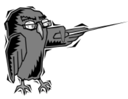
16 Spatial-temporal Bernoulli models for coral disease data

In this chapter we use data from Muller and van Woesik (2014) who used R-INLA to model the absence and presence of white-pox disease on *Acropora palmata* coral colonies in the Caribbean. We will use their data to illustrate the use of Bernoulli models with spatial and spatial-temporal correlation. We will minimise the technical explanation as this can be found in Chapter 15.



Prerequisite for this chapter: We assume that you are familiar with the material explained in Chapter 15. You also need to be familiar with Bernoulli GLMs; see Chapter 10.

16.1 Introduction

Infectious diseases are a major health threat for coral reefs. To protect healthy coral it is important to know whether a disease is contagious or whether it is due to environmental stress. White-pox disease, which may be caused by an infectious agent, has caused considerable coral mortality on Caribbean reefs.

Muller and van Woesik (2014) studied the absence and presence of white-pox disease on 68 *A. palmata* colonies in the US Virgin Islands. Their model contains a series of covariates that might affect the disease and also a spatial-temporal component that may indicate whether the disease is contagious (i.e. whether it was transmitted to nearby neighbours).

Between February 2003 and December 2009, 68 reef colonies in Haulover Bay on the northeast side of St. John, US Virgin Islands (Figure 16.1) were sampled on a monthly basis. The paper uses the following covariates: Northing, Easting, distance to nearest neighbour (DIST), previous incidence of disease (PIC), distance to previous infected colony (DFPIC), temperature (TEMP), solar insulation (IRR), and colony size (SIZE).

Data exploration indicates that there are no covariates with extremely small or large values, and there is no collinearity except for easting and northing. Muller and van Woesik (2014) did include these two variables as covariates in the fixed part of the model. This opens a more general discussion as to whether we should fit a model in which spatial coordinates are used as covariates and also use a spatial random field in the model. It has been suggested on the INLA discussion board to include spatial coordinates as covariates to capture the large-scale patterns and let the spatial correlation term capture any small-scale spatial patterns; apparently, this is common practice. We feel slightly uneasy about this because including the spatial random field and the spatial coordinates as covariates in the same model may add an extra level of collinearity. We will not use the two spatial coordinates as covariates, but you are free to change the model and include them. If you do this, we suggest using the cross product of longitude and latitude as the study area follows a diagonal line. Rotating the coordinates so that the rotated coordinates are either horizontal or vertical is another option.

The data are imported with the `read.csv` function. There is some R coding involved to get the years and month of sampling, but that is not shown here.

```
> CR <- read.csv(file = "CoralDisease.csv",
                 header = TRUE)
```

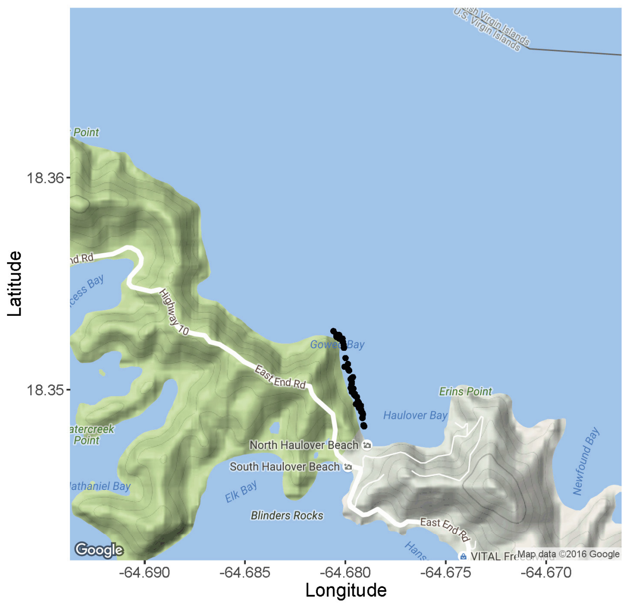


Figure 16.1. Position of 69 sampling locations.

16.2 Bernoulli model in R-INLA

Whenever we teach statistics courses on Bernoulli and binomial GLMs there is always some confusion. We therefore clarify the difference between a Bernoulli model and a binomial model before presenting the model for the coral data. If we toss a coin N times, and each time we record whether we have a head (which we label as success), and if each toss is independent, and if each toss has the same probability of a head, then we can use a binomial distribution for the number of heads (Y) out of N tosses. We can write this as $Y \sim B(\pi, N)$. If we only toss the coin once, then the number of trials is 1 and we use a Bernoulli distribution, which