CORDEX Flagship Pilot Study (FPS) southeast Africa: rainfall evaluation

REB

2024-06-10

Contents

1 Introduction

This document is an R-markdown script that blends R-code (contained in grey 'chunks' below), comments and results generated by running the R-code provided. It therefore makes the analysis transparent and adds replicability. This document can also be used as an example, serve as a training asset, and help people learning the utility and merit of the R-environment.

1.1 Rain gauge data evaluation

This is an evaluation of rain gauge data from southeastern Africa based on a comparison with the ERA5 reanalysis (Hersbach and Dee (2016), Hersbach et al. (2020)) obtained from Copernicus C3S https://cds.climate.copernicus.eu/. The analysis was also carried out for ERA5-Land by comparing ERA5 with ERA5-Land to see if there were large differences. The evaluation carried out here also was designed to be relevant for subsequent empirical-statistical downscaling work, and hence use downscaling techniques to evaluate the data.

A shared CORDEX FPS southeast Africa data set combines rain gauge observations from selected sites from several different countries. The embedded R-code shows how the data is read and processed before being analysed. In other words, this R-markdown provides the R-code and the results of the analysis, e.g. in terms of plots and figures.

1.1.1 Set-up R-envisonment and activate R-package

The analysis carried out here in R-studio makes us of an R-package called `esd`, and the following chunk of R code checks if it is installed - if not, tries to install it from GitHub.com. More information and documentation can be found on https://github.com/metno/esd/wiki.

```
## Install the esd package if it's not already installed:
install.esd <- ("esd" %in% rownames(installed.packages()) == FALSE)

if (install.esd) {
  ## Install esd from GitHub:
  install.devtools <- ("devtools" %in% rownames(installed.packages()) == FALSE)

  if (install.devtools) {
    print('Need to install the devtools package')
    ## You need online access.
    install.packages('devtools', dependencies = TRUE)
  }
  library(devtools)

  print('Now install the esd package')
  ## You need online access.
  install_github('metno/esd')
}

## Other R-packages/libraries
install.plotrix <- ("plotrix" %in% rownames(installed.packages()) == FALSE)
if (install.plotrix) install.packages('plotrix', dependencies = TRUE)
install.readxl <- ("readxl" %in% rownames(installed.packages()) == FALSE)
if (install.readxl) install.packages('readxl', dependencies = TRUE)
```

Once the R-packages have been installed, we need to activate them:

```
## CORDEX FPS southeast Africa TELECOM#5
## Tuesday, June 28. 09:30 - 12:00
## Rasmus.Benestad-at-met.no

## The following lines read an Excel file into R/R-studio:
## It's based on the import functionality in R-studio 'environment' panel

## First Excel files with station data has been downloaded on yor local computer.
## In this case: Downloads/MOZStationPrecip1979-2020.xlsx (but it may be easier to use
```

```r
## 'Import' in R-studio...)

## This package is installed directly from R-studio: Tools -> Install packages from CRAN
library(readxl)
## Available from: https://github.com/metno/esd/wiki/0.-How-to-install%3F
## Or https://drive.google.com/drive/folders/1mUkpH50DFISWzL71I4eXO17eQ7_Vi2wu
library(esd)
```

```
## Loading required package: ncdf4

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

## Registered S3 methods overwritten by 'esd':
##   method        from
##   subset.matrix base
##   subset.zoo    zoo
```

```r
library(plotrix)
library(RColorBrewer)
print(sessionInfo())
```

```
## R version 4.5.0 (2025-04-11)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 24.04.2 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.26.so;  LAPACK version 3.12.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=nb_NO.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=nb_NO.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=nb_NO.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=nb_NO.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Oslo
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] RColorBrewer_1.1-3 plotrix_3.8-4      esd_1.10.95        zoo_1.8-14
## [5] ncdf4_1.24         readxl_1.4.3
##
## loaded via a namespace (and not attached):
##  [1] digest_0.6.34     fastmap_1.1.1     xfun_0.41         cellranger_1.1.0
```

```
##  [5] lattice_0.22-5    knitr_1.45        htmltools_0.5.7   rmarkdown_2.25
##  [9] cli_3.6.2         grid_4.5.0        compiler_4.5.0    rstudioapi_0.15.0
## [13] tools_4.5.0       evaluate_0.23     yaml_2.3.8        rlang_1.1.3
```

```
#tmpdir <- '~/r-studio.tmp'
#dir.create(tmpdir)
#setwd(tmpdir)
```

1.1.2 Define year: October–September

The evaluation is based on annually aggregated data, but here the year is redefined as starting on October 1st and ending on September 30th the following year. The year can be redefined through the argument `start=year.start` in the esd-function `annual()`. Here we set it to October (the calculations are done on monthly aggregated data):

```
year.start='Oct'
```

1.1.3 Read the data from downloaded (Excel) spreadsheets

The following chunk of R-code reads rain gauge data from provided Excel/CSV files. The rain gauge data are from a set of observations that have been shared between the partners of the CORDEX Flagship Pilot Study (FPS) for southeast Africa. CORDEX (https://cordex.org/) is the Coordinated Downscaling Experiment under the World Climate Research Programme (WCRP; https://www.wcrp-climate.org/).

The use of the said shared data is restricted to within CORDEX FPS southeast Africa by mutual agreement. The data files from the partner countries have been formatted the same way, more or less - with first rows containing location names and coordinates. The rain gauges constitutes a selected sample, and there may be more in-situ data from this region that may be available from their respective national meteorological agencies.

```
if (!file.exists('precip.southeastAfrica.nc')) {
  print('Moz')
  year.start <- 'Oct'  # Hydrological year from October 1st to September 30th.
  precip <- read_excel("~/Downloads/MOZStationPrecip1979-2021.xlsx")
  #View(precip)

  ## The next few lines prepare the data: longitudes, latitudes, altitudes,
  ## time, and the precipitation:
  lon <- as.numeric(precip[1,-1])
  lat <- as.numeric(precip[2,-1])
  alt <- as.numeric(precip[3,-1])
  tim <- as.numeric(unlist(c(precip[-c(1:3),1])))
  pr <- as.numeric(unlist(c(precip[-c(1:3),-1])))
  ## Set negative rainfall to missing value:
  pr[pr < 0] <- NA
  ## Ensure the data object has correct dimensions (number of days, number of stations)
  dim(pr) <- c(length(tim),length(lon))

  ## The next few lines makes the data compatible with the esd-package:
  X <- zoo(pr,order.by=as.Date(tim-2,origin='1900-01-01'))
  X <- as.station(X,loc = colnames(precip)[-1],unit='mm',
                  param='precip',lon=lon,lat=lat,alt=alt,cntr=rep('Mozambique',length(lon)),
                  longname='Daily_rainfall',scr=rep('INAM',length(lon)))


  ## ------------ Add stations from Zimbabwe
```

```r
print('Zim')
precip <- read_excel('~/Downloads/Zimbabwe_Cordex Data.xls')
## The next few lines prepare the data: longitudes, latitudes, altitudes,
## time, and the precipitation:
lon <- as.numeric(precip[1,-1])
lat <- as.numeric(precip[2,-1])
alt <- as.numeric(precip[3,-1])
tim <- as.numeric(unlist(c(precip[-c(1:3),1])))
pr <- as.numeric(unlist(c(precip[-c(1:3),-1])))
## Set negative rainfall to missing value:
pr[pr < 0] <- NA
pr[pr > 1000] <- NA

## Ensure the data object has correct dimensions (number of days, number of stations)
dim(pr) <- c(length(tim),length(lon))

## The next few lines makes the data compatible with the esd-package:
Y <- zoo(pr,order.by=as.Date(tim-2,origin='1900-01-01'))
Y <- as.station(Y,loc = colnames(precip)[-1],unit='mm',
                param='precip',lon=lon,lat=lat,alt=alt,cntr=rep('Zimbabwe',length(lon)),
                longname='Daily_rainfall',src=rep('NA',length(lon)))
X <- combine.stations(X,Y)
#print(loc(X))
rm("Y")

## ------------ Add South Africa
print('SA')
precip <- read_excel('~/Downloads/SA_StationPrecip1979-2021.xlsx')
## The next few lines prepare the data: longitudes, latitudes, altitudes,
## time, and the precipitation:
lon <- as.numeric(precip[1,-1])
lat <- as.numeric(precip[2,-1])
alt <- as.numeric(precip[3,-1])
tim <- as.numeric(unlist(c(precip[-c(1:3),1])))
pr <- as.numeric(unlist(c(precip[-c(1:3),-1])))
## Set negative rainfall to missing value:
pr[pr < 0] <- NA
pr[pr > 1000] <- NA

## Ensure the data object has correct dimensions (number of days, number of stations)
dim(pr) <- c(length(tim),length(lon))

## The next few lines makes the data compatible with the esd-package:
Y <- zoo(pr,order.by=as.Date(tim-2,origin='1900-01-01'))
Y <- as.station(Y,loc = colnames(precip)[-1],unit='mm',
                param='precip',lon=lon,lat=lat,alt=alt,cntr=rep('South Africa',length(lon)),
                longname='Daily_rainfall',rep('SAWS',length(lon)))
X <- combine.stations(X,Y)
#print(loc(X))
rm("Y")

## ------------ Add Malawi
print('Malawi')
```

```
precip <- read.csv('~/Downloads/Malawi_Precip.csv')
## The next few lines prepare the data: longitudes, latitudes, altitudes,
## time, and the precipitation:
lon <- as.numeric(precip[1,-1])
lat <- as.numeric(precip[2,-1])
alt <- as.numeric(precip[3,-1])
tim <- as.Date(precip[-c(1:3),1],format="%m/%d/%Y")
pr <- as.numeric(unlist(c(precip[-c(1:3),-1])))
## Set negative rainfall to missing value:
pr[pr < 0] <- NA
pr[pr > 1000] <- NA

## Ensure the data object has correct dimensions (number of days, number of stations)
dim(pr) <- c(length(tim),length(lon))

## The next few lines makes the data compatible with the esd-package:
Y <- zoo(pr,order.by=tim)
Y <- as.station(Y,loc = colnames(precip)[-1],unit='mm',
                param='precip',lon=lon,lat=lat,alt=alt,cntr=rep('Malawi',length(lon)),
                longname='Daily_rainfall',rep('DCCMS',length(lon)))
X <- combine.stations(X,Y)
#print(loc(X))
rm("Y")

## ------------  Add Rwanda
print('Rwanda')
precip <- read_excel('~/Downloads/Rwanda_Rainfall_Tmax_Tmin_1981_1983.xlsx')
precip <- data.frame(precip)
## The next few lines prepare the data: longitudes, latitudes, altitudes,
## time, and the precipitation:
lon <- as.numeric(precip[1,-1])
lat <- as.numeric(precip[2,-1])
alt <- as.numeric(precip[3,-1])
tim <- as.Date(as.character(as.integer(precip[-c(1:2),1])),format="%Y%m%d")
pr <- as.numeric(unlist(c(precip[-c(1:2),-1])))
## Set negative rainfall to missing value:
pr[pr < 0] <- NA
pr[pr > 1000] <- NA

## Ensure the data object has correct dimensions (number of days, number of stations)
dim(pr) <- c(length(tim),length(lon))

## The next few lines makes the data compatible with the esd-package:
Y <- zoo(pr,order.by=tim)
Y <- as.station(Y,loc = colnames(precip)[-1],unit='mm',
                param='precip',lon=lon,lat=lat,alt=alt,cntr=rep('Rwanda',length(lon)),
                longname='Daily_rainfall',scr=rep('Meteo Rwanda',length(lon)))

X <- combine.stations(X,Y)

## Add Kenya
print("Add Kenya")
load('~/Downloads/Kenya_daily_rain_gauge.rda')
```

```r
X <- combine.stations(X,Y)
load('~/Downloads/kalro.rda')
X <- combine.stations(X,kalro)

## Add Tanzania
print("Add Tanzania")
precip <- read_excel('~/Downloads/TZprecip1979-2021.xlsx')
## The next few lines prepare the data: longitudes, latitudes, altitudes,
## time, and the precipitation:
lon <- as.numeric(precip[1,-1])
lat <- as.numeric(precip[2,-1])
alt <- as.numeric(precip[3,-1])
tim <- as.Date(as.numeric(unlist(c(precip[-c(1:3),1])))-2,origin='1900-01-01')
pr <- as.numeric(unlist(c(precip[-c(1:3),-1])))
## Set negative rainfall to missing value:
pr[pr < 0] <- NA
pr[pr > 1000] <- NA

## Ensure the data object has correct dimensions (number of days, number of stations)
dim(pr) <- c(length(tim),length(lon))

## The next few lines makes the data compatible with the esd-package:
ok <- is.finite(tim)
Y <- zoo(pr[ok,],order.by=tim[ok])
Y <- as.station(Y,loc = colnames(precip)[-1],unit='mm',
                param='precip',lon=lon,lat=lat,alt=alt,cntr=rep('Tanzania',length(lon)),
                longname='Daily_rainfall',scr=rep('TMA',length(lon)))

X <- combine.stations(X,Y)

## Add Madagaskar
print("Add Madagaskar")
precip <- read_excel('~/Downloads/Madagascar_DailyRainfall_FPS_1981_2021.xlsx')
## The next few lines prepare the data: longitudes, latitudes, altitudes,
## time, and the precipitation:
lon <- as.numeric(precip[1,-1])
lat <- as.numeric(precip[2,-1])
alt <- as.numeric(precip[3,-1])
tim <- as.Date(as.numeric(unlist(c(precip[-c(1:3),1])))-2,origin='1900-01-01')
pr <- as.numeric(unlist(c(precip[-c(1:3),-1])))
## Set negative rainfall to missing value:
pr[pr < 0] <- NA
pr[pr > 1000] <- NA

## Ensure the data object has correct dimensions (number of days, number of stations)
dim(pr) <- c(length(tim),length(lon))

## The next few lines makes the data compatible with the esd-package:
ok <- is.finite(tim)
Y <- zoo(pr[ok,],order.by=tim[ok])
Y <- as.station(Y,loc = colnames(precip)[-1],unit='mm',
                param='precip',lon=lon,lat=lat,alt=alt,cntr=rep('Madagascar',length(lon)),
                longname='Daily_rainfall',scr=rep('ENACTS DATABASE of Madagascar',length(lon)))
```

```r
X <- combine.stations(X,Y)

## Add Lesotho
precip <- read.csv('~/Downloads/Lesothodata19802020.csv')
lon <- as.numeric(precip[1,-1])
lat <- as.numeric(precip[2,-1])
alt <- as.numeric(precip[3,-1])
tim <- as.Date(precip[-c(1:3),1],origin='1900-01-01',format='%d/%m/%Y')
pr <- as.matrix(precip[-c(1:3),-1])
## Set negative rainfall to missing value:
pr[pr < 0] <- NA
pr[pr > 1000] <- NA
## Ensure the data object has correct dimensions (number of days, number of stations)
dim(pr) <- c(length(tim),length(lon))

## The next few lines makes the data compatible with the esd-package:
ok <- is.finite(tim)
Y <- zoo(pr[ok,],order.by=tim[ok])
Y <- as.station(Y,loc = colnames(precip)[-1],unit='mm',
                param='precip',lon=lon,lat=lat,alt=alt,cntr=rep('Lesotho',length(lon)),
                longname='Daily_rainfall',scr=rep('Lesotho Meteorological Services',length(lon)))

X <- combine.stations(X,Y)

## Add Burundi
precip <- read_excel("Downloads/Burundi Données_Prec_1991_2021.xlsx")
mburundi  <- read_excel("Downloads/station meteo.xlsx")

tim <- as.Date(precip$Date)
lon <- mburundi$`X= LONG`
lat <- mburundi$`Y=LAT`
alt <- mburundi$`Z=ELEVATION`
loc <- mburundi$`STATION NAME`
stid <- mburundi$ID
## Match record with metadata
cnames <- sub('TORA','MPOTA',toupper(colnames(precip))[-1])
srt <- rep(NA,length(cnames))
for (i in 1:length(srt)) srt[i] <- grep(substr(cnames[i],1,3),substr(loc,1,3))
print(rbind(cnames,loc[srt]))
d <- dim(as.matrix(precip)[,-1])
pr <- as.numeric(as.matrix(precip)[,-1])
## Set negative rainfall to missing value:
pr[pr < 0] <- NA
pr[pr > 1000] <- NA
dim(pr) <- d
Y <- zoo(pr,order.by=tim)
Y <- as.station(Y,loc = loc[srt],stid=stid[srt],unit='mm',
                param='precip',lon=lon[srt],lat=lat[srt],alt=alt[srt],
                cntr=rep('Burundi',length(lon)),longname='Daily_rainfall',
                scr=rep('HAKIZIMANA Jean Claude',length(lon)))
Y <- subset(Y,it=!duplicated(index(Y)))
X <- combine.stations(X,Y)
```

```r
  attr(X,'unit') <- 'mm'
  attr(X,'variable') <- 'precip'
  attr(X,'station_id') <- 1:length(loc(X))
  attr(X,'source') <- 'Southeast_Africa'
  ## Save the station data in a netCDF file so that the data can be examined through OpenClimateData
  ## <https://github.com/metno/OpenClimateData/>
  write2ncdf4(X,file='precip.southeastAfrica.nc')
} else X <- retrieve('precip.southeastAfrica.nc')
## Clean up: remove stations with little data
nv <- apply(X,2,'nv')
X <- subset(X,is=nv > 9000)
print(loc(X))
```

```
##   [1] "Pemba"                      "Lichinga"
##   [3] "Nampula"                    "Quelimane"
##   [5] "Tete"                       "Chimoio"
##   [7] "BeiraObs"                   "Inhambane"
##   [9] "Xai-Xai"                    "MaputoObs"
##  [11] "Beitbridge"                 "Belvedere"
##  [13] "Bulawayo Goetz"             "Gweru"
##  [15] "Hwange"                     "Matopos"
##  [17] "Chipinge"                   "Gokwe"
##  [19] "Mt Darwin"                  "Nyanga"
##  [21] "Bloemfontein WO"            "Bothaville - Balkfontein"
##  [23] "Brandvlei"                  "Cape Agulhas"
##  [25] "Cape St. Francis"           "Cape Town WO"
##  [27] "Cedara"                     "East London WO"
##  [29] "Irene WO"                   "Kimberley WO"
##  [31] "Laingsburg"                 "Marico"
##  [33] "Mount Edgecombe"            "Ottosdal"
##  [35] "Polokwane WO"               "Punda Maria"
##  [37] "Secunda"                    "Skukuza"
##  [39] "Upington WO"                "Warmbad Towoomba"
##  [41] "Chitipa"                    "Karonga"
##  [43] "Mzimba"                     "Kasungu"
##  [45] "KIA"                        "Salima"
##  [47] "Makoka"                     "Nkhotakota"
##  [49] "Dedza"                      "Mangochi"
##  [51] "Chileka"                    "Nkhatabay"
##  [53] "Bvumbwe"                    "Chichiri"
##  [55] "Mimosa"                     "Chikwawa"
##  [57] "Mchinji"                    "Makhanga"
##  [59] "Nsanje"                     "Naminjiwa"
##  [61] "Neno"                       "Mwanza"
##  [63] "Mpemba"                     "ZAagr"
##  [65] "WlkrsFerry"                 "Phalula"
##  [67] "Balaka"                     "Mzuzu"
##  [69] "Kbay"                       "Tembwe"
##  [71] "Nkhande"                    "Ntaja"
##  [73] "KIGALIAERO"                 "KAMEMBEAERO"
##  [75] "GISENYIAERO"                "Lodwar"
##  [77] "Mandera"                    "Kitale"
##  [79] "Kericho"                    "Kisii"
##  [81] "Narok"                      "Nyeri"
```

```
##  [83] "Dagoretti Corner"         "Machakos Agromet"
##  [85] "Voi"                      "Lamu"
##  [87] "Moi International Airpor" "NA"
##  [89] "Bukoba"                   "Musoma"
##  [91] "Mwanza"                   "Moshi"
##  [93] "Arusha"                   "Same"
##  [95] "Tabora"                   "Tanga"
##  [97] "Dodoma"                   "Morogoro"
##  [99] "Dar es Salaam"            "Sumbawanga"
## [101] "Mbeya"                    "Songea"
## [103] "Mtwara"                   "Antananarivo"
## [105] "Antsiranana"              "Fianarantsoa"
## [107] "Mahajanga"                "Toamasina"
## [109] "Toliary"                  "MOHALESHOEK"
## [111] "QACHASNEK"                "BUTHABUTHE"
## [113] "LERIBE"                   "MAFETENG"
## [115] "MALEFILOANE"              "MAPOTENG"
## [117] "MEJAMETALANA"             "MOKHOTLONG"
## [119] "MOSHOESHOEI"              "OXBOW"
## [121] "PHUTHIATSANA"             "QUTHING"
## [123] "SEMONKONG"                "THABATSEKA"
## [125] "BUJUMBURA (Aeroport)"     "GITEGA (Aerodrome)"
## [127] "CANKUZO"                  "GISOZI"
## [129] "MUYINGA"                  "MUSASA"
## [131] "NYAMUSWAGA"               "MPOTA (Tora)"
```

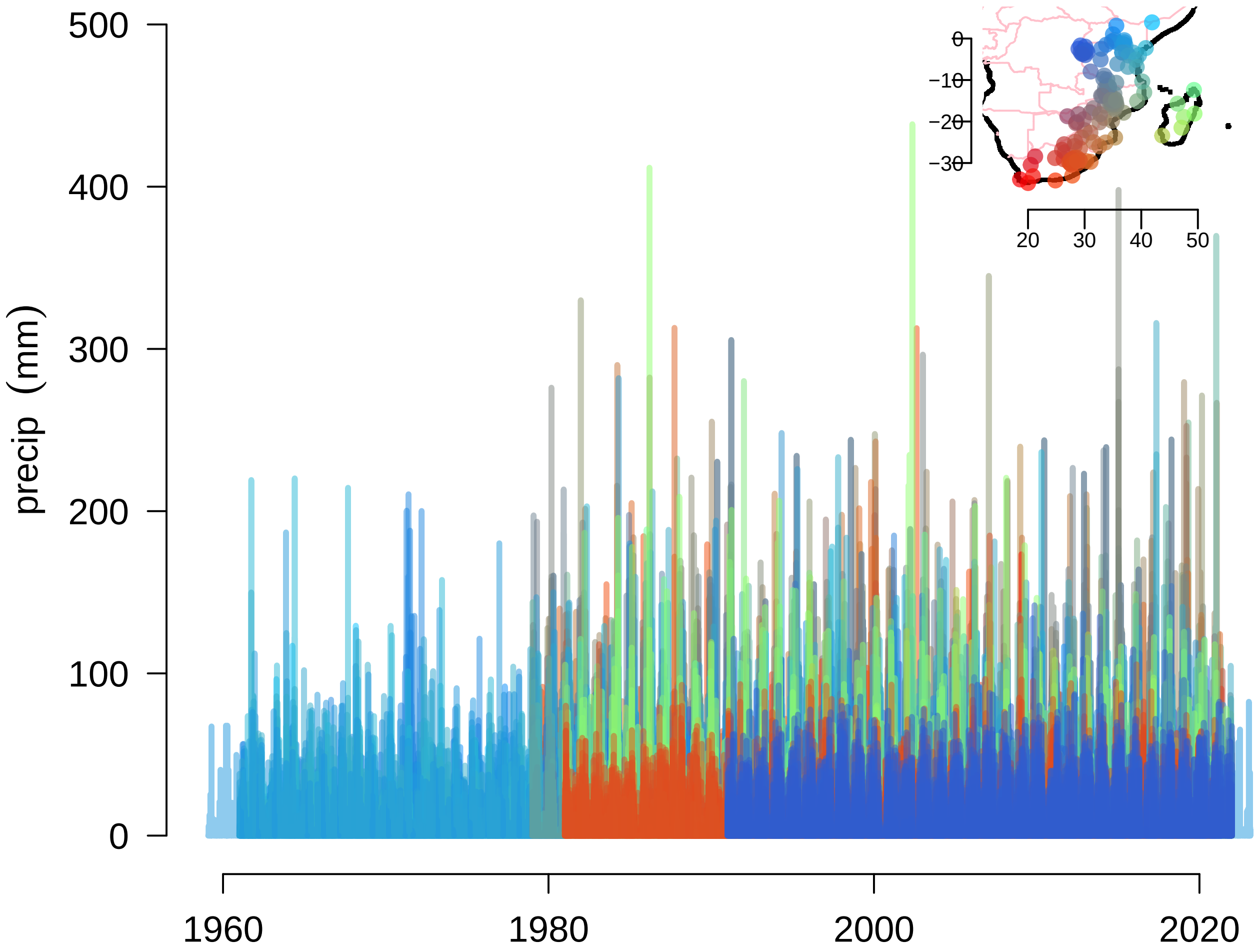
```
print(table(cntr(X)))
```

```
##
##      Burundi        Kenya      Lesotho   Madagascar       Malawi   Mozambique
##            8           13           15            6           32           10
##       Rwanda South Africa     Tanzania     Zimbabwe
##            3           20           15           10
```

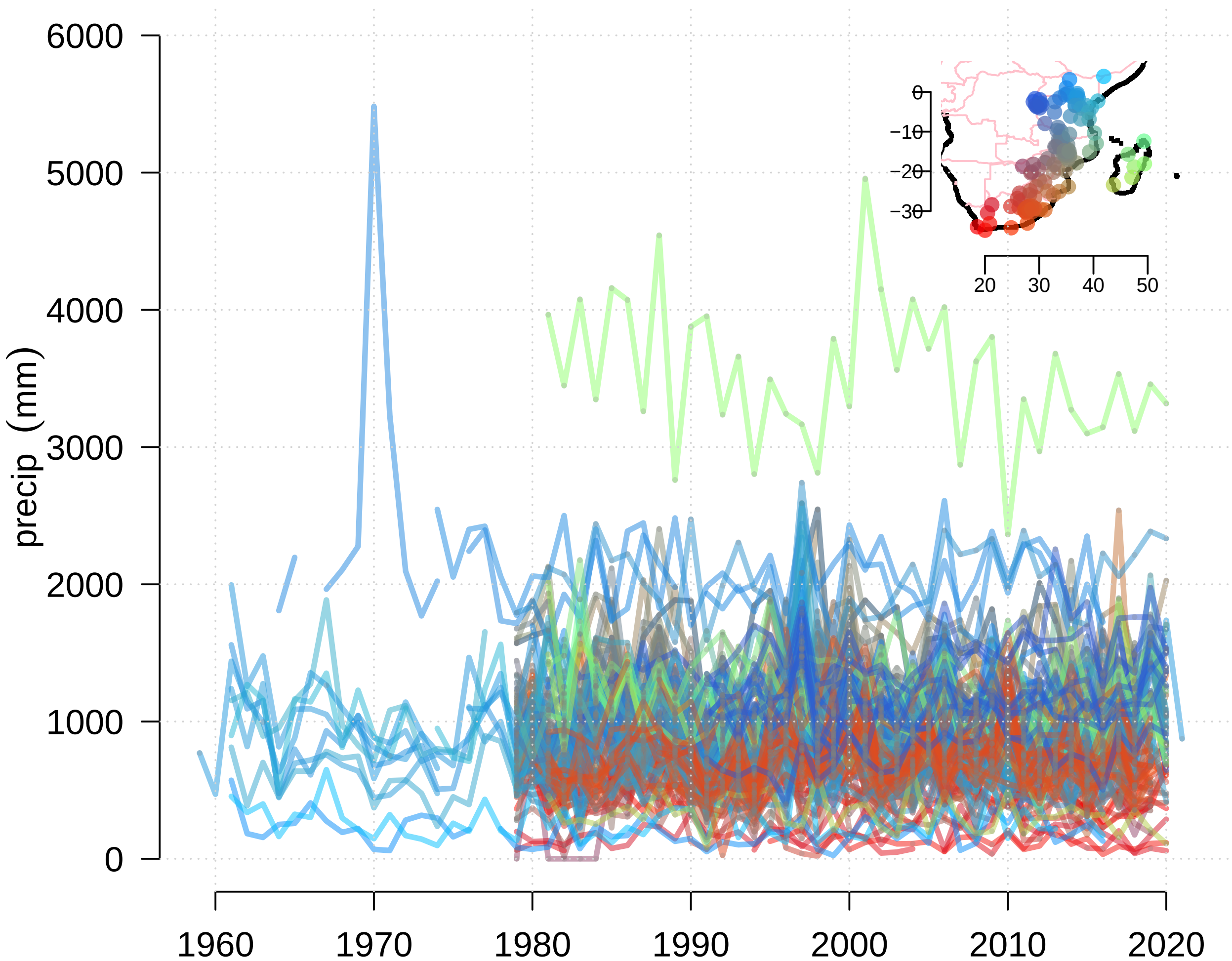
```
print(range(index(X)))
```

```
## [1] "1959-02-02" "2023-02-17"
```

```
rm("Y")


## Add annual rainfall totals from Madagascar: Mdg
# load('~/Downloads/Madagascar_annual-rainfall.rda')
# index(Mdg) <- year(Mdg)
# attr(Mdg,'country') <- rep('Madagascar',length(loc(Mdg)))

## Once in the esd-format, we can apply the esd-functionality
## e.g.
plot(X,new=FALSE)     ## Using the plot method for esd-station-objects...
```




```
## Also plot the annual sums - here we use the product between the mean and number
## of years due to years with missing days of data:
anntp <- 365.25*annual(X,FUN='mean',nmin=300,start=year.start)
plot(anntp,new=FALSE,ylim=c(0,6000))
grid()
```



A first line of quality control is to visualise the data. The chunk of R-code above reads and processes the data, and the plots give a first impression of the data quality in addition to its metadata (coordinates, country, location name).

Both the daily data as well as the annually aggregated statistics were plotted, as it's easier to see the differences between the sites from annual data. The colours of the time series in the figure match the colours of the symbols marking their location on the inserted map.

A plot of the annual rainfall (Oct 1st - Sept. 31st) reveals more systematic differences with Toamasina (Madagascar) having systematically higher annual rainfall than the other stations, but Kisii (Kenya) also have some suspect/interesting(?) data with the highest amount (5640 mm) in one year (1970) that can be described as an outlier.

Below is a detailed overview of location names, table of sites from each country and plot suggest that the data have been processed in an acceptable way and contain reasonable numbers.

```
maxanntp <-round(apply(anntp,2,max,na.rm=TRUE))
srt <- order(maxanntp)
print(cbind(loc(X),cntr(X),maxanntp)[srt,])
```

```
##                                                  maxanntp
## X.31   "Laingsburg"              "South Africa" "229"
## X.23   "Brandvlei"               "South Africa" "263"
## X.39   "Upington WO"             "South Africa" "559"
## X.76   "Lodwar"                  "Kenya"        "618"
## X.36   "Punda Maria"             "South Africa" "633"
```

```
## X.109 "Toliary"                    "Madagascar"   "648"
## X.24  "Cape Agulhas"               "South Africa" "740"
## X.26  "Cape Town WO"               "South Africa" "787"
## X.118 "MOKHOTLONG"                 "Lesotho"      "807"
## X.30  "Kimberley WO"               "South Africa" "864"
## X.34  "Ottosdal"                   "South Africa" "878"
## X.13  "Bulawayo Goetz"             "Zimbabwe"     "917"
## X.40  "Warmbad Towoomba"           "South Africa" "932"
## X.35  "Polokwane WO"               "South Africa" "942"
## X.15  "Hwange"                     "Zimbabwe"     "964"
## X.11  "Beitbridge"                 "Zimbabwe"     "977"
## X.21  "Bloemfontein WO"            "South Africa" "991"
## X.22  "Bothaville - Balkfontein"   "South Africa" "993"
## X.32  "Marico"                     "South Africa" "1015"
## X.25  "Cape St. Francis"           "South Africa" "1017"
## X.124 "THABATSEKA"                 "Lesotho"      "1022"
## X.115 "MALEFILOANE"                "Lesotho"      "1056"
## X.16  "Matopos"                    "Zimbabwe"     "1063"
## X.37  "Secunda"                    "South Africa" "1121"
## X.77  "Mandera"                    "Kenya"        "1121"
## X.117 "MEJAMETALANA"               "Lesotho"      "1125"
## X.123 "SEMONKONG"                  "Lesotho"      "1126"
## X.18  "Gokwe"                      "Zimbabwe"     "1138"
## X.110 "MOHALESHOEK"                "Lesotho"      "1144"
## X.44  "Kasungu"                    "Malawi"       "1152"
## X.111 "QACHASNEK"                  "Lesotho"      "1160"
## X.14  "Gweru"                      "Zimbabwe"     "1168"
## X.38  "Skukuza"                    "South Africa" "1172"
## X.43  "Mzimba"                     "Malawi"       "1172"
## X.45  "KIA"                        "Malawi"       "1175"
## X.125 "BUJUMBURA (Aeroport)"       "Burundi"      "1182"
## X.97  "Dodoma"                     "Tanzania"     "1188"
## X.114 "MAFETENG"                   "Lesotho"      "1193"
## X.100 "Sumbawanga"                 "Tanzania"     "1234"
## X.56  "Chikwawa"                   "Malawi"       "1240"
## X.94  "Same"                       "Tanzania"     "1248"
## X.122 "QUTHING"                    "Lesotho"      "1251"
## X.12  "Belvedere"                  "Zimbabwe"     "1260"
## X.50  "Mangochi"                   "Malawi"       "1260"
## X.29  "Irene WO"                   "South Africa" "1270"
## X.72  "Ntaja"                      "Malawi"       "1275"
## X.119 "MOSHOESHOEI"                "Lesotho"      "1286"
## X.19  "Mt Darwin"                  "Zimbabwe"     "1291"
## X.51  "Chileka"                    "Malawi"       "1292"
## X.116 "MAPOTENG"                   "Lesotho"      "1292"
## X.85  "Voi"                        "Kenya"        "1310"
## X.70  "Tembwe"                     "Malawi"       "1321"
## X.41  "Chitipa"                    "Malawi"       "1330"
## X.113 "LERIBE"                     "Lesotho"      "1368"
## X.49  "Dedza"                      "Malawi"       "1371"
## X.95  "Tabora"                     "Tanzania"     "1376"
## X.121 "PHUTHIATSANA"               "Lesotho"      "1388"
## X.5   "Tete"                       "Mozambique"   "1425"
## X.92  "Moshi"                      "Tanzania"     "1436"
```

```
## X.73  "KIGALIAERO"               "Rwanda"       "1439"
## X.112 "BUTHABUTHE"               "Lesotho"      "1459"
## X.98  "Morogoro"                 "Tanzania"     "1477"
## X.101 "Mbeya"                    "Tanzania"     "1479"
## X.58  "Makhanga"                 "Malawi"       "1481"
## X.27  "Cedara"                   "South Africa" "1491"
## X.28  "East London WO"           "South Africa" "1501"
## X.42  "Karonga"                  "Malawi"       "1525"
## X.84  "Machakos Agromet"         "Kenya"        "1530"
## X.129 "MUYINGA"                  "Burundi"      "1530"
## X.130 "MUSASA"                   "Burundi"      "1544"
## X.81  "Narok"                    "Kenya"        "1559"
## X.8   "Inhambane"                "Mozambique"   "1563"
## X.65  "WlkrsFerry"               "Malawi"       "1568"
## X.66  "Phalula"                  "Malawi"       "1570"
## X.9   "Xai-Xai"                  "Mozambique"   "1578"
## X.61  "Neno"                     "Malawi"       "1591"
## X.93  "Arusha"                   "Tanzania"     "1594"
## X.33  "Mount Edgecombe"          "South Africa" "1596"
## X.47  "Makoka"                   "Malawi"       "1618"
## X.71  "Nkhande"                  "Malawi"       "1625"
## X.90  "Musoma"                   "Tanzania"     "1629"
## X.127 "CANKUZO"                  "Burundi"      "1644"
## X.57  "Mchinji"                  "Malawi"       "1666"
## X.75  "GISENYIAERO"              "Rwanda"       "1667"
## X.91  "Mwanza"                   "Tanzania"     "1670"
## X.2   "Lichinga"                 "Mozambique"   "1671"
## X.120 "OXBOW"                    "Lesotho"      "1671"
## X.126 "GITEGA (Aerodrome)"       "Burundi"      "1676"
## X.6   "Chimoio"                  "Mozambique"   "1678"
## X.54  "Chichiri"                 "Malawi"       "1687"
## X.74  "KAMEMBEAERO"              "Rwanda"       "1692"
## X.103 "Mtwara"                   "Tanzania"     "1708"
## X.20  "Nyanga"                   "Zimbabwe"     "1714"
## X.46  "Salima"                   "Malawi"       "1725"
## X.68  "Mzuzu"                    "Malawi"       "1725"
## X.64  "ZAagr"                    "Malawi"       "1759"
## X.131 "NYAMUSWAGA"               "Burundi"      "1776"
## X.102 "Songea"                   "Tanzania"     "1781"
## X.1   "Pemba"                    "Mozambique"   "1839"
## X.67  "Balaka"                   "Malawi"       "1867"
## X.60  "Naminjiwa"                "Malawi"       "1869"
## X.132 "MPOTA (Tora)"             "Burundi"      "1870"
## X.17  "Chipinge"                 "Zimbabwe"     "1871"
## X.62  "Mwanza"                   "Malawi"       "1889"
## X.86  "Lamu"                     "Kenya"        "1889"
## X.105 "Antsiranana"              "Madagascar"   "1890"
## X.106 "Fianarantsoa"             "Madagascar"   "1890"
## X.53  "Bvumbwe"                  "Malawi"       "1891"
## X.3   "Nampula"                  "Mozambique"   "1899"
## X.78  "Kitale"                   "Kenya"        "1944"
## X.63  "Mpemba"                   "Malawi"       "1961"
## X.104 "Antananarivo"             "Madagascar"   "2012"
## X.83  "Dagoretti Corner"         "Kenya"        "2051"
```

```
## X.99  "Dar es Salaam"             "Tanzania"     "2066"
## X.48  "Nkhotakota"                "Malawi"       "2117"
## X.4   "Quelimane"                 "Mozambique"   "2128"
## X.59  "Nsanje"                    "Malawi"       "2135"
## X.82  "Nyeri"                     "Kenya"        "2173"
## X.107 "Mahajanga"                 "Madagascar"   "2177"
## X.128 "GISOZI"                    "Burundi"      "2256"
## X.88  "NA"                        "Kenya"        "2343"
## X.87  "Moi International Airpor"  "Kenya"        "2381"
## X.55  "Mimosa"                    "Malawi"       "2404"
## X.7   "BeiraObs"                  "Mozambique"   "2451"
## X.10  "MaputoObs"                 "Mozambique"   "2539"
## X.52  "Nkhatabay"                 "Malawi"       "2546"
## X.69  "Kbay"                      "Malawi"       "2546"
## X.96  "Tanga"                     "Tanzania"     "2592"
## X.79  "Kericho"                   "Kenya"        "2609"
## X.89  "Bukoba"                    "Tanzania"     "2740"
## X.108 "Toamasina"                 "Madagascar"   "4955"
## X.80  "Kisii"                     "Kenya"        "5483"
```

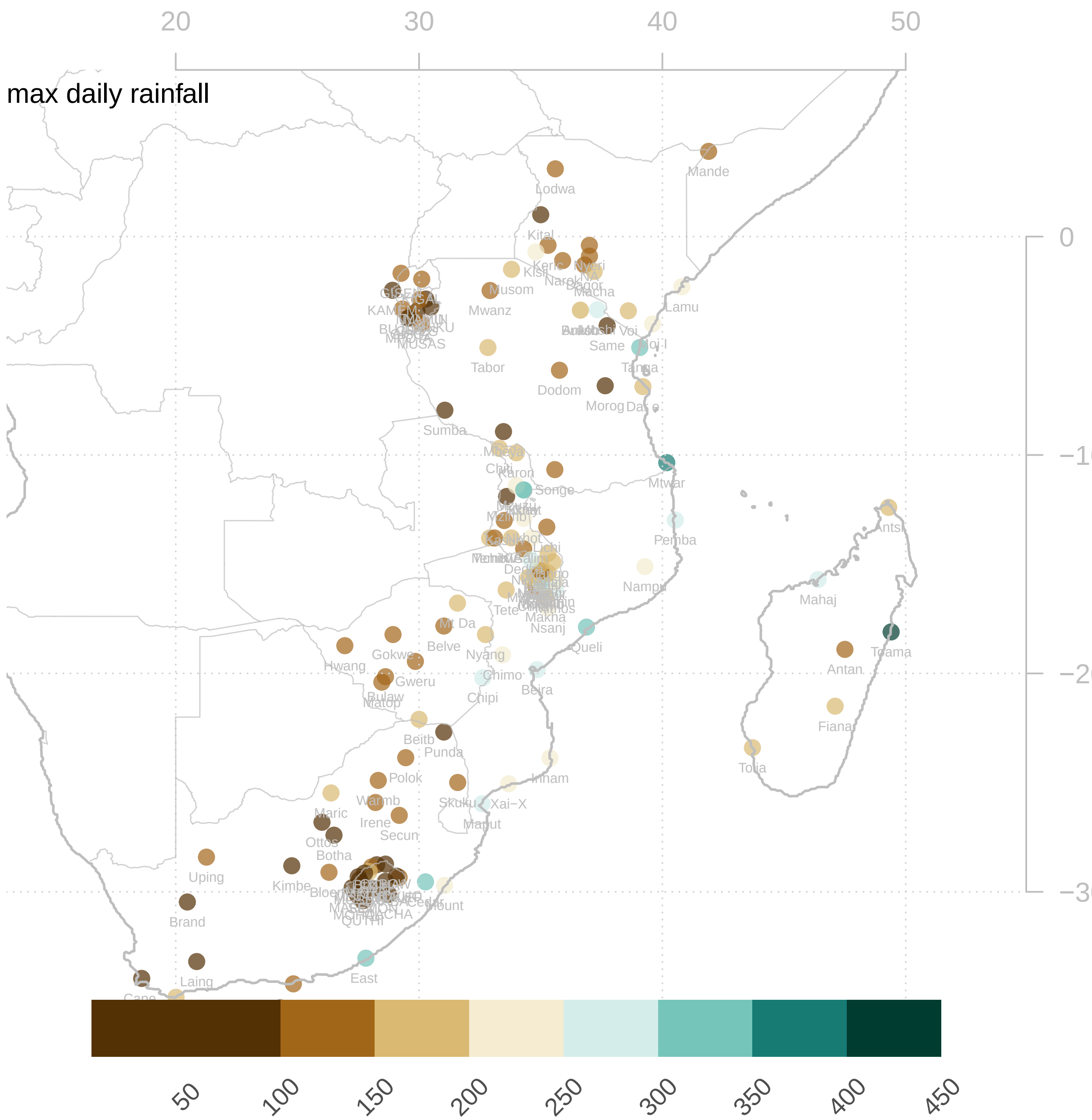
The chunk below shows the availability of the data - the white spaces indicate where there are gaps of missing data.

```
diagnose(X)
```

Southeast_Africa

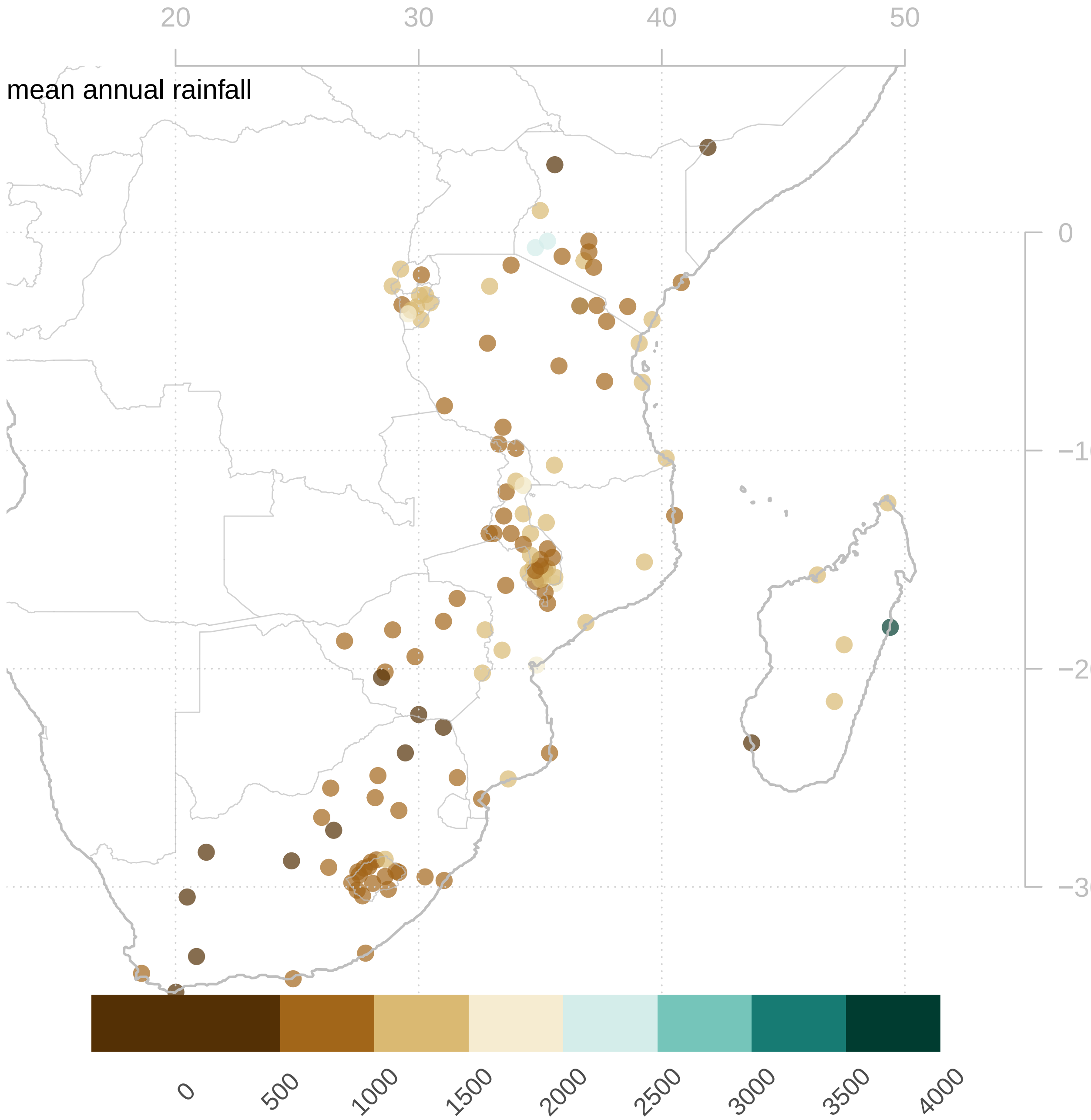
Another visual control is to plot a map of the average of the annual rainfall amounts - in this case, we look at the daily maximum values, and mean annual total rainfall. For the latter, we also set the range of values to suppress the weighting of outliers:

```
## Map of maximum values:
map(X,FUN='max',add.text = TRUE, colbar=list(pal='precip.ipcc'),main='max daily rainfall',
    new=FALSE,cex=1.25,cex.lab=0.5,border=TRUE)
```



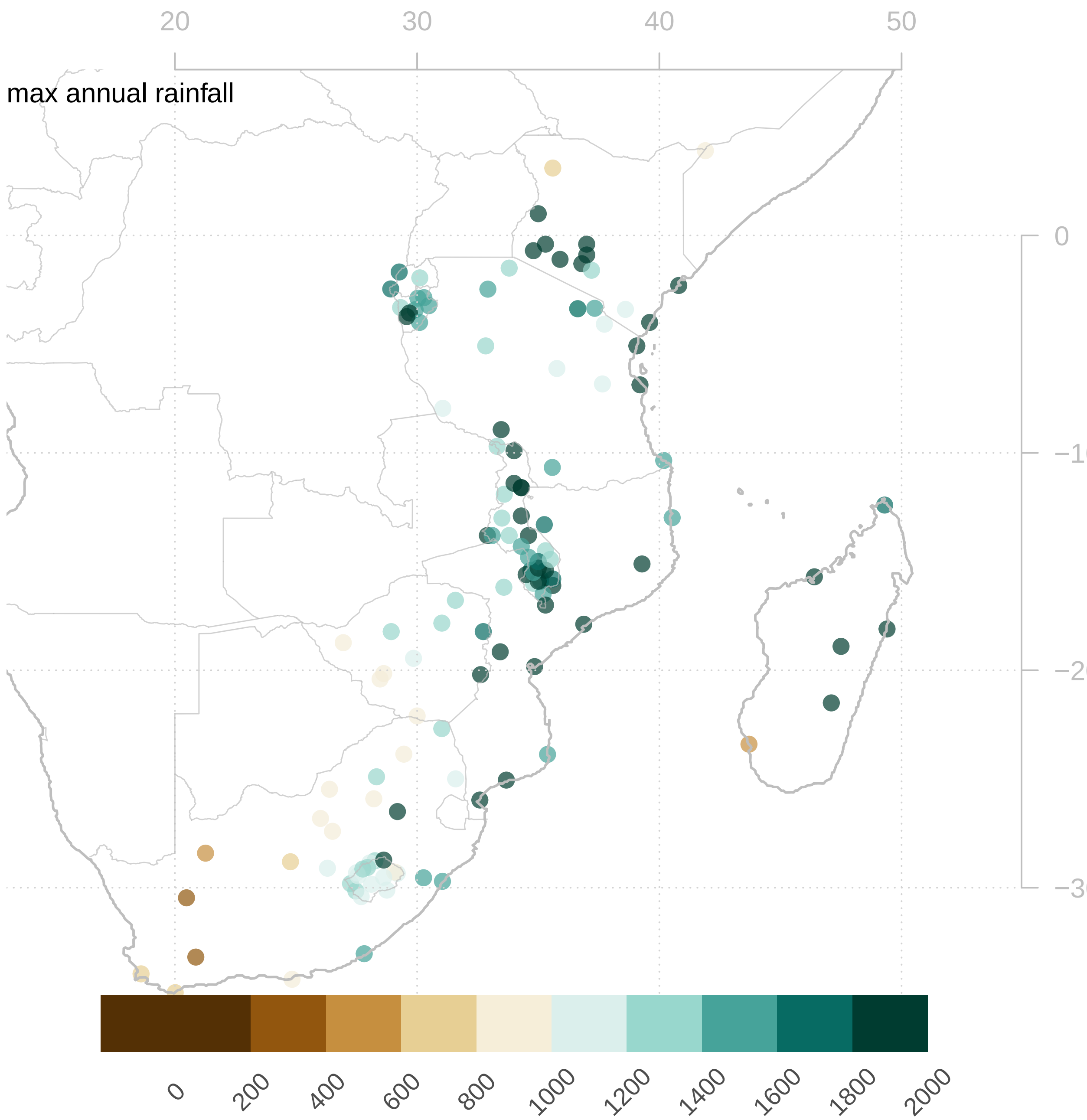

```
## Map of all stations without any constraints on the range of values:
map(365.25*annual(X,FUN='mean',nmin=30),FUN='mean',main='mean annual rainfall',
    colbar=list(pal='precip.ipcc'),new=FALSE,cex=1.25,border=TRUE)
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```




```
## Map of all stations with constraints on the range of values to disregard outliers:
map(365.25*annual(X,FUN='mean',nmin=30),FUN='max',main='max annual rainfall',
    colbar=list(breaks=seq(0,2000,by=200),pal='precip.ipcc'),new=FALSE,cex=1.25,border=TRUE)
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```



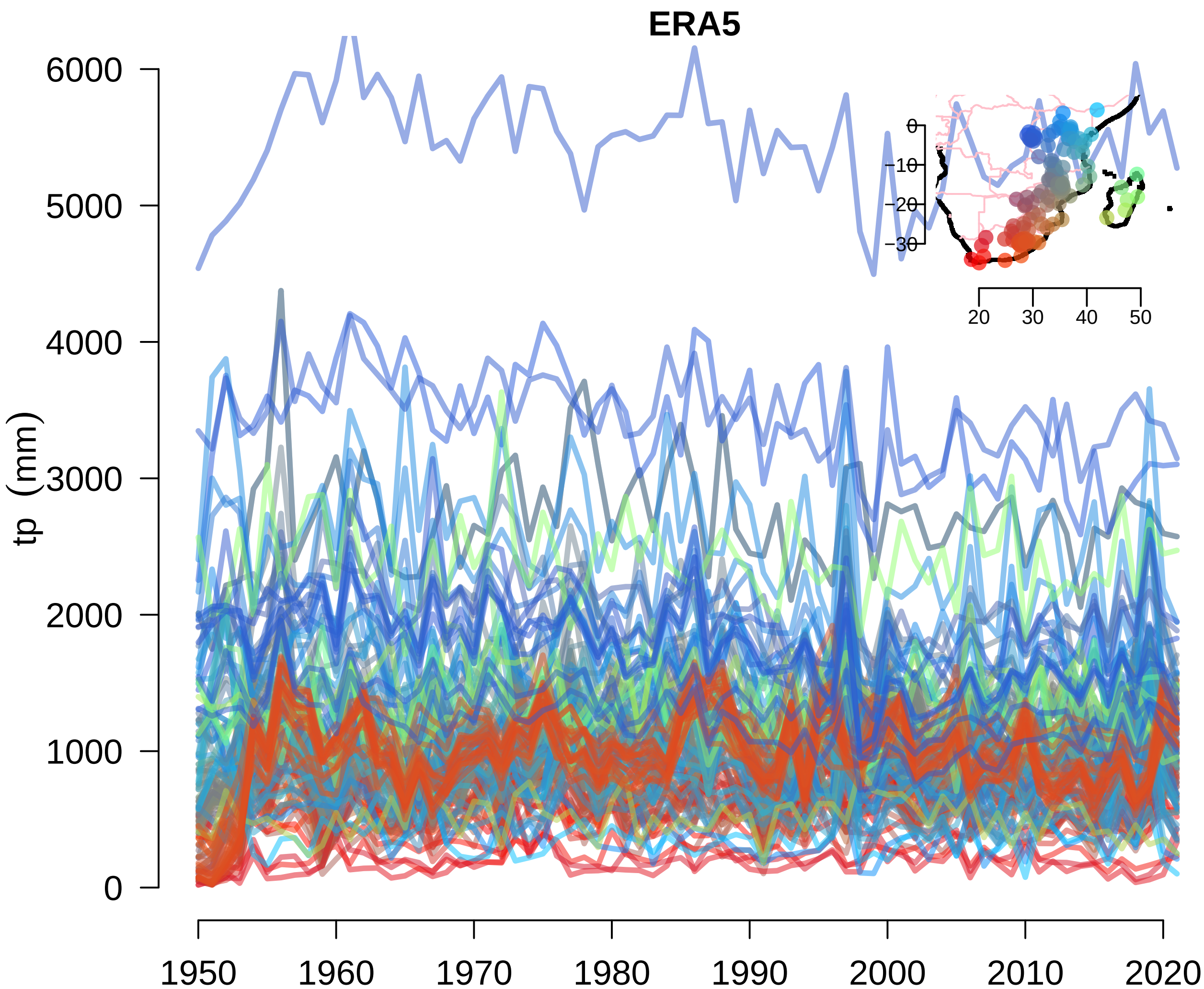
The maps indicate the sites with high rainfall seen in the plots above.

1.2 The ERA5 reanalysis.

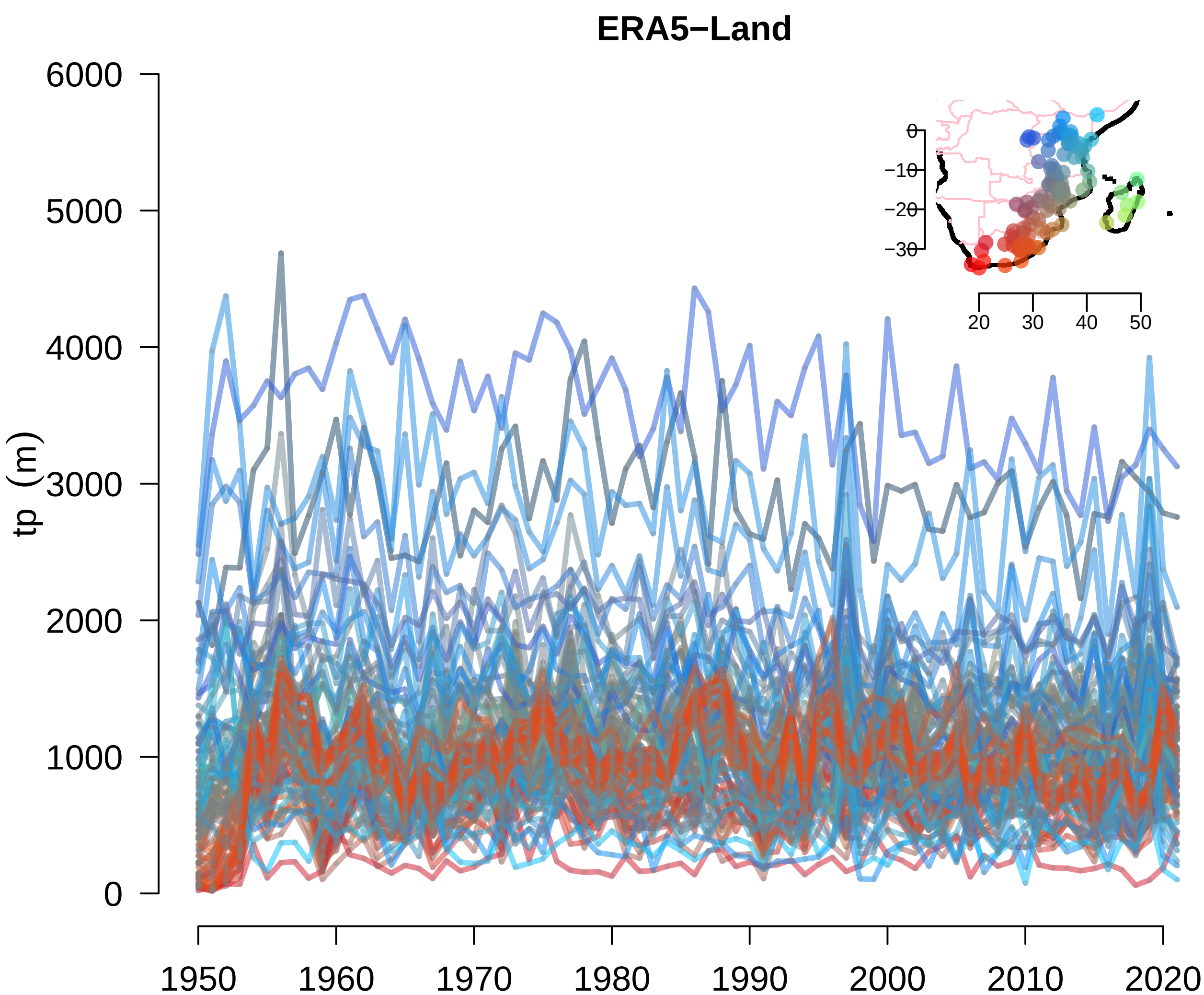
We read ERA5 monthly aggregated rainfall (monthly mean of daily amounts) which subsequently was aggregated to annual rainfall (Oct 1st - Sept. 31st). We based our evaluation on a comparison between the station-based rain gauge data and rainfall data from ERA5 interpolated to the same coordinates as the observational network (using the function ‘regrid()“).

We also took the ERA5-Land reanalysis and interpolate (bi-linear interpolation) to the same coordinates as the rain gauge measurements, but since its data volume is so much greater due to higher spatial resolution, we extract a smaller spatial domain. Hence, it did not cover all the stations (such as Madagascar and southern South Africa), and the comparison between ERA5 and ERA5-Land aggregated total rainfall will only be for those stations that are within the selected region. If ERA5 and ERA5-Land give similar results, then we can proceed with the evaluation of the rain gauge data against ERA5.

```
## Read data from ERA5 reanalysis - here we annually aggregated data that
## already have been processed through CDO - READ MONTHLY DATA
era.data <- 'era5.southeastern-Africa_mon.rda'
lons <- c(17,51); lats <- c(-36,4)
if (!file.exists(era.data)) {
  era5 <- retrieve('~/data/ERA5/ERA5_tp_mon_1950-2022.nc',lon=lons,lat=lats)
  save(era5,file=era.data)
} else load(era.data)
## Also get the ERA5-land data
eraland.data <- 'era5land.southeastern-Africa_mon.nc'
if (!file.exists(eraland.data)) {
  era5.land <- retrieve('~/data/ERA5/ERA5-land_tp_mon_1950-2022.nc',lon=lons,lat=lats)
  attr(era5.land,'dimensions') <-
    c(length(lon(era5.land)),length(lat(era5.land)),length(index(era5.land)))
  x.eraland <- 365.25*1000*annual(regrid(era5.land,is=X),FUN='mean',start=year.start)
  save(x.eraland,file=eraland.data)
} else load(eraland.data)
## Make sure that the monthly values are in mm/year - the original data is in m/day:
x.mon <- 30*1000*regrid(era5,is=X)
## Make sure that the annual values are in mm/year - the original data is in m/day:
era5 <- 365.25*1000*annual(era5,FUN='mean',start=year.start)
attr(era5,'unit') <- 'mm'
x <- regrid(era5,is=X)
attr(x,'unit') <- 'mm'
attr(x,'location') <- loc(X)
attr(x,'country') <- cntr(X)
RR.era5 <- x
attr(x.mon,'unit') <- 'mm'
attr(x.mon,'location') <- loc(X)
attr(x.mon,'country') <- cntr(X)
## Plot the extracted data - interpolated to the same coordinates as the observations:
plot(x,ylim=c(0,6000),new=FALSE,main='ERA5')
```




```
plot(x.eraland,ylim=c(0,6000),new=FALSE,main='ERA5-Land')
```



The ERA5 data extend further back in time (here 1950) than the rain gauges, but there appear to be differences between the years of high rainfall in the rain gauge observations and the reanalysis. Hence, the first impression is that there is not a close match between the two data sets.

For instance, the annual rainfall from ERA5 (1950-2022) did not indicate the 'outlier' local rainfall on Madagascar nor single years when the annual rainfall total spiked such as in Kenya. It's not so easy to tell whether the differences between the two data sources at a small number of sites clutter the figure and give the impression of a mismatch, or if it's the general rule for all sites.

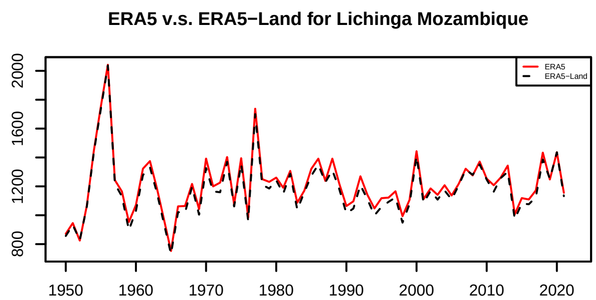
The plot of the extracted annual rainfall from ERA5 doesn't give a good basis for comparison, and we need to do the evaluation on a station-by-station basis where we compare the observations with the reanalysis data interpolated to the sale location.

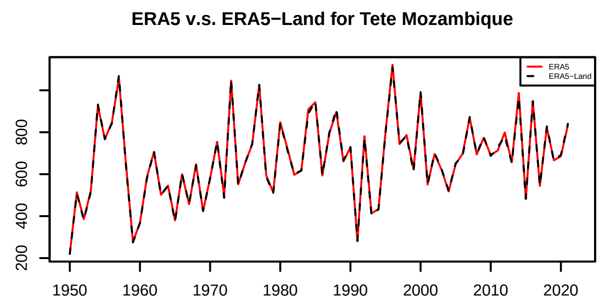
1.3 Compare ERA5 with ERA5 land

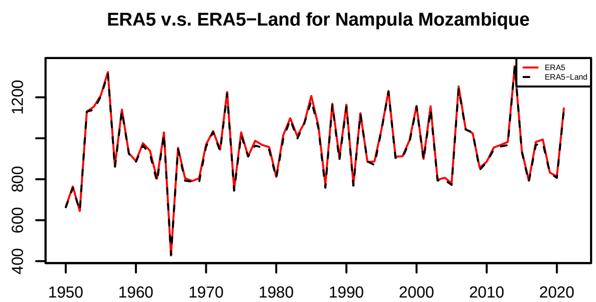
We compare the interpolated reults from ERA5 with ERA5-land

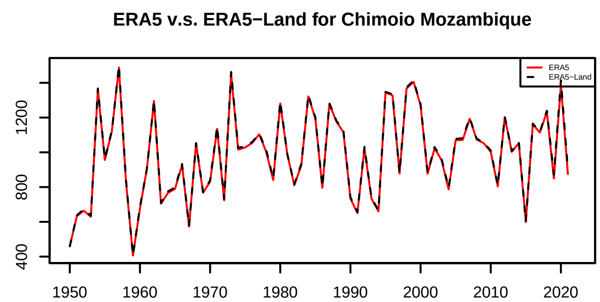
```
ns <- dim(x.eraland)[2]
nv <- apply(x.eraland,2,nv)
x.eraland <- subset(x.eraland,is=nv > 0)
par(mfcol=c(3,2),cex=0.5)
for (loc in loc(x.eraland)) {
  tp.era5 <- subset(x,is=list(loc=loc)); index(tp.era5) <- year(tp.era5)
```

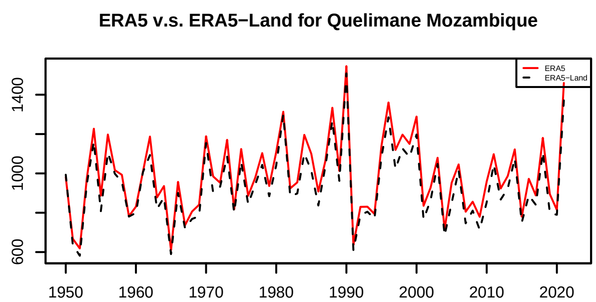
```
  tp.era5land <- subset(x.eraland,is=list(loc=loc)); index(tp.era5land) <- year(tp.era5land)
  if (sum(is.finite(tp.era5land))>30) {
    d <- dim(tp.era5)
    if (d[2]==1) tp.both <- combine.stations(tp.era5,tp.era5land) else
      tp.both <- combine.stations(subset(tp.era5,is=1),subset(tp.era5land,is=1),
                                  subset(tp.era5,is=2),subset(tp.era5land,is=2))
    plot(zoo(tp.both),lty=c(1,2),col=c('red','black'),plot.type='single',ylab='',
         xlab='',main=paste('ERA5 v.s. ERA5-Land for',loc(tp.era5),cntr(tp.era5)))
    legend('topright',c('ERA5','ERA5-Land'),lty=c(1,2),col=c('red','black'),cex=0.5)
  } else print(paste('No plot for',loc(tp.era5),cntr(tp.era5)))
}
```

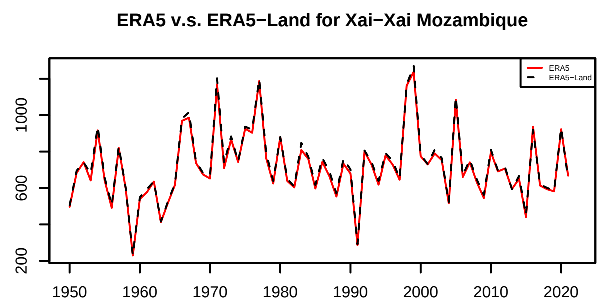




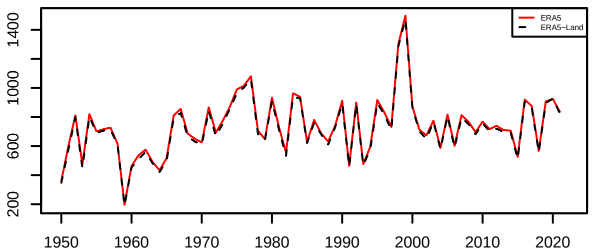




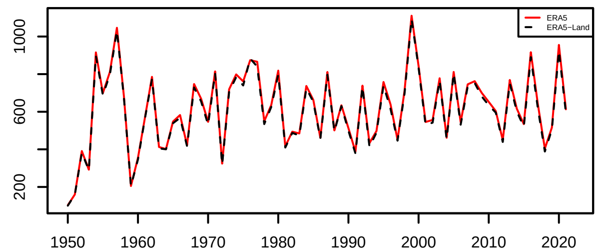




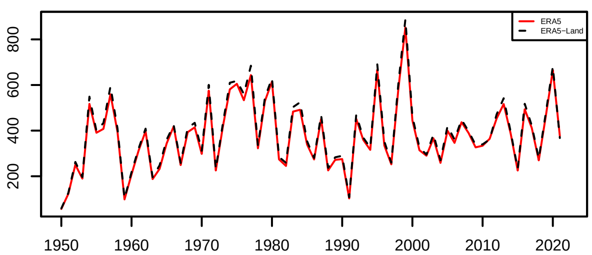
ERA5 v.s. ERA5−Land for MaputoObs Mozambique



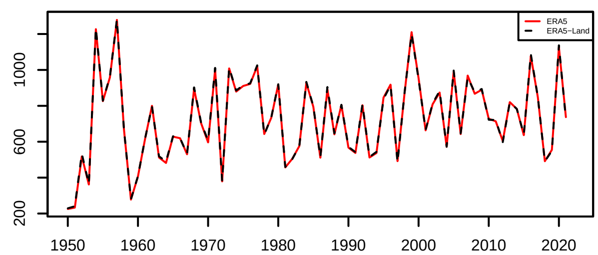
ERA5 v.s. ERA5−Land for Bulawayo Goetz Zimbabwe



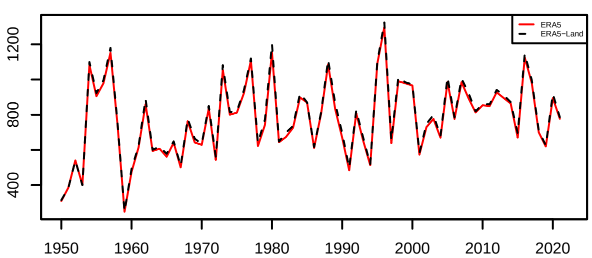
ERA5 v.s. ERA5−Land for Beitbridge Zimbabwe



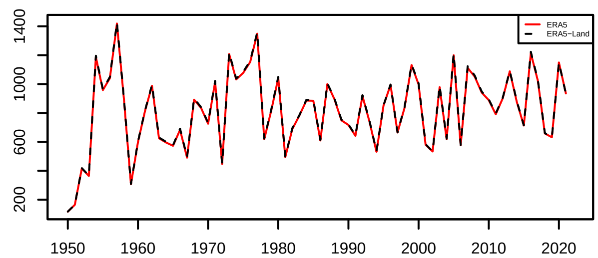
ERA5 v.s. ERA5−Land for Gweru Zimbabwe

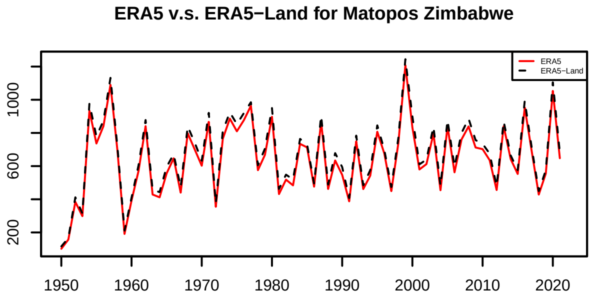


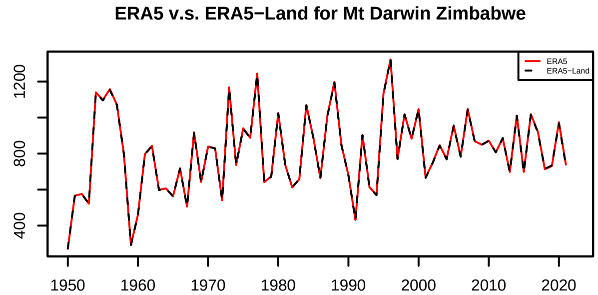
ERA5 v.s. ERA5−Land for Belvedere Zimbabwe

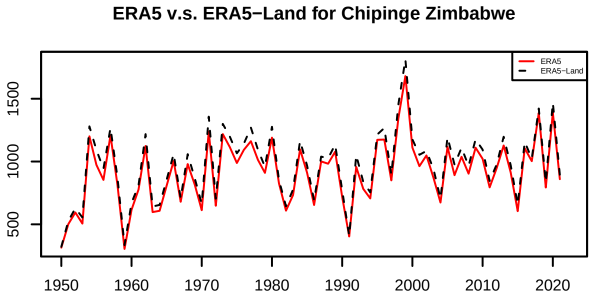


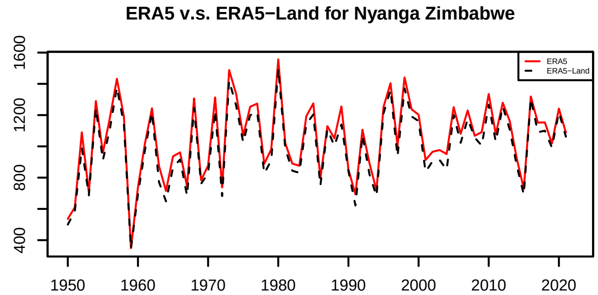
ERA5 v.s. ERA5−Land for Hwange Zimbabwe

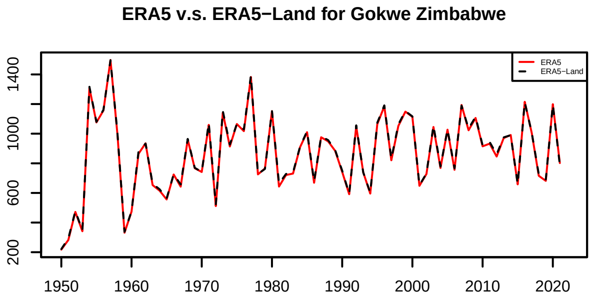


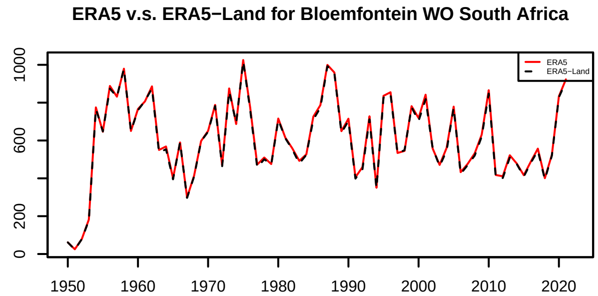

ERA5 v.s. ERA5−Land for Matopos Zimbabwe
ERA5
ERA5−Land
200
600
1000
1950
1960
1970
1980
1990
2000
2010
2020

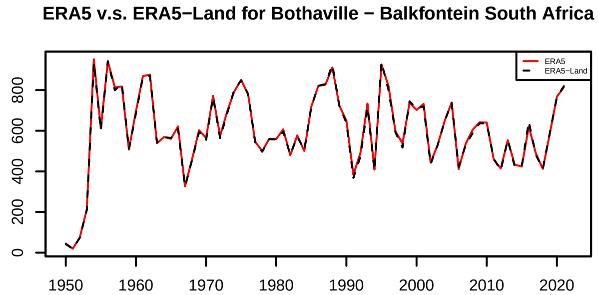

ERA5 v.s. ERA5−Land for Mt Darwin Zimbabwe
ERA5
ERA5−Land
400
800
1200
1950
1960
1970
1980
1990
2000
2010
2020

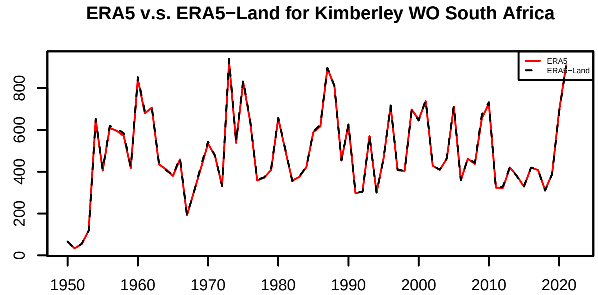

ERA5 v.s. ERA5−Land for Chipinge Zimbabwe
ERA5
ERA5−Land
500
1000
1500
1950
1960
1970
1980
1990
2000
2010
2020

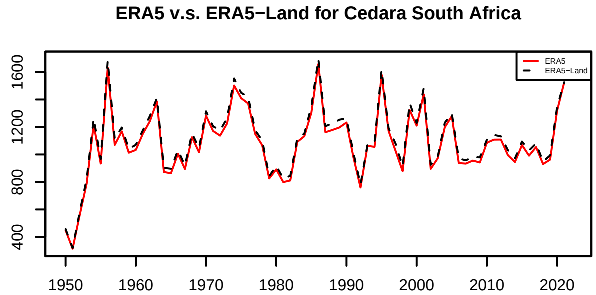

ERA5 v.s. ERA5−Land for Nyanga Zimbabwe
ERA5
ERA5−Land
400
800
1200
1600
1950
1960
1970
1980
1990
2000
2010
2020

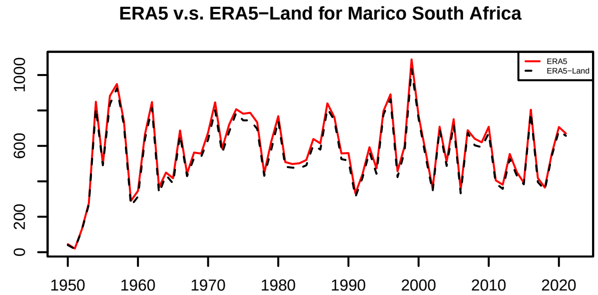

ERA5 v.s. ERA5−Land for Gokwe Zimbabwe
ERA5
ERA5−Land
200
600
1000
1400
1950
1960
1970
1980
1990
2000
2010
2020

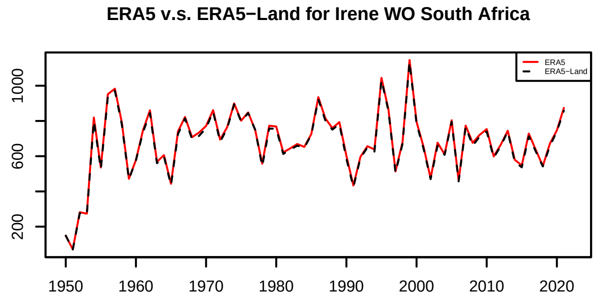

ERA5 v.s. ERA5−Land for Bloemfontein WO South Africa
ERA5
ERA5−Land
0
200
600
1000
1950
1960
1970
1980
1990
2000
2010
2020

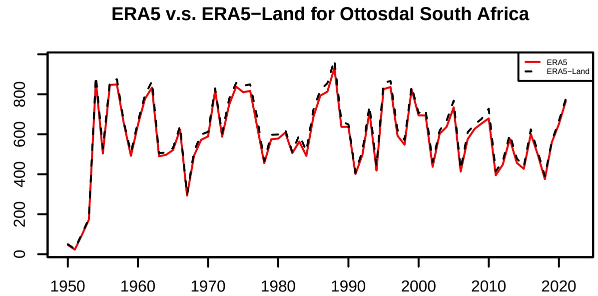

ERA5 v.s. ERA5−Land for Bothaville − Balkfontein South Africa
ERA5
ERA5−Land
0 200 400 600 800
1950 1960 1970 1980 1990 2000 2010 2020

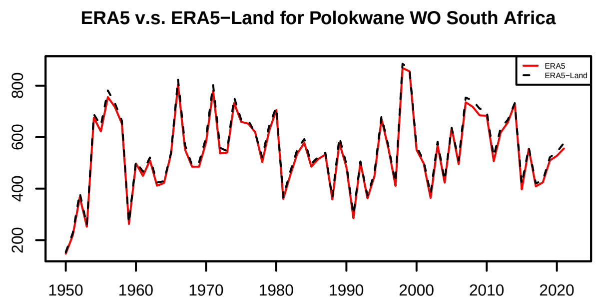

ERA5 v.s. ERA5−Land for Kimberley WO South Africa
ERA5
ERA5−Land
0 200 400 600 800
1950 1960 1970 1980 1990 2000 2010 2020

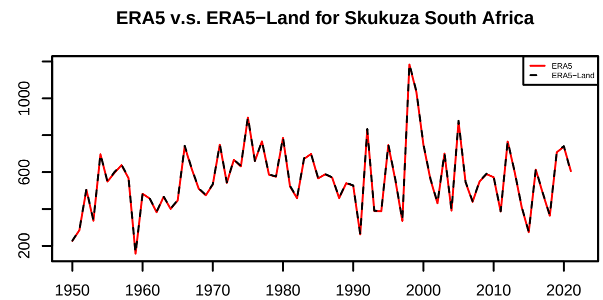

ERA5 v.s. ERA5−Land for Cedara South Africa
ERA5
ERA5−Land
400 800 1200 1600
1950 1960 1970 1980 1990 2000 2010 2020

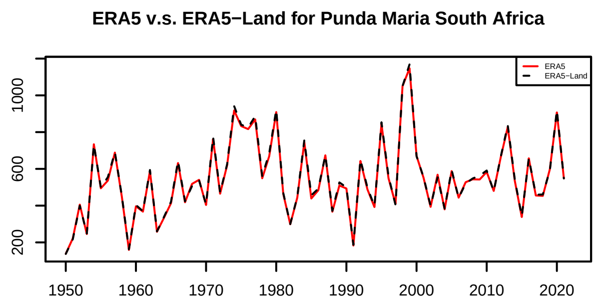

ERA5 v.s. ERA5−Land for Marico South Africa
ERA5
ERA5−Land
0 200 600 1000
1950 1960 1970 1980 1990 2000 2010 2020

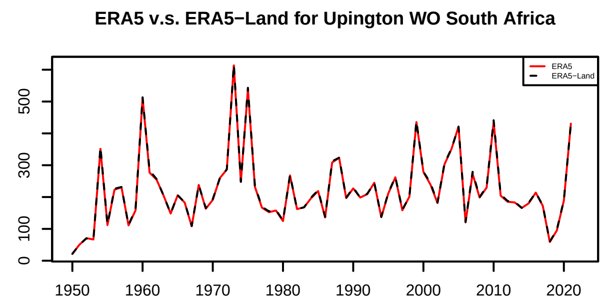

ERA5 v.s. ERA5−Land for Irene WO South Africa
ERA5
ERA5−Land
200 600 1000
1950 1960 1970 1980 1990 2000 2010 2020

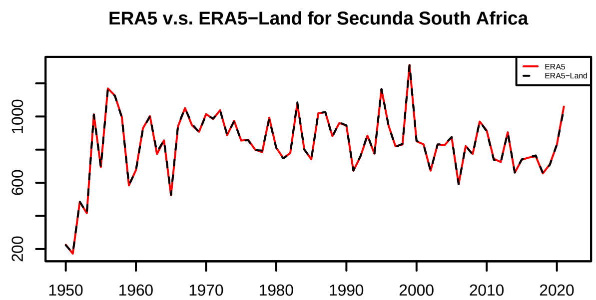

ERA5 v.s. ERA5−Land for Ottosdal South Africa
ERA5
ERA5−Land
0 200 400 600 800
1950 1960 1970 1980 1990 2000 2010 2020

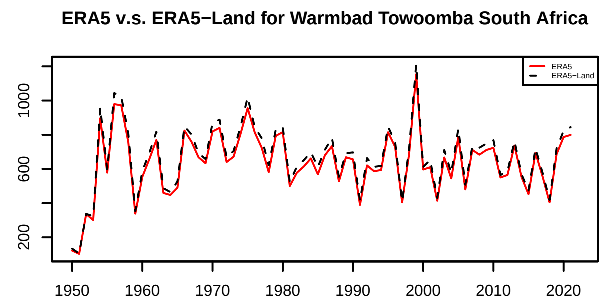

ERA5 v.s. ERA5−Land for Polokwane WO South Africa
200 400 600 800
1950 1960 1970 1980 1990 2000 2010 2020
ERA5
ERA5−Land


ERA5 v.s. ERA5−Land for Skukuza South Africa
200 600 1000
1950 1960 1970 1980 1990 2000 2010 2020
ERA5
ERA5−Land

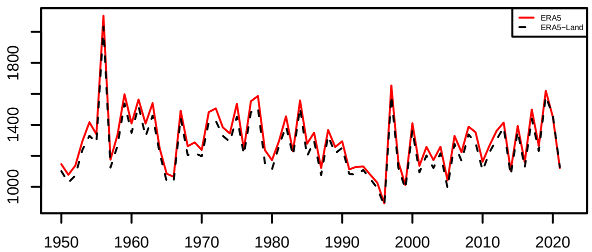

ERA5 v.s. ERA5−Land for Punda Maria South Africa
200 600 1000
1950 1960 1970 1980 1990 2000 2010 2020
ERA5
ERA5−Land


ERA5 v.s. ERA5−Land for Upington WO South Africa
0 100 300 500
1950 1960 1970 1980 1990 2000 2010 2020
ERA5
ERA5−Land

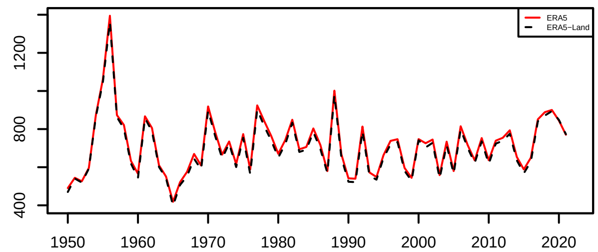

ERA5 v.s. ERA5−Land for Secunda South Africa
200 600 1000
1950 1960 1970 1980 1990 2000 2010 2020
ERA5
ERA5−Land


ERA5 v.s. ERA5−Land for Warmbad Towoomba South Africa
200 600 1000
1950 1960 1970 1980 1990 2000 2010 2020
ERA5
ERA5−Land

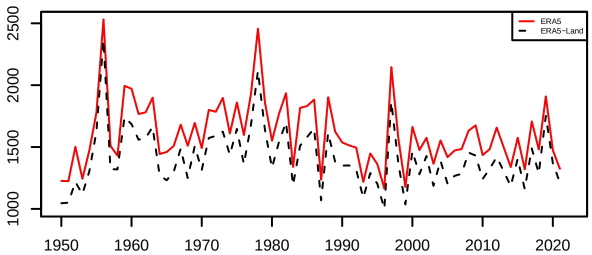
ERA5 v.s. ERA5−Land for Chitipa Malawi



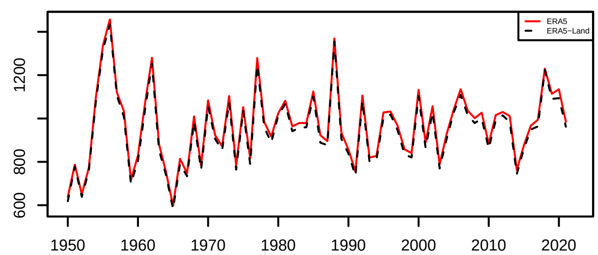
ERA5 v.s. ERA5−Land for Kasungu Malawi



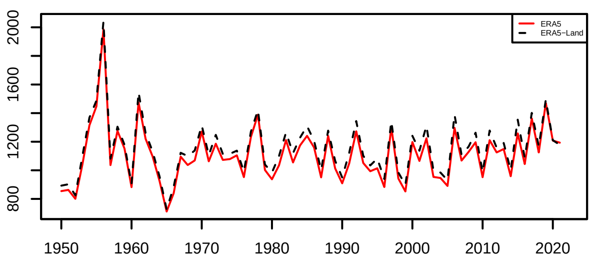
ERA5 v.s. ERA5−Land for Karonga Malawi



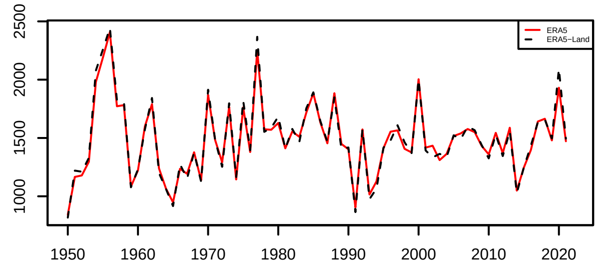
ERA5 v.s. ERA5−Land for KIA Malawi



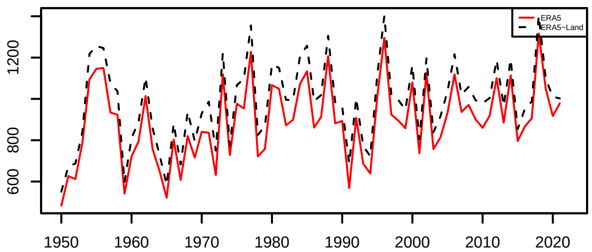
ERA5 v.s. ERA5−Land for Mzimba Malawi



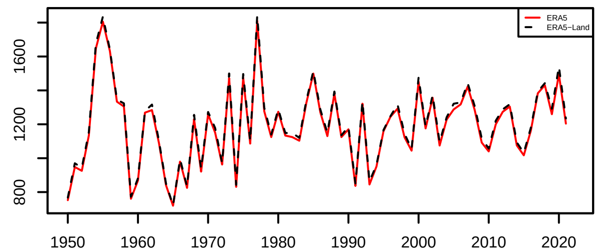
ERA5 v.s. ERA5−Land for Salima Malawi



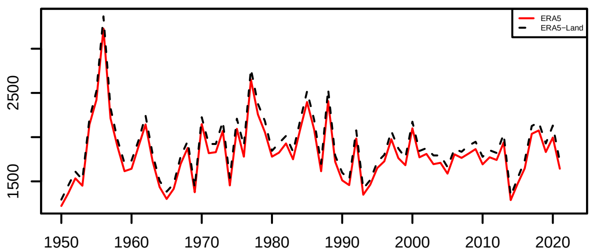
ERA5 v.s. ERA5−Land for Makoka Malawi



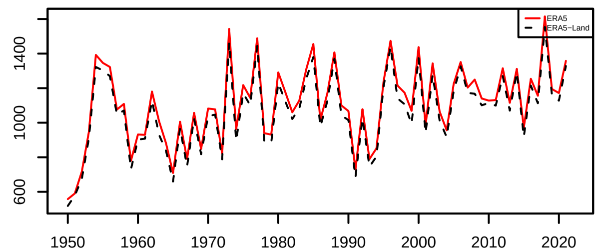
ERA5 v.s. ERA5−Land for Mangochi Malawi



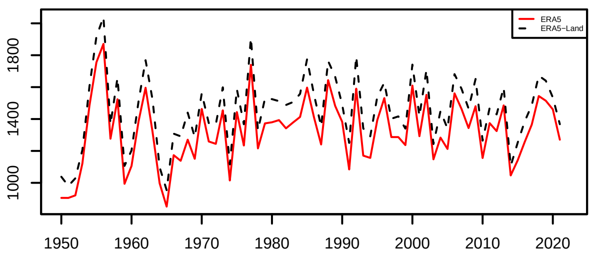
ERA5 v.s. ERA5−Land for Nkhotakota Malawi



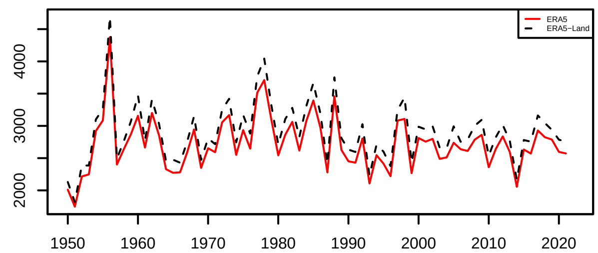
ERA5 v.s. ERA5−Land for Chileka Malawi



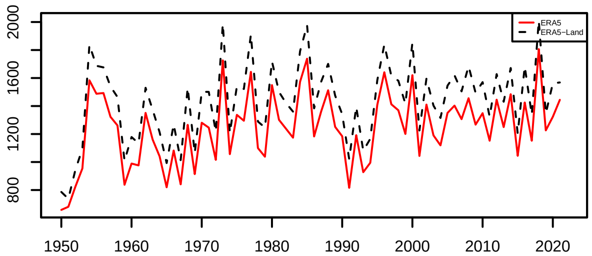
ERA5 v.s. ERA5−Land for Dedza Malawi



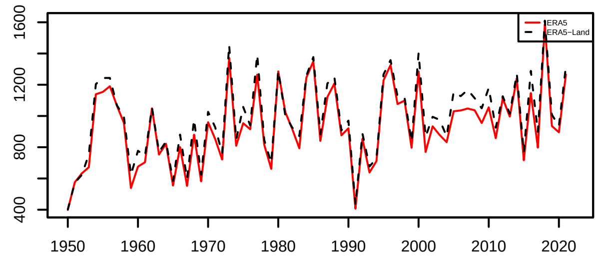
ERA5 v.s. ERA5−Land for Nkhatabay Malawi



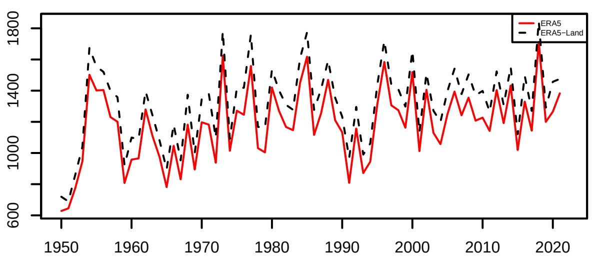
ERA5 v.s. ERA5−Land for Bvumbwe Malawi



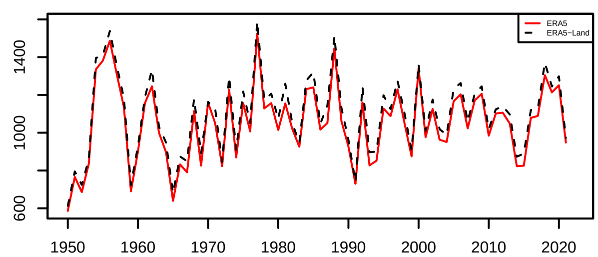
ERA5 v.s. ERA5−Land for Chikwawa Malawi



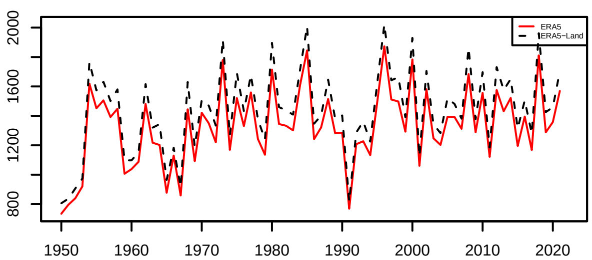
ERA5 v.s. ERA5−Land for Chichiri Malawi



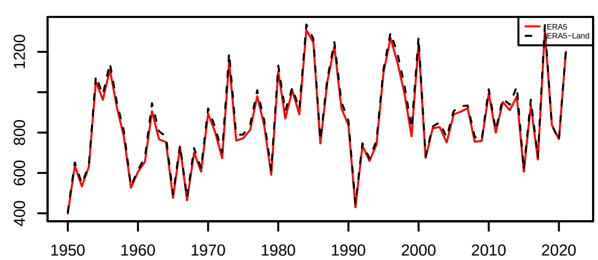
ERA5 v.s. ERA5−Land for Mchinji Malawi

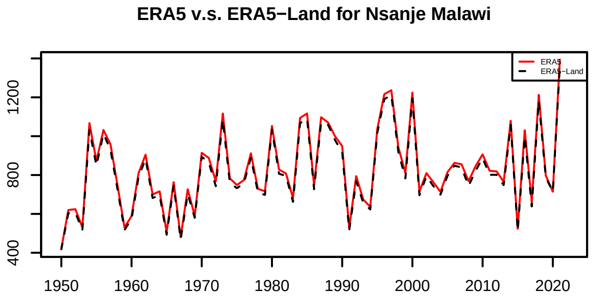


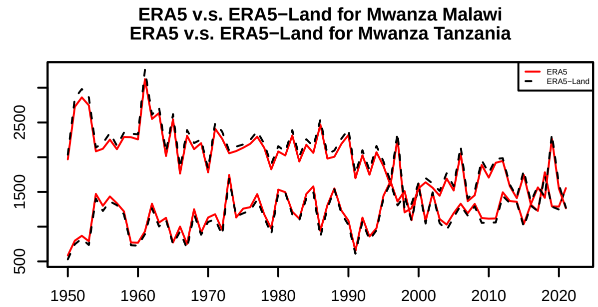
ERA5 v.s. ERA5−Land for Mimosa Malawi

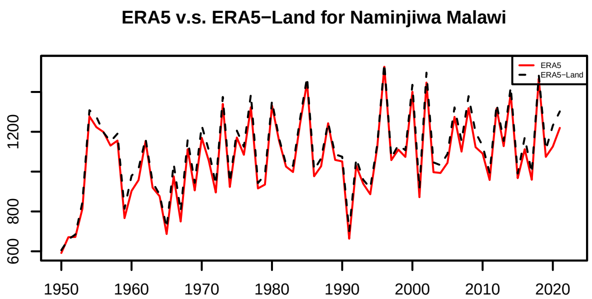


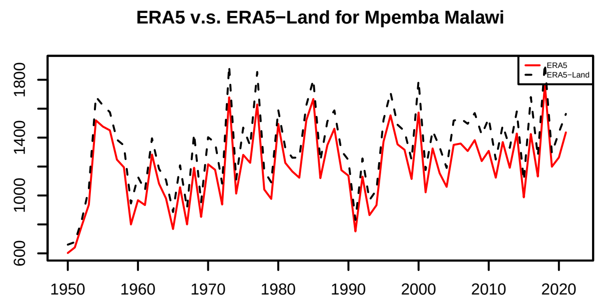
ERA5 v.s. ERA5−Land for Makhanga Malawi

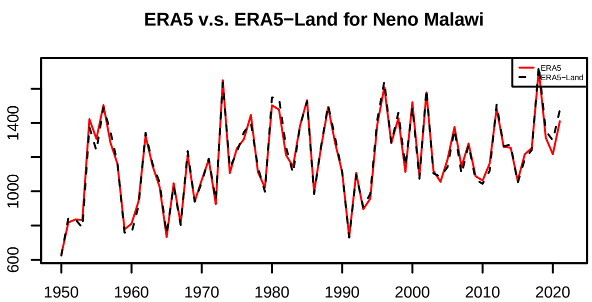


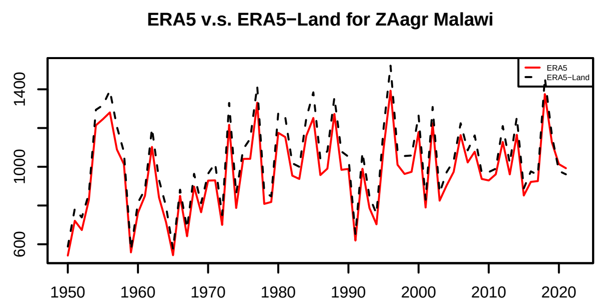

ERA5 v.s. ERA5−Land for Nsanje Malawi
ERA5
ERA5−Land
400
800
1200
1950
1960
1970
1980
1990
2000
2010
2020

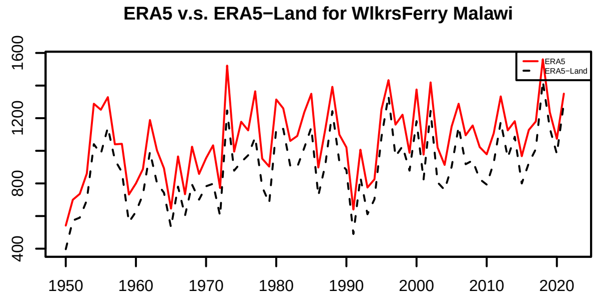

ERA5 v.s. ERA5−Land for Mwanza Malawi
ERA5 v.s. ERA5−Land for Mwanza Tanzania
ERA5
ERA5−Land
500
1500
2500
1950
1960
1970
1980
1990
2000
2010
2020

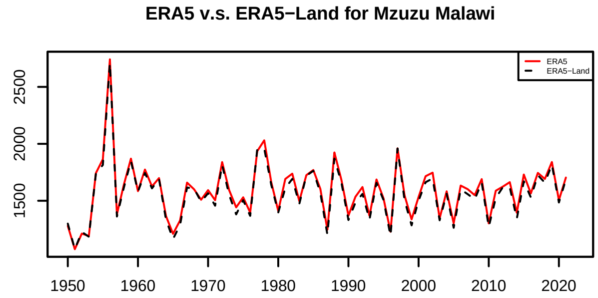

ERA5 v.s. ERA5−Land for Naminjiwa Malawi
ERA5
ERA5−Land
600
800
1200
1950
1960
1970
1980
1990
2000
2010
2020

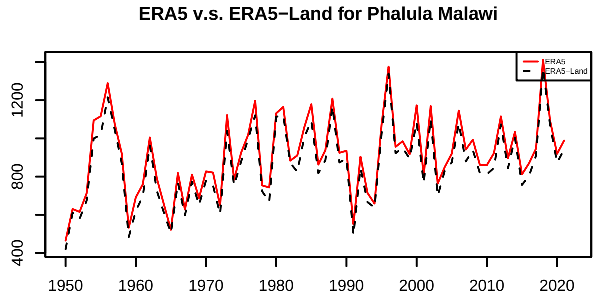

ERA5 v.s. ERA5−Land for Mpemba Malawi
ERA5
ERA5−Land
600
1000
1400
1800
1950
1960
1970
1980
1990
2000
2010
2020

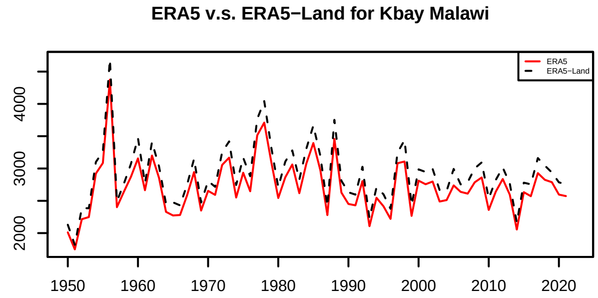

ERA5 v.s. ERA5−Land for Neno Malawi
ERA5
ERA5−Land
600
1000
1400
1950
1960
1970
1980
1990
2000
2010
2020

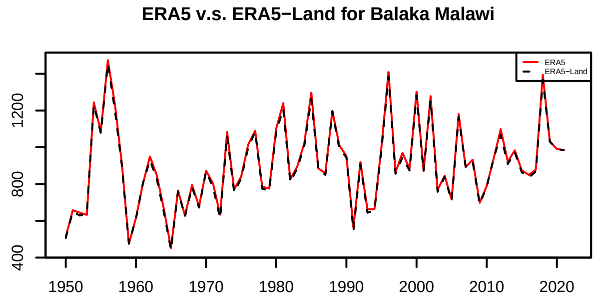

ERA5 v.s. ERA5−Land for ZAagr Malawi
ERA5
ERA5−Land
600
1000
1400
1950
1960
1970
1980
1990
2000
2010
2020

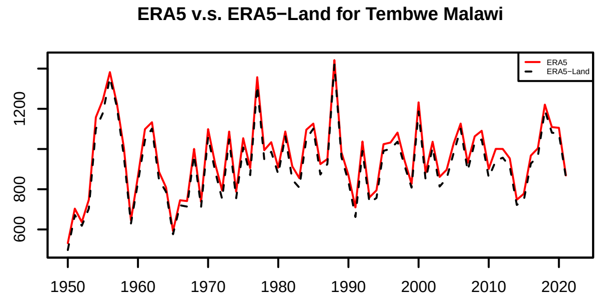

ERA5 v.s. ERA5−Land for WlkrsFerry Malawi
ERA5
ERA5−Land

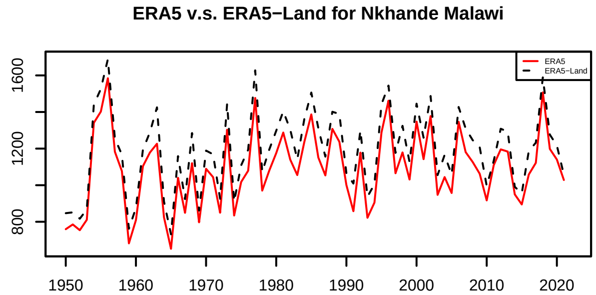

ERA5 v.s. ERA5−Land for Mzuzu Malawi
ERA5
ERA5−Land

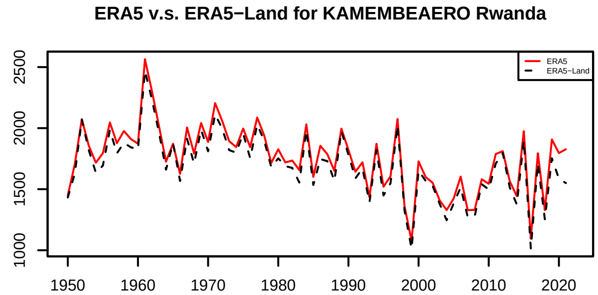

ERA5 v.s. ERA5−Land for Phalula Malawi
ERA5
ERA5−Land

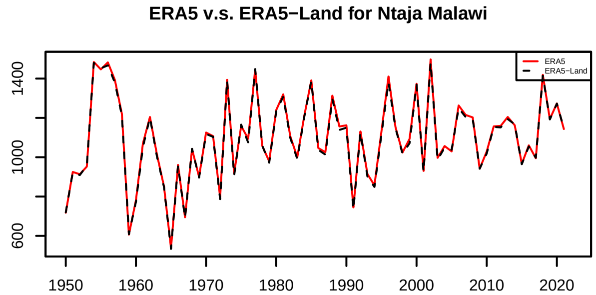

ERA5 v.s. ERA5−Land for Kbay Malawi
ERA5
ERA5−Land

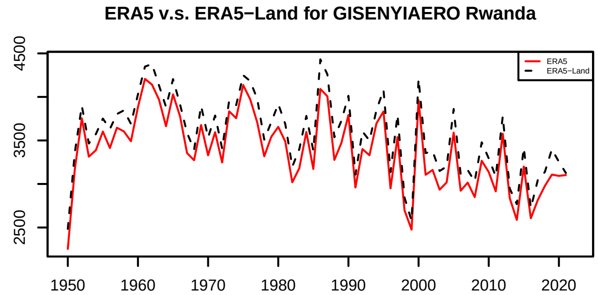

ERA5 v.s. ERA5−Land for Balaka Malawi
ERA5
ERA5−Land

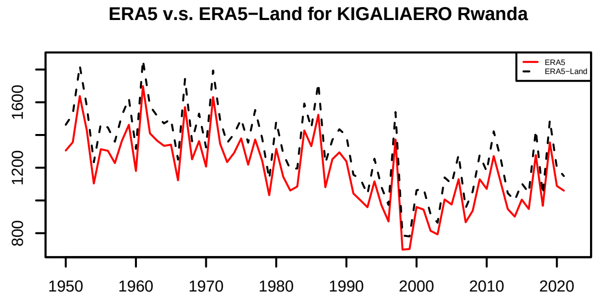

ERA5 v.s. ERA5−Land for Tembwe Malawi
ERA5
ERA5−Land

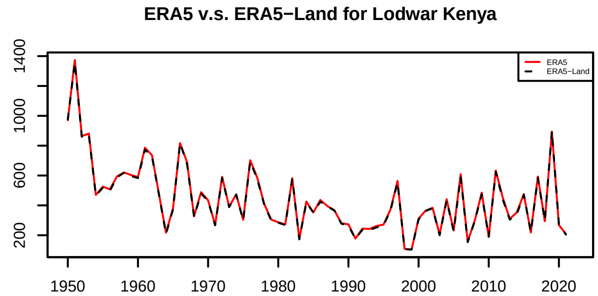

ERA5 v.s. ERA5−Land for Nkhande Malawi
ERA5
ERA5−Land
800
1200
1600
1950
1960
1970
1980
1990
2000
2010
2020


ERA5 v.s. ERA5−Land for KAMEMBEAERO Rwanda
ERA5
ERA5−Land
1000
1500
2000
2500
1950
1960
1970
1980
1990
2000
2010
2020

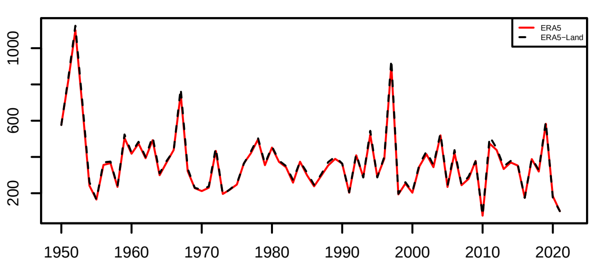

ERA5 v.s. ERA5−Land for Ntaja Malawi
ERA5
ERA5−Land
600
1000
1400
1950
1960
1970
1980
1990
2000
2010
2020


ERA5 v.s. ERA5−Land for GISENYIAERO Rwanda
ERA5
ERA5−Land
2500
3500
4500
1950
1960
1970
1980
1990
2000
2010
2020

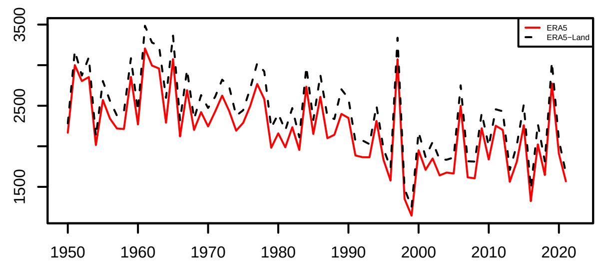

ERA5 v.s. ERA5−Land for KIGALIAERO Rwanda
ERA5
ERA5−Land
800
1200
1600
1950
1960
1970
1980
1990
2000
2010
2020


ERA5 v.s. ERA5−Land for Lodwar Kenya
ERA5
ERA5−Land
200
600
1000
1400
1950
1960
1970
1980
1990
2000
2010
2020

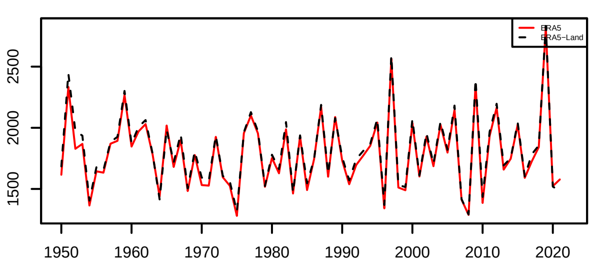
ERA5 v.s. ERA5−Land for Mandera Kenya



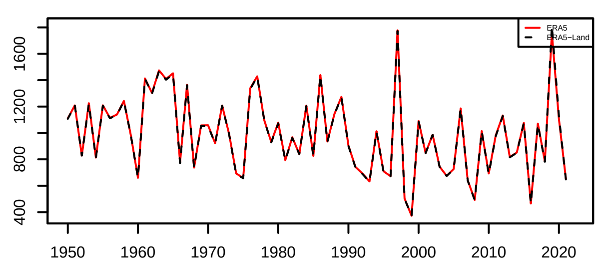
ERA5 v.s. ERA5−Land for Kisii Kenya



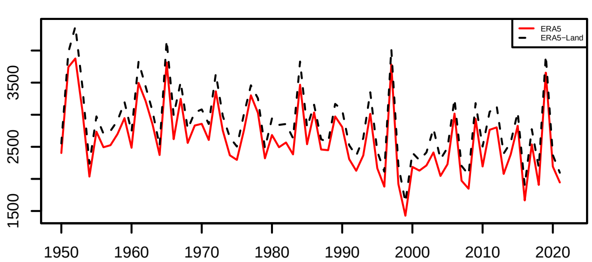
ERA5 v.s. ERA5−Land for Kitale Kenya



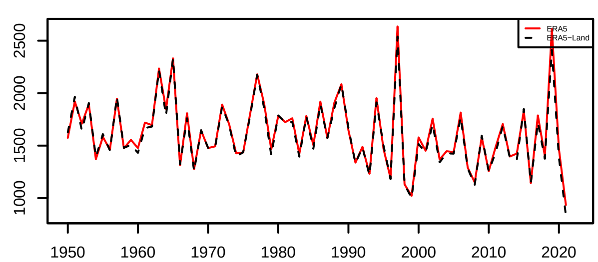
ERA5 v.s. ERA5−Land for Narok Kenya



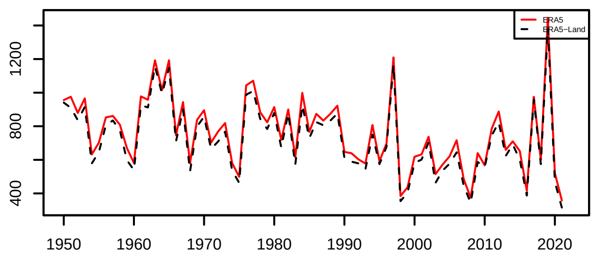
ERA5 v.s. ERA5−Land for Kericho Kenya



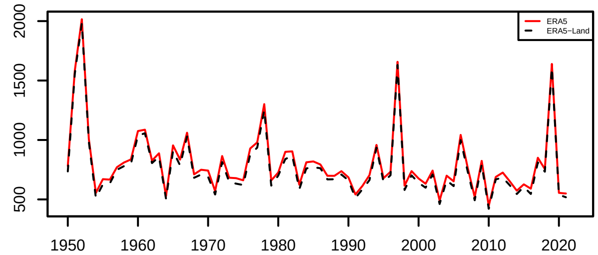
ERA5 v.s. ERA5−Land for Nyeri Kenya



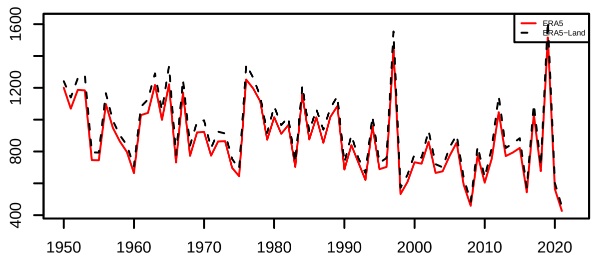
ERA5 v.s. ERA5−Land for Dagoretti Corner Kenya



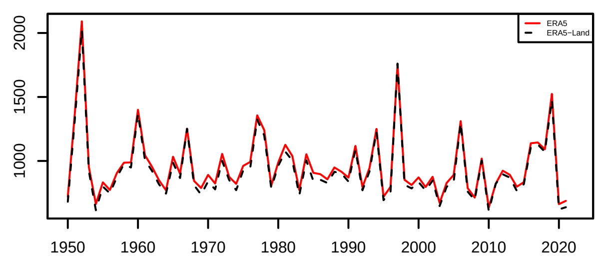
ERA5 v.s. ERA5−Land for Lamu Kenya



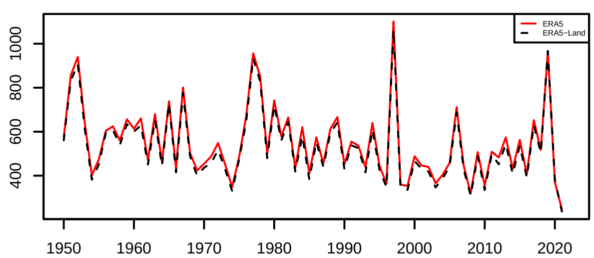
ERA5 v.s. ERA5−Land for Machakos Agromet Kenya



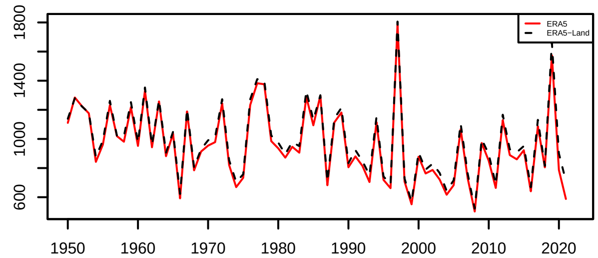
ERA5 v.s. ERA5−Land for Moi International Airpor Kenya



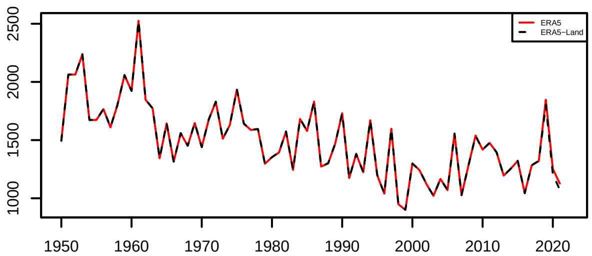
ERA5 v.s. ERA5−Land for Voi Kenya



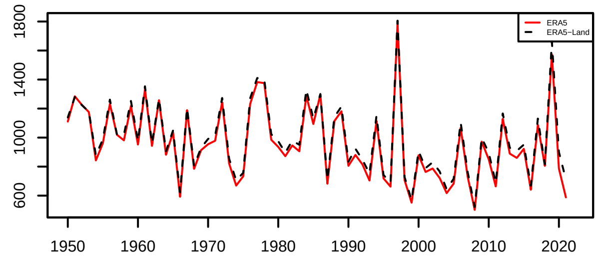
ERA5 v.s. ERA5−Land for Bukoba Tanzania



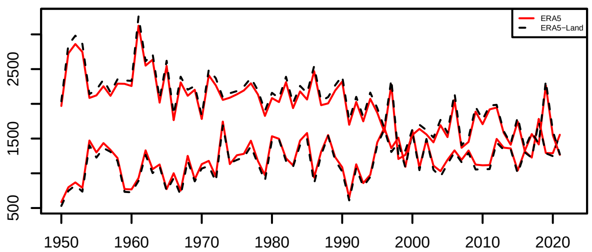
ERA5 v.s. ERA5−Land for Musoma Tanzania



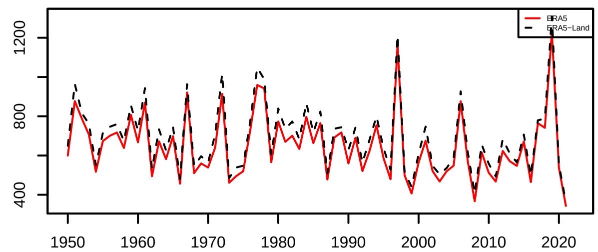
ERA5 v.s. ERA5−Land for Arusha Tanzania



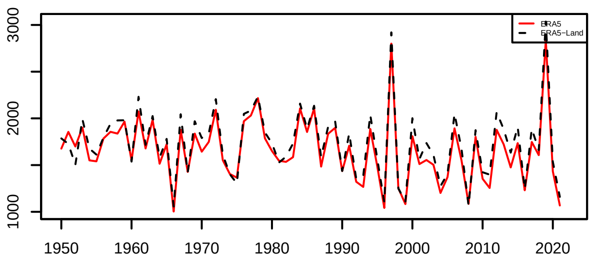
ERA5 v.s. ERA5−Land for Mwanza Malawi
ERA5 v.s. ERA5−Land for Mwanza Tanzania



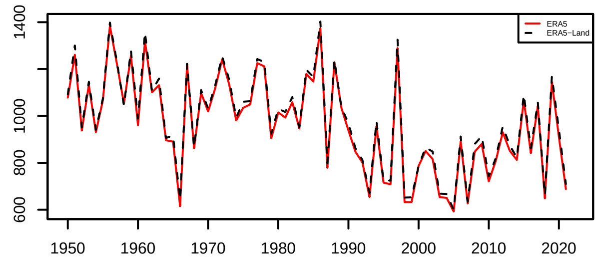
ERA5 v.s. ERA5−Land for Same Tanzania



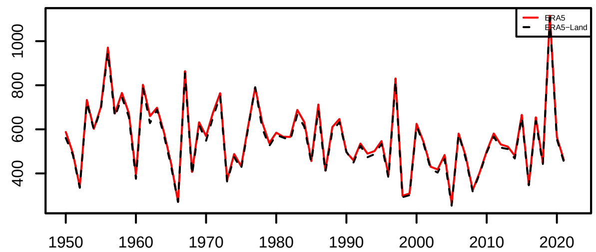
ERA5 v.s. ERA5−Land for Moshi Tanzania



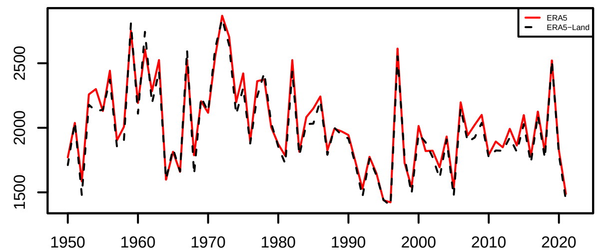
ERA5 v.s. ERA5−Land for Tabora Tanzania



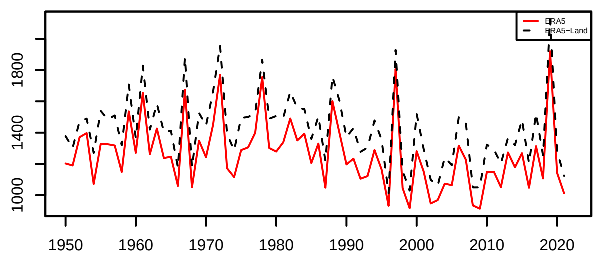
ERA5 v.s. ERA5−Land for Dodoma Tanzania



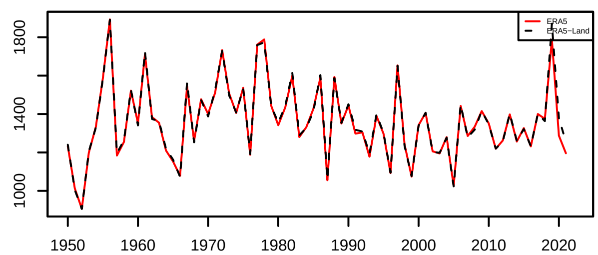
ERA5 v.s. ERA5−Land for Mbeya Tanzania



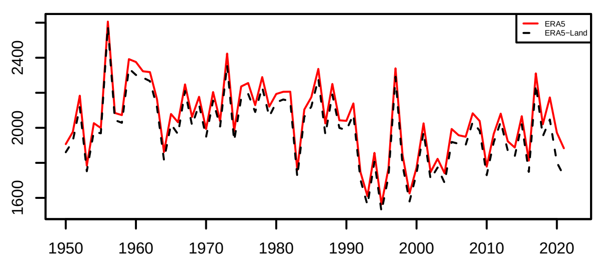
ERA5 v.s. ERA5−Land for Morogoro Tanzania



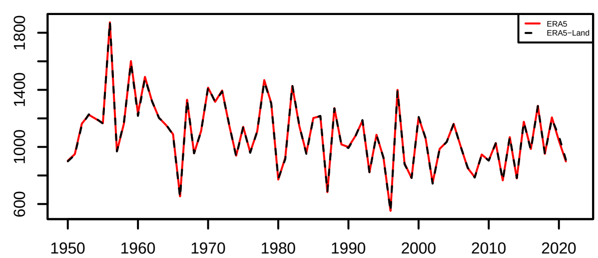
ERA5 v.s. ERA5−Land for Songea Tanzania

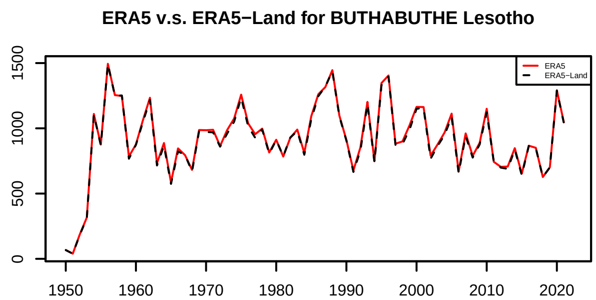


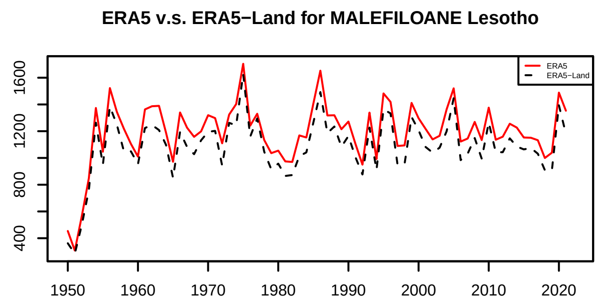
ERA5 v.s. ERA5−Land for Sumbawanga Tanzania

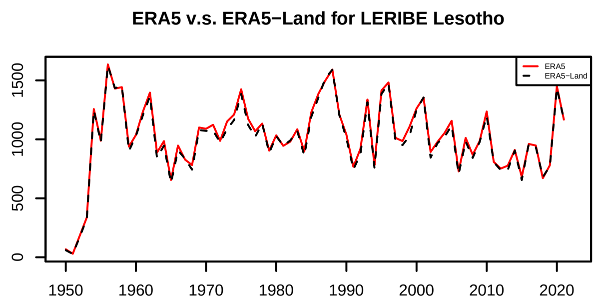


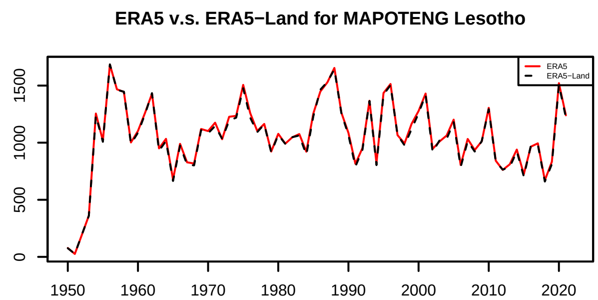
ERA5 v.s. ERA5−Land for Mtwara Tanzania

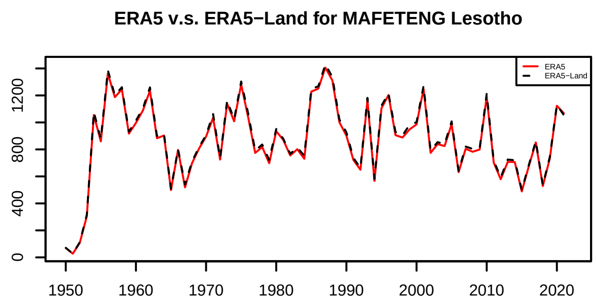


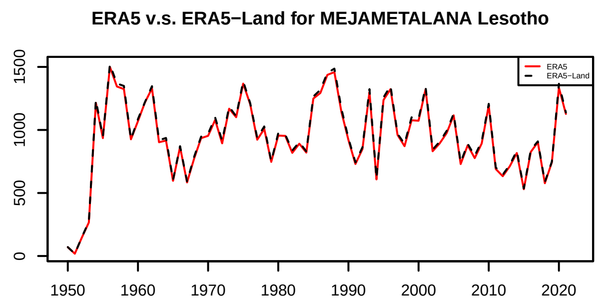

ERA5 v.s. ERA5−Land for BUTHABUTHE Lesotho
ERA5
ERA5−Land

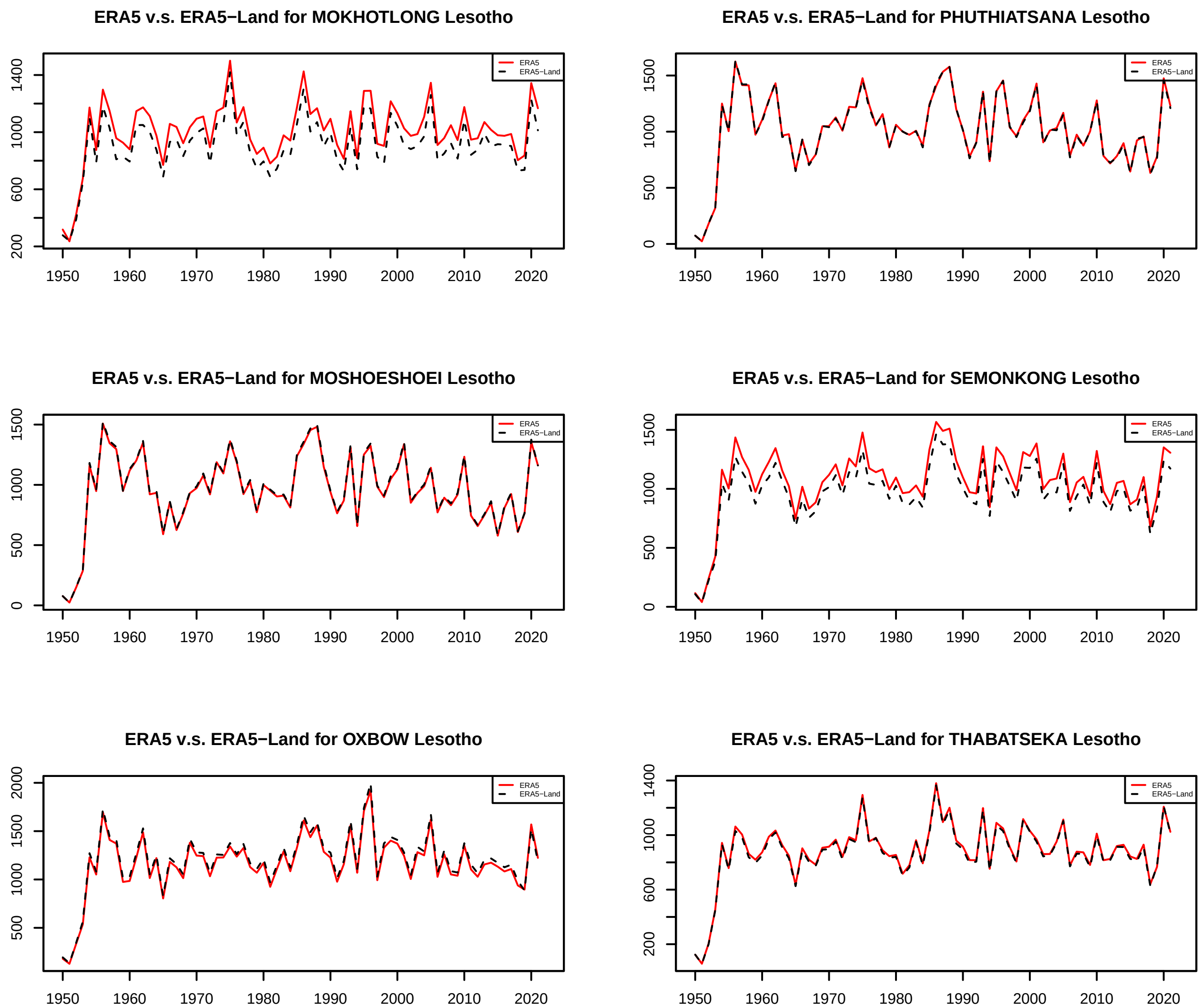

ERA5 v.s. ERA5−Land for MALEFILOANE Lesotho
ERA5
ERA5−Land


ERA5 v.s. ERA5−Land for LERIBE Lesotho
ERA5
ERA5−Land


ERA5 v.s. ERA5−Land for MAPOTENG Lesotho
ERA5
ERA5−Land


ERA5 v.s. ERA5−Land for MAFETENG Lesotho
ERA5
ERA5−Land


ERA5 v.s. ERA5−Land for MEJAMETALANA Lesotho
ERA5
ERA5−Land



The interpolated annual temperatures for the same sites as for the rain gauge measurements were in generally a good agreement between ERA5 and ERA5-Land, but there were some exceptions where the two reanalyses differed. Nevertheless, the results of the comparison indicated that an evaluation of rain gauge data against the ERA5 reanalysis will not be very different from an evaluation against ERA5-Land.

1.4 CHIRPS

Read the CHIRPS data and interpolate to the location of the rain gauges

```
filename <- '~/R/seafrica-chirps-xtract.rain.rda'
if (!file.exists(filename)) {
  lons <- range(lon(X)) + c(-1,1); lats <- range(lat(X)) + c(-1,1)
  fname <- '~/Downloads/CHIRPS_total_precipitation_day_0.25x0.25_africa_1989-2019_v2.0.nc'
  chirps <- retrieve(fname,lon=lons,lat=lats)
  #index(chirps) <- as.Date(index(chirps))
  index(chirps) <- seq(as.Date('1981-01-01'),as.Date('2019-12-31'),by=1)
  rr.chirps <- regrid(chirps,is=X)
  rr.ghcnd.chirps <- regrid(chirps,is=Y.ghcnd)
  plot(rr.chirps,new=FALSE)
  chirps5map <- map(annual(chirps,FUN='sum',start=year.start),
                    FUN='mean',colbar=list(breaks=breaks,pal='precip.ipcc'),
```

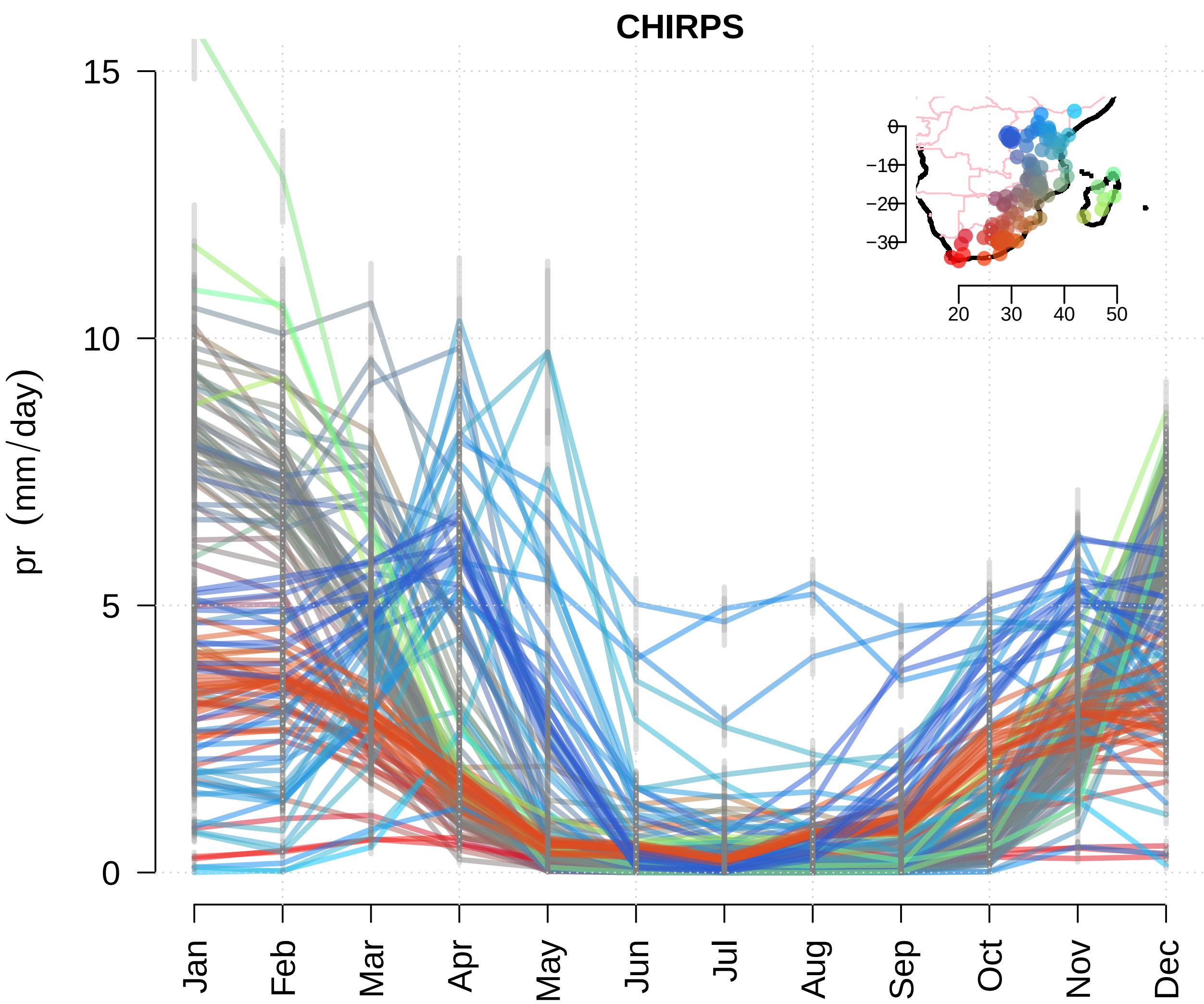
```
                 type='fill',new=FALSE,showaxis=FALSE,
                 main='October-September total rainfall (CHIRPS/rain gauges)')
  save(rr.chirps,rr.ghcnd.chirps,chirps5map,file=filename)
} else load(filename)
```

Inspect the rain seasons derived from CHIRPS.

```
plot(aggregate(rr.chirps,by=month,FUN='mean'),main='CHIRPS',ylim=c(0,15),new=FALSE)
```

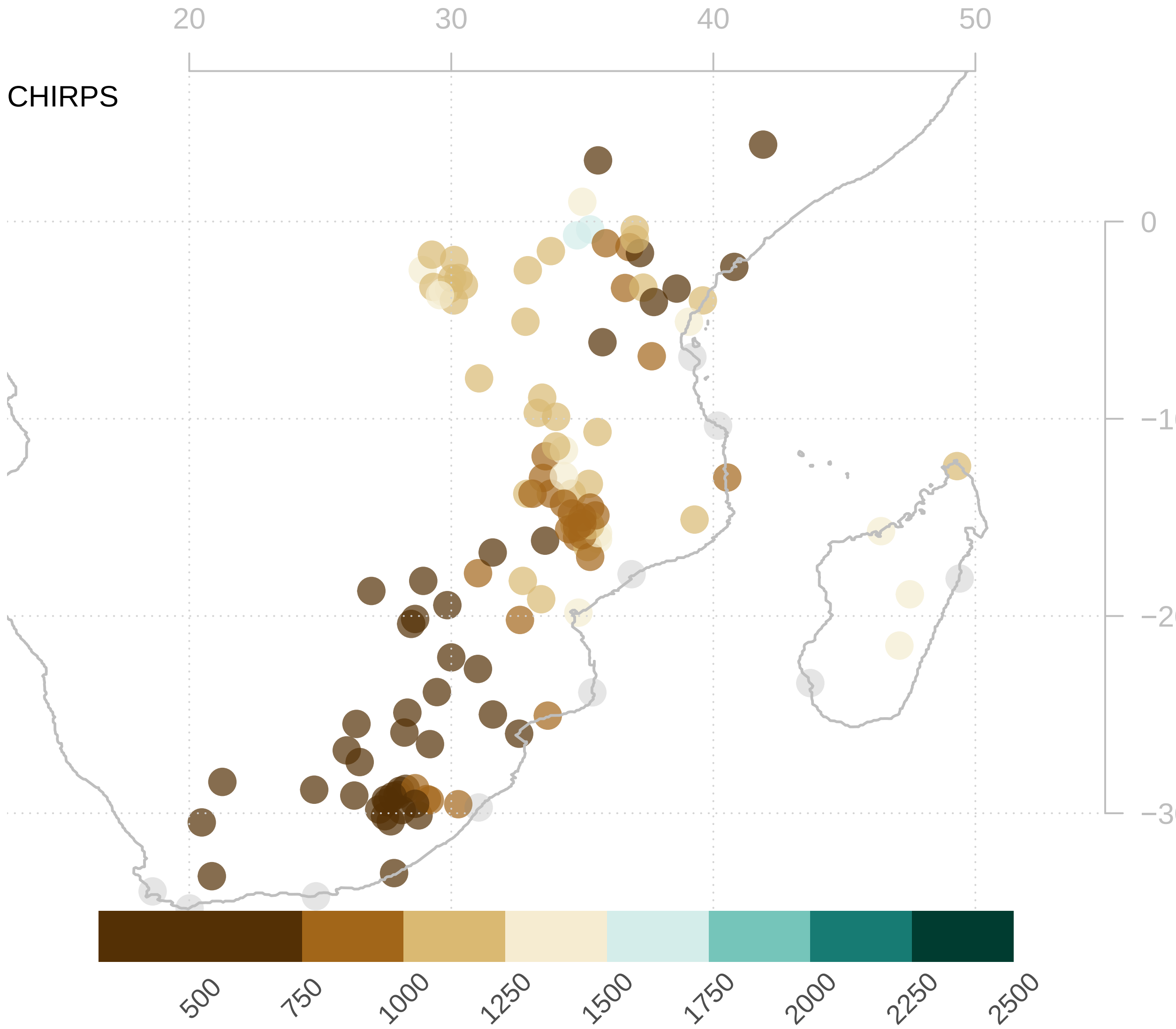
```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
grid()
```



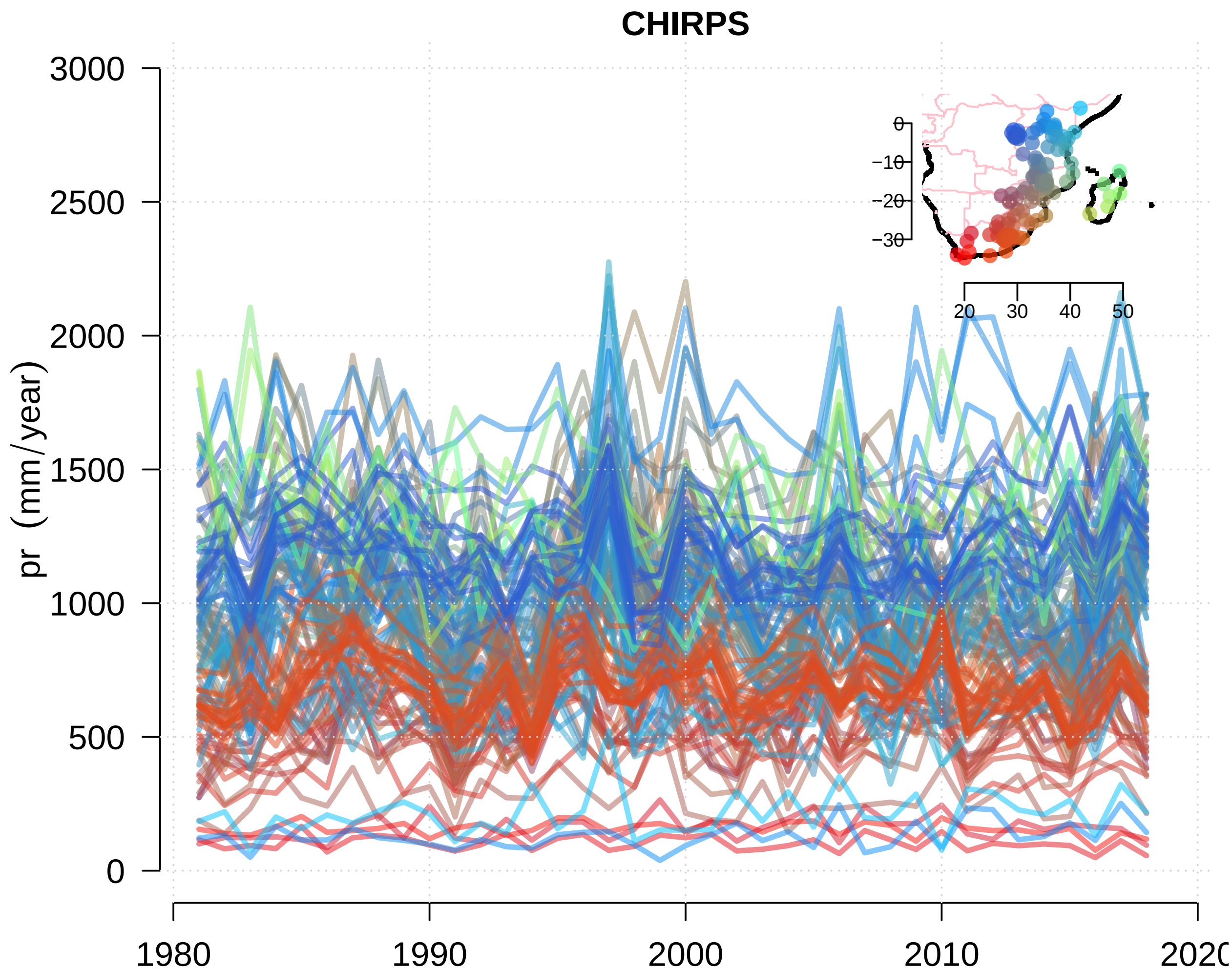
Inspect the geographical distribution of the mean annual rainfall

```
RR.chirps <- annual(rr.chirps,FUN='sum',start=year.start)
map(RR.chirps,FUN='mean',main='CHIRPS',colbar=list(breaks=seq(500,2500,by=250),
                                                   pal='precip.ipcc'),new=FALSE)
```



The grey symbols show missing data, as CHIPS is defined on land only and the interpolation fails if the sites are too close to the sea.

```
plot(RR.chirps,main='CHIRPS',new=FALSE,ylim=c(0,3000))
grid()
```




```
print(range(index(RR.chirps)))
```

```
## [1] 1981 2019
```

2 Results

2.1 Evaluation: a comparison between ERA5 reanalysis and in-situ rain gauge data

2.1.1 The mean annual cycle in rainfall

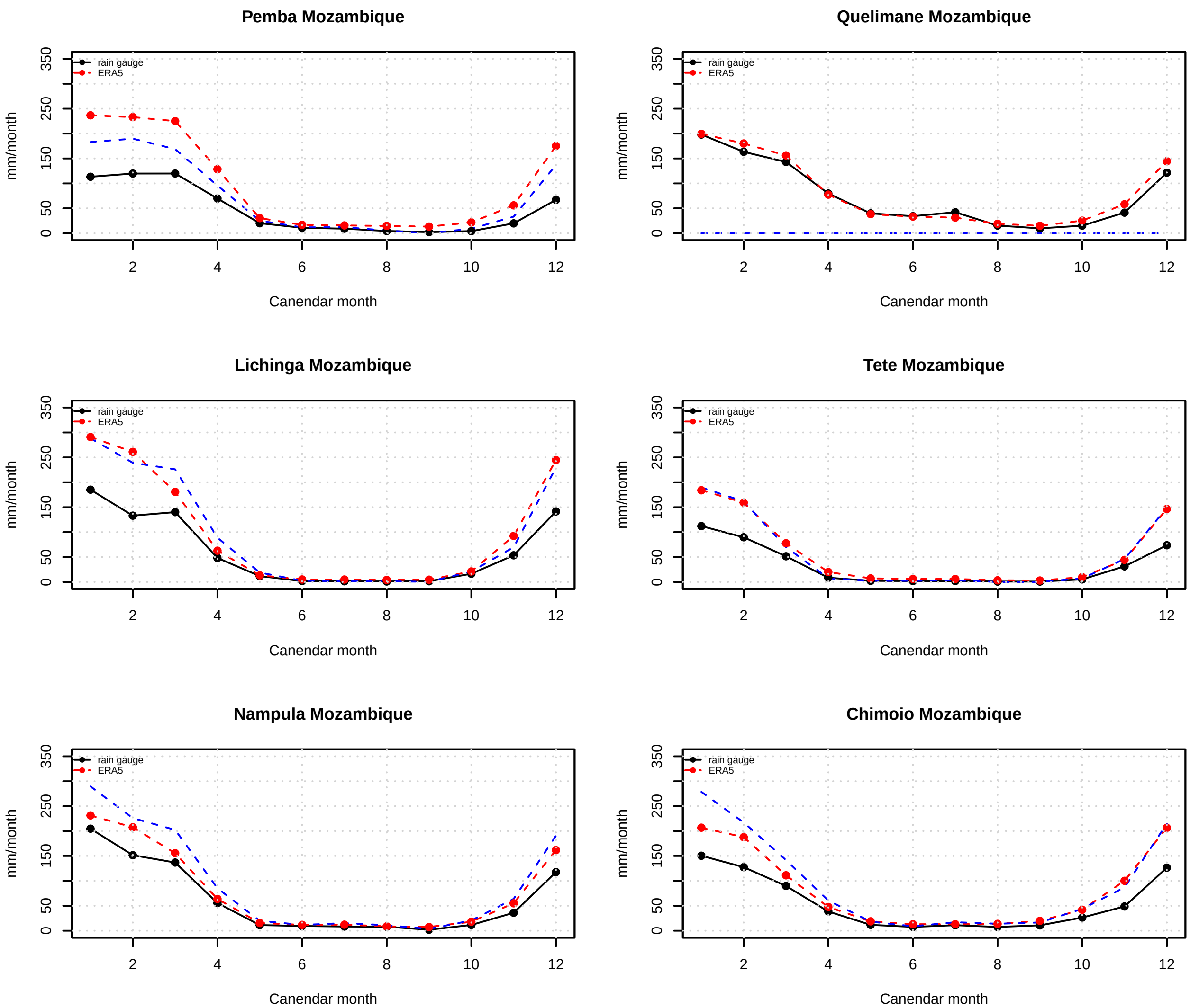
The mean seasonal cycle ('mac') is an important aspect of climatology and a comparison between ERA5 and rain gauges can give useful indication about whether the ERA5 reanalysis provides a description of the regional climatology that is consistent with the in-situ rain gauge data.

Below is a closer comparison between rain gauge climatology and interpolated ERA5 monthly data:

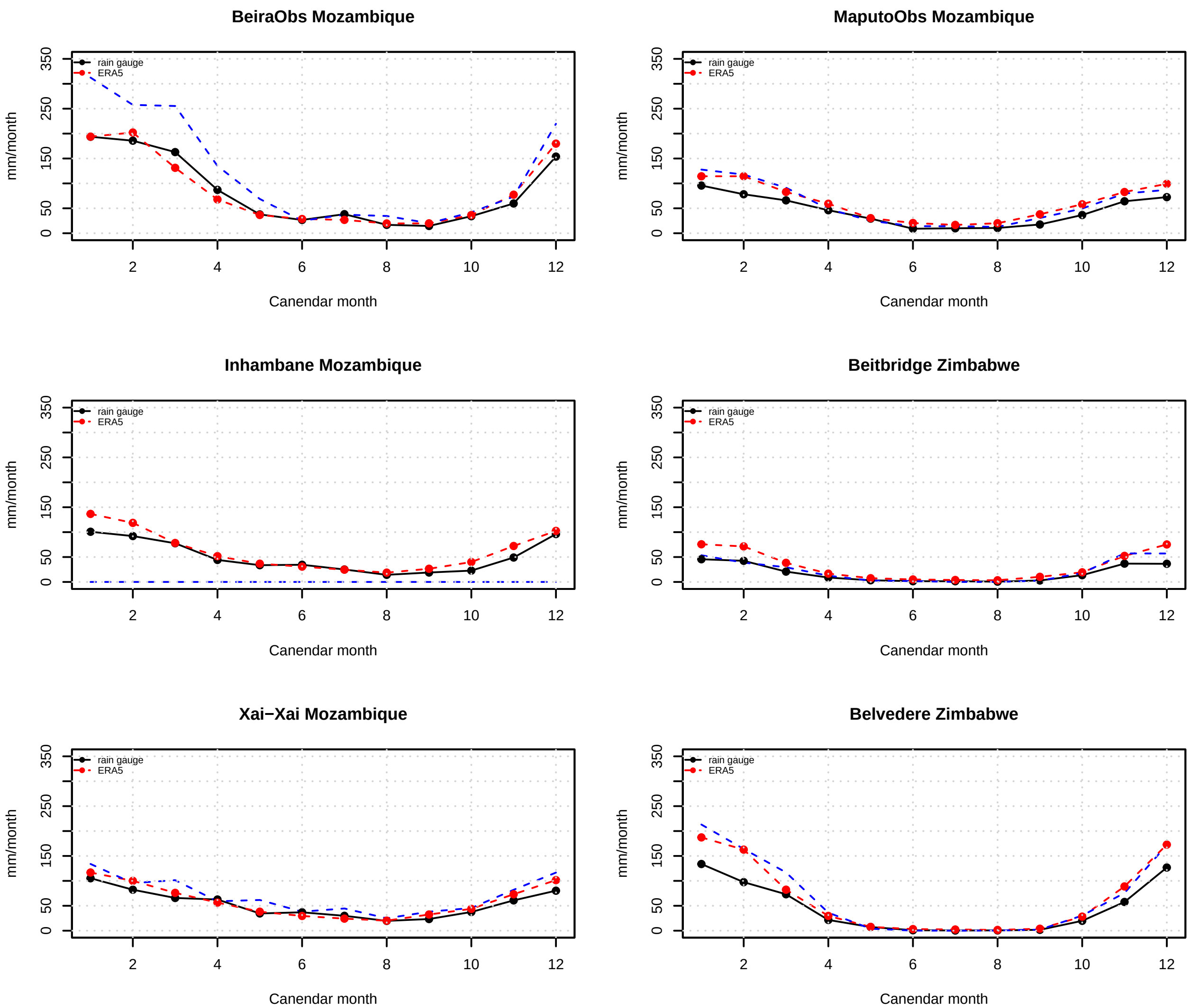
```
X <- subset(X,is=is.element(loc(X),loc(rr.chirps)))
stations <- loc(X)
print(length(stations))
```

```
## [1] 130
```

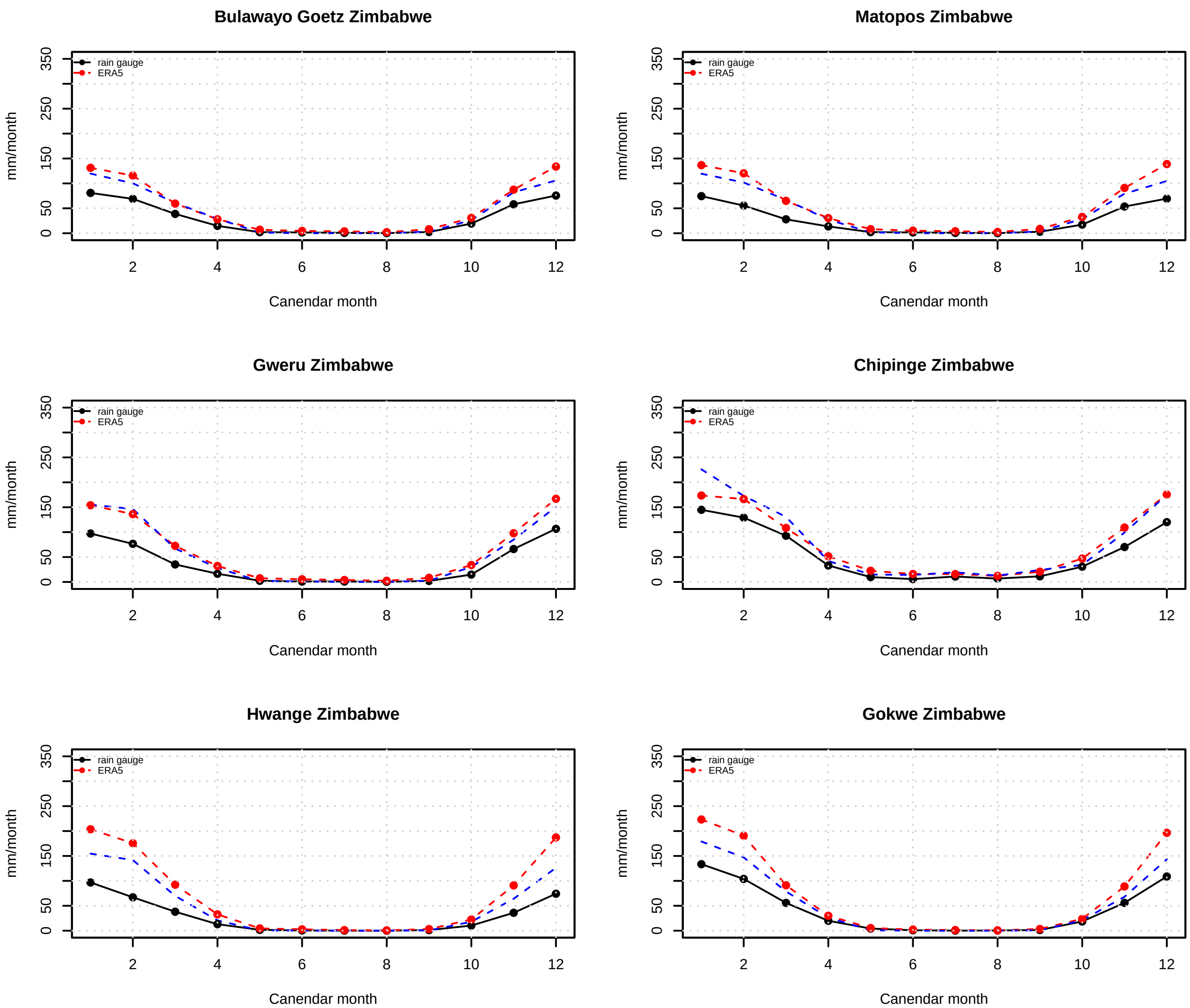
```r
rmse <- rep(NA,length(stations)); rmsep <- rmse
rmse.chirps <- rmse; rmsep.chirps <- rmse
par(mfcol=c(3,2),cex=0.5)
for (i in 1:length(stations)) {
  it <- matchdate(subset(x.mon,is=i),subset(X,is=i))
  x1 <- matchdate(as.monthly(subset(X,is=i),FUN='sum',na.rm=TRUE),x.mon)
  y1 <- matchdate(subset(x.mon,is=i),x1)
  z1 <- matchdate(as.monthly(subset(rr.chirps,is=i),FUN='sum',na.rm=TRUE),x.mon)
  coredata(y1)[is.na(coredata(x1))] <- NA
  x1 <- aggregate(zoo(x1),FUN='mean',by=month,na.rm=TRUE)
  y1 <- aggregate(zoo(y1),FUN='mean',by=month,na.rm=TRUE) # ERA5 is in units of m/day
  z1 <- aggregate(zoo(z1),FUN='mean',by=month,na.rm=TRUE)
  rmse[i] <- RMSE(coredata(x1),coredata(y1))
  rmsep[i] <- 100*rmse[i]/sum(x1)
  rmse.chirps[i] <- RMSE(coredata(z1),coredata(y1))
  rmsep.chirps[i] <- 100*rmse[i]/sum(z1)
  plot(merge(x1,y1,z1),plot.type='single',col=c('black','red','blue'),lty=c(1,2,2),
       main=paste(loc(subset(X,is=i)),cntr(subset(X,is=i))),
       ylab='mm/month',xlab='Canendar month',ylim=c(0,350))
  points(x1,pch=19); points(y1,col='red',pch=19)
  grid()
  legend('topleft',c('rain gauge','ERA5'),col=c('black','red'),pch=19,
         lty=1:2,bty='n',cex=0.6)
}
```

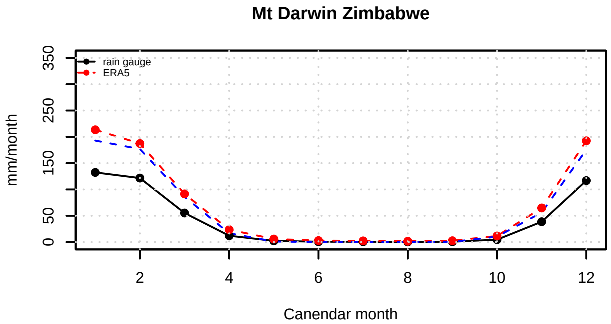


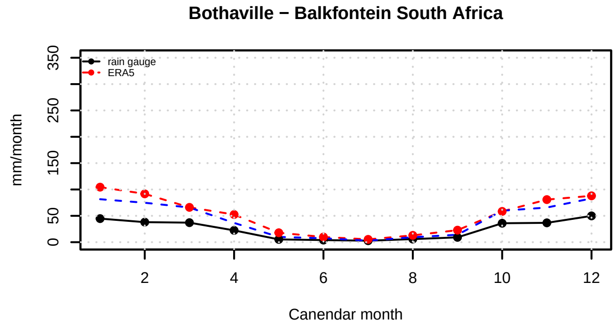
Pemba Mozambique
Quelimane Mozambique
Lichinga Mozambique
Tete Mozambique
Nampula Mozambique
Chimoio Mozambique
rain gauge
ERA5
mm/month
Canendar month
0
50
150
250
350
2
4
6
8
10
12

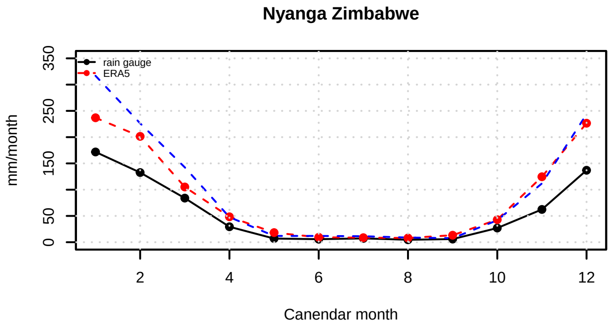


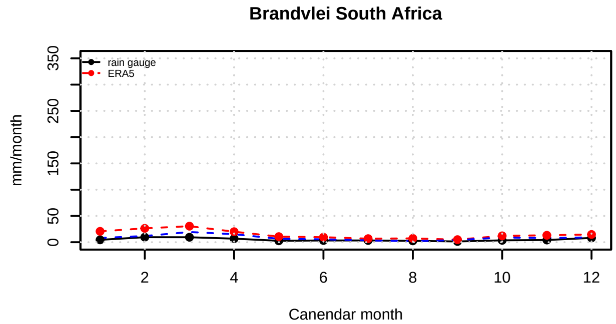
BeiraObs Mozambique
MaputoObs Mozambique
Inhambane Mozambique
Beitbridge Zimbabwe
Xai−Xai Mozambique
Belvedere Zimbabwe
rain gauge
ERA5
mm/month
Canendar month

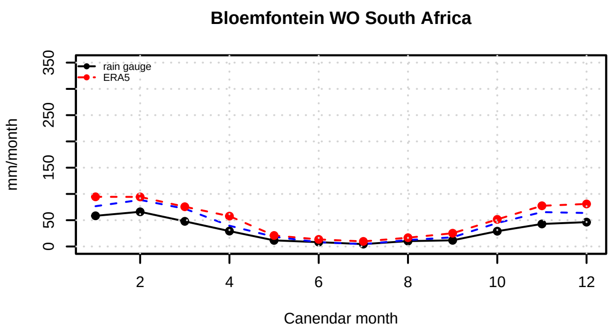

Bulawayo Goetz Zimbabwe
Matopos Zimbabwe
Gweru Zimbabwe
Chipinge Zimbabwe
Hwange Zimbabwe
Gokwe Zimbabwe
rain gauge
ERA5
mm/month
Canendar month
0
50
150
250
350
2
4
6
8
10
12

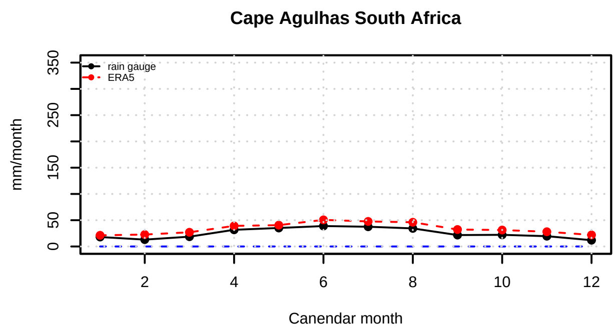

Mt Darwin Zimbabwe
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

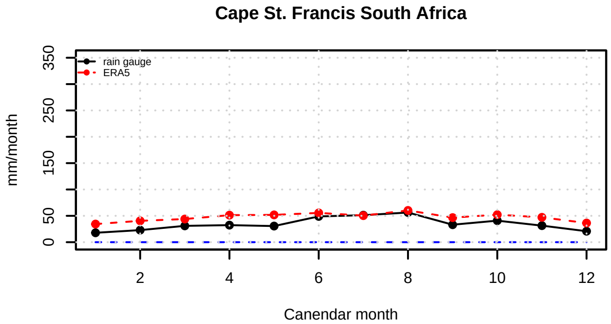

Bothaville − Balkfontein South Africa
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

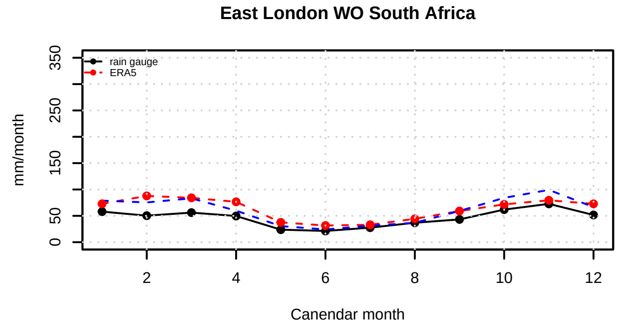

Nyanga Zimbabwe
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

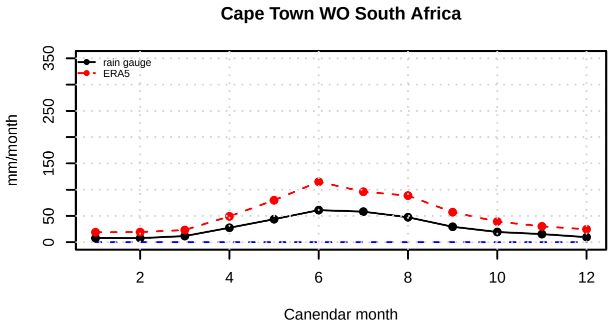

Brandvlei South Africa
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

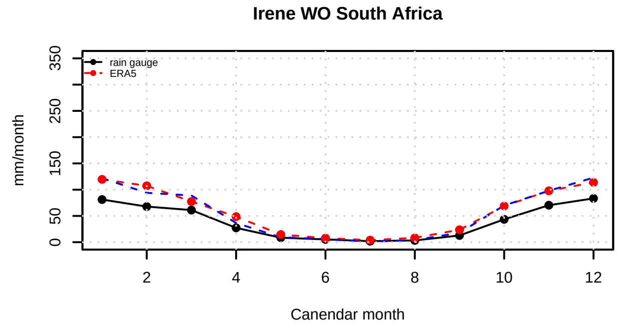

Bloemfontein WO South Africa
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

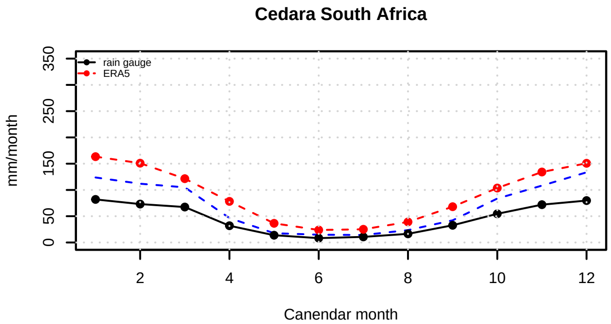

Cape Agulhas South Africa
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

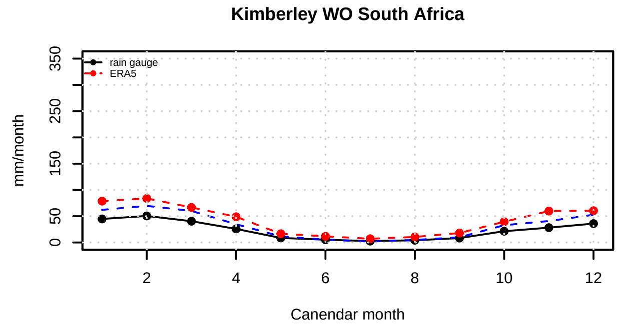

Cape St. Francis South Africa
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

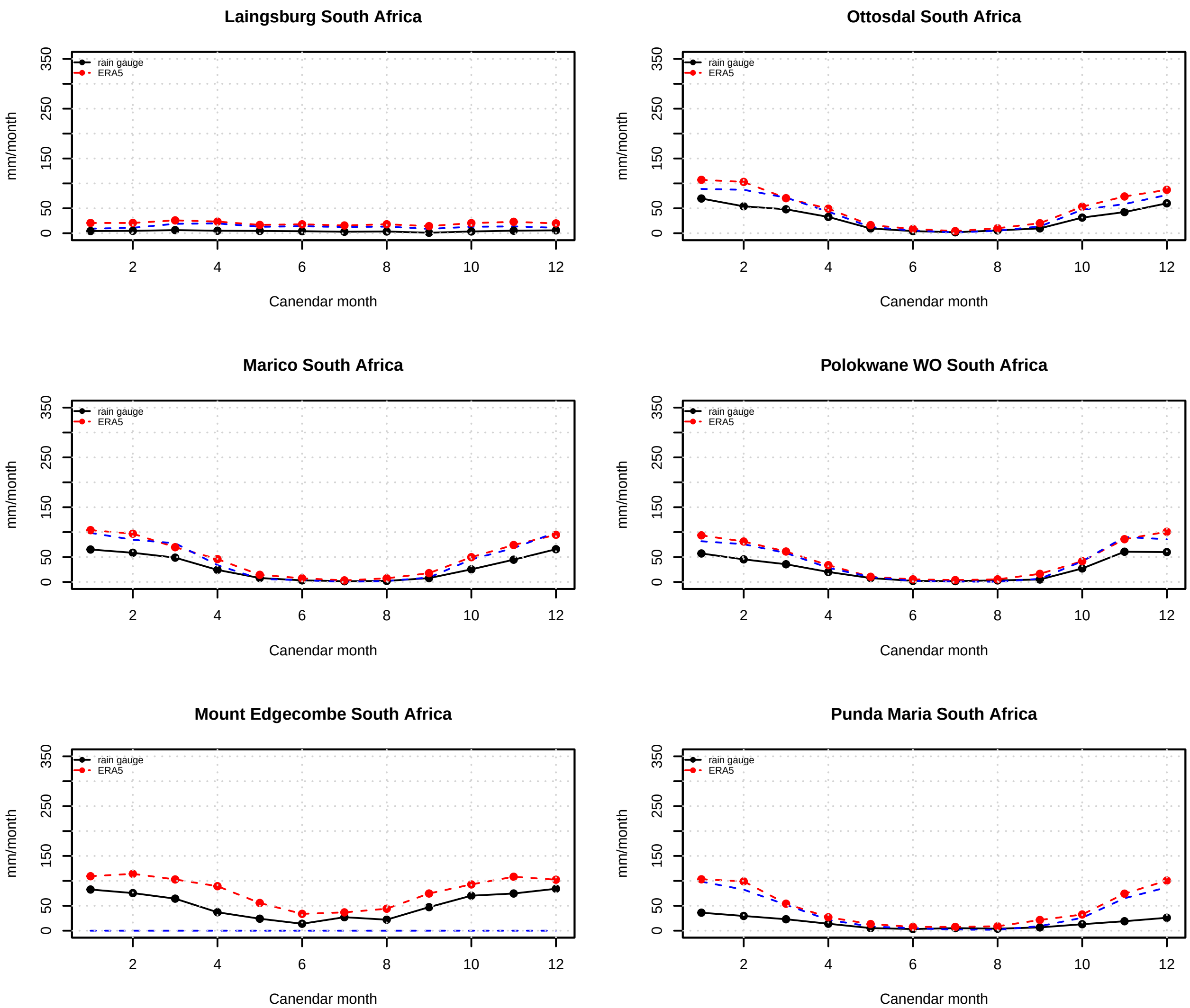

East London WO South Africa
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

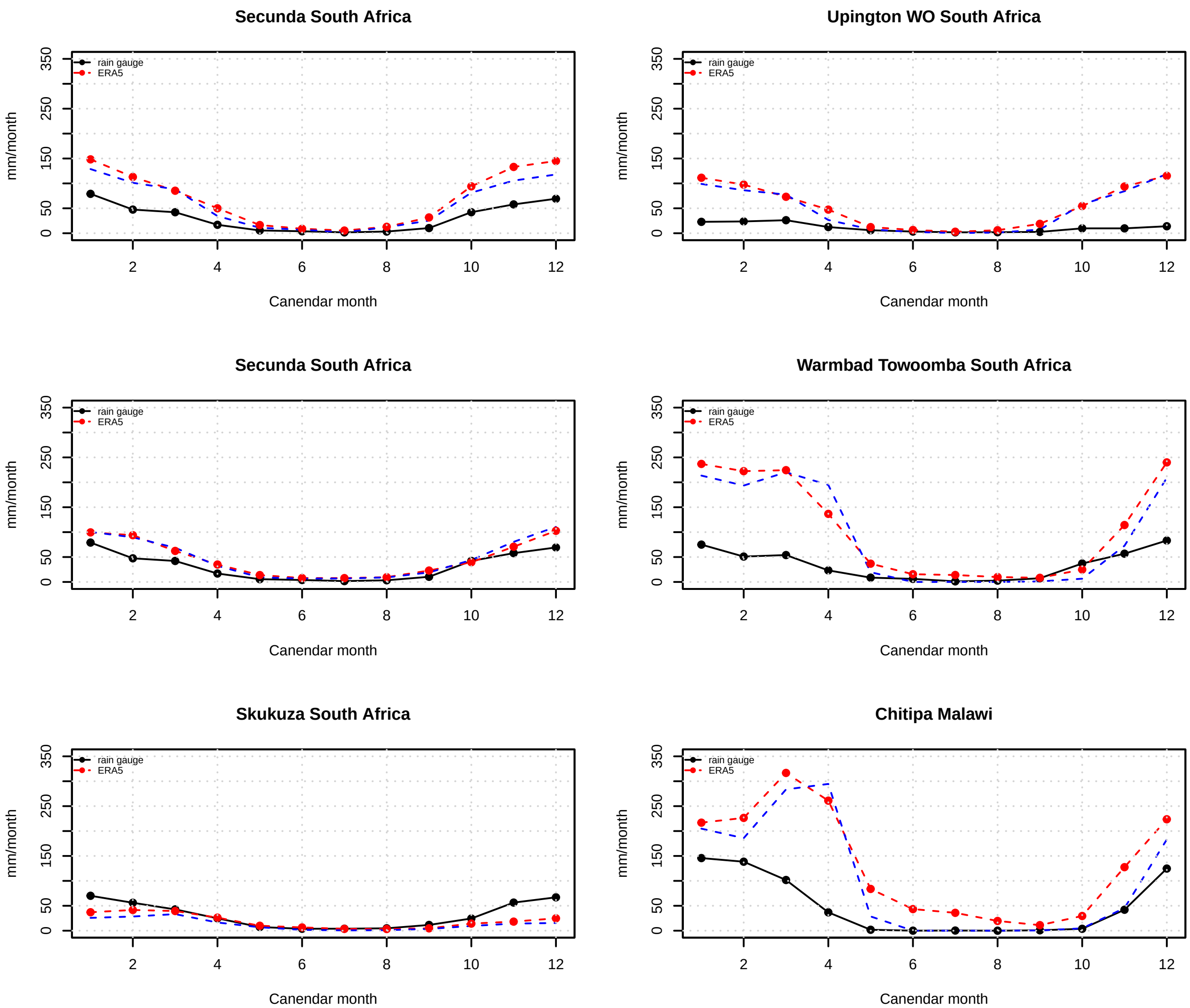

Cape Town WO South Africa
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

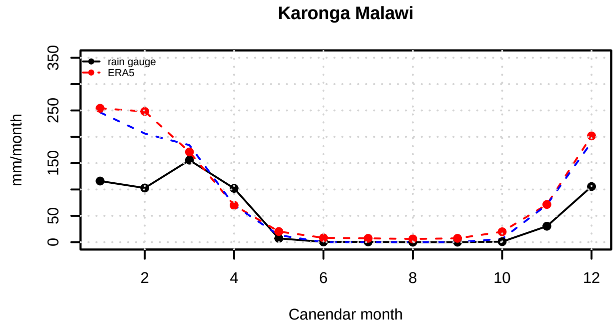

Irene WO South Africa
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

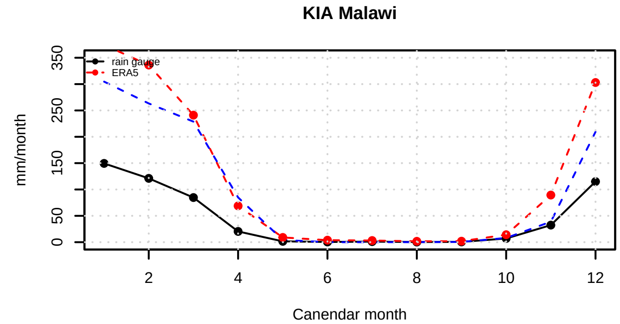

Cedara South Africa
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

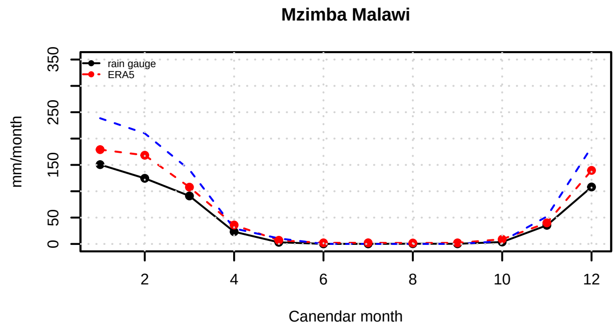

Kimberley WO South Africa
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

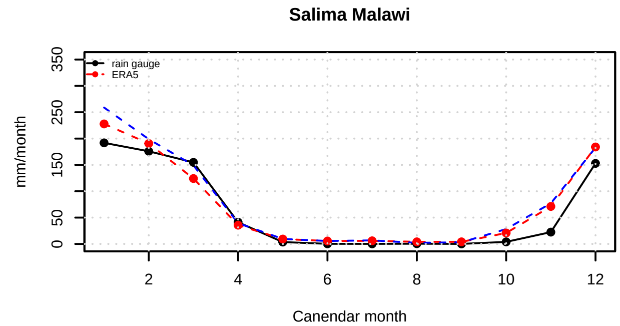

Laingsburg South Africa
Ottosdal South Africa
Marico South Africa
Polokwane WO South Africa
Mount Edgecombe South Africa
Punda Maria South Africa
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

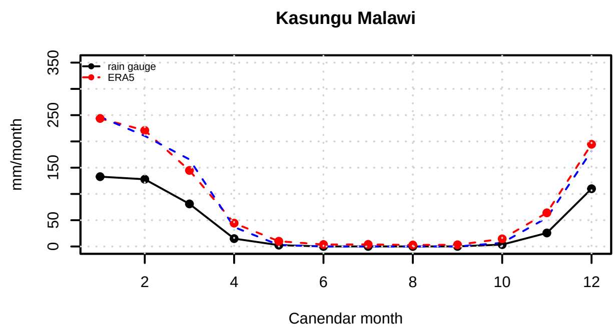

Secunda South Africa
Upington WO South Africa
Secunda South Africa
Warmbad Towoomba South Africa
Skukuza South Africa
Chitipa Malawi
rain gauge
ERA5
mm/month
Canendar month
0
50
150
250
350
2
4
6
8
10
12

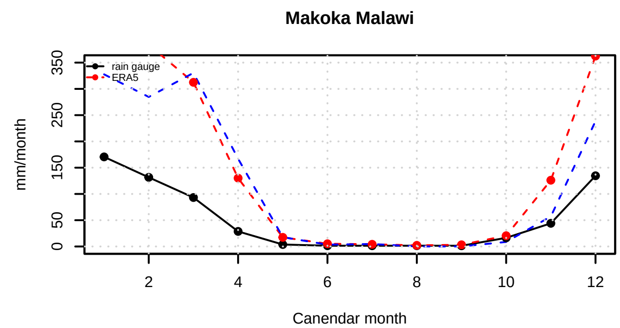

Karonga Malawi
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

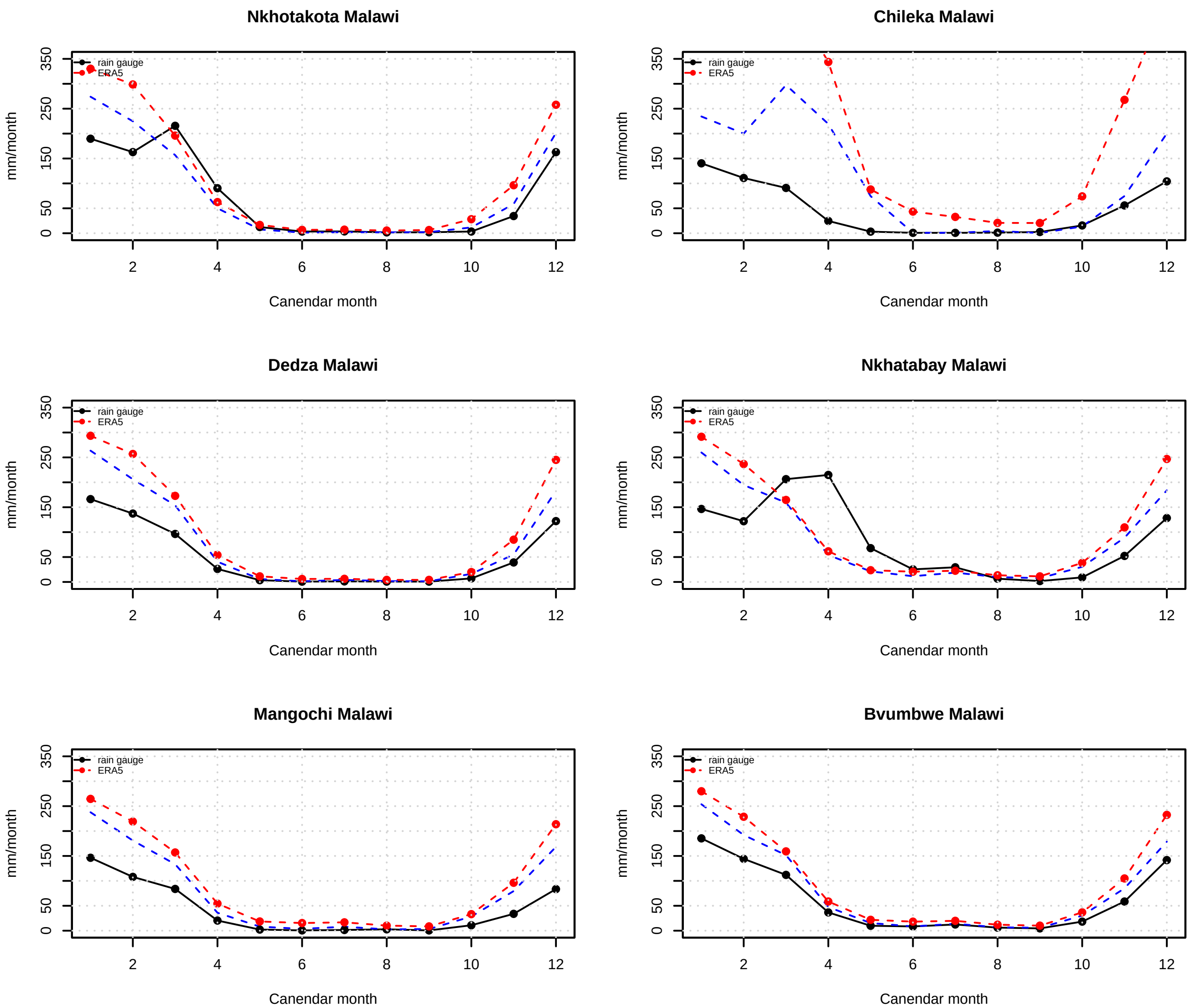

KIA Malawi
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

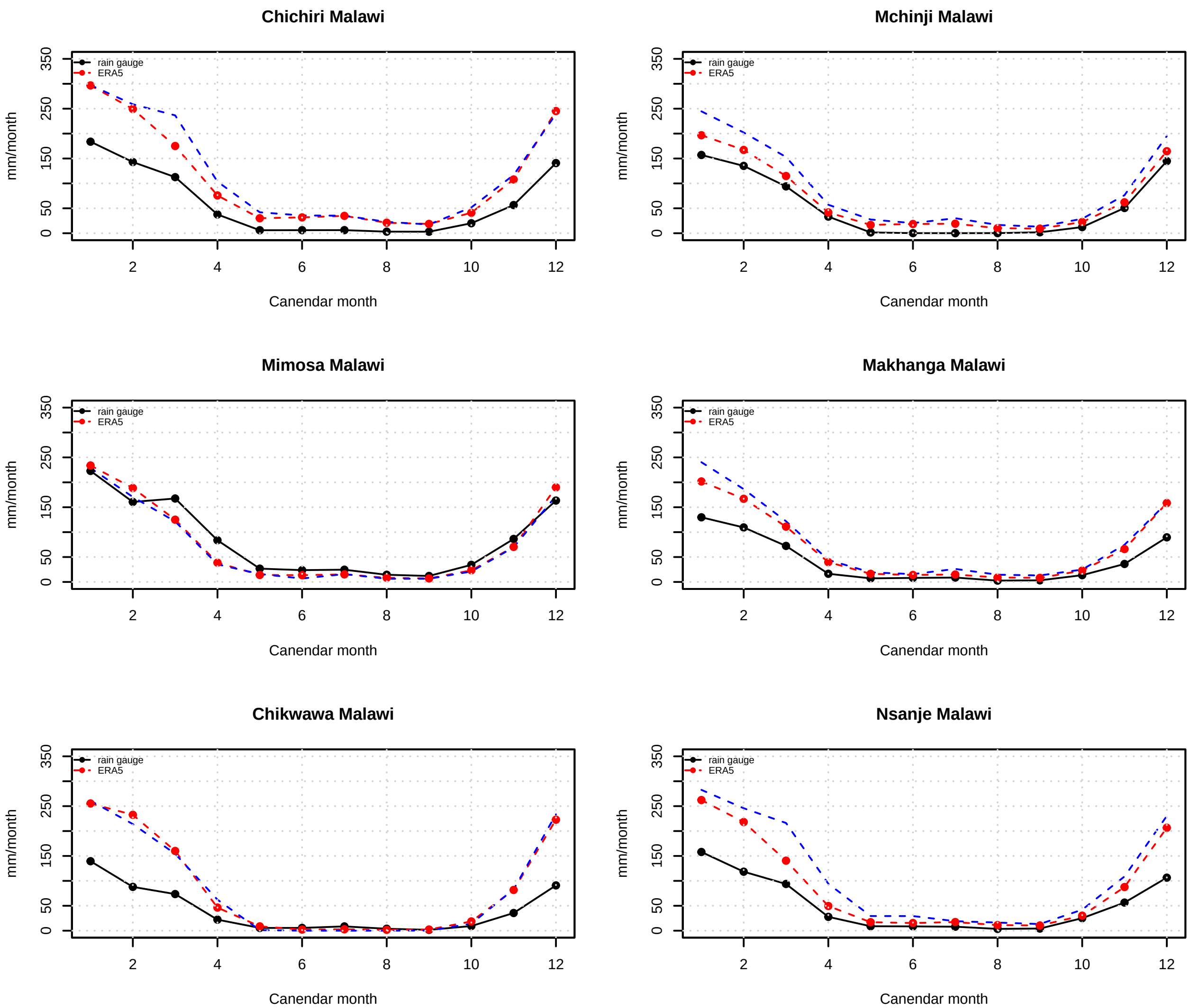

Mzimba Malawi
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

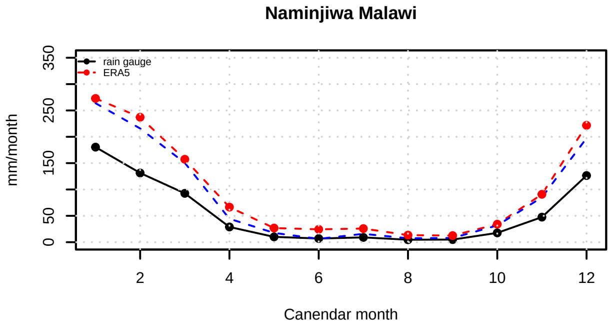

Salima Malawi
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

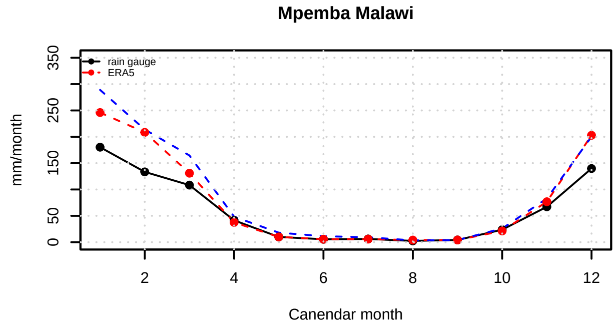

Kasungu Malawi
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

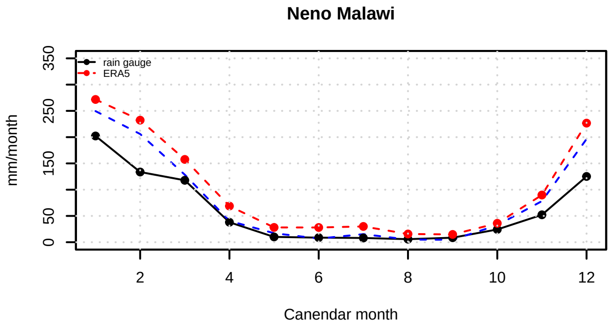

Makoka Malawi
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

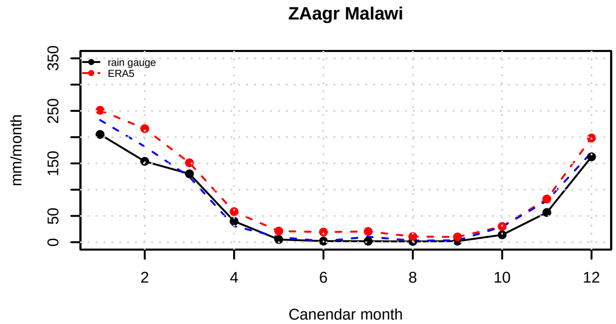

Nkhotakota Malawi
Chileka Malawi
Dedza Malawi
Nkhatabay Malawi
Mangochi Malawi
Bvumbwe Malawi
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

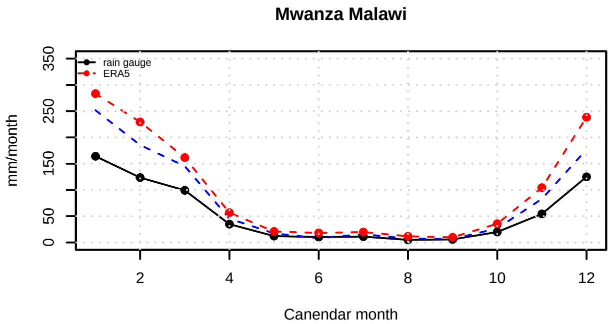

Chichiri Malawi
Mchinji Malawi
Mimosa Malawi
Makhanga Malawi
Chikwawa Malawi
Nsanje Malawi
rain gauge
ERA5
mm/month
Canendar month
0
50
150
250
350
2
4
6
8
10
12

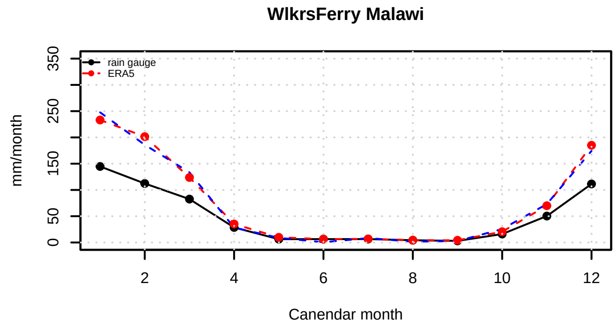

Naminjiwa Malawi
rain gauge
ERA5
mm/month
Canendar month

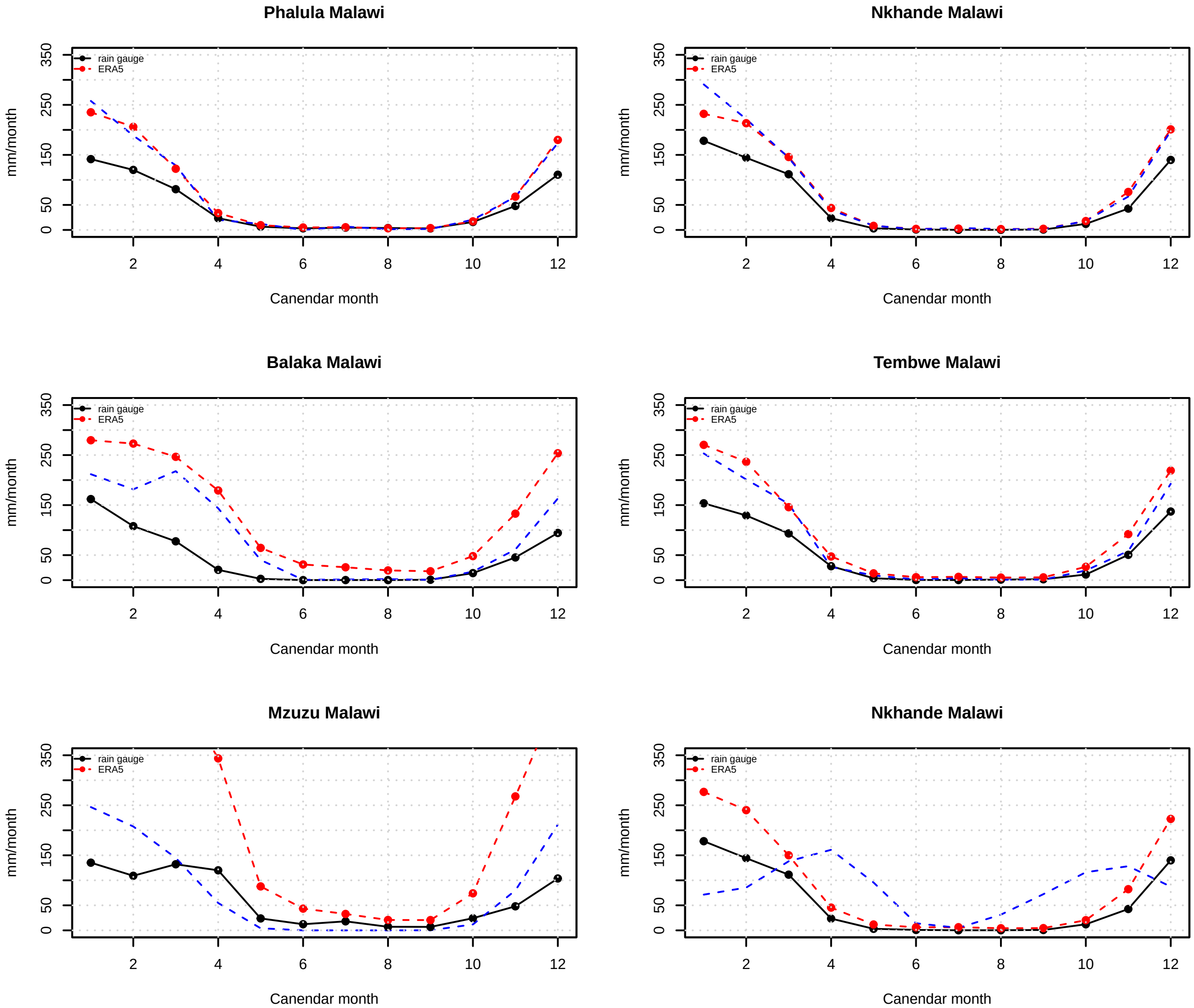

Mpemba Malawi
rain gauge
ERA5
mm/month
Canendar month


Neno Malawi
rain gauge
ERA5
mm/month
Canendar month

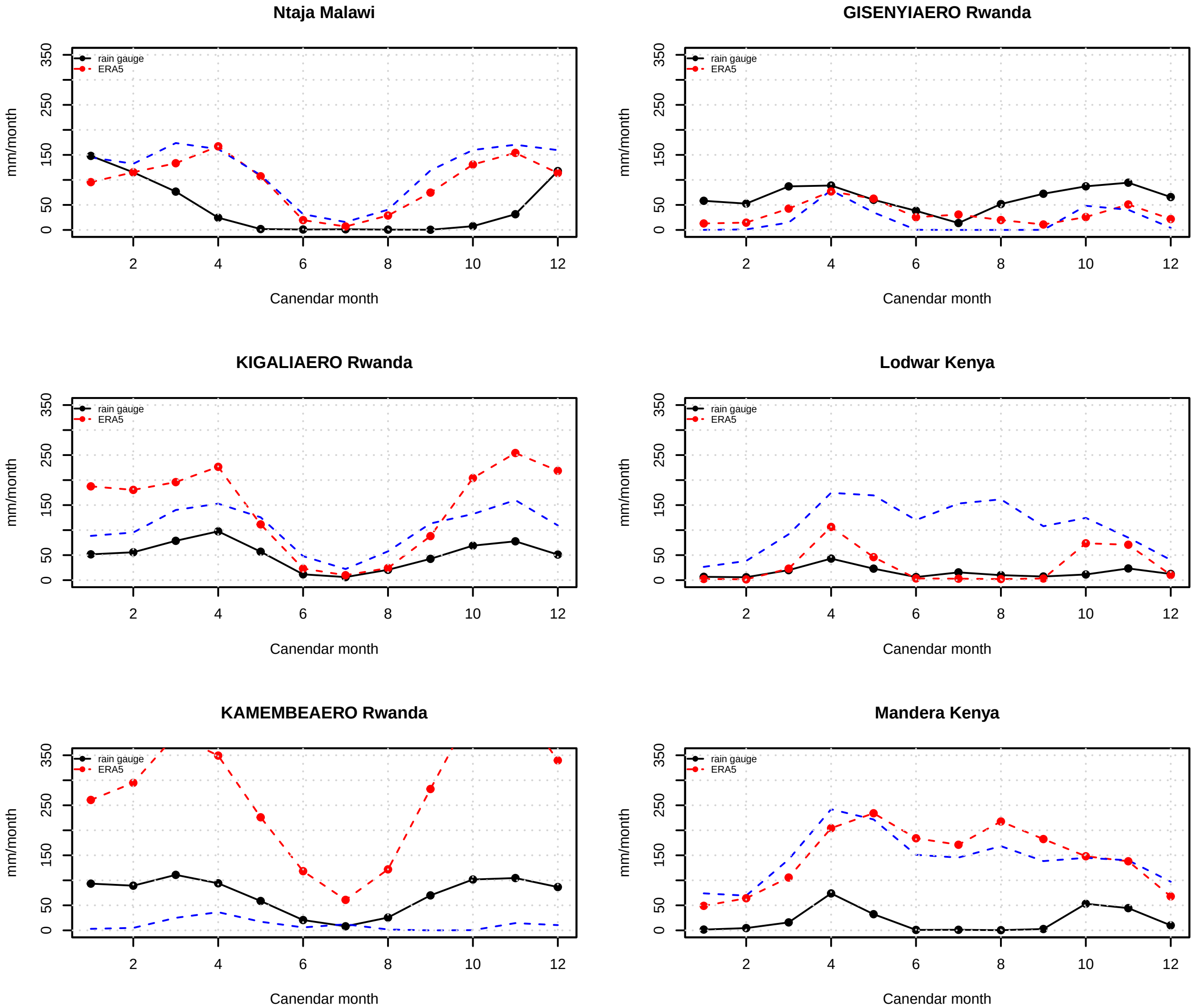

ZAagr Malawi
rain gauge
ERA5
mm/month
Canendar month

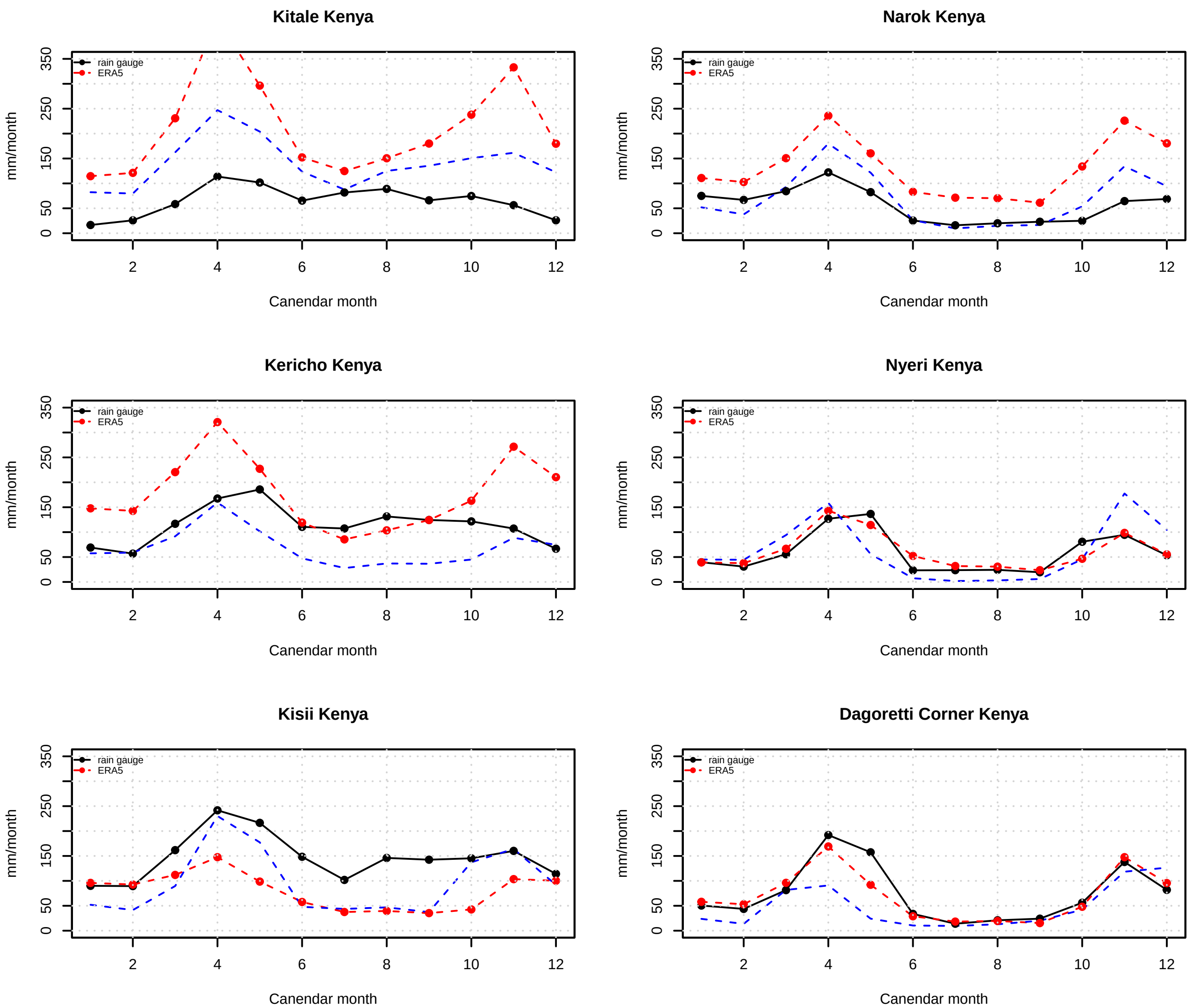

Mwanza Malawi
rain gauge
ERA5
mm/month
Canendar month

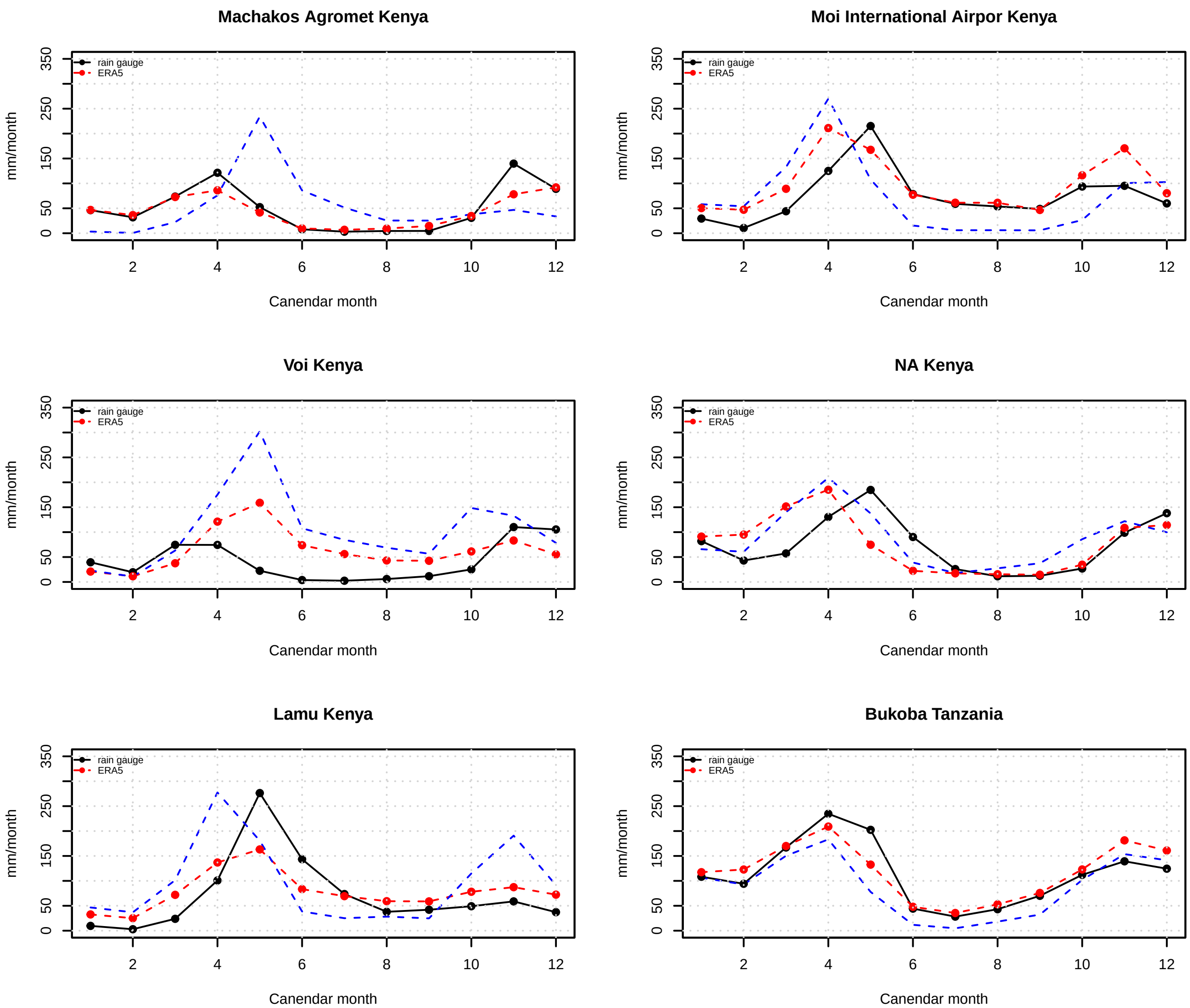

WlkrsFerry Malawi
rain gauge
ERA5
mm/month
Canendar month

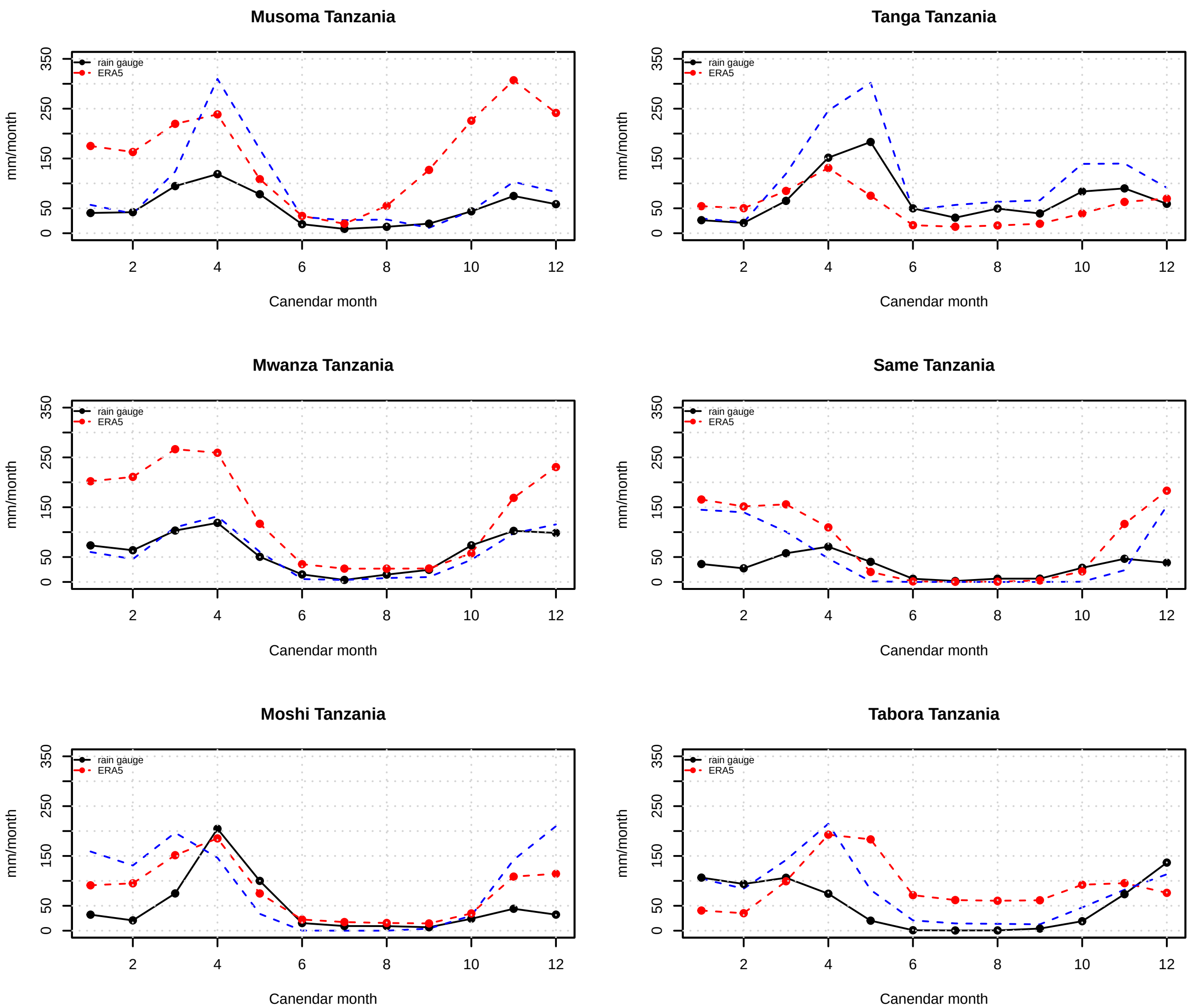


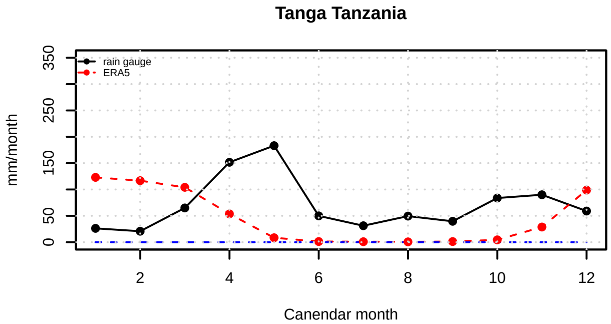
Phalula Malawi
Nkhande Malawi
Balaka Malawi
Tembwe Malawi
Mzuzu Malawi
Nkhande Malawi
rain gauge
ERA5
mm/month
Canendar month

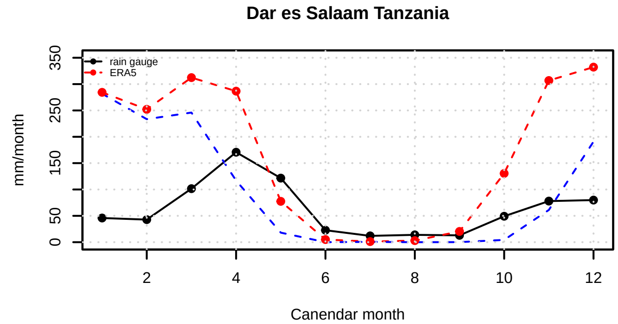

Ntaja Malawi
GISENYIAERO Rwanda
KIGALIAERO Rwanda
Lodwar Kenya
KAMEMBEAERO Rwanda
Mandera Kenya
rain gauge
ERA5
mm/month
Canendar month
0
50
150
250
350
2
4
6
8
10
12

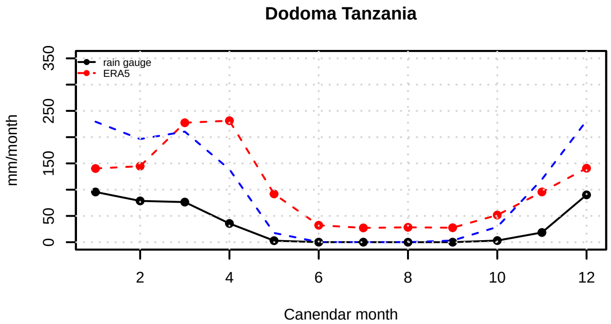

Kitale Kenya
Narok Kenya
Kericho Kenya
Nyeri Kenya
Kisii Kenya
Dagoretti Corner Kenya
rain gauge
ERA5
mm/month
Canendar month
0
50
150
250
350
2
4
6
8
10
12

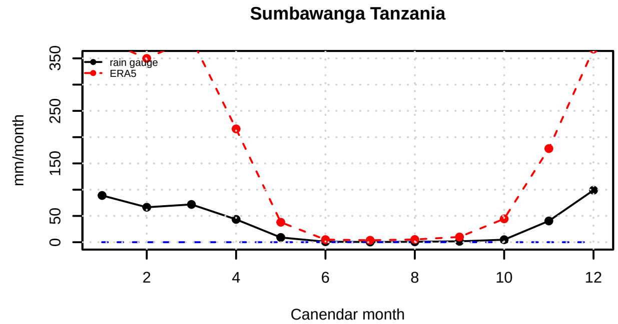

Machakos Agromet Kenya
Moi International Airpor Kenya
Voi Kenya
NA Kenya
Lamu Kenya
Bukoba Tanzania
rain gauge
ERA5
mm/month
Canendar month
0
50
150
250
350
2
4
6
8
10
12

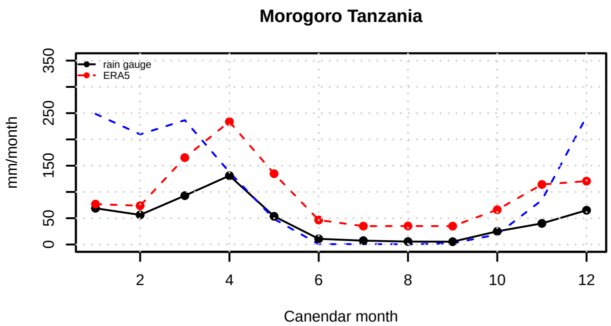

Musoma Tanzania
Tanga Tanzania
Mwanza Tanzania
Same Tanzania
Moshi Tanzania
Tabora Tanzania
rain gauge
ERA5
mm/month
Canendar month
0
50
150
250
350
2
4
6
8
10
12

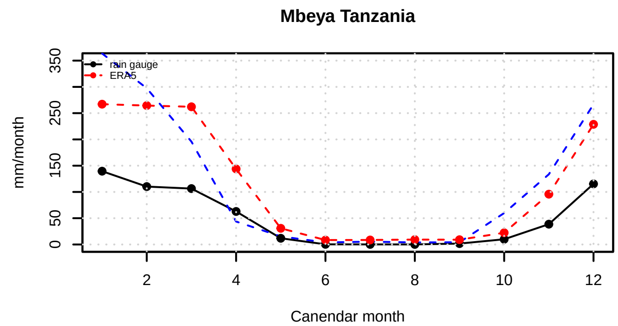

Tanga Tanzania
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

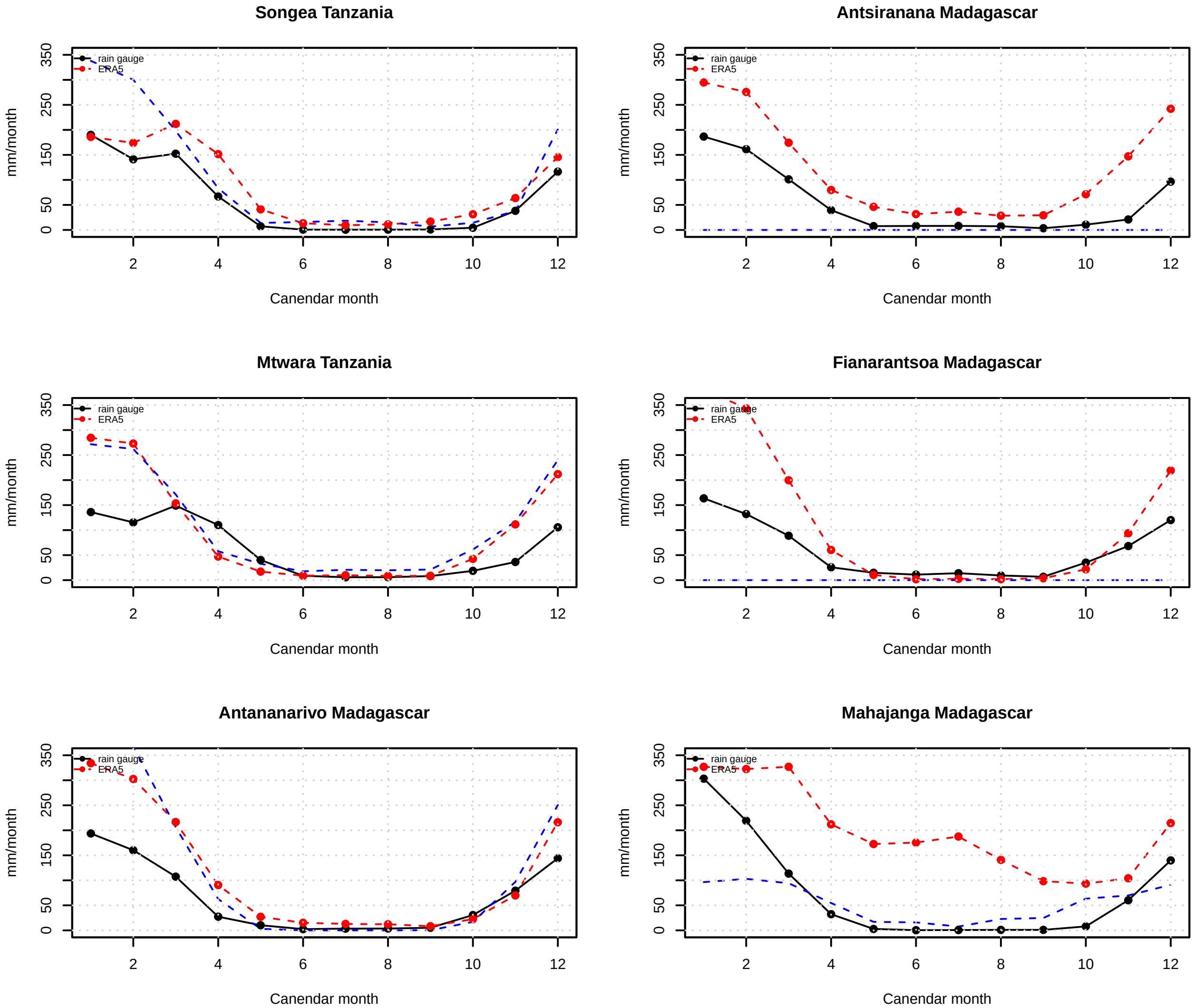

Dar es Salaam Tanzania
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

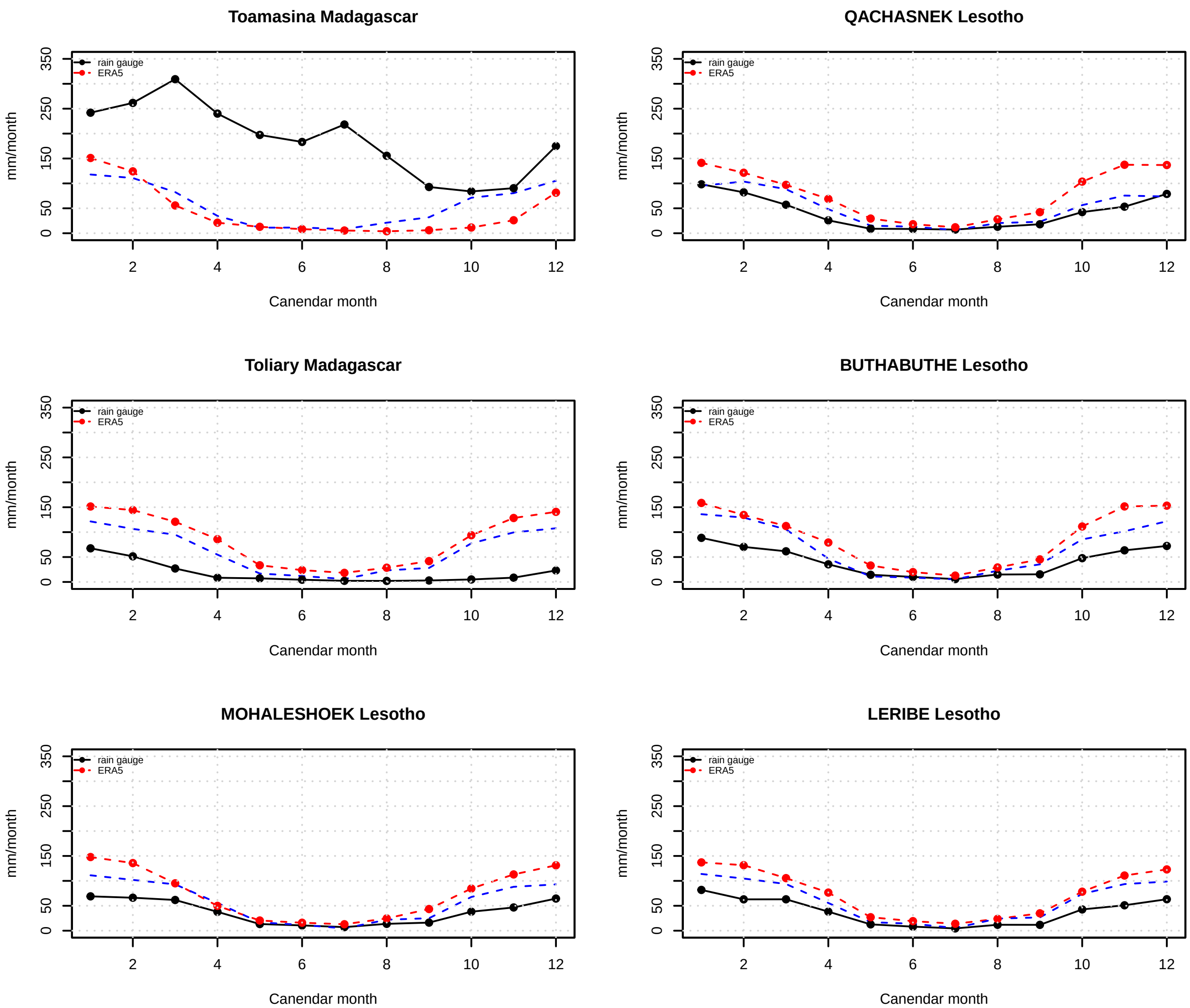

Dodoma Tanzania
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

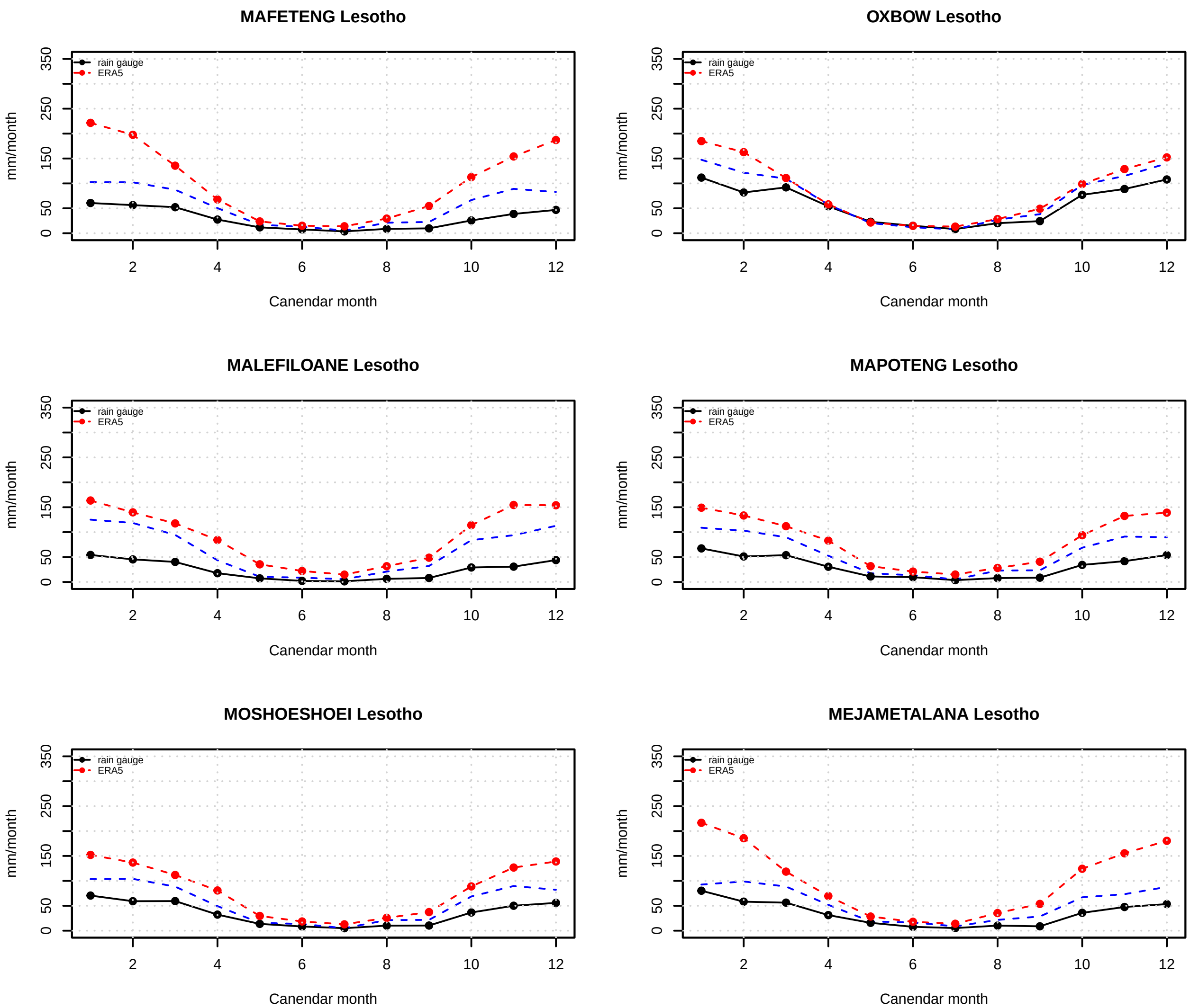

Sumbawanga Tanzania
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

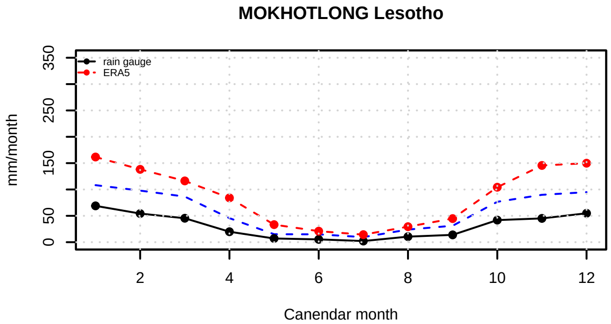

Morogoro Tanzania
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

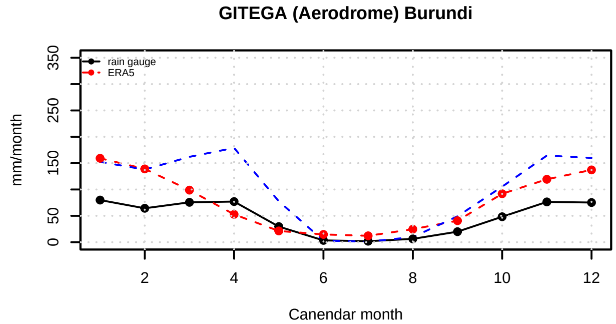

Mbeya Tanzania
mm/month
0
50
150
250
350
rain gauge
ERA5
2
4
6
8
10
12
Canendar month

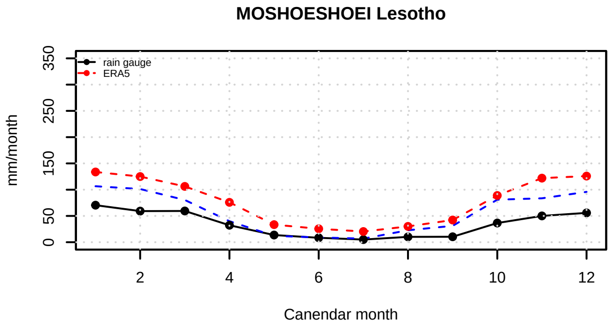

Songea Tanzania
Antsiranana Madagascar
Mtwara Tanzania
Fianarantsoa Madagascar
Antananarivo Madagascar
Mahajanga Madagascar
rain gauge
ERA5
mm/month
Canendar month
0
50
150
250
350
2
4
6
8
10
12

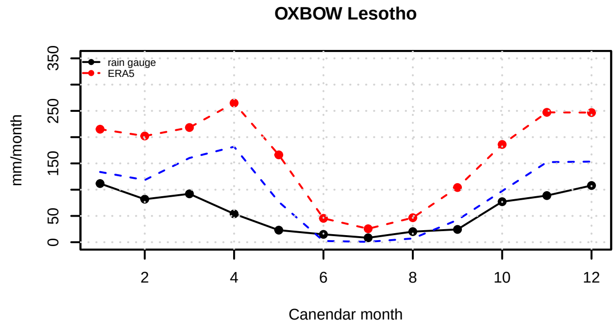

Toamasina Madagascar
QACHASNEK Lesotho
Toliary Madagascar
BUTHABUTHE Lesotho
MOHALESHOEK Lesotho
LERIBE Lesotho
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

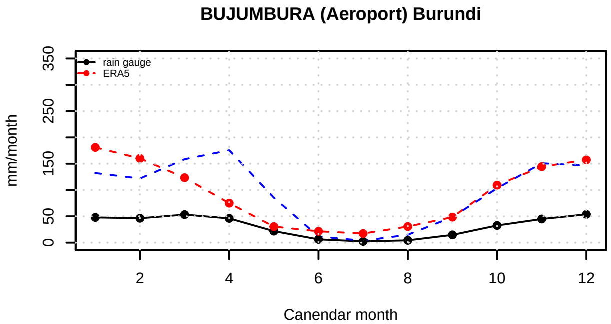

MAFETENG Lesotho
OXBOW Lesotho
MALEFILOANE Lesotho
MAPOTENG Lesotho
MOSHOESHOEI Lesotho
MEJAMETALANA Lesotho
rain gauge
ERA5
mm/month
0
50
150
250
350
2
4
6
8
10
12
Canendar month

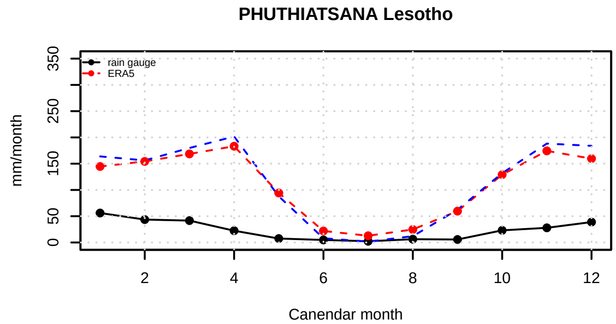




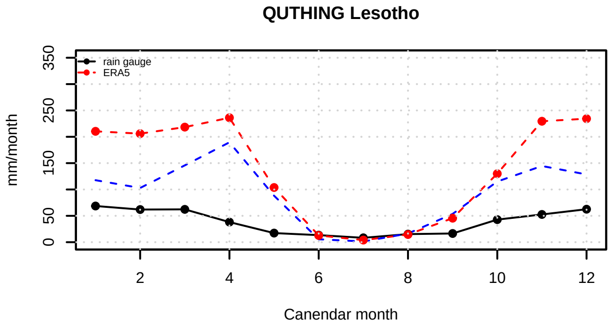


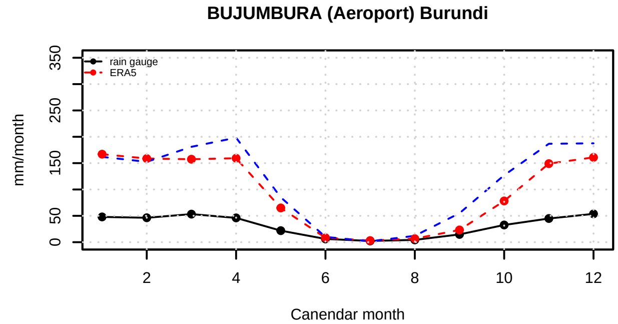


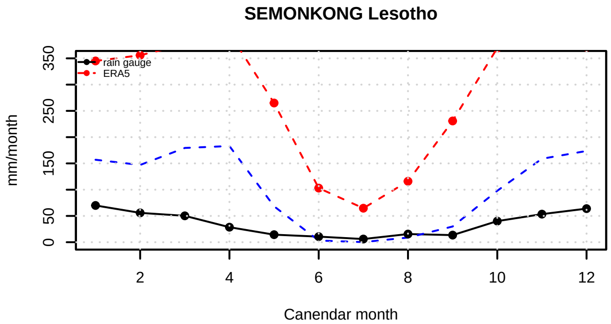


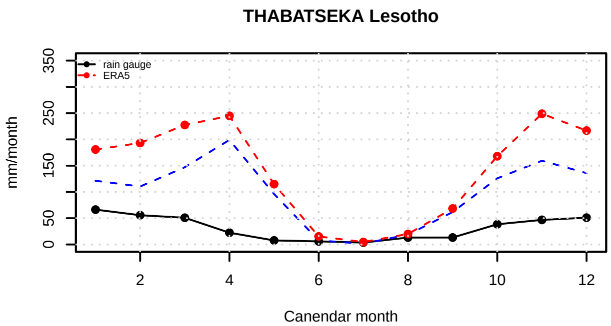



```
par(mfcol=c(1,1),cex=1)
```

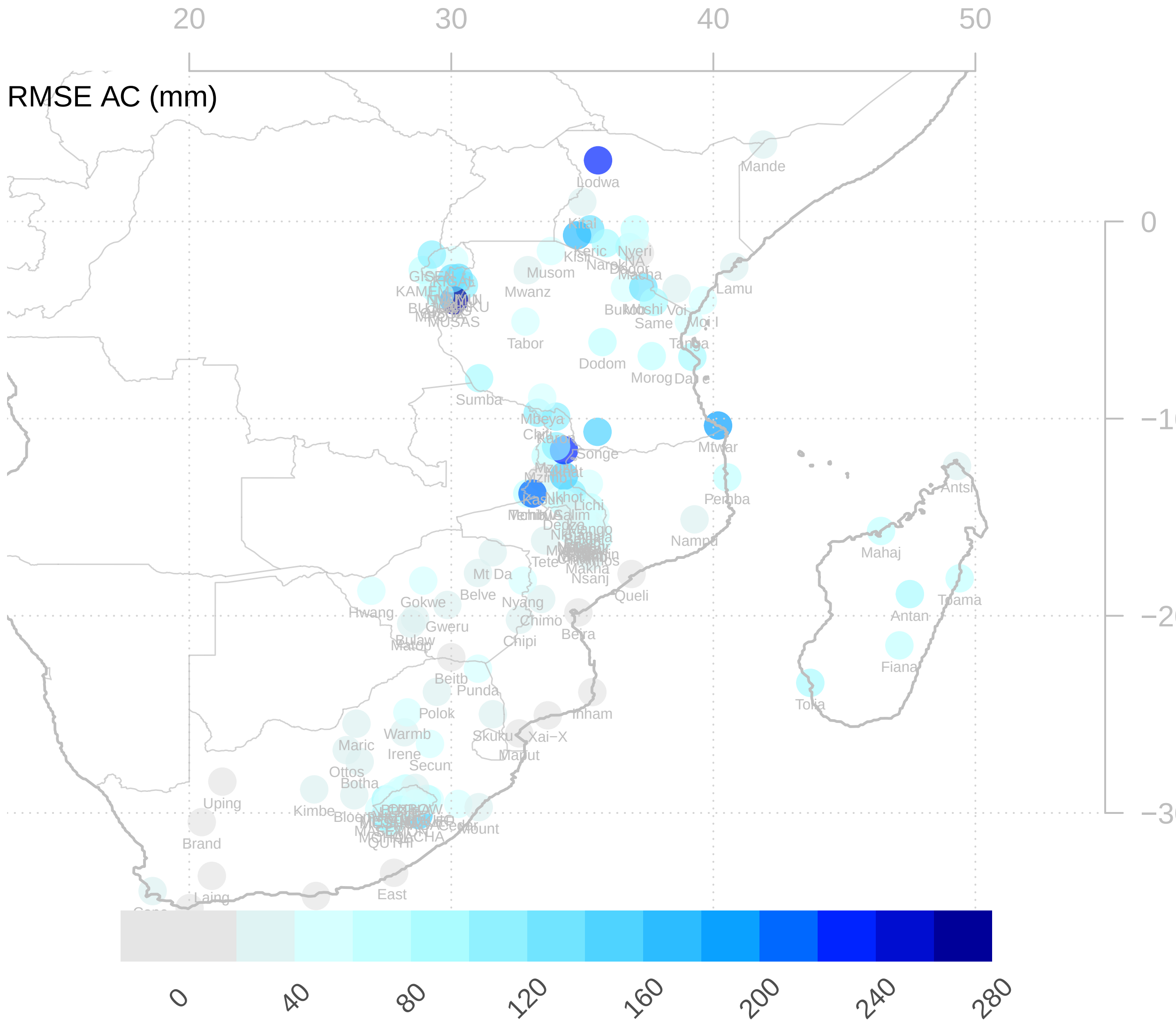




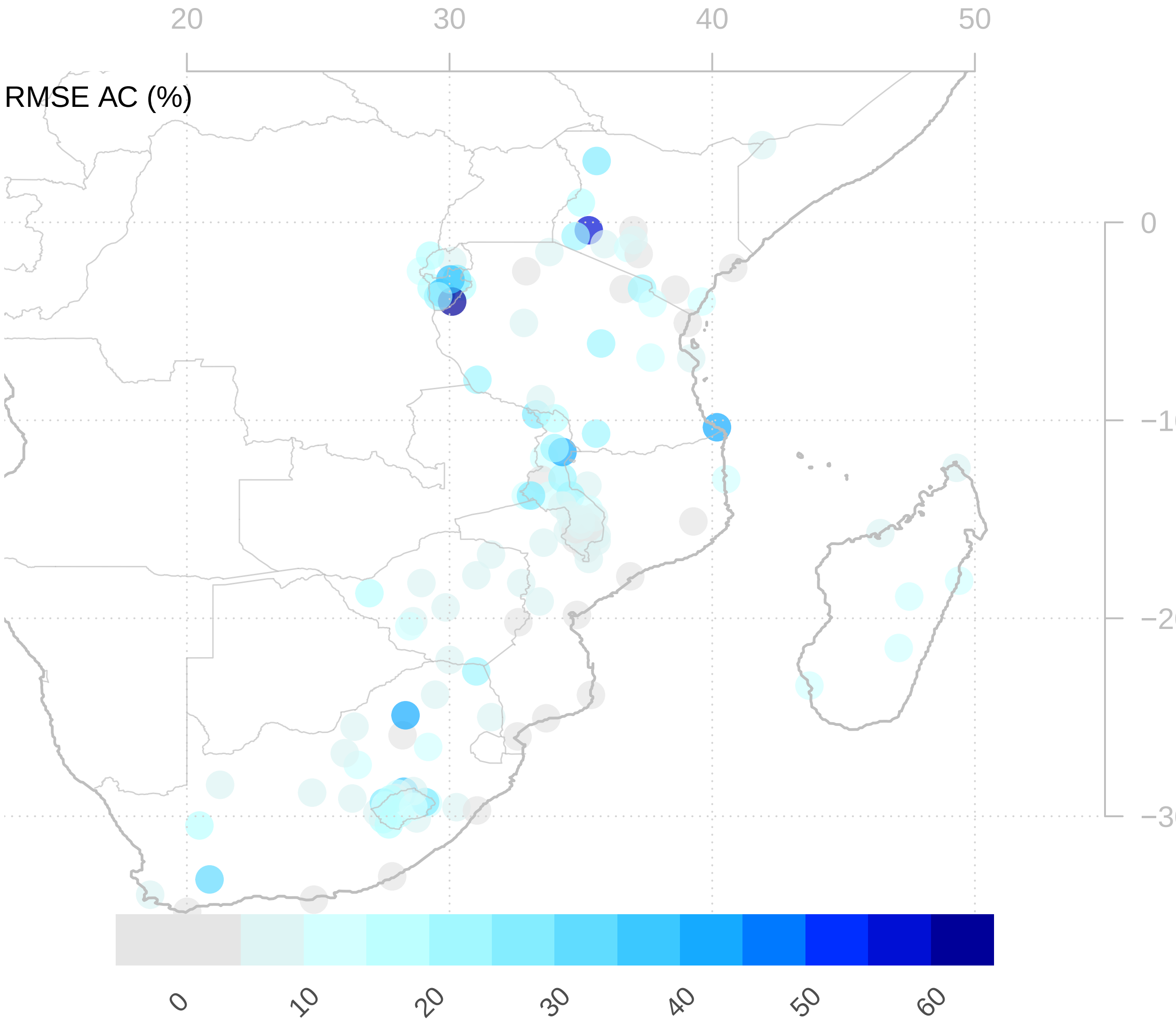





```
attr(X,'AC.RMSE') <- rmse
attr(X,'AC.RMSE.p') <- rmsep
attr(X,'AC.RMSE.chirps') <- rmse.chirps
attr(X,'AC.RMSE.p.chirps') <- rmsep.chirps
map(X,FUN='AC.RMSE',add.text = TRUE,main="RMSE AC (mm)",
    colbar=list(pal='precip'),new=FALSE,border=TRUE,cex.lab=0.5)
```




```
map(X,FUN='AC.RMSE.p',add.text = FALSE,main="RMSE AC (%)",
    colbar=list(pal='precip'),new=FALSE,border=TRUE)
```




```
srt <- order(rmsep)
print(cbind(stations,cntr(X),round(rmsep),round(rmse,1))[srt,])
```

```
##       stations
##  [1,] "Quelimane"                "Mozambique"   "1"  "11.3"
##  [2,] "BeiraObs"                 "Mozambique"   "2"  "15.3"
##  [3,] "Xai-Xai"                  "Mozambique"   "2"  "10.9"
##  [4,] "Mwanza"                   "Tanzania"     "2"  "28.7"
##  [5,] "Chikwawa"                 "Malawi"       "2"  "22.8"
##  [6,] "Machakos Agromet"         "Kenya"        "2"  "16"
##  [7,] "Voi"                      "Kenya"        "2"  "21.6"
##  [8,] "Inhambane"                "Mozambique"   "3"  "15.9"
##  [9,] "Cape Agulhas"             "South Africa" "3"  "9.1"
## [10,] "Makoka"                   "Malawi"       "3"  "22.8"
## [11,] "Nampula"                  "Mozambique"   "3"  "23.6"
## [12,] "Makhanga"                 "Malawi"       "3"  "20"
## [13,] "Cape St. Francis"         "South Africa" "3"  "14.2"
## [14,] "East London WO"           "South Africa" "3"  "19.1"
## [15,] "Kasungu"                  "Malawi"       "3"  "18.8"
## [16,] "Lamu"                     "Kenya"        "3"  "21.1"
```

```
## [17,] "MaputoObs"              "Mozambique"   "4"  "19"
## [18,] "WlkrsFerry"             "Malawi"       "4"  "28.9"
## [19,] "Chipinge"               "Zimbabwe"     "4"  "26"
## [20,] "Nyeri"                  "Kenya"        "4"  "78.7"
## [21,] "Bukoba"                 "Tanzania"     "4"  "41"
## [22,] "Tanga"                  "Tanzania"     "5"  "40.8"
## [23,] "ZAagr"                  "Malawi"       "5"  "34.8"
## [24,] "Mount Edgecombe"        "South Africa" "5"  "30.4"
## [25,] "Irene WO"               "South Africa" "5"  "22.9"
## [26,] "Antsiranana"            "Madagascar"   "5"  "36.3"
## [27,] "Mandera"                "Kenya"        "5"  "39.2"
## [28,] "Nkhande"                "Malawi"       "5"  "34.4"
## [29,] "NA"                     "Kenya"        "5"  "45.4"
## [30,] "Upington WO"            "South Africa" "5"  "19.8"
## [31,] "Skukuza"                "South Africa" "5"  "20.2"
## [32,] "OXBOW"                  "Lesotho"      "5"  "37.7"
## [33,] "Belvedere"              "Zimbabwe"     "5"  "29.5"
## [34,] "Chimoio"                "Mozambique"   "6"  "37.5"
## [35,] "Musoma"                 "Tanzania"     "6"  "51.8"
## [36,] "Nyanga"                 "Zimbabwe"     "6"  "43"
## [37,] "Marico"                 "South Africa" "6"  "23"
## [38,] "Ottosdal"               "South Africa" "7"  "24.1"
## [39,] "Bloemfontein WO"        "South Africa" "7"  "24.1"
## [40,] "Narok"                  "Kenya"        "7"  "91.4"
## [41,] "Chichiri"               "Malawi"       "7"  "49.9"
## [42,] "Mwanza"                 "Malawi"       "7"  "50.2"
## [43,] "Polokwane WO"           "South Africa" "7"  "22.6"
## [44,] "QACHASNEK"              "Lesotho"      "7"  "157.6"
## [45,] "Nsanje"                 "Malawi"       "7"  "37"
## [46,] "Dedza"                  "Malawi"       "8"  "66.4"
## [47,] "KIGALIAERO"             "Rwanda"       "8"  "49.6"
## [48,] "GITEGA (Aerodrome)"     "Burundi"      "8"  "42.5"
## [49,] "Phalula"                "Malawi"       "8"  "44.1"
## [50,] "Balaka"                 "Malawi"       "8"  "43.9"
## [51,] "Bulawayo Goetz"         "Zimbabwe"     "8"  "28.6"
## [52,] "Kimberley WO"           "South Africa" "8"  "21.7"
## [53,] "Lichinga"               "Mozambique"   "8"  "58.9"
## [54,] "Gweru"                  "Zimbabwe"     "8"  "33.6"
## [55,] "Bvumbwe"                "Malawi"       "8"  "81.5"
## [56,] "Mt Darwin"              "Zimbabwe"     "8"  "39.6"
## [57,] "Tabora"                 "Tanzania"     "8"  "47.5"
## [58,] "Cape Town WO"           "South Africa" "8"  "28.6"
## [59,] "Neno"                   "Malawi"       "8"  "56"
## [60,] "Beitbridge"             "Zimbabwe"     "9"  "18.2"
## [61,] "Mimosa"                 "Malawi"       "9"  "61.9"
## [62,] "Naminjiwa"              "Malawi"       "9"  "53.8"
## [63,] "MAFETENG"               "Lesotho"      "9"  "43.2"
## [64,] "Ntaja"                  "Malawi"       "9"  "55.7"
## [65,] "MAPOTENG"               "Lesotho"      "9"  "41.4"
## [66,] "Gokwe"                  "Zimbabwe"     "9"  "46.3"
## [67,] "Mpemba"                 "Malawi"       "9"  "61.7"
## [68,] "Mahajanga"              "Madagascar"   "9"  "71.8"
## [69,] "Cedara"                 "South Africa" "9"  "51.5"
## [70,] "Dar es Salaam"          "Tanzania"     "10" "81.1"
```

```
## [71,] "Tete"                       "Mozambique"   "10" "36.9" 
## [72,] "Mbeya"                      "Tanzania"     "10" "55.4" 
## [73,] "LERIBE"                     "Lesotho"      "10" "45"   
## [74,] "Fianarantsoa"               "Madagascar"   "10" "75.8" 
## [75,] "MALEFILOANE"                "Lesotho"      "11" "52.8" 
## [76,] "Mzimba"                     "Malawi"       "11" "66.5" 
## [77,] "KIA"                        "Malawi"       "11" "53.8" 
## [78,] "Bothaville - Balkfontein"   "South Africa" "11" "32.3" 
## [79,] "Mangochi"                   "Malawi"       "11" "67.5" 
## [80,] "Moi International Airpor"   "Kenya"        "11" "55.7" 
## [81,] "Matopos"                    "Zimbabwe"     "12" "36.8" 
## [82,] "THABATSEKA"                 "Lesotho"      "12" "47.9" 
## [83,] "Pemba"                      "Mozambique"   "12" "68.5" 
## [84,] "Toamasina"                  "Madagascar"   "12" "79.9" 
## [85,] "Morogoro"                   "Tanzania"     "12" "78.3" 
## [86,] "Secunda"                    "South Africa" "12" "47.3" 
## [87,] "Dagoretti Corner"           "Kenya"        "13" "85"   
## [88,] "Same"                       "Tanzania"     "13" "96.7" 
## [89,] "MOSHOESHOEI"                "Lesotho"      "13" "53.7" 
## [90,] "Chileka"                    "Malawi"       "14" "67.7" 
## [91,] "Toliary"                    "Madagascar"   "14" "98.5" 
## [92,] "Antananarivo"               "Madagascar"   "14" "85.5" 
## [93,] "Mchinji"                    "Malawi"       "15" "72"   
## [94,] "KAMEMBEAERO"                "Rwanda"       "15" "78.7" 
## [95,] "PHUTHIATSANA"               "Lesotho"      "16" "58.3" 
## [96,] "MOHALESHOEK"                "Lesotho"      "16" "137.6"
## [97,] "Kitale"                     "Kenya"        "16" "30.2" 
## [98,] "CANKUZO"                    "Burundi"      "17" "119.3"
## [99,] "SEMONKONG"                  "Lesotho"      "18" "64.7" 
## [100,] "Hwange"                    "Zimbabwe"     "18" "59.5" 
## [101,] "GISENYIAERO"               "Rwanda"       "18" "110.7"
## [102,] "Karonga"                   "Malawi"       "18" "107"  
## [103,] "Brandvlei"                 "South Africa" "19" "11.2" 
## [104,] "BUJUMBURA (Aeroport)"      "Burundi"      "20" "73.8" 
## [105,] "QUTHING"                   "Lesotho"      "20" "81"   
## [106,] "Mzuzu"                     "Malawi"       "20" "106.1"
## [107,] "Songea"                    "Tanzania"     "20" "153.8"
## [108,] "Dodoma"                    "Tanzania"     "21" "76.2" 
## [109,] "Moshi"                     "Tanzania"     "21" "128.4"
## [110,] "MPOTA (Tora)"              "Burundi"      "21" "80"   
## [111,] "Sumbawanga"                "Tanzania"     "22" "86.4" 
## [112,] "Kisii"                     "Kenya"        "22" "168.1"
## [113,] "Salima"                    "Malawi"       "22" "116.3"
## [114,] "Punda Maria"               "South Africa" "22" "40.6" 
## [115,] "Nkhotakota"                "Malawi"       "22" "141.1"
## [116,] "Chitipa"                   "Malawi"       "25" "102.7"
## [117,] "MEJAMETALANA"              "Lesotho"      "26" "90.2" 
## [118,] "Lodwar"                    "Kenya"        "26" "226"  
## [119,] "MOKHOTLONG"                "Lesotho"      "27" "76.2" 
## [120,] "MUYINGA"                   "Burundi"      "27" "122.9"
## [121,] "Tembwe"                    "Malawi"       "29" "213.7"
## [122,] "Laingsburg"                "South Africa" "31" "15.6" 
## [123,] "NYAMUSWAGA"                "Burundi"      "35" "133.1"
## [124,] "GISOZI"                    "Burundi"      "36" "100.2"
```

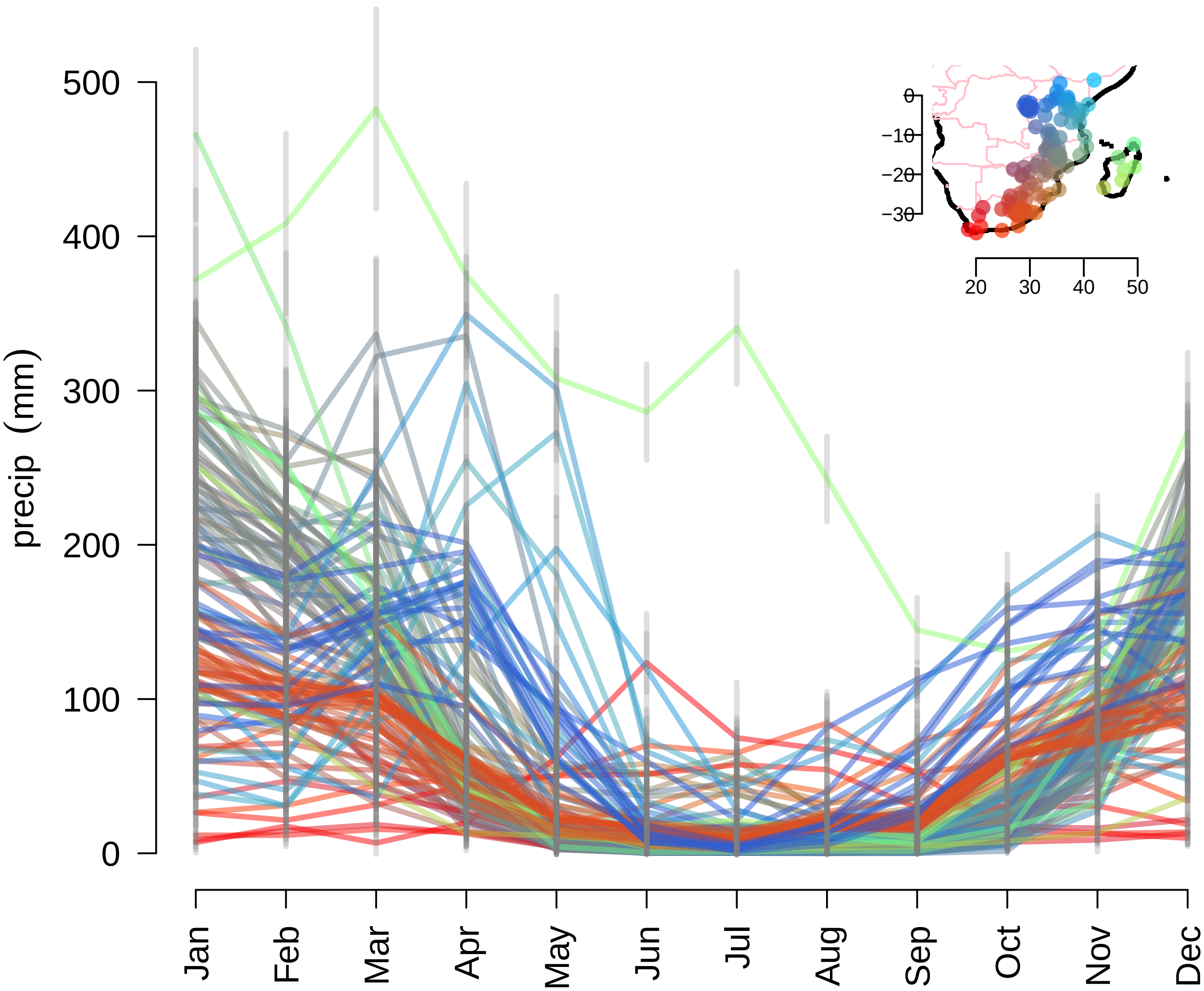
```
## [125,] "BUTHABUTHE"              "Lesotho"      "36" "76.4"
## [126,] "Warmbad Towoomba"        "South Africa" "41" "55.1"
## [127,] "Nkhatabay"               "Malawi"       "42" "228.9"
## [128,] "Mtwara"                  "Tanzania"     "42" "180.7"
## [129,] "Kericho"                 "Kenya"        "58" "139.8"
## [130,] "MUSASA"                  "Burundi"      "64" "269.6"
```

The comparison between climatologies based on rain gauges and the ERA5 renalysis for individual sites shows a large degree of similarities, but there are also some sites where the two data sets are inconsistent:

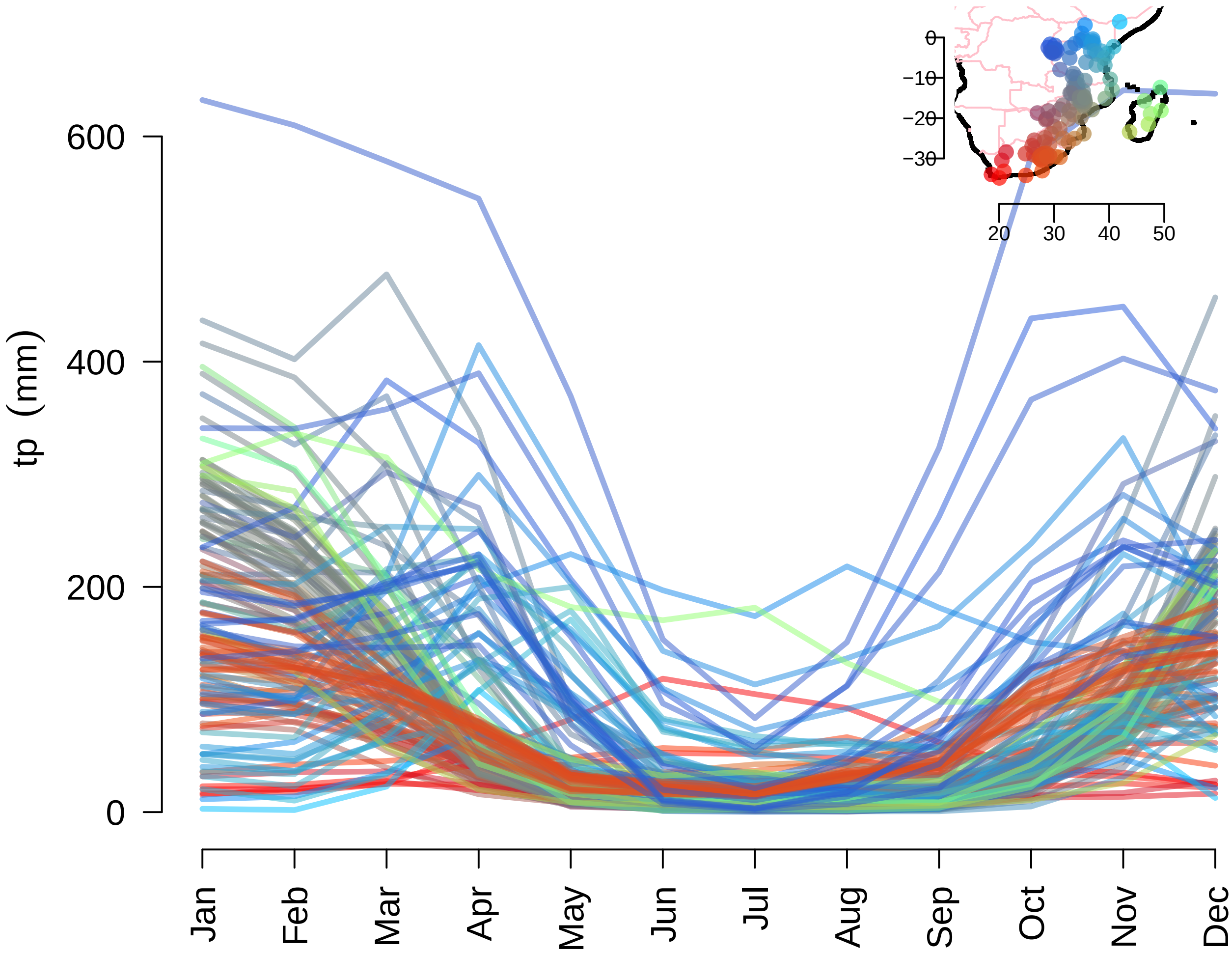
```
## Rain gauges
mac.obs <- aggregate(as.monthly(subset(X,it=c(1979,2021)),FUN='sum'),by=month)
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
plot(mac.obs,new=FALSE)
```




```
## Interpolated ERA5 data
mac.era5 <- aggregate(subset(x.mon,it=c(1979,2021)),by=month)
mac.era5 <- subset(mac.era5,is=is.element(loc(mac.era5),loc(mac.obs)))
plot(mac.era5,new=FALSE)
```




```
## Interpolated CHIRPS data
mac.chirps <- aggregate(as.monthly(subset(rr.chirps,it=c(1979,2021)),FUN='sum'),by=month)
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

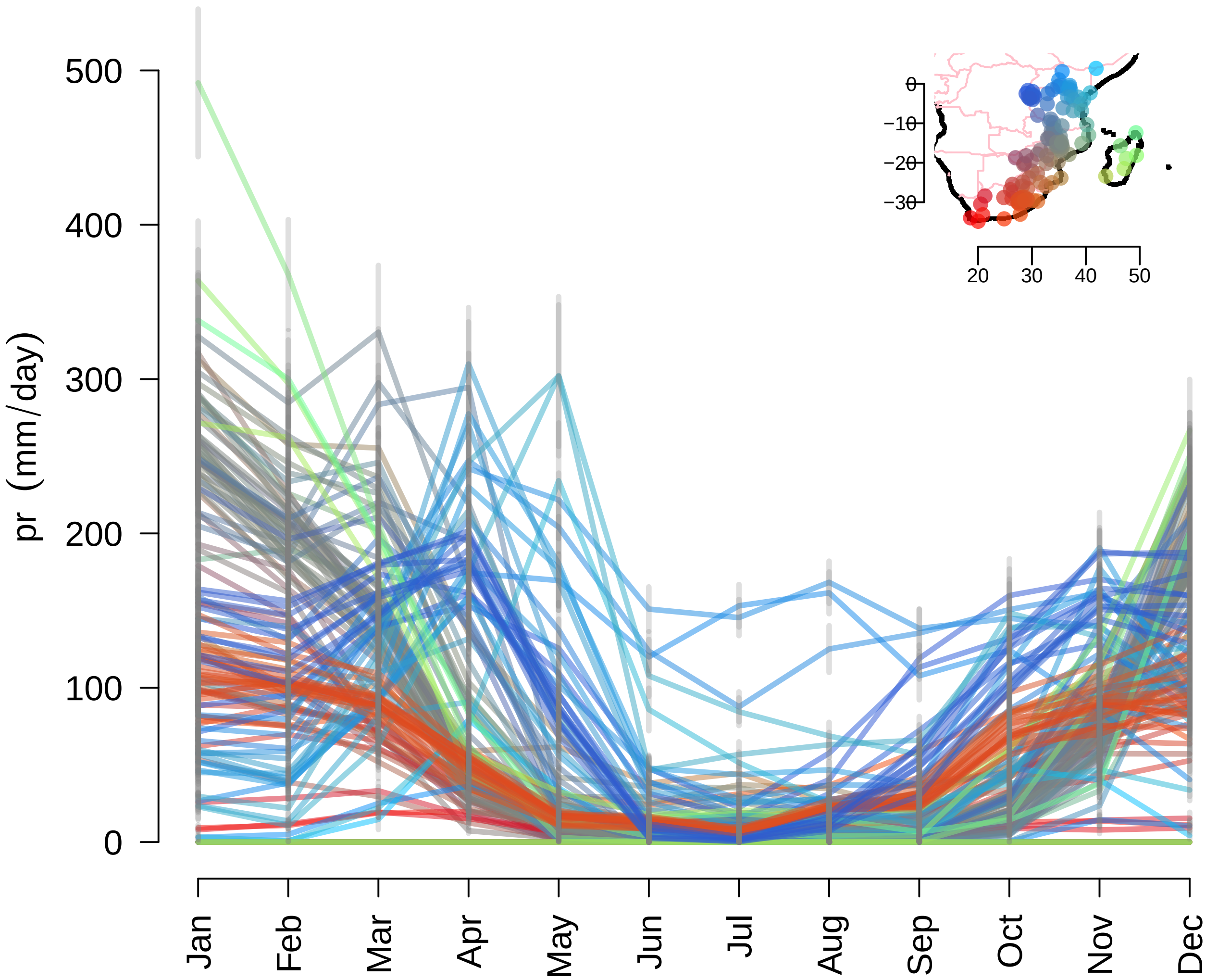
```
mac.chirps <- subset(mac.chirps,is=is.element(loc(mac.chirps),loc(mac.obs)))
set20 <- !is.finite(coredata(mac.chirps))
print(dim(mac.obs)); print(dim(mac.era5)); print(dim(mac.chirps))
```

```
## [1]  12 130
```

```
## [1]  12 130
```

```
## [1]  12 130
```

```
coredata(mac.chirps)[set20] <- 0
plot(mac.chirps,new=FALSE)
```



Skukuza (South Africa), Warmbad Towoomba (South Africa), Chitipa (Malawi), KIA (Malawi), Makoka (Malawi), Nkhatabay (Malawi), Chileka (Malawi), Mzuzu (Malawi), Kbay (Malawi), Balaka (Malawi), Ntaja (Malawi), KIGALIAERO (Rwanda), Mandera (Kenya), KAMEMBEAERO (Rwanda), Kitale (Kenya), Kericho (Kenya), Kisii (Kenya), Moi International Airpor (Kenya), Bukoba (Tanzania), Musoma (Tanzania), Same (Tanzania), Mwanza (Tanzania), Tabora (Tanzania), Moshi (Tanzania), Tanga (Tanzania), Dodoma (Tanzania), Dar es Salaam (Tanzania), Mtwara (Tanzania), Sumbawanga (Tanzania), Mahajanga (Madagascar) and Toamasina (Madagascar).

In many cases, there is a mismatch between the timing or the amplitude of the rainy seasons.

There may be various explanations for the mismatch between rain gauge observations, i.e. that the local measurements are influenced by local factors not present in the ERA5 simulations, different types of numbers (one almost point-source measurement and the other an average calculated for a ~31x31 km^2 grid box area), errors in the ERA5 product and errors in the readings/reporting of the local rainfall.

We can use principal component analysis (PCA; Wilks (2006)) to present regional patterns of different seasonal cycles in the rainfall:

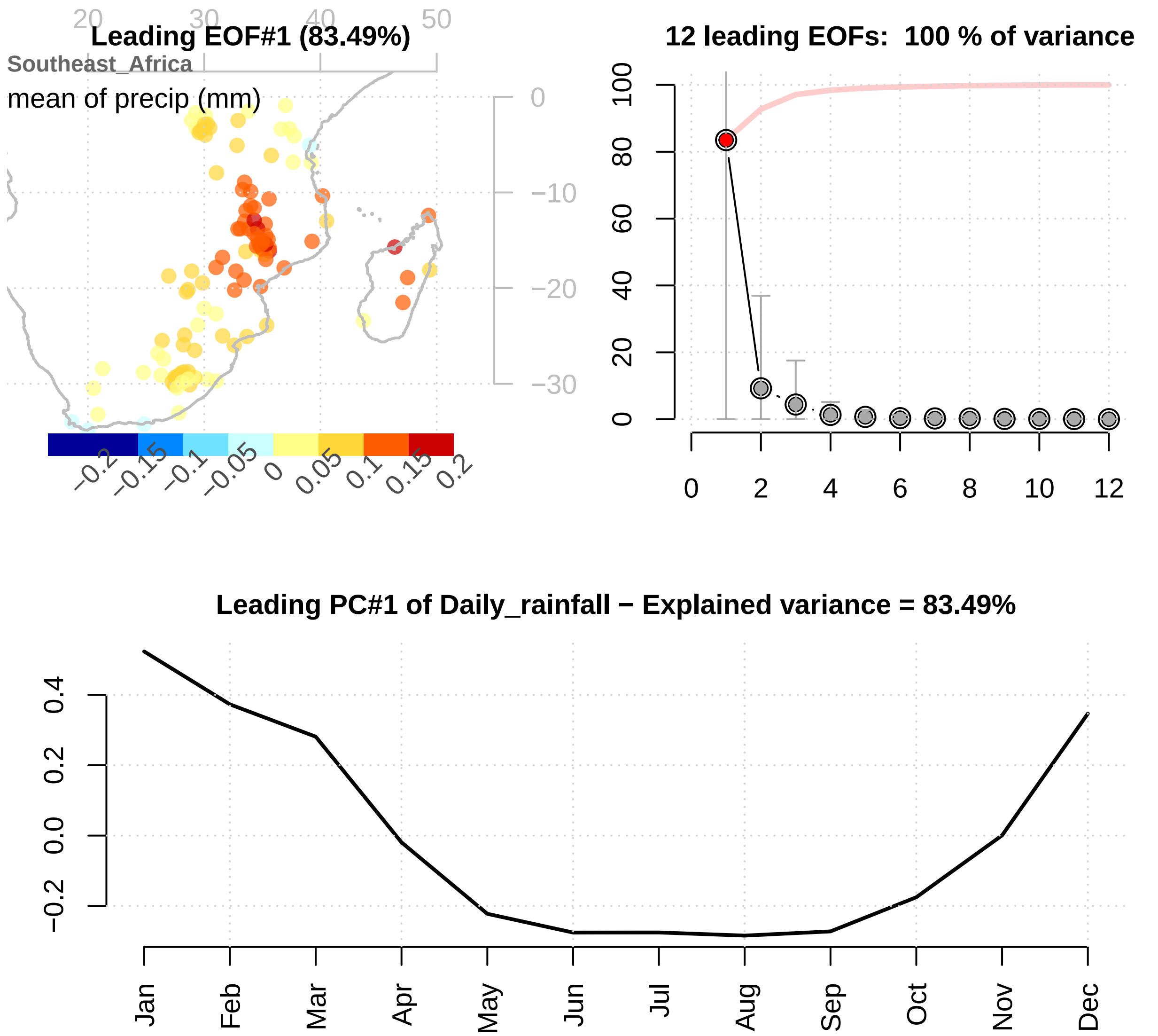
```
nv <- apply(mac.obs,2,'nv')
print(dim(mac.obs)); print(dim(mac.era5)); print(dim(mac.chirps))
```

```
## [1]  12 130
```

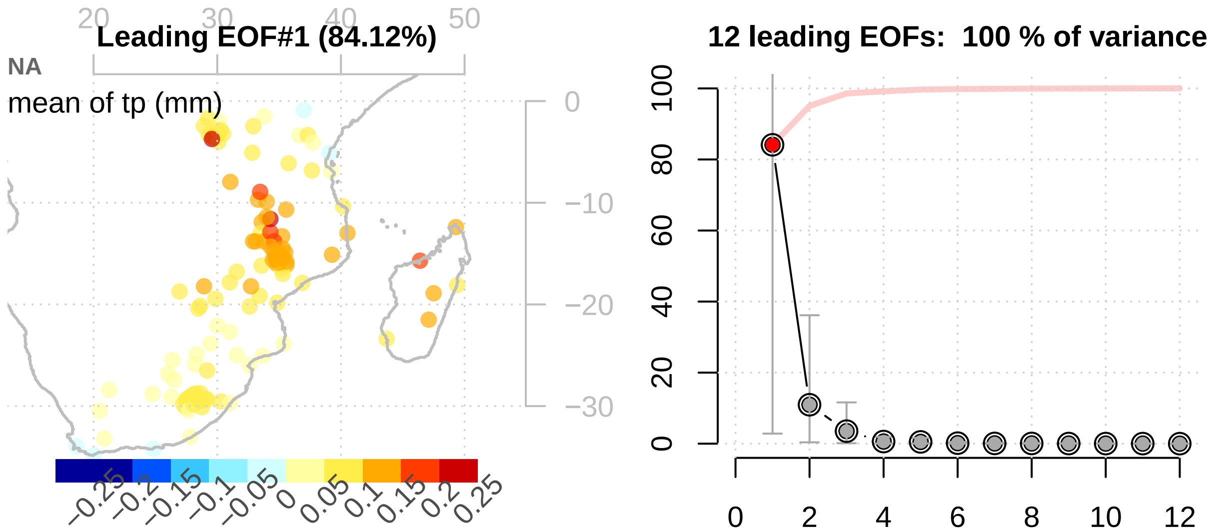
```
## [1]  12 130
```

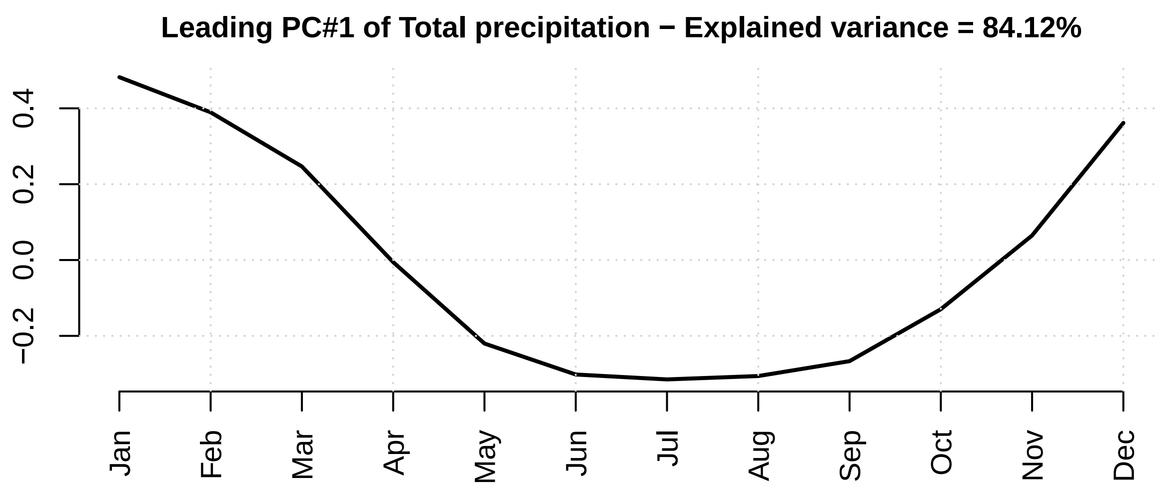
```
## [1]  12 130
```

```
mac.obs <- subset(mac.obs,is=nv==12)
mac.era5 <- subset(mac.era5,is=nv==12)
mac.chirps <- subset(mac.chirps,is=nv==12)
mac.chirps[!is.finite(mac.chirps)] <- 0 ## Fudge
pca.mac.obs <- PCA(mac.obs,n=12)
plot(pca.mac.obs,new=FALSE)
```

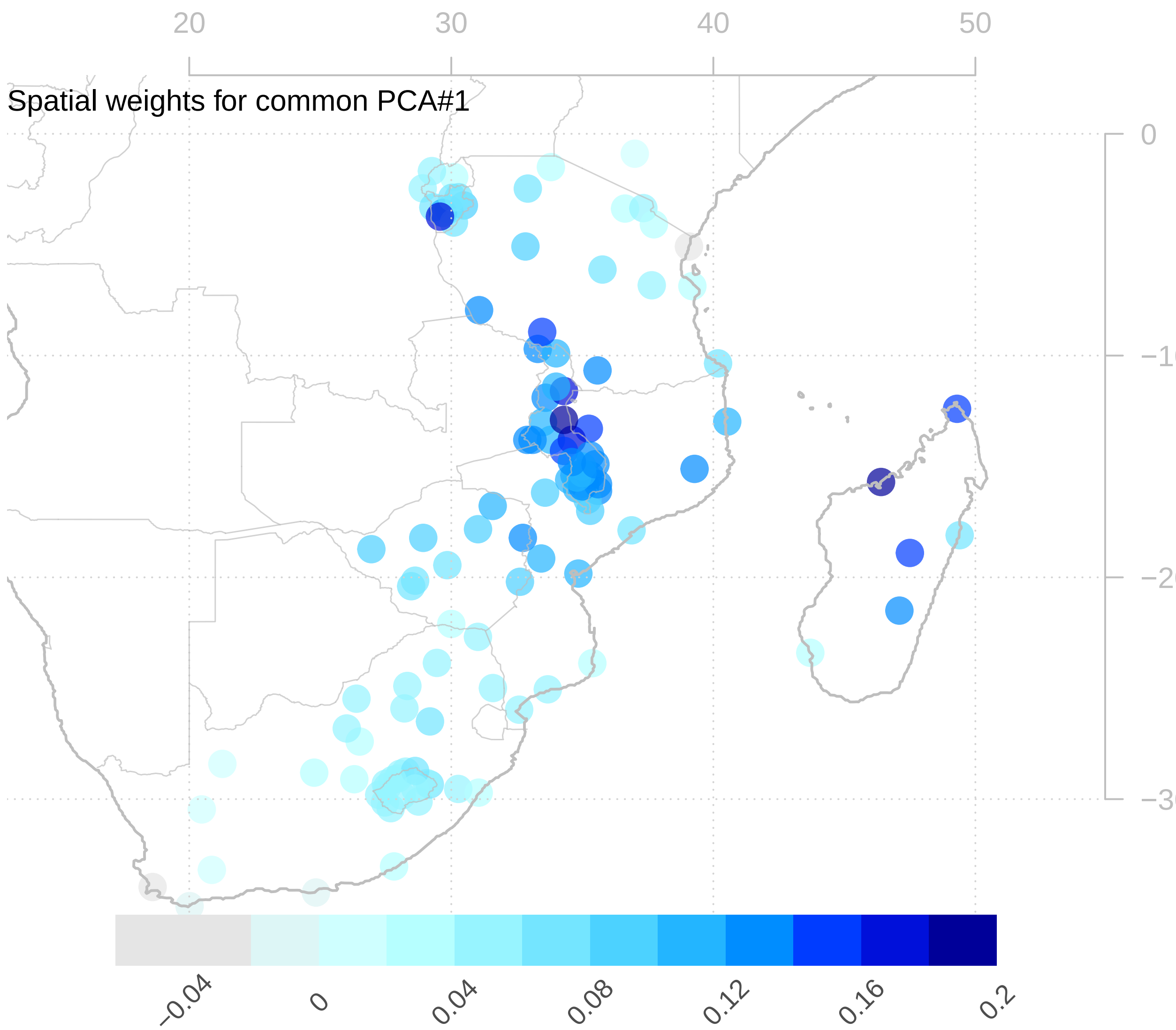



```
pca.mac.era5 <- PCA(mac.era5,n=12)
plot(pca.mac.era5,new=FALSE)
```

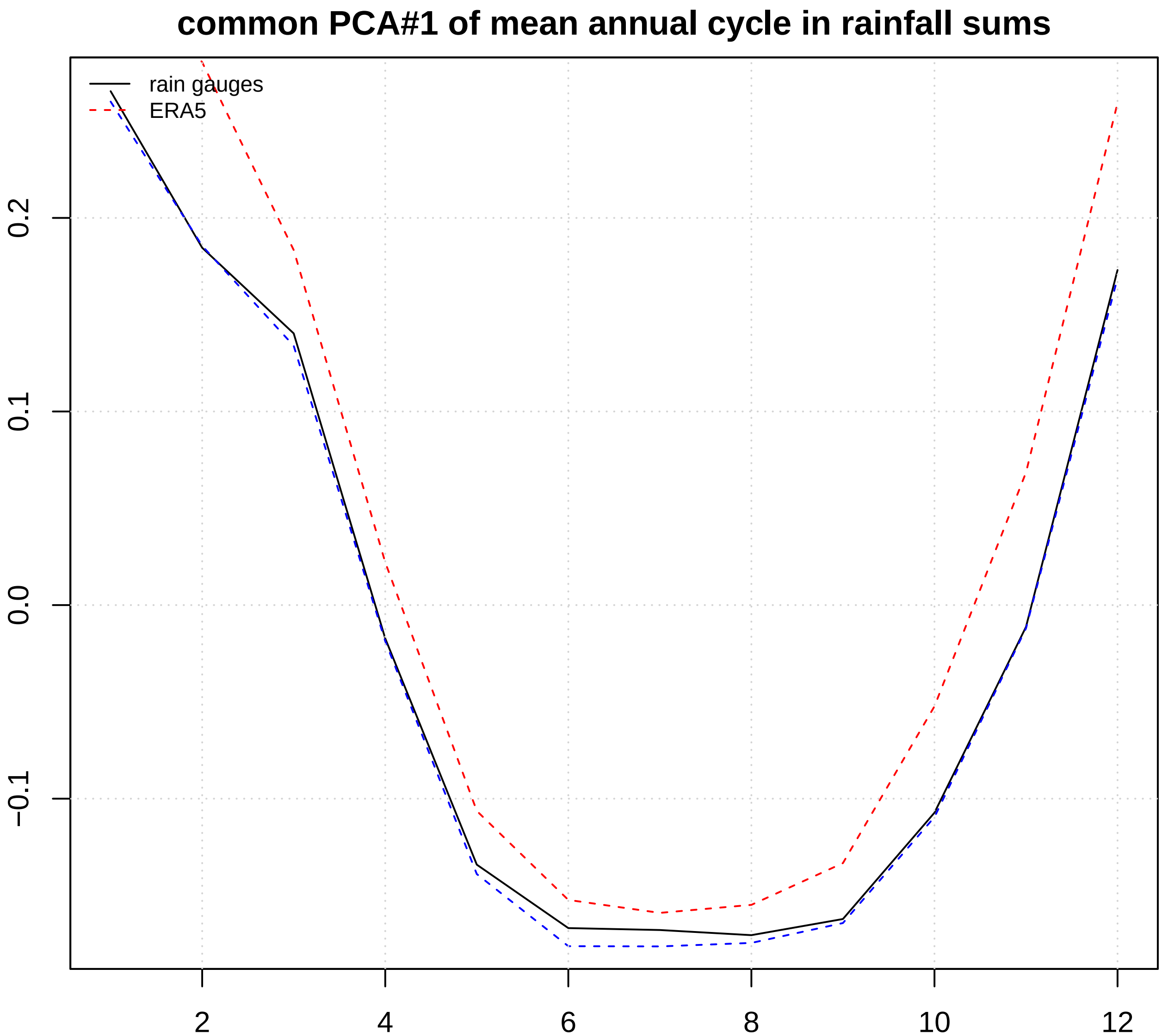





```
## 'Common' PCA
mac.both <- zoo(x=rbind(coredata(mac.obs),coredata(mac.era5),coredata(mac.chirps)),
                order.by=1:36)
mac.both <- attrcp(mac.obs,mac.both)
class(mac.both) <- class(mac.obs)
pca.both <- PCA(mac.both)
map(pca.both,main='Spatial weights for common PCA#1',border=TRUE)
```




```r
plot(zoo(pca.both[1:12,1]),main='common PCA#1 of mean annual cycle in rainfall sums',
     ylab='weight',xlab='Clendar month',cex.lab=0.5)
lines(zoo(coredata(pca.both)[13:24,1],order.by=1:12),col='red',lty=2)
lines(zoo(coredata(pca.both)[25:36,1],order.by=1:12),col='blue',lty=2)
grid()
legend('topleft',c('rain gauges','ERA5'),lty=c(1,2),col=c('black','red'),bty='n',cex=0.75)
```



The results of the PCA suggest mean seasonal cycles with similar geographical weights, variances and approximately similar temporal characteristics between the local rain gauge data and the ERA5 estimates. The ERA5 (red dashed) results indicate a slight wet bias.

The common PCA evaluation is inspired by the common EOF evaluation presented in Benestad et al. (2023). The post pronounced mean seasonal variations in the rainfall is seen around Malawi and on Madagascar. The leading PCA suggests that ERA5 captures most of the seasonal variations, but that ERA5 are suggests more rainfall during the rainy seasons.

2.1.2 Taylor diagram of annual data

A Taylor diagram can provide further information about the match between rain gauge data and ERA5, in this case the October–September annual rainfall.

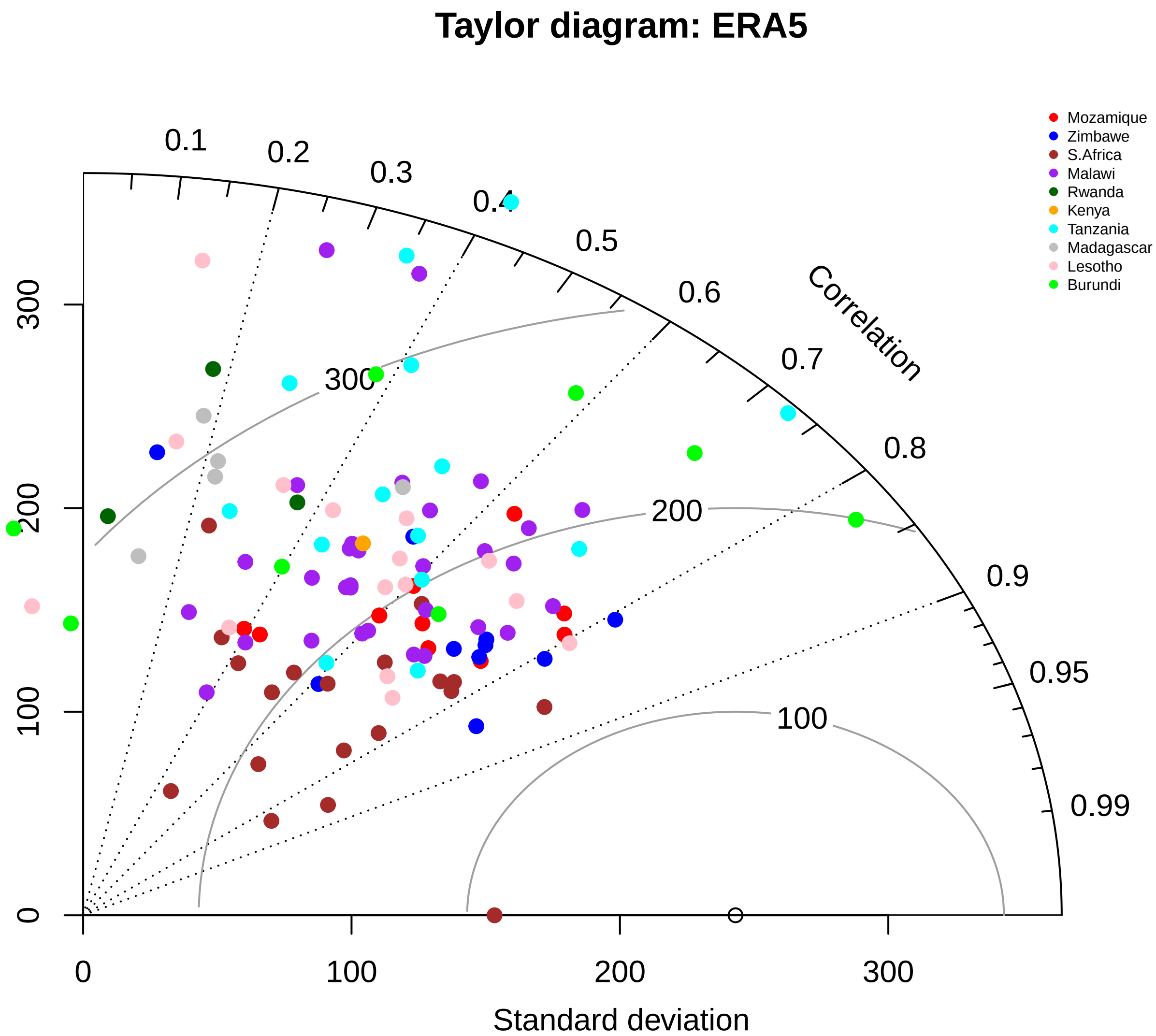
```
## BIG X holds the rain gauge data
ref <- matchdate(zoo(365.25*annual(X,start=year.start)),zoo(x))
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
## small x holds interpolated ERA5 data
mod <- matchdate(zoo(x),zoo(365.25*annual(X,start=year.start)))
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```r
cols <- cntr(X)
cols[cols=="Mozambique"] <- "red"
cols[cols=="Zimbabwe"] <- "blue"
cols[cols=="South Africa"] <- "brown"
cols[cols=="Malawi"] <- "purple"
cols[cols=="Rwanda"] <- "darkgreen"
cols[cols=="Kenya"] <- "orange"
cols[cols=="Tanzania"] <- "cyan"
cols[cols=="Madagascar"] <- "grey"
cols[cols=="Lesotho"] <- "pink"
cols[cols=="Burundi"] <- "green"
## Use the series with largest magnitude as a starting point
ix <- order(apply(mod,2,max),decreasing=TRUE)[1]
taylor.diagram(ref[,ix],mod[,ix],col=cols[ix],main='Taylor diagram: ERA5')
for (i in 1:dim(ref)[2]) taylor.diagram(ref[,i],mod[,i],col=cols[i],add=TRUE)
legend('topright',c('Mozamique','Zimbawe','S.Africa','Malawi','Rwanda','Kenya',
                    'Tanzania','Madagascar','Lesotho','Burundi'),cex=0.5,
       col=c('red','blue','brown','purple','darkgreen','orange','cyan','grey'
             ,'pink','green'),bty='n',pch=19)
```



The Taylor diagram also suggests that some rain gauge data from Malawi, Kenya, South Africa and Rwanda match poorly with ERA5. Furthermore, there is a substantial scatter in the match quality of rain gauge measurements within each country.

```
## BIG X holds the rain gauge data
ref2 <- matchdate(zoo(365.25*annual(X,start=year.start)),zoo(RR.chirps))
```

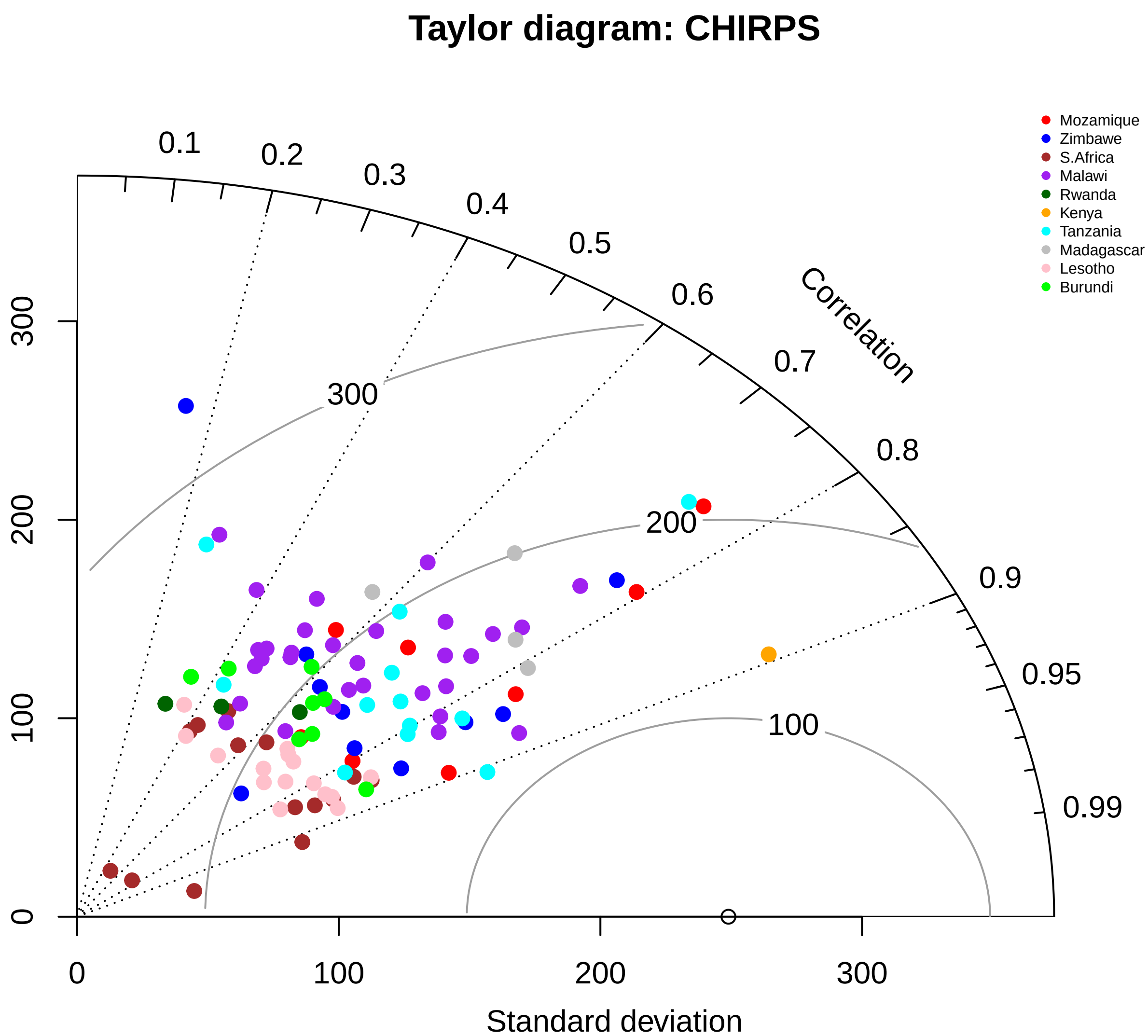
```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
## small x holds interpolated ERA5 data
mod2 <- matchdate(zoo(RR.chirps),zoo(365.25*annual(X,start=year.start)))
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

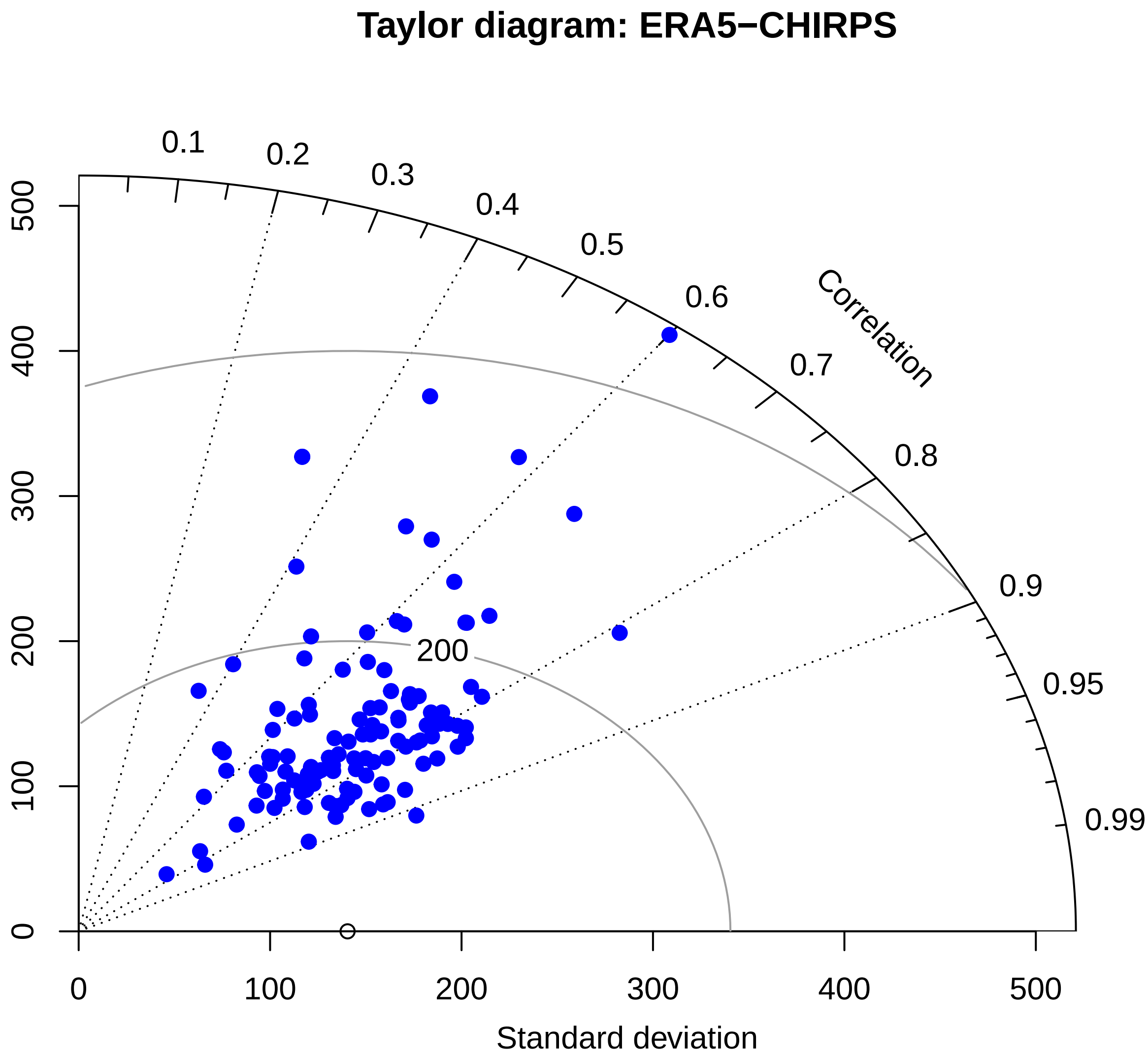
```
## Ensure same site
commonsites <- intersect(loc(X),loc(RR.chirps))
ref2 <- ref2[,is.element(loc(X),commonsites)]
mod2 <- mod2[,is.element(loc(RR.chirps),commonsites)]
cols2 <- cntr(subset(X,is=is.element(loc(X),commonsites)))
cols2[cols2=="Mozambique"] <- "red"
cols2[cols2=="Zimbabwe"] <- "blue"
cols2[cols2=="South Africa"] <- "brown"
```

```
cols2[cols2=="Malawi"] <- "purple"
cols2[cols2=="Rwanda"] <- "darkgreen"
cols2[cols2=="Kenya"] <- "orange"
cols2[cols2=="Tanzania"] <- "cyan"
cols2[cols2=="Madagascar"] <- "grey"
cols2[cols2=="Lesotho"] <- "pink"
cols2[cols2=="Burundi"] <- "green"
## Use the series with largest magnitude as a starting point
ix2 <- order(apply(mod2,2,max),decreasing=TRUE)[1]
taylor.diagram(ref2[,ix2],mod2[,ix2],col=cols2[ix2],main='Taylor diagram: CHIRPS')
for (i in 1:dim(ref2)[2]) taylor.diagram(ref2[,i],mod2[,i],col=cols2[i],add=TRUE)
legend('topright',c('Mozamique','Zimbawe','S.Africa','Malawi','Rwanda','Kenya',
                    'Tanzania','Madagascar','Lesotho','Burundi'),cex=0.5,
       col=c('red','blue','brown','purple','darkgreen','orange','cyan','grey'
             ,'pink','green'),bty='n',pch=19)
```




```
## BIG X holds daily rain gauge data
ref3 <- matchdate(zoo(RR.chirps),zoo(x))
## small x holds interpolated ERA5 data
mod3 <- matchdate(zoo(x),zoo(RR.chirps))
```

```r
## Ensure same site
commonsites <- intersect(loc(RR.chirps),loc(x))
ref3 <- ref3[,is.element(loc(RR.chirps),commonsites)]
mod3 <- mod3[,is.element(loc(x),commonsites)]
cols3 <- "blue"

## Use the series with largest magnitude as a starting point
ix3 <- order(apply(mod3,2,max),decreasing=TRUE)[1]
taylor.diagram(ref3[,ix3],mod3[,ix3],col=cols3,main='Taylor diagram: ERA5-CHIRPS')
for (i in 1:dim(ref3)[2]) taylor.diagram(ref3[,i],mod3[,i],col=cols3,add=TRUE)
```

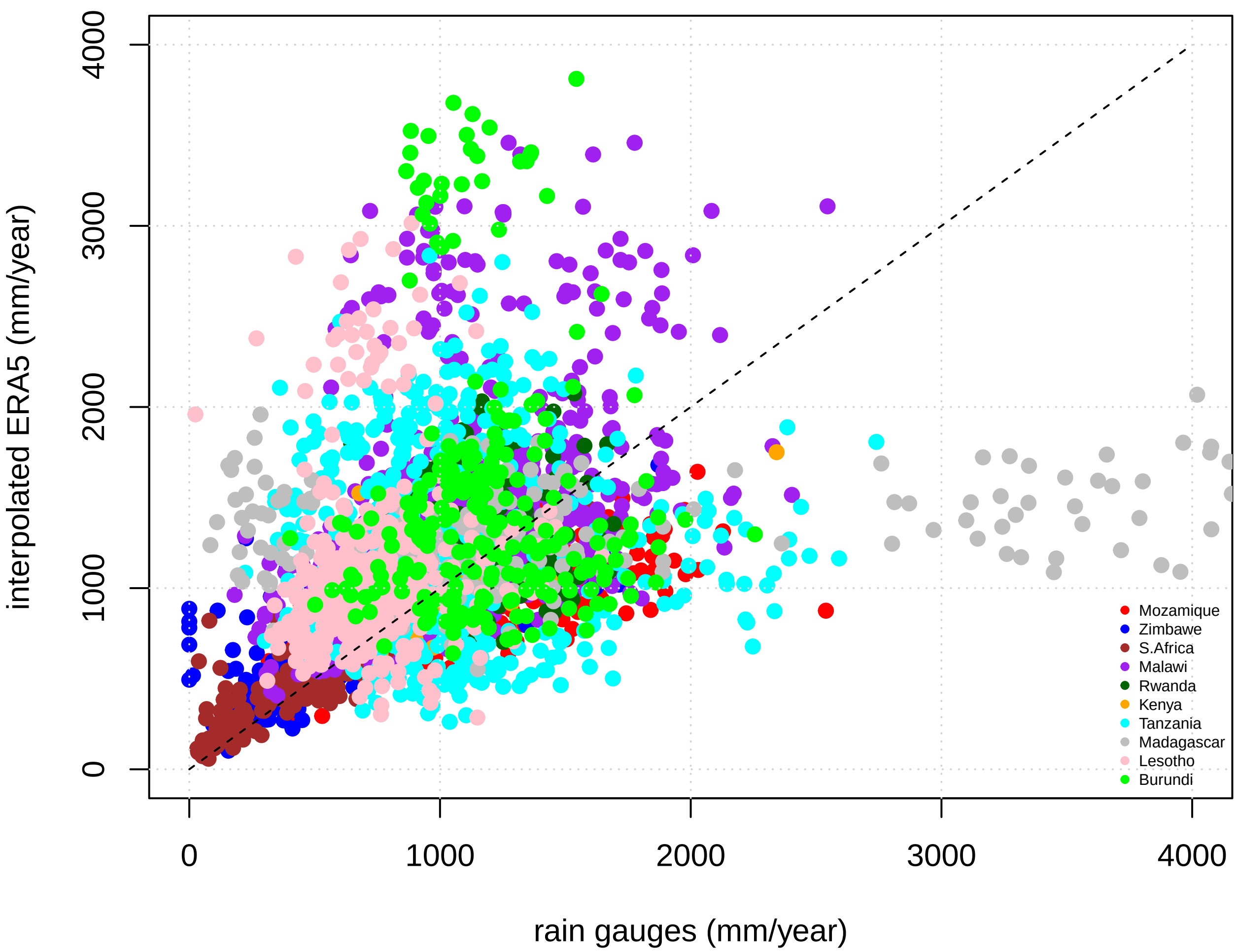


2.1.3 Scatter plot of annual rainfall

A scatter plot between October–September annual rainfall can provide further information about how closely they match.

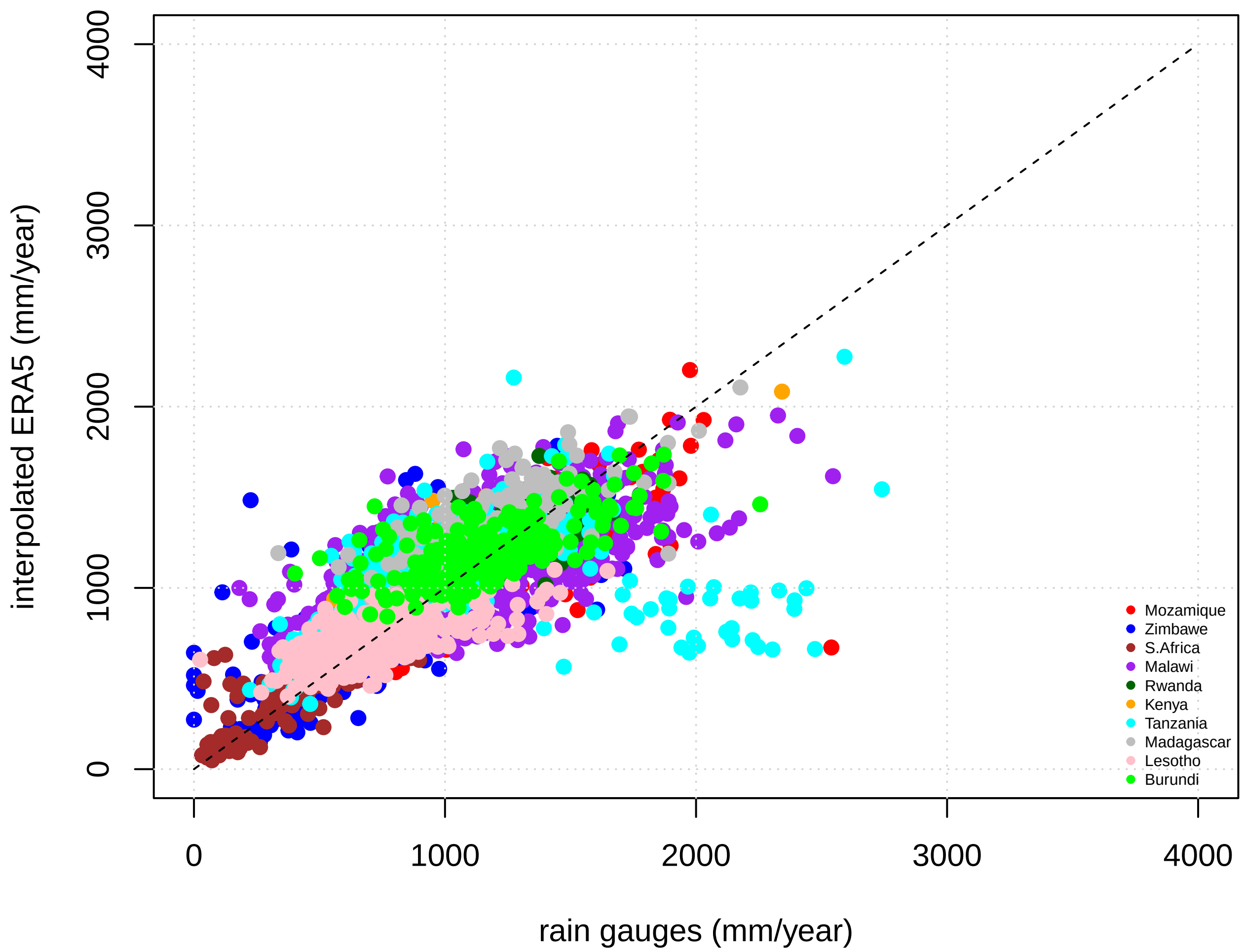
```r
plot(c(ref[,1]),c(mod[,1]),col=cols[1],xlim=c(0,4000),ylim=c(0,4000),pch=19,
     xlab='rain gauges (mm/year)',ylab='interpolated ERA5 (mm/year)')
for (i in 2:dim(ref)[2]) points(ref[,i],mod[,i],col=cols[i],pch=19)
grid()
```

```
lines(c(0,4000),c(0,4000),lty=2)
legend('bottomright',c('Mozamique','Zimbawe','S.Africa','Malawi','Rwanda','Kenya',
                       'Tanzania','Madagascar','Lesotho','Burundi'),cex=0.5,
       col=c('red','blue','brown','purple','darkgreen','orange','cyan','grey',
             'pink','green'),bty='n',pch=19)
```



The results indicate most pronounced differences for some of the annual rainfall totals from some rain gauges in Rwanda and Malawi where ERA5 indicates substantially higher amounts. Also there are some wet annual rainfall totals on Madagascar for which ERA5 indicates lower amounts. The data from Tanzania exhibited a fairly wide scatter with respect to ERA5.

```
plot(c(ref2[,1]),c(mod2[,1]),col=cols2[1],xlim=c(0,4000),ylim=c(0,4000),pch=19,
     xlab='rain gauges (mm/year)',ylab='interpolated ERA5 (mm/year)')
for (i in 2:dim(ref2)[2]) points(ref2[,i],mod2[,i],col=cols2[i],pch=19)
grid()
lines(c(0,4000),c(0,4000),lty=2)
legend('bottomright',c('Mozamique','Zimbawe','S.Africa','Malawi','Rwanda','Kenya',
                       'Tanzania','Madagascar','Lesotho','Burundi'),cex=0.5,
       col=c('red','blue','brown','purple','darkgreen','orange','cyan','grey',
             'pink','green'),bty='n',pch=19)
```



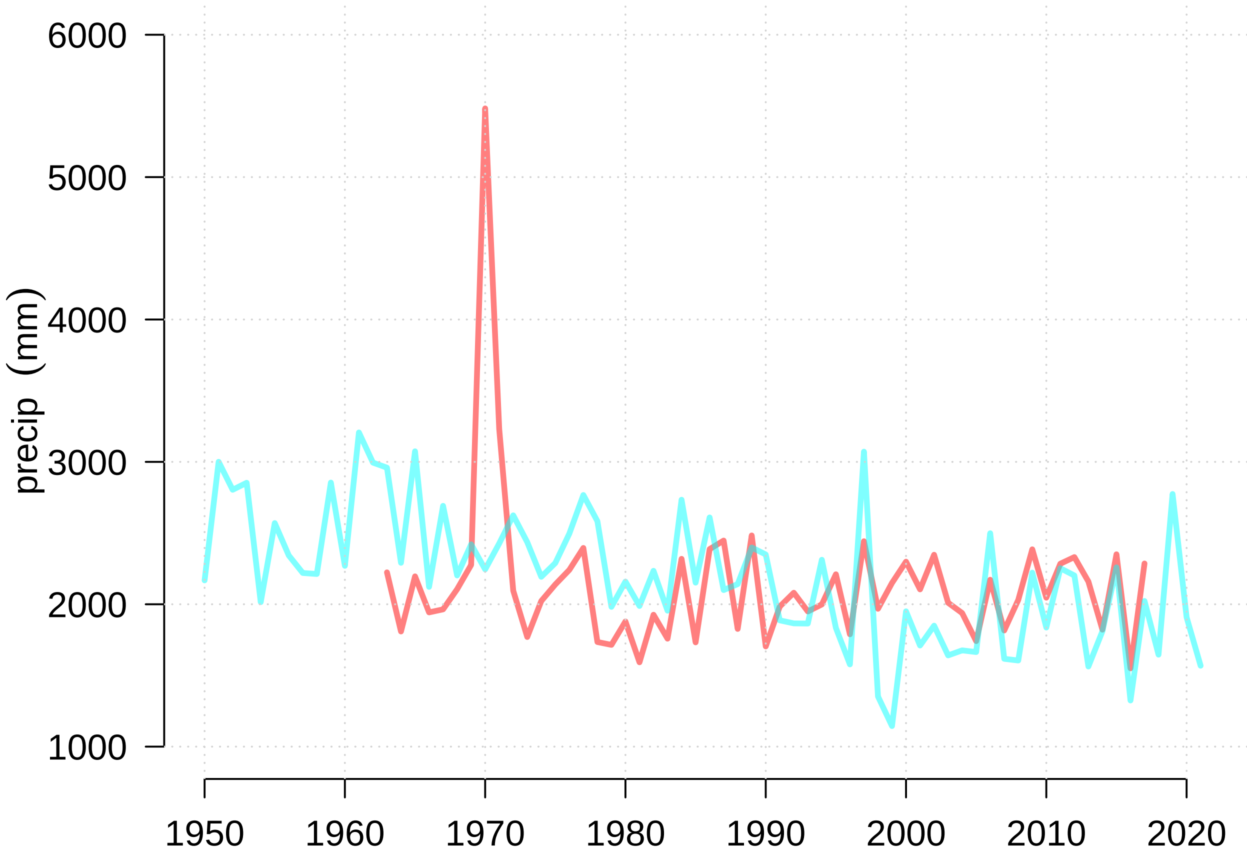
2.1.4 Closer evaluation of outlier data

We can look at some of the outliers such as Kisii in Kenya with exceptional rainfall in 1971 and compare the time series with ERA5:

```
kisii.raingauge <- 365.25*annual(subset(X,is=list(loc='Kisii')),FUN='mean',
                                   nmin=100,start=year.start)
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
index(kisii.raingauge) <-year(kisii.raingauge)
kisii.era5 <- subset(x,is=list(loc='Kisii'))
kisii <- combine.stations(kisii.raingauge,kisii.era5)
plot(kisii,new=FALSE,map.show=FALSE); grid()
```

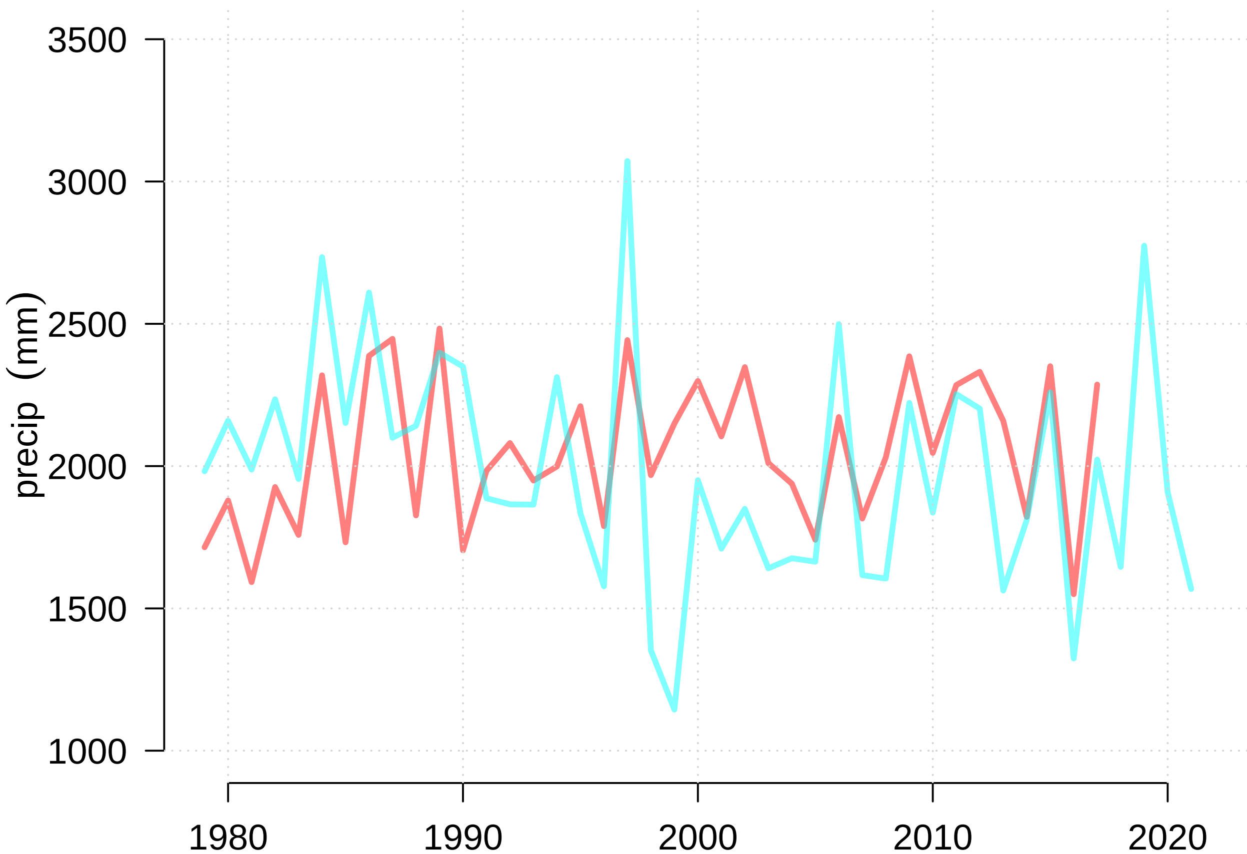


Year 1970 looged a spike in the annual rainfall that can not be seen in ERA5. If we limit our data to the period 1979–2022 we get the following results:

```
X <- subset(X,it=c(1979,2022))
x <- subset(x,it=c(1979,2022))
kisii.raingauge <- 365.25*annual(subset(X,is=list(loc='Kisii')),
                                  FUN='mean',nmin=100,start=year.start)
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
index(kisii.raingauge) <-year(kisii.raingauge)
kisii.era5 <- subset(x,is=list(loc='Kisii'))
kisii <- combine.stations(kisii.raingauge,kisii.era5)
plot(kisii,new=FALSE,map.show=FALSE); grid()
```



The two indicate a similar level but the year-to-year variations still don't match well.

2.1.5 Station-by-station evaluation of annual rainfall

In our evaluation we compared the rain gauge data with corresponding data from ERA5 interpolated to same coordinates by computing the root-mean-squared-error (rmse), the correlation, and the mean offset (bias $= \overline{x_{obs}} - \overline{x_{ERA5}}$).

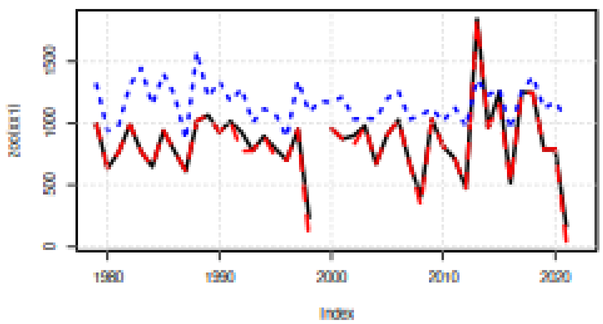
```
## Pick the annual precipitation from the first station
ns <- length(loc(X))
print(paste(ns,'stations'))
```

```
## [1] "130 stations"
```

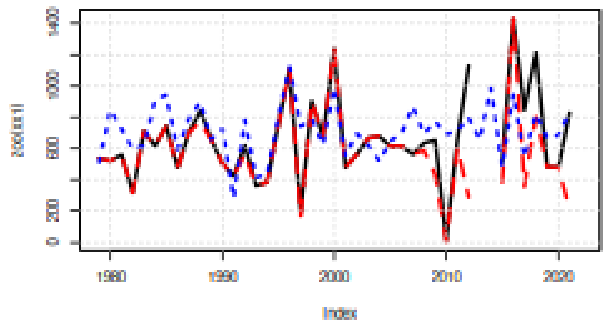
```
rmse <- rep(NA,ns); attr(rmse,'location') <- loc(X)
r <- rmse; offset <- rmse
rmsey <- rmse; ry <- rmse; offsety <- rmse
par(mfcol=c(4,2),cex=0.5)
for (is in 1:ns) {
  ## Because of missing data/days, we get more robust results by estimating the mean
  ## and then multiply by number of days per year.
  X1 <- 365.25*annual(subset(X,is=is),
                      FUN='mean',nmin=90,start=year.start)   # Station data - capital letters
  ## But if the lacking data are for the dry seasons, then it's better to choose the sum
  Y1 <- annual(subset(X,is=is),
               FUN='sum',nmin=90,start=year.start)           # Station data - capital letters
  x1 <- subset(x,is=is)                           # ERA5 reanalysis - small letters

  ## Estimate statistics such as RMSE, and correlation
```

```
  ## Here we merge station data with ERA5 reanalysis
  xy <- merge(zoo(X1[is.finite(X1)]),zoo(x1),all=FALSE)
  ok <- is.finite(xy[,1]) & is.finite(xy[,2])
  rmse[is] <- round(RMSE(xy[ok,1],xy[ok,2]))
  r[is] <- round(cor(xy[ok,1],xy[ok,2]),2)
  offset[is] <- round(mean(xy[ok,1]) - mean(xy[ok,2]))  ## Station mean - ERA5 mean
  xy <- merge(zoo(Y1[is.finite(X1)]),zoo(x1),all=FALSE)
  rmsey[is] <- round(RMSE(xy[ok,1],xy[ok,2]))
  ry[is] <- round(cor(xy[ok,1],xy[ok,2]),2)
  offsety[is] <- round(mean(xy[ok,1]) - mean(xy[ok,2]))  ## Station mean - ERA5 mean

  ## Show time series for each station: compare the two estimates with ERA5 data
  xx1 <- combine.stations(X1,Y1,x1)
  plot(zoo(xx1),main=paste0(loc(X1),' ',cntr(X1), ' RMSE=',rmse[is],'mm cor=',
                          r[is],' offset=',offset[is],'mm/year'),plot.type='single',
      lty=c(1,2,3),lwd=2,col=c('black','red','blue'))
  grid()
}
```

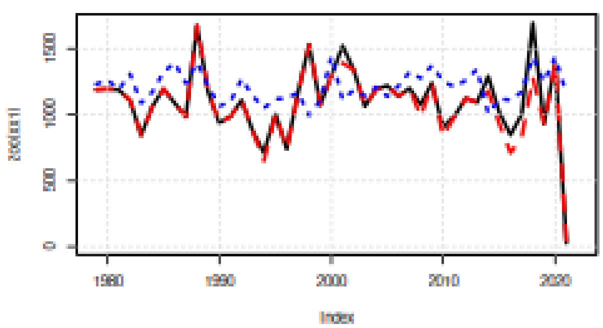
Pemba Mozambique RMSE=388mm cor=0.58 offset=-313mm/year



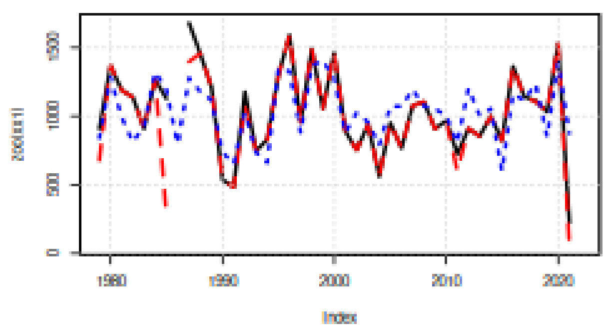
Tete Mozambique RMSE=228mm cor=0.58 offset=-53mm/year



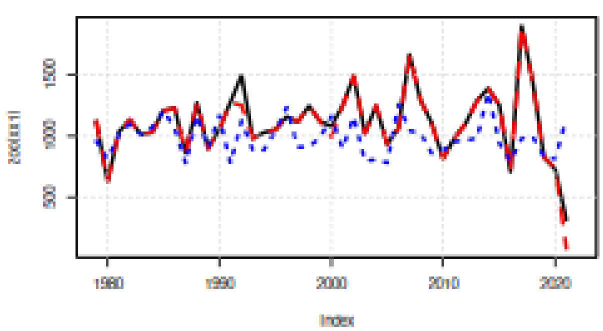
Lichinga Mozambique RMSE=275mm cor=0.36 offset=-110mm/year



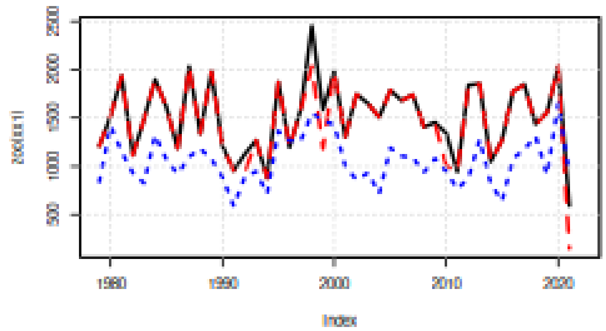
Chimoio Mozambique RMSE=204mm cor=0.75 offset=-5mm/year



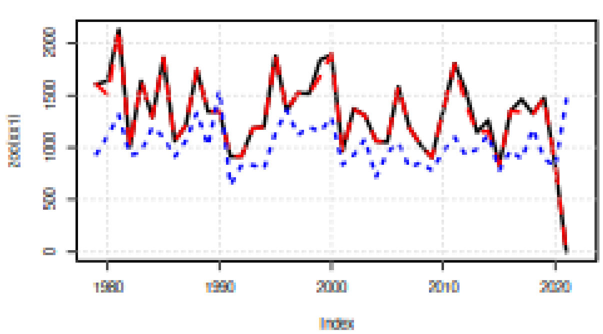
Nampula Mozambique RMSE=294mm cor=0.31 offset=125mm/year



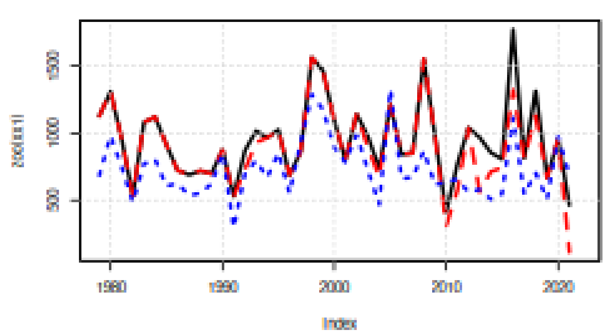
BeiraObs Mozambique RMSE=539mm cor=0.64 offset=458mm/year



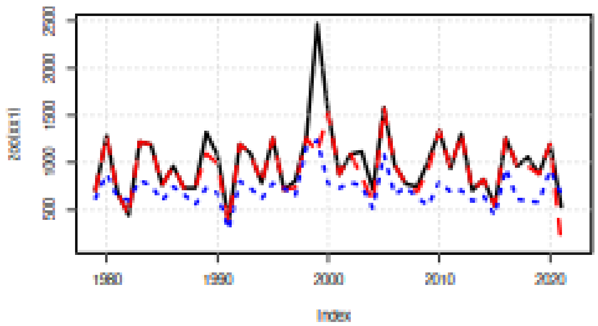
Quelimane Mozambique RMSE=476mm cor=0.32 offset=305mm/year



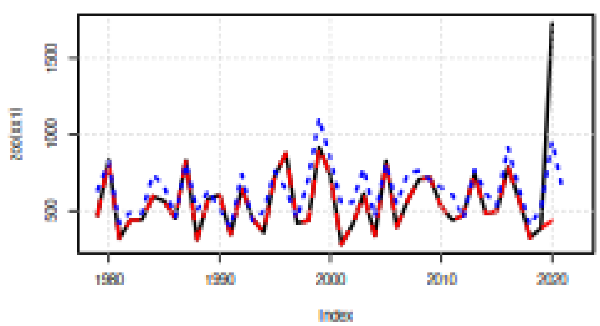
Inhambane Mozambique RMSE=290mm cor=0.74 offset=215mm/year



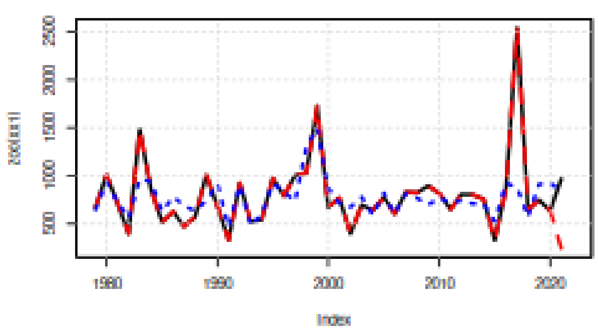
Xai-Xai Mozambique RMSE=376mm cor=0.81 offset=287mm/year



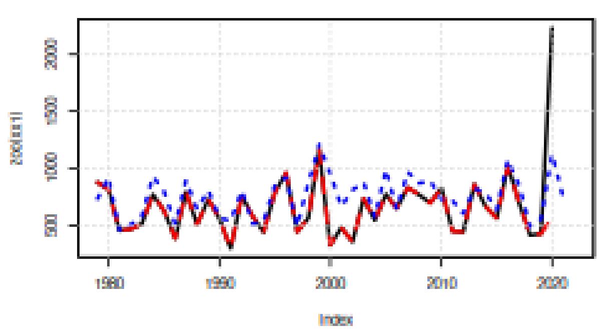
Bulawayo Goetz Zimbabwe RMSE=171mm cor=0.78 offset=-61mm/year



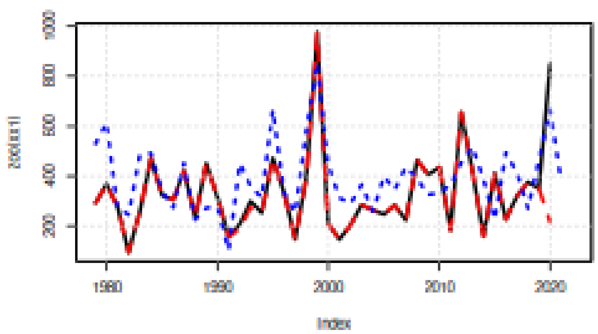
MaputoObs Mozambique RMSE=300mm cor=0.61 offset=26mm/year



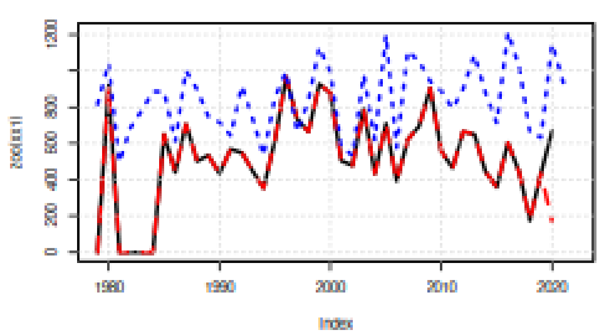
Gweru Zimbabwe RMSE=243mm cor=0.69 offset=-82mm/year



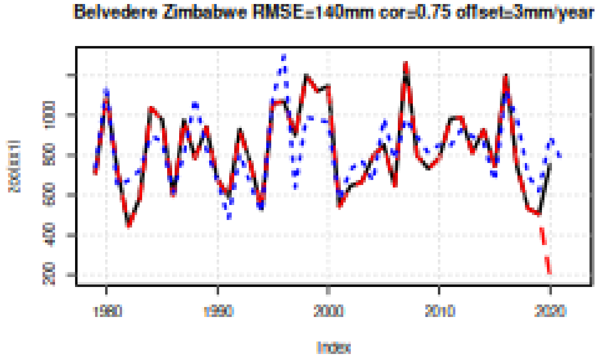
Beitbridge Zimbabwe RMSE=146mm cor=0.64 offset=-57mm/year



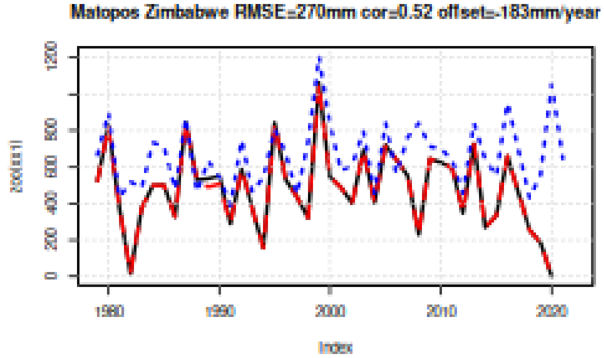
Hwange Zimbabwe RMSE=390mm cor=0.55 offset=-322mm/year

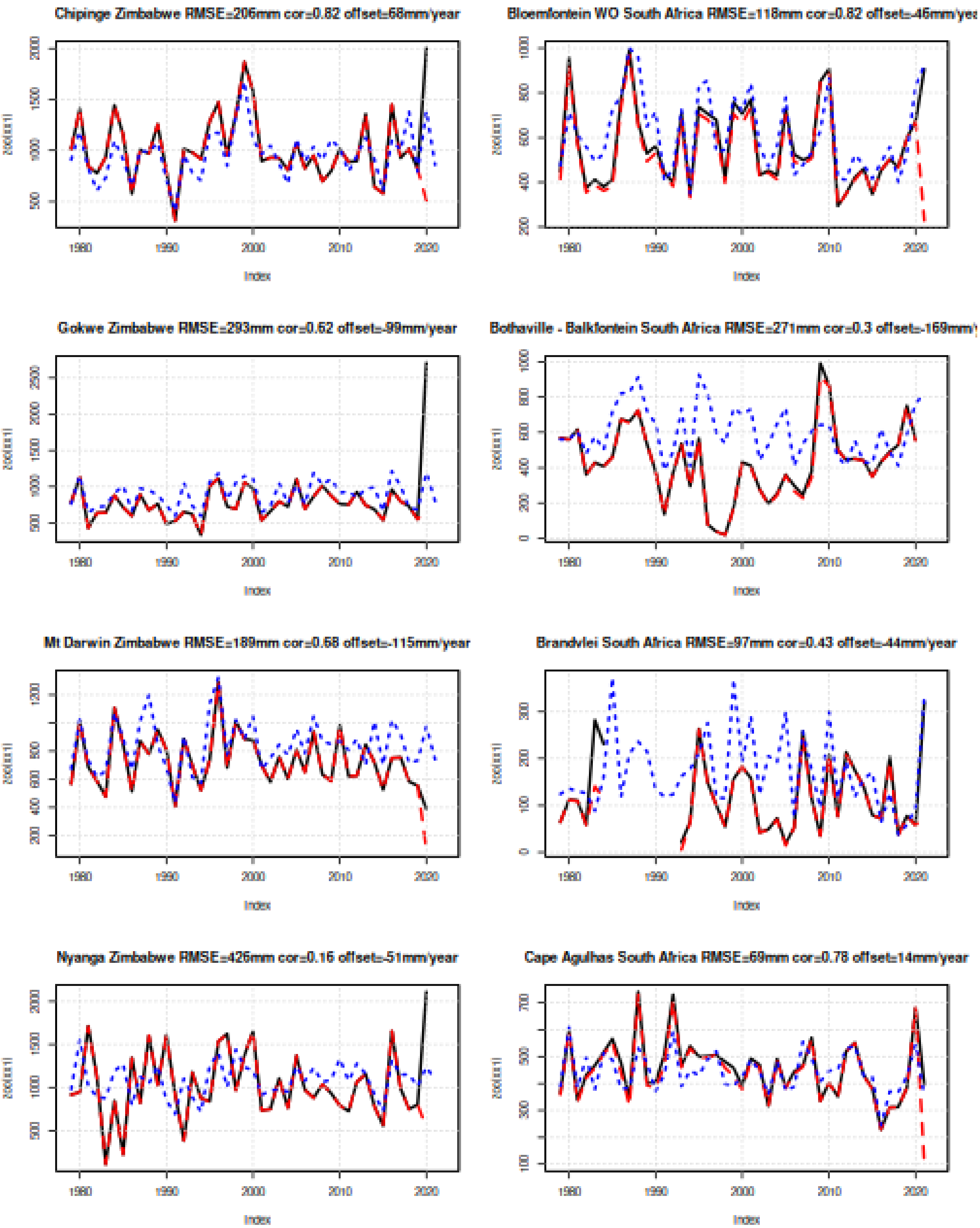


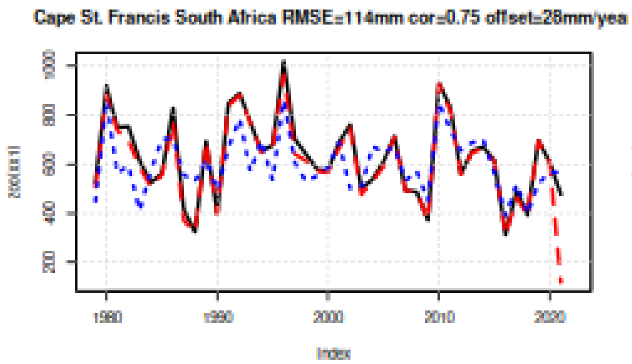
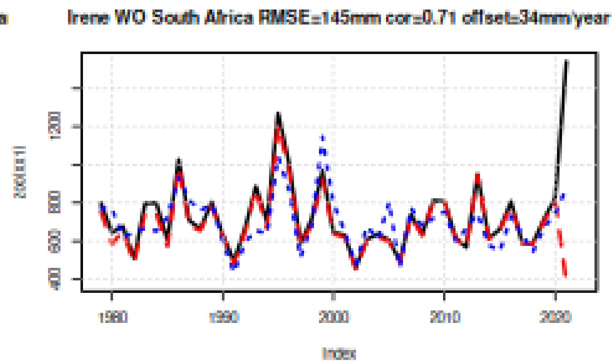
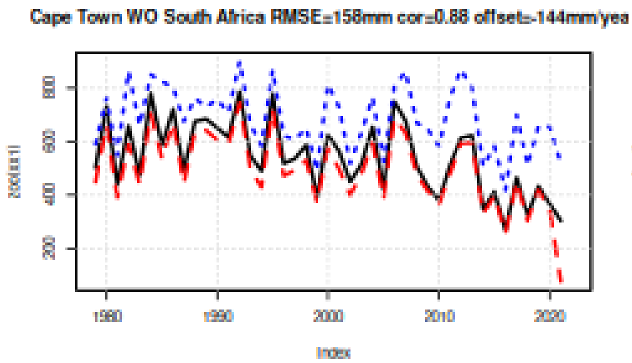
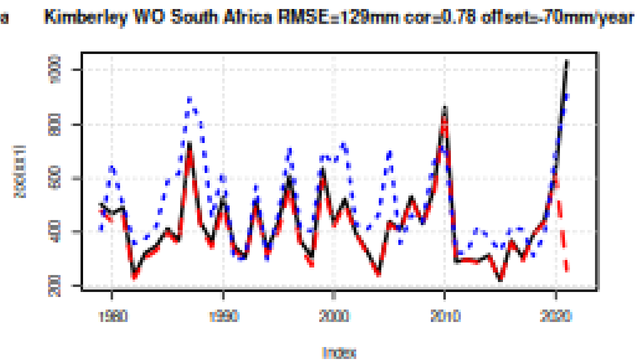
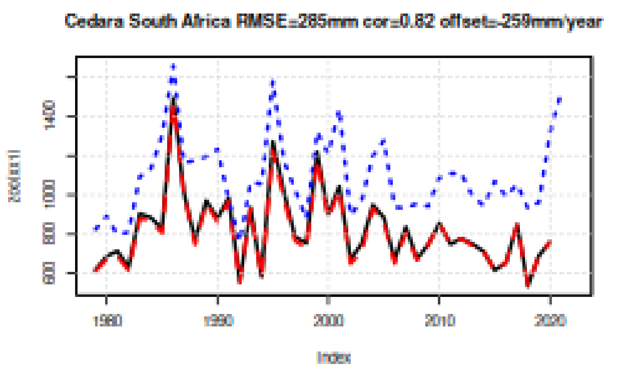
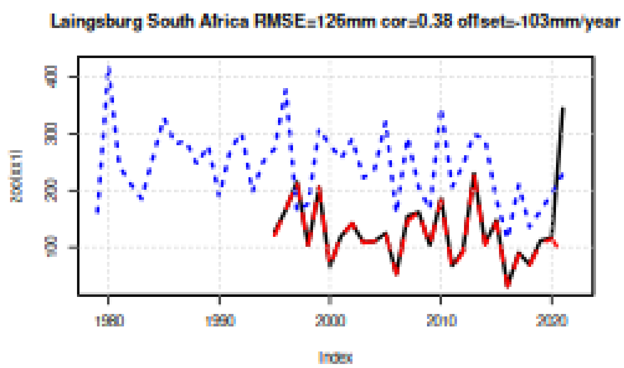
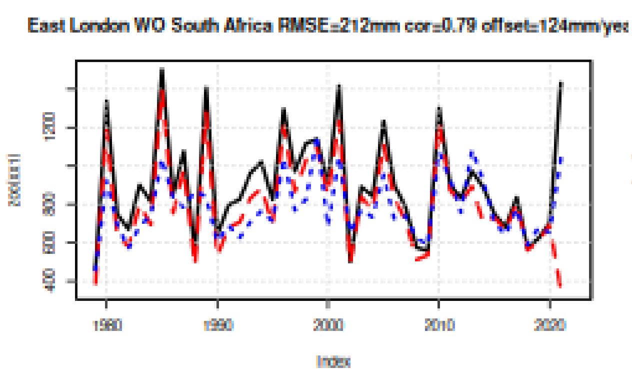
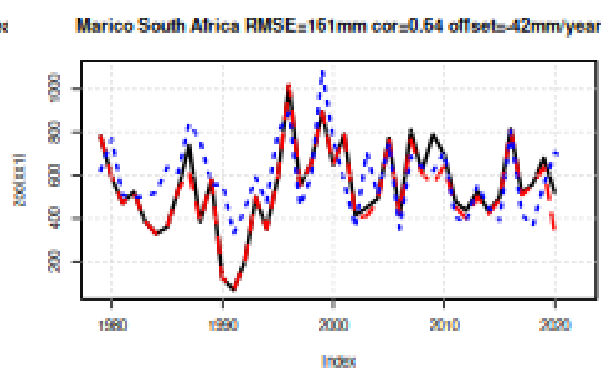
Belvedere Zimbabwe RMSE=140mm cor=0.75 offset=3mm/year

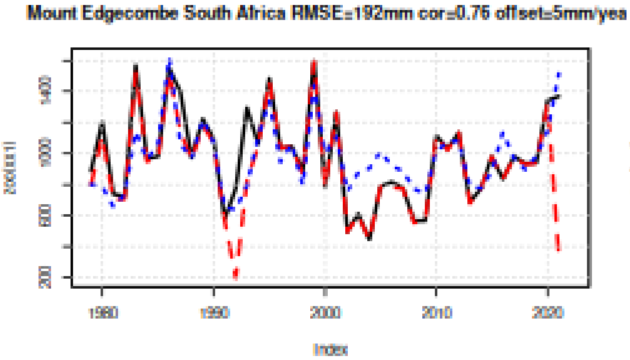


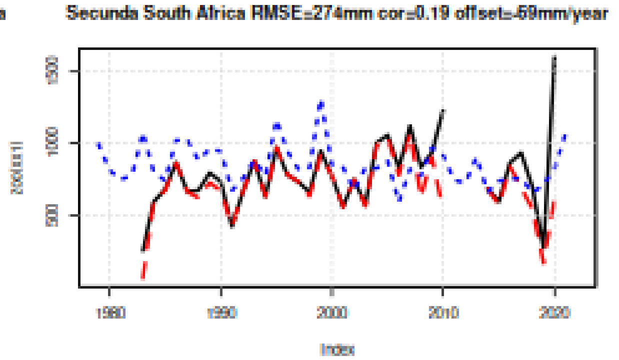
Matopos Zimbabwe RMSE=270mm cor=0.52 offset=-183mm/year

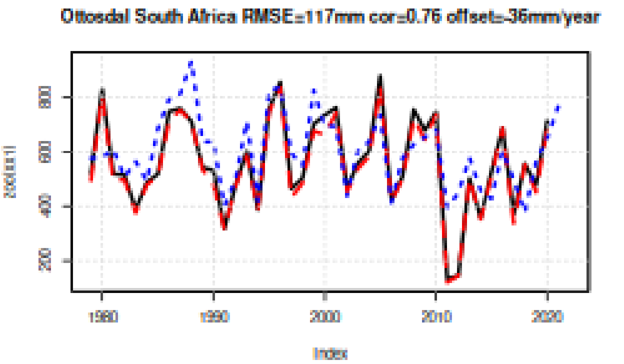


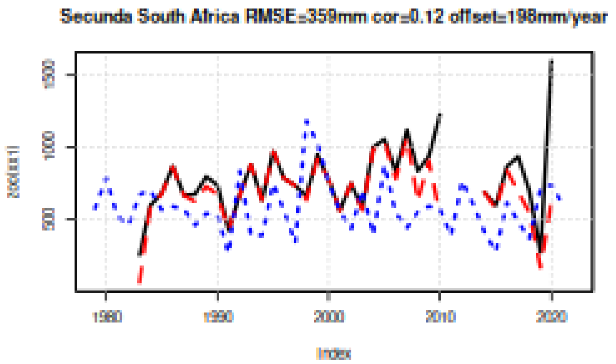

Chipinge Zimbabwe RMSE=206mm cor=0.82 offset=68mm/year
Bloemfontein WO South Africa RMSE=118mm cor=0.82 offset=-46mm/yea
Gokwe Zimbabwe RMSE=293mm cor=0.62 offset=-99mm/year
Bothaville - Balkfontein South Africa RMSE=271mm cor=0.3 offset=-169mm/
Mt Darwin Zimbabwe RMSE=189mm cor=0.68 offset=-115mm/year
Brandvlei South Africa RMSE=97mm cor=0.43 offset=-44mm/year
Nyanga Zimbabwe RMSE=426mm cor=0.16 offset=-51mm/year
Cape Agulhas South Africa RMSE=69mm cor=0.78 offset=14mm/year
Index
1980
1990
2000
2010
2020

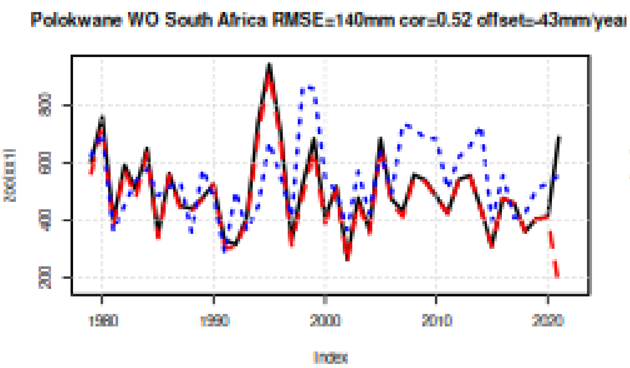

Cape St. Francis South Africa RMSE=114mm cor=0.75 offset=28mm/yea
200
400
600
800
1000
1980
1990
2000
2010
2020
Index

Irene WO South Africa RMSE=145mm cor=0.71 offset=34mm/year
400
600
800
1200
1980
1990
2000
2010
2020
Index

Cape Town WO South Africa RMSE=158mm cor=0.88 offset=-144mm/yea
200
400
600
800
1980
1990
2000
2010
2020
Index

Kimberley WO South Africa RMSE=129mm cor=0.78 offset=-70mm/year
200
400
600
800
1000
1980
1990
2000
2010
2020
Index

Cedara South Africa RMSE=285mm cor=0.82 offset=-259mm/year
600
800
1000
1400
1980
1990
2000
2010
2020
Index

Laingsburg South Africa RMSE=126mm cor=0.38 offset=-103mm/year
100
200
300
400
1980
1990
2000
2010
2020
Index

East London WO South Africa RMSE=212mm cor=0.79 offset=124mm/yea
400
600
800
1200
1980
1990
2000
2010
2020
Index

Marico South Africa RMSE=161mm cor=0.64 offset=-42mm/year
200
400
600
800
1000
1980
1990
2000
2010
2020
Index

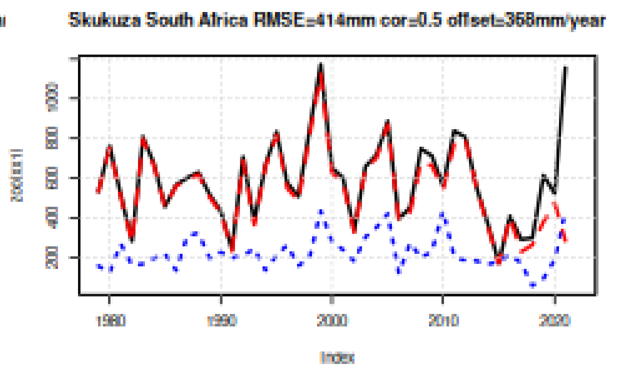

Mount Edgecombe South Africa RMSE=192mm cor=0.76 offset=5mm/yea
Index

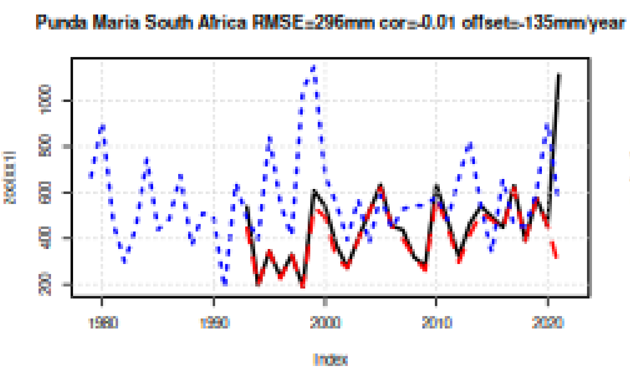

Secunda South Africa RMSE=274mm cor=0.19 offset=-69mm/year
Index

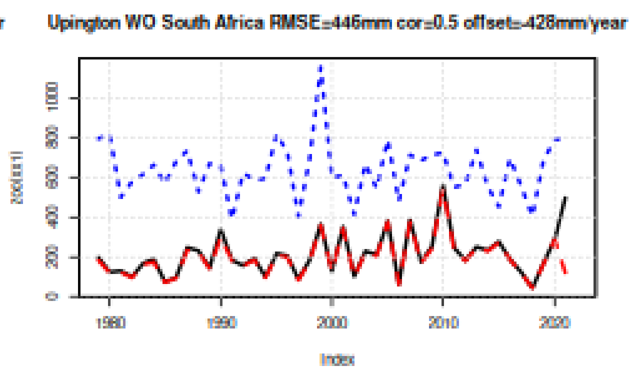

Ottosdal South Africa RMSE=117mm cor=0.76 offset=-36mm/year
Index

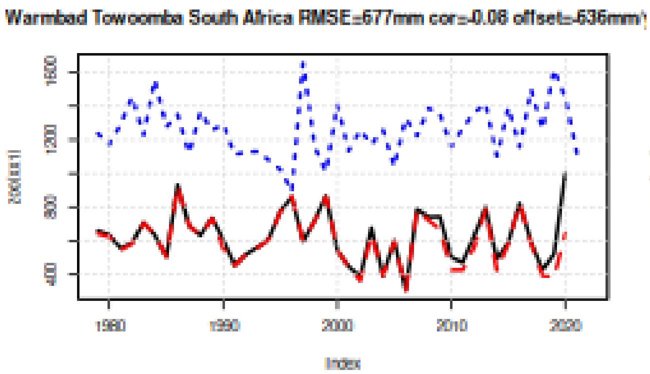

Secunda South Africa RMSE=359mm cor=0.12 offset=198mm/year
Index

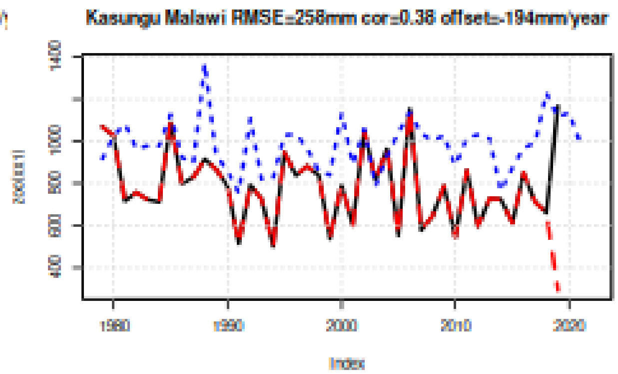

Polokwane WO South Africa RMSE=140mm cor=0.52 offset=-43mm/yeai
Index

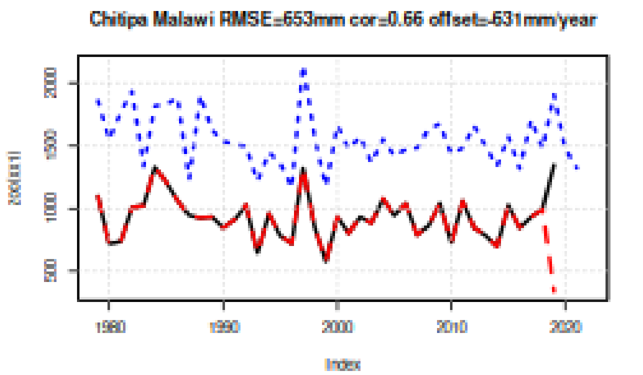

Skukuza South Africa RMSE=414mm cor=0.5 offset=368mm/year
Index

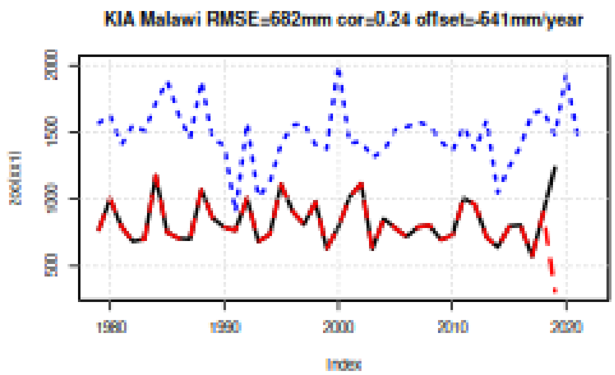

Punda Maria South Africa RMSE=296mm cor=-0.01 offset=-135mm/year
Index

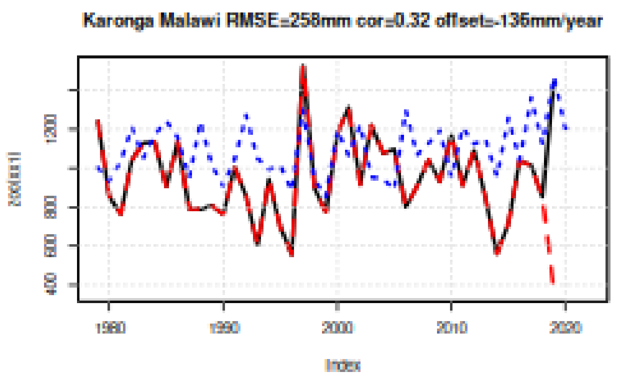

Upington WO South Africa RMSE=446mm cor=0.5 offset=-428mm/year
Index

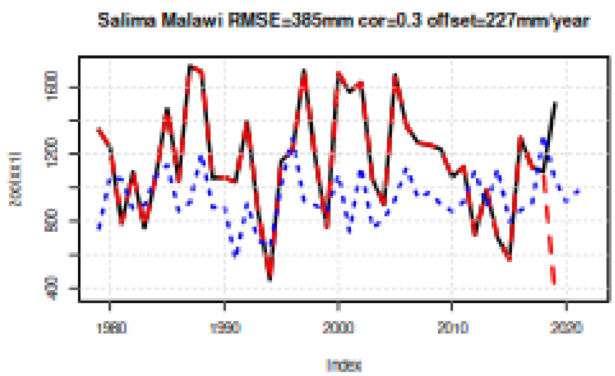

Warmbad Towoomba South Africa RMSE=677mm cor=-0.08 offset=-636mm/
Index

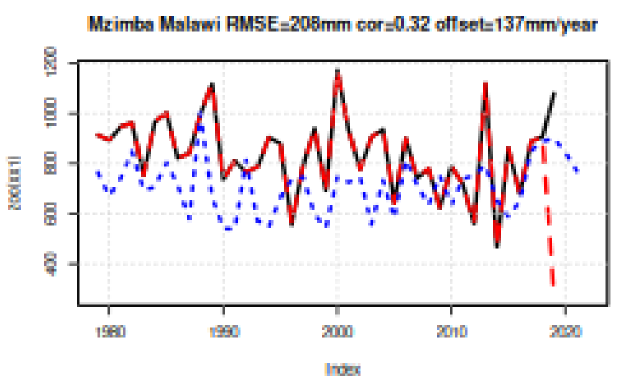

Kasungu Malawi RMSE=258mm cor=0.38 offset=-194mm/year
Index

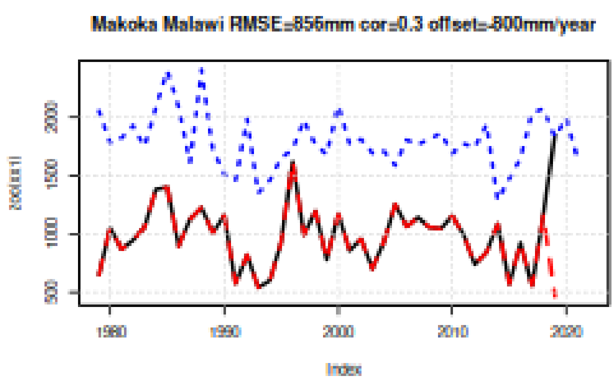

Chitipa Malawi RMSE=653mm cor=0.66 offset=-631mm/year
Index

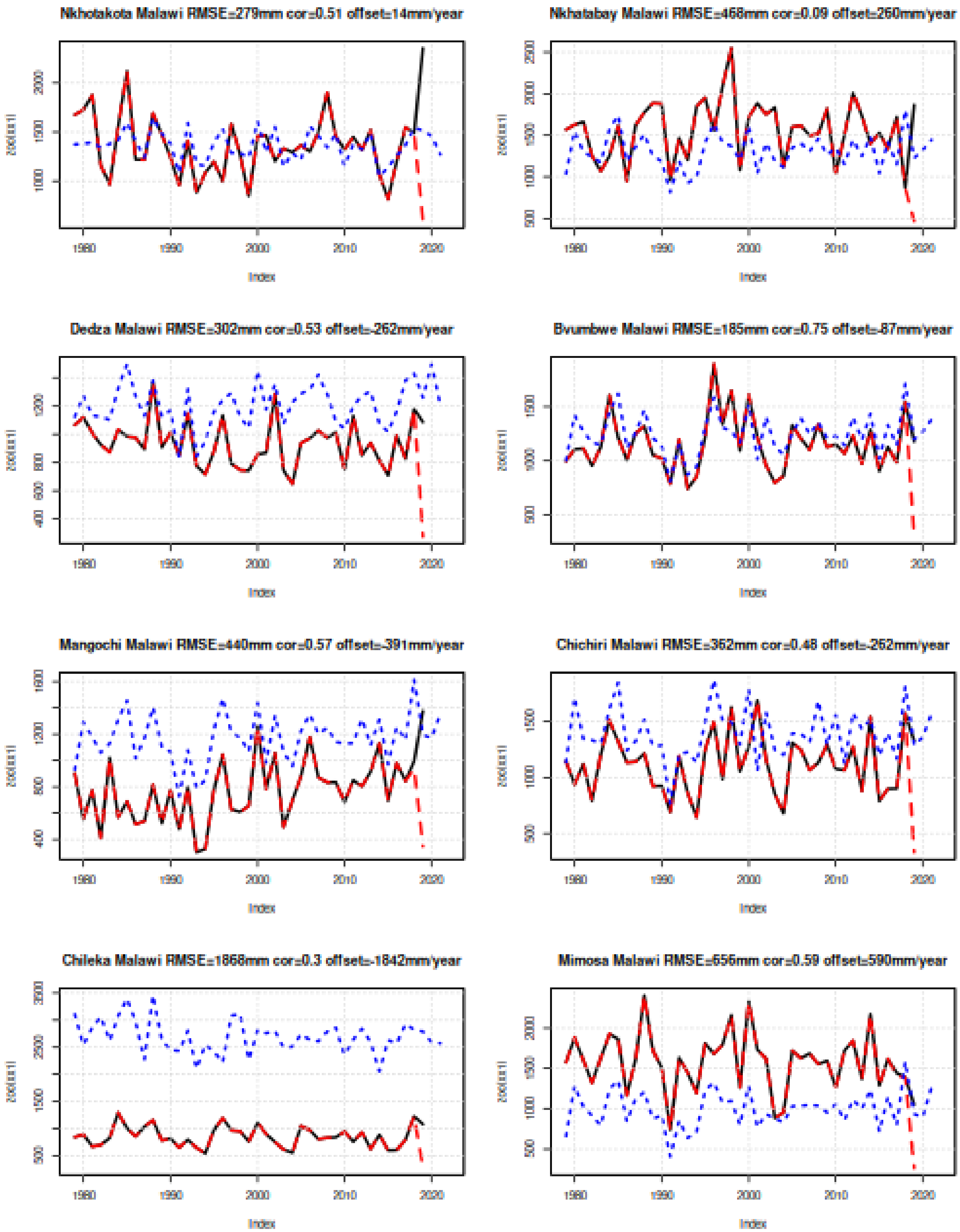

KIA Malawi RMSE=682mm cor=0.24 offset=-641mm/year
Index

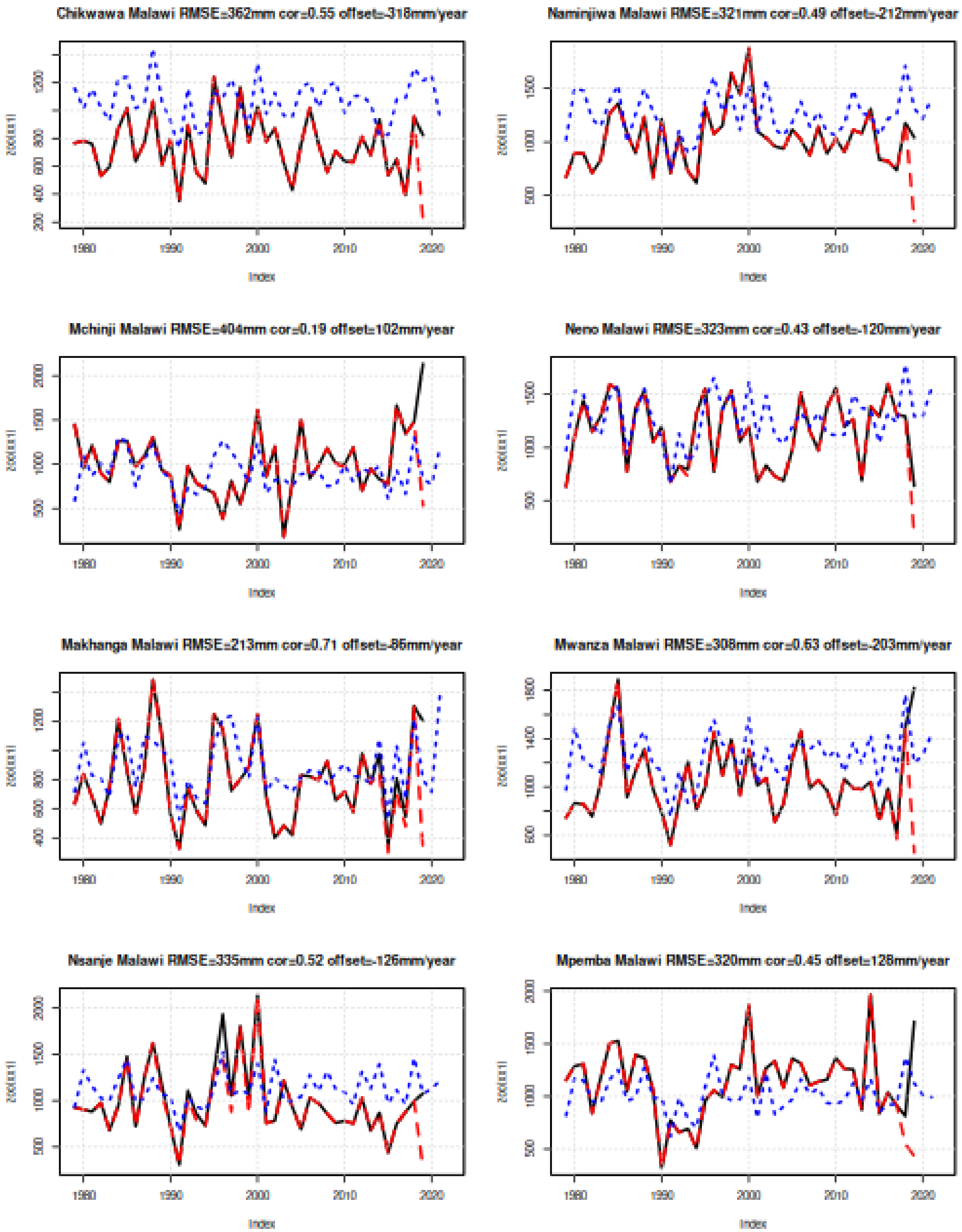

Karonga Malawi RMSE=258mm cor=0.32 offset=-136mm/year
Index

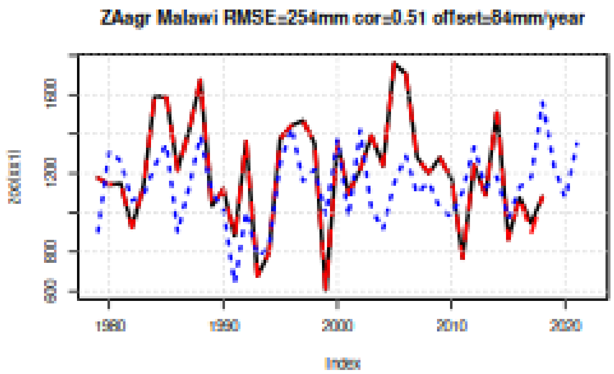
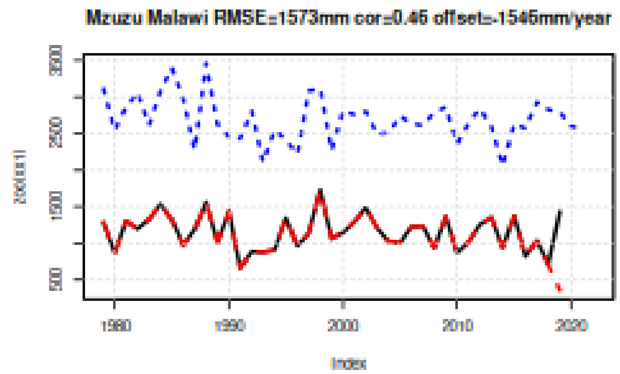
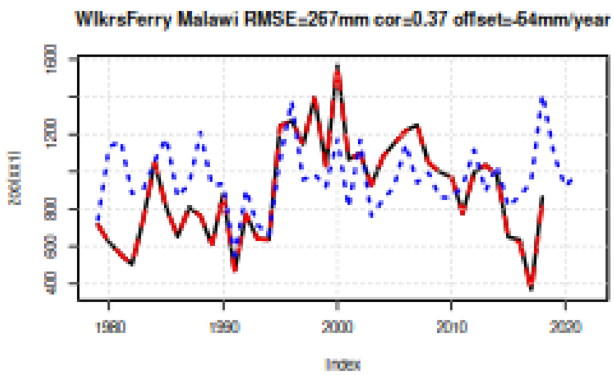
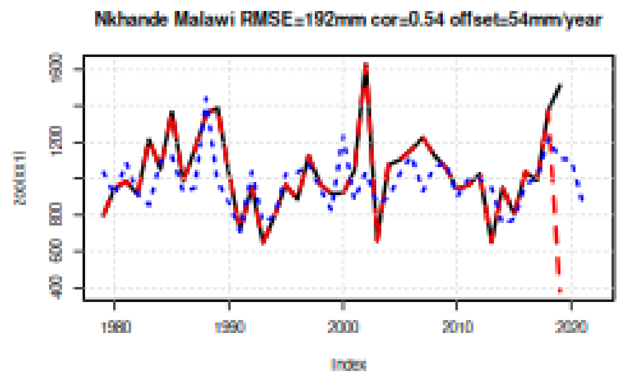
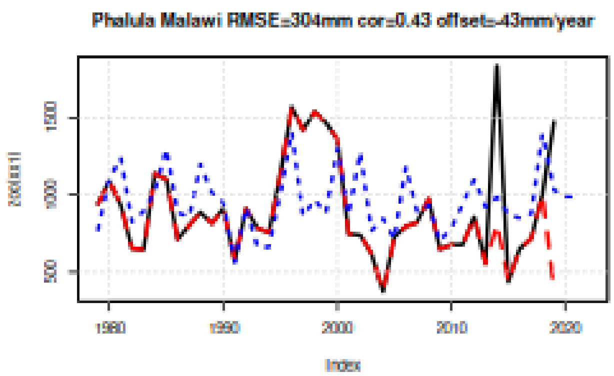
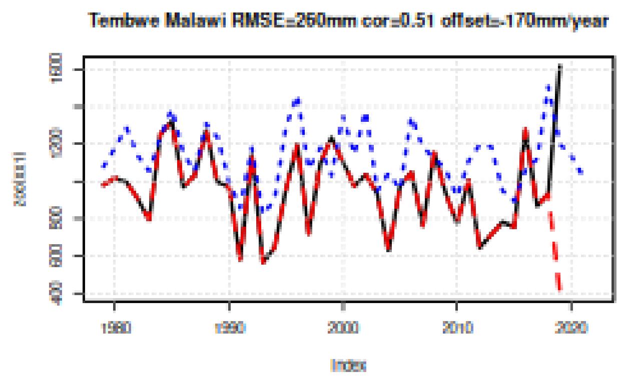
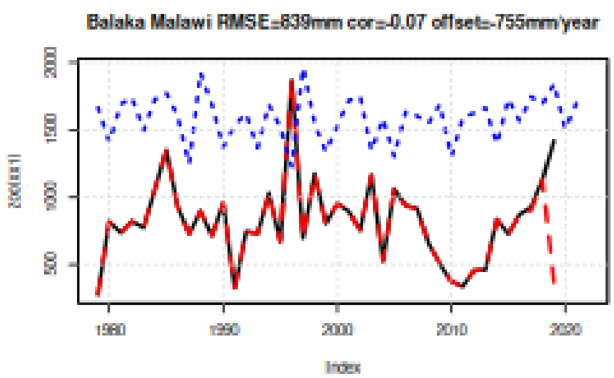
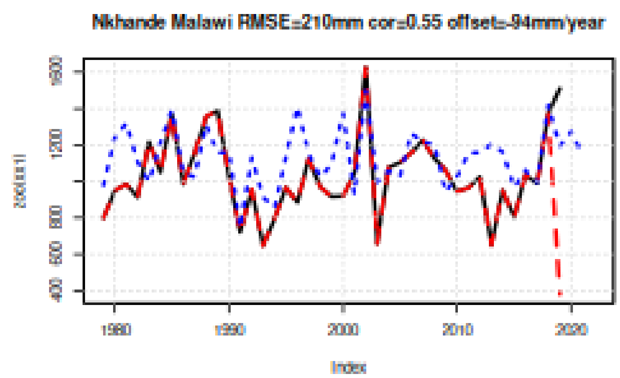

Salima Malawi RMSE=385mm cor=0.3 offset=227mm/year
Index


Mzimba Malawi RMSE=208mm cor=0.32 offset=137mm/year
Index

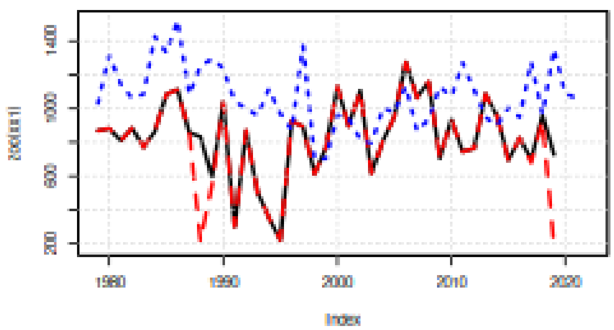

Makoka Malawi RMSE=856mm cor=0.3 offset=-800mm/year
Index


Nkhotakota Malawi RMSE=279mm cor=0.51 offset=14mm/year
Nkhatabay Malawi RMSE=468mm cor=0.09 offset=260mm/year
Dedza Malawi RMSE=302mm cor=0.53 offset=-262mm/year
Bvumbwe Malawi RMSE=185mm cor=0.75 offset=-87mm/year
Mangochi Malawi RMSE=440mm cor=0.57 offset=-391mm/year
Chichiri Malawi RMSE=362mm cor=0.48 offset=-262mm/year
Chileka Malawi RMSE=1868mm cor=0.3 offset=-1842mm/year
Mimosa Malawi RMSE=656mm cor=0.59 offset=590mm/year
1980
1990
2000
2010
2020
Index

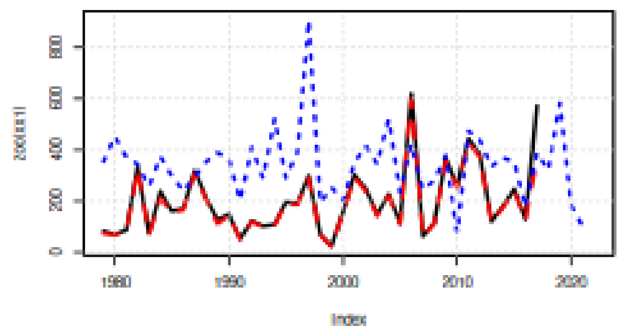

Chikwawa Malawi RMSE=362mm cor=0.55 offset=-318mm/year
Index
Naminjiwa Malawi RMSE=321mm cor=0.49 offset=-212mm/year
Index
Mchinji Malawi RMSE=404mm cor=0.19 offset=102mm/year
Index
Neno Malawi RMSE=323mm cor=0.43 offset=-120mm/year
Index
Makhanga Malawi RMSE=213mm cor=0.71 offset=-86mm/year
Index
Mwanza Malawi RMSE=308mm cor=0.63 offset=-203mm/year
Index
Nsanje Malawi RMSE=335mm cor=0.52 offset=-126mm/year
Index
Mpemba Malawi RMSE=320mm cor=0.45 offset=128mm/year
Index


ZAagr Malawi RMSE=254mm cor=0.51 offset=84mm/year
1980 1990 2000 2010 2020
Index

Mzuzu Malawi RMSE=1573mm cor=0.46 offset=-1546mm/year
1980 1990 2000 2010 2020
Index

WlkrsFerry Malawi RMSE=267mm cor=0.37 offset=-64mm/year
1980 1990 2000 2010 2020
Index

Nkhande Malawi RMSE=192mm cor=0.54 offset=54mm/year
1980 1990 2000 2010 2020
Index

Phalula Malawi RMSE=304mm cor=0.43 offset=-43mm/year
1980 1990 2000 2010 2020
Index

Tembwe Malawi RMSE=260mm cor=0.51 offset=-170mm/year
1980 1990 2000 2010 2020
Index

Balaka Malawi RMSE=639mm cor=-0.07 offset=-755mm/year
1980 1990 2000 2010 2020
Index

Nkhande Malawi RMSE=210mm cor=0.55 offset=-94mm/year
1980 1990 2000 2010 2020
Index

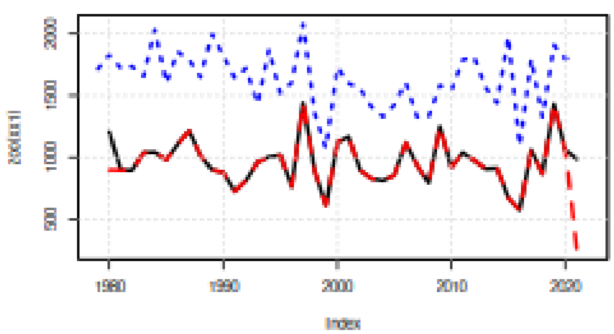
Ntaja Malawi RMSE=372mm cor=0.07 offset=-239mm/year



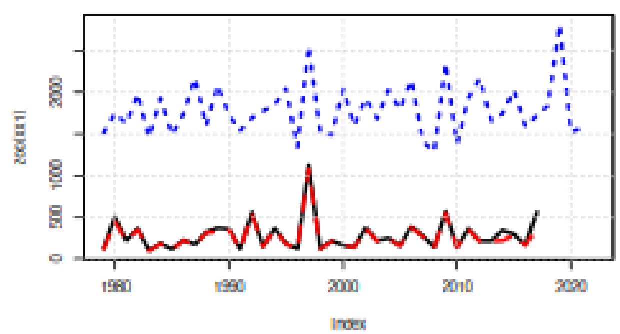
Lodwar Kenya RMSE=212mm cor=0.33 offset=-145mm/year



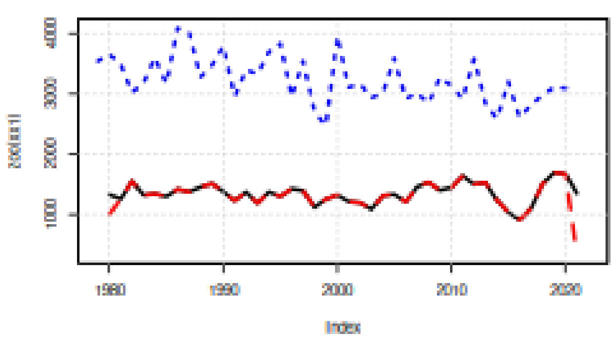
KIGALIAERO Rwanda RMSE=700mm cor=0.56 offset=-670mm/year



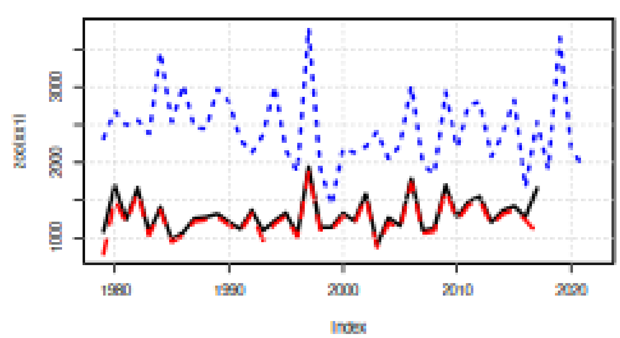
Mandera Kenya RMSE=1514mm cor=0.59 offset=-1497mm/year



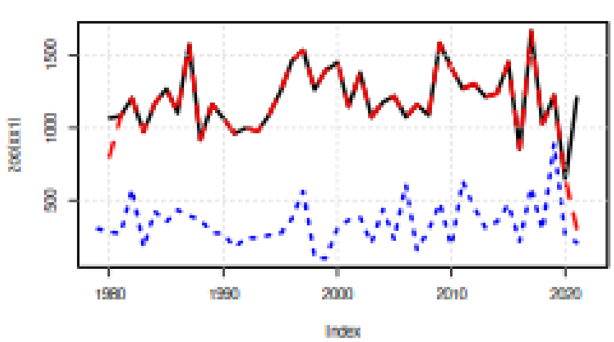
KAMEMBEAERO Rwanda RMSE=1920mm cor=0.19 offset=-1879mm/yea



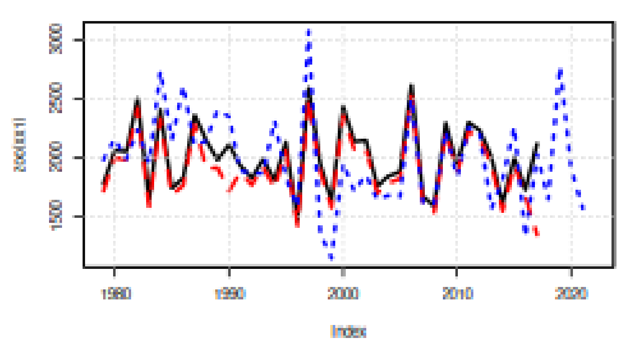
Kitale Kenya RMSE=1205mm cor=0.53 offset=-1136mm/year

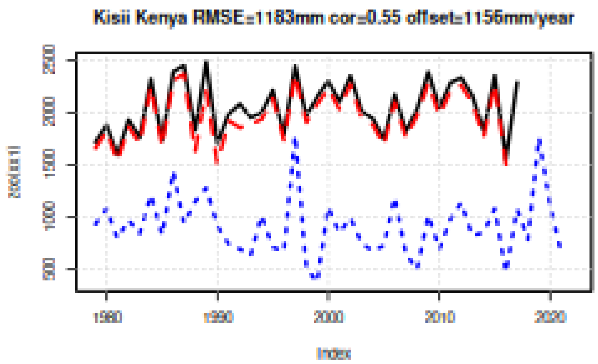


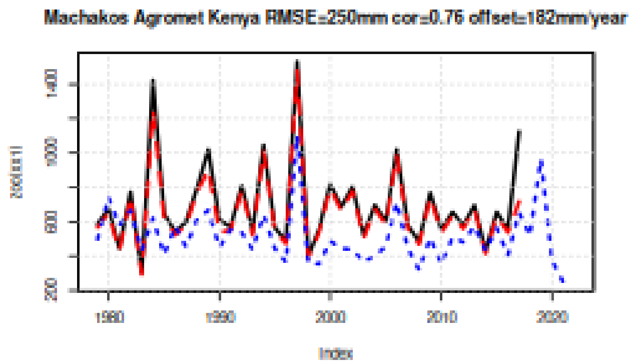
GISENYIAERO Rwanda RMSE=880mm cor=0.34 offset=854mm/year

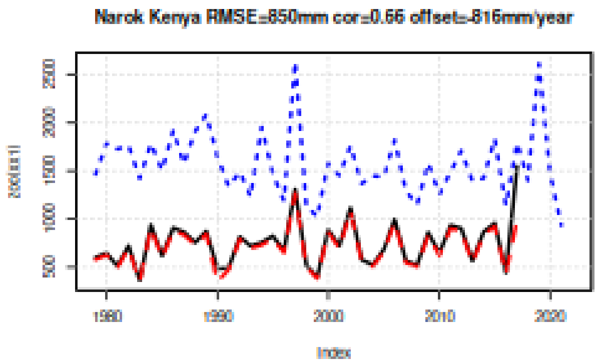


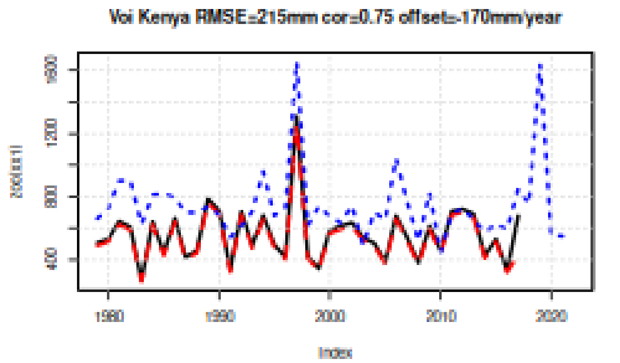
Kericho Kenya RMSE=306mm cor=0.62 offset=18mm/year

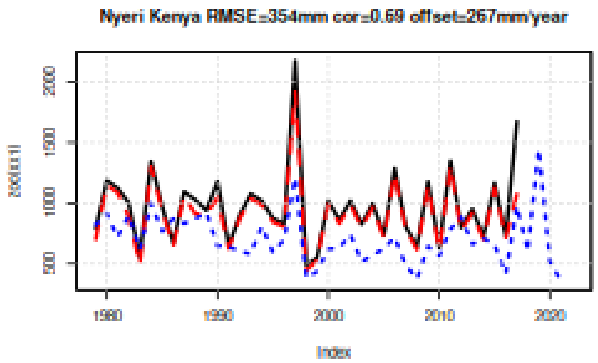


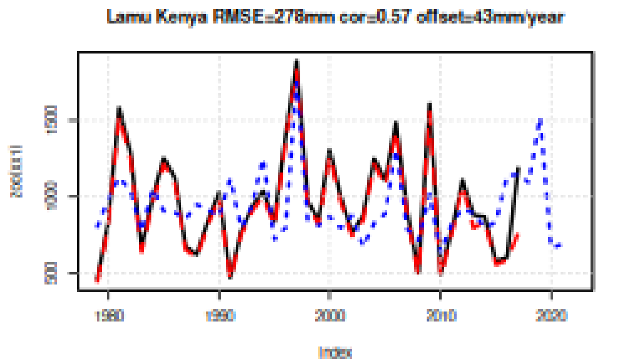

Kisii Kenya RMSE=1183mm cor=0.55 offset=1156mm/year
Index

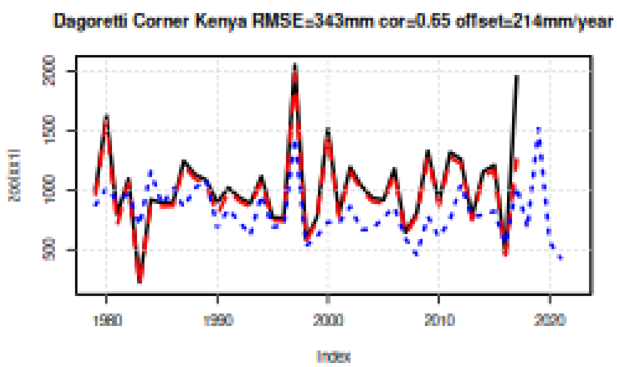

Machakos Agromet Kenya RMSE=250mm cor=0.76 offset=182mm/year
Index

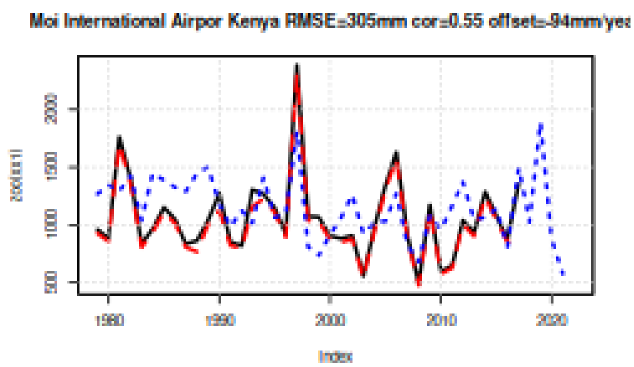

Narok Kenya RMSE=850mm cor=0.66 offset=-816mm/year
Index

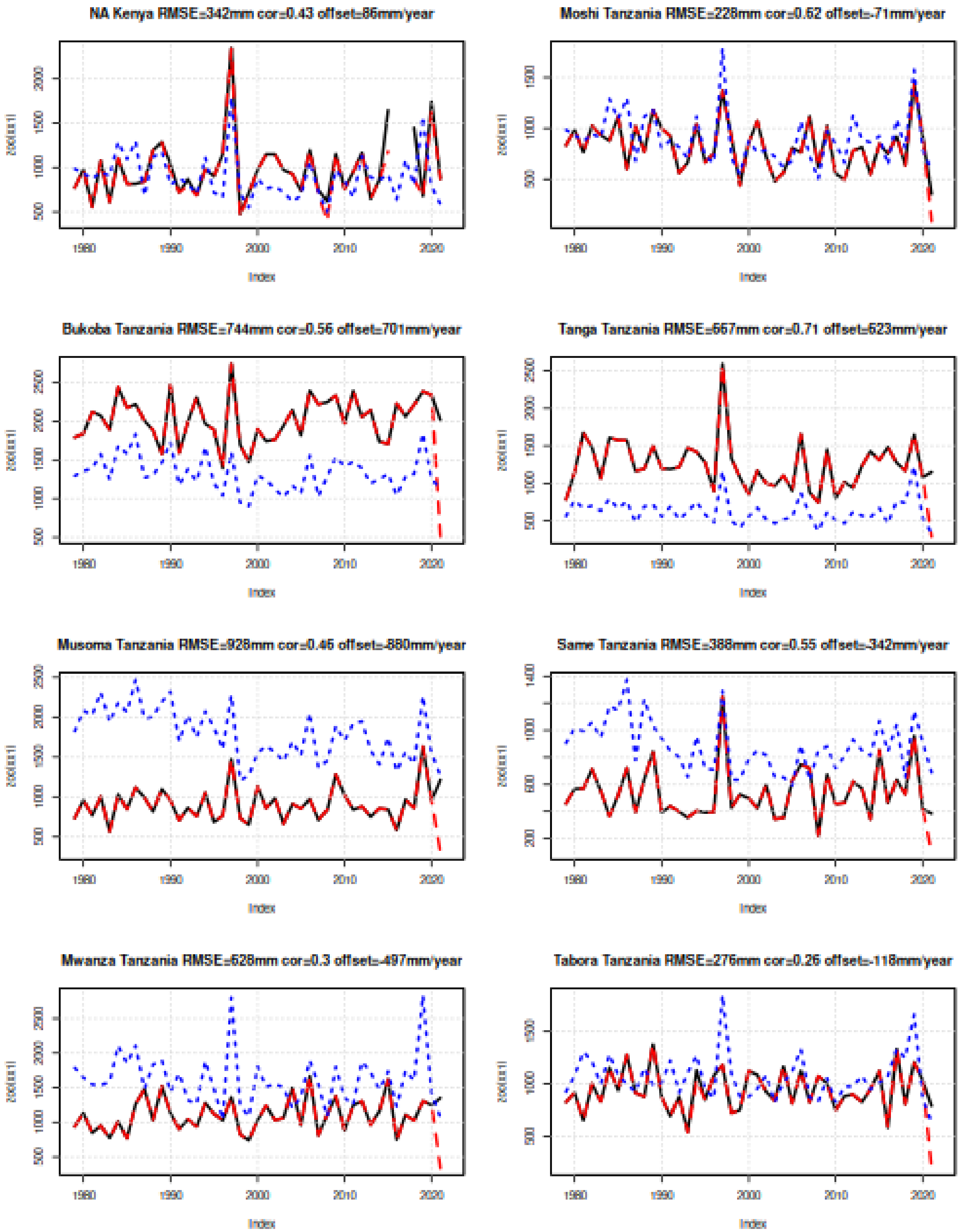

Voi Kenya RMSE=215mm cor=0.75 offset=-170mm/year
Index


Nyeri Kenya RMSE=354mm cor=0.69 offset=267mm/year
Index

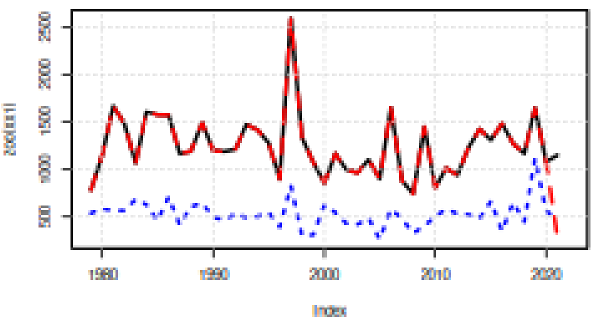

Lamu Kenya RMSE=278mm cor=0.57 offset=43mm/year
Index


Dagoretti Corner Kenya RMSE=343mm cor=0.65 offset=214mm/year
Index

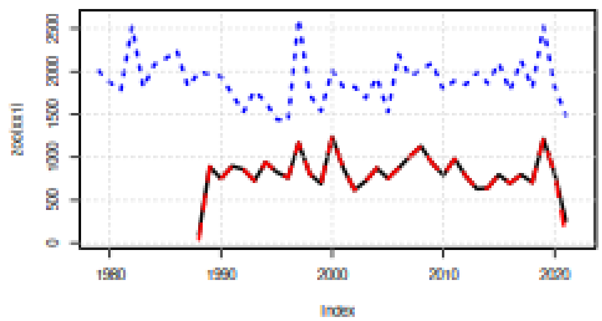

Moi International Airpor Kenya RMSE=305mm cor=0.55 offset=-94mm/yea
Index


NA Kenya RMSE=342mm cor=0.43 offset=86mm/year
Moshi Tanzania RMSE=228mm cor=0.62 offset=-71mm/year
Bukoba Tanzania RMSE=744mm cor=0.56 offset=701mm/year
Tanga Tanzania RMSE=667mm cor=0.71 offset=623mm/year
Musoma Tanzania RMSE=928mm cor=0.46 offset=-880mm/year
Same Tanzania RMSE=388mm cor=0.55 offset=-342mm/year
Mwanza Tanzania RMSE=628mm cor=0.3 offset=-497mm/year
Tabora Tanzania RMSE=276mm cor=0.26 offset=-118mm/year
Index

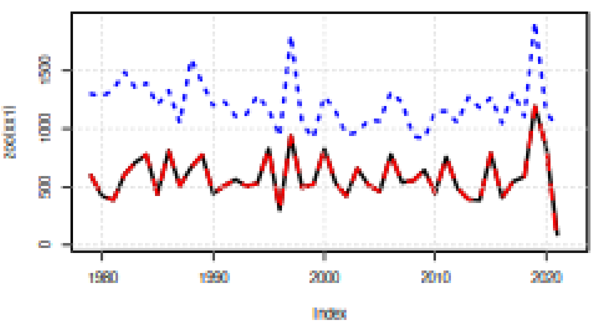
Tanga Tanzania RMSE=777mm cor=0.48 offset=722mm/year



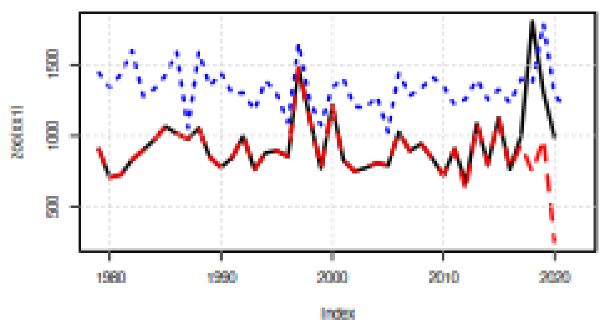
Sumbawanga Tanzania RMSE=1088mm cor=0.44 offset=-1056mm/year



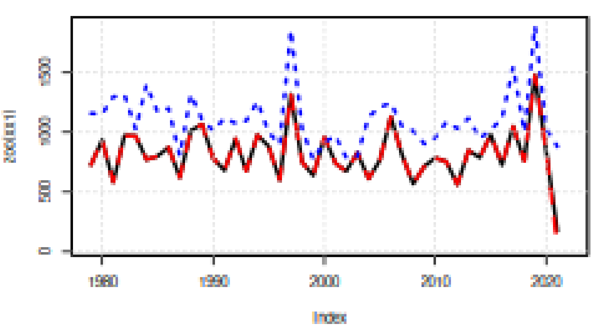
Dodoma Tanzania RMSE=649mm cor=0.61 offset=-624mm/year



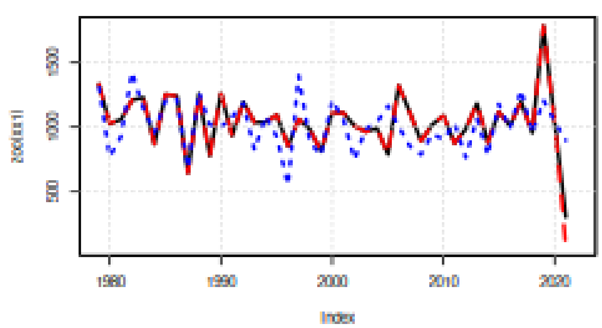
Mbeya Tanzania RMSE=449mm cor=0.42 offset=-400mm/year



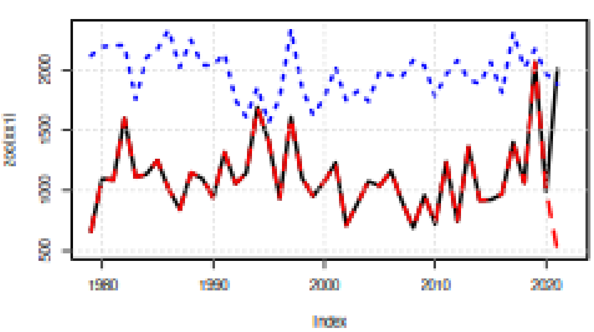
Morogoro Tanzania RMSE=343mm cor=0.68 offset=-291mm/year



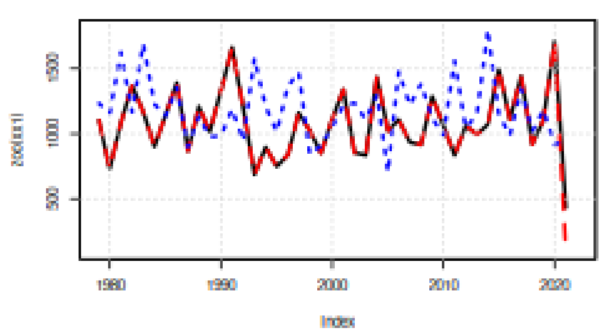
Songea Tanzania RMSE=205mm cor=0.53 offset=32mm/year

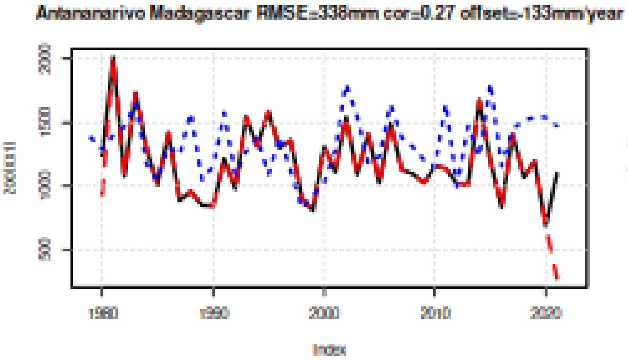


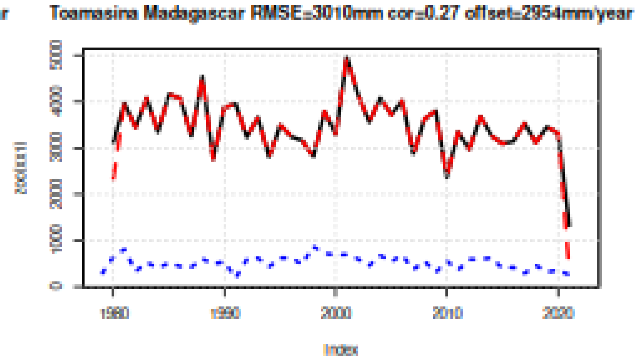
Dar es Salaam Tanzania RMSE=914mm cor=0.18 offset=-851mm/year

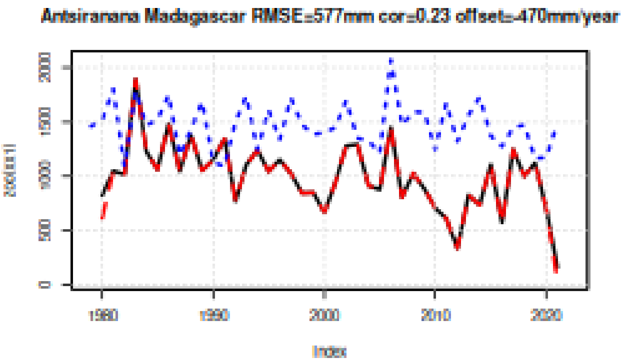


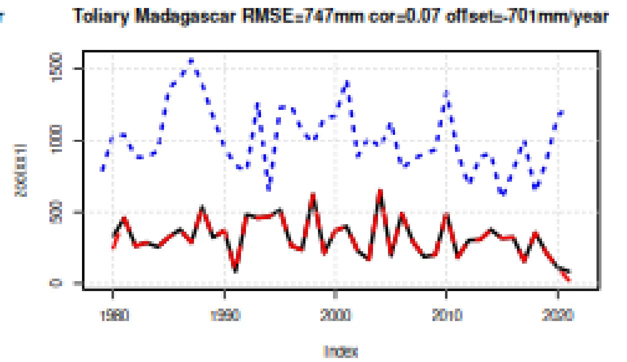
Mtwara Tanzania RMSE=362mm cor=0.01 offset=-106mm/year

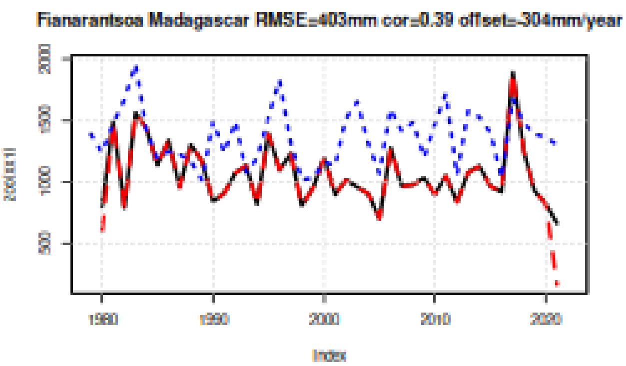


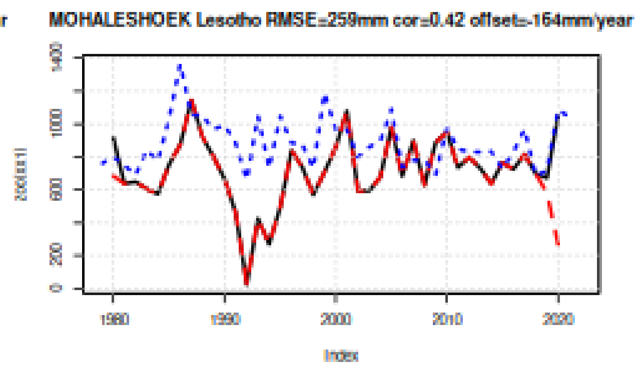

Antananarivo Madagascar RMSE=338mm cor=0.27 offset=-133mm/year
1980
1990
2000
2010
2020
Index

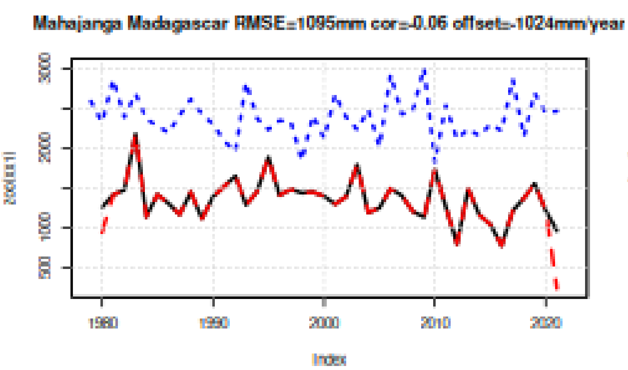

Toamasina Madagascar RMSE=3010mm cor=0.27 offset=2954mm/year
1980
1990
2000
2010
2020
Index

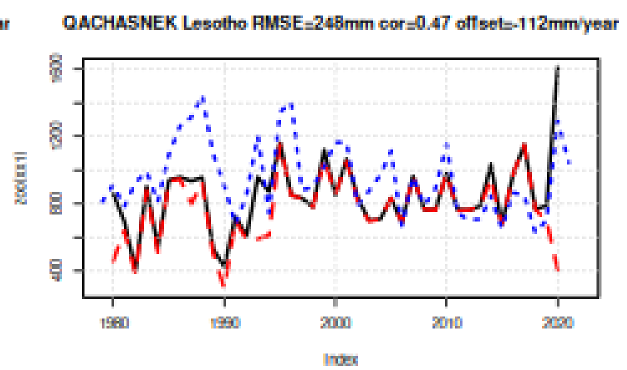

Antsiranana Madagascar RMSE=577mm cor=0.23 offset=-470mm/year
1980
1990
2000
2010
2020
Index

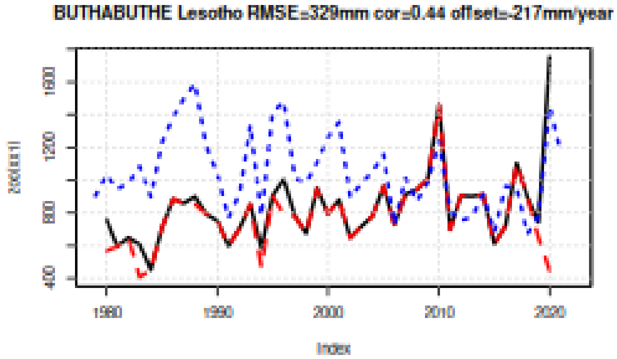

Toliary Madagascar RMSE=747mm cor=0.07 offset=-701mm/year
1980
1990
2000
2010
2020
Index

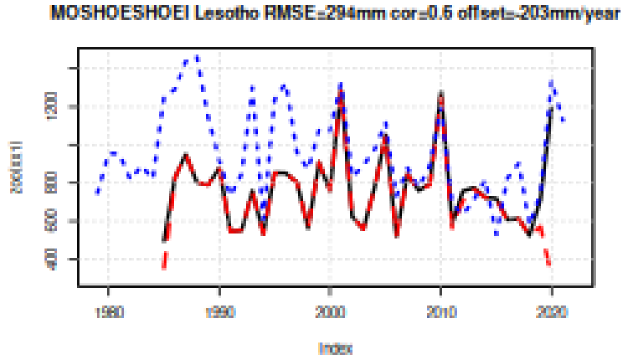

Fianarantsoa Madagascar RMSE=403mm cor=0.39 offset=-304mm/year
1980
1990
2000
2010
2020
Index

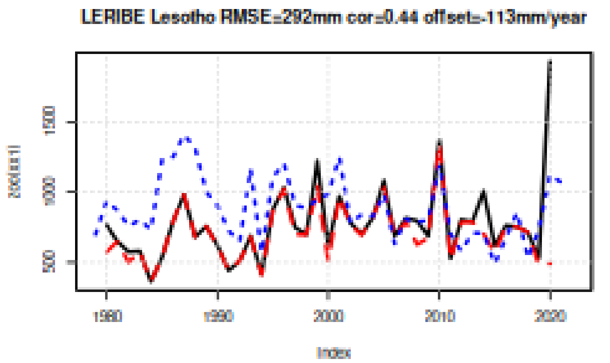

MOHALESHOEK Lesotho RMSE=259mm cor=0.42 offset=-164mm/year
1980
1990
2000
2010
2020
Index

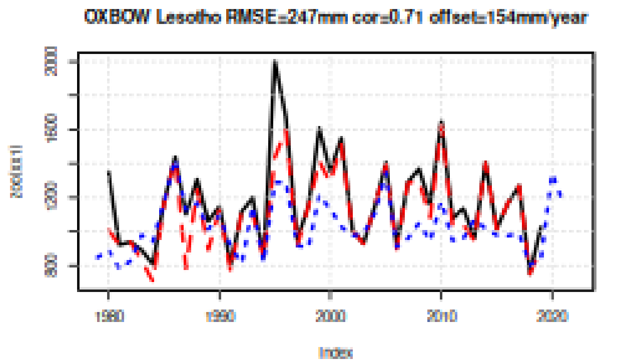

Mahajanga Madagascar RMSE=1095mm cor=-0.06 offset=-1024mm/year
1980
1990
2000
2010
2020
Index

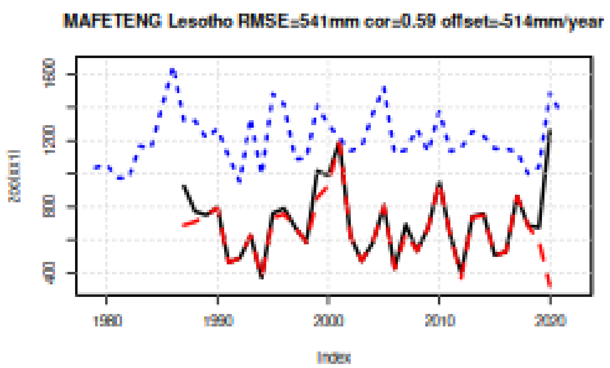

QACHASNEK Lesotho RMSE=248mm cor=0.47 offset=-112mm/year
1980
1990
2000
2010
2020
Index

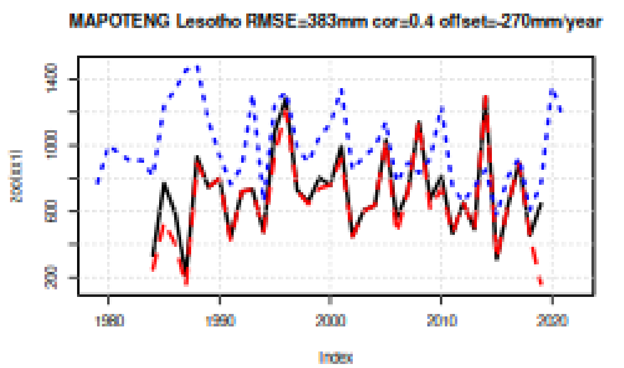

BUTHABUTHE Lesotho RMSE=329mm cor=0.44 offset=-217mm/year
1980
1990
2000
2010
2020
Index

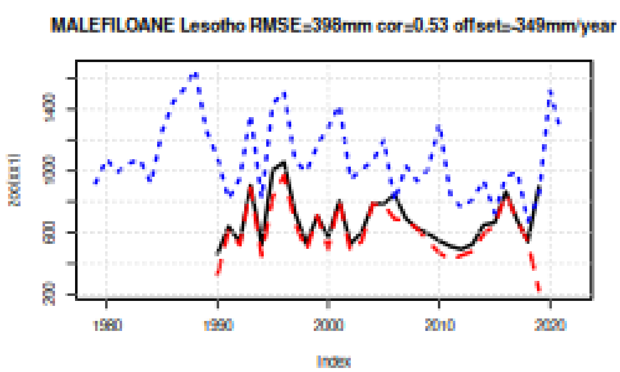

MOSHOESHOEI Lesotho RMSE=294mm cor=0.6 offset=-203mm/year
1980
1990
2000
2010
2020
Index

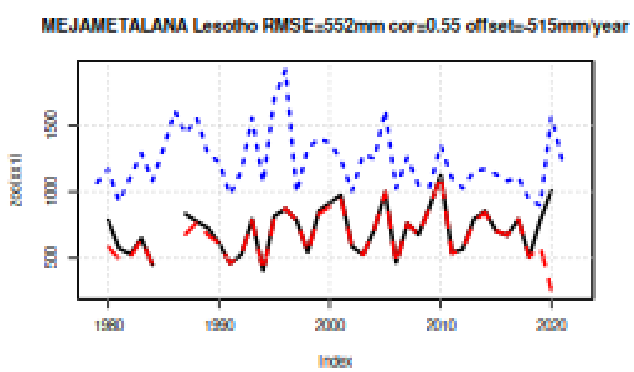

LERIBE Lesotho RMSE=292mm cor=0.44 offset=-113mm/year
1980
1990
2000
2010
2020
Index

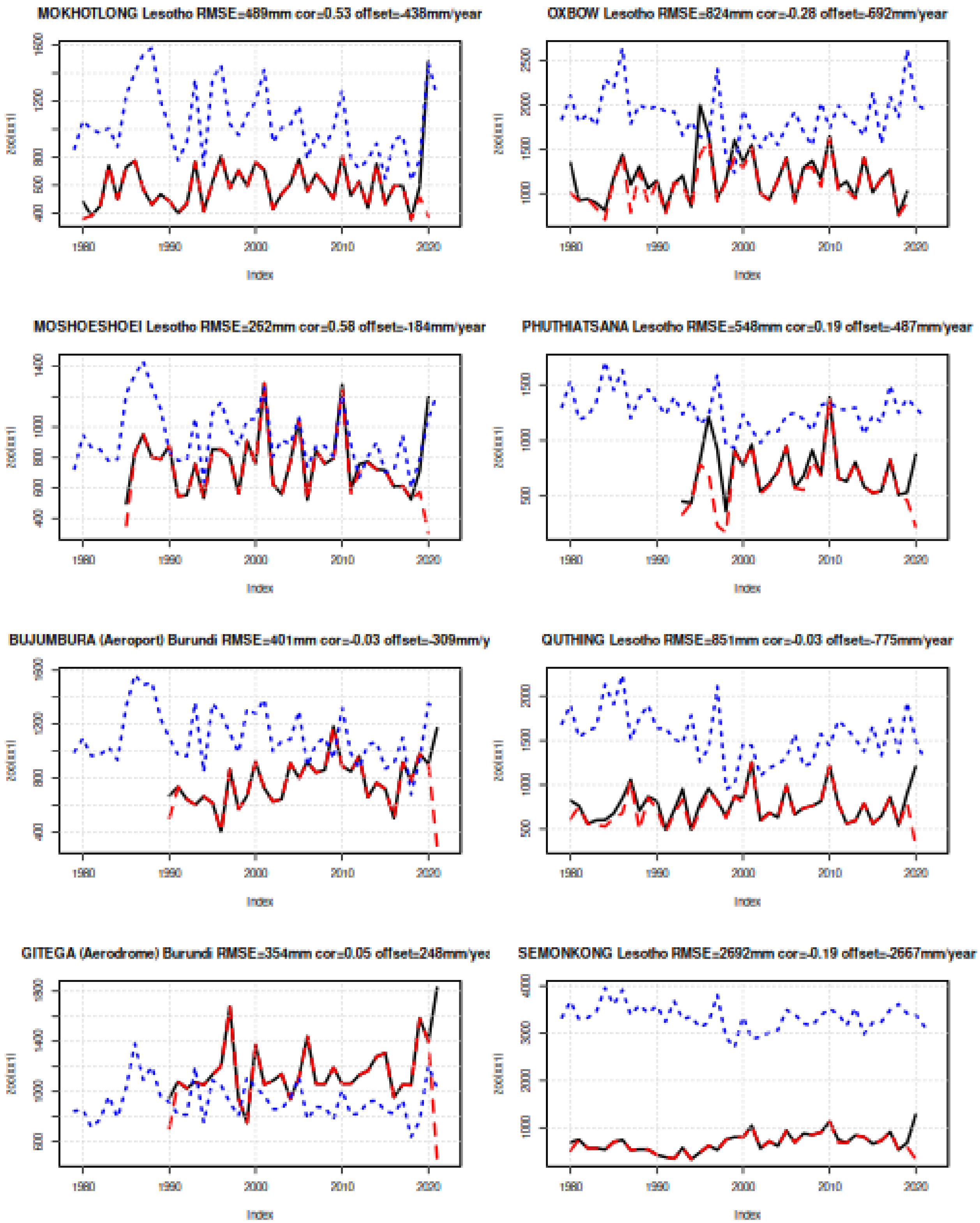

OXBOW Lesotho RMSE=247mm cor=0.71 offset=154mm/year
1980
1990
2000
2010
2020
Index


MAFETENG Lesotho RMSE=541mm cor=0.59 offset=-514mm/year
1980
1990
2000
2010
2020
Index


MAPOTENG Lesotho RMSE=383mm cor=0.4 offset=-270mm/year
1980
1990
2000
2010
2020
Index


MALEFILOANE Lesotho RMSE=398mm cor=0.53 offset=-349mm/year
1980
1990
2000
2010
2020
Index

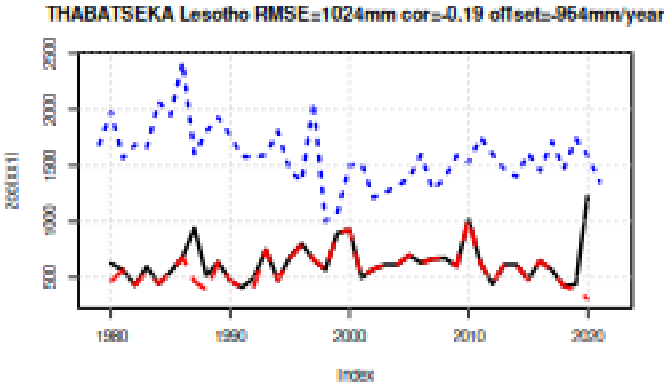

MEJAMETALANA Lesotho RMSE=552mm cor=0.55 offset=-515mm/year
1980
1990
2000
2010
2020
Index

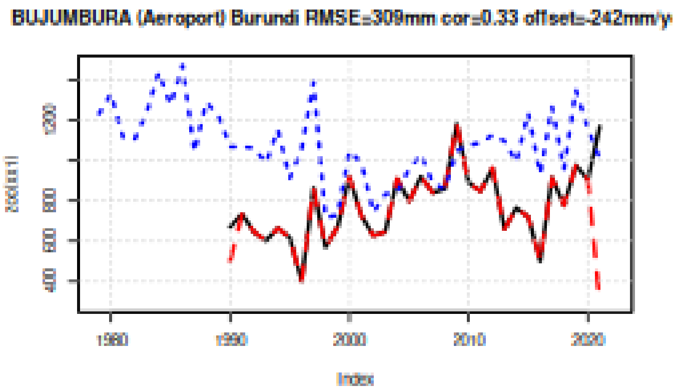



```
print(summary(r)); print(summary(ry))
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.2800  0.3200  0.5150  0.4599  0.6275  0.8800
```

```
##     Min.  1st Qu.  Median    Mean  3rd Qu.    Max.
## -0.2900  0.2700  0.4450  0.4083  0.6075  0.8600
```





The dashed blue curves represent the interpolated ERA5 data, the black curves represent annual rainfall based on the product between the mean and the number of days ($x_{tot} = n \times \overline{x}$), whereas the red dashed are estimates based on the summed values ($x_{tot} = \sum_d x_d$ where d are the days in a given year).

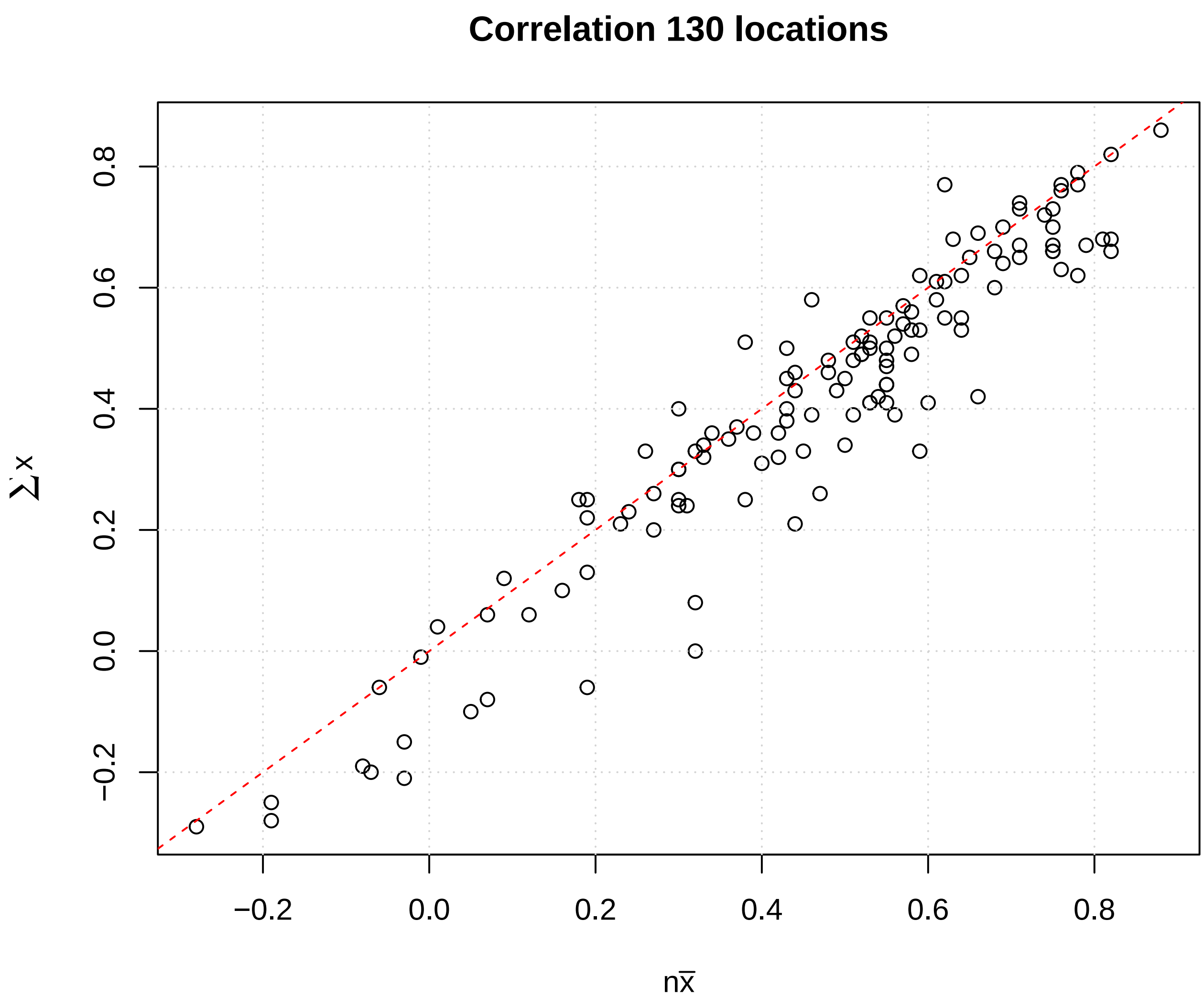
The reason for using two different ways of estimating annual totals is that there are missing days/data. If the missing days are random, then we get a better estimate for annual totals with the former, but if the missing days tend to be from dry periods, we should get more accurate results with the latter. The results here suggest that both give very similar estimates, and that we don't need to worry about the missing days.

The station-by-station comparison shows that interpolated ERA5 annual rainfall matches the corresponding results from rain gauges remarkably well for many sites, but also that there are some sites where there are substantial differences.

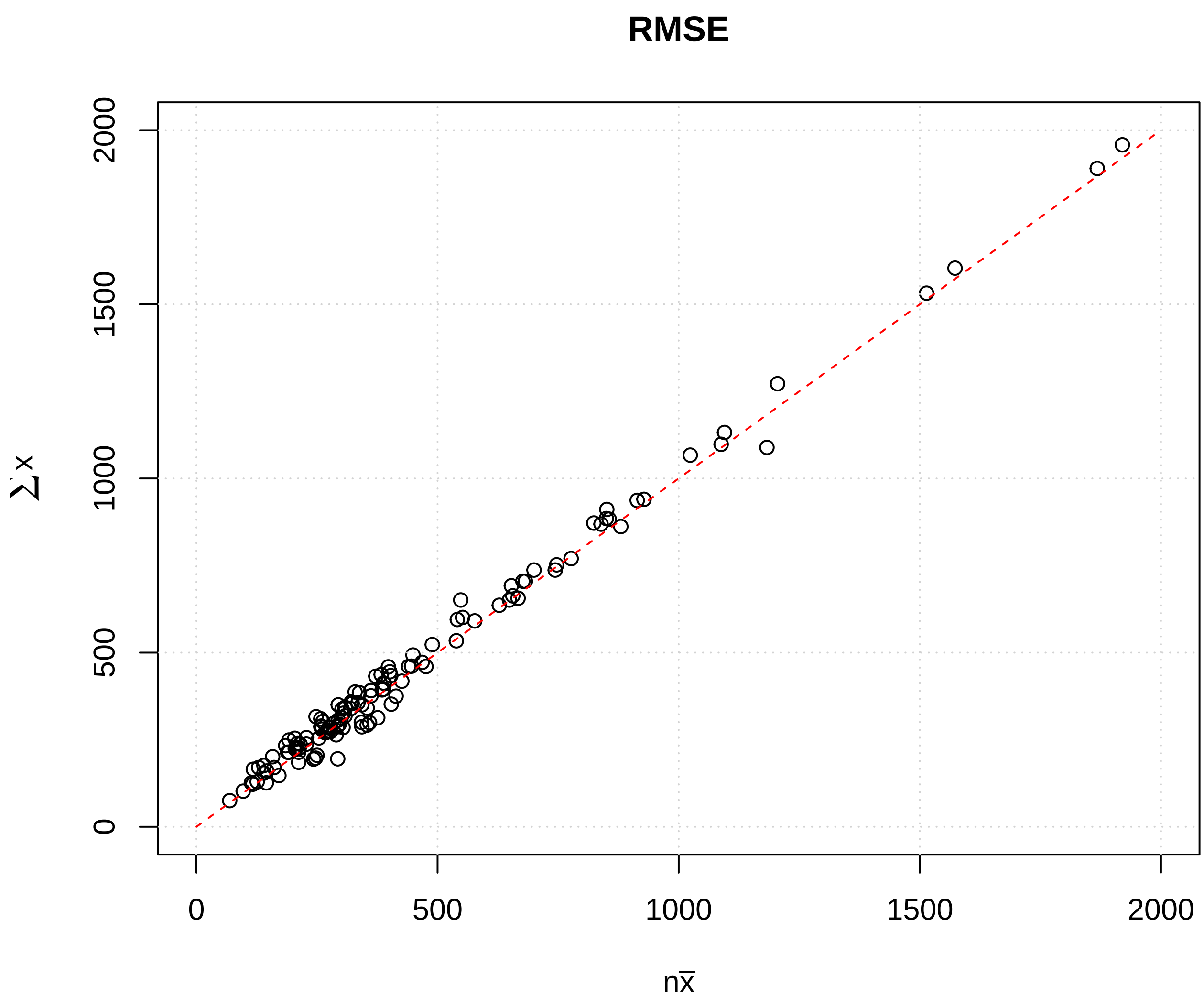
2.1.6 Evaluation of different approaches for estimating annual values.

We also compared the statistics between the two methods for estimating annual total precipitation in terms of correlation, RMSE and mean bias.

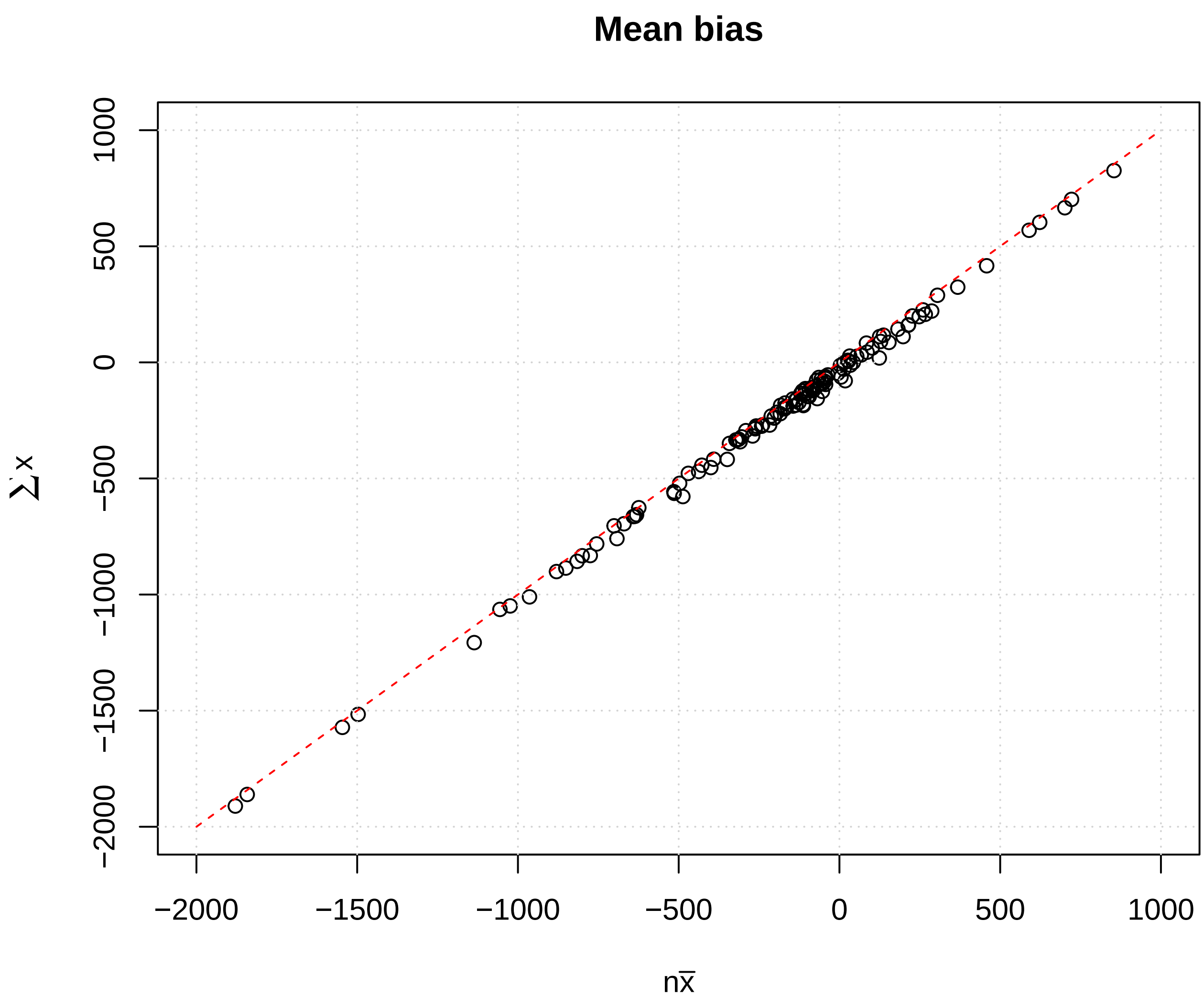
```r
par(mfcol=c(1,1),cex=1)
plot(r,ry,main=paste('Correlation',sum(is.finite(r)),'locations'),
     xlab=expression(n*bar(x)),ylab=expression(sum(x)))
grid()
lines(c(-1,1),c(-1,1),col='red',lty=2)
```




```
plot(rmse,rmsey,main='RMSE',xlab=expression(n*bar(x)),ylab=expression(sum(x)),
     xlim=c(0,2000),ylim=c(0,2000))
grid()

lines(c(0,2000),c(0,2000),col='red',lty=2)
```




```r
plot(offset,offsety,main='Mean bias',xlab=expression(n*bar(x)),
     ylab=expression(sum(x)),xlim=c(-2000,1000),ylim=c(-2000,1000))
grid()
lines(c(-2000,1000),c(-2000,1000),col='red',lty=2)
```




```
## fill in the missing values in ry
ii <- !is.finite(ry)
ry[ii] <- r[ii]
ii <- !is.finite(rmsey)
rmsey[ii] <- rmse[ii]
ii <- !is.finite(offsety)
offsety[ii] <- offset[ii]
```

This comparison suggested that our results were not sensitive to which method we used for estimating the annual rainfall total.

The chunk below stores the statistics from the evaluation together with the data itself in the shape of attributes. This is a nice way to ake the statistics follow the data in future analysis.

```
## Store the evaluation of stations as attributes of X

attr(X,'rmse') <- rmse
attr(X,'cor') <- r
attr(X,'offset') <- offset

## The annual rainfall totals - Jan-to-Dec
#atX <- combine.stations(365.25*annual(X,FUN='mean',nmin=300),Mdg)
```

```r
atX <- 365.25*annual(X,FUN='mean',nmin=100,start=year.start)
```

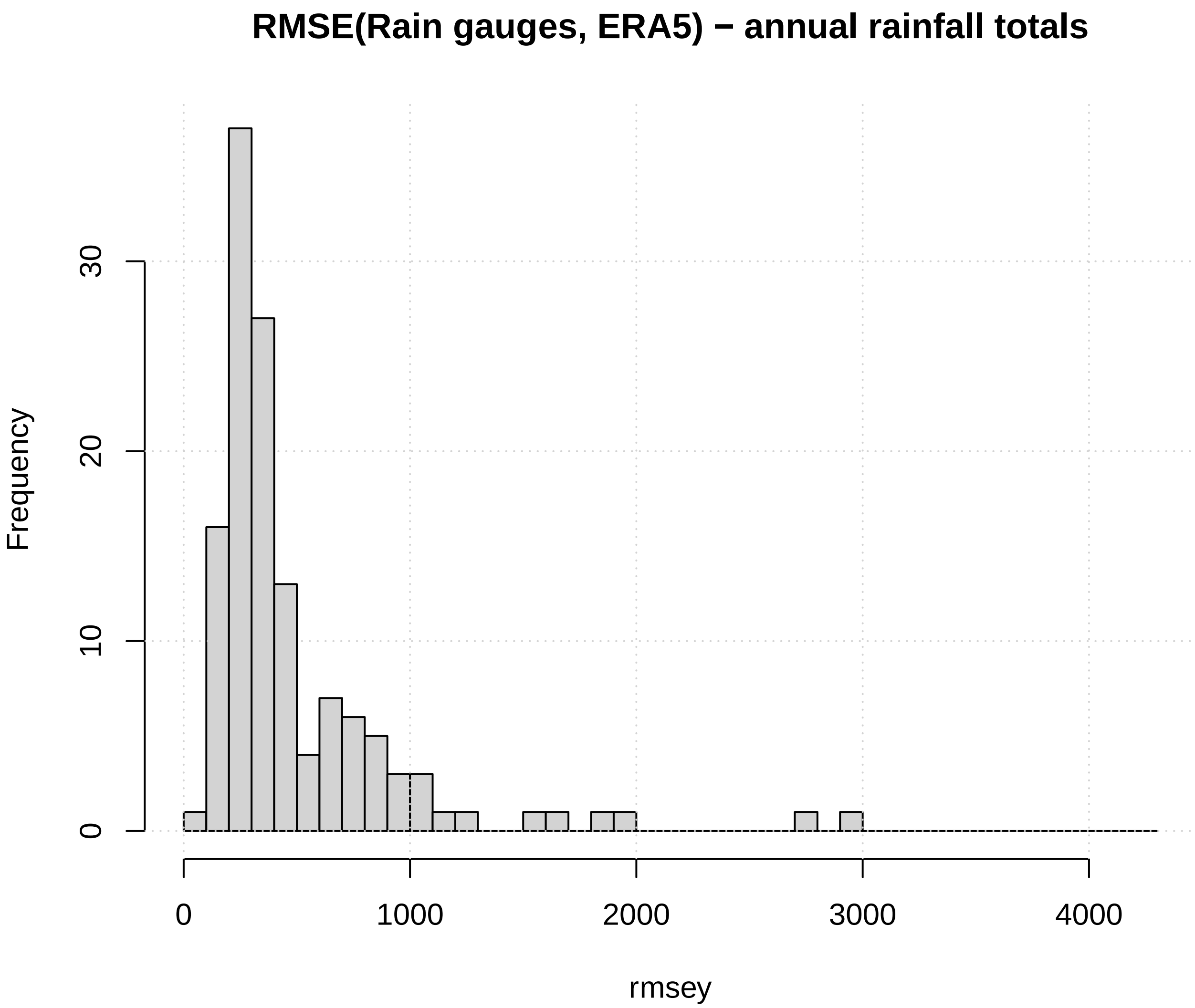
```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```r
attr(atX,'rmse') <- rmsey
attr(atX,'cor') <- ry
attr(atX,'offset') <- offsety
```

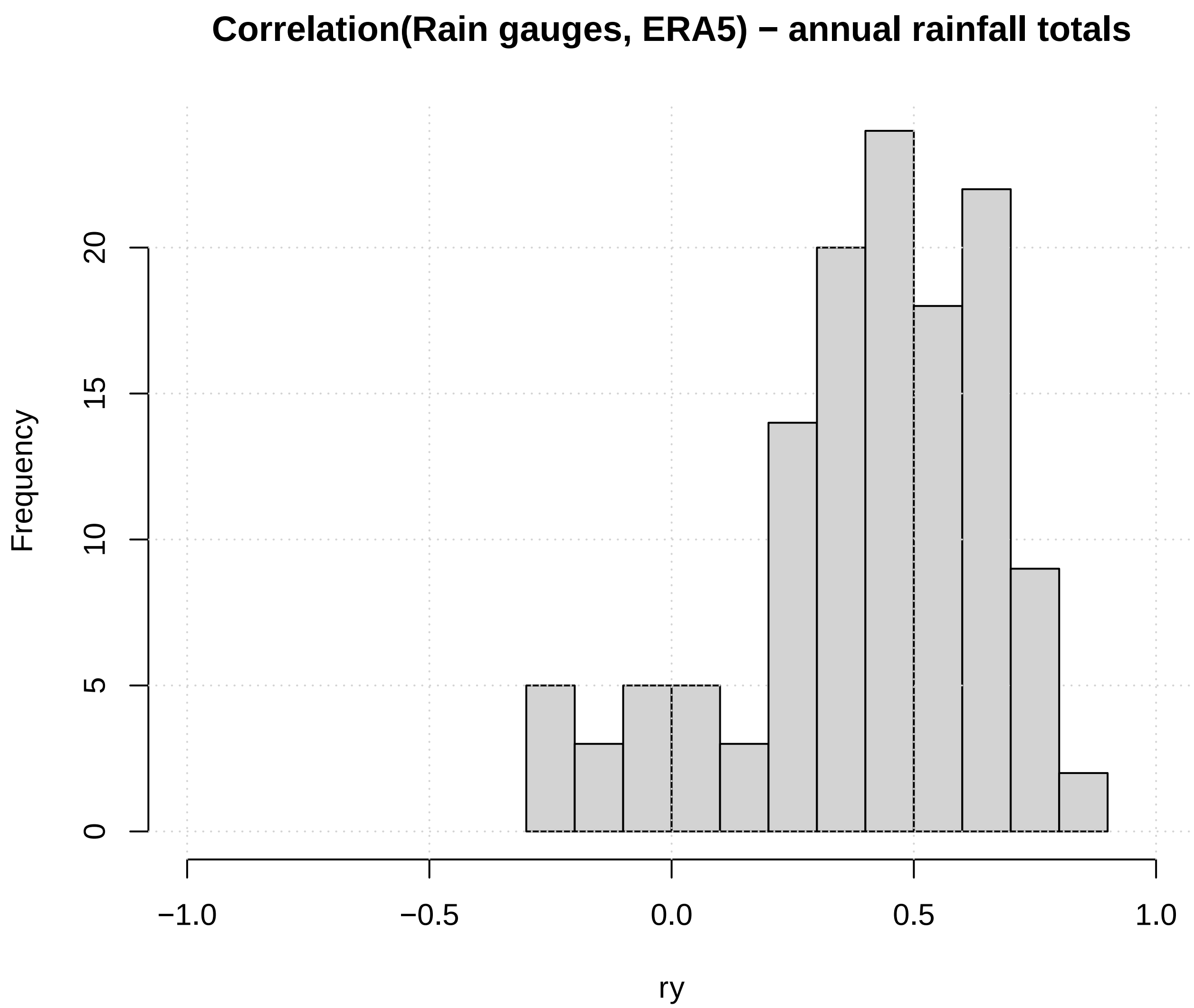
2.1.7 Histograms: metrics of similarity

Histograms give an overview of the over-all 'performance' of the rain gauges/ERA5 (which is the most representative?)

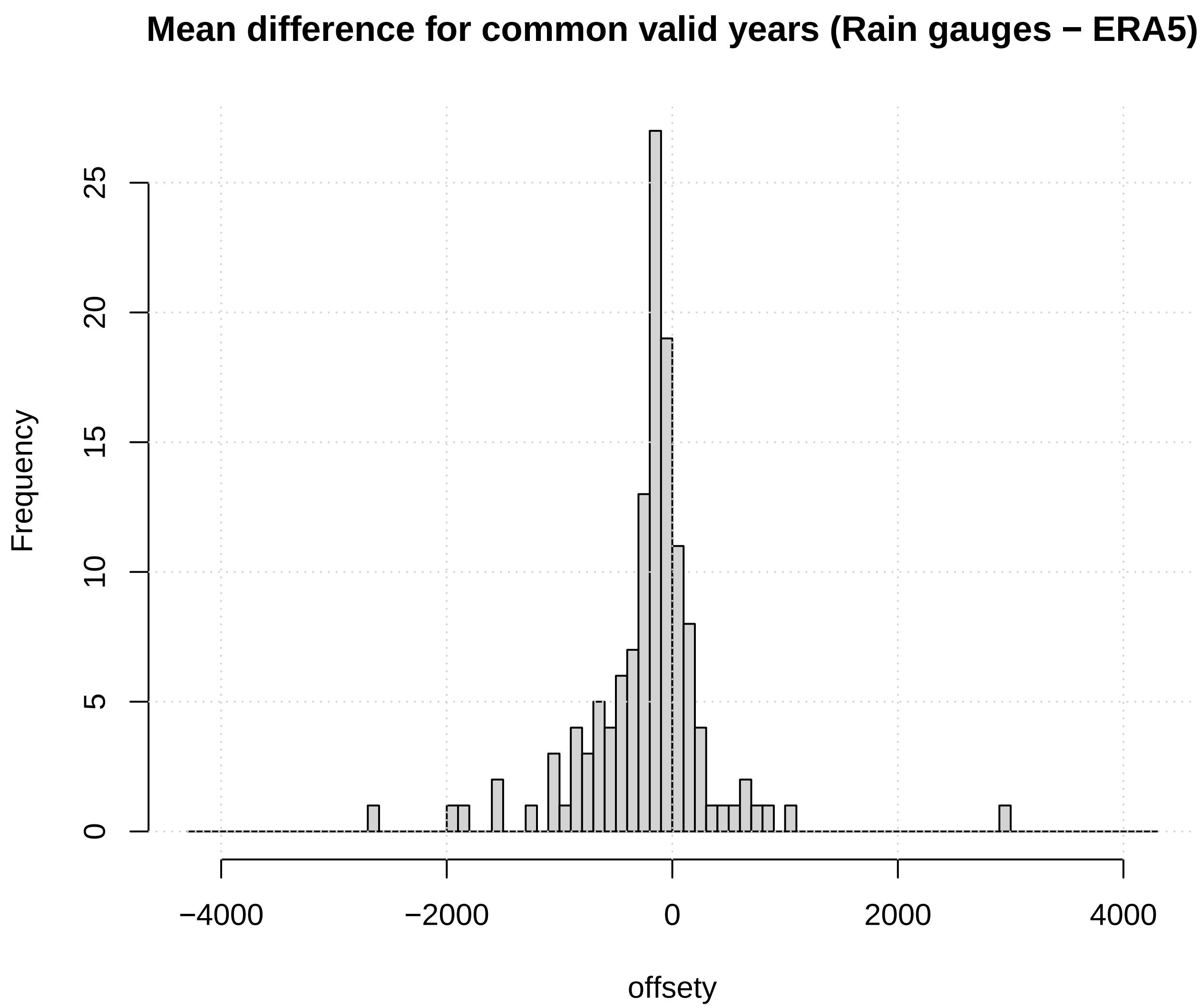
```r
par(mfcol=c(1,1),cex=1)
hist(rmsey,main='RMSE(Rain gauges, ERA5) - annual rainfall totals',
     breaks=seq(0,4300,by=100))
grid()
```



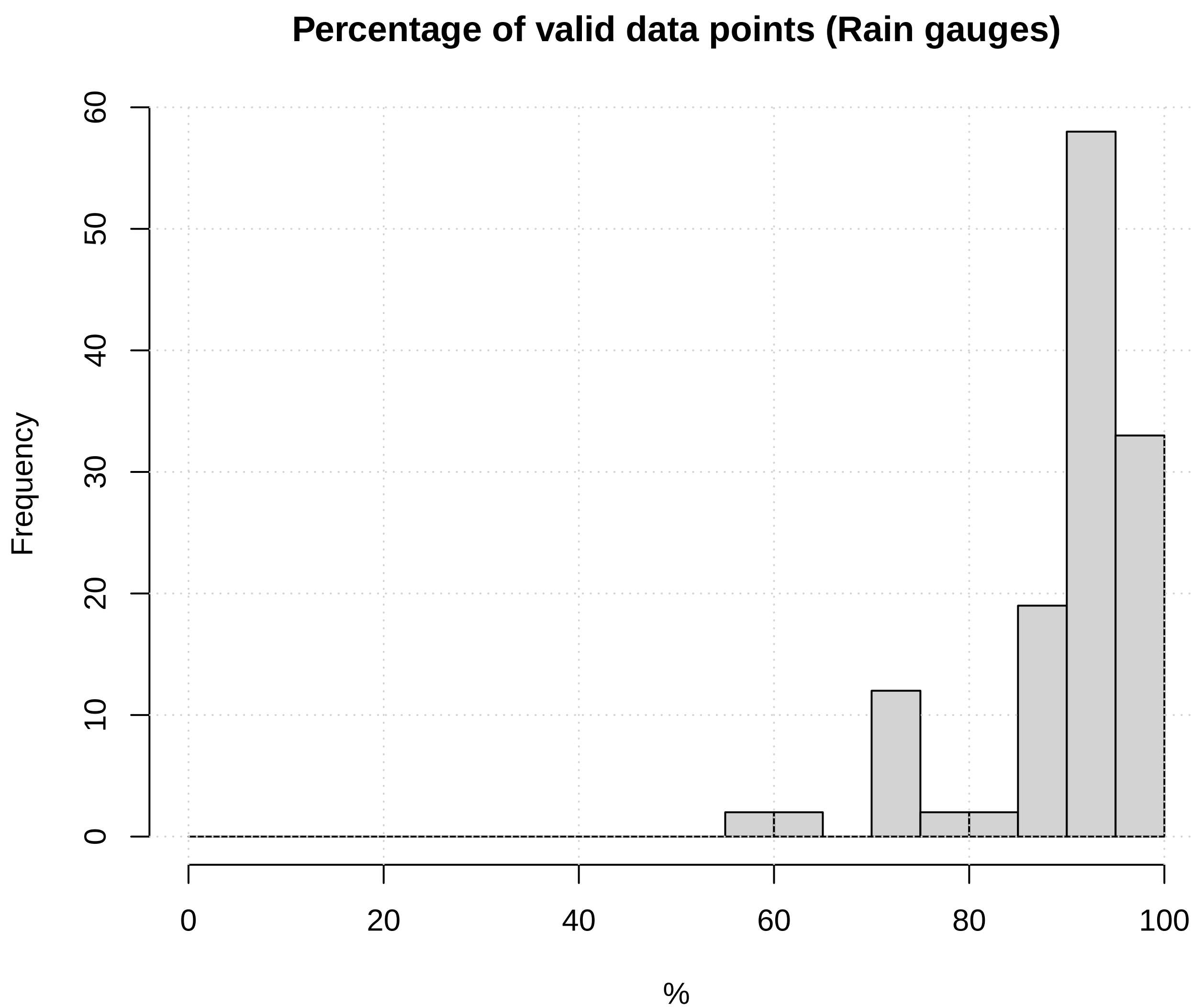

```r
hist(ry,main='Correlation(Rain gauges, ERA5) - annual rainfall totals',xlim=c(-1,1))
grid()
```




```
hist(offsety,main='Mean difference for common valid years (Rain gauges - ERA5)',
     breaks=seq(-4300,4300,by=100)); grid()
```




```
hist(100*apply(X,2,'nv')/length(index(X)),
     main='Percentage of valid data points (Rain gauges)',
     xlab='%',xlim=c(0,100),breaks=seq(0,100,by=5)); grid()
```



Only a small number of stations had RMSE greater than 500 mm, most had correlation above 0.5, and the offset was predominantly negative, suggesting that ERA5 had somewhat higher annual rainfall than the observations. Most of the selected rain gauges gave more or less a complete coverage.

2.1.8 CHIRPS statistics: RMSE, correlation, & bias

```
## Pick the annual precipitation from the first station
ns <- length(loc(X))
print(paste(ns,'stations'))
```

```
## [1] "130 stations"
```

```
rmse <- rep(NA,ns); attr(rmse,'location') <- loc(X)
r <- rmse; offset <- rmse
rmsey <- rmse; ry <- rmse; offsety <- rmse
for (is in 1:ns) {
  ## Because of missing data/days, we get more robust results by estimating the mean
  ## and then multiply by number of days per year.
  X1 <- 365.25*annual(subset(X,is=is),
                      FUN='mean',nmin=90,start=year.start)   # Station data - capital letters
  ## But if the lacking data are for the dry seasons, then it's better to choose the sum
```

```r
  Y1 <- annual(subset(X,is=is),
               FUN='sum',nmin=90,start=year.start)           # Station data - capital letters
  x1 <- subset(RR.chirps,is=is)                              # CHIRPS

  ## Estimate statistics such as RMSE, and correlation
  ## Here we merge station data with ERA5 reanalysis
  xy <- merge(zoo(X1[is.finite(X1)]),zoo(x1),all=FALSE)
  ok <- is.finite(xy[,1]) & is.finite(xy[,2])
  rmse[is] <- round(RMSE(xy[ok,1],xy[ok,2]))
  r[is] <- round(cor(xy[ok,1],xy[ok,2]),2)
  offset[is] <- round(mean(xy[ok,1]) - mean(xy[ok,2]))  ## Station mean - ERA5 mean
  xy <- merge(zoo(Y1[is.finite(X1)]),zoo(x1),all=FALSE)
  rmsey[is] <- round(RMSE(xy[ok,1],xy[ok,2]))
  ry[is] <- round(cor(xy[ok,1],xy[ok,2]),2)
  offsety[is] <- round(mean(xy[ok,1]) - mean(xy[ok,2]))  ## Station mean - ERA5 mean

}
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced

## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
print(summary(r)); print(summary(ry))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -0.1900  0.3500  0.5400  0.5002  0.7000  0.9300      10

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -0.3100  0.3300  0.5300  0.4836  0.6825  0.9400      10
```

```
attr(X,'rmse.chirps') <- rmse
attr(X,'cor.chirps') <- r
attr(X,'offset.chirps') <- offset

## The annual rainfall totals - Jan-to-Dec
#atX <- combine.stations(365.25*annual(X,FUN='mean',nmin=300),Mdg)
atX <- 365.25*annual(X,FUN='mean',nmin=100,start=year.start)
```

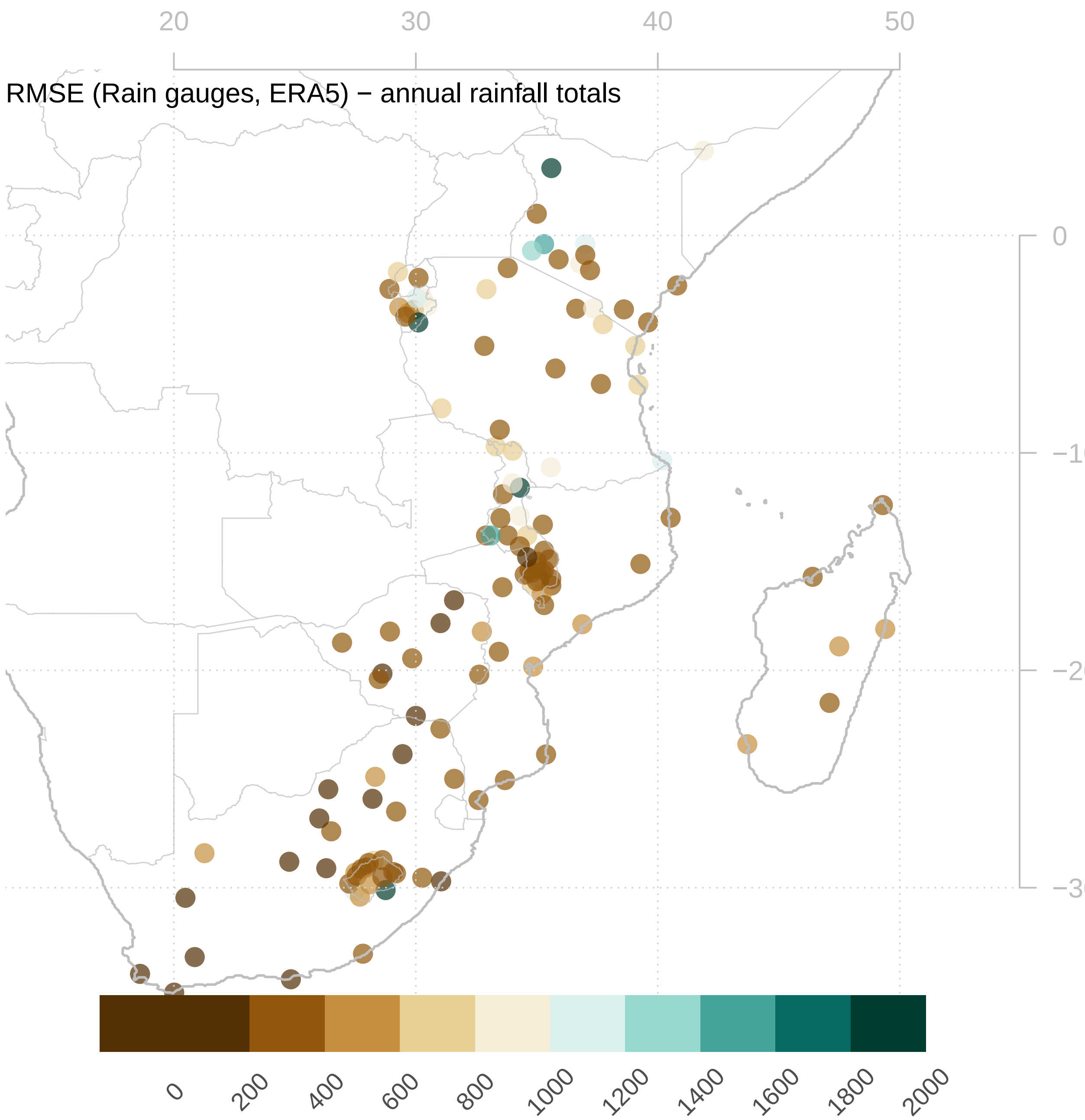
```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
attr(atX,'rmse.chirps') <- rmsey
attr(atX,'cor.chirps') <- ry
attr(atX,'offset.chirps') <- offsety
```

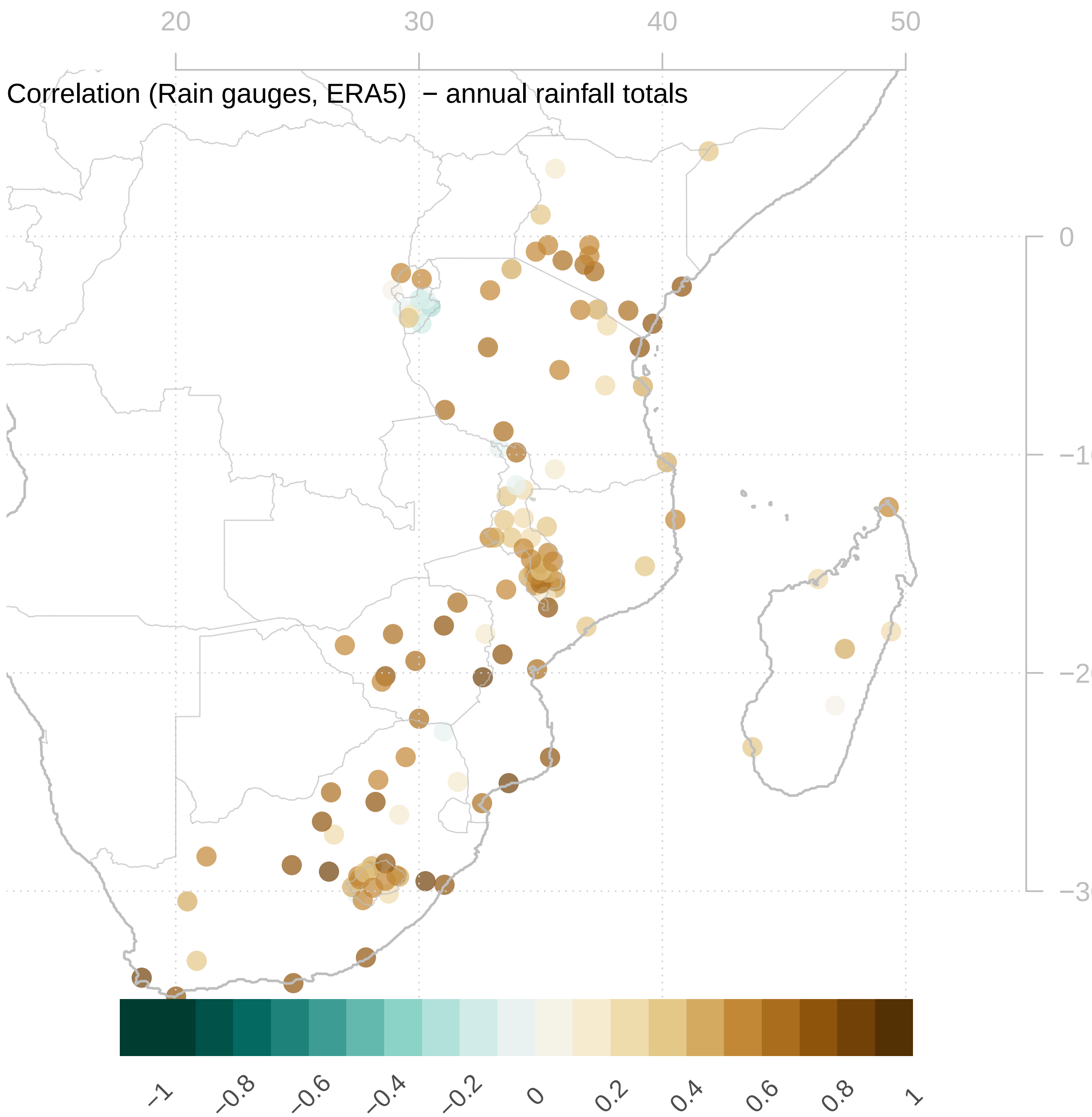
2.1.9 Maps of statistics

We can show how the similarity varies geographically through maps

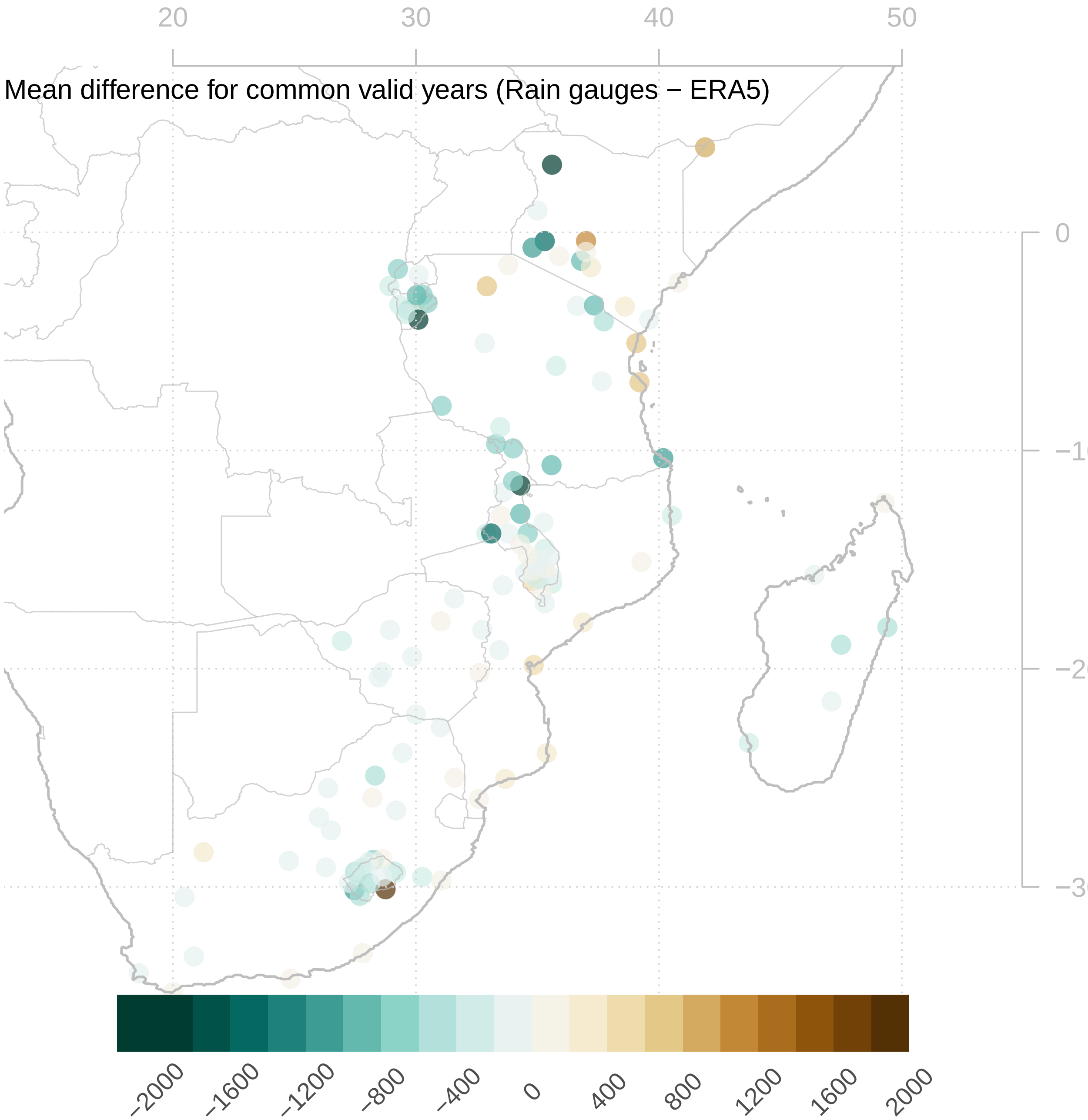
```
map(atX,FUN='rmse',add.text = FALSE,colbar=list(breaks=seq(0,2000,length=11),pal='precip.ipcc'),
    main='RMSE (Rain gauges, ERA5) - annual rainfall totals',cex=1.5,border=TRUE)
```




```r
map(atX,FUN='cor',add.text = FALSE,
    colbar=list(breaks=seq(-1,1,length=21), pal='precip.ipcc',rev=TRUE),border=TRUE,
    main='Correlation (Rain gauges, ERA5)  - annual rainfall totals',cex=1.5)
```




```
map(atX,FUN='offset',add.text = FALSE, border=TRUE,
    colbar=list(breaks=seq(-2000,2000,length=21), pal='precip.ipcc',rev=TRUE),
    main='Mean difference for common valid years (Rain gauges - ERA5)',cex=1.5)
```



The maps for RMSE suggests that a few stations from Rwanda, Malawi and Mozambique had some differences regarding interpolated rainfall statistics from ERA5. The statistics are presented for each location in the table below, sorted according to the RMSE (starting with greatest RMSE on the top):

```
rank <- order(rmsey,decreasing = TRUE)
nv <- apply(atX,2,'nv')
print(cbind(cntr(atX),rmsey,ry,offsety,nv)[rank,])
```

```
##                        rmsey  ry      offsety nv
## X.109 "Lesotho"        "2891" "0.14"  "2844"  "40"
## X.78  "Kenya"          "1467" "0.56"  "-1454" "39"
## X.75  "Kenya"          "1209" "0"     "1198"  "39"
## X.80  "Kenya"          "1125" "0.45"  "1092"  "39"
## X.77  "Kenya"          "1098" "0.79"  "-1089" "39"
```

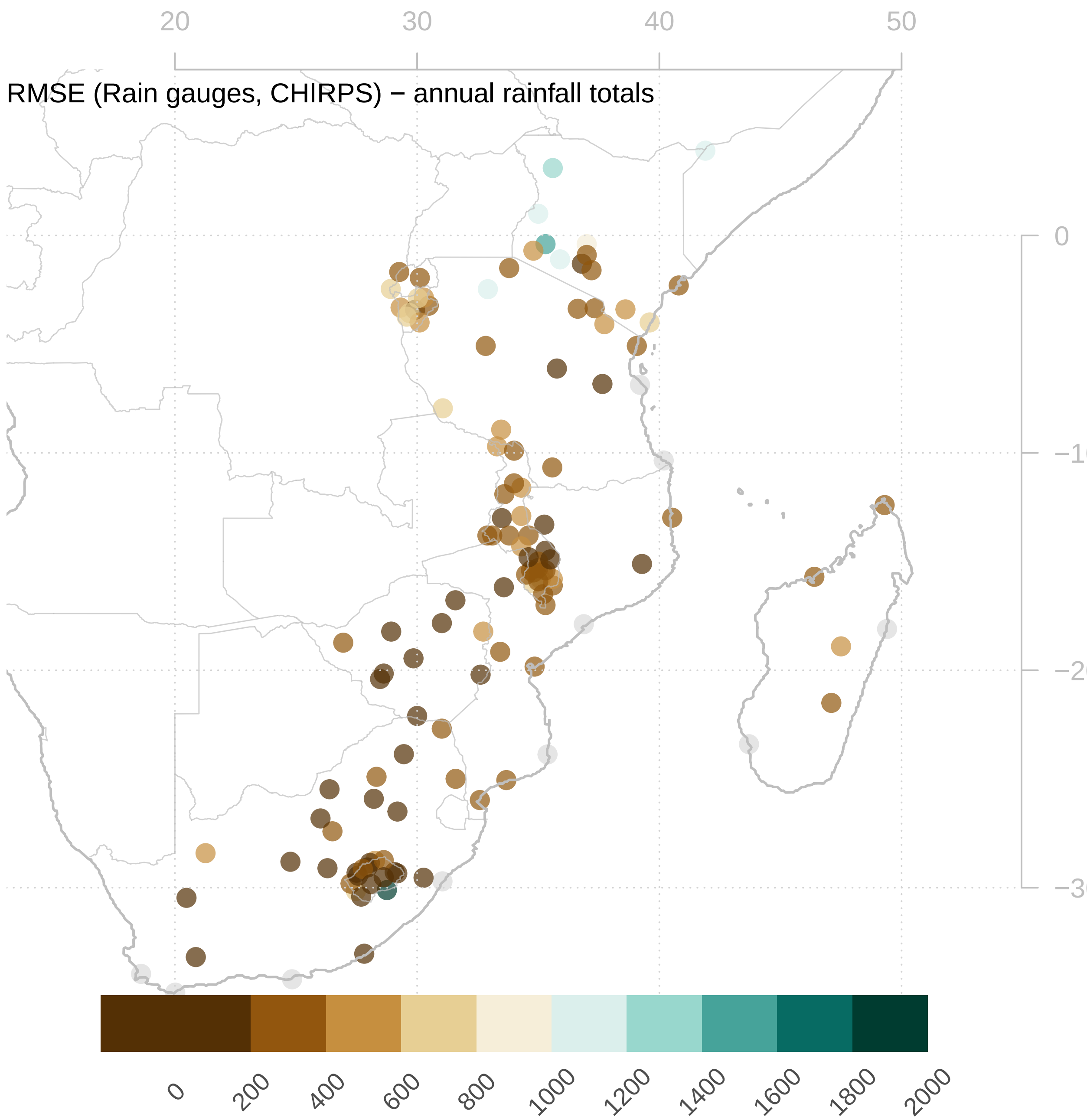
```
## X.76  "Kenya"        "1020" "0.36"  "1003"  "39"
## X.90  "Tanzania"     "1003" "0.3"   "959"   "42"
## X.81  "Kenya"        "871"  "0.33"  "819"   "39"
## X.126 "Burundi"      "775"  "-0.1"  "-710"  "31"
## X.108 "Lesotho"      "764"  "0.02"  "711"   "40"
## X.56  "Malawi"       "759"  "0.69"  "711"   "40"
## X.86  "Kenya"        "748"  "0.5"   "-698"  "39"
## X.73  "Rwanda"       "668"  "-0.18" "-599"  "41"
## X.53  "Malawi"       "631"  "0.23"  "514"   "40"
## X.130 "Burundi"      "626"  "0.24"  "-596"  "31"
## X.129 "Burundi"      "624"  "-0.31" "-582"  "31"
## X.98  "Tanzania"     "609"  "0.49"  "-592"  "33"
## X.128 "Burundi"      "575"  "-0.05" "-527"  "31"
## X.41  "Malawi"       "572"  "-0.04" "-531"  "40"
## X.102 "Madagascar"   "553"  "0.13"  "-493"  "41"
## X.48  "Malawi"       "516"  "0.53"  "-463"  "40"
## X.52  "Malawi"       "514"  "0.38"  "-471"  "40"
## X.92  "Tanzania"     "481"  "0.39"  "420"   "42"
## X.84  "Kenya"        "464"  "0.63"  "395"   "39"
## X.20  "Zimbabwe"     "460"  "0.16"  "-161"  "41"
## X.60  "Malawi"       "458"  "0.61"  "-362"  "40"
## X.79  "Kenya"        "447"  "0.69"  "-414"  "39"
## X.99  "Tanzania"     "444"  "0.6"   "-420"  "41"
## X.39  "South Africa" "443"  "0.29"  "398"   "42"
## X.110 "Lesotho"      "442"  "0.17"  "-412"  "40"
## X.127 "Burundi"      "435"  "0.09"  "-379"  "31"
## X.49  "Malawi"       "431"  "0.4"   "341"   "40"
## X.123 "Burundi"      "428"  "0.3"   "-388"  "31"
## X.65  "Malawi"       "397"  "0.62"  "331"   "40"
## X.87  "Kenya"        "395"  "0.38"  "-194"  "41"
## X.40  "South Africa" "393"  "0.22"  "-371"  "42"
## X.104 "Madagascar"   "384"  "-0.19" "-203"  "41"
## X.55  "Malawi"       "381"  "0.74"  "-332"  "40"
## X.105 "Madagascar"   "379"  "0.52"  "-287"  "41"
## X.46  "Malawi"       "366"  "0.53"  "-325"  "40"
## X.68  "Malawi"       "359"  "0.24"  "-216"  "40"
## X.42  "Malawi"       "351"  "0.47"  "-310"  "40"
## X.62  "Malawi"       "350"  "0.36"  "184"   "40"
## X.10  "Mozambique"   "345"  "0.57"  "107"   "42"
## X.57  "Malawi"       "343"  "0.41"  "-276"  "40"
## X.47  "Malawi"       "338"  "0.51"  "188"   "40"
## X.74  "Rwanda"       "329"  "0.56"  "-296"  "41"
## X.64  "Malawi"       "328"  "0.33"  "42"    "40"
## X.103 "Madagascar"   "320"  "0.3"   "-199"  "41"
## X.58  "Malawi"       "319"  "0.45"  "-100"  "40"
## X.88  "Tanzania"     "319"  "0.51"  "141"   "42"
## X.69  "Malawi"       "317"  "0.21"  "188"   "40"
## X.93  "Tanzania"     "317"  "0.36"  "-225"  "42"
## X.125 "Burundi"      "288"  "-0.16" "-6"    "31"
## X.118 "Lesotho"      "287"  "0.67"  "223"   "40"
## X.89  "Tanzania"     "273"  "0.61"  "-87"   "42"
## X.59  "Malawi"       "258"  "0.72"  "-168"  "40"
## X.38  "South Africa" "257"  "0.09"  "128"   "42"
## X.83  "Kenya"        "256"  "0.73"  "177"   "39"
```

```
## X.6   "Mozambique"   "250"  "0.72"  "-129"  "40"
## X.7   "Mozambique"   "246"  "0.73"  "16"    "42"
## X.51  "Malawi"       "245"  "0.46"  "-130"  "40"
## X.85  "Kenya"        "239"  "0.29"  "19"    "39"
## X.112 "Lesotho"      "238"  "0.46"  "167"   "33"
## X.67  "Malawi"       "237"  "0.57"  "-14"   "40"
## X.94  "Tanzania"     "236"  "0.75"  "-50"   "42"
## X.9   "Mozambique"   "234"  "0.68"  "112"   "42"
## X.91  "Tanzania"     "231"  "0.49"  "-127"  "42"
## X.66  "Malawi"       "229"  "0.54"  "18"    "40"
## X.45  "Malawi"       "228"  "0.35"  "-145"  "40"
## X.72  "Rwanda"       "226"  "0.21"  "28"    "41"
## X.119 "Lesotho"      "226"  "0.48"  "-16"   "26"
## X.43  "Malawi"       "225"  "0.14"  "-48"   "40"
## X.100 "Tanzania"     "219"  "0.5"   "-59"   "42"
## X.15  "Zimbabwe"     "218"  "0.55"  "-74"   "41"
## X.61  "Malawi"       "218"  "0.59"  "1"     "40"
## X.54  "Malawi"       "212"  "0.88"  "171"   "40"
## X.22  "South Africa" "210"  "0.42"  "-102"  "42"
## X.1   "Mozambique"   "206"  "0.76"  "-25"   "40"
## X.36  "South Africa" "206"  "0.1"   "-60"   "28"
## X.63  "Malawi"       "201"  "0.74"  "71"    "40"
## X.17  "Zimbabwe"     "199"  "0.77"  "41"    "41"
## X.96  "Tanzania"     "199"  "0.39"  "23"    "42"
## X.2   "Mozambique"   "197"  "0.64"  "-101"  "42"
## X.70  "Malawi"       "192"  "0.52"  "32"    "40"
## X.5   "Mozambique"   "185"  "0.76"  "-40"   "39"
## X.37  "South Africa" "182"  "0.37"  "-19"   "34"
## X.122 "Lesotho"      "181"  "0.5"   "64"    "40"
## X.113 "Lesotho"      "178"  "0.39"  "-16"   "29"
## X.3   "Mozambique"   "176"  "0.61"  "8"     "42"
## X.111 "Lesotho"      "174"  "0.54"  "17"    "40"
## X.28  "South Africa" "172"  "0.8"   "95"    "42"
## X.44  "Malawi"       "172"  "0.29"  "-40"   "40"
## X.116 "Lesotho"      "172"  "0.66"  "-126"  "40"
## X.71  "Malawi"       "170"  "0.58"  "23"    "39"
## X.82  "Kenya"        "167"  "0.83"  "-110"  "39"
## X.16  "Zimbabwe"     "165"  "0.62"  "-42"   "41"
## X.95  "Tanzania"     "165"  "0.62"  "-74"   "42"
## X.124 "Burundi"      "161"  "0.64"  "-70"   "31"
## X.117 "Lesotho"      "157"  "0.72"  "74"    "35"
## X.32  "South Africa" "149"  "0.63"  "-6"    "42"
## X.121 "Lesotho"      "148"  "0.64"  "-108"  "40"
## X.18  "Zimbabwe"     "145"  "0.78"  "90"    "41"
## X.120 "Lesotho"      "141"  "0.61"  "37"    "40"
## X.19  "Zimbabwe"     "134"  "0.7"   "27"    "41"
## X.11  "Zimbabwe"     "128"  "0.68"  "49"    "41"
## X.50  "Malawi"       "128"  "0.65"  "-8"    "40"
## X.114 "Lesotho"      "126"  "0.79"  "-3"    "35"
## X.115 "Lesotho"      "120"  "0.73"  "4"     "38"
## X.12  "Zimbabwe"     "117"  "0.84"  "21"    "41"
## X.14  "Zimbabwe"     "114"  "0.85"  "-37"   "41"
## X.34  "South Africa" "111"  "0.79"  "28"    "42"
## X.27  "South Africa" "105"  "0.85"  "-7"    "42"
```

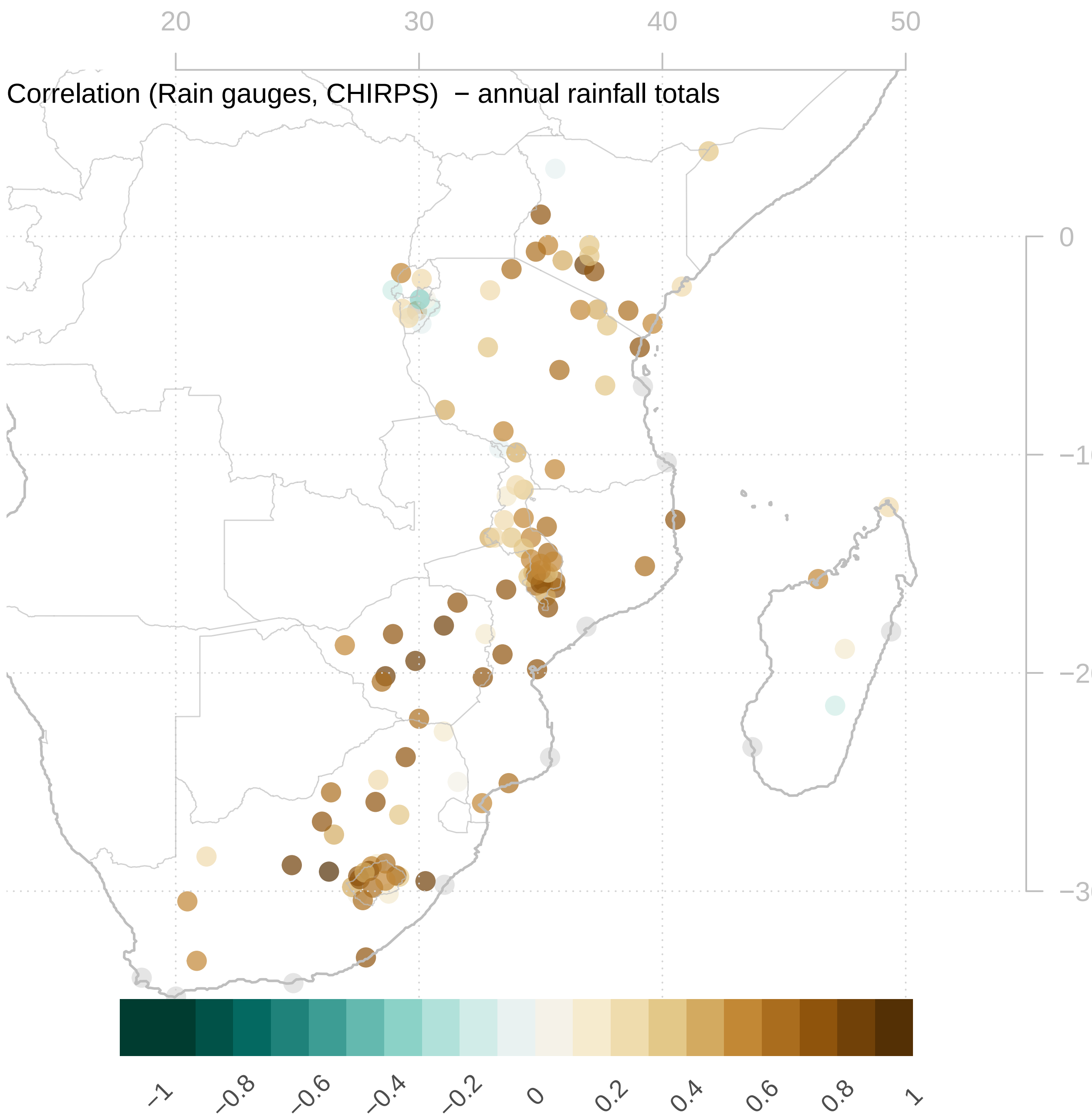
```
## X.29  "South Africa" "100"  "0.79"  "20"   "42"
## X.13  "Zimbabwe"     "91"   "0.86"  "18"   "41"
## X.35  "South Africa" "87"   "0.76"  "-1"   "42"
## X.30  "South Africa" "76"   "0.82"  "10"   "42"
## X.21  "South Africa" "75"   "0.94"  "22"   "42"
## X.23  "South Africa" "58"   "0.58"  "11"   "33"
## X.31  "South Africa" "56"   "0.53"  "-35"  "26"
## X.4   "Mozambique"   "NaN"  NA      "NaN"  "42"
## X.8   "Mozambique"   "NaN"  NA      "NaN"  "42"
## X.24  "South Africa" "NaN"  NA      "NaN"  "42"
## X.25  "South Africa" "NaN"  NA      "NaN"  "42"
## X.26  "South Africa" "NaN"  NA      "NaN"  "42"
## X.33  "South Africa" "NaN"  NA      "NaN"  "41"
## X.97  "Tanzania"     "NaN"  NA      "NaN"  "42"
## X.101 "Tanzania"     "NaN"  NA      "NaN"  "42"
## X.106 "Madagascar"   "NaN"  NA      "NaN"  "41"
## X.107 "Madagascar"   "NaN"  NA      "NaN"  "41"
```

The same stations that had high RMSE also had low correlation and large offsets. They also had fairly complete data records, so their poor match were not due to sample size.

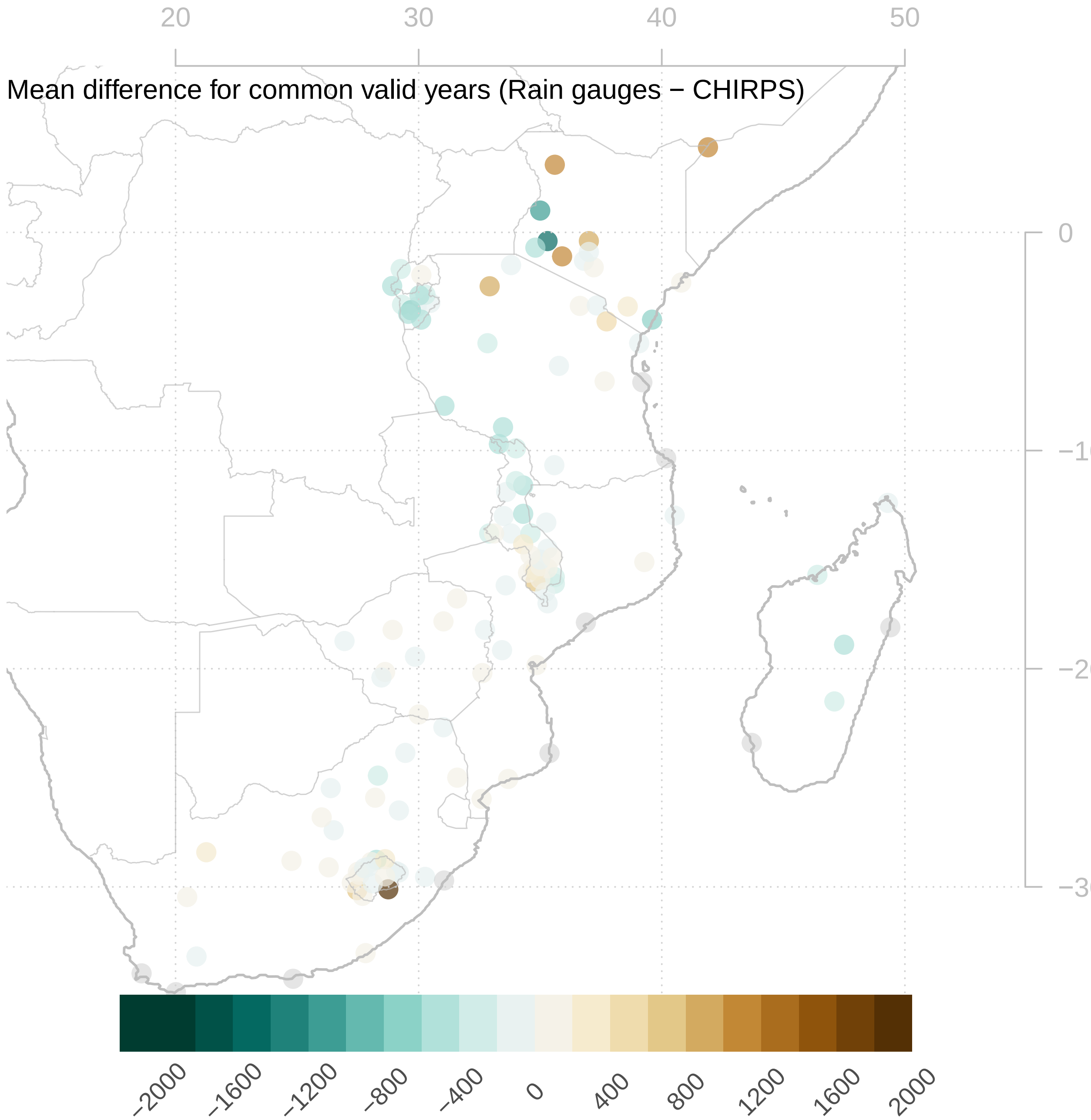
```
map(atX,FUN='rmse.chirps',add.text = FALSE,colbar=list(breaks=seq(0,2000,length=11),pal='precip.ipcc'),
    main='RMSE (Rain gauges, CHIRPS) - annual rainfall totals',cex=1.5,border=TRUE)
```




```r
map(atX,FUN='cor.chirps',add.text = FALSE,
    colbar=list(breaks=seq(-1,1,length=21), pal='precip.ipcc',rev=TRUE),border=TRUE,
    main='Correlation (Rain gauges, CHIRPS)  - annual rainfall totals',cex=1.5)
```



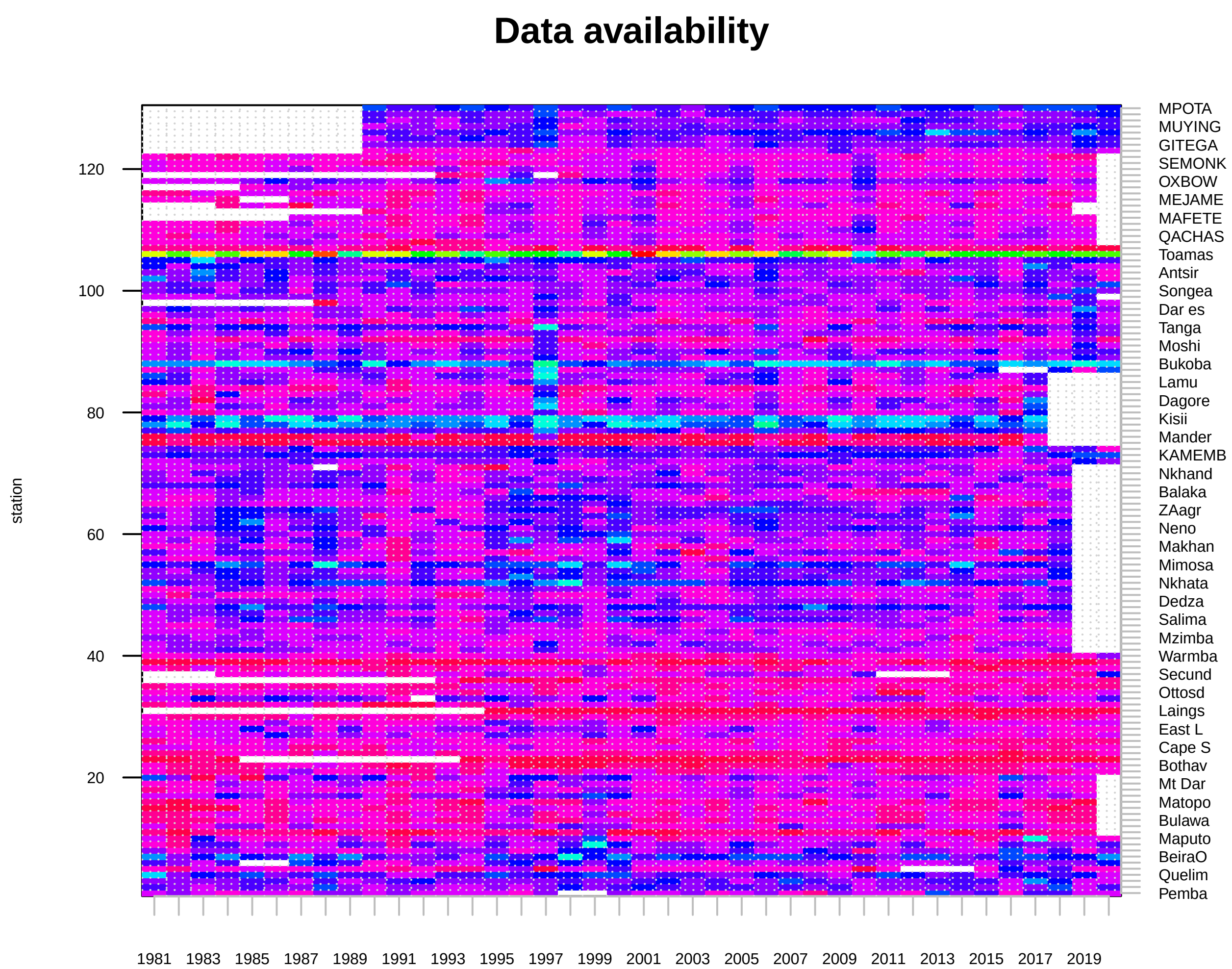

```
map(atX,FUN='offset.chirps',add.text = FALSE, border=TRUE,
    colbar=list(breaks=seq(-2000,2000,length=21), pal='precip.ipcc',rev=TRUE),
    main='Mean difference for common valid years (Rain gauges - CHIRPS)',cex=1.5)
```



2.1.10 Analysis of collective sites through PCA for mean precipitation

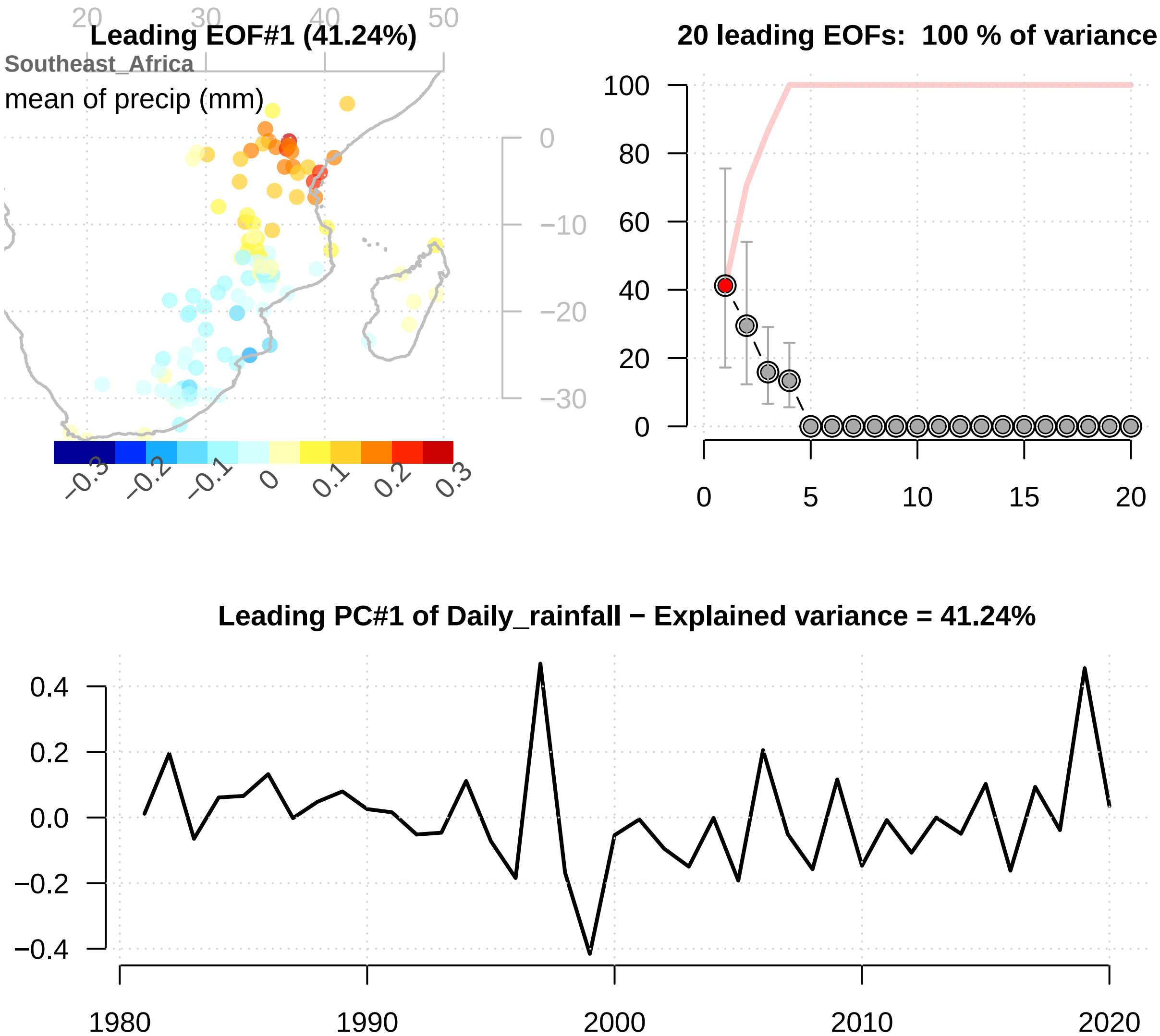
We used a regression-based empirical-statistical downscaling (ESD) to compare the rain gauge data with corresponding data from ERA5 interpolated to same coordinates. To do this for the whole group of stations, we applied principal component analysis (PCA). The PCA requires complete data series with no missing values, so we ran `diagnose()` to check the situation with missing data and used `pcafill()` to fill the gaps with interpolated data based on the assumption of a fixed spatio-temporal covariance structure (this filling in doesn't add any 'new information', but sweeps the problems under the carpet, so to speak).

```
## Estimate PCA for evaluation of downscaling potential
Z <- subset(atX,it=c(1981,2020))
diagnose(Z)
```

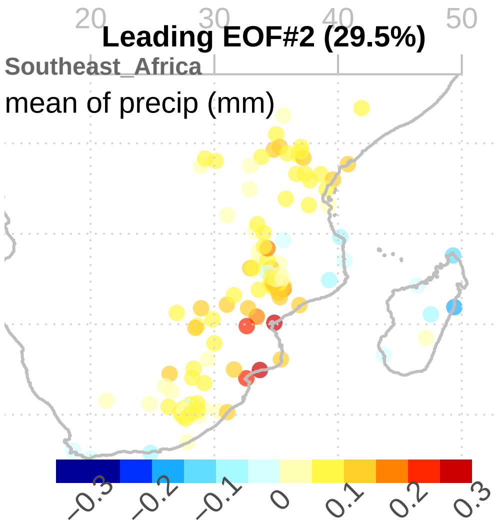


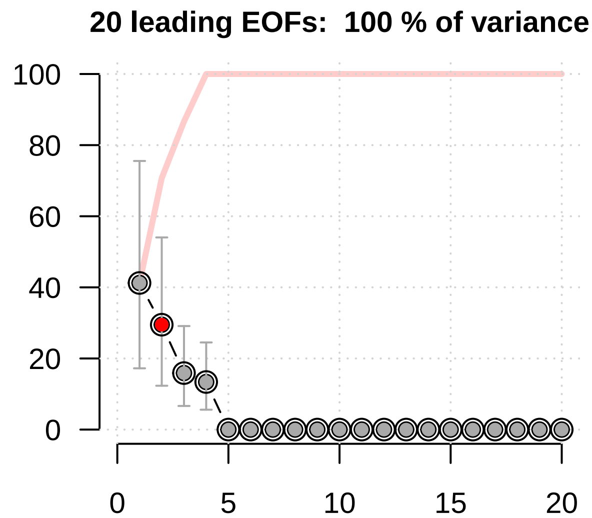
Southeast_Africa

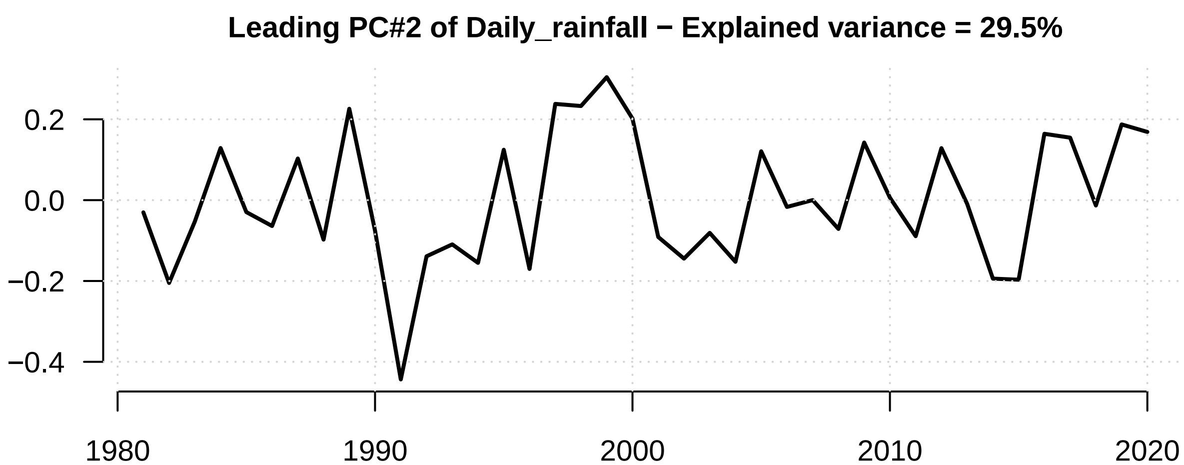
```
nv <- apply(Z,2,'nv')
Z <- subset(Z,is=nv >=33)
Y <- pcafill(Z)
pca <- PCA(Y)
plot(pca,new=FALSE)
```



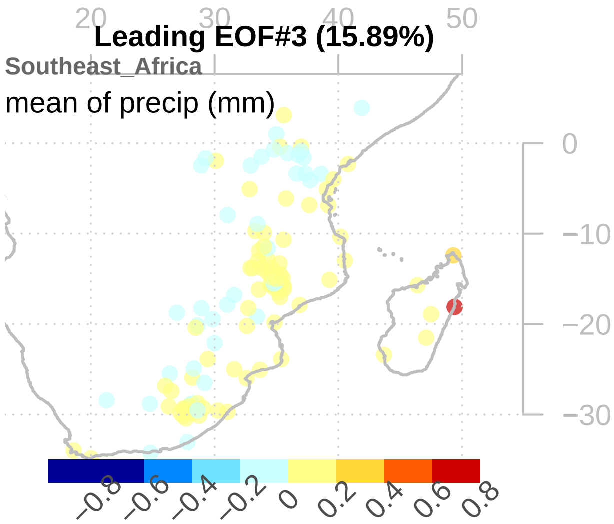

```
## Show the results for the second PCA
plot(pca,ip=2,new=FALSE)
```

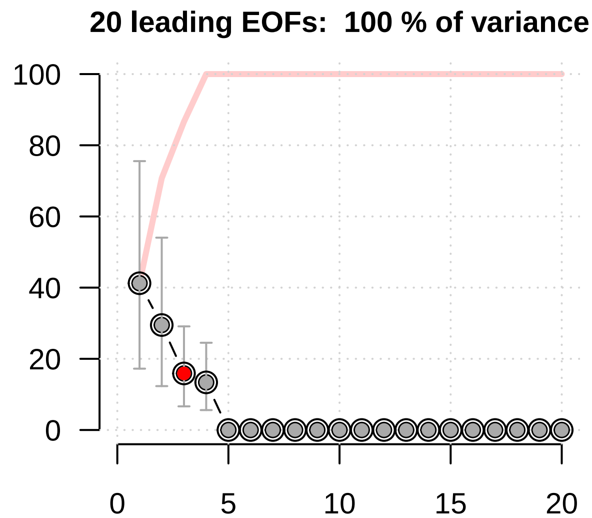


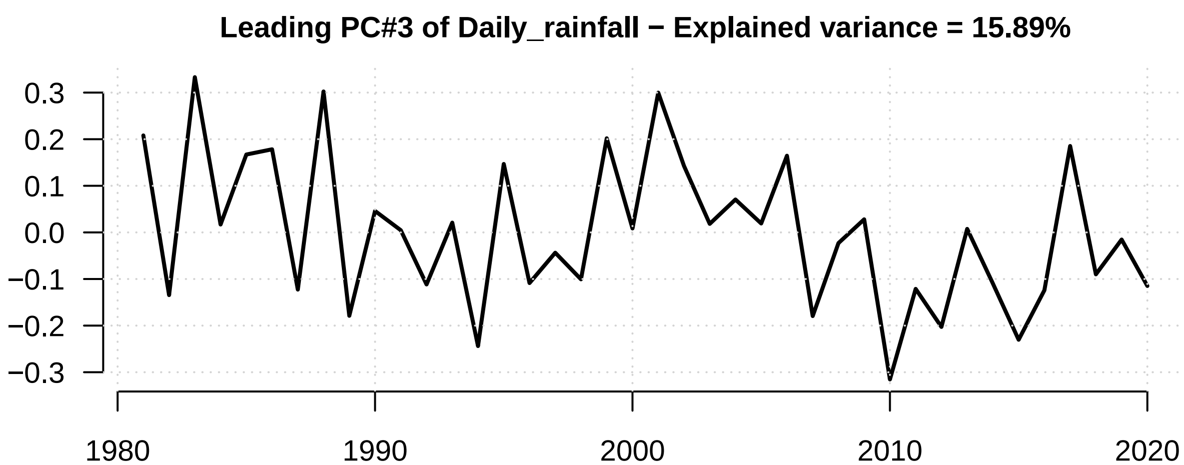




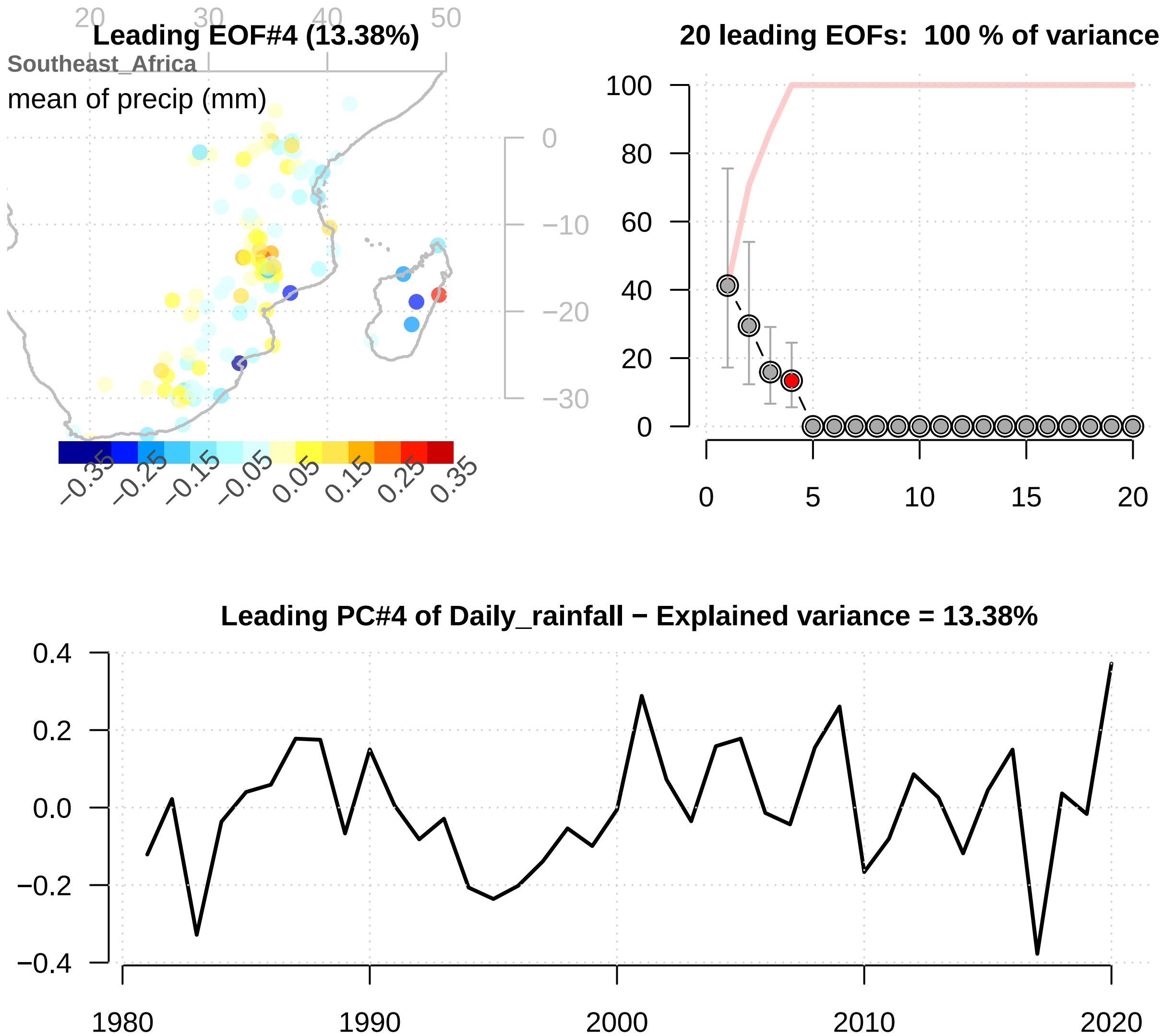

```
## Show the results for the third PCA
plot(pca,ip=3,new=FALSE)
```








```
## Show the results for the fourth PCA
plot(pca,ip=4,new=FALSE)
```



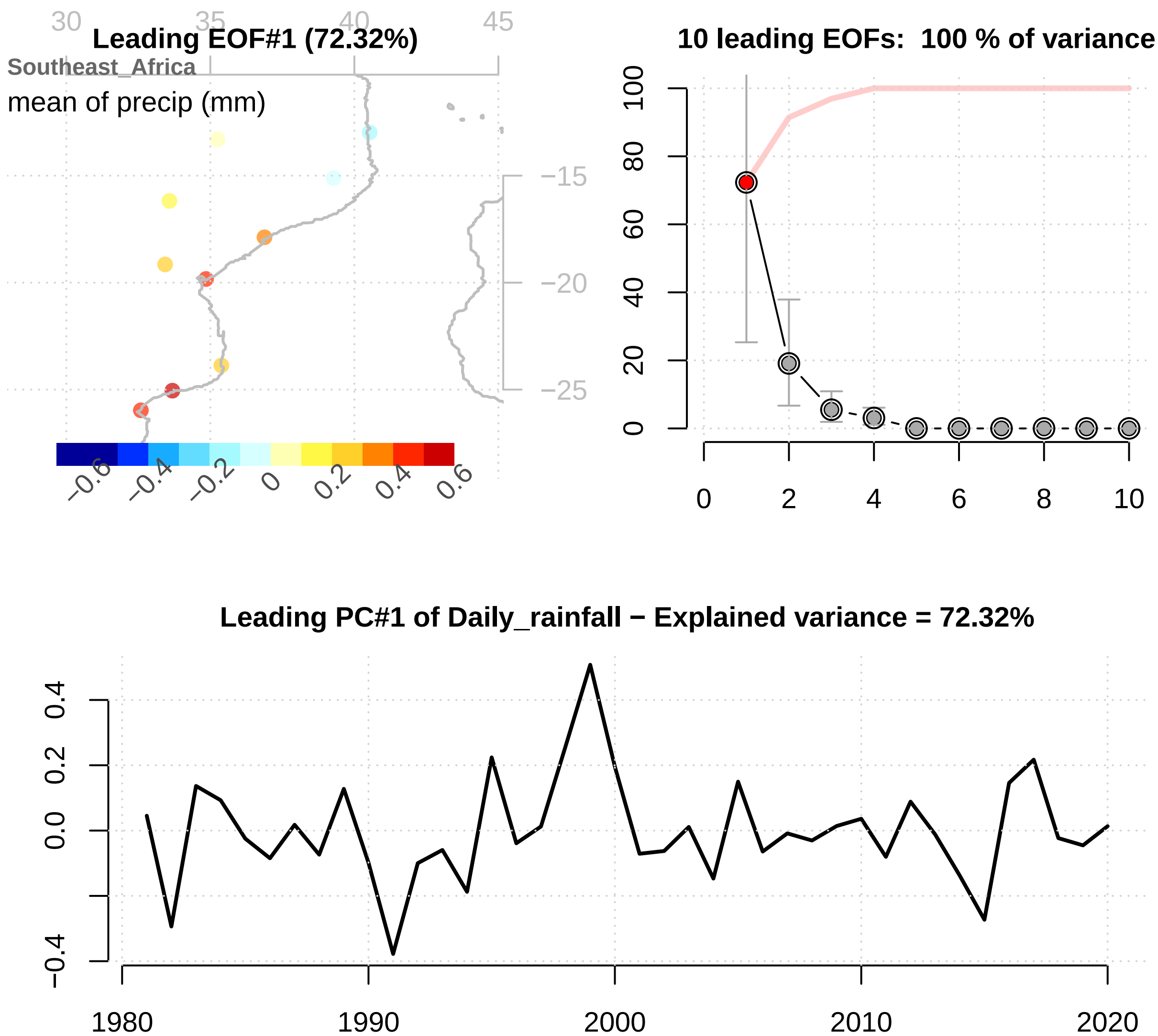
The eigenvalues from the PCA (upper left panels in some of the plots) indicated that the four leading modes sufficed to represent 100% of the variance in the annual rainfall totals for the region.

The leading PCA stands out whereas the three following modes have similar eigenvalues and are probably 'degenerate' (not well-resolved). The second mode is strongly dominated by one station (Xai-xai, Mozambique). The annual rainfall recorded in Maputo and Napui carried large weights in the third PCA mode and Maputo was an outlier in the fourth.

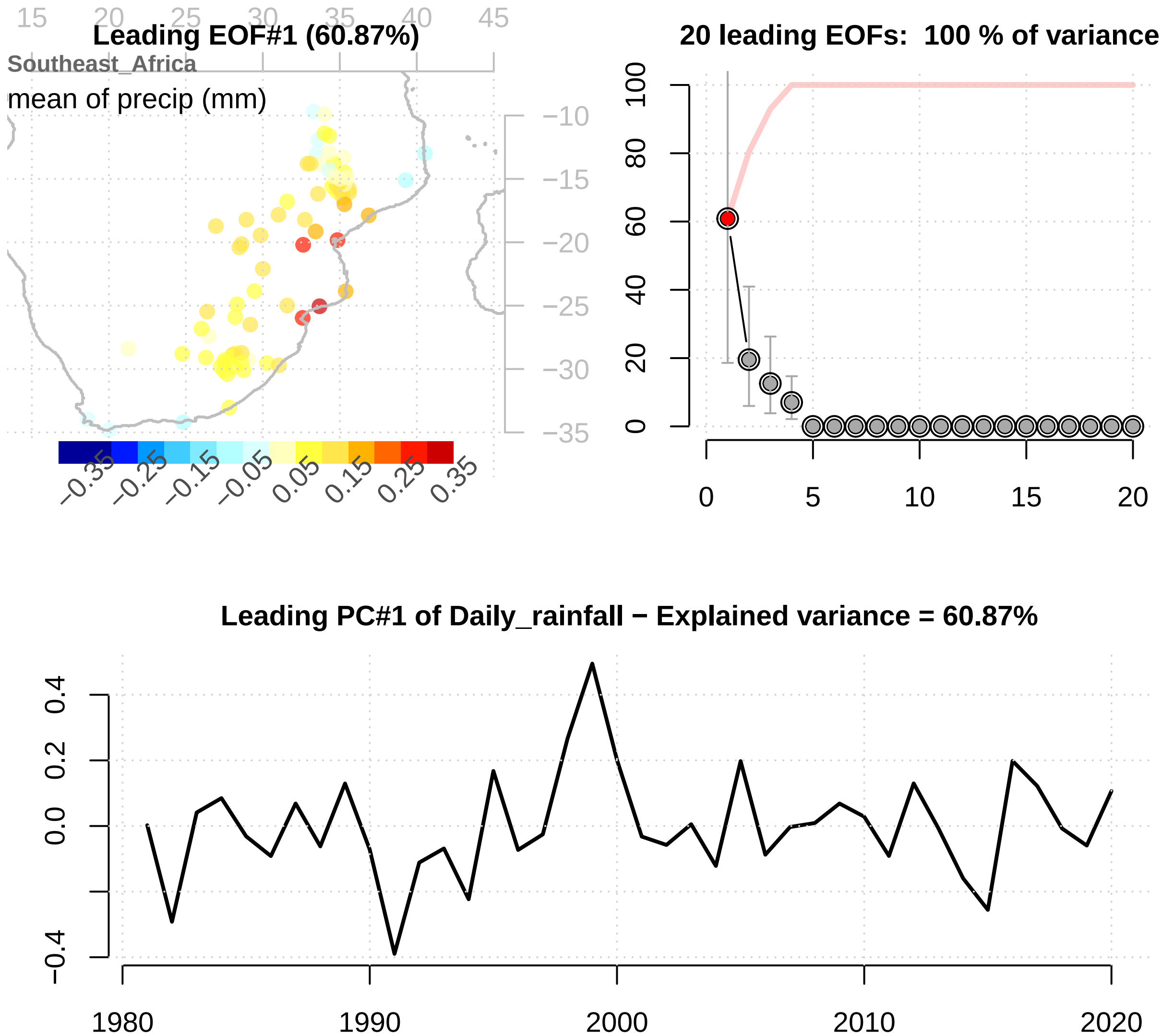
We also estimated PCA for different countries and sub-regions:

```
pca.moz <- PCA(subset(Y,is=list(cntr='Mozambique')))
pca.MMZAL <- PCA(subset(Y,is=list(cntr=c('Mozambique','Malawi','Zimbabwe','South Africa','Lesotho'))))
pca.zim <- PCA(subset(Y,is=list(cntr='Zimbabwe')))
pca.mal <- PCA(subset(Y,is=list(cntr='Malawi')))
pca.tan <- PCA(subset(Y,is=list(cntr='Tanzania')))
pca.saf <- PCA(subset(Y,is=list(cntr='South Africa')))
pca.KRB <- PCA(subset(Y,is=list(cntr=c('Kenya','Rwanda','Burundi'))))
pca.les <- PCA(subset(Y,is=list(cntr=c('Lesotho'))))
Y.mad <- subset(Y,is=list(cntr='Madagascar'))
pca.mad <- PCA(Y.mad)
## Repeat but with outlier site Toamasina exluded.
```

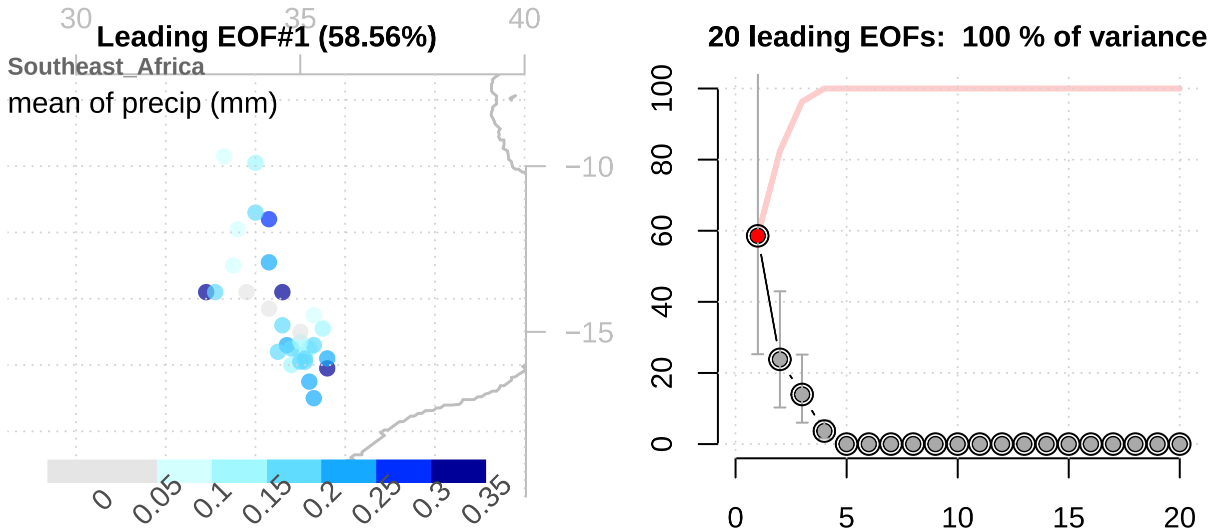
```
Y.mad2 <- subset(Y.mad,is=!is.element(loc(Y.mad),'Toamasina'))
pca.mad2 <- PCA(Y.mad2)
plot(pca.moz,new=FALSE)
```

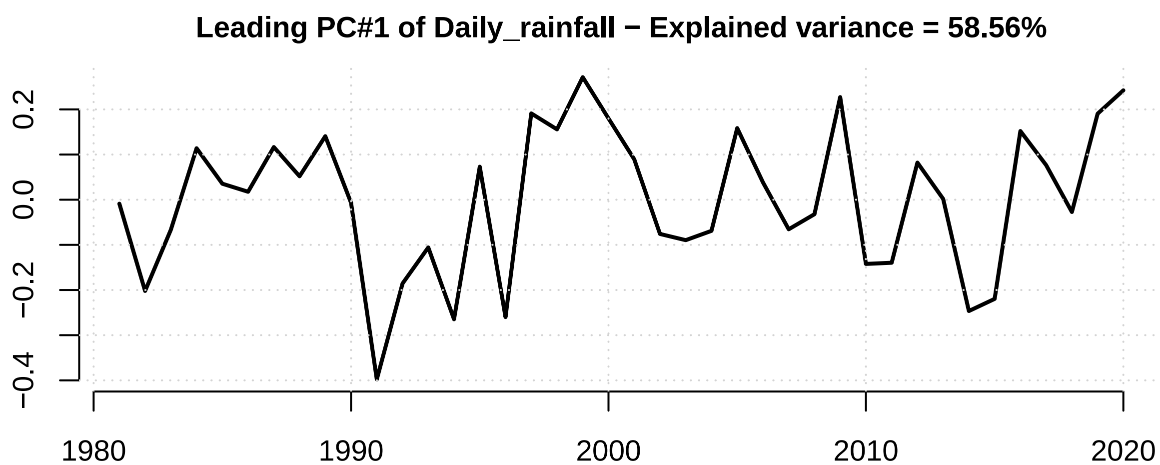



```
plot(pca.MMZAL,new=FALSE)
```

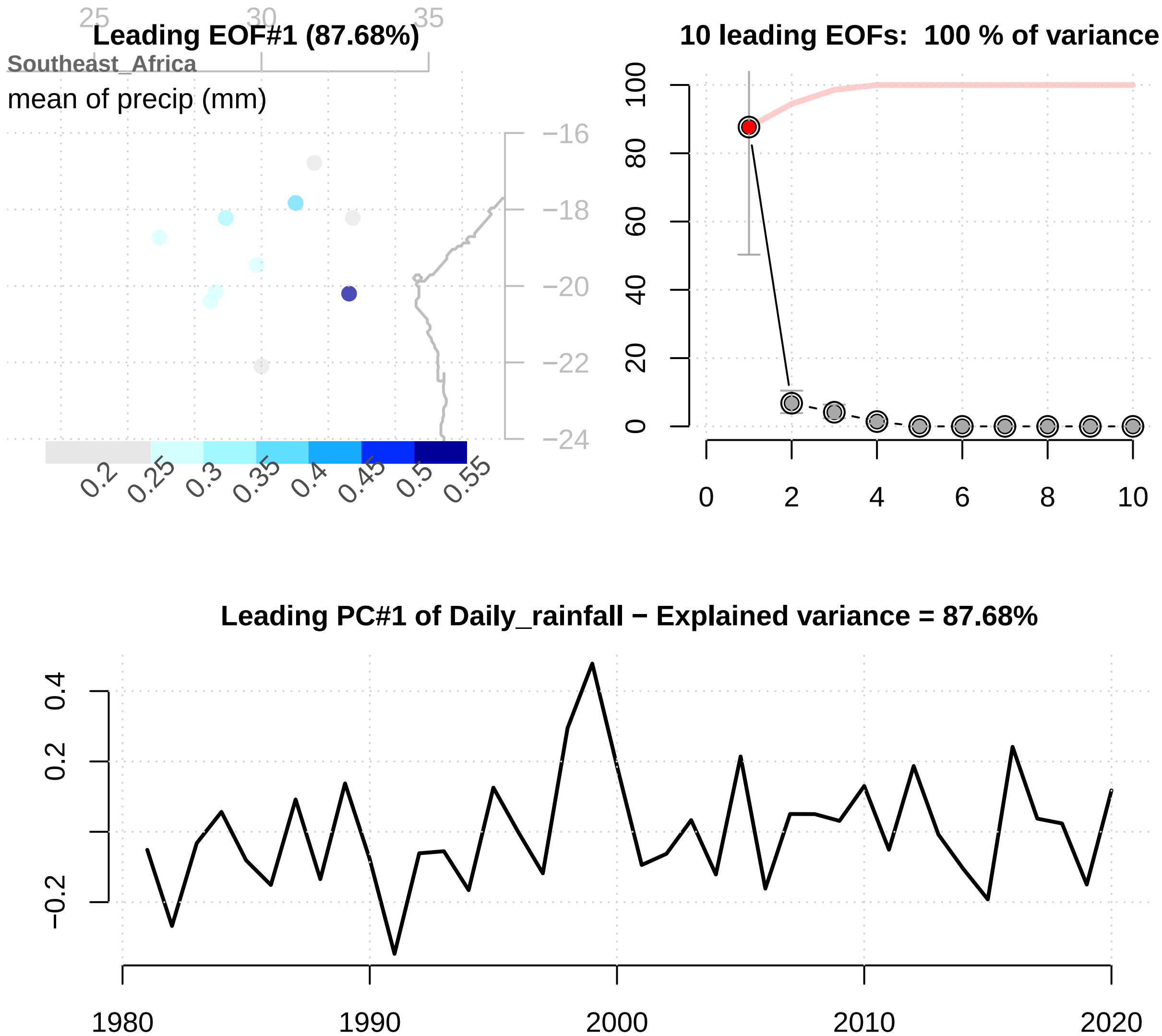



```
plot(pca.mal,new=FALSE)
```

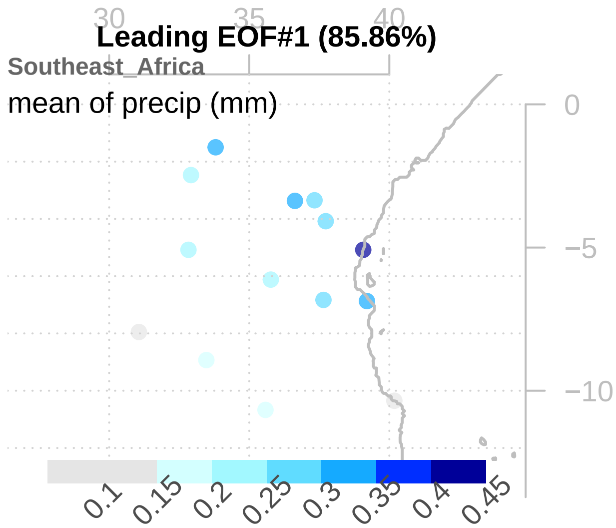


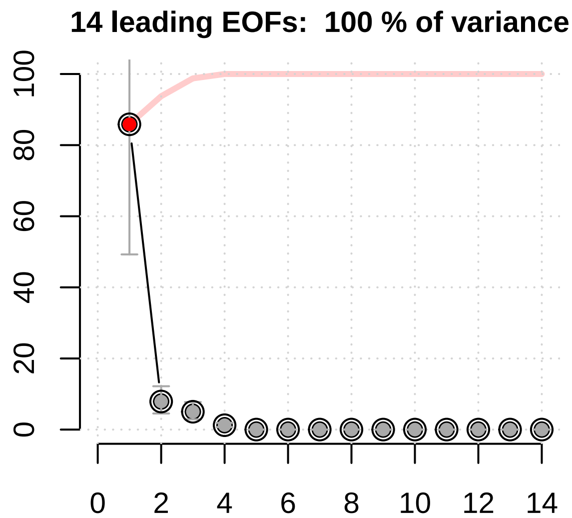


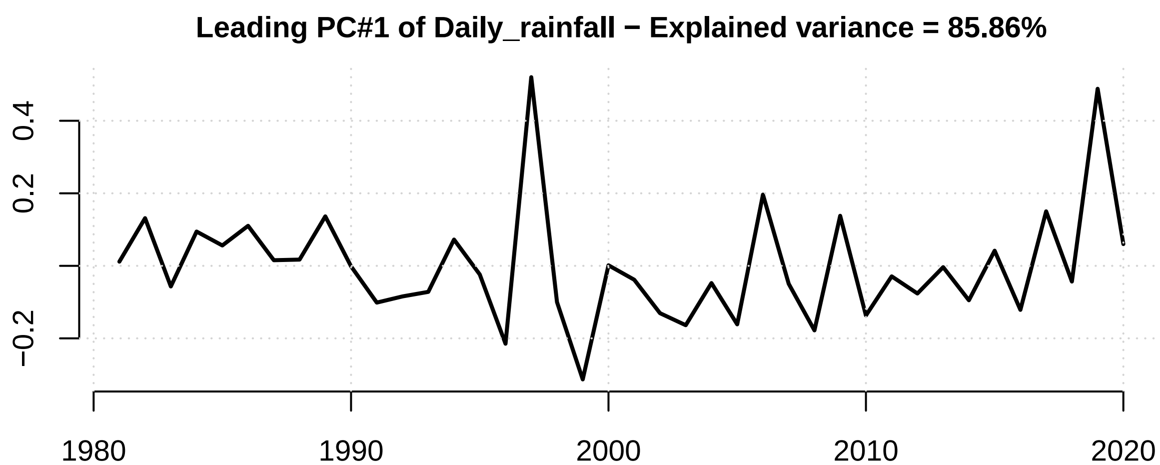

```
plot(pca.zim,new=FALSE)
```



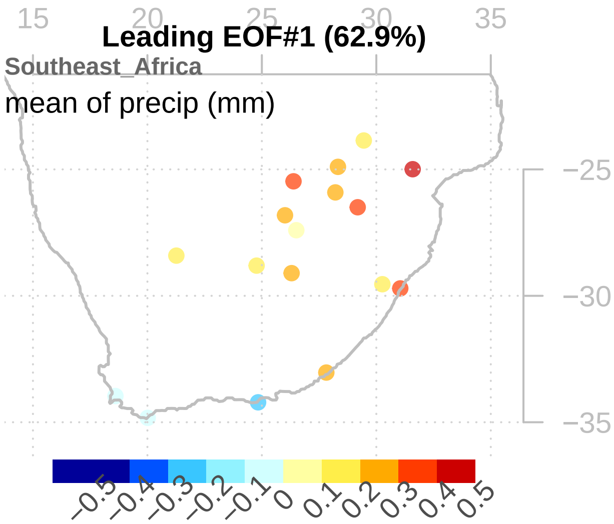

```
plot(pca.tan,new=FALSE)
```

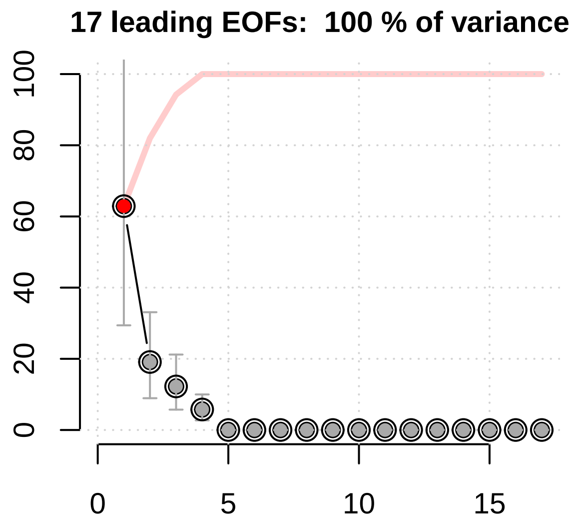


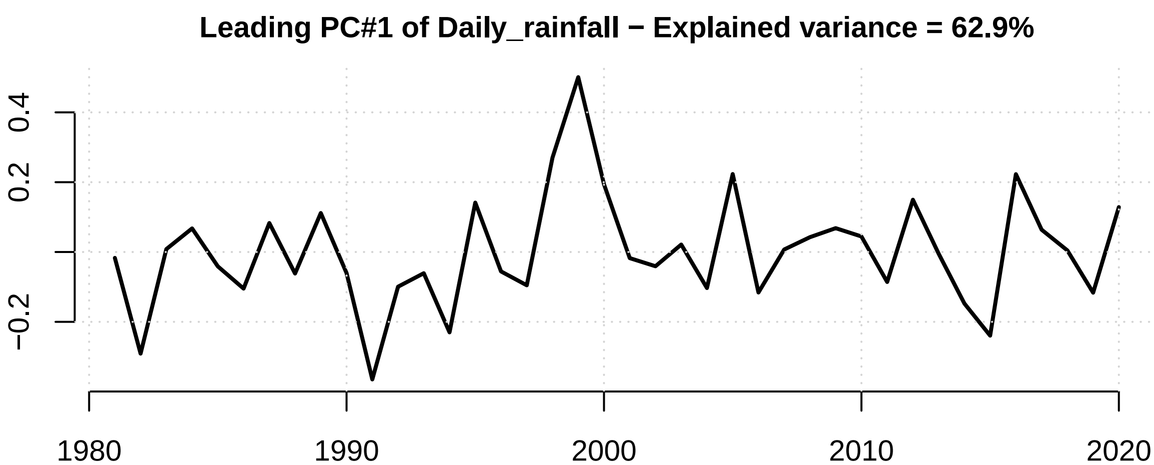




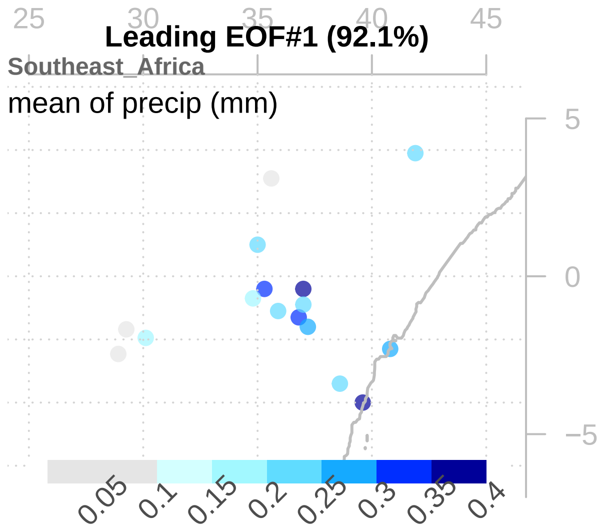

```
plot(pca.saf,new=FALSE)
```

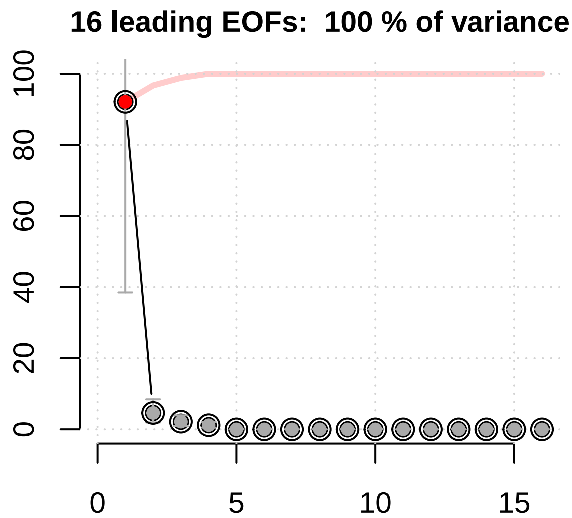


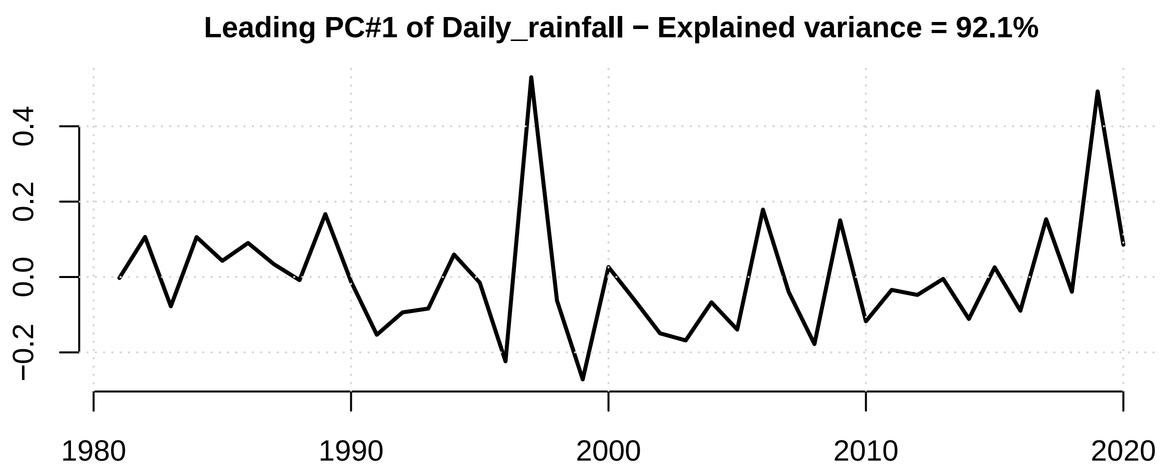




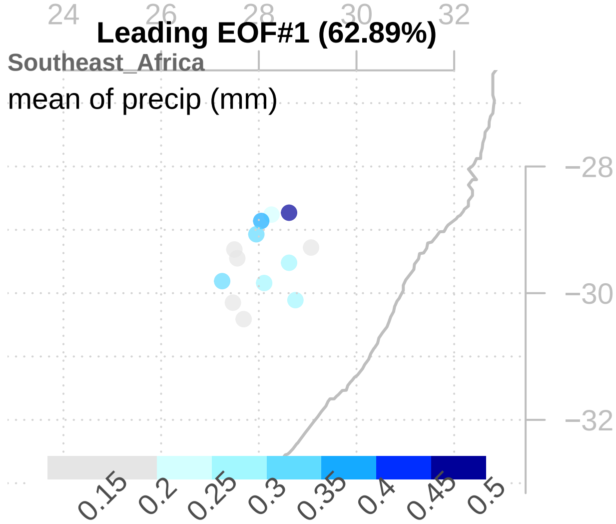

```
plot(pca.KRB,new=FALSE)
```

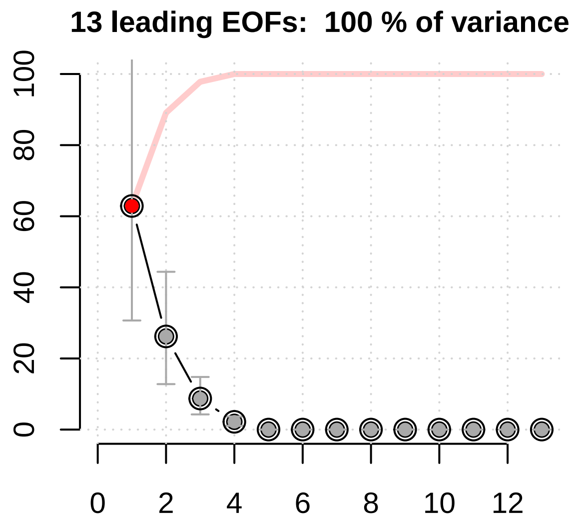


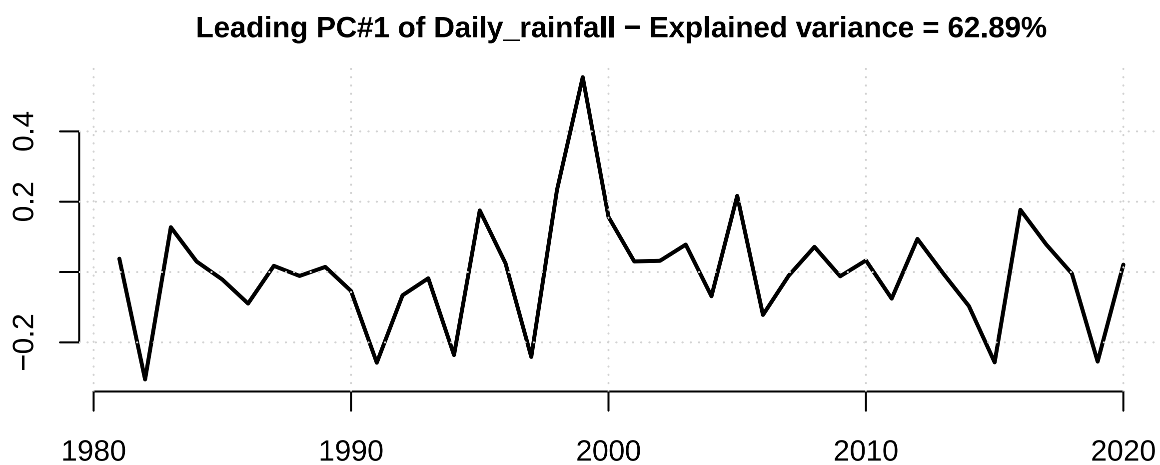




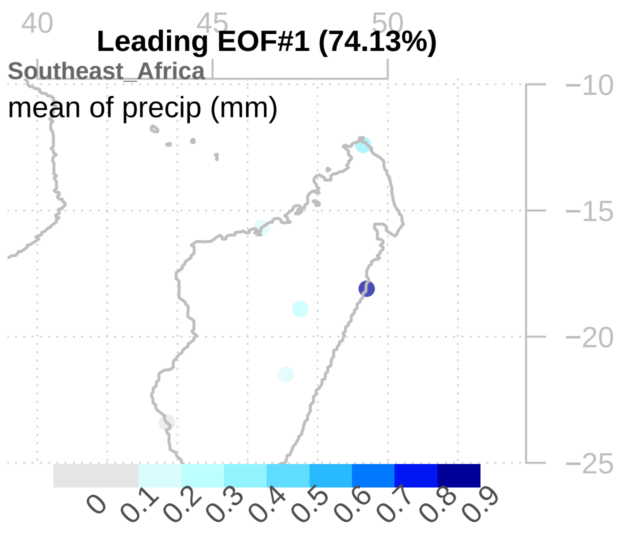

```
plot(pca.les,new=FALSE)
```

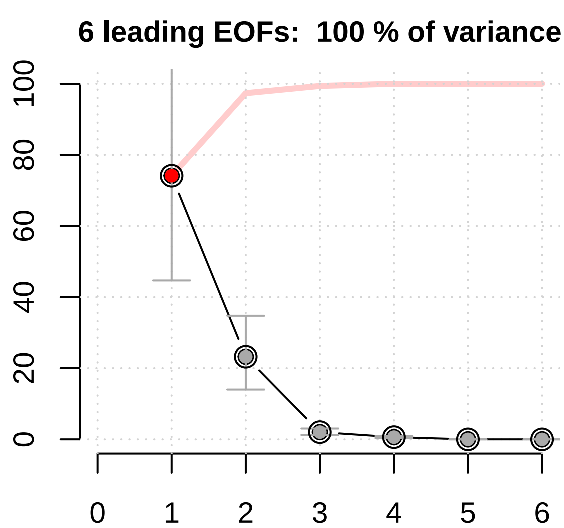


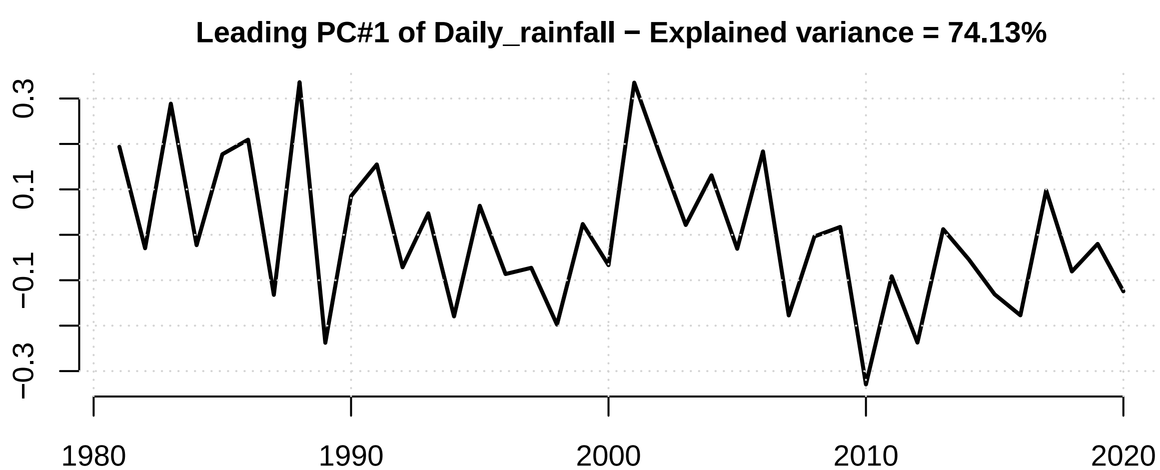





```
plot(pca.mad,new=FALSE)
```



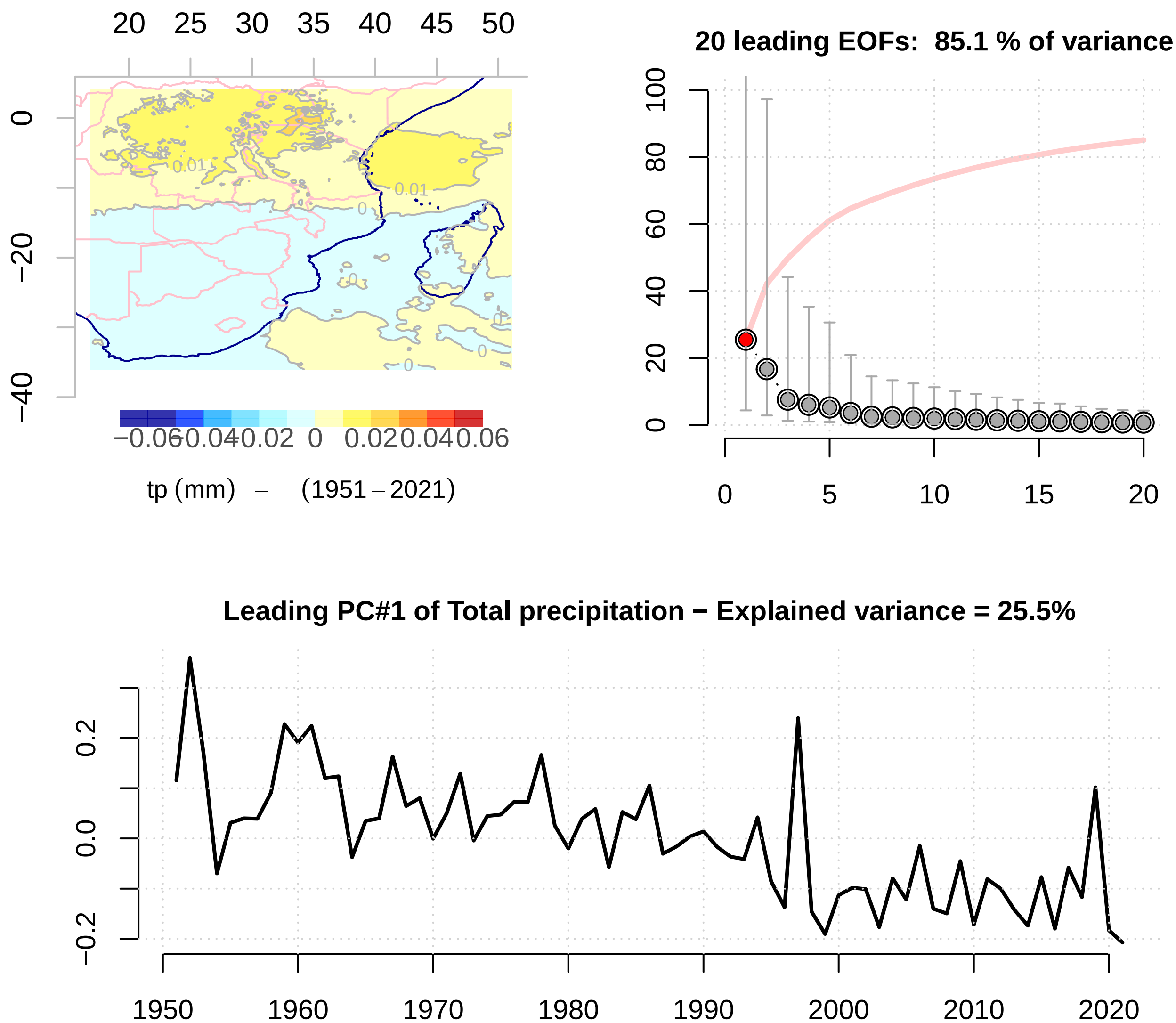




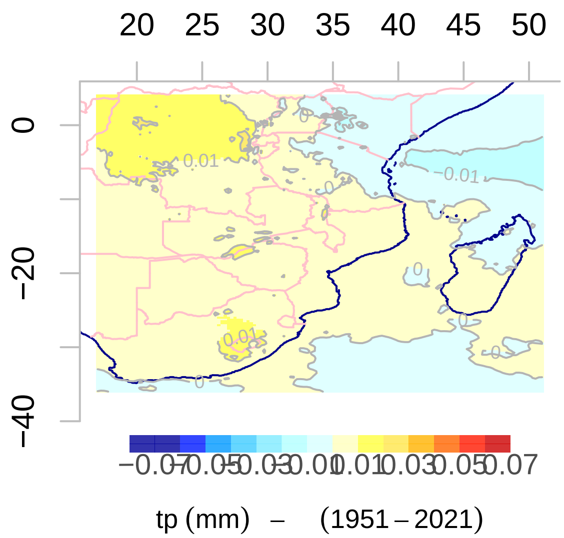
2.2 ERA5 used to represent large-scale predictors

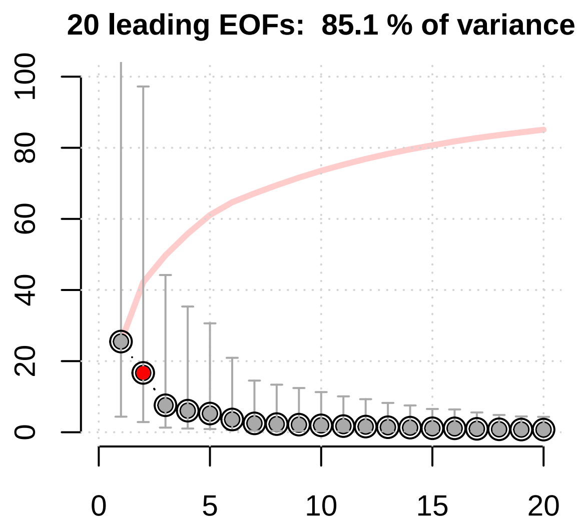
The predictor represents large-scale conditions used to estimate effects on smaller scales - the predictand. Empirical orthogonal functions (EOFs) can provide saline information about the large scales in an efficient way (small data volumes) and with favourable mathematical properties (the modes are orthogonal). They are related to the well-established mathematical concept known as eigenfunctions and capture the spatio-temporal covariance structure in the data.

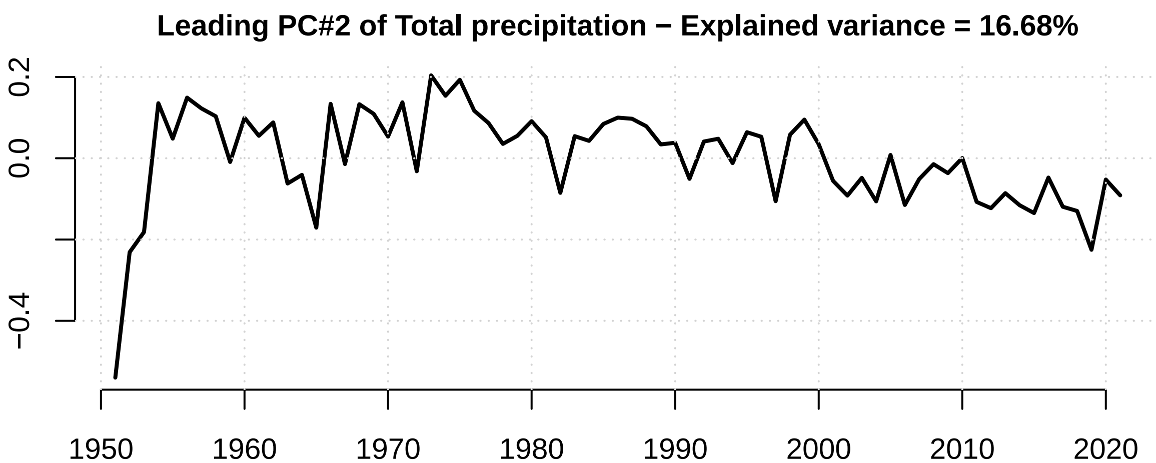
```
## Extract geographical region of interest to reduce data volume
#ERA5 <- subset(era5,is=list(lon=c(15,45),lat=c(-35,0)))
## Aggregate the annual mean rainfall (Oct-Sep)
ERA5 <- annual(era5,FUN='mean',start=year.start)
## Remove the start and end years because they don't include all months
ERA5 <- subset(ERA5,it=range(index(ERA5)) + c(1,-1))
## Calculate the EOFs of the annual rainfall
eof <- EOF(ERA5)
index(eof) <- year(eof)
plot(eof,new=FALSE)
```




```r
plot(eof,ip=2,new=FALSE)
```







The eigenvalues indicate that there are no recurring pattern of annual rainfall that dominates, but there is a diversity that is caught in a number of different modes. The 20 leading EOFs capture 87% of the variance and the leading mode about 26%. The leading mode is associated with a long-term trend and spikes in 1997 and 2019.

2.2.1 Downscaling-based evaluation

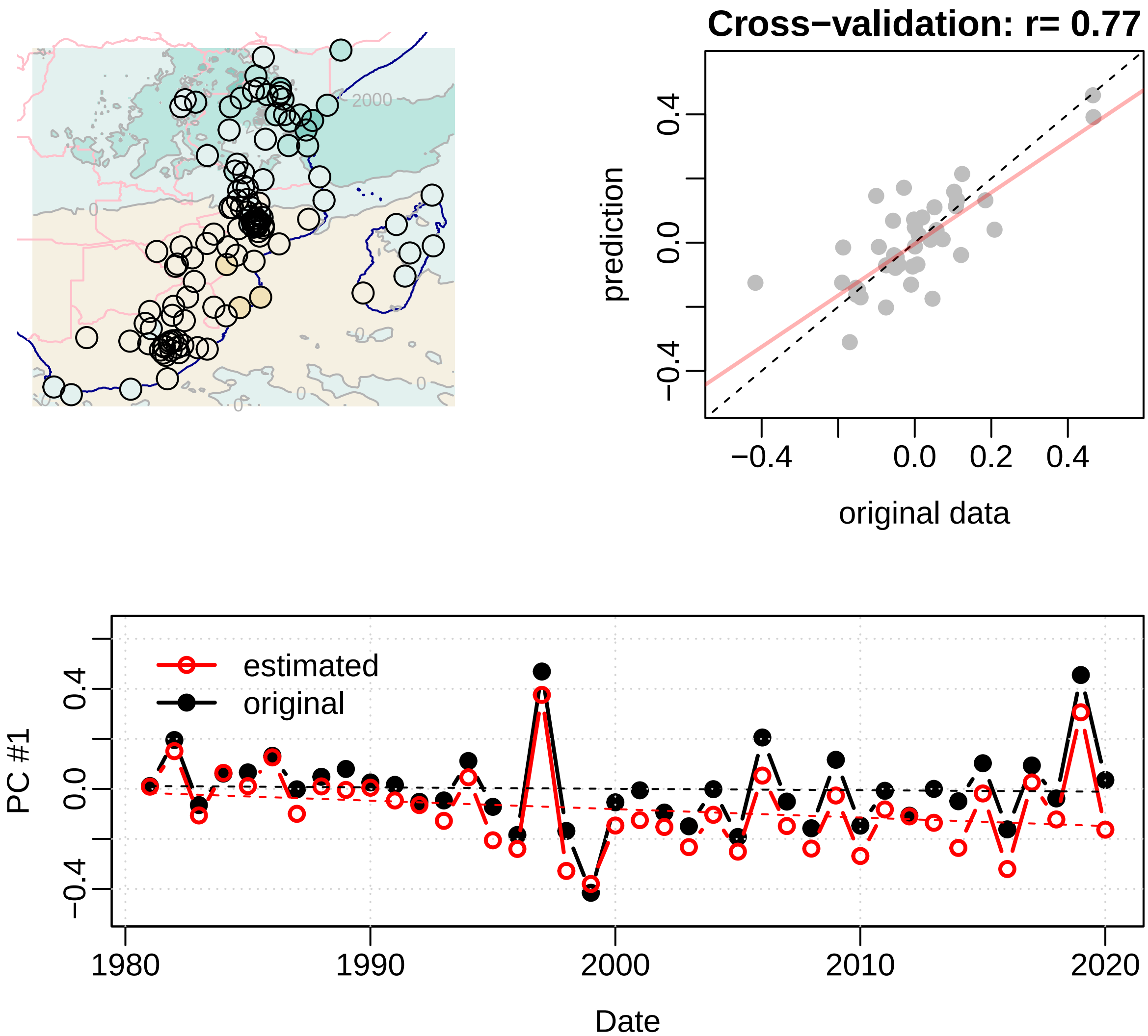
The following chunk uses regression-based ESD downscaling for comparing annually aggregated rain gauge data with corresponding data from ERA5: $\hat{y} = f(X)$ where the properties of $f(.)$ is used for the evaluation. The benefit of using PCA in ESD is that the emphasis is on all the stations as a group, rather than stations on an individual basis.

2.2.1.1 All participanting countries Diagnostics of the calibration of ERA5-based downscaling model for the annual rainfall totals (October-September).

```
## pca contains annual rainfall totals from rain gauges whereas eof contains fields of annual rainfall :
ds <- DS(pca,eof,ip=1:20)
```

```
##   |                                                                            |
```

```r
## Show the results for the leading PCA
plot(ds,new=FALSE)
```




```
## NULL
```

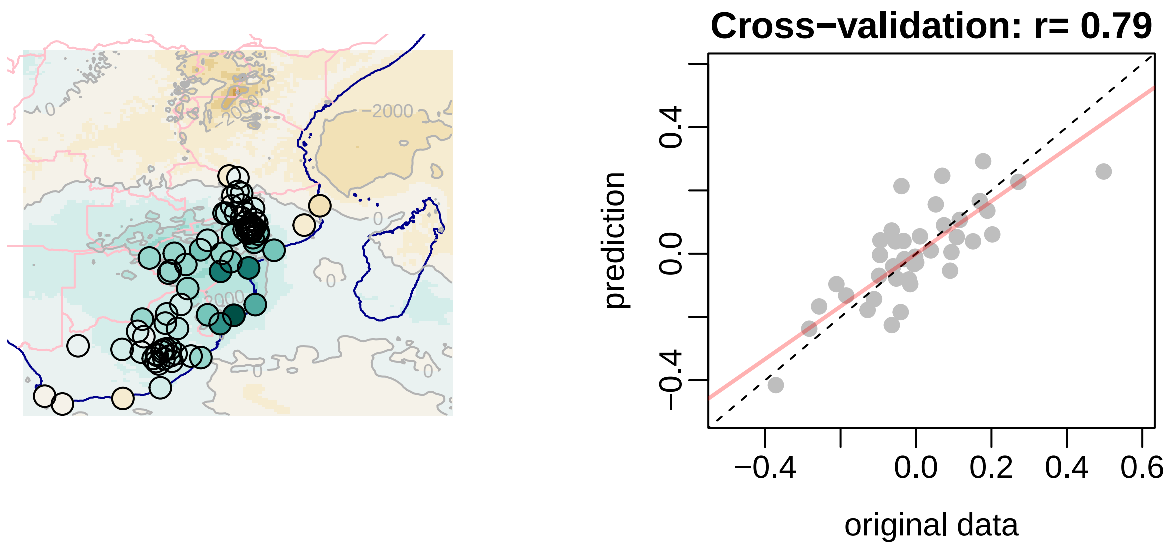
The cross-validation correlation for the leading PCA for all sites included here was 0.72.

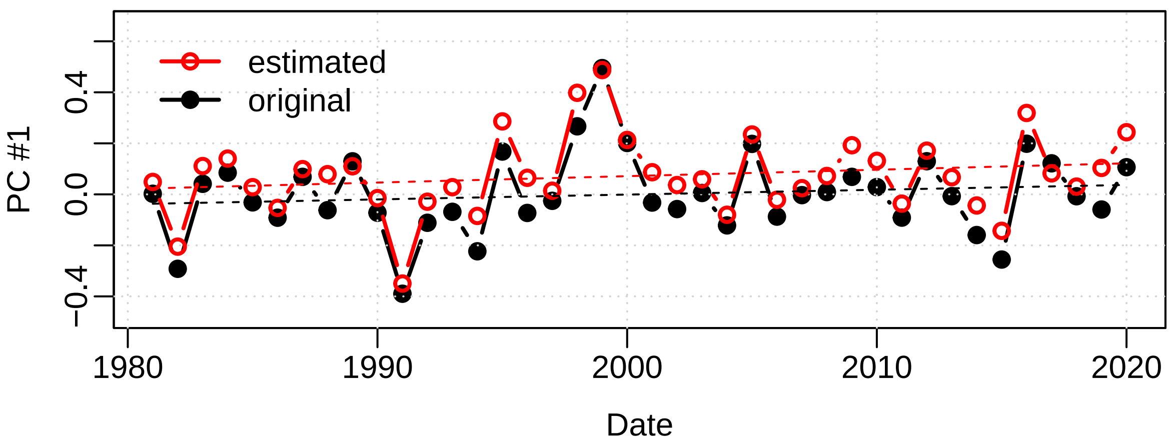
```r
## pca contains annual rainfall totals from rain gauges whereas eof contains fields of annual rainfall
ds.MMZAL <- DS(pca.MMZAL,eof,ip=1:20)
```

2.2.1.2 Mozambique, Malawi, Zimbabwe and South-Africa

```
##   |                                                                                |
```

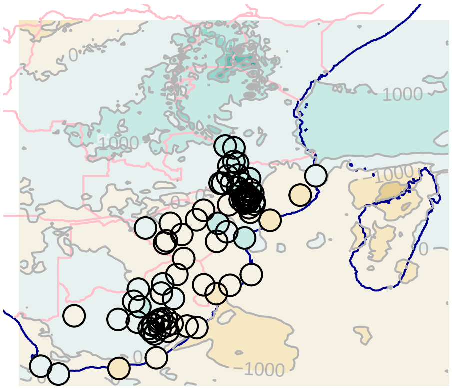
```r
## Show the results for the leading PCA
plot(ds.MMZAL,new=FALSE)
```

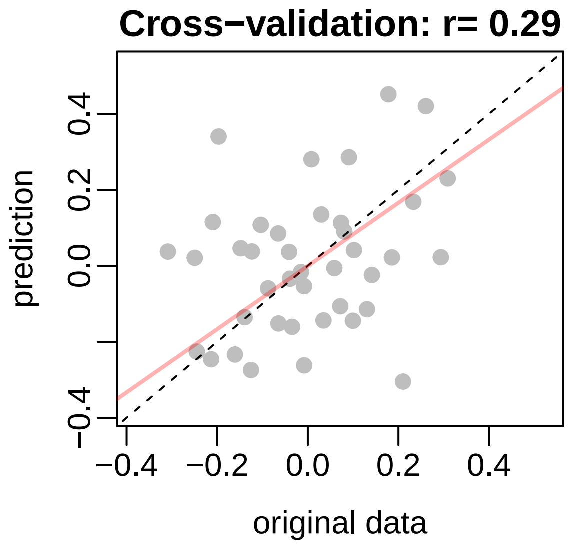


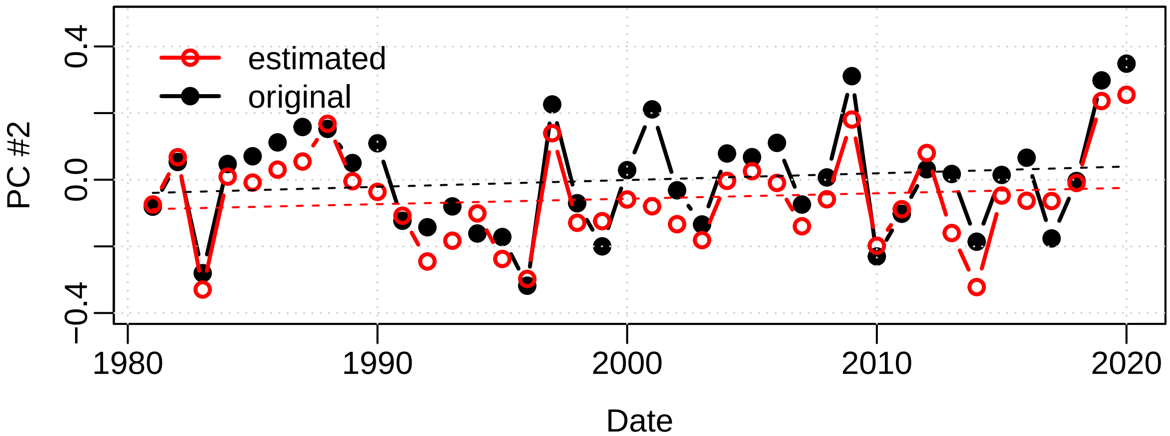



```
## NULL
```

```
## Show the results for the second PCA
plot(ds.MMZAL,ip=2,new=FALSE)
```

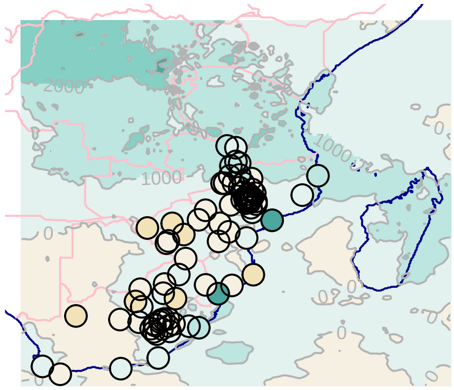


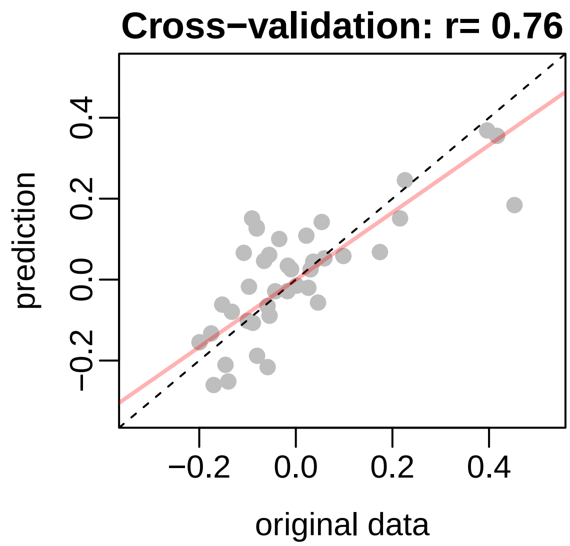


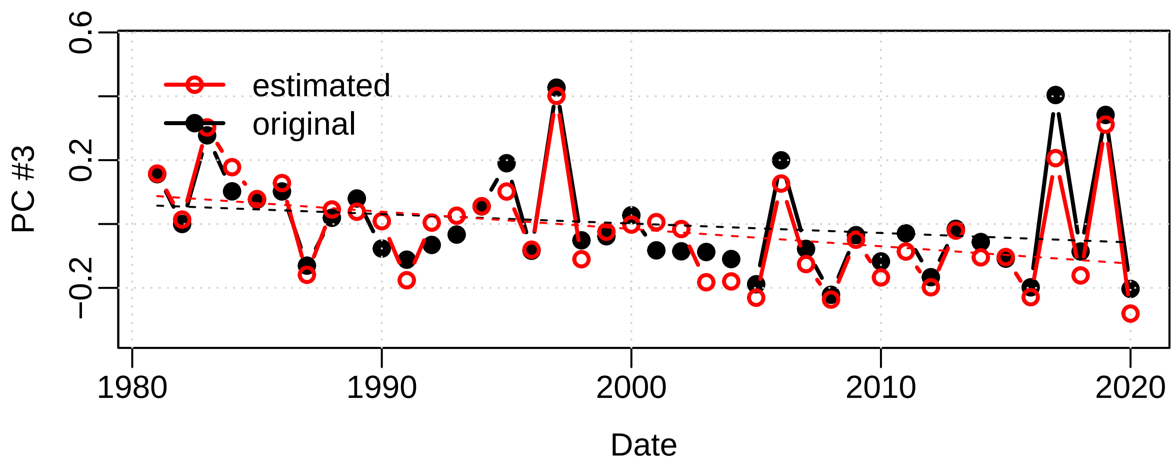



```
## NULL
```

```r
## Show the results for the third PCA
plot(ds.MMZAL,ip=3,new=FALSE)
```

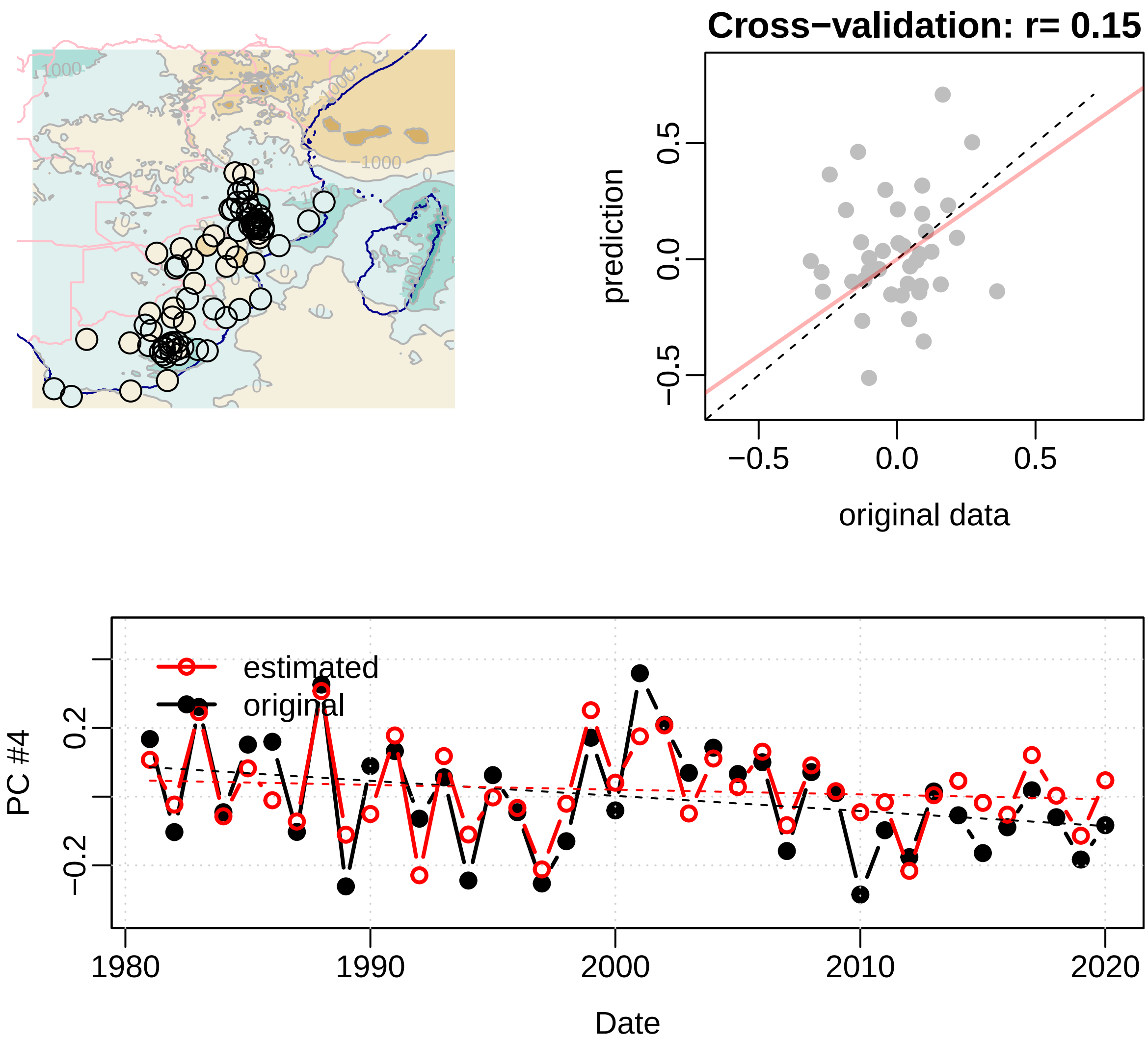







```
## NULL
```

```
## Show the results for the fourth PCA
plot(ds.MMZAL,ip=4,new=FALSE)
```




```
## NULL
```

The cross-validation for PCA mode 1 indicated high scores (correlation of 0.75) with greatest weights over southern Mozambique which corresponded with heavy weights in a similar regions for ERA5.

The observed and reproduced trends matched well (de-trended data were used for calibration but the trends in ERA5 were included in the reproduction). These results are encouraging for using the rain gauge data in ESD and making projections for the future.

The cross-validation results for PCA mode 2 was low (0.45), but not too bad for mode 3, and worse (negative cross-validation scores) for fourth mode. These modes are probably degenerate, represent more geographical differences, and do not individually account for a large fraction of the variance (together, they account for 41.3%).

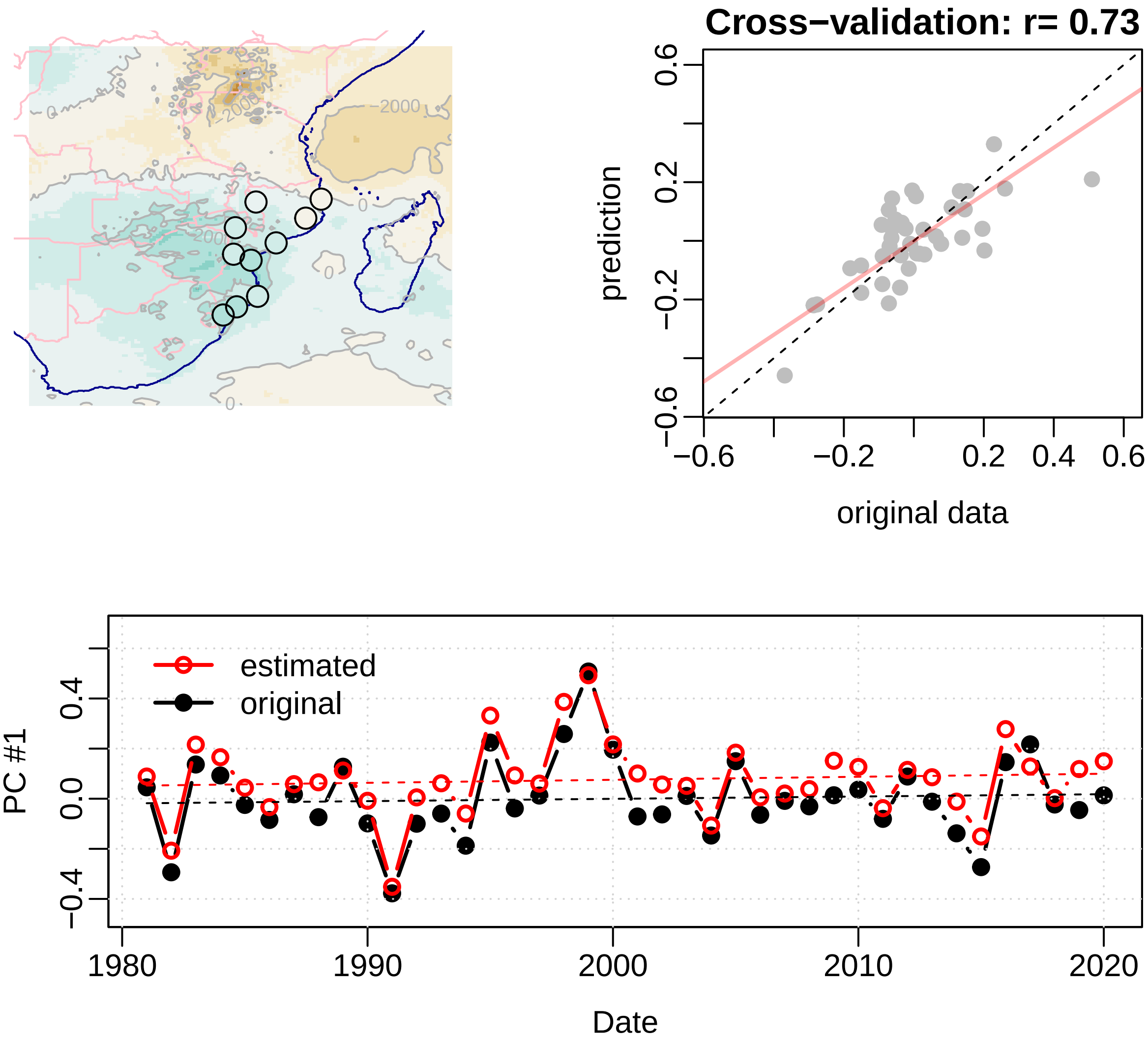
Nevertheless, the evaluation suggests that there is a generally good match between the observation-based annual rainfall from rain gauge data and ERA5, but that there also are stations which do not match ERA5 so well.

```
## pca contains annual rainfall totals from rain gauges whereas eof contains fields of annual rainfall
ds.moz <- DS(pca.moz,eof,ip=1:20)
```

2.2.1.3 Mozambique

```
##    |                                                                        |
```

```
## Show the results for the leading PCA
plot(ds.moz,new=FALSE)
```




```
## NULL
```

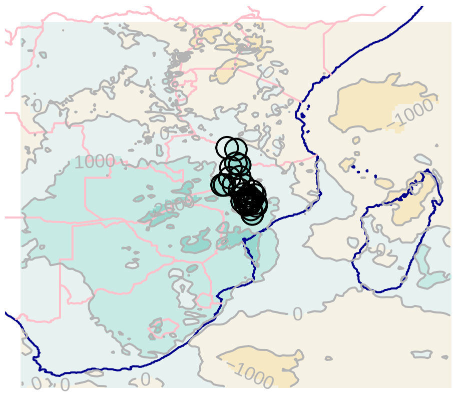
The cross-validation correlation for the leading PCA for Mozambique rain gauges was 0.72.

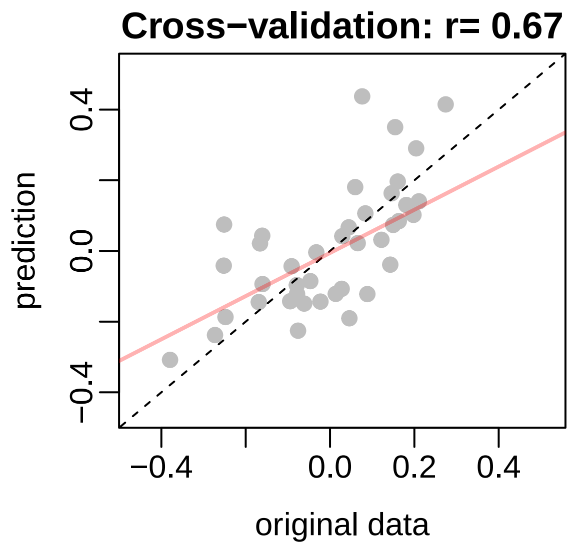
```
## pca contains annual rainfall totals from rain gauges whereas eof contains fields of annual rainfall
ds.mal <- DS(pca.mal,eof,ip=1:20)
```

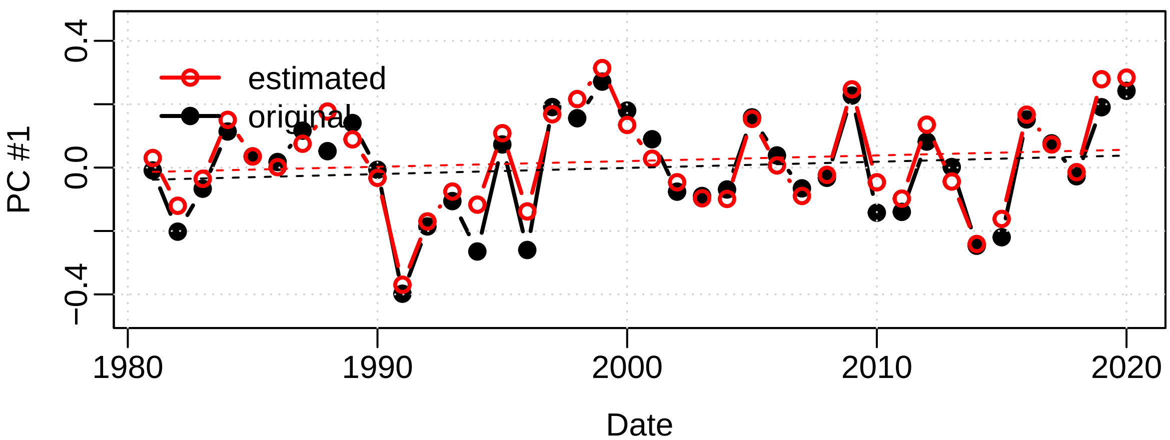
2.2.1.4 Malawi

```
##    |                                                                        |
```

```
## Show the results for the leading PCA
plot(ds.mal,new=FALSE)
```








```
## NULL
```

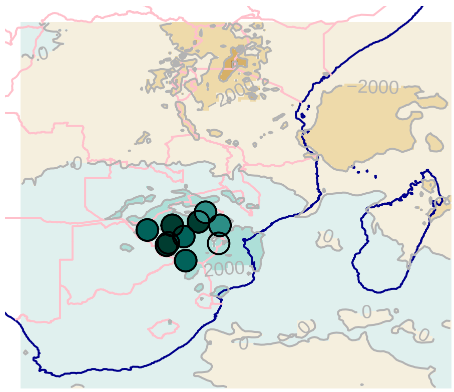
The cross-validation correlation for the leading PCA for Malawi rain gauges was 0.68.

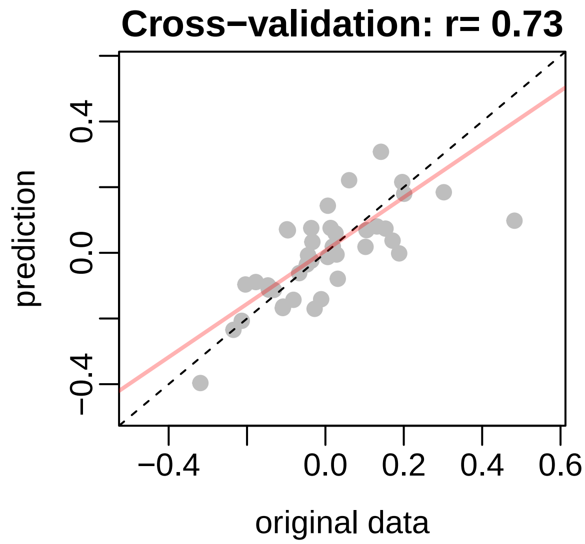
```
## pca contains annual rainfall totals from rain gauges whereas eof contains fields of annual rainfall 
ds.zim <- DS(pca.zim,eof,ip=1:20)
```

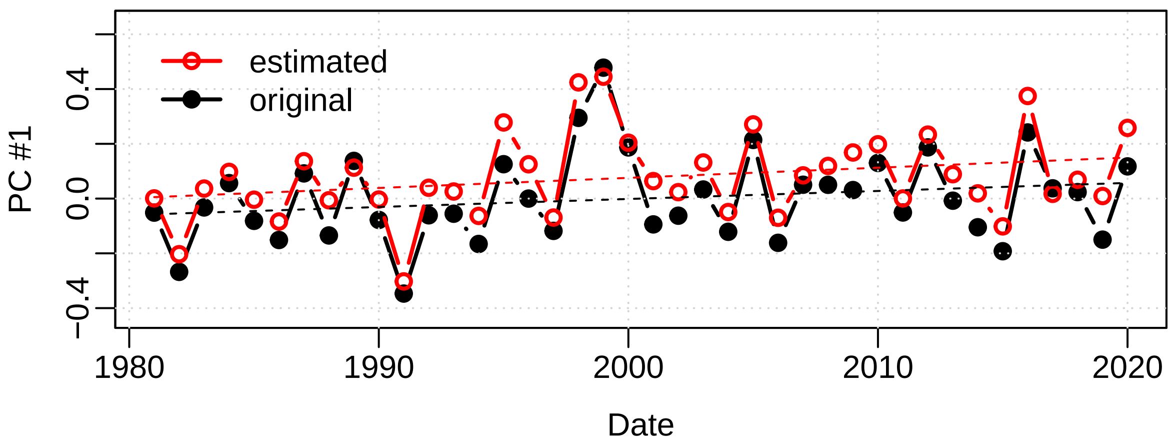
2.2.1.5 Zimbabwe

```
##   |                                                                                                    |
```

```
## Show the results for the leading PCA
plot(ds.zim,new=FALSE)
```








```
## NULL
```

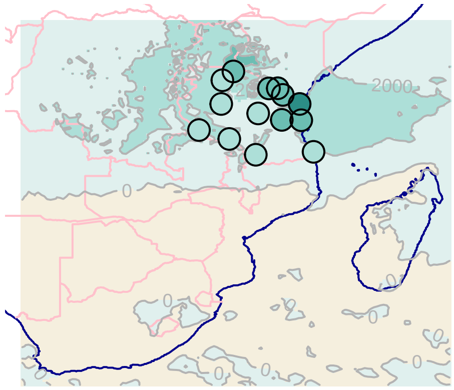
The cross-validation correlation for the leading PCA for Zimbabwe rain gauges was 0.70.

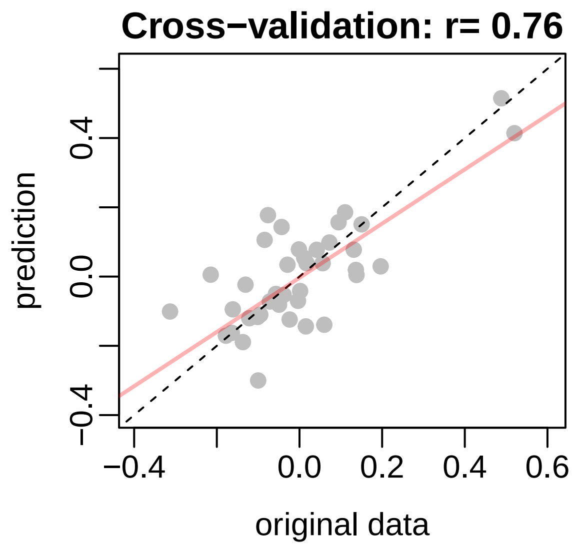
```
## pca contains annual rainfall totals from rain gauges whereas eof contains fields of annual rainfall 
ds.tan <- DS(pca.tan,eof,ip=1:20)
```

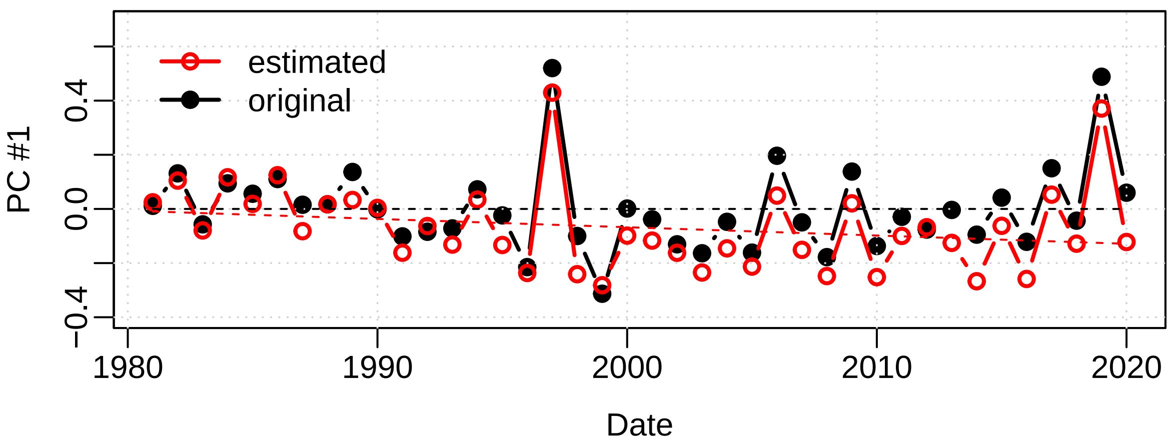
2.2.1.6 Tanzania

```
##   |                                                                                |
```

```
## Show the results for the leading PCA
plot(ds.tan,new=FALSE)
```








```
## NULL
```

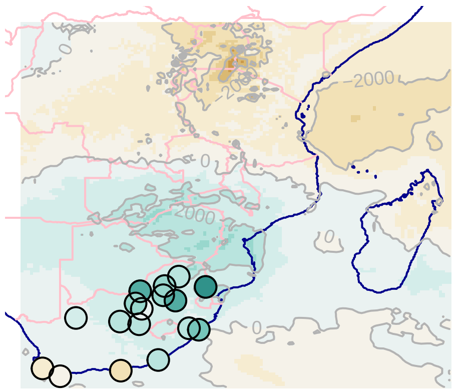
The cross-validation correlation for the leading PCA for Tanzania rain gauges was 0.76.The spatial predictor pattern had a different character as those cases shown above. Also, the leading pattern was dominated by two spikes, but the analysis only involved three locations, which may have affected the results. Nevertheless, the cross-validation correlation was fairly high.

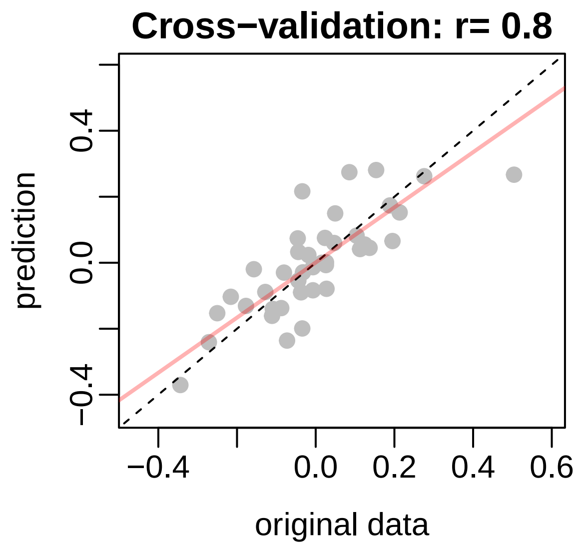
```
## pca contains annual rainfall totals from rain gauges whereas eof contains
## fields of annual rainfall from ERA5.
ds.saf <- DS(pca.saf,eof,ip=1:20)
```

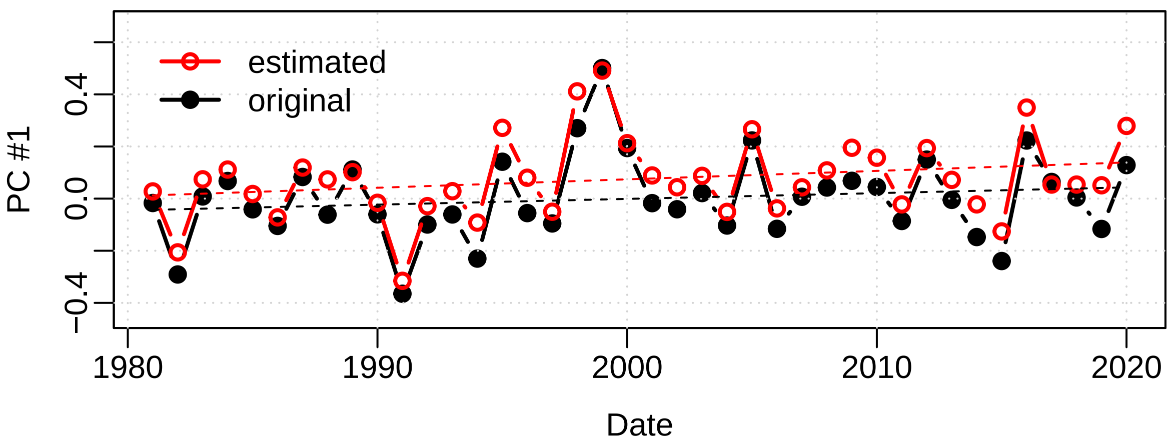
2.2.1.7 South Africa

```
##   |                                                                        |
```

```
## Show the results for the leading PCA
plot(ds.saf,new=FALSE)
```








```
## NULL
```

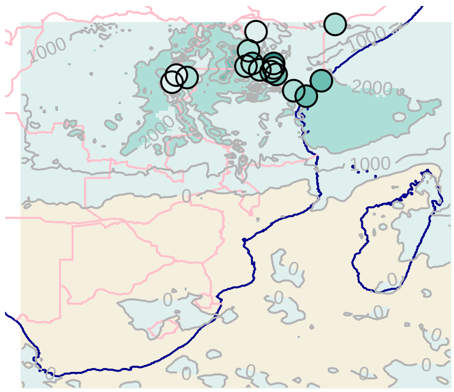
The cross-validation correlation for the leading PCA for South Afdrica rain gauges was 0.76.

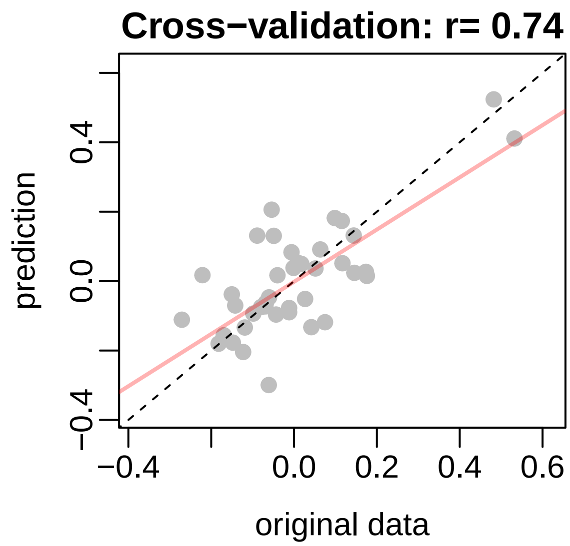
```
## pca contains annual rainfall totals from rain gauges whereas eof contains
## fields of annual rainfall from ERA5.
ds.KR <- DS(pca.KRB,eof,ip=1:20)
```

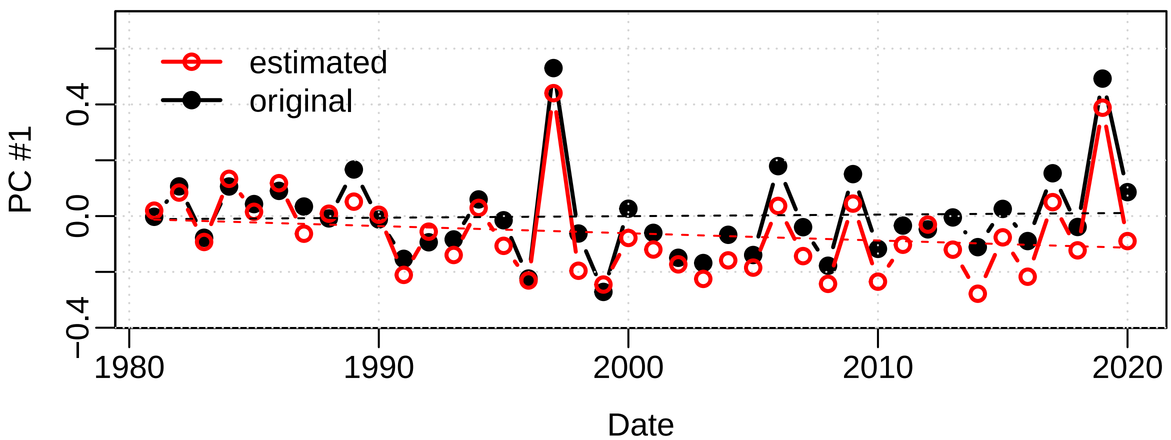
2.2.1.8 Kenya and Rwanda

```
##   |                                                                          |
```

```
## Show the results for the leading PCA
plot(ds.KR,new=FALSE)
```








```
## NULL
```

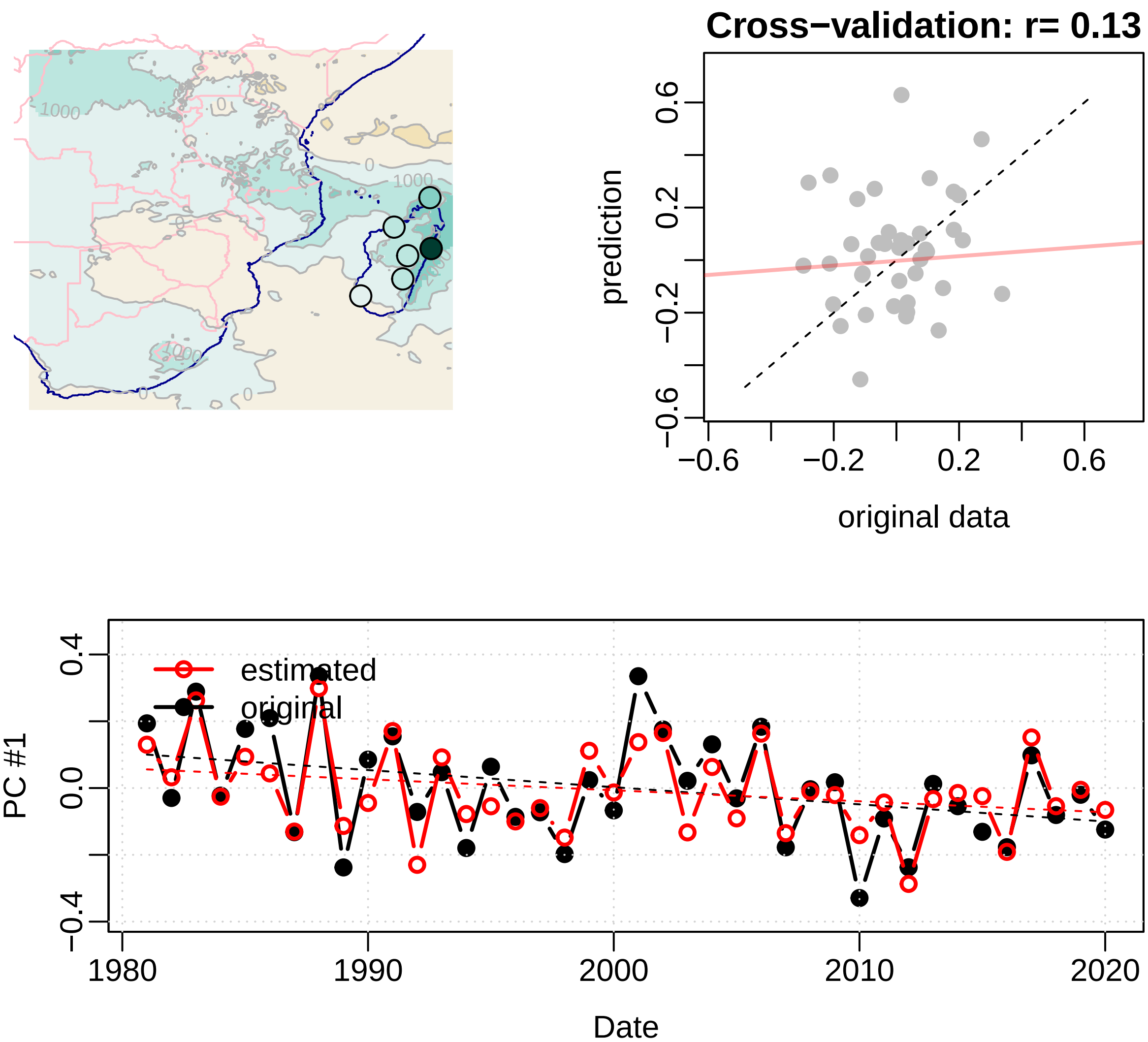
The cross-validation correlation for the leading PCA for Kenya and Rwanda rain gauges was 0.76. These results resembled those from Tanzania, derived from different locations but ones that have been subject to similar regional climate variations.

```
## pca contains annual rainfall totals from rain gauges whereas eof contains
## fields of annual rainfall from ERA5.
ds.mad <- DS(pca.mad,eof,ip=1:20)
```

2.2.1.9 Madagascar

```
##   |                                                                      |
```

```
## Show the results for the leading PCA
plot(ds.mad,new=FALSE)
```

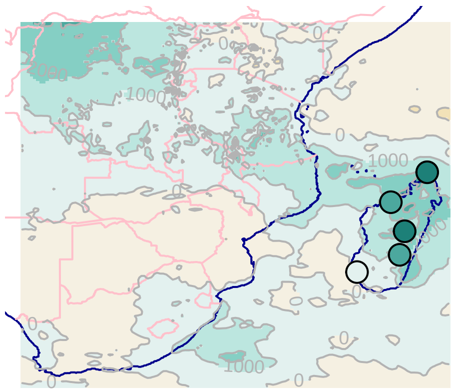


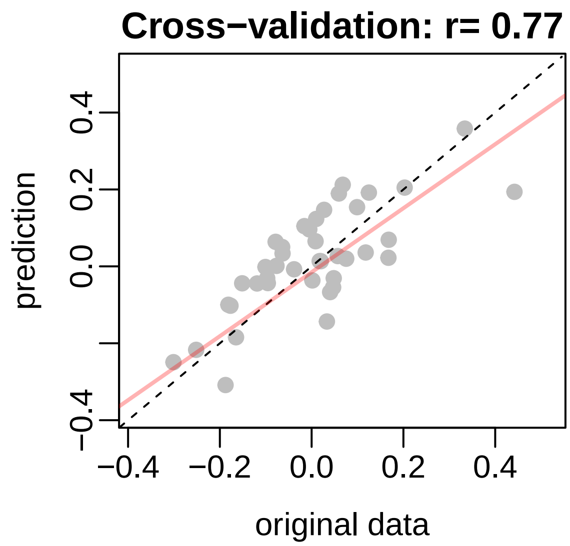

```
## NULL
```

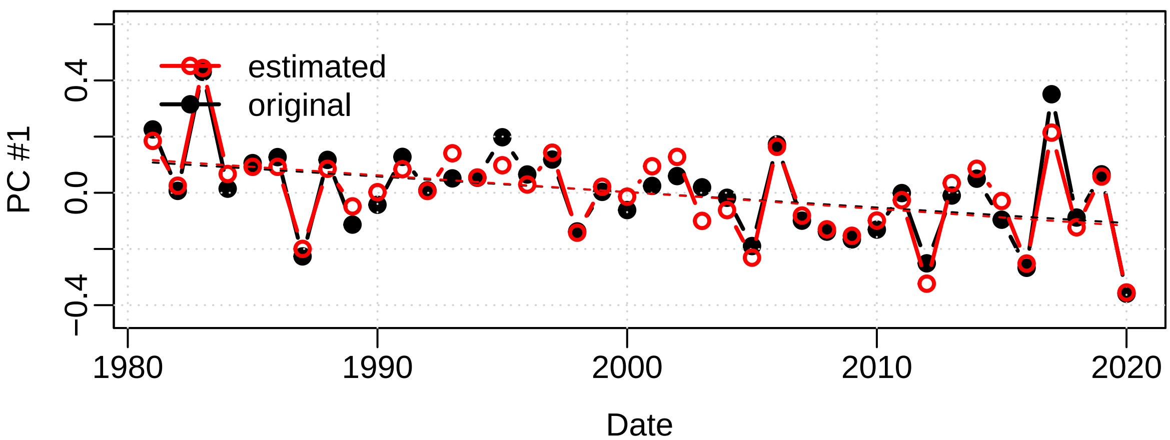
```r
ds.mad2 <- DS(pca.mad2,eof,ip=1:20)
```

```
##   |                                                                          |
```

```r
## Show the results for the leading PCA
plot(ds.mad2,new=FALSE)
```








```
## NULL
```

The cross-validation correlation for the leading PCA for Madagascar rain gauges was 0.34. This score was strongly influenced by Toamasina with much higher rainfall and a different climatology to the other sites in Madagaskar. Excluding Toamasina results in an improved agreement with ERA5, but the rain gauge data for Madagascar are still less comparable to ERA5 than those on the southeast African continent.

2.2.2 Save the results for total precipitation

We save the end results as R-binary:

```
attr(X,'longname')[] <- "Daily_rainfall"
unique <- !duplicated(paste(lon(X),lat(X)))
X <- subset(X,is=unique)
atX <- subset(atX,is=unique)
X <- subset(X,is=unique)
save(atX,X,file='readAfricanData.rda')
```

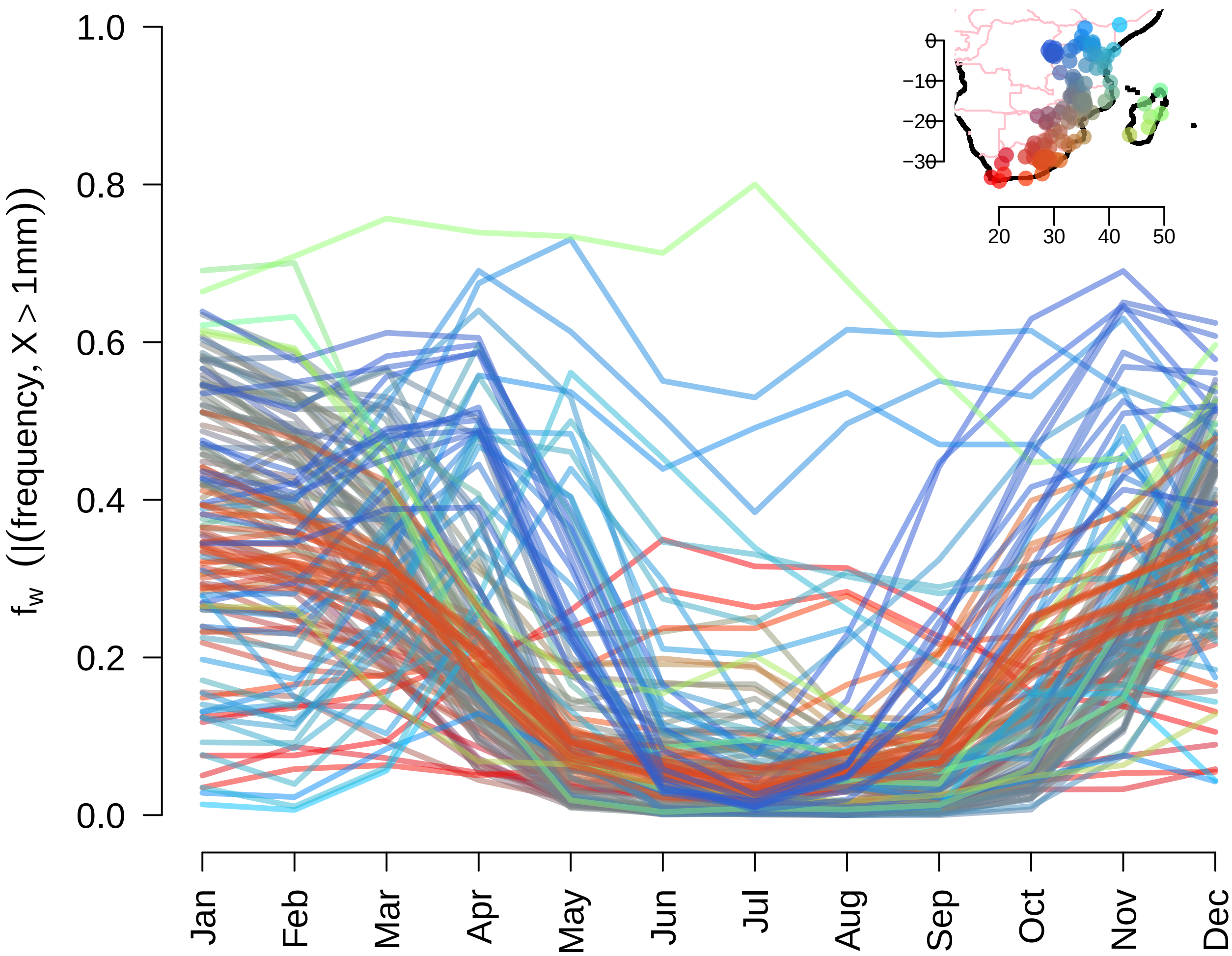
3 Evaluation of more advanced rainfall statistics

The remaining part of this analysis follows Benestad et al. (2024).

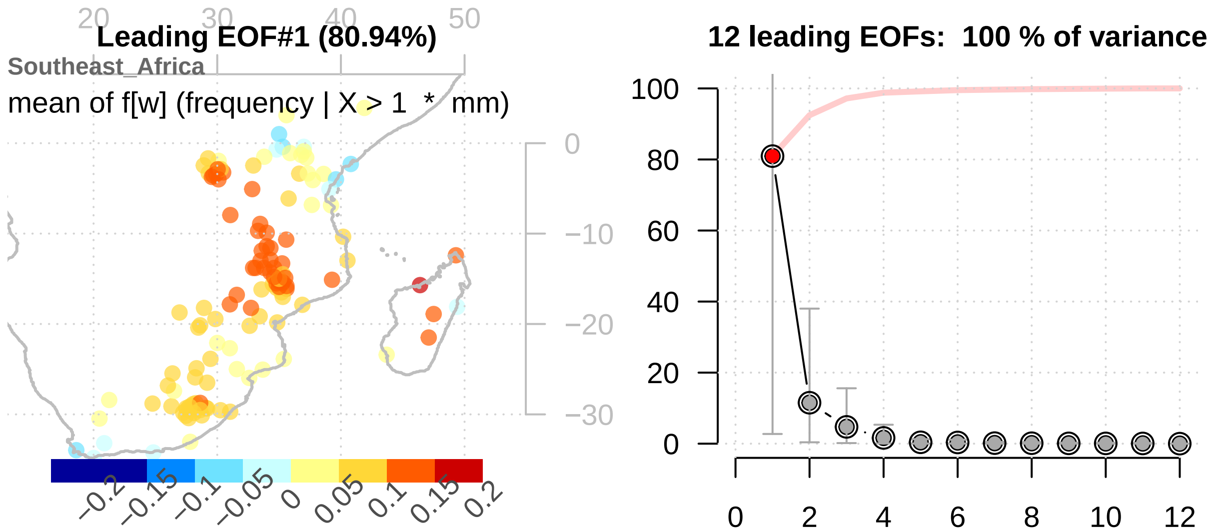
3.1 Wet-day frequency f_w

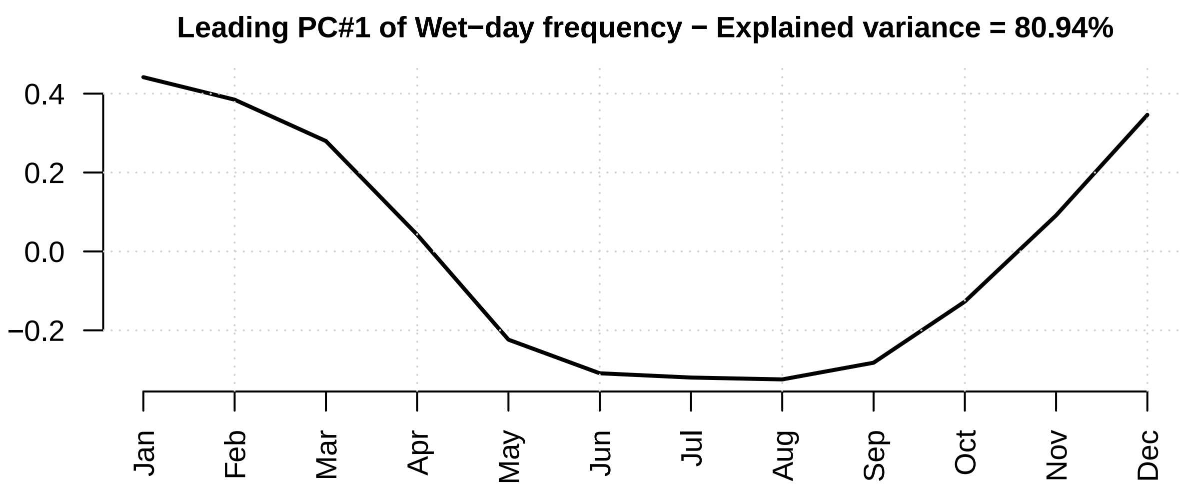
Since the mean the mean rainfall is the product of the frequency of rainy days and the mean rainfall intensity according to $\overline{x} = f_w \mu$, it's of interest to study variations in these respective factors.

```
fw.ac <- aggregate(X,by=month,FUN='wetfreq')
plot(fw.ac,new=FALSE)
```




```
## The information about the mean seasonal cycle in wet-day frequency can be summarised
##  through a PCA analysis
pca.fw.ac <- PCA(fw.ac,n=12)
plot(pca.fw.ac,new=FALSE)
```






```
fw <- annual(subset(X,it=c(1979,2020)),FUN='wetfreq',start=year.start)
nv <- apply(fw,2,'nv')
print(loc(X)[nv <= 33])
```

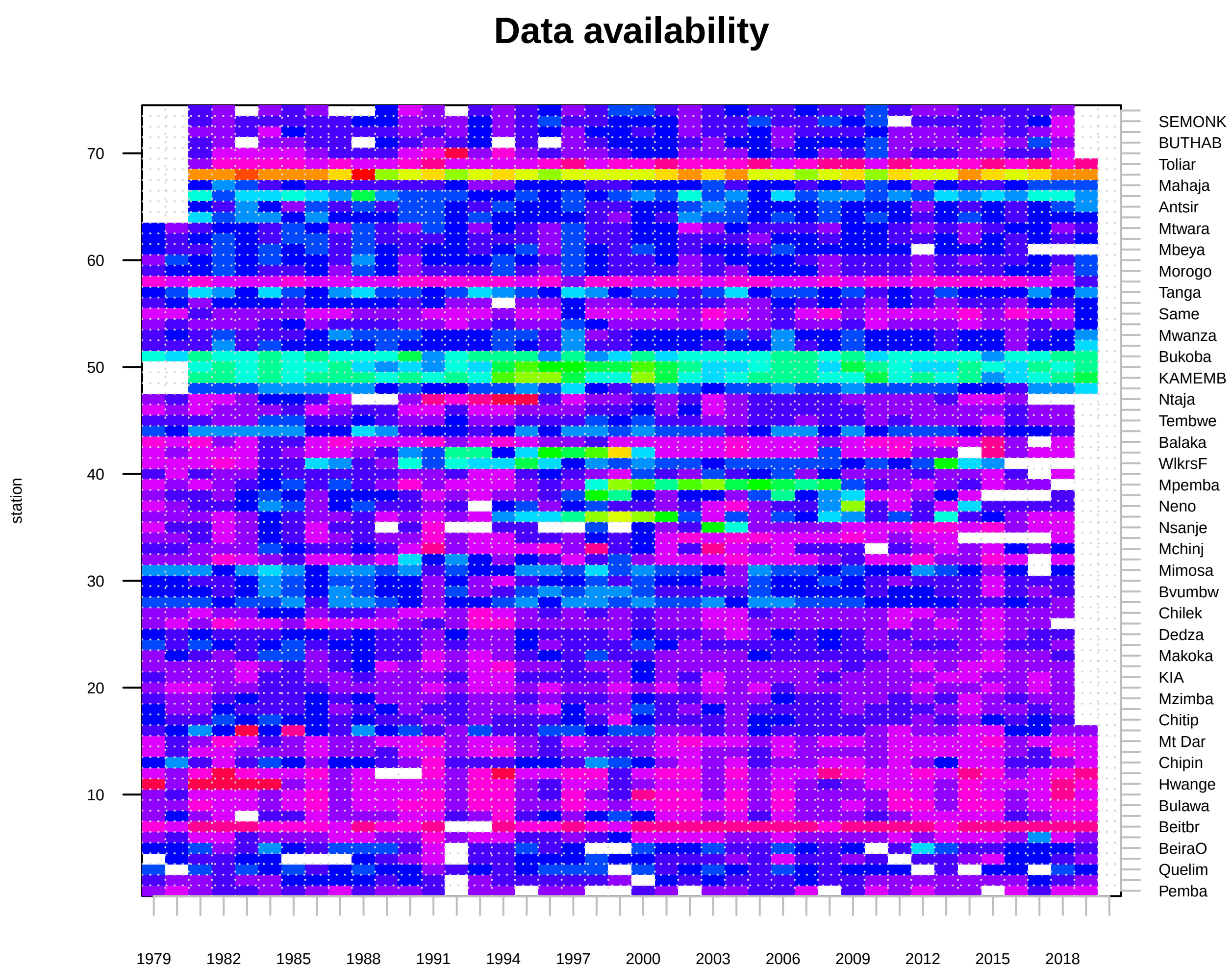
```
##  [1] "Lichinga"                "Tete"
##  [3] "Inhambane"               "Xai-Xai"
##  [5] "Bloemfontein WO"         "Bothaville - Balkfontein"
##  [7] "Brandvlei"               "Cape Agulhas"
##  [9] "Cape St. Francis"        "Cape Town WO"
## [11] "Cedara"                  "East London WO"
## [13] "Irene WO"                "Kimberley WO"
## [15] "Laingsburg"              "Marico"
## [17] "Mount Edgecombe"         "Ottosdal"
## [19] "Polokwane WO"            "Punda Maria"
## [21] "Secunda"                 "Skukuza"
## [23] "Upington WO"             "Warmbad Towoomba"
## [25] "Lodwar"                  "Mandera"
## [27] "Kitale"                  "Kericho"
## [29] "Kisii"                   "Narok"
```

```
## [31] "Nyeri"                       "Dagoretti Corner"
## [33] "Machakos Agromet"            "Voi"
## [35] "Lamu"                        "Moi International Airpor"
## [37] "NA"                          "Sumbawanga"
## [39] "QACHASNEK"                   "LERIBE"
## [41] "MAFETENG"                    "MALEFILOANE"
## [43] "MAPOTENG"                    "MEJAMETALANA"
## [45] "MOSHOESHOEI"                 "OXBOW"
## [47] "PHUTHIATSANA"                "QUTHING"
## [49] "BUJUMBURA (Aeroport)"        "GITEGA (Aerodrome)"
## [51] "CANKUZO"                     "GISOZI"
## [53] "MUYINGA"                     "MUSASA"
## [55] "NYAMUSWAGA"                  "MPOTA (Tora)"
```

```
fw <- subset(fw,is=(nv > 33))
nv <- apply(fw,2,'nv')
print(loc(X)[nv <= 33])
```

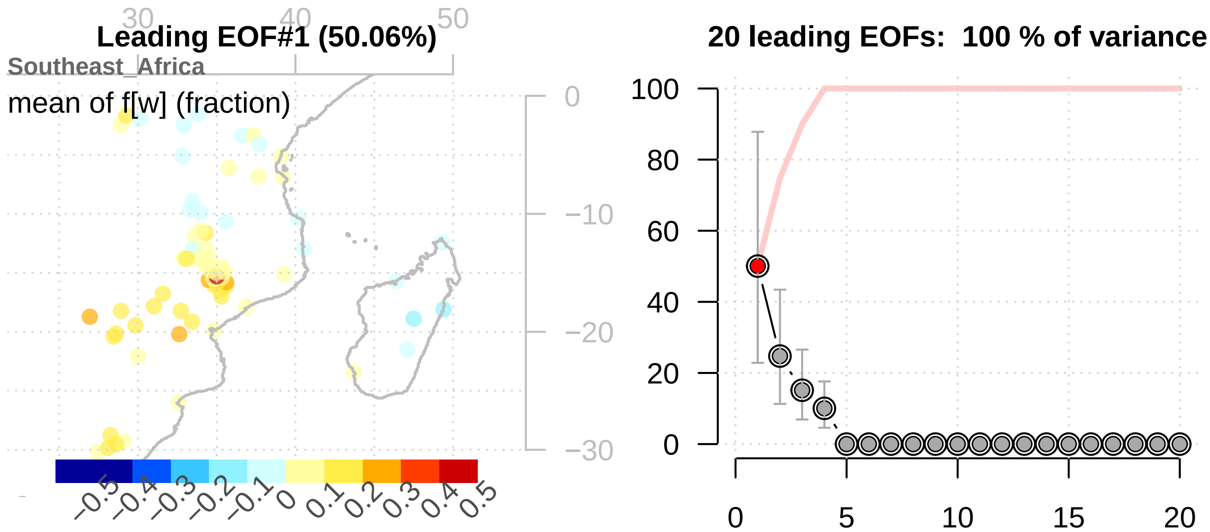
```
## character(0)
```

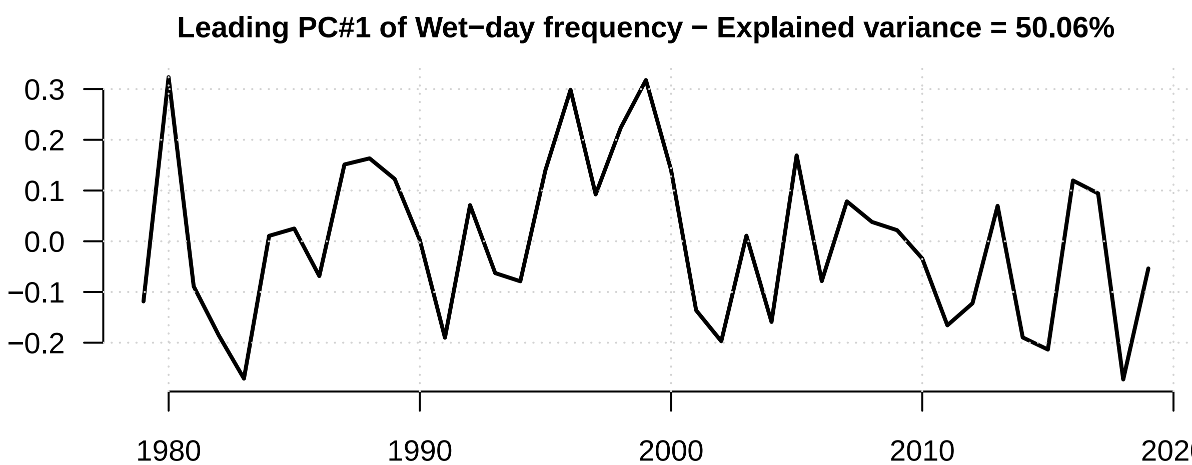
```
fw <- subset(fw,is=(nv > 33))
diagnose(fw)
```



Southeast_Africa

```r
pca.fw <- PCA(pcafill(fw))
plot(pca.fw,new=FALSE)
```





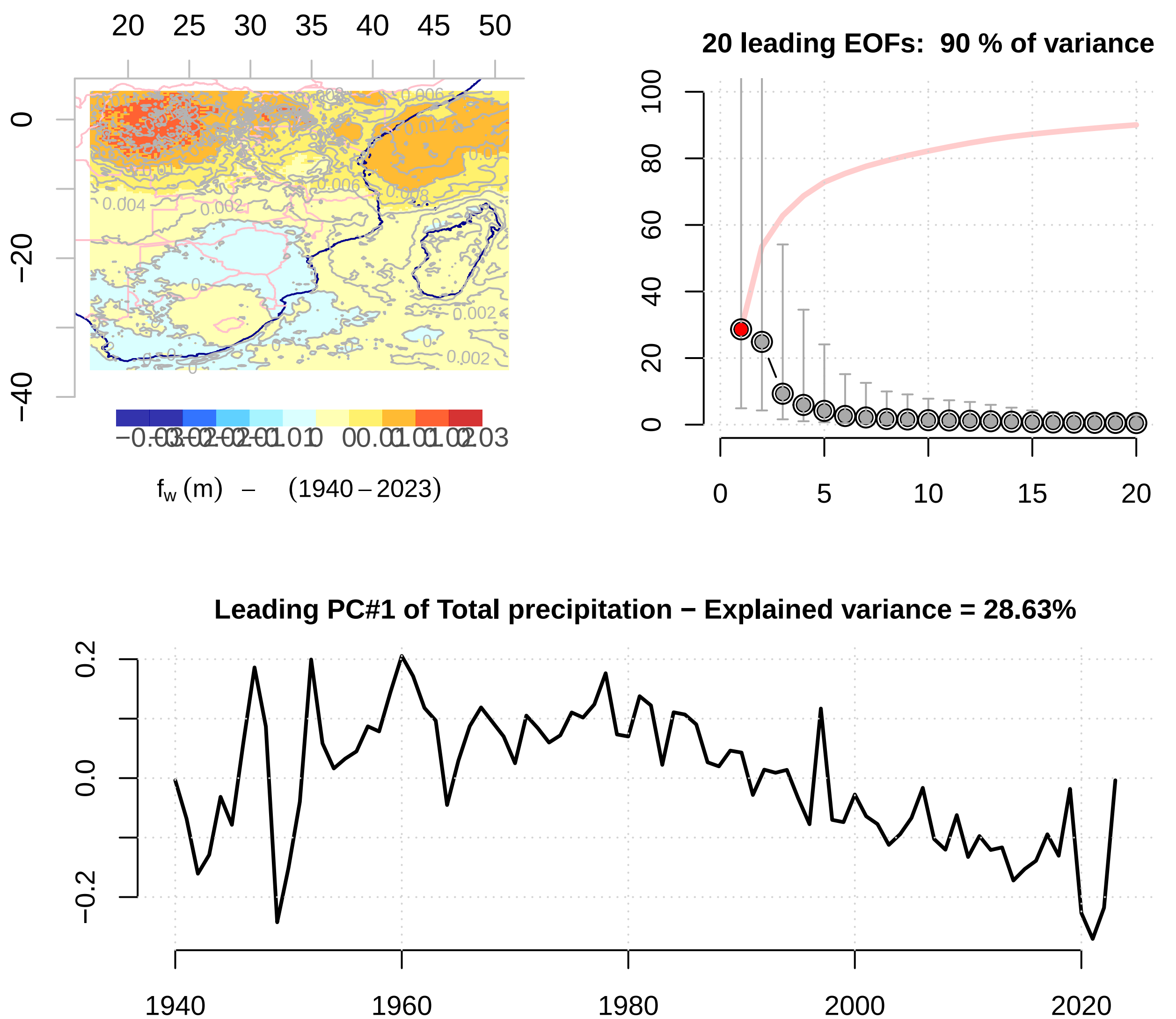
The wet-day frequency peaks around April-May closer to the equator and around December-February further south, except for south in South Africa where it is more steady through out the year. A PCA of the mean annual cycle in f_w indicates that one leading mode (82% of the variance) can capture much of the seasonal variation in the wet-day frequency, with strongest weights near 15°S that diminish towards lower and higher latitudes. PCA was also applied to October-September (yearly) wet-day frequencies, but due to missing data, some of the stations were excluded. The weights associated with the leading PCA for October-September f_w suggested that f_w on a yearly basis varies fairly coherently across the region covered by the rain gauge network.

3.1.1 Wet-day frequency f_w derived from ERA5

We examined the data through a common PCA and DS-based evaluation of f_w which involvedreading annual data from ERA5 and use the year October-September:

```
tmp.fw.file <- 'evalAfricaRainfall_fw.tmp.rda'
if (!file.exists(tmp.fw.file)) {
  FW <- retrieve('~/data/ERA5/ERA5_fw_mon.nc',lon=c(17,51),lat=c(-36,4))
  save(FW,file=tmp.fw.file)
} else load(tmp.fw.file)
attr(FW,'variable') <- 'f[w]'
eof.FW <- EOF(annual(FW,start=year.start))
```

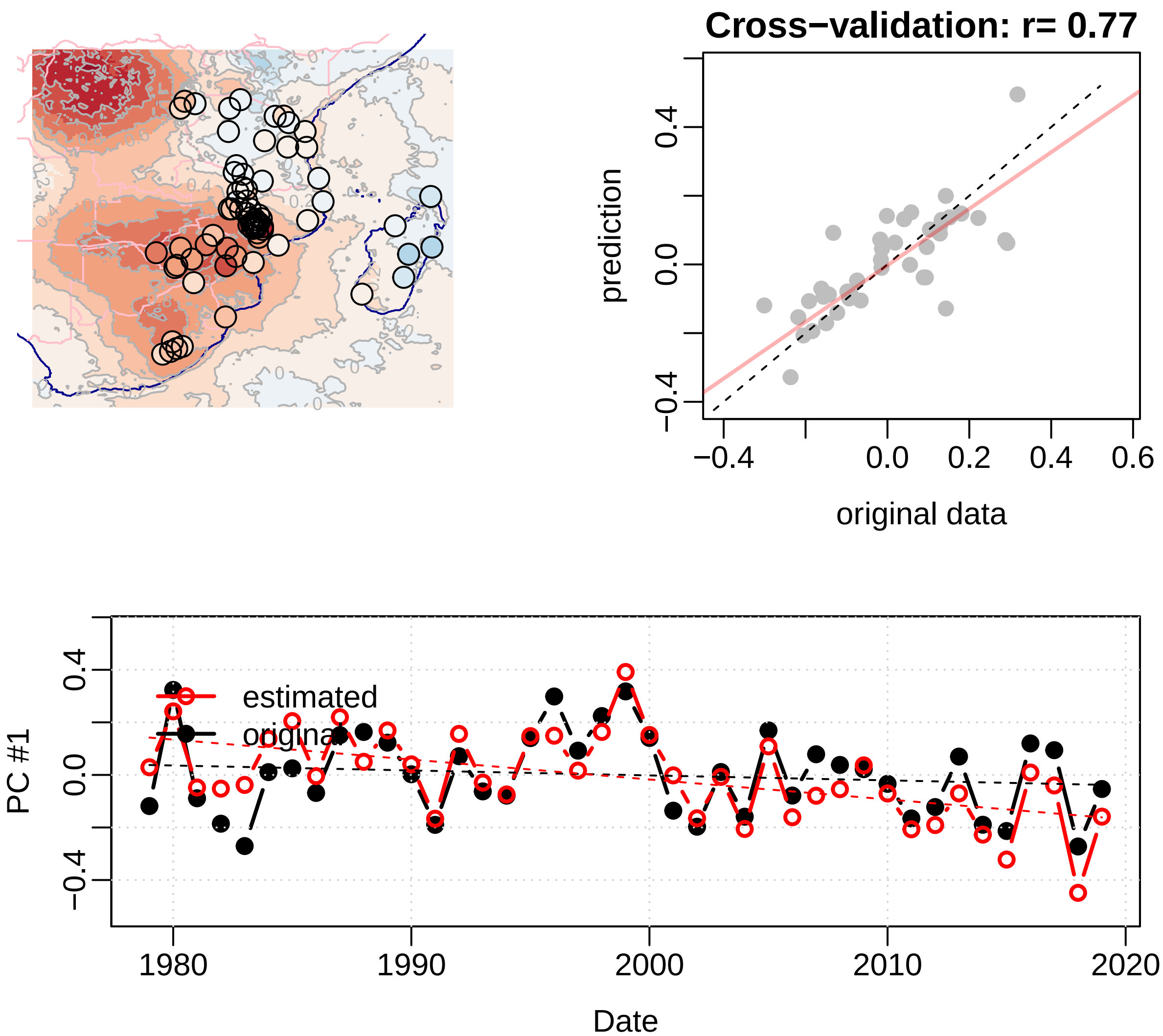
```r
plot(eof.FW,new=FALSE)
```




```r
ds.fw <- DS(pca.fw,eof.FW)
```

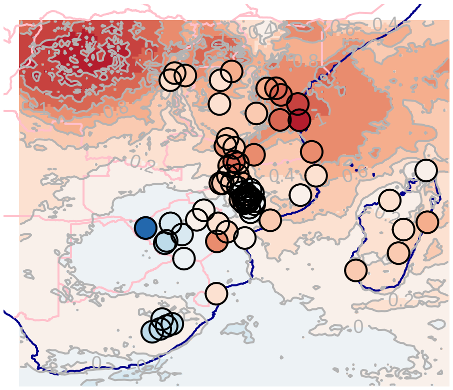
```
##    |                                                                      |
```

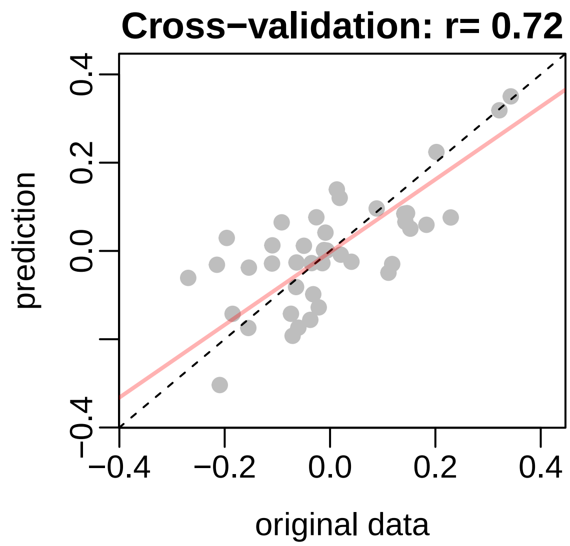
```r
plot(ds.fw,new=FALSE)
```

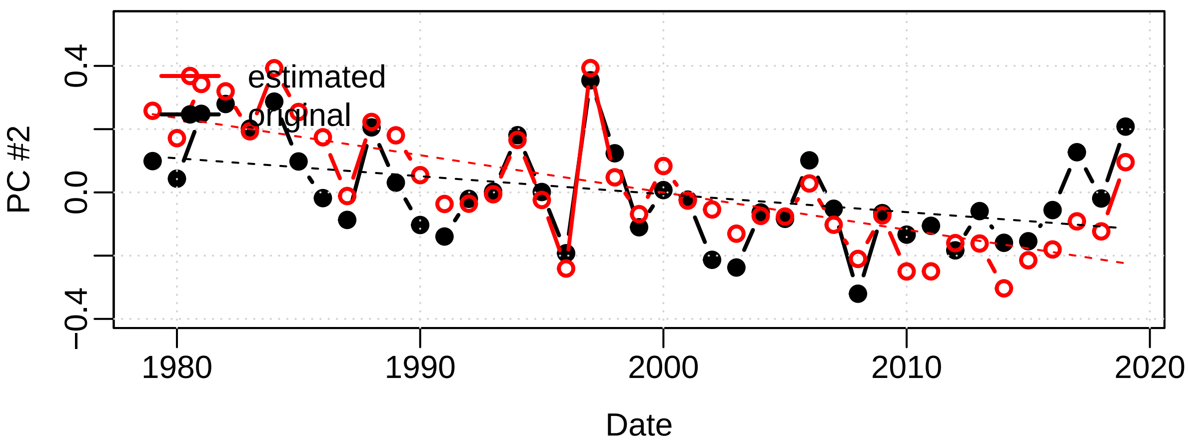



```
## NULL
```

```
plot(ds.fw,ip=2,new=FALSE)
```

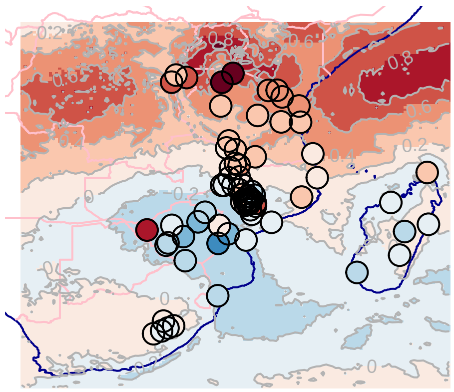


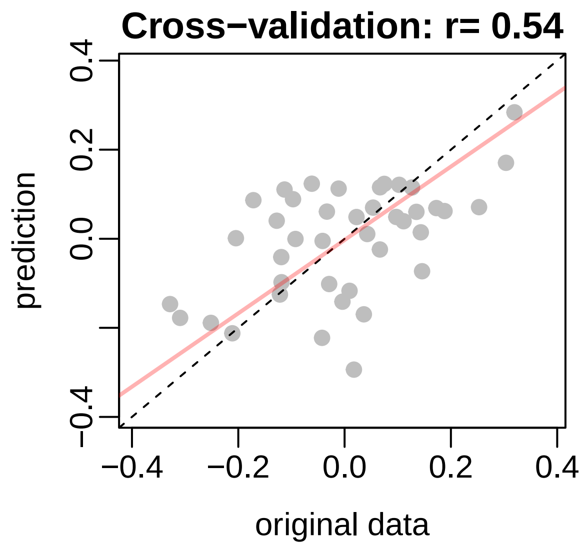


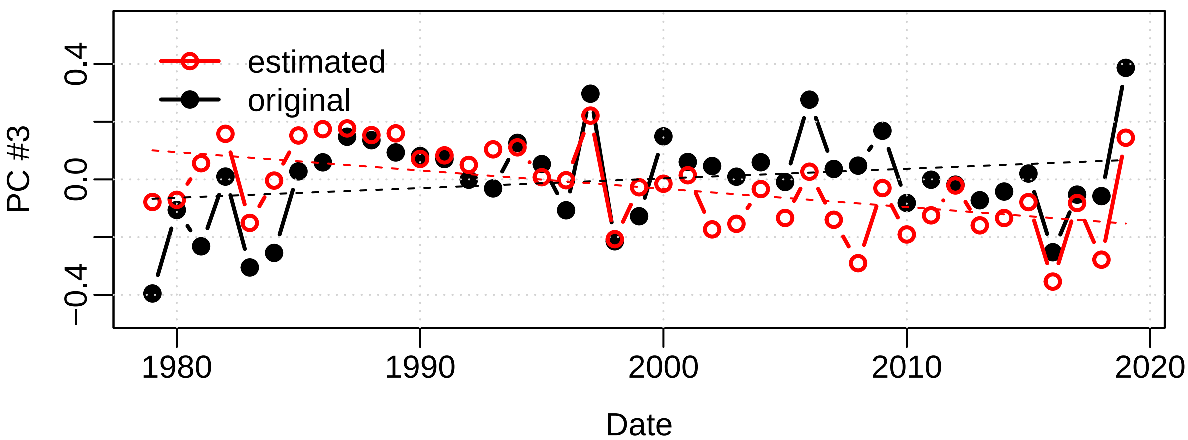



```
## NULL
```

```
plot(ds.fw,ip=3,new=FALSE)
```








```
## NULL
```

The evaluation on the annual (Oct-Sept) wet-day frequency indicated a good match between the aggregated statistics derived from the local rain gauge measurements and corresponding estimates derived from the ERA5 reanalysis in terms of the cross-validation correlation.

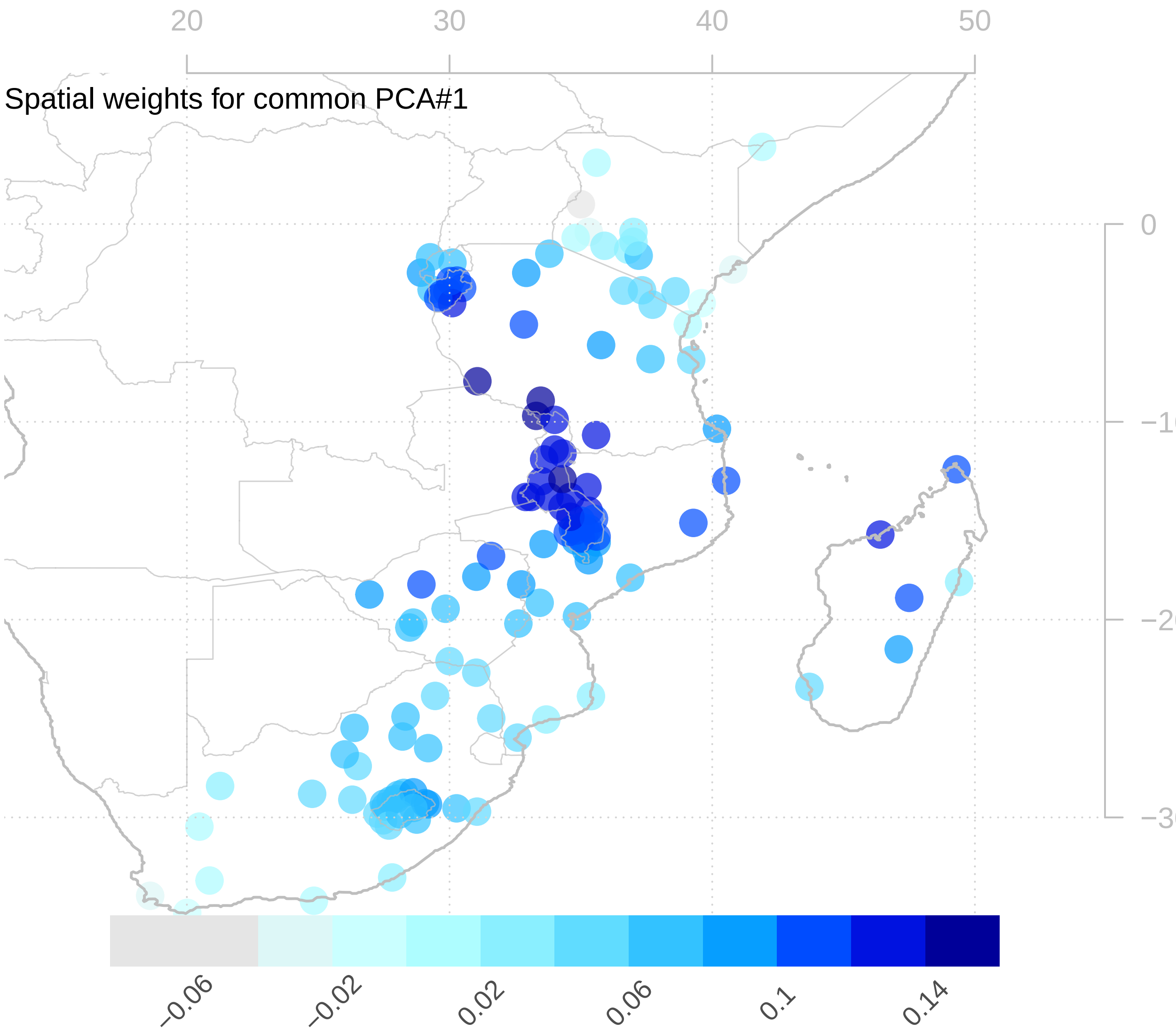
The chunk of R-code below was used to examine the mean annual cycle on he wet-day frequency:

```
FW.ac <- aggregate(regrid(FW,is=fw.ac),by=month,FUN='mean')
FW.ac.chirps <- aggregate(rr.chirps,by=month,FUN='wetfreq')
is <- is.element(loc(FW.ac.chirps),loc(FW.ac))
FW.ac.chirps <- subset(FW.ac.chirps,is=is)
set20 <- !is.finite(coredata(FW.ac.chirps))
coredata(FW.ac.chirps)[set20] <- 0
nv <- apply(fw.ac,2,'nv')
fw.mac.obs <- subset(fw.ac,is=nv==12)
fw.mac.era5 <- subset(FW.ac,is=nv==12)
fw.mac.chirps <- subset(FW.ac.chirps,is=nv==12)
## 'Common' PCA
fw.mac.both <- zoo(x=rbind(coredata(fw.mac.obs),coredata(fw.mac.era5),
                           coredata(fw.mac.chirps)),order.by=1:36)
fw.mac.both <- attrcp(fw.mac.obs,fw.mac.both)
```

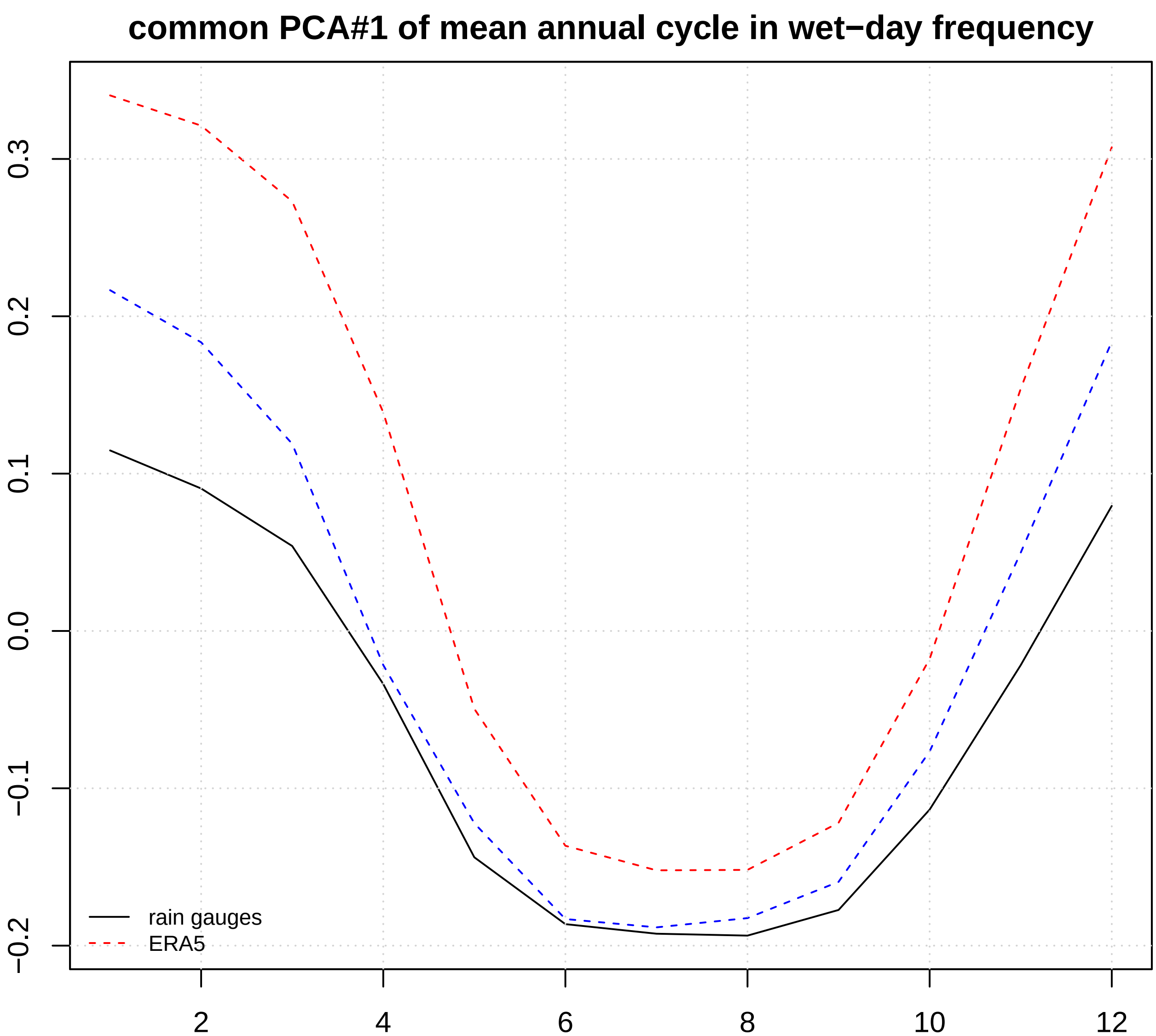
```r
class(fw.mac.both) <- class(fw.mac.obs)
fw.pca.both <- PCA(fw.mac.both)
100*attr(fw.pca.both,'eigenvalues')^2/attr(fw.pca.both,'tot.var')
```

```
##  [1] 76.41225373 11.65278616  4.98672849  2.93459034  1.21634894  0.74466466
##  [7]  0.61706451  0.27429981  0.25275865  0.16555704  0.11853098  0.10052964
## [13]  0.09201248  0.07954302  0.05589666  0.04261388  0.03640561  0.03142843
## [19]  0.03088957  0.02535986
```

```r
map(fw.pca.both,main='Spatial weights for common PCA#1',border=TRUE)
```




```r
plot(zoo(fw.pca.both[1:12,1]),main='common PCA#1 of mean annual cycle in wet-day frequency',
     ylab='weight',xlab='Clendar month',ylim=range(fw.pca.both[,1]))
lines(zoo(coredata(fw.pca.both)[13:24,1],order.by=1:12),col='red',lty=2)
lines(zoo(coredata(fw.pca.both)[25:36,1],order.by=1:12),col='blue',lty=2)
grid()
legend('bottomleft',c('rain gauges','ERA5'),lty=c(1,2),col=c('black','red'),bty='n',cex=0.75)
```



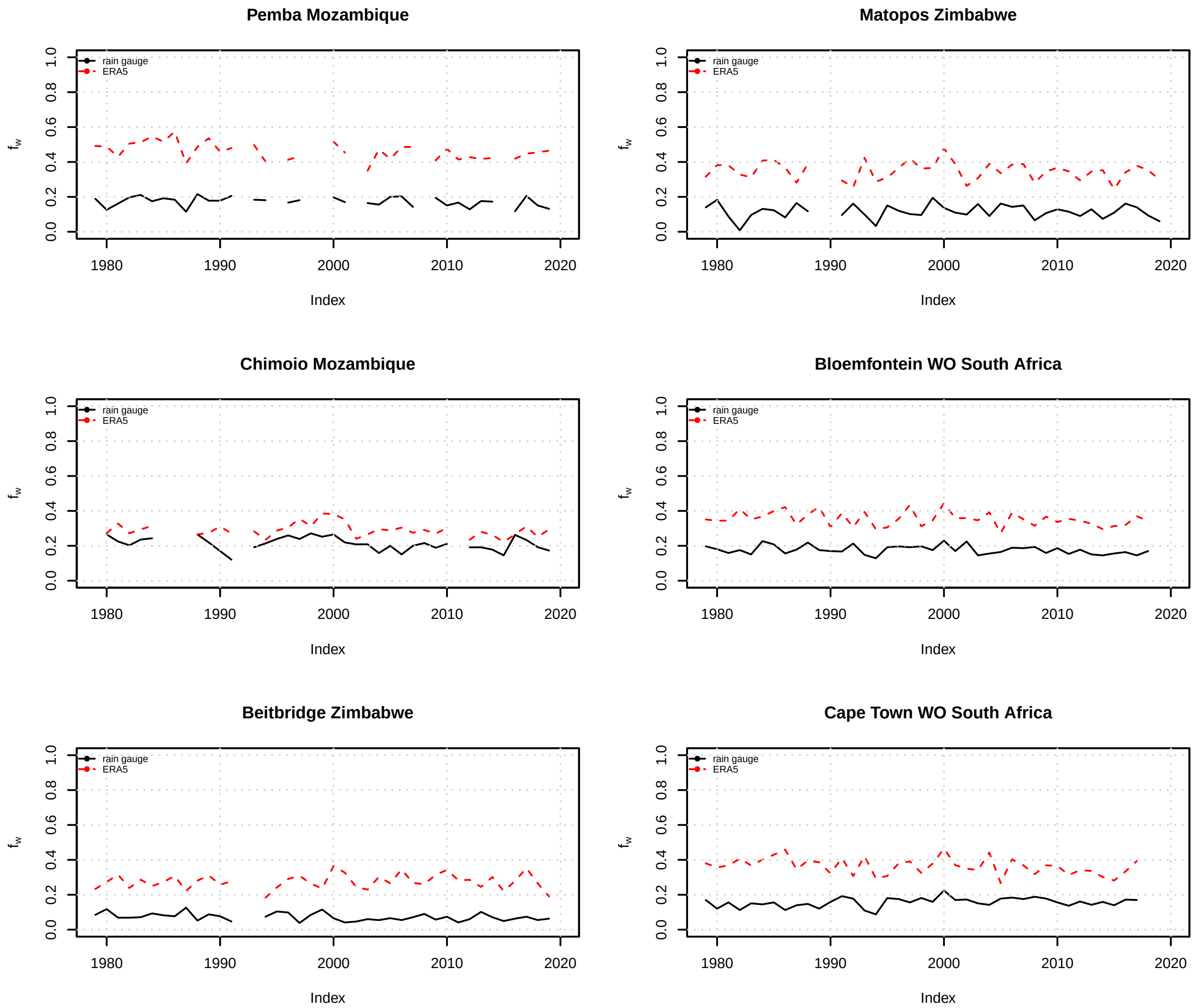
The results from the common PCA for the wet-day frequency f_w suggested that ERA5 has a mean seasonal cycle with too many rainy days over parts of southeastern Africa, despite fairly good match when it comes to total precipitation. The threshold for a wet-day in f_w was 1 mm/day, and discrepancies between ERA5 and rain gauge observations may be due to e.g. biases in low amounts (e.g a 'drizzle effect'). A mismatch between f_w from rain gauges and ERA5 can also be due to small-scale processes with heavy convective rainfall dominating and that the spatial scales of ERA5's grid boxes (~30 km) captures events that often are missed by a rain gauge.

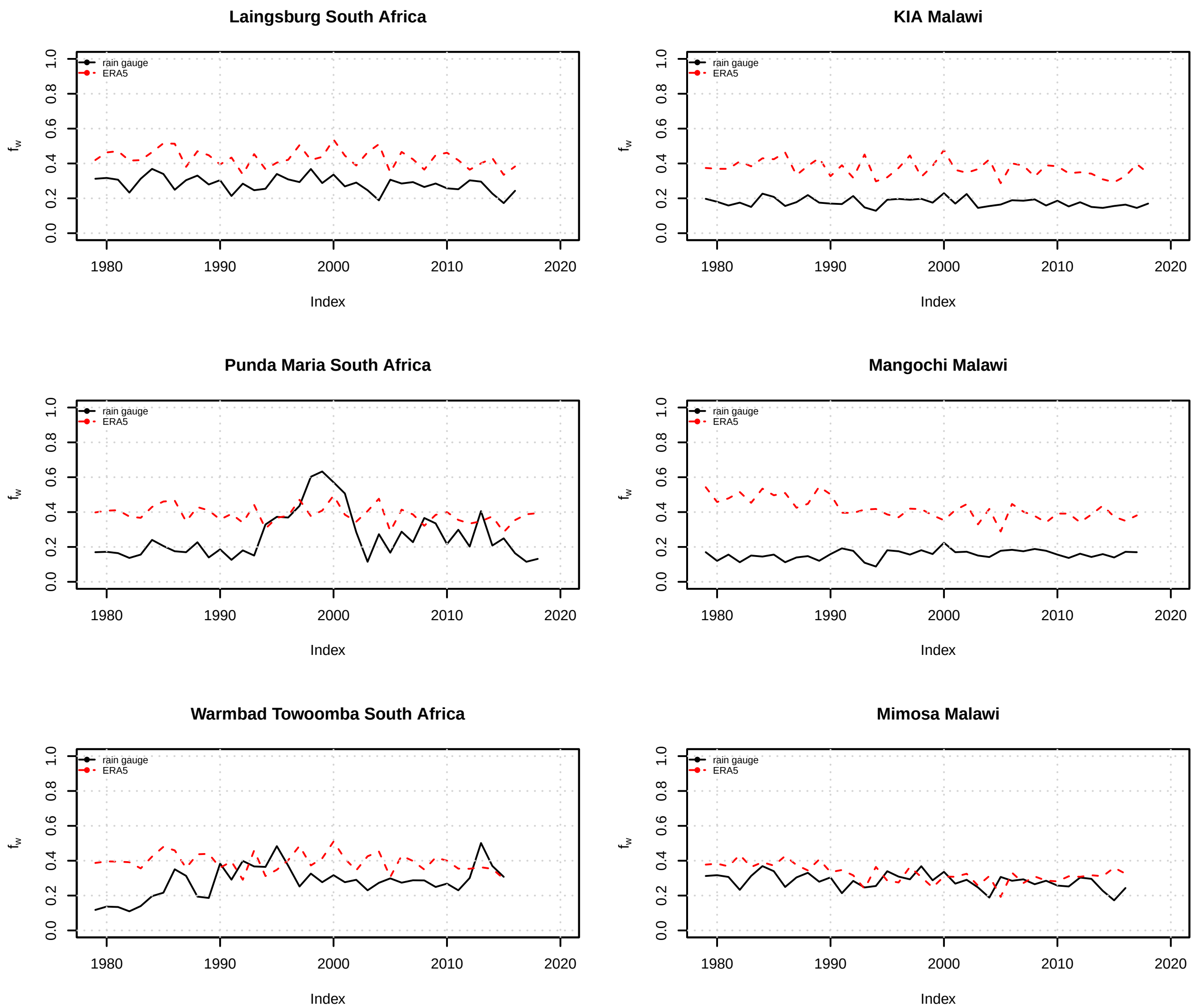
3.1.2 Time series of annual f

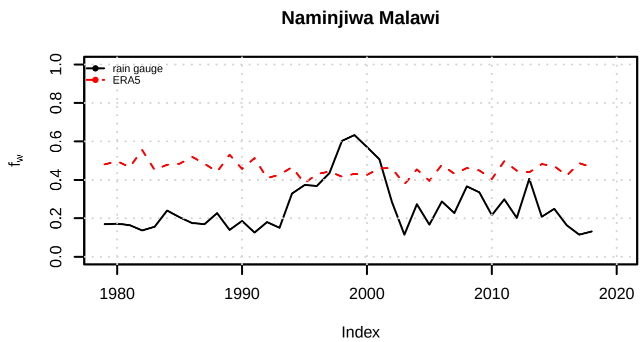
Compare f_w with similar statistics from ERA5 interpolated to same coordinates.

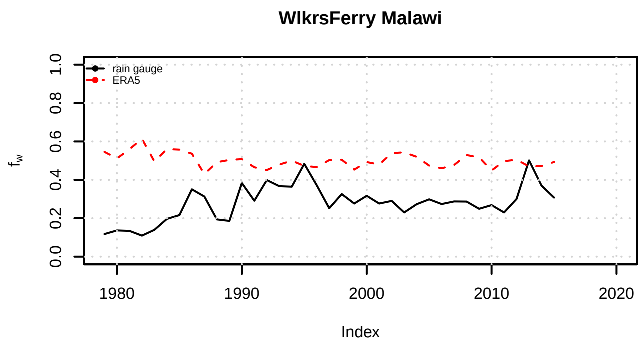
```
FW.x <- regrid(FW,is=fw)
par(mfcol=c(3,2),cex=0.5)
for (i in seq(1,dim(fw)[2],by=5)) {
  it <- matchdate(subset(fw,is=i),subset(FW.x,is=i))
  x1 <- matchdate(subset(fw,is=i),FW.x)
  index(x1) <- year(x1)
  y1 <- matchdate(subset(FW.x,is=i),fw)
```

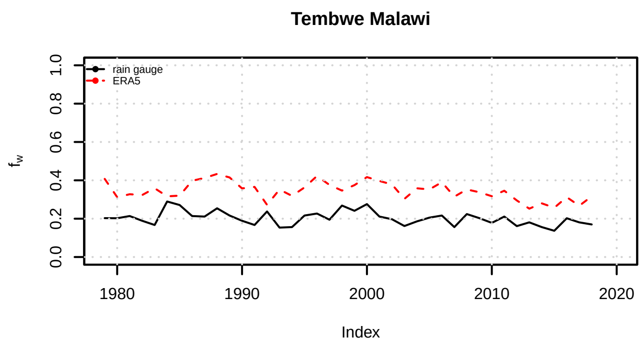
```
  coredata(y1)[is.na(coredata(x1))] <- NA
  rmse.fw <- RMSE(coredata(x1),coredata(y1))
  rmsep.fw <- 100*rmse[i]/sum(x1)
  plot(merge(x1,y1,all=TRUE),plot.type='single',col=c('black','red'),lty=c(1,2),
       main=paste(loc(subset(X,is=i)),cntr(subset(X,is=i))),
       ylab=expression(f[w]),ylim=c(0,1))
  grid()
  legend('topleft',c('rain gauge','ERA5'),col=c('black','red'),pch=19,
         lty=1:2,bty='n',cex=0.6)
}
```




Laingsburg South Africa
KIA Malawi
Punda Maria South Africa
Mangochi Malawi
Warmbad Towoomba South Africa
Mimosa Malawi
rain gauge
ERA5
f_w
0.0
0.2
0.4
0.6
0.8
1.0
1980
1990
2000
2010
2020
Index

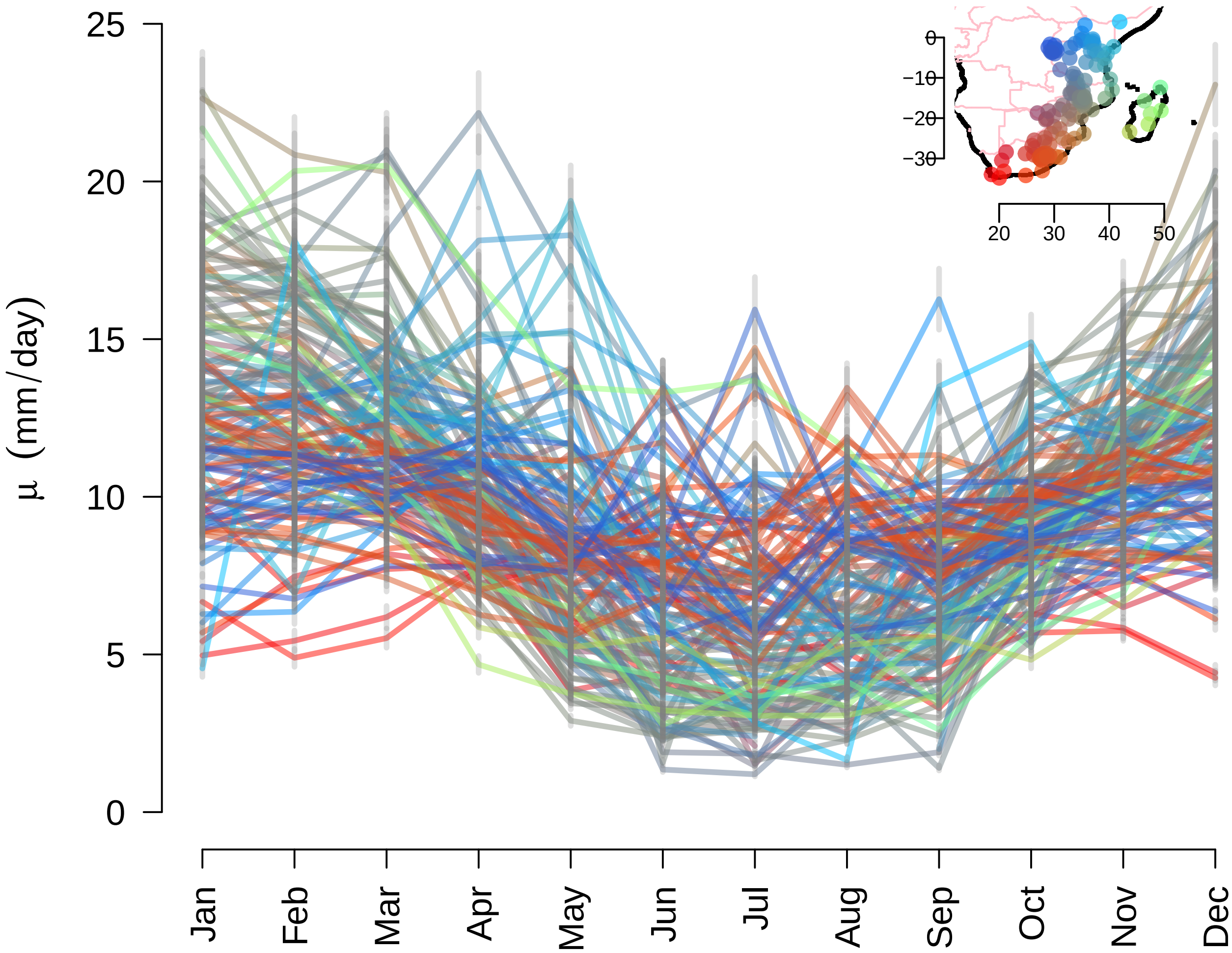




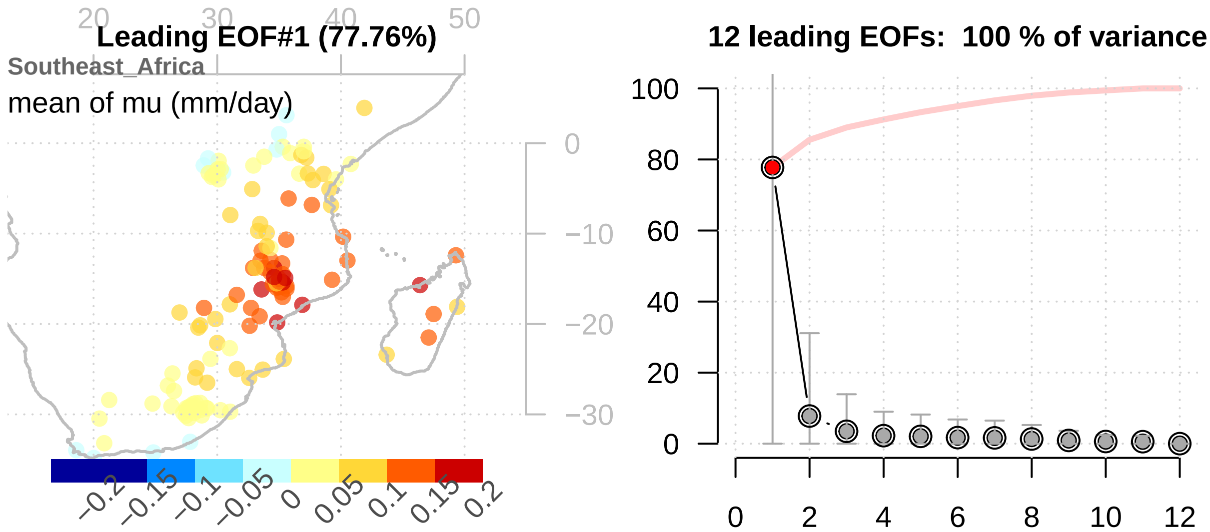


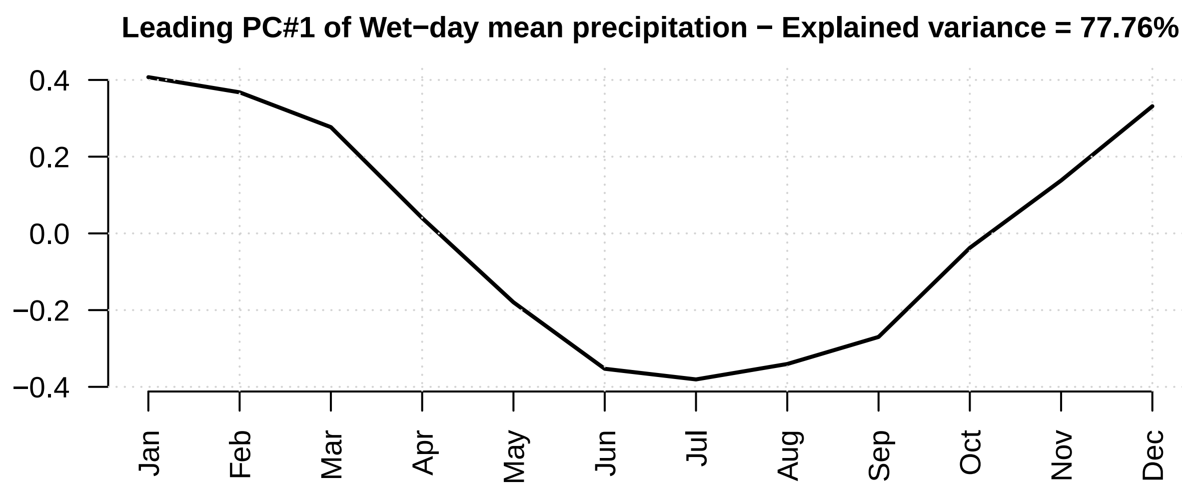
3.2 Mean rainfall intensity μ

```
mu.ac <- aggregate(X,by=month,FUN='wetmean')
plot(mu.ac,new=FALSE)
```




```
## The information about the mean seasonal cycle in wet-day frequency can be summarised
##  through a PCA analysis
pca.mu.ac <- PCA(mu.ac,n=12)
plot(pca.mu.ac,new=FALSE)
```






```r
mu <- annual(subset(X,it=c(1979,2020)),FUN='wetfreq',start=year.start)
nv <- apply(mu,2,'nv')
print(loc(X)[nv <= 33])
```

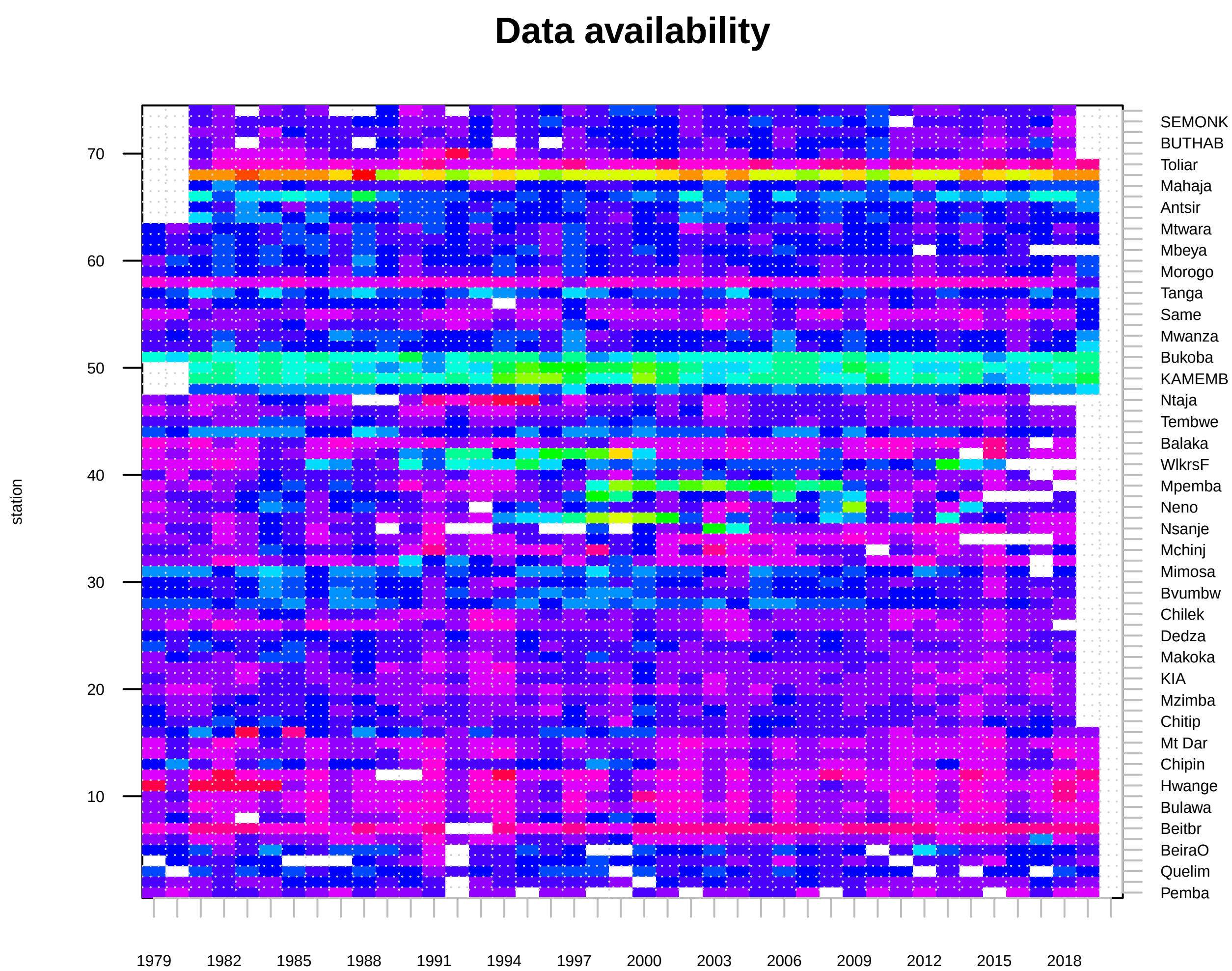
```
##  [1] "Lichinga"                "Tete"
##  [3] "Inhambane"               "Xai-Xai"
##  [5] "Bloemfontein WO"         "Bothaville - Balkfontein"
##  [7] "Brandvlei"               "Cape Agulhas"
##  [9] "Cape St. Francis"        "Cape Town WO"
## [11] "Cedara"                  "East London WO"
## [13] "Irene WO"                "Kimberley WO"
## [15] "Laingsburg"              "Marico"
## [17] "Mount Edgecombe"         "Ottosdal"
## [19] "Polokwane WO"            "Punda Maria"
## [21] "Secunda"                 "Skukuza"
## [23] "Upington WO"             "Warmbad Towoomba"
## [25] "Lodwar"                  "Mandera"
## [27] "Kitale"                  "Kericho"
## [29] "Kisii"                   "Narok"
```

```
## [31] "Nyeri"                      "Dagoretti Corner"
## [33] "Machakos Agromet"           "Voi"
## [35] "Lamu"                       "Moi International Airpor"
## [37] "NA"                         "Sumbawanga"
## [39] "QACHASNEK"                  "LERIBE"
## [41] "MAFETENG"                   "MALEFILOANE"
## [43] "MAPOTENG"                   "MEJAMETALANA"
## [45] "MOSHOESHOEI"                "OXBOW"
## [47] "PHUTHIATSANA"               "QUTHING"
## [49] "BUJUMBURA (Aeroport)"       "GITEGA (Aerodrome)"
## [51] "CANKUZO"                    "GISOZI"
## [53] "MUYINGA"                    "MUSASA"
## [55] "NYAMUSWAGA"                 "MPOTA (Tora)"
```

```r
mu <- subset(mu,is=(nv > 33))
nv <- apply(mu,2,'nv')
print(loc(X)[nv <= 33])
```

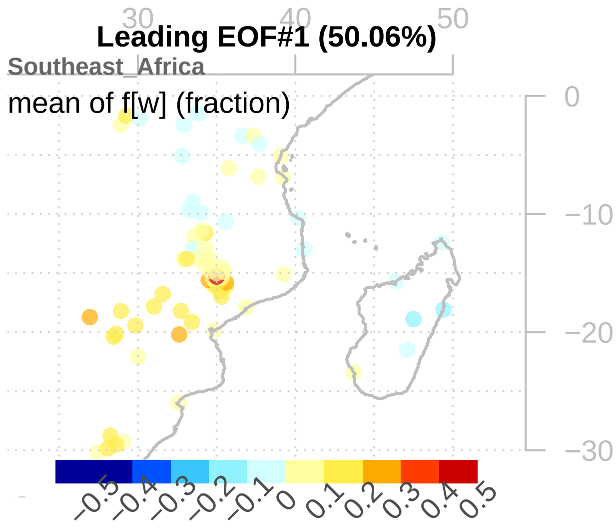
```
## character(0)
```

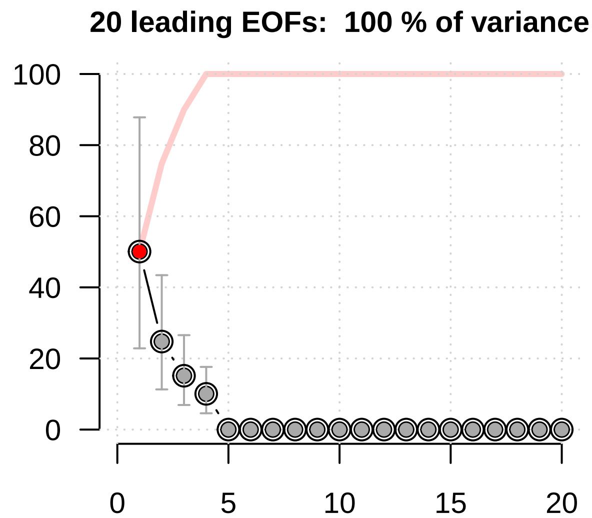
```r
mu <- subset(mu,is=(nv > 33))
diagnose(mu)
```

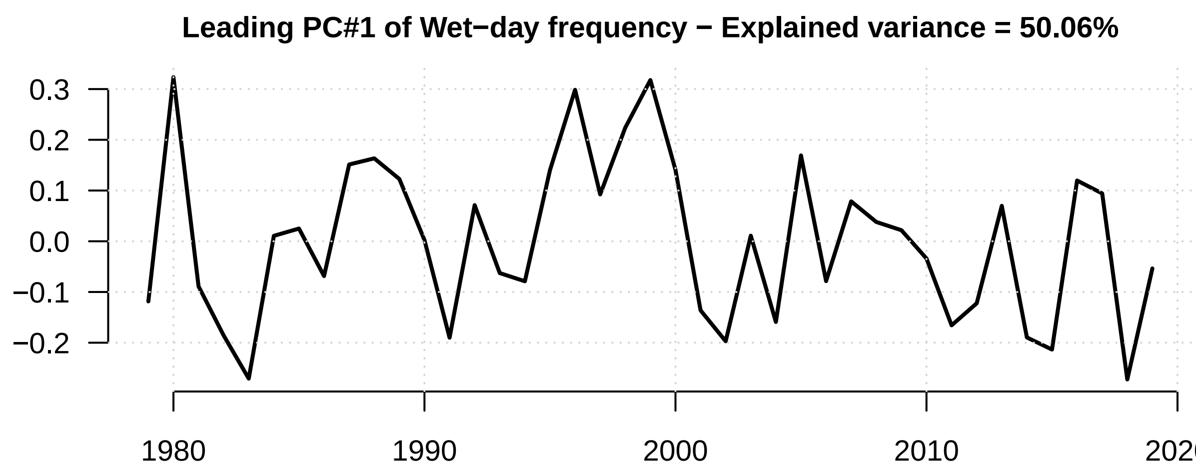


Southeast_Africa

```
pca.mu <- PCA(pcafill(mu))
plot(pca.mu,new=FALSE)
```







The analysis of the mean seasonal variations in μ had less distinctive peaks than for f_w, but nevertheless seasonal undulations with maximum magnitudes in the months October-April for the middle latitude part of the rain gauges. The sites in the very south had a more steady magnitude throughout the year. A PCA applied to the mean seasonal variations in μ exhibited many similar traits as f_w, also with strongest weights in the middle latitude range. For PCA applied to yearly October-September estimates for μ, the leading mode was associated with 47% of the variance with stronger weights towards the south.

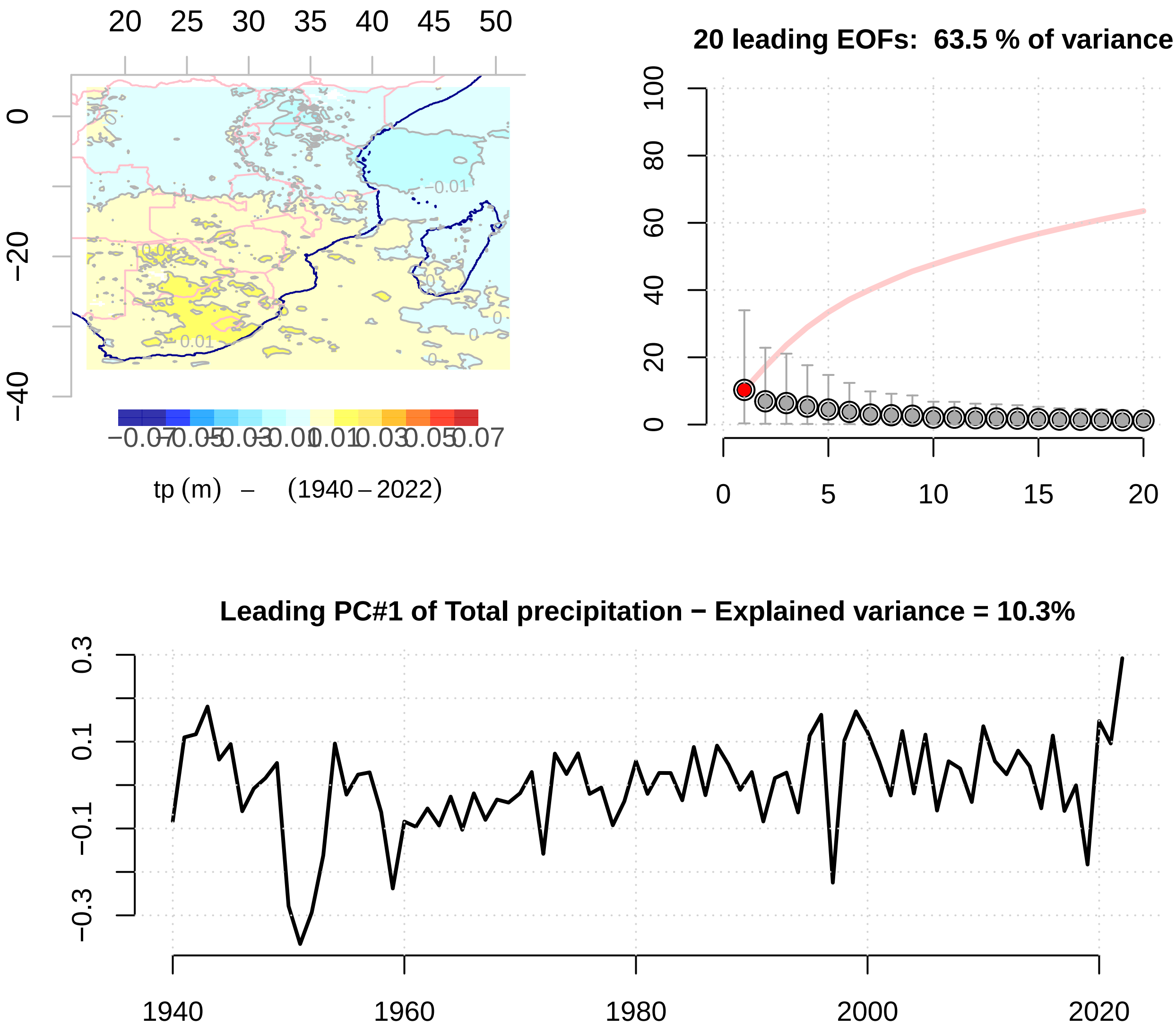
3.2.1 μ derived from ERA5

We carried out a DS-based evaluation of μ, for which we read annual data from ERA5 and use the year October-September:

```
tmp.mu.file <- 'evalAfricaRainfall_mu.tmp.rda'
if (!file.exists(tmp.mu.file)) {
  MU <- retrieve('~/data/ERA5/ERA5_mu_mon.nc',lon=c(17,51),lat=c(-36,4))
  save(MU,file=tmp.mu.file)
} else load(tmp.mu.file)
MU <- 1000*subset(MU,it=c(1940,2022))  ## units mm/day rather than m/day + year
## 2023 looked a bit dubious
eof.MU <- EOF(annual(MU,nmin=1,start=year.start))
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

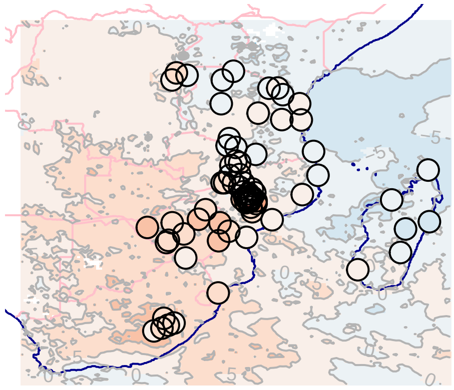
```r
plot(eof.MU,new=FALSE)
```

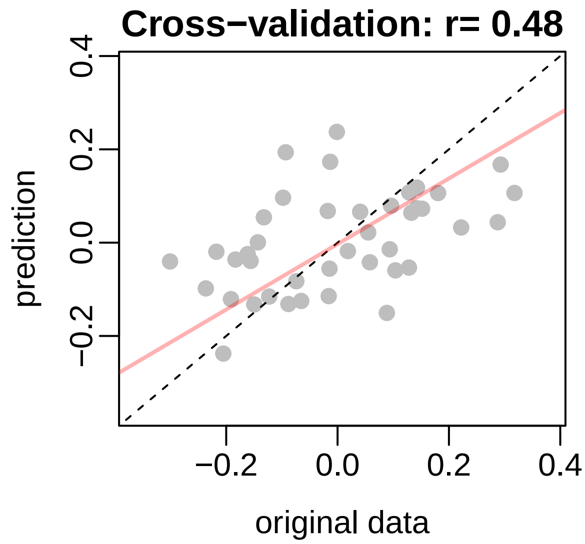


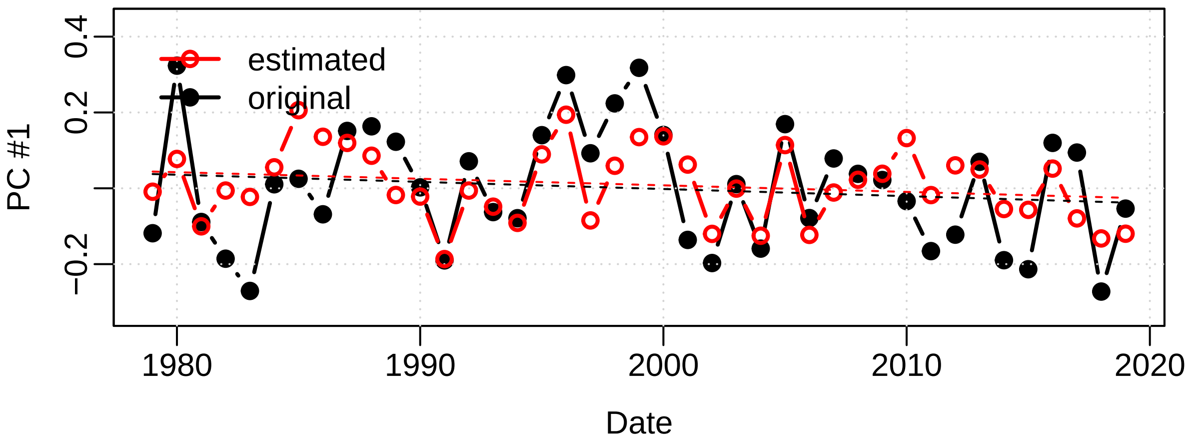

```r
ds.mu <- DS(pca.mu,eof.MU,ip=1:3)
```

```
##   |                                                                            |
```

```r
plot(ds.mu,new=FALSE)
```

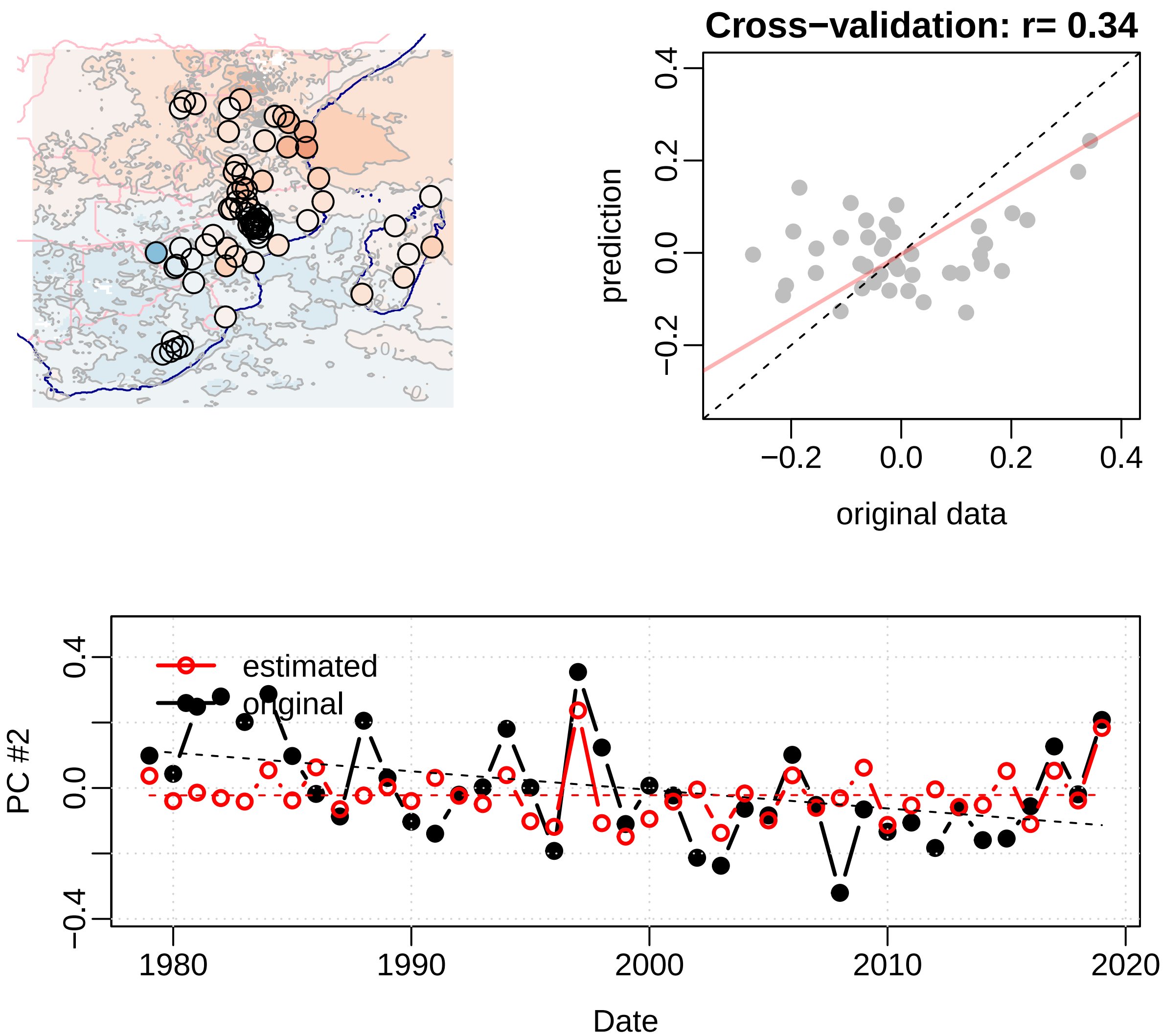






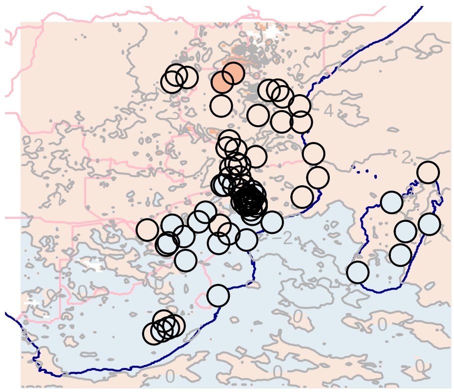

```
## NULL
```

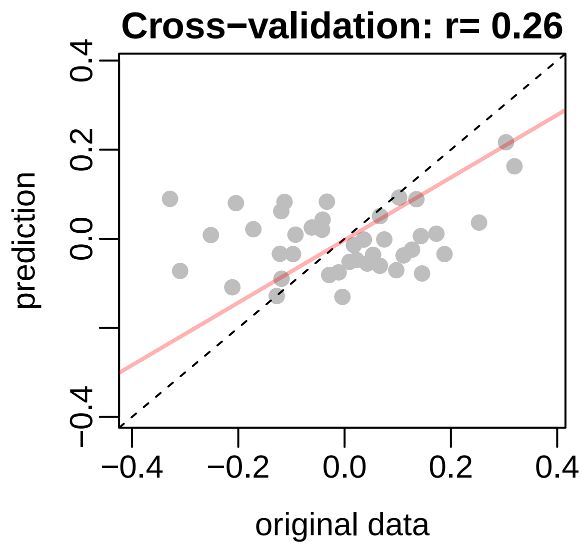
```
plot(ds.mu,ip=2,new=FALSE)
```

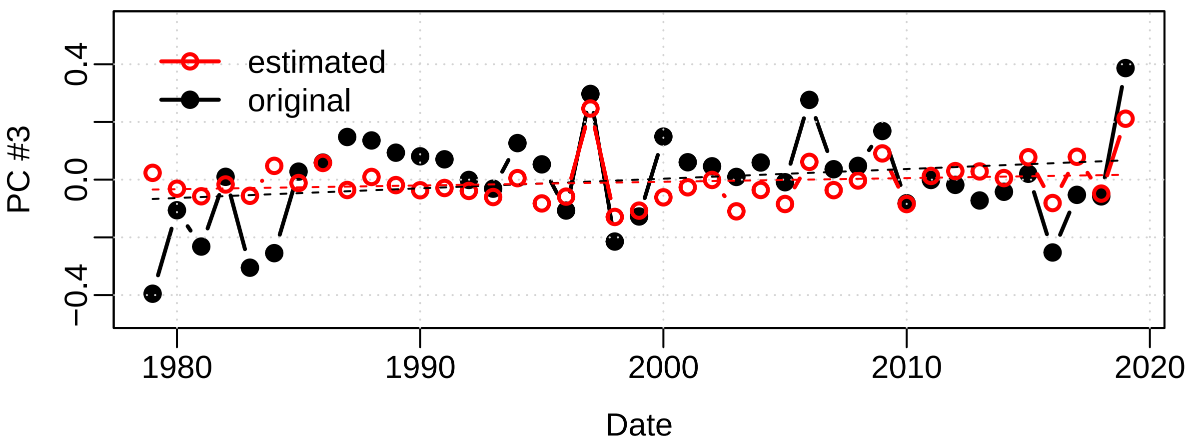



```
## NULL
```

```
plot(ds.mu,ip=3,new=FALSE)
```








```
## NULL
```

The spectrum of eigenvalues for the leading EOFs of annual μ was quite flat with the leading mode only accounting for 10% of the variance and the leading 20 EOFs accounting for 64% of the variance. One possible explanation is that μ is influenced by small-scale (mesoscale) phenomena such as convection, rather than synoptic scale processes.

```
MU.ac <- aggregate(regrid(MU,is=mu.ac),by=month,FUN='mean')
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
MU.ac.chirps <- aggregate(rr.chirps,by=month,FUN='wetmean')
```

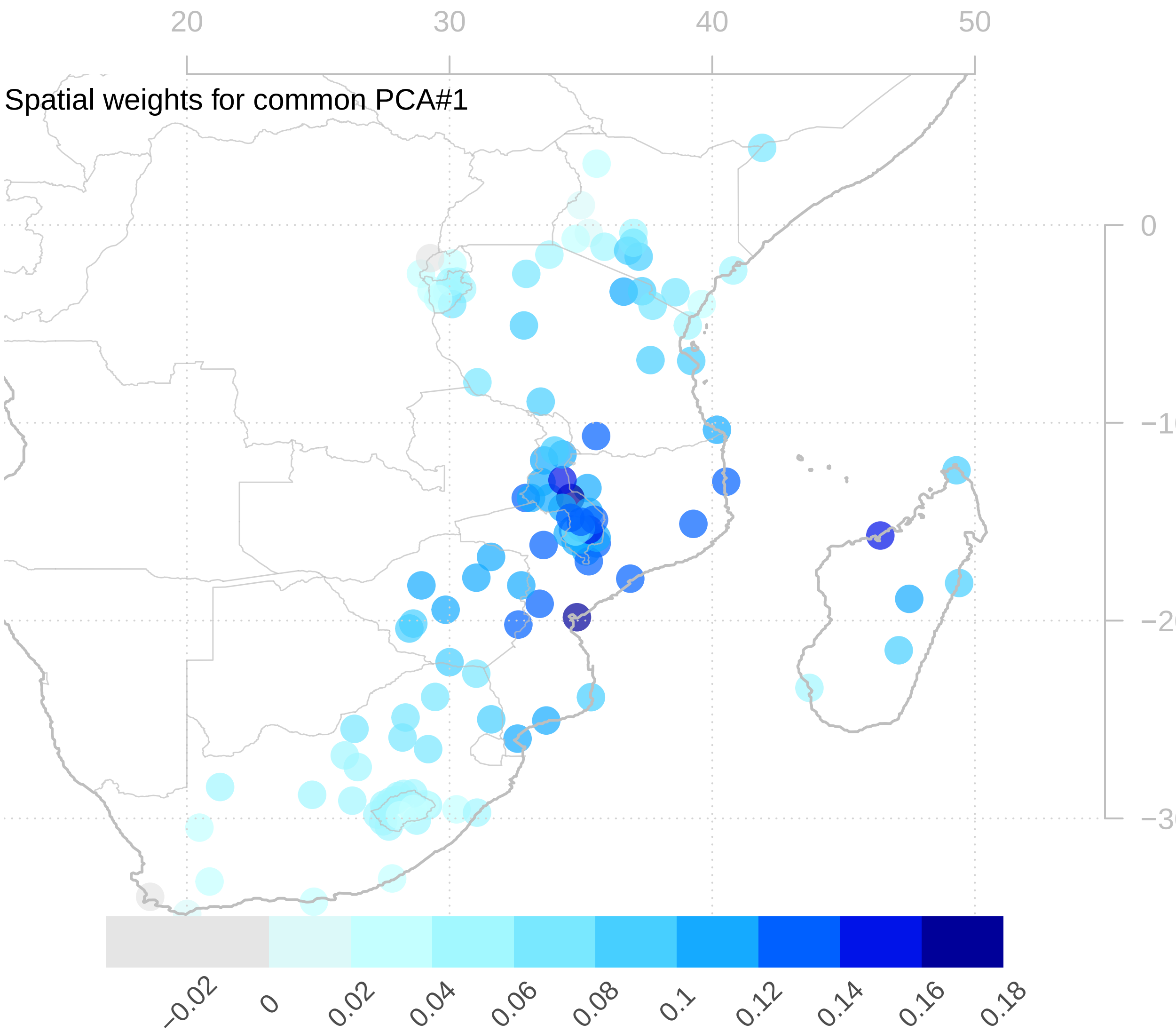
```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
is <- is.element(loc(MU.ac.chirps),loc(MU.ac))
MU.ac.chirps <- subset(MU.ac.chirps,is=is)
set20 <- !is.finite(coredata(MU.ac.chirps))
coredata(MU.ac.chirps)[set20] <- 0
nv <- apply(mu.ac,2,'nv')
mu.mac.obs <- subset(mu.ac,is=nv==12)
mu.mac.era5 <- subset(MU.ac,is=nv==12)
mu.mac.chirps <- subset(MU.ac.chirps,is=nv==12)
```

```
## 'Common' PCA
mu.mac.both <- zoo(x=rbind(coredata(mu.mac.obs),coredata(mu.mac.era5),
                           coredata(mu.mac.chirps)),order.by=1:36)
mu.mac.both <- attrcp(mu.mac.obs,mu.mac.both)
class(mu.mac.both) <- class(mu.mac.obs)
mu.pca.both <- PCA(mu.mac.both)
100*attr(mu.pca.both,'eigenvalues')^2/attr(mu.pca.both,'tot.var')
```

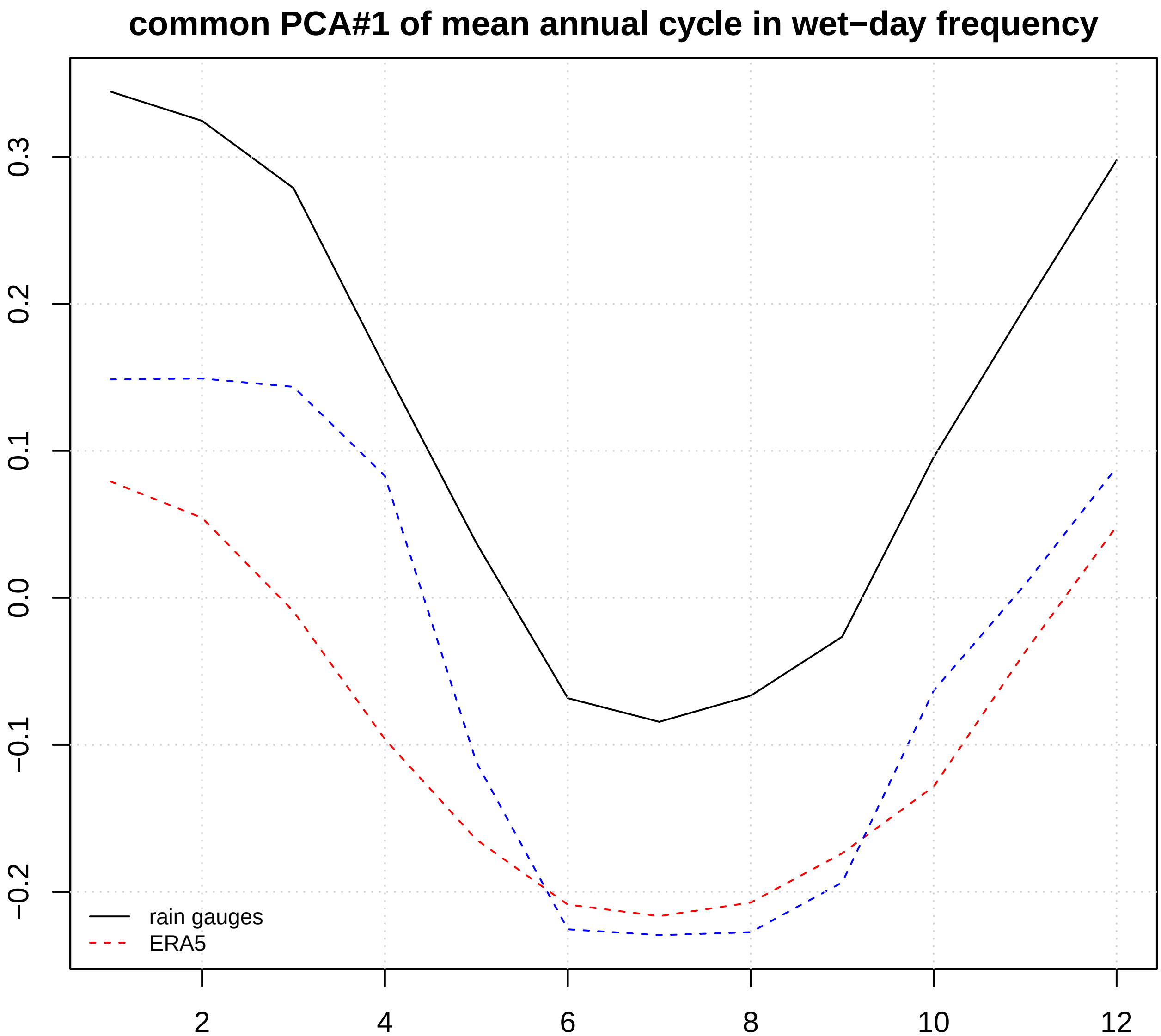
```
##  [1] 64.6723375 12.1565651  8.6803724  4.1304979  2.2428287  1.3300463
##  [7]  0.8915233  0.8076475  0.6900608  0.6545036  0.5127517  0.4323064
## [13]  0.4082747  0.3865848  0.2839827  0.2768029  0.2467369  0.1956355
## [19]  0.1623443  0.1522292
```

```
map(mu.pca.both,main='Spatial weights for common PCA#1',border=TRUE)
```




```
plot(zoo(mu.pca.both[1:12,1]),main='common PCA#1 of mean annual cycle in wet-day frequency',
     ylab='weight',xlab='Clendar month',ylim=range(mu.pca.both[,1]))
lines(zoo(coredata(mu.pca.both)[13:24,1],order.by=1:12),col='red',lty=2)
lines(zoo(coredata(mu.pca.both)[25:36,1],order.by=1:12),col='blue',lty=2)
grid()
```

```
legend('bottomleft',c('rain gauges','ERA5'),lty=c(1,2),col=c('black','red'),bty='n',cex=0.75)
```



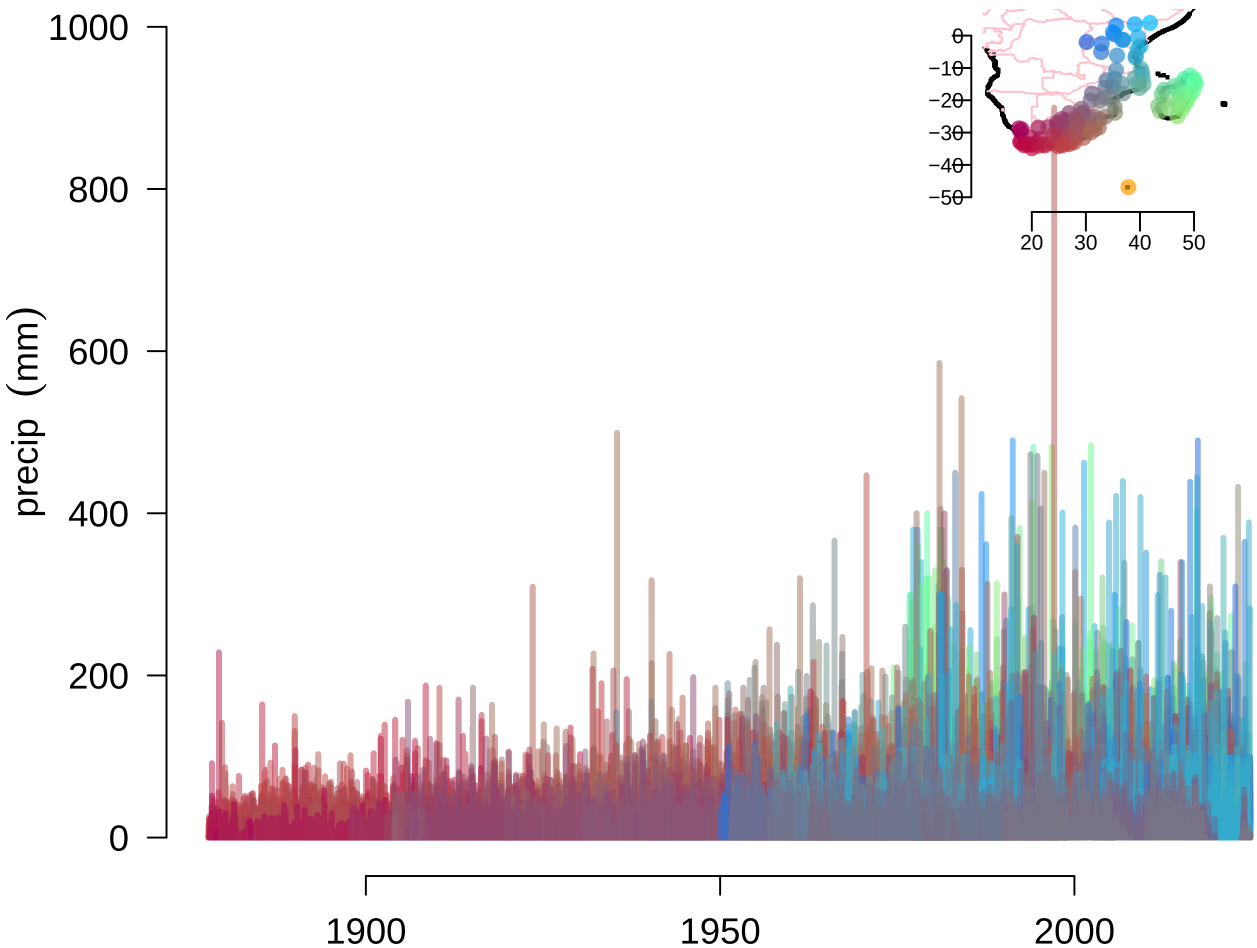
The results from the common PCA for the wet-day frequency μ suggested that the reanalysis generates rainfall with lower intensity than the rain gauges, despite giving a fairly good reproduction of the total rainfall. This is linked to the overestimation of f_w shown above.

3.3 GHCDN

The data from CORDEX can be compared with open access data from the GHCDN database https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily.

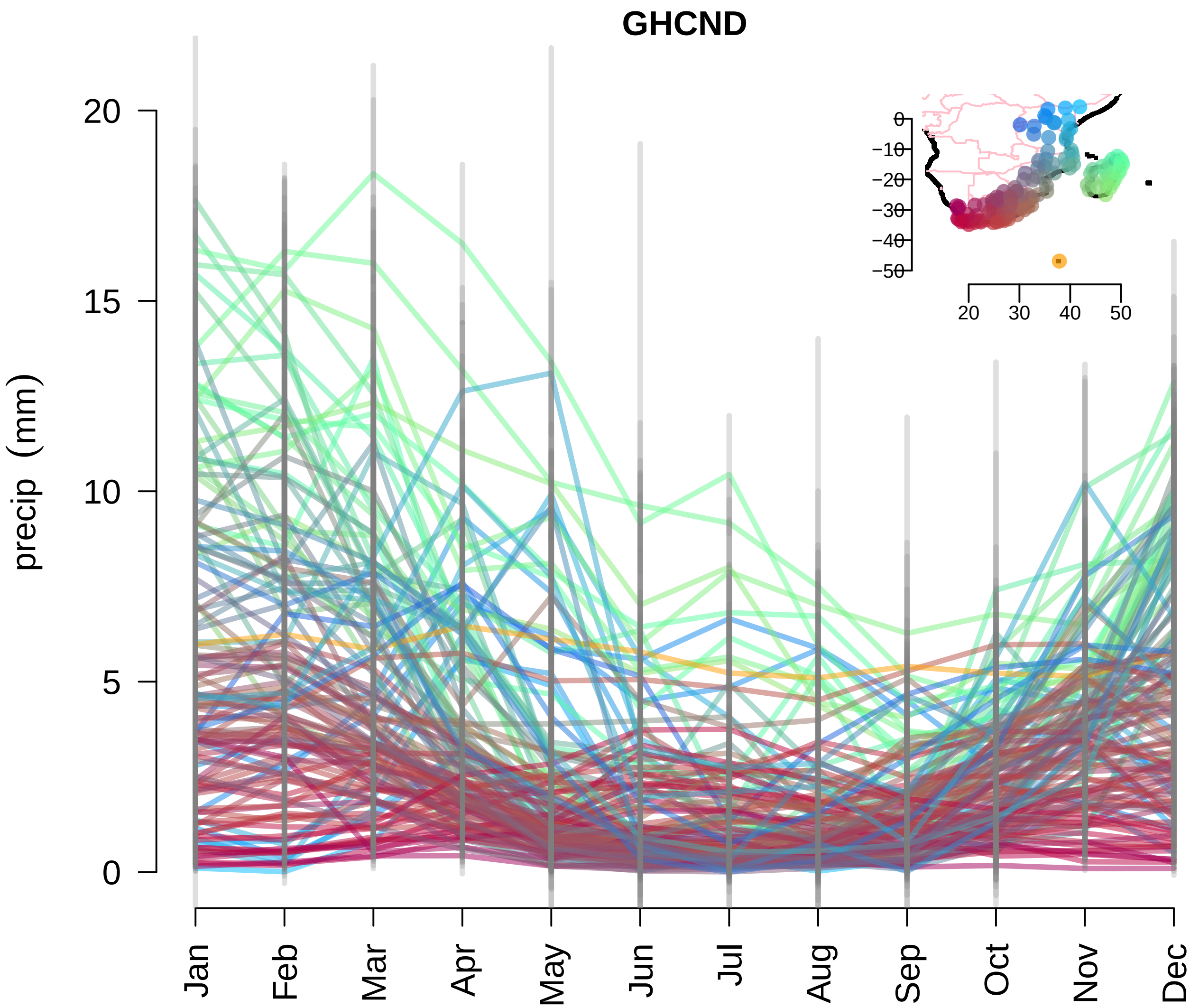
```
if (!file.exists('Y.ghcnd.seafrica.rda')) {
  cntr <- c(rownames(table(cntr(X))),'Burundi','Uganda')
  meta <- meta.GHCND(cntr=cntr)
  nyrs <- meta$end - meta$start
  meta <- meta[(nyrs>30) & (meta$end > 2000),]
  esd::map(meta,new=FALSE)
  Y.ghcnd <- station.GHCND(meta,param='precip')
  save(Y.ghcnd,file='Y.ghcnd.seafrica.rda')
```

```
} else load('Y.ghcnd.seafrica.rda')
plot(Y.ghcnd,new=FALSE)
```



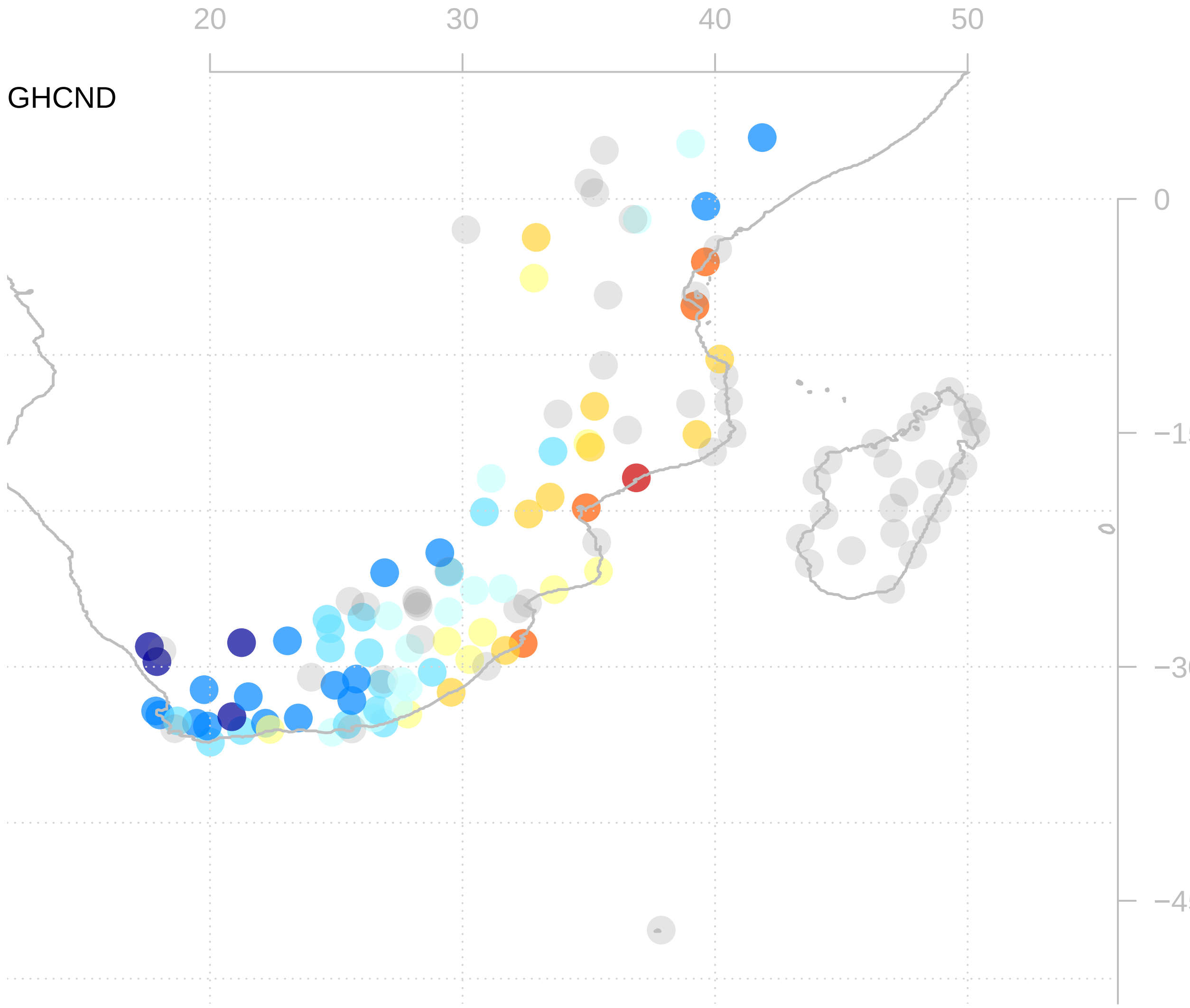

```
Y.ghcnd <- subset(Y.ghcnd,it=c(1979,2023))
plot(aggregate(Y.ghcnd,by=month,FUN='mean'),main='GHCND',new=FALSE)
```

```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

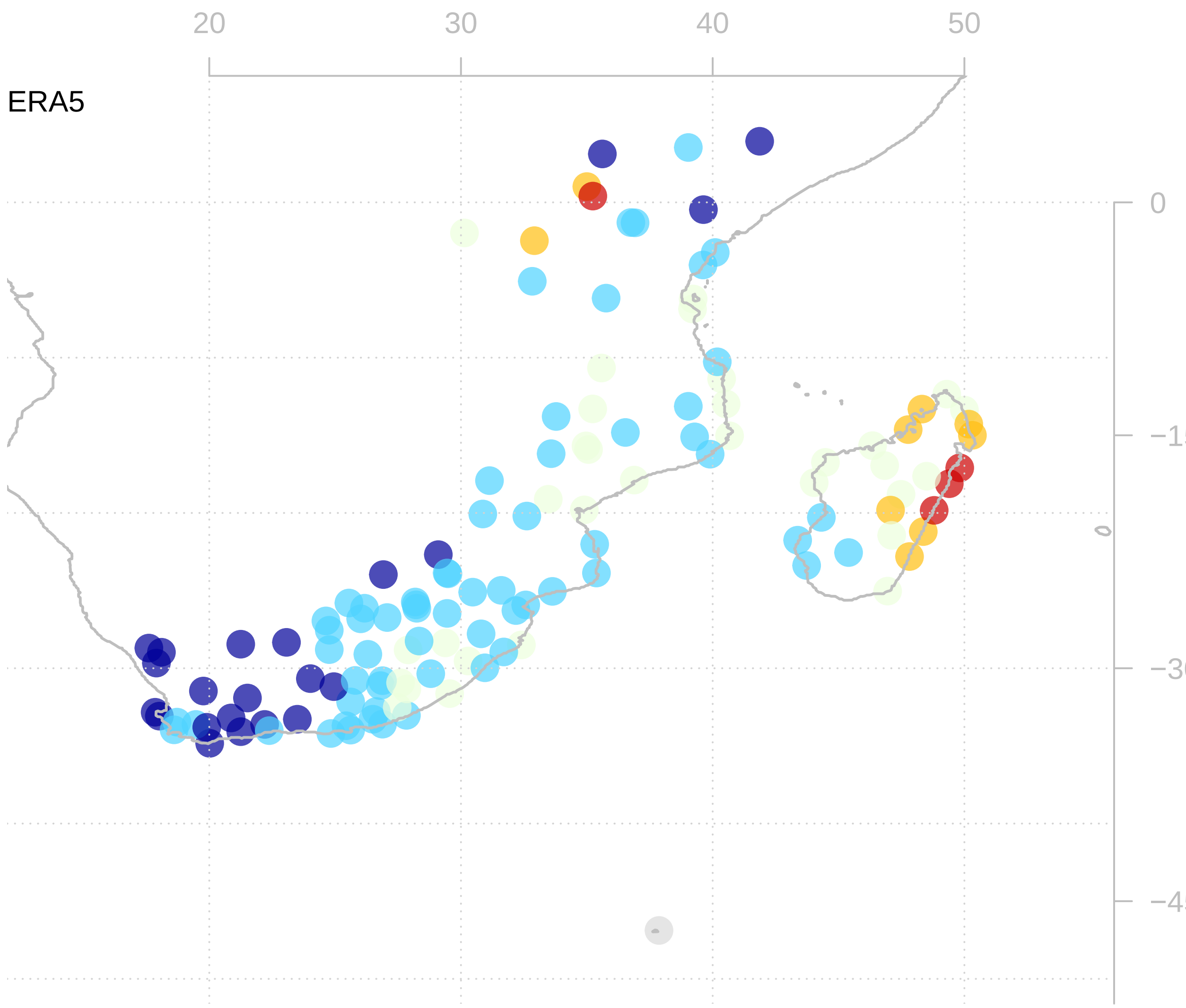


Compare the spatial distribution of rainfall between rain gauges and ERA5

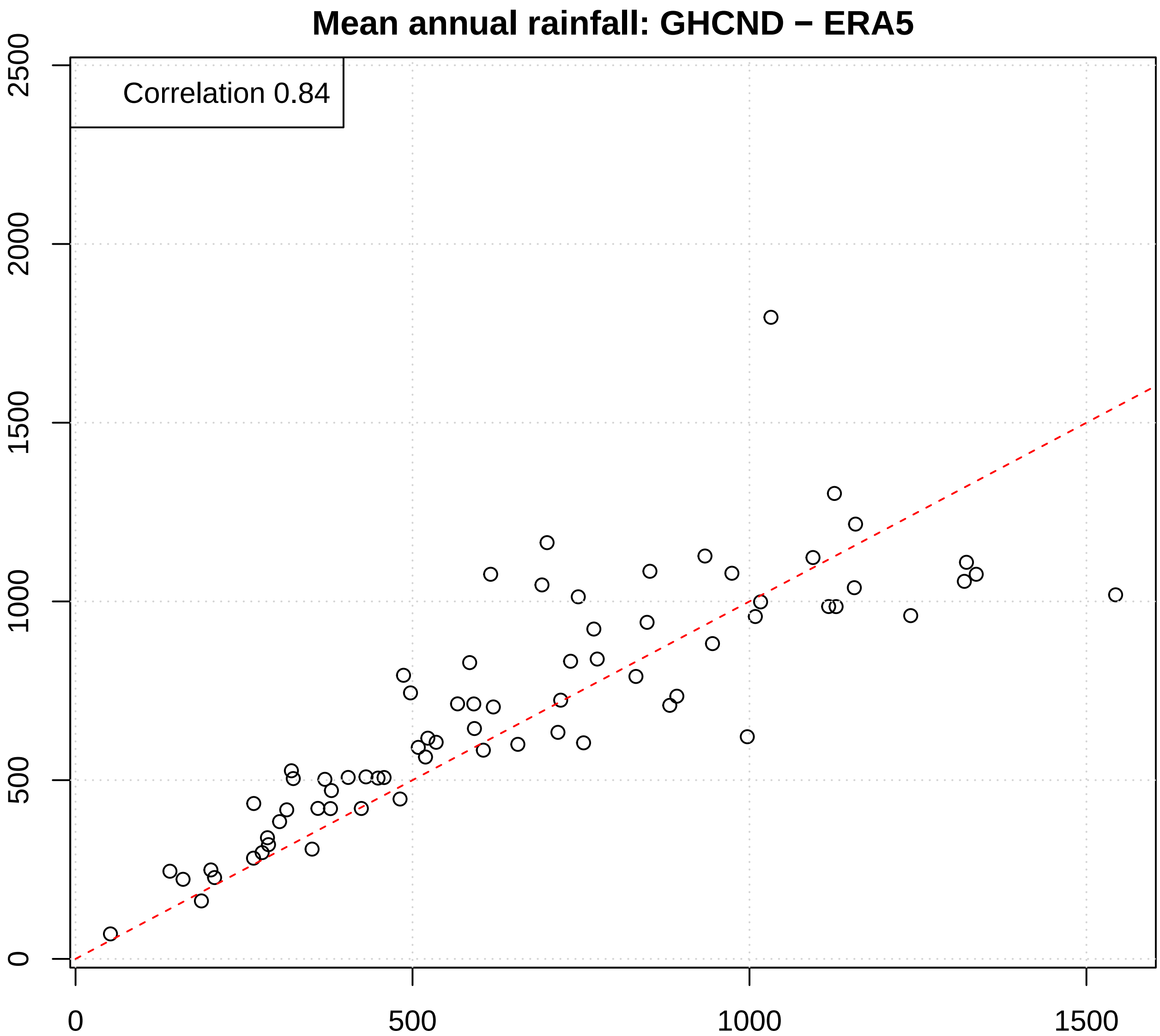
```
map(annual(Y.ghcnd,FUN='sum'),FUN='mean',main='GHCND',new=FALSE) -> m1
```




```
rr.era5.chirps <- regrid(era5,is=Y.ghcnd)
map(subset(rr.era5.chirps,it=c(1979,2023)),FUN='mean',main='ERA5',new=FALSE) -> m2
```




```
m1 <- c(coredata(m1)); m2 <- c(coredata(m2))
plot(m1,m2,main='Mean annual rainfall: GHCND - ERA5',
     xlab='GHCND (mm)',ylab='ERA5 (mm)')
grid()
ok <- is.finite(m1) & is.finite(m2)
legend('topleft',paste('Correlation',round(cor(m1[ok],m2[ok]),2)))
lines(c(0,2500),c(0,2500),lty=2,col='red')
```



A comparison between the mean annual rainfall from the same locations in ERA5 nd GHCND indicated a high correlation (0.84).

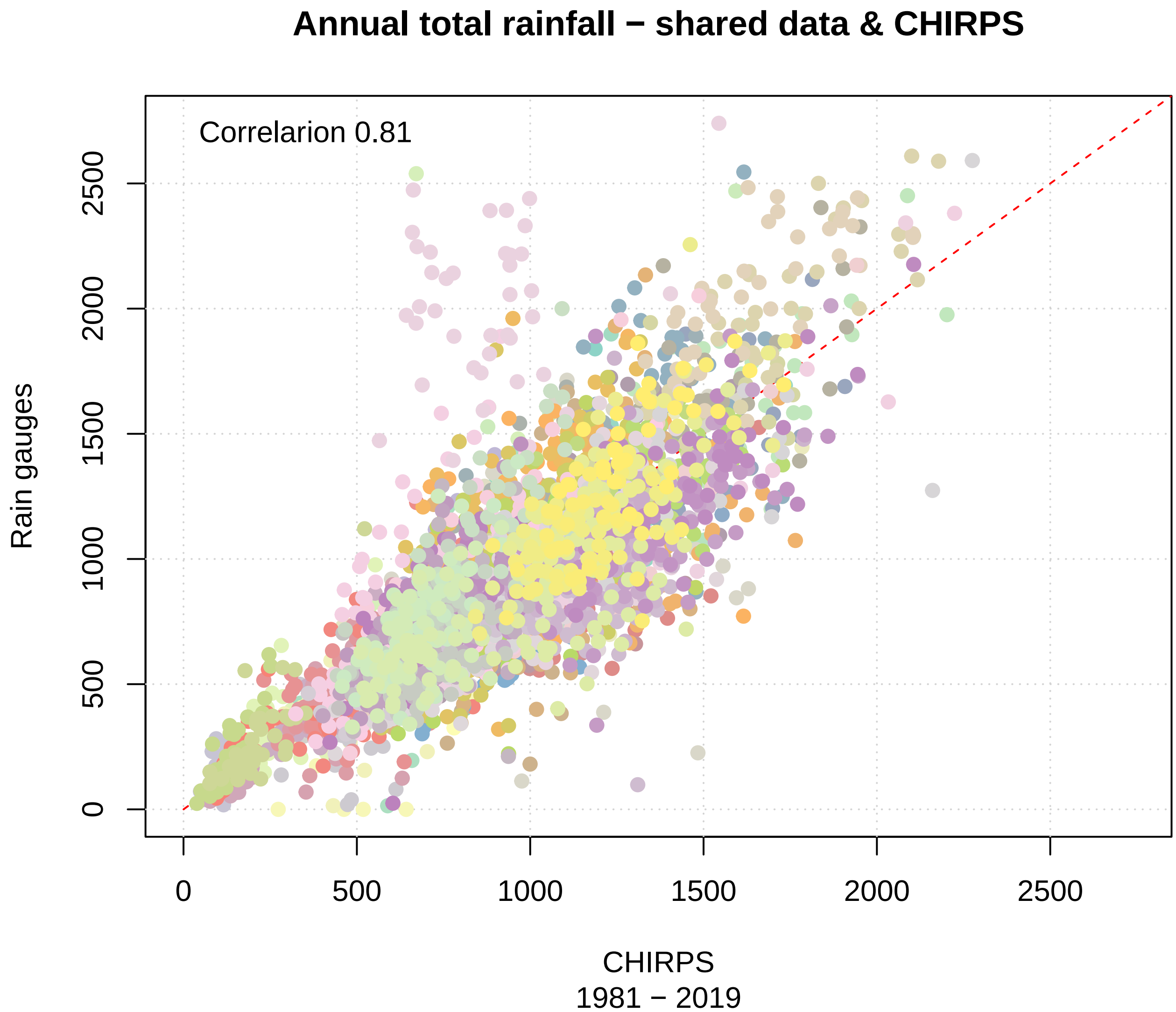
3.4 Compare with CHIRPS

```
ok <- (apply(RR.chirps,2,'nv') > 0)
print(paste(sum(ok),'sites with valid data'))
```

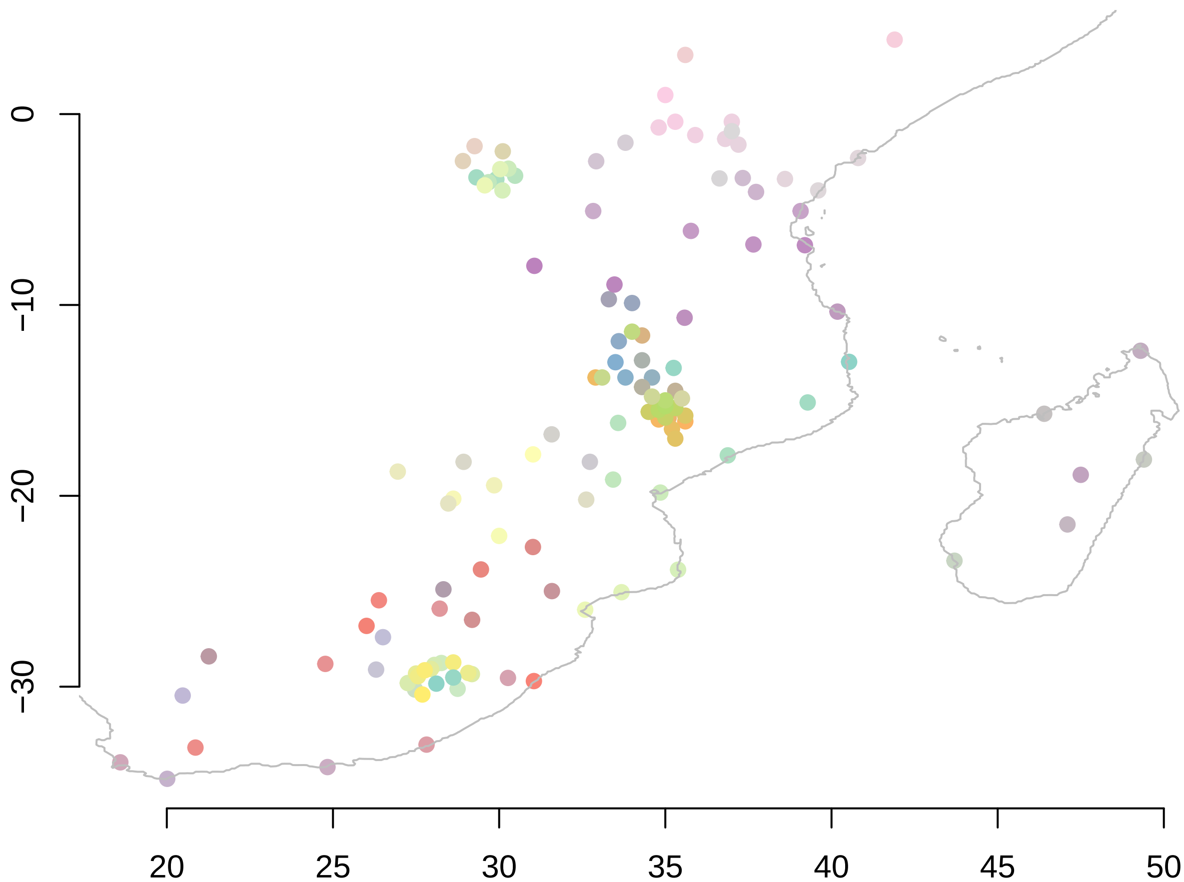
```
## [1] "120 sites with valid data"
```

```
XY <- merge(zoo(subset(RR.chirps,is=ok)),zoo(subset(atX,is=ok,it=c(1981,2019))),all=FALSE)
ns <- sum(ok)
palette <- colorRampPalette(brewer.pal(12, "Set3"))(ns)
r <- round(cor(c(coredata(XY)[,1:ns]),c(coredata(XY)[,(ns+1):(2*ns)]),use='complete.obs'),2)
plot(range(XY,na.rm=TRUE),range(XY,na.rm=TRUE),type='n',
     xlab='CHIRPS',ylab='Rain gauges',main='Annual total rainfall - shared data & CHIRPS',
     sub=paste(range(year(RR.chirps)),collapse=' - '))
grid()
legend('topleft',paste('Correlarion',r),bty='n')
```

```r
lines(c(0,3000),c(0,3000),lty=2,col='red')
for (is in 1:ns) points(coredata(XY)[,is],coredata(XY)[,is+ns],pch=19,col=palette[is])
```




```r
plot(lon(RR.chirps),lat(RR.chirps),pch=19,col=palette,bty='n',xlab='',ylab='')
data(geoborders)
lines(geoborders,col='grey')
```



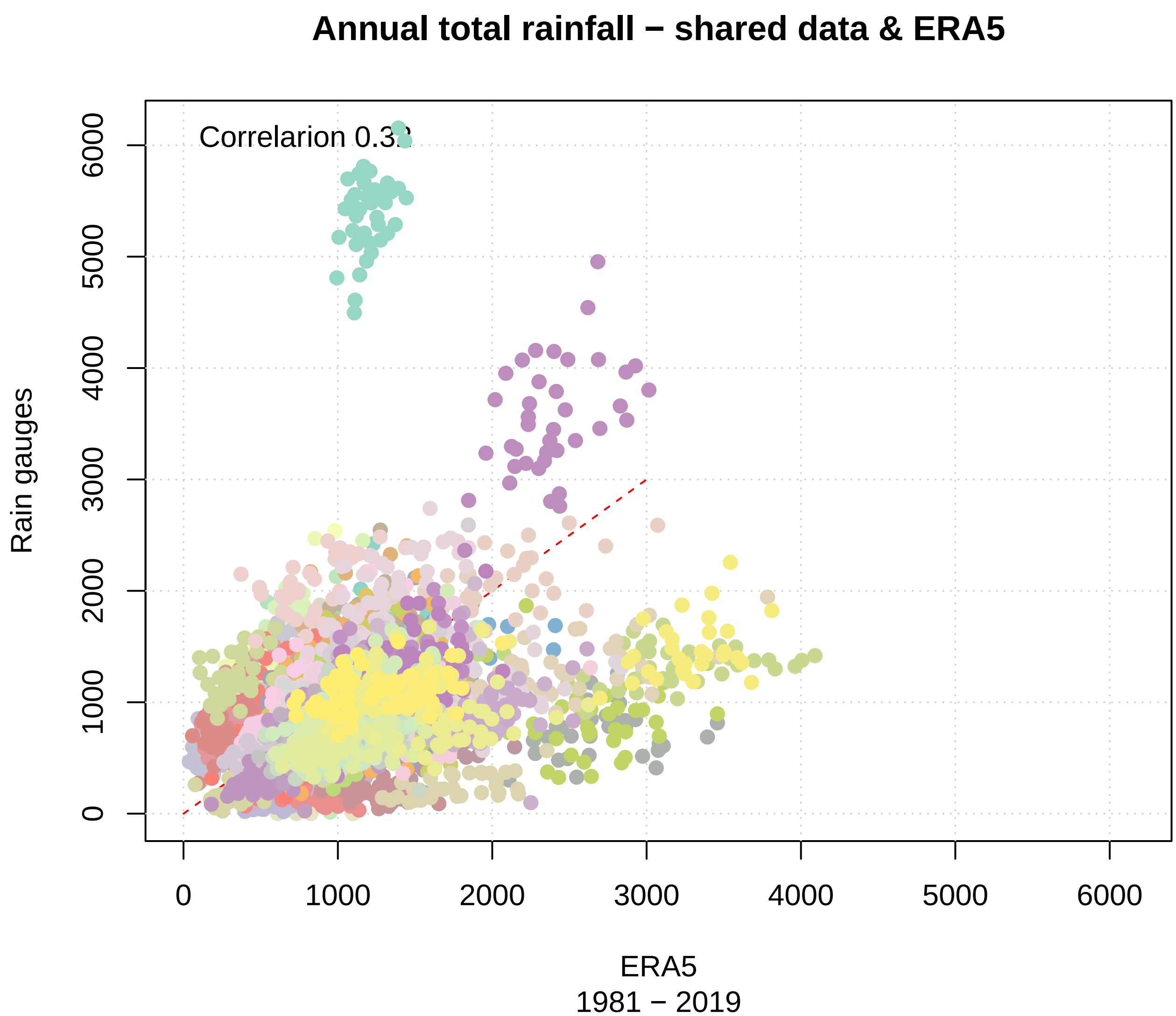

```
#text(lon(RR.chirps),lat(RR.chirps),loc(RR.chirps),cex=0.5)
```

The correlation between CHIRPS and the annual rainfall derived from the shared rain gauge data was high (0.81), albeit with a tendency of clustering according to the location of the measurements (cites). The colour of the data points match the locations for those shown on the map.

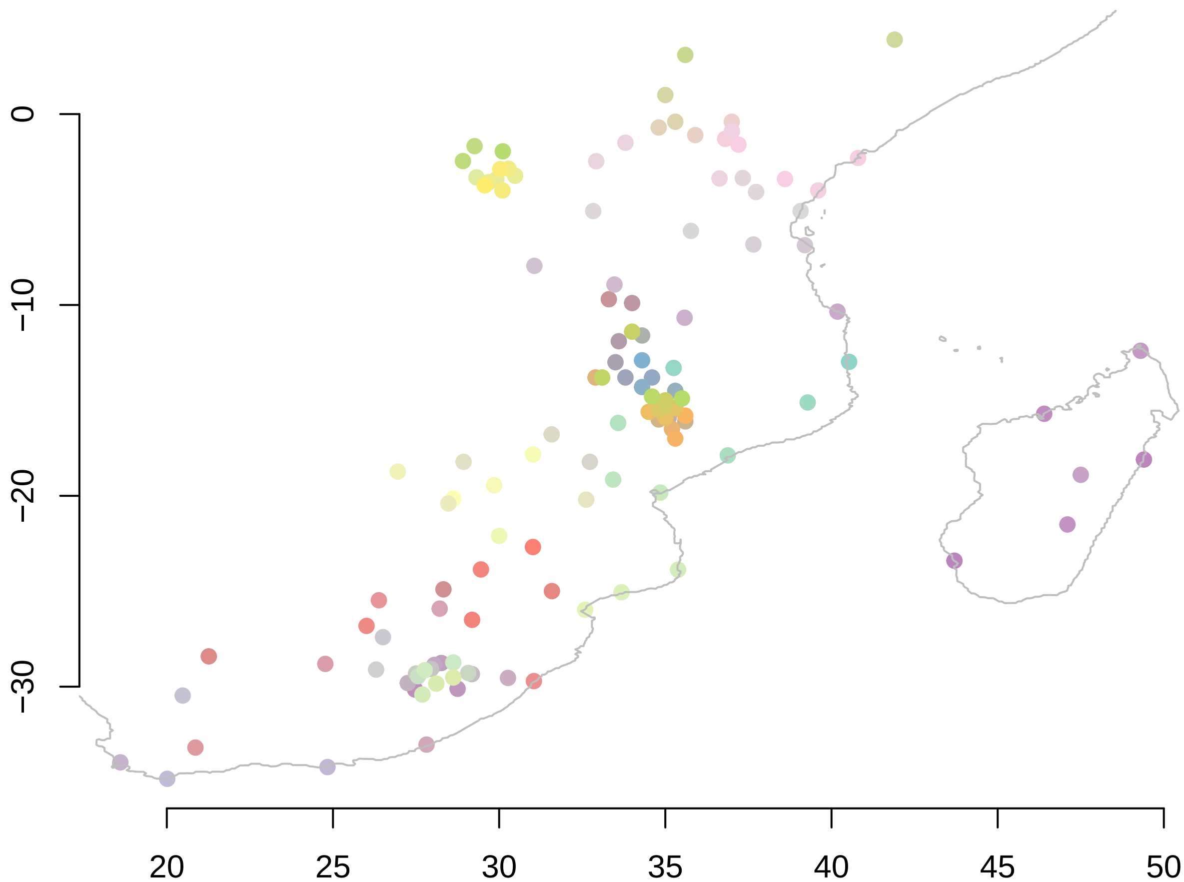
```
ok <- (apply(atX,2,'nv') > 0)
print(paste(sum(ok),'sites with valid data'))
```

```
## [1] "130 sites with valid data"
```

```
XY <- merge(zoo(subset(RR.era5,is=ok)),zoo(subset(atX,is=ok,it=c(1981,2019))),all=FALSE)
ns <- sum(ok)
palette <- colorRampPalette(brewer.pal(12, "Set3"))(ns)
r <- round(cor(c(coredata(XY)[,1:ns]),c(coredata(XY)[,(ns+1):(2*ns)]),use='complete.obs'),2)
plot(range(XY,na.rm=TRUE),range(XY,na.rm=TRUE),type='n',
     xlab='ERA5',ylab='Rain gauges',main='Annual total rainfall - shared data & ERA5',
     sub=paste(range(year(RR.chirps)),collapse=' - '))
grid()
legend('topleft',paste('Correlarion',r),bty='n')

lines(c(0,3000),c(0,3000),lty=2,col='red')
for (is in 1:ns) points(coredata(XY)[,is],coredata(XY)[,is+ns],pch=19,col=palette[is])
```




```
plot(lon(RR.chirps),lat(RR.chirps),pch=19,col=palette,bty='n',xlab='',ylab='')
data(geoborders)
lines(geoborders,col='grey')
```



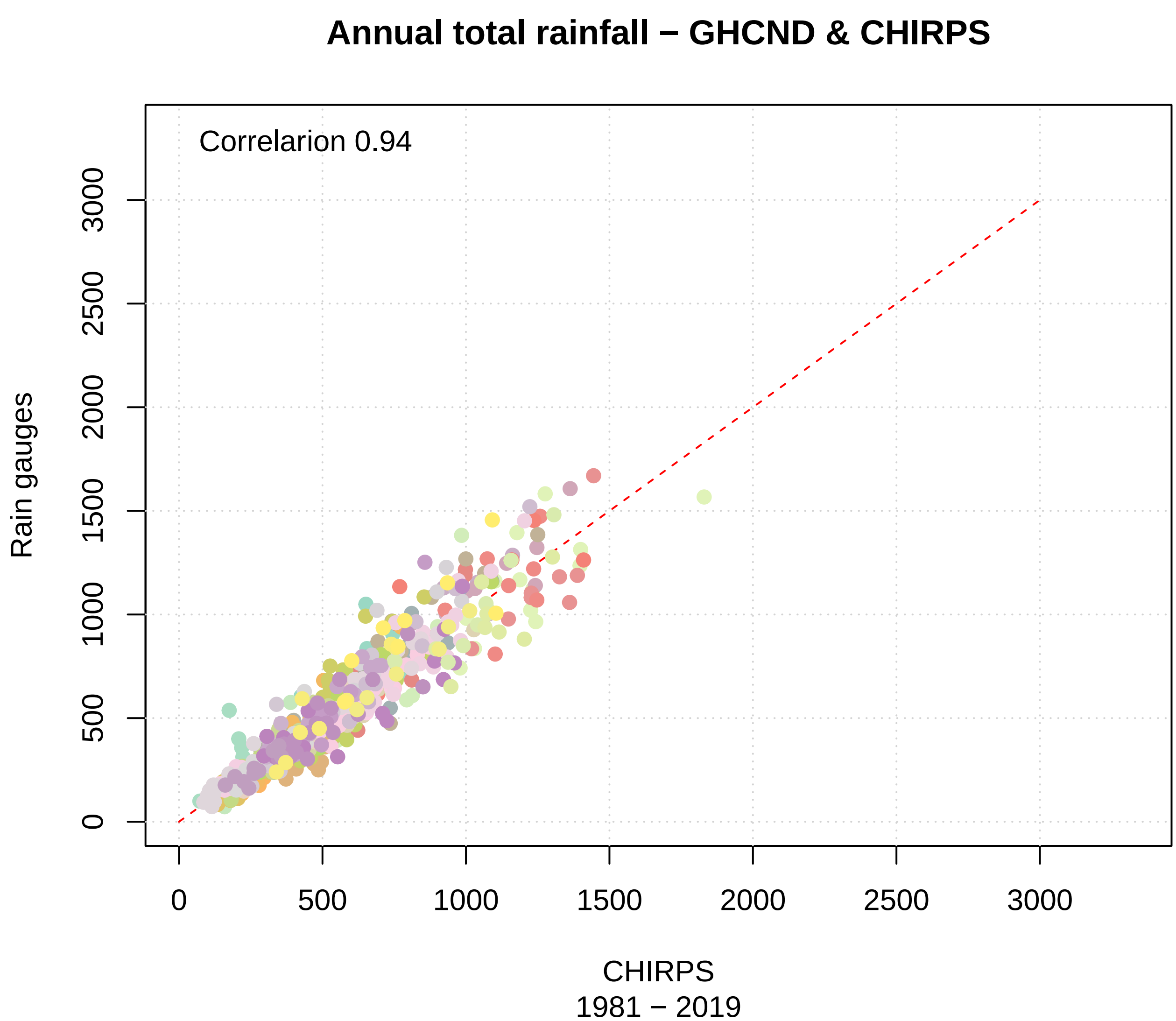

```
#text(lon(RR.chirps),lat(RR.chirps),loc(RR.chirps),cex=0.5)
```

There is a lower correlation (0.4) between the annual rainfall derived from rain gauges and ERA5 than between CHIRPS and the rain gauges (0.81). The comparison suggest that the data points cluster according to the sites from which the rain amount was measured.

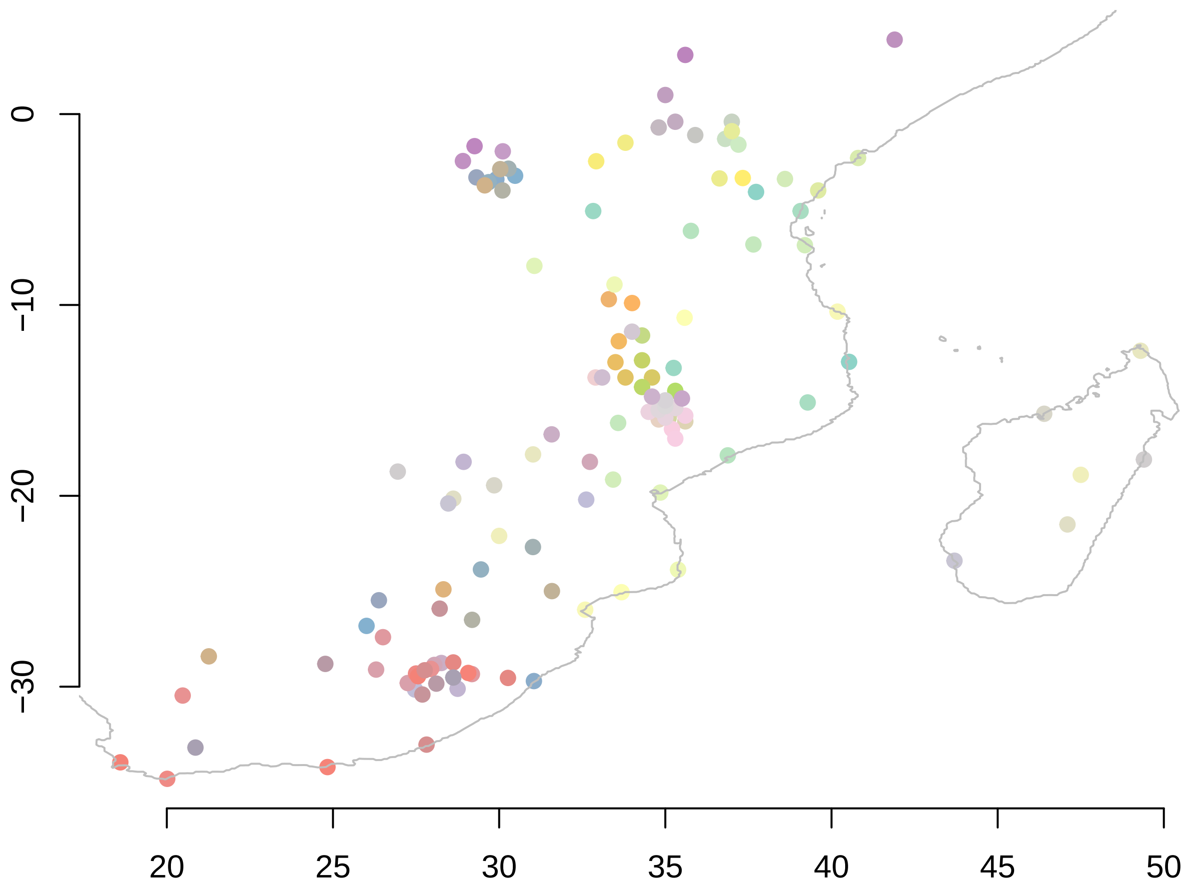
```
RR.ghcnd.chirps <- annual(rr.ghcnd.chirps,FUN='sum',start=year.start)
atY <- annual(Y.ghcnd,FUN='sum',start=year.start)
ok <- (apply(RR.ghcnd.chirps,2,'nv') > 0)
print(paste(sum(ok),'sites with valid data'))
```

```
## [1] "91 sites with valid data"
```

```
XY <- merge(zoo(subset(RR.ghcnd.chirps,is=ok)),zoo(subset(atY,is=ok)),all=FALSE)
ns <- sum(ok)
palette <- colorRampPalette(brewer.pal(12, "Set3"))(ns)
r <- round(cor(c(coredata(XY)[,1:ns]),c(coredata(XY)[,(ns+1):(2*ns)]),use='complete.obs'),2)
plot(range(XY,na.rm=TRUE),range(XY,na.rm=TRUE),type='n',
     xlab='CHIRPS',ylab='Rain gauges',main='Annual total rainfall - GHCND & CHIRPS',
     sub=paste(range(year(RR.chirps)),collapse=' - '))
grid()
legend('topleft',paste('Correlarion',r),bty='n')

lines(c(0,3000),c(0,3000),lty=2,col='red')
for (is in 1:ns) points(coredata(XY)[,is],coredata(XY)[,is+ns],pch=19,col=palette[is])
```



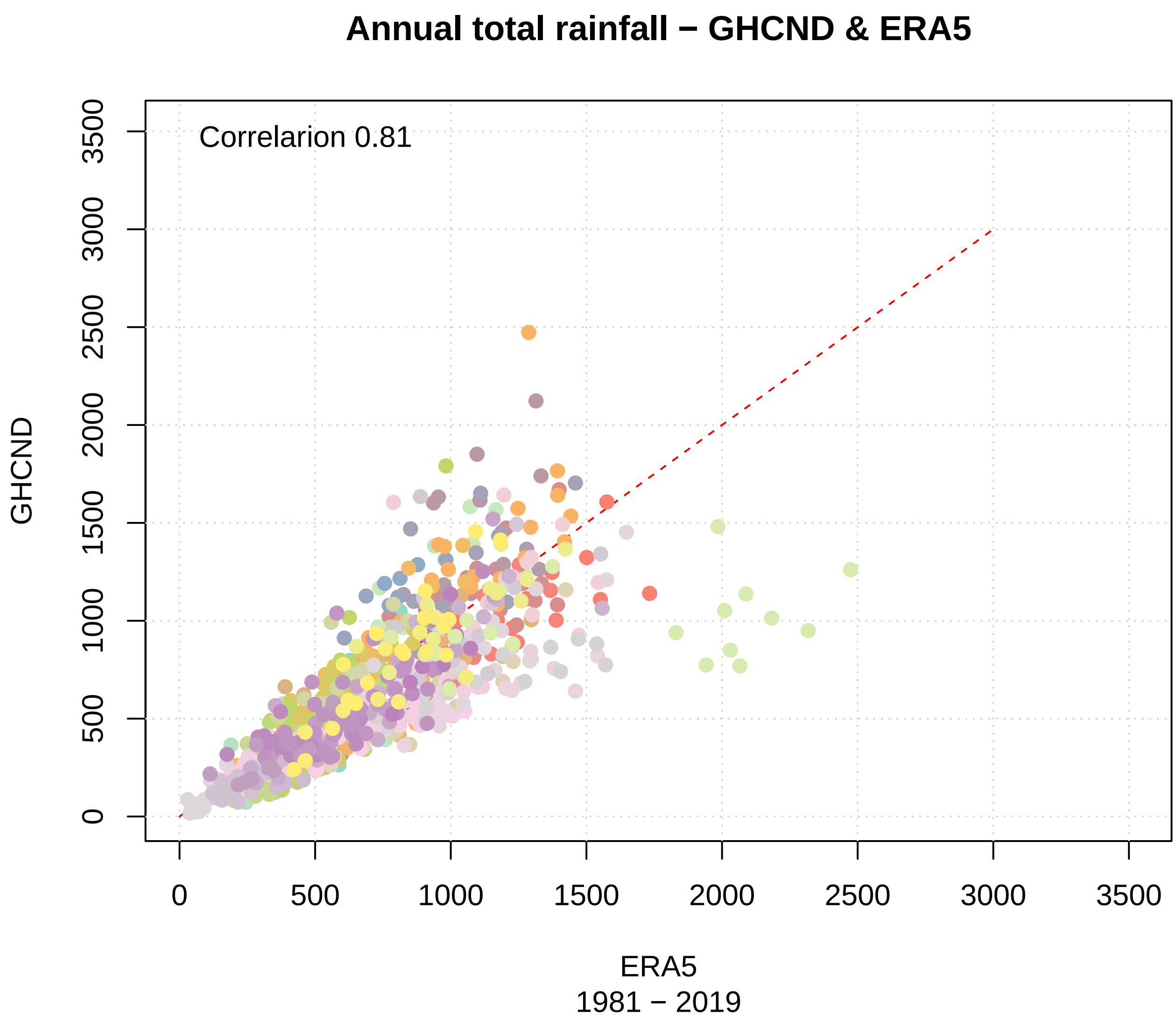

```
plot(lon(RR.chirps),lat(RR.chirps),pch=19,col=palette,bty='n',xlab='',ylab='')
data(geoborders)
lines(geoborders,col='grey')
```



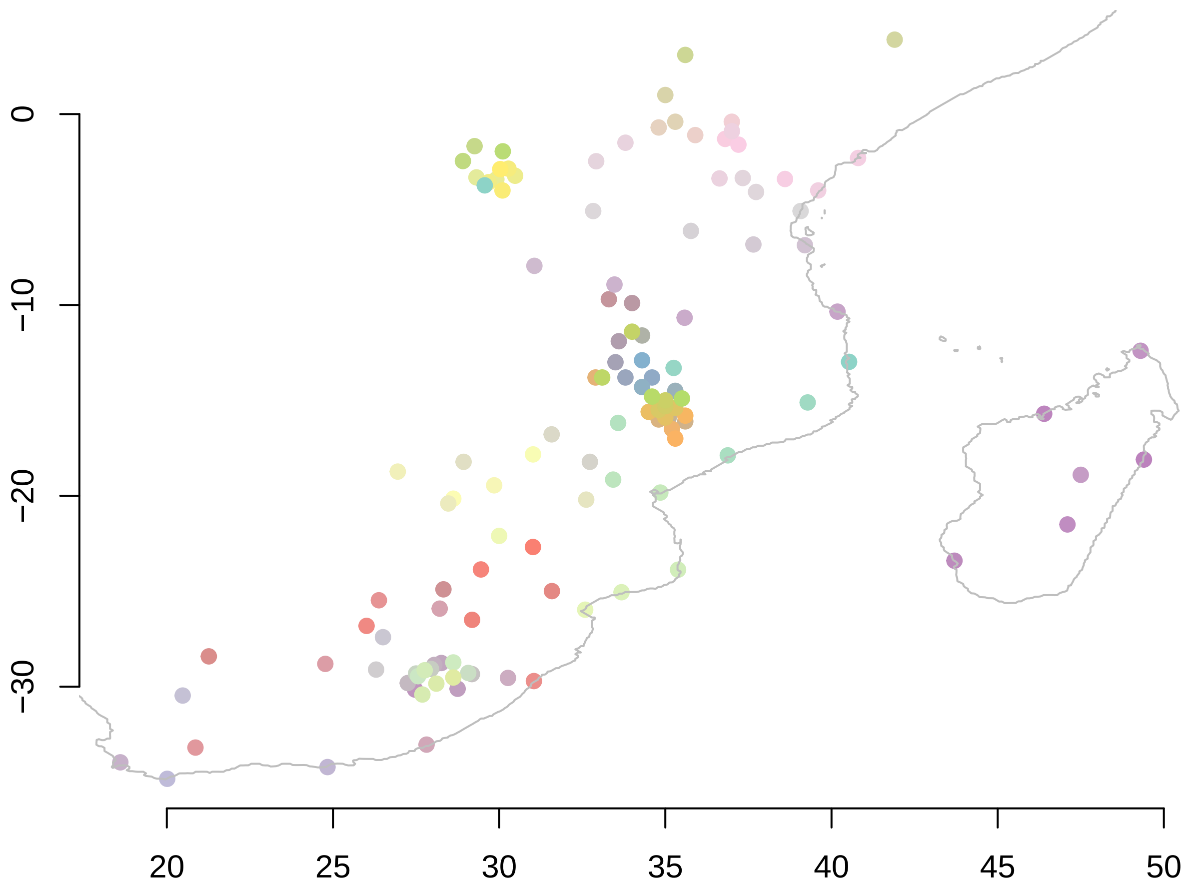

```
#text(lon(RR.chirps),lat(RR.chirps),loc(RR.chirps),cex=0.5)
```

There has a high correlation (0.94) between annual rainfall derived from the daily GHCND rain gauge data and CHIRPS, which may be related to how the CHIRPS data have been produced: CHIRPS incorporates our in-house climatology, CHPclim, 0.05° resolution satellite imagery, and in-situ station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring. (https://www.chc.ucsb.edu/data/chirps). Hence, the high correlation may possibly be due to the inclusion of the GHCND data in the production of CHIRPS. The comparison also shows that the data points don't cluster according to site of the GHCND measurements.

```
RR.ghcnd.era5 <- regrid(era5,is=Y.ghcnd)
XY <- merge(zoo(RR.ghcnd.era5),zoo(atY),all=FALSE)
ns <- dim(atY)[2]
palette <- colorRampPalette(brewer.pal(12, "Set3"))(ns)
r <- round(cor(c(coredata(XY)[,1:ns]),c(coredata(XY)[,(ns+1):(2*ns)]),use='complete.obs'),2)
plot(range(XY,na.rm=TRUE),range(XY,na.rm=TRUE),type='n',
     xlab='ERA5',ylab='GHCND',main='Annual total rainfall - GHCND & ERA5',
     sub=paste(range(year(RR.chirps)),collapse=' - '))
grid()
legend('topleft',paste('Correlarion',r),bty='n')

lines(c(0,3000),c(0,3000),lty=2,col='red')
for (is in 1:ns) points(coredata(XY)[,is],coredata(XY)[,is+ns],pch=19,col=palette[is])
```




```
plot(lon(RR.chirps),lat(RR.chirps),pch=19,col=palette,bty='n',xlab='',ylab='')
data(geoborders)
lines(geoborders,col='grey')
```




```
#text(lon(RR.chirps),lat(RR.chirps),loc(RR.chirps),cex=0.5)
```

The correlation between the annual rainfall from GHCND and ERA5 was high (0.81), but without any tendency of clustering around the rain gauge sites. It's interesting to note that the ERA5 data correlates better with the GHCND data than the shared rain gauge data.

```
## Need to make sure that the right stations are matched in the two data sets
locx <- tolower(loc(atX))
cntrx <- cntr(atX)
locy <- tolower(loc(atY))
cntry <- cntr(atY)
iX<- 1
atx <- NULL; aty <- NULL
for (loci in locx) {
  iy <- grep(loci,locy)
  jx<- match(loci,locx)
  cntr <- cntrx[match(loci,locx)]
  x1 <- subset(atX,is=list(loc=loci,cntr=cntr))
  if (length(iy)==1) {
    y1 <- subset(atY,is=list(loc=locy[iy],cntr=cntr))
    print(c(iX,iy))
    if (!is.null(cntr(y1))) {
      if (cntr(x1)==cntr(y1)) {
        print(c(loci,locx[iX],locy[iy],
                loc(x1),loc(y1),cntr(x1),cntr(y1)))
        if (is.null(atx)) {
          atx <- x1
```

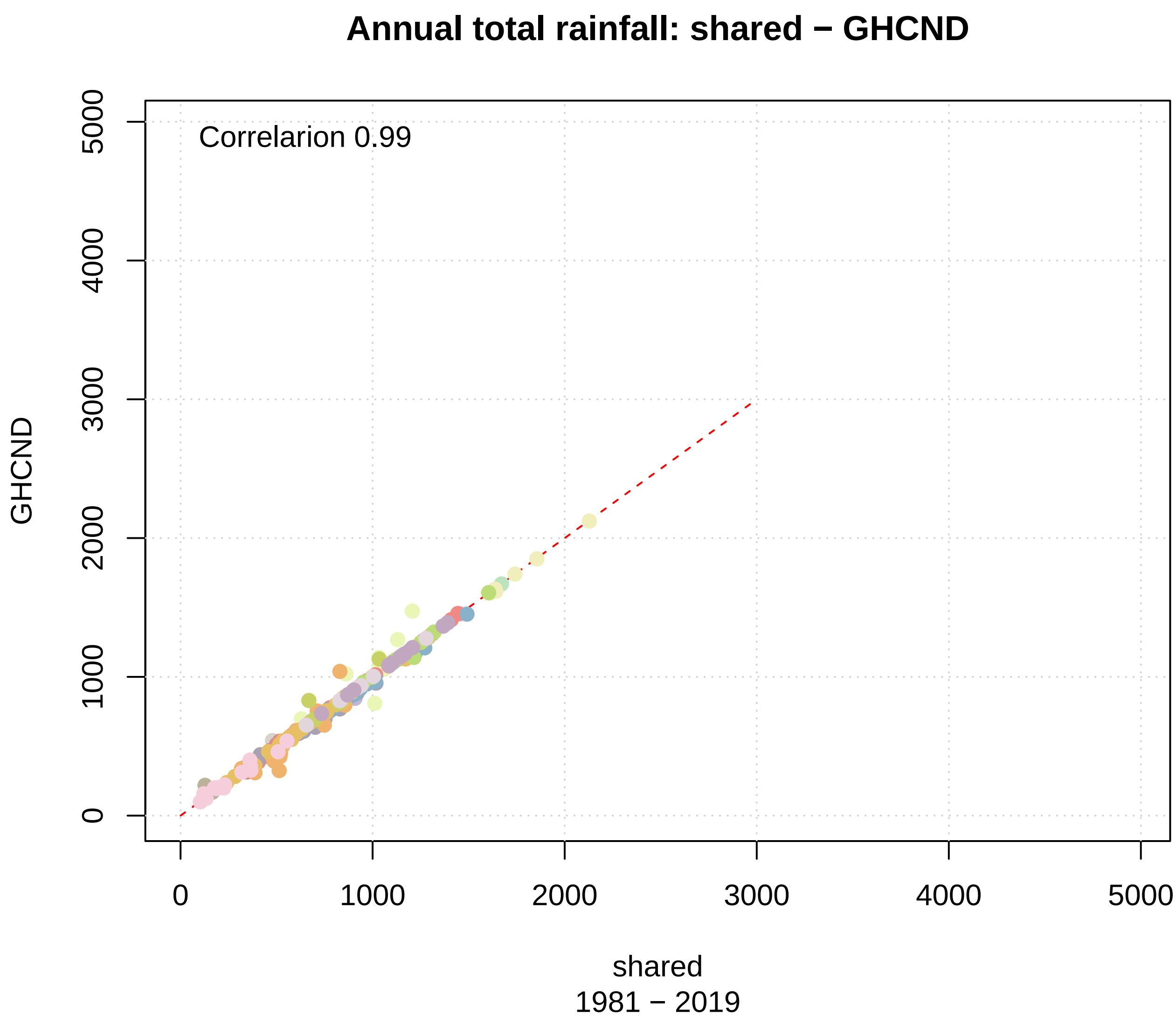
```
          aty <- y1
        } else  {
          atx <- combine.station(atx,x1)
          aty <- combine.station(aty,y1)
        }
      }
    }
  }
}
```

```
## [1]  1 38
## [1] "pemba"      "pemba"      "pemba, mz"  "Pemba"      "PEMBA, MZ"
## [6] "Mozambique" "Mozambique"
## [1]  1 39
## [1] "lichinga"     "pemba"        "lichinga, mz" "Lichinga"     "LICHINGA, MZ"
## [6] "Mozambique"   "Mozambique"
## [1]  1 40
## [1] "nampula"     "pemba"       "nampula, mz" "Nampula"     "NAMPULA, MZ"
## [6] "Mozambique"  "Mozambique"
## [1]  1 42
## [1] "quelimane"     "pemba"         "quelimane, mz" "Quelimane"
## [5] "QUELIMANE, MZ" "Mozambique"    "Mozambique"
## [1]  1 41
## [1] "tete"       "pemba"      "tete, mz"   "Tete"       "TETE, MZ"
## [6] "Mozambique" "Mozambique"
## [1]  1 43
## [1] "chimoio"     "pemba"       "chimoio, mz" "Chimoio"     "CHIMOIO, MZ"
## [6] "Mozambique"  "Mozambique"
## [1]  1 45
## [1] "inhambane"     "pemba"         "inhambane, mz" "Inhambane"
## [5] "INHAMBANE, MZ" "Mozambique"    "Mozambique"
## [1]   1 129
## [1] "chipinge"     "pemba"        "chipinge, zi" "Chipinge"     "CHIPINGE, ZI"
## [6] "Zimbabwe"     "Zimbabwe"
## [1]  1 62
## [1] "cape agulhas"     "pemba"            "cape agulhas, sf" "Cape Agulhas"
## [5] "CAPE AGULHAS, SF" "South Africa"     "South Africa"
## [1]  1 65
## [1] "cape st. francis"     "pemba"                "cape st. francis, sf"
## [4] "Cape St. Francis"     "CAPE ST. FRANCIS, SF" "South Africa"
## [7] "South Africa"
## [1]  1 91
## [1] "cedara"       "pemba"        "cedara, sf"   "Cedara"       "CEDARA, SF"
## [6] "South Africa" "South Africa"
## [1]  1 74
## [1] "laingsburg"     "pemba"          "laingsburg, sf" "Laingsburg"
## [5] "LAINGSBURG, SF" "South Africa"   "South Africa"
## [1]   1 103
## [1] "ottosdal"     "pemba"        "ottosdal, sf" "Ottosdal"     "OTTOSDAL, SF"
## [6] "South Africa" "South Africa"
## [1]  1 56
## [1] "skukuza"      "pemba"        "skukuza, sf"  "Skukuza"      "SKUKUZA, SF"
## [6] "South Africa" "South Africa"
## [1]  1 35
```

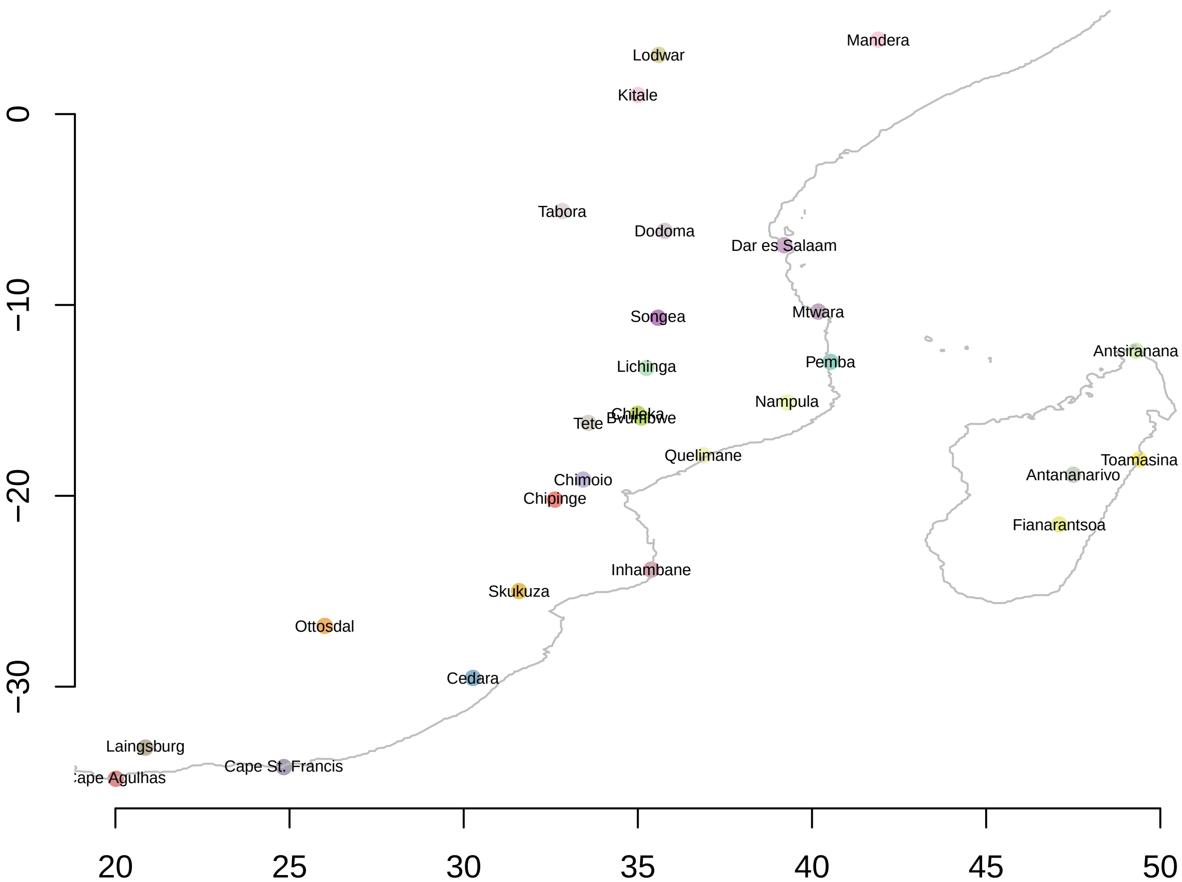
```
## [1] "chileka"     "pemba"       "chileka, mi" "Chileka"     "CHILEKA, MI"
## [6] "Malawi"      "Malawi"
## [1]  1 36
## [1] "bvumbwe"     "pemba"       "bvumbwe, mi" "Bvumbwe"     "BVUMBWE, MI"
## [6] "Malawi"      "Malawi"

## Warning in station.subset(x, it = it, is = is, verbose = verbose): Returning
## empty station set

## [1]   1 120
## [1] 1 1
## [1] "lodwar"     "pemba"      "lodwar, ke" "Lodwar"     "LODWAR, KE"
## [6] "Kenya"      "Kenya"
## [1] 1 3
## [1] "mandera"     "pemba"       "mandera, ke" "Mandera"     "MANDERA, KE"
## [6] "Kenya"       "Kenya"
## [1] 1 4
## [1] "kitale"     "pemba"      "kitale, ke" "Kitale"     "KITALE, KE"
## [6] "Kenya"      "Kenya"

## Warning in station.subset(x, it = it, is = is, verbose = verbose): Returning
## empty station set

## [1]   1 120
## [1]   1 121
## [1] "tabora"             "pemba"              "tabora airport, tz"
## [4] "Tabora"             "TABORA AIRPORT, TZ" "Tanzania"
## [7] "Tanzania"
## [1]   1 122
## [1] "dodoma"     "pemba"      "dodoma, tz" "Dodoma"     "DODOMA, TZ"
## [6] "Tanzania"   "Tanzania"
## [1]   1 123
## [1] "dar es salaam"             "pemba"
## [3] "dar es salaam airport, tz" "Dar es Salaam"
## [5] "DAR ES SALAAM AIRPORT, TZ" "Tanzania"
## [7] "Tanzania"
## [1]   1 124
## [1] "songea"     "pemba"      "songea, tz" "Songea"     "SONGEA, TZ"
## [6] "Tanzania"   "Tanzania"
## [1]   1 125
## [1] "mtwara"     "pemba"      "mtwara, tz" "Mtwara"     "MTWARA, TZ"
## [6] "Tanzania"   "Tanzania"
## [1]  1 13
## [1] "antananarivo"           "pemba"                  "antananarivo ivato, ma"
## [4] "Antananarivo"           "ANTANANARIVO IVATO, MA" "Madagascar"
## [7] "Madagascar"
## [1]  1 11
## [1] "antsiranana"     "pemba"           "antsiranana, ma" "Antsiranana"
## [5] "ANTSIRANANA, MA" "Madagascar"      "Madagascar"
## [1]  1 32
## [1] "fianarantsoa"     "pemba"            "fianarantsoa, ma" "Fianarantsoa"
## [5] "FIANARANTSOA, MA" "Madagascar"       "Madagascar"
## [1]  1 14
## [1] "toamasina"     "pemba"         "toamasina, ma" "Toamasina"
## [5] "TOAMASINA, MA" "Madagascar"    "Madagascar"
```

```
## Warning in station.subset(x, it = it, is = is, verbose = verbose): Returning
## empty station set

## Warning in station.subset(x, it = it, is = is, verbose = verbose): Returning
## empty station set

## Warning in station.subset(x, it = it, is = is, verbose = verbose): Returning
## empty station set
```

```
ns <- dim(atx)[2]
XY <- merge(zoo(atx),zoo(aty),all=FALSE)
palette <- colorRampPalette(brewer.pal(12, "Set3"))(ns)
r <- round(cor(c(coredata(XY)[,1:ns]),c(coredata(XY)[,(ns+1):(2*ns)]),use='complete.obs'),2)
plot(range(XY,na.rm=TRUE),range(XY,na.rm=TRUE),type='n',
     xlab='shared',ylab='GHCND',main='Annual total rainfall: shared - GHCND',
     sub=paste(range(year(RR.chirps)),collapse=' - '))
grid()
legend('topleft',paste('Correlarion',r),bty='n')

lines(c(0,3000),c(0,3000),lty=2,col='red')
for (is in 1:ns) points(coredata(XY)[,is],coredata(XY)[,is+ns],pch=19,col=palette[is])
```




```r
plot(lon(atx),lat(atx),pch=19,col=palette,bty='n',xlab='',ylab='')
data(geoborders)
lines(geoborders,col='grey')
text(lon(atx),lat(atx),loc(atx),cex=0.5)
```



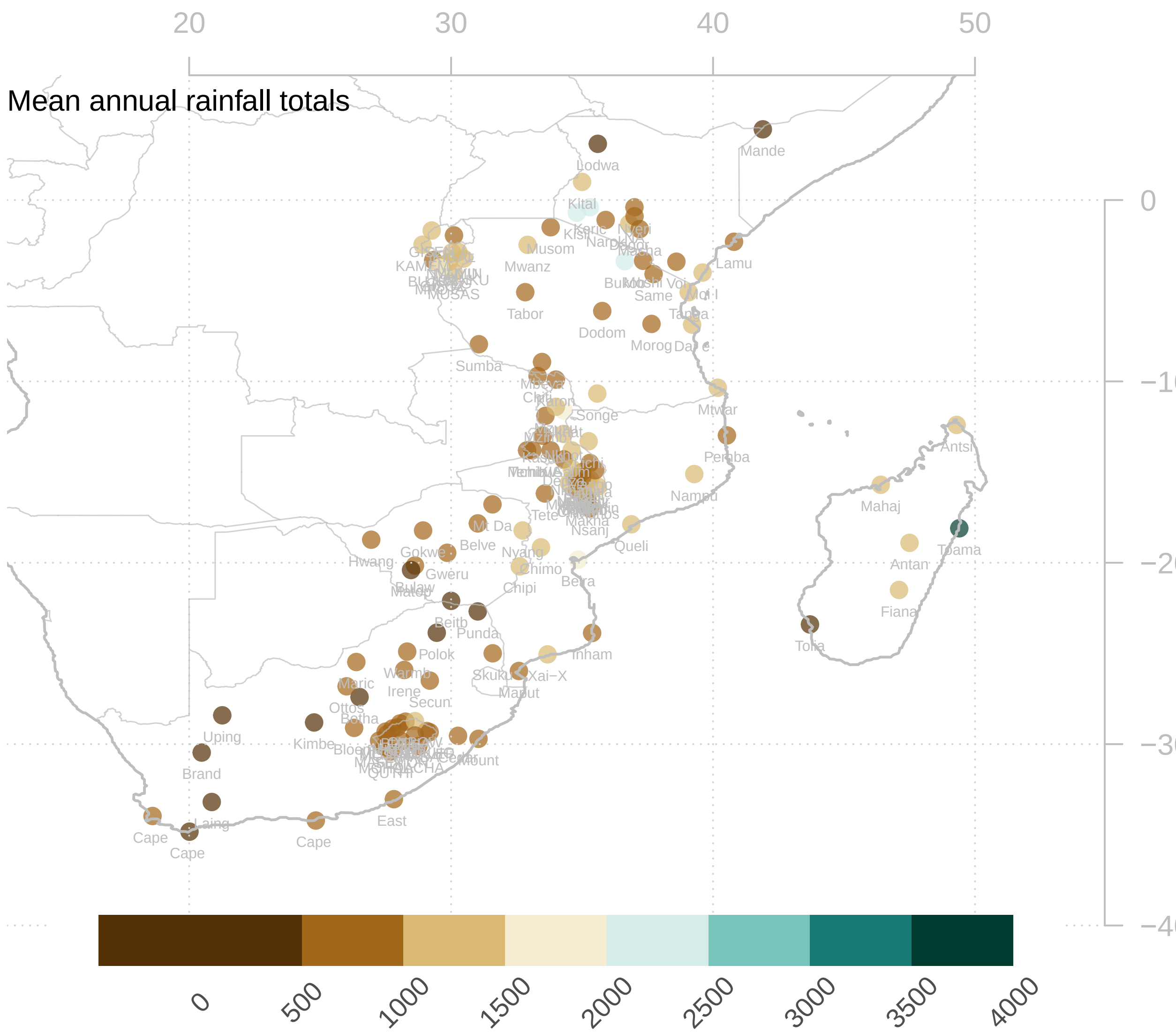
The correlation between corresponding stations from GHCND and the shared rain gauge data is high (0.99), which suggests that the differences reported above mainly are connected to data from other non-overlapping sites.

4 Figures for the manuscript

The following part generates the figures presented in the manuscript.

4.1 Figure 2.1

```
map(atX,FUN='mean',colbar=list(pal='precip.ipcc'),add.text=TRUE,cex.lab=0.5,cex=1.25,
    main='Mean annual rainfall totals',border=TRUE,ylim=c(-42,5))
```



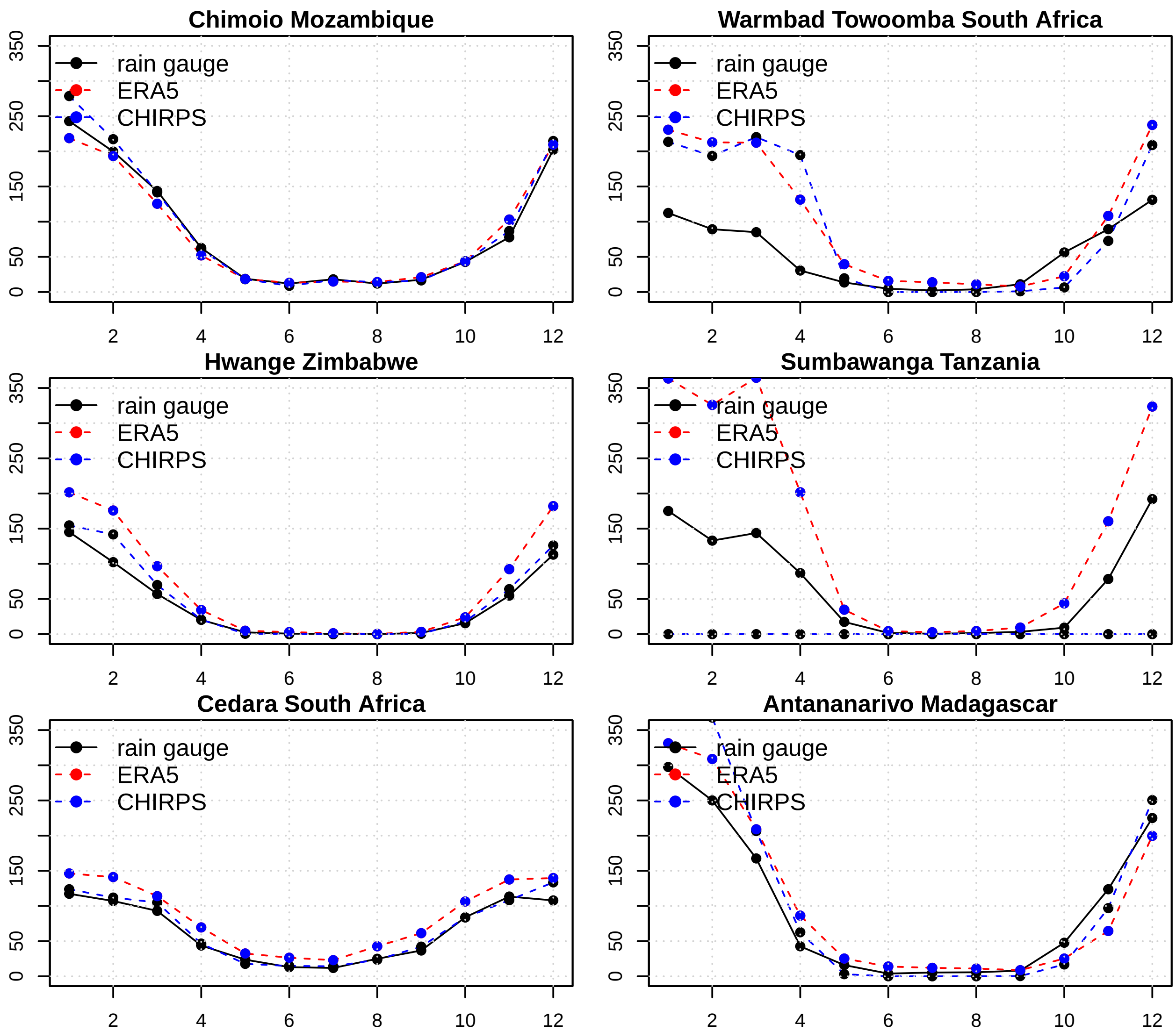
4.2 Figure 3.1

```
select <- grep('Mtwara|Chimoio|Hwange|Cedara|Chitipa|Mahajanga',loc(X),ignore.case = TRUE)
par(mfcol=c(3,2),mar=1.5*c(1.25,1.25,1,1),cex.labels=1.2)
```

```
## Warning in par(mfcol = c(3, 2), mar = 1.5 * c(1.25, 1.25, 1, 1), cex.labels =
## 1.2): "cex.labels" is not a graphical parameter
```

```
for (i in select) {
  it <- matchdate(subset(x.mon,is=i),subset(X,is=i))
  x1 <- matchdate(as.monthly(subset(X,is=i),FUN='sum'),x.mon)
  y1 <- matchdate(subset(x.mon,is=i),x1)
  z1 <- matchdate(as.monthly(subset(rr.chirps,is=i),FUN='sum',na.rm=TRUE),x.mon)
  coredata(y1)[is.na(coredata(x1))] <- NA
  x1 <- aggregate(zoo(x1),FUN='mean',by=month,na.rm=TRUE)
  y1 <- aggregate(zoo(y1),FUN='mean',by=month,na.rm=TRUE) # ERA5 is in units of m/day
  z1 <- aggregate(zoo(z1),FUN='mean',by=month,na.rm=TRUE)
  rmse[i] <- RMSE(coredata(x1),coredata(y1))
  rmsep[i] <- 100*rmse[i]/sum(x1)
```

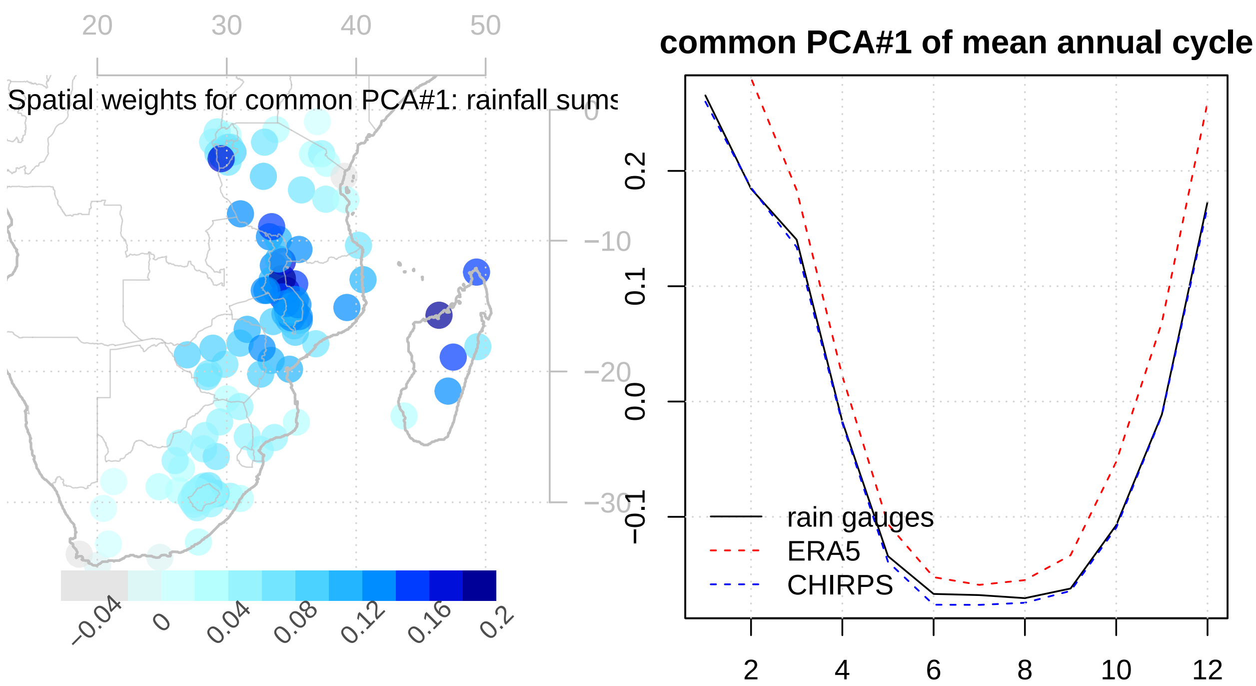
```
  plot(merge(x1,y1,z1),plot.type='single',col=c('black','red','blue'),lty=c(1,2,2),
       main=paste(loc(subset(X,is=i)),cntr(subset(X,is=i))),
       ylab='mm/month',xlab='Canendar month',ylim=c(0,350))
  points(x1,pch=19); points(y1,col='red',pch=19)
  points(z1,pch=19); points(y1,col='blue',pch=19)
  grid()
  legend('topleft',c('rain gauge','ERA5','CHIRPS'),col=c('black','red','blue'),
         pch=19, lty=c(1,2,2),bty='n',cex=1.2)
}
```



4.3 Figure 3.2

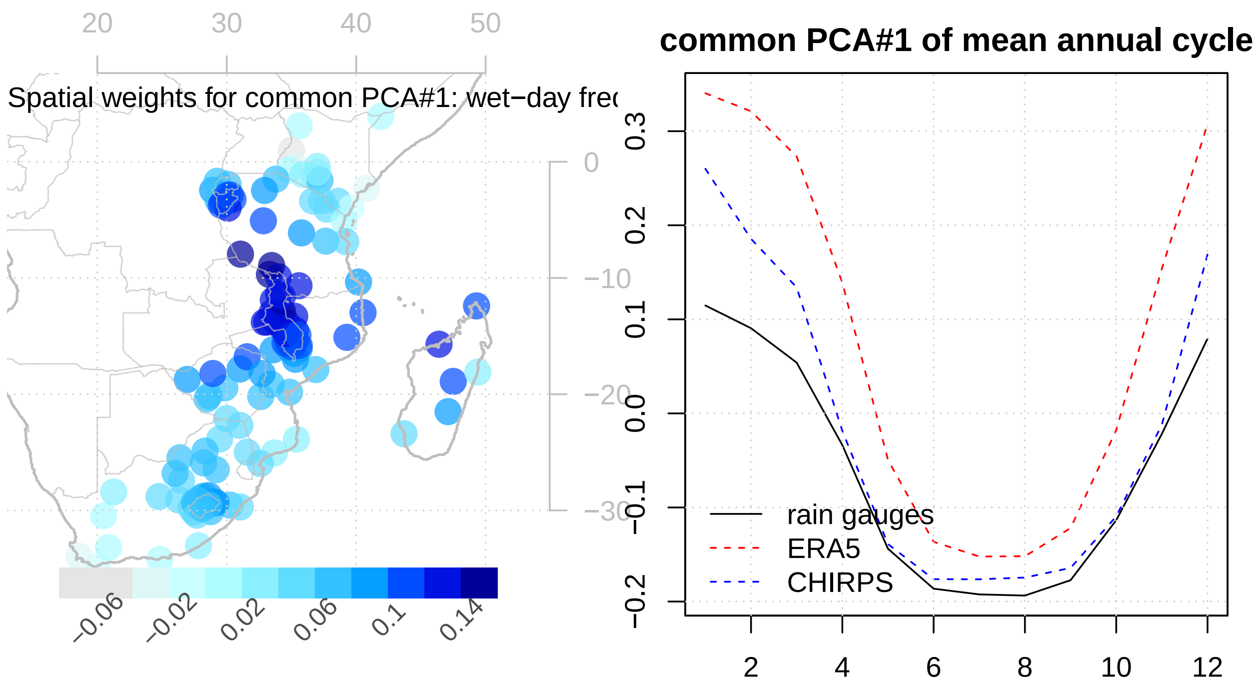
```
par(mfcol=c(1,2),mar=1.5*c(1,1,1,1))
map(pca.both,main='Spatial weights for common PCA#1: rainfall sums',
    new=FALSE,border=TRUE)
plot(zoo(pca.both[1:12,1]),main='common PCA#1 of mean annual cycle',
     ylab='weight',xlab='Clendar month')
lines(zoo(coredata(pca.both)[13:24,1],order.by=1:12),col='red',lty=2)
```

```
lines(zoo(coredata(pca.both)[25:36,1],order.by=1:12),col='blue',lty=2)
grid()
legend('bottomleft',c('rain gauges','ERA5','CHIRPS'),lty=c(1,2,2),
       col=c('black','red','blue'),bty='n')
```



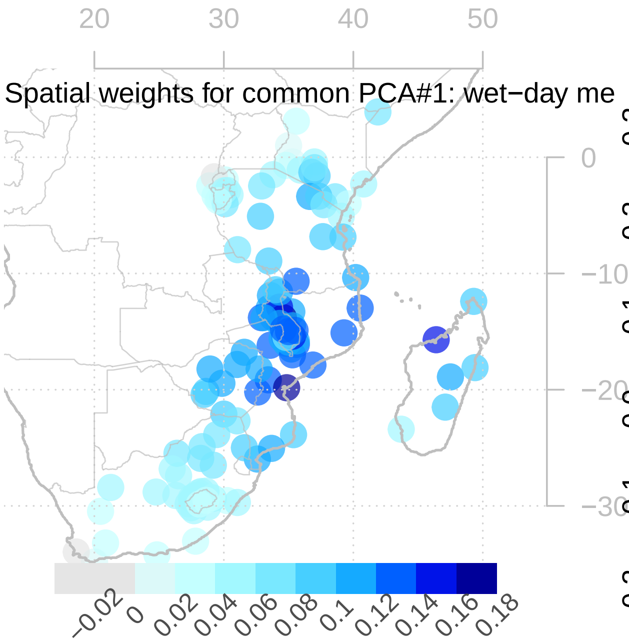
4.4 Figure 3.3

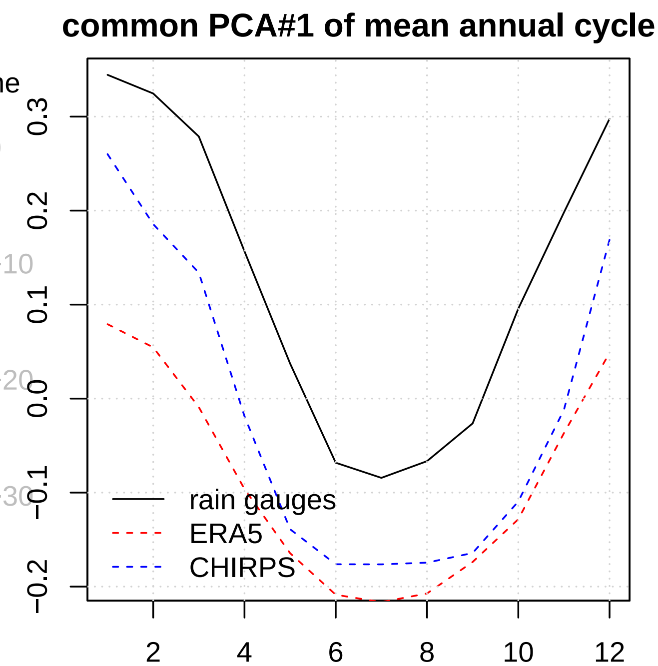
```
par(mfcol=c(1,2),mar=1.5*c(1,1,1,1))
map(fw.pca.both,main='Spatial weights for common PCA#1: wet-day frequency',
    new=FALSE,border=TRUE)
plot(zoo(fw.pca.both[1:12,1]),main='common PCA#1 of mean annual cycle',
     ylab='weight',xlab='Clendar month',ylim=range(fw.pca.both[,1]))
lines(zoo(coredata(fw.pca.both)[13:24,1],order.by=1:12),col='red',lty=2)
lines(zoo(coredata(pca.both)[25:36,1],order.by=1:12),col='blue',lty=2)
grid()
legend('bottomleft',c('rain gauges','ERA5','CHIRPS'),lty=c(1,2,2),
       col=c('black','red','blue'),bty='n')
```



4.5 Figure 3.4

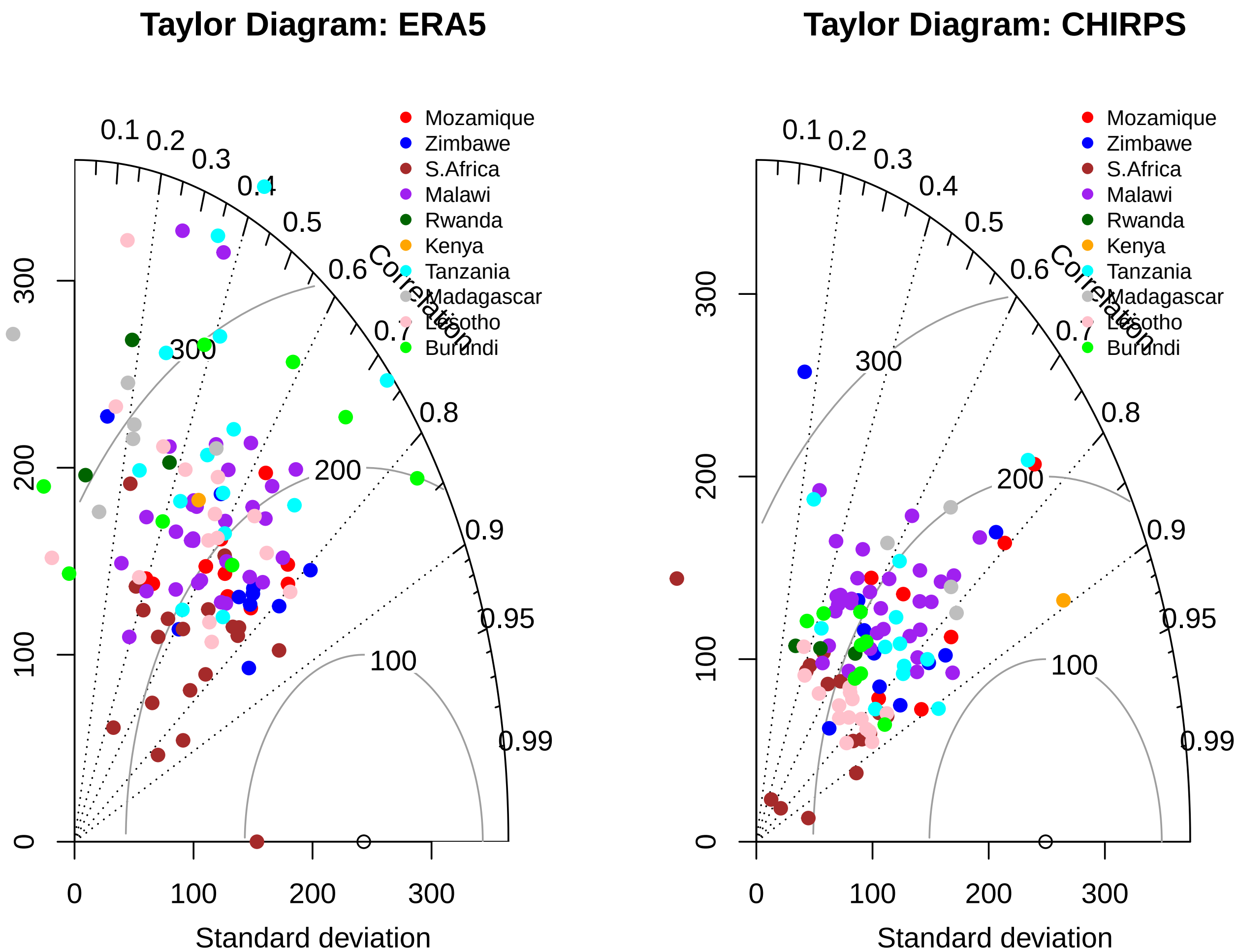
```
par(mfcol=c(1,2),mar=1.5*c(1,1,1,1))
map(mu.pca.both,main='Spatial weights for common PCA#1: wet-day mean precipitation',
    new=FALSE,border=TRUE)
plot(zoo(mu.pca.both[1:12,1]),main='common PCA#1 of mean annual cycle',
     ylab='weight',xlab='Clendar month',ylim=range(fw.pca.both[,1]))
lines(zoo(coredata(mu.pca.both)[13:24,1],order.by=1:12),col='red',lty=2)
lines(zoo(coredata(pca.both)[25:36,1],order.by=1:12),col='blue',lty=2)
grid()
legend('bottomleft',c('rain gauges','ERA5','CHIRPS'),lty=c(1,2,2),
       col=c('black','red','blue'),bty='n')
```





4.6 Figure 3.5

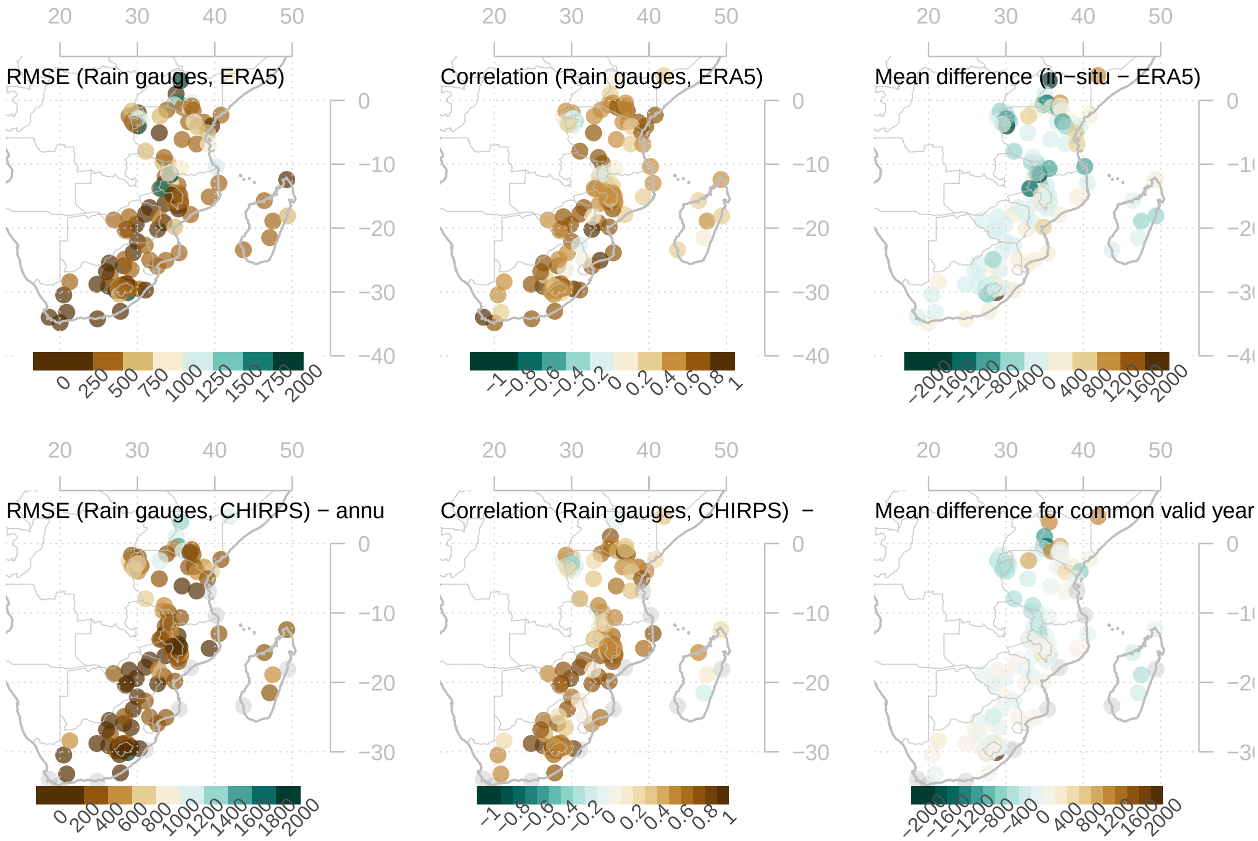
```
par(mfcol=c(1,2),mar=0.5*c(1,1,1,1))
taylor.diagram(ref[,ix],mod[,ix],col=cols[ix],main='Taylor Diagram: ERA5')
for (i in 1:dim(ref)[2]) taylor.diagram(ref[,i],mod[,i],col=cols[i],add=TRUE)
legend('topright',c('Mozamique','Zimbawe','S.Africa','Malawi','Rwanda','Kenya',
                    'Tanzania','Madagascar','Lesotho','Burundi'),cex=0.75,
       col=c('red','blue','brown','purple','darkgreen','orange','cyan','grey',
             'pink','green'),bty='n',pch=19)

taylor.diagram(ref2[,ix2],mod2[,ix2],col=cols2[ix2],main='Taylor Diagram: CHIRPS')
for (i in 1:dim(ref2)[2]) taylor.diagram(ref2[,i],mod2[,i],col=cols2[i],add=TRUE)
legend('topright',c('Mozamique','Zimbawe','S.Africa','Malawi','Rwanda','Kenya',
                    'Tanzania','Madagascar','Lesotho','Burundi'),cex=0.75,
       col=c('red','blue','brown','purple','darkgreen','orange','cyan','grey',
             'pink','green'),bty='n',pch=19)
```



4.7 Figure 3.6

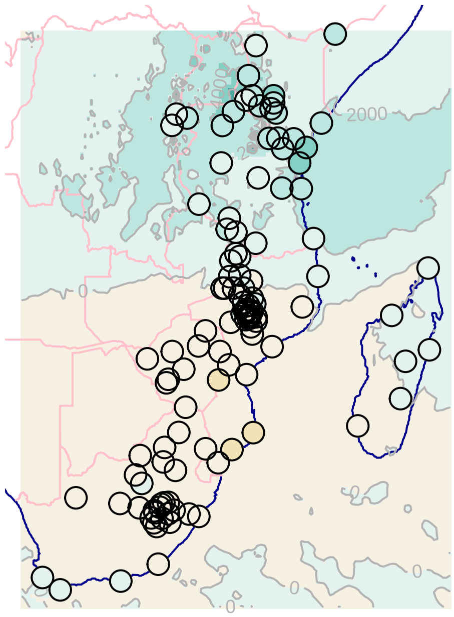
Larger fonts. Also for the CHIRPS data.

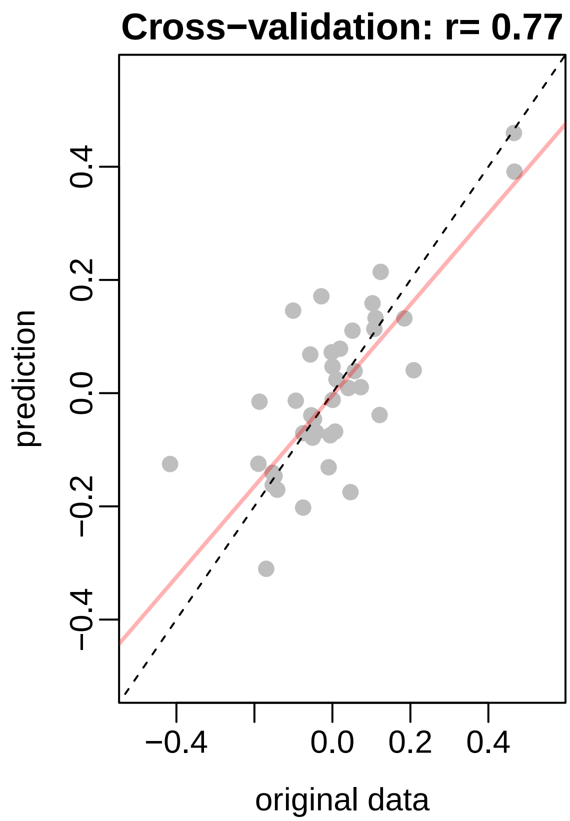
```
par(mfrow=c(2,3),mar=c(1,1,1,1),cex.axis=1.25, cex=0.95)
map(atX,FUN='rmse',add.text = FALSE,border=TRUE,
    colbar=list(breaks=seq(0,2000,by=250),pal='precip.ipcc'),ylim=c(-42,5),
    main='RMSE (Rain gauges, ERA5)',cex=1.5,new=FALSE)
map(atX,FUN='cor',add.text = FALSE,border=TRUE,ylim=c(-42,5),
    colbar=list(breaks=seq(-1,1,length=11), pal='precip.ipcc',rev=TRUE),
    main='Correlation (Rain gauges, ERA5)',cex=1.5,new=FALSE)
map(atX,FUN='offset',add.text = FALSE,border=TRUE,ylim=c(-42,5),
    colbar=list(breaks=seq(-2000,2000,length=11),pal='precip.ipcc',rev=TRUE),
    main='Mean difference (in-situ - ERA5)',cex=1.5,new=FALSE)

map(atX,FUN='rmse.chirps',add.text = FALSE,colbar=list(breaks=seq(0,2000,length=11),pal='precip.ipcc'),
    main='RMSE (Rain gauges, CHIRPS) - annual rainfall totals',cex=1.5,border=TRUE)
map(atX,FUN='cor.chirps',add.text = FALSE,
    colbar=list(breaks=seq(-1,1,length=21), pal='precip.ipcc',rev=TRUE),border=TRUE,
    main='Correlation (Rain gauges, CHIRPS)  - annual rainfall totals',cex=1.5)
map(atX,FUN='offset.chirps',add.text = FALSE, border=TRUE,
    colbar=list(breaks=seq(-2000,2000,length=21), pal='precip.ipcc',rev=TRUE),
    main='Mean difference for common valid years (Rain gauges - CHIRPS)',cex=1.5)
```

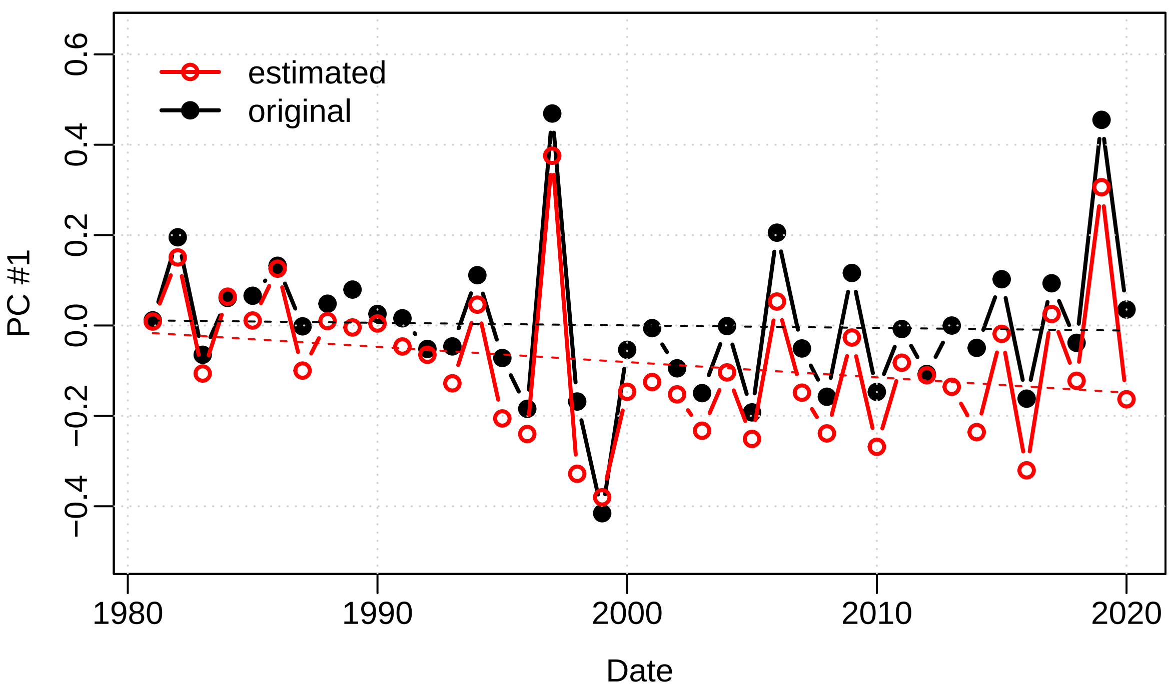


4.8 Figure 3.7

```
par(mfcol=c(1,1),mar=c(1,1,1,1))
## Show the results for the leading PCA
plot(ds,new=FALSE)
```


2000


Cross−validation: r= 0.77
prediction
original data
−0.4
−0.2
0.0
0.2
0.4


estimated
original
PC #1
Date
1980
1990
2000
2010
2020
−0.4
−0.2
0.0
0.2
0.4
0.6

```
## NULL
```

4.9 Figure 3.8

```
ptrend <- function(x,na.rm=TRUE) 100*trend.coef(x)/mean(x,na.rm=na.rm)
par(mfrow=c(1,3),cex=1.15,cex.main=1.15)
atX <- subset(atX,it=c(1981,2019))
print(range(index(atX)))
```

```
## [1] 1981 2019
```

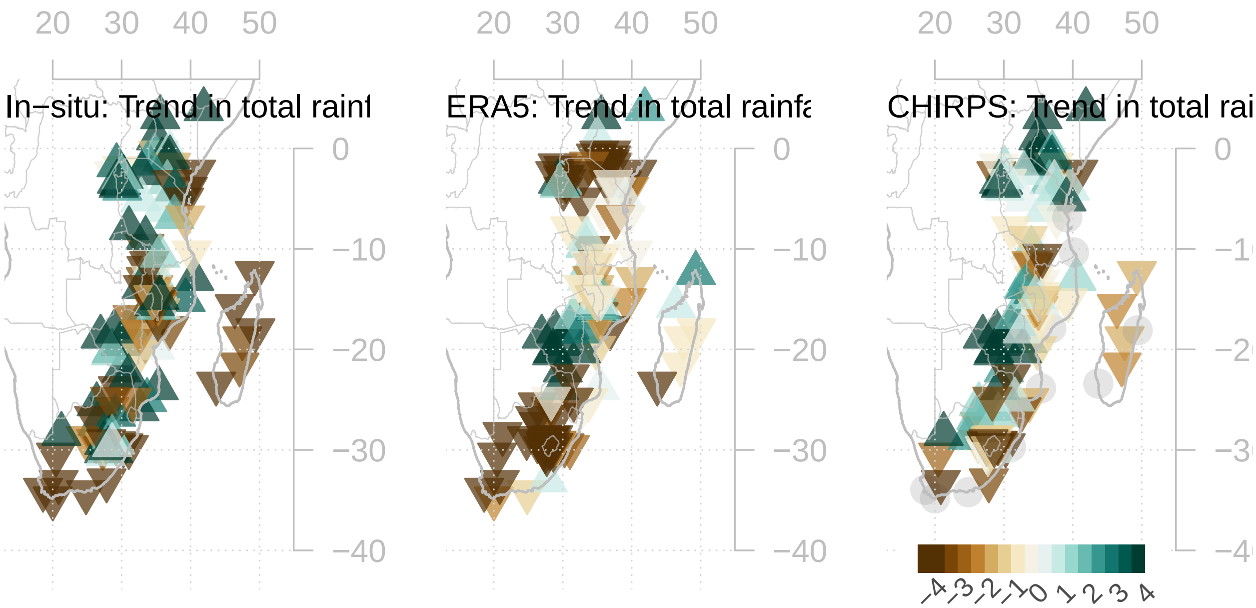
```
t.obs <- map(atX,FUN='ptrend',new=FALSE,border=TRUE,ylim=c(-42,5),
    main=paste('In-situ: Trend in total rainfall (%)',min(index(atX)),'-',max(index(atX))),
    colbar=list(show=FALSE,pal='precip.ipcc',breaks=seq(-4,4,by=0.5)),cex.lab=1)
## ERA5 - interpolate to same locations as rain gauges (again)
x <- 365.25*1000*annual(regrid(era5,is=X),FUN='mean',start=year.start)
## Make sure that the data cover the same years
x <- matchdate(x,atX)
print(range(index(x)))
```

```
## [1] 1981 2019
```

```
for (is in 1:dim(x)[2]) {
  zz <- subset(atX,is=is)
  isnv <- !is.finite(coredata(zz))
  coredata(x)[isnv,is] <- NA
  #print(paste(loc(zz),sum(isnv)))
}
## ERA5
t.era <- map(x,FUN='ptrend',new=FALSE,border=TRUE,ylim=c(-42,5),
    main=paste('ERA5: Trend in total rainfall (%)',min(index(atX)),'-',max(index(atX))),
    colbar=list(show=FALSE,pal='precip.ipcc',breaks=seq(-4,4,by=0.5)),cex.lab=1)

## CHIRPS
RR.chirps <- matchdate(RR.chirps,atX)
print(range(index(RR.chirps)))
```

```
## [1] 1981 2019
```

```
for (is in 1:dim(RR.chirps)[2]) {
  zz <- subset(atX,is=is)
  isnv <- !is.finite(coredata(zz))
  coredata(RR.chirps)[isnv,is] <- NA
  #print(paste(loc(zz),sum(isnv)))
}
t.chi <- map(RR.chirps,FUN='ptrend',new=FALSE,border=TRUE,ylim=c(-42,5),
    main=paste('CHIRPS: Trend in total rainfall (%)',min(index(atX)),'-',max(index(atX))),
    colbar=list(pal='precip.ipcc',breaks=seq(-4,4,by=0.5)),cex.lab=1)
```




```
## Correlation in trend estimates
## In-situ rain gauge v.s. ERA5
cor.test(coredata(t.obs),coredata(t.era))
```

```
##
##  Pearson's product-moment correlation
##
## data:  coredata(t.obs) and coredata(t.era)
## t = 1.4807, df = 128, p-value = 0.1411
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.04338331  0.29535850
## sample estimates:
##       cor
## 0.1297722
```

```
## In-situ rain gauge v.s. CHIRPS
i1 <- is.element(loc(atX),loc(RR.chirps))
i2 <- is.finite(t.chi)
cor.test(coredata(t.obs)[i1][i2],coredata(t.chi)[i2])
```

```
##
##  Pearson's product-moment correlation
##
## data:  coredata(t.obs)[i1][i2] and coredata(t.chi)[i2]
## t = 5.6482, df = 118, p-value = 1.14e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3075086 0.5916422
## sample estimates:
##       cor
## 0.4613226
```

```
## ERA5 v.s. CHIRPS
cor.test(coredata(t.era)[i1][i2],coredata(t.chi)[i2])
```

```
##
```

```
##  Pearson's product-moment correlation
##
## data:  coredata(t.era)[i1][i2] and coredata(t.chi)[i2]
## t = 6.3069, df = 118, p-value = 5.134e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3547972 0.6250915
## sample estimates:
##       cor
## 0.5021074
```

4.10 Figure 3.9

```
## Get the NINO3.4 index online
nino3.4 <- NINO3.4()
par(mfrow=c(1,1),mar=c(1,1,1,1))
calibrate <- data.frame(x=coredata(subset(annual(nino3.4,start=year.start),
                                          it=c(1981,2020))),
                        y=coredata(pca))
```

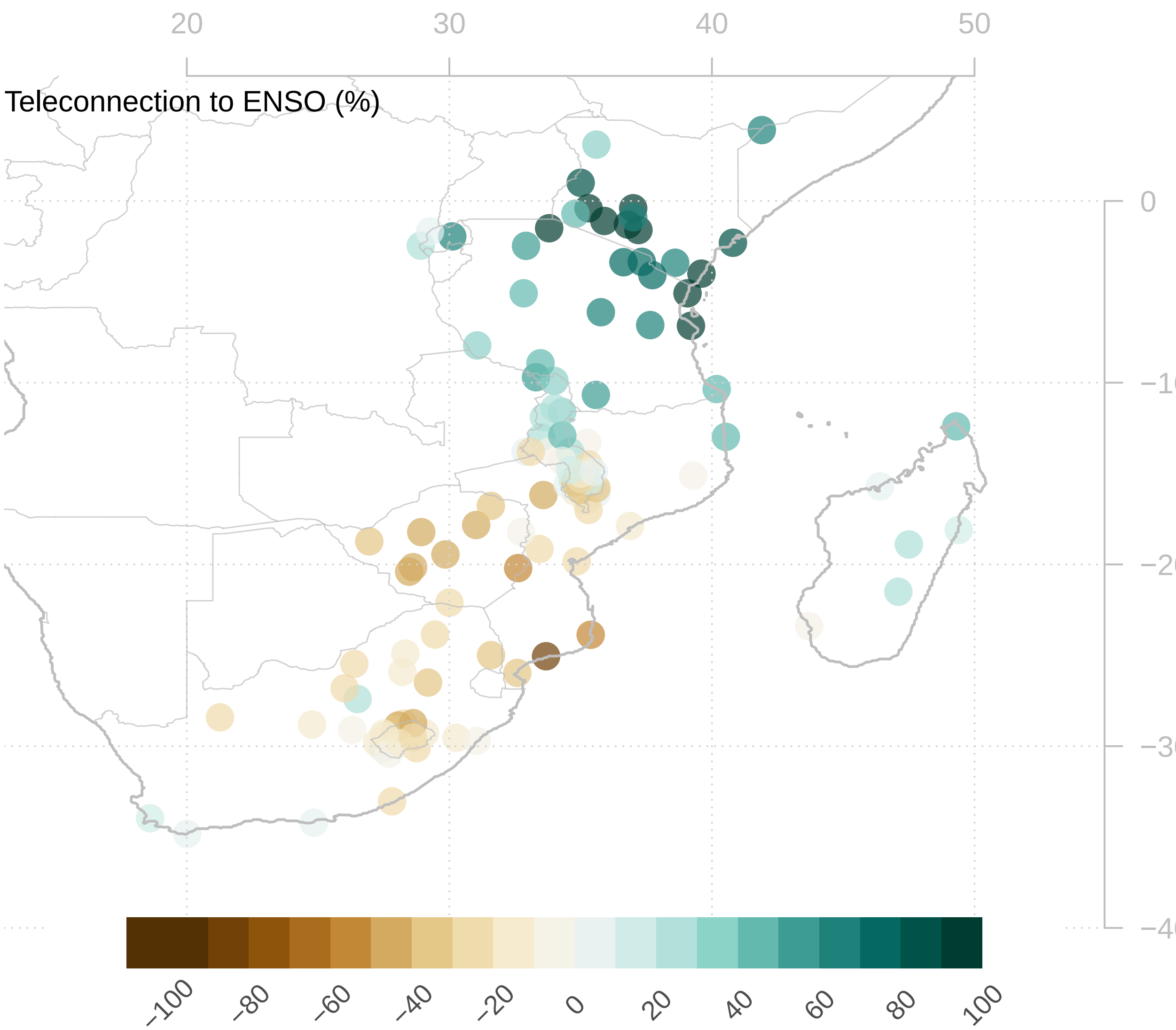
```
## Warning in sqrt(coredata(n) - 1): NaNs produced
```

```
## Fit the four leading modes which have eigenvalues above noise level
## seen in plot(pca) bove
fit <- lm(x ~ y.X.1 + y.X.2 + y.X.3 + y.X.14, data=calibrate)
print(summary(fit))
```

```
##
## Call:
## lm(formula = x ~ y.X.1 + y.X.2 + y.X.3 + y.X.14, data = calibrate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32357 -0.28519  0.06022  0.36537  0.83464
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.04466    0.08044 336.229  < 2e-16 ***
## y.X.1        1.66550    0.50855   3.275  0.00239 **
## y.X.2       -2.29547    0.50855  -4.514  6.9e-05 ***
## y.X.3       -0.49737    0.50855  -0.978  0.33478
## y.X.14      -0.07242    0.50872  -0.142  0.88762
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5085 on 35 degrees of freedom
## Multiple R-squared:  0.4782, Adjusted R-squared:  0.4186
## F-statistic: 8.019 on 4 and 35 DF,  p-value: 0.0001066
```

```
## Need to invert the weights, since the LHS of the equation holds the local rain data
weights <- 1/fit$coefficients
## Set all non-significant weights to zero
weights[-2] <- 0
pca.enso <- subset(pca,ip=1:4)
attr(pca.enso,'mean')[] <- 0
```

```
for (i in 1:4) pca.enso[,i] <- weights[1+i]
teleconnection <- pca2station(pca.enso)
teleconnection <- 100*teleconnection/attr(pca,'mean')
map(teleconnection,FUN='mean',new=FALSE,border=TRUE, main='Teleconnection to ENSO (%)',
    colbar=list(pal='precip.ipcc',breaks=seq(-100,100,by=10)),ylim=c(-42,5))
```

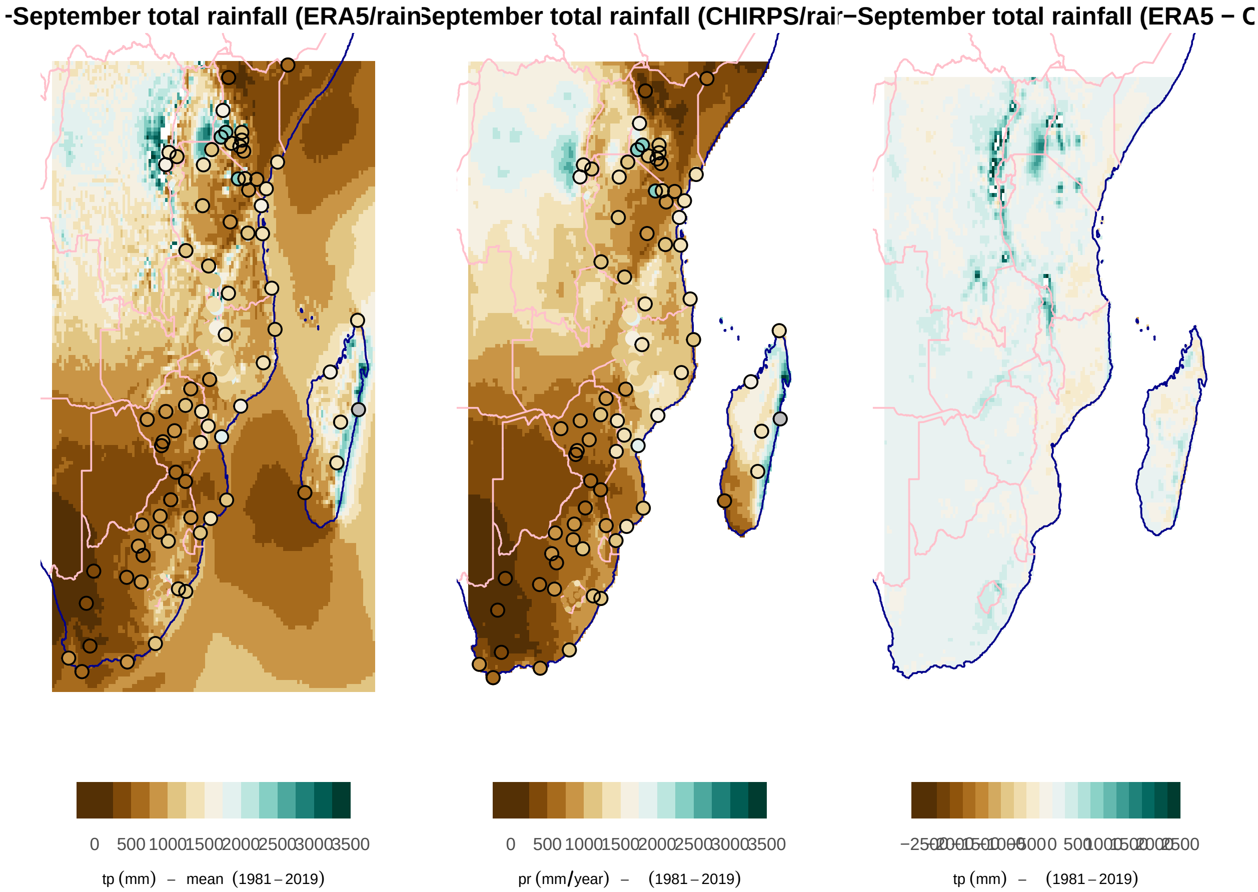


4.11 ## Figure 3.10

Produce map of ERA5 with rain gauge values superimposed

```
par(mfrow=c(1,3),mar=1.5*c(1.25,1,1,1))
## Map showing GCM mean precipitation
atX <- subset(atX,it=c(1979,2020))
atX <- pcafill(atX)
## Express ERA5 as annual totals for year starting in October and with units mm
ERA5 <- matchdate(annual(era5,FUN='mean',start=year.start),atX)
attr(ERA5,'units') <- 'mm'
mY <- colMeans(atX,na.rm=TRUE)
breaks <- seq(0,3500,by=250)
```

```r
era5map <- map(ERA5,FUN='mean',colbar=list(breaks=breaks,pal='precip.ipcc'),
               type='fill',new=FALSE,showaxis=FALSE,
               main='October-September total rainfall (ERA5/rain gauges)')
## Add rain gauge information
col <- rep('grey',length(mY))
cex <- rep(1.5,,length(mY))
cex[grep('Burundi|Lesotho',cntr(atX))] <- 0.5
for (i in 1:length(mY)) col[i] <- attr(era5map,'colbar')$col[breaks >= mY[i]][1]
col[is.na(col)] <- 'grey'
points(lon(atX),lat(atX),pch=19,cex=cex,col=col)
tight <- grep('Burundi|Lesotho|Malawi',cntr(atX))
points(lon(atX)[-tight],lat(atX)[-tight],pch=21,cex=cex,col='black')

## CHIRPS
map(chirps5map,colbar=list(breaks=breaks,pal='precip.ipcc'),
               type='fill',new=FALSE,showaxis=FALSE,
               main='October-September total rainfall (CHIRPS/rain gauges)')
points(lon(atX),lat(atX),pch=19,cex=cex,col=col)
tight <- grep('Burundi|Lesotho|Malawi',cntr(atX))
points(lon(atX)[-tight],lat(atX)[-tight],pch=21,cex=cex,col='black')

## ERA5 - CHIRPS
era5map <- regrid(era5map,is=chirps5map)
diffmap <- map(era5map - chirps5map,FUN='mean',colbar=list(breaks=seq(-2500,2500,by=250),pal='precip.ipc
               type='fill',new=FALSE,showaxis=FALSE,
               main='October-September total rainfall (ERA5 - CHIRPS)')
```



References

Benestad, R.E., Mezghani, A., Lutz, J., Dobler, A., Parding, K.M., Landgren, O.A., 2023. Various ways of using empirical orthogonal functions for climate model evaluation. Geoscientific Model Development 16, 2899–2913. https://doi.org/10.5194/gmd-16-2899-2023

Benestad, R.E., Parding, K.M., Dobler, A., 2024. Downscaling the probability of heavy rainfall over the Nordic countries. https://doi.org/10.5194/egusphere-2024-1463

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thepaut, J., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146, 1999–2049. https://doi.org/10.1002/qj.3803

Hersbach, H., Dee, D., 2016. ERA5 reanalysis is in production. ({ECMWF} {Newsletter} No. 147). ECMWF, Reading, United Kingdom.

Wilks, D.S., 2006. Statistical methods in the atmospheric sciences, 2nd ed. ed, International geophysics series. Academic Press, Amsterdam ; Boston.