Clean Discharge & Sampling Data

Tasks: Flowmeter Error Removal

PAZ

25 octobre 2016

Purpose

This document removes aberrant discahrge values in the flow meter data.

Used files:

1. **Alteck2016Debit.csv**

Produced file:

1. **hydroAlteck2016_NAs_R.csv** (Used for testing methods to replace missing values and smoothing the time series).

Required R-packages:

```
# Plotting functions
library("ggplot2")
library("scales")
library("tidyr")
```

```
## Warning: package 'tidyr' was built under R version 3.3.3
```

Working directory

```
# setwd("D:/Documents/these_pablo/Alteckendorf2016/R")
# setwd("/Users/DayTightChunks/Documents/PhD/Routput/Alteck/R")
# setwd("D:/Documents/these_pablo/Alteckendorf2016/00_TransparencyFolder/Discharge")
getwd()
```

```
## [1] "D:/Documents/these_pablo/Alteckendorf2016/HydrologicalMonitoring"
```

Import raw discharge data

```
dischargeAlteck = read.csv2("Data/Alteck2016Debit.csv", header = FALSE)

dischargeAlteck$Date = as.POSIXct(strptime(dischargeAlteck$V1,
                                           "%d/%m/%Y %H:%M",
                                           tz="EST"))
sum(is.na(dischargeAlteck$Date))
```

```
## [1] 0
```

```r
dischargeAlteck <- dischargeAlteck[!duplicated(dischargeAlteck$V1),]

dischargeAlteck = dischargeAlteck[,c(3,1:2)]
colnames(dischargeAlteck) <- c("Date", "DateCheck", "Q.m3Hrs")
head(dischargeAlteck)
```

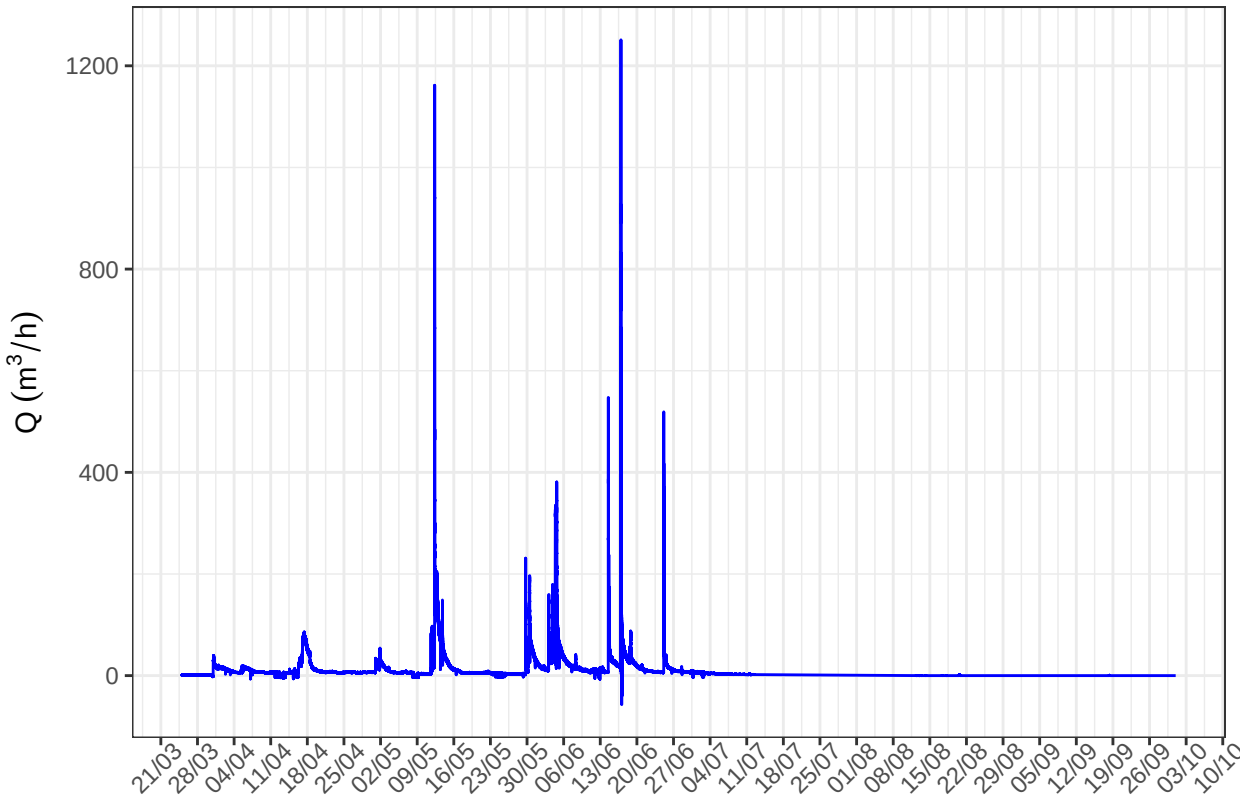
```
##                  Date        DateCheck Q.m3Hrs
## 1 2016-03-25 00:00:00 25/03/2016 00:00   1.256
## 2 2016-03-25 00:02:00 25/03/2016 00:02   1.219
## 3 2016-03-25 00:04:00 25/03/2016 00:04   1.192
## 4 2016-03-25 00:06:00 25/03/2016 00:06   1.212
## 5 2016-03-25 00:08:00 25/03/2016 00:08   1.195
## 6 2016-03-25 00:10:00 25/03/2016 00:10   1.219
```

```r
dischargeAlteck = subset(dischargeAlteck, Date < as.POSIXct("2016-10-01 00:00:00", tz="EST"))
```

Replace aberrant values with NA's

To evaluate which data needed correction and overview of the data was necessary:

```r
altp <- ggplot(dischargeAlteck, aes(x=Date, y=Q.m3Hrs))
altp + geom_line(colour = "blue") +
  theme_bw() +
  scale_x_datetime(breaks = date_breaks("weeks"), labels = date_format("%d/%m")) +
  theme(axis.text.x=element_text(angle = 45, hjust = 0.75)) +
  xlab("") +
  ylab(expression(paste("Q ",({m}^"3"/h))))
```



First, negative values were replaced with NA:

```
dischargeAlteck$Qna = dischargeAlteck$Q.m3Hrs
dischargeAlteck$Qna[dischargeAlteck$Q.m3Hrs <= 0.0] = NA
```

A number of subsections were then manually selected and unrealistic values converted to NA entries. Note that past May 30, only negative values have been removed (replaced by NA).

```
dischargeAlteck$Qna[dischargeAlteck$Q.m3Hrs < 1.0 &
                      dischargeAlteck$Date > as.POSIXct("2016-03-29 23:00:00 EST") &
                      dischargeAlteck$Date < as.POSIXct("2016-03-31 00:00:00 EST") ] = NA
dischargeAlteck$Qna[dischargeAlteck$Q.m3Hrs < 6 &
                      dischargeAlteck$Date > as.POSIXct("2016-04-01 23:00:00 EST") &
                      dischargeAlteck$Date < as.POSIXct("2016-04-04 00:00:00 EST") ] = NA
dischargeAlteck$Qna[dischargeAlteck$Q.m3Hrs < 7.5 &
                      dischargeAlteck$Date > as.POSIXct("2016-04-06 23:00:00 EST") &
                      dischargeAlteck$Date < as.POSIXct("2016-04-07 15:00:00 EST") ] = NA
dischargeAlteck$Qna[dischargeAlteck$Q.m3Hrs < 2.5 &
                      dischargeAlteck$Date > as.POSIXct("2016-04-9 00:00:00 EST") &
                      dischargeAlteck$Date < as.POSIXct("2016-04-17 00:00:00 EST") ] = NA
dischargeAlteck$Qna[dischargeAlteck$Q.m3Hrs < 2.5 &
                      dischargeAlteck$Date > as.POSIXct("2016-05-07 23:00:00 EST") &
                      dischargeAlteck$Date < as.POSIXct("2016-05-09 23:00:00 EST") ] = NA
dischargeAlteck$Qna[dischargeAlteck$Q.m3Hrs < 3.0 &
                      dischargeAlteck$Date > as.POSIXct("2016-05-22 23:00:00 EST") &
                      dischargeAlteck$Date < as.POSIXct("2016-05-30 23:00:00 EST") ] = NA
dischargeAlteck$Qna[dischargeAlteck$Q.m3Hrs < 100.0 &
```

```
                        dischargeAlteck$Date > as.POSIXct("2016-05-13 06:00:00 EST") &
                        dischargeAlteck$Date < as.POSIXct("2016-05-13 14:00:00 EST") ] = NA
```

The following is an example of the removed abberant values:

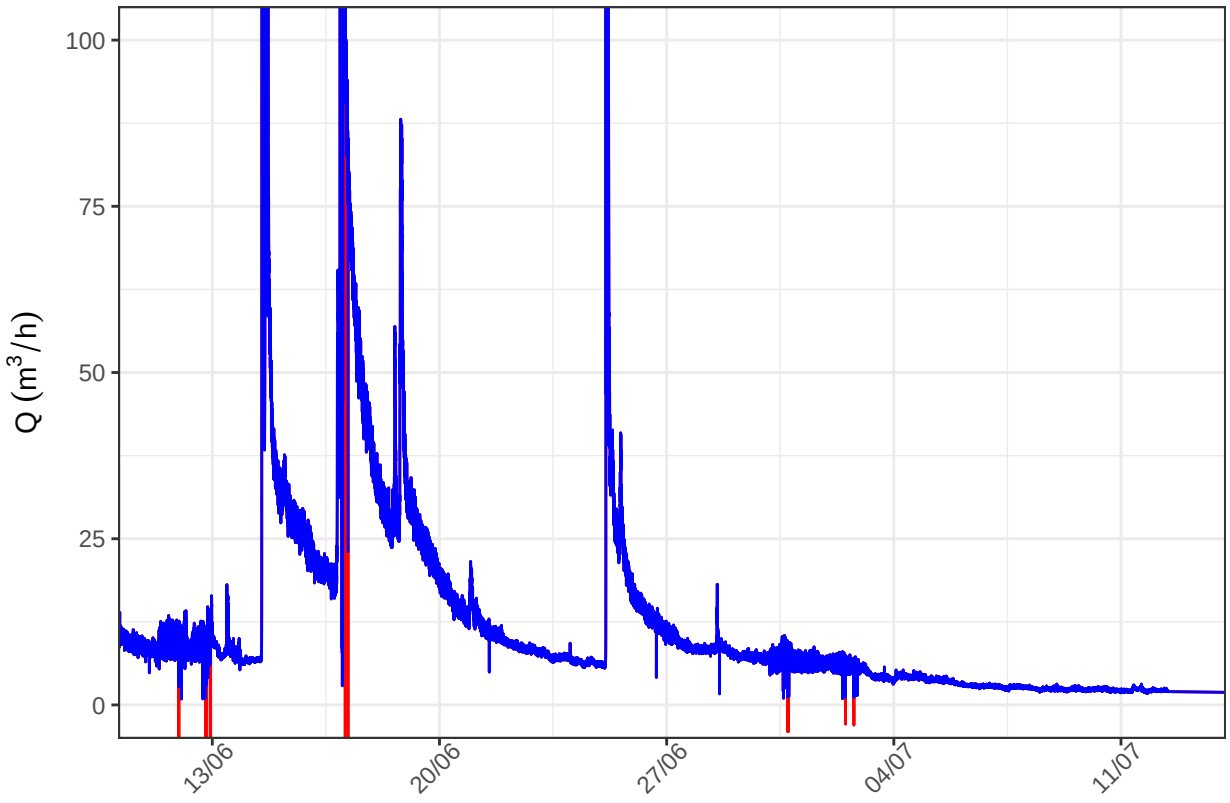
1. Example 1

Red geom_line has been omitted in final data.

```
altpTest = ggplot(dischargeAlteck, aes(x=Date)) +
  geom_line(aes(y= Q.m3Hrs), colour = "red") +
  geom_line(aes(y=Qna), colour = "blue") +
  coord_cartesian(xlim = c(as.POSIXct("2016-06-11 23:00:00 EST"),
                           as.POSIXct("2016-07-12 23:00:00 EST")),
                  ylim = c(0, 100)) + # no.1
  theme_bw() +
  scale_x_datetime(breaks = date_breaks("weeks"), labels = date_format("%d/%m")) +
  theme(axis.text.x=element_text(angle = 45, hjust = 0.75)) +
  #scale_y_continuous(trans=log_trans(), breaks=c(1,10,100,1000)) +
  xlab("") +
  ylab(expression(paste("Q ",({m}^"3"/h))))

altpTest
```

```
## Warning: Removed 9034 rows containing missing values (geom_path).
```



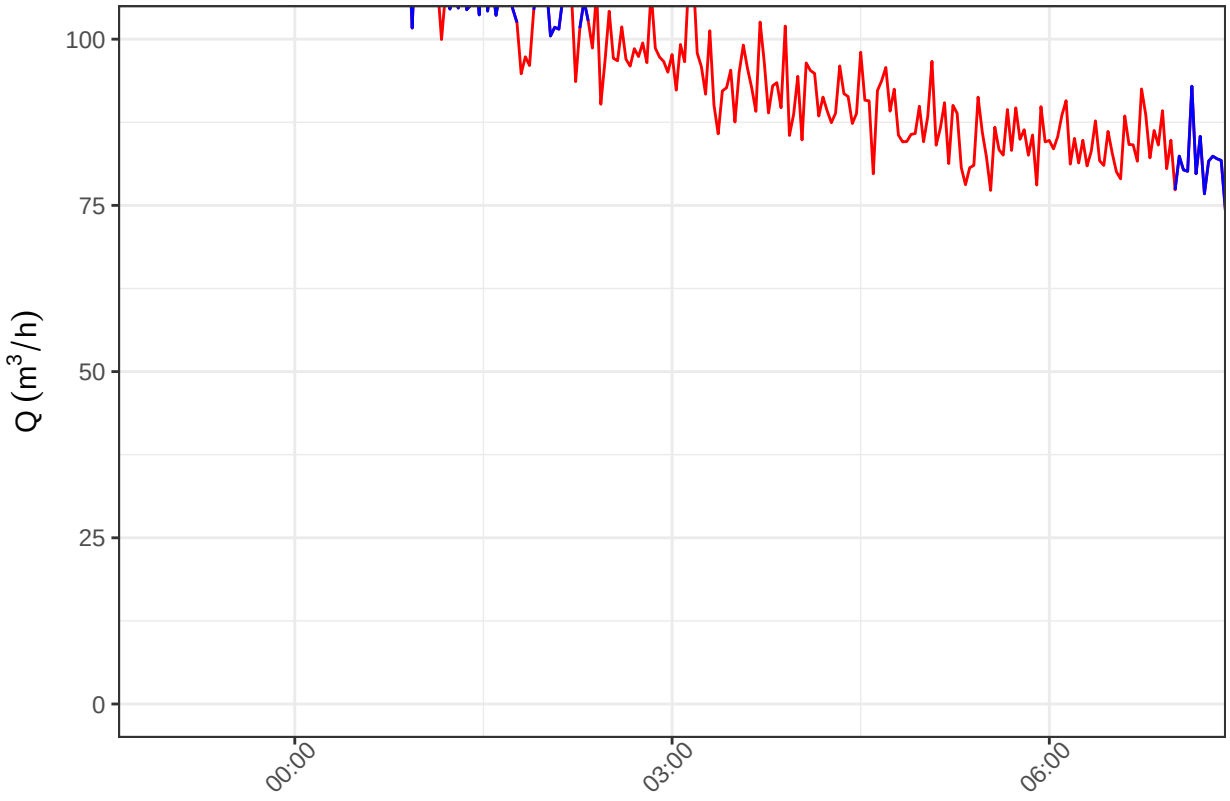
2. Example 2

Red geom_line has been ommitted infinal data.

```
altpTest = ggplot(dischargeAlteck, aes(x=Date)) +
  geom_line(aes(y= Q.m3Hrs), colour = "red") +
  geom_line(aes(y= Qna), colour = "blue") +
  coord_cartesian(xlim = c(as.POSIXct("2016-05-13 06:00:00 EST"),
                           as.POSIXct("2016-05-13 14:00:00 EST"))
                  ,
                  ylim = c(0, 100)
                  ) +
  theme_bw() +
  #scale_x_datetime(breaks = date_breaks("weeks"), labels = date_format("%d/%m")) +
  theme(axis.text.x=element_text(angle = 45, hjust = 0.75)) +
  #scale_y_continuous(trans=log_trans(), breaks=c(1,10,100,1000)) +
  xlab("") +
  ylab(expression(paste("Q ",({m}^"3"/h))))

altpTest
```

```
## Warning: Removed 9034 rows containing missing values (geom_path).
```




```
head(dischargeAlteck)
```

```
##                  Date        DateCheck Q.m3Hrs   Qna
## 1 2016-03-25 00:00:00 25/03/2016 00:00   1.256 1.256
## 2 2016-03-25 00:02:00 25/03/2016 00:02   1.219 1.219
## 3 2016-03-25 00:04:00 25/03/2016 00:04   1.192 1.192
```

```
## 4 2016-03-25 00:06:00 25/03/2016 00:06   1.212 1.212
## 5 2016-03-25 00:08:00 25/03/2016 00:08   1.195 1.195
## 6 2016-03-25 00:10:00 25/03/2016 00:10   1.219 1.219
```

Save files

```
head(dischargeAlteck)
```

```
##                  Date        DateCheck Q.m3Hrs   Qna
## 1 2016-03-25 00:00:00 25/03/2016 00:00   1.256 1.256
## 2 2016-03-25 00:02:00 25/03/2016 00:02   1.219 1.219
## 3 2016-03-25 00:04:00 25/03/2016 00:04   1.192 1.192
## 4 2016-03-25 00:06:00 25/03/2016 00:06   1.212 1.212
## 5 2016-03-25 00:08:00 25/03/2016 00:08   1.195 1.195
## 6 2016-03-25 00:10:00 25/03/2016 00:10   1.219 1.219
```

```
write.csv2(dischargeAlteck, "Data/hydroAlteck2016_NAs_R.csv", row.names = FALSE)
```