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
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
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

USGS
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.
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.
Water quality data analysis issues

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

- Assessments 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”
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
Approach

• Flexible statistical model to determine the expected value of concentration for any possible combination of date and discharge during the period of record.

• \( \text{E[flux]} = \text{E[Concentration]} \times \text{Discharge} \)
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
Choptank River, 293 km² watershed

WRTDS Example
“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)**
Locally Weighted Regression

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

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

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

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

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

Discharge in ft³/s


Mg/L

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
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!
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.
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.

Choptank River near Greensboro, MD  
Nitrate plus Nitrite, filtered, as N  
Locations of all available data
To organize the work, let's 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

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

Discharge in ft³/s

0.0 0.5 1.0 1.5 2.0

500 200 100 50 20


500 200 100 50 20

2009-Jan 2009-Jul 2010-Jan 2010-Jul 2011-Jan
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 $c \times Q$ 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 trend +40% 1985-2012

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

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.
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

Watershed

Streamgage & Sampling Location

Tidal River

Estuary

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.
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/
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.
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
```r
> 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)

<table>
<thead>
<tr>
<th></th>
<th>ConcLow</th>
<th>ConcHigh</th>
<th>Uncen</th>
<th>ConcAve</th>
<th>Julian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>1979-10-24</td>
<td>0.176</td>
<td>0.050</td>
<td>0.0000</td>
<td>47412</td>
</tr>
<tr>
<td>1st Qu.:</td>
<td>1989-03-18</td>
<td>0.900</td>
<td>0.900</td>
<td>1.0000</td>
<td>50845</td>
</tr>
<tr>
<td>Median:</td>
<td>1995-01-21</td>
<td>1.130</td>
<td>1.130</td>
<td>1.0000</td>
<td>52980</td>
</tr>
<tr>
<td>Mean:</td>
<td>1996-10-21</td>
<td>1.138</td>
<td>1.137</td>
<td>0.9986</td>
<td>53620</td>
</tr>
<tr>
<td>3rd Qu.:</td>
<td>2004-10-12</td>
<td>1.400</td>
<td>1.400</td>
<td>1.0000</td>
<td>56532</td>
</tr>
<tr>
<td>Max.:</td>
<td>2014-08-13</td>
<td>2.430</td>
<td>2.430</td>
<td>1.0000</td>
<td>60124</td>
</tr>
<tr>
<td>NA's:</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month</td>
<td>Min. : 1.000</td>
<td>Min. : 2.00</td>
<td>Min. : 1980</td>
<td>Min. : 1558</td>
<td>Min. : 1.00000</td>
</tr>
<tr>
<td>Day</td>
<td>1st Qu.: 3.000</td>
<td>1st Qu.: 83.75</td>
<td>1st Qu.: 1989</td>
<td>1st Qu.: 1671</td>
<td>1st Qu.: -0.62876</td>
</tr>
<tr>
<td>DecYear</td>
<td>Median : 6.000</td>
<td>Median : 157.00</td>
<td>Median : 1995</td>
<td>Median : 1741</td>
<td>Median : 0.19667</td>
</tr>
<tr>
<td>MonthSeq</td>
<td>Mean : 6.082</td>
<td>Mean : 169.23</td>
<td>Mean : 1997</td>
<td>Mean : 1762</td>
<td>Mean : 0.09121</td>
</tr>
<tr>
<td>SinDY</td>
<td>3rd Qu.: 9.000</td>
<td>3rd Qu.: 256.25</td>
<td>3rd Qu.: 2005</td>
<td>3rd Qu.: 1858</td>
<td>3rd Qu.: 0.79226</td>
</tr>
<tr>
<td>CosDY</td>
<td>Max. : 12.000</td>
<td>Max. : 364.00</td>
<td>Max. : 2015</td>
<td>Max. : 1976</td>
<td>Max. : 0.99992</td>
</tr>
</tbody>
</table>
> length(Sample$Date)
[1] 708
```
If the data are in a spreadsheet, it can be entered as a csv file with 3 columns:

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

<table>
<thead>
<tr>
<th></th>
<th>Sample Date</th>
<th>rem</th>
<th>PHOSPHOROUS TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5/28/1974</td>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>7/30/1974</td>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>8/21/1974</td>
<td></td>
<td>0.067</td>
</tr>
<tr>
<td>5</td>
<td>9/17/1974</td>
<td></td>
<td>0.045</td>
</tr>
<tr>
<td>6</td>
<td>10/22/1974</td>
<td></td>
<td>0.06</td>
</tr>
<tr>
<td>7</td>
<td>12/10/1974</td>
<td></td>
<td>0.032</td>
</tr>
<tr>
<td>8</td>
<td>1/21/1975</td>
<td></td>
<td>0.006</td>
</tr>
<tr>
<td>9</td>
<td>2/17/1975</td>
<td></td>
<td>0.041</td>
</tr>
<tr>
<td>10</td>
<td>3/18/1975</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>11</td>
<td>4/23/1975</td>
<td></td>
<td>0.355</td>
</tr>
</tbody>
</table>
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 designed 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)
EPA Storet Data from the Water Quality Portal

```r
> 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)

```

**Summary of Inorganic Nitrogen (Nitrate and Nitrite) Data**

<table>
<thead>
<tr>
<th>Date</th>
<th>ConcLow</th>
<th>ConcHigh</th>
<th>Uncen</th>
<th>ConcAve</th>
<th>Julian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>2005-01-24</td>
<td>0.041</td>
<td>0.0180</td>
<td>0.0090</td>
<td>56636</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>2009-02-08</td>
<td>3.658</td>
<td>0.1905</td>
<td>0.1905</td>
<td>58112</td>
</tr>
<tr>
<td>Median</td>
<td>2010-01-07</td>
<td>5.205</td>
<td>4.5950</td>
<td>4.5950</td>
<td>58446</td>
</tr>
<tr>
<td>Mean</td>
<td>2009-05-21</td>
<td>4.834</td>
<td>3.8710</td>
<td>3.8692</td>
<td>58215</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2011-03-03</td>
<td>6.560</td>
<td>6.2250</td>
<td>6.2250</td>
<td>58866</td>
</tr>
<tr>
<td>Max.</td>
<td>2011-11-28</td>
<td>11.400</td>
<td>11.4000</td>
<td>11.4000</td>
<td>59135</td>
</tr>
</tbody>
</table>

**NA's:** 8

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

<table>
<thead>
<tr>
<th>Date</th>
<th>Q</th>
<th>Julian</th>
<th>Month</th>
<th>Day</th>
<th>DecYear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.   :1979-10-01</td>
<td>Min. : 0.00991</td>
<td>Min. : 47389</td>
<td>Min. : 1.000</td>
<td>Min. : 1.0</td>
<td>Min. : 1980</td>
</tr>
<tr>
<td>1st Qu. :1988-06-30</td>
<td>1st Qu.: 0.96277</td>
<td>1st Qu.: 50584</td>
<td>1st Qu.: 4.000</td>
<td>1st Qu.: 93.0</td>
<td>1st Qu.: 1988</td>
</tr>
<tr>
<td>3rd Qu. :2005-12-28</td>
<td>3rd Qu.: 4.72891</td>
<td>3rd Qu.: 56975</td>
<td>3rd Qu.:10.000</td>
<td>3rd Qu.:275.0</td>
<td>3rd Qu.:2006</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MonthSeq</th>
<th>Qualifier</th>
<th>i</th>
<th>LogQ</th>
<th>Q7</th>
<th>Q30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.   :1558</td>
<td>Length:12782</td>
<td>Min. : 1</td>
<td>Min. : -4.61412</td>
<td>Min. : 0.01808</td>
<td>Min. : 0.09606</td>
</tr>
<tr>
<td>1st Qu.:1662</td>
<td>Class:character</td>
<td>1st Qu.: 3196</td>
<td>1st Qu.: -0.03794</td>
<td>1st Qu.: 1.00727</td>
<td>1st Qu.: 1.21102</td>
</tr>
<tr>
<td>Median :1767</td>
<td>Mode :character</td>
<td>Median : 6392</td>
<td>Median : 0.90161</td>
<td>Median : 2.63549</td>
<td>Median : 2.97421</td>
</tr>
<tr>
<td>Mean :1767</td>
<td>Mean : 6392</td>
<td>Mean : 0.78216</td>
<td>Mean : 4.17433</td>
<td>Mean : 4.17615</td>
<td></td>
</tr>
<tr>
<td>3rd Qu.:1872</td>
<td>3rd Qu.: 9587</td>
<td>3rd Qu.: 1.55370</td>
<td>3rd Qu.: 5.09804</td>
<td>3rd Qu.: 5.88802</td>
<td></td>
</tr>
</tbody>
</table>

> 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
Your site for streamflow data is 01491000.

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

This name will be used to label graphs and tables.

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

Choptank River near Greensboro, MD
The latitude and longitude of the site are: 38.99719, -75.78581 (degrees north and west).

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

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

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

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

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

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

Dissolved Nitrate, mg/L 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:

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

\[
> \text{multiPlotDataOverview(eList,qUnit=1)}
\]
We’ve gone to all this effort, let’s save our work

```r
> savePath<"~/Users/rhirsch/Desktop/"
> saveResults(savePath,eList)
```
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
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
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)

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
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
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)

CHOPTANK RIVER NEAR GREENSBORO, MD  Nitrate as N
Water Year
Mean Concentration (dots) & Flow Normalized Concentration (line)
> plotFluxHist(eList, fluxUnit = 8)

CHOPTANK RIVER NEAR GREENSBORO, MD  Nitrate as N
Water Year
Flux Estimates (dots) & Flow Normalized Flux (line)

Flux in $10^3$kg/yr

0  50  100  150  200  250  300

> 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
### CHOPTANK RIVER NEAR GREENSBORO, MD
Nitrate as N
Water Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Discharge (cfs)</th>
<th>Conc (mg/L)</th>
<th>FN_Conc</th>
<th>Flux (tons/yr)</th>
<th>FN_Flux</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>150.2</td>
<td>0.926</td>
<td>0.953</td>
<td>125.5</td>
<td>117</td>
</tr>
<tr>
<td>1981</td>
<td>78.3</td>
<td>1.004</td>
<td>0.963</td>
<td>72.6</td>
<td>119</td>
</tr>
<tr>
<td>1982</td>
<td>107.6</td>
<td>1.017</td>
<td>0.972</td>
<td>107.0</td>
<td>121</td>
</tr>
<tr>
<td>1983</td>
<td>176.1</td>
<td>0.998</td>
<td>0.984</td>
<td>144.4</td>
<td>124</td>
</tr>
<tr>
<td>1984</td>
<td>201.9</td>
<td>0.988</td>
<td>0.997</td>
<td>173.9</td>
<td>126</td>
</tr>
<tr>
<td>1985</td>
<td>53.6</td>
<td>1.055</td>
<td>1.011</td>
<td>53.8</td>
<td>129</td>
</tr>
<tr>
<td>1986</td>
<td>92.8</td>
<td>1.050</td>
<td>1.023</td>
<td>98.9</td>
<td>132</td>
</tr>
<tr>
<td>1987</td>
<td>119.1</td>
<td>1.064</td>
<td>1.043</td>
<td>124.7</td>
<td>135</td>
</tr>
<tr>
<td>1988</td>
<td>66.0</td>
<td>1.121</td>
<td>1.079</td>
<td>72.4</td>
<td>139</td>
</tr>
<tr>
<td>2007</td>
<td>151.2</td>
<td>1.395</td>
<td>1.373</td>
<td>177.1</td>
<td>168</td>
</tr>
<tr>
<td>2008</td>
<td>90.5</td>
<td>1.459</td>
<td>1.386</td>
<td>112.8</td>
<td>169</td>
</tr>
<tr>
<td>2009</td>
<td>130.0</td>
<td>1.404</td>
<td>1.402</td>
<td>147.9</td>
<td>170</td>
</tr>
<tr>
<td>2010</td>
<td>254.0</td>
<td>1.310</td>
<td>1.417</td>
<td>256.4</td>
<td>172</td>
</tr>
<tr>
<td>2011</td>
<td>185.2</td>
<td>1.417</td>
<td>1.431</td>
<td>179.0</td>
<td>174</td>
</tr>
<tr>
<td>2012</td>
<td>122.6</td>
<td>1.480</td>
<td>1.445</td>
<td>157.1</td>
<td>175</td>
</tr>
<tr>
<td>2013</td>
<td>226.0</td>
<td>1.376</td>
<td>1.460</td>
<td>241.1</td>
<td>177</td>
</tr>
<tr>
<td>2014</td>
<td>191.8</td>
<td>1.411</td>
<td>1.475</td>
<td>233.0</td>
<td>179</td>
</tr>
</tbody>
</table>
CHOPTANK RIVER NEAR GREENSBORO, MD
Nitrate as N
Water Year

### Concentration trends

<table>
<thead>
<tr>
<th>time span</th>
<th>change</th>
<th>slope</th>
<th>change</th>
<th>slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mg/L</td>
<td>mg/L/yr</td>
<td>%</td>
<td>%/yr</td>
</tr>
<tr>
<td>1980 to 1995</td>
<td>0.25</td>
<td>0.017</td>
<td>26</td>
<td>1.7</td>
</tr>
<tr>
<td>1980 to 2014</td>
<td>0.52</td>
<td>0.015</td>
<td>55</td>
<td>1.6</td>
</tr>
<tr>
<td>1995 to 2014</td>
<td>0.27</td>
<td>0.014</td>
<td>23</td>
<td>1.2</td>
</tr>
</tbody>
</table>

### Flux Trends

<table>
<thead>
<tr>
<th>time span</th>
<th>change</th>
<th>slope</th>
<th>change</th>
<th>slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tons/yr</td>
<td>tons/yr/yr</td>
<td>%</td>
<td>%/yr</td>
</tr>
<tr>
<td>1980 to 1995</td>
<td>33</td>
<td>2.2</td>
<td>29</td>
<td>1.9</td>
</tr>
<tr>
<td>1980 to 2014</td>
<td>62</td>
<td>1.8</td>
<td>53</td>
<td>1.6</td>
</tr>
<tr>
<td>1995 to 2014</td>
<td>29</td>
<td>1.5</td>
<td>19</td>
<td>1</td>
</tr>
</tbody>
</table>
EGRET produces a diagnostic plot to help spot serious flux problems with the model.

```
fluxBiasMulti(eList, fluxUnit=4)
```
This same type of plot can be used to look at other models, here the LOADEST7
Diagnostics and potential problems with estimating mean flux, see:


Bottom line, look at the fit before you use a statistical model!!!
How difficult is it to make those contour plots?

```r
> 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)

Choptank River near Greensboro, MD, Nitrate, filtered, as N
Season Consisting of Jun Jul Aug Sep Oct Nov Dec Jan
For Discharge < 50 Cubic Feet per Second

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

Choptank River near Greensboro, MD, Nitrate, filtered, as N
Season Consisting of Dec Jan Feb Mar Apr
For Discharge > 200 Cubic Feet per Second
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.
WRTDS representation of concentration as a function of time and discharge

Maumee River at Waterville OH  HU_SR_P as P, mg/L
Estimated Concentration Surface in Color
Black lines are 5 and 95 flow percentiles

Discharge in m$^3$/s


USGS
Maumee River at Waterville OH  HU_SRP as P, mg/L
Water Year
Mean Concentration (dots) & Flow Normalized Concentration (line)

Flux Estimates (dots) & Flow Normalized Flux (line)
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.
Maumee River, SRP - Green is WRTDS Flow Normalized Flux
Red, Brown and Black are three bootstrap replicate estimates of Flow Normalized Flux
Each bootstrap replicate can give us an estimate of change between any two years (say 1994 and 2014)
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”
Histogram of trend in SRP as P in mg/L
Flow Normalized Flux 1994 to 2014
Maumee River at Waterville, OH Water Year

WRTDS estimate of trend

Upward trend highly likely
The EGRETci software translates the bootstrap results into a set of words

<table>
<thead>
<tr>
<th>Frequency of upwards trends in the bootstrap replicates</th>
<th>Likelihood words</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.95, 1.0)</td>
<td>Upward trend is highly likely</td>
</tr>
<tr>
<td>(0.90, 0.95)</td>
<td>Upward trend is very likely</td>
</tr>
<tr>
<td>(0.67, 0.90)</td>
<td>Upward trend is likely</td>
</tr>
<tr>
<td>(0.33, 0.67)</td>
<td>Upward trend is about as likely as not</td>
</tr>
<tr>
<td>(0.10, 0.33)</td>
<td>Upward trend is unlikely</td>
</tr>
<tr>
<td>(0.05, 0.1)</td>
<td>Upward trend is very unlikely</td>
</tr>
<tr>
<td>(0.0, 0.05)</td>
<td>Upward trend is highly unlikely</td>
</tr>
</tbody>
</table>
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
Maumee River at Waterville OH  HU_SRP as P, mg/L
Estimated Concentration Versus Discharge Relationship at 3 specific dates

Discharge in m$^3$/s

Concentration in mg/L
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)

Maumee River at Waterville OH  HU_SR as P, mg/L
Estimated Concentration change from 1994 to 2015

Discharge in m$^3$/s

Jan1  Mar1  May1  Jul1  Sep1  Nov1  Jan1
Maumee River at Waterville OH
HU_SRP as P, mg/L
Water Year

### Concentration trends

<table>
<thead>
<tr>
<th>time span</th>
<th>change</th>
<th>slope</th>
<th>change</th>
<th>slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mg/L</td>
<td>mg/L/yr</td>
<td>%</td>
<td>%/yr</td>
</tr>
<tr>
<td>1995 to 2005</td>
<td>0.025</td>
<td>0.0025</td>
<td>56</td>
<td>5.6</td>
</tr>
<tr>
<td>1995 to 2015</td>
<td>0.036</td>
<td>0.0018</td>
<td>80</td>
<td>4</td>
</tr>
<tr>
<td>2005 to 2015</td>
<td>0.01</td>
<td>0.001</td>
<td>15</td>
<td>1.5</td>
</tr>
</tbody>
</table>

### Flux Trends

<table>
<thead>
<tr>
<th>time span</th>
<th>change</th>
<th>slope</th>
<th>change</th>
<th>slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10^6 kg/yr</td>
<td>10^6 kg/yr /yr</td>
<td>%</td>
<td>%/yr</td>
</tr>
<tr>
<td>1995 to 2005</td>
<td>0.19</td>
<td>0.019</td>
<td>60</td>
<td>6</td>
</tr>
<tr>
<td>1995 to 2015</td>
<td>0.23</td>
<td>0.011</td>
<td>71</td>
<td>3.6</td>
</tr>
<tr>
<td>2005 to 2015</td>
<td>0.038</td>
<td>0.0038</td>
<td>7.4</td>
<td>0.74</td>
</tr>
</tbody>
</table>
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?*
“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.”

rhirsch@usgs.gov

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