Correlation Analysis climate and rainfall

Willem Vervoort, Chun Liang, Floris van Ogtrop

17 December 2018

Introduction

This document only covers the analysis of the rainfall and climate in terms of correlation and basic statistics. This is a rewrite of the original paper analysis, but uses the SILO data rather than the BOM gridded data to check if this makes a difference.

load the climate index data

All the older climate index data

```
# climate indicator data
file1 <- "Data\\BoM\\SOI\\SOI.csv"
SOI <- read_csv(file1)
```

```
## Parsed with column specification:
## cols(
##   Year = col_integer(),
##   Jan = col_double(),
##   Feb = col_double(),
##   Mar = col_double(),
##   Apr = col_double(),
##   May = col_double(),
##   Jun = col_double(),
##   Jul = col_double(),
##   Aug = col_double(),
##   Sep = col_double(),
##   Oct = col_double(),
##   Nov = col_double(),
##   Dec = col_double()
## )
```

```
SOI <- gather(SOI,key="Month", value="SOI",Jan:Dec)
SOI <- SOI[order(as.Date(paste(SOI$Year,SOI$Month,
                                          "01",sep="-"), "%Y-%b-%d")),]
SOI_37Y <- SOI %>%
  filter(Year > 1978 & Year < 2016)
#SST34
file2 <- "Data\\BoM\\SST\\SST34_anomaly.nc"
SST34 <- nc_open(file2)
SST34_var <- ncvar_get(SST34,SST34$var[[1]])
nc_close(SST34)
SST34_df <- tibble(Date = seq.Date(as.Date("1856-01-01"),
                                   as.Date("2018-04-01"),
                                   by="month"),
                   SST34 = SST34_var)
SST34_37Y <- SST34_df %>%
```

```r
  filter(Date >= "1979-01-01" & Date < "2016-01-01")

# SST3
file3 <- "Data\\BoM\\SST\\SST3_anomaly.nc"
SST3 <- nc_open(file3)
SST3_var <- ncvar_get(SST3,SST3$var[[1]])
nc_close(SST3)
SST3_df <- tibble(Date = seq.Date(as.Date("1856-01-01"),
                                  as.Date("2018-04-01"),
                                  by="month"),
                  SST3 = SST3_var)
SST3_37Y <- SST3_df %>%
  filter(Date >= "1979-01-01" & Date < "2016-01-01")

#SST4
file4 <- "Data\\BoM\\SST\\SST4_anomaly.nc"
SST4 <- nc_open(file4)
SST4_var <- ncvar_get(SST4,SST4$var[[1]])
nc_close(SST4)
SST4_df <- tibble(Date = seq.Date(as.Date("1856-01-01"),
                                  as.Date("2018-04-01"),
                                  by="month"),
                  SST4 = SST4_var)
SST4_37Y <- SST4_df %>%
  filter(Date >= "1979-01-01" & Date < "2016-01-01")

# PDO
file5 <- "Data\\BoM\\PDO\\PDO.latest_1900-.txt"
pdo_mth <- read_table(file5)
```

```
## Parsed with column specification:
## cols(
##   YEAR = col_character(),
##   JAN = col_double(),
##   FEB = col_double(),
##   MAR = col_double(),
##   APR = col_double(),
##   MAY = col_double(),
##   JUN = col_double(),
##   JUL = col_double(),
##   AUG = col_double(),
##   SEP = col_double(),
##   OCT = col_double(),
##   NOV = col_double(),
##   DEC = col_double()
## )
```

```r
pdo_mth <- gather(pdo_mth,key="Month", value="PDO", JAN:DEC)
pdo_mth$YEAR <- substr(pdo_mth$YEAR,1,4)
pdo_mth <- pdo_mth[order(as.Date(paste(pdo_mth$YEAR,pdo_mth$Month,
                                       "01",sep="-"), "%Y-%b-%d")),]
PDO_37Y <- pdo_mth %>%
  filter(YEAR > 1978 & YEAR < 2016)
## deseasonalised PDO
```

```r
PDO_100 <- pdo_mth

#IOD
file7 <- "Data\\BoM\\IOD\\iod_jamstec.txt"
IOD <- read_table2(file7)
```

```
## Parsed with column specification:
## cols(
##   Date = col_character(),
##   West = col_double(),
##   East = col_double(),
##   `DMI(=West-East)` = col_double()
## )
```

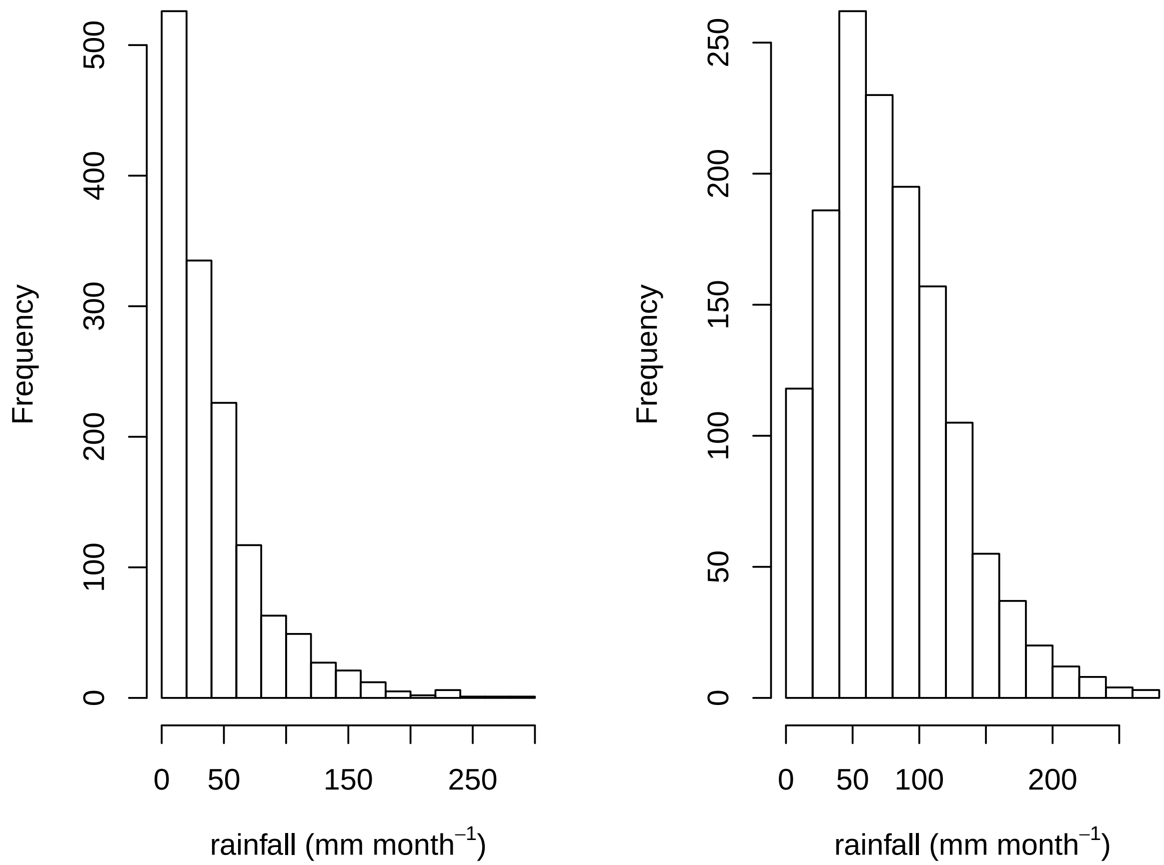
```r
IOD <- tibble(Date = seq.Date(as.Date("1870-01-16"),
                                    as.Date("2018-03-16"),
                                    by="month"),
              IOD = IOD$`DMI(=West-East)`)
IOD_37Y <- IOD %>%
  filter(Date > "1979-01-01" & Date < "2016-01-01")
# end of read data
# should write out useful data
#save.image("../../data/data_environment.Rdata")
```

Download and manage the rainfall data

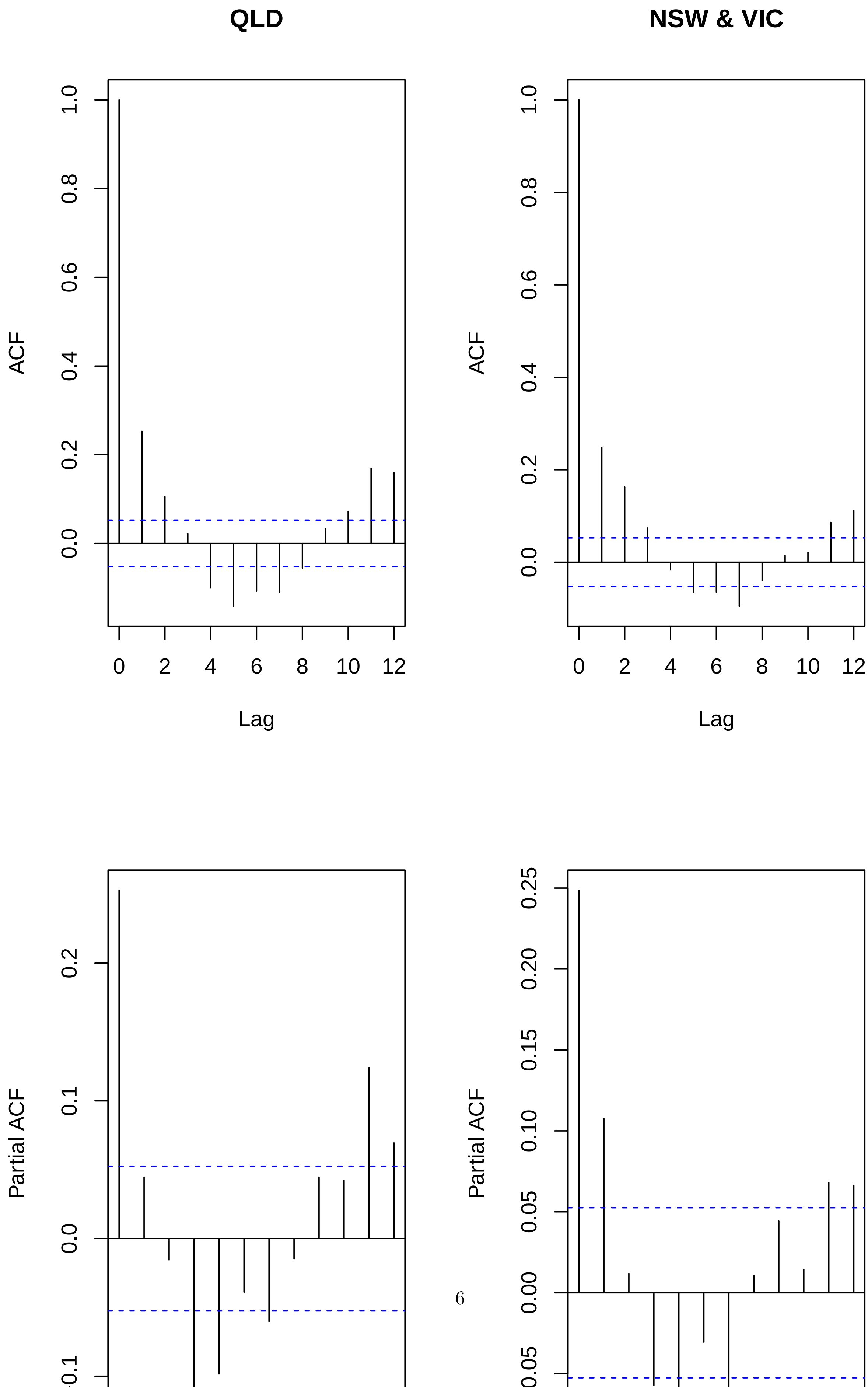
This deals with 100 years of data, deseasonalise and analyse correlations, subset in 37 years and 100 years

```r
# SILO data

## ----extract_Silo------------------------------------------------------
years <- seq(1901,2016,by=1)

SILOdata <- list()

# run a loop
for (i in 1:length(years)) {
   # url <- paste("https://s3-ap-southeast-2.amazonaws.com/silo-open-data/annual/monthly_rain/",
   #              years[i],
   #              ".monthly_rain.nc",sep="")
   # curl_download(url,
   #               paste("data/monthly_rainfall_",years[i],".nc",sep=""))
  SILOdata[[i]] <- brick(paste("data/monthly_rainfall_",years[i],".nc",
                          sep=""),varname="monthly_rain")
}

SILOdata__b <- do.call(stack, SILOdata)

#crop to extents
e111 <- extent(c(145.825, 148.625, -28.075, -25.575))
e222 <- extent(c(147.075, 148.925, -36.675, -35.075))

#  NSW
```

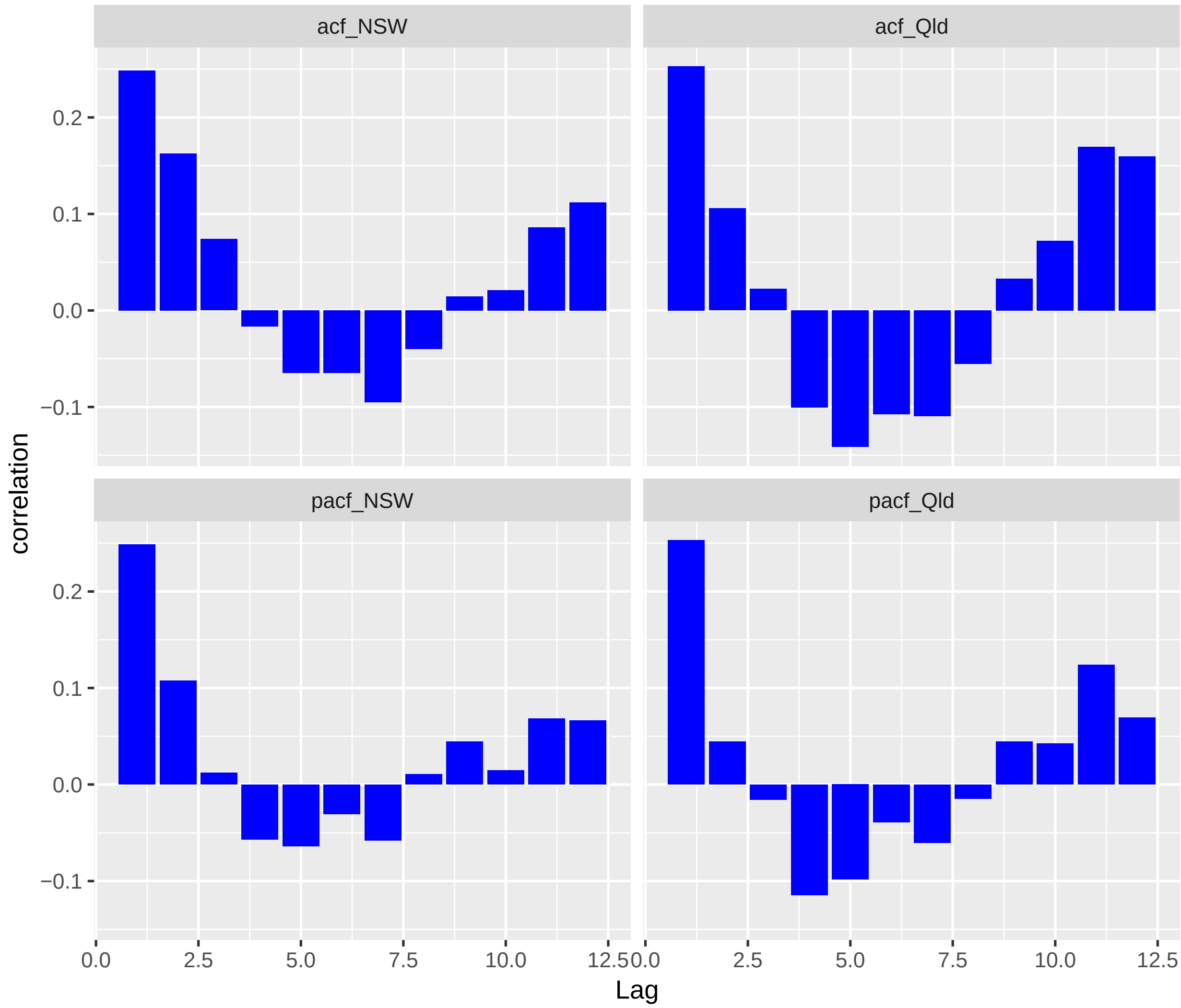
```
SILOdata.NSW <- crop(SILOdata__b, extent(e222))
#QLD
SILOdata.QLD <- crop(SILOdata__b, extent(e111))
# remove orig
rm(SILOdata)
rm(SILOdata__b)

# rewrite as a monthly dataframe
# NSW
rainNSW_t <- as.tibble(data.frame(Date =  seq.Date(as.Date("1901-01-01"),
                                   as.Date("2016-12-31"),by="month"),
                                  rain=t(SILOdata.NSW@data@values)))

# calculate mean by row
nsw_mean <- rainNSW_t %>%
  mutate(Means = rowMeans(.[,-1])) %>%
  dplyr::select(Date,Means)

nsw37_mean <- nsw_mean %>%
 filter(Date >= "1979-01-01")

# Qld
rainQld_t <- as.tibble(data.frame(Date =  seq.Date(as.Date("1901-01-01"),
                                   as.Date("2016-12-31"),by="month"),
                                  rain=t(SILOdata.QLD@data@values)))

# calculate mean by row
qld_mean <- rainQld_t %>%
  mutate(Means = rowMeans(.[,-1])) %>%
  dplyr::select(Date,Means)

qld37_mean <- qld_mean %>%
 filter(Date >= "1979-01-01")


rm(SILOdata.NSW)
rm(SILOdata.QLD)

# some simple stats graphs
#jpeg("../../figures/hist_rainfallSilo.jpg", width=960, height=480)
par(mfrow=c(1,2))
hist(qld_mean$Means,main="",xlab=expression(paste("rainfall (mm mont",h^-1,")",sep="")))
 hist(nsw_mean$Means,main="",xlab=expression(paste("rainfall (mm mont",h^-1,")",sep="")))
```




```
#dev.off()


# autocorrelation in rainfall
acf_df <- data_frame(lag = 1:12,
                     acf_Qld = acf(qld_mean$Means,
                                   lag.max=12,main="QLD")$acf[2:13,1,1],
                     acf_NSW = acf(nsw_mean$Means,
                                   lag.max=12,main="NSW & VIC")$acf[2:13,1,1],
                     pacf_Qld = pacf(qld_mean$Means,
                                     lag.max=12,main="")$acf[,1,1],
                     pacf_NSW = pacf(nsw_mean$Means,
                                     lag.max=12,main="")$acf[,1,1])
```


QLD
NSW & VIC
ACF
Lag
Partial ACF

```r
#jpeg("../../figures/autocorrelationSilo.jpg")
acf_df %>%
  gather(key="acf", value="cor", acf_Qld:pacf_NSW) %>%
  ggplot(aes(lag, cor)) + geom_bar(stat="identity", fill = "blue") +
  facet_wrap(~acf) + xlab("Lag") + ylab("correlation")
```




```r
#dev.off()
```

deseasonalise and correlations with climate indices

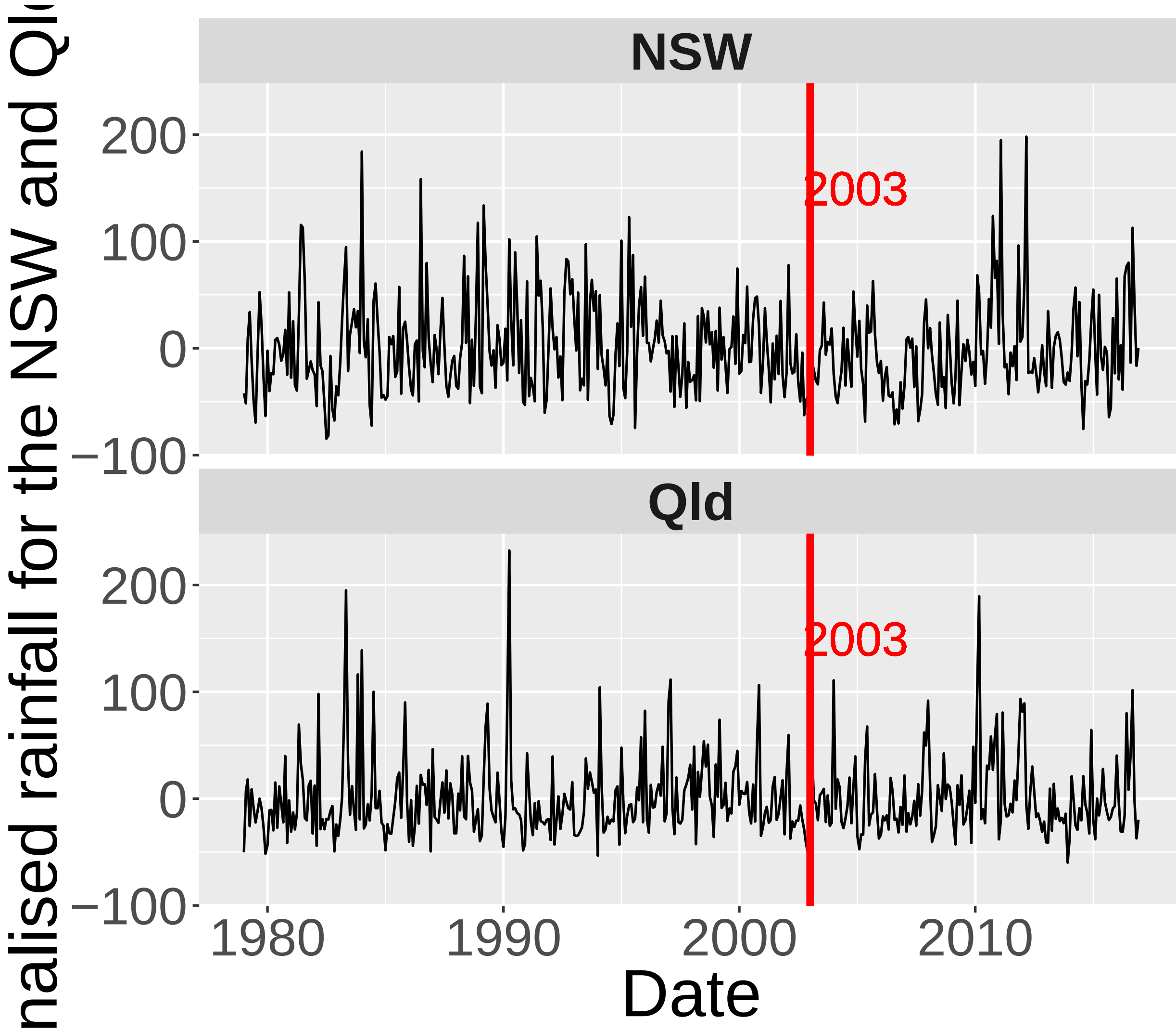
We can now deseasonalise using the package deseasonalise

```r
require(deseasonalize)
```

```
## Loading required package: deseasonalize
## Loading required package: lattice
## Loading required package: FitAR
## Loading required package: leaps
```

```
## Loading required package: ltsa
## Loading required package: bestglm
```

```
## deseasonalise & detrend rainfall: assume the deseasonalised monthly zonal rainfall is normal distribu

qld37_deseason <- ds(qld37_mean$Means,ic="AIC", type="monthly",
                     standardizeQ=F)$z
nsw37_deseason <- ds(nsw37_mean$Means,ic="AIC", type="monthly",
                     standardizeQ=F)$z
qld100_deseason <- ds(qld_mean$Means,ic="AIC", type="monthly",
                      standardizeQ=F)$z
nsw100_deseason <- ds(qld_mean$Means,ic="AIC", type="monthly",
                      standardizeQ=F)$z

# Show a plot
plot_df <- tibble(Dates = rep(qld37_mean$Date,2),
                  Resid = c(qld37_deseason,nsw37_deseason),
                  State = c(rep("Qld",length(qld37_deseason)),
                            rep("NSW",length(nsw37_deseason))))

#jpeg("../../figures/Rainfall_resid.jpg", width=960, height=960)
plot_df %>%
  ggplot(aes(Dates,Resid)) + geom_line() +
  geom_vline(xintercept = as.Date("2003-01-01"),
             col="red", size=1.5) +
  geom_text(aes(x=as.Date("2004-12-01"),y=150),
            label="2003", col="red", size=7) +
  facet_wrap(~State, ncol = 1) + xlab("Date") +
  ylab("Deseasonalised rainfall for the NSW and Qld locations") +
  theme(axis.text=element_text(size=rel(2)),
        axis.title=element_text(size=rel(2.5)),
        strip.text =element_text(size=rel(2), face="bold"))
```




```
#dev.off()
```

trends in the deseasonalised rainfall

Investigate if any of the detrended rainfall has a trend

```
# test whether any linear trends are significant
df_list <- list(qld37 = data.frame(z=qld37_deseason,
                           trend=1:length(qld37_deseason)),
                nsw37 = data.frame(z=nsw37_deseason,
                           trend=1:length(nsw37_deseason)),
                qld100 = data.frame(z=qld100_deseason,
                 trend=1:length(qld100_deseason)),
                nsw100 = data.frame(z=nsw100_deseason,
                 trend=1:length(nsw100_deseason)))
result <- lapply(df_list,
                 function(x) summary(lm(z~trend,
                                        data=x))$coefficients)
result
```

```
## $qld37
##                 Estimate Std. Error    t value  Pr(>|t|)
## (Intercept) -2.062987954 3.51082757 -0.5876073 0.5570880
## trend        0.009028394 0.01331348  0.6781392 0.4980289
##
## $nsw37
##                 Estimate Std. Error    t value  Pr(>|t|)
## (Intercept)  0.932965830 4.14356215  0.2251603 0.8219559
## trend       -0.004083001 0.01571289 -0.2598505 0.7950969
##
## $qld100
##                  Estimate  Std. Error    t value  Pr(>|t|)
## (Intercept) -0.4000354035 2.014688477 -0.1985594 0.8426364
## trend        0.0005743509 0.002505506  0.2292355 0.8187196
##
## $nsw100
##                  Estimate  Std. Error    t value  Pr(>|t|)
## (Intercept) -0.4000354035 2.014688477 -0.1985594 0.8426364
## trend        0.0005743509 0.002505506  0.2292355 0.8187196
```

None have a significant trend

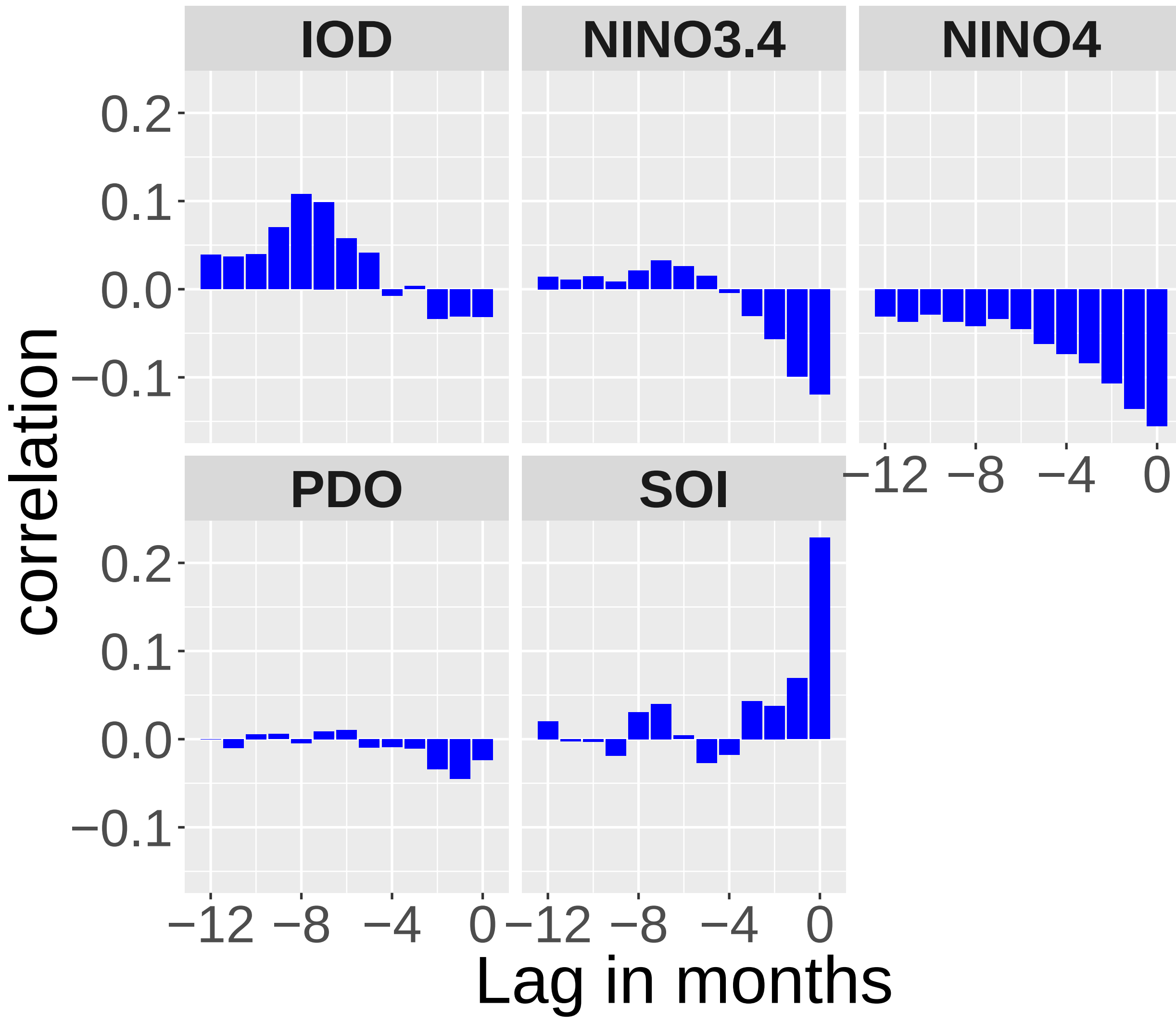
correlations with climate indices

Correlate the deseasonalised rainfall with the climate indices

```
# Figure 5
plot_df_qld <- tibble(Lag = ccf(as.vector(SOI_37Y$SOI),
                            as.vector(qld37_deseason),lag.max=12,
                            plot=F)$lag[,,1],
                  SOI = ccf(as.vector(SOI_37Y$SOI),
                            as.vector(qld37_deseason),lag.max=12,
                            plot=F)$acf[,,1],
                  NINO3.4 = ccf(as.vector(SST34_37Y$SST34),
                                as.vector(qld37_deseason),lag.max=12,
                                plot=F)$acf[,,1],
                  NINO4 = ccf(as.vector(SST4_37Y$SST4),
                              as.vector(qld37_deseason),lag.max=12,
                              plot=F)$acf[,,1],
                  PDO = ccf(as.vector(PDO_100$PDO),
                            as.vector(qld100_deseason),lag.max=12,
                            plot=F)$acf[,,1],
                  IOD = ccf(as.vector(IOD_37Y$IOD),
                            as.vector(qld37_deseason),lag.max=12,
                            plot=F)$acf[,,1])

#jpeg("../../figures/cor_qld.jpg", width=960, height=960)
plot_df_qld %>%
  gather(key="SST", value="correlation", SOI:IOD) %>%
  ggplot(aes(Lag,correlation)) +
  geom_bar(fill="blue", stat="identity") +
  facet_wrap(~SST) + xlab("Lag in months") + ylab("correlation") +
  xlim(c(-12.5,0.5)) +
  theme(axis.text=element_text(size=rel(2)),
```

```
axis.title=element_text(size=rel(2.5)),
strip.text =element_text(size=rel(2), face="bold"))
```

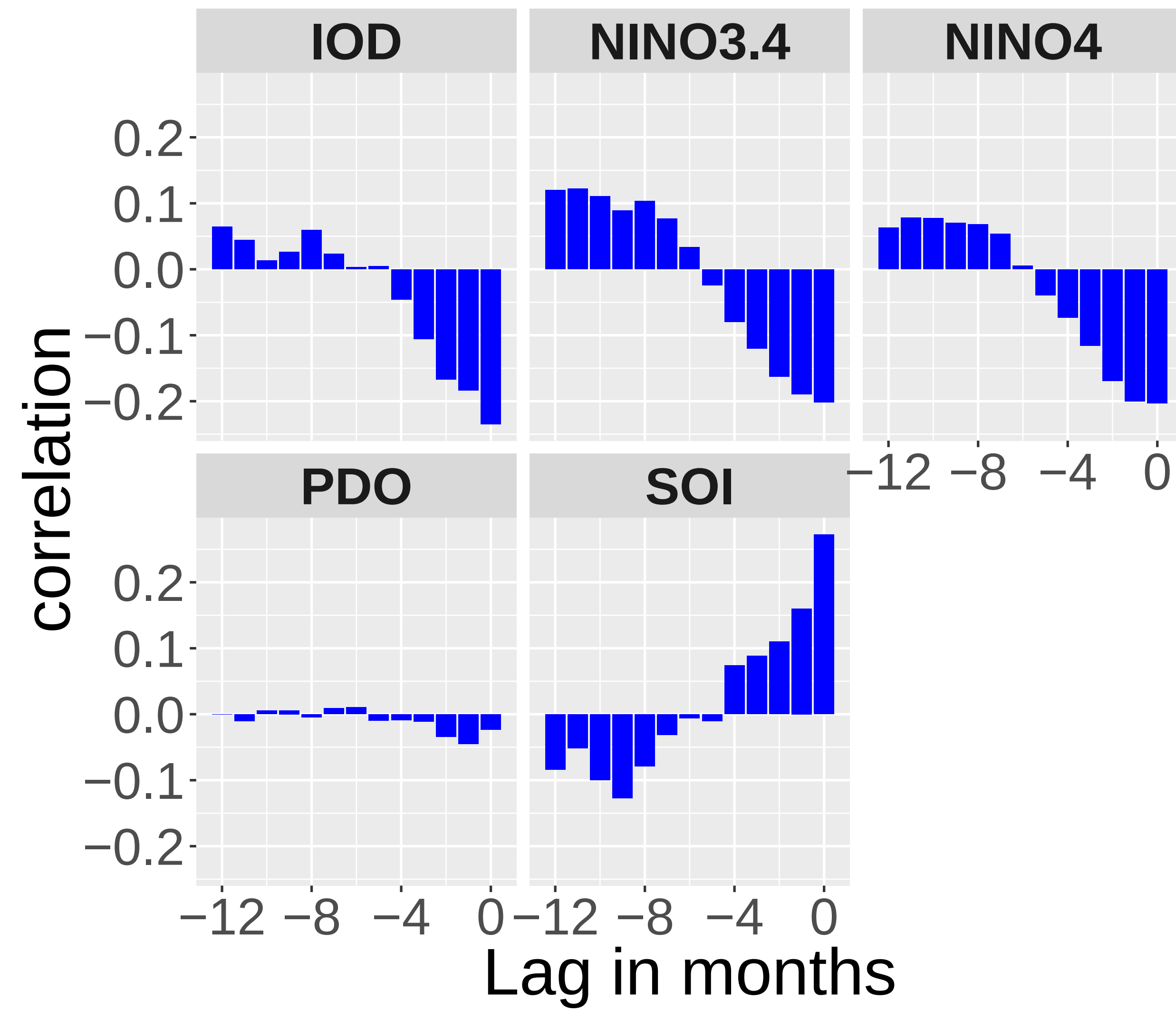
```
## Warning: Removed 60 rows containing missing values (position_stack).
```




```
#dev.off()

# to add and calculate ci: qnorm((1 + 0.95)/2)/sqrt(x$n.used)

plot_df <- tibble(Lag = ccf(as.vector(SOI_37Y$SOI),
                            as.vector(nsw37_deseason),lag.max=12,
                            plot=F)$lag[,,1],
                  SOI = ccf(as.vector(SOI_37Y$SOI),
                            as.vector(nsw37_deseason),lag.max=12,
                            plot=F)$acf[,,1],
                  NINO3.4 = ccf(as.vector(SST34_37Y$SST34),
                                as.vector(nsw37_deseason),lag.max=12,
                                plot=F)$acf[,,1],
                  NINO4 = ccf(as.vector(SST4_37Y$SST4),
                              as.vector(nsw37_deseason),lag.max=12,
```

```
                              plot=F)$acf[,,1],
                  PDO = ccf(as.vector(PDO_100$PDO),
                            as.vector(nsw100_deseason),lag.max=12,
                            plot=F)$acf[,,1],
                  IOD = ccf(as.vector(IOD_37Y$IOD),
                            as.vector(nsw37_deseason),lag.max=12,
                            plot=F)$acf[,,1])

#jpeg("../../figures/cor_nswvic.jpg", width=960, height=960)
plot_df %>%
  gather(key="SST", value="correlation", SOI:IOD) %>%
  ggplot(aes(Lag,correlation)) +
  geom_bar(fill="blue", stat="identity") +
  facet_wrap(~SST) + xlab("Lag in months") + ylab("correlation") +
  xlim(c(-12.5,0.5)) +
  theme(axis.text=element_text(size=rel(2)),
        axis.title=element_text(size=rel(2.5)),
        strip.text =element_text(size=rel(2), face="bold"))
```

```
## Warning: Removed 60 rows containing missing values (position_stack).
```




```
#dev.off()
```

From the paper:

Correlations between rainfall and each climate index were analysed. Rainfall in each study region was first deseasonalised and detrended using the seasonal decomposition function `ds` in the package `deseasonalise` in R [@Rstats2018]. Using detrended data gives a better indication of the underlying correlation rather than the correlation between trends in the data [@Smith2012]. The cross-correlations between the deseasonalised and detrended rainfall and the climatic indices were tested using the Pearson's product moment correlation method, assuming the relationships are linear. Although the optimal technique for exploring the correlation with each index could be different as described in @Risbey2009, the Pearson's method was applied to all indices for consistency. Because the PDO describes the multi-decadal SST with lower frequency [@MacDonald2005; @Zanchettin2008; @Kamruzzaman2011], instead of 37-year rainfall data, a longer period (115 years, from 1900 to 2015) was used to estimate the correlation with PDO, up to lag 24. For the other indices, the 37-year data was used.