Introduction to time series analysis

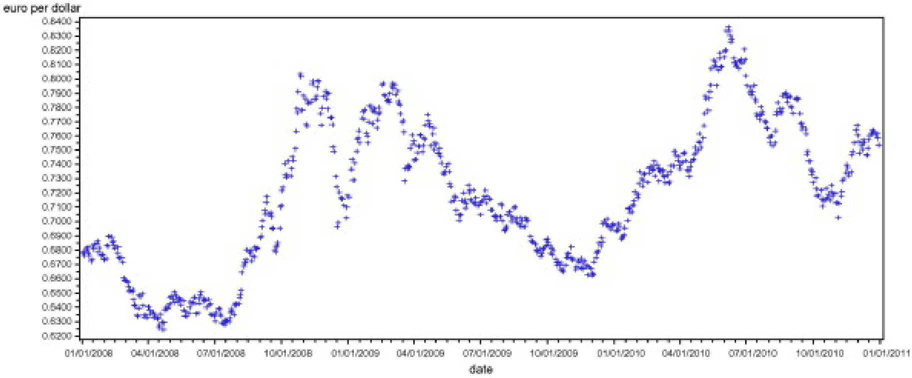
Statistical modelling: Theory and practice

Gilles Guillot

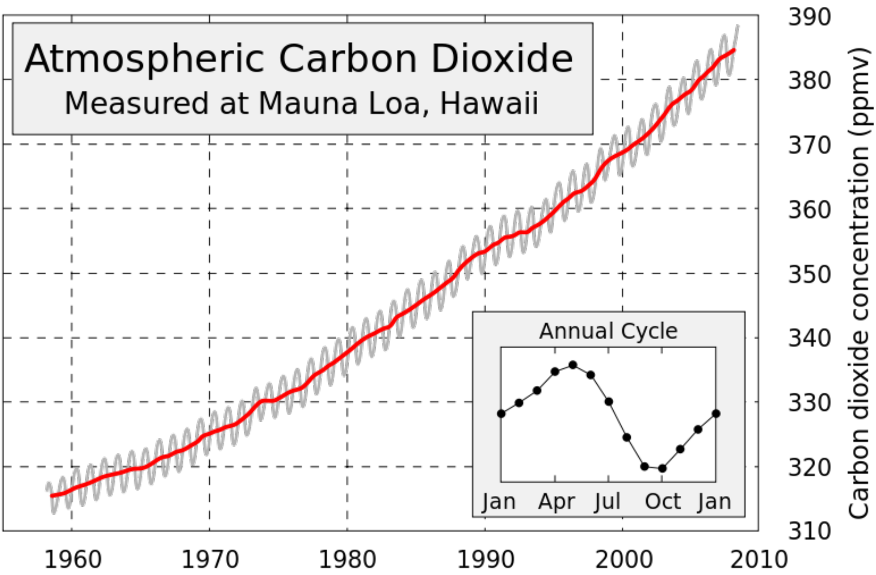
gigu@imm.dtu.dk

November 17, 2014

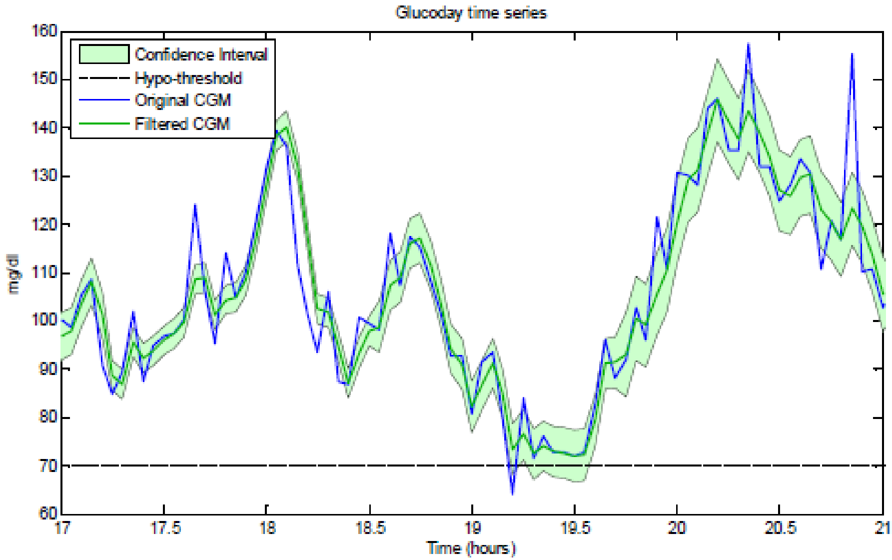
Finance data



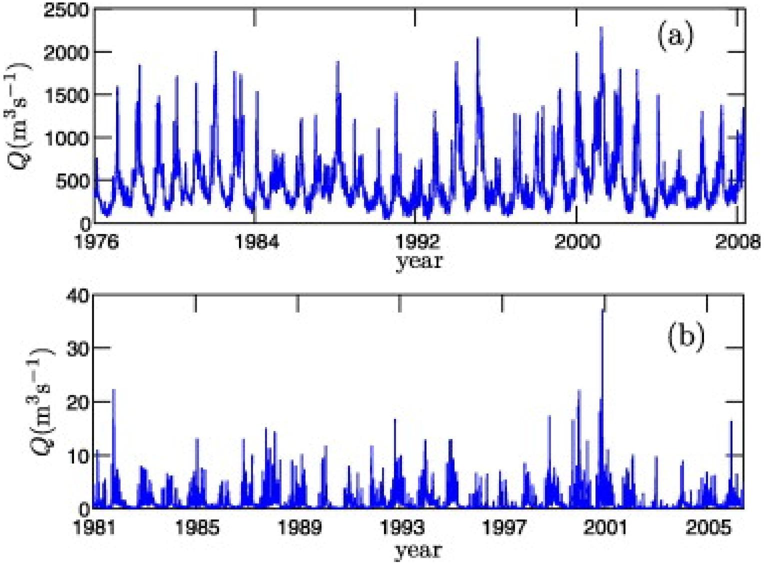
Geophysical data



Biomedical data



Civil engineering data



Different scientific/technical fields but common aspects in previous examples I

Common questions:

- A central question: prediction of future values
- Other questions:
  - is the phenomenon periodic? what is the period?
  - is there any trend ? can it be estimated (filtering problem)
  - is there a time structure?

Common data features:

- Observations come in a specific order (time! here obs. index not just a label!)
- There is a pattern in the variation of variables across time
- There is erratic variation away from the main "pattern"

Different scientific/technical fields but common aspects in previous examples II

Goal of the lecture: give a flavour of statistical tools to model variation across time.
In lectures on linear model, we had $y = \beta x + \varepsilon$. The focus was on the "pattern", the rest was i.i.d noise. In this lecture, the "residual variation" is as important as the main "trend".

Some definitions I

- **Time series:** a set of observations of a variable at different time points $(y_t)_{t=1,...,T} = (y_1, ..., y_T)$
- **Stochastic process:** a set of random variables observed at different time points $(Y_t)_{t=1,...,T} = (Y_1, ..., Y_T)$. If the random variables are mutually independent, the process is just a noise seen before.
- **Realization of a stochastic process:** a single set of observations $(y_t)_{t=1,...,T} = (y_1, ..., y_T)$.

Some definitions II

- **Continuous-time stochastic process:** model for a time series which is (or could be in principle) observed continuously in time.
Example:temperature in Lyngby
- **Discrete-time stochastic process:** model for a time series which is by nature defined only at some discrete time points. Example: exchange rate euro/dollar, GDP of a country, max. daily temperature
- **Sampling frequency:** number of observed data point per time unit
- **Support:** the time period over which the variable is averaged. Do not mix up average daily tempeartures with "instant" temperature measured at noon every day. Both are observed at the same frequency but not over the same time support.

First and second order characterisation of a stochastic process I

Definition: Mean function

For (Y_t) a stochastic process, the mean function is defined as

$$\mu(t) = E[Y_t]$$

First and second order characterisation of a stochastic process II

NB: If n realizations $y_t^{(k)}$ of (Y_t) have been observed (e.g. variation of glucose across time for n patients), then $\mu(t)$ can be estimated as $1/n \sum_{k=1}^{n} y_t^{(k)}$.
There is often only one realization available and the estimator above does not make sense. In this case the mean function is somehow arbitrary. Think of the mean exchange rate Dollar/Euro in November 20 2012?

Definition: Variance function

For (Y_t) a stochastic process, the variance function is defined as

$$\sigma^2(t) = Var[Y_t]$$

With several realizations at hand, one could estimate $\sigma^2(t)$ by $\widehat{\sigma}^2(t) = 1/(n-1)\sum_{k=1}^{n}(y_t^{(k)} - \bar{y}_t)^2$

Covariance function of a stochastic process I

Definition: Covariance function

For a stochastic process (Y_t), the covariance function is defined as

$$C(t, t') = Cov[Y_t, Y_t']$$

NB: $C(t, t) = Cov[Y_t, Y_t] = \sigma^2(t)$

Covariance function of a stochastic process II

Definition: Correlation function

For a stochastic process (Y_t), the correlation function is defined as

$$\rho(t,t') = Cor[Y_t, Y_t']$$

NB: $\rho(t,t) = Cor[Y_t, Y_t] = 1$

Stationarity I

Definition: First-order stationarity

A stochastic process (Y_t) is first-order stationary if its mean $\mu(t)$ does not depend on time.

Stationarity II

Definition: Second-order stationarity

A stochastic process (Y_t) is second-order stationary if the covariance between two observations depends only on the time lag between them.
Formally, there is a function K such that $C(t,t') = K(t-t')$.
In thise case, with a slight abuse, we often writes:
$C(t,t') = C(t-t') = C(k)$.

Gaussian stochastic processes I

Definition: Gaussian process

A stochastic process (Y_t) is Gaussian if for any susbet of time points $t_1, ..., t_n$ and any family of weights $\lambda_1, ..., \lambda_n$ the random variable $\sum_i \lambda_i Y_{t_i}$ is normally distributed.

Gaussian stochastic processes II

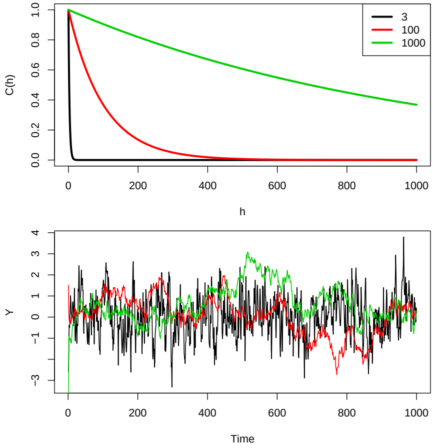
Theorem: Density of a non-degenerated Gaussian process

If the covariance matrix Σ of a Gaussian process (Y_t) observed at time points $t_1, ..., t_n$ is non-singular, $(Y_{t_1}, ..., Y_{t_n})$ admits a density of the form:

$$f(\boldsymbol{y}) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^t\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right]$$

Example of Gaussian processes

We consider below a second-order stationary Gaussian process with mean 0 and covariance function $C(t, t') = C(t - t') = C(k) = \exp(-|k|/\alpha)$. With three different values for the scale parameter $\alpha = 3, 100, 1000$



Realizations of three Gaussian stochastic processes with exponential covariances. Black curve: "short range" dependence ($\alpha = 3$). It is very erratic and the curve visits a broad range of values within a short period of time.

Estimating the covariance function

Assuming that Y_t is second-order stationary with covariance function $C(k)$ we can estimate the (constant) mean by

$$\hat{\mu} = \bar{Y} = \sum_{t=1}^{T} Y_t / T$$

and $C(k)$ by

$$\hat{C}(k) = \frac{1}{N_{t,t'}} \sum_{|t-t'|=k} (Y_t - \bar{Y})(Y_{t'} - \bar{Y})$$

where $N_{t,t'} = \#\{|t - t'| = k\}$

Estimating the covariance and correlation function with R

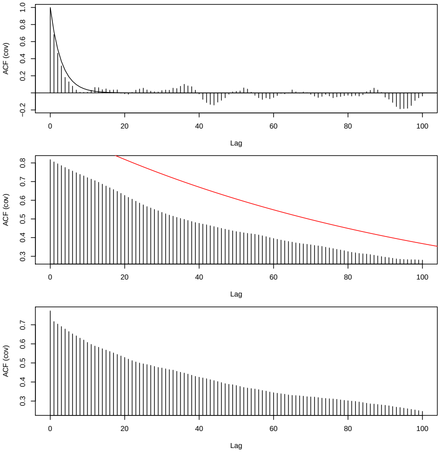
The R function `acf`:
`acf(y,type,lag.max,plot=TRUE)`
with `type='covariance'` or `type='correlation'`

Plot the empirical covariance or correlation.
The R object returned is of class `acf` and contains covariance or correlation values at different time lags.

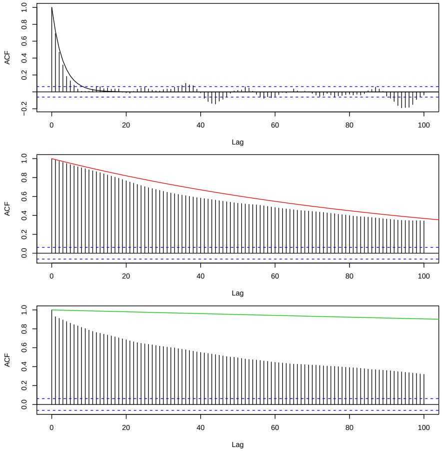
R code used to estimate the covariance function

```
par(mfrow=c(3,1),mar=c(4,4,1,2))
acf(y1,lag.max=100,type='cov')
lines(times,exp(-times/alpha1))

acf(y2,lag.max=100,type='cov')
lines(times,exp(-times/alpha2),col=2)

acf(y3,lag.max=100,type='cov')
lines(times,exp(-times/alpha3),col=3)
```

Estimated covariances



Estimated correlations



Estimating the covariance function, what for?

- Purely descriptive purpose: how far is the time series from a pure noise? At what lag does decorrelation occur?
- If $Z(t) = X(t) + Y(t)$ where Y and Z are unobserved processes with known covariance function, one can predict X_t from observations of Z_t at time $1, ..., T$ (filtering).
- If Y is observed at times $1, ..., T$ one can predict Y_{T+k} for any k.

The latter problems require some knowledge about the covariance structure (linear prediction methods).

Concluding remarks on second-order characteristics

- The covariance and correlation functions provide a <u>global</u> description of a stochastic process.
- They are powerfull but can become numerically demanding for large datasets
- The next slides provide an alternative modelling point of view based on a description of the <u>local</u> time structure.

First order auto-regressive models I

Intuitively,

- tomorrow's Euro/Dollar exchange rate is today's rate + a small pertubation.
- tomorrow's temperature is today's temperature + a small pertubation

Definition: First order auto-regressive process

A first order auto-regressive process (AR(1) for short) is a discrete-time process such that

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t$$

where c and ϕ are constant parameters and (ε_t) is an zero-mean i.i.d random noise.

Assuming that $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ makes the process above a Gaussian AR1.

Second order properties of AR(1) processes I

Under which conditions is an AR(1) process stationary?

- **Mean:** $Y_t = c + \phi Y_{t-1} + \varepsilon_t \Longrightarrow E(Y_t) = c + \phi E(Y_{t-1}) + 0$

  First-order starionarity requires that $\mu = c + \phi\mu$ hence $\mu = c/(1-\phi)$

Second order properties of AR(1) processes II

- **Variance:** $Y_t = c + \phi Y_{t-1} + \varepsilon_t \Longrightarrow V(Y_t) = \phi^2 V(Y_{t-1}) + \sigma_\varepsilon^2$

  Second-order stationarity requires that $\sigma_Y^2 = \phi^2 \sigma_Y^2 + \sigma_\varepsilon^2$.
  If $|\phi| < 1$ the above is achieved if $\sigma_Y^2 = \sigma_\varepsilon^2/(1-\phi^2)$

Conversely, for ϕ such that $|\phi| < 1$,
if $E(Y_0) = c/(1-\phi)$
and $V(Y_0) = \sigma_\varepsilon^2/(1-\phi^2)$,
then (Y_t) has stationary mean and variances.

Covariance function of an AR(1): For $k > 0$,

$$\begin{aligned} Cov(Y_t, Y_{t+k}) &= Cov(Y_t, c + \phi Y_{t+k-1} + \varepsilon_{t+k}) \\ &= \phi Cov(Y_t, Y_{t+k-1}) \end{aligned}$$

For $k = 1$, assuming stationarity of the variance, $Cov(Y_t, Y_{t+1}) = \phi \sigma_Y^2$

By induction, $Cov(Y_t, Y_{t+k}) = \phi^k \sigma_Y^2 \propto e^{k \ln \phi}$

The covariance decays at an exponential rate.

Conditional distribution of a Gaussian AR(1) process

Under the assumption that $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$,

the AR(1) structure $Y_t = c + \phi Y_{t-1} + \varepsilon_t$ implies the following conditional distribution:

$$Y_t | Y_{t-1} \sim \mathcal{N}(c + \phi Y_{t-1}, \sigma_\varepsilon^2)$$

Maximum likelihood inference for a Gaussian AR(1) process

Data: $y = (y_1, ..., y_T)$. Parameter: $\theta = (c, \phi, \sigma_\varepsilon)$.

$$\begin{aligned} L(y;\theta) &= L(y_1, ..., y_T; c, \phi, \sigma_\varepsilon) = f(y_1, ..., y_T) \\ &= f(y_1) f(y_2, ..., y_T | y_1) \\ &= f(y_1) \prod_{t=2}^{T} f(y_t | y_{t-1}) \\ &= f(y_1) \prod_{t=2}^{T} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} (y_t - c - \phi y_{t-1})^2 \right] \end{aligned}$$

The term $f(y_1)$ can be dealt with either by disregarding it or by making assumptions on this marginal distribution.
The log-likelihood can be maximized e.g. with the R function `optim`.

References

- H. Madsen, Time Series Analysis, Chapman & Hall/CRC, 280 pp., 2008.