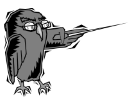
14 Time-series analysis in R-INLA

The scientific field of time-series analysis consists of such a wide variety of techniques that we could easily fill an entire book about this topic; see for example Harvey (1989), Chatfield (2003), Shumway and Stoffer (2017), and Durbin and Koopman (2012), among many others.

In this chapter we do not want to delve too deeply into time-series analysis techniques. Instead, we will touch upon methods available in R-INLA that can be used to identify trends. Some of these techniques will also be used in Chapters 15 and 16 which deal with spatial-temporal models.



Prerequisite for this chapter: You need to be familiar with the time-series analysis topics that were discussed in Chapter 3 and with linear regression and R-INLA as discussed in Chapter 8.

14.1 Simulation study

Let Y_t be a response variable that has been measured repeatedly over time, e.g. the biomass of a fish at time *t* or the number of birds in year *t*. We will assume that Y_t follows a certain distribution, given its mean value. For example we can use a normal distribution for the biomass of fish or a Poisson distribution for the number of birds. Actually, depending on the nature of the response variable we can use any distribution. It is the mean value of the distribution that is relevant for this chapter. We will model it as a function of a trend over time, and possibly also covariate effects. In its general form, using a normal distribution, the model is given by

$$\begin{aligned} Y_t &= \text{Intercept} + \text{Covariates}_i + \text{Trend}_t + \varepsilon_t \\ \varepsilon_t &\sim N\left(0, \sigma_\varepsilon^2\right) \end{aligned} \tag{14.1}$$

This looks very much like an ordinary linear regression model. But it depends on what we are going to do with the Trend_t component, whether this is indeed a linear regression model or whether it is something else.

Using a model in which the trend component is modelled as $Time \times \beta$ means that we indeed have an ordinary regression model, and this ultimately results in pseudoreplication when we apply the model to time-series data. To avoid this we will use a so-called random walk to model the trend. The simplest random walk model is given in Equation (14.2); we dropped the Covariates_i term.

$$Y_t = \text{Intercept} + \mu_t + \varepsilon_t$$
$$\mu_t = \mu_{t-1} + v_t \qquad (14.2)$$
$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad \text{and} \quad v_t \sim N(0, \sigma_v^2)$$

Let us for the sake of simplicity assume that the time index t refers to years. According to the model the response variable Y_t in year t is equal to a trend μ_t plus independent, identical, and normal distributed noise ε_t. The random walk trend itself is modelled as the trend from last year plus some independent, identical, and normal distributed noise v_t. The latter term has mean 0 and variance σ_v^2. In words, using the birds as an example, we have

Birds in year t = Trend in year t + pure noise ε_t
Trend in year t = Trend from last year + pure noise v_t

The model has two residual terms (the 'pure noise' components), namely the ε_t and the v_t, and each term has its own variance; these are σ_ε and σ_v respectively. To understand the roles of σ_ε and σ_v in the model we simulated 15 data sets; see Figure 14.1. These simulated data sets use different values for σ_ε and σ_v.

Based on the shapes of the curves we can conclude that a small value for σ_v implies that the v_t values tend to be small and consequently μ_t is close to μ_{t-1}. The resulting random walk trend μ_t is a smooth curve; see for example the simulated data sets 4–6. On the other hand, a large σ_v means that the v_ts can be large and μ_t can be rather different from μ_{t-1}. As a result we obtain trends that can change quite a lot (see simulated data sets 10–15).

The random walk trend is given by $\mu_t = \mu_{t-1} + v_t$, where $v_t \sim N(0, \sigma_v^2)$. The smaller the σ_v, the smoother the trend.

A small value of σ_ε means that the ε_t values are likely to be small, and as a result the variation of the response variable Y_t around the trend is small. Formulated differently, we have a good fit (see simulations 1–3 and 10–12). The opposite also holds; a large value for σ_ε means larger ε_ts, and therefore a large scatter around the trend (simulations 4–6 and 13–15).

It is relatively easy to add covariates to the model. These are added on the first line of the model. It is also possible to add daily, quarterly, seasonal, or cyclic components to the model.

Models with random walk trends are described in detail in Harvey (1989) and Durbin and Koopman (2012). More advanced applications can be found in Zuur et al. (2003a; 2003b; 2004) who applied a multivariate extension and dynamic factor analysis.