The EflowStats R package

Jessica Thompson[1] and Stacey Archfield[1]

[1] *United States Geological Survey*

May 13, 2014

Contents

1 Introduction to EflowStats

The EflowStats package was created to simplify the process of generating hydrologic indicator statistics using daily streamflow records. It has been specifically designed to work seamlessly with USGS NWIS data, and works with the NWCCompare package to compare modeled and observed streamflow data. This package is intended to be an update of the previously existing USGS NAHAT program (1) with additional statistics previously published by Archfield et al (2).

Quantifying the amount of ecologically necessary streamflow and determining the effect of changes in landscape and climate on flow are critical aspects to understanding the ecological health of rivers. A natural flow regime can be characterized using five components of streamflow: magnitude, frequency, duration, timing and rate fo change of streamflow. Statistics representing these five properties are commonly applied and the number of ecologically-relevant streamflow statistics (ERSS) available from various software packages is now in the hundreds. The EflowStats package is designed to calculate a comprehensive set of ERSS using the open-source R software environement and provides batch capabilities for calculation of ERSS for multiple sites simultaneously. EflowStats is both directly available as an R package and integrated into the USGS National Water Census Data Platform. EflowStats will enable easy, transparent and repeatable calculation of many of the most utilized ERSS for many scientists. The EflowStats package is designed to work seamlessly with USGS NWIS web-available data, and users may also load data from other sources (text files, spreadsheets). Section 2 provides examples of how one can calculate selected stats from USGS or other data.

For information on getting started in R and installing the package, see (4): Getting Started.

2 General Workflow

This example vignette first loads the relevant packages (assuming they are already installed) and the included sample data. For regular use, you would need to load the packages, but not the given example data.

First, it is a good practice to clear any existing objects from your working environment. In RStudio, in the upper right quadrant, there is a broom icon that says 'Clear'. Click this and select 'Yes' to remove objects. This allows you to being work with a clear workspace, negating the possibility of mis-identified variables.

```r
library(EflowStats)

##########################################
# Load sample data included with package:
daily_data<-dailyData
```

To calculate stats for USGS streamgage(s), enter gage id(s), start and end dates and a list of the groups of statistics to be calculated. The list in the example contains all possible statistics.

```r
##########################################
# calculate stats for a USGS streamgage
sites <- c("02177000","02178400")
startdate <- "2009"
enddate <- "2013"
stats="magnifSeven,magStat,flowStat,durStat,timStat,rateStat,otherStat"
```

Once this information is entered, the user runs the ObservedStatsUSGS function and maps the result to a named data frame.

```r
############################################
# calculate stats for USGS streamgage(s)
statsout <- ObservedStatsUSGS(sites,startdate,enddate,stats)
```

The retrieved data is now present as a dataframe called statsout in your R environment. This data frame has one row for each gage and one column for each statistic. An example of the first few columns is given below.

```
    site_no lam1Obs tau2Obs tau3Obs tau4Obs ar1Obs phaseObs
1 02177000   697.0    0.40    0.31    0.15  -0.14    -1.15
2 02178400   190.7    0.39    0.36    0.25  -0.09    -1.05
```

This dataframe may now be used within the R environment or saved as a file as shown.

```r
# save statsout to a tab-delimited file
output = "output.txt"
write.table(statsout, file = output, col.names = TRUE, row.names = FALSE,
            quote = FALSE, sep = "\t")
```

For daily discharge data not available via NWISWeb, statistics can be calculated by loading data into an R dataframe like the example "dailyData". This dataframe has two columns, the first named date, containing dates as characters in the format "YYYY-MM-DD" and the second column, discharge, containing the numeric discharge. This can be easily accomplished from most text file formats using "read.delim" or

"read.table". Then the user must enter a drainage area and a site id for each site, as well as the stats string.

```
###########################################################
# calculate stats for data from your own data file
drain_area=54
site_id="Test site"
daily_data<-dailyData
stats="magnifSeven,magStat,flowStat,durStat,timStat,rateStat,otherStat"
statsout <- ObservedStatsOther(daily_data,drain_area,site_id,stats)
```

The generated statsout dataframe has the same characteristics as that shown previously.

There is also a function available to plot the monthly means for a daily discharge timeseries. Using included sample data, the following code generates a dataframe of mean monthly discharge and then plots the values with a title containing the site id.

```
# plot monthly means for a daily discharge timeseries
qfiletempf<-sampleData
meanmonts<-monthlyMeanTs(qfiletempf)
plotMonthlyMean(meanmonts,'02178400')
```
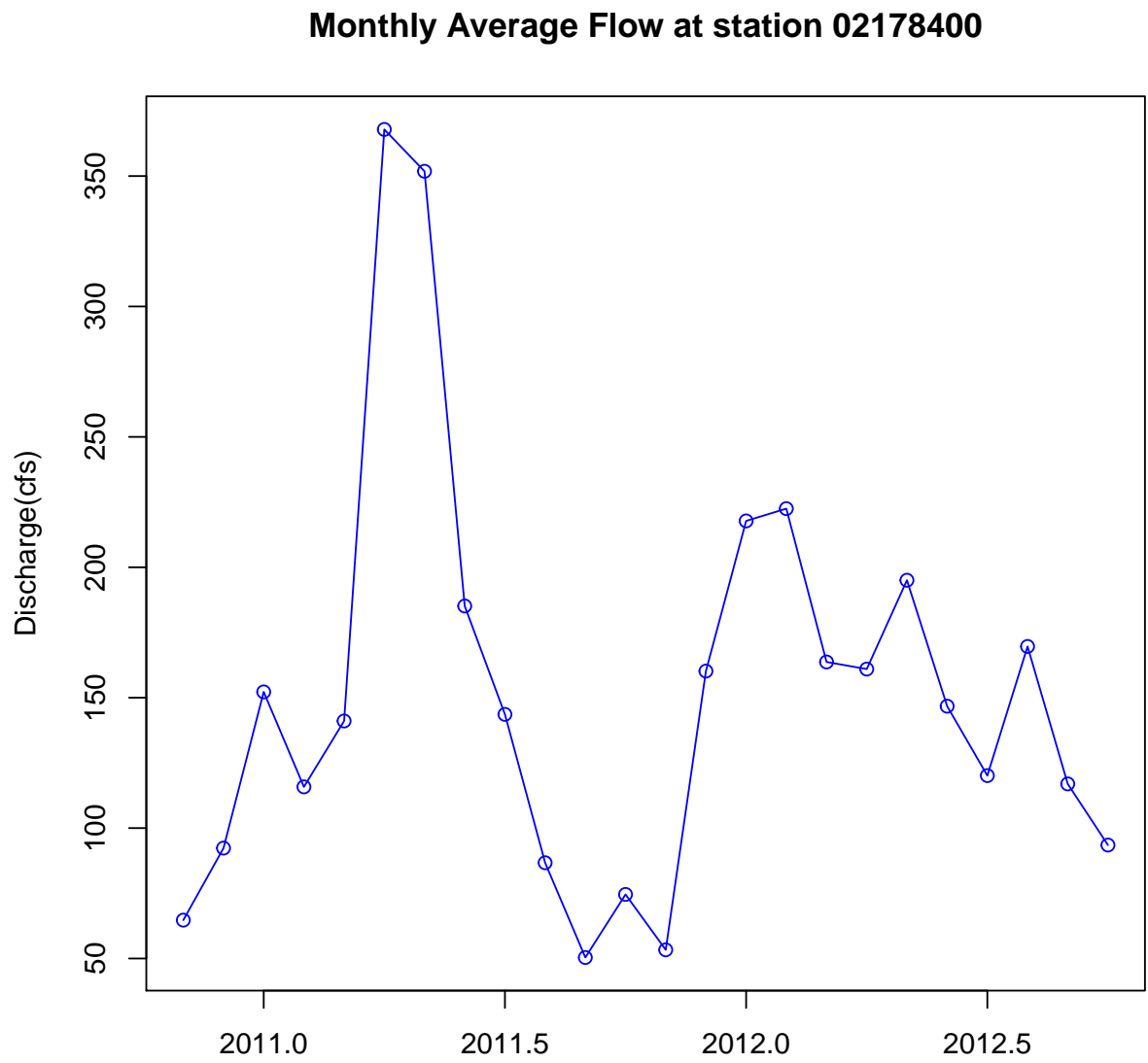
3

Monthly Average Flow at station 02178400

Figure 1: Monthly Average Flow at station 02178400

3 Stats Reference

Tables 1, 4, 5, 8, 9, 10 and 11 summarize the EflowStats functions:

Table 1: EflowStats magnitude functions

Function	Definition
ma1	Mean of the daily mean flow values for the entire flow record
ma2	Median of the daily mean flow values for the entire flow record
ma3	Mean (or median - use preference option) of the coefficients of variation (standard deviation/mean) for each year. Compute the coefficient of variation for each year of daily flows. Compute the mean of the annual coefficients of variation
ma4	Standard deviation of the percentiles of the logs of the entire flow record divided by the mean of percentiles of the logs. Compute the log(10) of the daily flows for the entire record. Compute the 5th, 10th, 15th, 20th, 25th, 30th, 35th, 40th, 45th, 50th, 55th, 60th, 65th, 70th, 75th, 80th, 85th, 90th and 95th percentiles for the logs of the entire flow record. Percentiles are computed by interpolating between the ordered (ascending) logs of the flow values. Compute the standard deviation and mean for the percentile values. Divide the standard deviation by the mean
ma5	The skewness of the entire flow record is computed as the mean for the entire flow record (ma1) divided by the median (ma2) for the entire flow record
ma6	Range in daily flows is the ratio of the 10-percent to 90-percent exceedence values for the entire flow record. Compute the 5-percent to 95-percent exceedence values for the entire flow record. Exceedence is computed by interpolating between the ordered (descending) flow values. Divide the 10-percent exceedence by the 90-percent value
ma7	Range in daily flows is computed in the same way as ma6 except using the 20-percent and 80-percent exceedence values. Divide the 20-percent exceedence value by the 80-percent value
ma8	Range in daily flows is computed in the same way as ma6 except using the 25-percent and 75-percent exceedence values. Divide the 25-percent exceedence value by the 75-percent value
ma9	Spread in daily flows is the ratio of the difference between the 90th and 10th percentile of the logs of the flow data to the log of the median of the entire flow record. Compute the log(10) of the daily flows for the entire record. Compute the 5th, 10th, 15th, 20th, 25th, 30th, 35th, 40th, 45th, 50th, 55th, 60th, 65th, 70th, 75th, 80th, 85th, 90th and 95th percentiles for the logs of the entire flow record. Percentiles are computed by interpolating between the ordered (ascending) logs of the flow values. Compute ma9 as (90th-10th)/log10(ma2)
ma10	Spread in daily flows is computed in the same way as ma9 except using the 20th and 80th percentiles
ma11	Spread in daily flows is computed in the same way as ma9 except using the 25th and 75th percentiles.
ma12.23	Means (or medians - use preference option) of monthly flow values. Compute the means for each month over the entire flow record. For example, ma12 is the mean of all January flow values over the entire record.
ma24.35	Variability (coefficient of variation) of monthly flow values. Compute the standard deviation for each month in each year over the entire flow record. Divide the standard deviation by the mean for each month. Take the mean (or median - use preference option) of these values for each month across all years.

Table 2: EflowStats magnitude functions continued

Function	Definition
ma36.40	Variability and skewness across monthly flows. ma36 - compute the minimum, maximum and mean flows for each month in the entire flow record. ma36 is the maximum monthly flow minus the minimum monthly flow divided by the median monthly flow. ma37 - compute the first (25th percentile) and the third (75th percentile) quartiles. ma37 is the third quartile minus the first quartile divided by the median of the monthly means. ma38 - compute the 10th and 90th percentiles for the monthly means. ma38 is the 90th percentile minus the 10th percentile divided by the median of the monthly means. ma39 - compute the standard deviation for the monthly means. ma39 is the standard deviation times 100 divided by the mean of the monthly means. ma40 - skewness in the monthly flows. ma40 is the mean of the monthly flow means minus the median of the monthly means divided by the median of the monthly means.
ma41.45	Annual runoff and the variability and skewness across annual flows. ma41 - compute the annual mean daily flows. ma41 is the mean of the annual means divided by the drainage area. ma42 is the maximum annual flow minus the minimum annual flow divided by the median annual flow. ma43 - compute the first (25th percentile) and third (75th percentile) quartiles for the annual means. ma43 is the third quartile minus the first quartile divided by the median of the annual means. ma44 - compute the 10th and 90th percentiles for the annual means. ma44 is the 90th percentile minus the 10th percentile divided by the median of the annual means. ma45 - skewness in the annual flows. ma45 is the mean of the annual flow means minus the median of the annual means divided by the median of the annual means.
ml1.12	Mean (or median - use preference option) minimum flows for each month across all years. Compute the minimum daily flow for each month over the entire flow record. For example ml1 is the mean of the minimums of all January flow values over the entire record.
ml13	Variability (coefficient of variation) across minimum monthly flow values. Compute the mean and standard deviation for the minimum monthly flows over the entire flow record. ml13 is the standard deviation times 100 divided by the mean minimum monthly flow for all years.
ml14.16	Minimum annual flow, low flow index and median annual minimum flow. ml14 - compute the minimum annual flow for each year. ml14 is the mean of the ratios of minimum annual flows to the median flow for each year. ml15 - low flow index. ml15 is the mean of the ratios of minimum annual flows to the mean flow for each year. ml16 - median of annual minimum flows. ml16 is the median of the ratios of minimum annual flows to the median flow for each year.
ml17	Base flow. Compute the mean annual flows. Compute the minimum of a 7-day moving average flow for each year and divide them by the mean annual flow for that year. ml17 is the mean (or median - use preference option) of those ratios.
ml18	Variability in base flow. Compute the standard deviation for the ratios of 7-day moving average flows to mean annual flows for each year. ml18 is the standard deviation times 100 divided by the mean of the ratios.
ml19	Base flow. Compute the ratios of the minimum annual flow for each year. ml19 is the mean (or median - use preference option) of these ratios times 100.
ml20	Base flow. Divide the daily flow record into 5-day blocks. Find the minimum flow for each block. Assign the minimum flow as a base flow for that block if 90 percent of that minimum flow is less than the minimum flows for the blocks on either side. Otherwise, set it to zero. Fill in the zero values using linear interpolation. Compute the total flow for the entire record and total base flow for the entire record. ml20 is the ratio of total base flow to total flow.
ml21	Variability across annual minimum flows. Compute the mean and standard deviation for the annual minimum flows. ml21 is the standard deviation times 100 divided by the mean.

Table 3: EflowStats magnitude functions continued

Function	Definition
ml22	Specific mean annual minimum flow. ml22 is the mean (or median - use preference option) of the annual minimum flows divided by the drainage area.
mh1.12	Mean (or median - use preference option) maximum flows for each month across all years. Compute the maximum daily flow for each month over the entire flow record. For example, mh1 is the mean of the maximums of all January flow values over the entire record.
mh13	Variability (coefficient of variation) across maximum monthly flow values. Compute the mean and standard deviation for the maximum monthly flows over the entire flow record. mh13 is the standard deviation times 100 divided by the mean maximum monthly flow for all years.
mh14	Median of annual maximum flows. Compute the annual maximum flows from monthly maximum flows. Compute the ratio of annual maximum flow to median annual flow for each year. mh14 is the median of these ratios.
mh15.17	High flow discharge index. mh15 is the 1-percent exceedence value for the entire record divided by the median flow for the entire record. mh16 is the 10-percent exceedence value for the entire record divided by the median flow for the entire record. mh17 is the 25-percent exceedence value for the entire record divided by the median flow for the entire record.
mh18	Variability across annual maximum flows. Compute the logs (log10) of the maximum annual flows. Find the standard deviation and mean for these values. mh18 is the standard deviation times 100 divided by the mean.
mh19	Skewness in annual maximum flows. Use the equation: (N2*sum(qm3)-3N*sum(qm)*sum(qm2)+2*sum(qm))3)/(N*(N-1)*(N-2)*stddev3) where N=number of years, qm=log10(annual maximum flows), stddev=standard deviation of the annual maximum flows.
mh20	Specific mean annual maximum flow. mh20 is the mean (or median - use preference option) of the annual maximum flows divided by the drainage area.
mh21	High flow volume index. Compute the average volume for flow events above a threshold equal to the median flow for the entire record. mh21 is the average volume divided by the median flow for the entire record.
mh22	High flow volume. Compute the average volume for flow events above a threshold equal to three times the median flow for the entire record. mh22 is the average volume divided by the median flow for the entire record.
mh23	High flow volume. Compute the average volume for flow events above a threshold equal to seven times the median flow for the entire record. mh23 is the average volume divided by the median flow for the entire record.
mh24	High peak flow. Compute the average peak flow value for flow events above a threshold equal to the median flow for the entire record. mh24 is the average peak flow divided by the median flow for the entire record.
mh25	High peak flow. Compute the average peak flow value for flow events above a threshold equal to three times the median flow for the entire record. mh25 is the average peak flow divided by the median flow for the entire record.
mh26	High peak flow. Compute the average peak flow value for flow events above a threshold equal to seven times the median flow for the entire record. mh26 is the average peak flow divided by the median flow for the entire record.
mh27	High peak flow. Compute the average peak flow value for flow events above a threshold equal to the 75th percentile value for the entire flow record. mh27 is the average peak flow divided by the median flow for the entire record.

Table 4: EflowStats frequency functions

Function	Definition
fl1.2	Low flood pulse count and variability in pulse count. fl1 - Compute the average number of flow events with flows below a threshold equal to the 25th-percentile value for the entire flow record. fl1 is the average (or median - use preference option) number of events. fl2 - Compute the standard deviation in the annual pulse counts for fl1. fl2 is 100 times the standard deviation divided by the mean pulse count.
fl3	Frequency of low pulse spells. Compute the average number of flow events with flow below a threshold equal to 5-percent of the mean flow values for the entire flow record. fl3 is the mean (or median - use preference option) number of events.
fh1.2	High flood pulse count and variability in pulse count. fh1 - Compute the average number of flow events with flows above a threshold equal to the 75th-percentile value for the entire flow record. fh1 is the mean (or median - use preference option) number of events. fh2 - compute the standard deviation in the annual pulse counts for fh1. fh2 is 100 times the standard deviation divided by the mean pulse count.
fh3	High flood pulse count. Compute the average number of days per year that the flow is above a threshold equal to three times the median flow for the entire record. fh3 is the mean (or median-use preference option) of the annual number of days for all years.
fh	High flood pulse count. Compute the average number of days per year that the flow is above a threshold equal to seven times the median flow for the entire record. fh4 is the mean (or median-use preference option) of the annual number of days for all years.
fh5	Flood frequency. Compute the average number of flow events with flows above a threshold equal to the median flow value for the entire flow record. fh5 is the mean (or median - use preference option) number of events.
fh6	Flood frequency. Compute the average number of flow events with flows above a threshold equal to three times the median flow value for the entire flow record. fh6 is the mean (or median - use preference option) number of events.
fh7	Flood frequency. Compute the average number of flow events with flows above a threshold equal to sevent times the median flow value for the entire flow record. fh7 is the mean (or median - use preference option) number of events.
fh8	Flood frequency. Compute the average number of flow events with flows above a threshold equal to the 25-percent exceedence value for the entire flow record. fh8 is the mean (or median - use preference option) number of events.
fh9	Flood frequency. Compute the average number of flow events with flows above a threshold equal to the 75-percent exceedence value for the entire flow record. fh9 is the mean (or median - use preference option) number of events.
fh10	Flood frequency. Compute the average number of flow events with flows above a threshold equal to the median of the annual minima for the entire flow record. fh10 is the mean (or median - use preference option) number of events.
fh11	Flood frequency. Compute the average number fo flow events with flows above a threshold equal to flow corresponding to a 1.67-year recurrence interval. fh11 is the mean (or median - use preference option) number of events.

Table 5: EflowStats duration functions

Function	Definition
dl1	Annual minimum daily flow. Compute the minimum 1-day average flow for each year. dl1 is the man (or median - use preference option) of these values
dl2	Annual minimum of 3-day moving average flow. compute the minimum of a 3-day moving average flow for each year. dl2 is the mean (or median - use preference option) of these values.
dl3	Annual minimum of 7-day moving average flow. Compute the minimum of a 7-day moving average flow for each year. dl3 is the mean (or median - use preference option) of these values.
dl4	Annual minimum of 30-day moving average flow. Compute the minimum of a 30-day moving average flow for each year. dl4 is the mean (or median - use preference option) of these values.
dl5	Annual minimum of 90-day moving average flow. Compute the minimum of a 90-day moving average flow for each year. dl5 is the mean (or median - use preference option) of these values.
dl6	Variability of annual minimum daily average flow. Compute the standard deviation for the minimum daily average flow. dl6 is 100 times the standard deviation divided by the mean.
dl7	Variability of annual minimum of 3-day moving average flow. Compute the standard deviation for the minimum 3-day moving averages. dl7 is 100 times the standard deviation divided by the mean.
dl8	Variability of annual minimum of 7-day moving average flow. Compute the standard deviation for the minimum 7-day moving averages. dl8 is 100 times the standard deviation divided by the mean.
dl9	Variability of annual minimum of 30-day moving average flow. Compute the standard deviation for the minimum 30-day moving averages. dl9 is 100 times the standard deviation divided by the mean.
dl10	Variability of annual minimum of 90-day moving average flow. Compute the standard deviation for the minimum 90-day moving averages. dl10 is 100 times the standard deviation divided by the mean.
dl11	Annual minimum daily flow divided by the median for the entire record. Compute the minimum daily flow for each year. dl11 is the mean of these values divided by the median for the entire record.
dl12	Annual minimum of 7-day moving average flow divided by the median for the entire record. Compute the minimum of a 7-day moving average flow for each year. dl12 is the mean of these values divided by the median for the entire record.
dl13	Annual minimum of 30-day moving average flow divided by the median for the entire record. Compute the minimum of a 30-day moving average flow for each year. dl13 is the mean of these values divided by the median for the entire record.
dl14	Low exceedence flows. Compute the 75-percent exceedence values for the entire flow record. dl14 is the exceedence value divided by the median for the entire record.
dl15	Low exceedence flows. Compute the 90-percent exceedence values for the entire flow record. dl15 is the exceedence value divided by the median for the entire record.
dl16.17	Low flow pulse duration and variability in low pulse duration. dl16 - Compute the average pulse duratino for each year for flow events below a threshold equal to the 25th-percentile value for the entire flow record. dl16 is the median of the yearly average durations. dl17 - compute the standard deviation for the yearly average low pulse durations. dl17 is 100 times the standard deviation divided by the mean of yearly average low pulse durations.

Table 6: EflowStats duration functions continued

Function	Definition
dl18	Number of zero-flow days. Count the number of zero-flow days for the entire flow record. dl18 is the mean (or median - use preference option) annual number of zero flow days.
dl19	Variability is number of zero-flow days. Compute the standard deviation for the annual number of zero-flow days. dl19 is 100 times the standard deviation divided by the mean annual number of zero-flow days.
dl20	Number of zero-flow months. Count the number of months in which there was no flow over the entire fow record.
dh1	Annual maximum daily flow. Compute the maximum of a 1-day moving average flow for each year. dh1 is the mean (or median - use preference option) of these values.
dh2	Annual maximum of 3-day moving average flows. Compute the maximum of a 3-day moving average flow for each year. dh2 is the mean (or median - use preference option) of these values.
dh3	Annual maximum of 7-day moving average flows. Compute the maximum of a 7-day moving average flow for each year. dh3 is the mean (or median - use preference option) of these values.
dh4	Annual maximum of 30-day moving average flows. Compute the maximum of a 30-day moving average flow for each year. dh4 is the mean (or median - use preference option) of these values.
dh5	Annual maximum of 90-day moving average flows. Compute the maximum of a 90-day moving average flow for each year. dh5 is the mean (or median - use preference option) of these values.
dh6	Variability of annual maximum daily flows. Compute the standard deviation for the maximum 1-day moving averages. dh6 is 100 times the standard deviation divided by the mean.
dh7	Variability of annual maximum 3-day moving average flows. Compute the standard deviation for the maximum 3-day moving averages. dh7 is 100 times the standard deviation divided by the mean.
dh8	Variability of annual maximum 7-day moving average flows. Compute the standard deviation for the maximum 7-day moving averages. dh8 is 100 times the standard deviation divided by the mean.
dh9	Variability of annual maximum 30-day moving average flows. Compute the standard deviation for the maximum 30-day moving averages. dh9 is 100 times the standard deviation divided by the mean.
dh10	Variability of annual maximum 90-day moving average flows. Compute the standard deviation for the maximum 90-day moving averages. dh10 is 100 times the standard deviation divided by the mean.
dh11	Annual maximum of 1-day moving average flows divided by the median for the entire record. Compute the maximum of a 1-day moving average flow for each year. dh11 is the mean of these values divided by the median for the entire record.
dh12	Annual maximum of 7-day moving average flows divided by the median for the entire record. Compute the maximum of a 7-day moving average flow for each year. dh12 is the mean of these values divided by the median for the entire record.
dh13	Annual maximum of 30-day moving average flows divided by the median for the entire record. Compute the maximum of a 30-day moving average flow for each year. dh13 is the mean of these values divided by the median for the entire record.

Table 7: EflowStats duration functions continued

Function	Definition
dh14	Flood duration. Compute the mean of the mean monthly flow values. Find the 95th percentile for the mean monthly flows. dh14 is the 95th percentile value divided by the mean of the monthly means.
dh15.16	High flow pulse duration and variability of high flow pulse duration. dh15 - Compute the average duration for flow events with flows above a threshold equal to the 75th percentile value for each year in the flow record. dh15 is the median of the yearly average durations. dh16 - Compute the standard deviation for the yearly average high pulse durations. dh16 is 100 times the standard deviation divided by the mean of the yearly average high pulse durations.
dh17	High flow duration. Compute the average duration of flow events with flows above a threshold equal to the median flow values for the entire flow record. dh17 is the mean (or median - use preference option) duration of the events.
dh18	High flow duration. Compute the average duration of flow events with flows above a threshold equal to three times the median flow values for the entire flow record. dh18 is the mean (or median - use preference option) duration of the events.
dh19	High flow duration. Compute the average duration of flow events with flows above a threshold equal to sevent times the median flow values for the entire flow record. dh19 is the mean (or median - use preference option) duration of the events.
dh20	High flow duration. Compute the 75th percentile value for the entire flow record. Compute the average duration of flow events with flows above a threshold equal to the 75th percentile value for the median annual flows. dh20 is the mean (or median - use preference option) duration of the events.
dh21	High flow duration. Compute the 25th percentile value for the entire flow record. Compute the average duration of flow events with flows above a threshold equal to the 25th percentile value for the median annual flows. dh21 is the mean (or median - use preference option) duration of the events.
dh22	Flood interval. Compute the flood threshold as the flow equivalent for a flood recurrence of 1.67 years. Determine the median number of days between flood events for each year. dh22 is the mean (or median - use preference option) of the yearly median number of days between flood events.
dh23	Flood duration. Compute the flood threshold as the flow equivalent for a flood recurrence of 1.67 years. Determine the number of days each year that the flow remains above the flood threshold. dh23 is the mean (or median - use preference option) of the number of flood days for years in which floods occur.
dh24	Flood-free days. Compute the flood threshold as the flow equivalent for a flood recurrence of 1.67 years. Compute the maximum number of days that the flow is below the threshold for each year. dh24 is the mean (or median - use preference option) of the maximum yearly no-flood days.

Table 8: EflowStats timing functions

Function	Definition
ta1.2	Constancy and predictability. ta1 - Constancy is computed via the formulation of Colwell (see example in Colwell, 1974). A matrix of values is compiled in which the rows are 11 flow categories and the columns are 365 (no February 29th) days of the year. The cell values are the number of times that a flow falls into a category on each day. The row totals, columns totals and grand total are computed. Using the equations for Shannon information theory parameters, constancy is computed as 1 - ((uncertainty with respect to state)/(log10(number of state))). ta2 - predictability is computed from the same matrix as constancy. It is computed as 1 - ((uncertainty with respect to interaction of time and state - uncertainty with respect to time)/(log10(number of state))).
ta3	Seasonal predictability of flooding. Divide years into 2-month periods (that is, Oct-Nov, Dec-Jan, and so forth). Count the number of flood days (flow events with flows ¿ 1.67-year flood) in each period over the entire flow record. ta3 is the maximum number of flood days in any one period divided by the total number of flood days.
tl1.2	Julian date of annual minimum and variability in julian date of annual minimum. tl1 - Determine the Julian date that the minimum flow occurs for each water year. Transform the dates to relative values on a circular scale (radians or degrees). Compute the x and y components for each year and average them all across all years. Compute the mean angle as the arc tangent of y-mean divided by x-mean. Transform the resultant angle back to Julian date. tl2 - Compute the coefficient of variation for the mean x and y components and convert to a date.
tl3	Seasonal predictability of low flow. Divide years into 2-month periods (that is Oct-Nov, Dec-Jan, and so forth). Count the number of low flow events (flow events with flows ¡= 5-year flood threshold) in each period over the entire flow record. tl3 is the maximum number of low flow events in any one period divided by the total number of low flow events.
tl4	Seasonal predictability of non-low flow. Compute the number of days that flow is above the 5-year flood threshold as the ratio of number of days to 365 or 366 (leap year) for each year. tl4 is the maximum of the yearly ratios.
th1.2	Julian date of annual maximum and variability in julian date of annual maximum. th1 - Determine the Julian date that the maximum flow occurs for each year. Transform the dates to relative values on a circular scale (radians or degrees). Compute the x and y components for each year and average them across all years. Compute the mean angle as the arc tangent of y-mean divided by x-mean. Transform the resultant angle back to Julian date. th2 - compute the coefficient of variation for the mean x and y components and convert to a date.
th3	Seasonal predictability of nonflooding. Computed as the maximum proportion of a 365-day year that the flow is less than the 1.67-year flood threshold and also occurs in all years. Accumulate nonflood days that span all years. th3 is maximum length of those flood-free periods divided by 365.

Table 9: EflowStats rate of change functions

Function	Definition
ra1	Rise rate. Compute the change in flow for days in which the change is positive for the entire flow record. ra1 is the mean (or median - use preference option) of these values.
ra2	Variability in rise rate. Compute the standard deviation for the positive flow changes. ra2 is 100 times the standard deviation divided by the mean.
ra3	Fall rate. Compute the change in flow for days in which the change is negative for the entire flow record. ra3 is the mean (or median - use preference option) of these values.
ra4	Variability in fall rate. Compute the standard deviation for the negative flow changes. ra4 is 100 times the standard deviation divided by the mean.
ra5	Number of day rises. Compute the number of days in which flow is greater than it was the previous day. ra5 is the number of positive-gain days divided by the total number of days in the flow record.
ra6	Change of flow. Compute the log10 of the flows for the entire flow record. Compute the change in log of flow for days in which the change is positive for the entire flow record. ra6 is the median of these log values.
ra7	Change of flow. Compute the log10 of the flows for the entire flow record. Compute the change in log of flow for days in which the change is negative for the entire flow record. ra7 is the median of these log values.
ra8.9	Number and variability of reversals. ra8 - compute the number of days in each year when the change in flow from one day to the next changes direction. ra8 is the mean (or median - use preference option) of the yearly values. ra9 - compute the standard deviation for the yearly reversal values. ra9 is 100 times the standard deviation divided by the mean.

Table 10: EflowStats other stats functions

Function	Definition
mean_flow	Mean of annual mean discharge values.
med_flow	Median of annual mean discharge values.
cv_flow	Coefficient of variation of annual mean discharge values.
cv_daily	Coefficient of variation of the daily discharge record.
flow_perc	Flow percentiles of the daily discharge record. Default is 10th, 15th, 25th, 50th, 75th and 90th percentiles, or others can be specified.

Table 11: EflowStats magnificent 7 functions

Function	Definition
lam1	Arithmetic mean
tau2	Coefficient of L-variation - analogous to coefficient of variation.
tau3	The third L-moment ratio or L-skew.
tau4	The fourth L-moment ratio or L-kurtosis.
ar1	AR1 correlation coefficient.
amplitude	Amplitude of the seasonal signal. Compute seasonality variables by first standardizing flows, using the fitting relation A*cos(2*pi*t) + B*sin(2*pi*t).
phase	Phase of the seasonal signal. Compute seasonality variables by first standardizing flows, the fitting relation A*cos(2*pi*t) + B*sin(2*pi*t).

4 Getting Started in R

This section describes the options for downloading and installing the EflowStats package.

4.1 New to R?

If you are new to R, you will need to first install the latest version of R, which can be found here: `http://www.r-project.org/`. There is also a useful USGS site for R help at `http://bwtst.usgs.gov/apps/R/index.html`.

There are many options for running and editing R code, one nice environment to learn R is RStudio. RStudio can be downloaded here: `http://rstudio.org/`. Once R and RStudio are installed, the dataRetrieval package needs to be installed as described in the next section.

At any time, you can get information about any function in R by typing a question mark before the functions name. This will open a file (in RStudio, in the Help window) that describes the function, the required arguments, and provides working examples.

```
library(EflowStats)
?mh19
```

To see the raw code for a particular code, type the name of the function:

```
mh19

function (x)
{
    annmax <- aggregate(x$discharge, list(x$wy_val), FUN = max,
        na.rm = TRUE)
    log_disch <- log10(annmax$x)
    sumq3 <- sum(log_disch^3)
    sumq2 <- sum(log_disch^2)
    sumq <- sum(log_disch)
    num_years <- length(unique(x$wy_val))
    qstd <- sd(log_disch)
    mh19 <- round(((num_years * num_years * sumq3) - (3 * num_years *
        sumq * sumq2) + (2 * sumq * sumq * sumq))/(num_years *
        (num_years - 1) * (num_years - 2) * qstd * qstd * qstd),
        digits = 2)
    return(mh19)
}
<environment: namespace:EflowStats>
```

4.2 R User: Installing EflowStats

Before installing EflowStats, the supporting packages must be first be installed:

```
install.packages(c("zoo","chron","doBy","XML","hydroGOF","lmomco","RCurl"))
install.packages("EflowStats",repos="http://usgs-r.github.com",type="source")
```

It is a good idea to re-start R after installing the package, especially if installing an updated version. Some users have found it necessary to delete the previous version's package folder before installing newer version of dataRetrieval. If you are experiencing issues after updating a package, trying deleting the package folder - the default location for Windows is something like this: C:/Users/userA/Documents/R/win-library/2.15/dataRetrieval, and the default for a Mac: /Users/userA/Library/R/2.15/library/dataRetrieval. Then, re-install the package using the directions above. Moving to CRAN should solve this problem.

After installing the package, you need to open the library each time you re-start R. This is done with the simple command:

```
library(EflowStats)
```

References

[1] Henriksen, J.A., Heasley, J. Kennen, J.G., and Nieswand, S., 2006, Users' manual for the Hydroecological Integrity Assessment Process software (including the New Jersey Assessment Tools): U.S. Geological Survey Open-File Report 2006-1093. 72 p. http://www.fort.usgs.gov/products/publications/21598/21598.pdf

[2] Archfield, S.A., J.G. Kennen, D.M. Carlisle, and D.M. Wolock. 2013. An Objective and Parsimonious Approach for Classifying Natural Flow Regimes at a Continental Scale. River Res. Applic. doi: 10.1002/rra.2710 http://onlinelibrary.wiley.com/doi/10.1002/rra.2710/abstract