23 Coral reef data and spatial barrier models

Alain F Zuur, Elena N Ieno, Adel Heenan, Ivor D Williams

In this chapter we will analyse data on flora and fauna associated with a living coral reef. These data were gathered for a long-term ecosystem monitoring program by professional divers who make visual observations of coverage of the coral reef benthos (seafloor) within cylindrical survey plots with a diameter of 15 m.

In most publications percentage coverage data are analysed using a linear regression model with a normal distribution, but that is incorrect as coverage data are always between 0% and 100% (or between 0 and 1), whereas the linear regression model may produce values outside this range. The correct model for such data is a beta distribution with, for example, a logistic link function. The beta distribution tends to be unfamiliar territory for most scientists so we provide a short introduction.

The divers also sample potential explanatory variables such as depth and information on the fish assemblage (e.g. species, number and size of all fishes observed within the survey plots). Fish biomass is derived from length and species-specific length-weight conversion parameters. Biomass has a non-linear relationship with nearly all benthic variables that we considered; hence we are in the world of generalised additive models (GAM) with a beta distribution.

We will use data collected from reefs around the islands of Tutuila and Aunu'u, which are part of American Samoa in the Southern Pacific Ocean. We have about 400 sites around these two islands, and the location of these sampling sites can be close to one another. This means that there is also a spatial element to the analysis.

For the available data set, the benthic species are grouped in various functional groups. Some groups have high coverage but one group has more than 30% of zeros. Hence, there is also a zero-inflation element.

Besides the beta distribution, non-linear relationships, zero inflation and spatial correlation, there is one more point that needs attention. The sampling locations are situated *around* two islands, and these islands act as barriers for benthic species. Two sampling sites may be separated by only 1 km in 'as the crow flies' distance, but a benthic species does not fly (unless it is in the belly of a bird). We will employ so-called barrier models to ensure that spatial correlation does not cover land.

The above paragraphs read as if this is a difficult chapter, but it is not. In fact, it is relatively simple.