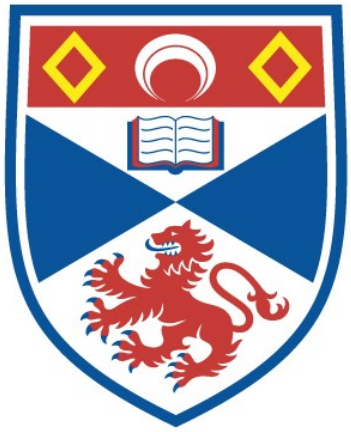
Expanding the use of spatial models in statistical ecology

Andrew Ernest Seaton

A thesis submitted for the degree of PhD
at the
University of St Andrews



2022

Candidate's declaration

I, Andrew Ernest Seaton, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 51,609 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2017.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date Signature of candidate

06/06/2021

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date Signature of supervisor

06/06/2021

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Andrew Ernest Seaton, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date Signature of candidate

06/06/2021

Date Signature of supervisor

06/06/2021

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Andrew Ernest Seaton, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date Signature of candidate

06/06/2021

Abstract

This thesis is focused on expanding the use of spatial modelling approaches for applications in ecology. Spatial ecology is about understanding the processes that give rise to spatial patterns in ecological data. In addition to developing a purely scientific understanding, insights into these processes are essential for the effective monitoring and conservation management of ecological systems.

However, for many ecological problems, the detectability of animals is imperfect, requiring the use of complex observation models that can account for this. In this thesis we focus on two such models: distance sampling and spatial capture-recapture (SCR). For both these models we incorporate spatially structured random effects to provide a non-parametric method for describing spatial variation in species' abundance, and to address the problem of spatial auto-correlation.

These complex models require the use of computationally efficient random effect structures and inference methods. In particular, we use a sparse stochastic partial differential equation (SPDE) approach as well as low rank penalised smoothing splines. We also draw links between these two approaches in order to illuminate the technically challenging results underpinning the SPDE approach. For inference in distance sampling models, we use a novel approach to achieve a one-stage model fit based on iterated model fitting using approximate Bayesian methods. For inference in SCR models, we use Laplace approximate maximum likelihood methods. We present models that have the necessary complexity to jointly model complex ecological and observation processes, as well as providing efficient methods to fit the models in practice. We conclude by discussing related avenues for future research that are motivated by applied problems in the field of spatial ecology.

Contents

1 Introduction **1**

1.1 Spatial modelling 8

1.1.1 The challenge of spatial auto-correlation 8

1.1.2 Spatially structured random effects 9

1.2 Point process models 14

1.3 Distance sampling 15

1.4 Spatial Capture-Recapture 16

1.5 Inference for spatial models 17

1.5.1 Integrated nested Laplace approximations 18

1.5.2 Laplace approximate maximum likelihood 20

1.6 A note on notation 20

1.7 Summary 21

2 Understanding the stochastic partial differential equation approach to smoothing **22**

2.1 Introduction 22

2.2 Basics of spline regression 24

2.2.1 Examples of basis-penalty smoothers 32

2.3 Basics of Gaussian random fields 34

2.4 Basics of stochastic partial differential equations 38

2.4.1 Solving the SPDE using finite element methods 42

2.5 The SPDE approach for the Matérn field 46
2.5.1 The connection with the smoothing penalty 48
2.5.2 Implementation in `mgcv` 52
2.6 Examples 54
2.6.1 Chlorophyll A in the Aral Sea 55
2.6.2 Coal Tit data in England and Wales 58
2.6.3 Bering Sea trawl data 60
2.7 Discussion 62

3 One-stage point transect distance sampling using iterated integrated nested Laplace approximations 64

3.1 Introduction 64
3.2 Study design 68
3.2.1 Study area and survey design 69
3.3 Distance sampling as a thinned point process 71
3.3.1 Overview of distance sampling methods 71
3.3.2 Model specification 75
3.3.3 Writing the model as a modified Poisson likelihood 77
3.3.4 Intensities for incomplete data 80
3.4 Iterated INLA 80
3.5 Spatially Structured Random Effect 82
3.6 Results 83
3.7 An alternative approach to predicting abundance 89
3.8 Communication of Results 93
3.8.1 Limitations of mapped summaries 94
3.8.2 Excursion sets and excursion functions 97
3.8.3 Estimating the information gained by sampling 100
3.9 Discussion 104

4 Flexible density models for spatial capture-recapture 108

4.1 Introduction . . . 110

4.2 Literature review . . . 113

4.3 The spatial capture-recapture model . . . 125

4.4 Inference using the Laplace approximation . . . 130

4.4.1 A brief introduction to automatic differentiation . . . 132

4.5 Basis-penalty random effects . . . 133

4.5.1 Low rank smoothers . . . 133

4.6 Low rank vs sparse effects simulation study . . . 138

4.6.1 Thin-plate regression splines simulation . . . 139

4.6.2 Matérn field simulation . . . 145

4.6.3 A note on integration schemes with the linear B-spline basis . 150

4.6.4 A note on prediction with Matérn field vs TPRS . . . 152

4.7 Case study: Louisiana black bear data . . . 153

4.8 Effects that account for complex spatial geometry . . . 158

4.8.1 Horseshoe simulation . . . 160

4.8.2 Case study: Boland leopard data . . . 165

4.9 Discussion . . . 170

5 Discussion 179

5.1 Chapter 2: Understanding the stochastic partial differential equation approach to smoothing . . . 180

5.1.1 Building mechanism into species distribution models . . . 180

5.2 Chapter 3: point transect distance sampling using iterated integrated nested Laplace approximations . . . 183

5.2.1 Functional responses . . . 184

5.2.2 Other detection models . . . 185

5.2.3 Parametric spatio-temporal dependencies . . . 187

5.2.4 A word of warning . . . 187

5.3 Chapter 4: Flexible density models for spatial capture-recapture . . . 188
5.3.1 One dimensional splines . . . 188
5.3.2 Alternative point process models . . . 189
5.3.3 Random effects on other components in the SCR model . . . 190
5.3.4 Alternative low rank methods . . . 191
5.4 Conclusions . . . 191

Appendices **193**

A Derivatives of the approximate SCR log likelihood **194**
A.1 First- and second-order derivatives . . . 194
A.2 A bespoke Laplace approximation implementation . . . 197

B Saturating functional responses using `inlabru` **200**

Acknowledgements

I am deeply indebted to my three supervisors: Janine Illian, David Borchers and Richard Glennie. When I began this journey I was relatively inexperienced in statistics and they showed me nothing but understanding and support as we worked together on this research. I am particularly thankful to Janine, who saw something in me when we first met, despite my inexperience in statistics, and agreed to take me on. From each of my supervisors I have learned a lot about how to approach statistical research but also, more importantly, how to be a kind, thoughtful and supportive colleague while doing this. I could never have completed this work without them.

I also consider myself to have two unofficial supervisors. David Miller has been a constant source of support, is always happy to chat about any topic, and has numerous times saved me from the depths of the debugging blues. Working with David taught me about how to do collaborative statistical research in an enjoyable and fruitful way and he has become a true friend. Finn Lindgren, my other unofficial supervisor, gave absolutely essential support for me when I was working on all the substantive chapters in this thesis. He provided endless support when I always seemed to have more questions about methods, software implementations and ideas for future research. I could not have completed this research without Finn's thoughtful and patient advice.

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) funding council. For each main chapter of my thesis I would like to thank, in addition to my supervisors and the EPSRC, the following people:

Chapter 2: Understanding the stochastic partial differential equation approach to smoothing

Richard Glennie and David Miller agreed to start our small reading group that began this work and took me under their wing to write a paper together on this work. Richard also agreed to become a supervisor following this. What began just for our own understanding soon developed into a piece of collaborative research. This was my first experience working in this way and it helped me enormously in my later research.

Finn Lindgren met with us for an afternoon to discuss this work, which resulted in substantial improvements to a draft paper which was later accepted for publication. Simon Wood also commented on a draft paper and suggested the improved method for parameterising the model in `mgcv`.

Chapter 3: Point transect distance sampling with iterated integrated nested Laplace approximations

Rick Camp of the United States Geological Survey and the University of St Andrews introduced me to the Hawai'ian forest bird survey and the wonderful 'ākepa, the endangered tropical bird that provided the case study for this work. We had many invaluable conversations over the years about this work, discussing the survey protocol, previous work analysing these data, and the conservation ecology of this species. The U.S. Fish and Wildlife Service (USFW), Hakalau Forest National Wildlife Refuge, and Steve Kendall, the Wildlife ecologist for USFW at Hakalau, were instrumental in collecting the data and making it publicly available.

David Miller discussed this work with me many times and helped to debug some of the many computational issues I ran into over the course of the research. Finn Lindgren provided relentless support to implement the model using his software, including quickly implementing bugfixes for errors we encountered, provided the useful advice on integration schemes for random fields, and suggested the information gain

metrics we considered for the posterior intensity field. Steve Buckland provided helpful comments on an early draft.

Chapter 4: Flexible density models for spatial capture-recapture

Ben Stevenson at the University of Auckland discussed his related work with me and we had many conversations about spatial capture-recapture (SCR) and random effects in general. He also helped me with coding in `C++`, suggesting more efficient implementations of the SCR likelihood, and for letting me pilfer many useful ideas from his existing `C++` code. Timo Adam at the University of St Andrews, also shared his SCR `C++` code with me, which led to some simplifications and improvements to own my code.

R. Chandler, J. D. Clark, C. Lowe, and K.O'Connell provided the Louisiana Black Bear data for a case study. The Cape Leopard Trust, Jeannie Hayward, and Anita Wilkinson provided the Boland leopard data for another case study. This data collection was partly funded through grants to the Cape Leopard Trust from the Hans Hoheisen Charitable Trust, Deutsche Bank Africa Foundation, and Leopard's Leap Wines. Thanks also to all the landowners in the Boland Mountain catchment and CapeNature for providing support and enabling access to their properties for the data collection.

Finn Lindgren shared his knowledge of potential pitfalls when choosing integration schemes for Gaussian random fields and met with me to discuss implementing his methods in SCR. David Miller met with me to discuss implementing his generalised smoothing spline method for complex spatial geometry and pointed me to the relevant parts of his code to implement this myself.

I would also like to thank everybody at the Centre for Research into Environmental and Ecological Modelling (CREEM) at the University of St Andrews for providing

such a welcoming environment for statistical research. The culture at CREEM is special and I am deeply thankful for everybody who has contributed to this, both past and present and, hopefully, into the future. Rhona Rodger and Phil Le Feurve deserve special mention for their patient and constant administrative support.

Thank you to my office mates over the years: Claudia Faustino, Filippo Franchini, Gui Bortolotto and Virginia Morera Pujol, for making our room such a lovely place to work. My thanks also to everyone who has attended the spatial statistics reading group over the years, too many to name here. It was one of the most enjoyable things about my time at CREEM. I learned so much and I'm so glad I decided to start the group. For me, the group represents the best of CREEM: a group of friendly people, open to all comers, working collectively, without ego, judgement or competitiveness, to better understand our craft.

None of this work would have been possible without the support of friends and family. I give my heartfelt thanks to Calum White and Roan Du Feu, both of whom I am unbelievably lucky enough to call my best friends since childhood. Of course they are much more efficient than me and have been quick to point out that they already have their doctorates. I thank them for for the teasing, it was a huge source of motivation for me. They are both inspiring to me and without their example I might not have considered this a valid path for me.

I am hugely indebted to my loving parents, Roxana and Nigel Seaton, for their constant support. They supported me when I decided to make this career change to statistical ecology, which has allowed me to stumble into this wonderful academic community that has brought me so much joy and purpose. They gave, and continue to give, so much of themselves to support our young family, taking time off work and changing career plans to support us through difficult times. I am also indebted to my brothers Daniel and Mark whom I love and who have, each in their own way, been role models for me. All the lessons, explicit and implicit, throughout my life that I learned from my parents and brothers are the reason I was able to complete

a PhD while raising a young family (during a pandemic!). Any resilience I have, I inherited from them.

I must also give enormous thanks to all my out-laws: Karen, David, Susan, Daniel, and Anne for everything they have done to support our family. Karen, David and Anne took us into their home during the darkest days of the pandemic and supported us, looking after our children so we could continue to work. Without their support we never would have made it through and we can never thank you enough.

All the energy and effort I put into this work is an extension of my love for my children, Felix and Mabel. The whole purpose of taking this path was to attempt to find a meaningful career that I can be proud of. This is not for myself really (although that is a bonus), but I was moved to do this so that I can be a happy, engaged and present father. It may not always have seemed this way, especially during the hardships of the past year, but I am much happier doing this than I would be doing almost anything else. I hope I can pass this happiness on to my children.

Finally, I give my deepest thanks of all to my partner Helen Traill. In 2019, when we were deciding whether to grow our family, I said, with reckless optimism, "There's never a good time, let's just go for it." We have since learned that some times are better than others. Having a baby during a pandemic, with on-again, off-again childcare, in a small two bedroom flat with no garden, with your partner camped at one end of the kitchen table suffering over his code and his thesis, is, in hindsight, less than ideal circumstances. I've dealt out more than my fair share of anxiety and stress over this past year and Helen has absorbed it all. Her love and support is the only reason I have reached the finish line with (I hope) my sanity intact. Without her, none of this would have been possible.

List of software

It is remarkable that I was able to complete this research exclusively using freely available software, most of it open-source. This section lists the software used to produce the results in this thesis. Although all these tools themselves depend on other software packages, to keep things simple I mention only software that I used directly and not their dependencies. I also omit software I used during exploratory work that did not end up directly contributing to the outputs in this thesis.

I used Ubuntu (versions 18.04 and 20.04), an open-source operating system using the Linux kernel and based on Debian, on my laptop and desktop machines as well as the Linux server hosted by the School of Mathematics and Statistics at the University of St Andrews.

The vast majority of the programming took place in `R` (R Core Team, 2017). To interface with `R`, I used `RStudio` (RStudio Team, 2020), `vim` (Moolenaar, 2019), and the `Nvim-R` plug-in (Alves de Aquino, 2021). The following `R` packages were used in the process of producing this thesis:

`dplyr` (Wickham et al., 2021)

`excursions` (Bolin and Lindgren, 2018, 2015)

`gdistance` (van Etten, 2017)

`ggmap` (Kahle and Wickham, 2013)

`ggplot2` (Wickham et al., 2020)

`here` (Müller, 2020)

`INLA` (Rue et al., 2009)

`inlabru` (Bachl et al., 2019)

`maptools` (Bivand and Lewin-Koh, 2020)

`MASS` (Ripley, 2021)

`mgcv` (Wood, 2017)

`patchwork` (Pedersen, 2020)

`RandomFields` (Schlather et al., 2015)

`raster` (Hijmans, 2020)

`rgdal` (Bivand et al., 2021)

`rgeos` (Bivand and Rundel, 2020)

`scales` (Wickham and Seidel, 2020)

`secr` (Efford, 2021)

`sf` (Pebesma, 2018)

`smoothr` (Strimas-Mackey, 2020)

`sp` (Pebesma and Bivand, 2021)

`tidyr` (Wickham, 2020)

`TMB` (Kristensen et al., 2016)

This thesis was typeset using LaTeXand TeXstudio (van der Zander et al., 2021).

Chapter 1

Introduction

This thesis is about understanding, developing and applying statistical methods to answer questions in spatial ecology. Spatial ecology is about understanding the processes that give rise to the spatio-temporal patterns that occur in ecological data. By building statistical models that provide a possible explanation for these patterns, inverse inference from observed data provides a means to generate knowledge about these hidden processes. Spatial ecologists want to learn about these processes for many reasons, from conservation management (Porfirio et al., 2014; Franklin, 2010b; Corsi et al., 1999) to biodiversity discovery (Williams et al., 2009; Guisan et al., 2006), predicting responses to climate change (Jarvie and Svenning, 2018; Iverson et al., 2011; Van der Putten et al., 2010) and more. The challenge for statisticians is to help ecologists make sense of these data with appropriate statistical models that can capture potentially rich and complex ecological processes. However, methods must also account for for non-trivial observation processes, requiring complex hierarchical models and many modelling decisions that have to be made along the way. This thesis presents contributions to each step of the applied statistician's workflow in addressing these challenges, from model conception and implementation through to inference method and communication of results.

Ecological data are naturally structured through space and time, with observa-

tions made closer together more similar than those made further apart. We cannot directly observe all of the animals all of the time, necessarily requiring sub-sampling in space and time as well as methods to account for imperfect detectability. The structure in the data we collect reflects the manner in which we made the observations as well as the ecological processes that are of scientific interest. Statistical methods must account for both the observation process and the, often latent, ecological processes that generate the observed data.

Within the field of spatial ecology there is a rich literature describing models for spatio-temporal data, which typically consist of the location and time a species was seen and other potentially relevant factors such as the local weather, climate and environmental conditions. Approaches to model these data vary from explicit mechanistic models (Bauer and Klaassen, 2013; Leroux et al., 2013; Rastetter et al., 2003) to purely correlative approaches such as machine learning (Rammer and Seidl, 2019; Christin et al., 2019) and everything in between.

The methods we use to observe animals also vary enormously. We might undertake a complete census of every tree species in a pre-defined plot, place a GPS device on an animal, or walk a pre-defined route through a nature reserve and note down everything we see, aware that we will probably miss a lot. We could leave motion-detecting cameras across the landscape and investigate the photographs later or we could ask people to download an application to their phone and note when they see particular species as they go about their daily lives. Each of these methods result in different data structures and therefore requires bespoke statistical models that can account for this. In statistical ecology we have to put as much effort into modelling the *observation process* as we do for modelling the *ecological process* we are interested in. Models for the observation process are often as rich and complex those for the ecological process. This thesis focuses on two observation models that have become a staple of monitoring programmes around the world: *distance sampling* (Buckland et al., 2015, 2004) and *spatial capture-recapture* (SCR) (Borchers

and Efford, 2008; Royle and Young, 2008).

Both these approaches were primarily developed with the aim of estimating the abundance of wild populations of animals. At the heart of distance sampling is the idea that animals further away from an observer are less likely to be detected. By estimating the number of animals that remain undetected, this approach allows us to estimate the total population size. SCR is based on a similar idea but also leverages the data where individual animals are identifiable, such as a leopard by the pattern of spots on its coat or by DNA analysis of hair or fecal samples.

In the years since their initial development, both distance sampling and SCR have been extended in various ways to incorporate complex spatio-temporal models for the distribution of animals across the landscape (Glennie et al., 2019; Yuan et al., 2017; Buckland et al., 2016; Borchers et al., 2014; Miller et al., 2013; Hedley and Buckland, 2004; Marques and Buckland, 2003). This thesis is another effort in this tradition. The benefit of these developments is that they allow the use of distance sampling and SCR data to generate spatially-explicit estimates of animal density. This allows ecologists and conservationists to investigate the possible causes of variation in density, such as responses to disturbance or conservation efforts, and to detect whether the distribution is changing over time.

A key tool in spatio-temporal modelling is the use of *spatially structured random effects* (see Section 1.1.2). These are flexible model components that can provide a non-parametric description of spatial patterns in data. Including these effects in models provides a way to address unexplained heterogeneity in the data that other model components cannot explain. Spatially structured random effects therefore act as insurance policy against unexplained spatial autocorrelation (Section 1.1.1) which can lead to bias and over-confidence in the explanatory power of other model components, such as the effects of spatial covariates. They also provide a means of visualising patterns in the data which can, in turn, generate new ecological hypotheses and spur future model development. They are, therefore, a fundamental

tool in spatial statistics.

In this thesis, we, for the first time, incorporate spatially structured random effects into an SCR model, allowing the use of flexible non-parametric models to describe spatial variation in animal abundance. We also incorporate spatially structured random effects in distance sampling models and present a method for *jointly* estimating the detectability and spatial distribution of animals from distance sampling data, a modelling process that is often split into two separate models. Miller et al. (2013) provides a summary of the two-stage approach, based on using Generalised Additive Models (GAMs) with an offset term to account for detectability estimated in a separate model. Royle et al. (2004) introduced a one-stage maximum likelihood approach based on discretising counts and distances into discrete classes that allows a multinomial formulation of the likelihood.

However, our work takes a Bayesian perspective and builds directly on Yuan et al. (2017), taking a point process view on distance sampling data, creating a single model that does not require observations to be binned into discrete counts or distance classes. We simplify the observation model used in Yuan et al. (2017), allowing the use of standard parametric families of detection functions, as is common in most distance sampling analyses, instead of the stochastic partial differential equation approach taken there. This requires a novel approach to inference based on a first-order Taylor expansion of the detection function and an iterated model fitting procedure.

The first topic we present, however, is not directly devoted to flexible modelling of distance sampling or SCR data, but rather how to *understand* one particular type of spatially structured random effect that has become a popular choice in spatial ecology. This effect is based on a stochastic partial differential equation (SPDE) representation of a spatially correlated Gaussian random field (Lindgren et al., 2011). While this approach has become popular, the literature on this is dense and requires a deep mathematical understanding that many applied statisticians do

not have. However, for a model to be interpretable, a fundamental step of any modelling process is for the practitioner to understand the model they would like to fit. We attempt to simplify the broad and disparate technical literature required to understand the SPDE approach and package it up in a way that applied statisticians and quantitative ecologists can understand.

In Chapter 2, *Understanding the stochastic partial differential equation approach to smoothing*, we draw links between Gaussian random fields specified via SPDEs and optimal smoothing splines (Wood, 2017; Wahba, 1990). Smoothing splines have been used in distance sampling models to describe spatial variation in species density (Miller et al., 2013) and unpenalised spline effects with a fixed degree of freedom have been implemented in SCR in the `secr` package (Efford, 2021). We show how to implement and fit SPDE models using the widely used penalised smoothing spline package `mgcv` (Wood, 2017), drawing links between two popular approaches to model animal density in spatial ecology.

The main results of this chapter have been published in Miller et al. (2019), for which I am joint first author. Chapter 2 is a longer-form version of this paper with more details, some corrections and clarifications, and presents some new case studies. The ideas presented here form the basis of the implementation of the random effects models in the subsequent chapters.

In Chapter 3, *One-stage point transect distance sampling using iterated integrated nested Laplace approximations*, we show how to fit a spatial SPDE model within the context of the distance sampling observation process. We do this using a novel approach to inference that is based on considering iterative model fits, taking a Taylor approximation of non-linear model components. This approach allows us to jointly estimate the spatial distribution and detectability of animals, a workflow that is usually split into two separate models, and shares uncertainty between the observation and spatial models. We also discuss the challenges in communicating uncertainty in animal density maps based on posterior log-Gaussian random fields

and suggest some new approaches to consider by presenting novel ways to summarise posterior Monte Carlo samples of the random field. In particular, we introduce some new metrics to measure information gain due to sampling effort and suggest the use of excursion sets (Bolin and Lindgren, 2018, 2015) as an improvement over pointwise summaries for each prediction location.

In Chapter 4, *Flexible density models for spatial capture-recapture*, we introduce penalised spatially structured random effects for animal density within SCR for the first time. We implement the SPDE approach to model animal density in SCR as well as implementing other random effects such as thin plate regression splines (Wood, 2003) and effects that can account for non-Euclidean spatial geometry (Miller and Wood, 2014). Here we use Laplace approximate maximum likelihood for inference, using automatic differentiation and implemented in the R package `Template Model Builder` (`TMB`) (Kristensen et al., 2016).

In a sense, this thesis represents many of the key steps in an applied statistician's workflow, which can be summarised in the following steps:

1) Understand your tools.

2) Pick appropriate tools to model the data you have.

3) Pick an appropriate inference method that is well-suited to the tools you picked.

4) Fit some models.

5) Check the fitted models. Can steps 1-3 be improved?

6) Effectively communicate the results of the modelling.

In reality, we have to iterate this procedure many times and explore many dead ends. The research presented in this thesis examined each of these aspects of the workflow. Chapter 2 is about understanding the SPDE approach to smoothing; Chapters 3 and 4 consider two popular approaches to working with wildlife survey

data with imperfect detection, in both cases using novel approaches to inference driven by the particular models we considered; and Chapter 3 has a particular focus on communicating uncertainty in species distribution maps.

An alternative title for the thesis could perhaps have been *Contributions to the workflow of applied spatial ecology.* The title we chose in the end is *Expanding the use of spatial models in statistical ecology.* This 'expanded use' is intended to describe both the novel application of spatial modelling methods to new areas of spatial ecology, as well as our expanded understanding of how the models work and how to communicate results based on these models.

The main benefit to understanding the SPDE approach was, for us, that this allowed us to implement similar models ourselves, in software packages that we are familiar with, and can flexibly apply in a wide variety of contexts. This understanding allowed us to detach the SPDE approach from any single particular software implementation, using both `mgcv` and `TMB` to fit SPDE models presented in this thesis. We hope this helps other practitioners make use of this approach using whichever software they are most familiar with.

The attractiveness of spatially structured random effects is that they provide a flexible model that is able to capture correlations in data for which we have no good covariate or mechanistic model to explain. In spatial ecology, there is nearly always some structure in the data that we cannot explain without the use of some flexible random effect. These methods act as an insurance policy against over-confidence in other aspects of the model, such as the effects of covariates or mechanistic model components. Spatially structured random effects are a fundamental part of the spatial statistician's toolkit and they should be available to use in the complex hierarchical spatial models that ecologists tend to need. They will not always be needed in every analysis, but when they are they should be straightforward to implement with efficient methods for inference. This is the motivation for implementing these methods for distance sampling and SCR.

The remainder of this introduction introduces the key concepts and modelling frameworks that are used throughout the rest of the thesis. Where necessary, these ideas are expressed more precisely elsewhere in the thesis.

1.1 Spatial modelling

1.1.1 The challenge of spatial auto-correlation

In most spatial modelling applications, models tend to include a *predictor* component, denoted here as $\eta(\boldsymbol{s})$, that depends on a spatial location $\boldsymbol{s}$, and usually has a linear additive form such as

$$\eta(\boldsymbol{s}) = \beta_0 + \sum_k \beta_k z_k(\boldsymbol{s}),$$

with intercept parameter β_0 and the effects of spatially indexed covariates $z_k(\boldsymbol{s})$ with coefficients β_k.

Given this predictor we consider a generic model of the form

$$y(\boldsymbol{s}) \sim \pi(y(\boldsymbol{s})|\eta(\boldsymbol{s})),$$

for data $y(\boldsymbol{s})$ with likelihood $\pi(y(\boldsymbol{s})|\eta(\boldsymbol{s}))$. Given a particular choice of likelihood, it is in considering various models for $\eta(\boldsymbol{s})$ that we attempt to understand how the data $y(\boldsymbol{s})$ arises. In a hierarchical model, $\eta(\boldsymbol{s})$ itself may have a model $\pi(\eta(\boldsymbol{s}))$, which could be viewed as a prior in a Bayesian context, some form of regularization in a penalised regression context, or some represent another data generating mechanisms.

The usual assumption for such a model is that observed data $y_1, \ldots, y_n$ are conditionally independent given $\eta_1, \ldots, \eta_n$. The implicit justification for this assumption is that $\eta(\boldsymbol{s})$ is a good model, in that it has sufficient richness that it can explain enough variation in the observed data that conditional independence is a reasonable

assumption to make. Violating this assumption risks introducing bias and can lead to artificially small uncertainty estimates. If the data $y_1, \ldots, y_n$ have some correlation that $\eta_1, \ldots, \eta_n$ cannot explain, then this has to be accounted for through some other means or by changing the model for $\eta(\boldsymbol{s})$. This is known as the problem of spatial auto-correlation (Dormann et al., 2007; Cliff and Ord, 1970).

One key way to address this is by adding a flexible model component to $\eta(\boldsymbol{s})$ that can absorb this residual correlation. This flexible component should be spatially correlated since its purpose is to address unexplained spatial correlation in the data. Such a model component we call a *spatially structured random effect*.

1.1.2 Spatially structured random effects

We denote a generic spatially structured random effect by $f(\boldsymbol{s})$. Since its purpose is to model spatial structure in the data, this effect is indexed by $\boldsymbol{s}$, representing a location in space. When such a model component is included we have a predictor of the form

$$\eta(\boldsymbol{s}) = \beta_0 + \sum_k \beta_k z_k(\boldsymbol{s}) + f(\boldsymbol{s}).$$

Typically $f(\boldsymbol{s})$ has a mean of zero and so $\mathbb{E}[\eta(\boldsymbol{s})]$ is just the model for the covariate effects. The question then is what kind of functions $f(\boldsymbol{s})$ should be allowed? It should be something that is flexible (able to 'bend' to match the data) but not too flexible (exactly fit everything by bending too much). In this thesis we consider effects defined on a continuous spatial domain and, in this context, there are two main approaches to specifying $f(\boldsymbol{s})$: *Gaussian random fields* (GRFs) (Rasmussen and Williams, 2005) and *smoothing splines* (Wood, 2017; Wahba, 1990). We will see later that these approaches are not really distinct but rather are intimately linked with each other.

The term 'random effect' comes with some baggage since it can mean many different things. Here we just use it to mean this flexible model component $f(\boldsymbol{s})$, that

is spatially correlated but doesn't have a fixed structure like a spatial covariate does. Historically the term random effect has been used to describe model parameters that themselves follow a probability distribution, such as i.i.d Gaussian variables, where the probability distribution is used to represent some super-population from which observations are drawn. More recently, the term has been applied in spatial statistics to describe a model component that has a probability distribution that induces spatial correlation. In this context there is no implied super-population for this model. Here we used the term spatially structured random effect to describe both GRFs and smoothing splines, although the terminology is probably more common in GRF applications than smoothing splines. For a more detailed discussion of the different interpretations of the term 'random effects', see Hodges (2019).

Gaussian random fields

One way to think about how to include a spatially structured random effect, $f(\boldsymbol{s})$, is to view this model component as a random function. We say $f(\boldsymbol{s})$ is a member of a space of functions and specify a probability distribution on this space to control the 'bendiness' of $f(\boldsymbol{s})$. Gaussian random fields (GRFs) (Rasmussen and Williams, 2005) are one way of doing this. A distribution on the function space is represented by specifying that the random function considered at any finite collection of locations, $[f(\boldsymbol{s}_1), \ldots, f(\boldsymbol{s}_n)]^\intercal$, has a multivariate Gaussian distribution. This distribution depends on the *covariance function* for the GRF, that allows us to construct the covariance matrix for any given $[f(\boldsymbol{s}_1), \ldots, f(\boldsymbol{s}_n)]^\intercal$. The covariance function is, therefore, central to the behaviour of $f(\boldsymbol{s})$ as this what determines the correlations between locations and thus, informally, the 'bendiness' of $f(\boldsymbol{s})$.

Smoothing splines

A similar and related approach is to specify $f(\boldsymbol{s})$ as a smoothing spline (Wood, 2017; Wahba, 1990). Under this paradigm we also specify a family of allowable functions

for $f(\boldsymbol{s})$. We then *penalise* bendy functions to stop over-fitting. This is done by specifying a *smoothing penalty*. One example of a way to measure bendiness is to consider the second derivatives of $f(\boldsymbol{s})$ and using a penalty of the form

$$\lambda \int \left(\frac{\partial f}{\partial s_1^2}\right)^2 + \left(\frac{\partial f}{\partial s_2^2}\right)^2 \mathrm{d}\boldsymbol{s},$$

where $\boldsymbol{s} = (s_1, s_2)$. The integral 'adds up' the bendiness of $f(\boldsymbol{s})$, as measured by the second-derivative with respect to each coordinate variable. The parameter $\lambda > 0$ represents its cost in terms of how it impacts on the likelihood $\pi(\eta(\boldsymbol{s}))$. Similar to the covariance function above, the choice of smoothing penalty is central as it determines what counts as smoothness and thus what kinds of functions are preferred. A key message of Chapter 2 is that the *covariance* of a random field and the *smoothness* of a spline effect, as defined by the smoothing penalty, are intimately related.

Approximate smoothing methods

Both the above approaches depend on specifying a space of functions for $f(\boldsymbol{s})$. However, many function spaces, such as spaces of differentiable or integrable functions, are infinitely large and have an infinite number of basic 'building blocks'. For example, any real-valued integrable function can be represented by its Fourier series $f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n\theta}$. Here the infinite set of functions $\{e^{2\pi i n\theta} \mid n \in \mathbb{Z}\}$ are the basic building blocks from which we can construct any integrable function by using appropriate coefficients c_n. However, for reasons of practical computation we have to consider a *finite* set of building blocks for a function space, also called the *basis* of the function space. The challenge is to do this in a way that still leaves a sufficiently rich function space for a given application.

For both GRFs and smoothing splines we choose a finite basis for $f(\boldsymbol{s})$ and represent it as $f(\boldsymbol{s}) = \sum_{j=1}^{M} \xi_j \phi_j(\boldsymbol{s})$. The functions $\phi_1, \ldots, \phi_M$ are the basic building blocks with which we represent any $f(\boldsymbol{s})$ in the space of functions spanned by this

basis by varying the coefficients $\xi_1, \ldots, \xi_M$. As we explain in Chapter 2, given a choice of a basis for $f(\boldsymbol{s})$, both the GRF and smoothing spline approach amount to placing a multivariate Gaussian distribution on the coefficients $\boldsymbol{\xi} := [\xi_1, \ldots, \xi_M]^T$.

Given this finite basis, various approaches have been developed for computationally efficient methods to work with these effects since, if M is large, as can often be required in spatial applications, then $\boldsymbol{\xi}$ can have a large and dense covariance matrix $\boldsymbol{\Sigma}$ that can be computationally challenging to work with. This has led to numerous efforts to approximate $\boldsymbol{\Sigma}$ by various methods. This thesis considers two approaches: construct a sparse approximation to $\boldsymbol{\Sigma}$ (or equivalently its inverse, the precision matrix), or, alternatively, create a low-rank approximation of $\boldsymbol{\Sigma}$. The SPDE approach (Lindgren et al., 2011) is an example of a way to induce sparsity. The thin plate regression spline, a common choice and the default in `mgcv` (Wood, 2017), is an example of a low-rank approximation approach. We use both of these methods in this thesis.

The SPDE effect

The SPDE approach is a method to approximate a mean zero GRF with a Matérn covariance function. The Matérn covariance function defined on a two-dimensional domain can be written as

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{2^{1-\nu}}{4\pi\kappa^2\tau^2\Gamma(\nu)}(\kappa\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|)^\nu K_\nu(\kappa\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|), \tag{1.1}$$

where ν, κ, τ are parameters and K_ν is the modified Bessel function of the second kind. The three parameters are not simultaneously identifiable (Zhang, 2004) and it is conventional to assume a value for ν which specifies the mean-square differentiability of the process. In this thesis we set $\nu = 1$ in all applications which is a common choice and is the default in `R-INLA`, the most popular implementation of the SPDE effect.

The Matérn covariance function is a popular choice in spatial statistics for many reasons. Stein (2012) emphasises the additional flexibility of the differentiability parameter ν that many other covariance functions do not have and Lindgren et al. (2011) mention the importance of a spatial Markov property that allows for computational methods that can leverage this. The Matérn covariance function decays with increasing distance between locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ with a rate of decay that depends on κ and τ. It is, therefore, a flexible model that is well-suited to modelling spatial correlation that decays as distance between observations increases.

Lindgren et al. (2011) constructs an approximation to a GRF with Matérn covariance by representing the random field as the solutions to an SPDE. Given this representation, numerical methods for solving SPDEs can be applied and doing this in a particular way results in a sparse precision matrix for the coefficient vector $\boldsymbol{\xi}$. We cover this in detail in Chapter 2.

Thin plate splines

The thin plate spline penalty in two dimensions is

$$\lambda \int \int \left(\frac{\partial^2 f}{\partial s_1^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial s_1 \partial s_2} \right)^2 + \left(\frac{\partial^2 f}{\partial s_1^2} \right)^2 \mathrm{d}s_1 \mathrm{d}s_2,$$

which is similar to the penalty example given above except for the additional cross-derivatives term. The thin plate spline basis is dependent on data, with a radially symmetric basis function defined on every unique data point. This means the model can be over-parameterised. To fit a computationally efficient version of the thin plate spline, Wood (2003) suggests taking a low rank approximation by considering the eigen-decomposition of the precision matrix and truncating this by taking the eigenvectors with the largest eigenvalues. This is known as the thin plate *regression* spline (TPRS). When the truncation is done appropriately, this results in a random effect with fewer parameters to estimate, but $f(\boldsymbol{s})$ is still flexible enough for practical

purposes. This is the basis for the implementation of thin plate splines in the smoothing spline R package `mgcv` (Wood, 2017) and, for our purposes, the low rank approximation was useful in implementing spatially structured random effects for SCR.

1.2 Point process models

Both the applications considered in this thesis, distance sampling and SCR, can be viewed through the lens of point process models. A *point pattern* is a random set of locations that typically represent an event or physical object. The location of trees in a rainforest or the epicentres of earthquakes are examples of point patterns. A *point process* model describes, probabilistically, the likelihood of a point pattern under that model. In spatial ecology, often the points represent the locations of animals when they are observed. There is a rich literature on applied point process models across many fields, Baddeley et al. (2015) and Illian et al. (2008) provide an overview.

All of the point process models considered in this thesis are based on the *Poisson process*. The *homogeneous* Poisson process is a model for complete spatial randomness, where locations of points in the pattern are uniformly distributed independently of one another. The *inhomogeneous* Poisson process allows points to be structured spatially but still conditionally independent, given a latent surface that describes the density of points. This latent surface is known as the *intensity function*.

A *Cox process* is similar to the inhomogeneous Poisson process, but the latent surface is itself stochastic. In the case of a *log-Gaussian* Cox process, the log of the intensity function is a GRF (Møller et al., 1998). It is this model, the log-Gaussian Cox process, that we use within distance sampling and SCR by adding spatially structured random effects, such as GRFs and smoothing splines, to the log intensity

function.

A key property of the Poisson and Cox process is that these models handle the imperfect detection of points in a natural way. If a Poisson process model (homogeneous or inhomogeneous) has intensity function $\lambda(\boldsymbol{s})$ and points are detected with probability $p(\boldsymbol{s})$, independently of one another, then the pattern of *detected* points is itself distributed according to a Poisson process model with intensity function $p(\boldsymbol{s})\lambda(\boldsymbol{s})$. The same holds if $\lambda(\boldsymbol{s})$ is a random field. In other words, independent imperfect detection of points does not require a fundamental reformulation of the model. Instead, we just need to adjust the intensity function by considering $p(\boldsymbol{s})$. Sometimes this is referred to as independent *thinning* of the points and we use the term *thinning probability* interchangeably with *detection probability*. It is this feature of the Poisson point process and log-Gaussian Cox process that we use to incorporate point process models into distance sampling and SCR and have made them a popular tool in species distribution modelling.

In this thesis we use point processes to model data on animal locations. In this context, the term animal *density* (the number of animals per unit area) is common in the literature. For our purposes, this is the same as the intensity function of the point process and we use both terms in the text. This is not to be confused with a probability density function and should be clear from context.

1.3 Distance sampling

Distance sampling (Buckland et al., 2015, 2004) is a family of related methods for estimating the abundance and spatial distribution of wild populations of animals. Distance sampling is based on the idea that animals further away from observers are harder to detect than animals that are nearer. This idea is implemented in the model as a *detection function* that depends on distance. Species at greater distances are harder to detect and the detection function therefore declines as distance increases.

The shape of this function, the rate of decline, is estimated from observed data.

Distance sampling data has also been incorporated into spatial models (Hedley and Buckland, 2004; Miller et al., 2013) to produce maps of spatially varying density of animals. Usually this is done by using detectability point estimates to create an offset vector to use within a generalised linear model (GLM) (McCullagh and Nelder, 2019) or generalised additive model (GAM) (Wood, 2017) with a count response variable. This requires us to bin the data into counts based on some discretisation of space.

In this thesis we take an alternative point process perspective, using the detection function to define the probability of detecting points, $p(\boldsymbol{s})$, and fitting point process models to the data. We jointly estimate the probability of detection along with the intensity $\lambda(\boldsymbol{s})$, fitting a single model, and requiring no binning of the point data into counts. This work builds on Yuan et al. (2017), who take a similar point process perspective but formulate the detection function as a random effect using an SPDE. Here, we instead use a parametric detection function as opposed to a random effect, as is more common in distance sampling applications. Yuan et al. (2017) apply their method to *line transect* distance sampling data of blue whales based on shipboard surveys in the eastern Pacific. Here we formulate the model for *point transect* data, applied to distance sampling data on an endangered tropical bird. This requires adjusting for the increased area sampled as distance increases and handling cases where the exact location of points is not known, but only the distance to a known observer location is recorded.

1.4 Spatial Capture-Recapture

Spatial capture-recapture (Borchers and Efford, 2008; Royle and Young, 2008) is related to distance sampling and is applied in contexts where the identity of individuals animals is knowable. When individually identifiable animals are seen multiple times

(or 'captured' and 'recaptured') these data can be used to estimate abundance (Amstrup et al., 2005). SCR is a method that leverages the spatial location of captures and recaptures to learn about the spatial ecology of animal populations. This is also based on a point process model, one that describes the distribution of *activity centres* across a region. Animals move around these activity centres and may or may not be detected by our survey efforts. In this setting, the further away an animal moves from its activity centre, the less likely it is to be detected. In this case, the model for $p(\boldsymbol{s})$ is more complicated than for distance sampling, but its purpose in the model amounts to the same thing: some animals remain undetected completely and we would like to estimate how many there are using the data we did observe.

1.5 Inference for spatial models

This section discusses two approaches to inference for spatial models that we use in this thesis. The distance sampling models are fitted using the approximate Bayesian method of integrated nested Laplace approximations (INLA) (Rue et al., 2009) which has become a popular choice in spatial statistics for its computational efficiency in implementing sparse random effects (Bakka et al., 2018).

The SCR model cannot be implemented in INLA and hence, for inference in this case, we use Laplace approximate maximum likelihood (LAML), marginalising out the spatially structured random effect. To do this integration approximately we use the Laplace approximation and automatic differentiation implemented in the R package `Template Model Builder` (`TMB`) (Kristensen et al., 2016).

Both methods, INLA and LAML, rely on the Laplace approximation which is a method for approximating the integral of a function by considering the second-order Taylor expansion around the mode of integrand. This approximation can be fairly accurate for uni-modal functions (Rue et al., 2009; Vonesh, 1996).

1.5.1 Integrated nested Laplace approximations

Rue et al. (2009) introduced an approximate method to compute posterior quantities of interest for *latent Gaussian models* (LGMs). A LGM is a model of the form

$$\begin{aligned} y_i|\eta_i, \theta &\sim \pi(y_i|\eta_i, \theta) \\ \eta_i &= \boldsymbol{\beta}^T \boldsymbol{z}_i + \sum_r f_r(u_i|\boldsymbol{\xi}) \\ [\boldsymbol{\beta}, \boldsymbol{\xi}]^T &\sim N(0, \boldsymbol{\Sigma}|\theta) \\ \theta &\sim \pi(\theta), \end{aligned}$$

where $\mathbb{E}(y_i|\eta_i, \theta) = g(\eta_i)$ for some appropriate link function $g(.)$. The predictor η_i depends on some 'fixed effect' covariates $\boldsymbol{z}_i$ with parameters $\boldsymbol{\beta}$, which for simplicity may also include an intercept term. In addition to these fixed effects, there are random effects $f_r(u_i)$ which have, collectively, a parameter vector $\boldsymbol{\xi}$. The main parameters of interest, $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$, are latent and have an Gaussian prior, hence the term *latent* Gaussian model. This prior may depend on parameters θ, which are sometimes referred to as hyper priors, and have prior distribution $\pi(\theta)$. In the types of models we consider in this thesis, the hyper parameters for the random effects are the smoothing parameter (in a spline regression context) or the parameters of a covariance function (in a GRF context).

INLA is a method for approximating marginal posterior distributions of interest for models that can be written as a LGM. These are typically $\pi(\beta_i|y)$ and $\pi(\theta_i|y)$. This is achieved by marginalising and using approximate methods to evaluate these integrals. The posterior for the hyper parameters is estimated as

$$\tilde{\pi}(\theta|y) \propto \frac{\pi(\hat{\boldsymbol{\beta}}, y, \theta)}{\pi_G(\hat{\boldsymbol{\beta}}|y, \theta)},$$

where $\hat{\boldsymbol{\beta}}$ is the mode of $\pi(\boldsymbol{\beta}, y, \theta)$ and $\pi_G(\boldsymbol{\beta}|y, \theta)$ is the Gaussian approximation of $\pi(\boldsymbol{\beta}|y, \theta)$ centred at $\hat{\boldsymbol{\beta}}$. Then the posterior of any specific hyper parameter can be

approximated by $\pi(\theta_i|y) = \int \tilde{\pi}(\theta|y)\mathrm{d}\theta_{-i}$. The posterior for the parameters of the predictor η, for now using $\boldsymbol{\beta}$ to represent all coefficients, including those for the random effects, can be approximated by

$$\tilde{\pi}(\beta_i|\theta, y) \propto \frac{\pi(\tilde{\boldsymbol{\beta}}, y, \theta)}{\pi_G(\tilde{\boldsymbol{\beta}}_{-i}|\beta_i, y, \theta)},$$

where $\tilde{\boldsymbol{\beta}}$ is the mode of $\pi(\boldsymbol{\beta}, y, \theta)$ where $\tilde{\beta}_i = \beta_i$, and $\pi_G(\boldsymbol{\beta}_{-i}|\beta_i, y, \theta)$ is the Gaussian approximation for $\boldsymbol{\beta}_{-i}|\beta_i, y, \theta$ here (i.e. for the joint posterior distribution of all other parameters in η except β_i). This is the Laplace approximation of $\pi(\beta_i|\theta, y)$.

This description of INLA borrows heavily from Wood (2020) which introduces a variation on INLA but has a very clear and accessible description of the method. INLA is implemented in the R package `R-INLA` (Rue et al., 2009) which has implemented a suite of likelihood functions for $\pi(y_i|\eta_i, \theta)$, as well as many different random effect models and hyper parameter priors. We mainly make use of the Poisson likelihood which we can use to approximate the log-Gaussian Cox process likelihood (see Chapter 3, Section 3.2).

`R-INLA` remains the most popular implementation of the SPDE approach given its ability to utilise sparsity when evaluating Gaussian densities that depend on a sparse precision matrix. It does this by using computationally efficient algorithms for working with the precision matrix that can leverage sparsity (Rue and Held, 2005). Using spatially-structured random effects can be computationally expensive, often resulting in dense prior precision matrices which typically have $\mathcal{O}(n^3)$ cost to factorise for an effect with n parameters. A key development in this area is the use of sparse Gaussian Markov random field (GMRF) representations of Gaussian random fields (GRFs) and `R-INLA` uses sparse matrix methods to factorise the precision matrix at a cost of $\mathcal{O}(n^{3/2})$.

1.5.2 Laplace approximate maximum likelihood

In a maximum likelihood setting, we can formulate the problem of inference with latent random effects in a similar way to the approximate Bayesian approach of INLA. In this perspective, we want to know the likelihood of $\boldsymbol{\beta}$ while treating the random effect parameters $\boldsymbol{\xi}$ as 'nuisance parameters' and integrating them out. We can write this as

$$\mathcal{L}(\boldsymbol{\beta}|y) = \int \pi(\boldsymbol{\beta}|\boldsymbol{\xi}, y)\pi(\boldsymbol{\xi}|y)\mathrm{d}\boldsymbol{\xi}. \tag{1.2}$$

When $\boldsymbol{\xi}$ is a GRF or smoothing spline effect, $\pi(\boldsymbol{\xi}|y)$ is a Gaussian density (see Chapter 2). We can approximate this integral using the Laplace approximation, which requires the mode of $\pi(\boldsymbol{\beta}|\boldsymbol{\xi}, y)\pi(\boldsymbol{\xi}|y)$. Note that this approach does not require the model to be a LGM as was necessary for INLA. We use this method for the SCR likelihood and implement it in `TMB`, using automatic differentiation (Kristensen et al., 2016) to calculate the gradients required for finding the mode with respect to $\boldsymbol{\xi}$ and also to maximise the Laplace approximate marginal likelihood and generate uncertainty estimates using the Hessian of the approximate marginal likelihood.

1.6 A note on notation

We use π to denote a generic probability distribution whose properties are implied by its arguments. For example, $\pi(X)$ is the probability distribution for X, $\pi(X|Y)$ is the conditional probability distribution for X given Y, $\pi(X, Y)$ is the joint probability distribution for X and Y, and so on. We also sometimes make use of the squared bracket notation where $[X]$ is the probability distribution function for X, $[X|Y]$ is the conditional probability distribution for X given Y, and so on. We use the term distribution here but in the text this notation could refer to a probability mass function or a probability density function and it should be clear from context what is intended.

1.7 Summary

To summarise, this thesis is about understanding and implementing spatially structured random effects within two common modelling frameworks in spatial ecology: distance sampling and spatial capture recapture. We present the SPDE approach in a comprehensive way that is understandable to applied statisticians without a strong background in the theory of SPDEs and Gaussian random fields. We then use this approach in a distance sampling application to jointly estimate the distribution and detectability of animals, using a novel approach to inference and introduce some new approaches for communicating uncertainty. Finally, for the first time, we use spatially structured random effects within SCR and present the various modelling decisions that need to be taken in order to do this, ending with a discussion on future research directions. This work represents contributions to each step of the applied statistics workflow, understanding the methods available to us, building appropriate models for a given application, choosing efficient and accurate means of inference, and effective communication of the modelling results.

Chapter 2

Understanding the stochastic partial differential equation approach to smoothing

2.1 Introduction

This chapter is concerned with developing an understanding of the relationship between two popular methods for modelling spatial and temporal structure in data: *penalised smoothing splines* (Wood, 2017) and *stochastic partial differential equations* (Oksendal, 2013) (SPDEs). The motivation is to better understand the approach introduced in Lindgren et al. (2011), who present a Gaussian Markov random field (Rue and Held, 2005) approximation for a Gaussian random field (GRF) defined by an SPDE. This has become known as the 'SPDE approach' in the literature and has been implemented in the `R-INLA` software package (Rue et al., 2009) and, more recently, in `Template Model Builder` (`TMB`) (Kristensen et al., 2016). Both these libraries offer a wide range of tools for flexible hierarchical modelling. However, a wider adoption of the SPDE approach in the applied spatial statistics literature has been hindered for two primary reasons. First, the mathematical concepts involved in

presenting the SPDE approach are advanced and more usually seen in the physical sciences and applied mathematics literature. These concepts and associated terms, and their relationships to other methods, are not necessarily well-known in the applied spatial statistics community. Second, the available software implementations, `R-INLA` and `TMB`, while flexible, are difficult to customise further, requiring a deep technical knowledge of the specific software packages, which could deter practitioners from adopting the SPDE approach.

The aim of this chapter is to address both these issues by drawing links between the SPDE approach and penalised smoothing splines Wood (2017). By doing this we hope that statisticians who are already familiar with smoothing splines, a common approach in applied spatial statistics, will be able to better understand the SPDE approach and implement it in software packages they are already familiar with. As an example of this we show how to implement the SPDE approach in the R package `mgcv` (Wood, 2017), which is commonly used to fit models with penalised regression splines, has a large base of users and is readily extended and customised for application.

As was mentioned in the introduction, much of this work is published in Miller et al. (2019), for which I am joint first author along with David Miller and Richard Glennie. This work grew out of a reading group I set up in order to better understand the SPDE approach, a method I intended to use throughout my PhD research. In doing this we realised that, by drawing links with methods we already had some familiarity with (smoothing splines), this understanding would be useful to others who are interested in implementing SPDE models.

The additional information in this chapter, over and above what is in Miller et al. (2019), can be summarised as follows: It contains more detailed and accessible introductions to smoothing splines, Gaussian random fields, SPDEs and finite element methods, and a detailed derivation of the finite element precision matrix for the SPDE effect which is essentially a more comprehensive version of the proof in

Appendix D.3.1 in Lindgren et al. (2011). We also include an additional discussion on the similarities and differences between smoothing penalties and SPDEs, and new examples of fitting the model using `mgcv` to real datasets that include using the Tweedie distribution which is currently not available in `R-INLA` [1] (Rue et al., 2009), the R package with the most commonly-used implementation of the SPDE approach.

2.2 Basics of spline regression

This section introduces smoothing splines and their key aspects that we later use to draw links with the SPDE approach. Spline regression is based on the idea that observations depend, in some way, on a latent function $\tilde{f}(\boldsymbol{s})$. We use $\boldsymbol{s}$ to indicate a location in space, that we assume is in some bounded domain $\mathcal{S}$, $\tilde{f}(\boldsymbol{s})$ to denote the 'true' function that gave rise to the observed data, $\hat{f}(\boldsymbol{s})$ to denote some point estimate of $\tilde{f}(\boldsymbol{s})$ and $f(\boldsymbol{s})$ to denote a general element of some space of functions $\mathcal{F}$.

The simplest case is where the data are an exact observation of $\tilde{f}$ at a finite number of locations. Then the data consist of $y_i = \tilde{f}(\boldsymbol{s}_i)$ for $i = 1, \dots n$. Ideally we want an estimate $\hat{f}(\boldsymbol{s})$ to be as simple as possible, fitting the data but without unnecessary flexibility. In order to do this we need some measure of flexibility that we can incorporate so that the model has some way to measure, and thus constrain, the flexibility of $f(\boldsymbol{s})$.

Definition (smoothing penalty)**.** Given a function $f(\boldsymbol{s})$ defined on a bounded domain $\mathcal{S}$, a *smoothing penalty* of $f(\boldsymbol{s})$ is an integral of the form

$$\int_{\mathcal{S}} [\mathcal{D} f(\boldsymbol{s})]^2 \, \mathrm{d}\boldsymbol{s},$$

for some choice of differential operator $\mathcal{D}$.

[1] The Tweedie distribution was added to the testing version of `R-INLA` in early 2021. However, it has not yet been fully tested and remains in development.

The choice of $\mathcal{D}$ can be thought of as the *definition* of smoothness for a given application. Depending on the choice of $\mathcal{D}$, different features of the function are considered as the measure of 'smoothness'. Derivatives describe rates of change of $f(\boldsymbol{s})$ and so by squaring these we get a non-negative measure of how much the function is changing at each location $\boldsymbol{s}$, which we consider over the whole domain (integrating the squared derivatives). A popular choice for $\mathcal{D}$ in one dimension is to use the second derivative which, for a function $f(t)$, uses $\mathcal{D} = \frac{\mathrm{d}^2}{\mathrm{d}t^2}$. In two dimensions, we could use the sum of the second-derivatives. For a function $f(\boldsymbol{s})$ this uses $\mathcal{D} = \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}$. The choice of $\mathcal{D}$ is context specific and, in many situations, there is not a clear and obvious choice that is superior to others. The smoothing penalty is the integral of this chosen measure of smoothness over the domain of the function.

Definition (interpolating spline)**.** Given exact (no error) observations, $y_i = \tilde{f}(\boldsymbol{s}_i)$ for $i = 1, \ldots, n$, the *interpolating spline* is the unique function $\hat{f}$ that minimises the smoothing penalty $\int_{\mathcal{S}} [\mathcal{D}f(\boldsymbol{s})]^2 \, \mathrm{d}\boldsymbol{s}$ subject to the constraint that $\hat{f}(\boldsymbol{s}_i) = \tilde{f}(\boldsymbol{s}_i)$ for $i = 1, \ldots, n$.

The interpolating spline is the smoothest function that passes through the observed data points, where the meaning of 'smoothest' is determined by the choice of $\mathcal{D}$. Here we take as given that $\hat{f}$ exists and is unique although this depends on some mild assumptions about $\tilde{f}$ and the number of observations n, which must be sufficiently large, where 'sufficient' depends on the order of $\mathcal{D}$. Kimeldorf and Wahba (1970b) cover the mathematical details.

Smoothing splines are similar to interpolating splines but consider the case where instead of exact observations of $\tilde{f}(\boldsymbol{s})$, we make observations with error, or perhaps never even observe it at all in cases where the smooth function is a latent model component. The aim here is to estimate a function $\hat{f}(\boldsymbol{s})$, or perhaps compute a posterior distribution on a space of functions $\mathcal{F}$ given the data, that is flexible enough to fit the data well, but not so flexible that it overfits and generalises poorly.

To introduce smoothing splines we consider the simplest case where we observe $\tilde{f}$ with i.i.d Gaussian error. We have data of the form

$$y_i = \tilde{f}(\boldsymbol{s}_i) + \epsilon_i \quad \text{for } i = 1, \dots n,$$

where each $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. We replace the exact conditions required by the interpolating spline (that $\hat{f}(\boldsymbol{s}_i) = \tilde{f}(\boldsymbol{s}_i)$) with some goodness-of-fit measure for an estimate $\hat{f}$. Often the natural choice here is to use the likelihood of the data given $\hat{f}$ but in this simple case we can take the expected sum of squares $\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \hat{f}(\boldsymbol{s}_i)\right]^2$ as the measure of fit. Then we can define the smoothing spline as follows:

Definition (smoothing spline)**.** For a given choice of smoothing penalty $\int \left[\mathcal{D}f(\boldsymbol{s})\right]^2 \mathrm{d}\boldsymbol{s}$ and smoothing parameter value $\lambda > 0$, the smoothing spline is the function $\hat{f}$ that minimises

$$\frac{1}{n}\sum_{i=1}^{n}\left[y_i - f(\boldsymbol{s}_i)\right]^2 + \lambda \int \left[\mathcal{D}f(\boldsymbol{s})\right]^2 \mathrm{d}\boldsymbol{s}.$$

Enforcing $\lambda > 0$ ensures the smoothing penalty is always costly in this minimisation problem. The ultimate 'cost' of the smoothing penalty is controlled by the smoothing parameter λ. Larger values of λ penalise more heavily than smaller values. For this simple Gaussian error model this expression is equivalent to a *penalised log-likelihood*. For a more general model, such as a exponential family generalised linear model (GLM) or some other hierarchical model, we can replace the sum of squares as the measure of fit with the log-likelihood for the data given the smooth function f. In the linear model with Gaussian error example above, the error variance σ^2 is absorbed by the smoothing parameter λ since the Gaussian likelihood for $\boldsymbol{y}$ is multiplied by the exponential of the smoothing penalty and so these two parameters are confounded. This makes sense because as unstructured noise in the data increases the smooth function needs less flexibility. Note that, for now, we are treating the smoothing parameter λ as fixed and known; we discuss approaches for the estimation of λ later.

A smoothing spline $\hat{f}$ must achieve the best fit it can to the data (minimising the expected sum of squares), under the constraint that more wiggly functions will make the smoothing penalty larger. Thus the smoothing spline is a trade off between fit and smoothness. For larger values of λ, the penalty is more severe, and the smoothing spline will be more smooth. For $\lambda = 0$, the function f is free to interpolate the observed data without any concern for smoothness. Again we take it as given that such a function exists Kimeldorf and Wahba (1970b). In practice, we do not know λ and therefore would like to estimate it.

Definition (penalised regression spline)**.** For a given choice of smoothing penalty $\int [\mathcal{D}f(\boldsymbol{s})]^2 \, \mathrm{d}\boldsymbol{s}$, the optimal smoothing spline is the function $\hat{f}$ and smoothing parameter value $\hat{\lambda}$ that together minimise

$$\frac{1}{n} \sum_{i=1}^{n} [y_i - f(\boldsymbol{s}_i)]^2 + \lambda \int [\mathcal{D}f(\boldsymbol{s})]^2 \, \mathrm{d}\boldsymbol{s}$$

i.e., the penalised regression spline is a smoothing spline for which the optimal smoothing parameter $\hat{\lambda}$ is also estimated. Inference for the smoothing parameter λ can be achieved by using REstricted maximum likelihood methods (REML) (Wood, 2011), generalized cross validation (GCV) (Wahba, 1990) or within a Bayesian framework (Lang and Brezger, 2004; Speckman and Sun, 2003). The method we use in this chapter is REML as implemented in `mgcv` using the Laplace approximation (Wood, 2017, 2011). The way the SPDE approach is implemented in a Bayesian context (e.g. in `R-INLA`) is essentially by defining a *prior* for $f(\boldsymbol{s})$ along with priors for associated smoothing parameters. We explore below the connection between smoothing penalties and priors that is the basis for drawing links between the SPDE approach and smoothing splines.

The above framework can be extended to more complicated, possibly hierarchical, models by using the log-likelihood to represent the goodness of fit to observed data. This leads to a general approach for including penalised regression

splines within a model. We denote the log-likelihood for f as $l(\boldsymbol{y}|f)$ and the smoothing penalty term as $J_{\mathcal{D}}(f|\lambda) := -\lambda \int [\mathcal{D}f(\boldsymbol{s})]^2 \mathrm{d}\boldsymbol{s}$. The joint log-likelihood is $l(\boldsymbol{y}|f,\lambda) = l(\boldsymbol{y}|f) + J_{\mathcal{D}}(f|\lambda)$. The REML approach for dealing with this joint likelihood is to integrate out f using the Laplace approximation and estimate λ using the marginal likelihood $l(\boldsymbol{y}|\lambda)$ (Wood, 2011).

A Bayesian framing of the problem is often introduced by considering the likelihood $\mathcal{L}(\boldsymbol{y}|f,\lambda) = \mathcal{L}(\boldsymbol{y}|f)\exp\left[J_{\mathcal{D}}(f|\lambda)\right]$ which is just taking the exponential of the joint log-likelihood given above (Wood, 2017). In this view the smoothing penalty is equivalent to a prior on f, i.e. there is a (potentially improper) density $\pi(f|\lambda)$ where

$$\pi(f|\lambda) \propto \exp\left(-\lambda \int \left[\mathcal{D}f(\boldsymbol{s})\right]^2 \mathrm{d}\boldsymbol{s}\right). \tag{2.1}$$

If a prior is also placed on λ then inference for f and λ can proceed by Bayes' theorem. However, there are some issues with this interpretation of penalised log-likelihoods. One is that not all penalties lead to a proper density as they imply covariance structures that are not full rank (Chapter 4 has a detailed discussion this). We clarify this point below in the context of having chosen a finite dimensional basis to represent $f(\boldsymbol{s})$. In general, if there are any basis functions that always have zero penalty (for example, a linear function in the basis for $f(\boldsymbol{s})$ would have zero second derivative) then the covariance matrix for the basis function coefficients will not be full rank. Therefore, the interpretation as a prior depends on the choice of basis and some choices result in an improper density.

Up to this point we have not been explicit about the space of functions $\mathcal{F}$ and so it is not immediately clear how to define a probability distribution over this space. The interpretation as a penalty or a prior depends primarily on the approach to inference, but not the model structure itself. Given a particular choice of smoothing penalty, this model component should not be described as a prior if inference is, for example, conducted by REML, as is the case in `mgcv` (Wood, 2011). In this framework, we do not estimate a posterior, just a point estimate based on maximising a marginal

likelihood and uncertainty around this point estimate. However, the shorthand understanding that penalties and priors are essentially playing the same role is a useful one to keep in mind.

In order to make sense of the penalty we need to say a bit more about the space of functions $\mathcal{F}$ from which we wish to obtain a point estimate for the optimal smoothing spline or, in the Bayesian case, a posterior distribution over $\mathcal{F}$. Function spaces can often contain an infinite number of basis elements (for example, the space of polynomials with real coefficients which has the infinitely large basis $\{1, x, x^2, x^3, \ldots\}$) but for computational reasons in applied statistics settings, the function space is usually more restricted and has a finite set of pre-specified basis functions. The span of this finite set of basis functions is then taken to be $\mathcal{F}$.

Definition (basis representation)**.** A space of functions $\mathcal{F}$ has a finite basis if there exists a finite set of functions $b_1, \ldots, b_K \subset \mathcal{F}$ such that any function $f \in \mathcal{F}$ can be written as

$$f(\boldsymbol{s}) = \beta_1 b_1(\boldsymbol{s}) + \ldots + \beta_K b_K(\boldsymbol{s})$$

for some unique set of coefficients $\beta_1, \ldots, \beta_K$. We call this the *basis representation* of f.

Restricting ourselves to spaces of functions with a finite basis allows us to write down the density on $\mathcal{F}$ that is implied by the penalty. Since the basis functions are fixed (non-stochastic), this amounts to specifying a distribution for $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_k]^T$ In practice the choice of basis should be sufficiently rich to address the particular application. If the latent function we are trying to estimate has a lot of flexibility then the basis needs to have enough functions to be able to capture this. For example, if a cubic polynomial shape is required, a basis consisting of a quadratic polynomial and a straight line will not be sufficient.

Given a basis representation, it is possible to write the integral in the penalty $J_{\mathcal{D}}(f|\lambda)$ as a matrix-vector product for cases where $\mathcal{D}$ is a linear differential oper-

ator. If this is the case then $\mathcal{D}f(\boldsymbol{s}) = \sum_k \beta_k \mathcal{D}b_k(\boldsymbol{s})$. In other words, the space of derivatives of functions in $\mathcal{F}$ also has a basis that can be constructed by applying $\mathcal{D}$ to each basis element of $\mathcal{F}$. In practical terms, since the basis is fixed, this means we only need to compute derivatives $\mathcal{D}b_1, \ldots, \mathcal{D}b_K$ once and can use these to compute the penalty.

Definition (penalty matrix)**.** Given a choice of basis $\{b_1, \ldots, b_K\}$ with coordinate vector $\{\beta_1, \ldots, \beta_K\}$, and a choice of linear differential operator $\mathcal{D}$, the *penalty matrix* $\boldsymbol{S}$ is the $K \times K$ matrix with entries $\boldsymbol{S}_{ij} = \int \mathcal{D}b_i(\boldsymbol{s})\mathcal{D}b_j(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$.

This implies that

$$\int_{\mathcal{S}} \left[\mathcal{D}f(\boldsymbol{s})\right]^2 \mathrm{d}\boldsymbol{s} = \boldsymbol{\beta}^T \boldsymbol{S} \boldsymbol{\beta}. \tag{2.2}$$

Therefore, given a basis representation for f, and a linear differential operator, the penalty can be expressed as a matrix-vector product. For any fixed choice of basis and differential operator, the matrix $\boldsymbol{S}$ only needs to be computed once and then the above matrix-vector formula can be used to compute the penalty for any vector of coefficients $\boldsymbol{\beta}$. In a sense, the matrix $\boldsymbol{S}$ contains all of what we mean by 'smoothness', expressed in terms of the building blocks of the function space. For this reason the penalty matrix is also sometimes referred to as the *smoothing matrix*. For example, consider the basis $\{1, t, t^2, t^3\}$ defined on the interval $[0, 1]$ along with the second derivative penalty with differential operator $\mathcal{D} = \frac{\mathrm{d}^2}{\mathrm{d}t^2}$. This basis spans the set of polynomials of degree ≤ 3. Then the smoothing matrix is a 4×4 matrix that has the form

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & S_{33} & S_{34} \\ 0 & 0 & S_{43} & S_{44} \end{bmatrix}.$$

The zeroes in this matrix follow from the fact that the second-derivative of 1 and t is zero. The other entries have the form $S_{33} = \int_0^1 2 \cdot 2\mathrm{d}t$, $S_{34} = \int_0^1 2 \cdot 6t\mathrm{d}t$ and so on.

In addition to this efficient computation of the penalty, we can now be more explicit by what we mean by a density on the space of functions $\mathcal{F}$ since this amounts to defining distributions on the coordinate vector space $\mathbb{R}^K$. The penalised likelihood can be now written as $l(\boldsymbol{y}|f) + \lambda\boldsymbol{\beta}^T\boldsymbol{S}\boldsymbol{\beta}$. This means we can replace the expression for the prior density $\pi(f|\lambda)$ with

$$\begin{aligned}\pi(\boldsymbol{\beta}|\lambda) &= \exp(-\lambda\boldsymbol{\beta}^T\boldsymbol{S}\boldsymbol{\beta}) \\ &\propto \mathcal{N}(0, \boldsymbol{\Sigma}) \end{aligned} \tag{2.3}$$

where $\boldsymbol{\Sigma} = (2\lambda\boldsymbol{S})^{-1}$. This makes it clear that the penalty implies a Gaussian prior density on $\mathbb{R}^K$ with a covariance matrix that is directly related to the smoothing penalty matrix. This is the first instance in this chapter where we make an explicit connection between the terms 'covariance' and 'smoothness'. For Gaussian random fields it is common to present random effects by directly referencing the covariance structure, usually by stating the covariance function. For smoothing splines the emphasis tends to be on stating the chosen smoothing penalty. This connection between covariance and smoothness is a theme that we return to later.

The above assumes that the penalty matrix $\boldsymbol{S}$ is invertible and leads to a positive semi-definite covariance matrix. If the penalty matrix is not invertible then $\boldsymbol{\Sigma}$ can be replaced with a pseudo-inverse and this results in an improper density. Some common choices of basis functions and smoothing penalties can lead to penalty matrices that are not full rank, as shown in the examples below and as was given in the example above with the polynomial basis. Note that the choice of basis as well as the choice of $\mathcal{D}$ both influence the structure of the resulting penalty matrix. For this reason, we sometimes use the term *basis-penalty smoothers* to emphasise the importance of both the basis and the penalty for the ultimate structure of the model. The penalty matrix $\boldsymbol{S}$ contains the information required to compute the smoothness of any given function represented as a choice of basis coefficients $\boldsymbol{\beta}$. If

inference is by Bayesian methods then this penalty can be interpreted as a (possibly improper) Gaussian prior, and the penalty matrix interpreted as precision matrix.

2.2.1 Examples of basis-penalty smoothers

This section introduces some commonly used basis-penalty smoothers. The aim is to give some examples in order to ground the above theory in some explicit models. All the examples in this section can be found in Chapter 4 of Wood (2017), which includes much more information than is presented here.

In one dimension the *cubic spline basis* is defined using a pre-specified set of locations $t_1, \ldots, t_K$, known as knots. The basis consists of a set of cubic polynomials defined on each interval $[t_i, t_j]$ and set to zero outside of this interval. In a cubic regression spline context, the coefficients of the basis functions are estimated with additional constraints that the spline function has continuous second derivatives at the knots (so the basis functions are joined together in a 'smooth' way) and boundary conditions $f''(t_1) = f''(t_n) = 0$. The second-derivative penalty in one dimension is the most common choice of penalty to use with a cubic spline basis. This leads to perhaps the most succinct statement of a penalised smoothing spline regression problem. Given data $\boldsymbol{y} = [y_1, \ldots, y_n]^T$ observed with Gaussian error and a cubic spline basis for the spline function f, the regression problem is to minimise

$$\sum_{i=1}^{n} [y_i - f(t_i)]^2 + \lambda \int f''(t)^2 \mathrm{d}t. \tag{2.4}$$

In this case we have smoothing penalty $J_{\mathcal{D}}(f|\lambda) = -\lambda \int f''(t)^2 \mathrm{d}t$ and $\mathcal{D} = \mathrm{d}^2/\mathrm{d}t^2$. Other formulations of cubic splines use basis functions that are defined over the whole observable interval. These basis functions are defined in more complicated way and so we omit them here for brevity. Cubic splines can also be defined to be cyclic so $f(t_1) = f(t_K)$. See Wood (2017) p149-152 for details.

Another popular choice of basis is the *B-spline* basis, where, similar to the cubic

spline basis, each basis function is non-zero over a finite subset of the domain. Below we will see an example of a B-spline basis of order one, which we call piece-wise linear, that is used in approximating solutions to SPDEs. Higher-order B-splines are more commonly used in the basis-penalty smoother paradigm. The use of B-spline bases led to the development of what Wood (2017) call *P-splines*, which stands for *penalised* B-splines. This approach uses a simple definition for the penalty matrix, known in this context as a difference matrix, and is an example of a penalty that is not derived from a differential operator. This smoothing penalty penalises the squared difference between the B-spline coefficients, which are usually placed at equal intervals. In this case the smoothing penalty is not derived from a differential equation and is

$$\boldsymbol{\beta}^T \begin{bmatrix} 1 & -1 & 0 & . & . \\ -1 & 2 & -1 & . & . \\ 0 & -1 & 2 & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix} \boldsymbol{\beta},$$

which is very straightforward to implement. P-splines have that disadvantage that if knots are not evenly spaced the penalty can be hard to interpret.

The final example we mention here is the *thin plate regression spline*. This is somewhat different in nature to the example above in that it is not necessary to choose a set of knot locations and the thin plate spline generalises to domains with more than one dimension. The thin plate basis places radially symmetric basis function on each unique combination of predictor variables. In the applications in this thesis this generally amounts to a basis function centred on each of the discretisation locations in a spatial domain. In addition to these radially symmetric basis functions there is also a set of *null-space* basis functions which are not penalised. These are functions for which the penalty always evaluates to zero. If the penalty involves second derivatives then a polynomial of degree one or zero would be in the

null-space. In two dimensions the thin plate smoothing penalty is

$$J(f|\lambda) = -\lambda \int \int \left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}\right)^2 \mathrm{d}x\mathrm{d}y$$

Since the basis has more basis functions than observations the computational cost can be high. This basis arises as an 'optimal' basis to use for spline interpolation problems in the space of twice-differentiable functions (see Wahba (1990)) and so is of a slightly different nature to other spline bases, which are typically chosen to be a sufficiently flexible approximation for a given application.

In order to turn this rich but unwieldy basis into something more useable, Wood (2003) introduced thin plate regression splines (TPRS) that use a truncated eigen-decomposition to approximate the full thin plate basis. This leads to substantial improvements in computational efficiency. This approach is presented in more detail and used the SCR context in Chapter 4.

2.3 Basics of Gaussian random fields

This section introduces the mathematical concepts and properties of Gaussian random fields (GRFs) that are relevant to applied spatial modelling. Rasmussen and Williams (2005) provide a good overview of the mathematical details of Gaussian random fields and Banerjee et al. (2014) cover important aspects in the applied spatial statistics setting. The presentation here takes inspiration from both.

Definition. (random field) A *random field* is a collection of random variables $\{\xi(\boldsymbol{s}); \boldsymbol{s} \in \mathcal{S}\}$ with index set $\mathcal{S}$.

This is a generalisation of a stochastic process, where the index set is usually the real line or the integers. For a random field, there are typically fewer restrictions on the index set. In spatial statistics the index set is usually a two-dimensional space of coordinates, such as longitude and latitude, for example.

Definition. (Gaussian random field) A random field $\{\xi(\boldsymbol{s}); \boldsymbol{s} \in \mathcal{S}\}$ is a *Gaussian random field* if, for any finite set $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\} \subset \mathcal{S}$, the random vector $[\xi(\boldsymbol{s}_1), \ldots, \xi(\boldsymbol{s}_n)]^T$ has a multivariate Gaussian probability distribution.

This definition of a random field is simple to state and an extension of the fact that a distribution on a finite dimensional space is completely described by the joint distribution function. Defining a random field in this way, for every possible finite collection of random variables, extends that notion to the case of an infinite collection of random variables. We use vector notation $\boldsymbol{\xi} = [\xi(\boldsymbol{s}_1), \ldots, \xi(\boldsymbol{s}_n)]^T$, usually suppressing the coordinates with the understanding it is a random vector defined at some finite set of locations. So $\mathbb{E}[\boldsymbol{\xi}]$ has i-th element $\mathbb{E}[\xi(\boldsymbol{s}_i)]$ for example. The covariance matrix for these multivariate Gaussian densities is determined by the covariance function of the Gaussian random field.

Definition. (covariance function) The *covariance function* of a Gaussian random field is the function

$$C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \mathrm{Cov}\,[\xi(\boldsymbol{s}_i), \xi(\boldsymbol{s}_j)]$$

for any pair of indices $\boldsymbol{s}_i, \boldsymbol{s}_j$

It follows from this definition that, for a GRF with covariance function $C(.)$, the random vector $\boldsymbol{\xi}$ has a multivariate Gaussian density and covariance matrix $\boldsymbol{\Sigma}$ with entries $(\boldsymbol{\Sigma})_{ij} = C(\boldsymbol{s}_i, \boldsymbol{s}_j)$. The covariance function is the rule by which covariance matrices can be constructed for the GRF at any finite set of locations.

In spatial statistics there is a focus on GRFs with particular properties for both the covariance function and the expected values of the Gaussian random field. We briefly cover these properties below.

Definition. (stationarity) A Gaussian random field is *stationary* (or *strongly stationary*) if its finite dimensional multivariate Gaussian distributions are invariant to

translation. i.e. the random vectors

$$[\xi(\boldsymbol{s}_1), \ldots, \xi(\boldsymbol{s}_n)]^T \quad \text{and} \quad [\xi(\boldsymbol{s}_1 + \boldsymbol{h}), \ldots, \xi(\boldsymbol{s}_n + \boldsymbol{h})]^T$$

have the same joint distribution for any choice of vector $\boldsymbol{h} \in \mathcal{S}$.

For a Gaussian random field with constant mean, the stationarity property implies that the covariance function depends only on the relative positions of the locations. That is to say, for any two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, and vector $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_j$, then the covariance function can be written as a function of $\boldsymbol{h}$, so that $C(\boldsymbol{s}_i, \boldsymbol{s}_j) = C(\boldsymbol{h})$. This property is known as *weak stationarity* or *second-order stationarity*. All of the Gaussian random fields considered in this thesis are both strongly and weakly stationary. Another related property is *isotropy* which is invariance with respect to rotation.

Definition. (isotropy) A Gaussian random field is *isotropic* if its finite dimensional multivariate Gaussian distributions are invariant to rotation. i.e. the random vectors

$$[\xi(\boldsymbol{s}_1), \ldots, \xi(\boldsymbol{s}_n)]^T \quad \text{and} \quad [\xi(\boldsymbol{R}\boldsymbol{s}_1), \ldots, \xi(\boldsymbol{R}\boldsymbol{s}_n)]^T$$

have the same joint distribution for any matrix $\boldsymbol{R}$ that defines a rotation.

The covariance function of a GRF that is both stationary and isotropic can be parameterised in terms of the distance between locations. In this case we can write $C(\boldsymbol{s}_i, \boldsymbol{s}_j) = C(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|)$ where $\|.\|$ denotes the Euclidean distance in $\mathcal{S}$. This reflects the fact that if the multivariate Gaussian densities are invariant to translation and rotation, then the only remaining information that could be relevant is the distance between locations.

Stationary and isotropic random fields with a *Matérn covariance* are of particular interest in spatial statistics. See Stein (2012) for a detailed investigation of this covariance function. Here we restrict ourselves to stating the function and

highlighting some properties.

Definition. (Matérn covariance) The Matérn covariance function for a two-dimensional domain is

$$C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \frac{2^{1-\nu}}{4\pi\kappa^2\tau^2\Gamma(\nu)}(\kappa\|\boldsymbol{s}_i - \boldsymbol{s}_j\|)^\nu K_\nu(\kappa\|\boldsymbol{s}_i - \boldsymbol{s}_j\|) \tag{2.5}$$

where ν, κ, τ are parameters and K_ν is the modified Bessel function of the second kind.

Since this covariance function depends only on the distance between locations, the Matérn covariance function implies that the GRF is stationary and isotropic (so long as the mean is constant). The parameter κ is an inverse range parameter, smaller values of κ give longer correlation range. The precision parameter τ and κ together determine the marginal variance of the random field which is $(4\kappa^2\tau^2)^{-1}$ for two-dimensional domains. In 2-dimensional cases, a common reparameterisation is to replace κ with a range parameter $\rho = \sqrt{8}/\kappa$. For more on the Matérn covariance parameters and the relationships between them see Chapter 6 in Blangiardo and Cameletti (2015). The parameter ν controls the mean-square differentiability of the process. Larger values of ν mean realisations of the process have more derivatives and so are 'smoother' in the differentiable sense (note this is not the typical use of the word smooth in this thesis). The three parameters, κ, τ and ν are not simultaneously identifiable (Zhang, 2004) and it is conventional to assume a value for ν and to infer the remaining parameters from data. This convention makes sense in most practical settings. It is hard to imagine a dataset at a high enough resolution for which we wish to infer the differentiability of the process, which is defined on infinitesimal neighbourhoods. We call a Gaussian random field with Matérn covariance function a *Matérn random field*.

Of particular interest to us in this chapter is a class of Gaussian random fields known as *Gaussian Markov random fields* (GMRFs). Our description of GMRFs is an extremely succinct summary of the detailed treatment given in Rue and Held

(2005). Given an undirected graph $G = (V, E)$, with verties V and edges E, a random field indexed by V is said to be a Markov random field with respect to G if any two finite subsets of variables are conditionally independent given a separating subset. i.e. a collection of random variables $\boldsymbol{\beta}_A = \{\boldsymbol{\beta}(\boldsymbol{s})|\boldsymbol{s} \in A \subset V\}$ is conditionally independent from $\boldsymbol{\beta}_B = \{\boldsymbol{\beta}(\boldsymbol{s})|\boldsymbol{s} \in B \subset V\}$ given $\boldsymbol{\beta}_S = \{\boldsymbol{\beta}(\boldsymbol{s})|\boldsymbol{s} \in S \subset V\}$, where every path from nodes in A to nodes in B passes through S. This is a generalisation of the Markov property for stochastic processes to higher-dimensional cases. We will see below, in the section describing the SPDE approach to modelling the Matérn field, examples where the graph G (also called a mesh) is constructed using Delauney triangulation and the Markov property is equivalent to the entries of the precision matrix being non-zero only for nodes that are connected by an edge. In other words, for a GMRF with precision matrix $\boldsymbol{Q}$, $(\boldsymbol{Q})_{ij} = 0$ if and only if nodes $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are not connected by an edge.

2.4 Basics of stochastic partial differential equations

This section introduces some basic definitions and properties of SPDEs. The focus is on a non-rigorous description that introduces key concepts only so far as they are needed to understand the subsequent sections of this chapter. For a more mathematically rigorous introduction to the theory of stochastic partial differential equations see, for example, Oksendal (2013).

First we consider a non-stochastic partial differential equation that takes the form $\mathcal{D}f(\boldsymbol{s}) = g(\boldsymbol{s})$. The differential operator $\mathcal{D}$ is applied to an unknown, but deterministic, function f which results in a known function g. This is a standard form of problem that is seen in many areas of mathematics and science where the problem is to solve the differential equation to find f, given that we know g and some boundary conditions. A canonical example of this kind of application is that

the position and velocity of an object at time t can be obtained as a solution to some differential equations that describe the movement dynamics, given some information about the location and dynamics at time $t = 0$.

One way to add stochasticity to this type of problem is to replace the known function $g(.)$ with a stochastic process. In this chapter we consider stochastic partial differential equations of the form

$$\mathcal{D}f(\boldsymbol{s}) = \epsilon(\boldsymbol{s}) \tag{2.6}$$

where $\epsilon(\boldsymbol{s})$ is a *Gaussian white noise process*.

Definition. (Gaussian white noise)

A stochastic process $\epsilon(\boldsymbol{s})$ defined on $\mathcal{S}$ is a *standard Gaussian white noise process* if $\epsilon(\boldsymbol{s})$ satisfies

$$\int_{\mathcal{S}} \phi(\boldsymbol{s})\epsilon(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \sim \mathcal{N}(0, \|\phi\|), \tag{2.7}$$

where $\|\phi\| = \int \phi(\boldsymbol{s})^2 \mathrm{d}\boldsymbol{s}$ and ϕ is any deterministic function defined on $\mathcal{S}$ for which the integral is well-defined.

This is an example of a *generalised* random field. This terminology is analogous to generalised functions (also known as distributions, in the non-statistical sense), whereby a generalised function is defined in terms of its relationship to other, more standard, functions. In this case, white noise only makes sense as a particular type of integration with respect to a deterministic function ϕ.

In practice, the white noise process is usually implemented considering a finite collection of functions, $\phi_1, \ldots, \phi_m$, which are sometimes known as *test functions*. Here we assume that the space of functions for ϕ is a suitable space for which the above integrals are well-defined. We also use the inner product notation $\langle f, g \rangle = \int f(\boldsymbol{s})g(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$ below for brevity. Given this definition of a Gaussian white noise process, the right-hand side of the SPDE is only meaningful when integrated with respect to some function ϕ. This means the SPDE itself is only meaningful when

both sides are integrated. This integrated form of the SPDE is known as the *weak* formulation. The *weak solution* is a random field f that satisfies

$$\langle \mathcal{D}f, \phi \rangle \stackrel{d}{=} \langle \epsilon, \phi \rangle \tag{2.8}$$

for any choice of function ϕ and where $\stackrel{d}{=}$ means equal in distribution. For a fixed and known ϕ the right hand side of (2.8) has a known distribution that follows directly from the definition of the white noise process. We assume that ϵ has unit variance. The variance of the stochastic process ϵ can be changed by multiplying by a constant but here we assume that this is incorporated as part of the differential operator $\mathcal{D}$. In Section 2.5 we include a precision parameter τ on the right hand side of the SPDE which is readily accounted for in the numerical methods used to solve the SPDE (Section 2.9). The function ϕ is known as a *test function*. From the definition of white noise we know these integrals are normally distributed with mean zero and variance $\|\phi\|$. Instead of a deterministic solution for f, replacing the right hand side with Gaussian white noise has the effect that the solution is now a Gaussian random field.

Given this SPDE representation of a Gaussian random field f, it is natural to ask what is the connection between the SPDE and the properties of the random field. It follows directly from the choice of the white noise process that f has mean zero everywhere. The connection between the SPDE and the covariance function is more complicated. We cover this informally with a brief outline that uses convolutions, Green's functions and Fourier theory.

One way to represent a solution to a SPDE of the form in equation (2.6) is by writing f as a convolution: $f(\boldsymbol{s}) = \int G(\boldsymbol{s} - \boldsymbol{u})\epsilon(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$. If G satisfies $\mathcal{D}G(\boldsymbol{s} - \boldsymbol{u}) = \delta(\boldsymbol{s} - \boldsymbol{u})$, where differentiation is with respect to $\boldsymbol{s}$ and δ is the Dirac delta function, then it is known as a Green's function. The Dirac delta function is a generalised function which has the property that taking the convolution with the delta function

amounts to evaluating the function at a single location. Formally, this can be written as $\int g(\boldsymbol{u})\delta(\boldsymbol{s}-\boldsymbol{u})\mathrm{d}\boldsymbol{u} = g(\boldsymbol{s})$ for any function g. This is the key property that means f will solve the SPDE. To see this note that

$$\begin{aligned}
\epsilon(\boldsymbol{s}) &= \int \delta(\boldsymbol{s}-\boldsymbol{u})\epsilon(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \\
&= \int \mathcal{D}G(\boldsymbol{s}-\boldsymbol{u})\epsilon(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \\
&= \mathcal{D}\left(\int G(\boldsymbol{s}-\boldsymbol{u})\epsilon(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right) \\
&= \mathcal{D}f(\boldsymbol{s})
\end{aligned}$$

Line three follows from Leibniz's rule for integrals since we assume $\mathcal{D}$ is linear and differentiation is with respect to $\boldsymbol{s}$ and not the variable of integration $\boldsymbol{u}$. The question now is how to find a G that satisfies this property. One way to do this is by using Fourier transform methods. This is a useful approach because differentiation and convolution have a simple representations in the frequency domain. Loosely speaking, differentiation corresponds to multiplication by a constant in the frequency domain and the convolution of two functions corresponds to the multiplication of their Fourier transforms. This means that the application of a linear differential operator to a function f can be represented as the Fourier transform of f multiplied by a constant (strictly speaking a polynomial of the frequency variable, the order and form of which depends on $\mathcal{D}$). Similarly, the convolution of two functions is the inverse transform of the product of their individual Fourier transforms.

These properties of the Fourier transform can be used to derive the covariance function that is implied by $\mathcal{D}$. First note that the Green's function can be derived by taking the Fourier transform of both sides of $\mathcal{D}G(\boldsymbol{s}-\boldsymbol{u}) = \delta(\boldsymbol{s}-\boldsymbol{u})$. The Fourier transform of the Dirac delta function is one everywhere and the Fourier transform of $\mathcal{D}G$ is the Fourier transform of G multiplied by some known polynomial. Then by rearranging the equation and taking the inverse Fourier transform we can derive

an expression for G. It also turns out that the covariance function can be written in terms of G.

$$\begin{aligned}
\text{Cov}[f(\boldsymbol{s}_i), f(\boldsymbol{s}_j)] &= \mathbb{E}[f(\boldsymbol{s}_i)f(\boldsymbol{s}_j)] \\
&= \mathbb{E}\left[\int G(\boldsymbol{s}_i - \boldsymbol{u})\epsilon(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \int G(\boldsymbol{s}_j - \boldsymbol{u})\epsilon(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right] \\
&= \mathbb{E}\left[\int G(\boldsymbol{s}_i - \boldsymbol{u})G(\boldsymbol{s}_j - \boldsymbol{u})\mathrm{d}\boldsymbol{u}\right] \\
&= \int G(\boldsymbol{s}_i - \boldsymbol{u})G(\boldsymbol{s}_j - \boldsymbol{u})\mathrm{d}\boldsymbol{u}
\end{aligned}$$

where the third line follows due to Ito's isometry (Oksendal, 2013). This derivation means the covariance function $C(\boldsymbol{s}_i, \boldsymbol{s}_j)$ can be derived directly from the SPDE by taking the convolution of the Green's function with itself. This makes explicit the link between the covariance function of the random field and the differential operator $\mathcal{D}$. A Gaussian random field may be specified using either a covariance function or via an SPDE. They are equivalent ways to specify the same mathematical structure.

Whittle (1954) applied the approach sketched above to a specific SPDE to show that its stationary solutions have Matérn covariance. This representation of the Matérn field is the basis for later work by Lindgren et al. (2011) who use this SPDE representation to derive a computationally efficient Gaussian *Markov* random field approximation to the Matérn field. We return to this specific example in section 2.5. First, we cover the numerical approaches that are used to solve SPDEs.

2.4.1 Solving the SPDE using finite element methods

In order to solve the SPDE we return to the weak solution $\langle \mathcal{D}f, \phi \rangle \stackrel{d}{=} \langle \epsilon, \phi \rangle$ given in (2.8). A solution to the SPDE is a Gaussian random field that satisfies the weak solution for any choice of test function ϕ. In practice, we choose a finite set of test functions $\{\phi_1, \ldots, \phi_m\}$ and arrive at an approximate solution. The test functions we consider here are based on finite element methods. These are a family

of related numerical algorithms for solving (S)PDEs. Chen (2005) gives a wide overview of these methods and Brenner and Scott (2008) provides a mathematical treatment including convergence and error rates. For the 2-dimensional examples in this chapter, we consider finite elements that are piece-wise linear functions defined with respect to a triangulation of the spatial domain. This triangulation defines a graph which is the basis for the GMRF approximation (recall from above that the definition of a Markov random field is always with respect to some graph). In keeping with the literature we refer to this graph as a *mesh*.

The test functions we consider here are piecewise linear functions defined with respect to this mesh. Each node of the mesh is associated with a basis function which takes the value 1 at the node and decreases linearly to zero at the neighbouring edges. Figure 2.1 shows an example mesh and two of these basis functions. Since these basis functions are zero everywhere except in the direct neighbourhood of the node, we have that $\langle \phi_i, \phi_j \rangle = 0$ if nodes i and j are not connected by an edge. We will see later that this has implications for the precision matrix for $\boldsymbol{\beta}$. Note that

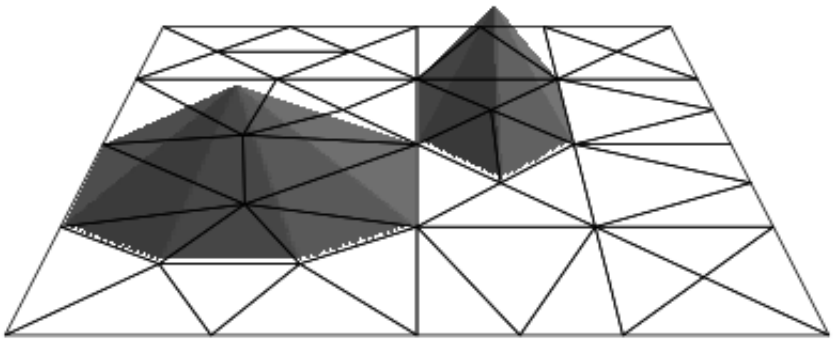


Figure 2.1: An example of a mesh and two of the associated piecewise linear basis functions. These basis functions take a value 1 at the central node, decrease linearly to zero at the neighbouring edges, and are zero everywhere else.

for any location that is not at a node, there are at least two basis functions that are non-zero at that location. This approach to choosing test functions is a way to 'cover' the entire spatial domain in such a way that these piece-wise linear basis functions can be used to approximate non-linear functions, as shown in figure 2.2. The mesh should be chosen at a sufficiently high resolution for the given application.

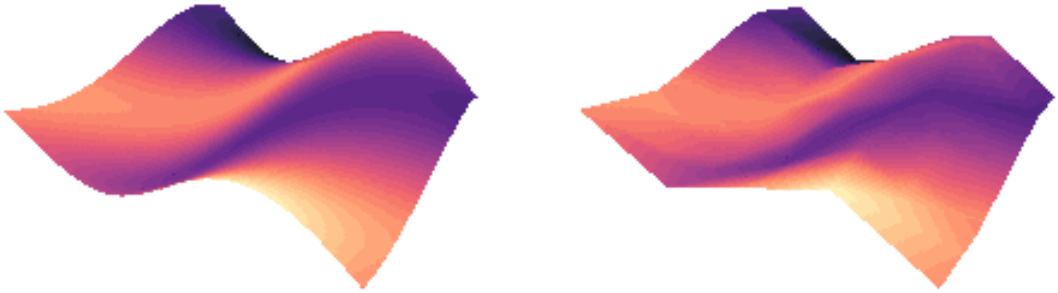


Figure 2.2: **Left:** A smooth function $\sin(x) - (y/\pi)^3$ **Right:** The piecewise linear approximation of $\sin(x) - (y/\pi)^3$ using the mesh-basis shown in figure 2.1.

If a rough approximation is all that is required then a coarse mesh, with relatively few nodes, will be sufficient. A higher resolution mesh can be used for applications where a closer approximation to the smooth function is required.

Given a choice of mesh and associated basis functions, the weak solution for the SPDE is

$$\langle \mathcal{D}f, \phi_j \rangle = \langle \epsilon, \phi_j \rangle \qquad \text{for } j = 1, \ldots, m$$

Up to this point we have been vague about what it means to apply a differential operator to a random field i.e. what exactly is meant by $\mathcal{D}f$. This becomes clear when

we define a finite-dimensional space of functions for f. Here we choose the basis for f to be the same as the set of test functions $\{\phi_1, \ldots, \phi_m\}$. This is referred to as the *Galerkin* solution by Lindgren et al. (2011). Choosing this finite-dimensional basis representation allows us to define the derivative of the random field using standard calculus. Let $f(\boldsymbol{s}) = \sum_{i=1}^{m} \beta_i \phi_i(\boldsymbol{s})$, where each $\phi_i(.)$ is a fixed, non-random function, and each β_i is a random coefficient. Then the derivative is $\mathcal{D}f(\boldsymbol{s}) = \sum_{i=1}^{m} \beta_i \mathcal{D}\phi_i(\boldsymbol{s})$. In other words, the space of derivatives of f has the basis $\{\mathcal{D}\phi_1, \ldots, \mathcal{D}\phi_m\}$. This means that we can represent $\mathcal{D}f$ by computing the derivatives of each of the fixed basis functions and, given a particular choice of mesh, these only need be computed once.

The weak formulation of the SPDE can be represented, given our choice of mesh, test functions, and basis functions for f, as the matrix-vector equation

$$\boldsymbol{D}\boldsymbol{\beta} = \boldsymbol{e} \tag{2.9}$$

where $\boldsymbol{D}$ is an $m \times m$ symmetric matrix with entries $\boldsymbol{D}_{ij} = \langle \mathcal{D}\phi_i, \phi_j \rangle$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_m]^T$ and $\boldsymbol{e}_i = \langle \epsilon, \phi_i \rangle$. One can think of this matrix-vector equation as a finite dimensional discretisation the SPDE. By the properties of the Gaussian white noise process and the fact that $\langle \phi_i, \phi_j \rangle = 0$ for $i \neq j$ it follows that $\boldsymbol{e} \sim \mathcal{N}(0, \boldsymbol{Q_e})$ where $\boldsymbol{Q_e}$ is diagonal with entries $(\boldsymbol{Q_e})_{ii} = \|\phi_i\|$. Since $\boldsymbol{D}$ is non-random this also implies that $\boldsymbol{\beta}$ has multivariate Gaussian density with mean zero and precision matrix $\boldsymbol{Q_\beta} = \boldsymbol{D^T Q_e D}$. Therefore, given a choice of differential operator, the SPDE essentially amounts to placing a multivariate Gaussian prior on the coordinate vector $\boldsymbol{\beta}$. Encoded in the precision matrix $\boldsymbol{Q_\beta}$ is the covariance structure implied by $\mathcal{D}$.

Recall that we arrived at a similar matrix-vector expression for the basis-penalty smoothers. Given a chosen basis and smoothing penalty, this results in a penalty matrix that encodes what we mean by smoothness. Similarly, given a choice of

SPDE and mesh, this results in a precision matrix that describes the covariance structure implied by the SPDE. At the heart of both approaches is the choice of a differential operator $\mathcal{D}$. For smoothing splines, the choice of $\mathcal{D}$ can be thought of as the 'definition of smoothness'. For Gaussian random fields, the choice of $\mathcal{D}$ determines the covariance function.

However, this analogy omits some important details. There is not always a covariance function that is associated with a particular basis-penalty choice for example. Some common choices of basis-penalty smoother, when formulated as an SPDE, would result in non-stationary random fields without succinct covariance functions. The thin-plate regression spline as implemented in `mgcv` is an example of this. However, for our purposes here, the relationship between the penalty matrix (for basis-penalty smoothers) and the precision matrix (for GMRFs) is a useful perspective from which to understand the SPDE method and how to implement it in GAM software such as `mgcv`.

2.5 The SPDE approach for the Matérn field

In this section we introduce the SPDE that defines the Matérn field and summarise the results in Lindgren et al. (2011) , using the perspective and notation introduced above. Whittle (1954) showed that the stationary solutions to the SPDE

$$(\kappa^2 - \nabla^2)^{\alpha/2} f(\boldsymbol{s}) = \frac{\epsilon(\boldsymbol{s})}{\tau} \tag{2.10}$$

have Matérn covariance (2.5) where $\alpha = \nu + d/2$ and d is the dimension of the domain. The operator ∇^2 is the Laplace operator which is defined, in two dimensions, as $\nabla^2 = \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}$. We refer to this SPDE as the *Matérn SPDE*. For most choices of α, the operator $\mathcal{D} = (\kappa^2 - \nabla)^{\alpha/2} f(\boldsymbol{s})$ is called a fractional differential operator since the exponent is a fraction. Here we investigate the special case where $\alpha = 2$ and so $\mathcal{D}$ is a linear differential operator. The finite element solutions for cases with

a fractional differential operator can be defined using recursive equations based on the finite element solutions for $\alpha = 1$ and $\alpha = 2$ (see equation 10 in Lindgren et al. (2011)).

The weak solution for the Matérn SPDE with $\alpha = 2$ is $\langle \kappa^2 f, \phi \rangle - \langle \nabla^2 f, \phi \rangle = \langle \epsilon/\tau, \phi \rangle$. This equation can be simplified by an application of Green's first identity and Neumann boundary conditions. The Neumann boundary conditions state that the first derivative at the boundary, in the direction orthogonal to the boundary, is zero. In this case, by Green's first identity we have $-\langle \nabla^2 f, \phi \rangle = \langle \nabla f, \nabla \phi \rangle$ and we can write the weak solution as $\langle \kappa^2 f, \phi_i \rangle + \langle \nabla f, \nabla \phi_i \rangle = \langle \epsilon/\tau, \phi_i \rangle$ for test functions $\phi_1, \ldots, \phi_m$. For $f = \sum_j \beta_j \phi_j$, this can be written as a matrix-vector equation of the form

$$\kappa^2 \boldsymbol{C}\boldsymbol{\beta} + \boldsymbol{G}\boldsymbol{\beta} = \boldsymbol{e}/\tau,$$

where $\boldsymbol{C}_{ij} = \langle \phi_i, \phi_j \rangle$, $\boldsymbol{G}_{ij} = \langle \nabla \phi_i, \nabla \phi_j \rangle$ and $\boldsymbol{e}_i = \langle \phi_i, \epsilon \rangle$. We can write this as $\boldsymbol{K}\boldsymbol{\beta} = \boldsymbol{e}/\tau$ where $\boldsymbol{K} = \kappa^2 \boldsymbol{C} + \boldsymbol{G}$. The precision matrix $\boldsymbol{Q}_{\boldsymbol{\beta}}$ can be derived from this equation, which we show below. We present a more detailed version of the proof in appendix D.3.1 in Lindgren et al. (2011), with the addition of the τ parameter which is omitted there and some additional explanation. We consider two functions which we represent using the finite element basis as $g = \sum_i g_i \phi_i$ and $h = \sum_i h_i \phi_i$. The derivation for $\boldsymbol{Q}_{\boldsymbol{\beta}}$ rests on deriving $\text{Cov}\left[\langle g, \mathcal{D}f \rangle, \langle h, \mathcal{D}f \rangle\right]$ in two different ways. Firstly, since $\mathcal{D}f = \epsilon/\tau$ we have

$$\begin{aligned}
\text{Cov}\left[\langle g, \mathcal{D}f \rangle, \langle h, \mathcal{D}f \rangle\right] &= \tau^{-2} \,\text{Cov}\left[\langle g, \epsilon \rangle, \langle h, \epsilon \rangle\right] \\
&= \tau^{-2} \,\mathbb{E}\left[\langle g, \epsilon \rangle \langle h, \epsilon \rangle)\right] \\
&= \tau^{-2} \,\langle g, h \rangle \\
&= \tau^{-2} \,\boldsymbol{g}^T \boldsymbol{C} \boldsymbol{h},
\end{aligned}$$

where the third line follows by an application of Ito's isometry. Secondly, note that $\langle g, \mathcal{D}f \rangle = \boldsymbol{g}^T \boldsymbol{K}\boldsymbol{\beta}$ and $\langle h, \mathcal{D}f \rangle = \boldsymbol{h}^T \boldsymbol{K}\boldsymbol{\beta}$. This follows directly from the same argu-

ment as above, applying Green's first identity and Neumann boundary conditions. Therefore,

$$\begin{aligned}\text{Cov}\left[\langle g, \mathcal{D}f\rangle, \langle h, \mathcal{D}f\rangle\right] &= \text{Cov}\left[\boldsymbol{g}^T\boldsymbol{K}\boldsymbol{\beta}, \boldsymbol{h}^T\boldsymbol{K}\boldsymbol{\beta}\right] \\ &= \boldsymbol{g}^T\boldsymbol{K}\text{Cov}\left[\boldsymbol{\beta}, \boldsymbol{\beta}\right]\boldsymbol{K}^T\boldsymbol{h}.\end{aligned}$$

Combining these two expressions, the $\boldsymbol{g}$ and $\boldsymbol{h}$ cancel to give $\text{Cov}\left[\boldsymbol{\beta}, \boldsymbol{\beta}\right] = (1/\tau^2)\boldsymbol{K}^{-1}\boldsymbol{C}\boldsymbol{K}^{-T}$ and so $\boldsymbol{Q}_{\boldsymbol{\beta}} = \tau^2\boldsymbol{K}^T\boldsymbol{C}^{-1}\boldsymbol{K}$. Plugging in $\boldsymbol{K} = \kappa^2\boldsymbol{C} + \boldsymbol{G}$ and rearranging we have

$$\boldsymbol{Q}_{\boldsymbol{\beta}} = \tau^2\left(\kappa^4\boldsymbol{C} + 2\kappa^2\boldsymbol{G} + \boldsymbol{G}^T\boldsymbol{C}^{-1}\boldsymbol{G}\right). \tag{2.11}$$

The matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ are sparse but $\boldsymbol{C}^{-1}$ is not. Lindgren et al. (2011) show that $\boldsymbol{C}$ can be replaced with a diagonal matrix whose diagonal entries are the row sums of $\boldsymbol{C}$ and that this replacement does not affect the rate of convergence of the finite element method to the true solution as the number of mesh nodes tends to infinity. Therefore, it is possible to construct a sparse approximation of the precision matrix.

2.5.1 The connection with the smoothing penalty

We have now introduced two approaches to flexible modelling, the basis-penalty smoothing approach and the Gaussian random field SPDE approach. Given an SPDE with linear differential operator $\mathcal{D}f = \epsilon$, we can apply the above finite element method approach to end up with an expression for $\boldsymbol{Q}_{\boldsymbol{\beta}} = \boldsymbol{D}^T\boldsymbol{Q}_{\boldsymbol{e}}\boldsymbol{D}$ where $\boldsymbol{D}_{ij} = \langle\mathcal{D}\phi_i, \phi_j\rangle$, and derived explicitly for the Matérn SPDE above. Similarly, from the optimal smoothing spline perspective, we can choose a penalty $J_{cD}(f|\lambda)$, choose a basis for f and derive the penalty term as a matrix-vector product $\lambda\boldsymbol{\beta}^T\boldsymbol{S}\boldsymbol{\beta}$ which we can interpret as being equivalent to placing a (possibly improper) multivariate Gaussian prior on $\boldsymbol{\beta}$. In this sense the matrix $\boldsymbol{S}$ is also a precision matrix for $\boldsymbol{\beta}$.

Therefore it seems natural to say that the precision matrix derived from the SPDE could also be derived from an appropriate smoothing penalty.

However, the connection between an SPDE formulation and the basis-penalty smoother is more complicated. The correspondence depends on the details of both the finite-dimensional approximation methods and the inference approach, frequentist or Bayesian. Taking the 'penalty-as-a-prior' viewpoint, we want to view $\boldsymbol{Q}_{\boldsymbol{\beta}}$ as a smoothing matrix and use this to define a smoothing penalty that we could have derived from $\int(\mathcal{D}f)^2$. But this is a simplification. To see the issues here we have to be explicit about the full space of functions, our approach to choosing a finite-dimensional approximation of this space, the meaning of the SPDE, and the connection between SPDEs and maximisers of penalised likelihoods. We illustrate these points using the approximate precision matrix derived above for the Matérn SPDE.

The correspondence between stochastic processes and smoothing splines was first identified by Kimeldorf and Wahba (1970a). However, this correspondence is expressed in a full, infinite dimensional, space of functions (specifically the space of functions with continuous derivatives C^k where k depends on the order of the differential operator) and does not depend on any decision to restrict the space of functions to a finite-dimensional sub-space. For simplicity consider the one-dimensional case where we have $y_i = f(t_i)$ for $i = 1, \ldots, n$. In this case Kimeldorf and Wahba show that the function

$$\hat{f} = \arg\min_f \{\sum_i [y_i - f(t_i)]^2 + J_{\mathcal{D}}(f)\}$$

is equivalent to $\mathbb{E}[f(t)|y_1, \ldots, y_n]$ where now, in a slight abuse of notation, $f(t)$ is a Gaussian random field with mean zero and covariance function K, which we could choose to specify using an SPDE. Note that this correspondence is between the *maximiser* of a penalised likelihood and the *posterior expectation* of a Gaussian random field.

The key to understanding the difference in these two perspectives is to be clear

about what function space we are using in each. The space of functions with continuous derivatives used in the optimal smoothing spline approach is not the same as the space of functions defined by the Gaussian random field. The properties of the GRF realisations are determined by the covariance function, which may or may not allow for continuously differentiable realisations. It is therefore not always easy to state a straightforward correspondence between smoothing penalties and GRF priors once specific choices for the finite-dimensional basis are made.

Defining f via the SPDE $\tau(\kappa^2 - \Delta)f = \epsilon$, the key aspect of this equation to note is that that Δf is not square integrable since Gaussian white noise is not square integrable. That is to say, no *realisation* of the process can be the optimal smoothing spline, since the penalty is not defined for these realisations. To see this, note that

$$\begin{aligned} J_{\mathcal{D}}(f|\kappa,\tau) &= \langle \tau(\kappa^2 - \Delta)f, \tau(\kappa^2 - \Delta)f \rangle \\ &= \tau^2 \left[\kappa^4 \langle f, f \rangle - 2\kappa^2 \langle f, \Delta f \rangle + \langle \Delta f, \Delta f \rangle \right], \end{aligned}$$

which involves the term $\langle \Delta f, \Delta f \rangle$ that is undefined if f is a solution the SPDE (recall that Gaussian white noise is a *generalised* random field that is only defined when integrating against a *deterministic* function, therefore squared white noise is not integrable). This penalty only has a meaning when considering f as a member of an appropriate function space. In this case, the finite element approximation has led to a GMRF whose realisations do not have a meaningful smoothing penalty. Kimeldorf and Wahba (1970a) show that if one only considers the posterior expectation $\hat{f} = \mathbb{E}\left[f(t)|y_1, \ldots, y_n\right]$, then this is the optimal smoothing spline. But the expectation of the process is not a solution to the SPDE. This means that when we justify our finite-dimensional approximation by considering the weak solution to an SPDE via the finite element method, we end up with a precision matrix that we could not derive from the smoothing penalty perspective.

The piece-wise linear basis functions themselves do not satisfy the SPDE either.

Choosing deterministic linear functions means $\Delta\phi_i = 0$ for all i and so the smoothing penalty would reduce to $\tau^2\kappa^4\|\phi_i\|$ for each basis function, thus capturing none of the behaviour of the derivatives. There is a large literature on understanding exactly how this approximation can still be useful even though the basis functions do not satisfy the SPDE themselves. Brenner and Scott (2008) cover the mathematical details of finite element methods in depth, and a large portion of the original Lindgren et al. (2011) paper is concerned with showing convergence rates for these methods applied to the Matérn SPDE.

Ultimately, there is a fundamental difference between considering the whole distribution of solutions to the SPDE or instead considering only the posterior expectation or optimal smoothing spline. This is because there is a difference between an appropriate choice of basis for approximating solutions to the SPDE and an appropriate choice of basis for representing C^k in the optimal smoothing spline framework. These are not always interchangeable; a good choice in one setting does not always translate to a sensible choice in the other. If we are interested only in the posterior expectation, then we can state the correspondence and say that there is a specific penalty which corresponds to the SPDE. These are subtle but important differences.

A full Bayesian analysis using the SPDE formulation should, in principle, be able to generate *realisations* from the posterior process $f|\boldsymbol{y}$, whereas the penalised likelihood approach can only generate a point estimate for $\mathbb{E}[f|\boldsymbol{y}]$, and uncertainty in this point estimate. The finite element basis works well enough in the context of a weak solution to the SPDE but is not meaningful in the context of evaluating a smoothing penalty into a smoothing matrix given some choice of basis.

A natural question to ask is whether, given these differences, is it valid to interpret the finite element precision $\boldsymbol{Q}_{\boldsymbol{\beta}}$ as a smoothing matrix? For the SPDE we have derived the approximation in such a way that we cannot arrive at the same place by writing the smoothing penalty form first and then derive the an approximate smoothing matrix. This is because the approximation was constructed for

the purpose of solving the weak SPDE, not just finding the posterior mode in a smoothing spline optimisation problem. We are co-opting this approximation and the dependence structure it implies and assuming it is a good choice within the penalised likelihood framework. This seems to work well enough in practice (see the examples in Section 2.6). But we have no mathematical analysis to justify this in the sense that we cannot start with a smoothing penalty, choose a basis and arrive at the same precision matrix as was derived in Section 2.5. However, it is also worth noting that we have no mathematical analysis that justifies the use of the SPDE or Matérn field in the first place since usually it is included as a non-mechanistic means to model spatial auto-correlation. Understanding these differences is important when it comes to understanding the approximation methods used to implement GRFs and smoothing splines. However, in practice, at least for the examples we have considered so far, we find that using SPDE precision matrix as though it were a penalty matrix leads to no practical differences in model estimation.

2.5.2 Implementation in `mgcv`

The `mgcv` R package is commonly used software for fitting models that involve basis-penalty smoothers using restricted maximum likelihood. New types of basis-penalty smoothers can be specified by writing functions that build the appropriate smoothing matrix, design matrix and other optional components. We can call functions used for the implementation of the SPDE approach in `R-INLA` within these functions to construct the appropriate matrices. Below is a sketch of the approach for implementing the Matérn SPDE in `mgcv`. Note that for the fully sparse finite element precision matrix, we use the notation $\boldsymbol{C}_0$ for the diagonal matrix which replaces the matrix $\boldsymbol{C}$ in equation (2.11), $\boldsymbol{G}_1 = \boldsymbol{G}$ (as above) and $\boldsymbol{G}_2 = \boldsymbol{G}^T\boldsymbol{C}_0^{-1}\boldsymbol{G}$.

1. **Mesh construction:** Create a finite element mesh using `INLA::inla.mesh.1d` or `INLA::inla.mesh.2d`.

2. **Finite element matrices:** Construct the three matrices, $\boldsymbol{C}_0$, $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ using `INLA::inla.mesh.fem` (denoted `c0`, `g1` and `g2` respectively in the object returned by this function)

3. **Projection matrix:** If required, construct the matrix $\boldsymbol{A}$ to map the basis functions to a set of locations (typically either observation locations or prediction locations), using `INLA::inla.spde.make.A`.

4. **Design matrix:** The design matrix is a combination of the design matrix for any fixed effects and the projection matrix $\boldsymbol{A}$.

5. **Inference:** Use REML to find the optimal κ, τ and $\boldsymbol{\beta}$.

Note that here we have two smoothing parameters τ and κ. To estimate them both simultaneously we can make use of the `paraPen` option in `mgcv` which accepts a matrix defining a linear reparameterisation of the log smoothing parameters. Letting $\lambda_1 = \tau^2\kappa^4$, $\lambda_2 = \tau^2\kappa^2$ and $\lambda_3 = \tau^2$ we have

$$\begin{aligned}\log\lambda_1 &= 2\log\tau + 4\log\kappa\\ \log\lambda_2 &= 2\log\tau + 2\log\kappa\\ \log\lambda_3 &= 2\log\tau\end{aligned}$$

and so, if $\boldsymbol{\theta} = [\log\tau, \log\kappa]^T$, we can represent this reparameterisation as

$$\log\boldsymbol{\lambda} = \boldsymbol{L}\boldsymbol{\theta} \quad \text{where} \quad \boldsymbol{L} = \begin{bmatrix} 2 & 4 \\ 2 & 2 \\ 2 & 0 \end{bmatrix}.$$

In this case, the smoothing parameters returned by `mgcv` are τ and κ but the reparameterisation is used internally. Note that `mgcv` will throw an error if there are more mesh nodes than observations. To get around this it is possible to construct

dummy observations to increase the size of the dataset passed to `gam()` and include a weight vector that gives weight 1 to real observations and 0 to dummy observations.

This parameterisation allows us to implement the SPDE effect directly in `mgcv` and is used in the subsequent examples.

2.6 Examples

This section contains three examples of fitting spatial models with the SPDE approach in `mgcv`. `mgcv` can fit penalised regression models using restricted maximum likelihood (Wood, 2017). This is contrast to the more commonly-used implementation of the SPDE approach in `R-INLA` which implements approximate Bayesian inference. The first example models chlorophyll data in the Aral sea and compares the predictions of the `mgcv` model with the same model fitted in `R-INLA`. The purpose of this is to show that the model fit is very similar with the two implementations; any small differences are likely due to differences in computational methods and choice of prior distributions for the smoothing parameters which `R-INLA` requires but `mgcv` does not allow. The second example fits a model to one year of the British Trust of Ornithology Garden Bird Feeding survey data. These data consist of a continuous positive response with zeroes and so are modelled using the Tweedie distribution, which is available in `mgcv` but not currently available in `R-INLA`. This demonstrates the advantage of being able to adapt the SPDE method to other software implementations. The third example is a fit to trawl data collected in the Bering sea by the National Oceanic and Atmospheric Administration (NOAA) of the United States which also uses the Tweedie distribution.

The purpose of these examples is not to present a full analysis of the data but rather to demonstrate the feasibility of fitting models with the SPDE approach using general purpose GAM fitting software. We do this to emphasise the validity of the perspective presented above, that linked SPDEs and smoothing splines, and

to provide an example implementing the SPDE approach in a new software and inference framework.

2.6.1 Chlorophyll A in the Aral Sea

This example fits a simple model with a Gaussian response to data on Chlorophyll measurements collected by the The National Aeronautics and Space Administration's SeaWifs satellite, averaged over the years 1998-2002. The dataset is available via the R package `gamair` which contains the datasets used in Wood (2017). The data consists of 485 measurements at unique locations in the Aral sea and three NA observations. The response variable is continuous and observation range from approximately 1.9 to 19.2. Plotting the data (Figure 2.3) reveals clear spatial structure which we model using the SPDE approach. The model was fitted using both `mgcv`

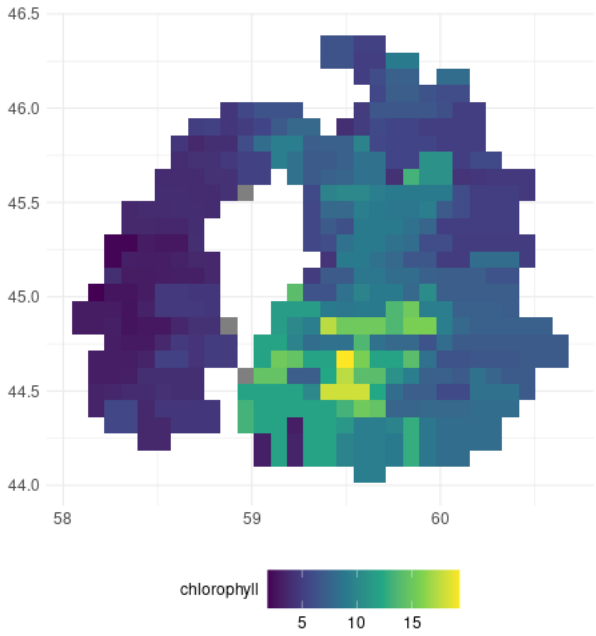


Figure 2.3: Observed chlorophyll measurements in the Aral sea, averaged over 1998-2002

and `R-INLA` to compare the two implementations as a sense check that the `mgcv` implementation is working. The INLA model was fitted using penalised complexity priors on the range and variance parameters of the SPDE (Fuglstad et al., 2019;

Simpson et al., 2017). The prior for the range was set to $\mathbb{P}(\text{SPDE range} < 0.1) = 0.1$ and the prior for the variance was set to $\mathbb{P}(\text{SPDE variance} < 4) = 0.1$. To ensure as fair a comparison as possible, the '`int.strategy = 'eb'`' setting was used in INLA. This setting uses the posterior mode of the SPDE parameters when approximating the subsequent Laplace approximations. We view this as roughly analogous to using only point estimates for the smoothing parameters in `mgcv`. The predicted mean response is shown in Figure 2.4. This shows the predicted mean response is very similar for both models, as we expected since they are approximately fitting the same model (except for priors).

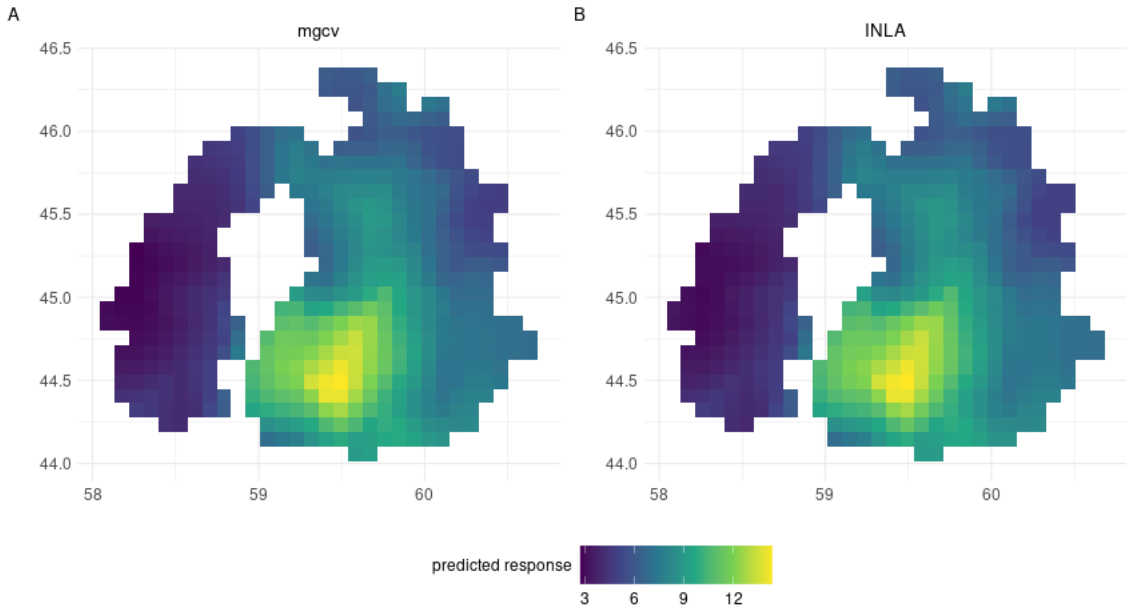


Figure 2.4: Predicted mean chlorophyll for both models. **A**: `mgcv` and **B**: `R-INLA`

To compare the uncertainty estimation between `mgcv` and `INLA` we used the 'Bayesian posterior covariance matrix' as returned by the `mgcv` function `vcov.gam()`. This uses the approach given in Wood (2006), which takes a Bayesian perspective on the smoothing penalty in order to derive a covariance matrix around the maximum likelihood estimate for the spline coefficients. For a Gaussian model the approach is exact in that the function returns the exact posterior covariance matrix that follows from Bayes' theorem (so long as one is willing to accept a possibly improper

prior). For the more general case of a non-Gaussian response, where the response distribution is in the exponential family, the large sample limit is considered which results in an approximate posterior covariance matrix that has a similar construction to the Gaussian case. There are clearly similarities with the approach taken in `INLA`, which uses repeated Laplace approximations to approximate the posterior, in comparison to the Gaussian large sample approximation used in `mgcv`. Here we restrict ourselves to a comparison of uncertainty only for this example rather than give a detailed mathematical comparison, which could be the focus of future research.

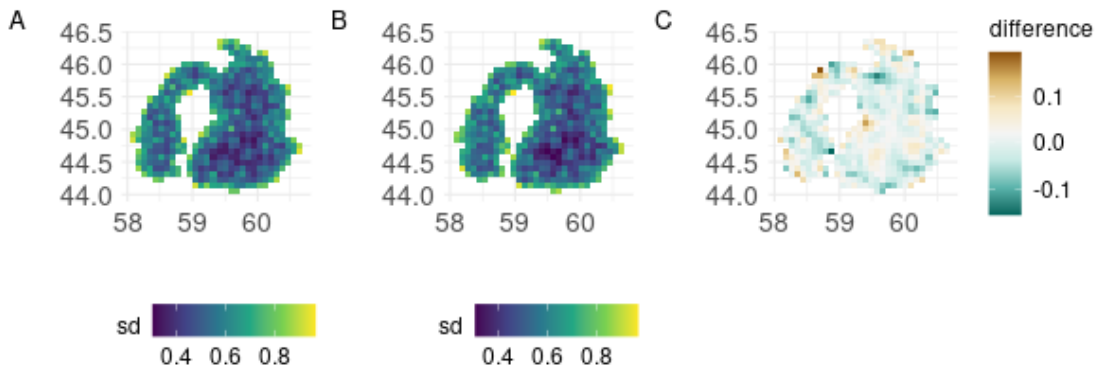


Figure 2.5: Posterior standard deviation in chlorophyll for the **A**: `mgcv` model, **B**: `R-INLA` model, **C**: the difference between the `mgcv` and `INLA` models.

Figure 2.5 plots the posterior standard deviation for the `mgcv` fit alongside the model fitted using `INLA`. Note, however, that `INLA` is fully Bayesian and so includes a prior distribution on the smoothing parameters. Therefore the models are not exactly equivalent. We can see some differences in the estimated standard deviation (Figure 2.5C) which may be due to this difference or the differences in the approximate inference mentioned above. The standard deviation maps appear qualitatively similar but the differences can be significant in some pixels. If the standard deviation is of particular interest in an analysis then the differences in model structure (priors

on smoothing parameters or not) and inference method (Laplace approximations or Gaussian large sample approximation) should be investigated to check whether these choices affect the conclusions of the analysis.

2.6.2 Coal Tit data in England and Wales

This example fits data derived from the British Trust for Ornithology's Garden Bird Feeding Survey which started in the winter of 1970/71 and runs to the present day. We use a cleaned subset of the data that was analysed in a joint-species modelling context in Swallow et al. (2016) and the data are available for download (Swallow et al., 2020). For our purposes, we selected data from a single year and a single species - the Coal Tit in 2004/05. The response is the mean weekly maximum bird count for each site in the survey. This data example was chosen because it has evidence of some spatial structure but also includes a substantial number of zeroes. Since the aim here is to demonstrate a smoothing application, the relatively sparse number of sites in Scotland and Northern Island were left out to avoid over-extrapolation in those regions, leaving only the records in England and Wales, where the coverage is more substantial (Figure 2.6A). This subset of the data consists of 220 observations of which 22 were zeroes and a maximum observed value of 5.2. Since the response is a non-negative continuous variable we choose the Tweedie distribution, which can allow for a point mass probability at zero as well as a continuous density for positive values. This is a response distribution that is available in `mgcv` but not in `R-INLA`.

The predicted mean weekly maximum bird count is shown in Figure 2.6B. There is some evidence that the model is failing to capture the largest observed values, perhaps undersmoothing and fitting to the bulk of the observations. The finite element mesh used to construct the Markov approximation is shown in Figure 2.7. In contrast to the chlorophyll A example, here the spline is used to describe a continuous latent surface that does not physically exist. Instead the spline is used to

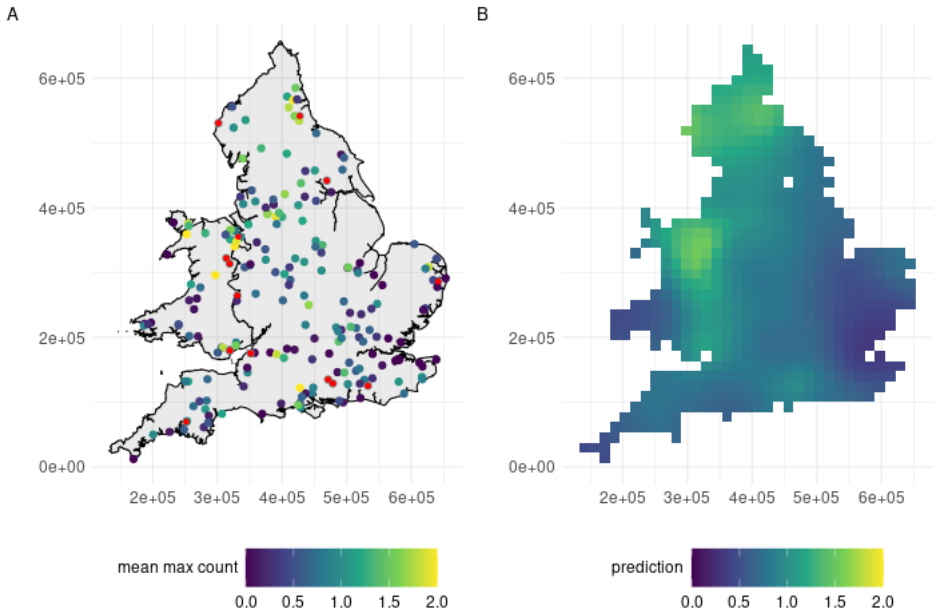


Figure 2.6: **A:** Coal Tit mean maximum weekly count for each site in the Garden Bird Feeding survey in England and Wales in 2004/05. The blue-yellow colour scale ranges from 0 to 2 and was chosen to convey the spatial structure in the majority of the observations. There were 14 observations greater than 2 which are marked in red. **B:** The mean predicted mean maximum weekly count on a prediction grid that covers England and Wales.

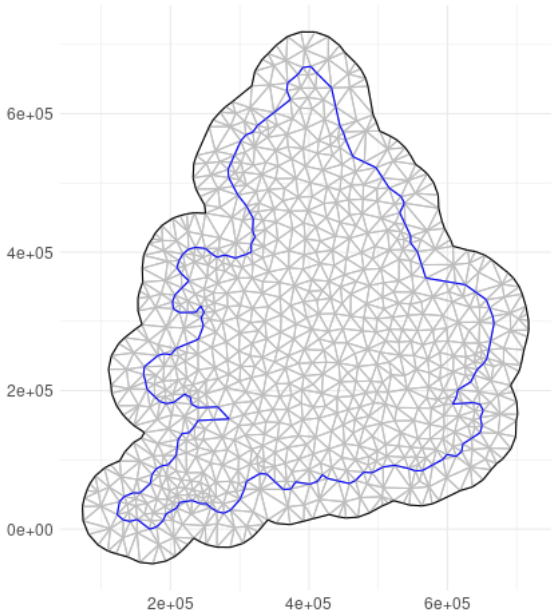


Figure 2.7: The finite element mesh used to construct the Markov approximation to the SPDE for the Coal Tit data analysis

model correlation between measurements made at discrete units (people's gardens).

This example shows we can use the SPDE approach in `mgcv` to use response distributions (in this case the Tweedie distribution) that are not available in `R-INLA`.

2.6.3 Bering Sea trawl data

NOAA's Alaska Fisheries Science Center regularly conducts bottom trawl surveys to assess the condition of groundfish and shellfish stocks in Alaskan waters to inform the management of these fisheries. The data are freely available to download from the NOAA website. We choose a subset of the data collected in the Bering sea. Similar to the above example, for ease of exposition we considered a single species in a single year - the Northern Rock Sole in 2016. Again this example was chosen as the data showed clear evidence of spatial structure suitable for smoothing. There are 309 observations in total, shown in Figure 2.8A. The response variable is NUMCPUE, the catch number per area the net swept (count per hectare). This is again a non-negative positive response variable so we modelled the data with the Tweedie distribution. The predicted response is shown in Figure 2.8B. The mesh used to construct the Markov approximation is shown in Figure 2.9.

Again this shows that our understanding of the SPDE approach has allowed us to fit models to real ecological data that we could not previously have fitted. We did this by using our understanding to implement the SPDE approach as a basis-penalty smoother in `mgcv` which allowed us to use the Tweedie distribution, a popular choice in spatial ecology due its ability to model continuous data with zeroes, as can often occur in ecological data.

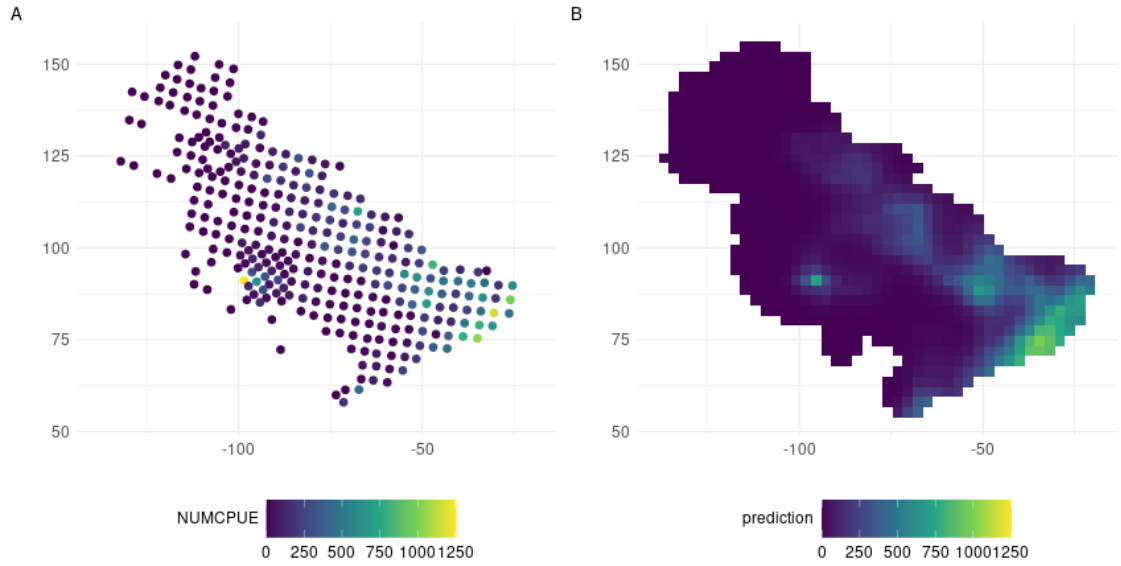


Figure 2.8: **A**: The observed number of northern rock sole caught per unit area swept by the net. **B**: The expected number of fish caught per unit area predicted by the model

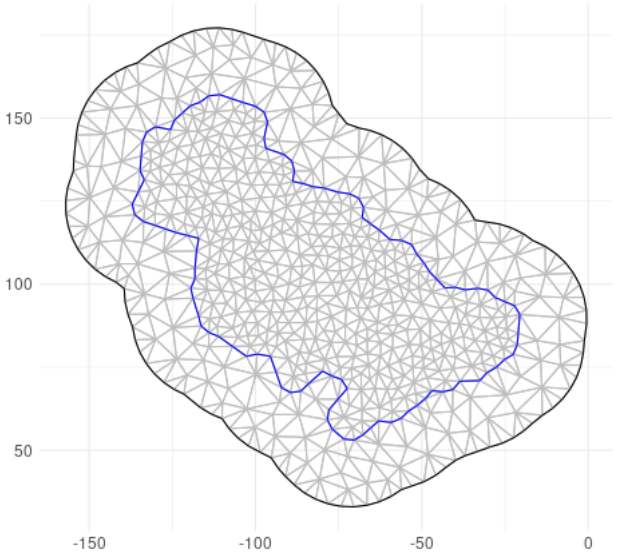


Figure 2.9: The finite element mesh used to construct the Markov approximation to the SPDE for the trawl data analysis

2.7 Discussion

These three examples have shown that it is possible to specify the SPDE model within `mgcv` and obtain results that are broadly speaking sensible. This opens up the possibility of combining the SPDE with the wide range of GAM modelling techniques available in `mgcv`.

This chapter has drawn links between two popular approaches for accounting for structure in data. The aim is to provide applied statisticians with an accessible understanding of what is equivalent between the two approaches and what is fundamentally different. This was done from the perspective of translating the SPDE literature into the language of basis-penalty smoothers. The reverse approach was taken in Yue et al. (2014) who show how adaptive smoothing splines can be implemented as SPDEs.

There exist other smoothing techniques with which comparisons can also be drawn. Smoothing splines, kriging, Gaussian random fields (specified more traditionally through covariance functions) and SPDEs have all been applied in spatial statistics contexts to achieve similar aims. The wide and varied histories of these approaches and the terminology used in the literature to describe them can lead to difficulties for practitioners when deciding which method is suitable for a given application. Taking an approximate precision matrix as the common currency between all these approaches is one way to view their commonalities and their differences. This can be summarised in the following workflow:

1. **Choose a covariance model:** either directly by choosing a covariance function or implicitly by a smoothing penalty or SPDE.

2. **Approximate the precision matrix**: for example by dimension reduction techniques (fixed rank kriging, thin plate splines) or induce sparsity (SPDE approach, B-splines).

3. **Conduct inference using a specific choice of software:** for example

with `mgcv` or `R-INLA` as considered in this article, or other software such as `TMB` (Kristensen et al., 2016), `Stan` (Carpenter et al., 2017) or `JAGS` (Plummer, 2017).

This conceptual separation of the modelling choices allows users to compare methods by identifying genuine differences in the covariance model, approximation technique and inference procedure. This may go some way to allow users to pick the tools best suited to them, given their existing expertise and familiarity with the methods and software packages.

The reason `R-INLA` and `TMB` are the go-to choices for fitting the SPDE approach is that these software packages make use of libraries that can take advantage of the sparsity in the precision matrix. This leads to efficient computation and the ability to fit models that other packages such as `mgcv`, `JAGS` and `Stan`, which do not utilise the sparsity, would struggle to fit with the same computing resources. Should these packages extend functionality to take advantage of sparse precision matrices, we would expect an increase in the number of models that use sparse approximation techniques, such as the SPDE approach, to follow. For large numbers of observations (though not necessarily large numbers of mesh nodes) there is the `bam()` function in `mgcv` which was developed to allow the use of datasets which contain upwards of tens of thousands of observations. An example of fitting the SPDE using `bam()` to high resolution land surface temperature data is given in Miller et al. (2019).

Chapter 3

One-stage point transect distance sampling using iterated integrated nested Laplace approximations

3.1 Introduction

This chapter demonstrates an application of the SPDE approach applied to the problem of estimating spatial variation in animal abundance. In particular, we use the SPDE effect to model the intensity of a point process that represents the locations of animal sightings. Often such sightings are made by observers who cannot detect all the animals perfectly and so, in order to estimate abundance, we require an estimate of the number of animals that remained undetected. Distance sampling (Buckland et al., 2015) is a popular approach to estimating the detectability of animals and is the focus of this chapter. We present a one-stage distance sampling model, simultaneously estimating the detectability and spatial distribution of animals. The spatial distribution of animals is modelled as a point process with an SPDE effect to describe spatial variation in animal density. The discussion and implementation of these methods is built around a case study analysing distance

sampling data collected for an endangered tropical bird species. This case study provides a useful context for the statistical methods discussed in this chapter and demonstrates their utility in applied spatial statistics.

The estimation of the size and spatial distribution of wild populations of animals is a critical objective within ecology and conservation (Schwarz and Seber, 1999). Estimates of these quantities are a critical step in generating evidence that contributes to answering questions in ecological science and conservation management. We present an analysis of wildlife survey data collected on a critically endangered Hawai'ian forest bird: the Hawai'i 'ākepa (hereafter 'ākepa; *Loxops coccineus*; nomenclature according to U.S. Fish and Wildlife Service, 1970), an endemic species whose population declined dramatically during the 20th century (U.S. Fish and Wildlife Service, 2006; Judge et al., 2018). The remaining population is the focus of sustained conservation efforts and monitoring is required to inform decision-makers about changes in the overall abundance and spatial distribution.

However, estimating the size and spatial distribution of animal populations presents many statistical challenges. Firstly, as in many ecological surveys, it is impossible to undertake a full census (i.e. the complete enumeration of all individuals within a defined study region). The existing population numbers in the thousands and lives in dense tropical forest that makes a census infeasible. For this reason, a monitoring survey must sub-sample appropriately in space and time. For the 'ākepa this takes the form of the Hawai'iForest Bird Survey (HFBS) (Scott et al., 1986), which is a large-scale, quantitative survey of Hawai'ian forest birds that, in the region with the largest 'ākepa population, consists of point transects located along randomly placed line transects.

Secondly, even at sampled locations, the detectability of animals is unknown. The 'ākepa survey estimates detectability using a point transect distance sampling protocol (Buckland et al., 2015) where, for each observation, the distance to a stationary observer is recorded. The approach assumes a parametric form of detection

function that decays with increasing distance. The detection function parameters are estimated from the observed distances, assuming the true density is uniform. This combination of spatio-temporal sub-sampling and unknown detectability means statistical methods used to analyse such data must model a complex observation process alongside the spatio-temporal distribution of animals.

Thirdly, a key aim of the analysis is to predict animal density at un-sampled locations. If density is modelled using spatial covariates the usual approach is to assume the relationship between animal density and spatial covariates is the same in the un-sampled locations as it is in the sampled locations. Predictions can then be based on estimating this relationship using the observed data and extrapolating to un-sampled regions. However, we usually have reason to believe there are drivers of the spatial distribution for which we have no explanatory covariates available or for which, in principle, no covariate could be constructed (e.g. the 'sociality' of the species or other behavioural characteristics). For the 'ākepa data, one example of an unexplained driver of the spatial distribution is a north-south gradient in density that has been investigated but for which no straightforward explanatory cause has been found (Camp et al., 2020). From a statistical perspective, this suggests the inclusion of a spatially-structured random effect in the model, which necessitates the use of appropriate statistical software capable of fitting such models.

Lastly, the above challenges mean the resulting statistical model is necessarily complex and it is therefore challenging to communicate results of the analysis. The data has been collected with the clear objective of monitoring the population and informing conservation management decisions. These decisions will be taken by stakeholders with a variety of statistical expertise who should have a good understanding of the full range of possible inferences implied by the model, even if they do not understand all the statistical technicalities involved. This final step of communicating results is therefore key and should be considered with as much care as the rest of the analysis. This requires the input of statisticians to ensure uncertainty is

appropriately incorporated into the key outputs of the analysis upon which decisions will be based.

In particular, we note that communicating uncertainty in maps of predicted animal density is challenging. Here we take a Bayesian approach to model estimation and, in this context, current approaches to visualise uncertainty in the posterior random field that describes animal density can be hard to interpret or mask certain properties that could be relevant. For example, maps of the posterior *mean* will always be smoother than *realisations* of the posterior intensity field itself; maps of the posterior coefficient of variation (CV) are hard to interpret when there is spatial variation in density and survey effort; and mapped posterior quantiles at each prediction location are tempting to interpret jointly when they should be viewed independently. These challenges mean careful consideration must be given to the communication of uncertainty in maps of animal density otherwise the risk is that key model outputs will be difficult, if not impossible, for even statistically adept audiences to interpret.

Here we provide an analysis of wildlife survey data that seeks to address all of these statistical challenges by presenting

(i) A model-based spatial point process perspective on point transect distance sampling, representing distance sampling as a thinning of a point processes.

(ii) A one-stage approximate Bayesian approach to inference to simultaneously estimate the observation model and spatial distribution by iterated model fits using integrated nested Laplace approximations (INLA) (Rue et al., 2009).

(iii) A computationally efficient Gaussian Markov random field (GMRF) spatially structured random effect specified using the SPDE approach (Lindgren et al., 2011) to account for unobserved drivers of the spatial distribution.

(iv) Examples of model evaluation and communication of results that take advantage of the one-stage approach by sampling from the joint posterior of all

model parameters. In particular, we consider various ways to work with samples of the posterior intensity field such as using excursions methods (Bolin and Lindgren, 2015) and some novel summary metrics to measure information gained by survey effort.

We demonstrate several novel contributions to the problem of species distribution modelling under unknown detectability. These include: a point process perspective that accounts for incomplete location information; a one-stage approximate Bayesian inference strategy; the use of excursion sets and functions to communicate uncertainty in predictive maps of animal density; and estimating the information gained by survey effort that is implied by the covariance of the posterior intensity field. We note that the challenges we have addressed are generic to many types of wildlife survey data, not only the 'ākepa survey, and can be tackled using a range of possible statistical approaches. Our analysis here is one possible choice of approach that we hope it is generalisable to other contexts.

The rest of the chapter proceeds as follows: (i) we describe the 'ākepa study-region and survey design; (ii) we present the perspective of distance sampling as a thinned point process; (iii) we describe the GMRF random effect and iterated INLA fitting procedure; (iv) we present the results of the analysis and discuss model evaluation and communication.

The data and code used to fit the model and produce the figures in this chapter is freely available to download at `https://github.com/ASeatonSpatial/ptds-public`.

3.2 Study design

The Hawai'i 'ākepa is an internationally and federally endangered Hawaiian honeycreeper (U.S. Fish and Wildlife Service, 1970; IUCN, 2016) that is endemic to Hawai'i Island, USA. Large-scale, quantitative surveys of Hawaiian forest birds and

their habitat commenced in the mid-1970s through the HFBS (Scott et al., 1986). Information from the HFBS is used to update the listing and delisting of endangered species, and establish preserves that coincided with native bird hotspots, including Hakalau Forest National Wildlife Refuge on Hawai'i Island (hereafter Hakalau). Hakalau is the first wildlife refuge established in Hawai'i with the primary purpose to protect, conserve and manage native forests for threatened and endangered bird and plant species. Due to its broad-scale coverage and robust design, the HFBS has become an invaluable resource used to determine changes in species' spatial distributions and population sizes over time.

During the 20th century 'ākepa declined dramatically due to habitat modification (Scott et al., 1986; Pratt, 1994), mosquito-transmitted avian diseases (Pratt, 1994; Atkinson et al., 1995), introduced predators (Lepson and Freed, 1997), and food resources competitors (Lepson and Freed, 1997). 'Ākepa has a global abundance of approximately 16,200 (95%CI 10,000–25,200) in five spatially distinct populations (Judge et al., 2018). Hakalau supports the largest 'ākepa population that, in 2012, was estimated at more than 11,000 birds (Camp et al., 2016). Maintaining and expanding the 'ākepa population at Hakalau is a critical conservation objective. Therefore, unbiased and precise abundance estimates are required by land and resource managers in order to evaluate past management actions and plan for future efforts.

3.2.1 Study area and survey design

Hakalau was established in 1985 to conserve 15,390 ha of montane forest habitat for native forest birds and rainforest plants. Annual forest bird surveys were initiated in 1987 to determine population status and track trends in abundance. Survey points were established along 14 line transects following a systematic, random design with point transects approximately 150 m apart on line transects located either 500 or 1,000 m apart. We limit our study area to the open-forest and closed-forest strata

of Hakalau (Figure 3.1), an extension of the area considered in Camp et al. (2010, 2016), who omitted the closed-forest stratum because it was not sampled in the early years of the 'ākepa surveys. For our analysis, we use data from a later year in which the closed-forest stratum was sampled and so we are able to include it in our analysis. The open-forest stratum was previously heavily grazed, and, since the removal of cattle in 1988, regeneration has proceeded naturally (Maxfield, 1998). The closed-forest stratum was historically least modified by grazing and is relatively intact forest habitat. To the north, the study area follows the refuge boundary while to the east it is bounded by a fence line (Fig. 3.1). The southern boundary is the same as that in Camp et al. (2010), chosen to exclude un-sample regions to the south of Hakalau but does not represent a physical boundary. To the west of the study area is pasture that is dominated by grass and is unsuitable habitat for 'ākepa.

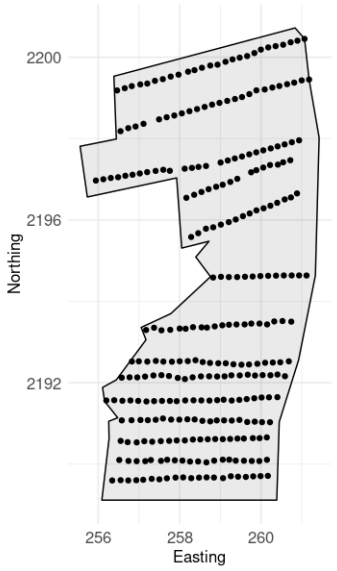


Figure 3.1: Study area showing the 2002 survey points (black dots) in Hakalau Forest National Wildlife Refuge, Hawai'i Island. Eastings and Northings are in kilometres.

The survey uses point-transect methods, recording the horizontal distances from observers to detected birds. Surveys commenced at dawn and continued until 11:00 or halted when weather conditions exceeded prescribed conditions that hindered detecting birds (light rain, and wind and gust strength >Baufort scale 3). During 8-minute counts, trained observers recorded the species, distance to the nearest metre

and detection type for each bird detected, along with the sampling conditions cloud cover, rain, wind strength, gust strength, and time of day each point was surveyed. Camp et al. (2010, 2016) provide a detailed description of Hakalau, the study area and the bird surveys.

The survey data is available in the open forest stratum is available for download (Camp, 2020) and the full time series data was analysed in a spatio-temporal context in (Camp et al., 2020). For the purposes of our analysis we select a single survey year from the 'ākepa time series (2002) that contains a broad sampling of the study area with sufficient numbers of detections to estimate detectability. In 2002, 289 points were sampled using point-transect distance sampling methods within the 4,603 ha study area of Hakalau (Fig. 3.1) during which 276 'ākepa were detected on 121 point transects. The number of detections within each point transect ranged from zero to six. We select data from a single year to demonstrate our approach in a simplified setting that does not require a temporal model component. However, our approach is readily extendible to a multi-year analysis which we discuss in more detail in Section 3.9.

3.3 Distance sampling as a thinned point process

3.3.1 Overview of distance sampling methods

Distance sampling methods aim to estimate abundance by using a spatially explicit sampling design and an assumed detection model to estimate the detectability of animals as a function of distance from transect (Buckland et al., 2004, 2015). Classic distance sampling approaches use a hybrid of design- and model-based inference to estimate population size. The probability of detection is modelled using a parametric equation and, given estimates of detectability, a randomized sampling design allows for the construction of Horvitz-Thompson-like estimators of animal density (Buckland et al., 2004; Horvitz and Thompson, 1952).

More recently, interest has focused on fully model-based approaches that include a spatially explicit model for animal density and allow for the use of non-random survey designs. These methods allow animal density to be associated with spatially-indexed covariates. Therefore, animal density can, in principle, be estimated or predicted for any subregion within the study area (Johnson et al., 2010; Miller et al., 2013; Buckland et al., 2016). This additional flexibility to make predictions for smaller, user-specified geographic units, and extrapolate relationships to unsurveyed regions has made model-based distance sampling a common approach in the literature (García-Barón et al., 2019; Herr et al., 2019; Breen et al., 2017; Williams et al., 2011; Stokes et al., 2010; Williams et al., 2006).

Model-based distance sampling has been implemented in a two-stage modelling framework. The `dsm` package (Miller et al., 2013), for example, provides tools to do this. In the first stage, detectability is estimated. In the second stage, detectability estimates from the first stage are used as an offset in a generalized additive model framework. Detections within sampling units are binned into counts and an appropriate response distribution is chosen to model these counts. Due to the often sparse nature of wildlife survey data, this may require consideration of over- or under-dispersion and zero-inflated distributions. The negative-binomial and Tweedie distributions are common choices here. This two-stage approach has come under the name *density surface models* and Miller et al. (2013) provide a review.

A key concern with the two-stage approach is the propagation of uncertainty from the detection model to the second-stage spatial model. Early attempts to address this focused on bootstrapping (Lahiri, 2003; Hedley and Buckland, 2004) but more recent work has pointed to potential difficulties with the bootstrapping approach, noting the difficulty of choosing the resampling units when combining spatially structured random effects and spatial bootstraps (Bravington et al., 2021; Williams et al., 2011). Instead, Bravington et al. (2021) propose avoiding bootstrapping by propagating error based on a second-order Taylor approximation of

detectability around the first-stage maximum-likelihood estimate.

Concerns about uncertainty propagation can be avoided by using one-stage modelling approaches. A popular one-stage maximum-likelihood method was introduced by Royle et al. (2004) and is implemented in the `R` package `unmarked` (Fiske and Chandler, 2011). This approach depends choosing discrete distance classes and binning the data into counts for discrete spatial units. This allows the likelihood to be written in a multinomial form and, by marginalising out site specific abundance, results in a Poisson likelihood that can be optimised to achieve one-stage inference. This approach rests heavily on the discretisation of both the distance data and the spatial location data. Below we present a point process perspective that does not require this discretisation step in either the distance or spatial location data.

Bayesian one-stage approaches have tended to use Markov chain Monte Carlo (MCMC) methods for inference along with data augmentation to model unobserved individuals or groups (Schmidt et al., 2012). Oedekoven et al. (2014) present a one-stage model that avoids data augmentation by specifying a combined likelihood of the detection and spatially-explicit count models and incorporated model uncertainty using reversible jump MCMC.

The only Bayesian one-stage analysis that does not use MCMC is, to the best of our knowledge, Yuan et al. (2017), who use INLA (Rue et al., 2009) and present an application to line transect distance sampling data. Yuan et al. (2017) also take a point process perspective and formulate the detection model as a thinning of a point pattern. Key to their approach is formulating the detection model as the solution to an SPDE. Choosing a B-spline basis and applying finite element methods results in a sparse prior precision matrix for the detection model coefficients which can then be incorporated into INLA using the same machinery as already exists for working with sparse precision matrices.

A major downside to this approach is that the solutions to the detection model SPDE are not necessarily monotonically decreasing, usually a key feature of detec-

tion functions. The authors' suggestion to reject any non-monotonically decreasing functions when sampling from the posterior is potentially computationally wasteful. Representing the observation model as an SPDE also differs from other distance sampling models that use a parametric model for the detection function. Such parametric models are readily extendible by including covariates and random effects to explain variation in detectability, something that does not seem straightforward to do using the SPDE formulation. These differences with traditional distance sampling methods have perhaps lead to low uptake amongst practitioners of this approach, despite the computational advantages of INLA and the benefits of a one-stage modelling process.

We present a related approach to the one presented in Yuan et al. (2017). We fit the model using INLA, realising the benefits of the one-stage fit and the computational efficiency over MCMC. However, we avoid the issues with the SPDE detection model, instead allowing the user to specify a parametric family of detection functions, such as the half-normal detection function. This approach of choosing a parametric family of detection functions will be more familiar to users of classical distance sampling methods than the SPDE specification. However, this parametric form results in components of the additive predictor that are non-linear in their parameters, making model fitting infeasible in the R-INLA (Rue et al., 2009) package implementation of INLA. We address this by using a method of iterated model fits based on a Taylor expansion of the non-linear model components and implemented in the `inlabru` package (Bachl et al., 2019).

We also apply these methods to point transect 'ākepa data, which differ from line transect methods as models need to account for the fact that the area surveyed increases with increasing distance from the transect. Our analysis is, to the best of our knowledge, the first analysis of point transect distance sampling data formulated as a thinned point process. However, the point process viewpoint is not new and has been taken numerous times in analyses of line transect data (Buckland et al., 2016;

Niemi and Fernández, 2010; Johnson et al., 2010; Waagepetersen and Schweder, 2006; Hedley and Buckland, 2004; Högmander, 1991; Stoyan, 1982).

3.3.2 Model specification

In this section we introduce the statistical model for the spatial pattern of animal locations and the observation process, including a detailed description of how this model can be written as a modified Poisson likelihood. We use this formulation to fit the model using the software package `inlabru` (Bachl et al., 2019), an extension to the `R-INLA` implementation of INLA, which implements the iterated INLA approach and is available through the Comprehensive R Archive Network (R Core Team, 2017).

We assume the locations of animals form a point pattern that follows a log-Gaussian Cox process with intensity process $\lambda(.)$, i.e. the intensity at location $\boldsymbol{s}$ is a random variable $\lambda(\boldsymbol{s})$. The log-Gaussian Cox process is a flexible approach that can include spatial covariates to model the mean intensity and a mean-zero spatially structured random effect to account for unexplained heterogeneity not captured by the covariates (Møller et al., 1998).

To account for the imperfect detection of points we specify a thinning probability function $g(\boldsymbol{s}) = \mathbb{P}(\text{a point at } \boldsymbol{s} \text{ is detected} \mid \text{a point is at } \boldsymbol{s})$. A key property of the log-Gaussian Cox process is that a realisation of a point process with intensity process $\lambda(\boldsymbol{s})$ that is thinned by thinning probability function $g(\boldsymbol{s})$ also follows a log-Gaussian Cox process with intensity given by $\tilde{\lambda}(\boldsymbol{s}) = \lambda(\boldsymbol{s})g(\boldsymbol{s})$.

Standard distance sampling approaches specify $g(\boldsymbol{s})$ as a function that declines with increasing distance. The type of distance measured depends on the type of survey. For line transects, the perpendicular distance to the transect line is used, and, for point transects, the relevant distance is the horizontal distance to the observer. For the remainder of the chapter we assume a point transect survey design and hence distance refers to the horizontal distance to the observer located at the

centre of the point transect.

The thinning probability function is specified as a parametric family of functions. For example, if $r(\boldsymbol{s})$ denotes the distance of a point at $\boldsymbol{s}$ from the observer, the half-normal thinning probability function is $g(\boldsymbol{s}|\sigma) = \exp(-r(\boldsymbol{s})^2/2\sigma^2)$, where $\sigma^2 > 0$ is a scale parameter to be estimated using observed distances. There are other parametric families of detection function, such as the hazard-rate and negative exponential functions, and some additional flexibility can be added to these models by including trigonometric series expansions (Buckland et al., 2015). In this analysis we assume a half-normal detection function with no covariates on detection parameters. However, covariates can, in principle, be included in a straightforward way.

The detection function parameters can only be estimated if an assumption is made about the intensity of the animal locations since, without such an assumption, detectability and intensity are confounded. The standard assumption in distance sampling is that the intensity is constant with respect to distances $r(\boldsymbol{s})$. Thus any observed deviations from uniformity can be attributed to detectability and not to variation in the intensity. This implies we need some way to formulate this standard distance sampling assumption within the context of the point process perspective.

The simplest way to achieve a uniform distribution with respect to $r(\boldsymbol{s})$ is to specify that the intensity itself is uniform with respect to $\boldsymbol{s}$ within each transect. However, it is possible to relax this slightly by allowing the intensity to be a linear function of $\boldsymbol{s}$ within each transect (see Section 3.3.3). Informally, if we imagine the observer looking in a particular direction with increasing intensity, this increase is directly offset by an equal decrease looking in the opposite direction, thus cancelling each other out and resulting in a uniform distribution for distances, irrespective of direction. We use this linear assumption for the intensity which works well with the piece-wise linear basis functions used for the SPDE effect and the small size of the point transects relative to the total study region.

A point transect distance sampling survey consists of a set of K point transects. We denote the k-th subset of space covered by the k-th point transect as $\Omega_k \subset \mathbb{R}^2$ and the total surveyed region is $\Omega = \cup_{k=1}^{K} \Omega_k$. For simplicity, we assume that all point transects are non-overlapping discs with equal radius W, as is the case in the HFBS survey. The probability of observing a point at location $\boldsymbol{s} \in \Omega_k$ given it is at $\boldsymbol{s} \in \Omega_k$ we denote $g_k(\boldsymbol{s})$. The probability of observing a point outside the surveyed region is zero. Since the point transects are non-overlapping, each location $\boldsymbol{s} \in \Omega$ is unambigiously associated with a single thinning probability function g_k. For example, for an observer at location $\boldsymbol{s}_k \in \Omega_k$, the centre of the disc, the thinning function is $g_k(\boldsymbol{s}) = \exp(-r_k(\boldsymbol{s})^2/2\sigma^2)$ where $r_k(\boldsymbol{s}) = \|\boldsymbol{s} - \boldsymbol{s}_k\|$. In practice the observer is always located at the centre of the point transect. The assumption of non-overlapping survey regions could be relaxed by including extra information such as the time of each observation such that each observation is unambiguously linked with a specific thinning probability function. The thinning probability function for any location $\boldsymbol{s} \in \Omega$ is then given by $g(\boldsymbol{s}) := g_{k(\boldsymbol{s})}(\boldsymbol{s})$ where $k(\boldsymbol{s})$ is an indexing function such that $k(\boldsymbol{s}) = k$ for $\boldsymbol{s} \in \Omega_k$.

The thinned log-Gaussian Cox process likelihood with observed points at locations $\boldsymbol{Y} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)^\intercal$ is then

$$\pi(\boldsymbol{Y}) = \exp\left(|\Omega| - \int_{\boldsymbol{s}\in\Omega} \lambda(\boldsymbol{s})g(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right) \prod_{i=1}^{n} \lambda(\boldsymbol{s}_i)g(\boldsymbol{s}_i), \tag{3.1}$$

where the intensity $\lambda(.)$ and thinning probability $g(.)$ both depend on parameters, omitted for readability, and $|\Omega|$ is the area of the total surveyed region.

3.3.3 Writing the model as a modified Poisson likelihood

Here we show how to approximate the thinned log Gaussian Cox process likelihood with a modified Poisson likelihood. The integral component, $\int_{\boldsymbol{s}\in\Omega} \lambda(\boldsymbol{s})g(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$, of the likelihood (3.1) does not have analytical solutions. However, replacing the integral

with a weighted sum approximation allows the log-likelihood to be approximated as a weighted Poisson log-likelihood. We adapt the approach in Simpson et al. (2016) for a fully observed point pattern to the point transect distance sampling context required here.

To evaluate the integral we use polar coordinates notation $\boldsymbol{s}_k(r,\theta) = \boldsymbol{s}_k + r\left[\cos\theta, \sin\theta\right]^T$ to represent locations in each sampling unit Ω_k using distance r from $\boldsymbol{s}_k$ and angle θ. In distance sampling applications the thinning function depends only on r and not on θ so we adopt the shorthand $g_k(r) = g(\boldsymbol{s}_k(r,\theta))$. The integral in (3.1) can be simplified using the assumption that $\lambda(\boldsymbol{s})$ is linear within each sampling unit. This assumption implies that, for any angle θ, $\lambda(\boldsymbol{s}_k(r,\theta)) + \lambda(\boldsymbol{s}_k(r,\theta+\pi)) = 2\lambda(\boldsymbol{s}_k)$. Therefore,

$$\begin{aligned}\int_0^{2\pi} \lambda(\boldsymbol{s}_k(r,\theta)) g_k(r) \mathrm{d}\theta &= \int_0^{\pi} \left[\lambda(\boldsymbol{s}_k(r,\theta)) + \lambda(\boldsymbol{s}_k(r,\theta+\pi))\right] g_k(r) \mathrm{d}\theta \\ &= 2\pi\lambda(\boldsymbol{s}_k) g_k(r).\end{aligned}$$

We note that this is a relaxation of the traditional assumption in distance sampling spatial models that the intensity is constant within each sampling unit, or at least approximated as constant when points are binned into discrete counts for count-type GLM approaches. A similar relaxation is also possible in the case of line transects, see (Yuan et al., 2017). This assumption of linear intensity within the transect complements the use of the piece-wise linear basis functions for the SPDE effect (see Section 3.5).

Integrating with respect to r and accounting for the change of variables gives $\int \lambda(\boldsymbol{s}) g(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} = \sum_{k=1}^K 2\pi\lambda(\boldsymbol{s}_k) \int_0^W r g_k(r) \mathrm{d}r$. For each sampling unit we approximate the one-dimensional integral $\int_0^W r g_k(r) \mathrm{d}r$ using a midpoint integration method with M integration locations $r_{k1}, \ldots, r_{kM}$ and associated weights $\alpha_{k1}, \ldots, \alpha_{kM}$. This gives

an approximation to the integral as

$$\int_{\boldsymbol{s}\in\Omega} \lambda(\boldsymbol{s})g(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \approx \sum_{k=1}^{K}\sum_{j=1}^{M} \tilde{\alpha}_{kj}\tilde{\lambda}(\boldsymbol{s}_{kj}),$$

where $\tilde{\alpha}_{kj} = 2\pi\alpha_{kj}r_{kj}$ and $\tilde{\lambda}(\boldsymbol{s}_{kj}) = \lambda(\boldsymbol{s}_k)g_k(r_{kj})$.

To simplify notation, we let $\tilde{\alpha}_k = (\alpha_{k1}, \ldots, \alpha_{kM})^\intercal$, $\tilde{\alpha} = (\tilde{\alpha}_1^\intercal, \ldots, \tilde{\alpha}_K^\intercal)^\intercal$, and, similarly, let $\tilde{\lambda}_k = (\tilde{\lambda}(s_{k1}), \ldots, \tilde{\lambda}(s_{kM}))^\intercal$, and $\tilde{\lambda}_{int} = (\tilde{\lambda}_1^\intercal, \ldots, \tilde{\lambda}_K^\intercal)^\intercal$, the vector of the intensity field evaluated at the integration locations. We also denote the intensity evaluated at the observation locations as $\tilde{\lambda}_{obs} = (\tilde{\lambda}(s_1), \ldots, \tilde{\lambda}(s_n))^\intercal$. Then the approximate log-likelihood is

$$\log \pi(\boldsymbol{Y}) \approx -\tilde{\alpha}^\intercal\tilde{\lambda}_{int} + 1^\intercal \log \tilde{\lambda}_{obs}. \tag{3.2}$$

This approximation can be expressed as a modified Poisson likelihood. To see this let $\eta = (\tilde{\lambda}_{int}^\intercal, \tilde{\lambda}_{obs}^\intercal)^\intercal$, $\alpha = (\tilde{\alpha}^\intercal, 0_{n\times 1}^\intercal)^\intercal$ and construct a vector of pseudo-observations $z = (0_{KM\times 1}^\intercal, 1_{n\times 1}^\intercal)^\intercal$. Then the approximate likelihood can be written as

$$\pi(\boldsymbol{Y}) \approx C \prod_{i=1}^{KM+n} \eta_i^{z_i}\exp(-\alpha_i\eta_i), \tag{3.3}$$

where C is a constant. This approximate likelihood is implemented in `inlabru` as a `"cp"` likelihood. This approximation is similar to what is often referred to as the 'Berman-Turner device' that is used to approximate a wide variety of point process likelihoods and pseudo-likelihoods (Berman and Turner, 1992; Baddeley and Turner, 2000), although here we do not use their assumption that the locations of observed points form a part of the quadrature scheme for the integration.

3.3.4 Intensities for incomplete data

In the above, we assume the data are complete records of animal locations. However, in many distance sampling surveys only the location of the observer and the distance to the observer are recorded. Data of this type can be analysed within a point process framework by deriving the appropriate intensity function for the incomplete data as we outline below.

Using the polar notation given above, for a detected point at location $\boldsymbol{s}_k(r, \theta) \in \Omega_k$, we consider the case where we observe r but not θ. It follows that the intensity for points at a distance r observed within sampling unit Ω_k is

$$\begin{aligned}\tilde{\lambda}_k(r) &:= \oint_{c_k(r)} \lambda(\boldsymbol{s}) g_k(r) \mathrm{d}\boldsymbol{s} \\ &= 2\pi r \lambda(\boldsymbol{s}_k) g_k(r), \end{aligned} \tag{3.4}$$

where $c_k(r)$ is a circle of radius r centred at $\boldsymbol{s}_k$ and the second line follows from changing to polar coordinates and the assumption of linear intensity within Ω_k. This intensity differs from the full data case through the 2π term, which accounts for the fact that we do not observe θ and the additional r term, which accounts for the increasing area surveyed at larger distances. The log-intensity for observed points is thus $\log \tilde{\lambda}(\boldsymbol{s}) = \log 2\pi r + \log \lambda(\boldsymbol{s}) + \log g(\boldsymbol{s})$ and so can be implemented using $\log 2\pi r$ as an offset term. As we noted above, the half-normal detection function $g(\boldsymbol{s})$ is non-linear in its parameters and this component of the additive predictor requires a novel approach to parameter estimation.

3.4 Iterated INLA

The half-normal detection function depends on the strictly positive parameter σ^2. We choose a log link and consider $g(\boldsymbol{s}|\phi)$, where $\log \sigma^2 = \phi$ and ϕ can take any real value, postive or negative. However, this model cannot be directly fitted using

INLA since $\log g(\boldsymbol{s}|\phi)$ is not linear in ϕ. To address this we use an approximate method based on a first-order Taylor approximation of the non-linear predictor. Let $\eta(\boldsymbol{u}) = \log \tilde{\lambda}(\boldsymbol{u})$ where $\boldsymbol{u}$ is the parameter vector of the predictor, which includes ϕ as well as any parameters for fixed and random effects. Let $\bar{\eta}_{\boldsymbol{u}_0}$ denote the first-order Taylor approximation of $\eta(\boldsymbol{u})$ at the point $\boldsymbol{u}_0$. That is

$$\bar{\eta}_{\boldsymbol{u}_0}(\boldsymbol{u}) = \eta(\boldsymbol{u_0}) + \boldsymbol{B}_{\boldsymbol{u}_0}(\boldsymbol{u} - \boldsymbol{u}_0),$$

where $\boldsymbol{B}_{\boldsymbol{u}_0}$ is the derivative matrix of $\eta(\boldsymbol{u})$ evaluated at $\boldsymbol{u}_0$. Since $\bar{\eta}_{\boldsymbol{u}_0}$ is linear in $\boldsymbol{u}$ it is possible to fit an approximate model in INLA by replacing the non-linear η with the linear Taylor approximation $\bar{\eta}_{\boldsymbol{u_0}}$.

The question then is how to choose the linearisation point $\boldsymbol{u}_0$. The approach taken here is via a fixed point iteration scheme. There are many possible choices here and we use the approach in the current implementation of `inlabru` [1] (Bachl et al., 2019). However, alternative approaches are likely to be added to `inlabru` in the future. We denote by $\bar{p}_{\boldsymbol{v}}$ the posterior distribution of the model parameters given the linearised model configuration at linearisation point $\boldsymbol{v}$ and define the functional

$$f(\bar{p}_{\boldsymbol{v}}) = \left\{ \arg\max_{\boldsymbol{u}_i} \bar{p}_{\boldsymbol{v}}(u_i|\boldsymbol{Y}), \quad i = 1, \ldots, n_{\boldsymbol{u}} \right\}, \tag{3.5}$$

where $n_{\boldsymbol{u}}$ is the length of $\boldsymbol{u}$. Note that each $\bar{p}_{\boldsymbol{v}}(u_i|\boldsymbol{Y})$ is approximated and the mode estimated during the model fitting using INLA, so these terms require no additional computation beyond fitting the model. We then seek a fixed point of this functional so that $\boldsymbol{u}_0 = f(\bar{p}_{\boldsymbol{u}_0})$, i.e. we find a linearisation location such that the mode of each (approximate) marginal posterior distribution directly coincides with this location. This fixed point is identified via an iterative scheme: given a current choice of linearisation point $\boldsymbol{v}$, estimate $\bar{p}_{\boldsymbol{v}}$ and set $\boldsymbol{u}_* := f(\bar{p}_v)$ to use as the new

[1] Since `inlabru` version 2.2.3 the option of using the joint 'empirical Bayes conditional mode' has been added. However, this analysis was completed before this option was available.

linearisation point. Iterate this procedure until a fixed point is identified within a chosen tolerance.

In practice the point $\boldsymbol{u}_*$ may be problematic and result in a linearised model configuration that is challenging to fit. Instead, this point is considered only as an initial candidate point. The next linearisation point is chosen by setting $\boldsymbol{u}_\alpha = (1-\alpha)\boldsymbol{v} + \alpha\boldsymbol{u}_*$ and finding the value of α that minimises $\|\eta(\boldsymbol{u}_\alpha) - \bar{\eta}_{\boldsymbol{v}}\|$. This means that the new linearisation point remains close to a feasible region for fitting, assuming the current point $\boldsymbol{v}$ is reasonable. We estimate α using an approximate line search method that avoids many potentially expensive evaluations of the non-linear predictor. This iterative method is implemented in the R package `inlabru` (Bachl et al., 2019).

3.5 Spatially Structured Random Effect

We use the computationally efficient SPDE effect (Lindgren et al., 2011) to describe spatial variation in 'ākepa density. The log-intensity for the point pattern of all animal locations (both detected and undetected) is given by

$$\log \lambda(\boldsymbol{s}) = \beta_0 + \xi(\boldsymbol{s}),$$

where β_0 is an intercept parameter and $\xi(\boldsymbol{s})$ is a zero-mean GRF with Matérn covariance. We use the GMRF approximation of $\xi(\boldsymbol{s})$ with piece-wise linear basis functions $\phi_1, \ldots, \phi_L$ defined on a finite element mesh with L nodes. Then $\xi(\boldsymbol{s}) = \sum_l \xi_l \phi_l(\boldsymbol{s})$ and the parameters $\xi_1, \ldots, \xi_L$ form a GMRF with sparse prior precision matrix $\boldsymbol{Q}_\xi = \frac{1}{\tau^2}(\kappa^4 \boldsymbol{C} + 2\kappa^2 \boldsymbol{G_1} + \boldsymbol{G_2})$ where $\boldsymbol{C}$, $\boldsymbol{G_1}$, $\boldsymbol{G_2}$ are all $L \times L$ sparse matrices (see Chapter 2, Section 2.5).

The parameters τ and κ control the shape and rate of decay of the Matérn covariance function. In order to specify priors on the Matérn covariance, we use a reparameterisation to range and marginal variance parameters. When $\nu = 1$ and

the domain is two-dimensional, the reparameterisation is $\rho = \sqrt{8}/\kappa$, for the range parameter, and $\sigma^2 = 1/(4\pi\kappa^2\tau^2)$, for the marginal variance parameter (Blangiardo and Cameletti, 2015).

This parameterisation is useful for setting penalised complexity priors (Simpson et al., 2017), an approach to prior specification that allows flexible model components to shrink towards a simple base model. For the Matérn field, the base model is the limiting case with $\sigma^2 = 0$ and $\rho = \infty$, since this would result in a random field that is almost surely zero everywhere, effectively removing the random field from the model. The purpose of penalised complexity priors is to place sufficient prior probability on a simple model to allow the random effect to remain essentially unused unless the observed data suggests it is required.

To specify the penalised complexity priors on the SPDE effect we set $\mathbb{P}(\sigma > 2) = 0.01$ and $\mathbb{P}(\rho < 130) = 0.01$. The value of 130 for the range parameter was selected based on the minimum distance between sampling locations. Since only the distance to the observer was recorded, and not the exact location of 'ākepa, we used the adjusted intensity given in equation (3.4).

3.6 Results

The predicted mean of the posterior intensity field is shown in Figure 3.2A along with a map of the coefficient of variation (CV) (Figure 3.2B) and standard deviation of the posterior intensity field (Figure 3.2C) on a regular prediction grid with each prediction cell area approximately 1.7 hectares. This shows a region of high intensity in the south and much lower intensity in the north, agreeing with a standard two-stage analysis of the same data (Camp et al., 2020). The CV plot shows that the CV in the posterior intensity field is lower in areas with greater sampling effort. There is a clear indication of preferential sampling with more survey effort in the south, where densities are higher, compared to the north. In the north, estimated

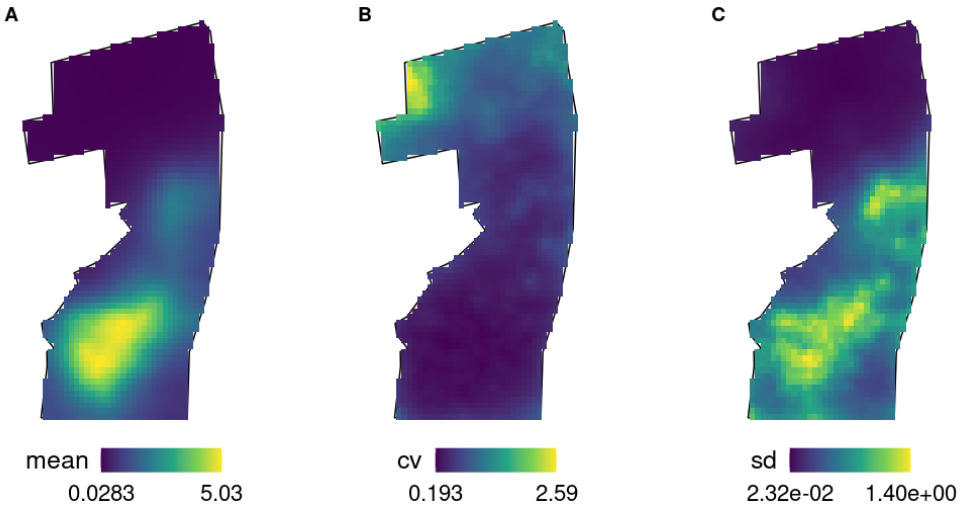


Figure 3.2: The A) mean, B) CV and C) standard deviation of the posterior intensity field. Units for the intensity field are per hectare.

intensities are lower, which also contributes to larger CV values and makes the CV map hard to interpret.

The standard deviation map shows stronger overall posterior variability in the south where intensity is larger. This is likely due to the assumed log-Gaussian relationship. This is due to the log-Gaussian relationship between the random effect and the density. If X is normally distributed with mean μ and variance σ^2 then $Z = \exp(X)$ is log-Gaussian with a mean-variance relationship given by $\text{Var}(Z) = \exp(\sigma^2 - 1)\mathbb{E}(Z)^2$. This implies that the variance will generally be higher when the expected density is higher and maps of the standard deviation will tent to have a similar pattern to the mean intensity.

We also map the lower and upper quantiles corresponding to a 95% credible interval for each prediction location (Figure 3.3). As we discuss below, these quantiles are independent summaries of the posterior intensity field at each prediction location. The chance of a realisation of the posterior field *simultaneously* taking the 0.095 quantile at each prediction location is vanishingly small, which makes these maps that show thousands of such quantiles side by side in a single image liable to be misinterpreted. We discuss this in more detail below.

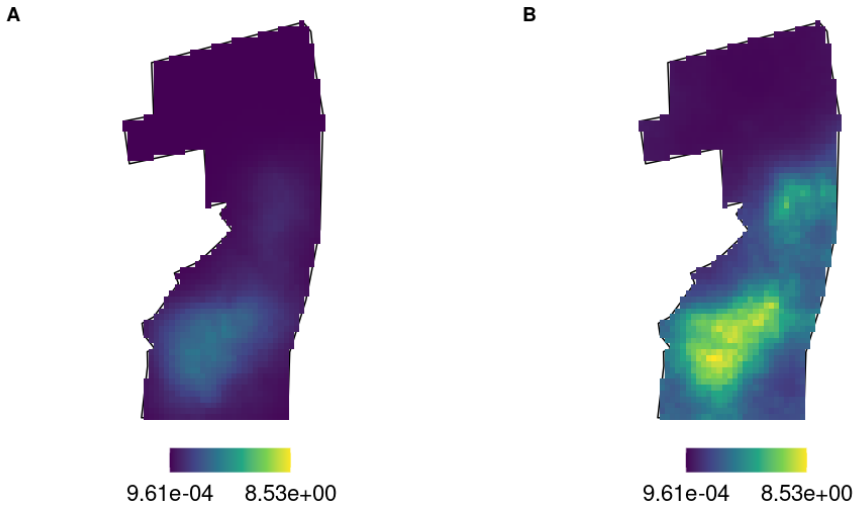


Figure 3.3: The predicted posterior intensity quantiles: A) 0.025 quantile B) 0.975 quantile. Both plots plots use the same colour scale. Units for the intensity field are per hectare.

The posterior detection function (Figure 3.4) shows that detectability drops to just under 0.25 at the maximum observable distance of 58 metres and also broadly agrees with a two-stage analysis. The posterior Matérn correlation function is shown

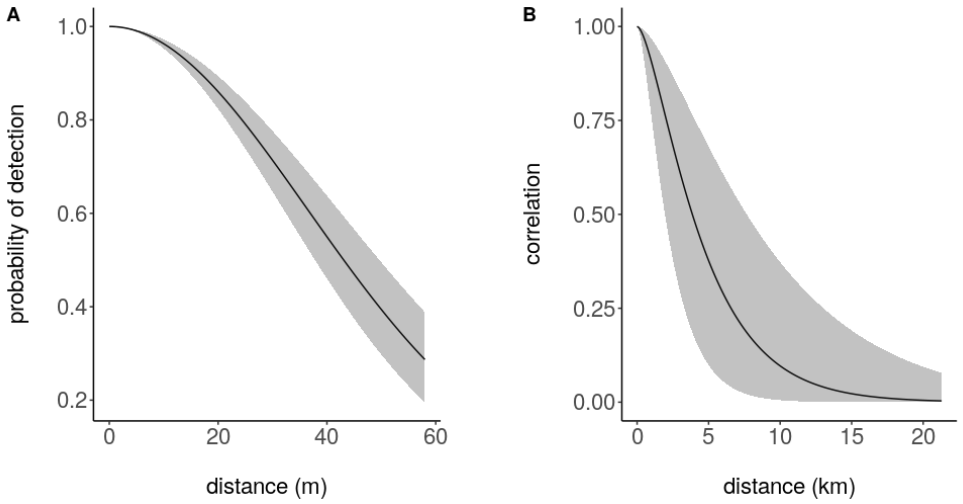


Figure 3.4: A) Posterior half-normal detection function. B) Posterior Matérn correlation function. The black line is the mean posterior estimate and the grey shaded area is 95% credible interval in both plots

in Figure 3.4; the grey regions show the 95% credible interval. Note that this correlation function shows positive correlation even at distances of 10,000 metres,

wider than the full length of the study region (approximately 8,000 metres). This is the correlation range on the log-intensity scale. On the intensity scale the empirical range of correlation is shorter than this.

To further evaluate the ability of the SPDE effect to capture the spatial heterogeneity in the observed data we generated 100 point patterns using the posterior intensity field and thinned them with the posterior detection function to create 1,000 simulated datasets to compare with the observed data. We calculated the pairwise distances between all observations for each of these datasets and compared the frequency of pairwise distances with those in the observed data (Figure 3.5). The plot

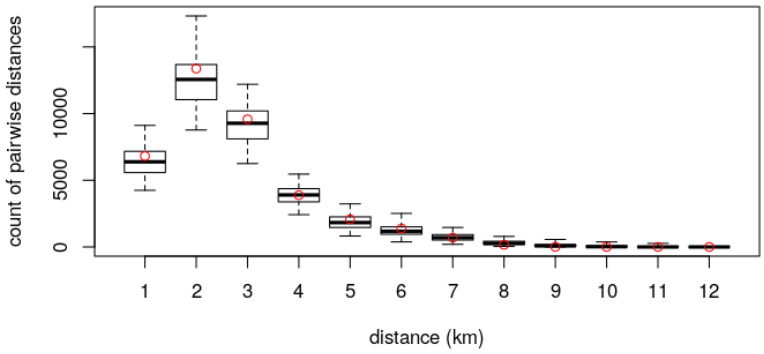


Figure 3.5: Boxplots of frequency of pairwise distances (tick mark values for end of each bin) based on 100 point patterns sampled from the posterior. Red circles: observed frequency of pairwise distances between observations

shows that the SPDE effect broadly captures the spatial clustering in the observed data although there is some evidence of under-predicting the strength of the clustering at distances between 1,000 and 2,500 m, with some evidence of negative bias here. This could perhaps be due to large areas of the study region having low density, which has a larger correlation range than higher density areas. Since the SPDE effect has only one range parameter it may struggle to predict in areas with shorter range correlations than the average correlation range across the whole region.

We note that Figure 3.5 is similar in spirit to Ripley's K-function (Ripley, 1976)

that is a common approach to investigating spatial structure in point patterns. Here we compare pairwise distances for simulated datasets with those in the observed data as a method to evaluate the model. Other methods of comparison could also be considered such as the average difference between predicted and observed counts at each point transect, for example. Working with posterior predictive datasets is a flexible approach to model evaluation that allows for innovative methods of comparison with observed data that can focus on particular features of the model.

We plot the posterior abundance estimate for the 'ākepa in Hakalau reserve using a Monte Carlo method that is based on sampling realisations of the posterior intensity field. Let n denote the abundance of the population within the region of interest

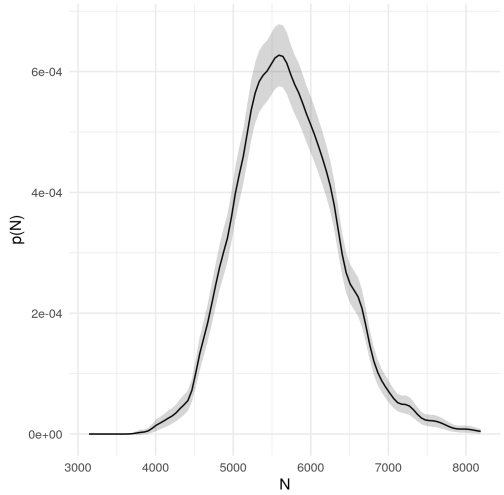


Figure 3.6: Posterior abundance. Grey area marks two Monte-carlo standard errors above and below the mean

and $N = \int \lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$ its corresponding random variable within our Bayesian framework. Integrating the mean of the posterior intensity field provides a point estimate for the abundance, i.e. the expected abundance. To generate a full posterior estimate for abundance we approximate the posterior $\pi(N|\boldsymbol{Y})$ by a Monte Carlo method. Taking m Monte Carlo samples of the posterior intensity field, $\lambda^{(1)}, \ldots, \lambda^{(m)}$, we estimate the posterior for the abundance as $\pi(N|\boldsymbol{Y}) \approx 1/m \sum_{i=1}^{m} \pi(N|\lambda = \lambda^{(i)})$. Each $\pi(N|\lambda = \lambda^{(i)})$ is a Poisson probability mass function with rate parameter

$\int_\Omega \lambda^{(i)}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$.

Figure 3.6 shows the approximate posterior for N with $m = 20,000$ with the grey shaded region the Monte Carlo error of the estimated posterior. This allows us to estimate the probability for any specific value of realised abundance n by estimated $\pi(N = n|\lambda)$. This approach can also be used to generate credible intervals and other posterior summaries of interest.

The abundance estimates cannot be directly compared to the results in Camp et al. (2020) since their two-stage model also includes a temporal correlation and we include an area of Hakalau that they had to exclude. Converting the above abundance estimates to density per hectare for the whole study region allows for some comparison with Figure 2 in Camp et al. (2020) which suggests that 'ākepa density was roughly between 0.5 and 1.5 birds per hectare in 2002 (considering both the design-based estimate and the spatio-temporal GAM estimate). In our analysis we find a credible region of roughly 0.9 to 1.7 birds per hectare, which is slightly different but overlaps significantly with the results in Camp et al. (2020).

The attractive feature of our Bayesian one-stage model is that, by using posterior Monte Carlo sampling, model summaries that are not based on sampling the detection model are just samples from the marginal posterior distribution that integrates over the detection function parameters. This means that all the outputs presented in the above, the summaries of the random field as well as the posterior abundance, include the uncertainty in the detection model in a natural way. There is no additional step required to incorporate detection function uncertainty beyond working with samples from the joint posterior.

3.7 An alternative approach to predicting abundance

In the above, the posterior for abundance is estimated by taking posterior realisations of the intensity at every prediction location within the study area. This predicts the intensity for prediction locations within point transects in the same manner as for those outside. However, we could use the fact that we have actually observed counts within point transects and so do not need to predict this. This section investigates this alternative approach for predicting overall abundance using the fitted 'ākepa model.

Let A denote the entire study region and the random variable of number of points within a bounded subset $B \subset A$ as $N(B)$. We have $N(B) \sim \text{Poisson}(\Lambda(B))$, where $\Lambda(B) := \int_{\boldsymbol{s} \in B} \lambda(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$, and we are interested in approximating the posterior $\pi(N(A)|\boldsymbol{Y})$. Recall also that the entire surveyed region is denoted Ω, the union of all point transects.

The approach we take is also based on working with samples from the posterior intensity field. Given sampled intensities $\lambda^{(1)}, \ldots, \lambda^{(M)}$, defined over the entire study region A, where each $\lambda^{(m)}$ is sampled from the posterior distribution of $\lambda|\boldsymbol{Y}$, we then approximate the posterior abundance via a Monte Carlo method similar to the above that used $\pi(N(A)|\boldsymbol{Y}) \approx 1/M \sum_{m=1}^{M} N(A|\lambda = \lambda^{(m)}, \boldsymbol{Y})$. However, an alternative approach is to apply the above method to only the *non-surveyed* region $A \setminus \Omega$ and to predict only the *undetected* counts within Ω, the region covered by the point transects.

The distribution of the number of undetected birds across all transects is Poisson distributed with rate parameter $\tilde{\Lambda}(\Omega) := \int_{\boldsymbol{s} \in \Omega} (1 - g(\boldsymbol{s})) \lambda(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$. Therefore, the distribution of the number of all *undetected* birds in the study region, whether through not being in a surveyed area (and so, in principle, undetectable) or within a point transect (in principle detectable but remaining undetected), is Poisson with

rate parameter $\Lambda(A \setminus \Omega) + \tilde{\Lambda}(\Omega)$. Then the distribution of the total abundance can be obtained by adding the observed abundance to this probability distribution.

In theory, since this now uses the observed counts instead of predicting them, this should have lower variance compared to the 'predict everywhere' approach. However, the point transects cover only 6.6% of the total study area, so the difference between the two approaches may be small.

To investigate whether this alternative approach makes a substantial difference to the results, we compared the approximate posteriors using both methods, defined using as much as possible the same numerical integration scheme to ensure that any differences were not due to different schemes. Recall each point transect is denoted Ω_k with centroid $\boldsymbol{s}_k$ and associated detection function $g_k(\boldsymbol{s})$. We can approximate, $\tilde{\Lambda}(\Omega)$, the rate parameter for birds undetected within the surveyed region Ω, as

$$\begin{aligned}
\tilde{\Lambda}(\Omega) &= \int_{\boldsymbol{s}\in\Omega} (1 - g(\boldsymbol{s}))\lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\
&= \sum_k \int_{\boldsymbol{s}\in\Omega_k} (1 - g_k(\boldsymbol{s}))\lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\
&\approx \sum_k 2\pi\lambda(\boldsymbol{s}_k) \int_0^W r(1 - g(r))\mathrm{d}r \\
&\approx \sum_k \sum_j 2\pi r_{kj}\alpha_{kj}\lambda(\boldsymbol{s}_k)(1 - g(r_{kj})) \qquad (3.6)
\end{aligned}$$

where locations r_{kj} and weights α_{kj} are constructed following the mid-point integration scheme described in Section 3.3.3.

The integration scheme for $\Lambda(A \setminus \Omega)$, the rate parameter for the abundance of birds in the unsurveyed region, is defined via a projection of integration locations to mesh nodes, taking into account the fact that some mesh nodes may be within or near to point transects and so should have a lower weight than in comparison to the full mesh integration scheme. Figure 3.7A shows the raw integration locations and associated weights for the integration scheme for a subset of the full study area A, alongside the integration locations for $A \setminus \Omega$ (Figure 3.7B).

These raw weights are then projected to mesh nodes using barycentric coordinates and the fact that the intensity is linear in each triangle of the mesh (see Bakka (2019) for a nice explanation of how this works). These projected integration weights are shown in Figure 3.7. Comparing the integration weights for the entire

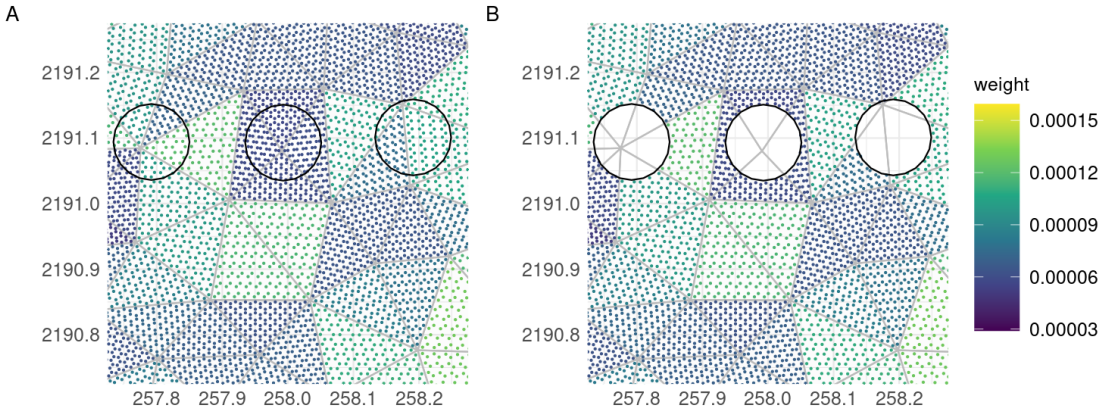


Figure 3.7: Integration locations and associated weights for integration over the whole study region (**A**) and for the whole study region minus the point transects (**B**). The black circles are point transects.

region versus the scheme with point transects removed, we can see lower weights at mesh nodes within or near point transects for this second integration scheme (Figure 3.8). This accounts for the fact that the intensity is not being integrated within the transects.

We use this scheme to approximate the integral of each sampled intensity over $A \setminus \Omega$ and combine it with the estimated rate parameters $\tilde{\Lambda}(\Omega)$ generated using the mid-point integration presented above. The posteriors estimated by this approach is shown in Figure 3.9 alongside the posterior using the 'predict everywhere' approach. This shows there is not a great deal of difference between the two approaches and the additional complication of using the observed counts may not be worth the trouble. The small differences may be because the observed counts are so low compared to predicted overall abundance (153 observed birds compared to a credible range of

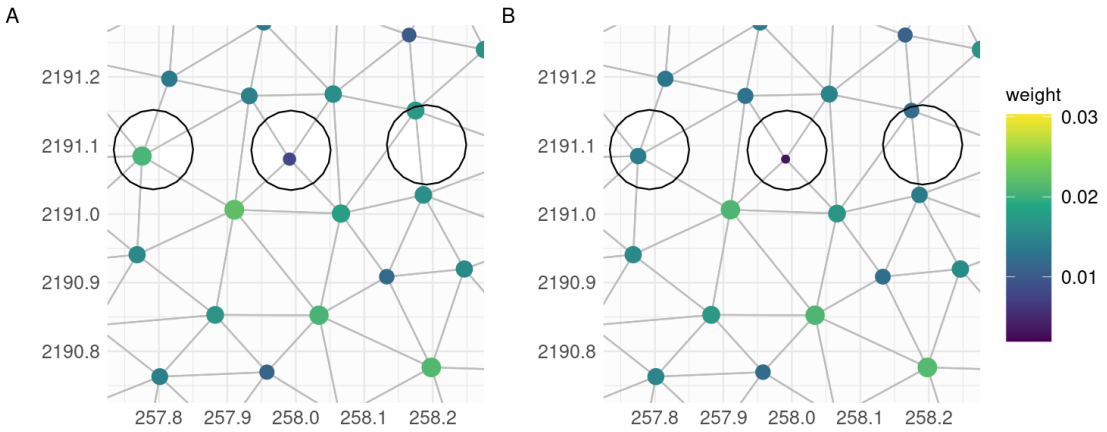


Figure 3.8: Integration weights projected to mesh nodes for integration over the whole study region (**A**) and integration over the whole study region minus the transects (**B**). The black circles are point transects. Note that the weights in **B** are smaller for mesh nodes within or near the point transects when compared to **A**, reflecting the fact that the intensity is not being integrated within transects in **B**.

4,000-8,000 or so for the total population in the study region) and that detectability was fairly high within point transects, with most birds detected. Coupled with the fact that the sampled region is only 6.6% of the total study region, it is perhaps not surprising that the resulting posterior estimates are almost identical.

In this specific application, we conclude that there is no much benefit to go through with the extra complication of using the observed counts over predicting everywhere. However, for other populations this difference may well be large enough to be worth accounting for, especially if a larger proportion of the region has been surveyed or observed abundance accounts for a larger proportion of the total population than was the case here. `inlabru` provides useful tools for this type of high-resolution quadrature scheme that was necessary here since transects were small compared to the resolution of the mesh and the size of the study area as a whole.

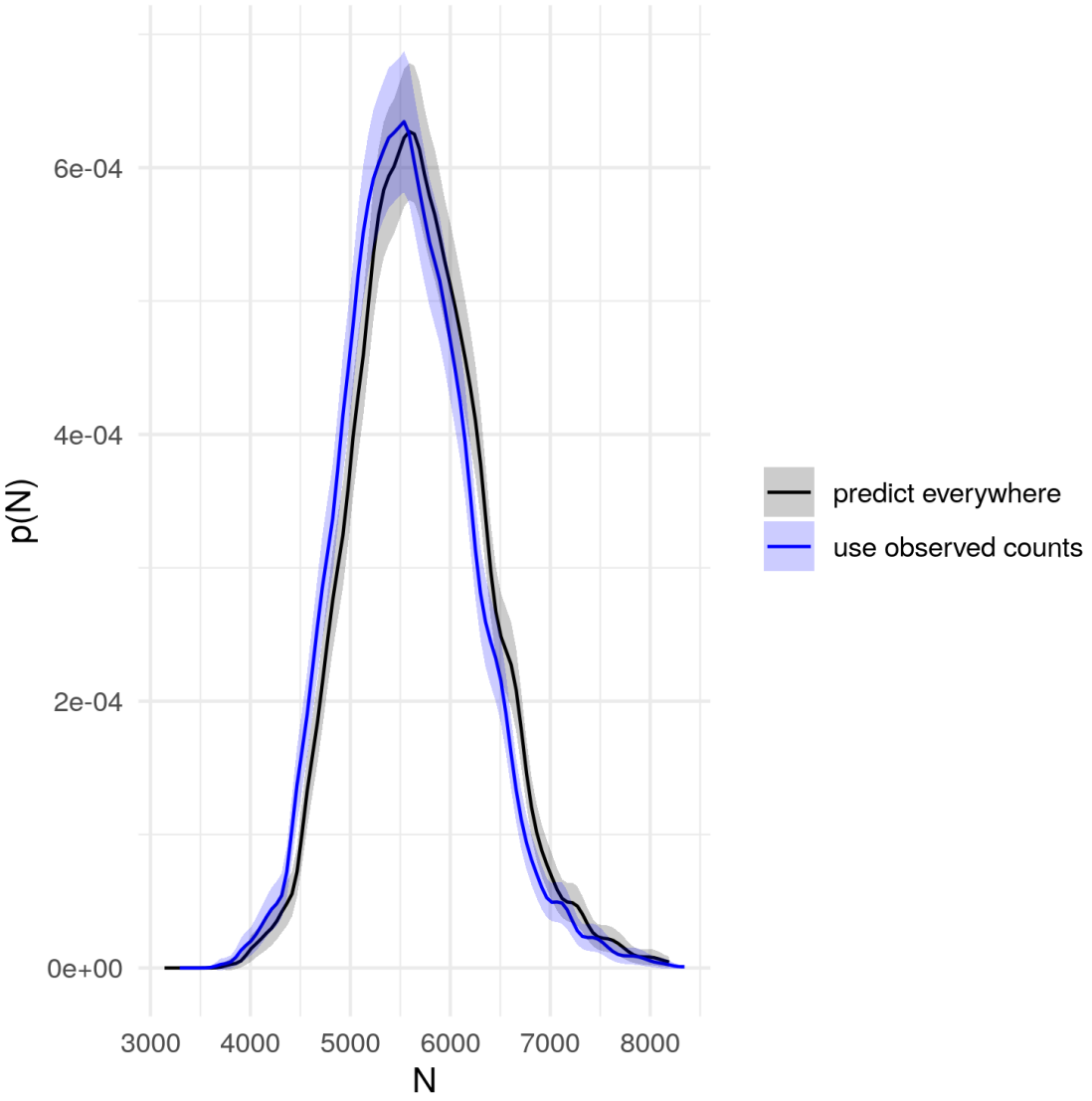


Figure 3.9: Comparing approximate posteriors for abundance using the two different approaches. Shaded regions are 2 Monte Carlo standard errors above and below the mean

3.8 Communication of Results

Our presentation of the results is broadly consistent with approaches taken in the species distribution modelling literature, although maps of predictive uncertainty (Figure 3.2 and Figure 3.3) are not always provided and it is common to see maps of point estimates without any accompanying communication of uncertainty. It is also common to report a point estimate of the expected abundance along with uncertainty in this point estimate. However, uncertainty in a point estimate will

tend to have lower variance than the variance of the random variable of interest. For this reason we chose to show the full abundance posterior in Figure 3.6.

Our intention in this section is to highlight limitations with mapped summaries of model predictions, including those presented in Section 3.6, and suggest ways to address these limitations. We note that all model outputs are based on sampling from the joint posterior of all model parameters and therefore naturally average over the uncertainty in the observation process, a key advantage of the one-stage approach.

3.8.1 Limitations of mapped summaries

The most common method to communicate uncertainty in a maps of animal density is to produce maps of some measure of predictive uncertainty, such as the standard deviation or CV (Fuller et al., 2018; Vallejo et al., 2017; Bradbury et al., 2014) that can be derived from the posterior predictive distribution for the model or bootstrapping in a maximum likelihood context. Another approach is to map quantiles or the probability of exceeding certain thresholds for each prediction location (Russell et al., 2016; Wilson et al., 2010). In Section 3.6, we present maps of the CV, standard deviation, and the 0.025 and 0.975 quantiles of the posterior intensity field that are all intended to communicate uncertainty in the spatial distribution of 'ākepa (Figures 3.2B, 3.2C, 3.3A, and 3.3B, respectively). Whilst maps that summarise spatial predictions are useful, they all mask certain properties that may be important when communicating the results of the analysis.

A map showing the posterior predicted mean across the study region (Figure 3.2A) is often the key output of a species distribution model. However, even this relatively innocuous summary statistic can mask important features of the random field since the posterior mean will always be smoother than actual realisations. Figure 3.10 shows three such realisations from the posterior intensity field. Note that each realisation has a finer-grained spatial structure than is shown in the posterior

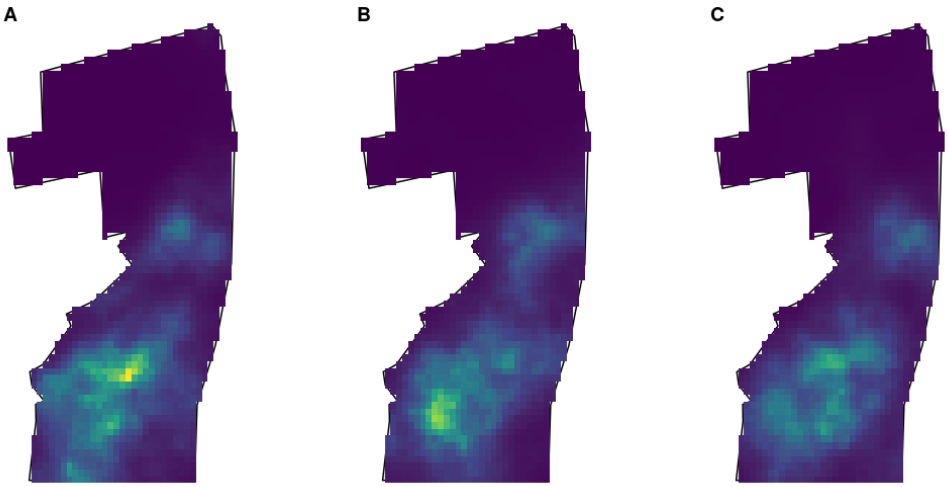


Figure 3.10: Three realizations of the posterior intensity field

mean (Figure 3.2A). Hence, considering this finer-grained structure, our interpretation is that the clustering of animals is stronger than might be expected if we only looked at the map of the posterior mean. This is one example of how a summary statistic can mask important features of the posterior random field.

For model evaluation we recommend plotting multiple realisations of the posterior field as these will more closely resemble the spatial structure of the observed data and thus what we might encounter in the field. The mean is only a summary of the posterior intensity field and can give a misleading sense of homogeneity, particularly in the high density area in the south, when compared to realisations of the mean. Presenting multiple realisations, perhaps as an animation, is also an effective way to communicate uncertainty. Bowman (2019) provide an interpolation method that preserves the mean and covariance structure of realisations that smoothly interpolates between realisations. This avoids abrupt changes that can occur when animating the raw realisations.

The CV map (Figure 3.2B) is also intended to communicate uncertainty. However, CV values will be higher in regions of low predicted intensity, particularly if the posterior standard deviation is relatively consistent across the study region or positively correlated with the intensity. The observed higher CV areas in such maps

will highlight regions of relatively lower intensity but not necessarily higher overall uncertainty. This is clearly the case in Figure 3.2B. However, there is also variable sampling effort within the study region and it is not clear what impact this has on the CV map. This means that the resulting map is hard to interpret.

The standard deviation (SD) map (Figure 3.2C) shows spatially varying SD values. However, care should be taken as to what the mapped values imply in the context of a given analysis. A default colour scale will nearly always show some spatial variation in uncertainty, with some regions of relatively high and low uncertainty. What such a map does not show is whether these differences actually matter in the context of the analysis. For example, it could be the case that the standard deviation values are so large as to make the predictions have an unacceptably high uncertainty everywhere. Or, on the other hand, perhaps differences between high and low uncertainty regions (where 'high' and 'low' are defined by the default colour scale for the map) are small enough to be negligible when it comes to making conclusions or decisions, in which case such a spatially varying map may not be a useful output to communicate to decision makers. These are two extreme examples. However, even in less extreme situations, the actual practical relevance of the SD values should be considered in order to be able to interpret the map.

The SD map suffers from another weakness which is that the SD values are linked to the predicted mean since the model assumes a log-Gaussian relationship. This means that the SD map usually looks similar to the mean intensity map. This relationship is the result of a modelling assumption and not necessarily reflective of the detectability of animals or survey effort. One might expect uncertainty to be lower in areas with greater survey effort, however for the 'ākepa data the opposite is the case. This is another limitation of the SD map.

The quantile maps (Figure 3.3) are also potentially difficult to interpret. The temptation is to perceive the maps as showing possible intensity surfaces that could have produced the observed data with one possible intensity surface (0.025 quantile

map) that signifies lower abundance and another surface (0.975 quantile map) that signifies higher abundance. However, presenting these quantiles together in a single map obscures the fact that it is vanishingly unlikely for all prediction locations to *simultaneously* achieve their 0.025 or 0.975 quantiles. A similar problem will occur for maps of the marginal probability of exceeding a threshold at each prediction location.

Presenting independent quantiles jointly in a single image is risky. We believe these caveats make the map difficult to use for most non-statistically trained audiences and even trained statisticians may misinterpret these quantile maps if they are not careful. To demonstrate the possible consequences of this we (incorrectly) treat the lower and upper quantile plots as though they are intensity functions and integrate them to obtain an expected abundance estimate of approximately 2,800 for the 0.025 quantile map and 10,200 for the 0.975 quantile map. A naive (and tempting) interpretation of these numbers is as lower and upper limits of the 95% credible interval for abundance. However, these abundance estimates are outside the support of the posterior for abundance (see Figure 3.6). This demonstrates that the tendency to interpret these maps as showing possible intensity surfaces that are consistent with the observed data can lead to interpretations that are inconsistent with the very model that generated the maps.

We conclude that, although all useful in their own way, each of the mapped summaries of th posterior intensity field suffer from some weaknesses, whether through masking certain properties of the random field or through being difficult to interpret. We next present some alternatives that avoid some, but not all, of these problems.

3.8.2 Excursion sets and excursion functions

One alternative to the quantile and standard deviation maps is to suggest that consideration should be given, *a priori*, to relevant values of animal density and acceptable levels of uncertainty given the context and aims of the analysis, and that

the predicted values of the posterior intensity field should be considered *jointly*, to avoid the problem of presenting independent quantiles side by side. Given pre-specified values of interest for the intensity, and an acceptable uncertainty level, these can then be used to construct summary maps that consider the joint probability of events across all prediction locations.

We demonstrate this perspective using excursions sets and excursion functions (Bolin and Lindgren, 2015), since excursions methods are a natural choice for Gaussian random fields. Excursion sets and excursion functions are based on the joint probability of events across a set of locations. For this reason they avoid the interpretability issues of the quantile or exceedance threshold maps. These methods also require the user to specify thresholds of interest for the random field and acceptable levels of uncertainty. They therefore also avoid the issue of a default colour scale potentially affecting the interpretation of results.

We give the technical definition of excursion sets and functions below. We use the same notation as Bolin and Lindgren (2015), which, for readers unfamiliar with these methods, can take some getting used to. However, the precise mathematical definitions are very useful when it comes to interpreting these methods in the context of generating maps that incorporate uncertainty in species distributions. We try to, as much as possible, give an informal explanation alongside the mathematical description.

The positive excursion set with level u for a function $f(\boldsymbol{s})$ with domain Ω is $A_u^+(f) = \{\boldsymbol{s} \in \Omega; f(\boldsymbol{s}) > u\}$, i.e. the set of all locations in Ω where f exceeds a threshold value u. For a random field, $\lambda(\boldsymbol{s})$, the positive excursion set with level u and probability $1 - \alpha$ is

$$E_{u,\alpha}^+(\lambda) = \arg\max_D \{|D| : \mathbb{P}\left[D \subset A_u^+(\lambda)\right] \geq 1 - \alpha\}.$$

Note that $A_u^+(f)$ specifies a set of locations for which a function $f(\boldsymbol{s})$ exceeds a

threshold value u for *every location* in the set. Therefore, the positive excursion set $E^+_{u,\alpha}(\lambda)$ is the *largest* such set of locations for which realisations of $\lambda(\boldsymbol{s})$ exceed the threshold u, *simultaneously* for all locations in the set, with a chosen probability level $1-\alpha$ that is set by the user. Negative excursion sets are similarly defined by considering the probability of being below a threshold value.

Excursion sets can be estimated by considering candidate sets for D of increasing size and a sequential integration scheme to estimate the required probabilities. A full description of the method can be found in Bolin and Lindgren (2015) and a software implementation is available in the `excursions` package (Bolin and Lindgren, 2018).

Figure 3.11A shows the positive excursion set with a level corresponding to 1 bird per hectare with probability 0.95 (hence $\alpha = 0.05$). This figure can be interpreted in a natural way as the largest region for which the intensity is greater than 1 bird per hectare for every location within the region, with probability 0.95.

This map could be used to define, for example, a 'core region' for the 'ākepa population, with particular conservation importance. Crucially, this depends on a clear mathematical description of what a 'core region' actually means in a statistical sense. Of course, any particular definition would be up for debate. Here we just intend to present an example of how a clear definition could potentially be used to produce a map that incorporates relevant definitions and uncertainties, and can be interpreted clearly.

To visualise how such maps change with different levels of uncertainty we can use the excursion function $F^+_u(\boldsymbol{s}) = \sup\{1-\alpha; \boldsymbol{s} \in E^+_{u,\alpha}\}$, which defines, for each location, the largest possible probability $1-\alpha$ for which that location would be in the excursion set defined using probability $1-\alpha$. i.e. if we allow greater uncertainty, this function shows which locations would be included in the excursion set.

The excursion function with a level corresponding to 1 bird per hectare is shown in Figure 3.11B. This figure can also be interpreted naturally. It shows the largest possible probability for each location to be a member of a set in which the intensity

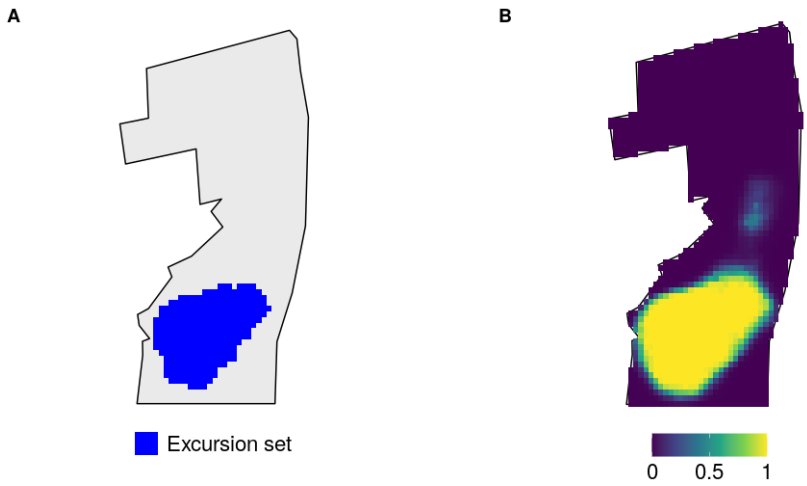


Figure 3.11: Left: The positive excursion set with a level corresponding to 1 bird per hectare and probability 0.95. Right: The positive excursion function with a level corresponding to 1 bird per hectare

exceeds a threshold simultaneously across all locations in the set. It is clear from the figure that regions on the edge of the excursion set would be included if the α value were allowed to increase slightly. There is a region to the north of the main population that would be included in the excursion set for 1 bird per hectare if the probability was allowed to be lower, around 0.5. However, for regions in the north there is essentially no probability level for which those locations would be included in $A_u^+(\lambda)$. This could be used to identify, for example, areas of potential importance that could be investigated further.

We chose the threshold of 1 bird per hectare and an error probability of 0.05 to demonstrate the approach. However, multiple values for each of these can be considered as a simple extension of this approach and we imagine that, in most contexts, there would be value in considering many different thresholds and uncertainty levels.

3.8.3 Estimating the information gained by sampling

Given we have a posterior random field describing the density of birds, it is natural to ask questions along the lines of "in which areas of the study region have we

learned more/less about the distribution of birds?". Given that there is a correlation between observations that decays with distance, we clearly learn more than just the density within each point transect. This correlation is what we use to extrapolate to un-sampled regions.

In this section, we consider an approach to answering these types of questions by defining various metrics that seek to quantify this idea that there is spatial variation in information gained due to spatial variation in sampling effort as well as the spatial correlation of the random field. However, we will see that the resulting metrics are largely driven by the log-Gaussian assumption of the model and, although inspired by similar approaches in the geostatistical literature, they do not easily carry over into the Poisson process setting.

The basic intuition is to compare the posterior random field informed by the data $\lambda|\boldsymbol{Y}, \theta$ and the random field conditioned only on the posterior covariance structure $\lambda|\theta$, where θ denotes the hyper-parameters of the SPDE as well as the intercept variable. The intercept should be included since the GRF and the intercept are not independent. We consider two quantities:

$$V_0(\boldsymbol{s}) = \mathbb{E}_{\theta|\boldsymbol{Y}}\left[\text{Var}(\lambda(\boldsymbol{s})|\theta)\right] + \text{Var}_{\theta|\boldsymbol{Y}}\left[\mathbb{E}(\lambda(\boldsymbol{s})|\theta)\right] \tag{3.7}$$

$$V_{\boldsymbol{Y}}(\boldsymbol{s}) = \mathbb{E}_{\theta|\boldsymbol{Y}}\left[\text{Var}(\lambda(\boldsymbol{s})|\theta, \boldsymbol{Y})\right] + \text{Var}_{\theta|\boldsymbol{Y}}\left[\mathbb{E}(\lambda(\boldsymbol{s})|\theta, \boldsymbol{Y})\right]. \tag{3.8}$$

Note that by the law of total variance $V_{\boldsymbol{Y}}(\boldsymbol{s})$ is just $\text{Var}(\lambda(\boldsymbol{s})|\boldsymbol{Y})$, the usual posterior variance given observed data $\boldsymbol{Y}$. In contrast, $V_0(\boldsymbol{s})$ is an adjusted version of this where the random field is considered only conditional on $\theta|\boldsymbol{Y}$ but not on the data directly. This is a form of null model describing a random field with the same posterior covariance structure but no observations directly informing λ. Note that $V_0(\boldsymbol{s}) \geq V_{\boldsymbol{Y}}(\boldsymbol{s})$ and so by constructing metrics of the form

$$I_{\text{Var}}(\boldsymbol{s}) = 1 - \frac{V_{\boldsymbol{Y}}(\boldsymbol{s})}{V_0(\boldsymbol{s})},$$

this gives a metric between 0 and 1 that describes how much the observations $\boldsymbol{Y}$ have informed the posterior random field at location $\boldsymbol{s}$. The subscript here indicates this metric is using the variance measure of uncertainty. We also consider two alternatives:

$$I_{\text{sd}}(\boldsymbol{s}) = 1 - \frac{\sqrt{V_{\boldsymbol{Y}}(\boldsymbol{s})}}{\sqrt{V_0(\boldsymbol{s})}}$$

$$I_{\text{cv}}(\boldsymbol{s}) = 1 - \frac{\sqrt{V_{\boldsymbol{Y}}(\boldsymbol{s})}/\mu_{\boldsymbol{Y}}(\boldsymbol{s})}{\sqrt{V_0(\boldsymbol{s})}/\mu_0},$$

where $\mu_{\boldsymbol{Y}}$ is the mean of $\lambda(\boldsymbol{s})|\boldsymbol{Y}, \theta$ and μ_0 is the mean of $\lambda|\theta$. These are the standard deviation and the coefficient of variation versions of the metric given above based on the variance. The idea behind considering all three is that the variance and standard deviation have a strong relationship with the mean due to the Poisson assumption.

The posterior variance, standard deviation and coefficient of variation are all estimable by posterior sampling in `inlabru` and so are straightforward to calculate. To estimate $V_0(\boldsymbol{s})$ we use a Monte Carlo sampling scheme based on the moments of the log-Gaussian distribution. This variance estimate can then be transformed as required for the I_{sd} and I_{cv} metrics. Let $\log \lambda|\theta \sim N(m, \sigma^2)$ where m and σ^2 are derivable from θ. Therefore, using the moments of the log-Gaussian distribution, we have

$$\begin{aligned} V_0(\boldsymbol{s}) &= \mathbb{E}_{\theta|\boldsymbol{Y}}\left[\text{Var}(\lambda(\boldsymbol{s})|\theta)\right] + \text{Var}_{\theta|\boldsymbol{Y}}\left[\mathbb{E}(\lambda(\boldsymbol{s})|\theta)\right] \\ &= \mathbb{E}_{\theta|\boldsymbol{Y}}\left[\exp(2m + \sigma^2)(\exp(\sigma^2) - 1)\right] + \text{Var}_{\theta|\boldsymbol{Y}}\left[\exp(m + \sigma^2/2)\right]. \end{aligned}$$

Given Monte-carlo samples $\theta^{(k)} \sim \theta|\boldsymbol{Y}$ for $k = 1, \ldots, K$ this can be estimated as

$$\begin{aligned} V_0(\boldsymbol{s}) &\approx \frac{1}{K}\sum_k \exp(2m_k + \sigma_k^2)(\exp(\sigma_k^2) - 1) + \\ &\quad \frac{1}{K}\sum_k \exp(2m_k + \sigma_k^2) - \frac{1}{K^2}\left[\sum_k \exp(m_k + \sigma_k^2/2)\right]^2 \\ &= \frac{1}{K}\sum_k \exp(2m_k + 2\sigma_k^2) - \frac{1}{K^2}\left[\sum_k \exp(m_k + \sigma_k^2/2)\right]^2. \end{aligned}$$

We use this approach to calculate the required quantities to compute the three metrics, $I_{\text{Var}}(\boldsymbol{s})$, $I_{\text{sd}}(\boldsymbol{s})$, and $I_{\text{CV}}(\boldsymbol{s})$, for each prediction location. Results are shown in Figure 3.12. The variance and standard deviation maps reveal similar patterns

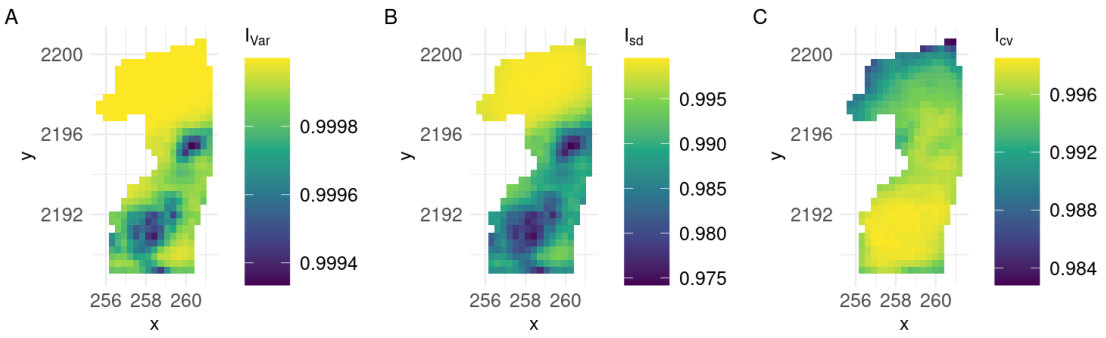


Figure 3.12: Comparing different information metrics. From left to right: I_{Var}, I_{sd} and I_{cv}.

.

to the standard map SD map (Figure 3.2B) that is heavily influence by the log-Gaussian assumption. The standard deviation metric looks more natural, in this example V_0 proved to be quite large (on order of 600,000) and so the metric is extremely close to 1 almost everywhere. Taking the square root of this (about 775) provided a number more comparable with the scale of the variation in the posterior intensity. Perhaps the most useful plot is I_{cv}. This shows that in the north, where

sampling effort is lower, we have less information than in the south, where sampling effort is higher.

One thing to note is that the posterior range of the random field is large relative to the size of the study region. For random fields where this range of correlation is lower, one might expect to see "bumps" of higher information around each sampled location. In this example, the large range of correlation may mask such a pattern. It would be interesting to apply these metrics in other settings to see whether this is the case. The exact scale of each metric is hard to interpret, for now only attempt to interpret these spatial differences in information gain in a relative sense (i.e. "more" or "less" information compared to other regions).

3.9 Discussion

This chapter presents a new approach to analysing point transect distance sampling data that incorporates several innovations in model specification, inference procedure, model evaluation and communication of results. These were applied to the particular problem of estimating the size and spatial distribution of populations of wild animals, but the approaches taken here will be useful to researchers who face similar questions and data structures in other application areas. In particular, we present a one-stage modelling approach to analyse imperfectly observed point transect data that incorporated a non-trivial spatio-temporal sub-sampling design, observation model, and spatially structured random effect. However, across applied spatial statistics there are many areas where similar ideas apply.

Our approach used the computationally efficient SPDE effect in combination with INLA to fit a one-stage Bayesian model in a fraction of the time it would take MCMC methods to fit the same model. The SPDE provided a fast and flexible method to estimate the spatial distribution of 'ākepa and, as we have seen, this random field component can lead to many different methods for investigating the fitted

model and communicating the results. We believe this could be a promising future direction for analysis of ecological datasets with non-trivial observation processes.

There are several natural extensions to the model and perhaps the most obvious is to include multiple years of survey data as was analysed in Camp et al. (2020). For this analysis, we only considered a single survey year as a simplified example of fitting distance sampling models using iterated INLA. Spatio-temporal extensions to this are possible, for example by considering space-time interaction models for the intensity function. `R-INLA` has the ability to fit space-time effects that can be represented as a Kronecker product of the relevant precision matrices (Blangiardo and Cameletti, 2015; Yuan et al., 2017), such as an interaction between a spatial SPDE effect and an auto-regressive temporal process. Blangiardo and Cameletti (2015) show how this can be done with numerous examples.

Another natural extension is to add complexity to the detection function. There is the potential to include explanatory covariates on detection function parameters. This is a common step in most distance sampling analyses. For example, we did not make use of the detection type (audio or visual) in our analysis, although this was recorded and detectability may vary depending on whether a bird is seen or heard. Other factors could also be considered, such as weather conditions, observer expertise, animal behaviour and morphological traits. Any additional parameters to the non-linear model components will need to be estimated using the linearised model form required by iterated INLA. It remains to be seen how complex the non-linear model component can become before this approach runs into problems. With additional complexity there is an increased risk of multi-modality and identifiability issues that may make the Laplace approximation inappropriate and could also lead to convergence issues for the iterated fitting procedure. We discuss this further in Chapter 5, Section 5.2.4.

In our analysis, we only briefly touch upon the problem of communicating uncertainty in maps and highlight the excursions methods as a new and positive addition

to this area. In our example, the threshold was chosen to illustrate the method but we believe that, for many analyses, discussions with relevant stakeholders could lead to agreement on a set of thresholds that are relevant given the aims and context of the analysis. This requires the input of both the relevant domain experts and statisticians so stakeholders understand the consequences of the thresholds and uncertainty levels they wish to consider.

The information gain metrics considered in Section 3.8.3 were largely unsatisfactory in the end, presenting not much more than a simple linear transformation of the original uncertainty maps (Figure 3.2). However, we believe the motivation for investigating these maps to be an important one. Presenting CV maps and SD maps without an additional understanding of how they are driven by model assumptions and not necessarily survey effort seems unsatisfying. Intuitively, it makes sense that areas with higher sampling effort should have lower uncertainty. However the opposite is the case in this analysis due to the preferential sampling of high density regions and the assumed log-Gaussian relationship. It remains an open question whether there is some measure of uncertainty that is purely a measure of sampling effort and spatial correlation but not an assumed distributional assumption for the random field. This would be a useful output for field ecologists to understand the implications of particular survey designs.

We end with a note comparing the analysis presented here and the more traditional two-stage approach. The one-stage thinned point process perspective can seem fundamentally different to a two-stage GAM fitted to count data. It can be tempting to think that, because the point process model does not require points to be binned into counts, this avoids some loss of information. However, when only distances to the observer are recorded this information loss is likely to minimal. This is because, for a single transect considered in isolation from others, there is no information in the data to inform within transect variation in intensity. Thus the key benefit to this approach is not necessarily the point process perspective, but

more the one-stage model and efficient inference method and the subsequent ease of working with posterior samples that average over observation process uncertainty. However, for datasets where the exact location of animals is recorded it could be that there is more advantage to the point process formulation since these data could be used to inform within transect variability and there would be a greater loss of data by binning into counts.

In summary, we have presented a novel framework for analysing point transect distance sampling data that introduces new approaches to model specification, inference procedure and communication of results. These developments will be of interest to those with similar data structures in spatial ecology as well as researchers in other areas who face related statistical challenges. In Chapter 5, we discuss some more avenues for future research inspired by some of the ideas presented here.

Chapter 4

Flexible density models for spatial capture-recapture

The aim of this chapter is to incorporate spatially structured random effects on animal density within spatial capture-recapture models. Spatial capture-recapture (SCR) (Borchers and Efford, 2008; Royle and Young, 2008) is a commonly used modelling framework to estimate the abundance and spatial distribution of animals using capture data on individually identifiable animals.

Although in the previous two chapters there has been a focus on using the SPDE effect, we will see in this chapter that this presents particular challenges for the SCR likelihood and inference procedure we use. Instead, we find there are some benefits to considering low-rank penalised regression splines on density instead of the sparse SPDE effect. Recall that, as discussed in Chapter 2, regression splines are closely linked to GRF effects and these modelling approaches play a near identical role in the model in the sense that we rarely have an a priori preference to a particular covariance structure or basis-penalty formulation in the context of species distribution modelling. These model components are included simply as a spatially correlated effect that addresses the problem of spatial autocorrelation. In this chapter, for reasons purely of computational efficiency, we prefer a low rank penalised regression

spline approach over a sparse GMRF.

Penalised regression splines are a go-to approach in spatial statistics to model non-linear heterogeneity in data (Cressie and Wikle, 2015; Wikle, 2015; Gelfand et al., 2010). Spline models have been implemented in SCR with a fixed degree of freedom with no smoothing parameter estimation in the package `secr` (Efford, 2021), a popular SCR software package. However, this involves choosing a small number of knots and does not estimate the smoothing parameters. Penalised smoothing splines have never been implemented within an SCR setting.

This chapter addresses this gap, implementing penalised smoothing splines with smoothing penalty estimation for SCR models and provides an overview of what to consider when fitting these models. We describe the contexts in which penalised splines on animal density should be considered, possible choices of penalised splines and their associated approximation methods, an approach to inference via Laplace approximate maximum likelihood, simulation studies and case studies fitting models to real data. The chapter ends with a reflection on the modelling choices available, situating the decisions taken in this chapter in a wider context, and suggestions for future research.

The chapter proceeds as follows: Section 4.1 introduces the key concepts of SCR and highlights the need to consider more flexible models of animal density within an SCR context; Section 4.2 reviews applications of SCR that aimed to investigate spatial variation in abundance and its possible causes; Section 4.3 provides a full statistical description of the SCR model; Section 4.4 describes the maximum likelihood inference method; Section 4.5 describes a variety of possible choices for the penalised regression spline; Section 4.6 presents a simulation study comparing two approaches, a sparse approximation and a low-rank one. We conclude that for the SCR likelihood, low-rank methods have some advantages; Section 4.7 applies a thin plate regression spline (TPRS) model to black bear hair snare data collected in Louisiana; Section 4.8 introduces a method for smoothing in regions with complex

spatial geometry; Section 4.9 ends the chapter by reflecting on the simulation studies and case studies, providing an overview of important things to consider when fitting penalised splines on density in SCR models.

4.1 Introduction

Spatial capture-recapture (SCR) methods were developed principally with the aim of estimating the density of wild animal populations. Similar to distance sampling, this approach estimates the detectability of animals as a key step in estimating the true density of all animals, detected or not. This section introduces the key concepts in SCR and the model components that have subsequently been developed to allow SCR models to answer a wider variety of ecological questions over and above estimating abundance. This section introduces the key concepts informally. A formal statistical description of the models is given in Section 4.3.

In an SCR context, the detectability of animals depends on distance via a detection function, just like in distance sampling. In distance sampling it is the distance to the observed animal that is assumed to affect detectability. In spatial capture-recapture models it is the distance to an *activity centre*. This activity centre can be thought of as a centre of gravity around which an animal moves and encounters detectors. There are thus two key components to the SCR model: first, the model that describes how many activity centres there are and how they are distributed across space; second, the model for the probability of detection as animals move around their activity centres. Both SCR and distance sampling methods idealise the behaviour of animals to a point location in two dimensional space. In distance sampling this point just represents the location of the animal when we saw it and this is fixed and does not change through time. In contrast, in an SCR model this point represents an *average location* of an individual over time, who we may have seen more than once on our detectors, which is the capture-recapture element to the

model. Note that, unlike distance sampling, this requires us to be able to identify individuals.

A common choice to model the detectability of animals as distance from the activity centre increases is the the half-normal function. This model, making the usual assumption of conditional independence between detections, given the activity centre location, says that an animal is most likely to be detected directly on top of the activity centre. Furthermore, detectability is radially symmetric around this, with no dependence through time. Being detected south of the activity centre has no implication for the next time step where the same animal can just as easily be detected to the north, for example. Naturally this simplified model of detectability has led to efforts to add more ecological realism to this model. Glennie et al. (2019) and Royle et al. (2016) allow activity centres to move over time and formulate this as a hidden Markov model. For this chapter we assume that activity centres are fixed for the duration of the survey.

SCR analyses in discrete time split the data up into a set of discrete *survey occasions* during which detections are summarised in some way. For example, for camera trap data this could be a count variable indicating the number of times an individual was detected on a specific detector during that occasion. Or for hair snare data it would be a binary variable, 1 if detected and 0 if not. The total survey window is known as a *session* during which the population is assumed to be closed and activity centres fixed (in conventional SCR analyses). Multi-session SCR models can be used to model multiple years worth of SCR data, where each year is a session assumed independent from the others. Sometimes the multi-session approach is used for spatially distinct regions that were surveyed in the same (or possibly different years). These models have a likelihood for each session with some parameters possibly shared across sessions.

The SCR model probabilistically describes two key processes: where are activity centres likely to be found (the point process model) and, given this, the consequences

for detectability at our detector locations (the detection function). This chapter is focused on the distribution of the activity centres themselves. Early spatial capture-recapture models used a homogeneous Poisson process to model the distribution of activity centres (Borchers and Efford, 2008; Royle and Young, 2008). Borchers and Efford (2008) also derived the likelihood for the inhomogeneous Poisson process.

The homogeneous Poisson process assumes the locations of animal activity centres are independent and randomly distributed across space. If the main aim is to estimate the total abundance within a region, SCR models can be fairly robust to violations of this homogeneity assumption (Efford and Fewster, 2013). Assuming a suitable survey design and avoiding over-extrapolation, these models tend to do well in estimating the total number of animals, even when assuming they have no spatial structure. Another way of putting it is the SCR model allows us to estimate the *average* density across a region, even if the true density is spatially varying.

However, often a survey is undertaken with broader aims than just estimating abundance. In these contexts the homogeneous assumption may no longer be sufficient. To give some examples, ecologists would like to be able to use SCR data to estimate species habitat associations, responses to disturbance, the effectiveness of management efforts, and identify hotspots of density that could be of key conservation concern. In these contexts, the model for the distribution of activity centres plays a key role and the homogeneity assumption may need to be relaxed. To address these research questions both the point process model and the detection model may need to be altered. If, for example, there is interest in whether a disturbance is causing individuals to ranger further afield, then clearly the movement model would be key to addressing this. On the other hand, if there is interest in how the spatial distribution and abundance of the population as a whole is potentially changing in response to a disturbance, i.e. changes in the distribution of activity centres, then the point process model may need to be adapted.

The most straightforward and natural extension of the homogeneous Poisson pro-

cess is to consider an inhomogeneous Poisson process by allowing the log-intensity to depend linearly on some spatially-referenced covariates. This model was also introduced in Borchers and Efford (2008) along with the homogeneous model. Section 4.2 reviews the SCR literature to investigate how spatially varying density has been addressed in published studies. Often this involves the inhomogeneous Poisson process but in a surprisingly large number of studies alternative models and methods were used.

One key assumption of the inhomogeneous Poisson process is that the locations of animals are assumed to be conditionally independent given the intensity. This assumption rests heavily on the idea that the available covariates, when added as independent linear fixed effects, provides a sufficiently flexible model to explain spatial heterogeneity. When this is not the case, and available covariates fail to sufficiently capture the spatial structure in animal locations, this assumption is perilous and it can lead to overconfidence in fixed effects estimates. In many contexts, the advice to avoid this problem is to include a spatially structured random effect that is flexible enough to account for any heterogeneity not accounted for by the fixed effects (Cressie and Wikle, 2015; Wikle, 2015; Gelfand et al., 2010). However, to date this approach has not been implemented in a spatial capture-recapture setting. This idea is the main focus of this chapter. Before introducing these methods we first summarise the SCR literature that focuses on applications where spatial variation in animal density is a key objective of the analysis.

4.2 Literature review

The purpose of this review is to highlight the types of questions ecologists and conservationists would like to be able to answer using SCR data and the current statistical approaches they take to answer them. This review is intended to situate the methods developed in this chapter within a broader set of aims and statistical

approaches. The focus is on any papers that wanted to, formally or informally, investigate spatial variation in animal density. The term animal density is vague in the context of SCR since it could refer to density of activity centres as well as the density of space-use around activity centres. For this reason we summarise the literature on both of these aspects of the SCR model. This distinction we return to in the discussion (Section 4.9) when considering whether a flexible density model for *activity centres* (the focus of this chapter) should be considered.

To find relevant papers in the literature we searched the Web of Science core collection database. The search criteria was for any papers under the topic "spatial capture recapture" OR "spatial capture-recapture" OR "spatially explicit capture-recapture" OR "spatially explicit capture recapture". The title of each paper was used to signify whether, explicitly or implicitly, the paper had any interested in spatially varying abundance. If there was any chance the paper may be relevant for this review then the abstract was read in full to decide whether the paper should be included in the review. Search terms such as "inhomogeneous density" and related phrases seemed to be unreliable for identifying these papers since many studies use different terminology and sometimes do not explicitly mention spatial variation in abundance other than implicitly through the stated aims of the study. This approach to reading the titles of all spatial capture-recapture papers seemed a better method although it did return a large number of search results.

If the abstract confirmed there was interest in spatially varying abundance (even if not stated explicitly) then I read the methods and results section of the paper and sometimes the rest of the paper if more context was necessary. I did this procedure for all search results from 1st January 2018 to present day (April 2020) which resulted in reading 58 papers in total. This cut-off date was arbitrary but seemed sufficient to gain an overview of the broad trends in the literature. There is, of course, the possibility I missed some important and interesting papers published before 2018.

The aims and objectives of the SCR studies that were included in this literature review can be grouped into a few broad categories:

1. To learn about the *spatial ecology* of the species, in the broadest sense of the term. Examples are understanding movement and activity patterns, species habitat associations, and inter-species dependencies.

2. To learn about the *population ecology* of the species, in the broadest sense of the term. Examples are estimating population growth rates, survival and recruitment rates, and spatial variation in these parameters.

3. To learn about *changes* in spatial distribution and population growth and possible causes of any change. Examples are estimating effects of management efforts, responses to possible disturbance, and responses to changes in climate or land-use patterns.

4. To identify areas of conservation importance. Examples are identifying population hotspots and areas where populations are declining.

These categories are not mutually exclusive; there is no clear dividing line between spatial and population ecology for example. But they serve to highlight the broad range of questions that SCR methods have been used to address.

Perhaps the most straightforward application of SCR to learn about the spatial ecology of a species is to fit an SCR model with an inhomogeneous Poisson process, using ecologically relevant spatial covariates as explanatory variables. Gaukler et al. (2020) investigate the effects of soil chemical properties on the density of small mammals, directly interpreting the fixed effect coefficients as indicating evidence for a causal relationship. This is a common approach: fit a model with possibly relevant fixed effects and interpret the estimated coefficient as directly relevant to an ecological hypothesis. Rather et al. (2021) do something similar but their hypotheses were more broad, seeking to identifying relevant spatial covariates from a

set of possible candidate covariates. They compared models using AIC and found that models that had similar support in terms of AIC were capable of predicting quite different levels of abundance. These differences were likely caused by using fixed effect estimates to predict density in areas that were not sampled or perhaps sampled with less effort. The ability to extrapolate to unsurveyed regions is one of the key motivations for learning about species-habitat relationships in a statistical framework. Mann et al. (2020) explore the habitat preference of Leopards by fitting an inhomogeneous Poisson process model with a 'habitat suitability' index as a covariate. This suitability index was also estimated from the same SCR data by fitting a MaxEnt model (Phillips et al., 2006) to detections at traps, ignoring the individual identities, and using the predicted surface as the covariate in the SCR analysis. Using multiple modelling frameworks, in this case SCR as well as MaxEnt, is a common approach that is discussed more below. In this case it seems questionable since it confounds the movement model and the distribution of activity centres. It also uses the data twice and so it seems unlikely their derived covariate would not be informative. Sutherland et al. (2018) fit an inhomogeneous model to estimate the effects of a pollutant on Mink density in two independent river systems. They also incorporate a non-Euclidean distance model to account for the geometry of the river systems. This is another example of directly interpreting a fixed effect estimate (pollution level) to address a specific hypotheses. These are just a few examples to illustrate the kind of papers that the inhomogeneous Poisson process has been used in. There are many more examples (McDonald et al., 2020; Furnas et al., 2020; Bajaru and Manakadan, 2020; Barrueto et al., 2020; Horn et al., 2020; Nelson et al., 2019; Welfelt et al., 2019; Havmøller et al., 2019; Loosen et al., 2019; Lamb et al., 2018; Berl et al., 2018). Examples of covariates considered in the above are habitat type, indices of abundance of other species (*e.g.* prey availability, predation pressure), distance to resources, and topographical and environmental conditions.

One of the key features of the inhomogeneous Poisson process is that the model

can be used to predict density in sub-regions of the total study area. This feature of the model seemed to only rarely be a motivation for using the inhomogeneous Poisson process, with only a few papers predicting density within sub-regions. Lamb et al. (2018) aim to estimate the effects of a habitat quality index on density and also predict density within distinct protected areas. They also produce a map of predicted density that appears to have sharp discontinuities, likely caused by discontinuities in explanatory covariates. A discontinuity may be ecologically realistic in some contexts but, if this is not the case, the model could perhaps benefit from some smoothing effect.

Many of these papers include figures that plot the predicted effect of a covariate on species density. Since the covariates are entering the model as linear fixed effects these effect curves appear as exponential curves on the density scale (assuming a log link). Often this is ecologically unrealistic, predicting exponentially greater or fewer numbers of animals as the values of the spatial covariate change, in the limit predicting zero or infinite density of animals with no optimal finite value for the covariate. This points towards the need to be able to consider more ecologically realistic relationships between density and spatial predictors. Incorporating penalised regression splines are one approach to doing this. One simple extension to linear fixed effects is to consider quadratic effects. Clare et al. (2019) used quadratic fixed effects to investigate forest disturbance on the distribution and abundance of American pine marten.

All of the above papers that fit an inhomogeneous Poisson process model used maximum likelihood methods. It seems that maximum likelihood is more commonly used than MCMC approaches, at least in the period considered here from 2018 to present, and there may perhaps be computational challenges involved in implementing inhomogeneous density models with current MCMC methods. Indeed one paper even used both MCMC and maximum likelihood methods to investigate spatial variation in different components of the SCR model (Barrueto et al., 2020). MCMC was

used to investigate the effects of spatial covariates on the encounter rate and, in the same paper, maximum likelihood was used to investigate spatially varying density.

Another approach to dealing with spatially varying intensity is to consider sites as statistically independent within a multi-session SCR model. The density in each session is assumed to be homogeneous and independent of the density in other sessions. This is also a common approach to investigate temporal trends if there is interest in population growth rates. Whether comparing estimates at different sites or through time, a key output of the modelling tends to be to plot the density estimates for each session. These plots are then interpreted in the context of the aims of the study. Note that for temporal trends this is in stark contrast to other approaches in population ecology which specifically model temporal dependence in abundance (e.g. Ricker growth, logistic growth) whereas the multi-session SCR approach treats sessions as conditionally independent given some shared parameters.

Broekhuis et al. (2021) investigate how densities fluctuate through time by fitting homogeneous multi-session models. To also investigate changes in spatial distribution they counted the number of predicted activity centre locations in two distinct regions and plotted these counts over time. This seems to be an ad hoc approach that, worringly, is based on interpreting hetetrogeneity in predictions from a model that assumed homogeneity.

Harihar et al. (2020) fit multi-session SCR to estimate density in three distinct regions that had different characteristics. The differences in density are then commented on as evidence for the effects of these different characteristics. In this paper there was specific interest in the effects of pastoral communities. They conclude that one site has low density due to a lack of functional connectivity with other sites and faces possible local extinction. This is a common approach in the literature. The characteristics of the sites are not included formally within the modelling framework as covariates. Instead the estimates from the model are interpreted within a wider context. The downside to doing this is the key inference of the paper is not

considered within a formal statistical framework. Some more papers that take this approach: Khanal et al. (2020) fit multi-session SCR models to investigate the influence of wild prey and livestock abundance on density; Kumar et al. (2019) compare temporal trends in two different protected areas that have received different levels of conservation efforts; Hearn et al. (2019) investigate anthropogenic disturbance at six sites but fitted models to each site separately and left one site out due to having a small sample size; Nelson et al. (2019) use multi-session SCR with an inhomogeneous model to investigate if relationships with covariates are different between sites; Espinosa et al. (2018) fit independent models to multiple sites with varying levels of accessibility to hunters; Tobler et al. (2018) fit independent models at two sites with different access to loggers. All of these studies had interest in spatially varying density but none of them formally included this in their model. For cases where sites are truly independent (*e.g.* in different countries in Tobler et al. (2018)) this multi-session approach seems reasonable. For other cases, where sites are nearby or even contiguous with one another (*e.g.* Kumar et al. (2019)), it may be problematic depending on whether the assumed independence of certain parameters is reasonable. A multi-session SCR model with homogeneous density assumed independent between sites and all other parameters shared across sessions is very close to a single session analysis with site as a factor covariate in an inhomogeneous model.

This final study mentioned above also includes an occupancy model fitted to the camera trapping data to investigate how the probability of detection varies with respect to distance to roads. This another theme common in the literature: to use SCR to estimate density and then some other modelling framework to investigate spatial patterns in the SCR data. Invariably this involves throwing away the information that identifies individuals and instead modelling detection rates (e.g. in a linear model) or presence/absence of the species during each occasion (e.g. in an occupancy model) or modelling total number of detections across all occasions (e.g. count variable in a binomial GLM). These other approaches, with the exception of

occupancy models, do not tend to model detectability. Sometimes the second model is informed by data collected in addition to capture history data. An example of this is using Resource selection functions (RSFs) (Johnson et al., 2006) informed by GPS telemetry data, which can form a complementary part of the analysis, in addition to the SCR model.

Kittle et al. (2021) aim to estimate space use within a protected area and investigate potential edge effects on the space use of individuals. They use homogeneous SCR for density estimation and then use encounter rate per day as a response variable in a linear model. They compare different linear models with covariates relating prey availability, distance to edge and habitat features. Models are compared by AIC and they use model averaging to synthesise estimates across models. The sampling scheme does not appear to be designed for the purpose of the study, there are very few detectors near the edge of the protected area and only one small segment of the boundary is surveyed. The linear model unnecessarily ignores the capture-recapture information when this could be incorporated directly into the encounter rate parameter in the SCR model. A smoothing spline on density could reveal edge effects by predicting lower density near the boundary and also reveal areas of greater uncertainty due to low sampling effort.

Yang et al. (2021) use SCR for abundance and then use an occupancy model to investigate spatial variation in the probability of detection. The authors note the problem of spatial auto-correlation in the occupancy analysis and use a restricted spatial regression model (Hughes and Haran, 2013) to address this. The problem of spatial autocorrelation is not mentioned in the SCR analysis. This suggests that while ecologists know spatial autocorrelation should be considered in spatial models, they lack the ability to do this with SCR models. If practitioners view SCR solely as a means to estimate abundance then it is likely they don't consider using spatial effects within SCR since the model is fairly robust at estimating overall abundance.

Wang et al. (2018) also do occupancy modelling with spatial smoothing and fit

multi-session SCR models to investigate differences between two sites defined as inside or outside a 'core area'. This appears to be a clear violation of the independence assumption of multi-session SCR since in this paper the core area is entirely contiguous with and surrounded by the other area that is assumed independent. This is a clear limitation of the multi-session approach and an inhomogeneous Poisson process with a 'core area' factor variable would have been a better choice. Penjor et al. (2018) also use an occupancy approach that considers spatial autocorrelation and a homogeneous SCR model purely to estimate density. The aim of the occupancy model is specifically to identify areas of conservation priority. However, since it is based purely on presence/absence it does not make use of the valuable capture-recapture data to identify spatial variation in abundance.

Lavery et al. (2020) use SCR to estimate abundance and also regress the total count of cat detections (ignoring recaptures) against spatial covariates within a binomial GLM. Rahman et al. (2018) fit a homogeneous model using MCMC and also do a MaxEnt analysis, viewing detections at traps as presences and adding pseudo-absences elsewhere. This seems questionable since the presences are the result of the sampling design as well as the space-use of the animals. It seems highly likely that any additional traps placed nearby traps that detected animals would also have been likely to detect animals which seems like a challenge for using pseudo-absences.

Allen et al. (2020) use a zero-inflated Poisson model on the number of detections at each trap. They also fit a homogeneous SCR model using MCMC and then use the posterior modes of each activity centre location to investigate spatial variation in density. This seems a contradictory approach that assumes a homogeneous density and then investigates predictions based on this assumption for signs of inhomogeneity. Lamichhane et al. (2019) do something similar by fitting a homogeneous density model using MCMC and then combining the posterior density of each detected activity centre location in some way to create a map of spatially varying density. This

approach seems akin to fitting a linear model and then interpreting a violation of the homoscedasticity assumption in the residuals as a key finding of the paper. There also may be some confusion about the difference between the posterior density for observed activity centre locations and the posterior for the intensity of the point process. Boulanger et al. (2018) do something similar but then do a second-stage analysis, modelling the predicted activity centre locations in a presence vs pseudo-absence model (which they call a RSF model) that did not account for uncertainty in the activity centre locations.

Another trend in the literature is the complementary use of telemetry data. Tirelli et al. (2019) use SCR data to estimate abundance and telemetry data to inform a RSF model. The two models are fitted separately and the SCR data is not used to infer any species-habitat relationships; Welfelt et al. (2019) jointly fit SCR with an inhomogeneous density and a RSF model to incorporate telemetry data. This study jointly modelled the data within a single statistical framework in which the SCR data was also used to model species-habitat associations. The specific relationship between the two models depends on what process the telemetry data is used to inform. Johnson (1980) defines a hierarchy of 'selection' processes that is still influential today. Selection (or preference) is defined as the difference between usage and overall availability, and one selection process is a lower-order than another if the latter is conditional on the former. Johnson (1980) suggest the following hierarchy:

1. **First-order selection**: the limits of the entire range of the population.
2. **Second-order selection**: the choice of home ranges of each individual in the population.
3. **Third-order selection**: the choice of resources to use within the home range of each individual.

The choices an individual makes within its home range (third-order) is conditional

on the location of the entire home range itself (second-order), because this defines what resources are available. Similarly, the choice of home range itself is conditional on the entire availability of feasible habitat that is accessible by the population. A typical RSF model is aimed at investigating third-order selection [1]

This hierarchy can also be mapped on to the different elements of an SCR model. For a constant density model, first-order selection is decided by assumption when specifying the spatial domain for the analysis. For a spatially-varying density model, first-order selection could be a blend of assumed and estimable model components since these models also have the flexibility to predict effectively zero density within the assumed spatial domain. In this case, the point process model accounts for both first- and second-order selection. The distribution of activity centre locations (home range centres) can reveal both the total extent of habitable regions as well as preference within those habitable regions. Third-order selection is the movement model around the activity centres, usually the conditionally independent 'teleporting' model implied by the detection function in most SCR analyses.

Within an SCR model, the covariates considered within the RSF model tend to be used to model the intercept parameter of the detection function. If a detector is placed in a preferable habitat within a home range then the individual is more likely to be detected there. This is the same order of selection (third-order) that RSF models informed by telemetry data seek to capture. This equivalence is the basis for incorporating telemetry data into SCR models. Linden et al. (2018) explicitly state that including telemetry data within an SCR analysis should be used to model third-order selection. Royle et al. (2018) also highlight the potential for SCR to unify datasets and concepts in movement, spatial and population ecology, one aspect of which is incorporating RSF models.

[1] The implied ordering of this hierarchy is open to debate. It seems reasonable to say that first- and second-order selection could *emerge* from third-order selection. An animal could make only local selection decisions and, from these movements, preference at larger scales is revealed. In this case one could say the order of conditioning should be reversed, first-order selection is conditional on second-order selection and second-order conditional on third-order. However, this framing is common in the literature and we think it can be useful to think of the SCR model in these terms.

Given this framing of selection, we can see that often the other modelling approahces included in these SCR papers are aimed at estimating third-order selection while accepting a homogeneity assumption on the second-order process. This may well be suitable in some cases. However, if there is more interest in larger-scale second-order selection, a flexible point process model may be preferable. This is the focus of this chapter.

One exception to assuming homogeneity with an RSF model is Loosen et al. (2019) who fit both inhomogeneous SCR and a RSF model based on SCR-derived data for presences and assumed pseudo-absences. They did model selection on the SCR density model but not the RSF model. Both modelling approaches agreed that land tenure was important but disagreed on the importance of the normalized difference vegetation index (NDVI). Given the framing given above we could say that this difference is potentially down to different covariates being relevant at different orders of selection. Laufenberg et al. (2018) use GPS and DNA-based SCR data in a two stage analysis. Firstly, the GPS data was used to identify relevant covariates in an RSF analysis. The second stage jointly used the GPS and SCR data in a combined model where the encounter rates and GPS data were modelled using the covariates identified in the first stage.

To end the review we discuss a paper published before January 1st 2018 but which deserves special mention as the only SCR paper of which we are aware that uses something other than an inhomogeneous or homogeneous Poisson process for the distribution of activity centres. Reich and Gardner (2014) define an SCR model for territorial species using a Strauss process. This allows for "repulsion" between points so, for example in the case of territorial species, activity centres are not likely to be found close to each other. The Strauss model presents computational challenges due to a high-dimensional integral with no closed form and inference in the paper is achieved using an approximation to the likelihood which allows for posterior sampling. The paper shows in a simulation study that accounting for this

can improve abundance estimates.

To summarise, there are many different motivations for considering variability in the space use of animals within SCR datasets. Many papers view SCR as being for abundance estimation only and use another modelling approach to consider the effects of spatial covariates. These other approaches seem to usually focus on third-order selection (space-use around activity centres) and not the distribution of activity centres themselves. For cases where second-order selection is specifically being considered, the inhomogeneous process is often used, typically with linear fixed effects and occasionally with quadratic effects, whose coefficients are interpreted through the lens of specific hypotheses. In some cases this can lead to unrealistic predictions of exponential relationships between covariates and density, and produce potentially unrealistic discontinuities in predicted density if factor covariates are used. In other cases, multi-session SCR was used, often in clear violation of the independence assumption. Some papers used methods to account for spatial auto-correlation in another model (e.g. GLM or occupancy) but did not in the SCR analysis. It is clear from this review that methods to account for spatial auto-correlation can play a key role in helping to make robust inferences from SCR data.

4.3 The spatial capture-recapture model

This section formally introduces the spatial capture-recapture model in discrete time. We assume we have K detectors at fixed locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K$, with T discrete survey occasions and n individuals detected. Each individual i has an activity centre at location $\boldsymbol{s}_i$. The connection between an activity centre at $\boldsymbol{s}_i$ and the probability of being detected on a trap at $\boldsymbol{x}_k$ on occasion t is via the detection function $p_{tk}(\boldsymbol{s}) = \mathbb{P}(\text{an individual at } \boldsymbol{s} \text{ is detected on trap } k \text{ during occasion } t)$. Usually this is a radially symmetric, continuous and monotonically decreasing function that depends on the Euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{x}_k$. The rate of decrease is estimated

from the recapture data as well as the intercept which models the probability of detection for a trap placed directly on top of an activity centre. One example of a detection function, that we use for the remainder of the chapter, is the *half-normal* detection function

$$p_{tk}(\boldsymbol{s}|p_0, \sigma) = p_0 \exp\left(-\frac{d_k(\boldsymbol{s})^2}{2\sigma^2}\right), \tag{4.1}$$

where $p_0 \in [0, 1]$ and $\sigma^2 > 0$. The intercept parameter p_0 is the probability of detecting an individual if the detector is placed at the same location as the activity centre. The scale parameter σ controls how quickly this probability decays as distance from the detector increases. Both p_0 and σ can be modelled with covariates and appropriate link functions to incorporate information that may affect the detection probability (*e.g.,* sex, age, weather conditions, detector type). The probability of detecting an individual with an activity centre at $\boldsymbol{s}$ at least once during the survey is

$$p(\boldsymbol{s}) = 1 - \prod_t \prod_k (1 - p_{tk}(\boldsymbol{s})), \tag{4.2}$$

suppressing parameter notation for for clarity. The probability of remaining undetected is $\prod_t \prod_k (1 - p_{tk}(\boldsymbol{s}))$. This probability is used in the likelihood to condition on the fact that we observe each individual at least once (we never observe the individuals we never detect) and also as a thinning probability to model the intensity of activity centres of *detected* individuals. This distribution of activity centres is modelled using a point process model. The most common choices are the homogeneous Poisson process with intensity λ and the inhomogeneous Poisson process with intensity function $\lambda(\boldsymbol{s})$. The point pattern of activity centres for *detected* individuals is a Poisson process with intensity $\tilde{\lambda}(\boldsymbol{s}) = p(\boldsymbol{s})\lambda(\boldsymbol{s})$ which implies the number of detected individuals within the study region A is Poisson with rate $\tilde{\Lambda} := \int_A \tilde{\lambda}(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$.

Each detected individual has capture history $\boldsymbol{\Omega}_i$ that consists of the data recorded about individual i on each trap and occasion. For a binary detector this capture history consists of binary variables ω_{itk}, indicating whether individual i was captured

on trap k on occasion t or not. For count detectors, such as camera traps, ω_{itk} records the number of times individual i was seen on trap k during occasion t. The set of all capture histories for detected individuals is $\boldsymbol{\Omega}$.

Let n denote the number of *detected* individuals. The SCR full likelihood (assuming, for the moment, that activity centres are observable) is the joint probability of observing n individuals with activity centres $\boldsymbol{S} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)^\intercal$ and capture histories $\boldsymbol{\Omega}$, given these individuals were detected at least once. Let $\delta_i = 1$ if individual i was detected at least once and $\delta_i = 0$ if not. Then for an observed dataset there are n detected individuals with $\delta_i = 1$ for all i since we do not observe individuals that are never detected.

Let $\boldsymbol{\Delta}_n = (\delta_1, \ldots, \delta_n)^\intercal$, then the full data likelihood is

$$\begin{aligned}[\boldsymbol{S}, \boldsymbol{\Omega}, \boldsymbol{\Delta}_n] &= [\boldsymbol{\Delta}_n][\boldsymbol{S}|\boldsymbol{\Delta}_n][\boldsymbol{\Omega}|\boldsymbol{S}, \boldsymbol{\Delta}_n] \\ &= [\boldsymbol{\Delta}_n] \prod_{i=1}^{n} [\boldsymbol{s}_i|\delta_i][\boldsymbol{\Omega}_i|\boldsymbol{s}_i, \delta_i],\end{aligned}$$

assuming independence between capture histories. We have $[\boldsymbol{\Delta}_n] = \text{Poisson}(\tilde{\Lambda})$ since this represents the model for the number of *detected* individuals. We also have

$$\begin{aligned}[\boldsymbol{\Omega}_i|\boldsymbol{s}_i, \delta_i] &= \frac{[\boldsymbol{\Omega}, \delta_i|\boldsymbol{s}_i]}{[\delta_i]} \\ &= \frac{[\boldsymbol{\Omega}_i|\boldsymbol{s}_i]}{[\delta_i]},\end{aligned}$$

since observing $\boldsymbol{\Omega}_i$ implies $\delta_i = 1$. Let $g_{itk}(\boldsymbol{s})$ denote the likelihood of observing ω_{itk} given individual i has activity centre $\boldsymbol{s}$. Note that $g_{itk}(\boldsymbol{s})$ is not conditioned on detection. Then assuming conditional independence between capture events we have $[\boldsymbol{\Omega}_i|\boldsymbol{s}_i]/[\delta_i] = \prod_t \prod_k g_{itk}(\boldsymbol{s}_i)/p(\boldsymbol{s}_i)$. The probability density of $\boldsymbol{s}_i$ given the fact that the individual was detected is $p(\boldsymbol{s}_i)\lambda(\boldsymbol{s}_i)/\tilde{\Lambda}$. Finally, since $\boldsymbol{S}$ is not observed,

we marginalise the likelihood which results in

$$[\boldsymbol{\Omega}, \boldsymbol{\Delta}_n] = \frac{\tilde{\Lambda}^n \exp(-\tilde{\Lambda})}{n!} \prod_{i=1}^{n} \int \frac{\prod_t \prod_k g_{itk}(\boldsymbol{s}_i)}{p(\boldsymbol{s}_i)} \frac{p(\boldsymbol{s}_i)\lambda(\boldsymbol{s}_i)}{\tilde{\Lambda}} \mathrm{d}\boldsymbol{s}_i$$
$$\propto \exp(-\tilde{\Lambda}) \prod_{i=1}^{n} \int \prod_t \prod_k g_{itk}(\boldsymbol{s}_i)\lambda(\boldsymbol{s}_i)\mathrm{d}\boldsymbol{s}_i. \tag{4.3}$$

As mentioned above, the exact form of g_{itk} depends on the type of data collected. Detectors can fall under three categories:

- **Single-catch**: these traps, such as a cage trap, have a maximum capacity of one and do not release any individual caught by the trap until the end of a sampling occasion.
- **Multi-catch**: these traps, such as mist nets, retain individuals until the end of a sampling occasion but do not have a maximum capacity (or at least, not one likely to be reached during a sampling occasion).
- **Proximity**: these traps, such as camera traps, do not retain individuals and do not have a maximum capacity.

These types of detector can generate a variety of response variables:

- **binary** variables record if an individual was detected or not.
- **count** variables record the number of times an individual was detected.
- **time** variables record the times of detection of each individual.

In this chapter we consider only discrete time SCR models and so describe the likelihoods for the binary and count variables. For proximity detectors, which generate binary response variables, each ω_{itk} is modelled as a conditionally independent Bernoulli random variable, given the activity centre, with probability of detection $p_{tk}(\boldsymbol{s})$. Then, in this case, the probability of observing capture history Ω_i, given the

activity centre is at $\boldsymbol{s}$, is

$$g_{itk}(\boldsymbol{s}) = \omega_{itk}^{p_{tk}(\boldsymbol{s})}(1-\omega_{itk})^{(1-p_{tk}(\boldsymbol{s}))}. \tag{4.4}$$

For count variables, each ω_{itk} is modelled as an independent Poisson random variable. The parameter of this Poisson distribution depends on p_{tk} and is most commonly represented in terms of the *encounter rate*. Because detections can happen multiple times within a single occasion, it is natural to consider the rate at which detections occur over time. The total number of detections in an occasion is then modelled as a Poisson random variable with a rate that is the integral of the encounter rate. Let h_{itk} denote the encounter rate of individual i on trap k during occasion t. This could potentially vary over time but here we assume it is constant within each survey occasion. Then the expected number of detections within occasion t on trap k, denoted H_{itk}, is just the product of duration of the occasion and h_{itk}. So for a count type detector we have

$$g_{itk}(\boldsymbol{s}) = \frac{H_{itk}^{\omega_{itk}}\exp(-H_{itk})}{\omega_{itk}!}. \tag{4.5}$$

The connection with the detection function is given by $p_{itk} = 1-\exp(-H_{itk})$. It is therefore possible to parameterise the model in terms of h_{itk} or p_{itk}. Datasets that include the exact time of detection can be analysed in a continuous time context by using this hazard formulation of the SCR model Borchers et al. (2014).

In this chapter we allow the intensity of activity centres to depend on spatially structured random effects. For the homogeneous Poisson process the intensity is a constant. Assuming a log link, it has the form $\log\lambda(\boldsymbol{s}) = \beta_0$. For an inhomogeneous Poisson process the intensity function can vary according to linear fixed effects of spatially indexed covariates $z_1, \ldots z_V$. In this case, $\log\lambda(\boldsymbol{s}) = \beta_0 + \sum_v \beta_v z_v(\boldsymbol{s})$. This allows the density of points to vary spatially. If we also wished to add a random effect $f(\boldsymbol{s})$, to account for spatial autocorrelation not explained by the fixed effects,

then the intensity function is

$$\log \lambda(\boldsymbol{s}) = \beta_0 + \sum_v \beta_v z_v(\boldsymbol{s}) + f(\boldsymbol{s}). \tag{4.6}$$

Given a basis $\{b_1(\boldsymbol{s}), \ldots, b_R(\boldsymbol{s})\}$ for $f(\boldsymbol{s})$ with associated coefficients $\boldsymbol{\xi} = (\xi_1 \ldots \xi_R)^\intercal$, the intensity depends on parameters $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$. The parameters $\boldsymbol{\xi}$ have an associated penalty which can be written as (proportional to) a multivariate Gaussian distribution. This defines a log-Gaussian Cox process (Møller et al., 1998) for the activity centre distribution, which generalises the inhomogeneous Poisson processes by allowing the intensity to be distributed as a Gaussian random field. We discuss various choices for the random effect $f(\boldsymbol{s})$ below in Section 4.5.

The likelihood (4.3) involves integrals that cannot be solved analytically and so are approximated using numerical integration. The integration scheme depends on a discretisation of space which is known in the SCR literature as a *mesh* or *mask*. Here we assume there are J discretisation locations, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_J$, with associated integration weights $\alpha_1, \ldots, \alpha_J$. Each integral in the likelihood uses the same integration scheme although this is not strictly necessary. The approximate likelihood is then

$$[\boldsymbol{\Omega}, n, \Delta] \approx \exp\left(-\sum_j \alpha_j p(\boldsymbol{s}_j)\lambda(\boldsymbol{s}_j)\right) \prod_{i=1}^{n} \left(\sum_j \alpha_j \prod_t \prod_k g_{itk}(\boldsymbol{s}_j)\lambda(\boldsymbol{s}_j)\right). \tag{4.7}$$

Usually the mesh is a regular grid and the integration weights are all the same value (the area of each grid cell). When implementing the SPDE effect below we also use the Delauney triangulation with associated weights as presented for the log-Gaussian Cox process likelihood in Simpson et al. (2016).

4.4 Inference using the Laplace approximation

We use the Laplace approximation method to approximate the integral of the likelihood (4.3) over $\boldsymbol{\xi}$. To give this in a general form we consider the case with multiple

random effects each given a basis representation and with associated parameters and penalties. We denote by $\boldsymbol{\xi}^{(j)}$ the parameters associated with the j-th random effect, gathered in a single parameter vector $\boldsymbol{\xi} = (\boldsymbol{\xi}^{(1)}, \ldots, \boldsymbol{\xi}^{(J)})$, and denote by $\boldsymbol{\theta}^{(-\boldsymbol{\xi})}$ all parameters that are not in the random effect parameter vector $\boldsymbol{\xi}$. Note that $\boldsymbol{\theta}^{(-\boldsymbol{\xi})}$ also includes the smoothing penalty parameters associated with the random effects, but not the basis coefficients themselves. Each random effect has density $\pi(\boldsymbol{\xi}^{(j)}|\boldsymbol{\theta}^{(j)})$, where $\boldsymbol{\theta}^{(j)}$ denotes the smoothing parameters for $\boldsymbol{\xi}^{(j)}$. Then, suppressing the $\boldsymbol{\Delta}$ notation for simplicity, the penalised likelihood is

$$[\boldsymbol{\Omega}, n|\boldsymbol{\theta}^{(-\boldsymbol{\xi})}] = \int [\boldsymbol{\Omega}, n|\boldsymbol{\theta}, \boldsymbol{\xi}] \prod_j \pi(\boldsymbol{\xi}^{(j)}|\boldsymbol{\theta}^{(j)}) \mathrm{d}\boldsymbol{\xi}, \tag{4.8}$$

The Laplace approximation of (4.8) is

$$\tilde{\mathcal{L}}(\boldsymbol{\Omega}, n|\boldsymbol{\theta}^{(-\boldsymbol{\xi})}, \hat{\boldsymbol{\xi}}) = (2\pi)^{p/2} |\boldsymbol{H}_{\boldsymbol{\theta}}|^{-1/2} [\boldsymbol{\Omega}, n|\boldsymbol{\theta}^{(-\boldsymbol{\xi})}, \hat{\boldsymbol{\xi}}] \prod_j \pi(\hat{\boldsymbol{\xi}}^{(j)}|\boldsymbol{\theta}^{(j)}), \tag{4.9}$$

where p is the length of $\boldsymbol{\xi}$, $\boldsymbol{H}_{\boldsymbol{\theta}}$ is the Hessian of $-\log\left([\boldsymbol{\Omega}, n|\boldsymbol{\theta}^{(-\boldsymbol{\xi})}, \hat{\boldsymbol{\xi}}] \prod_j \pi(\boldsymbol{\xi}^{(j)}|\boldsymbol{\theta}^{(j)})\right)$ with respect to $\boldsymbol{\xi}$, evaluated at $\hat{\boldsymbol{\xi}}$, the mode of $[\boldsymbol{\Omega}, n|\boldsymbol{\theta}^{(-\boldsymbol{\xi})}, \hat{\boldsymbol{\xi}}] \prod_j \pi(\boldsymbol{\xi}^{(j)}|\boldsymbol{\theta}^{(j)})$.

To compute the Laplace approximation we need the derivative of $[\boldsymbol{\Omega}, n|\boldsymbol{\theta}^{(-\boldsymbol{\xi})}, \hat{\boldsymbol{\xi}}] \prod_j \pi(\boldsymbol{\xi}^{(j)}|\boldsymbol{\theta})$ with respect to $\boldsymbol{\xi}$ in order to find the mode $\hat{\boldsymbol{\xi}}$ via numerical optimisation. Second derivatives are also required to calculate the required Hessian matrix $\boldsymbol{H}_{\boldsymbol{\theta}}$. Finally, the derivative with respect to $\boldsymbol{\theta}^{(-\boldsymbol{\xi})}$ of the Laplace approximation $\tilde{\mathcal{L}}(\boldsymbol{\Omega}; \boldsymbol{\theta}^{(-\boldsymbol{\xi})}, \hat{\boldsymbol{\xi}})$ is also required in order to perform maximum likelihood estimation on the non-random effect parameters. Although it is possible to write down the gradient for the inner optimisation and Hessian matrix (see Appendix A), the gradient of the Laplace approximate likelihood (4.9) with respect to $\boldsymbol{\theta}^{(-\boldsymbol{\xi})}$ is intractable. To compute these derivatives we use automatic differentiation using the R package `TMB` (Kristensen et al., 2016) which implements the Laplace approximation given a `C++` template of the log-likelihood.

4.4.1 A brief introduction to automatic differentiation

The basic insight of automatic differentiation is to view a (potentially complicated) mathematical function as a series of simple operations. For example the function $f(x) = 1/\sin(2x^4)$ can be written as

1. $x \mapsto x^4$
2. $x^4 \mapsto 2x^4$
3. $2x^4 \mapsto \sin(2x^4)$
4. $\sin(2x^4) \mapsto 1/\sin(2x^4)$.

Each step is itself a function and for all these elementary operations the derivative is known. The derivative of f can then be computed using these elementary derivatives by applying the chain rule. A list like the above is referred to as the graph (or tree) of the function since, for more complicated functions, it may have a branching structure. Higher order derivatives can be constructed by similar means. A key point is that, for a programming language like `C++`, all the elementary operations are known and have known derivatives.

When the function is a (log-)likelihood function we can use automatic differentiation to compute its derivatives, which are useful for optimisation algorithms and calculating variance estimates. Margossian (2019) review the application of automatic differentiation methods in statistics, with a particular focus on its importance for implementing Hamiltonian Monte Carlo. The R package `TMB` (Kristensen et al., 2016) implements automatic differentiation to compute the Laplace approximation given a `C++` template of the log-likelihood. A key feature of `TMB` is the optimisation routines it employs to speed up computation time. It does this mainly by allowing parallel computations and optimising the graph of the function by removing parts of the graph that are irrelevant to computing the required derivative. Note that for our approximate log-likelihood (the log of (4.7)) there are many operations in

the function. Each individual detected has an associated integral to evaluate which depends on a potentially high-resolution discretisation of space. We will see below that this complexity in the likelihood has consequences for using automatic differentiation with the SCR likelihood, with some models considered proving too large to hold in feasible memory limits during the optimisation step.

4.5 Basis-penalty random effects

This section describes some possible options for choice of smoothing spline. One of the main challenges in using basis-penalty random effects in complex hierarchical models is the computational burden of working with possibly large and dense multivariate Gaussian precision matrices. Evaluating the density of the random effect involves constructing the relevant precision matrix and then computing the determinant or log-determinant. If the random effect is parameterised using the covariance matrix then evaluating the density also involves computing its inverse, a potentially costly operation. Chapter 2 discusses in detail the SPDE approach to dealing with this computational challenge by constructing a sparse approximation to the dense precision matrix and using computational methods that can leverage this sparsity. However, we will see below that high-dimensional sparse precision matrices may not yield many computational gains in the context of using automatic differentiation with the SCR likelihood. One alternative to sparsity is to use a low-rank approximation. This has the advantage of requiring a relatively small number of random effects parameters which leads to precision matrices that are dense but low-dimensional.

4.5.1 Low rank smoothers

This section describes some options for low rank smoothers that are implemented in `mgcv` and can be easily constructed for use within `TMB`. It also discusses various

options for working with the null space basis associated with these effects.

Thin plate regression splines

A popular low rank smoother in spatial statistics is the thin plate regression spline (TPRS). The TPRS a is low rank approximation of thin plate splines (Duchon, 1977) and this approach has been implemented in the R package `mgcv` (Wood, 2003, 2017), a commonly-used package for fitting generalised additive models with penalised smoothing splines. This section briefly describes the TPRS and the low rank approximation method.

The thin plate spline penalty for a function $f(\boldsymbol{s})$ with $\boldsymbol{s} = (s_1, s_2)$ in a two-dimensional domain is

$$\int\int \left(\frac{\partial^2 f}{\partial s_1^2}\right)^2 + \left(\frac{\partial^2 f}{\partial s_1 s_2}\right)^2 + \left(\frac{\partial^2 f}{\partial s_2^2}\right)^2 \mathrm{d}s_1 \mathrm{d}s_2. \tag{4.10}$$

The thin plate spline penalty can be used for smoothing over domains of arbitrary dimension but the notation required to write down the general penalty is more cumbersome (Wood (2017), Section 5.5, has the details).

For thin plate splines the basis functions come in two varieties: basis functions that span the null space of the penalty (the penalty applied to these basis functions evaluates to zero) and a set of basis functions that have positive penalty. We call these the *null space basis* and the *penalised basis*, respectively. In two dimensions the null space basis for the thin plate penalty is $b_1(\boldsymbol{s}) = 1$, $b_2(\boldsymbol{s}) = s_1$, and $b_3(\boldsymbol{s}) = s_2$. Note that for identifiability reasons the constant basis function is usually removed in models that have an intercept parameter. This basis spans the set of polynomials of degree less than two. Such polynomials have second derivatives that are zero everywhere. Each of these functions (or any linear combination of them) have zero penalty and so the coefficients associated with these functions are not penalised during model fitting.

The penalised basis functions have a complicated form that we do not explicitly state here (see Wood (2017), Section 5.5). There are as many of these basis functions as there are unique combinations of covariate values in the data. In our context here, this is the number of mesh nodes since this represents the number of locations in the discretisation of space. The penalised basis amounts to placing a radially symmetric basis function centred on top of each mesh node. This can result in a large number of parameters if we want to use a high resolution discretisation of the domain. A solution to this is suggested by noting that the effective degrees of freedom of a fitted smoothing spline is usually much lower than the number of unique combinations of predictor variables (Wood, 2003, 2017). Therefore, it is possible that a low rank approximation may perform just as well, provided it is of sufficiently high rank to capture the required effective degrees of freedom.

Wood (2003) proposes taking an eigendecomposition of the penalty matrix and truncating the number of eigenvectors by selecting those with the largest eigenvalues. The motivation for doing this is that the largest eigenvalues are most relevant in computing the penalty. In `mgcv`, the default in two dimensions is to choose the first 30 vectors although this can be changed by the user. In two dimensions, this results in a low rank approximation that has 32 parameters in total (30 from the truncated basis and 2 from the null space) as opposed to the potentially thousands of parameters that would be required without taking a low rank approximation.

Duchon splines

Thin plate splines are a special case of a family of isotropic splines known as Duchon splines (Duchon, 1977). The generalisation of the penalty is constructed by expressing the penalty integral in the Fourier domain. The Duchon penalty can be derived by expressing the thin plate penalty using the Fourier transform and then augmenting the resulting integrand by multiplying with $\|\boldsymbol{u}\|^{2t}$, where $\boldsymbol{u}$ is the frequency variable and t is a user-selected parameter. The parameter t must be an integer

divided by two so that $2t$ is itself an integer value. If t is set to 0 then this recovers the usual thin plate penalty. The parameter t can be used to construct penalties with lower order derivatives but with optimal smoothing splines that are still continuous. If m is the order of the derivatives used and d the dimension of the domain then setting $t > d/2 - m$ ensures this continuity (Duchon, 1977). In Section 4.8 we consider the use of the Duchon splines to fit a smoothing spline in a higher dimensional Euclidean space that approximates the non-Euclidean geometry of a two-dimensional study region.

Low rank Matérn field

Kammann and Wand (2003) give an implementation of the Matérn random field in which the range parameter is not estimated but instead fixed to a pre-specified value. The authors recommend this to be the maximum displacement between any two locations in the domain although this is arbitrary and the suitability of this choice is context specific. This reduces the penalty to one that involves a single smoothing parameter. The truncated eigenvalue approximation approach used for the TPRS and Duchon splines can be applied to this special case and this is implemented in `mgcv` and accessed by using the `bs = 'gp'` argument.

Note that, in contrast to the SPDE approach, this implementation depends on choosing a fixed value for the range parameter, the basis functions do not have compact support, the smoothing penalty matrix is dense, and the effect includes a null-space basis similar to that for the TPRS and Duchon splines. This is, therefore, a very different approach to approximating the Matérn random field than the SPDE method.

Working with the null space

Using the standard TPRS results in a penalty matrix that is not full rank since some of the basis functions are not penalised. This effectively places an improper

prior on the random effect parameters. However, this presents a problem for the Laplace approximation. For an improper multivariate normal density $\pi(\boldsymbol{\xi})$ with precision matrix $\boldsymbol{Q}$, the Hessian of $\log \pi(\boldsymbol{\xi})$ is proportional to the rank deficient $\boldsymbol{Q}$ and so its determinant is zero. This means the normalising constant in the Laplace approximation does not properly account for the random effect. The `GMRFLib` (Rue and Follestad, 2003) software called by `TMB` to evaluate this multivariate normal density also uses the Cholesky factorisation which can only be applied to full-rank positive definite matrices and so attempting to use a singular precision matrix results in an error.

If the null space has dimension m then the precision matrix $\boldsymbol{Q}$ has m columns that are zero. There are two main ways to penalise the null-space and alter the precision matrix to make it full rank. Shrinkage smoothers, as described in Section 4.1.6 of Wood (2017), add a small value to the diagonal of $\boldsymbol{Q}$ for the columns that correspond to the null space. This value is chosen to be a fixed fraction of the smallest eigenvalue of the full-rank sub-matrix for the penalised basis. The result of this is that as the smoothing parameter tends to infinity the spline tends to zero everywhere, hence the term shrinkage smoother. This approach is implemented in `mgcv` using the `bs = 'ts'` basis and results in a full rank precision matrix that is appropriate to use with the Laplace approximation and `GMRFLib`. One disadvantage to this approach is that the scaling factor used to add a penalty to the diagonal is somewhat arbitrary and it can sometimes be challenging to find a value that penalises the null-space 'enough but not too much'.

The other approach to penalising the null space basis is to construct an additional penalty matrix and add this to the penalty. This comes at the cost of introducing a new smoothing parameter. This idea was explored in Marra and Wood (2012) and summarised in Wood (2017). Consider the eigen-decomposition of the precision matrix $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\intercal$ and denote by $\tilde{\boldsymbol{U}}$ the columns of $\boldsymbol{U}$ that correspond to the zero eigenvalues on the diagonal of $\boldsymbol{\Lambda}$. Then create a new penalty matrix $\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{U}}^\intercal$ with

associated smoothing parameter $\tilde{\tau}$. Setting the log penalty to the random effect to be $\tau \boldsymbol{\xi}^\intercal \boldsymbol{Q} \boldsymbol{\xi} + \tilde{\tau} \boldsymbol{\xi}^\intercal \tilde{\boldsymbol{Q}} \boldsymbol{\xi}$ results in a full rank penalty, with the second quadratic form specifically penalising functions that are not penalised by the first. Both these approaches to penalising the null space are considered below. The standard shrinkage smoother approach could lead to convergence issues during model fitting and this additional penalty matrix approach seemed to solve these problems.

4.6 Low rank vs sparse effects simulation study

The above section introduces several low rank smoothing approaches. However, the initial simulation studies we tried in this area were primarily focused on implementing the SPDE approach. This was for two reasons: first, it has already been implemented in `TMB`, which calls the same sparse matrix algebra library that `R-INLA` uses to make use of the sparsity in the precision matrix; second, the key message of Chapter 2 was that we can take the SPDE approach, commonly used in `R-INLA`, and use it elsewhere. Unfortunately the results of doing this were mixed, with models taking a long time to fit even with a coarse finite element mesh or not fitting at all and running into computational issues. In an attempt to alleviate these problems we turned to low rank smoothers. This section discusses two simulation studies to compare the low rank TPRS and the Lindgren et al. (2011) sparse implementation of the Matérn random field. In both simulation studies the intensity surface is simulated from the same model that is being used to fit to the data, so the Matérn effect is fitted to data that used a realisation of a Matérn field to simulate the data and similar for the TPRS. This was to provide a fair comparison of the performance of each approach under its own assumptions.

The purpose of these two simulation studies is not to present a direct comparison between the SPDE and the TPRS. Initial research in this area focused only on the SPDE effect but the computational challenges we encountered when doing this led

us to try an alternative approach, the low rank TRPS. The SPDE results presented here seemed to depend on particular parameter values and the intensity surface chosen to simulate data from. Other choices lead to models that we were unable to fit, usually due to memory errors that we could not resolve. Therefore, we present two simulation studies side by side as a proof of concept for each method but not as a direct comparison. The first simulation study shows that it is indeed feasible to fit the TPRS on density within SCR. The second simulation study shows that it is also possible to fit an SPDE effect if the mesh is of sufficiently low resolution, although in this case over 10% of simulations still failed.

4.6.1 Thin-plate regression splines simulation

The simulation study used a single fixed intensity surface created by sampling from the multivariate Gaussian density implied by the TPRS penalty. For each simulation, the spatial capture-recapture data was created by generating a point pattern of activity centre locations using this fixed intensity surface. Capture histories were then generated from these points with a fixed detection model. Proximity (binary) detectors and the half-normal detection function with no covariates were used. The intensity model was an intercept parameter plus the shrinkage TPRS effect which penalises the null space of the TPRS by adding a small value to the diagonal of the precision matrix. Starting values for the random effect parameters were chosen to help reduce computation time. The starting values were chosen by a pragmatic approach of fitting a Poisson count model with a shrinkage TPRS on space to counts derived from a constructed point pattern. This point pattern had the same number of points as given by an expected abundance estimate from a homogeneous model fit and the points were jittered around detectors in proportion to the number of individuals detected at each detector and using the estimated half-normal scale parameter. These points were then binned into counts and the starting values estimated by fitting a Poisson GAM with the shrinkage TPRS in `mgcv`. This is a

pragmatic way to get starting values by fitting to a point pattern that has roughly the correct number of points in roughly the right places. This could be improved in future by extracting the estimated activity centre locations from a homogeneous fit and using this with a point pattern model or also binned into counts. The binning into counts was necessary since `mgcv` does not implement a Poisson point process likelihood by default.

The true intensity and detector locations are shown in Figure 4.1. The intensity

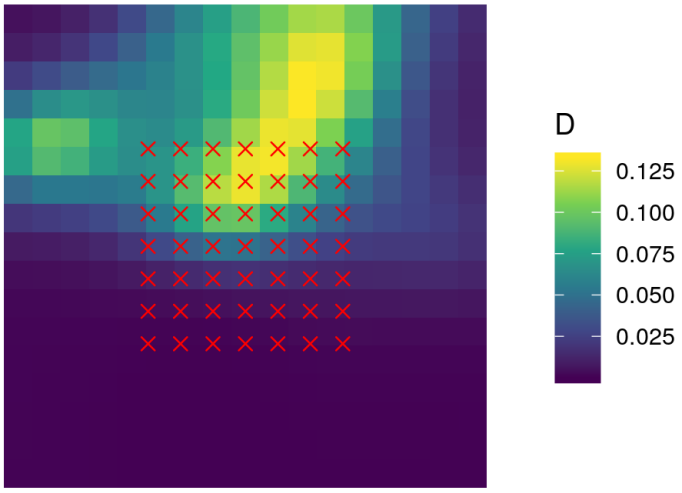


Figure 4.1: The true intensity surface and detector locations (red crosses) used in the simulation study.

function has most of the density to the north of the trapping array just by chance. We will see below that this has big implications for the abundance estimates. A total of 100 simulated datasets were generated from this intensity surface.

Figure 4.2 shows a summary of the simulation results for the detection model parameters and the expected abundance estimates. Estimation of detection function parameters appears relatively unbiased in these results. However, expected abundance estimates have positive bias for both the homogeneous and TRPS cases. The homogeneous estimates can be biased if the average density within the detectable region is different than outside. Looking at the true intensity in Figure 4.1, it seems

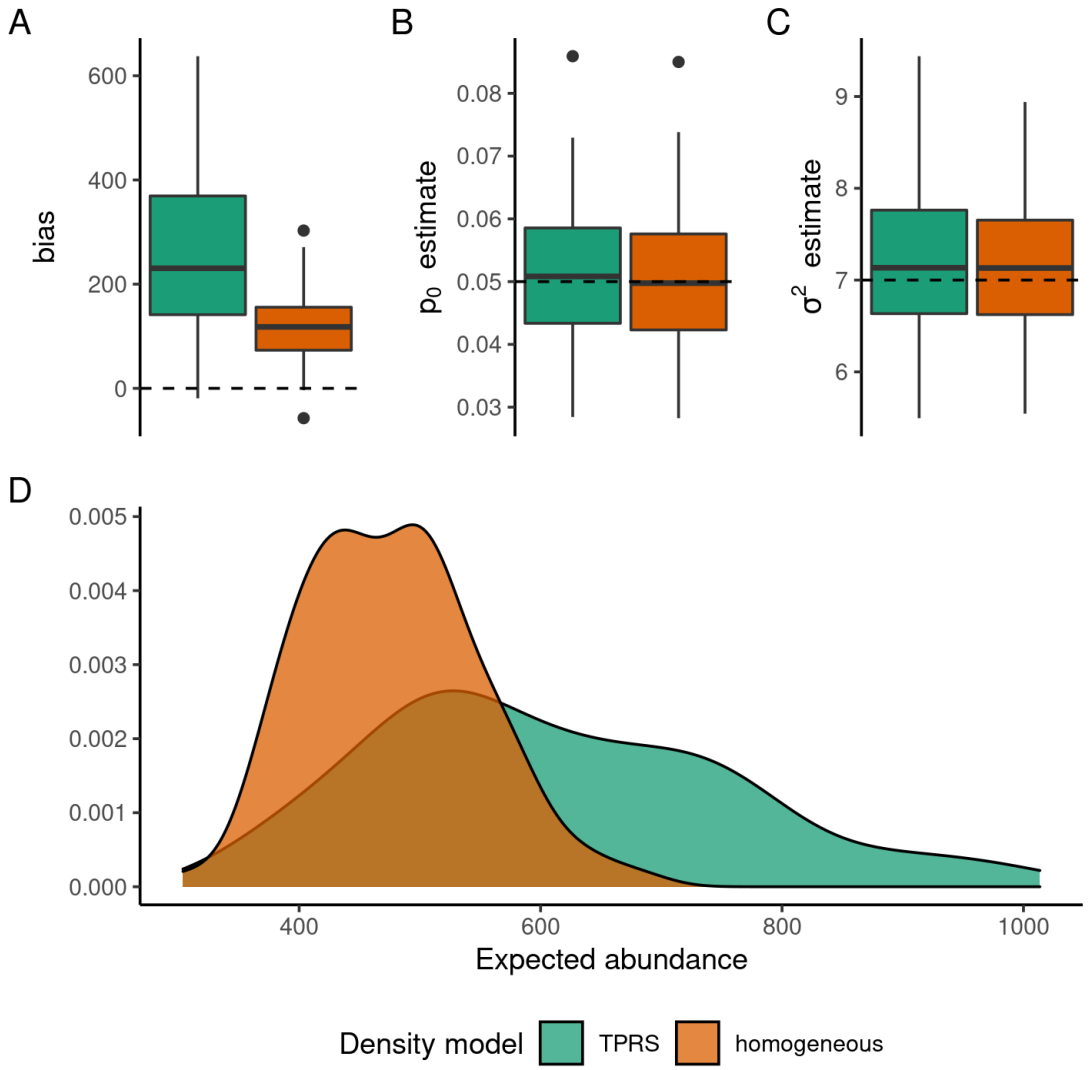


Figure 4.2: Summary of results of 100 simulations of the TRPS model compared to homogeneous density model. A) bias in expected abundance estimate (dashed line at zero) B) p_0 estimates (dashed line at the true value) C) σ^2 estimates (dashed line at the true value) D) smoothed histogram of abundance point estimates across the simulations.

that the density is, on average, higher within the detector array than outside. This results in overestimating abundance and highlights the importance of designing a representative sampling scheme. In the above simulation the trapping grid was fixed to be in the centre of the spatial domain. The intensity surface was chosen by simulating surfaces until there was likely to be 'a clear spatial pattern to fit to' in the generated SCR data. In other words, this surface was chosen specifically because

there was varying density within the detector array. This was inadvertently a case of preferential sampling (Diggle et al., 2010), although in this case the detector placement was fixed and the intensity surface was being preferentially sampled, whereas usually we would think about preferential placement of detectors.

The TPRS model is also positively biased and to a greater extent than the homogeneous model. This is for the same reason as the homogeneous model as well as the additional complication of the null space basis. It seems likely that the null space basis, which amounts to fitting a linear plane, would capture the north-south gradient. Then any predictions to the north in undetectable regions would be positively biased if the true intensity declines once beyond the range of detectability from the trapping grid. We investigate whether restricting the study region to smaller regions, closer to the detector array, influences this bias. To compare the estimated density with the true intensity used to simulate the data we use the integrated squared error (ISE), $\text{ISE}(\hat{\lambda}) = \int(\hat{\lambda}(\boldsymbol{s}) - \lambda(\boldsymbol{s}))^2 \text{d}\boldsymbol{s}$, for each intensity estimate $\hat{\lambda}$. Given the larger bias in expected abundance estimates for the TPRS model compared to the homogeneous model, we expect there should be, on average, a larger ISE for the TPRS model.

To investigate the effect of the null space functions on this error we also computed the ISE for two smaller regions that should be more constrained by observed data. To allow a direct comparison we divided the ISE for each region by the area of the region (since integrating over a larger area will necessarily generate a larger ISE), which we call average integrated squared error (AISE). One region was defined by buffering a short distance around the detector array (which in Figure 4.3 is labelled as the mid-region) and another smaller region defined by the extent of the detector array (which in Figure 4.3 is labelled as the small region). The null space basis functions are most constrained in the small region since here detectability is the highest. Figure 4.3 shows the AISE for the mean intensity prediction in each of these regions and compares this to the AISE for the homogeneous model. Figure

4.3A compares the AISE for the homogeneous model and TRPS prediction for the full-study region and confirms that the TRPS performs worse than the homogeneous model on average when considering the full study region, as we expected given the bias in the abundance estimates. However, Figure 4.3B shows that the TPRS model outperforms the homogeneous model when restricting our interest to regions more constrained by the observed data. The TRPS is better at estimating the density when considering only regions close to and within the detector array.

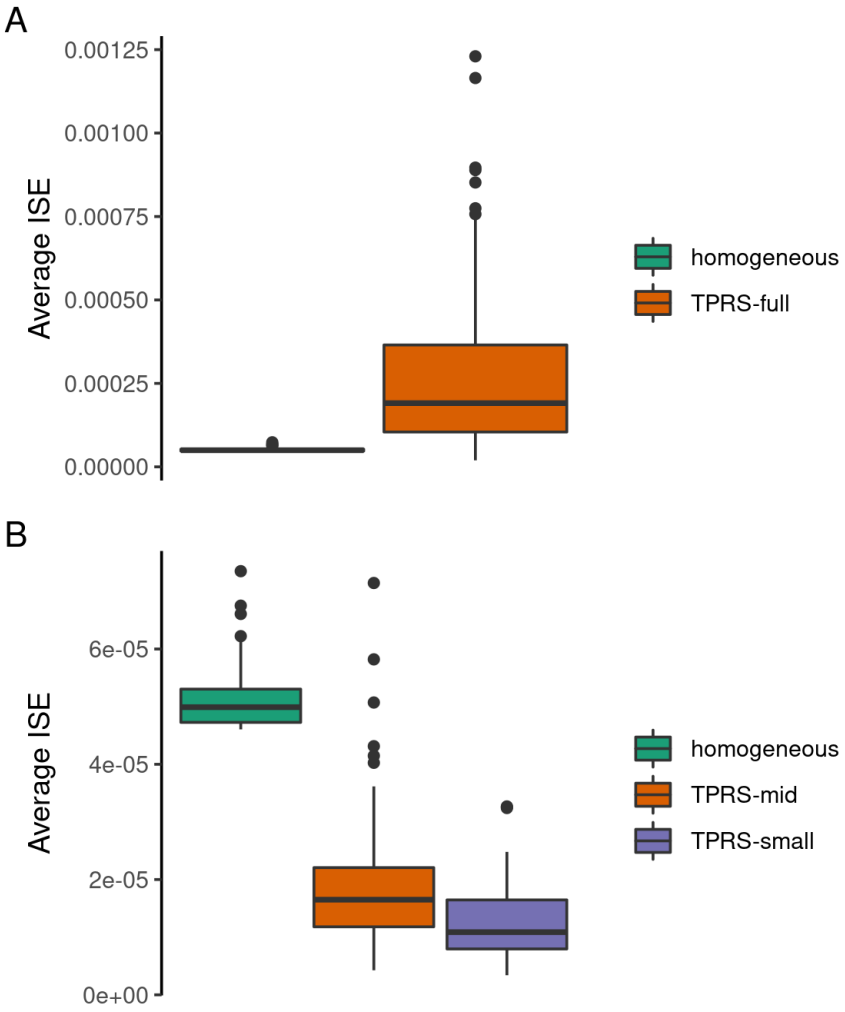


Figure 4.3: A) Average integrated squared error comparing the homogeneous model and the TRPS model over the full study region B) Average integrated squared error compared between the homogeneous model and the TRPS model over a mid-region defined by buffering the detector array and a small-region defined as the extent of the detector array

To more systematically investigate the effect of the null space and the relationship with sampling effort we also investigated the relationship between ISE and the probability of detection $p(\boldsymbol{s})$. The true detection function parameters used in the simulation gave $p(\boldsymbol{s})$ values that range between 0 and 0.72 over the whole study region. We considered regions defined by placing an exceedance threshold on $p(\boldsymbol{s})$. We considered thresholds of 0.01, 0.1, 0.25 and 0.5. As the threshold increases, these define smaller geographic regions that are more concentrated around the detector array. In each of these regions we consider the bias in abundance. The results are shown in Figure 4.4. This plot shows that even a very small threshold ($p(\boldsymbol{s}) > 0.01$)

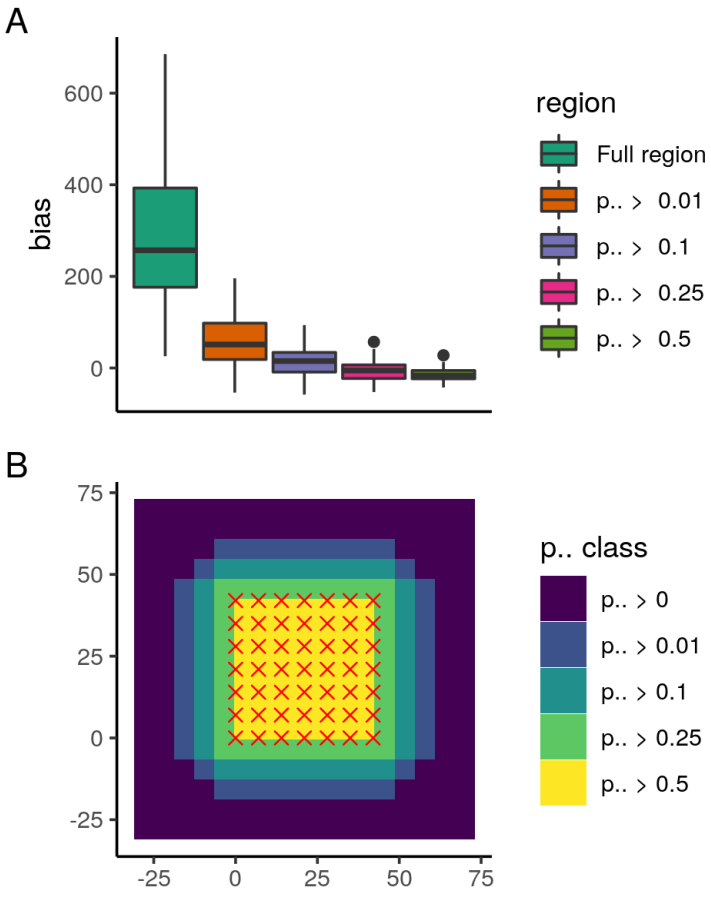


Figure 4.4: A) Bias in expected abundance estimate for different study regions defined using thresholds for overall probability of detection B) The regions corresponding to the probability of detection thresholds. As the threshold increases, the region shrinks towards the detector array

substantially reduces the bias in estimating abundance. Increasing the threshold reduces this bias further.

This points to a trade-off between bias in expected abundance and the area over which we want to predict the intensity. In order to minimise bias one may have to pay some price in reducing the area considered. This suggests a two stage modelling procedure. In the first stage, define a mesh over the largest plausible area and estimate the probability of detection. Then use these estimates to define a smaller region over which it is comparatively safer to estimate abundance when using the TPRS. Note that this simulation study is a clear case of a north-south gradient but in other datasets this may be less pronounced and so this prediction issue may be less severe. A pragmatic approach would be to plot just the null space basis function for any fitted spline to see the effect it would have on predictions. This relationship between the gradient of the linear plane and bias in intensity estimates means it is hard to give a threshold of detection that would work in all scenarios. A threshold of 0.01 seemed to do quite well in this simulation but this does not generalise to other contexts. It is also possible that alternative metrics, such as uncertainty in the density estimates, could be used to specify a threshold.

4.6.2 Matérn field simulation

The simulation study used a single fixed intensity surface, created by simulating from a Matérn field. For each simulation the data was created by generating a point pattern of activity centres from this fixed intensity surface. Capture histories were then generated from these activity centres using a fixed detection model. Proximity (binary) detectors and the half-normal detection function with no covariates were used. Starting values for the random effect parameters were chosen by fitting a log-Gaussian Cox process model to a constructed point pattern with the number of points chosen based on the expected abundance estimate from a homogeneous model fit. The points were jittered around detectors with the number of points at

each detector distributed in proportion to the number of individual detected at each detector. This is similar to the approach for TRPS but since the log-Gaussian Cox process with SPDE effect is already implemented in `inlabru` it was not necessary to bin the points into counts. The SPDE mesh, true intensity, and detector locations are shown in Figure 4.5.

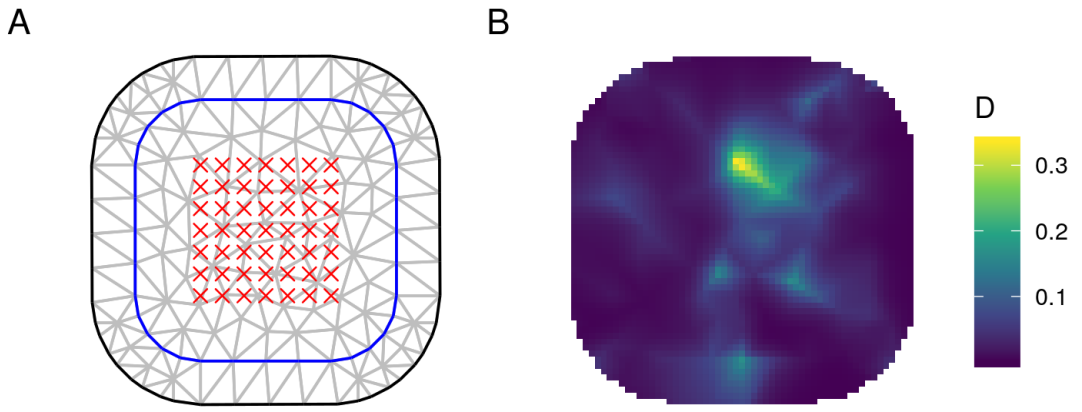


Figure 4.5: A) The SPDE mesh. The blue polygon marks the "study region" over which abundance is estimated. B) The true intensity function used. Red crosses mark detector locations in both

Figure 4.6 shows a summary of the simulation results for the estimation of abundance and the detection model parameters. This shows a general agreement between the Matérn model and homogeneous model when it comes to estimating detection function parameters p_0 and σ^2. There appears to be a slight positive bias in estimating σ^2 for the homogeneous model. However, there is a notable difference in abundance estimation. Both methods have bias (Figure 4.6A); the Matérn field has negative bias and the homogeneous model has positive bias. Figure 4.6D shows a smoothed histogram of the point estimates of expected abundance for both models. This is clearly multimodal although there is substantial overlap in the distributions.

To compare the estimated intensities with the true intensity we use the mean integrated squared error (ISE), calculated for both the Matérn field and the homo-

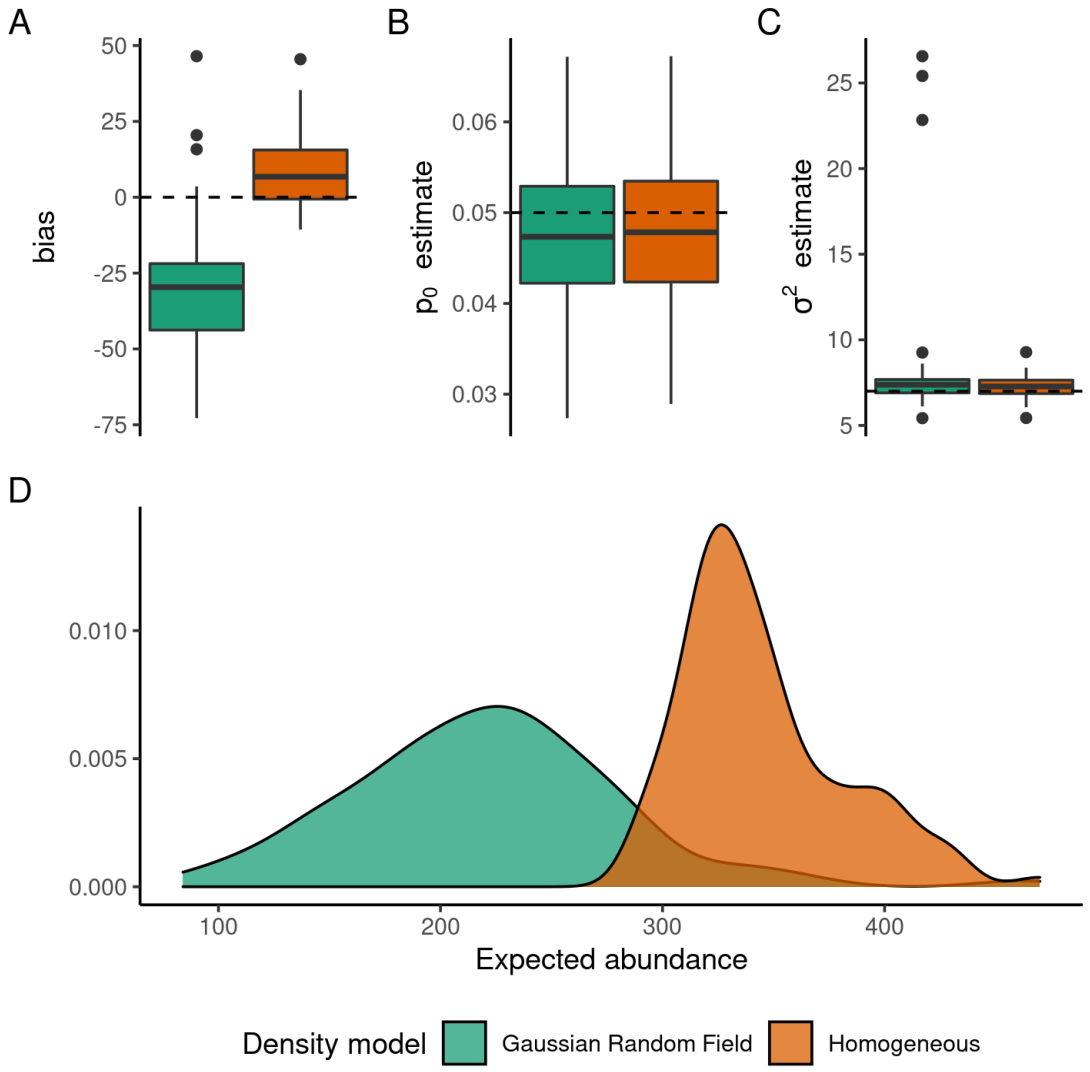


Figure 4.6: Summary of results of 100 simulations of the Matérn field density model compared to homogeneous density model. A) bias in expected abundance estimate (dashed line at zero) B) p_0 estimates (dashed line at the true value) C) σ^2 estimates (dashed line at the true value) D) smoothed histogram of abundance point estimates across the simulations.

geneous model. The results are in Figure 4.7. This shows there is much larger variation in the error for the Matérn field compared to the homogeneous model. On average the Matérn field has a lower mean integrated squared error. However, the large variance implies that it was not always superior to the homogeneous model. Inspecting a randomly sampled set of predicted mean intensity surfaces (Figure 4.8) reveals it is possible for the SPDE effect to be essentially zero (top right prediction).

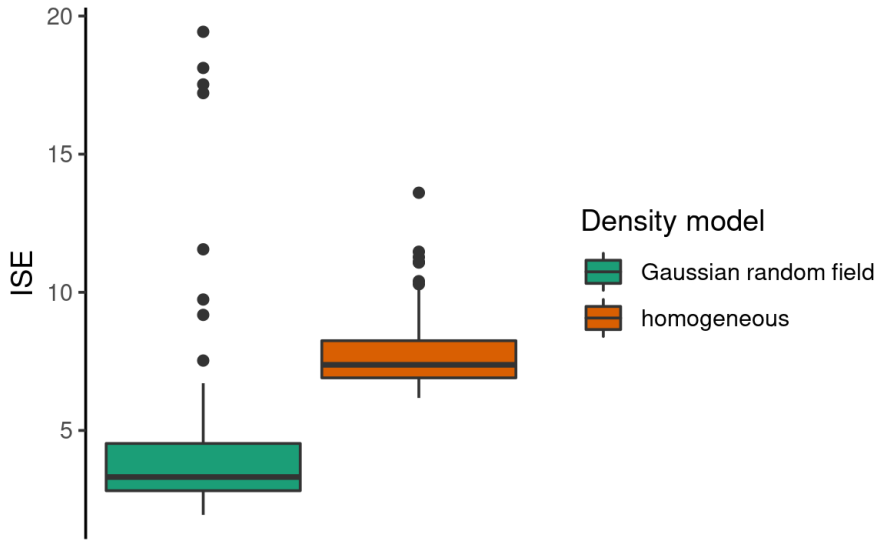


Figure 4.7: Integrated squared error for the Matérn field and homogeneous models.

There is some negative bias in abundance (Figure 4.6 A). There also seems to be some negative bias in the detection parameter p_0 (Figure 4.6B).

A cause for concern in generalising the results of this simulation study to real applications is the coarseness of the mesh. This mesh is much more coarse than would tend to be used in most applications of the SPDE effect. Problems with this may not appear in this simulation study since this coarse mesh was also used to generate the data but, in general, a higher resolution mesh gives a better GMRF approximation to the GRF. Here we were limited by computational issues if we used more than a few hundred mesh nodes. This is most likely caused by the automatic differentiation algorithm struggling to store the graphical representation of the likelihood function in memory. This is discussed more in Section 4.9. This problem occurred on a linux server with more than 80 gigabytes of RAM when attempting to use meshes with more than 300 or so mesh nodes.

Even with this simple mesh, 11 out of 100 simulations the model failed to fit. The consequence of this coarse mesh is that the true intensity (simulated from the same mesh as was used in fitting) looks strange and the effects of the triangulation are clear in model predictions (Figure 4.8), which are ecologically unrealistic and

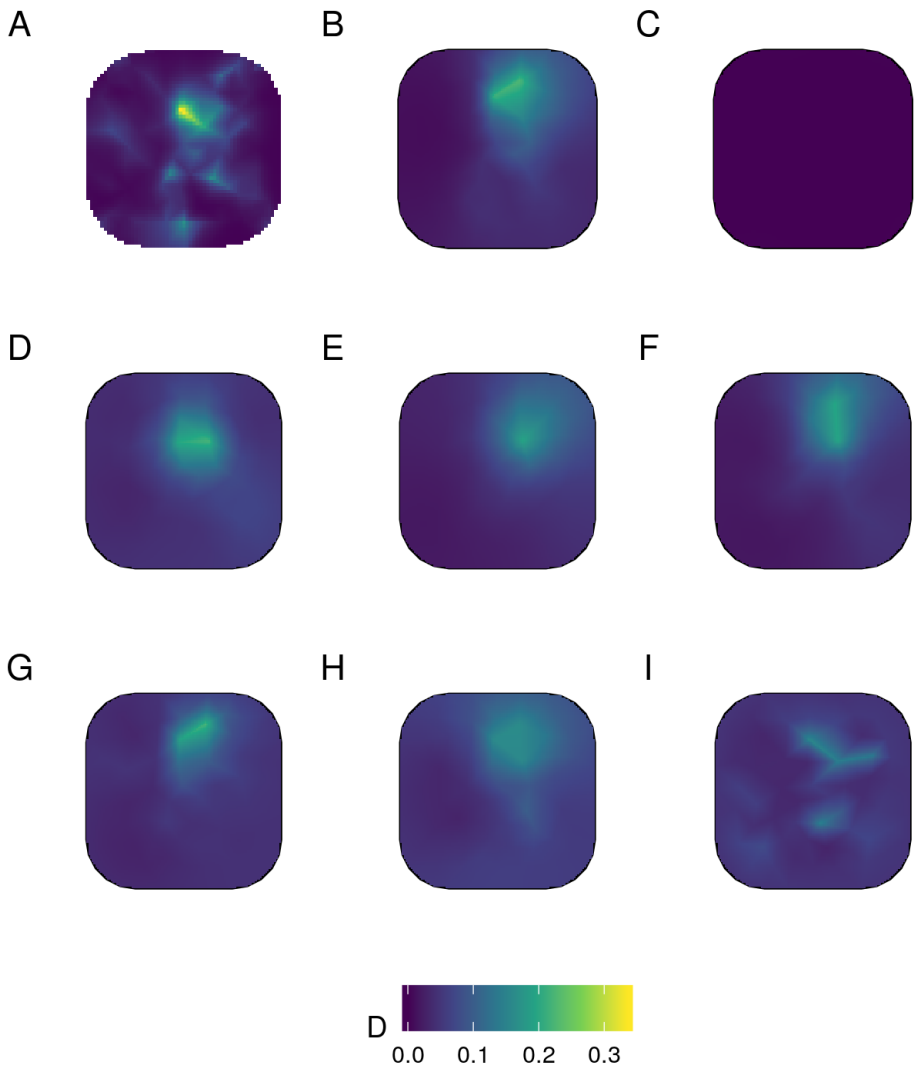


Figure 4.8: A) The true intensity surface used in the simulation study B,C,D) Three estimated intensity surfaces

purely an artefact of the mesh resolution. There also seems to be higher variance in the ISE for the SPDE effect than the homogeneous. This could be because a more flexible model can lead to less bias but more variance. For some simulated datasets, the homogeneous model did better than the SPDE effect in terms of ISE.

Figure 4.6C also indicates the model failed to estimate the scale parameter σ on three occasions. This suggests that using the SPDE is challenging within the SCR framework, at least as currently implemented using automatic differentiation

in `TMB`. A key conclusion based on the above two simulation studies is that low rank random effects may be computationally more feasible to use in the context of SCR models.

4.6.3 A note on integration schemes with the linear B-spline basis

Our early work on implementing the SPDE effect on density in SCR used a different integration scheme to the one used here. In the above simulation, the SPDE mesh nodes defined the quadrature points with weights calculated using the area of the hexagonal polygons in the dual mesh. This is the scheme suggested by Simpson et al. (2016) for the log-Gaussian Cox process likelihood. However, an alternative scheme we tried was to use a regular grid, as is common in SCR models, and to evaluate the intensity at these locations instead. We found this could lead to unstable estimates with some models predicting all the mass of the intensity function at just one or two mesh nodes. This can happen because of the interplay between the piece-wise linear basis functions and the integration scheme. The location of the mesh nodes and the location of quadrature points must be considered together. An inappropriate scheme can leave some flexibility for the likelihood to be maximised by increasing some basis coefficients to infinity without affecting the integral component of the likelihood that models the total number of points.

To see a toy example of an integration scheme issue, consider a one dimensional log-Gaussian Cox process on the interval $[0, 1]$. Then the one dimensional finite element basis function on this segment is $\log \lambda(s) = \eta(s) = \xi_0(1 - s) + \xi_1 s$. This is linear with $\eta(0) = \xi_0$, $\eta(1) = \xi_1$, and is the one-dimensional analogue of the pyramid shaped basis functions used in the 2D SPDE effect. The log-likelihood for a single observation at $y \in [0, 1]$ is

$$\mathcal{L} = -\int \exp(\eta(s))\mathrm{d}s + \eta(y)$$

Consider two integration schemes to approximate this integral. Scheme A is endpoint integration and Scheme B is midpoint integration. For scheme A we have integration locations at 0 and 1 with weight 1/2 at each. This gives approximate likelihood for a point at $y = 0.8$, for example, as

$$\mathcal{L}_A = -\frac{1}{2}\left[\exp(\xi_0) + \exp(\xi_1)\right] + \frac{1}{5}(\xi_0 + 4\xi_1).$$

For Scheme B we have one integration location at 0.5 with weight 1. This gives approximate likelihood

$$\begin{aligned}\mathcal{L}_B &= -\exp\left(\frac{\xi_0 + \xi_1}{2}\right) + \frac{1}{5}(\xi_0 + 4\xi_1) \\ &= -\exp(\xi_0/2)\exp(\xi_1/2) + \frac{1}{5}(\xi_0 + 4\xi_1)\end{aligned}$$

Since the observed point is at $y = 0.8$ we expect ξ_1 to be greater than ξ_0. For scheme A, as either ξ_0 or ξ_1 tends to infinity, the likelihood eventually decreases as a result of the approximate integral. This is not the case for scheme B. In this approximate likelihood ξ_1 can tend to infinity so long as ξ_0 compensates for this be tending towards minus infinity. This means the likelihood can be maximised by putting infinite weight at 1 (which is closer to 0.8) and letting x_0 tend to negative infinity which gives intensity 0 at location 0.

These issues with scheme A were encountered in the context of line transect distance sampling (where integration is along paths taken by observers) implemented for the log-Gaussian Cox process likelihood in `inlabru` and described in Yuan et al. (2017). Putting integration points within triangles (at a greater resolution than the mesh) gave the model this inappropriate flexibility. This is analogous to using a fine resolution regular grid (the standard in SCR analyses) together with the SPDE effect defined on the a more coarse triangulated mesh.

The integration scheme used in the simulation study is the same as that used

described in Simpson et al. (2016), using the mesh nodes as integration locations, and so avoids the problem with scheme B above. If we want to use a random effect with a B-spline basis within a SCR model then care should be taken to choose an integration scheme that is suitable. It is not clear to what extent these kinds of problems would carry over to a low rank smoothing spline. The problem seems to depend on the piecewise linear basis used for the SPDE effect which have compact support. In contrast, the low rank basis functions do not have compact support and may not run into this issue.

4.6.4 A note on prediction with Matérn field vs TPRS

This section discusses the difference in prediction behaviour between the low rank TPRS and the SPDE effect. The Matérn random field is stationary and isotropic which has implications for how predictions from this model behave when implemented with a B-spline basis. Note that with the Matérn simulation we did not have to restrict the predictions using $p(\boldsymbol{s})$ thresholds or some other method as we had to do for the TPRS simulation. This section discusses why this is the case.

Under a stationary and isotropic model, locations far enough away from the observations (such that the covariance between them is almost zero) will have predicted mean zero, i.e. observations have no effect on the expected value of the predictions far enough away from where the observations took place. In these region, the predicted mean is zero and, due to the stationarity, the covariance of the predictions will be the same as that which was estimated from the observed data. In other words, the smoothness in unobserved areas is assumed to be the same as the smoothness estimated within the observed region, but if we did not measure anything there then the random effect has a mean of zero. Note that this property of the GRF does not save us from having to design a sensible and representative sampling scheme. If the density in the observed region is higher than elsewhere then the intercept will lead to predictions that are biased higher, regardless of the fact that the random effect

has mean zero.

This is in contrast to the behaviour of predictions for the TPRS. The null space contains linear basis functions which, as prediction locations move further away from the observed region, will eventually dominate the predictions. In other words, all predictions on the log intensity scale will tend to plus or minus infinity, depending on the direction, as the distance increases away from the observed locations. This choice of basis means that the low rank implementation of the TPRS is non-stationary and non-isotropic. For this reason we need to take extra care when predicting density with these types of null-space basis functions.

4.7 Case study: Louisiana black bear data

This section presents a case study fitting a smoothing spline on density to SCR data collected on black bears in Louisiana. Given the computational challenges of using the SPDE approach, we decided to use the TPRS. In total 39 individuals were detected over 8 survey occasions. There were a total of 115 detectors. These were hair snares and so are a binary proximity detector type. Figure 4.9A shows the detector locations and the number of individuals detected on each trap and Figure 4.9B shows the total number of detections (including recaptures) on each trap. This data shows clear spatial structure, with a higher number of individuals detected and total number of detections in the north as compared to the south and west.

We fit a density model with an intercept parameter and TRPS effect. No covariates were used to model detection parameters. Figure 4.10 shows the estimated detection probability and estimated mean density, together with the standard deviation and coefficient of variation of the estimated mean density. Figure 4.10A shows that the smoothing spline has captured some of the spatial structure we expected from the summary data in Figure 4.9, with a higher density in the north as compared to elsewhere in the study region. The standard deviation (Figure 4.10B) is

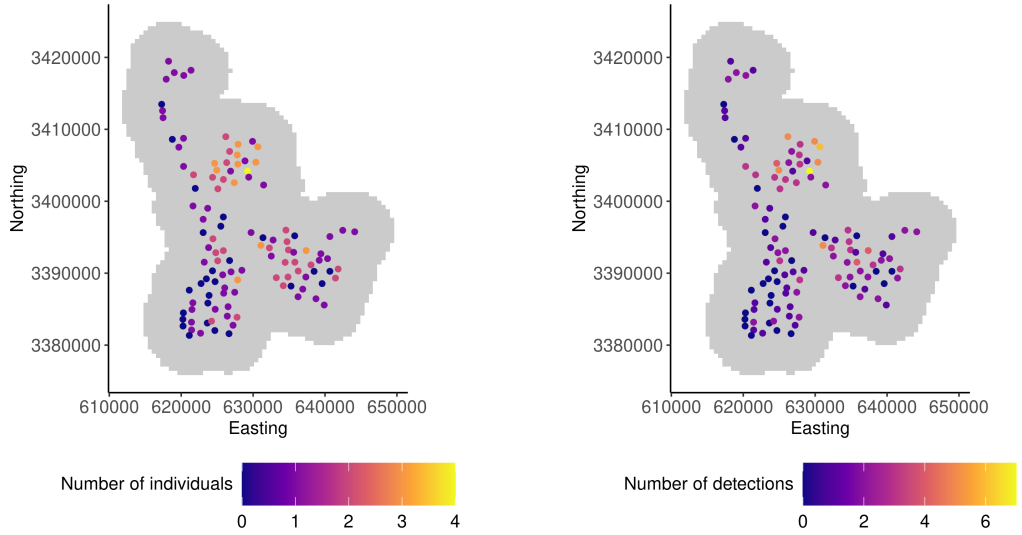


Figure 4.9: A) Detector locations and the number of unique individuals detected at each detector. Greyed area is the mesh used in all analyses. B) Detector locations and the total number of detections (including recaptures) at each detector.

also higher in this region as compared to elsewhere. This is due to the log-normal relationship between the random effect and the density. If X is normally distributed with mean μ and variance σ^2 then $Z = \exp(X)$ is log-normal with a mean-variance relationship given by $\text{Var}(Z) = \exp(\sigma^2 - 1)[\mathbb{E}(Z)]^2$. This implies that the variance will generally be higher when the expected density is higher. For this reason we also plot the coefficient of variation in Figure 4.10D.

Figure 4.10B shows the probability of detection with high levels of detectability in most locations in the mesh. This indicates that the null-space basis is not likely to be having an unwanted effect on predictions by extrapolating into regions not constrained by observed data. The overall abundance estimates from this model are comparable to a homogeneous model estimate, as are the estimates of the detection parameters.

We also compared this model with unpenalised TPRS fitted using `secr` (Efford, 2021). We fitted models with a pre-specified number of knots ranging from 9 to 28 and compared models using AIC and AICc. The results are shown in Figure 4.11. These figures show that there seems to be no additional benefit to model fit

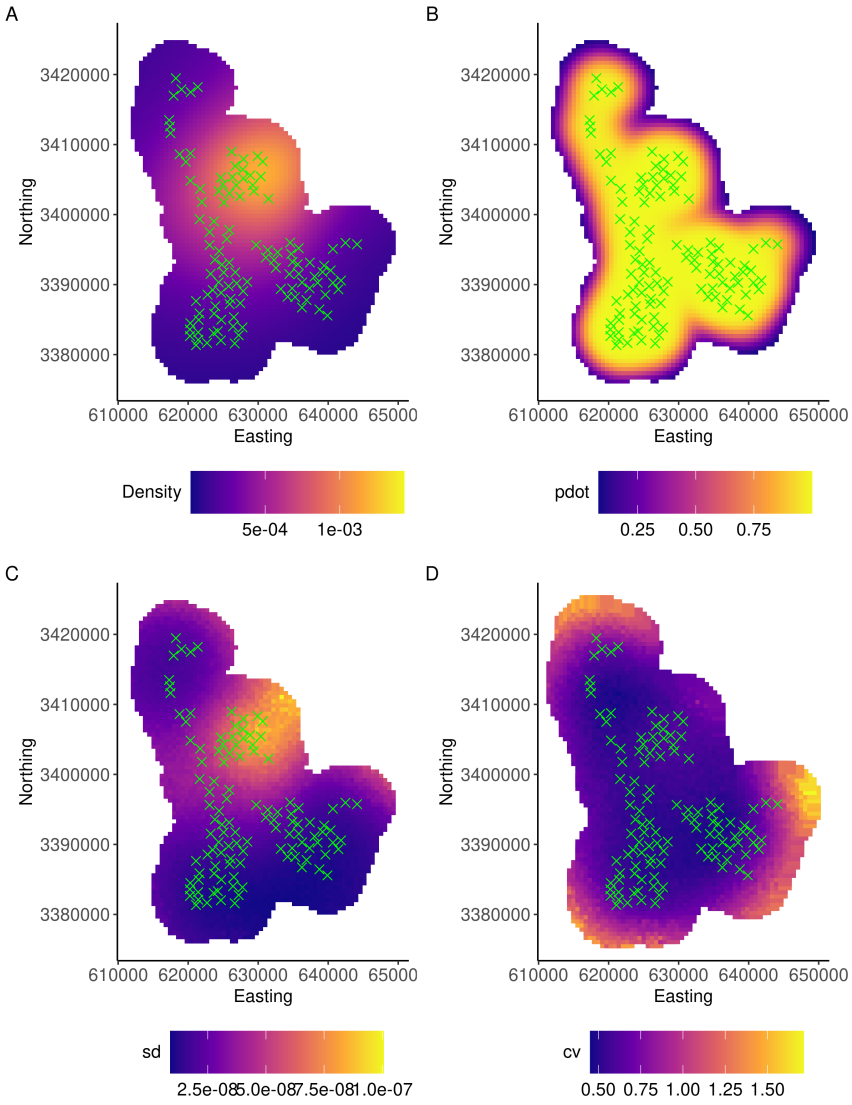


Figure 4.10: Summary of the predictions from the `TMB` model. A) Expected density B) Expected probability of detection C) Standard deviation of density prediction D) Coefficient of variation of density prediction. Green crosses are detector locations in all plots.

by increasing the number of knots with 9 knots being the lowest for both AIC and AICc. This is also preferred to a homogeneous model fit although for AICc the difference is quite small. The effective degrees of freedom of the random effect with design matrix $\boldsymbol{X}$ and precision matrix $\boldsymbol{Q}_\tau$ is $\text{tr}\left[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{Q}_\tau)\boldsymbol{X}^T\right]$, where τ is the smoothing parameter. The effective degrees of freedom for the TPRS in the `TMB` is approximately 24.8 of a possible maximum of 32 (30 truncated basis vectors and

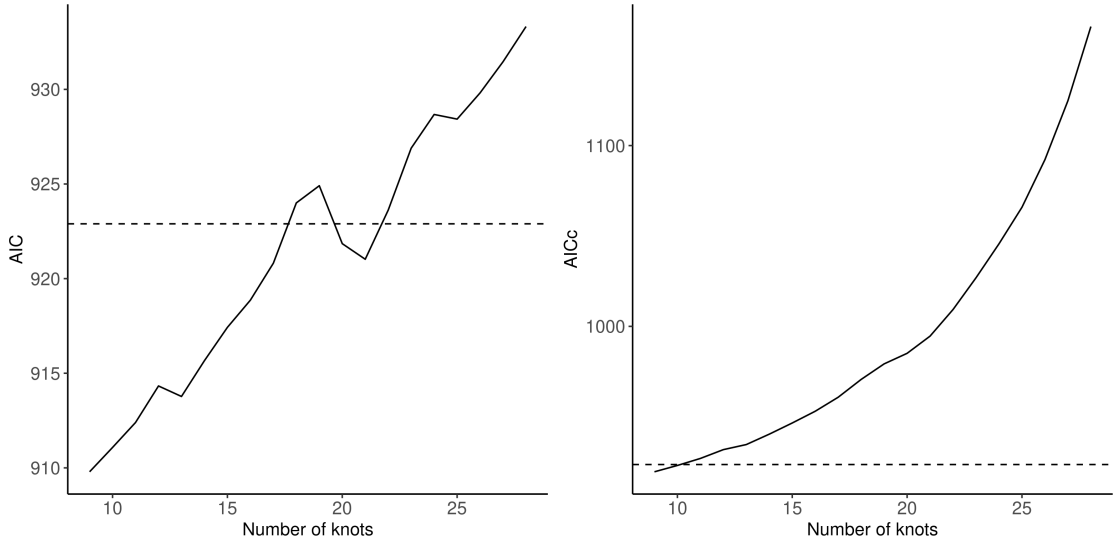


Figure 4.11: AIC of `secr` model fits with unpenalised TPRS effect and varying number of knots A) AIC B) AICc. Horizontal dashed line shows the AIC (or AICc) of the homogeneous density model

two null space functions). This is substantially larger than the degrees of freedom in the unpenalised spline selected by AIC which has 9 knots and two null-space bases and so has 11 degrees of freedom.

It is not clear why the effective degrees of freedom is higher in the penalised model fit. It seems that the penalty implied by AIC is stronger relative to a penalised regression spline with smoothing parameter estimation. We would not necessarily expect these to agree. However, the predicted density in the `TMB` model (Figure 4.10A) does not seem to have any notable increased flexibility over the unpenalised prediction with 9 knots (Figure 4.12A). This is a surprising result since AIC and AICc saw no additional benefit to increasing the degrees of freedom of the smooth and it does not appear as though the penalised fit gives qualitatively different estimates. However, it is possible that the first nine truncated eigenbasis vectors have sufficient flexibility to produce qualitatively similar fits to the 24.8 effective degrees of freedom of the penalised fit.

Another explanation is that the point estimate for the smoothing parameter might not be representative of the variability in the likelihood of this parameter.

If lower values of τ are also quite likely, this could reduce the effective degrees of freedom. A summary of the unpenalised model fit with 9 knots using `secr` is shown in figure 4.12. This shows broadly similar trends in the predicted density surface although the high density region is less radially symmetric as highlighted in figure 4.12D which compares the unpenalised and penalised models.

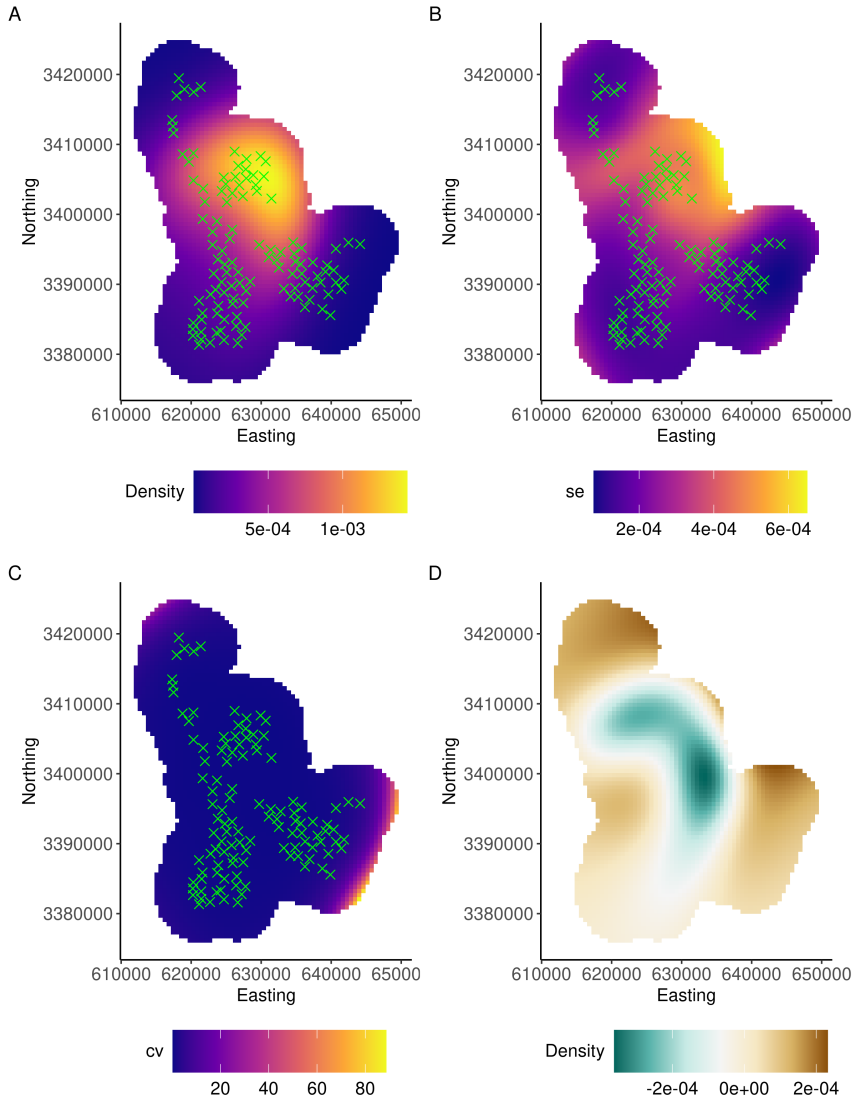


Figure 4.12: Summary of the predictions from the `secr` model A) Expected density B) Standard error of expected density C) Standard error divided by expected density D) `TMB` expected density minus `secr` expected density

This case study has shown that it is possible to use penalised TPRS to estimate a density surface from real world SCR data. This density surface appears to be

consistent with the spatial structure in the data measured at detector locations. It also produces comparable abundance estimates and detection parameter estimates to models that assume a homogeneous density. However, when compared to an unpenalised fit we notice some differences. The best model selected by AIC seems to have a more variable density surface but has fewer degrees of freedom. It is not clear from this case study that by allowing a penalised spline we have achieved a significantly different estimate from the unpenalised spline.

4.8 Effects that account for complex spatial geometry

Within spatial ecology there has been substantial interest in developing spatial random effects that can account for the geometry of the spatial region (Bakka et al., 2019; Scott-Hayward et al., 2014; Miller and Wood, 2014; Wood et al., 2008). This section investigates using such random effects within an SCR model.

There are three approaches that have been widely used within spatial ecology contexts, though never within SCR. The soap film smoother (Wood et al., 2008) uses a smoothing penalty that is responsive to the shape of the boundary of the spatial domain and boundary conditions which can be assumed fixed or estimated. This is implemented in `mgcv` with a B-spline basis. The complex region spatial smoother (CReSS) (Scott-Hayward et al., 2014) also uses a radially symmetric B-spline basis. This approach uses model averaging to vary the number of knots and the scale of the basis functions. Bakka et al. (2019) adapted the SPDE approach for the Matérn field (Lindgren et al., 2011) to account for barriers that effectively stop any correlation across them. This also uses a linear B-spline basis implemented using finite element methods.

Given that these approaches all rely on a B-spline basis which may prove difficult to implement in SCR using automatic differentiation, for similar reasons to

the computational issues encountered with the SPDE above, we considered a less well-known approach that is based on low rank Duchon splines. Miller and Wood (2014) call this approach *generalised distance splines* (GDS). The approach is based on using a higher-dimensional Euclidean space that approximates a lower dimensional non-Euclidean geometry. The idea is that in this higher dimension space the Euclidean distance between points matches the non-Euclidean distance in the lower dimensional domain. Then one can use a smoother in this Euclidean space to smooth between locations. The way this higher dimensional space is constructed is by applying classical multidimensional scaling (Gower, 1966; Mardia, 1978) to the matrix $\boldsymbol{D}$, where $\boldsymbol{D}_{ij}$ is the non-Euclidean distance between discretisation locations i and j.

One can think of the matrix $\boldsymbol{D}$ as containing a discrete representation of the non-Euclidean geometry of the spatial domain. Classical multidimensional scaling is an approach that seeks to approximate this geometry. It does this by representing a set of M distances with locations in an $M-1$ dimensional Euclidean space. This space is then approximated using a truncated Eigen decomposition, keeping the first p principal components. Miller and Wood (2014) suggest Duchon splines (Duchon, 1977) are a natural smoother to consider in higher dimensions and use the low rank implementation in `mgcv`. Initial exploratory models tried to implement this using the 'shrinkage smoother' approach but found that the likelihood was unstable and often led to errors in the optimiser. The arbitrary scaling factor in the shrinkage smoother approach could lead to precision matrices that were numerically singular. Our extra penalty approach avoided this problem. Below we present a simulation study on a horseshoe shaped domain and a case study on leopard data that was the motivation for developing this approach.

4.8.1 Horseshoe simulation

We tested the generalised distance spline (GDS) using the horseshoe domain that is common to demonstrate non-Euclidean smoothing methods (Wood et al., 2008; Scott-Hayward et al., 2014; Miller and Wood, 2014) and is often referred to as Ramsay's horseshoe after it was introduced by Ramsay (2002). The function defined on the horseshoe in those papers was assumed to be a Gaussian response and could take negative values. To turn this function into a valid intensity surface we added a positive intercept term and adjusted the gradient of the slope to ensure the function was positive everywhere. Detectors were placed in the domain using a stratified random sampling domain and then fixed for all simulations. Detection parameters were also set to fixed values for all simulations and chosen to give good probability of detection across the domain. Figure 4.13 shows the true intensity, detector placement and probability of detection used in all simulations.

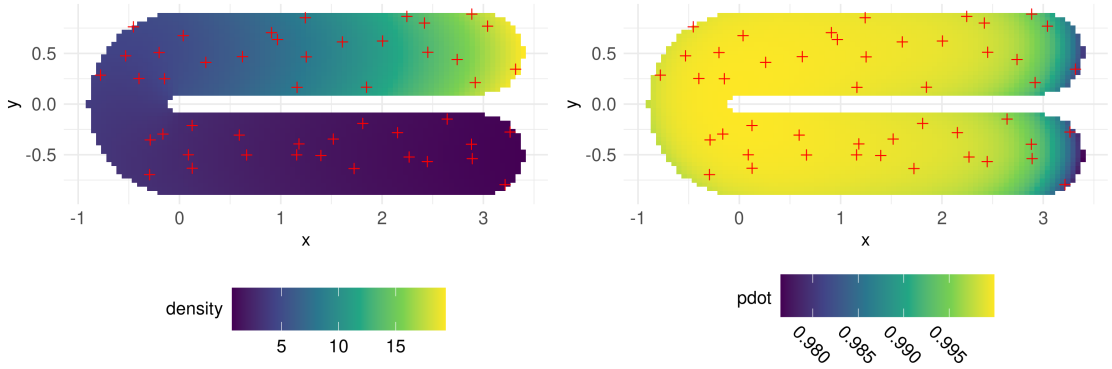


Figure 4.13: A) Detector locations and the true intensity surface B) The true probability of detection in the domain

The non-Euclidean distance was taken to be the shortest path distance as calculated using the R package `gdistance` (van Etten, 2017). Following the results of the simulation study in Miller and Wood (2014) we set $p = 3$, using a three di-

mensional Euclidean space to approximate the two dimensional horseshoe domain, which was the dimension selected by lowest AIC in their papers. The mesh used had approximately 3000 discretisation locations defined on a regular grid.

A total of 50 simulations were run. For each simulation a point pattern of activity centres was sampled using the true intensity. Capture histories were then generated based on this point pattern using a fixed detection model for all simulations. For each set of capture histories two models were fitted: the GDS and the TPRS on x and y that ignores non-Euclidean distances. This led to a total of 100 models being fitted to 50 datasets which took approximately 2.5 days to run on a linux server. The expected number of individuals detected for each simulation is approximately 30. For each simulation the true abundance was recorded and compared against the point estimate. In all models the detection function used the shortest path distance. The detectors were defined to be count type. For this reason the detection function was chosen to be the hazard half-normal as this is a natural way to model count responses in SCR models (see Section 4.3). This has two parameters to estimate: λ_0 a positive encounter rate parameter, and σ, a positive scale parameter. Both parameters were modelled using a log link.

The performance of the two types of smoothers is summarised in Figure 4.14. The GDS smoother outperforms the TPRS according to integrated squared error of the point estimate of intensity (Figure 4.14A), although both methods seem to have outliers with substantially higher integrated squared error. Both models show a slight negative bias in estimating abundance although it is small as a percentage of the total population size. The GDS smoother sometimes had some larger negative bias in abundance estimates (Figure 4.14B). The reason for the negative bias is not certain, although it is likely related to the slight positive bias in estimating the encounter rate λ_0 (Figure 4.14C). Both the GDS and TPRS model perform similarly here with one notable outlier for both.

The two models also show similar performance in estimating σ (Figure 4.14D).

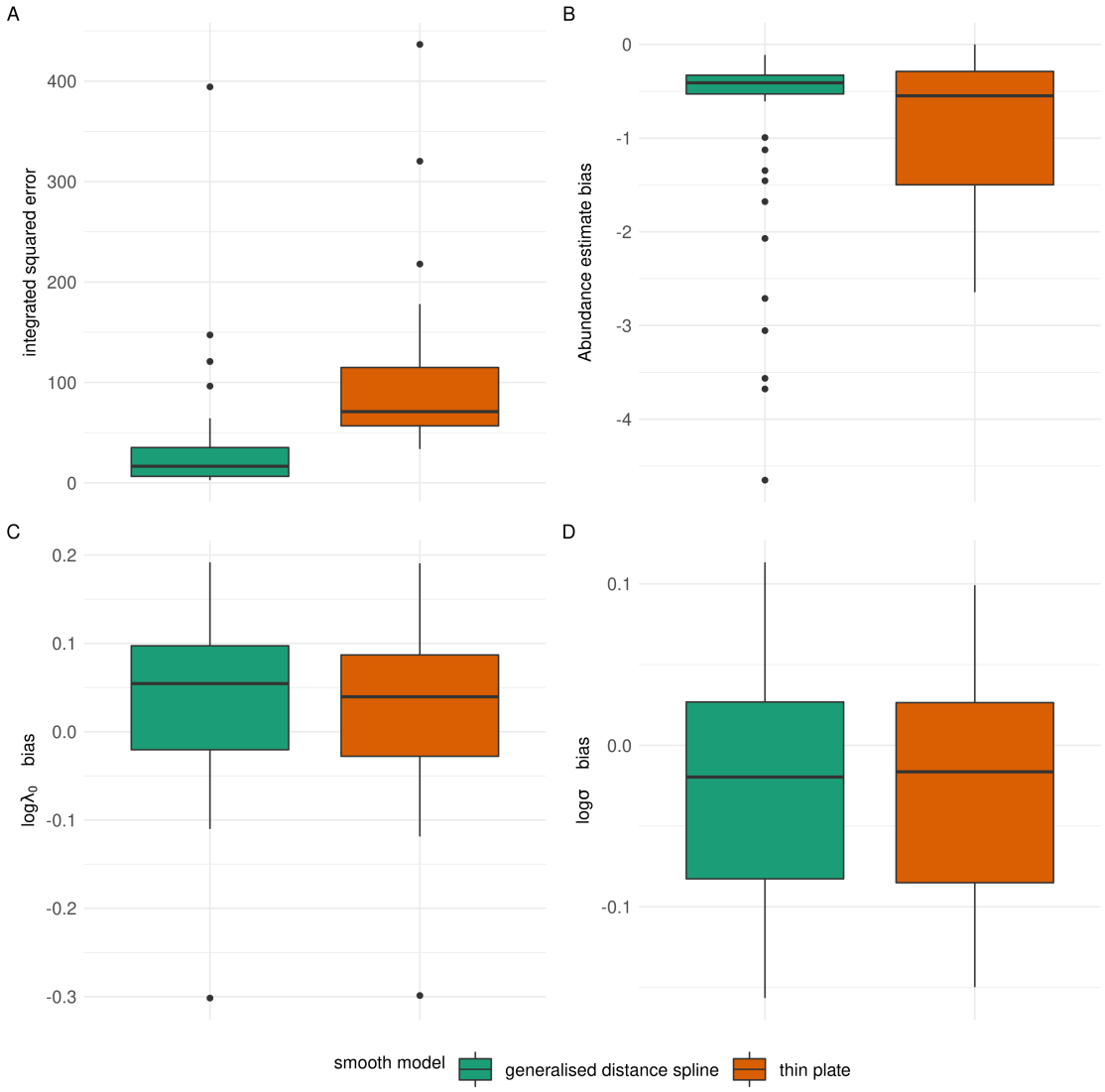


Figure 4.14: A) Comparison of integrated squared error B) Bias in abundance estimate C) Bias in $\log \lambda_0$ D) Bias in $\log \sigma$

All in all the results show that the model does fairly well at estimating abundance and detection parameters but that there can be some simulated data sets for which the model does poorly when compared to the average across all simulations. The GDS does a better job than the TPRS in estimating the true intensity surface as measured by integrated squared error (Figure 4.14A). This is confirmed in Figure 4.15 which compares the average prediction across all simulations for the two smoothing methods. The GDS smoother clearly does a better job on average of

estimating the intensity although, as indicated in the other plots, for some datasets both methods can do a poor job. It is not clear the reason for this but, in part, it may be due to poor identifiability of the smoothing parameters.

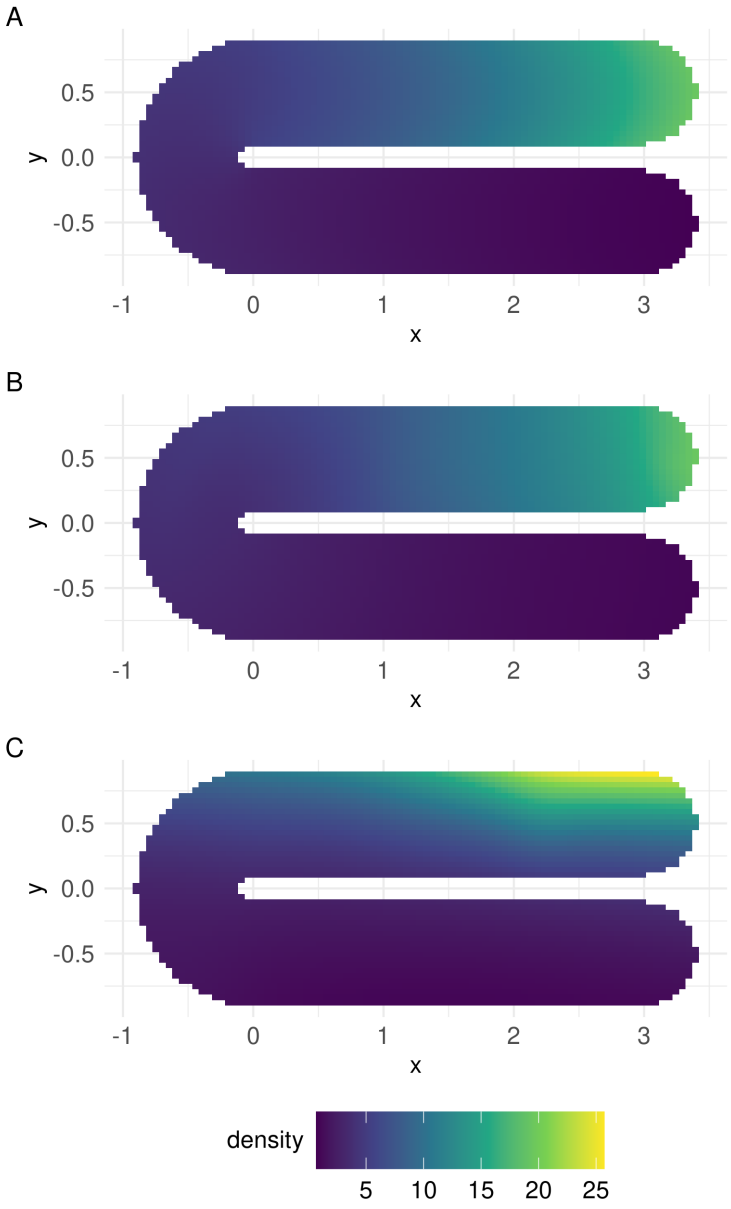


Figure 4.15: A) True intensity surface B) GDS average predicted intensity across all simulations C) TPRS average predicted intensity across all simulations

Figure 4.16 shows scatter plots of the estimated values of the smoothing param-

eters. The smoothing parameters are τ_1, associated with the Duchon penalty and τ_2, the additional smoothing parameter to penalise the null space of the Duchon penalty. Plots are shown on the log scale. The plot for τ_1 shows that the maximum likelihood estimate was drawn to two very different regions of the parameter space, one centred around zero and another around fifteen. Note that fifteen on the log scale implies that the shrinkage was very strong on the spline coefficients, shrinking them essentially to zero. For the GDS this is consistent with the fact that the true function is a linear gradient that increases "along the horseshoe" and so in the higher dimensional Euclidean space the linear basis function is able to fit well to this when projected back to 2 dimensions. However, for the TPRS it does not have the extra

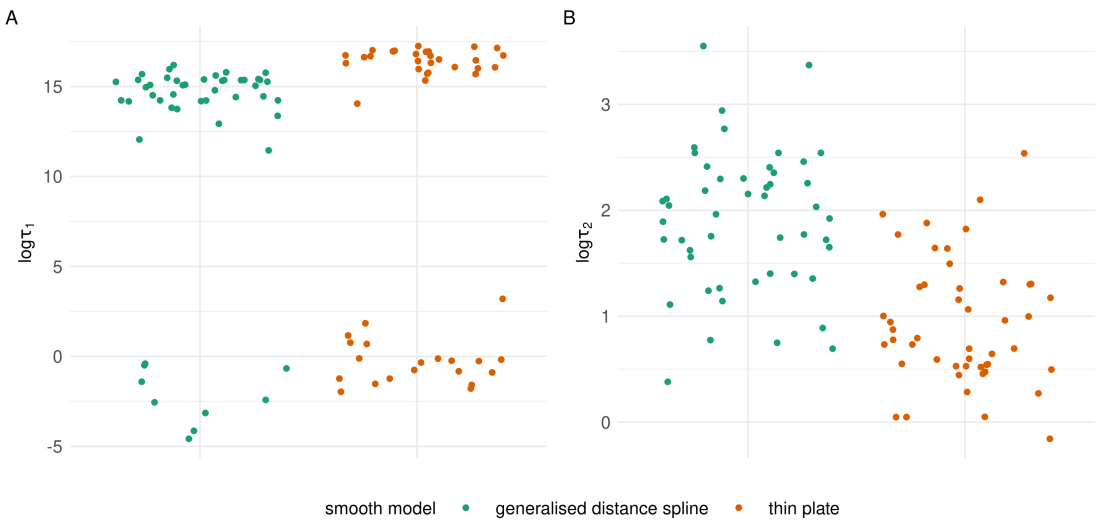


Figure 4.16: A) estimates for smoothing parameter $\log \tau_1$ B) estimates for null-space smoothing parameter $\log \tau_2$.

flexibility which is perhaps the reason for the incorrect north-south gradient in the average prediction in Figure 4.15C.

One possible reason that in some simulations the optimiser explored a different region of parameter space for τ_1 is that zero was chosen as the starting value for this parameter and the model was stuck in a local minimum. A possible approach to avoid this would be to fit the same smooth model to data collected at detectors

(such as the encounter rate) to achieve an approximate estimate of smoothness and to fix this value in the full SCR model. This may be insufficient however since data collected at detectors is the combination of two processes, the point process and the detection process, and so may not provide a suitable guideline for smoothness in the point process intensity.

In summary, it is possible to fit smoothing splines within SCR models that respect the non-Euclidean distances between mesh node locations. In this simple domain the model performed better than a naive alternative in the TPRS spline. There are indications in the poorer quality estimates that the likelihood is sensitive to specific qualities of the datasets generated or that the inference method can struggle to fully explore the parameter space. However, on average, across all simulated datasets, the GDS model performed better.

4.8.2 Case study: Boland leopard data

We fit the GDS model to data on leopards collected in the Boland area in the Western Cape province of South Africa. In this region there is a complex mix of available habitat and protected areas. For the purposes of this analysis we ignore the available covariate information other than the shape of the protected area. GPS tagging data indicates that individuals rarely leave this boundary (Figure 4.18). The complex shape of the protected area and in particular an area of unsuitable habitat in the in the southern region suggests that standard Euclidean distance may be inappropriate. We refer to this region as the 'hole' in the domain, see Figure 4.18. We used the shortest path distance as defined on a raster discretisation of the study region. This was used both in setting up the penalty matrix for the GDS and within the detection function.

The trapping array consisted of 92 camera traps deployed for 260 occasions. A total of 43 individuals were detected during the course of the survey. Figure 4.17 shows a summary of the capture histories. Figure 4.17A shows that there is some

spatial structure in the number of individuals detected and that the camera trap in the far north detected the most individuals. Figure 4.17B shows the total number of detections at each detector and shows a slightly different spatial pattern with more detections in the region north of the hole. The model was fitted with $p = 5$, using a 5 dimensional Euclidean space to approximate the non-Euclidean distances in the 2 dimensional domain and an extra penalty was used to penalise the null space.

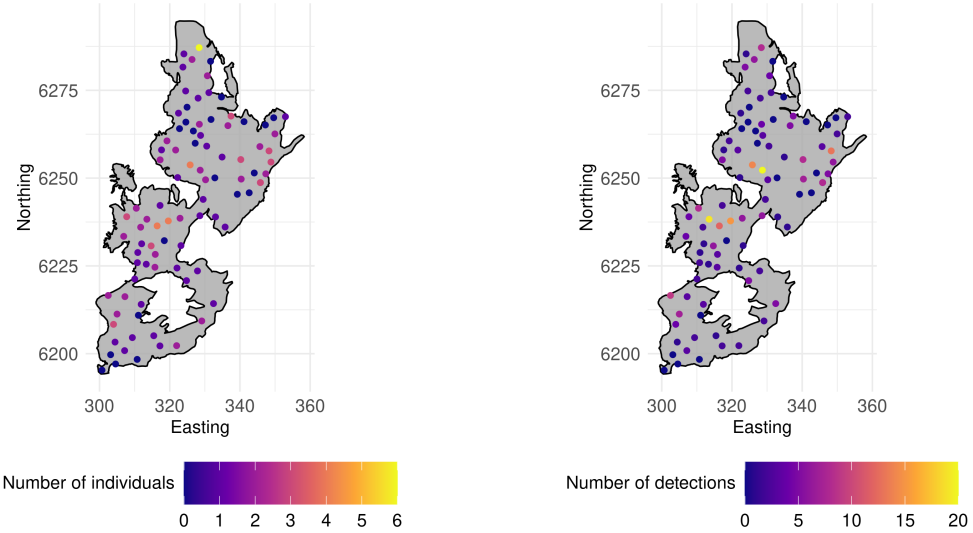


Figure 4.17: A) Detector locations and the number of unique individuals detected at each detector. B) Detector locations and the total number of detections (including recaptures) at each detector.

Figure 4.19 summarises the results of the model predictions. We can see that the null space functions seem to influence the mean prediction (Figure 4.19A) again, with the model predicting the highest density in the far north and lowest in the far south. Comparing this to the number of individuals detected on each trap it looks like this might be influenced by the detector in the north that detected the most individuals. The middle region just north of the hole also shows some of the spatial structure we might expect.

Figure 4.19B shows that detectability was estimated to be quite high in all regions of the domain. Figure 4.19C shows the median prediction based on 500 Monte Carlo samples using the Hessian matrix at the maximum likelihood estimate

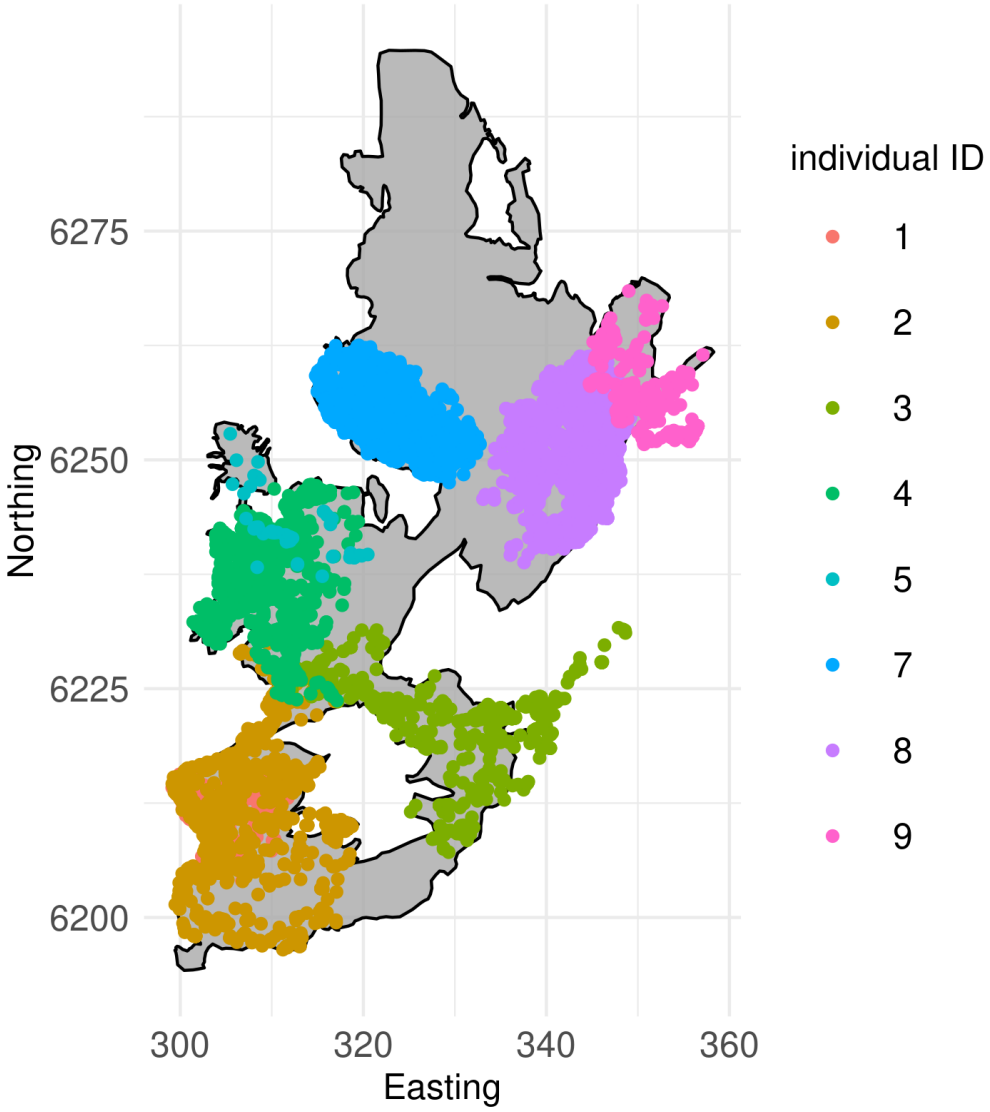


Figure 4.18: A) GPS tag locations of 9 individuals in the study region in 2012-14.

that is also returned by `TMB`. This seems very similar to the mean prediction in Figure 4.19A. Figure 4.19D plots the interquantile range which is defined to be the difference between the 0.975 and 0.025 quantiles, estimated using the same Monte Carlo samples as used to calculate the median. This shows there is greatest uncertainty in the regions that have the lowest probability of detection and also high uncertainty in the north where small changes in the gradient of the linear basis function can lead to large differences in the value the basis function takes.

To compare with a non-Euclidean smoother we also fitted a TPRS on density.

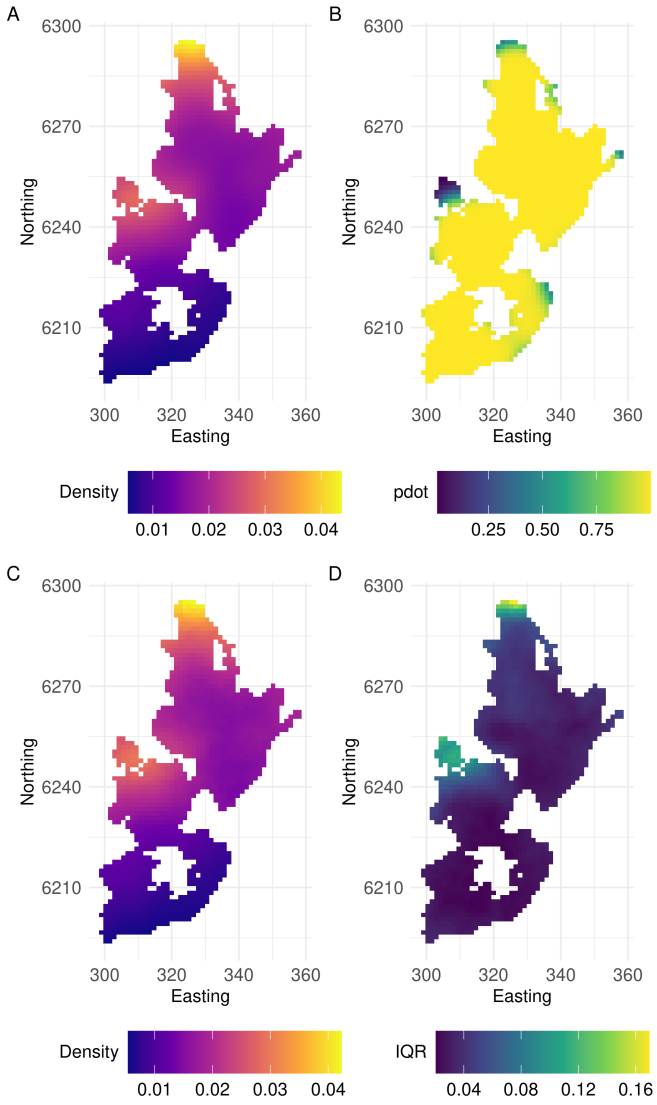


Figure 4.19: Summary of the GDS model predictions A) Expected density point estimate B) Probability of detection C) Median density based on 500 Monte Carlo samples D) Interquantile range (difference between 0.975-th and 0.025-th quantiles)

All other components of the model remained the same, including the non-Euclidean distances in the detection model. The mean prediction is shown in Figure 4.20A. Similar to the GDS, the null space basis is playing a predominant role in the predictions, fitting a clear north-south linear gradient to the data. However, it is less well able to capture the higher density to the north of the hole and there is evidence that the model has inappropriately smoothed across a gap in the study region domain. Figure 4.20B shows the difference between the TPRS prediction and the

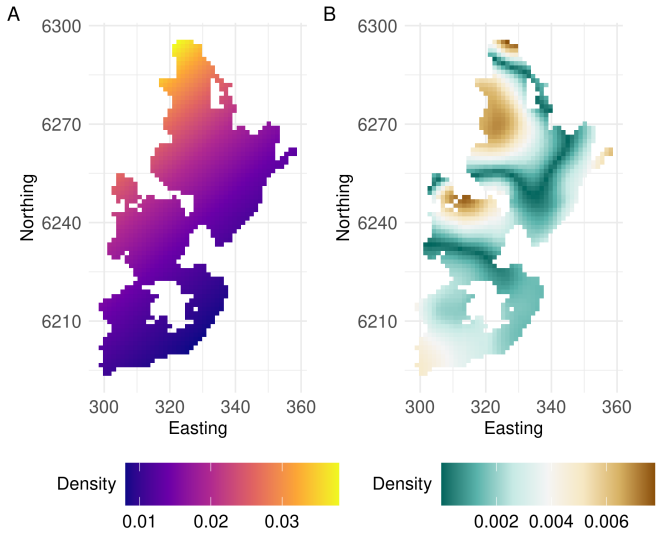


Figure 4.20: A) Mean prediction for TPRS model B) Difference between TPRS mean and GDS mean predictions

GDS prediction. This highlights that area to the north of the hole as an area where the predictions differ. There is also disagreement in the far north. The GDS seems to more closely match expectations from plotting the capture history data but both approaches still have problematic behaviour with the linear null space basis function, fitting to a north-south gradient that seems questionable given the summaries of the data (Figure 4.17). Figure 4.21 shows a boxplot of the 500 Monte Carlo samples of the two smoothing parameters. This plot shows that the maximum likelihood estimate for the null space parameter is quite large and is essentially not penalising the null space at all. The variability in this parameter is also very large. Although the variance on the log scale is comparable to the other smoothing parameter after taking the exponential this leads to much greater variance due to the difference in scale. This suggests the likelihood is not particularly informative for identifying the null space smoothing parameter.

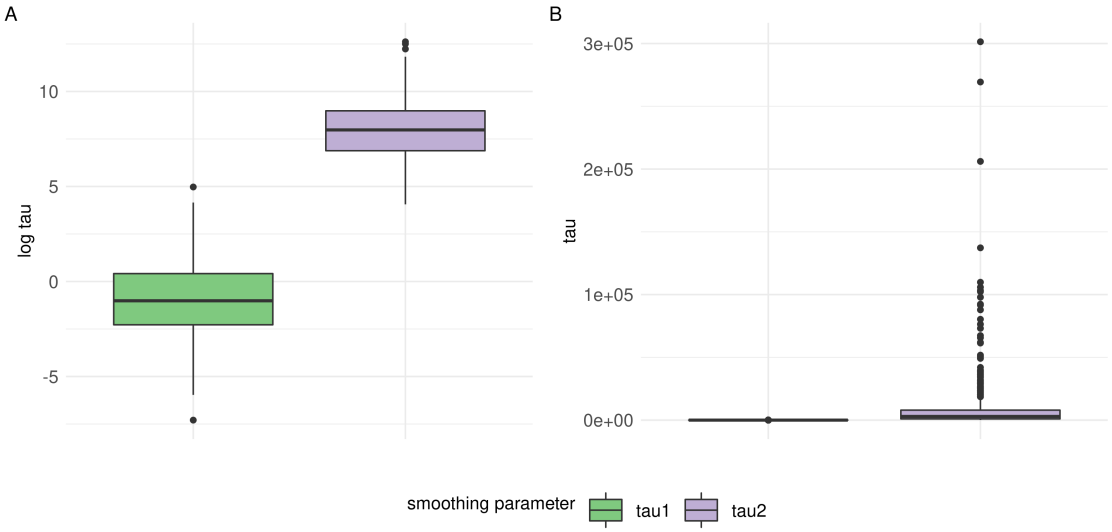


Figure 4.21: A) Smoothing parameter Monte Carlo samples on the log scale B) Smoothing parameter Monte Carlo samples on the smoothing parameter scale

4.9 Discussion

In this discussion we survey the process of fitting smoothing splines on density in SCR taken in this chapter. The aim is to situate the particular modelling decisions taken here amongst other possible options and provide pointers to important things to consider. We break this down into sections whether or not to use random effects on density, the types of random effects that could be used and their prediction behaviour in extrapolating to unobserved regions, and possible alternative approaches to inference.

We have, for the first time, demonstrated the feasibility of using penalised regression splines on density in SCR models but perhaps the main research value in doing this is to highlight the decisions that need to be taken and challenges that need to be addressed. The results from the case studies and simulation studies in this chapter are mixed. There seems to be some potential for this approach but there are many possible choices to make and many accompanying potential pitfalls. There is no guarantee that the predictions from such models will be robust or ecologically sensible. The purpose of this discussion is to provide a kind of roadmap,

highlighting the path taken here and possible alternative routes to take. There are many possibilities for future research that we also summarise here.

Is it worth the trouble to fit a spline on density?

An important question to ask is whether it is worth the additional complexity in model structure and inference procedure, and the associated computational costs, of considering a spatially structured random effect on density within an SCR model. We suggested in the introduction to this chapter that these types of model components are a standard part of the spatial statistician's toolkit and should therefore be available to use within SCR. There are, however, some particular features of SCR datasets that mean it may be of limited use in specific contexts. The first issue to consider is sample size. Since SCR is a popular method to survey hard-to-detect species the sample sizes are often much lower than in other point process applications in ecology. We tried adding TPRS effects on density to datasets with as low as 14 individuals and found no evidence that the smooth effect was preferred over a homogeneous model. Unless the clustering is strong and activity centre locations identified to a high degree of accuracy, then it is unlikely there will be much information in the dataset to inform a smoothing spline on density. There is also likely to be limited ability to identify anything other than very strong fixed effects.

Another issue with SCR data relates to sampling design. There is more likely to be spatial structure in the dataset if survey effort is placed in areas of relatively low density as well as high density. Unfortunately this is not what many SCR surveys seek to do. Given the costs involved of purchasing, maintaining, deploying and recovering detectors it can seem, naturally, undesirable to place detectors where the ecologists believe no or very few detections will be made. The result of this is that density can be very similar in the surveyed region if the sampling only took place in a relatively homogeneous environment. Therefore one should consider both the sample size and the sampling design to get a feel for whether there is likely to

be spatial structure in the dataset. Of course plotting summaries of the capture histories as we did for the two case studies above will also give a good indication of this. A low sample size, with fewer than 20 detected individuals, say, would typically be viewed as insufficient to fit anything but relatively simple point process models. Add to that the additional uncertainty of the SCR observation process and it is clear that in many contexts there will not be enough information in the data to identify anything more complicated than a homogeneous Poisson process. Preferentially sampling high density will also mute any heterogenous signal in the data and may have worrying implications for extrapolation to unobserved regions since the model has no way to learn what conditions are associated with lower densities.

A third thing one should consider is which spatial process in the SCR model should be smoothed. A key feature of SCR models is that they combine two spatial processes which occur at different scales: the point process model (second-order selection in the parlance of Johnson (1980)) for the location of activity centres and the movement model around these (third-order selection). Care should be taken to think carefully about which process is most relevant to the questions at hand. It may be one or the other or both. It is not hard to imagine scenarios in which other parameters in the model may benefit from being spatially correlated. If, for example, the scale of movement parameter σ is modelled with a smoothing spline in space then this would imply that animals near each other have similar movement patterns. This seems ecologically reasonable since these movement patterns are at least partially the result of local conditions (which are spatially correlated).

A spline on this process could also model potentially complicated processes such as interactions with other individuals (e.g. density dependent movement). A population that is tightly packed in one region may have different movement patterns to another region where individuals have more room to roam freely. The same argument could apply to adding a smoothing spline to the encounter rate parameter. Stevenson et al. (2021) propose using Gaussian random fields on the detection model

and point out that spatial-correlation in detections is almost inevitable due to the movement patterns of individuals as they encounter detectors. An individual does not really have the ability to teleport around its activity centre.

One should therefore consider carefully which ecological process is the primary target of inference. If the main process of interest is the intensity process then considering a smoothing spline seems a prudent step to avoid problems of spurious significance when interpreting fixed effects and smoothing over discontinuities in predictor variables (if appropriate). There may be identifiability issues in fitting random effects on the encounter rate, movement scale and the intensity at the same time and future research could investigate this.

Which type of spline to use?

After deciding whether or not to use a smoothing spline at all, the next question is which type of smoothing spline to use? There is somewhat of a gap in the spatial ecology literature with regards to this question. The general approach seems to be to use something that one is familiar with in a software package that one already knows how to use. Since the model component is not intended as a mechanistic description of the ecological processes, the differences between smoothing splines may have minimal consequences in terms of the role the effect plays in the model. However, as we have seen here, it can have big consequences for whether inference is feasible or not, due to the differences in approximation methods. In this chapter one key aspect was to choose a low dimension option. This was because the likelihood could prove too complex for automatic differentiation, at least as currently implemented in `TMB`, and resulted in memory problems during function optimisation. The reason for this is that an SCR likelihood involves an integral for every detected individual. If this is approximated by a sum of a large number of integration locations, each of which involves a potentially large number of spline basis coefficients, the graph of the function can become too large to store in memory.

When implementing the SPDE simulation study it was not possible to use a mesh with more than a few hundred nodes without hitting a memory error. The triangulation used was much more coarse than in many applications that use the SPDE. A key result in Lindgren et al. (2011) is that the Gaussian Markov random field approximation converges (in probability) to the Gaussian random field defined by the SPDE as the resolution of the mesh increases. In other words, the more mesh nodes, the better the approximation. That being said, the practical consequences of the mesh are not with respect to the accurate approximation of a Matérn field but whether the chosen GMRF can provide a good fit to the data.

Using such a coarse mesh resulted in predictions where the triangular basis functions were clearly visible, a clear downside to communication of results since these triangles appear only as a modelling artefact. So care should be taken to consider whether the number of parameters is feasible to fit with `TMB` and if the number of mesh nodes is too great then perhaps a low dimension option would be better. This was the reason low-dimension approaches were taken in the case studies and the non-Euclidean smoothing application. It may be the case that a B-spline basis with a number of knots in the low hundreds is sufficient for the application at hand but this should be checked. Even if it is possible to fit using `TMB` the actual fitting time may be much longer than a low dimensional alternative that gives qualitatively similar results. Sparse smoothers are another approach to gaining computational efficiency, one example of which the SPDE approach. However there may be only modest benefits to using sparse effects within SCR since the likelihood involves integrating over the location of each point and the model has a non-sparse dependence between the likelihoods for each individual's capture history. We discuss this more below when considering alternatives to automatic differentiation for inference.

Another aspect to consider is the geometry of the area the animals can use. One should consider whether non-Euclidean distances are most appropriate and choose a smoothing method that is appropriate to this. In this chapter we considered a low

rank option for doing this but there may be others. The more commonly used soap film smoother (Wood et al., 2008) and barrier model (Bakka et al., 2019) both use B-splines and were not used here for the same reason the SPDE was not used — there were too many parameters for `TMB` to handle.

Prediction behaviour

The simulation study for the TPRS (Section 4.6) revealed some important aspects of SCR survey design and smoothing spline prediction behaviour. The main cause of bias in the abundance estimates when extrapolating into unobserved regions of space was the linear null space basis function. This means that the model will predict exponentially increasing (assuming a log link) density in one particular direction and any extrapolation in that direction is likely to give biased intensity and abundance estimates. Uncertainty will also be greater in this direction since small changes in the gradient of the linear function will result in large changes in density. Correspondingly there will be ever decreasing uncertainty extrapolating in the opposite direction.

This behaviour was the reason for considering the two options for adding a penalty to the null space although the results of doing this were mixed. In the Boland leopards example one can clearly see a strong effect of the linear basis function that does not seem appropriate (Figure 4.17A). A nice advantage of using a mean zero GRF effect is that predictions have mean zero in un-sampled regions (using the term sample vaguely here to also include 'effectively sampled' implied by the smoothness of the latent process of interest). When using a low rank spline with global null space basis functions this is not the case. A prudent approach suggested by Figure 4.3 is to restrict predictions to areas above some threshold for probability of detection. In this case the likelihood of the observed data should, in theory at least, do something to constrain the global basis functions.

A related issue is the problem of ensuring an appropriate sampling design for extrapolating into unobserved regions. The problem of preferential sampling was

mentioned above in the context of dampening any spatial structure in the observed data but it is perhaps much more of a concern when it comes to extrapolation to unobserved regions. Intuitively, if you want to understand what leads to low density then you need to sample in low density regions. When choosing the true intensity surface in both simulations, random fields were sampled until there was substantial spatial variation that one could detect with the detector array, to give the random effects something to fit to. Thus this was doing preferential sampling in an indirect way, keeping the detector array fixed while simulating possible intensity functions. The result of this was positive bias in estimating abundance in both the TPRS and the SPDE methods since density was higher, on average, within the surveyed region as compared to the outside. When using random effects on density this should be considered carefully. Model based inference does not save us from important considerations of preferential and representative sampling.

Inference method

In this chapter, we chose to incorporate penalised random effects using a Laplace approximate maximum likelihood approach implemented using `TMB` with automatic differentiation to compute the required derivatives. This led us to the previously mentioned memory problems when considering random effects with more than a few hundred parameters. This presents a challenge for using high-resolution integration schemes with highly flexible random effects. In the models considered in this chapter, spatially structured random effects with a few hundred parameters and 30-60 individuals seemed to be about the limit, given available, reasonably capable, computing resources. This led to the use of low dimensional smoothers and the subsequent challenges involving penalising the null space and prediction behaviour.

One clear alternative would be to attempt to fit these models using MCMC. However, there was a startling lack of examples in the literature of even the relatively simple inhomogeneous Poisson process being fitted using MCMC methods. This is

a gap in the literature. It may be that current MCMC samplers used to fit SCR model perform poorly with estimating spatial structure in the point process. If the MCMC schemes struggle with fixed effects on density then they are likely to also struggle with random effects. The SCR literature seems to have a clear dividing line: if inhomogeneous density is called for then maximum likelihood seems to be the only option at the moment. Future research should seek to address this gap for MCMC methods. There could be benefits to exploring flexible density models using MCMC methods since these approaches often provide summaries of the model posterior that reveal important features of the likelihood such as identifiability issues and multi-modality that may be missed when applying maximum likelihood (and Laplace approximation) methods.

It could also be possible to develop a bespoke piece of software to implement the Laplace approximation for SCR models and avoid the memory problems encountered using automatic differentiation with `TMB`. It is possible to derive the first- and second-derivatives of the approximate likelihood (4.7) with respect to the spline basis parameters in order to implement the Laplace approximation directly without using automatic differentiation (see Appendix A). However, the derivatives of the Laplace approximation of the marginal likelihood seems mathematically intractable. It could, however, be possible to use gradient-free optimisation methods to maximise the Laplace approximate marginal likelihood. A pilot study using the Nelder-Mead method implemented in the `R` function `optim` did not converge but there are many alternatives. Details of the gradients used and the way this was implemented can be found in Appendix A.

One recent development is a new method for inference for *extended* latent Gaussian models, where more complicated dependence structures are allowed in the data likelihood component of the model (Stringer et al., 2021). This builds on the approach taken in INLA and extends the class of models that can be fitted. We mentioned above the problem of maintaining sparsity between model parameters in

the SCR likelihood given the shared parameters between each individual's capture history likelihood. This more complicated structure destroys the sparsity and also means the SCR likelihood is not possible to implement in `R-INLA` as it currently stands. Approximate methods for extended latent Gaussian models, as introduced in Stringer et al. (2021), may be a promising alternative to MCMC for Bayesian inference with SCR. The SCR likelihood appears to be within the class of extended latent Gaussian models and this could be a possible avenue of future research.

Conclusion

In summary, this chapter has, for the first time, introduced penalised random effects on the density of animals within an SCR model. This represents an extension to the types of point process models that can be considered within SCR and brings a valuable and commonly used tool in spatial statistics to this class of models. Using automatic differentiation for inference was a pragmatic choice that had the benefits of avoiding the need to derive mathematically challenging derivatives. However, this led to some computational challenges for random effects with many parameters as was the case with the SPDE approach and B-splines effects on space in general. To address this we made extensive use of low rank approximations for the thin plate spline and Duchon splines.

In order to use the Laplace approximation it was necessary to find ways to penalise the null space to create a full rank precision matrix. We covered two methods, the 'add a number to the diagonal' and the 'add an extra penalty' method and found the extra penalty to be more numerically stable in simulations. This was particularly the case for the GDS which allows for smoothing effects that respect complex spatial geometry. We ended with a survey of all the decisions taken in this chapter and possible alternatives. There remain many outstanding research questions in this area on topics from model choice, sample size, preferential sampling, choice of random effect, approximation method and how to achieve inference.

Chapter 5

Discussion

To conclude the thesis we summarise the key results of each chapter in turn and reflect on how this research could be taken forward. The overarching aim of developing methods in spatial ecology is to allow ecologists to address key questions in their field and this is reflected in the discussion here which is, in large part, motivated by applied problems. Each chapter has, in its own way, natural next steps to consider for future work, driven by the applied work of spatial ecologists as well as the purely statistical challenges involved.

A key theme of the thesis is the incorporation of spatially structured random effects into ecological models with complex observation processes. We hope this thesis has contributed to making spatially structured random effects a key part of a spatial ecologist's toolkit by expanding our understanding of how these methods work, presenting computationally efficient approaches for achieving inference within the context of complex observation processes, highlighting issues around communicating results from such models, and, in this discussion, presenting various avenues for future research.

5.1 Chapter 2: Understanding the stochastic partial differential equation approach to smoothing

This chapter presented an application of finite element methods to an SPDE whose stationary solutions have Matérn covariance. The key message of the chapter is that the SPDE, after choosing a basis representation, can be reformulated as a multivariate Gaussian prior model with a precision matrix that encodes the correlation structure implied by the SPDE. It is natural to then ask, could finite element methods also be applied to other SPDEs in statistical ecology? If so, it is possible that the computational efficiency that has made INLA so popular in applications of spatial statistics could be extended to a wider class of models. Below we describe a few promising areas where this approach could be beneficial and help to address a broad range of ecological questions.

5.1.1 Building mechanism into species distribution models

Species distribution models (SDMs) (Franklin, 2010a) seek to explain the spatio-temporal patterns of space use by species, populations and individuals. Distance sampling and SCR can be viewed as observation models that, at their heart, contain SDMs. One limitation of correlative species distribution models (SDMs) is that they only model correlations between predictors and the expected value of a response variable, but do not provide a scientifically plausible mechanistic model of the ecological processes which generated the observed data. Most SDMs are formulated as a GLM-type model, with a response variable that is assumed conditionally independent given a predictor. This predictor usually consists of an additive linear combination of model components, such as an intercept parameter, and fixed and random effects. No ecologist believes that this additive linear predictor is a math-

ematical description of an actual ecological process that occurs in nature. Rather, it is a convenient method to explore *relationships* between covariates and response variables. This may or may not allow for an interpretation that contributes to our understanding of ecological theory. Such an interpretation depends on many factors other than the model structure itself, such as survey or experimental design and specific contextual knowledge of the problem.

On the other hand, mechanistic models do not adopt a predictor expression as a convenient statistical framework to represent processes in spatial ecology, but instead attempt to use a direct mathematical description of the ecological processes of interest. Efforts in this area have tended to be agent-based models (Wilensky and Rand, 2015; McLane et al., 2011; Janssen and Ostrom, 2006) or based on differential equations (Busenberg, 2012; Holmes et al., 1994) and often achieving inverse inference from observed data can be challenging for such models.

There is, however, an exciting middle ground between these two approaches of correlative and mechanistic SDMs. Louvrier et al. (2020) combine the two approaches and incorporate a diffusion process, expressed as a differential equation, into a GLM-based framework to model the recolonisation process of wolves in eastern France. The model is based on a PDE that describes the inflow of new individuals to the region and the outflow (diffusion) to neighbouring regions in space. The parameters of the model have a clear interpretation: the rate at which new individuals arrive and the rate at which the population spreads across the landscape. This spatial process is also coupled with a logistic growth temporal process. This, similarly, is based on a differential equation model whose parameters have a clear interpretation. Louvrier et al. (2020) use the method of lines (Schiesser, 2012) to approximate the diffusion PDE and logistic growth ODE and arrive at a matrix-vector equation to approximate to the differential equations. Inference was by MCMC with the Gibbs algorithm using the `JAGS` package (Plummer, 2017).

Louvrier et al. (2020) provides a clear inspiration for similar work that could

be undertaken using finite element methods and potentially fitted with INLA or some other statistical software such as `TMB`. Instead of using the methods of lines to approximate the differential equations, one could apply finite element methods and derive the associated precision matrix for these mechanistic models. This precision matrix could incorporated into `R-INLA`, for example, by using the `rgeneric` framework that allows users to define custom model components (Chapter 11 of Gomez-Rubio (2020) provides an example of this for an intrinsic conditional autoregressive effect). This would allow access to many of the other useful features of `R-INLA` such as the ability to create more complex random effects by combining model components. For example, the parameters of the diffusion model for the recolonisation process of wolves in Louvrier et al. (2020) could also be modelled as a random effect, perhaps due to sex or season.

Contrast the above example of spatial diffusion and temporal logistic growth with, for example, a spatio-temporal random field with a separable space-time covariance function. One example would be a Matérn random field in space with an auto-regressive model through time. This is also a potential candidate model that could be suitable to investigate changes in the spatial distribution of wolves recolonising France, as was the aim of Louvrier et al. (2020). This model may be able to fit the data well (or it may not), but in either case it would not have the clear interpretability of the mechanistic model. The auto-regressive parameter tells us something about correlation through time but is not a direct replacement for a population growth model founded on ecological theory. Similarly, a spatio-temporal random field may capture something about the spatial diffusion of wolves across France, but any estimates of spatial rates of spread across the landscape would have to be derived post-hoc from posterior samples of the random field, and not included explicitly in the model. This is not necessarily a bad approach but the clear interpretation of a mechanistic model is appealing.

One criticism of mechanistic models in ecology is that the clear interpretation

can come at the cost of requiring a relatively simplistic model. Michelot et al. (2021) attempt to address criticism this by modelling the parameters of stochastic differential equations themselves using random effects such as smoothing splines. This seems to be a kind of middle ground between interpretability and flexibility. Michelot et al. (2021) use the Euler-Maruyama discretisation of a diffusion process and formulate the likelihood in such a way to make use of the Markov property of the diffusion process. They implement their approach in `TMB`. It is possible that finite element methods would also be useful here and the Markov property of the diffusion process is also a property of the Matérn SPDE that Lindgren et al. (2011) so successfully applied these methods to.

It seems that there may be a useful role that finite element methods can play to incorporate mechanism into SDMs. The perspective taken in this chapter, that SPDEs can be reduced to a prior (perhaps sparse) precision matrix, which can be used in many different software packages, could be useful for further research on statistical inference for flexible mechanistic models in spatial ecology. The advantage of building mechanism into GLM-type models is that the inference methods for GLMs are well established and there are many software packages and approaches that could be used to do this.

5.2 Chapter 3: point transect distance sampling using iterated integrated nested Laplace approximations

This chapter showed how to fit a one-stage distance sampling model in the R package `inlabru`, which makes use of a novel approach to inference using iterated integrated nested Laplace approximations. This extends the class of latent Gaussian models that can be fitted using INLA to a class of *conditional* latent Gaussian models, where

the model is an LGM conditional on particular values of some parameters. In the distance sampling case, the model could be fitted using standard INLA conditional on the parameters of the detection function.

This means, therefore, that `inlabru` can be used to fit models that standard INLA cannot. The class of conditional LGMs is not as rich as the models that could be specified in a full Bayesian probabilistic programming language such as `Rstan` (Stan Development Team, 2020) or `JAGS` (Plummer, 2017), but this additional flexibility does bring some new opportunities to fit complex models with non-linear parametric effects in spatial ecology. The benefits of doing this in `inlabru` is that these more complex models can now take advantage of the greater computational efficiency of INLA over MCMC. This section discusses some possibilities for using the iterated INLA method of inference in statistical ecology applications.

5.2.1 Functional responses

In the distance sampling application we used a parametric form of detection function to model an observation process. Similar parametric forms can also be used to model complex covariate effects, known as *functional responses*. A functional response captures the idea that the overall abundance of a resource or risk will affect animals' responses to it (Holling, 1959). For example, there may be diminishing returns to the value of a resource and so this could be modelled using a parametric equation that has an asymptote. Chapter 2 of Matthiopoulos et al. (2020) describes an example of a 'saturating response' with the effect of a covariate z modelled as

$$f(z) = \alpha(1 - \exp(-\gamma z^c)) \quad , \tag{5.1}$$

with parameters $\alpha, \gamma, c > 0$. This equation asymptotes at the value of α and the shape of the function is controlled by parameters γ and c. In a very simple simulation example, `inlabru` managed to recover this function from simulated count data

generated using (5.1) (see Appendix B).

There is a rich literature on functional responses with many possible effects designed to model a wide variety of response behaviour (Rosenbaum and Rall, 2018; Englund et al., 2011; Real, 1979; Oaten and Murdoch, 1975). There is considerable potential to use software like `inlabru` to develop a suite of more ecologically realistic model components to use within SDM models. This allows the predictor expression we use in these regression type models to capture more ecological realism. Importantly, fitting these models using the iterated INLA approach allows us to retain much of the computational efficiency that has made INLA so popular whilst also allowing more realistic and interpretable models.

Another example of a functional response is a multi-variable response that considers trade-offs between resources. For example, perhaps an animal prefers resource A over resource B, but if resource A is not available it will still use resource B. There is the potential to model this kind of interaction effect within a conditional LGM estimated using `inlabru`. This type of interaction between covariates has been addressed in the prey-choice literature (Smout et al., 2010) but we are not aware of its use in species distribution modelling in general. A similar idea is tackled in Matthiopoulos et al. (2011) where they formulate a varying coefficients model that uses the average conditions of all covariates to model the coefficients of each covariate individually.

5.2.2 Other detection models

This section considers observation processes other than distance sampling that could be implemented within `inlabru`.

The inhomogeneous Poisson process has become a key method for species distribution modelling using presence only data. Peel et al. (2019) and Fithian et al. (2015) both take this perspective and highlight that the key property of the Poisson process is that thinning the points means that presence only data can be viewed as a

realisation of an inhomogeneous Poisson process with intensity $\lambda(\boldsymbol{s})p(\boldsymbol{s})$. However, they do not model $p(\boldsymbol{s})$ as a probability which takes values between zero and one. Instead they use a log link, adding 'detectability covariates' to the linear predictor for $\log \lambda(\boldsymbol{s})$, and fitting the models using standard GLM software. However, this is means that the effect of the detectability covariates is not bounded between zero and one.

Using the iterated INLA approach implemented in `inlabru`, it is possible to add a non-linear model component, such as the logit of a linear combination of detectability covariates, within this predictor. This would bound the effect of the detectability covariates between zero and one. It is not clear the consequences of inappropriately assuming a log link for $p(\boldsymbol{s})$. In Peel et al. (2019) and Fithian et al. (2015) they simulate from the model that uses a log link so it is hard to say from these papers what the possible consequences are if the true data generating process actually modelled $p(\boldsymbol{s})$ as a probability. Future work could address this by simulating data using an appropriate thinning probability model and fitting both the log link and logit link models using `inlabru`. It seems likely that overall abundance estimates would be affected by incorrectly using a log link, although relative abundance estimates may be more robust.

Another common detection process in statistical ecology is *selectivity analysis*, which is used to model the probability that a fishing trawler net will capture a fish, dependent on the size of the fish. The larger fish are more likely to be trapped and the smaller fish more likely to escape through holes in the net.

Various parametric equations have been proposed to describe this type of relationship (Galbraith et al., 1994; Millar and Fryer, 1999), including a logistic linear model. Often the selectivity of specific types of gear is modelled separately and then these estimates are used to adjust field survey data. `inlabru` allows the user to define random effects on non-linear model components, for example using a random intercept term for each survey vessel and trawling gear type. There is also

support for implementing multiple likelihoods to allow the selectivity and spatial distribution of fish stocks to be estimated jointly.

These are just two examples of detection models that ecologists use on a regular basis and could be implemented in `inlabru` and there are likely to be many more.

5.2.3 Parametric spatio-temporal dependencies

One of the most popular features of INLA is the ability to fit spatio-temporally varying random fields with a wide variety of covariance structures. Whilst these model components are highly flexible and can model a number of ecological processes, they do not have a clear interpretation in terms of how they can contribute to ecological theory. There is the potential to explore defining model components with parametric forms of dependence that, conditionally, form a LGM. For example, logistic growth through time could be coupled with a spatial model. This could potentially lead to insights into ecological parameters of interest such as spatial variability in carrying capacity. A model like this could, in principle, be fitted using `JAGS` or `RStan`, but the appeal of `inlabru` is that we can make use of the speed of `R-INLA` whilst also extending the class of models we can consider. It may also be possible to use parametric spatial dependencies, such as those often used in compartmental models (Brauer, 2008), and specify these as conditional LGMs.

5.2.4 A word of warning

Whilst the ability of iterated INLA to allow the fitting of more complex models seems promising, future research is needed to understand how well this approximate inference procedure does in practice. The approximation error for LGMs fitted using INLA is minimal (Rue et al., 2009) but this has not been investigated for this wider class of conditional LGMs. More complex models will, in general, have a greater tendency to run into issues of identifiability and multi-modality in the posteriors which could make the Laplace approximation unsuitable. It is also possible that

iterated INLA could fail silently.

Therefore, caution is needed when developing new applications with the iterated INLA approach. A pragmatic strategy would be to investigate the behaviour of models using MCMC methods that allow a full exploration of the posterior distribution, which would reveal issues such as multi-modality that could cause problems for the Laplace approximation. Comparisons could then be made with the approximate posteriors generated by iterated INLA to judge whether the more computational efficient approximate approach is justifiable within the given context.

In summary, the iterated INLA approach has considerable potential to be used in a wide variety of models that are of interest to spatial ecologists. It provides an half-way house between the speed of INLA and the full flexibility of probabilistic programming languages like `JAGS` and `RStan`. Work in this are should be cautious and informed by simulation studies since the accuracy and appropriateness of using iterated INLA has not been fully investigated on the class of conditional LGMs. Challenges such as multi-modal posteriors could go undetected and as models become more complex this becomes more likely.

5.3 Chapter 4: Flexible density models for spatial capture-recapture

5.3.1 One dimensional splines

In keeping with the theme of this thesis, this chapter primarily investigated the use of random effects to model spatial autocorrelation. However, there are other components of the SCR model that could benefit from penalised random effects models.

A one-dimensional spline effect on a covariate to explain density would allow ecologists to consider more complicated functional responses to environmental con-

ditions within an SCR model. This offers greater flexibility than the usual log-linear fixed effect. These one-dimensional splines typically require fewer knots than the two-dimensional spatial effects and so may avoid some of the computational challenges encountered for random effects on density. Future work could investigate this through simulations and applied case studies, providing examples can be found of real applications where a non-linear effect of a covariate is justified.

5.3.2 Alternative point process models

There are, however, some limitations as to how far we can go with random effects on the intensity of a log-Gaussian Cox process that we used to model the distribution of activity centres. The log-Gaussian Cox process is a flexible model that can capture unexplained clustering in animal locations but it cannot, however, model repulsion between points. One alternative was addressed by Reich and Gardner (2014) who modelled repulsion between points for a territorial species using a Strauss process. There is a rich literature on point process models beyond the Poisson process and its variants (Baddeley et al., 2015; Illian et al., 2008) and the full breadth of these possibilities have not been considered within SCR. Some types of dependence between points are hard to incorporate into a log-Gaussian Cox process model, such as repulsion between points or self-exciting processes. For some SCR applications there may be some benefit to implementing other types of point processes that better fit the ecology. Often point processes that incorporate more complicated dependencies between points are fitted using pseudo-likelihood approaches (Baddeley and Turner, 2000). Future work could investigate whether `TMB` could be applied to these cases which could also open the door to combining these models with the random effects structures presented here.

In addition to new point process models there is also scope to consider point processes defined networks. For example, Sutherland et al. (2018) applied SCR methods to data collected on a river network. There is a growing literature on point

processes defined on networks which we briefly describe. Defining random effects for the intensity on a network requires consideration of how correlations propagate through a branching structure (Baddeley et al., 2017) and there are a variety of methods to handle this (Rakshit et al., 2019, 2017; McSwiggan et al., 2017). For a river network there is also a natural direction to correlation due to the flow of the river which needs to be accounted for (Santos-Fernandez et al., 2021; Rasmussen and Christensen, 2021). There are many potential options for implementing a random effect for an intensity function defined on a network. For those effects that result in constructing a precision (or smoothing) matrix, the approach taken in this chapter could be fruitful.

5.3.3 Random effects on other components in the SCR model

Of course, the SCR model is more than just the intensity function for the distribution of activity centres. Other components of the model could also benefit from flexible random effects models. Stevenson et al. (2021) point out that capture-recapture data will almost always have some un-modelled heterogeneity in detections because the assumption of conditional independence between detections is likely to be violated due the way in which animals move and encounter detectors. Stevenson et al. (2021) address this using a latent random field approach similar to that taken in this thesis.

The success of using random fields to model varying detection as well as density suggests that a general software package that allows the use of penalised random effects on all major parameters in the SCR model, the encounter rate, scale parameter and density surface, would be a powerful toolkit for practitioners to build rich and complex models for SCR data. In doing so they will be able to account for the complex spatio-temporal dependencies in the data by using computationally efficient and flexible random effects. This would represent a significant development in SCR applications, bringing valuable tools from the spatial statistics literature

into the SCR setting.

However, this additional complexity may come at a cost. When considering spatially structured random effects on all major parameters of the SCR model, the issue of identifiability has to be considered. The use of non-parametric effects to describe both the detectability of individuals and their spatial distribution seems likely to lead to confounding unless some simplifying assumptions could be made. There may be some model structures and, in a Bayesian setting, prior formulations, that avoid this problem. This should be an avenue of future research.

5.3.4 Alternative low rank methods

To conclude, we also highlight that there are many approaches to implementing spatially structured random effects other than those we considered here, which were essentially the low-rank random effects implemented in `mgcv`. This was a practical choice given our familiarity with the software and meant we were able to use `mgcv` to generate the required smoothing matrices and then use these within `TMB`. There are, however, other low-rank random effects that could be considered. One recent contribution is a low rank GRF approach based on a truncation of the spectral representation of the covariance function (Riutort-Mayol et al., 2020).

5.4 Conclusions

As the above discussion shows, there are many possible future avenues for research on the topics presented in this thesis. Statistical ecology requires advanced statistical methods to account for complex spatio-temporal dependencies in ecological processes and observation methods. This thesis has focused on implementing computationally efficient spatially structured random effects in distance sampling and SCR, providing valuable flexible modelling approaches that can model complex spatial dependencies in ecological data. In doing this, we also presented a novel perspective on the SPDE

approach, a technically challenging model, that we hope helps lead to a better understanding of this approach amongst applied quantitative ecologists. We also focused on the issue of communicating uncertainty in models that use such effects. This remains a rich area of research, with many open questions and opportunities for applied statisticians to have an impact of the field of spatial ecology.

We hope this thesis is an example of how to usefully contribute to statistical methods for spatial ecology through developing a better understanding of the methods we use, constructing appropriately complex models that can handle the unique challenges of observational ecological datasets, using computationally efficient inference methods that allow these models to be estimable in practice, and considering how to communicate the outputs from these models. It has been an honour to work on all of these aspects of applied statistical ecology and an experience from which I learned so many helpful things. I hope, if you've somehow made it this far to read through to the end of this thesis, that it was helpful to you too.

Appendices

Appendix A

Derivatives of the approximate SCR log likelihood

A.1 First- and second-order derivatives

Choose M integration locations $\boldsymbol{s} = (s_1, \ldots, s_M)^\intercal$ with weight α_j at location s_j. The approximate log-likelihood is

$$\log\{\pi(\Omega|\lambda)\} = -\tilde{\Lambda} + \sum_{i=1}^{n} \log\left[\sum_{j=1}^{M} \alpha_j \exp\left\{\sum_t \sum_k \log[g_{itk}(s_j)] + \log\lambda_j\right\}\right], \quad \text{(A.1)}$$

where

$$\tilde{\Lambda} = \sum_{j=1}^{M} \alpha_j \lambda_j p_{\cdot\cdot}(s_j) = \sum_{j=1}^{M} \alpha_j p_{\cdot\cdot}(s_j) e^{\log\lambda_j},$$

and $\lambda_j = \lambda(s_j)$. Writing $\tilde{\Lambda}$ like this is useful for differentiating with respect to $\log\lambda_j$ below. The log-intensity is

$$\log\lambda(s_j) = \beta + \sum_{r=1}^{R} \xi_r \Psi_r(s_j)$$

where β is an intercept parameter and there are R basis functions $\psi_1, \ldots, \psi_R$ for the random effect and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_R)^\intercal$ have precision matrix $\boldsymbol{Q}$. Let matrix $\boldsymbol{A}$ have (j, r)-th entry $A_{jr} = \Psi_r(s_j)$ then the intensity evaluated at each integration location is $\log \boldsymbol{\lambda} = \beta + \boldsymbol{A}\boldsymbol{\xi}$.

For simplicity I suppress all other parameter notation other than $\boldsymbol{\xi}$ in the likelihoods. The marginal likelihood, integrating out $\boldsymbol{\xi}$, is

$$\begin{aligned} \pi(\Omega) &= \int \cdots \int \pi(\Omega|\boldsymbol{\xi}) f_\xi(\boldsymbol{\xi})\, \partial\xi_1 \cdots \partial\xi_R \\ &= \int \cdots \int \exp\left\{\log\left[\pi(\Omega|\boldsymbol{\xi})\right] + \log f_\xi(\boldsymbol{\xi})\right\} \partial\xi_1 \cdots \partial\xi_R. \end{aligned}$$

Let $g(\boldsymbol{\xi}) = \log\left[\pi(\Omega|\boldsymbol{\xi})\right] + \log f_\xi(\boldsymbol{\xi})$. Then the Laplace approximation of the marginal likelihood is

$$\tilde{\pi}(\Omega) = \sup_{\boldsymbol{\xi}} \left\{\exp[g(\boldsymbol{\xi})]\right\} \frac{(2\pi)^{\frac{R}{2}}}{|\boldsymbol{H}|^{\frac{1}{2}}}, \tag{A.2}$$

where $\boldsymbol{H}$ is the matrix of second derivatives of $-g(\boldsymbol{\xi})$ with respect to $\boldsymbol{\xi}$ evaluated at the mode of $g(\boldsymbol{\xi})$.

Derivatives of $g(\boldsymbol{\xi})$

Let $h(\boldsymbol{\xi}) = \log\left[\pi(\Omega|\boldsymbol{\xi})\right]$ and so $g(\boldsymbol{\xi}) = h(\boldsymbol{\xi}) + \log f_\xi(\boldsymbol{\xi})$. The derivatives of $\log f_\xi(\boldsymbol{\xi})$ are straighforward since $f_{\boldsymbol{\xi}}$ is the pdf of a GMRF. The pdf of a GMRF with zero mean and precision matrix $\boldsymbol{Q}$ is proportional to $\exp(-\frac{1}{2}\boldsymbol{\xi}^\intercal \boldsymbol{Q} \boldsymbol{\xi})$. Therefore we have

$$\frac{\partial \log f_{\boldsymbol{\xi}}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = -\boldsymbol{Q}\boldsymbol{\xi} \tag{A.3}$$

$$\frac{\partial^2 \log f_{\boldsymbol{\xi}}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^2} = -\boldsymbol{Q}. \tag{A.4}$$

Computing derivatives of $h(\boldsymbol{\xi})$ requires repeated applications of the chain rule.

Since $h(\boldsymbol{\xi})$ is a function of $\log \lambda_j$ for $j = 1, ..., M$. Then by the chain rule:

$$\begin{aligned}\frac{\partial h(\boldsymbol{\xi})}{\partial \xi_r} &= \sum_j \frac{\partial h(\boldsymbol{\xi})}{\partial \log \lambda_j} \frac{\partial \log \lambda_j}{\partial \xi_r} \\ &= \sum_j \frac{\partial h(\boldsymbol{\xi})}{\partial \log \lambda_j} \Psi_r(s_j). \end{aligned} \tag{A.5}$$

The derivatives of $h(\boldsymbol{\xi})$ with respect to $\log \lambda_j$ are

$$\begin{aligned}\frac{\partial h(\boldsymbol{\xi})}{\partial \log \lambda_j} &= -\alpha_j \lambda_j p_{..}(s_j) + \sum_{i=1}^{n} \frac{\alpha_j \exp\left\{\sum_t \sum_k \log[g_{itk}(s_j)] + \log\left[\lambda(s_j)\right]\right\}}{\sum_{j^*=1}^{M} \alpha_{j^*} \exp\left\{\sum_t \sum_k \log[g_{itk}(s_{j^*})] + \log\left[\lambda(s_{j^*})\right]\right\}} \\ &= -\alpha_j \lambda_j p_{..}(s_j) + \sum_{i=1}^{n} \frac{\alpha_j \exp\left\{\sum_t \sum_k \log[g_{itk}(s_j)] + \log \lambda_j\right\}}{\sum_{j^*=1}^{M} \alpha_{j^*} \exp\left\{\sum_t \sum_k \log[g_{itk}(s_{j^*})] + \log[\lambda_{j^*}]\right\}} \\ &= -\alpha_j \lambda_j p_{..}(s_j) + \sum_{i=1}^{n} \frac{\alpha_j \lambda_j \prod_t \prod_k g_{itk}(s_j)}{\sum_{j^*=1}^{M} \alpha_{j^*} \lambda_{j^*} \prod_t \prod_k g_{itk}(s_{j^*})} \\ &= -\alpha_j \lambda_j p_{..}(s_j) + \alpha_j \lambda_j \left\{ \sum_{i=1}^{n} \frac{\prod_t \prod_k g_{itk}(s_j)}{\sum_{j^*=1}^{M} \alpha_{j^*} \lambda_{j^*} \prod_t \prod_k g_{itk}(s_{j^*})} \right\}. \end{aligned} \tag{A.6}$$

To obtain the second derivatives apply the chain rule again

$$\begin{aligned}\frac{\partial^2 h(\boldsymbol{\xi})}{\partial \xi_r \xi_v} &= \sum_{j_2} \frac{\partial^2 h(\boldsymbol{\xi})}{\partial \log \lambda_{j_2} \partial \xi_r} \Psi_v(s_{j_2}) \\ &= \sum_{j_2} \sum_{j_1} \frac{\partial^2 h(\boldsymbol{\xi})}{\partial \log \lambda_{j_2} \partial \log \lambda_{j_1}} \Psi_r(s_{j_1}) \Psi_v(s_{j_2}), \end{aligned} \tag{A.7}$$

where j_2 is the index associated with applying the chain rule to obtain the second derivatives and j_1 similarly is associated with computing the first derivatives of h. Therefore we need to compute $\partial^2 h(\boldsymbol{\xi}) / \partial \log \lambda_{j_2} \partial \log \lambda_{j_1}$ for all combinations of j_1 and j_2. For the diagonal elements of the Hessian, when $j_2 = j_1$, we have

$$\begin{aligned}\frac{\partial^2 h(\boldsymbol{\xi})}{\partial \log \lambda_j^2} &= -\alpha_j \lambda_j p_{..}(s_j) + \alpha_j \lambda_j \left\{ \sum_{i=1}^{n} \frac{\prod_t \prod_k g_{itk}(s_j)}{\sum_{j^*=1}^{M} \alpha_{j^*} \lambda_{j^*} \prod_t \prod_k g_{itk}(s_{j^*})} \right\} \\ &\quad + \alpha_j^2 \lambda_j \sum_{i=1}^{n} \frac{\left[\prod_t \prod_k g_{itk}(s_j)\right]^2}{\left[\sum_{j^*=1}^{M} \alpha_{j^*} \lambda_{j^*} \prod_t \prod_k g_{itk}(s_{j^*})\right]^2}. \end{aligned} \tag{A.8}$$

For the off-diagonal elements, when $j_2 \neq j_1$, we have

$$\frac{\partial^2 h(\boldsymbol{\xi})}{\partial \log \lambda_{j_2} \partial \log \lambda_{j_1}} = \alpha_{j_1}\alpha_{j_2}\lambda_{j_1}\lambda_{j_2} \sum_{i=1}^{n} \frac{\{\prod_t \prod_k g_{itk}(s_{j_1})\}\{\prod_t \prod_k g_{itk}(s_{j_2})\}}{\left[\sum_{j^*=1}^{M} \alpha_{j^*}\lambda_{j^*} \prod_t \prod_k g_{itk}(s_{j^*})\right]^2} \quad \text{(A.9)}.$$

The above expressions can be simplified using matrix notation. Let $\boldsymbol{A}_r = [\Psi_r(s_1) \dots \Psi_r(s_M)]^\intercal$ and let $\boldsymbol{H}_{\log\lambda}$ be the matrix whose (j,k)-th element is $\partial^2 h(\boldsymbol{\xi})/\partial \log[\lambda_j] \partial \log[\lambda_k]$. Then

$$\begin{aligned} \frac{\partial^2 h(\boldsymbol{\xi})}{\partial \xi_r \xi_v} &= \sum_{j_2}\sum_{j_1} \frac{\partial^2 h(\boldsymbol{\xi})}{\partial \log[\lambda_{j_2}] \partial \log[\lambda_{j_1}]} \Psi_r(s_{j_1}) \Psi_v(s_{j_2}) \\ &= \boldsymbol{A}_r^\intercal \boldsymbol{H}_{\log\lambda} \boldsymbol{A}_v. \end{aligned}$$

Then we can write the Hessian of $h(\boldsymbol{\xi})$, denoted $\boldsymbol{H}_h$, as

$$\boldsymbol{H}_h = \boldsymbol{A}^\intercal \boldsymbol{H}_{\log\lambda} \boldsymbol{A}. \quad \text{(A.10)}$$

Then the Hessian of $g(\boldsymbol{\xi})$ is $\boldsymbol{H} = \boldsymbol{H}_h - \boldsymbol{Q}$. Note that when the integration locations are the same as the knot locations for the random effect (as was used in the SPDE implementation) then $\boldsymbol{A}$ is the identity matrix.

A.2 A bespoke Laplace approximation implementation

This section provides one possible way to implement the above derivatives in computer code by seeking to re-write the equations given above in a more concise way.

Constructing $\boldsymbol{H}_{\log\lambda}$

Seeking to simplify equations (A.8) and (A.9), which define the elements of $\boldsymbol{H}_{\log\lambda}$, let $\boldsymbol{G}$ be the matrix with

$$G_{ij} = \alpha_j \lambda_j \prod_t \prod_k g_{itk}(s_j).$$

The row sums of $\boldsymbol{G}$ give the denominator indexed by j^* in (A.8) and (A.9). Denote the vector $\boldsymbol{G^*} = \boldsymbol{G}\mathbf{1}_M$. Then the i-th element of $\boldsymbol{G^*}$ is $G_i^* = \sum_{j^*=1}^{M} \alpha_{j^*}\lambda_{j^*} \prod_t \prod_k g_{itk}(s_{j^*})$. Then (A.9) becomes (letting $j_1 = j$, $j_2 = k$)

$$\begin{aligned} \frac{\partial^2 h(\boldsymbol{\xi})}{\partial \log[\lambda_j] \partial \log[\lambda_k]} &= \alpha_j \alpha_k \lambda_j \lambda_k \sum_{i=1}^{n} \frac{\{\prod_t \prod_k g_{itk}(s_j)\}\{\prod_t \prod_k g_{itk}(s_k)\}}{\left[\sum_{j^*=1}^{M} \alpha_{j^*}\lambda_{j^*} \prod_t \prod_k g_{itk}(s_{j^*})\right]^2} \\ &= \sum_{i=1}^{n} \frac{G_{ij}G_{ik}}{(G_i^*)^2}, \end{aligned} \tag{A.11}$$

and (A.8) can be written as:

$$\frac{\partial^2 h(\boldsymbol{\xi})}{\partial \log \lambda_j^2} = -\alpha_j \lambda_j p_{\cdot\cdot}(s_j) + \sum_{i=1}^{n} \frac{G_{ij}}{G_i^*} + \sum_{i=1}^{n} \frac{\alpha_j (G_{ij})^2}{(G_i^*)^2}. \tag{A.12}$$

So given vectors $\boldsymbol{\alpha}$, $\boldsymbol{\lambda}$, $\boldsymbol{p}_{\cdot\cdot}$, $\boldsymbol{G}^*$, and the matrix $\boldsymbol{G}$ (all of which are relatively straightforward to construct) then we can use equations (A.11) and (A.12) to create $\boldsymbol{H}_{\log\lambda}$ (Note: In `R` I did this using loops as I couldn't see any other way but in a compiled language this wouldn't matter). To create $\boldsymbol{G}$, first create $\boldsymbol{B}$ with (i,j)-th element given by

$$\begin{aligned} b_{ij} &= \prod_t \prod_k g_{itk}(s_j) \\ &= \exp\left[\sum_t \sum_k \log g_{itk}(s_j)\right], \end{aligned}$$

and, letting $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{\alpha} \odot \boldsymbol{\lambda})$, where $\odot$ denotes the element-wise product, then

$\boldsymbol{G} = \boldsymbol{BD}$. The entries of $\boldsymbol{B}$ have to be calculated to evaluate the likelihood anyway to this comes at minimal extra cost.

The gradient of $h(\boldsymbol{\xi})$ with respect to $\log \boldsymbol{\lambda}$, given for a single location s_j in equation (A.6), is then

$$\frac{\partial h(\boldsymbol{\xi})}{\partial \log \boldsymbol{\lambda}} = \boldsymbol{\alpha} \odot \boldsymbol{\lambda} \odot \left(-\boldsymbol{p}_{\cdot\cdot} + \sum_{i=1}^{n} \frac{B_i}{\boldsymbol{G}_i^*} \right), \tag{A.13}$$

where B_i is the i-th row of the matrix $\boldsymbol{B}$ and $\boldsymbol{G}_i^*$ is the i-th element of vector $\boldsymbol{G}^*$. The derivative of $h(\boldsymbol{\xi})$ with respect to $\boldsymbol{\xi}$ is then

$$\frac{\partial h(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \boldsymbol{A}^\intercal \frac{\partial h(\boldsymbol{\xi})}{\partial \log \boldsymbol{\lambda}}. \tag{A.14}$$

Therefore the gradient of $g(\boldsymbol{\xi})$ with respect to $\boldsymbol{\xi}$ is

$$\frac{\partial g(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \boldsymbol{A}^\intercal \frac{\partial h(\boldsymbol{\xi})}{\partial \log \boldsymbol{\lambda}} - \boldsymbol{Q}\boldsymbol{\xi}. \tag{A.15}$$

This then provides all the building blocks necessary to estimate the mode of $g(\boldsymbol{\xi})$ and construct its Hessian in order to compute the Laplace approximate marginal likelihood (A.2).

Appendix B

Saturating functional responses using `inlabru`

This appendix presents a proof of concept for the use of non-linear parametric effects in species distribution models fitted using `inlabru`.

For the purposes of demonstrating the idea, we choose one specific effect, known as a *saturating response*. This is an effect that has diminishing returns as the covariate value increases. It could, for example, represent a food source that once available in sufficient quantities, provides no additional benefit to a species. We consider an effect of the form

$$f(z) = \alpha(1 - \exp(-\gamma z^c)), \tag{B.1}$$

for covariate z with parameters $\alpha, \gamma, c > 0$. This equation asymptotes at the value of α and the shape of the function is controlled by parameters γ and c. This is clearly non-linear in the parameters α, γ and c. For simplicity we assume c is a fixed constant.

We generated a covariate effect (Figure B.1A) and then calculated the saturating effect of this covariate (Figure B.1B) with $\alpha = 5$ and $\gamma = 0.6$ and c assumed fixed at $c = 1$. We used a log link and added an intercept parameter to generate Poisson

rate parameters (Figure B.1C) from which we generated an observed set of counts (Figure B.1D). The model was estimated using `bru(..., family = "Poisson")` and the fitting completed in seconds.

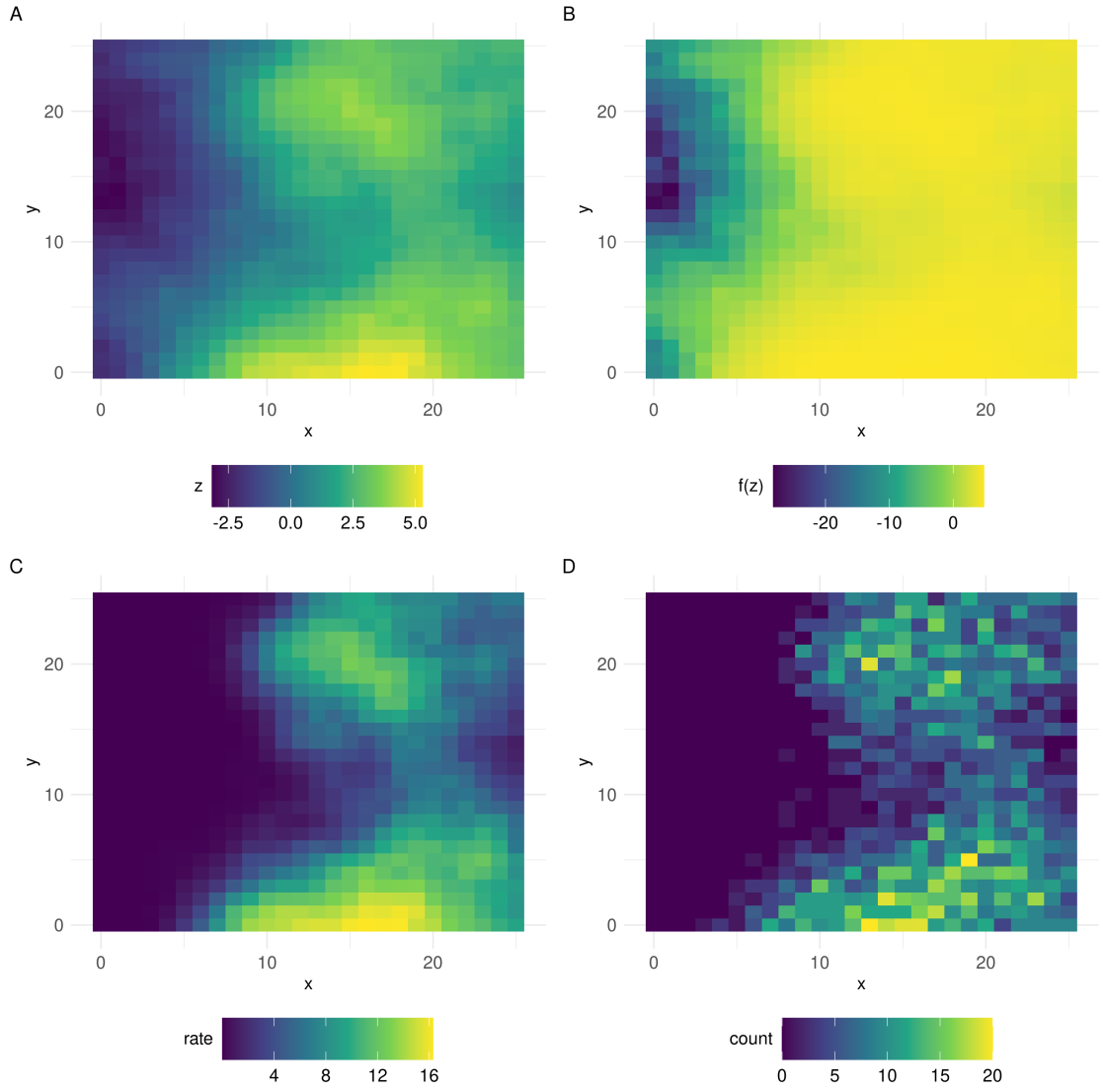


Figure B.1: **A**: A simulated covariate; **B**: The saturating effect of the simulated covariate; **C**: The Poisson rates that depend on this saturating effect (using a log link); **D**: Simulated counts.

Figure B.2 summarises the posterior effect of z estimated by the model and shows good agreement with the true effect used to simulate the data (blue line). This shows that non-linear functional responses are, in principle, estimable using the iterated INLA approach implemented in `inlabru`.

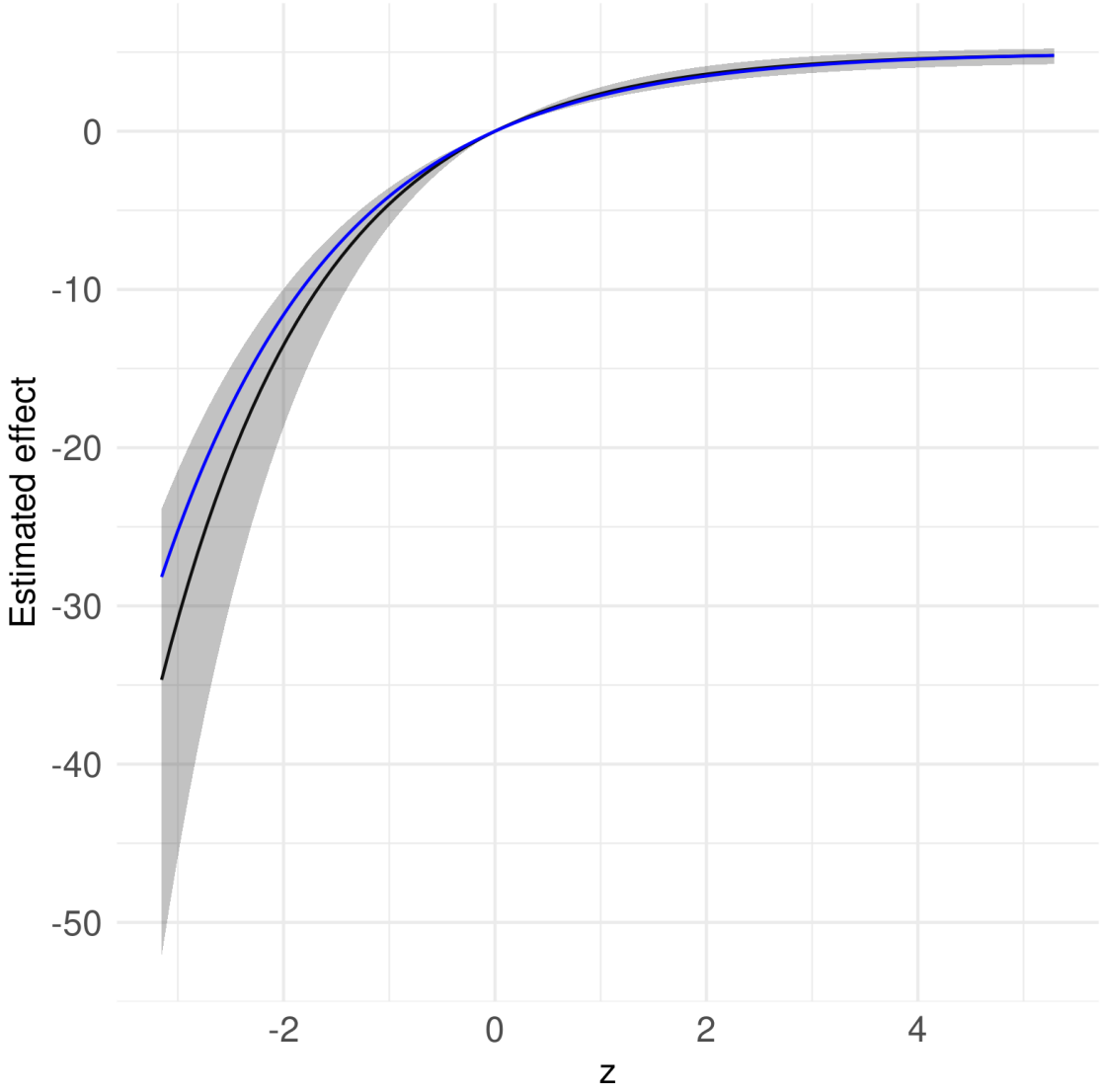


Figure B.2: Black line: Mean posterior effect of z; Grey shaded area: 0.025 and 0.975 pointwise credible intervals for the posterior effect of z; Blue line: The true effect of z used to simulate the data

Bibliography

Allen, M. L., S. Wang, L. O. Olson, Q. Li, and M. Krofel (2020). Counting cats for conservation: Seasonal estimates of leopard density and drivers of distribution in the Serengeti. *Biodiversity and Conservation 29*(13), 3591–3608.

Alves de Aquino, J. (2021). Nvim-R. `https://github.com/jalvesaq/Nvim-R`.

Amstrup, S. C., T. L. McDonald, and B. F. Manly (Eds.) (2005). *Handbook of Capture-Recapture Analysis*. Princeton University Press.

Atkinson, C., K. Woods, R. J. Dusek, L. Sileo, and W. Iko (1995). Wildlife disease and conservation in hawaii: pathogenicity of avian malaria (*Plasmodium relictum*) in experimentally infected iiwi (vestiaria coccinea). *Parasitology 111*(S1), S59–S69.

Bachl, F. E., F. Lindgren, D. L. Borchers, and J. B. Illian (2019). Inlabru: An R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution 0*(0).

Baddeley, A., G. Nair, S. Rakshit, and G. McSwiggan (2017). "Stationary" point processes are uncommon on linear networks. *Stat 6*(1), 68–78.

Baddeley, A., E. Rubak, and R. Turner (2015). *Spatial Point Patterns: Methodology and Applications with r*. CRC Press.

Baddeley, A. and R. Turner (2000). Practical Maximum Pseudolikelihood for Spatial Point Patterns. *Australian & New Zealand Journal of Statistics 42*(3), 283–322.

Bajaru, S. B. and R. Manakadan (2020). Population Characteristics and Microhabitat Preference of Critically Endangered Kondana Soft-Furred Rat (Millardia kondana) in the Northern Western Ghats, India. *Mammal Study 45*(2), 111–122.

Bakka, H. (2019). How to solve the stochastic partial differential equation that gives a Mat\'ern random field using the finite element method. *arXiv:1803.03765 [stat]*.

Bakka, H., H. Rue, G.-A. Fuglstad, A. Riebler, D. Bolin, J. Illian, E. Krainski, D. Simpson, and F. Lindgren (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics 10*(6), e1443.

Bakka, H., J. Vanhatalo, J. B. Illian, D. Simpson, and H. Rue (2019). Non-stationary Gaussian models with physical barriers. *Spatial Statistics 29*, 268–288.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press.

Barrueto, M., M. A. Sawaya, and A. P. Clevenger (2020). Low wolverine (Gulo gulo) density in a national park complex of the Canadian Rocky Mountains. *Canadian Journal of Zoology 98*.

Bauer, S. and M. Klaassen (2013). Mechanistic models of animal migration behaviour – their diversity, structure and use. *Journal of Animal Ecology 82*(3), 498–508.

Berl, J. L., K. F. Kellner, E. A. Flaherty, and R. K. Swihart (2018). Spatial Variation in Density of White-footed Mice Along Edges in Fragmented Habitat. *American Midland Naturalist 179*(1), 38–50.

Berman, M. and T. R. Turner (1992). Approximating Point Process Likelihoods with Glim. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 41*(1), 31–38.

Bivand, R., T. Keitt, and B. Rowlingson (2021). *rgdal: Bindings for the Geospatial Data Abstraction Library*. R package version 1.5-19.

Bivand, R. and N. Lewin-Koh (2020). *maptools: Tools for Handling Spatial Objects*. R package version 1.0-2.

Bivand, R. and C. Rundel (2020). *rgeos: Interface to Geometry Engine - Open Source (GEOS)*. R package version 0.5-5.

Blangiardo, M. and M. Cameletti (2015). *Spatial and Spatio-Temporal Bayesian Models with R - INLA*. John Wiley & Sons.

Bolin, D. and F. Lindgren (2015, January). Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*(1), 85–106.

Bolin, D. and F. Lindgren (2018, September). Calculating Probabilistic Excursion Sets and Related Quantities Using Excursions. *Journal of Statistical Software 86*(1), 1–20.

Borchers, D., G. Distiller, R. Foster, B. Harmsen, and L. Milazzo (2014). Continuous-time spatially explicit capture–recapture models, with an application to a jaguar camera-trap survey. *Methods in Ecology and Evolution 5*(7), 656–665.

Borchers, D. L. and M. G. Efford (2008). Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics 64*(2), 377–385.

Boulanger, J., S. E. Nielsen, and G. B. Stenhouse (2018). Using spatial mark-recapture for conservation monitoring of grizzly bear populations in Alberta. *Scientific Reports 8*(1), 5204.

Bowman, A. W. (2019). Graphics for uncertainty. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 182*(2), 403–418.

Bradbury, G., M. Trinder, B. Furness, A. N. Banks, R. W. G. Caldow, and D. Hume (2014, September). Mapping Seabird Sensitivity to Offshore Wind Farms. *PLOS ONE 9*(9), e106366.

Brauer, F. (2008). Compartmental models in epidemiology. In F. Brauer, P. van den Driessche, and J. Wu (Eds.), *Mathematical Epidemiology*, Lecture Notes in Mathematics, pp. 19–79. Berlin, Heidelberg: Springer.

Bravington, M. V., D. L. Miller, and S. L. Hedley (2021). Variance propagation for density surface models. *Journal of Agricultural, Biological and Environmental Statistics 26*(2), 306–323.

Breen, P., A. Cañadas, O. Ó. Cadhla, M. Mackey, M. Scheidat, S. C. V. Geelhoed, E. Rogan, and M. Jessopp (2017, May). New insights into ocean sunfish (Mola mola) abundance and seasonal distribution in the northeast Atlantic. *Scientific Reports 7*(1), 1–9.

Brenner, S. and R. Scott (2008). *The Mathematical Theory of Finite Element Methods* (Third ed.). Texts in Applied Mathematics. New York: Springer-Verlag.

Broekhuis, F., N. B. Elliot, K. Keiwua, K. Koinet, D. W. Macdonald, N. Mogensen, D. Thuo, and A. M. Gopalaswamy (2021). Resource pulses influence the spatio-temporal dynamics of a large carnivore population. *Ecography 44*(3), 358–369.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas (2004, February). *Advanced Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press.

Buckland, S. T., C. S. Oedekoven, and D. L. Borchers (2016, March). Model-Based Distance Sampling. *Journal of Agricultural, Biological, and Environmental Statistics 21*(1), 58–75.

Buckland, S. T., E. A. Rexstad, T. A. Marques, and C. S. Oedekoven (2015). *Distance Sampling: Methods and Applications*. Methods in Statistical Ecology. Springer International Publishing.

Busenberg, S. (2012). *Differential Equations and Applications in Ecology, Epidemics, and Population Problems*. Elsevier.

Camp, R. J. (2020). Hakalau Forest National Wildlife Refuge, Hawaii Akepa point-transect survey, 2002: U.S. Geological Survey data release. https://doi.org/10.5066/P9Q9UXMZ.

Camp, R. J., K. W. Brinck, P. M. Gorresen, and E. H. Paxton (2016). Evaluating abundance and trends in a hawaiian avian community using state-space analysis. *Bird Conservation International 26*(2).

Camp, R. J., D. L. Miller, L. Thomas, S. T. Buckland, and S. J. Kendall (2020). Using density surface models to estimate spatio-temporal changes in population densities and trend. *Ecography 43*(7), 1079–1089.

Camp, R. J., T. K. Pratt, P. M. Gorresen, J. J. Jeffrey, and B. L. Woodworth (2010). Population Trends of Forest Birds at Hakalau Forest National Wildlife Refuge, Hawai'i. *The Condor 112*(2), 196–212.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan: A probabilistic programming language. *Journal of Statistical Software 76*(1), 1–32.

Chen, Z. (2005). *Finite Element Methods and Their Applications*. Springer Science & Business Media.

Christin, S., É. Hervet, and N. Lecomte (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution 10*(10), 1632–1644.

Clare, J., S. T. McKinney, E. M. Simons-Legaard, J. E. DePue, and C. S. Loftin (2019). Satellite-detected forest disturbance forecasts American marten population decline: The case for supportive space-based monitoring. *Biological Conservation 233*, 336–345.

Cliff, A. D. and K. Ord (1970). Spatial autocorrelation: A review of existing and new measures with applications. *Economic Geography 46*, 269–292.

Corsi, F., E. Duprè, and L. Boitani (1999). A large-scale model of wolf distribution in Italy for conservation planning. *Conservation Biology 13*(1), 150–159.

Cressie, N. and C. K. Wikle (2015). *Statistics for Spatio-Temporal Data*. John Wiley & Sons.

Diggle, P. J., R. Menezes, and T.-l. Su (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 59*(2), 191–232.

Dormann, C. F., J. M. McPherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz, W. D. Kissling, I. Kühn, R. Ohlemüller, P. R. Peres-Neto, B. Reineking, B. Schröder, F. M. Schurr, and R. Wilson (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography 30*(5), 609–628.

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In W. Schempp and K. Zeller (Eds.), *Constructive Theory of Functions of Several Variables*, Lecture Notes in Mathematics, Berlin, Heidelberg, pp. 85–100. Springer.

Efford, M. (2021). *secr: Spatially explicit capture-recapture models*. R package version 4.4.4.

Efford, M. G. and R. M. Fewster (2013). Estimating population size by spatially explicit capture–recapture. *Oikos 122*(6), 918–928.

Englund, G., G. Öhlund, C. L. Hein, and S. Diehl (2011). Temperature dependence of the functional response. *Ecology Letters 14*(9), 914–921.

Espinosa, S., G. Celis, and L. C. Branch (2018). When roads appear jaguars decline: Increased access to an Amazonian wilderness area reduces potential for jaguar conservation. *Plos One 13*(1), e0189740.

Fiske, I. and R. Chandler (2011). Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software 43*, 1–23.

Fithian, W., J. Elith, T. Hastie, and D. A. Keith (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods in Ecology and Evolution 6*(4), 424–438.

Franklin, J. (2010a). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press.

Franklin, J. (2010b). Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions 16*(3), 321–330.

Fuglstad, G.-A., D. Simpson, F. Lindgren, and H. Rue (2019). Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association 114*(525), 445–452.

Fuller, L., M. Shewring, and F. M. Caryl (2018, May). A novel method for targeting survey effort to identify new bat roosts using habitat suitability modelling. *European Journal of Wildlife Research 64*(3).

Furnas, B. J., R. H. Landers, R. G. Paiste, and B. N. Sacks (2020). Overabundance

of Black-Tailed Deer in Urbanized Coastal California. *The Journal of Wildlife Management 84*(5), 979–988.

Galbraith, R. D., R. J. Fryer, and K. M. S. Maitland (1994, June). Demersal pair trawl cod-end selectivity models. *Fisheries Research 20*(1), 13–27.

García-Barón, I., M. Authier, A. Caballero, J. A. Vázquez, M. B. Santos, J. L. Murcia, and M. Louzao (2019). Modelling the spatial abundance of a migratory predator: A call for transboundary marine protected areas. *Diversity and Distributions 25*(3), 346–360.

Gaukler, S. M., S. M. Murphy, J. T. Berryhill, B. E. Thompson, B. J. Sutter, and C. D. Hathcock (2020). Investigating effects of soil chemicals on density of small mammal bioindicators using spatial capture-recapture models. *Plos One 15*(9), e0238870.

Gelfand, A. E., P. Diggle, P. Guttorp, and M. Fuentes (2010). *Handbook of Spatial Statistics*. CRC Press.

Glennie, R., D. L. Borchers, M. Murchie, B. J. Harmsen, and R. J. Foster (2019). Open population maximum likelihood spatial capture-recapture. *Biometrics 75*(4), 1345–1355.

Gomez-Rubio, V. (2020). *Bayesian Inference with INLA*. CRC Press.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika 53*(3-4), 325–338.

Guisan, A., O. Broennimann, R. Engler, M. Vust, N. G. Yoccoz, A. Lehmann, and N. E. Zimmermann (2006). Using niche-based models to improve the sampling of rare species. *Conservation Biology 20*(2), 501–511.

Harihar, A., B. Pandav, M. Ghosh-Harihar, and J. Goodrich (2020). Demographic

and ecological correlates of a recovering tiger (Panthera tigris) population: Lessons learnt from 13-years of monitoring. *Biological Conservation 252*, 108848.

Havmøller, R. W., S. Tenan, N. Scharff, and F. Rovero (2019). Reserve size and anthropogenic disturbance affect the density of an African leopard (Panthera pardus) meta-population. *PLOS ONE 14*(6), e0209541.

Hearn, A. J., J. Ross, H. Bernard, S. A. Bakar, B. Goossens, L. T. B. Hunter, and D. W. Macdonald (2019). Responses of Sunda clouded leopard Neofelis diardi population density to anthropogenic disturbance: Refining estimates of its conservation status in Sabah. *Oryx 53*(4), 643–653.

Hedley, S. L. and S. T. Buckland (2004, June). Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics 9*(2), 181.

Herr, H., N. Kelly, B. Dorschel, M. Huntemann, K.-H. Kock, L. S. Lehnert, U. Siebert, S. Viquerat, R. Williams, and M. Scheidat (2019). Aerial surveys for Antarctic minke whales (Balaenoptera bonaerensis) reveal sea ice dependent distribution patterns. *Ecology and Evolution 9*(10), 5664–5682.

Hijmans, R. J. (2020). *raster: Geographic Data Analysis and Modeling*. R package version 3.4-5.

Hodges, J. S. (2019). *Richly parameterized linear models: additive, time series, and spatial models using random effects*. Chapman and Hall/CRC.

Högmander, H. (1991). A Random Field Approach to Transect Counts of Wildlife Populations. *Biometrical Journal 33*(8), 1013–1023.

Holling, C. S. (1959). Some characteristics of simple types of predation and parasitism. *The Canadian Entomologist 91*(7), 385–398.

Holmes, E. E., M. A. Lewis, J. E. Banks, and R. R. Veit (1994). Partial differential equations in ecology: Spatial interactions and population dynamics. *Ecology 75*(1), 17–29.

Horn, P. E., M. J. R. Pereira, T. C. Trigo, E. Eizirik, and F. P. Tirelli (2020). Margay (Leopardus wiedii) in the southernmost Atlantic Forest: Density and activity patterns under different levels of anthropogenic disturbance. *PLOS ONE 15*(5), e0232013.

Horvitz, D. G. and D. J. Thompson (1952). A Generalization of Sampling Without Replacement From a Finite Universe. *Journal of the American Statistical Association 47*(260), 663–685.

Hughes, J. and M. Haran (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(1), 139–159.

Illian, D. J., P. A. Penttinen, D. H. Stoyan, and D. D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons.

IUCN (2016). The IUCN Red List of Threatened Species.

Iverson, L. R., A. M. Prasad, S. N. Matthews, and M. P. Peters (2011). Lessons learned while integrating habitat, dispersal, disturbance, and life-history traits into species habitat models under climate change. *Ecosystems 14*(6), 1005–1020.

Janssen, M. A. and E. Ostrom (2006). Empirically based, agent-based models. *Ecology and Society 11*(2).

Jarvie, S. and J.-C. Svenning (2018). Using species distribution modelling to determine opportunities for trophic rewilding under future scenarios of climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences 373*(1761), 20170446.

Johnson, C. J., S. E. Nielsen, E. H. Merrill, T. L. McDONALD, and M. S. Boyce (2006). Resource selection functions based on use-availability data: Theoretical motivation and evaluation methods. *The Journal of Wildlife Management 70*(2), 347–357.

Johnson, D. H. (1980). The comparison of usage and availability measurements for evaluating resource preference. *Ecology 61*(1), 65–71.

Johnson, D. S., J. L. Laake, and J. M. Ver Hoef (2010, March). A Model-Based Approach for Making Ecological Inference from Distance Sampling Data. *Biometrics 66*(1), 310–318.

Judge, S. W., R. J. Camp, P. J. Hart, and S. T. Kichman (2018). Population estimates of the Endangered Hawai'i 'Ākepa (Loxops coccineus) in different habitats on windward Mauna Loa. *Journal of Field Ornithology 89*(1), 11–21.

Kahle, D. and H. Wickham (2013). ggmap: Spatial visualization with ggplot2. *The R Journal 5*(1), 144–161.

Kammann, E. E. and M. P. Wand (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 52*(1), 1–18.

Khanal, G., C. Mishra, and K. R. Suryawanshi (2020). Relative influence of wild prey and livestock abundance on carnivore-caused livestock predation. *Ecology and Evolution 10*(20), 11787–11797.

Kimeldorf, G. S. and G. Wahba (1970a). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics 41*(2), 495–502.

Kimeldorf, G. S. and G. Wahba (1970b). Spline functions and stochastic processes. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 32*.

Kittle, A. M., A. C. Watson, and P. K. L. Samaranayake (2021). Edge effects and distribution of prey forage resources influence how an apex predator utilizes Sri Lanka's largest protected area. *Journal of Zoology n/a*(n/a).

Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. M. Bell (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software 70*(1), 1–21.

Kumar, U., N. Awasthi, Q. Qureshi, and Y. Jhala (2019). Do conservation strategies that increase tiger populations have consequences for other wild carnivores like leopards? *Scientific Reports 9*(1), 14673.

Lahiri, S. N. (2003). *Resampling methods for dependent data*. Springer Science & Business Media.

Lamb, C. T., G. Mowat, A. Reid, L. Smit, M. Proctor, B. N. McLellan, S. E. Nielsen, and S. Boutin (2018). Effects of habitat quality and access management on the density of a recovering grizzly bear population. *Journal of Applied Ecology 55*(3), 1406–1417.

Lamichhane, B. R., H. Leirs, G. A. Persoon, N. Subedi, M. Dhakal, B. N. Oli, S. Reynaert, V. Sluydts, C. P. Pokheral, L. P. Poudyal, S. Malla, and H. H. de Iongh (2019). Factors associated with co-occurrence of large carnivores in a human-dominated landscape. *Biodiversity and Conservation 28*(6), 1473–1491.

Lang, S. and A. Brezger (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics 13*(1), 183–212.

Laufenberg, J. S., H. E. Johnson, P. F. Doherty, and S. W. Breck (2018). Compounding effects of human development and a natural food shortage on a black bear population along a human development-wildland interface. *Biological Conservation 224*, 188–198.

Lavery, T. H., M. Alabai, P. Holland, C. Qaqara, N. Vatohi, T. H. Lavery, M. Alabai, P. Holland, C. Qaqara, and N. Vatohi (2020). Feral cat abundance, density and activity in tropical island rainforests. *Wildlife Research 47*(8), 660–668.

Lepson, J. and L. Freed (1997). Akepa (*Loxops coccineus*). In A. Pool and F. Gill (Eds.), *The birds of North America, No. 294*, Number 294, pp. 1–23. Philadelphia, PA, USA: The Academy of Natural Sciences.

Leroux, S. J., M. Larrivée, V. Boucher-Lalonde, A. Hurford, J. Zuloaga, J. T. Kerr, and F. Lutscher (2013). Mechanistic models for the spatial spread of species under climate change. *Ecological Applications 23*(4), 815–828.

Linden, D. W., A. P. K. Siren, and P. J. Pekins (2018). Integrating telemetry data into spatial capture-recapture modifies inferences on multi-scale resource selection. *Ecosphere 9*(4), e02203.

Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(4), 423–498.

Loosen, A. E., A. T. Morehouse, and M. S. Boyce (2019). Land tenure shapes black bear density and abundance on a multi-use landscape. *Ecology and Evolution 9*(1), 73–89.

Louvrier, J., J. Papaïx, C. Duchamp, and O. Gimenez (2020). A mechanistic–statistical species distribution model to explain and forecast wolf (Canis lupus) colonization in South-Eastern France. *Spatial Statistics 36*, 100428.

Mann, G. K. H., M. J. O'Riain, and D. M. Parker (2020). A leopard's favourite spots: Habitat preference and population density of leopards in a semi-arid biodiversity hotspot. *Journal of Arid Environments 181*, 104218.

Mardia, K. V. (1978). Some properties of clasical multi-dimensional scaling. *Communications in Statistics - Theory and Methods 7*(13), 1233–1241.

Margossian, C. C. (2019). A review of automatic differentiation and its efficient implementation. *WIREs Data Mining and Knowledge Discovery 9*(4), e1305.

Marques, F. F. C. and S. T. Buckland (2003). Incorporating covariates into standard line transect analyses. *Biometrics 59*(4), 924–935.

Marra, G. and S. N. Wood (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics 39*(1), 53–74.

Matthiopoulos, J., J. Fieberg, and G. Aarts (2020). *Species-Habitat Associations: Spatial Data, Predictive Models, and Ecological Insights*. University of Minnesota Libraries Publishing.

Matthiopoulos, J., M. Hebblewhite, G. Aarts, and J. Fieberg (2011). Generalized functional responses for species distributions. *Ecology 92*(3), 583–589.

Maxfield, B. (1998). Hakalau Forest National Wildlife Refuge. *Endangered Species Bulletin 23*, 26–27.

McCullagh, P. and J. A. Nelder (2019). *Generalized Linear Models* (Second ed.). Boca Raton: Routledge.

McDonald, P. J., A. Stewart, M. A. Jensen, and H. W. McGregor (2020). Topographic complexity potentially mediates cat predation risk for a critically endangered rodent. *Wildlife Research 47*(7-8), 643–648.

McLane, A. J., C. Semeniuk, G. J. McDermid, and D. J. Marceau (2011). The role of agent-based models in wildlife ecology and management. *Ecological Modelling 222*(8), 1544–1556.

McSwiggan, G., A. Baddeley, and G. Nair (2017). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics 44*(2), 324–345.

Michelot, T., R. Glennie, C. Harris, and L. Thomas (2021). Varying-coefficient stochastic differential equations with applications in ecology. *Journal of Agricultural, Biological and Environmental Statistics*.

Millar, R. B. and R. J. Fryer (1999). Estimating the size-selection curves of towed gears, traps, nets and hooks. *Reviews in Fish Biology and Fisheries 9*(1), 89–116.

Miller, D. L., M. L. Burt, E. A. Rexstad, and L. Thomas (2013, July). Spatial models for distance sampling data: Recent developments and future directions. *Methods in Ecology and Evolution 4*(11), 1001–1010.

Miller, D. L., R. Glennie, and A. E. Seaton (2019). Understanding the stochastic partial differential equation approach to smoothing. *Journal of Agricultural, Biological and Environmental Statistics*.

Miller, D. L. and S. N. Wood (2014). Finite area smoothing with generalized distance splines. *Environmental and Ecological Statistics 21*(4), 715–731.

Møller, J., A. R. Syversveen, and R. P. Waagepetersen (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics 25*(3), 451–482.

Moolenaar, B. (2019). `https://www.vim.org/`.

Müller, K. (2020). *here: A Simpler Way to Find Your Files*. R package version 1.0.1.

Nelson, D. L., K. F. Kellner, and R. K. Swihart (2019). Rodent population density and survival respond to disturbance induced by timber harvest. *Journal of Mammalogy 100*(4), 1253–1262.

Niemi, A. and C. Fernández (2010, September). Bayesian Spatial Point Process Modeling of Line Transect Data. *Journal of Agricultural, Biological, and Environmental Statistics 15*(3), 327–345.

Oaten, A. and W. W. Murdoch (1975). Functional response and stability in predator-prey systems. *The American Naturalist 109*(967), 289–298.

Oedekoven, C. S., S. T. Buckland, M. L. Mackenzie, R. King, K. O. Evans, and L. W. Burger (2014, June). Bayesian Methods for Hierarchical Distance Sampling Models. *Journal of Agricultural, Biological, and Environmental Statistics 19*(2), 219–239.

Oksendal, B. (2013). *Stochastic Differential Equations: An Introduction with Applications*. Springer Science & Business Media.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal 10*(1), 439–446.

Pebesma, E. and R. Bivand (2021). *sp: Classes and Methods for Spatial Data*. R package version 1.4-5.

Pedersen, T. L. (2020). *patchwork: The Composer of Plots*. R package version 1.1.1.

Peel, S. L., N. A. Hill, S. D. Foster, S. J. Wotherspoon, C. Ghiglione, and S. Schiaparelli (2019). Reliable species distributions are obtainable with sparse, patchy and biased data by leveraging over species and data types. *Methods in Ecology and Evolution 10*(7), 1002–1014.

Penjor, U., D. W. Macdonald, S. Wangchuk, T. Tandin, and C. K. W. Tan (2018). Identifying important conservation areas for the clouded leopard Neofelis nebulosa in a mountainous landscape: Inference from spatial modeling techniques. *Ecology and Evolution 8*(8), 4278–4291.

Phillips, S. J., R. P. Anderson, and R. E. Schapire (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling 190*(3), 231–259.

Plummer, M. (2017). *JAGS Version 4.3.0 user manual*.

Porfirio, L. L., R. M. B. Harris, E. C. Lefroy, S. Hugh, S. F. Gould, G. Lee, N. L. Bindoff, and B. Mackey (2014). Improving the use of species distribution models in conservation planning and management under climate change. *PLOS ONE 9*(11), e113749.

Pratt, H. D. (1994). Avifaunal change in the Hawaiian Islands, 1893–1993. *Studies in Avian Biology 15*, 103–118.

R Core Team (2017). *R: A Language and Environment for Statistical Computing. Version 3.4.1*. Vienna, Austria: R Foundation for Statistical Computing.

Rahman, D. A., P. Rianti, M. Muhiban, A. Muhtarom, U. M. Rahmat, Y. Santosa, and S. Aulagnier (2018). Density and spatial partitioning of endangered sympatric Javan leopard (Felidae) and dholes (Canidae) in a tropical forest landscape. *Folia Zoologica 67*(3-4), 207–219.

Rakshit, S., T. Davies, M. M. Moradi, G. McSwiggan, G. Nair, J. Mateu, and A. Baddeley (2019). Fast Kernel Smoothing of Point Patterns on a Large Network using Two-dimensional Convolution. *International Statistical Review 87*(3), 531–556.

Rakshit, S., G. Nair, and A. Baddeley (2017). Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics 22*, 129–154.

Rammer, W. and R. Seidl (2019). Harnessing deep learning in ecology: An example predicting bark beetle outbreaks. *Frontiers in Plant Science 10*.

Ramsay, T. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(2), 307–319.

Rasmussen, C. E. and C. K. I. Williams (2005). *Gaussian Processes for Machine Learning*. The MIT Press.

Rasmussen, J. G. and H. S. Christensen (2021). Point processes on directed linear networks. *Methodology and Computing in Applied Probability 23*(2), 647–667.

Rastetter, E. B., J. D. Aber, D. P. C. Peters, D. S. Ojima, and I. C. Burke (2003). Using mechanistic models to scale ecological processes across space and time. *BioScience 53*(1), 68–76.

Rather, T. A., S. Kumar, and J. A. Khan (2021). Density estimation of tiger and leopard using spatially explicit capture–recapture framework. *PeerJ 9*, e10634.

Real, L. A. (1979). Ecological determinants of functional response. *Ecology 60*(3), 481–485.

Reich, B. J. and B. Gardner (2014). A spatial capture-recapture model for territorial species. *Environmetrics 25*(8), 630–637.

Ripley, B. (2021). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. R package version 7.3-54.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability 13*(2), 255–266.

Riutort-Mayol, G., P.-C. Bürkner, M. R. Andersen, A. Solin, and A. Vehtari (2020). Practical Hilbert space approximate Bayesian Gaussian processes for probabilistic programming. *arXiv:2004.11408 [stat]*.

Rosenbaum, B. and B. C. Rall (2018). Fitting functional responses: Direct parameter estimation by simulating differential equations. *Methods in Ecology and Evolution 9*(10), 2076–2090.

Royle, J. A., D. K. Dawson, and S. Bates (2004). Modeling abundance effects in distance sampling. *Ecology 85*(6), 1591–1597.

Royle, J. A., A. K. Fuller, and C. Sutherland (2016). Spatial capture–recapture models allowing Markovian transience or dispersal. *Population Ecology 58*(1), 53–62.

Royle, J. A., A. K. Fuller, and C. Sutherland (2018). Unifying population and landscape ecology with spatial capture–recapture. *Ecography 41*(3), 444–456.

Royle, J. A. and K. V. Young (2008). A hierarchical model for spatial capture–recapture data. *Ecology 89*(8), 2281–2289.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, PBC.

Rue, H. and T. Follestad (2003). GMRFLib: A C-library for fast and exact simulation of Gaussian Markov random fields.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications*. CRC Press.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(2), 319–392.

Russell, D. J. F., G. D. Hastie, D. Thompson, V. M. Janik, P. S. Hammond, L. A. S. Scott-Hayward, J. Matthiopoulos, E. L. Jones, and B. J. McConnell (2016). Avoidance of wind farms by harbour seals is limited to pile driving activities. *Journal of Applied Ecology 53*(6), 1642–1652.

Santos-Fernandez, E., J. M. V. Hoef, E. E. Peterson, J. McGree, D. Isaak, and K. Mengersen (2021). Bayesian spatio-temporal models for stream networks. *arXiv:2103.03538 [stat]*.

Schiesser, W. E. (2012). *The Numerical Method of Lines: Integration of Partial Differential Equations*. Elsevier.

Schlather, M., A. Malinowski, P. J. Menck, M. Oesting, and K. Strokorb (2015). Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software 63*(8), 1–25.

Schmidt, J. H., K. L. Rattenbury, J. P. Lawler, and M. C. Maccluskie (2012). Using distance sampling and hierarchical models to improve estimates of Dall's sheep abundance. *The Journal of Wildlife Management 76*(2), 317–327.

Schwarz, C. J. and G. A. F. Seber (1999). Estimating animal abundance: Review III. *Statistical Science 14*(4), 427–456.

Scott, J. M., S. Mountainspring, F. L. Ramsey, and C. B. Kepler (1986). Forest bird communities of the Hawaiian Islands: their dynamics, ecology, and conservation. *Studies in Avian Biology, No. 9 9*, 1–431.

Scott-Hayward, L. A. S., M. L. Mackenzie, C. R. Donovan, C. G. Walker, and E. Ashe (2014). Complex region spatial smoother (CReSS). *Journal of Computational and Graphical Statistics 23*(2), 340–360.

Simpson, D., J. B. Illian, F. Lindgren, S. H. Sørbye, and H. Rue (2016). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika 103*(1), 49–70.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science 32*(1), 1–28.

Smout, S., C. Asseburg, J. Matthiopoulos, C. Fernández, S. Redpath, S. Thirgood, and J. Harwood (2010). The functional response of a generalist predator. *PLoS ONE 5*(5).

Speckman, P. L. and D. Sun (2003). Fully bayesian spline smoothing and intrinsic autoregressive priors. *Biometrika 90*(2), 289–302.

Stan Development Team (2020). RStan: the R interface to Stan. R package version 2.21.2.

Stein, M. L. (2012). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media.

Stevenson, B. C., R. M. Fewster, and K. Sharma (2021). Spatial correlation structures for detections of individuals in spatial capture-recapture models. *Biometrics*.

Stokes, E. J., S. Strindberg, P. C. Bakabana, P. W. Elkan, F. C. Iyenguet, B. Madzoké, G. A. F. Malanda, B. S. Mowawa, C. Moukoumbou, F. K. Ouakabadio, and H. J. Rainey (2010, April). Monitoring Great Ape and Elephant Abundance at Large Spatial Scales: Measuring Effectiveness of a Conservation Landscape. *PLOS ONE 5*(4), e10294.

Stoyan, D. (1982). A Remark on the Line Transect Method. *Biometrical Journal 24*(2), 191–195.

Strimas-Mackey, M. (2020). *smoothr: Smooth and Tidy Spatial Features*. R package version 0.1.2.

Stringer, A., P. Brown, and J. Stafford (2021). Fast, scalable approximations to posterior distributions in extended latent Gaussian models. *arXiv:2103.07425 [stat]*.

Sutherland, C., A. K. Fuller, J. A. Royle, M. P. Hare, and S. Madden (2018). Large-scale variation in density of an aquatic ecosystem indicator species. *Scientific Reports 8*(1), 8958.

Swallow, B., S. Buckland, R. King, and M. Toms (2020). Data from: Assessing

factors associated with changes in the numbers of birds visiting gardens in winter: Are predators partly to blame?

Swallow, B., S. T. Buckland, R. King, and M. P. Toms (2016). Bayesian hierarchical modelling of continuous non-negative longitudinal data with a spike at zero: An application to a study of birds visiting gardens in winter. *Biometrical Journal 58*(2), 357–371.

Tirelli, F. P., F. D. Mazim, P. G. Crawshaw, A. P. Albano, C. Espinosa, D. Queirolo, F. L. Rocha, J. B. Soares, T. C. Trigo, D. W. Macdonaldb, M. Lucherini, and E. Eizirik (2019). Density and spatio-temporal behaviour of Geoffroy's cats in a human-dominated landscape of southern Brazil. *Mammalian Biology 99*(1), 128–135.

Tobler, M. W., R. Garcia Anleu, S. E. Carrillo-Percastegui, G. Ponce Santizo, J. Polisar, A. Zuñiga Hartley, and I. Goldstein (2018). Do responsibly managed logging concessions adequately protect jaguars and other large and medium-sized mammals? Two case studies from Guatemala and Peru. *Biological Conservation 220*, 245–253.

U.S. Fish and Wildlife Service (1970). Notice of proposed rule making (conservation of endangered specie and other fish and wildlife). *Federal register 35*(50 CFR Part 17), 13519–13520.

U.S. Fish and Wildlife Service (2006). *Revised recovery plan for Hawaiian forest birds.* Region 1, Portland, Oregon, USA: U.S. Fish and Wildlife Service.

Vallejo, G. C., K. Grellier, E. J. Nelson, R. M. McGregor, S. J. Canning, F. M. Caryl, and N. McLean (2017). Responses of two marine top predators to an offshore wind farm. *Ecology and Evolution 7*(21), 8698–8708.

Van der Putten, W. H., M. Macel, and M. E. Visser (2010). Predicting species distribution and abundance responses to climate change: Why it is essential to

include biotic interactions across trophic levels. *Philosophical Transactions of the Royal Society B: Biological Sciences 365*(1549), 2025–2034.

van der Zander, B., J. Sundermeyer, D. Braun, and T. Hoffmann (2021). Texstudio. `https://www.texstudio.org/`.

van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software 76*(1), 1–21.

Vonesh, E. F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika 83*(2), 447–452.

Waagepetersen, R. and T. Schweder (2006). Likelihood-based inference for clustered line transect data. *Journal of Agricultural, Biological, and Environmental Statistics 11*(3), 264.

Wahba, G. (1990). *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics.

Wang, T., J. A. Royle, J. L. D. Smith, L. Zou, X. Lu, T. Li, H. Yang, Z. Li, R. Feng, Y. Bian, L. Feng, and J. Ge (2018). Living on the edge: Opportunities for Amur tiger recovery in China. *Biological Conservation 217*, 269–279.

Welfelt, L. S., R. A. Beausoleil, and R. B. Wielgus (2019). Factors associated with black bear density and implications for management. *The Journal of Wildlife Management 83*(7), 1527–1539.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika 41*(3-4), 434–449.

Wickham, H. (2020). *tidyr: Tidy Messy Data*. R package version 1.1.2.

Wickham, H., W. Chang, L. Henry, T. L. Pedersen, K. Takahashi, C. Wilke, K. Woo, H. Yutani, and D. Dunnington (2020). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.3.3.

Wickham, H., R. François, L. Henry, and K. Müller (2021). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.6.

Wickham, H. and D. Seidel (2020). *scales: Scale Functions for Visualization*. R package version 1.1.1.

Wikle, C. K. (2015). Modern perspectives on statistics for spatio-temporal data. *WIREs Computational Statistics 7*(1), 86–98.

Wilensky, U. and W. Rand (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NETLogo*. MIT press.

Williams, J. N., C. Seo, J. Thorne, J. K. Nelson, S. Erwin, J. M. O'Brien, and M. W. Schwartz (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions 15*(4), 565–576.

Williams, R., S. Hedley, and P. Hammond (2006, February). Modeling Distribution and Abundance of Antarctic Baleen Whales Using Ships of Opportunity. *Ecology and Society 11*(1).

Williams, R., S. L. Hedley, T. A. Branch, M. V. Bravington, A. N. Zerbini, and K. P. Findlay (2011). Chilean Blue Whales as a Case Study to Illustrate Methods to Estimate Abundance and Evaluate Conservation Status of Rare Species. *Conservation Biology 25*(3), 526–535.

Wilson, T. L., J. B. Odei, M. B. Hooten, and T. C. E. Jr (2010). Hierarchical spatial models for predicting pygmy rabbit distribution and relative abundance. *Journal of Applied Ecology 47*(2), 401–409.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(1), 95–114.

Wood, S. N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics 48*(4), 445–464.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(1), 3–36.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (second ed.). CRC Press.

Wood, S. N. (2020). Simplified integrated nested Laplace approximation. *Biometrika 107*(1), 223–230.

Wood, S. N., M. V. Bravington, and S. L. Hedley (2008). Soap film smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(5), 931–955.

Yang, H., B. Xie, G. Zhao, Y. Gong, P. Mou, J. Ge, and L. Feng (2021). Elusive cats in our backyards: Persistence of the North Chinese leopard (Panthera pardus japonensis) in a human-dominated landscape in central China. *Integrative Zoology 16*(1), 67–83.

Yuan, Y., F. E. Bachl, F. Lindgren, D. L. Borchers, J. B. Illian, S. T. Buckland, H. Rue, and T. Gerrodette (2017). Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *The Annals of Applied Statistics 11*(4), 2270–2297.

Yue, Y. R., D. Simpson, F. Lindgren, and H. Rue (2014). Bayesian adaptive smooth-

ing splines using stochastic differential equations. *Bayesian Analysis 9*(2), 397–424.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association 99*(465), 250–261.