

Data preparation

Willem Vervoort, Michaela Dolk & Floris van Ogtrop

2017-01-29

```
# root dir
knitr::opts_knit$set(root.dir = "C:/Users/rver4657/ownCloud/Virtual Experiments/VirtExp")
```

This rmarkdown document and the resulting pdf are stored on github. All directories (apart from the root working directory) refer to the directories in this repository

Introduction

This document is related to the manuscript “Disentangling climate change trends in Australian streamflow” (vervoort et al.), submitted to Journal of Hydrology. This document outlines the preparation of the original data into the dataframes that have been analysed in the project. The decision making on how stations were identified, is outlined in the methods of the submitted manuscript. This document is aimed at documenting the code of the analysis.

sources of data

As outlined in the manuscript, the original data were sourced from the following locations: Streamflow is from the Bureau of Meteorology (BOM) hydrological reference stations <http://www.bom.gov.au/water/hrs/> Rainfall and temperature were obtained from the BOM gridded data and the BOM station data <http://www.bom.gov.au/climate/data-services/>

Reading in the data

The data consists of comma delimited (csv) files, as downloaded from the websites. All data cover the period 1970 - 2010. The following flow stations were used:

Table 1: Stations used in this project (continued below)

Name.used.in.this.project	Number	Region	Catchment.area.smaller.250km2.
COTT	410730	ACT	130
RUTH	219001	NSW	14
CORA	215004	NSW	166
ELIZ	G8150018	NT	96
COCH	113004A	QLD	95
COEN	922101B	QLD	170
SCOT	A5030502	SA	29
HELL	312061	TAS	101
NIVE	304497	TAS	174
MURR	405205	VIC	106
SOUT	225020A	VIC	10
YARR	614044	WA	80
DOMB	607155	WA	116

Latitude	Longitude	Rain.St	HQTmax.st
-35.59	148.8	70316	70351
-36.59	149.4	69003	70351
-35.15	150	69049	68072
-12.61	131.1	14149	14015
-17.74	145.6	31083	34084
-13.96	143.2	27005	27045
-35.1	138.7	23734	23373
-41.42	145.7	96023	96003
-42.03	146.4	91065	96003
-37.41	145.6	88028	85072
-37.83	146.4	85238	85072
-32.81	116.2	9538	9021
-34.58	116	9590	9518

define decades to analyze

```
study_period_decades <- c("70_80", "80_90", "90_00", "00_10")
decade_start <- c(as.Date("1/1/1970", format="%d/%m/%Y"),
  as.Date("1/1/1980", format="%d/%m/%Y"),
  as.Date("1/1/1990", format="%d/%m/%Y"),
  as.Date("1/1/2000", format="%d/%m/%Y"))
decade_end <- c(as.Date("31/12/1979", format="%d/%m/%Y"),
  as.Date("31/12/1989", format="%d/%m/%Y"),
  as.Date("31/12/1999", format="%d/%m/%Y"),
  as.Date("31/12/2010", format="%d/%m/%Y"))

# define the overall period
start_date <- as.Date("1970-01-01")
end_date <- as.Date("2010-12-31")
```

read in the daily stream flow data

This includes conversion from ML/day (as indicated on the source website) to mm to match the rainfall data and to use in models. This means that the data need to be scaled to the catchment size:

- convert ML/day to mm
- 1 ML = 10^6 L = 10^6 dm³ is 10^9 cm³ is 10^{12} mm³
- 1 km² = 10^6 m² = 10^{12} cm² = 10^{14} mm²
- ML/day to mm → flow/area(km²)/100 = mm

```
# read in the flow data and convert to zoo
for (i in seq_along(Stations[,1])) {
  temp <- read.csv(paste("data/Original streamflow data/", Stations[i,2],
    "_daily_ts2.csv", sep=""))
  year <- substr(as.character(temp$Date), nchar(as.character(temp$Date))-1,
    nchar(as.character(temp$Date)))
  Dates <- as.Date(paste(substr(as.character(temp$Date), 1,
    nchar(as.character(temp$Date))-2),
    ifelse(as.numeric(year)>=50,paste("19",year,sep=""),
  assign(paste(Stations[i,1], "_daily_flow", sep=""),
    zoo(temp$Q/(Stations[i,4]),order.by=Dates))
```

```

}
#####

# use zoo to merge all catchments to use same time interval
flow_zoo<-merge(COTT_daily_flow, RUTH_daily_flow, CORA_daily_flow,
               ELIZ_daily_flow, COCH_daily_flow, COEN_daily_flow,
               SCOT_daily_flow, HELL_daily_flow, NIVE_daily_flow,
               MURR_daily_flow, SOUT_daily_flow, YARR_daily_flow,
               DOMB_daily_flow)

# limit to 1970 - 2010
flow_zoo <- window(flow_zoo, start=start_date, end=end_date)

# Also create a dataframe for flow
flow_data_70_10<-data.frame(Date=time(flow_zoo), coredata(flow_zoo))
colnames(flow_data_70_10)[2:14] <- Stations[,1]
#####

```