

Seasonal Analysis in EGRET

USGS Office of Water Information



Introduction

This document describes how to obtain seasonal information from the R package EGRET. For example, we might want to know the fraction of the load that takes place in the winter season (say that is December, January, and February). We can look at the seasonal information for a single year, or averages over several years, or in terms of flow normalized fluxes.

Getting started

First, you need to have installed and loaded the EGRET package. Then, you'll need need to create an eList object. See the EGRET vignette or user guide [here](#) for more information on how to use your own water quality data in EGRET. Once the eList object has been created, run the modelEstimation function to create a WRTDS model.

For this post, we will use the Choptank River as an example. There is an example data set included in EGRET with measured nitrate in the Choptank River. Here is an example of how to load that example data:

```
library(EGRET) eList <- Choptank_eList
```

To start with lets look at the results calculated for complete water years.

```
tableResults(eList) ## ## Choptank River ## Inorganic nitrogen (nitrate and nitrite) ## Water Year ## ## Year Discharge Conc  
FN_Conc Flux FN_Flux ## cms mg/L 10^6 kg/yr ## ## 1980 4.25 0.949 1.003 0.1154 0.106 ## 1981 2.22 1.035 0.999 0.0675  
0.108 ## 1982 3.05 1.036 0.993 0.0985 0.110 ## 1983 4.99 1.007 0.993 0.1329 0.112 .... ## 2006 3.59 1.382 1.362 0.1409 0.147  
## 2007 4.28 1.408 1.382 0.1593 0.149 ## 2008 2.56 1.477 1.401 0.1008 0.149 ## 2009 3.68 1.409 1.419 0.1328 0.149 ## 2010  
7.19 1.323 1.438 0.2236 0.149 ## 2011 5.24 1.438 1.457 0.1554 0.148
```

Looking at the last column of these results we see that, for example, the flow normalized flux in water year 2010 is estimated to be 0.149 106 kg/year. Now, let's say we had a particular interest in the winter season which we define here as the months of

December, January, and February. Note that some lines (1984-2005) were removed from this blog post simply to save space.

The next step is to establish what season you are interested in looking at. All functions in EGRET can be done on the "water year" (Oct-Sept), the calendar year (Jan-Dec), or any set of sequential months. To define what period of analysis (PA) to use, there is a function setPA. The setPA function has two arguments:

- paStart is the number of the calendar month that is the start of the season.
- paLong is the length of the season in months (it can be any number from 1 to 12).

For example let's say we want to consider the winter, defined here as December through February, the starting month (paStart) would be 12, and the length (paLong) would be 3:

```
eList <- setPA(eList, paStart = 12, paLong = 3)
```

Now lets view our results when we are focusing on the winter season.

```
tableResults(eList) ## ## Choptank River ## Inorganic nitrogen (nitrate and nitrite) ## Season Consisting of Dec Jan Feb ## ##
Year Discharge Conc FN_Conc Flux FN_Flux ## cms mg/L 10^6 kg/yr ## ## 1980 4.220 1.10 1.11 0.1403 0.156 ## 1981 1.960
1.20 1.13 0.0735 0.161 ## 1982 5.057 1.16 1.15 0.1764 0.165 ## 1983 3.504 1.21 1.16 0.1310 0.169 .... ## 2006 5.843 1.51 1.51
0.2675 0.214 ## 2007 5.417 1.55 1.53 0.2419 0.216 ## 2008 2.436 1.62 1.55 0.1217 0.217 ## 2009 2.711 1.66 1.56 0.1396 0.216
## 2010 13.779 1.26 1.57 0.4161 0.215 ## 2011 3.369 1.66 1.57 0.1669 0.215
```

Note that now the estimated flow normalized flux is 0.215 106 kg/year. Let's take a moment to think about that in relation to the previous result. What it is saying is that the flux (mass per unit time) is greater during this three month period than the average flux for the entire water year. That makes sense, some seasons will have higher than average fluxes and other seasons will have lower than average fluxes.

Now, it might be that we want to think about these results, not in terms of a flux but in terms of mass alone. In the first result (water year) the mass for the year is 0.149 106 kg. But for the winter season we would compute the mass as $0.215 * 90 / 365$ to give us a result of 0.053 106 kg. To get this result we took the annual rate (0.215) and divided by the number of days in the year (365) to get a rate in mass per day and then multiplied by the number of days in the season (90) which is the sum of the length of those three months (31 + 31 + 28) to simply get a total mass.

We have taken pains here to run through this calculation because users have sometimes been confused, wondering how the flux for a part of the year can be greater than the flux for the whole year. Depending on the ultimate objective of the analysis one might want to present the seasonal results in either of these ways (mass or mass per unit time).

Next we will look at descriptions of change.

Seasonal Changes

Let's use the tableChange function to explore the change from 1990 to 2010. We will do it first for the full water year and then for the winter season.

```
eList <- setPA(eList, paStart = 10, paLong = 12) tableChange(eList, yearPoints = c(1990,2010)) ## ## Choptank River ##
Inorganic nitrogen (nitrate and nitrite) ## Water Year ## ## Concentration trends ## time span change slope change slope ## mg/L
mg/L/yr % %/yr ## ## 1990 to 2010 0.31 0.016 28 1.4 ## ## ## Flux Trends ## time span change slope change slope ## 10^6
kg/yr 10^6 kg/yr /yr % %/yr ## 1990 to 2010 0.02 0.00099 15 0.76 eList <- setPA(eList, paStart = 12, paLong = 3)
tableChange(eList, yearPoints = c(1990,2010)) ## ## Choptank River ## Inorganic nitrogen (nitrate and nitrite) ## Season
Consisting of Dec Jan Feb ## ## Concentration trends ## time span change slope change slope ## mg/L mg/L/yr % %/yr ## ##
1990 to 2010 0.24 0.012 18 0.89 ## ## ## Flux Trends ## time span change slope change slope ## 10^6 kg/yr 10^6 kg/yr /yr %
%/yr ## 1990 to 2010 0.02 0.001 10 0.51
```

What we see is a fairly large difference between the concentration trends for the winter as compared to the whole year (concentration rose more steeply for the winter than it did for the year on average). But what we are going to focus on here is the trend in flux. The first thing to note is that the change from 1990 to 2010 is identical for the winter season and the year as a whole. That is, the change in the flow normalized flux for the full water year is +0.020 106 kg/year (it went from 0.129 to 0.149) and then, when we look at the winter season the change is also +0.020 106 kg/yr (from 0.195 to 0.215). So the change for this season over the 20 year period is essentially the same as the change for the entire water year. In other words, the results tell us that although the change in flux (mass per unit time) for the winter is the same as for the full year, the change in the mass for the winter season is about 25% of the change for the full year (because the winter consists of about 25% of the days in the year). Thus, we can conclude that the winter change is not atypical of the changes for the other parts of the year.

The results shown above also express the change as a slope (either 0.001 or 0.00099 virtually identical to each other) and these are simply the change results divided by the number of years. The next entry in the tableChange output is for change expressed as %. Here we see a big difference between the winter and the whole year. The whole year shows an increase of + 15 % over the 20 years, while the winter season shows an increase of 10 %. That is because the same amount of increase 0.02 106 kg / year is being compared to a the smaller number (1990 flow normalized annual flux of 0.129 106 kg/year) in the first table and compared to a larger number (seasonal flux of 0.195 106 kg/year) in the second table. So, even though the change is focused equally in the winter and non-winter months, the percentage change for the winter is smaller than the percentage change for the whole year.

Seasonal Load Fraction

Next, we can think about the seasonal load fraction.

You will need to read in two new functions called setupSeasons and setupYearsPlus designed for this purpose. You can copy them from here and paste them into your workspace (all as a single copy and paste) or you can create an .R file from them that you will source each time you want to use them. The functions use the package dplyr, a package that is useful for general data exploration

and manipulation.

```
library(dplyr) setupSeasons <- function(eList){ Daily <- eList$Daily SeasonResults <- setupYearsPlus(Daily, paLong =
eList$INFO$paLong, paStart = eList$INFO$paStart) AnnualResults <- setupYearsPlus(Daily, paLong = 12, paStart =
eList$INFO$paStart) %>% filter(Counts >= 365) #To make sure we are getting full years divideBy <- 1000000 annualPctResults
<- AnnualResults %>% mutate(FluxYear = Flux*Counts/divideBy, FNFluxYear = FNFlux*Counts/divideBy) %>%
select(FluxYear, FNFluxYear, Year) seasonPctResults <- SeasonResults %>% mutate(FluxSeason = Flux*Counts/divideBy,
FNFluxSeason = FNFlux*Counts/divideBy) %>% left_join(annualPctResults, by="Year") %>% mutate(pctFlux =
100*FluxSeason/FluxYear, pctFNFlux = 100*FNFluxSeason/FNFluxYear) %>% select(-Q, -Conc, -Flux, -FNFlux, -FNConc, -
Counts) %>% rename(seasonStart = paStart, seasonLong = paLong) return(seasonPctResults) } setupYearsPlus <- function
(Daily, paLong = 12, paStart = 10){ monthsToUse <- seq(paStart, length=paLong) monthsToUse[monthsToUse > 12] <-
monthsToUse[monthsToUse > 12] - 12 crossesYear <- paLong + (paStart - 1) > 12 AnnualResults <- Daily %>% mutate(Year =
as.integer(format(Date, "%Y"))) %>% filter(Month %in% monthsToUse) %>% mutate(Year = if(crossesYear){ ifelse(Month >=
paStart, Year + 1, Year) } else { Year }) %>% group_by(Year) %>% summarise(DecYear = mean(DecYear, na.rm = TRUE), Q =
mean(Q, na.rm = TRUE), Conc = mean(ConcDay, na.rm = TRUE), Flux = mean(FluxDay, na.rm = TRUE), FNConc =
mean(FNConc, na.rm = TRUE), FNFlux = mean(FNFlux, na.rm = TRUE), Counts = sum(!is.na(ConcDay))) %>% mutate(paLong
= paLong, paStart = paStart) return(AnnualResults) }
Simply use the loaded eList to calculate these seasonal load fractions. Let's go back to the winter season (Dec-Feb):
eList <- setPA(eList, paStart = 12, paLong = 3) seasonPctResults <- setupSeasons(eList)
```

Looking at your results

What you now have is a data frame called seasonPctResults. The columns it contains are the following:

variable	Definition
DecYear	Decimal Year of the mid-date of the season
Year	Calendar Year of mid-date of the year
FluxYear	Estimated flux for the year in millions of kg
FNFluxYear	Flow Normalized flux for the year in millions of kg
FluxSeason	Estimated flux for the season in millions of kg
FNFluxSeason	Flow Normalized flux for the season in millions of kg
pctFlux	Season flux as a percentage of Annual Flux
pctFNFlux	FlowNormalized Seasonal Flux as a percent of Flow Normalized Annual Flux
seasonLong	Length of the Season in Months
seasonStart	Starting Month of the Season, 1=January

Plotting the time series

We can make a graph showing the percentage flux (estimated annual and flow normalized). Note, this workflow uses base R plotting functions. You could also use the EGRET function genericEGRETDotPlot to automatically to pick some plotting styles that are consistent with other EGRET plots.

```
plotTitle = paste("Seasonal Flux as a Percent of Annual Flux\n", eList$INFO$shortName, eList$INFO$paramShortName, "\nSolid
line is percentage of flow normalized flux") par(mar=c(5,6,4,2) + 0.1,mgp=c(3,1,0)) plot(seasonPctResults$DecYear,
seasonPctResults$pctFlux,pch=20, yaxs="i",ylim = c(0,100),las=1,tck=.01, xlab="Year",ylab="Percentage of Annual Flux",
main=plotTitle,cex=1.5) lines(seasonPctResults$DecYear,seasonPctResults$pctFNFlux,col="green",lwd=2) axis(3, labels =
FALSE,tck=.01) axis(4, labels = FALSE,tck=.01)
```

We can interpret this example graph as follows. The winter flux of nitrate fluctuates a good deal from year to year. From a low of around 10% to a high of around 60% but the mean percentage hasn't changed much over the years. It is around 35% of the annual total flux.

Computing averages over a period of years

Let's say we wanted to answer the question, what percentage of the annual total flux moved in the winter season during the years 2000 through 2010. We can answer that question with a simple set of calculations.

Keep in mind, the way we are defining "year" is what year the ending year of the period of analysis fell. So, for this analysis, the full 2010 "year" is from Dec. 2009 through the end of November 2010.

- Filter the data frame seasonPctResults for the years 2000 - 2010.
- Now we can compute the sum of the annual fluxes for those years and the sum of the seasonal fluxes for those years, and then get our answer by taking the ratio and multiplying by 100.

```
years00_10 <- filter(seasonPctResults, Year >= 2000, Year <= 2010) sumYears <- sum(years00_10$FluxYear) sumSeasons <-
sum(years00_10$FluxSeason) avePct <- 100 * sumSeasons / sumYears
```

The total flux for all years in the period of interest in millions of kg is sumYears = 1.7297595.

The total seasonal flux for all years of the period of interest in millions of kg is `sumSeasons = 0.6099614`.

The percentage of the total flux for the years 2000 through 2010 that was transported in the winter months is `avePct = 35.2627852`.

This can be determined for any set of years simply by changing Year values in the filter function. So, for the years 1990-1999:

```
years90_99 <- filter(seasonPctResults, Year >= 1990, Year <= 1999) c("sumYears" = sum(years00_10$FluxYear), "sumSeasons"
= sum(years00_10$FluxSeason), "avePct" = 100 * sumSeasons / sumYears) ## sumYears sumSeasons avePct ## 1.7297595
0.6099614 35.2627852
```

Questions

- Robert M. Hirsch
- Laura DeCicco

Information on USGS-R packages used in this post:

EGRET: Exploration and Graphics for RivEr Trends: An R-package for the analysis of long-term changes in water quality and streamflow, including the water-quality method Weighted Regressions on Time, Discharge, and Season (WRTDS) `dataRetrieval`:

This R package is designed to obtain USGS or EPA water quality sample data, streamflow data, and metadata directly from web services.
