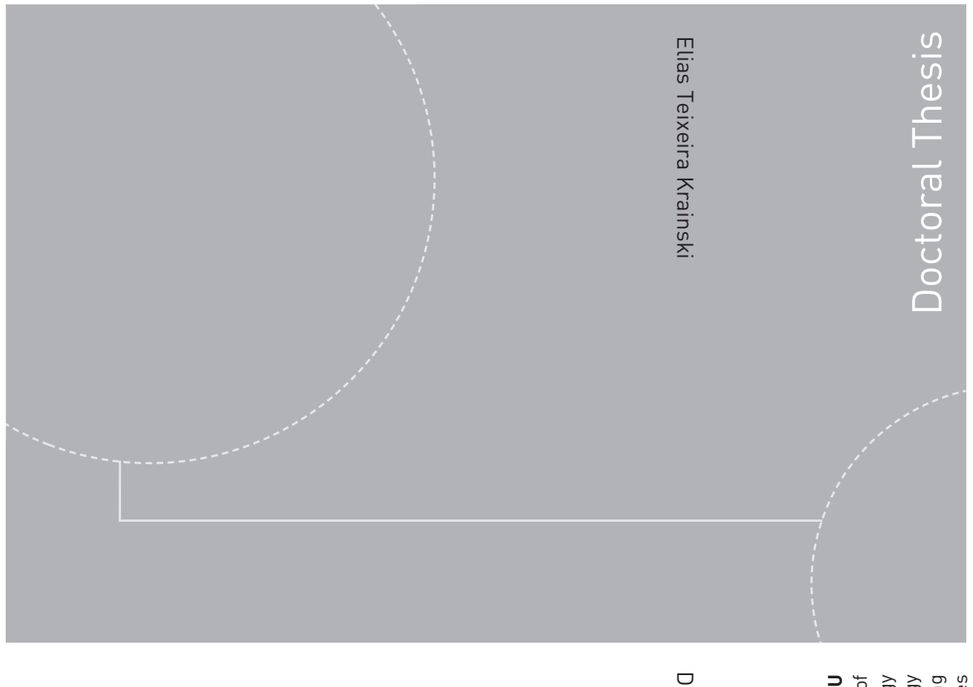
ISBN 978-82-326-2976-3 (printed version)
ISBN 978-82-326-2977-0 (electronic version)
ISSN 1503-8181



Doctoral theses at NTNU, 2018:093

Elias Teixeira Krainski

Statistical Analysis of Space-time Data: New Models and Applications

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Doctoral theses at NTNU, 2018:093

NTNU

NTNU
Norwegian University of
Science and Technology

Elias Teixeira Krainski

Statistical Analysis of Space-time Data: New Models and Applications

Thesis for the degree of Philosophiae Doctor

Trondheim, March 2018

Norwegian University of Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology
and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-326-2976-3 (printed version)
ISBN 978-82-326-2977-0 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2018:093

Printed by Skipnes Kommunikasjon as

Acknowledgments

If I start talking about something, I would like to give details.

By end of 2012 Håvard Rue wrote to my colleague Paulo Justiniano Ribeiro Júnior about a project on space-time and asked about me. In the same week my wife and I knew that we would have a new baby. Håvard arranged things and I went alone to Trondheim for a two and half month stay in the beginning of 2013 to get things started. We moved in August 2013 to Trondheim with our two daugthers. I thank Håvard for this. This ended up being a wonderful choice in several ways. It was quite a life experience and we all liked it. Norway will always be in our hearts.

I thank Håvard for being a good planner and good at connecting people. He knows how to create a good working environment and how to treat people well. I was fortunate to have him as my advisor. Several small points are still missing here, however certainly contributed to my training. Thanks also for understanding my limitations. Thank you for kindness again in arranging my short stay in KAUST at the end of 2017 for me to finish my thesis. He was clever in guiding me on how to organize this work and advising me how to present it in a reasonable way.

My thanks to Anne Kajander who helped to find a place to live in the beginning and when we moved. Seems that she knew and always solved practicalities. Thanks also to the other staff in the department for making a good working environment.

Thanks to Thiago who was there before me doing his PhD. For all the lunch times we had together, and sometimes together with other Brazilians around. For the dinners as well. For you and Gabriella introducing us to other Brazilians in Trondheim. Thanks for all the Brazilians we meet in Trondheim. For all the dinner and parties we had together. You really made our stay better.

I also thank Geir-Arne and Haakon for being good colleagues while we shared an office. I definitely learned something from you both during our many discussions together. Also, thanks to the other colleagues for the

discussions around the INLA group and for the fellowship we had together: Håvard, Xiangping, Rikke, Jingyi, Morten and Alexander.

I also want to thank all of those people who contributed in some way to my work. Dan, who knows how to ask questions. Finn, the background he gave me for part 1. Andrea, for joining in part 2. To Aurelie, Damaris, Ana Isabel, Jed and Marco for driving my enjoyable work in part 3. To Laurie for reviewing my English (anything still not perfect is my fault) and motivating an application for the models in part 1. Thank you all for the the partnership.

Thanks to my wife and my daughters for their love. To my parents and brothers for all the support you gave in different ways in this period.

Thanks to my colleagues at UFPR for letting me take this opportunity. Specifically to my colleagues at LEG for your friendly, stimulating discussions and for taking my teaching load. And especially to Paulo who taught me a lot and for giving me more time to finish my thesis in the end.

I am still afraid to have forgotten some things...

Elias Teixeira Krainski
Curitiba, March 5, 2018

> Consuel in the heart of man is like deep water; but a man of understanding will draw it out. **PROVERBS 20:5**

> There is nothing better for a man, than that he should eat and drink, and that he should make his soul enjoy good in his labour. This also I saw, that it was from the hand of God. **ECCLESIASTES 2:24**

Preface

This thesis is divided into four parts.

In the first part considers the stochastic partial differential equation - SPDE approach to define non-separable space-time models. A review of this approach for modeling in one and two dimensions by summarizing the results in Lindgren et al. (2011). This is them considered when working with the space-time models. The main objective considered in the work with the space-time models was to study marginal properties of the implied multivariate Gaussian distributions. This was made through the spectral properties which was also considered when building Gaussian Markov random field representations.

The second part is a draft of a paper on a class of space-time intrinsic models for areal data. It considers the issue of the needing of a lot of constraints when fitting these models. The usual sum-to-zero ones are computationally expensive to work with when the dimension is big. Two approaches were considered in order to avoid the sum-to-zero constraints. A simulation study was performed in order to compare the approaches and an application to real data was also considered.

The third part presents a collection of contributed work in several applications of space and space-time models. Having in mind that useful methods in practice can be better applied through a partnership between statisticians and researchers. It was the case to be very stimulating working with experts. Statisticians can can learn the weakness and robustness of statistical methods and refine them through a collaborative work.

Seems that didactic material available online is useful for the spreading of new methods. With this in mind, the last part is an educative material on spatial and space-time modeling with SPDE and using the Integrated Nested Laplace Applications, Rue et al. (2009). It shows how use several features of the **INLA** package to extend the modeling in some different directions. For example, joint modeling of more than one outcome, each with a different likelihood. This material has been used several times.

Though all of it I learned so many things. I have learned a bit of stochastic process when working with the models in the first part of the thesis. Before considering the approach in part 2, I have hit the problem over years. The work introduced in part 3 tauch me about real world and was enjoyable. The last part reveals something like: "as soon as I learn something, I start teaching it."

Elias Teixeira Krainski
Curitiba, March 5, 2018

> If the iron be blunt, and he do not whet the edge, then must he put to more strength: but wisdom is profitable to direct.
> **ECCLESIASTES 10:10**

Contents

Chapter 1

Introduction

In many research fields data are collected from a phenomena that evolves over space and time. Some statistical models have been proposed in the literature to model these kinds of data in order to understand the phenomena observed considering both the spatial and temporal processes explicitly.

In this work we refer to these models as space-time [1].

A statistical model is an empirical stochastic model that is proposed to model a set of observed data. A space-time statistical model is a kind of statistical model that explicitly takes space-time variation into account. Such variation may be due to the structure of the data collection resulting from repeated measurements on individuals that occur near in time or space or both.

A statistical model assumes a statistical distribution, or likelihood function, for the outcome, $\boldsymbol{y}$. Usually we build a model for the expected value of $\boldsymbol{y}$ based on two kinds of effects. The first relates to causal effects in the response/outcome from factors/covariates of interest which are called fixed effects. The second includes the structure of the data and is used to account for extra variation in the outcome that is not explained by the fixed effects. These two effects also form the basis for the analysis of longitudinal data, Diggle et al. (2013).

This kind of mixed model is translated into the linear predictor as

$$\boldsymbol{\eta} = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}$$

where $\boldsymbol{F}$ is a $n \times p + 1$ known matrix with the first column with ones and the last p columns including information from p covariates, $\boldsymbol{\beta}$ the regression

[1] We prefer the term *space-time* to *spacetime*, *spatio-temporal* or *spatiotemporal* since it is the most frequent in both the British National and the Contemporary American English, corpuses as seen at `http://corpus.byu.edu` on February 2016.

coefficients, or fixed effects, $\boldsymbol{Z}$ is a $n \times m$ known projection matrix for the m dimensional random effects vector $\boldsymbol{b}$. We have $\boldsymbol{F}$ and $\boldsymbol{Z}$ known and $\boldsymbol{\beta}$ and $\boldsymbol{b}$ unknown.

An important concept in a statistical model is how $\boldsymbol{y}$ depends on $\boldsymbol{\eta}$. The key idea is to consider that given a particular $\boldsymbol{\eta}_i$ the observation $\boldsymbol{y}_i$ is a random variable. It is equivalent to say that given a condition $\boldsymbol{F}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}$ we can assume a probability density function or probability mass distribution for $\boldsymbol{y}_i$. This is also called the likelihood function assumed for $\boldsymbol{y}$. If we have a Gaussian likelihood, $\boldsymbol{y}|\boldsymbol{\eta}, \tau_y \sim \text{Normal}(\boldsymbol{\eta}, \tau_y^{-1}\boldsymbol{I})$, τ_y is the likelihood parameter to measure the uncertainty of $\boldsymbol{y}$ given $\boldsymbol{\eta}$.

There are cases when an assumed likelihood has no additional parameter. There are also cases when it is neccesary to specify a link function between the expected value for $\boldsymbol{y}$ and $\boldsymbol{\eta}$ such as when we have a Poisson likelihood. In these cases the log-link function is usually specified such that $\boldsymbol{\lambda} = \exp(\boldsymbol{\eta})$ for $\boldsymbol{y}|\boldsymbol{\eta} \sim \text{Poisson}(\boldsymbol{\lambda})$ and there is no additional parameter.

The main interest is to infer $\boldsymbol{\beta}$, $\boldsymbol{b}$ and possible additional parameters. Here, we briefly describe the inference for $\boldsymbol{\beta}$ and $\boldsymbol{b}$. In most cases a Gaussian distribution is assumed for the random effect $\boldsymbol{b}$ and for $\boldsymbol{\beta}$ as well. We can therefore write the linear predictor as

$$\boldsymbol{\eta} = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$$

where $\boldsymbol{A} = [\boldsymbol{F}\ \boldsymbol{Z}]$ and $\boldsymbol{x} = [\boldsymbol{\beta}^T\ \boldsymbol{b}^T]^T$ and we can treat $\boldsymbol{\beta}$ and $\boldsymbol{b}$ jointly.

When the likelihood is Gaussian and given $\boldsymbol{x} \sim N(0, \boldsymbol{Q}(\theta))$, the distribution for $\boldsymbol{x}$ after observing $\boldsymbol{y}$ is also Gaussian with precision equal to $\boldsymbol{Q}_{x|y}(\theta) = \boldsymbol{Q}(\theta) + \tau_y\boldsymbol{A}^T\boldsymbol{A}$. The expected value of $\boldsymbol{x}$ given the data, $\boldsymbol{\mu}_{x|y}$, is obtained by solving $\boldsymbol{Q}_{x|y}\boldsymbol{\mu}_{x|y} = \boldsymbol{y}$. The key ingredient to have computationally efficient inference for $\boldsymbol{x}$ is for $\boldsymbol{Q}_{x|y}(\theta)$, not only $\boldsymbol{Q}(\theta)$ to be sparse. When $\boldsymbol{y}|\boldsymbol{\eta}$ is not Gaussian, the Integrated Nested Laplace Approximation - INLA, Rue et al. (2009), can be used taking the benefits of the sparsity on $\boldsymbol{Q}_{x|y}(\theta)$. The implementation in **INLA** package also includes $\boldsymbol{\eta}$ in $\boldsymbol{x}$ as well as adding a $\tau_\eta\boldsymbol{I}$ matrix in $\boldsymbol{Q}(\theta)$ as well, where τ_η is a fixed big number, Rue et al. (2017).

This modelling approach is powerful as we can consider several kinds of stochastic models for $\boldsymbol{b}$ and for β as well. This can be seen through the contributions described in this thesis. We divided these contributions into four chapters and summarise them under the following headings:

1. **Non-separable space-time models**: In this chapter we consider models based on non-separable Stochastic Partial Differential Equations (SPDE), and illustrate its marginal properties and Gaussian Markov representations;

2. **Computational trick**: In this chapter, we propose two approaches to avoid the computationally expensive model constraints when fitting space-time intrinsic interaction models;

3. **Applications of Bayesian hierarchical models**: This chapter details several projects where we have analysed data in partnership with researchers from several diverse research fields. These collaborations have led to a number of co-authored papers as well.

4. **How to fit**: In the final chapter, we present the educational contribution of this work in the form of a tutorial on the SPDE based models and a book chapter.

We introduce each chapter in the next four sections.

1.1 Introduction to Part I

Space-time statistical modeling is a research field with several areas of ongoing research detailed in recent books about space-time statistical models. Finkenstadt et al. (2006) present a collection of work on models for point process, geostatistics and areal data. Cressie and Wikle (2011) focus on the dynamic geostatistical models and Montero et al. (2015) focus on geostatistical models. We review some of this work in the beginning of chapter 2. This chapter also includes an intuitive introduction to the SPDE approach and the GMRF representation from Lindgren et al. (2011) as a background for the other chapters.

The use of SPDE's for modeling data is not a new approach. There are examples in the time series literature and in Cressie and Wikle (2011) for the space-time case. Since the spectral density can be obtained from an SPDE, it can be considered directly to analyse data, as in Fuentes (2007). For the space-time case there is the approach in Sigrist et al. (2015). An important fact is that the solution of a linear SPDE is a Markov random field, Rozanov (1977). This is the case for fields whose spectral density reciproca is a polynomial function of the frequencies.

The Matérn covariance function is probably the most used one in geostatistics. Therefore, the approach in Lindgren et al. (2011) was an impetus for using SPDE's in spatial statistics because it gives the explicit link between a linear SPDE and the Matérn covariance function. Additionally, it was important from a practical point of view, because the Finite Element Method - FEM was considered to build a GMRF representation and that it was made available for practitioners in the **INLA** package, Rue et al. (2017).

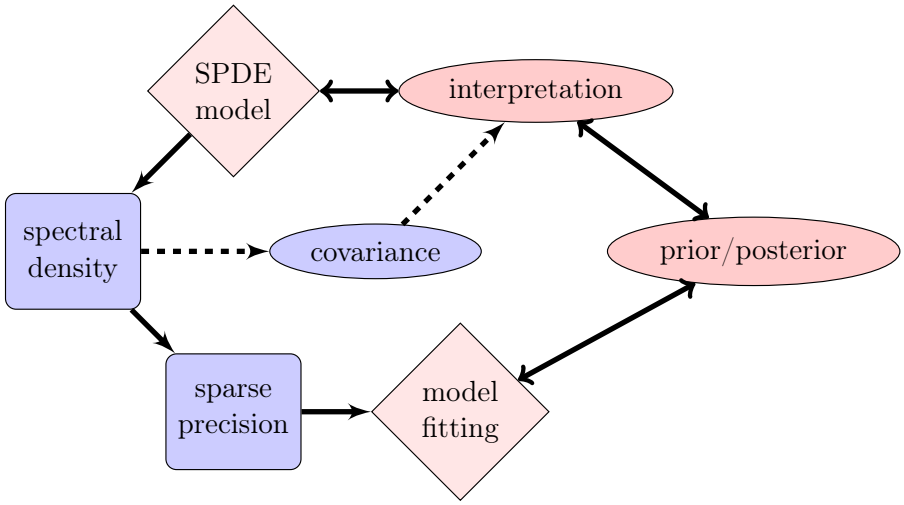


Figure 1.1: Work flow with the SPDE model formulation.

Lindgren et al. (2011) have established an approach to work with SPDE based models that can be schematically represented in Figure 1.1.

The SPDE itself has parameters that are related to the local properties of the model which are translated into the conditional distributions in the GMRF representation. One may find the physical interpretation from the SPDE enough and build the *prior* distributions for the model parameters. Others may need to see the marginal properties of the model. In this case the covariance can be computed from the spectral density even though it is not actually needed for the modeling.

Although there is no need to compute the covariance when working with SPDE's to fit the model for data, understanding the marginal properties of the model is important as they provide additional information about the process. The covariance can be computed from the spectral density and we consider it in chapter 5. We consider the spectral density of the non-separable models and the spatial marginal and temporal marginal densities in order to compute the space-time covariance and marginal spatial and temporal ones as well. The marginal properties of the models are them considered to build a map to the model parameters. This work is also important to guide practitioners to in prior elicitation whe fitting the model for real data.

We consider the simplest space-time model one can derive using an SPDE in chapter 3. We consider the heat equation model and provide the statistical background including work done on the marginal properties of this model.

We then review the GMRF representation based on the finite volume method and provide a representation based on the finite element method. We show how these representations work in a scenario.

In the chapter 4 we consider an model proposed in Stein (2005). This model was purported to have good properties, however it failed on having an easy to compute covariance. Thus, we propose an SPDE that gives the same spectral density and providing a GMRF representation we can overcome this issue.

Chapter 5 is on building models from an iterated version of the heat equation with spatially correlated driving noise. The marginal properties of the models were considered as having an implied covariance function. We consider the sphere as the spatial domain for the intrinsically stationary and the damped versions. The damped is considered also in $\mathbb{R}^2$. We show how the marginal variance, marginal spatial correlation and marginal temporal correlation are driven by the model parameters.

1.2 Introduction to Part 2

There are some models for modeling space-time phenomena where the spatial setting is a set of non-intersecting areas and the time setting is also discrete. The easiest way to model this case is to consider an interaction between a spatial model and a temporal one. One approach is to consider an interaction between a spatial model and a first order autoregression over time Martínez-Beneito et al. (2008), Vivar and Ferreira (2009), Rushworth et al. (2014), Ruiz-Cárdenas et al. (2012) and Blangiardo and Cameletti (2015). However, none of these studies pay attention to the main spatial or temporal effect.

Knorr-Held (2000) considered a model where the main effects are considered and proposed four interaction cases for the space-time part. However, three of such cases need to have sum-to-zero constraints that may become computationally demanding when the number of areas or time steps increase. We revisit these models and propose a way to avoid these linear constraints in order to be able to fit these models even when the number of areas or time steps are large.

1.3 Introduction to Part 3

The third part of this thesis introduces collaborative work involving applications of the Bayesian hierarchical models to practical problems conducted during the PhD. Three were on the analysis of fisheries bycatch data, one on fishery occurrence modeling, four on the analysis of old-age survival data,

one on tuberculosis disease mapping, one on joint modeling of the prevelance of wasting, stunting, and underweight in children in Somalia and one about the space-time relation between Twitter data and violent crime in Amsterdam.

Fisheries bycatch

A set of three pieces of work are introduced in Section 6.1, these are three out of seven chapters in a PhD titled "Elasmobranch Bycatch in the Canadian Northwest Atlantic and Arctic Adjacent Seas: Composition, Biogeography, and Mitigation" defended by Aurelie Cosandey-Godin on March, 12 2015 at the department of Biology, Dalhousie University, Halifax - Canada. One of these chapters was published in the *Canadian Jornal of Fisheries and Aquatic Sciences* in 2014, see Cosandey-Godin et al. (2014).

This thesis focused on the issue of bycatch or non-target catch in Canadian commercial fisheries. In Canada, bycatch data is collected by at-sea observer programs who provide information on the number (or weight) of non-target species for a certain percentage of the fishing effort. These data are spatially referenced by fishing location and typically also include information on the target species, fishing methods, and some environmental variables such as depth and water temperature.

My main contribution to these chapters was to develop and apply space-time models to the study of bycatch and to identify bycatch hotspots for conservation purposes using the INLA/SPDE approach. These statistical tools proved to be an efficient and powerful means to analyse, predict, and evaluate management strategies to mitigate fisheries bycatch.

Fishery occurrence

The paper about collective memories and fish distribution shape, Macdonald et al. (2017) will be part of the thesis work by Jed I. MacDonald at the Faculty of Life and Environmental Science, University of Iceland, Reykjavík, Iceland. In this paper a series of Bayesian space-time occurrence models were considered to investigate wintering dynamics over 23 years, using point-referenced fishery and survey records from Icelandic waters.

Old-aged survival and Tuberculosis

Four studies are introduced in Section 6.3 for which my contribution was to set, describe and fit Bayesian hierarchical models for the analysis of old-aged survival data. These were four out of nine papers in the PhD work

entitled "Long-lived lives: The role of the contextual determinants" done by Ana Isabel Correia Ribeiro. She defended her thesis on April, 4 2016, at the University of Porto, Porto - Portugal, for a PhD degree in Public Health.

Similar work introduced in Section 6.4 was done for the analysis of Tuberculosis data.

Wasting, stunting and underweight

In the work presented in section 6.5, my contribution was to develop a way to do joint modeling of the prevalence of wasting, stunting and underweight among children in Somalia. The statistical model considered three spatial continuous random fields shared across the linear predictor of each one of the three outcomes. This work is part of the PhD thesis of Damaris K Kinyoki who defended her thesis in May, 2017 which is publicly available at the url `http://wrap.warwick.ac.uk/93209/`. The paper is published in the BMJ Open journal, Kinyoki et al. (2016), and is free available at the url `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4785320/`.

Twitter data and violent cime

In this work presented in section 6.6 a joint space-time model of tweets and crime data was developed to study relationships between the two datasets. The data comprised tweets collected containing some key words and the violent crime data both in Amsterdam. These data were geo-localized and its exact time annotated as well. The joint model developed had a shared space-time Gaussian field when considering the log-Gaussian Cox process. This work was published in the Preproceedings of the 29th Benelux Conference on Artificial Intelligence - BNAIC 2017, Stam et al. (2017).

1.4 Introduction to Part 4

It is very nice to propose models with good properties. However, it is useful to provide the code in order to make the application of such models in practice straightforward. Additionally, tutorials illustrating such applications can be seen as neccessary to complement new methodology. This is true as the majority of practitioners are not able to do all of the work alone by only looking into the methodological papers. Therefore, we as a field can work together to present examples of how interesting new methods can be used in practice. To complete the circle, it is common that practitioners using an implementation ask questions and propose new analytical challenges that guide future developments.

The last part of this thesis is a tentative step in the direction of providing material for educating researchers and practitioners to use SPDE based models. We started by working with the basic model and some extensions to provide some examples of possible applications. Not long after that we were facing specific questions from practitioners that motivated new examples.

This section presents two pieces of work. The first is the "The R-INLA tutorial on SPDE models". It consists of a collection of examples on how to fit models that contains an SPDE based random effect. The second is a book chapter on some of the more complex models for space or space-time data, from spatially misaligned data (Gaussian or not), a hurdle gamma model, dynamic models and lowering the model time resolution on a space-time model.

This was some work that required a significant amount of effort. An effort was given to make the code as simple as possible and to illustrate one point at a time. Also, we had to make sure that every detail was made clear, especially for the book chapter. This book chapter is included at the end of this part.

Part I

Non-separable space-time models

Chapter 2

Space-time modeling introduction

In this chapter we briefly review the literature on space-time statistical models in Section 2.1.1. We then provide an intuitive background on the methodology that is basis for the work on the next chapters in this first part of the thesis.

2.1 Literature review in space-time models

Space-time modeling is an area of ongoing research. Even though there are several available statistical models in the literature, there is still room for new developtments, for example for modeling over the sphere, Porcu et al. (2017). We briefly review the developments made so far in this section split into separable and non-separable space-time models.

2.1.1 Separable space-time models

This approach is the first step to move from one to two dimensional models for the space-time models. Once a temporal and spatial model is defined, then separable space-time models are easy to obtain. The spectral density for a separable space-time process can be written as a product between a purely temporal spectral density and a purely spatial spectral density. In the kernel convolution approach, one can use a kernel that splits into a purely temporal smoother times a purely spatial one Higdon (2002). When one considers covariance, the Kronecker product between a temporal correlation matrix and a spatial covariance matrix defines a separable space-time

model at a finite set of space-time locations. This approach has an appealing computational advantage from the Kronecker product properties for factorizations and determinants and is considered in several softwares.

There is a popular way to consider dynamic models as

$$\boldsymbol{\eta}_t = \boldsymbol{M}\boldsymbol{\eta}_{t-1} + \boldsymbol{w}_t$$

where $\boldsymbol{M}$ defines the temporal evolution structure of the model and $\boldsymbol{w}_t$ is a spatially structured innovation error vector. One widely used inference approach considers a model based on covariance for the innovation error and computations are done using the Kalman filter/smoother algorithm or other modifications Cressie and Wikle (2011). Some non-separable dynamic models linked to non-separable SPDEs along with discretizations using the finite differences approach can be fitted considering modified Kalman filter based algorithms, Cressie and Wikle (2011).

The one and two dimensional SPDE's can be considered to build separable space-time SPDE based models and can be used as an alternative way to define some space-time models. The idea is that one can use a one-dimensional SPDE to describe the temporal evolution and another SPDE to model the driving noise as spatially correlated. Therefore we have an SPDE with partial derivatives with respect to time and driving white noise in time correlated over space modeled considering another SPDE with partial derivatives taken over space, Jones and Zhang (1997). That is a combination of the models presented in previous sections,

$$(\frac{\delta}{\delta t} + \varphi)u(\boldsymbol{s}, \delta t) = \xi(\boldsymbol{s}, \delta t) \tag{2.1}$$

$$\tau(\kappa^2 - \Delta)^{\alpha/2}\xi(\boldsymbol{s}, \delta t) = \mathcal{W}(\boldsymbol{s}, \delta t)\ . \tag{2.2}$$

The FEM approach can be considered for a discretization of $u(\boldsymbol{s}, t)$ over both time and space, Lindgren and Rue (2015). The temporal process at each spatial location discretized as in Section 2.3.1 and the the spatial process at a time as in Section 2.3.2. This gives a precision matrix that is the Kronecker product $(\varphi\boldsymbol{D} + \boldsymbol{H}) \otimes \boldsymbol{Q}_\alpha(\tau, \kappa)$. This model has direct connection with dynamic models and the single parameter evolution matrix, $\boldsymbol{M} = \rho\boldsymbol{I}$, for $\rho = \mathrm{e}^{-\varphi\delta t}$, and the precision of the entire vector is the Kronecker product between the precision for the AR1 model and the precision for $\boldsymbol{w}$, Cameletti et al. (2012).

2.1.2 Non-separable space-time models

Non-separable models can be built considering any of the five ways we have listed in the beginning. Basically, non-separable space-time models are those

having covariance that can not be written as a Kronecker product between the temporal correlation and the spatial covariance. Similarly, the spectral density of non-separable space time models cannot be written as a product of a purely temporal spectral density and a purely spatial one.

When considering kernel convolution approach, non-separable space time models are those where the kernel function does not splits into a purely temporal kernel function times a purely spatial kernel function. That is, the flexibility lies on the definition of the kernel function.

Most of the approaches for space-time modeling considers the covariance directly, Cressie and Huang (1999), Gneiting (2002), Fonseca and Steel (2011). However, the main limitation of working with this approach is that it is simply too computational demanding as the non-separable models do not factorize into smaller blocks of matrices. One common approach is to consider low rank approximations, Rodrigues and Diggle (2010).

Stein (2013) defined three desired properties for a space-time model

1. have “any degree of smoothness in space and any (possibly different) degree of smoothness in time";
2. “be smoother away from the origin than it is at the origin"; and
3. covariance “computed accurately and efficiently".

The separable models fail to fulfill the second requirement and the Stein's model only follows when α_1 and α_2 are integers.

However, he also wrote that

> “As best as I am aware, no existing class of generalized covariance functions satisfies all three of these requirements."

This quote is motivated by the fact that his model class proposed in Stein (2005), and the particular case in Stein (2013), fail to have easy to compute covariances.

One may think how important are these three desired properties. The first desired property is related to one kind of model flexibility, the smoothness. However, smoothness is one property that is hard to infer from real data. Usually a space-time model is just one component of the model for the observed data model and frequently an independent noise is also present. Although it would be important to consider the correct smoothness of the process in order to have the other terms of the models not affected. For example, a smoother process plus noise may be equivalent to a less smoother process plus a smaller nugget effect. An additional point is when considering parameter uncertainty where the actual result would be a mixture of predictions weighted by each parameter value density.

The second property is related to high frequency properties. Again another term in the data model will be required in order for the model to be purely identified. However, when the aim is just to capture the "big picture" of the space-time variation of the phenomena the high resolution would not be so important. This is also the case when dealing with very large amounts of data.

The third desired property is not important when considering fitting procedures that do not need the covariance to be computed. However, we do agree that it is important to seek for a computationally efficient fitting process. Towards this goal, the GMRF representation is an important approach.

In this work we consider models that are built based on stochastic differential equations - SDE. We review some models that are already presented in the literature and build a discretization that gives a Gaussian Markov Random Field - GMRF representation. That is we start with the stochastic heat equation in chapter 3 for the case when the space is one-dimensional. An iterated version is then considerd in chapter 4 as a SPDE approach for the model class proposed in Stein (2005). The case with spatially correlated driving noise is considered is presented in chapter 5.

2.2 Stochastic modeling ways

Let a stochastic process $u(\boldsymbol{l})$ that evolves over a domain indexed by $\boldsymbol{l}$. We can have $\boldsymbol{l} = t : t$ for time $\boldsymbol{l} = \boldsymbol{s} : \boldsymbol{s}$ for space or $\boldsymbol{l} = (\boldsymbol{s}, t)$ for space-time. There are at least five ways to define a stochastic process, listed below.

1. **Covariance**: In this case, the covariance matrix for a finite set of locations $\boldsymbol{l}_1, ..., \boldsymbol{l}_n$ is

$$V(\boldsymbol{l}, \boldsymbol{l}') = \text{Cov}(u(\boldsymbol{l}), u(\boldsymbol{l}')) \tag{2.3}$$

   and should be positive definite. A simple covariance function is

$$V(\boldsymbol{l}, \boldsymbol{l}') = \sigma_u^2 \text{e}^{\kappa ||\boldsymbol{l} - \boldsymbol{l}'||}$$

   where $||\boldsymbol{l} - \boldsymbol{l}'\ ||$ is the Euclidean distance, σ_u^2 is the marginal variance and κ is a scale parameter. For a review on covariance functions see Abrahamsen (1997).

2. **Spectral density**: The process is specified as

$$u(\boldsymbol{l}) = \int_{-\infty}^{\infty} \text{e}^{i\boldsymbol{w}\boldsymbol{l}} dZ(\boldsymbol{w}), \tag{2.4}$$

where $Z(\boldsymbol{w})$ is a stochastic spectral process with spectral distribution $F(\boldsymbol{w})$, Lindgren (2013). Also, $F(\boldsymbol{w}) = \int_{s=-\boldsymbol{\infty}}^{\boldsymbol{w}} f(\boldsymbol{s})\partial \boldsymbol{s}$ where $f(\boldsymbol{w}) \geq 0$ is a spectral density. The Bochner's theorem states the following link to the covariance function:

$$V(\boldsymbol{h}) = \int \mathrm{e}^{i\boldsymbol{w}\boldsymbol{h}} dF(\boldsymbol{w}). \tag{2.5}$$

When considering discrete spectra and frequencies ω_k, a one dimensional process can be written as a sum of cosine functions

$$u(t) = \sum_{k=1}^{N} A_k cos(\omega_k t + \phi_k)$$

where A_k and ϕ_k are random amplitudes and phases, respectively.

3. **Kernel convolution**: In this case the process is a smoothed/filtered noise, written as

$$u(\boldsymbol{l}) = \int k(\boldsymbol{l} - \boldsymbol{u})\mathcal{W}(\boldsymbol{l})d\boldsymbol{u} \tag{2.6}$$

where $\mathcal{W}(.)$ is a white noise, $k(.) \geq 0$ is a kernel function such that $\int k(\boldsymbol{u})d\boldsymbol{u} < \infty$ and $\int k^2(\boldsymbol{u})d\boldsymbol{u} < \infty$, Higdon (2002). In this case

$$V(\boldsymbol{l}, \boldsymbol{l}') = \int_{\boldsymbol{u}} k(\boldsymbol{l} - \boldsymbol{u})k(\boldsymbol{l}' - \boldsymbol{u})\partial \boldsymbol{u}. \tag{2.7}$$

4. **Stochastic partial differential equations** (SPDE) or stochastic ordinary differential equations for the unidimensional case, are stochastic versions of differential equations. Consider the following SPDE

$$\mathcal{L}_1 u(\boldsymbol{l}) = \mathcal{L}_2 \mathcal{W}(\boldsymbol{l}) \tag{2.8}$$

where $\mathcal{L}_1$ and $\mathcal{L}_2$ are differential operators and $\mathcal{W}(.)$ is a white noise process. Replacing $\mathcal{L}_1$ by a constant, it becomes similar to the kernel convolution approach, however, with a very particular kernel function. The autoregressive moving average model class in the one dimensional case is another particular case. The ARMA process can be written in this way. An example in the spatial case is in Bolin and Lindgren (2011). Our interest is when $\mathcal{L}_2$ is replaced by a constant. In this case, the process is Markovian, Rozanov (1977).

5. **Conditional distributions** are a more statistical sound approach and are usually considered in discrete settings, either for discrete time or for

areal (spatial) data. A Gaussian model of this class is called a Gaussian Markov Random Field - GMRF which inherits sparse precision matrices and allows for efficient computations, Rue and Held (2005). However, there are continuous domain processes that are direct linked to GMRF's, Lindgren et al. (2011), and also approximations, Rue and Tjelmeland (2001) and Datta et al. (2016).

A process is usually considered by its marginal properties and local properties. The covariance is a marginal property of the model while there are conditional properties. The kernel convolution approach is closely related to the covariance, Eq. 2.7. The SPDE and conditional approach consider the dynamics of the process since it considers local behaviour of the process. The spectral density approach has connection with the marginal properties through Bochner's theorem, Eq. 2.5. However it also give information about the process dynamics. The power spectra of an SPDE can be obtained and one can use the Bochner's theorem to compute the covariance. But, there is not always an analytical solution for the integrals involved.

When the reciprocal of a spectral density is a polynomial the process is Markov, which is the case when $\mathcal{L}_2$ in Eq. 2.8 is replaced by a constant, Rozanov (1977). A Markov process can be discretized into a GMRF that is represented by a set of conditional distributions, Rue and Held (2005). This allows for fast computations which will be explored in this work.

2.3 GMRF representations for 1D and 2D SPDE

We briefly review the SPDE apprach in the following sections.

2.3.1 One dimensional case

We start with a one dimensional model that is usually defined from a difference equation and will provide a link with a stochastic ordinary differential equation. The first order auto-regressive model, named AR1 model, can be defined from the following equation

$$u_t - \rho u_{t-1} = \sigma w_t \tag{2.9}$$

where $w_t \sim N(0, 1)$.

The marginal distribution for this process at a finite set of time knots, $\boldsymbol{u} = \{u_1, ..., u_n\}$ can be obtained. Considering $u_1 \sim N(0, \sigma^2/(1-\rho^2))$, the stationary distribution, and $u_t \sim N(\rho u_{t-1}, \sigma^2)$ we have

$$\pi(\boldsymbol{u}) = (2\pi\sigma^2)^{-n/2}\sqrt{1-\rho^2}\,\exp^{-\frac{1}{2\sigma^2}[(1-\rho^2)u_1^2+(u_2-\rho u_1)^2+...+(u_n-\rho u_{n-1})^2]}\ . \tag{2.10}$$

The sum inside the exponential can be written in matrix form as

$$\begin{aligned} & u_1^2 - \rho^2 u_1^2 + u_2^2 - 2\rho u_1 u_2 + \rho^2 u_1^2 + ... + u_n^2 - 2\rho u_{n-1} u_n + \rho^2 u_{n-1}^2 \\ = & \sum_{i=1}^{n} u_i^2 - 2\rho \sum_{i=2}^{n} u_i u_{i-1} + \rho^2 \sum_{i=2}^{n-1} u_i^2 \\ = & \boldsymbol{u}^T [\mathcal{D}_n^{(1,1+\rho^2,1)} + \mathcal{T}_n^{(0,-\rho)}] \boldsymbol{u} \end{aligned} \tag{2.11}$$

where $\mathcal{D}_n^{(a,b,c)}$ defines an n dimensional diagonal matrix with a being the first element, c the last element and all elements as b. The matrix $\mathcal{T}_n^{(a,b)}$ is a tridiagonal matrix with all the elements in the diagonal equal to a and all the other non-zero elements equal to b, represented as

$$\mathcal{D}_n^{(a,b,c)} = \begin{bmatrix} a & & & & \\ & b & & & \\ & & \ddots & & \\ & & & b & \\ & & & & c \end{bmatrix}_{n \times n} \quad \text{and} \quad \mathcal{T}_n^{(a,b)} = \begin{bmatrix} a & b & & & & \\ b & a & b & & & \\ & & \ddots & & & \\ & & & b & a & b \\ & & & & b & a \end{bmatrix}_{n \times n} . \tag{2.12}$$

One benefit of dealing with a sparse matrix comes from taking sparsity into account when doing the Cholesky factorization and using it to compute the determinant. An interesting fact in this parametrization for the AR1 model is that it does not even need the Cholesky factorization as we have that $|\mathcal{D}_n^{(1,1+\rho^2,1)} + \mathcal{T}_n^{(0,-\rho)}| = 1 - \rho^2$, see Eq. (2.10).

The model in Eq. 2.9 can be considered in the continuous case as

$$\tau \left[\varphi + \frac{\partial}{\partial t} \right] u(\delta t) = \mathcal{W}(\delta t) \tag{2.13}$$

where φ is a scaling parameter, τ is a local precision parameter, δt is the time frame and $\mathcal{W}(\delta t)$ is a continuous unit variance white noise. For details, see chapter 3 of Jones (1993). The corresponding parametrization for the correlation at the distance δt is $\mathrm{e}^{-\varphi \delta t}$, the exponential correlation function. If this continuous process is sampled at equally spaced intervals $t_1, t_2, ...,$ with $t_2 - t_1 = \delta t$, then $\rho = \mathrm{e}^{-\varphi \delta t}$. If $\delta t = 1$ we have $\rho = \mathrm{e}^{-\varphi}$.

The following SPDE

$$\tau \left(\varphi^2 - \frac{\partial^2}{\partial t^2} \right)^{1/2} u(\delta t) = \mathcal{W}(\delta t) \tag{2.14}$$

actually corresponds to the same model when the driving noise is Gaussian. This can be seen when considering the spectral representation for $\frac{d}{dt}$ as $i\omega$

and for $\frac{d^2}{dt^2}$ as $-\omega^2$. Therefore we have $\mathcal{F}(\frac{\partial}{\partial t}) = |i\omega| = |\omega|$ and $\mathcal{F}(\frac{\partial^2}{\partial t^2}) = |-\omega^2| = |\omega|^2$ which gives the same spectral density as

$$f(\omega) = (\tau 2\pi(\varphi^2 + \omega^2))^{-1} . \tag{2.15}$$

Therefore we can follow the discretization approach in Lindgren et al. (2011). Let

$$\Phi(t) = [\phi_1(t), ..., \phi_{N_\phi}(t)]$$

be the temporal basis functions. Then define

$$\boldsymbol{H}^{(0)} = \boldsymbol{D} = \langle \Phi, \Phi \rangle, \ \boldsymbol{H} = \langle \nabla\Phi, \nabla\Phi \rangle \text{ and } \boldsymbol{H}^{(k)} = \boldsymbol{D}^{-1}\boldsymbol{H}^{(k-1)}. \tag{2.16}$$

This results in the joint distribution at the N_ϕ set of time knots being Gaussian with a precision matrix equal to $\tau(\varphi^2\boldsymbol{D} + \boldsymbol{H})$.

Let $0 = u_0 < u_1 < ... < u_n < u_{n+1} = T$ a set of (ordered) knots spanning the time domain $(0, T)$. We can look at the case when piecewise linear basis functions are considered and equally spaced knots such that $a = u_i - u_{i-1}$ for $i = 0, ..., n+1$. With Dirichlet boundary condition, the distribution is given at $u_1, ..., u_n$ with $\boldsymbol{D} = a\boldsymbol{I}_n$, for $\boldsymbol{I}_n$ the identity matrix of dimension n, and $\boldsymbol{H} = \mathcal{T}_n^{(2a,-a)}$. Using Neumann boundary condition, the distribution is given at $u_0, ..., u_{n+1}$ and considering these $m = n+2$ time knots, $\boldsymbol{D} = \mathcal{D}_m^{(a/2,a,a/2)}$ and $\boldsymbol{H} = \mathcal{D}_m^{(a,2a,a)} + \mathcal{T}_m^{(0,-a)}$.

An interesting particular case is the limit case when $\rho = 1$, which corresponds to having $\varphi = 0$. This is then the first order random walk, Rue and Held (2005). When considering the Neumann boundary condition and $a = 1$, we have the precision matrix as $\tau\boldsymbol{H}$ and Eq. (2.11) becomes equal to $\tau\boldsymbol{u}^T\boldsymbol{H}\boldsymbol{u}$. A computational detail in this case is that the determinant of $\boldsymbol{H}$ can be avoided in most of the computations.

2.3.2 Two dimensional case

A simple model in two dimensions is the Laplace equation, where the second derivative of the process with respect to each coordinate is zero as in

$$\left(\frac{\partial^2}{\partial s_1} + \frac{\partial^2}{\partial s_2}\right) u(\boldsymbol{s}) = 0 \quad \text{or} \quad \nabla^2 u(\boldsymbol{s}) = 0 \quad \text{or} \quad \Delta u(\boldsymbol{s}) = 0$$

where $\nabla^2 = \Delta$ is the Laplace operator or Laplacian. One may understand it better considering a numeric approximation for the Laplacian, as

$$\Delta f(x, y) \approx \frac{f(x-h, y) + f(x+h, y) + f(x, y-h) + f(x, y+h) - 4f(x, y)}{h^2}$$

where the value at a point is contrasted with the average over the neighborhood allowing a local second order degree curvature.

A stochastic version was considered in Whittle (1954) in the context of lattice data named as Simultaneous Autoregressions - SAR model. A general version was considered in Lindgren et al. (2011) as

$$\tau(\kappa^2 - \Delta)^{\alpha/2} u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \tag{2.17}$$

where κ is a scale parameter and α is a smoothness parameter. When $\mathcal{W}(\boldsymbol{s})$ is a Gaussian white noise process, it was shown in Lindgren et al. (2011) that the solution for this SPDE is a Gaussian random field with Matérn covariance, Matérn (1960). The case when $\mathcal{W}(\boldsymbol{s})$ is non-Gaussian was considered in Bolin (2014) and Wallin and Bolin (2015).

Lindgren et al. (2011) proposed the use of the Finite Element Method (FEM) approach. All the details are in the Appendices of this paper. We will summarize it very briefly here. Let

$$\Psi(\boldsymbol{s}) = [\psi_1(\boldsymbol{s}), ..., \psi_{N_\psi}(\boldsymbol{s})]$$

be a a set of N_ψ basis functions. Consider the following matrices

$$\tilde{\boldsymbol{C}} = \langle \Psi, 1 \rangle, \quad \boldsymbol{C} = \langle \Psi, \Psi \rangle, \quad \boldsymbol{G} = \langle \nabla\Psi, \nabla\Psi \rangle \tag{2.18}$$

and $\boldsymbol{G}^{(k)} = \boldsymbol{G}(\boldsymbol{C}^{-1}\boldsymbol{G})^{(k-1)}$, for $k = 1, 2, ...$ and set $\boldsymbol{G}^{(0)} = \boldsymbol{C}$. The appendix A.2 of Lindgren et al. (2011) contains the expressions for such matrices for the case of $\Psi(\boldsymbol{s})$, considered as piecewise linear functions at the nodes of a mesh triangulation. Given τ, κ and α, the precision matrix $\boldsymbol{Q}_\alpha(\tau, \kappa)$ at the set of N_ψ mesh nodes is

$$\begin{aligned}
\boldsymbol{Q}_1(\tau, \kappa^2) &= \tau \boldsymbol{K}_{\kappa^2} = \kappa^2 \boldsymbol{C} + \boldsymbol{G} && (2.19)\\
\boldsymbol{Q}_2(\tau, \kappa^2) &= \tau \boldsymbol{K}_{\kappa^2} \boldsymbol{C}^{-1} \boldsymbol{K}_{\kappa^2} = \tau[\kappa^4 \boldsymbol{C} + 2\kappa^2 \boldsymbol{G} + \boldsymbol{G}^{(2)}] && (2.20)\\
\boldsymbol{Q}_\alpha(\tau, \kappa^2) &= \tau \boldsymbol{K}_{\kappa^2} \boldsymbol{C}^{-1} \boldsymbol{Q}_{\alpha-2}(1, \kappa^2) \boldsymbol{C}^{-1} \boldsymbol{K}_{\kappa^2} \text{ for } \alpha = 3, 4, ... && (2.21)
\end{aligned}$$

The power of this result is that the structure of the precision matrix only depends on α. This is convenient when doing computations with numerical linear algebra for sparse matrices as it implies that the reordering does not change.

The $\alpha = 2$ case corresponds to the considered elementary model in Whittle (1954). The $\alpha = 1$ case corresponds to the model based on conditional autoregressions - CAR, proposed in Besag (1974) for discrete domains and has been widely used. The $\kappa = 0$ case still has valid random measures, Appendix C.3 in Lindgren et al. (2011), and for $\alpha = 1$ and $\kappa = 0$ it corresponds to the Wijs process, Besag and Mondal (2005).

Another appealing point is that a bigger α does not correspond to a smoother process as it is directly related to the smoothness parameter in the Matérn covariance function. In the precision term, we have it consider a wider neighborhood for bigger α. The choice of α between 1 or 2 is equivalent to choosing between CAR or SAR in the discrete case.

When α is not an integer, the field is no longer a Markov field, Rozanov (1977). This limitation on integer α can be bypassed by considering the approximation to the best Markov approximation, suggested in the reply to the original paper, Lindgren et al. (2011). Also, recently it was proposed to combine the discretization using FEM in space with a rational approximation, Bolin and Kirchner (2017). Yet another recent proposal is to consider FEM and a quadrature approximation of an integral representation, Bolin et al. (2017a) and Bolin et al. (2017b).

Chapter 3

Stochastic heat equation and GMRF representations

In this chapter we consider space-time non-separable models derived from the heat equation. We will start with a simple heat equation model and its damped version, which is one of those considered in Heine (1955), and consider a variation that allows for higher spatial dimensions as in Jones and Zhang (1997). A Gaussian Markov representation is provided.

3.1 The stochastic heat equation and variations

The heat equation describes how heat spreads over space and time. It can be written using the following homogeneous equation

$$\frac{\partial u(\mathbf{s},t)}{\partial t} = \gamma \Delta u(\mathbf{s},t) \tag{3.1}$$

where $u(\mathbf{s},t)$ is the temperature at the d-dimensional location $\mathbf{s}$ and time t, Δ is the Laplace operator (Laplacian) and γ is the thermal diffusivity parameter. This parameter expresses how fast the heat spreads over space. An alternative parametrization can be made considering

$$\rho \frac{\partial u(\mathbf{s},t)}{\partial t} = \Delta u(\mathbf{s},t) \tag{3.2}$$

where $\rho = 1/\gamma$ and is a persistency parameter over time.

The stochastic version of the heat equation can be written as

$$\tau \left[\rho \frac{\partial}{\partial t} - \Delta \right] u(\mathbf{s},t) = \mathcal{W}(\mathbf{s},t) \tag{3.3}$$

where $\mathcal{W}(\mathbf{s},t)$ is a space-time white noise term with unit variance and τ is a local precision parameter that controls the size of the difference between $\rho\frac{\partial}{\partial t}u(\boldsymbol{s},t)$ and $\Delta u(\boldsymbol{s},t)$ when $\tau\to\infty$ Eq. (3.3) approaches Eq. (3.1).

This model is only for the variation over time and space and the properties do not change with the addition of a constant to the process. The variation over time and space are not separable and one can only modulate how much comes from time and space by the parameter γ (or ρ). That is, the stochastic heat equation defines an intrinsic random field.

Considering $\lambda\in\mathbb{R}^d$ the spatial frequency and $\omega\in\mathbb{R}$ the temporal frequency the spectral density can be written as

$$f_{\mathrm{h}}(\lambda,\omega)=\frac{1}{(2\pi)^{d+1}\tau(\rho^2\omega^2+\|\lambda\|^4)} \tag{3.4}$$

which has an intrinsicness point at $f(0,0)$. Working with intrinsic models in real applications is usually considered to add a valid linear constraint. For example, a sum-to-zero linear constraint is adequate as we only have a point of intrinsicness. We will consider two GMRF representations for this model and the empirical covariance under a constraint in Section 3.2.

A damped version of this model can be written as

$$\tau\left[\rho\frac{\partial}{\partial t}-\Delta+\kappa^2\right]u(\mathbf{s},t)=\mathcal{W}(s,t) \tag{3.5}$$

and the spectral density can be written as

$$f_{\mathrm{hd}}=\frac{1}{(2\pi)^{d+1}\tau[\rho^2\omega^2+(\kappa^2+\|\lambda\|^2)^2]} \tag{3.6}$$

which is finite as $\kappa>0$.

This model was considered in Heine (1955) for $d=1$. The space-time correlation at time lag l and spatial distance r is

$$\mathrm{corr}(s,t)=\frac{1}{2}\left[\mathrm{e}^{-r^*}\mathrm{Erfc}(l^*-\frac{r^*}{2\sqrt{l^*}})+\mathrm{e}^{r^*}\mathrm{Erfc}(l^*+\frac{r^*}{2\sqrt{l^*}})\right] \tag{3.7}$$

where $l^*=l\kappa^2/\rho$ and $r^*=\kappa r$ and Erfc is the error function that can be written as

$$\mathrm{Erfc}(x)=\frac{2}{\sqrt{\pi}}\int_0 x\mathrm{e}^{-t^2}\partial t$$

and is more common in statistics books. The model decays faster as κ increases or ρ decreases, as shown in Figure 3.1. The impact of κ on the decay rate is bigger than the impact of ρ. This is evident when we compare

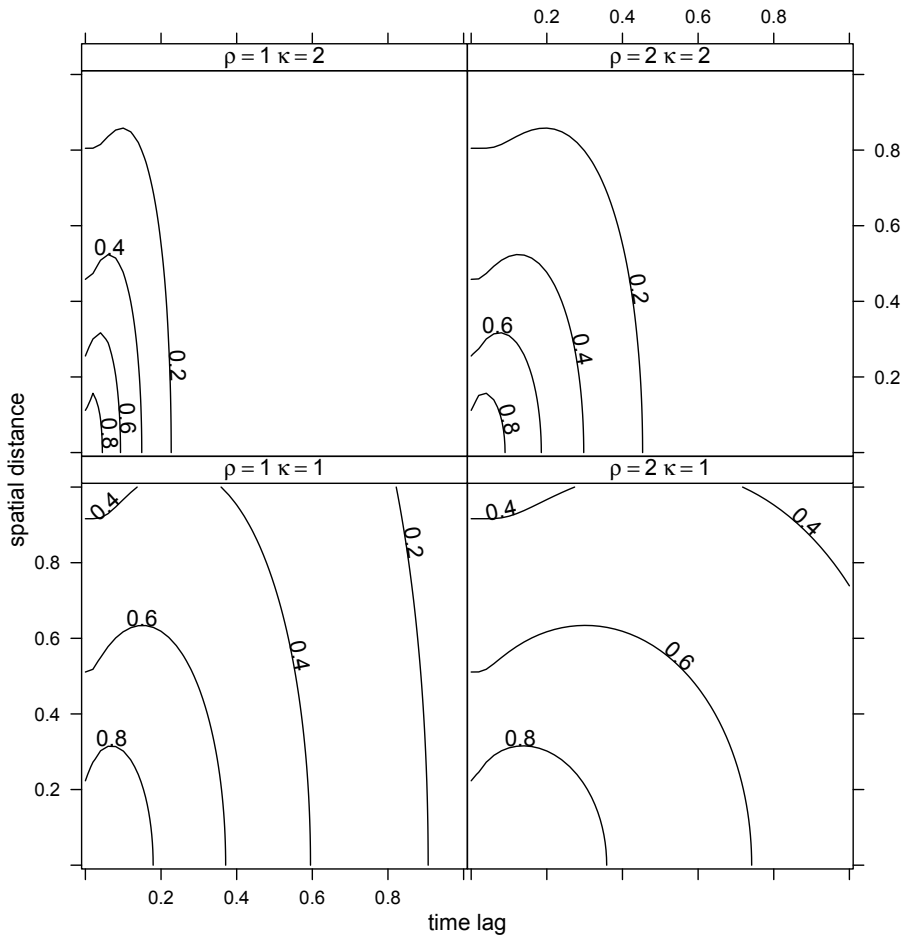


Figure 3.1: Space-time correlation function of the one dimensional damped heat equation model.

the decay for the ($\rho = 2$, $\kappa = 1$) and ($\rho = 1$, $\kappa = 2$) cases. An important feature is that the correlation dies faster along $(0, l)$ or $(r, 0)$ than along (r, l) which is a characteristic of positive non-separability.

The marginal spatial spectral density is obtained integrating over the temporal frequencies as

$$2\int_0^\infty f_{\mathrm{h}}(\lambda,\omega)\partial\omega = 1/(4\pi\rho\tau(\kappa^2+\lambda^2)) \; .$$

Therefore, the spatial marginal covariance is obtained as

$$2\int_0^\infty \frac{\cos(r\lambda)}{4\pi\rho\tau(\kappa^2+\lambda^2)}\partial\lambda = \frac{\mathrm{e}^{-r\kappa}}{4\rho\tau\kappa} \; .$$

This is the well known exponential correlation function with range $1/\kappa$ and marginal variance $1/(4\rho\tau\kappa)$. Setting $r^* = 0$ in 3.1, the temporal correlation is Erfc(l^*), Jones and Zhang (1997), which depends on κ and ϕ. These marginal correlation functions are shown in Figure 3.2 for some parameter values.

We usually consider $2/\kappa$ as a practical range for the exponential correlation function because the correlation at this distance is $\exp(-2) = 0.1353$. The Erfc function decays as the argument increases. ρ is inversely proportional to the temporal lag and κ^2 is directly proportional. Therefore ρ/κ^2 acts as a range parameter for the temporal marginal correlation. In fact, at the distance of $\sqrt{2\pi}\rho/\kappa^2$ the correlation is Erfc($\sqrt{2\pi}$) = 0.1134 and this expression can be considered as a practical temporal range for the marginal temporal correlation. These two correlation values are shown in each plot of Figure 3.2 as a dotted horizontal lines.

This model in one-dimensional space has potential applications for modeling phenomena on a river or road, that evolves over time. However, a possible application will be when considering two dimensional Gaussian process with different degree of smoothness in each direction. Gaussian process are considered in modeling varing regression coefficients in the machine learning literature, Rasmussen and Williams (2006). Varying regression coefficients are also considered in the context of Generalized Additive Models (GAM), Wood (2006). The advantage of the GMRF representation provided here is that it allows for efficient computations.

Jones and Zhang (1997) proposed the following SPDE

$$\tau\left[(\kappa-\Delta)^{\alpha_1/2}+\rho\frac{\partial}{\partial t}\right]u(\mathbf{s},t) = \mathcal{W}(\mathbf{s},t) \tag{3.8}$$

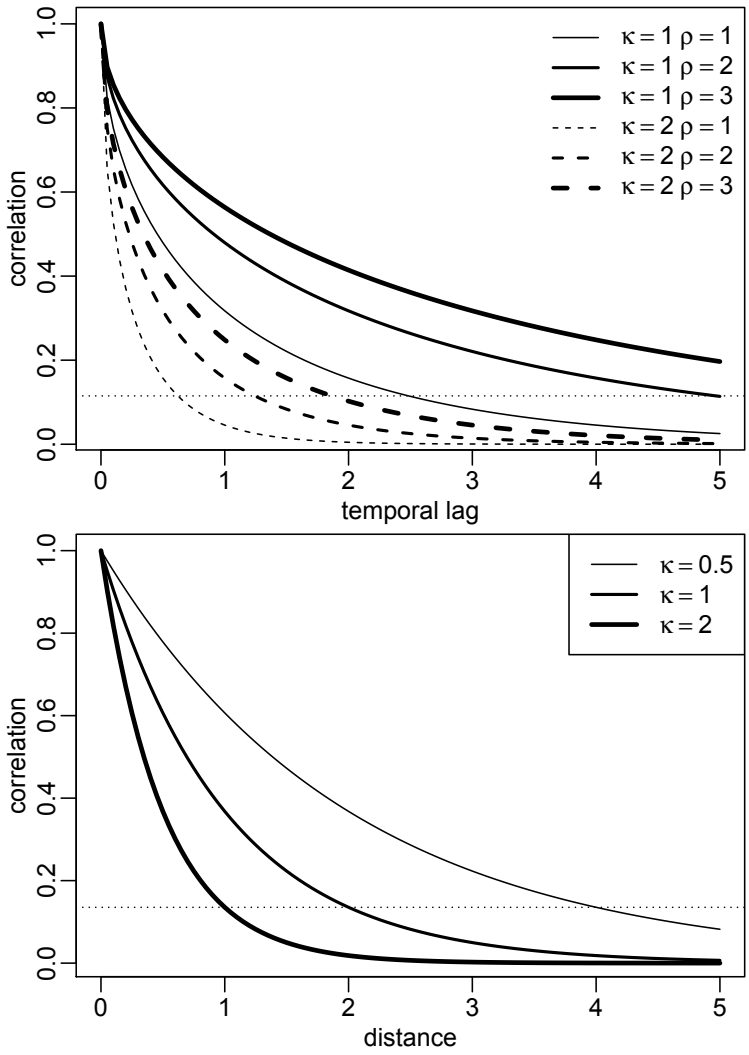


Figure 3.2: Marginal temporal covariance for different ρ and κ values (top) and marginal spatial covariance for different κ values (bottom).

which does enough smoothing when $\alpha_1 > d$. The spectral density is

$$f_{\text{hda}} = \frac{1}{(2\pi)^{d+1}\tau[\rho^2\omega^2 + (\kappa^2 + \|\lambda\|^2)^{\alpha_1}]}. \tag{3.9}$$

The marginal spatial spectral density is

$$2\int_0^\infty f_{\text{hda}}(\lambda,\omega)\partial\omega = 1/[4\pi^d\rho\tau(\kappa^2 + \|\lambda\|^2)^{\alpha_1/2}]$$

which corresponds to the spectral density of the Matérn covariance with smoothness equal to $(\alpha_1 - d)/2$ and range parameter $1/\kappa$. The marginal temporal correlation is

$$\mathrm{e}^{-l^*} - \sqrt{\pi l^*}\mathrm{Erfc}(\sqrt{l^*})$$

as in Jones and Zhang (1997). This model is a particular case of the model class proposed in Stein (2005) considered in Chapter 4.

3.2 Two GMRF representations for the intrinsically stationary stochastic heat equation model

In this section we consider two GMRF representations for the space-time model in Eq. (3.3). The first uses the Finite Volume Method (FVM) as proposed in Fuglstad (2010) and we will only present the results. The second uses the Finite Element Method (FEM).

3.2.1 Using the Finite Volume Method

The Finite Volume Method is comonly considered to solve problems of conservative laws of forms like the heat equation, Eymard et al. (2000). It approximates the average integral value on a reference value rather than pointwize approximations. Fuglstad (2010) considered the model given in Eq. (3.3) in the one-dimensional space domain to derive the precision matrix considering the FVM. We only present the results here and refer readers to Fuglstad (2010) for details of the methods.

It was considered $[0, L]$ as the spatial domain and $[0, T]$ as the time domain. Neumann boundary conditions were considered as $\frac{\partial}{\partial t}u(0,t) = 0, \quad t \in [0,T]$ and $\frac{\partial}{\partial t}u(L,t) = 0, \quad t \in [0,T]$. The approximation divides both spatial and temporal domains into regular grids. That is $0 < t_1 < ... < t_M = T$, $k = t_1 - t_0$, and $0 = s_0 < ... < s_N = L$. Further, define $d_i = s_{i+1} - s_i$, for $i = 1, ..., N-1$, and define spatial cells as

$E_1 = [s_1, s_1 + d_1/2]$, $E_N = [s_N - d_{N-1}/2, s_N]$ and $E_i = [s_i - d_i/2, s_i + d_i/2]$ for $i = 2, ..., N-1$.

Let a diagonal matrix $\boldsymbol{V} = \text{diag}(V_1, ..., V_N)$ where V_i is the length of the interval E_i. Then define a tridiagonal matrix $\boldsymbol{H}^{(V)}$ as

$$\boldsymbol{H}_{ij}^{(V)} = \begin{cases} -\frac{1}{d_i} & \text{if} \quad i = j = 1 \\ -(\frac{1}{d_{i-1}} + \frac{1}{d_i}) & \text{if} \quad 1 < i = j < N \\ -\frac{1}{d_{N-1}} & \text{if} \quad i = j = N \\ \frac{1}{d_i} & \text{if} \quad j = i+1 \text{ or } j = i-1 \\ 0 & \text{otherwise.} \end{cases}$$

Additionally, let $\boldsymbol{B} = \boldsymbol{V} - k\gamma\boldsymbol{H}^{(V)}$ and $\boldsymbol{B}^{(2)} = \boldsymbol{B}^T\boldsymbol{V}^{-1}\boldsymbol{B}$. The resulting precision matrix considers a precision matrix $\boldsymbol{A}$ for the initial state, at time t_0. The joint precision for the process at the space-time discretization grid is given by

$$\frac{\tau}{k}\begin{bmatrix} \boldsymbol{V} + \frac{k}{\tau}\boldsymbol{A} & -\boldsymbol{B} & 0 & \dots & 0 \\ -\boldsymbol{B}^T & \boldsymbol{V} + \boldsymbol{B}^{(2)} & -\boldsymbol{B} & & 0 \\ 0 & \ddots & \ddots & \vdots & \\ \vdots & & -\boldsymbol{B}^T & \boldsymbol{V} + \boldsymbol{B}^{(2)} & -\boldsymbol{B} \\ 0 & \dots & 0 & -\boldsymbol{B}^T & \boldsymbol{B}^{(2)} \end{bmatrix} \tag{3.10}$$

which has dimension equals to $N(M+1) \times N(M+1)$.

We can consider a small case to visualize what the precision looks like. Given $L = T = 5$, $k = 1$, $N = 6$ and $\tau = \gamma = 1$ the resulting precision matrix is Figure 3.3. It shows a global tridiagonal pattern, that is the diagonal and the interaction between the vector process at time t and $t-1$. The main diagonal is surrounded by two non-zero elements in each side due to the pattern of the $\boldsymbol{B}^{(2)}$ matrix. The interaction between time t and time $t-1$ is surrounded by one element on each side, which is due to the $-\boldsymbol{B}$ matrix that is tridiagonal.

3.2.2 Using the Finite Element Method

The FEM is a popular technique in applied mathematics and engeniering among other numerical methods for working with Partial Differential Equations, Quarteroni (2009). A particular feature of this technique is its flexibility of working with complicated geometries, for example non-regular spatial domains. Therefore it is interesting to consider it for spatial statistics.

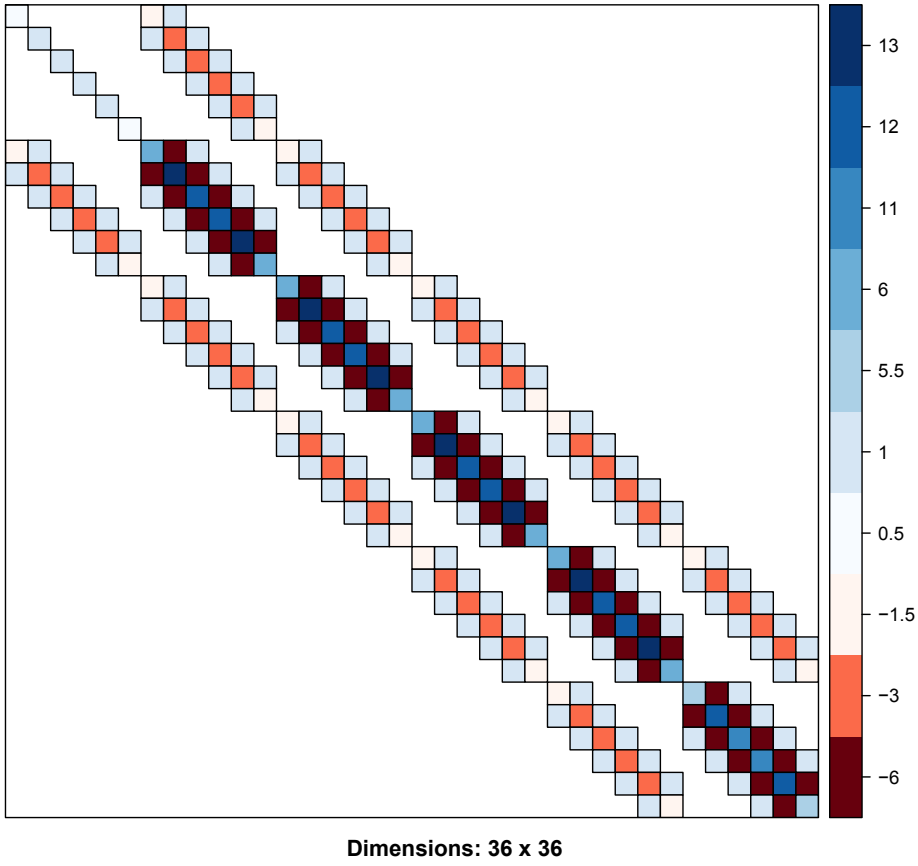


Figure 3.3: Precision matrix from FVM.

The discretization approach in this section considers the spatial discretization as in Lindgren et al. (2011). The spatial discretization was considered to write the process as a system of temporal SDEs. This is them decoupled and the generalized eigenvalue/eigenvector problem is considered to find the precision matrix. This diagonalization technique based in the generalized eigenvalue problem is common to solve differential equations. These steps are detailed next and follows from Lindgren et al. (2016).

First we consider the set of m spatial basis functions $\Psi(.)$ and the matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ as in Lindgren et al. (2011) and as briefly outlined in Section 2.3.2. We write the process $u(\boldsymbol{s}, t)$ as a function of the spatial basis

functions and the process at time t: $u(\boldsymbol{s},t) = \Psi(\boldsymbol{s})\boldsymbol{u}.(t)$ as

$$\langle \Psi(.), \frac{\partial}{\partial t}\mathbf{u}(.,t) - \gamma\Delta\mathbf{u}(.,t)\rangle_\Omega = \langle \Psi, \Psi\rangle_\Omega \frac{\partial}{\partial t}\mathbf{u}.(t) + \gamma\langle\nabla\Psi, \nabla\Psi\rangle_\Omega \boldsymbol{u}.(t)$$

The set of basis function inherits the spatial properties of the process. We can write the following coupled system of stochastic differential equations,

$$\boldsymbol{C}\frac{\partial}{\partial t}\mathbf{u}.(t) + \gamma\boldsymbol{G}\mathbf{u}.(t) = \tau\boldsymbol{C}\mathcal{W}.(t), \tag{3.11}$$

where $\mathcal{W}.(t)$ is a vector process with covariance equals to $|\delta t|\boldsymbol{C}^{-1}$.

Let $\boldsymbol{U}$ be a matrix of eigenvectors for the generalized eigenvalue problem

$$\boldsymbol{G}\boldsymbol{U} = \boldsymbol{C}\boldsymbol{U}\boldsymbol{\Lambda},$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{N_\Psi})$ is a diagonal matrix with the eigenvalues. The eigenvalues are chosen so that $\boldsymbol{U}^T\boldsymbol{C}\boldsymbol{U} = \boldsymbol{I}$ and $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{C}^{-1}$. Introduce $\boldsymbol{z}.(t) = \boldsymbol{U}^T\boldsymbol{C}\boldsymbol{u}.(t)$, and multiply Eq. (3.11) from the left, yielding

$$\frac{\partial}{\partial t}\mathbf{z}.(t) + \gamma\Lambda\mathbf{z}.(t) = \mathcal{W}^{\mathbf{z}}.(t), \tag{3.12}$$

where $\mathcal{W}^{\mathbf{z}}.(t)$ is a vector process with covariance $|\delta t|(\boldsymbol{U}^T\boldsymbol{C}\boldsymbol{U})^{-1} = |\delta t|\boldsymbol{I}$. We now have the system in Eq. (3.11) decoupled into m independent processes

$$\frac{\partial}{\partial t}z_k(t) + \gamma\lambda_k\boldsymbol{z}_k(t) = \tau^{-1}\mathcal{W}(t) \tag{3.13}$$

where $\mathcal{W}(t)$ is a white noise process.

Each temporal process $z_k(t)$ can be considered as in Eq. 2.13 or Eq. 2.14. Following Section 2.3.1 we have the temporal precision matrix as

$$\tau(\boldsymbol{H} + \gamma\boldsymbol{D})$$

for the spatial basis weights for $z_k(t)$. The joint precision for all the $z_k(t)$ processes $\boldsymbol{z}$ is

$$\boldsymbol{Q}_z = \tau(\boldsymbol{H}\otimes\boldsymbol{I} + \gamma\boldsymbol{D}\otimes\Lambda^2).$$

Transforming back to $\boldsymbol{u} = \boldsymbol{U}\boldsymbol{z}$ we have its precision as

$$\begin{aligned}\boldsymbol{C}\boldsymbol{U}\boldsymbol{Q}_z\boldsymbol{U}^T\boldsymbol{C} &= \tau\boldsymbol{C}\boldsymbol{U}(\boldsymbol{H}\otimes\boldsymbol{I} + \gamma\boldsymbol{D}\otimes\Lambda^2)\boldsymbol{U}^T\boldsymbol{C} \\ &= \tau(\boldsymbol{H}\otimes\boldsymbol{C} + \gamma\boldsymbol{D}\otimes\boldsymbol{G}^{(2)})\end{aligned} \tag{3.14}$$

where it was considered that $\boldsymbol{C}\boldsymbol{U}\Lambda^k\boldsymbol{U}^T\boldsymbol{C} = \boldsymbol{C}(\boldsymbol{C}^{-1}\boldsymbol{G})^k$.

We consider the same scenario as in the previous section, where we have $L = T = 5$, $k = 1$, $N = 6$ and $\tau = \gamma = 1$. It gives the precision matrix shown in Figure 3.4 considering Neumann boundary conditions both for the time and space domains. The overall tridiagonal pattern is from $\boldsymbol{H}\otimes\boldsymbol{C}$ and the main diagonal and surrounding terms from $\boldsymbol{D}\otimes\boldsymbol{G}^{(2)}$. Notice that $\boldsymbol{G}^{(2)}$ has band width equal to five.

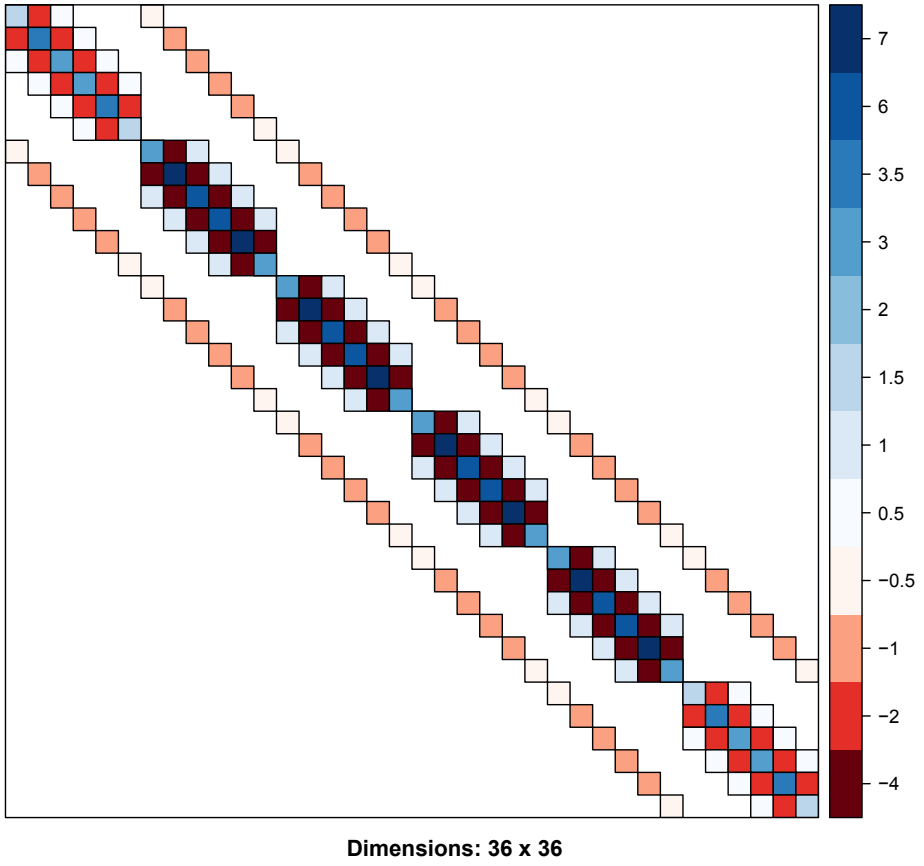


Figure 3.4: Precision matrix from FEM.

3.2.3 Fixed initial condition example

Since this is an intrinsically stationary model, one needs to set at least one constraint in order to compute the covariance for the constrained field. Following Fuglstad (2010) we consider $T = 10$, $L = 5$, $k = 1/100$, $N = 400$ and a constraint in the initial state. That is, the process at t_0 is

$$\boldsymbol{u}(x,0) = \begin{cases} 10 & \text{if} \quad 0 \leq x < \frac{5}{2} \\ -10 & \text{if} \quad \frac{5}{2} \leq x \leq 5 \end{cases} .$$

We compare the marginal mean and marginal variance from both GMRF representations under this scenario. The mean conditional on this starting

condition and the marginal variance for both discretization approaches is shown in Figure 3.5

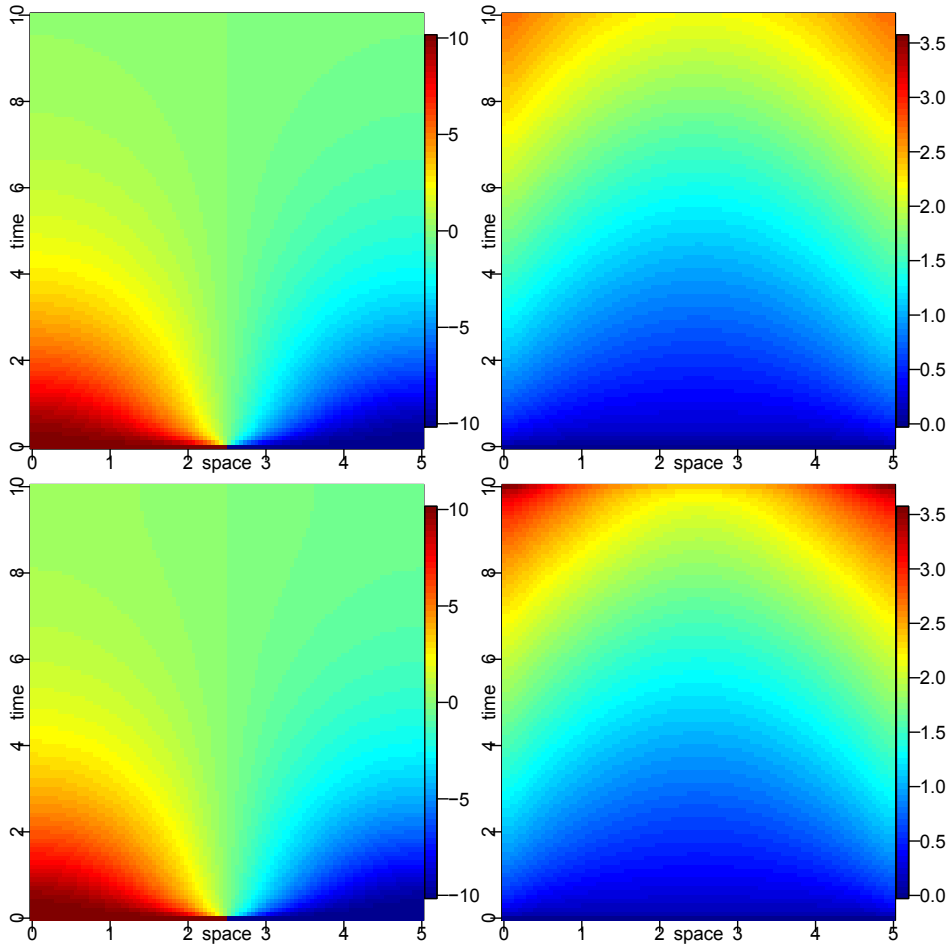


Figure 3.5: Marginal mean from the FVM discretization (top left) and marginal variance (top right). Marginal mean from the FEM discretization (bottom left) and marginal variance (bottom right).

We can also draw samples from the initial conditions considering each of the derived precision matrix structures as shown in Figure 3.6.

3.3 Constrained process covariance example

The spectral density for the model in Eq. (3.3) is

$$f(\omega, \lambda) = \frac{1}{(2\pi)^2 \tau(\omega^2 + \gamma\lambda^4)} \tag{3.15}$$

which is a particular case of the proposed model in Stein (2013), where expressions for the generalized covariance were derived.

Since the process is intrinsic, the variance is not finite. However, it is intrinsically stationary and we can compute the covariance for a valid linear transformation. In this case, we can consider a sum-to-zero constraint in the process, that is $\int_R \int_R u(s,t)dsdt = 0$. The constrained process empirical covariance is the generalized inverse of the precision matrix from the discretization.

Consider the scenario when $T = L = 5$ and $n = m + 1$, that is, the same time and spatial window size with the same discretization knots. In order to understand the effect of γ we inspect the empirical covariance of the constrained process for γ equal to 1/10, 1 and 10.

In this section we consider the generalized inverse of the precision matrix from the FEM discretization in order to visualize some marginal properties of the model. We can see the marginal variance computed for the constrained process in Figure 3.7 considering different values for γ.

The first point to notice is the effect of the sum to zero constraint. Under this constraint the marginal variance is not constant over the domain, similar to what one can observe for random walk models. The behaviour is that the marginal variance is lower in the center of the space-time domain and lager near the boundary. This is clear also when looking at it shown over time in the left plots in Figure 3.8 or over space in the right plots in this Gigure. This is due to the different degree of smoothness along time and space. Considering the case of $\gamma = 1$ we can see that the marginal variance is near one at the spatial boundaries (either 0 or 5) for t between 2 and 3. However, at the time boundaries (either 0 or 5) it is near 0.5 for s between 2 and 3. In resume, it varies more over space when looking at the same time point than over time when looking at the same spatial location.

For $\gamma = 1$ we have the case where the SPDE has no coefficient multiplying by the spatial operator, as we already have for the time derivative. This case is shown in the plots in the center of Figure 3.7. What is shown in these plots is the effect of the smoothing order. We observe that the second order smoothing causes more variation in the process because the marginal variance varies more over space than over time.

We can see that when $\gamma = 0.1$ the marginal variance at each spatial knot does not vary so much as shown in the bottom left plot. However, the marginal variance does vary more at each time knot as seen in the mid left plot. That is, the marginal variance varies mainly over space. The joint variation is shown in the top plots for different γ values. As γ increases the marginal variance starts to vary more over time and less over space. This result makes sense as with higher γ we have a bigger spatial range and, consequently, less variation over space for the process. One point here is that the non-separability happens already in the marginal variance as λ affects the marginal variance over both space and time.

Figure 3.9 presents the correlation over space for a specific time point and over time at a specific spatial point. A key feature shown in this Figure is the feature due the difference in the degree of smoothness of the process over space and over time. Looking at the case when $\gamma = 1$, shown in the mid left and mid right plots in Figure 3.9, one can see that the spatial correlation is smoother than the temporal correlation because it goes less fast from 1 to -1 along the spatial direction.

Another point to be noticed in Figure 3.9 is the effect of the γ parameter. The γ parameter is proportional to the spatial range because for bigger values of it the correlation decay is slower. Thus, the correlation over space is higher as γ increases. However, when the range is relatively big with respect to the size of the area, as in the case of $\gamma = 10$, we observe less variation. The opposite is true for the temporal correlation, as γ increases the temporal range decreases and this is a feature of the non-separability of the implied covariance structure of this model.

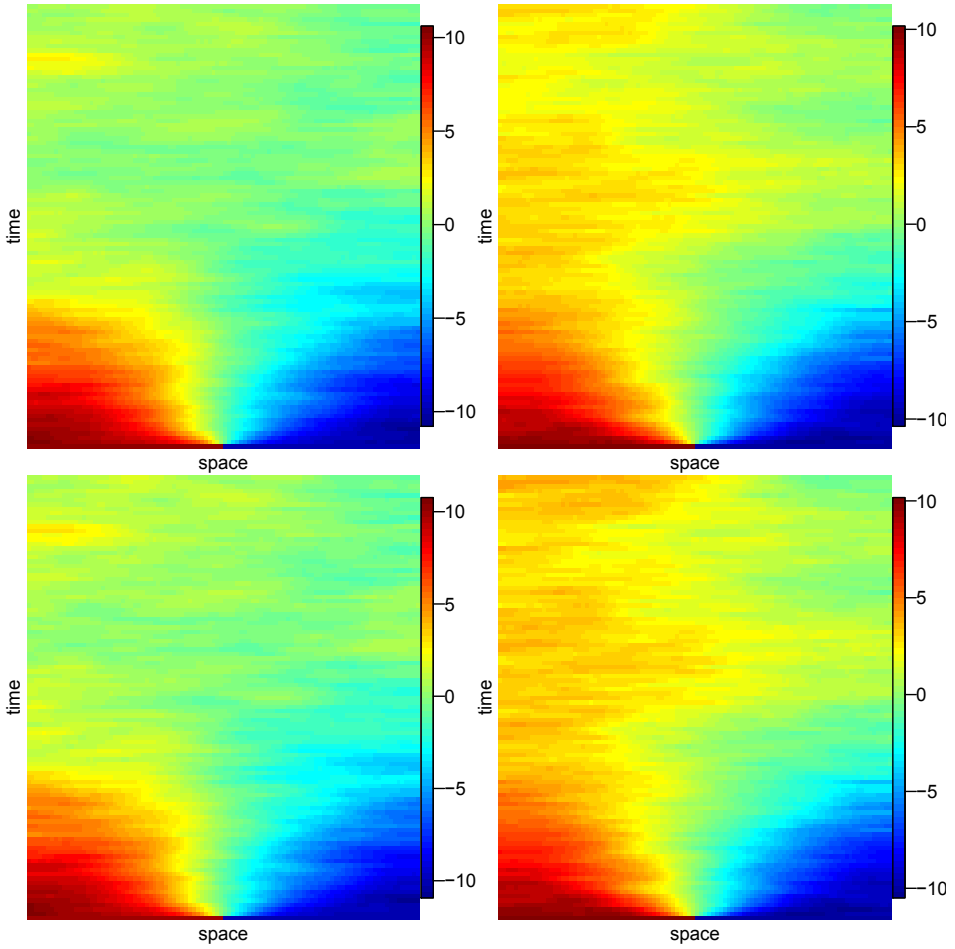


Figure 3.6: Two samples considering the initial condition and the FVM discretization (top) and two samples considering the initial condition and the FEM discretization (bottom).

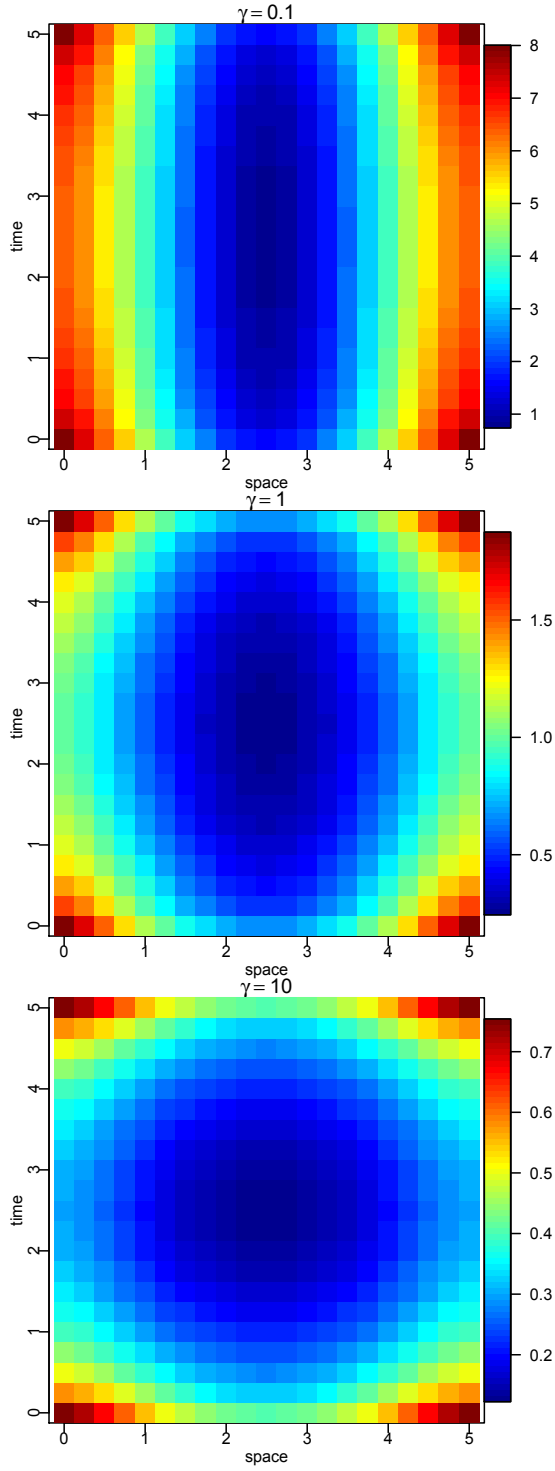


Figure 3.7: Constrained process marginal variance at each space-time discretization point for some values of γ.

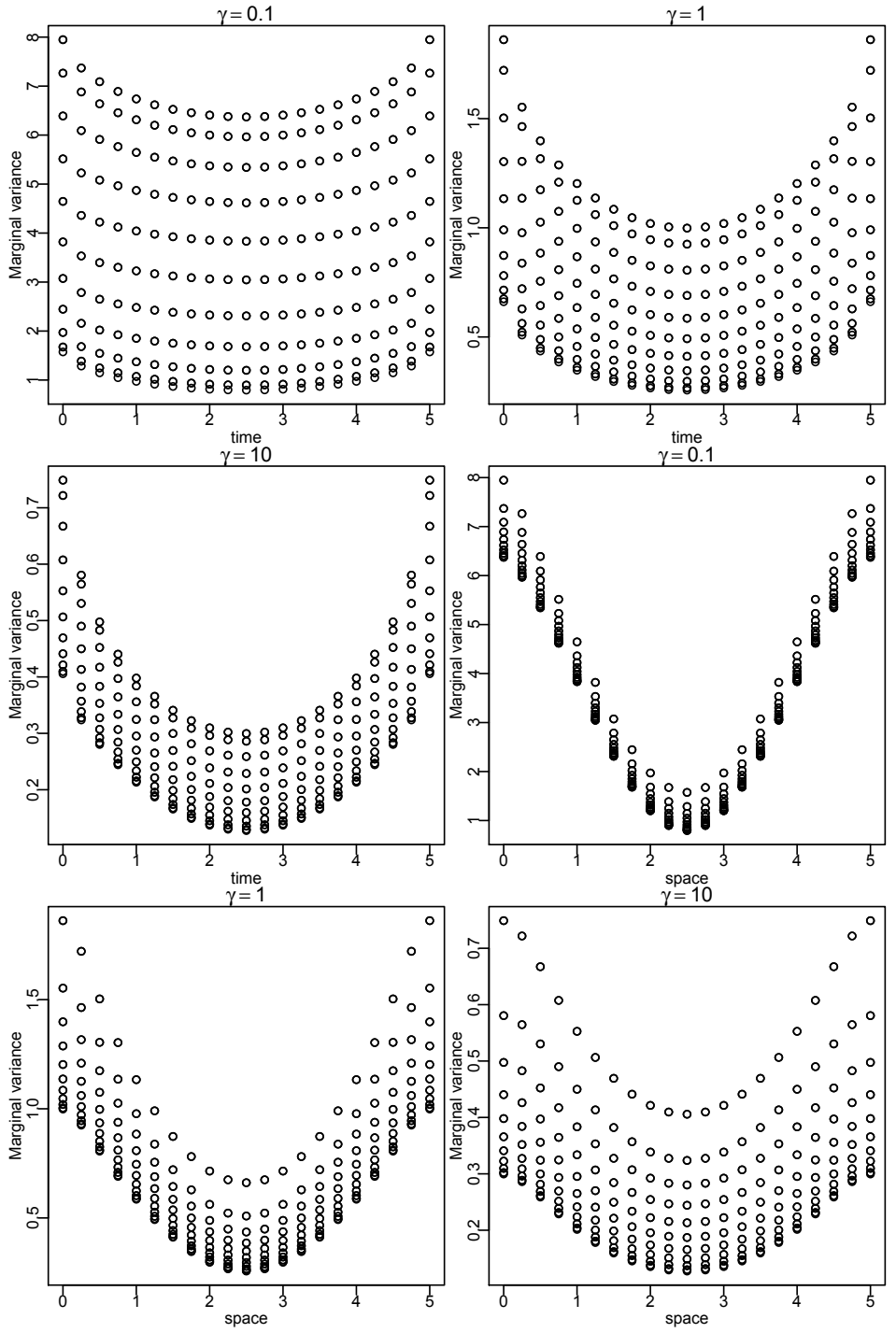


Figure 3.8: Marginal variance shown over time at each spatial location for different values of γ (left plots). Marginal variance over space at each time point for different values of γ (right plots).

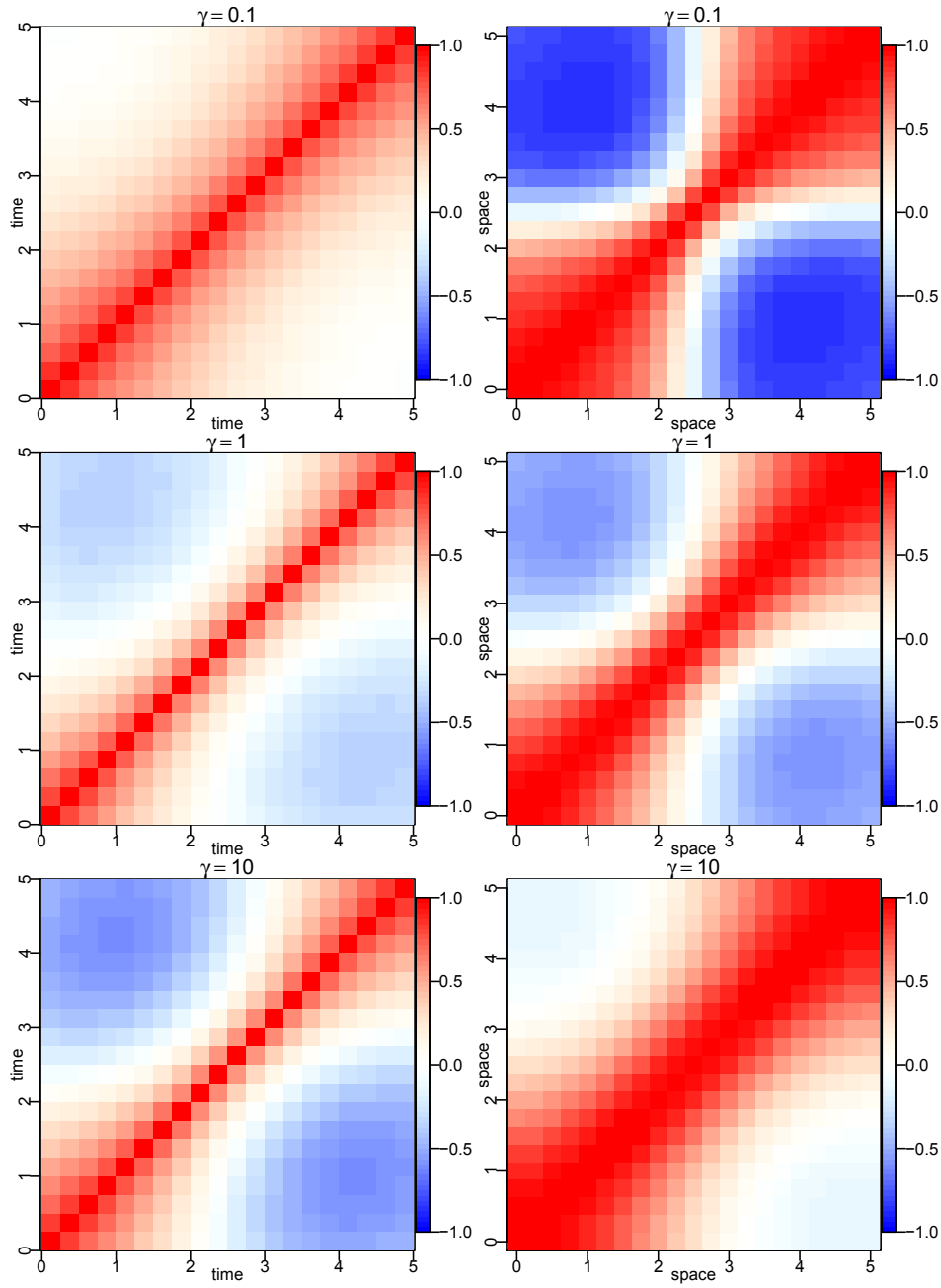


Figure 3.9: Temporal covariance for space fixed at $\lfloor n/2+1 \rfloor$ (left plots) and spatial covariance for time fixed at $\lfloor m/2+1 \rfloor$ (right plots).

Chapter 4

Stein's model SPDE and a GMRF representation

In this Chapter we consider the model proposed in Stein (2005). We provide a SPDE formulation of this model by proposing an SPDE along with a spectral representation that gives the same spectral density of Stein's model. We consider a GMRF representation for particular cases of the smoothness parameters of this model.

4.1 Stein's model

Stein (2005) proposed a non-separable model starting from the following spectral density

$$f(\boldsymbol{w}_1, w_2) = \{c_1(a_1^2 + \|\boldsymbol{w}_1\|^2)^{\alpha_1} + c_2(a_2^2 + |w_2|^2)^{\alpha_2}\}^{-\nu} \tag{4.1}$$

where c_1 and c_2 are scaling parameters, a_1 and a_2 are inverse range parameters and α_1, α_2 and ν are smoothness parameters. It is needs $c_1 > 0$ and $c_2 > 0$ and to have it integrable it is needed $a_1^2 + a_2^2 > 0$ and a minimum smoothing condition stated by

$$d_1/\alpha_1 + d_2/\alpha_2 < 2\nu. \tag{4.2}$$

For the space-time case we consider $\boldsymbol{w}_1$ the two dimensional frequency in space and w_2 the one dimensional temporal frequency, which gives $d_1 = 2$ and $d_2 = 1$. Therefore, the required smoothing order over space is twice as over time.

Without loss of generality we can set $a_2 = 0$ and consider $x = |w_2|$ and use 3.241.4 of Gradshteyn and Ryzhik (2007) to compute the marginal

spectral density as

$$\begin{aligned} f(\boldsymbol{w}_1) &= 2\int_0^\infty \frac{\partial x}{(c_1(a_1^2+\|\boldsymbol{w}_1\|^2)^{\alpha_1}+c_2x^{2\alpha_2})^\nu} \\ &= \frac{\Gamma(\frac{1}{2\alpha_2})\Gamma(\nu-\frac{1}{2\alpha_2})}{\Gamma(\nu)\alpha_2 c_2^{\frac{1}{2\alpha_2}} c_1^{\nu-\frac{1}{2\alpha_2}} (a_1^2+\parallel \boldsymbol{w}_2 \parallel^2)^{\alpha_1(\nu-\frac{1}{2\alpha_2})}} \end{aligned} \tag{4.3}$$

which means that the marginal spatial covariance is Matérn with scale a_1 and smoothness $\alpha_1(\nu-1/(2\alpha_2))-1$. Similarly for time, by setting $a_1 = 0$ we have the marginal temporal covariance as Matérn with scale a_2 and smoothness $\alpha_2(\nu-1/(2\alpha_1))-1/2$. Thus, with $\alpha_1 = \alpha_2 = 1$ and $\nu = 2$, we have marginal spatial as Exponential and marginal temporal as Matérn with smoothness equals to 1.

This knowledge is useful because all the knowledge from the Matérn covariance can be considered to help understanding and interpreting the model parameters. For example, for choosing α_1, α_2 and ν. Notice that setting $\nu = 2$, $\alpha_1 = 1$ and $\alpha_2 = 1/2$ would give the Exponential temporal covariance case. However, the smoothing condition is not verified for this case. For $\nu = 2$ and $\alpha_1 = \alpha_2 = 1$ the smoothing condition is satisfied and it give both marginal covariances, spatial and temporal, as Matérn with smoothness equals 1.

Analytical expression for the space-time covariance is not possible to compute for all the range of the smoothness parameters. It can be computed only few particular cases. The case when $\nu = 1$ and $\alpha_2 = 1$ we have Exponential temporal covariance for each spatial frequency. When $a_2 = 0$, $c1 = 1$ and $\alpha_1 = 2$, $\alpha_2 = 1$, and $\nu = 2$, the analytical expression for the covariance was given in Eq. (6) in Stein (2005).

The key point of our work with this model is based in two results from Rozanov (1977). First, having α_1 and α_2 integers and ν half-integer gives the reciprocal of Eq. (4.1) a polynomial in $\boldsymbol{w}_1$ and w_2 and thus, the associated process is a Markovian field. Second, a solution of a linear SPDE is a Markovian random field. Therefore, we are working with a Markovian space-time process and there will be explicit GMRF representations associated, like in Lindgren et al. (2011).

The heat equation considered in Chapter 3 is a particular case, for $d_1 = 1$. Another particular case of this model, when $\nu = 1 = \alpha_2 = 1$ and $\alpha_1 > 2$, was considered in Jones and Zhang (1997) and Kelbert et al. (2005).

We will have a GMRF representation for this model in Section 4.3. This provides a computationally efficient way for model inference. Then, by replacing the Stein's third desired property by

3. fitting process computationally efficient

we have a space-time model fulfilling all the three Stein's like desired properties.

4.2 The SPDE formulation

Let the following SPDE

$$\left(c_1(a_1^2 - \Delta)^{\alpha_1} + c_2(a_2^2 - \frac{\partial^2}{\partial t^2})^{\alpha_2}\right)^{\nu/2} u(\boldsymbol{s}, t) = (2\pi)^{d+1}\mathcal{W}(\boldsymbol{s}, t) \qquad (4.4)$$

where $(\mathbf{s}, t)$ is the space-time coordinate for $\mathbf{s} \in \mathbb{R}^d$ and $t \in \mathbb{R}$, $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{s_i^2}$ and c_1, c_2, a_1, a_2, α_1, α_2 and ν the model parameters in Eq. 4.1.

One way of defining fractional operators is to consider the spectral representations as $\mathcal{F}(-\partial^2/\partial t^2)^a = |\omega^2|^a = |\omega|^{2a}$ and $\mathcal{F}(-\Delta)^a = \|\lambda^2\|^a = \|\lambda\|^{2a}$. Thus, considering $d = 2$ and $\mathcal{W}(\boldsymbol{s}, t)$ to be unit variance space-time white noise we arrive at the spectral density in Eq. 4.1.

We have some particular cases of this model in Table 4.2 considering different values for α_1 and α_2 and the condition needed for ν. The three first cases, when one or both is $1/2$ it is required $\nu > 2$. For having $\nu = 2$ in such cases one can them consider a spatially correlated driving noise, as for the models in Chapter 5. The next three cases it is required $\nu > 1$ and we can have, for example, $\nu = 2$ and it does not implies interaction terms in the spectrum. Thus, in the $\nu = 2$ case both the spatial and temporal derivatives are collapsed into the same noise giving a completely non-separable space-time model.

An important fact for having integer values of α_1, α_2 and ν is that the spectral density is a reciprocal of a polynomial and the associated process is Markovian, Rozanov (1977). For such cases there may be a Markov representation. Thus having Gaussian driving noise we can consider a GMRF representation which allows for efficient computations in the fitting process. Then, by replacing the third Stein's desired property by

3. fitting process computationally efficient

we have a space-time model class that fulfilling all the desired (like) properties.

Setting $a_2 = 0$ in Eq. (4.4) we have the $c_2|w_2|^{2\alpha_2}$ term in Eq. (4.1). However, if we replace the negated second order time derivative in the SPDE by the positive first order one we have the $c_2|w_2|^{\alpha_2}$ instead. Therefore,

Table 4.1: Condition on ν for α_1 and α_2

α_1, α_2	Operator	ν
$1/2, 1/2$	$\left(c_1(a_1^2 - \Delta)^{1/2} + c_2(a_2^2 - \frac{\partial^2}{\partial t^2})^{1/2}\right)^{\nu/2}$	> 3
$1/2, 1$	$\left(c_1(a_1^2 - \Delta)^{1/2} + c_2(a_2^2 - \frac{\partial^2}{\partial t^2})\right)^{\nu/2}$	$> 5/2$
$1, 1/2$	$\left(c_1(a_1^2 - \Delta) + c_2(a_2^2 - \frac{\partial^2}{\partial t^2})^{1/2}\right)^{\nu/2}$	> 2
$1, 1$	$\left(c_1(a_1^2 - \Delta) + c_2(a_2^2 - \frac{\partial^2}{\partial t^2})\right)^{\nu/2}$	$> 3/2$
$1, 2$	$\left(c_1(a_1^2 - \Delta) + c_2(a_2^2 - \frac{\partial^2}{\partial t^2})^2\right)^{\nu/2}$	$> 5/4$
$2, 1$	$\left(c_1(a_1^2 - \Delta)^2 + c_1(a_2^2 - \frac{\partial^2}{\partial t^2})\right)^{\nu/2}$	> 1
$2, 2$	$\left(c_1(a_1^2 - \Delta)^2 + c_2(a_2^2 - \frac{\partial^2}{\partial t^2})^2\right)^{\nu/2}$	$> 3/4$

doubling α_2 when considering the alternative first order time operator can give the same power spectrum. Thus, two different operators can give the same spectral density but it is a fundamentally different operator. This fact matters when having non-Gaussian driving noise because the distributions actually become different depending on how the operators are defined, Bolin (2014).

An alternative parametrization for the SPDE in Eq. 4.4 is

$$\tau[(1 - \gamma\Delta)^{\alpha_1} + (1 - \rho\frac{\partial^2}{\partial t^2})^{\alpha_2}]^{\nu/2} u(\boldsymbol{s}, t) = (2\pi)^{-(d+1)} \mathcal{W}(\boldsymbol{s}, t) \tag{4.5}$$

where there is one parameter less and the γ and ρ parameter can be easier to interpret.

4.3 Obtaining the precision matrix

We now follow the steps in the sub-section 3.2.2. The first step is to project the spatial properties to the spatial basis functions Ψ in order to have the a system of stochastic differential equations.

We can take the case when $\alpha_2 = 1$ and $\nu = 2$ to work with the following space-time SPDE

$$(c_1(a_1^2 - \Delta)^{\alpha_1} + c_2(a_2^2 - \frac{\partial^2}{\partial t^2}))u_(\mathbf{s}, t) = \mathcal{W}(\mathbf{s}, t)$$

and we have the following system of SDEs

$$c_1\mathbf{R}(a_1, \alpha_1)\mathbf{u}.(t) + c_2 a_2^2 \mathbf{C}\mathbf{u}.(t) - c_2\mathbf{C}\ddot{\mathbf{u}}.(t) = \mathbf{C}\mathcal{W}.(t) \tag{4.6}$$

where $\mathbf{R}(a_1, \alpha_1)$ is the precision matrix for a spatial Matérn field with scale a_1 and smoothness $\alpha_1 - 1$, as in Lindgren et al. (2011). For $\alpha_1 = 1$ we have $\mathbf{R}(a_1, 1) = a_1^4\mathbf{C} + 2a_1^2\mathbf{G} + \mathbf{G}^{(2)}$.

Then we consider the generalized eigenvalue problem and introduce $\mathbf{x}.(t) = \mathbf{U}^\top\mathbf{C}\mathbf{z}.(t)$. Multiplying Eq.(4.6) by $\mathbf{U}^\top$ from the left we arrive at a polynomial on $\mathbf{\Lambda}$ of order $2\alpha_1$ multiplying $\mathbf{x}.(t)$ plus a term with second order time derivative for $\mathbf{x}$. When $\alpha_1 = 1$ we have

$$(c_1a_1^4 + c_2a_2^2)\mathbf{x}.(t) + 2c_1a_1^2\Lambda\mathbf{x}.(t) + c_1\Lambda^2\mathbf{x}.(t) - c_2\ddot{\mathbf{x}}.(t) = \mathcal{E}^x.(t)$$

With respect to time, the above expression is a system of SDEs, $z_k(t)$, each one with the operator of the form $c_2(a_k/c_2 - \frac{\partial^2}{\partial t^2})$, where $a_k = (c_1a_1^4 + c_2a_2^2) + 2c_1a_1^2\lambda_k + c_1\lambda_k^2$. Considering the temporal basis functions ψ and the $\mathbf{D}$ and $\mathbf{H}$ matrices as defined in the sub-section 3.2.2 and the scale as a_k/c_2 we have precision of the form $c_2((a_k/c_2)^2\boldsymbol{D} + 2a_k/c_2\boldsymbol{H} + \boldsymbol{H}^{(2)})$.

With $a = c_1a_1^4 + c_2a_2^2$, $b = 2c_1a_1^2$ and so $a_k = a + b\lambda_k + c_1\lambda_k^2$ we have the precision for the basis weights for $x_k(t)$ written as

$$\begin{aligned}\boldsymbol{Q}_k \;=\; & (a^2 + 2ab\lambda_k + (2ac + b^2)\lambda_k^2 + 2bc\lambda_k^3 + c_1^2\lambda_k^4)/c_2\boldsymbol{D} + \\ & 2(a + b\lambda_k + c_1\lambda_k^2)\boldsymbol{H} + c_2\boldsymbol{H}^{(2)}\end{aligned}$$

Thus, the joint precision for all the $x_k(t)$ processes is

$$\begin{aligned}&(a^2\mathbf{D}\otimes\mathbf{I} + 2ab\boldsymbol{D}\otimes\Lambda + (2ac_1{+}b^2)\boldsymbol{D}\otimes\Lambda^2 + 2bc_1\boldsymbol{D}\otimes\lambda^3 + c_1^2\boldsymbol{D}\otimes\Lambda^4)/c_2 + \\ &2a\boldsymbol{H}\otimes\boldsymbol{I} + 2b\boldsymbol{H}\otimes\Lambda + c_1\boldsymbol{H}\otimes\Lambda^2 + c_2\boldsymbol{H}^{(2)}\otimes\boldsymbol{I}\ .\end{aligned}$$

Transforming back to $\boldsymbol{u} = \boldsymbol{U}\boldsymbol{x}$ we have its precision as

$$\begin{aligned}&a^2/c_2\boldsymbol{D}\otimes\boldsymbol{C} + 2ab/c_2\boldsymbol{D}\otimes\boldsymbol{G} + (2ac_1 + b^2)/c_2\boldsymbol{D}\otimes\boldsymbol{G}^{(2)} + \\ &2bc_1/c_2\boldsymbol{D}\otimes\boldsymbol{G}^{(3)} + c_1^2/c_2\boldsymbol{D}\otimes\boldsymbol{G}^{(4)} + \\ &2a\boldsymbol{H}\otimes\boldsymbol{C} + 2b\boldsymbol{H}\otimes\boldsymbol{G} + 2c_1\boldsymbol{H}\otimes\boldsymbol{G}^{(2)} + c_2\boldsymbol{H}^{(2)}\otimes\boldsymbol{C}\ .\end{aligned} \tag{4.7}$$

Something can be said about the implied conditional distributions for the discretization looking to the above matrix expression and link it to the SPDE. First, setting $a_2 = 0$ only changes a and so it seems to be redundant to have two damping parameters. The first term is a diagonal matrix and the elements in the first and second lines are related to the spatial interactions as we have the $\boldsymbol{D}$ matrix, which is diagonal, from the temporal. Notice that we have c_2 in these two first lines and that dividing all the terms.

The first and last terms in the last line are the only temporal interaction terms and the two remaining are the space-time interaction terms. c_1 controls some of the temporal interaction as it is in a and the space-time interaction. c_2 is directly multiplying $\boldsymbol{H}^{(2)}$.

4.4 Particular case: iterated heat equation

We can consider an iterated heat equation SPDE as follows

$$\left((a_1^2 - \Delta)^{\alpha_1/2} + \sqrt{c_2}\frac{\partial}{\partial t}\right)^{\nu} u(\boldsymbol{s}, t) = \mathcal{W}(s, t)\ . \tag{4.8}$$

The spectral density is

$$f(\boldsymbol{w_1}, w_2) = \left((a_1^2 + \parallel \boldsymbol{w}_1 \parallel^2)^{\alpha_1} + c_2 w_2^2\right)^{-\nu} / (2\pi)^3$$

which is a particular case of having $\alpha_2 = 1$ in Eq. (4.1), without the constant $(2\pi)^3$.

Considering the minimum smoothing condition in Eq. (4.2) we do need $\nu > 1/2 + 1/\alpha_1$. We can choose $\alpha_1 = 3$ in order to have $\nu = 1$ and no space-time interactions in the spectra. For $\nu = 1$ we already have in Eq. (3.8) the SPDE proposed by Jones and Zhang (1997) and the need to have $\alpha_1 > 2$.

Without going into the details we can define the precision matrix that matches the spectral density. The precision matrix for this case is

$$\boldsymbol{D} \otimes \boldsymbol{R}(a_1, \alpha_1) + c_2 \boldsymbol{H} \otimes \boldsymbol{C}$$

In the next chapter we work with the case when the driving noise is spatially correlated. Them an additional $\boldsymbol{R}(.)$ term is introduced in the precision, see 5.4.

Chapter 5

Iterated heat equation and coloured driving noise

In this chapter we consider a class of non-separable space-time models derived from the iterated heat equation with spatially correlated driving noise. This is a particular case of the transport and diffusion equation mentioned in Lindgren et al. (2011) and considered in Sigrist et al. (2015) because we do not work with the transport term.

The model defined for the spatial and temporal coordinates $\mathbf{s}$ and t can be written as follows

$$\begin{aligned} \tau(\rho\tfrac{\partial}{\partial t} + 1 - \gamma_x\Delta)^{\alpha_t} u(\mathbf{s},t) &= \mathcal{E}(\mathbf{s},t) \\ (1-\gamma_{\mathcal{E}}\Delta)^{\alpha_{\mathcal{E}}/2}\mathcal{E}(\mathbf{s},\delta t) &= \mathcal{W}_{\mathcal{E}}(\mathbf{s},\delta t) \end{aligned}$$

where $\mathcal{W}_{\mathcal{E}}$ is space-time white noise, $\mathcal{E}(\mathbf{s},t)$ is a stationary white noise process in time and spatially correlated, α_t and $\alpha_{\mathcal{E}}$ are smoothness parameters which will be usually fixed when estimating the other model parameters, τ, ρ, γ_x and $\gamma_{\mathcal{E}}$ which are real positives. τ is a precision parameter, ρ is the time correlation parameter and both γ_x and $\gamma_{\mathcal{E}}$ are spatial range parameters.

The main purpose of this chapter is to study the marginal properties of the model. We actually do not need to compute covariances for fitting the models, because the precision matrix is actually easier to compute from an linear SPDE model and it is much better to work with due its sparsity. However, the knowledge about the meaning of the parameters in the marginal properties can help to understand better the process and the parameters as well. Furthermore, this knowledge is useful for defining prior distributions. Also, in some cases one needs to sample from the model and them it is really needed to have an idea of which parameter combinations to choose for.

We have a particular model having not the damping term in the first equation and thus the iterated intrinsically stationary heat equation in Section 5.1. The iterated damped heat equation, as defined above, is considered in Section 5.2. These two sections consider the sphere as the spatial domain. For the damped case we also have the case when the spatial domain is the $\mathbb{R}$ in Section 5.3. The precision matrix for these models taken from the discretization is shown in Section 5.4. An application of these equations to model daily global temperature data is considered in section 5.5.

5.1 Intrinsically stationary iterated heat equation

The model considered in this section is based on an iterated intrinsically stationary heat equation with a stationary spatially correlated driving noise. It can be written as

$$\tau(\rho\frac{\partial}{\partial t} - \Delta)^{\alpha_t} u(\mathbf{s},t) = \mathcal{E}(\mathbf{s},t) \quad (5.1)$$

$$(1-\gamma_{\mathcal{E}}\Delta)^{\alpha_{\mathcal{E}}/2}\mathcal{E}(\mathbf{s},\delta t) = \mathcal{W}_{\mathcal{E}}(\mathbf{s},\delta t) \quad (5.2)$$

where $\mathcal{W}_{\mathcal{E}}$ is a space-time white noise term, $\mathcal{E}(\mathbf{s},t)$ is a stationary process with white noise properties in time and spatially correlated, α_t and $\alpha_{\mathcal{E}}$ are smoothness parameters which will be usually fixed when estimating the other model parameters, τ, ρ and $\gamma_{\mathcal{E}}$ which are real positives and our main interest. Notice that we are considering the range parameter $\gamma_{\mathcal{E}}$ instead of the scale parameter $\kappa_{\mathcal{E}}$ for the Matérn driving noise.

From equations, Eq. (5.1) and Eq. (5.2), one can see that these parameters are related to the overall precision, temporal dependency and spatial dependency, respectively. However, it is necessary to actually map the marginal variance, temporal range and spatial range to these parameters in order to understand their contribution in the model.

5.1.1 Spectral densities

When considering the sphere as the spatial domain, we have discrete spatial frequencies k, $k = 0, 1, ...$ and $\lambda_k = k(k+1)$. With a continuous time domain and with temporal frequencies $\omega \in \mathbb{R}$, the spectral representation is given by

$$(i\rho\omega + \lambda_k)^{\alpha_t}\hat{z}(k,\omega) = \tau^{1/2}\hat{\mathcal{E}}(k,\omega)$$

$$(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_t/2}\hat{\mathcal{E}}(k,\omega) = \hat{\mathcal{W}}_{\mathcal{E}}(k,\omega)$$

and the spectral densities are obtained as follows

$$\begin{aligned}(\rho^2\omega^2+\lambda_k^2)^{\alpha_t}S_u(\lambda_k,\omega) &= \tau^{-1}S_{\mathcal{E}}(\lambda_k,\omega)\\ (1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}S_{\mathcal{E}}(\lambda_k,\omega) &= \frac{1}{2\pi}\end{aligned}$$

giving the spectral density as

$$S_u(\lambda_k,\omega)=\frac{1}{2\pi\tau(\rho^2\omega^2+\lambda_k^2)^{\alpha_t}(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}} \tag{5.3}$$

which has intrinsicness at $S_u(0,0)$.

We can rewrite Eq. (5.3) as

$$S_u(k,\omega)=\frac{1}{\tau(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\rho^{2\alpha_t}}\frac{1}{2\pi(\omega^2+(\lambda_k/\rho)^2)^{\alpha_t}}\,. \tag{5.4}$$

to easily see that for each spatial frequency $k>0$ we have a Matérn temporal covariance with smoothness α_t and scale parameter λ_k/ρ, Lindgren (2013), page 226. As ρ/λ_k is the corresponding range parameter, we have smaller temporal range for higher spatial frequencies. Thus, the local spatial behaviour does have less effect on the temporal range than the overall spatial behaviour.

The marginal temporal spectral density can be computed considering Eq. (5.4) and summing over the spatial frequencies

$$\begin{aligned}S_u(\omega) &= \sum_{k=0}^{\infty}\frac{2k+1}{4\pi}S_u(k,\omega)\\ &= \frac{1}{8\pi^2\tau\rho^{2\alpha_t}}\sum_{k=0}^{\infty}\frac{2k+1}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(\omega^2+\frac{\lambda_k^2}{\rho^2})^{\alpha_t}}\,.\end{aligned} \tag{5.5}$$

The marginal spatial spectral density is obtained by integrating with respect to the temporal domain, that is

$$\begin{aligned}S_u(k) &= \int_{\mathbf{R}}S_u(k,\omega)d\omega\\ &= \frac{\Gamma(\alpha_t-1/2)}{(4\pi)^{1/2}\Gamma(\alpha_t)\tau\rho(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\lambda_k^{2\alpha_t-1}}\end{aligned} \tag{5.6}$$

which has intrinsicness at $S_u(0)$. This marginal spatial process has local variance inversely proportional to $\tau\rho$.

5.1.2 Covariance

As the spatial domain is the sphere, the space-time covariance can be computed considering Legendre polynomials, Wahba (1981), as

$$\mathrm{Cov}(\mathbf{s}, \mathbf{s}^{'}, t, t^{'}) = \sum_{k=0}^{\infty} \frac{2k+1}{4\pi} P_k(\mathbf{s}^T \mathbf{s}^{'}/R) \int_{\mathbf{R}} S_u(k,\omega) \mathrm{e}^{i\omega(t^{'}-t)} d\omega \tag{5.7}$$

where $\mathbf{s}^T\mathbf{s}'$ is the cosine of the angle between the points $\mathbf{s}$ and $\mathbf{s}'$, R is the radius of the sphere (we will consider $R = 1$) and $P_k(\cdot)$ are the Legendre polynomials.

Since Eq. (5.6) has intrinsicness at $\boldsymbol{S}_u(0)$, we can consider the covariance for the non-intrinsic part, that is considering $k > 0$. It does correspond to the covariance for a constrained process $\boldsymbol{Z}$, for example, constrained to sum-to-zero. The space-time covariance for this case is

$$\frac{\Gamma(\alpha_t - 1/2)}{8\pi\sqrt{\pi}\Gamma(\alpha_t)\tau\rho} \sum_{k=1}^{\infty} \frac{(2k+1)P_k(\mathbf{s}^T\mathbf{s}^{'})}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\lambda_k^{2\alpha_t-1}} \mathcal{M}_{\alpha_t-1/2}(|t-t'|, \lambda_k/\rho) \tag{5.8}$$

where $\mathcal{M}_\nu(h,\kappa)$ is the Matérn correlation function given by

$$\mathcal{M}_\nu(h,\kappa) = \frac{(h\kappa)^\nu K_\nu(h\kappa)}{2^{\nu-1}\Gamma(\nu)} \tag{5.9}$$

where h is the distance (or lag), κ is the scale parameter (inverse of the range parameter), ν is the smoothness parameter and $K_\nu(.)$ is the *modified Bessel function of the second kind* of order ν, Lindgren (2013), page 226.

For the particular case of $\alpha_t = 1$, we have an exponential time covariance for each spatial frequency. In this case the space-time covariance is

$$\frac{1}{8\pi\tau\rho} \sum_{k=1}^{\infty} \frac{(2k+1)P_k(\mathbf{s}^T\mathbf{s}^{'})}{\lambda_k(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}} \mathrm{e}^{-|t-t'|\lambda_k/\rho} \tag{5.10}$$

We can see the space-time covariance function for the constrained to sum-to-zero process shown in Figure 5.1. The ρ and $\gamma_{\mathcal{E}}$ parameters are temporal range and spatial range parameters. Thus, increasing these parameter slow the decay of the correlation. However, as the model is not separable, we also have that increasing $\gamma_{\mathcal{E}}$ decreases the range of the temporal correlation.

One can consider the marginal properties of the process in order to have a better understanding of the process. This can be done by analyzing the spatial covariance at a time point, the temporal covariance at a spatial location, and the marginal variance as well. One can get the marginal properties

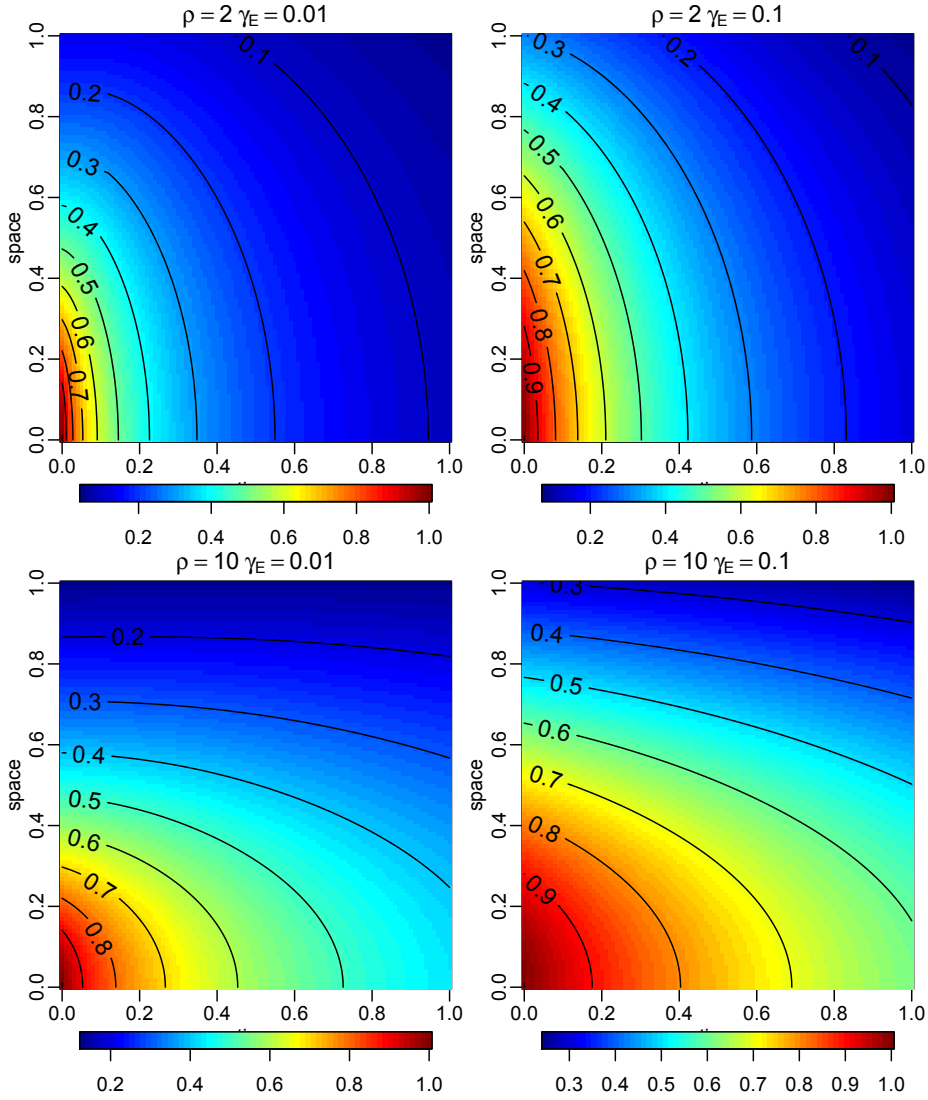


Figure 5.1: Space-time correlation for some values of ρ and $\gamma_{\mathcal{E}}$.

by using the marginal spectral densities or by considering $\mathbf{s} = \mathbf{s}'$ or $t = t'$ in Eq. (5.7). The marginal covariance can be computed considering $\mathbf{s} = \mathbf{s}'$ and $t = t'$.

As α_t and $\alpha_{\mathcal{E}}$ are usually fixed, we will consider a few combinations of the smoothness parameters, that is considering the four combinations from the $\alpha_t \in \{1, 2\}$ and $\alpha_{\mathcal{E}} \in \{1, 2\}$.

Marginal variance

We can compute the marginal variance considering the marginal (Matérn) variance for each spatial frequency, that is $v_k = \sigma_k^2 / a_k$ where

$$\sigma_k^2 = 1/(\tau(1 + \gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\rho^{2\alpha_t})$$

and

$$a_k = \frac{2\sqrt{\pi}\Gamma(\alpha_t)}{\Gamma(\alpha_t - 1/2)} \left(\frac{\lambda_k}{\rho}\right)^{2(\alpha_t - 1/2)} .$$

giving

$$v_k = \frac{\Gamma(\alpha_t - 1/2)}{2\sqrt{\pi}\Gamma(\alpha_t)\tau\rho} \frac{1}{(1 + \gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\lambda_k^{2\alpha_t - 1}} . \tag{5.11}$$

For $\alpha_t = 1$, v_k term simplifies to

$$\frac{1}{2\tau\rho} \frac{1}{\lambda_k(1 + \gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}} .$$

One can then use the marginal variance in equation 5.8 for $\mathbf{s} = \mathbf{s}'$ and $t = t'$. Then one has the non-intrinsic part of the marginal variance as

$$\sum_{k=1}^{\infty} \frac{2k+1}{4\pi} v_k = \frac{\Gamma(\alpha_t - 1/2)}{8\pi\sqrt{\pi}\Gamma(\alpha_t)\tau\rho} \sum_{k=1}^{\infty} \frac{2k+1}{(1 + \gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\lambda_k^{2\alpha_t - 1}} \tag{5.12}$$

where we have contributions from all the model parameters and that τ and ρ contribute in the same way. Due to this fact, we have plotted this marginal variance as a function of $\tau\rho$ in Figure 5.2.

Each plot in Figure 5.2 considers a different combination of the smoothness parameters α_t and $\alpha_{\mathcal{E}}$. The plots in this figure show the log linear relationship between $\tau\rho$ and the marginal variance. The τ, ρ and $\gamma_{\mathcal{E}}$ parameters are all inversely related to the marginal variance (they are in the denominator of (5.12). The decay rate is the same for different values of $\gamma_{\mathcal{E}}$ and for different cases of α_t and $\alpha_{\mathcal{E}}$. When changing from $\alpha_t = 1$ to $\alpha_t = 2$, the impact from $\gamma_{\mathcal{E}}$ is bigger.

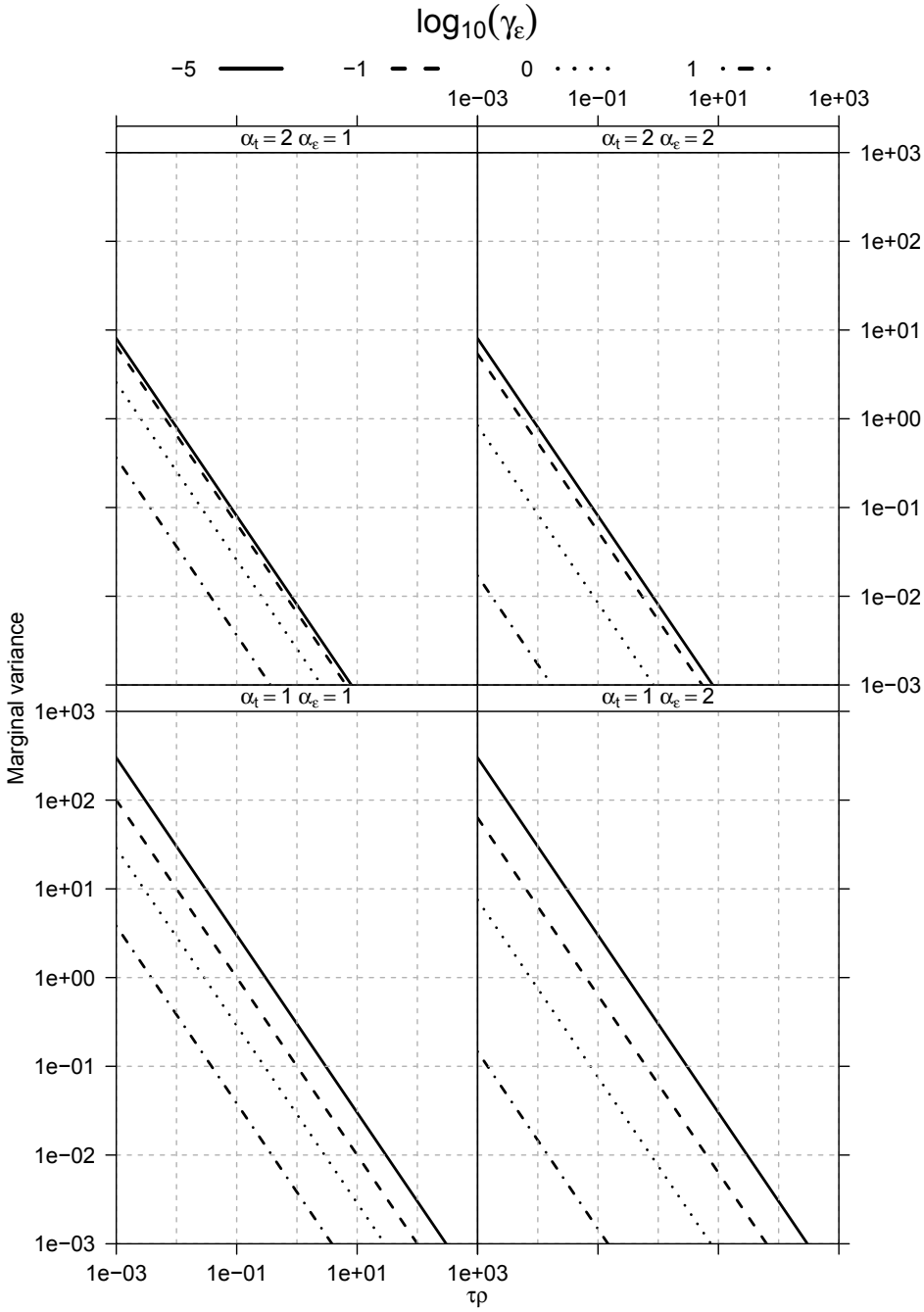


Figure 5.2: Marginal variance as a function of the model parameters.

Marginal spatial correlation

The spatial covariance at a point can be computed from (5.8) considering $t = t^{'}$, which is the same as the marginal spatial covariance. The non-intrinsic marginal spatial covariance part is

$$\text{Cov}(\mathbf{s}, \mathbf{s}^{'}, t, t) = \sum_{k=1}^{\infty} \frac{2k+1}{4\pi} P_k(\mathbf{s}^T \mathbf{s}^{'}) v_k \tag{5.13}$$

where v_k is the variance from the Matérn temporal covariance at spatial frequency k, Eq. (5.11). Notice that the v_k term is a function of all the model parameters.

It is useful to look at the correlation function instead. When computing the correlation, the $1/(4\pi)$ term in Eq. (5.8) cancels. Eq (5.11) shows that the term $\frac{\Gamma(\alpha_t - 1/2)}{2\sqrt{\pi}\Gamma(\alpha_t)\tau\rho}$ does not depend on k and also cancels. The remaining term from (5.11) is $1/[\lambda_k^{2\alpha_t - 1}(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}]$. It shows that the spatial marginal correlation does not depends on τ neither on ρ.

The spatial marginal correlation non-intrinsic part can be written as

$$\left[\sum_{k=1}^{\infty} \frac{(2k+1)P_k(\mathbf{s}^T\mathbf{s}')}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\lambda_k^{2\alpha_t - 1}}\right] \left[\sum_{k=1}^{\infty} \frac{(2k+1)}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\lambda_k^{2\alpha_t - 1}}\right]^{-1}. \tag{5.14}$$

The Legendre polynomial $P_1(\cdot)$ dominates the other order polynomials and the correlation can assume negative values, as shown in Figure 5.3.

Each plot in Figure 5.3 considers a different combination of the α_t and $\alpha_{\mathcal{E}}$ parameters. Different values of $\gamma_{\mathcal{E}}$ were considered in each plot. The horizontal dashed gray lines were drawn at 0 and 0.139 to a practical reference as the Matérn correlation at the distance of $\sqrt{8\nu}/\kappa$ is near 0.139.

The plots in Figure 5.3, shows that the spatial marginal correlation can assume negative values, due to behavior of the Legendre polynomials. Additionally, there is no short spatial range which is explained by the fact that the spatial operator in Eq. (5.1) has a coefficient equal to one, which is relative to the unit radius sphere. However, what we actually see in Figure 5.3 is in combination with the spatial range from the driving noise, which is $\sqrt{8(\alpha_{\mathcal{E}} - 1)}$. One can shorten the spatial range for the case when $\alpha_t = 1$, considering $\gamma_{\mathcal{E}}$ values up to 1, as for bigger values the pattern is similar to the $\alpha_t = 2$ case. When $\alpha_t = 2$ the spatial operator from the heat equation dominates.

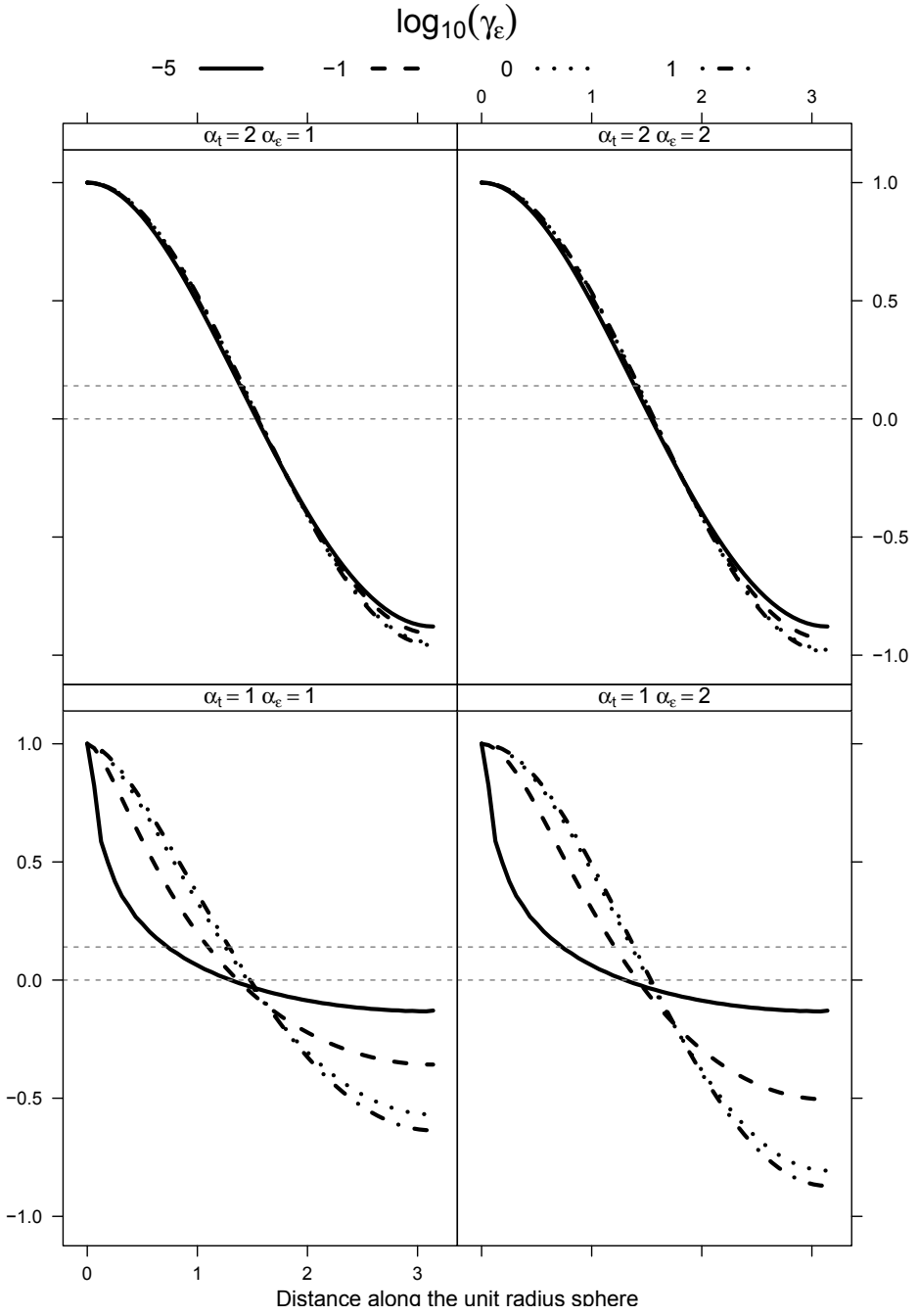


Figure 5.3: Marginal spatial correlation as a function of the great circle distance for some combination of the model parameters.

Marginal temporal correlation

The temporal covariance at a point, or the marginal temporal covariance, can be computed from (5.8) considering $s = s'$, which gives $P_k(\mathbf{s}^T\mathbf{s}) = 1$. Dividing it by the marginal variance gives the temporal correlation at a spatial point, or the marginal temporal correlation. The non-intrinsic part of the marginal temporal correlation is

$$\left[\sum_{k=1}^{\infty} \frac{(2k+1)\mathcal{M}_{\alpha_t\text{-}\frac{1}{2}}(|t\text{-}t'|, \frac{\lambda_k}{\rho})}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\lambda_k^{2\alpha_t\text{-}1}}\right]\left[\sum_{k=1}^{\infty} \frac{(2k+1)}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\lambda_k^{2\alpha_t\text{-}1}}\right]^{-1} \tag{5.15}$$

and we can see that only τ does not affect it. In Figure 5.4 the marginal temporal corelation is shown as a function of the time lag for some combination of the other parameters.

In Figure 5.4 we have plots for different values of α_t and $\alpha_{\mathcal{E}}$. The lines were drawn considering different combinations between $\gamma_{\mathcal{E}}$ and ρ. For $\alpha_t = 2$ there is almost no influence from $\gamma_{\mathcal{E}}$. When $\alpha_t = 1$, a smaller $\gamma_{\mathcal{E}}$ means the decay of the marginal temporal correlation is steeper.

One of the horizontal dashed gray lines is drawn at 0.139, in order to relate to the Matérn practical range. It depends directly on ρ and for $\alpha_t = 2$, it is slightly larger than 8 for $\rho = 5$ and below 2 for $\alpha_t = 1$. However, for $\alpha_t = 1$ it depends considerably on $\gamma_{\mathcal{E}}$, that is when $\gamma_{\mathcal{E}}$ increases, the marginal temporal range also increases.

The $\alpha_t = 1$ and $\alpha_{\mathcal{E}} = 2$ case

We now have a closer look at the case when $\alpha_t = 1$ and $\alpha_{\mathcal{E}} = 2$. The non-separability is seen in this case in the effect of $\gamma_{\mathcal{E}}$ on the temporal marginal correlation. Different values of $\gamma_{\mathcal{E}}$ are considered for each value of ρ, shown in each plot of Figure 5.5.

In each plot in Figure 5.5, when $\gamma_{\mathcal{E}} = 1$, the practical temporal marginal range is a bit lower than ρ and when $\gamma_{\mathcal{E}}$ decreases, it also decreases. For $\gamma_{\mathcal{E}} = 0.1$, the marginal temporal range is near 80% of ρ, for $\gamma_{\mathcal{E}} = 0.01$ it is near $\rho/2$ and for $\gamma_{\mathcal{E}} = 10^{-5}$, it is a bit bigger than $\rho/5$.

5.1.3 Marginal properties to model parameters

The spatial range is not affected by τ and ρ and a mapping from the spatial range to $\gamma_{\mathcal{E}}$ is one-to-one as shown in Figure 5.6.

We also notice limits in the spatial marginal practical range in Figure 5.3. Before making the map we visualize the limits as a function of τ in the left

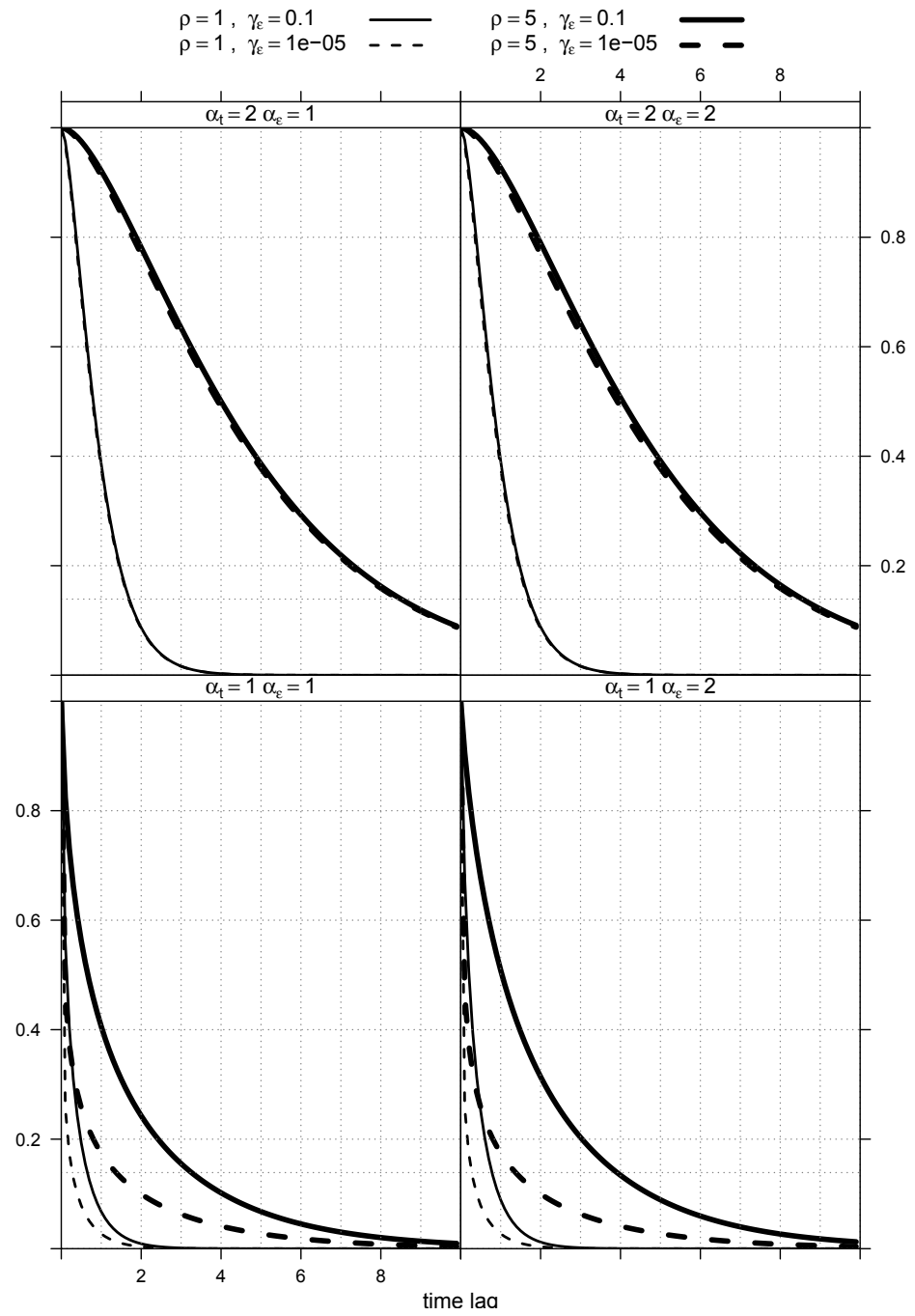


Figure 5.4: Marginal temporal correlation as a function of the temporal lag for some values of the model parameters.

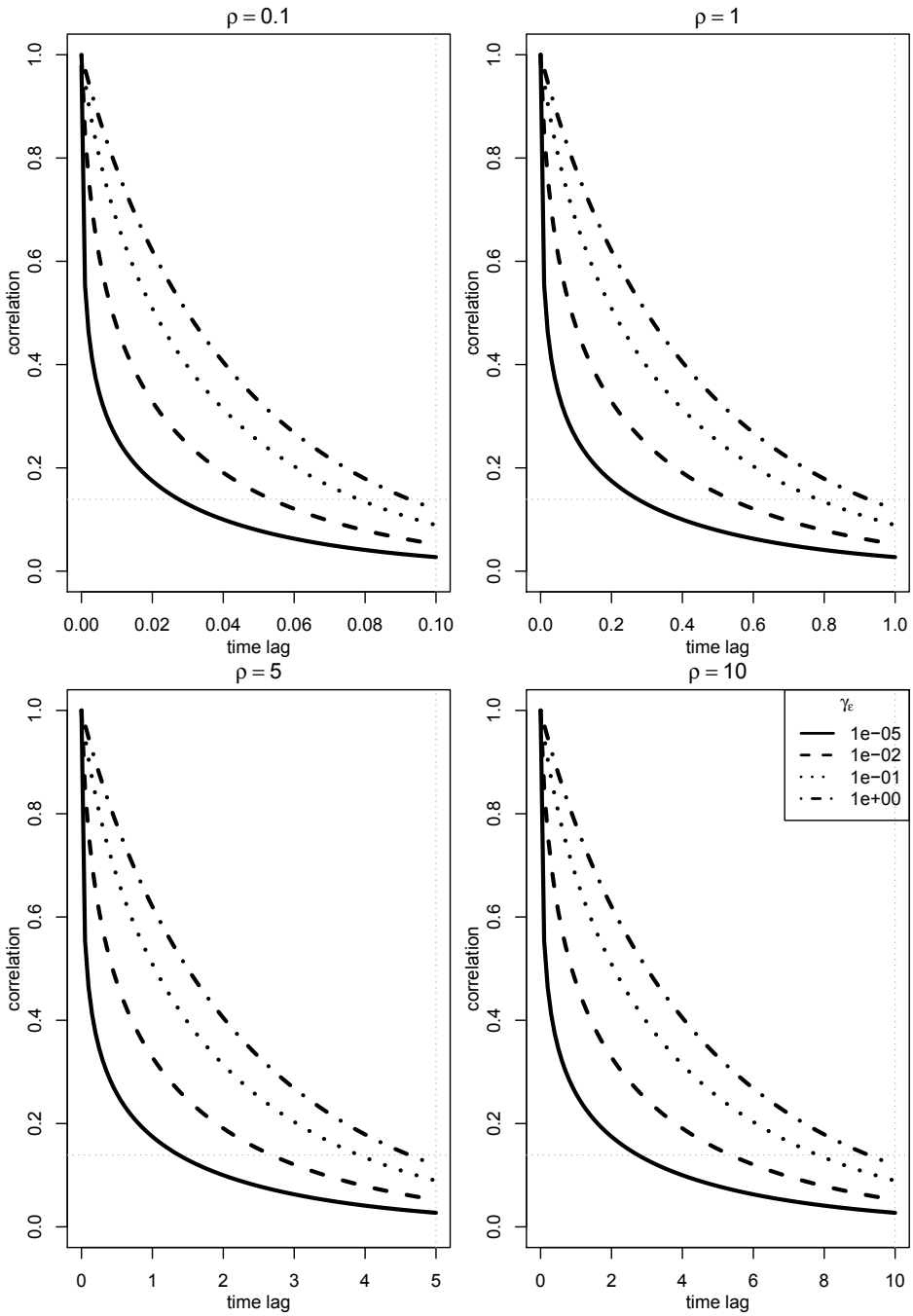


Figure 5.5: Marginal temporal correlation for the $\alpha_t = 1$ and $\alpha_{\mathcal{E}} = 2$ case.

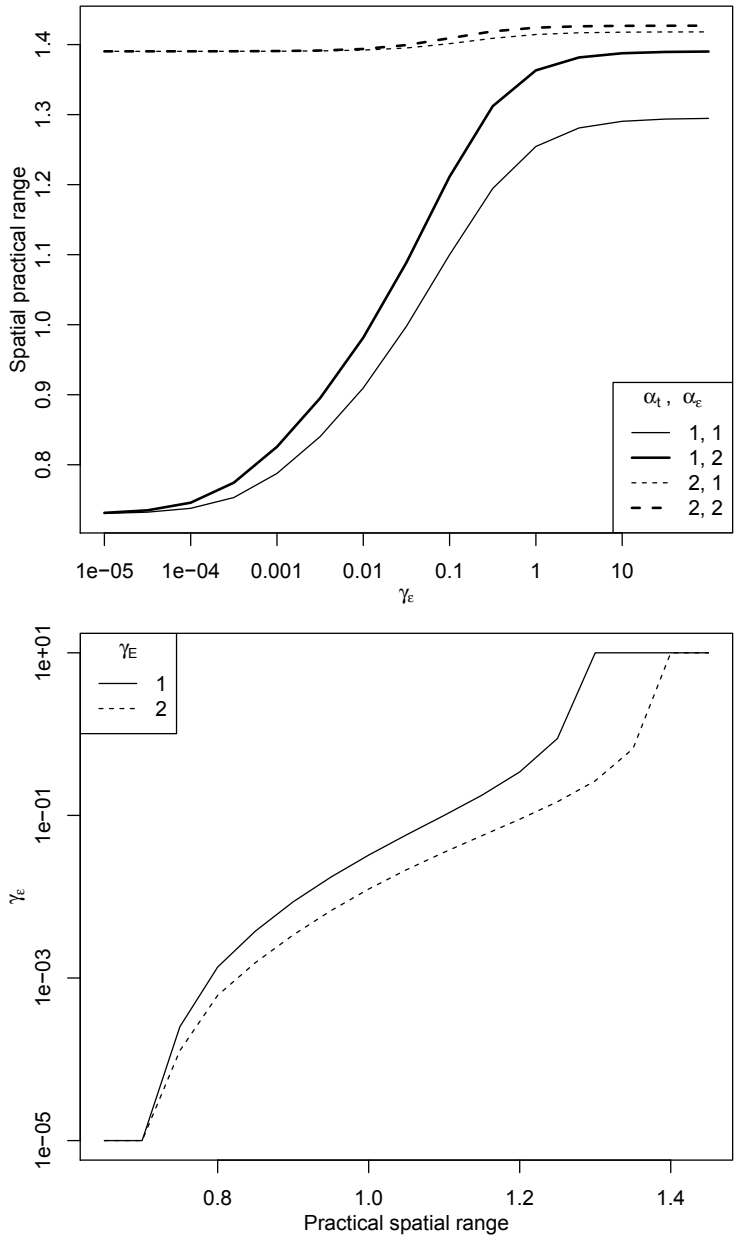


Figure 5.6: Marginal spatial practical range as function of $\gamma_{\mathcal{E}}$ (left) and map from the marginal spatial practical range to $\gamma_{\mathcal{E}}$ for $\alpha_t = 1$ (right).

plot in Figure 5.6. For the case of $\alpha_t = 2$, the range is always near 1.4. When $\alpha_t = 1$ the minimum range is around 0.75, and when $\alpha_{\mathcal{E}} = 1$ the maximum range is around 1.295 and when $\alpha_{\mathcal{E}} = 2$ it is around 1.39.

The function to map from the marginal practical spatial range considering $\alpha_t = 1$ is shown in the right plot in Figure 5.6. $\gamma_{\mathcal{E}}$ has considerable influence in the temporal practical range, for $\alpha_t = 1$. That is lower values for $\gamma_{\mathcal{E}}$ imply a steeper decay in both spatial and temporal marginal correlation.

The temporal range depends on $\gamma_{\mathcal{E}}$, mainly when $\alpha_t = 1$ and we plot the mapping for different values of $\gamma_{\mathcal{E}}$ and each of the four smoothness scenarios in Figure 5.7.

The practical temporal range as a function of ρ is approximately a line. In log scale the practical temporal range for both axes is actually a line, as shown in Figure 5.8. Therefore, the mapping can be considered by a linear regression in log scale, where the coefficient is a function of $\gamma_{\mathcal{E}}$.

The mapping procedure is as follows. First, choose $\gamma_{\mathcal{E}}$ from the desired marginal spatial range. Then use it to choose ρ given the desired marginal temporal range. At the end, consider ρ and $\gamma_{\mathcal{E}}$ to find τ given a desired marginal variance. This last mapping is shown in Figure 5.9 for some values of ρ and $\gamma_{\mathcal{E}}$, for the four scenarios from the smoothness parameters.

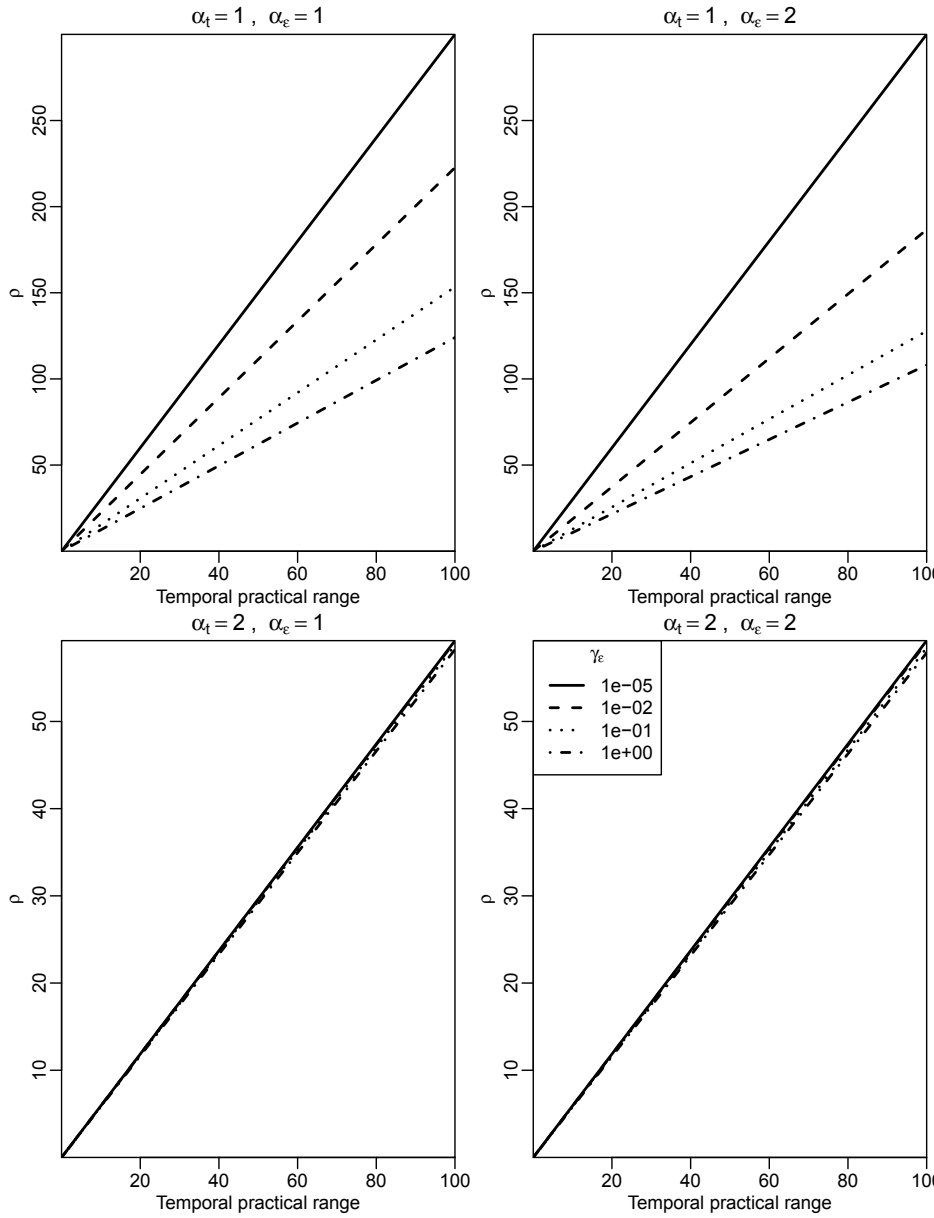


Figure 5.7: Map from temporal practical range to ρ.

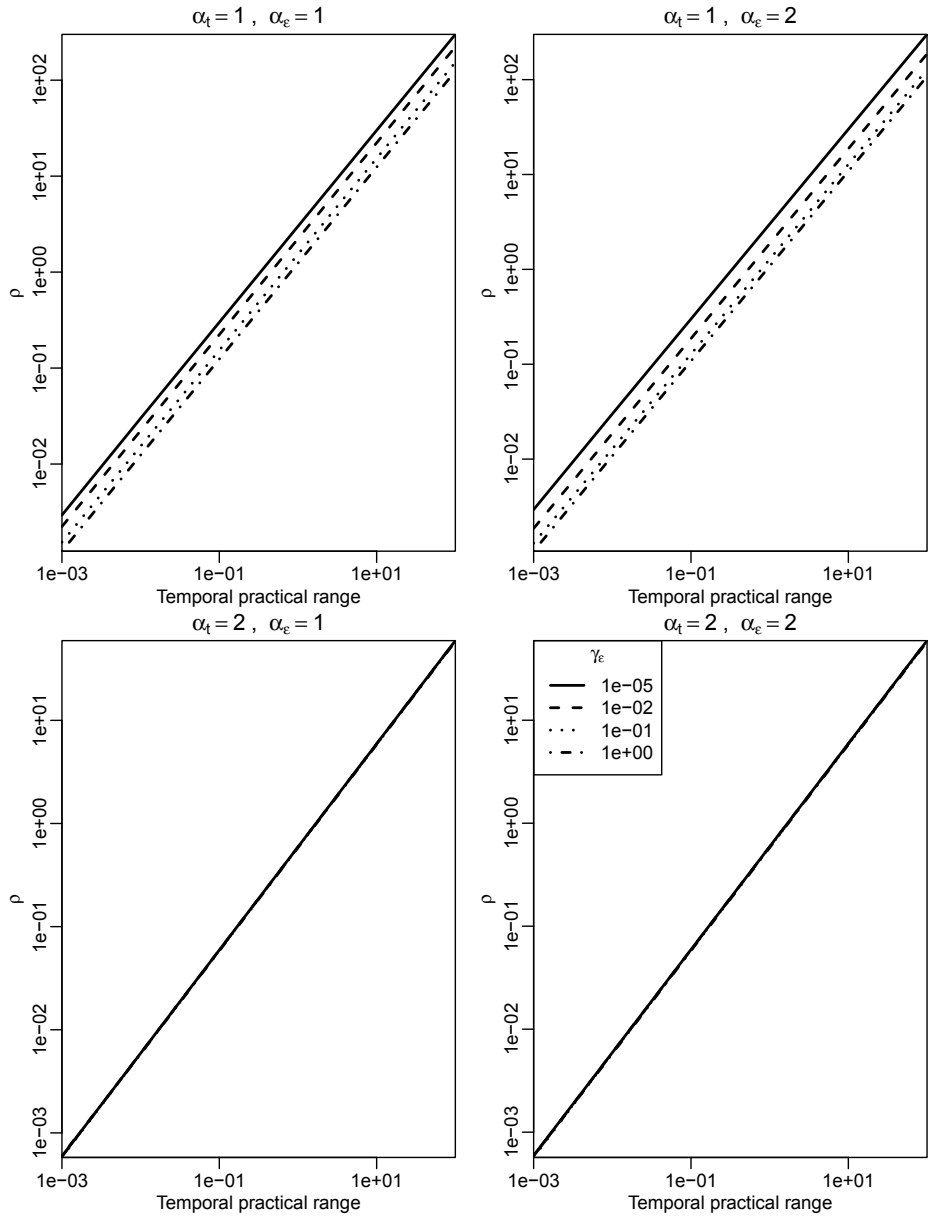


Figure 5.8: Map from temporal practical range to ρ in log scale.

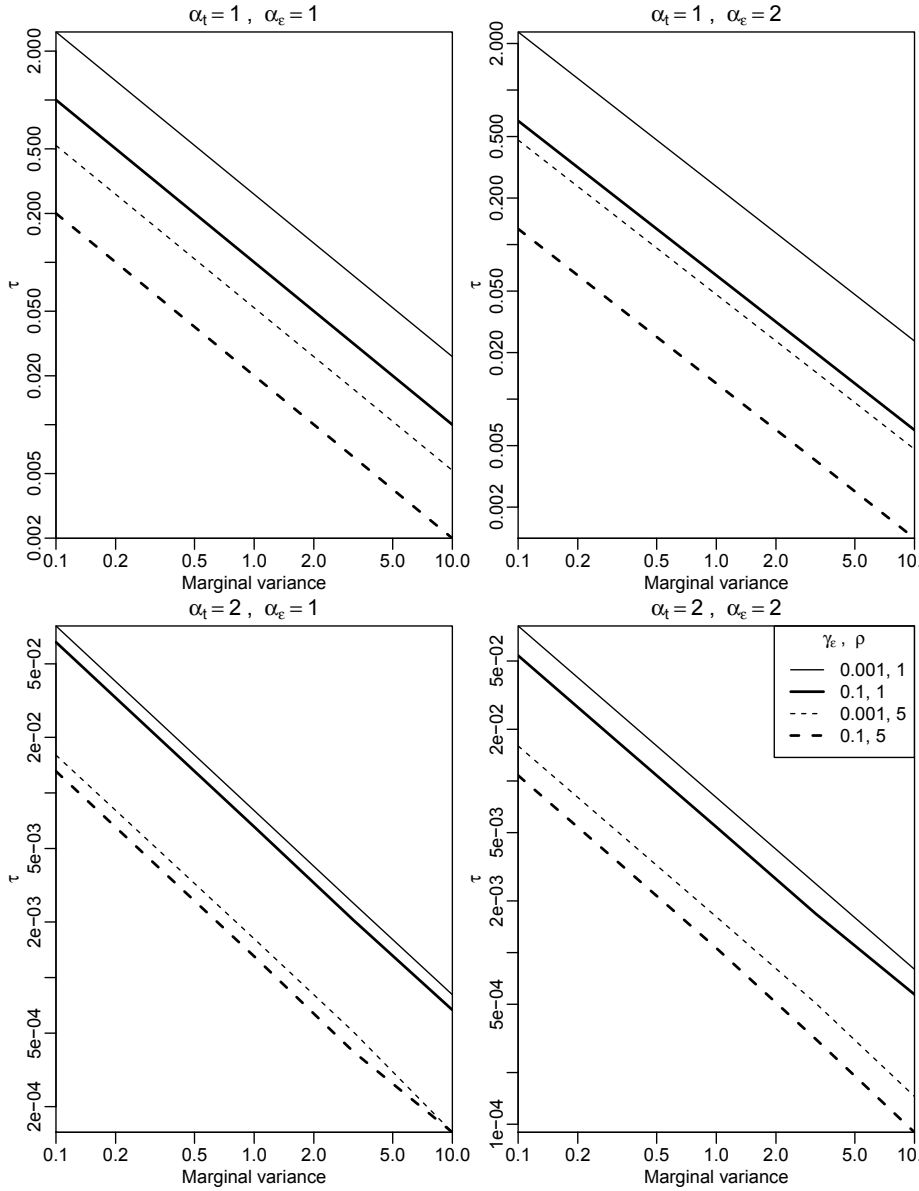


Figure 5.9: Mapping from marginal variance to τ.

5.2 Damped iterated heat equation

In this section we consider a model based in the iterated damped stationary heat equation with a stationary spatially correlated driving noise. The SPDE can be written as

$$\begin{aligned} \tau(\rho\frac{\partial}{\partial t} + 1 - \gamma_x\Delta)^{\alpha_t} u(\mathbf{s},t) &= \mathcal{E}(\mathbf{s},t) && (5.16)\\ (1-\gamma_{\mathcal{E}}\Delta)^{\alpha_{\mathcal{E}}/2}\mathcal{E}(\mathbf{s},\delta t) &= \mathcal{W}_{\mathcal{E}}(\mathbf{s},\delta t) && (5.17)\end{aligned}$$

where $\mathcal{W}_{\mathcal{E}}$ is a space-time white noise process with unit variance, $\mathcal{E}(\mathbf{s},t)$ is a stationary process noise not correlated over time (white in time) and and spatially correlated (coloured), α_t and $\alpha_{\mathcal{E}}$ are smoothness parameters which will usually be fixed when estimating the other model parameters, and τ, ρ, γ_x and $\gamma_{\mathcal{E}}$ parameters are real positives.

We are primarily insterested in the τ, ρ, γ_x and $\gamma_{\mathcal{E}}$ parameters. τ is the overall precision, ρ is the time persistency parameter and both γ_x and $\gamma_{\mathcal{E}}$ are spatial range parameters. We can also consider a mapping from these parameters to the practical parameters for the marginal variance, the temporal range and the spatial range to better understand the process marginal properties. However, before we go in this direction we consider the spectral density.

5.2.1 Spectral densities

If the sphere is the spatial domain, we have discrete spatial frequencies k, $k = 0, 1, ...$ and $\lambda_k = k(k+1)$. With continuous time domain and temporal frequencies $\omega \in \mathbb{R}$, the spectral representation is given by

$$\begin{aligned} \tau(\rho|i\omega| + 1 + \gamma_x\lambda_k)^{\alpha_t}\hat{z}(k,\omega) &= \hat{\mathcal{E}}(k,\omega)\\ (1-\gamma_{\mathcal{E}}\lambda_k)^{\alpha_t/2}\hat{\mathcal{E}}(k,\omega) &= \hat{\mathcal{W}}_{\mathcal{E}}(k,\omega)\,.\end{aligned}$$

The spectral densities are obtained as follows

$$\begin{aligned} \tau(\rho^2\omega^2 + (1+\gamma_x\lambda_k)^2)^{\alpha_t} S_u(\lambda_k,\omega) &= S_{\mathcal{E}}(\lambda_k,\omega)\\ (1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}} S_{\mathcal{E}}(\lambda_k,\omega) &= \frac{1}{2\pi}\end{aligned}$$

giving the spectral density as

$$\begin{aligned} S_u(\lambda_k,\omega) &= \frac{1}{2\pi\tau(\rho^2\omega^2 + (1+\gamma_x\lambda_k)^2)^{\alpha_t}(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}} && (5.18)\\ &= \frac{1}{\tau(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\rho^{2\alpha_t}}\frac{1}{2\pi(\omega^2 + (1+\gamma_x\lambda_k)^2/\rho^2)^{\alpha_t}} && (5.19)\end{aligned}$$

where for each spatial frequency k we have a Matérn temporal covariance with smoothness α_t and scale parameter $(1 + \gamma_x\lambda_k)/\rho$, Lindgren (2013), page 226. As $(1 + \gamma_x\lambda_k)/\rho)$ is the corresponding range parameter, we have smaller temporal range for higher spatial frequencies. Therefore, we also have in this model the fact that the local spatial behaviour has less impact on the temporal range. However, the damping term does attenuate the effect of γ_x in this behaviour for this model.

The marginal temporal spectral density can be computed considering Eq. (5.19) and by summing over the spatial frequencies

$$\begin{aligned} S_u(\omega) &= \sum_{k=0}^{\infty} \frac{2k+1}{4\pi} S_u(k,\omega) \\ &= \frac{1}{8\pi^2\tau\rho^{2\alpha_t}} \sum_{k=0}^{\infty} \frac{2k+1}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(\omega^2 + \frac{(1+\gamma_x\lambda_k)^2}{\rho^2})^{\alpha_t}} . \end{aligned} \tag{5.20}$$

The marginal spatial spectral density is obtained by integrating with respect to the temporal domain

$$\begin{aligned} S_u(k) &= \int_{\mathbf{R}} S_u(k,\omega) d\omega \\ &= \frac{\Gamma(\alpha_t - 1/2)}{(4\pi)^{1/2}\Gamma(\alpha_t)\tau\rho(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(1+\gamma_x\lambda_k)^{2\alpha_t-1}}. \end{aligned} \tag{5.21}$$

One can see that when $\gamma_x = \gamma_{\mathcal{E}}$ the marginal spatial covariance is Matérn with range equals to γ_x and smoothness equals to $\alpha_{\mathcal{E}} + 2\alpha_t$.

Therefore, all the knowledge about the Matérn covariance function can be used to work with this model. This knowledge is very useful because one can use it to define model parameters for doing simulations from the model. Additionally, one can think on this marginal property in order to define prior distribution for the spatial range parameter. The $\tau\rho$ term acts like the inverse of the variance parameter in the Matérn covariance function.

5.2.2 Covariance

We can compute the space-time covariance using Eq. (5.7) where we can now consider the sum from $k = 0$. The space-time covariance considering spatial locations $\boldsymbol{s}$ and $\boldsymbol{s}'$ and time points t and t' is

$$\frac{\Gamma(\alpha_t - 1/2)}{8\pi\sqrt{\pi}\Gamma(\alpha_t)\tau\rho} \sum_{k=0}^{\infty} \frac{(2k+1)P_k(\mathbf{s}^T\mathbf{s}')\mathcal{M}_{\alpha_t-\frac{1}{2}}(|t-t'|, (1+\gamma_x\lambda_k)/\rho)}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(1+\gamma_x\lambda_k)^{2\alpha_t-1}} \tag{5.22}$$

which is the sum of temporal Matérn temporal correlations for each k considering a scale equal to $(1 + \gamma_x\lambda_k)/\rho$.

When $\alpha_t = 1$ the space-time covariance simplifies to

$$\frac{1}{8\pi\tau\rho}\sum_{k=0}^{\infty}\frac{(2k+1)P_k(\mathbf{s}^T\mathbf{s}')}{(\gamma_x^2+\lambda_k)(\gamma_{\mathcal{E}}^2+\lambda_k)^{\alpha_{\mathcal{E}}}}\mathrm{e}^{-|t-t'|(\gamma_x^2+\lambda_k)/\rho} \tag{5.23}$$

as we have an exponential correlation function for each spatial frequency.

We can see the correlation plot shown in Figure 5.10 considering $\alpha_t = 1$ and $\alpha_{\mathcal{E}} = 2$ and $\gamma_x = \gamma_{\mathcal{E}}$ with some configuration of $\gamma_{\mathcal{E}}$ and ρ. These range parameters are easy to interpret as they are directly related to the decay of the correlation function in each direction and this relation is very related to what happens in the Matérn correlation function. Additionally, from the marginal spatial spectral density we have that the ρ parameter does not affects the marginal spatial covariance and there is no effect from ρ in the correlation decay along the time lag. This is not the case for the marginal temporal spectral density as we have γ_x and $\gamma_{\mathcal{E}}$ inside the summation in Eq. (5.20). However, the effect from these parameters are not big in the plots in Figure 5.10.

Marginal variance

We can compute the marginal variance considering the marginal (Matérn) variance for each spatial frequency, that is $v_k = \sigma_k^2/a_k$ where

$$\sigma_k^2 = 1/(\tau(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}\rho^{2\alpha_t})$$

and

$$a_k = \frac{2\sqrt{\pi}\Gamma(\alpha_t)}{\Gamma(\alpha_t - 1/2)}\left(\frac{1+\gamma_x\lambda_k}{\rho}\right)^{2(\alpha_t-1/2)}.$$

giving

$$v_k = \frac{\Gamma(\alpha_t - 1/2)}{2\sqrt{\pi}\Gamma(\alpha_t)\tau\rho}\frac{1}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(1+\gamma_x\lambda_k)^{2\alpha_t-1}}. \tag{5.24}$$

For $\alpha_t = 1$, v_k term simplifies to

$$\frac{1}{2\tau\rho}\frac{1}{(1+\gamma_x\lambda_k)(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}}.$$

Considering Eq. (5.22) for $\mathbf{s} = \mathbf{s}'$ and $t = t'$, the marginal variance is

$$\sum_{k=0}^{\infty}\frac{2k+1}{4\pi}v_k = \frac{\Gamma(\alpha_t - 1/2)}{8\pi\sqrt{\pi}\Gamma(\alpha_t)\tau\rho}\sum_{k=1}^{\infty}\frac{2k+1}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(1+\gamma_x\lambda_k)^{2\alpha_t-1}} \tag{5.25}$$

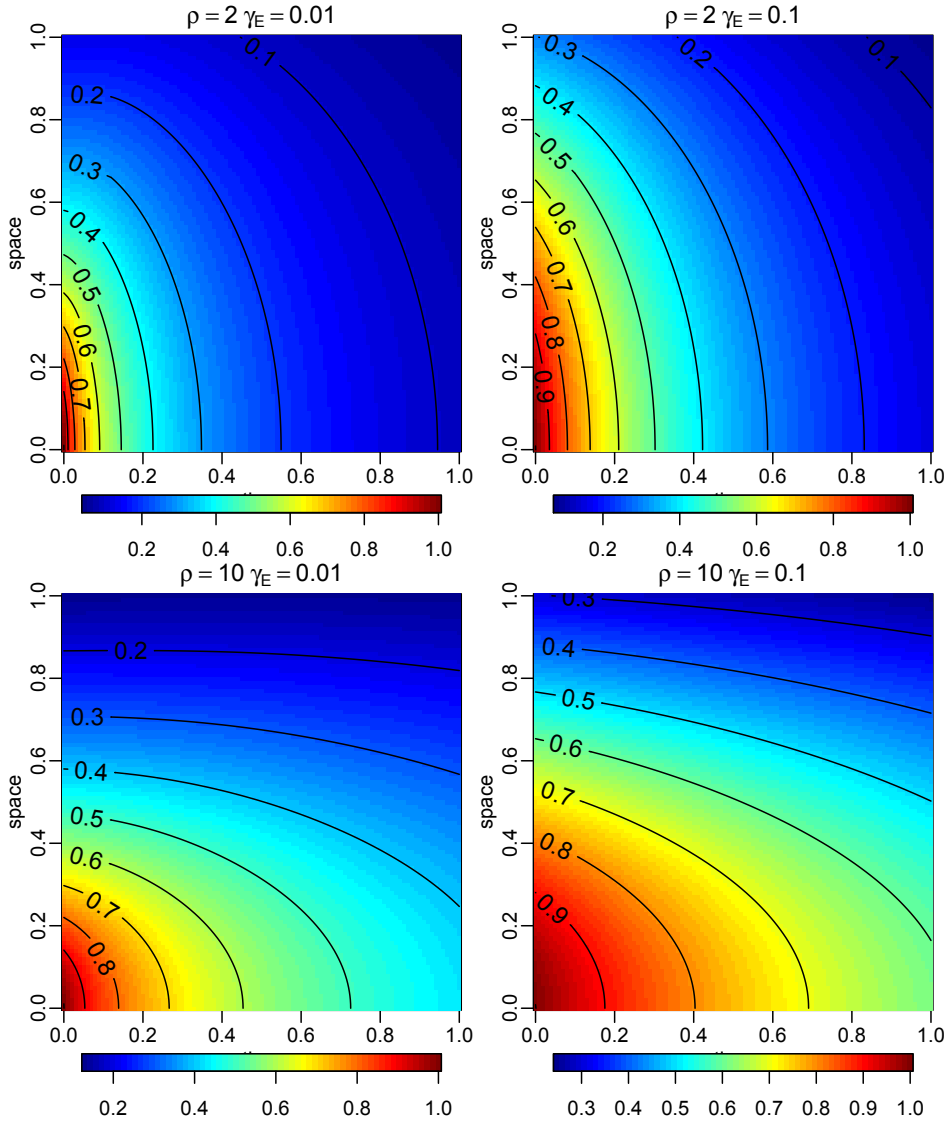


Figure 5.10: Space-time correlation for the case of $\alpha_t = 1$ and $\alpha_{\mathcal{E}} = 2$ for some values of the other model parameters.

where we have contributions from all the model parameters. This is indirectly proportional to γ_x and $\gamma_{\mathcal{E}}$, directly inversely proportional to $\rho\tau$. Therefore, we show the log of the marginal variance as a function of $\log(\tau\rho)$ in Figure 5.11.

Each plot in Figure 5.11 considers a different combination of the smoothness parameters α_t and $\alpha_{\mathcal{E}}$. The plots in this figure shows the log linear relationship between $\tau\rho$ and the marginal variance. The decay rate is the same for different values of $\gamma_{\mathcal{E}}$ and for different cases of α_t and $\alpha_{\mathcal{E}}$. When changing from $\alpha_t = 1$ to $\alpha_t = 2$, the impact from $\gamma_{\mathcal{E}}$ is greater. The important point here is to notice that for $\tau\rho$ near one the marginal variance is also near one.

Marginal spatial correlation

The spatial covariance at a point can be computed from (5.22) considering $t = t^{'}$, which is the marginal spatial covariance. This can be computed with

$$\text{Cov}(\mathbf{s}, \mathbf{s}^{'}, t, t) = \sum_{k=0}^{\infty} \frac{2k+1}{4\pi} P_k(\mathbf{s}^T \mathbf{s}^{'}) v_k \tag{5.26}$$

where v_k is the variance from the Matérn temporal covariance at spatial frequency k, Eq. (5.24). Notice that the v_k term is a function of all the model parameters.

It is useful to look at the correlation function instead. When computing the correlation, the $1/(4\pi)$ term in Eq. (5.22) cancels. In eq (5.24) the term $\frac{\Gamma(\alpha_t - 1/2)}{2\sqrt{\pi}\Gamma(\alpha_t)\tau\rho}$ does not depend on k and also cancels. The remaining term from (5.24) is $1/[(1+\gamma_x\lambda_k)^{2\alpha_t - 1}(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}]$. It means that the spatial marginal correlation does not depends on τ neither on ρ.

The spatial marginal correlation between locations $\boldsymbol{s}$ and $\boldsymbol{s'}$ can be written as

$$\left(\sum_{k=0}^{\infty} \frac{(2k+1)P_k(\mathbf{s}^T\mathbf{s}')}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(1+\gamma_x\lambda_k)^{2\alpha_t 1\text{-}1}}\right)\left(\sum_{k=0}^{\infty} \frac{(2k+1)}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(1+\gamma_x\lambda_k)^{2\alpha_t 1\text{-}1}}\right)^{-1} \tag{5.27}$$

where the Legendre polynomials $P_k(\cdot)$ are evaluated considering the cosine of the distance in a path along the sphere. The shortest path along a unit radius sphere (the actual distance) is from 0 to π. The Legendre polynomials can assume negative values, $P_k(.) \in (-1, 1)$. However, for $k = 0$ it is constant equals to 1 which corresponds to the lowest λ_k and this dominates the summation terms giving an positive correlation values, as shown in Figure 5.12.

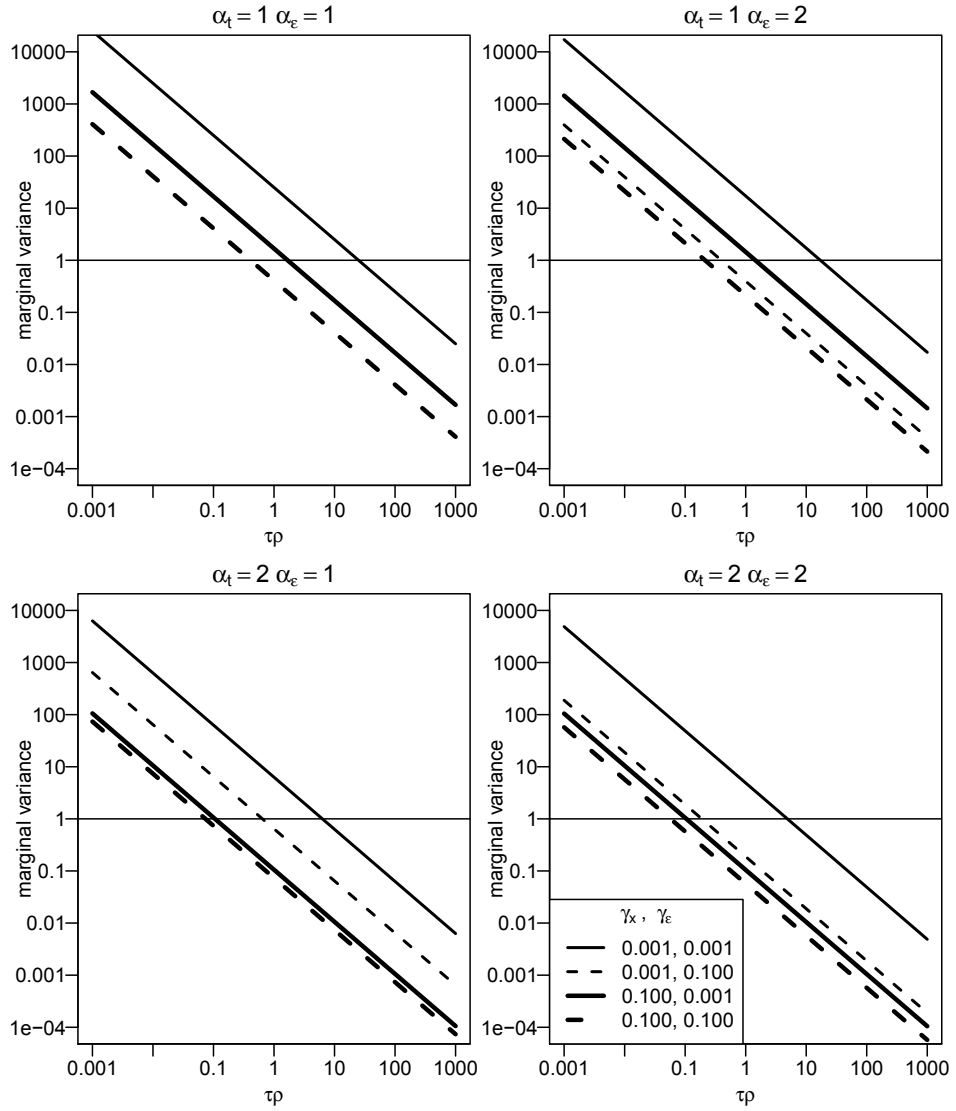


Figure 5.11: Marginal variance as a function of $\tau\rho$ for some values of the other model parameters.

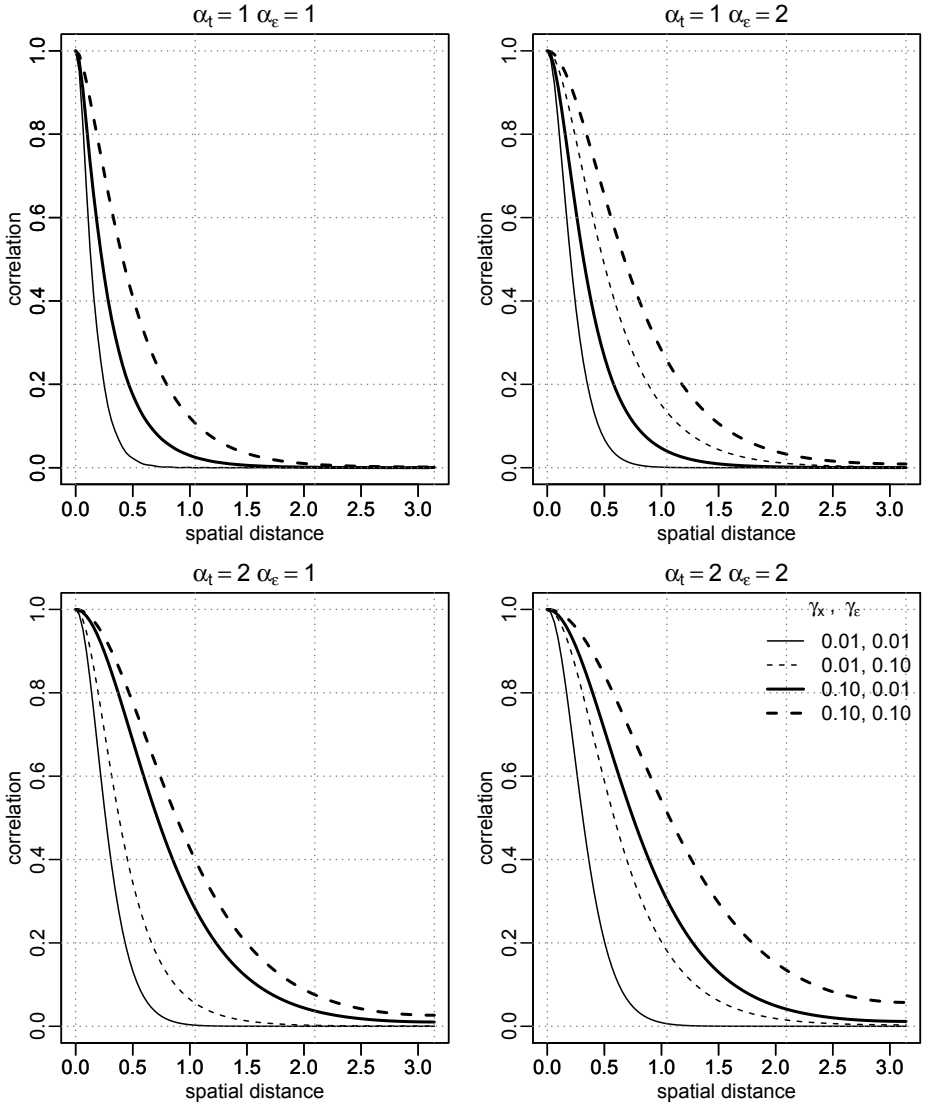


Figure 5.12: Marginal spatial correlation as a function of the great circle distance for some combination of the model parameters.

Each plot in Figure 5.12 considers a different combination of the α_t and $\alpha_{\mathcal{E}}$ parameters. Different values of $\gamma_{\mathcal{E}}$ were considered in each plot. We can see how flexible is the marginal spatial correlation as we have great change in its decay by changing the range parameters. The decay is also affected by the smoothness parameters. Notice that for $\alpha_t = \alpha_{\mathcal{E}} = 1$ we have the correlation function for $\gamma_x = 0.001$ and $\gamma_{\mathcal{E}} = 0.1$ the same as for $\gamma_x = 0.1$ and $\gamma_{\mathcal{E}} = 0.001$.

Marginal temporal correlation

The temporal covariance at a point, or the marginal temporal covariance, can be computed from (5.22) considering $s = s^{'}$, which gives $P_k(\mathbf{s}^T\mathbf{s}) = 1$. Dividing it by the marginal variance gives the temporal correlation at a spatial point, or the marginal temporal correlation. The marginal temporal correlation is

$$\left(\sum_{k=0}^{\infty} \frac{(2k+1)\mathcal{M}_{\alpha_t-\frac{1}{2}}(|t-t'|, \frac{1+\gamma_x\lambda_k}{\rho})}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(1+\gamma_x\lambda_k)^{2\alpha_t-1}}\right)\left(\sum_{k=0}^{\infty} \frac{(2k+1)}{(1+\gamma_{\mathcal{E}}\lambda_k)^{\alpha_{\mathcal{E}}}(1+\gamma_x\lambda_k)^{2\alpha_t-1}}\right)^{-1} \tag{5.28}$$

and we can see that only τ does not affect it. In Figure 5.13 the marginal temporal correlation was shown as a function of the time lag for some combination of the other parameters.

In Figure 5.13 we have four plots, each one for different combination values of α_t and $\alpha_{\mathcal{E}}$. We can see that these smoothness parameters do have effect on the marginal temporal correlation. Increasing these parameters does increases the marginal temporal range. The lines were drawn considering different combinations between γ_x, $\gamma_{\mathcal{E}}$ and ρ. For $\alpha_t = 2$ the effect from γ_x and $\gamma_{\mathcal{E}}$ is small. When $\alpha_t = 1$, a smaller values for γ_x or $\gamma_{\mathcal{E}}$ increases the temporal range. This effect from $\gamma_{\mathcal{E}}$ is bigger than from γ_x.

The horizontal dashed gray lines is drawn at 0.139 in order to relate to the Matérn practical range. This is the correlation at the distance equals the square root of the smoothness times the range parameter. Thus, increasing the smoothness the practical range and we are also seeing this effect for the marginal temporal correlation.

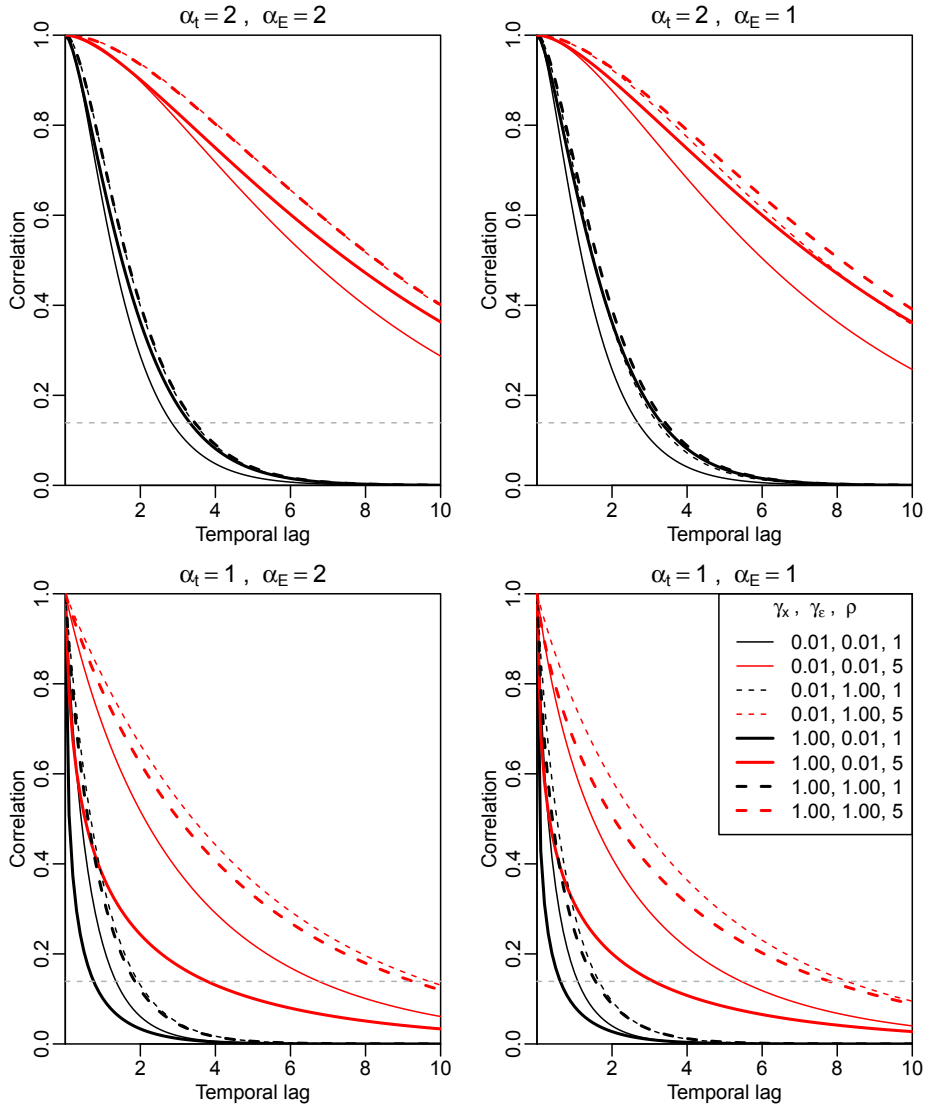


Figure 5.13: Marginal temporal correlation as a function of the temporal lag for some values of the model parameters.

5.3 Iterated damped heat equation model in $\mathbb{R}^d$

In this section we consider the iterated damped heat equation model considering the space domain as the $\mathbb{R}^d$ and looking closely at the case when $d = 2$. The spectral density is

$$s_u(\boldsymbol{\lambda}, \omega) = \frac{1}{(2\pi)^{d+1}\tau[\rho^2\omega^2 + (1 + \gamma_x \parallel \boldsymbol{\lambda} \parallel^2)^2]^{\alpha_t}(1 + \gamma_{\mathcal{E}} \parallel \boldsymbol{\lambda} \parallel^2)^{\alpha_{\mathcal{E}}}} \quad (5.29)$$

for $\boldsymbol{\lambda} \in \mathbb{R}^d$.

5.3.1 Marginal spatial covariance

Integrating out ω from Eq. 5.29 we have the spatial marginal spectral density

$$\begin{aligned} s_u(\boldsymbol{\lambda}) &= 2\int_0^\infty s_z(\boldsymbol{\lambda}, \omega)\partial\omega \\ &= \frac{\Gamma(\alpha_t - 1/2)}{2^{d+1}\pi^{d+1/2}\tau\rho(1 + \gamma_x \parallel \boldsymbol{\lambda} \parallel^2)^{2\alpha_t - 1}(1 + \gamma_{\mathcal{E}} \parallel \boldsymbol{\lambda} \parallel^2)^{\alpha_{\mathcal{E}}}} \quad (5.30) \end{aligned}$$

and considering $\gamma_x = \gamma_{\mathcal{E}}$ gives a marginal spatial Matérn spectral density. For this case we have that at each time point the spatial covariance is Matérn with smoothness equal to $\alpha_{\mathcal{E}} + 2\alpha_t$, range parameter equals to γ_x. Additionally, considering $\gamma_x = \gamma_{\mathcal{E}}$, from the Matérn covariance function we have the marginal variance equals to

$$\frac{\Gamma(\alpha_t - 1/2)\Gamma(\alpha_{\mathcal{E}} + 2\alpha_t - 1 - d/2)\gamma_x^{-2\alpha_{\mathcal{E}} - 4\alpha_t + 2 + d}}{2^{1+d/2}\pi^{\frac{d+1}{2}}\Gamma(\alpha_{\mathcal{E}} + 2\alpha_t - 1)\Gamma(\alpha_t)\tau\rho}.$$

Similarly to the previous model, this knowledge is important for working with this model.

5.3.2 Marginal temporal covariance

Since the model is isotropic over space, Eq (5.29) can be written as function of $\lambda = \parallel \boldsymbol{\lambda} \parallel$. However, to compute analytical expression for the marginal temporal covariance involves to integrate an expression that is not easy to work with.

Considering the case when $\alpha_t = 1$ and $\alpha_{\mathcal{E}} = 2$, the marginal temporal spectral density is obtained as follows

$$s_u(\omega) = 2\pi\int_0^\infty |\lambda| s_u(\lambda, \omega)\partial\lambda \quad = \quad \frac{2\arctan(\frac{1}{\rho\omega}) - \pi + 2\rho\omega}{4\rho^3\gamma_{\mathcal{E}}^{-1}\omega^3(2\pi)^{(d+1)}} \quad (5.31)$$

We have to analyze the behavior when $\omega \to 0$ in order to conclude about intrinsicness. Using the L'Hôpital's rule as

$$\frac{\partial[2\arctan(\frac{1}{\rho\omega}) - \pi - 2\rho\omega]/\partial\omega}{\partial[4\rho^3\gamma_{\mathcal{E}}^{-2}\omega^3(2\pi)^{(d+1)}]/\partial\omega} = \frac{1}{4(2\pi)^{d+1}\gamma_{\mathcal{E}}^{-2}(\rho\omega^2+1)},$$

which gives that $s_u(\omega) > 0$ for all ω, and particularly that $\lim_{\omega\to 0} s_u(\omega)$ is a constant. Therefore, at each spatial point the temporal process is not intrinsic.

We have already considered the marginal spectral density to understand the marginal properties. For the case of $s_u(\omega)$, since the analytical expression is not straightforward to make conclusions, we can compare its standardized version with the standardized version of the spectral density of some common models. Thus, we show the behavior of $s_u(\omega)$ is shown in Figure 5.14. We also include the standardized spectral density for the Exponential and Gaussian covariance functions. We have that the marginal temporal covariance of the studied model decays more slowly and with a heavier tail than the Exponential, Matérn with $\nu = 0.5$, and faster than the Gaussian at the origin and has heavier tail as well.

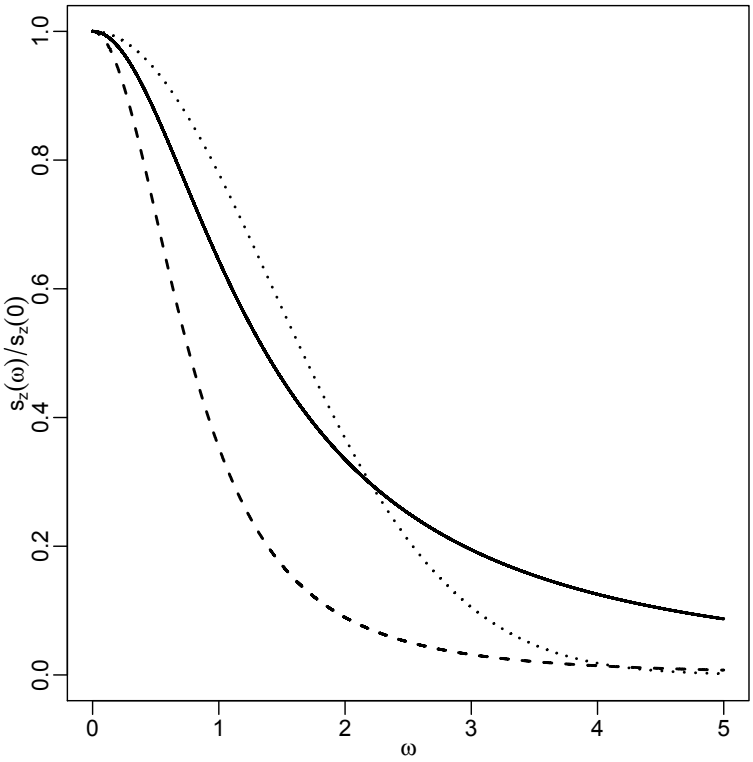


Figure 5.14: Scaled spectral density (continuous line), Matérn with smoothness equals one spectral density (dashed line) and the Gaussian (dotted line).

5.4 Discretization and precision matrix

The discretization approach consider the results in Lindgren et al. (2011). It can be considered the approach to write it as a system of differential equations, use the generalized eigenvalue problem to find the solution as in Lindgren et al. (2016) as we have considered in the previous two chapters. We can instead do find it heuristically by matching the precision matrix to the power spectra.

Let us write the model as follows

$$\begin{aligned} \tau(\rho\tfrac{\partial}{\partial t} + a - \gamma_x\Delta)^{\alpha_t} u(s,t) &= \mathcal{E}(s,t) \\ (1-\gamma_{\mathcal{E}}\Delta)^{\alpha_{\mathcal{E}}/2}\mathcal{E}(s,\delta t) &= \mathcal{W}_{\mathcal{E}}(s,\delta t) \end{aligned}$$

where a can be either zero, for the intrinsically stationary, or one, for the damped version. This gives the following spectral density

$$s_u(\boldsymbol{\lambda},\omega)^{-1} \propto \left(\rho^2\omega^2 + (a+\gamma_x \parallel \boldsymbol{\lambda} \parallel^2)^2\right)^{\alpha_t} (1+\gamma_{\mathcal{E}} \parallel \boldsymbol{\lambda} \parallel)^{\alpha_{\mathcal{E}}}$$

For the case when $\alpha_t = 1$ we have

$$s_u(\boldsymbol{\lambda},\omega)^{-1} \propto \rho^2\omega^2(1+\gamma_{\mathcal{E}} \parallel \boldsymbol{\lambda} \parallel)^{\alpha_{\mathcal{E}}} + (a+\gamma_x \parallel \boldsymbol{\lambda} \parallel^2)^2(1+\gamma_{\mathcal{E}} \parallel \boldsymbol{\lambda} \parallel)^{\alpha_{\mathcal{E}}}$$

5.4.1 The case with no damping

With no damping ($a = 0$) the spectral density for $\alpha_t = 1$ has the form

$$\rho^2\omega^2(1+\gamma_{\mathcal{E}} \parallel \boldsymbol{\lambda} \parallel)^{\alpha_{\mathcal{E}}} + \gamma_x^2 \parallel \boldsymbol{\lambda} \parallel^4 (1+\gamma_{\mathcal{E}} \parallel \boldsymbol{\lambda} \parallel)^{\alpha_{\mathcal{E}}} .$$

The first term translates into $\rho^2 \boldsymbol{H} \otimes \boldsymbol{R}_{\alpha_{\mathcal{E}}}(\gamma_{\mathcal{E}})$ where $\boldsymbol{R}_{\alpha_{\mathcal{E}}}(\gamma_{\mathcal{E}})$ is the precision matrix associated to the driving noise precision defined as

$$\boldsymbol{R}_1(\gamma) = (\boldsymbol{C} + \gamma\boldsymbol{G}) \text{ and } \boldsymbol{R}_2(\gamma) = (\boldsymbol{C} + 2\gamma\boldsymbol{G} + \gamma^2\boldsymbol{G}^{(2)}).$$

The second term translates into $\gamma_x \boldsymbol{D} \otimes \boldsymbol{G}^{(2)} \boldsymbol{R}_{\alpha_{\mathcal{E}}}$. Thus,

$$Q_{1,\alpha_{\mathcal{E}}} = \tau(\rho^2 \boldsymbol{H} \otimes \boldsymbol{R}_{\alpha_{\mathcal{E}}}(\gamma_{\mathcal{E}}) + \gamma_x \boldsymbol{D} \otimes \boldsymbol{G}_s^{(2)} \boldsymbol{R}_{\alpha_{\mathcal{E}}}(\gamma_{\mathcal{E}}))$$

The $\alpha_{\mathcal{E}} = 1$ case is possible with the current `spde3`[a] specification:

$$\begin{aligned} \boldsymbol{Q}_{1,1} &= \tau[\rho^2 \boldsymbol{H} \otimes \boldsymbol{C}(\boldsymbol{I}+\gamma_{\mathcal{E}}\boldsymbol{C}^{-1}\boldsymbol{G}) + \boldsymbol{D} \otimes \boldsymbol{G}^{(2)}(\boldsymbol{I}+\gamma_{\mathcal{E}}\boldsymbol{C}^{-1}\boldsymbol{G})] \\ &= \tau[\rho^2 \boldsymbol{H} \otimes \boldsymbol{C} + \rho^2\gamma_{\mathcal{E}}\boldsymbol{H} \otimes \boldsymbol{G} + \boldsymbol{D} \otimes \boldsymbol{G}^{(2)} + \gamma_{\mathcal{E}}\boldsymbol{D} \otimes \boldsymbol{G}^{(3)}] \\ &= \tau\gamma_{\mathcal{E}}[\rho^2 \boldsymbol{H} \otimes \boldsymbol{G} + \rho^2/\gamma_{\mathcal{E}}\boldsymbol{H} \otimes \boldsymbol{C} + \boldsymbol{D} \otimes \boldsymbol{G}^{(3)} + \\ &\quad (\boldsymbol{G} \otimes \boldsymbol{D}^{1/2})^T \frac{1}{\gamma_{\mathcal{E}}} (\boldsymbol{D}^{1/2} \otimes \boldsymbol{G})], \text{ the actual way considered.} \end{aligned}$$

[a] Specification using `INLA:::inla.spde3.generic()` in the **INLA** package

5.4.2 The damped heat equation case

With damping ($a = 1$) the spectral density for $\alpha_t = 1$ has the form

$$\rho^2\omega^2(1+\gamma_{\mathcal{E}} \parallel \boldsymbol{\lambda} \parallel)^{\alpha_{\mathcal{E}}} + (1+\gamma_x^2 \parallel \boldsymbol{\lambda} \parallel^2)^2(1+\gamma_{\mathcal{E}} \parallel \boldsymbol{\lambda} \parallel)^{\alpha_{\mathcal{E}}}) \, .$$

The first term translates as before while the second term translates into $\boldsymbol{D} \otimes \boldsymbol{R}_2(\gamma_x)\boldsymbol{R}_{\alpha_{\mathcal{E}}}(\gamma_{\mathcal{E}})$. Thus we have that

$$Q_{1,\alpha_{\mathcal{E}}} = \tau(\rho^2 \boldsymbol{H} \otimes \boldsymbol{R}_{\alpha_{\mathcal{E}}}(\gamma_{\mathcal{E}}) + \boldsymbol{D} \otimes \boldsymbol{R}_2(\gamma_x)\boldsymbol{R}_{\alpha_{\mathcal{E}}}(\gamma_{\mathcal{E}}))$$

5.5 Application to daily temperature data

We consider the daily temperature data available at `ftp.ncdc.noaa.gov/pub/data/ghcn/daily`. We considered the period from January, 1 2017 to June, 30 2017. In this period there is data at 12862 stations. The locations for these stations is shown in Figure 5.15. We can see that the locations are spread in a very irregular way. Even thou it only considers on land stations, there are countries with small number of stations and those with several, the USA for example.

The main computational task in this model is the evaluation of the log determinant of the resulting precision matrix. The state of the art in doing so for one-core machines is implemented in the **INLA** package, `r-inla.org`. However, there is a limitation in the memory use, since it is considered the Cholesky decomposition. Thus, this application is just a proof of concept. This is because that in order to do a proper analysis we still need further developments in numerical solution for general large sparse matrix decompositions.

We can check how much can be done in a single core with over a hundred of Gigabyte RAM machine. When doing the Cholesky factorization, is is considered the reordering of the precision matrix that gives the lowest number of non-zeros elements in the Cholesky factor, nnz. We considered different mesh resolutions to analyse different number of days of data. One case is the first 10 days of data with 10 time knots, so no reduction in the time resolution. A second case is the 183 days of data with time knots at each 10 days, $k = 10$. A third case considered the 31 days in January with one knot at each day. Thus we are considering the increasing number of data points when having the 10 and 31 days case at the same time that considering more data and lower temporal resolution, with 180 days and $k = 10$. The number of non-zeros in the Cholesky factor versus nnz for these three cases is shown in the top plot in Figure 5.16.

Figure 5.15: Locations of the stations considered.

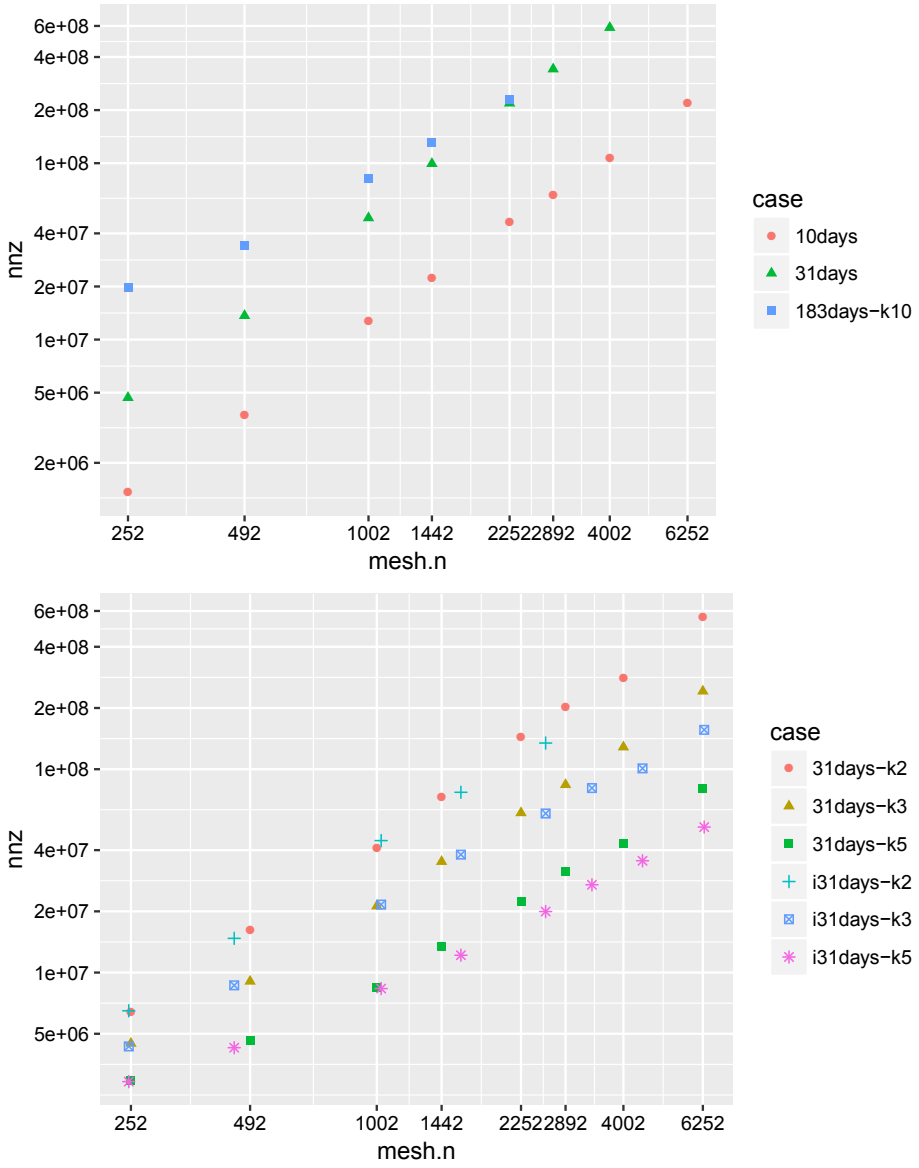


Figure 5.16: The number of non-zeros in the Cholesky factor versus the number of noes in the mesh for some cases.

In the three cases shown in the top plot of Figure 5.16 we can see that increasing the number of mesh nodes increases the number of non-zeros delivers a larger nnz for all the cases. Also, that more data also increases nnz. However, we can see that is worth of reducing the time resolution when having more data to reduce nnz. The case with 183 days and $k = 10$ actually gives 18 time knots which is less than 31 days and therefore the nnz increases slowly for increasing the number of nodes in the mesh.

In the bottom plot in Figure 5.16 we have two set of cases for working with 31 days of data. We consider regular meshes, with triangles with same size everywhere, and irregular meshes, with triangle sizes smaller where there stations. For these two set of cases we have $k = 2$, $k = 3$ and $k = 5$. We can see that using irregular meshes is better to reduce the nnz.

The mean temperature for the day t at location $\boldsymbol{s}$, $y(\boldsymbol{s}, t)$ were modelled considering the following model

$$y(\boldsymbol{s}, t) \sim N(\beta \text{Altitude} + f(\text{latitude}, \text{month}) + \boldsymbol{A}u(\boldsymbol{s}_0, t_0), \tau_y^{-1})$$

where we have a fixed effect of altitude, a smoothed effect of latitude for each month, and the space-time effect. For the later, we have $\boldsymbol{A}$ being the projector matrix from the spatial mesh and temporal knots to the data locations, $u(\boldsymbol{s}_0, t_0)$ is the process at the mesh nodes and time knots and τ_u is the precision for independent error which is also called as the nugget effect. The model assumed for $u(\boldsymbol{s}, t)$ is the model based in the intrinsically stationary heat equation, having $\alpha_t = 1$ and $\alpha_{\mathcal{E}} = 1$.

The results we are going to show consider 183 days of data with $k = 10$ and a very crude mesh with only 362 nodes. We also have fitted a separable model, the one in Cameletti et al. (2012), just to have something to compare with. Three goodness-of-fit measures were considered, the popular deviance information criteria - DIC, the recently proposed widely applicable information criteria - WAIC and logarithm of the cross-validated predictive ordinates, CPO. These are shown in Table 5.1. We can see that these were similar among the two different models.

space-time model	DIC	WAIC	CPO
separable	10687588.44	10686084.08	5343048.83
non-separable	10687770.09	10686244.53	5343126.70

Table 5.1: Goodness-of-fit measures.

The summary for the posterior marginal distribution of τ_y and the log of the three parameters of the space-time model is shown in Table 5.2.

Parameter	mean	sd	2.5% Q.	Median	97.5% Q.	mode
τ_y	0.07	0.00	0.07	0.07	0.07	0.07
$\log(\tau)$	-12.63	0.07	-12.79	-12.62	-12.53	-12.58
$\log(\rho)$	7.18	0.05	7.11	7.18	7.30	7.15
$\log(\gamma)$	-5.68	0.07	-5.84	-5.68	-5.55	-5.66

Table 5.2: Summary of the posterior distribution for the model parameters.

With the fitted parameters we can drawn the marginal correlation functions. In Figure 5.17 we have both the marginal spatial correlation and the marginal temporal correlation. We considered the posterior mean and to drawn the continuous lines and the quantiles to drawn the dashed lines. We can see a very different behavior for the two kind of space-time models considered.

The standard deviation of $u(\boldsymbol{s}, t)$ at the mesh nodes were computed. This statistic was projected for the globe and shown in Figure 5.18 for two time knots. We can see that the uncertainty is inversely proportional to the density of stations.

In the Figure 5.19 we can see the mean of the posterior marginal distribution for $u(\boldsymbol{s}, t)$ for the first 10 time knots, and in Figure 5.20 for the last eight ones. As the spatial resolution of the model is small, we mainly have the effect of latitude and season as it is bigger near the equatorial line and becomes warmer in the north as time goes from January to June and colder in the south.

5.5.1 Appendix: the code to fit the model

We insert bellow the code considered to fit the model.

```
### function to define the model
ciheat11 <- function(
  cmd=c("graph", "Q", "mu", "initial", "log.norm.const",
        "log.prior", "quit"), theta = NULL, args = NULL) {
    interpret.theta <- function(n, theta)
        return(list(gs=exp(theta[1]), gt=exp(theta[2]),
                    ge=exp(theta[3])))
    graph <- function(n, theta) return(iheat11.objects$graph)
    Q <- function(n, theta) {
        gamma <- exp(theta)
        q <- gamma[1]*gamma[3]*(
```

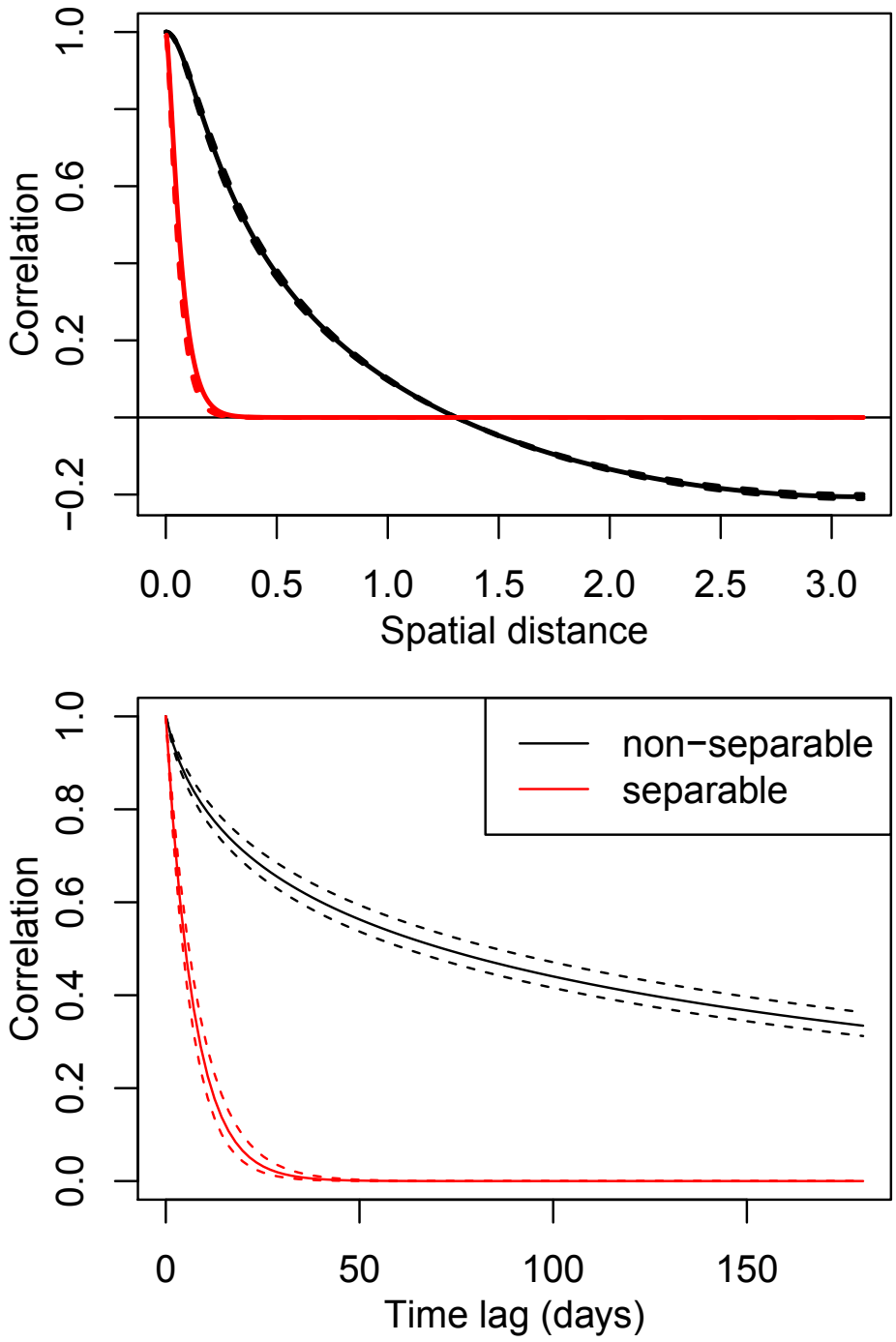


Figure 5.17: Marginal spatial correlation (left) and the marginal spatial correlation (right).

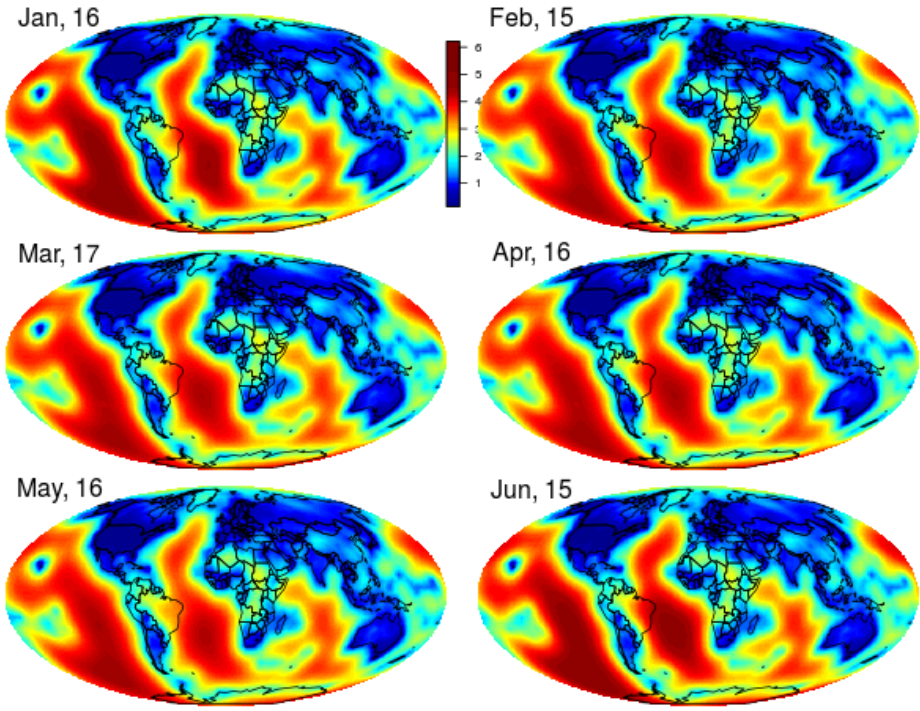


Figure 5.18: Standard deviation of the posterior mean for $u(\boldsymbol{s}, t)$ at six time points.

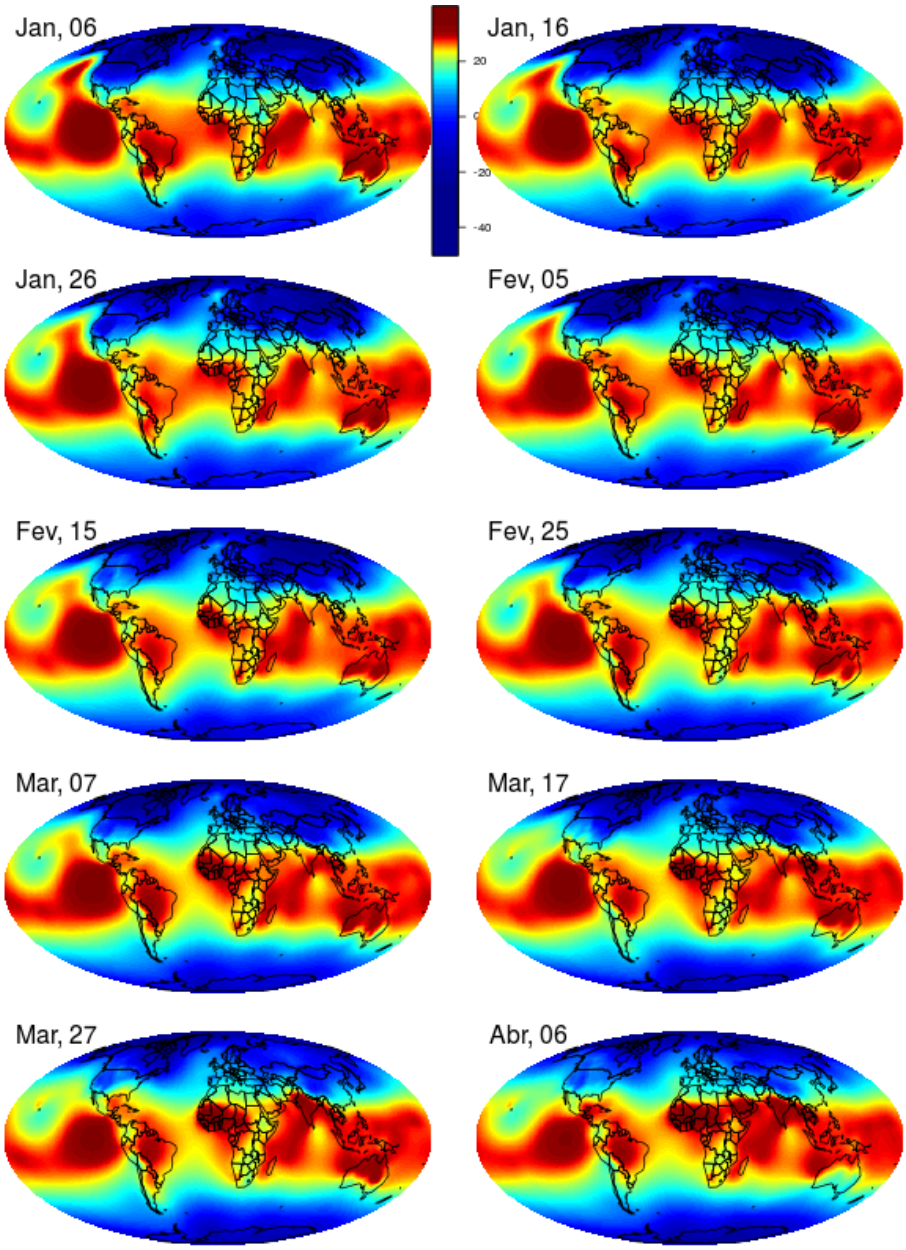


Figure 5.19: Mean of the posterior mean for $u(\boldsymbol{s}, t)$ at the first 10 time knots.

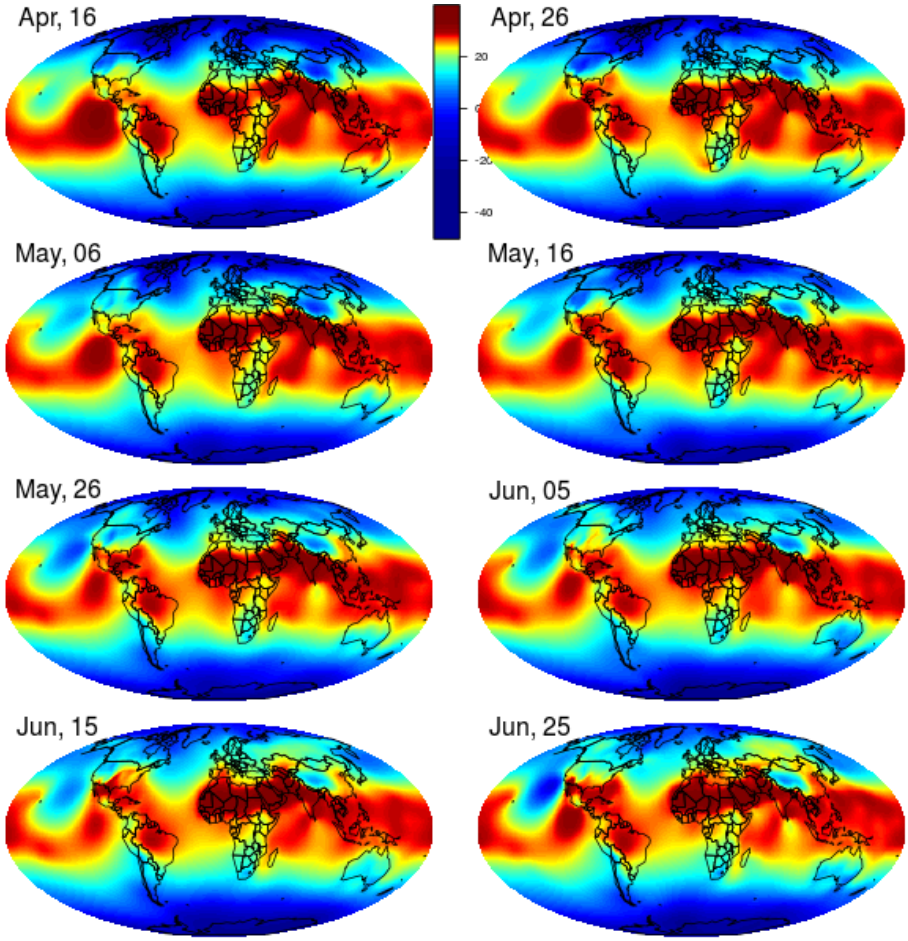


Figure 5.20: Mean of the posterior mean for $u(\boldsymbol{s}, t)$ at the last eight time knots.

```
          (gamma[2]^2)*iheat11.objects$M0 +
          (gamma[2]^2/gamma[3])*iheat11.objects$M1 +
          iheat11.objects$M2 + (1/gamma[3])*iheat11.objects$M3)
        return(q)
    }
    mu <- function(n, theta) return(numeric(0))
    log.norm.const <- function(n, theta) return(numeric(0))
    log.prior <- function(n, theta) ### N(0, 1)^3
        return(sum(dnorm(theta, 0.0, 1, log=TRUE)))
    initial <- function(n, theta) return(c(0,0,0))
    quit <- function(n, theta) return(invisible())
    cmd <- match.arg(cmd)
    val <- do.call(cmd, args=list(n=as.integer(args$n),
                   theta=theta))
    return(val)
}

### load the dataset
load('data/tmed17.RData')

### define the time knots and one-dimensional mesh over it
(days <- sort(unique(tmed$day)))
(ndays <- length(days)); k <- 10
(gt <- seq(trunc(k/2)+1, ndays, k));(nk <- length(gt))
tmesh <- inla.mesh.1d(gt, degree=1)

### convert the coordinates: Mollweide projection
coo.mollweide <- inla.mesh.map(cbind(tmed$long, tmed$lat),
                               'longlat', inverse=TRUE)

### build the mesh
mesh <- inla.mesh.create(globe=10)

### build the space-time projector matrix
Ast <- inla.spde.make.A(mesh, coo.mollweide,
                        group=tmed$day, group.mesh=tmesh)

#### non-separable model objects definition
fe <- inla.mesh.fem(mesh, order=3)
tfe <- inla.mesh.fem(tmesh)
```

```
nm <- mesh$n*tmesh$n

iheat11.objects <- list(
    M0=kronecker(tfe$g1, fe$g1),
    M1=kronecker(tfe$g1, fe$c0),
    M2=kronecker(tfe$c0, fe$g3),
    M3=kronecker(tfe$c0, fe$g2))
iheat11.objects$graph <-
   (Reduce('+', lapply(iheat11.objects, abs))>0)+0

riheat11 <- inla.rgeneric.define(
   ciheat11, n=as.integer(mesh$n*tmesh$n),
   iheat11.objects=iheat11.objects)

## formula
fns <- y ~ 0 +
    f(st, model=riheat11, values=1:nm)

### data stack
datst <- inla.stack(data=list(y=tmed$tmed),
                    A=list(Ast),
                    effects=list(st=1:ncol(Ast)),
                    tag='tmedst')

nsres <- inla(fns, data=inla.stack.data(datst),
              control.predictor=list(A=inla.stack.A(datst)),
              num.threads=1)
```

Bibliography

Abrahamsen, P. (1997). A review of Gaussian random fields and correlation functions. Norwegian Compting Center report No. 917.

Apolinário, D., Ribeiro, A. I., Krainski, E. T., Sousa, P., Abranches, M., and Duarte, R. (2017). Tuberculosis inequalities and socio-economic deprivation in portugal. *International Journal of Tuberculosis and Lung Disease*, 21(7):784–789.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *JRSS-B*, 36(2):192–236.

Besag, J. and Mondal, D. (2005). First-order intrinsic autoregressions and the de wijs process. *Biometrika*, 92(4):909–920.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.

Blangiardo, M. and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. Wiley.

Bolin, D. (2014). Spatial Matérn fields driven by non-Gaussian noise. *Scandinavian Journal of Statistics*, 41(3):557–579.

Bolin, D. and Kirchner, K. (2017). The SPDE approach for Gaussian random fields with general smoothness. *submitted*.

Bolin, D., Kirchner, K., and M., K. (2017a). Numerical solution of fractional elliptic stochastic PDEs with spatial white noise. *subimitted*.

Bolin, D., Kirchner, K., and M., K. (2017b). Weak convergence of Galerkin approximations for fractional elliptic stochastic PDEs with spatial white noise. *subimitted*.

Bolin, D. and Lindgren, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to globla ozone mapping. *The Annals of Applied Statistics*, 5(1):523–550.

Bolin, D. and Lindgren, F. (2014). Lindgren, excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):85–106.

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2012). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *Advances in Statistical Analysis*, 97(2):109–131.

Cosandey-Godin, A., Krainski, E. T., Worm, B., and Flemming, J. M. (2014). Applying bayesian spatiotemporal models to fisheries bycatch in the canadian arctic. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(2):186–197.

Cressie, N. and Huang, H. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *JASA*, 94(448):1330–1340.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Society*.

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2013). *Analysis of Longitudinal Data*. Oxford University Press, 2nd edition.

Eymard, R., Thierry, G., and Raphaèle, H. (2000). *Handbook of numerical analysis*, volume 7, chapter Finite Volume Methods, pages 713–1018. Zoe Kruze.

Finkenstadt, B., Held, L., and Isham, V., editors (2006). *Statistical Methods for Spatio-Temporal Systems*. CRC Press.

Fonseca, T. C. O. and Steel, M. F. J. (2011). A general class of nonseparable spaceâĂŞtime covariance models. *Environmetrics*, 22(2):224–242.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Jornal of the American Statistical Association*, 102(447):321–331.

Fuglstad, G. (2010). Approximating solutions of stochastic differential equations with Gaussian Markov random fields. Technical report, NTNU. Specialization Project.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *JASA*, 97(458):590–600.

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*. Elsevier, 7th edition.

Heine, V. (1955). Models for two-dimensional stationary stochastic processes. *Biometrika*, 42(1):170–178.

Higdon, D. (2002). *Quantitative methods for current environmental issues*, chapter Space and space-time modeling using process convolutions, pages 37–56. Springer.

Jones, R. H. (1993). *Longitudinal data with serial correlation: A state-space approach*. Springer, 1 edition.

Jones, R. H. and Zhang, Y. (1997). *Modeling longitudinal and spatially correlated data*, chapter Models for continuous stationary space-time processes, pages 289–298. Springer.

Kelbert, M. Y., Leonenko, N. N., and Ruiz-Medina, M. D. (2005). Fractional random fields associated with stochastic fractional heat equations. *Adv. Apply. Prob.*, 37(1):108–133.

Kinyoki, D. K., Kandala, N., Manda, S. O., Krainski, E. T., Fuglstad, G., Moloney, G. M., Berkley, J. A., and Noor, A. M. (2016). Assessing comorbidity and correlates of wasting and stunting among children in Somalia using cross-sectional household surveys: 2007 to 2010. *BMJ Open*, 85(1):164–176.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19:2555–2567.

Lindgren, F., Guttorp, P., and Rue, H. (2016). A stochastic model for global temperature and climate using weather data. *in preparation*.

Lindgren, F. and Rue, H. (2015). Bayesian spatial and spatio-temporal modelling with R-INLA. *Journal of Statistical Software*.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *JRSS-B*, 73(4):423–498.

Lindgren, G. (2013). *Stationary Stochastic Srocess: Theory and Applications*. CRC Press.

Macdonald, J. I., Logemann, K., Krainski, E. T., Sigurðsson, Þ., Beale, C. M., Huse, G., Hjøllo, S. S., and Marteinsdóttir, G. (2017). Can collective memories shape fish distributions? a test, linking space-time occurrence models and population demographics. *Ecography*.

Martínez-Beneito, M. A., López-Quilez, A., and Botella-Rocamora, P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine*, 27(10):2874–2889.

Matérn, B. (1960). *Spatial variation: Stochastic models and their application o some problems in forest surveys and other sampling investigations*. PhD thesis, Faculty of Mathematics and Natural Sciences of the University of Stockholm.

Montero, J., Fernandéz-Avilés, G., and Mateu, J. (2015). *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*. Wiley.

Porcu, E., Alegria, A., and Furrer, R. (2017). Modeling temporally evolving and spatially globally dependent data. *submitted*.

Quarteroni, A. (2009). *Numerical Models for Differential Problems*. Springer-Verlag Italia.

Rasmussen, C. E. and Williams, K. I. (2006). *Gaussian Processes for Machine Learning*. the MIT Press.

Ribeiro, A. I. C., Krainski, E. T., Carvalho, M. S., and de Pina, M. d. (2016). Where do people live longer and shorter lives? an ecological study of old-age survival across 4404 small areas from 18 european countries. *Journal of Epidemiology Community Health*, 70(6):561–568.

Riebler, A., Søorbye, S. H., Simpson, D., and Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165.

Rodrigues, A. and Diggle, P. (2010). A class of convolution-based models for spatio-temporal processes with non-separable covariance structure. *Scandinavian Journal of Statistics*, 37(4):553–567.

Rozanov, J. A. (1977). Markov random fields and stochastic partial differential equations. *Math. USSR Sbornik*, 32(4)):515–534.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. Monographs on Statistics & Applied Probability. Boca Raton: Chapman and Hall.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.

Rue, H., Riebler, A. I., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with inla: A review. *Annual Review of Statistics and Its Application*, 4:395–421.

Rue, H. and Tjelmeland, H. (2001). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, 29(1):31–49.

Ruiz-Cárdenas, R., Krainski, E. T., and Rue, H. (2012). Direct fitting of dynamic models using integrated nested Laplace approximations - INLA. *Computational Statistics & Data Analysis*, 56(6):1808 – 1828.

Rushworth, A., Lee, D., and Mitchell, R. (2014). A spatio-temporal model for estimating the long- term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and Spatio-temporal Epidemiology*, 1(10):29–38.

Sigrist, F., R., K. H., and Stahel, W. A. (2015). Stochastic partial differential equation based modelling of large spacetime data sets. *Journal of the Royal Statistical Society, Series B*, 77(1):3–33.

Simpson, D. P., Illian, J. B., Lindren, F., Sørbye, S. H., and Rue, H. (2015). Going off grid: Computationally efficient inference for log-gaussian cox processes. *to be published*.

Stam, M., Charlotte, G., van Breda, W., and Krainski, E. (2017). Assessing the spatiotemporal relation between Twitter data and violent crime. In *Preproceedings of the 29th Benelux Conference on Artificial Inteligence - BNAIC 2017*.

Stein, M. L. (2005). Space-time covariance functions. *JASA*, 100(469):310–321.

Stein, M. L. (2013). On a class of space-time intrinsic random functions. *Bernoulli*, 19(2):287–408.

Vivar, J. C. and Ferreira, M. A. R. (2009). Spatio-temporal models for Gaussian areal data. *Journal of Computational and Graphical Statistics*, 18(3):658–674.

Wahba, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16.

Wallin, J. and Bolin, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scandinavian Journal of Statistics*, 42(3):872–890.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(3/4):434–449.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Press.

Part II

Non-centered space-time models

A computationally efficient approach to account for intrinsic space-time variation in disease mapping

Elias T. Krainski, Andrea Riebler and Håvard Rue

January 11, 2018

Abstract

A main interest in space-time modeling is to define models which account for temporal changes in the spatial pattern, so called space-time variation. In the disease mapping literature there are either models considering an overall spatial variation, an overall temporal variation plus a space-time variation or models that only consider the space-time variation excluding the main effects. Intrinsically stationary models are often considered in the first case and proper autoregressive models in the latter case. For the ease of interpretation it is beneficial to use both main and interaction effects in the model. However, if the number of areas and/or number of considered time points are large this gets computationally challenging due to the required number of linear constraints needed for the interaction term to ensure identifiability of the main effects. We propose an alternative formulation avoiding linear constraints entirely and trading them for computationally cheaper linear combinations in order to retain the usual effect estimates. We illustrate the performance of this new approach based on simulated data and compare it to the original approach. The results show that out approach reduces dramatically the computational cost while leaving the relative risk estimates unchanged. Furthermore, we model space-time variation of infant mortality in the Paraná state in Brazil from 1996 to 2015. The **R** code using INLA to fit all models is provided.

Disease Mapping, INLA, Space-time Interaction

1 Introduction

In disease mapping the analyzed outcome is usually the number cases or deaths due to some disease observed for several non-overlapping but neighbouring areas. These counts are frequently observed for several periods over time. The expected number of cases can be computed considering the population structure at each area and time. The ratio between the observed and expected number of cases is termed Standardized Mortality Ratio (SMR) or Standardized Incidence Ration (SIR). In the simplest case the log of SMR or SIR is modelled linearly but more often generalised linear models are considered. A Bayesian framework is attractive to borrow strength between neighbouring regions (Lawson, 2013).

There are several models proposed in the literature for the space-time variation in the relative risk. The main approach is to combine models for variation over space with models for variation over time, the so called space-time interaction of the main spatial and temporal effects (Clayton, 1996). When both main effects are fixed effects, the interaction between them will also be a fixed effect. When either of them is random the interaction is

random as well. When both are random, the resulting space-time precision/covariance matrix is proportional to the Kronecker product between the corresponding precision/covariance matrices.

The interaction between a fixed time effect and a random spatial effect was considered in Bernardinelli et al. (1995), resulting in a regression on the time index with coefficients varying for each area. Assunção et al. (2001) added an additional quadratic term for time also weighted by area-specific weights. The use of B-splines with spatially structured coefficients was proposed in MacNab and Dean (2001) and MacNab and Dean (2002). These temporal trends are easier to interpret and have the flexibility to vary, smoothly, over space. However, a random space-time interaction effect is actually more flexible because it is not required to follow any specific trend imposed by the fixed term. The first approach in this direction was proposed by Waller et al. (1997), where a spatially structured effect at each time period was considered.

Knorr-Held (2000) proposed a general random effects modeling approach in discrete time and discrete space resulting in four different interaction types. The idea is to combine one of the two temporal effects, either structured or unstructured, with one of the two spatial effects. The type I interaction is obtained when both unstructured main effects are combined, the type 2 interaction when the structured temporal component is assumed to interact with the unstructured spatial component, the type 3 when the structured spatial component is assumed to interact with the unstructured temporal component and finally the type 4 is obtained when both structured effects are combined. Note that the interactions type 2, 3 and 4 are intrinsic, i.e. their precision matrices do not have full rank. The resulting precision matrix has rank deficiency equal to the number of areas for type 2, number of time points for type 3, and number of time points plus number of areas minus one, when using a first order random walk over time, or minus two, when using a second order random walk for time, for type 4. Sum-to-zero constraints can be applied to make the model proper (Rue and Held, 2005). The use of such hard but also *soft-constraints*, where linear constraints are specified with some uncertainty, requires solving a system with size equal to the number of constraints (Rue and Held, 2010). This computation is not only required once but several times during the fitting process. When considering a Markov chain Monte Carlo framework, adjustment of sum-to-zero constraints is needed for every sample, and using the INLA approach (Rue et al., 2009) for every posterior function evaluation. Thus, for models with several sum-to-zero constraints it becomes prohibitive.

One option to avoid sum-to-zero constraints but keep the model flexibility is to consider a proper distribution for the random space-time interaction. This approach was considered in Martínez-Beneito et al. (2008), Vivar and Ferreira (2009), Ruiz-Cárdenas et al. (2012), Rushworth et al. (2014) and Blangiardo and M. (2015), Ch. 8. The resulting models are easy to fit in current available software and there the model estimation can be done in a computationally efficient way. However, the interaction and the main effects are confounded leading to difficulties in parameter interpretation.

In this paper, we propose an alternative approach which leads to identifiable and interpretable parameter estimates, but still allows for a computational efficient model estimation. From our simulation studies we found that we can still choose which of the four space-time intrinsic interactions

are suitable for the dataset being analyzed in a much shorter time than when considering the original sum-to-zero constraints. The idea is to account for the null space of the intrinsic prior distributions of the random effect only after the model estimation. This is achieved by either considering a contrast parameterization or by making the prior proper through the addition of a small diagonal term in the prior precision matrix. The null space is adjusted for using linear combinations involving the posterior marginals of the random effects.

The paper is structured as follows. In Section 2 we review the Knorr-Held's type 4 interaction model and the corresponding required identifiability constraints. In Section 3 we propose to avoid the constraints and use instead linear combinations to account for the null space of the space-time type 2, 3 and 4 intrinsic interaction models. In Section 4 we use scaled intrinsic models to have interpretable and comparable parameterization across different model components. A simulation study is presented in Section 5. We applied these models to analyze the infant mortality rate in Paraná state in Brazil, showing the results in Section 6. We end discussing our findings in Section 7.

2 On the space-time intrinsic interactions

In this section we review an approach to model data collected at a set of n areas over m time points, y_{it}, $i = 1, ..., n$ and $t = 1, ..., m$. Within a generalised linear model framework it is common to model the linear predictor η_{it}, which is linked to the expected value of $\boldsymbol{y}$ via a suitable link function $g(\cdot)$, so that $\mathrm{E}(\boldsymbol{y}) = g^{-1}(\boldsymbol{\eta})$. Assume

$$\eta_{it} = \text{other effects} + v_t + s_i + d_{it} \tag{1}$$

where v_t is a temporal main effect common for all areas, s_i is a spatial main effect common for all times and d_{it} is a space-time interaction effect.

2.1 The main effects

Knorr-Held (2000) considered a BYM type model (Besag et al., 1991) for both the main effects $\boldsymbol{v}$ and $\boldsymbol{s}$. The temporal main effect is $\boldsymbol{v} = \boldsymbol{\gamma} + \boldsymbol{\alpha}$, where $\boldsymbol{\gamma}$ is an unstructured random effect and $\boldsymbol{\alpha}$ is a structured random effect. Similarly, the main spatial effect was represented as $\boldsymbol{s} = \boldsymbol{\phi} + \boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is structured and $\boldsymbol{\phi}$ is not. All random effects are assumed to be a priori zero mean Gaussian with precision matrices $\tau_\gamma \boldsymbol{I}$ for $\boldsymbol{\gamma}$, $\tau_\alpha \boldsymbol{K}$ for $\boldsymbol{\alpha}$, $\tau_\phi \boldsymbol{I}$ for $\boldsymbol{\phi}$, and $\tau_\theta \boldsymbol{R}$ for $\boldsymbol{\theta}$. Here, $\boldsymbol{I}$ matrix denotes the identity matrix with the adequate dimension and $\boldsymbol{K}$ and $\boldsymbol{R}$ denote the temporal and spatial structure matrices as defined below.

The structure matrix for the structured temporal part, $\boldsymbol{\alpha}$, can be written as $\boldsymbol{K} = \tilde{\boldsymbol{H}} - \boldsymbol{H}$ where $\boldsymbol{H}$ is the temporal neighborhood structure matrix and $\tilde{\boldsymbol{H}}$ is the diagonal matrix with diagonal elements equal to the row sum

of $\boldsymbol{H}$. Here we assume a random walk of order 1 (RW1), which this leads to

$$\boldsymbol{K} = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ & 2 & & & \\ & & \ddots & & \\ & & & 2 & \\ & & & & 1 \end{bmatrix} - \begin{bmatrix} & 1 & & & \\ 1 & & 1 & & \\ & \ddots & & \ddots & \\ & & 1 & & 1 \\ & & & 1 & \end{bmatrix}.$$

Analogously, the structure matrix for the spatially structured effect, $\boldsymbol{\theta}$, can be formulated as $\boldsymbol{R} = \tilde{\boldsymbol{G}} - \boldsymbol{G}$, where $\boldsymbol{G}$ is the spatial neighborhood structure, defined as

$$\boldsymbol{G}_{i,j} = \begin{cases} 1 & \text{if} \quad j \sim i \\ 0 & \text{otherwise,} \end{cases}$$

where $j \sim i$ means "area j is neighbor to area i", and $\tilde{\boldsymbol{G}}$ is a diagonal matrix with elements equal to the row sum of $\boldsymbol{G}$.

2.2 The space-time interactions

For the space-time interaction term, d_{it}, four cases were proposed in Knorr-Held (2000). The type 4 is the most complex and is obtained when the precision matrix is

$$\begin{aligned} \tau_d \boldsymbol{K} \otimes \boldsymbol{R} &= \tau_d(\tilde{\boldsymbol{H}} - \boldsymbol{H}) \otimes (\tilde{\boldsymbol{G}} - \boldsymbol{G}) \\ &= \tau_d(\tilde{\boldsymbol{H}} \otimes \tilde{\boldsymbol{G}} - \tilde{\boldsymbol{H}} \otimes \boldsymbol{G} - \boldsymbol{H} \otimes \tilde{\boldsymbol{G}} + \boldsymbol{H} \otimes \boldsymbol{G}) \end{aligned} \tag{2}$$

Here, the first term is diagonal with the number of spatial neighbors times the number of temporal neighbors, the second term accounts for the spatial neighborhood weighted by the number of temporal neighbors, the third for the temporal neighborhood weighted by the number of spatial neighbors, and the last term accounts for the space-time neighborhood structure, specifying spatial neighbors at time neighbors (a kind of second order neighbors). The conditional distribution of a pixel d_{it} is given in Knorr-Held (2000), section 2.2.4, and corresponds to

$$N\left(\frac{\tilde{\boldsymbol{H}}_{tt}\sum_{j\sim i} d_{jt} + \tilde{\boldsymbol{G}}_{ii}\sum_{k\sim t} d_{ik} - \sum_{j\sim i,k\sim t} d_{jk}}{\tilde{\boldsymbol{H}}_{tt}\tilde{\boldsymbol{G}}_{ii}}, \frac{1}{\tau_d \tilde{\boldsymbol{H}}_{tt}\tilde{\boldsymbol{G}}_{ii}}\right) \tag{3}$$

where $j \sim i$ denotes the set of neighboring areas to area i and $k \sim t$ the set of neighboring time points to time t.

The interaction type 2 corresponds to the case in which the precision matrix for $\boldsymbol{d}$ is given by $\tau_d \boldsymbol{K} \otimes \boldsymbol{I}$. This means in terms of Eq. (2) that the matrix $\tilde{\boldsymbol{G}}$ in the first and third term is replaced by an identity matrix and the second and last term are not present. Thus each area has an specific structured temporal effect and the conditional mean in (3) reduces to $\sum_{k\sim t} d_{ik}/\tilde{\boldsymbol{H}}_{tt}$.

The interaction type 3 implies that precision matrix for $\boldsymbol{d}$ is $\tau_d \boldsymbol{I} \otimes \boldsymbol{R}$. In this case, the matrix $\tilde{\boldsymbol{H}}$ in the first and second term of Eq. (2) are replaced by an identity matrix and the last two terms disappear. It means that at each time, there is a time specific structured spatial effect and the conditional mean in (3) reduces to $\sum_{j\sim i} d_{jt}/\tilde{\boldsymbol{G}}_{tt}$.

2.3 The null space and constraints

We start by considering the two structured main effects and extend for the type 2, 3 and 4 space-time interactions. We just give an overview of the matter with focus in the practical application. All the details are in Chapter 3 of Rue and Held (2005).

The precision matrix of a first order intrinsic Gaussian Markov random field (IGMRF) is given by a precision parameter multiplied with a structure matrix. This matrix has rank deficiency equal to one which means that its null space has dimension one as well. It implies that the precision matrix has one eigenvalue, which is equal to zero. The associated eigenvector spans the null space of the precision structure and has entries that are all equal to the same constant. The specific value of this constant does not matter and the eigenvector can actually be considered to be a vector of ones, $\mathbf{1}$. Here, we can consider it to be $\frac{1}{n}\mathbf{1}$.

The practical implication is that the addition of a constant to all elements of a first order IGMRF does not change its density. Therefore it is invariant to the addition of a constant. One practical way to make the IGMRF identifiable is to impose a sum-to-zero constraint, or equivalently require to have zero mean. Thus, the sum-to-zero constraint is accounting for its null space.

The interaction type 2 implies that each time series in $\boldsymbol{d}$ is one dimensional a first order IGMRF, such as α. The structure matrix is $\boldsymbol{K} \otimes \boldsymbol{I}$, with $\boldsymbol{I}$ being a n-dimensional identity matrix. This results in a total rank deficiency equal to n. The null space is the Kronecker product between the null space of the temporal IGMRF and the n dimensional identity matrix, named $\mathbf{M}2 = \frac{1}{n}\mathbf{1}_{m\times 1} \otimes \boldsymbol{I}_n$. This leads to n sum-to-zero constraints since one constraint is required for each area, or each one of the the n time series in $\boldsymbol{d}$.

The interaction type 3 specifies one spatial field at each time point t. Thus, $\boldsymbol{d}$ is a spatially structured first order IGMRF, such as θ. The rank deficiency of $\mathbf{I}_m \otimes \boldsymbol{R}$ is equal to m (assuming we have a connected graph). The resulting null space can be defined as $\mathbf{M}3 = \mathbf{I}_m \otimes \mathbf{1}_{n\times 1}\frac{1}{m}$ and the constraints apply to the each of the m spatial fields in $\boldsymbol{d}$.

The interaction type 4 is now an IGMRF along both directions, temporal and spatial. The rank deficiency of $\boldsymbol{K}\otimes\boldsymbol{R}$ is $n+m-1$. Practically speaking, each of the n time series needs to sum-to-zero as well as each spatial field at each of $m-1$ time points. The needed constraints correspond to those for the interaction type 2 plus $m-1$ of those for interaction type 3. Or the other way around, namely $n-1$ of those for interaction type 2 plus all for interaction type 3.

In practice often so-called *soft constraints* are considered, see Rue and Held (2005) and Rue and Held (2010). These means that the linear combinations are specified with some uncertainty, which is equivalent to posterior conditioning on observations of these linear combinations. To be more specific, it implies conditioning a vector $\boldsymbol{w}$ of length k whereby $\boldsymbol{w} \mid \boldsymbol{\delta} \sim N(\boldsymbol{M}\boldsymbol{d}, \mathcal{V})$. Here, $\boldsymbol{M}$ is a matrix of dimension $k \times (n \cdot m)$, where each row specifies one of the k required sum-to-zero constraints. Further, $\mathcal{V} = v\boldsymbol{I}$ and $v > 0$ is small. The aim is to evaluate properties of the conditional distribution for $(\boldsymbol{d} \mid \boldsymbol{w})$, which is Gaussian with precision matrix equal to

$$\boldsymbol{Q} + \boldsymbol{M}^T\mathcal{V}^{-1}\boldsymbol{M}\ , \tag{4}$$

see Eq. 2.34 in Rue and Held (2005). (Here, we use $\boldsymbol{Q}$ to denote a general precision matrix for $\boldsymbol{d}$). However, $\boldsymbol{M}^T\mathcal{V}^{-1}\boldsymbol{M}$ is a dense matrix making the direct work with this precision matrix infeasible.

An alternative approach is to work with the unconstrained prior distribution and correct the distribution later to be constrained, as in Eq. 12.21 in Rue and Held (2010). The evaluation of the conditional density of δ given the *soft constraint* requires to work with the covariance matrix of the constraint given δ, which is equal to $\mathbf{M}^T\mathbf{Q}^{-1}\mathbf{M}+\mathbf{V}$ which is a dense matrix of dimension k. Therefore the additional cost is to factorize this matrix, and the cost of k linear constraints is $O(nk^2)$. The $\mathbf{Q}^{-1}$ is considered through its Cholesky factor which is already available in the model fitting. This computation is required once for each new hyperparameter in an MCMC algorithm or each hyperparameter configuration in the INLA algorithm.

3 Avoiding the sum-to-zero constraints

In this section we present two approaches to avoid the computation of the sum-to-zero constraints while doing the model fitting. The first approach considers a contrast parameterization and the second the addition of a small diagonal to the precision matrix. For both cases the constrained effect can be computed after the fitting process considering the null space in a set of linear combinations.

3.1 The contrast parameterization

The contrast parameterization is usually considered when working with factors as fixed effects in linear regression. In this case one level is set as the reference level and there are coefficients for each one of the remaining levels measuring the contrast effect from the reference level. This same strategy can be considered for IGMRF in order to prevent intrinsic aliasing, Clayton (1996).

Let $\mathbf{x}$ be a random vector following a first order IGRMF with precision matrix $\mathbf{Q}_x$ that has rank deficiency equal to k. The contrast parameterization for a first order IGMRF implies setting k elements of $\mathbf{x}$ equal to zero. Let I be a vector of length k identifying the elements of $\mathbf{x}$ that will be set equal to zero. When considering k elements of x to be equal to zero, the conditional distribution precision matrix of the remaining elements under this condition, $\mathbf{x}|\mathbf{x}_I=\mathbf{0}$, is just the original precision matrix but without the I lines and I columns.

Let $\mathbf{x}^{(I)}$ be a vector with same length of $\mathbf{x}$ where the elements a the I entries are set to zero and the remaining elements have a precision matrix equal to the precision matrix of $\mathbf{x}|\mathbf{x}_I=\mathbf{0}$. The null space of $\mathbf{Q}_x$, $\mathbf{M}_x$, can be considered to define linear combinations when computing the overall constrained random vector, $\mathbf{x}|\mathbf{M}_x\mathbf{x}=\mathbf{0}$. These linear combinations can be defined as $\mathbf{I}-\mathbf{M}_x$.

Next we illustrate the application of this idea for an RW1 model and consider the linear combinations for the space-time case.

3.1.1 A one dimensional example in time

Consider a RW1 for $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)^\top$, with precision matrix $\tau_\alpha \boldsymbol{K}$. Instead of constraining $\boldsymbol{\alpha}$ to sum to zero, i.e. $\alpha \mathbf{1} = 0$, one can set one of the elements of $\boldsymbol{\alpha}$, α_l say, to be zero. The conditional distribution of $\boldsymbol{\alpha} \mid \alpha_l = 0$ is Gaussian where the precision matrix is the original precision matrix omitting the l-th row and the l-th column, see Section 16.6.3 in Clayton (1996).

The implication is that the re-parameterized random walk on $\boldsymbol{\alpha}^{(l)} = (\alpha_1^{(l)}, \ldots, \alpha_m^{(l)})^\top$ models the differences $\alpha_1 - \alpha_l, ..., \alpha_m - \alpha_l$. When modeling data with a Poisson likelihood and log-link function, this corresponds to modelling the log-risk at each time with respect to time l, instead of with respect to the average period effect as it is done using the original sum-to-zero constrained $\boldsymbol{\alpha}$. In fact $\alpha_1^{(l)}, \ldots, \alpha_{l-1}^{(l)}, \alpha_{l+1}^{(l)}, \ldots, \alpha_m^{(l)}$ are shifted from the original effects $\alpha_1, \ldots, \alpha_{l-1}, \alpha_{l+1}, \ldots, \alpha_m$ by α_l and α_l itself is missing. However, the original effects α_t can be obtained based on the reduced set $\alpha_t^{(l)}$ using

$$\alpha_t = \alpha_t^{(l)} - \frac{1}{m} \sum_{t=1}^{m} \alpha_t^{(l)} \tag{5}$$

which is a linear combination defined on $\boldsymbol{\alpha}^{(l)}$.

Let us consider a sub-dataset from the data analyzed in Section 6. The observed and expected yearly number of deaths among children less than one year old in Curitiba city from 1996 to 2015 are shown in the top plot of Figure 1 together with the posterior mean and 95% credibility interval for the infant mortality rate. We illustrate the constraint parameterization considering two scenarios. In the first, we set α_1 equal to zero and in the second α_{10}, which corresponds to the years 1996 and 2005, respectively.

The posterior for $\boldsymbol{\alpha}$ and the 95% credibility interval are shown as continuous line in the central panel of this Figure. The re-parameterized version of it for $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\alpha}^{(10)}$ is shown in the same plot as dashed and dotted lines, respectively. The medians are shifted with respect to the median curve for $\boldsymbol{\alpha}$, by the median value of $-\alpha_1$ and $-\alpha_{10}$, respectively. The quantile based 95% credibility interval infer the difference between the risk at time t with respect to the reference time and do not compare with the credibility interval properties of $\boldsymbol{\alpha}$. The linear combinations to re-compute the original $\boldsymbol{\alpha}$ are shown in the bottom plot together with $\boldsymbol{\alpha}$ obtained considering the sum-to-zero constraint. Looking at this plot, one can see only numerical differences between these quantities computed using the three different parameterizations. The same applies for the linear predictor quantiles, shown in the top panel.

An equivalent soft constraint can be formulated that makes $\boldsymbol{M}^T \mathcal{V}^{-1} \boldsymbol{M}$ in Eq. (4) diagonal. This happens because now $\boldsymbol{M}$ is a row-vector of length m, in which only element l is different from zero. Thus, $\boldsymbol{M}^T \mathcal{V} \boldsymbol{M}$ is a diagonal matrix of dimension $m \times m$. Of note, here in the one dimensional case we only have one constraint, i.e. $k = 1$, so that $\mathcal{V}^{-1}$ reduces to a number, namely $1/v$. Consequently, the structure matrix for $\boldsymbol{\alpha}$ using this *soft-constraint* has a high value, namely $1/v$, added to the element (l, l):

$$(\boldsymbol{K}^{(l)})_{i,j} = \begin{cases} \boldsymbol{K}_{ij} + 1/v & \text{if} \quad i = j = l \\ \boldsymbol{K}_{i,j} & \text{otherwise.} \end{cases} \tag{6}$$

and $\boldsymbol{K}^{(l)}$ is positive definite for $v > 0$.

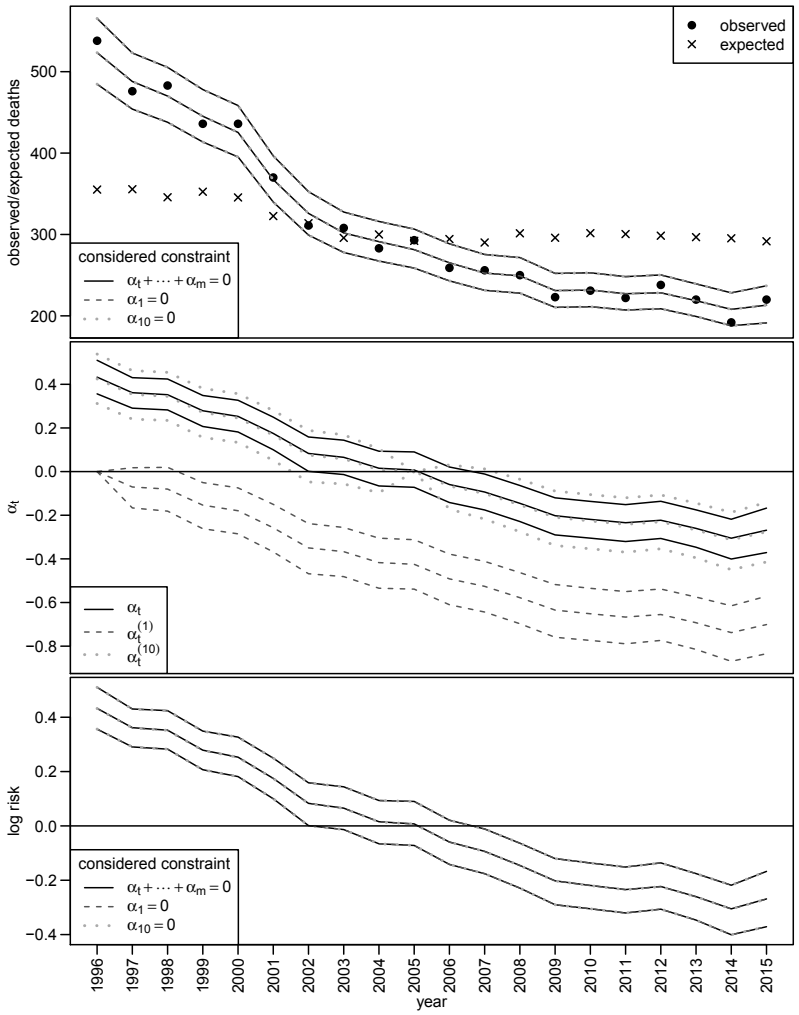


Figure 1: Observed and expected children deaths in Curitiba and the posterior 2.5% , 50% and 97.5% quantiles for the infant mortality rate considering the three parameterizations (top). The 2.5%, 50% and 97.5% quantiles for α_t, $\alpha_t^{(1)}$ and $\alpha_t^{(10)}$ (mid). The 2.5%, 50% and 97.5% quantiles for the log-relative temporal risk computed from each parameterization (bottom).

3.1.2 A spatio-temporal example

Applying the contrast parameterization in a spatio-temporal setting, we do need to set one time point l as a reference for the interaction type 2, one area r as the reference area for the interaction type 3, and both, one time point l and one area r, as the reference levels for the interaction type 4. Compared to the original structure matrix, the new structure matrices will have n rows and columns removed for type 2, m for type 3 and $n + m - 1$ for type 4 assuming a RW1 for time.

A special case occurs when there are no main intrinsic effects present in the model formulation. In this case, only one overall sum-to-zero constraint is needed to make a potential intercept in the linear predictor identifiable. This element refers simultaneously to the area and time reference and only this row and column are dropped from the precision matrix in (2). In the *soft constraint* version the value $1/v$ will be added to the element $k \times l$ of its diagonal.

The constrained sum-to-zero d_{it} effects can be obtained analogously to Equation (5). For the interaction type 2 we define the following linear combinations

$$d_{it} = d_{it}^{(l)} - \frac{1}{m}\sum_{t=1}^{m} d_{it}^{(l)} \ , \tag{7}$$

given that $d_{it}^{(l)} = 0$ if $t = l$. Similarly for the interaction type 3 we have

$$d_{it} = d_{it}^{(k)} - \frac{1}{n}\sum_{i=1}^{n} d_{it}^{(k)}. \tag{8}$$

For the interaction type 4 we have to account for both directions and add a correction term as

$$d_{it} = d_{it}^{(k,l)} - \frac{1}{m}\sum_{t=1}^{m} d_{it}^{(k,l)} - \frac{1}{n}\sum_{i=1}^{n} d_{it}^{(k,l)} + \frac{1}{n \cdot m}\sum_{i=1}^{n}\sum_{t=1}^{m} d_{it}^{(k,l)}. \tag{9}$$

What about the linear combinations to regain the main effects $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$?

These linear combinations are actually related to the null space of the original precision matrix, which we will use in a different way in the following section. Let $\mathbf{M}$ be the null space of the precision matrix of the unconstrained space-time random effect of one type. The linear combinations can be written considering the matrix $\boldsymbol{C} = (\mathbf{I} - \mathbf{M}\mathbf{M}^{\top})$ as $\boldsymbol{d} = \boldsymbol{C}\boldsymbol{d}^{(k,l)}$.

The corresponding *soft constraint* implies the addition of the value $1/v$ to more than one element of the structure matrix of the space-time interaction effect. The precision structure for the re-parameterized $\boldsymbol{d}$ assuming an interaction type 4 is the original $\boldsymbol{K} \otimes \boldsymbol{G}$ with a high value added to the diagonal corresponding to the time point l or area r. For example, if $k = l = 1$, we have to add $1/v$ to the first n elements in the diagonal as well as to elements $n + 1$, $2 \cdot n + 1, \ldots, (m - 1) \cdot n + 1$. For the interaction type 2, the value $1/v$ is added to the n elements of the precision matrix corresponding to the area set as reference, while for the interaction type 3 it will be added to the m elements corresponding to the reference time point.

3.2 Addition of a small diagonal

The contrast parameterization might be criticized for the need to select reference categories. Here, we consider an alternative approach. The idea

is based on an appealing argument for constructing IGMRF's as a liming case, see end of Section 3.3.2 and Eq. (3.37) in Rue and Held (2005). In this construction, a small value $\gamma > 0$ is added to the entire diagonal of the precision matrix of the IGMRF making it positive definite.

Theorem 1. *Let* $\mathbf{x}$ *be a* n*-dimensional IGMRF with precision matrix* $\mathbf{Q}$ *that has null space as the columns of the* $n \times k$ *matrix* $\mathbf{M}$*, which are the eigenvectors of* $\mathbf{Q}$ *with corresponding eigenvalues equals to zero. Let* $\mathbf{x}^{(\gamma)}$ *be a GMRF with precision matrix* $\tilde{\mathbf{Q}} = \mathbf{Q} + \gamma\mathbf{I}$*. The strong and weak norm of the difference between the precision matrix of* $\mathbf{x} \mid \mathbf{M}\boldsymbol{x} = 0$ *and the precision of* $(\mathbf{I} - \mathbf{M}\mathbf{M}^T)\mathbf{x}^{(\gamma)}$ *are* γ *and* $\gamma\sqrt{(n-k)/n}$*, respectively. Thus the difference between* $\mathbf{x} \mid \mathbf{M}\boldsymbol{x} = 0$ *and* $(\mathbf{I} - \mathbf{M}\mathbf{M}^T)\mathbf{x}^{(\gamma)}$ *is controlled by* γ*.*

Proof. Given in the appendix at section 8.1. □

The matrix $\boldsymbol{M}$ appears in the linear combination $\boldsymbol{C}\boldsymbol{x}$ where $\boldsymbol{C} = \boldsymbol{I} - \mathbf{M}\mathbf{M}^T$, illustrated in Equations (5), (7), (8) and (9). Thus, if one considers the prior precision matrix $\tilde{\mathbf{Q}} + \gamma\mathbf{I}$ to fit the model, the result can then be considered to compute $\mathbf{x}^{(\gamma)}$ considering the posterior for $\mathbf{x}$, in order to account for the null space. Therefore, one can avoid to account for the computationally expensive constraints during the model fitting process. Notice that even though adding a diagonal to $\mathbf{Q}$ will make the prior distribution proper, the model will be still over-parameterized because the null space remains unchanged.

4 Model scaling and prior distributions

In this section we discuss two key ideas that make the modelling with IGMRFs easier in terms of parameter interpretation and comparison between different model components.

1. Scaling an IGMRF to make its hyperparameter marginally interpretable and also comparable between different IGMRF effects in the model.

2. Prior distributions that are interpretable and practical for the applied scientist.

4.1 Scaling an IGMRF

Any IGMRF has a local precision parameter, that controls the differences between neighbours. Larger values indicate a stronger relationship, i.e. more smoothing, while lower values indicate a weaker relationship, i.e less smoothing. This parameter depends on the size and connectivity of the graph that underlies the IGMRF, so that precision parameters of IGMRFs that have a different size or are based on different graphs are not marginally comparable. To make them comparable the structure matrix needs to be scaled so that the marginal variances have a geometric average which is equal to one, see Sørbye and Rue (2014) for details.

In the model given in Equation (1) a spatial main effect $\boldsymbol{s}$ is included which splits in two components, a structured effect $\boldsymbol{\theta}$ and an unstructured effect $\boldsymbol{\phi}$. The structured effect $\boldsymbol{\theta}$ follows an IGMRF. It has zero mean and precision matrix $\tau_\theta\mathbf{R}$. The scaled precision matrix is given by $\tau_\theta\boldsymbol{R} \cdot \sigma^2_{GV}(\boldsymbol{R})$,

where $\sigma_{GV}(\boldsymbol{R})$ denotes the generalized variance of $\boldsymbol{R}$ as proposed by Sørbye and Rue (2014):

$$\sigma^2_{GV}(\boldsymbol{R}) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log(\text{diag}(\boldsymbol{R}^{-}))\right).$$

Here, $\boldsymbol{R}^{-}$ denotes the generalized inverse of $\boldsymbol{R}$. It implies that the marginal variance of each element of a random vector modeled as a scaled IGMRF is one.

We then follow Riebler et al. (2016) and reparameterise the classical BYM model as

$$s_i = \frac{1}{\sqrt{\tau_s}}(\sqrt{\omega_s}\theta_i + \sqrt{1-\omega_s}\phi_i) \tag{10}$$

where $\boldsymbol{\phi} \sim N(\mathbf{0}, \boldsymbol{I})$. As both components $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are standardized, $\tau_s > 0$ is a precision parameter and $0 < \omega_s < 1$ is a weight parameter defining the proportion of the main spatial effect that is explained by the structured component. The benefit from this parameterization is that we can consider $\sigma_s = 1/\sqrt{\tau_s}$ as a marginal standard deviation parameter.

For the temporal effect $\boldsymbol{\alpha}$ we scale the corresponding structure matrix $\mathbf{K}$, and for the space-time interaction term we we use the scaled versions of the structure matrices $\boldsymbol{R}$ and $\boldsymbol{K}$ in the Kronecker product. Thus, the precision parameter of the structured temporal effect $\boldsymbol{\alpha}$ can be interpreted as the inverse of the marginal variance giving standard deviation parameter $\sigma_\alpha = 1/\sqrt{\tau}_\alpha$, and the space-time interaction term precision can be interpreted as the inverse of the marginal variance giving standard deviation parameter $\sigma_d = 1/\sqrt{\tau}_d$.

Scaling all three IGRMF model components will allow us to compare the posterior estimates of their precision (variance) parameters between them.

4.2 Definition of hyperprior distributions

Scaling the IGMRFs does not only allow us a comparison of the hyperparameter estimates, but also facilitates the assignment of hyper prior distributions. This is because the precision parameter of a scaled IGMRF is the marginal precision. Thus the square root of its inverse is the marginal standard deviation which measures the size of the effect in the scale of the linear predictor.

Here, we consider the penalized complexity (PC) prior framework proposed by Simpson et al. (2016). The main idea under this framework is to consider a base model and the model component being considered as a more complex case. In this idea, the parameter being considered measures the deviation from the base model. For the random effects case, the base model is the straight line, which has zero variance. Thus, the variance (or standard deviation) measures deviation from the base model towards a more complex one.

One nice feature of this way of setting prior distributions is that the base model has positive density a priori. Therefore, this kind of prior supports the Occam's razor as the more parsimonious model is supported a priori. Let $\nu > 0$ being the parameter that controls deviations from the base model, implying that the bigger ν the more complex is the model. The PC-prior idea considers priors of the form $P(\nu > u) = \alpha$ and thus we have to choose u and α.

We can use the BYM prior for the main temporal effect v. One alternative is to consider the first order autoregressive model, AR1. One advantage in this model is when the parameterization is considered for the innovation drawn from the stationary distribution. In this case the ρ parameter is the lag one correlation or a persistence parameter and the precision parameter, τ_v, is the marginal precision. Thus, we have $\sigma_v = 1/\sqrt{\tau}_v$ as the marginal standard deviation and the PC-prior for this parameter is defined considering $P(\sigma_v > 0.5) = 0.5$. We also consider the prior for the standard deviation of the other random effects with $u = 0.5$ and $\alpha = 0.5$.

The ρ parameter, the persistence parameter, was considered having also a PC-prior considering the base model as $\rho = 1$. In the AR1 model with innovations with distribution equals the stationary one, when $\rho = 0$ we have white noise and for $\rho = 1$ we have the straight line, there is no innovations. Thus we can choose the base model as the case for $\rho = 1$. This means that the PC-prior penalizes deviations from the straight line. Details about this prior are in Simpson et al. (2016). The PC-prior parameters considered for ρ considers that $P(\rho > 0) = 0.9$.

In the main spatial effect the mixing parameter, ϕ defines the proportion of the variance due to the structured part. The PC-prior for this parameter depends on the eigenvalues of the structure matrix, see Appendix A3 in Simpson et al. (2016). Our choice considers the prior given the condition that $P(\phi > 0.5) = 0.5$.

These prior distributions for the five model parameters are shown in Figure 4 together with the posterior distributions obtained in the application of modelling infant mortality data in Paraná.

5 Simulation study

We perform three simulation cases with four scenarios each. In the first two cases we considered the Ireland map divided into 26 areas and in the third we considered the Columbus map containing 49 areas, both available in the **spde R** package, Bivand et al. (2012). In the three cases we have considered the number of time points to be equal 20. A Gaussian likelihood was considered for the first case and a Poisson for the other two cases.

In each case we considered four scenarios for the three precision parameters: τ_v, τ_s and τ_d. In the first and third case it was considered τ_v, τ_s, τ_d as (1, 9, 9), (9,1,9), (9,9,1) and (9,9,9) while in the second we have (1, 16,16), (16,1,16), (16,16,1) and (16,16,16). Thus, because the parameterization of the random effects is as in the previous section, these precision parameters are marginal and so we can compare them and have the last scenario where all the three effects are small and the first three scenarios when, respectively, the main temporal effect is bigger, the main spatial effect is bigger and the space-time interaction effect is bigger. We considered 250 data sets for each of these four combination of the precision parameters and so 1000 samples for each case in total.

The PC-prior for these precision parameters was considered with median in the value used to sample. Thus, for example, for τ_d in the third scenario we have $P(\sigma_d > 1) = 0.5$ and $P(\sigma_d > 1/3) = 0.5$ for the other three scenarios. The ρ and ϕ parameter were considered both equal 0.75, with prior as in the section 4. Thus $P(\rho > 0) = 0.9$ and $P(\phi > 0.5) = 0.5$.

The type 4 space-time interaction was considered to simulate the space-time effect. The type 1, 2, 3 and 4 were fitted considering the original sum-to-zero and both the contrast parameterization and the diagonal add approach were considered for the interaction types 2, 3 and 4 as well. Thus, for each simulated dataset we have 10 fitting results in total. The computer considered in this study is an old machine with 32 cores, each one an AMD Opteron(tm) Processor 6136 with only 800MHz and the cache with only 512KB. Our strategy to run the simulation study was to use only one thread to fit each model and the 1000 replications were done in parallel. Thus, the reported CPU time is at least three times bigger than one can have with a common laptop and much smaller in a desktop with several cores.

We summarize the results here and present them in more details in the Appendix 8.2.

In all the cases we have seen that the two approaches can be considered with the objective of selecting the best space-time interaction type. This is because the DIC, WAIC and CPO were smaller for the interaction type 4 in general, which was the structure considered to simulate the datasets.

When the space-time effect is relatively bigger than a main effect, the contrast parameterization approach causes some impact in the hyperparameters for the main effects. This is because the space-time effect at one area or one time point is assigned by the contrast parameterization to the respective main effect.

The add diagonal approach makes the estimation for τ_d, ρ, τ_s and ϕ affected too much as the space-time interaction structure takes over the main temporal and spatial variations. So when having the interaction type 4 all these four hyperparameters are affected, when having the interaction type 3, τ_s and ϕ are affected, and when having the interaction type 2, τ_v and ρ are affected.

However, these two approaches can be considered to fit the relative risks without problems since the mean squared errors - MSE for these are not affected by the approach.

6 Application

In this section consider the number of born and children deaths in 399 municipalities in Paraná State, Brazil, from 1996 to 2015. This data is freely available in the web site of the Informatics Department of SUS, *Departamento de Informática do SUS - DATASUS*, at `http://datasus.saude.gov.br`.

The observed rates time series for each area are shown in Figure 2. This provides an overall impression of the trend, since too many time series are in one plot. The rates mainly vary from near zero up to 4%, however, some areas presents rates above 4% at some years. This variation is expected to happen when the population at risk is not big. There is an overall decay over the period, as seen clearly in the rate computed over the entire State for each year, plotted in white line.

The maps shown in Figure 3 were drawn using the same classification color, given in the legend in the map for the last year. For each map, the number of areas with rate in each class is shown in the top right side of each map. Due to the decay pattern observed in Figure 2, the number of areas with higher rates decreases over the years, and the number of areas with

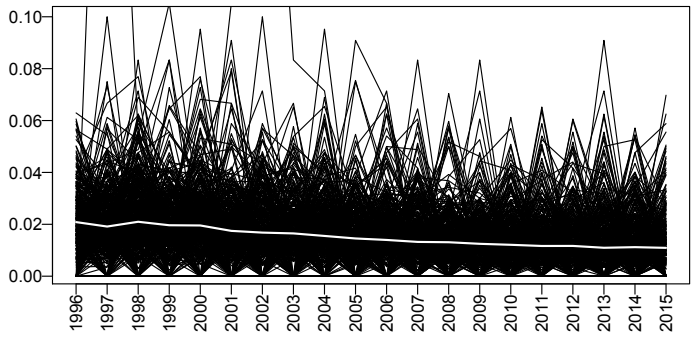


Figure 2: Time series of the observed infant mortality rates at each area.

lower rates increases.

We fit the models considered in Section 2 considering the original constrained prior distribution for the random effects, the contrast parameterization and the add diagonal approach considering the scaling and prior distributions as defined in Section 4.

In Table 1 we have the DIC, WAIC and CPO differences with respect to the model with the lower value, and the CPU time in minutes. The computing time was dramatically reduced by considering the contrast parameterization or the diagonal add approach. Instead of take over a week to fit the type 2 and type 4 with the sum-to-zero constraints, we have results in near half hour.

Model	DIC	WAIC	CPO	CPU (min)
1	190.70	163.00	135.40	3.70
2	49.50	51.90	32.90	14139.80
3	134.40	135.50	86.90	55.20
4	0.00	0.00	0.00	15309.80
2d	238.00	256.20	141.00	9.50
3d	142.60	140.40	90.70	25.80
4d	193.00	211.50	111.20	29.70
2c	170.40	194.30	89.50	11.70
3c	141.70	148.00	90.60	25.70
4c	89.00	102.10	43.20	33.80

Table 1: DIC, WAIC and CPO differences to the lowest and the CPU time.

The lowest DIC, WAIC and CPO was obtained for the interaction type 4 with the sum-to-zero constraints. These values were 29280, 29333.7 and 14686.4, respectively. We observe in Table 1 that the second lowest DIC, WAIC and CPO are when considering the interaction type 2 with sum-to-zero constraints and the third with the interaction type 4 with contrast parameterization.

If we consider the four interaction types with the sum-to-zero contrast, the choice is the type 4. If we consider only the interaction type 1 and the others with the contrast parameterization, the type 4 would be the choice. However, considering the type 1 and the others with the add diagonal

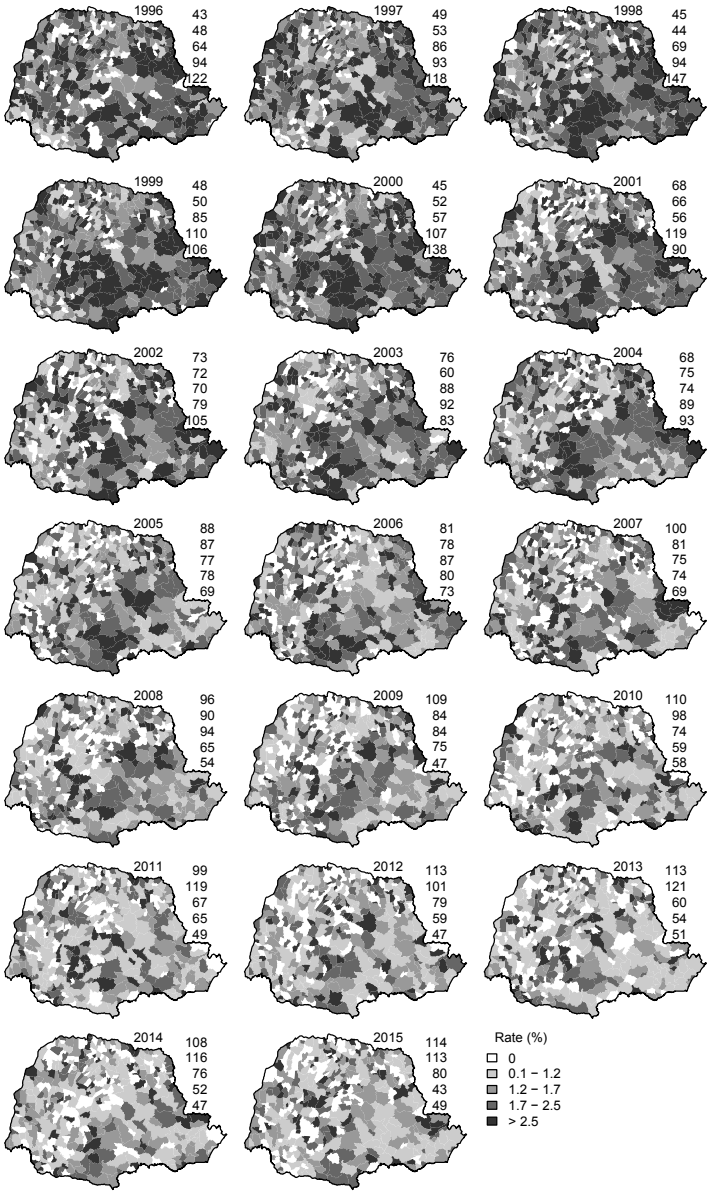


Figure 3: Maps of the observed infant mortality rates at each area and year.

approach, the choice would be the interaction type 3.

We have the summary for the posterior marginal distribution of each one of the hyperparameters considering the model fitted with the interaction type four and the sum-to-zero constraints in Table 2. First, it suggests that the two main effects have similar marginal precision and the space-time interaction has a bigger one. Thus, the main spatial terms are bigger than the space-time one. The main temporal term is very strongly structured because ρ is near one. This is in line of what we can observe in Figure 2. For the main spatial effect we have that the structured part is more important than the unstructured.

Parameter	Mean	S.D.	2.5% Q	Median	97,5% Q.	Mode
τ_v	27.65	18.46	5.30	23.48	74.66	14.56
ρ	0.96	0.03	0.90	0.97	0.99	0.98
τ_s	29.16	4.56	20.97	28.92	38.83	28.52
ϕ	0.82	0.10	0.60	0.84	0.96	0.88
τ_d	69.19	9.76	52.22	68.41	90.47	66.80

Table 2: Summary of the posterior marginal distribution for each one of the hyperparameters.

We have the prior and posterior distributions for each one of the hyperparameters in Figure 4.

The relative main temporal risk can be seen in Figure 5. This shows the decrease we have seen in Figure 2.

The relative main spatial risk can be seen in Figure 6. We have here that the West and and Northwest side are the region with an overall relative lower risk rates. The south and northeast sides are the worse areas. Curitiba, the capital of Paraná, has an overall relative low risk.

The relative space-time risk can be seen in Figure 7. We can see how it changed over time. In the begin of the period we have the South, center and some municipalities at the Northeaster, Northwest and near the ocean, extreme east side, with relatively higher risks. Them it slowly changed over time and more recently we have relatively higher risks at the Northwest side.

7 Discussion

The two proposed approaches to avoid the expensive sum-to-zero constraints were evaluated in a simulation study. It showed a great computational time reduction when the number of areas is more than few dozens. The savings is of the order of days in computational time.

The relative risk can be computed accounting for the null space in the linear combinations. These linear combinations considered to compute the relative risks showed to be accurate when comparing with the true relative risk in all the three simulation cases.

The hyperparameters of the main effects are affected by the approaches considered. Thus the proposed approaches can be used as screening tools when performing complex space-time models. This will deliver results in a short time and make the practitioner more freely to evaluate among different kind of model structures even when there is a lot of constraints needed.

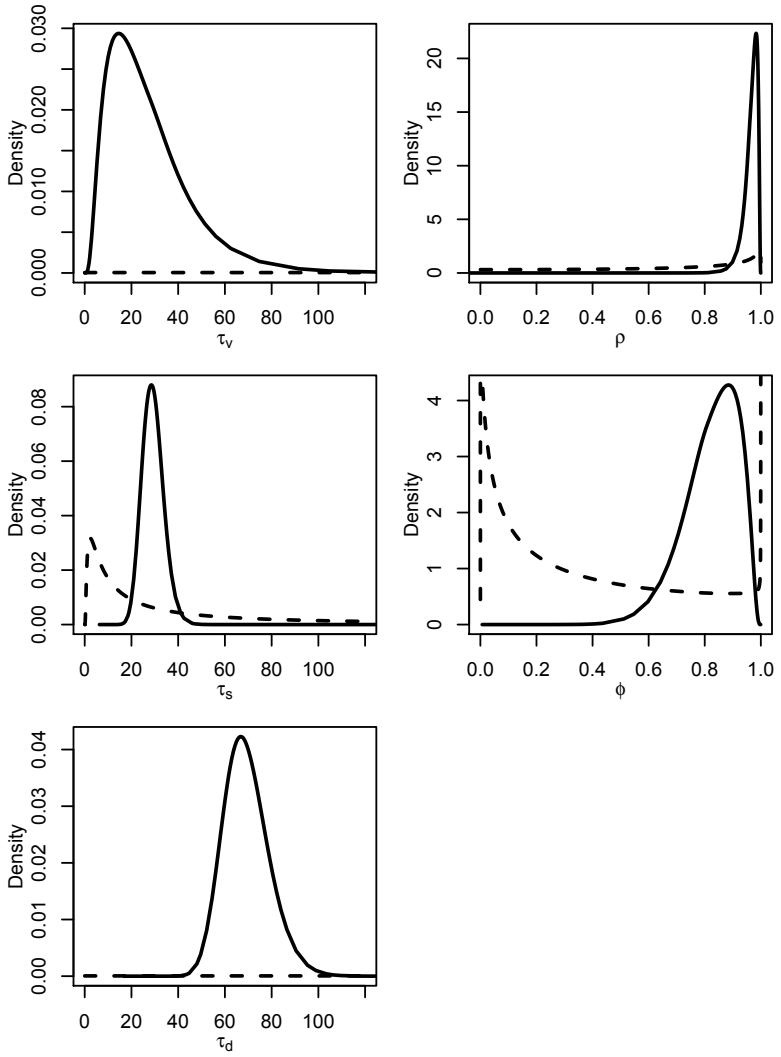


Figure 4: Density for the prior (dashed line) and posterior (continuous line) distribution of each hyperparameter.

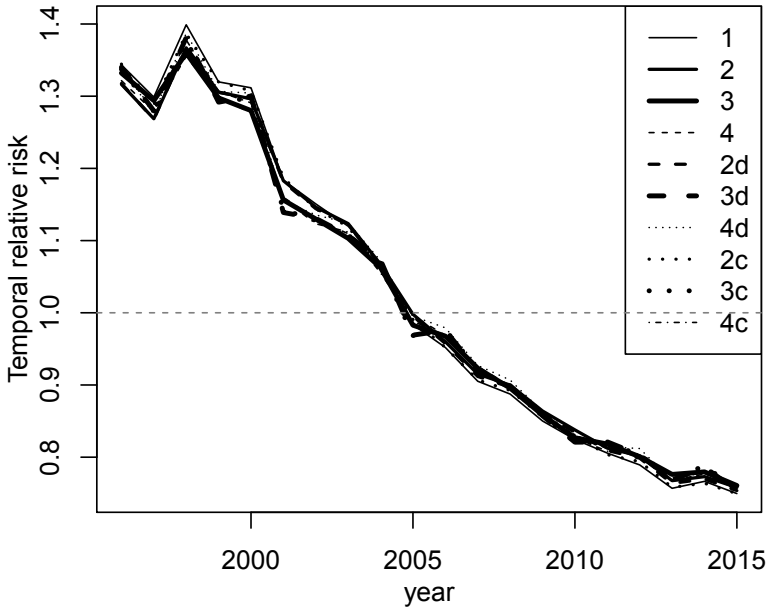


Figure 5: Posterior mean of the temporal relative risk considering all the 10 models.

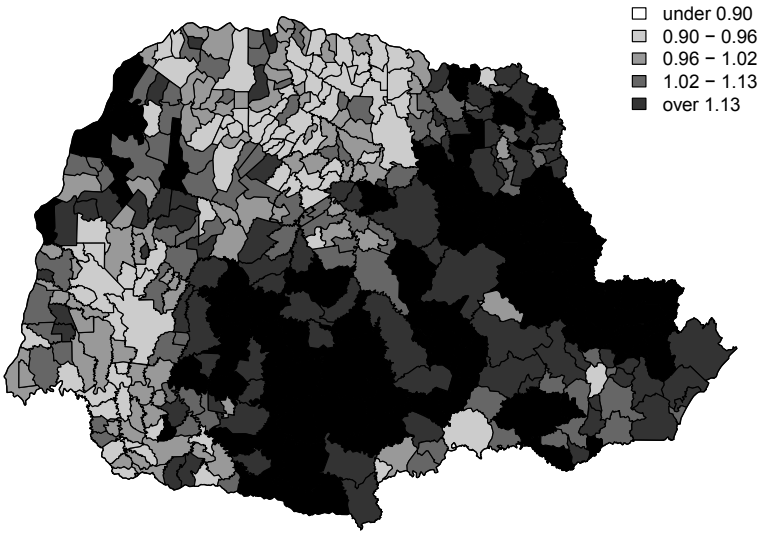


Figure 6: Posterior mean of the spatial relative risk considering model 4.

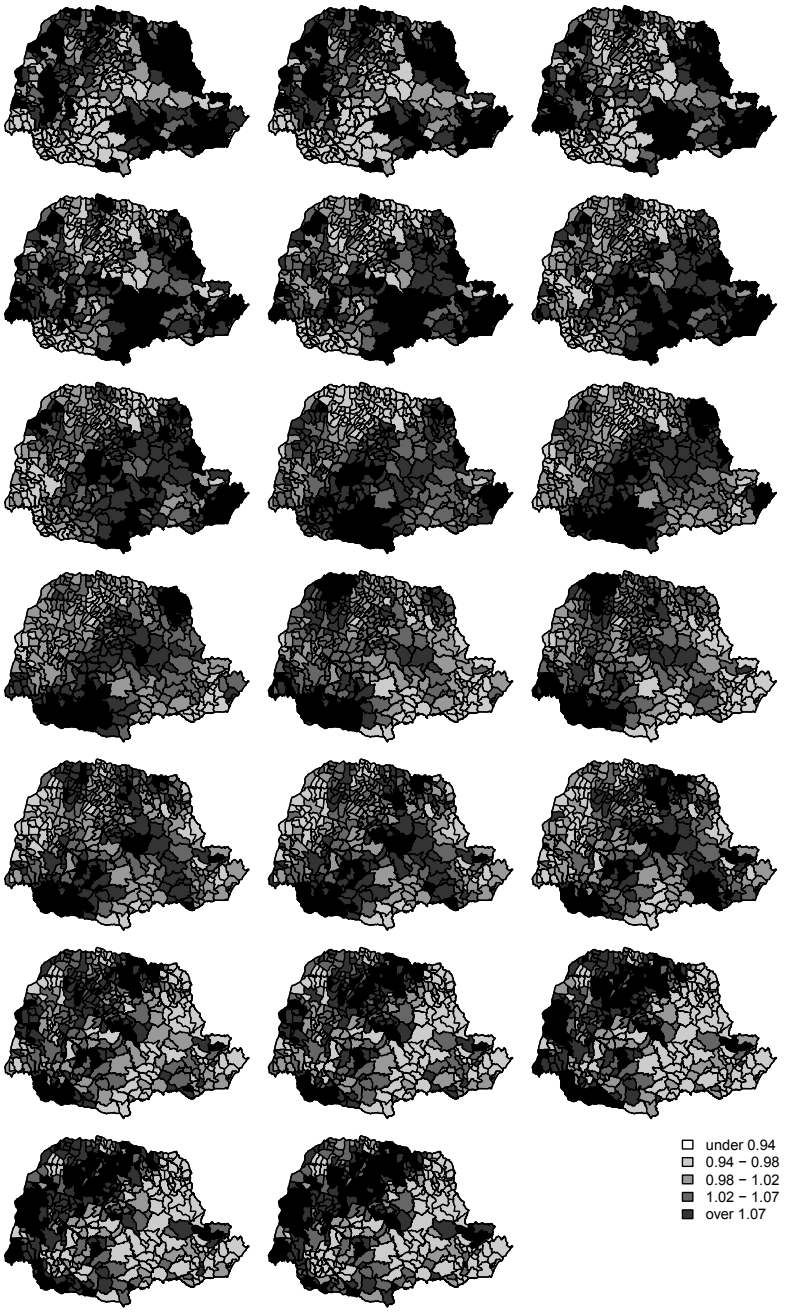


Figure 7: Posterior mean of the space-time relative risk considering model 4.

We suggest to consider the scaled intrinsic models in order to facilitate the prior assignment and to compare the different model components. The scaled random effects standard deviation can also be compared to regression coefficients of standardized covariates. So, it is useful to compare these two kind of effects being fitted together in a model.

References

Assunção, R. M., Reis, I. A., and Oliveira, C. D. (2001). Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian spacetime model. *Statistics in Medicine*, 20(15):2319–2335.

Bernardinelli, L., Clayton, D. G., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, 21–22(14):2433–2443.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). 43(1):1–59.

Bivand, R., with contributions by Micah Altman, Anselin, L., ao, R. A., Berke, O., Bernat, A., Blanchet, G., Blankmeyer, E., Carvalho, M., Christensen, B., Chun, Y., Dormann, C., Dray, S., Halbersma, R., Krainski, E., Legendre, P., Lewin-Koh, N., Li, H., Ma, J., Millo, G., Mueller, W., Ono, H., Peres-Neto, P., Piras, G., Reder, M., Tiefelsdorf, M., and Yu., D. (2012). *spdep: Spatial dependence: weighting schemes, statistics and models*. R package version 0.5-55.

Blangiardo, M. and M., C. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley.

Clayton, D. G. (1996). *Markov Chain Monte Carlo in Practice*, chapter Generalized Linear Mixed Models, pages 275–301. Chapman & Hall.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19:2555–2567.

Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC press.

MacNab, Y. C. and Dean, C. B. (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics*, 57(3):949–956.

MacNab, Y. C. and Dean, C. B. (2002). Spatio-temporal modelling of rates for the construction of disease maps. *Statistics in Medicine*, 21(3):347–358.

Martínez-Beneito, M. A., López-Quilez, A., and Botella-Rocamora, P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine*, 27(10):2874–2889.

Riebler, A., Søorbye, S. H., Simpson, D., and Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Monographs on Statistics & Applied Probability. Boca Raton: Chapman and Hall.

Rue, H. and Held, L. (2010). *Discrete spatial variation*, pages 171–200. Taylor and Francis.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.

Ruiz-Cárdenas, R., Krainski, E. T., and Rue, H. (2012). Direct fitting of dynamic models using integrated nested Laplace approximations — INLA. *Computational Statistics & Data Analysis*, 56(6):1808 – 1828.

Rushworth, A., Lee, D., and Mitchell, R. (2014). A spatio-temporal model for estimating the long- term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and Spatio-temporal Epidemiology*, 1(10):29–38.

Simpson, D., Rue, H., Martins, T. G., Riebler, A., and Sørbye, S. (2016). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28.

Sørbye, S. and Rue, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, 8.

Vivar, J. C. and Ferreira, M. A. R. (2009). Spatio-temporal models for Gaussian areal data. *Journal of Computational and Graphical Statistics*, 18(3):658–674.

Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92(438):607–617.

8 Appendix

8.1 Proof appendix

A background for this result is the alternative limit argument for the first-order IGRMF in the end of Section 3.3 in Rue and Held (2005).

Let a IGMRF, $\mathbf{x} = \{x_1, ..., x_n\}$, with precision matrix $\mathbf{Q}$ that has eigenvalue/vector pairs written as

$$(0, \boldsymbol{w}_1), ..., (0, \boldsymbol{w}_k), (\lambda_{k+1}, \boldsymbol{e}_{k+1}), ..., (\lambda_n, \boldsymbol{e}_n)$$

and we have that $\boldsymbol{Q}$ has rank deficiency equals k and the vectors $\boldsymbol{w}_1, ..., \boldsymbol{w}_k$ spans its null space. We can write

$$\mathbf{Q} = \sum_{i=k+1}^{n} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T \text{ and } \mathbf{Q}^- = \sum_{i=k+1}^{n} \lambda_i^{-1} \boldsymbol{e}_i \boldsymbol{e}_i^T$$

where $\mathbf{Q}^-$ is the generalized inverse of $\mathbf{Q}$. A practical interpretation of $\mathbf{Q}^-$ is that it does corresponds to the covariance of the constrained version of $\boldsymbol{x}$, $\mathbf{C}\boldsymbol{x}$, where $\mathbf{C} = \mathbf{I} - \mathbf{M}\mathbf{M}^T$ for $\mathbf{M} = [\boldsymbol{w}_1 \dots \boldsymbol{w}_k]$.

Let $x^{(\gamma)}$ to be a GMRF with precision $\tilde{\mathbf{Q}} = \mathbf{Q} + \gamma\mathbf{I}$ which is a positive definite matrix with eigenvalue/vector pairs as

$$(\gamma, \boldsymbol{w}_1), ..., (\gamma, \boldsymbol{w}_k), (\gamma + \lambda_{k+1}, \boldsymbol{e}_{k+1}), ..., (\gamma + \lambda_n, \boldsymbol{e}_n) \ .$$

Therefore we can write

$$\tilde{\mathbf{Q}} = \gamma \mathbf{M}\mathbf{M}^T + \sum_{i=k+1}^{n} (\gamma + \lambda_i) \boldsymbol{e}_i \boldsymbol{e}_i^T \text{ and } \tilde{\mathbf{Q}}^{-1} = \frac{1}{\gamma}\mathbf{M}\mathbf{M}^T + \sum_{i=k+1}^{n} \frac{1}{\gamma + \lambda_i} \boldsymbol{e}_i \boldsymbol{e}_i^T \ .$$

Define the linear combination $\mathbf{C}x^{(\gamma)}$. The covariance of $\mathbf{C}x^{(\gamma)}$ is

$$\begin{aligned} \mathbf{C}\tilde{\mathbf{Q}}^{-}\mathbf{C}^T &= [\mathbf{I} - \mathbf{M}\mathbf{M}^T][\frac{1}{\gamma}\mathbf{M}\mathbf{M}^T + \sum_{i=k+1}^{n} \frac{1}{\gamma + \lambda_i} \boldsymbol{e}_i \boldsymbol{e}_i^T][\mathbf{I} - \mathbf{M}\mathbf{M}^T] \\ &= [\mathbf{I} - \mathbf{M}\mathbf{M}^T][\sum_{i=k+1}^{n} \frac{1}{\gamma + \lambda_i} \boldsymbol{e}_i \boldsymbol{e}_i^T] = \sum_{i=k+1}^{n} \frac{1}{\gamma + \lambda_i} \boldsymbol{e}_i \boldsymbol{e}_i^T \end{aligned}$$

provided that $\mathbf{M}^T\mathbf{M} = \mathbf{I}$ and $\mathbf{e}_i^T\mathbf{M} = \mathbf{0}$.

The difference between the precision of $\mathbf{C}x$ and the precision of $\mathbf{C}x^{(\gamma)}$ is

$$\sum_{i=k+1}^{n} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T - \sum_{i=k+1}^{n} (\gamma + \lambda_i) \boldsymbol{e}_i \boldsymbol{e}_i^T = \gamma \sum_{i=k+1}^{n} \boldsymbol{e}_i \boldsymbol{e}_i^T \ .$$

The strong norm of this difference is γ while the weak norm is $\gamma\sqrt{(n-k)/n}$. Thus, letting $\gamma \to 0^+$ gives the precision of $\mathbf{C}\mathbf{x}^{(\gamma)}$ equivalent to the precision of $\mathbf{C}\mathbf{x}$.

8.2 Simulation results

8.2.1 Results for the first case

In Figure 8 we have the computing time in seconds. The type 1 takes less time to fit while the type 4 is the more expensive which is expected due the model structure and the number of constraints. In this small scenario, some saving in the CPU time due to the new approaches of contrast parametrization or diagonal add only happens for the type 3 interaction.

We also have considered the number of evaluations of the $p(\theta|y)$ during the optimization step to find the posterior mode for the hyperparameters. We checked this because it can, sometimes, be a not easy task when the model is not well specified. We have around 200 to 400 as number of function evaluations for among most of the 10 fitting results. When considering the diagonal add for type 3 it was a bit less and when considering the diagonal add for type 4 it was around 100 function evaluations.

We have considered common goodness of fit statistics, the deviance information criteria - DIC, the widely applicable information criteria - WAIC and the negative sum of the cross validated conditional predictive ordinate - CPO. For each dataset we considered the difference to the lower value among the 10 models fitted. Thus the difference for the best fit for each simulated data is zero.

The DIC is shown in Figure 9 considering each one of the four scenarios. The difference from the best fit is near zero in all the cases. It being a bit off zero for type 1 and type 3. As we simulated from the type 4, this was expected that this would have the best fit.

For the WAIC differences we have in Figure 10 that there were no differences among all the 10 models fitted for each data. Similarly for the CPO as shown in Figure 11, all the 10 models fitted delivered similar results.

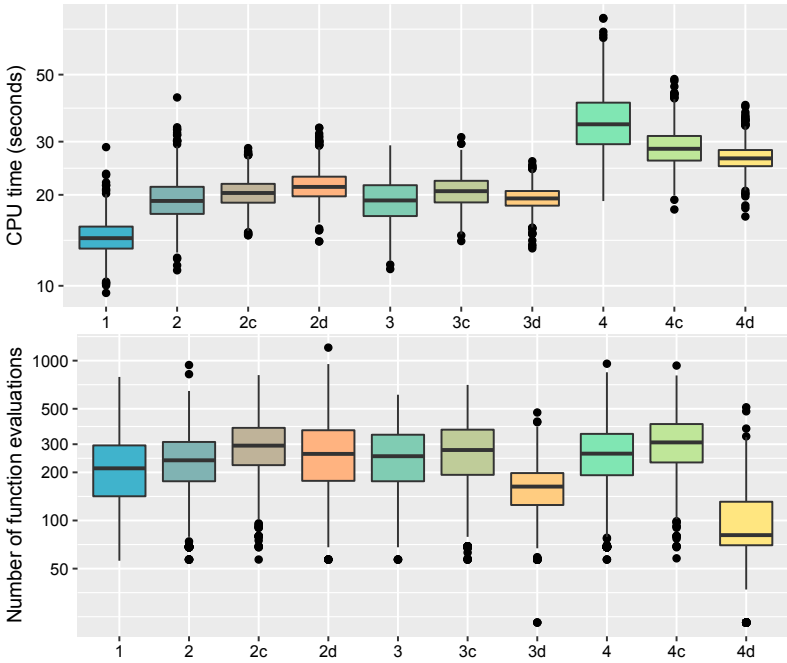


Figure 8: CPU time and number of function evaluations for the first case.

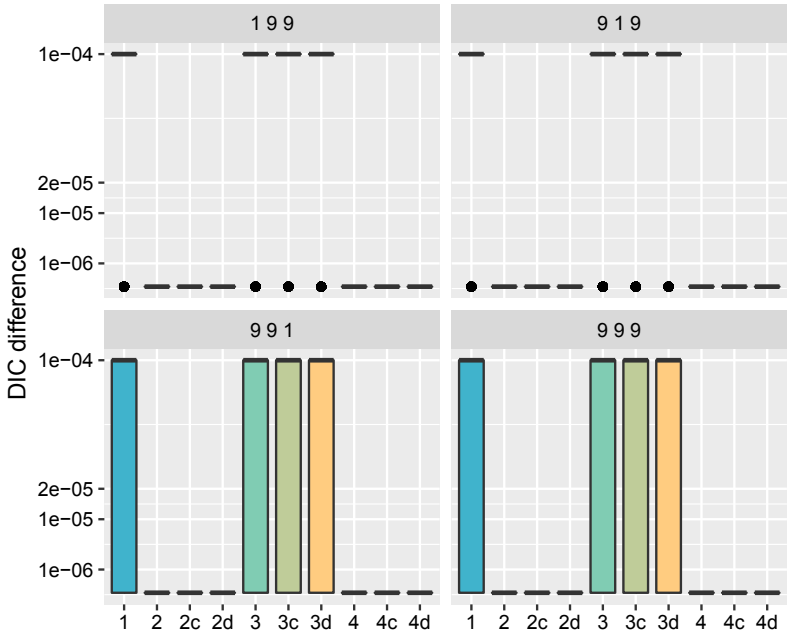


Figure 9: DIC differences considering for each scenario in the first case.

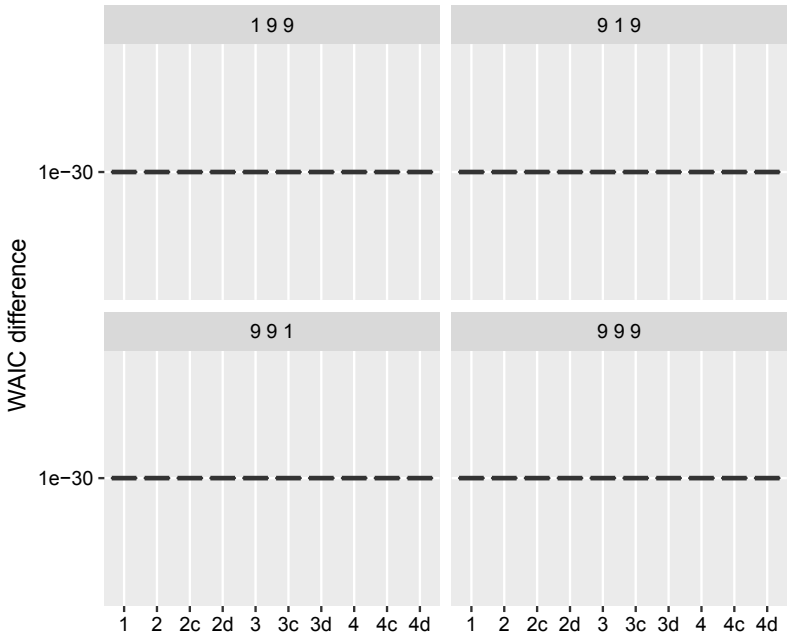


Figure 10: WAIC differences considering for each scenario in the first case.

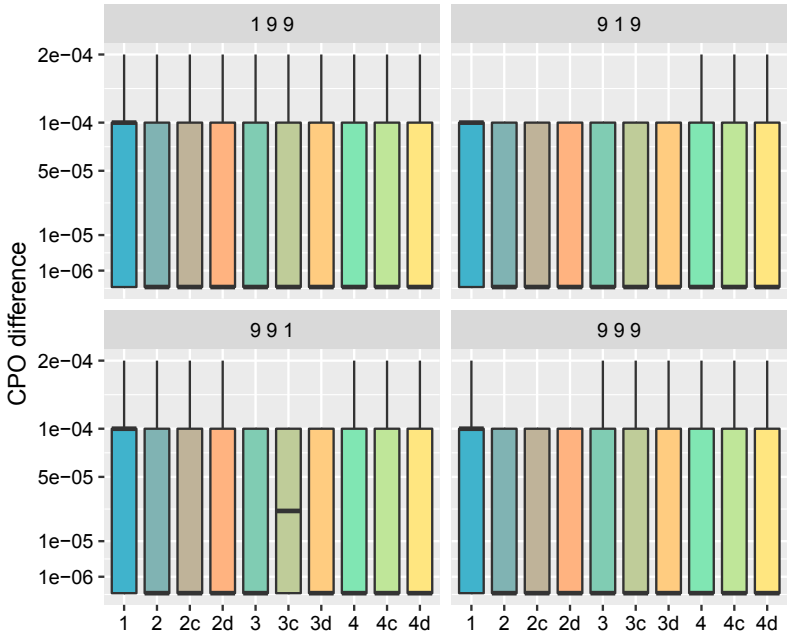


Figure 11: CPO differences considering for each scenario in the first case.

We also considered the mean squared error between the posterior mean of the random effect and the simulated. For the main temporal random effect we can see in Figure 12 that when this effect is bigger than the otters, the errors are also bigger and equally bigger among all the 10 models fitted. Similarly when the main effect or the space-time effect has the higher variance. When the space-time effect has the higher variance, we have the type 1 and type 2 fitted models delivering a bit higher errors for the main temporal effect.

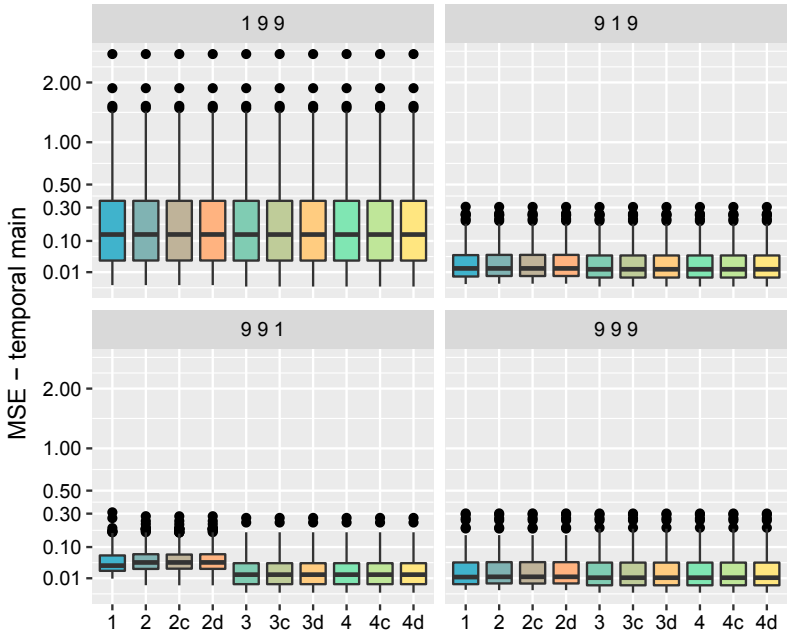


Figure 12: Mean squared error for the main temporal random effect.

For the main spatial effect errors, shown in Figure 13, we have it bigger when it has the bigger variance or when the space-time has it and we consider type 1 or type 3 interaction type. Considering type 1 or type 3 also delivers a bit bigger error for the main spatial effect when the main temporal effect is bigger or when all the three effects has the same precision.

is shown in Figure 26. It was smaller when the fitted model considers the type 4 space-time interaction, the same considered to simulate the data. This is the similar no matter the approach being considered to deal with the constraints. The bigger differences were for the case when this effect was bigger than the other two, with $\tau_d = 1$, when the fitted model considers type 1 or type 2.

We also considered the posterior mean for the five hyperparameters, τ_v, ρ, τ_s, ϕ and τ_d. We can start looking for the posterior means of τ_d, shown in Figure 15. We have that for most of the models being fitted, posterior mean of τ_v is close to the value used to simulate the data. Exception happens when considering the diagonal add approach and when the space-time effect is the biggest. When adding the small diagonal for the space-time precision prior we have this effect takes all the main temporal effect and so the higher

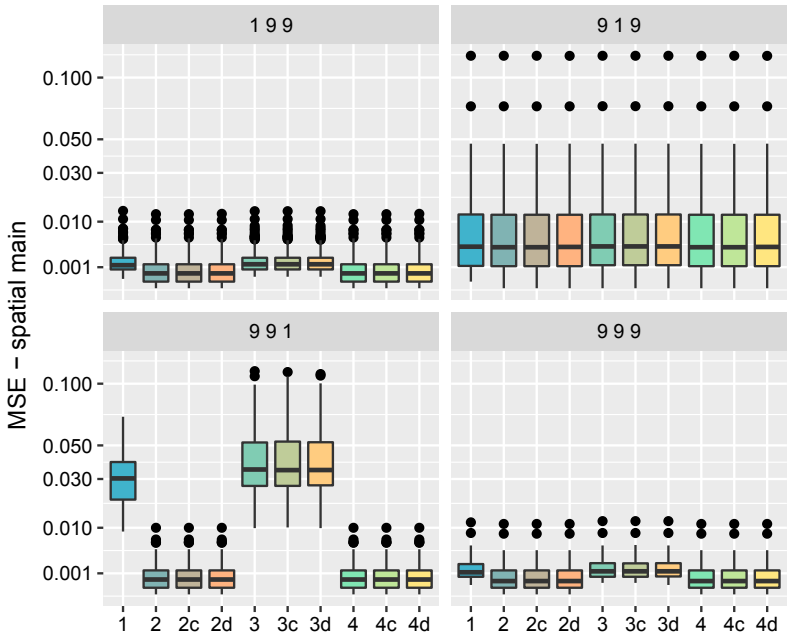


Figure 13: Mean squared error for the main spatial random effect.

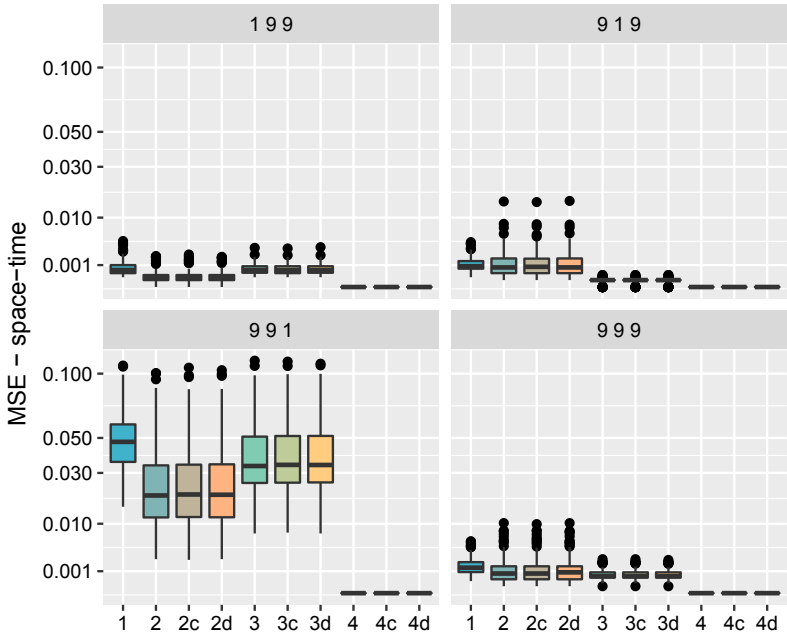


Figure 14: Mean squared error for the space-time random effect.

values for τ_v. This fact also happens for most of the other models being fitted when the space-time effect is bigger, only not happening for the type 4 with constraint during the fitting process.

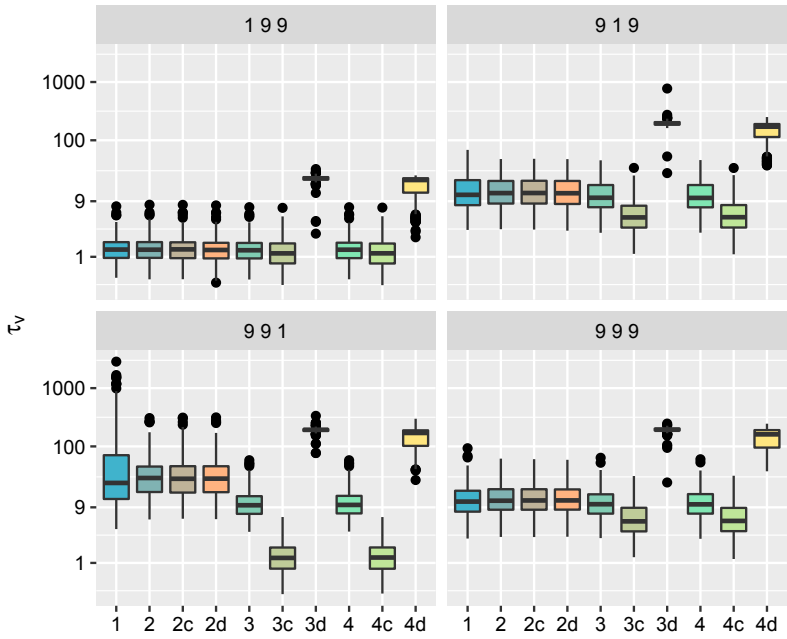


Figure 15: Posterior mean for τ_v for the first case.

We have to remember that when considering the contrast approach, the main effects being fitted does includes also the space-time effect for the set to zero time. This fact can be clearly seen when looking in Figure 15 when the main temporal effect is the biggest and in Figure 17 when the main spatial effect is the biggest. For those two situations we have almost no bias in the posterior for the respective precision parameter. Actually, this was the best fit among the others.

In Figure 16 we have the posterior means for ρ. The results shown here inherits the facts discussed for τ_v. Thus, we can see a near zero value of ρ when using the diagonal add approach for the interaction type 3 and 4. When the temporal effect is the biggest the other approaches delivered similar results. When the spatial is the case or when all the effects are of the same size the type 1 and using the contrast approach for type 2 and type 4 delivered a bit bigger posterior mean. When the space-time effect is the biggest, the posterior mean for ϕ is smaller for type 2, which is the space-time effect with temporal structure and spatially not structured.

The results for τ_s are shown in Figure 17. We have its posterior mean smaller when considering the contrast approach for type 2 and type 4 for the case when the main temporal is the biggest. This fact is more pronounced when the space time effect is the biggest. When the main spatial effect is the biggest we have the posterior mean for τ_s off when considering the diagonal add approach. When the space-time effect is the biggest we have it correct only for the type 2 and type 4 using the original approach. For the

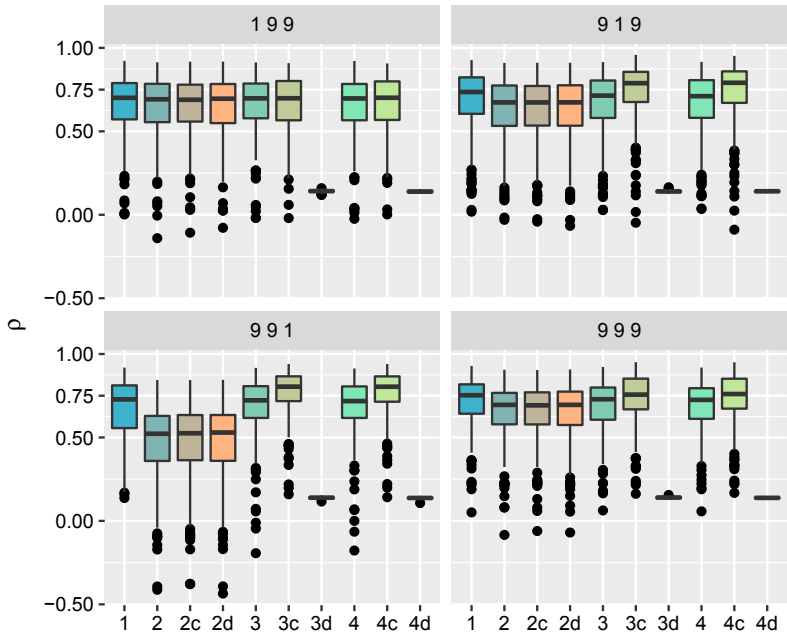


Figure 16: Posterior mean for ρ for the first case.

contrast approach τ_s suffers from the effect that we have the addition of the space-time effect at one time, the one considered to be zero. The two main facts, the higher τ_s when using the diagonal and the lower when using the contrast are also present when all the effects have the same size.

The results for ψ are shown in Figure 18. In all the fitted models we have it concentrated lower the value used to simulate the datasets. Similarly to the facts for ρ when using the diagonal add approach, were we have it now lower for type 3 and type 4, we have it here lower for type 2 and type 4 as the space-time effect takes all the spatial variation.

The precision parameter for the space-time interaction does not suffer any of the problems reported for the other parameters, as shown in Figure 19. We have no difference from considering the original sum-to-zero approach to the contrast parametrization or to the diagonal add approach. The differences are only due to the choice of its interaction type and we see that when it is type 4 we have no bias.

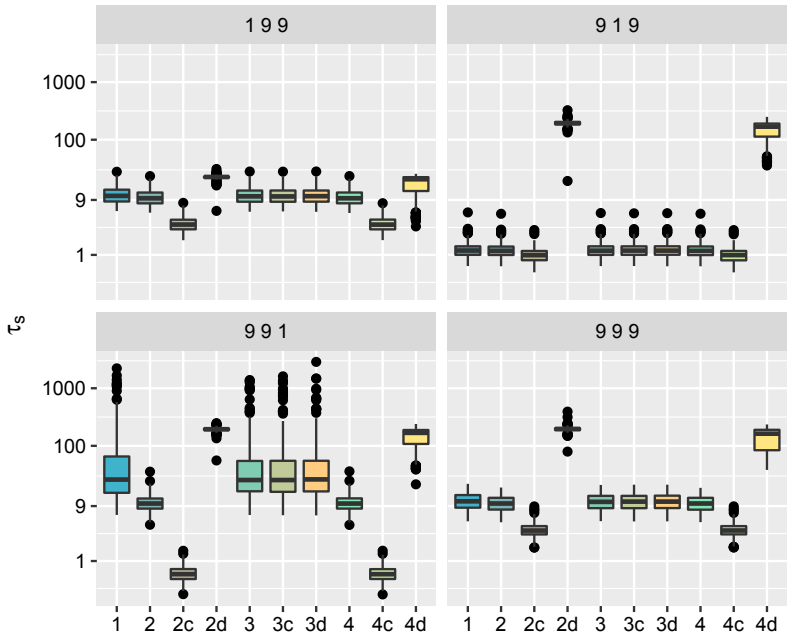


Figure 17: Posterior mean for τ_s for the first case.

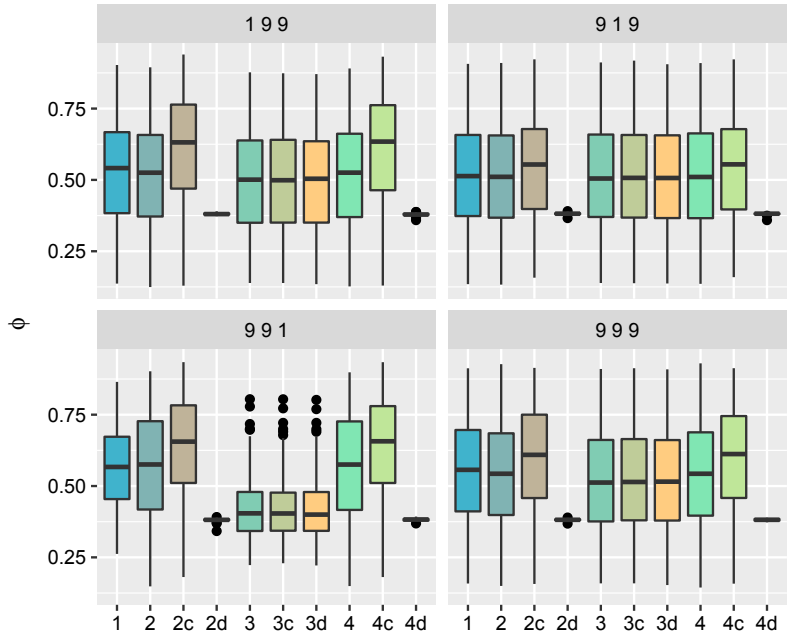


Figure 18: Posterior mean for ϕ for the first case.

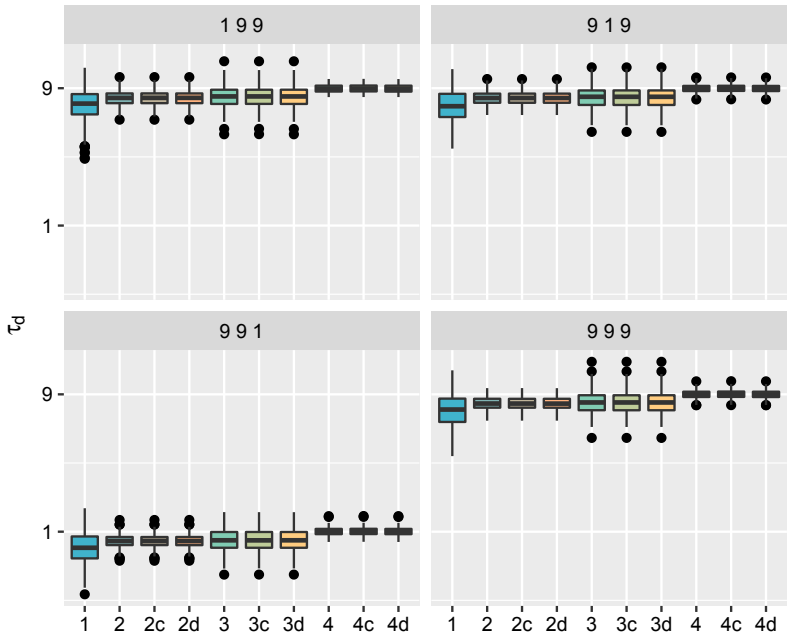


Figure 19: Posterior mean for τ_d for the first case.

8.2.2 Simulation results for the second case

In Figure 20 we have the computing time in seconds. The type 1 takes less time to fit while the for the interaction type 2, 3 and 4 with the sum-to-zero constraints are more expensive. The number of areas and time points in this case is the same as in the former. However, we have non-Gaussian likelihood and so the second Laplace approximation is computationally demanding when accounting for the sum-to-zero constraints. This happens even thou it is still not a lot of areas and times. Thus, already for this relatively small problem there is some computational advantage by avoiding the sum-to-zero constraints. We observe that when considering it the CPU time is approximately three times higher than when considering the contrast parametrization or the add diagonal approach. The number of function evaluations of the approximation for $p(\theta|y)$, bottom plot in Figure 8, was around 200 for all the 10 models fitted, just a bit smaller for type 4 when considering the sum-to-zero constraints or the add diagonal approach.

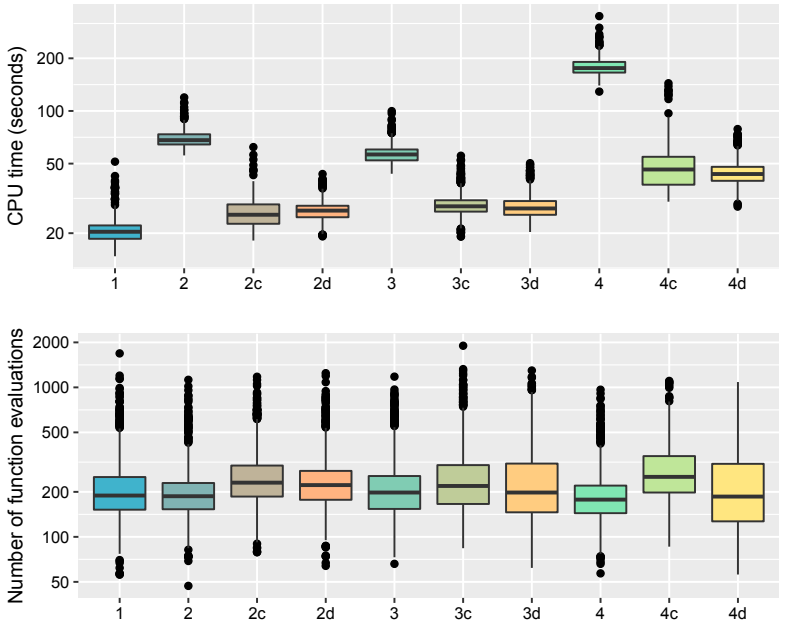


Figure 20: CPU time and number of function evaluations for the second case.

The results for the DIC differences to the lowest for each dataset is shown in Figure 21. We have that the interaction type 4 had the lowest DIC in general. However, when considering the contrast approach it was around 10 DIC units bigger and when considering the diagonal add around one DIC units bigger. Thus, even these differences, avoiding the sum-to-zero approaches still signs to the correct space-time structure.

The results for the WAIC differences are in Figure 22. Similarly to the DIC results it signs to the correct space-time structure no matter the approach being considered to fit the model. However, seems that WAIC

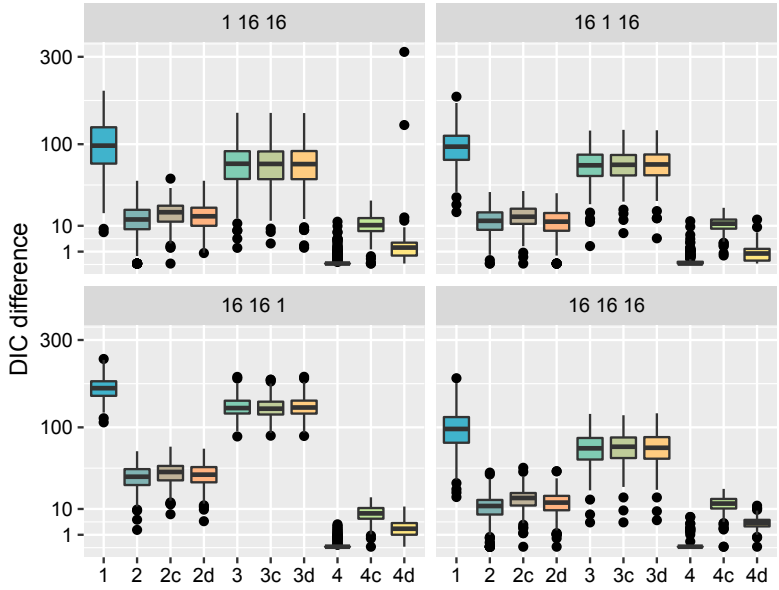


Figure 21: DIC differences considering for each scenario in the second case.

points out the correct model more clearly, specially when considering the contrast parametrization approach for most of the scenarios.

The results for the CPO are shown in Figure 23. These are similar to the DIC results.

The MSE results for the main temporal random effect we can see in Figure 24. These are very similar considering the different approaches to fit the models. It was a bit bigger only when the space-time effect is bigger than the others and it was considered the interaction type 1 and type 2.

The main spatial effect error results are shown in Figure 25. There is almost no difference for the scenario when the spatial effect is the biggest. For the case when the space-time effect is bigger, the better results where when considering the interaction type 1 and type 2.

The space-time interaction error is shown in Figure 26. It was smaller when the fitted model considers the type 4 space-time interaction in each scenario. The bigger differences were for the scenario when this effect was bigger and smaller differences for the scenario when the three effects were simulated with the same marginal precision.

The results for the posterior mean for the model hyperparameters are similar to the first case. We have it for τ_v are shown in Figure 27, for ρ in Figure 28, for τ_s in Figure 29, for ϕ in Figure 30, for τ_d in Figure 31.

These results are similar than the ones for the firs case. We also have the parameters for the main effects affected when considering the small diagonal add approach. Also, these are a bit off when the contrast parametrization was considered.

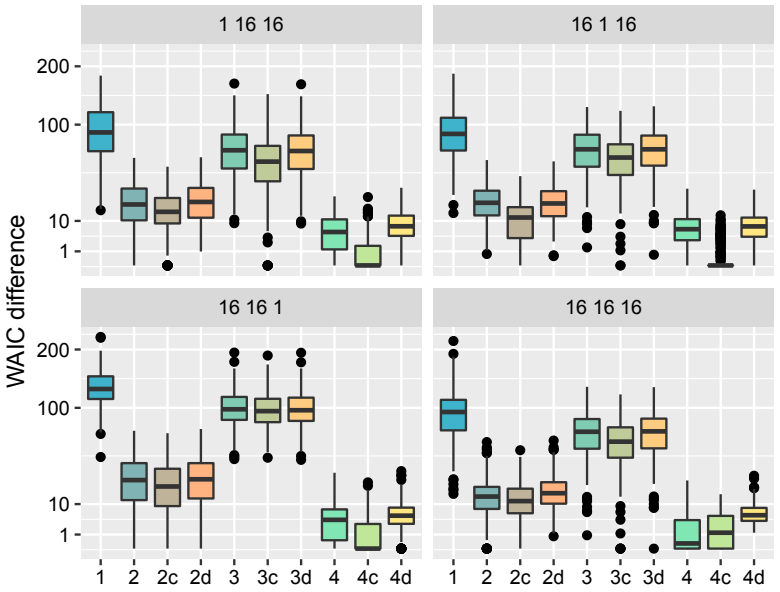


Figure 22: WAIC differences considering for each scenario in the second case.

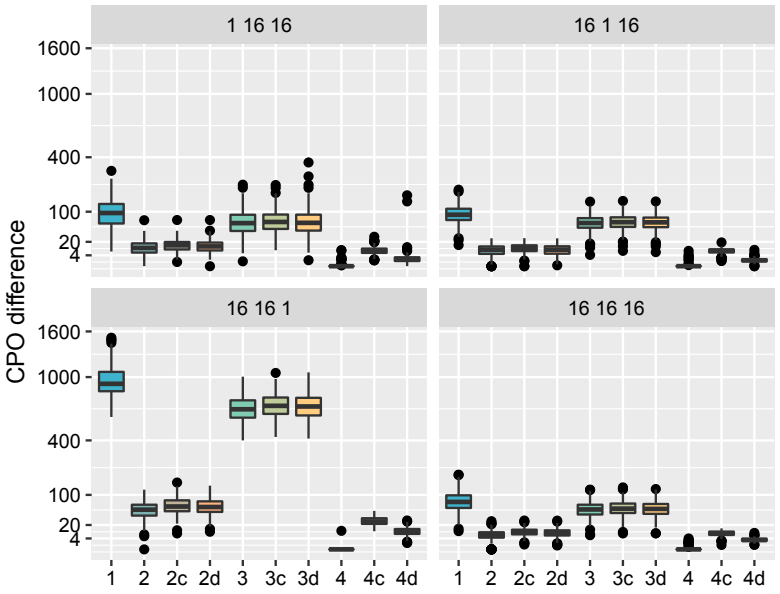


Figure 23: CPO differences considering for each scenario in the second case.

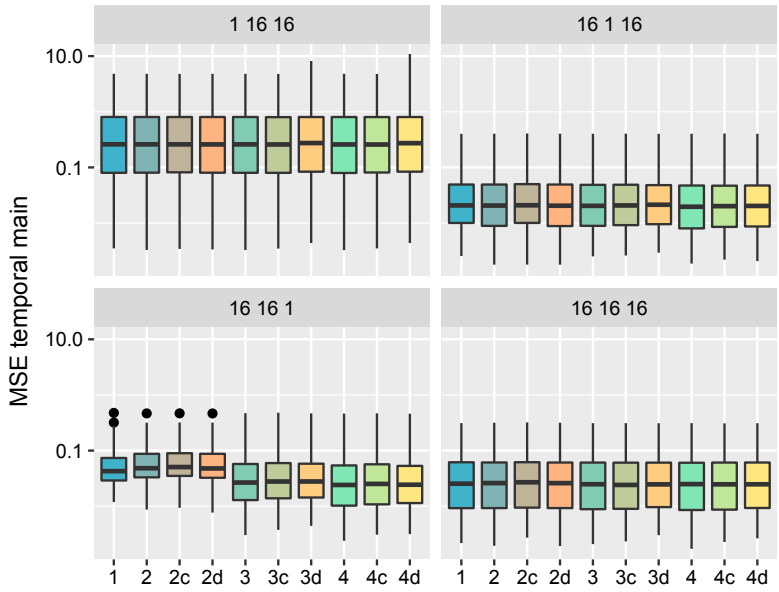


Figure 24: Mean squared error for the main temporal random effect in the second case.

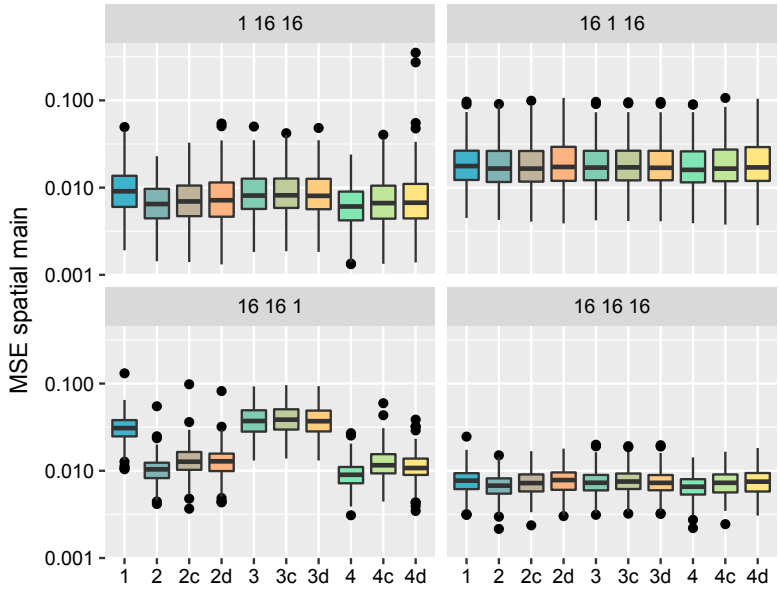


Figure 25: Mean squared error for the main spatial random effect in the second case.

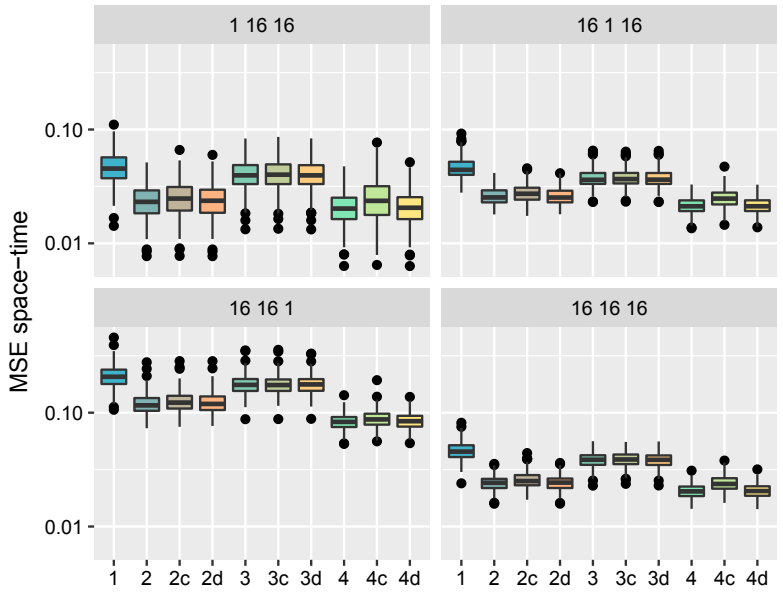


Figure 26: Mean squared error for the space-time random effect in the second case.

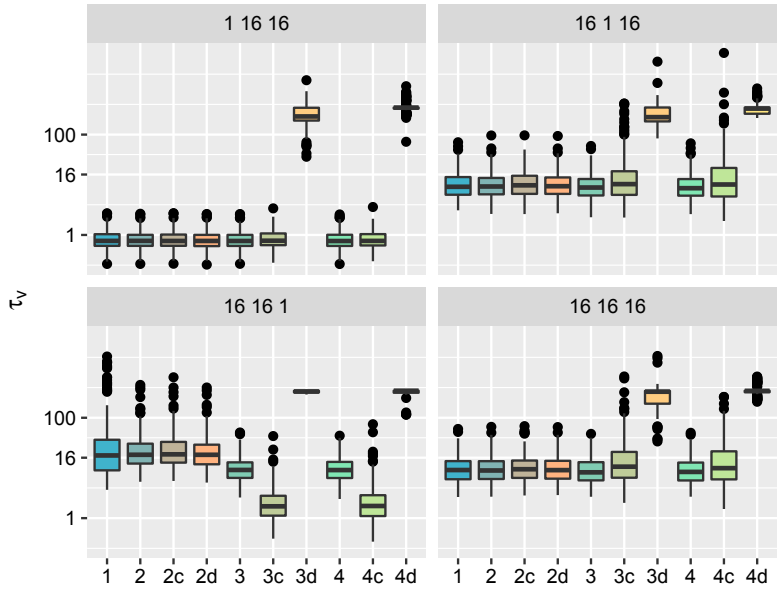


Figure 27: Posterior mean for τ_v for the second case.

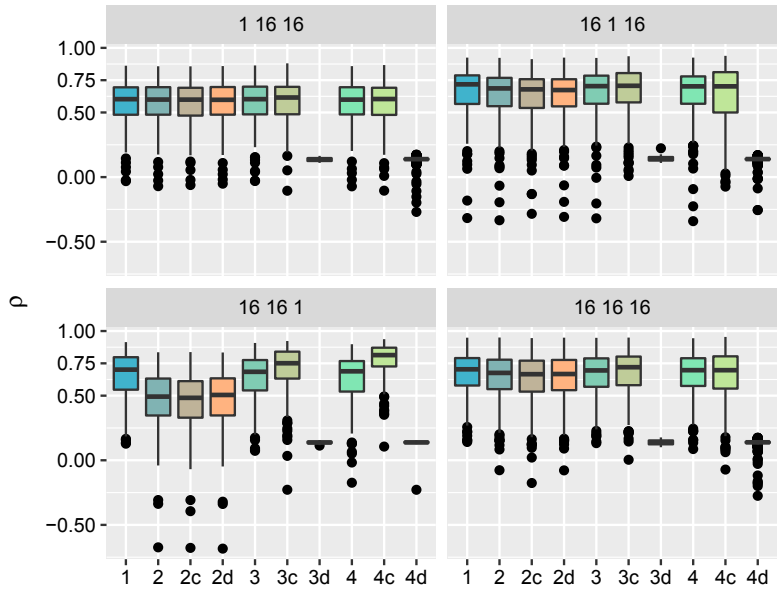


Figure 28: Posterior mean for ρ for the second case.

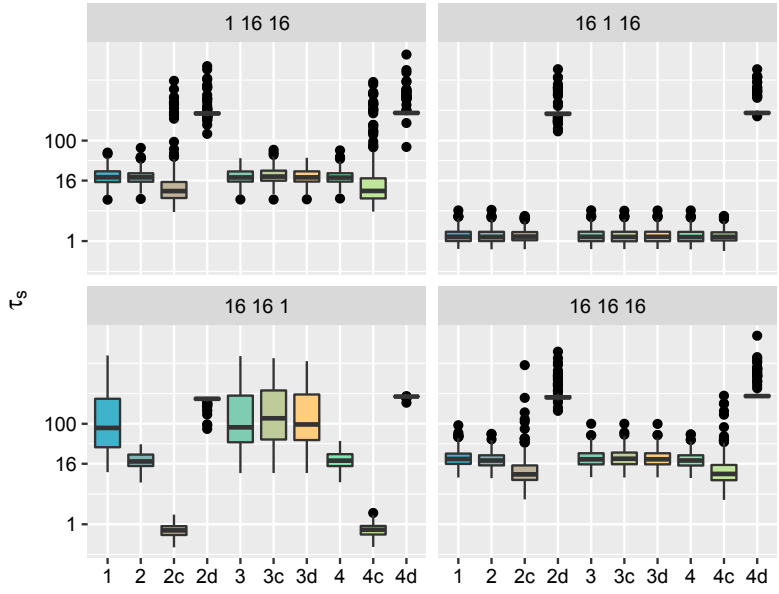


Figure 29: Posterior mean for τ_s for the second case.

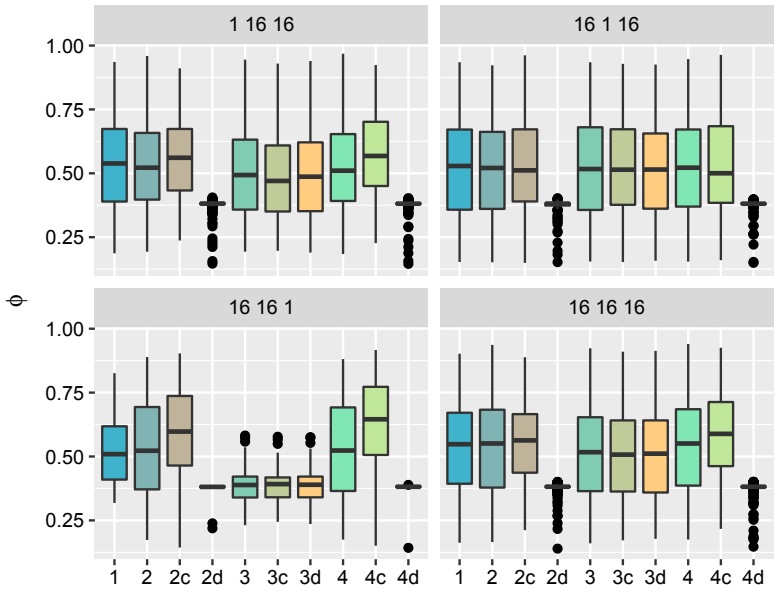


Figure 30: Posterior mean for ϕ for the second case.

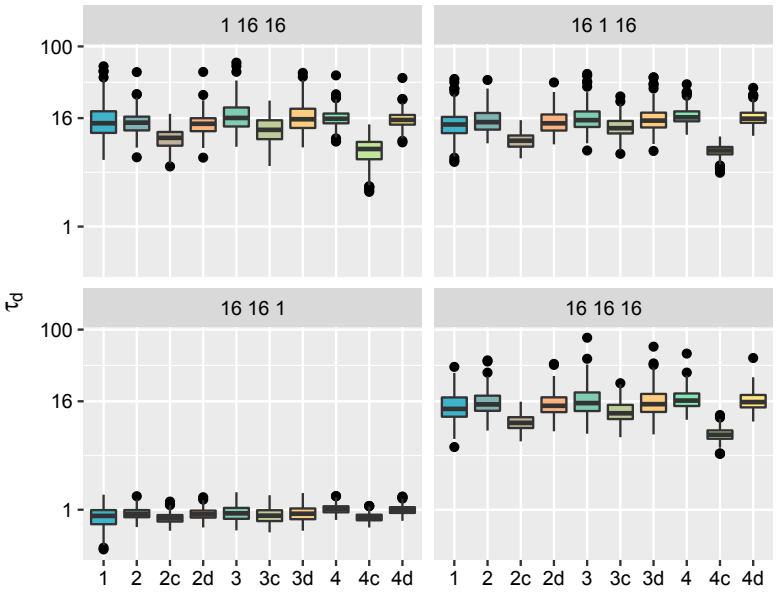


Figure 31: Posterior mean for τ_d for the second case.

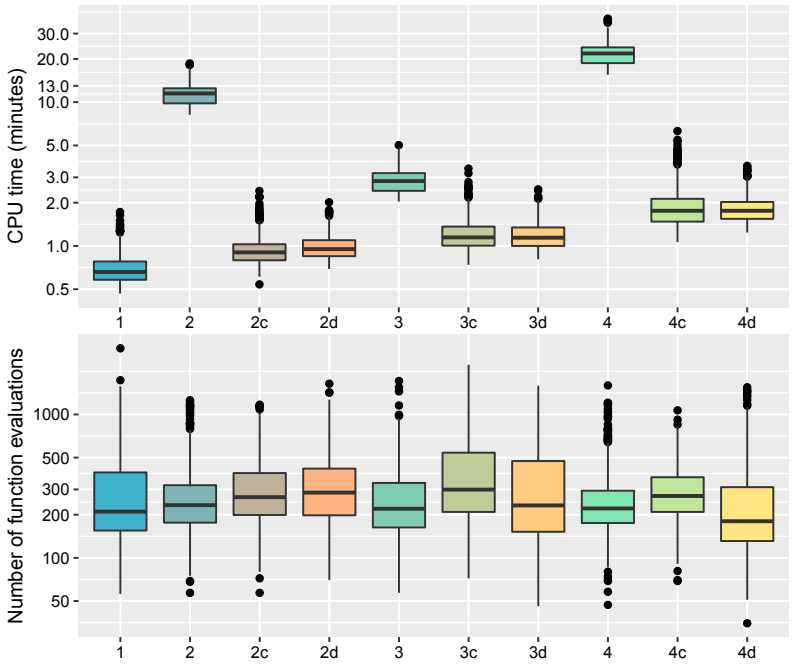


Figure 32: CPU time and number of function evaluations for the third case.

8.2.3 Simulation results for the third case

In Figure 32 we have the computing time in minutes. The type 1 takes less time to fit while the type 4 is the most expensive to fit within each approach as it has a precision structure less sparse. The proposed approaches, the contrast parametrization and the diagonal add approaches were twice as fast for type 3. For type 2 and 4 the CPU saving is very expressive in this scenario. For the original sum-to-zero constraints it took around 10 to 14 minutes for type 2 and around 20 or more for type 4 while considering the proposed approaches it took less than 2 minutes. The number of function evaluations were near the same for all the cases.

Considering the DIC, WAIC and CPO, shown in Figures 33, 34 and 35, respectively. We have that the smaller values were always for type 4, and this should be the case as the datasets were simulated considering this model. Additionally, the bigger values were mainly for type 1. The good result was that both the contrast parametrization and add diagonal approaches delivered these statistics near to the sum-to-zero constraint approach. Even thou a difference near 10 for DIC and WAIC computed using both proposed approaches, one will be still choosing the right interaction type. Using the CPO this choice is even more clearly.

The mean squared error - MSE for the random effects is shown in Figure 36. The MSE for the main temporal effect is bigger for the case when it has bigger marginal variance compared to the other two effects, as expected. All the approaches within each structured interaction type delivers approximately the same MSE for the main spatial effect.

The MSE results for the third case is shown in Figure 37. The spatial

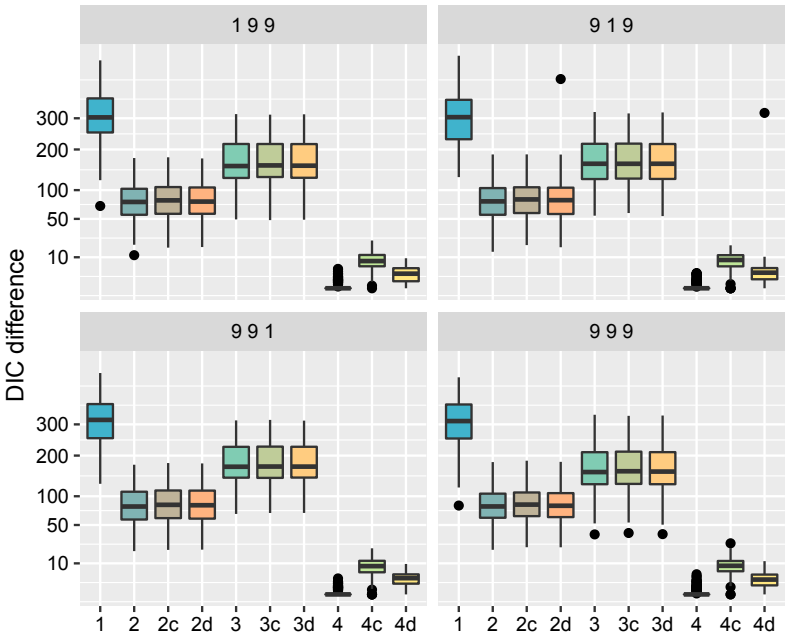


Figure 33: DIC differences for the third case.

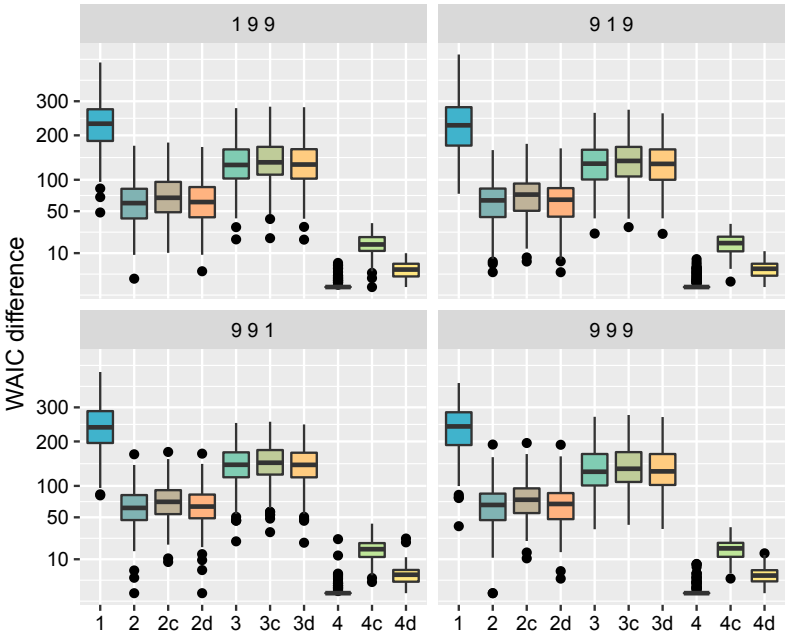


Figure 34: WAIC differences for the third case.

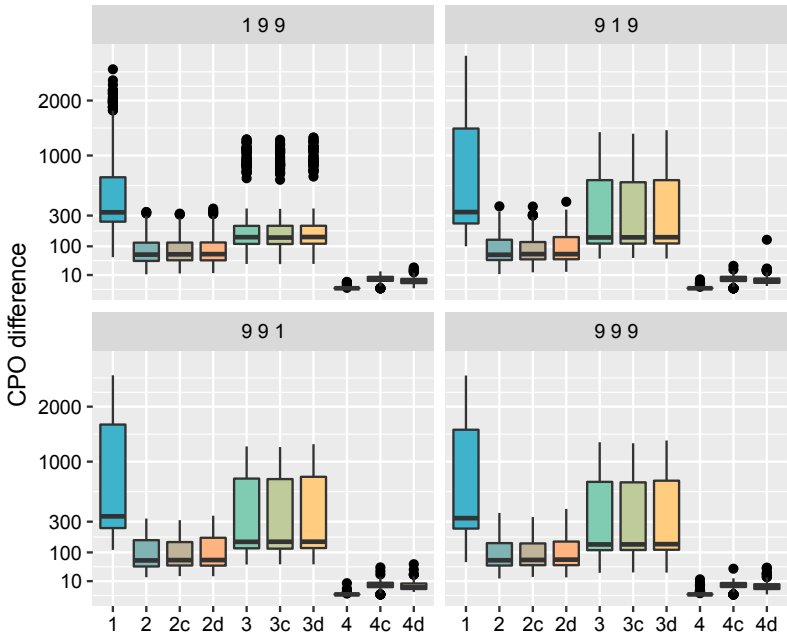


Figure 35: CPO differences for the third case.

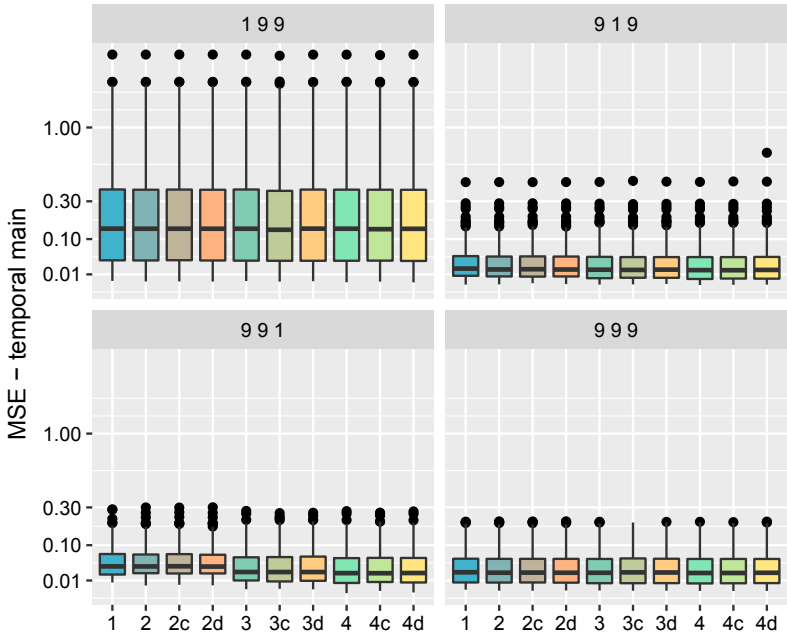


Figure 36: MSE for the main temporal effect in the third case.

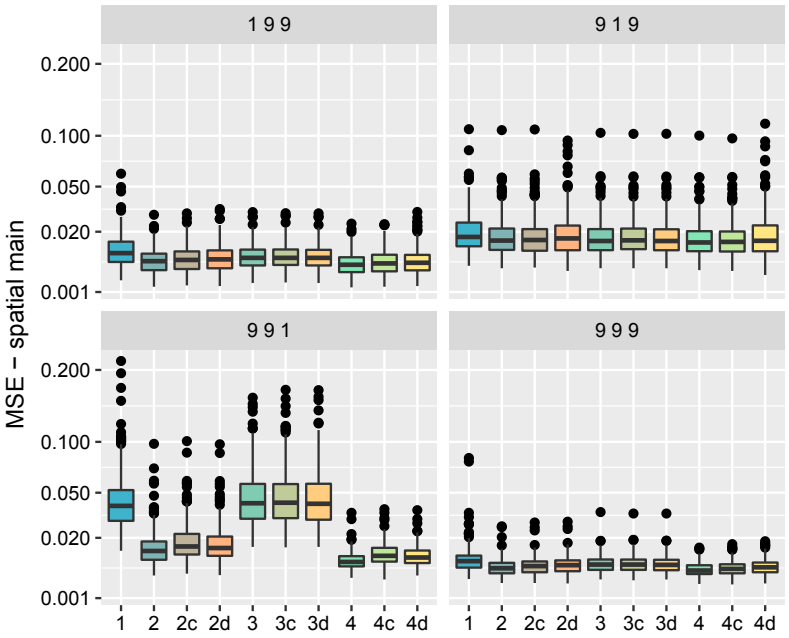


Figure 37: MSE for the main spatial effect in the third case.

main effect MSE was bigger for type 1 and 3 in the case where the space-time effect has the biggest marginal variance. For the other scenarios this difference was small.

The MSE results for the space-time interaction effect is shown in Figure 38. The smaller MSE values were when the fitted model was the type 4. This result is irrespectively to the approach being considered to fit. As expected, the bigger differences were for the case when this effect was the biggest, with $\tau_d = 1$. Comparing the different model fitting approaches within the space-time interaction structure, no expressive differences were found. Thus, the linear combinations considered were able to fit the original effect.

We have the posterior mean results for τ_v in Figure 39, for ρ in Figure 40, for τ_s in Figure 41, for ϕ in Figure 42 and for τ_d in Figure 43. As in the other two cases, the posterior mean τ_d was near the value considered to simulate for all the models fitted and approaches, 9 in the first two scenarios, 1 in the third and 9 in the last scenario. For the other parameters, the results are similar to the other two cases. An improvement here is that it was more closer to the value used to simulate. Even for the ϕ parameter which showed posterior mean results smaller in the two other cases and in this third case, except when the space-time type is 1 or 3 and the diagonal approach is considered, it is more closer to 0.75.

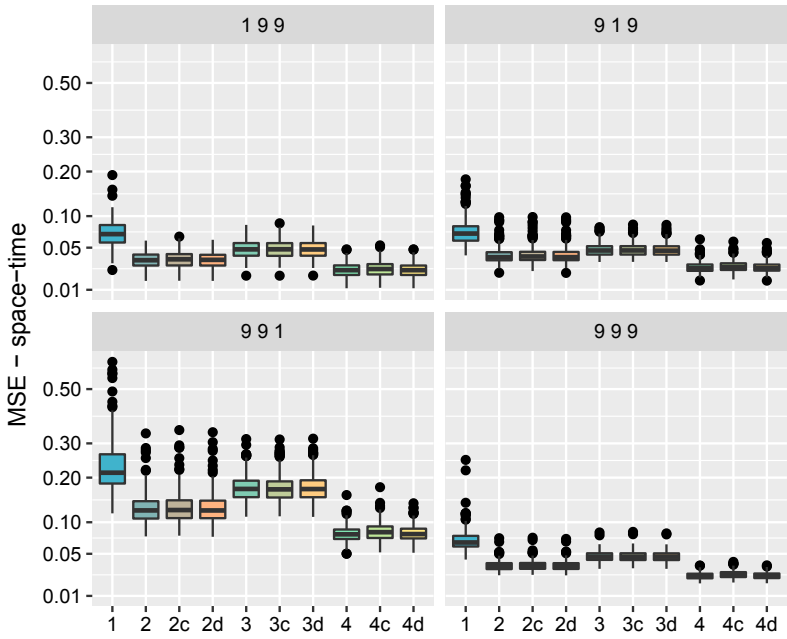


Figure 38: MSE for the space-time effect in the third case.

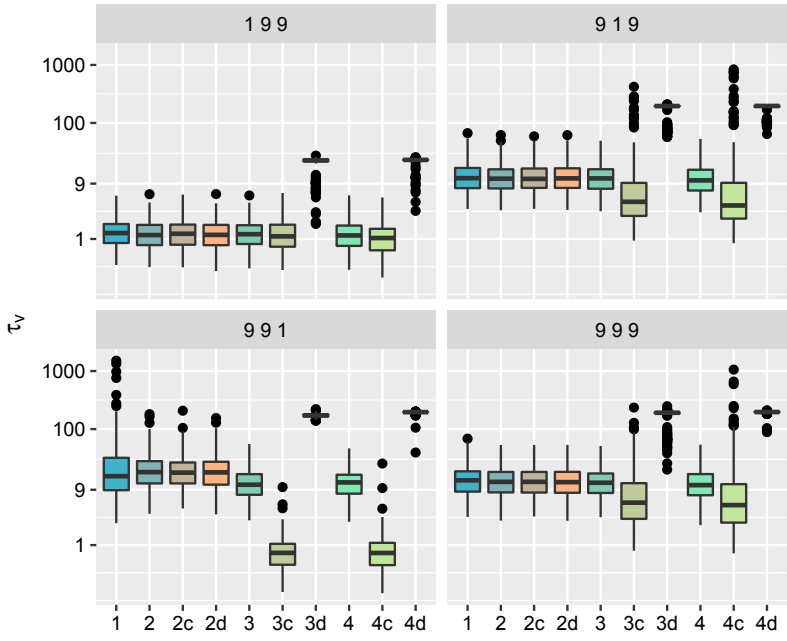


Figure 39: Posterior mean results for τ_v in the third case.

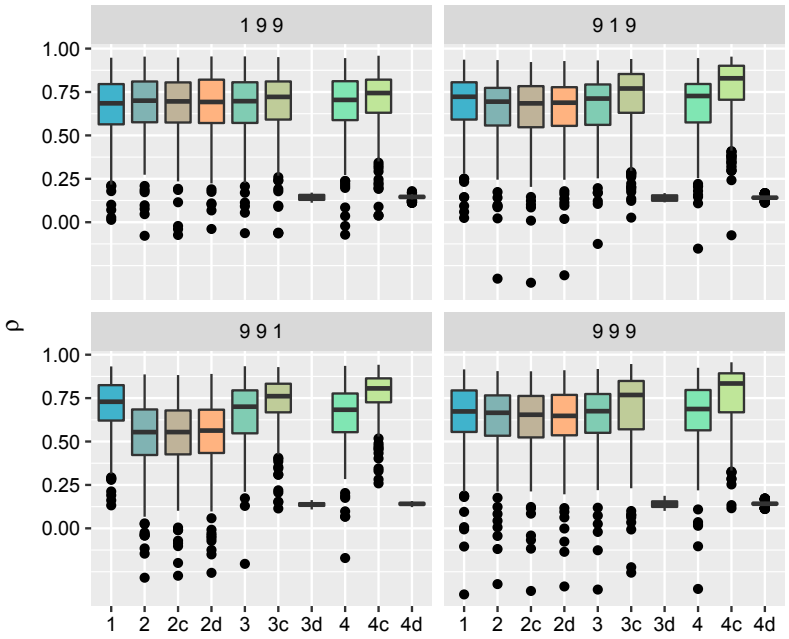


Figure 40: Posterior mean results for ρ in the third case.

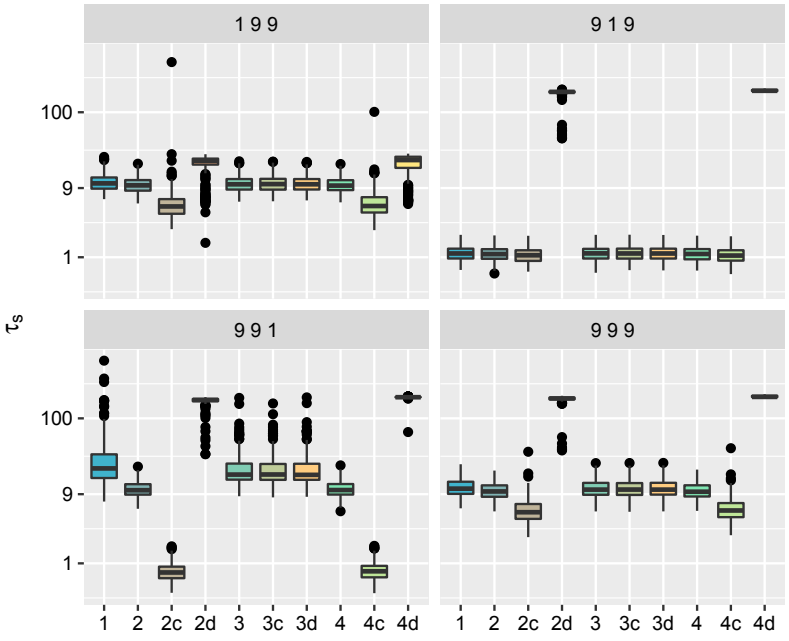


Figure 41: Posterior mean results for τ_s in the third case.

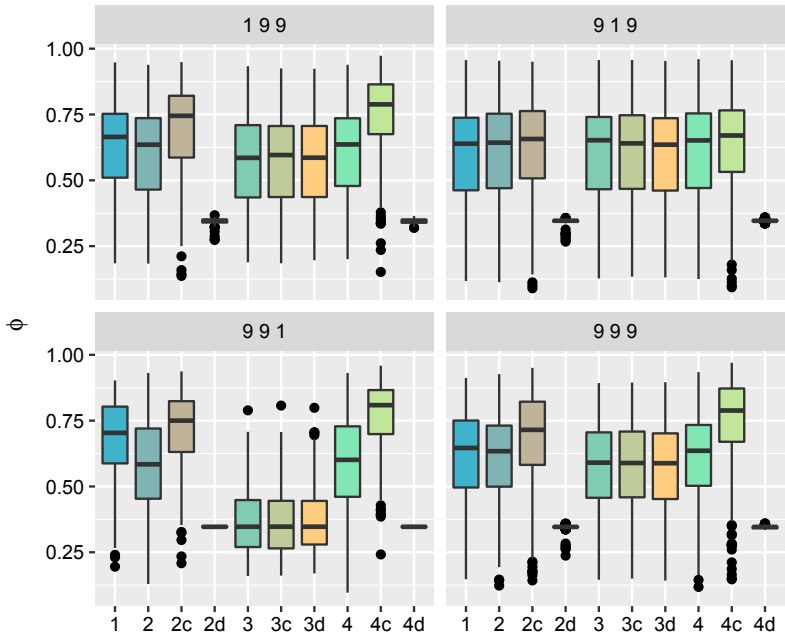


Figure 42: Posterior mean results for ϕ in the third case.

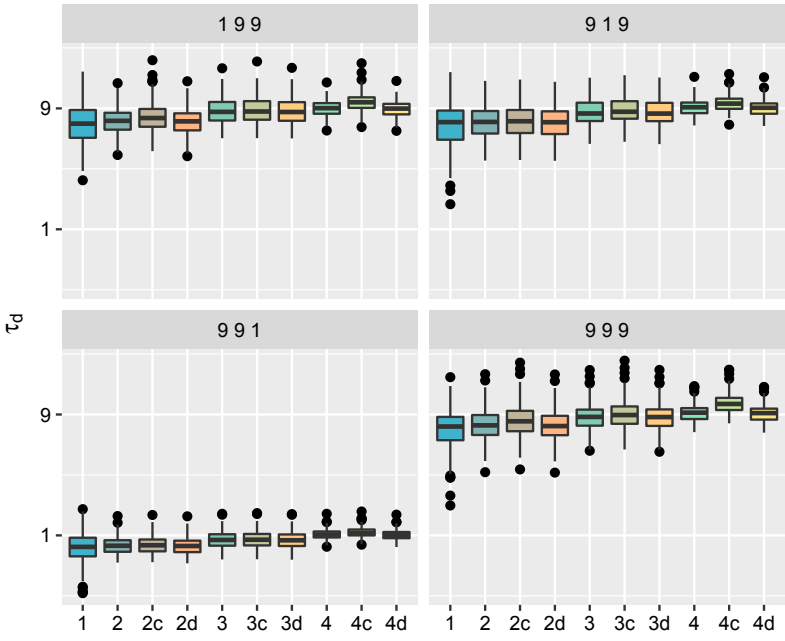


Figure 43: Posterior mean results for τ_d in the third case.

Part III

Application papers

Chapter 6

Application papers

In this chapter we introduce each of the applications of Bayesian hierarchical spatial and space-time models that I have been involved with. Seven of these projects have already been published in peer reviewed journals of the related area and one has been published in a conference preproceedings. This chapter ends with the published version of those projects introduced in Section 6.1.1 and in Section 6.3.3. The first one is in accordance with the authors rights policy stated in `http://www.nrcresearchpress.com/page/authors/information/rights`. The second is published in an open journal.

6.1 Space-time modelling of bycath fishery in Canada

The direct impacts of overfishing on target stocks are being increasingly addressed. However, unwanted bycatch and discarding of non-target species remains a key challenge of contemporary fisheries management. The following sub-sections describes the work that focused on analysing and identifying bycatch hotspots of a subclass of cartilaginous fish - elasmobranchs - which includes sharks, skates, rays and chimeras in the Canadian Northwest Atlantic waters and Arctic adjacent seas. The first subsection focuses on a deep-water shark species captured accidentally in Greenland halibut (Reinhardtius hippoglossoides) fishery in Canadian Arctic waters in the regions of the Baffin Bay and Davis Strait. The second subsection considers the bycatch of pelagic shark in the Northwest Atlantic Canadian pelagic longline fishery and finally, the third is a synthesis of all elasmobranch discards in the Canadian Northwest Atlantic and Arctic adjacent seas.

6.1.1 Bayesian spatiotemporal models to fisheries bycatch in the Canadian Artic

In this paper space-time models were applied to analyse the number of Greenland shark (Somniosus microcephalus) bycatch in Baffin Bay and Davis Strait located between the Northeast coast of Canada and Greenland. The data considered covered the period from 2008 to 2011. The fishing season is typically short due to the ice-cover and spans from July to November.

Therefore, an interesting point considered in the statistical analysis was to model the space-time effect over consecutive months and replicated over the years. One of the models considered in the analysis accounted for this effect. The space-time precision matrix was

$$I_4 \otimes R_m \otimes R_s$$

where Rs is the spatial precision matrix, R_m is the temporal structure for the correlation between consecutive month within each year and I_4 is a four dimensional identity matrix to define the replication for the four years of the within year space-time model.

In the analysis, we included several covariates, some with a linear effect and another with smoothed effect modeled with a random walk prior.

To select the best predictive model, different space time random effects and likelihoods - Poisson or negative binomial, and three different ways to consider zero-inflation were evaluated. All the different models were compared according to the deviance information criteria. A cross-validation test was also performed where the mean square error between the log of positive outcomes and the respective linear predictor predictive mean was compared [1]. This was computed considering 1100 observations randomly selected from the total of 1641 available.

The paper was published in the *Canadian Journal of Fisheries and Aquatic Sciences*, Cosandey-Godin et al. (2014) and is included in this chapter. Supplementary material including similar data and R code developed are available to the public on the *R-INLA* website and on the fist author's GitHub repository (`https://github.com/GodinA/cjfas-bycatch-INLA-SPDE`).

[1] This sentence feels like it ends early? Did you mean to add "was compared"

6.1.2 Predicting shark bycatch hotspots in a pelagic longline fishery

In this work, bycatch models were built similarly using at-sea observer data, but one additional dataset was introduced to estimate total bycatch rates. In Canada fishermen logbook data is collected in the pelagic longline fishery. It provides the most complete census of all fishing activities and contains information about the area fished, fishing effort, vessel and gear specifications and species landed, but does not record discards. The at-sea observer data from scientific observers fills this gap as it records all catch, including bycath that is discarded. However, the at-sea observer programs typically covers only 5% to 10% of the total fishing effort.

Hence, at-sea observer data was used to build a model accounting for covariates and the space-time variation. The logbook data was then considered as a prediction scenario and the predictions were done considering the fitted model. Models and predictions were analysed for three main pelagic shark bycatch in this Canadian fishery, namely porbeagle (*L*amma nasus), blue sharks (*P*rionace glauca) and shortfin mako (*I*surus oxyrinchus).

One important objective of this chapter was to indentify areas where bycatch risks were higher than average, i.e., identifying bycatch hotspots there is a based in the predictive posterior distribution. To achieve this, Bolin and Lindgren (2014) methodology was applied. At least three types of hotspots maps (excursion sets) can be generated:

1. individual hotspots map per time (year in our case);

2. one summary map showing combined hotspots over the whole time-period combining the largest areas where bycatch threshold was exceeded at least in one year,

3. one summary map showing only hotspots that were consistent over the whole time-period, i.e., largest area where the bycatch threshold was exceeded every time.

The results were shown to be important for dynamic management over space and time as time was also considered to define varying hotspots. This work was submitted for publication to *Conservation Biology* on November 14, 2014 but was since rejected and will be re-submitted with improvements.

6.1.3 Elasmobranch discards in the Northwest Atlantic and Arctic Adjacent seas: Composition and biogeography

The main objective of this chapter was again to identify bycatch hotspots and estimate total bycatch, in this case, specifically discards levels. However, the scope and complexity of the dataset was much greater considering that all fishing activities within the Canadian Northwest Atlantic and Arctic Adjacent Seas were considered. In total, now 55 fishing gear configurations, categorized into one of 11 major gear categories were included in the analysis.

From a conservation perspective, the results presented in this chapter provide the most comprehensive analysis of the distribution of elasmobranch discards and cumulative impacts across fisheries in the Canadian Northwest Atlantic waters and Arctic adjacent seas and included 17 years (1996-2012) of data. The emphasis was placed on deep-water species, which remain poorly studied. Twenty of the most frequently caught elasmobranch bycatch species were analysed in the chapter.

Therefore, the distribution pattern and dynamics differed considerably. To further facilitate the use of the space-time models efficiently (shorten the computation time) for this large geographical area, we aggregated the data considering Dirichlet Tessellation tiles.

From a statistical point of view, the main difference was that the data considered was not count data but weight and as such, two-part models were developed - one to model the occurrence and a subsequent model to model continuous positive discarded weight. Therefore the predictions were done considering samples drawn from the predictive distribution of both models. The statistical analysis included covariate information that were important to identify species specific hotspots and estimate cumulative discards.

6.2 Can collective memories shape fish distributions? A test, linking space-time occurrence models and population demographics

A particular interest in this paper was in the role of spatial memory in shaping distribution patterns. The dataset consists of 48724 occurrence/absence records of Atlantic herring *Clupea harengus* from logbook data of fishing activity around Iceland over 23 years.

Covariates reflecting local-scale environmental factors including temporally-lagged prey biomass and recent fishing activity, and through an index capturing distributional persistence over time, we derived two proxies for spatial memory of past wintering sites. Incorporating space-time cor-

relation structure and time-varying regression coefficients improved model performance. The capacity to predict distribution patterns one step ahead has important implications for the spatial management of herring stocks throughout the North Atlantic, and for other species exhibiting some homing tendency.

This paper was published in Ecography, Macdonald et al. (2017). An extensive supplementary material includes a similar dataset, the code used and some outputs for reproducing figures and tables.

6.3 Spatial modelling of log-age survival rate in Europe

In this section we introduce the work done for analysing old age survival in Europe. The first analysis considered data from 18 European countries and a spatial model was derived for each gender and for two time periods to model the *logit* of the probability parameter in the clustered/clumped Binomial model and excursions maps were produced to find areas with low or hight rates. The second one consider five countries and it was considered a smoothed covariate effect for each country and gender. The third only considered municipalities of Portugal and more covariates with smoothed interaction with gender. The fourth considered data from census areas in Porto and focused in model effect from neighbourhood on the old age survival.

6.3.1 Where do people live longer and shorter lives? An ecological study of old-age survival across 4404 small areas from 18 European countries

The main purpose in this work was to estimate the 10-year survival rate to express the population aged 75-84 years old who reached 85-94 years of age, considering census data from years 1991, 2001 and 2011 at 4404 small geographical areas in 18 European countries.

The statistical analysis to model the survival rate for women and men jointly consider a spatial effect for each of the two different 10 years periods, 1991 to 2001 and 2001 to 2011. In France, the decennial censuses took place in 1990 and 1999 so there is an interval of 9 years in the first period. Because of that difference, we used the clustered/clumped Binomial model, Cbinomial(m,n,p). To use this model, we have to consider the deaths instead survival counts. This model is defined as follows. Suppose that 'z' represents the death probability of a person and it is Binomial(n,p) distributed, where

'p' is the death probability of that person after one year and 'n' is the number of years of the period under analysis. Then, to define the variable 'u', we assume 0 when z=0 and 1 if $z > 0$. It gives $\mathrm{P}(u = 0) = (1 - p)^n$ and $\mathrm{Prob}(u = 1) = 1 - (1 - p)^n$. It happens that for each area we have 'm' persons in the beginning of the period. So, let $y = u_1 + \ldots + u_k$ the number of persons who died at the end of the period of 'n' years. This variable, y, is Cbinomial(m,n,p). In our study, we have $n = 9$ for France in the first period under analysis and $n = 10$ for all other countries and periods.

There were tested 80 different structure models. The basic model was just a Generalized Linear Model, as we have modeled the rate considering gender and year only. The other models were built taking into account the spatial structure,with and without interactions with gender and year. Moreover, an independent (unstructured) random effect was considered, and we allowed it to vary by area 'i', by area and gender 'ki', by year and area 'it', and by area, gender and year 'kit'. So, we created five scenarios to evaluate the unstructured random effect (URE) 'u': one without it and four different variations units for the URE. Considering the spatially structured random effect (SSRE) 's', we considered the scenario without SSRE and fifteen scenarios for having any kind of SSRE. By combining these $5x16$ scenarios we ended up fitting 80 models to our data.

The SSRE models we evaluated consider different interaction structures from a basic model and the four groups formed by gender and year. The basic model is just the Besag's model varying over areas 'i', considering the graph defined by the neighbourhood structure from the map. So, when considering this model and the URE by area 'i' we have the well know BYM model, see Besag et al. (1991), and the linear predictor considering gender, year, u_i and s_i can be written as

$$\eta_{kit} = \beta_0 + \beta_1 \mathrm{gender}_{kit} + \beta_2 \mathrm{year}_{kit} + s_i + u_i \ .$$

The model with minor DIC was the one with SSRE effect correlated across gender (the same for each year) and URE for each gender, area and year.

We considered the excursions method, see Bolin and Lindgren (2014), to find areas with significant hight and low survival old-age rate. The areas with low or hight survival rate for each gender were determined and showed into maps.

This paper was published in the *Journal of Epidemiology Community Health*, Ribeiro et al. (2016).

6.3.2 The association between socioeconomic deprivation and old-age survival in European small areas - a cross-national analysis

In this work it was considered part of the data from the paper in previous section in order to account for the effect from the deprivation index available for the set of 1911 small areas of five European countries. It was considered the 2001 and 2011 census data.

In the statistical model the outcome was the 85-94 years old population in 2011 at area i, y_i. They were considered the survivors and the 75-84 years old population in 2001 at area i, n_i were considered as the number of trials in the Binomial likelihood. Therefore we have p_i as the survival rate which was modeled considering

$$logit(p_i) = \eta_i = \beta x_i + f(x_i) + s_i$$

where s_i is a spatial random effect considered as

$$s_i = \tau(\sqrt{\phi}u_i + \sqrt{1-\phi}v_i$$

as described in Riebler et al. (2016), βx_i represent the linear effect from gender and country and $f(x_i)$ represent the non-linear/smooth deprivation index effect which depends on country and gender. Therefore we have $f(x_i) = e_{i,k,l}$, where $j = \{1, 2\}$ denotes gender, k =country and $e = \{1, ..., 10\}$ denotes deprivation class.

The deprivation effect smoothed effect model consider a random walk prior over the 10 deprivation classes replicated for each country and gender. Therefore it can be represented as one curve over deprivation for each country and gender combination: 10 curves in total. This accounts for interaction of the deprivation effect with country and gender. The posterior mean and credibility intervals showed interesting interaction effects, as shown in Figure 6.3.2. These interaction effects were related to the welfare regimes of each country.

The paper is under review.

6.3.3 Socioeconomic deprivation, healthcare access and environment on old-age survival in Portugal

In this work it was considered a subset of the data in the work on the previous section, those from 228 municipalities in Portugal. Two additional covariates were considered in addition to the to the deprivation index. A healthcare access index based on availability of long-term care and social

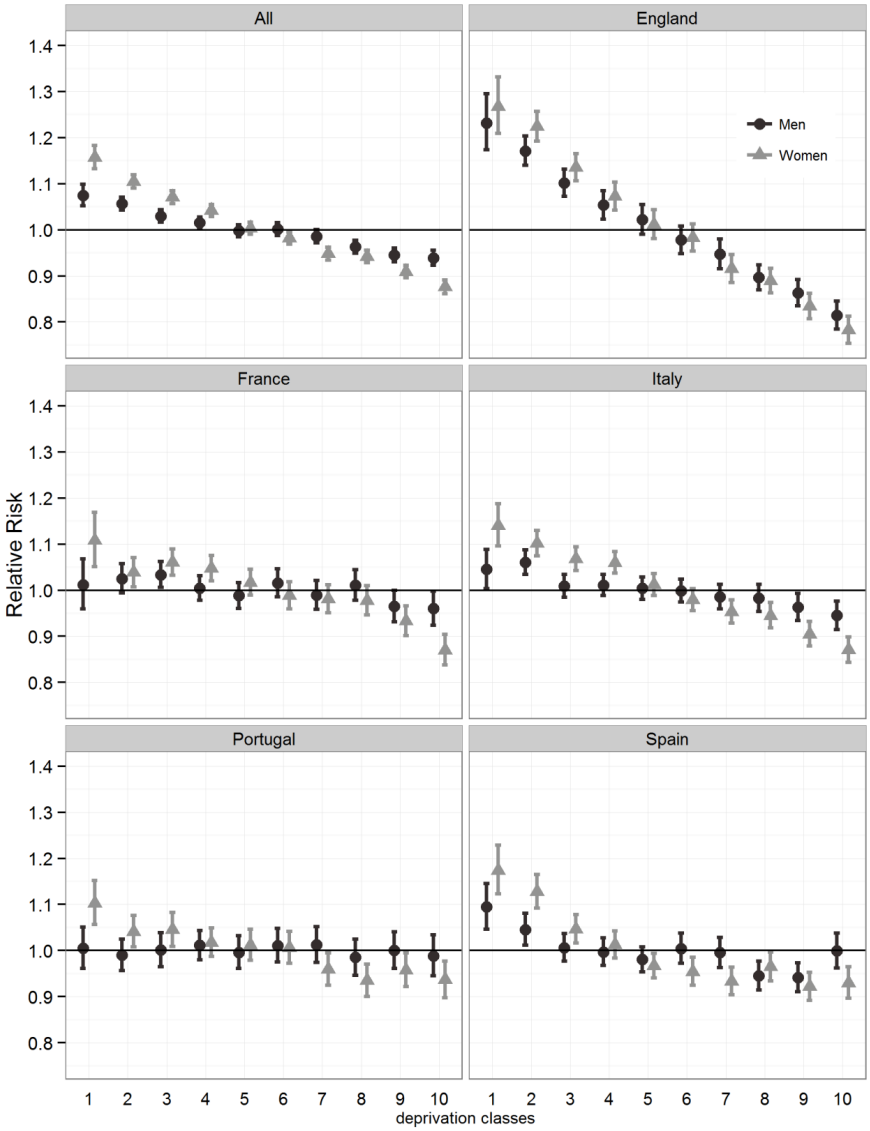


Figure 6.1: Interaction effect between deprivation, gender and Country.

support facilities, geographical access to healthcare facilities, and availability of health professionals.

The statistical model consider a smoothed effect for the three covariates by means of a random walk prior on the 10 classes for each of the three covariates replicated for gender to account for the interaction with gender. The spatial random effect were considered as in the previous work, as described in Riebler et al. (2016).

This paper was published in the *Geospatial Health* an Open Jornal and we include it at the end of this chapter. The URL for this paper is `http://geospatialhealth.net/index.php/gh/article/view/581/0`.

6.3.4 The influence of socioeconomic, biogeophysical and built environment on old-age survival in a Southern European city

In this work a fine spatial resolution analysis considering 109 aggregated census block groups in the city of Porto. The outcome were the population aged 85-94 years old in 2011 and the population aged 75-94 years old in 2001. The socioeconomic deprivation index was considered along with neighbourhood characteristics by means of a physical environment index and a walkability index. These physical environmental is based in air pollution variables, climate and greenspace availability. The walkability index consider residential density, street connectivity and land use mix.

For each of the three covariates it was considered smoothed a effect by means of a random walk prior replicated for gender to account interaction effects between gender and each covariate. The spatial random effect were considered as in the previous work, as described in Riebler et al. (2016). The excursion method was also considered to find areas with low or hight old-age survival rate.

6.4 Tuberculosis inequalities and socio-economic deprivation in Portugal

The objective of this paper was to investigate Tuberculosis (TB) spatial distribution and the association to socio-economic deprivation. We analysed all TB cases notified in Portugal from 2010 to 2014, as counts by municipalities, disaggregated by gender and quinquennial age groups.

We did not find a statistically significant association between TB notifications and the economic deprivation index. However, it was found for

components of this index named as "percentage of manual workers" and "percentage unemployed". It was defined 20 cases per 100 000 as the threshold to determine high-/low-risk municipalities. A significance level of 0.05 was set up when applying the excursions methodology to find the high- and low-risk areas.

The paper was published in the *International Journal of Tuberculosis and Lung Disease*, Apolinário et al. (2017).

6.5 Assessing comorbidity and correlates of wasting and stunting among children in Somalia using cross-sectional household surveys: 2007 to 2010

This work consider data collected at individual level and spatially referenced by the survey clusters in Somalia over years 2007 to 2010. A total of 73778 children aged 6-59 months from 1066 survey clusters were considered. The main objective was to analyse factor to explain wasting, stunting and underweight.

The proposed model for the the *logit* of the probability of each outcome k, $k = \{1, 2, 3\}$, in each survey cluster i. It can be represented as follows

$$\text{logit}(p_{k,i}) \quad = \quad \alpha_{k,i} + s_{k,i}$$

where $\alpha_{i,k}$ represents the fixed effect part and $s_{k,i}$ represents the random effect part. The fixed part consider effects from covariates to each outcome as particular to each one while the spatial effect has common components described next.

The spatial effect for each outcome is based on a set of three spatial random field in a way that each one is present in the linear predictor for two outcomes. It can be represented as

$$\begin{aligned} s_{1,i} &= u_{1,i} + \beta_{2,1} u_{2,i} \\ s_{2,i} &= u_{2,i} + \beta_{3,2} u_{3,i} \\ s_{3,i} &= u_{3,i} + \beta_{1,3} u_{1,i} \ . \end{aligned}$$

where u_1 is the shared component for wasting and stunting; u_2 is the shared component for stunting and underweight; u_3 is the shared component for wasting and underweight. The $\beta 2, 1$, $\beta 3, 2$ and $\beta 1, 3$ parameters measures the sharing spatial effect between wasting and stunting, between stunting and underweight and between wasting and underweight, respectively.

The paper is published in the BMJ Open journal, Kinyoki et al. (2016), and is free available at the url `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4785320/`.

6.6 Relation between Tweets and violent crime

In this work crime data and Tweets were collected over two separate time periods of 68 and 108 days within the city of Amsterdam. Tweets containing three kind of key words were considered. The first related to physical

violence, the second with aggressive moods states and the third related to substance abuse.

For each kind of key word Tweets it was considered as a realization of a log-Gaussian Cox point process, considering the likelihood approximation approach proposed by Simpson et al. (2015) and extending it for the space-time case. The linear predictor for each one was considered as

$$\eta(s,t)_k = \alpha_k + \epsilon(s,t)_k$$

for $k = 1, 2, 3$, α_k an intercept for the key work type k, $\epsilon(s,t)_k$ being the k-replica of a space-time continuous Gaussian process modeled assuming a prior distribution with spatially marginal Matérn covariance structure and marginally temporal first order autoregressive.

The linear predictor for the crime data was modeled as

$$\eta(s,t)_0 = \alpha_0 + \epsilon(s,t)_0 + \beta_k \epsilon(s,t)_k$$

were the α_0 is the intercept, $\epsilon(s,t)_0$ is a space-time Gaussian random field particular for violent crimes and β_k is a coefficient to capture the association to each kind of Tweet. Thus, β_k different from zero indicates association. The posterior distributions for β_1 and β_2 were off zero in the two periods while all the three were off zero for the first period.

This work was published in the Preproceedings of the 29th Benelux Conference on Artificial Inteligence - BNAIC 2017, Stam et al. (2017).

NRC Research Press

ARTICLE

Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic

Aurelie Cosandey-Godin, Elias Teixeira Krainski, Boris Worm, and Joanna Mills Flemming

Abstract: Understanding and reducing the incidence of accidental bycatch, particularly for vulnerable species such as sharks, is a major challenge for contemporary fisheries management. Here we establish integrated nested Laplace approximations (INLA) and stochastic partial differential equations (SPDE) as two powerful tools for modelling patterns of bycatch through time and space. These novel, computationally fast approaches are applied to fit zero-inflated hierarchical spatiotemporal models to Greenland shark (*Somniosus microcephalus*) bycatch data from the Baffin Bay Greenland halibut (*Reinhardtius hippoglossoides*) gillnet fishery. Results indicate that Greenland shark bycatch is clustered in space and time, varies significantly from year to year, and there are both tractable factors (number of gillnet panels, total Greenland halibut catch) and physical features (bathymetry) leading to the high incidence of Greenland shark bycatch. Bycatch risk could be reduced by limiting access to spatiotemporal hotspots or by establishing a maximum number of panels per haul. Our method explicitly models the spatiotemporal correlation structure inherent in bycatch data at a very reasonable computational cost, such that the forecasting of bycatch patterns and simulating conservation strategies becomes more accessible.

Résumé : Comprendre et minimiser les répercussions relatives aux prises accessoires accidentelles, et plus particulièrement celles touchant les espèces vulnérables, tel le requin, constituent de nos jours un défi de taille en gestion halieutique. Par la présente recherche, nous soumettons que les approximations intégrées et agglomérées de Laplace (INLA) ainsi que les équations différentielles partielles stochastiques (EDPS), représentent de puissants outils de modélisation de la répartition des prises accessoires dans l'espace et le temps. Ces nouvelles approches informatiques, efficaces et innovantes sont ici utilisées afin de reproduire des modèles spatiotemporels hiérarchiques à inflation zéro en les appliquant aux données relatives aux prises accessoires du requin du Groenland (Somniosus microcephalus) dans la pêche au filet maillant du flétan noir (Reinhardtius hippoglossoides) de la baie de Baffin. Les résultats obtenus démontrent que les prises accessoires du requin du Groenland sont dispersées dans l'espace et le temps et varie de manière considérable selon les années. Certains facteurs quantifiables, tels le nombre de panneaux de filet par trait et la quantité de prises totale de flétans noirs, ainsi que certaines caractéristiques physiques, telle la bathymétrie, influent sur la fréquence des prises accessoires du requin du Groenland. Conséquemment, les probabilités de prises accessoires pourraient être réduites en limitant l'accès de pêche dans les endroits spatiotemporels à haut risque ou encore, en établissant une quantité maximum de panneaux de filet par trait. Ces méthodes modélisent d'une façon nette et précise la structure des corrélations spatiotemporelles inhérentes aux données sur les prises accessoires, et ce, à un coût informatique des plus raisonnables, rendant désormais possible et accessible le calcul des prévisions quant au nombre de prises accessoires ainsi que la planification de stratégies efficaces en conservation des espèces.

Introduction

High levels of bycatch are a predominant problem in many fisheries around the world and contribute greatly to broader concern about overfishing (Kelleher 2005; Davies et al. 2009). Bycatch commonly refers to the part of the catch that is not a legal target of the fishery; it may be retained and landed but is often discarded (dead or alive) (FAO 2011). It tends to be particularly problematic for long-lived marine megafauna such as sharks, marine turtles, seabirds, and marine mammals (Lewison et al. 2004; Hall et al. 2000). Under the ecosystem approach to fisheries management, a core objective is to reduce and eliminate bycatch (Pikitch et al. 2004; Garcia et al. 2003). One of the first steps in addressing bycatch issues is to identify and prioritize key conservation and management areas (Kirby and Ward 2014). These priority areas are often referred to as "hotspots" and are locations where bycatch patterns indicate abnormally high risk (e.g., Huang and Yeh 2011; Cambiè et al. 2012; Roe et al. 2014). Detailed information on bycatch patterns and their drivers can help establish effective spatial management, such as time–area closures and spatially explicit gear restrictions and catch quotas. These tools are increasingly used in marine resource management to better integrate multiple and often divergent economical and environmental objectives (Dunn et al. 2011; Douvere 2008). However, such regulations require some understanding of the spatiotemporal dynamics of the system.

Bycatch data are most often collected by at-sea observer programs and composed of the presence and absence (either count or mass) of nontarget species georeferenced by fishing location. They typically also contain information concerning the target species, vessel and gear specification, fishing effort, and environmental information like depth and sea surface temperature. Like many fishery datasets, bycatch data are characterized by complicated statistical features, such as excess of zeros, nonlinearity and nonconstant variance structure, and spatiotemporal correlation

Received 2 April 2014. Accepted 22 September 2014.

A. Cosandey-Godin. Department of Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada; WWF-Canada, Halifax, NS B3J 1P3, Canada.
E.T. Krainski. Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.
B. Worm. Department of Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada.
J.M. Flemming. Department of Mathematics and Statistics, Dalhousie University, Halifax, NS B3H 4R2, Canada.
Corresponding author: Aurelie Cosandey-Godin (e-mail: godina@dal.ca).

Published at www.nrcresearchpress.com/cjfas on 29 September 2014.

(Ciannelli et al. 2008). These characteristics violate the assumptions underlying basic statistical techniques such that more sophisticated models are required. Statistical tools that explicitly model the sources of zero observations (Martin et al. 2005) are commonly used in fisheries science and custom practice in bycatch studies (e.g., Minami et al. 2007; Barlow and Berkson 2012; Murray and Orphanides 2013).

To deal with nonlinearity and nonconstant variance, bycatch data are often modelled with well-established techniques used in catch-per-unit-effort (CPUE) standardization and stock assessments. These methods include generalized linear models (GLMs) (e.g., Megalofonou et al. 2009; Jannot and Holland 2013), generalized additive models (GAMs) (e.g., Minami et al. 2007; Murray and Orphanides 2013), and, to a lesser extent, generalized linear mixed models (GLMMs) (e.g., Trebilco et al. 2010) and generalized additive mixed models (GAMMs) (Bjorge et al. 2013). For a detailed discussion of these models in fisheries research, see Venables and Dichmont (2004).

Other techniques rely on matching animal telemetry and fleet distribution data to infer spatial overlap and derive bycatch predictions (e.g., Harden and Willard 2012; Roe et al. 2014; McClellan et al. 2009). These techniques offer advantages over fishery-dependent data but are limited to species with available tracking information. Bayesian hierarchical models have also been proposed for estimating bycatch probabilities (Gardner et al. 2008; Sims et al. 2008; Moore and Read 2008), but have not been widely adopted, likely because of high computational costs and complex estimation routines. However, these methods do represent very powerful approaches for dealing with complex ecological datasets with multiple sources of uncertainty (Cressie et al. 2009) and are readily used in other areas of fisheries science (e.g., Rivot et al. 2008; Yu and Leung 2010; Harley and Myers 2001).

Hierarchical models can simplify complex interactions by allowing parameters to vary at more than one level via the introduction of random effects. The expected value of the response is then expressed conditional on these random effects. Mixed models, such as GLMMs and GAMMs, are examples, the first being a fairly straightforward extension of linear regression (Venables and Dichmont 2004). Trebilco et al. (2010) modelled seabird bycatch in the eastern Australian tuna and billfish pelagic longline fishery using GLMMs with a random effect for each fishing vessel. Ortiz and Arocha (2004) standardized CPUE indices of billfish bycatch in the Venezuelan tuna longline fishery with all year interactions treated as random effects. Hierarchical approaches are very well-suited to nested data, such as bycatch data, where, for example, fishing sets are sampled from a trip, sampled from a vessel, which is part of a larger fleet. In this setting, errors associated with both the data and uncertainties about the ecological process are included, which results in more robust statistical inference (Cressie et al. 2009; Wikle 2003). The advantages of using hierarchical Bayesian models emerge more so as complexity increases, when, for example, spatiotemporal variability needs to be modelled explicitly (Cressie et al. 2009). The Bayesian framework also offers the advantage of providing full inference, such that model parameters and uncertainty can be quantified, which has great utility in applied conservation (Wade 2000; Wintle et al. 2003).

The ad hoc approach often taken in bycatch studies is to model space using a nonrandom factor variable, like a 5° × 5° grid cell (e.g., Brodziak and Walsh 2013) or geographical fishing boundaries (e.g., Bjorge et al. 2013; Barlow and Berkson 2012). Others commonly include one or two geographic coordinates (latitude and (or) longitude) in their models (e.g., Yeh et al. 2013; Jiménez et al. 2009; Orphanides 2010). Representing latitude and longitude as continuous variables offers the advantage that the data are not isolated in separate units, but these variables are still only incorporated into models as fixed effects and, as such, do not include spatial dependency. When fishing locations are georeferenced in space and the main research questions revolve around spatial predictions, the most appropriate statistical approaches are geostatistics-based models, which intrinsically incorporate the first law of geography: "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Hierarchical Bayesian models extend the concept of multilevel structure to include a spatial random effect (Gaussian random field, GRF). This random field is a stochastic process indexed in space that essentially represents all spatially explicit processes that may have an effect on the bycatch pattern. This is the real advantage of these models; they are built to approximate and include uncertainties with the entire bycatch phenomena as opposed to only uncertainty associated with discretely observed data. In so doing, bycatch hotspots cannot only be rigorously identified, they can also be better forecasted for management planning (Clark et al. 2001).

Hierarchical Bayesian models have traditionally relied on Markov chain Monte Carlo (MCMC) simulation techniques, which are computationally expensive and technically challenging, consequently limiting their use. However, a new statistical approach is now readily available, namely integrated nested Laplace approximations (INLA) via the R-INLA package (http://www.r-inla.org). INLA methodology and its powerful application to modelling complex datasets has recently been introduced to a wider nontechnical audience (Illian et al. 2013). As opposed to MCMC simulations, INLA uses an approximation for inference and hence avoids the intense computational demands, convergence, and mixing problems sometimes encountered by MCMC algorithms (Rue and Martino 2007). Moreover, included in R-INLA, the stochastic partial differential equations (SPDE) approach (Lindgren et al. 2011) is another statistical development that models GRFs much faster (similar to kriging) as well as constructs flexible fields that are better adept to handle datasets with complex spatial structure (Lindgren 2013). This is often the case with fisheries data, since fishermen tend to target particular fishing grounds, resulting in clustered spatial patterns and large regions without any values. Together, these new statistical methods and their implementation in R allows scientists to fit considerably faster and more reliably complex spatiotemporal models (Rue et al. 2009).

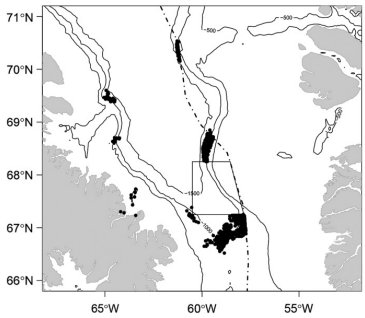
The aim of this paper is to analyze bycatch data using hierarchical Bayesian spatiotemporal models fitted using these two novel techniques. We present an analysis of Greenland shark (*Somniosus microcephalus*) bycatch in the Canadian gillnet fishery in Baffin Bay. We demonstrate how our approach can yield answers to the ubiquitous questions behind bycatch studies, that is, to (*i*) identify times and areas of higher bycatch risk (which may give insight into the species biology, i.e., suitable habitat); (*ii*) identify environmental and fisheries drivers affecting bycatch rates; and (*iii*) identify plausible mitigation measures. In this paper, we first fully describe the Greenland shark bycatch data, briefly discuss relevant aspects of INLA and SPDE, and specify all models that were investigated. We then address model selection, inference, and goodness of fit. Details on the final model are provided in Appendix A, and all R codes and datasets are available on the R-INLA website. In closing, we discuss future opportunities for the INLA framework in relation to bycatch studies.

Materials and methods

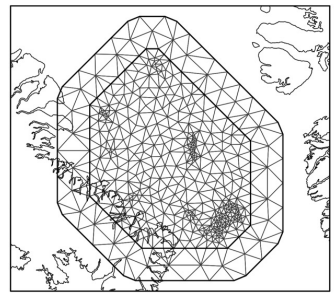
Greenland shark bycatch dataset

Baffin Bay and Davis Strait are two large basins between Nunavut's Baffin Island and Greenland that connect the Arctic and Atlantic oceans. These regions sustain the only large-scale commercial fisheries in Canada's Arctic. A Greenland halibut (*Reinhardtius hippoglossoides*) fishery in Baffin Bay began in 1996 as a small exploratory fishery but has been expanding greatly since 2001 (DFO 2014). Greenland sharks are commonly caught as bycatch and

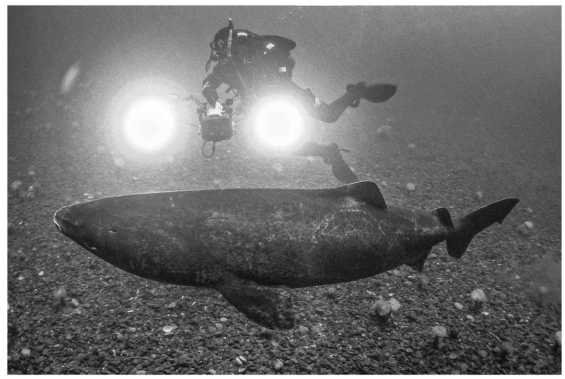
Fig. 1. (*a*) Spatial distribution of the observed Greenland halibut gillnet fishing hauls, location of the fishing closure (solid line), and Canadian Exclusive Economic Zone (dashed line); (*b*) mesh used to calculate the Gaussian (Markov) random field in the SPDE approach; and (*c*) image of a free-swimming Greenland shark.



(a) Map of Baffin Bay with isobaths (metres)



(b) Constrained refined Delaunay triangulation



(c) *Somniosus microcephalus* ©Jeffrey Gallant

discarded (MacNeil et al. 2012). Currently, estimates of their biomass, productivity, and fishing mortality are unknown, which limits the ability to predict fishing impact on the Greenland shark population. As a result, there is concern that the species could become overexploited because of presumed low productivity; hence, a precautionary approach to their management is advocated (Davis et al. 2013). Investigating bycatch data, one of the only available sources of information for the species, can provide insights into the shark's biology and help establish appropriate management efforts.

At-sea observers are assigned to monitor the Greenland halibut fishery (DFO 2014). Observers record location of the net (at the start of the haul), target species (*R. hippoglossoides*) and mass captured, number of gillnet panels, mesh size, soak duration, depth at which the net was set, and bycatch mass and count per haul. Count information has only recently (since 2008) been recorded for pelagic species, including Greenland sharks.

The observer dataset used for this case study represents 79.2% of the total fishing effort (number of hauls) of the Baffin Bay gillnet fleet over the 4-year period (2008–2011). Data were collected by a Newfoundland and Labrador-based observer company and obtained from Fisheries and Oceans Canada (DFO).

Gillnet vessels tend to be relatively small in size (65 feet; 1 foot = 0.304 m) and are allowed to carry a maximum of 500 nets per trip (DFO 2014). Fishermen usually set several gillnets a day with on average about 40–50 panels in each gillnet (DFO 2005). Since 2007, a large portion of the Greenland shelf 11 750 km in the southeast was closed to fishing for the protection of narwhal (*Monodon monoceros*) and deepsea corals (DFO 2007). Gillnet effort is concentrated in areas to the north and south of the closure (Fig. 1). The fishery has expanded over the years, for example, the northeast of Baffin Bay (71°N) is a new fishing ground where the Greenland shelf projects into Canadian waters and provides suitable depths for halibut fishing.

There were 1647 hauls from 26 trips, spanning 2008 to 2011, with all observations occurring between the months of July and November. Three hauls were discarded because there was no information on the mass of Greenland halibut captured; none of these hauls had shark bycatch. Another three hauls were inside the closure or outside the 200 nautical mile (1 n.mi. = 1.852 km) Canadian Exclusive Economic Zone (EEZ) and hence were also disregarded. In addition, 25 hauls had recorded mass of Greenland sharks but no associated counts. The majority of these (21 records) were less than 300 kg. For each of these entries, we assumed a count of one, which makes sense given the average mass of a shark in the others hauls (127 kg). Hereinafter, a total of 1641 hauls were included in the analysis (Table 1). We refer to each haul as a fishing geolocation, 147 (9%) of which recorded bycatch of Greenland shark; counts varied from 1 to 40 sharks per haul.

Fig. 2. Relationship between Greenland shark bycatch counts and covariates of interest. Results of fitting both a Poisson regression (solid line) and a generalized additive model (broken line) are shown.

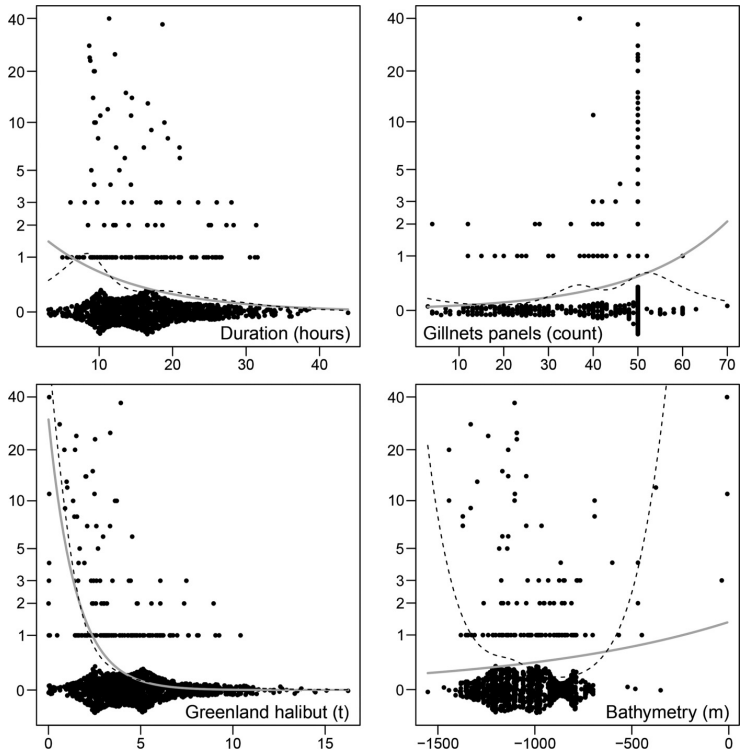


Table 1. Summary of the dataset showing the total number of hauls with bycatch (>0), number of observed hauls, mass (kg), and counts of Greenland shark bycatch per year.

Year	Hauls with bycatch	Hauls	Mass (kg)	Counts
2008	46	411	5 027	66
2009	34	378	8 765	83
2010	20	374	4 435	21
2011	47	478	92 425	400

In addition to the observed data, ocean depth was extracted from the National Oceanic and Atmospheric Administration (NOAA) global relief model ETOPO1 (Amante and Eakins 2014). Spatial data were projected from latitude–longitude locations on the surface of a sphere into locations on a plane using NAD83 (CSRS) – UTM zone 19N, a coordinate reference system suitable for use in Canada between 66°W and 72°W. The relationship between shark bycatch and four variables of potential interest (duration (soak time in hours and decimal minute), number of gillnet panels, total catch of Greenland halibut (in metric tonnes, t), and the bathymetry (metres, m)) were included in the models. Exploratory analysis revealed no clear relationships between Greenland shark bycatch and these covariates, with the exception of some evidence of a nonlinear relationship with bathymetry (Fig. 2).

Hierarchical spatiotemporal model structure

Similar to a GLM framework, the response, in our case the observed Greenland shark bycatch (count) at a particular fishing geolocation and time, is assumed to have a distribution that belongs to the exponential family, and the parameters of the family (ϕ) are linked to a structured additive predictor η through a link function $g(\cdot)$ such that $g(\phi) = \eta$, where the linear predictor η is defined, in our case, as follows:

$$\eta = \beta_0 + \beta_1 \text{Duration} + \beta_2 \text{Panels} + \beta_3 \text{TC.tspp} + \sum_{k=1}^{K} f_k(\text{Bathymetry}) \boldsymbol{w}_k + f(\cdot) \quad (1)$$

where β_0 is the intercept, and β_1, β_2, and β_3 are the (linear) regression coefficients for our covariates: duration or soak time in hours and decimal minutes (Duration), number of gillnet panels (Panels), and catch (t) of Greenland halibut (TC.tspp), respectively. The function f_k is the sum of smooth functions defining the random effect of bathymetry (m), where regression coefficients vary with bathymetry values (K values), and $\boldsymbol{w}_k$ is a vector of known bathymetry values defined for each of the fishing geolocations. This is equivalent to a smooth function used in GAMs. $f(\cdot)$ is a semiparametric function defining the spatiotemporal random effect included in the model (Table 2). All of these components form the (nonobservable) latent field defined as $\boldsymbol{\theta} = \{\beta_0, \boldsymbol{\beta}, \boldsymbol{f}\}$, where $\boldsymbol{\beta}$ and $\boldsymbol{f}$ are the covariates and smooth functions, respectively, included in the linear predictor with their appropriate priors (ψ).

The latent field is characterized by a joint normal (Gaussian) multivariate distribution with mean $\mathbf{0}$ and precision matrix $\mathbf{Q}(\psi)$

Table 2. Spatiotemporal correlation structures considered.

Identifier	Spatiotemporal structure
m0	None
m1	Constant: $f(\mathbf{s}, t) = f(s)$ is a Matérn correlation structure with $\nu = 1$, where scale and variance need to be estimated
m2	Different each year: $f(\mathbf{s}, t)$ is a Matérn correlation structure like model m1 but with a different realization every year
m3a	Correlated in consecutive years: $f(\mathbf{s}, t)$ is a combination of model m1 with additional correlation structure between neighbouring years
m3b	Exchangeable correlation between years: $f(\mathbf{s}, t)$ is similar to model m3a but correlation structure between years is repeated (i.e., the correlation between years 1 and 2 is the same as that between years 1 and 4)
m4	Different for each month: similar to model m2 but the time index is month rather than year
m5a	Correlated in consecutive months of the same year: same structure to model m3a for month
m5b	Exchangeable correlation between months of the same year (e.g., 2009–2007 has the same correlation with 2009–2008 as with 2009–2009)

(i.e., $\boldsymbol{\theta} \sim N[\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\psi})]$). Each observation y_i depends on a linear combination of a subset of the elements of $\boldsymbol{\theta}$ defined as follows:

$$(2) \qquad y_i|\boldsymbol{\theta}, \boldsymbol{\psi} \sim p\left(y_i \middle| \sum_j A_{ij}\theta_j, \psi\right)$$

The term A_{ij} is the generic element of an observation matrix $\mathbf{A}$ defined by the SPDE approach. Each y_i is independent and identically distributed given the latent field $\boldsymbol{\theta}$. Simplistically, the SPDE method allows one to fully evaluate the continuous GRF as a **discretely** indexed random process (i.e., a Gaussian Markov Random Field; Lindgren and Rue 2013), and it does so by subdividing the domain D (the area of the ocean where the fishing fleet is active) into triangular tiles, creating an index mesh (Lindgren and Rue 2013). Further built-in R-INLA commands are then used to construct the observation matrix $\mathbf{A}$ that extracts the values of the spatiotemporal random field at the measurement locations and time points used for the parameter estimation. The likelihood is linked to the latent field through η^* ($\eta^* = \mathbf{A}\eta$) as

$$(3) \qquad p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{i=1}^{n} p(y_i|\eta^*, \psi)$$

Since the Greenland shark bycatch data are counts characterized by many zeros, we evaluated Poisson and negative binomial distributions, as well as their zero-inflated versions. R-INLA offers different forms of zero-inflation, namely Types 0, 1, and 2. In a nutshell, Type 0 is a hurdle model that treats the response variable as being in a "perfect state" where the probability of bycatch can only be positive (i.e., does not include zero), whereas Types 1 and 2 are mixture models (2 being an extension of 1, that allows for additional zero probability), which describe the probability of being in an "imperfect state" where positive events (e.g., bycatch) may occur but are not certain and, as such, include both zero and nonzero values. For more details refer to Martin et al. (2005) and the R-INLA website. Table 2 lists all spatiotemporal correlation structures evaluated in our models. Please see Cameletti et al. (2012) and the R-INLA website for further documentation on these latent random field models.

Inference, goodness of fit, and prediction

All analyses were performed using R (R Development Core Team 2013), specifically the R-INLA package (Rue et al. 2009). The INLA procedure, in accordance with the Bayesian approach, calculates the marginal posterior distribution of all random effects and parameters involved in the model. There are different options offered in R-INLA with which to approximate the posterior marginal distributions; we used the most accurate one, the Laplace (Martins et al. 2013). We also used the default and recommended settings for priors (Held et al. 2010). These priors are vague priors or approximations of "non-informative" priors, which have little influence on the posterior distributions; hence, results are mostly derived from the data (similar to a frequentist approach). Prior sensitivity tests were conducted on the final best model. Further information on priors is provided in Appendix A.

To use the SPDE approach, the first step is to create a mesh on which the GRF is to be built; this mesh defines the spatial domain of interest (Ω; Fig. 1*b*). This is straightforward in R-INLA, but still requires some tuning. The mesh function creates by default a constrained refined Delaunay triangulation (CRDT) for the set of spatial locations provided (i.e., uses the observed fishing locations as the triangle nodes). If desired, the mesh could be derived from another set of points, for example, a regular grid. However, using the discrete fishing locations offers precision and efficiency. With CRDT, smaller size triangles can be defined in areas that have been sampled (fished) and larger ones in areas with no information (no fishing). This saves computational costs and increases the accuracy of the spatial field where there is fishing. Best mesh designs have more regular-shaped triangles and include some outer extension to avoid the "boundary effect" (increased variance at borders; Lindgren and Rue 2013). Different mesh designs were evaluated to investigate their effects during model selection.

Best candidate models were selected based on deviance information criterion (DIC; Spiegelhalter et al. 2002). Further model selection was performed on the three best models using cross-validation; 10 samples, each with 1100 observations, were randomly drawn from the dataset (total of 1641 observations) and fitted with each model; the remaining 541 observations were used for validation. DIC values and the mean squared error (MSE) using the log of the positive observed values and the linear predictor were calculated for each cross-validation. Note that MSE was only computed for the positive counts (bycatch of shark), since it is bycatch events that we are interested in predicting accurately.

A final model was chosen for model inference and prediction of bycatch hotspots for 2008–2011. Estimated bycatch can be predicted over the whole spatial domain (Ω) determined by the mesh (Fig. 1*b*). However, in our case study, there are large areas of Baffin Bay that are not fished. To avoid predicting bycatch in areas where fishing is unlikely, we first created a lattice of 1 km × 1 km grid cells over Ω and selected only the cells within a 5 km radius around our observed fishing locations. Note that these numbers (1 and 5 km) are arbitrary; we could choose a larger or smaller grid cell or radius, since we can predict bycatch everywhere in Ω. However, predictions on finer grids and larger areas of Ω will be more computationally intensive.

Bathymetry for each 1 km × 1 km grid cell was extracted from the NOAA TOPO1 raster. Since Greenland halibut catch also fluctuated in space and time, we interpolated catch for each grid cell using a weighted mean, where the weights are proportional to exp(–distance/a), where a is equal to 0.5 km, that is inversely proportional to the distance. Spatial predictions of the expected

Table 3. DIC values for all models tested.

	m0	m1	m2	m3a	m3b	m4	m5a	m5b
Poisson	2630.46	1400.51	1171.36	1176.61	1177.74	1144.19	1139.26	1143.51
zpoisson0	1678.52	1394.45	1375.65	1380.46	1376.68	1376.66	1370.33	1368.60
zpoisson1	1665.22	1321.45	1178.56	1180.23	1180.35	1151.52	1153.06	1137.73
zpoisson2	1859.05	1352.43	1186.43	1189.52	1187.79	1160.64	1165.65	NA
nbinom	1442.40	1297.25	1182.62	1183.76	1181.94	1149.13	1157.12	1163.13
znbinom0	1484.98	1415.29	1389.23	1391.14	1390.26	1385.27	1388.60	1400.54
znbinom1	1442.79	1298.10	1185.19	1186.59	1184.92	1154.65	1160.51	1168.08
znbinom2	1448.39	1292.72	1172.73	1174.23	1173.77	1153.38	1159.47	1168.11

Note: Poisson and negative binomial with 0, 1, 2 are referring to the zero-inflated types. Note that model m5b with zero-inflated Poisson 2 had a very large DIC suggesting that the model was inappropriate (NA).

bycatch for each year were then obtained by running the final model with the mean values of the other covariates (Duration = 15.94 and Panels = 43.33) and the grid values of the bathymetry and halibut catch. Results were then plotted on the 1 km × 1 km grid.

R scripts and datasets are available on the R-INLA website (http://www.r-inla.org).

Results

Model selection

DIC results of all models tested are presented in Table 3. Models ran from less than 10 s to 45 min on an Apple Darwin 10.8.0 (64-bit) platform with 2.4 GHz Intel Core 2 Duo Processor. Models fitted with a more elaborate spatiotemporal random field performed better than those with no random field (m0) or one that is constant over time (m1), irrespective of the likelihood. Models with spatial correlation among years (m3a and m3b) did not perform as well as the others, which treated years as independent.

Correlation among months was very high (around 90%), such that m5 models were very similar. For further model selection, m2, m4, and m5a were investigated. DIC and MSE results from these models' validation simulations are shown in Fig. 3. All models with the exception of their Type 0 zero-inflated Poisson and negative binomial versions have similar DIC values (Fig. 3). However, MSE results show that the zero-inflated Type 2 Poisson and negative binomial likelihood performs better with all three models (i.e., MSE values are smaller). However, when these models were tested with different mesh designs with more triangulations and different boundaries, the Poisson likelihood's DIC values were very sensitive (large fluctuations with zero values), whereas the negative binomial likelihoods were stable irrespective of the mesh used. For this reason, the zero-inflated Type 2 negative binomial likelihood and the most parsimonious mesh (simplest boundary and lesser number of triangulations) were selected.

All three models provide similar prediction performance, as seen from the MSE boxplots in Fig. 3. Model m4 is not a practical model for predictions purpose, since each month is an independent realization of the random field. Under m2, we have a common spatial pattern for months within each year and no correlation between years. Under m5a, we have a high correlation between consecutive months, within each year. Note that this may have arisen because of observed hauls in consecutive months but very close in time. For this reason and because it is a simpler model, we selected m2 as our final model. This model includes a different realization of a spatial correlation structure every year. The model ran for less than 6 min on our platform. Inference, posterior means, and simulations of fitting model m2 with a zero-inflated negative binomial (Type 2) model are presented below. Additional details on the structure of this final model can be found in Appendix A.

Inference and prediction

Results for all parameters are shown in Table 4 and Fig. 4. Duration of the hauls (soak time) had no significant effect. The number of gillnet panels was positively correlated with the amount of bycatch. For every 10 panels, the expected bycatch increased by approximately 30%. Total halibut catch was negatively correlated, such that higher expected bycatch was associated with hauls where less Greenland halibut was captured. Shallower waters (<1000 m) were associated with higher bycatch, whereas deeper waters (>1000 m) were associated with less bycatch; however, the credible intervals were large (Fig. 4). The spatial correlation range

Fig. 3. Model evaluation and selection. Shown are deviance information criterion (DIC) and mean squared error (MSE) boxplots under different likelihood families and spatiotemporal structures, considering different samples of the data. Grey shades refer to m2 (dark grey), m4 (medium grey), and m5 (light grey) models. Note that m5 refers to the m5a model.

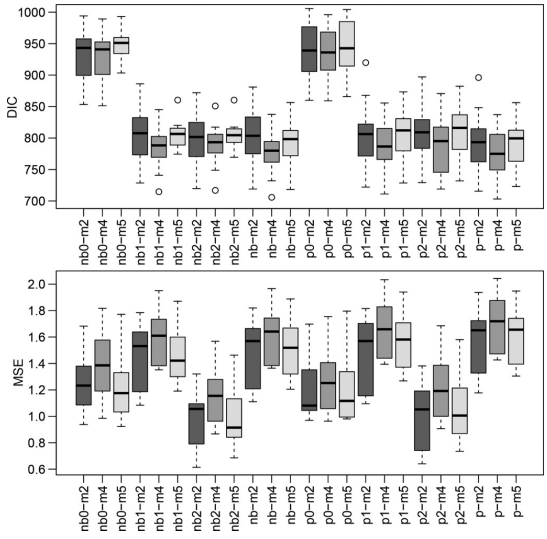


Table 4. Posterior estimates (mean and 95% credibility interval) of model parameters with negative binomial zero-inflated Type 2 likelihood and m2 spatiotemporal structure.

Parameters	Mean	$Q_{0.025}$	$Q_{0.975}$
Duration (hours and decimal minutes)	0.0053	−0.0170	0.0275
Number of gillnet panel	0.0271	0.0100	0.0446
Total halibut catch (t)	−0.1023	−0.1814	−0.0250
n	2.0316	0.9754	3.2489
α	0.7557	0.3753	1.1876
Nominal range (km)	174.7362	75.2938	291.5277

Fig. 4. Mean and the 2.5% and 97.5% quantiles for the posterior distribution of the bathymetry smoothed regression effect.

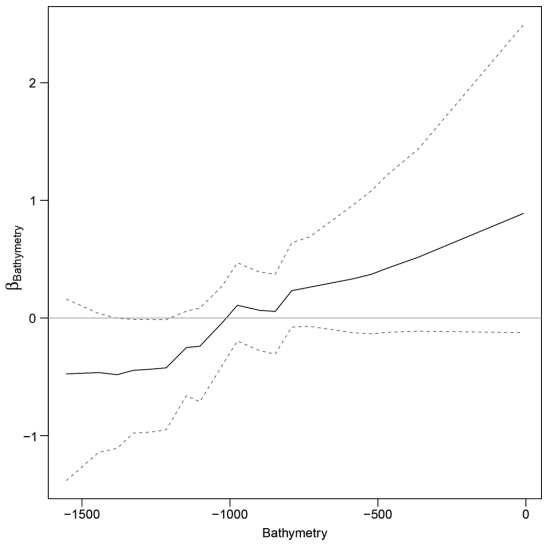
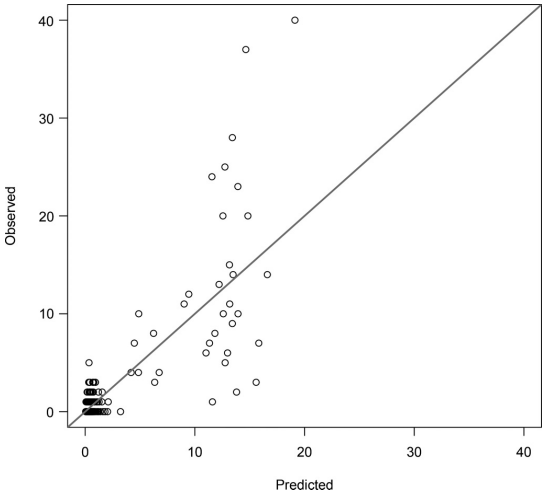


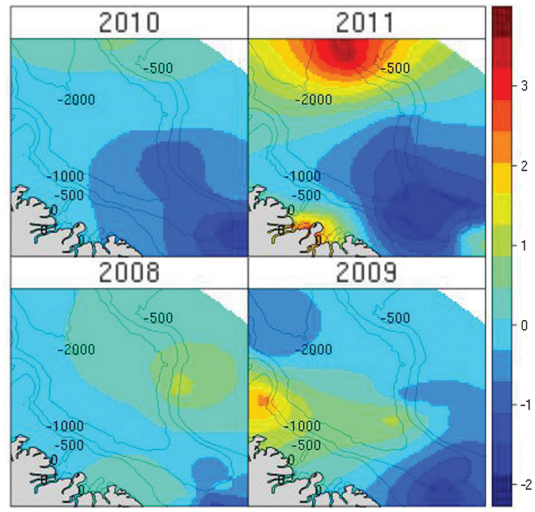
Fig. 5. Observed versus predicted Greenland shark bycatch (counts) of the final model, including a negative binomial zero-inflated Type 2 likelihood and m2 spatiotemporal structure.



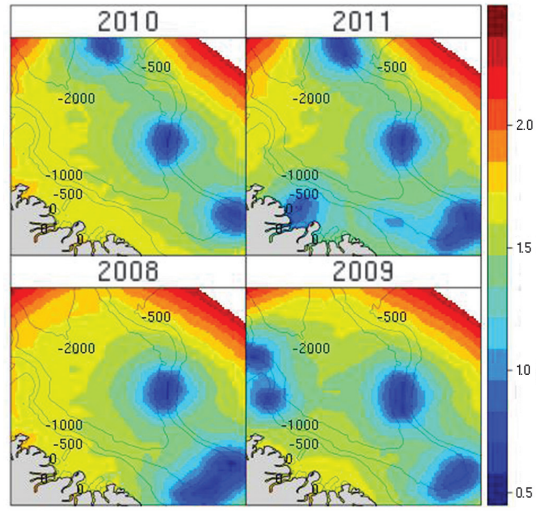
(nominal range) was 175 km with 95% credible interval (75, 292). The posterior mean of the overdispersion parameter (n) was 2.03 and the 95% credible interval was (0.98, 3.25). Since this overdispersion parameter was significantly greater than zero, we can conclude that there was evidence that bycatch was clustered. The posterior mean of α was 0.76 with 95% credible interval (0.4, 1.19). This indicated that the extra probability of zero was also significantly dependent on the linear predictor (i.e., greater values of the linear predictor resulted in less zeros).

Predicted versus observed bycatch of Greenland shark are shown in Fig. 5. The model was able to predict with more accuracy small to medium bycatch events (1–10 sharks per haul) but underestimated rare catch of large numbers of sharks (>20 sharks). This is expected since the zero-inflated negative binomial distribution is a generalization of the Poisson distribution, which assumes a variance equal to the mean, such that larger means are associated with larger errors. The mean and standard deviation of the yearly spatial random field are shown in Fig. 6. The final model included a different realization of the GRF every year, and accordingly, the

Fig. 6. Summary of the spatial random effect (Gaussian random field) every year included in the final model with negative binomial zero-inflated Type 2 likelihood and m2 spatiotemporal structure. For the coloured version of this figure, refer to the Web site at http://www.nrcresearchpress.com/doi/full/10.1139/cjfas-2014-0159.

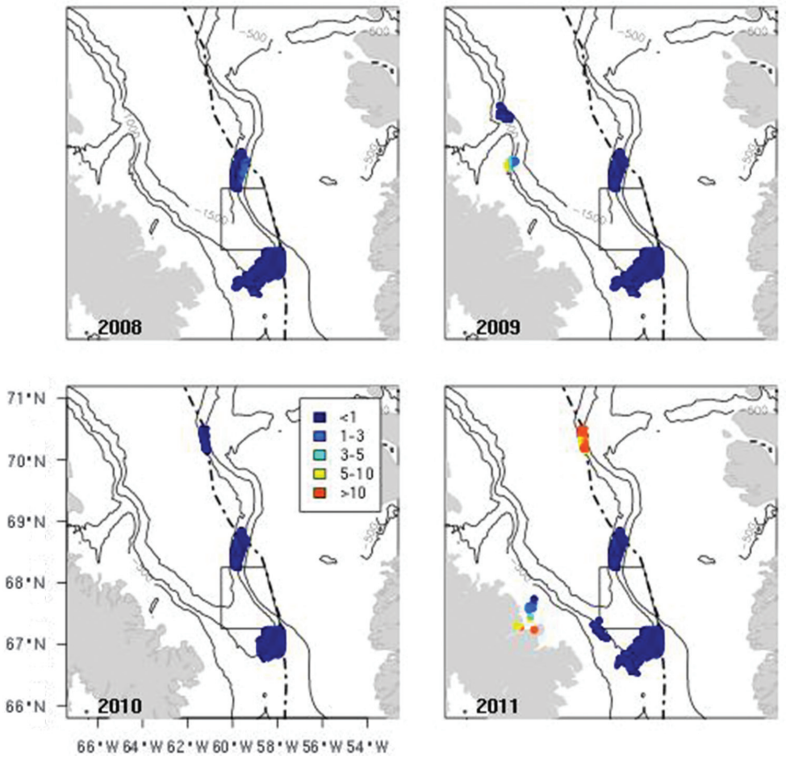


(a) Mean



(b) Standard deviation

Fig. 7. Simulated posterior means of the relative risk of Greenland shark bycatch (counts) in the Baffin Bay Greenland halibut gillnet fishery for the years 2008–2011. For the coloured version of this figure, refer to the Web site at http://www.nrcresearchpress.com/doi/full/10.1139/cjfas-2014-0159.



spatial effects differ from year to year. The southeastern part of Baffin Bay has always been negatively correlated with Greenland shark bycatch, while the waters near the coast (fished in 2009, 2011) and the northeastern fishing ground (fished in 2010, 2011) were positively correlated with Greenland shark bycatch. Standard deviation patterns for the spatial random field are driven by the amount of information; there is reduced uncertainty where fishing occurred. Prediction of bycatch risk (expected mean counts) of Greenland shark is shown in Fig. 7. Each year is different, but high levels of bycatch remain the same throughout the year (i.e., no month effect). Shallower waters observed in 2011 and the northernmost area in 2011 were associated with nearly 100 times higher bycatch risk, calculated as the exponential differences between the mean linear predictor values of these regions: exp[3.1 – (–1.5)].

Discussion

The objective of this study was to detail a computationally efficient and statistically powerful approach to analyze spatially explicit bycatch data and provide an example of its application using the case study of Greenland shark bycatch in Canadian gillnet fisheries in Baffin Bay. The first part of this discussion is concerned with the biological results and management applications, while the second part focuses on the model and modelling approach and future applications to bycatch studies.

Our results provide novel insight into the pattern of Greenland shark bycatch in gillnet fisheries. There is evidence that fishing locations less than 175 km apart are spatially correlated, hence sharing similar underlying processes (e.g., oceanographical, biological, fisheries). In our case, most of the fishing clusters are spatially dependent, with the exception of the southernmost and northernmost fishing clusters. This suggests that spatially explicit bycatch management could potentially be considered for these two regions. Greenland shark bycatch patterns are distinct from year to year, but hotspots remain the same for the fishing season. Given that the fishing season is currently short, limited by the ice-free season during the summer months, it is sensible that bycatch hotspots remain constant throughout the year. Maps show clear evidence of bycatch hotspots (100 times higher risk), particularly in shallower areas in coastal fjords near Broughton Island as well as to the northeast of Baffin Bay. Some of these coastal areas (Scott Inlet and Sam Ford Trough) may be nursery grounds for the species (N. Hussey, unpublished data), which may explain the higher incidence of bycatch in these areas. The northeast hotspot, on the other end, appeared to be associated with larger (>200 kg) individuals (A.C. Godin, unpublished data), which were present across a broad geographic area in the region (Campana et al. 2013). The northeast hotspot was only present in 2011, although was fished as well in 2009. This area may be of some biological importance to the species, but future bycatch monitoring and tracking studies are required here. Greenland shark bycatch is clustered; where bycatch occurs, it is more likely to catch more than one shark in the same haul, which supports recent tracking data suggesting that Greenland sharks have schooling behavior (N. Hussey, unpublished data). Note that we assume the count of a single animal in hauls with observed

bycatch mass but missing count information, such that this "clustering" parameter is most likely larger than estimated.

Given the increasing commercial fishing interest in the region (Christiansen et al. 2014) and the recent expansion of gillnet fisheries, and high postrelease mortality associated with gillnets (usually >70%; e.g., Manire et al. 2001; Thorpe and Frierson 2009), precautionary management that aims at minimizing bycatch of Greenland shark is timely. Our modelling approach allows forecasting of bycatch patterns on a year-by-year basis at a high spatial resolution. In our case study, hauls set in shallower waters, with more gillnet panels, and less halibut yield resulted in more shark bycatch. Hence, sensible bycatch management options could be to limit fishing in hotspots (e.g., using time–area closure or spatially explicit discard quotas), set a maximum number of gillnet panels per haul, or limit fishing to deep waters (>1000 m). The relationship with low halibut yield associated with higher shark bycatch is less intuitive. Greenland halibut may display some avoidance behaviors when confronted with higher abundance of predators, but more likely, habitat conditions along the Baffin coast (e.g., temperature, salinity) are not as favourable for Greenland halibut as they are for Greenland sharks. The above is by no means an exhaustive list of mitigation options, but highlights those that can be directly inferred from our model results. We caution the readers that the analysis is based on a limited dataset with only 4 replicated years. The data were statistically challenging with overdispersion, zero-inflation, and discontinuous temporal and spatial domains (i.e., patchy fishing pattern), and not all fishing grounds were fished every year. Our case study was not attempting to describe the full cost–benefit of bycatch mitigations, but rather to present an effective modelling scheme for inference and predictions, which could eventually be used to quantitatively assess the effect of spatially explicit management scenarios.

Our analysis captures the variations in fishing effort as well as the correlation with bathymetry and fishery variables, in addition to the spatiotemporal random effect. This spatiotemporal random field was necessary to rigorously capture the heterogeneity of Greenland shark bycatch. In general, bycatch studies always find that spatial and temporal factors are very influential factors for bycatch distribution, while other factors are of lesser importance (e.g., Megalofonou et al. 2009; Winter et al. 2011). Hotspots are sensitive to the structure of the data; Viana et al. (2012) showed that omitting the hierarchical and spatial and temporal structure in the modelling pattern of discards in the Irish Sea greatly affects the ability to characterize hotspots. This strongly indicates that there is a need to better incorporate space (and time) within the modelling scheme when analyzing bycatch data. However, there has been a slow progress in this area, most likely due to the unique difficulties arising from the spatial structure of fisheries data and the large computational burden of fitting complex spatiotemporal models.

The proposed Bayesian hierarchical spatiotemporal models are extremely powerful and flexible, specifically when the focus of the study is to quantify the spatial magnitude and uncertainty of this process. This is because the latent field is explicitly specified in the hierarchical model, such that information can be directly extracted from the model. To our knowledge, no applied bycatch studies to date have included such random processes to better capture the spatial and temporal correlation in the data (but see Sims et al. 2008; Viana et al. 2012). This is somewhat surprising considering the fact that the use of GRFs is very popular in the field of epidemiology (Best et al. 2005), where the research interest is analogous to bycatch studies: describing the variation of a biological process over space and time.

However, it is our hope that INLA, SPDE, and their R interface will allow fisheries bycatch scientists to explore these sophisticated models, which can now be more easily fitted. INLA avoids the vast computational demands of MCMC methods, and SPDE directly models georeferenced data rather than gridding the data first, which necessarily results in a substantial loss of information. Moreover, probability maps can easily be obtained from the posterior distribution, and results are more intuitive and interpretable (i.e., the probability is explicit) for nonstatisticians than p values (Wade 2000). Unfortunately, many fisheries receive very little observer coverage, mostly because of the expense associated with such programs. Hence, observer-based bycatch datasets are often of poor quality, and maps of raw bycatch rates can be misleading (Sims et al. 2008). In such cases, it is especially important to include the hierarchical structure and spatiotemporal heterogeneity to better estimate total bycatch and hotspot locations for the fleet of interest. The framework we developed here allows for this and can be extended to complex situation where multispecies bycatch occurrences are of interest (Sims et al. 2008). Moreover, since the computational cost is low and time constraints not an issue, scientists can now simulate a large number of cost–benefit scenarios with which to investigate best management options.

R-INLA is continuously evolving, greatly extending the scope of models available to applied scientists. These methods have now been applied in health sciences (Bessell et al. 2010; Li et al. 2012; Musenge et al. 2012; Wilking et al. 2012), climate research (Cameletti et al. 2012), ecology (Johnson et al. 2011; Haas et al. 2011; Holand et al. 2013; Illian et al. 2013), and recently in marine ecology research (Muñoz et al. 2012). The package interface resembles the glm function in R, such that scientists already familiar with these common tools should find it straightforward to use R-INLA. An advantage of the Bayesian approach is that prior knowledge can be incorporated into the model to get more precise posterior distributions. Instead of using noninformative priors (default priors in R-INLA), user-friendly commands for prior specificity are currently undergoing development (H. Rue, personal communication). Moreover, there is great interest in bycatch studies in combining datasets from different sources (e.g., oceanographic data) to better predict high incidental catch. Each subset of data may be described by a different family of distributions. Fitting models with different likelihoods is nearly impossible with most popular R packages, but is implemented in R-INLA. See Martins et al. (2013) for a complete overview of INLA's new features.

The models that we have proposed here are not yet widely used in fisheries and aquatic sciences. To this end, we have demonstrated the merits of using Bayesian hierarchical spatiotemporal models, for bycatch studies in particular, but have gone further to present an effective way of fully implementing these models (R-INLA). Bycatch studies tend to have a strong ecological focus, seeking insight into the causal drivers of bycatch frequency and spatial distribution. However, increasing demand for mapped products for fisheries spatial planning is requiring scientists to look into new statistical techniques that can explicitly include the inherent uncertainties associated with the observations, sampling, models, and parameters, such that accurate predictions of bycatch can be obtained and included in decision-making. Under a Bayesian hierarchical model, all uncertainties (e.g., resulting from sampling bias and uncertainty due to the variability across space (and time)) are incorporated, and as such it is straightforward to obtain posterior predictive probability distributions for nonsampled areas as demonstrated in our case study. By making available both the data and the code necessary to implement the models for our particular application, we hope to facilitate the uptake of these powerful approaches by marine scientists.

Acknowledgements

We acknowledge DFO, especially Margaret Treble, who provided in-depth details on the Greenland halibut fishery, and Todd Inkpen and Carol Ann Peters, who supplied the observer data. We are greatly indebted to Havard Rue for hosting A.C.G. and for generous advice and suggestions throughout this project. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through an Industrial Post-

graduate Scholarship to A.C.G., a Discovery Grant to J.M.F., and a Canada Strategic Network Grant (NETGP 375118-08) to J.M.F. via the Ocean Tracking Network. In addition, A.C.G. received funding from WWF–Canada, and E.T.K. received funds from the Research Council of Norway.

References

Amante, C., and Eakins, B.W. 2014. ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis [online]. NOAA Tech. Memo. NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M. Available from http://www.ngdc.noaa.gov/mgg/global/global.html [accessed 29 October 2014].

Barlow, P.F., and Berkson, J. 2012. Evaluating methods for estimating rare events with zero-heavy data: a simulation model estimating sea turtle bycatch in the pelagic longline fishery. Fish. Bull. **110**(3): 344–360.

Bessell, P.R., Matthews, L., Smith-Palmer, A., Rotariu, O., Strachan, N.J., Forbes, K.J., Cowden, J.M., Reid, S.W., and Innocent, G.T. 2010. Geographic determinants of reported human Campylobacter infections in Scotland. BMC Public Health, **10**: 423. doi:10.1186/1471-2458-10-423. PMID:20633277.

Best, N., Richardson, S., and Thomson, A. 2005. A comparison of Bayesian spatial models for disease mapping. Stat. Methods Med. Res. **14**(1): 35–59. doi:10.1191/0962280205sm388oa. PMID:15690999.

Bjorge, A., Skern-Mauritzen, M., and Rossman, M.C. 2013. Estimated bycatch of harbour porpoise (*Phocoena phocoena*) in two coastal gillnet fisheries in Norway, 2006–2008. Mitigation and implications for conservation. Biol. Conserv. **161**: 164–173. doi:10.1016/j.biocon.2013.03.009.

Brodziak, J., and Walsh, W.A. 2013. Model selection and multimodel inference for standardizing catch rates of bycatch species: a case study of oceanic whitetip shark in the Hawaii-based longline fishery. Can. J. Fish. Aquat. Sci. **70**(12): 1723–1740. doi:10.1139/cjfas-2013-0111.

Cambiè, G., Sánchez-Carnero, N., Mingozzi, T., Muiño, R., and Freire, J. 2012. Identifying and mapping local bycatch hotspots of loggerhead sea turtles using a GIS-based method: implications for conservation. Mar. Biol. **160**(3): 653–665. doi:10.1007/s00227-012-2120-5.

Cameletti, M., Finn Lindgren, F., Simpson, D., and Rue, H. 2012. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. Adv. Stat. Anal. **97**(2): 109–131. doi:10.1007/s10182-012-0196-3.

Campana, S.E., Fisk, A.T., and Klimley, A.P. 2013. Movements of Arctic and northwest Atlantic Greenland shark (*Somniosus microcephalus*) monitored with archival satellite pop-up tags suggest long-range migrations. Deep Sea Res Part II Top. Stud. Oceanogr. [In press.] doi:10.1016/j.dsr2.2013.11.001.

Christiansen, J.S., Mecklenburg, C.W., and Karamushko, O.V. 2014. Arctic marine fishes and their fisheries in light of global change. Global Change Biol. **20**(2): 352–359. doi:10.1111/gcb.12395. PMID:24105993.

Ciannelli, L., Fauchald, P., Chan, K., Agostini, V., and Dingsor, G. 2008. Spatial fisheries ecology: recent progress and future prospects. J. Mar. Syst. **71**(3–4): 223–236. doi:10.1016/j.jmarsys.2007.02.031.

Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D.H., and Wear, D. 2001. Ecological forecasts: an emerging imperative. Science, **293**(5530): 657–660. doi:10.1126/science.293.5530.657. PMID:11474103.

Cressie, N., Calder, C.A., Clark, J.S., Ver Hoef, J.M., and Wikle, C.K. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecol. Appl. **19**(3): 553–570. doi:10.1890/07-0744.1. PMID:19425416.

Davies, R.W.D., Cripps, S.J., Nickson, A., and Porter, G. 2009. Defining and estimating global marine fisheries bycatch. Mar. Pol. **33**(4): 661–672. doi:10.1016/j.marpol.2009.01.003.

Davis, B., VanderZwaag, D.L., Cosandey-Godin, A., Hussey, N.E., Kessel, S.T., and Worm, B. 2013. The conservation of the Greenland shark (*Somniosus microcephalus*): setting scientific, law, and policy coordinates for avoiding a species at risk. J. Int. Wildl. Law Pol. **16**(4): 300–330. doi:10.1080/13880292.2013.805073.

DFO. 2005. Fishery management plan Greenland halibut. NAFO Subarea 0 2006–2008. Fisheries and Oceans Canada, Winnipeg, Man.

DFO. 2007. Development of a closed area in NAFO 0A to protect narwhal overwintering grounds, including deep-sea corals. Fisheries and Oceans Canada, Central and Arctic Region.

DFO. 2014. Integrated fishery management plan Greenland halibut (*Reinhardtius hippoglossoides*) Northwest Atlantic Fisheries Organization Subarea 0 Effective 2014.

Diggle, P.J., and Ribeiro, P.J., Jr. 2007. Model-based geostatistics. Springer, New York.

Douvere, F. 2008. The importance of marine spatial planning in advancing ecosystem-based sea use management. Mar. Pol. **32**(5): 762–771. doi:10.1016/j.marpol.2008.03.021.

Dunn, D.C., Boustany, A.M., and Halpin, P.N. 2011. Spatio-temporal management of fisheries to reduce by-catch and increase fishing selectivity. Fish Fish. **12**(1): 110–119. doi:10.1111/j.1467-2979.2010.00388.x.

FAO. 2011. International guidelines on bycatch management and reduction of discards [online]. Food and Agriculture Organization of the United Nations, Rome. Available from http://www.fao.org/docrep/015/ba0022t/ba0022t00.pdf.

Garcia, S.M., Zerbi, A., Aliaume, C., Do Chi, T., and Lasserre, G. 2003. The ecosystem approach to fisheries. Issues, terminology, principles, institutional foundations, implementation and outlook. FAO Fisheries Technical Paper, FAO, Rome, Italy.

Gardner, B., Sullivan, P.J., Epperly, S., and Morreale, S.J. 2008. Hierarchical modeling of bycatch rates of sea turtles in the western North Atlantic. Endangered Species Res. **5**(December): 279–289. doi:10.3354/esr00105.

Haas, S.E., Hooten, M.B., Rizzo, D.M., and Meentemeyer, R.K. 2011. Forest species diversity reduces disease risk in a generalist plant pathogen invasion. Ecol. Lett. **14**(11): 1108–1116. doi:10.1111/j.1461-0248.2011.01679.x. PMID:21884563.

Hall, M.A., Alverson, D.L., and Metuzals, K.I. 2000. By-Catch: Problems and Solutions. Mar. Pollut. Bull. **41**(1–6): 204–219. doi:10.1016/S0025-326X(00)00111-9.

Harden, L.A., and Willard, A.S. 2012. Using spatial and behavioral data to evaluate the seasonal bycatch risk of diamondback terrapins *Malaclemys terrapin* in crab pots. Mar. Ecol. Progr. Ser. **467**: 207–217. doi:10.3354/meps09958.

Harley, S.J., and Myers, R.A. 2001. Hierarchical Bayesian models of length-specific catchability of research trawl surveys. Can. J. Fish. Aquat. Sci. **58**(8): 1569–1584. doi:10.1139/f01-097.

Held, L., Schrodle, B., and Rue, H. 2010. Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. *In* Statistical modelling and regression structures. pp. 111–131.

Holand, A.M., Steinsland, I., Martino, S., and Jensen, H. 2013. Animal models and integrated nested Laplace approximations. G3: Genes Genomes Genet. **3**(8): 1241–1251. doi:10.1534/g3.113.006700. PMID:23708299.

Huang, H.-W., and Yeh, Y.-M. 2011. Impact of Taiwanese distant water longline fisheries on the Pacific seabirds: finding hotspots on the high seas. Anim. Conserv. **14**(5): 562–574. doi:10.1111/j.1469-1795.2011.00462.x.

Illian, J.B., Martino, S., Sørbye, S.H., Gallego-Fernández, J.B., Zunzunegui, M., Paz Esquivias, M., and Travis, J.M.J. 2013. Fitting complex ecological point process models with integrated nested Laplace approximation. Methods Ecol. Evol. **4**(4): 305–315. doi:10.1111/2041-210x.12017.

Jannot, J.E., and Holland, D.S. 2013. Identifying ecological and fishing drivers of bycatch in a U.S. groundfish fishery. Ecol. Appl. **23**(7): 1645–1658. doi:10.1890/12-2225.1.

Jiménez, S., Domingo, A., and Brazeiro, A. 2009. Seabird bycatch in the Southwest Atlantic: interaction with the Uruguayan pelagic longline fishery. Polar Biol. **32**: 187–196. doi:10.1007/s00300-008-0519-8.

Johnson, D.S., London, J.M., and Kuhn, C.E. 2011. Bayesian inference for animal space use and other movement metrics. J. Agric. Biol. Environ. Stat. **16**(3): 357–370. doi:10.1007/s13253-011-0056-8.

Kelleher, K. 2005. Discards in the worlds marine fisheries. An update [online]. Available from http://www.fao.org/docrep/008/y5936e/y5936e00.htm.

Kirby, D.S., and Ward, P. 2014. Standards for the effective management of fisheries bycatch. Mar. Pol. **44**: 419–426. doi:10.1016/j.marpol.2013.10.008.

Lewison, R.L., Crowder, L.B., Read, A.J., and Freeman, S.A. 2004. Understanding impacts of fisheries bycatch on marine megafauna. Trends Ecol. Evol. **19**(11): 598–604. doi:10.1016/j.tree.2004.09.004.

Li, Y., Brown, P., Rue, H., al-Maini, M., and Fortin, P. 2012. Spatial modelling of lupus incidence over 40 years with changes in census areas. J. R. Stat. Soc. Ser. C Appl. Stat. **61**(1): 99–115. doi:10.1111/j.1467-9876.2011.01004.x.

Lindgren, F. 2013. Continuous domain spatial models in R-INLA. ISBA Bull. **19**(4): 1–8.

Lindgren, F., and Rue, H. 2013. Bayesian spatial and spatio-temporal modelling with R-INLA. Journal of Statistical Software. [In press.]

Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. **73**(4): 423–498. doi:10.1111/j.1467-9868.2011.00777.x.

MacNeil, M.A., McMeans, B.C., Hussey, N.E., Vecsei, P., Svavarsson, J., Kovacs, K.M., Lydersen, C., Treble, M.A., Skomal, G.B., Ramsey, M., and Fisk, A.T. 2012. Biology of the Greenland shark *Somniosus microcephalus*. J. Fish Biol. **80**(5): 991–1018. doi:10.1111/j.1095-8649.2012.03257.x. PMID:22497371.

Manire, C., Hueter, R., Hull, E., and Spieler, R. 2001. Serological changes associated with gill-net capture and restraint in three species of sharks. Trans. Am. Fish. Soc. **130**(6): 1038–1048. doi:10.1577/1548-8659(2001)130<1038:SCAWGN>2.0.CO;2.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., and Possingham, H.P. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol. Lett. **8**(11): 1235–1246. doi:10.1111/j.1461-0248.2005.00826.x. PMID:21352447.

Martins, T.G., Simpson, D., Lindgren, F., and Rue, H. 2013. Bayesian computing with INLA: new features. Comput. Stat. Data Anal. **67**: 68–83. doi:10.1016/j.csda.2013.04.014.

McClellan, C.M., Read, A.J., Price, B.A., Cluse, W.M., and Godfrey, M.H. 2009. Using telemetry to mitigate the bycatch of long-lived marine vertebrates. Ecol. Appl. **19**(6): 1660–1671. doi:10.1890/08-1091.1. PMID:19769110.

Megalofonou, P., Damalas, D., Deflorio, M., and De Metrio, G. 2009. Modeling environmental, spatial, temporal, and operational effects on blue shark bycatches in the Mediterranean long-line fishery. J. Appl. Ichthyol. **25**: 47–55. doi:10.1111/j.1439-0426.2009.01221.x.

Minami, M., Lennert-Cody, C.E., Gao, W., and Román-Verdesoto, M. 2007. Mod-

eling shark bycatch: The zero-inflated negative binomial regression model with smoothing. Fish. Res. **84**: 210–221. doi:10.1016/j.fishres.2006.10.019.

Moore, J.E., and Read, A.J. 2008. A Bayesian uncertainty analysis of cetacean demography and bycatch mortality using age-at-death data. Ecol. Appl. **18**(8): 1914–1931. doi:10.1890/07-0862.1. PMID:19263888.

Muñoz, F., Pennino, M.G., Conesa, D., López-Quílez, A., and Bellido, J.M. 2012. Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. Stoch. Environ. Res. Risk Assess. **27**(5): 1171–1180. doi:10.1007/s00477-012-0652-3.

Murray, K.T., and Orphanides, C.D. 2013. Estimating the risk of loggerhead turtle *Caretta caretta* bycatch in the US mid-Atlantic using fishery-independent and -dependent data. Mar. Ecol. Progr. Ser. **477**: 259–270. doi:10.3354/meps10173.

Musenge, E., Chirwa, T.F., Kahn, K., and Vounatsou, P. 2012. Bayesian analysis of zero inflated spatiotemporal HIV/TB child mortality data through the INLA and SPDE approaches: Applied to data observed between 1992 and 2010 in rural North East South Africa. Int. J. Appl. Earth Obs. Geoinform. **22**: 86–98. doi:10.1016/j.jag.2012.04.001.

Orphanides, C.D. 2010. Protected species bycatch estimating approaches: estimating harbor porpoise bycatch in U.S. northwestern Atlantic gillnet fisheries. J. Northw. Atl. Fish. Sci. **42**: 55–76. doi:10.2960/J.v42.m647.

Ortiz, M., and Arocha, F. 2004. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. Fish. Res. **70**(2–3): 275–297. doi:10.1016/j.fishres.2004.08.028.

Pikitch, E.K., Santora, C., Babcock, E.A., Bakun, A., Bonfil, R., Conover, D.O., Dayton, P., Doukakis, P., Fluharty, D., Heneman, B., Houde, E.D., and Link, J. 2004. Ecosystem-based fishery management. Science, **305**: 346–347. doi:10.1126/science.1098222. PMID:15256658.

R Development Core Team. 2013. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.r-project.org/.

Rivot, E., Prévost, E., Cuzol, A., Baglinière, J.-L., and Parent, E. 2008. Hierarchical Bayesian modelling with habitat and time covariates for estimating riverine fish population size by successive removal method. Can. J. Fish. Aquat. Sci. **133**: 117–133. doi:10.1139/F07-153.

Roe, J.H., Morreale, S.J., Paladino, F.V., Shillinger, G.L., Benson, S.R., Eckert, S.A., Bailey, H., Tomillo, P.S., Bograd, S.J., Eguchi, T., Dutton, P.H., Seminoff, J.A., Block, B.A., and Spotila, J.R. 2014. Predicting bycatch hotspots for endangered leatherback turtles on longlines in the Pacific Ocean. Proc. R. Soc. B Biol. Sci. **281**: doi:10.1098/rspb.2013.2559. PMID:24403331.

Rue, H., and Held, L. 2005. Gaussian Markov random fields: theory and applications [online]. Chapman and Hall CRC, Boca Raton, Fla. Available from http://www.math.ntnu.no/~hrue/GMRF-book/.

Rue, H., and Martino, S. 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. J. Stat. Plann. Infer. **137**(10): 3177–3192. doi:10.1016/j.jspi.2006.07.016.

Rue, H., Martino, S., and Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. **71**(2): 319–392. doi:10.1111/j.1467-9868.2008.00700.x.

Sims, M., Cox, T., and Lewison, R. 2008. Modeling spatial patterns in fisheries bycatch: improving bycatch maps to aid fisheries management. Ecol. Appl. **18**(3): 649–661. doi:10.1890/07-0685.1. PMID:18488624.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B Stat. Methodol. **64**(4): 583–639. doi:10.1111/1467-9868.00353.

Thorpe, T., and Frierson, D. 2009. Bycatch mitigation assessment for sharks caught in coastal anchored gillnets. Fish. Res. **98**(1–3): 102–112. doi:10.1016/j.fishres.2009.04.003.

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. **2**(46): 234–240. doi:10.2307/143141.

Trebilco, R., Gales, R., Lawrence, E., Alderman, R., Robertson, G., and Baker, G.B. 2010. Characterizing seabird bycatch in the eastern Australian tuna and billfish pelagic longline fishery in relation to temporal, spatial and biological influences. Aquat. Conserv. Mar. Freshw. Ecosyst. **20**(5): 531–542. doi:10.1002/aqc.1115.

Venables, W., and Dichmont, C. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fish. Res. **70**(2–3): 319–337. doi:10.1016/j.fishres.2004.08.011.

Viana, M., Jackson, A.L., Graham, N., and Parnell, A.C. 2012. Disentangling spatio-temporal processes in a hierarchical system: a case study in fisheries discards. Ecography, **36**(5): 569–578. doi:10.1111/j.1600-0587.2012.07853.x.

Wade, P.R. 2000. Bayesian methods in conservation biology. Conserv. Biol. **14**(5): 1308–1316. doi:10.1046/j.1523-1739.2000.99415.x.

Wikle, C.K. 2003. Hierarchical models in environmental science. Int. Stat. Rev. **71**(2): 181–199. doi:10.1111/j.1751-5823.2003.tb00192.x.

Wilking, H., Höhle, M., Velasco, E., Suckau, M., and Eckmanns, T. 2012. Ecological analysis of social risk factors for Rotavirus infections in Berlin, Germany, 2007–2009. Int. J. Health Geogr. **11**: 37. doi:10.1186/1476-072X-11-37. PMID:22929067.

Winter, A., Jiao, Y., and Browder, J.A. 2011. Modeling low rates of seabird bycatch in the US Atlantic long line fishery. Waterbirds, **34**(3): 289–303. doi:10.1675/063.034.0304.

Wintle, B.A., McCarthy, M.A., Volinsky, C.T., and Kavanagh, R.P. 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. Conserv. Biol. **17**(6): 1579–1590. doi:10.1111/j.1523-1739.2003.00614.x.

Yeh, Y.-M., Huang, H.-W., Dietrich, K.S., and Melvin, E. 2013. Estimates of seabird incidental catch by pelagic longline fisheries in the South Atlantic Ocean. Anim. Conserv. **16**(2): 141–152. doi:10.1111/j.1469-1795.2012.00588.x.

Yu, R., and Leung, P. 2010. A Bayesian hierarchical model for modeling white shrimp (*Litopenaeus vannamei*) growth in a commercial shrimp farm. Aquaculture, **306**(1–4): 205–210. doi:10.1016/j.aquaculture.2010.04.028.

Appendix A

Model

In the generalized linear model (GLM) framework, we define a link function that maps the mean of the response to the linear predictor. For example, under the assumption that the response variable has a negative binomial (NB) distribution and that the bycatch at location $\mathbf{s}$ at time t is equal to k with dispersion parameter n, then we can write $P[Y(\mathbf{s}, t) = k]$ as

$$\mathrm{NB}[k, n, \theta(\mathbf{s}, t)] = \frac{\Gamma(k + n)}{\Gamma(n)\Gamma(k + 1)}\theta(\mathbf{s}, t)[1 - \theta(\mathbf{s}, t)]$$

The mean of the above distribution, $\mu(\mathbf{s}, t) = n\frac{1-\theta(\mathbf{s}, t)}{\theta(\mathbf{s}, t)}$, is linked to the linear predictor $\eta(\mathbf{s}, t)$ by $\mu(\mathbf{s}, t) = \exp[\eta(\mathbf{s}, t)]$, where

$$\eta(\mathbf{s}, t) = \boldsymbol{X}(\mathbf{s}, t)\boldsymbol{\beta} + f(\mathbf{s}, t)$$

Note that in the above expression, $\mathbf{X}(\mathbf{s}, t)$ is the design matrix with p fixed covariates, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)$ is the regression coefficients vector, and $f(\mathbf{s}, t)$ introduces the spatiotemporal structure via random effects by considering a Gaussian random field (that is spatially and temporally correlated) as per Cameletti et al. (2012). To proceed, we consider $f(\mathbf{s}, t)$ to be a realization of a continuously indexed spatial process (random field) changing in time denoted by

$$\{f(\mathbf{s}, t) : (\mathbf{s}, t) \in D \subseteq \Re^2 \times \Re\}$$

These realizations are used to make inference about the process and to predict it at desired locations. Usually, we deal with a Gaussian field (GF) that is completely specified by its mean and spatiotemporal covariance function $\mathrm{cov}[f(\mathbf{s}, t), f(\mathbf{s}', t')] = \sigma^2\mathrm{cov}[(\mathbf{s}, t), (\mathbf{s}', t')]$, defined for each $(\mathbf{s}, t)$ and $(\mathbf{s}', t')$ in $\Re^2 \times \Re$. Unfortunately, implementation of a GF suffers from the so-called "big n problem" because it requires the factorization of the covariance matrix to compute the likelihood (see Diggle and Ribeiro 2007); this arises particularly with large datasets in space and time. One solution is to represent a GF as a **discretely** indexed random process (i.e., a Gaussian Markov random field, GMRF; see Rue and Held 2005). This proposal is based on the work of Lindgren et al. (2011), where an explicit link between GFs and GMRFs is proven through the use of stochastic partial differential equations (SPDE). This is the approach taken here.

By assuming a zero-inflated negative binomial distribution rather than a standard negative binomial distribution, we are allowing for additional probability of $Y(\mathbf{s}, t) = 0$. For example, under the Type 2 formulation, we have

$$P[Y(\mathbf{s}, t) = k] = p(\mathbf{s}, t)I_{k=0} + [1 - p(\mathbf{s}, t)] \times \mathrm{NB}[k, n, \theta(\mathbf{s}, t)]$$

where $p(\mathbf{s}, t)$ is the extra probability added to the negative binomial for $k = 0$ (a similar relationship exists between the Poisson and zero-inflated Poisson distribution). The extra probability of zero, $p(\mathbf{s}, t)$, depends on the linear predictor $\eta(\mathbf{s}, t)$ (i.e., the extra

probability depends on the covariates and on the spatiotemporal random effect). That is

$$p(\mathbf{s}, t) = 1 - (\exp[\eta(\mathbf{s}, t)]/\{1 + \exp[\eta(\mathbf{s}, t)]\})^{\alpha}$$

This means that the probability of zero bycatch at location **s** at time t is inversely proportional to the linear predictor $\eta(\mathbf{s}, t)$, which makes sense because under the negative binomial distribution, for example, we have $\mu(\mathbf{s}, t) = \exp[\eta(\mathbf{s}, t)]$ (i.e., $\eta(\mathbf{s}, t)$ is proportional to the expected number of sharks that were recorded as bycatch). We also have $\pi(\mathbf{s}, t) = \exp[\eta(\mathbf{s}, t)]/\{1 + \exp[\eta(\mathbf{s}, t)]\}$ proportional to $\eta(\mathbf{s}, t)$, and writing this in one equation, we arrive at

$$P[Y(\mathbf{s}, t) = k] = [1 - \pi(\mathbf{s}, t)^{\alpha}]_{k=0} + [\pi(\mathbf{s}, t)^{\alpha}] \times \mathrm{NB}[k, n, \theta(\mathbf{s}, t)]$$

The parameter α can be thought of as a hyperparameter that controls both the impact of the linear predictor on the extra probability at zero as well as that for the positive counts. When $\alpha = 0$ we have $\pi(\mathbf{s}, t)^{\alpha} = 1$ and simply $P[Y(\mathbf{s}, t) = k] = \mathrm{NB}[k, n, \theta(\mathbf{s}, t)]$, namely, no **extra** probability at zero. Because $\pi(\mathbf{s}, t)^{\alpha} \in (0, 1)$ and $\alpha > 0$ if α increases, $\pi(\mathbf{s}, t)$ also increases and the extra probability of zero decreases.

Priors

We used the default priors and hyperparameters currently implemented in R-INLA. For further information, refer to the R-INLA documentation available on the website (http://www.r-inla.org). Note that this is an active area of research for the R-INLA team.

Table A1. Summary of the default priors used in the model tested.

$\log(n)$	loggamma(1, 1)
$\log(\alpha)$	N(0.693, 1)
β_0	N(0, ∞)
β_j, $j = 1, 2, 3$	N(0, 1/0.001)
τ_1	Gamma(1, 0.00005)
θ_1	N(0, 10)
θ_2	N(0, 1)
$\log\left(\frac{1+\rho}{1-\rho}\right)$	N(0, 0.15)

For the overdispersion parameter n (size), represented as $\log(n)$, the prior is a loggamma(1, 1). For the α parameter on the Type 2 zero-inflated model, we assume N(0.693, 1) for $\log(\alpha)$.

Every components of the latent field $\boldsymbol{\theta} = \{\beta_0, \boldsymbol{\beta}, \boldsymbol{f}\}$ have priors. For the smooth function of bathymetry, we assumed a random walk of order one, which is defined in terms of a Gaussian distribution N(0, $\tau_1\mathbf{R}_1$), where $\mathbf{R}_1$ is the (fixed and known) structure matrix, and τ_1 is the precision parameter. To model the spatial correlation, we assume a spatial Matérn correlation using the SPDE approach, with parameters $\theta_1 = \log(\tau)$ and $\theta_2 = \log(\kappa)$ and priors defined in Lindgren and Rue (2013).

The SPDE framework allows the easy linkage of different GRFs in time and in this way develop a dynamic model where the spatial field evolves according to time-varying processes, for example, an autoregressive process. For these dynamic models (i.e., m3a, m3b, m5a, and m5b), we also had an additional correlation parameter ρ for time, $\log\left(\frac{1+\rho}{1-\rho}\right)$, with a Gaussian prior.

The influence of socioeconomic deprivation, access to healthcare and physical environment on old-age survival in Portugal

Ana Isabel Ribeiro,[1-4] Elias Teixeira Krainski,[5,6] Marilia Sá Carvalho,[7] Maria de Fátima de Pina[2,3,8,9]

[1]EPIUnit-Instituto de Saúde Pública, Universidade do Porto; [2]i3S-Instituto de Investigação e Inovação em Saúde, Universidade do Porto; [3]INEB-Instituto de Engenharia Biomédica, Universidade do Porto; [4]Department of Clinical Epidemiology, Predictive Medicine and Public Health, University of Porto Medical School, Portugal; [5]The Norwegian University for Science and Technology, Trondheim, Norway; [6]Departamento de Estatística, Universidade Federal do Paraná, Curitiba; [7]PROCC-Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro; [8]ICICT/FIOCRUZ-Instituto de Comunicação e Informação Científica e Tecnológica em Saúde/Fundação Oswaldo Cruz, Rio de Janeiro; [9]CARTO-FEN/UERJ-Departamento de Engenharia Cartográfica, Faculdade de Engenharia da Universidade do Estado do Rio de Janeiro, Brazil

Abstract

Spatial inequalities in old-age survival exist in Portugal and might be associated with factors pertaining to three distinct domains: socioeconomic, physical environmental and healthcare. We evaluated the contribution of these factors on the old-age survival across Portuguese municipalities deriving a surrogate measure of life expectancy, a 10-year survival rate that expresses the proportion of the population aged 75-84 years old who reached 85-94. As covariates we used two internationally comparable multivariate indexes: the European deprivation index and the multiple physical environmental deprivation index. A national index was developed to evaluate the access to healthcare. Smoothed rates and odds ratios (OR) were estimated using Bayesian spatial models. Socioeconomic deprivation was found to be the most relevant factor influencing old-age survival in Portugal [women: least deprived areas OR=1.132(1.064-1.207); men OR=1.044(1.001-1.094)] and explained a sizable amount of the spatial variance in survival, especially among women. Access to healthcare was associated with old-age survival in the univariable model only; results lost significance after adjustment for socioeconomic circumstances [women: higher access to healthcare OR=1.020(0.973-1.072); men OR=1.021(0.989-1.060)]. Physical environmental deprivation was unrelated with old-age survival. In conclusion, socioeconomic deprivation was the most important determinant in explaining spatial disparities in old-age survival in Portugal, which indicates that policy makers should direct their efforts to tackle socioeconomic differentials between regions.

Correspondence: Ana Isabel Ribeiro, EPIUnit – Instituto de Saúde Pública, Universidade do Porto, Rua das Taipas 135, 4050-600, Porto, Portugal.
Tel: +351.222061820 - Fax: +351.222061821.
E-mail: ana.isabel.ribeiro@ispup.up.pt

Key words: Inequalities; Ageing; Spatial analysis; Geographic Information Systems; Portugal.

Acknowledgments: this work was supported by Portuguese funds through FCT – Fundação para a Ciência e a Tecnologia in the framework of project UID/BIM/04293/2013. AIR would also like to thank to FCT – Fundação para a Ciência e a Tecnologia for the PhD grant SFRH/BD/82529/2011. MSC was supported by CNpQ (309692/2013-0) and FAPERJ (E-26/203.557/2014). The EPIUnit, Institute of Public Health, University of Porto, Portugal (UID/DTP/04750/2013) is funded by the FCT – Fundação para a Ciência e a Tecnologia.

Received for publication: 29 April 2017.
Revision received: 14 July 2017.
Accepted for publication: 17 July 2017.

Geospatial Health 2017; 12:581
doi:10.4081/gh.2017.581

Introduction

In high-income countries, premature mortality has plateaued at very low levels, and, consequently, old-age survival is now the mechanism that regulates life-expectancy (Rossi *et al*., 2013). Old-age survival is then a good general indicator of population health and development (Huisman *et al*., 2004). Although overall mortality variation decreased, survivors have become increasingly heterogeneous with respect to mortality risk, which suggest that mortality inequalities are shifting to older ages as survival in early life improves (Huisman, *et al*., 2004; Engelman *et al*., 2010).

Despite being a relatively small country, spatial inequalities have been identified in Portugal (Santana, 2015), particularly regarding cardiovascular disease (Ferreira-Pinto *et al*., 2012), fractures (de Pina *et al*., 2008), cancer (Alves *et al*., 2016), suicide (Santana *et al*., 2015a), tuberculosis (Apolinário *et al*., 2017) and all-cause mortality (Santana *et al*., 2015b). Moreover, recently, important spatial inequalities in the distribution of old-age sur-

vival have been revealed (Ribeiro *et al*., 2016a, 2016b). A myriad of factors can account for the spatial inequalities in old-age survival that include a complex network of factors of different natures affects population health over the time (Ribeiro, *et al*., 2016b).

Socioeconomic factors may play an important role in explaining these spatial differentials, as the association between socioeconomic position and health is one of the oldest and most solid findings in public health (Mackenbach *et al*., 2008). However, diverse studies have shown that in southern Europe the association between health and socioeconomic deprivation tends to be rather modest compared to western Europe (Mackenbach, *et al*., 2008; Gotsens *et al*., 2013; Borrell *et al*., 2014; Hoffmann *et al*., 2014; Mari-Dell'Olmo *et al*., 2015). It is then crucial to explore the influence of other key health determinants, such as the physical environment (the material features that surrounds population and includes physical, chemical and biological agents external to the human body (Porta, 2001) and access to healthcare, *i.e*. the ability to obtain appropriate health services when needed (Obrist *et al*., 2007).

Regarding, physical environment, there is considerable evidence that the characteristics of the physical environment contribute to extend or shorten life expectancy among older adults (Takano *et al*., 2002; Lv *et al*., 2011; Robine *et al*., 2012). From all age groups, the elderly are certainly the most affected by the impact of climate extremes (Yu *et al*., 2012) and air pollution (Bell *et al*., 2013). It is also important to note that detrimental physical environments are not randomly distributed. Several studies have found that physical and socioeconomic deprivations coincide in space (the so-called environmental injustice) (Fecht *et al*., 2015). Consequently, these two items should be taken into account when addressing health inequalities (Lee, 2002). Similarly, access to healthcare is a vital aspect, especially at older ages, due to the heavy burden of chronic diseases and to higher susceptibility to infections. The likelihood of surviving beyond a certain age is certainly affected by the use of healthcare resources (Vogt and Vaupel, 2015). The importance of healthcare is patent in several European studies, which reveals that the fast increase in old-age survival of the past decades can be mostly attributed to improvements in healthcare (Mackenbach *et al*., 2011; Peters *et al*., 2015; Vogt and Vaupel, 2015).

These intricate relations between population health and socioeconomic, physical and healthcare factors can only be understood with the use of theoretically sound and validated indicators that grasp the multifactorial nature of these influences (Wills and Briggs, 1995). Multivariate ecological indexes of socioeconomic deprivation are becoming common place (Department for Communities and Local Government, 2011; Guillaume *et al*., 2016) as epidemiology research shifts from its traditional biomedical focus to an eco-social approach. Yet, multivariate indexes about the physical environment and access to healthcare (at least specific to older populations) are still uncommon. Only recently has this kind of indexes become available in Portugal. In 2016, a multivariate index of socioeconomic deprivation was created under robust methods and theories (Guillaume *et al*., 2016; Ribeiro *et al*., 2017) and started to be used to study the link between deprivation and health outcomes at the individual level (Antunes *et al*., 2016) and also at the ecological level (Ribeiro *et al*., 2016a). In 2015, a multivariate index of physical environment deprivation (MEDIx) for Portuguese municipalities was developed using sound and internationally validated methodologies (Ribeiro *et al*., 2015). Significant and plausible associations between this measure and mortality were found showing its potential to understand the role of physical environment in diverse health outcomes.

In this study we aimed to evaluate the role of the socioeconomic, physical environmental and healthcare factors on old-age survival in Portugal. We derived a measure of old-age survival for the 278 municipalities of Portugal based on census data. As covariates, we used evidence-based ecological indexes.

Materials and Methods

Study area

The study was conducted in Continental Portugal (which excludes the archipelagos of Madeira and Azores) using municipalities as units of analysis. Municipalities are commonly the smallest unit for health data dissemination and, apart from the large urban areas, they tend to be homogeneous in terms of social and economic profile. Two hundred and seventy eight municipalities exist in Continental Portugal with an average population of 36,143 inhabitants in 2011.

Old age survival

Because life expectancy and mortality data in old ages was not disclosed for municipalities, we derived a measure of old-age survival that expresses the probability of the people aged 75-84 years to survive an additional ten years, *i.e*. surpass the average life expectancy (Ribeiro *et al*., 2016b):

$$r_i = \frac{y_i}{n_i} \quad \text{Eq. 1}$$

where r_i is the ten-year survival rate, i(=1,…,278) the area, y the population aged 85-94 years old in 2011 and n the population aged 75-84 years old ten years before (in 2001). This and similar indicators are straightforward and understandable metrics to estimate survival at advanced ages in small areas, which solely require a time series of population census data (Poulain *et al*., 2004; Magnolfi *et al*., 2007; Ribeiro *et al*., 2016b).

Covariates

Socioeconomic deprivation

The European Deprivation Index (EDI) was used to classify small areas according to their level of socioeconomic deprivation. It was constructed in three steps using both individual and area level census data as has been detailed elsewhere (Guillaume *et al*., 2016). In brief: i) construction of an individual level indicator of deprivation based on the European Union Statistics on Income and Living Conditions (EU-SILC) information, which is available at http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions; ii) identification of variables available both at the individual level (EU-SILC) and at the area level (2001 national population census); and iii) determination, at the individual level, whether the set of area level variables from the census selected at step 2 were associated with the indicators of individual deprivation created in step 1.

The associated census variables were then included in the EDI formula, whose final score was based upon the weighted sum of

these variables. The weights were the regression coefficients measuring the association between the indicator of individual deprivation and the variables from the census that were also available at the individual level identified in step 2. The score for Portugal was based upon the weighted sum of the following variables expressed as percent: overcrowded households; households with no bath or shower; household with no indoor flushing toilets; households occupied by non-owners; women aged ≥65; individuals with low education level; individuals in low income occupations; and individuals unemployed as discussed by Ribeiro *et al*. (2017).

The EDI index was normalised and then classified into ten classes (C_1 – the least deprived to C_{10} – the most). Cut-offs for these theoretical deciles were defined based on standard deviations from the overall mean and customised so that the classes included a more even number of observations (cut-offs = -1.28, -0.84, -0.52, -0.25, 0, 0.25, 0.52, 0.84, and 1.28). This approach avoids the well-known problems of using empirical quintiles, which assume homogeneity of risk within groups (Bennette and Vickers, 2012). The geographic EDI score distribution across Portuguese municipalities is shown in Figure 1 (classes correspond to quintiles instead of deciles to facilitate visualisation).

Access to healthcare

Some measures of access to healthcare have been developed in Portugal, but these were for the overall population (not exclusively for the elderly) and/or they only accounted for hospital services (Polzin *et al*., 2014; Santana, 2015). Therefore we derived an index of access to healthcare for the older population starting by retrieving all datasets on healthcare availability and accessibility (the two domains of healthcare access for which data are available). Variables were obtained at the municipality level for the year 2001 (whenever possible) and for Continental Portugal from two data sources: Hospitals and Primary Care Centers Surveys from Instituto Nacional de Estatística (INE) - National Institute of Statistics (INE, 2001a, 2001b) and Social Map from the Ministry of Solidarity, Employment and Social Security (Carta social, 2008). From 49 datasets, those with too many missing/censored and zero values were discharged (n=16 datasets selected).

We calculated the rates to express the population exposure to the variables created (n=16), which were then characterised and transformed to become more normally distributed. Subsequently, bivariate correlations were computed to identify variables excessively correlated and therefore discharged (n=10 variables selected). Finally, principal component analysis was run to derivate a summary measure expressing access to healthcare in each municipality. The three principal components that explained 72% of the variability in the latent variable access to healthcare were as follows. First, availability of long-term care and social support facilities including the capacities of the day-care centres; the nursing

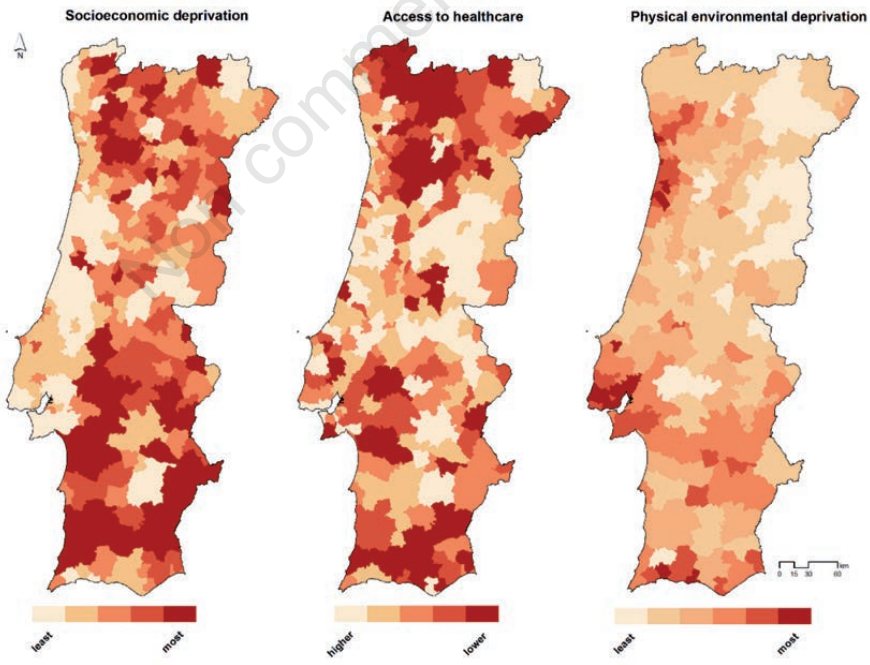


Figure 1. Spatial distribution of socioeconomic deprivation, access to healthcare and physical environmental deprivation in Continental Portugal.

homes; and home care. Second, availability and geographical accessibility to healthcare facilities - population weighted mean distance to public hospitals (maternities and paediatric hospitals excluded) computed using the Network Analyst extension of ArcGIS 10.4.1 (ESRI, Redlands, CA, USA) and a street network dataset provided courtesy of ESRI; primary care centres including extensions; pharmacies; and mobile pharmacy posts. Third, availability of health professionals – medical doctors by place of residence; nurses by place of work; dentists by place of residence; and pharmacists by place of work.

For each municipality, *i*, each component score was multiplied by the proportion of the variation explained.

$$HCA\ score_i = 0.25340 * 1st\ component_i + 0.23857 * 2nd\ component_i + 0.22924 * 3rd\ component_i \quad \text{(Eq. 2)}$$

Similarly to socioeconomic deprivation, after standardisation, the index was categorised into 10 classes based on theoretical deciles. The geographic distribution of the index across Portuguese municipalities is shown in Figure 1 (again classes correspond to quintiles).

Physical environmental deprivation

The measure of multiple physical environmental deprivation (PT-MEDIx) was built at the municipality level using data from the years 2001 and 2011 and developed in four stages fully described elsewhere (Ribeiro *et al*., 2015). The PT-MEDIx covered five dimensions of the physical environment: air pollution (particulate matter, nitrogen dioxide, carbon monoxide); climate (temperature); drinking water quality (trihalomethanes and nitrates); green space availability; and industry proximity. Municipalities in the highest quintile of exposure received a score of +1 for harmful factors and -1 for beneficial factors. The PT-MEDIx of each municipality equalled the sum of these scores and ranged from -1 (the least environmental deprivation) to +4 (the most). We treated all factors as equal contributors to environmental deprivation because any weighting would be arbitrary without robust evidence (Richardson *et al*., 2010; Pearce *et al*., 2011). The geographic distribution of PT-MEDIx across the Portuguese municipalities is shown in Figure 1.

Statistical model

Bayesian Hierarchical Spatial models were used to estimate the effect of each covariate in old-age survival. We assumed that the response variable, number of survivors in each area *i* and gender *j* would (Y_{ij}) follow a binomial distribution, where p_{ij} is an unknown survival rate and n_{ij} the population aged 75-84 years old ten years before:

$$Y_{ij} \sim Bin(n_{ij}, p_{ij}) \quad \text{(Eq. 3)}$$

The logit of the survival rate is modelled considering gender and the interaction between the covariates and gender x_{ij} and area:

$$\text{logit}(p_{ij}) = \eta_{ij} = gender_j + f_j(x_i) + s_i \quad \text{(Eq. 4.1)}$$

where η_{ij} is the linear predictor, $gender_j$ an intercept specific for each gender, $f_j(x_i)$ the gender-specific effect of each covariate (which assumes the value x_i for the area *i*), and s_i the area-specific effect. The function f_i assumes a nonlinear effect of the covariates, which is modelled as a first order random-walk prior over the covariates' classes, *i.e*. a normal distribution, whose mean at each class is an average over the neighbouring classes (Martino and Rue, 2009). As x_i was categorised into classes, we can simplify $f_j(x_i)$ into e_{lj}, which denotes the effect of covariate class *l* for gender *j*.

The area-specific effect s_i was modelled considering a Besag, York and Mollie's (BYM) model (Besag *et al*., 1991) with a parametrisation as suggested by Dean and colleagues (2001):

$$s_i = \tau(\sqrt{\varphi} * u_i + \sqrt{1-\varphi} * v_i) \quad \text{(Eq. 4.2)}$$

where υ_i is the structured effect and v_i the unstructured effect. The υ_i effect was scaled in order to make the model more intuitive and interpretable, so that φ expresses the proportion of the spatial effect due to the structured part and $1/\tau$ is the marginal variance of s_j. A penalised complexity prior was considered.

Considering the model defined by equations 3 and 4.1, the exponential of the gender main effect is the ratio between the men's odds of survival and women's, *i.e*. the gender odds ratio (OR). The exponential of e_{lj} is the ratio between the odds of survival of the covariate class and gender and the overall odds for the entire population – the covariate and gender-specific OR. For example, an OR of 1.15 in a certain covariate class (*e.g*., the least socioeconomically deprived) and gender (*e.g*., women) means that for that gender and covariate class the odds of survival is 15% higher than the overall odds survival of the entire population of that gender. OR and 95% credible intervals (95% CrI) were derived from their posterior means and quantiles. An OR would be considered significantly higher or lower if its 95% CrI does not include the value 1. Posterior distributions were obtained using the Integrated Nested Laplace Approximation (INLA), which was implemented in the R INLA library (Rue *et al*., 2009).

The model produced by equation 4.1 was our final, but we started with a simple model, where we only included the gender-specific intercept and the spatial effect and then introduced each covariate successively. The order of entering each covariate was based on the strength and significance of associations observed in the univariable models. These models were also run to assess the unadjusted influence of each covariate. Three measures of goodness of fit, Deviance Information Criteria (DIC), Watanabe-Akaike information criterion (WAIC) and Conditional Predictive Ordinate, (CPO) were used to compare models. The relative reduction in the variance of spatial effect (τ) was also evaluated to ascertain to what extent covariates contributed to explaining the spatial variation of old-age survival. The presence of interactions between covariates was also tested. Finally, we did explore the correlation between the covariates by computing the Pearson's correlation coefficient (r).

Results

On average, the old-age survival rates were 32.3% (maximum=39.7; minimum=27.5) among men and 43.7% (67.2; 34.5) among women. The presence of spatial inequalities in the distribution of survival is depicted in Figure 2, showing a nearly two-fold difference between areas. In general terms, higher survival rates were concentrated in the North and Central coasts of the country and in the urban municipalities, whereas the lowest were found

concentrated in the South and in the northern inland.

In Figure 3, the survival rates are represented as a function of each covariate. In general, survival rates decreased with socioeconomic deprivation in both genders. On the other hand, survival rates neither seemed to increase nor decrease in a clear fashion according to healthcare access and physical environmental deprivation. We found a significant but moderate correlation between the covariates: physical environmental deprivation was negatively associated with socioeconomic deprivation (r= -0.288, P<0.001) and access to healthcare was also negatively associated with socioeconomic deprivation (r= -0.344, P<0.001).

The results obtained with the univariable and multivariable models are shown in Table 1 (univariable), Table 2 (multivariable, men) and Table 3 (multivariable, women). Among women, in the multivariable model, a rather linear association between old-age survival and socioeconomic deprivation was observed and this association persisted even after the inclusion of the remaining covariates [least deprived areas OR=1.132 (1.064-1.207)] (Table 3). In the null model (no covariates, only spatial effect), the percent variability attributed to the spatial random effect was 65%. We then added one variable each time to assess its impact on old-age survival. The variance attributed to the spatial effect was reduced by 31.4% after including socioeconomic deprivation in the model. The inclusion of the remaining variables did neither cause much change in the adjustment parameters nor in the variance explained by the spatial effect. After adjustment for socioeconomic deprivation the association with healthcare access, which was slightly associated with survival in the univariable model (Table 1), was no longer significantly associated with survival [lowest access to healthcare OR=0.961 (0.901-1.009)]. Similarly, no significant association was found between survival and physical environmental deprivation and no interaction effects were observed between covariates. For men we reached similar results although the magnitude of the associations was considerably smaller.

Socioeconomic deprivation was significantly associated with survival among men [OR=1.044 (1.001-1.094)] (Table 2). In the null model (no covariates, only spatial effect), the percent variability attributed to the spatial random effect was 51%. Compared to what was observed for women, the reduction of spatial variance caused by the introduction of socioeconomic deprivation in men was comparatively smaller (13.1% *vs* 31.4% in women). As with women, neither access to healthcare nor environmental deprivation were significantly associated with old age survival for men.

Discussion

In this study we aimed at investigating the contribution of three important multidimensional determinants on old-age survival in Portuguese municipalities. We found that socioeconomic deprivation was the most relevant factor, explaining a considerable share of the spatial variance in old-age survival, especially among women. Despite evidence showing that physical environment and

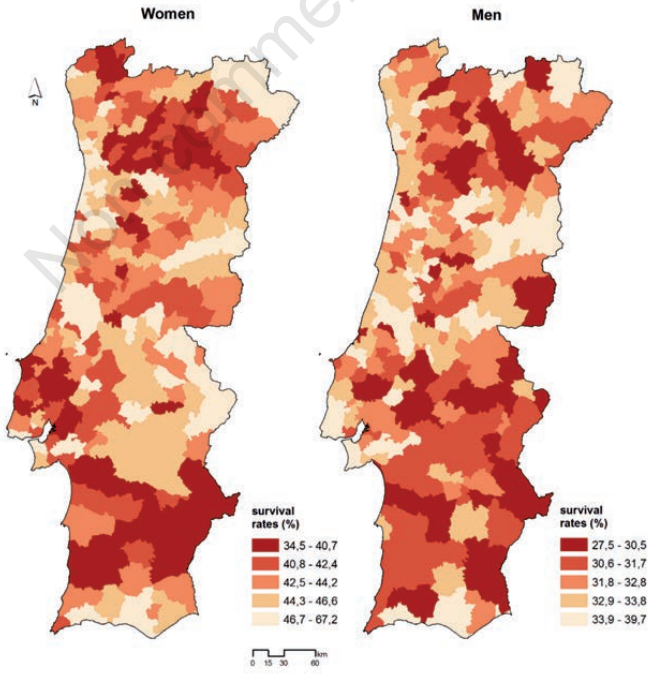


Figure 2. Spatial distribution of old-age survival rates posterior means in Continental Portugal.

healthcare do affect older people's health (Takano *et al*., 2002; Lv *et al*., 2011; Robine *et al*., 2012; Vogt and Vaupel, 2015), those factors did not play such a major influence as socioeconomic deprivation for older people's chances of survival.

To date very few studies have compared the relative importance of factors pertaining to different domains, such as social, economic and physical environmental ones (Domínguez-Berjón *et al*., 2010; Ferreira-Pinto *et al*., 2012; Hood *et al*., 2016). These studies have used indicators (outcomes and covariates) that are not directly comparable to ours, but they all acknowledged that socioeconomic deprivation had the strongest effect. The prominent role of socioeconomic deprivation in shaping a population's health has

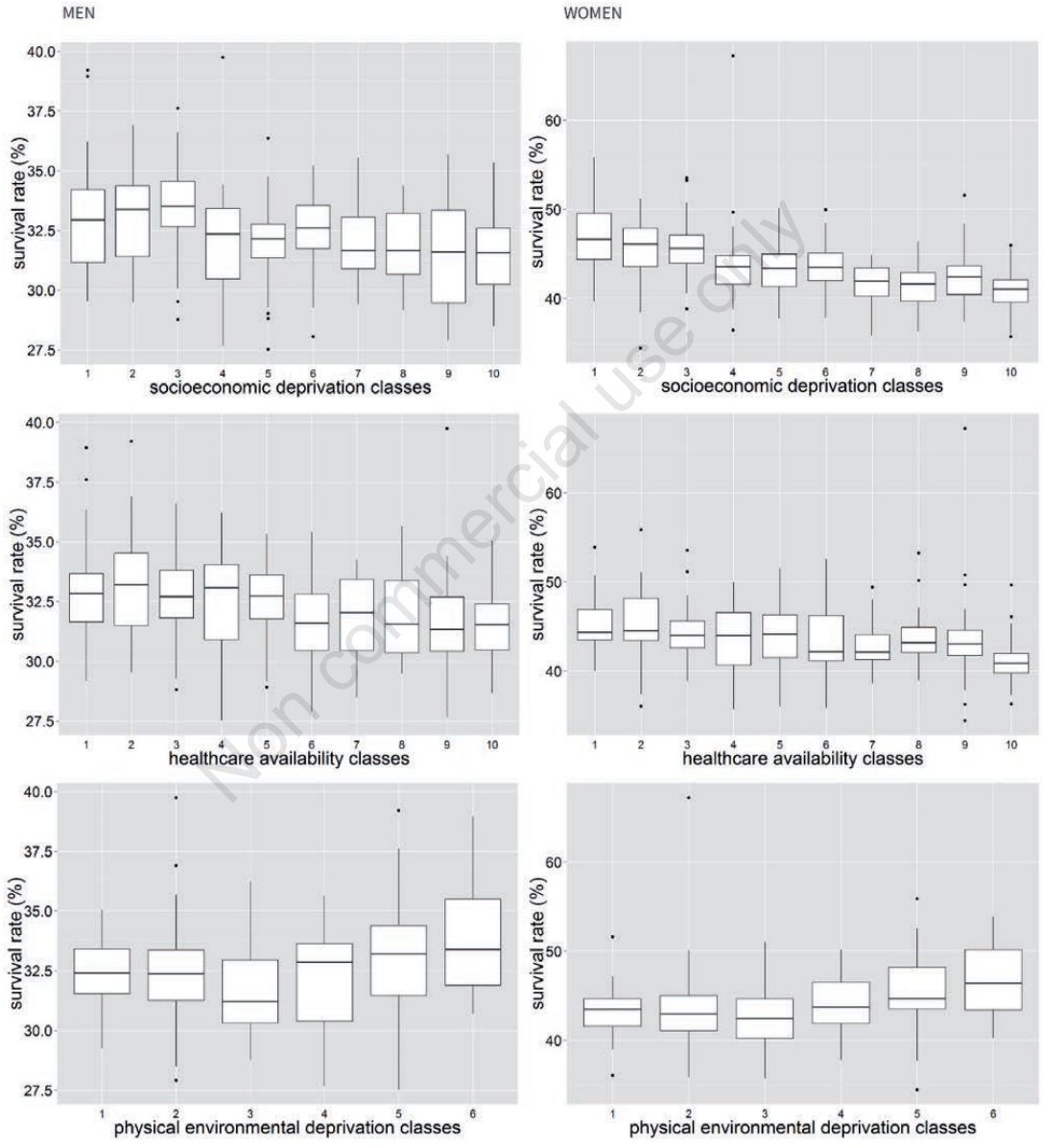


Figure 3. Old-age survival rates posterior means as a function of socioeconomic deprivation, access to healthcare and physical environmental deprivation in Continental Portugal.

been the matter of discussion for centuries and its influence is still observed in our time. For instance, in the United Kingdom and the United States, socioeconomic deprivation seems to explain most of the spatial inequalities in health and life expectancy (Woods *et al*., 2005; Hood *et al*., 2016). The latter authors report the following relative contributions to health in the United States of socioeconomic factors (47%), health behaviour (34%), clinical care (16%) and the physical environment (3%). However, in our study the association between old-age survival and socioeconomic deprivation was significant but its contribution in explaining spatial effects was modest, only 31% among women and 13% in men. It is plausible that other factors, which we did not measure, counterbalanced the effects of socioeconomic deprivation (*e.g*., formal and informal social and economic support).

In the city of Porto in northern Portugal, we conducted a similar study (Ribeiro *et al*., 2016a), where we analysed the impact of the socioeconomic deprivation, built and physical environment on old-age survival across the neighbourhoods of the city. In this study, we found that more than 41% of the differences between neighbourhoods could be attributed to the socioeconomic characteristics of the neighbourhoods. We also found that the measures that described the physical environmental characteristics (built and biogeophysical) of the neighbourhoods were not associated with old-age survival at all. In other words, socioeconomic factors are also the biggest drivers of the spatial differentials in old-age survival in Porto. Although the associations were similar to those we found in the present study that entails the entire country, it is important to point out that the percent of variability explained by socioeconomic deprivation was considerably lower in Portugal as a whole (31% in women and 13% in men) than in Porto. This is likely a consequence of using municipalities as unit of analysis in the present study. These are relatively large areas that might not suffice to detect and capture spatial inequalities and associations whenever there is considerable within-area variability in outcomes and covariates.

Our current study is ecological in nature and, consequently, we could not ascertain causal relations neither the mechanism by which socioeconomic aspects affects survival. There are numerous theories trying to conceptualise that. One of the most relevant is the (neo)material model, which states that most deprived people have poorer health due to lack of material conditions at home and in the living context (work, school, neighbourhood, region) being particularly relevant (Skalická *et al*., 2009). However, our results do not fully accord to this theory. For instance, and contrasting to other studies (Pearce *et al*., 2010), we did not find socioeconomic deprivation to be directly related to physical environmental deprivation expressed by us as a combination of different exposures (green space availability, air/water pollution, climate). Indeed, we observed the exact opposite (negative correlation, r=-0.288, P<0.001) showing taht affluent areas had the poorest physical environmental conditions. But, we did find evidence that healthcare is less available in more deprived areas (negative correlation, r=-0.354, P<0.001) suggesting that some form of environmental unfairness exists in Portugal, as observed in another national study (Nogueira, 2010).

Physical environment did not affect old-age survival in our study, which was also found in the Porto study (Ribeiro *et al*., 2016a). Poor physical environments were concentrated in affluent urbanised areas (where there is a concentration of pollution sources, such as industry and traffic) but the positive influence of having good material resources and facilities (*e.g*., healthcare, jobs, housing conditions, *etc*.) might conceal the detrimental effects of living in a more hazardous environment. The absence of an association with physical environmental aspects might also be attributed to the mortality patterns in very old population strata; among the oldest, the top mortality cause is cardiovascular disease (CVD, responsible for over 40% of the deaths after the 85 years old). In a previous work about the development of PT-MEDIX (Ribeiro *et al*., 2015), we did not find a significant association between physical environmental deprivation and CVD, but, on the other hand, we observed a strong dose-response relation with cancer mortality. Therefore, the impact of physical environment might be modest among the oldest population groups that are less affected by cancer mortality.

Despite the evidence stating that access to healthcare play an important role in extending life expectancy and survival (Mackenbach *et al*., 2011; Peters *et al*., 2015; Vogt and Vaupel, 2015) we found that after controlling for socioeconomic characteristics, access to healthcare was no longer associated with old age-survival. Indeed a dozen of studies have found that, when compared with socioeconomic factors, access to healthcare play a much smaller role (Ferreira-Pinto *et al*., 2012; Kim, 2014; Kim and Kim, 2014). Social and economic conditions adversely affect people's ability to access healthcare (*e.g*., ability to pay for travelling and medical costs) and to understand health information (Hood *et al*., 2016), which may exert a much stronger influence than the

Table 1. Univariable associations among old-age survival and socioeconomic deprivation, access to healthcare and physical environmental deprivation (men).

	OR (95% CrI)	
	Men	Women
Socioeconomic deprivation		
1*	*1.051 (1.008-1.101)*	*1.146 (1.079-1.222)*
2	*1.042 (1.008-1.080)*	*1.087 (1.038-1.138)*
3	*1.040 (1.004-1.086)*	*1.083 (1.031-1.142)*
4	1.004 (0.971-1.035)	1.029 (0.986-1.074)
5	0.974 (0.935-1.007)	0.994 (0.951-1.037)
6	0.991 (0.959-1.026)	0.986 (0.943-1.033)
7	0.991 (0.954-1.031)	*0.937 (0.884-0.986)*
8	0.983 (0.947-1.021)	*0.925 (0.876-0.971)*
9	0.969 (0.931-1.005)	*0.940 (0.894-0.991)*
10°	0.959 (0.915-1.001)	*0.903 (0.850-0.955)*
Access to healthcare		
1ǂ	1.029 (0.993-1.070)	1.036 (0.981-1.094)
2	*1.036 (1.005-1.074)*	*1.055 (1.010-1.107)*
3	1.023 (0.995-1.056)	1.037 (0.994-1.088)
4	1.010 (0.982-1.040)	1.004 (0.957-1.046)
5	1.003 (0.975-1.034)	1.014 (0.971-1.065)
6	0.983 (0.948-1.011)	0.986 (0.938-1.030)
7	0.986 (0.957-1.014)	0.978 (0.934-1.019)
8	0.982 (0.949-1.012)	0.997 (0.952-1.057)
9	0.975 (0.942-1.004)	0.976 (0.934-1.023)
10§	0.976 (0.937-1.012)	*0.923 (0.862-0.982)*
Physical environment		
-1^	0.979 (0.931-1.022)	0.960 (0.901-1.014)
0	0.992 (0.961-1.023)	0.965 (0.923-1.004)
1	0.974 (0.937-1.004)	0.967 (0.922-1.006)
2	0.992 (0.958-1.021)	1.012 (0.974-1.055)
3	1.026 (0.992-1.070)	1.042 (0.996-1.035)
4§	1.039 (0.989-1.106)	1.058 (0.992-1.147)

OR, odds ratio; CrI, credible interval. *Least deprived; °most deprived; ǂhigher; §lower; ^least environmentally deprived; §most environmentally deprived. Statistically significant results are in italics.

availability, quantity and/or geographical accessibility to healthcare (the components of healthcare access captured in our index (Obrist *et al*., 2007) in the Portuguese context.

The main limitation of this study is related with the use of aggregated data. Scale might have influenced our results. We conducted this analysis at the municipality level and this unit can have as few as 1,830 inhabitants in Portugal or hold over 500,000 inhabitants, and 30% of the Portuguese population resides in municipalities with >150,000 inhabitants (INE, 2016). Consequently, we might have failed to detect important associations and inequalities. This may also explain the different proportion of explained variability we observed in Porto (41%) compared with Portugal as a whole (between 13 and 31%). In Porto, we had the opportunity of using a much smaller geographical unit, which is more appropriate for small-area studies of environment and health, since this approach minimises within-area variation, is better to control for potential confounding across areas and captures slight variations in outcomes and covariates (Elliott and Savitz, 2008). However, due to lack of high-resolution data for the entire country, we could only use municipalities as the unit of analysis. The Modifiable Areal Unit Problem (MAUP) is another potential source of bias. A different arrangement of the spatial units might have yielded different results.

Another plausible limitation was that our study was grounded on the assumption that people have lived in the same area during 10 years. However, results from census and original research support the belief that our results are not driven by migration patterns: only about 6% of the Portuguese reported to reside in another geographical unit five years ago and, according to Tatsiramos and colleagues (2006), migration of those aged ≥75 years is infrequent, especially in southern Europe (~1.0 to 1.5%). Moreover, using data from EPIPorto cohort, a landmark epidemiological cohort study in Portugal that has been ongoing for over 15 years (http://ispup.up.pt/research/research-structures/cohorts/), we were able to estimate the frequency of residential mobility in a large Portuguese city. The cohort was constituted in 1999-2003 com-

Table 2. Association among old-age survival and socioeconomic deprivation, access to healthcare and physical environmental deprivation (men).

	Model 1 OR (95% CrI) (socioeconomic deprivation only)	Model 2 OR (95% CrI) (plus healthcare)	Model 3 OR (95% CrI) (plus healthcare and physical environment)
Socioeconomic deprivation			
1*	*1.051 (1.008-1.101)*	*1.043 (1.001-1.092)*	*1.044 (1.001-1.094)*
2	*1.042 (1.008-1.080)*	*1.035 (1.001-1.073)*	*1.035 (1.001-1.073)*
3	*1.040 (1.004-1.086)*	1.034 (0.999-1.079)	1.033 (0.999-1.077)
4	1.004 (0.971-1.035)	1.006 (0.975-1.037)	1.006 (0.976-1.037)
5	0.974 (0.935-1.007)	0.978 (0.939-1.009)	0.977 (0.939-1.008)
6	0.991 (0.959-1.026)	0.991 (0.960-1.023)	0.989 (0.957-1.020)
7	0.991 (0.954-1.031)	0.991 (0.956-1.029)	0.989 (0.954-1.026)
8	0.983 (0.947-1.021)	0.986 (0.952-1.002)	0.987 (0.952-1.022)
9	0.969 (0.931-1.005)	0.973 (0.935-1.008)	0.974 (0.936-1.008)
10°	0.959 (0.915-1.001)	0.967 (0.922-1.008)	0.970 (0.926-1.011)
Access to healthcare			
1#		1.020 (0.988-1.061)	1.021 (0.989-1.060)
2		1.025 (0.997-1.062)	1.024 (0.997-1.060)
3		1.016 (0.991-1.049)	1.016 (0.991-1.047)
4		1.005 (0.978-1.038)	1.005 (0.980-1.033)
5		1.003 (0.977-1.033)	1.002 (0.977-1.030)
6		0.986 (0.951-1.010)	0.987 (0.957-1.014)
7		0.989 (0.960-1.013)	0.988 (0.960-1.012)
8		0.987 (0.955-1.015)	0.987 (0.957-1.014)
9		0.983 (0.950-1.011)	0.984 (0.952-1.011)
10§		0.987 (0.949-1.022)	0.986 (0.950-1.021)
Physical environment			
-1^			0.995 (0.950-1.039)
0			1.001 (0.972-1.032)
1			0.974 (0.938-1.004)
2			0.991 (0.959-1.019)
3			1.017 (0.985-1.058)
4$			1.023 (0.978-1.082)
DIC	2268.58	2269.45	2269.15
WAIC	2263.53	2264.69	2263.71
CPO	-1066.65	-1070.99	-1072.39
Reduction of spatial effect**	13.1	18.7	23.7

OR, odds ratio; CrI, credible interval; DIC, deviance information criteria; WAIC, Watanabe-Akaike information criterion; CPO, conditional predictive ordinate. *Least deprived; °most deprived; #higher; §lower; ^least environmentally deprived; $most environmentally deprived; **percent reduction in the variance of the spatial effect (structured and unstructured). Statistically significant results are in italics.

prising a representative sample of 2,485 adults (≥18 years of age) residing in Porto municipality (Ramos *et al*., 2004). From wave 1 (1999-2003) to wave 2 (2005-2008) of the cohort, a 6-year period, 6.5% (n=162) of the participants changed their neighbourhood of residence, but this percentage was significantly lower among the 75 years olds (n=4; mobility 2.0%). Furthermore, evidence suggests that residential mobility is most likely to cause an underestimation of spatial inequalities and socioeconomic effects (Bryere *et al*., 2015).

Finally, the location of nursing homes might have lower survival among their community residents, and this could have influenced the analysis (Shah *et al*., 2013). Presence of nursing homes, in particular in municipalities, could lead to an underestimation of old-age survival. However, we did not find any match between the spatial distribution of the areas of high and low survival and that of Portuguese nursing homes. Finally, we were not able to evaluate the role of other potentially important aspects, like social support or certain features of the built environment, which might account for the remaining spatial effect that our statistical model was not able to explain. Our study has numerous strengths as well. First, very few studies have dealt with three important determinants of human health and survival: socioeconomic deprivation, physical environment and access to healthcare. More importantly, we have used robust measures that express how advantaged or disadvantaged small areas are in terms of socioeconomic circumstances, physical environment and access to healthcare. These measures were constructed based on sound theories and methods, which allow us to be confident about the study findings. The EDI and PT-MEDIX were built for other countries, with which data our results are internationally comparable. Also, multivariable indexes contribute to a better understanding and monitoring of multidimensional phenomena as they measure the cumulative burden of health detrimental factors at population level (Wills and Briggs, 1995; Corvalán *et al*., 2000). Statistically speaking, we have used robust spatial statistics that allowed us to account for the small number problem and extract the *true* spatial pattern of old-age survival in Portugal. Finally, we also accounted for spatial autocorrelation as we employed a BYM framework to model the spatial effect.

Table 3. Association among old-age survival and socioeconomic deprivation, access to healthcare, and physical environmental deprivation (women).

	Model 1 OR (95% CrI) (socioeconomic deprivation only)	Model 2 OR (95% CrI) (plus healthcare)	Model 3 OR (95% CrI) (plus healthcare and physical environment)
Socioeconomic deprivation			
1*	*1.146 (1.079-1.222)*	*1.135 (1.068-1.210)*	*1.132 (1.064-1.207)*
2	*1.087 (1.038-1.138)*	*1.080 (1.032-1.131)*	*1.080 (1.032-1.129)*
3	*1.083 (1.031-1.142)*	*1.074 (1.025-1.132)*	*1.072 (1.023-1.128)*
4	1.029 (0.986-1.074)	1.032 (0.990-1.077)	1.032 (0.990-1.077)
5	0.994 (0.951-1.037)	0.994 (0.952-1.035)	0.994 (0.952-1.035)
6	0.986 (0.943-1.033)	0.981 (0.939-1.026)	0.978 (0.937-1.022)
7	*0.937 (0.884-0.986)*	*0.940 (0.889-0.987)*	*0.941 (0.890-0.987)*
8	*0.925 (0.876-0.971)*	*0.931 (0.889-0.976)*	*0.933 (0.886-0.978)*
9	*0.940 (0.894-0.991)*	*0.942 (0.897-0.991)*	*0.943 (0.899-0.991)*
10°	*0.903 (0.850-0.955)*	*0.916 (0.862-0.969)*	*0.920 (0.866-0.973)*
Access to healthcare			
1‡		1.020 (0.971-1.073)	1.020 (0.973-1.072)
2		1.032 (0.994-1.083)	1.032 (0.994-1.081)
3		1.021 (0.984-1.070)	1.020 (0.984-1.066)
4		0.994 (0.947-1.032)	0.996 (0.950-1.032)
5		1.014 (0.977-1.067)	1.011 (0.975-1.061)
6		0.987 (0.941-1.025)	0.988 (0.944-1.025)
7		0.979 (0.933-1.015)	0.980 (0.935-1.015)
8		1.004 (0.964-1.062)	1.003 (0.964-1.058)
9		0.993 (0.954-1.038)	0.992 (0.954-1.036)
10§		0.957 (0.896-1.008)	0.961 (0.901-1.009)
Physical environment			
-1^			0.993 (0.945-1.041)
0			0.987 (0.952-1.019)
1			0.972 (0.928-1.006)
2			1.003 (0.971-1.040)
3			1.020 (0.985-1.067)
4§			1.025 (0.974-1.093)
DIC	2481.10	2481.43	2481.53
WAIC	2471.92	2473.54	2472.88
CPO	-1252.19	-1255.65	-1257.34
Reduction of spatial effect**	31.4	33.5	35.6

OR, odds ratio; CrI, credible interval; DIC, deviance information criteria; WAIC, Watanabe-Akaike information criterion; CPO, conditional predictive ordinate. *Least deprived; °most deprived; ‡higher; §lower; ^least environmentally deprived; §most environmentally deprived; **percent reduction in the variance of the spatial effect (structured and unstructured). Statistically significant results are in italics.

Conclusions

Important spatial inequalities in the distribution of old-age survival across Portuguese municipalities were discovered. Socioeconomic deprivation was found to be the most important determinant of old-age survival. However, further studies are needed to identify the unaccounted factors that might explain spatial differentials in old-age survival. Our results suggest policy makers should direct their efforts to tackle socioeconomic differentials between regions and guarantee equitable distribution of the healthcare resources.

References

Alves J, Nunes C, Perelman J, 2016. Socio-economic inequalities in tobacco-related diseases in Portugal: an ecological approach. Publ Health 130:36-42.

Antunes L, Mendonça D, Bento MJ, Rachet B, 2016. No inequalities in survival from colorectal cancer by education and socioeconomic deprivation - a population-based study in the North Region of Portugal, 2000-2002. BMC Cancer 16:608.

Apolinário D, Ribeiro AI, Krainski E, Sousa P, Abranches Mm Duarte R, 2017. Tuberculosis inequalities and socioeconomic deprivation. Int J Tuberc Lung Dis 21:784-9.

Bell ML, Zanobetti A, Dominici F, 2013. Evidence on vulnerability and susceptibility to health risks associated with short-term exposure to particulate matter: a systematic review and meta-analysis. Am J Epidemiol 178:865-76.

Bennette C, Vickers A, 2012. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. BMC Med Res Method 12:21-2.

Besag J, York J, Mollié A, 1991. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 43:1-20.

Borrell C, Mari-Dell'olmo M, Palencia L, Gotsens M, Burstrom BO, Dominguez-Berjon F, Rodriguez-Sanz M, Dzurova D, Gandarillas A, Hoffmann R, Kovacs K, Marinacci C, Martikainen P, Pikhart H, Corman D, Rosicova K, Saez M, Santana P, Tarkiainen L, Puigpinos R, Morrison J, Pasarin MI, Diez E, 2014. Socioeconomic inequalities in mortality in 16 European cities. Scand J Public Health 42:245-54.

Bryere J, Pornet C, Dejardin O, Launay L, Guittet L, Launoy G, 2015. Correction of misclassification bias induced by the residential mobility in studies examining the link between socioeconomic environment and cancer incidence. Cancer Epidemiol 39:256-64.

Carta Social, 2008. Capacidade das respostas sociais. In Gabinete de Estratégia e Planeamento, Ministério da Solidariedade, Emprego e Segurança Social, Lisboa, Portugal.

Corvalán C, Briggs D, Zielhuis G, 2000. Decision-making in environmental health - from evidence to action. E & FN Spon, London, UK.

de Pina MF, Alves SM, Barbosa M, Barros H, 2008. Hip fractures cluster in space: an epidemiological analysis in Portugal. Osteopor Int 19:1797.

Dean CB, Ugarte MD, Militino AF, 2001. Detecting interaction between random region and fixed age effects in disease mapping. Biometrics 57:197-202.

Domínguez-Berjón MF, Gandarillas A, Segura del Pozo J, Zorrilla B, Soto MJ, López L, Duque I, Marta MI, Abad I, 2010. Census tract socioeconomic and physical environment and cardiovascular mortality in the Region of Madrid (Spain). J Epidemiol Commun Health 64:1086-93.

Elliott P, Savitz DA, 2008. Design issues in small-area studies of environment and health. Environ Health Perspect 116:1098-104.

Engelman M, Canudas-Romo V, Agree EM, 2010. The implications of increased survivorship for mortality variation in aging populations. Popul Dev Rev 36:511-39.

Fecht D, Fischer P, Fortunato L, Hoek G, de Hoogh K, Marra M, Kruize H, Vienneau D, Beelen R, Hansell A, 2015. Associations between air pollution and socioeconomic characteristics, ethnicity and age profile of neighbourhoods in England and the Netherlands. Environ Pollut 198:201-10.

Ferreira-Pinto LM, Rocha-Gonçalves F, Teixeira-Pinto A, 2012. An ecological study on the geographic patterns of ischaemic heart disease in Portugal and its association with demography, economic factors and health resources distribution. BMJ Open 2:e000595.

Gotsens M, Mari-Dell'Olmo M, Perez K, Palencia L, Martinez-Beneito MA, Rodriguez-Sanz M, Burstrom B, Costa G, Deboosere P, Dominguez-Berjon F, Dzurova D, Gandarillas A, Hoffmann R, Kovacs K, Marinacci C, Martikainen P, Pikhart H, Rosicova K, Saez M, Santana P, Riegelnig J, Schwierz C, Tarkiainen L, Borrell C, 2013. Socioeconomic inequalities in injury mortality in small areas of 15 European cities. Health Place 24:165-72.

Guillaume E, Pornet C, Dejardin O, Launay L, Lillini R, Vercelli M, Marí-Dell'Olmo M, Fernández Fontelo A, Borrell C, Ribeiro AI, Pina MFd, Mayer A, Delpierre C, Rachet B, Launoy G, 2016. Development of a cross-cultural deprivation index in five European countries. J Epidemiol Commun Health 70:493-9.

Hoffmann R, Borsboom G, Saez M, Mari Dell'Olmo M, Burstrom B, Corman D, Costa C, Deboosere P, Dominguez-Berjon MF, Dzurova D, Gandarillas A, Gotsens M, Kovacs K, Mackenbach J, Martikainen P, Maynou L, Morrison J, Palencia L, Perez G, Pikhart H, Rodriguez-Sanz M, Santana P, Saurina C, Tarkiainen L, Borrell C, 2014. Social differences in avoidable mortality between small areas of 15 European cities: an ecological study. Int J Health Geogr 13:8.

Hood CM, Gennuso KP, Swain GR, Catlin BB, 2016. County health rankings: relationships between determinant factors and health outcomes. Am J Prev Med 50:129-35.

Huisman M, Kunst AE, Andersen O, Bopp M, Borgan JK, Borrell C, Costa G, Deboosere P, Desplanques G, Donkin A, Gadeyne S, Minder C, Regidor E, Spadea T, Valkonen T, Mackenbach JP, 2004. Socioeconomic inequalities in mortality among elderly people in 11 European populations. J Epidemiol Commun Health 58:468-75.

INE, 2001a. Inquérito aos centros de saúde. Instituto Nacional de Estatística, Lisboa, Portugal.

INE, 2001b. Inquérito aos hospitais. Instituto Nacional de Estatística, Lisboa, Portugal.

INE, 2016. População residente (N.º) por local de residência (NUTS - 2013), sexo e grupo etário; anual - INE, estimativas anuais da população residente. Instituto Nacional de Estatística, Lisboa, Portugal.

Kim J, 2014. Association between social factors of health ageing and longevity: determinants of the longevity index (LI) in

OECD Countries. Ageing Int 39:97-105.
Kim JI, Kim G, 2014. Factors affecting the survival probability of becoming a centenarian for those aged 70, based on the human mortality database: income, health expenditure, telephone, and sanitation. BMC Geriatrics 14:113.
Lee C, 2002. Environmental justice: building a unified vision of health and the environment. Environ Health Perspect 110:141-4.
Lv J, Wang W, Li Y, 2011. Effects of environmental factors on the longevous people in China. Arch Gerontol Geriatr 53:200-5.
Mackenbach JP, Slobbe L, Looman CW, van der Heide A, Polder J, Garssen J, 2011. Sharp upturn of life expectancy in the Netherlands: effect of more health care for the elderly? Eur J Epidemiol 26:903-14.
Mackenbach JP, Stirbu I, Roskam AJ, Schaap MM, Menvielle G, Leinsalu M, Kunst AE, 2008. Socioeconomic inequalities in health in 22 European countries. New Engl J Med 358:2468-81.
Magnolfi SU, Petruzzi E, Pinzani P, Malentacchi F, Pazzagli M, Antonini FM, 2007. Longevity index (LI%) and centenarity index (CI%): new indicators to evaluate the characteristics of aging process in the Italian population. Arch Gerontol Geriatr 44:271-6.
Mari-Dell'Olmo M, Gotsens M, Palencia L, Burstrom B, Corman D, Costa G, Deboosere P, Diez E, Dominguez-Berjon F, Dzurova D, Gandarillas A, Hoffmann R, Kovacs K, Martikainen P, Demaria M, Pikhart H, Rodriguez-Sanz M, Saez M, Santana P, Schwierz C, Tarkiainen L, Borrell C, 2015. Socioeconomic inequalities in cause-specific mortality in 15 European cities. J Epidemiol Commun Health 69:432-41.
Martino S, Rue H, 2009. Implementing approximate bayesian inference using integrated nested Laplace approximation: a manual for the inla program. Available from: https://www.math.ntnu.no/~hrue/GMRFsim/manual.pdf
Nogueira HG, 2010. Deprivation amplification and health promoting resources in the context of a poor country. Soc Sci Med 70:1391-5.
Obrist B, Iteba N, Lengeler C, Makemba A, Mshana C, Nathan R, Alba S, Dillip A, Hetzel MW, Mayumana I, Schulze A, Mshinda H, 2007. Access to health care in contexts of livelihood insecurity: a framework for analysis and action. PLoS Medicine 4:e308.
Pearce JR, Richardson EA, Mitchell RJ, Shortt NK, 2010. Environmental justice and health: the implications of the socio-spatial distribution of multiple environmental deprivation for health inequalities in the United Kingdom. Trans Inst Br Geogr 35:522-39.
Pearce JR, Richardson EA, Mitchell RJ, Shortt NK, 2011. Environmental justice and health: A study of multiple environmental deprivation and geographical inequalities in health in New Zealand. Soc Sci Med 73:410-20.
Peters F, Nusselder W, Reibling N, Wegner-Siegmundt C, Mackenbach J, 2015. Quantifying the contribution of changes in healthcare expenditures and smoking to the reversal of the trend in life expectancy in the Netherlands. BMC Public Health 15:1-9.
Polzin P, Borges J, Coelho A, 2014. An extended kernel density two-step floating catchment area method to analyze access to health care. Environ Plann B Plann Design 41:717-35.
Porta M, 2001. A dictionary of epidemiology. Oxford University Press, Oxford, UK.
Poulain M, Pes GM, Grasland C, Carru C, Ferrucci L, Baggio G, Franceschi C and Deiana L, 2004. Identification of a geographic area characterized by extreme longevity in the Sardinia island: the AKEA study. Exp Gerontol 39:1423-9.
Ramos E, Lopes C, Barros H, 2004. Investigating the effect of non-participation using a population-based case-control study on myocardial infarction. Ann Epidemiol 14:437-41.
Ribeiro AI, Krainski ET, Autran R, Teixeira H, Carvalho MS, de Pina MdF, 2016a. The influence of socioeconomic, biogeophysical and built environment on old-age survival in a Southern European city. Health Place 41:100-9.
Ribeiro AI, Krainski ET, Carvalho MS, Pina MdFd, 2016b. Where do people live longer and shorter lives? An ecological study of old-age survival across 4404 small areas from 18 European countries. J Epidemiol Commun Health 70:551-6.
Ribeiro AI, Mayer A, Miranda A, Pina MF, 2017. The Portuguese version of the European deprivation index: an instrument to study health inequalities. Acta Médica Portuguesa 30:17-25.
Ribeiro AI, Pina MdFd, Mitchell R, 2015. Development of a measure of multiple physical environmental deprivation. After United Kingdom and New Zealand, Portugal. Eur J Pub Health 25:610-7.
Richardson EA, Mitchell R, Shortt NK, Pearce J, Dawson TP, 2010. Developing summary measures of health-related multiple physical environmental deprivation for epidemiological research. Environment and Planning A 42:1650-68.
Robine J-M, Herrmann FR, Arai Y, Willcox DC, Gondo Y, Hirose N, Suzuki M, Saito Y, 2012. Exploring the impact of climate on human longevity. Exp Gerontol 47:660-71.
Rossi IA, Rousson V, Paccaud F, 2013. The contribution of rectangularization to the secular increase of life expectancy: an empirical study. Int J Epidemiol 42:250-8.
Rue H, Martino S, Lindgren F, Simpson D, Riebler A, 2009. INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation. R Package Version.
Santana P, 2015. A Geografia da Saúde da População. Evolução nos últimos 20 anos em Portugal Continental. Centro de Estudos em Geografia e Ordenamento do Território (CEGOT), University of Coimbra, Coimbra, Portugal.
Santana P, Costa C, Cardoso G, Loureiro A, Ferrão J, 2015a. Suicide in Portugal: Spatial determinants in a context of economic crisis. Health Place 35:85-94.
Santana P, Costa C, Marí-Dell'Olmo M, Gotsens M, Borrell C, 2015b. Mortality, material deprivation and urbanization: exploring the social patterns of a metropolitan area. Int J Equity Health 14:55.
Shah SM, Carey IM, Harris T, DeWilde S, Cook DG, 2013. Mortality in older care home residents in England and Wales. Age Ageing 42:209-15.
Skalická V, van Lenthe F, Bambra C, Krokstad S, Mackenbach J, 2009. Material, psychosocial, behavioural and biomedical factors in the explanation of relative socio-economic inequalities in mortality: evidence from the HUNT study. Int J Epidemiol 38:1272-84.
Takano T, Nakamura K, Watanabe M, 2002. Urban residential environments and senior citizens' longevity in megacity areas: the importance of walkable green spaces. J Epidemiol Commun Health 56:913-8.
Tatsiramos K, 2006. Residential mobility and housing adjustment of older households in Europe. Available from:

http://ftp.iza.org/dp2435.pdf
Vogt TC, Vaupel JW, 2015. The importance of regional availability of health care for old age survival - Findings from German reunification. Popul Health Metr 13:26.
Wills J, Briggs D, 1995. Developing indicators for environment and health. World Health Stat Quart 48:155-63.
Woods LM, Rachet B, Riga M, Stone N, Shah A, Coleman MP, 2005. Geographical variation in life expectancy at birth in England and Wales is largely explained by deprivation. J Epidemiol Commun Health 59:115-20.
Yu W, Mengersen K, Wang X, Ye X, Guo Y, Pan X, Tong S, 2012. Daily average temperature and mortality among the elderly: a meta-analysis and systematic review of epidemiological evidence. Int J Biometeorol 56:569-81.

Part IV

Educational work

Chapter 7

Tutorial on SPDE models and book chapter

In this chapter we introduce the tutorial about SPDE models in INLA and the book chapter containing four kinds of example on modeling with INLA.

7.1 The R-INLA tutorial on SPDE models

The purpose of writing "The R-INLA tutorial on SPDE models" was to help spread the SPDE modeling approach and to show the flexibility of its implementation in the R-INLA package to build more complex models to analyse data. This was in line with the idea of educating researchers and practitioners on how to use models already developed as a kind of way to prepare them for the more complex SPDE models to be developed. It seems to be useful for practitioners and has become popular, with more than 10,000 downloads in 2016.

Development of this tutorial began in March, 1 2013 with a toy simulated example and an example looking at rainfall data in the Paraná state in Brazil. Then we started adding more complex examples to provide sample code for doing similar analysis in published papers and others to answer questions posted in the r-inla mailing list or privately. The case of modeling misaligned data and hurdle Gamma were taken out from this tutorial and published as part of the book Chapter presented in the next Section. The current version is structured into five Chapters.

Next, we include its abstract and table of contents.

Abstract

This tutorial will show you how to fit models that contains at least one effect specified from an SPDE using the 'R-INLA'. Up to now, it can an SPDE based model can be applied to model random effects over continuous one- or two- dimensional domains. However, the theory works for higher dimensional cases. The usual application is data whose geographical location is explicitly considered in the analysis. This tutorial explores 'R-INLA' functionalities by using examples. It starts with simple models and increases in complexity.

In Chapter 1 includes a section introducing the random field models and the Matérn class. We illustrate some features of this class in figures. The we introduce the main results in [Lindgren et al., 2011] intuitively linking to images of the matrices involved being computed for a illustrative small case with few spatial locations. We show how to fit a geostatistical model for a simulated data, the toy example, covering from the mesh building, model definition, data preparation, showing results, doing predictions and considering results from different meshes. We also show how to build a mesh considering non-convex domains, spatial polygons objects and domains with holes or physical boundaries.

In Chapter 2 we consider three examples. We consider the daily average rainfall from rainfall collected at 616 gauge stations in the Paraná state in Brasil over year 2011. For this data we show a detailed analysis including code to compute geographical covariates, smoothed regression an prediction. The second example in this Chapter consider survival analysis for the Leukaemia dataset, analysed in [Henderson et al., 2003]. We show how to consider the parametric Weibull case and also the non-parametric Cox proportional hazard case. For this case we have the implementation internally considers a new structure of the data in order to perform a Poisson regression. The last example in this Chapter considers simulated data to illustrate the approach of modeling the SPDE model parameters by a regression which is the case of having covariates in the covariance, proposed in [Ingebrigtsen et al., 2014].

In Chapter 3 we have a collection of examples were copy random fields to model two or more outcomes jointly. It includes a measurement error model in order to account for spatially structured measurement error in a covariate. A coregionalization model consider the case for three outcomes were the fist outcome is in the linear predictor for the second one and both are in the predictor for the third outcome, as proposed in [Schimdt and Gelfand, 2003]. An example considering copying a part or the entire linear predictor from one outcome in a linear predictor to another one ends this chapter. It shows a slight different way from the coregionalization model to jointly model three outcomes.

The log Cox point process model is considered in Chapter 4. In this case we show how to fit a Log-cox point process using the direct approximation for the likelihood as proposed in [Simpson et al., 2016]. We also take the opportunity to show how to consider the joint modeling of the process and the locations, under the preferential sampling as proposed in [Diggle et al., 2010].

Finally, Chapter 5 presents several cases to example analysis of space-time data. We start by an example having discrete time domain as in [Cameletti et al., 2012]. We extend this example considering the time as continuous by considering time knots along with temporal function basis functions for projection. We also extend the coregionalization example for the space-time in this Chapter. The space-time model is also applied for modeling regression coefficients in a dynamic regression example having the regression coefficients varying over space-time. We consider the space-time version of the log-Cox point process model for a dataset and also illustrates an approach to deal whit the case of having a large space-time point process data.

Since this tutorial is more a collection of examples, one should start with the tutorial marked as **Read this first!** at the tutorials link in the R-INLA web page, `http://www.r-inla.org`, more precisely at `http://www.r-inla.org/examples/tutorials/spde-tutorial-from-jss`. If you are in a rush to fit a simple geostatistical model, we made a short tutorial without the details as a vignette in the **INLA**. Thus one can have it just typing `vignette(SPDEhowto)` for a two dimensional example or `vignette(SPDE1d)` for a one dimensional example. We built a Shiny application to help one to understand the mesh building. It depends on the **shiny** package. This application opens by typing `demo(mesh2d)`.

Contents

7.2 Book chapter "Advanced modeling"

The purpose of this contribution was to publish two examples from an earlier versions of the SPDE tutorial and other examples on space-time models for areal data. We end up expanding the example dealing with misaligned data into one for Gaussian likelihood and the other for non-Gaussian. These examples consider two correlated SPDE models considering the copy strategy available in **INLA** to model multivariate outcomes with a common shared model. In the example, either the semicontinuous model or the hurdle Gamma model is applied to modeling daily rainfall using the copy strategy giving one SPDE model shared between two linear predictors, one for the rain occurrence, considered as a Bernoulli process, and the other for the positive rain amount, considered as a Gamma process of the continuous valued outcome. The examples considering areal data were two on dynamic space-time models one considering the group strategy in INLA to build Kronecker precision models and other considering the data augumented strategy. The last example considers the case when one wants to build a space-time model considering knots over time in order to lower the time resolution.

This contribution is the 8th Chapter in Blangiardo and Cameletti (2015). The cover of this book is shown in Figure 7.2. We present the table of contents, the preface and this chapter in the next pages.

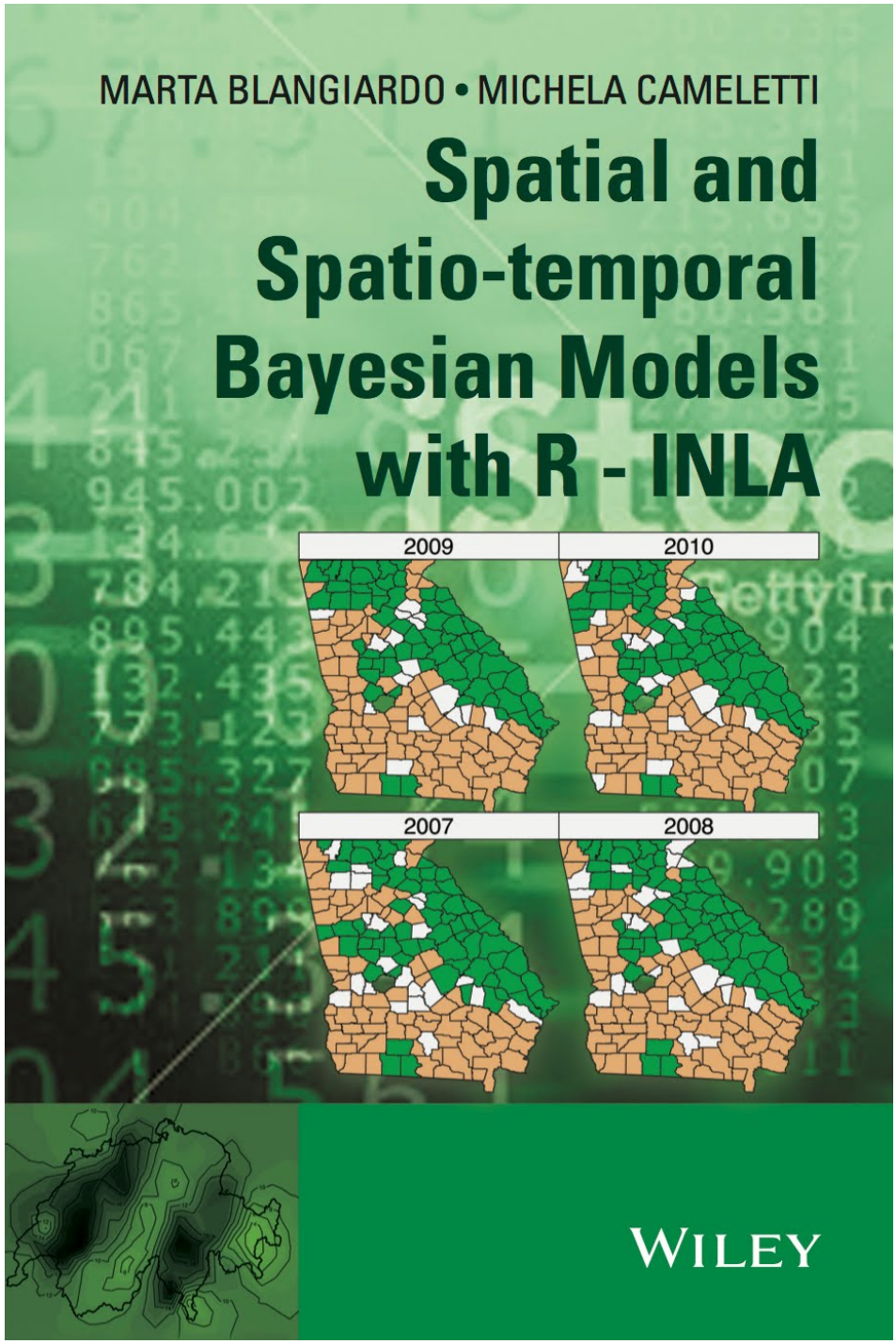


Figure 7.1: Cover of the "Spatial and Spatio-temporal Bayesian Models with R-INLA" book

Contents

Preface

This book presents the principles of Bayesian theory for spatial and spatio-temporal modeling, combining three aspects: (1) an introduction to Bayesian thinking and theoretical aspects of the Bayesian approach, (2) a focus on the spatial and spatio-temporal models used within the Bayesian framework, (3) a series of practical examples which allow the reader to link the statistical theory presented to real data problems. All the examples are coded in the `R` package `R-INLA`, and based on the recently developed integrated nested Laplace approximation (INLA) method, which has proven to be a valid alternative to the commonly used Markov Chain Monte Carlo (MCMC) simulations.

The book starts with an introduction in Chapter 1, providing the reader with the importance of spatial and spatio-temporal modeling in several fields, such as social science, environmental epidemiology, and infectious diseases epidemiology. We then show why Bayesian models are commonly used in these fields and why we focus on the INLA approach. We also describe the datasets which will be used in the rest of the book, providing information on the topics that will be used as illustration.

As all the examples are run in `R`, in Chapter 2 we introduce the basic concepts of the `R` language. Chapter 3 describes the Bayesian methods: first we introduce the paradigms of this approach (i.e., the concepts of prior and posterior distributions, Bayes theorem, conjugacy, how to obtain the posterior distribution, the computational issues around Bayesian statistics for conjugated and non conjugated models). We also include a small section on the differences between the frequentist and the Bayesian approach, focusing on the different interpretation of confidence intervals, parameters, and hypothesis testing.

Chapter 4 discusses the computational issues regarding Bayesian inference. After the Monte Carlo method is introduced, we consider MCMC algorithms, providing some examples in `R` for the case of conjugated and non conjugated distributions. The focus of the chapter is the INLA method, which is a computationally powerful alternative to MCMC algorithms. In particular, the `R-INLA` library is described by means of a small tutorial and of a step-by-step example.

Then in Chapter 5 we present the Bayesian modeling framework which is used in the fields introduced in Chapter 1 and focuses on regression models (linear and generalized linear models). In this context, we introduce the concept of exchangeability and explain how this is used to predict values from variables of interest, a

topic which will be expanded later in the chapters on spatial and spatio-temporal modeling. The last section of this part is devoted to introducing hierarchical models.

Chapter 6 focuses on models for two types of spatial processes: (1) area level—introducing disease mapping models and small area ecological regressions (including risk factors and covariates) and then presenting zero inflated models for Poisson and Binomial data; (2) point level—presenting Bayesian kriging through the stochastic partial differential equations (SPDE) approach and showing how to model observed data and also to predict for new spatial locations. Chapter 7 extends the topics treated in Chapter 6 adding a temporal dimension, where we also include the time dimension in the models.

Finally, Chapter 8 introduces new developments within INLA and focuses on the following advanced applications: when data are modeled using different likelihoods, when missing data are present in covariates, a spatio-temporal model with dynamic evolution for the regression coefficients, and a spatio-temporal model for high-frequency data on time where a temporal resolution reduction is needed.

We would like to thank many people who helped with this book: Håvard Rue for his precious contribution, his endless encouragement and for introducing us to Elias Krainski, who became involved in the book; Finn Lindgren, Aurelie Cosandey Godin and Gianluca Baio for reading drafts of the manuscript and providing useful comments; Philip Li, Ravi Maheswaran, Birgit Schrödle, Virgilio Gómez-Rubio, and Paola Berchialla, who provided some of the datasets; finally, a huge thank to our families who have supported us during all this time.

We hope that this book can be helpful for readers at any level, wanting to familiarize or increase their practice and knowledge of the INLA method. Those who are approaching the Bayesian way of thinking for the first time could follow it from the beginning, while those who are already familiar with `R` and Bayesian inference can easily skip the first chapters and focus on spatial and spatio-temporal theory and applications.

Marta Blangiardo and Michela Cameletti

Is not included due to copyright