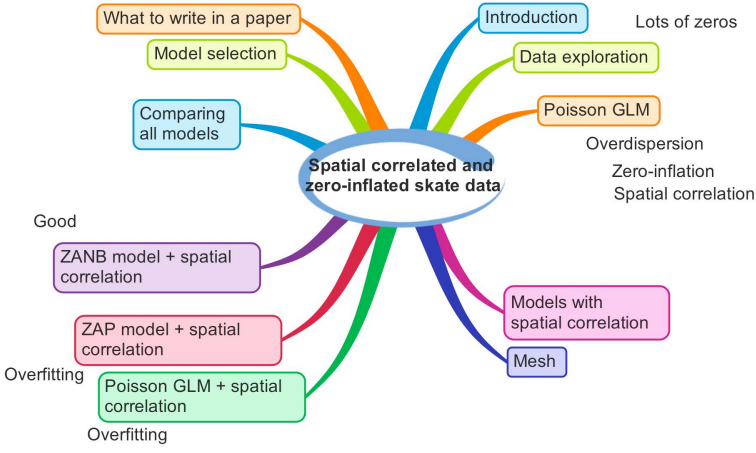
19 Spatial correlated and zero-inflated skate data

In Chapter 18 we used a skate data set to explain the underlying theory of zero-inflated Poisson (ZIP) and zero-altered Poisson (ZAP) models, and we also showed how to implement these models in R-INLA. In this chapter we will use the skate data again and extend the models with spatial correlation.

The flowchart below shows the key components of this chapter. We will start with an introduction and data exploration, followed by the application of a Poisson generalised linear model (GLM). As we will see, this model is overdispersed and to fix this we are faced with a world of options. After trying a series of models that give a perfect fit we eventually end up with a zero-altered negative binomial (ZANB) GLM.



Prerequisite for this chapter: We assume that you are familiar with the material explained in Chapters 10 and 13 (applying Poisson GLM with spatial correlation in R-INLA) of Volume I and Chapter 18 (applying ZAP models in R-INLA) of Volume II.

19.1 Introduction

Zuur et al. (2012a) used the same skate data and applied ZIP models on three different skate species, namely *Sympterygia bonapartii* (smallnose fanskate), *Rioraja agassizi* (Rio skate) and *Atlantoraja castelnaui* (spotback skate). During the preparation of this chapter, we played around with the data from all three species. It turned out that the analyses of the Rio skate and spotback skate were reasonably straightforward, but the

analyses of the smallnose fanskate gave us some serious trouble. Instead of presenting a chapter with an easy-to-analyse species, we decided to present the analyses of the smallnose fanskate data and show the problems we encountered and how we solved them.

The data were collected from the Rio de la Plata basin, a major river basin of the Argentine coast (Figure 19.2). The data are imported in R with the following code.

```
> Skate <- read.table(file = "Skate2.txt",
                      header = TRUE,
                      dec = ".")
```

We also load all required packages for this chapter and source our support function.

```
> library(lattice)
> library(INLA)
> library(sp)
> library(gstat)
> library(ggplot2)
> library(fields)
> library(maps)
> library(maptools)
> library(mapdata)
> library(dismo)
> library(rgeos)
> data("worldHiresMapEnv")
> source("HighstatLibV11.R")
```

The species data are called SB, RA and AC (referring to their Latin names) in the data file. The variable SweptArea is the area surveyed and represents the sampling effort per site. We can either use the log-transformed sampling effort as an offset variable (Zuur et al. 2013) or use sampling effort as a covariate. Here, we go with the second approach. Other covariates are year (sampling took place in 1998, 1999, 2003 and 2005), month (November and December), latitude, longitude, depth, temperature, salinity and bottom type (with the levels mud, sand and mud, sand, sand/shells/rest).

We first define the categorical covariates in R. We express the UTM coordinates in kilometers.

```
> Skate$fYear       <- factor(Skate$Year)
> Skate$fMonth      <- factor(Skate$Month)
> Skate$fBottomType <- factor(Skate$BottomType,
                      levels = c(1, 2, 3, 4),
                      labels = c("Mud", "SanMud",
                                 "Sand", "Rest"))
> Skate$Xkm <- Skate$X / 1000
> Skate$Ykm <- Skate$Y / 1000
```