locally Weighted Regression

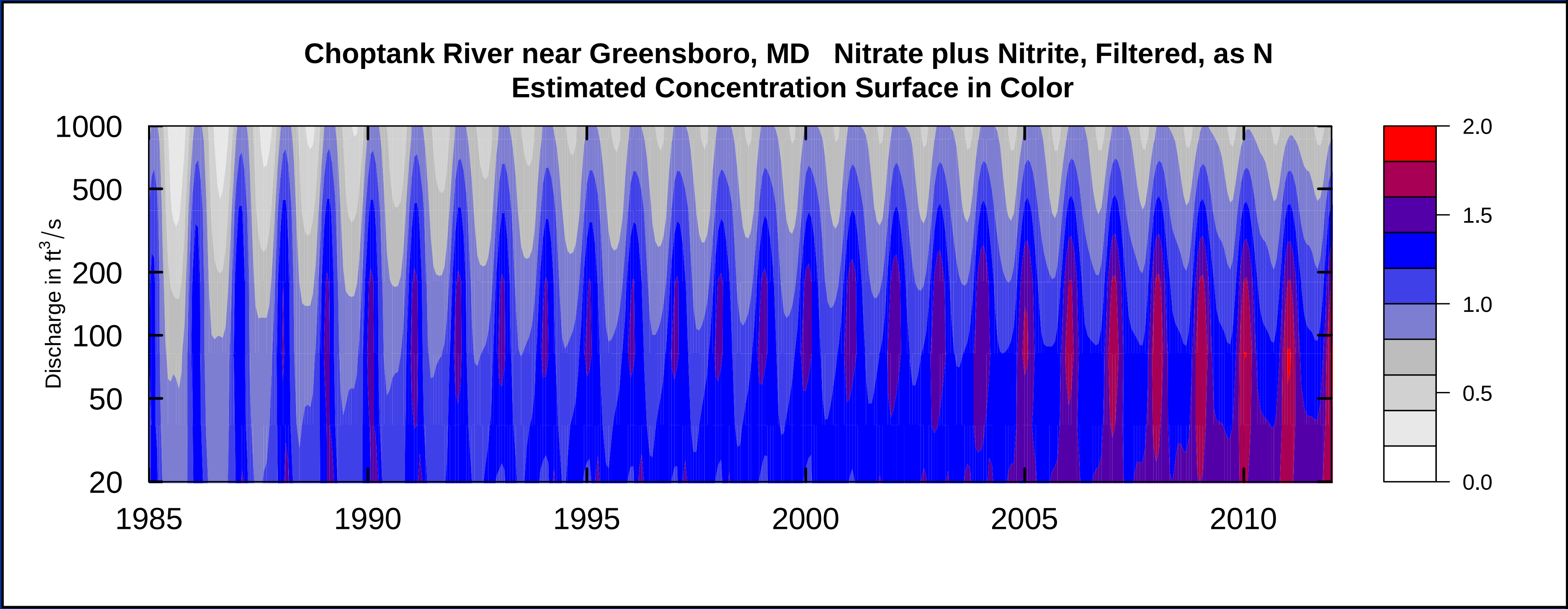
For any location in time - discharge space (*t* and *Q*) we assume that concentration (*c*) follows this model

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

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

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.

WRTDS view of the evolving behavior of nitrate

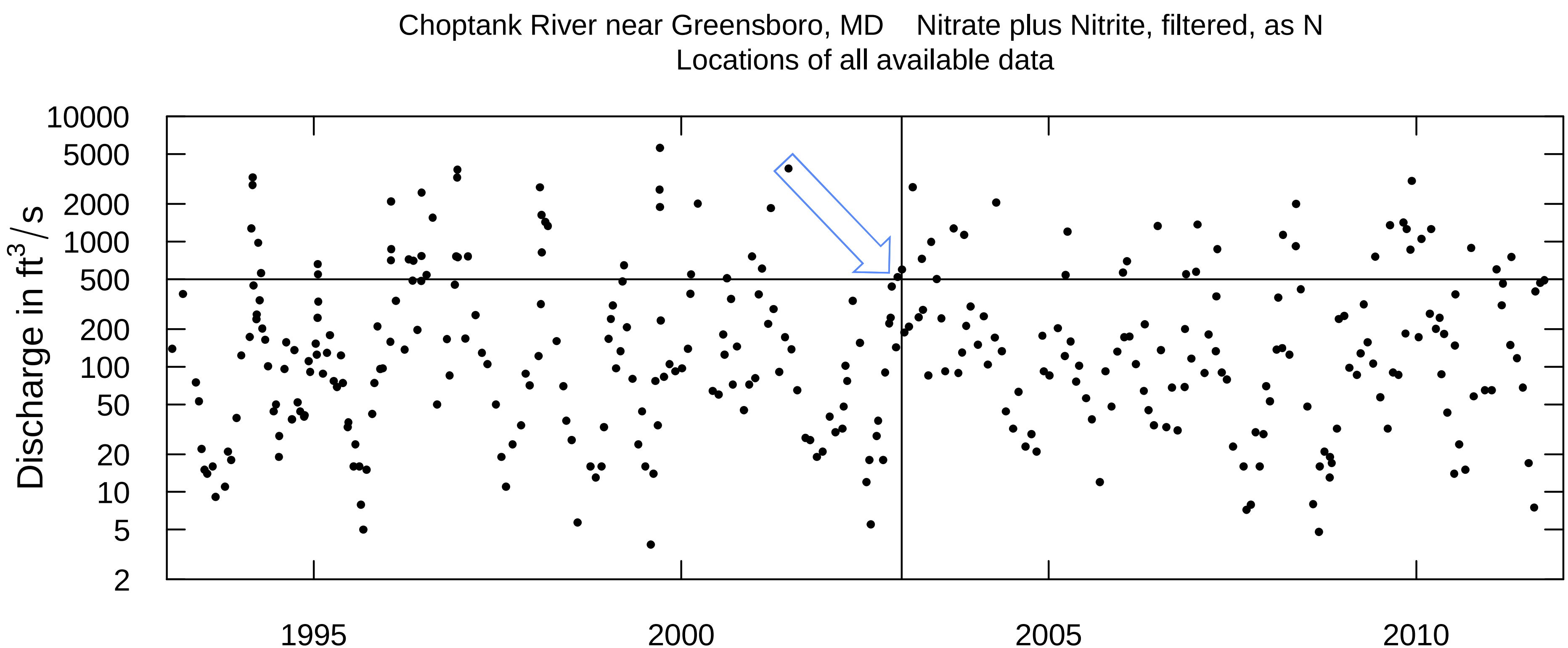


How is this surface created?

USGS

Every dot is a data point from 1993 to 2012

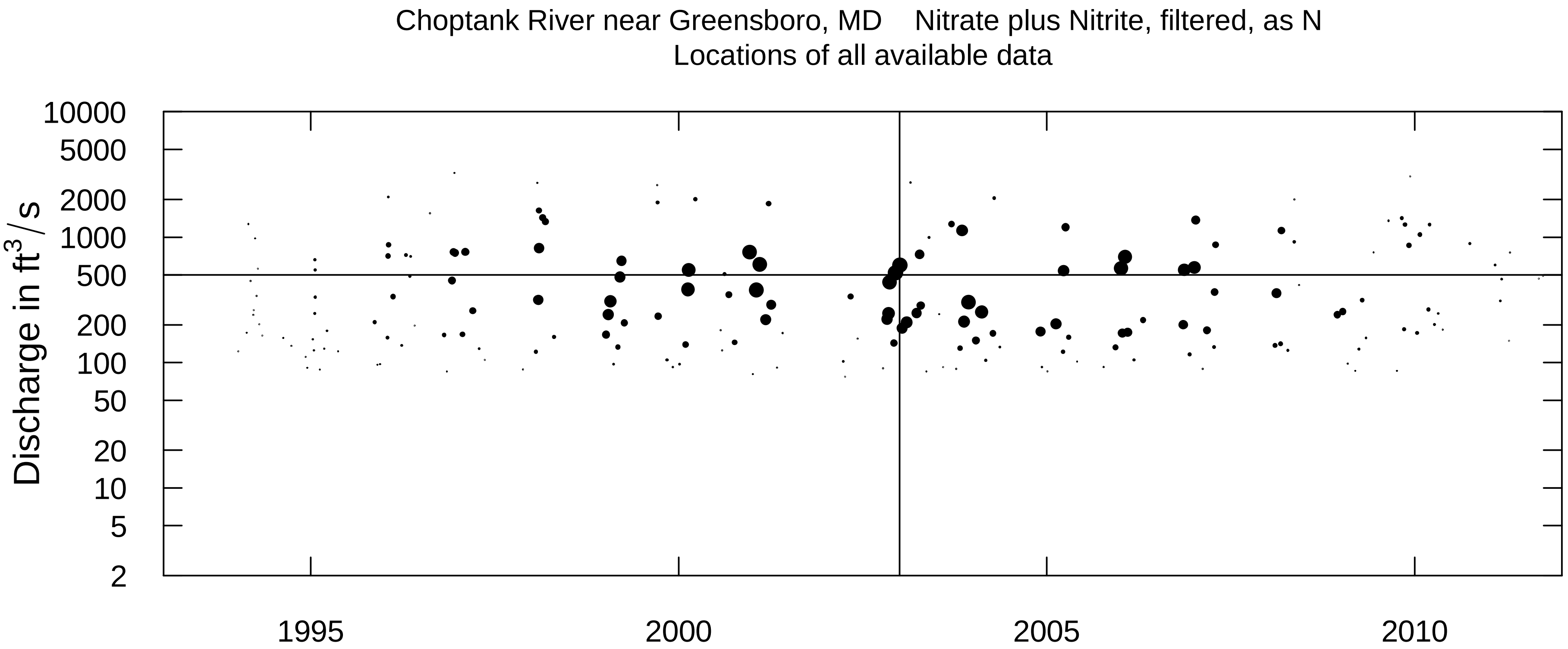
Let's say we want to use the data to estimate the expected value of concentration for January 1, 2003 at Q=500 cfs



USGS

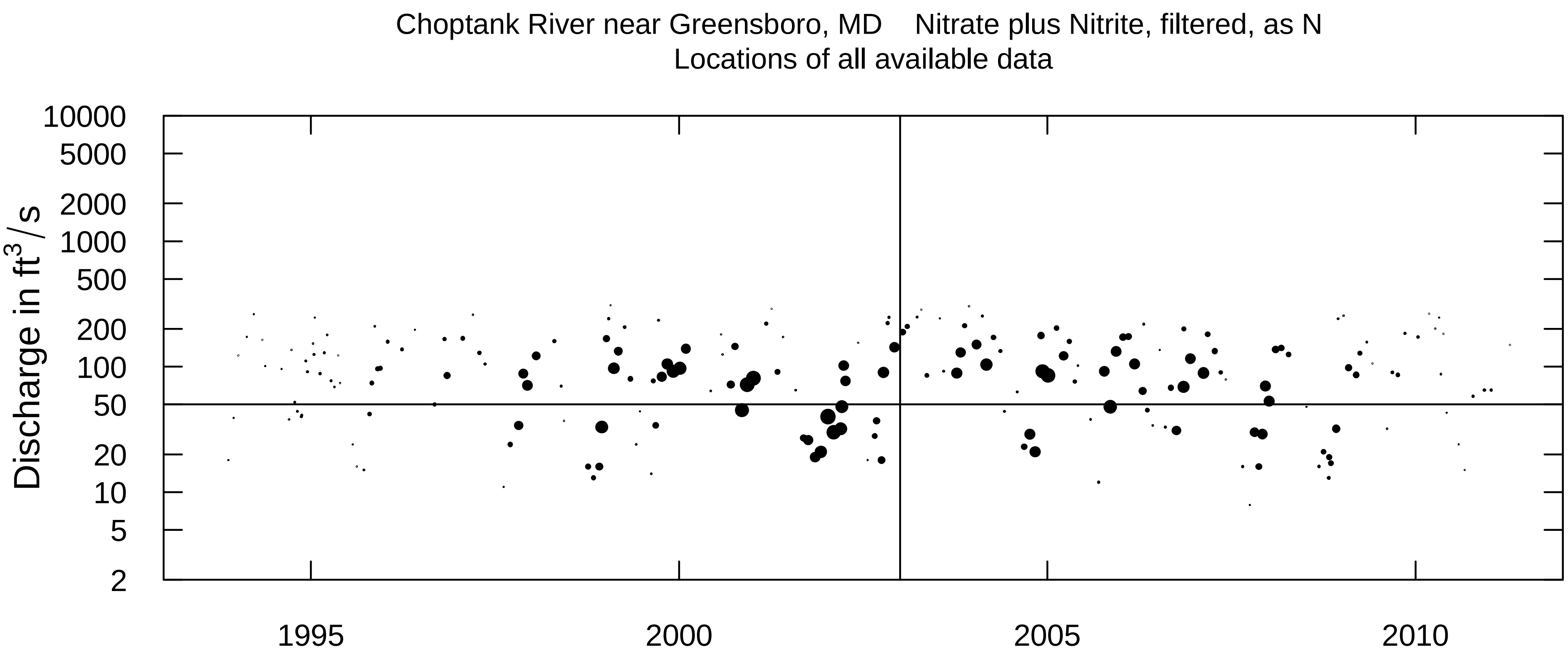
How do we set the weights for the regression?

- 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



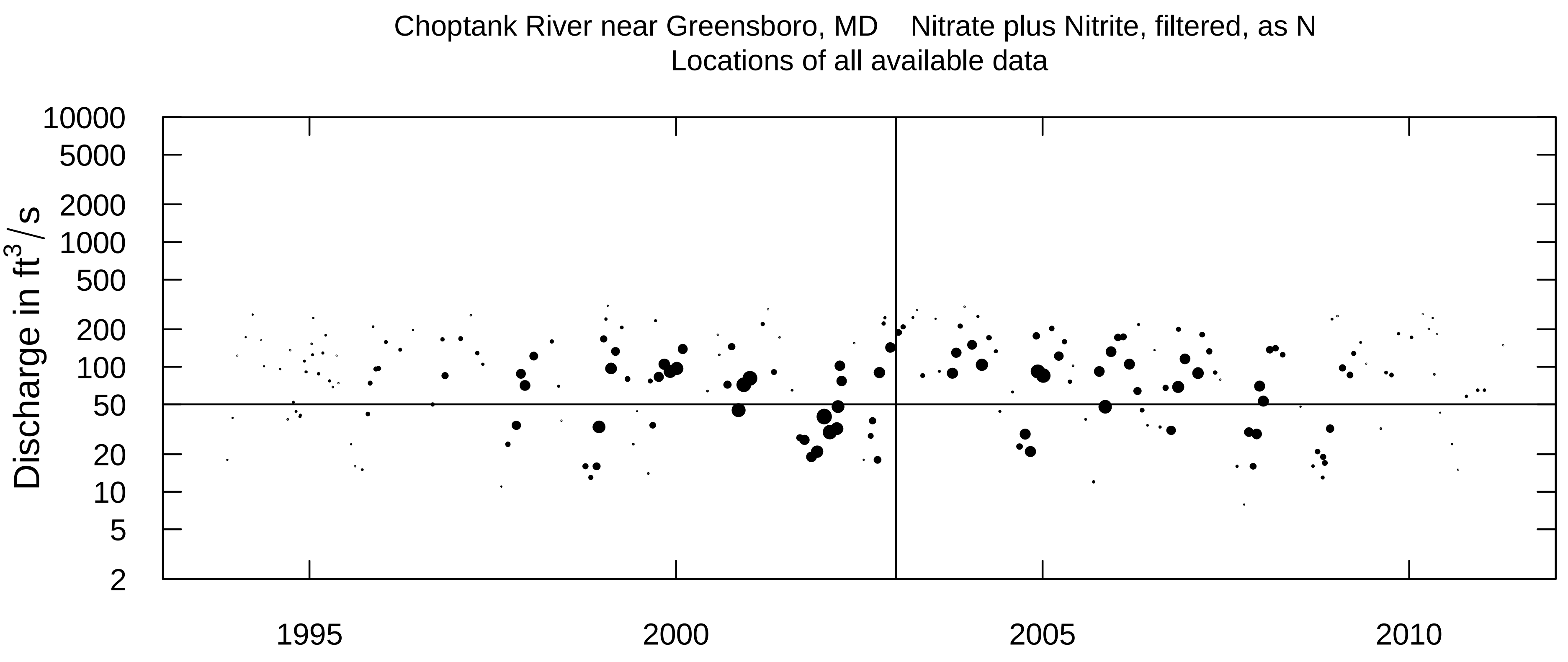
What if we wanted to make an estimate for January 1, 2003 but for Q = 50 cfs

Redo the weights for distance from that point

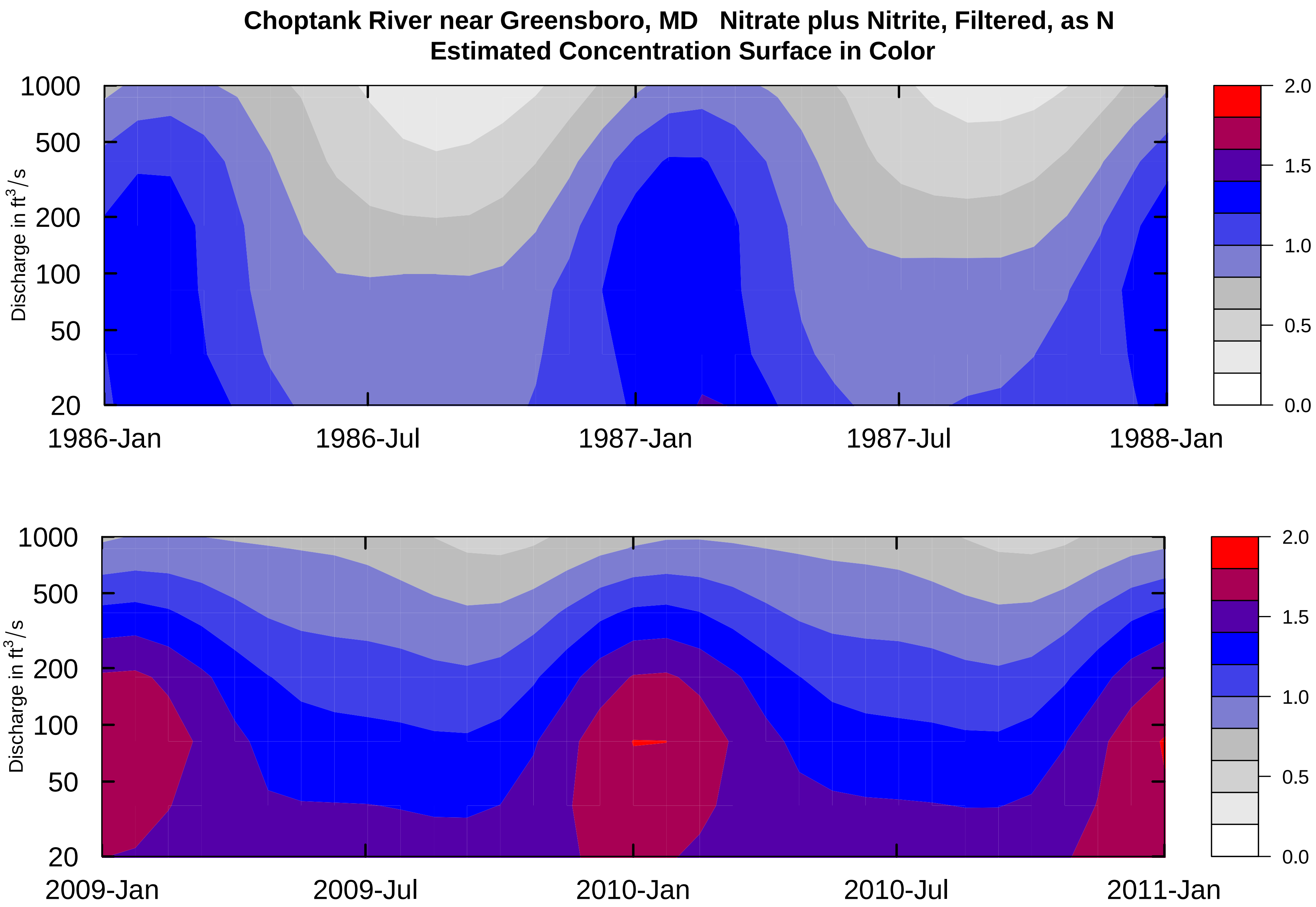


USGS

- To organize the work, lets make estimates for a fine mesh of points in this space.
- We will do it at 14 Q values and 177 time values, for a grid of 2,478 points.



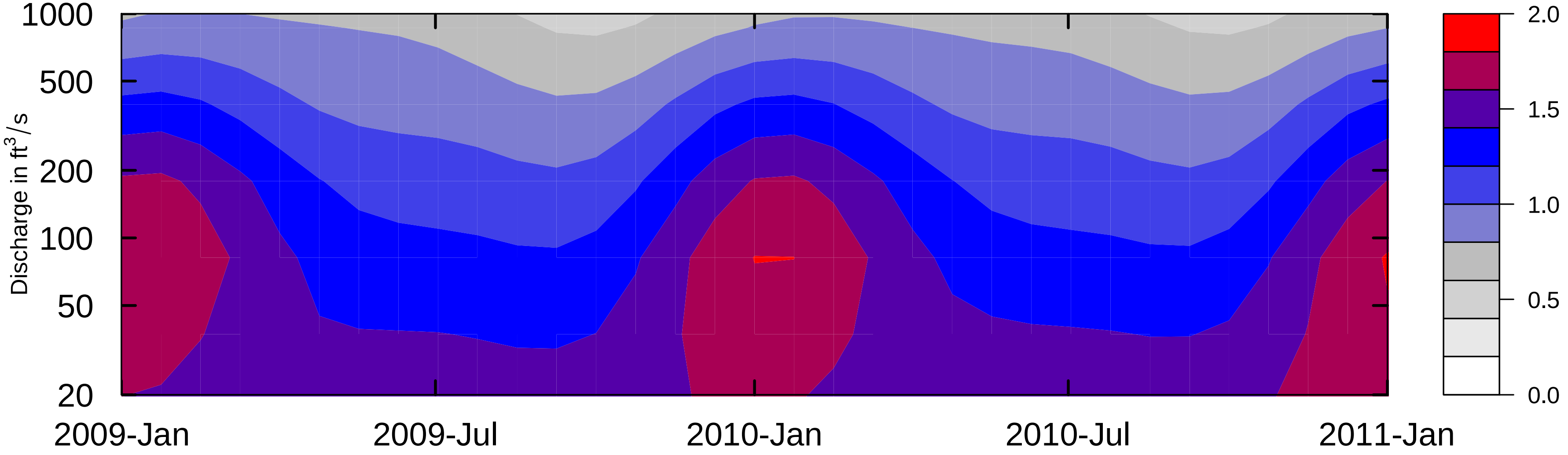
Here are two, more detailed looks at this surface



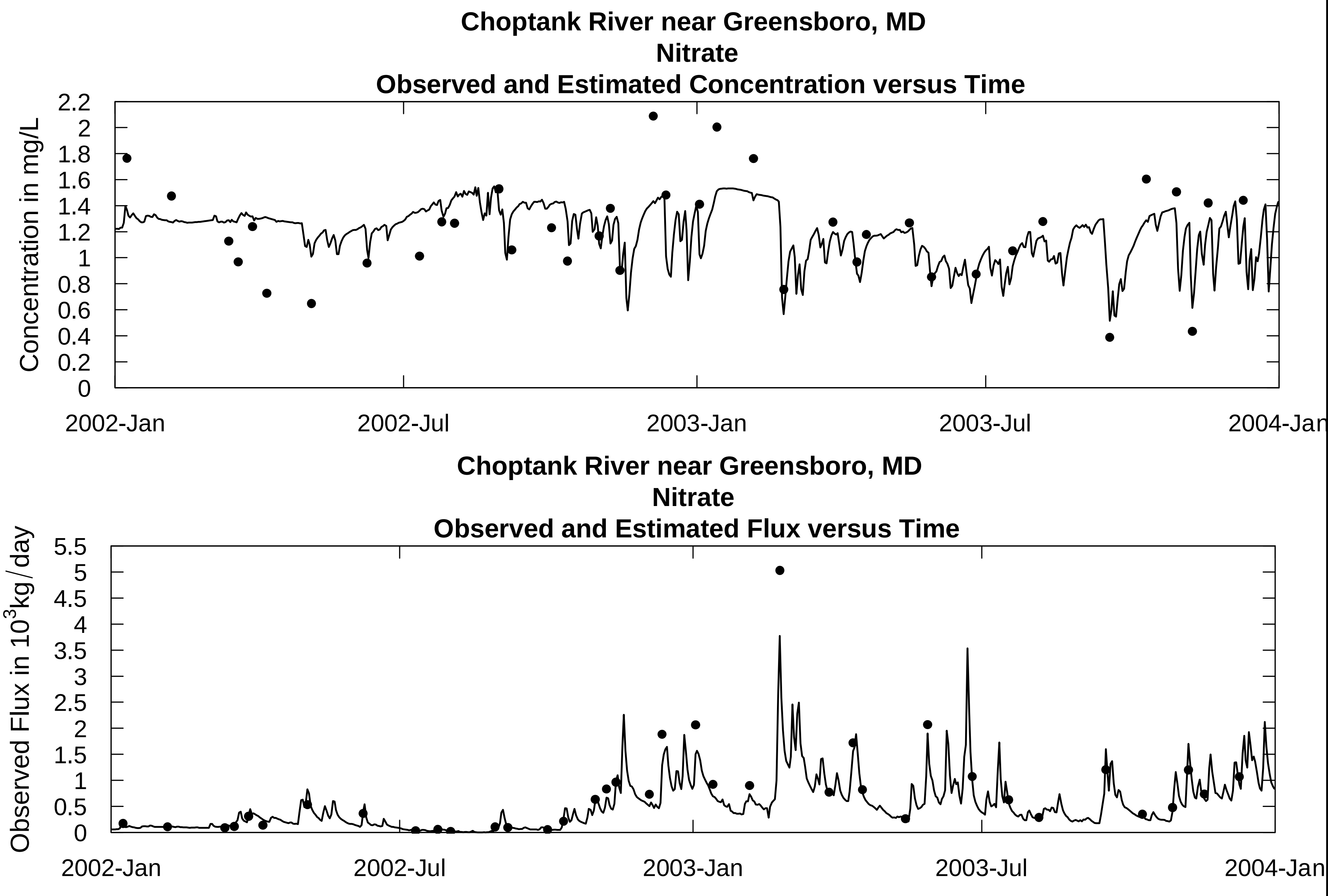
Now, for every one of 10,227 days in the record from 1985 through 2012:

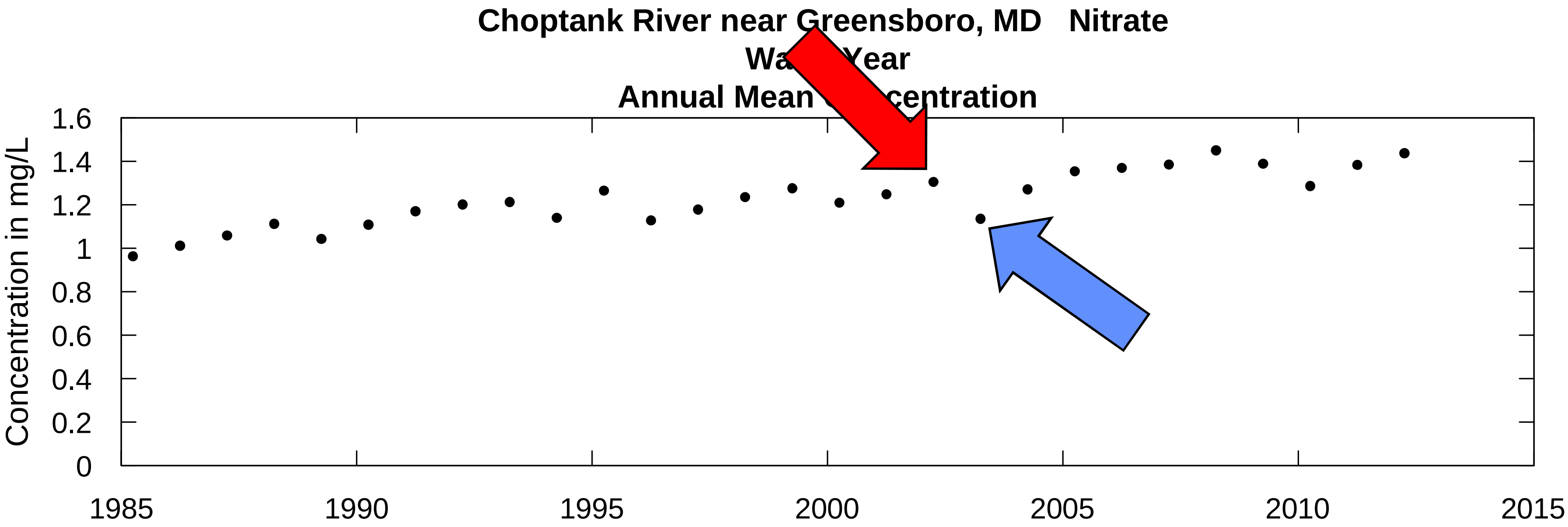
We can use the date and the observed discharge to compute the expected value of concentration.

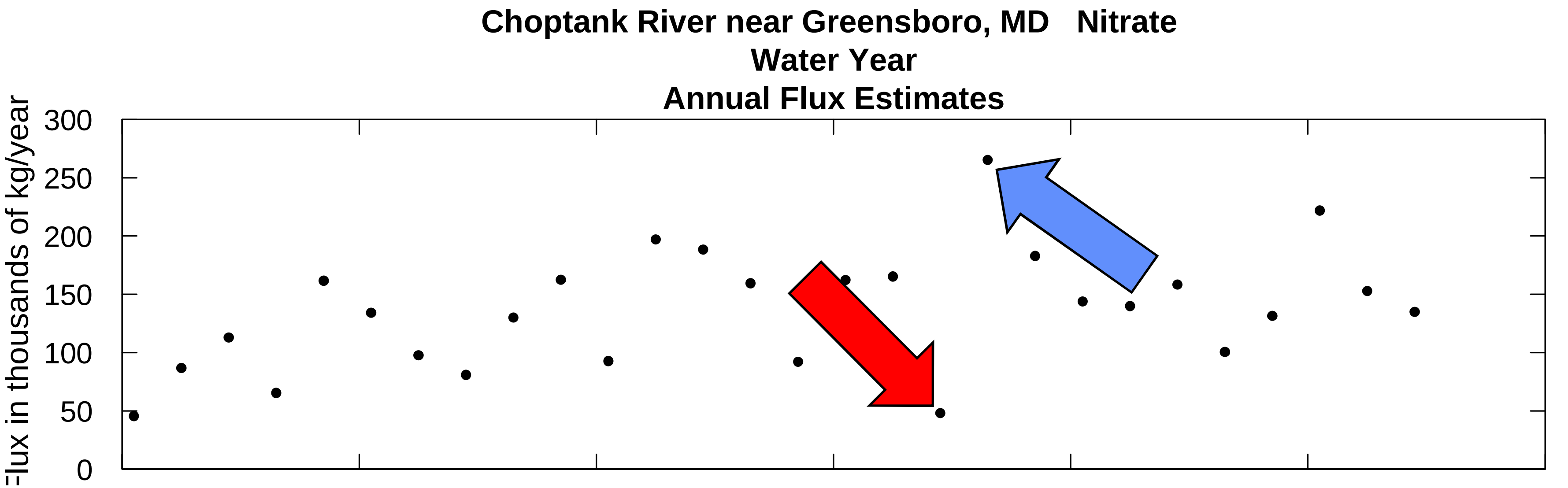
From that value we can compute the expected value of flux.



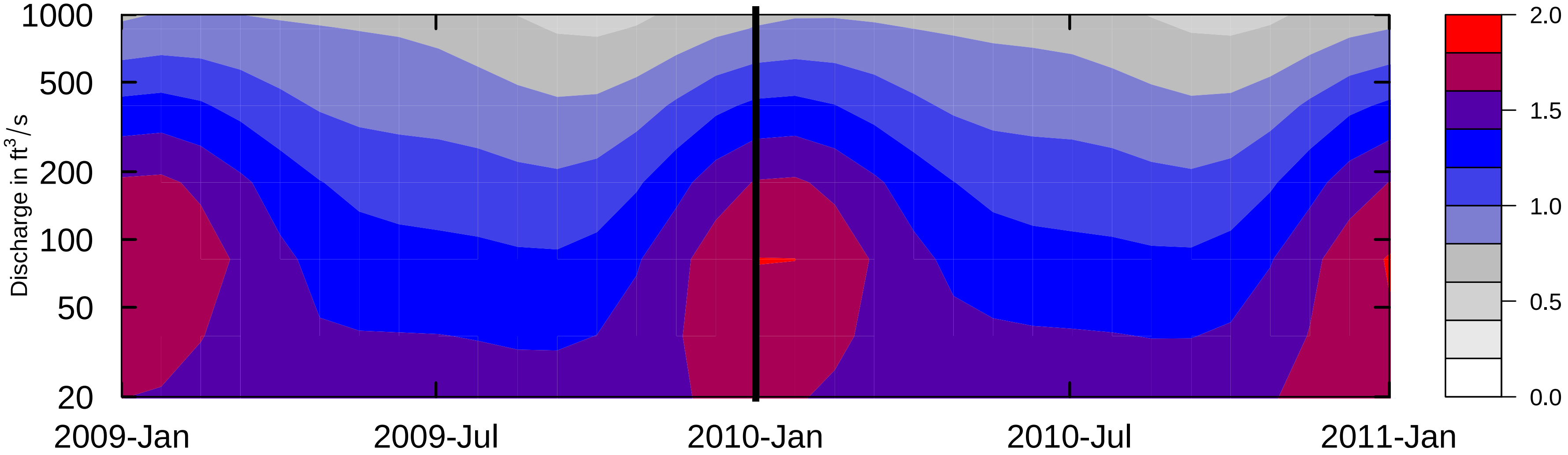
Then we can sum these estimates by year to compute estimates of annual mean concentration & annual mean flux







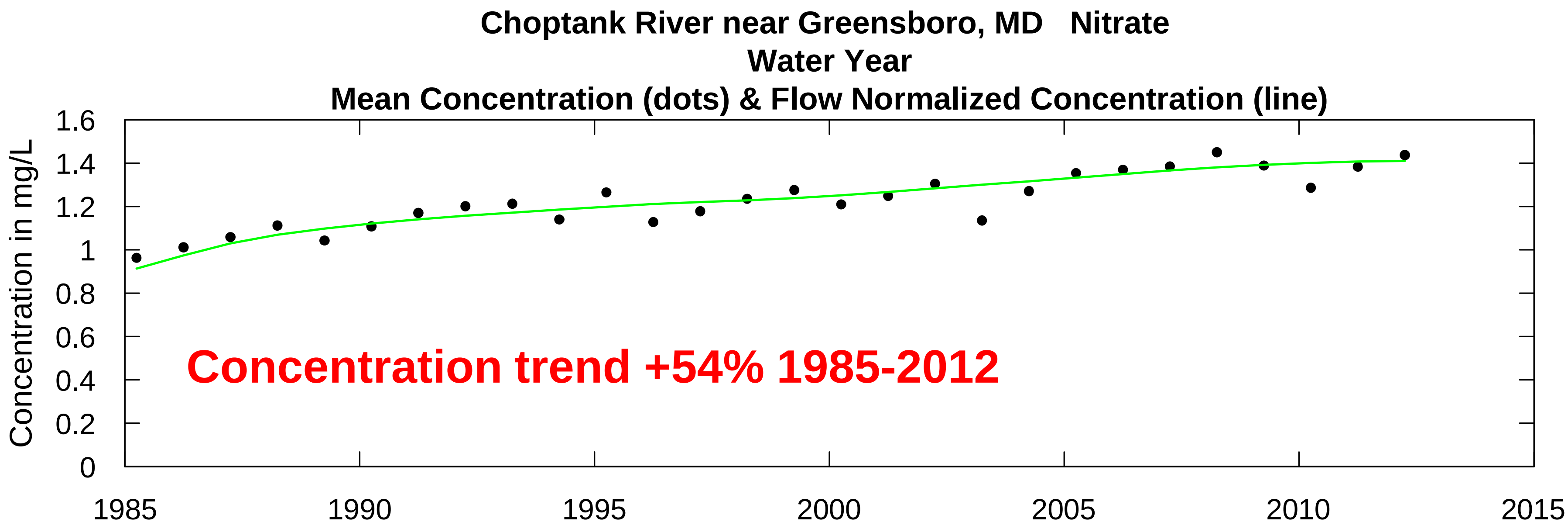
Can we filter out this flow-driven variation to see the underlying change?



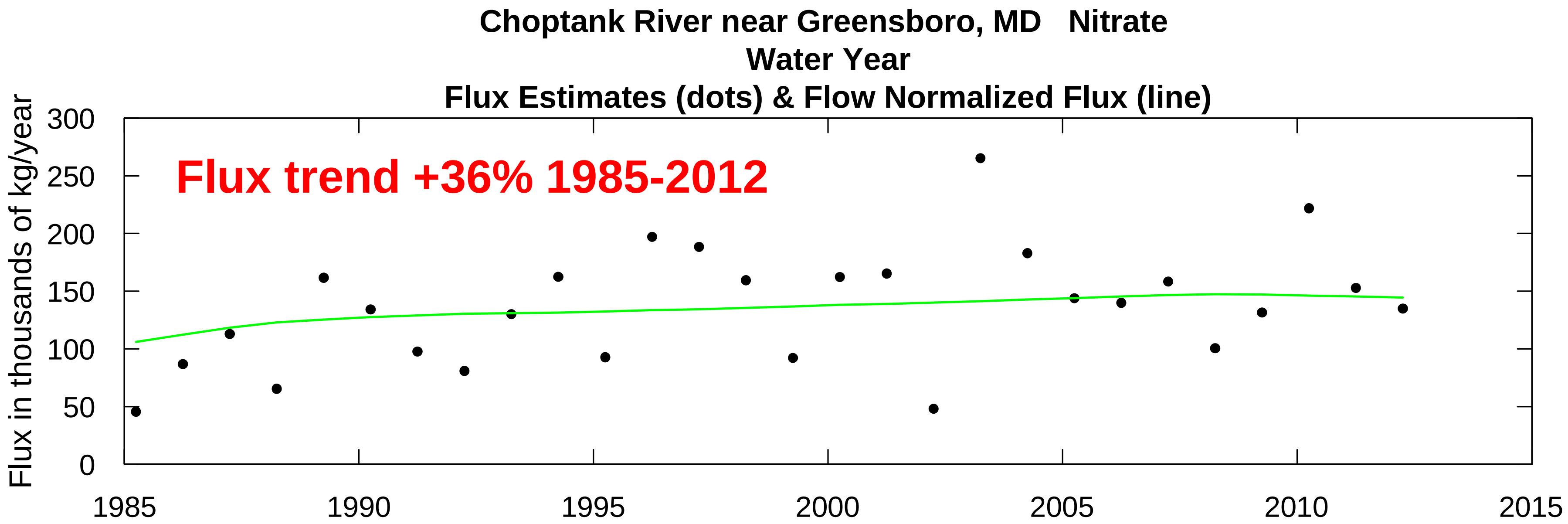
The "flow normalized concentration" on any given day is: c=f(Q,T) integrated over the probability distribution of Q for that day of the year.

Flow normalized flux is just c x Q integrated over discharge.

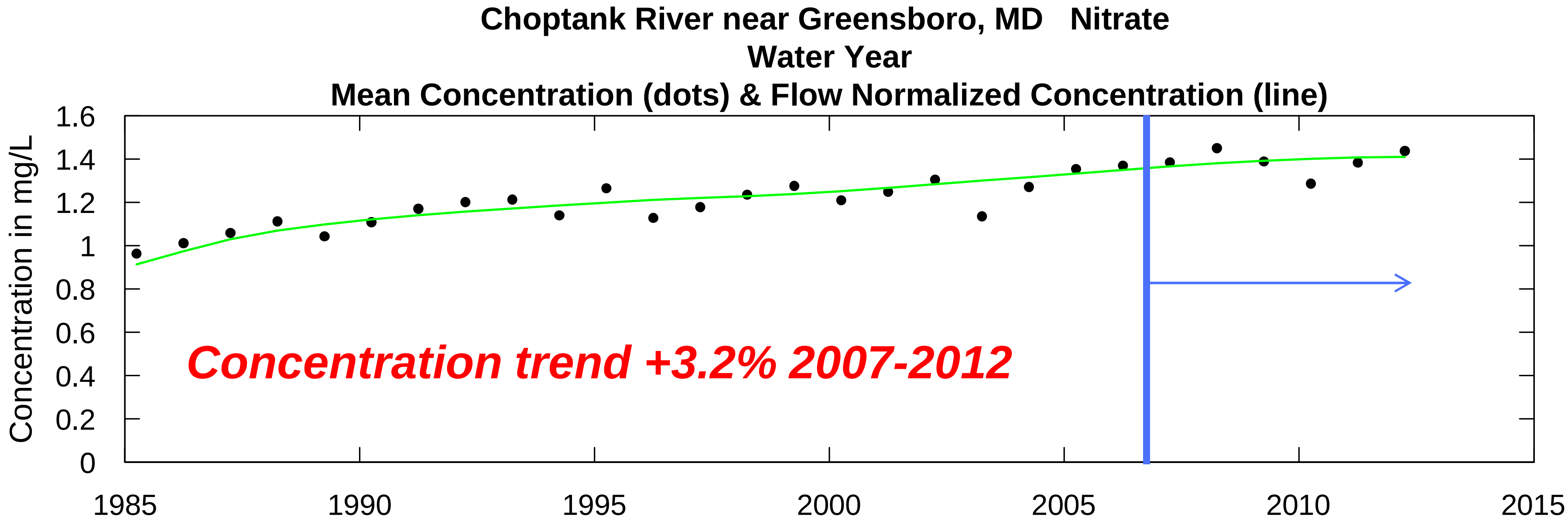
Sum those over the year to get annual flow-normalized mean concentration and flux.



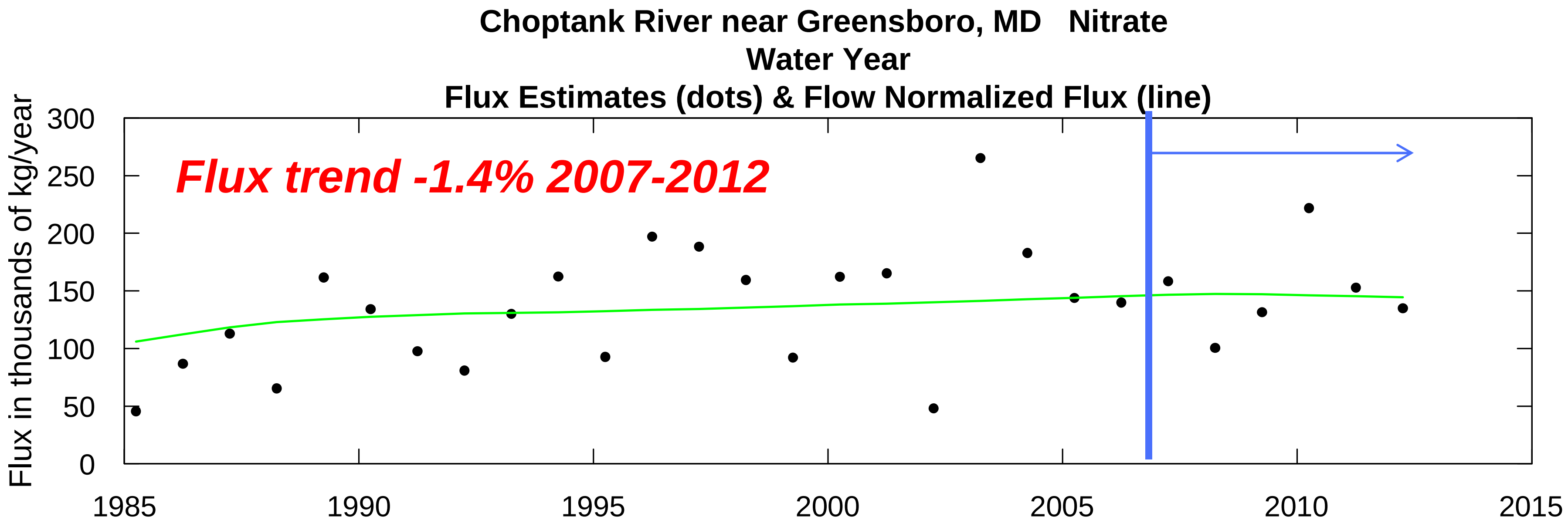
Choptank River near Greensboro, MD Nitrate
Water Year
Mean Concentration (dots) & Flow Normalized Concentration (line)
Concentration in mg/L
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Concentration trend +54% 1985-2012
1985
1990
1995
2000
2005
2010
2015



Choptank River near Greensboro, MD Nitrate
Water Year
Flux Estimates (dots) & Flow Normalized Flux (line)
Flux in thousands of kg/year
300
250
200
150
100
50
0
Flux trend +36% 1985-2012
1985
1990
1995
2000
2005
2010
2015



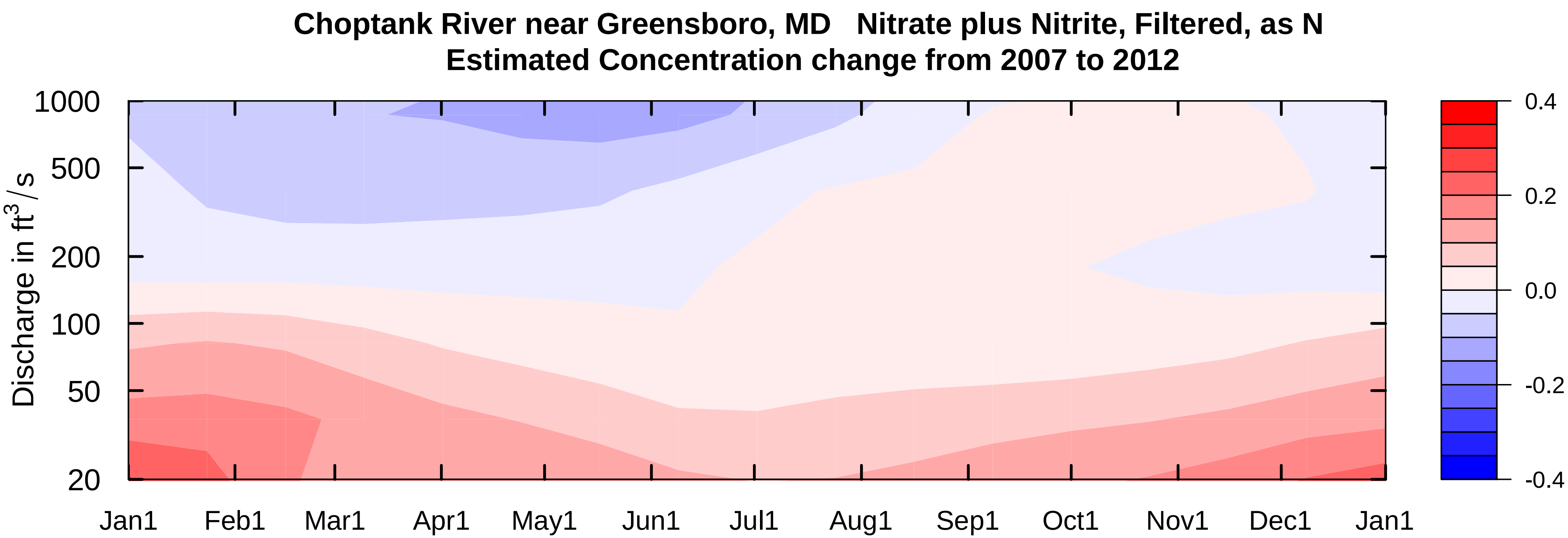
Choptank River near Greensboro, MD Nitrate
Water Year
Mean Concentration (dots) & Flow Normalized Concentration (line)
Concentration in mg/L
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1985
1990
1995
2000
2005
2010
2015
Concentration trend +3.2% 2007-2012



Choptank River near Greensboro, MD Nitrate
Water Year
Flux Estimates (dots) & Flow Normalized Flux (line)
Flux in thousands of kg/year
300
250
200
150
100
50
0
1985
1990
1995
2000
2005
2010
2015
Flux trend -1.4% 2007-2012

Look at changes in just the last few years.

This is a graphic of differences 2007 to 2012



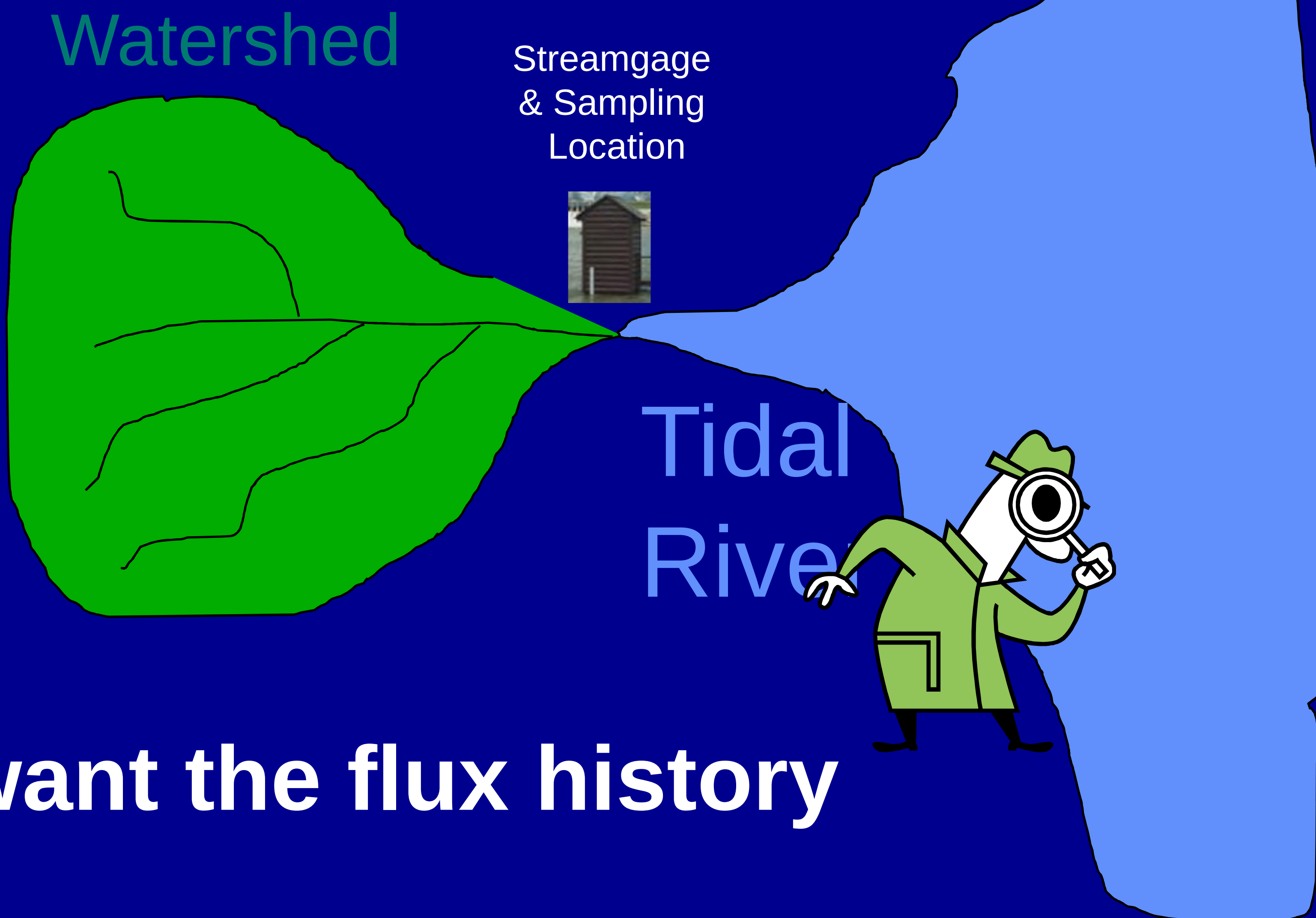
Hypothesis, cover crops are helping at higher flows particularly in the winter. Low flows are still responding to legacy of nitrate enriched groundwater.

Why all this complexity?

Different products for different purposes

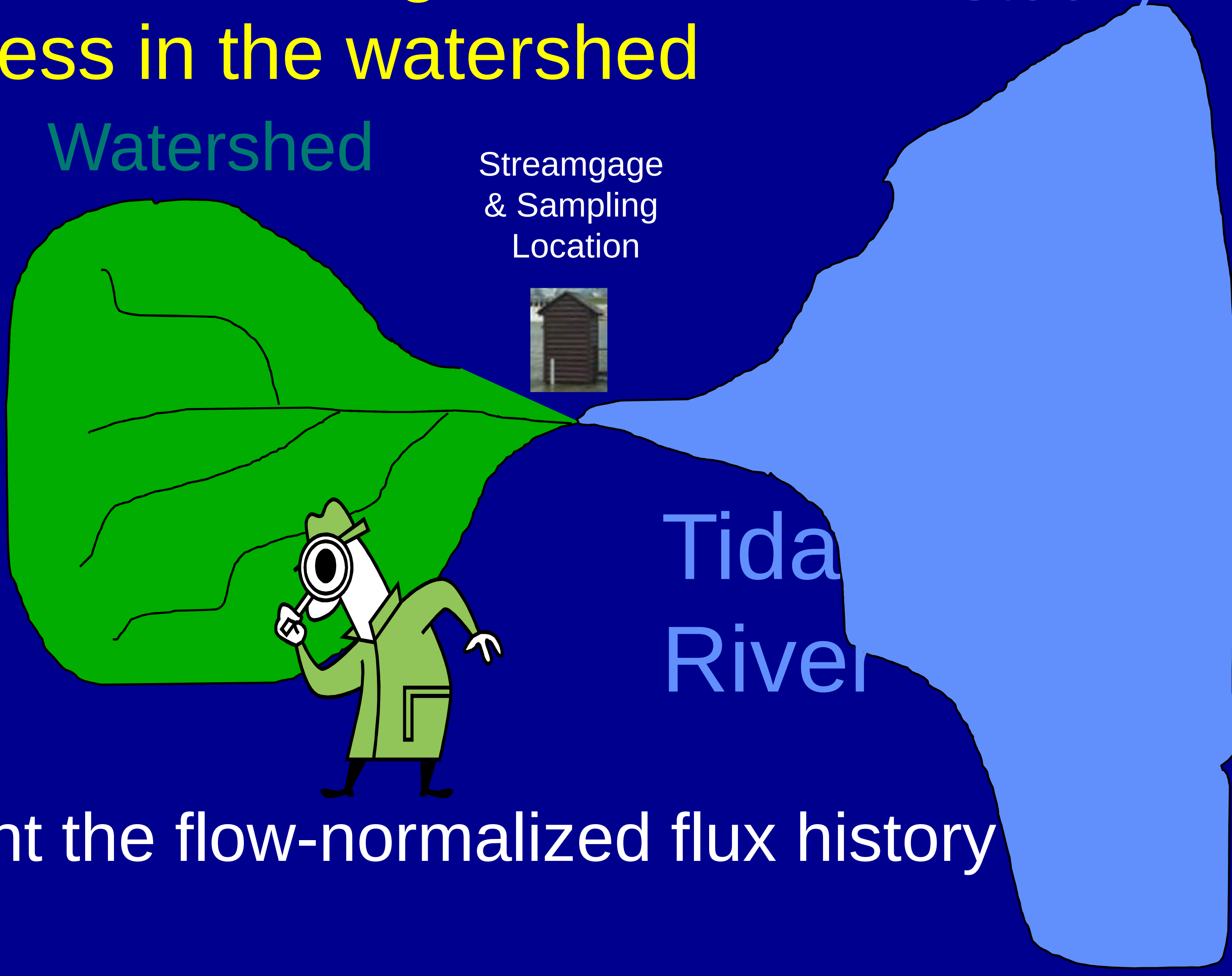
- Concentration versus flux
- Actual history versus flow-normalized history

For understanding impact on the estuary ecosystem



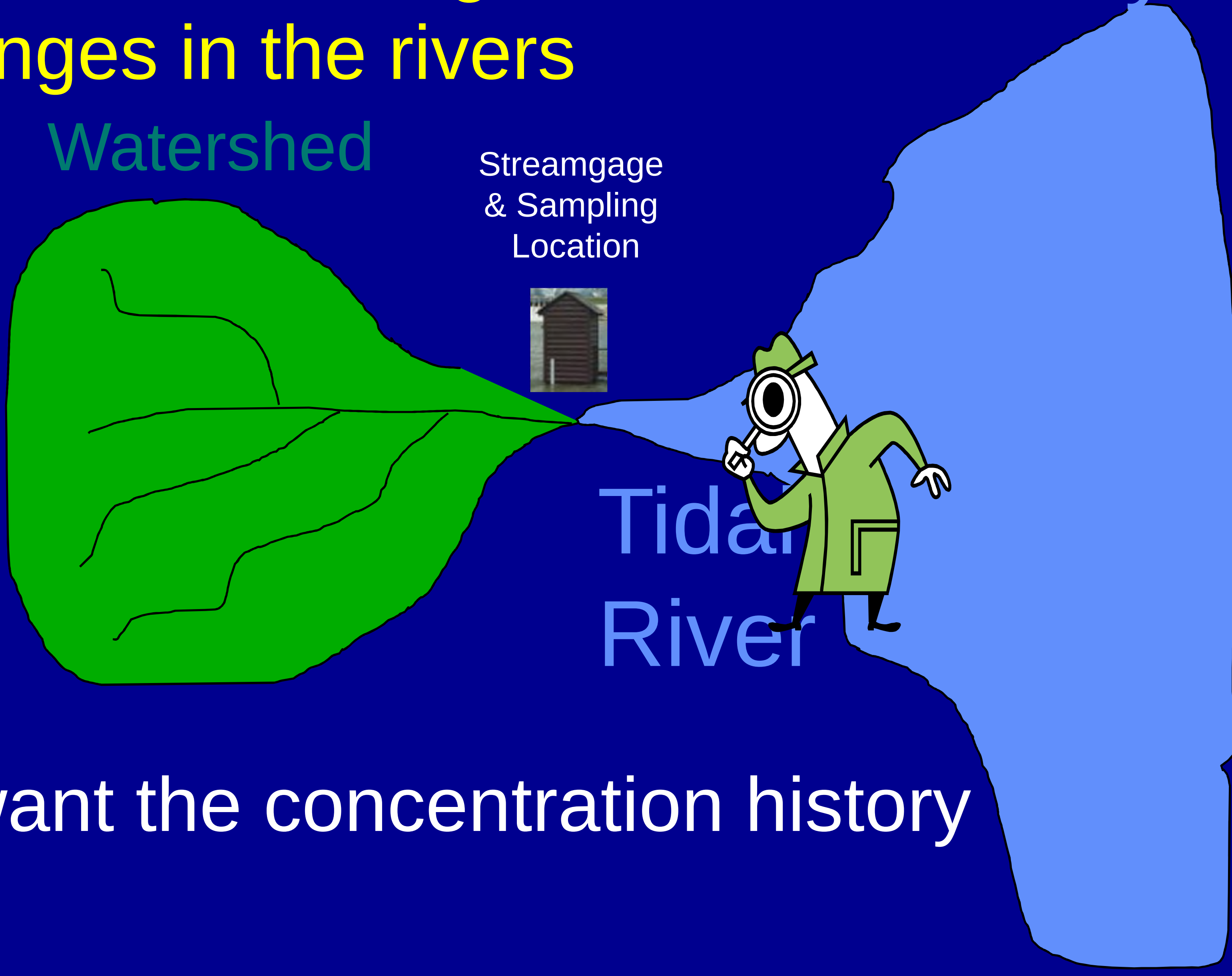
We want the flux history

For understanding progress in the watershed



We want the flow-normalized flux history

For understanding the changes in the rivers

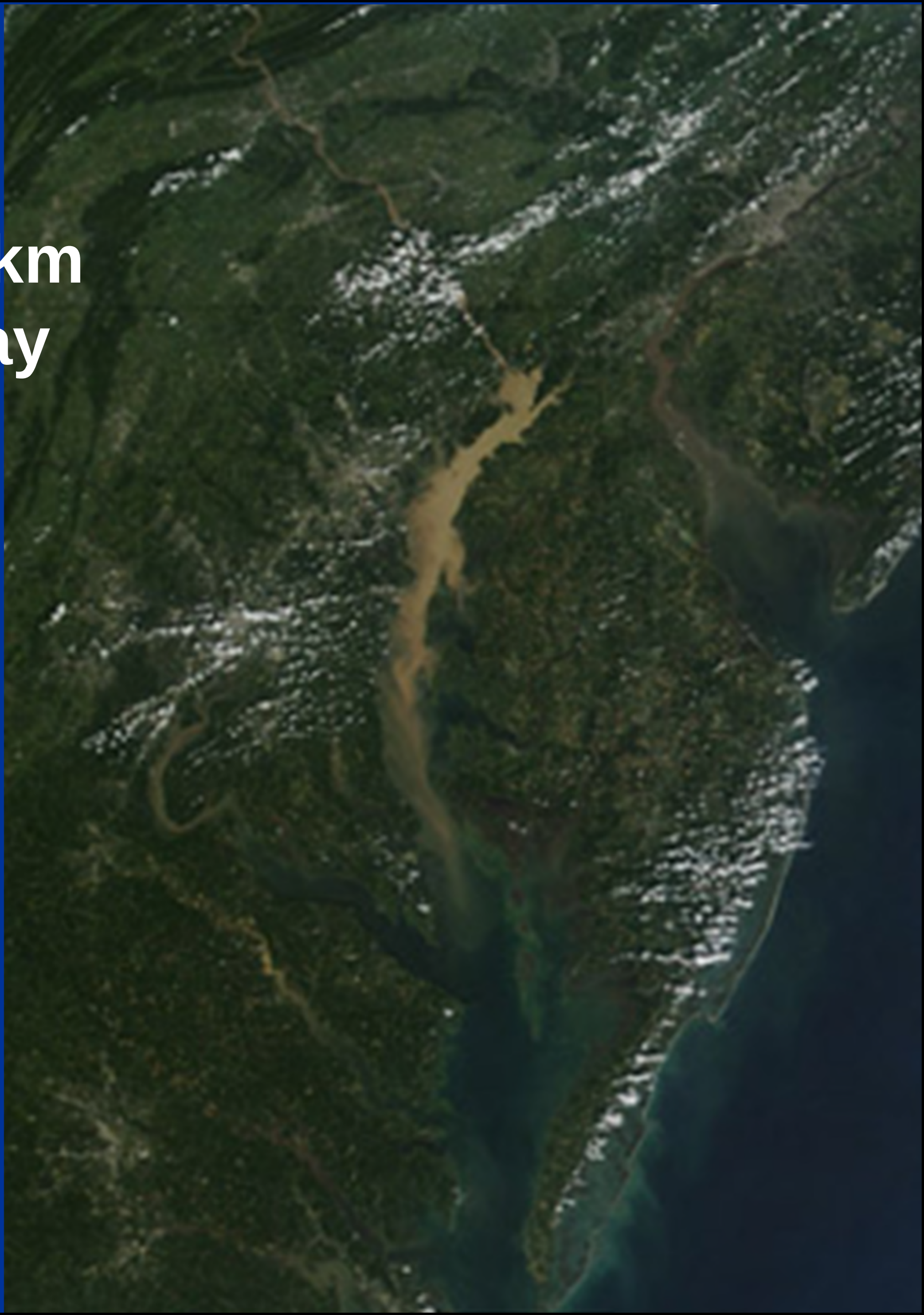


We want the concentration history

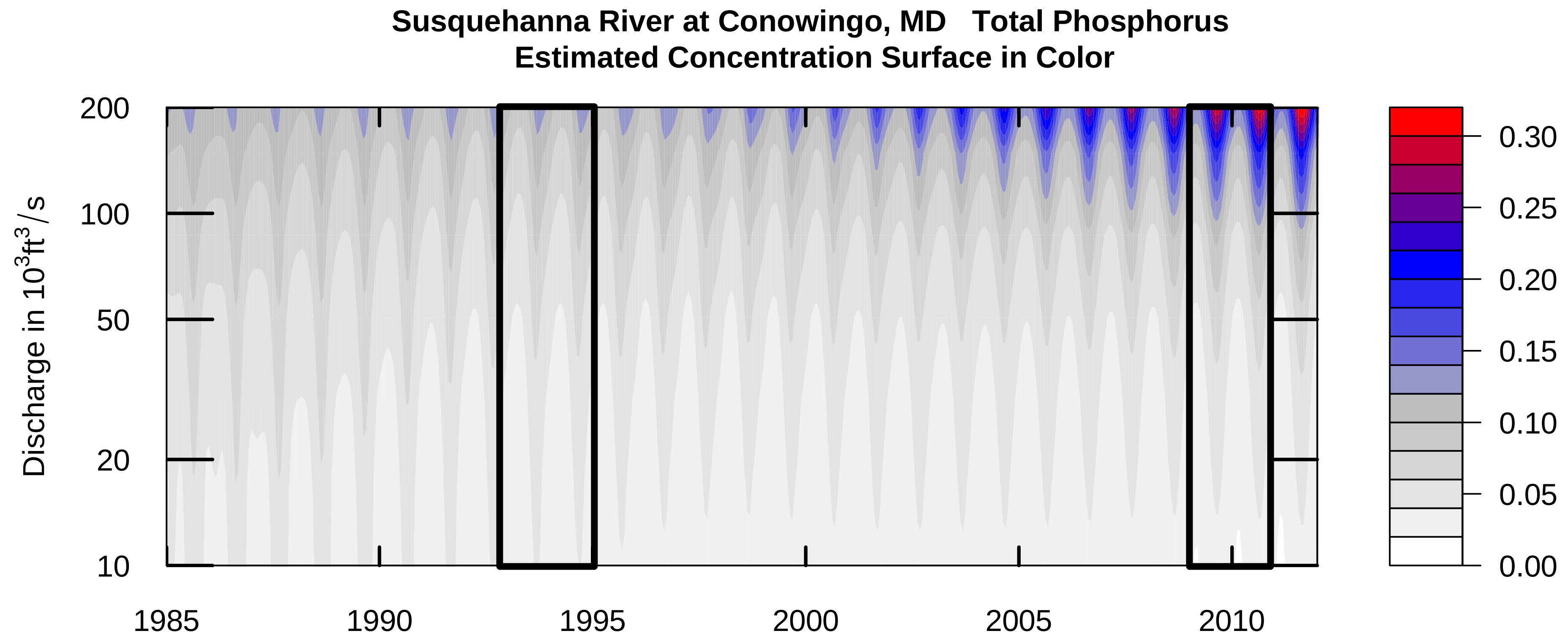
Susquehanna, 70,000 km^2 watershed

Sediment plume from Susquehanna watershed
Plume extends over 150 km down the Chesapeake Bay
Carrying:
Sediment,
Phosphorus,
Nitrogen

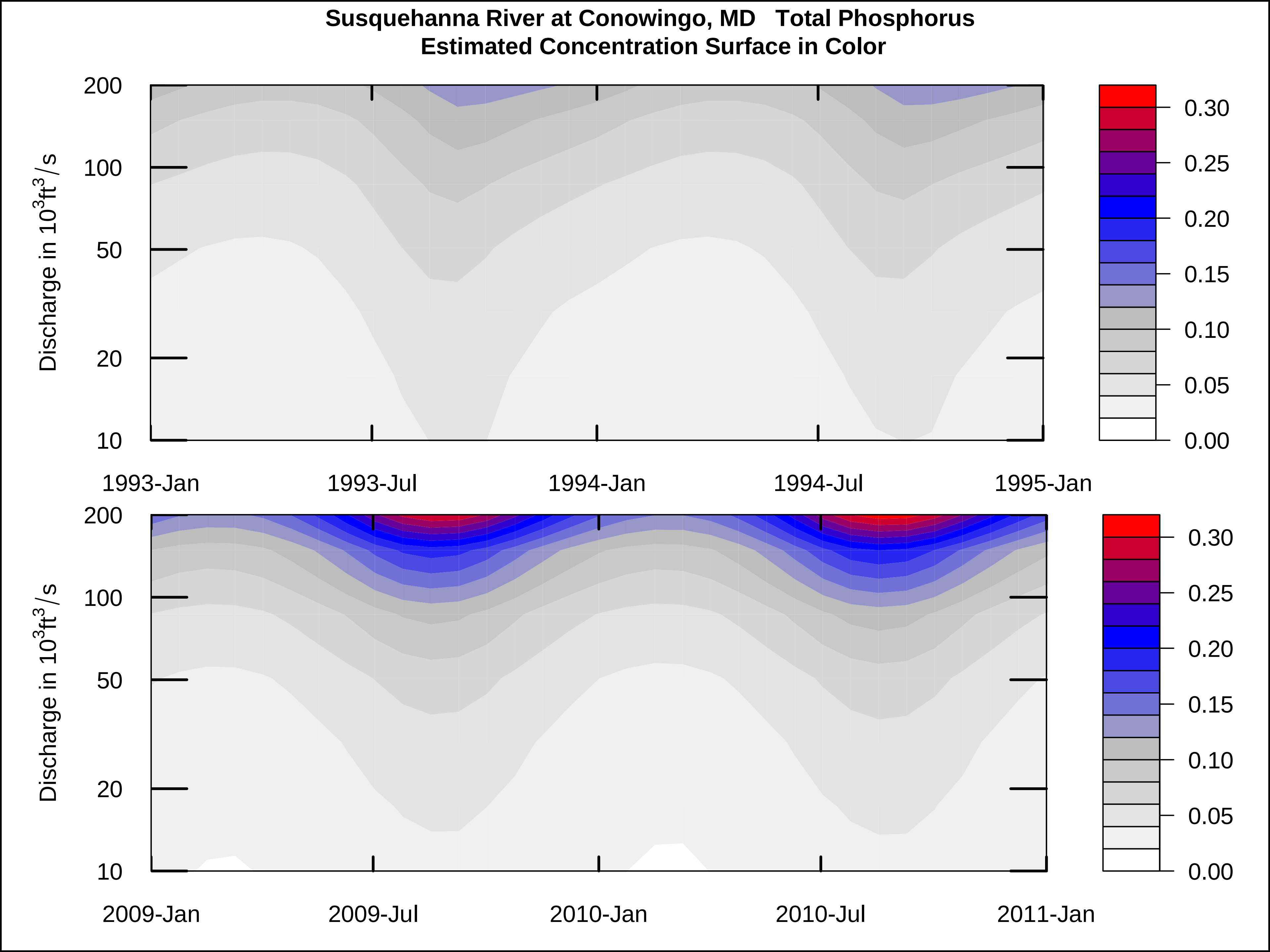
From the watershed and from storage in Conowingo Reservoir



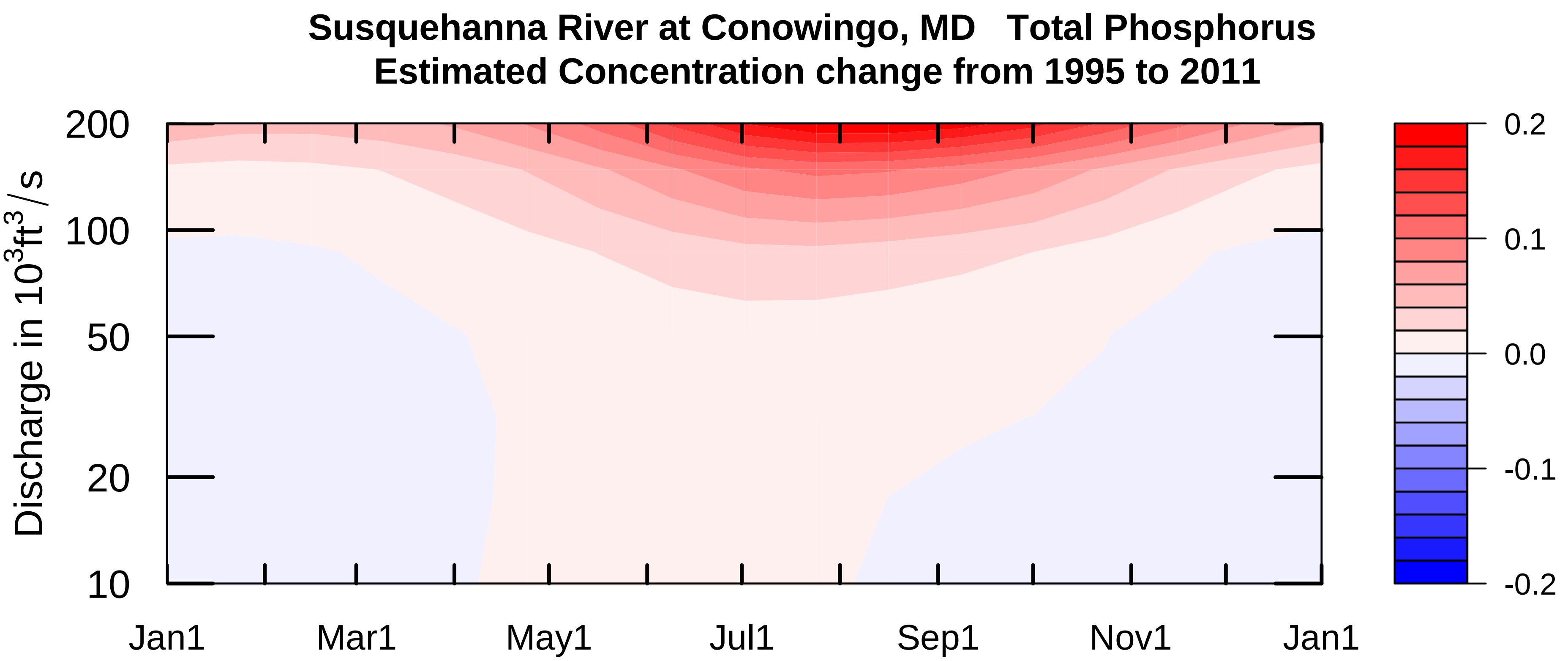
USGS



Conowingo Dam during Tropical Storm Lee, September 2011, Reservoir is rapidly filling, Trap efficiency in decline



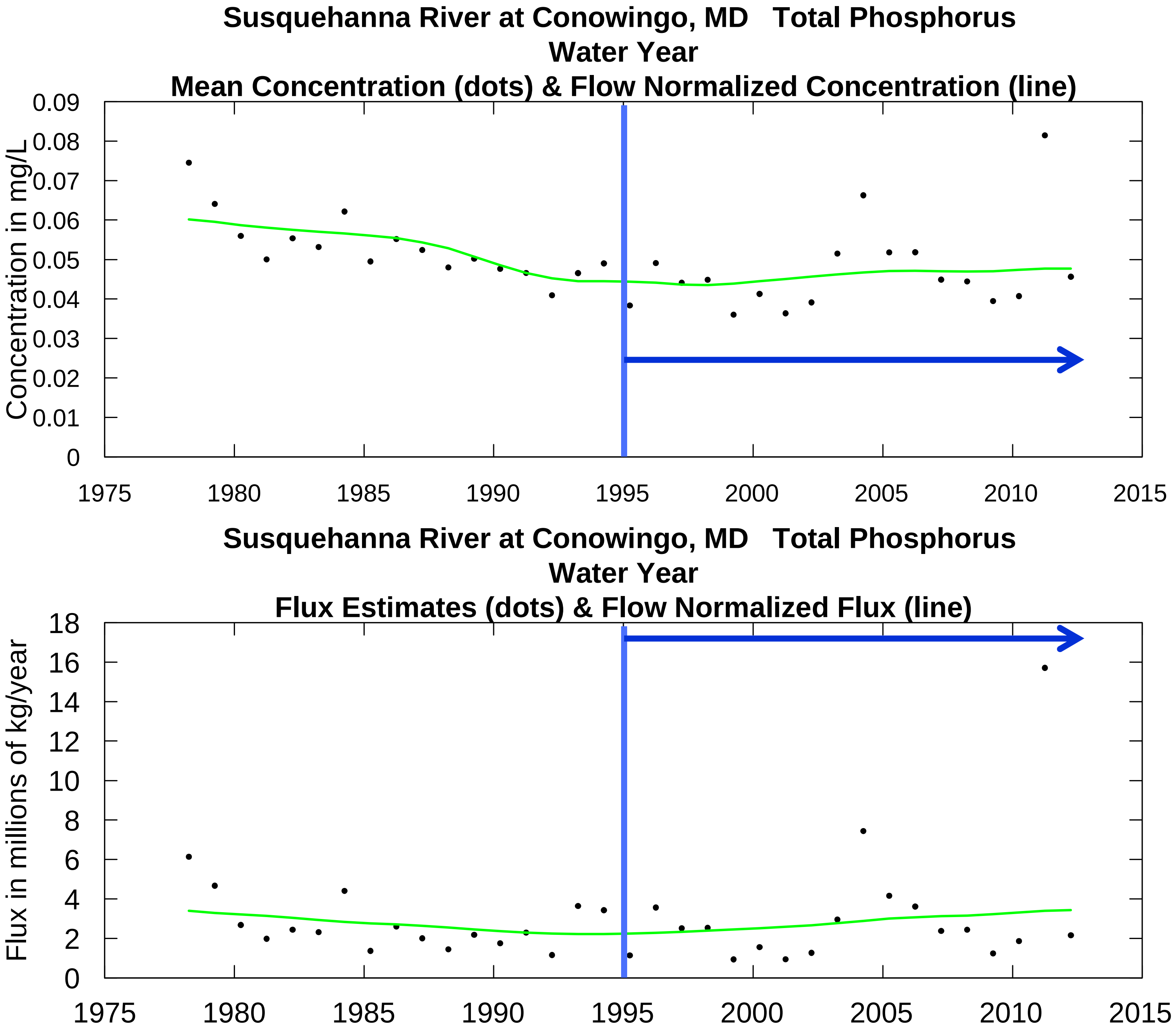
Susquehanna River at Conowingo, MD Total Phosphorus
Estimated Concentration Surface in Color
Discharge in $10^3 ft^3/s$
200
100
50
20
10
1993-Jan
1993-Jul
1994-Jan
1994-Jul
1995-Jan
0.30
0.25
0.20
0.15
0.10
0.05
0.00
2009-Jan
2009-Jul
2010-Jan
2010-Jul
2011-Jan


Susquehanna River at Conowingo, MD   Total Phosphorus
Estimated Concentration change from 1995 to 2011
Discharge in 10^3ft^3/s
200
100
50
20
10
Jan1
Mar1
May1
Jul1
Sep1
Nov1
Jan1
0.2
0.1
0.0
-0.1
-0.2

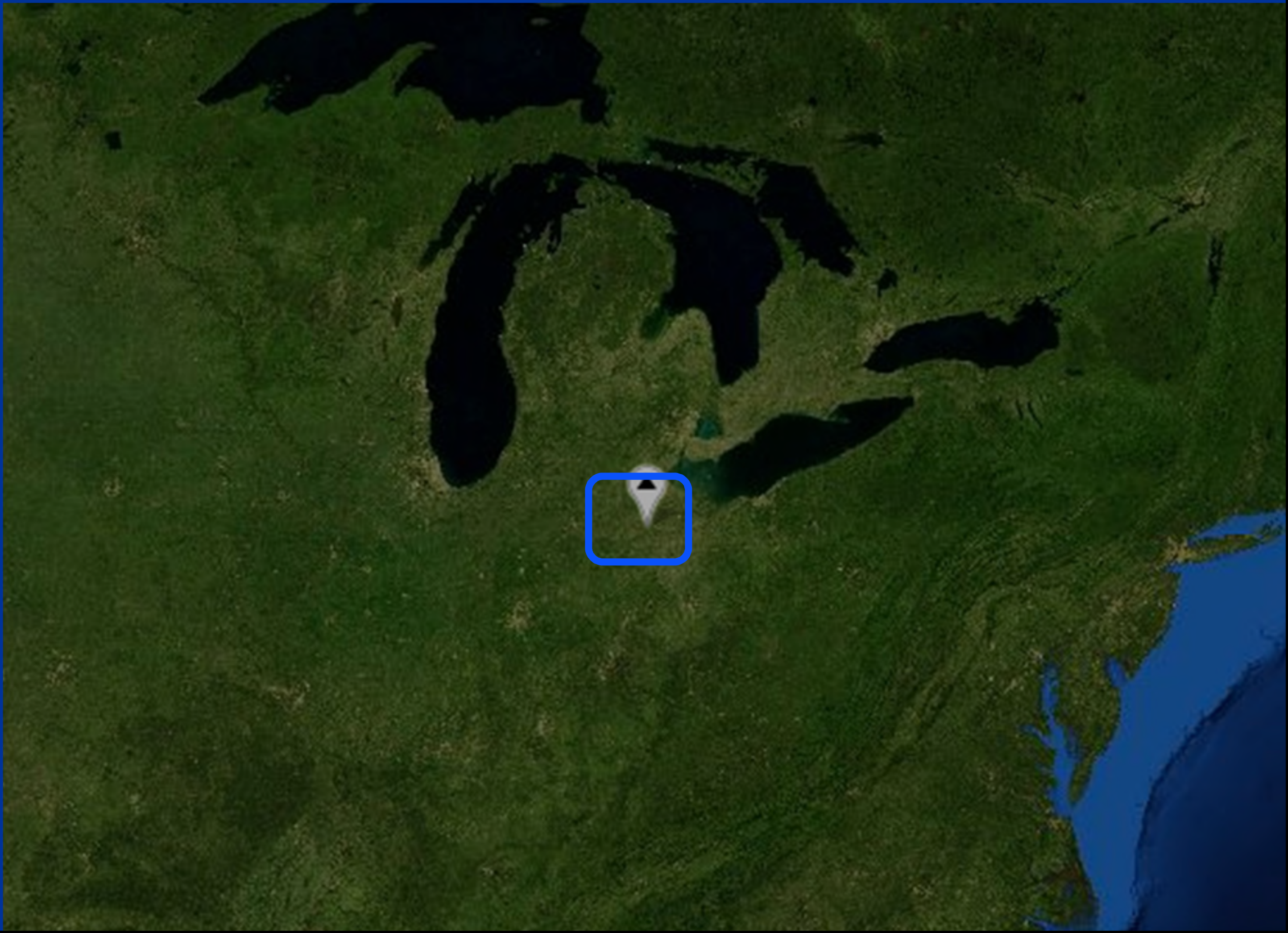
Change from 1995-2012

0.4%/yr

3.1%/yr



Maumee River – 16,000 km^2 Tributary to Lake Erie



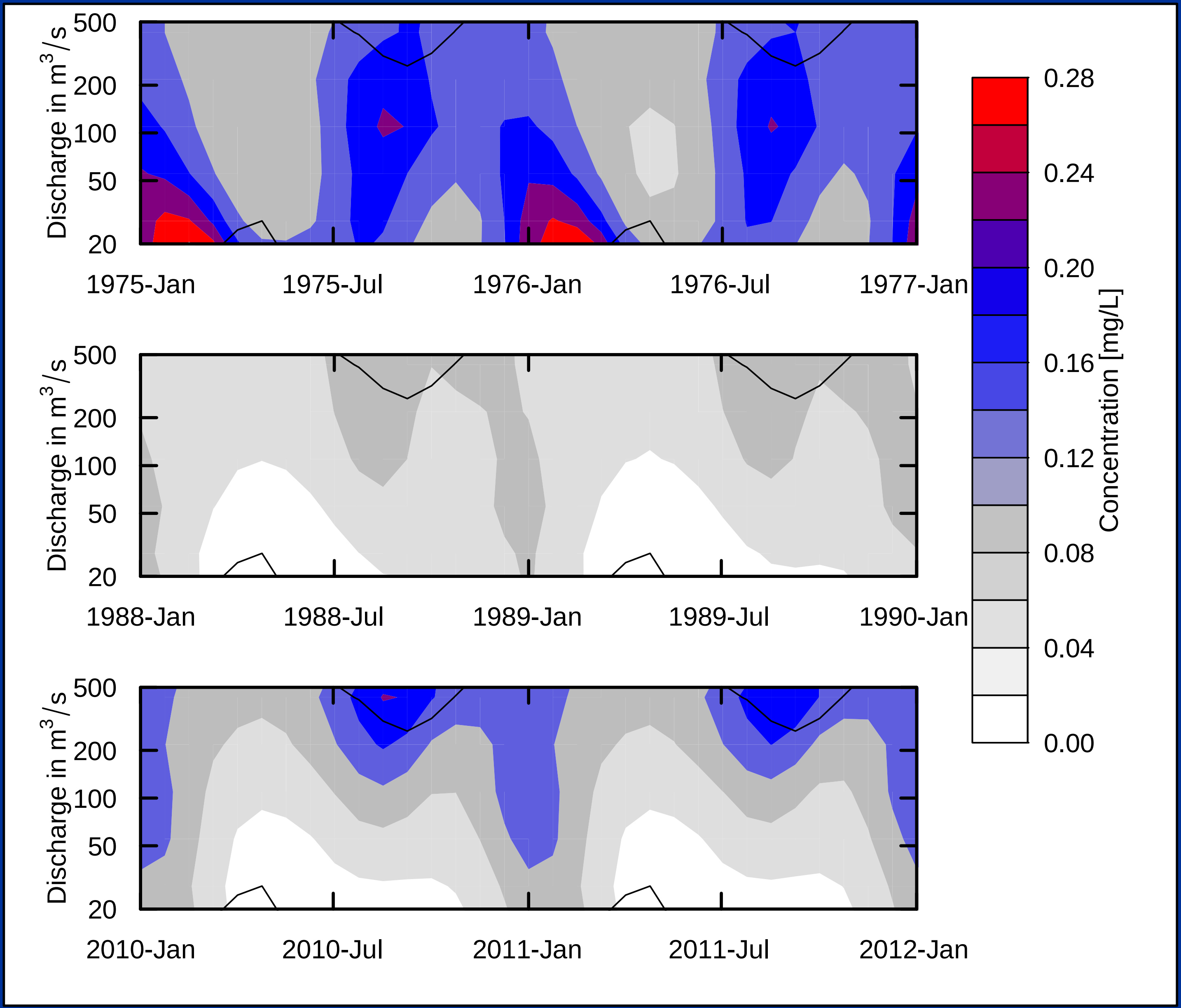
Cyanobacter – Lake Erie

Dissolved Reactive Phosphorus, Maumee River, at Waterville, OH

Mid 1970's

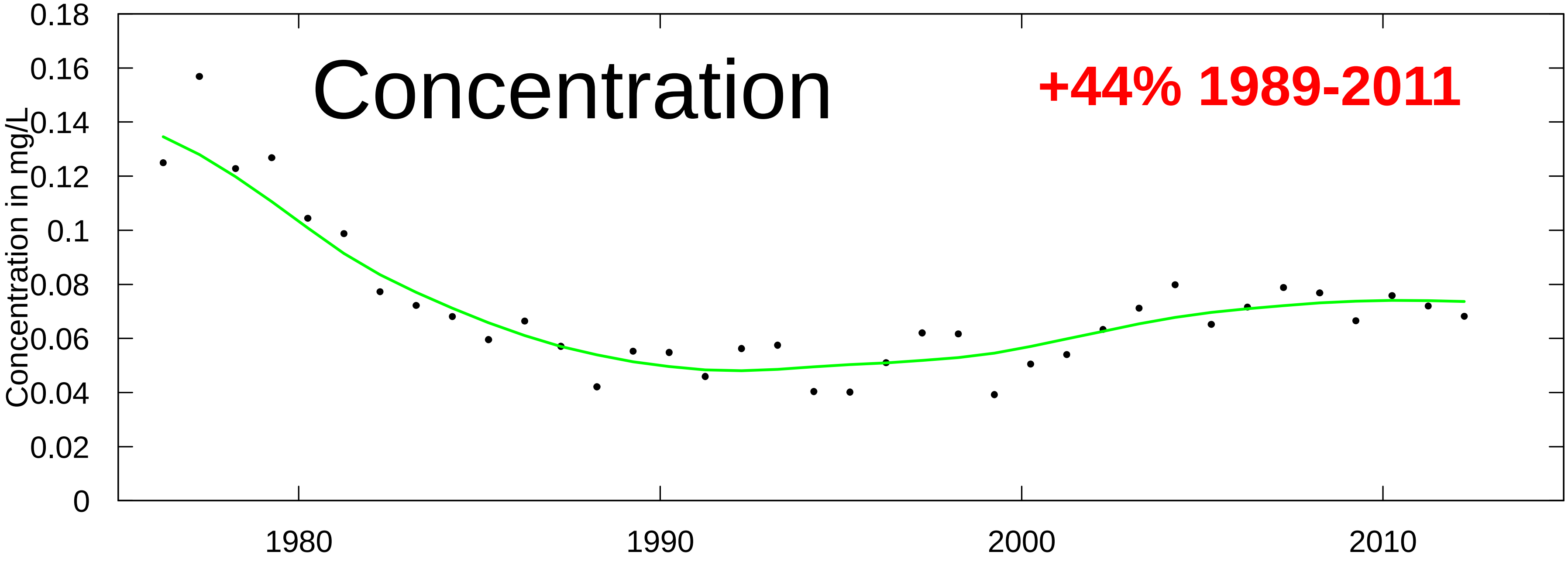
Late 1980's

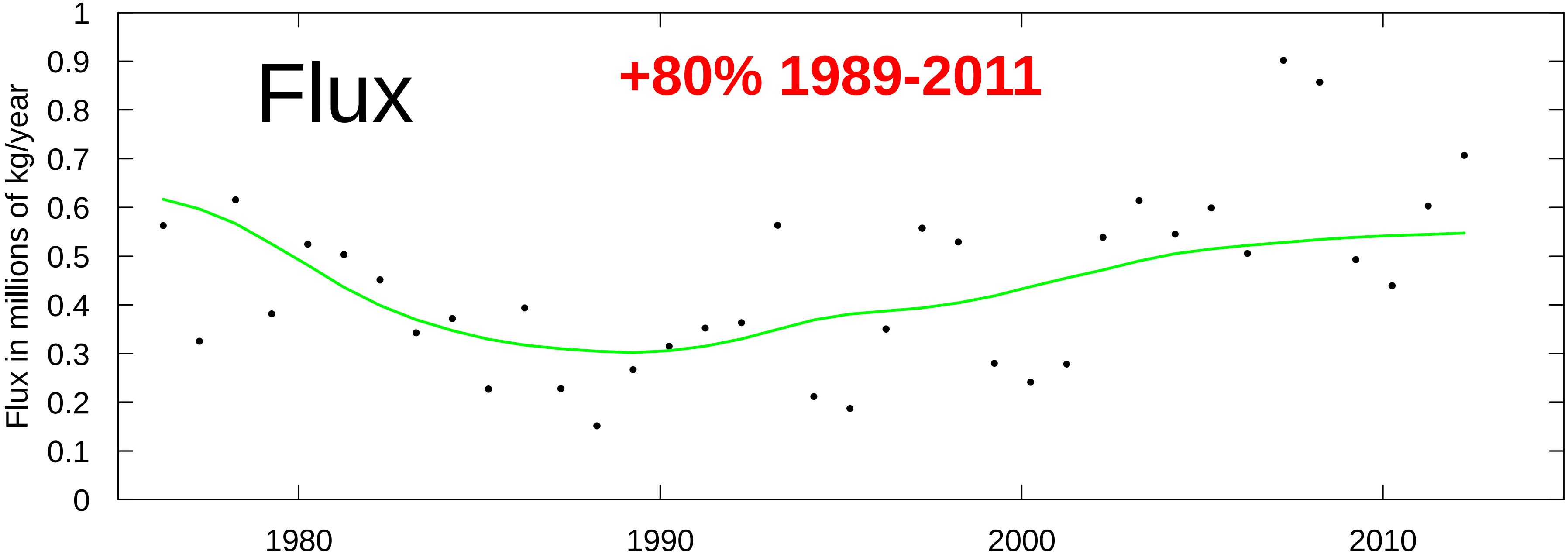
Early 2010's



USGS

Dissolved Reactive Phosphorus, Maumee River, at Waterville, OH





dataRetrieval & EGRET (Exploration and Graphics for RivEr Trends)

- dataRetrieval: ingests the data from multiple sources and puts them in a form suitable for analysis by EGRET or generic use of R
- EGRET includes the WRTDS computations: for trend analysis, flux computation, and exploration
- Also includes analysis of streamflow trends
- Produces a wide range of graphical and tabular outputs

dataRetrieval

- Can acquire data from web services (USGS or Storet) as well as from user-supplied files
- Includes capability for sample data, daily discharge, other daily values, sensor data, and meta data about site and parameter
- Structures the data to be conveniently used by the EGRET software

USGS

Getting Started

- Need to install R (freely downloaded from http://cran.us.r-project.org/
- Next install the EGRET & dataRetrieval packages & the other R packages they depend on:
- In R, run the following commands:
  - install.packages(c("zoo","survival","stringr", "fields","spam","XML","Rcurl","plyr", "reshape2"))
  - install.packages("dataRetrieval",repos=http://usgs-r.github.com,type="both")
  - install.packages("EGRET",repos="http://usgs-r.github.com",type="both")

USGS

Getting Started 2

- Then each time packages are used, they need to be loaded, using the commands

  library(dataRetrieval)

  library(EGRET)

- Once this is done you will have access to help and to the package vignettes.
- To get help with a function (such as the function getSampleData) just type ?getSampleData

How can we enter data

- For the water quality sample data
  - From USGS web services
  - From EPA-Storet web services
  - From a user supplied file
- For the daily discharge data
  - From USGS web services
  - From a user supplied file
- For the meta-data
  - From USGS web services
  - From user entries

```
> library(dataRetrieval)
> library(EGRET)
> siteID <- "01491000"
> parameter_cd <- "00631"
> startDate <- "1979-09-01"
> endDate <- "2011-09-30"
> Sample <- getSampleData(siteID,parameter_cd,startDate,endDate)
> summary(Sample)
```

```
> library(dataRetrieval)
> library(EGRET)
> siteID <- "01491000"
> parameter_cd <- "00631"
> startDate <- "1979-09-01"
> endDate <- "2011-09-30"
> Sample <- getSampleData(siteID, parameter_cd, startDate, endDate)
> summary(Sample)
      Date                ConcLow          ConcHigh          Uncen           ConcAve          Julian          Month      
 Min.   :1979-09-25   Min.   :0.176   Min.   :0.050   Min.   :0.0000   Min.   :0.025   Min.   :47383   Min.   : 1.00  
 1st Qu.:1989-01-06   1st Qu.:0.880   1st Qu.:0.880   1st Qu.:1.0000   1st Qu.:0.880   1st Qu.:50775   1st Qu.: 3.00  
 Median :1994-03-15   Median :1.100   Median :1.100   Median :1.0000   Median :1.100   Median :52668   Median : 6.00  
 Mean   :1995-05-23   Mean   :1.124   Mean   :1.122   Mean   :0.9985   Mean   :1.122   Mean   :53102   Mean   : 6.07  
 3rd Qu.:2002-05-29   3rd Qu.:1.400   3rd Qu.:1.400   3rd Qu.:1.0000   3rd Qu.:1.400   3rd Qu.:55665   3rd Qu.: 9.00  
 Max.   :2011-09-29   Max.   :2.430   Max.   :2.430   Max.   :1.0000   Max.   :2.430   Max.   :59075   Max.   :12.00  
                      NA's   :1                                                                                       
      Day            DecYear        MonthSeq        SinDY               CosDY          
 Min.   :  2.00   Min.   :1980   Min.   :1557   Min.   :-1.00000   Min.   :-0.999963  
 1st Qu.: 82.25   1st Qu.:1989   1st Qu.:1669   1st Qu.:-0.63727   1st Qu.:-0.672949  
 Median :157.00   Median :1994   Median :1731   Median : 0.17928   Median : 0.000000  
 Mean   :169.03   Mean   :1995   Mean   :1745   Mean   : 0.08233   Mean   : 0.003879  
 3rd Qu.:256.75   3rd Qu.:2002   3rd Qu.:1830   3rd Qu.: 0.78412   3rd Qu.: 0.722117  
 Max.   :364.00   Max.   :2012   Max.   :1941   Max.   : 1.00000   Max.   : 0.999668  

> length(Sample$Date)
[1] 654
> 
```

> Daily <- getDVData(siteID,"00060",startDate,endDate)

```
>
>
> Daily <- getDVData(siteID,"00060",startDate,endDate)
There are 11718 data points, and 11718 days.
> summary(Daily)
      Date                  Q                  Julian          Month             Day             DecYear
 Min.   :1979-09-01   Min.   :  0.00991   Min.   :47359   Min.   : 1.000   Min.   :  1.0   Min.   :1980
 1st Qu.:1987-09-08   1st Qu.:  0.93446   1st Qu.:50288   1st Qu.: 4.000   1st Qu.: 93.0   1st Qu.:1988
 Median :1995-09-15   Median :  2.40693   Median :53218   Median : 7.000   Median :184.0   Median :1996
 Mean   :1995-09-15   Mean   :  4.08210   Mean   :53218   Mean   : 6.529   Mean   :183.9   Mean   :1996
 3rd Qu.:2003-09-22   3rd Qu.:  4.61565   3rd Qu.:56147   3rd Qu.:10.000   3rd Qu.:275.0   3rd Qu.:2004
 Max.   :2011-09-30   Max.   :246.35656   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          NA's   :29
> length(Daily$Q)
[1] 11718
> |
```

Censored values

All concentration data are treated as intervals.

- •Let's say reported concentration is 1 mg/L
- •We code this as: ConcLow = 1.0 and ConcHigh = 1.0
- •The interval for this data point is then 1.0 to 1.0

- •For a value reporte as "less than 1.0 mg/L"
- •We code this as: ConcLow = NA and ConcHigh = 1.0
- •The interval for this data point is then 0.0 to 1.0

All of the "weighted regressions" in WRTDS are really "survival regression" (the function survreg in R) which is design for data reported as an interval.

USGS

Censored values and compound analytes

Sometimes an 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**

The two analyte values were reported as <0.2 and 0.9 mg/L respectively. Therefore, this data point has ConcLow = 0.9 and ConcHigh = 1.1.

- **The conventional left-censored approach calls this (0,1.1)**
- **WRTDS calls this (0.9 to 1.1)**

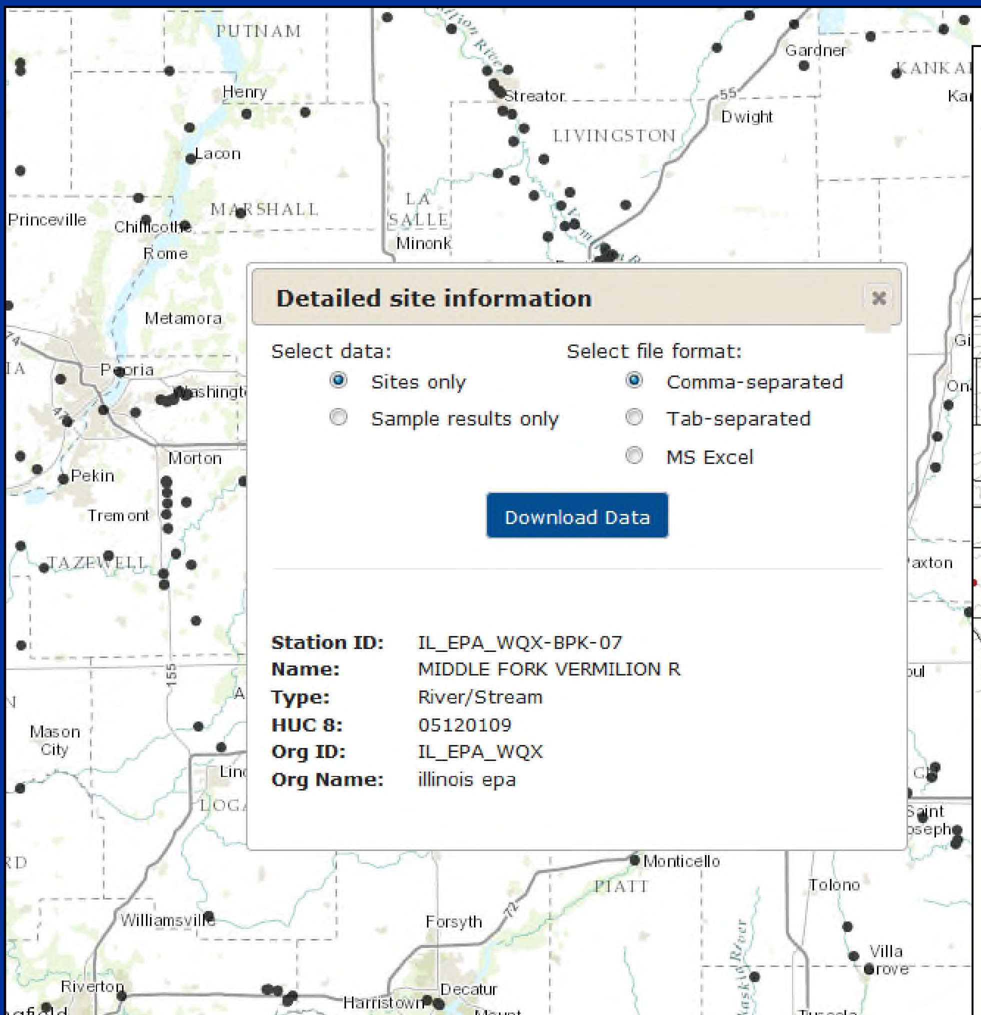
USGS

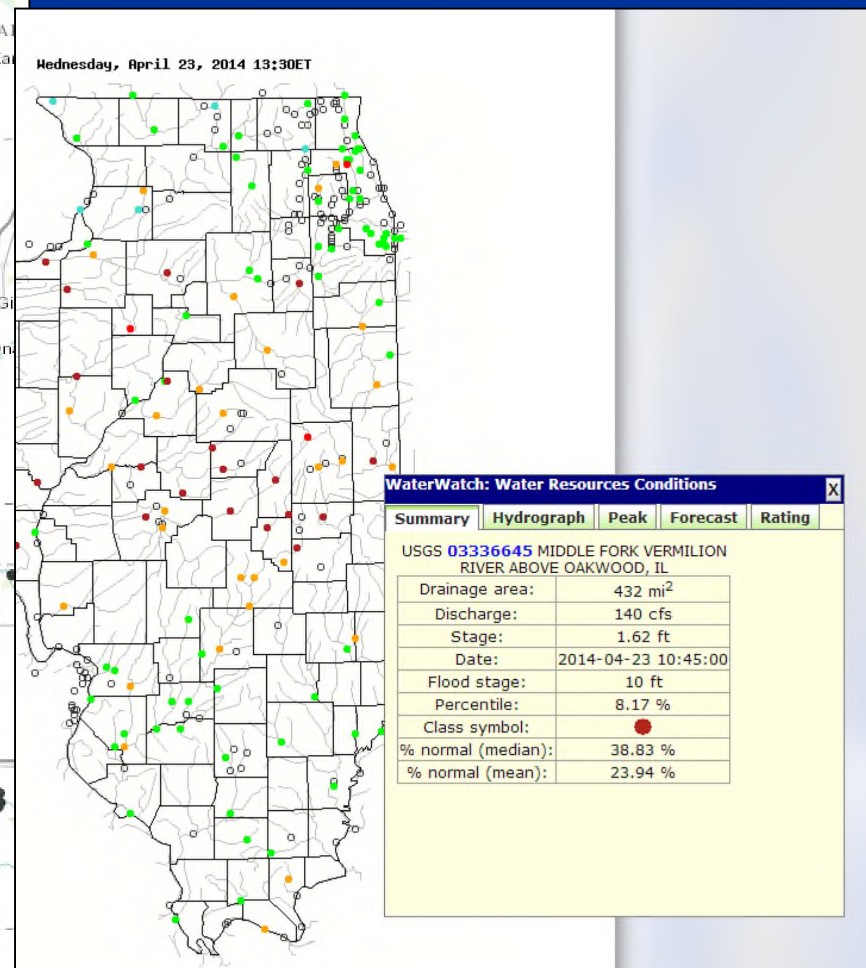
For Non-USGS data from the Water Quality Portal

> siteID <- “IL_EPA_WQX-BPK-07”
> characteristicName <- “Inorganic nitrogen (nitrate and nitrite)”
> startDate <- “2005-01-01”
> endDate <- “2012-01-01”
> Sample <- getSTORETSampleData(siteID,characteristicName,startDate,endDate)
➢summary(Sample)

```
>
>
> siteID <- "IL_EPA_WQX-BPK-07"
> characteristicName <- "Inorganic nitrogen (nitrate and nitrite)"
> startDate <- "2005-01-01"
> endDate <- "2012-01-01"
> Sample <- suppressWarnings(getSTORETSampleData(siteID,characteristicName,startDate,endDate))
> summary(Sample)
      Date               ConcLow          ConcHigh            Uncen         ConcAve             Julian
 Min.   :2005-01-24   Min.   : 0.041   Min.   : 0.0180   Min.   :0.0   Min.   : 0.0090   Min.   :56636
 1st Qu.:2009-02-08   1st Qu.: 3.658   1st Qu.: 0.1905   1st Qu.:1.0   1st Qu.: 0.1905   1st Qu.:58112
 Median :2010-01-07   Median : 5.205   Median : 4.5950   Median :1.0   Median : 4.5950   Median :58446
 Mean   :2009-05-21   Mean   : 4.834   Mean   : 3.8710   Mean   :0.8   Mean   : 3.8692   Mean   :58215
 3rd Qu.:2011-03-03   3rd Qu.: 6.560   3rd Qu.: 6.2250   3rd Qu.:1.0   3rd Qu.: 6.2250   3rd Qu.:58866
 Max.   :2011-11-28   Max.   :11.400   Max.   :11.4000   Max.   :1.0   Max.   :11.4000   Max.   :59135
                      NA's   :8
     Month            Day            DecYear        MonthSeq         SinDY               CosDY
 Min.   : 1.000   Min.   : 10.0   Min.   :2005   Min.   :1861   Min.   :-0.997643   Min.   :-0.99908
 1st Qu.: 4.000   1st Qu.: 96.5   1st Qu.:2009   1st Qu.:1910   1st Qu.:-0.742051   1st Qu.:-0.69469
 Median : 6.500   Median :184.0   Median :2010   Median :1921   Median :-0.008572   Median :-0.16188
 Mean   : 6.425   Mean   :179.5   Mean   :2009   Mean   :1913   Mean   :-0.011740   Mean   :-0.07631
 3rd Qu.: 9.000   3rd Qu.:256.2   3rd Qu.:2011   3rd Qu.:1934   3rd Qu.: 0.739359   3rd Qu.: 0.62395
 Max.   :12.000   Max.   :349.0   Max.   :2012   Max.   :1943   Max.   : 0.998674   Max.   : 0.98673

> length(Sample$Date)
[1] 40
>
>
```

Find a nearby USGS streamgage that can provide appropriate discharge data to use with these water quality data



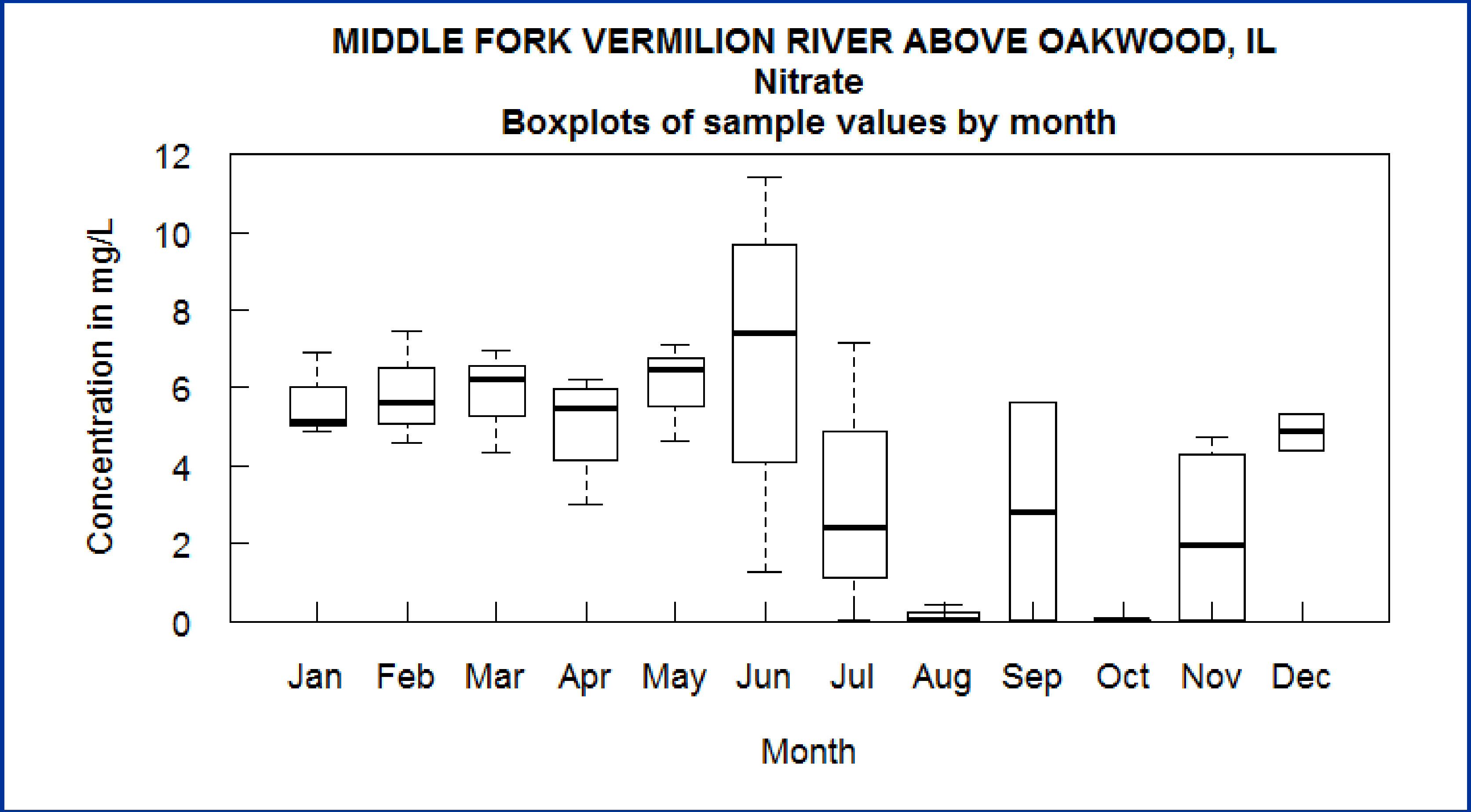


Retrieve daily discharge data

> siteID <- “03336645”
> Daily <- getDVData(siteID,"00060",startDate,endDate)

```
>
> siteID <- "03336645"
> Daily <- getDVData(siteID,"00060",startDate,endDate)
There are 2557 data points, and 2557 days.
> summary(Daily)
      Date                  Q                  Julian           Month             Day            DecYear        MonthSeq
 Min.   :2005-01-01   Min.   :  0.1727   Min.   :56613   Min.   : 1.000   Min.   :  1.0   Min.   :2005   Min.   :1861
 1st Qu.:2006-10-02   1st Qu.:  1.4725   1st Qu.:57252   1st Qu.: 4.000   1st Qu.: 93.0   1st Qu.:2007   1st Qu.:1882
 Median :2008-07-02   Median :  4.6723   Median :57891   Median : 7.000   Median :184.0   Median :2009   Median :1903
 Mean   :2008-07-02   Mean   : 12.2350   Mean   :57891   Mean   : 6.522   Mean   :183.7   Mean   :2009   Mean   :1903
 3rd Qu.:2010-04-02   3rd Qu.: 12.4877   3rd Qu.:58530   3rd Qu.:10.000   3rd Qu.:275.0   3rd Qu.:2010   3rd Qu.:1924
 Max.   :2012-01-01   Max.   :322.8120   Max.   :59169   Max.   :12.000   Max.   :366.0   Max.   :2012   Max.   :1945

  Qualifier              i                LogQ               Q7                Q30
 Length:2557        Min.   :   1   Min.   :-1.7560   Min.   :  0.1905   Min.   : 0.2219
 Class :character   1st Qu.: 640   1st Qu.: 0.3869   1st Qu.:  1.6768   1st Qu.: 2.3293
 Mode  :character   Median :1279   Median : 1.5416   Median :  5.2548   Median : 7.6092
                    Mean   :1279   Mean   : 1.4538   Mean   : 12.2107   Mean   :11.9463
                    3rd Qu.:1918   3rd Qu.: 2.5247   3rd Qu.: 14.1463   3rd Qu.:19.9598
                    Max.   :2557   Max.   : 5.7771   Max.   :151.2524   Max.   :55.4217
                                                     NA's   :6          NA's   :29
> length(Daily$Q)
[1] 2557
>
>
>
```

Data can then be used for EGRET or other analysis   this plot > boxConcMonth( )



Unit values retrieval (not used by EGRET)

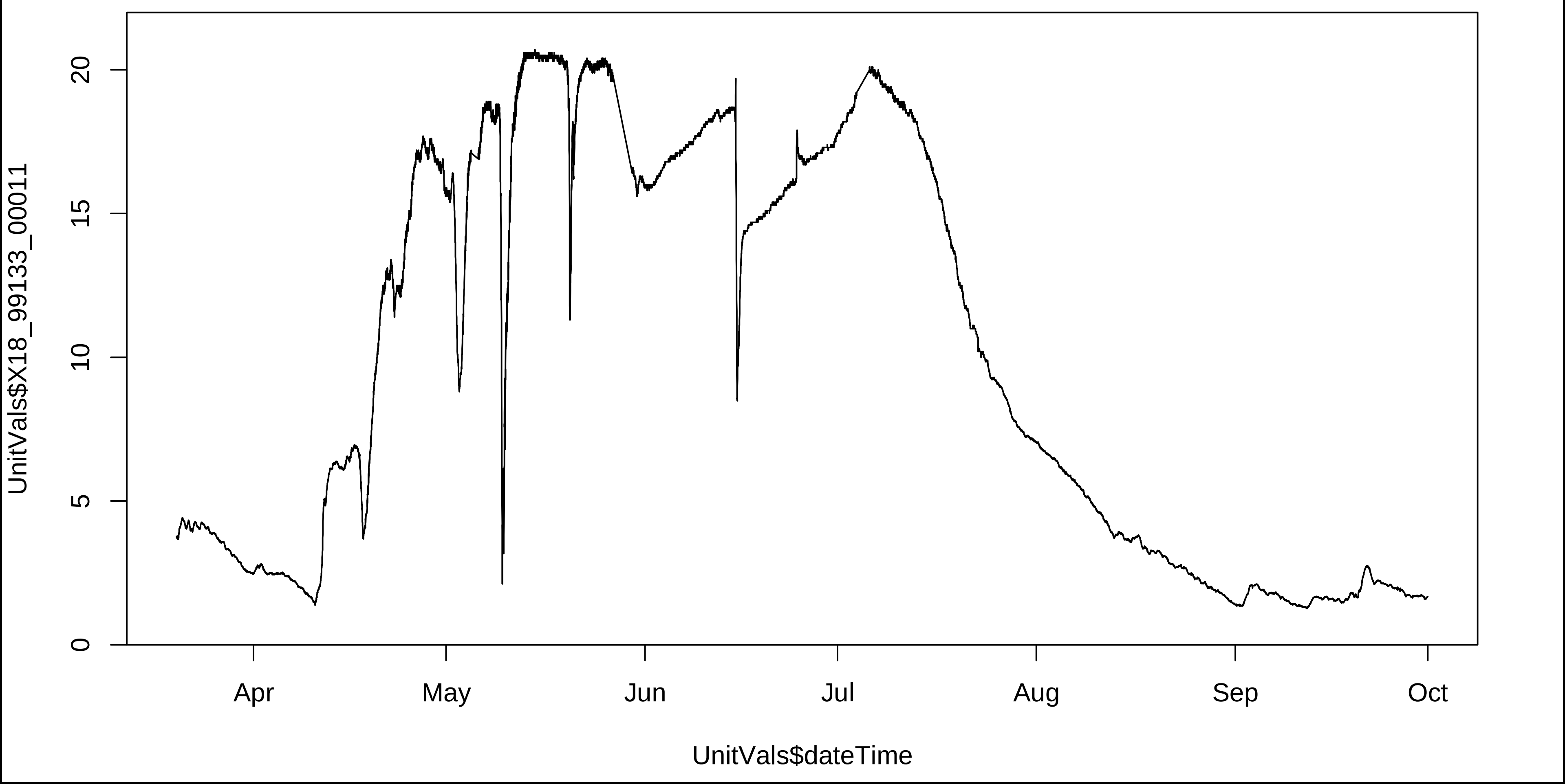
- Raccoon River at Van Metre, IA
- Nitrate sensor data
- March – Sept 2013

```
> siteID <- “05484500”
> parameterCD <- “99133”
> UnitVals <- retrieveUnitNWISData (siteID, parameterCD, “2013-03-01”, “2013-09-30”)
> Summary(UnitVals)
```

USGS

```
> summary(UnitVals)
   agency              site               dateTime                    tz_cd        X18_99133_00011  X18_99133_00011_cd
 Length:17720       Length:17720       Min.   :2013-03-20 00:00:00   CST:17720   Min.   : 1.260   A:17720
 Class :character   Class :character   1st Qu.:2013-05-08 16:11:15               1st Qu.: 2.540
 Mode  :character   Mode  :character   Median :2013-06-27 23:52:30               Median : 7.250
                                       Mean   :2013-06-26 23:57:31               Mean   : 9.651
                                       3rd Qu.:2013-08-15 15:33:45               3rd Qu.:16.900
                                       Max.   :2013-09-30 23:45:00               Max.   :20.700
```

```
> plot(UnitVals$dateTime,UnitVals$X18_99133_00011,type="l",ylim=c(0,22),yaxs="i")
```



Back to our main EGRET example, metadata is next

- The getMetaData function is used to retrieve, input, and save metadata about the current analysis
- It defines what data you have and where they came from
- Inputs are used for labels on figures & tables
- Inputs are used to name your saved workspaces, using abbreviations

> INFO<-getMetaData(siteID,parameter_cd)

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):

<cr>

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

USGS

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

It is helpful to set up a constiuent 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

USGS

If you are using supplied data, you still must run the getMetaData command:

```
> INFO <- getMetaData()
```

The program will then prompt you to enter metadata about your site and study.

All metadata is voluntary except the following required fields:

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

USGS

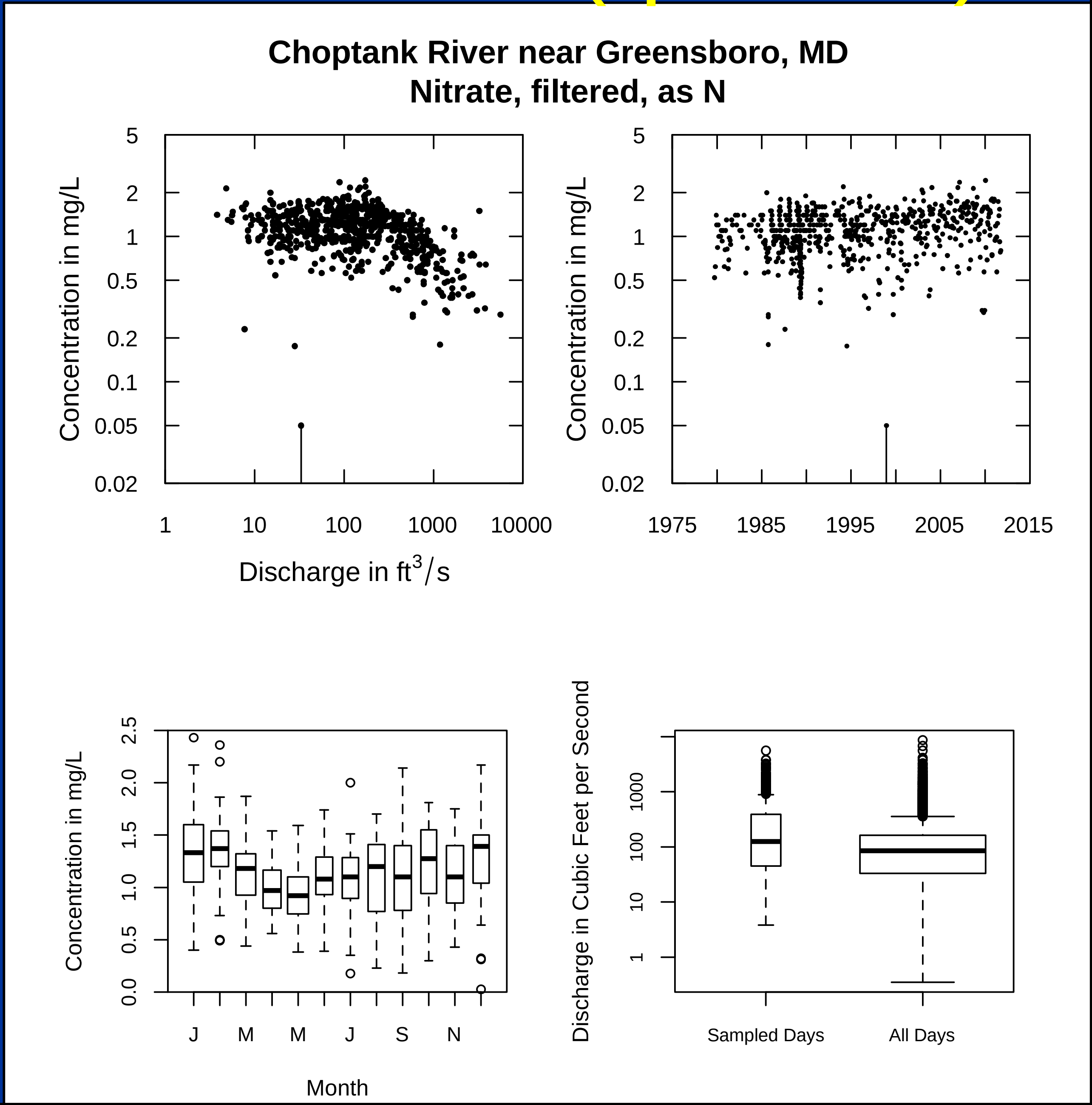
Two more commands before we can start our analysis of the data

> 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 654 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 %

> multiPlotDataOverview(qUnit=1)

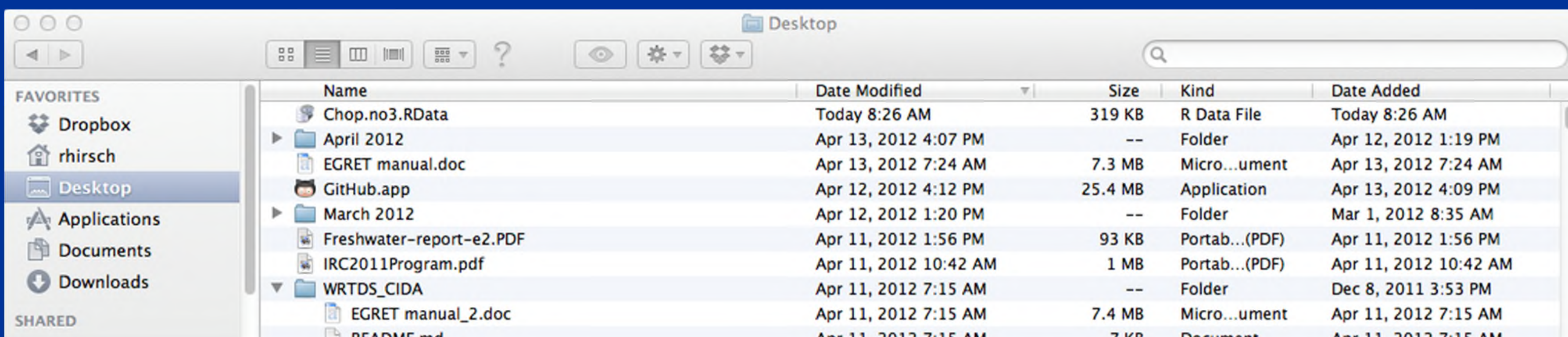
Let's look at the data before we proceed, the function is:



USGS

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

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



Save it over and over as you proceed and add results

We now have 3 data frames

- Sample (654 rows, 14 columns)
- Daily (11,718 rows, 12 columns)
- INFO (1 row, 42 columns)

> modelEstimation()

- Runs the model in cross-validation mode
- Estimates the “surface” for concentration as a function of time and discharge
- Uses the surface to compute daily values of
  - Concentration
  - Flux
  - Flow-normalized concentration
  - Flow-normalized flux
- Adds those to the Daily data frame

USGS

User has choices about some of the parameters of the WRTDS model

Now what is in Daily?
It is a data frame that has dimension (11718, 19)

```
> 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
```

"Period of Analysis" concept in EGRET.

- 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 Period
paLong = length of Period, in months

USGS

Period of analysis set up

Say we want calendar year

```
INFO <- setPA(paStart = 1, paLong=12)
```

Say we want April, May, June

```
INFO <- setPA(paStart = 4, paLong = 3)
```

Default is water year

USGS

Units in EGRET

Everything stored as:

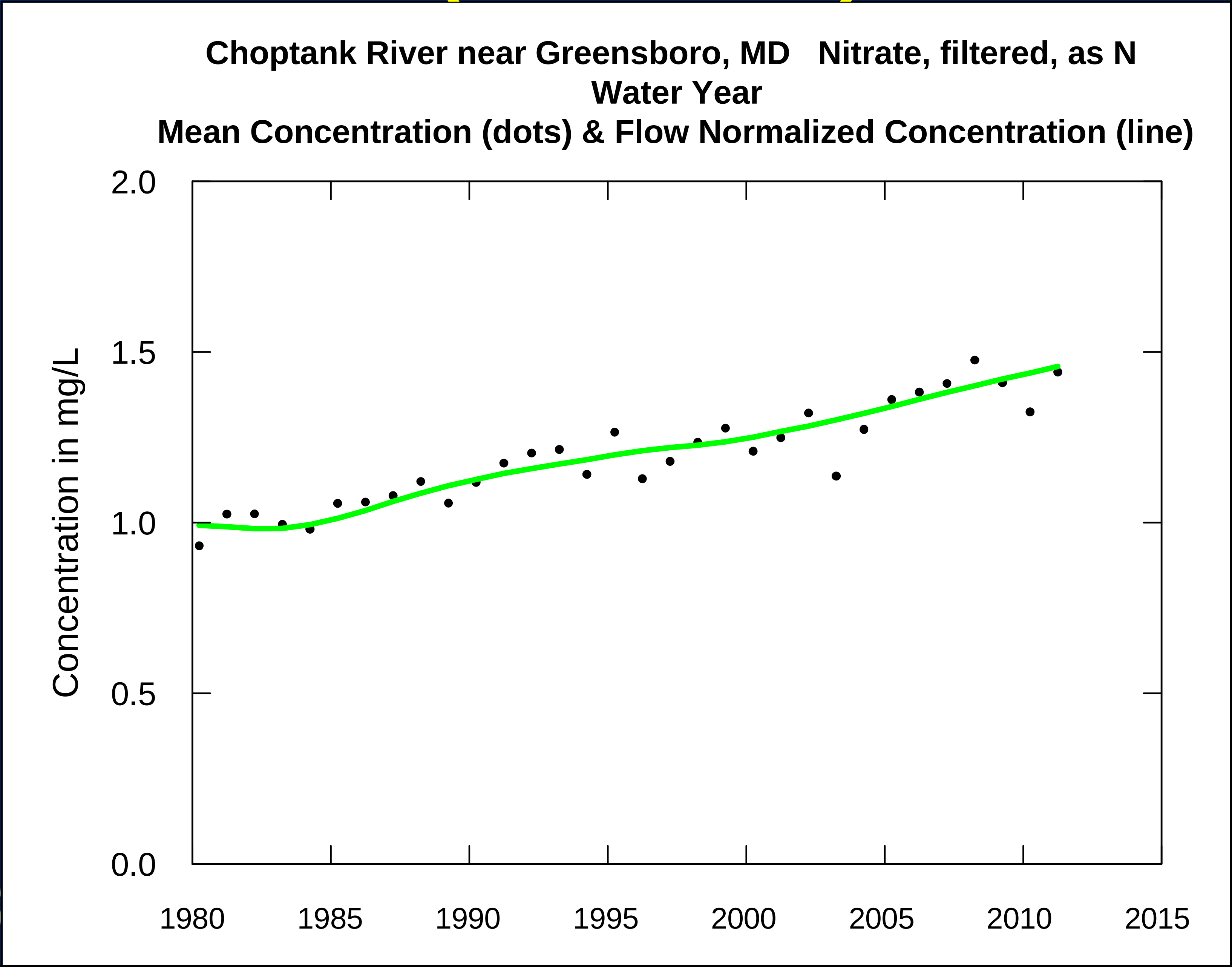
m^3/s, kg/day, or mg/L

But each graphic or table has a wide choice of units (English and SI) that the user can select

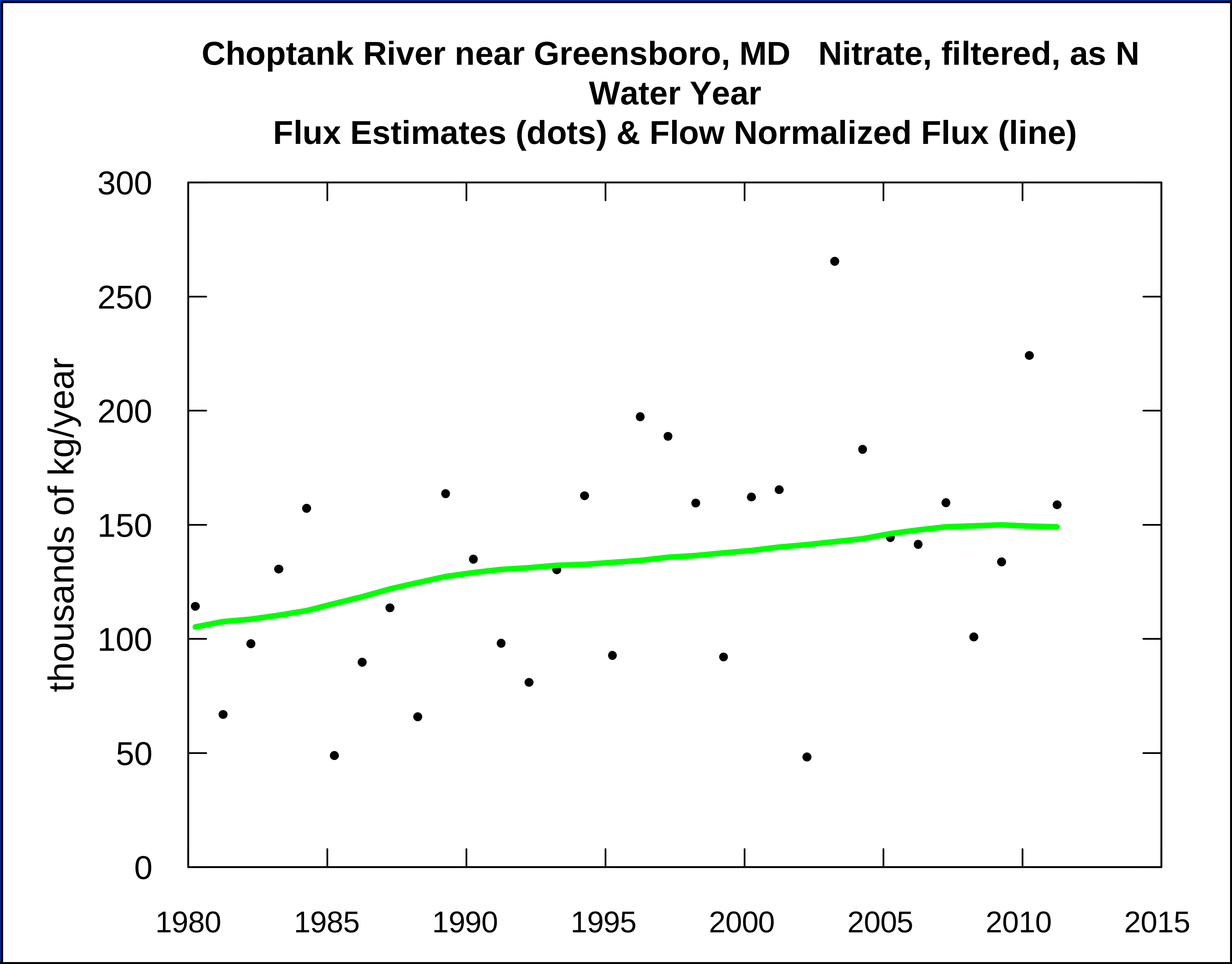
Now lets see some trend results

USGS

> plotConcHist(1980,2012)



> plotFluxHist(1980,2012,fluxUnit=8)



USGS

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

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

Year	Discharge cfs	Conc mg/L	FN_Conc mg/L	Flux tons/yr	FN_Flux tons/yr
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

USGS

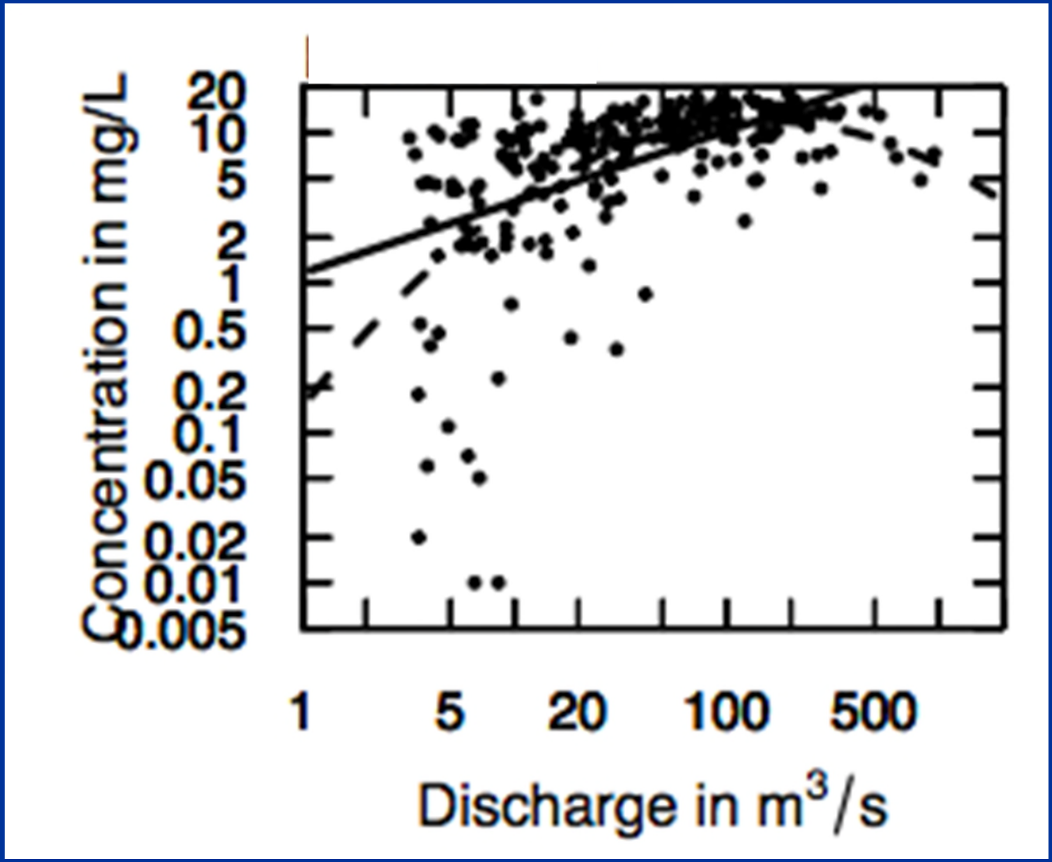
How to see trends as numbers

```
yearPoints = c(1980,1995,2011)

tableChange(fluxUnit=5,yearPoints)
```

```
> 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
```

USGS

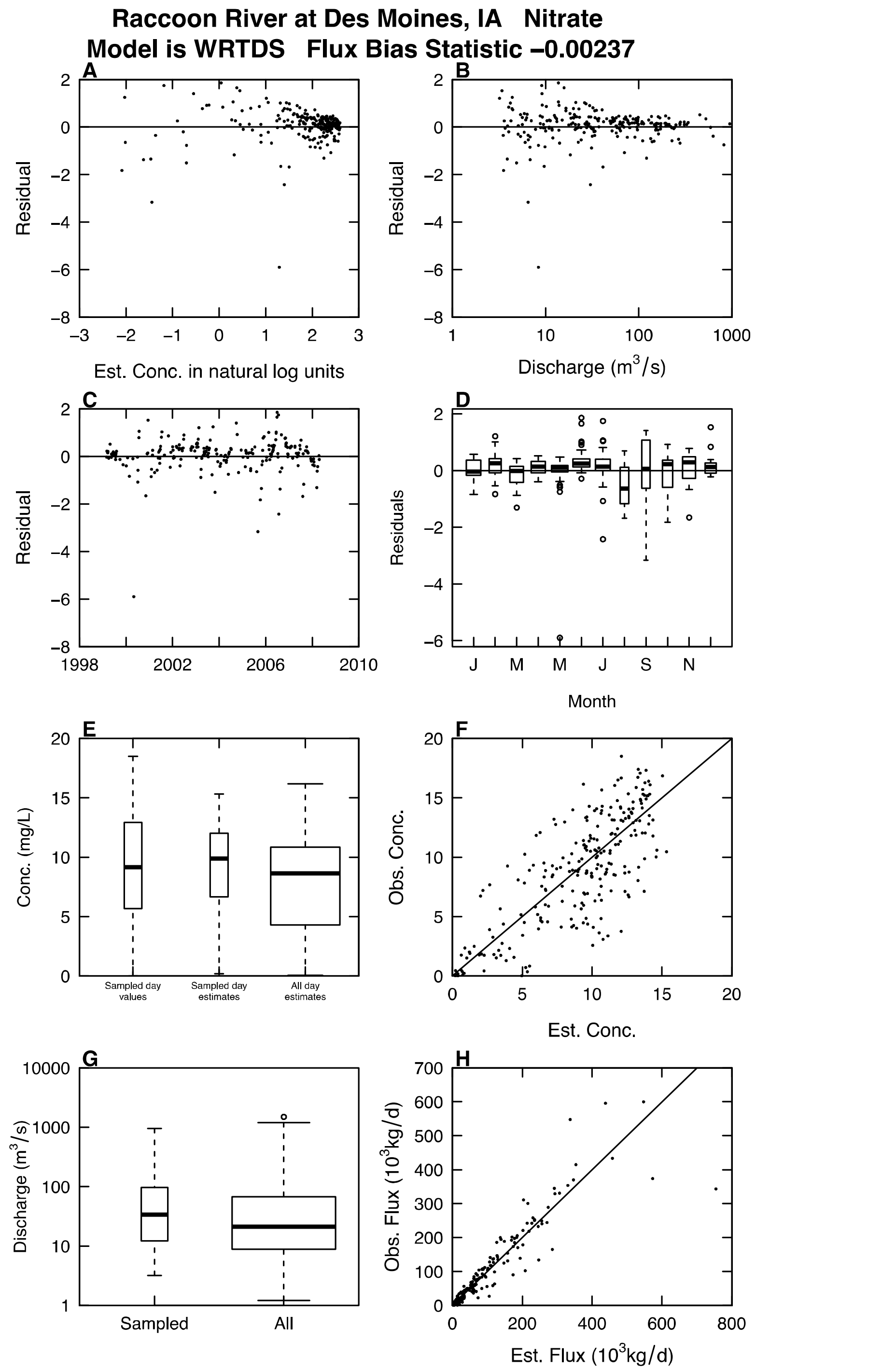
I'm going to switch data sets to Nitrate for the Raccoon River at Des Moines Iowa



USGS

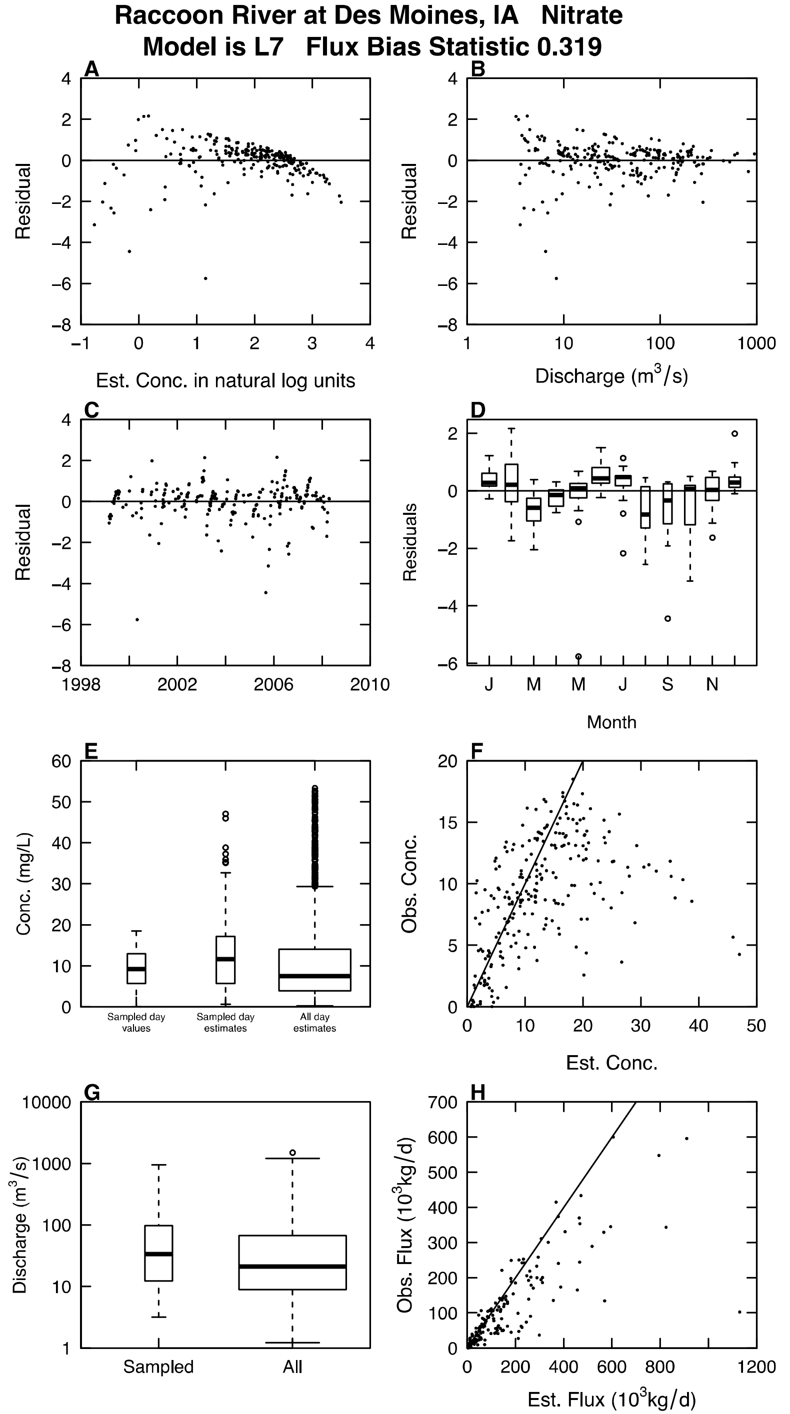
EGRET produces a diagnostic plot to help spot serious problems with the model

fluxBiasMulti(fluxUnit=4)



This same type of plot can be used to look at other models, here the LOADEST7

fluxBiasMulti(fluxUnit=4)



Diagnostics and potential problems with estimating mean flux, see:

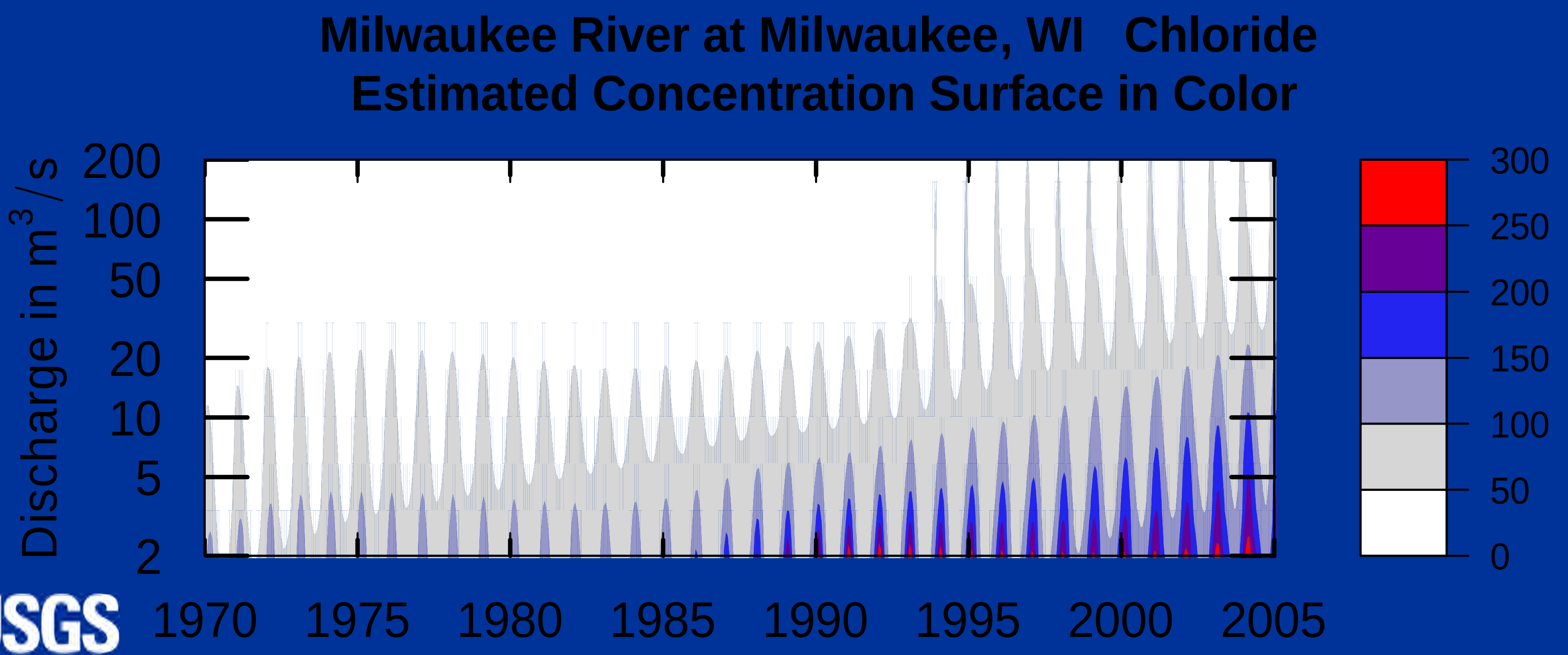
Hirsch, R.M., 2014, Large biases in regression-based constituent flux estimates: causes and diagnostics. Journal of the American Water Resources Association, in press.

Bottom line, look at the fit before you use a statistical model!!!

USGS

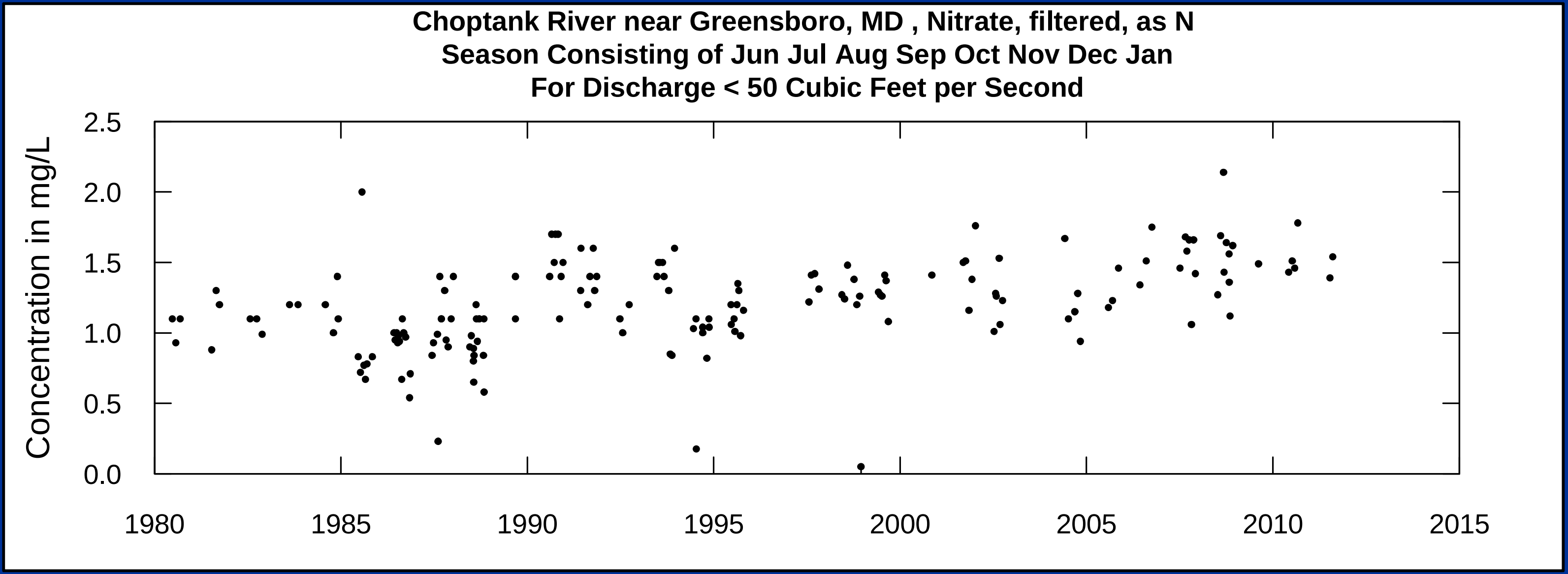
How difficult is it to make those contour plots?

```
>plotContours(yearStart=1970, yearEnd=2005,
qBottom=2, qTop=200, qUnit=2,
contourLevels=seq(0,300,50))
```

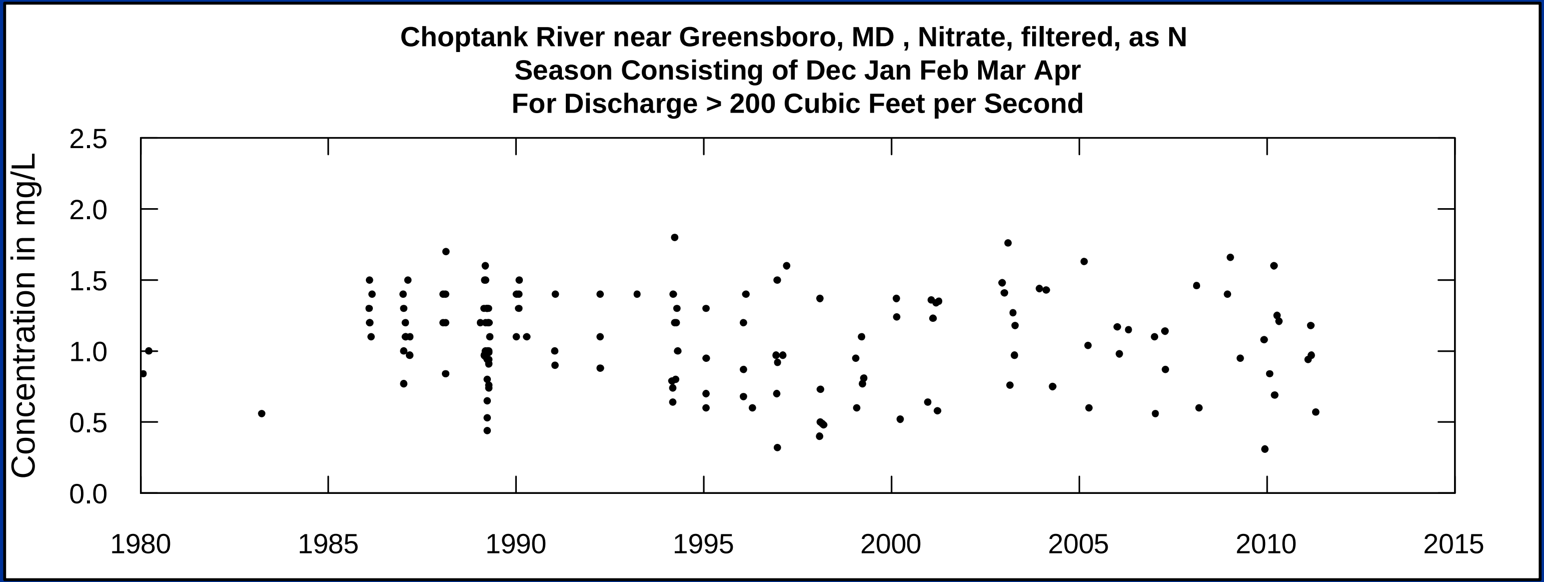


USGS

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



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



When all is said and done:

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.

dataDelivery and EGRET

- **Information and software available at: https://github.com/USGS-R/EGRET/wiki**

USGS