Course Notes Advanced SWAT: Checking the input water balance

Willem Vervoort

27-09-2017

Introduction

This is an introduction into developing SWAT CUP inout files using R, written for the "How do I use satellite and global reanalysis data for hydrological simulations in SWAT?" workshop in Montevideo between 7 - 11 August 2017, jointly organised by the University of Sydney, IRI (the University of Columbia) and INIA, Uruguay.

Checking the input data water balance

An important and simple check as part of the SWAT calibration is to check the input data water balance. This is slightly tricky as we have no real estimate of actual ET, we can only estimate potential ET using either Hargreaves, or Penman Monteith or Priestley Taylor. As a result we are probably overestimating the ET component and would expect the water balance to be mostly negative.

In order to calculate the water balance, we need to derive the potential ET from the original weather station data. To simplify this, I have written a helper function that given a dataframe with Minimum and Maximum Temperature, Humidity, Solar Radiation and Windspeed will calculate the potential ET using three different ET functions (Penman Monteith, Priestley Taylor and Hargreaves). Each of these functions has their own limitations and you are encouraged to check the literature about this.

Reading in the data

The first step is to read in the data from the weather station and rainfall stations and the flow data. I will once again use the package `zoo` to merge all the data and make them into timeseries.

```
library(tidyverse)
```

```
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
library(zoo)
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
# Example for Santa Lucia
# Read in the flow data in cumecs
Flowdata <- readRDS("data/SantaLucia.RDS")
Flow <- zoo(Flowdata$flow/(5171*10^6)*86400*1000,
            order.by = as.Date(Flowdata$Date))

# read in the weather data:
WD <- read.csv("data/weather/INIALasBrujas_1983-2016.csv")

# There are two different dates: 31-12-2016 and 1/1/1983
# Need to convert to standard dates
Dates <- ifelse(grepl("-",WD$Fecha)==TRUE,
                as.character(as.Date(WD$Fecha,"%d-%m-%Y")),
                as.character(as.Date(WD$Fecha,"%d/%m/%Y")))

# Add additional rainfall data
# read in the precipitation data
```

```r
Pdata <- read.csv("data/weather/Precipitacion_SantaLucia.csv",
                  na.strings="NaN")
# create dates from the dates in the file
Dates2 <- as.Date(paste(Pdata[,1],Pdata[,2],
                        Pdata[,3],sep="-"))

# find the stations that have 90% of data after 1995
Pdata_1995 <- Pdata[Dates2 >= "1995-01-01",]
result <- rep(0,(ncol(Pdata_1995)-3))
for (i in 4:ncol(Pdata_1995)) {
  result[i-3] <- sum(ifelse(is.na(Pdata_1995[,i]),1,0))/nrow(Pdata_1995)
}

# result indicates the fraction of NA data for the stations
# throw out all the columns and rows where result >0.1
Pdata_new <- Pdata_1995[,-(which(result>0.1)+3)]
Dates1995 <- Dates2[Dates2 >= "1995-01-01"]


# calculate the average across the stations
Rainfall_avg <- apply(Pdata_new[,4:ncol(Pdata_new)],1,mean,na.rm=T)
# zoo data
Rainfall_z <- zoo(Rainfall_avg, order.by = Dates1995, frequency=1)

# wind data needs to be in m/s
wind_z <- zoo(WD[,5]*1000/86400, order.by =Dates)

# humidity
hum_z <- zoo(WD[,6], order.by =Dates)

# maxT
maxT_z <- zoo(WD[,3], order.by =Dates)

# minT
minT_z <- zoo(WD[,4], order.by =Dates)

# solar radiation needs to be in MJ/m2
# 1 cal cm-2 day-1 = 4.1868 10-2 MJ m-2 day-1
# http://www.fao.org/docrep/X0490E/x0490e0i.htm
slr_z <- zoo(WD[,2]*0.041868, order.by =Dates)
```

Note that in the above we have calculated the average rainfall across the catchment from the station data with less than 10% missing data past 1995. So for rainfall we only have data post 1995.

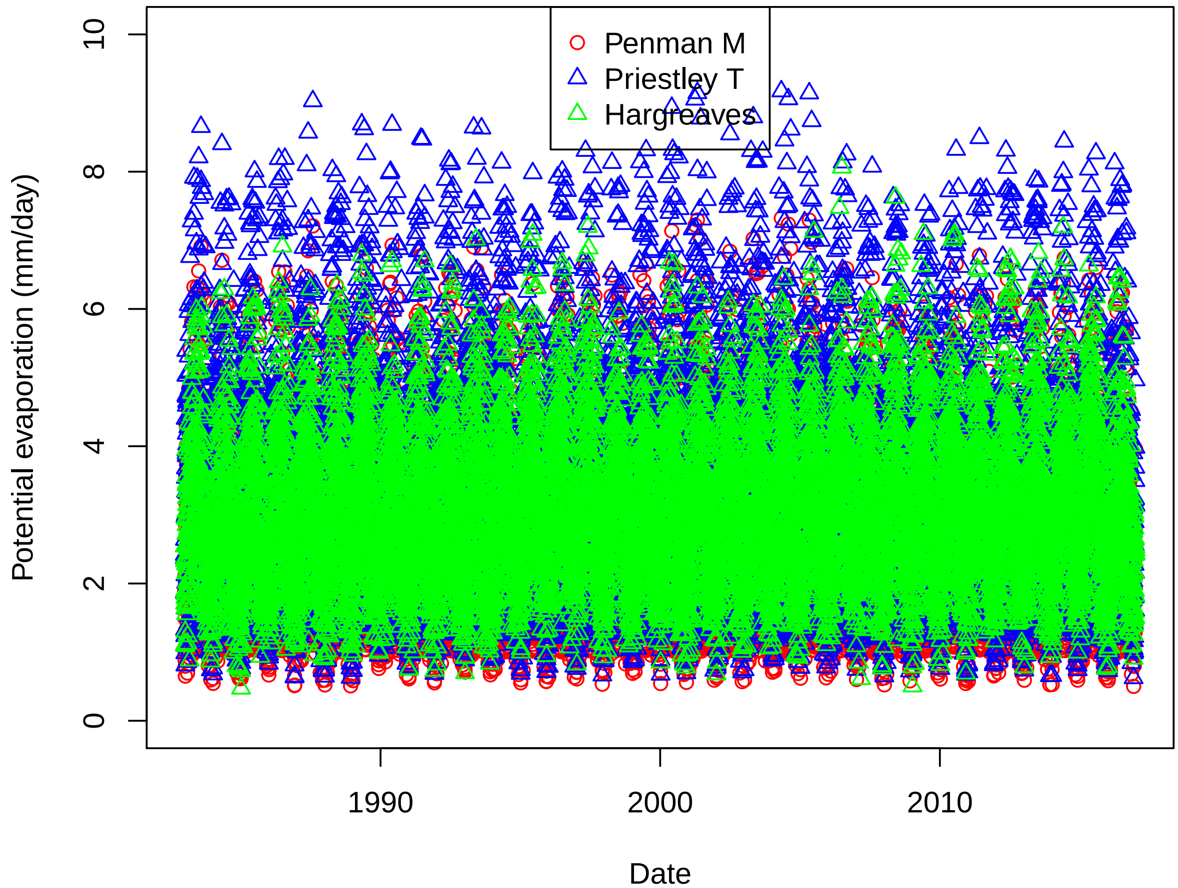
Calculating potential ET

As indicated, for this we need to load the `ETfun.R` script that has the potential ET calculations for the three different methods. Please check the function for the detail

```r
source("functions/ETfun.R")
```

Now we need to reformat the data to conform the input requirements of the ETfun. This function can take both minimum and maximum relative humidity data, but in this case we only have the average.

```r
# Create a data.frame to feed into ET.fun for potET
# only one relative humidity, no max and min
ETinput <- merge(maxT_z, minT_z, hum_z, hum_z, slr_z, wind_z)


# Calculate all the ET output using ET.fun
ETout <- ET.fun(ETinput,elev=30, lat=34)
```




```r
names(ETout)
```

```
## [1] "MaxT"      "MinT"      "MaxRelH"   "MinRelH"   "Solar.Rad" "Windspeed"
## [7] "PM.Ep"     "PT.Ep"     "HG.Ep"
```

```r
ETout_z <- zoo(ETout[,7:9], order.by=as.Date(time(ETout)))
```

Water balance calculation

Once we have calculated potential ET we construct a series of water balance calculations, all slightly different due to the different potential ET calculations. I am focussing here on the annual water balance, but the calculation can be easily repeated for the monthly water balance

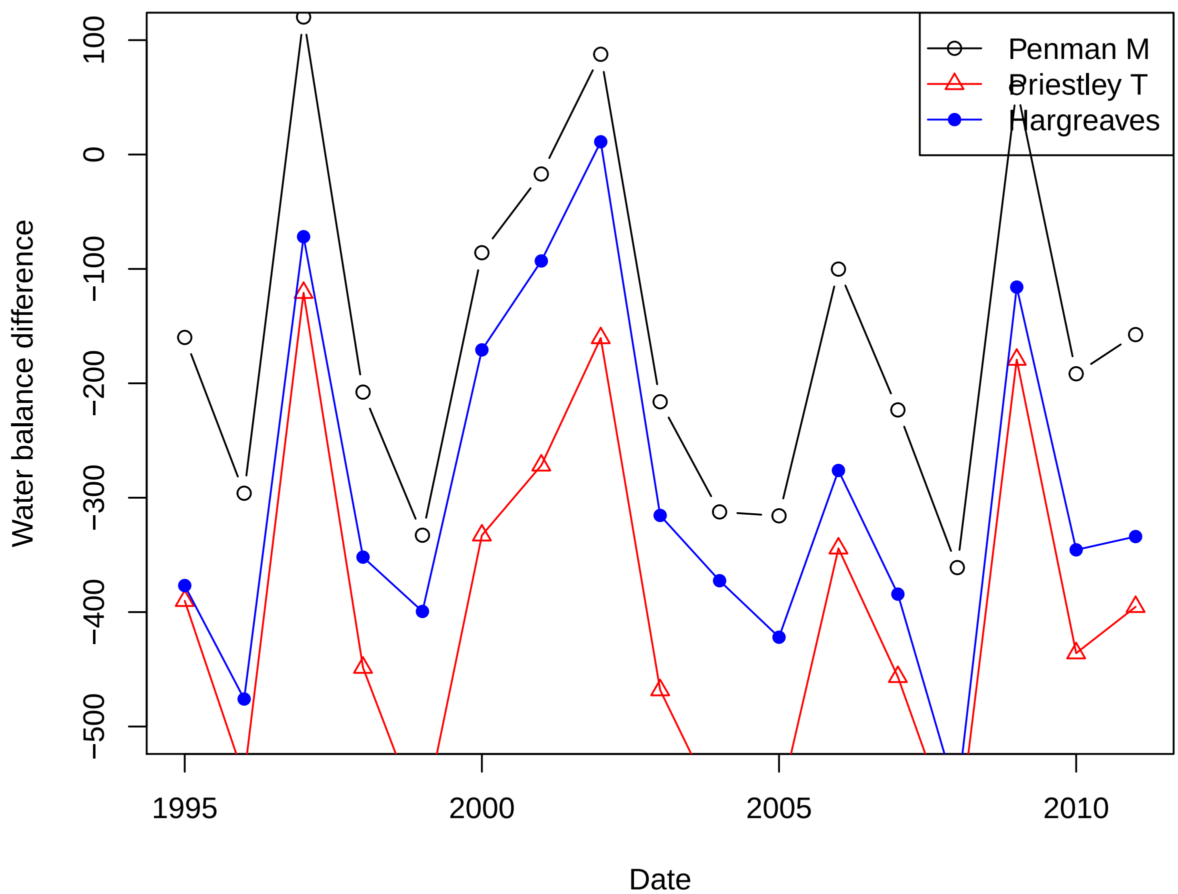
```r
# first combine data into one dataframe (start with Penman M ET)
WBdata_PM <- merge(Flow, ETout_z$PM.Ep,Rainfall_z, all=F)
names(WBdata_PM) <- c("Flow", "PMEp", "Rain")
# Priestley Taylor
WBdata_PT <- merge(Flow, ETout_z$PT.Ep,Rainfall_z, all=F)
names(WBdata_PT) <- c("Flow", "PTEp", "Rain")
# Hargreaves
WBdata_HG <- merge(Flow, ETout_z$HG.Ep,Rainfall_z, all=F)
names(WBdata_HG) <- c("Flow", "HGEp", "Rain")

# summarise to annual data
WBdata_annual_PM <- aggregate(WBdata_PM,
                          list(year=format(time(WBdata_PM),"%Y")),
                          sum,na.rm=T)
# show the result
head(WBdata_annual_PM)
```

```
##           Flow      PMEp      Rain
## 1995 221.36616  932.1177  993.6029
## 1996 109.67137 1003.4686  817.0571
## 1997  63.66472  972.4794 1156.3571
## 1998 251.62664  973.6971 1017.6324
## 1999 566.44022 1020.2451 1253.9000
## 2000 504.85482  997.6984 1416.7115
```

```r
WBdata_annual_PT <- aggregate(WBdata_PT,
                          list(year=format(time(WBdata_PT),"%Y")),
                          sum,na.rm=T)
WBdata_annual_HG <- aggregate(WBdata_HG,
                          list(year=format(time(WBdata_HG),"%Y")),
                          sum,na.rm=T)
```

We can now plot the waterbalance resultant (essentially the change in storage): $Rain - ET - Flow = \delta S$

```r
plot(time(WBdata_annual_PM),
     with(WBdata_annual_PM,Rain - PMEp - Flow), type="b",
     xlab = "Date", ylab = "Water balance difference",
     ylim=c(-500,100))
lines(time(WBdata_annual_PT),
     with(WBdata_annual_PT,Rain - PTEp - Flow), col="red")
points(time(WBdata_annual_PT),
     with(WBdata_annual_PT,Rain - PTEp - Flow), col="red",pch=2)
lines(time(WBdata_annual_HG),
     with(WBdata_annual_HG,Rain - HGEp - Flow), col="blue")
points(time(WBdata_annual_HG),
     with(WBdata_annual_HG,Rain - HGEp - Flow), col="blue",pch=16)
legend("topright",c("Penman M", "Priestley T", "Hargreaves"),
       lty = 1, pch = c(1,2,16), col=c(1,"red","blue"))
```



From the result we can see that all water balances are predominantly negative, which can be explained by using potential ET rather than actual ET. We can also see that the uisng the Priestley Taylor potential ET results in the largest negative water balances, so we can probably ignore this and not use this in SWAT.

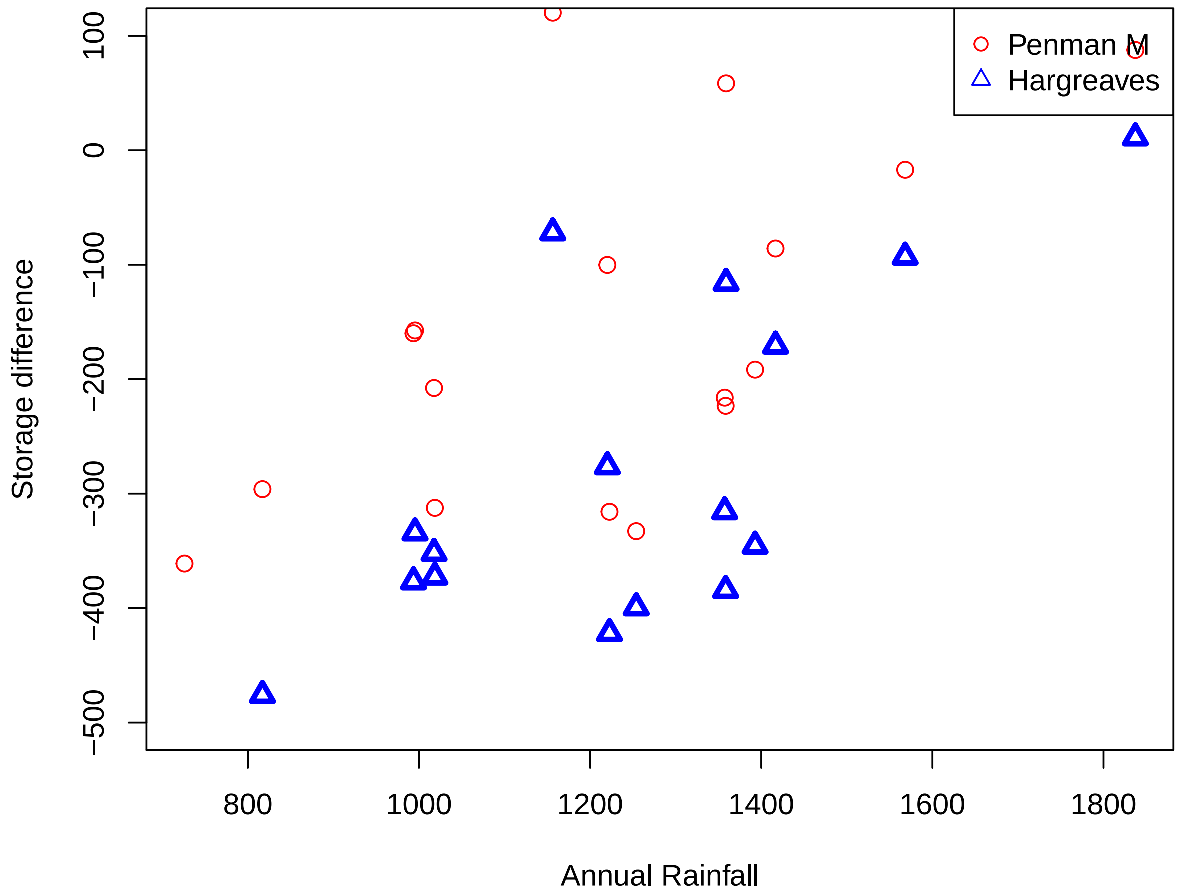
plots of rainfall and storage and runoff

Another plot that might help understand the differences is a plot of annual rainfall against the water balance difference, and a plot of annual rainfall versus annual flow (just to check whether there is no strong non-linearity in the relationship)

Rainfall and storage

```
plot(WBdata_annual_PM$Rain,
     with(WBdata_annual_PM,Rain - PMEp - Flow),
     pch=1, col= "red", lwd=3, cex=1.2,
     xlab="Annual Rainfall", ylab="Storage difference",
     ylim=c(-500,100))
points(WBdata_annual_HG$Rain,
     with(WBdata_annual_HG,Rain - HGEp - Flow), pch=2, col="blue",
     lwd=3, cex=1.2)
```

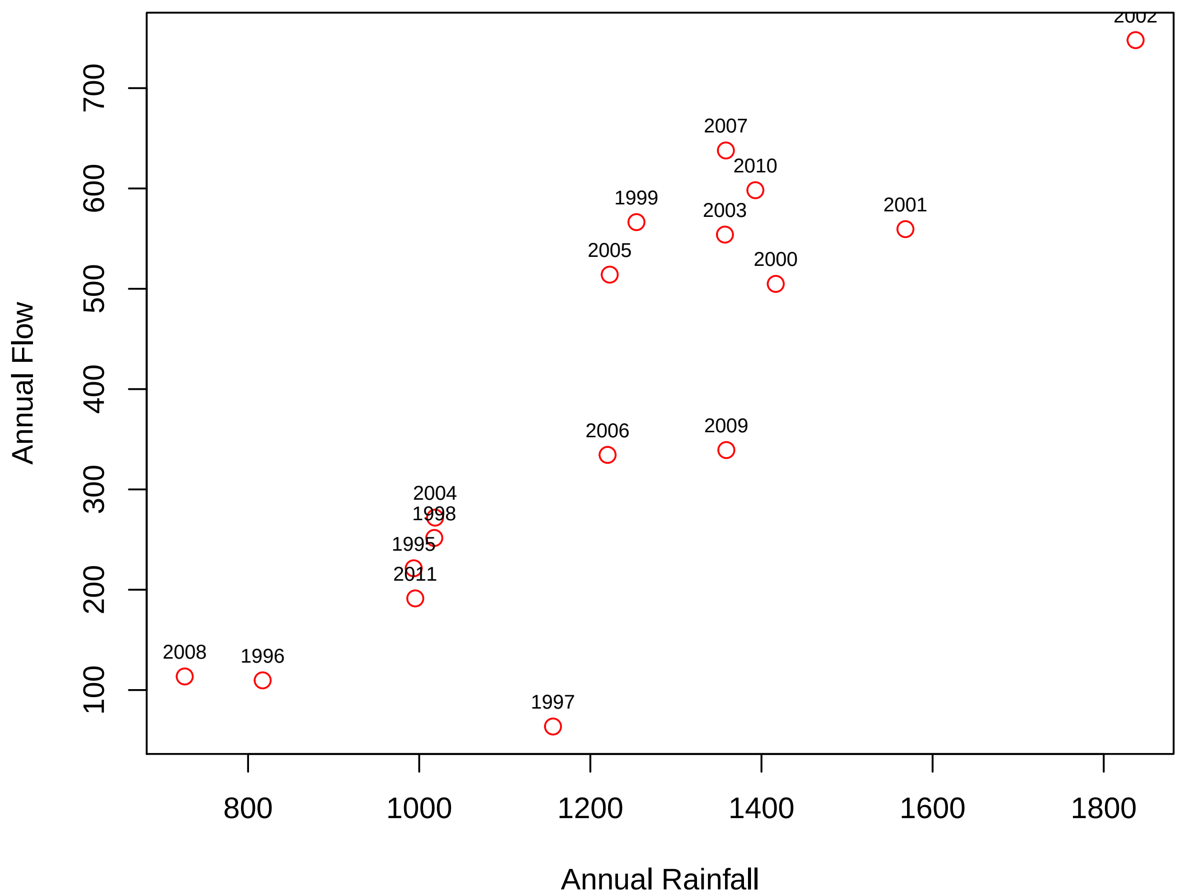
```
legend("topright",c("Penman M",  "Hargreaves"),
        pch = c(1,2), col=c("red","blue"))
```



This indicates a positive relationship between storage and annual rainfall, which is what you would expect, drier years would have more negative storage balances, while wetter years would have more positive storage balances.

Rainfall and runoff

```
plot(WBdata_annual_PM$Rain,
     WBdata_annual_PM$Flow,
     pch=1, col= "red", lwd=3, cex=1.2,
     xlab="Annual Rainfall", ylab="Annual Flow")
text(WBdata_annual_PM$Rain, WBdata_annual_PM$Flow,
     labels=time(WBdata_annual_PM),
     cex= 0.7, pos=3)
```



This shows the expected postive relationship between rainfall and streamflow, but with some years deviating from the general relationships, these might be worth investigating.