Merging Discharge & Sampler Data

PAZ

26 octobre 2016

Purpose

This document merges inputed (i.e. corrected) flowmeter data and automatic sampler data.

Used files:

1. **hydroAlteck2016_smooth_R.csv**
2. **prelev_20160713.csv**

Produced file:

1. **hydroAlteck2016_R.csv** (Used for plotting Sample and Discharge data together).

Required R-packages:

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

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/Alteckendorf2016/00_TransparencyFolder"
```

Import inputted discharge data

```
dischargeAlteck = read.csv2("Data/hydroAlteck2016_smooth_R.csv")
head(dischargeAlteck)
```

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

```
## 5 2016-03-25 00:12:00 25/03/2016 00:12   1.217 1.217   1.217   1.217
## 6 2016-03-25 00:14:00 25/03/2016 00:14   1.230 1.230   1.230   1.230
##      Q.HW1            Q.HW2
## 1 1.248600            1.182
## 2 1.237280 1.15424267729696
## 3 1.232224 1.17062590682503
## 4 1.224779 1.15615409458726
## 5 1.223623 1.17724053690379
## 6 1.222299 1.17698892559366
```

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

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

```
naDates = dischargeAlteck[is.na(dischargeAlteck$Date == TRUE),]

duplicateAlteck <- dischargeAlteck[duplicated(dischargeAlteck$DateCheck),]
duplicateAlteck
```

```
## [1] Date      DateCheck Q.m3Hrs   Qna       Qapprox   Qinterp   Q.HW1
## [8] Q.HW2
## <0 rows> (or 0-length row.names)
```

Import raw sampler data (March 25th to Jul 12th)

```
samplesAlteck = read.csv2("Data/prelev_20160713.csv", header = FALSE)

head(samplesAlteck)
```

```
##                 V1 V2
## 1 25/03/2016 12:04  1
## 2 26/03/2016 08:33  1
## 3 27/03/2016 06:04  1
## 4 28/03/2016 02:52  2
## 5 28/03/2016 22:37  2
## 6 30/03/2016 06:20  1
```

```
samplesAlteck = samplesAlteck[samplesAlteck$V2 != 0, ]
samplesAlteck$Date = as.POSIXct(strptime(samplesAlteck$V1,
                                         "%d/%m/%Y %H:%M",
                                         tz="EST"))
sum(is.na(samplesAlteck$V1))
```

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

```
samplesAlteck = samplesAlteck[,c(3,1:2)]
colnames(samplesAlteck) <- c("Date", "DateCheck", "sampleQ")

sum(is.na(samplesAlteck$Date))
```

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

```
samplesAlteck = samplesAlteck[order(samplesAlteck$Date),]

head(samplesAlteck)
```

```
##                  Date        DateCheck sampleQ
## 1 2016-03-25 12:04:00 25/03/2016 12:04       1
## 2 2016-03-26 08:33:00 26/03/2016 08:33       1
## 3 2016-03-27 06:04:00 27/03/2016 06:04       1
## 4 2016-03-28 02:52:00 28/03/2016 02:52       2
## 5 2016-03-28 22:37:00 28/03/2016 22:37       2
## 6 2016-03-30 06:20:00 30/03/2016 06:20       1
```

Merge the Discharge and the Samples' dataframes

To merge the two data.frames, we need to correcting minutes in the sample data, some of which took place during odd minutes.

1. Identify the odd minutes in a temporary data set to discard

```
discard = merge(dischargeAlteck, samplesAlteck, by = "Date", all = T)

# How many missing Discharge values resulting from the merge?
sum(is.na(discard$Date))
```

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

```
sum(is.na(discard$Q.m3Hrs))
```

```
## [1] 190
```

```
naQs = discard[is.na(discard$Q.m3Hrs == TRUE),]

naQs$Date = naQs$Date+60

naQs = naQs[,c("Date", "DateCheck.y")]

head(naQs)
```

```
##                    Date      DateCheck.y
## 976  2016-03-26 08:34:00 26/03/2016 08:33
## 2839 2016-03-28 22:38:00 28/03/2016 22:37
## 4359 2016-03-31 01:16:00 31/03/2016 01:15
## 4432 2016-03-31 03:40:00 31/03/2016 03:39
## 4451 2016-03-31 04:16:00 31/03/2016 04:15
## 4628 2016-03-31 10:08:00 31/03/2016 10:07
```

```r
head(dischargeAlteck)
```

```
##                  Date        DateCheck Q.m3Hrs   Qna Qapprox Qinterp
## 1 2016-03-25 00:04:00 25/03/2016 00:04   1.192 1.192   1.192   1.192
## 2 2016-03-25 00:06:00 25/03/2016 00:06   1.212 1.212   1.212   1.212
## 3 2016-03-25 00:08:00 25/03/2016 00:08   1.195 1.195   1.195   1.195
## 4 2016-03-25 00:10:00 25/03/2016 00:10   1.219 1.219   1.219   1.219
## 5 2016-03-25 00:12:00 25/03/2016 00:12   1.217 1.217   1.217   1.217
## 6 2016-03-25 00:14:00 25/03/2016 00:14   1.230 1.230   1.230   1.230
##      Q.HW1            Q.HW2
## 1 1.248600            1.182
## 2 1.237280 1.15424267729696
## 3 1.232224 1.17062590682503
## 4 1.224779 1.15615409458726
## 5 1.223623 1.17724053690379
## 6 1.222299 1.17698892559366
```

2. Add these odd-date markers to the flow-meter data (note that Date column remains as even minutes)

```r
# Merge new dates to discharge data
hydroAlteck2016 = merge(dischargeAlteck, naQs, by = c("Date"), all = T)
head(hydroAlteck2016)
```

```
##                  Date        DateCheck Q.m3Hrs   Qna Qapprox Qinterp
## 1 2016-03-25 00:04:00 25/03/2016 00:04   1.192 1.192   1.192   1.192
## 2 2016-03-25 00:06:00 25/03/2016 00:06   1.212 1.212   1.212   1.212
## 3 2016-03-25 00:08:00 25/03/2016 00:08   1.195 1.195   1.195   1.195
## 4 2016-03-25 00:10:00 25/03/2016 00:10   1.219 1.219   1.219   1.219
## 5 2016-03-25 00:12:00 25/03/2016 00:12   1.217 1.217   1.217   1.217
## 6 2016-03-25 00:14:00 25/03/2016 00:14   1.230 1.230   1.230   1.230
##      Q.HW1            Q.HW2 DateCheck.y
## 1 1.248600            1.182        <NA>
## 2 1.237280 1.15424267729696        <NA>
## 3 1.232224 1.17062590682503        <NA>
## 4 1.224779 1.15615409458726        <NA>
## 5 1.223623 1.17724053690379        <NA>
## 6 1.222299 1.17698892559366        <NA>
```

```r
# Check number of odd-minute dates, should be 0:
sum(is.na(hydroAlteck2016$Q.m3Hrs))
```

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

```r
# Fill in the rest  of the Target dates (even)
hydroAlteck2016$DateCheck.S <- ifelse(is.na(hydroAlteck2016$DateCheck.y),
                                      as.character(hydroAlteck2016$DateCheck),
                                      as.character(hydroAlteck2016$DateCheck.y))
hydroAlteck2016$DateCheck.y <- NULL

# Create common column name in samples' target column (i.e. DateCheck.S)
samplesAlteck <- samplesAlteck[, c("DateCheck", "sampleQ")]
colnames(samplesAlteck) <- c("DateCheck.S", "sampleQ")
head(samplesAlteck)
```

```
##        DateCheck.S sampleQ
## 1 25/03/2016 12:04       1
## 2 26/03/2016 08:33       1
## 3 27/03/2016 06:04       1
## 4 28/03/2016 02:52       2
## 5 28/03/2016 22:37       2
## 6 30/03/2016 06:20       1
```

3. Merging the two tables

```
hydroAlteck2016 = merge(hydroAlteck2016, samplesAlteck, by = c("DateCheck.S"), all = T)

# Checks
sum(is.na(hydroAlteck2016$Date))
```

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

```
anyDuplicated(hydroAlteck2016$Date)
```

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

```
sum(is.na(hydroAlteck2016$Q.m3Hrs))
```

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

```
head(hydroAlteck2016)
```

```
##        DateCheck.S                Date        DateCheck Q.m3Hrs   Qna
## 1 01/04/2016 00:00 2016-04-01 00:00:00 01/04/2016 00:00   17.12 17.12
## 2 01/04/2016 00:02 2016-04-01 00:02:00 01/04/2016 00:02   14.71 14.71
## 3 01/04/2016 00:04 2016-04-01 00:04:00 01/04/2016 00:04   13.82 13.82
## 4 01/04/2016 00:06 2016-04-01 00:06:00 01/04/2016 00:06   14.58 14.58
## 5 01/04/2016 00:08 2016-04-01 00:08:00 01/04/2016 00:08   13.62 13.62
## 6 01/04/2016 00:10 2016-04-01 00:10:00 01/04/2016 00:10   14.48 14.48
##   Qapprox Qinterp    Q.HW1            Q.HW2 sampleQ
## 1   17.12   17.12 14.63129 16.0163558897583      NA
## 2   14.71   14.71 15.12903 16.9994185259239      NA
## 3   13.82   13.82 15.04522 14.8463828626439      NA
## 4   14.58   14.58 14.80018  13.860730304389      NA
## 5   13.62   13.62 14.75614 14.4885280729906      NA
## 6   14.48   14.48 14.52891 13.6487756027902      NA
```

```
class(hydroAlteck2016$Date)
```

```
## [1] "POSIXct" "POSIXt"
```

```
# Order by date
hydroAlteck2016 = hydroAlteck2016[order(hydroAlteck2016$Date),]
```

Create a "Type"" column to point to Sampling times during plotting

```r
hydroAlteck2016$Type = ifelse(is.na(hydroAlteck2016$sampleQ), "Discharge", "Sample")

head(hydroAlteck2016)
```

```
##            DateCheck.S                Date        DateCheck Q.m3Hrs   Qna
## 60042 25/03/2016 00:04 2016-03-25 00:04:00 25/03/2016 00:04   1.192 1.192
## 60043 25/03/2016 00:06 2016-03-25 00:06:00 25/03/2016 00:06   1.212 1.212
## 60044 25/03/2016 00:08 2016-03-25 00:08:00 25/03/2016 00:08   1.195 1.195
## 60045 25/03/2016 00:10 2016-03-25 00:10:00 25/03/2016 00:10   1.219 1.219
## 60046 25/03/2016 00:12 2016-03-25 00:12:00 25/03/2016 00:12   1.217 1.217
## 60047 25/03/2016 00:14 2016-03-25 00:14:00 25/03/2016 00:14   1.230 1.230
##       Qapprox Qinterp    Q.HW1            Q.HW2 sampleQ      Type
## 60042   1.192   1.192 1.248600            1.182      NA Discharge
## 60043   1.212   1.212 1.237280 1.15424267729696      NA Discharge
## 60044   1.195   1.195 1.232224 1.17062590682503      NA Discharge
## 60045   1.219   1.219 1.224779 1.15615409458726      NA Discharge
## 60046   1.217   1.217 1.223623 1.17724053690379      NA Discharge
## 60047   1.230   1.230 1.222299 1.17698892559366      NA Discharge
```

Saving

```r
write.csv2(hydroAlteck2016, "Data/hydroAlteck2016_R.csv", row.names = F)
```