

17 Introduction to Volume II

In 2013 we started our journey into the world of R-INLA. During the last 5 years the development of R-INLA has made great strides, and our enthusiasm for this tool has increased commensurately. In 2016 we wrote Volume I of *Beginner's Guide to Spatial, Temporal and Spatial-Temporal Ecological Data Analysis with R-INLA*, which was published in 2017. The title of that volume is *Using GLM and GLMM*.

Between 2012 and 2016 we also wrote two non-INLA books on the analysis of zero-inflated data (Zuur et al. 2012a; 2016), and two books on generalised additive (mixed) models (Zuur 2012b; Zuur et al. 2014). In these four books we presented a large number of analyses of real data sets. But there were always data sets with spatial or spatial-temporal dependency that we couldn't properly analyse due to lack of good statistical tools. That is no longer the case, thanks to R-INLA!

In Volume I we introduced R-INLA, but we were only able to cover regression models, generalised linear models (GLM), linear mixed effects models and generalised linear mixed effects models (GLMM) in combination with spatial, temporal and spatial-temporal dependency. In this volume, we go one step further and extend the regression models, GLMs and GLMMs with smoothers and/or zero-inflation components. We also discuss how to deal with natural barriers. In the last chapter we explain how to analyse areal and lattice data.

The title of this volume is therefore *Using GAMs and Zero-Inflated Models*. The prerequisite for this volume is all the material from Volume I.



Upon reading this volume you will be able to implement GLMs, GLMMs, GAMs and GAMMs with spatial correlation, spatial-temporal correlation and zero inflation in R-INLA.

17.1 Outline of Volume II

In Chapter 18 we will explain how to deal with zero-inflated data. We introduce so-called zero-inflated Poisson (ZIP) models, zero-inflated negative binomial (ZINB) models, zero-altered Poisson (ZAP) models and zero-altered negative binomial (ZANB) models.

In Chapter 19 we extend the ZIP, ZINB, ZAP and ZANB models with spatial correlation. Both these chapters use a skate data set from South America. In the appendix accompanying Chapter 19 we also explain how to manipulate maps and create spatial polygons (e.g. for coastlines).

In Chapter 20 we revisit a data set with which we have been battling since 2006. It is about begging behaviour of owl nestlings. In Zuur (2009a) we applied linear mixed effects models on it, and in Zuur et al.

(2012a) we analysed it with a zero-inflated GLMM. Thanks to R-INLA we finally cracked this data set and apply a zero-inflated GAMM.

In Chapter 21 we analyse sandeel count data. This work came out of a consultancy project that we carried out for Wageningen Marine Research (The Netherlands) in 2017. Although the setup of the experiment is simple (approximately 400 sites sampled once per year, for 4 years), analysing these data and writing this chapter took about 30 days. This should give you an idea about the complexity of the statistical tools (zero-inflated GAMMs + spatial-temporal correlation) that we discuss in this book.

Chapter 22 is about zero-inflated bird densities sampled in the Labrador Sea, located between the Labrador Peninsula (Eastern Canada) and Greenland. This chapter is about the analysis of zero-inflated continuous data with spatial correlation. A zero-altered gamma model with spatial correlation is used.

In Chapter 23 we analyse coral reef data sampled around an island. A lot of misery comes together in this chapter: smoothers, zero-inflation and spatial dependency that should not cross land as benthic species that live in a coral reef do not walk over land! We will discuss barrier models (Bakka et al. 2018) which ensure that spatial correlation seeps around a barrier (in this case an island).

Up to Chapter 23 all data sets analysed were geostatistical data and not areal or lattice data. The reason for this is that most ecological data is geostatistical. In Chapter 24 we analyse aggregated tornado data in 102 counties in Illinois. This is areal data. We will use various CAR models (e.g. iCAR, BYM, BYM2) for zero-inflated spatial and spatial-temporal correlated data.

In all our books we use real data sets, and occasionally we simulate some data sets to illustrate something. For most data sets this has typically resulted in relatively short chapters of approximately 20 pages. The data sets used in this volume are far more challenging than anything we have dealt with previously. As a result, the chapters are considerably longer, typically 40–50 pages. Initially, we tried to shorten the text by omitting R code, but this didn't work. The (shorter) text without documented R code became text in which the reader quickly lost track of the plot. We therefore added the R code again. This does mean longer chapters, but because seeing is believing, it clarifies the material.

17.2 Availability of the R code and data

All R code and data sets are available on the website for this book. We explain how to open the R files in the preface of Volume II.