

Session 3.1: Introduction to temporal modelling

Spatial and Spatio-Temporal Bayesian Models with R-INLA, University of São Paulo

28 September 2022

Learning objectives

After this lecture you should be able to

- Explain why time also matter and describe features of time series data
- Understand the difference between stationary and nonstationary temporal processes
- Describe basic temporal models
- Know the key functions to implement temporal models through the R-INLA package

Some of the topics treated in this lecture are presented in Chapter 8 of the book **Bayesian inference with INLA** by Virgilio Gómez-Rubio (link: <https://becarioprecario.bitbucket.io/inla-gitbook/index.html>).

Outline

1. Time series
2. Features of time series
3. Stationarity
4. Basic temporal models

Time series

Introduction

- We live in a complex world and it is often not sufficient to consider just snapshots of a spatial process at a given time.

Introduction

- We live in a complex world and it is often not sufficient to consider just snapshots of a spatial process at a given time.
- Time also matter: the behaviour from one time point to the next is important and many data that we deal with in spatial analysis are actually both spatial and temporal in nature.

Introduction

- We live in a complex world and it is often not sufficient to consider just snapshots of a spatial process at a given time.
- Time also matter: the behaviour from one time point to the next is important and many data that we deal with in spatial analysis are actually both spatial and temporal in nature.
- Similarly to spatial dependence, it is sometime necessary to model temporal dependence on data and parameters.

Introduction

- We live in a complex world and it is often not sufficient to consider just snapshots of a spatial process at a given time.
- Time also matter: the behaviour from one time point to the next is important and many data that we deal with in spatial analysis are actually both spatial and temporal in nature.
- Similarly to spatial dependence, it is sometime necessary to model temporal dependence on data and parameters.
- Unlike space, the temporal data hold a **natural order**.

Example of time series studies from environmental epidemiology

- In environmental epidemiology, time series have been widely used, notably for investigating the short-term associations between exposures such as air pollution or weather variables, and health outcomes such as cardiovascular and respiratory morbidity and mortality.
- Typically, for both exposure and outcome, data are available at regular time intervals (e.g. daily pollution levels and daily mortality counts) and the aim is to explore short-term associations between them.

Time series

- A **time series** is a set of observations taken sequentially in time.
- Depending on different applications, data may be collected hourly, daily, weekly, monthly, yearly, and so on.
- A time series that can be recorded continuously in time, is said to be **continuous**, while a time series that is taken only at specific time intervals is said to be **discrete**. We will work mainly with discrete time series data.
- We use notation such as $\{Z_t : t \in \mathcal{D}_t\}$. Henceforth, we assume that $\mathcal{D}_t = \{0, 1, \dots\}$ and we refer to $\{Z_t : t = 0, 1, \dots\}$ as a time series.
- The natural (temporal) ordering in the time series creates an internal structure in the data, that shows, commonly, dependence in the observations, *such that values in the present depend upon observations available in the past*.

Autocorrelation or serial correlation

- Autocorrelation → the correlation of a variable with itself.
- **Space**: the correlation between the value of the variable at two different locations (or areas).
- **Time series**: the values of a variable at time t depends on the value of the same variable at time $t - h$, where h is the **time-lag separation**.
- Thus, autocorrelation is also sometimes called *lagged correlation* or *serial correlation*, which refers to the correlation between members of a series of numbers arranged in time.

Time series: mean, autocovariance and autocorrelation

A time series $\{Z_t\}$ has:

- **Mean function:** $\mu_t = E(Z_t)$
- **Autocovariance function:** $C(t, r) = \text{cov}(Z_t, Z_r) = E\{(Z_t - \mu_t)(Z_r - \mu_r)\}$
- **Autocorrelation function:** $\rho(t, r) = \frac{C(t, r)}{\sqrt{C(t, t)C(r, r)}}$ where $\rho(t, r) \in [-1, 1]$.

Notice that:

- The mean indicates the trend of the series
- The autocovariance function summarizes how the process co-varies across different time lags, after accounting for the mean function. We have: $C(t, r) = C(r, t)$
- The variance is a special case of the autocovariance in which $C(t, t) = \text{var}(Z_t) = \sigma_t^2$

Features of time series

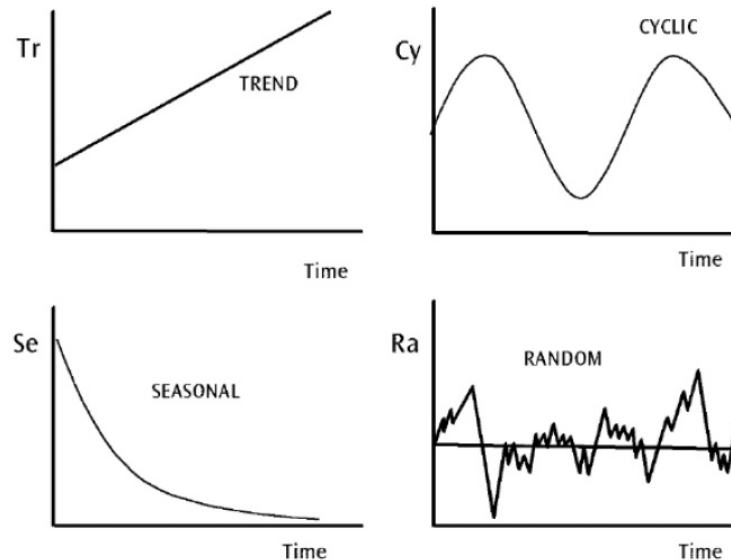
Components of time series

- Time series analysis typically presents challenges, as it exhibits **patterns** and **irregular fluctuations**.

Patterns can be specified as:

- Trend**, that is the most common time series feature to account for and refers to long-term change in the mean level;
- Seasonal variation**, which refers to periodic fluctuations which occur periodically within a year;
- Cyclic changes**, which are recurrent rise and fall that are not of fixed period and are over a period longer than one year.

Irregular fluctuations are variations that are short in duration, following not regularity in the occurrence.



Stationarity

- In studying time series, a very important concept is given by **stationary**, that refers to the stability of the statistical properties of the process through time.
- Broadly speaking, a stationary process is one whose statistical properties do not change over time.
- There are two important forms of stationarity:
 - **strong stationarity**;
 - **weak (or second order) stationarity**.

Strong and weak stationarity

A time series is said to be **strongly stationary** if

- for any finite sequence of times t_1, t_2, \dots, t_n and any temporal lag h the probability distribution of the vector $(Z_{t_1}, \dots, Z_{t_n})'$ is identical to the probability distribution of the vector $(Z_{t_1+h}, \dots, Z_{t_n+h})'$.
- *In words: all aspects of the process's behavior are unaffected (unchanged) by a shift in time.*

A time series is said to be **weak (or second order) stationary** if

- $E(Z_t) = \mu$, i.e. the mean is constant for all t
- $\text{var}(Z_t) = \sigma^2$, i.e. the variance does not depend on t
- $\text{Cov}(Z_t, Z_r) = C(t - r)$, i.e. the autocovariance depends only on the elapsed time between t and r and not their actual location
- *In words: weak stationarity (only) concerns the shift-invariance of first and second moments of a process.*

Basic temporal Models

White noise process [1]

- A white noise process is a sequence of independent normally identically distributed random variables
- The term **noise** is due to the fact that there's no pattern, just random variation
- The **Gaussian white noise** is defined as:

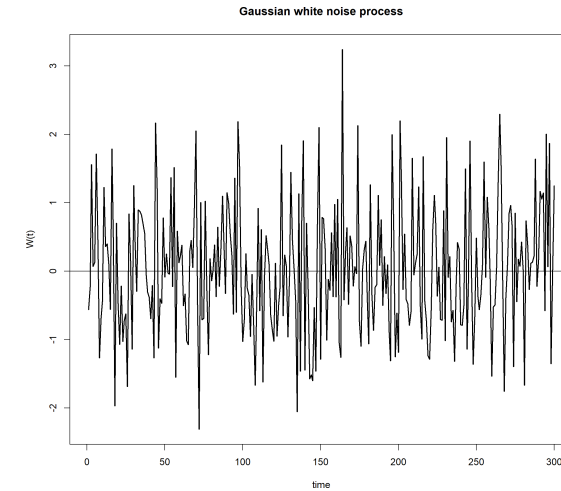
$$W_t \stackrel{iid}{\sim} N(0, \sigma_W^2)$$

- This process is stationary:
 - $E[W_t] = 0$, i.e. the expectation is always constant and equal to zero
 - $var(W_t) = \sigma_W^2$, i.e. the variance is constant
 - $cov(W_t, W_r) = 0$ for $t \neq r$, i.e. the covariance is zero at all lags
- Note, iid stands for **independent and identically distributed**. An iid model assumes that observations on a phenomenon are taken under identically conditions and that each observation is taken independently of any other.

White noise process [2]

- Realization of a Gaussian white noise process $W_t \stackrel{iid}{\sim} N(0, \sigma_W^2)$

```
> set.seed(123) # set random number seed
> W = rnorm(300) # generate iid normal random variates
> ts.plot(W, main="Gaussian white noise process",
+         xlab="time", ylab="W(t)",
+         col="black", lwd=2)
> abline(h=0)
```



- This figure shows that there are no discernible patterns and the distribution is completely random
- In R-INLA an iid Gaussian random effect is specified with the model `iid`
- To obtain details about the model `iid` we can type `inla.doc("iid")`

Random walk (RW) [1]

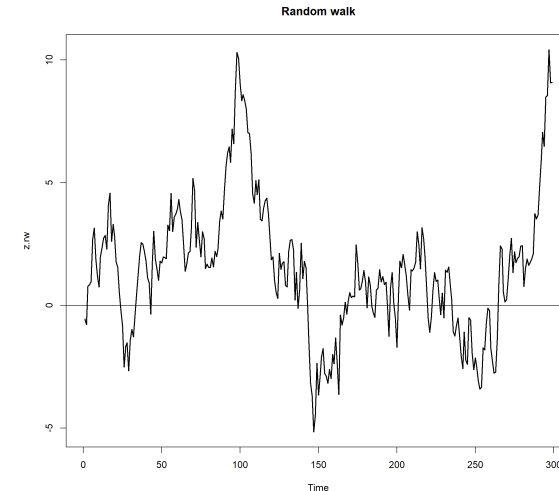
- The random walk (RW) describes how an observation directly depends upon one or more previous measurements plus a white noise process
- The **random walk of order 1, RW1**, is defined as:

$$Z_t = Z_{t-1} + W_t$$

where W_t is a white noise process.

- Realization of a RW1

```
> set.seed(123) # set random number seed
> Z0 <- 0 #Z0 is fixed
> T <- 300
> W <- c(Z0 + rnorm(T-1))
> z.rw <- cumsum(W) # compute cumulative sum
> ts.plot(z.rw, main="Random walk", lwd=2, col="k")
> abline(h=0)
```



- The RW1 only models the difference of levels on consecutive time points: $Z_t - Z_{t-1} = W_t$

Random walk (RW) [2]

- The RW is a non-stationary process (i.e. observations in a random walk are dependent on time)
- For a RW1, by recursively substitution, starting from $t = 1$, we have:

$$\begin{aligned}Z_1 &= Z_0 + W_1 \\Z_2 &= Z_1 + W_2 = Z_0 + W_1 + W_2 \\&\vdots \\Z_t &= Z_0 + W_1 + \dots + W_t \\&= Z_0 + \sum_{j=1}^t W_j\end{aligned}$$

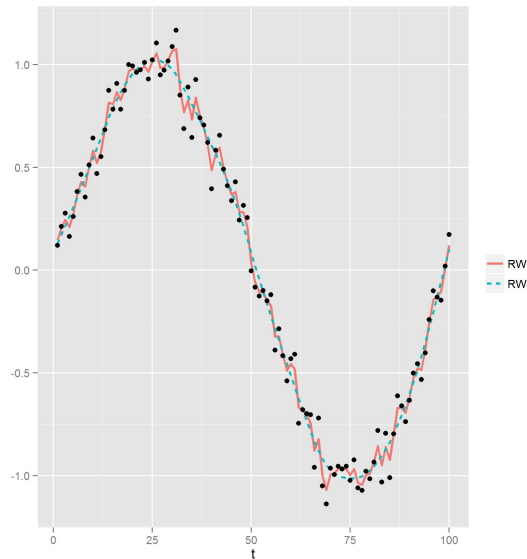
Hence, the first order moment (or the expected value) for this process is equal to:

$$E(Z_t) = Z_0 + \sum_{j=1}^t E(W_j) = Z_0, \text{ which is independent of } t.$$

The variance is $\text{var}(Z_t) = \text{var}\left(\sum_{j=1}^t W_j\right) = \sum_{j=1}^t \sigma_W^2 = t\sigma_W^2$, which depends on t . Thus the random walk process $\{Z_t\}$ is not stationary.

Random walk (RW) [3]

- The **random walk of order 2, RW2**, is defined as: $Z_t = 2Z_{t-1} - Z_{t-2} + W_t$
- The RW2 only models a linear combination of levels on consecutive time points: $Z_t - 2Z_{t-1} + Z_{t-2} = W_t$



- The parameterization of the RW1 model in R-INLA is $Z_t | Z_{t-1} \sim N(Z_{t-1}, \sigma_Z^2)$, and it is implemented through the specification `rw1`.
- The parameterization of the RW2 model in R-INLA is $Z_t | Z_{t-1}, Z_{t-2} \sim N(2Z_{t-1} - Z_{t-2}, \sigma_Z^2)$, and it is implemented through the specification `rw2`.

Autoregressive (AR) process [1]

- The **autoregressive process of order (p), AR(p)** is a time series model where the original data is expressed as a function of its previous values in time
- It is defined as:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + W_t$$

where:

- W_t is a Gaussian error term with mean zero and variance σ_W^2 (i.e. a Gaussian white noise process)
 - $\{\phi_i : i = 1, \dots, p\}$ is a sequence of unknown autoregressive parameters
- This class of models is called autoregressive because Z_t is regressed on past terms of the same process
 - The simplest model is given by the **AR1** (i.e. $p=1$) and is defined as:

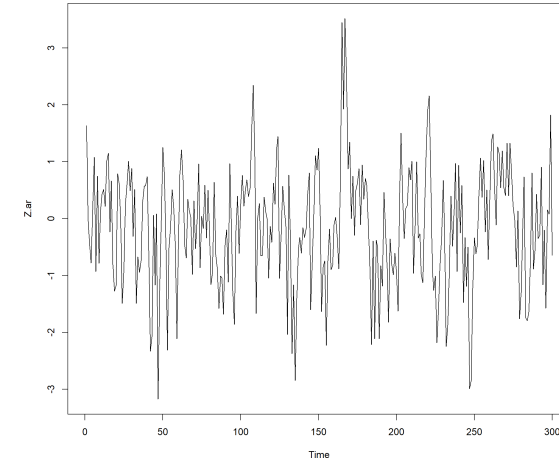
$$Z_t = \rho Z_{t-1} + W_t$$

where $|\rho| < 1$ is the unknown temporal correlation term

Autoregressive (AR) process [2]

- Realization of an AR1 process

```
> set.seed(121) # set random number seed  
> Z.ar <- arima.sim(model=list(ar=.5), n=300)  
> plot.ts(Z.ar)
```



- The AR1 process can be written as an infinite series of white noise random variables. Since $E(W_t) = 0$ and $var(W_t) = \sigma_w^2$, it follows that $E(Z_t) = 0$ and $var(Z_t) = \frac{\sigma_w^2}{1-\rho^2}$, which does not depend on t , thus the process is stationary. Note that if $\rho = 1$, the process is a random walk.
- In R-INLA, the AR1 model is implemented through the model specification `ar1`, while an AR model of arbitrary order is implemented through the specification `ar`.
- To obtain details about the specification of the AR1, and more in the general about the AR model, we can type `inla.doc("ar1")` and `inla.doc("ar")`.

To wrap up

A stationary time series **doesn't exhibit trend or seasonality**:

- Observations do not tend upwards or downwards
- Variance does not increase or decrease with time
- Observations do not tend to be large in some periods than others

References

- Broemeling D.L. (2019), Bayesian Analysis of Time Series, CRC Press
- Cressie N. and Wikle C.K. (2011), Statistics for spatio-temporal data, Wiley
- Zivot E. and Wang J. (2006), Modeling Financial Time Series with S-PLUS, Springer, 2nd ed.