Time Series Toolbox

Birgit Erni

SEEC, Department of Statistical Sciences, University of Cape Town

25 April 2019

SEEC - Statistics in Ecology, Environment and Conservation

Outline: Time Series in Practice

- When and why do we need time series models?
- Basic models and definitions: white noise, AR1, MA, random walk, stationarity.
- 3 approaches to time series modelling: ARIMA, Regression, Structural time series / state-space models

Aim today: understand basic difficulties with time series, construct a few simple but useful models

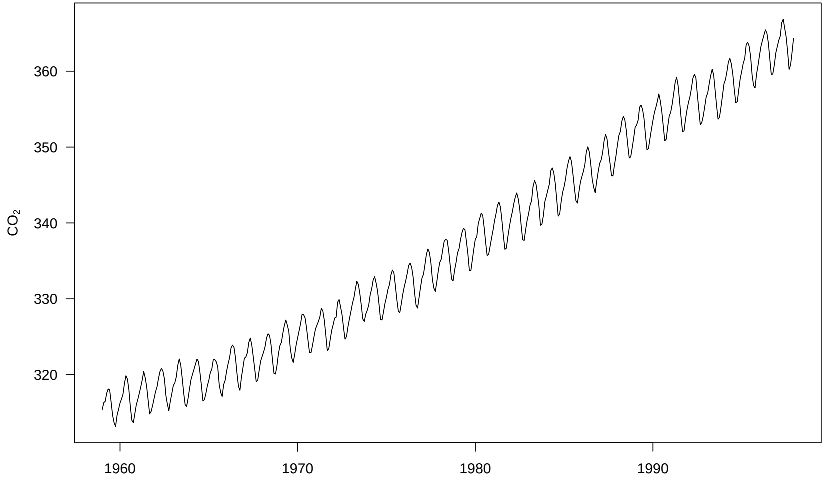
References

- Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: Principles and Practice, 2nd edition, OTexts: Melbourne, Australia. https://otexts.com/fpp2
- Applied Time Series Analysis for Fisheries and Environmental Sciences. E. E. Holmes, M. D. Scheuerell, and E. J. Ward. (2019). https://nwfsc-timeseries.github.io/atsa-labs/index.html

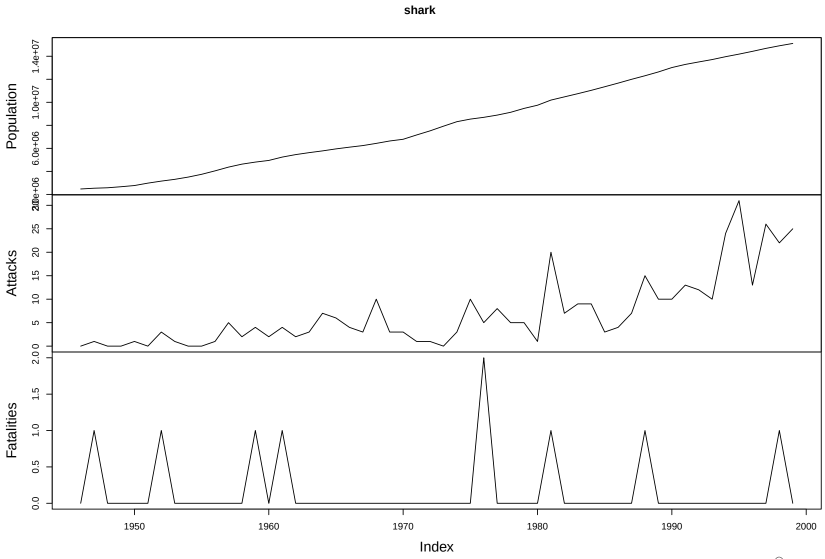
More technical:

- Shumway, R.H. and Stoffer, D.S., 2017. Time series analysis and its applications: with R examples. Springer. https://www.stat.pitt.edu/stoffer/tsa4/

Motivating Example: Mauna Loa Atmospheric CO2 Concentration



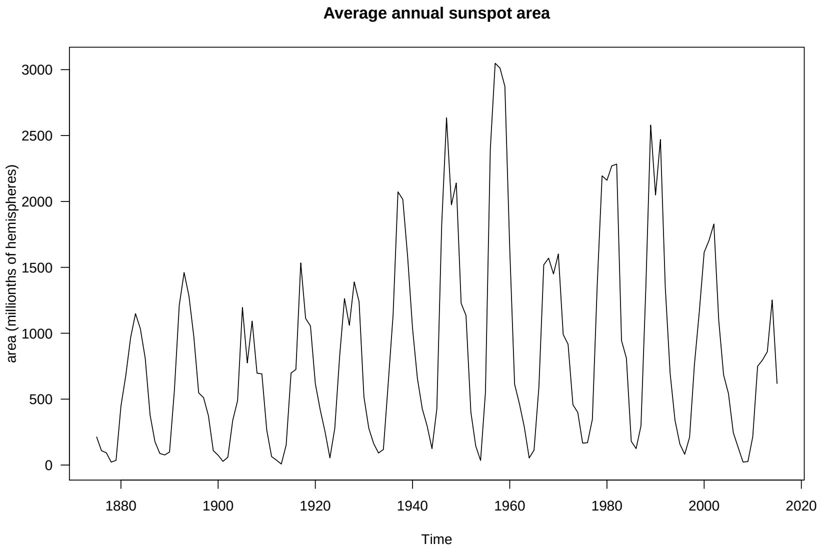
Shark Attacks in Florida



Financial

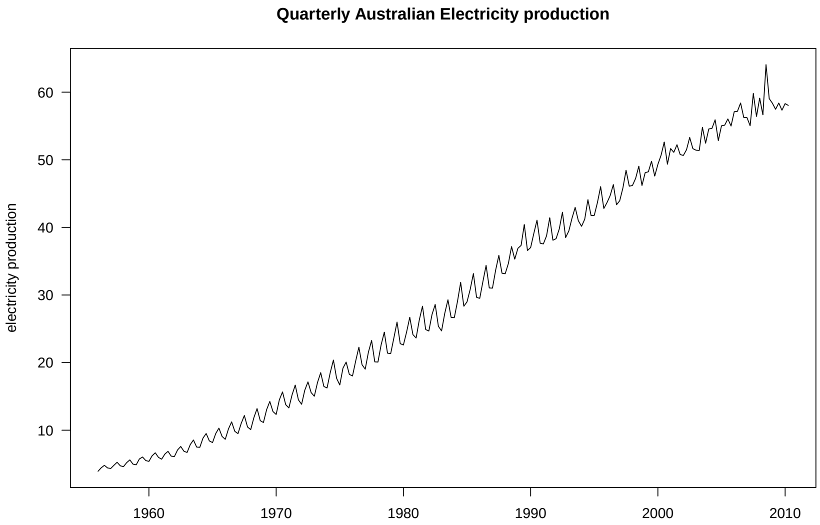
SEEC – Statistics in Ecology, Environment and Conservation

Sunspot Area



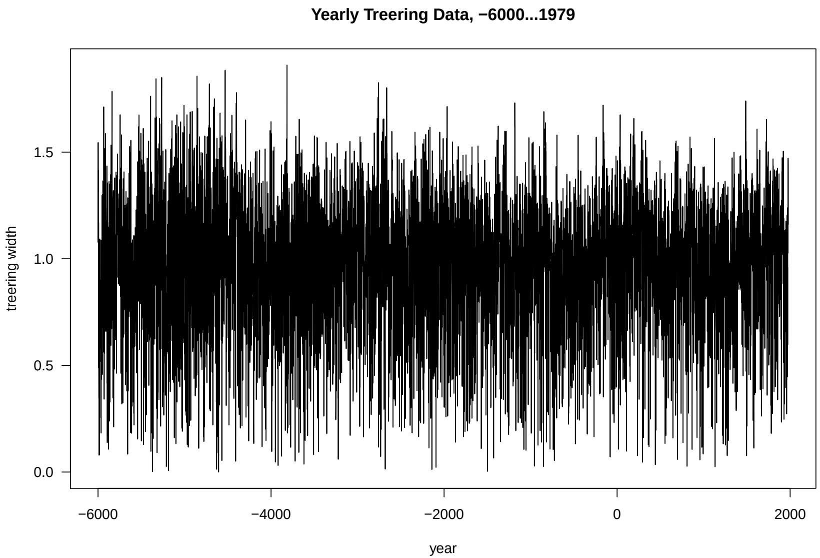
SEEC - Statistics in Ecology, Environment and Conservation

Electricity Production



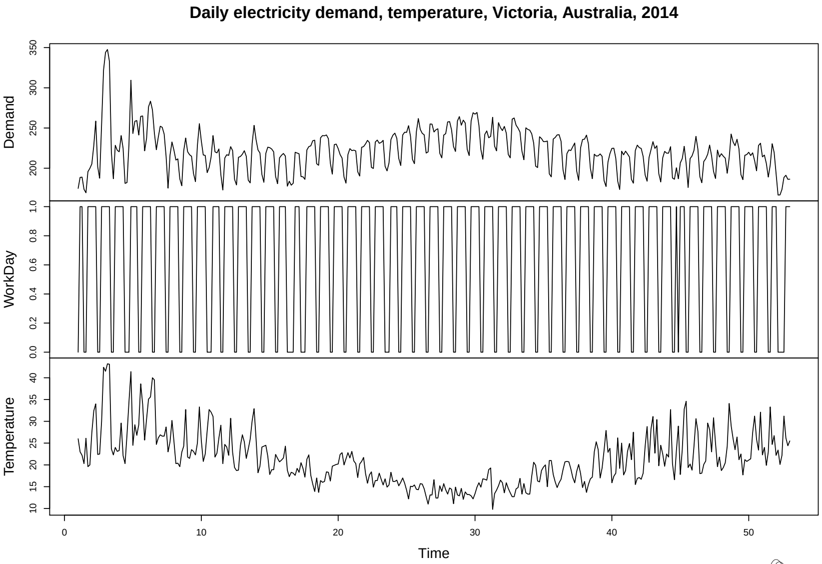
SEEC – Statistics in Ecology, Environment and Conservation

Treerings



SEEC - Statistics in Ecology, Environment and Conservation

Electricity Demand



SEEC - Statistics in Ecology, Environment and Conservation

Goals of Time Series Analysis

- prediction / forecast
- impact of single event
- study causal patterns

Time Series in R

```
## CO2 data

##          Jan    Feb    Mar    Apr    May    Jun    Jul
## 1959 315.42 316.31 316.50 317.56 318.13 318.00 316.39 31
## 1960 316.27 316.81 317.42 318.87 319.87 319.43 318.01 31
##          Nov    Dec
## 1959 314.66 315.43
## 1960 314.84 316.03

##
##  treering data

## Time Series:
## Start = -6000
## End = -5981
## Frequency = 1
##  [1] 1.345 1.077 1.545 1.319 1.413 1.069 0.489 1.171 0.8
## [12] 0.846 0.837 0.079 0.829 0.919 0.776 0.081 0.876 0.2
```

Time Series in R

Create time series of quarterly data:

```r
ts(rnorm(40), frequency = 4, start = c(2019, 2))
```

```
##             Qtr1        Qtr2        Qtr3        Qtr4
## 2019              0.33524522 -0.71583544 -0.13751188
## 2020  1.03208627  1.26466849  0.88699606  0.26858419
## 2021  1.10018262 -0.58865896  0.14023571 -0.27432658
## 2022 -0.22098499 -0.54758319  2.49355165  1.61537688
## 2023 -0.52070523  1.11406196 -0.08867364  0.21284585
## 2024 -0.45450847 -1.49247939 -0.08907184 -1.73229935
## 2025 -0.60372515  1.12784551  1.11462281 -0.45584863
## 2026 -0.97076347  0.35040089 -1.95474722 -0.63967372
## 2027 -0.40198452  0.10506850  0.04510855  0.84300729
## 2028  0.70200183  0.40813321  0.56816479  0.88684596
## 2029 -0.44914980
```

When and why do we need time series models?

When there is *auto-correlation in the residuals* (after modelling trends, seasonality, effects of explanatory variables).

If we ignore autocorrelation:

- standard error estimates are wrong
- predictions and prediction intervals are wrong

Terminology / Definitions

time series: $y_1, ..., y_t$

autocorrelation: correlation with previous values,

Basic time series processes

AR process

$$\text{AR(p):} \qquad x_t = \phi_1 x_{t-1} \phi_2 x_{t-2} + ... + e_t$$

$$\text{AR1:} \qquad x_t = \phi x_{t-1} + e_t$$

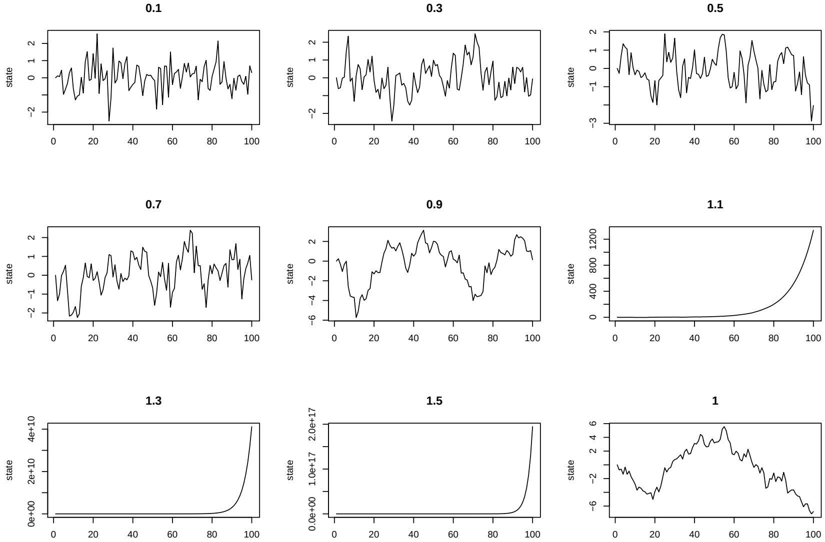
Why would anything behave like this?

I see x_{t-1} as a measure of everything that was not measured explicitly at previous time step.

Random walk

$$x_t = x_{t-1} + e_t$$

AR1 processes, different ϕ



Stationarity

- mean, variance, correlations stay constant over time

Why is stationarity important?

There is a single observation per time point. If mean and variance are different for every point, we can't estimate mean and variance, correlation or model parameters.

AR1 processes are stationary if $|\phi| < 1$.

non-stationary means

mean changes, variance changes, seasonality present, correlation changes

MA(q) process

$$y_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t$$

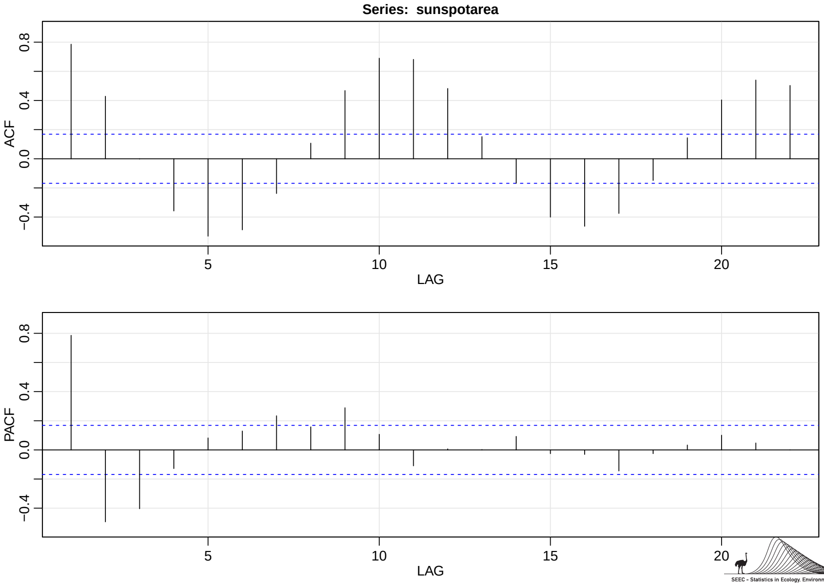
sum of previous shocks / events

White noise

identically, independently distributed, mean 0, no autocorrelation

ACF and PACF

autocorrelation function, partial autocorrelation function



Three approaches to time series modelling

1. ARIMA, very briefly

arima(p, d, q)

AR(p), d = difference order, MA(q)

If y_t is not stationary then $y_t - y_{t-1}$ sometimes is (first order differences).

Three approaches to time series modelling

2. Regression

Ignore or model auto-correlation in errors like this:

$$y_t = \beta_0 + \beta_1 x_t + \nu_t$$

$$\nu_t = \phi \nu_{t-1} + e_t$$

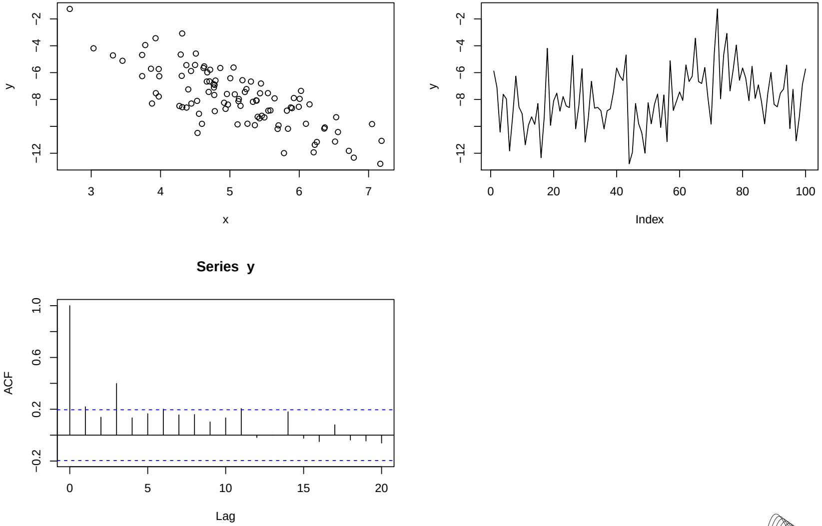
Not this:

$$y_t = \beta_0 + \beta_1 x_t + \phi y_{t-1} + e_t$$

problem: β_1 is not change in response per unit change in x

https://robjhyndman.com/hyndsight/arimax/

Simulate some data with autocorrelation, fit OLS regression model



ARMA errors

arima with xreg models autocorrelation in errors

```
a1 <- arima(y, order = c(0, 0, 0), xreg = x)
a2 <- arima(y, order = c(1, 0, 0), xreg = x)
```

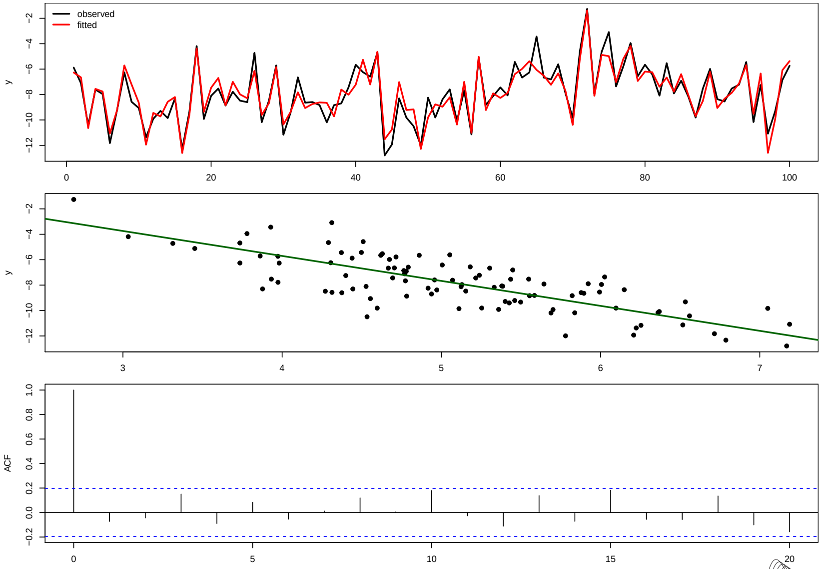
```
##
## Call:
## arima(x = y, order = c(1, 0, 0), xreg = x)
##
## Coefficients:
##          ar1  intercept        x
##       0.8238     2.1524  -1.9644
## s.e.  0.0544     0.5223   0.0640
##
## sigma^2 estimated as 0.5757:  log likelihood = -114.85,
```

true values: $\beta_0 = 3, \beta_1 = -2, \phi = 0.9, \sigma^2 = 0.64$

Regression without ARMA errors

```
##
## Call:
## arima(x = y, order = c(0, 0, 0), xreg = x)
##
## Coefficients:
##       intercept        x
##          1.4424  -1.8305
## s.e.     0.7726   0.1494
##
## sigma^2 estimated as 1.842:  log likelihood = -172.43,
```

ARMA errors


observed
fitted
y
−2
−4
−6
−8
−10
−12
0
20
40
60
80
100
3
4
5
6
7
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20

Terminology

fitted values:

$\hat{y}_t | y_1, y_2, \ldots, y_{t-1}$

usually one-step-ahead forecast, using model

residuals:

$y_t - \hat{y}_t$

forecasts:

$y_{t+h} | y_1, y_2, \ldots, y_t$

Point forecasts need **PREDICTION INTERVALS**

Alternative for regression with ARMA errors: GLS

GLS: generalized least squares

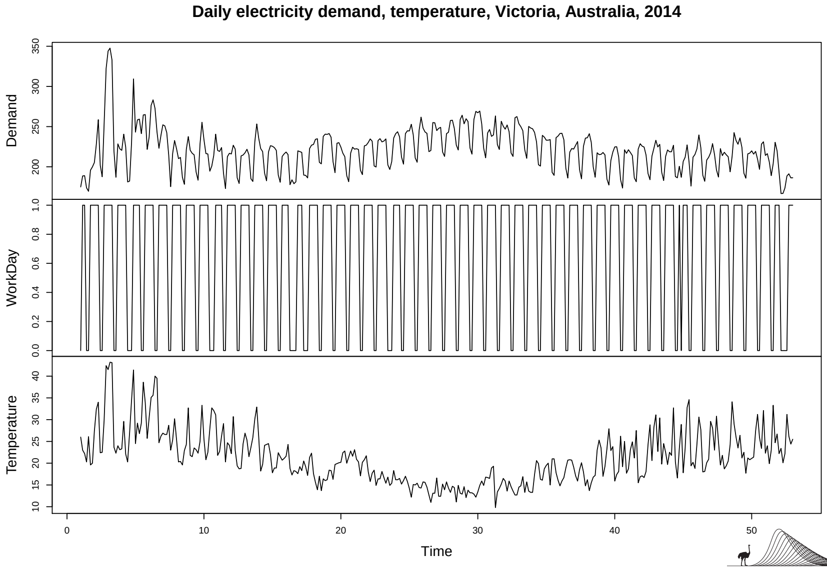
linear regression with correlated errors

```
library(nlme)
g2 <- gls(y ~ x, correlation = corAR1(form = ~ 1))
```

More flexible alternative for regression with ARMA errors: GAMs with correlated errors

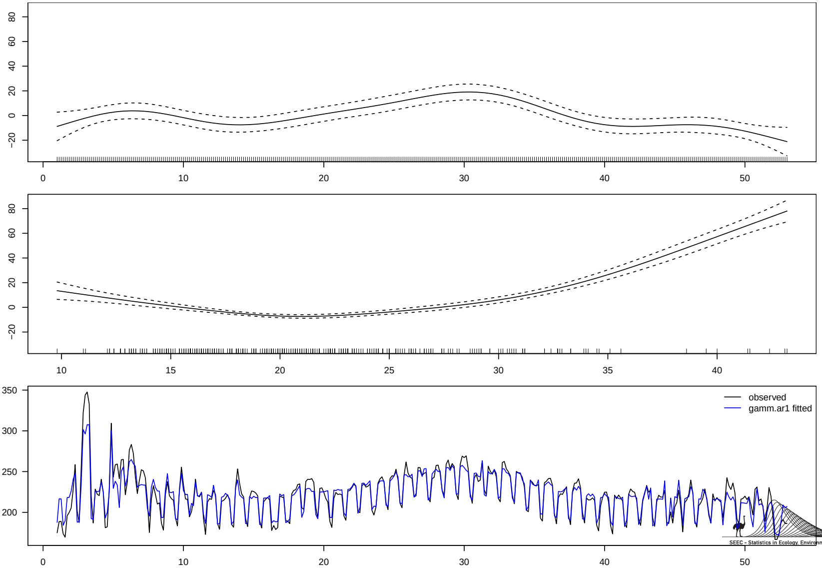
`gamm` (mgcv) will fit a GAM, and allow different correlation structures for the errors.

Example: Daily electricity demand (weekday, maximum temperature)



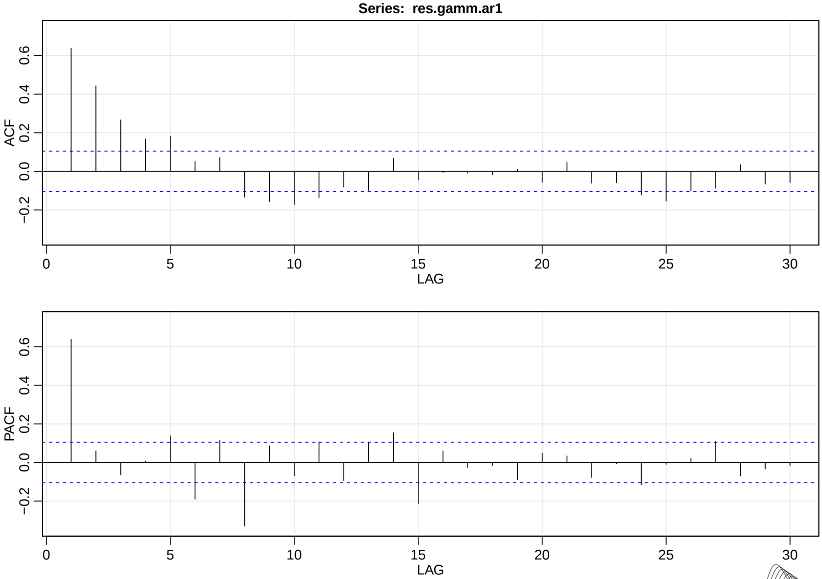
GAMs with correlated errors

```
gamm.ar1 <- gamm(demand ~ s(week) + s(temp) + weekd,
                 correlation = corAR1(), method = "REML")
```



SEEC - Statistics in Ecology, Environment and Conservation

GAMM residuals



SEEC - Statistics in Ecology, Environment and Conservation

3rd approach: Structural Time Series Models

Trend, season, error modelled explicitly. Similar to regression. Easier to understand.

$$y_t = T_t + S_t + X_t + R_t$$

T_t = trend

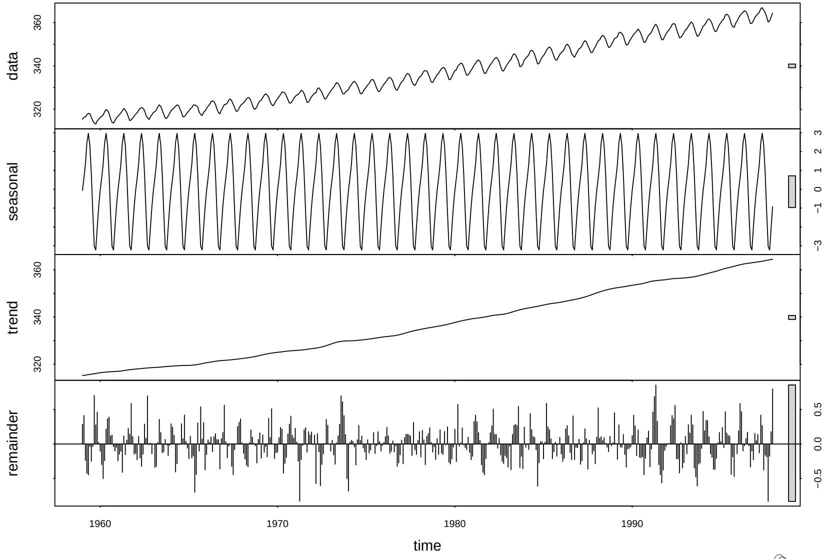
S_t = seasonality

R_t = remainder

X_t = regression terms

International Encyclopedia of Statistical Science:
https://link.springer.com/referenceworkentry/10.1007%2F978-3-642-04898-2_577

decompose



SEEC - Statistics in Ecology, Environment and Conservation

Structural Time Series Models

linear Gaussian state-space models for univariate time series based on a decomposition of the series into components.

Some basic models:

1. Local Level:

$$\mu_t = \mu_{t-1} + \xi_t, \qquad \xi_t \sim N(0, \sigma^2_\xi)$$

with observations:

$$y_t = \mu_t + e_t, \qquad e_t \sim N(0, \sigma^2_e)$$

trend / mean is a random walk, errors independent

Structural Time Series Models

2. Local Linear Trend:

$$\mu_t = \mu_{t-1} + \beta_t + \xi_t, \qquad \xi_t \sim N(0, \sigma_\xi^2)$$

$$\beta_t = \beta_{t-1} + w_t \qquad w_t \sim N(0, \sigma_w^2)$$

dynamic / time-varying trend

observations as before

Structural Time Series Models

3. Basic Structural Model

$$y_t = \mu_t + \tau_t + e_t, \qquad e_t \sim N(0, \sigma_e^2)$$

τ_t is the seasonal component with dynamics

$$\tau_t = -\sum_{s=1}^{S-1} \tau_{t-s} + w_t, \qquad w_t \sim N(0, \sigma_w^2)$$

or

$$\tau_t = \alpha_1 cos(\frac{2\pi t}{\omega}) + \alpha_2 sin(\frac{2\pi t}{\omega})$$

State-space model

model the components, can change over time

$$y_t = \underbrace{\mu_t}_{\text{trend}} + \underbrace{\tau_t}_{\text{seasonal}} + \underbrace{X_t\beta_t}_{\text{regression}} + e_t$$

$$\mu_t = \mu_{t-1} + \underbrace{\delta_t}_{\text{slope}} + u_t$$

$$\delta_t = \delta_{t-1} + w_t$$

$$e_t, u_t, w_t$$

independent, (Gaussian) white noise

JAGS

```
## JAGS model for simple state-space model
## local level

model {

  ## Prior for error of observation process
  sigma.obs ~ dunif(0, 10)
  tau.obs <- pow(sigma.obs, -2)

  ## Prior for error of state process
  sigma.proc ~ dunif(0, 10)
  tau.proc <- pow(sigma.proc, -2)

  ## State process
  mu[1] ~ dnorm(y0, 0.001)

  for (t in 2:T) {
    r[t] ~ dnorm(0, tau.proc)
    mu[t] <- mu[t-1] + r[t]
  }
```

JAGS

```
  ## Observation process / model / likelihood
  for (t in 1:T) {
    y[t] ~ dnorm(mu[t], tau.obs)
  }
}
```

https://maialesosky.files.wordpress.com/2016/02/bayesian-state-space-model-applications-for-time-series-analysis-1.pdf

JAGS: electricity demand, local level state-space model