



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

MRC

Laboratory of
Molecular Biology

Time Series Analysis

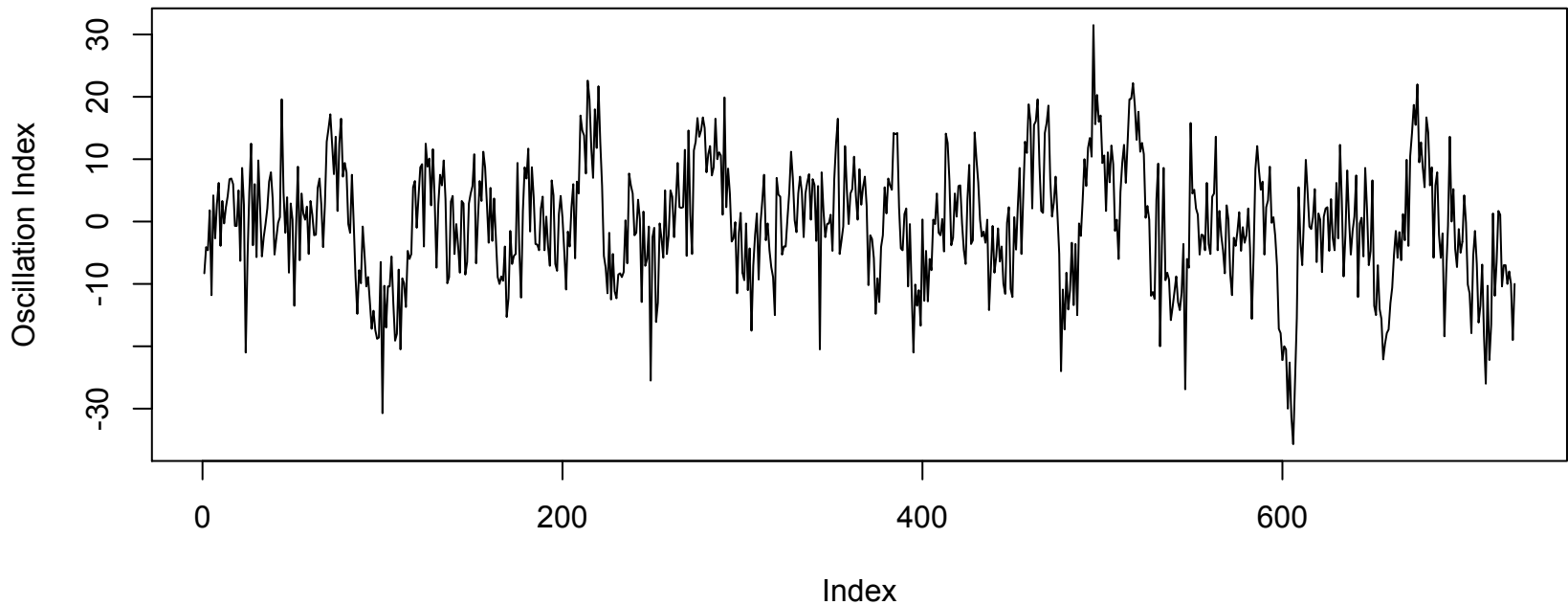
10th of May 2018

R. Nicholls / D.-L. Couturier / M. Fernandes

Time series analysis

Example: Monthly Southern Oscillation Index

Monthly difference in sea-surface air pressure between Darwin and Tahiti



Used for predicting rainfall in parts of Australia

Time series analysis

A time series is a process in which a given observation depends on other datapoints in the same series.

Linear regression models:

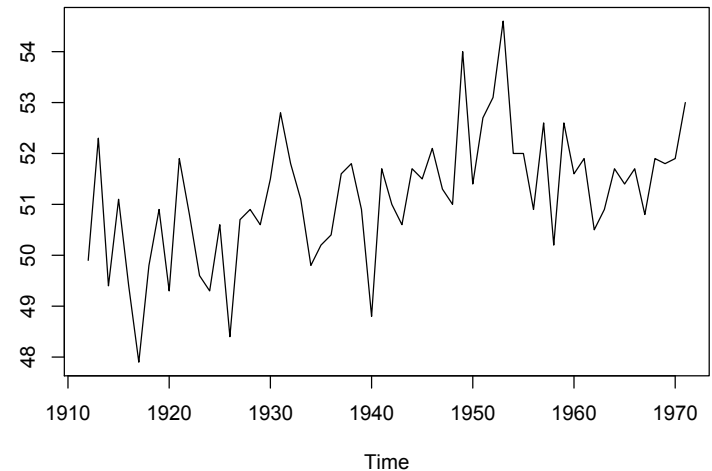
- Response variable (y)
- Independent variables (x)

Time series:

- Single process (y)

Idea:

- Exploit correlation in order to understand and model data
- Potentially forecast likelihood of future events



Time series analysis

When analysing time series, we are interested in how two values in the series – separated by k time-steps – affect each other.

k^{th} autocovariance:

$$\gamma_k = E(y_t - \mu)(y_{t-k} - \mu) \quad \text{K: Lag}$$

Average covariance between pairs of values that are k time steps apart in the series.

Since these are dependent on the scale of the process, these need to be standardised:

k^{th} autocorrelation:

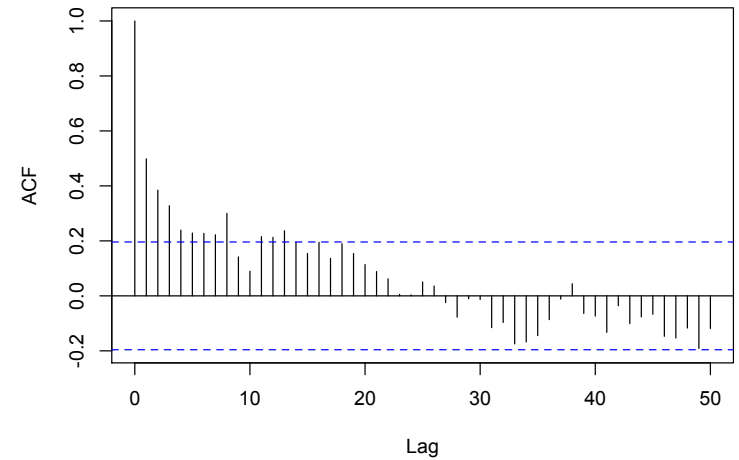
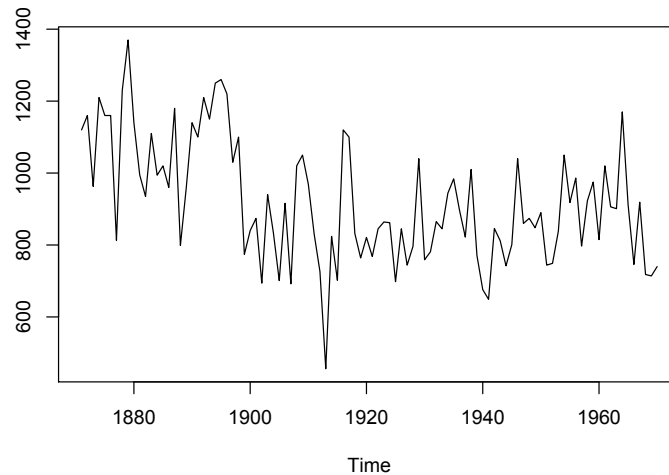
$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

The autocorrelation function is useful for characterising time series.

Time series analysis

Autocorrelation function:

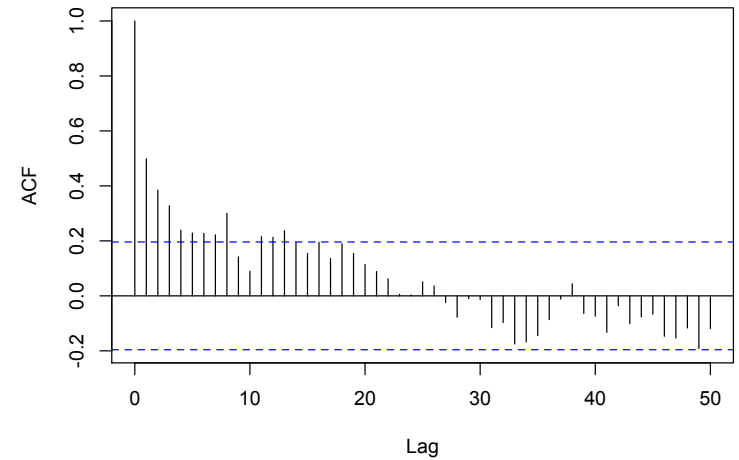
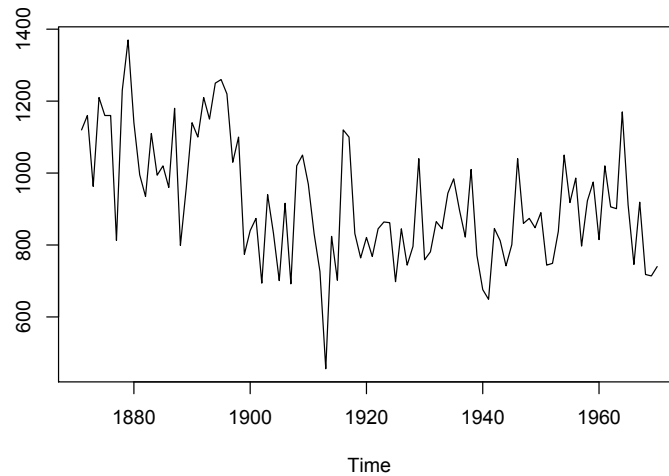
Nile annual flow:



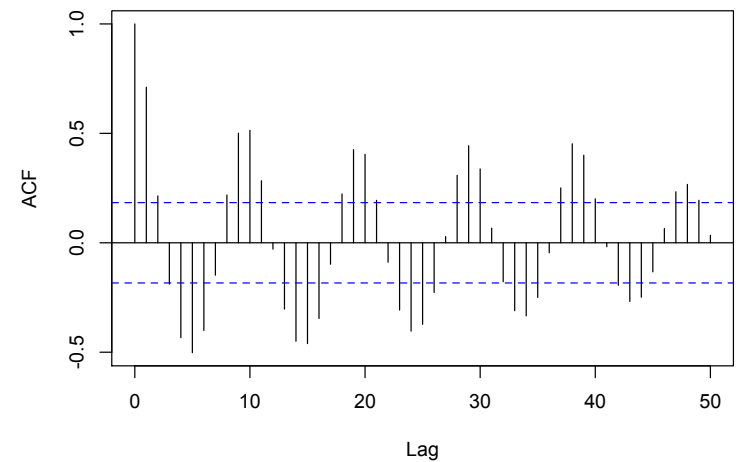
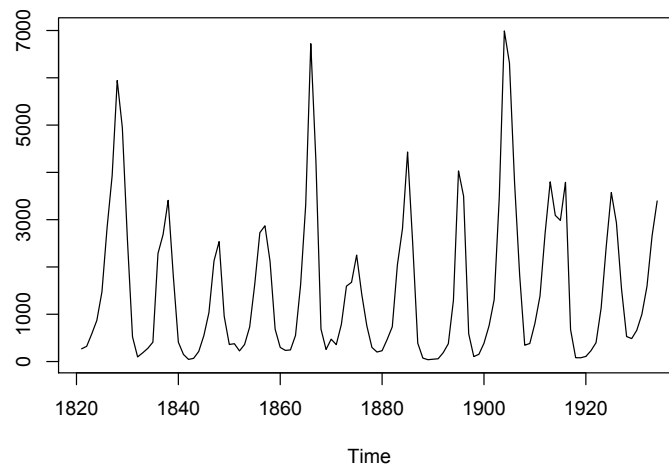
Time series analysis

Autocorrelation function:

Nile annual flow:



Lynx trappings:



Time series analysis

Autoregressive (AR) time series models:

AR(1):
$$y_t = c + \varphi_1 y_{t-1} + \varepsilon_t$$

Time series analysis

Autoregressive (AR) time series models:

AR(1): $y_t = c + \varphi_1 y_{t-1} + \varepsilon_t$

AR(2): $y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$

Time series analysis

Autoregressive (AR) time series models:

$$\text{AR}(1): \quad y_t = c + \varphi_1 y_{t-1} + \varepsilon_t$$

$$\text{AR}(2): \quad y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$$

$$\text{AR}(p): \quad y_t = c + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$$

Similarities to multiple regression model, except for the dependencies
Parameters estimated using least squares or maximum likelihood

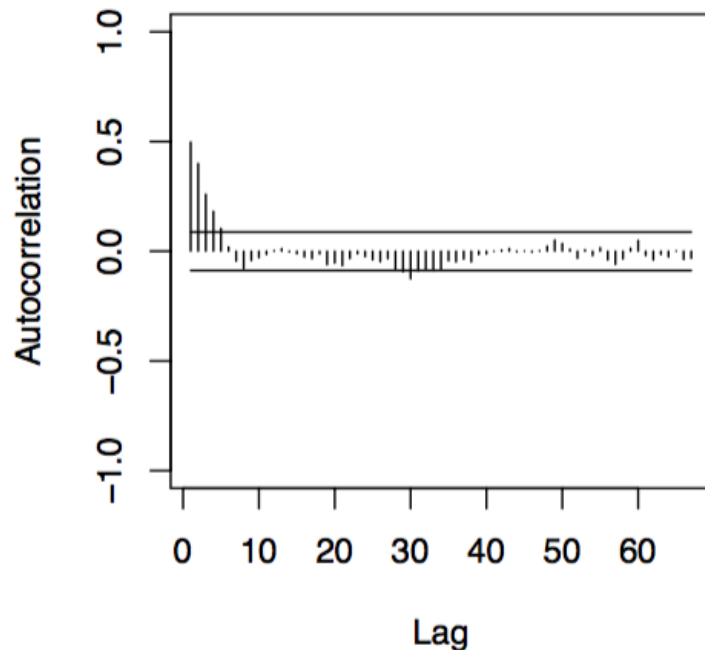
Assumptions:

- Independent Gaussian errors
- Covariance stationary process (trend doesn't change over time)

Time series analysis

Autoregressive (AR) time series models:

AR(2) with $c=0$, $\phi_1=0.4$ and $\phi_2=0.2$

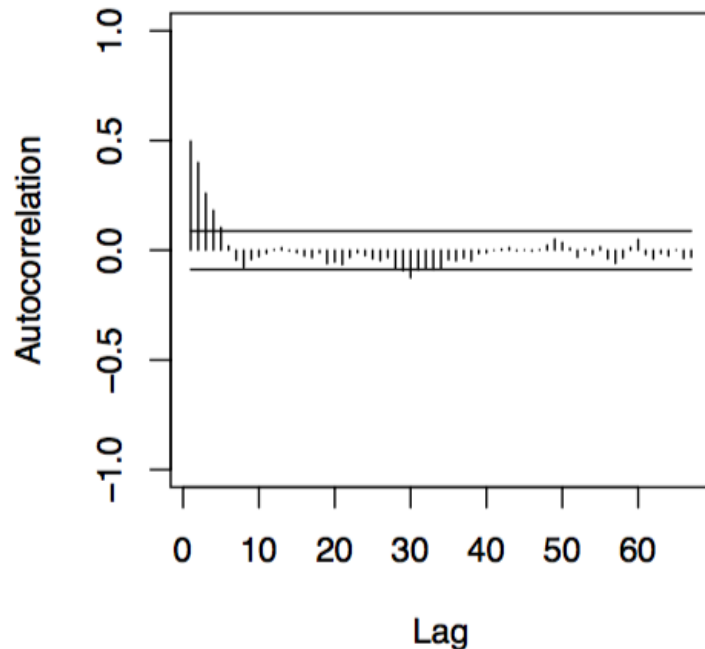


How to interpret ACF?

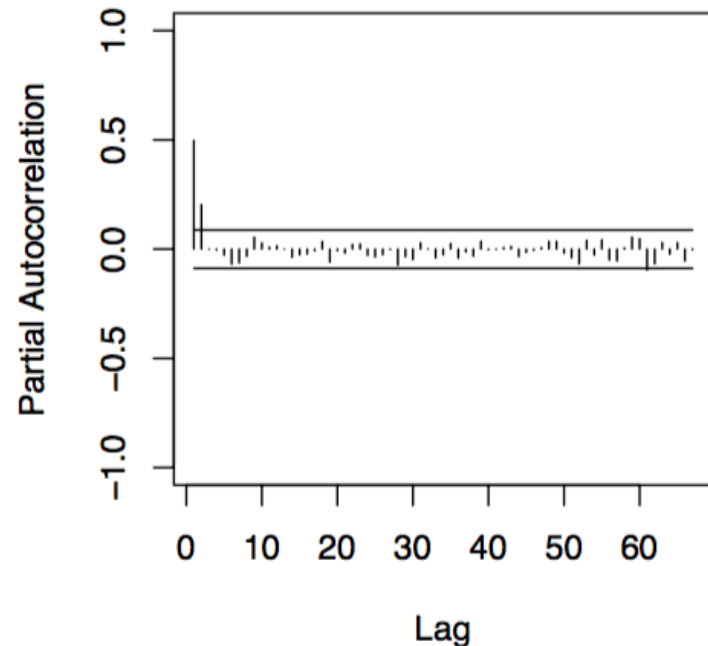
- Positive parameters: ACF should decay, not oscillate.
- Should decay gradually until within the confidence interval, then stay there.
- Can't infer order...

Time series analysis

Autoregressive (AR) time series models:



AR(2) with $c=0$, $\varphi_1=0.4$ and $\varphi_2=0.2$



Partial autocorrelation function: $\alpha(p) = \varphi_p$ from a $AR(p)$ model

Parsimonious modelling:

- First try $AR(1)$, then $AR(2)$, etc. until $H_0: \alpha(p) = 0$ is not rejected.
- Failure to reject leads us to conclude that $AR(p)$ is more appropriate than $AR(p-1)$.

Time series analysis

Moving Average (MA) time series models:

MA(1):
$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Time series analysis

Moving Average (MA) time series models:

MA(1): $y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$

MA(2): $y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$

Time series analysis

Moving Average (MA) time series models:

$$\text{MA}(1): \quad y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$\text{MA}(2): \quad y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

$$\text{MA}(q): \quad y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

Unlike multiple regression model there are multiple error terms
However, the current state is only ever dependent on a known no. of previous states

Since the current state only depends on the previous q states,
the ACF should suddenly drop to zero, unlike AR(p) processes

Time series analysis

More general models:

Auto Regressive, Moving Average:

$$\text{ARMA}(p,q): \quad y_t = c + \underbrace{\varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p}}_{\text{AR}(p)} + \varepsilon_t + \underbrace{\theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}}_{\text{MA}(q)}$$

Time series analysis

More general models:

Auto Regressive, Moving Average:

$$\text{ARMA}(p,q): \quad y_t = c + \underbrace{\varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p}}_{\text{AR}(p)} + \varepsilon_t + \underbrace{\theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}}_{\text{MA}(q)}$$

Auto Regressive, Integrated, Moving Average:

$$\text{ARIMA}(p,1,q): \quad x_t = y_t - y_{t-1} \quad \text{then model as ARMA}(p,q)$$

Time series analysis

More general models:

Auto Regressive, Moving Average:

$$\text{ARMA}(p,q): \quad y_t = c + \underbrace{\varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p}}_{\text{AR}(p)} + \varepsilon_t + \underbrace{\theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}}_{\text{MA}(q)}$$

Auto Regressive, Integrated, Moving Average:

$$\text{ARIMA}(p,1,q): \quad x_t = y_t - y_{t-1} \quad \text{then model as ARMA}(p,q)$$

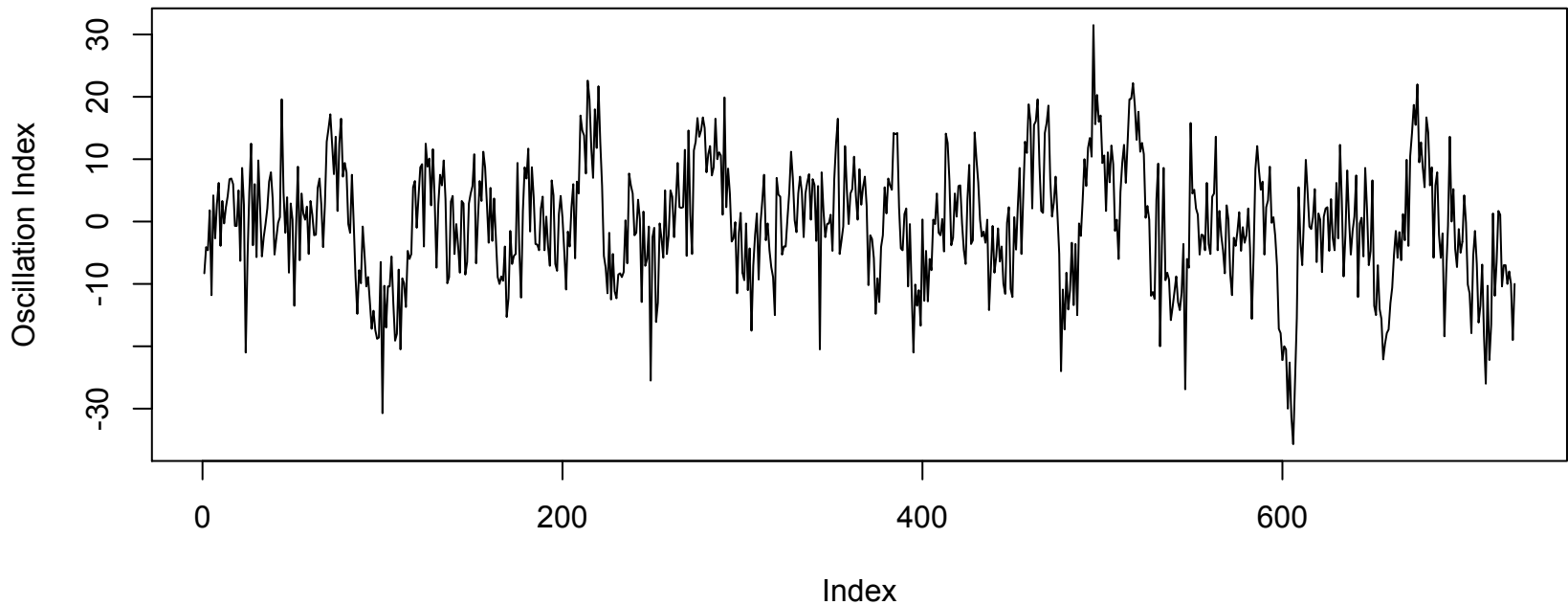
$$\text{ARIMA}(p,d,q): \quad x_t = \nabla^d y_t \quad \text{take } d^{\text{th}} \text{ order differences}$$

Considering ARIMA models can be a useful “transformation” if assumptions are violated

Time series analysis

Example: Monthly Southern Oscillation Index

Monthly difference in sea-surface air pressure between Darwin and Tahiti

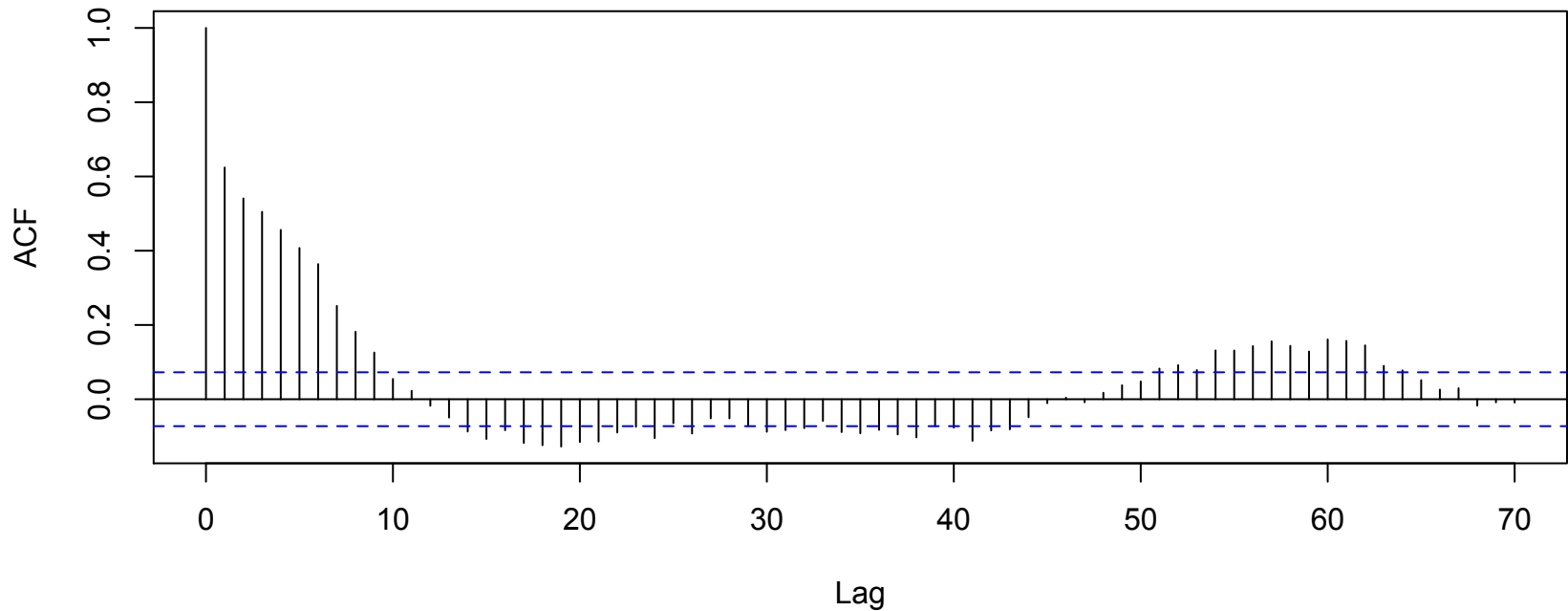


Used for predicting rainfall in parts of Australia

Time series analysis

Example: Monthly Southern Oscillation Index

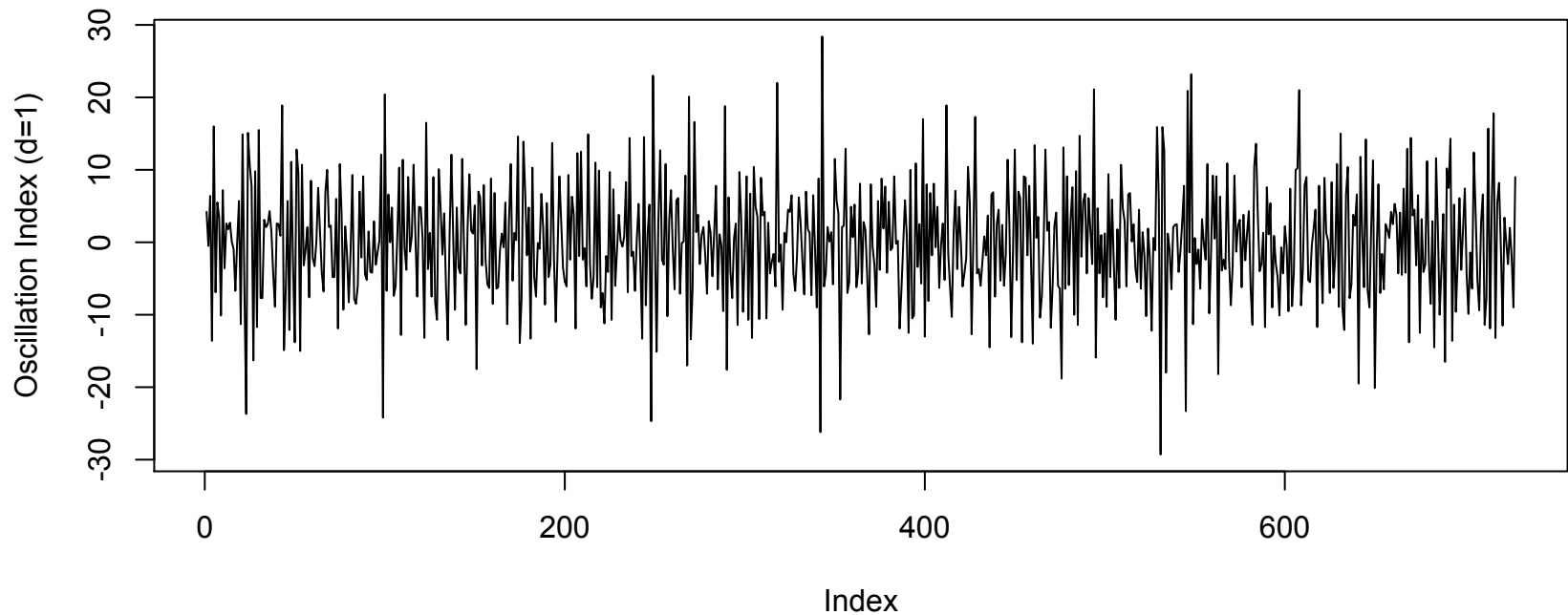
Monthly difference in sea-surface air pressure between Darwin and Tahiti



Time series analysis

Example: Monthly Southern Oscillation Index

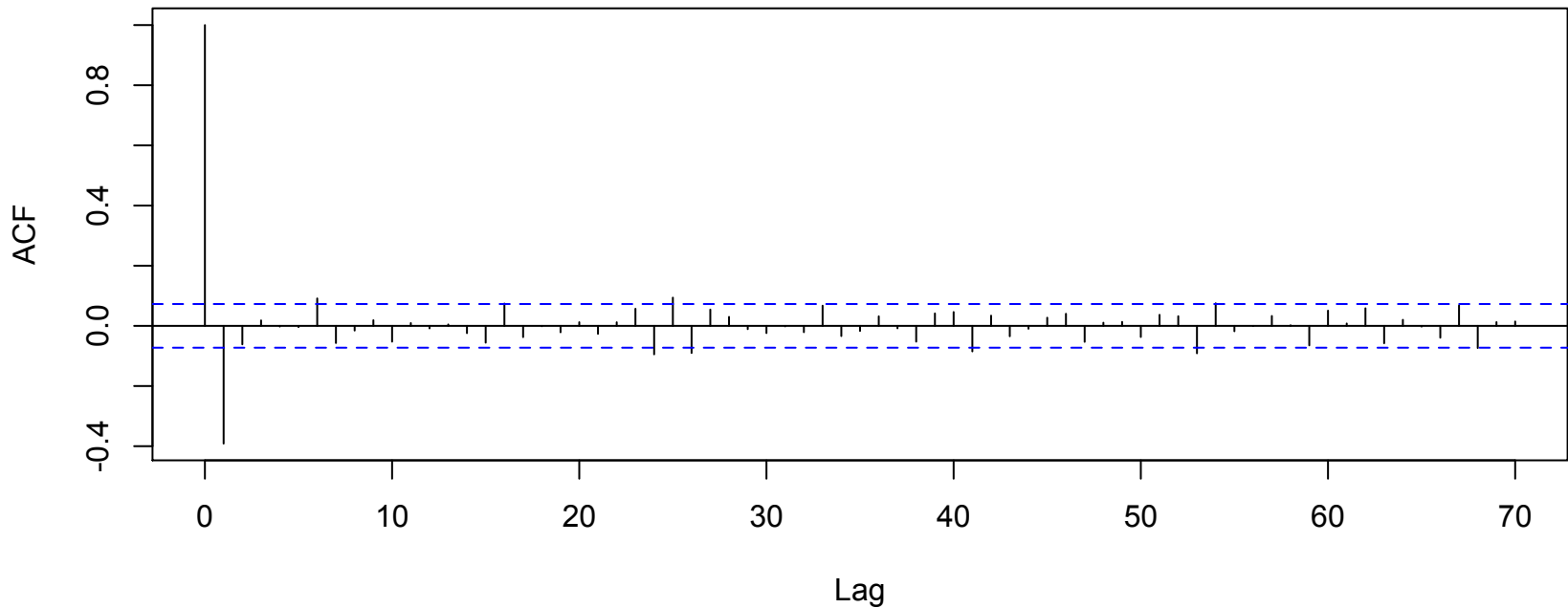
Monthly difference in sea-surface air pressure between Darwin and Tahiti



Time series analysis

Example: Monthly Southern Oscillation Index

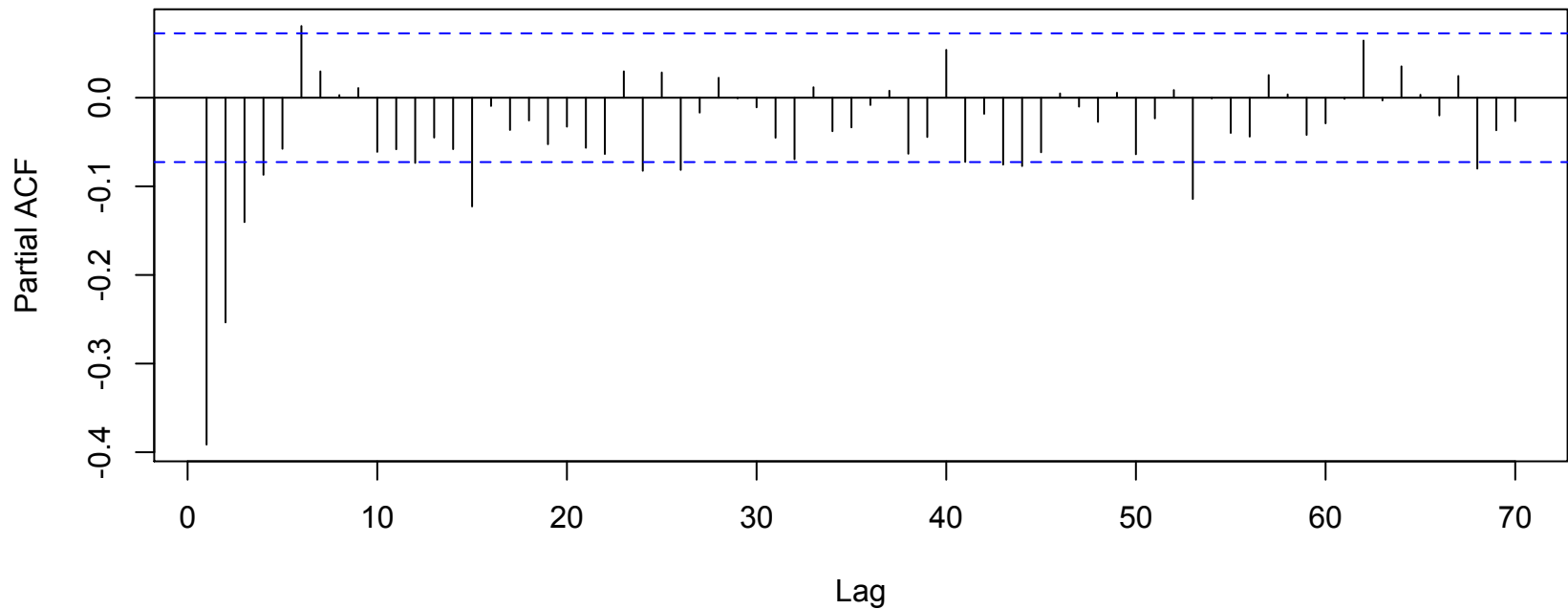
Monthly difference in sea-surface air pressure between Darwin and Tahiti



Time series analysis

Example: Monthly Southern Oscillation Index

Monthly difference in sea-surface air pressure between Darwin and Tahiti



Time series analysis

Example: Monthly Southern Oscillation Index

Monthly difference in sea-surface air pressure between Darwin and Tahiti

Try ARIMA(0,1,1) model:

```
arima(x = x$Index, order = c(0, 1, 1))
```

Coefficients:

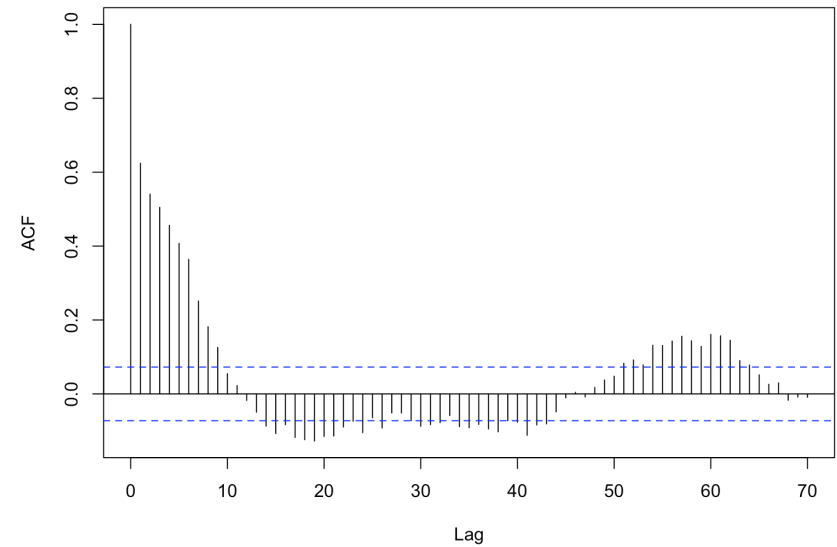
```
      ma1  
      -0.5579  
s.e.    0.0308
```

sigma^2 estimated as 52.94: log likelihood = -2477.98, aic = 4959.96

Time series analysis

R functions:

`acf(x,lag.max=70)`



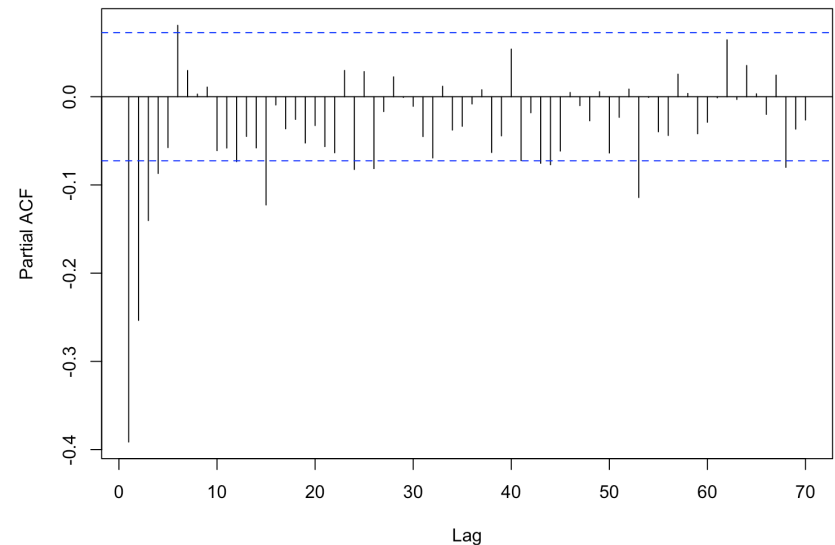
Time series analysis

R functions:

`acf(x,lag.max=70)`

`diff(x)`

`pacf(diff(x),lag.max=70)`



Time series analysis

R functions:

acf(x,lag.max=70)

diff(x)

pacf(diff(x),lag.max=70)

arima(x,order=c(0,1,1))

##

Call:

arima(x = x[Index], order = c(0, 1, 1))

##

Coefficients:

ma1

-0.5579

s.e. 0.0308

##

sigma^2 estimated as 52.94: log likelihood
= -2477.98, aic = 4959.96