# Overview of WRTDS and the EGRET and EGRETci Packages

WRTDS = Weighted Regressions on Time, Discharge and Season
EGRET = Exploration and Graphics for RivEr Trends

EGRETci = Confidence Intervals for EGRET

Robert M. Hirsch
USGS
2016-05-02

# These are two R-packages

Free, quality assured, open-source, platform independent, documented, and based on methods that appear in peer reviewed journal articles

EGRET is for data exploration: it depends on dataRetrieval and it implements WRTDS

EGRETci is for uncertainty analysis for WRTDS results. It depends on EGRET

USGS

# Outline of the presentation

Motivations for the packages

The WRTDS concept

How EGRET works

How EGRETci works

USGS

# Motivation for EGRET: 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) Produce graphs and tables

USGS

# Guiding ideas for WRTDS

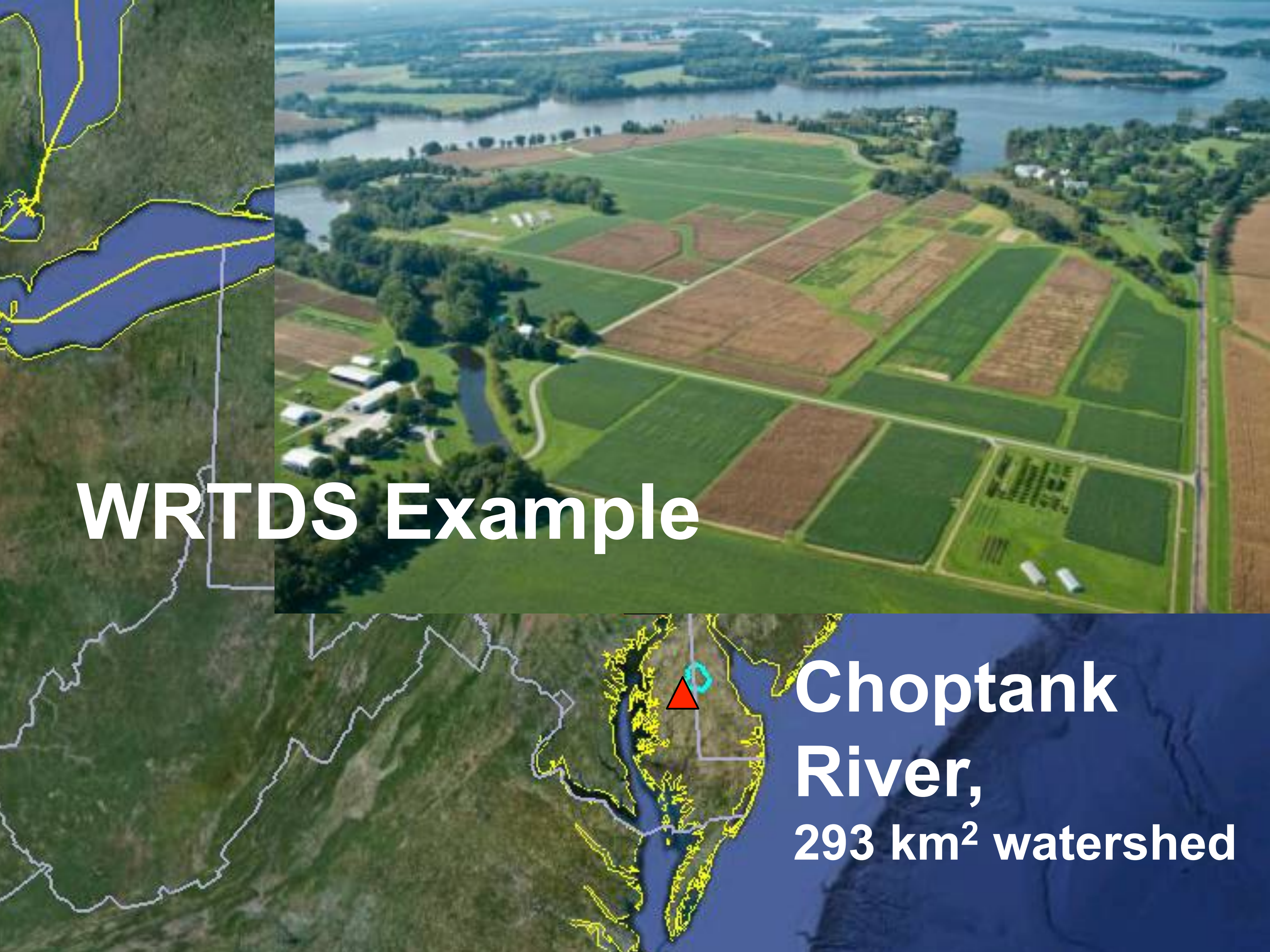
- Describe the evolving behavior of the watershed. No mathematical straight-jacket!!
- Estimate both concentration & flux (averages as well as trends).
- Estimate the actual history but also a flow-normalized history.
- Resolve a serious bias in flux estimates.
- Be quantitative but also exploratory.

USGS

# Data requirements

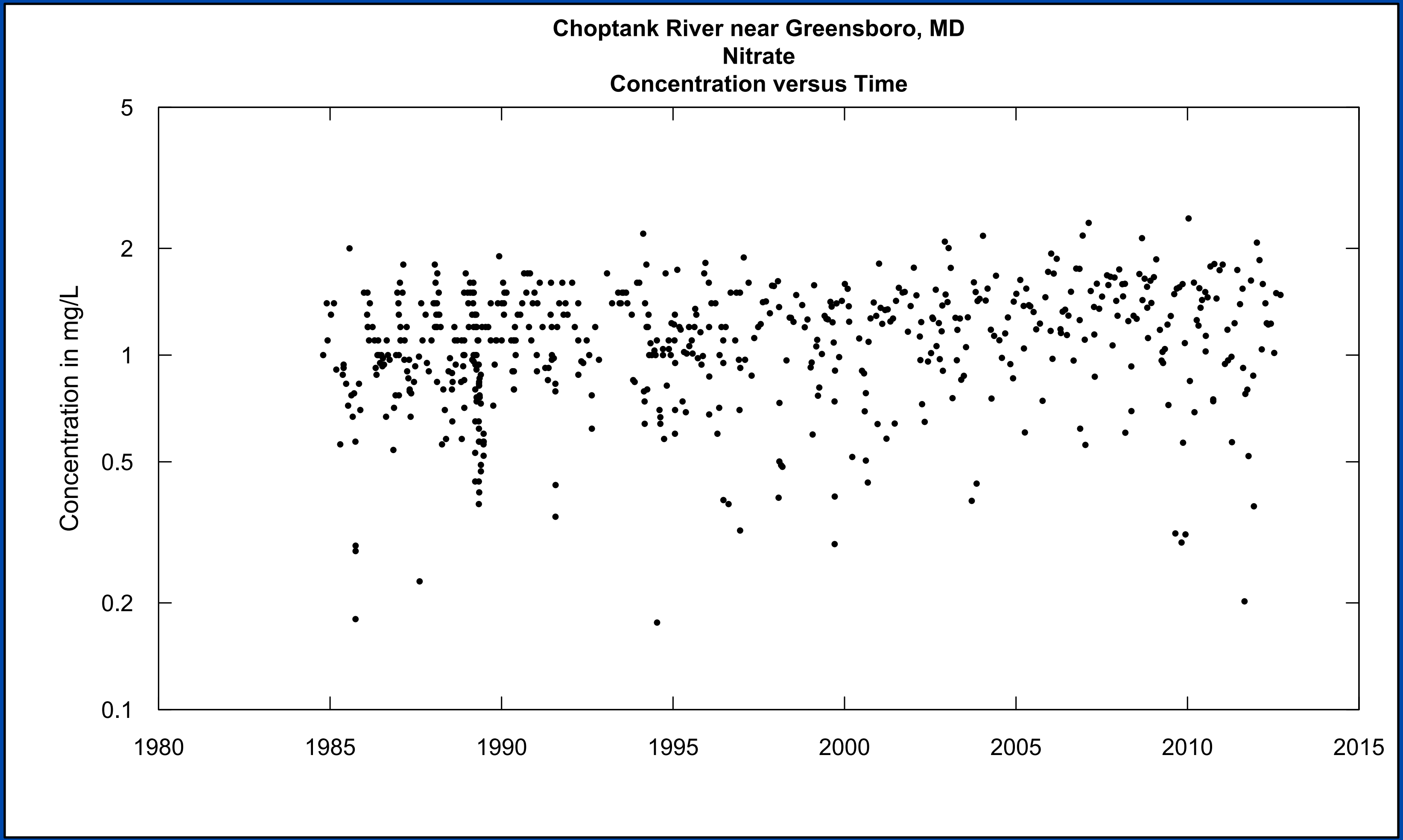
- Low intra-day variability (not flashy)
- Requires a complete daily discharge record
- Intended for >200 samples, but has been used for some purposes with as few as 60 samples
- Water quality samples cover most of the discharge range
- For trend studies: 20+ years, but can do less
- For average flux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

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




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

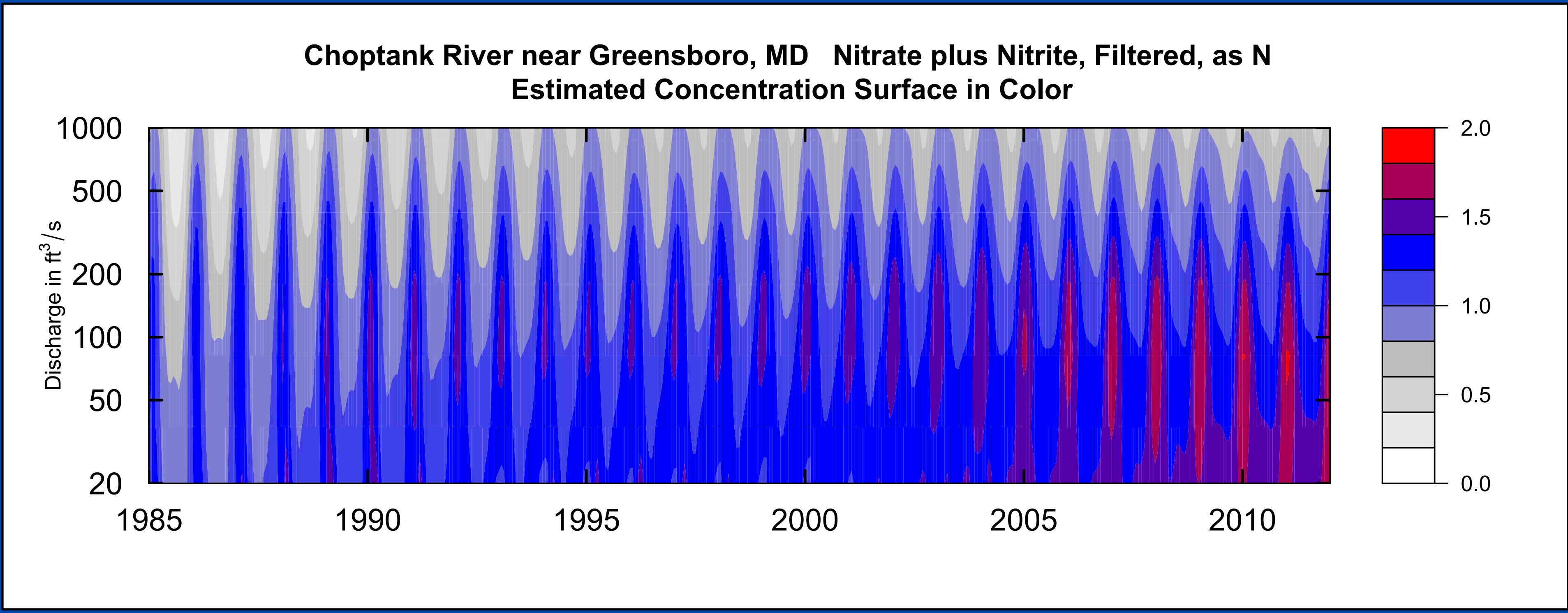
**For any location in time - discharge space (*t* and *Q*) we assume that concentration (*c*) follows this model**

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

**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

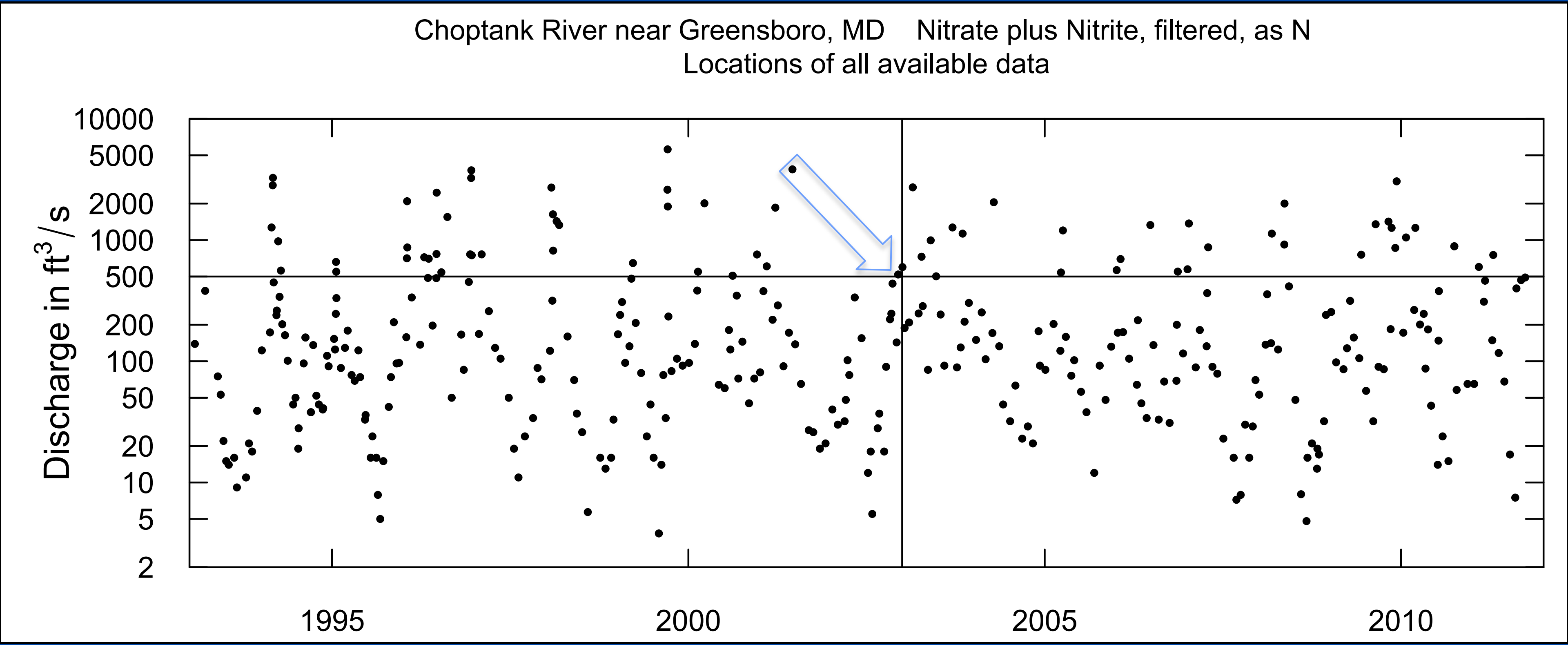



## How is this surface created?

USGS

# Every dot is a data point from 1993 to 2012

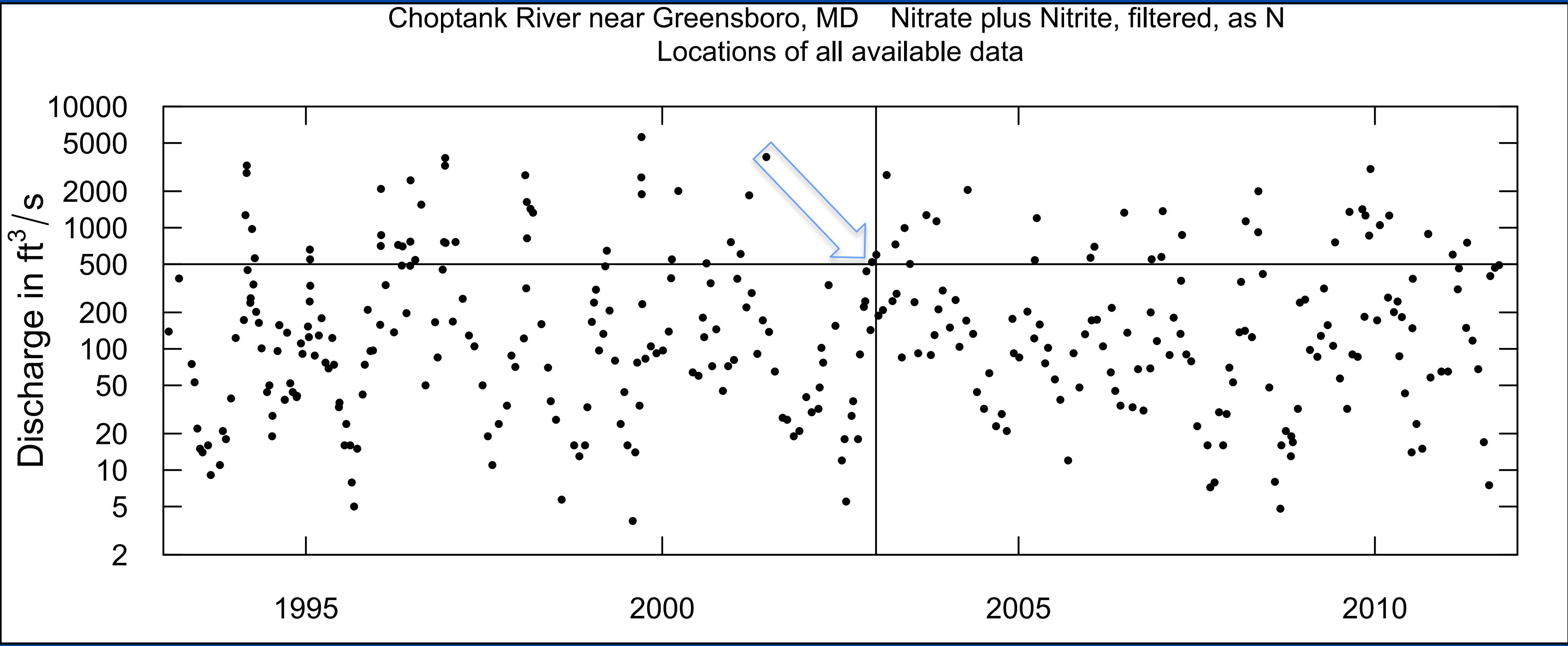
# Let's say we want to use the data to estimate the expected value of concentration for January 1, 2003 at Q=500 cfs




USGS

**The principle is this:**

**Do a weighted regression at this point. The weights on each observation are related to their "distance" from Jan 1 2015 at 500 cfs**



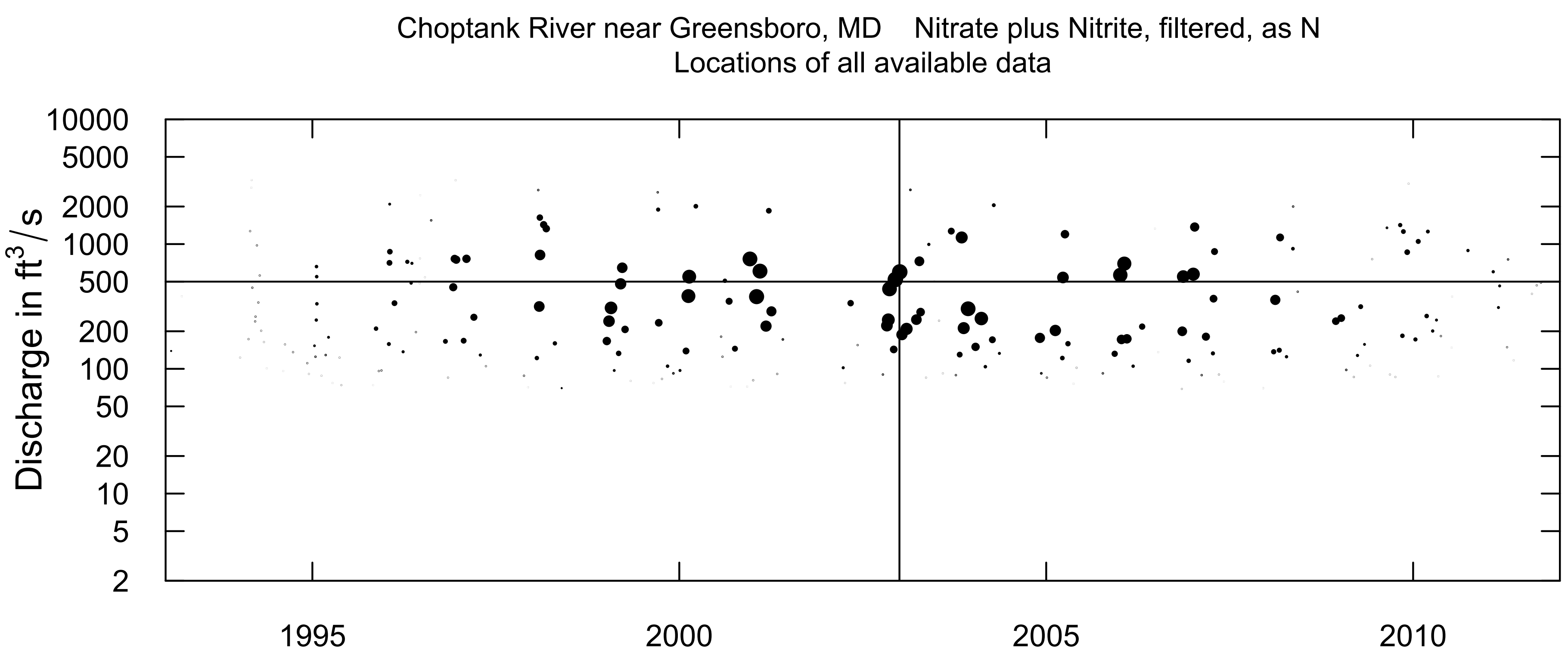

**Distance in time, in log(Q), and season.**

**Now move to the next point and do it all over again.**

USGS

How do we set the weights for the regression?

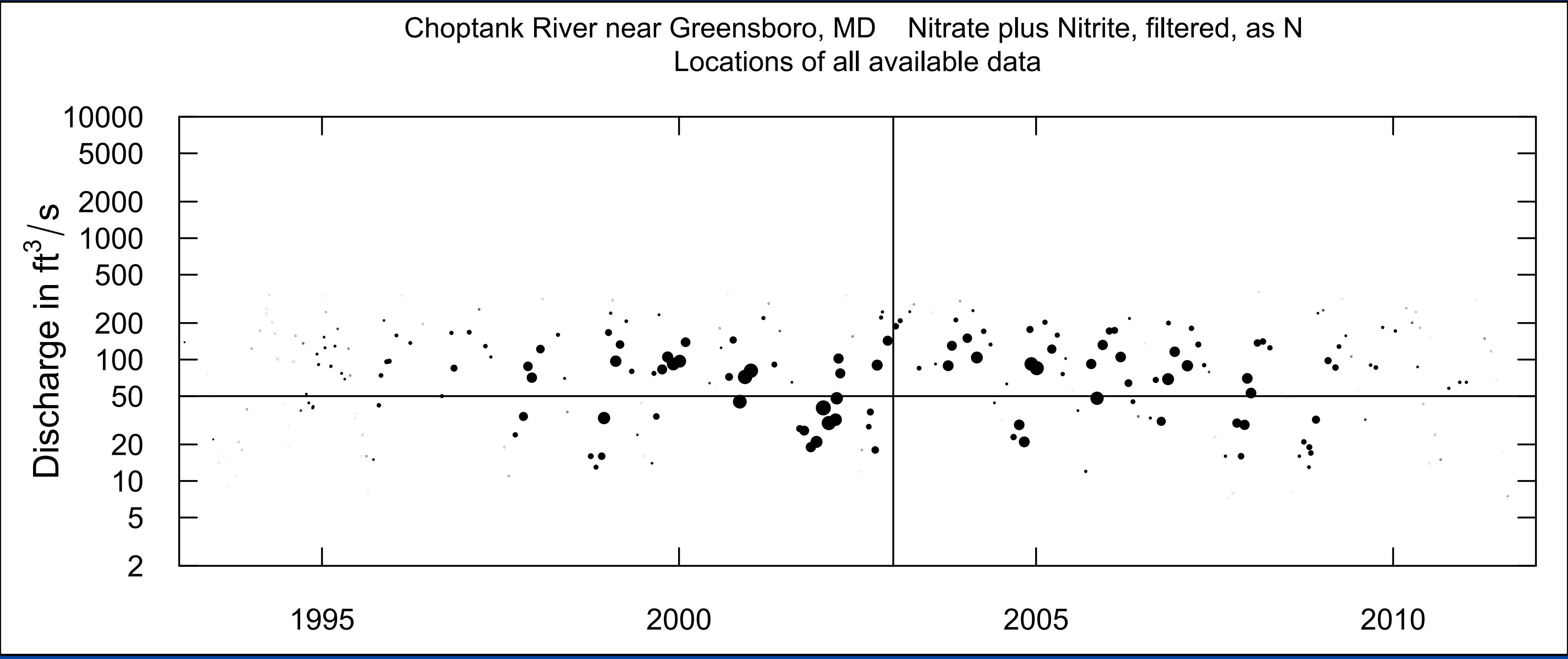
- These are the same points we just saw, but the radius of the dot is proportional to weight assigned to that point for purposes of estimating concentration for January 1, 2003 at Q=500 cfs
- The weight depends on distance in: time, log discharge, and season from January 1, 2003 at Q = 500 cfs




USGS

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

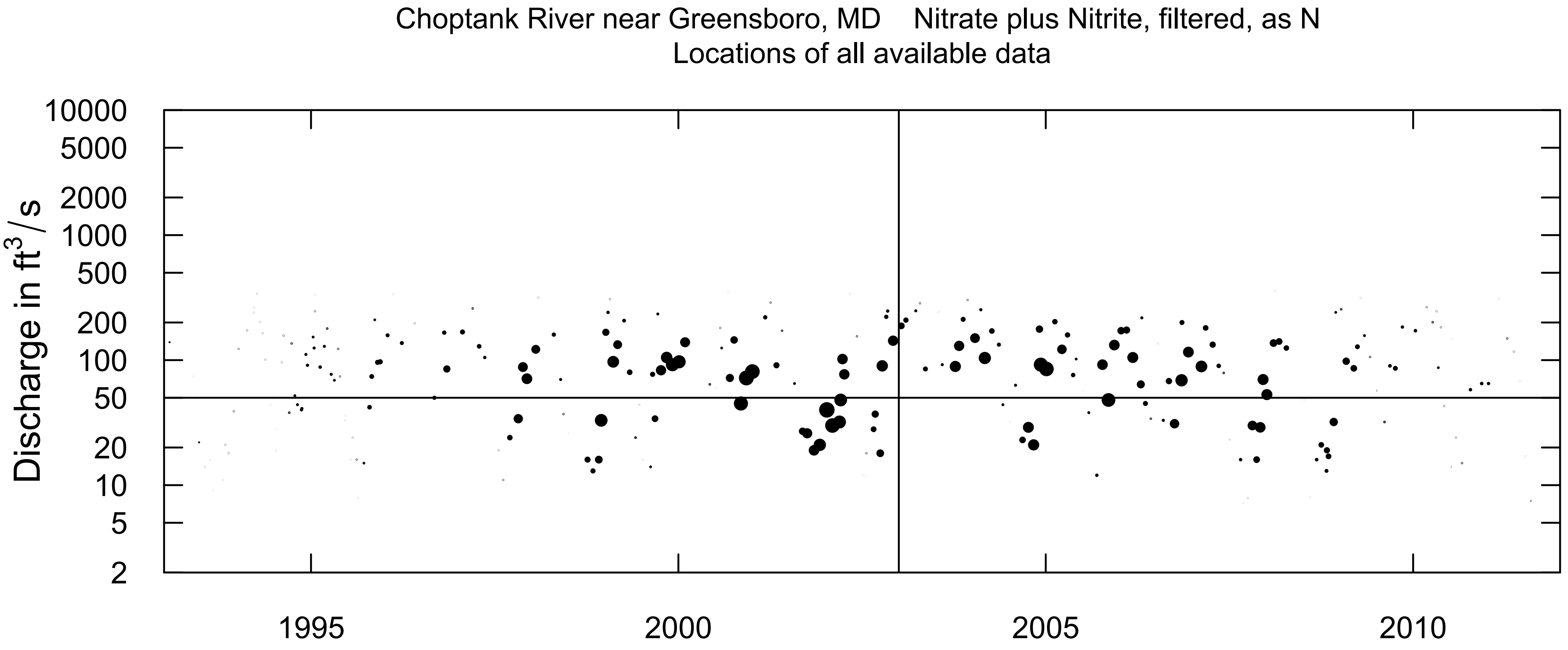
# Redo the weights for distance from that point




USGS

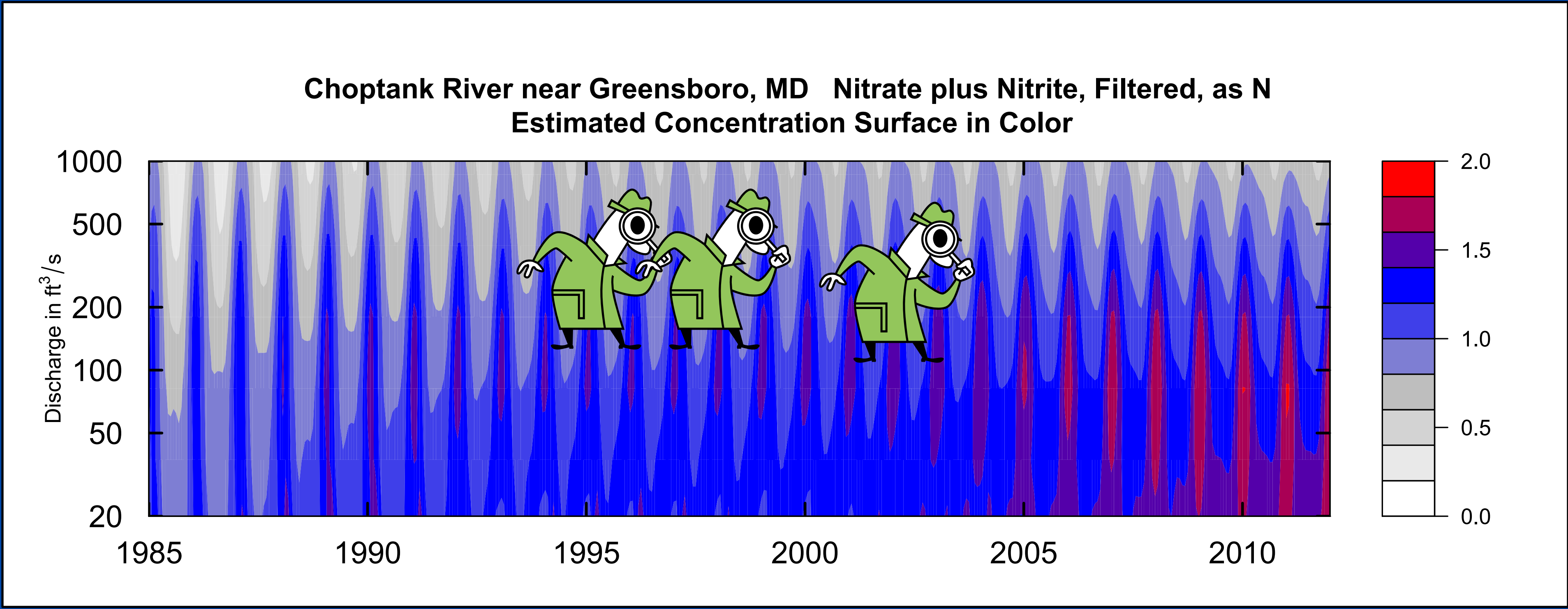
# To organize the work, lets make estimates for a fine mesh of points in this space.

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




USGS

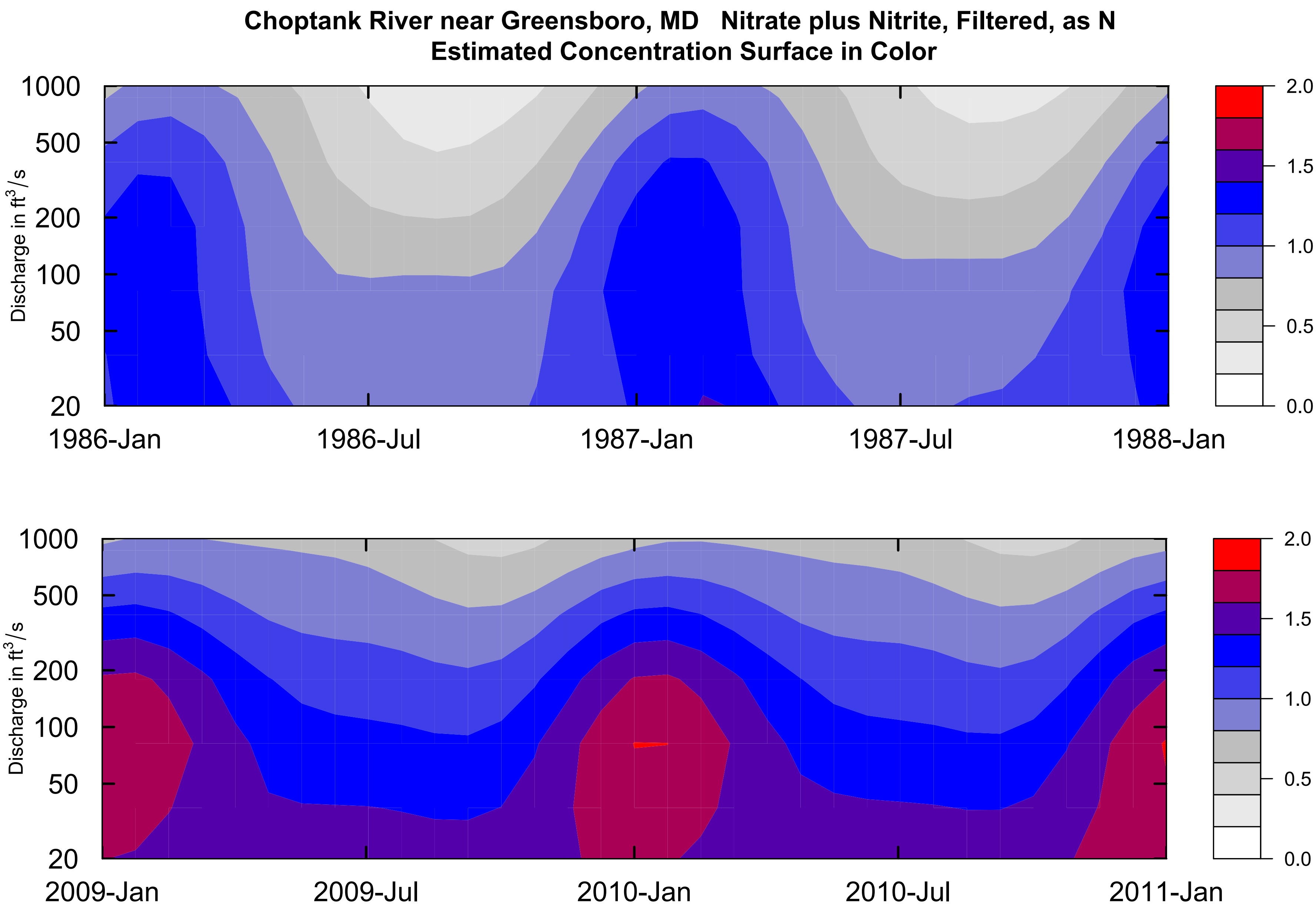
# This kind of weighted regression gets done about 6000 times to form this whole surface!!




**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.**

USGS

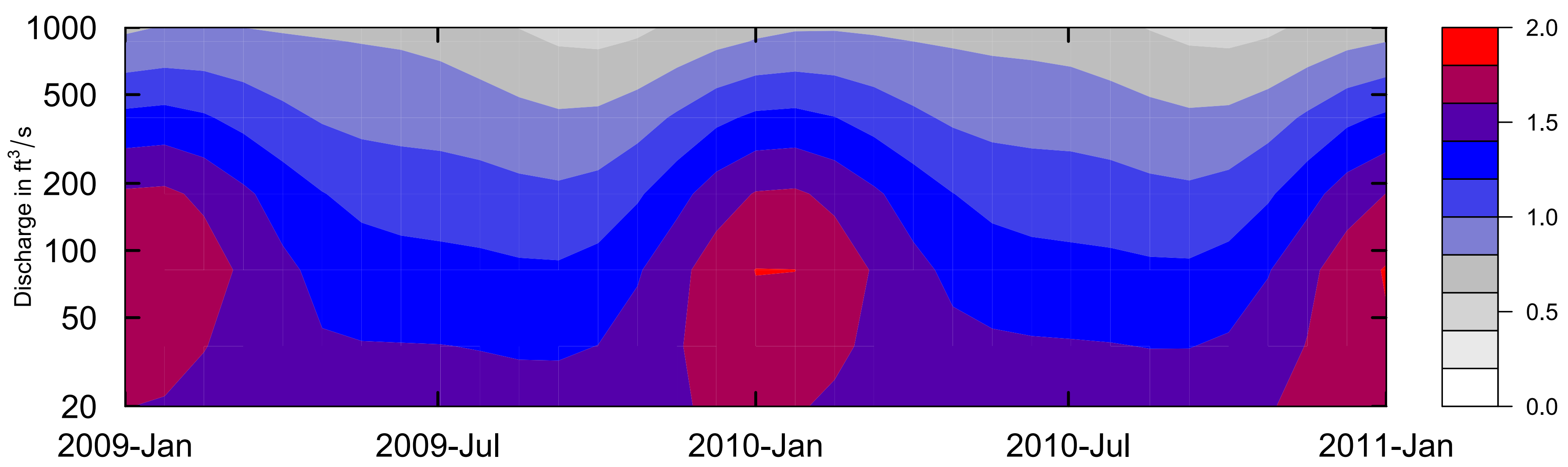
# Here are two, more detailed looks at this surface



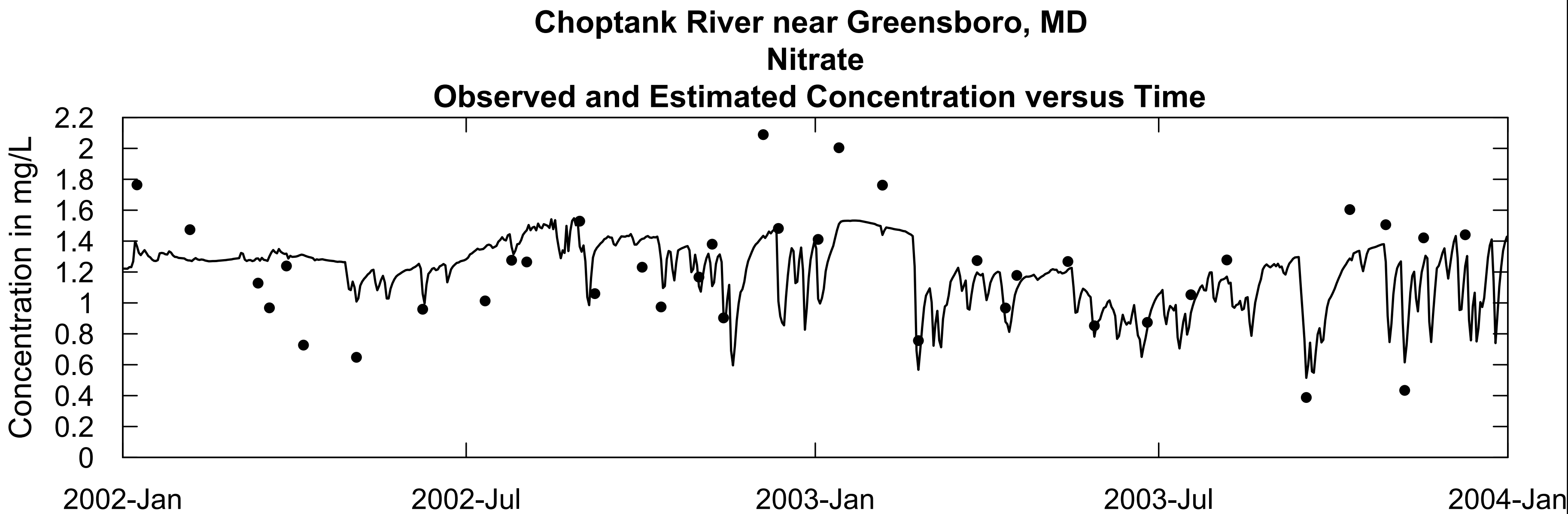
Now, for every one of 10,227 days in the record from 1985 through 2012:

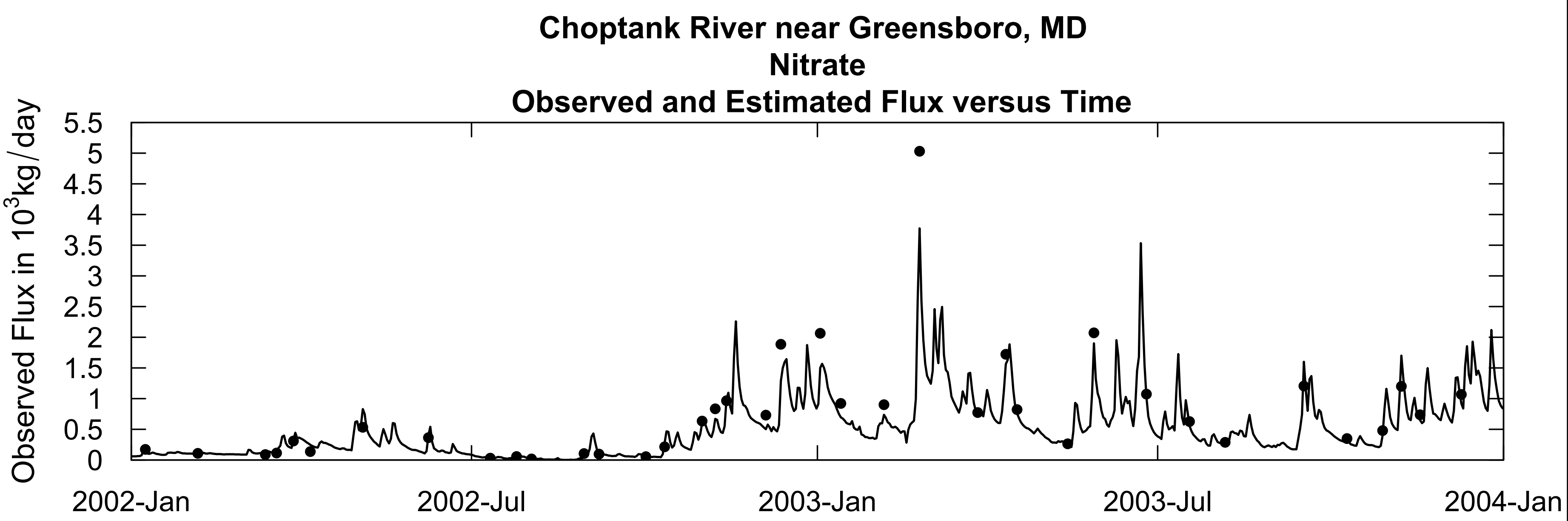
We can use the date and the observed discharge to compute the expected value of concentration.

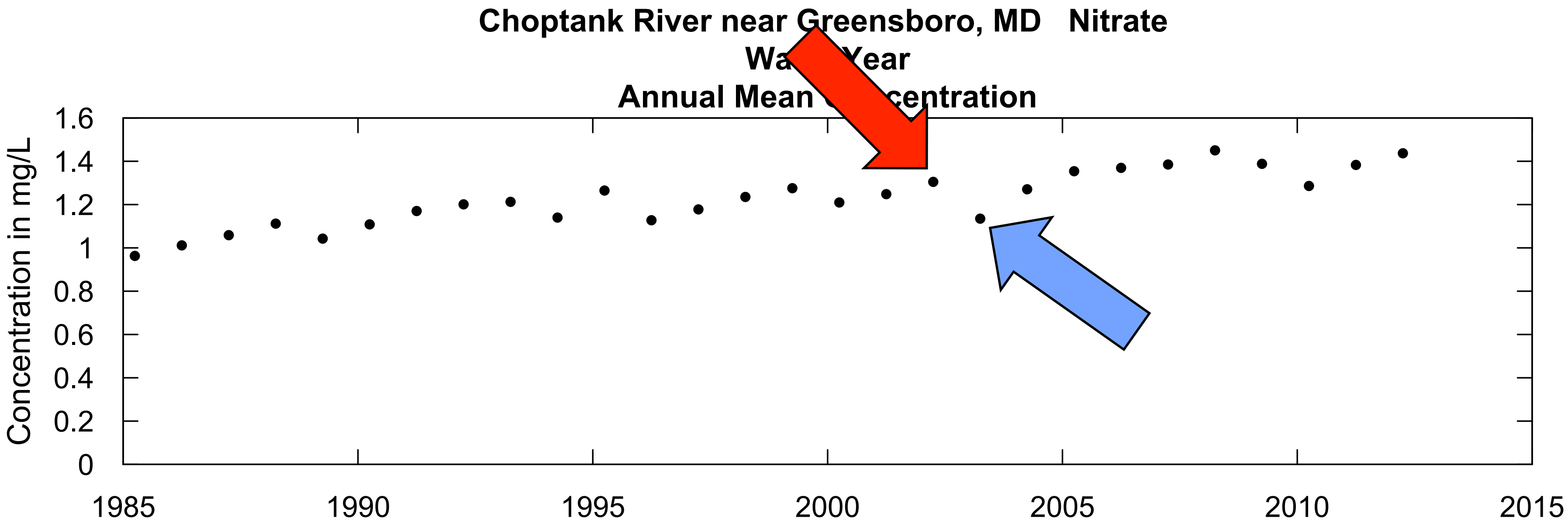
From that value we can compute the expected value of flux.

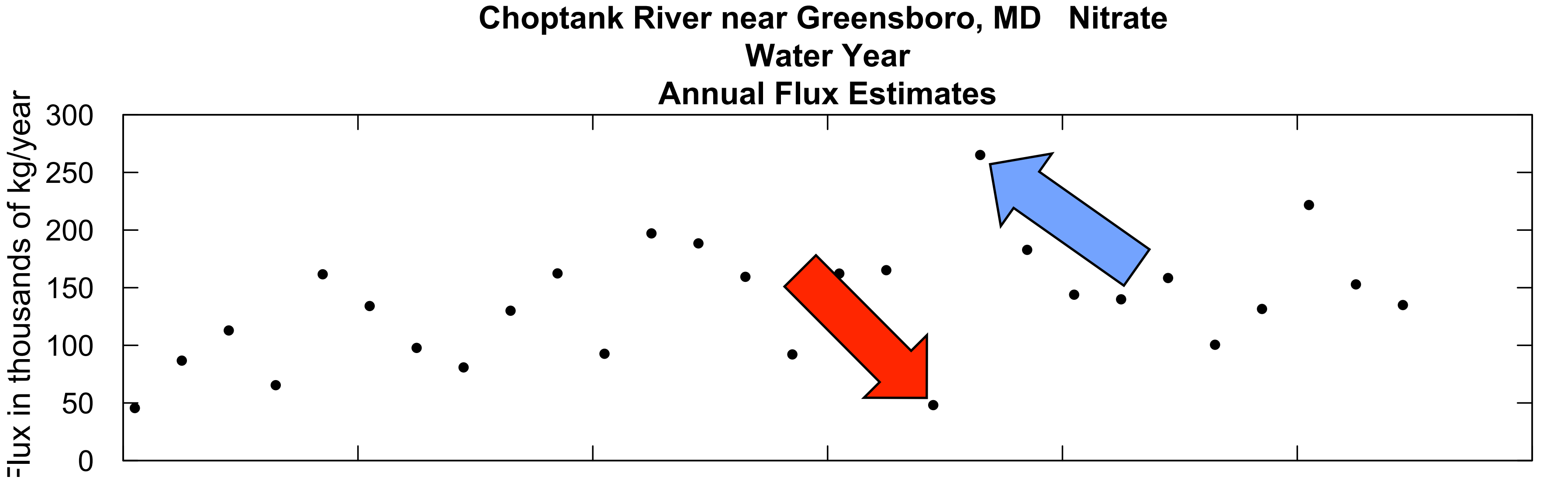


## Then we can sum these estimates by year to compute estimates of annual mean concentration & annual mean flux

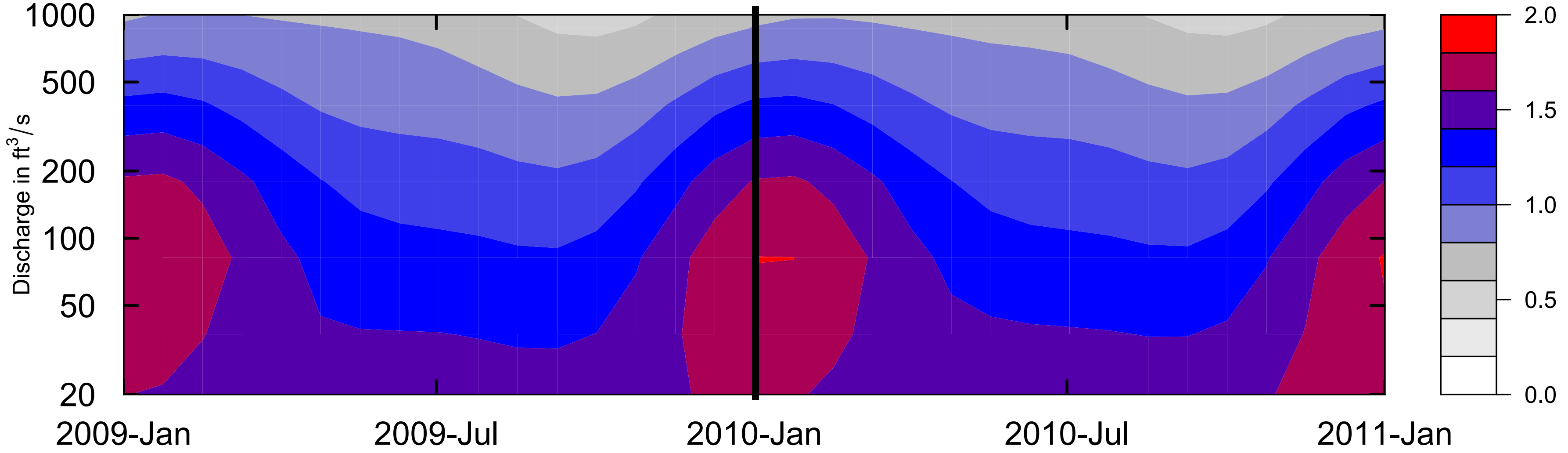




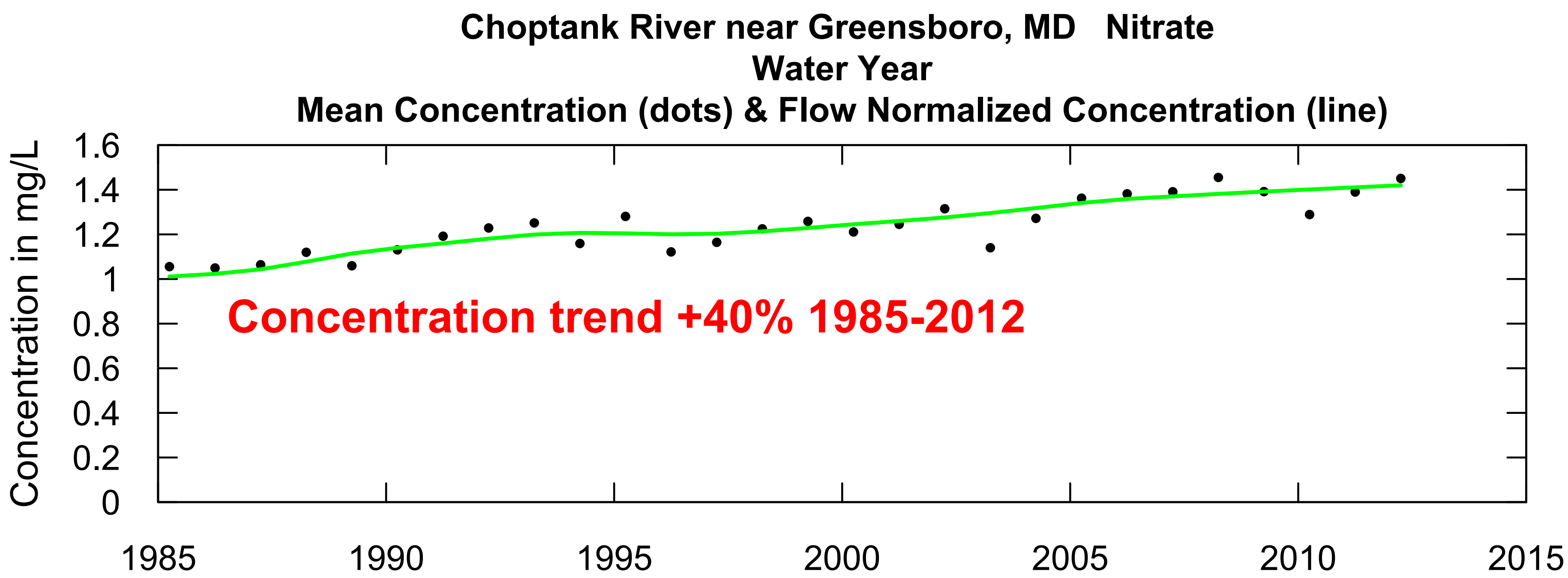

**Can we filter out this flow-driven variation to see the underlying change?**

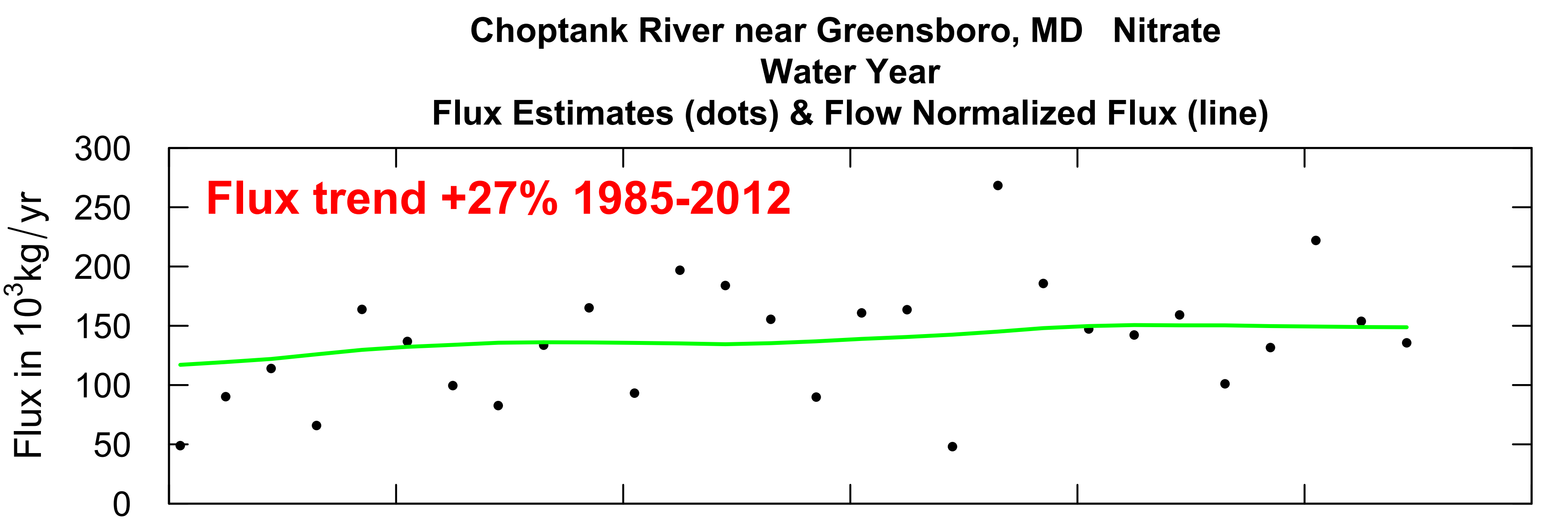



**The “flow normalized concentration” on any given day is: c=f(Q,T) integrated over the probability distribution of Q for that day of the year.**

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

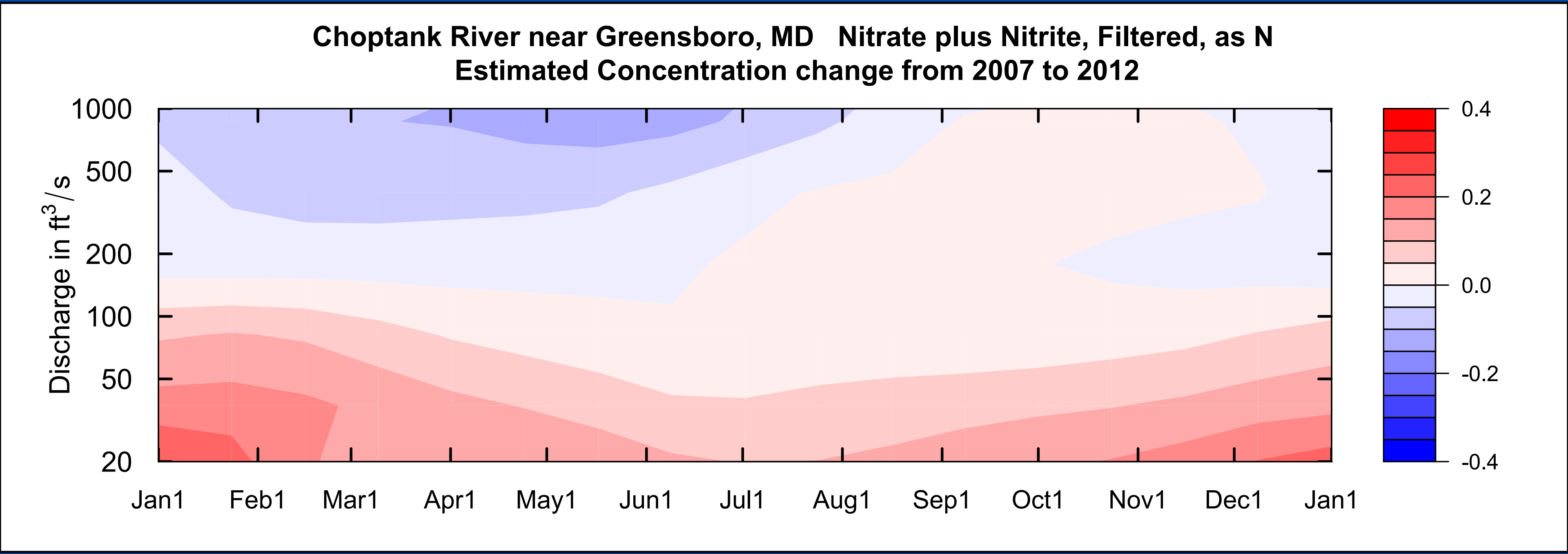
**Sum those over the year to get annual flow-normalized mean concentration and flux.**


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 10^3 kg/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.**

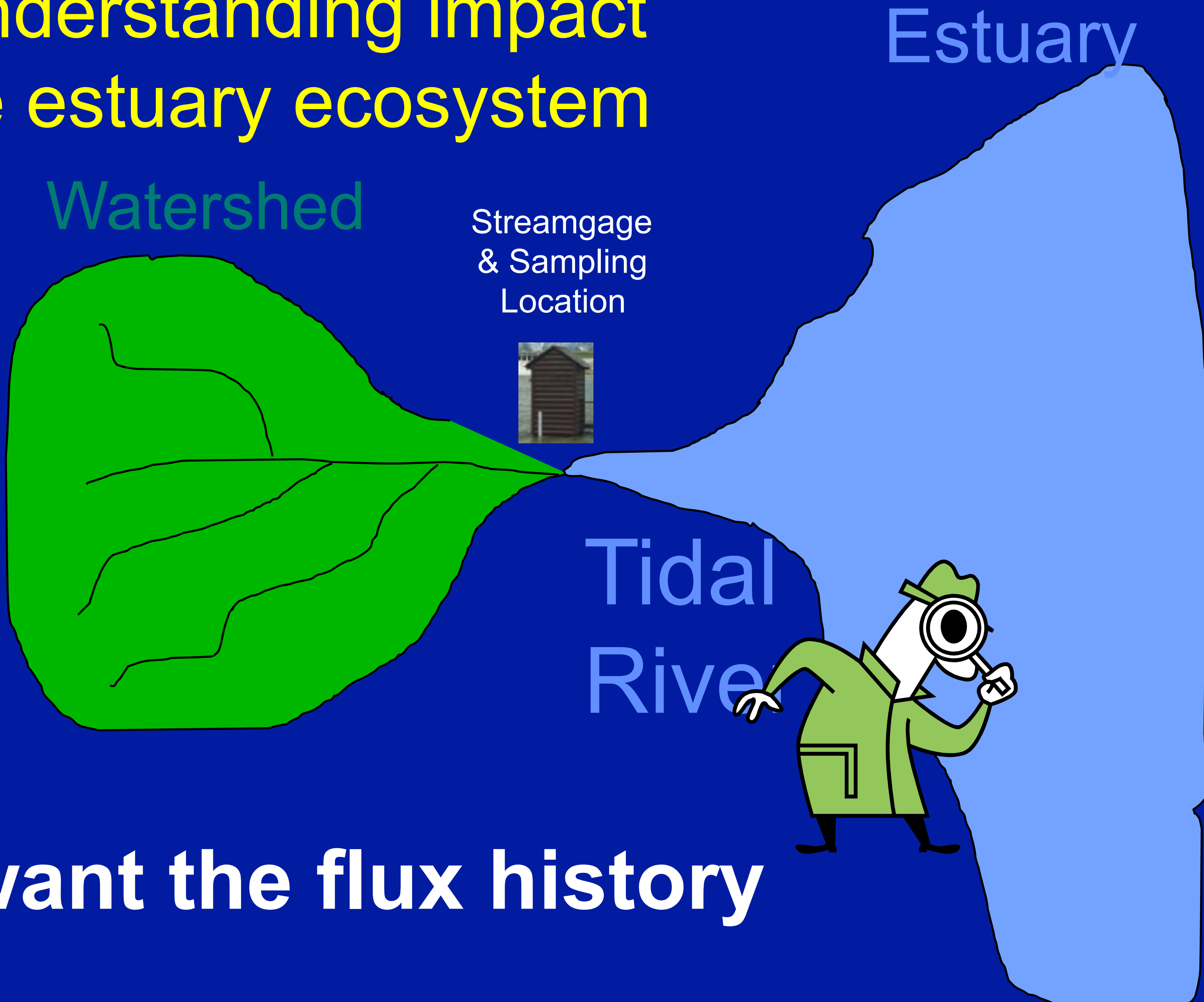
USGS

# Why all this complexity?

Different products for different purposes

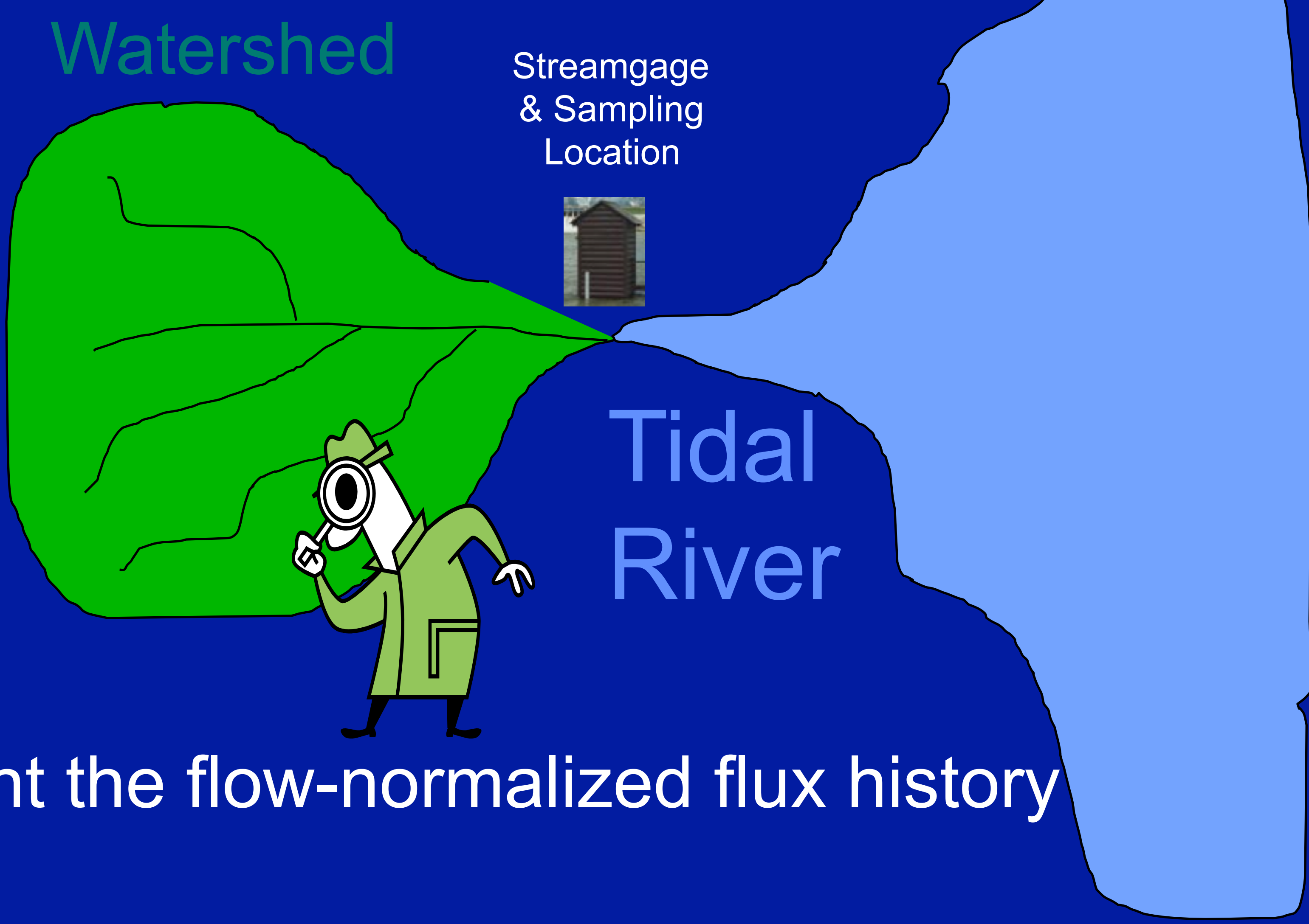
- Concentration versus flux
- Actual history versus flow-normalized history

# For understanding impact on the estuary ecosystem



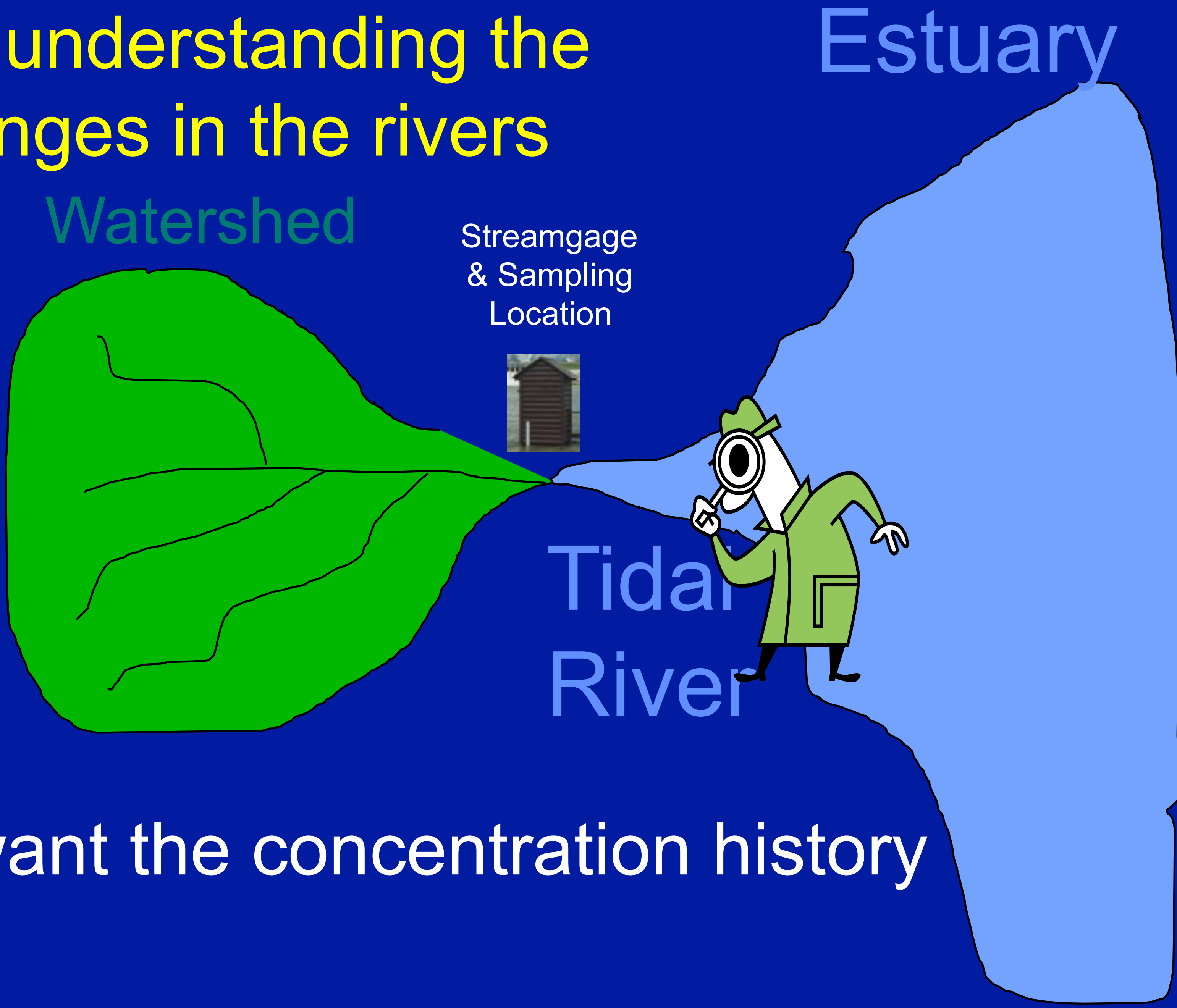

**We want the flux history**

# For understanding progress in the watershed



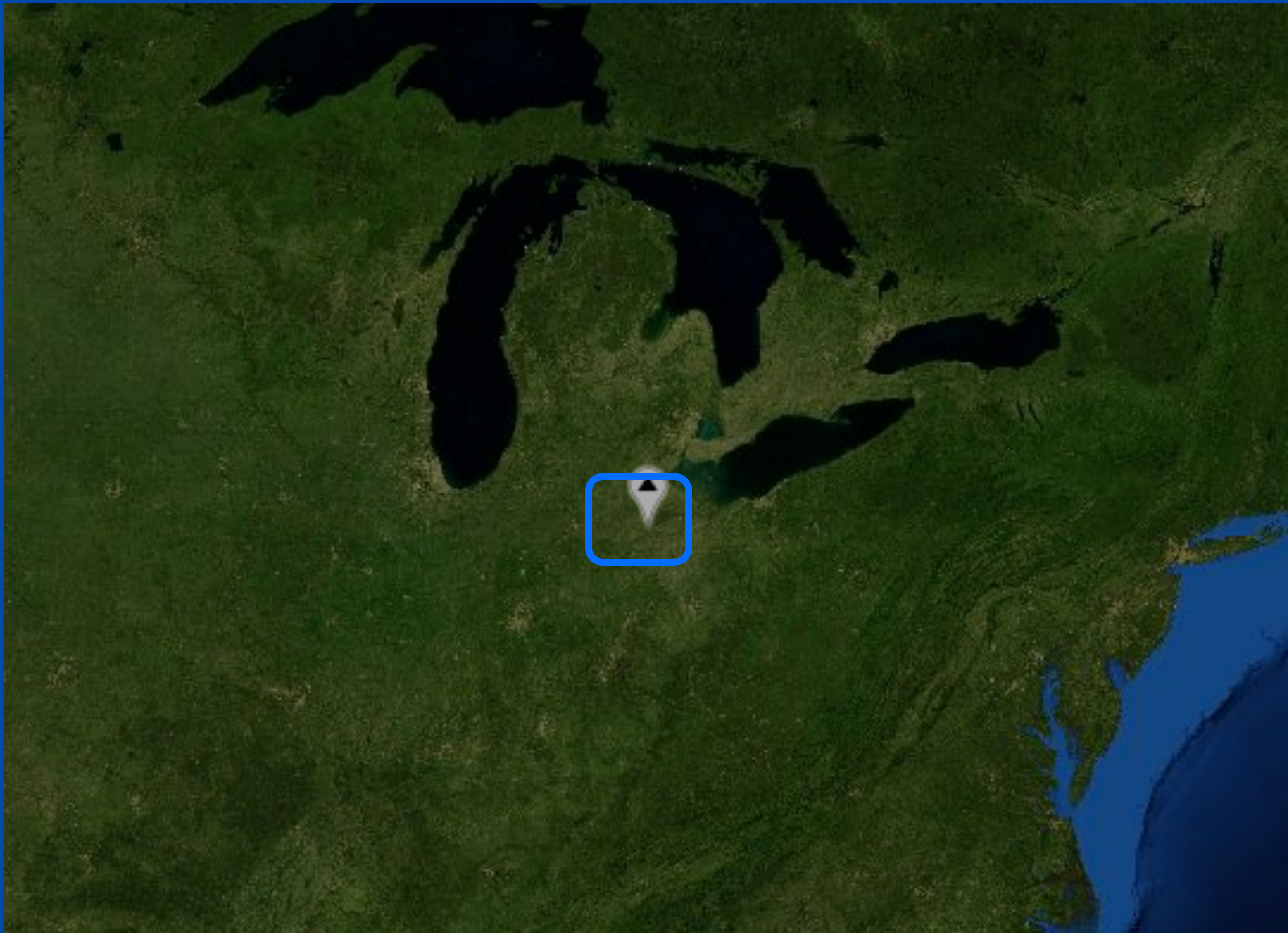

We want the flow-normalized flux history

# For understanding the changes in the rivers




We want the concentration history

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



USGS

# Cyanobacter – Lake Erie

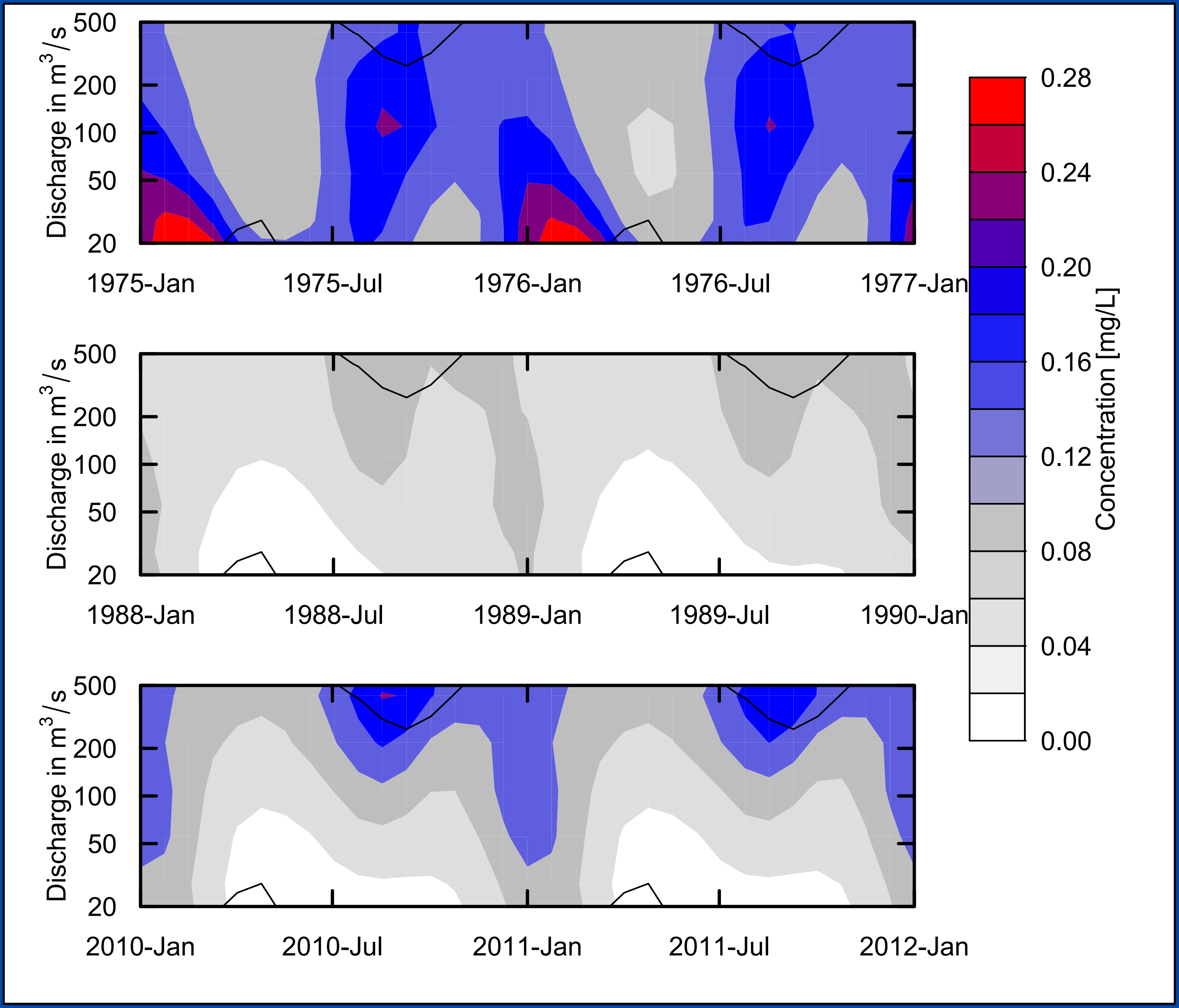
USGS

# Dissolved Reactive Phosphorus, Maumee River, at Waterville, OH

**Mid 1970's**

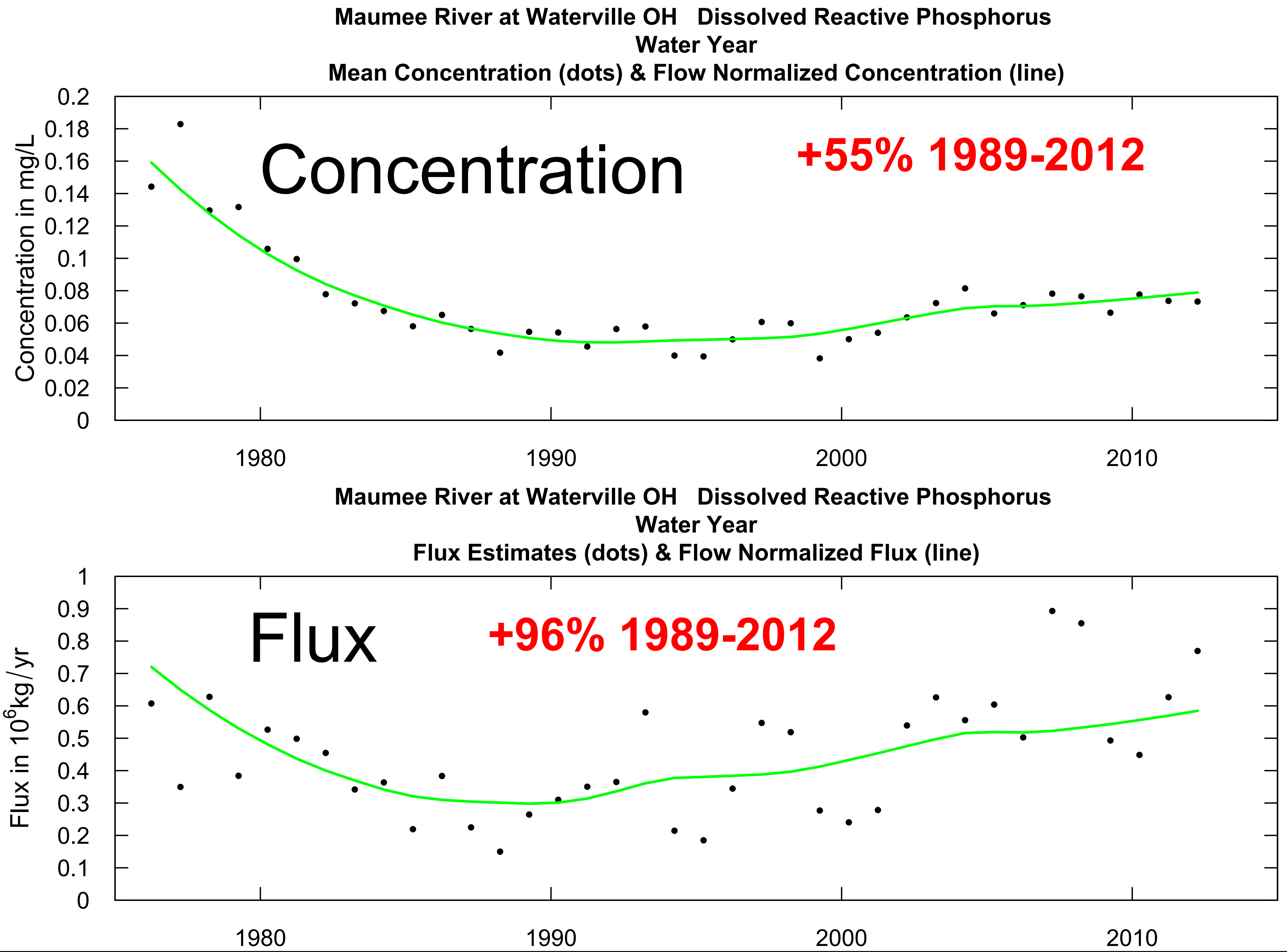
**Late 1980's**

**Early 2010's**




USGS

# Dissolved Reactive Phosphorus, Maumee River, at Waterville, OH



# 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)
```

You are ready to go

check out our new developments at:

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

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 readNWISSample) just type ?readNWISSample

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

**The result: we have created a data frame of 708 rows (one per sample) with columns for: Date, Concentration, Days since January 1, 1850, Month of the year, Day of the year, Decimal year, sine and cosine of time of year, and censoring information.**

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.

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 at Conowingo, 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<-getWQPSample(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(site,"00060",startDate,endDate)
```

The result: we have created a data frame of 12,782 rows (one per day) with columns for:

Date, Discharge, Days since January 1, 1850, Month of the year, Day of the year, Decimal year, mean flow for past 7 days, mean flow for past 30 days

# Storing the metadata

- For NWIS data

  INFO<-readNWISInfo(site,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

```
> INFO<-readNWISInfo(site,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):

<cr>
```

USGS

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):**

**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 command:**

**> INFO <- readUserInfo()**

**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)

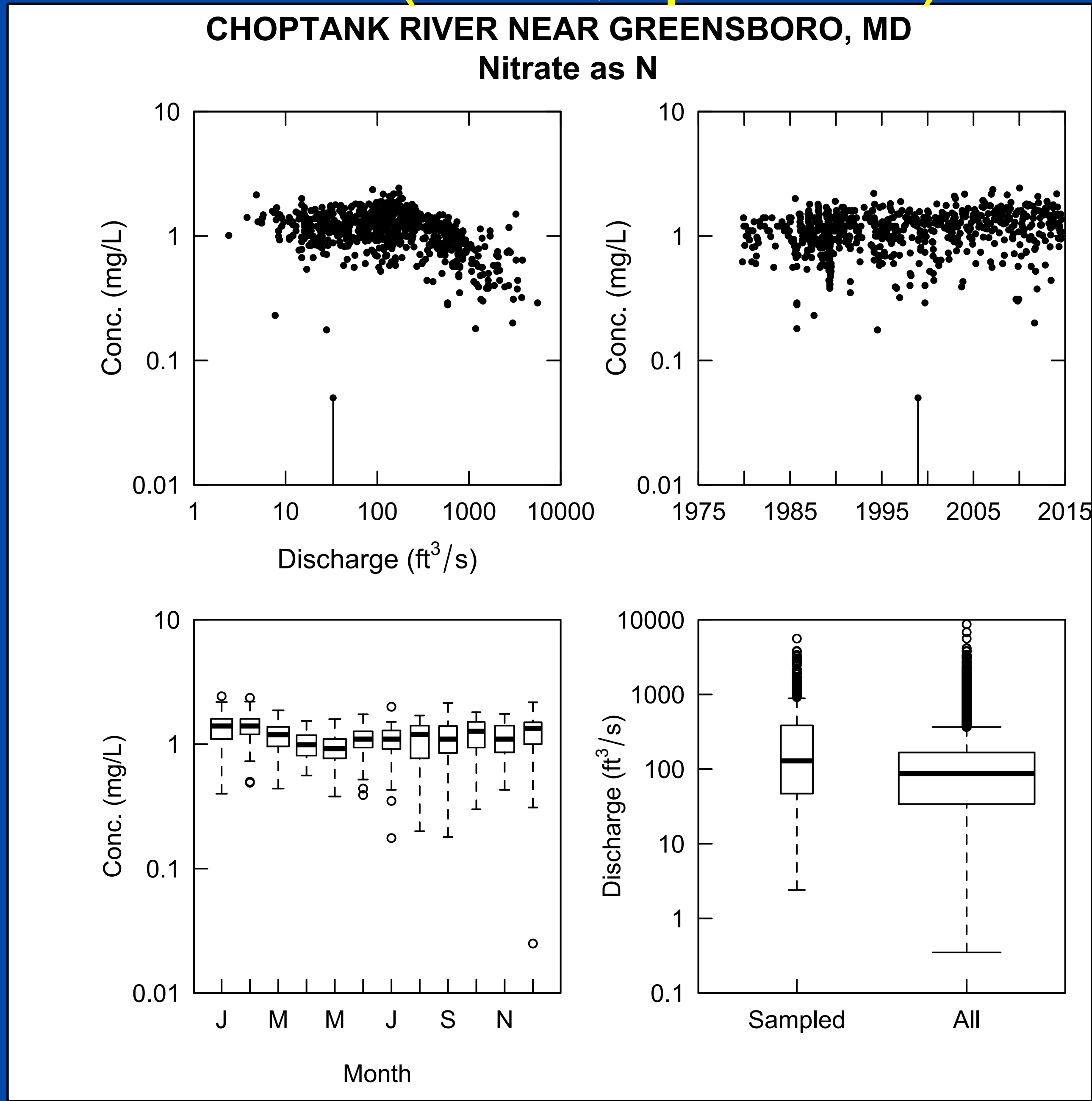
```
> eList <- mergeReport(INFO,Daily,Sample)

 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 %
```

**Now, look at your data.**

**No excuses!!**

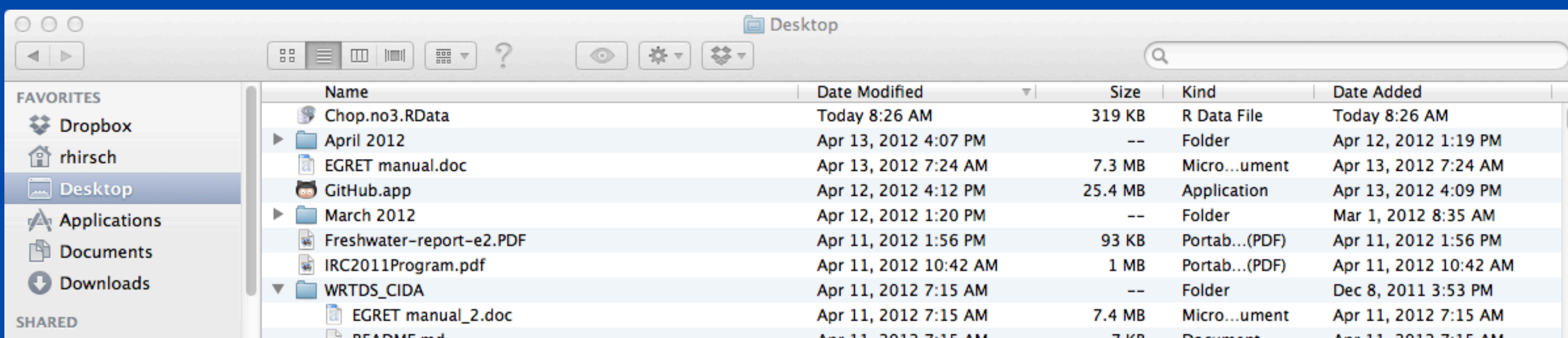
USGS

# > multiPlotDataOverview(eList, qUnit=1)



# 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, bound together in eList

- Sample (708 rows, 14 columns)
- Daily (12,782 rows, 12 columns)
- INFO (1 row, 53 columns)

USGS

# > modelEstimation(eList)

- 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

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

USGS

# Now what is in Daily?
# It now has dimensions (12782, 19)

**It has added columns for daily estimates of:**

**log of concentration,**

**standard error of the log of concentration,**

**concentration,**

**flux,**

**flow-normalized concentration,**

**flow-normalized flux**

USGS

## Now what is in Sample?
## It now has dimensions (708, 14)

It has added columns for "leave-one-out cross validation" estimates of the following for each sampled day:

log of concentration,

standard error of the log of concentration,

concentration,

USGS

# "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:

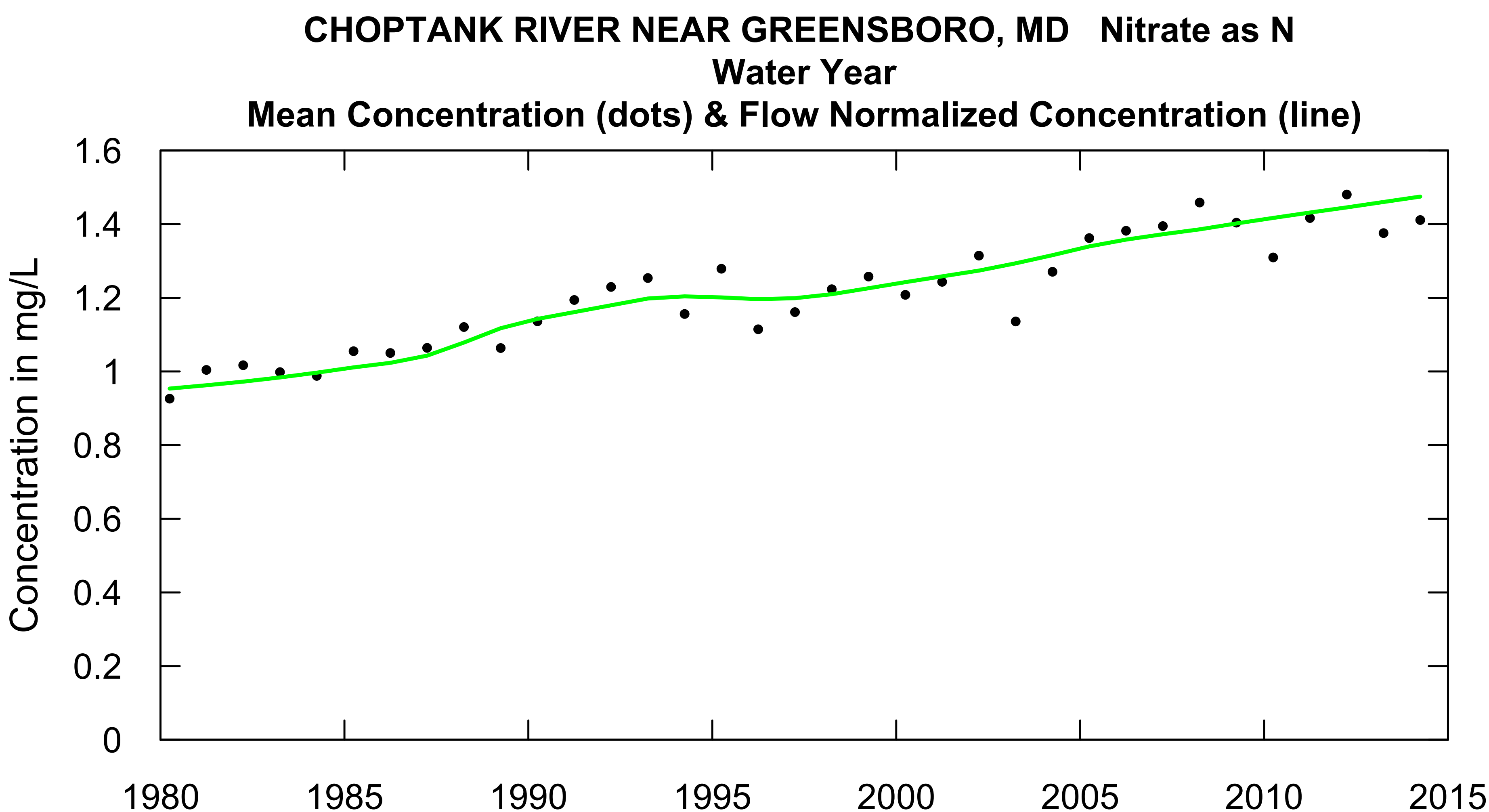
$m^3/s$, kg/day, or mg/L

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

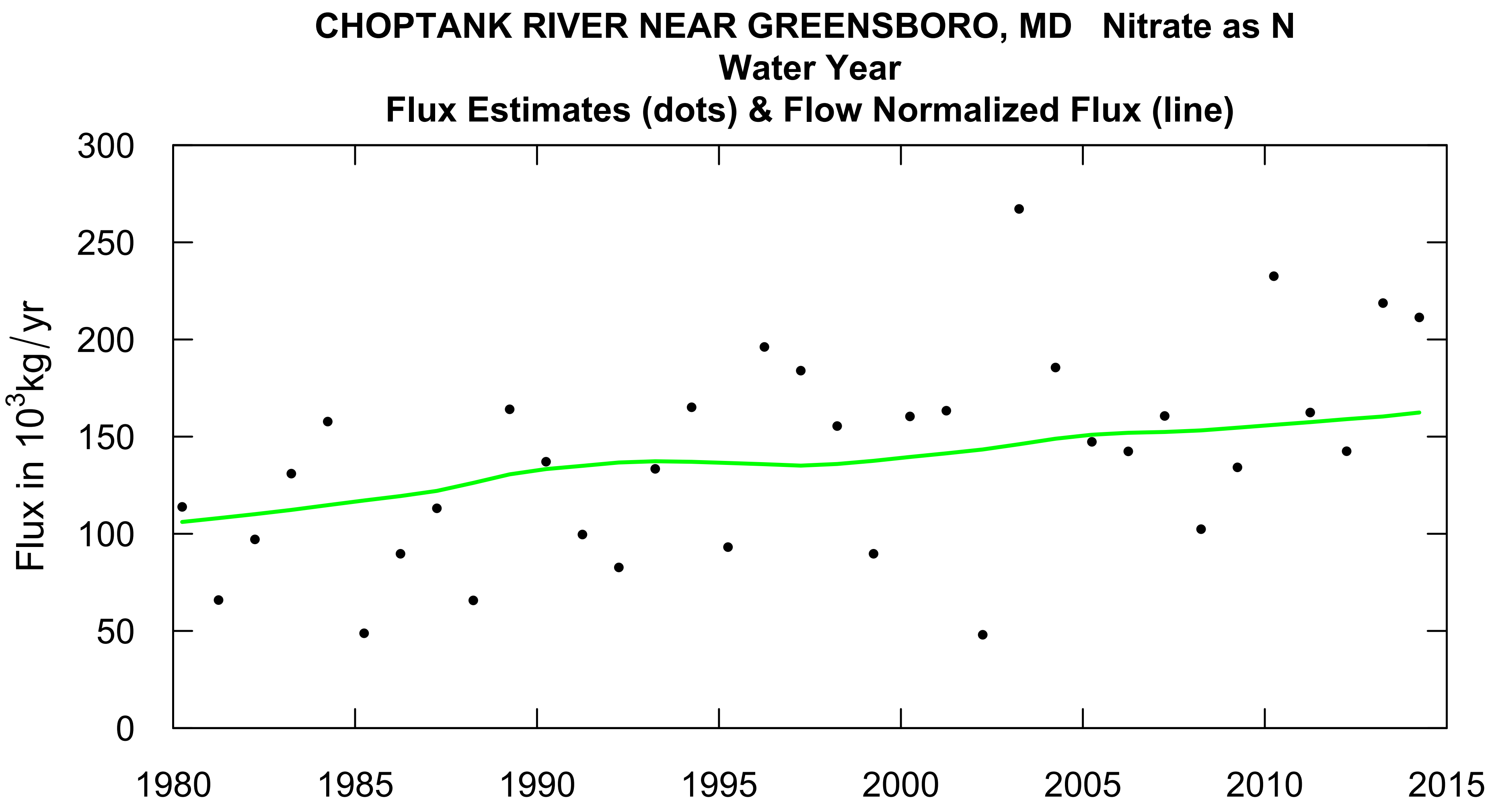
Now lets see some trend results

USGS

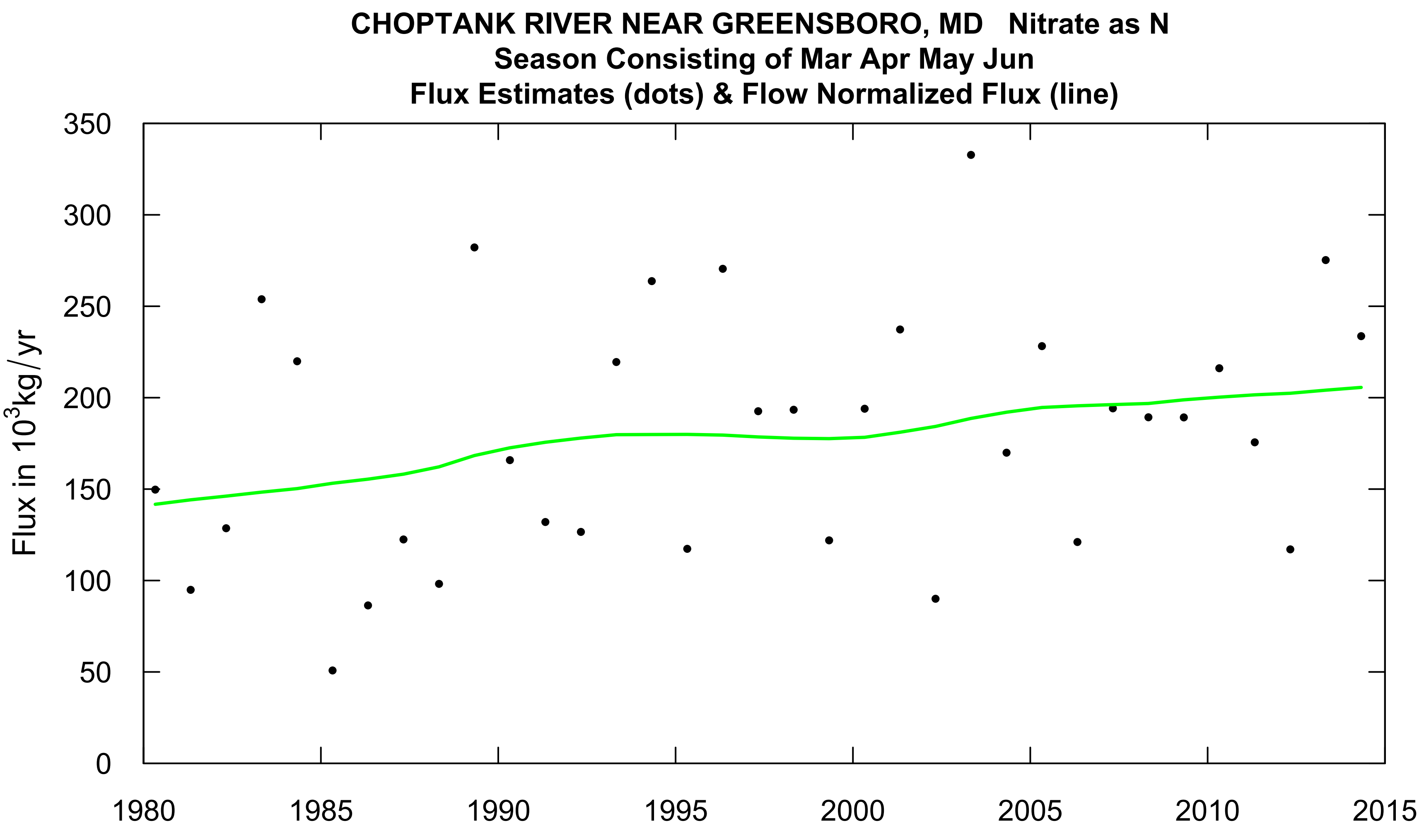
# > plotConcHist(eList)



# > 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 maximum scale
- Use log scale
- Change colors
- Save image as .png or .pdf
- …

The EGRET vignette provides information on "Extending Plots Past Defaults", p.47

USGS

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

CHOPTANK RIVER NEAR GREENSBORO, MD
Nitrate as N
Water Year

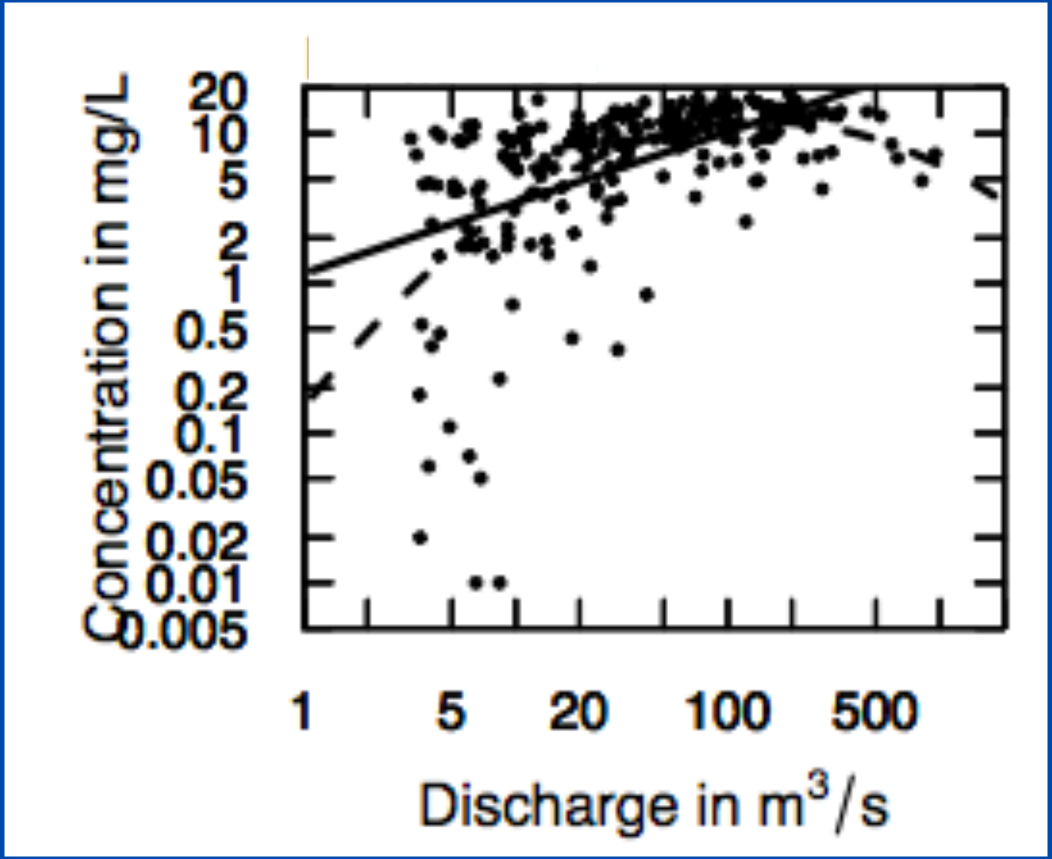
| Year | Discharge | Conc | FN_Conc | Flux | FN_Flux |
|---|---|---|---|---|---|
| | cfs | mg/L | | tons/yr | |
| 1980 | 150.2 | 0.926 | 0.953 | 125.5 | 117 |
| 1981 | 78.3 | 1.004 | 0.963 | 72.6 | 119 |
| 1982 | 107.6 | 1.017 | 0.972 | 107.0 | 121 |
| 1983 | 176.1 | 0.998 | 0.984 | 144.4 | 124 |
| 1984 | 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

The "flux bias problem": Some regression-based models (such as LOADEST) can produce flux estimates with very large biases (+ or -)

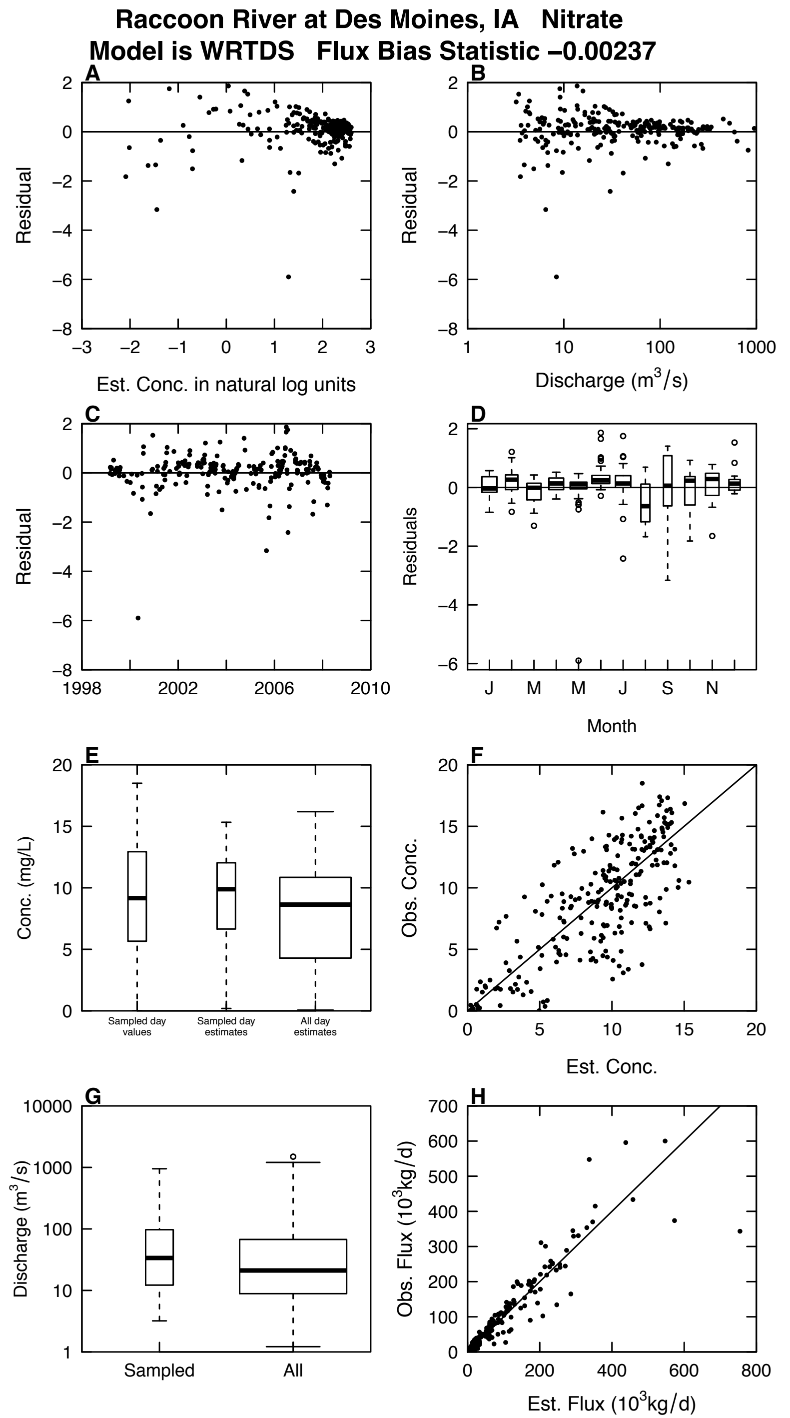
I'm going to switch data sets to Nitrate for the Raccoon River at Des Moines Iowa




USGS

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

fluxBiasMulti(eList, fluxUnit=4)



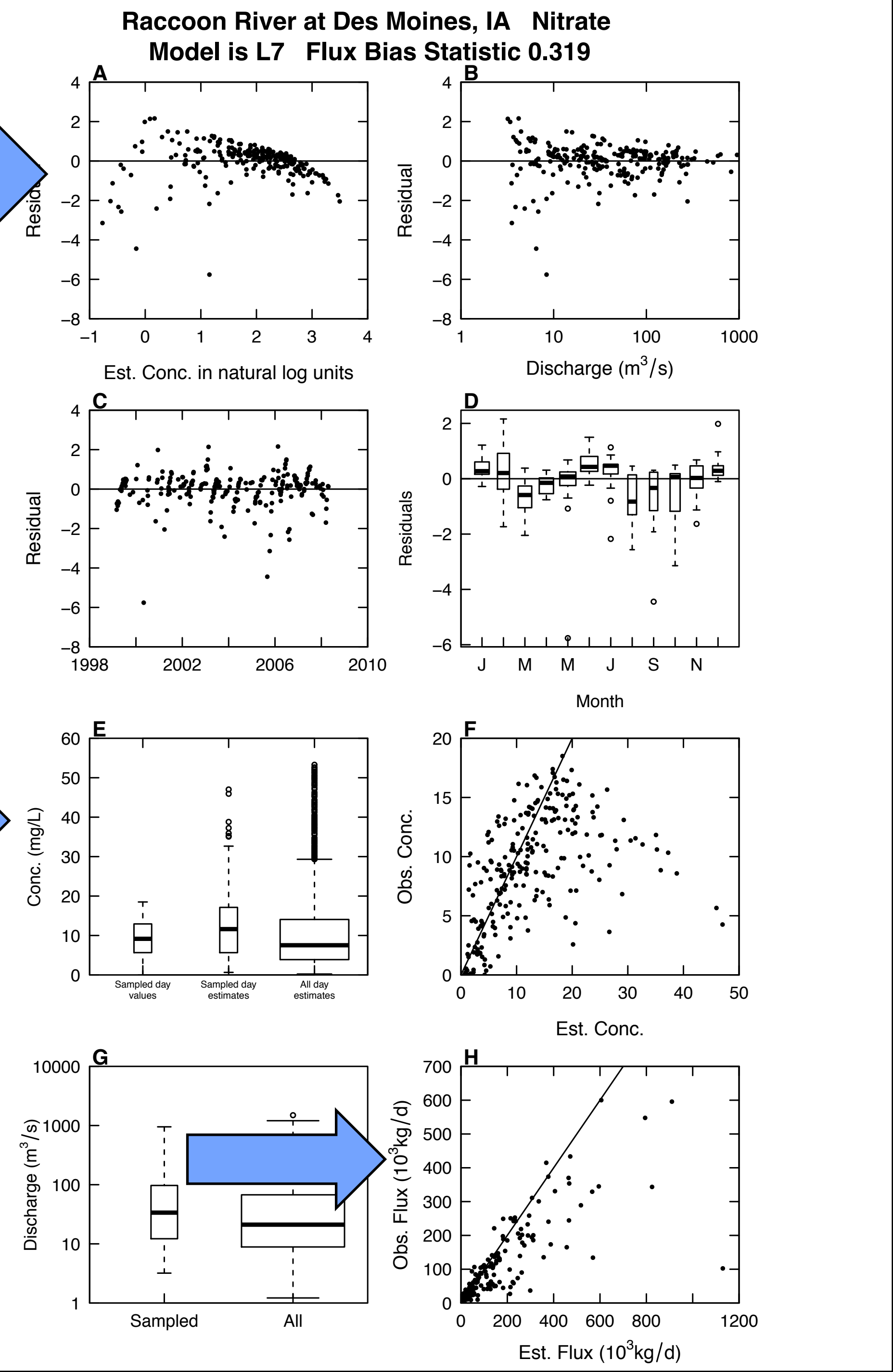

USGS

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:

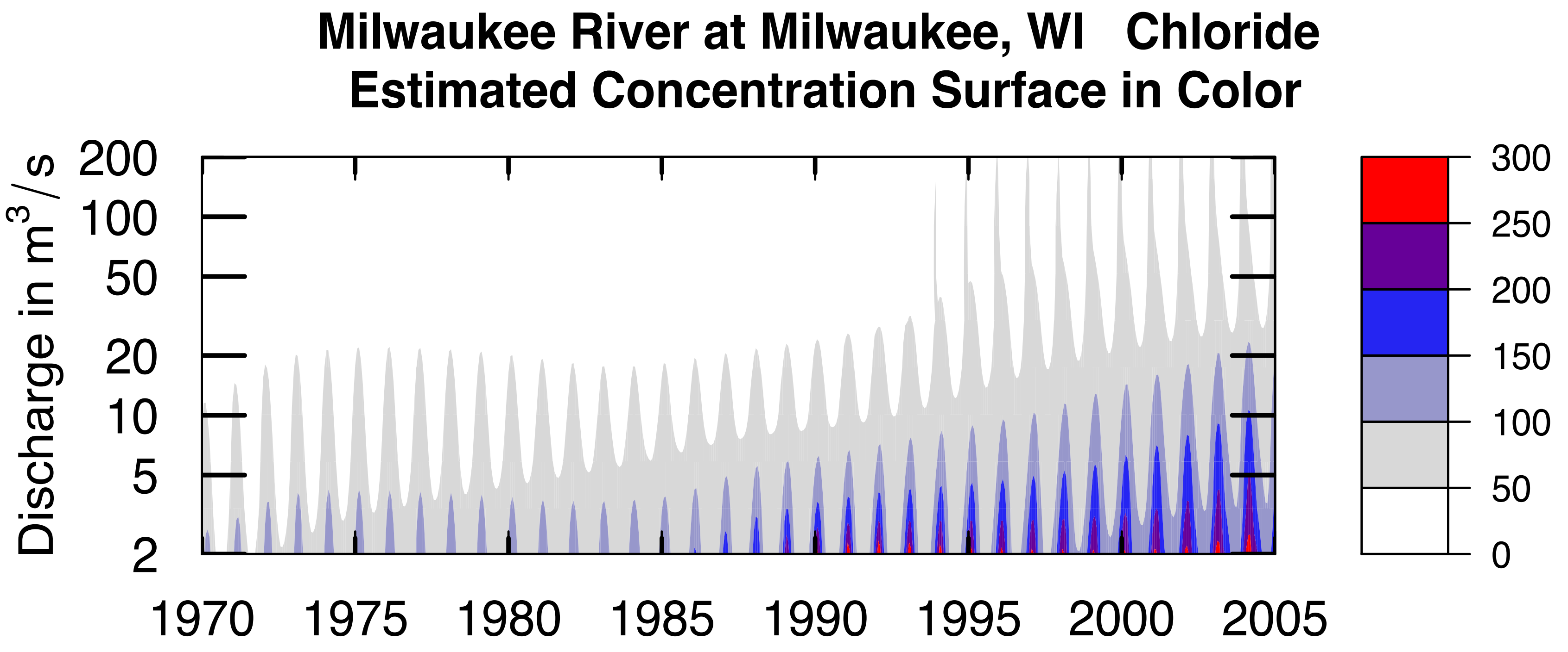
Hirsch, R.M., 2014, Large biases in regression-based constituent flux estimates: causes and diagnostics. Journal of the American Water Resources Association.

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

USGS

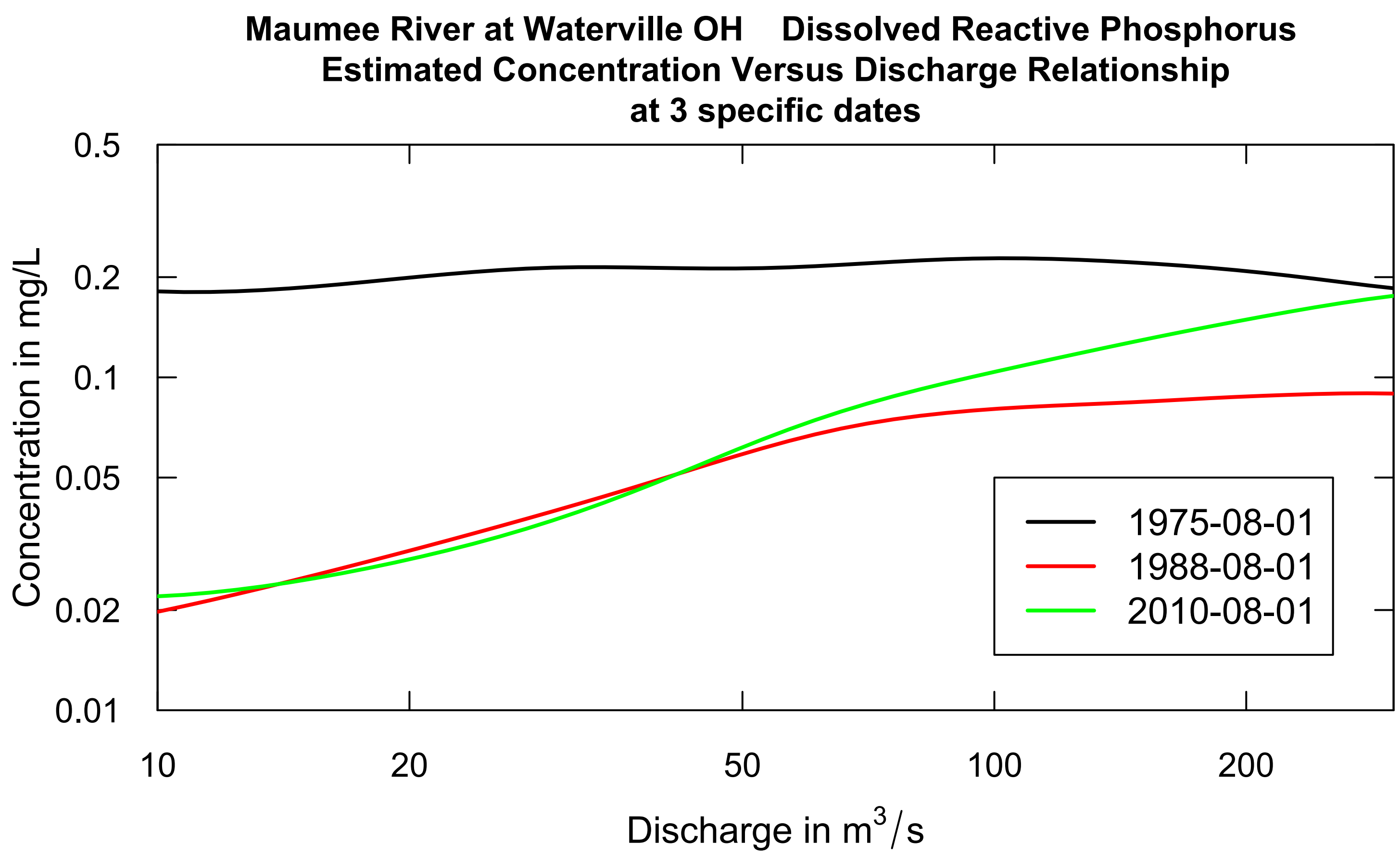
# How difficult is it to make those contour plots?

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

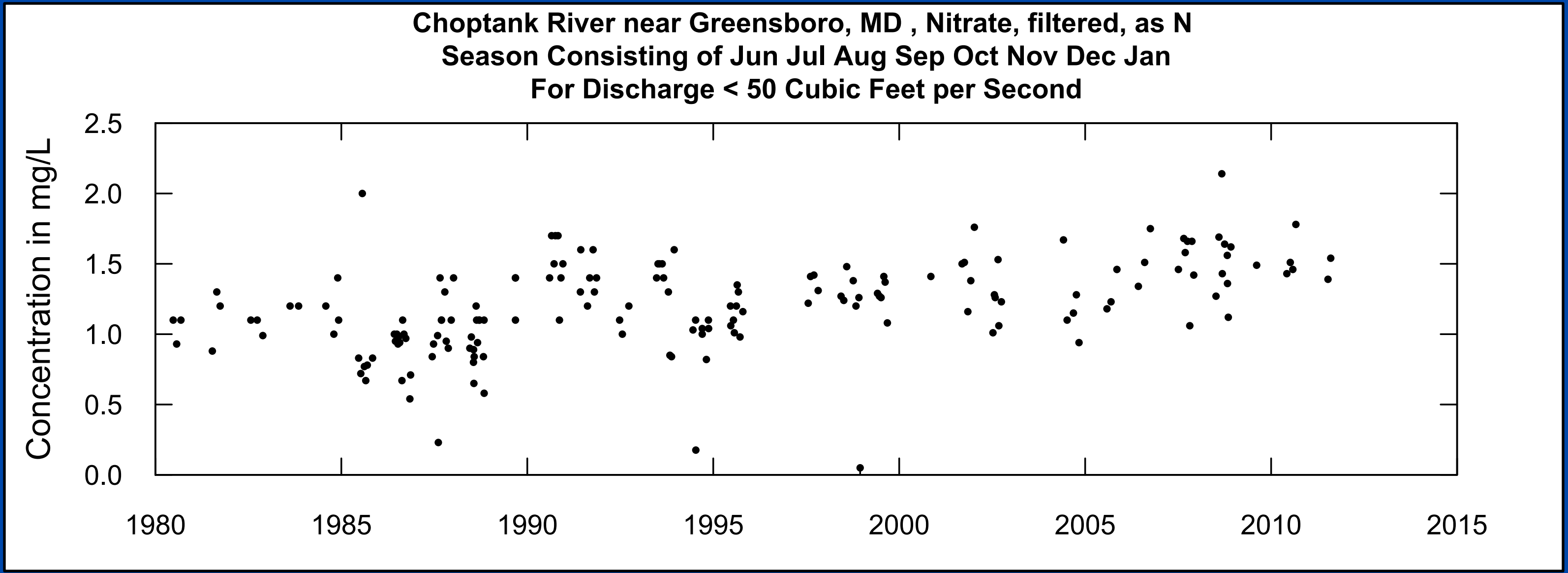


# There are many more graphics, for example

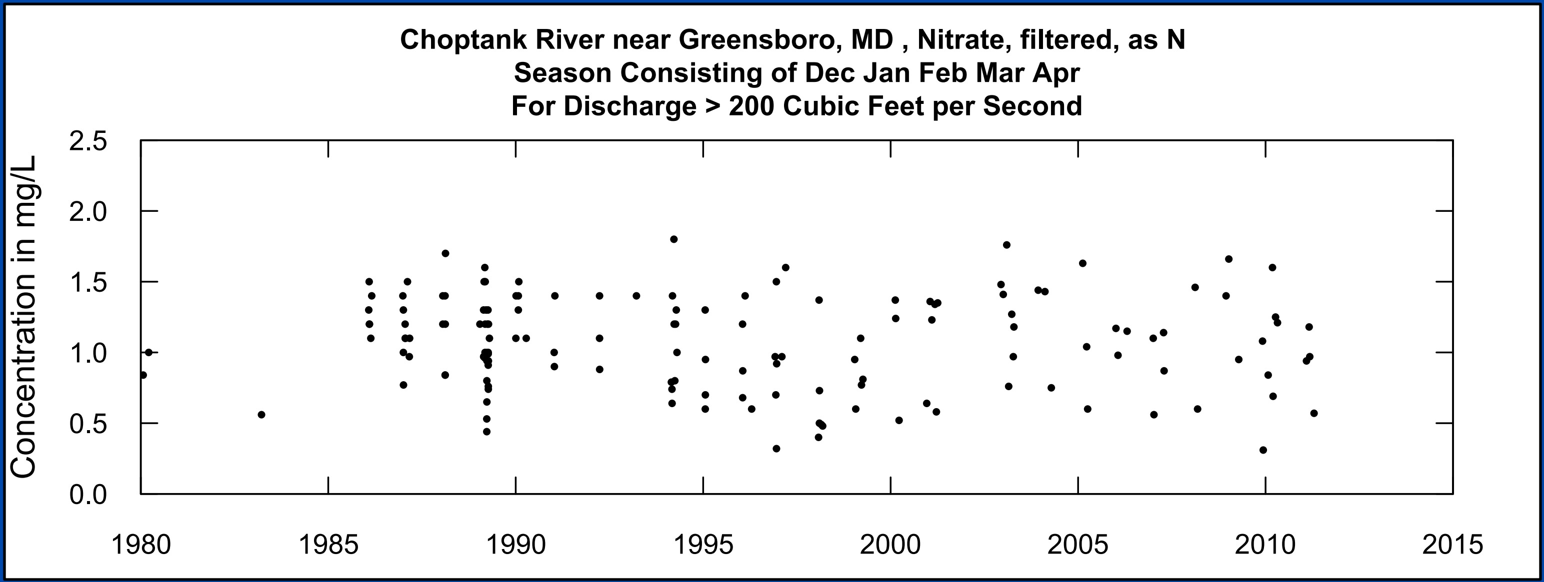
**> plotConcQSmooth(eList,"1975-08-01”, "1988-08-01", "2010-08-01", qLow=10, qHigh=300, qUnit=2, logScale=TRUE, legendLeft=100, legendTop=0.05)**



> 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)




USGS

# 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

# Based on published paper

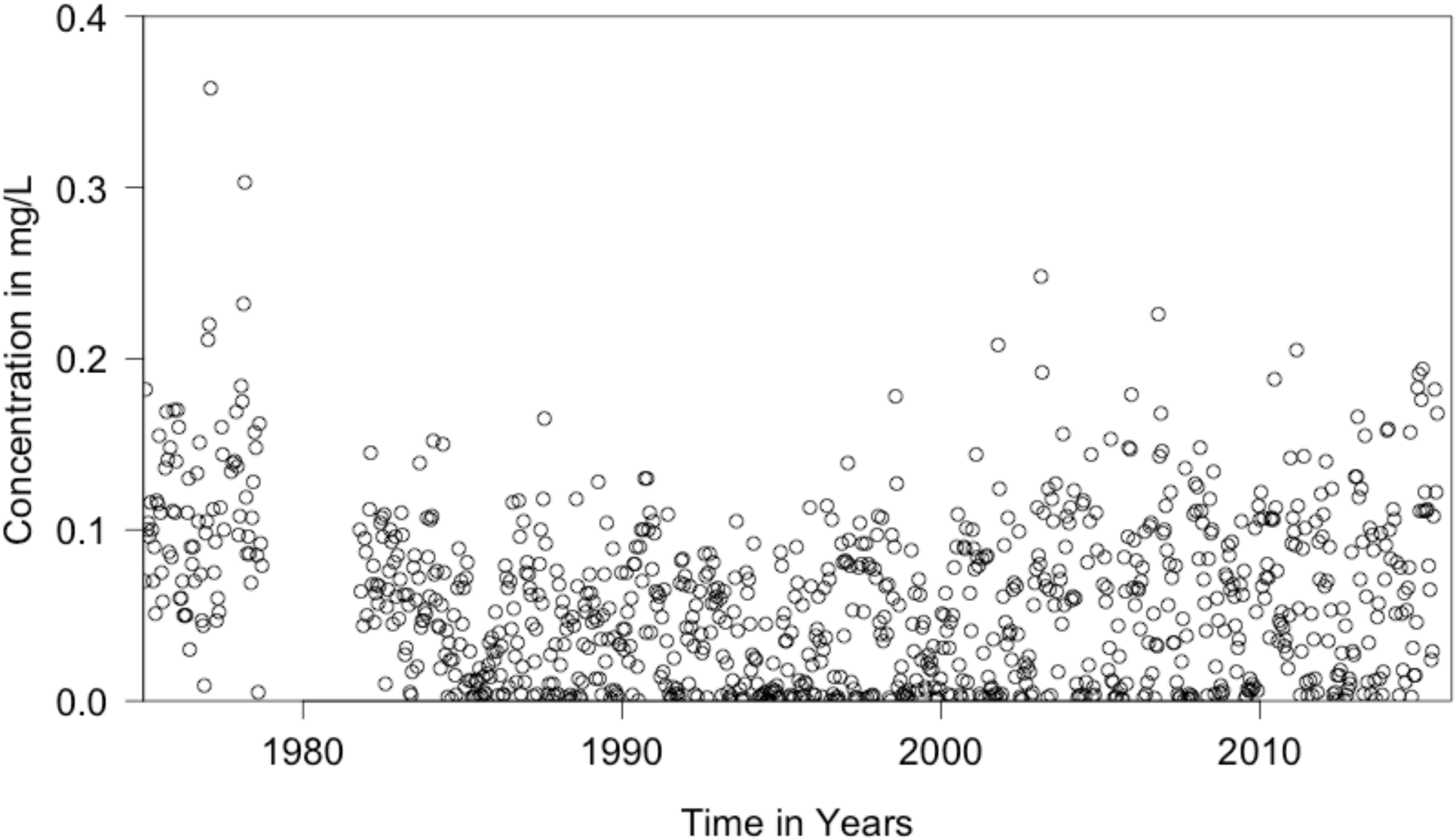
Hirsch, Robert M., Archfield, Stacey A., and DeCicco, Laura A., 2015,

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

Environmental Modelling and Software, 73, 148-166.

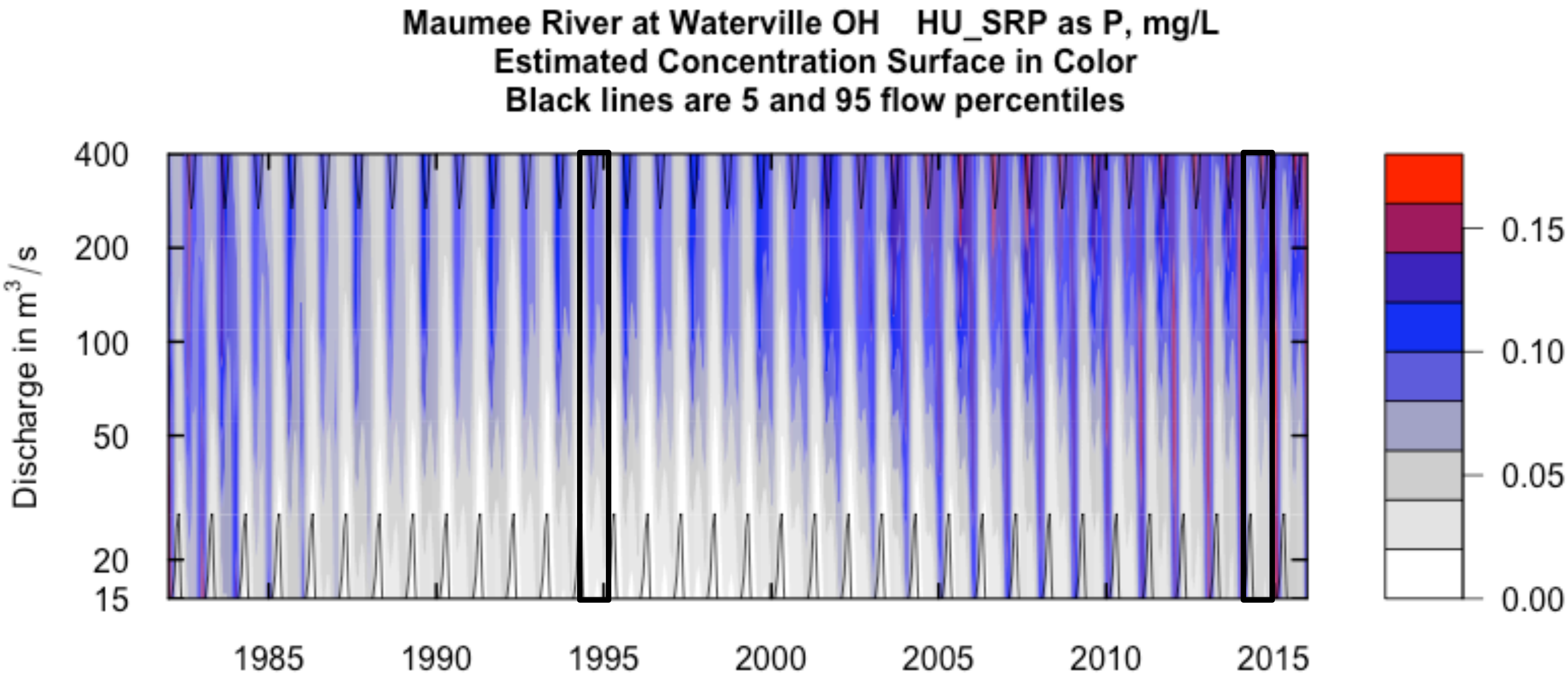
USGS

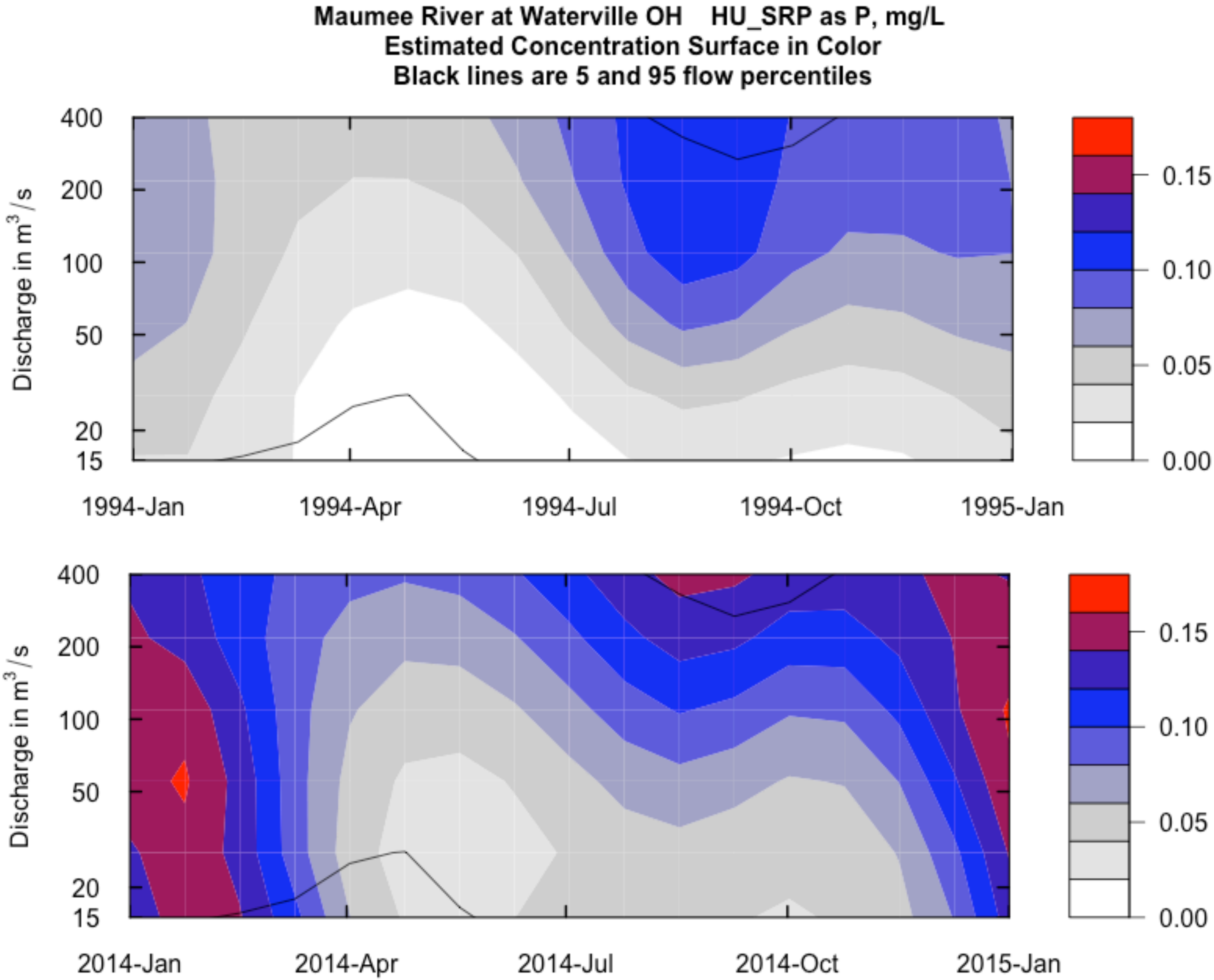
## Raw Sample of SRP, Maumee River



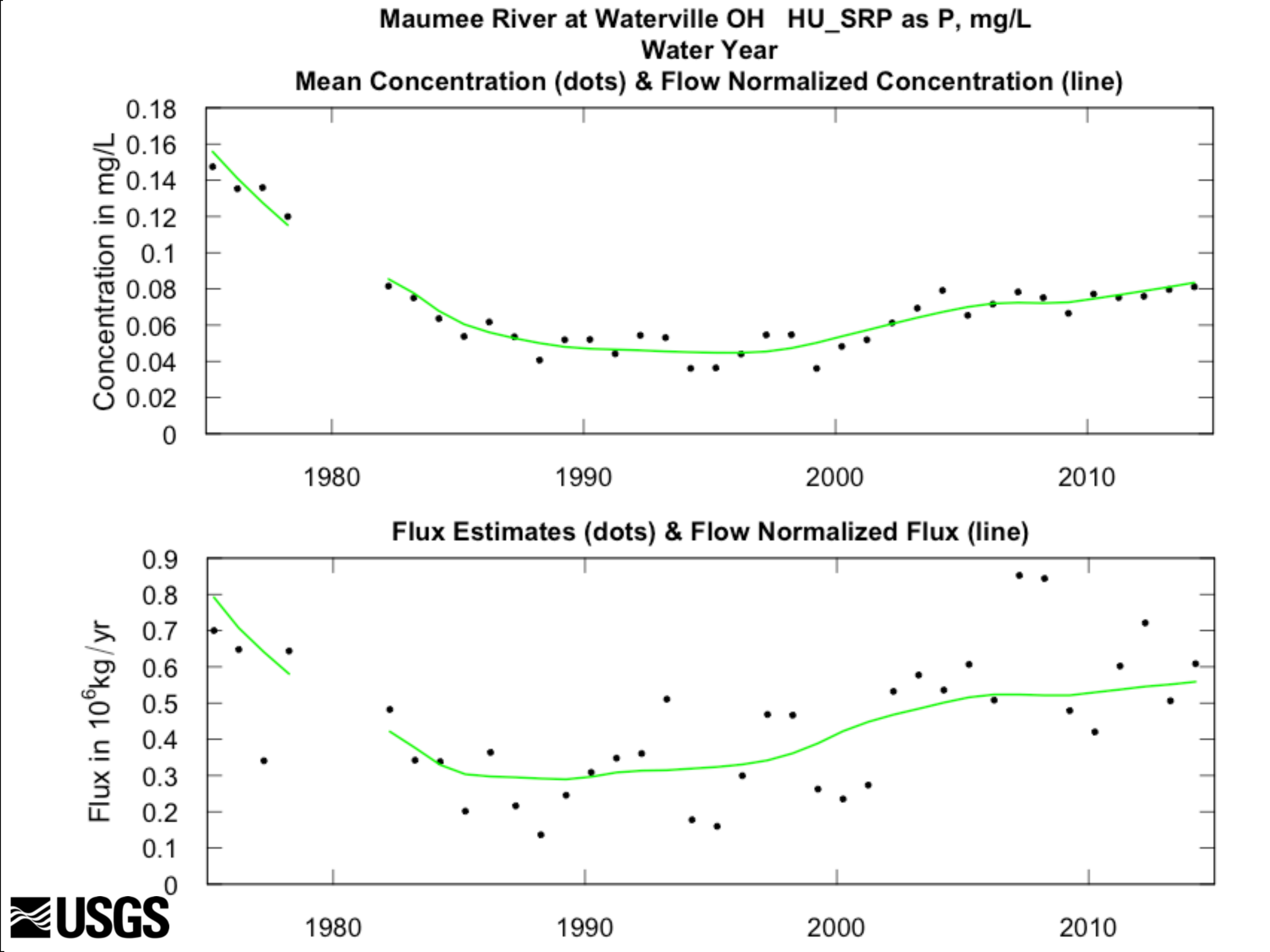

USGS

# WRTDS representation of concentration as a function of time and discharge





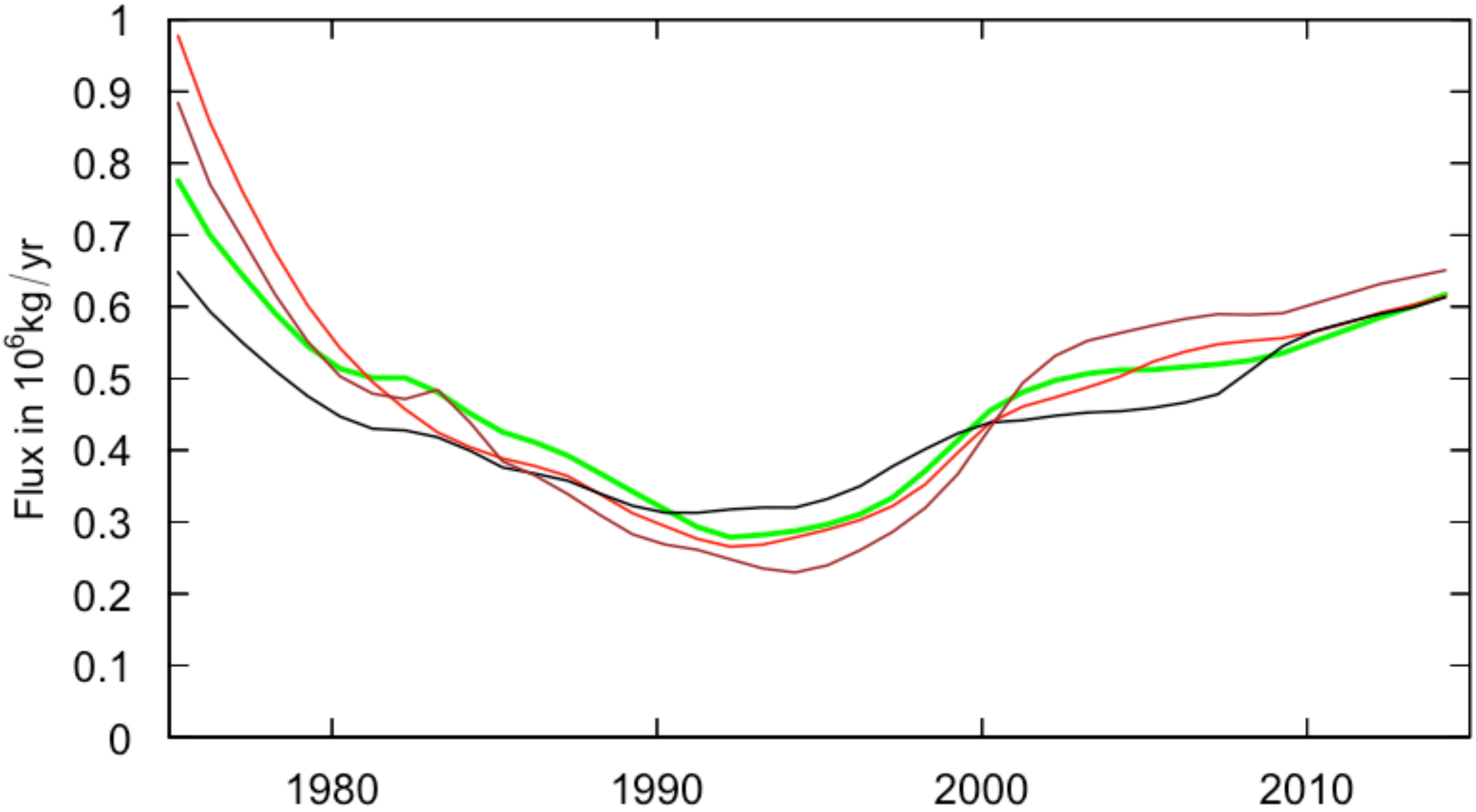
Maumee River at Waterville OH HU_SRP as P, mg/L
Estimated Concentration Surface in Color
Black lines are 5 and 95 flow percentiles
Discharge in m3 / s
400
200
100
50
20
15
1994-Jan
1994-Apr
1994-Jul
1994-Oct
1995-Jan
0.15
0.10
0.05
0.00
Discharge in m3 / s
400
200
100
50
20
15
2014-Jan
2014-Apr
2014-Jul
2014-Oct
2015-Jan
0.15
0.10
0.05
0.00


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

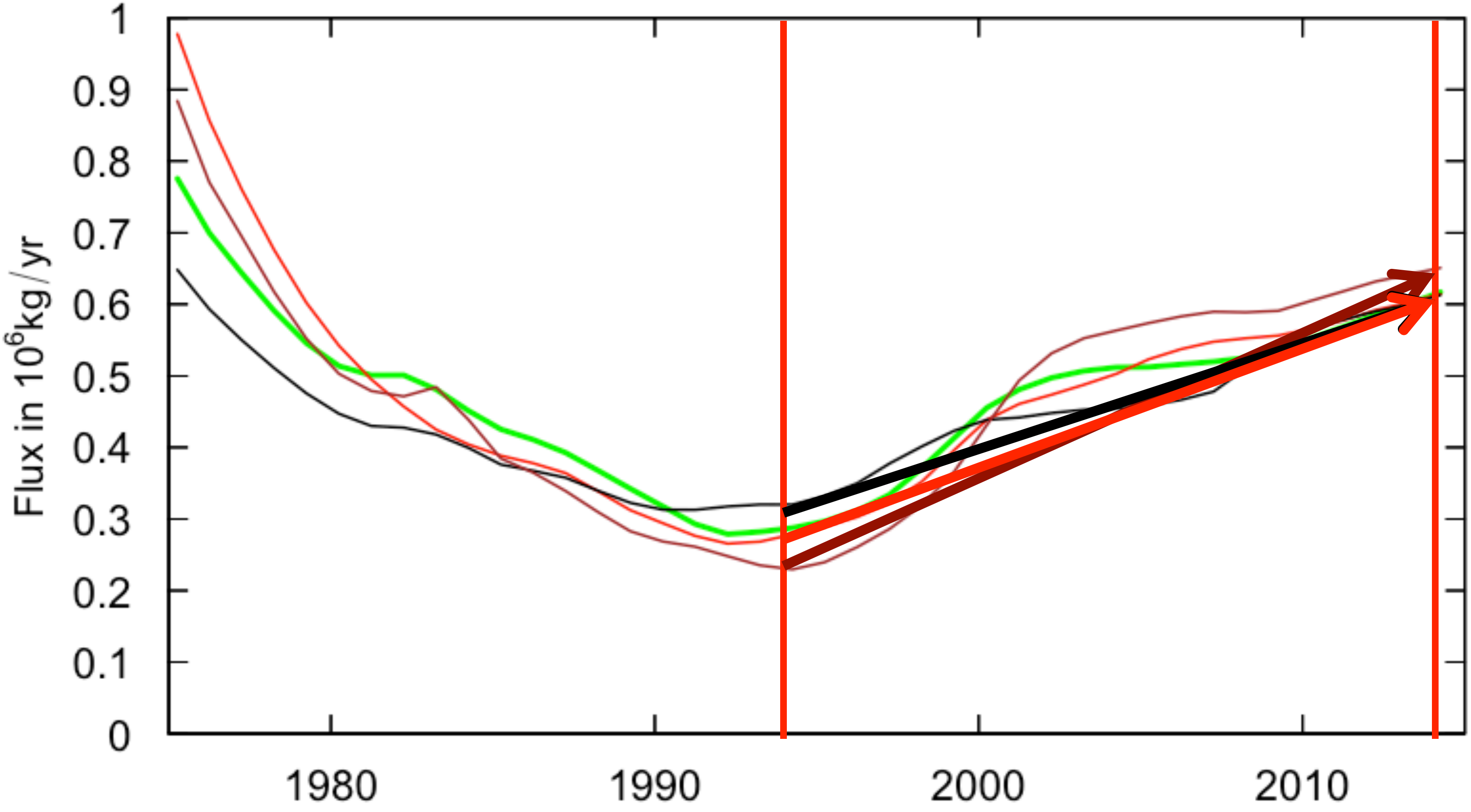
- Resample the data set, by 200 day blocks, with replacement
- Conduct the WRTDS estimation process for each replicate
- Uncertainty of the trend magnitude is determined from a sample of bootstrap estimates for the selected trend period.

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



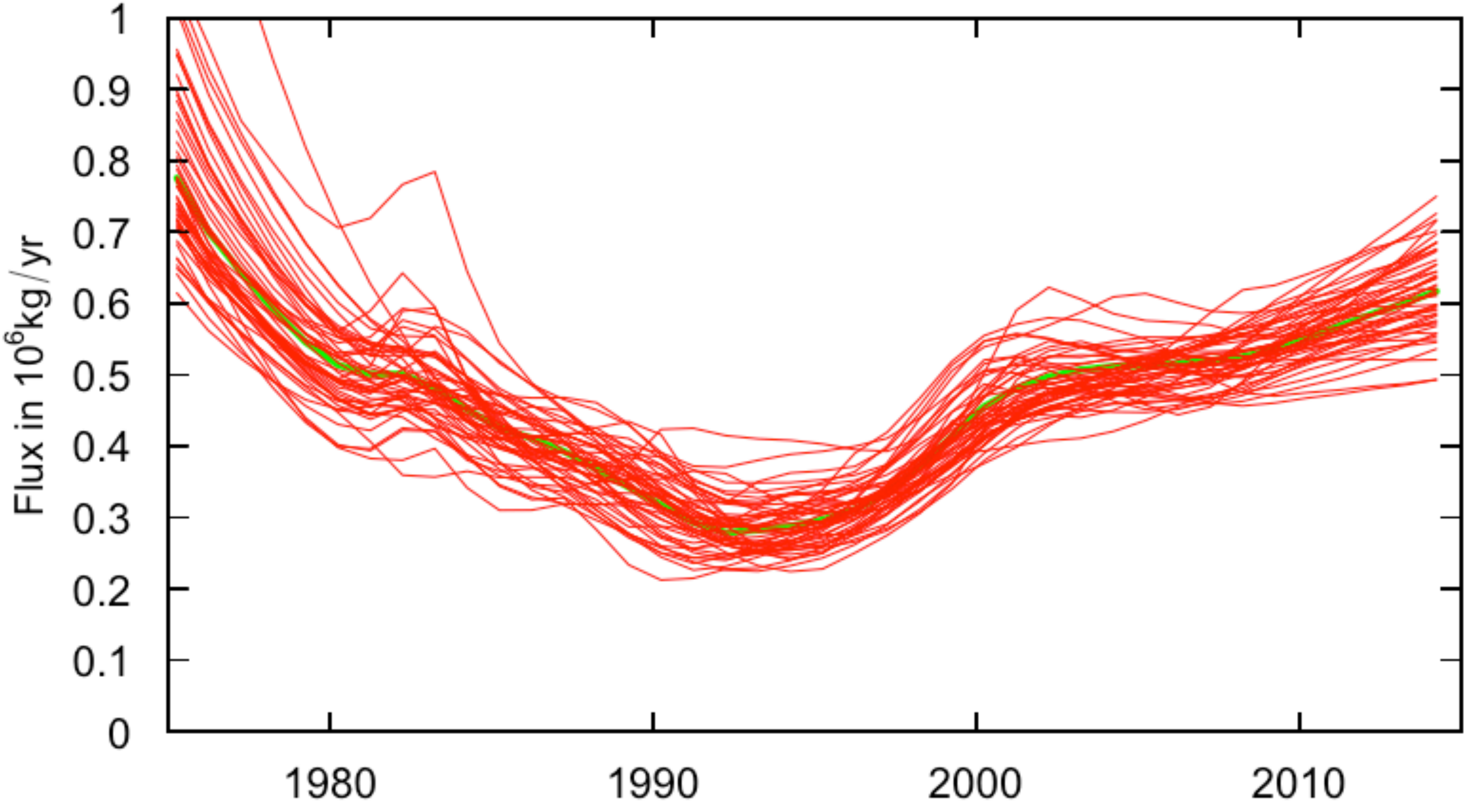

USGS

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




USGS

# 50 bootstrap replicates

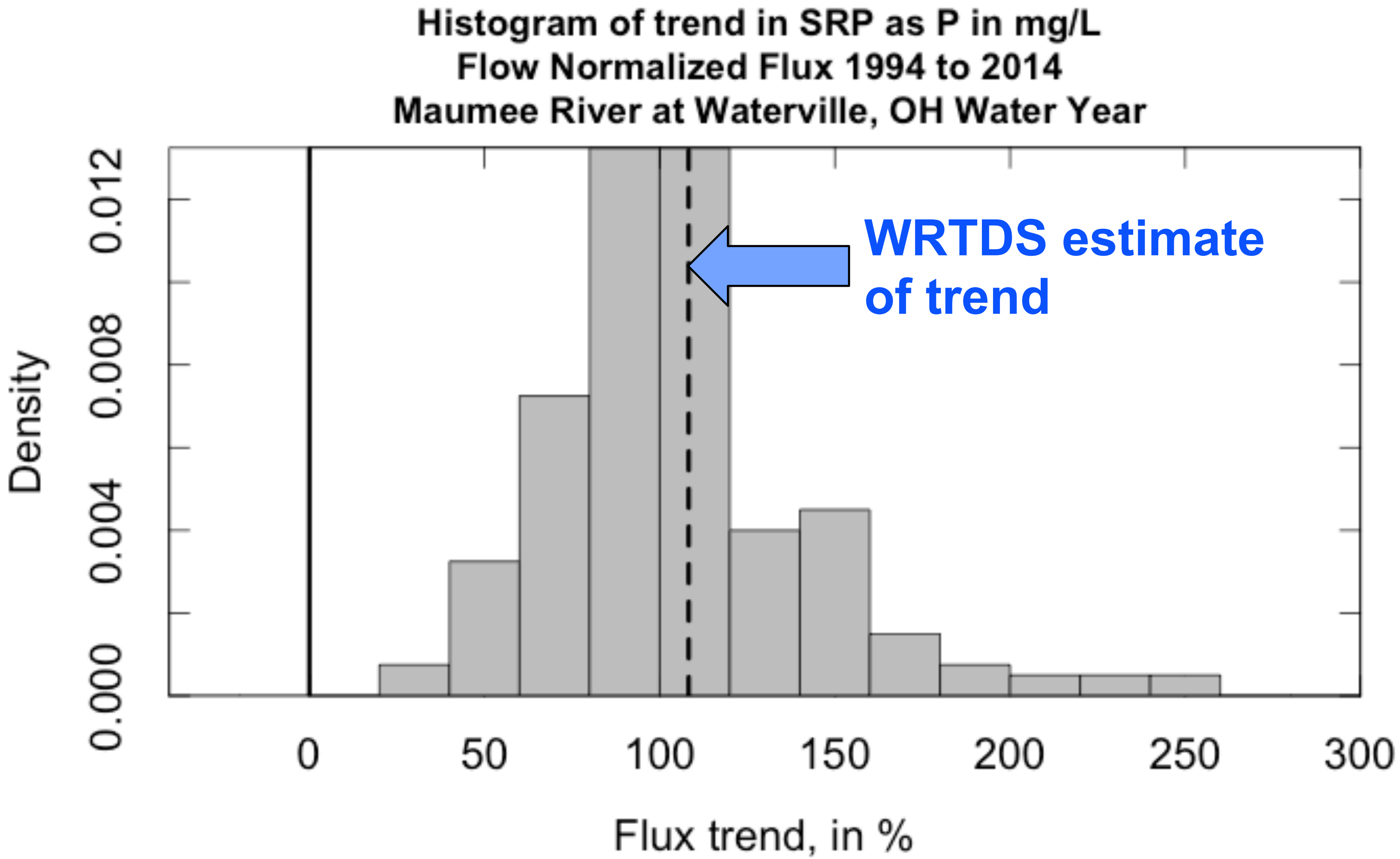



USGS

# 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 |

# 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

USGS

# EGRETci output looks like this:

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

Water Year

  Bootstrap process, for change from Water Year 2004 to Water Year 2014
                   data set runs from WaterYear 1975 to Water Year 2015

  Bootstrap block length in days 200
  bootBreak is 200  confStop is 0.7

 WRTDS estimated concentration change is  0.0185  mg/L
 WRTDS estimated flux change is           0.1051  10^6 kg/yr
```

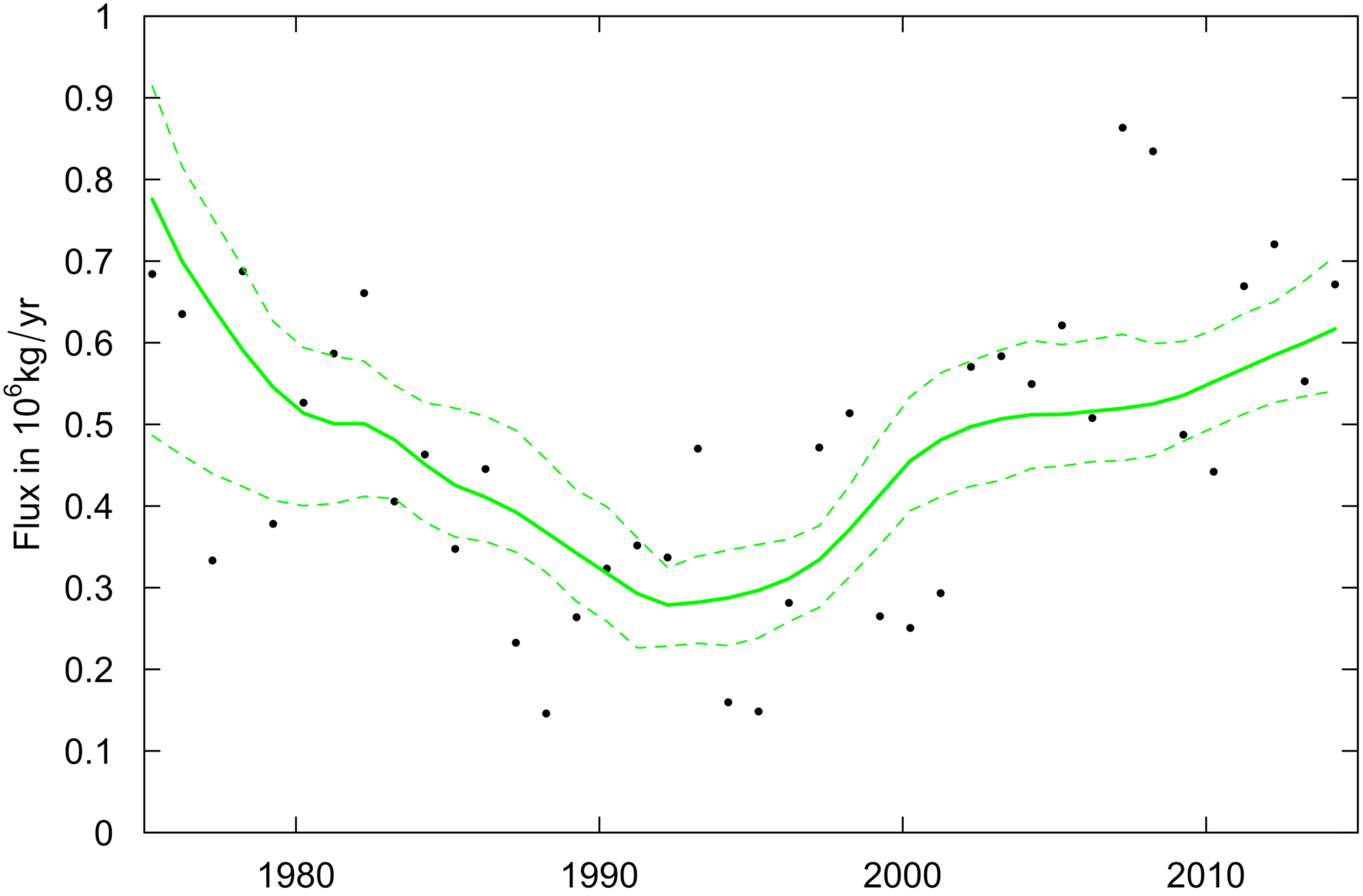
USGS

# Results expressed in text form like this.
# Also output as a data frame for creating summaries

```
Should we reject Ho that Flow Normalized Concentration Trend = 0 ? Do Not Reject Ho
  best estimate is  0.0185 mg/L
  Lower and Upper 90% CIs -0.00129  0.03426
  also 95% CIs-0.00691  0.03593
  and 50% CIs  0.01187  0.02451
  approximate two-sided p-value for Conc      0.11
  Likelihood that Flow Normalized Concentration is trending up =      0.943
  is trending down =     0.0572

 Should we reject Ho that Flow Normalized Flux Trend = 0 ? Reject Ho
  best estimate is   0.1051 10^6 kg/year
  Lower and Upper 90% CIs  0.00293  0.19341
  also 95% CIs -0.02608  0.22192
  and 50% CIs  0.05854  0.14531
  approximate two-sided p-value for Flux     0.077
  Likelihood that Flow Normalized Flux is trending up = 0.963
  is trending down= 0.0373

 Upward trend in concentration is very likely
 Upward trend in flux is highly likely
 Downward trend in concentration is very unlikely
 Downward trend in flux is highly unlikely
```

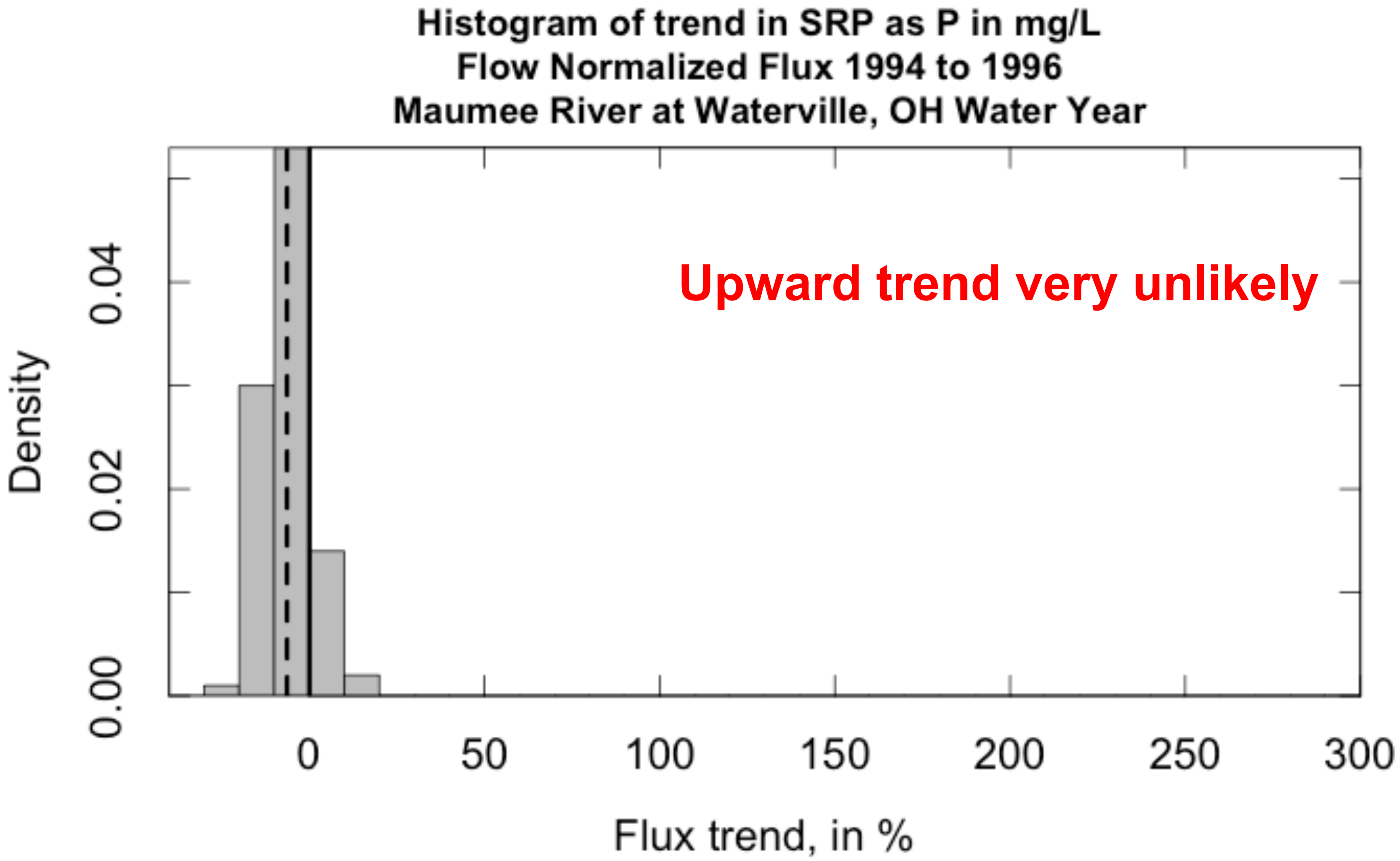
USGS

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

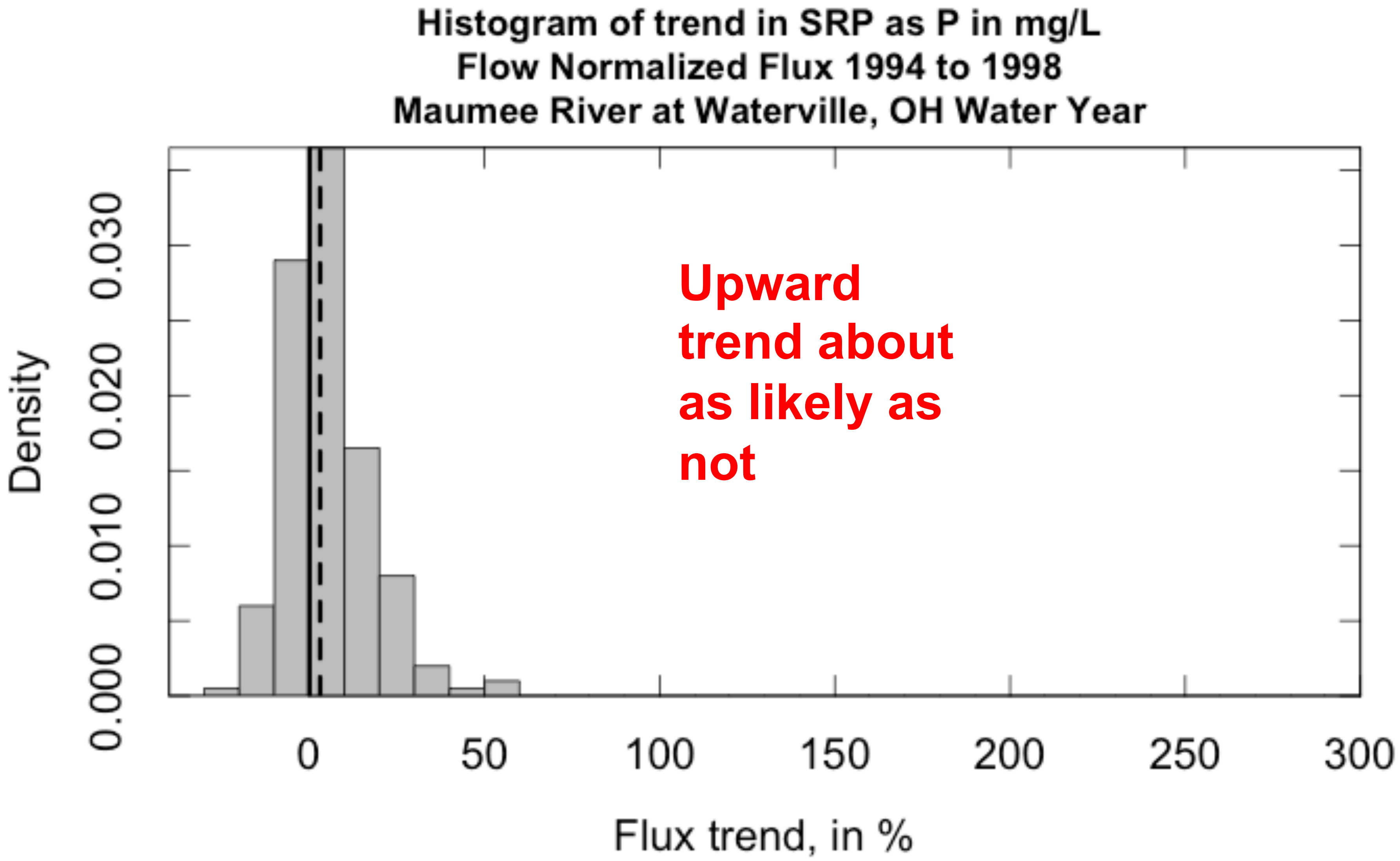


# Let’s simulate the use of the test as if it were done every two years.

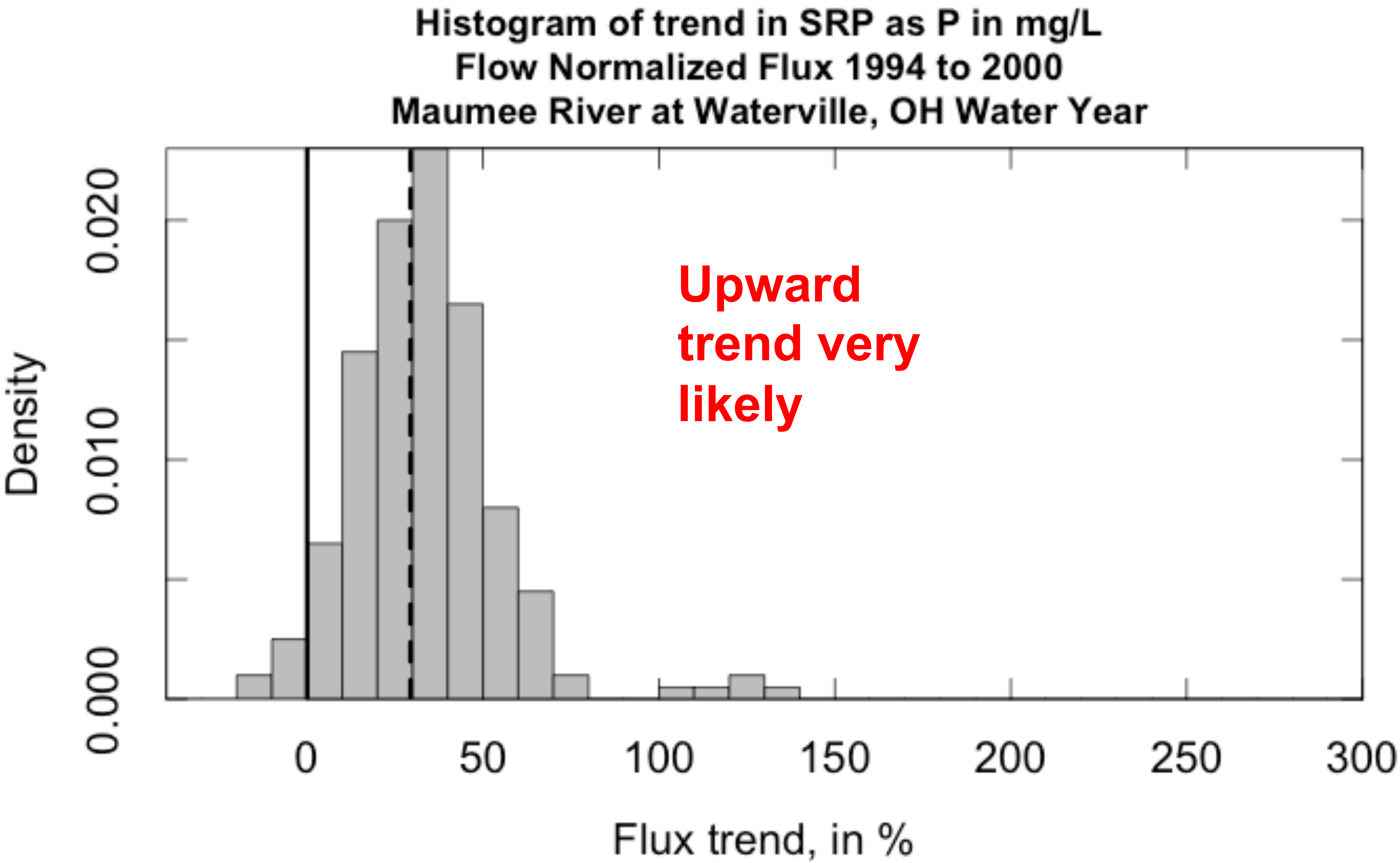
- Each time we will evaluate the trend from Water Year 1994 to the ending year using all the data we have to date.


Histogram of trend in SRP as P in mg/L
Flow Normalized Flux 1994 to 1996
Maumee River at Waterville, OH Water Year
Upward trend very unlikely
Density
0.00
0.02
0.04
0
50
100
150
200
250
300
Flux trend, in %

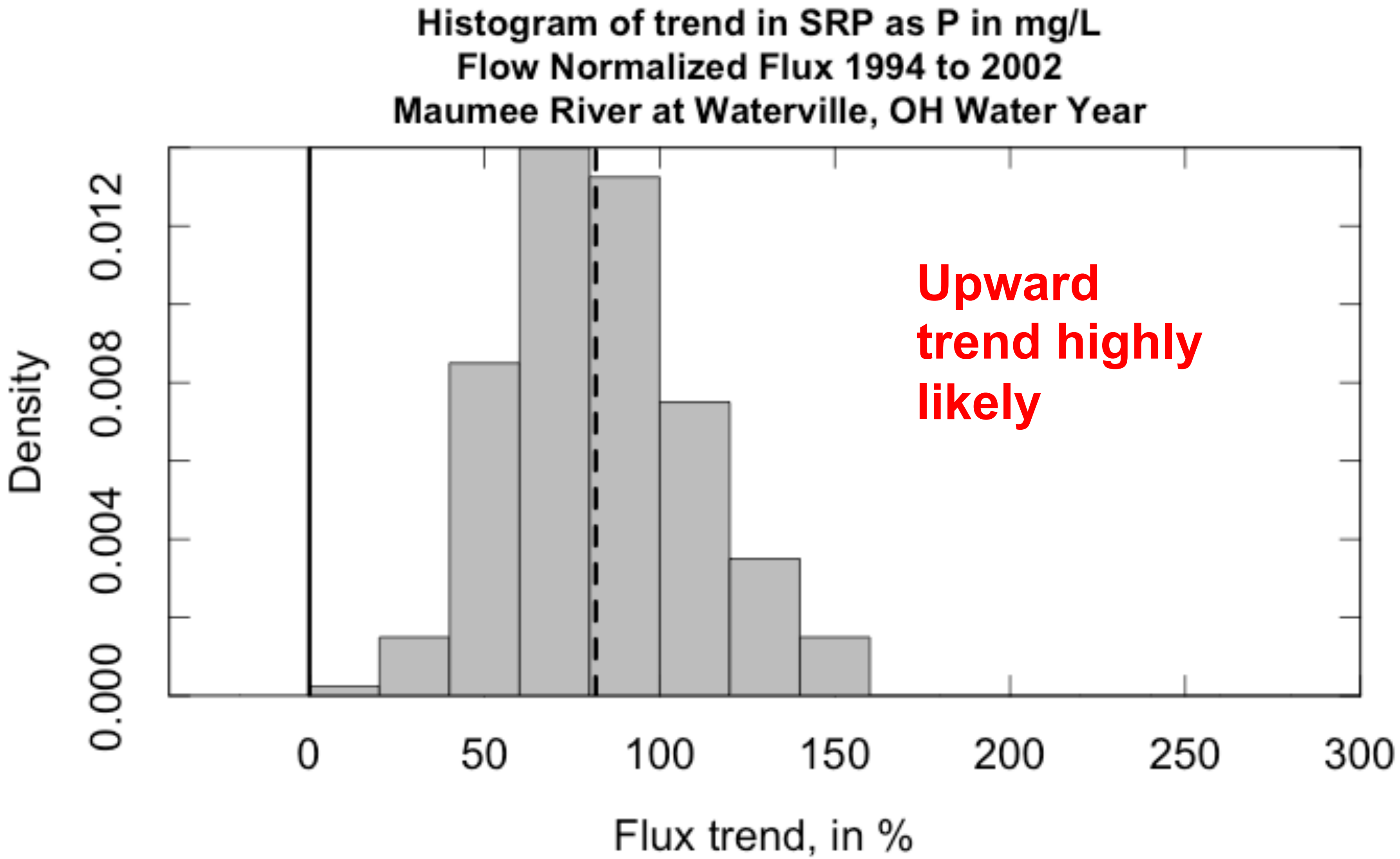
USGS


Histogram of trend in SRP as P in mg/L
Flow Normalized Flux 1994 to 1998
Maumee River at Waterville, OH Water Year
Density
0.000
0.010
0.020
0.030
0
50
100
150
200
250
300
Flux trend, in %
Upward trend about as likely as not

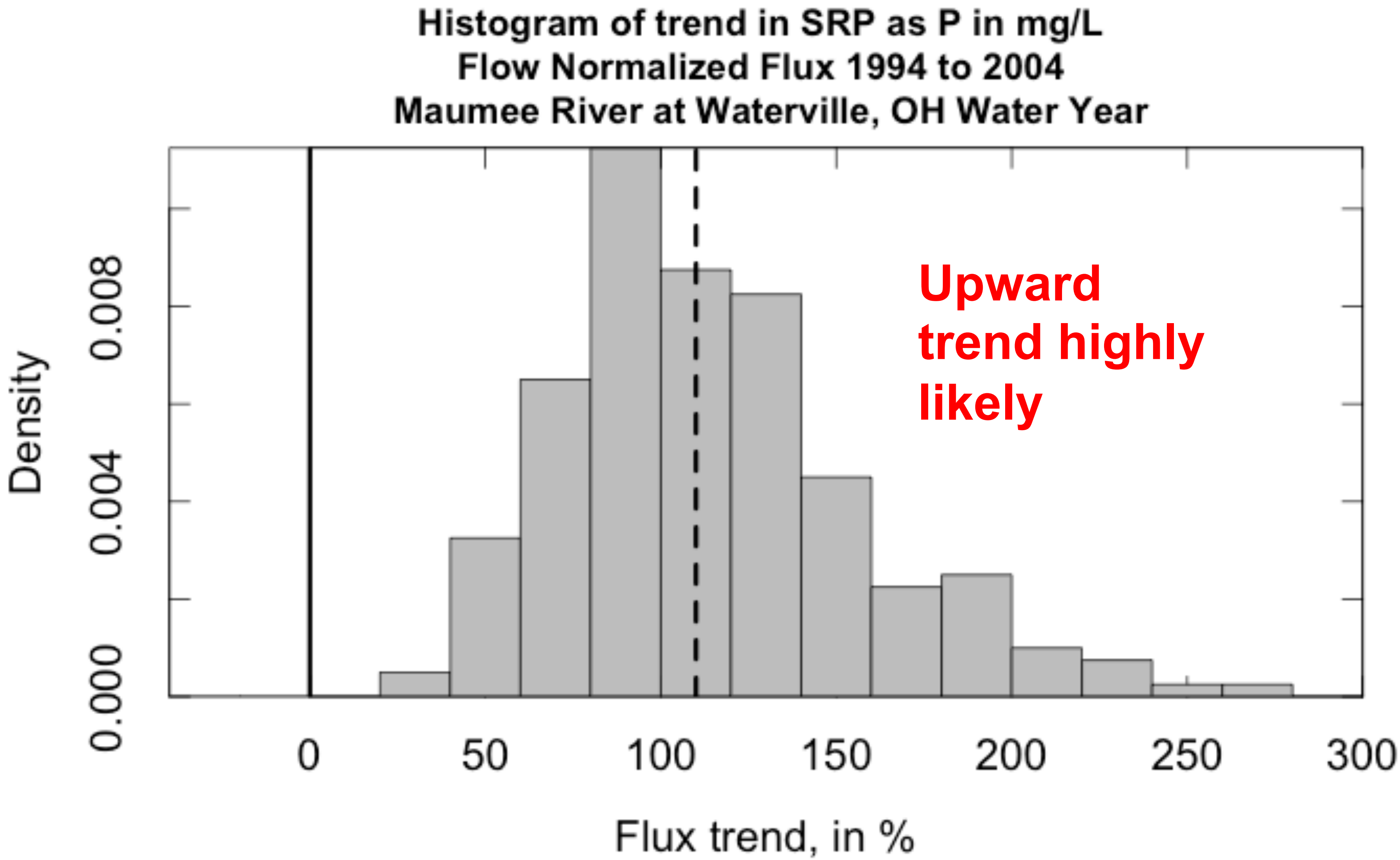
USGS


Histogram of trend in SRP as P in mg/L
Flow Normalized Flux 1994 to 2000
Maumee River at Waterville, OH Water Year
Density
0.000
0.010
0.020
0
50
100
150
200
250
300
Flux trend, in %
Upward
trend very
likely

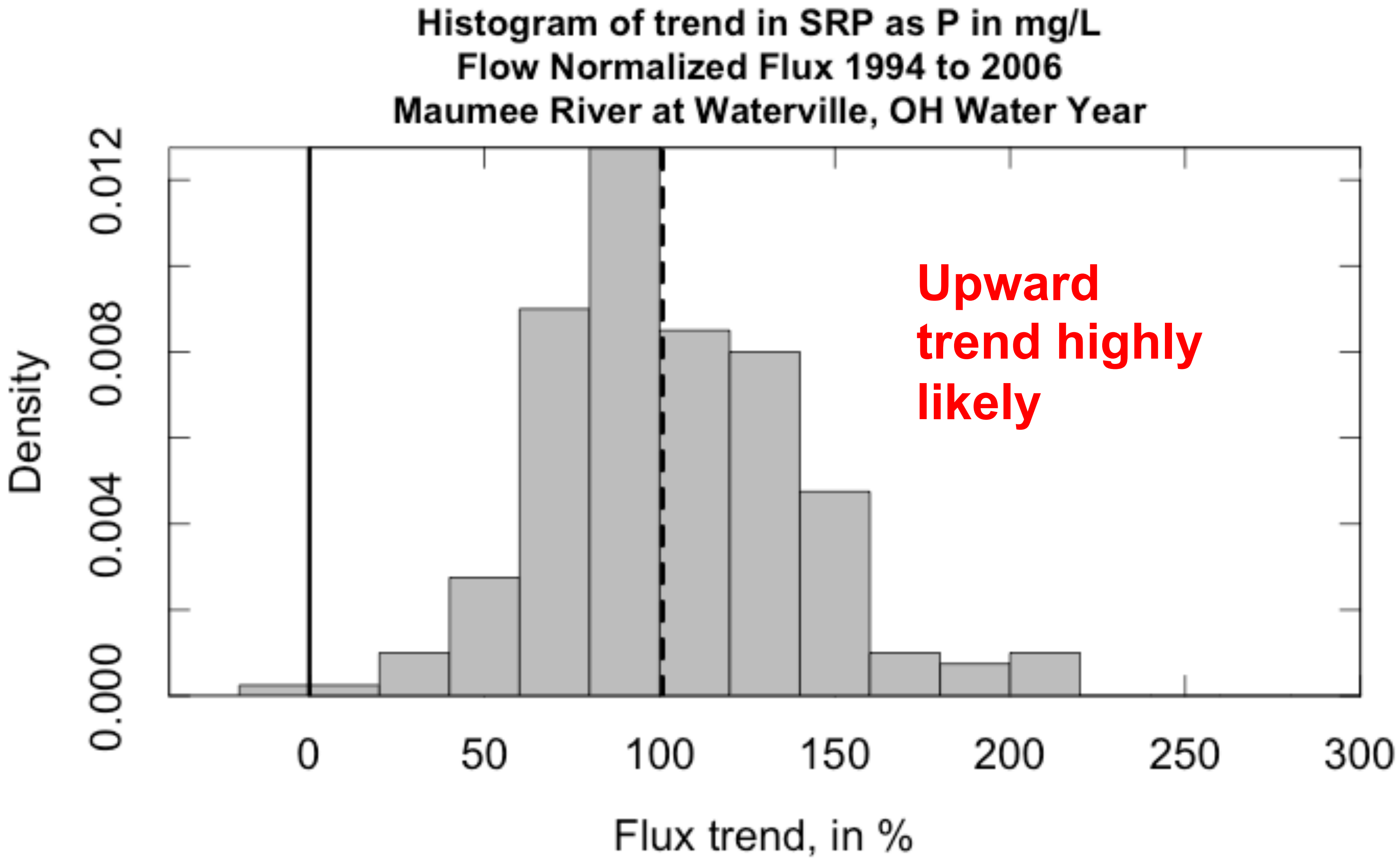
USGS


Histogram of trend in SRP as P in mg/L
Flow Normalized Flux 1994 to 2002
Maumee River at Waterville, OH Water Year
Density
0.000
0.004
0.008
0.012
0
50
100
150
200
250
300
Flux trend, in %
Upward
trend highly
likely

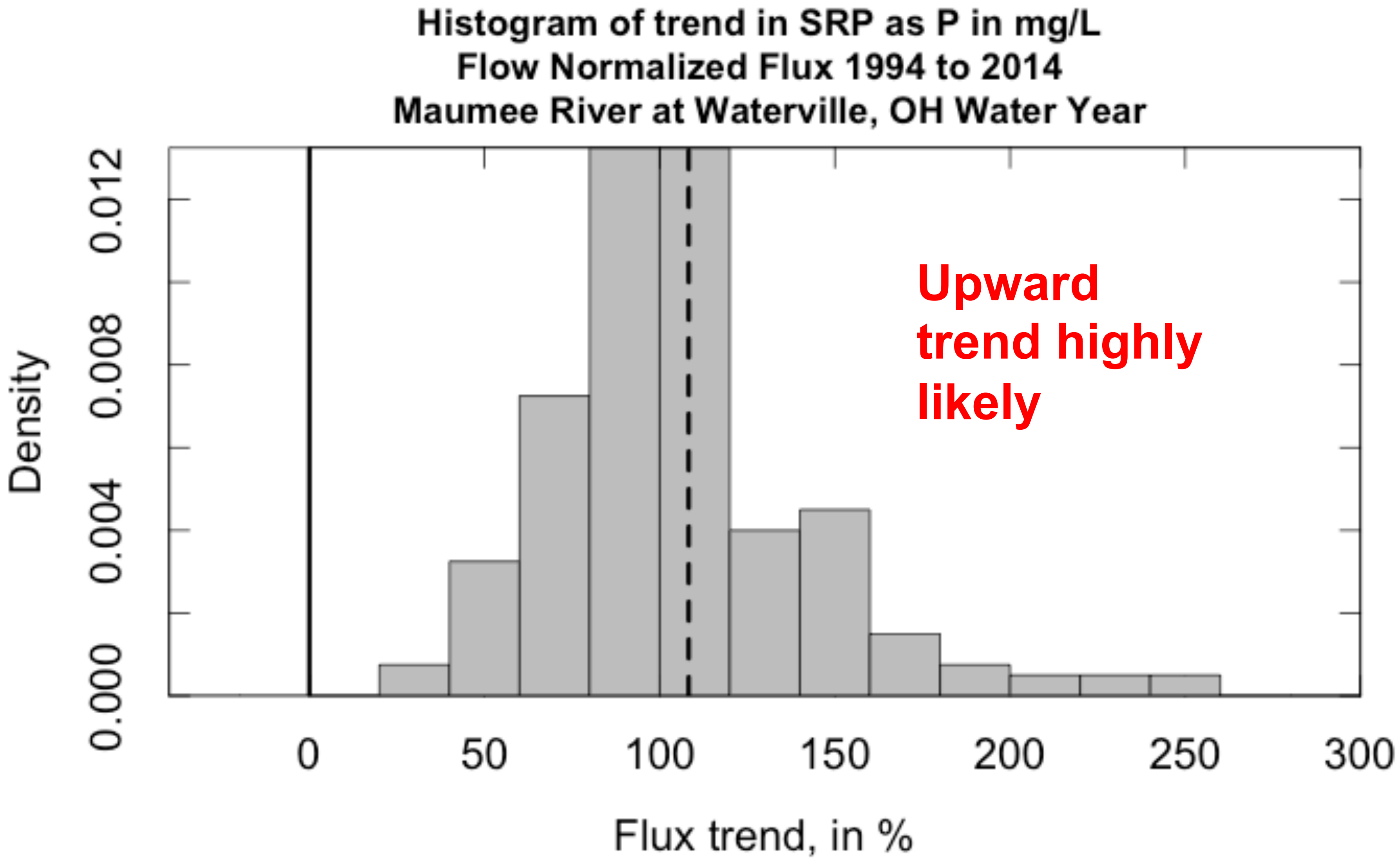
USGS


Histogram of trend in SRP as P in mg/L
Flow Normalized Flux 1994 to 2004
Maumee River at Waterville, OH Water Year
Upward
trend highly
likely
Density
0.000
0.004
0.008
0
50
100
150
200
250
300
Flux trend, in %

USGS


Histogram of trend in SRP as P in mg/L
Flow Normalized Flux 1994 to 2006
Maumee River at Waterville, OH Water Year
Density
0.000
0.004
0.008
0.012
0
50
100
150
200
250
300
Flux trend, in %
Upward trend highly likely

USGS



Histogram of trend in SRP as P in mg/L
Flow Normalized Flux 1994 to 2014
Maumee River at Waterville, OH Water Year
Density
0.000
0.004
0.008
0.012
0
50
100
150
200
250
300
Flux trend, in %
Upward trend highly likely


USGS

# Example of use in a network context

- Shows three categories:
  - Upward (highly likely, very likely, likely)
  - No trend (about as likely as not)
  - Downward (highly likely, very likely, likely)

- Couple that with information about:
  - yield
  - change in yield
  - change in percent

USGS

# Loads to the Chesapeake

- Slides are from Doug Moyer, USGS, Virginia Water Science Center: coordinator of the Non-tidal network for Chesapeake Bay Program.
- Data collection by several agencies
- Load and trend in load results are at http://cbrim.er.usgs.gov/summary.html

USGS

# Total Nitrogen per Acre Loads

Total nitrogen loads range from 1.19 to 33.4 lbs/ac with an average load of 7.33 lbs/ac

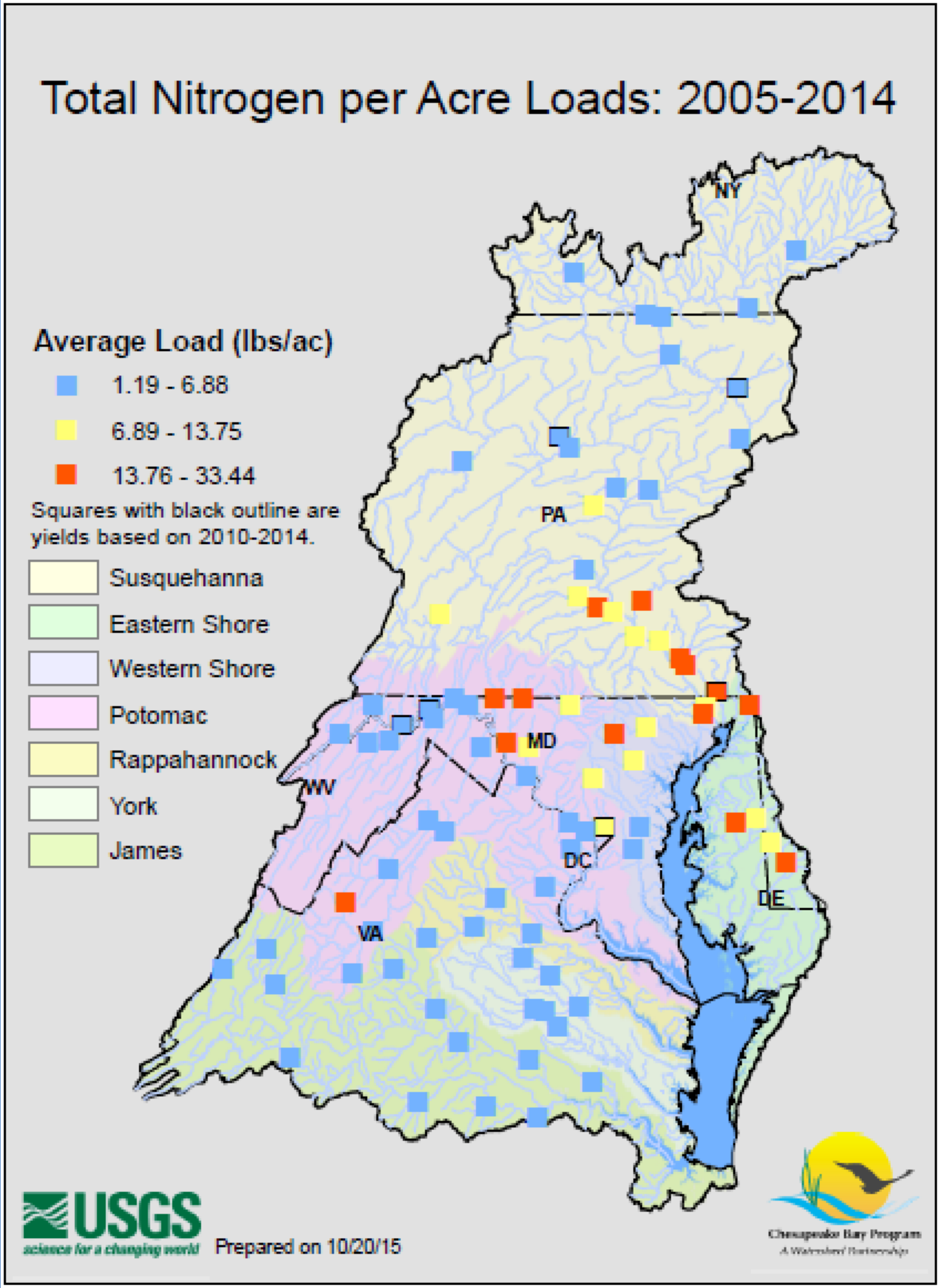
3 Categories of Loads:

(1) Low =
≤ 6.88 lbs/ac
52 of 81 stations

(2) Medium =
> 6.88 to ≤ 13.75
15 of 81 stations

(3) High Yields = ≥ 13.76
14 of 81 stations

USGS



# Total Nitrogen per Acre Loads and Trends: 2005-2014

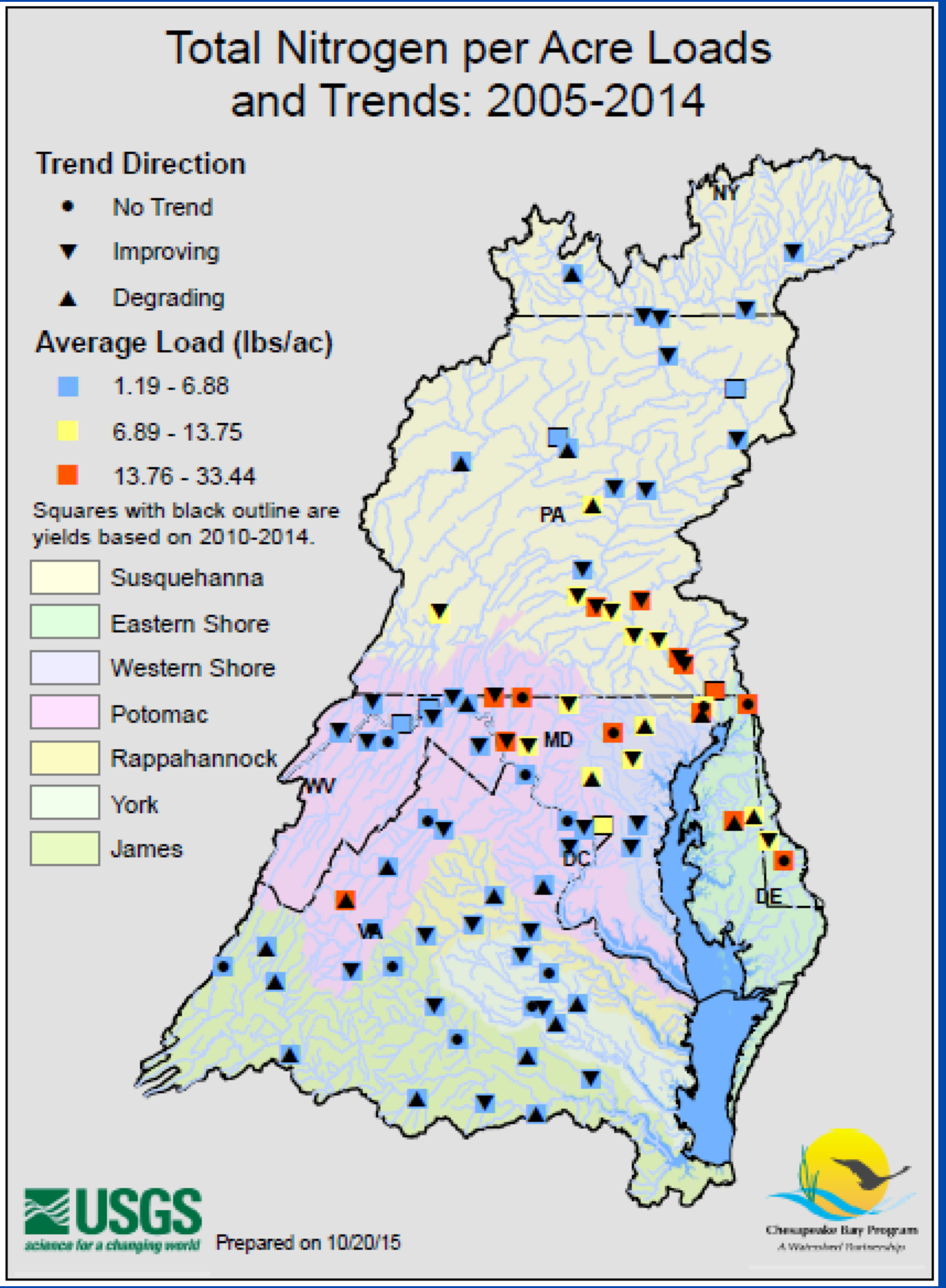
Improving Trends = 44 of 81 (54%)
Degrading Trends = 22 of 81 (27%)
No Trend = 15 of 81 (19%)

Of the 14 stations with the highest per acre loads for Total Nitrogen:

- 6 have improving trends
- 3 have degrading trends
- 4 have no trends
- 1 has insufficient data for trends

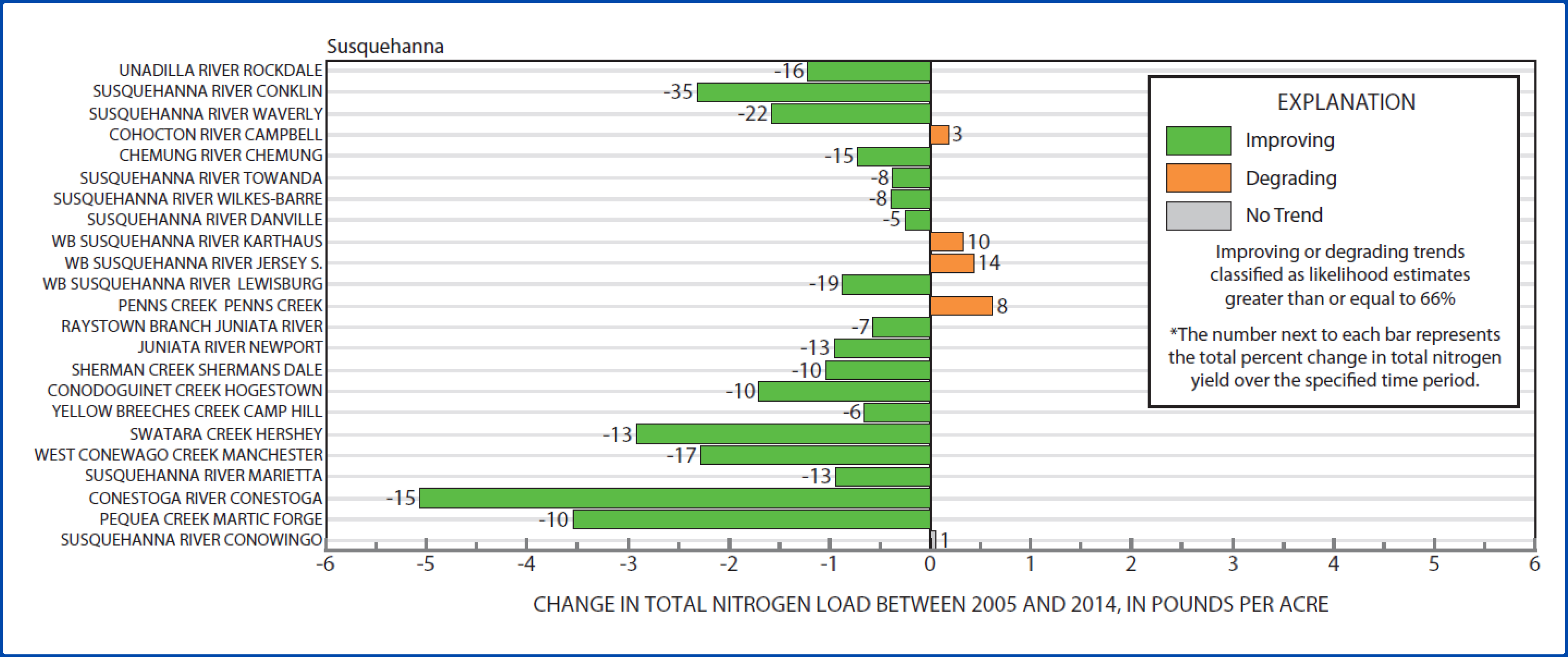
Results by major basins

USGS



# Changes in Nitrogen per Acre Loads: 2005-2014

## *Example from the Susquehanna Watershed*




USGS

# Changes in Nitrogen per Acre Loads: 2005-2014

Trend in load network is the first of its kind

Improving Stations

*Range = -0.10 to -5.07 lbs/ac*

*Median = -0.68 lbs/ac (-10.0%)*

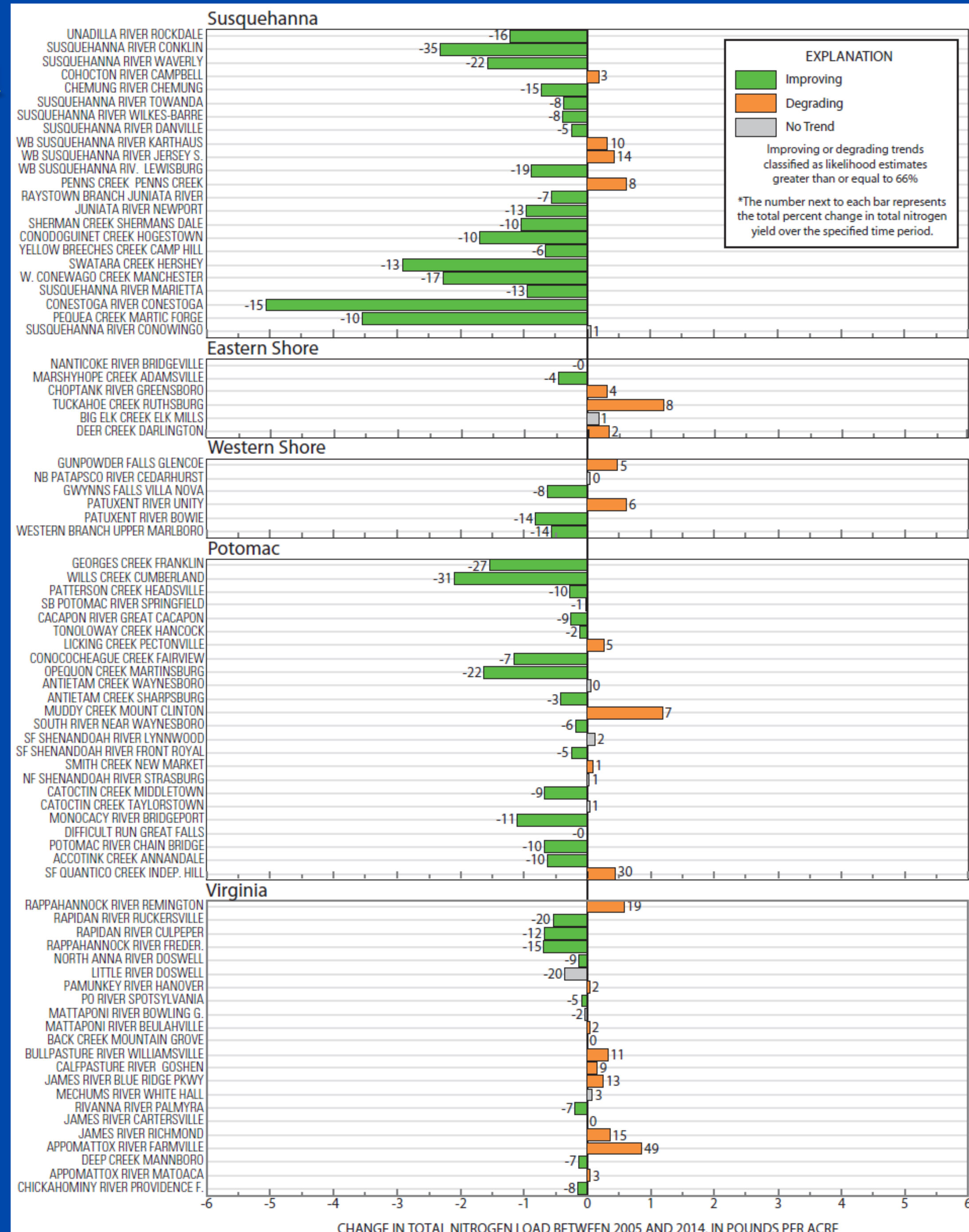
Degrading Stations

*Range = 0.04 to 1.21 lbs/ac*

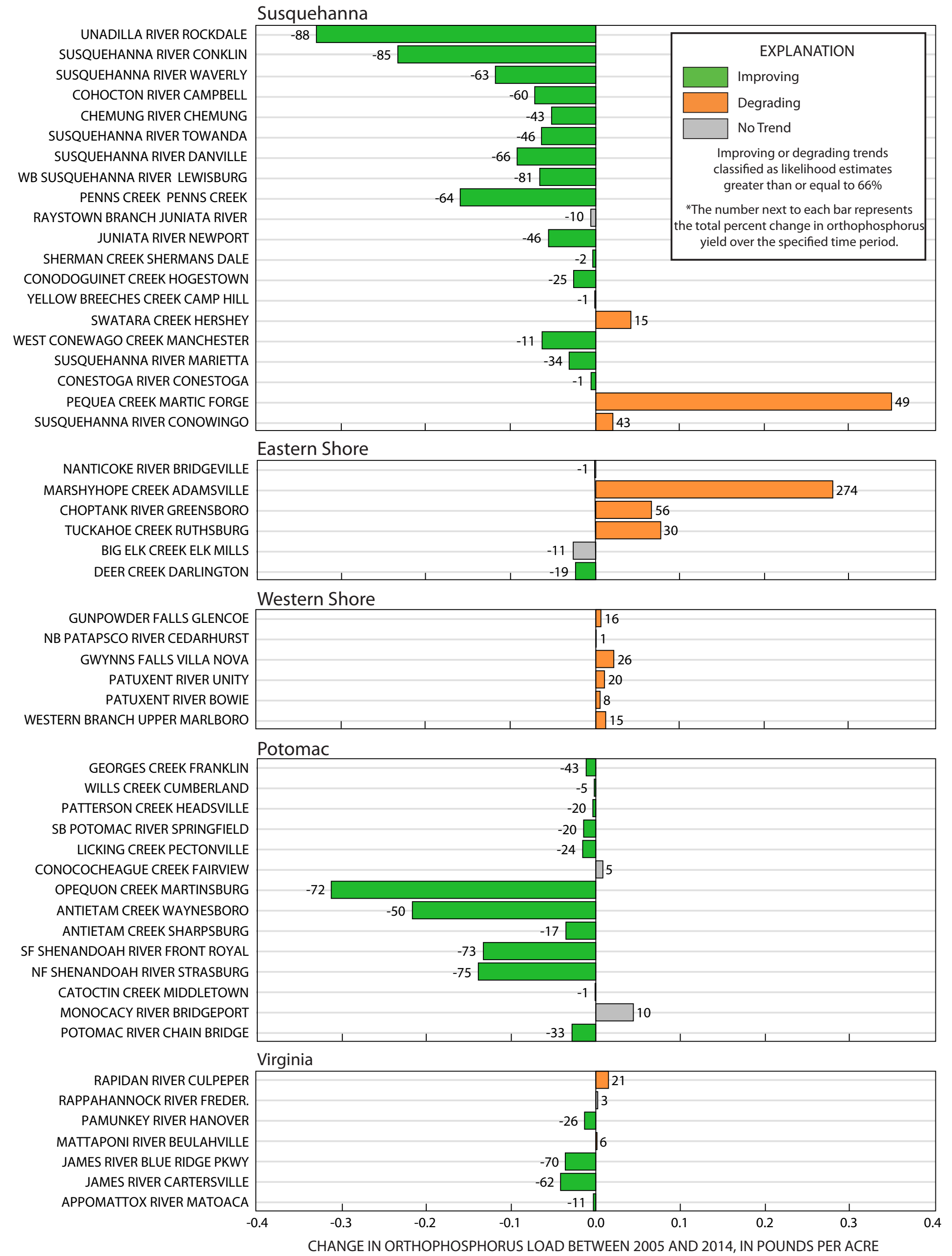
*Median = 0.33 lbs/ac (7.84%)*

Download figure:
http://cbrim.er.usgs.gov/maps.html

USGS



# Changes in Orthophosphorus per Acre Loads: 2005-2014




USGS

USGS
science for a changing world

Intranet Home
Find A Person
Search Intranet

# Interactive Map of USGS Water-Quality Loads and Trends Information in the Chesapeake Bay Watershed



**Trends** ◉ **Loads** ○ **Yields** ○

Constituent: DIP

Time Period: Short Term

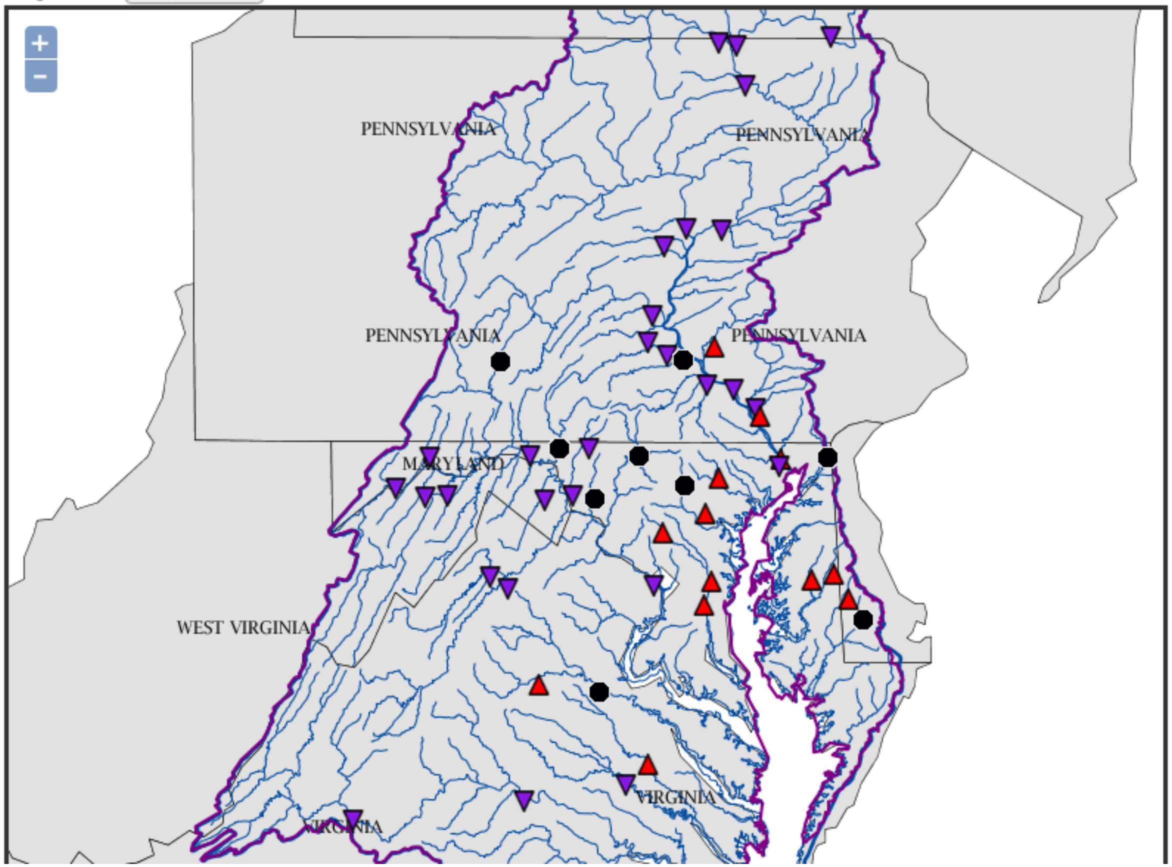
Toggle Station Table

Background: Simple

**Trends Legend**

▼ Improving, Decreasing Load
▲ Degrading, Increasing Load
● No Trend

Click on a station on the map to see more information.
All results presented are through the **2014** water year.




### What Are Trends in Loads?

Trends in sediment and nutrient loads (expressed as yields) describe whether relative water-quality conditions (independent of flow) are improving, degrading, or not changing. The trend results provided on this Web page are our best tool for linking watershed management to water-quality change. Trends are computed for:

- Short Term: The last 10 years of record (2005-2014) for each site; and
- Long Term: The period of record for each station having more than 25 years of data.

USGS

USGS science for a changing world

Intranet Home
Find A Person
Search Intranet

# Interactive Map of USGS Water-Quality Loads and Trends Information in the Chesapeake Bay Watershed

**Navigate Menu**

Home

**Background**

Introduction

Methods

Glossary

Bibliography

**Results and Maps**

Load and Trend Summary

Interactive Map

Tables and Figures

**Downloads**

Model Input Data

Load Table

Yield Table

Trend Results Table

**Contacts**

Partners

USGS Chesapeake Activities

Technical Contacts

Web Administrator

**Trends** (selected) **Loads** **Yields**

Constituent: DIP

Time Period: Short Term

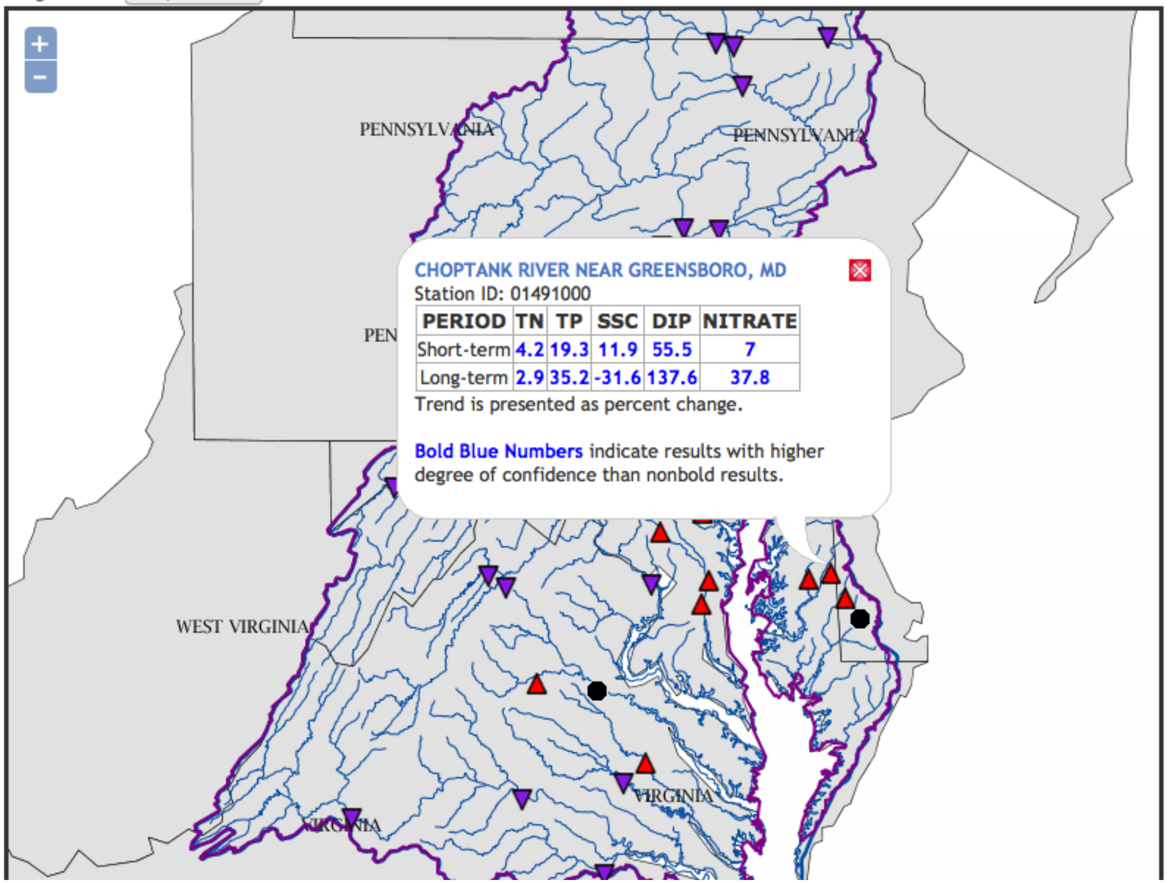
Toggle Station Table

Background: Simple

**Trends Legend**

- ▼ Improving, Decreasing Load
- ▲ Degrading, Increasing Load
- ● No Trend

Click on a station on the map to see more information.
All results presented are through the **2014** water year.




CHOPTANK RIVER NEAR GREENSBORO, MD
Station ID: 01491000

| PERIOD | TN | TP | SSC | DIP | NITRATE |
|---|---|---|---|---|---|
| Short-term | **4.2** | **19.3** | **11.9** | **55.5** | **7** |
| Long-term | **2.9** | **35.2** | **-31.6** | **137.6** | **37.8** |

Trend is presented as percent change.

**Bold Blue Numbers** indicate results with higher degree of confidence than nonbold results.

## What Are Trends in Loads?

Trends in sediment and nutrient loads (expressed as yields) describe whether relative water-quality conditions (independent of flow) are improving, degrading, or not changing. The trend results provided on this Web page are our best tool for linking watershed management to water-quality change. Trends are computed for:

- Short Term: The last 10 years of record (2005-2014) for each site; and
- Long Term: The period of record for each station having more than 25 years of data.

USGS

- WRTDS Bootstrap Test provides a measure of uncertainty in trend magnitudes, robust to:
  - Non-linear and non-monotonic trends
  - Changing seasonal and discharge relationships
  - Serial correlation
- Flexible options for expressing uncertainty

USGS

# 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?*
- https://github.com/USGS-R/EGRET/issues

USGS

# Information about EGRET

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

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

"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."

USGS