Bayesian spatio-temporal hierarchical modeling:

Bycatch in the Barents Sea shrimp fishery and North Atlantic windiness

Olav Nikolai Breivik

Dissertation presented for the degree of
Philosophiae Doctor (PhD)

Department of Mathematics
University of Oslo
August 2016

© Olav Nikolai Breivik, 2016

*Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 1788*

ISSN 1501-7710

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Acknowledgments

I started working on my thesis in August 2012 at the University of Oslo (UiO) with Ida Scheel as my main supervisor, and Bent Natvig and Arne Bang Huseby as co-supervisors. When Ida Scheel went on maternity leave, Geir Storvik also came in as a co-supervisor. I have enjoyed and am very grateful for their help and support. The collaboration with Arne has been limited in the thesis, but I have enjoyed publishing two papers with him within the field of risk analysis that are not included in this thesis.

I am also very grateful to Erik Vanem. He finished his PhD in statistics at the University of Oslo in 2013 and we collaborated in the beginning of my PhD and wrote one article together about spatio-temporal modeling of long term wind speed changes in the North Atlantic. When I started my PhD I had not done any research on spatial or spatio-temporal statistics before, and our collaboration in the end of his PhD project was of great help.

Fortunately, I have enjoyed help from people with expertise in the biological field of my research. My main project has been dedicated to bycatch in the Barents Sea shrimp fishery, and especially I want to thank Kjell Nedreaas. He is working as a scientist at the Institute of Marine Research (IMR) in Bergen, Norway, and his help has been important for understanding the biological aspects of our research. Furthermore, I would like to thank the people working as inspectors at the Norwegian Directorate of Fisheries Monitoring and Surveillance Service for both discussions and for collecting the data used in the three first papers in this thesis. It is they who have the best intuition about bycatch, and for four days I had the opportunity to discuss my research with one of the inspectors, Frank Kristoffersen, while we were trawling after shrimps in the North of Norway.

I want to thank my family for always backing me up and believing in me. For this I am very lucky. The main part of this thesis is dedicated to fishery in the far north, and my passion for fishery goes back to my childhood when I every summer went out fishing in the North of Norway with my grandfather, uncle, mother, father and cousins. Looking back, I am very grateful for those experiences which lead to motivation to work on my project about fishery in the Barents Sea.

List of papers

Paper I

Breivik, O. N., Storvik, G., and Nedreaas, K. (2016). Latent Gaussian models to decide on spatial closures for bycatch management in the Barents Sea shrimp fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(8): 1271-1280.

Paper II

Breivik, O. N., Storvik, G., and Nedreaas, K. (2017). Latent Gaussian models to predict historical bycatch in commercial fishery. *Fisheries Research*, 185: 62-72.

Paper III (Technical report)

Modeling excess zero count data using R-INLA, applied to bycatch of cod and redfish in the Barents Sea shrimp fishery

Paper IV

Vanem, E., Breivik O. N. (2013). Bayesian hierarchical modeling of North Atlantic windiness. *Natural Hazards and Earth System Sciences*, 13(3):545-557.

Contents

Acknowledgments	i
List of papers	ii
1 Introduction	1
2 Methodology	3
2.1 Model construction	3
2.1.1 Latent correlation structures	3
2.2 Inference	5
2.2.1 Bayesian inference	6
2.3 Model selection and validation	9
2.3.1 Cross-validation with a K-fold procedure	9
2.3.2 Deviance information criteria (DIC)	10
2.3.3 Bayes factor	10
3 Background: Application and motivation	11
3.1 Bycatch in the Barents Sea shrimp fishery	11
3.1.1 Monitoring of the shrimp fishery	11
3.1.2 Historical bycatch prediction	13
3.2 Wind speed in the North Atlantic	14
4 Summary of papers	15
4.1 Paper I	15
4.1.1 Effects found important for prediction	15
4.1.2 Management	16
4.2 Paper II	16
4.3 Paper III (Technical report)	17
4.4 Paper IV	18
5 Discussion	19
5.1 Continuous spatio-temporal modeling with R-INLA	19
5.1.1 Smoothed temporal AR(1)-structure	20
5.2 Cleaning of the survey data	22
5.3 Usage of the bycatch model for regulation	23
References	24
Papers I-IV	29

1 Introduction

Randomness constantly shapes the world we live in, and we strive to understand relations between observations. Observations are often naturally ordered in space and time, and in such situations spatio-temporal statistics is essential to give scientific insight into hidden structures in the data. In the thesis we will illustrate two applications of spatio-temporal statistics. Paper I, paper II and paper III are dedicated to bycatch in the Barents Sea shrimp fishery, which is the main topic of the thesis. Paper IV is about long term wind speed changes in the North Atlantic. In both of these applications it is reasonable to assume latent spatio-temporal correlation structures, and incorporating these correlation structures is crucial for understanding the uncertainty, making correct conclusions and performing predictions.

Observations close to each other tend to be more similar. For example people living close to each other tend to have the same dialect, and dialects tend to differ between generations. When including dependence structures in space and time in our statistical analysis, we enter the area of spatial and spatio-temporal statistics. Such structures, when present, might be extremely important to include in the analysis. Paper II illustrates that a prediction procedure without the inclusion of spatio-temporal correlation structure underestimates the uncertainty in the specific application.

Inference based on models with latent correlation structures is typically computationally costly. With the growing computational power, many flexible and general methods for modeling random structures have therefore been developed or become increasingly popular. Gaussian random fields are of great importance for spatio-temporal statistics since they can represent fundamental underlying structures between observations, while at the same time hold practical theoretical results which are utilized for fast computation. In the thesis we will explore the use of Gaussian random fields to approximate underlying correlation structures that give a better insight into the problem at hand.

In the thesis we introduce Bayesian spatio-temporal hierarchical models and estimate our hierarchical spatio-temporal models by using Markov chain Monte Carlo (MCMC) or integrated nested Laplace approximations (INLA) (Rue et al., 2009; Martins et al., 2013). For a model to be both computationally feasible and well specified, relatively few parameters that are computationally hard to estimate should be included. At the same time, those parameters must be able to capture important structures describing relations of interest. Important parameters in the model are often correlated, and such correlation complicates the joint estimation of the posterior distribution. One advantage of INLA is its efficient exploration of the posterior of correlated parameters. Creating a method which use the INLA-technique from scratch typically induces more work compared to creating a MCMC, but fortunately there are user friendly packages implemented in R (R Core Team, 2014), which can perform approximations of a wide specter of latent Gaussian models. We will use the R-INLA (<http://www.r-inla.org>) (Rue et al., 2009; Martins et al., 2013) package.

1. INTRODUCTION

The three first articles concerning bycatch are written for a broad specter of readers interested in marine science and/or statistics. Most of the questions answered in these papers are easy to understand. For example the main question in the first paper is “What is the bycatch rate of shrimp trawling in a certain area at a certain time?”, and in paper II the main question is “What is the yearly bycatch of commercial shrimp trawlers in the Barents Sea?”. To make the proposed answers easy to follow and to understand, the results and parameters are elaborated with possible biological interpretations.

2 Methodology

Statistical models utilize data to give scientific insight into areas of applied research. In this thesis insight is achieved by constructing Bayesian spatio-temporal models which utilize rough intuitive structures in space and time. Statistical modeling in the thesis consists mainly of three general steps which are elaborated in this chapter. First a model is constructed. The model, or a set of different models, are constructed by assuming simplifying structures, and expertise about the problem at hand is preferably incorporated for modeling structures in the data. Secondly the parameters are estimated with an inference procedure, and it is important that the inference procedure is efficient, in particular when the model is complex. Thirdly *one* model is selected and evaluated. When constructing the model, many simplifying assumptions are needed, and the validation part is important for justification of the assumptions and the conclusions.

2.1 Model construction

Generalized linear regression models are frequently used to understand relations between observations. A generalized linear regression model is on the form

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} \quad (2.1)$$

$$\mathbf{y} \sim f(\cdot; \boldsymbol{\eta}), \quad (2.2)$$

where $\mathbf{y} = \{y_1, \dots, y_n\}$ is a vector of the responses, $f(\cdot; \boldsymbol{\eta})$ is the data distribution, $\boldsymbol{\eta}$ is the linear predictor, $\mathbf{X}$ is the design matrix, $\boldsymbol{\beta}$ is a vector of the regression coefficients, and $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_n\}$ is a vector of random variables.

All the four papers in this thesis use latent Gaussian generalized linear models (2.1), i.e. that $\boldsymbol{\gamma}$ is multivariate Gaussian distributed. Observations close to each other are typically more similar. If such an existing structure is not modeled satisfactory through the explanatory variables in (2.1), it can be included as latent structures in $\boldsymbol{\gamma}$ through correlation functions.

2.1.1 Latent correlation structures

The random part, $\boldsymbol{\gamma}$, in (2.1) explains variations which are not explained by the explanatory variables and the regression coefficients. It represents something that we do not directly observe, thereof the name latent structures. By assuming certain properties of these latent structures, we are able to model them. In this thesis we assume that the random terms, $\boldsymbol{\gamma}$, are Gaussian and model the latent dependence structures as a sum of latent *Gaussian random fields*.

Definition 2.1.1. A stochastic process $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbf{D} \subset \mathbb{R}^d\}$ is a Gaussian random field if and only if for any $k \geq 1$ and set $\mathbf{s}_1, \dots, \mathbf{s}_k \in \mathbf{D}$, $(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_k))$ is multivariate Gaussian distributed.

In spatio-temporal statistics, the dimension of the Gaussian field, d , in definition 2.1.1 is typically equal to 1, 2 or 3, and referring to dimensions in time and space. The correlation structures in the Gaussian random field included in (2.1) are selected such that they can capture realistic spatial, temporal and spatio-temporal structures in the data.

For computational efficiency, conditioned independence assumptions, i.e. Markov assumptions, can be made on the Gaussian random field by defining *neighborhood structures*.

Definition 2.1.2. Let $\{Z(s_1), \dots, Z(s_n)\}$ be a Gaussian field, then s_k is a neighbor of s_i if the conditional distribution of $Z(s_i)$, given all the other site values, depends functionally on $Z(s_k)$. Also define

$$N_i = \{k : k \sim i\}, \quad (2.3)$$

where $k \sim i$ means that k is a neighbor of i .

A Gaussian random field with a neighborhood structure has the Markov property due to the conditional independence, and is therefore called a Gaussian Markov random field (GMRF).

The time efficiency induced by the introduction of a neighborhood structure is caused by the following result (see e.g. Rue & Held (2005) for a proof):

Theorem 2.1.1. Let $\mathbf{Q}$ be the precision matrix of a given Gaussian random field with elements Q_{ij} . Then $Q_{ij} = 0$ if and only if $j \notin N_i$.

The sparseness of the precision matrix depends therefore on the complexity of the neighborhood structure. A sparse precision matrix results in the ability to use fast matrix calculation algorithms with low memory usage for Gaussian random fields. This is discussed further in section 2.2.1.

Separable covariance functions are often used in spatio-temporal statistics for computational convenience.

Definition 2.1.3. A stochastic process $\{Z(\cdot; \cdot)\}$ is said to have a separable spatio-temporal covariance function if, for all $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^2$, $t_1, t_2 \in \mathbb{R}$, we obtain

$$\text{cov}\left(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)\right) = \text{cov}^s(\mathbf{s}_1, \mathbf{s}_2) \text{cov}^t(t_1, t_2), \quad (2.4)$$

where $\text{cov}^s(\cdot)$ and $\text{cov}^t(\cdot)$ are spatial and temporal covariance functions, respectively.

If $\mathbf{C}^{s,t}$ is a separable covariance matrix in space and time, we see directly from definition 2.1.3 that

$$\mathbf{C}^{s,t} = \mathbf{C}^s \otimes \mathbf{C}^t, \quad (2.5)$$

where $\mathbf{C}^s$ and $\mathbf{C}^t$ are spatial and temporal covariance matrices and $\otimes$ is the Kronecker product. By standard matrix calculations the precision matrix is then given by $\mathbf{Q}^{s,t} = \mathbf{Q}^s \otimes \mathbf{Q}^t$, where $\mathbf{Q}^s$ and $\mathbf{Q}^t$ are the precision matrices in space and time. By including neighborhood structures in both space and time, it is seen from Theorem 2.1.1 that the separable spatio-temporal covariance structure inherits a joint version of the neighborhood structure in space and time.

2.2. Inference

For simplicity, *stationary* and *isotropic* assumptions are often made on covariance functions.

Definition 2.1.4. A spatio-temporal covariance function is called *stationary* if it only depends on the distance vectors in time and space. It is further called *isotropic* if it only depends on the euclidian distances, that is there exists a function C such that

$$\text{cov}\left(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)\right) = C\left(\|\mathbf{s}_1 - \mathbf{s}_2\|, |t_1 - t_2|\right). \quad (2.6)$$

The isotropic assumption might not be realistic in applications if the underlying physical process changes in space and time. Note that if the isotropic assumption is not realistic, there might exist a linear transformation of the locations in the random field which makes the isotropic assumption on the transformed random field realistic (page 128 Cressie & Wikle, 2011).

A Gaussian field with neighborhood structure as in definition 2.1.2 is defined on a given set of locations. Observations are often in continuous time and space, and preferred underlying correlation structures are often continuous. However, modeling of a continuous correlation structure by defining the latent locations at every observation location may be inconvenient for computational reasons. There are procedures to utilize the sparse precision structure of a GMRF defined on a grid for modeling observations in continuous time and space. For example the area of interest can be divided into sub-areas which are defined as the locations in the GMRF, and for simplicity the latent effect can be assumed constant within the sub-areas. To accommodate for variations within sub-areas, the locations in the GMRF can be defined as the vertices enclosing the sub-areas, and the random values inside a sub-area can be defined as a linear combination of the GMRF values at the locations which enclose the sub-area (Lindgren et al., 2011; Banerjee et al., 2008). Banerjee et al. (2008) introduced the *predictive process* which is defined as such a linear combination with weights given as a function of distances from the locations in the GMRF and the correlation function. However, one must be cautious when defining the locations in the GMRF. The distances from the observed locations to the locations in the GMRF inflict the marginal variance due to the linear combinations (Banerjee et al., 2008), which results in an overestimated nugget effect. This topic will be further discussed and elaborated in subsection 5.1.1.

2.2 Inference

After a parametric model has been constructed, data is used to estimate the unknown parameters in the model. Classic text book inference procedures, such as the least square method and the maximum likelihood method, have intuitive and easy to understand results, e.g. estimates of the uncertainty of the parameters in the model (Casella & Berger, 2002; Devore & Berk, 2007). Most of these classic results are based on the assumption of independent observations. However, the consequences of violating the independence assumption are not always easy to understand, and thereby may such procedures be used too often without critical discussion. In this thesis Bayesian inference is used for modeling dependence structures and for estimating the parameters.

2.2.1 Bayesian inference

In a Bayesian framework we let the parameters, θ , be random variables and introduce a prior belief on them. Ideally, this prior belief should be stated before looking at the data to avoid using data twice in the inference. By Bayes theorem (Bayes & Price, 1763), the joint posterior distribution of the parameters and the latent variables is given by

$$p(\theta, \mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\theta, \mathbf{x})p(\theta, \mathbf{x})}{p(\mathbf{y})}, \quad (2.7)$$

where $\mathbf{x}$ is the latent variables, $p(\mathbf{y}|\theta, \mathbf{x})$ is the likelihood, $p(\theta, \mathbf{x})$ is the prior, $p(\mathbf{y})$ is the normalizing constant and $\mathbf{y}$ is the observations. Posterior estimates are obtained through combining the posterior distribution with a loss function. The most common point estimates are given by the posterior mean, median or mode. These estimates are equivalent to the Bayesian estimators obtained with L_2 , L_1 and 0 – 1 loss functions, respectively. Credibility intervals are then constructed from the posterior distribution to give insight into the uncertainty.

The computational complexity of calculating the posterior (2.7) depends on the complexity of the model. By using a model with a prior distribution which is conjugate to the likelihood, an explicit analytically formula for the posterior distribution can be derived and used for inference. In many practical applications, such models are not satisfactory, and calculation or approximation of the posterior is done numerically. However, the calculation of $p(\mathbf{y})$ is, in relatively complex models, infeasible by numerical integration. Fortunately $p(\mathbf{y})$ is just a constant with respect to the unknown parameters of interest, and there exist methods for estimating the posterior distribution without evaluating $p(\mathbf{y})$.

Latent random variables are included to model dependence structures. The number of latent variables included and their correlation structures are crucial for the computational complexity of the model. By defining these correlation structures through sparse precision matrices as described in section 2.1.1, we are able to use common efficient algorithms for inference (Rue, 2001; Rue & Held, 2005).

Markov Chain Monte Carlo

The dominant methodology for estimating the posterior distribution is by sampling using Markov Chain Monte Carlo (MCMC) methods. A common approach is to combine the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) and the Gibbs sampler (Geman & Geman, 1984). Typically a Gibbs sampler is constructed with Metropolis-Hastings steps. Suppose we want to sample from $\mathbf{u}|\mathbf{y}$, and $\mathbf{u}$ is divided into k different components or sub-vectors, $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_k)$. Such a situation naturally rises in Bayesian hierarchical modeling, where the joint posterior is factorized in more simple terms. Each sample of the Gibbs sampler cycles through the sub-vectors of $\mathbf{u}$ and samples conditional on each other. For each iteration, $t = 1, \dots, B$, each $\mathbf{u}_j^{(t)}$ is sampled from

$$\pi(\mathbf{u}_j^{(t)}|\mathbf{u}_{-j}^{(t)}, \mathbf{y}), \quad (2.8)$$

2.2. Inference

where $\mathbf{u}_{-j}^{(t)}$ represents the current values of $\mathbf{u}$ except for $\mathbf{u}_j^{(t)}$:

$$\mathbf{u}_{-j}^{(t)} = (\mathbf{u}_1^{(t)}, \dots, \mathbf{u}_{j-1}^{(t)}, \mathbf{u}_{j+1}^{(t-1)}, \dots, \mathbf{u}_k^{(t-1)}). \quad (2.9)$$

The Gibbs sampler assumes accessible forms on the conditional posteriors (2.8). If this is not the case for all the parameters, a Metropolis-Hastings step can be used for sampling. The basic Metropolis-Hastings step is as follows: A proposal for a new sample, $\mathbf{u}_{0,j}^{(t)}$ of $\mathbf{u}_j$ is selected from a proposal function $q(\mathbf{u}_{0,j}^{(t)}|\mathbf{u}_j^{(t-1)})$. The proposal is then selected as the new sample $\mathbf{u}_j^{(t)}$ with probability

$$\alpha(\mathbf{u}_{0,j}^{(t)}|\mathbf{u}_{-j}^{(t)}, \mathbf{y}) = \min \left(1, \frac{\pi(\mathbf{u}_{0,j}^{(t)}|\mathbf{u}_{-j}^{(t)}, \mathbf{y})q(\mathbf{u}_j^{(t-1)}|\mathbf{u}_{0,j}^{(t)})}{\pi(\mathbf{u}_j^{(t-1)}|\mathbf{u}_{-j}^{(t)}, \mathbf{y})q(\mathbf{u}_{0,j}^{(t)}|\mathbf{u}_j^{(t-1)})} \right). \quad (2.10)$$

If $\mathbf{u}_{0,j}^{(t)}$ is not selected, $\mathbf{u}_j^{(t)}$ is set to $\mathbf{u}_j^{(t-1)}$.

Strong correlation structures in the posterior of the parameters and latent variables can make exploration of the posterior computationally challenging. The construction of efficient MCMC algorithms is an engineering art form, and there are many techniques to shorten the exploration time for the posterior distribution. To shorten the exploration time the correlated parameters should preferably be blocked, and sample from the conditional joint distribution of the block as one step of the Metropolis-Hastings algorithm. Preferably we would like to have as little correlation between the blocks as possible. If there are strong correlations between blocks the sampler might be improved by delaying the updating of one block until the next block also has been updated (Rue & Held, 2005). The selection of proposal function in the Metropolis-Hastings algorithm (2.10) is crucial for the time efficiency, and the correlation between the parameters can be used to guide the proposal function $q(\cdot)$. For example the Metropolis adjusted Langevin algorithm (MALA) (Roberts & Stramer, 2002) utilizes the gradient of the posterior for fast exploration, and RMALA (Girolami & Calderhead, 2011) utilizes a Riemann manifold in the MALA algorithm for a dynamic selection of step sizes of the proposals.

The MCMC method is a general method for calculating the posterior of the parameters. It does not require any specific assumptions on the model, and the researcher is free to construct the model before creating the inference routine. This is an extremely appealing part of the MCMC routine and is a main reason for its popularity. However, as discussed in the previous paragraph, the MCMC method can be slow, and creation of an efficient MCMC remains an engineering art form.

Integrated nested Laplace approximations

If the model is additive latent Gaussian, integrated nested Laplace approximations (INLA) can be used for efficient inference of the posterior distribution (Rue et al., 2009). The INLA-technique consists of two main time consuming parts. First the posterior mode of the hyperparameters is found. This is done by maximizing the Laplace approximation

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \approx \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} \quad (2.11)$$

typically with a Newton routine. Here $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation of $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. Then an area with relatively high posterior density of the hyperparameters is explored with a grid procedure. The second time consuming part is to approximate the latent field (including the regression coefficients) for every set of explored hyperparameters. The computational complexity of approximating the latent field depends on the data distribution. If the response is Gaussian, the posterior of the latent field is Gaussian, and the approximation is exact. However, if the data distribution is skewed or has heavy tails a Gaussian approximation of the latent field tends to be inaccurate and a version of Laplace approximation should be applied in this step as well:

$$\tilde{\pi}_{\text{LA}}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{\text{GG}}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-i}=\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})}. \quad (2.12)$$

Here, $\tilde{\pi}_{\text{GG}}$ is the Gaussian approximation to $\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$, and $\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})$ is the mode of $\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$. The full Laplace approximation of the latent fields (2.12) is however time consuming, and satisfactory approximations can be achieved with third order Taylor approximation for several popular skewed likelihoods, e.g. Poisson and negative binomial (Rue et al., 2009). After the distribution of the latent field given the observations and hyperparameters is approximated, the uncertainty in the hyperparameters is integrated out (Rue et al., 2009; Martins et al., 2013):

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}_{\text{LA}}(x_i|\boldsymbol{\theta}_k, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y}) \Delta_k, \quad (2.13)$$

and hence the name integrated nested Laplace approximations. Here Δ_k is the area weight corresponding to the grid exploration of the posterior distribution of the hyperparameters.

The INLA-routine is implemented with many choices of likelihoods and correlation structures in the R-package R-INLA (Rue et al., 2009; Martins et al., 2013). The R-INLA routine is a user friendly package that makes it possible to use a large amount of latent Gaussian models for different applications by little effort of a researcher familiar to latent Gaussian models. The package can also be used as a black box for researchers in other fields than statistics, and due to the wide specter of possible application it has become a popular toolbox in many areas of research.

The output of the inference functions in the user friendly R-INLA package is constructed such that the user can investigate many aspects of the model, e.g. sample from the posterior distributions of the parameters. This makes R-INLA appealing since it broadens its range of applications. Furthermore, several choices are constructed for the user on how detailed the inference should be done. For example there are several options on how to accommodate for the uncertainty in the hyperparameters (2.11). The standard procedure is the central composite design (Wu & Hamada, 2011). If the user has reasons to believe that the variability is dominated by

2.3. Model selection and validation

the likelihood, the uncertainty in the hyperparameters can be omitted for computation efficiency by using the posterior mode. Neglecting the uncertainty in the hyperparameters has especially large impact on the computation time if the latent field requires time consuming approximations, such as if the data distribution is t -distributed and the full Laplace approximation (2.12) is required. At the present moment (August 20th, 2016) the main R-INLA paper, Rue et al. (2009), has 1309 citations and is an ongoing success story for fast Bayesian inference.

Computational complexity

The computational complexity of common efficient algorithms using Gaussian random fields, such as conditional sampling and evaluation of likelihood, is typically given as a function of the bandwidth of the precision matrix (Rue, 2001). For example the computational complexity of calculating the Cholesky decomposition of an arbitrary squared matrix is in the order N^3 , where N is the dimension of the matrix. By using efficient calculation procedures for sparse matrices, the calculation time reduces to the order $N * B^2$, where B is the bandwidth of the matrix (Rue, 2001). The memory needed to store a matrix can also be expressed as a function of the bandwidth. To store a $N \times N$ matrix, we only need to store $N * B$ elements. The bandwidth of the precision matrix depends on the neighborhood structure, a smaller neighborhood structure results in a smaller bandwidth. Therefore, by assuming a Markov structure on the Gaussian random field with small neighborhood structure, efficient inference can be achieved with respect to computation time and memory usage. For an elaboration of several computational efficient algorithms for GMRF we refer to Rue (2001); Rue & Held (2005).

2.3 Model selection and validation

For a given application, several statistical models are typically proposed, and only one of them is to be selected. Furthermore, the purpose of statistical modeling is to give insight through data, and it is important to give an objective scientific reasoning of whether insight is achieved. Intuitive criteria for good models are that they should be able to give relatively high probability to the observed data, have predictive power and not be too complex. Methods for performing model selection and validation represent a large ongoing research area. This section elaborates on three procedures used in the thesis for selection of covariance structures, covariates and observation models.

2.3.1 Cross-validation with a K-fold procedure

Cross-validation investigates the prediction performance of the model by using parts of the data for inference, and parts of the data for comparing predictions versus observations. Ideally, separate sets should be used for choosing the model, estimating the parameters and for comparison of predictions versus observations, respectively. As data is often scarce and model selection is often time consuming, for convenience a K-fold procedure is often used with the model selected using all the data. A K-fold cross-validation procedure divides the data into K sets, and iteratively uses each set as a test set and the others as training sets (Hastie et al., 2009, page 241). In this thesis, we performed K-fold cross-validation on the model, which was selected using all the

data. The training and test sets are chosen randomly in paper I. In paper II and paper III, we use clustered test and training sets to make the validation more representative for the application of the research.

2.3.2 Deviance information criteria (DIC)

The DIC favors models which give a good fit to the data, and penalizes complex models, which is also the intuition behind other information criteria, e.g. AIC (Akaike, 1974) and BIC (Schwarz et al., 1978). The DIC penalizes complex models through the “Effective number of parameters” given by

$$p_D = E_{\theta|y}[-2 \log(\pi(y|\theta))] + 2 \log[\pi(y|\tilde{\theta}(y))], \quad (2.14)$$

where $\tilde{\theta}(y)$ is the estimated parameter vector. To quantify the goodness of fit, the deviance is defined as $D(\theta) = -2 \log(\pi(y|\theta))$. The DIC-value is then defined as

$$\text{DIC} = D(\tilde{\theta}) + 2p_D. \quad (2.15)$$

The DIC-value was introduced in (Spiegelhalter et al., 2002) and has been extensively used ever since. See Spiegelhalter et al. (2014) for a short elaboration of the DIC, its usage in scientific papers, and a discussion of weaknesses and modifications.

2.3.3 Bayes factor

The Bayes factor (Gelfand, 1996) is the fraction of the marginal likelihoods of two models:

$$\text{BF} = \frac{\pi(y|\text{Model 1})}{\pi(y|\text{Model 2})}. \quad (2.16)$$

Note that the Bayes factor is sensitive to the priors since they are integrated out in both numerator and denominator in (2.7). The Bayes factor should thereby be used with caution, and preferably the user should have some knowledge of the prior contributions to the marginal likelihoods.

Pseudo Bayes factor (Gelfand, 1996) is not sensitive to the priors and can be used as an alternative to the Bayes factor. The intuition behind the pseudo Bayes factor criteria is that the cross validation densities, $\pi(y_i|y_{-i}, \text{Model})$, should be fairly large when the model is good. The pseudo Bayes criteria is defined by:

$$\text{PBF} = \prod_{i=1}^n \frac{\pi(y_i|y_{-i}, \text{Model 1})}{\pi(y_i|y_{-i}, \text{Model 2})}. \quad (2.17)$$

Observe that the PBF typically does not penalize complex models as much as BF since the posteriors of the unimportant parameters are typically concentrated at locations where they give little contribution to predictions. The PBF is, however, sometimes dependent on tail behavior, which can make it difficult to calculate, and it must therefore be used with caution.

3 Background: Application and motivation

3.1 Bycatch in the Barents Sea shrimp fishery

In the Barents Sea the cold arctic currents meet the warmer Atlantic currents (page 47 Jakobsen & Ozhigin, 2011), generating vertical flows which give rise to a rich flora of species. One of the species that has a direct commercial value is the shrimp (*Pandalus borealis*). The trawling after shrimp occurs at the seabed at around 200-400 meters depth where the shrimp concentration is highest (Jakobsen & Ozhigin, 2011, page 172). Fig. 3.1 shows an illustration of a shrimp trawl equipment. Notice the grid which sorts out the larger fish. The grid was imposed in 1992/1993 (ICES, 1994) to reduce bycatch, and has been mandatory ever since. The trawl is held open at the seabed with the help of two large heavy metal doors that scrape the seabed at each side of the trawl (not included in Fig. 3.1). The size and shape of the opening of the trawl varies with several factors, e.g. the weather, the number of trawls behind the boat and the maneuvers of the captain.

The Norwegian Monitoring and Surveillance Service (MSS) continuously conduct observations representative for the commercial fisheries in the Barents Sea and adjunct waters. At present time there are employed full time 17 observers at MSS (Rolf Harald Jensen, personal communication, May 3, 2016). The main task of the observers is to observe the amount of bycatch in the commercial fishery, and thereby regulate the fishery given the rules constructed by the Norwegian Directorate of Fisheries. If it is observed that an area typically has more than a certain amount of bycatch of juvenile fish per kilo-gram shrimp, the area is temporally closed for shrimp fishing. Today this ratio is 0.8 for cod (*Gadus morhua*), 2.0 for haddock (*Melanogrammus aeglefinus*) and 0.3 for redfish (*Sebastes norvegicus* and *Sebastes mentella*) (Fiskeridirktoratet, 2005).

3.1.1 Monitoring of the shrimp fishery

The practice of closing and opening of fishing areas has been an important contribution to the Norwegian fishery management since 1983 in order to reduce bycatch of juvenile fish in the Barents Sea shrimp fishery. With time the shrimp fishery has shifted towards using larger industrialized boats, while at the same time the number of vessels has been reduced. The price for hiring commercial fishing boats for monitoring purposes has thereby increased to a level which is challenging for MSS. The Norwegian Directorate of Fisheries has therefore requested more research on data collected by the MSS in order to obtain a more cost effective regulation procedure, while at the same time maintaining or improving the quality of the regulation.

The main objective of our research on bycatch ratio predictions is to create a model that can be

3. BACKGROUND: APPLICATION AND MOTIVATION

Figure 3.1: *Illustration of a shrimp trawl.*

implemented and used for real time regulation of the Barents Sea shrimp fishery. An automatic real time regulation of the fishery would be of great value for MSS since it could be used to optimize their resource allocations. The research that is presented in this thesis has therefore the potential to be of great value for the Norwegian society, as it could contribute to a sustainable fishery in the far north. In addition, an implemented automatic data driven regulation procedure would be short term beneficial for the fishermen, since the fishing season would be more predictable.

Currently a simple ratio estimator (Scheaffer et al., 1996, page 204), which does not fully utilize all data available, is in use for regulating the shrimp fishery. When MSS suspects that there is a high ratio of bycatch in a certain area, an inspector joins or rents a trawler and counts the number of juvenile cod caught as bycatch by new trawl hauls in that area. The bycatch ratio is then estimated by dividing the total number of juvenile cod by the total catch of shrimps. Based on this estimation a subjective decision is made whether to close the area. After an area is held closed for some time (often some months), data from new trawl hauls are collected and a decision is made whether to open or not.

Looking abroad, there are different rules and procedures regarding real time closures to reduce bycatch in commercial fisheries. For example, the Scottish North Sea demersal fishery and the Icelandic cod fishery are regulated by closing areas when high bycatch ratios are observed. In these fisheries the areas are automatically reopened after 14 to 21 days (Little et al., 2015). We have found no data driven justification of these time intervals. In the US the regulation procedure is fundamentally different. The regulation in the North-America typically relies on collective cooperation and on voluntarily contribution by the commercial fishery (Little et al., 2015). As far as we know, no regulating routine has been implemented using Bayesian spatio-temporal statistics. It appears reasonable that a Bayesian model may be able to produce reliable predictions with high precision, which is an important motivation for this application.

3.1. Bycatch in the Barents Sea shrimp fishery

3.1.2 Historical bycatch prediction

Good prediction procedures for aggregated historical commercial bycatch are important ingredients for knowledge about the impact of Barents Sea shrimp fisheries on fish stock abundances. Bycatch might have an impact on the economy in fisheries involving the fish species that are caught as bycatch, and historical bycatch predictions can be used to give insight into economic losses caused by shrimp trawling. Redfish is now marked as *critically endangered* of going extinct in the Norwegian Red List for Species (Kålås et al., 2015), illustrating the need for good algorithms to predict the damage caused on the redfish populations due to bycatch. Cod is on the other hand marked as *least concern*, and is thereby assumed not to be in danger of extinction. However, precautions are important for a sustainable fishery.

Commercial trawlers are obliged to report commercial catches together with information regarding the catch, such as location, the date and the trawling equipment used. These data are processed by the Norwegian Directorate of Fisheries and given to the Norwegian Marine Research Institute in Bergen, Norway. Currently, a simple ratio procedure is in use for producing official historical commercial bycatch estimates in the Barents Sea shrimp fishery (Ajiad et al., 2007; Hylen & Jacobsen, 1987; ICES, 2015). This ratio procedure assumes that the observed bycatch ratios are representative for the commercial bycatch ratios. The estimates are then produced by scaling the ratios with the total commercial shrimp catch:

$$\hat{B}_{A,t}^{\text{ratio}} = \frac{\sum_{i=1}^n b_{i,A,t}}{\sum_{i=1}^n c_{i,A,t}} C_{A,t} = R_{A,t} C_{A,t}. \quad (3.1)$$

Here $(c_{i,A,t}, b_{i,A,t})$ are the i th observed target catch and bycatch in the survey data from area A at time interval t , $R_{A,t}$ is then the observed bycatch ratio in area A and time interval t , and $C_{A,t}$ is the total commercial target catch in area A at time interval t . The historical bycatch in a larger time interval can then be estimated in the whole Barents Sea as $\sum_A \sum_t R_{A,t} C_{A,t}$. If the area has no observed bycatch ratios, observations from larger areas and time intervals are used (Ajiad et al., 2007).

Both the commercial fishery and the surveillance service use the same type of trawling equipment. However, the commercial shrimp trawlers have economic incentives to trawl in areas with high density of shrimps. The surveillance service, on the other hand, trawls at areas where shrimp trawling occurs without the same economic incentives. As the surveillance observations are not based on the same economic incentives as the commercial fishery to search for higher shrimp densities, the assumption of representative ratios in (3.1) may be questionable. Furthermore, areas used for the estimation in (3.1) often contain few observed trawl hauls compared with the commercial fishery. It is thereby reasonable that the ratio method is not robust, is biased and has a large variance which is difficult to quantify. Versions of the ratio method are typically used as the standard procedure for estimating historical bycatch of other fisheries worldwide (Vinther, 1999; Ye et al., 2000; Ye, 2002; Walmsley et al., 2007; Davies et al., 2009; Amandè et al., 2010). Since the ratio method has several obvious and important flaws, there is a need for pushing the frontiers on bycatch research by establishing sophisticated and good prediction procedures for historical bycatch in order to improve the insight into damage caused by the bycatch.

3.2 Wind speed in the North Atlantic

Previous research indicates a significant trend of increasing wave height in the North Atlantic (Vanem et al., 2012). Waves are mainly generated by energy transfer from the atmosphere to the ocean due to wind friction on the sea surface, see e.g. Talley (2011). In our research we investigated whether there is a significant increase in wind speed in the same area as identified by increased wave heights. The findings in this research may thereby give an indication whether the wave height increase is due to local wind changes or is caused by energy transferred by swells. The latent Gaussian model used in this this research was also used for detecting increased wave heights in Vanem et al. (2012), and is based on a general latent Gaussian model introduced in Wikle et al. (1998).

4 Summary of papers

All the four papers in this thesis are based on Bayesian spatio-temporal hierarchical models with correlation structures modeled as latent Gaussian random fields with algorithms that utilize sparse structures imposed on the Gaussian random fields. The models in paper I, paper II and paper III are estimated with use of R-INLA. Paper II and paper III are further based on paper I, which is a Bayesian extension of the model introduced in Aldrin et al. (2012). Even though our models for bycatch are of general types and are similar to models used for e.g. modeling air pollution (Cameletti et al., 2013) and weather prediction (Finley et al., 2012), we argue that they are also tailored for the purpose of bycatch predictions since most of the parameters have biological interpretations for describing bycatch. The model in paper IV is estimated with a Gibbs sampler with Metropolis Hastings steps. This model was proposed as a general spatio-temporal model in Wikle et al. (1998), and further used in e.g. Natvig & Tvette (2007); Vanem et al. (2012) for predicting earthquakes and wave heights, respectively.

4.1 Paper I

Paper I introduces a Bayesian hierarchical spatio-temporal latent Gaussian model for predicting bycatch ratios. The model is applied on bycatch of juvenile cod in the Barents Sea shrimp fishery, and it assumes that the occurrence of shrimp and juvenile cod can be modeled by linked regression models containing several covariates and a sum of random effects modeled as Gaussian fields. This summary is divided into two parts. First we summarize the effects that were found important for prediction, and then we summarize possibilities of the research from a management perspective.

4.1.1 Effects found important for prediction

It is concluded that the bycatch ratio of cod depends on a seasonal effect, trawling equipment, the zero-group abundance and time on day. It is furthermore shown that the amount of bycatch is dependent on the shrimp catch. All the covariates which are shown to be important are given a biological justification. As an example, for shrimp fishermen it is common knowledge that it is more difficult to catch shrimp during night at certain times of the year. This is probably due to the vertical migration pattern of the shrimps which depends on light conditions (Hopkins et al., 1993). This night effect was, however, shown to have a small effect on the bycatch ratio since the cod bycatch is also reduced during night time trawl.

The random effects are intended to approximate the underlying latent dependence structures. The estimated spatial random effect indicates that some locations typically have more catch or bycatch than others. This is reasonable since the shrimps are known to be concentrated at frontal

zone areas (Jakobsen & Ozhigin, 2011, page 173), and the juvenile cod have a specific migration pattern which has clear similarities with the posterior mean of the spatial effect (Jakobsen & Ozhigin, 2011, page 227). The estimated temporal random effect indicates that some years typically have more catch or bycatch than others, which is reasonable since the number of juvenile cod present in the Barents Sea varies from year to year (Jakobsen & Ozhigin, 2011, page 565). For both shrimp and bycatch a clear spatio-temporal interaction is observed, indicating that observations close to each other tend to be more equal.

4.1.2 Management

The main objective of this research is to construct a model which can be utilized for real time regulation by the MSS. Paper I illustrates that the model is capable of predicting reasonable bycatch ratios for the shrimp fishery, and thereby has the potential to be of great value for management purposes. Bycatch management in the Barents Sea is conducted by closing and opening of areas where commercial fishery occurs. Paper I illustrates both these two aspects of management through data driven decision making. Through an example, the paper shows how the model can be used for closing an area. After the area is closed, the paper shows how the model can be used to reopen the area without new expensive observations.

4.2 Paper II

A prediction procedure for commercial bycatch utilizing the Bayesian spatio-temporal bycatch model introduced in paper I is proposed. The purpose of paper II is to predict historical commercial bycatch, which is important for insight into the damage caused by the bycatch. A log-Gaussian likelihood is used in paper I. However, after a comment from a reviewer, the data model was modified from log-Gaussian to zero-inflated negative binomial in paper II. Paper II furthermore differs from paper I by utilizing two sources of data, fishery data and survey data. The inclusion of the fishery data in paper II implies a much larger data set, resulting in new computational difficulties with use of the spatio-temporal correlation structure in paper I.

The prediction of historical bycatch is accomplished in two steps. First the survey data is used for estimating the model. Then a prediction procedure is constructed outside of R-INLA with use of the estimated model and by combining the survey data with the fishery data. The prediction procedure utilizes the uncertainty in the parameters, and by use of Markov integration the predictions with uncertainties are obtained. Yearly and quarterly bycatch predictions in the period 1994 to 2006 are constructed with use of the Bayesian hierarchical spatio-temporal model.

Estimates of historical bycatch are typically made with the ratio and effort based method (Davies et al., 2009; Vinther, 1999; Ye et al., 2000; Amandè et al., 2010; Ye, 2002; Walmsley et al., 2007). We compare the model based procedure proposed with these two methods, and make a strong case that our prediction procedure produces more reliable historical bycatch predictions.

Paper II introduces a new method for predicting historical bycatch, and gives new insights into the damages caused by bycatch in the Barents Sea shrimp fishery. The paper further illustrates how easily a complex Bayesian spatio-temporal model can answer a question which is of huge

4.3. Paper III (Technical report)

interest in fisheries research. In most of the years our Bayesian predictions are in agreement with the ratio and effort based methods. However, in some years our predictions differ clearly from the ratio and effort based estimates. In these years we argue that our method is the most trustworthy, as it is more robust when few observed trawl hauls are present.

In paper II, the yearly historical bycatch is illustrated as percentage of abundance estimates of one year old cod in the Barents Sea. The predictions indicate that a relatively small proportion of the juvenile cod are taken as bycatch, which may hence illustrate the success of the current regulation regime elaborated in section 3.1.1.

The zero-inflated negative binomial data distribution used in paper II was tested and discarded in paper I. This was done based on a the pseudo Bayes factor, see sub-section 2.3.3. In the process of writing Paper II, it was observed that this validation criterion was dependent on tail behavior, and was estimated inaccurately in paper I without warning from R-INLA. At the R-INLA home page, under frequently asked questions (www.r-inla.org/faq), there is now given a warning which elaborates that leave out one densities used in the pseudo Bayes factor can be estimated inaccurately without a warning from R-INLA, and must therefore be used with caution. This warning message was constructed after the observed inaccurate leave out one densities in paper I were discussed with the R-INLA developer Håvard Rue. In paper II it was observed that the zero-inflated version of the model in paper I improves the predictions by removing an observed bias, and the data distribution was therefore modified.

4.3 Paper III (Technical report)

Ecological count data often contains an abundance of zero values. For example the data used in paper I and II contains 18.5% zero counts of cod bycatch, and in the same trawl hauls there were reported 57.4 % zero counts of redfish. When modeling observations with an abundance of zero values, it is typically assumed that there is one structure which explains all or some of the zero counts, and another structure which describes the rest of the counts. In paper II, these two structures are assumed with a link connecting them by *one common* spatio-temporal linear predictor. Paper III goes further by comparing such models with an unlinked model which uses *separate* spatio-temporal linear predictors for each of the two structures. The data for redfish bycatch contains many zero counts, and it is therefore reasonable that the procedure for modeling the zero-probability is of special importance for redfish. In paper III, both bycatch of cod and redfish are modeled.

As in paper I and paper II, the inference in paper III is performed with use of R-INLA. We show that linking or not linking has a potential important impact on the prediction performance. It is illustrated that a badly specified link function may result in biased predictions, and may make the model inadequate for estimating the prediction uncertainty. For the cod bycatch data, the best linked model (the same model as used in paper II) performs very similar to the model without link. However, for the redfish bycatch data, the model without link seems to perform best because of more accurate predictions.

4.4 Paper IV

Paper IV was written together with a former PhD student and colleague at the University of Oslo, Erik Vanem. The paper is based on his former research on long term wave height changes in the North Atlantic (Vanem et al., 2012), where a significant temporal increase in wave heights was observed. This increase may be caused by locally increased wind speed or swells. By using the same model as in Vanem et al. (2012), estimated with a Gibbs sampler with Metropolis Hastings steps, no significant increase of monthly maximum wind speed is observed. This indicates that the increase of wave heights may be explained by energy transferred by swells. The model used is a relatively complex hierarchical spatio-temporal model, where most of the complexity lies in the spatial and the short term spatio-temporal interaction. Such structures are important to consider for estimation of the uncertainty. For the parameter of interest, which is a long term trend, a linear trend is investigated. For details of the calculation of the posterior distributions, see Wikle et al. (1998); Vanem et al. (2012).

5 Discussion

The main objective of this thesis has been to model the bycatch of juvenile fish in the Barents Sea shrimp fishery, and this discussion is divided into three parts. The first part is a discussion of continuous spatio-temporal modeling in R-INLA, and elaborates a procedure using a smoothed temporal AR(1)-structure. The second part is about the importance of careful data collection. The third part is a discussion of the possibilities for implementing the models constructed for bycatch to perform real time regulation in the Barents Sea shrimp fishery with respect to bycatch.

5.1 Continuous spatio-temporal modeling with R-INLA

Short term spatio-temporal interaction correlation functions for shrimp catch and bycatch are reasonable to assume to be continuous. In the thesis we model shrimp and bycatch with a continuous exponential spatio-temporal correlation structure. Currently, there exists no such implemented procedure in R-INLA for unstructured data in space and time. However, a generic class is available, where the *precision matrix* is given by $\mathbf{Q} = \tau\mathbf{C}$, where τ^{-1} is the marginal variance and $\mathbf{C}$ is fully specified by the user. In paper I, II and III a procedure was constructed outside of R-INLA which uses this generic class combined with an explicit formula for the covariance function. This procedure finds the posterior mode of the range parameters in the spatio-temporal interaction by running R-INLA several times. Such a specification is computationally inconvenient since the correlation matrix needs to be inverted outside of R-INLA for each proposal of its range parameters. Furthermore, to make the precision matrix sparse, some values in the precision matrix are truncated to zero. How much the truncation influence the correlation structure is data dependent, and especially depends on the correlation lengths in space and time compared to how scattered the observations are. In paper I, different truncation procedures were investigated to validate whether such a possible influence could be neglected.

For fast inference, it is crucial to have an explicit formula for a sparse precision matrix defining the correlation structure. There exists an implemented procedure for spatio-temporal modeling in R-INLA which uses an AR(1)-structure in time, and Matern covariance structure in space:

$$\xi_r = a\xi_{r-1} + \omega_r, \quad \omega_r \sim N(0, \tilde{\Sigma}) \quad r = 0, \dots, T. \quad (5.1)$$

Here $\tilde{\Sigma}$ is the precision matrix in space with Matern covariance function, and the reader is referred to equation (6) in paper I for further details. See e.g. Cameletti et al. (2013); Finley et al. (2012) for two applications of (5.1). Such an autoregressive spatio-temporal correlation structure is quite similar to the structure used in paper I, II and III, and was tested and discarded in paper I. The reason why the continuous correlation structure is preferred compared to (5.1) in paper I, may be due to the fact that the underlying structure is better approximated continuously compared to a rough discretization in time. The next subsection elaborates a procedure using R-

INLA with a continuous correlation structure in space-time by smoothing the temporal AR(1)-structure.

5.1.1 Smoothed temporal AR(1)-structure

The correlation structure given in (5.1) is implemented in R-INLA where the spatial part of the separable correlation structure is approximated as continuous by a linear approximation of a GMRF (Lindgren et al., 2011; Cameletti et al., 2013). With the same reasoning, an approximation of a continuous spatio-temporal correlation function in time and space can be performed by including a linear smoothing of the temporal part. Define the Gaussian field

$$\delta(t) = g_1(u)\xi_r + g_2(u)\xi_{r+1}, \quad (5.2)$$

where $u = \frac{t-r\Delta_t}{\Delta_t}$, Δ_t is the discretization length in time and r is an integer chosen such that $t \in [r\Delta_t, (r+1)\Delta_t)$. Notice that this smoothing is achievable in R-INLA with given weight functions, g_1 and g_2 , by manipulating the A-matrix described in Cameletti et al. (2013). Natural choices of weighting functions g_1 and g_2 are

$$g_1(u) = 1 - u \quad \text{and} \quad g_2(u) = u, \quad (5.3)$$

which are one dimensional versions of the weights in Lindgren et al. (2011); Cameletti et al. (2013). The structure (5.2) with use of these weights is further referred to as a *linear smoothed Ar(1)-structure*.

The temporal smoothing discussed above is quite similar to the predictive processes introduced in Banerjee et al. (2008). The correlation function (5.1) is not defined between the discrete time points. To introduce a predictive process, define $\{\xi'(t)\}$ to be a continuous extension of (5.1) with the temporal term in the separable spatio-temporal correlation function replaced by $a^{|t_1-t_2|/\Delta_t}$. See the appendix in paper I for an explicit formula of the correlation function given by the spatio-temporal structure defined by (5.1). A predictive process with knots at the original locations in (5.1) can be defined as

$$\tilde{\xi}'(t) = E[\xi'(t)|\xi'(r\Delta_t), \xi'([r+1]\Delta_t)], \quad (5.4)$$

where r is an integer chosen such that $t \in [r\Delta_t, (r+1)\Delta_t)$ (Banerjee et al., 2008). The predictive process (5.4) turns out to be identical to (5.2) with weight functions

$$g_1(u) = \frac{a^u(1 - a^{2(1-u)})}{1 - a^2} \quad \text{and} \quad g_2(u) = \frac{a^{1-u}(1 - a^{2u})}{1 - a^2}. \quad (5.5)$$

The previous result can be proved with the conditional expectation formula for multivariate Gaussian random variables. Fig. 5.1 shows the weights produced by (5.3) and by the predictive processes (5.5) with different selection of autoregressive coefficients. We observe that the weights given by the predictive process in Banerjee et al. (2008) are approximately similar to (5.3) when there is a strong autoregressive correlation. This indicates that using a linear smoothed Ar(1)-structure, which is easily achievable with R-INLA, produces similar results as

5.1. Continuous spatio-temporal modeling with R-INLA

Figure 5.1: Illustration of the weights in (5.2) with use of linear weights (black line), and weights produced by the predictive process (red line) with different autoregressive parameters. In a) $a = 0.7$, b) $a = 0.5$ and c) $a = 0.25$.

using the predictive process, which is not easily achievable with R-INLA.

Since the Gaussian process (5.2) is a weighted average with sum of weights equal to 1 with use of (5.3) and less than one with use of (5.5), the marginal variance is smaller between the locations in (5.1). Fig. 5.2 illustrates by how large factor the marginal variance decreases between the locations with use of the weights (5.3) and (5.5) for different autoregressive parameters. To accommodate for the underestimated marginal variance, Finley et al. (2009) suggested to add independent random terms such that the marginal variance is stabilized. Adding such independent noise can be done within R-INLA (with predefined marginal variance). However, this independent variation should preferably be explained through the spatio-temporal latent field, and we do not know the exact effect caused on the estimation by such a replacement.

As the two spatio-temporal structures (5.1) and (5.2) differ from each other, the autoregressive parameters may be estimated differently. By using simulations of $\{\xi'(t)\}$, we have observed that the discrete version (5.1) typically estimates stronger autoregressive correlation compared to the linear smoothed Ar(1)-structure. We find this reasonable, as the weights in (5.3) are linear while the correlation decreases exponentially. Also for the shrimp and bycatch models in paper I the autocorrelation is estimated smaller by using the smoothed AR(1)-structure. The autoregressive parameter estimate reduces from 0.72 (0.64,0.78) to 0.60 (0.51,0.67) for the shrimp catch and from 0.62 (0.54,69) to 0.32 (0.19,0.44) for the bycatch in paper I, where the intervals are 95% credibility intervals.

We have predicted historical bycatch both with use of (5.1) and the linear smoothed Ar(1)-structure. The predictions were then approximately similar to those obtained in paper II (with time discretized in 14 days). This indicates that the smoothing of the AR(1) process is not important for predicting historical bycatch. However, the model selection criteria in Paper I favored the linear smoothed Ar(1)-structure compared with (5.1), but not compared with the selected continuous correlation structure.

Figure 5.2: Illustration of the reduction in marginal variance caused by the linear combination in (5.2) with use of linear weights (black line), and the weight produced by the predictive process (red line), with different autoregressive parameters. In a) $a = 0.7$, b) $a = 0.5$ and c) $a = 0.25$.

5.2 Cleaning of the survey data

The survey data used for bycatch in paper I, paper II and paper III include fewer observations compared to Aldrin et al. (2012). The reason for this is due to newly found inconsistencies and absence of information in the manually collected survey data. When the data was collected, the observers had to fill out a comprehensive form with up to around 100 numbers and characters (Mjanger et al., 2013) for every trawl haul. These numbers and characters were further processed on shore. When collecting and processing the large amount of data, it is reasonable that human errors occur, and much time has been spent in the work with this thesis, filtering out observations that were obviously wrong. For example, some observations were taken by the same vessel at locations in time and space such that the vessel had to travel unrealistically fast, and some observations lacked e.g. the starting or stopping time down to minutes scale. Such obviously incorrect observations are easy (but time consuming) to find, and were for simplicity removed before doing the inference. We want to emphasize that 99.2% of the data collected by the MSS were used in paper I. With respect to the large amount of information affiliated to each observation, this reflects the good quality of the data collected by the MSS. After estimating the model, leave out one densities were used for investigating which observations the model found surprising in search of more inconsistencies. It turned out that 14 observations of shrimp catches were extremely surprising, and thereby indicated a need for further manual inspection. When comparing these observations with the raw data, it was observed that they were processed such that they were incorrectly given in tons instead of kilo-grams.

When the survey data were collected in the Barents Sea and adjunct waters, mainly for the purposes of short term monitoring, the inspectors may not have foreseen that statistical models were later to be constructed based on the observations. The time spent on discovering inconsistencies in the data reminds us on the importance of carefully collecting and processing data, no matter what the short term reasons for the sampling and processing is.

5.3 Usage of the bycatch model for regulation

Paper I constructs a model which can be used for regulating the Barents Sea shrimp fishery with respect to bycatch. The paper illustrates several procedures for how the model can be used in practical management applications. In the light of paper II and paper III, it is reasonable that the Hurdle model in paper III should be used for regulation with a similar procedure as in paper I. An implementation of the model must be done with caution not to jeopardize a sustainable fishery. Furthermore, we want to emphasize that the model constantly needs the addition of new observations in order to give updated information to the user through the spatio-temporal interaction and the yearly effect.

The shrimp fishery has in the last decades become more industrialized, and thereby more expensive to regulate. To hire a shrimp trawler in the Barents Sea can cost as much as 472 000 NOK a day (Kjell Nedreaas, personal communication, January 13, 2016). If such an expedition lasts for three weeks, the costs will be approximately 10 million NOK, which comes in addition to other expenses, e.g. salary to the inspectors. These numbers illustrate that a data driven automatic regulation procedure can be of great importance for MSS with respect to optimal resource allocation.

In the Barents Sea there are several other fisheries which are regulated by the MSS, e.g. the cod fishery and pollock (*Pollachius virens*) fishery. It seems reasonable that the procedure introduced in the presented work can give good predictions of bycatch (or proportion of undersized fish) also in these fisheries. However, the models constructed in this thesis must be generalized to other fisheries with caution. For example the pollock is known to be concentrated locally at areas depending on the seabed surface (personal communication, Rolf Harald Jensen) indicating that there is a geophysical reason to question the stationary and isotropic assumptions utilized for bycatch prediction of cod and redfish.

At present time (August, 2016), a working group, with mandate given by the Norwegian Directorate of Fisheries, is writing a report on how the regulation of fisheries can be achieved more cost efficiently in the Barents Sea. One chapter of the report discusses the possibilities made available by the research in this thesis. Since the bycatch models are shown to produce reliable predictions with uncertainty, it is realistic that the models, or modifications of the models, can and will be utilized for real time regulations. The inspectors at the MSS have the best intuition and knowledge about bycatch and regulation routines. And at in the end it is they who should decide whether the predictions made by the models are trustworthy enough for regulation. We believe that the MSS would draw benefits from using an interface based on the research in this thesis, and be better equipped to regulate the large area that the Barents Sea and adjunct waters spans.

From a North American perspective, the regulation by MSS is viewed as a strict top to bottom management (Little et al., 2015). In North America, it is typically the commercial fisheries who report the bycatch in real time, and they argue that avoiding a strict top to bottom management gives incentives to cooperate with the government and to produce correct observations (Little et al., 2015). The model introduced in this thesis can contribute to such cooperation. The model can be used to continuously construct maps with color codes describing locations with high

probability of giving much bycatch. The maps can then be used as a tool for the commercial fishery to avoid such areas, and to remind the fishery of the controversial bycatch problem. Such an application, with use of a Bayesian spatio-temporal model, was also suggested in Ward et al. (2015). At the end of the day it is the total amount of fish killed as bycatch which is the important issue of bycatch management, no matter whether it is strictly regulated by law or by incentives.

References

- AJIAD, A., AGLLEN, A., NEDREAAS, K. & KVAMME, C. (2007). NAFO/ICES Pandalus Assessment Group Meeting .
- AKAIKE, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* **19**, 716–723.
- ALDRIN, M., MORTENSEN, B., STORVIK, G., NEDREAAS, K., AGLLEN, A. & AANES, S. (2012). Improving management decisions by predicting fish bycatch in the Barents Sea shrimp fishery. *ICES Journal of Marine Science: Journal du Conseil* **69**, 64–74.
- AMANDÈ, M. J., ARIZ, J., CHASSOT, E., DE MOLINA, A. D., GAERTNER, D., MURUA, H., PIANET, R., RUIZ, J. & CHAVANCE, P. (2010). Bycatch of the European purse seine tuna fishery in the Atlantic Ocean for the 2003–2007 period. *Aquatic Living Resources* **23**, 353–362.
- BANERJEE, S., GELFAND, A. E., FINLEY, A. O. & SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 825–848.
- BAYES, M. & PRICE, M. (1763). An essay towards solving a problem in the doctrine of chances. By the late rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a letter to John Canton, A.M.F.R.S. *Philosophical Transactions (1683-1775)* , 370–418.
- CAMELETTI, M., LINDGREN, F., SIMPSON, D. & RUE, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis* , 1–23.
- CASELLA, G. & BERGER, R. L. (2002). *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA.
- CRESSIE, N. & WIKLE, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons.
- DAVIES, R., CRIPPS, S., NICKSON, A. & PORTER, G. (2009). Defining and estimating global marine fisheries bycatch. *Marine Policy* **33**, 661–672.
- DEVORE, J. L. & BERK, K. N. (2007). *Modern mathematical statistics with applications*. Cengage Learning.

- FINLEY, A. O., BANERJEE, S. & GELFAND, A. E. (2012). Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes. *Journal of geographical systems* **14**, 29–47.
- FINLEY, A. O., SANG, H., BANERJEE, S. & GELFAND, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Computational statistics & data analysis* **53**, 2873–2884.
- FISKERIDIREKTORATET (2005). Forskrift om utøvelse av fisket i sjøen (in Norwegian). <https://lovdata.no/dokument/SF/forskrift/2004-12-22-1878>.
- GELFAND, A. E. (1996). Model determination using sampling-based methods. *Markov chain Monte Carlo in practice*, 145–161.
- GEMAN, S. & GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 721–741.
- GIROLAMI, M. & CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 123–214.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., HASTIE, T., FRIEDMAN, J. & TIBSHIRANI, R. (2009). *The elements of statistical learning*, vol. 2. Springer.
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.
- HOPKINS, C., SARGENT, J. & NILSSEN, E. (1993). Total lipid content, and lipid and fatty acid composition of the deep-water prawn *Pandalus borealis* from Balsfjord, northern Norway: growth and feeding relationships. *Marine Ecology-Progress Series* **96**, 217–217.
- HYLEN, A. & JACOBSEN, J. (1987). Estimation of cod taken as by-catch in the Norwegian fishery for shrimp north of 69 N. *ICES CM*.
- ICES (1994). Report of the Arctic Fisheries Working Group, Copenhagen, 24 August - 2 September 1993.
- ICES (2015). Report of the Arctic Fisheries Working Group (AFWG), 23-29 April 2015 Hamburg, Germany.
- JAKOBSEN, T. & OZHIGIN, V. K. (2011). *The Barents Sea-ecosystem, resources, management. Half a century of Russian-Norwegian cooperation*. Tapir Akademisk Forlag.
- KÅLÅS, J., VIKEN, Å., HENRIKSEN, S. & SKJELSETH, S. (2015). The 2015 Norwegian red list for species. *Norwegian Biodiversity Information Centre, Norway*.
- LINDGREN, F., RUE, H. & LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 423–498.

REFERENCES

- LITTLE, A. S., NEEDLE, C. L., HILBORN, R., HOLLAND, D. S. & MARSHALL, C. T. (2015). Real-time spatial management approaches to reduce bycatch and discards: experiences from Europe and the United States. *Fish and Fisheries* **16**, 576–602.
- MARTINS, T. G., SIMPSON, D., LINDGREN, F. & RUE, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis* **67**, 68–83.
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics* **21**, 1087–1092.
- MJANGER, H., HESTENES, K., SVENDSEN, B. V. & WENNECK, T. D. L. (2013). *Håndbok for prøvetaking av fisk og krepsdyr (in Norwegian)*.
- NATVIG, B. & TVETE, I. F. (2007). Bayesian hierarchical space–time modeling of earthquake data. *Methodology and Computing in Applied Probability* **9**, 89–114.
- R CORE TEAM (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- ROBERTS, G. O. & STRAMER, O. (2002). Langevin diffusions and Metropolis-Hastings algorithms. *Methodology and computing in applied probability* **4**, 337–357.
- RUE, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 325–338.
- RUE, H. & HELD, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.
- RUE, H., MARTINO, S. & CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 319–392.
- SCHEAFFER, R., MENDENHALL III, W. & OTT, R. L. (1996). *Elementary survey sampling, Fifth Edition*. Duxbury Press.
- SCHWARZ, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics* **6**, 461–464.
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. & LINDE, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 485–493.
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. & VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 583–639.
- TALLEY, L. D. (2011). *Descriptive physical oceanography: an introduction*. Academic press.

- VANEM, E., HUSEBY, A. B. & NATVIG, B. (2012). A Bayesian hierarchical spatio-temporal model for significant wave height in the North Atlantic. *Stochastic environmental research and risk assessment* **26**, 609–632.
- VINTHER, M. (1999). Bycatches of Harbour Porpoises (*Phocoena phocoena*, L.) in Danish set-net fisheries. *Journal of Cetacean Research and Management* .
- WALMSLEY, S. A., LESLIE, R. W. & SAUER, W. H. (2007). Bycatch and discarding in the South African demersal trawl fishery. *Fisheries Research* **86**, 15–30.
- WARD, E. J., JANNOT, J. E., LEE, Y.-W., ONO, K., SHELTON, A. O. & THORSON, J. T. (2015). Using spatiotemporal species distribution models to identify temporally evolving hotspots of species co-occurrence. *Ecological Applications* **25**, 2198–2209.
- WIKLE, C. K., BERLINER, L. M. & CRESSIE, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics* **5**, 117–154.
- WU, C. J. & HAMADA, M. S. (2011). *Experiments: planning, analysis, and optimization*, vol. 552. John Wiley & Sons.
- YE, Y. (2002). Bias in estimating bycatch-to-shrimp ratios. *Aquatic Living Resources* **15**, 149–154.
- YE, Y., ALSAFFAR, A. & MOHAMMED, H. (2000). Bycatch and discards of the Kuwait shrimp fishery. *Fisheries Research* **45**, 9–19.