STAT 443 Assignment 4

Derek Situ

2022-04-03

Question 1

1a

The power spectrum of $\{X_t\}_{t\in\mathbb{Z}}$ is

$$f(\omega) = \frac{1}{\pi}\left[\gamma(0) + 2\sum_{k=1}^{\infty}\gamma(k)cos(\omega k)\right]$$

where

$$\gamma(0) = 1\sigma^2 + 0.7^2\sigma^2 + 0.2^2\sigma^2 + 0.1^2\sigma^2 = 1.54$$

since $\sigma^2 = 1$. Then,

$$\gamma(1) = 0.7 + 0.2 \times 0.7 - 0.1 \times 0.2 = 0.82,$$
$$\gamma(2) = 0.2 - 0.1 \times 0.7 = 0.13,$$
$$\gamma(3) = -0.1,$$

and

$$\gamma(k) = 0, \forall k > 3.$$

Then

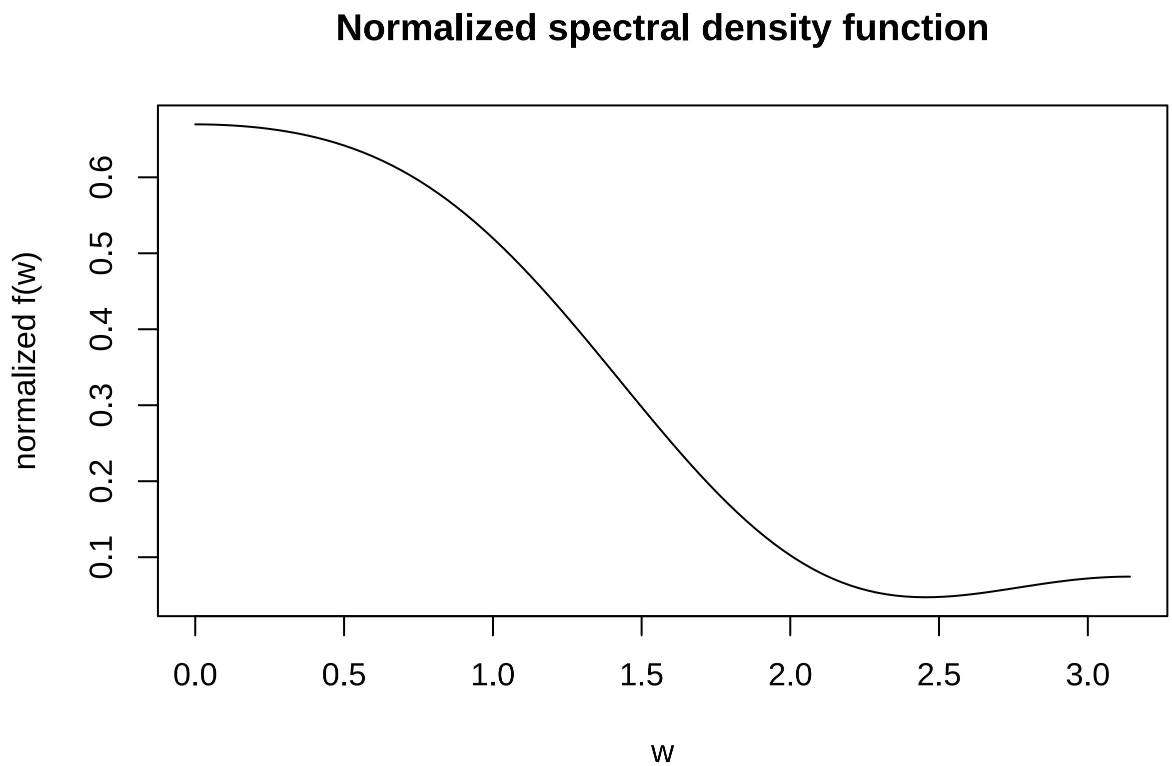
$$f(\omega) = \frac{1}{\pi}\left[1.54 + 2(0.82cos(\omega) + 0.13cos(2\omega) - 0.1cos(3\omega))\right].$$

1b

$$f^*(\omega) = \frac{f(\omega)}{\gamma(0)} = \frac{1}{1.54\pi}\left[1.54 + 2(0.82cos(\omega) + 0.13cos(2\omega) - 0.1cos(3\omega))\right]$$

1c

```
omega <- seq(0, pi, by = 0.01*pi)
norm_spect <- 1/(1.54*pi) * (1.54 + 2*(0.82*cos(omega) + 0.13*cos(2*omega) -
                                        0.1*cos(3*omega)))
plot(x = omega, y = norm_spect,
     type = "l",
     xlab = "w",
     ylab = "normalized f(w)",
     main = "Normalized spectral density function")
```



The normalized spectral density function is high for low frequencies and decreases as omega increases.

Question 2

2a

Proof: $f_W(\omega) = f_X(\omega) + f_Y(\omega)$

From the definition of a spectrum,

$$f_W(\omega) = \frac{1}{\pi}\left[Var(W_t) + 2\sum_{k=1}^{\infty} Cov(W_{t+k}, W_t)cos(\omega k)\right]$$

where

$$\begin{aligned} Var(W_t) = Var(X_t + Y_t) &= Var(X_t) + Var(Y_t) + 2Cov(X_t, Y_t) \\ &= Var(X_t) + Var(Y_t) \end{aligned}$$

by independence of X and Y, and

$$Cov(W_{t+k}, W_t) = Cov(X_{t+k} + Y_{t+k}, X_t + Y_t)$$
$$= Cov(X_{t+k}, X_t) + Cov(Y_{t+k}, Y_t) + Cov(X_{t+k}, Y_t) + Cov(Y_{t+k}, X_t)$$
$$= Cov(X_{t+k}, X_t) + Cov(Y_{t+k}, Y_t)$$

by independence of X and Y. Then,

$$f_W(\omega) = \frac{1}{\pi}\left[Var(X_t) + Var(Y_t) + 2\sum_{k=1}^{\infty} Cov(X_{t+k}, X_t)cos(\omega k) + 2\sum_{k=1}^{\infty} Cov(Y_{t+k}, Y_t)cos(\omega k)\right]$$

$$= \frac{1}{\pi}\left[Var(X_t) + 2\sum_{k=1}^{\infty} Cov(X_{t+k}, X_t)cos(\omega k)\right] + \frac{1}{\pi}\left[Var(Y_t) + 2\sum_{k=1}^{\infty} Cov(Y_{t+k}, Y_t)cos(\omega k)\right]$$

$$= f_X(\omega) + f_Y(\omega).$$

Proof: $W_t = X_t + Y_t$ is stationary

Assume $\{X_t\}$ and $\{Y_t\}$ are independent and stationary. Then $\forall t$, $E(X_t) = \mu_1$ and $E(Y_t) = \mu_2$ for some constants μ_1, μ_2. Also, $Var(X_t), Var(Y_t) < \infty$, and $Cov(X_t, X_{t+h}) = g_1(h)$ and $Cov(Y_t, Y_{t+h}) = g_2(h)$. We want to show that $\{W_t\}$ has a constant mean, a finite variance, and an autocovariance function that only depends on the lag h. We see that $E(W_t) = \mu_1 + \mu_2$ which is a constant, satisfying the first criterion. We also have $Var(W_t) = Var(X_t) + Var(Y_t)$ by independence of $\{X_t\}$ and $\{Y_t\}$, which is the sum of two finite variances, therefore is finite itself. Thus the second criterion is satisfied. Finally, we saw in the previous proof that $Cov(W_t, W_{t+h}) = Cov(X_{t+k}, X_t) + Cov(Y_{t+k}, Y_t) = g_1(h) + g_2(h)$ by independence of $\{X_t\}$ and $\{Y_t\}$, which only depends on the lag h, therefore the third and last criterion is satisfied.

2b

Assume X and Y are independent and use the result from (2a) to get

$$f_W(\omega) = f_X(\omega) + f_Y(\omega)$$

where

$$f_Y(\omega) = \frac{1}{\pi}\left[Var(Y_t) + 2\sum_{k=1}^{\infty} Cov(Y_{t+k}, Y_t)cos(\omega k)\right]$$

$$= \frac{\sigma^2}{\pi}$$

since for a white noise process Y_t, $Cov(Y_{t+k}, Y_t) = 0, \forall k \neq 0$. Also,

$$f_X(\omega) = \frac{1}{\pi}\left[Var(X_t) + 2\sum_{k=1}^{\infty} Cov(X_{t+k}, X_t)cos(\omega k)\right]$$

where

$$Var(X_t) = Var(\frac{1}{1 + 0.5B} Z_t)$$

$$= Var(Z_t) + 0.5^2 Var(Z_{t-1}) + 0.5^4 Var(Z_{t-2}) + ...$$

$$= \frac{\sigma^2}{1 - 0.5^2},$$

and

$$Cov(X_{t+k}, X_t) = E\left(\sum_{i=0}^{\infty} -0.5^i Z_{t-i} \times \sum_{j=0}^{\infty} -0.5^j Z_{t+k-j}\right)$$

$$= \sigma^2 \sum_{i=0}^{\infty} -0.5^{2i+k}$$

$$= \sigma^2 \frac{-0.5^k}{1 - 0.5^2}$$

recalling that $E(Z_{t-i}, Z_{t-j}) \neq 0$ when $i = j$. Then,

$$f_X(\omega) = \frac{1}{\pi}\left[\frac{\sigma^2}{1 - 0.5^2} + 2\sum_{k=1}^{\infty} \frac{-0.5^k}{1 - 0.5^2}\sigma^2 cos(\omega k)\right].$$

With these expressions for $f_Y(\omega)$ and $f_X(\omega)$, we see that

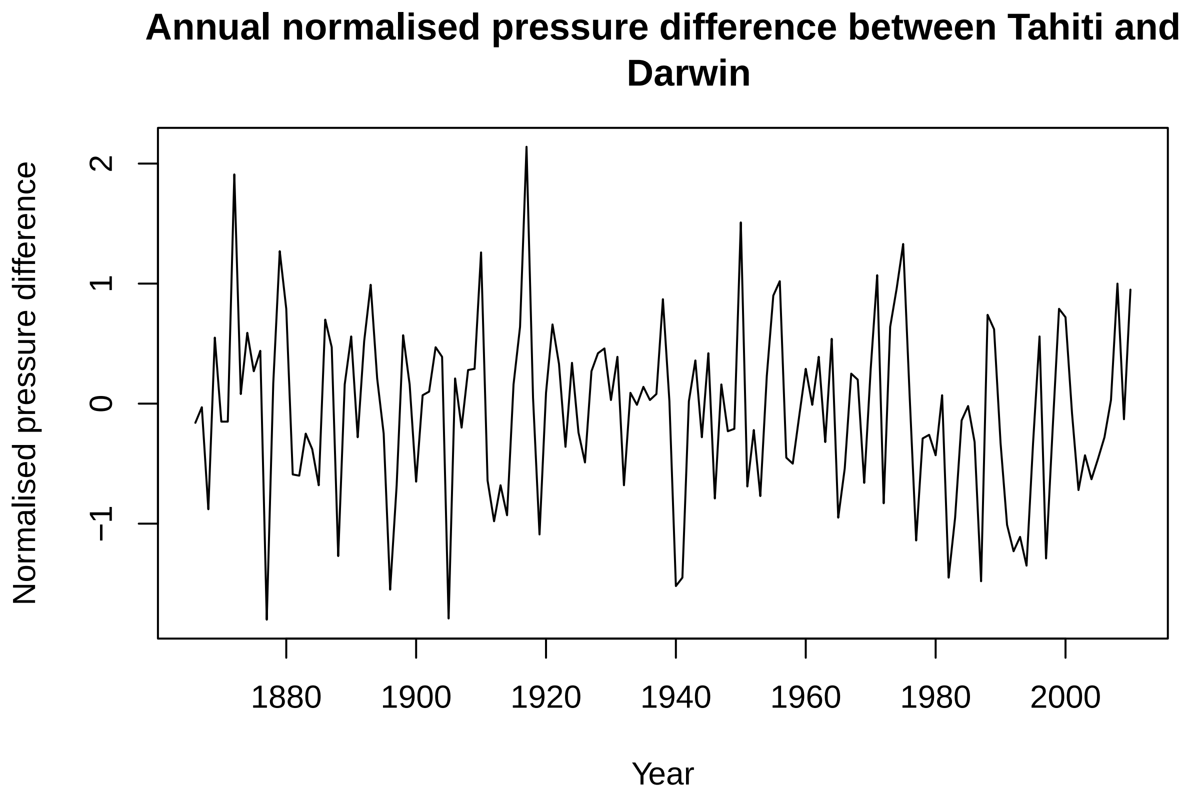
$$f_W(\omega) = f_Y(\omega) + f_X(\omega)$$

$$= \frac{\sigma^2}{\pi} + \frac{1}{\pi}\left[\frac{\sigma^2}{1-0.5^2} + 2\sum_{k=1}^{\infty}\frac{-0.5^k}{1-0.5^2}\sigma^2 cos(\omega k)\right].$$

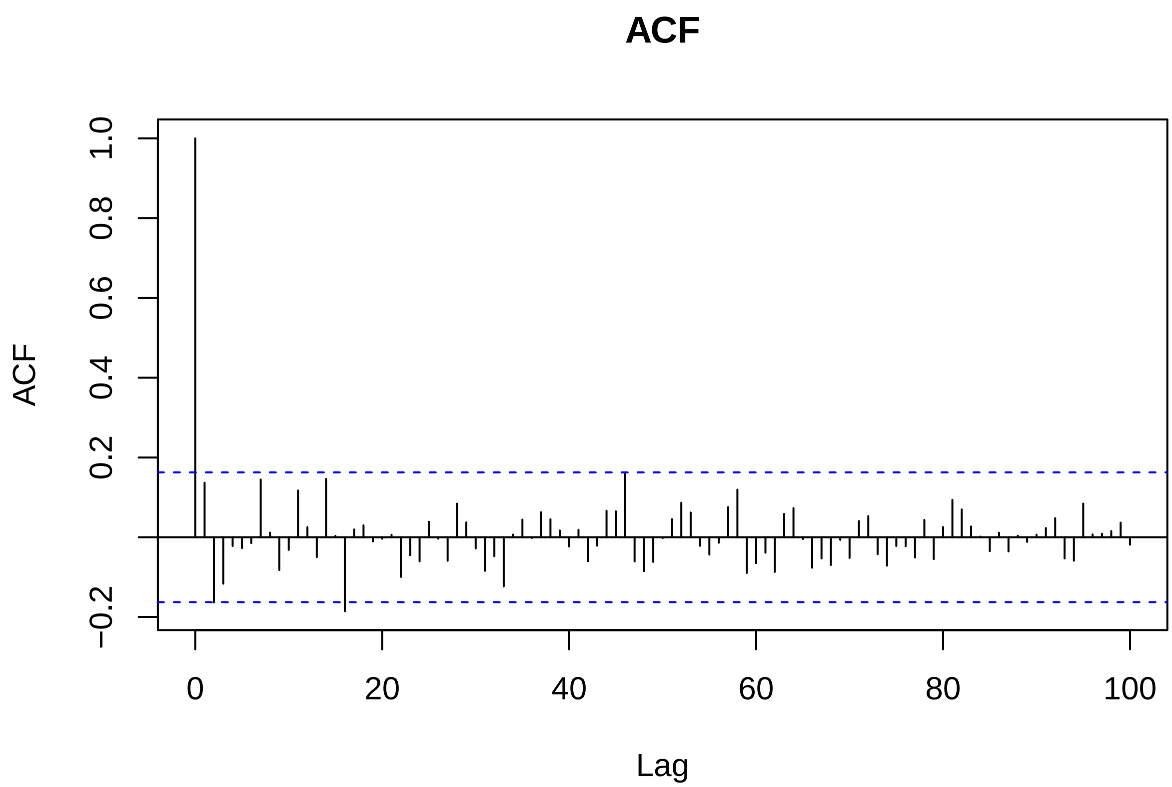
Question 3

3a

```r
soi <- read.delim("soi.txt", header = TRUE, sep = "")
annual <- ts(soi$annual, start=soi$year[1], end = soi$year[length(soi$year)])
plot(annual,
     main = "Annual normalised pressure difference between Tahiti and
     Darwin",
     xlab = "Year",
     ylab = "Normalised pressure difference")
```



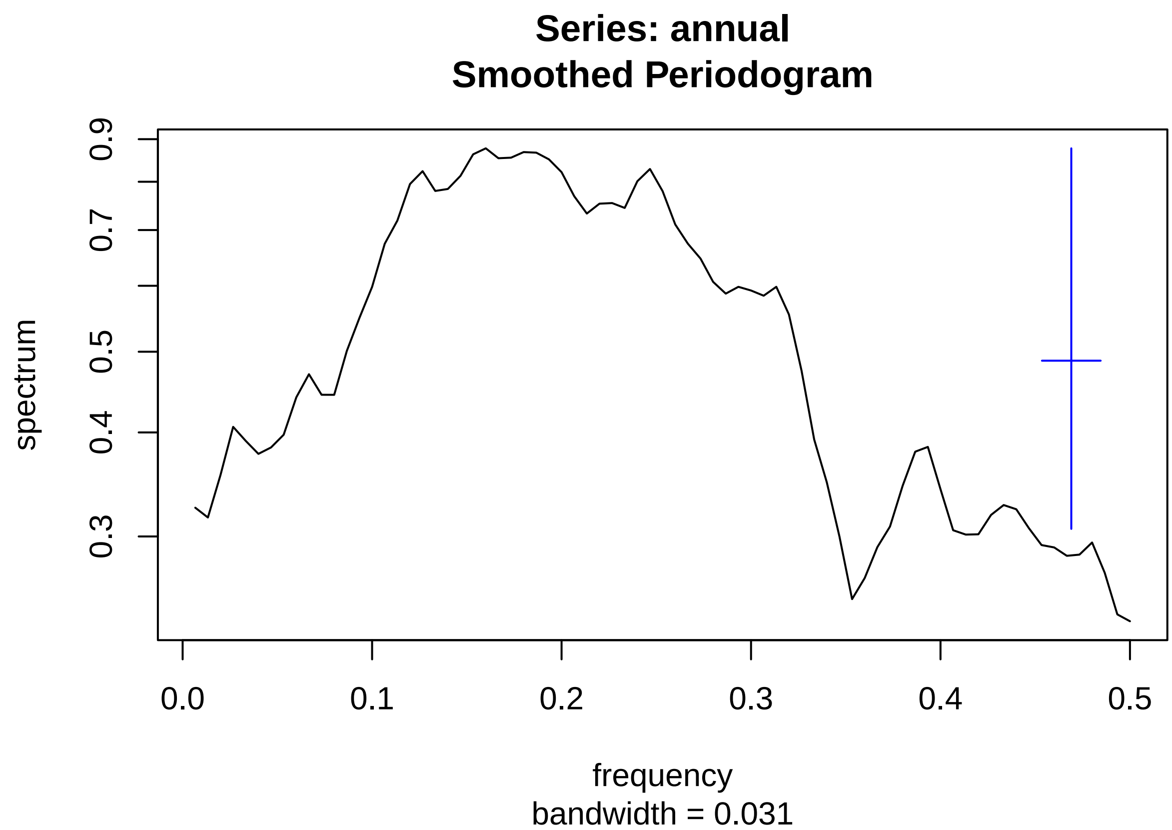

```r
acf(annual, lag.max = 100, main = "ACF")
```



From the plot, we see that the data appears to have some cyclic variation. The acf tails off without pattern.

3b

```
pgram <- spec.pgram(annual, spans = sqrt(2 * length(annual)))
```




```
pgram$freq[which.max(pgram$spec)] # frequency of the model where periodogram is max
```

```
## [1] 0.16
```

```
pgram$freq[which.max(pgram$spec)] * 2*pi # angular frequency where periodogram is max
```

```
## [1] 1.00531
```

The periodogram generally increases until it reaches its maximum, and then generally decreases. It reaches a maximum at an angular frequency of

$$\omega = 0.16(2\pi) = 0.32\pi$$

and the wavelength is

$$\ell = \frac{2\pi}{\omega} = \frac{2\pi}{0.16(2\pi)} = \frac{1}{0.16} = 6.25$$

years.

3c

```
# Inputs: a time series 'ts', and a constant 'p' in {0, 1, ..., N/2}
# Output: the angular frequency corresponding to p for the time series ts
```

```r
get_freq <- function(ts, p) {
  (2 * pi) * (p / length(ts))
}

get_freq(annual, 10)
```

```
## [1] 0.4333231
```

The output of the function is 0.4333231.

3d

```r
# gets the f statistic of a model
get_fstat <- function(model) {
  summary(model)[["fstatistic"]][["value"]]
}

# identifies values of p giving significant Fourier frequencies,
#   and puts them in the vector P
t <- c(1:length(annual))
P <- c()

for (p in 1:73) {

  model <-
    lm(annual ~ cos(get_freq(annual, p) * t) + sin(get_freq(annual, p) * t),
       data = soi)

  if (get_fstat(model) > qf(0.95, df1 = 2, df2 = 145 - 3)) {
    P <- c(P, p)
  }
}

P
```

```
## [1] 16 20 23 25 41
```

The following values of p give significant Fourier frequencies: 16, 20, 23, 25, 41

3e

```r
model_data <-
  data.frame(
    annual = soi$annual,
    cos_term_1 = cos(get_freq(annual, P[1]) * t),
    cos_term_2 = cos(get_freq(annual, P[2]) * t),
    cos_term_3 = cos(get_freq(annual, P[3]) * t),
    cos_term_4 = cos(get_freq(annual, P[4]) * t),
```

```r
    cos_term_5 = cos(get_freq(annual, P[5]) * t),
    sin_term_1 = sin(get_freq(annual, P[1]) * t),
    sin_term_2 = sin(get_freq(annual, P[2]) * t),
    sin_term_3 = sin(get_freq(annual, P[3]) * t),
    sin_term_4 = sin(get_freq(annual, P[4]) * t),
    sin_term_5 = sin(get_freq(annual, P[5]) * t)
  )

model <-
  lm(annual ~ cos_term_1 + cos_term_2 + cos_term_3 + + cos_term_4 +
       cos_term_5 + sin_term_1 + sin_term_2 + sin_term_3 + sin_term_4 +
       sin_term_5,
     data = model_data)
summary(model)
```

```
##
## Call:
## lm(formula = annual ~ cos_term_1 + cos_term_2 + cos_term_3 +
##     +cos_term_4 + cos_term_5 + sin_term_1 + sin_term_2 + sin_term_3 +
##     sin_term_4 + sin_term_5, data = model_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.61999 -0.41957  0.05496  0.44940  1.74653
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.05683    0.05344  -1.063  0.28953
## cos_term_1   0.12266    0.07558   1.623  0.10693
## cos_term_2   0.21402    0.07558   2.832  0.00534 **
## cos_term_3  -0.07272    0.07558  -0.962  0.33767
## cos_term_4  -0.15147    0.07558  -2.004  0.04707 *
## cos_term_5   0.21868    0.07558   2.893  0.00445 **
## sin_term_1  -0.22334    0.07558  -2.955  0.00369 **
## sin_term_2  -0.03790    0.07558  -0.501  0.61689
## sin_term_3   0.20607    0.07558   2.727  0.00726 **
## sin_term_4  -0.16460    0.07558  -2.178  0.03117 *
## sin_term_5  -0.14669    0.07558  -1.941  0.05437 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6435 on 134 degrees of freedom
## Multiple R-squared:  0.2673, Adjusted R-squared:  0.2127
## F-statistic:  4.89 on 10 and 134 DF,  p-value: 5.207e-06
```

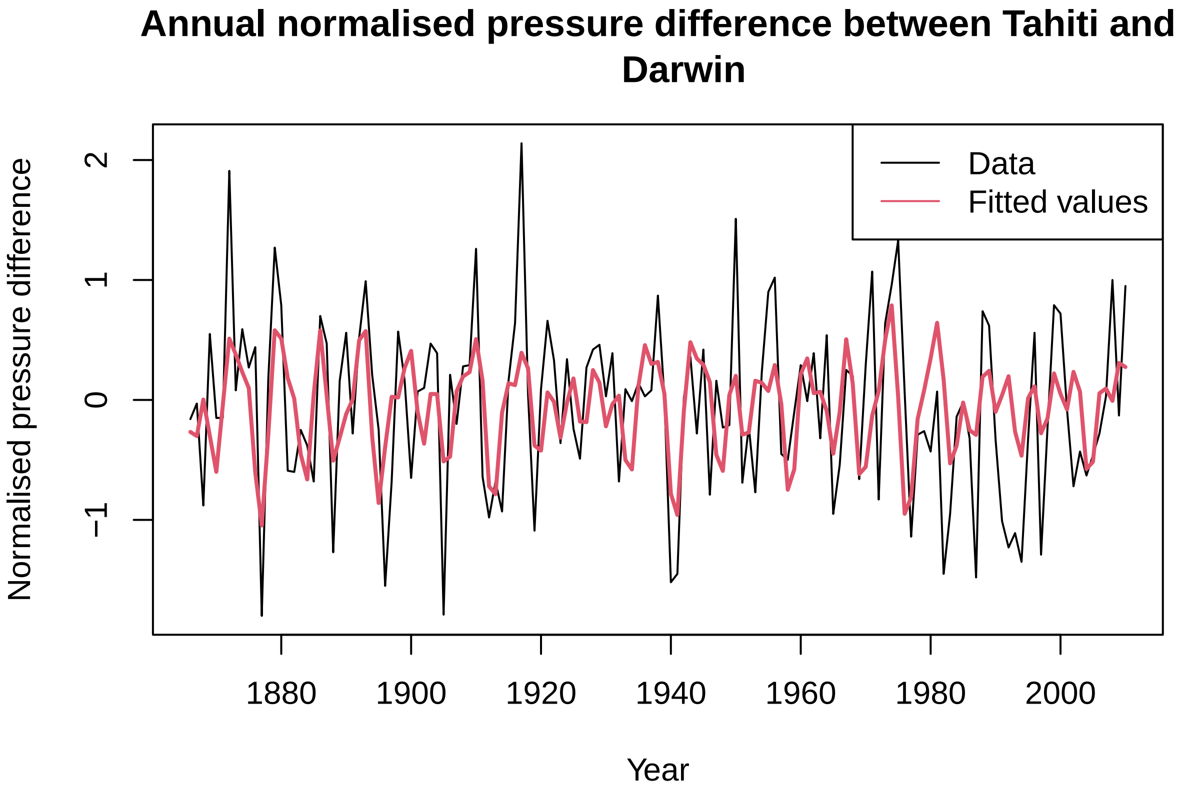
The estimated coefficients for the linear model are summarized below:

	Estimates
alpha0	-0.0568276
alpha1	0.1226644
alpha2	0.2140222
alpha3	-0.0727230

	Estimates
alpha4	-0.1514678
alpha5	0.2186812
beta1	-0.2233396
beta2	-0.0378970
beta3	0.2060698
beta4	-0.1645952
beta5	-0.1466859

3f

```
plot(annual,
     main = "Annual normalised pressure difference between Tahiti and
     Darwin",
     xlab = "Year",
     ylab = "Normalised pressure difference")
lines(x = soi$year,
      y = model$fitted.values,
      type = "l",
      col = 2,
      lwd = 2)
legend("topright",
       legend = c("Data", "Fitted values"),
       col = c(1, 2),
       lwd = 1)
```



Annual normalised pressure difference between Tahiti and Darwin
Data
Fitted values
Normalised pressure difference
−1
0
1
2
1880
1900
1920
1940
1960
1980
2000
Year