MATH3802
Chapter 1: Introduction

Peter Thwaites

with thanks to Stuart Barber for previous years' notes and slides

January 2017

Admin details

Lecturer Dr Peter Thwaites.
Contact details Email `P.A.Thwaites@leeds.ac.uk`. Room 8.27.
Resources These are available on the VLE.
Lectures/classes Mon 9am, RSLT 22; Weds 10am RSLT 24.
Assessment 2 hour end of semester examination (80%) and an assessed practical (20%).
Practical Assessed practical in Teaching week 6, on the Monday pm (MATH3802), or Tuesday am (MATH5802).
Exercises Approximately 4 exercise sheets will be given out (roughly every other week).

Outline syllabus In this module we will study time series, i.e. random quantities which depend on time, including

trend and smoothing;

autocovariance and autocorrelation;

autoregressive (AR) and moving average (MA) models;

ARMA and ARIMA models;

estimation, assessment and forecasting.

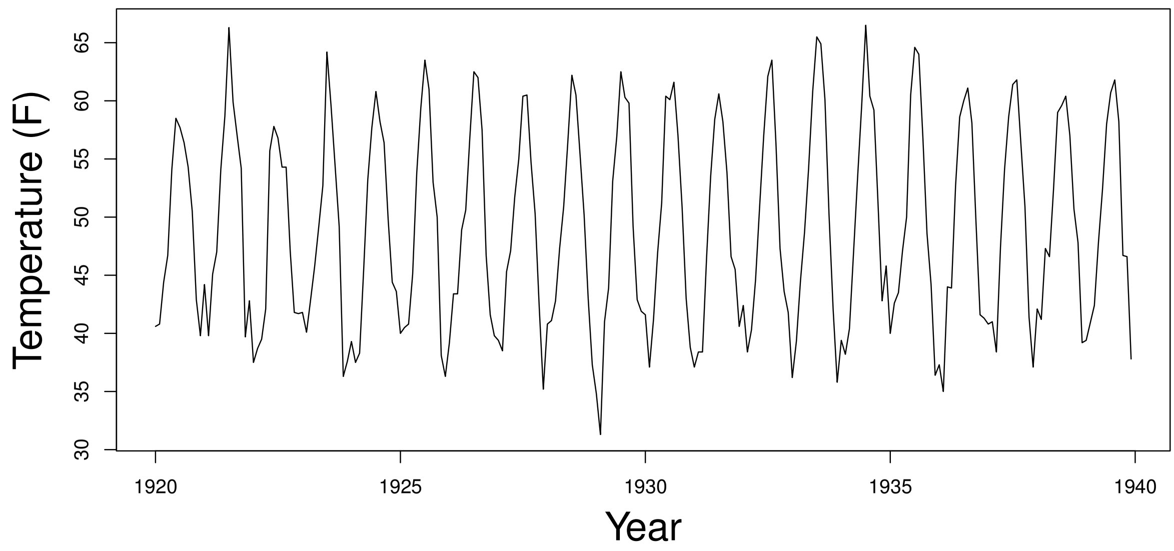
Books The presentation of the module will be self-contained, i.e. it will not be necessary to buy/borrow a book for the module. However, there are several good books if you want to do some background reading, including

C. Chatfield, *The Analysis of Time Series: an Introduction*.

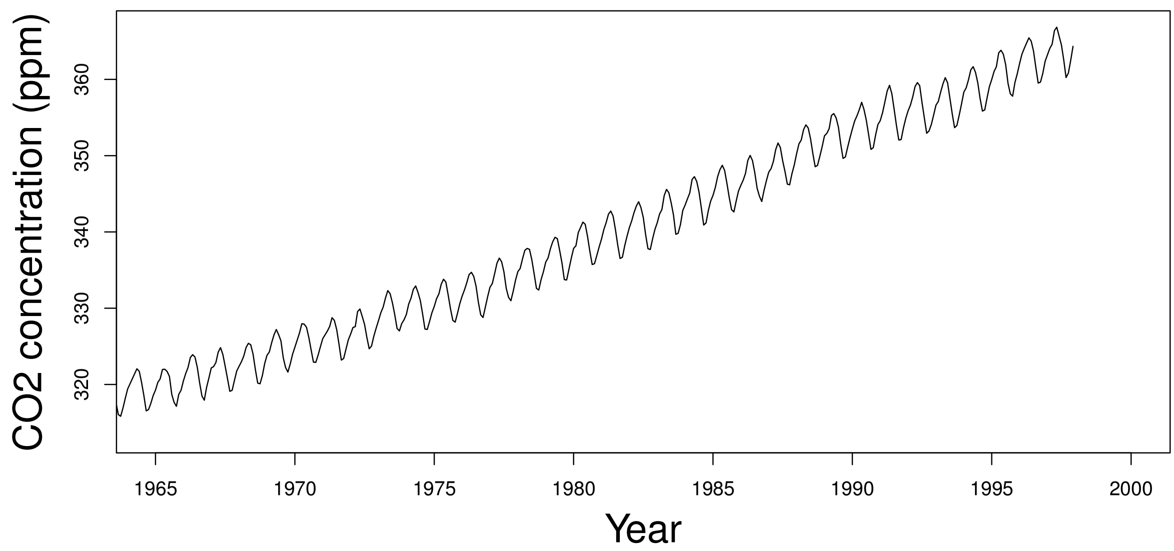
1. Introduction

- A *time series* $\{X_t\}$ consists of observations X_t made at times $t = t_1, \ldots, t_n$ (e.g. every day). These data are influenced by systematic effects and randomness. Typically observations at nearby times are "similar".
- We wish to describe the data (e.g. plot X_t against t), suggest a model (estimate parameters and test validity), and predict future behaviour of the time series (forecasting).
- The initial stages of our analysis will be to identify, model, and remove any *trend* or *seasonal effects* that might be present in the data.
- First, consider some traditional example time series.

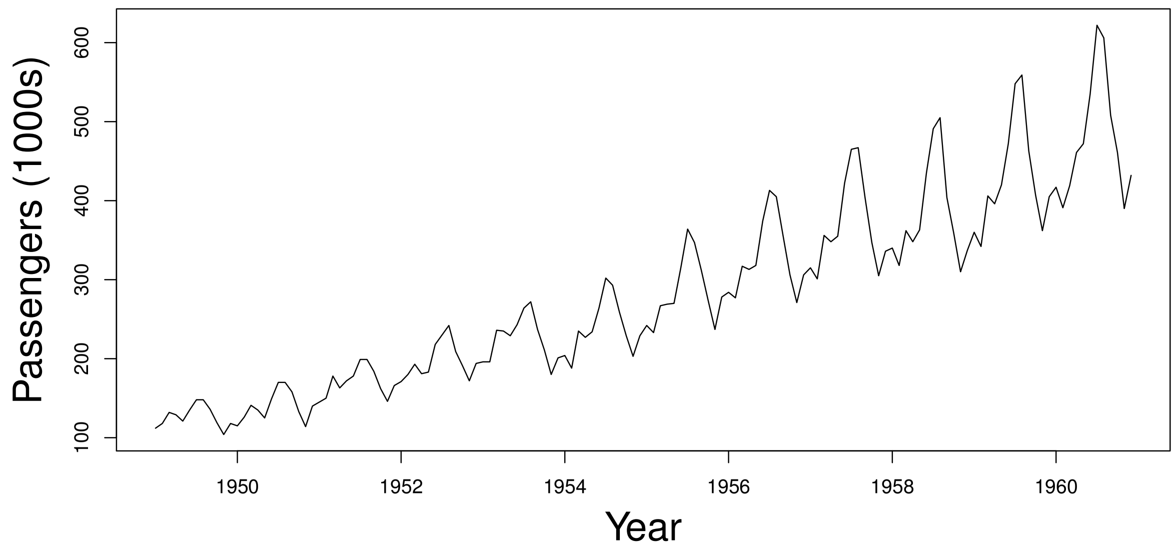
Nottingham Castle air temperature



Mauna Loa CO_2 concentration



Airline passengers



1.1 Trend

- The *trend* of a time series is a slow change in mean level.
- Consider the model

$$X_t = \mu(t) + \varepsilon_t, \tag{1}$$

  where
  - $\mu(t)$ is a (deterministic) trend as a function of t,
  - ε_t is a (random) fluctuation about the trend at time t, $E(\varepsilon_t) = 0$.
- For example, consider
  - $\mu(t) = \alpha + \beta t$,
  - $\mu(t) = \alpha + \beta \exp(\gamma t)$,
  - $\mu(t) = \alpha + \beta \log t \ldots$,

  $\vdots$

  where $\alpha, \beta, \ldots \in \mathbb{R}$.

Fitting a trend

- Suppose we have observed values $X_1, \ldots, X_n$ at times $t_1, \ldots, t_n$ and we believe a linear trend is appropriate.
- We can estimate α, β by linear regression, i.e. by minimising

$$r(\alpha, \beta) = \sum_{i=1}^{n} (X_i - \alpha - \beta t_i)^2.$$

This yields

$$\widehat{\alpha} = \overline{X} - \widehat{\beta}\,\overline{t} \qquad \text{and} \qquad \widehat{\beta} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(t_i - \overline{t})}{\sum_{i=1}^{n} (t_i - \overline{t})^2}.$$

- The estimated trend is thus

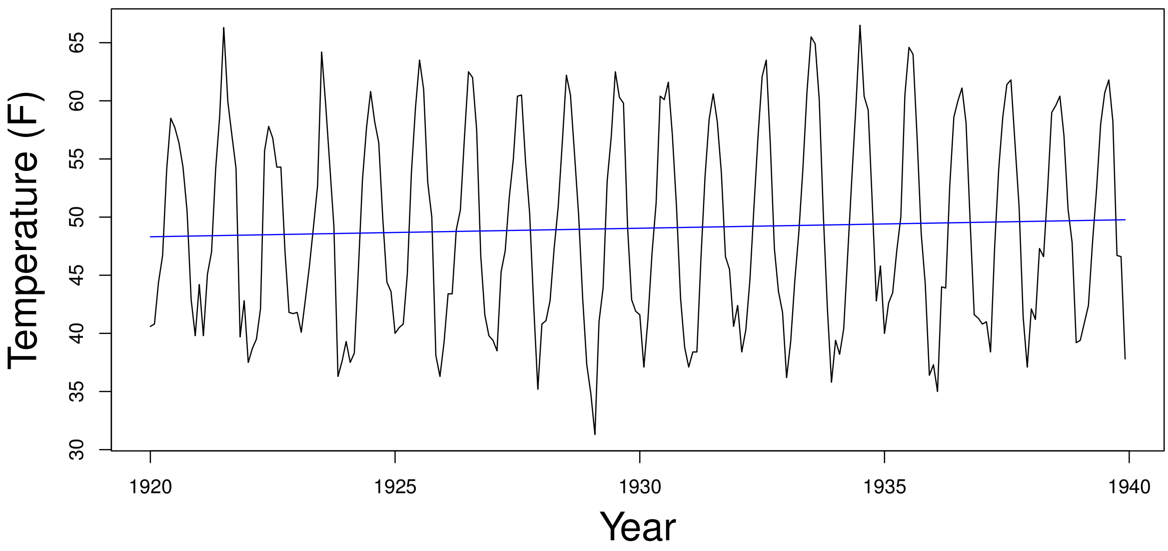
$$\widehat{\mu}(t) = \widehat{\alpha} + \widehat{\beta} t.$$

Remarks

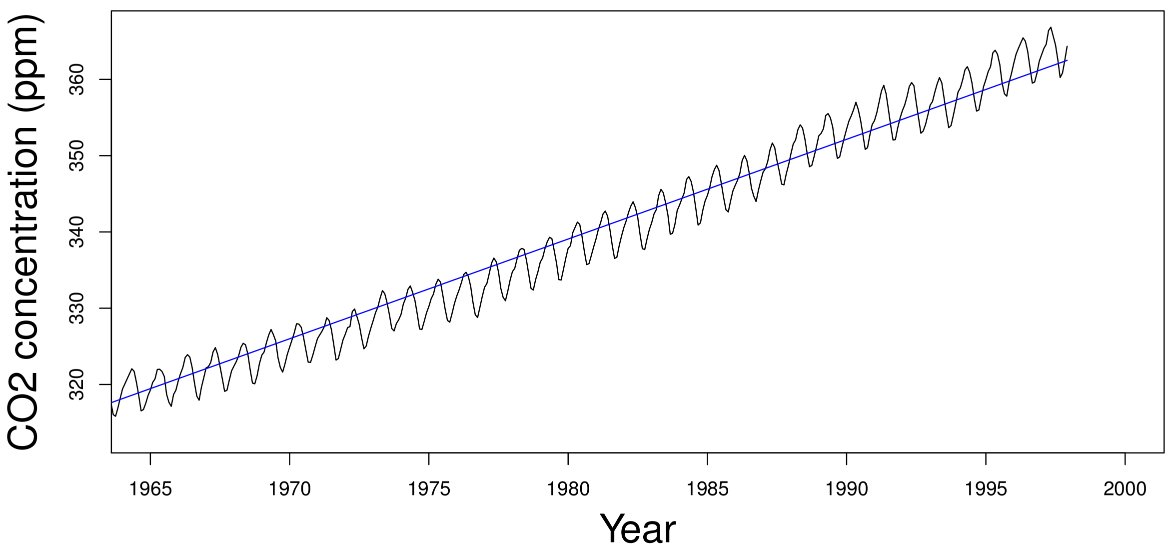
- If a trend is present in the data, the usual summary statistics (mean, variance, etc) will not be very useful.
- After a trend is fitted, the *residuals* $Y_{t_i} = X_{t_i} - \widehat{\mu}(t_i)$ can be analysed further, e.g. by estimating $\text{var}(Y_{t_i})$.
- We can predict future behaviour of X by considering $\widehat{\mu}(t)$ for $t > t_n$. Once we have analysed the residuals, we can improve the forecast by considering $\widehat{\mu}(t) + \widehat{Y}_t$ for $t > t_n$.
- Note that forecast error depends on the variance of our estimates. For example, if the ε_t are i.i.d. $N(0, \sigma^2)$, then

$$\text{var}(\widehat{\mu}(t)) = \sigma^2 \left(\frac{1}{n} + \frac{(t - \overline{t})^2}{\sum_{i=1}^{n} (t_i - \overline{t})^2} \right).$$

Nottingham Castle air temperature



Mauna Loa CO_2 concentration



Airline passengers

We investigate how we can fit a trend with R. See "RcodeforAirlineTimeSeries" on the VLE.

1.2 Seasonal effects

- *Seasonal effects* are periodic components of a time series which repeat with a fixed frequency.
- We might revise model (1) to be

$$X_t = \mu(t) + s(t) + \varepsilon_t, \tag{2}$$

where
$\mu(t)$ — trend;
$s(t)$ is a seasonal effect, with $s(t) = s(t+a)\ \ \forall t$ where a is the period length;
ε_t — residuals/fluctuations.

- If we assume $t = 0, 1, 2, \ldots$ and $a = n \in \mathbb{N}$, then we have

$$s = (s_0\ s_1\ \ldots\ s_{n-1}\ s_0\ s_1\ \ldots\ s_{n-1}\ \ldots)$$

so $s(t) = s_{t \bmod n}$. To find s we have to estimate $s_0, s_1, \ldots, s_{n-1}$.

- We can write

$$X_t = \mu(t) + s_0\delta_{t,0} + \cdots + s_{n-1}\delta_{t,n-1} + \varepsilon_t,$$

where

$$\delta_{t,i} = \begin{cases} 1 & \text{if } t \text{ is in season } i \\ 0 & \text{else.} \end{cases}$$

- The parameters $s_0, \ldots, s_{n-1}$ can be estimated using linear regression.
- For example, if t is seasonal in quarters of a year and if the period length is one year, then we have to estimate

  s_0 (spring),
  s_1 (summer),
  s_2 (autumn), and
  s_3 (winter).
- For the airline data, we can estimate s_0 (Jan) — s_{11} (Dec).

1.3 Summary

- We start the analysis of a time series by identifying the trend (if any) in the model $X_t = \mu(t) + \varepsilon_t$. We then analyse the residuals $Y_t = X_t - \widehat{\mu}(t)$ further. Finally, any results about the $\{Y_t\}$ need to be converted back by adding $\widehat{\mu}(t)$.
- For example, a forecast for X_t is given by the forecast for $Y_t + \widehat{\mu}(t)$.
- After the trend is removed, we remove the periodic components (if any) in the model $Y_t = s(t) + \varepsilon_t$. We then analyse $Z_t = Y(t) - \widehat{s}(t)$ further. Convert results about the $\{Z_t\}$ into results about $\{Y_t\}$ by adding $\widehat{s}$.