Spatio–temporal methods in environmental epidemiology

:Lecture 7 & 8

1.

Outline

1. Temporal processes
2. Spatial processes
    - Point referenced data
    - Area data
    - Point process data
3. Spatio–temporal processes

NOTE: R code for examples will be posted with Lab 2.

Spatio–temporal process modeling & analysis

Spatio-temporal modeling

Currently very active research & applications area.

- Relationship between **deaths & atmospheric particulate concentrations** [e.g. London Fog]
- **Climate modeling** - 1000s of sites for temperature or precipitation
- Location, location, location: **house prices**
- **Used car prices**
- Strain gauges on the **space station**
- **Fires** in tall wooden buildings
- **Lightning strikes** & forest fires
- **Acid rain**

Spatio-temporal modeling: General data categories

Time - usually discrete index, $t = 1, \ldots, T$.
Spatial locations indexed by $s \in D$.

- **Point referenced data:** D = continuum or dense spatial grid; measurements made at irregular network of locations.
  E.g: ozone field
- **Lattice processes:** D = not necessarily regular grid of areal regions or specified locations D where measurements are made.
  E.g: death counts per county; centroids = lattice points
- **Point processes:** Measurements or “marks”. made at randomly selected points in continuum D
  E.g: lightning strikes

Spatio-temporal modeling: General approach

Hierarchical modeling:

- Parameter model - random in our Bayesian context
- Process model - random (e.g. ozone concentrations varying continuously over time) or deterministic (e.g. chemical transport models for ozone given by differential equations)
- Measurement model - (e.g. hourly ozone)

Spatio-temporal modeling: General approach

Hierarchical modeling: Alternate formulation with
[X] = probability distribution of X

- $[parameters] = [\theta]$
- $[process|parameters] = [Y \mid \theta]$
- $[measurement|process, parameters] = [Z \mid Y, \theta]$

Spatio-temporal modeling: General approach

Examples:

1. Measurement model: $Z_{st} = Y_{st} + \epsilon_{st}$. Here parameters describe distributions of Y and ϵ.
2. $[Z_{st} \mid Y_{st}, \theta] = Poisson(\beta Y_{st})$ where Y_{st} denotes daily maximum 8 hour moving average concentration and Z denotes hospital admissions that day for region in which s is located.

Temporal processes

Temporal processes: introduction

Why model them?

- *To undertand them*
- *To characterize/remove regional level temporal components in spatial modeling* when interest lies on the sources of spatial variation. Spatial modeling can then focus on the variation in process residuals due to spatial components e.g. hazardous waste sites.
- *To forecast them:*
    - volcanic ash e.g. recent Iceland incident
    - nuclear power radiation leak: mobile monitors; plume dispersion models; long term effects need spatial not temporal modeling
    - Ozone is a serious health hazard so municipalities commonly issue 24 hour ozone forecasts.

Example- ozone fields in US

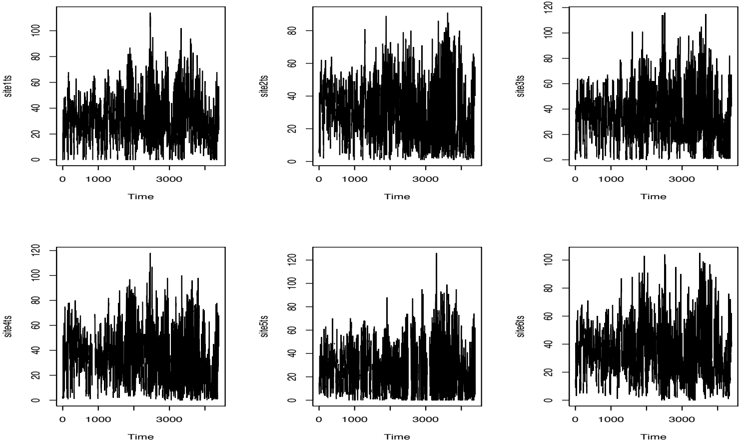
Ground level ozone:

- Colorless gas produced through photochemistry; sunlight; high temperatures; NO_x emissions from motor vehicles esp during morning, evening commute
- One of the criteria pollutants regulated by the US Clean Air Act (1970)
- Nasty stuff; reduces respiratory capacity; attacks susceptible people, young and old, people with lung diseases, asthma, emphysema, COPD
- Lead to deaths, hospital admissions, school abscences

Next slide shows 24 hour cycles for time series of hourly ozone concentrations. The one after that the seasonal effects - the latter vary from year–to–year

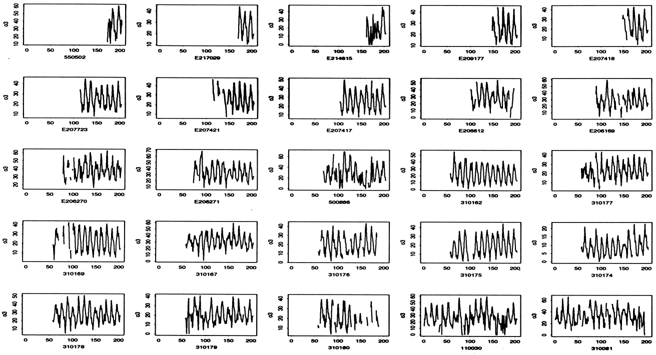
Example- ozone fields in US

Time series plots: Hourly concentrations at 6 O3 monitoring sites, Eastern USA Note **24 hour cycles.**



Example - ozone fields in BC

Time series plots: Monthly measurements at 25 O3 sites in BC. Note **seasonality** and different start dates.



Examples - lessons learned

- Temporal processes can have both regular (systematic) and irregular components.
- Regular components include
  - time trends
  - spatial trends
  - periodic components eg 24 hour cycles
  - day of the week effect
  - leap year effect
- Irregular components are randomly distributed around the regular patterns
- Measurements can be continuous as with ozone or discrete as with daily hospital admission counts for asthma attacks.
- Regular temporal component can vary across space
- Regular spatial component can vary over time.
- Measurements made at monitoring sites - they can start up at varying times, leading to the staircase structure in the data matrix.

Temporal processes: the regular components

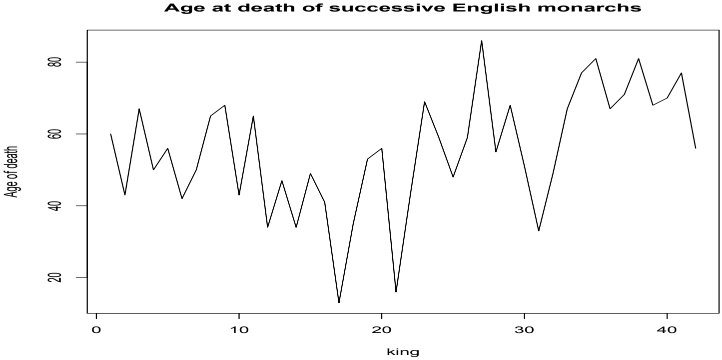
EDA

Bayesians and frequentists alike will ordinarily need to explore the measurements in a time series $\{Z_t\}$. Here we have dropped the s and assume we are working at a specific site or location.

EDA - Example: Death dates of successive English monarchs

The computational methods used here and in the sequel are described in the Lab material for this lecture. The time series in this example represents the age at death of successive English monarchs. The king's number is plotted on the horizontal axis. The age of death is on the vertical axis.

EDA - Example: Death dates of successive English monarchs



EDA - Example: Death dates of successive English monarchs

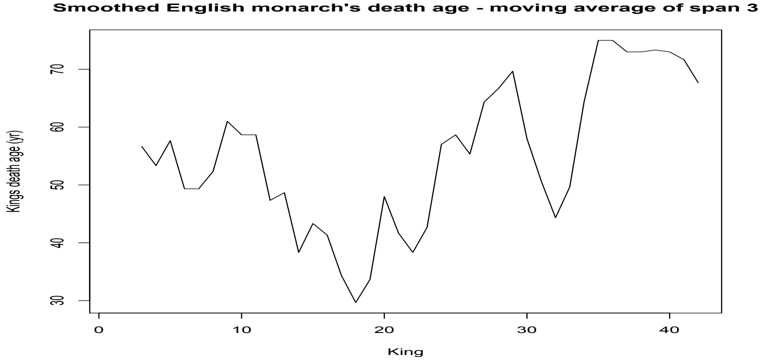
Observe the systematic decline in the age of death up to about monarch 17. Then a steady uptrend. This would not be a so-called stationary time series because of that trend component. The irregular variation around that trend does look fairly stationary, suggesting an additive measurement model with trend + irregular variation.
We next estimate the trend by smoothing the data. Lots of ways of doing that:

- moving average - statistical properties easy to assess
- loess - more complext
- Shumway's 19 day symmetrically weighted moving average

These are called lo-pass filters since they do not allow the irregular variation through.

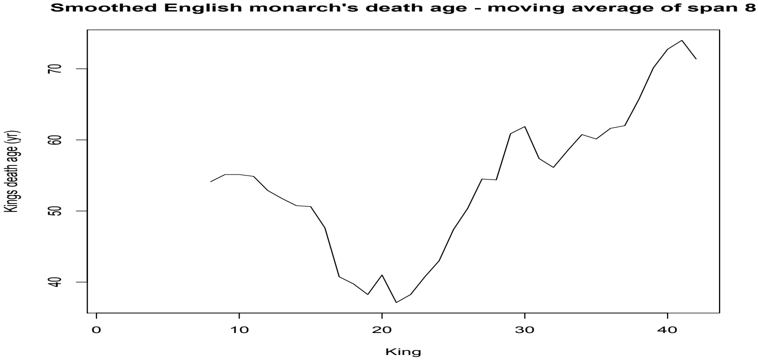
EDA - Example: Death dates of successive English monarchs

We start with a 3 Monarch moving average. The trend is clearer but more smoothing needed.



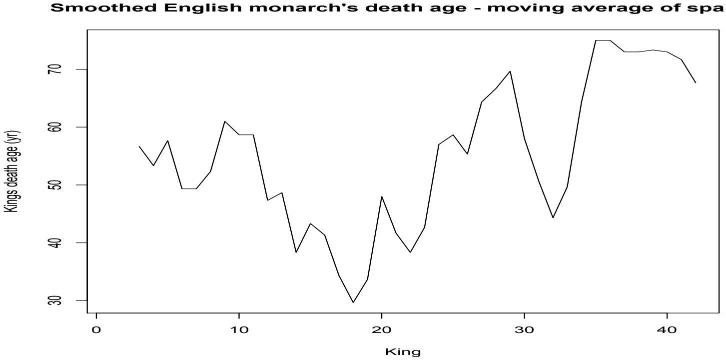
EDA - Example: Death dates of successive English monarchs

An 8 monarch span does much better. A curious trend which can now be removed to get the irregular component.



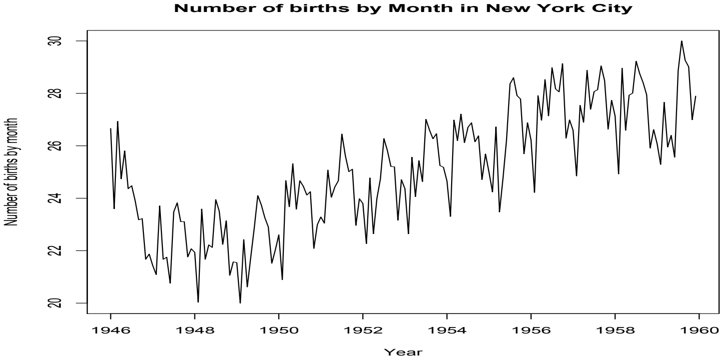
EDA - Example: Death dates of successive English monarchs

We start with a 3 Monarch moving average. The trend is clearer but more smoothing needed.



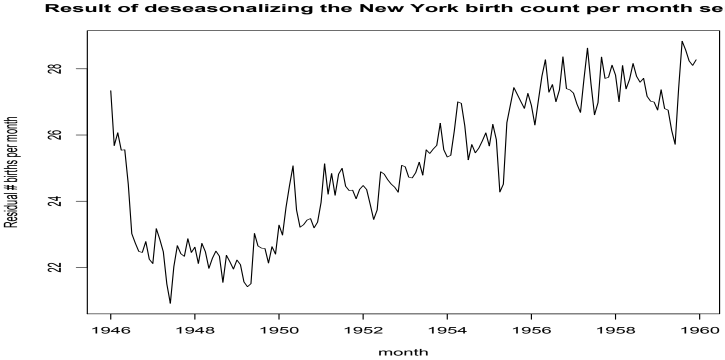
EDA - Example: Monthly birth numbers in New York City

This series exhibits both a trend and seasonal component.



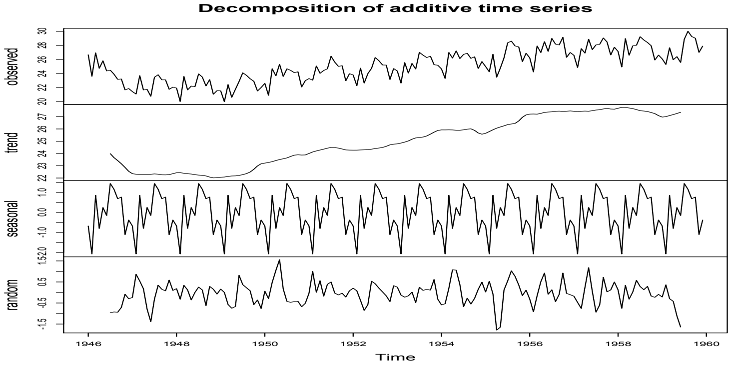
EDA - Example: Monthly birth numbers in New York City

Lets desasonalize the data to see what is going on.



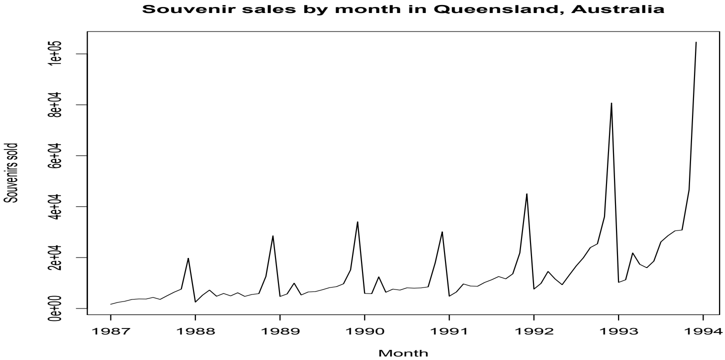
EDA - Example: Monthly birth numbers in New York City

The result has a trend plus an irregular, stationary component. We can decompose the series directly as you will see in the lab.



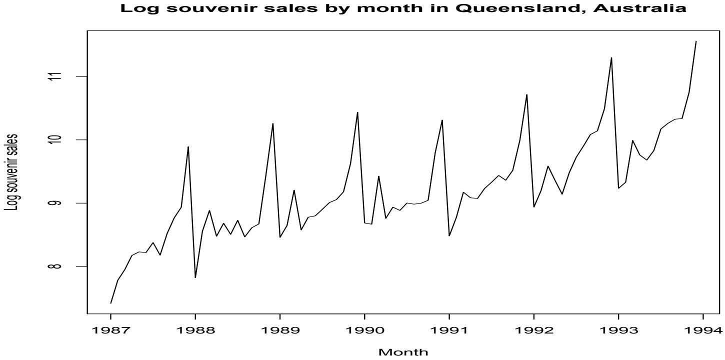
EDA - Example: Monthly souvenir sales in Queensland

A data transformation may be helpful in modelling the regular components as seen in this example. Clear seasonal pattern - but varying amplitude makes modeling tricky.



EDA - Example: Monthly souvenir sales in Queensland

A logarithmic transformation shows an additive seasonal component.



Summary re regular components

1. All the standard EDA techniques can be used with some ingenuity to identify the seasonal components
2. Once the temporal process is understood, more formal analyses can proceed. e.g Bayesian analysis
3. Regular components are often the biggest source of variation.
4. However, the irregular components are commonly correlated. These can power-up future forecasting. Anyway, they must be accounted for in the statistical analysis.
5. So now we turn to the problem of modelling them.

Note: The Nyquist frequency–

–highest frequency that can be fit to time series data –2 measurements per cycle when representing discrete process by Fourier (sine, cosine) expansion:

$$Y_t = \sum_{k=1}^{p_\alpha} \alpha_k \phi_k(t).$$

Example:. temperature measured daily at noon implies smallest cycle that can be included be a two days. To represent daily cycles needs 2 measurements per day.

Aliasing: Related issue false frequencies due to too scarce measurements. Measuring temperature at noon every 3rd day detects 6 day cycle at best and falise frequency of 1/6 of a cycle per day
Conclusion: Measure twice as often as the highest conceivable frequency to avoid the problem.

Temporal processes: the irregular components

Fundamental considerations in their analysis

1. Time series analysis has long been concerned with modeling the irregular components in a single realization of a stochastic series. In contrast longitudinal data analysis concerns replicates of short series. The two have evolved in very different ways.
2. With only a single realization, certain assumptions are needed.
   Ergodicity: Population parameters can be estimated from a single correlated realization - correlation dies fast!
   Weak stationarity: means roughly that $Y_{t_1+h} - Y_{t_1}$ and $Y_{t_2+h} - Y_{2_1}$ are identically distributed so variances and covariance can be estimated.
3. Similar ideas involved in the analysis of irregular components of spatial series.
4. Classical time and spatial series analysis have tended to focus on second order properties.

Temporal processes: moments

1. Notation: $t_1 : t_n = (t_1, \ldots, t_n)$
2. CDF: $F_{t_1:t_n}(y_1, \ldots, z_n) \equiv P(Y_{t_1} \leq z_1, \ldots, Y_{t_n} \leq z_n)$
3. Expectation: $\mu(t) = E[Y_t]$
4. Variance: $\sigma^2(t) = Var[Y_t] = E[Y_t - \mu(t)]^2$
5. Covariance: $C_Y(t_1, t_2) = C(t_1, t_2) = E[(Y_{t_1} - \mu(t_1))(Y_{t_2} - \mu(t_2))]$.
6. Autocorrelation $\rho(t_1, t_2 = C(t_1, t_2)/\sqrt{C(t_1, t_1)C(t_2, t_2)}$

Temporal processes: moments

1. Notation: $t_1 : t_n = (t_1, \ldots, t_n)$
2. CDF: $F_{t_1:t_n}(y_1, \ldots, y_n) \equiv P(Y_{t_1} \leq y_1, \ldots, Y_{t_n} \leq y_n)$
3. Expectation: $\mu_Y(t) = E[Y_t]$
4. Covariance: $C_Y(t_1, t_2) = C(t_1, t_2) = E[(Y_{t_1} - \mu(t_1))(Y_{t_2} - \mu(t_2))]$.
5. Variance: $\sigma_Y^2(t) = C_Y(t, t)$
6. Autocorrelation $\rho(t_1, t_2) = C(t_1, t_2)/\sigma_Y(t_1)\sigma_Y(t_2)$

Temporal processes: Stationarity

Stationarity comes in variety of forms.

Strong stationarity: $Y_{t_1:t_n} \sim Y_{t_1+\tau:t_n+\tau}$ for any integer τ. This is unrealistically strong.

Weak stationarity: $C(t,\tau) = C(\tau - t)$ for any t, τ pair. Most commonly used assumption, but requires process to have second moments. As well $\mu_Y(t) \equiv \mu_Y$, $\sigma_Y^2(t) \equiv \sigma_Y^2$.

Spatial processes: Analogous concepts apply to spatial processes.

Discrete time: Analogous results for discrete time. Equi–spaced sampling points. Integer valued without loss of generality.

Temporal processes: Bochner's theorem

We now assume a weakly stationary temporal process at integer valued sampling points. The next theorem concerns the covariance function which must be positive definite by defn (except in degenerate cases): $\sum_i \sum_j a_i a_j C(t_i - t_j) >$ for any vectors of constants $a_1 : a_m$ and time points $t_1 : t_m$

Bochner's theorem: C is a positive definite covariance function for a stationary process implies the existence of a spectral distribution function $F(\omega), \ -1/2 \leq \omega \leq 1/2$ such that

$$C(\tau) = \int_{-1/2}^{1/2} \exp\{2\pi i \omega \tau\} dF(\omega), \mid \tau \mid = 0, 1, \ldots$$

Corollary: $\sum_{\tau=-\infty}^{\tau=\infty} \mid C(\tau) \mid < \infty \Rightarrow dF(\omega) = f(\omega) d\omega$

Temporal processes: Bochner's theorem

Remarks:

1. Very important result. It can be used to construct temporal and spatial covariance functions by specifying f. Finding legitimate covariance functions is difficult in practice.
2. Extends to continuous time and spatial processes in a natural way.
3. States that the covariance can be expanded in terms of sines and cosines since $\exp\{2\pi i\omega\tau\} = cos(2\pi\omega\tau) + isin(2\pi\omega\tau)$. Thus $\omega = 1$ says that as τ goes from 0 to 1, cos goes from 1 to 1 to complete one period of oscillation. The sin from 0 to 0. So the frequency is 1, ie. one cycle per unit of time.
4. $f(\omega)$ is called the spectral density. It describes how power is distributed over the various frequencies ω. Later we will see that Mercer's theorem says that the process itself must be expandable in terms of sins and coss of various frequencies.
5. Examples to follow.

Temporal processes: Spectral representation theorem

A beautiful representation of stationary processes: It says that an such process is a sum of sines and cosines with random coefficients. More formally over a continuum of frequencies ω:

$$Y_t = \int_{-\infty}^{\infty} e^{i\omega t}\, dU(\omega)$$

where U is a complex values process with orthogonal increments meaning that if $\omega_1 < \omega_2 < \omega_3 < \omega_4$:

$$E[U(\omega_2) - U(\omega_1)][\overline{U(\omega_4) - U(\omega_3)}] = 0$$

Also $E[dU(\omega) d\overline{U(\omega)}] = dF(\omega)$.

Temporal processes: Spectral representation theorem

NOTES:

- Theorem says stationary processes are sums of sines and cosines
- The representation theorem provides of practical way of building stationary processes.
- Both this theorem and Bochner's extend to the spatial and spatio–temporal case
- The discrete version of this theorem replaces the integral with one over $[-\pi, \pi]$.

Temporal processes: Model types

White noise process

For time discrete time

$$Z_t = W_t \sim (\mu_W, \sigma_W^2),\ t = 1, \ldots,$$

where the $\{W_t\}$ are independent and identically distributed. Here $C(\tau) = \sigma_W^2 I\{\tau = 0\}$ so $f(\omega) \equiv \sigma_W^2$ for $-1/2 \leq \omega \leq 1/2$. Exercise. Usually $\mu_W = 0$

Random walks

For discrete time t

$$Y_t = Y_{t-1} + W_t,\ t = 1, \dots,$$

where W is a white noise process. Thus $\mu_Y(t) = t\mu_W$ and $\sigma_Y(t) = t\sigma_W^2$. Exercise. So W cannot be stationary.
Remark: The log of a stock's daily price (in standardized currency) is usually modelled as a random walk.

Autoregressive processes

These processes capture local dependence in time through a Markov like model. The do not capture long range dependence, which can be an important determinant of variation in some processes. Such dependence may persist over days or centuries. Quite a specialized topic, not to be covered in this course.

Autoregressive processes

AR(1) process. For time t & fixed spatial location s

$$Y_t = \alpha_1 Y_{t-1} + W_t,\ t =, \ldots, 0, 1, \ldots,$$

Here $\alpha_1 = \rho(t-1, t)$ for all t (stationary process); $\{W_t\}$ iid zero mean sequence

Multivariate version MAR(1).

$$\mathbf{Z}_t = \boldsymbol{\alpha}_1 \mathbf{Y}_{t-1} + \mathbf{W}_t,\ t = 1, \ldots,$$

These definitions have obvious extensions to higher dimensions an Bayesian implementations.

Exercise: Find a recursive relation for the autocorrelation function of an AR(2) process.

Moving average processes

The independent evolutionary bumps $\{W_t\}$ in the AR models, can represent shocks to a system such as an earthquake or a weather system. In that case they may take time to subside. This leads to the moving average process MA(2) in the case of the second order:

$$Y_t = W_t + \beta_1 W_{t-1} + \beta_2 W_{t-2}.$$

ARMA processes

ARMA(p,q) combines an AR(p) process and an MA(q) process. E.g ARMA(1,1) would be

$$Y_t = \alpha_1 Y_{t-1} + W_t + \beta_1 W_{t-1}$$

State space models

Generalizes the ARMA process model:
measurement model:

$$Z_t = F_t Y_t + W_t,\ W_t \sim N(0, V)$$

process model:

$$Y_t = G_t Y_{t-1} + \omega_t, \omega_t \sim N(0, W)$$

parameter model:

$$[Y_0, V, W]$$

A serious example follows when we get to spatio–temporal modeling in the course.

Model fitting

Lots of different approaches have been taken including:

- method of moments
- least squares
- likelihood based methods
- Bayesian methods

These have been implemented in the various software packages.

Example: Ergodicity implies for a stationary process suggests with additive measurement error:

$$\hat{\mu}(t) \equiv \frac{\sum_{t=1}^{T} Z_t}{T}$$

$$\hat{C}(\tau) = \frac{\sum_{t=1}^{T-\tau} (Z_{t+\tau} - \hat{\mu})(Z_t - \hat{\mu})}{T}$$

Temporal processes: forecasting

General approaches:

Ignore corr: Approach 1: ignore autocorrelation- e.g. exponential smoothing

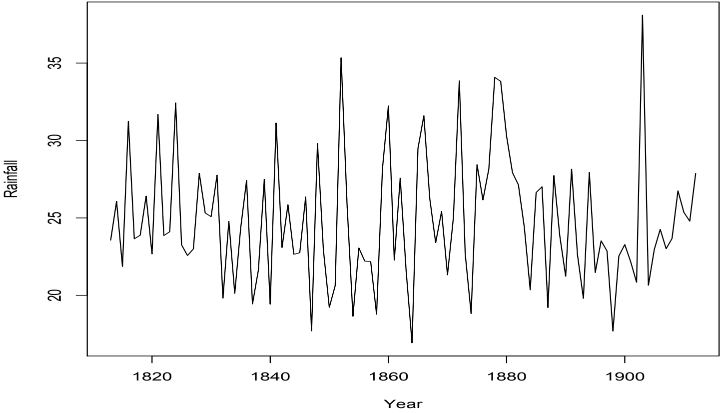
Include corr: Approach 2: include and exploit autocorrelation – e.g. Box-Jenkins

Ignoring autocorrelation

Use the series to date to fit coefficients in a forecasting model e.g. moving average and use that. Approach associated with Holt and Winters. But ignore autocorrelation in the series. Holt's approach applies with trend but no seasonality. Holt-Winters includes both.

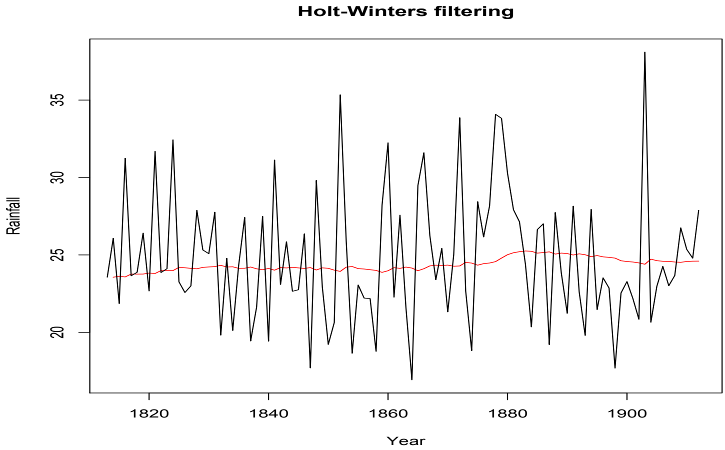
Including autocorrelation: Precipitation

Data source: Hyndman. **Note:** no trend or seasonality suggests exponential smoothing forecaster.



Ignoring autocorrelation: Precipitation

Plot depicts in red the successive forecasts using exponential smoothing.



Including autocorrelation: ARIMA processes

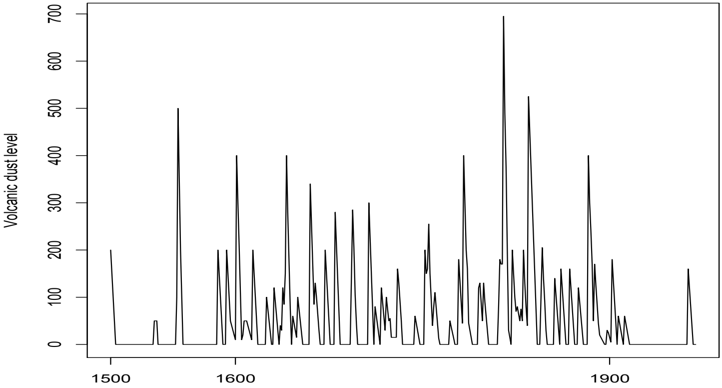
These combine AR and MA processes to get the ARMA(p,q) process.
From the last slides we would get an ARMA(1,2) process.

ARIMA

- ARIMA(p,d,q) means differencing the original series often enough – d times – to get a stationary series which is modelled as a ARMA(p,q) process. Process removes regular components: trends, seasonality, etc, without explicitly modeling them. ARMA parameters estimated. Series on original scale can then be reconstructed by ”integrating” the differences to do the forecasting.
- Library(forecast) has a routine for fitting an ARIMA process.

Including autocorrelation: Volcanic ash

Data source: Hyndman. **Note:** Complex pattern resulting from successive eruptions. Large spikes suggest possibility of moving average components. But no obvious regular components suggests ARIMA with $d = 0$.



Including autocorrelation: Volcanic ash

Next we study the correlogram = autocorrelation function plot.

Including autocorrelation: Volcanic ash

And the partial correlogram plot. Since in acf points to significance to lag 3 and pacf keeps significance lag 2, suggests a ARMA(2,0) or maybe ARMA(0,3)

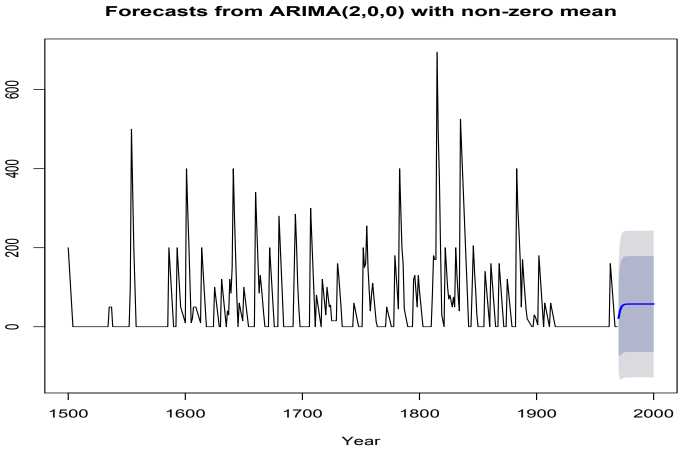
Including autocorrelation: Volcanic ash

Conclusion: For simplicity choose ARMA(2,0) with smaller number of parameters:

$$Y_t - \mu = \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + W_t$$

Including autocorrelation: Volcanic ash

We can now forecast future ash levels using the ARIMA(2,0,0) model with the result below. The next step would be diagnostic analysis.



Temporal processes: A modern view

Phystat modeling

Briefly combines deterministic models with statistical models. Overcomes deficiencies of each in a synergistic way. In particular, allows uncertainties in the deterministic models to be represented and thus adds error bars to predictions and estimates of parameters in this models. At the same time gives statistical models a “backbone” by incorporating physical knowledge. More discussion later in the course. Here we start setting the foundations.

Dynamic processes

Reference: Cressie and Wikle (2011). Excellent reference.

Temporal processes are dynamic processes that change over time. Are characterized by

$$\frac{\partial Y_t}{\partial t} = \dot{Y}_t = H[Y(t)]$$

in continuous time and

$$\nabla Y_t = Y_t - Y_{t-1} = H[Y_t]$$

in discrete time.

Notes:

1. H and Y_0 known and fixed makes Y deterministic. The Y usually written in l.c. y.
2. Nonlinear H can lead to chaotic behavior in deterministic systems. In that case small changes in Y_0 can lead to huge variations in $Y(t)$ - the *butterfly effect*.

Dynamic processes - DP

Notes (cont'd):

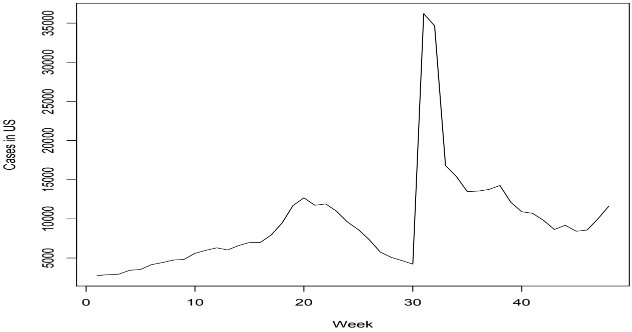
3. Y_0 can be random. Moreover system can be subject to random environmental disturbances so that

$$\frac{\partial Y_t}{\partial t} = \dot{Y}_t = H[Y(t)] + \eta_t$$

where η is random and independent of Y. A Bayesian approach makes parameters in a deterministic system described by H random. In all these cases we get a random process even though we start with a conceptually deterministic one. Hence Bayesian hierarchical models are very fashionable even in the physical sciences.

Infectious disease as a DP: Example

The US Center for Disease Control maintains surveillance program on the incidence of influenza cases. Figure shows results over the 2008 - 2009 season



Infectious disease as a DP: Modeling

Reference: Leduc, H. (2011). Estimation de paramètres dans des modéles d'épidémies. Mémoire de maitrise, Département de mathématiques, UQAM.

Thesis is based on the SIR (susceptibles–infected–recovered) model for infectious diseased going back to 1927. Here: x_t, y_t z_t denote the numbers in the 3 categories and they must sum to N the population size.

Dynamic process model:

$$
\begin{aligned}
\dot{x}_t &= -\beta_N x_t y_t \\
\dot{y}_t &= \beta_N x_t y_t - \gamma_N y_t \\
\dot{z}_t &= \gamma_N y_t
\end{aligned}
$$

Infectious disease as a DP: Modeling

Leduc elegantly models the process by combining a stochastic and deterministic process with random X_t, Y_t, and Z_t replacing their deterministic counterparts. Since Y is determined by X and Z only these are stochastically modelled. In particular

- X_t is a non homogeneous pure death process with with rate $\beta_N^* x_t y_t$
- Z_t is a non homogeneous Poisson (birth) process with rate γy_t

Notes:

4. Methods for estimating the model parameters are developed and applied

5. Cressie and Wikle devote two long chapters to dynamic systems modeling and inference for their parameters.

Temporal processes: Summary

We have explored:

- classical time series modeling:
    - foundations
    - exploratory data analysis with associated software
    - classical models up to state state modeling
- a modern perspective
    - dynamic systems including deterministic models
    - limitations of physical models
    - infectious diseases and how they might be phystat modelled.