22 Zero-inflated continuous seabird data

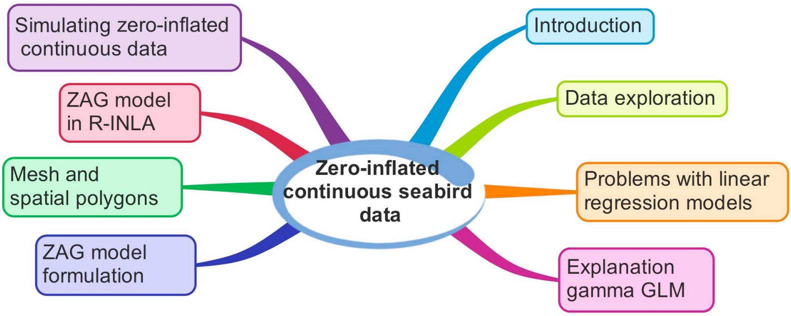
Alain F Zuur, Elena N Ieno, David A Fifield,
April Hedd, Carina Gjerdrum

In Chapters 18–21 we discussed models for zero-inflated count data. We considered Poisson generalised linear models (GLM), negative binomial GLMs, zero-inflated Poisson (ZIP) models, zero-altered Poisson (ZAP) models, zero-inflated negative binomial (ZINB) models and zero-altered negative binomial (ZANB) models.

In this chapter we focus on zero-inflated continuous data and apply a so-called zero-altered gamma (ZAG) model. The ZAG model is a 'sister' of the ZAP and ZANB models.

The data that we will analyse are seabird densities sampled at a large number of spatial locations in the Labrador Sea, located between the Labrador peninsula (Eastern Canada) and Greenland. This means that besides the zero-inflation component, there is also a spatial element in the analysis. And to make matters even worse, there are two covariates that have a non-linear relationship with seabird densities, which means that we also need smoothers.

The flowchart below shows the key elements of this chapter. We will start with a short introduction and data exploration. We then show why linear regression applied on the original data and also for log-transformed data doesn't work for this data set. We explain the principle of a ZAG model (which requires knowledge of gamma GLM, and we will explain this as well) and implement it in R-INLA. In the last section we show how to simulate zero-inflated continuous data with the ZAG model so that we can assess whether the model produces data comparable to the sampled data.



`

Prerequisite for this chapter: Knowledge of ZAP models (Chapter 18), how to include spatial dependency (Chapter 19) and using smoothers in R-INLA (Chapter 20) is required.

22.1 Introduction

The data come from the Eastern Canada Seabirds At Sea (ECSAS) database (Environment and Climate Change Canada 2016, Gjerdrum et al. 2012). The portion of data here covers the Labrador Sea to the edge of Canada's Exclusive Economic Zone.

Data were collected from vessels of opportunity in the Labrador Sea in the Fall (September and October) of 2012–2014 during continuous (nominally) 5-min survey segments, so that consecutive segments are close in both space and time. The data have been pre-processed to account for detectability using distance sampling; see Fifield et al. (2017) for details. Each data row contains the estimated density of all seabirds in that segment in birds/km^2 along with a number of environmental covariates thought to affect seabird density. These covariates are given below.

- Lat: The unprojected latitude (WGS84) of the observation.
- Lon: The unprojected longitude (WGS84) of the observation.
- DateTime: The date and time of the observation (GMT)
- SST: Sea surface temperature assigned to the observation (derived from remote sensing)
- EKE: Eddy kinetic energy (as a proxy for oceanic fronts) assigned to the observation (derived from remote sensing)
- SSH: Sea surface height anomaly assigned to the observation (derived from remote sensing)
- Depth: Water depth at location of observation (in meters)
- Slope: Slope of the ocean floor at the location of the observation
- SSTGrad: Gradient (slope) of sea surface temperature at the location of the observation (derived from remote sensing)
- Dist1000: Distance to the continental shelf edge (taken as the 1000 m depth contour)

The underlying question is which of these covariates is important for bird densities.

22.2 Data exploration

In this section we apply a short data exploration. We start by importing the data from a csv file.

```
> SB <- read.csv("LabradorSeaBirdsV2.csv",
                  header = TRUE)
```