

Olivier Gimenez

***Bayesian Analysis of
Capture-Recapture Data with
Hidden Markov Models***
Theory and Case Studies in R



Contents



List of Figures



List of Tables



Welcome

Welcome to the online version of the book *Bayesian Analysis of Capture-Recapture Data with Hidden Markov Models – Theory and Case Studies in R*.

The HMM framework has gained much attention in the ecological literature over the last decade, and has been suggested as a general modelling framework for the demography of plant and animal populations. In particular, HMMs are increasingly used to analyse capture-recapture data and estimate key population parameters (e.g., survival, dispersal, recruitment or abundance) with applications in all fields of ecology.

In parallel, Bayesian statistics is well established and fast growing in ecology and related disciplines, because it resonates with scientific reasoning and allows accommodating uncertainty smoothly. The popularity of Bayesian statistics also comes from the availability of free pieces of software (WinBUGS, OpenBUGS, JAGS, Stan, NIMBLE) that allow practitioners to code their own analyses.

This book offers a Bayesian treatment of HMMs applied to capture-recapture data. You will learn to use the R package NIMBLE which is seen by many as the future of Bayesian statistical ecology to deal with complex models and/or big data. An important part of the book consists in case studies presented in a tutorial style to abide by the “learning by doing” philosophy.

I’m currently writing this book, and I welcome any feedback. You may raise an issue here¹, amend directly the R Markdown file that generated the page you’re reading by clicking on the ‘Edit this page’ icon in the right panel, or email me². Many thanks!

¹<https://github.com/oliviergimenez/banana-book/issues>

²<mailto:olivier.gimenez@cefe.cnrs.fr>

Olivier Gimenez. Written in Montpellier, France and Athens, Greece.
Last updated: August 11, 2023

License

The online version of this book is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License³.

The code is public domain, licensed under Creative Commons CC0 1.0 Universal (CC0 1.0)⁴.

³<http://creativecommons.org/licenses/by-nc-nd/4.0/>

⁴<https://creativecommons.org/publicdomain/zero/1.0/>

Preface

Why this book?

To be completed. Why and what of capture-recapture data and models, with fields of application.⁵ Brief history of capture-recapture, with switch to state-space/hidden Markov model (HMM) formulation. Flexibility of HMM to decompose complex problems in smaller pieces that are easier to understand, model and analyse. From satellite guidance to conservation of endangered species. Why Bayes? Also three of my fav research topics – capture-recapture, HMM and Bayes statistics – let's enjoy this great cocktail together.

Who should read this book?

This book is aimed at beginners who're comfortable using R and write basic code (including loops), as well as connoisseurs of capture-recapture who'd like to tap into the power of the Bayesian side of statistics. For both audiences, thinking in the HMM framework will help you in confidently building models and make the most of your capture-recapture data.

⁵Watch out nice Johnny Ball's video <https://www.youtube.com/watch?v=tyX79mPm2xY>.

What will you learn?

The book is divided into five parts. The first part is aimed at getting you up-to-speed with Bayesian statistics, NIMBLE, and hidden Markov models. The second part will teach you all about capture-recapture models for open populations, with reproducible R code to ease the learning process. In the third part, we will focus on issues in inferring states (dealing with uncertainty in assignment, modelling waiting time distribution). The fourth part provides real-world case studies from the scientific literature that you can reproduce using material covered in previous chapters. These problems can either i) be used to cement and deepen your understanding of methods and models, ii) be adapted for your own purpose, or iii) serve as teaching projects. The fifth and last chapter closes the book with take-home messages and recommendations, a list of frequently asked questions and references cited in the book. **Likely to be amended after feedbacks.**

What won't you learn?

There is hardly any maths in this book. The equations I use are either simple enough to be understood without a background in maths, or can be skipped without prejudice. I do not cover Bayesian statistics or even hidden Markov models fully, I provide just what you need to work with capture-recapture data. If you are interested in knowing more about these topics, hopefully the section Suggested reading at the end of each chapter will put you in the right direction. There are also a number of important topics specific to capture-recapture that I do not cover, including closed-population capture-recapture models (?), and spatial capture-recapture models (?). These models can be treated as HMMs, but for now the usual formulation is just fine. **There will be spatial considerations in the Covariates chapter w/ splines and CAR. I'm not sure yet about SCR models (R. Glennie's Biometrics paper on**

HMMs and open pop SCR will not be easy to Bayes transform and implement in NIMBLE).

Prerequisites

This book uses primarily the R package NIMBLE, so you need to install at least R and NIMBLE. A bunch of other R packages are used. You can install them all at once by running:

```
install.packages(c(
  "magick", "MCMCvis", "nimble", "pdftools",
  "tidyverse", "wesanderson"
))
```

Acknowledgements

To be completed.

How this book was written

I am writing this book in RStudio⁶ using bookdown⁷. The book website⁸ is hosted with GitHub Pages⁹, and automatically updated after every push by Github Actions¹⁰. The source is available from GitHub¹¹.

The version of the book you're reading was built with R version 4.2.3 (2023-03-15) and the following packages:

package	version	source
magick	2.7.4	CRAN (R 4.2.0)
MCMCvis	0.15.5	CRAN (R 4.2.0)
nimble	0.13.1	CRAN (R 4.2.0)
pdftools	3.3.3	CRAN (R 4.2.0)
tidyverse	2.0.0	CRAN (R 4.2.0)
wesanderson	0.3.6	CRAN (R 4.2.0)

⁶<http://www.rstudio.com/ide/>

⁷<http://bookdown.org/>

⁸<https://oliviergimenez.github.io/banana-book>

⁹<https://pages.github.com/>

¹⁰<https://github.com/features/actions>

¹¹<https://github.com/oliviergimenez/banana-book>

About the author

My name is Olivier Gimenez (<https://oliviergimenez.github.io/>). I am a senior (euphemism for not so young anymore) scientist at the National Centre for Scientific Research (CNRS) in the beautiful city of Montpellier, France.

I struggled studying maths, obtained a PhD in applied statistics a long time ago in a galaxy of wine and cheese. I was awarded my habilitation (<https://en.wikipedia.org/wiki/Habilitation>) in ecology and evolution so that I could stop pretending to understand what my colleagues were talking about. More recently I embarked in sociology studies because hey, why not.

Lost somewhere at the interface of animal ecology, statistical modeling and social sciences, my so-called expertise lies in population dynamics and species distribution modeling to address questions in ecology and conservation biology about the impact of human activities and the management of large carnivores. I would be nothing without the students and colleagues who are kind enough to bear with me.

You may find me on Twitter (<https://twitter.com/oaggimenez>), GitHub (<https://github.com/oliviergimenez>), or get in touch by email¹².

¹²<mailto:olivier.gimenez@cefe.cnrs.fr>



Part I

I. Foundations





Introduction



Part II

II. Transitions



Introduction



1

Survival

1.1 Introduction

In this fourth chapter, you will learn about the Cormack-Jolly-Seber model that allows estimating survival based on capture-recapture data. You will also see how to deal with covariates to try and explain temporal and/or individual variation in survival. This chapter will also be the opportunity to illustrate how to incorporate prior information to improve model inference.

1.2 The Cormack-Jolly-Seber (CJS) model

In chapter ??, we introduced a capture-recapture model with constant survival and detection probabilities which we formulated as a HMM and fitted to data in NIMBLE. Historically, however, it was a slightly more complicated model that was first proposed – the so-called Cormack-Jolly-Seber (CJS) model – in which survival and recapture probabilities are time-varying. This feature of the CJS model is useful to account for variation due to environmental conditions in survival or to sampling effort in detection. Schematically the CJS model can be represented this way: