Investigating Baseflow and Observed Discharge

Wes Porter

August 31, 2017

#### Introduction

The goal of this workflow is to investigate if any observable trends exist in our observed data - namely the baseflow and the observed discharge (baseflow removed).

The observed discharge data that we are using to compare our simulated discharge results comes from USGS gage station:

07288500 BIG SUNFLOWER RIVER AT SUNFLOWER, MS

Before we can begin analyzing the observed discharge data to see if there are any trends that could affect our calibration-validation process, we need to import the raw data into R.

in\_loc <- file.path("E:","Wes","Work","USDA","raw","Mississippi","Ms\_BaseflowRemoval","Discharge\_Analysis\_AUGUST\_8\_2017","R\_input")  
  
# Read in USGS observed total streamflow  
obs <- read.csv(file.path(in\_loc,"observed\_discharge.csv"),stringsAsFactors = FALSE)  
obs$Date <- as.Date(obs$Date, '%m/%d/%Y')

.

#### Baseflow

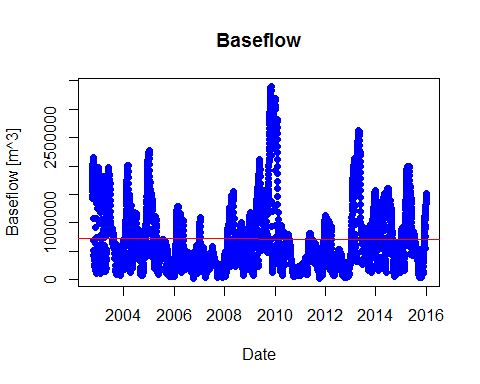
Once we have the data imported, we can use the hydrostats package to calculate the baseflow.

# Calculate baseflow from Observed  
obs\_bf <- select(obs,Date,Discharge)   
colnames(obs\_bf) <- c('Date','Q')  
obs\_bf <- baseflows(obs\_bf, ts='daily')

.

We can then plot the baseflow and add a linear regression line to see if there's a visual trend to the data.

x\_lab <- "Date"  
y\_lab <- "Baseflow [m^3]"  
  
plot(obs\_bf$Date, obs\_bf$bf, main = "Baseflow", xlab = x\_lab, ylab = y\_lab, col = "blue", pch = 16)  
  
lin\_reg <- lm(obs\_bf$bf ~ obs\_bf$Date)  
  
# Add line - Intercept of linear reg and slope  
abline(lin\_reg$coefficients[1],lin\_reg$coefficients[2], col = "red")

 .

At first glance it looks like there isn't a trend. So let's look at the summary statistics of our linear regression to verify what the plot is showing.

summary(lin\_reg)

##   
## Call:  
## lm(formula = obs\_bf$bf ~ obs\_bf$Date)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -700988 -456224 -163383 261651 2683878   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 769621.643 90385.168 8.515 <2e-16 \*\*\*  
## obs\_bf$Date -3.407 6.256 -0.545 0.586   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 607900 on 4837 degrees of freedom  
## Multiple R-squared: 6.134e-05, Adjusted R-squared: -0.0001454   
## F-statistic: 0.2967 on 1 and 4837 DF, p-value: 0.586

.

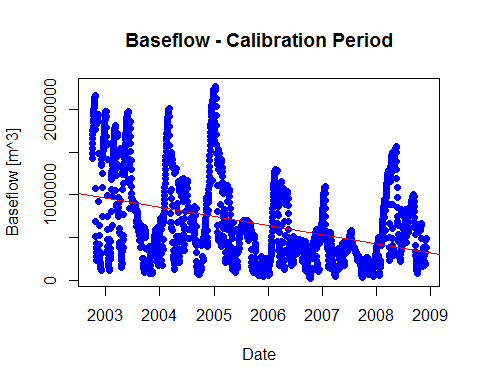
In this case we are only interested in seeing if there is a response in Baseflow across time, so the statistics for 'obs\_bf$Date' are what we should focus our consideration on. The statistics that we're most interested in are the 't value', 'Pr(>|t|)', and the 'F-statistic'.

The t-value indicates how many standard deviations our slope coefficient (-3.407) is from zero. The p-value suggests how likely it is that we would observe a relationship between the Date and Baseflow due to random chance. The F-statistic indicates whether there is a relationship between predictor and response variables. The further the F-statistic from 1, the more likely that a relationship exists.

When looking at Baseflow for the entire period 2002-2016, we have a t-value of -0.544, a p-value of 0.586, and a F-statistic of 0.2967 which indicate that there is not a statistically significant relationship between Baseflow and time. This supports our earlier graphical prognosis that there isn't a significant positive or negative trend in baseflow from 2002 to 2016.

But this doesn't tell us if there are trends to the baseflow during our calibration and validation periods. If we break up the baseflow into two parts based on the calibration period 2002-2008, and validation period 2008-2016, we can look to see if trends exits across shorter time periods.

# Baseflow Calibration period  
  
bf\_cal <- filter(obs\_bf,Date <= '2008-12-01')  
  
plot(bf\_cal$Date, bf\_cal$bf, main = "Baseflow - Calibration Period", xlab = x\_lab, ylab = y\_lab, col = "blue", pch = 16)  
  
fit\_cal <- lm(bf\_cal$bf ~ bf\_cal$Date)  
abline(lm(bf\_cal$bf ~ bf\_cal$Date), col = "red")

 .

When we look at baseflow during the calibration period, a downward sloping trend appears.

summary(fit\_cal)

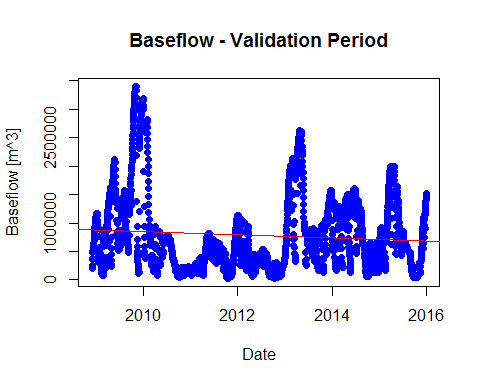
##   
## Call:  
## lm(formula = bf\_cal$bf ~ bf\_cal$Date)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -853524 -330395 -109274 318173 1526524   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4506479.72 202383.27 22.27 <2e-16 \*\*\*  
## bf\_cal$Date -294.24 15.44 -19.05 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 476800 on 2251 degrees of freedom  
## Multiple R-squared: 0.1389, Adjusted R-squared: 0.1385   
## F-statistic: 363 on 1 and 2251 DF, p-value: < 2.2e-16

.

The summary statistics of the linear regression report the same negative trend of baseflow over time during the calibration period. Our t-value (-19.05) is considerably far from zero and our p-value (2e-16) is extremely small indicating that there is a significant relationship between the Baseflow response and date.

The next question is whether the same downward trend for baseflow is observed during the validation period.

# Baseflow Validation period  
  
bf\_val <- filter(obs\_bf,Date >= '2008-12-01')  
  
plot(bf\_val$Date, bf\_val$bf, main = "Baseflow - Validation Period", xlab = x\_lab, ylab = y\_lab, col = "blue", pch = 16 )  
  
fit\_val <- lm(bf\_val$bf ~ bf\_val$Date)  
abline(lm(bf\_val$bf ~ bf\_val$Date), col = "red")

 .

Again it looks like we have a negative trend, though the slope here appears to be less steep than that of the calibration period.

summary(fit\_val)

##   
## Call:  
## lm(formula = bf\_val$bf ~ bf\_val$Date)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -795628 -502462 -157491 260437 2556283   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1922341.42 274689.84 6.998 3.29e-12 \*\*\*  
## bf\_val$Date -73.84 17.69 -4.173 3.10e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 672100 on 2585 degrees of freedom  
## Multiple R-squared: 0.006693, Adjusted R-squared: 0.006309   
## F-statistic: 17.42 on 1 and 2585 DF, p-value: 3.099e-05

.

According to our statistical results, we have a significant downward trend in baseflow over time during the validation period (p-value = 3.10e-05).

In both the calibration period and validation period, the baseflow is trending downward. Perhaps irrigation is a contributing factor to the decrease in baseflow during these periods. However when zooming out and considering the entire period from 2002-2016, there is **not** an observable trend due to peaks in baseflow especially during 2009 and early 2010.

.

#### Observed Discharge

Now lets look to see if we observe any trends over time in the observed discharge [m^3] data.

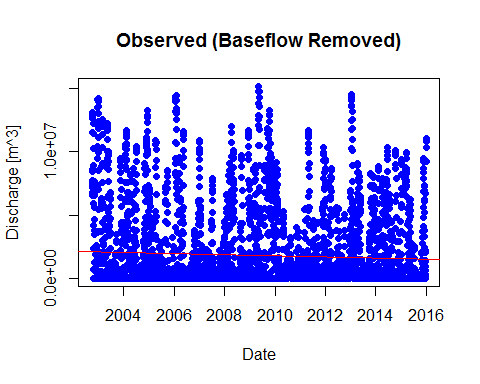
First we need to remove the baseflow from our observed discharge reported from the Big Sunflower gage station.

# Observed (Baseflow removed)  
  
obs\_bf[[5]] <- obs\_bf[[2]] - obs\_bf[[3]]  
colnames(obs\_bf)[5] <- 'Q\_bf\_removed'  
obs\_bf\_removed <- select(obs\_bf,Date,Q\_bf\_removed)  
colnames(obs\_bf\_removed) <- c('Date','Discharge')

.

Once the baseflow has been subtracted out of the observed discharge, we can plot it.

y\_lab2 <- "Discharge [m^3]"  
plot(obs\_bf\_removed$Date, obs\_bf\_removed$Discharge, main = "Observed (Baseflow Removed)", xlab = x\_lab, ylab = y\_lab2, col = "blue", pch = 16)  
  
obs\_fit <- lm(obs\_bf\_removed$Discharge ~ obs\_bf\_removed$Date)  
abline(lm(obs\_bf\_removed$Discharge ~ obs\_bf\_removed$Date), col = "red")

 .

It appears as though there is a slight negative trend in observed discharge from 2002-2016.

summary(obs\_fit)

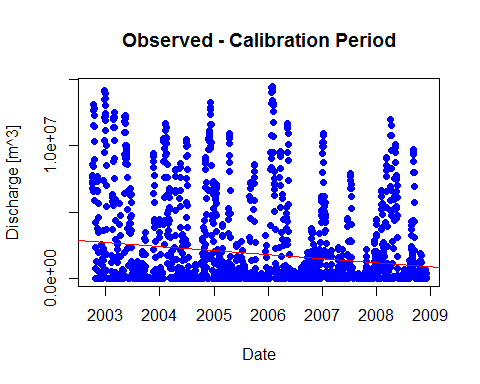
##   
## Call:  
## lm(formula = obs\_bf\_removed$Discharge ~ obs\_bf\_removed$Date)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2106401 -1844660 -1607606 501933 13350154   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3482563.56 464878.23 7.491 8.06e-14 \*\*\*  
## obs\_bf\_removed$Date -114.82 32.17 -3.569 0.000362 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3126000 on 4837 degrees of freedom  
## Multiple R-squared: 0.002626, Adjusted R-squared: 0.00242   
## F-statistic: 12.74 on 1 and 4837 DF, p-value: 0.0003621

.

The summary statistics appear to tell the same story. Our low p-value indicates a highly significant relationship between observed discharge and time. The f-statistic (12.74) is also relatively far from 1, indicating that the downward trend is likely not a random occurrence.

We can go a step further with our investigation and break the observed discharge down into calibration and validation periods like we did for the baseflow.

# Observed Calibration Period  
  
obs\_cal <- filter(obs\_bf\_removed,Date <= '2008-12-01')  
  
plot(obs\_cal$Date, obs\_cal$Discharge, main = "Observed - Calibration Period", xlab = x\_lab, ylab = y\_lab2, col = "blue", pch = 16)  
  
obs\_cal\_fit <- lm(obs\_cal$Discharge ~ obs\_cal$Date)  
abline(lm(obs\_cal$Discharge ~ obs\_cal$Date), col = "red")

 .

The trend is downward sloping for the calibration period as expected given the overall downward trend for the entire period 2002-2016.

summary(obs\_cal\_fit)

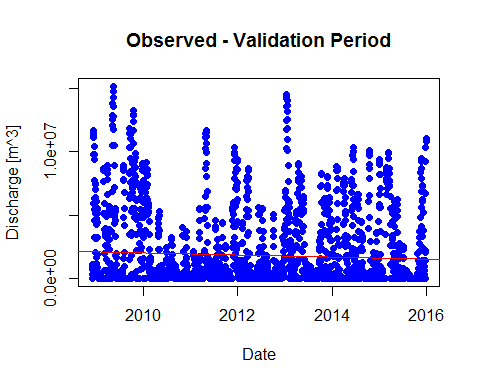
##   
## Call:  
## lm(formula = obs\_cal$Discharge ~ obs\_cal$Date)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2737284 -1902773 -1252880 298399 12729647   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 12659529.6 1364299.0 9.279 < 2e-16 \*\*\*  
## obs\_cal$Date -827.9 104.1 -7.952 2.88e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3214000 on 2251 degrees of freedom  
## Multiple R-squared: 0.02732, Adjusted R-squared: 0.02689   
## F-statistic: 63.23 on 1 and 2251 DF, p-value: 2.879e-15

.

We have a significant p-value and much higher F-statistic. The Date slope coefficient (-827.9) has also steepened.

A similar downward trend should occur in the validation period.

# Observed Validation Period  
  
obs\_val <- filter(obs\_bf\_removed,Date >= '2008-12-01')  
  
plot(obs\_val$Date, obs\_val$Discharge, main = "Observed - Validation Period", xlab = x\_lab, ylab = y\_lab2, col = "blue", pch = 16)  
  
obs\_val\_fit <- lm(obs\_val$Discharge ~ obs\_val$Date)  
abline(lm(obs\_val$Discharge ~ obs\_val$Date), col = "red")



summary(obs\_val\_fit)

##   
## Call:  
## lm(formula = obs\_val$Discharge ~ obs\_val$Date)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2134231 -1844057 -1586085 766623 13085159   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5403786.69 1230391.78 4.392 1.17e-05 \*\*\*  
## obs\_val$Date -230.02 79.25 -2.902 0.00373 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3010000 on 2585 degrees of freedom  
## Multiple R-squared: 0.003248, Adjusted R-squared: 0.002863   
## F-statistic: 8.424 on 1 and 2585 DF, p-value: 0.003734

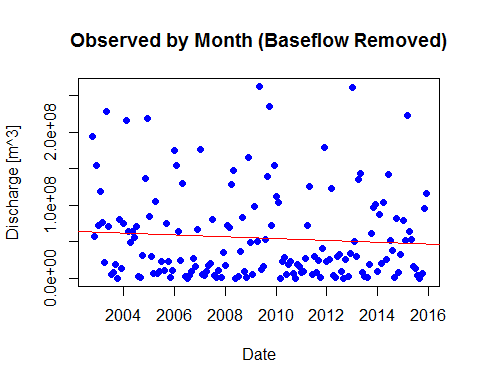
.

When we look at the plot and statistics for the validation period 2008-2016, the downward slope is still significant but not as steep.

#### Observed - Monthly

If we were to look at the data aggregated monthly, the same overall trends should hold true.

obs\_bf\_removed\_bymonth <- obs\_bf\_removed  
obs\_bf\_removed\_bymonth$Date <- floor\_date(obs\_bf\_removed\_bymonth$Date, "month")  
obs\_bf\_removed\_bymonth <- aggregate(Discharge~Date, data=obs\_bf\_removed\_bymonth, FUN=sum)  
  
plot(obs\_bf\_removed\_bymonth$Date, obs\_bf\_removed\_bymonth$Discharge, main = "Observed by Month (Baseflow Removed)", xlab = x\_lab, ylab = y\_lab2, col = "blue", pch = 16)  
  
obs\_fit <- lm(obs\_bf\_removed\_bymonth$Discharge ~ obs\_bf\_removed\_bymonth$Date)  
abline(lm(obs\_bf\_removed\_bymonth$Discharge ~ obs\_bf\_removed\_bymonth$Date), col = "red")

 .

#### Summary

When considering baseflow from 2002 to 2016, there is no positive or negative trend to the data over time. However when split into calibration (2002-2008) and validation (2008-2016) periods, downward trends emerge.

The overall trend for the observed discharge (baseflow removed) did show a a negative trend with discharge over time. As with the baseflow, the downward trend was less severe during the validation period than the calibration period. This makes sense considering the baseflow and observed discharge are directly correlated.

It would be useful to understand in what ways processes such as irrigation might be contributing to these downward trends, in order to better calibrate our model parameters.