Supporting material for ‘A simple equation to study changes in rainfall statistics’

Rasmus Benestad
June 12, 2019

Supporting material

This document provides additional analysis for the study on the probability of heavy daily rainfall to confirm its veracity and the computer code used to generate the results for the sake of transparency and replicability. The data presented here can also be explored with an on-line tool at https://ocdp.met.no: they are daily rain gauge records from the Norwegian Meteorological Institute’s climate archive, the ECA&D data from https://www.ecad.eu/dailydata/predefinedseries.php, daily Global Historical Climate Network (GHCND) accessed through the open-source esd-package https://github.com/metno/esd (ftp://ftp.ncdc.noaa.gov/pub/data/ghcn), and from CLARIS LPB for data from Argentine.

R Markdown

This is an R Markdown document and is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

```r
## The following lines will generate a pure R-script from the code chunks in this document
## (e.g. for generating the individual PDF-files holding the figures):
## Extract just the R-code for the generation of the graphics used for the figures seperately.
## Check if you need to get the devtools-package:
install.knitr <- ('knitr' %in% rownames(installed.packages())) == FALSE

if (install.knitr) {
  print('Need to install the knitr package')
  ## You need online access.
  install.packages('knitr')
}
library(knitr)
purl('rainequation.Rmd', output='rainequation.R')
```

Data processing set-up

Activate the esd package and set some parameters for guiding the processing.

```r
## The rain equation paper: rasmus.benestad@met.no 2019-01-02
## Check netCDF files with daily rain gauge data for Europe, North America and Australia:
## Interval with best coverage: subset
## Select those stations where number of valid data points = length of record for the selected subset
## Trend analysis: mu, fw, P(X>x), sigma,
##
print('Evaluation of "the rain equation"')
```

## [1] "Evaluation of \"the rain equation\""
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
##
## Attaching package: 'esd'
## The following object is masked from 'package:base':
##
## subset.matrix
require(grDevices)
sstep <- 300  ## Size of the chunk of stations to read - the data is split up in order to cope with
## its volume due to memory restrictions.
nmin <- 300
x0 <- 1

## Algorithm to estimate the moderate return-values empirically
RVxyr <- function(x,t=10*365.25,na.rm=TRUE) {
  x[x<0] <- NA
  if (na.rm) x <- x[is.finite(x)]
  if (length(x) < t) return(NA)
  xx <- ceiling(max(x,na.rm=na.rm))
  h <- hist(x,breaks=seq(0,xx+10,by=1),plot=FALSE)
  cy <- cumsum(h$density)
  rv10yr <- approx(x=cy,y=h$mids,xout=1-1/t)$y
  return(rv10yr)
}

## Get the 24-hr precipitation data: find the interval with most complete data coverage for trend
## analysis for the same interval of all. Also collect the test results of the rain equation from
## scatter plots.

## Set working directory
setwd '~/R')

Fig 1: example of the simple equation

This work aims to evaluate the equation $Pr(X > x) = f_w e^{-x/\mu}$ that estimates the probability for heavy
daily precipitation based on two key parameters: the wet-day frequency $f_w$ and the wetday mean $\mu$. Here a
wet-day is define as days with $X > 1mm$. We use the notation where capital $X$ refers to a random variable
whereas lower-case $x$ refers to a particular value/threshold.

The following figure shows an example of the use of the equation $Pr(X > x) = f_w e^{-x/\mu}$ for Bjørnholt:
## Time series example

data("bjornholt")
rt <- test.rainequation(bjornholt, new=FALSE)

The "rain equation" for BJØRNHOLT

\[ Pr(X > x) = f_w e^{-x/\mu} \]
\[ \sum H(X - x)/n \]

Aggregate statistics for a large number of stations

The following chunk of R-code loops over daily rain gauge data from a large set of stations and repeats the calculations as in the example for a set of rainfall thresholds and extracts the results together with the observed data.

The data is presented in https://ocdp.met.no but only the Norwegian data are openly available from https://github.com/metno/OpenClimateData. The station data are stored in netCDF files that have been generated with esd::write2ncdf4.station().

files <- list.files(path="/OpenClimateData/data", pattern='precip', full.names = TRUE)

if (!file.exists('rainequation.rda')) {
  print('Some lengthy calculations')
  j <- 0
  if (file.exists('rainequation.tmp.rda')) {

## [1] "Correlation: 0.85"
```r
load('rainequation.tmp.rda')
else {
  files <- c(files[5],files[-5]) ## ECAD first
  print(files)
}
for (file in files) {
  print(file)
  z <- retrieve.stationsummary(file)
  ns <- dim(z)[1]
  iss <- seq(1,ns,by=sstep)
  for (is in iss) {
    if (is != max(iss)) ii <- is:(is+sstep-1) else ii <- is:ns
    Y <- retrieve(file,is=ii)
    nvg <- zoo(apply(Y,1,FUN='nv'),order.by = index(Y))
    sr <- scatterplot.rainequation(Y,x0=c(10,20,30,40,50))
    if (j==0) {
      ## The first results
      NVG <- nvg
      SR <- sr
      Z <- z
      Meta <- paste(loc(Y),stid(Y),cntr(Y))
    } else {
      ## Combine with previous results
      NVG <- NVG + nvg
      SR <- mapply(c, SR, sr, SIMPLIFY=FALSE)
      Z <- rbind(Z,z)
      Meta <- c(Meta,paste(loc(Y),stid(Y),cntr(Y)))
    }
    j <- j + 1
  }
  files <- files[-grep(file,files)]
  save(files,j,Z,SR,NVG,Meta,file='rainequation.tmp.rda')
}
save(SR,NVG,Z,Meta,file='rainequation.rda'
file.remove('rainequation.tmp.rda')
else {
  print('Use saved preprocessed data: rainequation.rda')
  load('rainequation.rda')
}
```

## [1] "Use saved preprocessed data: rainequation.rda"

print('Read the summary statistics (quick)')
```
## [1] "Read the summary statistics (quick)"
for (i in 1:length(files)) {
  print(files[i])
  z <- retrieve.stationsummary(files[i])
  if (i==1) Z <- z else Z <- rbind(Z,z)
}
```

## [1] "/home/rasmusb/OpenClimateData/data/precip.Africa.nc"

4
xCntr <- function(x) {
  n <- length(x);
  for (i in 1:n) if (!is.na(is.numeric(x[i]))) ix <- i
  return(paste(x[ix:n], collapse = ' '))
}

## Look for potentially duplicated stations
print('Search for duplicated stations')

## [1] "Search for duplicated stations"

dup <- duplicated(Meta)
print(table(dup))

## dup
## FALSE TRUE
## 9812 5

print(Meta[dup])

## [1] "KOUMAC (NLLE-CALEDO NC000091577 NEW CALEDONIA [FRANCE]"
## [2] "GISBORNE AERODROME NZ000093292 NEW ZEALAND"
## [3] "KWAJA RMW00040604 MARSHALL ISLANDS"
## [4] "MAJURO WBAS AP RMW00040710 MARSHALL ISLANDS"
## [5] "RAOUL ISL/KERMADEC NZ000093994 NEW ZEALAND"

sp <- strsplit(Meta[!dup], ' ')
sc <- unlist(lapply(sp, xCntr))
cntrs <- table(sc)
print(cntrs)

## sc
## (BRAZZ [DENMARK] [FRANCE] AFRI AFRICA
## # 11 52 3 7 174
## # ALBANIA ALGERIA ANGOLA ARABIA Argentina
## # 2 24 1 4 74
## # ARGENTINA ARMENIA AUSTRALIA AUSTRIA AZERBAIJAN
## # 19 6 377 11 41
## # BARBADOS BELARUS BELGIUM BENIN BOLIVIA
## # 1 55 1 7 8
## # BOTSWANA BRAZIL BURUNDI CAMEROON CANADA
## # 19 238 2 1 116
## # CHAD CHILE COLOMBIA CROATIA CUBA
## # 4 11 5 8 1
## # CYPRUS D’IVOIR DENMARK ECUADOR EGYPT
## # 14 13 22 2 14
## # ERITREA ESTONIA ETHIOPIA FASO FEDERATION
## # 2 2 10 8 1066
Check the trends in wet-day frequency

To see how robust the wet-day frequency is with respect to the threshold of 1 mm/day, we repeat the trend analysis with 2 mm/day and plot the trend estimates of the two.

```r
print(paste(length(Meta[!dup]),'stations from',length(cntrs),'countries'))
```

## [1] "9812 stations from 136 countries"

Check the trends in wet-day frequency

To see how robust the wet-day frequency is with respect to the threshold of 1 mm/day, we repeat the trend analysis with 2 mm/day and plot the trend estimates of the two.

```r
print('Compare trends for wet-day frequency with thresholds of 1 and 2 mm/day respectively')
```

## [1] "Compare trends for wet-day frequency with thresholds of 1 and 2 mm/day respectively"
if (!file.exists('wetfreq.robustness.test.rda')) {
files <- list.files(path=~OpenClimateData/data',pattern='precip',full.names = TRUE)
for (i in 1:length(files)) {
  pre <- retrieve.station(files[i])
  fw1 <- trend.coef(annual(pre,'wetfreq',threshold=1))
  fw2 <- trend.coef(annual(pre,'wetfreq',threshold=2))
  if (i==1) {FW1 <- fw1; FW2 <- fw2} else {FW1 <- c(FW1,fw1); FW2 <- c(FW2,fw2)}
}
save(FW1,FW2,file='wetfreq.robustness.test.rda')
} else load('wetfreq.robustness.test.rda')
ok <- is.finite(FW1) & is.finite(FW2)
plot(FW1,FW2,main='Wet-day frequency trend evaluation: change in frequency per decade',
     xlab=expression(x[0]==1*mm/day),ylab=expression(x[0]==2*mm/day),
     sub=paste(length(FW1),'station series; correlation=',round(cor(FW1[ok],FW2[ok]),2)))
grid()
fw12fit <- lm(FW2 ~ FW1)
print(summary(fw12fit))

## Call:
## lm(formula = FW2 ~ FW1)
##
## Residuals:
## Min 1Q Median 3Q Max
## -0.194628 -0.001358 -0.000050 0.001555 0.128827
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.784e-04 9.259e-05 3.006 0.00265 **
## FW1 7.666e-01 3.638e-03 210.737 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009114 on 9809 degrees of freedom
## (412 observations deleted due to missingness)
## Multiple R-squared: 0.8191, Adjusted R-squared: 0.8191
## F-statistic: 4.441e+04 on 1 and 9809 DF, p-value: < 2.2e-16
abline(fw12fit,col='red',lty=2)
Wet-day frequency trend evaluation: change in frequency per decac

Data overview

The next chunk of code examines the availability of the data and organises some of the data (the range of longitudes)

```r
print('Original size of the data set')

## [1] "Original size of the data set"
print(dim(SR)); print(dim(Z))

## NULL
## [1] 10223 55
print('Get some additional data')

## [1] "Get some additional data"
data("geoborders")
Z$longitude[Z$longitude > 180] <- Z$longitude[Z$longitude > 180] - 360
## Sort after number of valid data
srt_nv <- order(Z$number.valid)
Z <- Z[srt_nv,]
## Make sure to weed out any accidental duplicate
print('Check for duplicates in the station summary statistics')

## [1] "Check for duplicates in the station summary statistics"
```
lonlat <- paste(Z$longitude, Z$latitude)
dd <- duplicated(lonlat)
  #print(table(dd));
  print(table(cntr(Z)[dd])); print(src(Z)[dd])
## < table of extent 0 >
## NULL
Z <- Z[!dd,]

## How the number of valid data varies over time
par(bty='n')
plot(NVG, ylim=c(0, 8000), main='Number of valid rain gauge measurements'); grid()

Number of valid rain gauge measurements

## The global coverage for all rain gauge data used to test Pr(X>x)
esd::map(Z, FUN='mean', cex='number.valid', pal='precip', new=FALSE)

## Warning in par(par0): graphical parameter "cin" cannot be set
## Warning in par(par0): graphical parameter "cra" cannot be set
## Warning in par(par0): graphical parameter "csi" cannot be set
## Warning in par(par0): graphical parameter "cxy" cannot be set
## Warning in par(par0): graphical parameter "din" cannot be set
## Warning in par(par0): graphical parameter "page" cannot be set
The results of the test

The following figure shows the results of the comparison between the estimated likelihood and observed frequency of 24-hr precipitation amounts above the given thresholds. We repeated the comparison for four different thresholds, and the number of data points (1,003,096) was so large that a traditional scatter plot would make the size of the PDF-document too large. We present the results as a heatmap in stead with a logarithmic z-axis (colour scaling):

```r
## Compare calculations with observations
ok <- is.finite(SR$x) & is.finite(SR$y)

main <- expression(paste('Evaluation of the equation', 'Pr(X > x) = f[w] * e^{-x/\mu}'))
xlab <- expression(paste('Observed frequency: ', 'sum(H(X - x))/n'))
ylab <- expression(paste('Estimated probability: ', 'Pr(X > x) = f[w] * e^{-x/\mu}'))
breaks <- seq(0,6,by=0.5)
scatter(SR$x,SR$y,main=main,xlab=xlab,ylab=ylab,log=TRUE,breaks=breaks,
col=heat.colors(length(breaks)-1),sub=paste('Correlation=',round(cor(SR$x[ok],SR$y[ok]),3),', N=10'))
```
Evaluation of the equation \( \Pr(X > x) = f_w e^{-x/\mu} \)

The results from the test revealed a close 1-to-1 relationship between the majority of the points, with a correlation of 0.983. The conclusion from this analysis is that we can use \( \Pr(X > x) = f_w e^{-x/\mu} \) gives a close description of the likelihood for heavy 24-hr precipitation if we know \( f_w \) and \( \mu \).

Preparing daily precipitation data for trend analysis

We are interested in how the precipitation statistics has changed over the historical period with daily observations. The next chunk of computer code aggregates the data and estimates the key parameters \( f_w, \mu \) and return values \( x = \mu \ln(T, f_w) \).

```r
# Trend analysis: only stations with sufficient data:
if (!file.exists('rainequation.trend.rda')) {
  j <- 0
  RV <- list()
  for (tau in 1:10) {
    print(paste('tau=', tau))
    files <- list.files(path='~/OpenClimateData/data', pattern='precip', full.names = TRUE)
    files <- c(files[7], files[-7]) ## METNOD first
    for (file in files) {
      print(file)
      z <- retrieve.stationsummary(file)
      ns <- dim(z)[1]
      iss <- seq(1, ns, by=sstep)
      for (is in iss) {
        if (is != max(iss)) ii <- is:(is+sstep-1) else ii <- is:ns
      }
    }
  }
```
Y <- subset(retrieve(file,is=ii),it=c(1961,2018))
mu <- annual(Y,FUN='wetmean')
mum <- aggregate(Y,FUN='wetmean',month)
fw <- annual(Y,FUN='wetfreq')
ny <- apply(coredata(mu),2,'nv')
fwm <- aggregate(Y,FUN='wetfreq',month)
rv <- apply(Y,2,'RVxyr',t=tau*365.25)
## Only use stations with more than 50 years with valid data for return values
rv[ny < 50] <- NA
cal.rt <- data.frame(x=rv,
                     y=as.numeric(-colMeans(mu,na.rm=TRUE)*log(1/(colMeans(fw,na.rm=TRUE)*tau*365.25))))
ok <- is.finite(cal.rt[,1]) & is.finite(cal.rt[,2])
cal.rt <- cal.rt[ok,]
if (j==0) {
  ## The first results
  if (tau==1) {
    ## Only aggregate fw and mu once for one return-interval
    MU <- mu
    FW <- fw
    Mu <- mum
    Fw <- fwm
  }
  RV[[paste0('tau.',tau)]] <- cal.rt
}
else {
  ## Combine with previous results
  ## Only aggregate fw and mu once for one return-interval
  if (tau==1) {
    MU <- combine.stations(MU,mu)
    FW <- combine.stations(FW,fw)
    Mu <- combine.stations(Mu,mum)
    Fw <- combine.stations(Fw,fwm)
  }
  RV[[paste0('tau.',tau)]] <- rbind(RV[[paste0('tau.',tau)]],cal.rt)
}
j <- j + 1
}
tstid <- table(stid(MU))
tloc <- table(loc(MU))
print(tau); print(dim(MU)); print(summary(as.numeric(rv)))
print(summary(as.numeric(tstid))); print(summary(as.numeric(tloc)))
files <- files[-grep(file,files)]
plot(RV[[paste0('tau.',tau)]]))
print(dim(MU))
}
attr(RV,'description') <- 'Empirical 10-year-return-value'
save(MU,FW,Mu,Fw,RV,file='rainequation.trend.rda')
} else load('rainequation.trend.rda')

The distilled statistics is saved in a temporary file and used for the subsequent analysis.
Synchronise the aggregated data

Estimate the number of valid data points, keep only the stations with more than 50 years of data, and make sure that the data matrices with annual \( f_w \) and \( \mu \) represent the same locations (there were a few locations where there were differences). Also estimate the total annual precipitation amounts rather than the mean daily amounts.

```r
## Inspect the results (sanity check):
print('Sanity checks')

## [1] "Sanity checks"
par0 <- par()
#diagnose(MU) ## Skip this, as it takes so much memory.
## Only keep time records with more than 50 years of data.
nv <- apply(MU, 2, 'nv'); MU <- subset(MU, is=nv >=50)
nv <- apply(FW, 2, 'nv'); FW <- subset(FW, is=nv >=50)

## Make sure to have the same sites for MU and FW:
FW <- subset(FW, is=is.element(loc(FW), loc(MU)))
MU <- subset(MU, is=is.element(loc(MU), loc(FW)))

## Make sure to remove potential duplicates
print('Weed out duplicated sites based on lon-lat coordinates')
lonlat <- paste(lon(MU), lat(MU))
dd <- duplicated(lonlat)
print(table(dd))

## dd
## FALSE TRUE
##  1875  103
FW <- subset(FW, is=!dd)
MU <- subset(MU, is=!dd)
nv <- apply(MU, 2, 'nv')

print('Number of years and stations')

## [1] "Number of years and stations"
print(dim(MU)); print(dim(FW))

## [1] 58 1875
## [1] 58 1875

print('Number of years with data:')

## [1] "Number of years with data:"
print(summary(apply(MU, 2, 'nv')))
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 50.00 52.00 55.00 54.54 57.00 58.00

sc <- table(cntr(MU))
print(sc)

##
## AMERICAN SAMOA [UNITED S Argentina AUSTRALIA
## 1 23 69
## BOSNIA AND HERZEGOVINA CANADA CROATIA
## 1 3 13
## CYPRUS DENMARK ESTONIA
## 3 3 2
## FINLAND FRANCE FRENCH GUIANA [FRANCE]
## 64 35 1
## GERMANY ICELAND IRELAND
## 280 3 1
## ISRAEL LUXEMBOURG MARSHALL ISLANDS
## 4 1 2
## NETHERLANDS NORWAY PORTUGAL
## 270 238 1
## ROMANIA RUSSIAN FEDERATION SLOVAKIA
## 23 262 4
## SLOVENIA SPAIN SWEDEN
## 9 47 232
## SWITZERLAND UKRAINE UNITED KINGDOM
## 7 4 1
## UNITED STATES
## 268

print(paste('Records from',length(sc),'countries with at least 50 years between 1961 and 2018'))

## [1] "Records from 31 countries with at least 50 years between 1961 and 2018"

## The mean precipitation
X <- 365.25*FW*MU
X <- attrcp(MU,X); class(X) <- class(MU)
attr(X,'variable') <- 'precip'
attr(X,'unit') <- 'mm'
print('The data size after synchronisation and only using series longer than 50 years:')

## [1] "The data size after synchronisation and only using series longer than 50 years:"

print(dim(X))

## [1] 58 1875

Overview of the general rainfall statistics

The following chunk of code provides diagnostics for inspection of the data. The diagnostics will only include
time series that are 50 years or longer.

## Set up a data frame with data that will be displayed as maps:
print('Maps')

## [1] "Maps"
Q <- data.frame(longitude=lon(X), latitude=lat(X), number.valid=nv)
class(Q) <- c('stationsummary', 'data.frame')

esd::map(X, FUN='mean', plot=FALSE) -> X.mean
Q$X.mean <- as.numeric(X.mean)

esd::map(Q, FUN='X.mean', cex='number.valid', cex0=0.75, pal='precip', new=FALSE)

## Warning in par(par0): graphical parameter "cin" cannot be set
## Warning in par(par0): graphical parameter "cra" cannot be set
## Warning in par(par0): graphical parameter "csi" cannot be set
## Warning in par(par0): graphical parameter "cxy" cannot be set
## Warning in par(par0): graphical parameter "din" cannot be set
## Warning in par(par0): graphical parameter "page" cannot be set

X.mean (mean= 796, sd=428 [25, 3586])
### Warning in par(par0): graphical parameter "cin" cannot be set
### Warning in par(par0): graphical parameter "cra" cannot be set
### Warning in par(par0): graphical parameter "csi" cannot be set
### Warning in par(par0): graphical parameter "cxy" cannot be set
### Warning in par(par0): graphical parameter "din" cannot be set
### Warning in par(par0): graphical parameter "page" cannot be set

**mu** (mean= 7.1, sd=2.8 [2.8, 18.6])

```
esd::map(FW,FUN='mean',plot=FALSE) -> fw
Q$fw <- as.numeric(fw)
esd::map(Q,FUN='fw',cex='number.valid',cex0=0.75,rev=TRUE,new=FALSE)
```

### Warning in par(par0): graphical parameter "cin" cannot be set
### Warning in par(par0): graphical parameter "cra" cannot be set
### Warning in par(par0): graphical parameter "csi" cannot be set
### Warning in par(par0): graphical parameter "cxy" cannot be set
### Warning in par(par0): graphical parameter "din" cannot be set
### Warning in par(par0): graphical parameter "page" cannot be set
fw (mean = 0.31, sd = 0.09 [0.01, 0.64])

```r
esd::map(X, FUN = 'trend', plot = FALSE) -> dx.dt
Q$dx.dt <- as.numeric(dx.dt)
esd::map(Q, FUN = 'dx.dt', cex = 'number.valid', cex0 = 0.75, rev = TRUE, new = FALSE)

## Warning in par(par0): graphical parameter "cin" cannot be set
## Warning in par(par0): graphical parameter "cra" cannot be set
## Warning in par(par0): graphical parameter "csi" cannot be set
## Warning in par(par0): graphical parameter "cxy" cannot be set
## Warning in par(par0): graphical parameter "din" cannot be set
## Warning in par(par0): graphical parameter "page" cannot be set
```
dx.dt (mean = 13, sd = 23 [-96, 150])

esd::map(MU,FUN = 'trend', plot = FALSE) -> dmu.dt
Q$dmu.dt <- as.numeric(dmu.dt)
esd::map(Q,FUN = 'dmu.dt', cex = 'number.valid', cex0 = 0.75, rev = TRUE, new = FALSE)

## Warning in par(par0): graphical parameter "cin" cannot be set
## Warning in par(par0): graphical parameter "cra" cannot be set
## Warning in par(par0): graphical parameter "csi" cannot be set
## Warning in par(par0): graphical parameter "cxy" cannot be set
## Warning in par(par0): graphical parameter "din" cannot be set
## Warning in par(par0): graphical parameter "page" cannot be set
**esd::map(FW,FUN='trend',plot=FALSE) -> dfw.dt**

Q$dfw.dt <- as.numeric(dfw.dt)
Q$dfw.dt[abs(Q$dfw.dt) > 0.03] <- NA
esd::map(Q,FUN='dfw.dt',cex='number.valid',cex0=0.75,rev=TRUE,new=FALSE)

## Warning in par(par0): graphical parameter "cin" cannot be set
## Warning in par(par0): graphical parameter "cra" cannot be set
## Warning in par(par0): graphical parameter "csi" cannot be set
## Warning in par(par0): graphical parameter "cxy" cannot be set
## Warning in par(par0): graphical parameter "din" cannot be set
## Warning in par(par0): graphical parameter "page" cannot be set
The size of the symbols in these maps indicate the length of the timeseries for the given location, taking the circle radius to be $\propto \sqrt{L}$ where $L$ is the length of the timeseries.

The statistics confirm that the greatest annual precipitation totals are near the coasts (ranging from 25 to 3568 mm/year), mean annual $\mu$ is smallest in the interior mid-to-high latitudes and greatest in warm regions affected by the coast, and $f_w$ is greatest along the west coast of northern Europe and some Pacific islands.

The greatest trend in the annual total precipitation is seen near the coast of eastern North America and western Europe. The trend ranges from -96 mm/year to 168 mm/year per decade, with an average of 13 mm/year and a standard deviation of 24 mm/year per decade.

The trend in $\mu$ varies between -0.83 mm/day per decade to 1.21 mm/day per decade with a mean of 0.08 mm/day/decade and a standard deviation of 0.14 mm/day per decade. There is a more pronounced trend in the eastern part of North America, Argentina and in maritime-influenced regions of northern Europe.

For $f_w$ there is a clear geographical pattern in the trend, with a general increase at higher latitudes and decrease in the lower latitudes (the sub-tropics). The mean is 0 with a standard deviation of 0.01/decade (fraction, not percentage) and a range of -0.03 to +0.04 per decade.

### Proportional change (%)

The following chunk estimates $f_w$ and $\mu$ in terms of percents of the reference period 1961–1990:

```r
## Change & trends in percentage for clearer visualisation
print('Percentage trends')
```
The mean seasonal cycle

The mean annual cycle was inspected (not shown) and it looked realistic/good. There are similarities in the mean seasonal cycle in the rainfall statistics within different regions, where the variations follow the seasons. The strongest proportional seasonal variations (in %) in the wet-dat frequency (%) is over dry regions with low frequencies. Likewise, the strongest proportional seasonal variaatoins (in %) in the wet-dat mean precipitation (%) is over the interior continents.

The probability of rainfall more than 50 mm/day

The following chunk of code presents the probability of more than 50 mm/day $Pr(X > 50)$ and its tend:

```
Pr.X.gt.50 <- 100*zoo(coredata(FW)*exp(-50/coredata(MU)),order.by=index(MU))
## Ignore outliers
zpr <- coredata(Pr.X.gt.50); zpr[zpr > 2] <- NA; zpr -> coredata(Pr.X.gt.50)
Pr.X.gt.50 <- attrcp(MU,Pr.X.gt.50); class(Pr.X.gt.50) <- class(MU)
attr(Pr.X.gt.50,'variable') <- 'probability'
attr(Pr.X.gt.50,'unit') <- '%'
## Mean
esd::map(Pr.X.gt.50,FUN='mean',plot=FALSE) -> Pr.gt.50
Q$Pr.gt.50 <- as.numeric(Pr.gt.50)
esd::map(Q,FUN='Pr.gt.50',cex='number.valid',cex0=0.75,pal='precip',new=FALSE)
```

```
## Warning in par(par0): graphical parameter "cin" cannot be set
## Warning in par(par0): graphical parameter "cra" cannot be set
## Warning in par(par0): graphical parameter "csi" cannot be set
## Warning in par(par0): graphical parameter "cxy" cannot be set
## Warning in par(par0): graphical parameter "din" cannot be set
## Warning in par(par0): graphical parameter "page" cannot be set
```
Pr.gt.50 (mean= 0.11, sd=0.25 [0, 1.61])

## Trend

```r
nvy <- apply(Pr.X.gt.50,2,ts1

esd::map(Pr.X.gt.50,FUN='trend',plot=FALSE) -> dPrdt

## Ignore outliers

zpr <- coredata(dPrdt); zpr[abs(zpr) > 0.1] <- NA; zpr[nvy < 50] <- NA; zpr -> coredata(dPrdt)

## Express changes in probability as % relative to the mean probability

dPr.dt <- 100*dPrdt/coredata(Pr.gt.50)
decrpr <- dPr.dt < 0; incrpr <- dPr.dt > 0

Q$dPr.dt <- as.numeric(dPr.dt)

esd::map(Q,FUN='dPr.dt',cex='number.valid',cex0=0.75,rev=TRUE,new=FALSE)
```

## Warning in par(par0): graphical parameter "cin" cannot be set

## Warning in par(par0): graphical parameter "cra" cannot be set

## Warning in par(par0): graphical parameter "csi" cannot be set

## Warning in par(par0): graphical parameter "cxy" cannot be set

## Warning in par(par0): graphical parameter "din" cannot be set

## Warning in par(par0): graphical parameter "page" cannot be set
The greatest trend probabilities is found in regions influenced by the marine environment and generally higher humidity. The bulk of the estimates indicate increasing trends, but there are also some outliers with large
negative trend in Iberia.

Variance according to the exponential distribution

The following chunk of code quantifies the variance $\sigma^2$ of the 24-hr precipitation based on the probability density function $f(x)$ and the integral $\sigma^2 = \int x^2 f(x) \, dx$ and the solution to this expression $\sigma^2 = f_w \left[ \mu^2 + (x_0^2 + 2\mu x_0 + 2\mu^2) e^{-x_0/\mu} \right]$ (using rain gauge data with more than 30 years and hence a larger set than that with stations longer than 50 years). Also we use the expression derived from the trend in the variance $d\sigma^2/dt = \left[ \mu^2 + (x_0^2 + 2\mu x_0 + 2\mu^2) e^{-x_0/\mu} \right] dfw/dt + f_w \left[ 2\mu + (4x_0 + 4\mu + x_0^3/\mu^2 + 2x_0^2/\mu) e^{-x_0/\mu} \right] d\mu/dt$.

The mean variance is greatest in southeastern North America, Argentina, India, Indonesia, parts of Brazil, northern Australia around the Mediterranean, parts of Africa, parts of Central America, and some Pacific islands. The trend in $\sigma^2$ has mainly been increasing variance and the analysis indicates a number of short records (small symbol size) in Africa and Latin America with an apparent reduction.

Test of trend in the probability for heavy daily precipitation

The following R-code was intended for testing the decomposition of the trends in to terms with trends in $f_w$ and $\mu$. We used scatter plots to display the influence of the change in the frequency and mean intensity and to relate these to the trends in the annual total precipitation and the likelihood $Pr(X > x)$:

```r
## Get consistent scale with the mean values:
dfw.dt <- 0.1*as.numeric(dfw.dt)
dmu.dt <- 0.1*as.numeric(dmu.dt)

## Test the trends:

## Trend in the mean precipitation:
dX.dt <- dfw.dt*mu + fw*dmu.dt
## Trend in probability
dfx.dt <- dfw.dt*exp(-50/\mu) + 50*fw/\mu^2*exp(-50/\mu)*dmu.dt

pal <- colscal(n=100,col='t2m',rev=TRUE)

## Trend in annual precipitation totals:

## Check trends in mean:
par(new=FALSE,mar=c(6.1, 6.1, 4.1, 2.1))
plot(as.numeric(dx.dt),3652.5*as.numeric(dx.dt),pch=19,col=rgb(0.5,0.5,0,0.3),
     main='Test of trends mean precipitation',
     xlab=expression(paste(' Trend: ',frac(d*x,dt),' (mm/year per decade) ')),
     ylab=expression(paste(' Trend in intensity: ',frac(d*f[w],d*t)*mu+f[w]*frac(d*mu,dt),' (mm/year per decade)')))
lines(c(-100,100),c(-100,100),lty=2)
lines(c(-100,100),c(0,0),lty=1);
lines(c(0,0),c(-100,100),lty=1);
grid()
```

24
Test of trends mean precipitation

Trend: \( \frac{dx}{dt} \) (mm/year per decade)

Trend in intensity: \( \frac{d^2f w}{dt^2} \)  
\( \mu + \frac{d^2f w}{dt^2} \) (mm/year per decade)

\[
\text{print} \left( \text{cor.test} \left( \text{as.numeric} \left( \frac{dx}{dt} \right), 3652.5 \ast \text{as.numeric} \left( \frac{dX}{dt} \right) \right) \right)
\]

## Pearson's product-moment correlation
##
## data:  \text{as.numeric}(\frac{dx}{dt}) \text{ and } 3652.5 \ast \text{as.numeric}(\frac{dX}{dt})
## t = 554.89, \text{df} = 1873, \text{p-value} < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9966856 0.9972342
## sample estimates:
## cor
## 0.9969723

## Set color scheme according to trend in probability
dpmax <- \text{max}(\text{abs}(\frac{dx}{dt})); \text{incrx} \leftarrow \frac{dx}{dt} \ast 0
\text{col} \leftarrow \text{pal}[\text{trunc}(\text{approx}(x=\text{seq}(-\text{dpmax}, \text{dpmax}, \text{length}=100), y=1:100, xout=\frac{dx}{dt})$y)]

\text{par(new=FALSE,mar=c(6.1, 6.1, 4.1, 2.1))}
\text{plot}(10 \ast \text{as.numeric}(\frac{df w}{dt} \ast \mu), 10 \ast \text{as.numeric}(\frac{fw \ast d\mu}{dt}), \text{pch}=19, \text{col}=\text{col},
\text{main='Decomposed trends in mean precipitation'},
\text{xlab=expression(\text{paste('Trend in frequency: } ', \frac{d^2f w}{dt^2}, ' \ast \mu, \text{ (mm/decade)}'))},
\text{ylab=expression(\text{paste('Trend in intensity: } f[w] \ast \frac{d^2f w}{dt^2}, \text{ (mm/decade)}'))})
\text{points}(10 \ast \text{as.numeric}(\frac{df w}{dt} \ast \mu)[\text{incrx}], 10 \ast \text{as.numeric}(\frac{fw \ast d\mu}{dt})[\text{incrx}],
\text{col=rgb(0,0,0,0.2))}
\text{lines}(2 \ast \text{c(-1,1)}, 2 \ast \text{c(-1,1)}, \text{lty=2})
\text{lines}(2 \ast \text{c(-1,1)}, 2 \ast \text{c(0,0)}, \text{lty=1});
Decomposed trends in mean precipitation

The first scatter plot confirms that the trend in mean is the sum of the terms related to the trend in $f_w$ and $\mu$ respectively. In other words, the trend in the mean precipitation shows that it agrees well with the sum of the terms related to the changes in the mean intensity and frequency.

The second scatter plot suggests that an increase in the total annual amount can take place even if either $f_W$ or $\mu$ is decreasing, but not both. There is no tendency that either $f_w$ or $\mu$ plays a dominant role for the trend in the total amounts.

Trend in probability for heavy daily precipitation

```
## Trend in probability:

## Set color scheme according to trend in probability
dpmax <- max(abs(dPrdt),na.rm=TRUE)
col <- pal[trunc(approx(x=seq(-dpmax,dpmax,length=100),y=1:100,xout=dPrdt)$y)]

par(new=FALSE,mar=c(6.1, 6.1, 4.1, 2.1))
plot(as.numeric(dPrdt),100*as.numeric(dfx.dt)*10,pch=19,col=col,
     main='Test of trend in probability of heavy precipitation by components',
     xlab=expression(paste('Trend in Pr(',X','>: ',frac(d*Pr(',X','>x)',dt),')', '(%/decade)'),
     ylab=expression(paste('Trend by components: ',
                          frac(d*f[w],dt)*e^{-x/mu} + frac(f[w]*x,mu^2)*e^{-x/mu}*frac(d*mu,dt),') (%/decade)'))
```
Printed text is not available.
The second figure provides a check of the composition of the trends in the probability of heavy daily precipitation in terms of expressions relating to trends in the mean intensity and frequency. The estimates tend to fall along the diagonal and hence verify the composition of the trend in the probability.

The third plot compares the trend in probability with the two components $df_w/dt$ and $d\mu/dt$, and it is evident that the trends in the mean intensity $\mu$ dominates over $f_w$ when it comes to the effect on the probability of days with heavy rainfall. Moreover, the trend in precipitation has mainly been due to increasing mean intensity. The probability has increased also when the mean frequency has decreased.

Change in variance, intensity and frequency

The following scatter plot relates the trend in $\sigma^2$ to trends in $f_w$ and $\mu$ respectively:

```r
## Trend in variance:
## Trend in variance (sigma^2)
ds2.dt <- mu^2*dfw.dt + ( mu^2 + (x0^2 + 2*mu*x0 + 2*mu^2)*exp(-x0/mu) )*dfw.dt +
  fu*(2*mu + (4*x0 + 4*mu + x0^3/mu^2 + 2*x0^2/mu)*exp(-x0/mu ))*dmu.dt
sum(ds2.dt>0)/sum(is.finite(ds2.dt))*100
```

### [1] 78.56
### Set color scheme according to trend in probability

dpmax <- max(abs(ds2.dt)); incrs2 <- ds2.dt > 0
col <- pal[trunc(approx(x=seq(-dpmax,dpmax,length=100),y=1:100,xout=ds2.dt)$y)]

par(new=FALSE,mar=c(6.1, 6.1, 4.1, 2.1))
plot(10*as.numeric(mu^2*dfw.dt + ( mu^2 + (x0^2 + 2*mu*x0 + 2*mu^2)*exp(-x0/mu) )*dfw.dt),
10*as.numeric(fw*(2*mu + (4*x0 + 4*mu + x0^3/mu^2 + 2*x0^2/mu)*exp(-x0/mu) )*dmu.dt),pch=19,col=col:
main=expression(paste('Decomposed trends in precipitation variance ',sigma^2)),
xlab=expression(paste('Trend in frequency: ',f(d*mu/dt),' (mm^2/decade)'))),
ylab=expression(paste('Trend in intensity: ',f(dfw/dt),' (mm^2/decade)')))
points(10*as.numeric(mu^2*dfw.dt + ( mu^2 + (x0^2 + 2*mu*x0 + 2*mu^2)*exp(-x0/mu) )*dfw.dt)[incrs2],
10*as.numeric(fw*(2*mu + (4*x0 + 4*mu + x0^3/mu^2 + 2*x0^2/mu)*exp(-x0/mu) )*dmu.dt)[incrs2],
col=rgb(0,0,0,0.2))
lines(0.2*c(-1,1),0.2*c(-1,1),lty=2)
lines(0.2*c(-1,1),0.2*c(0,0),lty=1);
lines(c(0,0),0.2*c(-1,1),lty=1);
grid()

---

Trends in the mean intensity has been a dominating cause for the trend in variance, but there are also some cases where the variance has increased with a slight decrease in the mean intensity. The variance has increased in many cases despite a decrease in the mean frequency.

#### Change in variance, intensity and frequency

The following scatter plot relates the trend in $x_{10}$ to trends in $f_w$ and $\mu$ respectively:
## Trend in variance:

```r
dx10.dt <- mu/fw*dfw.dt + log(fw*3652.5)*dmu.dt
```

## Set color scheme according to trend in probability

```r
dpmax <- max(abs(dx10.dt)); incrs3 <- dx10.dt > 0
col <- pa[trunc(approx(x=seq(-dpmax,dpmax,length=100),y=1:100,xout=dx10.dt)y)]
```

```r
par(new=FALSE,mar=c(6.1, 6.1, 4.1, 2.1))
plot(10*as.numeric(mu/fw*dfw.dt),10*as.numeric(log(fw*3652.5)*dmu.dt),pch=19,col=col,
    main=expression(paste('Decomposed trends in return-value ',x[tau])),
    xlab=expression(paste('Trend in frequency: ',mu/fw*frac(d*fw,dt),'} (mm/day per decade)'),
    ylab=expression(paste('Trend in intensity: ',ln(tau*fw)*frac(d*mu,dt),'} (mm/day per decade'))))
points(10*as.numeric(mu/fw*dfw.dt)[incrs3],
    10*as.numeric(log(fw*3652.5)*dmu.dt)[incrs3],
    col=rgb(0,0,0,0.2))
lines(0.2*c(-1,1),0.2*c(-1,1),lty=2)
lines(0.2*c(-1,1),0.2*c(0,0),lty=1);
lines(c(0,0),0.2*c(-1,1),lty=1);
grid()
```

### Explanation for trends in the probability of heavy rainfall

The following code produces pie charts that visualises the proportion of cases with increasing or decreasing likelihood of heavy precipitation and the association with the trends in $f_w$ and $\mu$:  

![Decomposed trends in return-value $x_\tau$](chart.png)
## Summary statistics:
probs.50 <- rep('Increasing',length(dPr.dt)); probs.50[decrpr] <- 'Decreasing'
p.ip <- round(100*sum(incrpr,na.rm=TRUE)/sum(is.finite(incrpr)))
p.dp <- round(100*sum(decrpr,na.rm=TRUE)/sum(is.finite(decrpr)))
print(table(probs.50))

## probs.50
## Decreasing Increasing
## 430 1445
incr.prob.50.decr.mu <- (probs.50=='Increasing') & dmu.dt < 0
decr.prob.50.decr.mu <- (probs.50=='Decreasing') & dmu.dt < 0
print(table(incr.prob.50.decr.mu))

## incr.prob.50.decr.mu
## FALSE TRUE
## 1802 73
incr.prob.50.incr.mu <- (probs.50=='Increasing') & dmu.dt > 0
decr.prob.50.incr.mu <- (probs.50=='Decreasing') & dmu.dt > 0
print(table(incr.prob.50.incr.mu))

## incr.prob.50.incr.mu
## FALSE TRUE
## 503 1372
p.ip.dm <- round(100*sum(incr.prob.50.decr.mu,na.rm=TRUE)/sum(probs.50=='Increasing',na.rm=TRUE))
p.ip.im <- round(100*sum(incr.prob.50.incr.mu,na.rm=TRUE)/sum(probs.50=='Increasing',na.rm=TRUE))
p.dp.dm <- round(100*sum(decr.prob.50.decr.mu,na.rm=TRUE)/sum(probs.50=='Decreasing',na.rm=TRUE))
p.dp.im <- round(100*sum(decr.prob.50.incr.mu,na.rm=TRUE)/sum(probs.50=='Decreasing',na.rm=TRUE))
print(c(p.ip.dm,p.ip.im))

## [1] 5 95
incr.prob.50.decr.fw <- (probs.50=='Increasing') & dfw.dt < 0
decr.prob.50.decr.fw <- (probs.50=='Decreasing') & dfw.dt < 0
print(table(incr.prob.50.decr.fw))

## incr.prob.50.decr.fw
## FALSE TRUE
## 1473 402
incr.prob.50.incr.fw <- (probs.50=='Increasing') & dfw.dt > 0
decr.prob.50.incr.fw <- (probs.50=='Decreasing') & dfw.dt > 0
print(table(incr.prob.50.incr.fw))

## incr.prob.50.incr.fw
## FALSE TRUE
## 832 1043
p.ip.df <- round(100*sum(incr.prob.50.decr.fw,na.rm=TRUE)/sum(probs.50=='Increasing',na.rm=TRUE))
p.ip.if <- round(100*sum(incr.prob.50.incr.fw,na.rm=TRUE)/sum(probs.50=='Increasing',na.rm=TRUE))
p.dp.df <- round(100*sum(decr.prob.50.decr.fw,na.rm=TRUE)/sum(probs.50=='Decreasing',na.rm=TRUE))
p.dp.if <- round(100*sum(decr.prob.50.incr.fw,na.rm=TRUE)/sum(probs.50=='Decreasing',na.rm=TRUE))
print(round(100*sum(incr.prob.50.decr.fw,na.rm=TRUE)/sum(is.finite(incr.prob.50.decr.fw)))))

## [1] 21
fincr <- sqrt(sum(is.element(probs.50, 'Increasing'))/sum(!is.na(probs.50)))
fdecr <- sqrt(sum(is.element(probs.50, 'Decreasing'))/sum(!is.na(probs.50)))

print(paste('Increasing Pr:', sum(incr.prob.50.incr.mu & incr.prob.50.incr.fw),
    'stations w incr. mu',
    sum(incr.prob.50.incr.mu), 'w incr. fw',
    sum(incr.prob.50.incr.fw), 'w decr. mu',
    sum(incr.prob.50.decr.mu), 'w decr. fw',
    sum(incr.prob.50.decr.fw)))

## [1] "Increasing Pr: 992 stations w incr. mu 1372 w incr. fw 1043 w decr. mu 73 w decr. fw 402"

print(paste('Decreasing Pr:', sum(decr.prob.50.incr.mu & decr.prob.50.incr.fw),
    'stations w incr. mu',
    sum(decr.prob.50.incr.mu), 'w incr. fw',
    sum(decr.prob.50.incr.fw), 'w decr. mu',
    sum(decr.prob.50.decr.mu), 'w decr. fw',
    sum(decr.prob.50.decr.fw)))

## [1] "Decreasing Pr: 46 stations w incr. mu 103 w incr. fw 193 w decr. mu 327 w decr. fw 237"

## Graphical presentation
par(mar=c(1,1,2,1))
nf <- layout(matrix(c(rep(1,12),2,2,2,2,3,3,4,4,4,5,5,5), 7, 4, byrow = TRUE))
pie(c(sum(incrpr, na.rm=TRUE), sum(!incrpr, na.rm=TRUE)), radius=1,
    main='Trend in Pr(X > 50 mm/day)',
    col=c('blue', 'brown'),
    labels=c(paste0(p.ip, '% increasing'), paste0(p.dp, '% decreasing')))
pie(c(sum(incr.prob.50.decr.mu), sum(incr.prob.50.incr.mu)),
    main=expression(paste('Increasing Pr(X > 50 mm/day): ', mu)), radius=1.2*fincr,
    col=c('brown', 'blue'),
    labels=c(paste0(p.ip.dm, '% decreasing'), paste0(p.ip.im, '% increasing')))
pie(c(sum(incr.prob.50.decr.fw), sum(incr.prob.50.incr.fw)),
    main=expression(paste('Increasing Pr(X > 50 mm/day): ', f[w])), radius=1.2*fincr,
    col=c('brown', 'blue'),
    labels=c(paste0(p.ip.df, '% decreasing'), paste0(p.ip.if, '% increasing')))
pie(c(sum(decr.prob.50.decr.mu), sum(decr.prob.50.incr.mu)),
    main=expression(paste('Decreasing Pr(X > 50 mm/day): ', mu)), radius=1.2*fdecr,
    col=c('brown', 'blue'),
    labels=c(paste0(p.dp.dm, '% decreasing'), paste0(p.dp.im, '% increasing')))
pie(c(sum(decr.prob.50.decr.fw), sum(decr.prob.50.incr.fw)),
    main=expression(paste('Decreasing Pr(X > 50 mm/day): ', f[w])), radius=1.2*fdecr,
    col=c('brown', 'blue'),
    labels=c(paste0(p.dp.df, '% decreasing'), paste0(p.dp.if, '% increasing')))
76% of the locations indicate an increasing trend in the likelihood of heavy rainfall (more than 50 mm/day) compared to 24% with decreasing probability. Only 4% of the sites with increasing probability also indicate a negative trend in $\mu$, whereas 21% of the sites with increasing probability also had decreasing trends in $f_w$.

**Crude estimation of low 24-hr precipitation return values**

The simple expression for the probability of daily precipitation exceeding a given threshold can be used to derive an expression for a return value $x = \mu \ln(f w)$. We tested this expression for the 10-year return value against empirical frequencies based on the observations in the next chunk of code:

```r
par(new=FALSE, mar=c(5.1, 5.1, 4.1, 2.1), bty='n', xaxt='s', yaxt='s', fig=c(0,1,0,1))
nrv <- length(RV)
bias.factors <- matrix(rep(NA, 4*nrv), nrv, 4)
for (itau in 1:nrv) {
  cal.rt <- RV[[itau]]
  ## bias-adjusted
  rt.adjust <- lm(x ~ y, data=cal.rt)
bias.factors[itau,] <- summary(rt.adjust)$coefficients[1:4]
  print(summary(rt.adjust))
  plot(cal.rt$x, cal.rt$y, pch=19, col=rgb(0.5, 0.5, 0.3),
       main=expression(paste('Bias-adjusted return-value from ', Pr(X>x)==f[w]*e^{-x/\mu}))),
       xlab='Empirical estimate (mm/day)', ylim=c(0, 200), ylim=c(0, 200),
       ylab=expression(paste(\mu \phantom{0} \ln(f[w]*tau), ' (mm/day)')),
       sub=paste('Correlation', round(cor(cal.rt$x, cal.rt$y), 3))
       points(cal.rt$x, predict(rt.adjust), col=rgb(0, 0, 0, 0.2))
```

33
Bias--adjusted return--value from $\Pr(X > x) = f_w e^{-x/\mu}$

Empirical estimate (mm/day)
Corelation 0.953

### Call:
```r
lm(formula = x ~ y, data = cal.rt)
```
## Residuals:
## Min 1Q Median 3Q Max
## -23.924 -3.547 -0.705 2.685 109.333
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.22010  0.46013  -9.172  <2e-16 ***
## y            1.29586  0.01113   116.467  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.474 on 1976 degrees of freedom
## Multiple R-squared: 0.8728, Adjusted R-squared: 0.8728
## F-statistic: 1.356e+04 on 1 and 1976 DF, p-value: < 2.2e-16

Bias–adjusted return–value from Pr(X > x) = f_w e^{-x/\mu}

Empirical estimate (mm/day)
Correlation 0.934
$$\text{Bias-adjusted return-value from } Pr(X > x) = f_w e^{-x/\mu}$$

![Graph showing the relationship between $\mu \ln(f_w \tau)$ and empirical estimate (mm/day).](image)

**Corelation 0.922**

### Call:
```
Call: lm(formula = x ~ y, data = cal.rt)
```

### Coefficients:
```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.49792  0.61895  -7.267  5.27e-13 ***
y           1.34529  0.01326 101.462 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Residual standard error: 10.03 on 1976 degrees of freedom
---

### Model Output:
```
Residuals: Min 1Q Median 3Q Max
-31.960 -4.900 -0.890 3.734 127.608

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.55109  0.55888  -8.143  6.74e-16 ***
y           1.32871  0.01257 105.725 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.065 on 1976 degrees of freedom
```

Multiple R-squared: 0.8498, Adjusted R-squared: 0.8497
F-statistic: 1.118e+04 on 1 and 1976 DF, p-value: < 2.2e-16
## F-statistic: 1.029e+04 on 1 and 1976 DF, p-value: < 2.2e-16

Bias-adjusted return-value from $\Pr(X > x) = f_w e^{-x/\mu}$

Empirical estimate (mm/day)
Corelation 0.916

## Call:
## lm(formula = x ~ y, data = cal.rt)
##
## Residuals:
##    Min     1Q  Median     3Q    Max
## -33.273 -5.411  -0.926  4.085 141.207
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.67377   0.68270  -6.846  1.01e-11 ***
##         y    1.36368   0.01411  96.671 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.06 on 1976 degrees of freedom
## Multiple R-squared:  0.8255, Adjusted R-squared:  0.8254
## F-statistic: 9345 on 1 and 1976 DF, p-value: < 2.2e-16
Bias−adjusted return−value from $\Pr(X > x) = f_w e^{-x/\mu}$
Bias–adjusted return–value from \( \Pr(X > x) = f_w e^{-x/\mu} \)

Empirical estimate (mm/day)

Correlation 0.901

## Call:
\[
\text{lm(formula = x ~ y, data = cal.rt)}
\]
## Residuals:

|       | Min  | 1Q   | Median | 3Q   | Max  |
|-------|------|------|--------|------|------|
| Min   | -39.977 | -6.490 | -1.101 | 4.580 | 153.717 |

## Coefficients:

|               | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | -4.40936 | 0.79372    | -5.555  | 3.15e-08 *** |
| y             | 1.38156  | 0.01557    | 88.743  | < 2e-16 *** |

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.85 on 1976 degrees of freedom
Multiple R-squared:  0.7994, Adjusted R-squared:  0.7993
F-statistic: 7875 on 1 and 1976 DF,  p-value: < 2.2e-16
Bias-adjusted return-value from \( \Pr(X > x) = f_w e^{-x/\mu} \)

```
## Call:
## lm(formula = x ~ y, data = cal.rt)
##
## Residuals:
##    Min     1Q   Median     3Q    Max
##-42.413 -6.795  -1.223  4.749 160.456
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) -4.13182   0.83478  -4.95 8.07e-07 ***
##       y     1.38595   0.01605   86.35  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.51 on 1976 degrees of freedom
## Multiple R-squared: 0.7905, Adjusted R-squared: 0.7904
## F-statistic: 7457 on 1 and 1976 DF, p-value: < 2.2e-16
```
Bias-adjusted return-value from \( \Pr(X > x) = f_w e^{-x/\mu} \)

Empirical estimate (mm/day)

\[ \mu \ln(f_w \tau) \text{ (mm/day)} \]

## Call:
```
## lm(formula = x ~ y, data = cal.rt)
##```

## Residuals:
```
##    Min      1Q  Median      3Q     Max
## -45.019  -7.287  -1.367   4.837  171.519
##```

## Coefficients:
```
##                Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -3.94252   0.89721  -4.394 1.17e-05 *** 
## y             1.39247   0.01696  82.126  < 2e-16 *** 
##                ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Residual standard error: 14.52 on 1976 degrees of freedom
## Multiple R-squared:  0.7734, Adjusted R-squared:  0.7733
## F-statistic: 6745 on 1 and 1976 DF, p-value: < 2.2e-16
Bias-adjusted return-value from \( \Pr(X > x) = f_w e^{-x/\mu} \)

### Call:
```
Call:
  lm(formula = x ~ y, data = cal.rt)
```

### Residuals:
```
  Min  1Q Median  3Q  Max
-47.080 -7.451 -1.599 4.873 303.212
```

### Coefficients:
```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.86085   1.00073 -3.858  0.000118 ***
y          1.40071   0.01863  75.201 < 2e-16 ***
```
```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Residual standard error: 16.19 on 1976 degrees of freedom
### Multiple R-squared:  0.7411, Adjusted R-squared:  0.7409
### F-statistic: 5655 on 1 and 1976 DF,  p-value: < 2.2e-16
The simple expression for the return value gives estimates that are highly correlated with the empirical estimates, but with a low bias. However, with a bias-adjustment based on a scaling factor it can give a crude estimate of the return values based on estimates of $f_w$ and $\mu$, even if the series is fairly short. The discrepancy is due to the heavy-tail characteristics of the daily precipitation and the thinner tail of the exponential distribution. As long as the return period is not too far out, then a scaling factor may be used to relate the probabilities of the different pdfs. The scaling factor is expected to vary with the return interval.

```r
rownames(bias.factors) <- names(RV)
colnames(bias.factors) <- c('Intercept', 'y', 'Std.error intercept', 'Std.error y')
```

```r
bias.cal <- data.frame(x=log(1:10), y=bias.factors[, 2])
bias.lm <- lm(y ~ x, data=bias.cal)
## Adjustment-model details for the return-values:
print(summary(bias.lm))
```
The scaling-factors for bias-corrections are close to linear with respect to $\log(\tau)$ and hence we can extrapolate their values for return-intervals beyond 10 years. The variation of the scaling-factor with return-interval is expected to reflect the differences between the actual tail of the distribution of the 24-hr precipitation and the exponential distribution.

**Dependence of bias-adjustment on return value**

The regression coefficients used to fit the predicted return values to the empirical estimates were expected to depend with the return interval. Here we examine how they vary with different return value $\tau$ from 1 to 10 years.
## Compare the bias factors for the different return values:

```r
plot(rep(1:nrv,3),
     c(bias.factors[,1],bias.factors[,1]+bias.factors[,3],bias.factors[,1]-bias.factors[,3]),
     pch=c(rep(19,nrv),rep(2,nrv),rep(6,nrv)),main='Offset',ylab='',xlab='Years')
for (i in 1:nrv)
     lines(rep(i,2),c(bias.factors[i,1]+bias.factors[i,3],bias.factors[i,1]-bias.factors[i,3]))
ggrid()
```

## Fit the regression coefficients for the slope

```r
#cal.bias <- data.frame(x=1:nrv,y=bias.factors[,2])
#biasfit <- glm(y ~ x, data=cal.bias, family='poisson') # Use a log-link
#print(summary(biasfit))

plot(rep(1:nrv,3),
     c(bias.factors[,2],bias.factors[,2]+bias.factors[,4],bias.factors[,2]-bias.factors[,4]),
     pch=c(rep(19,nrv),rep(2,nrv),rep(6,nrv)),main='Scaling',ylab='',xlab='Years')
for (i in 1:nrv)
     lines(rep(i,2),c(bias.factors[i,2]+bias.factors[i,4],bias.factors[i,2]-bias.factors[i,4]))
#lines(1:nrv,exp(predict(biasfit)),col='red',lty=2)
ggrid()
```
When it comes to the offset (constant term), the best-fit process returned low numbers (less than 10) compared to the the actual values. The offset for the longer return intervals improved monotonously with longer times, but the scaling factor was best for the 1-year return values and increased to 5-year return values, after which it reached a plateau.

**Rate of change in the 10-year-return-value**

It is possible to derive an expression of the rate of change in the return values based on the simple expression $x = \mu \ln(T_f)$ and express the trends (rate of change) in the return values as a function of the trends in the wet-day mean precipitation and the wet-day frequency. Here the units are mm/day per decade for one set of estimates and in percentage for the proportional change.

```r
mFW <- colMeans(FW, na.rm = TRUE)
mMU <- colMeans(MU, na.rm = TRUE)
tFW <- trend.coef(FW)
tMU <- trend.coef(MU)
dx10.dt <- log(3652.5 * mFW) * tMU + mMU / mFW * tFW
print(summary(dx10.dt))
```

```
# Min. 1st Qu. Median Mean 3rd Qu. Max.
# -5.2075 0.1026 0.5435 0.5683 1.0023 6.8697
```

```
print(paste("Fraction of increasing return-values: ", round(100 * sum(dx10.dt > 0) / sum(is.finite(dx10.dt)))))
```

```
[1] "Fraction of increasing return-values: 79"
```
## Proportional change:
\[
\text{dx10.x} \leftarrow \frac{\text{tMU}}{\text{mMU}} + \frac{\text{tFW}}{\text{mFW}} \times \log(3652.5 \times \text{mFW})
\]

print(summary(dx10.x))

## | Min.  | 1st Qu. | Median | Mean   | 3rd Qu. | Max.   |
---|-------|---------|--------|--------|---------|--------|
## | -0.75453 | -0.02054 | 0.05744 | 0.04981 | 0.12863 | 0.93694 |

Q$\text{dx10yr.dt} \leftarrow \text{as.numeric(dx10.dt)}$

\[
\text{esd::map}(Q, \text{FUN}=\text{`dx10yr.dt', cex='number.valid', cex0=0.75, rev=TRUE, new=FALSE})
\]

## Warning in par(par0): graphical parameter "cin" cannot be set
## Warning in par(par0): graphical parameter "cra" cannot be set
## Warning in par(par0): graphical parameter "csi" cannot be set
## Warning in par(par0): graphical parameter "cxy" cannot be set
## Warning in par(par0): graphical parameter "din" cannot be set
## Warning in par(par0): graphical parameter "page" cannot be set

\[
\text{dx10yr.dt (mean= 0.57, sd=0.98 [-5.21, 6.87])}
\]

\[
\text{Q$\text{dx10yr.x10yr} \leftarrow \text{as.numeric(dx10.x)}$
\]

\[
\text{esd::map}(Q, \text{FUN}=\text{`dx10yr.x10yr', cex='number.valid', cex0=0.75, rev=TRUE, new=FALSE})
\]

## Warning in par(par0): graphical parameter "cin" cannot be set
## Warning in par(par0): graphical parameter "cra" cannot be set
## Warning in par(par0): graphical parameter "csi" cannot be set
The results indicate a general increase in the 10-year-return-value ranging from 22mm/day to 155mm/day per decade. The greatest changes are found in the southern states of the USA, around the Mediterranean, parts of Argentina and the east coast of Australia, based on the trends in the wet-day mean precipitation and the wet-day frequency.

For the results of propositional change, we see a typical increase in the 10-year-return-value of around 1.5%/decade before bias-adjustment, which would indicate about 2%/decade for the best estimate.

**Ideal case**

According to the expression \( \frac{\delta x}{x} = \delta \mu / \mu + \ln(T_f) \delta f_w / f_w \), the return-value \( x \) changes proportionally with changes in the wet-day mean precipitation \( \mu \) and in a more complicated fashion for changes in the wet-day frequency \( f_w \). We can explore the effect of trends in the key precipitation parameters \( \mu \) and \( f_w \) on the return-value through a simple idealised experiment:

\[
dX.X <- \text{function}(\text{dmu.mu,df.f,f,tau}) \quad \text{return}(\text{dmu.mu + df.f*log(tau*f))}
\]

```r
# Assume trends
dmu.mu <- c(0.01, 0, -0.01)
df.f <- c(0.01,0,-0.01)
f <- c(0.6,0.2)
tau <- (1:10)*365.25
```
n <- length(dmu.mu)*length(df.f)*length(f)
m <- length(tau)
x.x <- matrix(rep(NA,n*m),n,m)
cols <- rep(c('blue','grey','red'),length(df.f)*length(f))
ltys <- (((1:n)-1)/%/%length(dmu.mu))%%(length(df.f))+1
lwds <- rep(3,n); lwds[(n/2+1):n] <- 1
rnames <- rep('',n)
for (i in 1:n) {
  jdm <- (i-1)%/%length(dmu.mu) + 1
  jfd <- ((i-1)%/%length(dmu.mu))%%(length(df.f)) + 1
 jf <- (i-1)%/%(n/length(f)) + 1
  rnames[i] <- paste0('dmu.mu=',dmu.mu[jdm],',',
                      'dfw.fw=',df.f[jfd],',',
                      'f=',f[jf])
  print(paste(i,rnames[i],',
             'mu:',jdm,',jfd,',
             'f:',jf,'- col=',cols[i],'
             'lty=',ltys[i],'
             'lwd=',lwds[i],'
  ))
  }

for (j in 1:m) x.x[i,j] <- 100*dX.x(dmu.mu[jdm],df.f[jfd],f[jf],tau[j])

x.x[is.finite(x.x)] <- NA
rownames(x.x) <- rnames

plot(range(tau)/365.25,range(x.x,na.rm=TRUE),type='n',
     main='Proportional trend in return-values',xlab=expression(paste(tau,' years'))),
     ylab=expression('%'),ylim=c(-20,10))
grid()
for (i in 1:n) lines(tau/365.25,x.x[i,],col=cols[i],lty=ltys[i],lwd=lwds[i])

legend(2,-8,c(expression(paste(frac(delta*mu,mu),' = 1%'))),
         expression(paste(frac(delta*mu,mu),' = 0%'))),
         expression(paste(frac(delta*mu,mu),' = -1%'))),
         col=c('blue','grey','red'),lty=c(1),bty='n',cex=0.7)
legend(5,-8,c(expression(paste(frac(delta*f,f)',' = 1%'))),
         expression(paste(frac(delta*f,f)',' = 0%'))),
         expression(paste(frac(delta*f,f)',' = -1%'))),
         lty=1:3,bty='n',cex=0.7)
legend(7,-8,c('f = 60%','f = 20%'),lty=1,lwd=c(3,1),bty='n',cex=0.7)
The results suggest that trends in $f_w$ have a more pronounced effect on the return-values of the longer return-intervals than the shorter ones. The effect of trends in $\mu$ is to shift the entire curve vertically, and if there is not trend in $f_w$, the effect is similar for all return-values. The trend in $f_w$ also has an effect on the vertical position of the curves, and also has an effect on all return values.

```r
# Proportional trend in return-values

tau <- seq(1,500,by=10)*365.25
xmu <- seq(5,30,by=5)
xfw <- seq(0.01,1,by=0.1)
plot(range(tau)/365.25,c(0,max(xmu)*log(max(tau)*max(xfw)))),type='n',
     main='Return-value',xlab='year',ylab='return value (mm/day)'
grid()
cols=heat.colors(length(xmu))
for (i in 1:length(xmu)) {
  for (j in 1:length(xfw)) {
    lines(tau/365.25,xmu[i]*log(tau*xfw[j]),col=cols[i],lwd=2)
  }
}
cols=cm.colors(length(xfw))
for (i in 1:length(xmu)) {
  for (j in 1:length(xfw)) {
    lines(tau/365.25,xmu[i]*log(tau*xfw[j]),col=cols[j],lty=2)
  }
}
```
xmu <- seq(0,30,by=1)
xfw <- seq(0,1,by=0.01)
TAU10 <- matrix(rep(NA,length(xmu)*length(xfw)), length(xmu),length(xfw))
for (i in 1:length(xmu)) {
  for (j in 1:length(xfw)) {
    TAU10[i,j] <- 1.4*xmu[i]*log(3652.5*xfw[j])
  }
}
image(xmu,xfw,TAU10,main='10-year-return-value',
xlab=expression(mu),ylab=expression(f[w]))
points(colMeans(MU,na.rm=TRUE),colMeans(FW,na.rm=TRUE),pch=19,cex=0.75,col=rgb(1,1,1,0.05))
contour(xmu,xfw,TAU10,add=TRUE)
grid()
The estimates for 10-year return-value varies with both wet-day frequency and wet-day mean precipitation, and bias-adjusted $\mu \ln(T f_w)$ may exceed 300 mm/day for conditions where $\mu = 30 \text{mm/s}$ and $f_w > 0.6$.

**Number of record-breaking events**

The number of record-breaking events provides an additional indication of the trends in extreme events and information about weather the tails of the statistical distribution are constant.

```r
# Number of record-breaking days with rainfall

print(summary(Z$records))
```

|        | Min. | 1st Qu. | Median | Mean  | 3rd Qu. | Max. | NA's |
|--------|------|---------|--------|-------|---------|------|------|
|        | 0.0942 | 0.8930 | 1.1173 | 1.3159 | 1.4001 | 13.8137 | 2694 |

```r
Z$records[Z$records > 5] <- NA

hist(100*Z$records,breaks=seq(0,1500,by=25),xlim=c(0,600),col='grey',
     main='Ratio of number of records compared to stationary climate',xlab='%',
     sub=paste('Mean=',round(mean(Z$records,na.rm=TRUE)),', sigma=',round(sd(Z$records,na.rm=TRUE))))
grid()
```
Ratio of number of records compared to stationary climate

\begin{verbatim}
esd::map(Z, FUN='records', cex='number.valid', cex0=0.75, rev=TRUE, new=FALSE)

## Warning in par(par0): graphical parameter "cin" cannot be set
## Warning in par(par0): graphical parameter "cra" cannot be set
## Warning in par(par0): graphical parameter "csi" cannot be set
## Warning in par(par0): graphical parameter "cxy" cannot be set
## Warning in par(par0): graphical parameter "din" cannot be set
## Warning in par(par0): graphical parameter "page" cannot be set
\end{verbatim}
A comparison between the counted number of record-breaking daily rainfall amounts and the expected number for a timeseries of same length for a variable that is independently and identically distributed (iid) in terms of percentage suggests that the upper tail of the daily rainfall has stretched to higher values over the recent past.

The geographical distribution of the estimates of daily precipitation variance $\sigma^2$ indicated high values generally near coasts and rain forest regions with observations, but also in a band across the USA.

The analysis of the rain gauge data $\frac{d\sigma^2}{dt}$ indicated that the variance has been increasing over most of the world and that there are only a few sites with a decrease.

Using the results for sites with short time series.

We can investigate to see how many years are necessary for obtaining a good estimate of $\mu$ and $f_w$.

```r
wm <- annual(bjornholt,FUN='wetmean')
wf <- annual(bjornholt,FUN='wetfreq')
plot(cumsum(wm)/1:length(wm),new=FALSE); grid()
lines(range(index(wm)),rep(mean(wm,na.rm=TRUE),2),lty=2)
```
plot(cumsum(wf)/1:length(wf), new=FALSE); grid()
lines(range(index(wf)), rep(mean(wf, na.rm=TRUE), 2), lty=2)
For Bjørnholt, the minimum length is about 40 years to get a good estimate of $\bar{\mu}$, but estimate for $f_w$ seems to be non-stationary.

The effect of changing sample size on block-maximum statistics.

The following chunk demonstrates how the block-maximum varies with sample size:

```r
n <- 10*seq(1,1000,length=100)
q <- rep(NA,length(n))
for (i in 1:length(n)) {
  q[i] <- max(rnorm(n[i]))
}
plot(n,q,main='Block-maximum vs sample size for identical distribution',
     xlab='sample size',ylab='block maximum')
abline(lm(q ~ n),lty=2,col='red')
grid()
```
Block–maximum vs sample size for identical distribution