



New Tools for Water Quality Data Access, Trend and Load Analysis

An overview of the USGS R Packages:

dataRetrieval and EGRET

Robert M. Hirsch

USGS

2014-10-07

This information is preliminary and is subject to revision. It is being provided to meet the need for timely best science. The information is provided on the condition that neither the U.S. Geological Survey nor the U.S. Government may be held liable for any damages resulting from the authorized or unauthorized use of the information.

dataRetrieval

Retrieves water data for use in R

Some functions designed specifically as inputs to EGRET

Preliminary Information-Subject to Revision. Not for Citation or Distribution

EGRET: Exploration & Graphics for RivEr Trends

Flow history analysis

Water quality graphics

WRTDS: Weighted Regressions on Time Discharge & Season: for water quality trends and fluxes

WRTDS for exploration

These packages have been a joint effort for 3 years

- Laura De Cicco of the CIDA group in Wisconsin is the primary author of dataRetrieval and has made huge contributions to all aspects, including trying to retrain a “Paleo Code Writer”
- Also thanks to our colleague reviewers: Jeff Chanat, VA WSC and Jessica Thompson, WI WSC
- Crucial support from OWQ, NAWQA, OWI, and CIDA

Preliminary Information-Subject
to Revision. Not for Citation or
Distribution

Outline

Motivations for the packages

The WRTDS concept and examples of results

Overview of dataRetrieval

How EGRET works, doing WRTDS analysis and producing graphs and tables

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

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

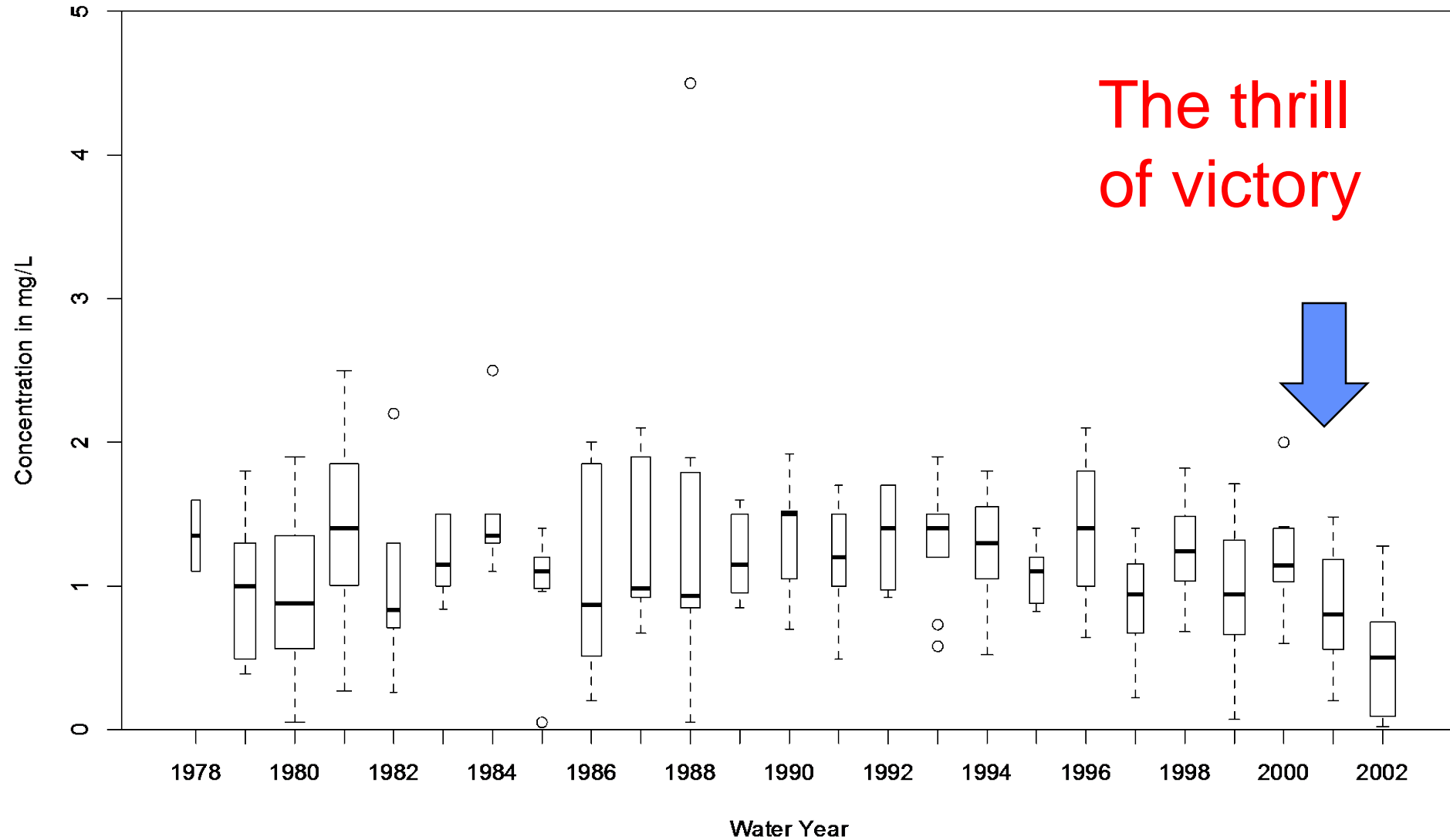
**How do we come to understand what
is happening to water quality in
large watersheds.**

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?**

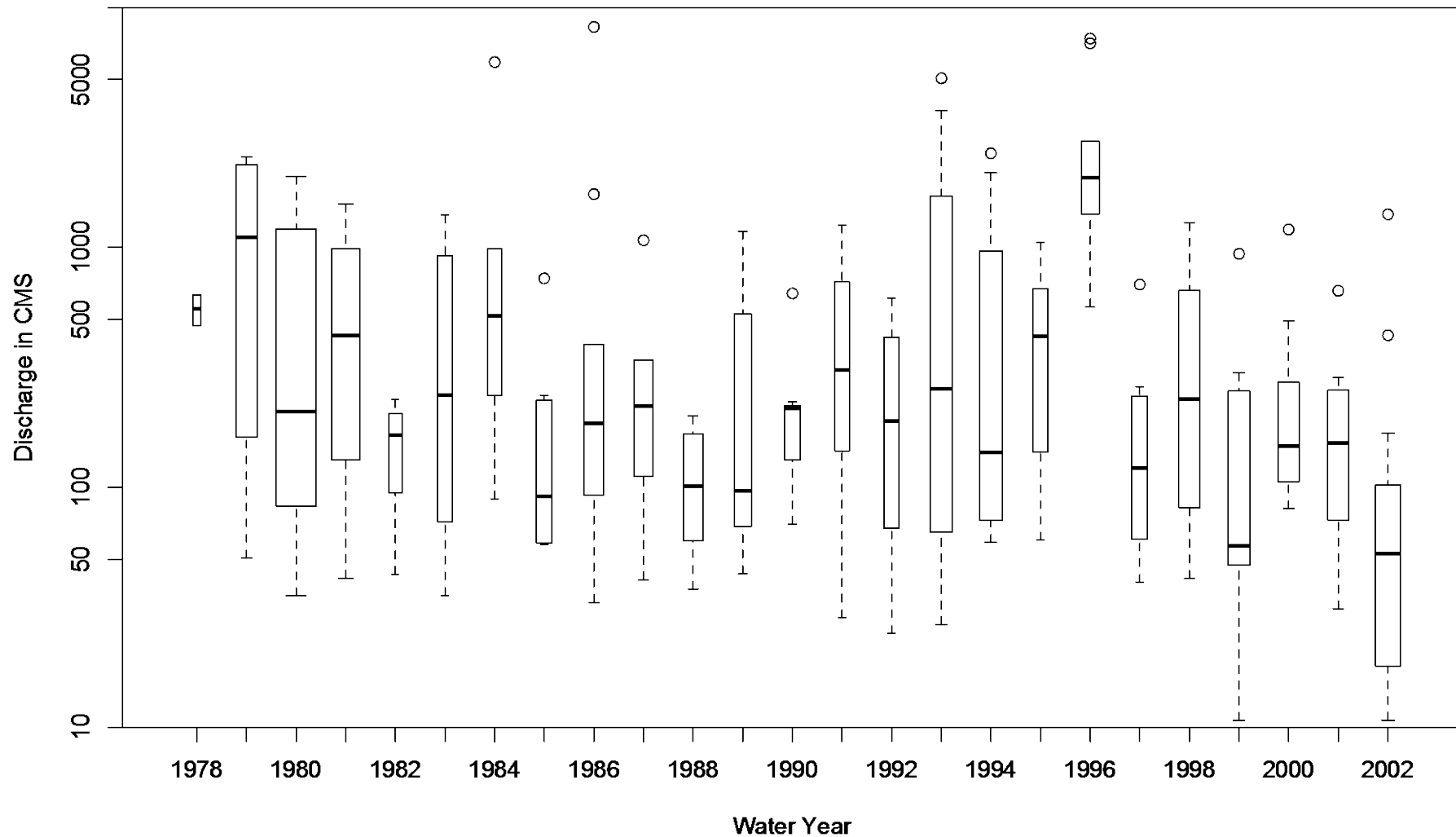
Potomac River at Chain Bridge, Washington DC
Box plot of sample values by Water Year
Nitrate as N



Data through September 2002

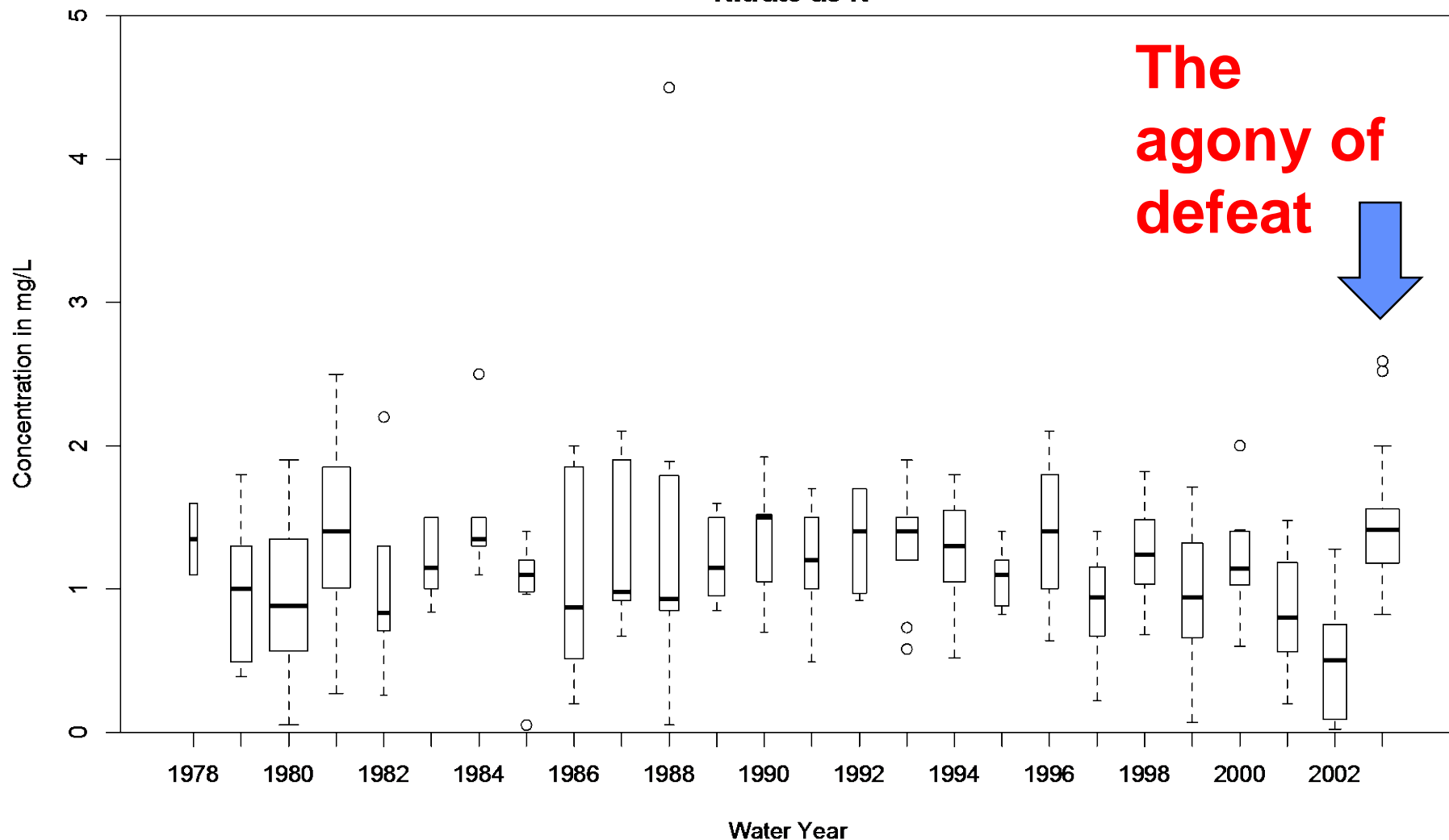
Preliminary Information-Subject
to Revision. Not for Citation or
Distribution

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



Preliminary Information-Subject
to Revision. Not for Citation or
Distribution

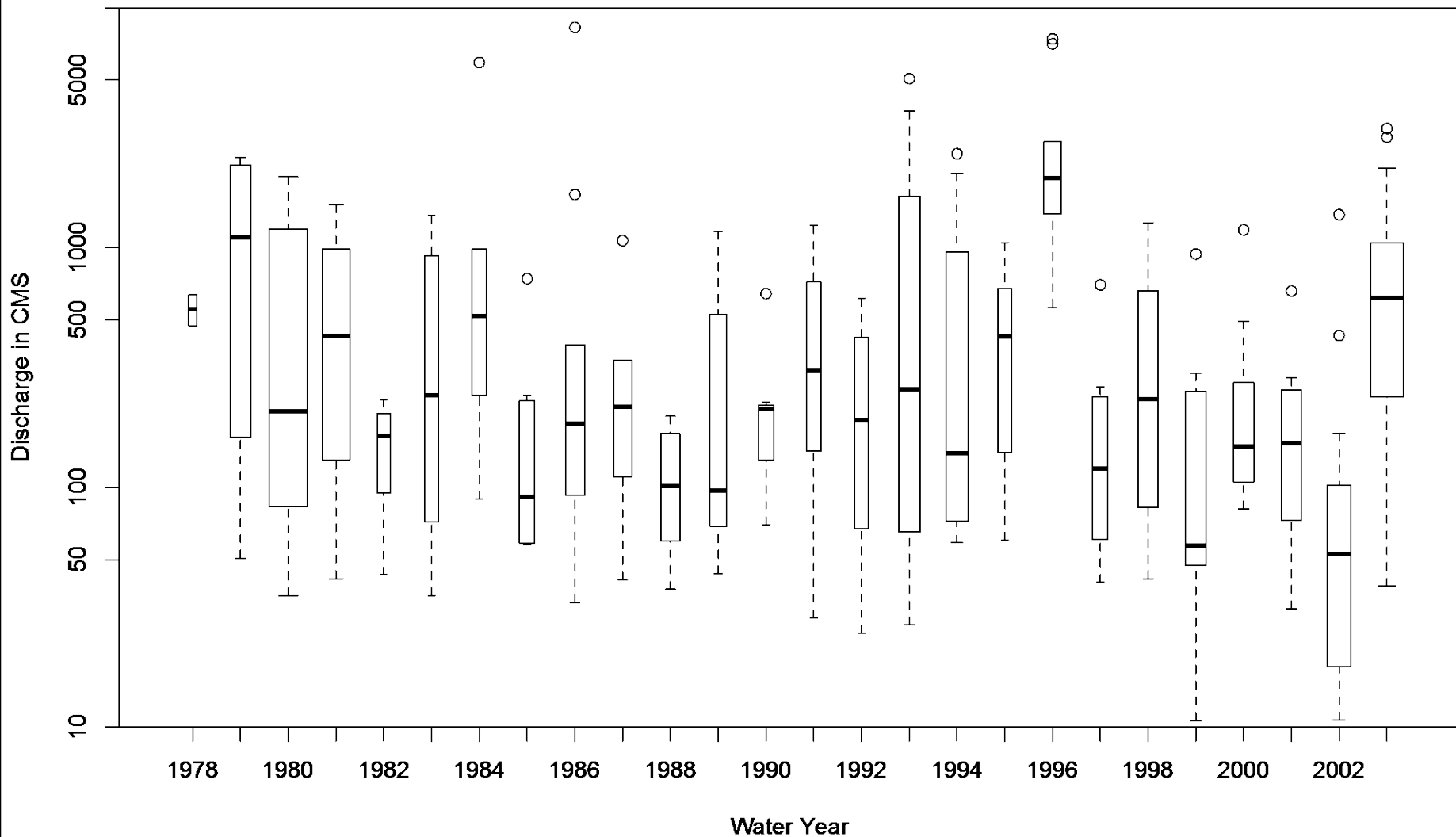
Potomac River at Chain Bridge, Washington DC
Box plot of sample values by Water Year
Nitrate as N



Data through September 2003

Preliminary Information-Subject
to Revision. Not for Citation or
Distribution

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



Preliminary Information-Subject
to Revision. Not for Citation or
Distribution

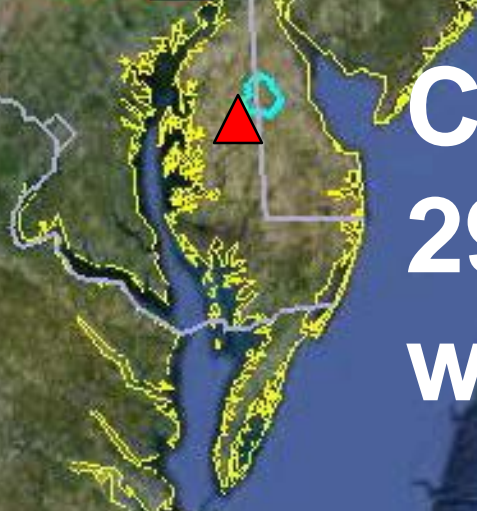
Motivations for the method

- 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.

Data requirements

Preliminary Information-Subject to Revision. Not for Citation or Distribution

- 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.

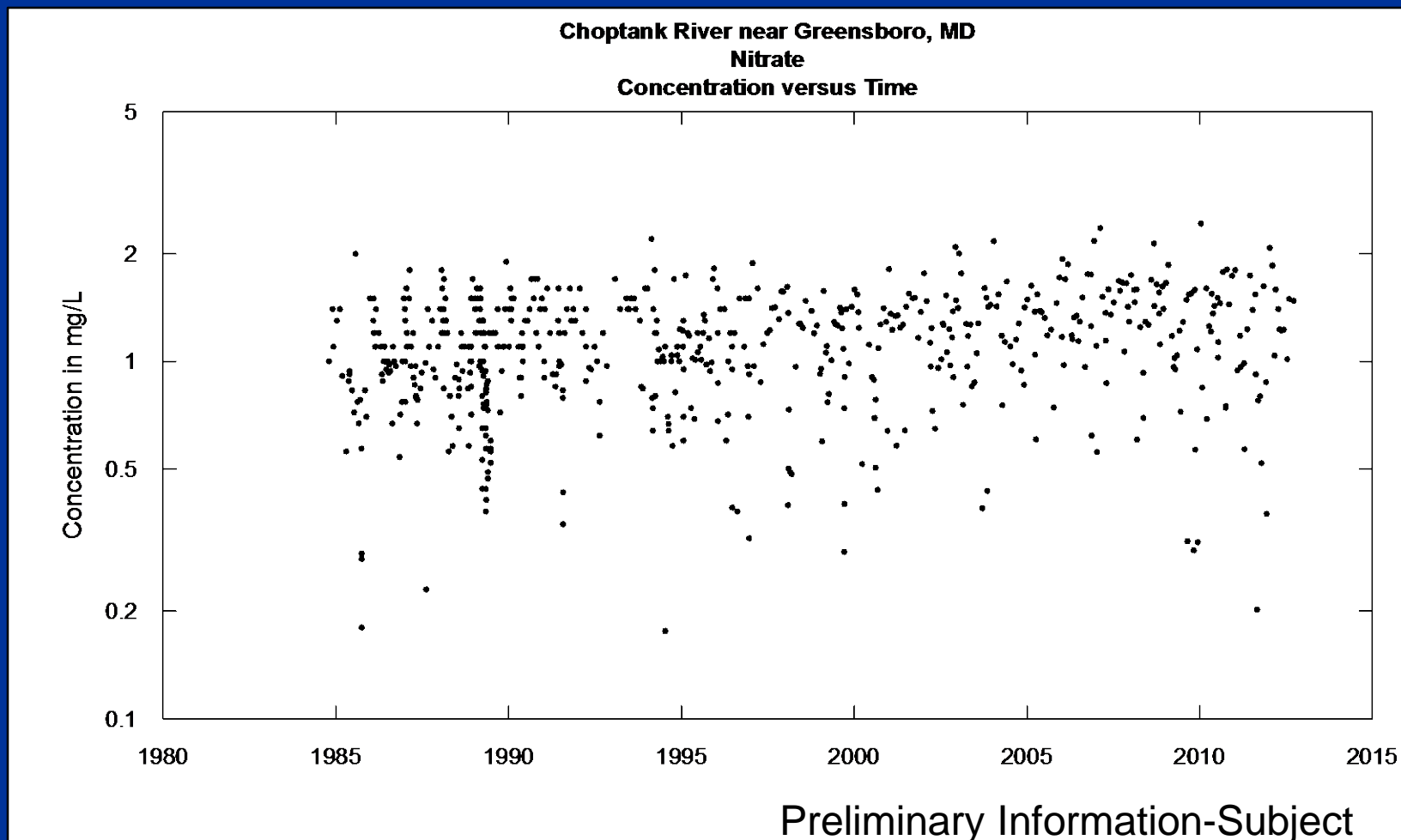


**Choptank,
293 km²
watershed**

Preliminary Information-Subject
to Revision. Not for Citation or
Distribution

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

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



Preliminary Information-Subject
to Revision. Not for Citation or
Distribution

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)**

Locally Weighted Regression

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

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

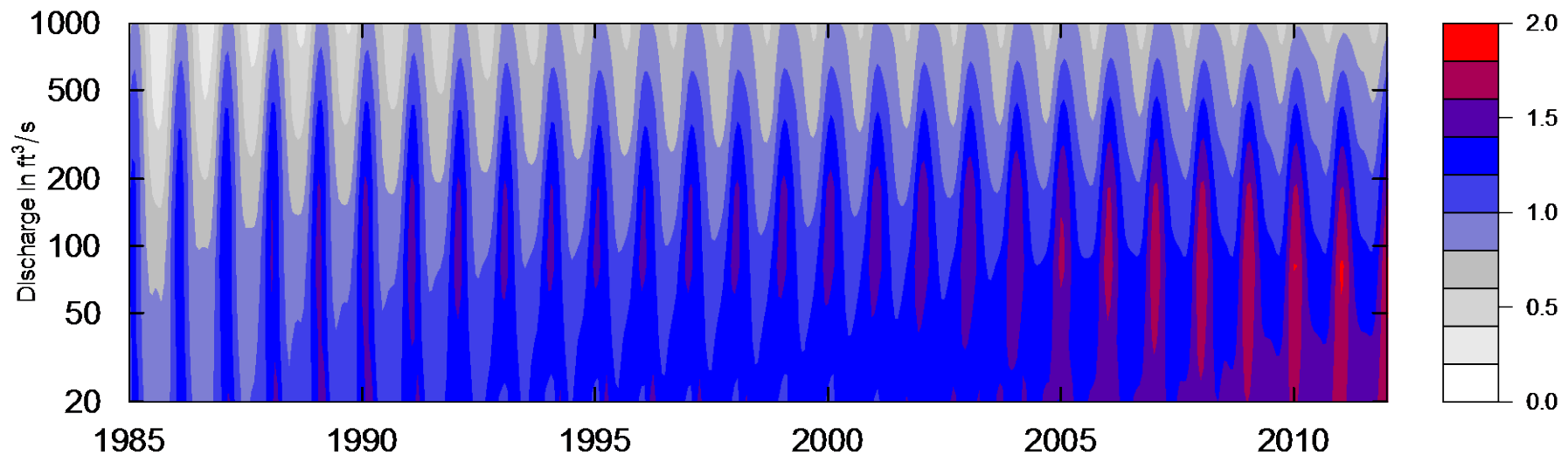
But the coefficients should be smoothly changing as we move through the space

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.

Preliminary Information-Subject to
Revision. Not for Citation or Distribution

WRTDS view of the evolving behavior of nitrate

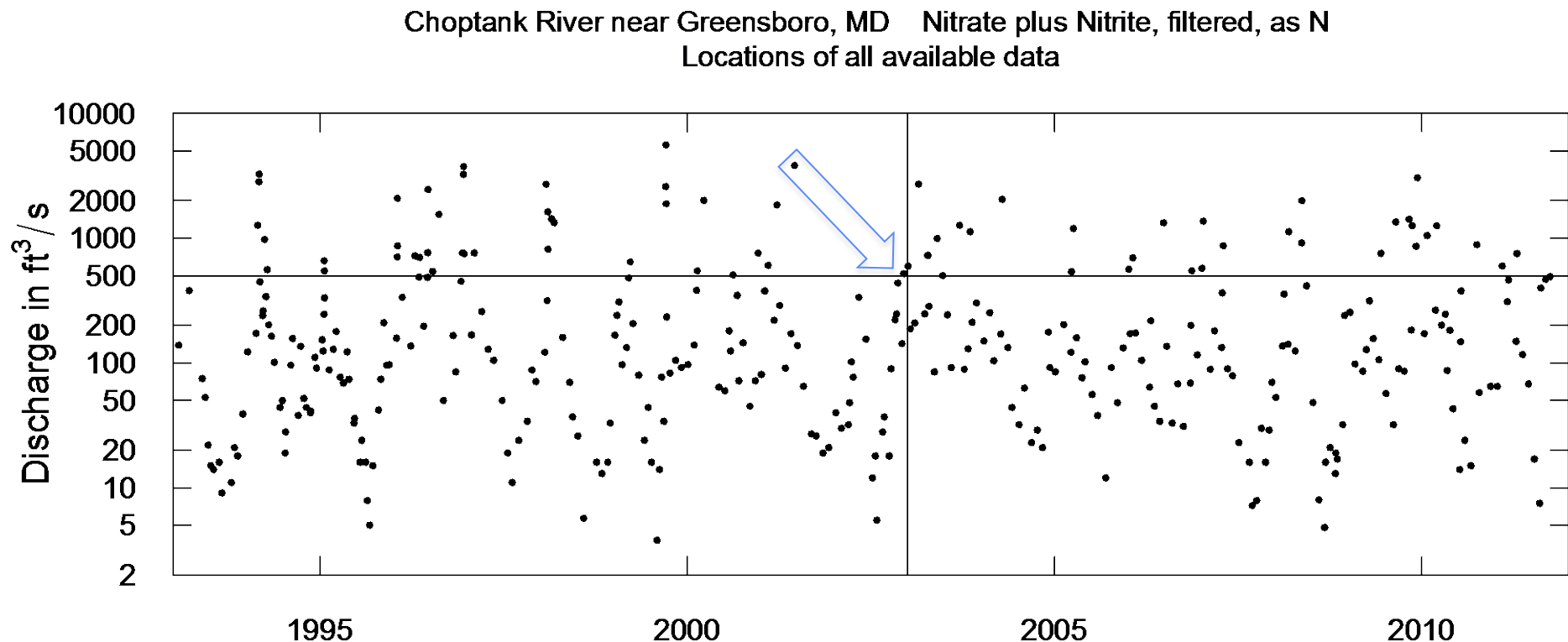
Choptank River near Greensboro, MD Nitrate plus Nitrite, Filtered, as N
Estimated Concentration Surface in Color



How is this surface created?

Every dot is a data point from 1993 to 2012

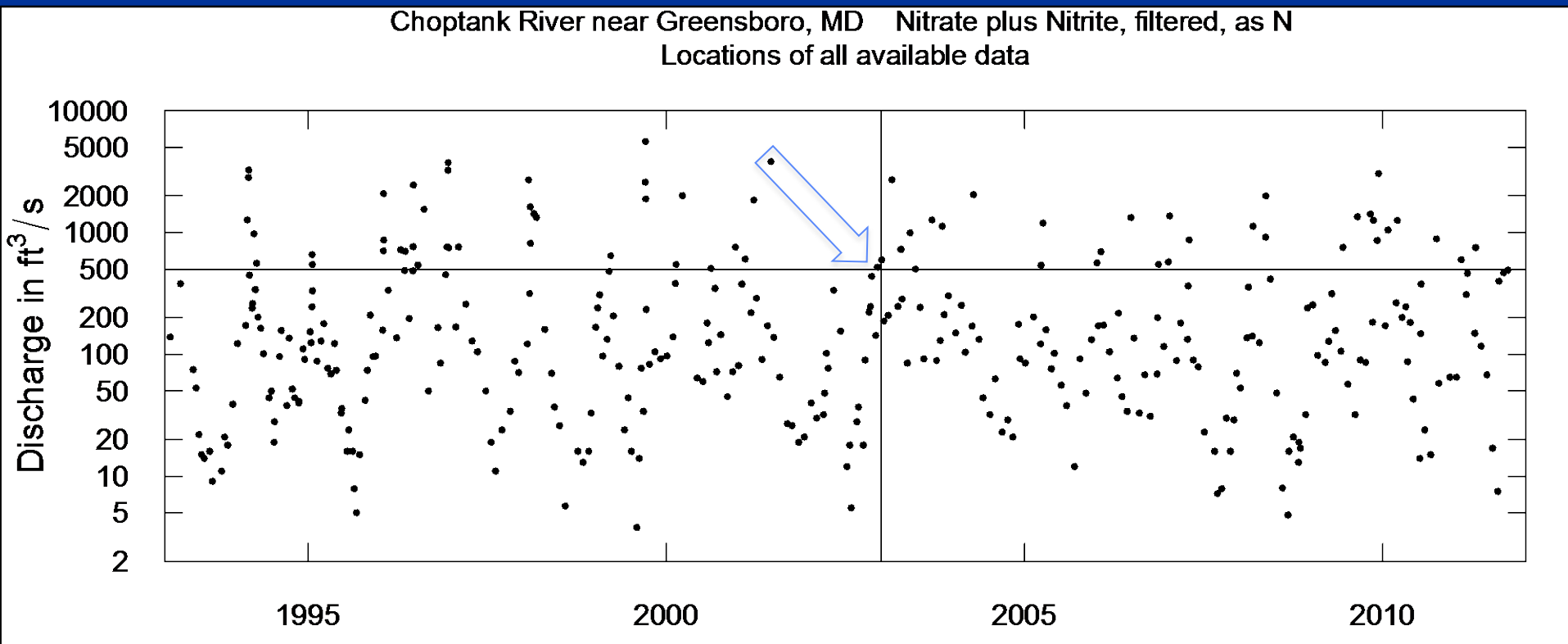
Let's say we want to use the data to estimate the expected value of concentration for January 1, 2003 at Q=500 cfs



Preliminary Information-Subject
to Revision. Not for Citation or
Distribution

The principle is this:

Do a weighted regression at this point. The weights on each observation are related to their “distance”

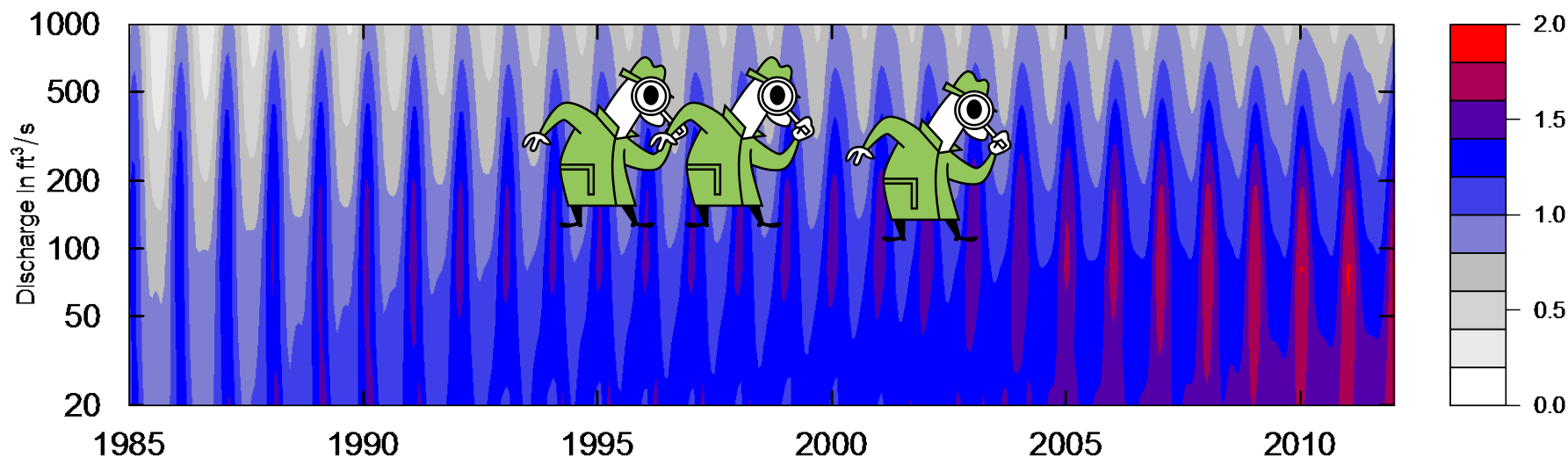


Distance in time, in $\log(Q)$, and season.

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

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

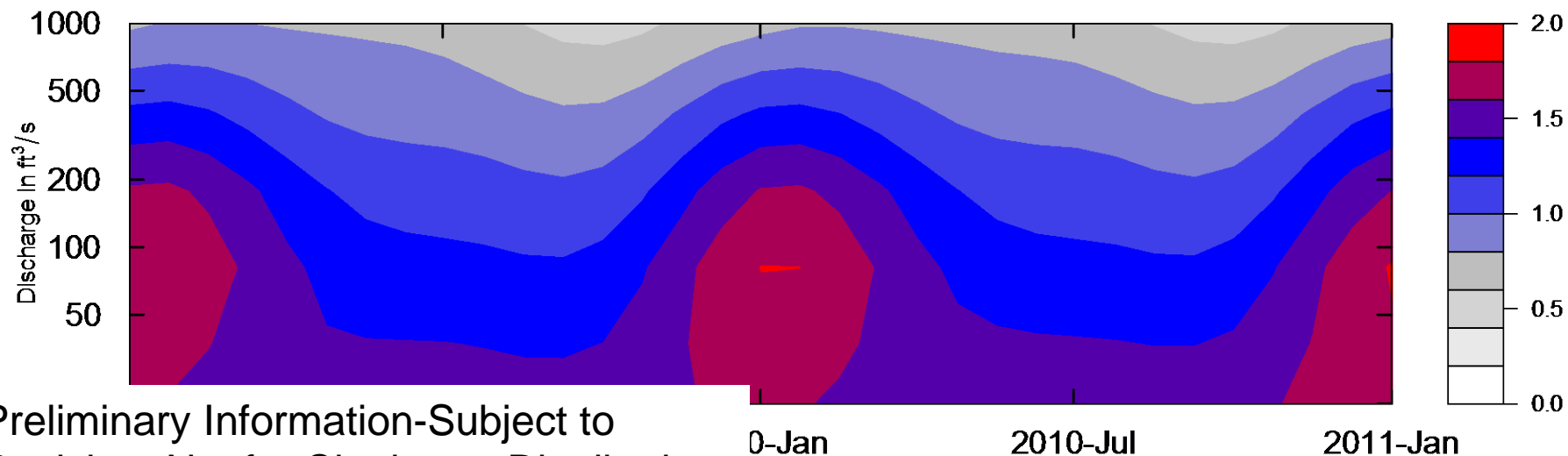
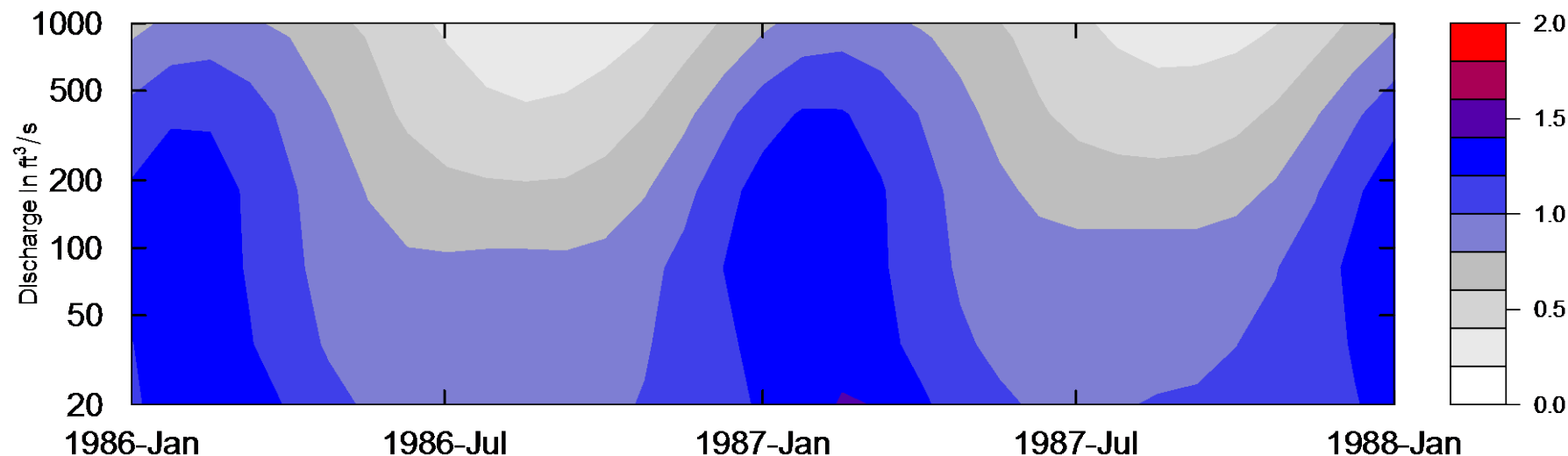
**Choptank River near Greensboro, MD Nitrate plus Nitrite, Filtered, as N
Estimated Concentration Surface in Color**



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

Choptank River near Greensboro, MD Nitrate plus Nitrite, Filtered, as N
Estimated Concentration Surface in Color



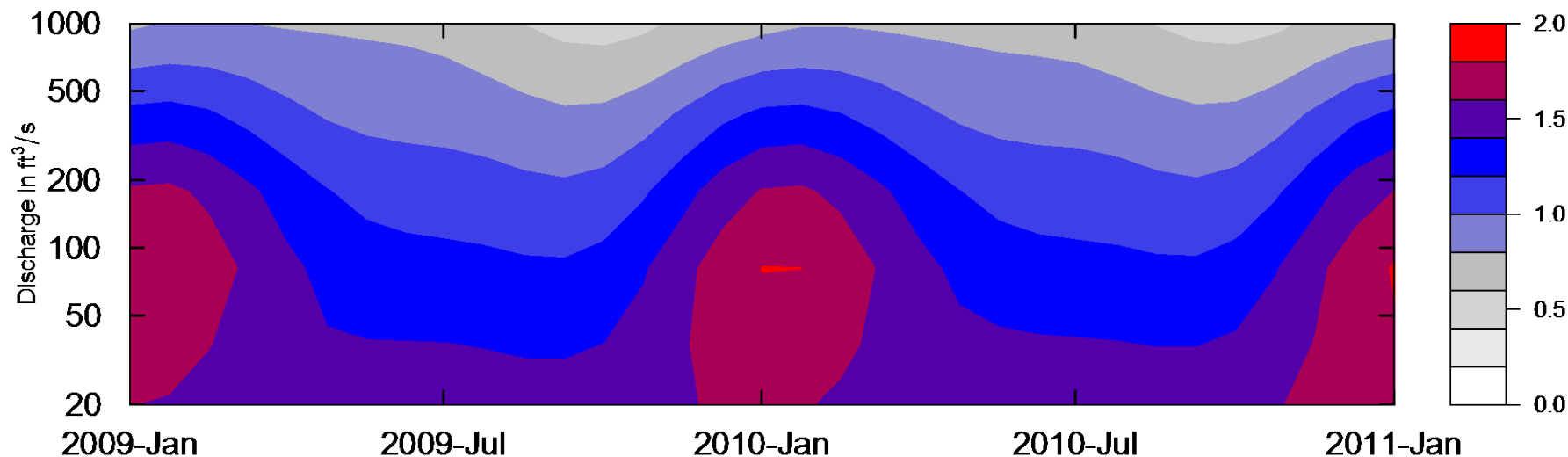
Preliminary Information-Subject to
Revision. Not for Citation or Distribution

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.

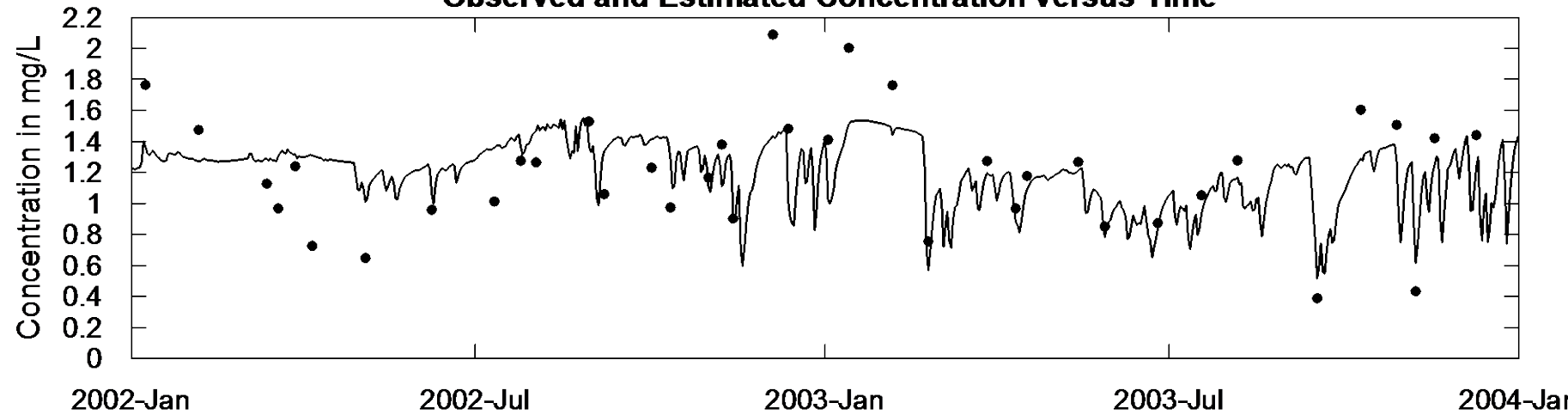
Preliminary Information-Subject to
Revision. Not for Citation or Distribution



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

**Choptank River near Greensboro, MD
Nitrate**

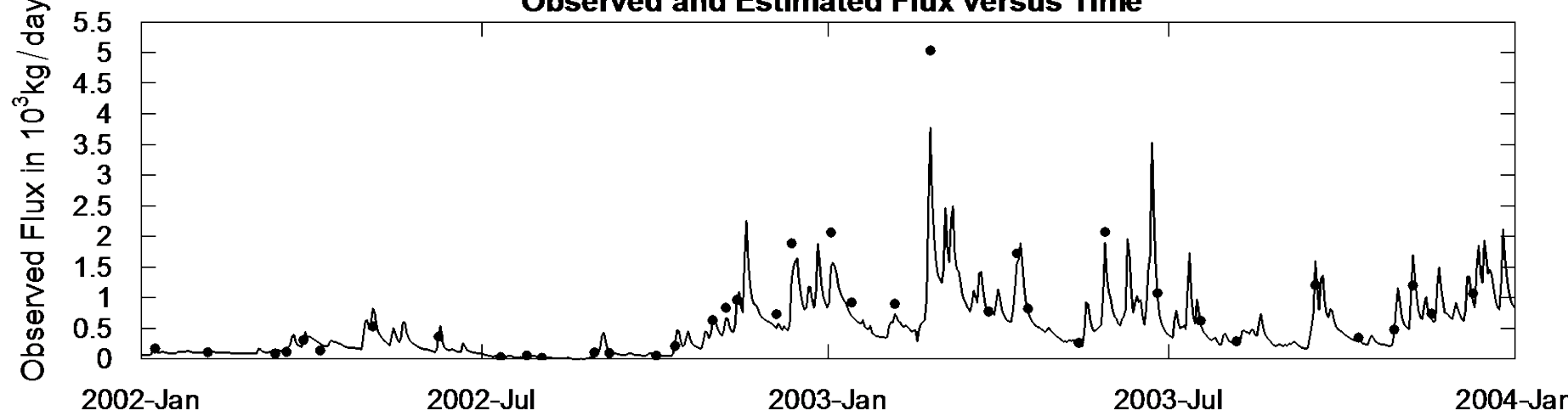
Observed and Estimated Concentration versus Time



Preliminary Information-Subject to
Revision. Not for Citation or Distribution

**Choptank River near Greensboro, MD
Nitrate**

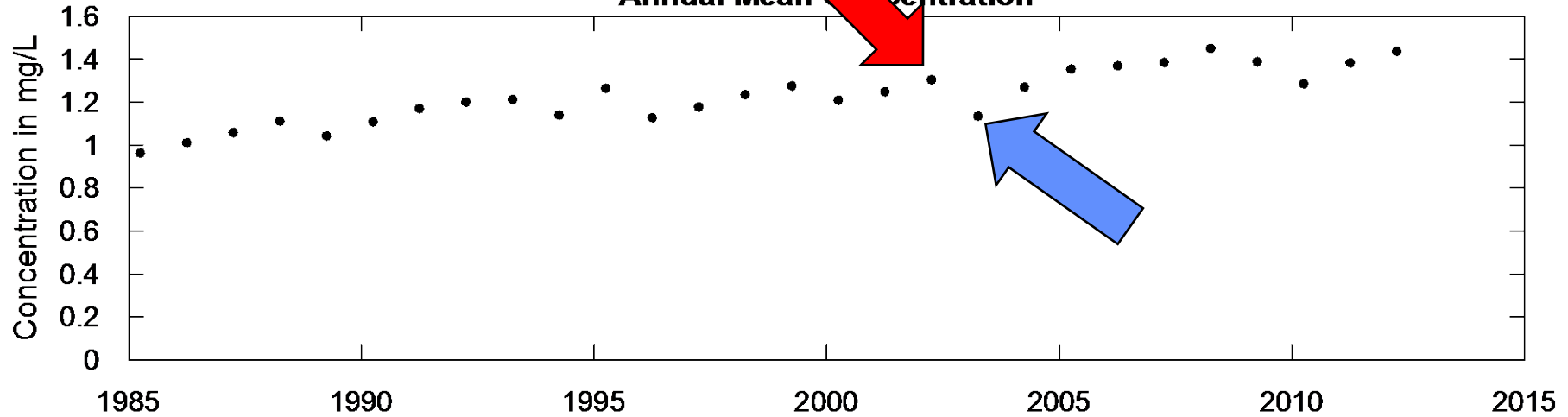
Observed and Estimated Flux versus Time



Choptank River near Greensboro, MD Nitrate

Water Year

Annual Mean Concentration

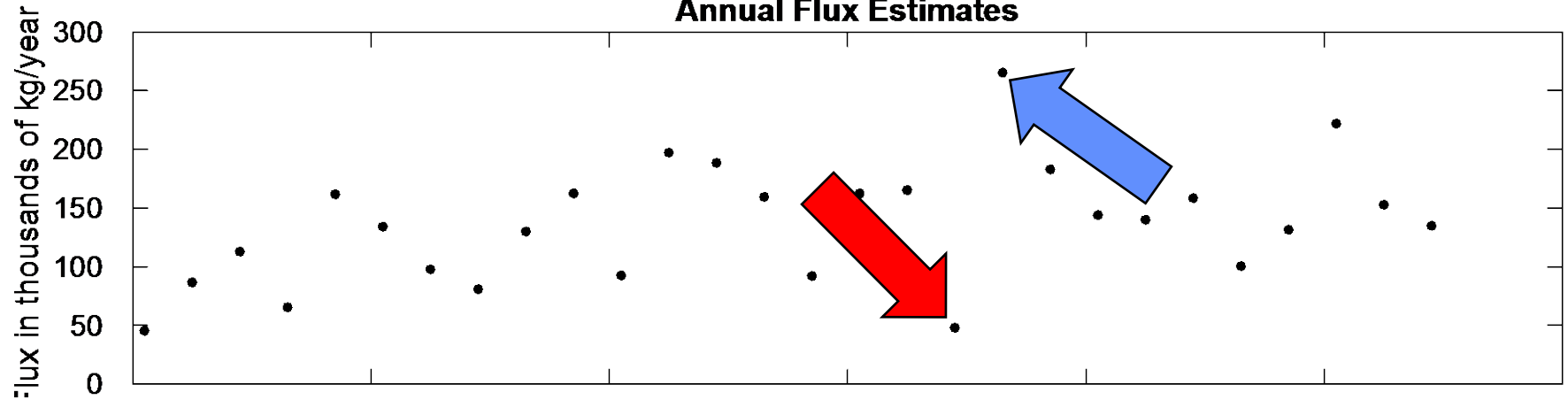


Preliminary Information-Subject to
Revision. Not for Citation or Distribution

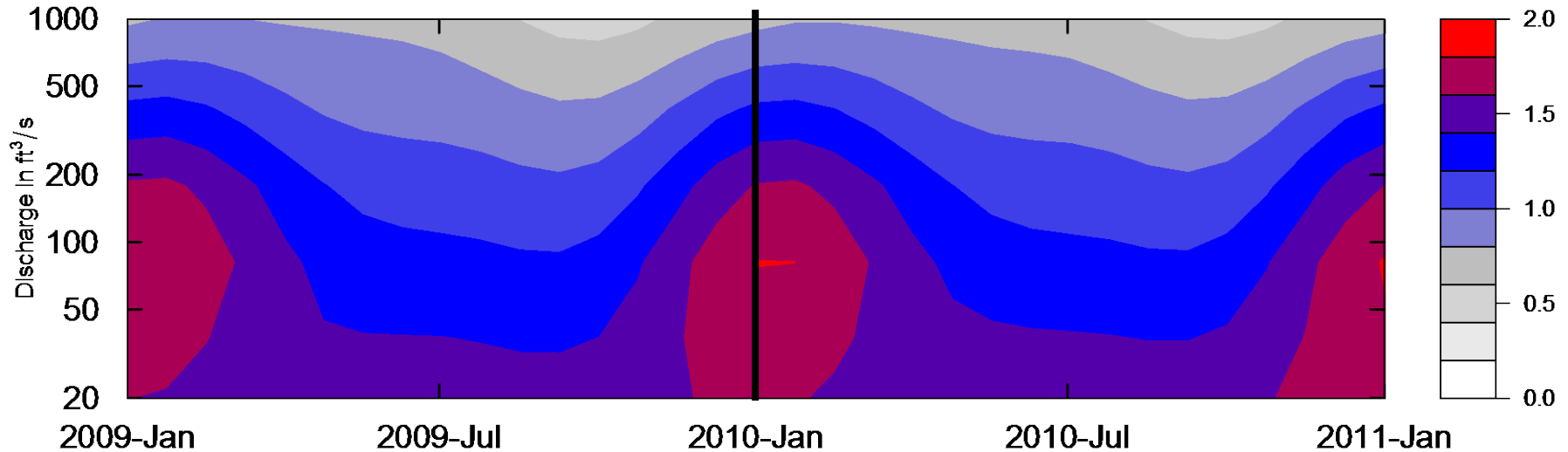
Choptank River near Greensboro, MD Nitrate

Water Year

Annual Flux Estimates



**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.**

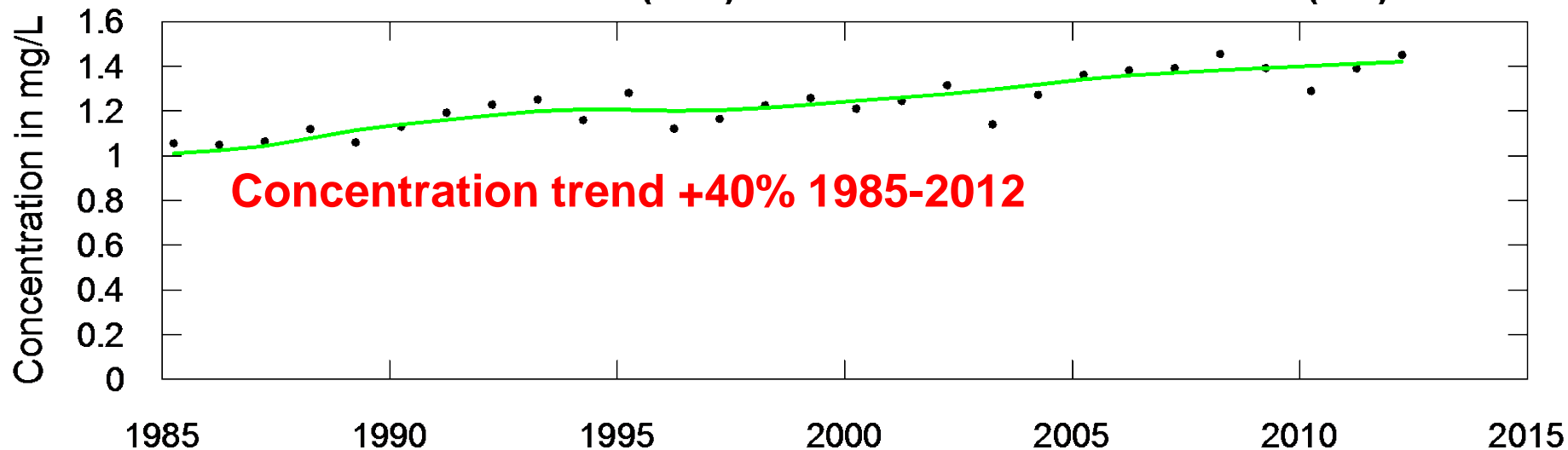
**Flow normalized flux is just $c \times Q$ integrated over
discharge.**

Preliminary Information-Subject to
Revision. Not for Citation or Distribution

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

Choptank River near Greensboro, MD Nitrate
Water Year

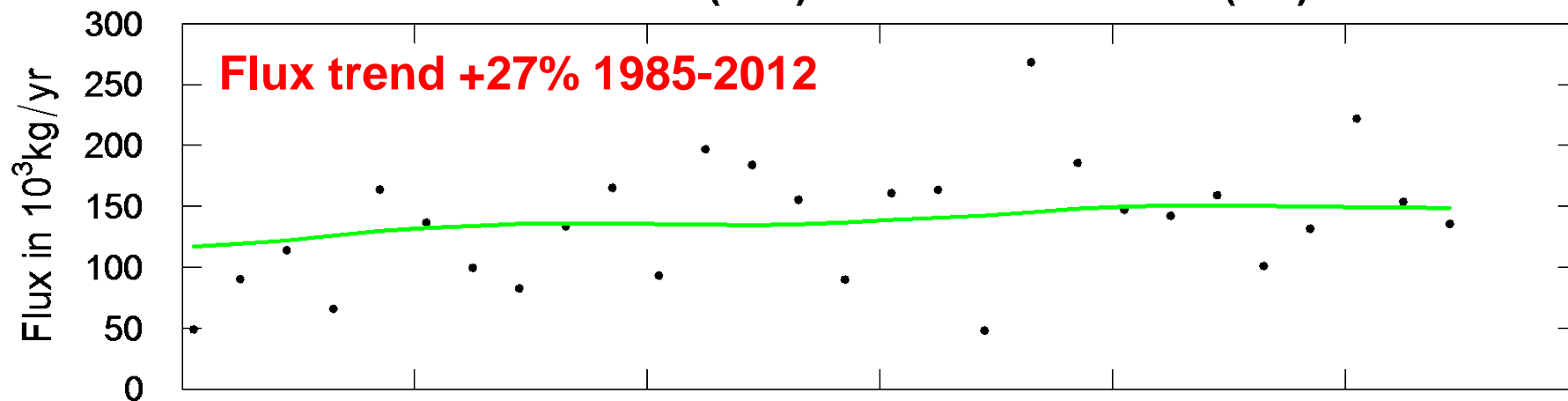
Mean Concentration (dots) & Flow Normalized Concentration (line)



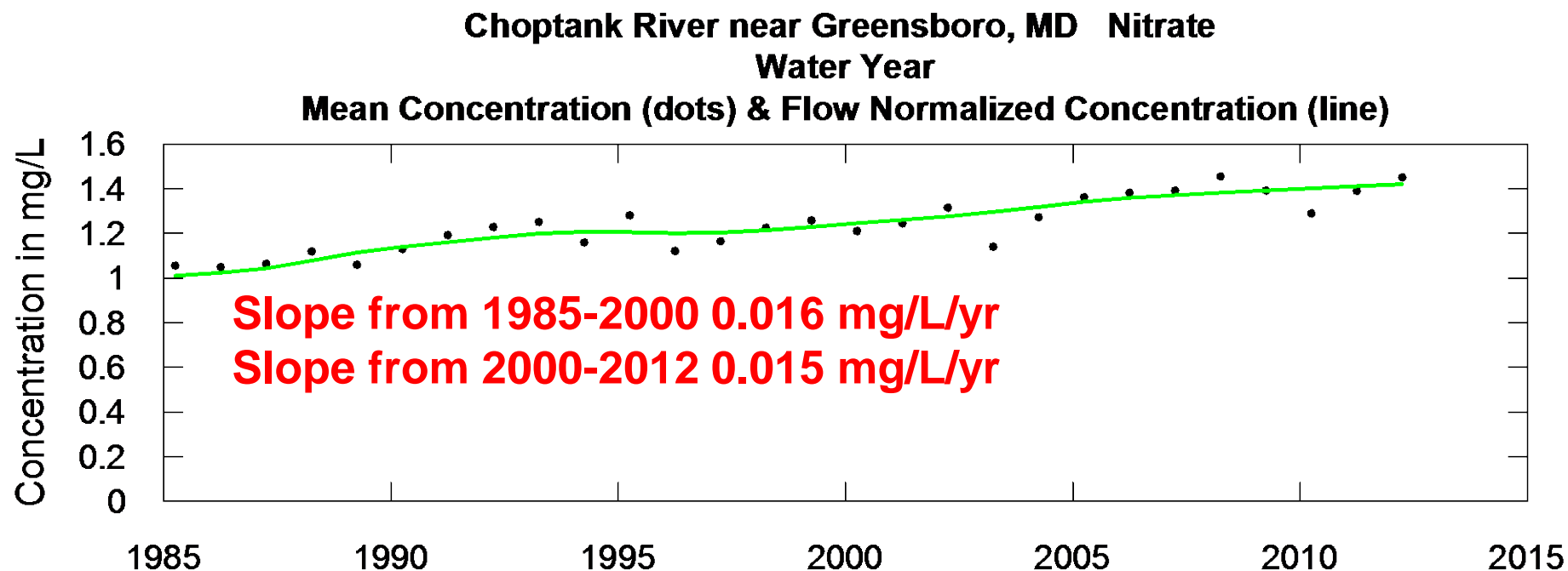
Preliminary Information-Subject to
Revision. Not for Citation or
Distribution

Choptank River near Greensboro, MD Nitrate
Water Year

Flux Estimates (dots) & Flow Normalized Flux (line)

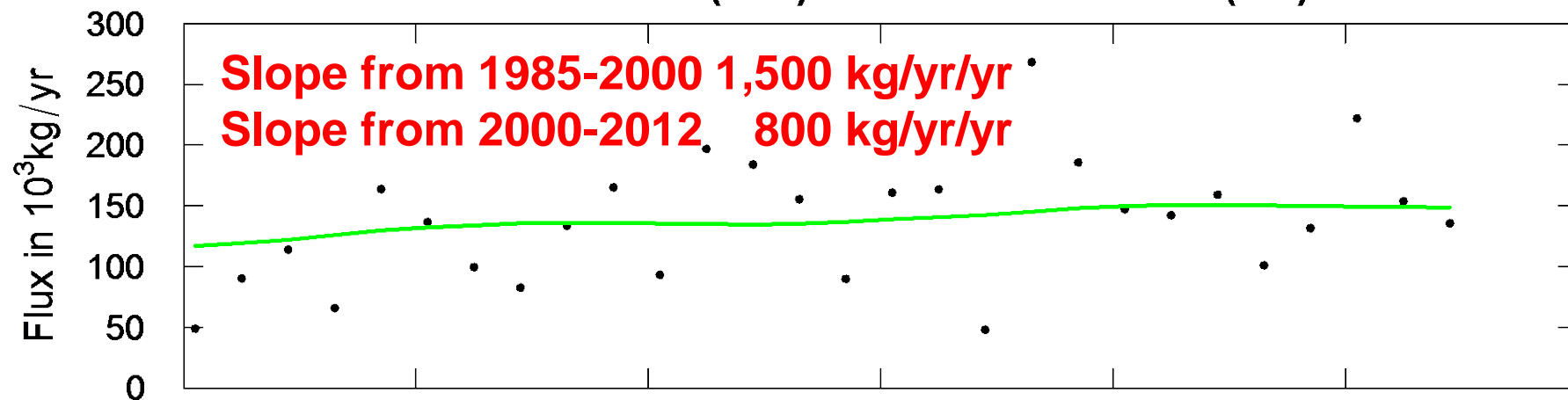


Hints that rates of increase have slowed?



Preliminary Information-Subject to
Revision. Not for Citation or
Distribution

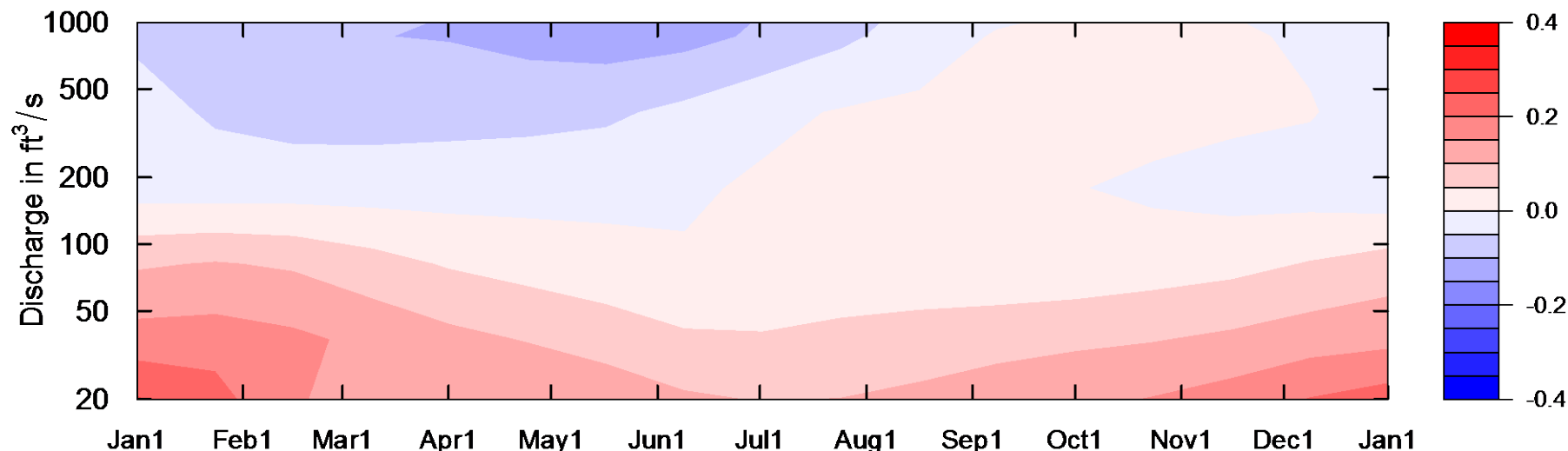
Choptank River near Greensboro, MD Nitrate
Water Year
Flux Estimates (dots) & Flow Normalized Flux (line)



Look at changes in just the last few years.

This is a graphic of differences 2007 to 2012

Choptank River near Greensboro, MD Nitrate plus Nitrite, Filtered, as N
Estimated Concentration change from 2007 to 2012



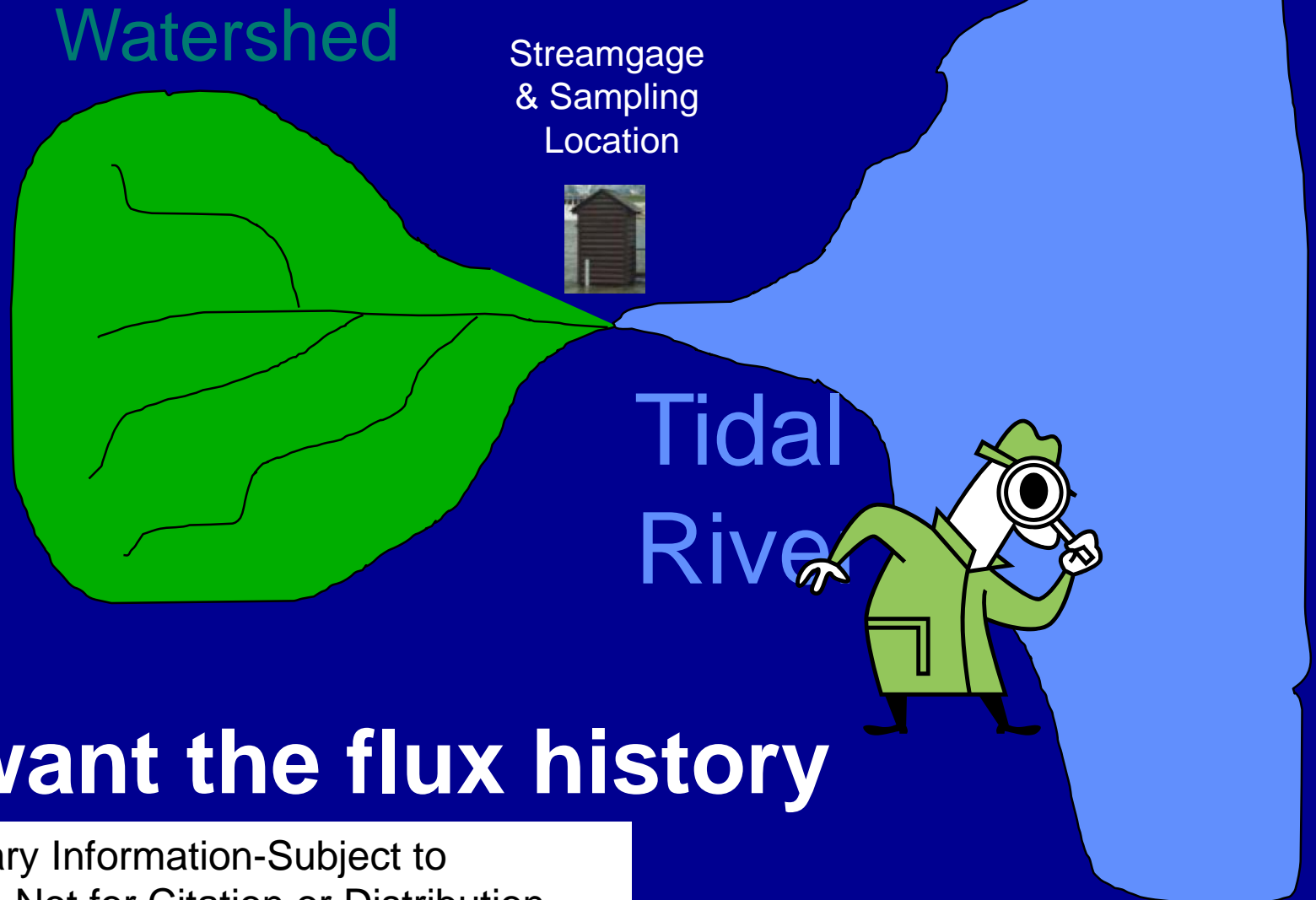
Hypothesis, cover crops are helping at higher flows particularly in the winter. Low flows are still responding to legacy of nitrate enriched groundwater.

Why all this complexity?

Different products for different purposes

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

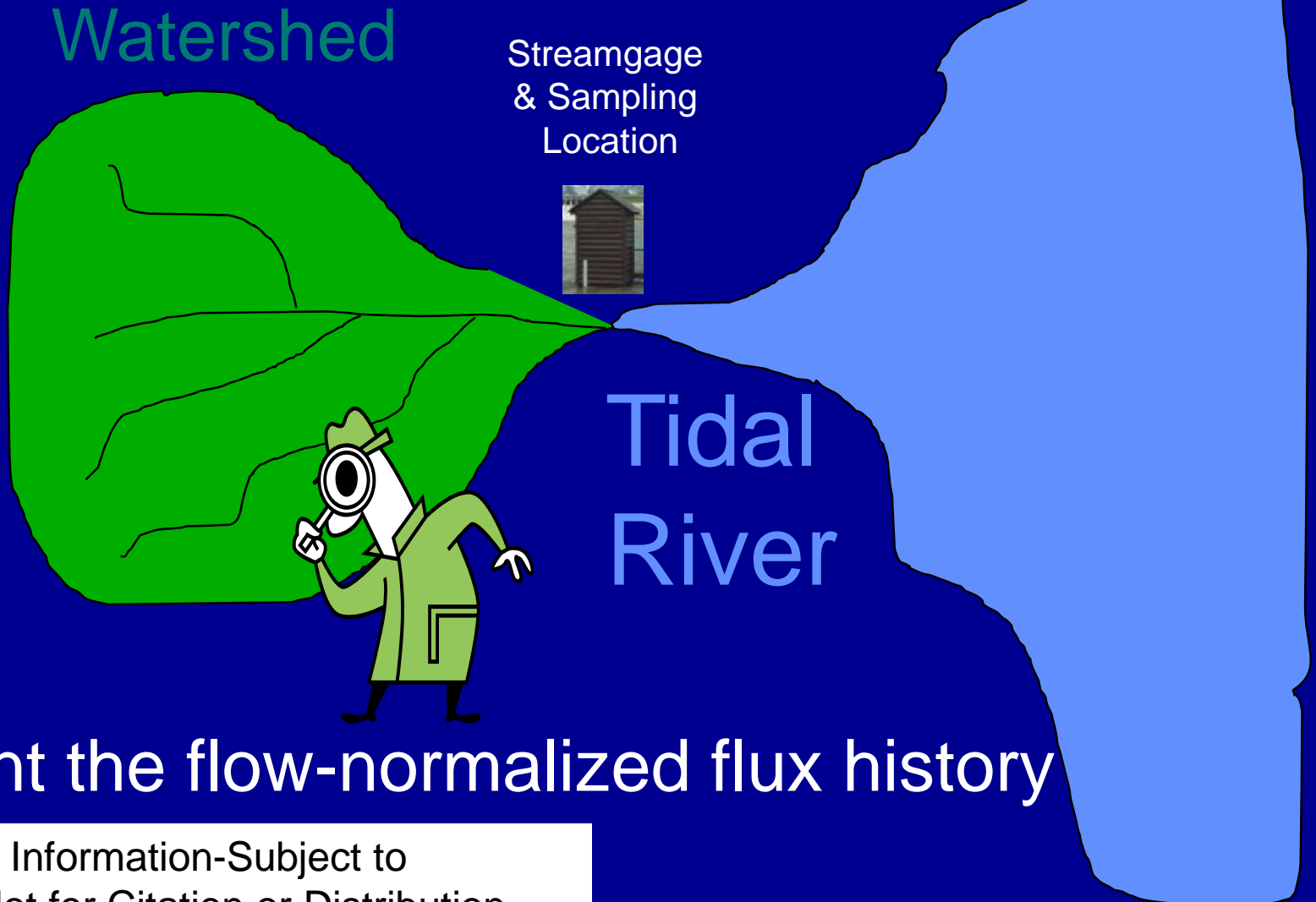
For understanding impact on the estuary ecosystem



We want the flux history

Preliminary Information-Subject to
Revision. Not for Citation or Distribution

For understanding progress in the watershed



We want the flow-normalized flux history

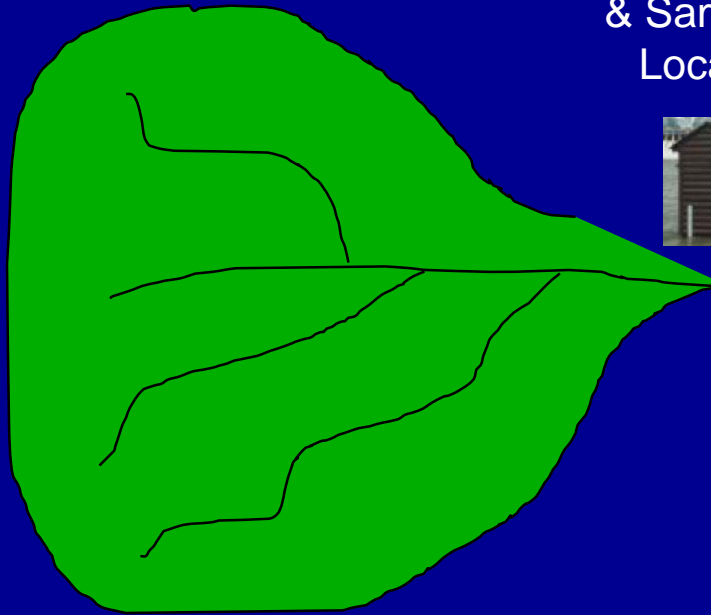
Preliminary Information-Subject to
Revision. Not for Citation or Distribution

For understanding the
changes in the rivers

Estuary

Watershed

Streamgage
& Sampling
Location



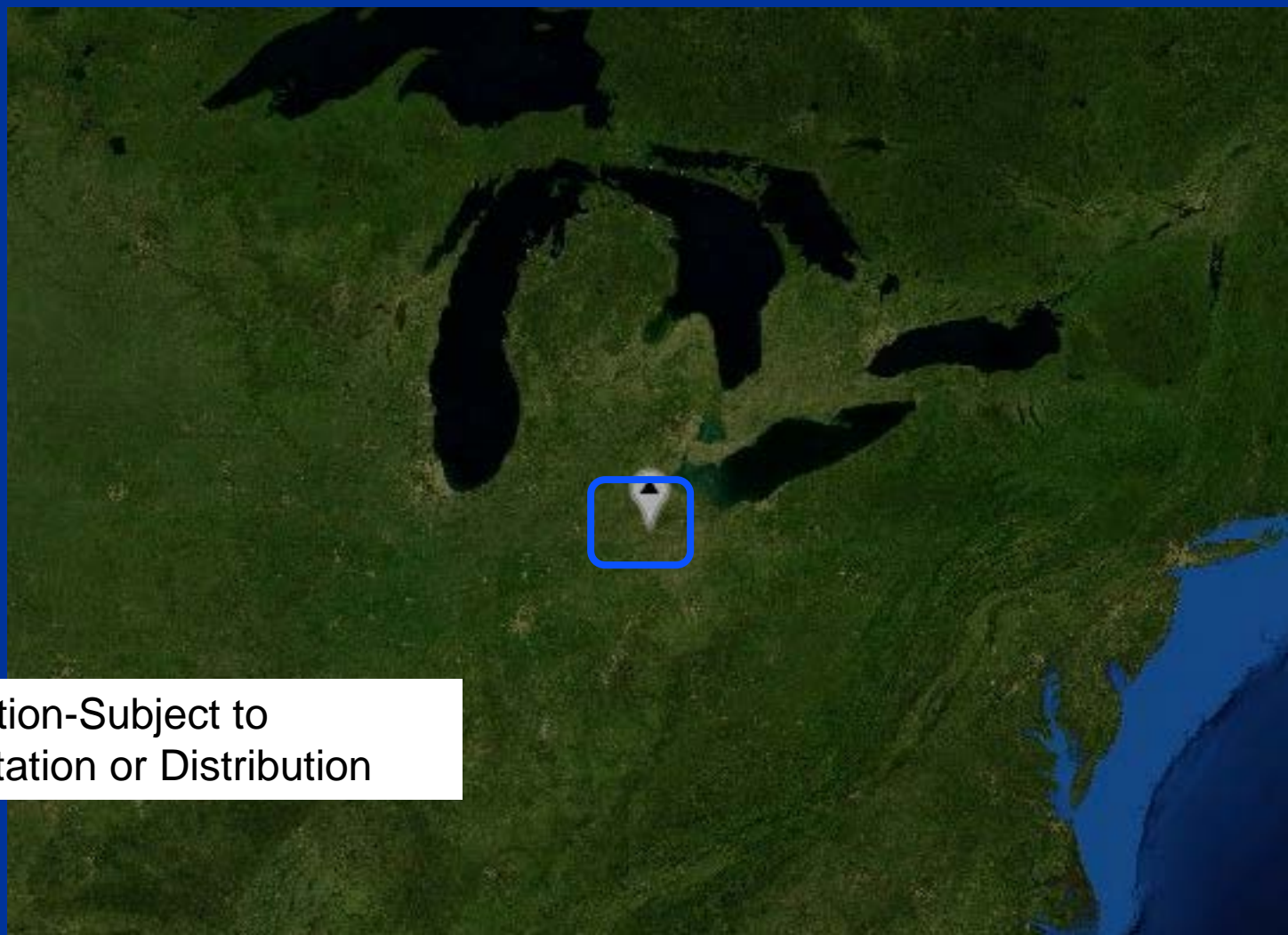
Tidal
River



We want the concentration history

Preliminary Information-Subject to
Revision. Not for Citation or Distribution

Maumee River – 16,000 km² Tributary to Lake Erie



Preliminary Information-Subject to
Revision. Not for Citation or Distribution

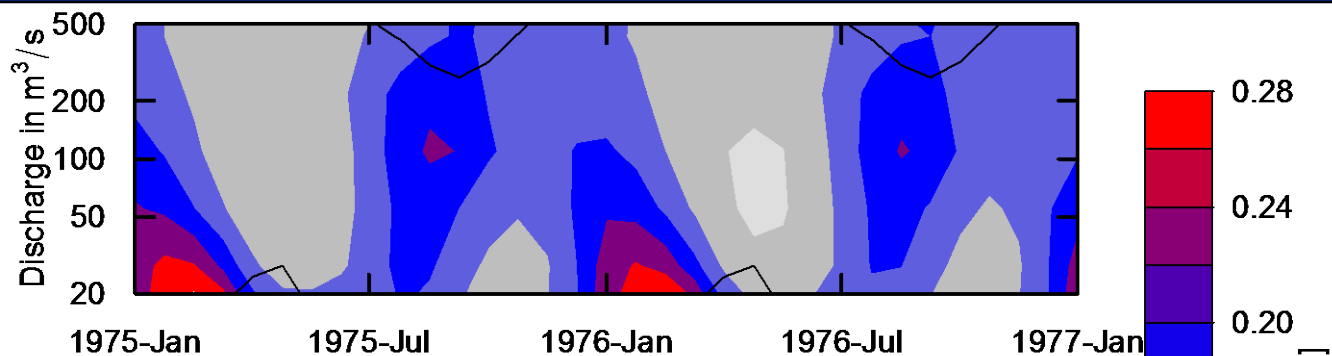
Cyanobacter – Lake Erie

Preliminary Information-Subject to
Revision. Not for Citation or Distribution

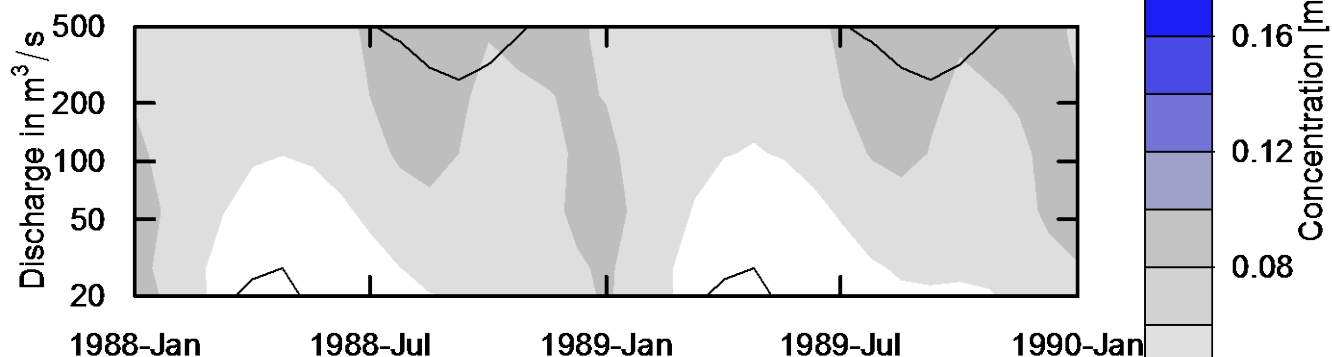


Dissolved Reactive Phosphorus, Maumee River, at Waterville, OH

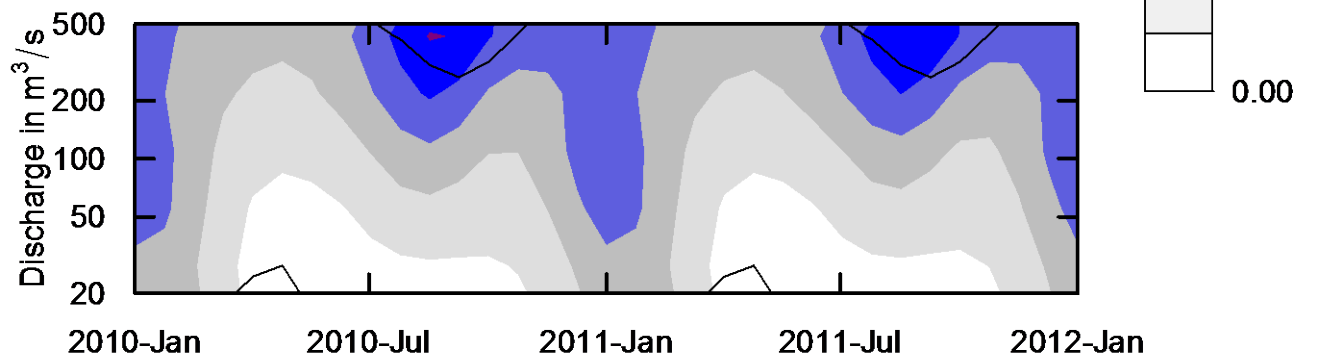
Mid 1970's



Late 1980's



Early 2010's

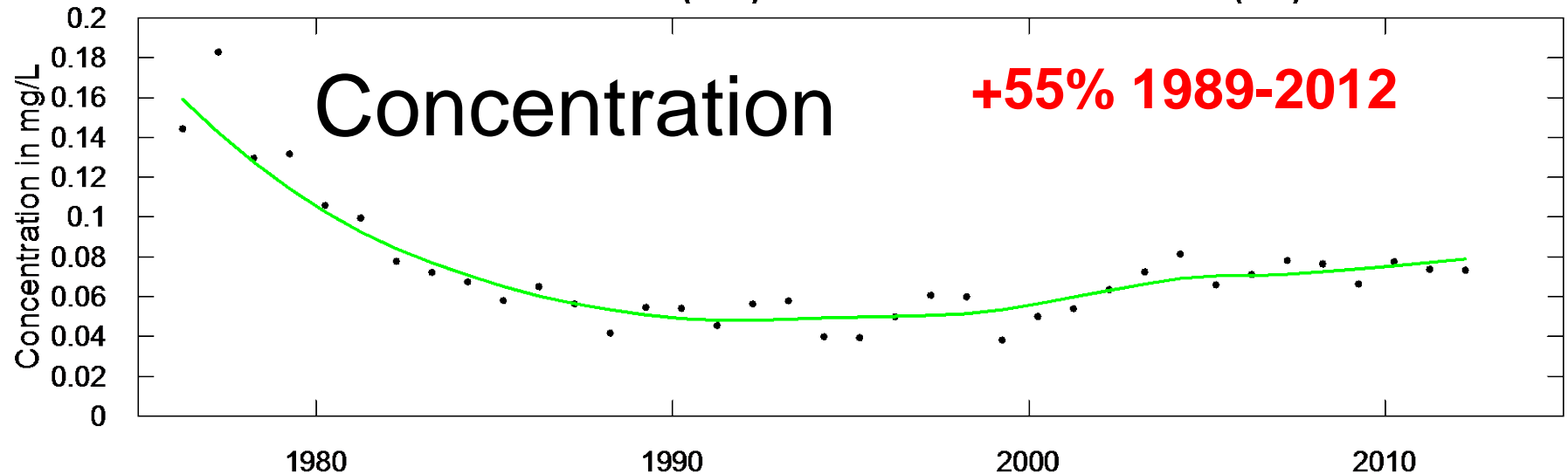


Dissolved Reactive Phosphorus, Maumee River, at Waterville, OH

Maumee River at Waterville OH Dissolved Reactive Phosphorus

Water Year

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

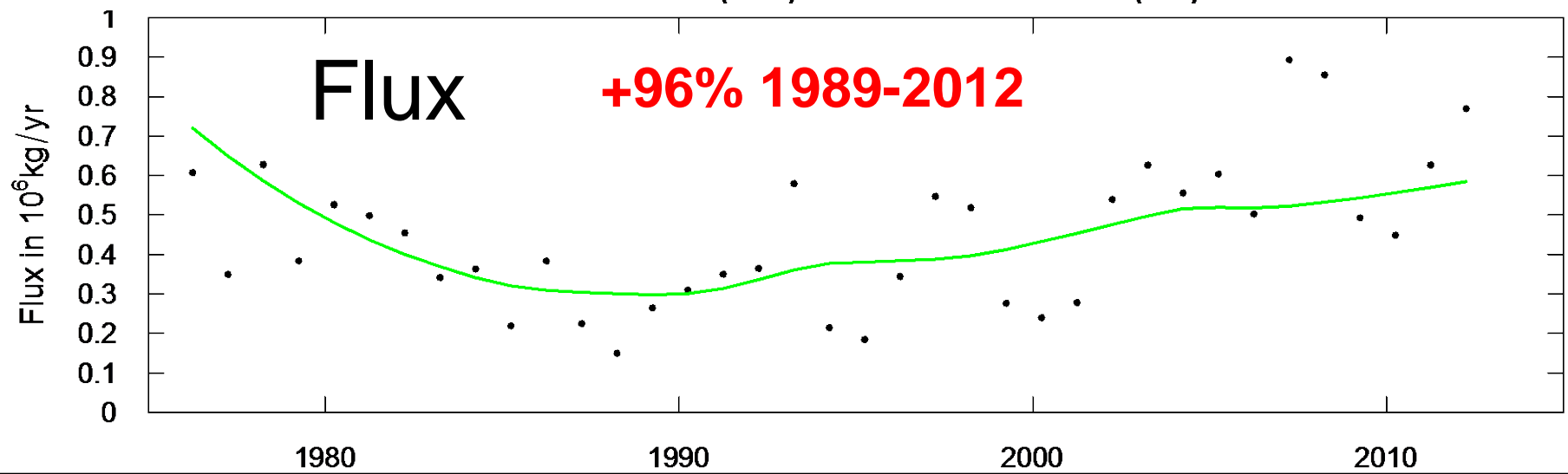


Preliminary Information-
Subject to Revision. Not for
Citation or Distribution

Maumee River at Waterville OH Dissolved Reactive Phosphorus

Water Year

Flux Estimates (dots) & Flow Normalized Flux (line)



dataRetrieval

- Brings data from various sources into R
- Organizes it
- Some of the functions organize for EGRET uses

dataRetrieval functions:

By information source and purpose

Information Source	Site Query	Meta Data	Data
NWIS			
Water Quality Portal			
User-supplied files			

dataRetrieval functions:

By information source and purpose

Information Source	Site Query	Meta Data	Data
NWIS	getNWISSites getNWISDataAvailability	getNWISInfo getNWISSiteInfo getNWISPcodeInfo	getNWISData getNWISDaily getNWISSample getNWISdvData getNWISunitData getNWISqwData
Water Quality Portal	getWQPSites	getWQPInfo	getWQPSample getWQPqwData getWQPData
User-supplied files		getUserInfo	getUserDaily getUserSample

Those functions shown in red are designed specifically to provide data frames suitable for use by the EGRET package

Getting Started

- Need to install R (freely downloaded from <http://cran.us.r-project.org/>)
- Next install the EGRET & dataRetrieval packages per the instructions on our wiki page.

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

Will soon be on CRAN (Contributed Packages) ☐

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 `getNWISSample`) just type `?getNWISSample`

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 <- getNWISSample(siteNumber,parameterCD,startDate,endDate)
> summary(Sample)

```

```

> Sample<-getNWISSample("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

```



Preliminary Information-Subject to
Revision. Not for Citation or Distribution

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 reported 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 $\text{ConcLow} = 0.9$ and $\text{ConcHigh} = 1.1$.

- The conventional left-censored approach calls this $(0, 1.1)$
- WRTDS calls this $(0.9 \text{ to } 1.1)$

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
```

```
➤ Daily <- getNWISDaily(siteNumber,"00060",
  startDate,endDate)
```

```
> Daily<-getNWISDaily("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<-getNWISInfo(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


```
> INFO<-getNWISInfo(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):

<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

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

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

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

no3

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

```
> INFO <- getUserInfo()
```

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

Preliminary
Information-Subject to
Revision. Not for
Citation or Distribution

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

```
> Sample<-mergeReport()
```

```
> Sample<-mergeReport()
```

```
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
```

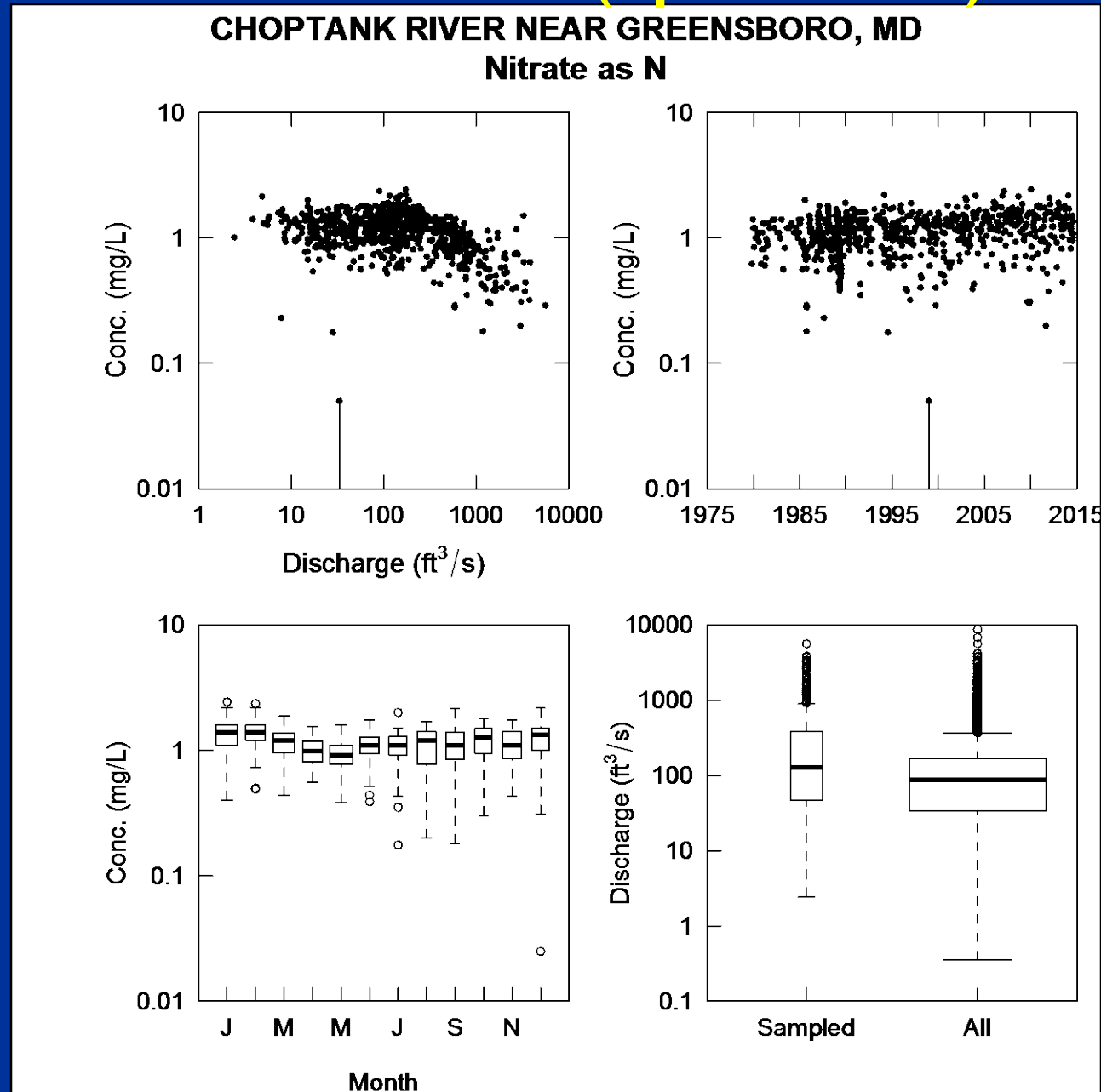
```
Concentration: Minimum, mean and maximum 0.05 1.1 2.4
```

```
Percentage of the sample values that are censored is 0.14 %
```

> multiPlotDataOverview(qUnit=1)

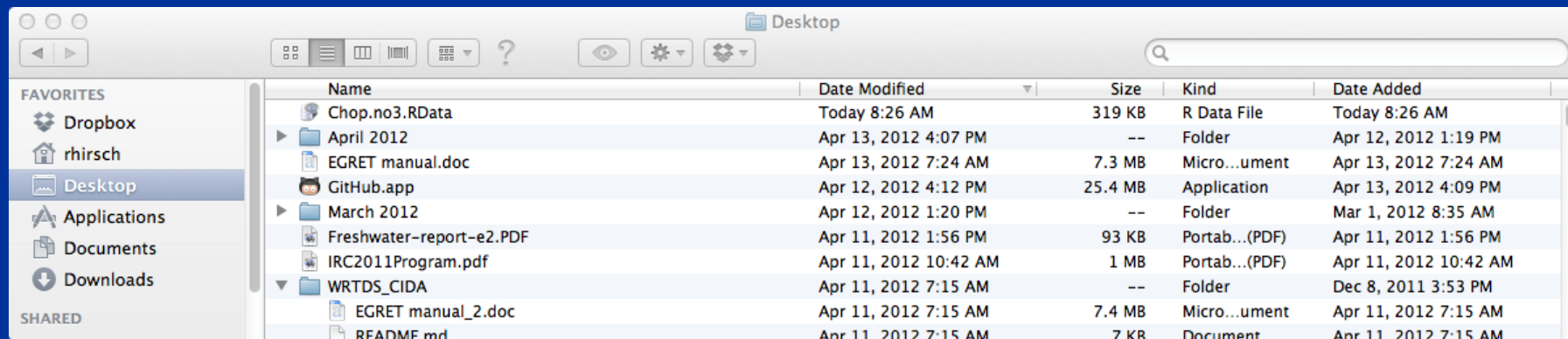
Let's look at
the data
before we
proceed, the
function is:

Preliminary
Information-Subject to
Revision. Not for
Citation or Distribution



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

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



Chop.no3.RData

Save it over and over as
you proceed and add
results

Preliminary Information-Subject to
Revision. Not for Citation or Distribution

We now have 3 data frames

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

A short digression into other things you can do with dataRetrieval, not related to EGRET

Unit values retrieval (not used by EGRET)

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

```

> Unit<-getNWISunitData("05484500",parameterCd=c("99133","00060"),"2013-03-01","2013-09-30")
> summary(Unit)
  agency_cd      site_no      datetime      tz_cd      X01_00060_00011
Length:19568   Length:19568   Min.    :2013-03-10 00:00:00   Length:19568   Min.    : 108
Class :character   Class :character   1st Qu.:2013-04-30 00:26:15   Class :character   1st Qu.: 331
Mode  :character   Mode  :character   Median :2013-06-21 00:22:30   Mode  :character   Median : 1320
                                   Mean  :2013-06-20 14:54:09      Mean  : 3370
                                   3rd Qu.:2013-08-10 23:18:45    3rd Qu.: 4180
                                   Max.  :2013-09-30 23:45:00      Max.  :24600
                                   NA's  :86

X01_00060_00011_cd X18_99133_00011 X18_99133_00011_cd
Length:19568      Min.    : 1.260   Length:19568
Class :character   1st Qu.: 2.540   Class :character
Mode  :character   Median : 7.250   Mode  :character
                                   Mean  : 9.651
                                   3rd Qu.:16.900
                                   Max.  :20.700
                                   NA's  :1848

```

We have 19,568 time steps

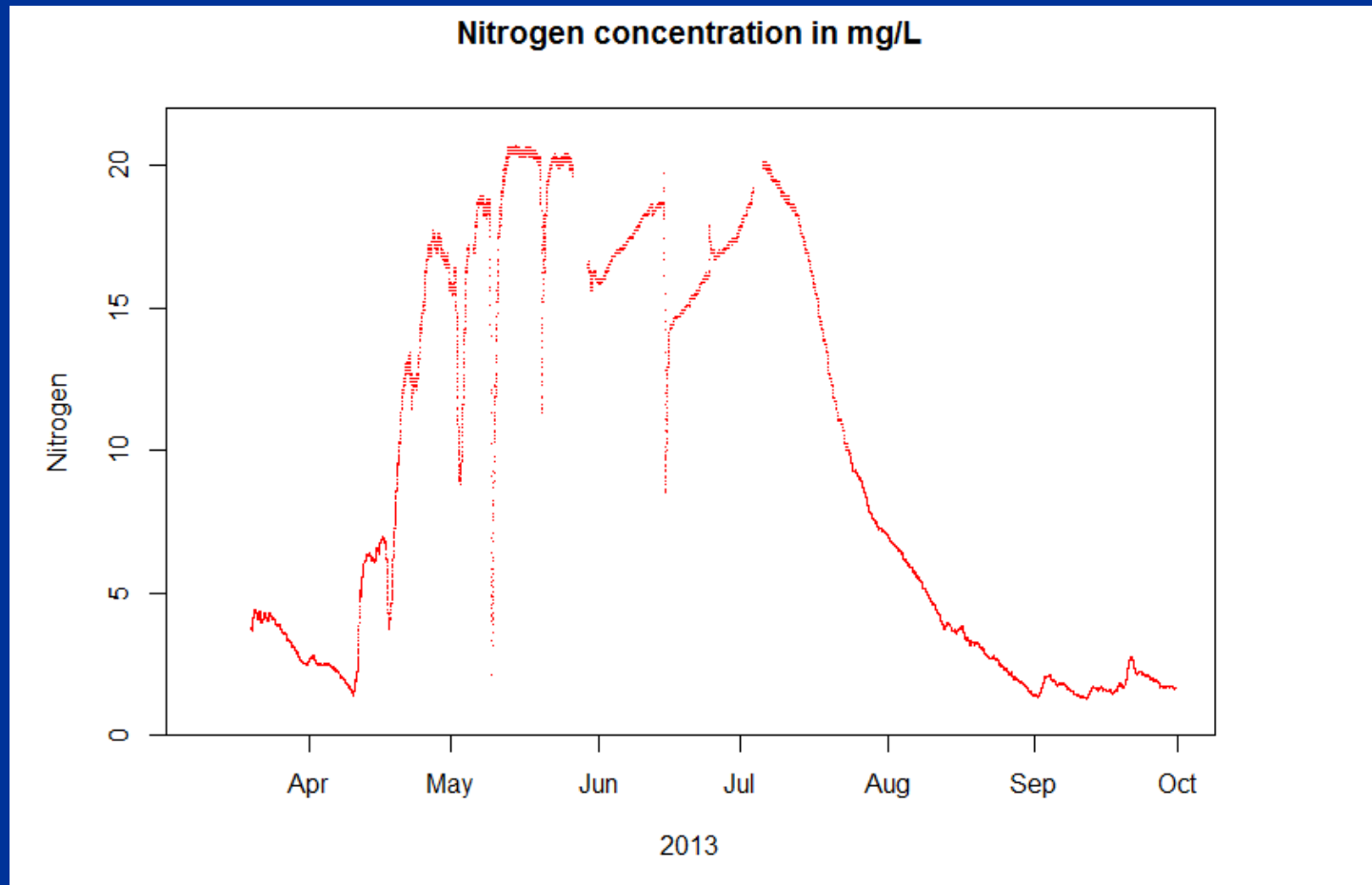
Discharge missing at 86 of them

Nitrate missing at 1848 of them

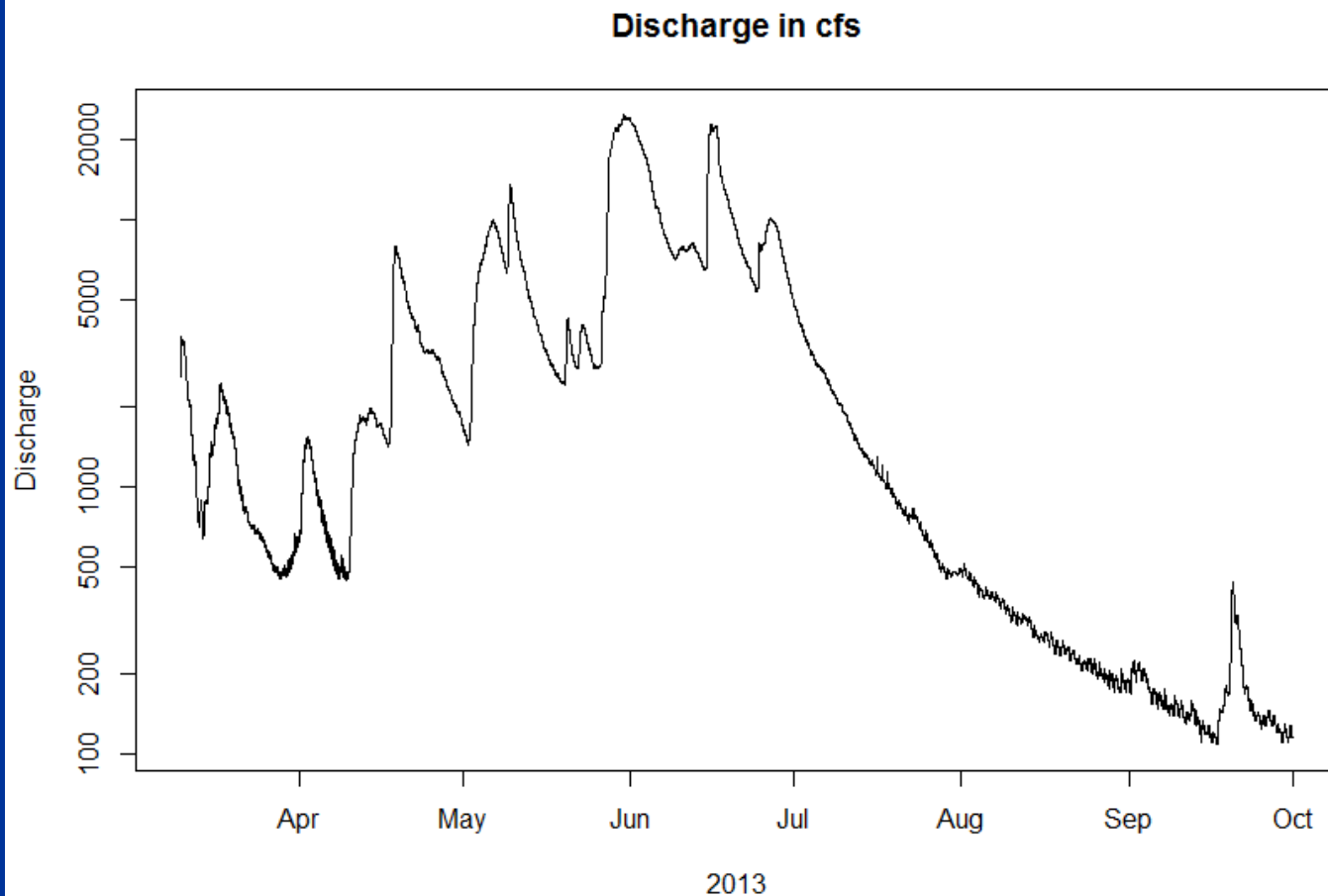


Preliminary Information-Subject to
Revision. Not for Citation or Distribution

```
> with(Unit, plot(datetime,X18_99133_00011,  
  col="red",pch=".",ylim=c(0,22),yaxs="i",  
  xlab="2013",ylab="Nitrogen",  
  main="Nitrogen concentration in mg/L"))
```



```
> with(Unit, plot(datetime, X01_00060_00011,  
  col="black", log="y", type="l",  
  xlab="2013", ylab="Discharge",  
  main="Discharge in cfs"))
```



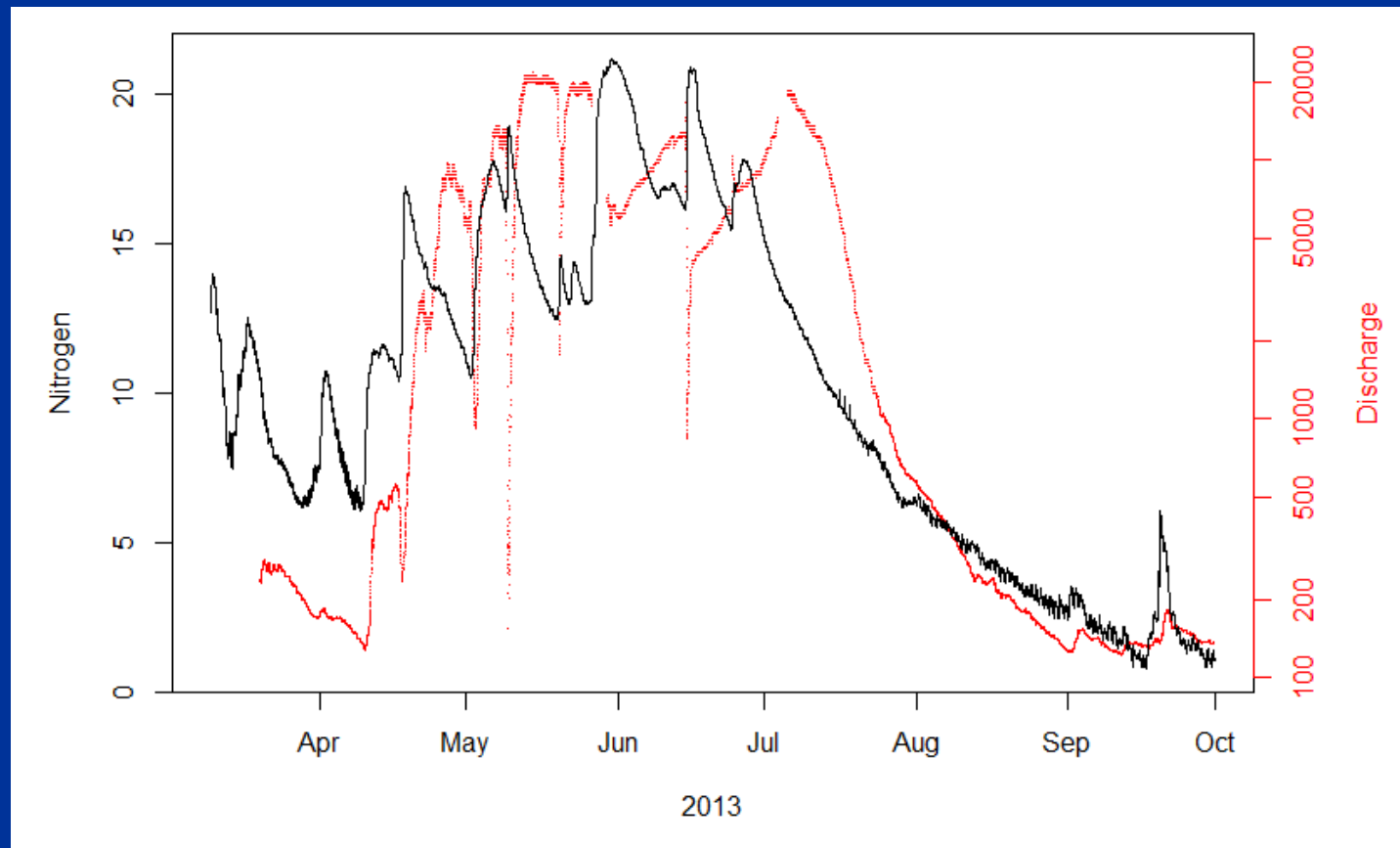
```

> par(mar=c(5,5,5,5))
with(Unit, plot(datetime,X18_99133_00011,
                 col="red",pch=".",ylim=c(0,22),yaxs="i",
                 xlab="2013",ylab="Nitrogen"))
> par(new=TRUE)
> with(Unit, plot(datetime,X01_00060_00011,
                 col="black",log="y",type="l",
                 xaxt="n",yaxt="n",axes=FALSE,ylab="",xlab=""))
> axis(4,col="red",col.axis="red")
> mtext("Discharge",side=4,line=3,col="red")

```

Many options for graphics and modeling of concentration and discharge together. Not a part of EGRET – but it must be in a next generation of EGRET!

Preliminary
Information-
Subject to
Revision. Not for
Citation or
Distribution



Now back to what is in EGRET

- This is where EGRET estimates the WRTDS model and applies it

> 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

Preliminary Information-Subject to
Revision. Not for Citation or Distribution

**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 (12782, 19)

```
> 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

yHat	SE	ConcDay	FluxDay	FNConc	FNFlux
Min. : -1.470422	Min. :0.1303	Min. :0.2485	Min. : 1.245	Min. :0.8072	Min. : 74.77
1st Qu.: -0.004537	1st Qu.:0.2066	1st Qu.:1.0370	1st Qu.: 100.207	1st Qu.:1.0756	1st Qu.: 176.68
Median : 0.147315	Median :0.2348	Median :1.2147	Median : 258.263	Median :1.2287	Median : 328.73
Mean : 0.133796	Mean :0.2583	Mean :1.2131	Mean : 380.318	Mean :1.2151	Mean : 375.83
3rd Qu.: 0.277517	3rd Qu.:0.2888	3rd Qu.:1.3785	3rd Qu.: 508.028	3rd Qu.:1.3363	3rd Qu.: 559.70
Max. : 0.595483	Max. :0.7169	Max. :1.8551	Max. :5741.182	Max. :1.7822	Max. :1013.80

“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

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

Units in EGRET

Everything stored as:

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

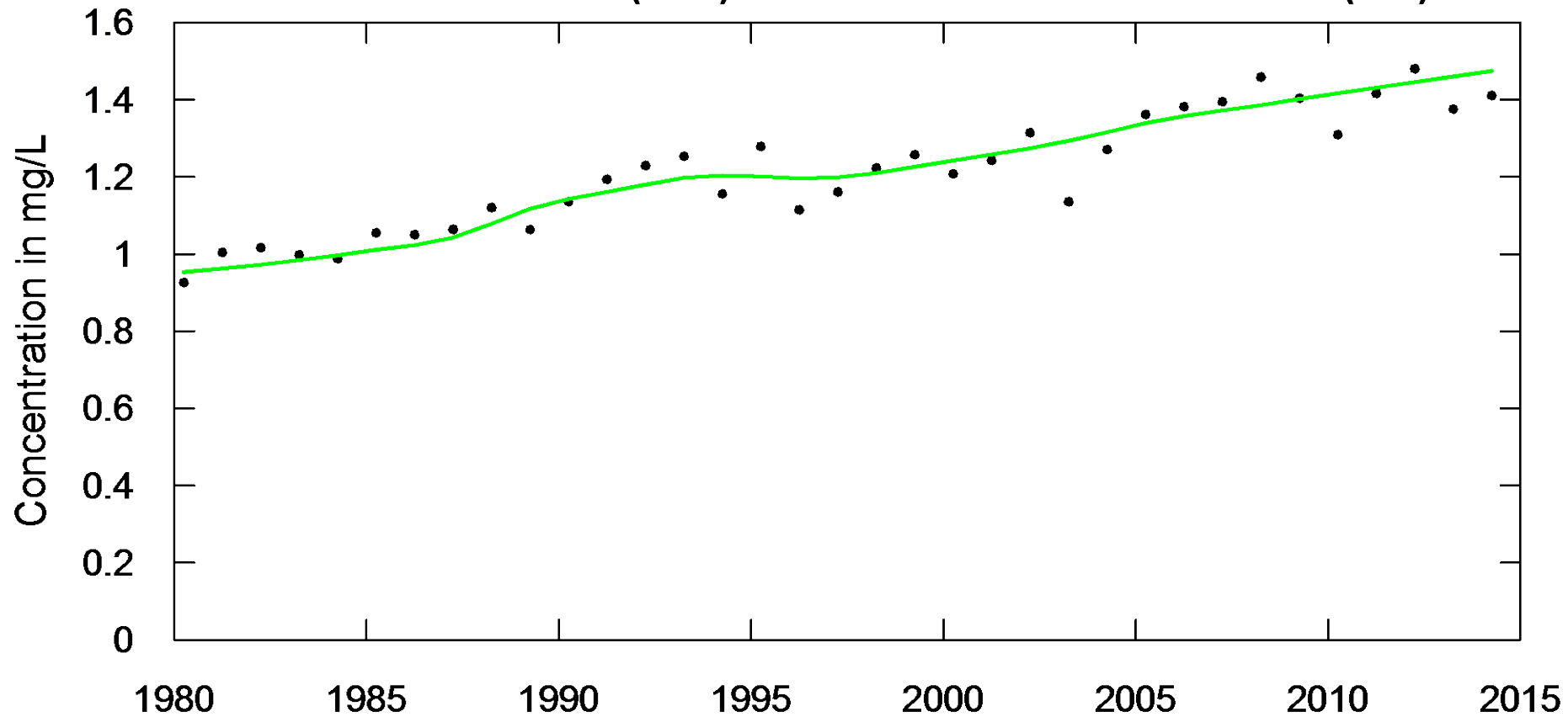
But each graphic or table has a wide choice of units (English and SI) that the user can select

Now lets see some trend results

```
> plotConcHist()
```

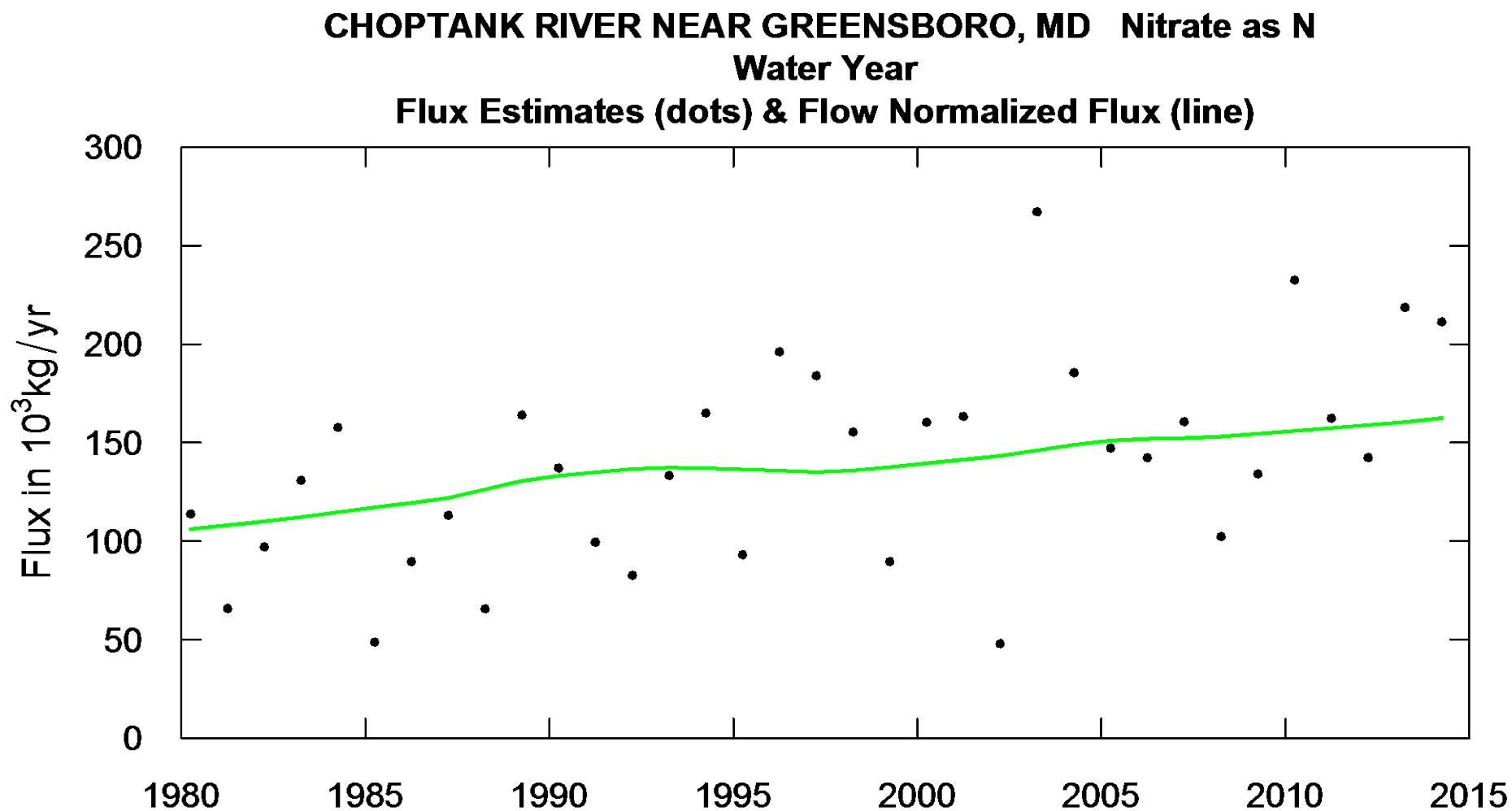
CHOPTANK RIVER NEAR GREENSBORO, MD Nitrate as N
Water Year

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



Preliminary Information-Subject to
Revision. Not for Citation or Distribution

> plotFluxHist(fluxUnit=8)



Preliminary Information-Subject to
Revision. Not for Citation or Distribution

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

CHOPTANK RIVER NEAR GREENSBORO, MD

Nitrate as N

Water Year

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

Preliminary
Information-
Subject to
Revision. Not
for Citation or
Distribution


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

CHOPTANK RIVER NEAR GREENSBORO, MD

Nitrate as N

Water Year

Concentration trends

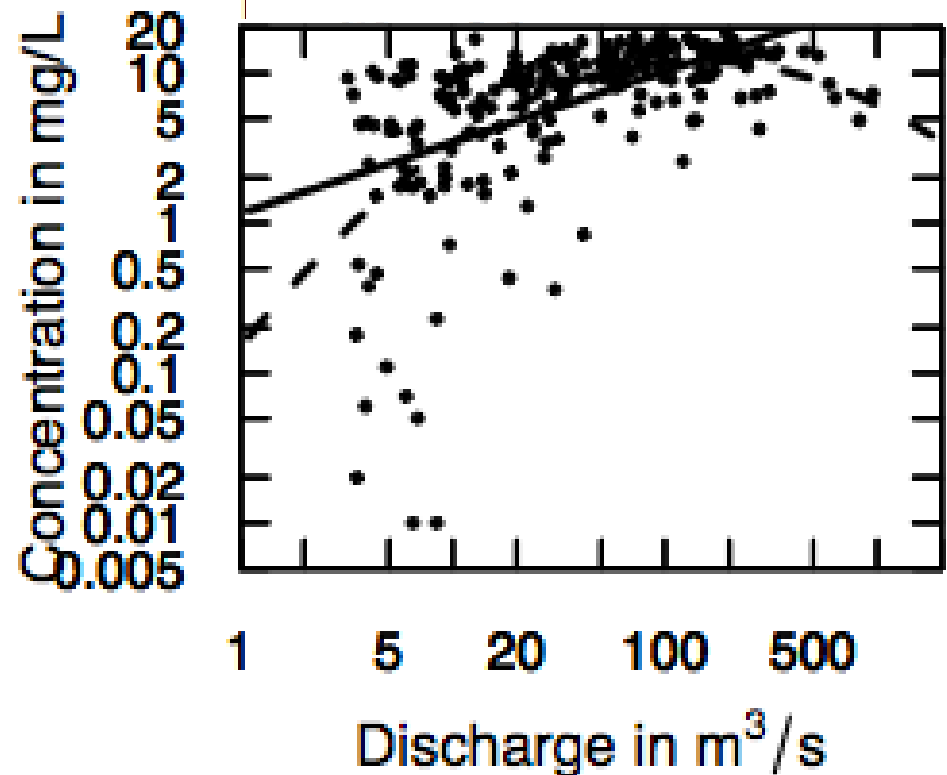
time span			change mg/L	slope mg/L/yr	change %	slope %/yr
1980	to	1995	0.25	0.017	26	1.7
1980	to	2014	0.52	0.015	55	1.6
1995	to	2014	0.27	0.014	23	1.2

Flux Trends

time span			change tons/yr	slope tons/yr /yr	change %	slope %/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

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

Preliminary
Information-
Subject to
Revision. Not for
Citation or
Distribution

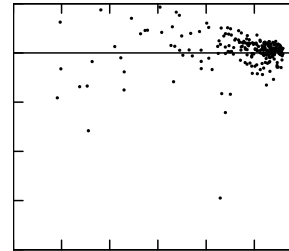


**EGRET
produces a
diagnostic
plot to help
spot serious
problems
with the
model**

fluxBiasMulti(fluxUnit=4)

Preliminary Information-Subject
to Revision. Not for Citation or
Distribution

trate
Model is WRTDS Flux Bias Statistic -0.00237

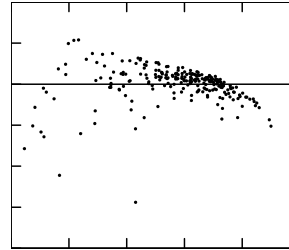


This same
type of plot
can be used
to look at
other
models,
here the
LOADEST7

`fluxBiasMulti(fluxUnit=4)`

Preliminary Information-
Subject to Revision. Not for
Citation or Distribution

n River at Des Moines, IA Nitrate
Model is L7 Flux Bias Statistic 0.319



Diagnostics and potential problems with estimating mean flux, see:

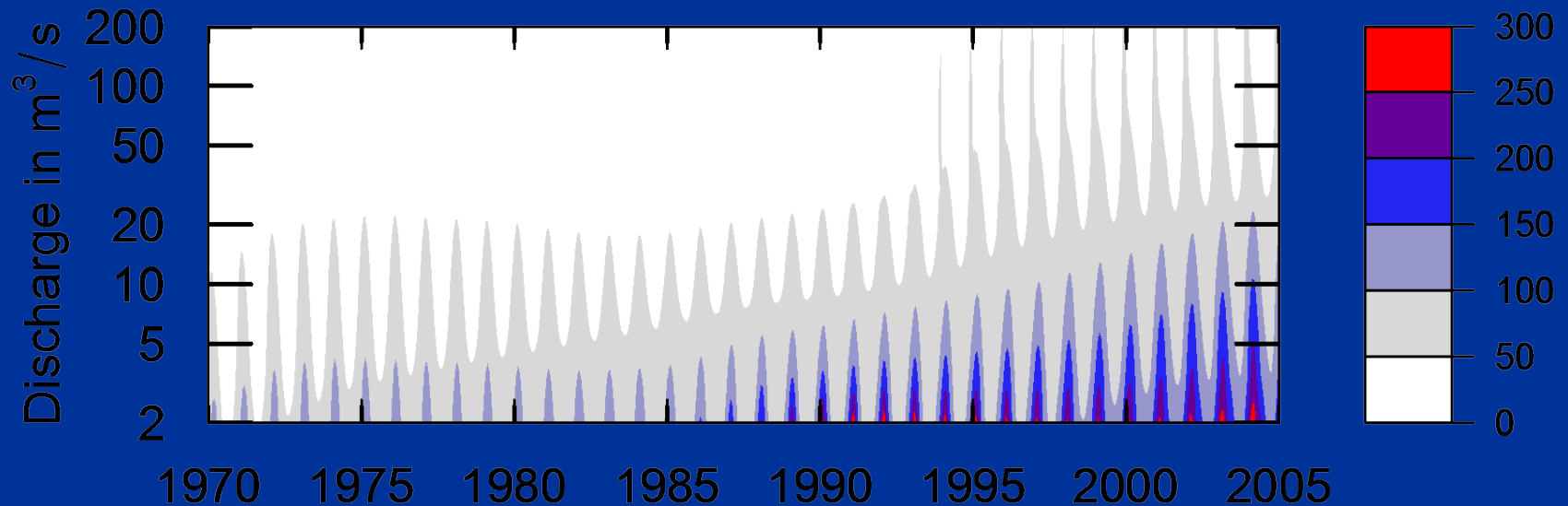
Hirsch, R.M., 2014, Large biases in regression-based constituent flux estimates: causes and diagnostics. Journal of the American Water Resources Association.

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

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))
```

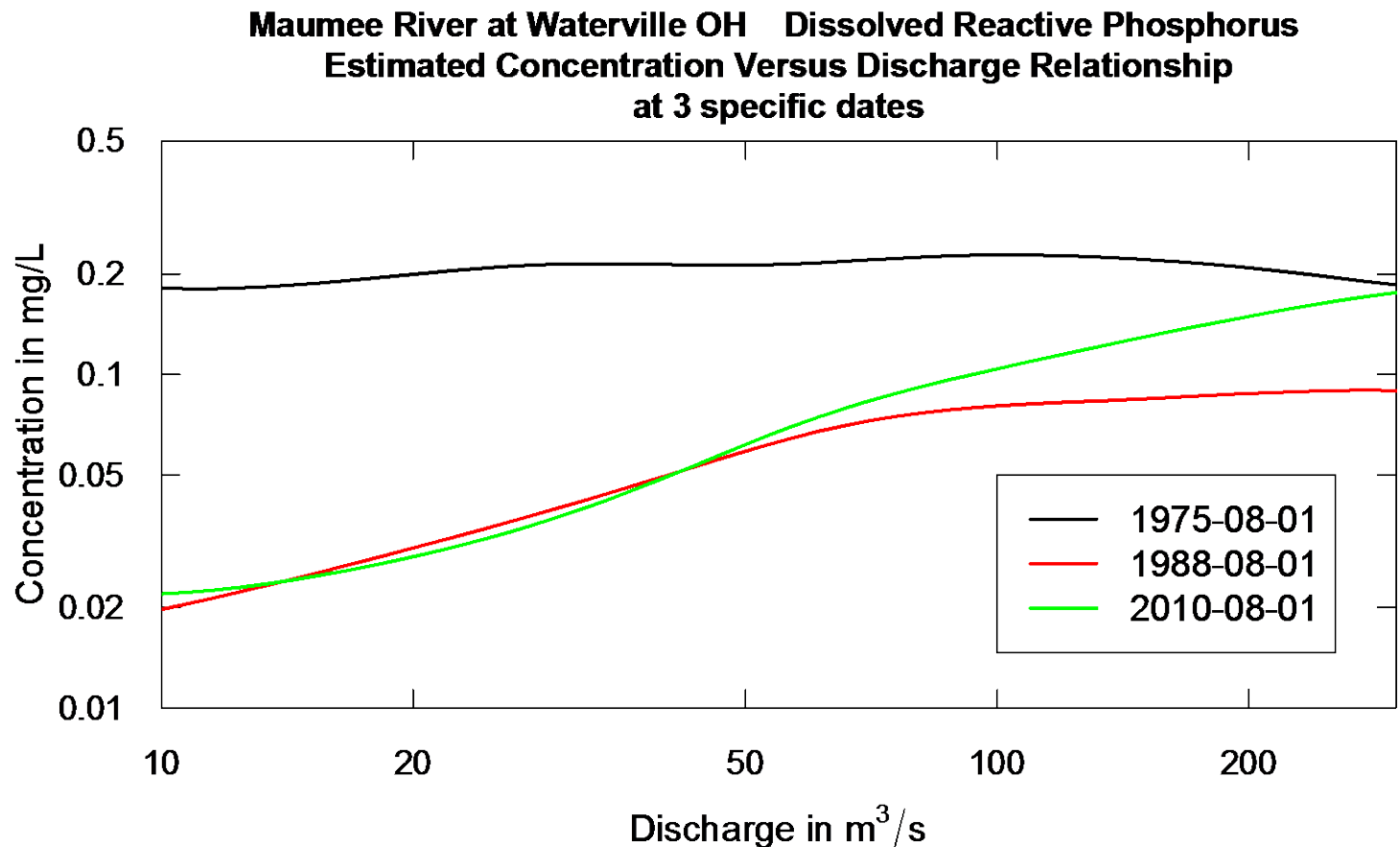
Milwaukee River at Milwaukee, WI Chloride
Estimated Concentration Surface in Color



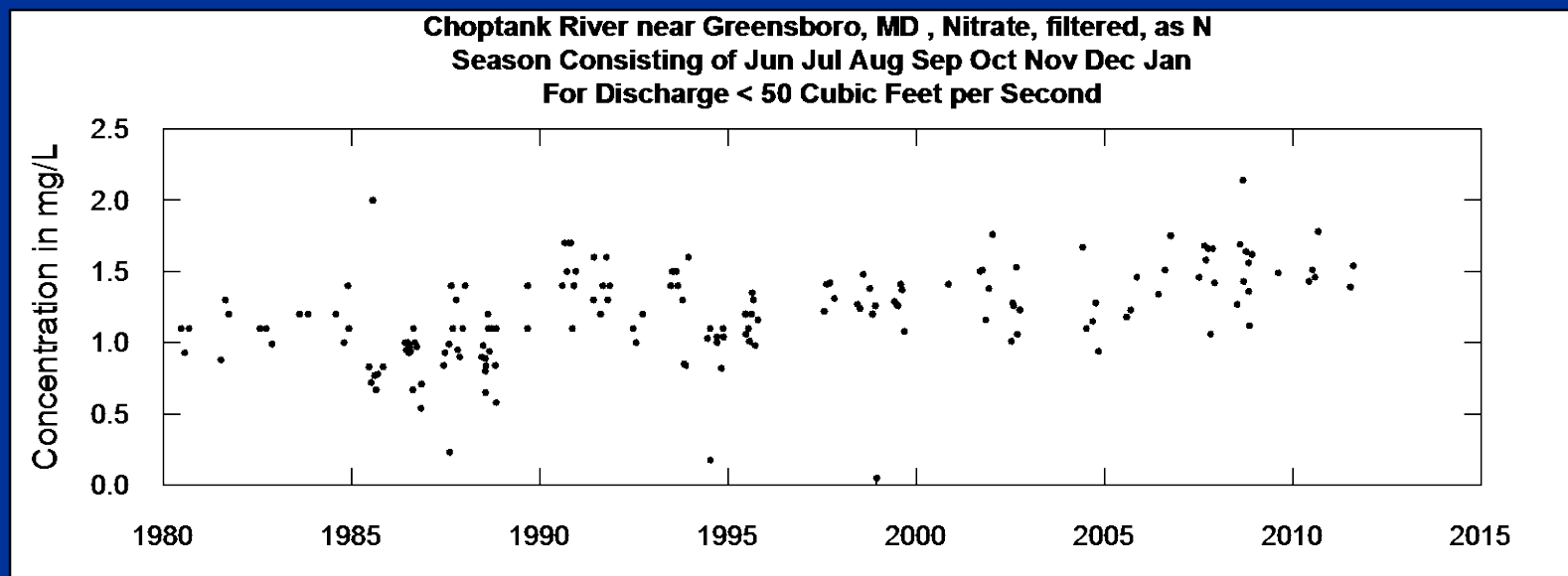
There are many more graphics, for example

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

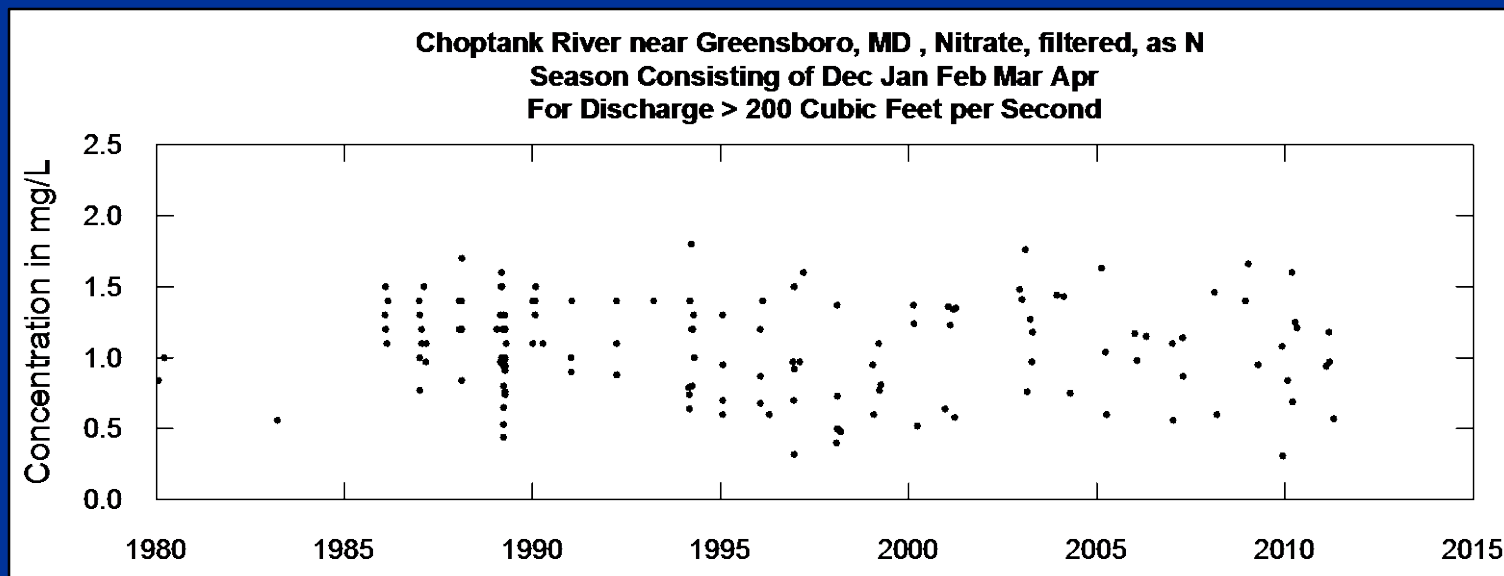
Preliminary
Information-
Subject to
Revision. Not
for Citation or
Distribution



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



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



Preliminary
Information-
Subject to
Revision. Not
for Citation or
Distribution

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**

Preliminary Information-Subject to
Revision. Not for Citation or Distribution

and poring over the records.

Anticipated enhancements

- Significance levels and confidence intervals for trends
- Dealing with ephemeral streams
- Estimation of trends in frequency of exceedances of threshold values
- Dealing with nonstationarity in Q
- Improved estimates of individual year fluxes

dataDelivery and EGRET

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

A huge thanks to
Laura De Cicco for
making this dream
a reality



Three more dataRetrieval slides. If I still have time

```
> sta<-"01578310"
> param<-c("00665","00671","80154")

> Conowingo<-getNWISqwData(sta,param,"","")
```

```
> summary(Conowingo)
```

dateTime	site	qualifier_00665	value_00665	qualifier_00671
Min. :1978-01-29 11:15:00	Length:1213	Length:1213	Min. :0.01000	Length:1213
1st Qu.:1987-04-15 12:00:00	Class :character	Class :character	1st Qu.:0.03200	Class :character
Median :1990-04-04 12:00:00	Mode :character	Mode :character	Median :0.05000	Mode :character
Mean :1993-07-12 00:51:08			Mean :0.07134	
3rd Qu.:2000-11-02 11:45:00			3rd Qu.:0.08000	
Max. :2014-08-05 12:15:00			Max. :2.31000	
			NA's :117	

value_00671	qualifier_80154	value_80154
Min. :0.00000	Length:1213	Min. : 1.00
1st Qu.:0.00400	Class :character	1st Qu.: 8.00
Median :0.00800	Mode :character	Median : 15.00
Mean :0.01114		Mean : 47.84
3rd Qu.:0.01200		3rd Qu.: 37.75
Max. :0.13700		Max. :3680.00
NA's :119		NA's :107

What have I got:

- a data frame of 1213 samples from site 01578310 (Susquehanna River at Conowingo, MD),
- It contains all the Total Phosphorus, Dissolved Orthophosphorus, and Suspended Sediment sample data

What do I want? I want only the samples that had all three of these parameters measured in the sample.

```
> ConowingoAll<-na.omit(Conowingo)
```

```
> summary(ConowingoAll)
```

dateTime	site	qualifier_00665	value_00665	qualifier_00671
Min. :1978-11-28 10:30:00	Length:978	Length:978	Min. :0.01000	Length:978
1st Qu.:1987-08-20 08:31:15	Class :character	Class :character	1st Qu.:0.03200	Class :character
Median :1992-04-02 11:52:30	Mode :character	Mode :character	Median :0.05000	Mode :character
Mean :1994-09-16 19:08:36			Mean :0.07263	
3rd Qu.:2002-05-30 16:26:15			3rd Qu.:0.08000	
Max. :2014-06-03 10:45:00			Max. :2.31000	

value_00671	qualifier_80154	value_80154
Min. :0.00000	Length:978	Min. : 1.00
1st Qu.:0.00400	Class :character	1st Qu.: 8.00
Median :0.00800	Mode :character	Median : 14.00
Mean :0.01096		Mean : 45.49
3rd Qu.:0.01200		3rd Qu.: 34.00
Max. :0.13700		Max. :3680.00

Done: now I can start to explore the relationships of TP, DOP, SS and time

```

> ConowingoAll$ratio<-
ConowingoAll$value_00665/ConowingoAll$value_80154
> ConowingoLow<-subset(ConowingoAll,value_80154<=14)
>
plot(ConowingoLow$dateTime,ConowingoLow$ratio*1000,log="y",xlab=
"Year",ylab="mg TP / g Sediment",main="Susquehanna River at
Conowingo\nRatio of TP to Suspended Sediment\n for samples with
< 14 mg/L SS which is the long term median")

```

Preliminary Information-
Subject to Revision. Not for
Citation or Distribution

