

# Homogenization of ECA&D daily series with Climatol

Jose A. Guijarro

*State Meteorological Agency (AEMET), Balearic Islands Office, Spain*

Version 1.2 (July 2020)



This guide and the referred software have been prepared and released as part of the Work Package 3 of the INDECIS project (see deliverable 3.3), which is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR), with co-funding by the European Union's Horizon 2020 research and innovation programme.



This guide is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

# Contents

- 1. Introduction** **3**
  
- 2. Prerequisites** **3**
  
- 3. Methodology** **4**
  - 3.1. Start R and read the homogenization functions . . . . . 4
  - 3.2. Preparation of the ECA&D input files . . . . . 4
  - 3.3. Preparation of the reanalysis reference series . . . . . 5
  - 3.4. Homogenization of the ECA&D daily series . . . . . 7
  
- 4. Acknowledgments** **8**

# 1. Introduction

The Work Package 3 (WP3) of the INDECIS project is due to provide quality controlled and homogenized daily series of the Essential Climate Variables (ECV) stored at the European Climate Assessment & Dataset (ECA&D) to be used by the rest of the teams.

This guide explains how to proceed to prepare input files suitable for the homogenization functions of Climatol from the ECAD files and homogenize them obtaining the resulting series in CSV files.

The ECV processed in this guide are the following (in units as stored at ECA&D):

- CC : Cloud Cover (oktas)
- FG : Wind Speed ( $0.1 \text{ m s}^{-1}$ )
- HU : Relative Humidity (1 %)
- PP : Sea Level Pressure (0.1 hPa)
- RR : Precipitation Amount (0.1 mm)
- SS : Sunshine (0.1 hours)
- TN : Minimum Temperature (0.1 °C)
- TX : Maximum Temperature (0.1 °C)

## 2. Prerequisites

All the needed homogenization functions, written in R, have been gathered in a source file named [homogenECAD.R](#), which consists in a subset of the functions of the Climatol package (candidate release 4.0) plus additional functions to convert the input series from the ECA&D to the Climatol format, add reanalysis series to them, and extract the homogenized series from the Climatol results.

The whole procedure runs in a personal computer under a Linux operating system, with the bash interpreter and a modern version of the R environment installed. Also advisable is to have the CDO (Climate Data Operators) installed, needed to calculate the reanalysis wind speed from its U, V components.

The R working directory must contain all files and data directories involved in the process:

**[homogenECAD.R](#)** : The code containing the needed functions.

**[homogenECAD-VV.Rsh](#)** : R scripts to homogenize the different variables, where **VV** stands for CC, FG, HU, PP, RR, SS, TN and TX.

**[stations](#)** : Directory containing the ECA&D stations metadata files, one per variable, with names [CC\\_stations.txt](#), [FG\\_stations.txt](#), etc.

**CC, FG, HU, PP, RR, SS, TN, TX** : Directories containing the ECA&D series in CSV format and extension `.txt`. No other file with that extension must be present in those directories.

**20CR** : Directory containing 20CR reanalysis annual files of the needed surface variables (`dswrf.sfc`, `prate`, `pres.sfc`, `rhum.sig995`, `tcdc.eatm`, `tmax.2m`, `tmin.2m`, `uwnd.10m` and `vwnd.10m`).

### 3. Methodology

The operational procedure consists in these steps:

1. Start R and read the homogenization functions.
2. Convert the quality controlled series from ECA&D format to the Climatol format.
3. Prepare reanalysis reference series in Climatol format.
4. Apply the homogenization function to the series.

#### 3.1. Start R and read the homogenization functions

Go to the `homogenECAD` directory and issue the commands (anything after `#` is just a comment):

```
R #start R
wdir='~/homogenECAD' #set here your working directory (can be another name)
setwd(wdir) #change to the chosen working directory
source('homogenECAD.R') #read the homogenization and auxiliary functions
#also set these parameters for the rest of the process:
yeari=1950 #initial year to homogenize
yearf=2019 #final year to homogenize
yrean=2014 #final year of the available reanalysis
```

#### 3.2. Preparation of the ECA&D input files

In the first place we will gather all ECA&D station metadata in a single file with Climatol format. All the ECA&D station metadata files must be placed in the 'stations' directory hanging from our working directory 'homogenECAD', with names `CC_stations.txt`, `FG_stations.txt`, etc (one file per every variable to homogenize). Then we issue the command:

```
ecadstations()
```

Then we must prepare the input files for every variable. All ECA&D data files for a single variable VV must be located in a directory named VV, where VV stands for the two letter acronym of the variables (CC, FG, etc). This is accomplished with the commands:

```
minyr=10 #set the minimum no. of years with data in the series
for(VV in c('CC','FG','HU','PP','RR','SS','TN','TX'))
  ecad2climatol(VV,yeari,yearf,minyr) #create the files for Climatol
```

### 3.3. Preparation of the reanalysis reference series

[ftp://ftp.cdc.noaa.gov/Datasets/20thC\\_ReanV2c/Dailies](ftp://ftp.cdc.noaa.gov/Datasets/20thC_ReanV2c/Dailies) was the server from where the 20CR reanalysis daily series were downloaded as yearly NetCDF files, which were placed in the directory [20CR](#). All available years from the first of the period to homogenize must be present in this directory. As an example, this is the list for the year 2014:

- `dswrf.sfc.2014.nc` : Downward solar radiation flux
- `prate.2014.nc` : Precipitation rate
- `pres.sfc.2014.nc` : Pressure at surface
- `rhum.sig995.2014.nc` : Relative Humidity (sigma level 0.995)
- `tcdc.eatm.2014.nc` : Total cloud cover
- `tmax.2m.2014.nc` : Maximum temperature at 2 m above the ground
- `tmin.2m.2014.nc` : Minimum temperature at 2 m above the ground
- `uwnd.10m.2014.nc` : U wind component at 10 m above the ground
- `vwnd.10m.2014.nc` : V wind component at 10 m above the ground

As the wind is available as components U and V, we must calculate the module. One of the simplest ways is with the help of the CDO (Climate Data Operators). Once installed in your system, you can call them from R with:

```
for(year in yeari:yrean) system(sprintf('cdo cname,uwnd,wspd -sqrt -add
  -sqr -selname,uwnd uwnd.10m.%d.nc -sqr -selname,vwnd vwnd.10m.%d.nc
  wspd.10m.%d.nc', year, year, year))
```

Now we have to add to the Climatol files the closest reanalysis series to any ECA&D station with the help of the function `rean2climatol`:

```

setwd(sprintf('%s/CC',wdir)) #----- CC (octas):
fp='../20CR/tcdc.eatm.%d.nc' #file pattern
rean2climatol('CC',yeari,yearf,'tcdc',fp,ndec=0,b=8/100)
setwd(sprintf('%s/FG',wdir)) #----- FG (0.1 m/s):
fp='../20CR/wspd.10m.%d.nc' #file pattern
rean2climatol('FG',yeari,yearf,'wspd',fp,ndec=0,b=10)
setwd(sprintf('%s/HU',wdir)) #----- HU (%):
fp='../20CR/rhum.sig995.%d.nc' #file pattern
rean2climatol('HU',yeari,yearf,'rhum',fp)
setwd(sprintf('%s/PP',wdir)) #----- PP (0.1 hPa):
fp='../20CR/pres.sfc.%d.nc' #file pattern
rean2climatol('PP',yeari,yearf,'pres',fp,ndec=0,b=.1)
setwd(sprintf('%s/RR',wdir)) #----- RR (0.1 mm):
fp='../20CR/prate.%d.nc' #file pattern
rean2climatol('RR',yeari,yearf,'prate',fp,ndec=0,b=24*3600*10)
setwd(sprintf('%s/TN',wdir)) #----- TN (0.1 °C):
fp='../20CR/tmin.2m.%d.nc' #file pattern
rean2climatol('TN',yeari,yearf,'tmin',fp,ndec=0,a=-273,b=10)
setwd(sprintf('%s/TX',wdir)) #----- TX (0.1 °C):
fp='../20CR/tmax.2m.%d.nc' #file pattern
rean2climatol('TX',yeari,yearf,'tmax',fp,ndec=0,a=-273,b=10)
setwd(sprintf('%s/SS',wdir)) #----- SS (0.1 hours):
fp='../20CR/tcdc.eatm.%d.nc' #file pattern
rean2climatol('SS',,yeari,yearf,'tcdc',fp,ndec=0)
#But cloudiness must be converted to sunshine hours:
leerdat('SS.r',yeari,yearf)
fech=seq(as.Date(sprintf('%s-01-01',yeari)),as.Date(sprintf('%s-12-31',
yearf)),1)
r1=min(which(substr(est.c[,4],1,1)=='*')) #first reanalysis series
for(k in r1:ne) { #convert cloudiness to sunshine duration (in 0.1 hours):
  it=insolteor(est.c[k,2],fech)
  dat[,k]=.9*(100-dat[,k])*it/10
}
write(dat,'SS.r_1901-2018.dat',ncolumns=10)
setwd(wdir) #return to the working directory

```

However, as it will be more efficient to keep the reference series in separate files, reanalysis series will be detached from the ECA&D series and saved apart:

```

nd2 <- length(seq(as.Date(sprintf('%d-01-01',yeari)),
as.Date(sprintf('%d-12-31',yean)),1))
for(VV in c('CC','FG','HU','PP','RR','SS','TN','TX')) {
  cat(' ',VV)
  setwd(sprintf('%s/%s',wdir,VV))
  leerdat(sprintf('%s.r',VV),yeari,yearf)
  r1=min(which(substr(est.c[,4],1,1)=='*')) #first reanalysis series
  write(dat[1:nd2,r1:ne],sprintf('%s20CR_%d-%d.dat',VV,yeari,yean))
}

```

```

write.table(est.c[r1:ne,], sprintf('%s20CR_%d-%d.est', VV, yeari, yrean),
            row.names=F, col.names=F)
}
cat('\n')

```

### 3.4. Homogenization of the ECA&D daily series

The ECA&D daily series can now be homogenized, one at a time, using the closest reanalysis series as reference, with the help of R scripts prepared for every variable. These scripts can be run one by one, or all at the same time, taking advantage of modern multi-core processors, with the bash (after quitting the R session) commands:

```

#redefine parameters for bash:
wdir=~ /homogenECAD #or your designated working directory
yeari=1950 #initial year to homogenize
yearf=2019 #final year to homogenize
yrean=2014 #final year of the available reanalysis
cd $wdir
#now launch the homogenization in unattended background:
for VV in CC FG HU PP RR SS TN TX; do
    nohup ./homogenECAD-$VV.Rsh $yeari $yearf $yrean &
done

```

After some 5 days of processing time (less if using a more powerful server), results can be found in each VV directory in four text files with names:

- VV\_1950-2018\_out.csv : List of rejected outliers.
- VV\_1950-2018\_brk.csv : List of corrected breaks.
- VV\_1950-2018.csv : Homogenized series (by rows).
- VV\_1950-2018.est : Station coordinates, codes and names.

One out of every 100 series will output also auxiliary files aux-N\_1950-2014\* (N being the correlative series number) to allow monitoring a sample of the homogenization processes.

## 4. Acknowledgments

Project INDECIS is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

ECA&D and EURO-CORDEX are also acknowledged, since their data are crucial for this project.

Support for the Twentieth Century Reanalysis Project version 2c dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office.