Time Series Analysis

Gavin Simpson

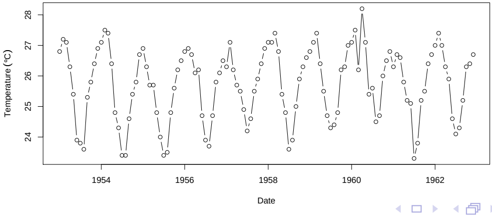
Institute of Environmental Change and Society
University of Regina

Adelaide, February 2017

Outline

Introduction

- A Time Series is a collection of observations made sequentially through time
- A continuous time series is one where observations are made continuously through time
  - *Continuous* refers to the measurement of observations not the type of variable that is observed
- A discrete time series is one where the observations are taken at specific time points
  - Sampling points are generally equally spaced in time
- **deterministic** vs. **stochastic**



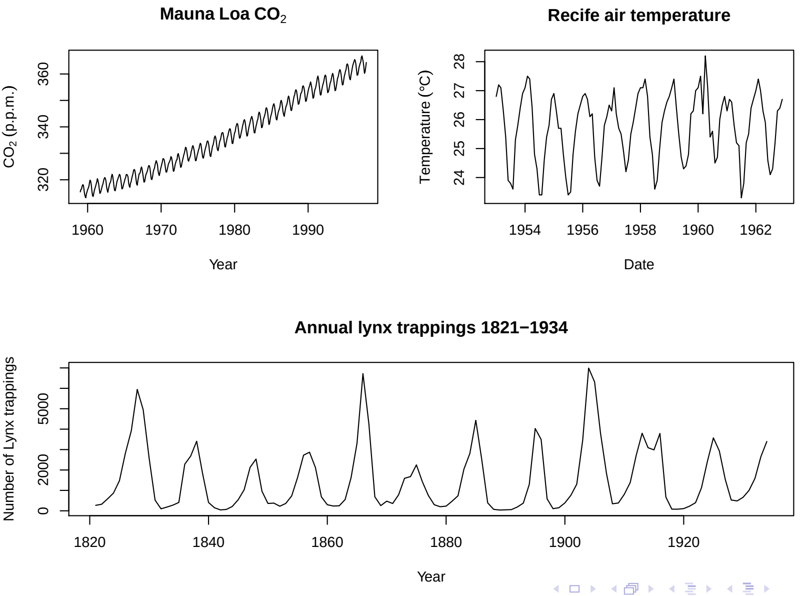
Objectives of time series analysis

- Description
  - Time plots of observations; a simple way to describe temporal patterns in a time series
  - regular **seasonal effects** or **cyclicity**, presence of a **trends**, **outliers**, **sudden changes** or **breaks**
- Explanation
  - Observations on one variable in time may be used to explain the variation in another series
  - May help understand the **mechanisms** that generated a given time series
- Prediction
  - Given an observed time series one may want to predict future values of the series — also called **forecasting**
- Control
  - Time series often collected to improve *control* over a physical process
  - Monitoring to alert when conditions exceed an *a priori* determined threshold

Descriptive Techniques — types of variations

- Traditional time series methods are often concerned with decomposing variation in a time series in components representing **trend**, **seasonal** or other **cyclic** variation. Remaining variation is attributed to **irregular** fluctuations
- **Seasonal variation**
  - Variation that is **annual** in period
  - Easily estimated if of interest, or removed — deseasonalised
- **Cyclic variation**
  - Variation that is **fixed** in period — diurnal temperature variation
  - Oscillations without a fixed period but are predictable to some extent
- **Trend**
  - **Long term change in the mean level**
  - Trend is a function of the length of the record
- Other **irregular** fluctuations
  - Variation remaining after removal of trend and cyclic variations
  - May or may not be **random**

Types of variation



Transformations 1

- Transform time series for similar reasons as for any other type of data
  - to stabilise the variance
    - If trend present and variance of series increases with mean; $\log$ transform
    - If no trend but variance increases with mean then a transformation is of little use
  - to make seasonal component additive
    - If seasonal component increases with the mean in presence of a trend, said to be **multiplicative**
    - Transform (e.g. $\log$) to make the seasonal component constant from year to year; **additive**
    - Transformation will only stabilise the variance if the error term is also multiplicative
  - to make the data normally distributed
    - Model building usually assumes data are normally distributed
    - 'spikes' in the time plot will show up as skew in the distribution — can be difficult to remove
- Inherently difficult, however…

Transformations 2

- Seasonal components
  - Additive: $X_t = m_t + S_t + \varepsilon_t$
  - Multiplicative: $X_t = m_t S_t + \varepsilon_t$
  - Multiplicative: $X_t = m_t S_t \varepsilon_t$
  - Only the latter will be improved by a transformation
- A transformation that makes the seasonal component additive may fail to stabilise the variance
- As such we may not be able to achieve all the aims on previous slide
- A model constructed on transformed data less useful than one fitted to raw data
  - May be more difficult to interpret to models fitted to transformed data
  - Forecasts need to be back transformed
- Avoid transformation where possible, though use them if they make physical sense (e.g. $\log$ or square root for abundances or percentages)

Differencing

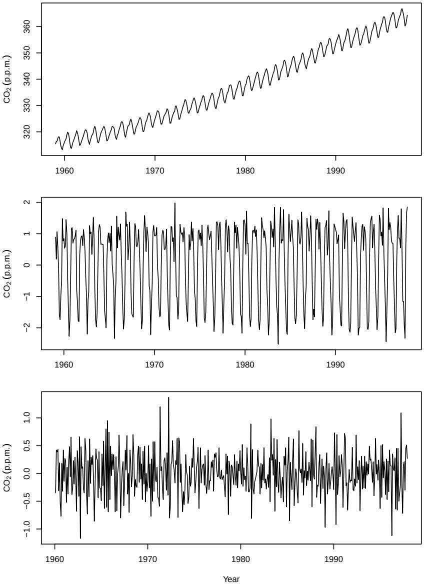
- **Differencing** is a special type of filtering useful for removing trends and seasonality to produce a stationary series
- First order differencing; new series formed by subtracting x_{t-1} from x_t

$$\nabla x_t = x_t - x_{t-1}$$

- **Seasonal differencing**; e.g. for monthly data

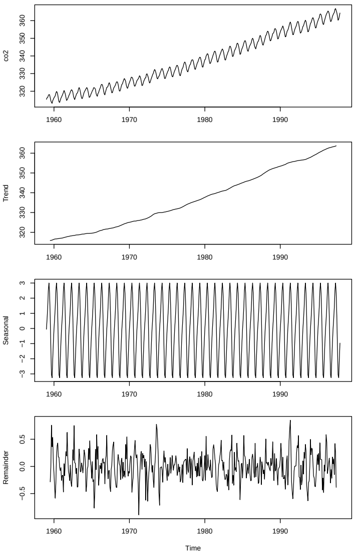
$$\nabla_{12} x_t = x_t - x_{t-12}$$

- Raw CO_2 data (upper); $\nabla_1 CO_2$ (middle); $\nabla_1\{\nabla_{12} CO_2\}$ (lower)



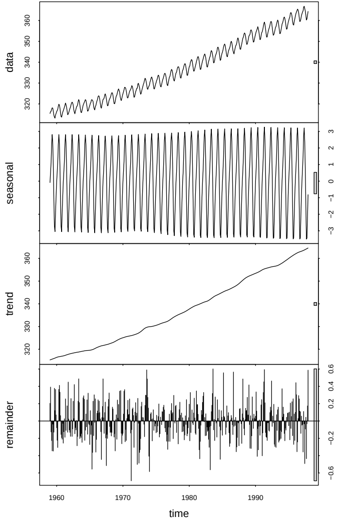
Decomposing time series — classical approach

- Decompose series into **trend**, **seasonal**, and **random** components
- $x_t = \mathrm{Trend}_t + \mathrm{Seasonal}_t + \mathrm{remainder}_t$
- Moving average filter used to identify the trend
- Compute seasonal component as the average over the detrended series of each period (e.g. month or quarter)
- Seasonal component is formed from the period averages repeated to match the length of the original series
- Random component is the remainder once the trend and the seasonal components have been subtracted from the original series



Decomposing time series — LOESS approach (STL)

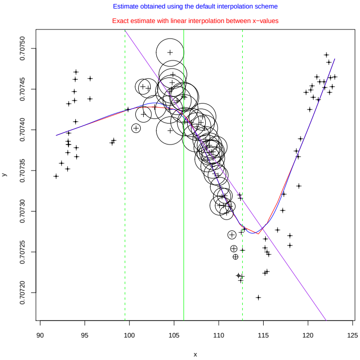
- Decompose series into **trend**, **seasonal**, and **random** components using LOESS
- The seasonal component is found by LOESS smoothing of the seasonal sub-series (e.g. series of January values)
- x_t is deseasonalised and this series is smoothed to find the trend
- Overall level subtracted from seasonal series and added to the trend
- This process is repeated a few times until convergence
- Remainder is the residuals of the trend + seasonal components



Lowess — Locally weighted regression

Locally weighted regression scatterplot smoother

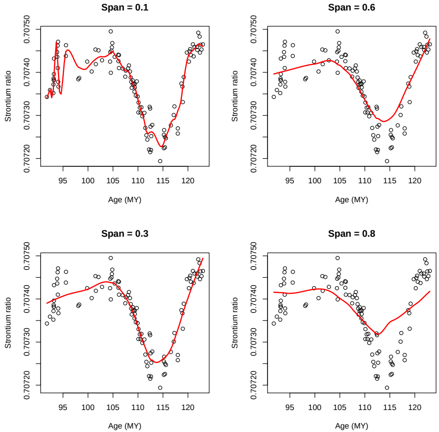
- Decide how smooth relationship should be (span or size of bandwidth window)
- For target point assign weights to observations based on adjacency to target point
- Fit linear (polynomial) regression to predict target using weighted least squares; repeat
- Compute residuals & estimate robustness weights based on residuals; well-fitted points have high weight
- Repeat Loess procedure with new weights based on robustness and distance weights



Try different span and degree of polynomial to optimise fit

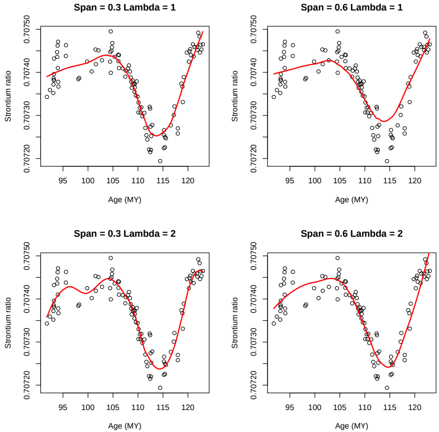
Lowess — Locally weighted regression

- Two key choices in Loess
- α is the span or bandwidth parameter, controls the size of the window about the target observation
- Observation outside the window have 0 weight
- Larger the window the more global the fit — smooth
- The smaller the window the more local the fit — rough
- λ is the degree of polynomial using the the weighted least squares
- $\lambda = 1$ is a linear fit, $\lambda = 2$ is a quadratic fit



Lowess — Locally weighted regression

- Two key choices in Loess
- α is the span or bandwidth parameter, controls the size of the window about the target observation
- Observation outside the window have 0 weight
- Larger the window the more global the fit — smooth
- The smaller the window the more local the fit — rough
- λ is the degree of polynomial using the the weighted least squares
- $\lambda = 1$ is a linear fit, $\lambda = 2$ is a quadratic fit



Lowess — Locally weighted regression

"In any specific application of LOESS, the choice of the two parameters α and λ must be based upon a combination of judgement and trial and error. There is no substitute for the latter"

Cleveland (1993) Visualising Data. AT&T Bell Laboratories

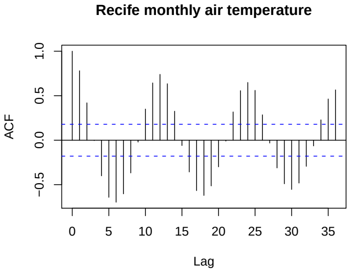
- CV can be used to optimise α and λ to guard against overfitting the local pattern by producing too rough a smoother or missing local pattern by producing too smooth a smoother
- Loess is perhaps most useful as an exploratory technique as part of EDA
- Cleveland, W.S. (1979) J. Amer. Stat. Assoc. **74**, 829–836
- Cleveland, W.S. (1994) The Elements of Graphing Data. AT&T Bell Laboratories
- Efron, B & Tibshirani, R (1981) Science **253**, 390–395

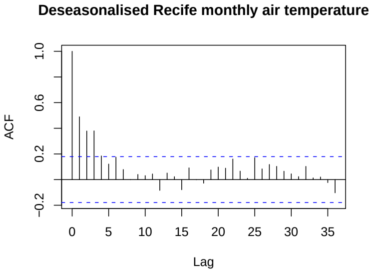
Autocorrelation function

- **Sample autocorrelation coefficients** are an important guide to the properties of time series
- Measure the correlation between observations at different distances apart

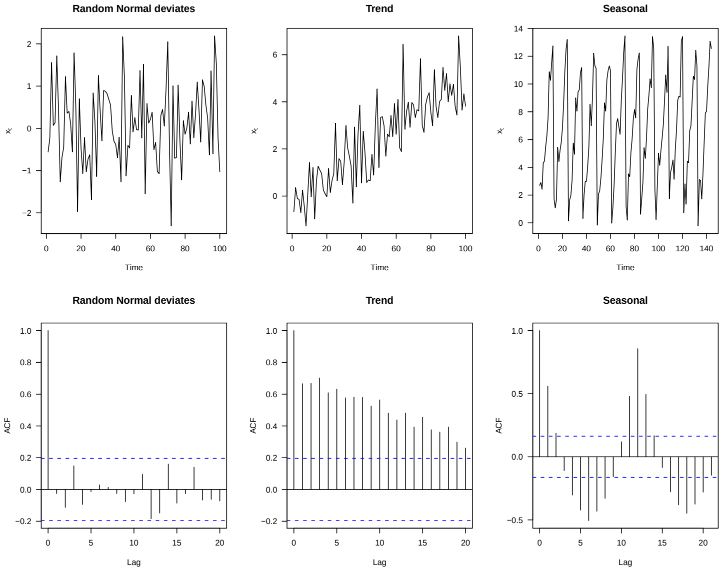
$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2}$$

- Computed for small number of lags k
- Use min. of 36 lags to view several seasonal cycles
- Dashed lines drawn at $\pm 2/\sqrt{n}$ enclose insignificant correlations
- Display on a **correlogram**





Example correlograms 1

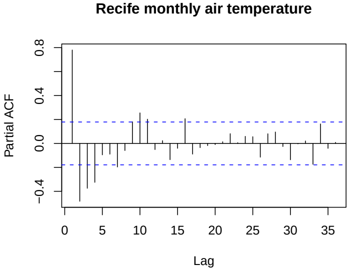


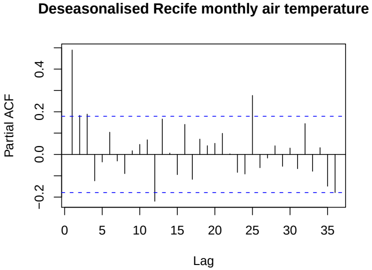
Partial autocorrelation function

- If x_t and x_{t-1} are highly correlated then x_{t-1} and x_{t-2} will also be highly correlated
- Because x_t and x_{t-2} are highly correlated with x_{t-1}, it is likely that x_t and x_{t-2} are also highly correlated
- It would be nicer if we could estimate correlation between x_t and x_{t-2} after removing the effect of x_{t-1}
- This is the **partial autocorrelation**
- The partial autocorrelation α_k is obtained as coefficient β_k from the regression

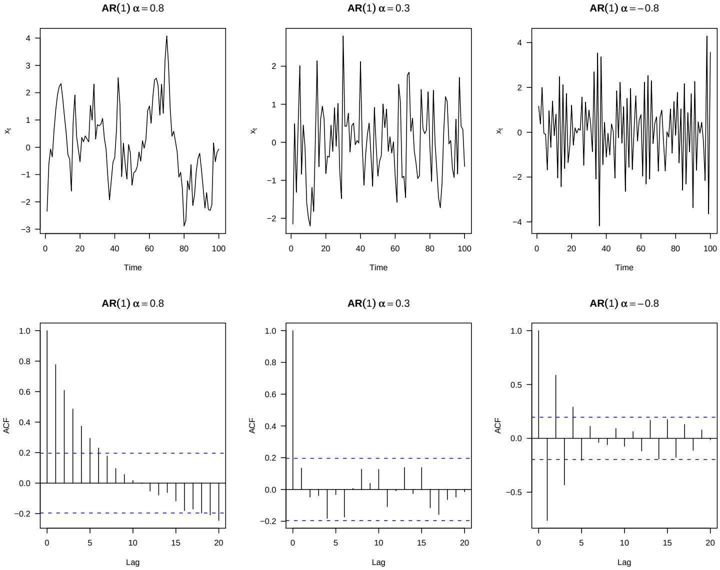
$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_k x_{t-k}$$

- This is an **autoregressive** (AR) process of order k



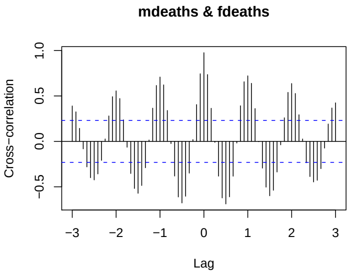


Example correlograms 2



Cross-correlation function

- **Sample cross-correlation function** measures the correlation between observations of two series at different lags



$$r_{xy}(k) = \begin{cases} \dfrac{1}{n} \dfrac{\sum\limits_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y})}{s_x s_y} & k = 0, 1, \ldots, n-1 \\ \dfrac{1}{n} \dfrac{\sum\limits_{t=1-k}^{n} (x_t - \bar{x})(y_{t+k} - \bar{y})}{s_x s_y} & k = -1, -2, \ldots, -(n-1) \end{cases}$$

Outline

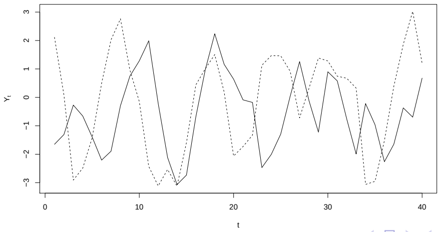
Stochastic Trends

- Tend to think of a trend as a deterministic one polluted by noise

$$Y_t = \beta_0 + \beta_1 \mu_t + \varepsilon_t$$

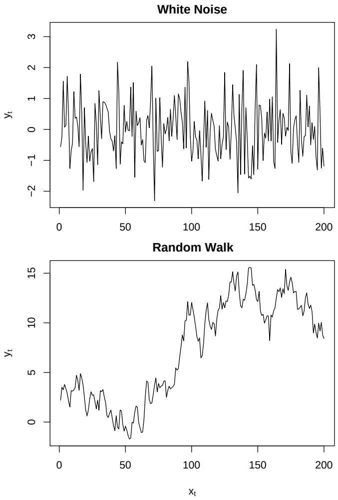
- Other types of trend may be at work; stochastic trends
- Here, variation in a time series is determined solely via dependence between successive observations rather than a fixed, deterministic trend
- Two stochastic trends fitted from the model

$$Y_t = 0.9959 Y_{t-1} + -0.5836 Y_{t-2} + \varepsilon_t$$



Useful time series models — purely random processes

- Several stochastic processes are useful for modelling time series
- A **purely random process** consists of mutually independent random variables, distributed normal with zero mean and variance σ^2
- Such a process has constant mean and variance
- Often termed **white noise**
- Different values are uncorrelated; $\rho(k) = 1$ if $k = 0$, otherwise $\rho(k) = 0$
- As the mean and autocovariance function do not depend on time, the process is second order stationary
- Independence implies the series is strictly stationary



Useful time series models — random walks

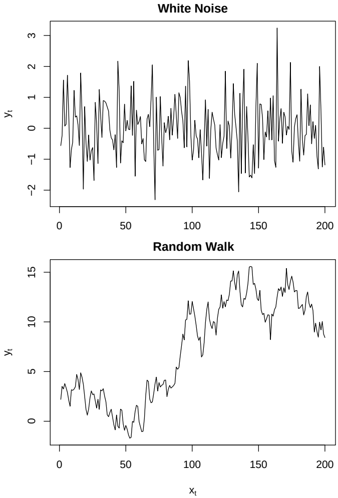
- Several stochastic processes are useful for modelling time series
- Suppose $\{z_t\}$ is a purely random process with mean μ and variance σ_z^2
- A time series is said to be a **random walk** if

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

- If started at 0 when $t = 0$ then $x_1 = z_1$ and

$$x_t = \sum_{i=1}^{t} z_i$$

- x_t is the cumulative sum of the random process up to time t
- The first differences of a random walk yield a purely random process

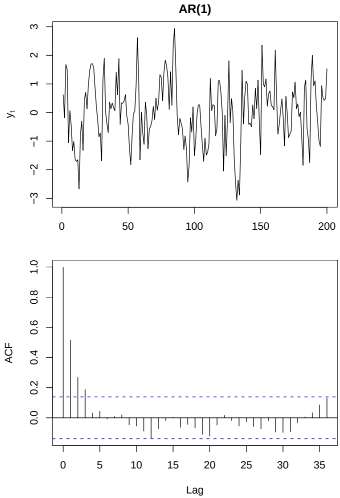


Useful time series models — autoregressive processes

- Several stochastic processes are useful for modelling time series
- A series x_t is said to be an **autoregressive** process if

$$x_t = \alpha_0 x_{t-1} + \cdots + \alpha_p x_{t-p} + z_t$$

- x_t is a function of past observations plus a purely random process (z_t)
- An **AR(p)** is a function of the p previous observations — said to be of order p
- AR(1) is the simplest such function, where $x_t = \alpha_1 x_{t-1} + z_t$
- An AR(1) is also called a **Markov process**
- Can have a non-zero mean and then the AR(p) contains an intercept (mean) term

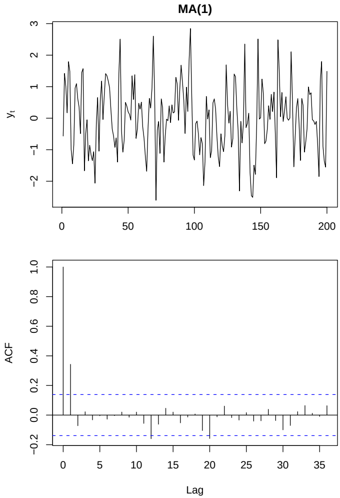


Useful time series models — moving average processes

- Several stochastic processes are useful for modelling time series
- A series x_t is said to be a **moving average** process if

$$x_t = \beta_0 z_t + \beta_1 z_{t-1} + \cdots + \beta_q z_{t-q}$$

- x_t is modelled as a function of the current and past values of a purely random process
- An **MA(q)** is of order q
- MA(1) is the simplest such function, where $x_t = \beta_0 z_t + \beta_1 z_{t-1}$
- Can have a non-zero mean and then the MA(q) contains an intercept (mean) term
- The ACF of a MA(q) has a sharp cut-off

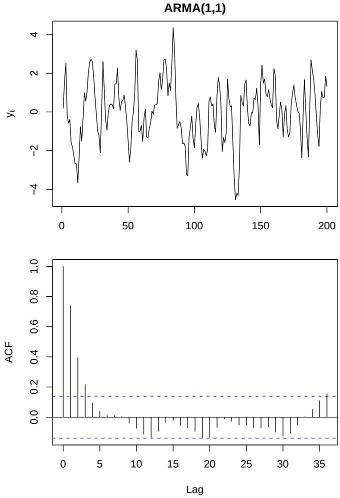


Useful time series models — ARMA models

- An **autoregressive moving average** process combines both AR(p) and MA(q) terms into a general model for time series

$$x_t = \sum_{l=1}^{p} \alpha_l x_{t-l} + z_t + \sum_{j=1}^{q} \beta_j z_{t-j}$$

- In shorthand we have ARMA(p,q)
  - p refers to the order of the AR process
  - q refers to the order of the MA process

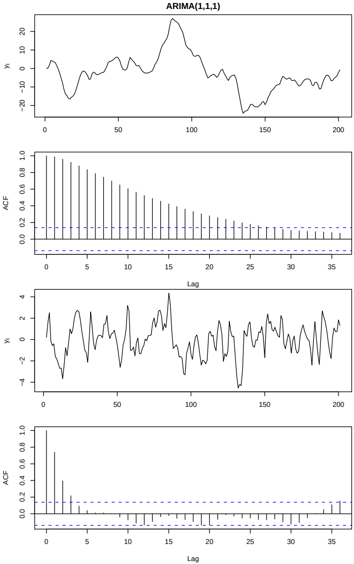


Useful time series models — ARIMA models

- An **autoregressive integrated moving average** process combines both AR(p) and MA(q) terms, *and* differencing into a general model for time series

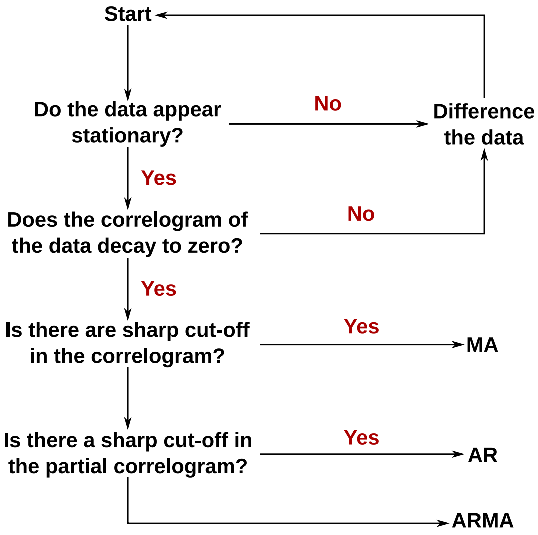
$$\nabla^d x_t = \sum_{l=1}^{p} \alpha_l \nabla^d x_{t-l} + z_t + \sum_{j=1}^{q} \beta_j z_{t-j}$$

- In shorthand we have ARIMA(p,d,q)
  - p refers to the order of the AR
  - q refers to the order of the MA
  - d refers to the order of the differencing applied to the original x_t
- First-order differencing is usually sufficient so generally $d = 1$
- A random walk can be regarded as an ARIMA(0,1,0)



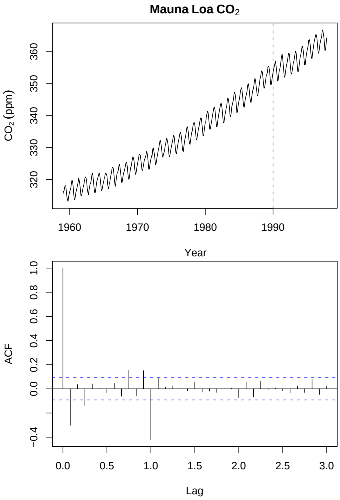
Choosing between ARMA models

- First step; are the data stationary?
- If not difference them
- Then compute the correlogram (ACF)
- If sharp cut-off then MA
- If not, compute partial-ACF
- If sharp cut-off then AR
- If not, ARMA
- If differenced, then use ARIMA choosing AR, MA or ARMA selected above
- If seasonal data, use SARIMA



Example — Mauna Loa CO_2 concentrations

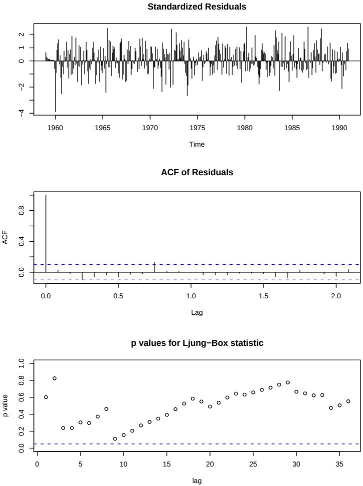
- CO_2 concentrations (ppm) measured at Mauna Loa 1959–1997
- Develop an appropriate SARIMA model for these data
- Fit model for 1957–1990
- Features:
  - Trend (differencing)
  - Seasonal component (seasonal differencing require)
  - Sharp cut-off in ACF suggests MA
  - ∇^{12} not removed all seasonal component; SAR or SMA
- Fit several models (36) and select using BIC
- p(0-2), d(1), q(0-2), P(0-1), D(1), Q(0-1), $s = 12$



Example — Mauna Loa CO_2 concentrations

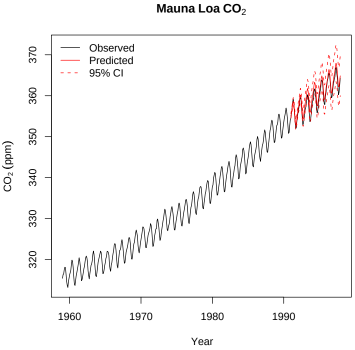
- Optimal model has BIC $= 151.46$
  - $p(0)$, $d(1)$, $q(1)$, $P(0)$, $D(1)$, $Q(1)$, $s = 12$
- Diagnostics suggest no major problems with residuals

```
Call:
arima(x = CO2, order = c(0, 1, 1),
      seasonal = c(0, 1, 1))

Coefficients:
          ma1     sma1
      -0.3605  -0.8609
s.e.   0.0545   0.0313

sigma^2 estimated as 0.08031:
log likelihood = -66.8,  aic = 139.61
```



Example — Mauna Loa CO_2 concentrations

- Using the fitted SARIMA model, predict for the years 1991-1997
- Predicted values are in general agreement with the observed trend and seasonality
- Model over predicts for the "unobserved" period slightly
- The observed data, however lie within the 95% confidence interval of the predictions



Outline

Regression models for time series

- The ARIMA family of models allows us to model properties of a single time series
- They help us to understand the stochastic processes that might underlie the observations
- They don't help explain which factors may be driving the observed time series
- If we have additional time series on explanatory variables we can use these to try to explain the response time series
- Can extend ARIMA model to include exogenous variables ARIMAX...
- ... but regression provides a more familiar and powerful way of modelling time series and the effects of predictor variables
- ARIMA-type models assume equally-spaced observations; can have missing data, and hence an irregular sequence
- This means they are of limited use for lots of ecological and palaeoecological data

Regression models for time series

- Ordinary least squares regression makes assumptions about the model residuals — i.i.d.
  - Residuals are **independent** and **identically** distributed
  - Normally distributed, with mean 0 and known variance σ^2
- GLMs and GAMs allow us to relax the distributional assumptions to take account of Poisson or binomial data, etc.
- To model time series with regression techniques we need to account for the lack of independence of the observations in some way
- We can extend the linear regression case through the use of **generalised least squares** GLS
- Going further, we can extend GLS to use smoothers and model non-normal responses using **generalised additive mixed models** GAMMs
- This is achieved, primarily, by relaxing the assumptions about the variance, σ^2

Assumptions of least squares regression

1. The linear model correctly describes the functional relationship between y and X
   - If violated the estimate of predictor variances (σ^2) will be inflated
   - Incorrect model specification can show itself as patterns in the residuals
2. x_i are measured without error
   - Allows us to isolate the error component as random variation in y
   - Estimates $\hat{\beta}$ will be biased if there is error in X — often ignored!
3. For any given value of x_i, the sampled y_i values are independent with normally distributed errors
   - Independence and normality of errors allows us to use parametric theory for confidence intervals and hypothesis tests on the F-ratio.
4. Variances are constant along the regression line/model
   - Allows a single constant variance σ^2 for the variance of the regression line/model
   - Non-constant variances can be recognised through plots of residuals (amongst others) — i.e. residuals get wider as the values of y increase.

Generalised Least Squares

- The familiar least squares regression model can be written

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_i x_i \qquad \varepsilon \sim N(0, \sigma^2 \mathbf{\Lambda})$$

- $\mathbf{\Lambda} \equiv \mathbf{I}$
- $\mathbf{I}$ is the identity matrix
- When multiplied by σ^2 we get the following covariance matrix

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \qquad \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

- σ^2 same for all observations (variance), and all are independent (0 covariance)

Generalised Least Squares

- In least squares, the β are estimated by

$$\hat{\beta} = (X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}} y$$

- If we now allow for correlated errors and set $\sigma^2 \mathbf{I}$ from the previous slide to be $\Sigma_{\varepsilon\varepsilon}$, the coefficients in GLS are estimated by

$$\hat{\beta} = (X^{\mathsf{T}} \Sigma_{\varepsilon\varepsilon}^{-1} X)^{-1} X^{\mathsf{T}} \Sigma_{\varepsilon\varepsilon}^{-1} y$$

- We now need to choose a simple enough form for $\Sigma_{\varepsilon\varepsilon}$ so that we can estimate it and all the other parameters in the model from the data in a parsimonious manner
- As $\Sigma_{\varepsilon\varepsilon}$ is not known, estimation of model is done by ML
- It is worth noting that if the diagonal of $\Sigma_{\varepsilon\varepsilon}$ contains different values it indicates different variances for the residuals

Generalised Least Squares — correlated errors

- We can assume that the covariance of two errors depends only on their separation in time
- In which case we can use the autocorrelation function and define the autocorrelation at lag s as ρ_s, the correlation between two errors that are separated by $|s|$ time periods
- This results in an error covariance matrix with the following pattern

$$\boldsymbol{\Sigma}_{\varepsilon\varepsilon} = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{pmatrix} = \sigma^2 \mathbf{P}$$

- This allows quite a flexible correlation structure, but comes at the costs of estimating n distinct parameters (σ^2 and $\rho_1, \cdots, \rho_{n-1}$)
- Too many!

Generalised Least Squares — correlated errors

- To simplify the model further, we restrict the order of the autocorrelations to a small number of lags
- Usually the first-order AR process is used: $\varepsilon_s = \rho\varepsilon_{s-1} + \eta_s$
- The correlation between two errors at times t and s is

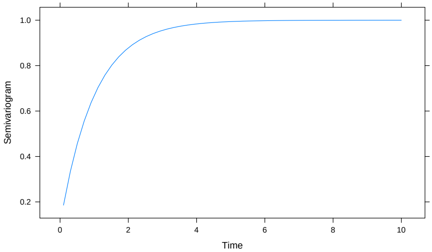
$$\mathrm{cor}(\varepsilon_s \varepsilon_t) = \begin{cases} 1 & \text{if} s = t \\ \rho^{|t-s|} & \text{else} \end{cases}$$

- This results in an error covariance matrix with the following pattern

$$\mathbf{\Sigma}_{\varepsilon\varepsilon} = \sigma^2_\varepsilon \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix} = \sigma^2 \mathbf{P}$$

Smoothing and correlated errors

- Can also use MA or ARMA processes for the model errors
- This is fine for equally-spaced observations, for irregularly spaced observations we need to turn to spatial correlation structures
- Use a 1-D spatial correlation structure to model correlations between two errors that are arbitrary distances apart in time
- Several structures, simplest is Exponential spatial correlation structure
- In 1-D, this is also known as the Continuous-time AR(1) (CAR(1))



Additive models

- Additive models are a generalization of linear models that replace the sum of regression coefficients $\times$ covariates by a sum of smooth functions of the covariates X
- Such a model has the following form

$$y = \beta_0 + \sum_{p=1}^{j} f_j(X_j) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 \mathbf{\Lambda})$$

- where the f_j are arbitrary smooth functions
- Additive models are more flexible than the linear model but remain interpretable as the f_j can be plotted to show the marginal relationship between the predictor and the response

Using statistical models on times series data — trends

- Approach follows closely that of Ferguson et al (2006, 2007)
- A linear model for a trend component in the data might be

$$y = \beta_0 + \beta_1 time + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 \mathbf{\Lambda})$$

- An additive model for a trend component in the data might be

$$y = \beta_0 + f_1(time) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 \mathbf{\Lambda})$$

- We can compare these two models to select between a linear or smooth (non-linear) trend
- We can also test for the presence of a trend by comparing this model to a null model

$$y = \beta_0 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 \mathbf{\Lambda})$$

- Model testing is does via **likelihood ratio test** (LRT) and information statistics

Using statistical models on times series data — trends

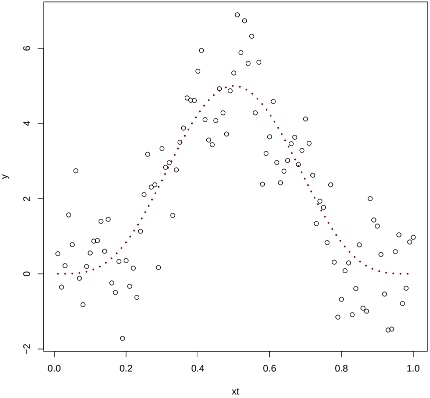
- Additionally, we can include additional predictor variables into the linear predictor to model changes in the response time series as a function of the explanatory variables
- We can also test whether the autocorrelation structure is required using LRT
- As these models are just regression models, use existing, well-developed theory for fitting models
- Using smoothers allows very flexible models to be fitted that include non-linear trends, seasonal smoothers etc.
- However, fitting these models in software is demanding and not for the faint hearted

Smoothing and correlated errors

- Data generated from the model

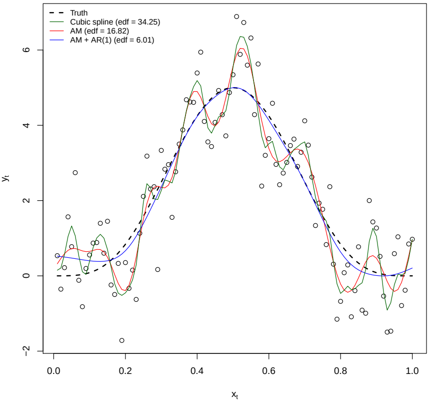
$$y_t = (1280 \times x_t^4) \times (1 - x_t)^4$$

- Added AR(1) noise; $\alpha = 0.3713$
- Task is to retrieve the trend y_t from the noisy data
- The data exhibit a non-linear trend and we can use a smoother to model this feature of the data
- A problem in smoothing is selecting the bandwidth or complexity of fitted spline
- The usual methods for smoothness selection assume the observations are independent



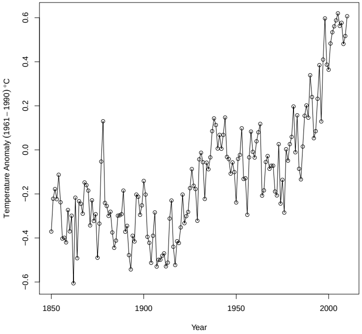
Smoothing and correlated errors

- Fit three separate models to the data
  1. Cubic smoothing spline (GCV)
  2. Additive model (GCV)
  3. Additive model with AR(1) errors (LMM)
- The two models that assume independent errors over fit the data
- Identify structure that is sampling artefact of the data at hand
- Additive model with AR(1) errors fits actual trend will
- $\hat{\phi} = 0.403$ (0.169, 0.594 95% CI)



NH Global Temperature

- Much interest in the patterns shown in temperature records, esp. in the most recent period
- Classic diagram used in the most recent IPCC Assessment Reports demonstrating trends and rates of change in temperature
- Mann and colleagues filtered the timeseries so had to pad the series at the ends to allow estimates of trends and rates in most recent period
- Padding the series done in several ways, but all involved inventing data for most recent period
- Can we do better with a regression model? — AM

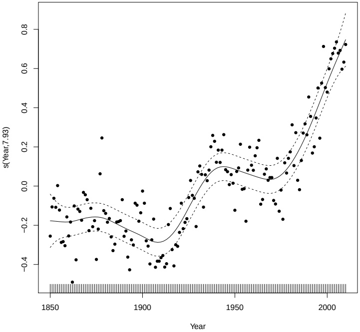


NH Global Temperature — additive model

- Annual mean global NH temperature anomaly (1960-1990)
- Fitted additive model of form

$$y = \beta_0 + f_1(year) + \varepsilon,\ \varepsilon \sim N(0, \sigma^2 \mathbf{\Lambda})$$

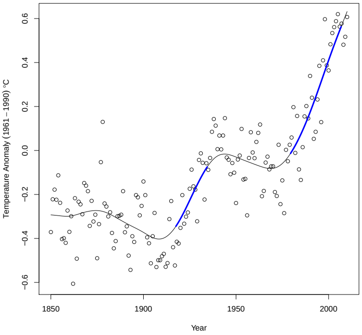
- Estimated the model with $\mathbf{\Lambda} \equiv \mathbf{I}$



- $\mathbf{\Lambda}$ assumed to be simple $\mathrm{AR(p)}$ or $\mathrm{MA(q)}$ or combinations for p and $q \in (1, 2, 3)$
- Fitted ARMAs to the residuals of this model to identify best model for residuals — AR(1)
- Refitted AM with an AR(1) correlation structure
- Checked for residual correlation

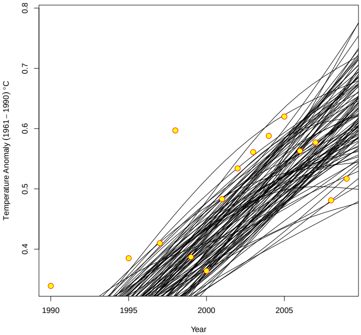
NH Global Temperature — additive model

- Interested in rates of change in temperature
- Also in the trend in the most recent period
- Can estimate the first derivatives of the fitted trend to show periods where the first derivative is statistically different from 0
- Use the method of finite differences
- Colour the fitted trend according to periods of significant change
  - Red — significantly decreasing
  - Blue — significantly increasing
  - (*c.f.* Sizer)



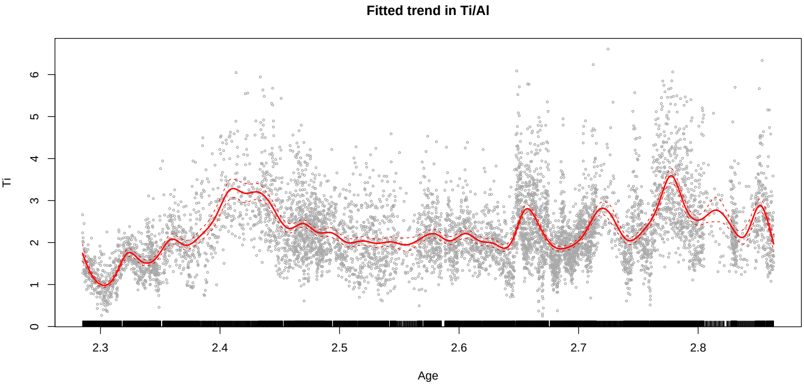
NH Global Temperature — additive model

- Uncertainty in fitted trend?
- Recall that the smoother is a spline for which we estimate coefficients $\hat{\beta}^s$
- Each $\hat{\beta}^s$ estimated with uncertainty
- The set of $\hat{\beta}^s$ form a MVN distribution
- Simulate new values for $\hat{\beta}^s$ from the MVN to generate trends consistent with the fitted model
- Sampling from the posterior distribution of the model parameters
- Only a tiny proportion of 1000 samples from the posterior distribution suggest that, given these data, there is little support for claims that the planet is cooling



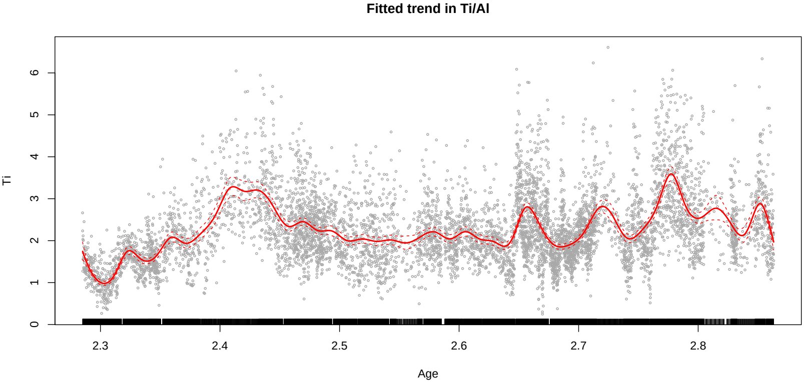
Katy's crazy amounts of data

- Lots, and lots, and lots of XFR data
- Model takes a week to fit
- How can we deal with data like this?



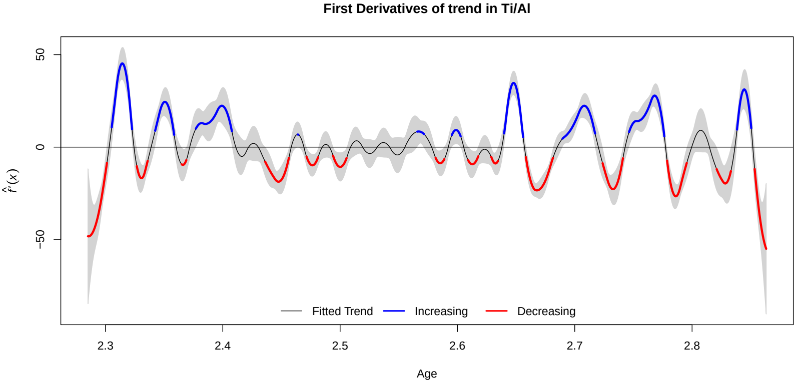
Katy's crazy amounts of data

- Realise that we might have to be subjective here
- Fix the degree of smoothness — or force use of fewer DF, and
- Use and adaptive smoother



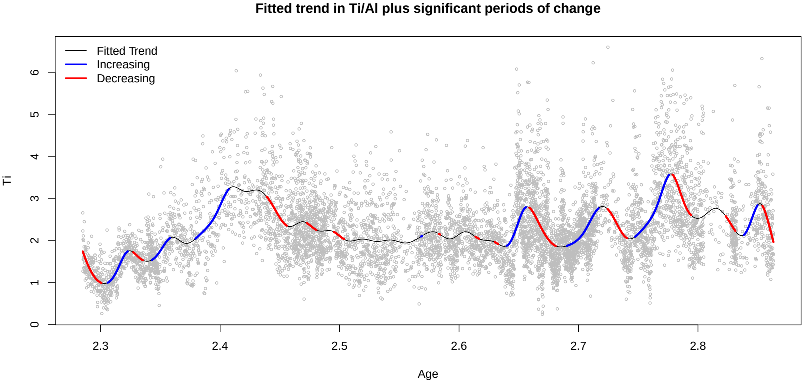
Katy's crazy amounts of data

- Realise that we might have to be subjective here
- Fix the degree of smoothness — or force use of fewer DF, and
- Use and adaptive smoother



Katy's crazy amounts of data

- Realise that we might have to be subjective here
- Fix the degree of smoothness — or force use of fewer DF, and
- Use and adaptive smoother



Protocols for fitting GLS and (G)AMM

- Model selection now involves finding the correct fixed effects formulation *and* the correct specification for the model errors
- A protocol for model selection could take the following form
  1. Fit the fixed effects model you think is plausible without the autocorrelation — don't worry too much at this stage about getting a minimal, adequate model for the fixed effects
  2. Now add the autocorrelation structure to the model and refine that — LRT to see if the correlation is required
  3. Finally, revisit the fixed effects and remove variables not required
- When fitting the correlation structure, don't worry about getting this part exactly correct
- The aim is to add a structure that plausibly accounts for the autocorrelation, not to model it exactly
- In general, this means AR(1) for equally-spaced observations and CAR(1) for unequally-spaced observations

Outline

Spectral Analysis

- Spectral analysis: methods of estimating the spectral density function, or spectrum, of a given time series
- Spectral analysis can be used to detect periodic signals corrupted by noise
- Periodic signals; a repeating pattern in a series is periodic, with **period** equal to the length of the pattern
- The sine wave is the fundamental periodic signal in mathematics
- Joseph Fourier (1768–1830) showed that good approximations to most periodic signals can be achieved using sums of sine waves
- Spectral analysis is based on sine waves and a decomposition of variation in series into waves of various frequencies

Sine waves

- A sine wave of frequency ω, amplitude A, and phase ψ for time t is

$$A \sin(\omega t + \psi)$$

- A general sine wave can be expressed as a weighted sum of sine and cosine functions

$$A \sin(\omega t + \psi) = A \cos(\psi) \sin(\omega t) + A \sin(\psi) \cos(\omega t)$$

- A sampled sine wave of any amplitude and phase can be fitted by a linear regression model with the sine and cosine functions as predictor variables

Sine waves

- Suppose we have a time series of length n, $\{x_t : t = 1, \ldots, n\}$ (n is even)
- Fit time series regression with x_t as response and $n - 1$ predictor variables

$$\cos\left(\frac{2\pi t}{n}\right), \sin\left(\frac{2\pi t}{n}\right), \cos\left(\frac{4\pi t}{n}\right), \sin\left(\frac{4\pi t}{n}\right)$$

$$\cos\left(\frac{6\pi t}{n}\right), \sin\left(\frac{6\pi t}{n}\right), \ldots \cos\left(\frac{2(n/2-1)\pi t}{n}\right), \sin\left(\frac{2(n/2-1)\pi t}{n}\right)$$

$$\cos(\pi t)$$

Sine waves

- Estimated coefficients denoted by $a_1, b_1, a_2, b_2, \ldots, a_{n/2-1}, b_{n/2-1}, a_{n/2}$
- As many coefficients as observations
- No degrees of freedom for errors
- a_0 is the intercept, and is the mean of x
- Lowest frequency is one cycle (or 2π radians) per record length, $2\pi/n$ radians per sampling interval (RPSI)
- General frequency, m cycles per sampling interval ($2\pi m/n$ RPSI), m is integer between 1 and $n/2$
- highest frequency, 0.5 cycles per sampling interval (π RPSI), $n/2$ cycles in the series
- Sine wave that makes m cycles in series length is the mth harmonic
- Amplitude of mth harmonic is $A_m = \sqrt{a_m^2 + b_m^2}$

Raw Periodogram

- Amplitude of mth harmonic is $A_m = \sqrt{a_m^2 + b_m^2}$
- Parseval's Theorem expresses variance of a series as a sum of $n/2$ components at integer frequencies $1, \dots, n/2$

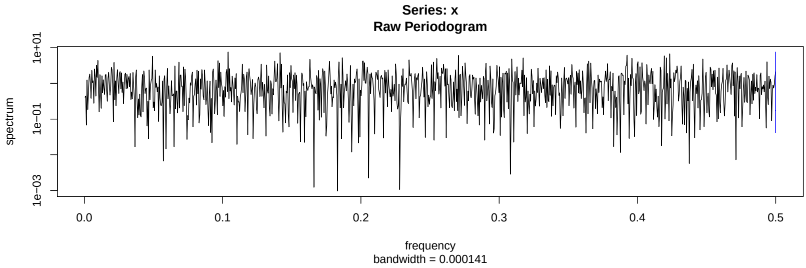
$$\mathrm{Var}(x) = \frac{1}{2} \sum_{m=1}^{(n/2)-1} A_m^2 + A_{n/2}^2$$

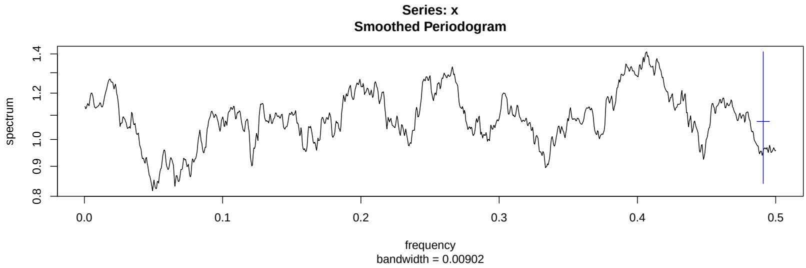
- In general, instead of via a regression, the calculations above are usually performed with the fast fourier transform (FFT) algorithm
- A plot of A_m^2 as spikes against m is a Fourier line spectrum
- Raw periodogram is produced by joining the tips of the spikes in the Fourier line spectrum and scaling area equal to the variance

Smoothed Periodogram

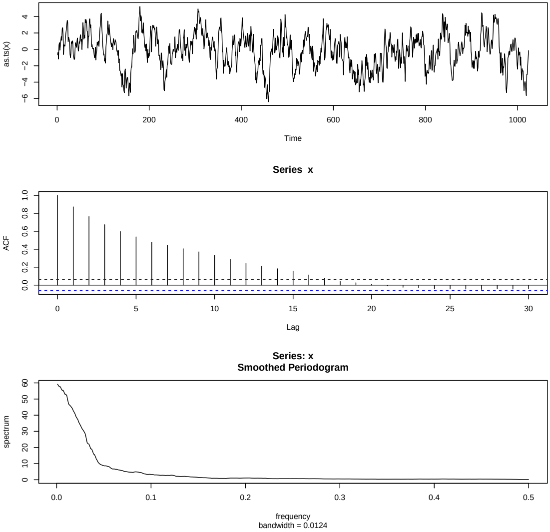
- Periodogram distributes variance over frequency but has two drawbacks
  - Precise set of frequencies is arbitrary, depends on series length
  - Periodogram does not get smoother as series length increases — just gets more packed
- The remedy to this is to smooth the periodogram
- Smooth the spikes of the Fourier line spectrum using moving averages before joining the tips
- Smoothed periodogram also known as the (sample) spectrum
- Smoothing will reduce the heights of the peaks and excessive smoothing will blur the features we are interested in
- In practice try several degrees of smoothing

Example Periodograms — white noise

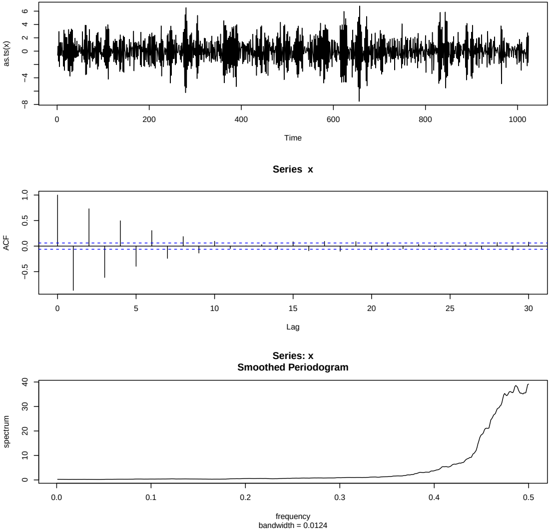




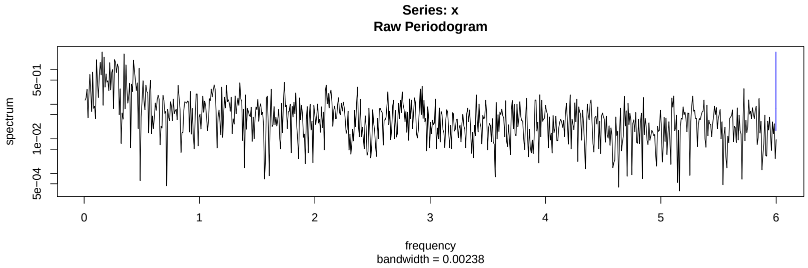
Example Periodograms — AR(1) v1

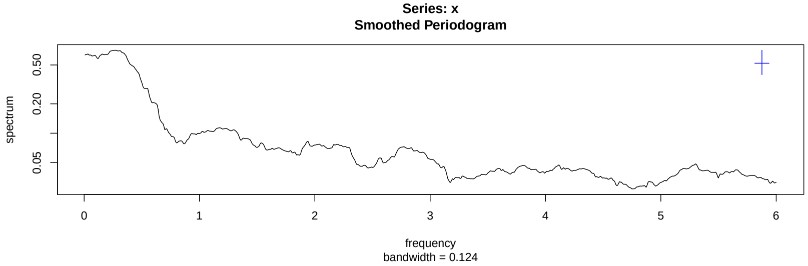


Example Periodograms — AR(1) v2



Southern Oscillation Index

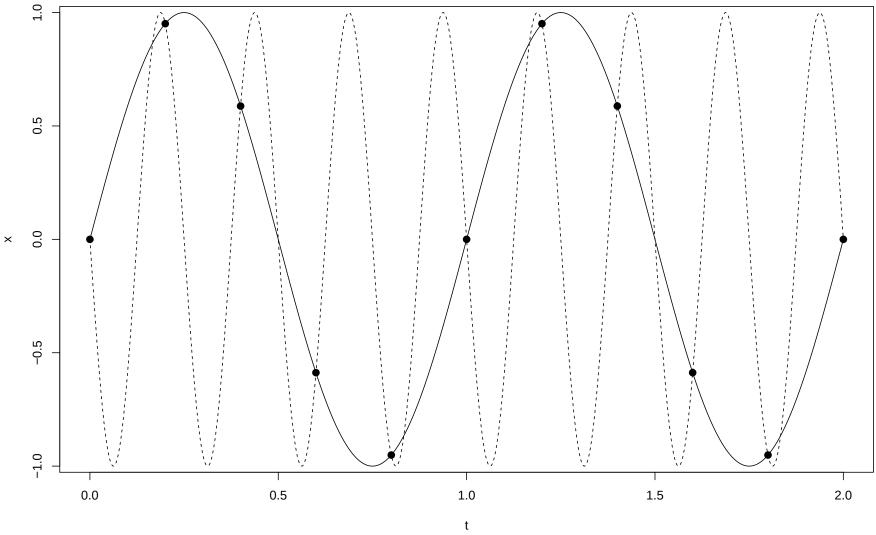




Aliasing and the Nyquist frequency

- Many time series are discrete measurements of a continuous process
- Important to sample at high enough frequency to capture highest frequency oscillations in process
- If sampling frequency is too low, miss information
- Also, real high frequency variation will show up as lower frequency variation
- This is known as aliasing
- The Nyquist frequency is half the sampling frequency and is the maximum frequency we can recover from the data series collected

Aliasing



Further reading

- Andersen et al (2008) Ecological thresholds and regime shifts: approaches to identification. Trends in Ecology and Evolution **24**(1), 49–57.
- Chandler & Scott (2011) Statistical Methods for Trend Detection and Analysis in the Environmental Sciences. Wiley-Blackwell.
- Chatfield (2004) *The Analysis of Time Series: An Introduction*. 6th Edition. Chapman & Hall/CRC.
- Cowpertwait & Metcalfe (2009) *Introductory Time Series with R*. Springer.
- Diggle (1990) *Time Series; A Biostatistical Introduction*. Oxford University Press.
- Ferguson et al (2007) Assessing ecological responses to environmental change using statistical models. *Journal of Applied Ecology*, **45(1)**, 193–203.
- Fox (2008) *Applied Regression Analysis and Generalized Linear Models*. Sage. (Chapter 16)
- Wood (2006) *Generalized Additive Model: An Introduction with R*. Chapman & Hall/CRC.
- Zuur et al (2007) *Analysing Ecological Data*. Springer.
- Zuur et al (2009) *Mixed Effects Models and Extensions in Ecology with R*. Springer.