



Norwegian University of
Science and Technology

A Statistical Approach to Spatial Mapping of Temperature Change

Ingeborg Gullikstad Hem

Master of Science in Physics and Mathematics

Submission date: June 2017

Supervisor: Ingelin Steinsland, IMF

Co-supervisor: Geir-Arne Fuglstad, IMF

Norwegian University of Science and Technology
Department of Mathematical Sciences

Preface

This paper is the assignment of the 30-credit points course TMA4905 - Statistics, Master's Thesis, spring 2017. The thesis concludes my five-year study time at the Norwegian University of Science and Technology.

I would like to say thank you to my supervisor, Professor Ingelin Steinsland, for her guidance in the work on my thesis. She is always positive, and her excitement during the meetings is very motivating. She never sees problems, only solutions, and that is comforting for a stressed master student.

A huge thank you is also given to Geir-Arne Fuglstad. He has been a co-supervisor for my thesis, and has provided valuable help, guidance and advises this spring. He has been very patient, and has been eager to help me with programming, INLA, statistical theory and anything else I have needed help with. Thanks to Geir-Arne and his knowledge about almost everything, I have been able to do much more than I imagined in the beginning, which has led to both a better understanding of Bayesian statistics, and a thirst for more knowledge.

Thank you!

Ingeborg Gullikstad Hem, June 2017

Abstract

In this thesis, spatio-temporal temperature trends are estimated based on monthly average temperatures from 503 observation locations in the southern half of Norway. The time period studied is 1960 to 2016. A latent Gaussian model is proposed, where spatial Gaussian random fields and temporal polynomials of second degree are used to model the temperature trends. A Bayesian approach is taken for the inference, and the integrated nested Laplace approximations methodology is used to carry out the inference. The model is easy to interpret, has interpretable results, is able to predict missing observations, and successfully estimates temperature trends that correspond to other research results. Five sets of different prior distributions are used on the model components, each month of the year is modelled individually, and all months have the same prior distributions for each set of priors.

In general, the temperature trend is increasing over the entire southern half of Norway according to the model, and the five sets of prior distributions give similar results. Twelve different spatio-temporal temperature trends, one for each month of the year, are obtained, and all months of the year have experienced an increase in temperature from 1960 to 2016. In some areas and some months, the temperature has decreased as well. The increase is largest during the winter months, and smallest during the summer months. The model is computationally fast; the inference is carried out in approximately 20 minutes for each month. The range parameters of the Gaussian random fields are given extra attention, and a minor prior sensitivity analysis on the range parameters is carried out. The sensitivity analysis reveals that to achieve the interpretable results, the range parameters, especially for the warmest months of the year, must to a large extent be controlled by the prior distribution.

Sammendrag

I denne oppgaven er rom-tid-temperaturtrender beregnet fra 1960 til 2016 i den sørlige halvdelen av Norge. Månedlige gjennomsnittstemperaturer fra 503 observasjonslokasjoner er brukt. En latent Gaussisk modell brukes, der temperaturtrendene estimeres ved bruk av romlige Gaussiske stokastiske felter og temporale andregradspolynomer. En Bayesiansk tilnærming og inferens brukes, og inferensen utføres med metodikken "integrated nested Laplace approximations". Modellen er tolkbar, gir tolkbare resultater, er i stand til å predikere manglende observasjoner, og estimerer temperaturtrender som stemmer overens med andre forskningsresultater. Fem ulike sett med apriorifordelinger brukes på modellkomponentene, hver måned i året modelleres individuelt, og alle måneder har de samme apriorifordelingene for hvert sett av fordelinger.

I henhold til modellen er temperaturtrenden økende i den sørlige halvdelen av Norge. De fem settene med apriorifordelinger gir like resultater. Resultatene består av blant annet tolv forskjellige rom-tid-temperaturtrender, en for hver måned, og alle månedene i året har hatt områder med temperaturøkning i perioden 1960 til 2016. I tillegg har noen områder i noen måneder hatt en nedgang i månedlig gjennomsnittstemperatur. Temperaturøkningen er størst i vintermånedene, og minst i sommermånedene. Beregningsmessig er modellen rask, da inferensen utføres på omtrent 20 minutter for hver måned. Det gis ekstra oppmerksomhet til rekkeviddeparameterne til de Gaussiske stokastiske feltene, og en sensitivitetsanalyse utføres på apriorifordelingene til rekkeviddeparameterne. Sensitivitetsanalysen avdekker at for å oppnå tolkbare resultater må rekkeviddeparameterne i stor grad være kontrollert av apriorifordelingen. Dette gjelder spesielt for de varmeste månedene i året.

Contents

1	Introduction	1
2	Data and initial explanatory analysis	5
2.1	Monthly average temperature data	5
2.2	Initial data exploration	9
3	Background	13
3.1	Bayesian inference	13
3.1.1	Bayesian hierarchical models	13
3.1.2	Latent Gaussian models	14
3.2	Spatial statistics	15
3.2.1	Gaussian random fields	15
3.2.2	Matérn covariance function	16
3.2.3	Gaussian Markov random fields	17
3.2.4	Stochastic partial differential equations	18
3.3	Prior distributions	19
3.3.1	Penalized complexity priors	20
3.3.2	Penalized complexity priors on Gaussian random fields	21
3.3.3	log-Gaussian priors on Gaussian random fields	22
3.3.4	Other prior distributions	22
3.4	Fast inference for Bayesian models	23
3.4.1	Integrated nested Laplace approximations methodology	24
3.4.2	SPDE approach in R-INLA	26
4	Method	29
4.1	Comments on R-INLA	29
4.2	The model	30
4.3	Prior choices	33
4.3.1	Priors on the fixed effects and precision	33

4.3.2	Priors on the Gaussian random fields	34
4.4	Model summary	36
5	Results	39
5.1	Numerical problems	40
5.2	Results differing between models	42
5.2.1	Insufficient models	42
5.2.2	The preferred model	46
5.3	Results similar in all models	51
5.4	Predicting missing observations	62
5.5	Prior sensitivity	64
6	Discussion	67
7	Conclusion	71
8	Bibliography	73
A	Additional results	77

Chapter 1

Introduction

The objective of this thesis is to estimate spatio-temporal temperature trends using monthly average temperatures from the southern part of Norway. Measurements from 503 weather stations from 1960 to 2016 are used. The model and the model results should be interpretable, all available data should be used, and the inference should be computationally fast. Spatio-temporal temperature trends can give an idea on how the climate is changing, where the changes occur, and to some degree what we might expect in the future (Houghton, 2015). As humanity is facing rapid changes in the climate with the highest global average temperature for several hundred years (Hartmann *et al.*, 2013), to model the past and to predict the future is important.

Hem (2016) has investigated temperature trends in time only, using a random walk of second order and time series from 27 observation locations in Norway. Even though spatial relationships is not the focus, clear spatial patterns can be seen. Trend estimation in time only is in other words already investigated, and the objective of this thesis is to utilize the spatial dependencies in the temperature data when estimating and modelling the temperature trends. When utilizing spatial dependencies, the temperature data can contain missing observations and thus more data can be included in the analysis.

There are several ways to model a temporal temperature trend over space. The ideal way is one of the most complex alternatives: climate models. In climate models, differential equations are used to model physical relationships in the atmosphere, and require enormous amounts of processing power. Super computers are usually used to do the modelling (Houghton, 2015). Simpler alternatives include non-separable statistical models with spatio-temporal covariance func-

tions on random fields, separable statistical models with temporal processes with spatial innovations, and the approach used in this thesis: separable statistical models with simplified spatio-temporal interactions (Blangiardo and Cameletti, 2015). A Bayesian hierarchical model is used, as the Bayesian approach can model uncertainty in all parameters in the model. The Bayesian approach also includes possibilities to estimate parameters based on prior knowledge, and not only based on the data as in frequentistic statistics. The Bayesian hierarchical models are both intuitive and practical when carrying out inference. They yield numerous possibilities when doing inference due to their flexible nature (Rue *et al.*, 2009), and since Bayesian statistics is frequently used, multiple computational tools are developed to do fast inference on Bayesian hierarchical models. Thus, a Bayesian analysis is desired. In this thesis, latent Gaussian models which have structured additive predictors have been chosen, due to their intuitive interpretation and the possibilities to do fast inference on such models.

To model a spatio-temporal temperature trend using latent Gaussian models means that Gaussian random fields can be used for the spatial modelling. The Gaussian random fields (GRFs) take care of the spatial dependencies, while other model components can be added to model the trend in time. These can also be combined as spatio-temporal interactions. GRFs are used in the model in this thesis. Two problems arise with this model approach; the first is the amount of data. It is an advantage to have large data sets, but only if the data can be processed in reasonable time. Large data sets imply large matrices that must be factorized when doing inference, which often is time-consuming (Lindgren *et al.*, 2011). The other problem is that a Bayesian analysis is desired, and not a frequentistic one, which means more complex models. There is a need for a computational tool, preferably an existing one, that can carry out the inference.

The integrated nested Laplace approximations (INLA) methodology, proposed by Rue *et al.* (2009), is used to carry out the inference. The common approach when doing inference for latent Gaussian models is Markov chain Monte Carlo (MCMC) (Blangiardo and Cameletti, 2015). INLA can do fast inference on latent Gaussian models with Markov structure, which is suitable for the modelling of this thesis. INLA utilizes that discretized random fields, Gaussian Markov random fields to be more exact, with sparse matrices can be factorized in much less time than dense matrices, which solves the problem with large datasets. The GRFs included in the model are approximated using an approach proposed by Lindgren *et al.* (2011): the stochastic partial differential equation (SPDE) approach, which is implemented in INLA. INLA and the SPDE approach can carry out the inference in reasonable time, even though such models are computational intensive to work with.

In Bayesian inference, the unknown parameters are given *prior distributions*, to include prior information about the parameters in the model (Cressie and Wikle, 2011). Ideally, experts in the fields relevant for the model have extensive knowledge about the prior uncertainty in the model, but since prior distributions are not easily elicited from experts for all models and approaches, statisticians must choose the prior distributions in more ad-hoc ways. Some knowledge about the model is still required, but as the unknown parameters can be assigned prior distributions, the parameters get some, but not too much freedom. The prior knowledge available for the model in this thesis is used to choose prior distributions and requirements for these priors.

The aim of this thesis is to estimate spatio-temporal temperature trends based on Norwegian monthly average temperature series. All available data should be used, also time series with missing data, and temperature trends for spatial locations where no data is collected are desired. The approach is to use latent Gaussian models in a Bayesian setting. If and how interpretable results can be obtained from the model is studied. The model will not be used for forecasting, but shall fit polynomial temperature curves with spatially varying coefficients to the observed data. The inference should preferably be fast to carry out. A detailed explanation of the model is given in later chapters, and for now, all the information about the model given is that it consists of several components, as fixed effects and spatial fields, where different variations of this model are created by changing the prior distributions. As several different prior distributions are used, minor prior sensitivity analyzes of some parameters, such as range parameters of the GRFs, are carried out. More on this in later chapters.

Note that the approach, the model, and the tools used in this thesis could be replaced with other alternatives. This is discussed towards the end. However, this thesis is written during one single semester, and as the focus is on the estimated temperature trends, other models and approaches is outside the scope of this thesis.

The monthly average temperature data used in the inference and a preliminary analysis of the data are presented in Chapter 2. Chapter 3 contains all necessary background theory for creating the model and to carry out the inference. In Chapter 4 a presentation of the model itself is then given, the results are presented in Chapter 5, and a discussion and a conclusion are given in Chapters 6 and 7, respectively.

Chapter 2

Data and initial explanatory analysis

The temperature data set is now presented and briefly discussed, and is followed by a preliminary explanatory analysis of the data.

2.1 Monthly average temperature data

The data used are monthly average temperature observations from 503 observation locations in the southern part of Norway over 57 years. The data were downloaded from Meteorologisk Institutt, see <http://eklima.met.no>¹, and consists of time series from 1960 to 2016. Figure 2.1 shows the 503 observation locations. UTM32 coordinates are used, and the unit 100 kilometers is used to accommodate a computational constructive and intuitive unit. The altitude for each location is also provided, and varies from at sea level to 2000 meters above sea level.

The time series are not complete, there are missing observations. Only time series with at least one measurement for all the twelve months in a year are included in the analysis, but the series have no other restrictions. 503 locations and 57 years results in more than 340,000 possible measurements, and about 100,000 of these are observed. Figure 2.2 is a plot of the number of observations for each

¹Data downloaded February 2nd and 3rd 2017.

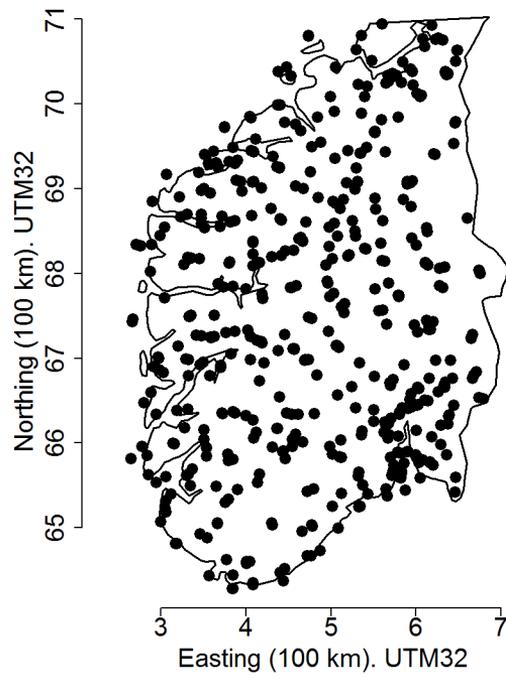


Figure 2.1: The 503 observation locations of the monthly average temperatures.

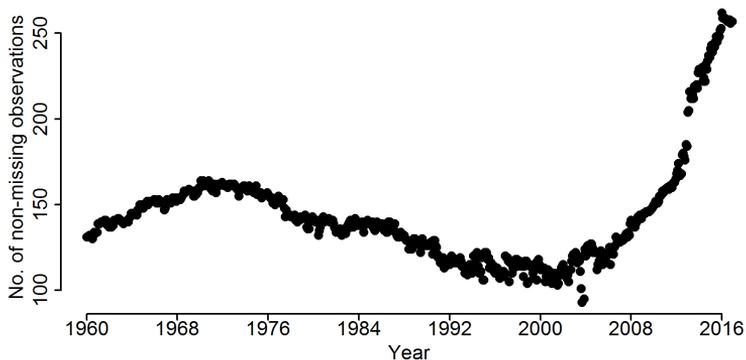
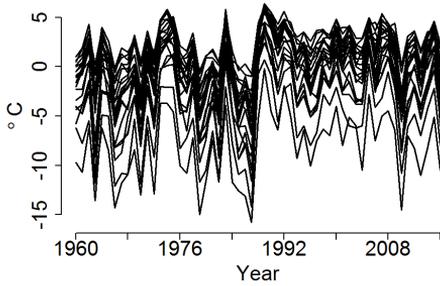


Figure 2.2: Number of (non-missing) temperature observations, plotted against time. The total number of observation locations is 503.

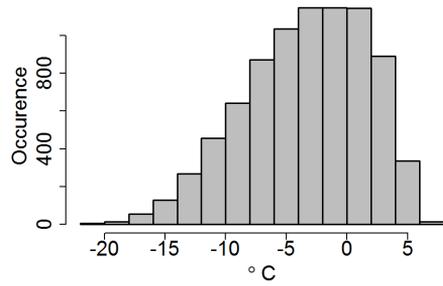
month from 1960 to 2016, and shows that since around year 2000, the number of observation locations has increased rapidly.

Figure 2.3 has plots describing the monthly average temperatures for January. A selection of January monthly mean temperature time series can be seen in Figure 2.3a, and a histogram of all observed measurements is included in Figure 2.3b. They show that the January temperature varies much between years, and that it spans a large interval over the 57 years. Figures 2.3c and 2.3d use, for each observation location, the average January temperature from 1960 to 2016. Missing data are removed from the calculation of the average. The first shows these 57-year-averages as a function of altitude; the relationship looks linear. The second figure is a spatial plot of the 57-year-averages with a prominent spatial pattern. The coastal areas are warmer than the inland in January.

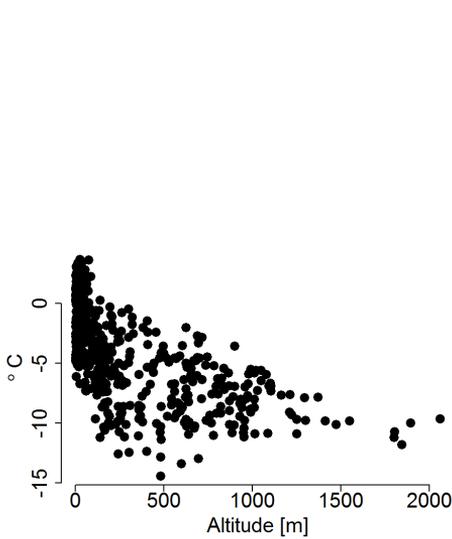
Figure 2.4 shows a boxplot for each of the twelve months of the year, based on the observed monthly average temperatures. The boxes are defined by the 25 % and 75 % quantiles, and the whiskers of the boxes are 1.5 times the interquartile range (the 75 % quantile minus the 25 % quantile) or the minimum or maximum values if there are no outliers. All months have cold outliers, while only the summer months have warm outliers. In addition, the summer months have smaller interquartile ranges and whiskers, which means less variability in the data for those months. The winter months have the most variability in the data. Due to the difference in variability in average monthly temperature between months, each month are treated individually.



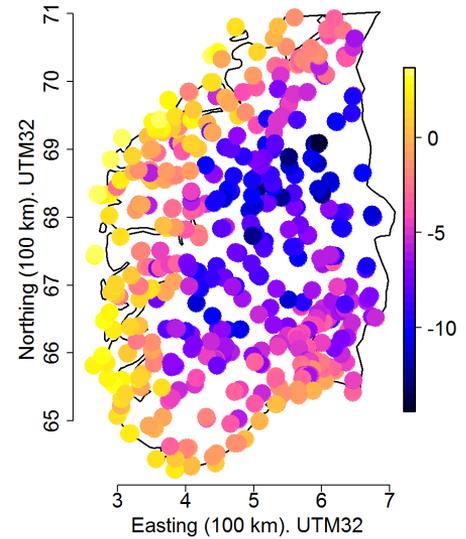
(a) Observed monthly temperature in January for 20 locations from 1960 to 2016. All the displayed time series are complete.



(b) A histogram of the observed temperature measurements for January.



(c) The 57-year-average January temperature from 1960 to 2016 for each location plotted against the corresponding altitude.



(d) The 57-year-average January temperature (in degrees Celsius) from 1960 to 2016 for each location plotted spatially.

Figure 2.3: January monthly average temperatures diagnostics. When the average is calculated, missing data are removed from the calculation.

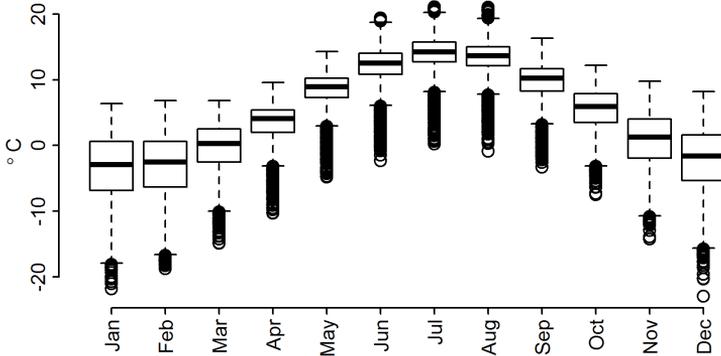


Figure 2.4: Boxplot of monthly average temperatures for all locations and all years.

2.2 Initial data exploration

The natural starting point for modelling the temperature trend in the monthly average temperatures is a simple model. January monthly temperature observations are modelled using the linear regression model

$$y_{j,i} = \beta_{0,i} + a_i t_j + b_i t_j^2 + \varepsilon_{j,i}, \quad i = 1, \dots, n, \quad j = 1, \dots, T, \quad (2.1)$$

where $y_{j,i}$ is the value observed in location s_i in year j . $\beta_{0,i}$ is an intercept, a_i is a linear trend coefficient and b_i is a quadratic trend coefficient. a_i and b_i are multiplied with a time vector $\mathbf{t} = (t_j)$ and \mathbf{t}^2 , respectively. \mathbf{t} goes from 0 to 1 with T equally spaced time point values, which means that the trend curve $a_i \mathbf{t} + b_i \mathbf{t}^2$ begins at zero degrees Celsius in 1960 and ends at $a_i + b_i$ after 57 years in 2016, with $T = 57$. This results in one $\beta_{0,i}$, one a_i and one b_i for observation location.

Some restrictions are put on the data used to perform the linear regression in Equation (2.1). First, January measurements only are used. This is done to avoid problems with seasonal changes between months. This results in up to 57 measurements in each time series. In addition, each time series must have at least one observation the first ten years, and at least one observation the last ten years of the time period. At last, each time series need to have more than 5 observations. This is done to avoid trend curves with a good fit only at a time observations are made. 51 of the 503 time series fulfill these requirements, which is about 10 % of the total number of time series.

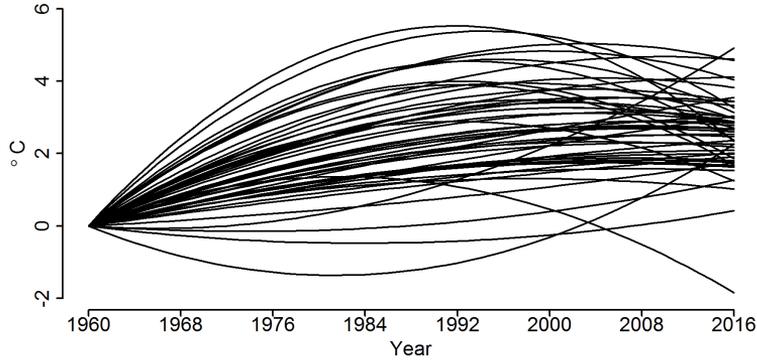


Figure 2.5: The trend curves $a_i t + b_i t^2$, using the coefficients from the linear regression in Equation (2.1). The data is January mean temperature from 1960 to 2016.

Figure 2.5 shows the temperature trend curves $a_i t + b_i t^2$ for the time series fulfilling the requirements listed above. The intercept is not included to make the comparison of the trend curves simpler. Most of the lines agree on the shape, only a few stands out, but none are equal. This indicates that the temperature trend differs between locations, and that there might be spatial dependence in the data. Figure 2.6 shows the sum of a_i and b_i in each location i , which corresponds to the temperature change from 1960 to 2016. Note that this does not tell anything about the temperature evolution. A pattern can be seen; the temperature along the west coast has had a smaller increase than the temperature in the eastern part of Norway. The pattern indicates that this linear model might not be sufficient; a spatio-temporal component is needed.

A spatial model where spatial correlation can be included has the benefit that all time series, also those with few measurements, can be provided to the model since missing observations can be predicated using time series from nearby observation locations. The linear model can only use 10 % of the available temperature time series. It is a pity to declare so much data useless, and in a spatial model, all time series are useful. A spatial model is also able to predict the temperature trend at locations where there are made no observations at all, which gives more complete results. The linear model is too simple and should be extended with spatial components, with one spatial field for each coefficient ($\beta_{0,i}$, a_i and b_i) in addition to the constant coefficients.

The spatial model is in this thesis taken to be a latent Gaussian model, which in addition to including spatial Gaussian random fields, can contain independent

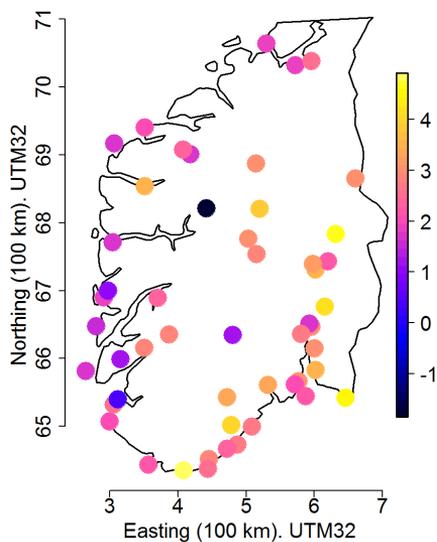


Figure 2.6: Sum of the coefficient estimates $a_i + b_i$ for the linear regression model in Equation (2.1), plotted spatially. The locations are independent. The unit is degrees Celsius.

and identically distributed (iid) components in time, space, or both. An iid-effect in time might be used to correct for unusually cold or unusually warm years. The spatial model is much more complex than the linear regression model, and to use a more complex model means that a more complex tool is needed to do fit the models.

Chapter 3

Background

This chapter is dedicated to all background theory needed to carry out the inference of a model with spatio-temporal components. It is divided into four main topics; Bayesian inference, spatial statistics, prior distributions, and computational tools.

3.1 Bayesian inference

The first main topic is Bayesian inference, what it is, and how it can be used.

3.1.1 Bayesian hierarchical models

Statistical inference is what statisticians call to draw conclusions in the presence of uncertainty (Cressie and Wikle, 2011), and means among others to estimate unknown parameters, and to do predictions of unknown random quantities. This can be done using a statistical model, and in this thesis the focus is on hierarchical models.

Hierarchical statistical models, also called *latent models*, are used to form complex dependence structures through simple components split in multiple levels or layers. Cressie and Wikle (2011) give the following description of a hierarchical model: A data model is at the top, then a process model and at last a parameter model at the bottom. The data model expresses the distribution of the data given an unknown, hidden process. The process model models the uncertainty of

this hidden process. The hidden process can itself consist of several submodels, with uncertainties expressed through conditional probabilities at sublevels. The last layer, the parameter model, consists of probability distributions of all the unknown parameters in the model. These three layers make a *Bayesian hierarchical model*. See e.g. Cressie and Wikle (2011) for a more thorough explanation of Bayesian hierarchical models.

Let $\pi(a)$ be the probability distribution of the random variable A , and let $\pi(\cdot|\cdot)$ be a conditional probability density. Then the Bayesian hierarchical models (BHMs) can be specified by

Data model: $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$

Process model: $\pi(\mathbf{x}|\boldsymbol{\theta})$

Parameter model: $\pi(\boldsymbol{\theta})$

where \mathbf{y} represents the data, \mathbf{x} the underlying process and $\boldsymbol{\theta}$ all unknown parameters. The parameter model corresponds to the *prior distribution*. This gives a joint distribution

$$\pi(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) = \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

and perhaps the most interesting result for this thesis, the conditional distribution of \mathbf{x} and $\boldsymbol{\theta}$ given the data \mathbf{y} , also called the joint *posterior distribution*

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{y})}$$

from Bayes' theorem, see e.g. Casella and Berger (2002). $\pi(\mathbf{y})$ is the marginal distribution of the data model. When doing inference using a Bayesian hierarchical model, the inference is called Bayesian inference. In Bayesian inference, the inference is carried out in a Bayesian setting using prior knowledge and observed data to estimate posterior information (Cressie and Wikle, 2011).

3.1.2 Latent Gaussian models

Structured additive regression models is a class of models where the observation variables y_i are assumed to be from an exponential family. The Gaussian distribution is one of the members of this family, and then the mean μ_i is modelled using a structured additive predictor $\eta_i = \mu_i \forall i$, i.e., $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ is Gaussian with mean $\boldsymbol{\eta}$. For other, non-Gaussian distributions this relationship is not linear, but this will not be discussed.

The predictor η_i is a sum of functions of covariates

$$\eta_i = \beta_0 + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \sum_{k=1}^{n_f} f_k(v_{ki}) + \varepsilon_i \quad (3.1)$$

where β_0 is the intercept, $\{\beta_j\}$ are coefficients of the fixed effects \mathbf{z} , $\{f_j(\cdot)\}$ are unknown functions of the covariates \mathbf{v} , and ε represents unstructured terms. Thanks to the variety of different forms the functions $\{f_j(\cdot)\}$ can take, as a random walk, a spatial field, or an autoregressive process, this yields flexible models with many applications (Rue *et al.*, 2009).

The class of latent Gaussian models is a subset of the Bayesian hierarchical models with a structured additive predictor (Rue *et al.*, 2009) (see Equation (3.1)), where all elements of the predictor η_i are assumed to be Gaussian (Rue *et al.*, 2009). These models have parameters, denoted $\boldsymbol{\theta}$, belonging to the elements of a latent Gaussian field consisting of $\beta_0, \{\beta_j\}, \{f_j(\cdot)\}$ and ε , denoted \mathbf{x} . Note that since η_i is a sum of Gaussian variables, it is itself Gaussian.

3.2 Spatial statistics

As the functions $\{f_j(\cdot)\}$ can be spatial effects, the latent Gaussian processes can be spatial. The second main topic is spatial statistics, and Gaussian random fields and their properties are introduced.

3.2.1 Gaussian random fields

Gaussian random fields (GRFs) have an important role in the statistical society, and are commonly used to model continuous spatial phenomena (Ingebrigtsen, 2014). Stein (1999) describes a Gaussian random field as a real-valued random field $\{X_{\mathbf{s}} : \mathbf{s} \in \mathbb{R}^d\}$ where all finite-dimensional distributions are Gaussian. Gaussian probability distributions are fully specified by their expectation and variance, and thus it is sufficient to specify the mean function $\boldsymbol{\mu}(\mathbf{s}_i)$ and the covariance function $C(\mathbf{s}_i, \mathbf{s}_j)$ for all locations \mathbf{s}_i and \mathbf{s}_j in the domain $\mathcal{D} \in \mathbb{R}^d$ of the field. The expectation has no restrictions, but the covariance function must be positive definite in order to ensure the existence of all the finite-dimensional distributions (Stein, 1999).

A GRF can have several properties, which can be seen easily through the covariance function. A *stationary* GRF has a covariance function dependent on the *difference* in position of two locations only, and not the absolute positions

of each locations. In addition, the mean is a constant function for stationary GRF. An *isotropic* GRF has a covariance function dependent on only the Euclidean distance between the two locations (Lindgren *et al.*, 2011; Abrahamsen, 1997).

For more information on Gaussian random fields, and random fields in general, see e.g. Stein (1999) or Abrahamsen (1997).

3.2.2 Matérn covariance function

Usually the covariance function is chosen from a parameterized family of covariance functions. The most important family of covariance functions is according to Stein (1999) the *Matérn* family. He proposes the Matérn family as a better alternative than common covariance function families as the exponential or Gaussian, due to a flexibility that allows for any degree of differentiability for the random field. Based on this and a thorough analysis, Stein (1999, page 14) concludes with "Use the Matérn model".

A family of covariance functions is usually defined by a marginal variance $\sigma^2 > 0$ and a range $\rho > 0$. The Matérn family is in addition defined by a parameter $\nu > 0$ which controls the degree of differentiability of the field. Thus, the smoothness of the field can be varied without changing the covariance function family, which is an advantage.

Let $|\mathbf{s}_i - \mathbf{s}_j|$ denote the Euclidean distance between the locations \mathbf{s}_i and \mathbf{s}_j in \mathbb{R}^d . The Matérn covariance function is then given by

$$C(\mathbf{s}_i, \mathbf{s}_j) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\sqrt{8\nu}}{\rho} |\mathbf{s}_i - \mathbf{s}_j| \right)^\nu K_\nu \left(\frac{\sqrt{8\nu}}{\rho} |\mathbf{s}_i - \mathbf{s}_j| \right), \quad \mathbf{s}_i, \mathbf{s}_j \in \mathbb{R}^d, \quad (3.2)$$

where $K_\nu(\cdot)$ is the modified Bessel function of second kind and order ν , σ^2 is the marginal variance, and $\rho > 0$ is the range parameter (Lindgren *et al.*, 2011). The range is by Lindgren *et al.* (2011) defined to be the distance where correlations are close to 0.1, and the same definition used here. It can be shown that the Matérn covariance function has the exponential as a special case ($\nu = \frac{1}{2}$), and the Gaussian as a limiting case ($\nu \rightarrow \infty$).

The covariance function $C(\mathbf{s}_i, \mathbf{s}_j)$ can for a given set of locations $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ be used to compute a covariance matrix $\mathbf{\Sigma}$ of size $n \times n$. This covariance matrix has a corresponding precision matrix \mathbf{Q} , which is the inverse of $\mathbf{\Sigma}$. The precision matrix is always possible to compute for a GRF; since the covariance function is positive definite (Rue and Held, 2005), the covariance matrix is also positive definite, and hence invertible.

3.2.3 Gaussian Markov random fields

Continuous Gaussian random fields are popular and can be used to model many natural phenomena. When using GRFs with observations in a given set of locations, the precision matrix \mathbf{Q} of the field must be factorized to solve linear systems of equations and calculating the determinant $\det|\mathbf{Q}|$. The cost of factorizing a dense covariance matrix of size $n \times n$ is $\mathcal{O}(n^3)$, and the computation time needed is often unreasonable (Lindgren *et al.*, 2011). To avoid this high cost, the spatial GRF can be replaced by a spatial Gaussian Markov random field (GMRF), where the factorization of the covariance matrix has a cost of $\mathcal{O}(n^{3/2})$ for two-dimensional spatial GMRFs due to the sparse precision matrix of a spatial GMRF. This makes calculations feasible even for millions of observation locations.

Let x_i and x_j be random variables. x_i and x_j are independent if and only if $\pi(x_i, x_j) = \pi(x_i)\pi(x_j)$, and this is denoted $x_i \perp x_j$. They are *conditionally* independent if and only if $\pi(x_i, x_j | \mathbf{x}_{-ij}) = \pi(x_i | \mathbf{x}_{-ij})\pi(x_j | \mathbf{x}_{-ij})$, which is denoted $x_i \perp x_j | \mathbf{x}_{-ij}$, where \mathbf{x}_{-ij} denotes all of \mathbf{x} except x_i and x_j .

An important result when working with GMRFs is the *Markov property* (Rue and Held, 2005). The Markov property comes in three forms; the pairwise, the local, and the global. Let \mathbf{x} be a set of random variables, and \mathcal{G} a labelled graph with vertices (nodes) \mathcal{V} and edges \mathcal{E} between the node i and j if and only if x_i and x_j are conditionally dependent, i.e., no edge when they are conditionally independent. An edge between node i and node j is denoted $\{i, j\}$.

The *pairwise* Markov property says that for $i \neq j$, x_i and x_j are independent conditioned on \mathbf{x}_{-ij} if $\{i, j\} \notin \mathcal{E}$, i.e., if there is no edge between node i and node

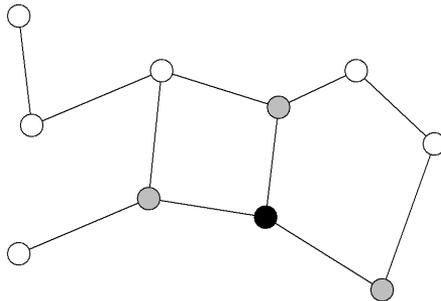


Figure 3.1: An illustration of the local Markov property. The black node and the white nodes are independent given the grey nodes in the neighborhood of the black node.

j. Hence, the pairwise Markov property automatically holds for the graph \mathcal{G} . Given the nodes in the neighborhood of node i , the *local* Markov property says that node i is independent of any node not in this neighborhood, for all $i \in \mathcal{V}$. The graph in Figure 3.1 illustrates the local Markov property: The grey nodes are in the neighborhood of the black node, which makes the black node and the white nodes independent given the grey nodes. The *global* Markov property says that two non-empty sets \mathbf{x}_A and \mathbf{x}_B are independent conditioned on a third set \mathbf{x}_C separating \mathbf{x}_A and \mathbf{x}_B , where \mathbf{x}_A , \mathbf{x}_B and \mathbf{x}_C are disjoint. It can be shown that if \mathbf{x} is a GMRF, the three Markov properties are equivalent (Rue and Held, 2005).

Rue and Held (2005) give in Chapter 2.2, Definition 2.1, the following definition of a GMRF:

Definition 1 (Gaussian Markov random field). *A random vector $\mathbf{x} = (x_1, \dots, x_d)^T$ is called a GMRF with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q} > 0$, if and only if its density has the form*

$$\pi(\mathbf{x}) = (2\pi)^{-d/2} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^d$$

and

$$Q_{ij} \neq 0 \iff \{i, j\} \in \mathcal{E} \text{ for all } i \neq j.$$

In a GMRF, conditional independence $x_i \perp x_j | \mathbf{x}_{-ij}$ is equivalent with $Q_{ij} = 0$. Note that this does not mean that x_i and x_j are marginally independent as well. See e.g. Rue and Held (2005) for more on conditional independence.

The labelled, undirected graph \mathcal{G} can be used to find the structure of the precision matrix \mathbf{Q} . Note that since the graph is undirected, the covariance matrix and hence the precision matrix will be symmetric (Rue and Held, 2005), which it will be in a Gaussian distribution. If a graph \mathcal{G} is fully connected, the corresponding precision matrix \mathbf{Q} is completely dense. To factorize dense matrices has a high cost, but to factorize sparse matrices, where most of the non-diagonal elements are zero, is much cheaper (Rue and Held, 2005).

3.2.4 Stochastic partial differential equations

As a dense precision matrix of a Gaussian random field has a factorizing cost of $\mathcal{O}(n^3)$ they are not computationally practical to use, and cannot be used directly for a high number of observation locations. A computationally efficient way to approximate a GRF with the Matérn covariance function is by using *stochastic*

partial differential equations (Lindgren *et al.*, 2011). A stochastic partial differential equation (SPDE) is a partial differential equation with stochastic terms. The stochastic terms are Gaussian white noise.

An SPDE must be solved numerically and needs to be discretized. The finite element method (FEM) is a good alternative, with an irregular triangulation instead of a regular grid. The irregular triangulation has two benefits worth mentioning: Avoiding interpolation of observation locations to the nearest grid point, and allowing for irregular boundaries and finer resolution where needed (Lindgren *et al.*, 2011). The triangulation consists of a set of non-intersecting triangles where at most a common corner (vertex) or edge of any two triangles meet, if they meet at all (Lindgren *et al.*, 2011). The SPDE is solved in these vertices, and the solution can be interpolated on the triangles. A given triangulation is defined by piecewise linear basis functions ϕ_g , being 1 in vertex g , and zero in all other vertices. Several conditions can be used to define the triangulation, such as maximum length of each triangle edge or the minimum triangle angle. It is common to choose the triangulation so the smallest triangle angle is maximized, which gives a Delaunay triangulation (Lindgren *et al.*, 2011).

In Bayesian modelling of spatial fields, one particularly useful SPDE is the linear fractional SPDE

$$\left(\frac{8}{\rho^2} - \Delta\right) \left(\frac{\rho}{4\sqrt{2\pi}\sigma} u(\mathbf{s})\right) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} = (s_1, s_2) \in \mathbb{R}^2, \quad \rho, \sigma > 0, \quad (3.3)$$

where \mathcal{W} is spatial Gaussian white noise with unit variance (Lindgren *et al.*, 2011). Note that the dimension is set to $d = 2$. The solution to this particular SPDE is a Gaussian random field $u(\mathbf{s})$ with the Matérn covariance function. The parameters can be directly related to the parameters of the Matérn covariance function in Equation (3.2); σ^2 is the marginal variance and ρ is the range parameter. ν has been set equal to 1. Δ is the Laplacian $\Delta = \sum_{i=1}^2 \frac{\partial^2}{\partial s_i^2}$. Lindgren *et al.* (2011) have used this to propose a new method for doing calculations with GRFs in Bayesian modelling. More on this in Section 3.4.2.

3.3 Prior distributions

Any probability distribution can be used as prior distributions, but some will be more appropriate for given parameters than others. For some problems and models there exist guidelines for what priors that are the most suitable. However, this does not exist for the model used in this thesis. When using computational tools for Bayesian inference, such as R-INLA, default prior distributions are implemented. These are seldom appropriate (Simpson *et al.*, 2017), and thus the

prior distributions used in the model in this thesis are investigated and justified before chosen.

In this section, the relatively new concept of penalized complexity priors are introduced and discussed, first in general and then for spatial fields. Then other possible prior distributions that can be used for spatial fields are presented. At last a short presentation of prior distributions required for other model components is included.

3.3.1 Penalized complexity priors

The penalized complexity (PC) prior is a proper prior which penalizes increased complexity in the model (Simpson *et al.*, 2017). The PC prior aims to reduce overfitting, i.e., not include a model component that should not be present. An overfitting prior will create a more flexible model than perhaps necessary, and the base model will not have support in the posterior (Simpson *et al.*, 2017). Simpson *et al.* (2017, p. 5), defines a *base model* as the "simplest" model for a given model component. The base model usually corresponds to the component not being present in the model at all. Further they state that a prior that overfits, in this setting, is a prior with zero mass at the base model. If this is the case, the parameter will be forced to be in the model even though it should not be present.

Simpson *et al.* (2017) give four principles which outline the approach of constructing PC priors for a univariate parameter. A summary of them follows:

1. **Occam's razor.** Simpler models, i.e., the base models, are preferred until there is enough support for a more complex model, and to deviate from the base model should be penalized. Based on the prior alone, the base model is the preferred model.
2. **Measure of complexity.** The Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951) is used as a measure of increased complexity. For the base model g and the more flexible model f , the KLD is defined as

$$\text{KLD}(f||g) = \int f(\mathbf{x}) \log \left(\frac{f(\mathbf{x})}{g(\mathbf{x})} \right) d\mathbf{x},$$

and is the information lost when the base model is used to approximate the more flexible model. When used in practice, the distance measure $t(f||g) = \sqrt{2\text{KLD}(f||g)}$ is used to measure the distance between models g and f .

3. **Constant rate penalization.** For a given $t = t(f||g)$, the penalty for changing the flexible model an additional distance δ from the base model should be independent of t , with a constant decay-rate r . The prior $\pi_t(\cdot)$ of t then satisfies

$$\frac{\pi_t(t + \delta)}{\pi_t(t)} = r^\delta, \quad \delta, t \geq 0, \quad 0 < r < 1.$$

This constant rate penalty assumption implies using an exponential prior on the distance t , such that $\pi_t(t) = \lambda \exp(-\lambda t)$ for $r = \exp(-\lambda)$. Then the prior, on the original space, becomes

$$\pi(\xi) = \pi_t(t(\xi)) \left| \frac{\partial t(\xi)}{\partial \xi} \right|$$

for a connection $t(\xi)$ between the distance and the parameter ξ . The mode of the prior π_t is at $t = 0$, i.e., at the base model.

4. **User-defined scaling.** The user needs to have some idea on either the size of the parameter ξ , or a property of the corresponding model component. λ from above can be found by choosing some U and α so

$$P(Q(\xi) > U) = \alpha.$$

U is a user-defined bound that specifies a tail event, and α is the probability of this event. $Q(\xi)$ is some transformation of the parameter ξ .

These principles can be extended to the multivariate $\boldsymbol{\xi} \in \mathcal{M}$ for a subset \mathcal{M} of a smooth n -dimensional manifold, with base model $\boldsymbol{\xi} = \mathbf{0} \in \mathcal{M}$. For more details on this, see Simpson *et al.* (2017).

3.3.2 Penalized complexity priors on Gaussian random fields

The PC prior can be used on Gaussian random fields (GRF). Fuglstad *et al.* (2017) have derived the PC prior for a GRF with the Matérn covariance function, and their work is used as the foundation for this section. They have derived the PC prior for dimension d and smoothness parameter ν . For $d = 2$ and $\nu = 1$, the Matérn covariance function is given as

$$C(\mathbf{s}_i, \mathbf{s}_j) = \sigma^2 \left(\frac{\sqrt{8}}{\rho} |\mathbf{s}_i - \mathbf{s}_j| \right) K_1 \left(\frac{\sqrt{8}}{\rho} |\mathbf{s}_i - \mathbf{s}_j| \right), \quad \mathbf{s}_i, \mathbf{s}_j \in \mathbb{R}^2,$$

where \mathbf{s}_i and \mathbf{s}_j are locations as before, σ^2 is the marginal variance, and ρ is the range. The base model corresponding to the PC prior for GRFs is $\rho = \infty$ and

$\sigma = 0$, which (in the limit) gives a field with the same value in all locations and no uncertainty. The results of Fuglstad *et al.* (2017) necessary for this thesis is the joint PC prior for the standard deviation and range, given by

$$\pi(\sigma, \rho) = \lambda_\sigma \lambda_\rho \rho^{-2} \exp(-\lambda_\sigma \sigma - \lambda_\rho \rho^{-1}). \quad (3.4)$$

The prior is specified by $P(\sigma > \sigma_0) = \alpha_\sigma$ and $P(\rho < \rho_0) = \alpha_\rho$ which gives

$$\lambda_\sigma = -\frac{\log(\alpha_\sigma)}{\sigma_0} \text{ and } \lambda_\rho = -\log(\alpha_\rho)\rho_0.$$

Note that the marginal prior distributions easily can be extracted from the joint prior distribution in Equation (3.4). The range is inverse exponential distributed with parameter λ_ρ , and the standard deviation is exponential distributed with parameter λ_σ . See Fuglstad *et al.* (2017) for full derivations of the prior for general d and ν .

3.3.3 log-Gaussian priors on Gaussian random fields

An alternative prior to use on the random field parameters is the two-tailed log-Gaussian distribution family. Both the range ρ and the standard deviation σ of the Matérn covariance function of a GRF are then given log-Gaussian priors, which can be rewritten to a multivariate prior. This also allows for prior correlation between the two parameters, but this will not be given any attention.

For a log-Gaussian variable X , the marginal distribution is denoted $\text{logGaussian}(\mu, \sigma^2)$ and given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left(-\frac{1}{2\sigma^2}(\log(x) - \mu)^2\right),$$

with mean $e^{\mu+\sigma^2/2}$ and variance $e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$. The corresponding distribution of $\log(X)$ is the Gaussian with mean μ and variance σ^2 ,

$$f_{\log(X)}(\log(x)) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(\log(x) - \mu)^2\right).$$

3.3.4 Other prior distributions

In the parameter model, variances of model components are included, and they need prior distributions. One common choice is to use a gamma prior on the inverse variances, i.e., the precisions, due to the relationship to the Gaussian

likelihood; in Bayesian inference the gamma prior is the conjugate prior (see e.g. Givens and Hoeting (2013)) for a Gaussian likelihood with known mean, but unknown precision. Denote the precision τ , then the gamma distribution of τ is given by

$$\pi(\tau) = \frac{b^a}{\Gamma(a)} \tau^{a-1} \exp(-b\tau), \quad \tau > 0$$

where the hyperparameters $a > 0$ and $b > 0$ are the shape parameter and the rate parameter, respectively. The gamma distribution has mean a/b and variance a/b^2 . The precision can be given a large prior variance compared to the mean, by letting $b < 1$, and a small prior variance by letting $b > 1$. For $a = 1$ the gamma distribution becomes the exponential distribution. A distribution on the precision may be difficult to interpret, and can be transformed into the distribution of the more interpretable standard deviation using the transformation formula (see e.g. Casella and Berger (2002)). For the precision $\tau \sim \text{Gamma}(a, b)$, the standard deviation $\sigma = 1/\sqrt{\tau}$ has the distribution

$$\pi(\sigma) = \frac{2b^a}{\Gamma(a)} \sigma^{-2a-1} \exp\left(-\frac{b}{\sigma^2}\right).$$

Another possible choice for the prior precision is the PC prior. For the precision of a Gaussian component, the density is

$$\pi(\tau) = \frac{\lambda}{2} \tau^{-3/2} \exp\left(-\lambda\tau^{-1/2}\right), \quad \tau > 0$$

for $\lambda = -\log(\alpha)/U > 0$ (Simpson *et al.*, 2017). U and α are determined by choosing values $P(\sigma > U) = \alpha$, $U > 0$ and $0 < \alpha < 1$, for $\sigma = 1/\sqrt{\tau}$ being the corresponding standard deviation. This leads to σ being exponentially distributed with rate λ , $\pi(\sigma) = \lambda \exp(-\lambda\sigma)$.

Since latent Gaussian models require all elements of the latent field \mathbf{x} to be Gaussian, any fixed effects in such a model have a Gaussian prior distribution, where the prior mean and variance of the fixed effects are chosen.

3.4 Fast inference for Bayesian models

To carry out the inference, the Integrated nested Laplace approximations (INLA) methodology is used. The INLA methodology is implemented in the statistical computing program **R**, where it is called **R-INLA**. The GRFs in the model are approximated using the solution of the SPDE in Equation (3.3), and this approach is integrated in **R-INLA**. In this section, the INLA methodology and the SPDE approach are explained, and for more information see www.r-inla.org.

3.4.1 Integrated nested Laplace approximations methodology

The integrated nested Laplace approximations (INLA) methodology proposed by Rue *et al.* (2009) is a non-sampling based method for doing fast Bayesian inference on latent Gaussian models. INLA utilizes that GMRFs with sparse precision matrices give computational benefits due to the Markov property (see Section 3.2.3). The models that INLA can do inference on are latent Gaussian models where the latent field \mathbf{x} , see Section 3.1.2, is a GMRF (Rue *et al.*, 2009).

For a distribution $\pi(\cdot)$, the Laplace approximation

$$\int_a^b \pi(x) dx \approx \pi(x^*) \sqrt{2\pi\sigma^{2*}} (\Phi(b) - \Phi(a))$$

gives an accurate result, especially away from the tails of $\pi(x)$. The mode of $\pi(x)$ and thus also of $\log \pi(x)$ is $x^* = \operatorname{argmax}_x (\log \pi(x))$, $\sigma^{2*} = -1 / \left. \frac{\partial^2 \log \pi(x)}{\partial x^2} \right|_{x=x^*}$, and $\Phi(a)$ is the cumulative density function of $\text{Gaussian}(x^*, \sigma^{2*})$ evaluated in a . This result is used in the INLA methodology to compute approximate non-standard posterior marginal distributions deterministically.

The notation used here is the same as in Section 3.1.2, but a short recap is nevertheless included here. Let $\mathbf{y} = (y_1, \dots, y_n)$ be a vector of observations, where y_i , $i = 1, \dots, n$ is assigned a likelihood from the exponential distribution family. Since the observations are assumed to be Gaussian in this thesis, the explanation of INLA is for Gaussian likelihoods only. Several steps are then simplified, see e.g. Rue *et al.* (2009) for a full explanation of the INLA methodology. The distribution assigned to \mathbf{y} is specified by the linear predictor $\boldsymbol{\eta} = \boldsymbol{\mu}$ with $\mu_i = \eta_i \forall i$. The linear predictor η_i is as given in Equation (3.1) and in INLA, it is required that η_i is Gaussian conditioned on the parameters $\boldsymbol{\theta}$.

The components of η_i in Equation (3.1) are called latent components, and are gathered in the latent field $\mathbf{x} = (\boldsymbol{\eta}, \beta_0, \boldsymbol{\beta}, \mathbf{f})$. The parameters describing the latent components and the likelihood are denoted $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$.

The observations \mathbf{y} are assumed to be conditionally independent, as in a GMRF, which gives a likelihood $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ that can be written as

$$\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^n \pi(y_i|\eta_i, \boldsymbol{\theta}).$$

This means that each observation y_i depends on one linear predictor η_i of the latent field only. INLA requires that \mathbf{x} is a GMRF, and hence the multivariate

density function of \mathbf{x} is

$$\pi(\mathbf{x}) = \text{Gaussian}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{Q}_{\mathbf{x}}^{-1})$$

where $\boldsymbol{\mu}_{\mathbf{x}}$ is the mean of \mathbf{x} . The precision matrix $\mathbf{Q}_{\mathbf{x}}$ is specified by the parameters $\boldsymbol{\theta}$, and must be sparse. INLA also requires that the number of parameters $\boldsymbol{\theta}$ is small, Rue *et al.* (2017) propose 2 to 5, and not more than 20.

The joint posterior distribution of \mathbf{x} and $\boldsymbol{\theta}$, $\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})$, is proportional to the product of the likelihood and the joint prior distribution of \mathbf{x} and $\boldsymbol{\theta}$ (from Bayes' theorem, see Section 3.1.1 and 3.1.2), where the product can be calculated directly. The results of interest are the marginal posterior distributions of the latent field and of the parameters: $\pi(x_i | \mathbf{y})$ and $\pi(\theta_k | \mathbf{y})$. $\pi(\boldsymbol{\theta} | \mathbf{y})$ can be obtained by

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \frac{\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})}{\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \propto \frac{\pi(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) \pi(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \Bigg|_{\mathbf{x} = \mathbb{E}(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta})} =: \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) \quad (3.5)$$

and can be used to find the marginals $\pi(\theta_k | \mathbf{y})$. When $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$ is obtained, the marginal posterior distributions of the latent field, $\pi(x_i | \mathbf{y})$, can be approximated with the integral

$$\tilde{\pi}(x_i | \mathbf{y}) = \frac{\int \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) \tilde{\pi}(x_i | \boldsymbol{\theta}, \mathbf{y}) d\boldsymbol{\theta}}{\int \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}}.$$

This is done using numerical integration with a set of relevant integration points $\{\boldsymbol{\theta}^{(j)}\}$ and corresponding weights $\{\Delta_j\}$,

$$\tilde{\pi}(x_i | \mathbf{y}) \approx \frac{\sum_j \tilde{\pi}(\boldsymbol{\theta}^{(j)}, \mathbf{y}) \tilde{\pi}(x_i | \boldsymbol{\theta}^{(j)}, \mathbf{y}) \Delta_j}{\sum_j \tilde{\pi}(\boldsymbol{\theta}^{(j)} | \mathbf{y}) \Delta_j}. \quad (3.6)$$

The formulas in Equations (3.5) and (3.6) are the basis for the INLA methodology. The procedure in INLA is as follows:

1. Explore $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$ to obtain a good set of integration points $\{\boldsymbol{\theta}^{(j)}\}$ for the numerical integration. This grid exploration can be done in several ways; Rue *et al.* (2009) proposes the central composite design (CCD) strategy when the number of parameters $\boldsymbol{\theta}$ is more than four, but the CCD strategy gives good results in short time which makes this strategy good also for fewer parameters. In the CCD strategy, the $\boldsymbol{\theta}$ -space is reparametrized to find different values of $\boldsymbol{\theta}$, which is then used to find the optimal values. This is done by:
 - (a) The mode $\boldsymbol{\theta}^*$ of $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$ is found by optimizing $\log(\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}))$. For a Gaussian likelihood, this mode is the mode of the true posterior $\pi(\boldsymbol{\theta} | \mathbf{y})$.

- (b) The negative Hessian \mathbf{H} of $\log(\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}))$ at the modal configuration is calculated.
- (c) Then the eigen-decomposition $\mathbf{H}^{-1} = \boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{V}^T$ is computed.
- (d) The new variable $\boldsymbol{\zeta}$ is defined such that $\boldsymbol{\theta}(\boldsymbol{\zeta}) = \boldsymbol{\theta}^* + \mathbf{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\zeta}$ is fulfilled.

The negative Hessian and the mode are used to select some relevant points in the $\boldsymbol{\theta}$ -space for performing a second-order approximation to a response variable. When the grid has been explored, the posterior marginals $\tilde{\pi}(\theta_k|\mathbf{y})$ for all $k = 1, \dots, K$ are found using an interpolation algorithm based on $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ and $\{\boldsymbol{\theta}_k^{(j)}\}$ (Blangiardo and Cameletti, 2015).

2. Then for each value in $\{\boldsymbol{\theta}_k^{(j)}\}$, the conditional posteriors $\tilde{\pi}(x_i|\boldsymbol{\theta}^{(j)}, \mathbf{y})$ are evaluated on a grid of values for x_i , and the marginal posteriors $\tilde{\pi}(x_i|\mathbf{y})$ can be obtained using numerical integration as in Equation (3.6).

When all the marginal distributions are computed, all results are found. As this methodology does not involve sampling or any other stochastic operations, the results will be the same every time the inference is carried out, unlike e.g. MCMC algorithms.

To summarize, the INLA methodology consists of three steps; first the posterior marginals of the parameters are approximated, second the conditional posteriors of the latent components are approximated, and at last the two first steps are combined using numerical integration to approximate the posterior marginals of the latent components.

3.4.2 SPDE approach in R-INLA

The stochastic partial differential equation (SPDE) approach by Lindgren *et al.* (2011) utilizes that the solution to the SPDE in Equation (3.3) is a Gaussian random field with the Matérn covariance function, which can be approximated by a discretely indexed Gaussian Markov random field. As GMRFs are important and heavily utilized in INLA, the SPDE approach fits the INLA framework well. The SPDE approach exploits the best properties of a GRF and the best of a GMRF; the flexible and interpretable structure of the former with few and meaningful parameters, and the computational benefits of the latter.

The SPDE is in R-INLA solved using the finite element method (FEM) on a Delaunay triangulation (Lindgren *et al.*, 2011). The observation locations are used as vertices where appropriate, and additional vertices are added to satisfy the chosen constraints on the triangulation. The triangulation gives a mesh where the SPDE is solved, and the mesh is the domain of the resulting GRF.

The GRF is a part of the model through the predictor $\boldsymbol{\eta}$. Since the predictor consists of measurements from the observation locations, each η_i must be linked to the vertices in the mesh. A projector matrix \mathbf{A} is used to transform $\boldsymbol{\eta}$, and the SPDE model can then be treated as an indexed random effect (Lindgren and Rue, 2015). \mathbf{A} is created using the basis functions ϕ_g defining the mesh. The new predictor $\boldsymbol{\eta}^* = \mathbf{A}\boldsymbol{\eta}$ replaces the old $\boldsymbol{\eta}$, and the likelihood is then linked to the latent field through $\boldsymbol{\eta}^*$ instead of $\boldsymbol{\eta}$. This must be done for each GRF in the model.

Since the SPDE is solved using the FEM, boundary conditions are needed, and they are in R-INLA set to be Neumann boundary conditions (Lindgren *et al.*, 2011). These boundary conditions might not be the appropriate one in all cases, and then the solution at and near the boundary is not necessarily correct. Near the boundary, the estimated variance will be larger than the rest of the domain. Lindgren and Rue (2015) give a "rule of thumb": at a distance corresponding to the range ρ from the boundary the boundary effects will be negligible. This means that the boundary should preferably be at a certain distance from the observation locations and the domain of interest, and this can be fulfilled by using a mesh extending outside the domain of interest. Another solution to avoid problems around the boundary is to use other boundary conditions. This is not implemented in R-INLA and hence not studied.

To extend the domain to a larger area leads to a larger number of triangles and vertices, which makes the computation more time-consuming. Since the solution of the SPDE outside the domain of interest is not important for other reasons than to avoid boundary effects, the mesh can be divided into an inner part (the domain of interest) and an outer part. Then the resolution in the outer part of the mesh can be lower than the resolution of the inner part, which reduces the computation time. Keep in mind that the shape of the triangles near the inner boundary should not be too narrow, so the difference in resolution cannot be too large. The Delaunay triangulation will to some extent aim to have smooth transitions between small and large triangles (Lindgren *et al.*, 2011), but caution must be taken by the user.

The process of choosing a mesh might be difficult. When aware of the boundary effects, the user can create one inner and one outer mesh, with appropriate resolutions. Even though the observation locations are used as vertices, some triangles might get a high density of observation locations. As long as the number of observation locations in one triangle is small, this will not be a problem, but it is necessary to use a resolution avoiding many observation locations inside one single triangle. In R-INLA, the mesh must be specified using either a set of locations or a domain, and for both the inner and outer mesh, either the largest allowed triangle edge length, or the maximum number of vertices allowed in the

mesh must be provided. These conditions are not necessarily absolute, but will be adhered to best possible. Several additional choices can be made, for more on this see www.r-inla.org.

This explanation of the SPDE approach in **R-INLA** is only concerning the main aspects, for the complete picture the reader is referred to Lindgren *et al.* (2011) and Lindgren and Rue (2015).

Chapter 4

Method

4.1 Comments on R-INLA

The INLA methodology implemented in R (R-INLA) is used to carry out the inference in this thesis. This is a complex library, but naturally restrictions exist which affects the possibilities of the model. The amount of prior distributions implemented is as of 2017 limited, and to simplify the implementation of the model only built-in priors are used in this thesis. Both the PC prior and the log-Gaussian prior for SPDE models are implemented in R-INLA, as well as the PC prior and the gamma prior for precisions of Gaussian variables. For fixed effects, the Gaussian prior is implemented. Other alternatives exist, but are not considered.

Looking at the way the PC prior for the range ρ of a GRF is constructed, at least one possible problem arises. The range can be restricted with either an upper bound or a lower bound, and the latter is preferred, but not both. If one wishes to use a prior that restricts the range from both sides, the PC prior might not be a good choice. The right tail of the inverse exponential distribution, which is the marginal prior for ρ in two dimensions and $\nu = 1$, is heavy. The base model for the range is at $\rho = \infty$, which favors large range values, and is not a good alternative for all data. The PC prior on the standard deviation of a GRF is for $d = 2$ and $\nu = 1$ the exponential distribution, which has lighter tails than the inverse exponential, but also here the prior has only one tail. The base model is $\sigma = 0$, which may give problems if a field with large variations over short distances is preferred. For a large range, the boundary effects will affect the model, which is not desired. Thus, an alternative to the PC prior is considered; a log-Gaussian

prior which can be restricted from both sides. Both prior distribution families are investigated.

The model studied is now presented. First, the model and its components are introduced, and then prior distributions on the parameters are discussed and justified. In the end, three different combinations of prior distributions are presented.

4.2 The model

Let $j = 1, \dots, T$ denote year, $m = 1, \dots, 12$ denote month of the year, where $m = 1$ is January, $m = 2$ is February and so on, and \mathbf{s}_i denote observation location. The observed temperature measurement for year j and month m , and in observation location \mathbf{s}_i is denoted $y_{j,m,i}$. $i = 1, \dots, n$, where $n = 503$, but only some are observed in each month and year.

The likelihood of the temperatures is assumed to be Gaussian. There are several reasons for this: The interpretation of the parametrization in the Gaussian distribution family is straight forward; the mean and variance are used directly as parameters. The Gaussian distribution is computationally fast to work with, and for complex models this is an advantage. The monthly average temperature data is the average of daily observations, which will at least be close to Gaussian. At last, the posterior distribution of the likelihood is not as interesting as the posterior distribution of the underlying process, which includes the temperature trend, and then there are no strong arguments against the Gaussian distribution.

The observations are monthly average temperatures, denoted $y_{j,m,i}$ for year j , month m and location \mathbf{s}_i , and are assumed to be Gaussian conditionally on the mean, with precision τ_y ,

$$y_{j,m,i} \sim \text{Gaussian}(\mu_{j,m,i}, \tau_y^{-1}), \quad i = 1, \dots, n, \quad m = 1, \dots, 12, \quad j = 1, \dots, T.$$

The mean $\mu_{j,m,i}$ is modelled using the predictor $\eta_{j,m,i} = \mu_{j,m,i}$. The predictor $\boldsymbol{\eta}$ is now used instead of the mean $\boldsymbol{\mu}$. The linear regression model from Section 2.2, Equation (2.1), is used as a starting point when selecting the model components. An intercept and a spatial field to capture spatial variations for each month are included to make a present temperature trend more prominent. This field has constraints integrating it to zero, and can be thought of as a spatial intercept for the temporal process. The altitude of each location is included as a linear effect as the temperature decreases almost linearly with height (Andrews, 2010), which makes this a suitable assumption. Figure 2.3c from Section 2.1 shows a

linear relationship between the altitude and the observed temperature, and a fixed effect is the simplest way of including altitude in the model.

To model the temperature trend in time, linear and quadratic basis functions are used. This is done using scalars as fixed effects common in all locations, one for the linear and one for the quadratic basis function. In addition, two spatial fields capturing the linear and quadratic spatial variation in time are included. Both these fields have integrate-to-zero-constraints. The scalars can then be thought of as the "intercepts" of these fields, or as the average linear and quadratic change in the whole domain during the time period. At last, a temporal independent and identically distributed (iid) component is included as an effect in time to account for an unusually temperate month in the whole domain.

For a given year j , month m and location \mathbf{s}_i , let $\beta_{0,m}$ and $u_m(\mathbf{s}_i)$ denote scalar and spatial intercepts, respectively, h_i denote altitude with coefficient β_1 , a_m and b_m denote scalars and $u_{A,m}(\mathbf{s}_i)$ and $u_{B,m}(\mathbf{s}_i)$ denote spatial fields used to model the temperature trend, and $\varepsilon_{j,m}$ denote the temporal iid-effect. The predictor of the model is given as

$$\eta_{j,m,i} = \beta_{0,m} + \beta_1 h_i + u_m(\mathbf{s}_i) + a_m t_{j,m} + b_m t_{j,m}^2 + u_{A,m}(\mathbf{s}_i) t_{j,m} + u_{B,m}(\mathbf{s}_i) t_{j,m}^2 + \varepsilon_{j,m}, \quad j = 1, \dots, T, \quad m = 1, \dots, 12, \quad i = 1, \dots, n$$

where $\mathbf{t} = (t_{j,m})$ is a vector of time points. For easy interpretation of the results, \mathbf{t} has equal spaced elements from 0 to 1, and is of the same length as the time series with monthly average temperatures. \mathbf{t} then has no unit. The field u_m is the random field included to account for spatial variations in the monthly average temperature each year, i.e., areas that are always colder or warmer in a given month m are taken into account, giving a temperature trend not depending too much on spatial variation occurring every year. $u_{A,m}$ and $u_{B,m}$ are the random fields estimating the spatio-temporal trend. $u_{A,m}$ is the deviation in each location from the linear change modelled by a_m , and $u_{B,m}$ is the deviation in each location from the quadratic change modelled by b_m .

All fixed effects ($\beta_{0,m}$, β_1 , a_m and b_m) are assumed Gaussian with zero mean. $\varepsilon_{j,m}$ is the temporal iid-effect, included to capture unstructured variation not relevant for the trend itself. $\varepsilon_{j,m}$ is equal in all locations \mathbf{s}_i for a given point in time, and is assumed Gaussian with mean 0 and precision τ_ε . The temporal iid-effect is constrained with

$$\sum \varepsilon_{j,m} = 0, \quad \sum \varepsilon_{j,m} t_{j,m} = 0 \quad \text{and} \quad \sum \varepsilon_{j,m} t_{j,m}^2 = 0$$

to prevent it from capturing the linear and quadratic effects $a_m \mathbf{t}$ and $b_m \mathbf{t}^2$ are included to account for. The spatial fields u_m , $u_{A,m}$ and $u_{B,m}$ are all assumed to

be Gaussian random fields, with zero mean and the Matérn covariance function presented in Section 3.2.2.

Each month is treated individually, i.e., the index m is held constant. This means that the vector \mathbf{t} has length $T = 57$. The intercept $\beta_{0,m}$ and the field u_m would take care of the seasonality if all months were included in the same model, but due to the difference in temperature variability between months there is reason to believe that the temperature trend differs as well (see Figure 2.4). Thus a_m and b_m , and $u_{A,m}$ and $u_{B,m}$, are assumed unique for each month, and only τ_y , β_1 and $\varepsilon_{j,m}$ are common for all months. The fixed effect β_1 is assumed to be mostly driven by the data, so is the precision of the observations τ_y , and $\varepsilon_{j,m}$ is the temporal iid-effect and has no correlation between the time points. A common model for all months and years would with these assumptions contain 36 Gaussian random fields, which leads to a number of parameters well over the recommended limit of 20 (Rue *et al.*, 2017). To have the observation precision, one fixed effect and one iid-effect as the only common components for the whole model is unnecessary and would likely lead to a higher total runtime than when each month is treated individually.

u_m is measured in degrees Celsius. $u_{A,m}(\mathbf{s}_i)$ and $u_{B,m}(\mathbf{s}_i)$ are representing change in both space and time, but as \mathbf{t} is scaled to have no unit, these fields have unit degrees Celsius as well. Keep in mind that all three fields are constrained to integrate to zero.

The GRFs $u_{A,m}$ and $u_{B,m}$ are spatio-temporal interactions. They are simple to add to the model using R-INLA, but an option for this exact approach does not exist. To include these spatio-temporal interactions, the projector matrix \mathbf{A} from $\boldsymbol{\eta}^* = \mathbf{A}\boldsymbol{\eta}$ (see Section 3.4.2) must be linked to the time-weights \mathbf{t} for each spatial field. For a given month, a given field is equal for all years (as there is only one GRF for each basis function), but the connection to $\boldsymbol{\eta}$ and thus \mathbf{y} changes with time. To get the model in R-INLA link the GRFs to $\boldsymbol{\eta}$ and \mathbf{y} correctly, \mathbf{A} is simply multiplied with the vector \mathbf{t} with the time weights, where each weight is repeated n times; one time for each observation location, regardless of the number of missing observations.

To investigate different prior distributions, varying priors on the model components are used. Before presenting the specific choices, the prior distributions used are justified. To simplify the notation, the month index m is excluded when not discussing a specific month, with one exception: u_m keeps the index to clearly distinguish it from u_A and u_B . In the model, all twelve months are treated

individually, and the predictor in the simplified notation is

$$\eta_{j,i} = \beta_0 + \beta_1 h_i + u_m(\mathbf{s}_i) + at_j + bt_j^2 + u_A(\mathbf{s}_i)t_j + u_B(\mathbf{s}_i)t_j^2 + \varepsilon_j, \quad j = 1, \dots, T, \quad i = 1, \dots, n \quad (4.1)$$

for each month $m = 1, \dots, 12$, where $y_{j,i} \sim \text{Gaussian}(\eta_{j,i}, \tau_y^{-1})$.

4.3 Prior choices

It is a difficult task to give prior distributions to the parameters of a statistical model. A prior is the distribution of a given parameter before any data are observed, and a guide for what priors to use when does not always exist. In some cases there is common practice to use certain priors, but this is not the situation here. However, some prior information usually exists, and can be used to choose suitable priors on at least some model components. For other components the priors must be chosen based on what the parameter or component should do in the model. For instance, Norwegian monthly average temperatures are analyzed in this thesis, and monthly averages lower than -25 or higher than 25 degrees Celsius has not occurred in the time period from 1960 to 2016 as can be seen in Figure 2.4 in Section 2.1. This is information that can be used to put priors on the parameters. Note that even though each month is treated individually, all are given the same prior distributions on all model parameters.

Both prior distribution families and the hyperparameters used for the distributions must be decided. In this thesis, three groups of prior distribution families are studied, yielding three model variations, where the prior sensitivity is investigated to some degree. All variations of the model contain the same components. All priors are justified, but none are elicited from experts. A pragmatic, engineering approach is used in addition to the prior knowledge, due to numerical problems occurring when fitting models with certain prior distribution. The prior distributions found using an engineering approach still meet the chosen requirements of the priors.

4.3.1 Priors on the fixed effects and precision

The fixed effects β_0 , a and b are given Gaussian priors with mean 0 and variance 100. The coefficients are without any restrictions, but all have the unit degrees Celsius and none of them should become smaller than about -30 or larger than +30 for Norwegian temperature data. β_0 is the intercept and values outside [-30,30] is not likely, while a temperature trend modelled using linear or quadratic

coefficients outside this interval is not desired. More specific information about β_0 , a and b is not known prior to data collection, and since most of the mass of a Gaussian distribution (more than 99 %) lies within three standard deviations of the mean, this prior seem appropriate. The altitude coefficient β_1 is also given a Gaussian prior with mean 0, but has variance 1000. The unit of this coefficient is degrees Celsius per meter, and intuitively it is small and mostly driven by the data, which makes a strong prior unnecessary.

The precision of the Gaussian observations, τ_y , is given a gamma prior. For the Gaussian observations, the precision accounts for the measurement error made for all observations, and is given a prior with parameters $a = b = 0.5$. This corresponds to a mean of 1 and a variance of 2. This parameter is assumed to depend mostly on the data, but will not become especially large (this is equivalent to a small variance) since there will always be measurement error, and a prior with a larger mean does is not considered necessary.

The precision of the temporal iid-effect τ_ε is given two different priors. One of them is the gamma distribution. The hyperparameters for this prior are more difficult to choose than for τ_y ; they depend on how the temperature changes "randomly" from one year to the next for a given month. The hyperparameter values for the gamma prior for τ_ε are $a = 1$ and $b = 0.9$. The corresponding prior for the standard deviation of the temporal iid-effect has most of the distribution mass between 0 and 4, which is reasonable for monthly average temperature for any month of the year when measurement error is accounted for (see the boxplot in Figure 2.4 in Section 2.1).

The other prior used on τ_ε is the PC prior. The PC prior for precision of a Gaussian random variable is presented in Section 3.3.4, and the parameter λ is decided by giving an upper tail probability to the corresponding standard deviation, $P(\sigma > U) = \alpha$. Since the PC prior penalizes the model for including the component it is used on, the base model is $\sigma_\varepsilon = 0$, i.e., not have a temporal iid-effect in the model. What prior the standard deviation of the temporal iid-effect should have is difficult to decide, and the tail probability is set to be $P(\sigma > 5) = 0.01$, to avoid that the temporal iid-effect component becomes too big, but at the same time give it the possibility to be significantly bigger than zero.

4.3.2 Priors on the Gaussian random fields

How the results change with different priors on the parameters of the Gaussian random fields u_A and u_B , and especially the ranges, is investigated. All three GRFs are assumed to have zero mean, which means the covariance function

parameters (standard deviation σ and range ρ) are the ones that require prior distributions. What priors the standard deviation and range for each field should get depends on the random field and what it is supposed to do in the model; in this model the field u_m , and u_A and u_B have different purposes and should probably have different priors. But deducing which priors are the most appropriate is not straight forward.

The range and standard deviations of the random fields u_m , u_A and u_B are given different priors between models; both different families of distributions and different hyperparameters. The PC prior and the log-Gaussian prior are tested, and possibilities for both are discussed. The joint PC prior for the range and standard deviation of a GRF with Matérn covariance function for $\nu = 1$ and $d = 2$ is presented in Section 3.3.2, and the joint prior can be split into marginal priors; ρ and σ are not correlated. With a log-Gaussian prior, both ρ and σ are log-Gaussian, and can be dependent in the prior. In this thesis, they are assumed independent in the prior.

Recall that u_m is a GRF included to capture spatial variations occurring every year for a given month. Thus the field u_m should have a short range, relative to the observation domain, so small areas with lower or higher average temperature than the surroundings, are detected by the field. The standard deviation is assumed to be small since variations between years and decades are accounted for in other components of the model. The range should however give all locations at least one neighbor, and the prior must be set accordingly. Due to this, the range and standard deviation for u_m are ideally given priors with small probabilities of being large, and the range should have a small probability of being small as well.

u_A is a linear and u_B is a quadratic spatio-temporal interaction. To give the standard deviation and especially the range of the GRFs specific prior distributions is not straight forward due to minimal amount of prior knowledge on the parameters. Thus the parameters of these two fields are given the same prior distributions in all models. It is reasonable to have a range smaller than the domain size, and large enough to give all locations at least one neighbor, just as for u_m . More specific values than this is difficult to find, due to a lack of prior knowledge on the parameters. The standard deviation gets a prior similar to the standard deviation prior for u_m .

The domain of observation locations can be seen in Figure 2.1 in Section 2.1. The largest distance from one observation location to the closest neighbor is around 50 kilometers, and the largest distance between two observation locations is approximately 700 kilometers. Having posterior mean range values smaller than 50 kilometers or larger than 700 kilometers is not desirable; hence the prior

distributions to be studied are those for which most of the mass is in this interval. A smaller range than 50 kilometers is almost equivalent to white noise in space, a larger range than 700 kilometers leads to a slowly varying field compared to the domain size, and in both cases, the need for the random field is reduced. In addition, the boundary effects may affect the results when the range is too large (Lindgren and Rue, 2015). The standard deviations of all three GRFs are assumed to be around or less than 3 degrees Celsius. Since observation noise and temporal iid-effects are accounted for by other model components, it is assumed that no larger standard deviation is needed.

For ρ_A and ρ_B , a more specific interval for the prior distributions than [50, 700] kilometers of the prior distribution is difficult to find. For ρ_m , however, the interval can be more specific based on the desired properties of this GRF. This field is included in the model to capture spatial variation occurring every year in a given month m , and for the field to be able to include small areas with different temperature, only prior distributions of ρ_m with most of the mass between 50 and 300 kilometers are considered.

4.4 Model summary

Different prior distributions are used to create five variations of the model with predictor in Equation (4.1). The observations are assumed to be Gaussian, and the priors on τ_y , β_0 , β_1 , a and b are kept equal in all model variations. That leaves the priors on the range and standard deviation of the three random fields, and the precision of the temporal iid-effect to vary between models. The following list is a summary of the prior distribution families and hyperparameters that are equal in all models:

- $\tau_y \sim \text{Gamma}(0.5, 0.5)$,
- $\beta_0, a, b \sim \text{Gaussian}(0, 100)$, and
- $\beta_1 \sim \text{Gaussian}(0, 1000)$.

Three main model variations are studied with the following priors:

Model 1 PC prior on the field parameters and gamma prior for the precision of the temporal iid-effect,

Model 2 log-Gaussian priors on the field parameters and gamma prior for the precision of the temporal iid-effect, and

Model 3 log-Gaussian priors on the field parameters and PC prior for the precision of the temporal iid-effect.

To simplify the notation when using PC prior distributions, the notation $PC(U, \alpha)$ is used to denote the PC prior with tail event U with corresponding probability α , see Sections 3.3.2 and 3.3.4. Keep in mind that interpretation of U and α differs for different parameters. Note that even though the model variations are usually referred to as different models, only the prior distributions differ between them, all variations consist of the same model components.

The hyperparameters used for each distribution family is based on Section 4.3 together with a pragmatic engineering approach, and are found in Table 4.1. ρ_m and σ_m are the parameters of u_m , ρ_A and σ_A are the parameters of u_A , and the same for u_B . The unit of the range is given in 100 kilometers, which is used to get an easy interpretable distance scale, and the unit of the standard deviations are degrees Celsius, for all three Gaussian random fields. The distribution families named model 3 (log-Gaussian priors on ρ and σ for all three GRFs, and PC prior on τ_ε) has three different sets of hyperparameters, which is used to among others a minor prior sensitivity analysis. The three variations of model 3 are referred to as *submodels* 3a, 3b and 3c. This yield five model variations.

As Table 4.1 shows, the parameters of the Gaussian random fields u_A and u_B are always given the same prior distributions on both range and standard deviation. Keep in mind that the PC prior is specified by a lower bound for range, and an upper bound for standard deviations. The log-Gaussian distribution has mean and variance $e^{\mu+\sigma^2/2}$ and $e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$, respectively. A posterior distribution

Param.	Model 1	Model 2		
ρ_m	PC(2, 0.1)	logG(0.4, 0.08)		
σ_m	PC(3, 0.1)	logG(0.3, 0.2)		
ρ_A, ρ_B	PC(2, 0.1)	logG(1.2, 0.02)		
σ_A, σ_B	PC(3, 0.1)	logG(0.3, 0.2)		
τ_ε	Gamma(1, 0.9)	Gamma(1, 0.9)		
Model 3				
Param.	Submodel 3a	Submodel 3b	Submodel 3c	
ρ_m	logG(0.4, 0.08)	logG(0.4, 0.08)	logG(0.4, 0.08)	
σ_m	logG(0.3, 0.2)	logG(0.3, 0.2)	logG(0.3, 0.2)	
ρ_A, ρ_B	logG(1.2, 0.18)	logG(1.2, 0.08)	logG(1.2, 0.02)	
σ_A, σ_B	logG(0.3, 0.3)	logG(0.3, 0.3)	logG(0.3, 0.3)	
τ_ε	PC(5, 0.01)	PC(5, 0.01)	PC(5, 0.01)	

Table 4.1: Prior distribution families and hyperparameters in the different model variations. $PC(\cdot, \cdot)$ denotes the PC prior, which has different interpretations between parameters, $\logG(\cdot, \cdot)$ is the log-Gaussian prior.

does not only depend on the prior, but the data and other model components as well. Prior distributions provide some knowledge about a given parameter to the model, and ideally, for the model used in this thesis, the posterior has some common features with the prior, but a lower standard deviation.

Chapter 5

Results

In this chapter, all results from the inference are presented. Some discussion on details for the different models is included, while a more thorough discussion is given in the next chapter. As many of the results are similar for all models and submodels, only the results that stand out are presented for each individual model. The results that are unique for each model are mainly the range parameters and only those are displayed for each model and submodel. Supplementary figures can be found in Appendix A, and the individual graphs are referred to when needed.

The Gaussian random fields represented using the SPDE method require meshes. The same mesh is used in all model variations, and can be seen in Figure 5.1. See Figure 2.1 in Section 2.1 for the specific observation locations. The observation locations are located inside the inner part of the mesh, and the outer part is included to avoid unwanted boundary effects. The total number of nodes in the mesh is 2480.

Note that even though the priors are put on the *precisions*, when the results are presented the *standard deviations* are plotted and discussed, due to easier interpretation of the latter. To clarify which months specific results come from, the subscript m is re-introduced on parameters and components when needed. If month m is not specified, the parameter or model component in general is referred to. Note that u_m denotes the GRF in general, while u_1 denote the field for January. In addition, a superscript is introduced to separate the five model variations, i.e., $\rho_{A,3}^{[2]}$ is the range belonging to the GRF $u_{A,3}^{[2]}$ for model 2 and March (third month of the year). When the superscript is absent, it is the parameter in general that is discussed, and not for a particular model. The

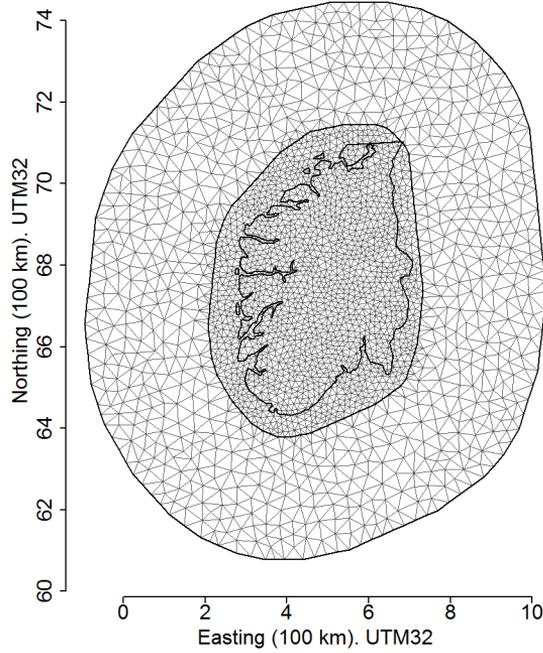


Figure 5.1: The mesh used when solving the SPDEs in all model variations and for all GRFs. The shape of the southern half of Norway is added to the mesh.

superscripts used are 1, 2, 3, 3a, 3b and 3c, where 3 refers to model 3 in general, with no specific hyperparameters.

One of the desired properties of the model is a low computation time. The inference is for a given month carried out in between 15 and 40 minutes. The runtime will naturally depend on the computer, and on other processes occurring at the same time as the inference is carried out. The runtime also differs between models; model 1 and 2 are slower than model 3. About 20 minutes is the average computation time required for submodels 3a, 3b and 3c.

5.1 Numerical problems

Numerical problems during the inference have sometimes occurred when INLA is calculating the negative Hessian \mathbf{H} of the log-likelihood, see steps 1a, 1b and

1c from the explanation of INLA in Section 3.4.1. The algorithm in R is not able to successfully find the mode of the log-likelihood, which means the mode found does not correspond to a maximum. This leads to an incorrect calculation of the negative Hessian which leads to wrong estimates of posterior marginals. There can be several reasons for why such problems occur. As an example, consider a two-dimensional log-likelihood which is very steep on one side of the true mode, and has a slow decline on the other side. Then the numerical optimization algorithm might find the wrong mode due to the odd shape of the log-likelihood. Other shapes of the likelihood leading to difficulties finding the mode also exists. Whether it is a problem with the shape of the log-likelihood or something else that goes wrong is not investigated further; when the INLA algorithm is not able to find the correct mode, the model with those specific hyperparameters for a given prior distribution family has been discarded.

That numerical problems occur for certain prior distributions has become a part of choosing prior distributions. Some of the parameters in the model require stronger priors than others because there is not enough information about them in the data, and it is this lack of information that leads to numerical problems. As weak priors are not usable on all model components, the prior choices are limited to stronger priors, and some of these strong priors are chosen using an engineering approach given the guidelines from Section 4.3. This is not the optimal way of doing Bayesian inference, but the resulting prior distributions used follows the guidelines, and the stronger prior distributions give interpretable model results. None of the five model variations with prior distributions presented in Table 4.1 in Section 4.4 have experienced numerical problems.

The range parameters of the GRFs, and especially ρ_A and ρ_B , are given more attention than the other parameters of the model. Recall that a posterior range much smaller than 0.5 or much larger than 7 is not acceptable in the model in this thesis. Thus, a model having posterior mean of any range parameter outside $[0.5, 7]$ is not considered a good model. For the field u_m , the range should preferably be smaller than 3. A model with range values far outside these limits is regarded as insufficient.

Note that the problems with calculation of the Hessian could have been solved by e.g. letting some of the unknown parameters θ be fixed, but the focus is on finding working prior distributions instead. Problems concerning locating the mode and/or calculating the Hessian will be referred to merely as numerical problems from now on. No further attention will be given.

The results from all model variations are now presented. The posterior range distributions are presented for each model individually, and in Section 5.3 the rest of the results are included, using submodel 3b as a representative for all five

models. The models are presented and described in Section 4.4, but a short recap is nevertheless included here. All models consist of the components in Equation (4.1). Common for all models is Gaussian priors on fixed effects and a gamma prior on the observation precision. Model 1 is given PC priors on all ranges and standard variations, and a gamma prior on the precision of the temporal iid-effect. Model 2 is given log-Gaussian priors on all ranges and standard deviations, and a gamma prior on the precision of the temporal iid-effect. Model 3 is given log-Gaussian priors on all ranges and standard deviations, and a PC prior on the precision of the temporal iid-effect. The specific hyperparameter values can be found in Table 4.1.

5.2 Results differing between models

The results differing between the five model variations are now presented. The model variations deemed as insufficient (models 1 and 2) are presented first, while the satisfying model 3 with submodels follows.

5.2.1 Insufficient models

Model 1 has the PC prior on the Gaussian random field parameters. Since the PC priors are restricted from one side only and the model without the parameter is the most probable, these priors lead to too large posterior mean values of ρ_A and ρ_B . Model 2 is to a large extent limited due to numerical problems, which makes a prior sensitivity analysis difficult, and the prior distributions need to be stronger than what is necessary for model 3. Hence neither model 1 nor model 2 fulfill all desired properties of the model. Even though some are fulfilled, they are both regarded as insufficient. When the results of these two models are presented, only the most interesting results are displayed. Several of the parameters have similar posteriors in all five models and submodels, and those not included here will be later.

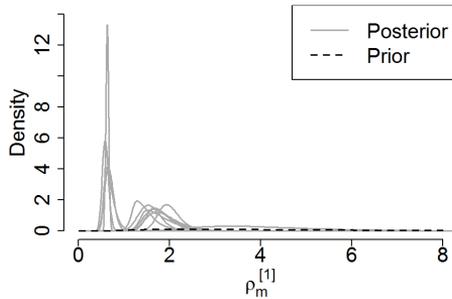
Model 1

Model 1 has the following prior distributions on the GRF parameters and the precision of the temporal iid-effect:

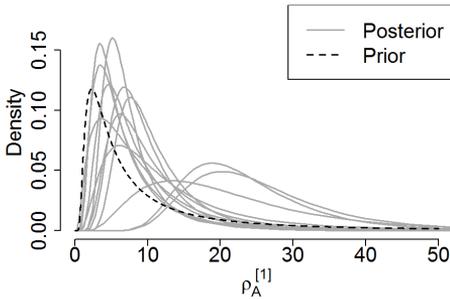
$$\rho_m^{[1]}, \rho_A^{[1]}, \rho_B^{[1]} \sim \text{PC}(2, 0.1), \sigma_m^{[1]}, \sigma_A^{[1]}, \sigma_B^{[1]} \sim \text{PC}(3, 0.1), \tau_\varepsilon^{[1]} \sim \text{Gamma}(1, 0.9),$$

i.e., all ranges and all standard deviations of the GRFs have the same prior distribution. Note that the PC prior is not using the lower limit of 0.5 on the range, but a larger number. This is to get a slightly stronger prior to avoid numerical problems, and is not violating the chosen guidelines.

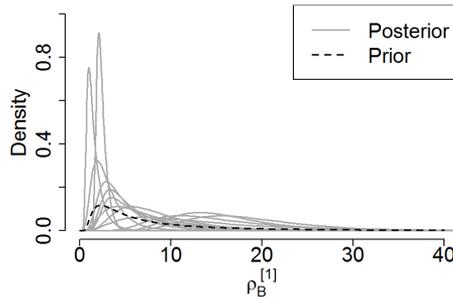
The PC prior of the range is preferably restricted with a lower bound. The range of any GRF is for this model not intended to be much larger than 700 kilometers (corresponding to 7 in the chosen unit), and an upper bound is especially important for the range prior for this model. However, an upper bound does not fit into the PC prior framework, where a large range is favored. It is expected before doing any inference that this model does not fit the temperature data particularly well, and this is confirmed by the results.



(a) Prior and posterior distributions of the range parameter $\rho_m^{[1]}$.



(b) Prior and posterior distributions of the range parameter $\rho_A^{[1]}$.



(c) Prior and posterior distributions of the range parameter $\rho_B^{[1]}$.

Figure 5.2: Prior and posterior marginal distributions of the range parameters of the three GRFs for model 1. Posteriors of all twelve months of the year are plotted, but which is which is not indicated.

Figure 5.2 shows the posterior marginal distributions of $\rho_m^{[1]}$, $\rho_A^{[1]}$ and $\rho_B^{[1]}$ for model 1, along with the prior distribution. The most interesting result from these graphs is how much the posterior distributions for a given parameter differs between months. This justifies the choice of treating each model individually.

All the ranges $\rho_m^{[1]}$ in Figure 5.2a have a prior where the probability of being smaller than 2 is 0.1. Despite of that, the majority of the ranges u_m^1 have modes at smaller range values than 2, and several are at smaller values than 1. None are smaller than 0.5, and only June has a posterior range with mode larger than 3. Thus, this parameter has a satisfying posterior for almost all months.

$\rho_A^{[1]}$ and $\rho_B^{[1]}$ in Figures 5.2b and 5.2c, however, have posteriors with much larger modes than intended. The mode of $\rho_{A,7}^{[1]}$ (July) is almost 20, corresponding to 2000 kilometers and three times the largest distance between two observation locations. Then the boundary effects affect the results, which is not ideal, and the resulting GRF varies slowly in space with a large variance. For some months, the posterior ranges $\rho_A^{[1]}$ have modes smaller than 7, which are acceptable, but this is not true for all months. The posterior marginals $\rho_B^{[1]}$ do not have as extreme modes values as $\rho_A^{[1]}$, but the posteriors are not satisfying for all months. For both ranges, the summer months do in general have larger posterior ranges than the winter months.

The rest of the posterior marginals are not displayed here, some are included in Appendix A and are discussed later. They look similar to the posterior marginals of the other model variations. Even though model 1 is sufficient for some months of the year, it is not sufficient for *all* months, and will not be used to anything else than comparison with other models from now on.

Due to the use of the PC prior on all GRF parameters, model 1 is included in this thesis even though it does not fulfill all the requirements a sufficient model is decided to fulfill. Before giving any restrictions on the posterior distributions of the parameters, the idea of the PC prior fits the model well due to the lack of prior knowledge on the random fields. The problem is that the PC prior favors the simplest model without the fields, which is one of the advantages of the PC prior, while it is desired to have all three fields in the model. The PC prior framework is not suitable for all the random fields of this model, as it contradicts the decision of having all of u_m , u_A and u_B present in the model.

Model 2

Model 2 has the following prior distributions on the GRF parameters and the precision of the temporal iid-effect:

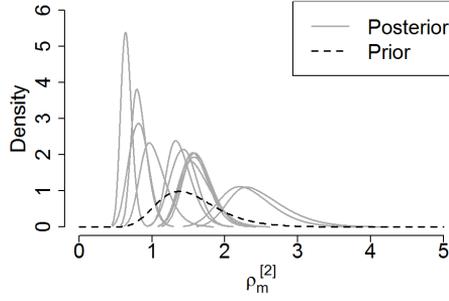
$$\rho_m^{[2]} \sim \text{logGaussian}(0.4, 0.08), \rho_A^{[2]}, \rho_B^{[2]} \sim \text{logGaussian}(1.2, 0.02), \\ \sigma_m^{[2]}, \sigma_A^{[2]}, \sigma_B^{[2]} \sim \text{logGaussian}(0.3, 0.2), \tau_\varepsilon^{[2]} \sim \text{Gamma}(1, 0.9).$$

The prior distribution for the range parameter of $u_m^{[2]}$ is different from the priors on $u_A^{[2]}$ and $u_B^{[2]}$. The priors for the range parameters of the two latter fields are strong, which is required to avoid numerical problems. The numerical problems indicate that the model may be overparameterized, it is then difficult to identify the parameters, and to make the model work it is necessary to use a strong prior.

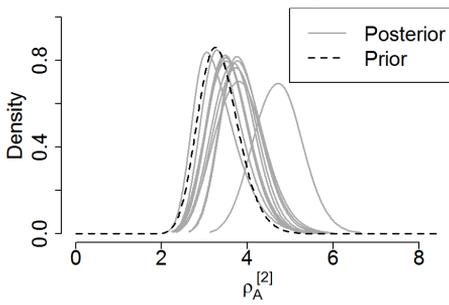
Figure 5.3 shows the prior and posterior marginal distributions of the ranges $\rho_m^{[2]}$, $\rho_A^{[2]}$ and $\rho_B^{[2]}$ for model 2. The posterior ranges of $\rho_m^{[2]}$ in Figure 5.3a vary more than the posteriors of $\rho_m^{[1]}$ in Figure 5.2a, but also here all posterior modes lie above 0.5, and no now posterior modes are larger than 3. With the log-Gaussian prior, it is possible to use stronger prior distributions than with the PC prior, which here leads to more satisfying posterior distributions on the range parameters of $u_A^{[2]}$ and $u_B^{[2]}$. All posteriors $\rho_m^{[2]}$ have a lower standard deviation than the prior, which means that the data \mathbf{y} has removed some uncertainty from the parameter.

The posterior marginals for $\rho_A^{[2]}$ and $\rho_B^{[2]}$ are found in Figures 5.3b and 5.3c. Both parameters are given the same prior, which is strong with 99 % of the mass between 2.3 and 4.8. Therefore, it is not a surprise that these posterior ranges have modes well within the interval $[0.5, 7]$. The posterior distributions are similar to each other, which is a result of the strong prior. Only one posterior stands out from Figures 5.3b and 5.3c; August, which in both plots has the largest mode.

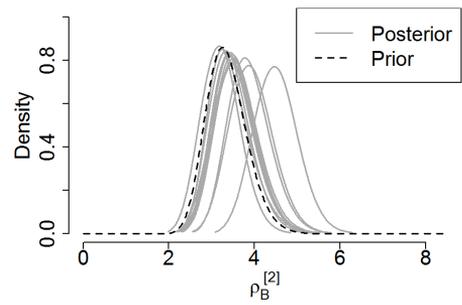
As for model 1, the posterior distributions of the other parameters do not have any features out of the ordinary and are discussed later, with some graphs included in Appendix A. Even though this model gives desired results, the procedure of setting priors was difficult due to the occurrence of numerical problems when doing inference with R-INLA. This limits the prior distribution choices and makes a prior sensitivity analysis problematic. Therefore model 2 is as model 1 deemed insufficient and the investigation of this model ends here.



(a) Prior and posterior distributions of the range parameter $\rho_m^{[2]}$.



(b) Prior and posterior distributions of the range parameter $\rho_A^{[2]}$.



(c) Prior and posterior distributions of the range parameter $\rho_B^{[2]}$.

Figure 5.3: Prior and posterior marginal distributions of the range parameters of the three GRFs for model 2. Posteriors of all twelve months of the year are plotted, but which is which is not indicated.

5.2.2 The preferred model

Model 3 has log-Gaussian priors on the range and standard deviation of the Gaussian random fields and the PC prior on the precision of the temporal iid-effect, with three different sets of hyperparameters denoted submodel 3a, 3b and 3c. The GRFs u_A and u_B have parameters that are difficult to estimate, and model 3 is chosen above the other models because the prior distributions for model 3 are strong enough to avoid parameter values leading to numerical problems during the inference.

The prior on the temporal iid-effect precision $\tau_\varepsilon^{[3]}$ is kept constant through the three submodels 3a, 3b and 3c, so are the priors on both range and standard

deviation of $u_m^{[3]}$ and the priors on the standard deviations $\sigma_A^{[3]}$ and $\sigma_B^{[3]}$. The prior distributions on the fixed effects and the observation precision $\tau_y^{[3]}$ are kept constant here as in models 1 and 2. This leaves the ranges on $u_A^{[3]}$ and $u_B^{[3]}$, which are the only prior distributions changing between the submodels 3a, 3b and 3c. The priors on $\rho_A^{[3]}$ and $\rho_B^{[3]}$ are changed simultaneously, and for a given submodel they have the same prior. Note that only the hyperparameters differs between the submodels, the prior distribution family is the same. The prior distributions on $\rho_A^{[3]}$ and $\rho_B^{[3]}$ in the submodels are, compared to each other, weak, medium, and strong for submodels 3a, 3b and 3c, respectively. $u_m^{[3]}$ denotes the GRF u_m in model 3 in general, not relating it to one of the specific submodels 3a, 3b or 3c. The three submodels are now presented separately with the unique results first, and then some results common for all five models and submodels are presented, using submodel 3b as a representative for the others.

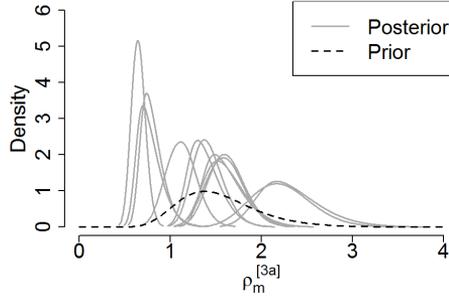
Submodel 3a

Submodel 3a has the following prior distributions on the GRF parameters and the precision of the temporal iid-effect:

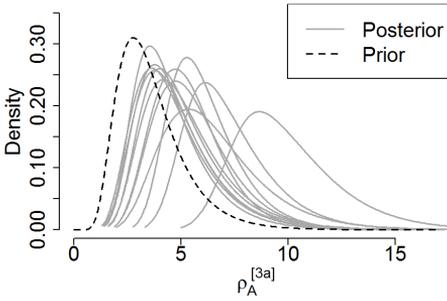
$$\begin{aligned}\rho_m^{[3a]} &\sim \text{logGaussian}(0.4, 0.08), \quad \rho_A^{[3a]}, \rho_B^{[3a]} \sim \text{logGaussian}(1.2, 0.18), \\ \sigma_m^{[3a]} &\sim \text{logGaussian}(0.3, 0.2), \quad \sigma_A^{[3a]}, \sigma_B^{[3a]} \sim \text{logGaussian}(0.3, 0.3), \\ \tau_\varepsilon^{[3a]} &\sim \text{PC}(5, 0.01)\end{aligned}$$

The posterior marginal distributions of $\rho_m^{[3a]}$, $\rho_A^{[3a]}$ and $\rho_B^{[3a]}$ can be seen in Figures 5.4a, 5.4b and 5.4c, respectively. The first mentionable detail is that even though the range of $u_m^{[3a]}$ is slightly weaker than in model 2, the posterior distributions of $\rho_m^{[3a]}$ in Figure 5.4a are almost indistinguishable from $\rho_m^{[2]}$ in Figure 5.3a. This indicates that ρ_m is, at least to some extent, driven by the data.

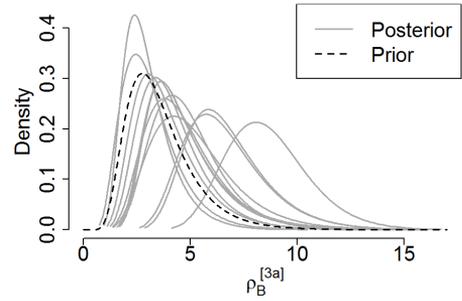
This submodel has weak priors for $\rho_A^{[3]}$ and $\rho_B^{[3]}$ relative to submodels 3b and 3c, with 99 % of the mass between 1.1 and 9.9. It is worth noting that all posterior distributions $\rho_A^{[3]}$ have larger modes than the prior, and also have larger standard deviations. The only month that does not fulfill the wanted properties of the range parameters ρ_A and ρ_B is August, which has a mode above 7 for both parameters. Even though that makes this submodel unsatisfying, it is not far away from being acceptable, and a slightly stronger prior on ρ_A and ρ_B will give a satisfying model.



(a) Prior and posterior distributions of the range parameter $\rho_m^{[3a]}$.



(b) Prior and posterior distributions of the range parameter $\rho_A^{[3a]}$.



(c) Prior and posterior distributions of the range parameter $\rho_B^{[3a]}$.

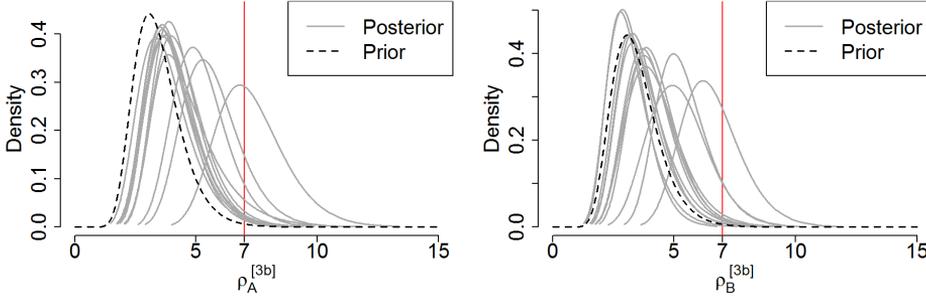
Figure 5.4: Prior and posterior marginal distributions of the range parameters of the three GRFs for submodel 3a. Posteriors of all twelve months of the year are plotted, but which is which is not indicated.

Submodel 3b

Submodel 3b has the following prior distributions on the GRF parameters and the precision of the temporal iid-effect:

$$\begin{aligned} \rho_m^{[3b]} &\sim \text{logGaussian}(0.4, 0.08), \quad \rho_A^{[3b]}, \rho_B^{[3b]} \sim \text{logGaussian}(1.2, 0.08), \\ \sigma_m^{[3b]} &\sim \text{logGaussian}(0.3, 0.2), \quad \sigma_A^{[3b]}, \sigma_B^{[3b]} \sim \text{logGaussian}(0.3, 0.3), \\ \tau_\varepsilon^{[3b]} &\sim \text{PC}(5, 0.01). \end{aligned}$$

Figure 5.5 shows the posterior marginal distributions of $\rho_A^{[3b]}$ and $\rho_B^{[3b]}$ for submodel 3b. Here the priors for ρ_A and ρ_B are slightly stronger than in submodel 3a, with 99 % of the mass between 1.9 and 5.7, and Figures 5.5a and 5.5b show



(a) Prior and posterior distributions of the range parameter $\rho_A^{[3b]}$.

(b) Prior and posterior distributions of the range parameter $\rho_B^{[3b]}$.

Figure 5.5: Prior and posterior marginal distributions of the range parameters of $u_A^{[3b]}$ and $u_B^{[3b]}$ for submodel 3b. Posteriors of all twelve months of the year are plotted, but which is which is not indicated.

that the both parameters and all months have posterior range modes less than 7. Note that the posterior of $\rho_A^{[3b]}$ with the largest mode (which is August) is skewed and has more mass in the right tail than the left, which gives a mean slightly above 7, but as the mean is so close to 7, and the mode is below, it is accepted. As for submodel 3a, the posterior distributions of the range $\rho_A^{[3b]}$ have both larger modes and standard deviations than the prior. Again, August is the month with the highest posterior range for both u_A and u_B .

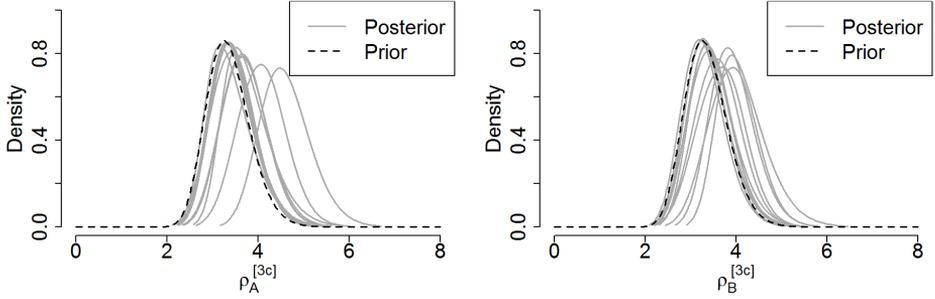
The posterior marginals of $\rho_m^{[3b]}$ are not included here, as they are so similar to the posterior marginals of $\rho_m^{[3a]}$ that they are barely distinguishable. They are however included in Figure A.1 in Appendix A.

Submodel 3c

Submodel 3c has the following prior distributions on the GRF parameters and the precision of the temporal iid-effect:

$$\begin{aligned} \rho_m^{[3c]} &\sim \text{logGaussian}(0.4, 0.08), \quad \rho_A^{[3c]}, \rho_B^{[3c]} \sim \text{logGaussian}(1.2, 0.02), \\ \sigma_m^{[3c]} &\sim \text{logGaussian}(0.3, 0.2), \quad \sigma_A^{[3c]}, \sigma_B^{[3c]} \sim \text{logGaussian}(0.3, 0.3), \\ \tau_\varepsilon^{[3c]} &\sim \text{PC}(5, 0.01) \end{aligned}$$

The last submodel studied in this thesis has the strongest priors on ρ_A and ρ_B of the submodels 3a, 3b and 3c, which are the same priors used on $\rho_A^{[2]}$ and $\rho_B^{[2]}$



(a) Prior and posterior distributions of the range parameter $\rho_A^{[3c]}$. (b) Prior and posterior distributions of the range parameter $\rho_B^{[3c]}$.

Figure 5.6: Prior and posterior marginal distributions of the range parameters of $u_A^{[3c]}$ and $u_B^{[3c]}$ for submodel 3c. Posteriors of all twelve months of the year are plotted, but which is which is not indicated.

in model 2. The posterior distributions of $\rho_A^{[3c]}$ and $\rho_B^{[3c]}$ can be seen in Figure 5.6. That this is a strong prior is clear; the majority posterior distributions are barely distinguishable from the prior distribution for both range parameters, and the posteriors have almost no mass in the distribution at range 7. Figure 5.6a displays the posteriors of $\rho_A^{[3c]}$, which are deviating more from the prior than the posteriors of $\rho_B^{[3c]}$ seen in Figure 5.6b. The other model components affect the posteriors of ρ_A and ρ_B ; the posteriors for these parameters for model 2 in Figures 5.3b and 5.3c differ from those for submodel 3c, even though the prior is the same. For completeness, it is mentioned that as in submodels 3a and 3b, also here August is the month with the largest ranges, both ρ_A and ρ_B . The posterior distributions of $\rho_m^{[3c]}$ are omitted from this chapter due to the strong resemblance to $\rho_m^{[3a]}$, and can be seen in Figure A.1.

Since a posterior range less than 700 kilometers has been such an important requirement, one could here be wondering about why model 1 and submodel 3a are included in this thesis at all. The reason is simple: To show that a range of 700 kilometers or less is not guaranteed for all months. Most months have posterior ranges (for all GRFs) less than 700 kilometers, but the summer months with relatively little variation in the data have posterior ranges of u_A and u_B that are too large. In Bayesian inference, a prior distribution should not control a parameter too much, but the decided restrictions state that the posterior means of ρ_A and ρ_B should be less than 700 kilometers. A discussion should thus not be about whether a model is unfairly discarded, but if the demand of a posterior range less than 700 kilometers should be removed. As this has been thoroughly

justified, the decision about discarding models with too large posterior range stands.

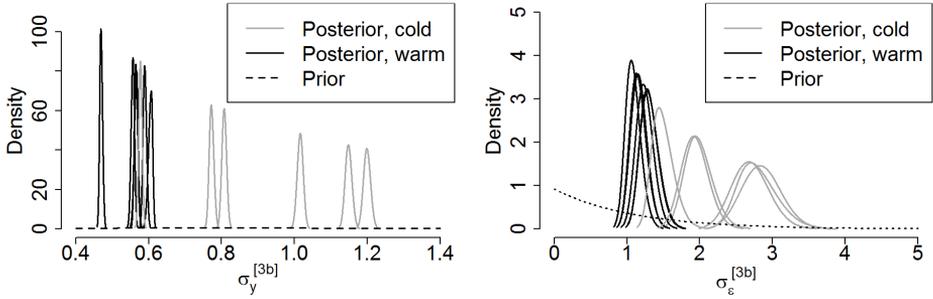
5.3 Results similar in all models

Until now, only the posterior marginal distributions of the range parameters have been discussed. Now the rest of the parameter posteriors are presented and discussed, but due to the many similarities between the models, only the results from one of the models are shown in most cases. Submodel 3b is chosen to represent the posteriors.

Due to posterior range values of $u_A^{[3b]}$ and $u_B^{[3b]}$ that fulfill the desired requirements, without a too strong prior, submodel 3b is considered to be the best model. A strong prior is not necessarily negative, and is in certain cases needed, but the difference in the posteriors between the months of the year becomes smaller for stronger priors (see e.g. Figures 5.4b, 5.5a and 5.6a), and these differences between months are an important part of the result.

The posterior standard deviations of the observations, σ_y , and of the temporal iid-effect, σ_ε , for submodel 3b are presented first. They can be seen in Figure 5.7. Both are heavily dependent on the data, and especially for σ_y , there are almost no differences between the models. See Figure A.2 for the posterior of σ_y for the other four models. Figure 5.7a shows the posterior marginals of $\sigma_y^{[3b]}$ for all months in submodel 3b. To visualize the difference between the colder and warmer months of the year, the months are divided into two groups; the cold and warm months. The cold months are defined as October to March, and the warm months as April to September. The difference is clear: The cold months have higher observation standard deviations than the warm months. This is as expected, as the cold months have more variation in measurements than the warm months, see the boxplot in Figure 2.4 in Section 2.1. The only exception is $\sigma_{y,10}^{[3b]}$ for October, which has a mean of about 0.6 degrees Celsius.

Figure 5.7b displays the posterior marginals of σ_ε . Also here the warm months have lower standard deviations than the cold. This parameter has different prior distribution through the models. This naturally affects the posterior some, but the differences between models 1 and 2, and model 3, are small. The posterior distributions of σ_ε for model and submodels 1, 2, 3a and 3c can be seen in Figure A.3. The standard deviation of the temporal iid-effect is mostly driven by the data.



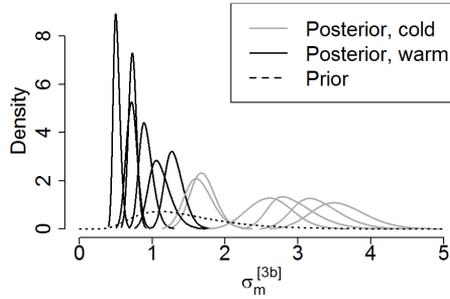
(a) Prior and posterior distributions of the standard deviation $\sigma_y^{[3b]}$ of the observations. The months are divided into two groups.

(b) Prior and posterior distributions of the standard deviation $\sigma_\varepsilon^{[3b]}$ of the temporal iid-effect. The months are divided into two groups.

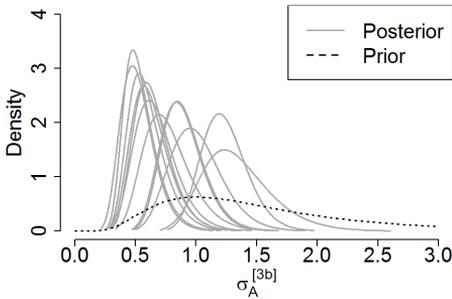
Figure 5.7: Prior and posterior marginal distributions of the standard deviations of the observations and the temporal iid-effect for all twelve months of the year for submodel 3b. The warm months are taken to be April to September, while the cold are October to March.

Figure 5.8 displays the posterior standard deviations of the three Gaussian random fields for submodel 3b, and A.4 display the same posteriors for models 1, 2 and submodels 3a and 3c. These posterior distributions vary between the models, the resemblance between all but model 1 is prominent, but model 1 clearly differs from the others. Model 1 has the PC prior on the standard deviation of all fields, while models 2 and 3 have the log-Gaussian, and a difference in posterior distributions is not surprising. In general, the posterior distributions for model 1 have larger modes of σ_m , σ_A and σ_B for a given month, than the other models.

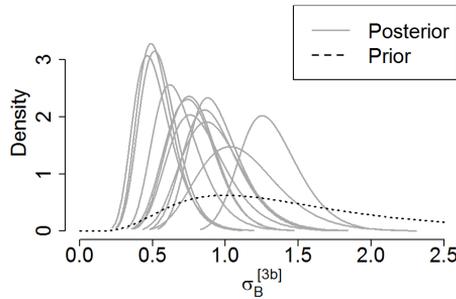
The posterior distributions of $\sigma_m^{[3b]}$ are found in Figure 5.8a. As for σ_y and σ_ε , the cold (October to March) and warm (April to September) months have the largest and smallest posterior means of $\sigma_m^{[3b]}$, respectively, and this is also the case for the other models (see Figure A.4). For the posterior standard deviations of u_A and u_B , however, no such pattern is apparent, and the months are not divided into groups, neither here nor in the appendix. The posteriors of $\sigma_A^{[3b]}$ and $\sigma_B^{[3b]}$ can be seen in Figures 5.8b and 5.8c, respectively. The standard deviations are in general higher for σ_m than the two others, and this is true for all models. Recall that August often had the highest range value for both ρ_A and ρ_B in most models. Here, August has one of the highest, if not the highest, standard deviation of u_A and u_B in all models.



(a) Prior and posterior distributions of the standard deviation $\sigma_m^{[3b]}$ of the GRF $u_m^{[3b]}$. The months are divided into two groups.



(b) Prior and posterior distributions of the standard deviation $\sigma_A^{[3b]}$ of the GRF $u_A^{[3b]}$. Posteriors of all twelve months of the year are plotted, but which is which is not indicated.



(c) Prior and posterior distributions of the standard deviation $\sigma_B^{[3b]}$ of the GRF $u_B^{[3b]}$. Posteriors of all twelve months of the year are plotted, but which is which is not indicated.

Figure 5.8: Prior and posterior marginal distributions of the standard deviations of the Gaussian random fields for all twelve months of the year for submodel 3b. The warm months are taken to be April to September, while the cold are October to March.

The remaining parameters not presented are the fixed effects β_0 , a , b and β_1 . The latter is the coefficient for the altitude in the model, and has unit degrees Celsius per meter. The prior of β_1 is Gaussian(0, 1000) in all models, and despite the weak prior, all models highly agree on the posterior distributions of β_1 for each month. Figure 5.9a shows the mean values of $\beta_1^{[3b]}$ with 0.025 and 0.975 quantiles for submodel 3b for each month. However, if β_1 from one of the other models had been used instead, there would not be any visible change in the plot, see Figure

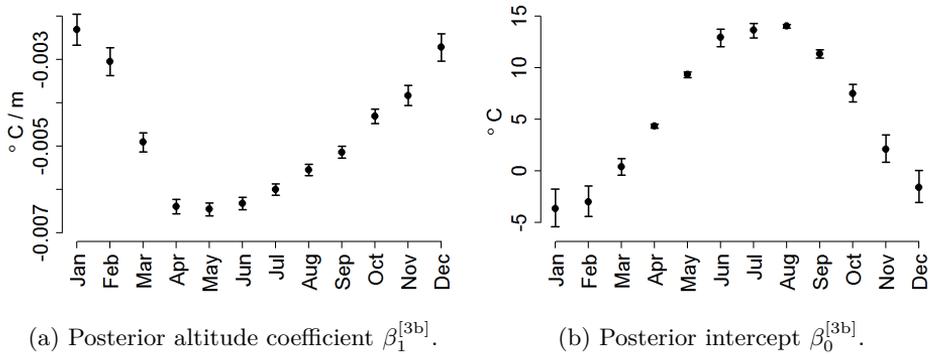


Figure 5.9: Posterior mean values and 0.025 and 0.975 quantiles for the fixed effects $\beta_1^{[3b]}$ and $\beta_0^{[3b]}$ for all twelve months of the year for submodel 3b.

A.5 in Appendix A. The coefficient values are small, but $\beta_1 = -0.005$ implies a difference of 0.5 degrees Celsius over 100 meters in altitude. The coefficient clearly should be in present in the model.

The intercept $\beta_0^{[3b]}$ for submodel 3b is displayed with 0.025 and 0.975 quantiles in Figure 5.9b. Figure A.6 displays the intercept with quantiles for the four remaining models (1, 2, 3a and 3c). The prior is Gaussian with zero mean and a variance of 100. The estimates of the coldest four months are the most uncertain, and the resemblance between this plot and the boxplot in Section 2.1, Figure 2.4, is clear. The other models agree on the posterior mean values, and all but model 1 also agree on the quantiles. Model 1 has for most months larger quantiles than the others, and has more uncertain estimates of β_0 (see Figure A.6).

The linear and quadratic temperature trend coefficients, a and b , respectively, are not included graphically here, but can for all five models be seen in Figures A.7 (for a) and A.8 (for b). Their values are not particular interesting alone, and thus only some appreciable details are included. Models 2 and 3 highly agree on both the mean values of the coefficient estimates and their 0.025 and 0.975 quantiles measuring the estimate certainty. Model 1 agrees to some degree on the mean values, but the quantile values are much larger than in the other models, and only one of the twelve months have a significant estimate of $a^{[1]}$, unlike models 2 and 3 that has only two non-significant estimates, where both estimates are close to zero. This again justifies the decision about discarding model 1. Non-significant in this context means that 0 is somewhere in the interval between the 0.025 and 0.975 quantiles.

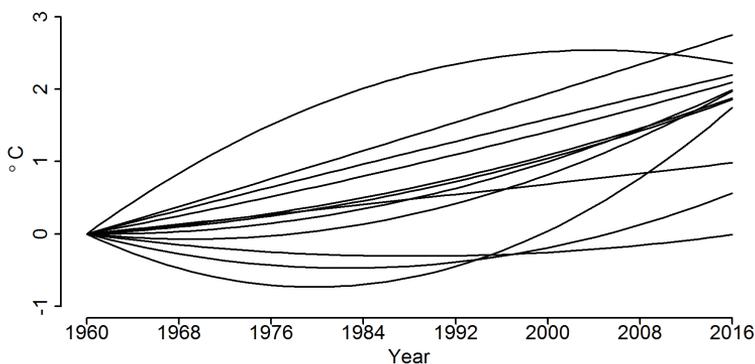


Figure 5.10: Posterior temperature trend $a^{[3b]}t + b^{[3b]}t^2$ for submodel 3b. The posterior trend of all twelve months of the year are plotted, but which is which is not indicated.

As all parameters in the model now have been presented, it is time for the estimates of the Gaussian random fields and the temperature trends. The simplest version of the posterior temperature trend, $at + bt^2$, is plotted in Figure 5.10, where as usual submodel 3b represents all models. As the mean estimates of a and b for a given month are similar in all models, naturally the trend curves are as well. Model 1 has values that differs from the others, which leads to slightly different curves. The trend curves from the other models are not included in the thesis, but the parameters a and b for all models (including submodel 3b) can be seen in Figures A.7 and A.8, respectively, in Appendix A. The trend curves in Figure 5.10 supports the decision about treating each month individually: the curves have various shapes, and the estimated temperatures in 2016 spans almost 3 degrees Celsius. June ends at the lowest temperature of all months in 2016 at 0 degrees Celsius, while January has had the largest increase and ends at almost 2.4 degrees. Remember that Figure 5.10 displays the curves made from the fixed effects only, and that spatial variations are not yet considered. This leads to an incomplete picture of the temperature trend, which is greater in some areas in the southern half of Norway, and smaller in others.

The need for both the linear and quadratic temperature trend coefficient could be questioned here, as some curves are almost linear, while others would manage with only the quadratic coefficient. However, this model should be applicable to *all* months of the year, with the same components and prior distributions. To remove one of them would clearly benefit some months, but it would affect other

months negatively. Then it is good that the coefficients will be estimated to zero if they are superfluous.

There are three Gaussian random fields in each model; u_m , u_A and u_B . Since each month is treated individually, this results in 180 different random fields with both mean and standard deviation, and naturally not all will be presented. Despite of the varying posterior ranges in models 1, 2 and 3, the means of the fields are almost identical, which removes the necessity of displaying the results from each model. Again the results for submodel 3b are used to represent all models. In addition, only the January and July fields are shown here and discussed in detail, to represent the cold and the warm months of the year. Just as it is the sum of the linear and quadratic temperature trend coefficients that is of interest, it is the sum of u_A and u_B that gives the best impression of the temperature trend in time. This means that in total four GRFs are displayed here, with mean and standard deviation.

It is natural to begin with the random fields $u_1^{[3b]}$ and $u_7^{[3b]}$. u_m has the role of capturing spatial variations occurring every year, independent of time, to make the temperature trend more prominent and easier to interpret. Since an intercept β_0 is included in all models, the field is constrained to integrate to zero. This means that u_m accounts for the spatial variation in temperature, and not the actual temperature. Keep in mind that the field is integrated to zero on the domain mesh, and not only on the observed area displayed. Figure 5.11 displays the mean and standard deviations of $u_m^{[3b]}$ for January and July for submodel 3b. Note that the intercept is not included in the spatial maps, and the reader is referred to Figure 5.9b for the posterior mean and quantiles of $\beta_0^{[3b]}$. Figures 5.11a and 5.11b show the mean and standard deviations of $u_1^{[3b]}$, respectively. The mean of the January field $u_1^{[3b]}$ spans a large interval, from nearly -11 to 5.5 degrees Celsius. The posterior means of the GRF u_1 (i.e., January) for models 1, 2 and submodels 3a and 3c are displayed in Figure A.9. As can be seen, the differences between the models are minimal (see also Figure 5.11a) for a given month. The posterior mean of u_m for the other months are almost identical in all models as well, but due to the large number of figures required this is not included in the thesis. For u_m , both the mean and standard deviation are similar between all models.

In the winter, the spatial pattern is prominent, with warmer areas around the coast and colder areas in the inland. The standard deviation of $u_1^{[3b]}$ is mostly lower than 1.7, and becomes larger far from observation locations. It is, not surprisingly, smallest where the temperature observations have been made. The posterior mean and standard deviation of the July field $u_7^{[3b]}$ can be seen in Figures 5.11c and 5.11d, respectively. The posterior mean of $u_7^{[3b]}$ spans a smaller tem-

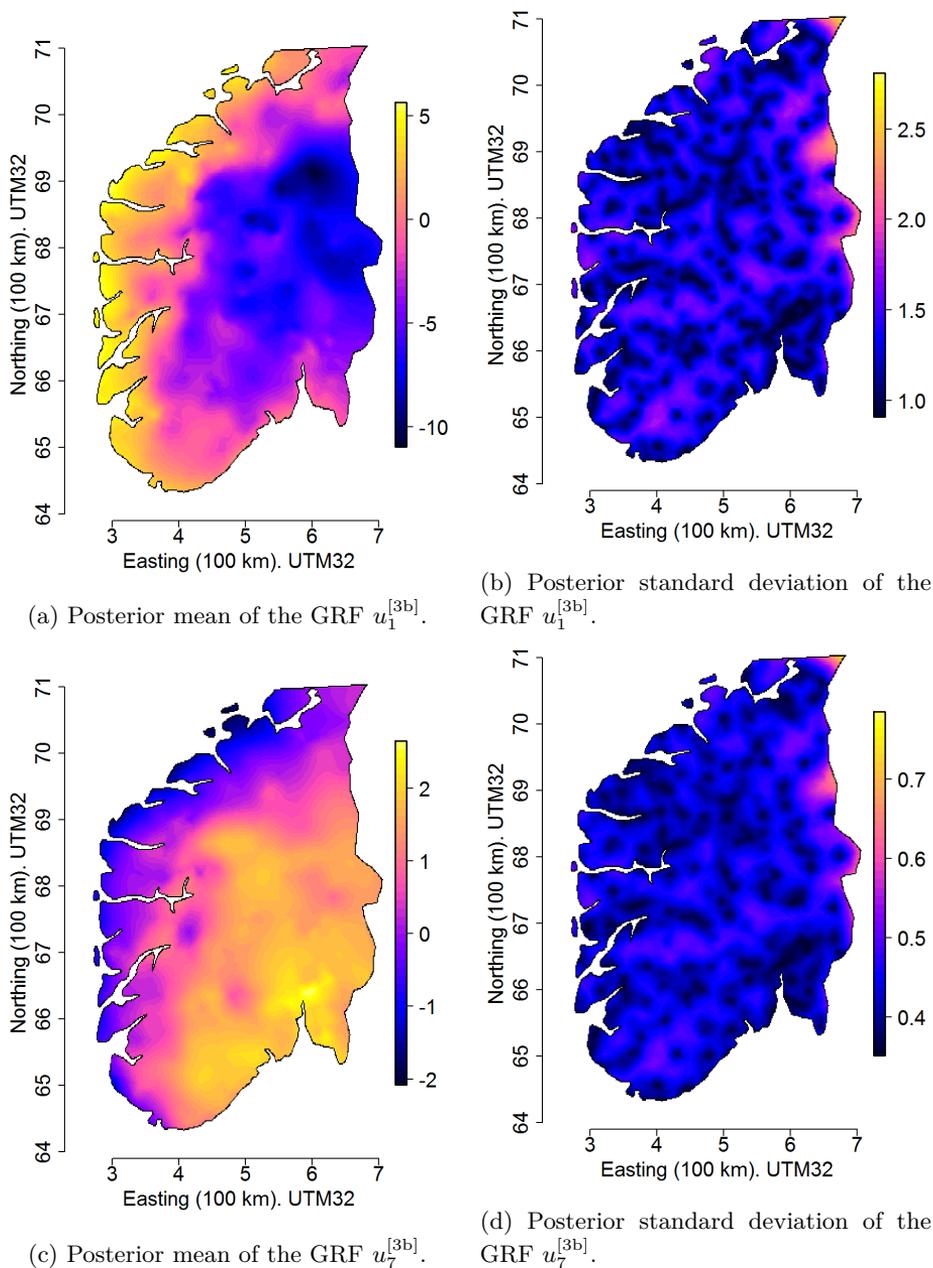


Figure 5.11: The mean and standard deviation of $u_m^{[3b]}$ in degrees Celsius for January and July, for submodel 3b. The intercept $\beta_0^{[3b]}$ is *not* included.

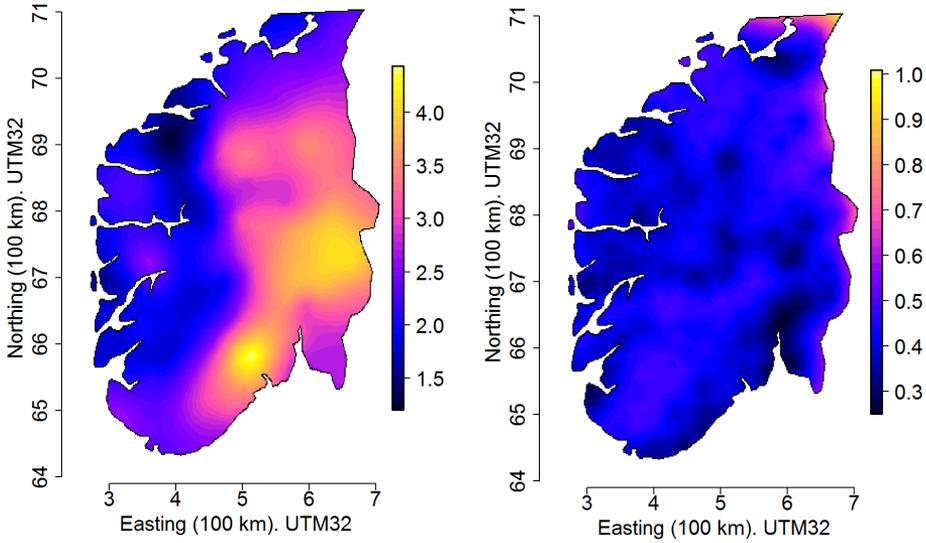
perature interval than $u_1^{[3b]}$, with values from -2 to 2.6 degrees Celsius. The cold areas are now the west coast, while the inland and the south coast are warmer. As for $u_1^{[3b]}$, the standard deviation is smallest in the observation locations, but now most areas have a standard deviation less than 0.5.

The coastal areas have the warmest January monthly average temperature and the coldest July temperature, see Figures 5.11a and 5.11c. This pattern is to some extent apparent in the temperature trends as well: The inland does in general experience a more rapid and a larger temperature increase than the coastal areas. This can have many reasons; the high heat capacity of the oceans may be one of them (Houghton, 2015). Water have higher heat capacity than air, and the ocean temperature changes slower than air. This might keep the land temperature more stable through the seasons in coastal areas than the inland, and lead to a slower temperature increase.

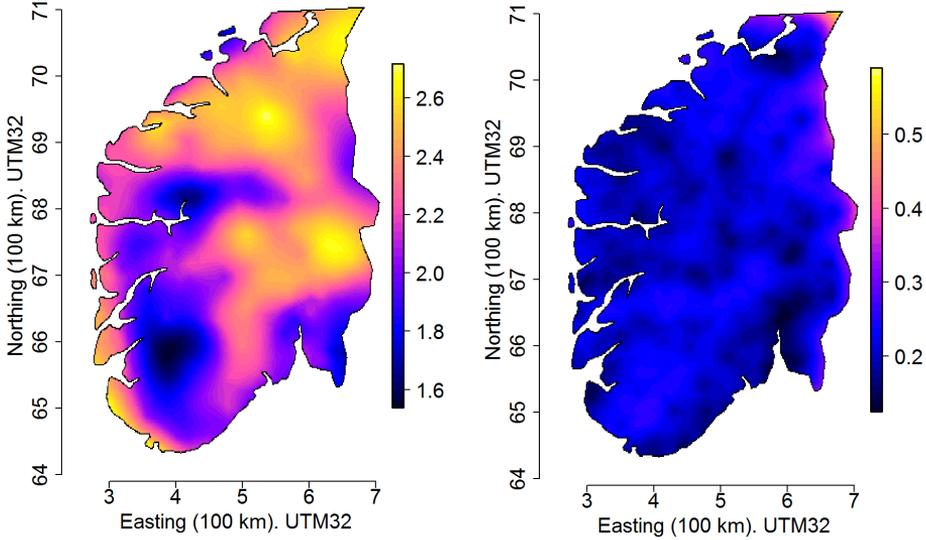
The perhaps most interesting results from the models are the posterior Gaussian random fields in Figure 5.12: The mean and standard deviation of the spatio-temporal temperature trends. The results displayed are for submodel 3b for January and July. Figure 5.12a is the mean of $a_1^{[3b]} + b_1^{[3b]} + u_{A,1}^{[3b]} + u_{B,1}^{[3b]}$, i.e., the temperature in January 2016 where the spatio-temporal trend started at zero everywhere in 1960. The western areas (including the coast) have had a smaller increase in temperature than the inland. The posterior means of $a_1 + b_1 + u_{A,1} + u_{B,1}$ for the other four models for January are displayed in Figure A.10. The differences between the plots are minimal, just as the differences between u_1 in Figure A.9 are minimal. The standard deviations of u_A and u_B (not included graphically in the thesis) are larger for all months for model 1 than for the other four models, just as for the fixed effects a and b (see Figure A.7 and A.8).

The January temperature increase during the 57-year time period is for some areas greater than 4 degrees Celsius, and even the areas with the smallest increase has experienced an increase of almost 1.5 degrees Celsius. No areas have experienced a decrease in monthly average temperature during the 57 years. The standard deviation is less than 0.6 degrees for most of the domain, with up to 1 degree in areas far away from observations. Thus, it is extremely unlikely that more than a few areas have not experienced a January temperature increase during those 57 years when using a second order polynomial temperature trend as in this model.

The posterior mean and standard deviation of $a_7^{[3b]} + b_7^{[3b]} + u_{A,7}^{[3b]} + u_{B,7}^{[3b]}$ for July for submodel 3b can be seen in Figures 5.12c and 5.12d, respectively. The temperature increase for July is not as large as for January. Some western areas and the southern coast has had the smallest July temperature increase. The majority of the domain has had an increase of 2.2 degrees Celsius or more, and



(a) Posterior mean of $a_1^{[3b]} + b_1^{[3b]} + u_{A,1}^{[3b]} + u_{B,1}^{[3b]}$.
 (b) Posterior standard deviation of $a_1^{[3b]} + b_1^{[3b]} + u_{A,1}^{[3b]} + u_{B,1}^{[3b]}$.



(c) Posterior mean of $a_7^{[3b]} + b_7^{[3b]} + u_{A,7}^{[3b]} + u_{B,7}^{[3b]}$.
 (d) Posterior standard deviation of $a_7^{[3b]} + b_7^{[3b]} + u_{A,7}^{[3b]} + u_{B,7}^{[3b]}$.

Figure 5.12: The posterior means and standard deviations of $a^{[3b]} + b^{[3b]} + u_A^{[3b]} + u_B^{[3b]}$ in degrees Celsius for January and July, for submodel 3b.

with a standard deviation of around 0.3 degrees for most of the domain, it is unlikely that any areas have had no increase or a decrease.

The maps in Figure 5.12 strongly agrees with both the decision to include spatial components in the model, and to treat all months of the year individually. The spatial difference in temperature trend is prominent. Just as the temperature trend varies between months, as seen in Figure 5.10, the spatial variations in temperature trend varies as well. Figure 5.13 displays the mean of the spatial temperature trend ($a + b + u_A + u_B$) for all months of the year for submodel 3b. The large variation in how much the temperature has increased between months is remarkable, but as it requires solid knowledge on weather and climate to investigate the reasons behind these differences further, this will not be studied in this thesis. The only months experiencing a decrease are May, June, and October, and only some areas in the domain have had a decrease. All other months of the year have experienced temperature increases. February has had the largest increase, with almost 5 degrees Celsius in the eastern part of Norway. The eastern part of Norway has experienced an increase in the monthly average temperature from 1960 to 2016 in all months of the year.

As the maps in Figure 5.12 represent the end in time of the temperature trends, it seems appropriate to include how the trend evolves from 1960 to 2016. This would require four dimensions, two spatial, the temperature and time, and animations are needed to fully display the results. Animations of the January and July temperature trend in time and space for submodel 3b can be found here: <http://folk.ntnu.no/ingebogh/>. Note that the year 1960 is defined to be the beginning of the trend, and is zero everywhere. Both the mean and the standard deviation of the temperature trend are animated, and which year is the current year is found in the animation. The temperature increase begins rapidly in the inland, while the west coast is a bit behind; in January, the temperature increase begins in the east, and "spreads" to the rest of the southern half of Norway. The same is happening for July, but the increase begins more to the south and is not as rapid as the increase for January. The end of the animation are the maps displayed in Figure 5.12. Note that the color scale indicating temperature differs between the animation and Figure 5.13. As the animations is not included in the thesis directly, and are just referred to for completeness, they will not be discussed further. It is however recommended to look at the animations to get the complete picture of the results.

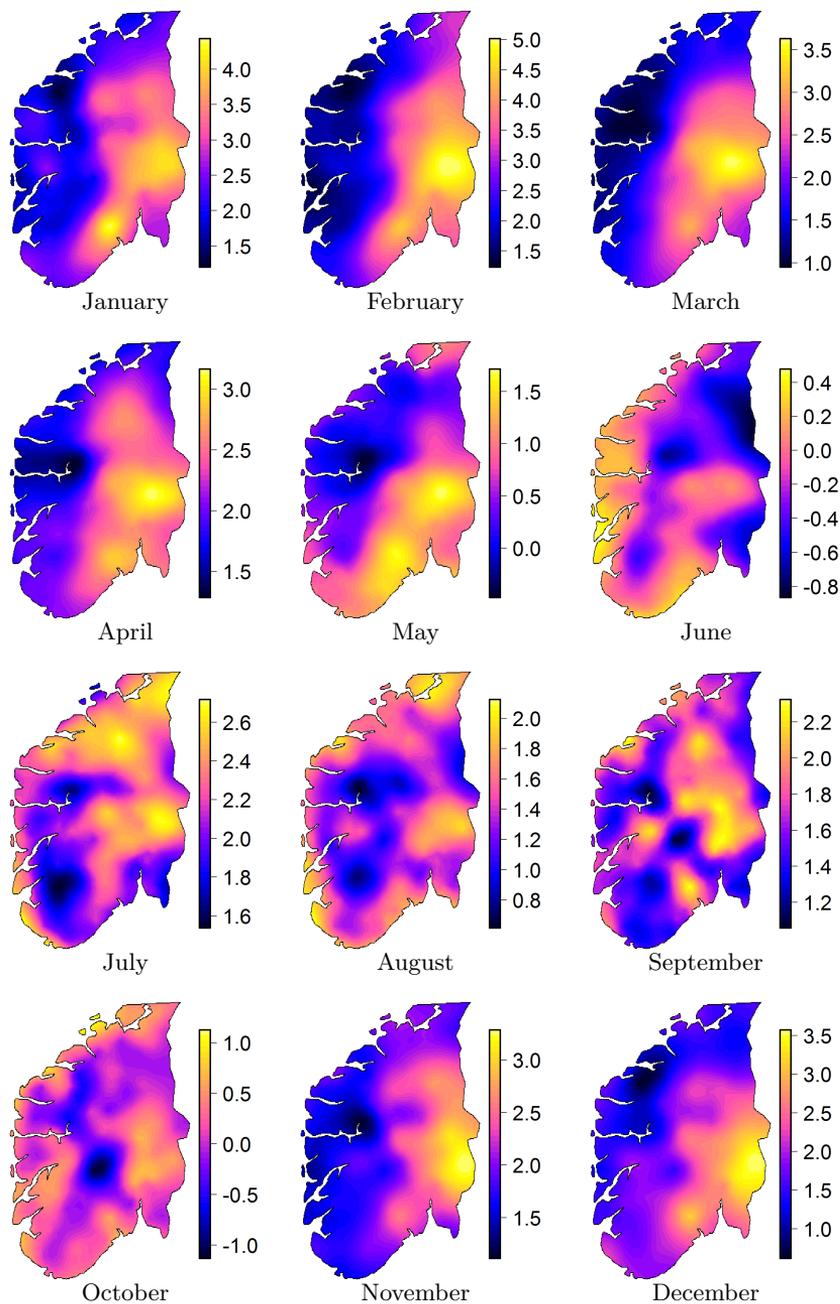


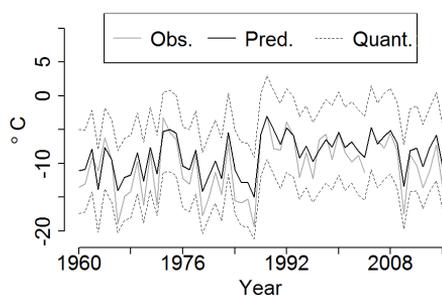
Figure 5.13: The posterior means of $a + b + u_A + u_B$ in degrees Celsius for all months for submodel 3b. Which month is which is found below each map. The maps correspond to the temperature change since 1960 in 2016 where the spatio-temporal trend started at zero everywhere in 1960. The axes are removed from the maps for a tidier figure, but have the same coordinate system as the maps in e.g. Figure 5.12.

5.4 Predicting missing observations

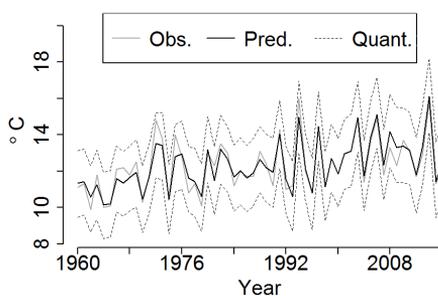
One of the main reasons for using a complex model with spatial components instead of a linear regression model, is that a spatial model can use all the available data even though the time series are incomplete, since such a model can predict the missing observations. To investigate how accurate the model can predict, the monthly average temperatures from two observation locations with close to complete time series in January and July are removed, one at a time, and the differences between the true and estimated time series are studied. Only the Drevsjø series is removed when predictions for Drevsjø are made, and same for Blindern. The observation locations chosen are Drevsjø in Hedmark county, and Blindern in Oslo county. The former is the observation location with the longest distance to the closest neighbor in the whole data set, and has only two neighbors in a 50 kilometers radius around itself. Blindern, on the other hand, has as many as 41 neighbors in a 50 kilometers radius. The monthly average temperature series from Drevsjø are not entirely complete, and is missing the year 2005 in the January series, and the years 1997, 1998 and 2000 to 2002 in the July series.

Figure 5.14 displays the original temperature series at Drevsjø and Blindern for January and July, and the predicted time series with corresponding 0.025 and 0.975 quantiles. Figures 5.14a and 5.14b show the Drevsjø observations and predictions for January and July, respectively, and Figure 5.14c and 5.14d show the same for Blindern for January and July, respectively. There is no surprise that the predictions for Blindern are better than for Drevsjø, as Blindern has many observation locations in close proximities, while Drevsjø does not. There is no clear difference in the quality of the predictions between the January temperature and the July temperature, but as the former spans a temperature interval three times as large as the latter, the standard deviations of the January predictions naturally are larger than the standard deviations of the July predictions. The observed temperature lies within the 0.025 and 0.975 quantiles of all four predicted time series.

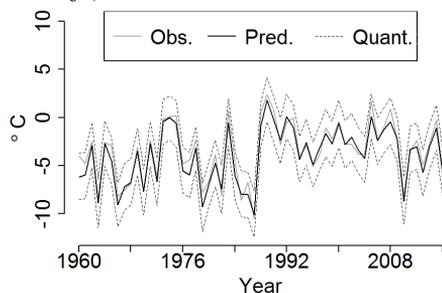
The predictions of the Drevsjø observations (Figures 5.14a and 5.14b) are despite of the large standard deviations fairly accurate. The most extreme values have been underestimated or overestimated, depending on whether the year was unusually warm or cold, respectively, but the predicted temperature curves follows the observed. The predictions of the Blindern observations (Figures 5.14c and 5.14d) are in both January and July less uncertain than the Drevsjø, with accurate prediction means.



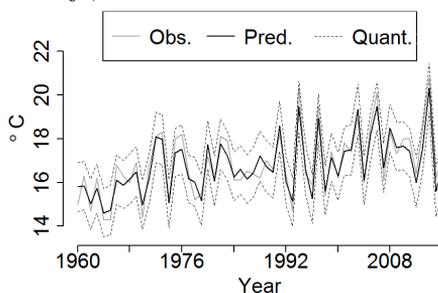
(a) Observations and predictions with 0.025 and 0.975 quantiles for January at Drevsjø, Hedmark.



(b) Observations and predictions with 0.025 and 0.975 quantiles for July at Drevsjø, Hedmark.



(c) Observations and predictions with 0.025 and 0.975 quantiles for January at Blindern, Oslo.



(d) Observations and predictions with 0.025 and 0.975 quantiles for July at Blindern, Oslo.

Figure 5.14: Observed and predicted monthly average temperatures for submodel 3b, for two observation locations for January and July. The predictions for each time series are made with the observed temperature series removed from the model. When predicting the Drevsjø temperature, the Blindern time series is included, and vice versa. The 0.025 and 0.975 quantiles of the predicted observations are included in the graphs.

The graphs in Figure 5.14 confirm that the model is highly capable of predicting missing observations using temperatures from nearby observation locations. The more information in nearby locations, the more accurate the predictions are, but even for observation locations that lie far from others the predictions are sufficient. This illustrates one of the huge advantages with a spatial model; now all available data can be used, and the temperature where no observations are made can be predicted by the model. In the linear regression model in Equation

(2.1) in Section 2.2, only a fraction of the available temperature data could be used. This problem is solved with the spatial model.

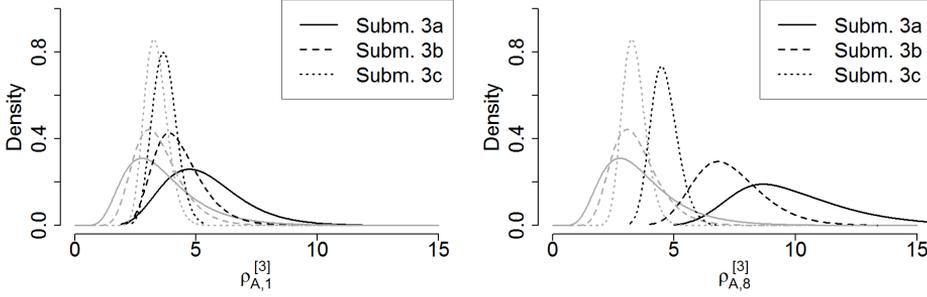
5.5 Prior sensitivity

As the prior distributions in submodels 3a, 3b and 3c are from the same distribution family, but with different hyperparameters on ρ_A and ρ_B , it is possible to analyze the prior sensitivity to some extent. The three submodels 3a, 3b and 3c represent weak, medium, and strong priors on ρ_A and ρ_B , respectively, and even though they have already been presented, they are now compared in the same graphs to easier see the posterior differences. As usual, January is used to represent the cold months, but now August represents the warm months instead of July, as August has by far the largest mean range in almost all the GRFs in the three submodels. Figure 5.15 consists of four graphs: the prior and posterior of $\rho_A^{[3]}$ and $\rho_B^{[3]}$ for January and August for submodels 3a, 3b and 3c.

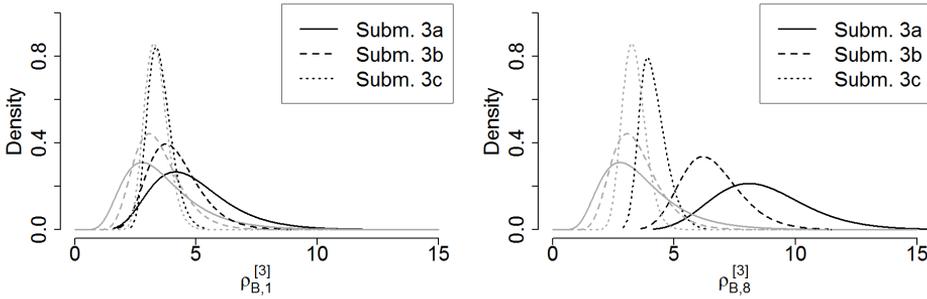
Common in all four plots in Figure 5.15 is that the posterior range has a larger mode than the prior for a given model, and that the posterior has a larger standard deviation than the corresponding prior. The posteriors of $\rho_A^{[3]}$ for January and August can be seen in Figures 5.15a and 5.15b, respectively. They have slightly larger modes and standard deviations than $\rho_B^{[3]}$ for a given month and model. The posteriors of the latter can be seen in Figure 5.15c for January and Figure 5.15d for August. Clearly, the posterior highly depends on the prior for both parameters. The posteriors have smaller mean values when the priors are stronger and no longer allow large posterior means. The posterior January and August ranges both seem to have large means and modes if they can, and when the possibilities of having large mean values are constricted by the prior, they almost reluctantly follow the prior to a smaller mean range value. Here it becomes clear that the strong prior is necessary to achieve a posterior mean range less than 7. August, and to some extent the other warmest months of the year as well, get estimated ranges that are larger than desired when the prior is too weak (see Section 5.2).

January has posterior ranges according to the prior, but as Figures 5.15a and 5.15c show, the posterior range has relatively equal modes in all submodels, independent of the prior. This is also the case for most months of the year, with some warm months as exceptions. Thus, the months with a satisfying posterior for a weak prior are not affected much by a stronger prior other than getting smaller standard deviation of the posterior, while the unsatisfactory posterior ranges for the remaining months are affected and results in an acceptable model.

The standard deviations of the GRFs $u_A^{[3]}$ and $u_B^{[3]}$ are the only other parameters changing significantly when the priors on $\rho_A^{[3]}$ and $\rho_B^{[3]}$ are changed, which can be seen in Figure A.4. In general, the posteriors become less uncertain and have smaller modes with stronger priors on the ranges ρ . In addition, the posteriors of σ_A and σ_B are similar for a given model and month.



(a) Prior and posterior distributions of $\rho_{A,1}^{[3a]}$, $\rho_{A,1}^{[3b]}$ and $\rho_{A,1}^{[3c]}$. The grey lines are the prior distributions, and the black are the posterior distributions. (b) Prior and posterior distributions of $\rho_{A,8}^{[3a]}$, $\rho_{A,8}^{[3b]}$ and $\rho_{A,8}^{[3c]}$. The grey lines are the prior distributions, and the black are the posterior distributions.



(c) Prior and posterior distributions of $\rho_{B,1}^{[3a]}$, $\rho_{B,1}^{[3b]}$ and $\rho_{B,1}^{[3c]}$. The grey lines are the prior distributions, and the black are the posterior distributions. (d) Prior and posterior distributions of $\rho_{B,8}^{[3a]}$, $\rho_{B,8}^{[3b]}$ and $\rho_{B,8}^{[3c]}$. The grey lines are the prior distributions, and the black are the posterior distributions.

Figure 5.15: Prior and posterior marginal distributions of $\rho_A^{[3]}$ and $\rho_B^{[3]}$ for January and August for submodels 3a, 3b and 3c.

Chapter 6

Discussion

The proposed spatio-temporal model is interpretable and has interpretable results, the model successfully estimates the desired temperature trends, and clearly show an increase in temperature that varies in space. As shown in Section 5.4, the model is highly capable of predicting missing observations accurately, and with prediction uncertainty. With the Bayesian approach predictions are calculated with uncertainty, which includes the model parameter uncertainty. All available data can be used, as the data set does not have to be complete when the inference is carried out. This was a problem with the linear regression model from Section 2.2, which is solved by using a spatio-temporal model.

The model results are not only interpretable, they are also consistent with measurements and estimates made by others. The increase is for some months greater than yearly average measurements show, and larger than global trends (see e.g. Houghton (2015)), but the increase is consistent with projections made by the Intergovernmental Panel on Climate Change (IPCC, 2013), and with measurements, which both show that northern latitudes experience more temperature increase than the global average. In addition, it is discovered that from 1960 to 2016 it is the coldest months of the year (December to March) that have experienced the largest temperature increase. May, June, and October have experienced a decrease in some areas, while the eastern part of Norway has experienced an increase in all months of the year. There are large variations in both the spatial pattern and the magnitude of the temperature changes between the months of the year.

The months are treated individually in the model in this thesis. The differences in posterior distributions for most parameters supports this decision, and so does

the different mean and standard deviations of the three Gaussian random fields. The posterior parameter distributions differ between months, but for a given month, the posteriors usually are similar to the month before and the following month. This is also true for the posterior mean of the spatial temperature trend (see Figure 5.12).

The model experiences numerical problems for certain prior distributions. This complicates the process of choosing prior distributions to some extent, as strong priors are required to avoid numerical problems. The numerical problems indicate an overparametrized model in the sense of having too many parameters describing the same quantity, and are mainly occurring due to the GRFs and their parameters and hence the SPDE models. The advantage of the SPDE approach is the discretization process with irregular grids, and other approaches will not necessarily lead to less numerical problems, as large posterior ranges will still be estimated from the data. The SPDE models might amplify the problems when the range is large due to among others the boundary conditions, but a long range is not desired in this thesis. Strong prior distributions on the range lead to interpretable posterior ranges, and in those cases the SPDE approach does not cause trouble. One of the priors that are too weak on the GRF parameters is the PC prior, and even though there exist hyperparameters leading to a working model with the PC prior, the base model of the PC prior contradicts the desire of having all three GRFs in the model. The PC prior is however suitable for the temporal iid-effect, as the base model without this model component is reasonable.

The model in this thesis is computationally fast to do inference on using INLA, and as all months are treated individually the inference can be carried out for all months at the same time, yielding a reasonable computation time, especially when the amount of data is taken into account. The inference takes about 20 minutes for each month. Other model approaches than the one used in this thesis naturally exists, and one of them is separable models with temporal processes and spatial innovations. In such models the basis functions used in this thesis can be replaced with e.g. a random walk of second order, which requires less parameters, and gives a smooth and flexible trend curve that can capture yearly events, which the polynomial of second degree cannot. This is however a more time-consuming model to do inference on, it is more difficult to interpret, and a trend in the data will be more difficult to detect as the trend curve has much more freedom. The model in this thesis is on the other hand both computationally fast and interpretable with interpretable results and a clear temperature trend.

That this spatio-temporal model is considered to be sufficient for its purposes does not mean there is no room for improvement. The numerical problems indicate that the model might be overparameterized with too many parameters describing the same quantity, which can be solved in several ways. As the parameters of the

GRFs, and especially of u_A and u_B , describe similar quantities, it is natural to aim to reduce the number of field parameters. The numerical problems occurring due to these parameters would be avoided if the parameter values were fixed, as this approach does not include prior and posterior distributions. The parameter values are however unknown and difficult to give fixed values to, such an approach is not Bayesian, and is not a good solution. A better solution is to model the range and standard deviation for the GRFs estimating the temperature trend together, i.e., have one common range and one common standard deviation parameter for u_A and u_B . This may lead to even faster inference, and also opens for possibilities of including more basis functions than one linear and one quadratic as the number of parameters will not increase. Each random field must still have a unique posterior mean and standard deviation.

Another way to make the modelling less complex is to model several months together. As the results for subsequent months are similar, the average of two or three months could be examined together, instead of one by one. This will not give any advantages besides having less results, which in some cases could be a disadvantage, but less results might be easier to interpret. Further work yielding more precise results includes using a skewed likelihood. As the temperature all around the globe is changing, the distribution of the temperature changes as well (Houghton, 2015). Then a distribution where the mass of the tails can be changed individually for each tail might be more appropriate. This will however lead to more parameters and a more time-consuming model. If the number of parameters from other model components can be reduced, this is a step towards more accurate results.

Chapter 7

Conclusion

The model in this thesis is successfully able estimate spatio-temporal temperature trends for incomplete data sets, as it accurately predicts missing observations. An interpretable model with interpretable results is achieved, partially due to strong prior distributions. Especially the range parameters of the Gaussian random fields require strong priors to give interpretable posterior distributions. The estimated spatio-temporal temperature trends show large variations between months; May, June and October have areas with a decrease in temperature from 1960 to 2016, whereas the other nine months of the year have experienced an increase only. The biggest change is the February monthly average temperature which in the eastern part of Norway have increased almost 5 degrees Celsius during the 57-year time period, according to this model.

Despite of numerical problems occurring during the inference for certain prior distributions, the model is satisfactory. The model is a computationally fast alternative to using e.g. a separable model with temporal processes with spatial innovations. It is simple to implement using **R-INLA**, and thus simple to replicate. The model is easy to extend in time with more temperature data, and to other types of data. The trend can be modelled using other basis functions than linear and quadratic, but caution must be taken to avoid an overparameterized model. To let all fields that are connected to the basis functions have common parameters may avoid overparameterization, and might also reduce the occurrence of numerical problems.

The model can be used to do an initial analysis when wondering if a data set has a spatio-temporal trend, and can be extended to model other trends than polynomial of second degree as long as the number of parameters in the model is

kept low. If some of the suggestions given in the discussion can avoid overparameterization and numerical problems, the idea of using basis functions to model data trends in space and time have potential to become useful.

In conclusion, even though the model in this thesis is somewhat difficult to use due to numerical problems, the model is able to predict missing observations, estimate temperature trends, and it gives interpretable results. It is thus sufficient for its purposes. Due to the reasonable computation time of the model, it is a good starting point when investigating the existence of a spatio-temporal trend in a data set.

Further work includes, but is not limited to, investigating ways to reduce the number of parameters in the model, studying if other types of data can be analyzed with this approach, and use other basis function to model spatio-temporal data trends.

Chapter 8

Bibliography

- Abrahamsen, P., (1997). *A review of Gaussian random fields and correlation functions*. Norsk Regnesentral.
- Andrews, D. G., (2010). *An Introduction to Atmospheric Physics*. Cambridge University Press, second edition. ISBN 9780521693189.
- Blangiardo, M. and Cameletti, M., (2015). *Spatial and Spatio-temporal Bayesian Models with R - INLA*. Wiley.
- Casella, G. and Berger, R. L., (2002). *Statistical Inference*. Duxbury, second edition.
- Cressie, N. and Wikle, C. K., (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Ltd., Hoboken, New Jersey, USA.
- Fuglstad, G. A., Simpson, D., Lindgren, F., and Rue, H., (2017). Constructing Priors that Penalize the Complexity of Gaussian Random Fields. Available at: <https://arxiv.org/abs/1503.00256>. Preprint.
- Givens, G. and Hoeting, J., (2013). *Computational Statistics*. Wiley, 2nd edition.
- Hartmann, D., Tank, A. K., Rusticucci, M., Alexander, L., Brönnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., A. Kaplan, B. S., Thorne, P., Wild, M., and Zhai, P., (2013). Observations: Atmosphere and Surface. In Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., (eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter 2,

- pages 159–254. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Available at: www.climatechange2013.org.
- Hem, I. G. Latent Gaussian models for analysing monthly temperature in Norway. Unpublished specialization project, (2016). Available at: http://folk.ntnu.no/ingeboogh/analysing_temperature_norway_hem.pdf.
- Houghton, J., (2015). *Global Warming: The Complete Briefing*. Cambridge University Press, fifth edition.
- Ingebrigtsen, R., (2014). *Bayesian spatial modelling of non-stationary processes and misaligned data utilising Markov sproperties for computational efficiency*. PhD thesis, Norwegian University of Science and Technology.
- IPCC, (2013). Annex I: Atlas of Global and Regional Climate Projections. In Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., (eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter AI, pages 1311–1394. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Available at: www.climatechange2013.org.
- Kullback, S. and Leibler, R. A., (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.
- 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. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.
- Lindgren, F. K. and Rue, H., (2015). Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, 63:1–25.
- Rue, H. and Held, L., (2005). *Gaussian Markov Random Fields. Theory and Applications*. 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.

- Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H., (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32(1):1–28.
- Stein, M. L., (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer New York.

Appendix A

Additional results

As the five model variations described and presented in this thesis yields a large number of graphs and maps, most are not included in the result-chapter (Chapter 5). Usually, the results for submodel 3b are presented there, and here some of the same results for models 1, 2 and submodels 3a and 3c are included for completeness. This makes comparison of results between models possible.

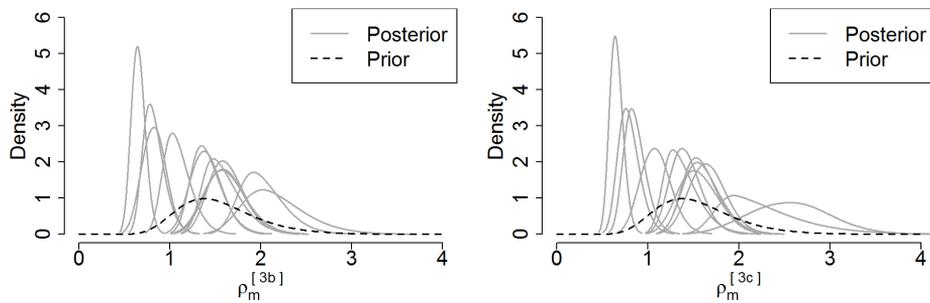


Figure A.1: Prior and posterior marginal distributions of the range ρ_m for submodels 3b and 3c. Posteriors of all twelve months of the year are plotted, but which is which is not indicated.

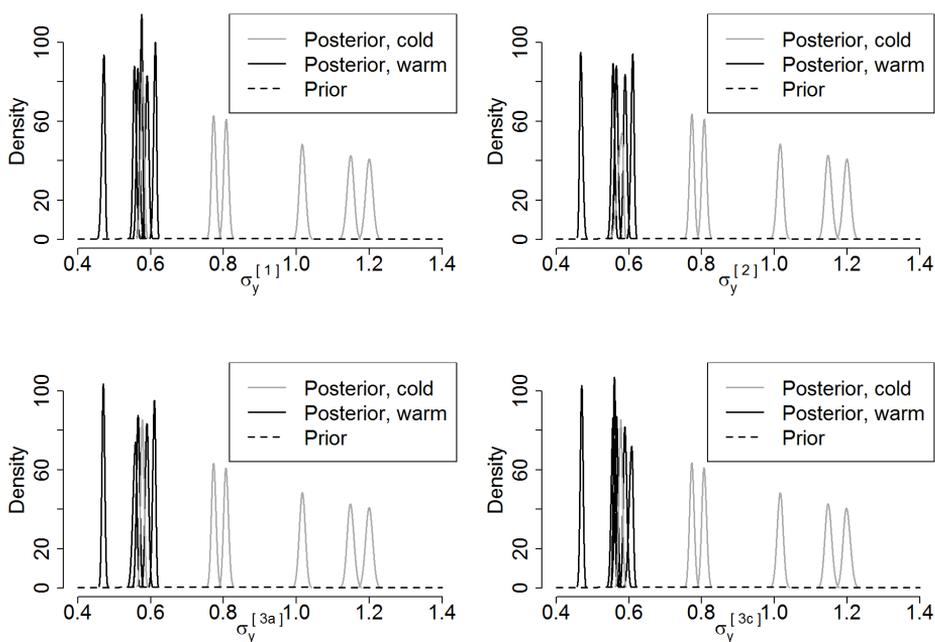


Figure A.2: Prior and posterior marginal distributions of the standard deviation σ_y of the observations for all twelve months of the year for models 1, 2 and submodels 3a and 3c. The months are divided into groups of warm and cold months. The warm months are taken to be April to September, while the cold are October to March.

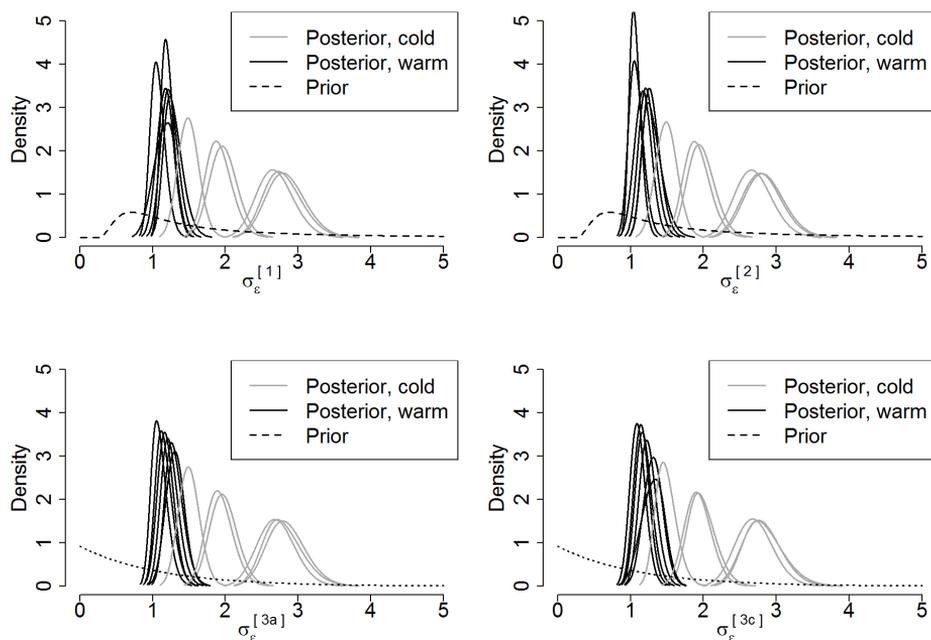


Figure A.3: Prior and posterior marginal distributions of the standard deviation σ_ϵ of the temporal iid-effect for all twelve months of the year for model 1, 2 and submodels 3a and 3c. The months are divided into groups of warm and cold months. The warm months are taken to be April to September, while the cold are October to March.

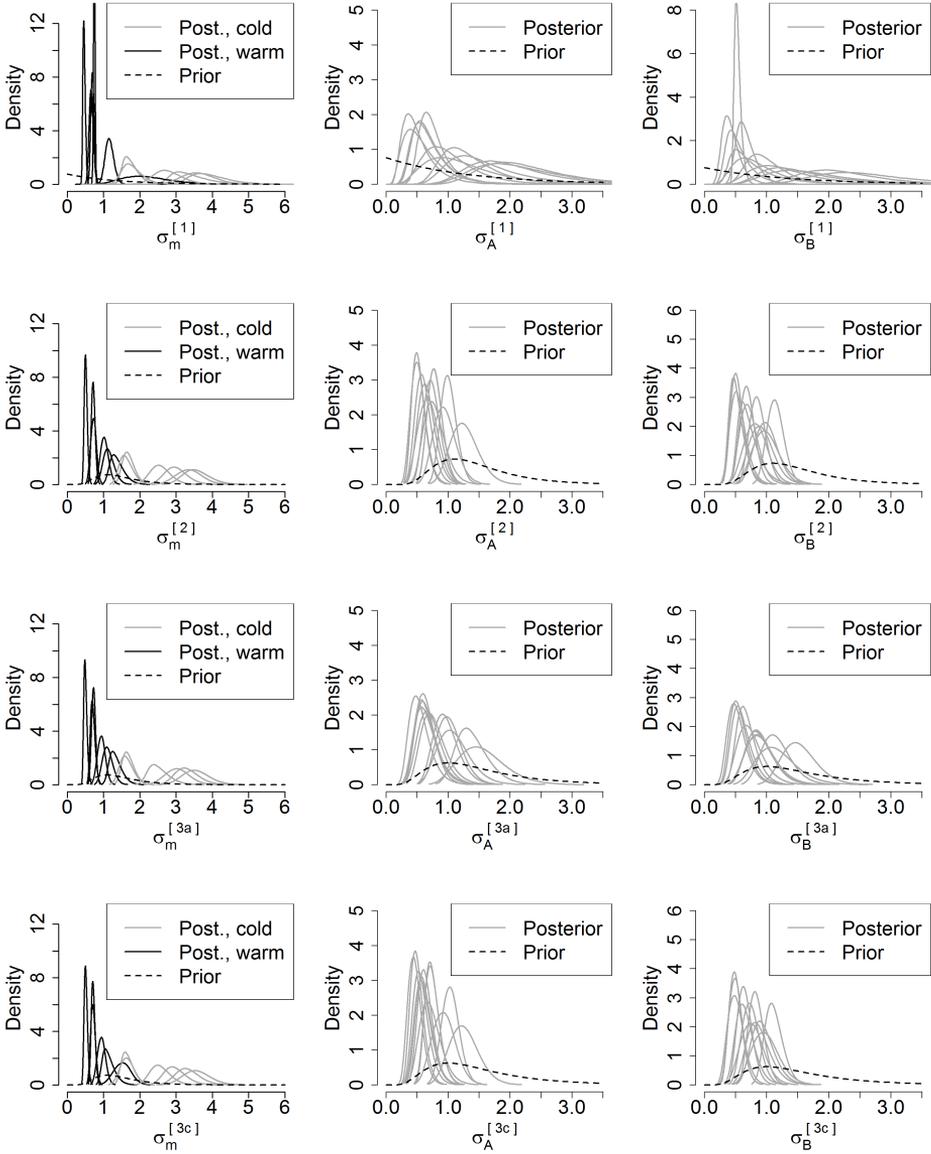


Figure A.4: Prior and posterior marginal distributions of the standard deviations σ_m , σ_A and σ_B for the Gaussian random fields u_m , u_A and u_B , respectively, for all twelve months of the year for models 1, 2 and submodels 3a and 3c. The posteriors of σ_m are divided into two groups: the warm group contains the month April to September, while the cold group contains October to March. The posteriors of σ_A and σ_B have no indications of which month is which.

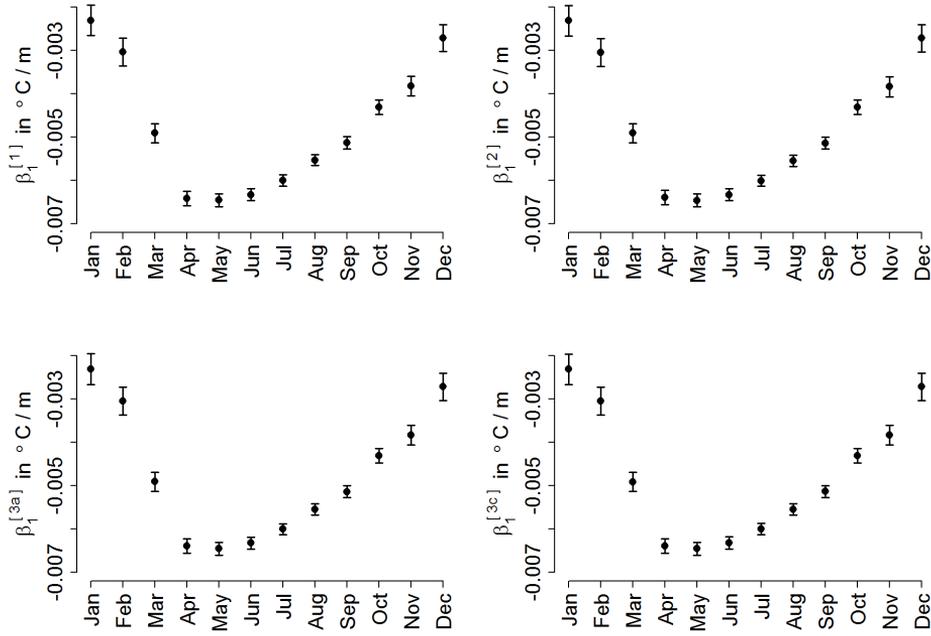


Figure A.5: The posterior altitude coefficient β_1 for all twelve months of the year for models 1, 2 and submodels 3a and 3c. The mean is plotted along with the 0.025 and 0.975 quantiles.

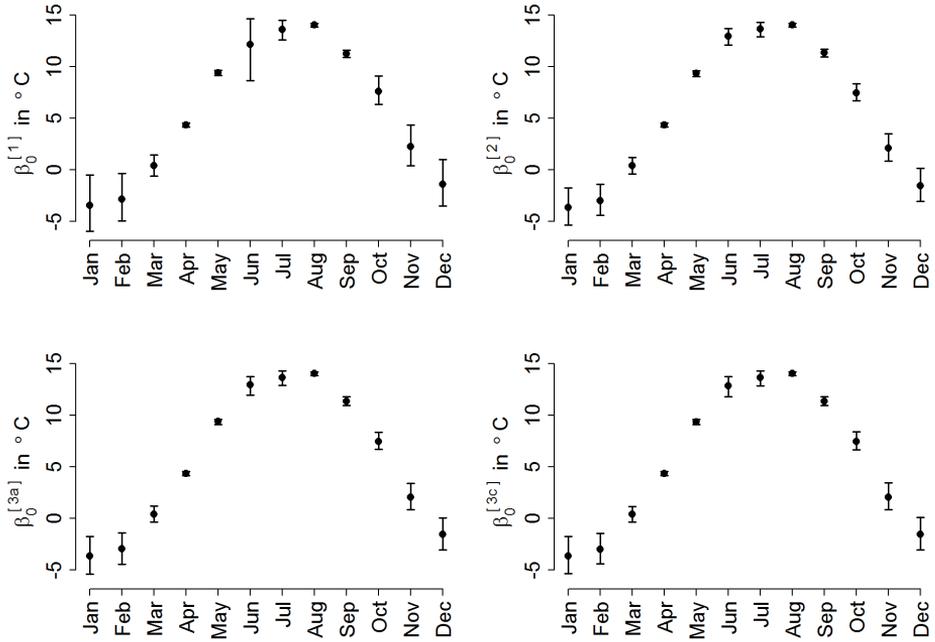


Figure A.6: The posterior intercept β_0 for all twelve months of the year for models 1, 2 and submodels 3a and 3c. The mean is plotted along with the 0.025 and 0.975 quantiles.

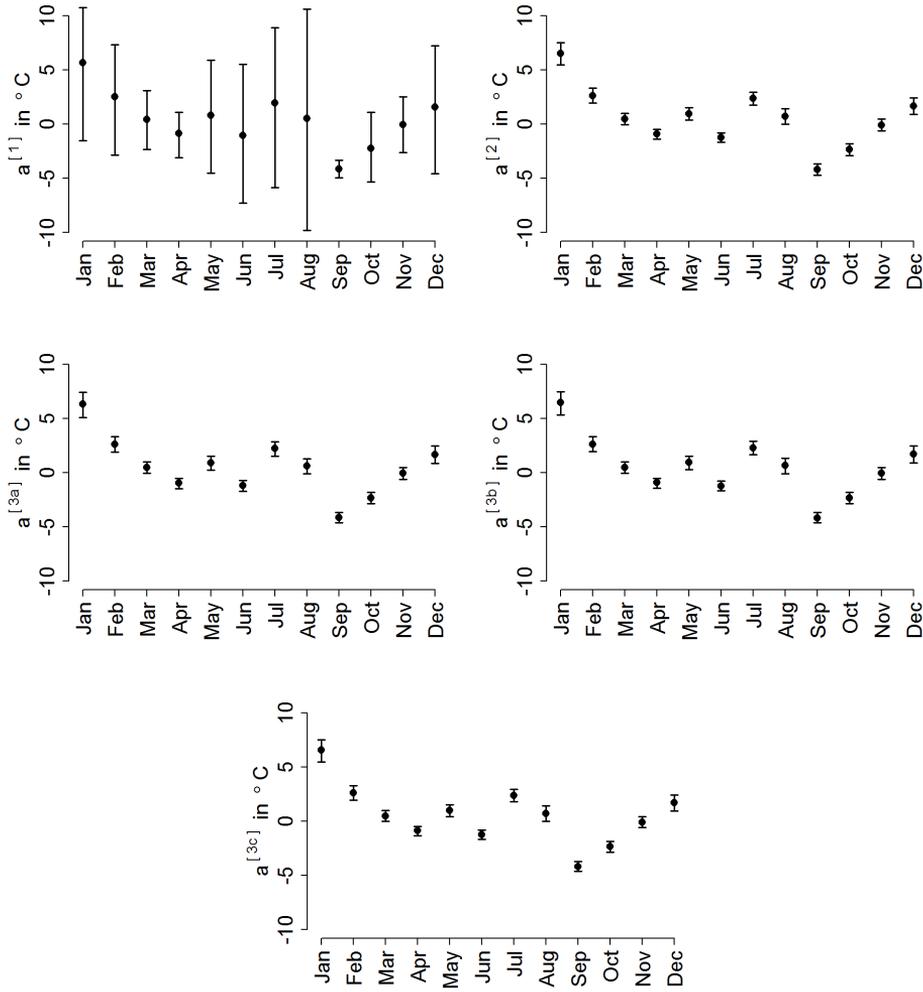


Figure A.7: The posterior linear temperature trend coefficient a for all twelve months of the year for all models (model 1 and 2, and submodels 3a, 3b and 3c). The mean is plotted along with the 0.025 and 0.975 quantiles.

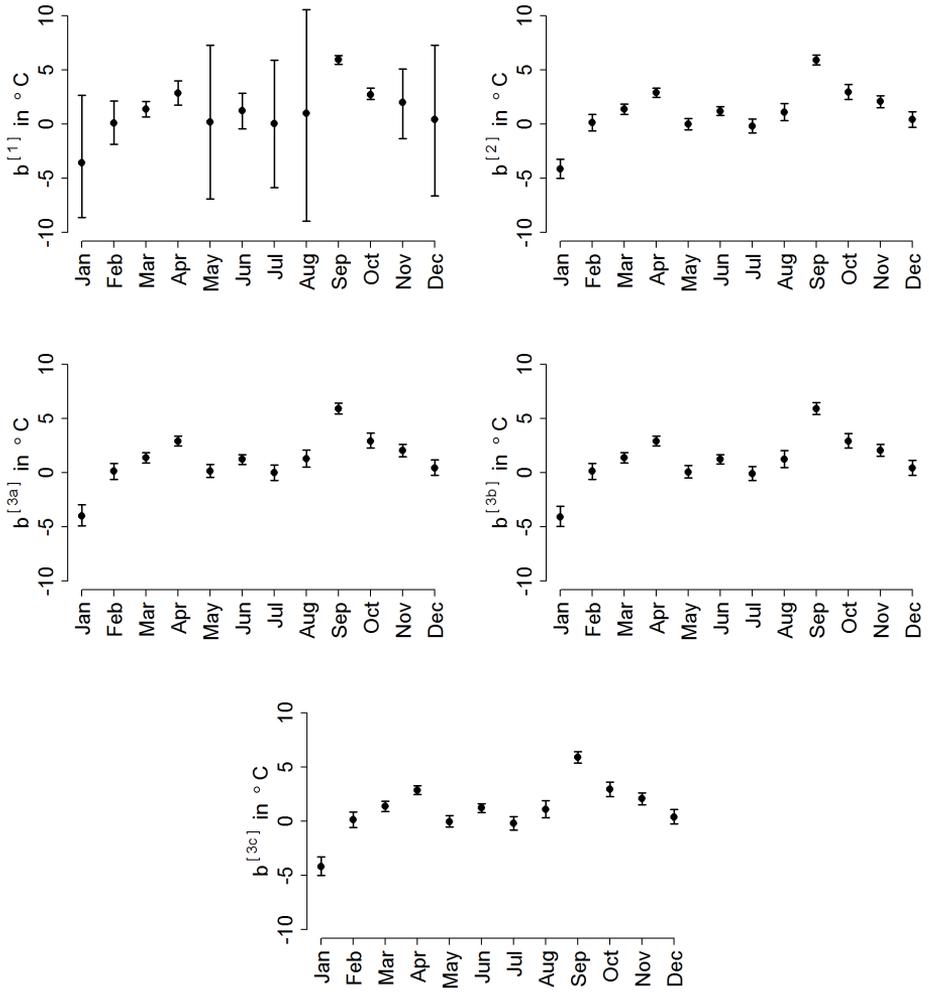


Figure A.8: The posterior quadratic temperature trend coefficient b for all twelve months of the year for all models (model 1 and 2, and submodels 3a, 3b and 3c). The mean is plotted along with the 0.025 and 0.975 quantiles.

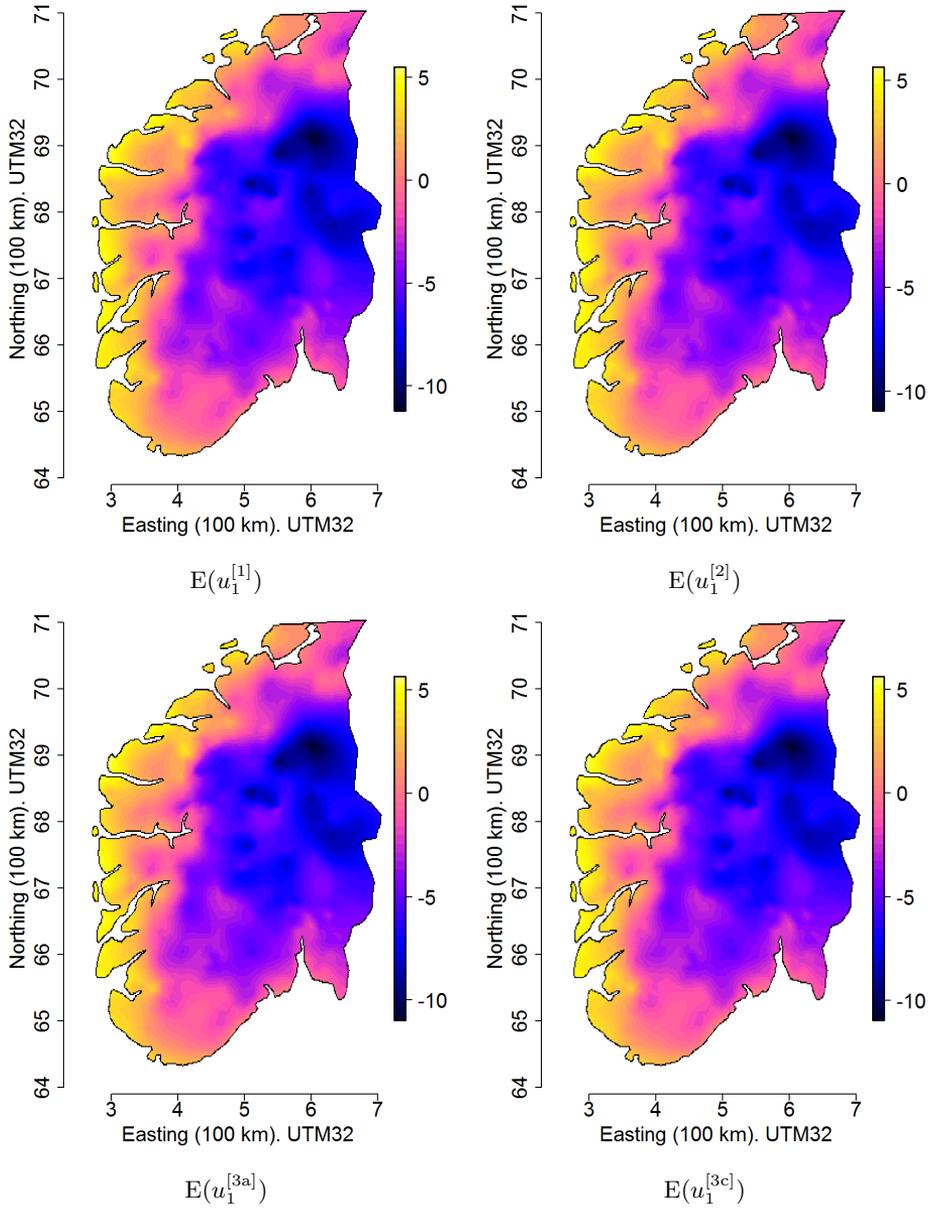


Figure A.9: The posterior means of the Gaussian random field u_1 in degrees Celsius for January for models 1, 2 and submodels 3a and 3c. The intercepts β_0 are *not* included, and can be seen in Figure A.6. The maps correspond to the spatial variation in average temperature.

