Overview of WRTDS and the EGRET and EGRETci Packages

WRTDS = Weighted Regressions on Time, Discharge and Season
EGRET = Exploration and Graphics for RivEr Trend
EGRETci = Confidence Intervals for EGRET

Robert M. Hirsch, Research Hydrologist
US Geological Survey
2017-04-06

Outline of the presentation

1. Motivations for the WRTDS method and the two related R-software packages
2. The WRTDS concept
3. How EGRET works
4. How EGRETci works

USGS

Motivation:
Quote 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

EGRET (Exploration and Graphics for RivEr Trends):

1) Obtain and organize: Sample data, daily discharge data, and meta-data

2) Use the WRTDS method to explore evolving water quality conditions

3) Explore streamflow trends

4) Produce graphs and tables

WRTDS overview

- Data sets: >100 water quality samples, daily discharge for every day
- Uses the sample values to build a statistical model of concentration for any combination of discharge, season, and year.
- Use model to compute mean values & trends.

USGS

Guiding ideas for WRTDS

- 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.
- Avoid a potential bias in flux estimates.
- Be quantitative but also exploratory.

USGS

Water quality data analysis issues

- Highly related to streamflow and season
- Highly skewed
- Sometimes censored

- Assessements of progress can be easily obscured by the random, but persistent, patterns of wet and dry years
- I call this: "The thrill of victory, the agony of defeat"

USGS

Analysis issues addressed by WRTDS

- Trends can be different across seasons
- Trends can be different across flows
- Trends shouldn't be restricted to be linear or monotonic
- We want a highly flexible model of how daily concentration varies as a function of time, discharge, and season

USGS

Approach

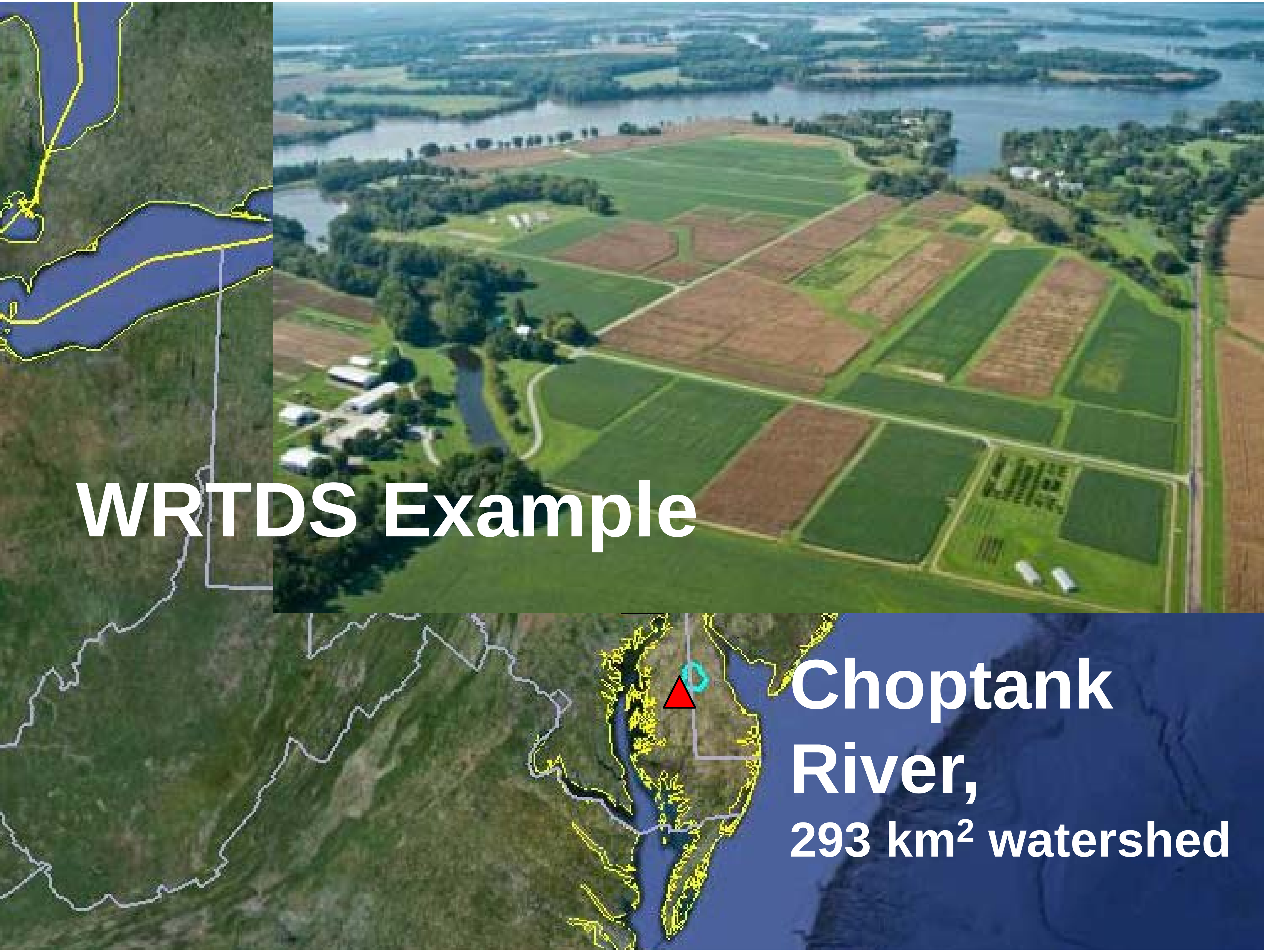
- Flexible statistical model to determine the expected value of concentration for any possible combination of date and discharge during the period of record.
- E[flux] = E[Concentration] * Discharge

USGS

Data requirements

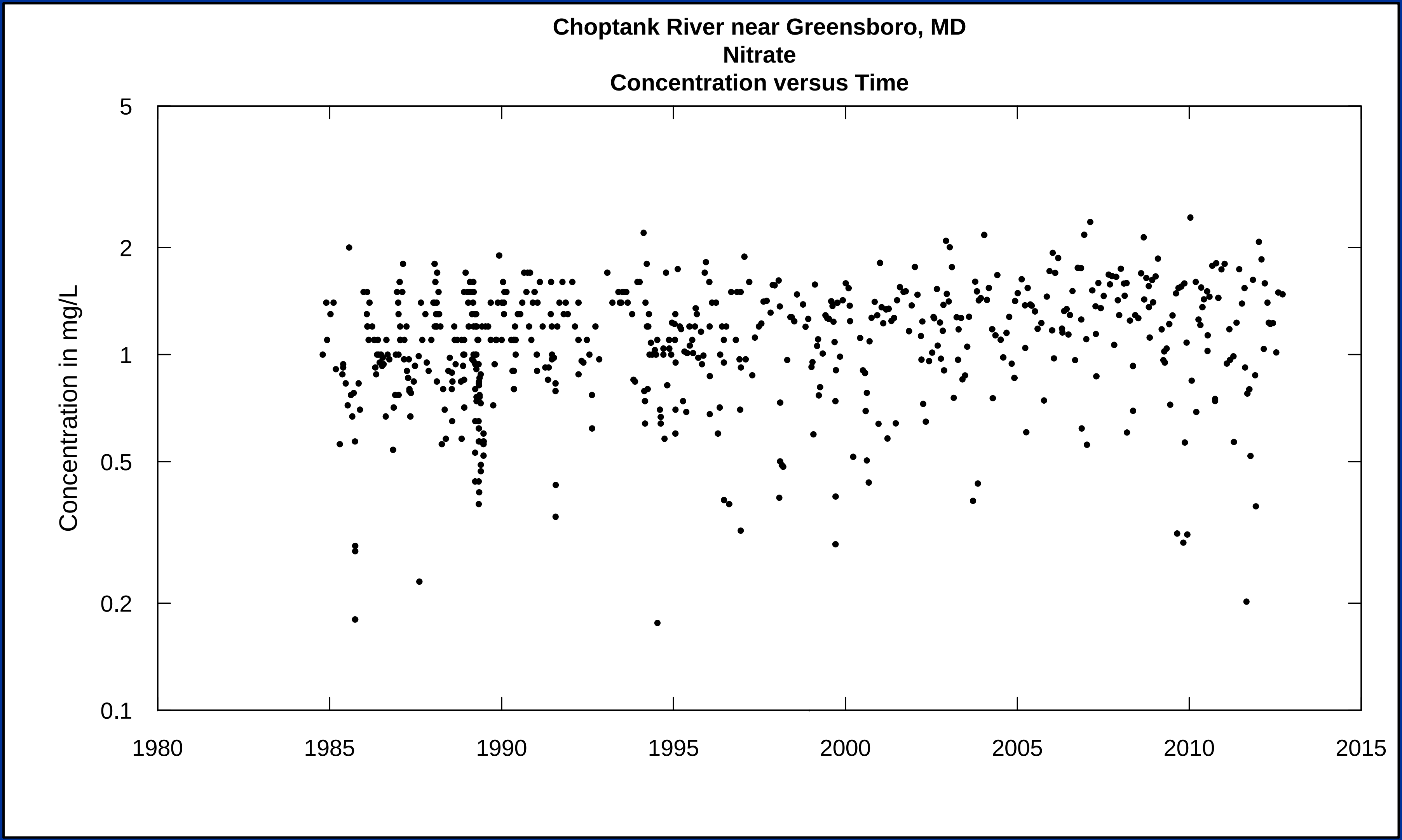
- Requires a complete daily discharge record
- Streamflow can’t be too flashy (low intra-day variability)
- Works best with >100 samples
- Water quality samples should cover most of the discharge range
- For trends: 10 or more years of data
- For average flux: 5 or more years of data

USGS


WRTDS Example
Choptank River,
293 km^2 watershed

"Data without models are chaos, but models without data are fantasy"

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



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

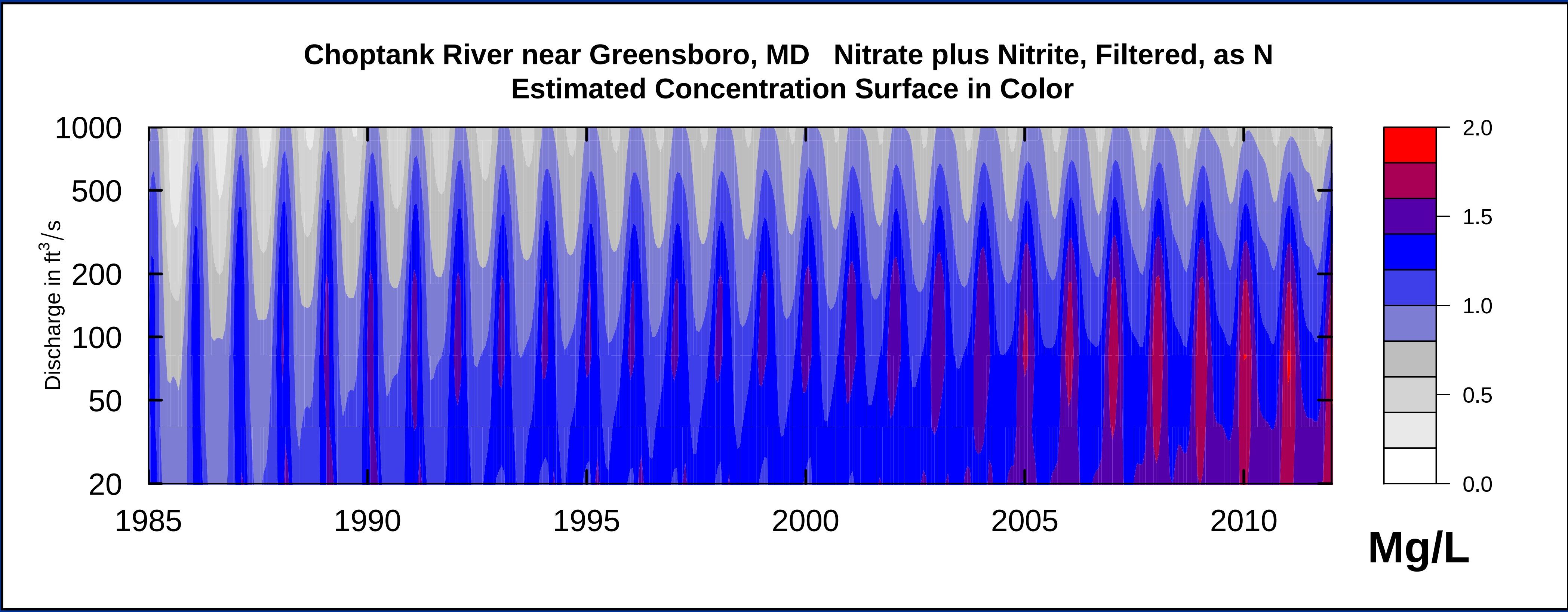
For any location in time - discharge space (*t* and *Q*) we assume that concentration (*c*) follows this model

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

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

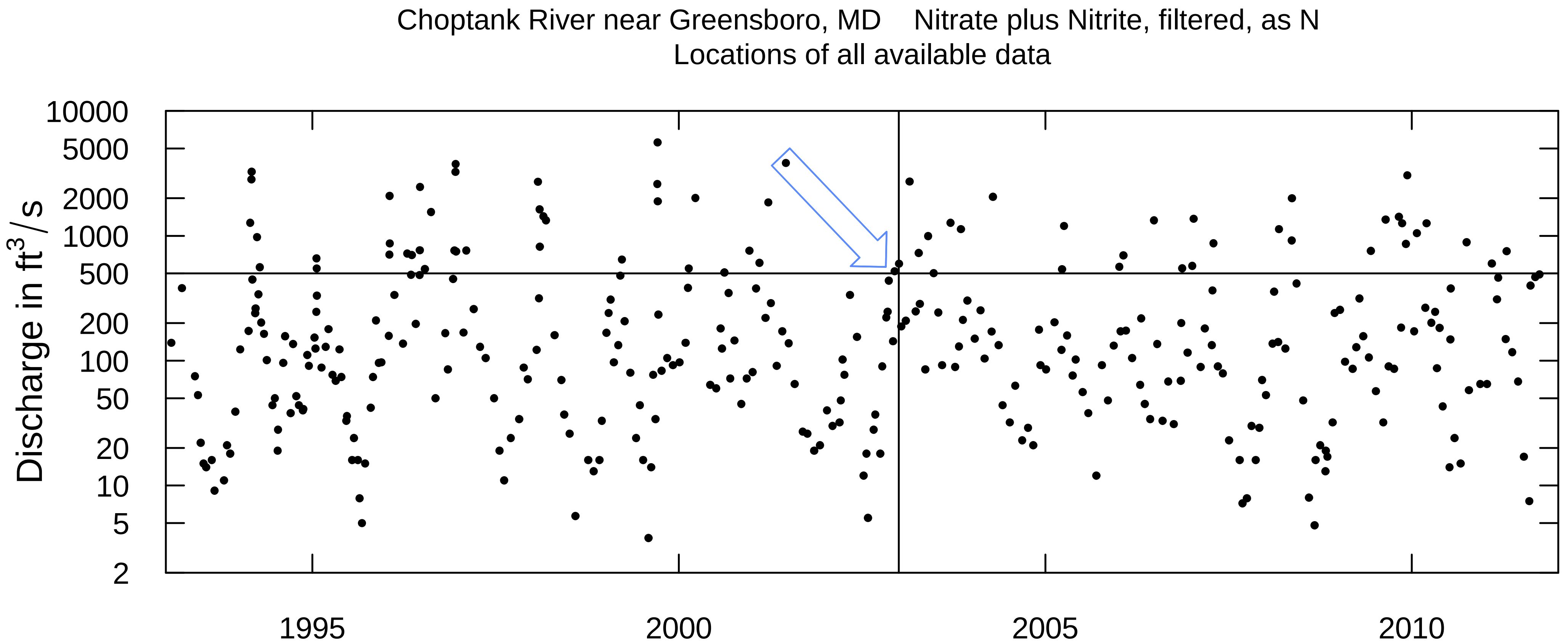


How is this surface created?

USGS

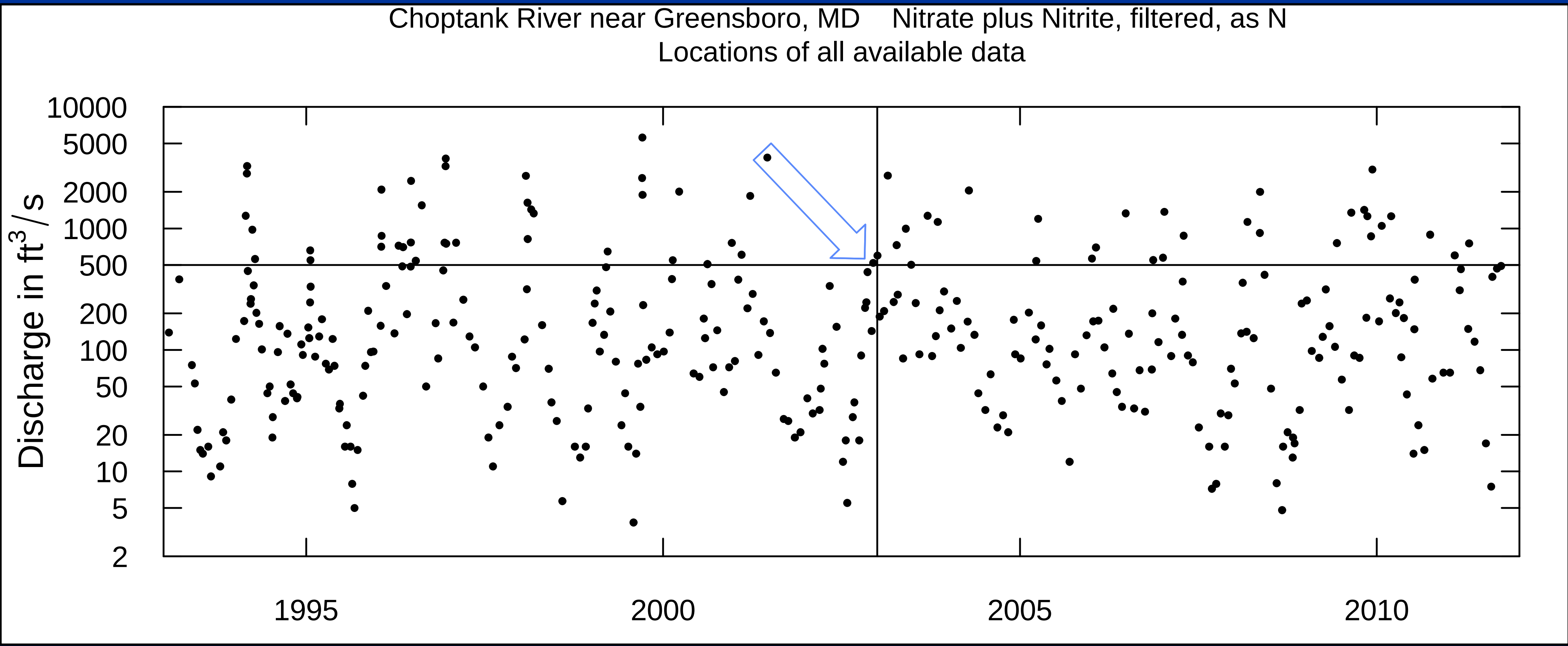
Every dot is a data point from 1993 to 2012

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



The principle is this:

Do a weighted regression at this point. The weights on each observation are related to their “distance” from Jan 1 2003 at 500 cfs



Weights based on distance in time, in log(Q), and season.

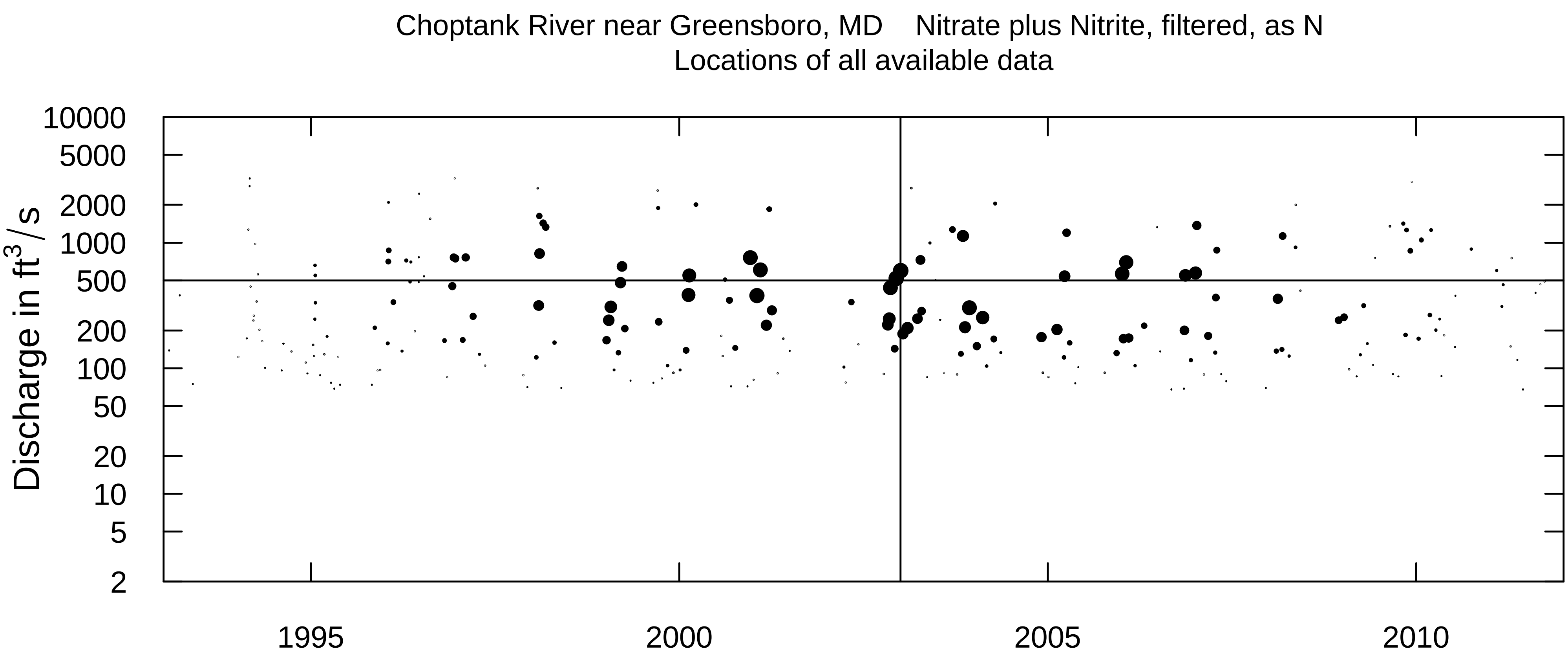
Weights come from tri-cubed weight function.

The weight for each observation is the product of these three weights.

All of this just to estimate the surface at this one point!

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

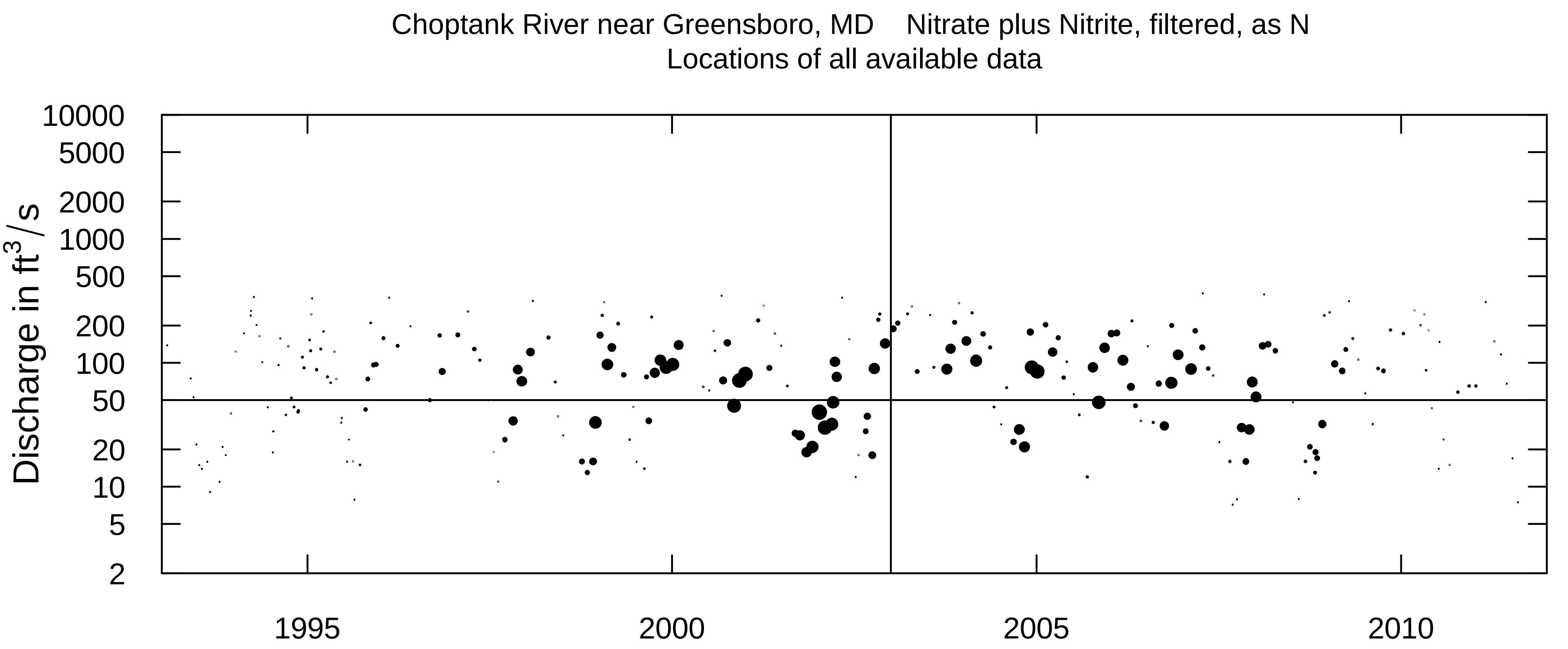


USGS

Now, on to another point in the space.

For example an estimate for January 1, 2003 but for Q = 50 cfs

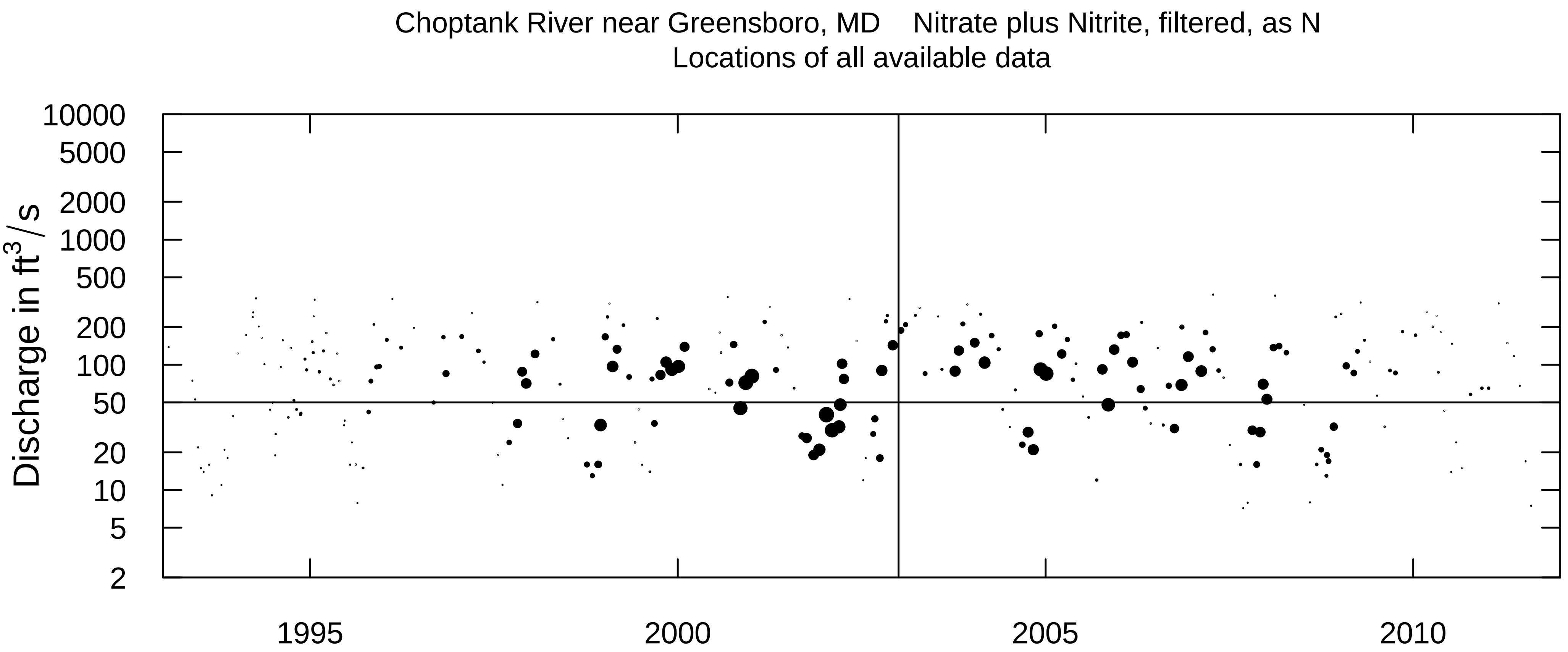
Redo the weights based on distance from that point and redo the regression



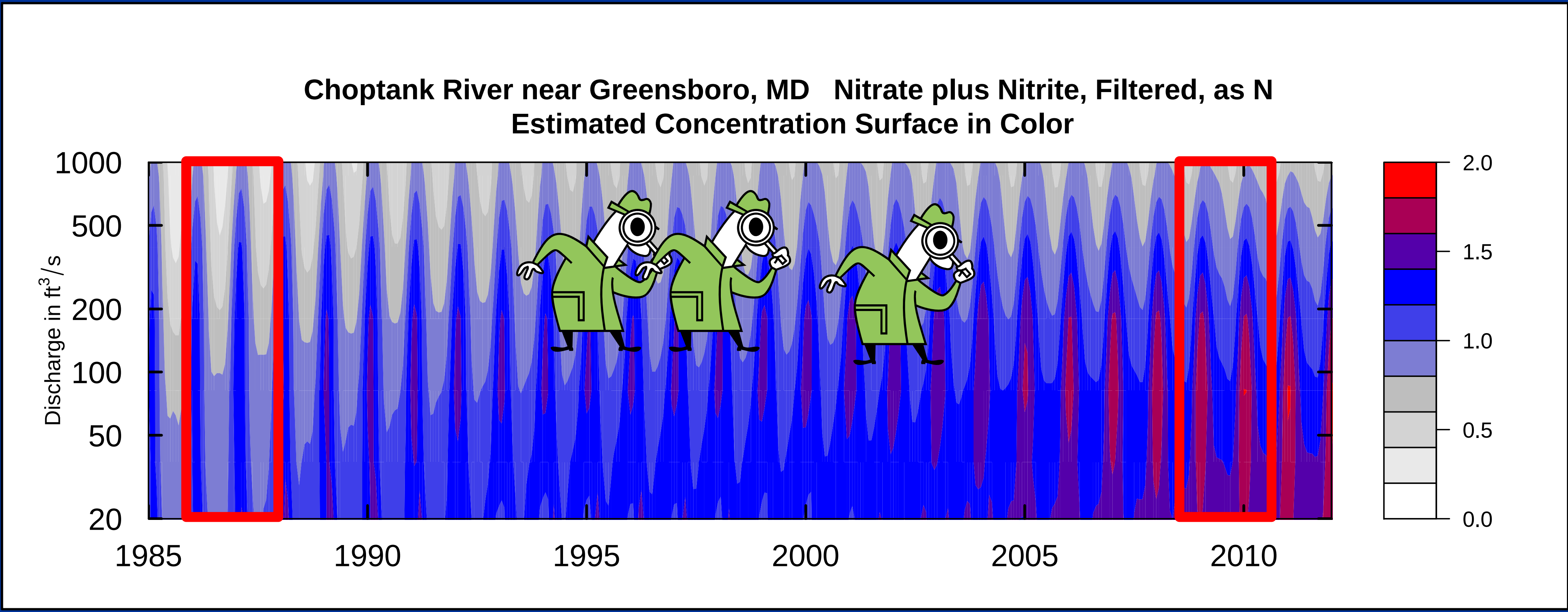
USGS

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

14 Q values x 16 times per year for the period of record

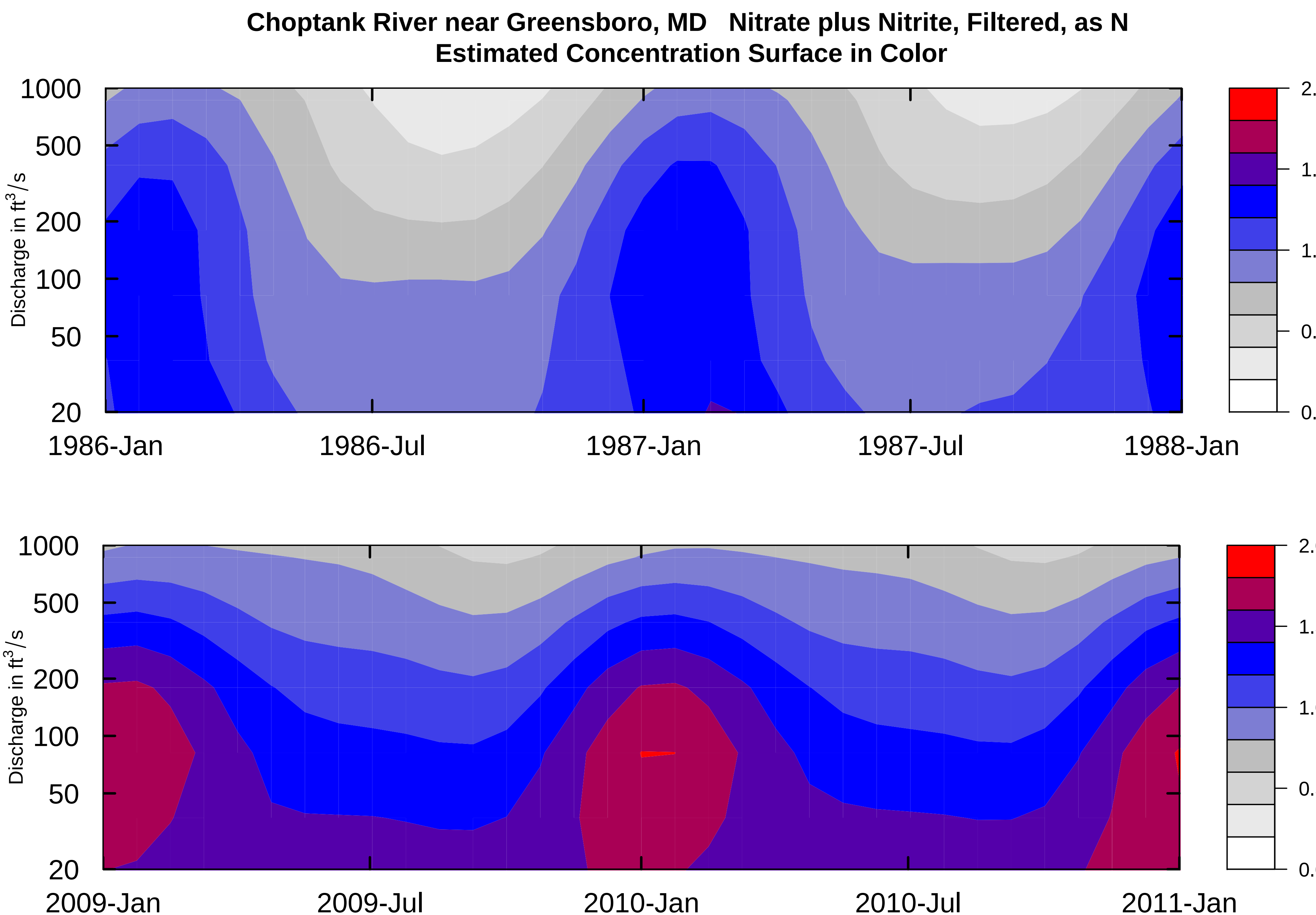


This kind of weighted regression gets done about 6000 times to form this whole surface!!



You must be kidding. This is a ton of computations!!
That's right! But it's what we need to make order out of chaos.

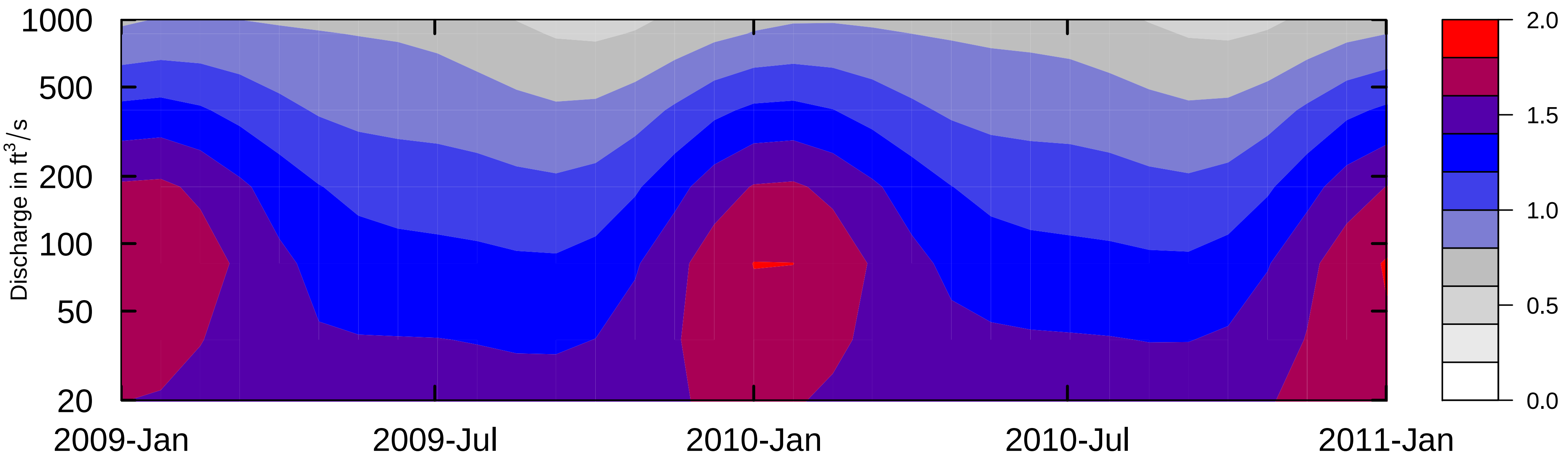
Here are two, more detailed looks at this surface



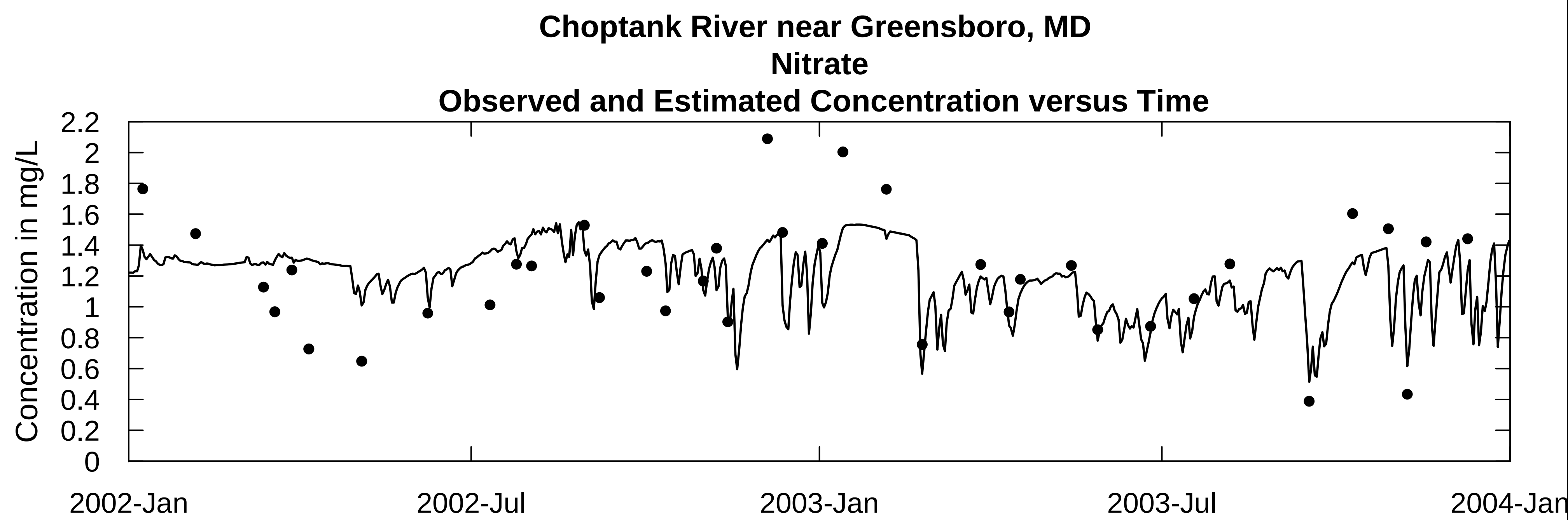
Now, for every one of 10,227 days in the record from 1985 through 2012:

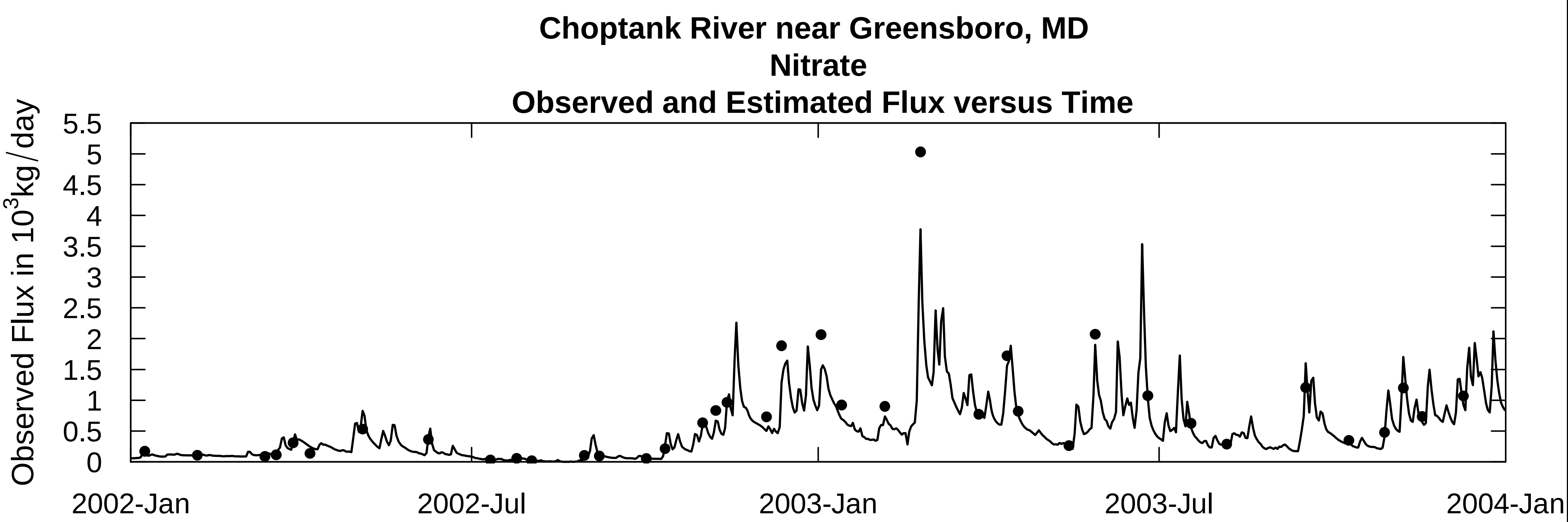
We can use the date and the observed discharge to compute the expected value of concentration.

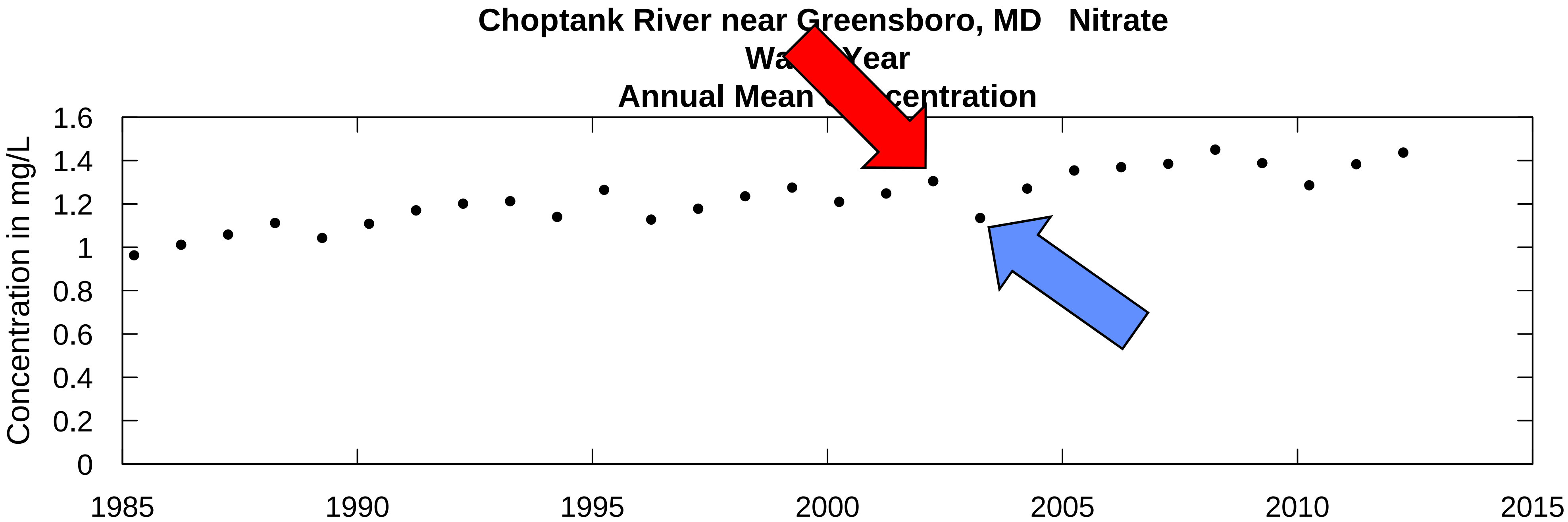
From that value we can compute the expected value of flux.

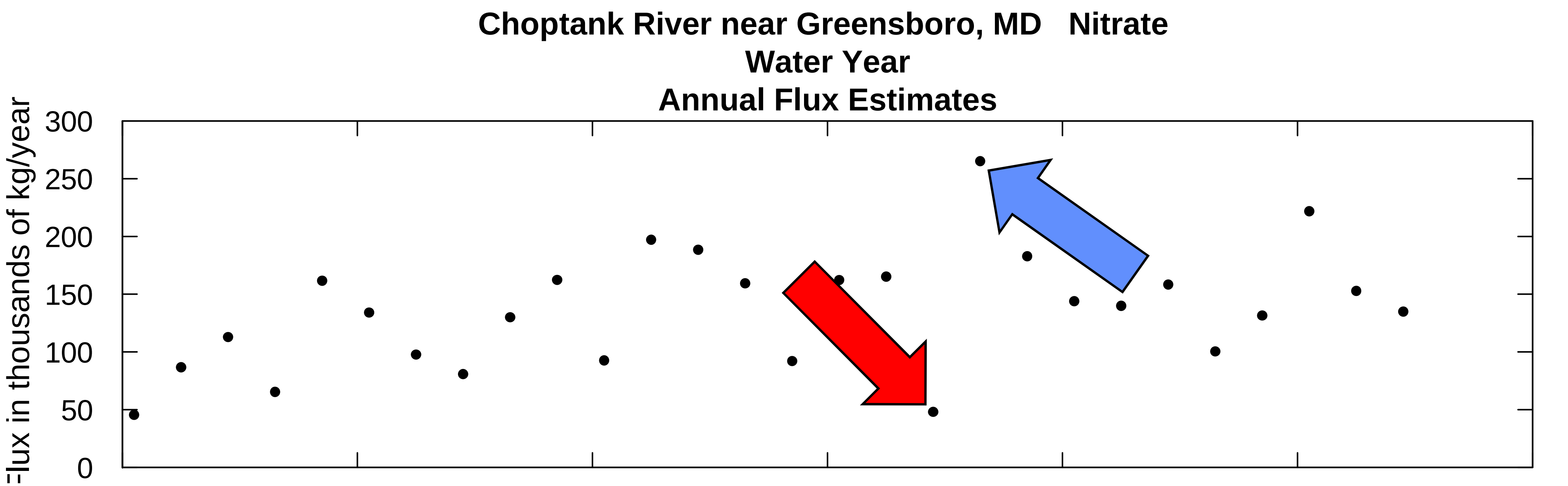


Then we can sum these estimates by year to compute estimates of annual mean concentration & annual mean flux

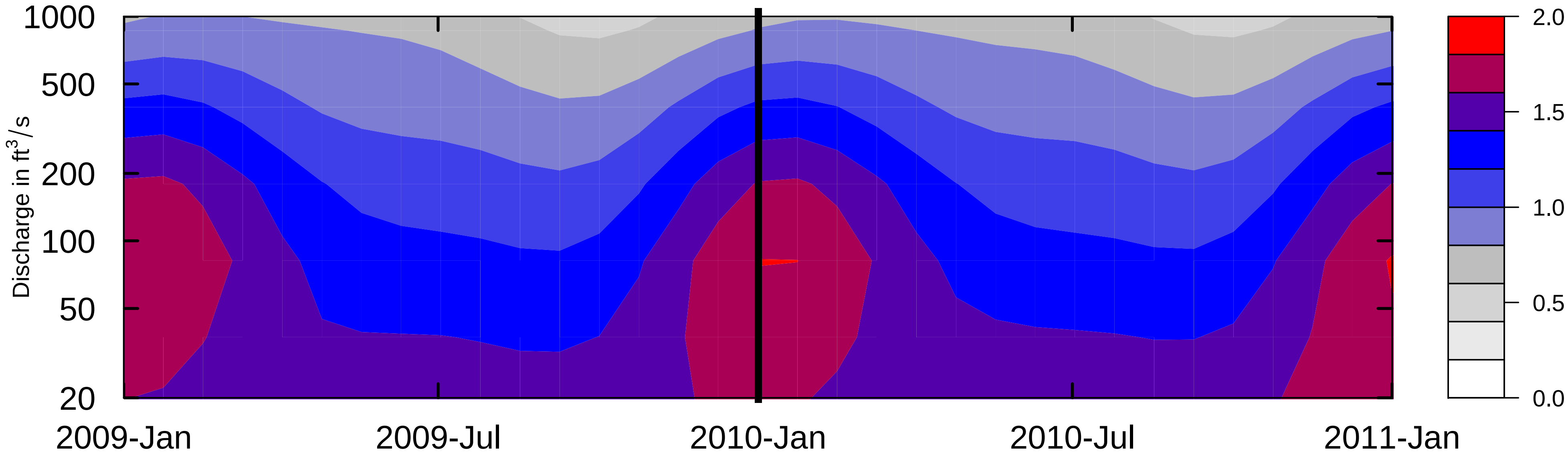








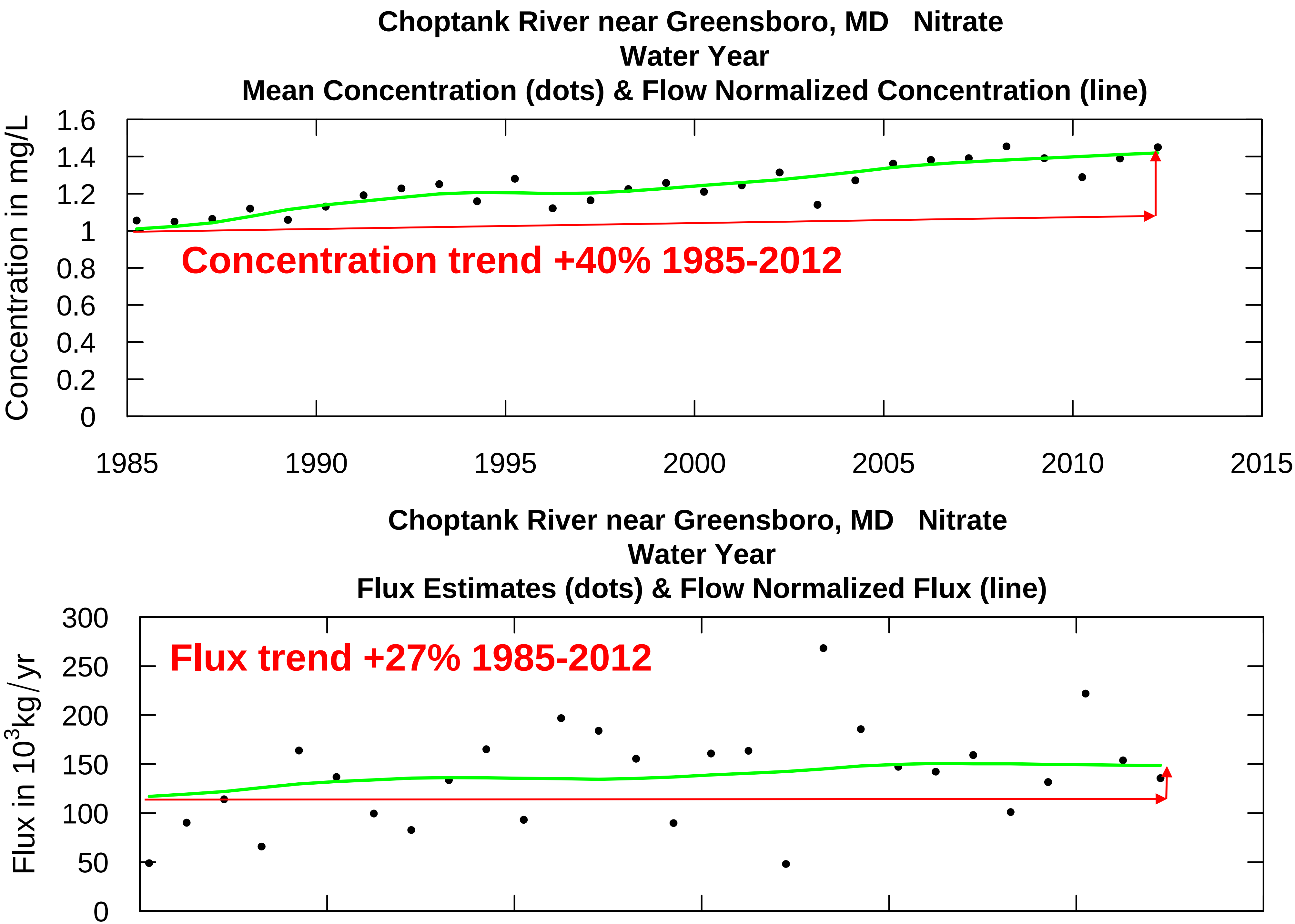
Can we filter out this flow-driven variation to see the underlying change?



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.

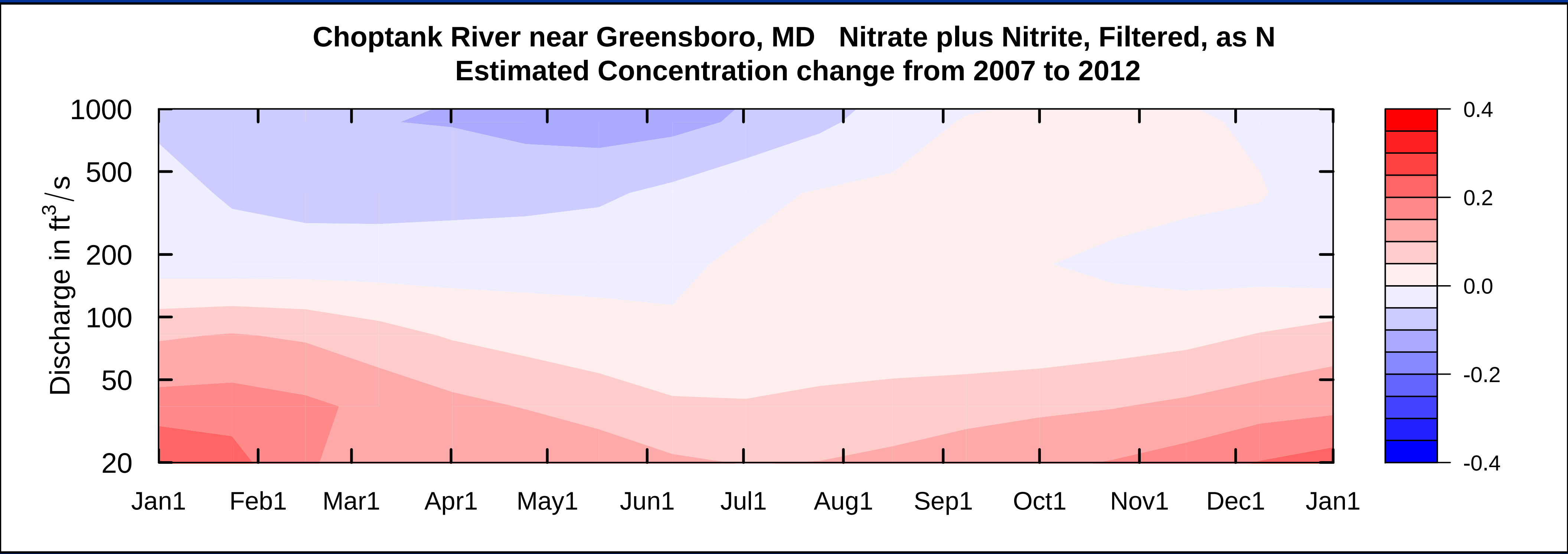
Sum those over the year to get annual flow-normalized mean concentration and flux.



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 +40% 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 10^3kg/yr
300
250
200
150
100
50
0
Flux trend +27% 1985-2012

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.

USGS

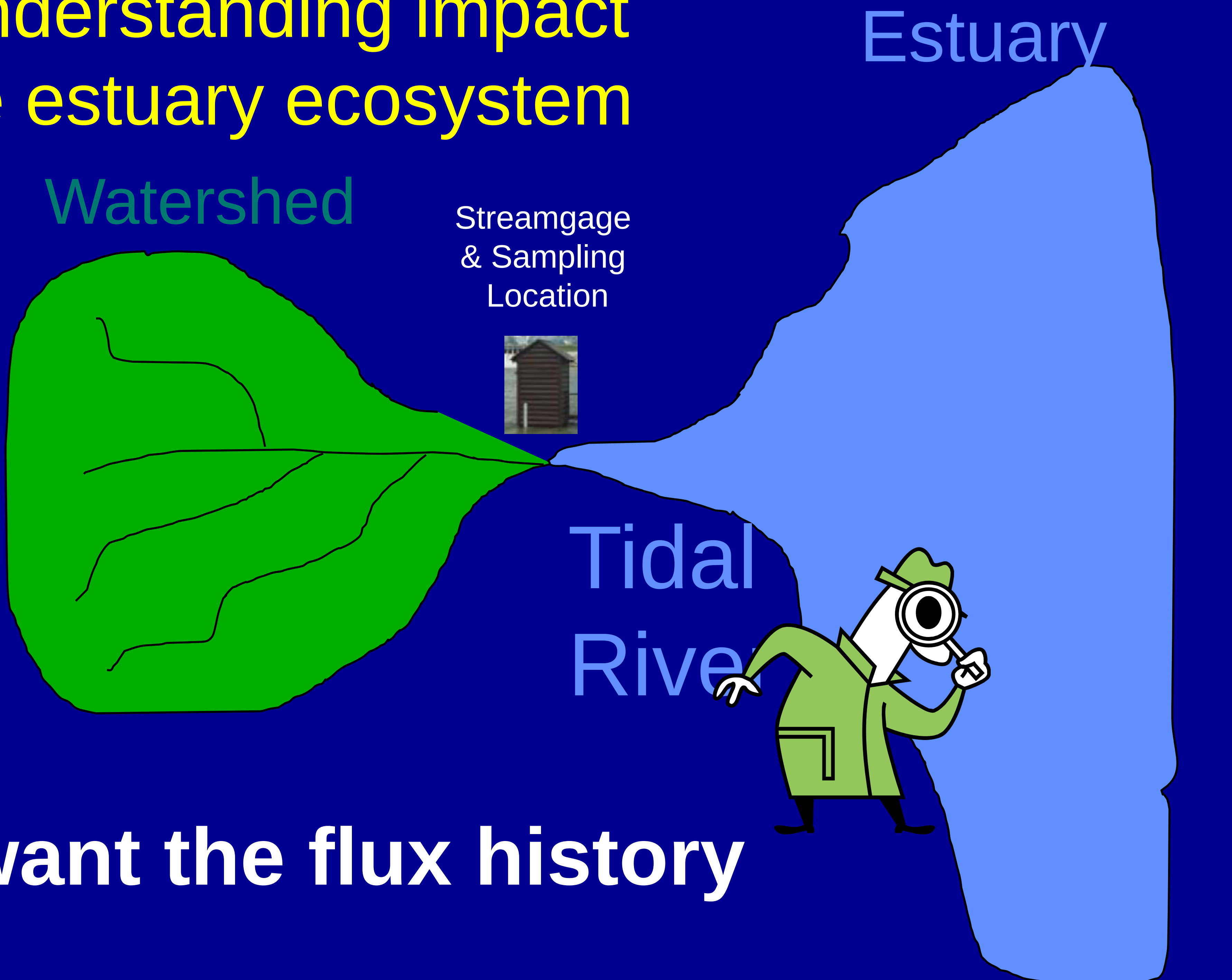
Why all this complexity?

Different products for different purposes

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

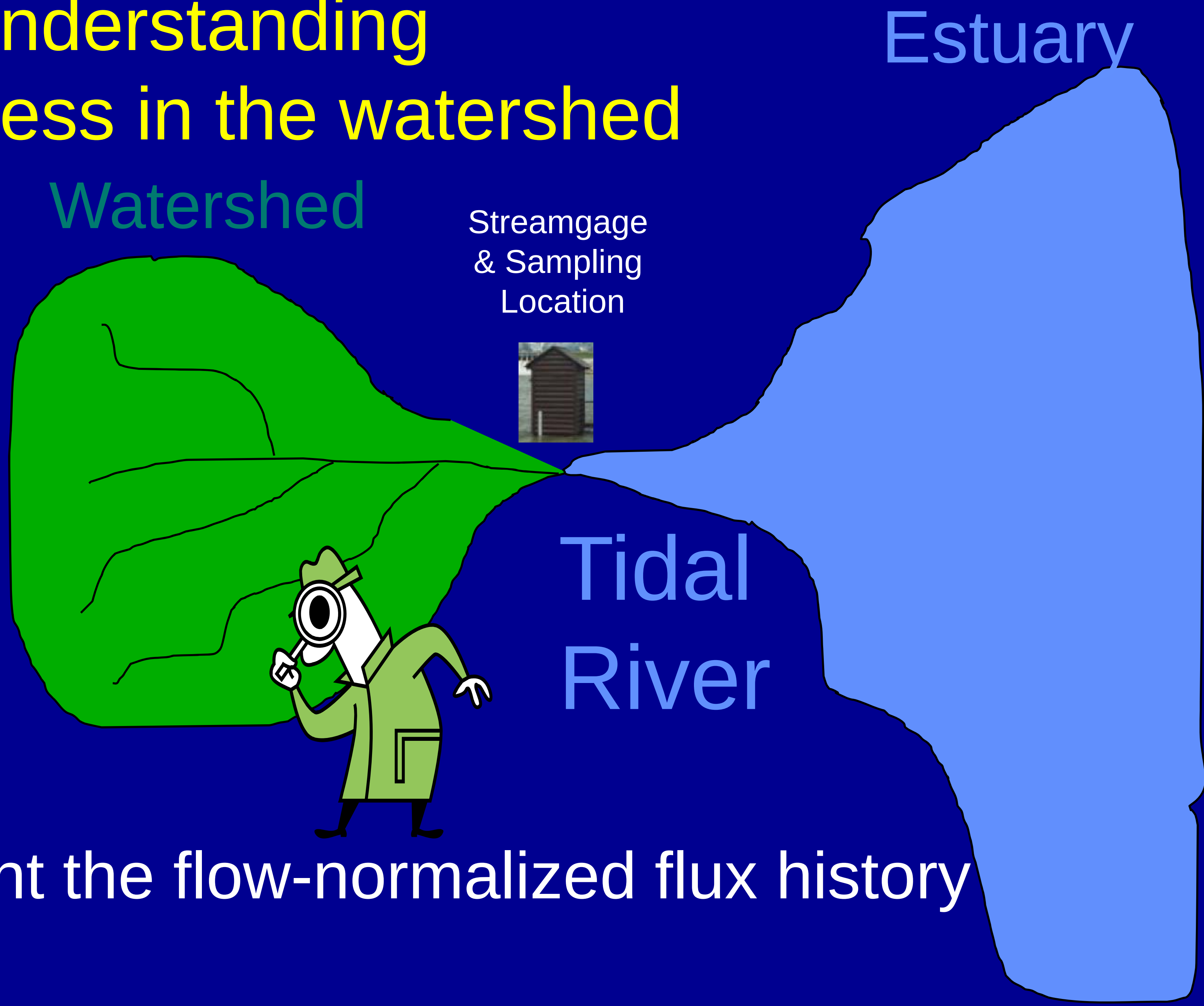
USGS

For understanding impact on the estuary ecosystem

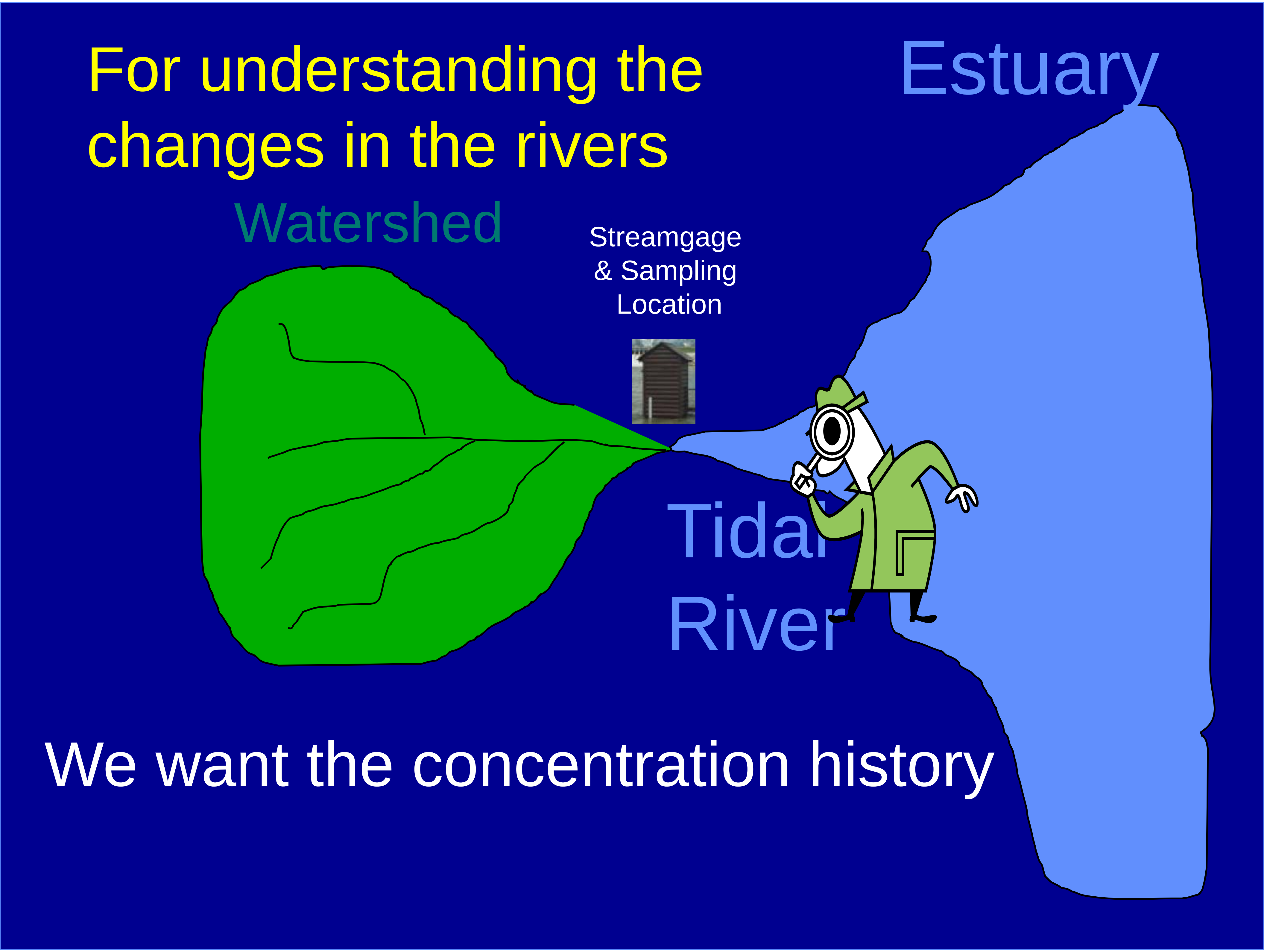


We want the flux history

For understanding progress in the watershed



We want the flow-normalized flux history


For understanding the changes in the rivers
Estuary
Watershed
Streamgage & Sampling Location
Tidal River
We want the concentration history

The software: how do I get it?

- **Need to install R (freely downloaded from http://cran.us.r-project.org/) on your computer**

- **Once you start R, you can load the software:**

```
install.packages("EGRET", "EGRETci")
library(EGRET)
library(EGRETci)
```

check out our new developments at:

https://github.com/USGS-R/EGRET/wiki

see also https://owi.usgs.gov/blog/

USGS

Using EGRET

- For each session the code needs to be loaded:

  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 readUserSample) just type ?readUserSample

- For this workshop I’m running through the steps as if we are in interactive mode. All of this can be done in batch mode.

USGS

How can we enter data

- For the water quality sample data
  - From USGS web services
  - From Water Quality Portal (for STORET)
  - 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 or Water Quality Portal
  - From user entries

```
> library(dataRetrieval)
> library(EGRET)
> siteNumber <- "01491000"
> parameterCd <- "00631"
> startDate <- "1979-10-01"
> endDate <- "2014-09-28"
> Sample<- readNWISSample(siteNumber,parameterCD,startDate,endDate)
> summary(Sample)
```

```
> Sample<-readNWISSample("01491000","00631","1979-10-01","2014-09-28")
> summary(Sample)
      Date                ConcLow         ConcHigh          Uncen            ConcAve          Julian
 Min.   :1979-10-24   Min.   :0.176   Min.   :0.050   Min.   :0.0000   Min.   :0.025   Min.   :47412
 1st Qu.:1989-03-18   1st Qu.:0.900   1st Qu.:0.900   1st Qu.:1.0000   1st Qu.:0.900   1st Qu.:50845
 Median :1995-01-21   Median :1.130   Median :1.130   Median :1.0000   Median :1.130   Median :52980
 Mean   :1996-10-21   Mean   :1.138   Mean   :1.137   Mean   :0.9986   Mean   :1.137   Mean   :53620
 3rd Qu.:2004-10-12   3rd Qu.:1.400   3rd Qu.:1.400   3rd Qu.:1.0000   3rd Qu.:1.400   3rd Qu.:56532
 Max.   :2014-08-13   Max.   :2.430   Max.   :2.430   Max.   :1.0000   Max.   :2.430   Max.   :60124
                      NA's   :1
     Month             Day             DecYear        MonthSeq          SinDY              CosDY
 Min.   : 1.000   Min.   :  2.00   Min.   :1980   Min.   :1558   Min.   :-1.00000   Min.   :-1.000000
 1st Qu.: 3.000   1st Qu.: 83.75   1st Qu.:1989   1st Qu.:1671   1st Qu.:-0.62876   1st Qu.:-0.686704
 Median : 6.000   Median :157.00   Median :1995   Median :1741   Median : 0.19667   Median :-0.021513
 Mean   : 6.082   Mean   :169.23   Mean   :1997   Mean   :1762   Mean   : 0.09121   Mean   :-0.001613
 3rd Qu.: 9.000   3rd Qu.:256.25   3rd Qu.:2005   3rd Qu.:1858   3rd Qu.: 0.79226   3rd Qu.: 0.700615
 Max.   :12.000   Max.   :364.00   Max.   :2015   Max.   :1976   Max.   : 0.99992   Max.   : 0.999667

> length(Sample$Date)
[1] 708
```

If the data are in a spreadsheet, it can be entered as a csv file with 3 columns:

```
> library(EGRET)
> Sample <- readUserSample(savePath,fileName="TP.csv")
```

	A	B	C
1	Sample Date	rem	PHOSPHOROUS TOTAL
2	5/28/1974		0.07
3	7/30/1974		0.05
4	8/21/1974		0.067
5	9/17/1974		0.045
6	10/22/1974		0.06
7	12/10/1974		0.032
8	1/21/1975		0.006
9	2/17/1975		0.041
10	3/18/1975		0.04
11	4/23/1975		0.355

USGS

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.

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

EPA Storet Data from the Water Quality Portal

```
> siteNumber<-"IL_EPA_WQX-BPK-07"
> characteristicName<-"Inorganic nitrogen (nitrate and nitrite)"
> startDate<-"2005-01-01"
> endDate<-"2013-12-31"
> Sample<-readWQPSample(siteNumber,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.   :-0.997917   Min.   :-0.99867
 1st Qu.: 4.000   1st Qu.: 96.5   1st Qu.:2009   1st Qu.:1910   1st Qu.:-0.739146   1st Qu.:-0.69630
 Median : 6.500   Median :184.0   Median :2010   Median :1921   Median : 0.000000   Median :-0.14961
 Mean   : 6.425   Mean   :179.5   Mean   :2009   Mean   :1913   Mean   :-0.009202   Mean   :-0.07491
 3rd Qu.: 9.000   3rd Qu.:256.2   3rd Qu.:2011   3rd Qu.:1934   3rd Qu.: 0.740889   3rd Qu.: 0.62203
 Max.   :12.000   Max.   :349.0   Max.   :2012   Max.   :1943   Max.   : 0.999250   Max.   : 0.98666

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

USGS

Daily <- readNWISDAILY(siteNumber,"00060",startDate,endDate

```
> Daily<-readNWISDaily("01491000","00060","1979-10-01","2014-09-28")
There are 12782 data points, and 12782 days.
> summary(Daily)
      Date                  Q                   Julian           Month             Day              DecYear
 Min.   :1979-10-01   Min.   :  0.00991   Min.   :47389   Min.   : 1.000   Min.   :  1.0   Min.   :1980
 1st Qu.:1988-06-30   1st Qu.:  0.96277   1st Qu.:50584   1st Qu.: 4.000   1st Qu.: 93.0   1st Qu.:1988
 Median :1997-03-30   Median :  2.46357   Median :53780   Median : 7.000   Median :184.0   Median :1997
 Mean   :1997-03-30   Mean   :  4.17317   Mean   :53780   Mean   : 6.522   Mean   :183.7   Mean   :1997
 3rd Qu.:2005-12-28   3rd Qu.:  4.72891   3rd Qu.:56975   3rd Qu.:10.000   3rd Qu.:275.0   3rd Qu.:2006
 Max.   :2014-09-28   Max.   :246.35656   Max.   :60170   Max.   :12.000   Max.   :366.0   Max.   :2015

    MonthSeq      Qualifier                  i                 LogQ                  Q7                  Q30
 Min.   :1558   Length:12782       Min.   :    1   Min.   :-4.61412   Min.   : 0.01808   Min.   : 0.09606
 1st Qu.:1662   Class :character   1st Qu.: 3196   1st Qu.:-0.03794   1st Qu.: 1.00727   1st Qu.: 1.21102
 Median :1767   Mode  :character   Median : 6392   Median : 0.90161   Median : 2.63549   Median : 2.97421
 Mean   :1767                      Mean   : 6392   Mean   : 0.78216   Mean   : 4.17433   Mean   : 4.17615
 3rd Qu.:1872                      3rd Qu.: 9587   3rd Qu.: 1.55370   3rd Qu.: 5.09804   3rd Qu.: 5.88802
 Max.   :1977                      Max.   :12782   Max.   : 5.50678   Max.   :84.00395   Max.   :25.47478
                                                                      NA's   :6          NA's   :29
> length(Daily$Date)
[1] 12782
```

Storing the metadata

- For NWIS data INFO<-readNWISInfo(siteNumber,parameterCD)
- Similar function for the Water Quality Portal
- The contents of INFO are used to label tables and figures as well as document the site and constituent information
- Creates a system of abbreviations to keep track of workspace files

USGS

```
> INFO<-readNWISInfo(siteNumber,parameterCd)

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

USGS

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):

Dissolved Nitrate, mg/L 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

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

> INFO <- readNWISInfo(“”,””)

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

> eList<-mergeReport(INFO,Daily,Sample)

```
> eList<-mergeReport(INFO,Daily,Sample)

 Discharge Record is 12782 days long, which is 35 years
 First day of the discharge record is 1979-10-01 and last day is 2014-09-28
 The water quality record has 708 samples
 The first sample is from 1979-10-24 and the last sample is from 2014-08-13
 Discharge: Minimum, mean and maximum 0.00991 4.17 246
 Concentration: Minimum, mean and maximum 0.05 1.1 2.4
 Percentage of the sample values that are censored is 0.14 %
```

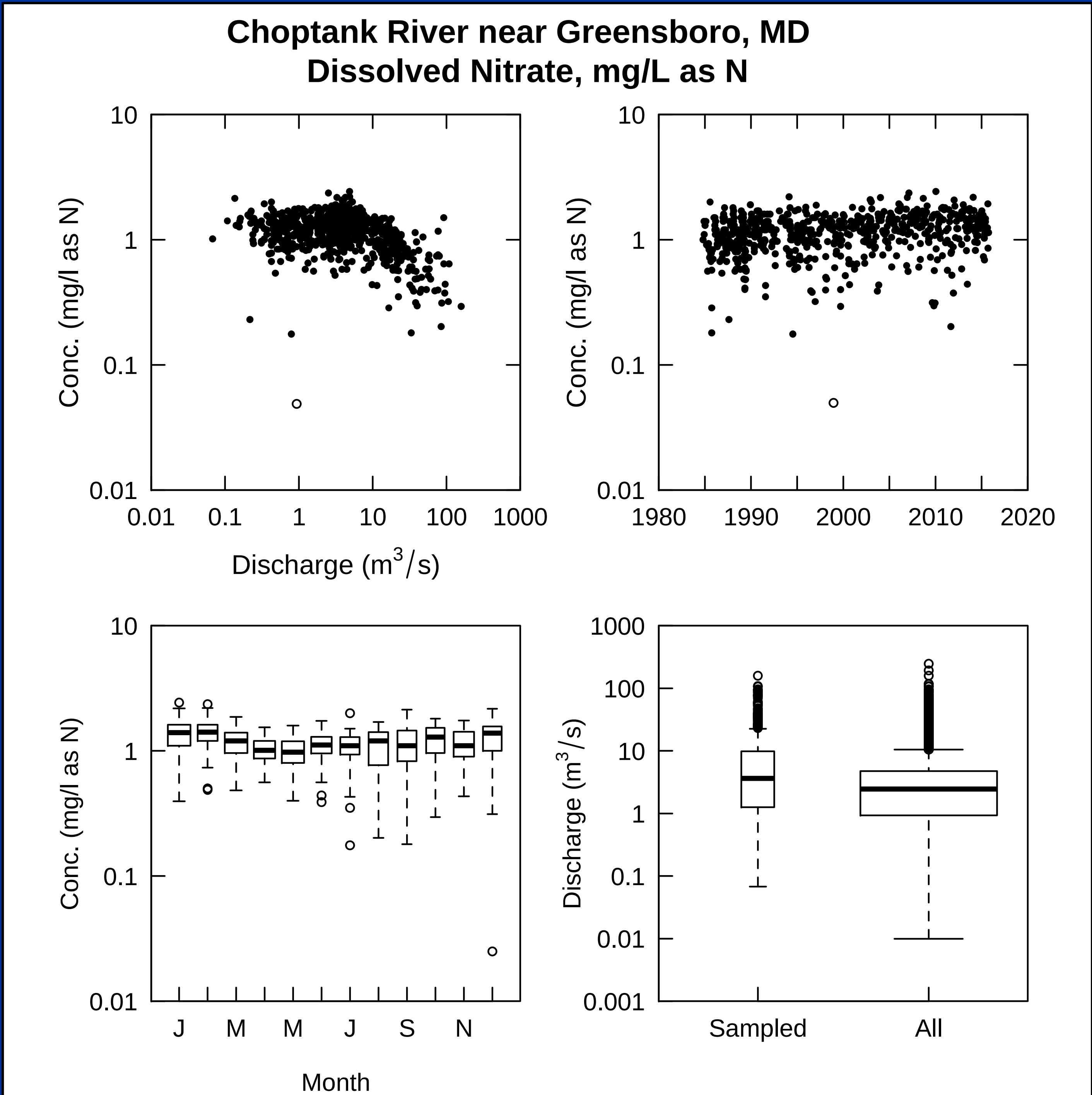
eList is a named list that contains the 4 objects that contain all the data and results:

INFO, Daily, Sample, and surfaces

USGS

> multiPlotDataOverview(eList,qUnit=1)

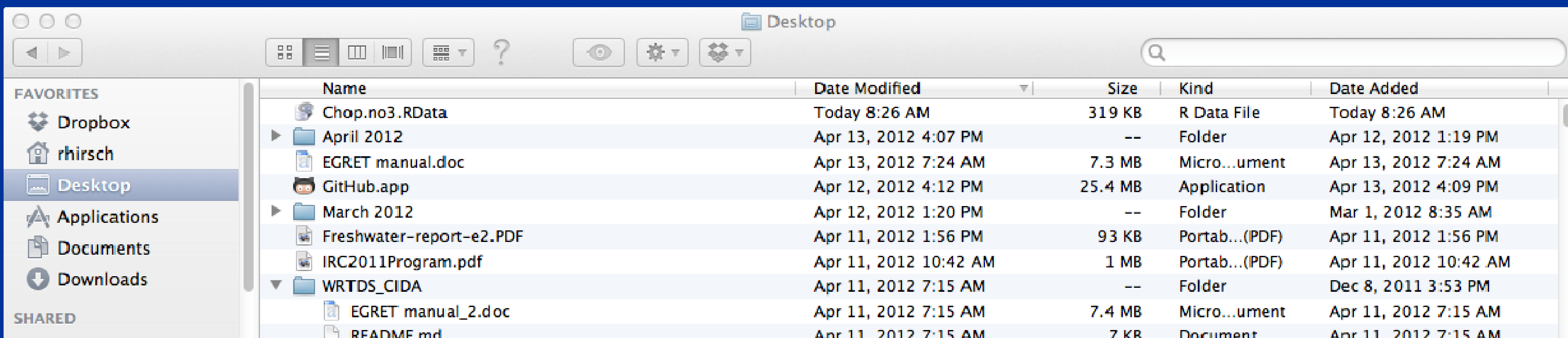
Let's look at the data before we proceed, the function is:



USGS

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

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



Save it over and over as you proceed and add results

We now have 3 data frames

- Sample (651 rows, 15 columns)
- Daily (11,323 rows, 12 columns)
- INFO (1 row, 53 columns)

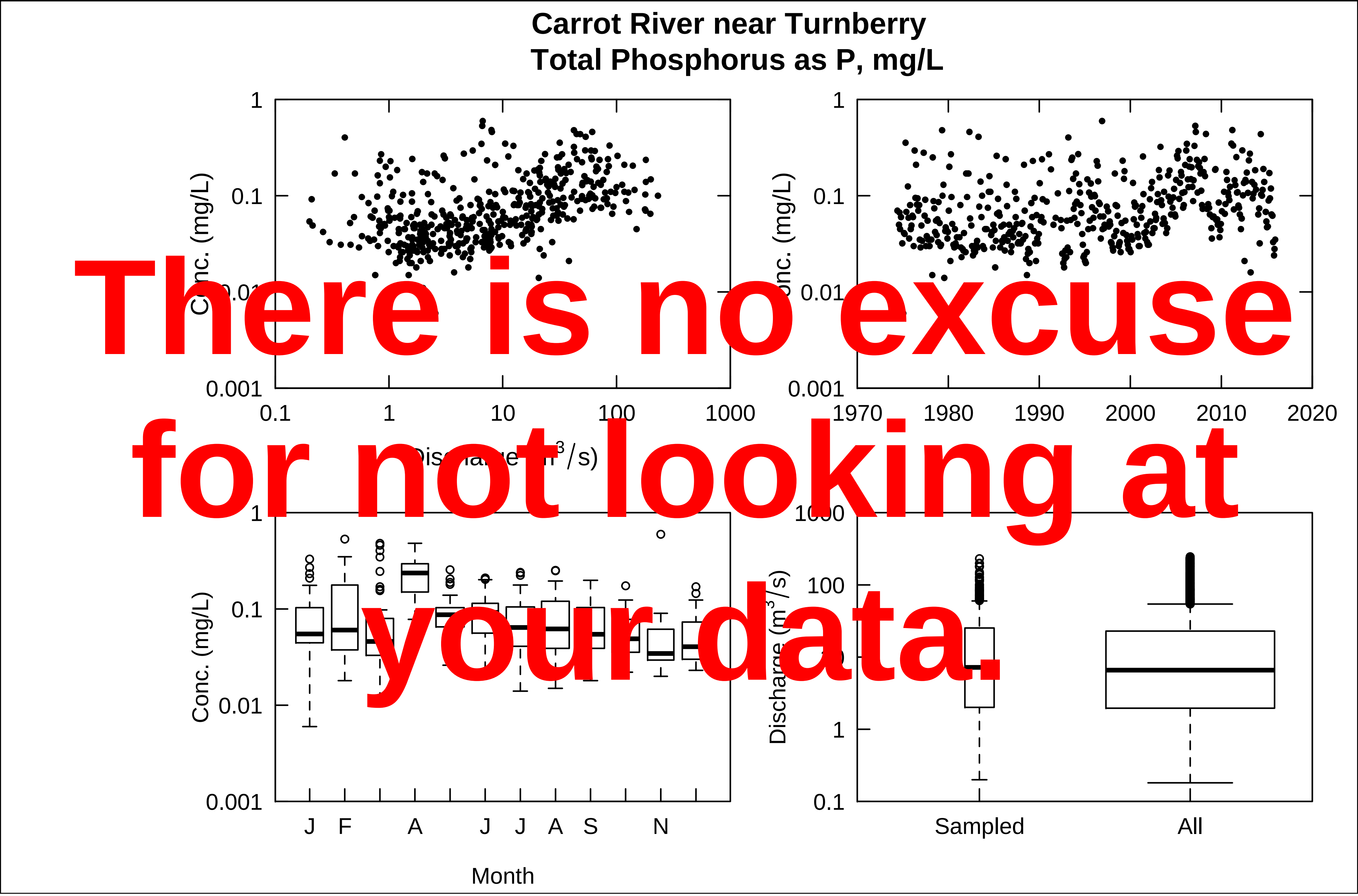
They are all stored in the named list called eList

A comment about the EGRET data frames

- Having a standard set of variable names and units and other conventions has great value
- People can share data sets and not have to explain the structure
- EGRET is a great platform for many analyses well beyond WRTDS

USGS

> multiPlotDataOverview(eList)



There is no excuse for not looking at your data.

> eList <- modelEstimation(eList)

- Runs the model in cross-validation mode
- Estimates the "concentration surface"
- Uses the surface to compute daily values of
  - Concentration
  - Flux
  - Flow-normalized concentration
  - Flow-normalized flux
- Adds those to the Daily data frame

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

USGS

We now have 3 data frames, bound together in eList

- Sample (651 rows, 18 columns)
- Daily (11,323 rows, 18 columns)
- INFO (1 row, 68 columns)
- plus the surface represented by the contour plot (14 x 513)

USGS All can be saved as “eList”

"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

eList <- setPA(eList, paStart = 1, paLong=12)

Say we want April, May, June

eList <- setPA(eList, paStart = 4, paLong = 3)

Default is water year

USGS

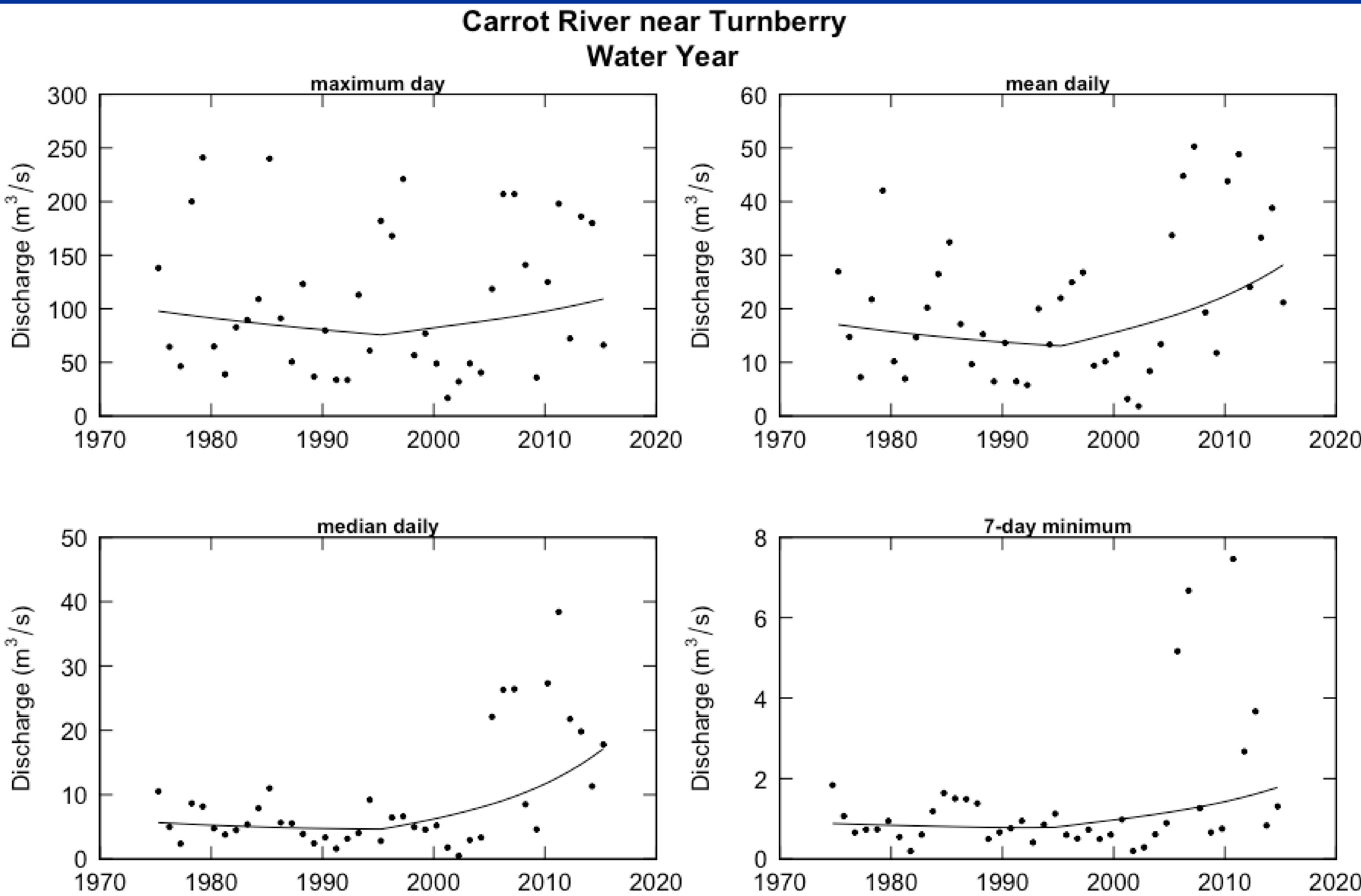
Units in EGRET

Everything stored as:

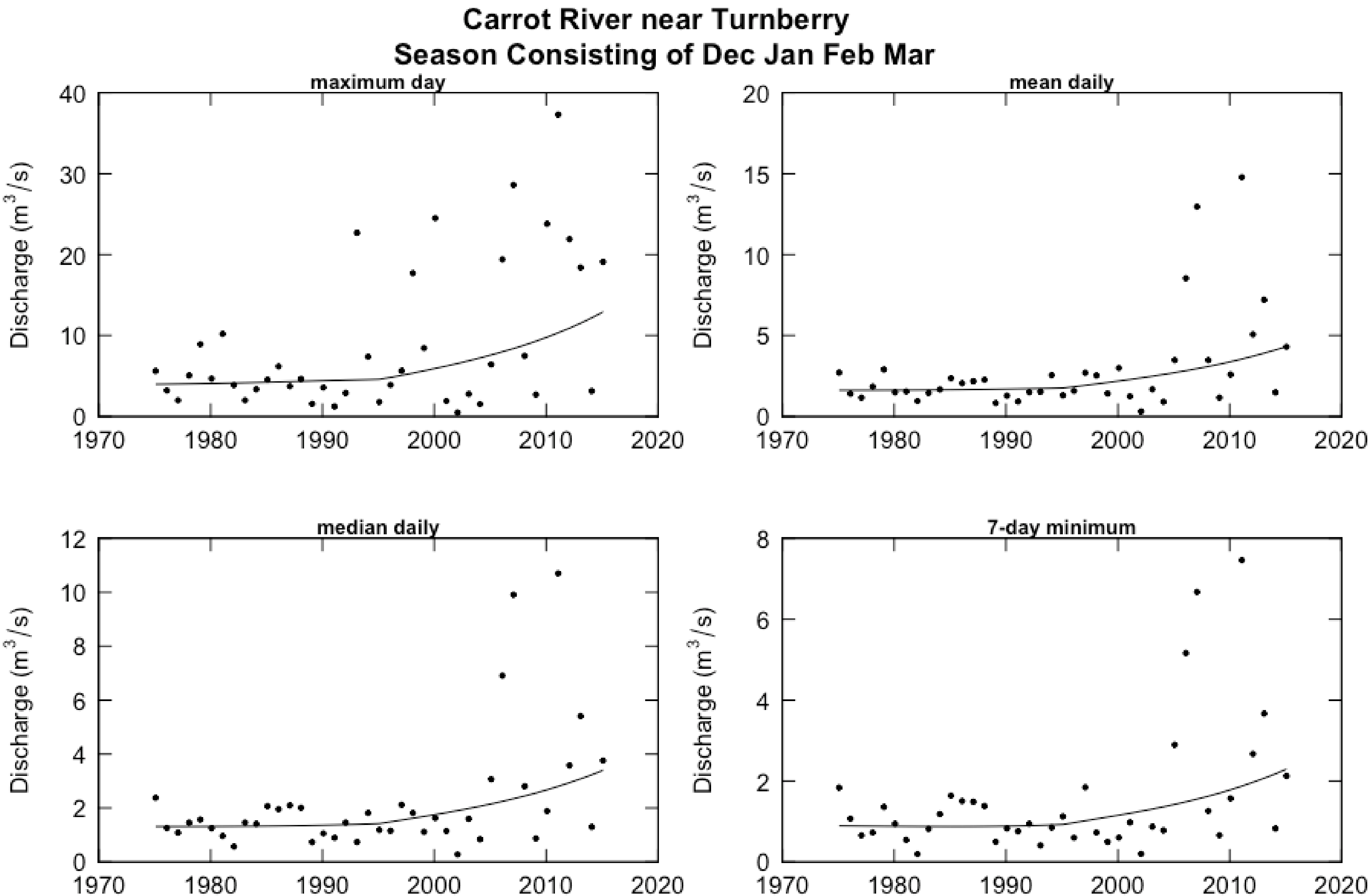
m^3/s, kg/day, or mg/L

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

Brief digression to flow history in EGRET plotFourStats(eList, qUnit=2)



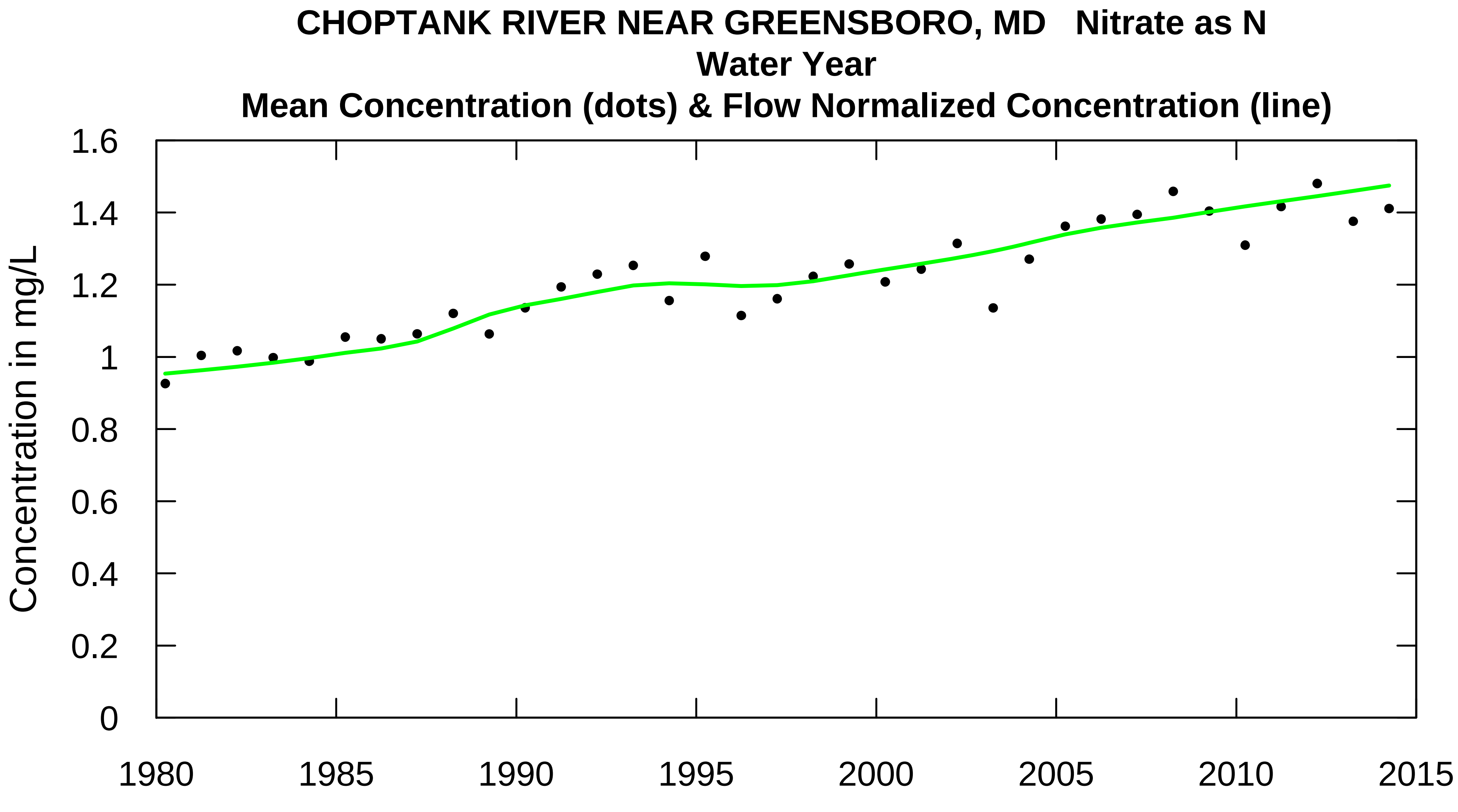
Let's look at winter flows
paStart = 12, paLong = 4



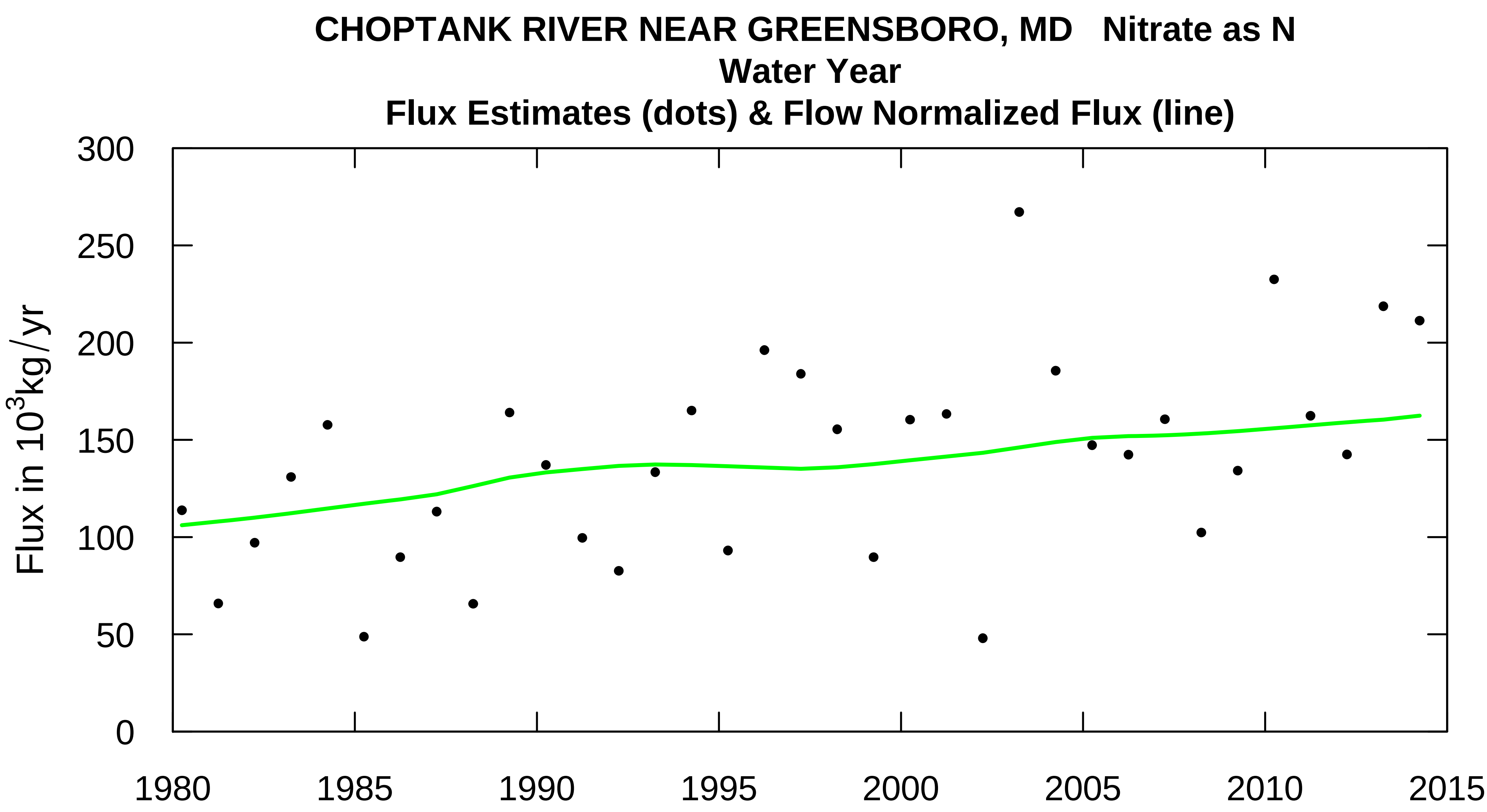
Back to WRTDS trend results

Choptank River example

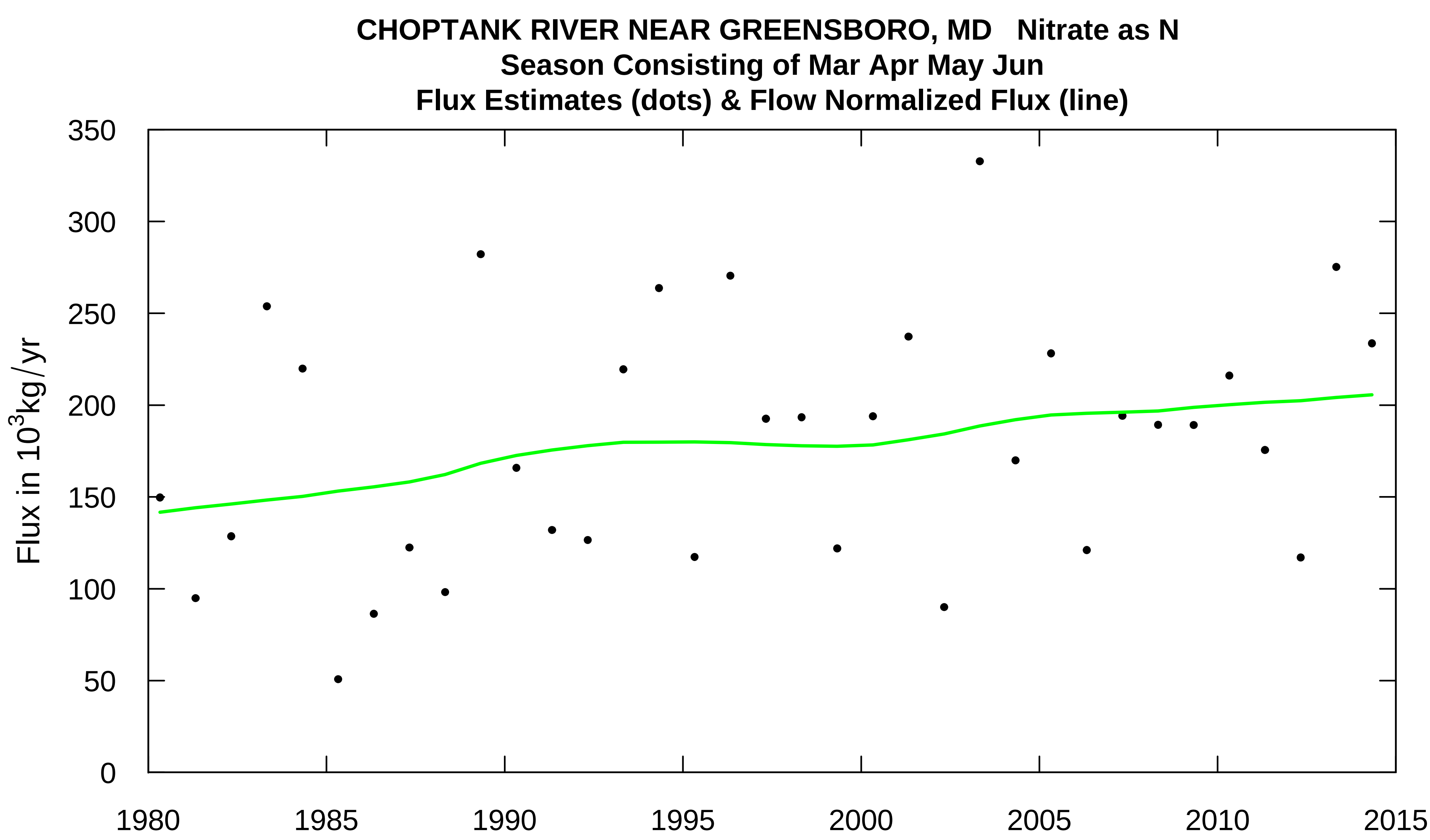
> plotConcHist(eList)



> plotFluxHist(eList, fluxUnit = 8)



```
> eList <- setPA(paStart = 3, paLong = 4)
> plotFluxHist(fluxUnit = 8)
```



Graphics options

- Print or not print the title
- Change font sizes
- Set axis maximum
- Use log scale
- Change colors
- Save image as .png or .pdf
- …

USGS

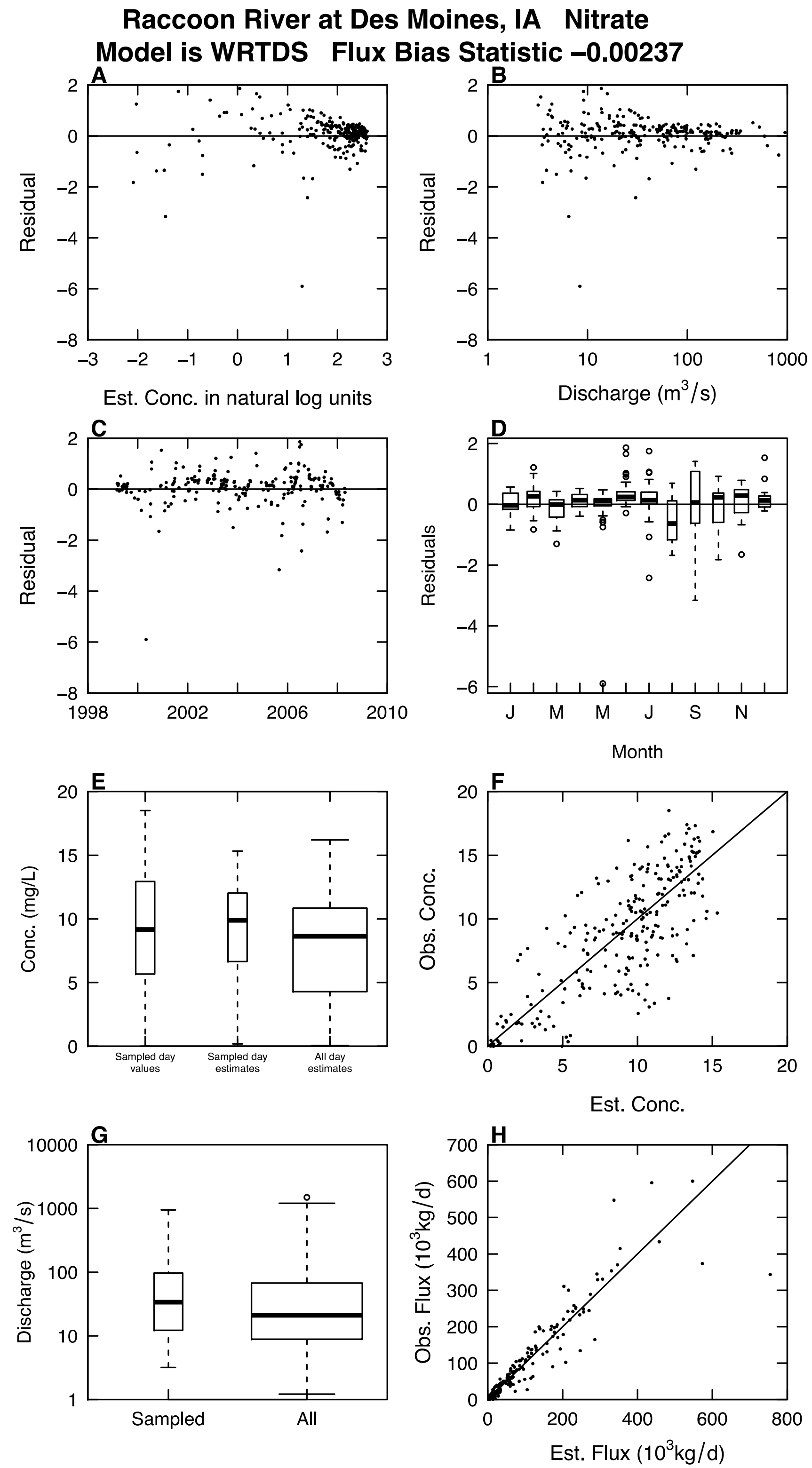
```
> tableResults(eList, qUnit = 1, fluxUnit = 5)

   CHOPTANK RIVER NEAR GREENSBORO, MD
   Nitrate as N
   Water Year

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

   1980     150.2     0.926     0.953     125.5        117
   1981      78.3     1.004     0.963      72.6        119
   1982     107.6     1.017     0.972     107.0        121
   1983     176.1     0.998     0.984     144.4        124
   1984     201.9     0.988     0.997     173.9        126
   1985      53.6     1.055     1.011      53.8        129
   1986      92.8     1.050     1.023      98.9        132
   1987     119.1     1.064     1.043     124.7        135
   1988      66.0     1.121     1.079      72.4        139
     .
     .
     .
   2007     151.2     1.395     1.373     177.1        168
   2008      90.5     1.459     1.386     112.8        169
   2009     130.0     1.404     1.402     147.9        170
   2010     254.0     1.310     1.417     256.4        172
   2011     185.2     1.417     1.431     179.0        174
   2012     122.6     1.480     1.445     157.1        175
   2013     226.0     1.376     1.460     241.1        177
   2014     191.8     1.411     1.475     233.0        179
```

```
> tableChange(eList, fluxUnit=5,yearPoints=c(1980,1995,2014))

   CHOPTANK RIVER NEAR GREENSBORO, MD
   Nitrate as N
   Water Year

           Concentration trends
   time span          change     slope    change     slope
                        mg/L   mg/L/yr         %      %/yr

 1980  to  1995         0.25     0.017        26       1.7
 1980  to  2014         0.52     0.015        55       1.6
 1995  to  2014         0.27     0.014        23       1.2


                 Flux Trends
   time span             change          slope          change          slope
                     tons/yr      tons/yr   /yr           %          %/yr
 1980  to  1995              33            2.2              29            1.9
 1980  to  2014              62            1.8              53            1.6
 1995  to  2014              29            1.5              19              1
```

USGS

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

fluxBiasMulti(eList, fluxUnit=4)

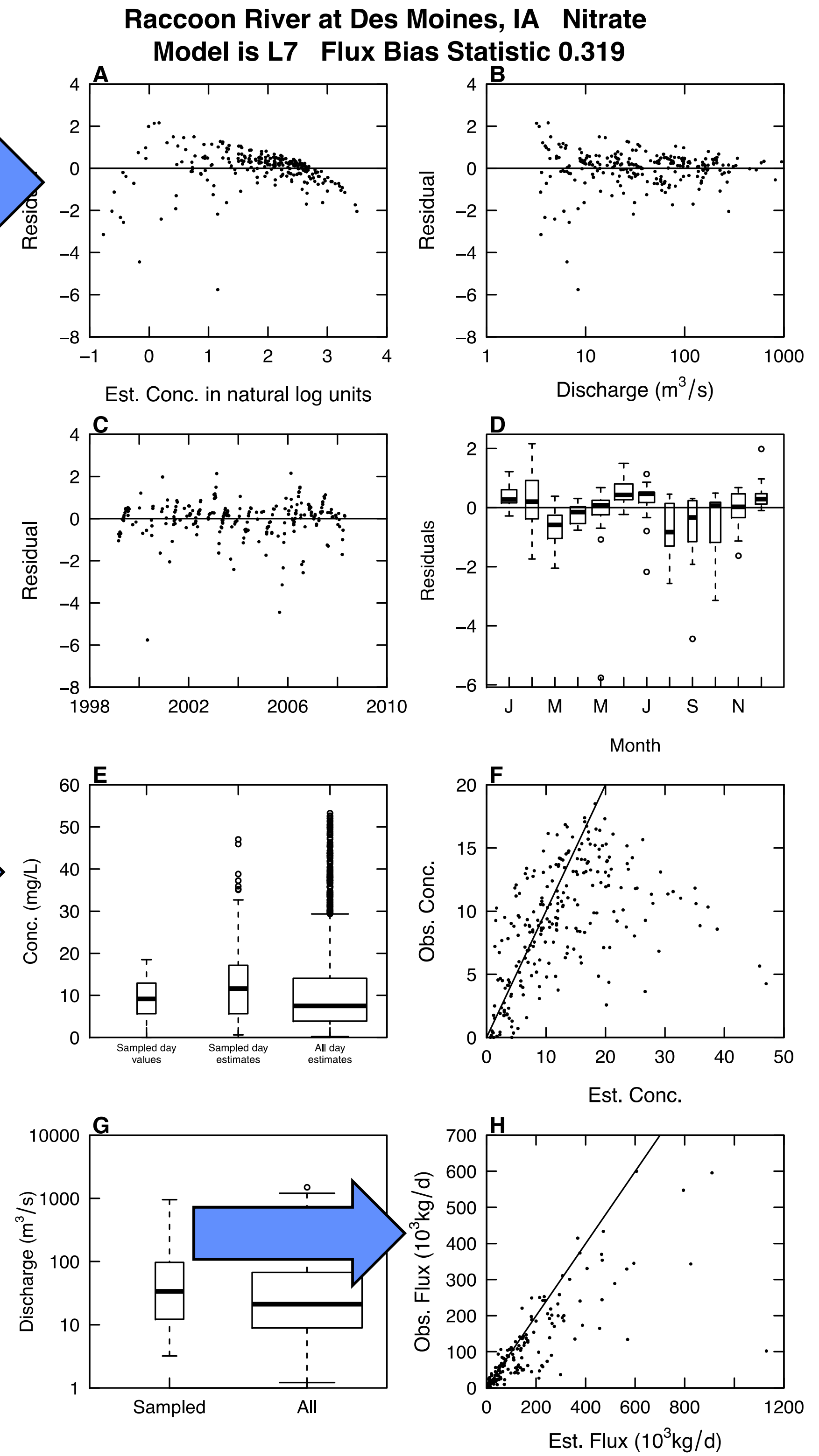


curvature

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

Extreme predictions

Flux bias



Diagnostics and potential problems with estimating mean flux, see:

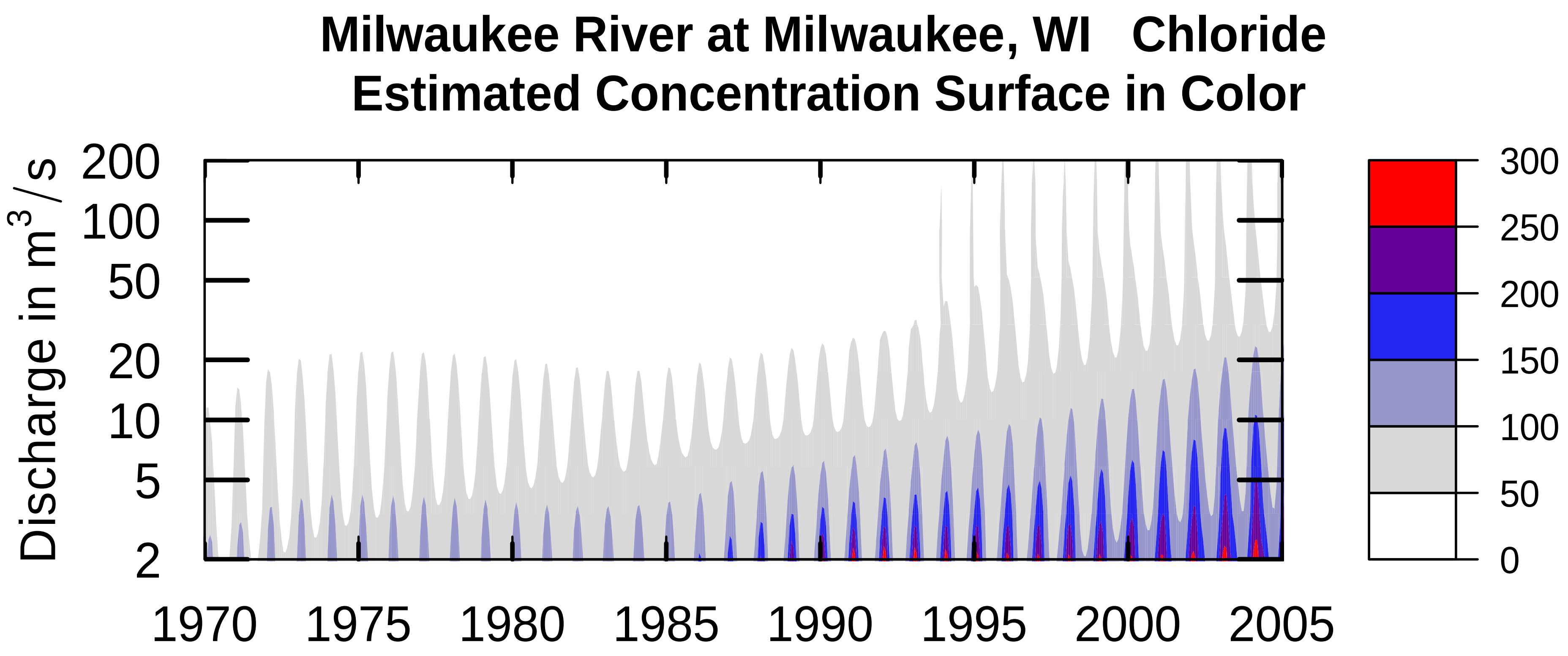
Hirsch, R.M., 2014, Large biases in regression-based constituent flux estimates: causes and diagnostics. Journal of the American Water Resources Association.

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

USGS

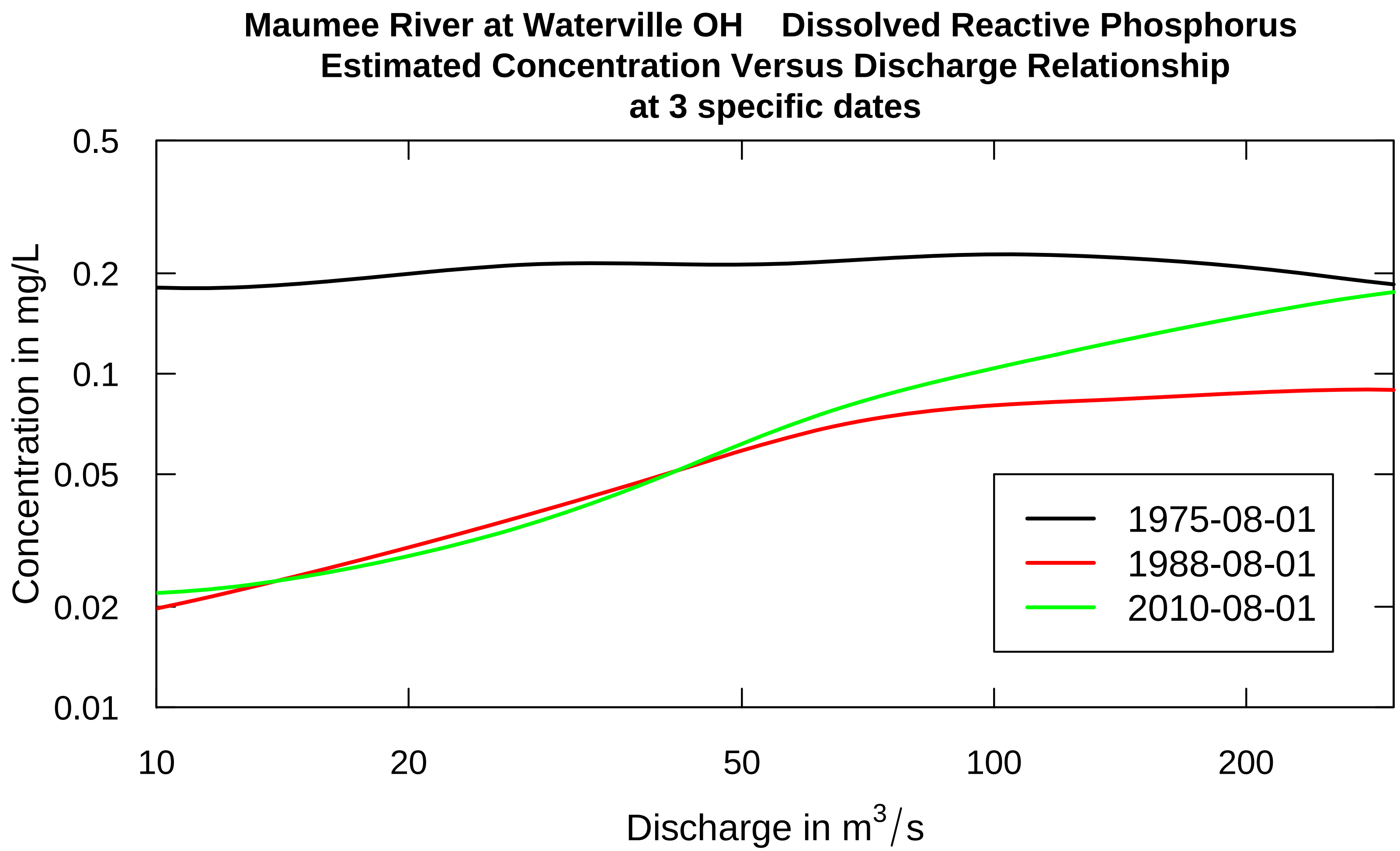
How difficult is it to make those contour plots?

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

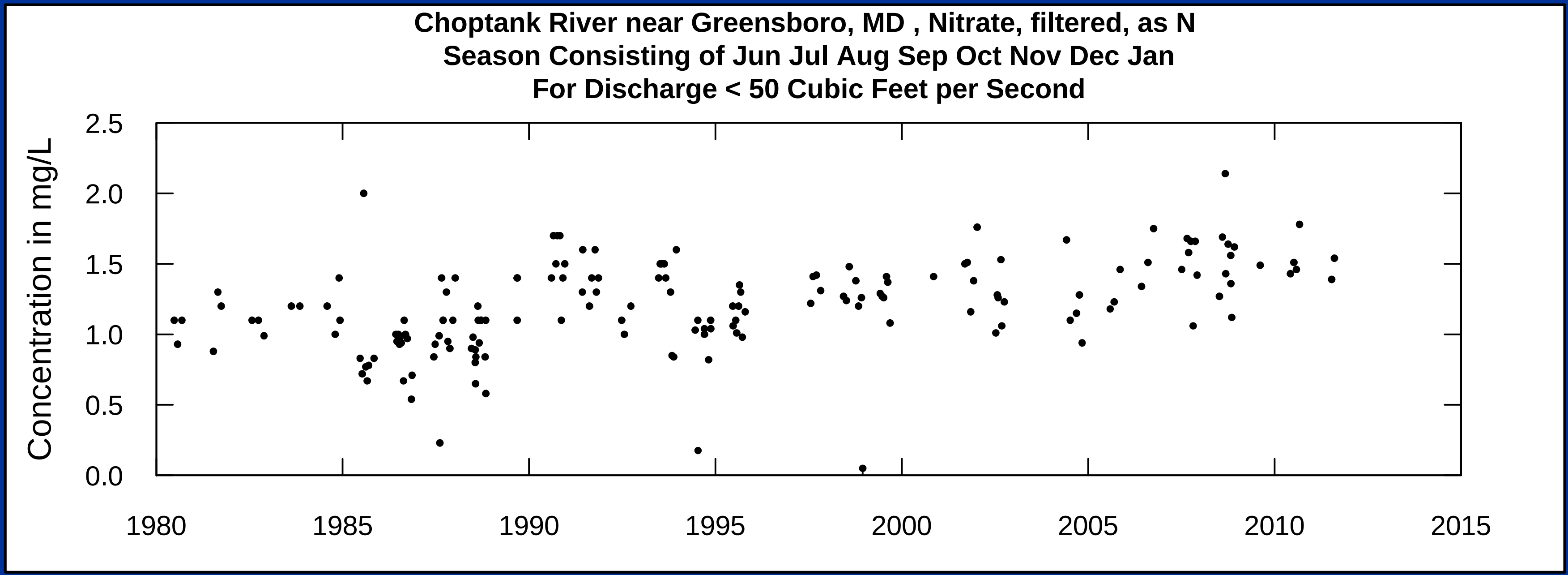


There are many more graphics, for example

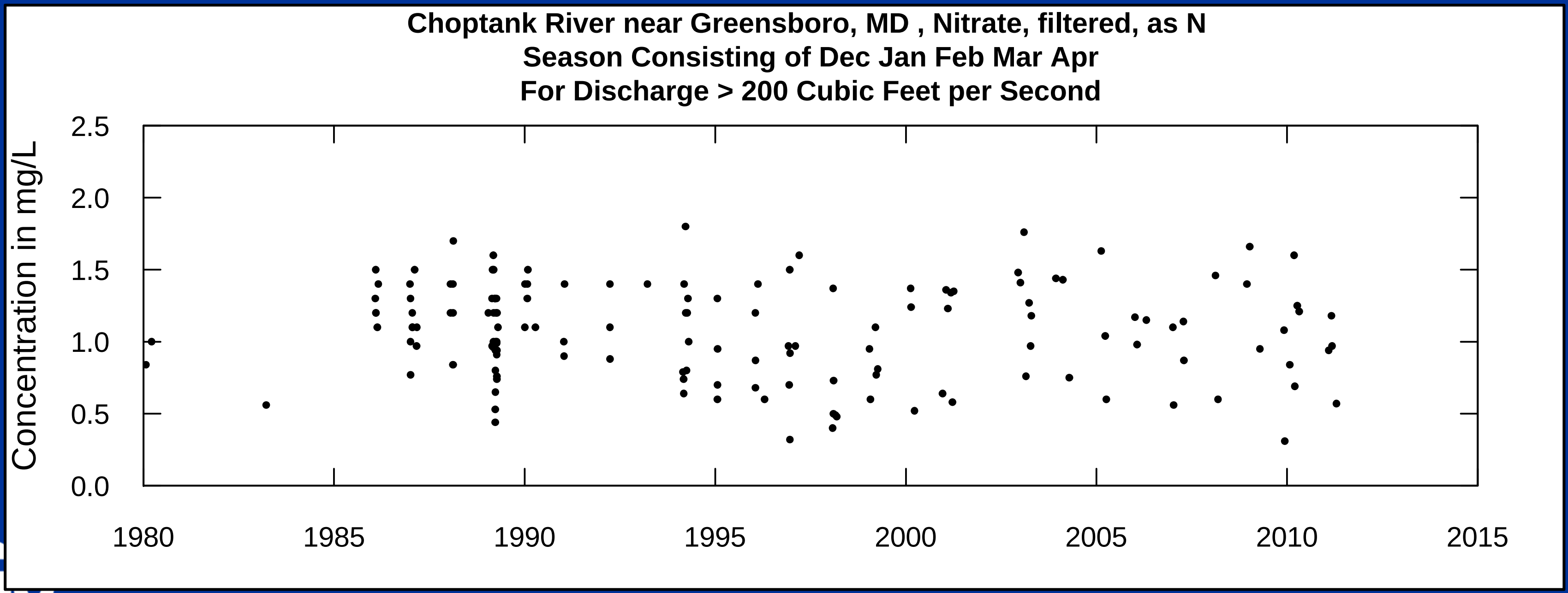
> plotConcQSmooth(eList,"1975-08-01”, "1988-08-01", "2010-08-01", qLow=10, qHigh=300, qUnit=2, logScale=TRUE, legendLeft=100, legendTop=0.05)



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



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



Uncertainty analysis: WRTDS Bootstrap Test (wBT) in EGRETci package

- WRTDS developed as an exploratory data analysis method
- Users liked it, but wanted to bring in formal analysis of uncertainty on the trend results

USGS

Based on published paper

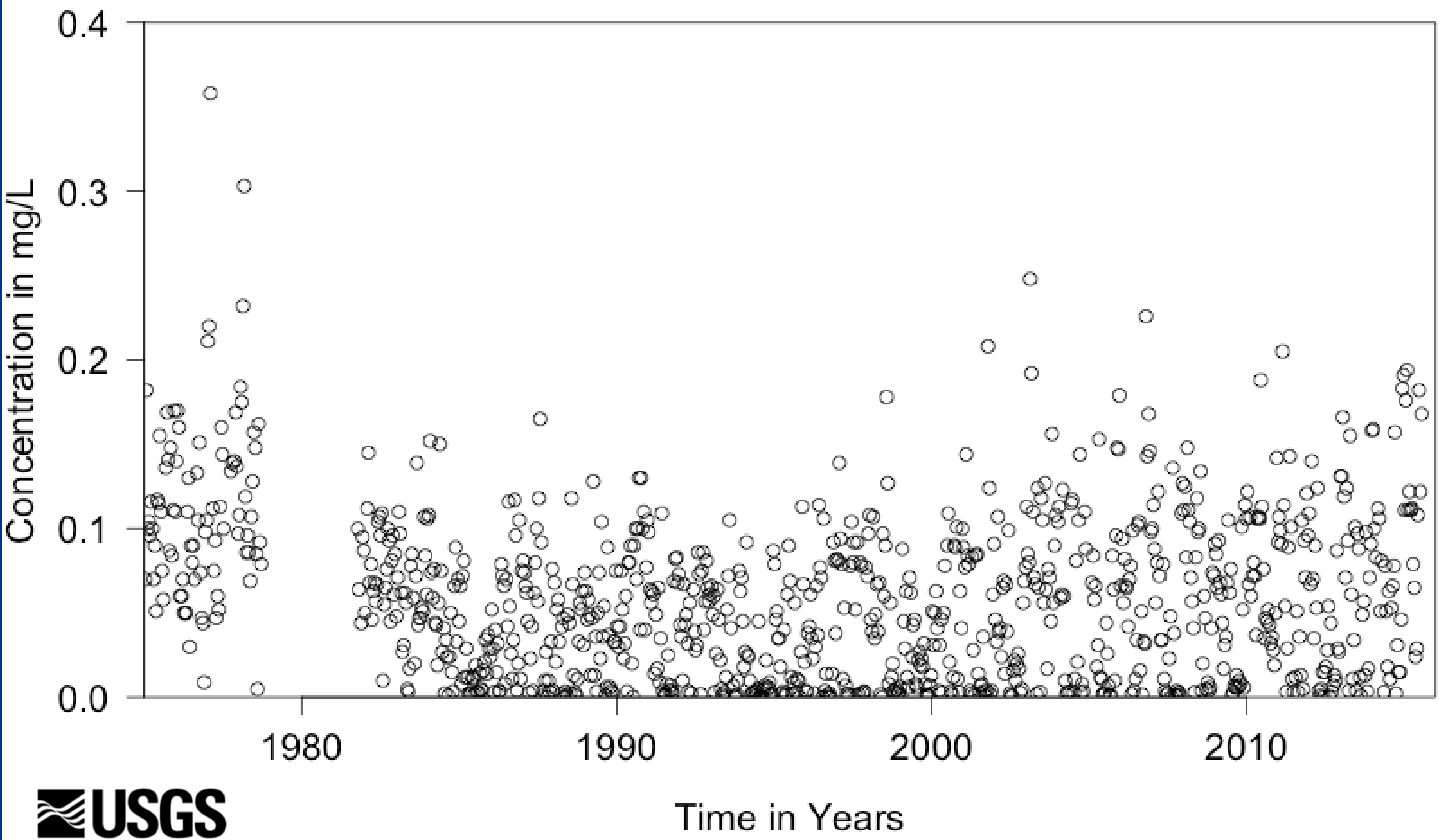
Hirsch, Robert M., Archfield, Stacey A., and DeCicco, Laura A., 2015,

"A bootstrap method for estimating uncertainty of water quality trends"

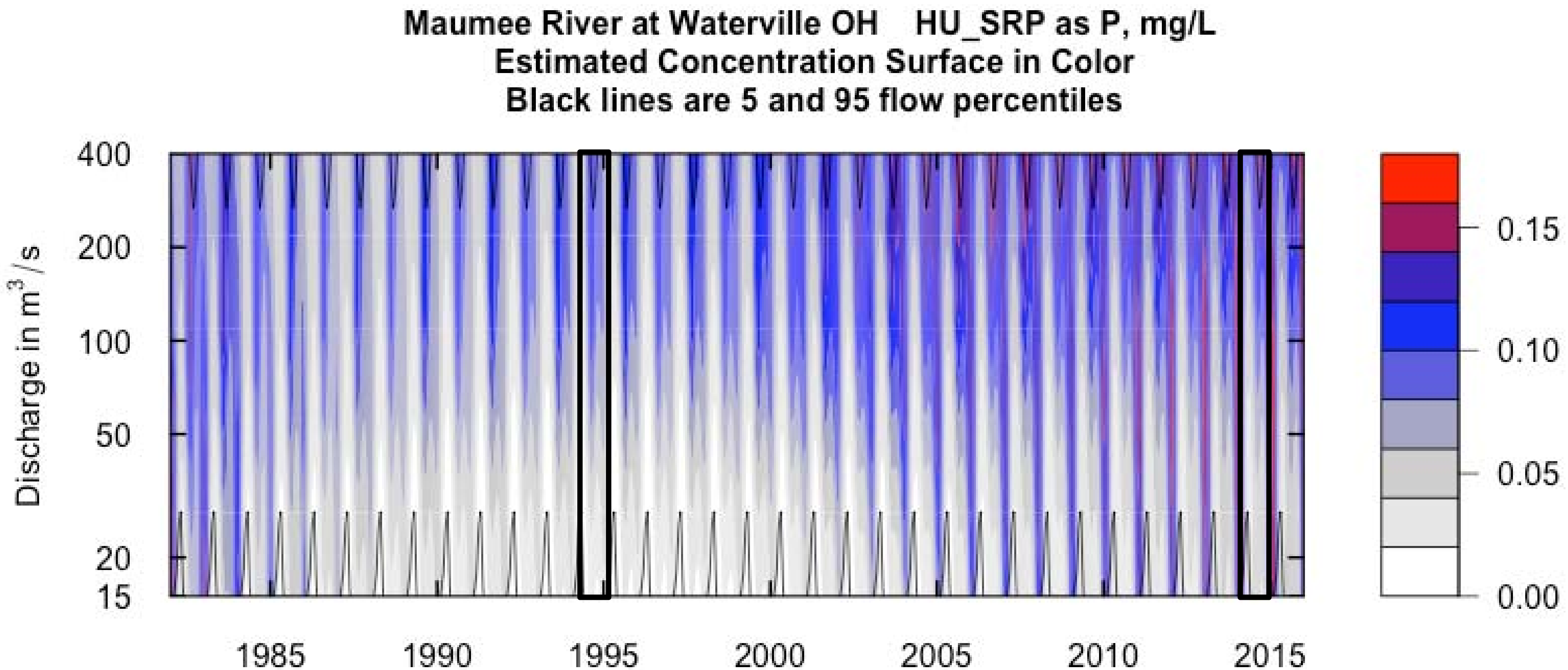
Environmental Modelling and Software, 73, 148-166.

USGS

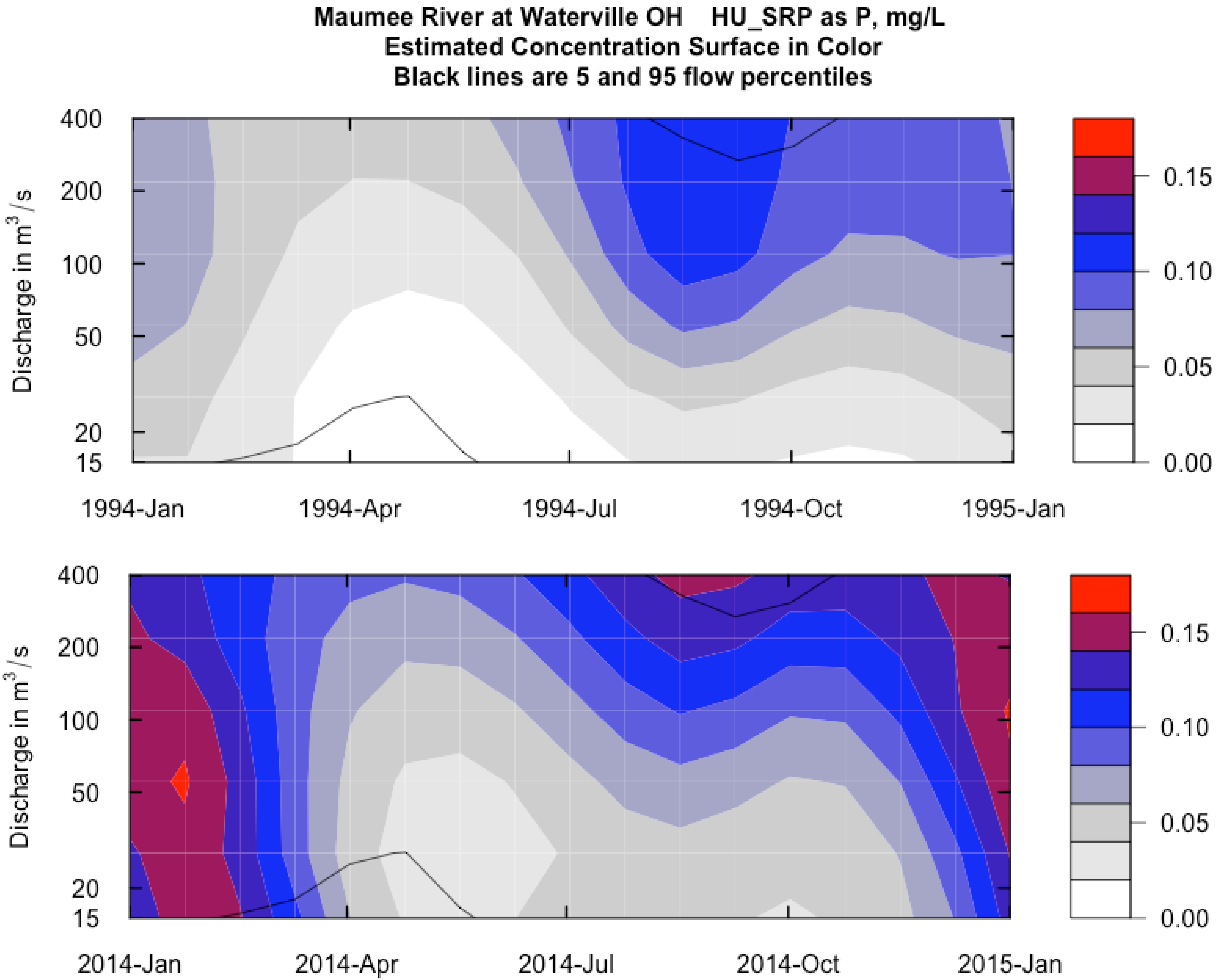
Raw Sample of SRP, Maumee River



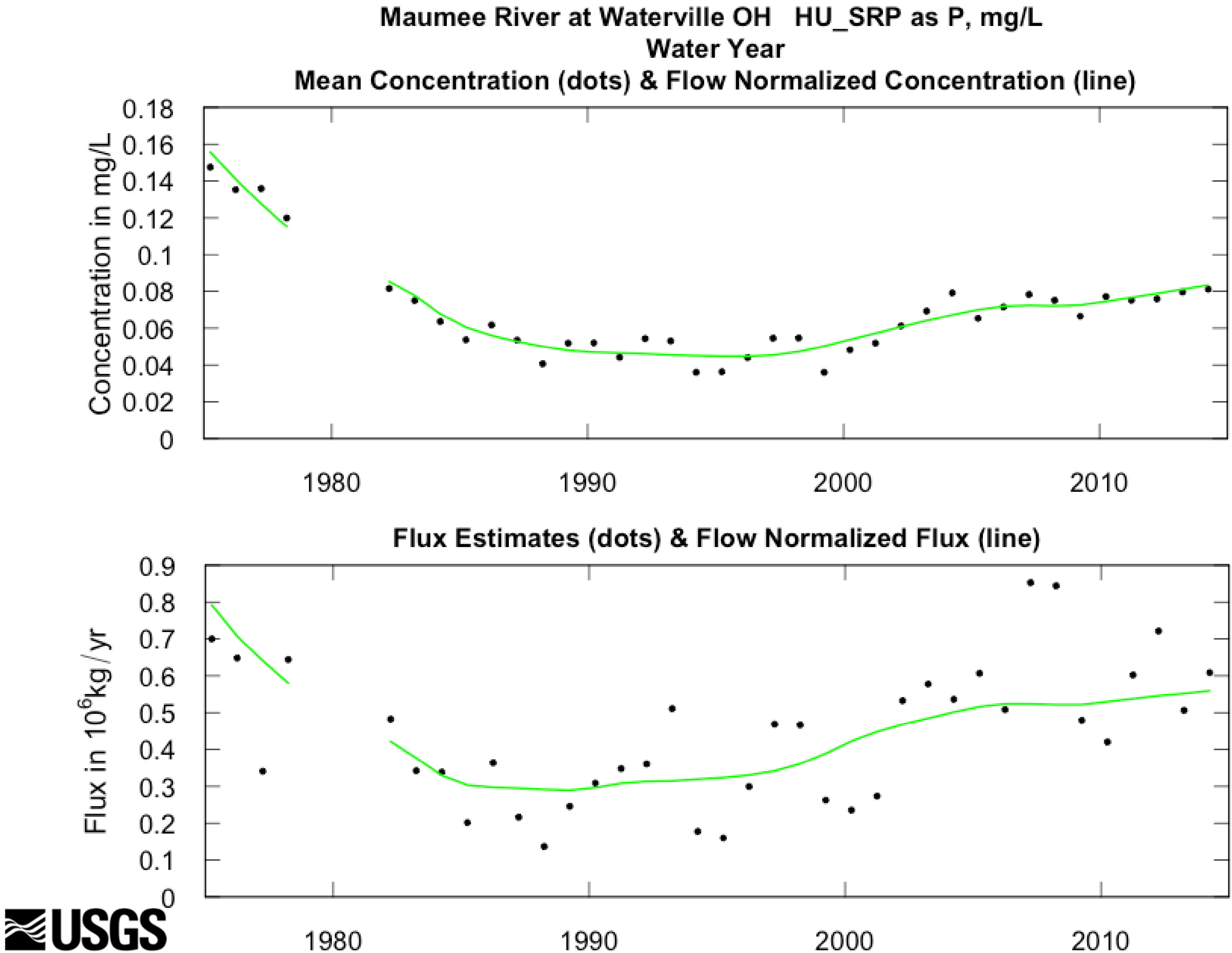
WRTDS representation of concentration as a function of time and discharge



USGS



Maumee River at Waterville OH HU_SRP as P, mg/L
Estimated Concentration Surface in Color
Black lines are 5 and 95 flow percentiles
Discharge in m^3/s
400
200
100
50
20
15
1994-Jan
1994-Apr
1994-Jul
1994-Oct
1995-Jan
0.15
0.10
0.05
0.00
2014-Jan
2014-Apr
2014-Jul
2014-Oct
2015-Jan



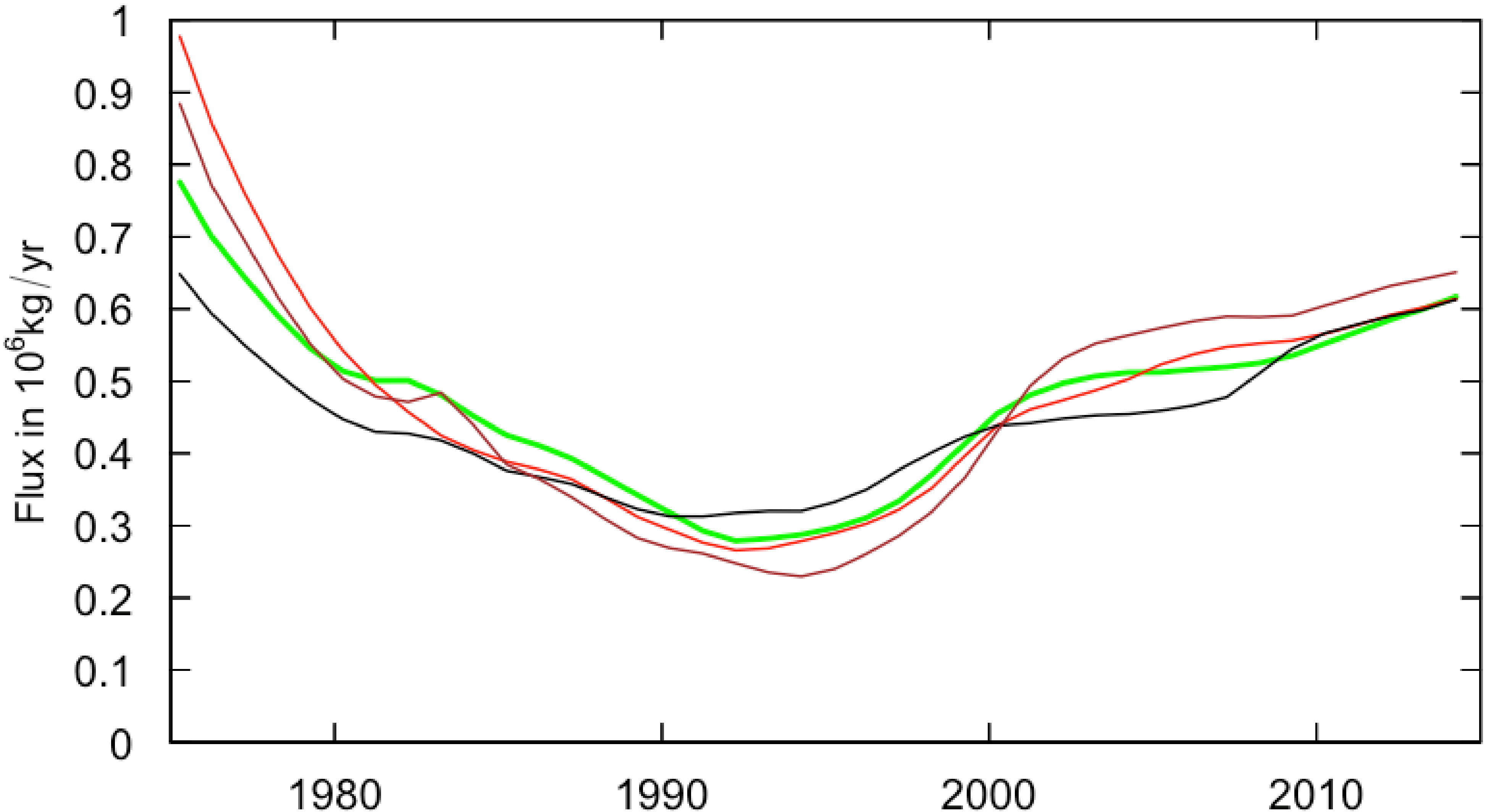
Maumee River at Waterville OH HU_SRP as P, mg/L
Water Year
Mean Concentration (dots) & Flow Normalized Concentration (line)
Concentration in mg/L
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1980
1990
2000
2010
Flux Estimates (dots) & Flow Normalized Flux (line)
Flux in 10^6 kg / yr
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1980
1990
2000
2010
USGS

Use a Bootstrap method to evaluate uncertainty

- Resample the data set, by 200 day blocks, with replacement
- Conduct the WRTDS estimation process for each replicate
- Uncertainty of the trend magnitude is determined from the set of bootstrap estimates for the selected trend period.

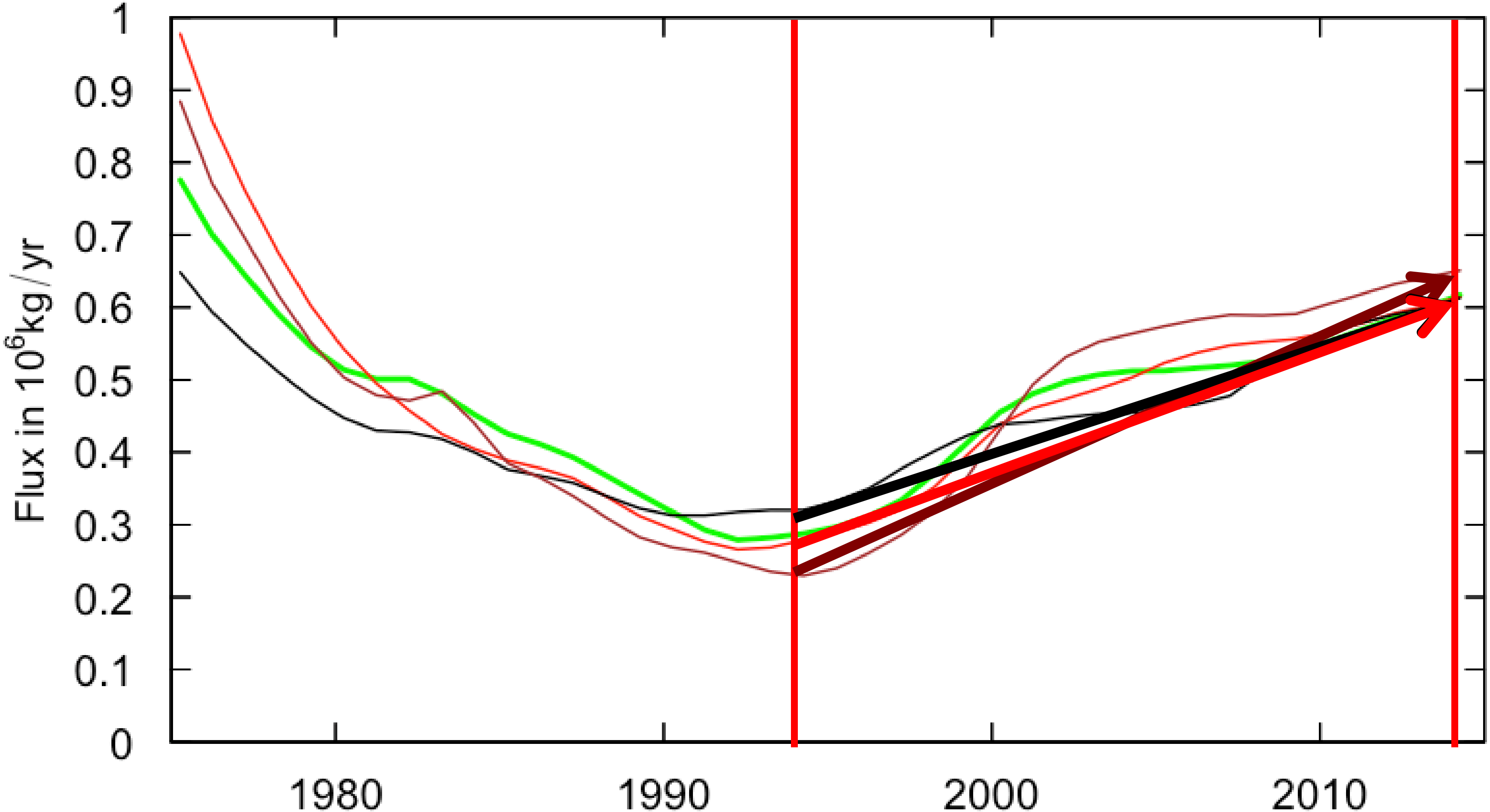
USGS

Maumee River, SRP - Green is WRTDS Flow Normalized Flux
Red, Brown and Black are three bootstrap replicate estimates of Flow Normalized Flux



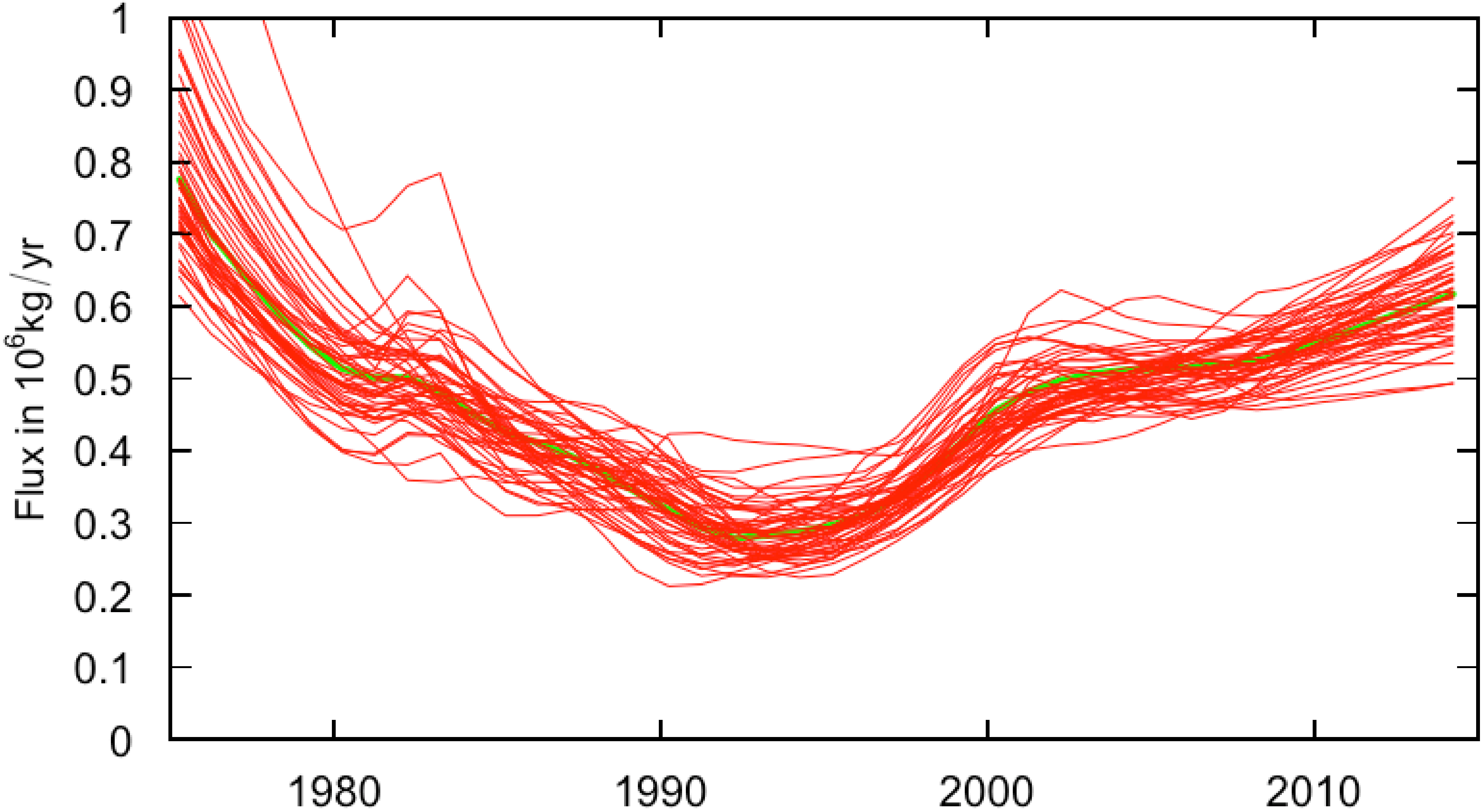
USGS

Each bootstrap replicate can give us an estimate of change between any two years (say 1994 and 2014)



USGS

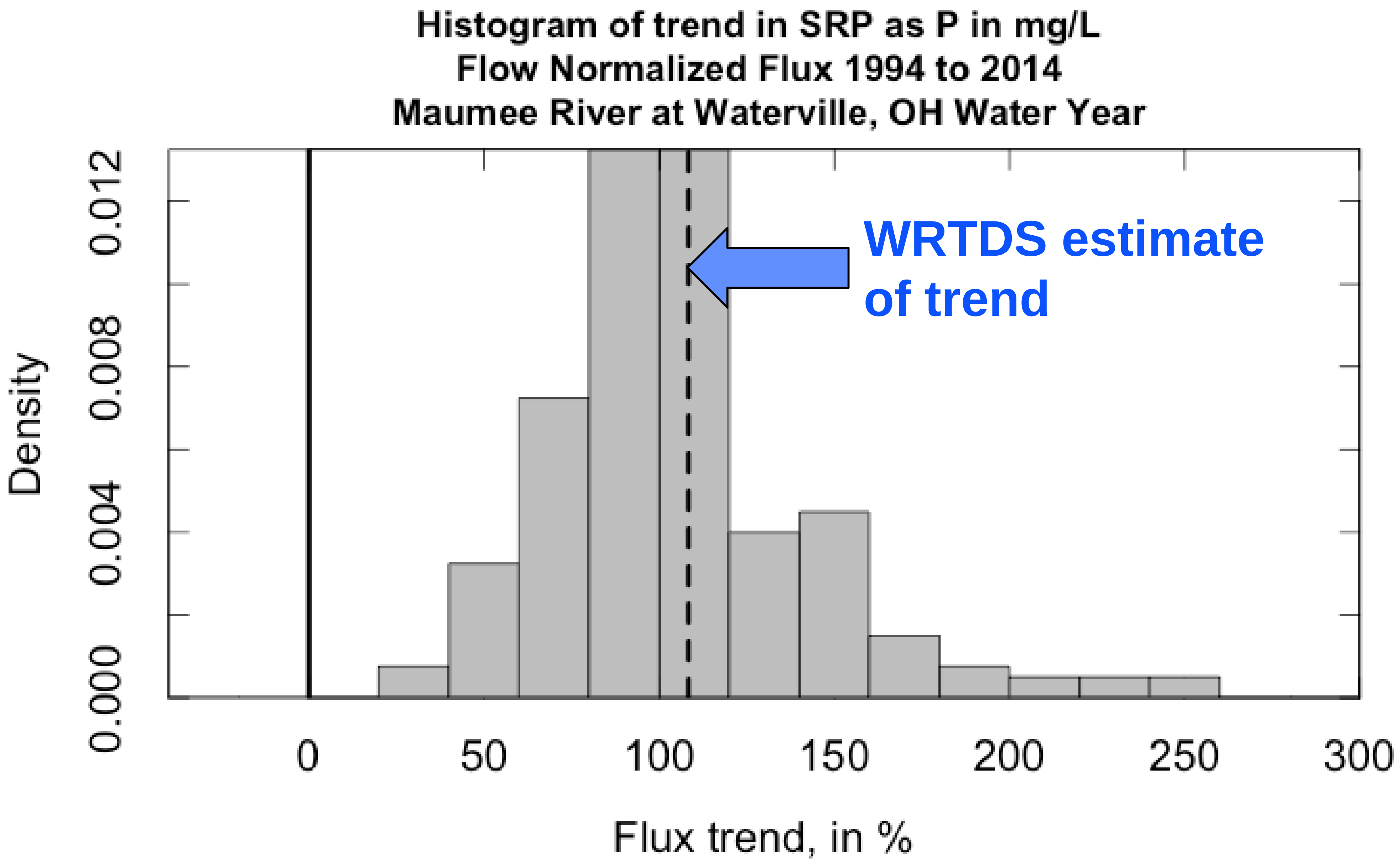
50 bootstrap replicates



Two ways to convey an answer to the question: Is there a trend?

- Conventional p-value approach (reject H_o or do not reject H_o)
- Describe the results in terms of "likelihood of uptrend" or "likelihood of downtrend"

USGS



USGS

Upward trend highly likely

The EGRETci software translates the bootstrap results into a set of words

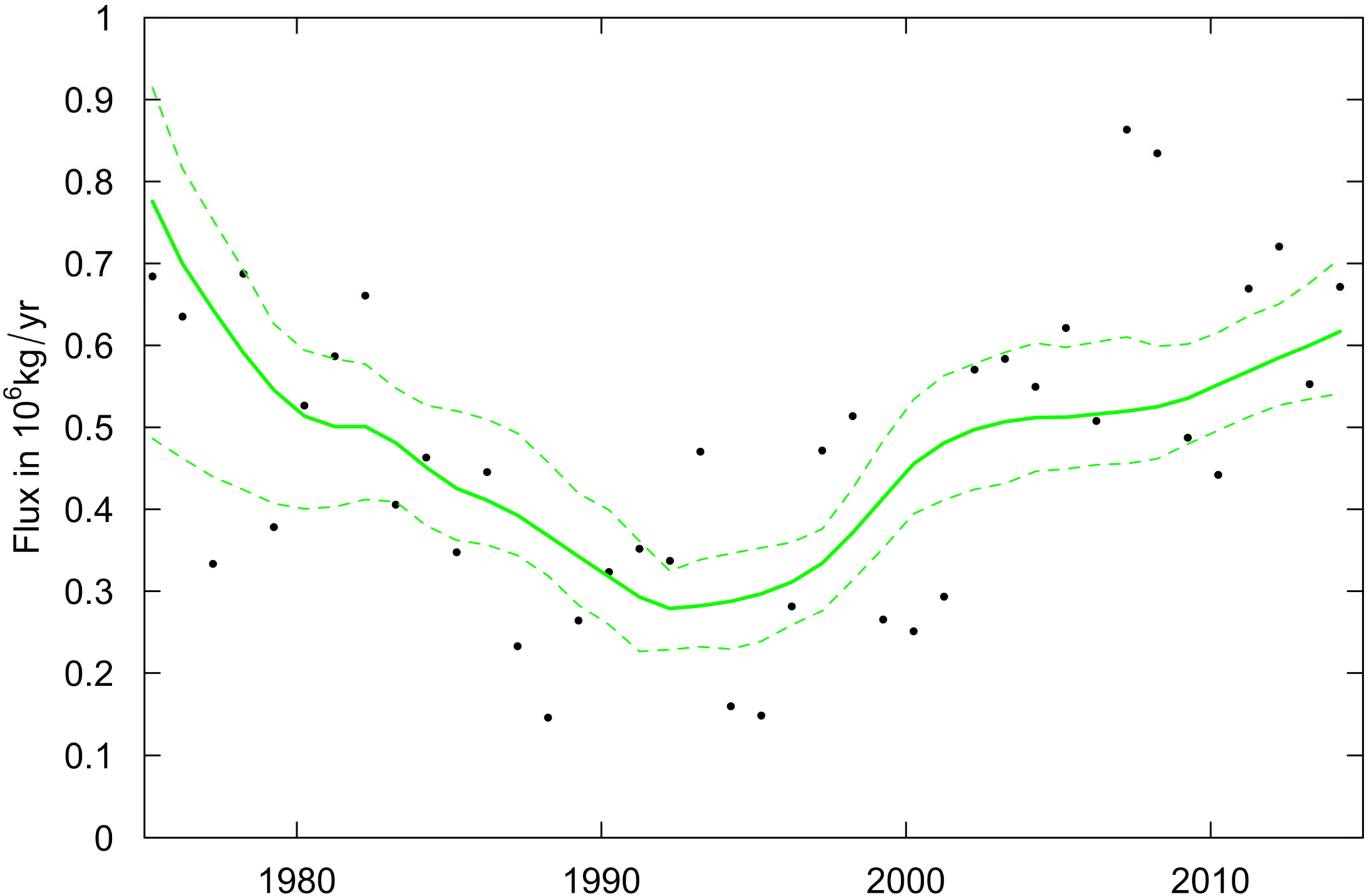
Frequency of upwards trends in the bootstrap replicates	Likelihood words
(0.95, 1.0)	Upward trend is highly likely
(0.90, 0.95)	Upward trend is very likely
(0.67, 0.90)	Upward trend is likely
(0.33, 0.67)	Upward trend is about as likely as not
(0.10,0.33)	Upward trend is unlikely
(0.05, 0.1)	Upward trend is very unlikely
(0.0, 0.05)	Upward trend is highly unlikely

USGS

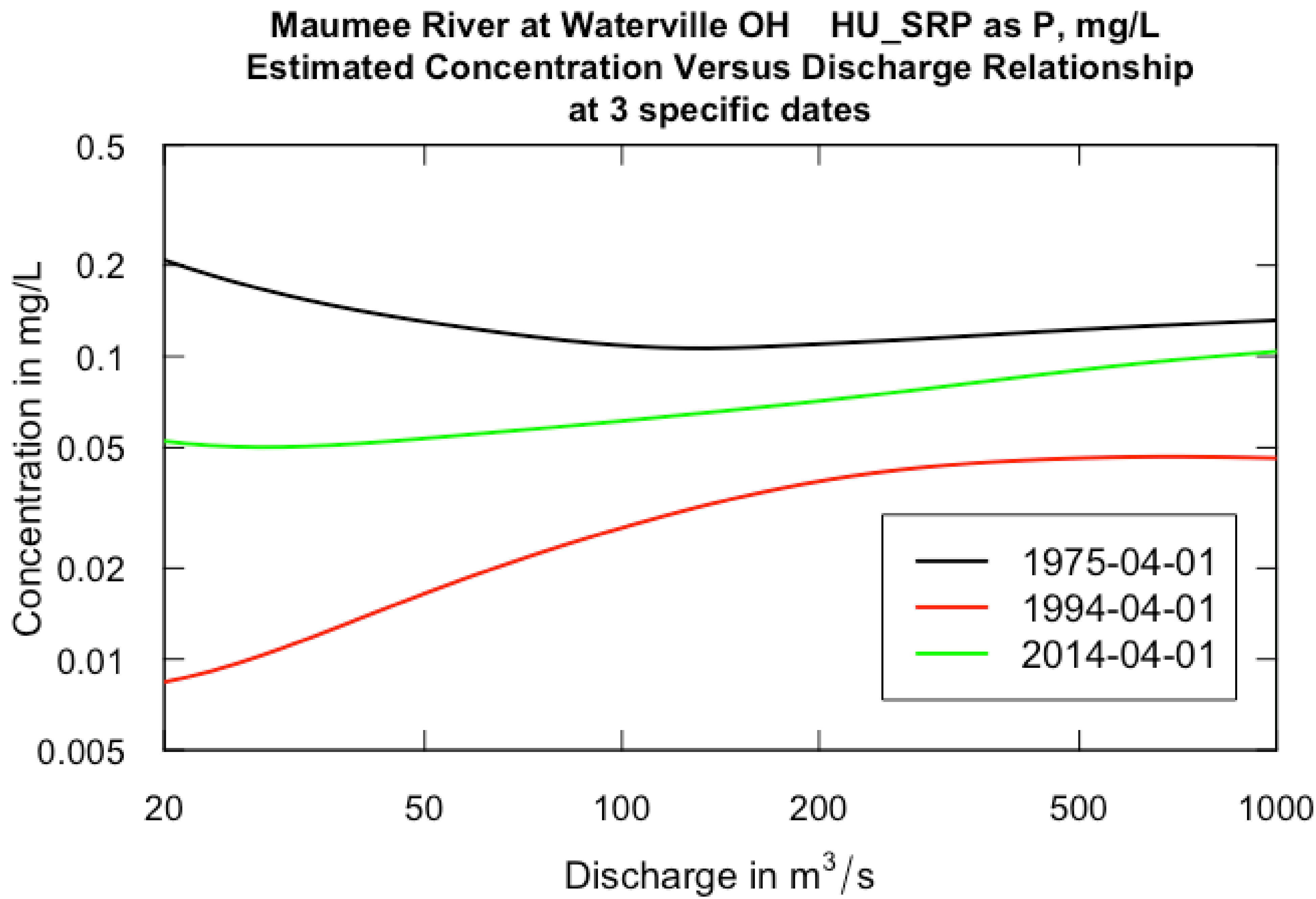
The EGRETci package can also give us confidence intervals

- Various confidence intervals for the change over a specific time
- Graphical confidence intervals for the entire period of record

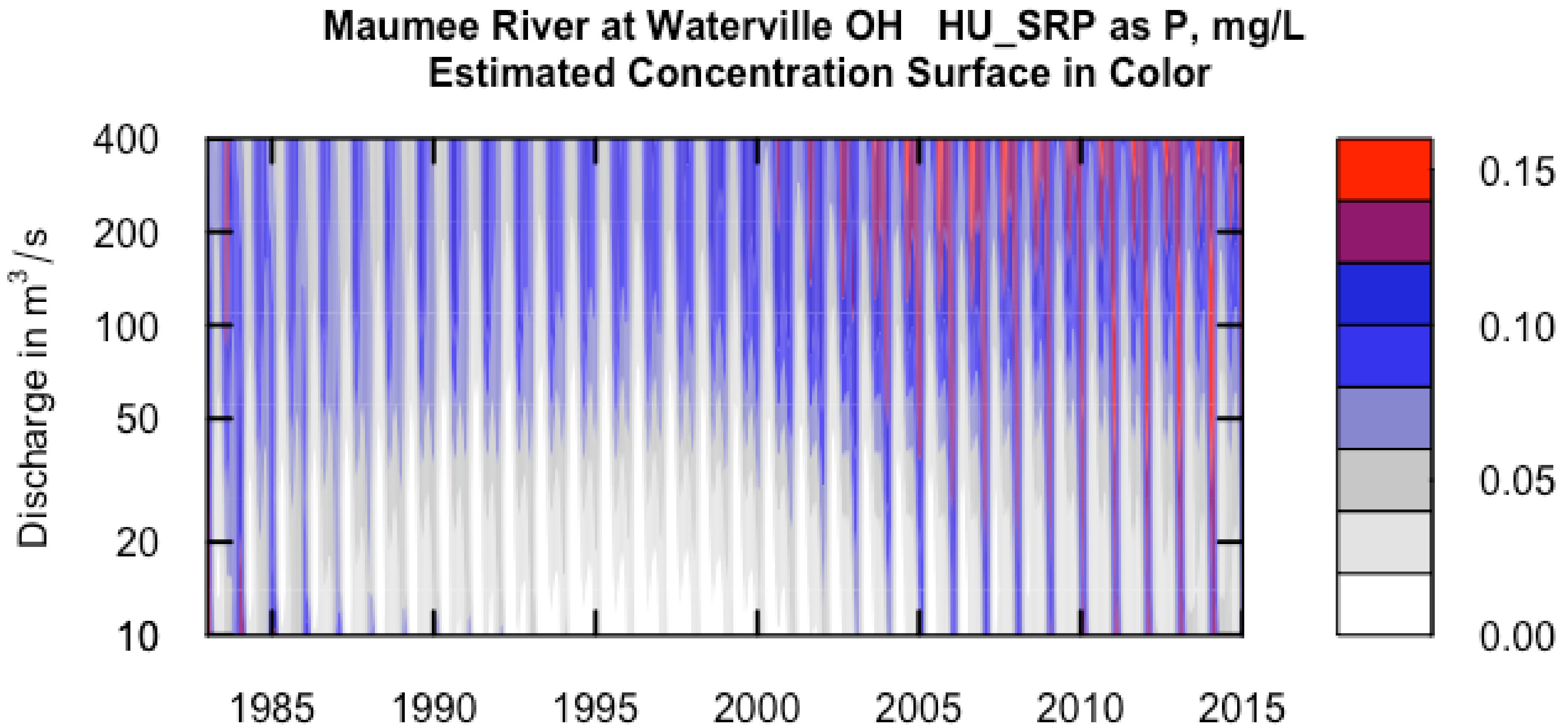
Maumee River, SRP Flux
90% Confidence Intervals, based on 200 bootstrap replicates

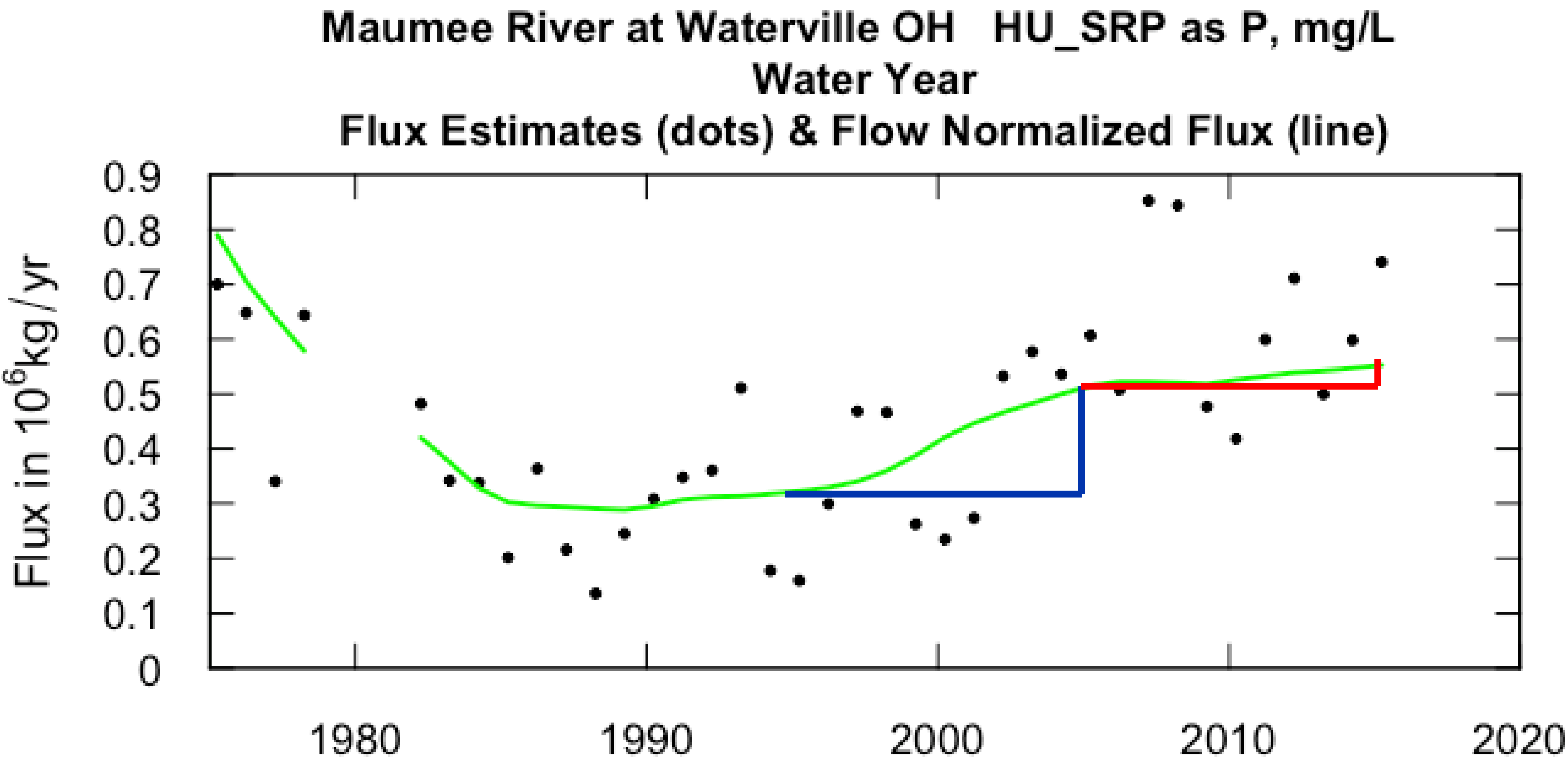


```
plotConcQSmooth(eList,"1975-04-01","1994-04-01","2014-04-01",qLow=20,qHigh=1000,qUnit=2,logScale=TRUE,legendLeft=250,legendTop=0.03)
```



```
plotContours(eList, yearStart=1983, yearEnd=2015, qBottom=10, qTop=400,
contourLevels=seq(0,0.16,0.02), flowDuration=FALSE)
```

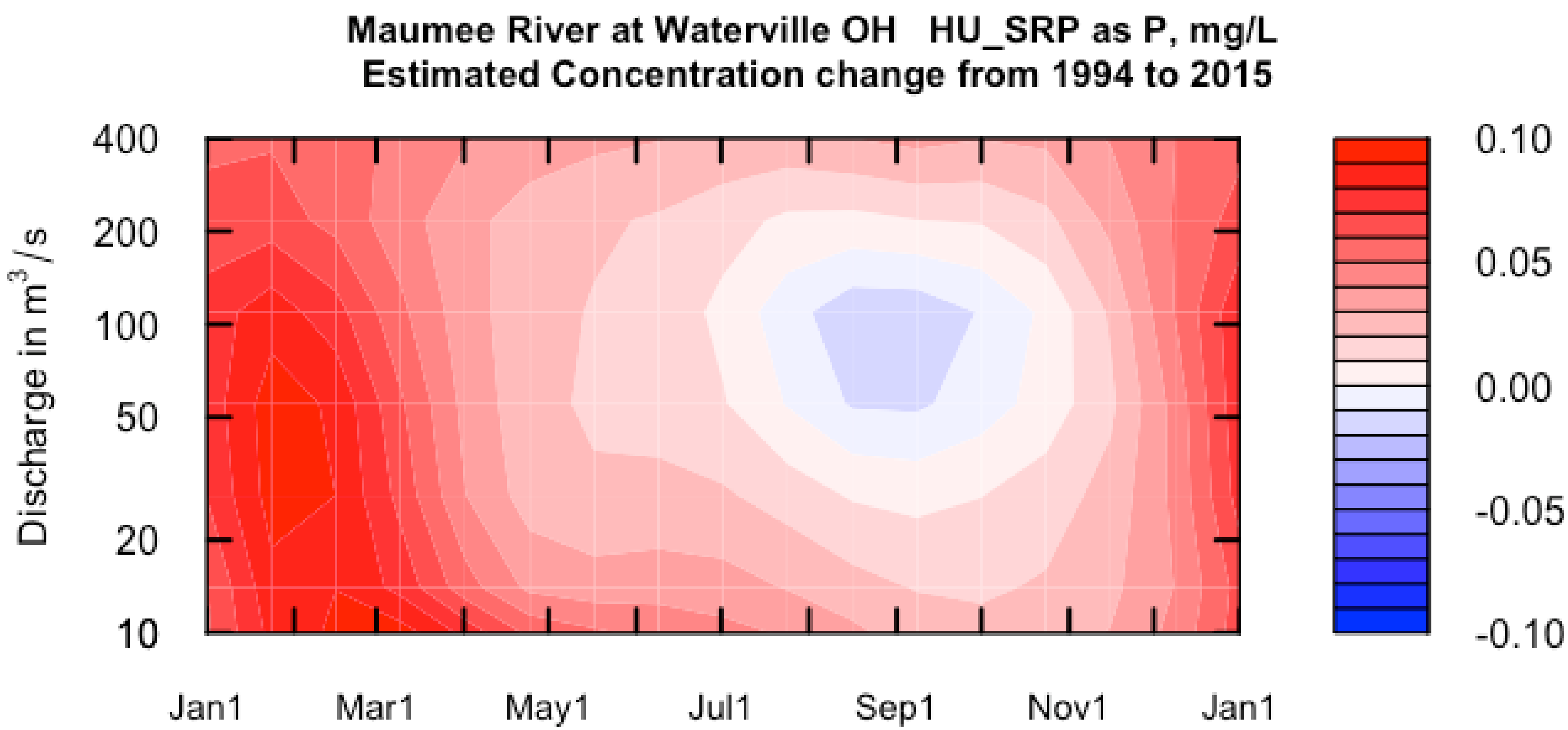




Trend from 1995 to 2005 = +6% per year

Trend from 2005 to 2015 = +0.7% per year

```
plotDiffContours(eList, year0  =1994, year1 = 2015, qBottom = 10, qTop = 400,
maxDiff = 0.1, flowDuration = FALSE)
```



```
> plotFluxHist(eList)


> tableChange(eList,yearPoints=c(1995,2005,2015))

   Maumee River at Waterville OH
   HU_SRP as P, mg/L
   Water Year

           Concentration trends
   time span       change     slope    change     slope
                     mg/L   mg/L/yr        %        %/yr

 1995  to  2005     0.025    0.0025       56       5.6
 1995  to  2015     0.036    0.0018       80         4
 2005  to  2015      0.01     0.001       15       1.5



                 Flux Trends
   time span            change         slope          change          slope
                  10^6 kg/yr    10^6 kg/yr /yr         %              %/yr
 1995  to  2005           0.19          0.019            60               6
 1995  to  2015           0.23          0.011            71             3.6
 2005  to  2015          0.038         0.0038           7.4            0.74
```

Anticipated enhancements to WRTDS and EGRET package

- **Dealing with ephemeral streams**
- **Estimation of trends in frequency of exceedances of threshold values**
- **Dealing with nonstationarity in Q**
- **Improved estimates of yearly fluxes**
- ***Users ideas?***

USGS

Information about EGRET

- https://github.com/USGS-R/EGRET/wiki

“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.”

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

rhirsch@usgs.gov

USGS