Modelling spatial data in R with CARBayes

Part 3: Spatio-temporal modelling with CARBayesST

Duncan Lee and Eilidh Jack

GEOMED 2017

1. Introduction

This session will cover the following:

- Goals of a spatio-temporal analysis.
- Different types of spatio-temporal models to address different questions.

Spatio-temporal modelling

- In its simplest setting, spatio-temporal modelling extends the spatial setting by having data for the same set of K areal units for multiple consecutive time periods.
- Suppose data are available for $t = 1, \ldots, N$ consecutive time periods, then you have KT observations in total.
- When doing the modelling, each variable needs to be stored as a single vector of length KT, where the first K observations relate to time 1, the next K observations relate to time 2 and so on.
- Within each block of K data points for a single time period, the ordering of the spatial units must be the same.

Notation

The notation used extends naturally from that used in purely spatial data, where k denotes the areal unit (space) and t denotes the time period. So we have:

- Y_{kt} is the measure of disease / health for area k and time period t.
- E_{kt} is the population size or indirectly standardised expected number of disease cases (for discrete outcomes) for area k and time period t.
- $\mathbf{x}_{kt} = (1, x_{kt2}, \ldots, x_{ktp})$ is a vector of p covariates for area k and time period t.

We illustrate the Poisson model below, but binomial and Gaussian models are similar extensions to their spatial only versions.

Example - Poisson model

If the disease data Y_{kt} for area k and time period t is a count of the number of events, then an appropriate Poisson model is given by:

$$\begin{aligned} Y_{kt} &\sim \text{Poisson}(E_{kt}\theta_{kt}), \\ \ln(\theta_{kt}) &= \beta_1 + \beta_2 x_{kt2} + \ldots + \beta_p x_{ktp} + \phi_{kt}, \end{aligned}$$

where θ_{kt} is the risk of having the event in area k and time period t and is modelled by:

- x_{ktj} - Known area-level covariate risk factors such as smoking rates, etc.
- ϕ_{kt} - Random effects that allow for spatio-temporal correlation and trends in the disease data due to unmeasured confounding.

Implementing spatio-temporal models

- These models are most often fitted in a Bayesian setting, using Markov chain Monte Carlo (MCMC) simulation.
- The `R` package `CARBayesST` can fit these types of models, and a number of different formulations are possible depending on what type of spatio-temporal pattern you wish to detect.
- This is a sister package to `CARBayes`, and thus it is designed to look very similar.
- The software can fit Poisson, binomial and Gaussian models, which thus allow one to model both continuous and discrete data.
- These models differ in how they represent ϕ_{kt}, which represents the unmeasured spatio-temporal pattern in the disease / health data.

Goals of a spatio-temporal data analysis

There are two main goals when modelling spatio-temporal areal unit health data:

- Estimate the spatio-temporal pattern in disease risk and identify high-risk areas, overall trends, etc.
- Estimate the effects of covariate risk factors on disease risk.

If the first of these is your goal then the choice of model will depend on the type of space-time patterns you wish to detect.

2. Types of models

A large variety of spatio-temporal models have been proposed in the literature, and `CARBayesST` currently can fit 6 different model types. The most commonly used ones are:

- `ST.CARanova()` - overall space and time effects.
- `ST.CARar()` - evolving spatial surface over time.
- `ST.CARlinear()` - linear time trends for each area.

The choice of model depends on the goal of the analysis.

ST.CARanova()

Schematically, this model represents the spatio-temporal variation in ϕ_{kt} as:

$$\phi_{kt} \;=\; \text{Spatial effect}_k + \text{temporal effect}_t + \text{Interaction}_{kt}$$

where

- Spatial effect$_k$ - is an overall spatial effect common to all time periods.
- temporal effect$_t$ - is an overall temporal trend common to all spatial areas.
- Interaction$_{kt}$ - are independent space-time interactions. Essentially they capture any deviation from the overall spatial and temporal main effects.

Why use `ST.CARanova()`?

This model is useful if:

- The goal of the analysis is to estimate:
  - An average region-wide temporal trend in time.
  - An average spatial pattern common to all time periods.
  - Local deviations from these overall spatial and temporal trends.
- One believes that the structure in the data can be represented well by overall spatial and temporal main effects.
- One wishes to see, by means of the interaction terms, which areas and time periods deviate from the overall pattern.

The interaction terms can be removed from the model if required.

ST.CARar()

Schematically, this model represents the spatial surface at time t given by $\boldsymbol{\phi}_t = (\phi_{1t}, \ldots, \phi_{Kt})$ as:

$$\boldsymbol{\phi}_t \;=\; \gamma \boldsymbol{\phi}_{t-1} + \text{spatially correlated error}_t$$

- So the spatial surface at time t is equal to a proportion of the spatial surface at the previous time interval $t-1$ plus spatially correlated error.
- The spatially correlated error is modelled by the CAR model proposed by *Leroux et al. (2000)* that we discussed earlier.

Why use `ST.CARar()`?

This model is useful if:

- The goal of the analysis is to estimate the effect of covariates on disease, and the spatial correlation is a nuisance (not of interest) and just needs to be modelled.
- If you expect the spatial surface in disease risk to evolve slowly over time, rather than having the same spatial surface for all time periods.
- If the spatio-temporal dynamics in risk are unknown, then this model makes no assumption about the shapes of any trends. It only assumes things don't change too quickly in time or space.

ST.CARlinear()

Schematically, this model represents the spatio-temporal variation in ϕ_{kt} as:

$$\phi_{kt} \;=\; \alpha_k + \delta_k t$$

a separate linear trend in time for each area. Here

- α_k - is the intercept parameter of the time trend for area k.
- δ_k - is the slope parameter of the time trend for area k.

The intercept and slope parameters (α_k, δ_k) for area k are both correlated across space by the CAR prior proposed by *Leroux et al. (2000)* that we discussed earlier.

Why use `ST.CARlinear()`?

This model is useful if:

- The goal of the analysis is to estimate a separate (but correlated) linear time trend in disease risk for each area.
- One wishes to see which areas show a significantly decreasing trend over time, which areas show a significantly increasing trend over time and which show no significant change.
- One wishes to see whether areas that start off with high / low disease risk are more likely to exhibit an increased / decreased disease risk in time.