Applied Time Series Analysis
FISH 507

Eric Ward

Mark Scheuerell

Eli Holmes

Introductions

- Who are we?
- Who & why you're here?
- What are you looking to get from this class?

Days and Times

- Lectures

  When: Tues & Thurs from 1:30-2:50

  Where: FSH 203

- Computer lab

  When: Thurs from 3:00-3:50

  Where: FSH 207

Grading

- Weekly homework (30% of total)
  - Assigned Thurs at the end of computer lab
  - Due by 5:00 PM the following Tues
  - Based on material from lecture & computer lab
- Research project & paper (40% of total)
  - Must involve some form of time series model(s)
  - Due by 11:59 PM PST on March 10
- Two anonymous peer-reviews (20% of total)
  - One review each for 2 colleague's papers
  - Due by 11:59 PM PST on March 16

Expectations for final project

- Research paper or thesis chapter that you can turn into a peer-reviewed publication
- Ideally a solo effort, but you can work in pairs
- Focus on applied time series analysis
  - Univariate or multivariate
- Short format similar to "Report" in *Ecology* or "Rapid Communication" in *CJFAS*
  - Max of 20 pages, inclusive of refs, tables, figs, etc
  - 12-pt font, double-spaced throughout

Don't have any time series data?

- RAM Legacy
  http://ramlegacy.marinebiodiversity.ca/
- RAM's Stock-Recruitment Database
  http://www.mscs.dal.ca/~myers/welcome.html
- Global Population Dynamics Database
  http://www3.imperial.ac.uk/cpb/databases/gpdd
- NOAA NWFSC Salmon Population Summary
  https://www.webapps.nwfsc.noaa.gov/apex/f?p=261:home:0
- SAFS
  - Alaska Salmon Program
  - Lake Washington plankton

Course topics

Week 1: Decomposition, covariance, autocorrelation

Week 2: Autoregressive & moving-average models, model estimation

Week 3: Univariate & multivariate state-space models

Week 4: Covariates & seasonal effects; model selection

Week 5: Dynamic linear models

Week 6: Forecasting & dynamic factor analysis

Week 7: Multistage & non-Gaussian models

Week 8: Detection of outliers & perturbation analysis

Week 9: Spatial effects & hierarchical models

Week 10: Presentations of final projects

An introduction to time series and their analysis

Mark Scheuerell

FISH 507 – Applied Time Series Analysis

3 January 2017

Topics for today (lecture)

- Characteristics of time series (ts)
  - What is a ts?
  - Classifying ts
  - Trends
  - Seasonality (periodicity)
- Classical decomposition

What is a time series?

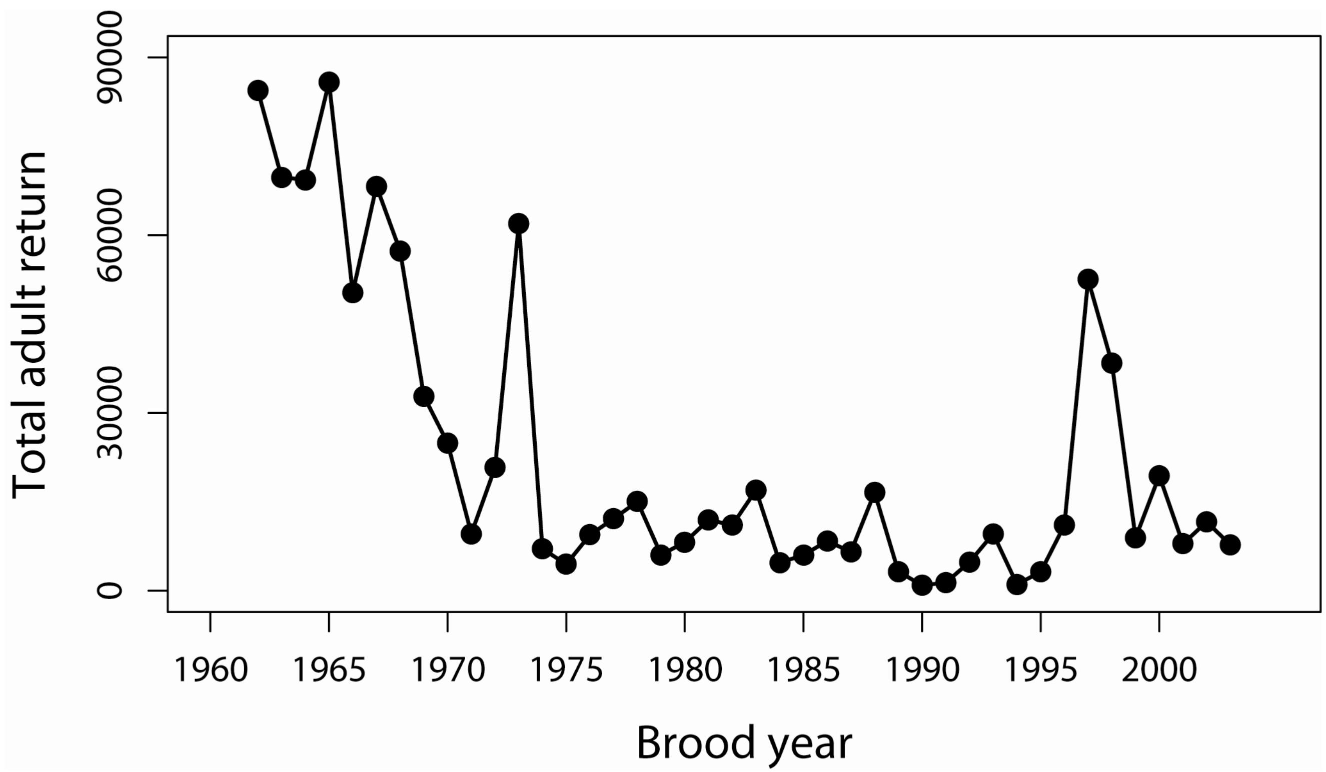
- A *time series* (ts) is a set of observations taken sequentially in time
- A ts can be represented as a set

  $\{x_t : t = 1,2,3,\ldots,n\} = \{x_1,x_2,x_3,\ldots,x_n\}$
- For example,

  {10,31,27,42,53,15}

Example of a time series

Number of wild spr/sum Chinook salmon returning to the Snake R



Classification of time series (I)

I. By some index set

  A. Interval across real time $x(t)$; $t \in [1.1,2.5]$

  B. Discrete time x_t

    1. Equally spaced; $t = \{1,2,3,4,5\}$
    2. Equally spaced w/ missing values; $t = \{1,2,4,5,6\}$
    3. Unequally spaced; $t = \{2,3,4,6,9\}$

Classification of time series (II)

II. By underlying process

   A. Discrete (eg, total # of fish caught per trawl)

   B. Continuous (eg, salinity, temperature)

Classification of time series (III)

III. By number of values recorded

   A. Univariate/scalar (eg, total # of fish caught)

   B. Multivariate/vector (eg, # of each spp of fish caught)

Classification of time series (IV)

IV. By type of values recorded

A. Integer (eg, # of fish in 5 min trawl = 2413)

B. Rational (eg, fraction of unclipped fish = 47/951)

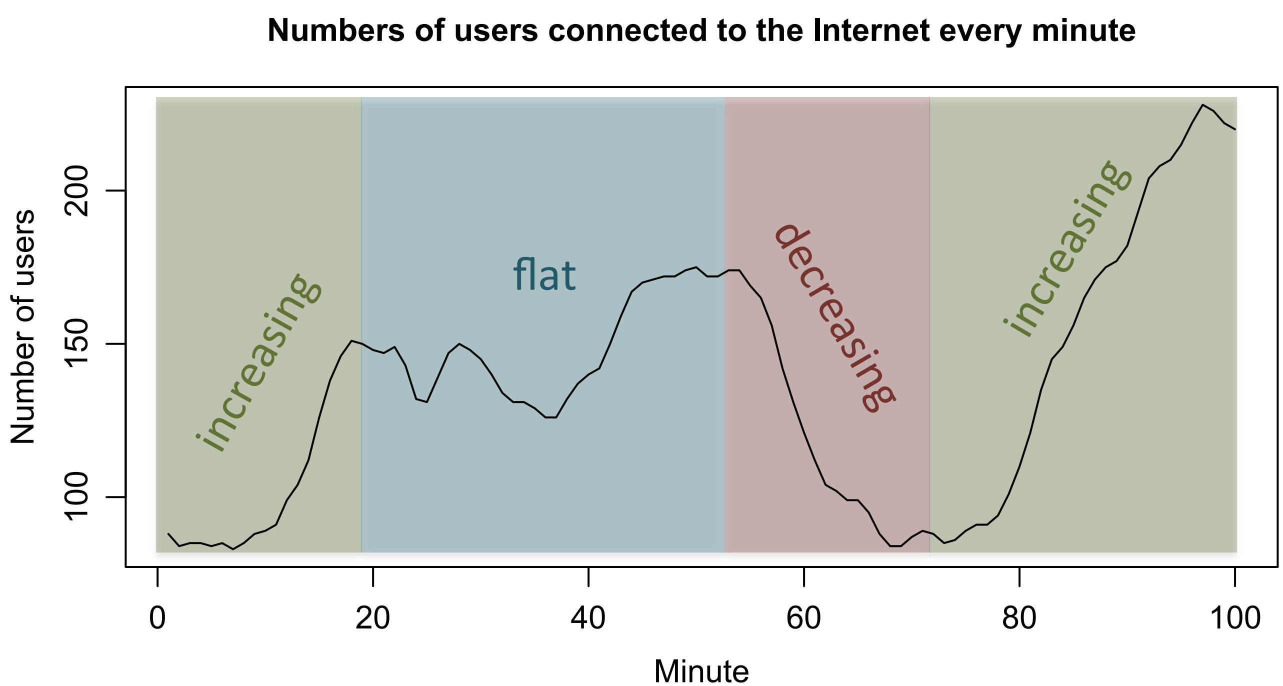
C. Real (eg, fish mass = 10.2 g)

D. Complex (eg, $\cos[2\pi*2.43] + i \sin[2\pi*2.43]$)

Statistical analyses of time series

- Most statistical analyses are concerned with estimating properties of a population from a sample
- Time series analysis, however, presents a different situation
- Although we could vary the length of an observed sample, it is often impossible to make multiple observations at a given time
- For example, one can't observe today's closing price of Microsoft stock more than once
- This makes conventional statistical procedures, based on large sample estimates, inappropriate

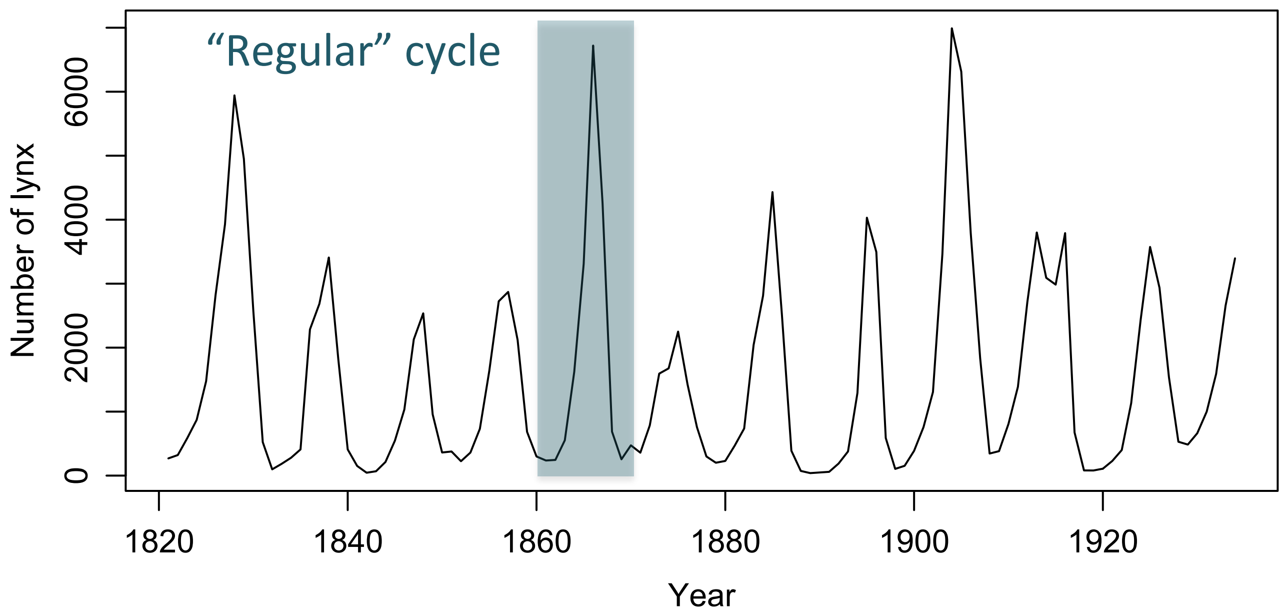
Examples of time series



How would we describe this ts?

Examples of time series

Annual numbers of lynx trapped in Canada from 1821-1934



How would we describe this ts?

What is a time series model?

- A *time series model* for $\{x_t\}$ is a specification of the joint distributions of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is thought to be a realization

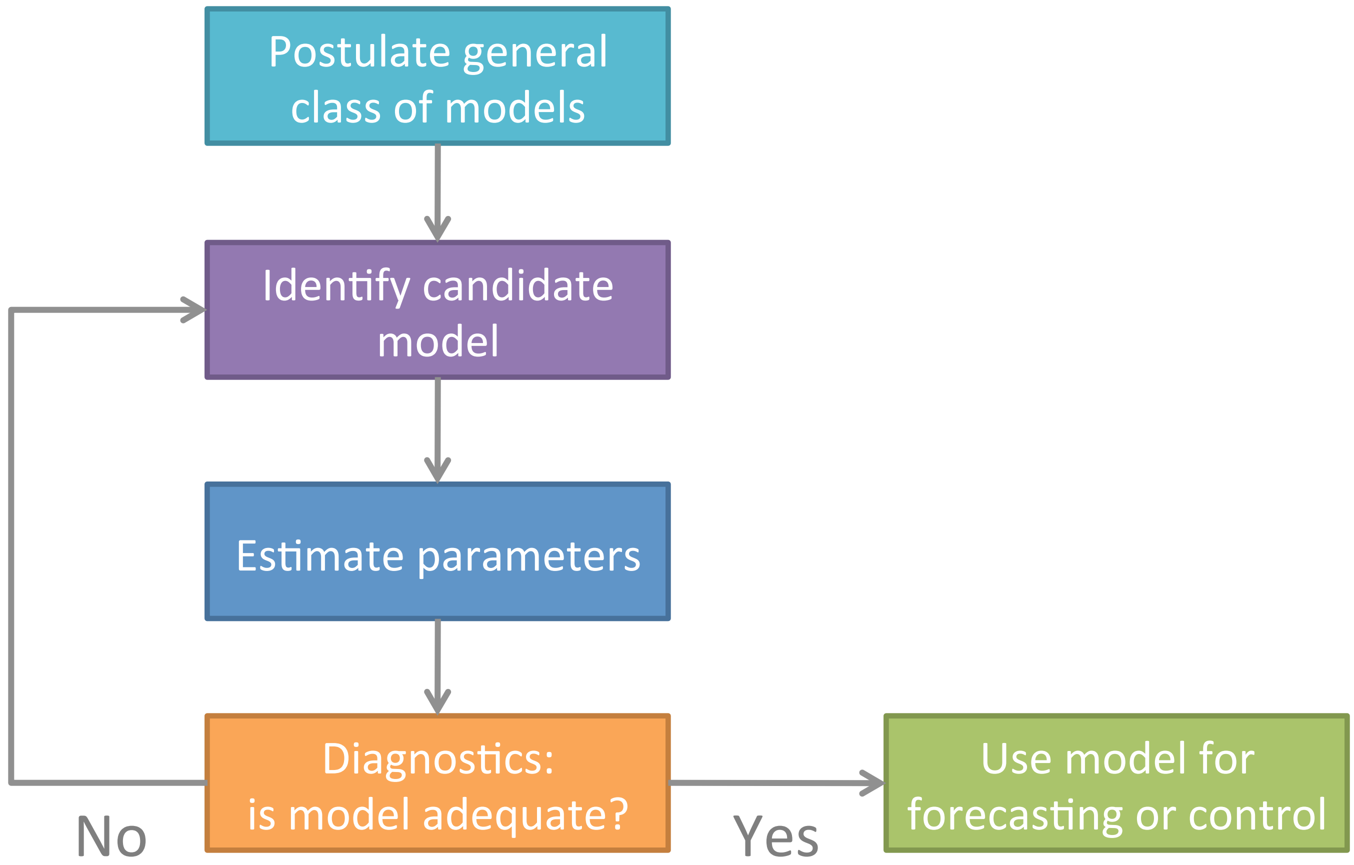
- For example,

“white” noise: $x_t = w_t$ and $w_t \sim N(0,1)$

autoregressive: $x_t = x_{t-1} + w_t$ and $w_t \sim N(0,1)$

Iterative approach to model building

Also known as the "Box-Jenkins Approach"



Classical decomposition of time series

- *Classical decomposition* of an observed time series is a fundamental approach in time series analysis
- The idea is to decompose a time series $\{x_t\}$ into a trend (m_t), a seasonal component (s_t), and a remainder (e_t)

$$x_t = m_t + s_t + e_t$$

Linear filtering of time series

- Beginning with the trend (m_t), we need a means for extracting a “signal”

- A common method is to use linear filters

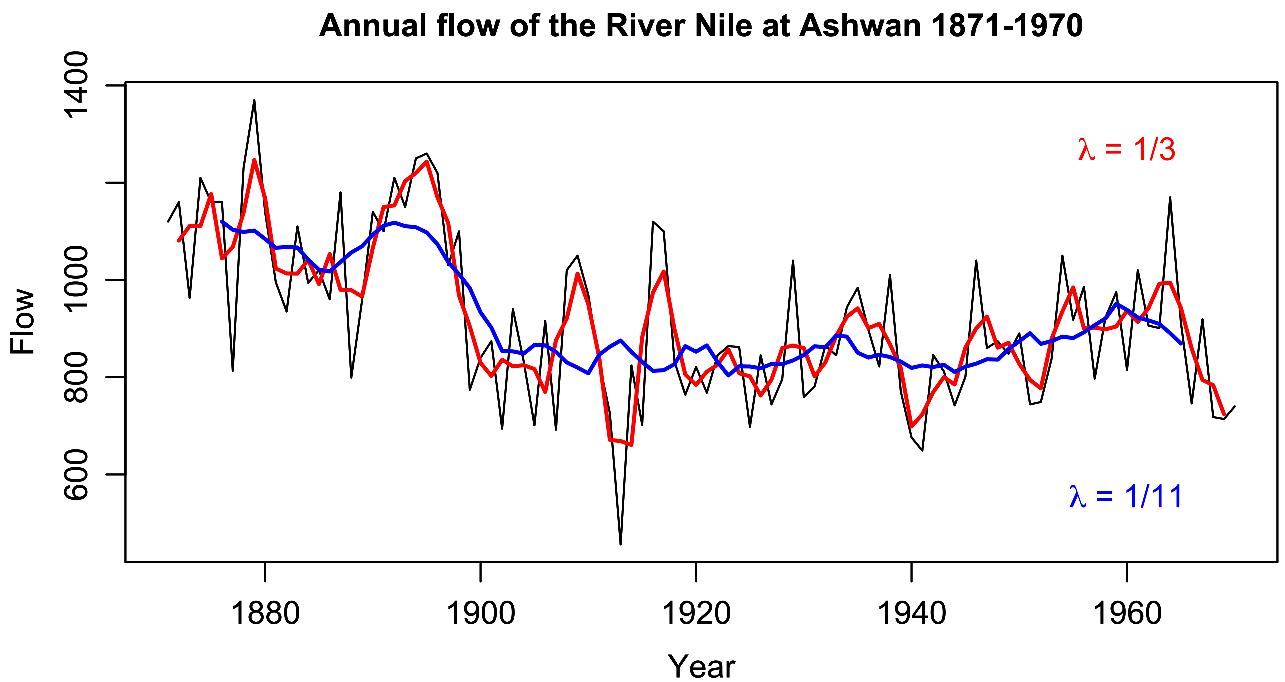
$$m_t = \sum_{i=-\infty}^{\infty} \lambda_i x_{t+i}$$

- For example, moving averages with equal weights

$$m_t = \sum_{i=-a}^{a} \frac{1}{2a+1} x_{t+i}$$

(FYI, this is what Excel does)

Example of linear filtering

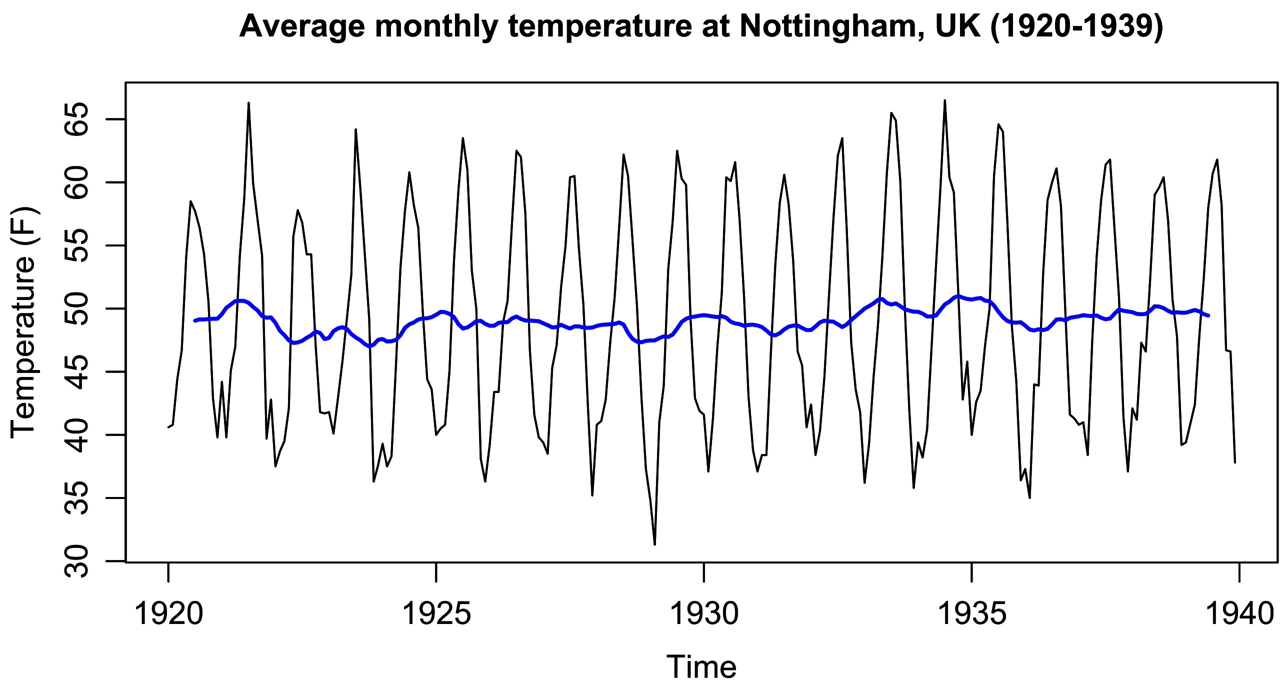


Linear filtering of time series

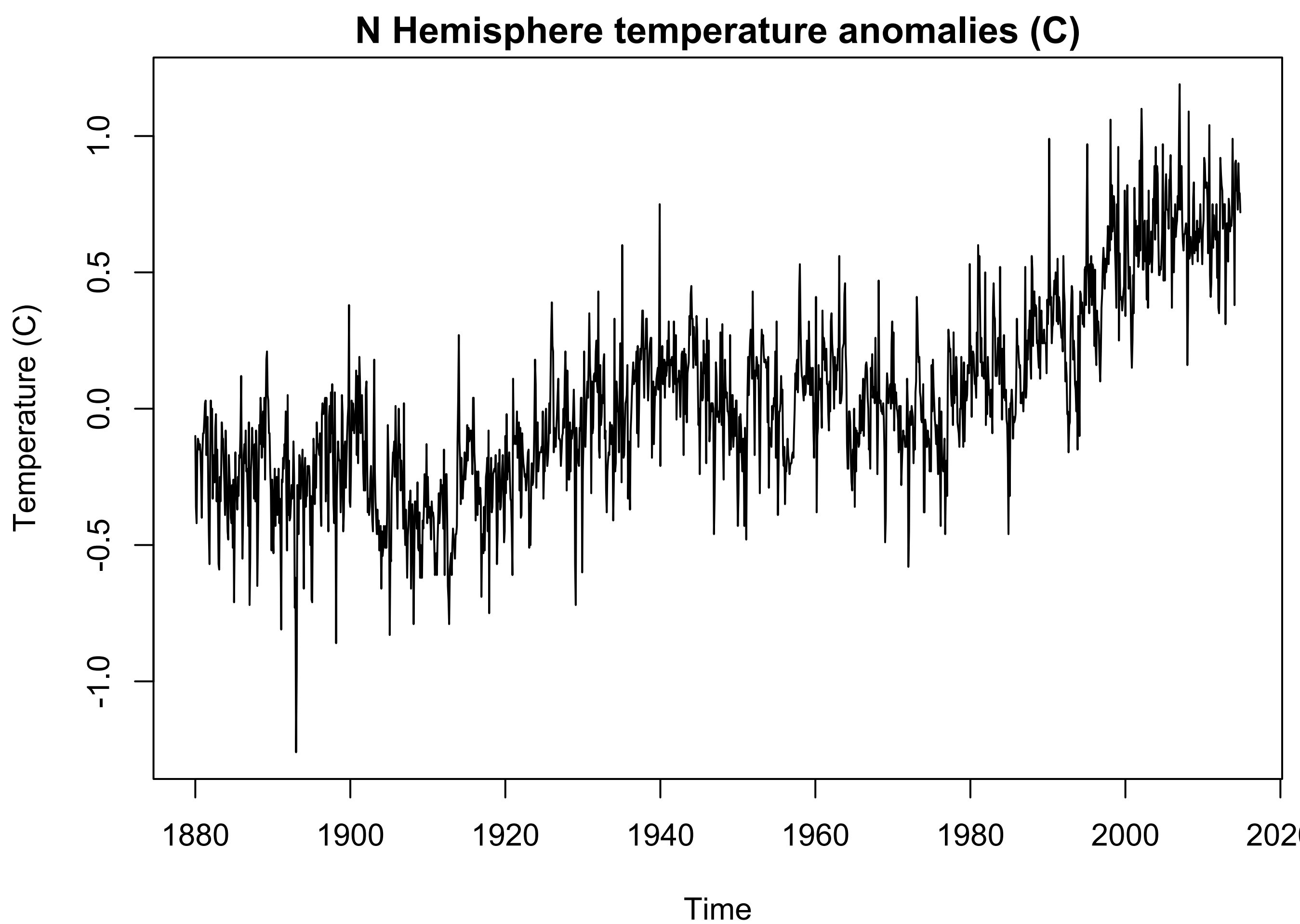
- Consider case where season is based on 12 months & ts begins in January (t=1)
- Monthly averages over year will result in t = 6.5 for m_t (which is not good)
- One trick is to average (1) the average of Jan-Dec & (2) the average of Feb-Jan

$$m_t = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + \cdots + x_{t-1} + x_t + x_{t+1} + \cdots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}$$

Example of linear filtering

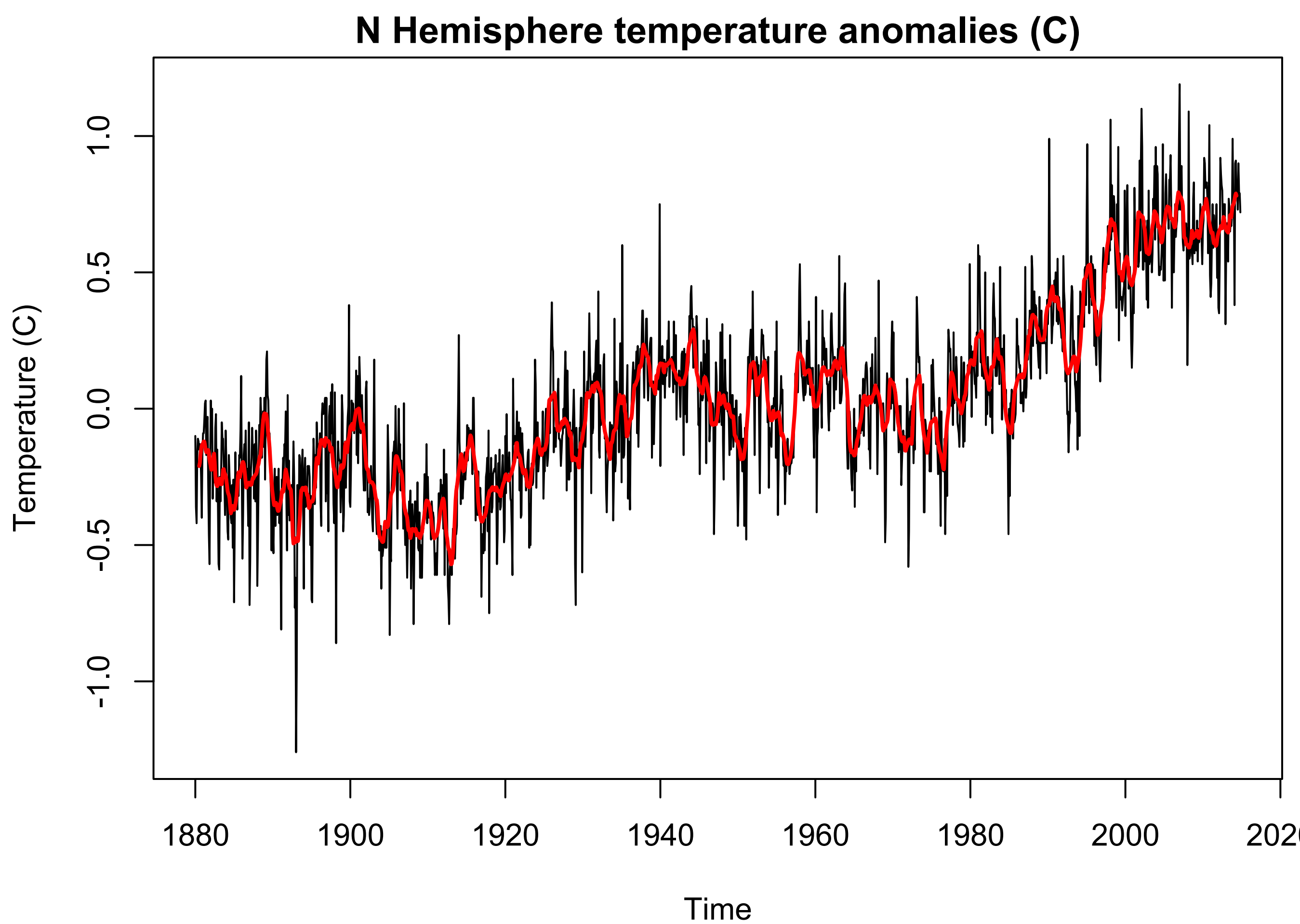


Example of linear filtering



Data from http://www.ncdc.noaa.gov/

Example of linear filtering

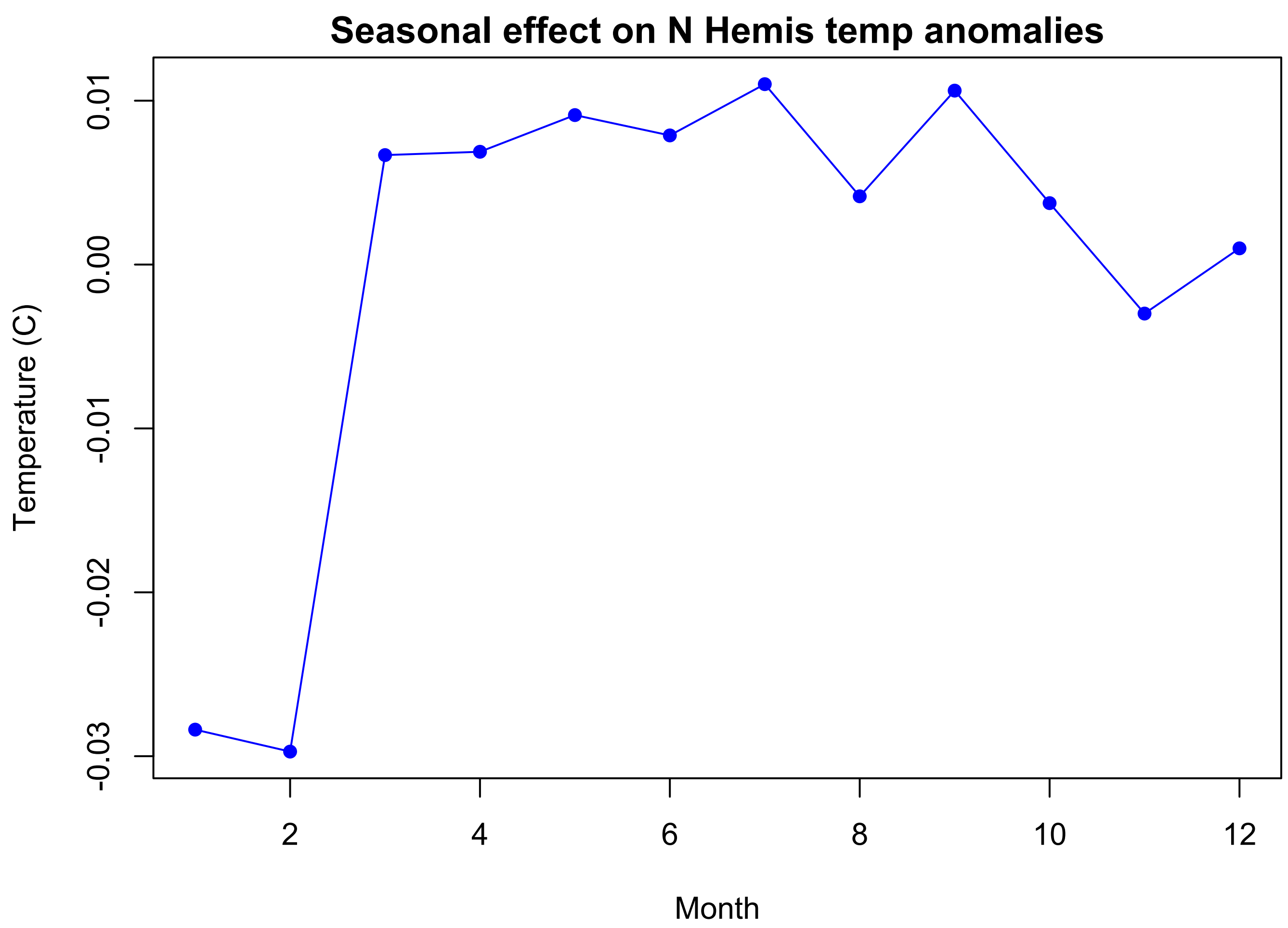


Decomposition of time series

Now that we have an estimate of m_t, we can get estimate of s_t simply by subtraction:

$$\hat{s}_t = x_t - \hat{m}_t$$

Example of linear filtering

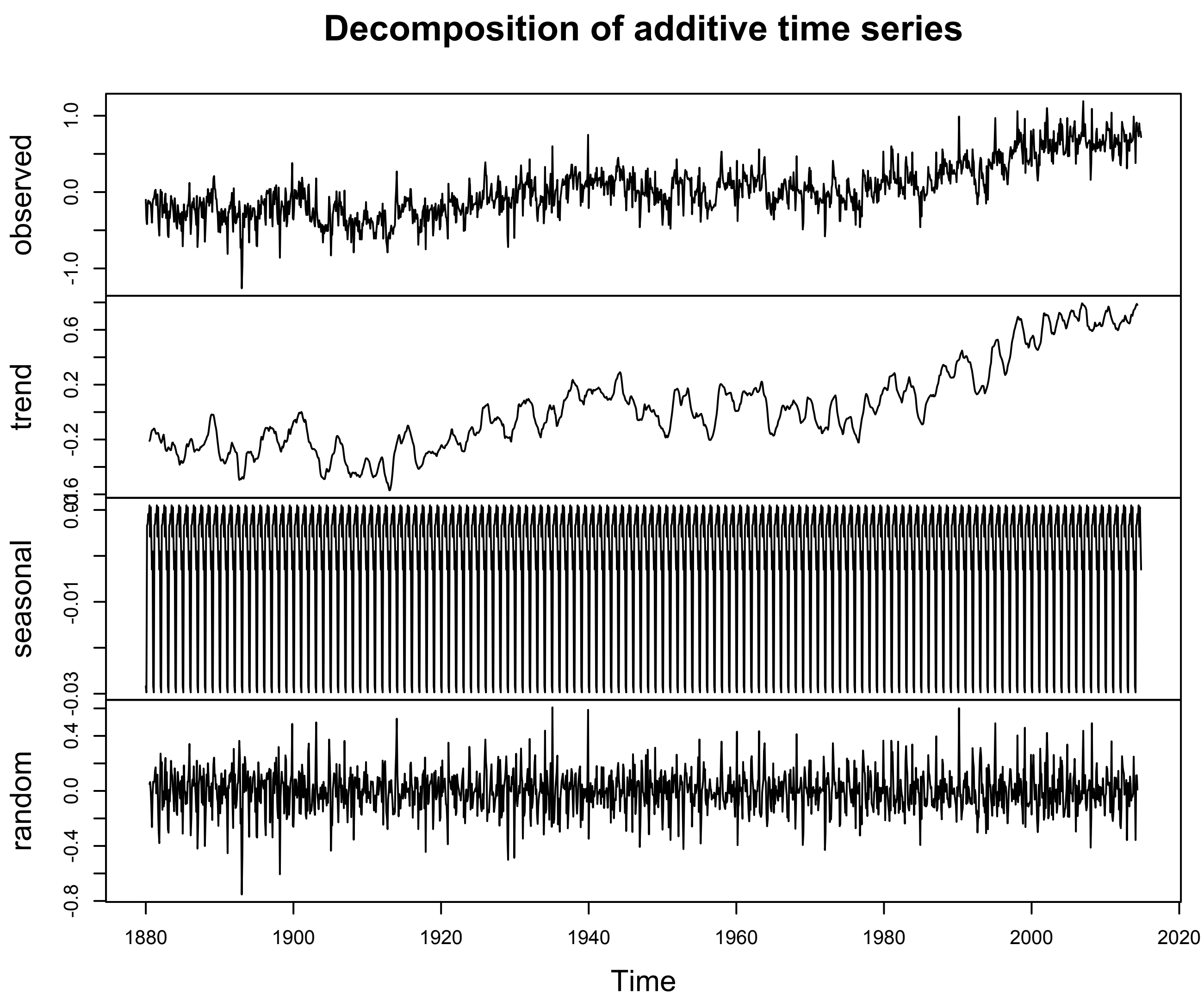


Decomposition of time series

Now that we have an estimate of s_t, we can get estimate of e_t simply by subtraction:

$$\hat{e}_t = x_t - \hat{m}_t - \hat{s}_t$$

Example of linear filtering



Notes on decomposition

- Obtaining a “model” for a ts via decomposition is easy, but…
- You don’t get a formula with which to obtain forecasts
- Let’s look at an alternative

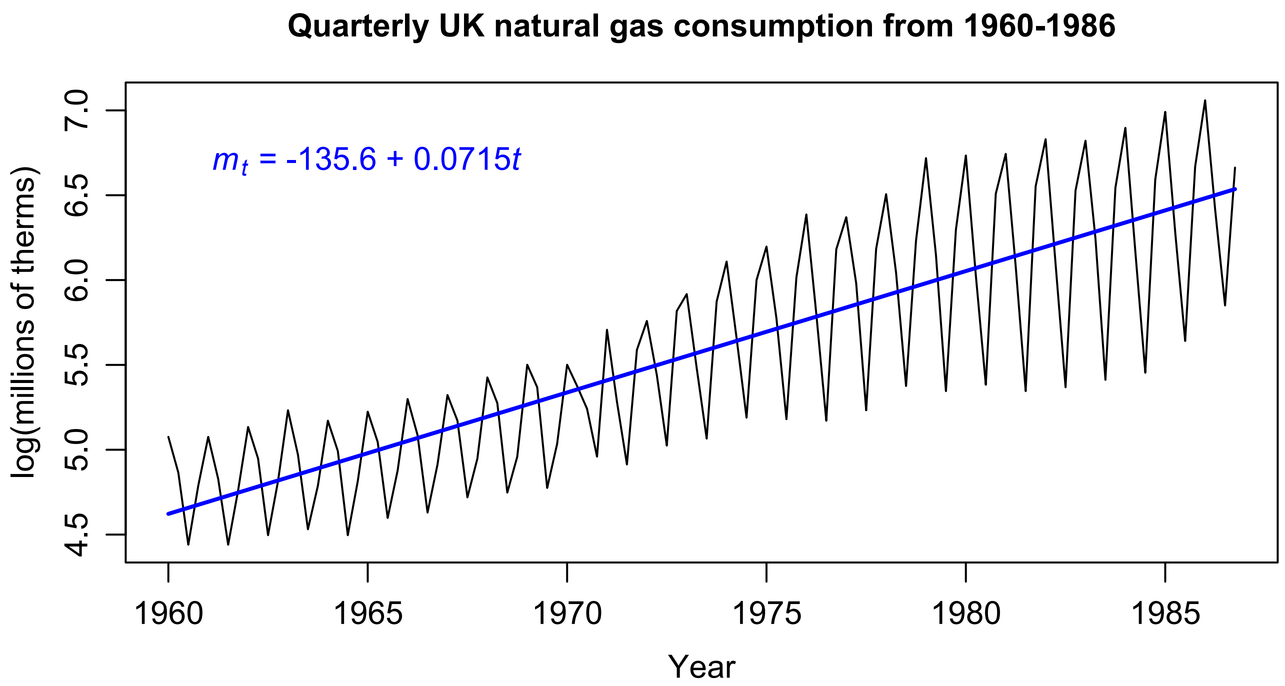
Example of linear trend fitting

- A simple method for trend extraction is to use linear regression

$$m_t = \alpha + \beta t + e_t$$

- Note: the *t* index here could be a non-integer in cases with seasonal data

Example of linear trend fitting



Decomposition of time series

- Another means for extracting a trend is via nonparametric regression models (eg, LOESS)
- see R pkg `stl`

Moving on with decomposition

- We have decomposed our time series into a trend plus remainder ($s_t + e_t$)

$$x_t = (-135.6 + 0.0715t) + s_t + e_t$$

- Now let's consider the seasonal part

Example of linear trend fitting

- One method is to use fixed effects (eg, ANOVA)

$$x_t = m_t + s_t + e_t$$

$$x_t = (\alpha + \beta t) + s_t + e_t$$

Example of linear trend fitting

- Adding in a model for season (ie, quarters)

$$x_t = -137 + 0.072 \cdot t + s\left(q\left(t\right)\right) + e_t$$

$$q\left(t\right) = 4\left(t - \lfloor t \rfloor + 1\right) - 3$$

This is the "floor" function

- So, for example, if q = 10.25:

$$q\left(10.25\right) = 4\left(10.25 - \lfloor 10.25 \rfloor + 1\right) - 3$$

$$q\left(10.25\right) = 4\left(10.25 - 10 + 1\right) - 3$$

$$q\left(10.25\right) = 4\left(1.25\right) - 3$$

$$q\left(10.25\right) = 5 - 3 = 2$$

Example of linear trend fitting

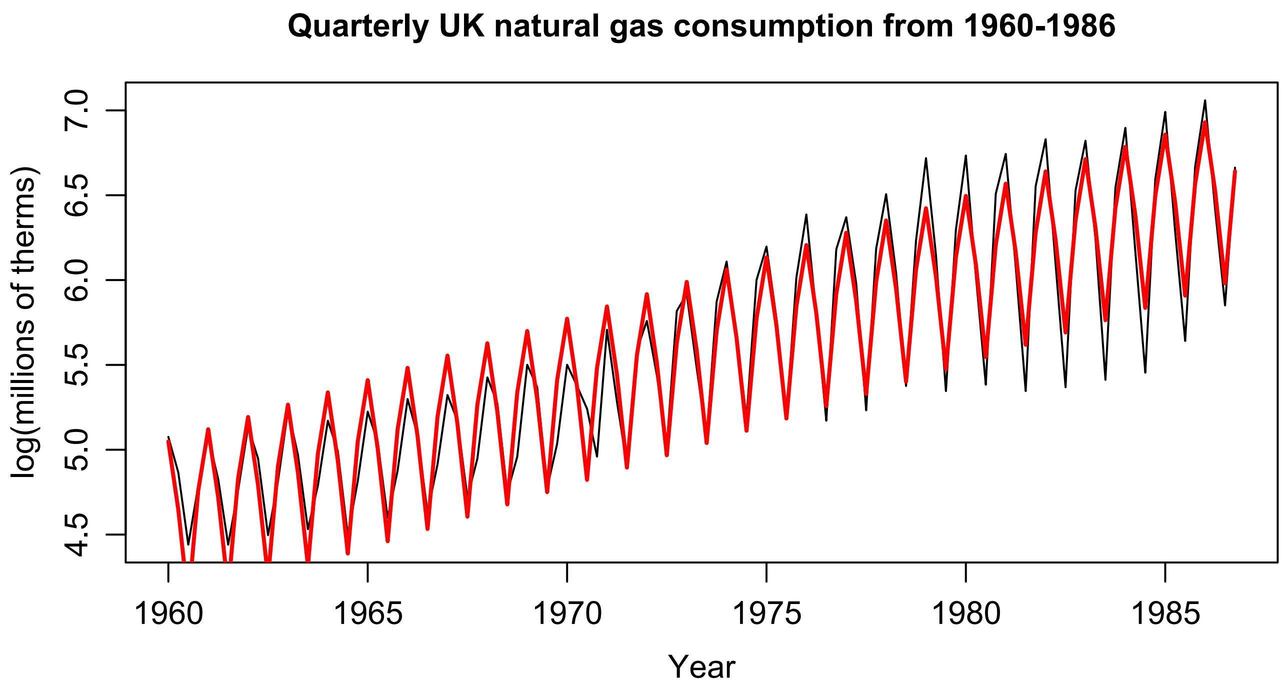
- Our final decomposition model

$$x_t = -137 + 0.072 \cdot t + \theta\left(q\left(t\right)\right) + e_t$$

$$q\left(t\right) = 4\left(t - \lfloor t \rfloor + 1\right) - 3$$

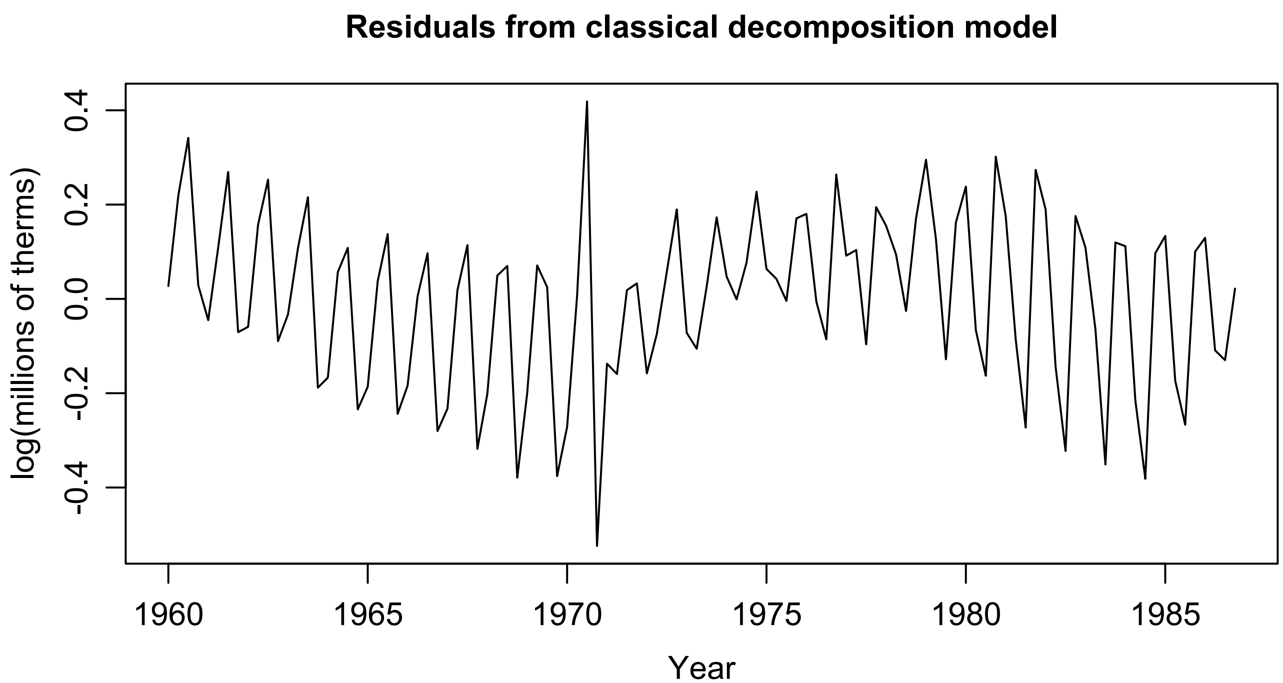
$$s\left(q\left(t\right)\right) = \begin{cases} 0 & \text{if } q = 1 \\ -0.42 & \text{if } q = 2 \\ -0.99 & \text{if } q = 3 \\ -0.34 & \text{if } q = 4 \end{cases}$$

Example of trend + season fitting

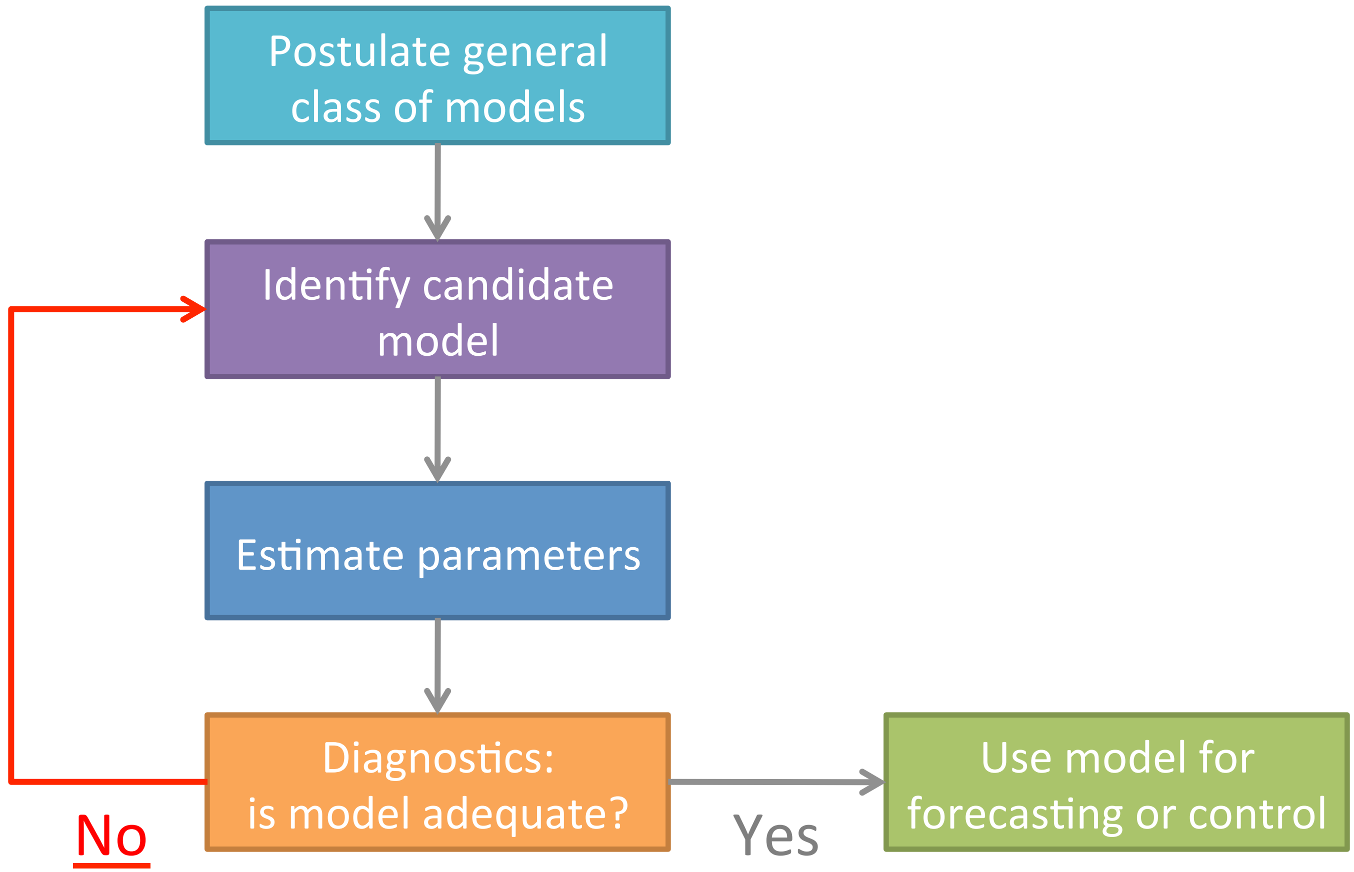


Are the residuals stationary?

- The goal with decomposition is to reduce the time series to a trend, season & stationary residuals



Iterative approach to model building



Summary

- This was a *brief* overview—there is *lots* of stuff we didn't cover
- Please ask for help/guidance if you're looking for more details, other R code, etc