NTNU

Norwegian University of
Science and Technology

Spatial Non-Stationary Models for Precipitation with Elevation in the Dependency Structure

A Case Study of Annual Precipitation in Hordaland

Jorid Ødegård

Master of Science in Physics and Mathematics
Submission date: June 2017
Supervisor: Ingelin Steinsland, IMF

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

In this work, we construct spatial statistical models for interpolation of precipitation in areas characterised by orographic precipitation. The models are built as latent Gaussian models. We use the stochastic partial differential equation (SPDE) approach to spatial modelling to reduce the computational cost. The methodology integrated nested Laplace approximation (INLA) is used for inference and interpolations. The aim of the study is manifold. Firstly, the study has applied purposes. We aim to construct a good model for interpolation of precipitation in areas characterised by orographic precipitation, and for prediction of the total precipitation in catchment areas, i.e., the areal precipitation. Such a model should be able to quantify the uncertainty of the interpolations, and have a good predictive performance. It is also desired to get a better understanding of the physical precipitation process in this kind of terrain, e.g., how precipitation varies with elevation. Secondly, the study has a statistical purpose. We aim to obtain a better knowledge about non-stationary and stationary modelling of spatial processes. In particular, we wish to examine how large degree of non-stationarity there has to be in a process, before it is detectable and relevant for the predictive performance. We compare a stationary model and a non-stationary model with dependency structure varying with elevation. Simple toy examples are used to explore the consequences of having dependency structure that varies with elevation. A case study is carried out, using annual observations of precipitation in Hordaland. It is also performed a simulation study, in order to further explore the effect of non-stationarity, and its impact on predictions of precipitation. The results showed that a non-stationary model has a slightly better predictive performance when doing interpolations of precipitation in areas characterised by orographic precipitation. However, there are still some large errors when using the non-stationary model to predict areal precipitation in catchments located at high elevations. The results also showed that the predictive performance of both models became noticeably better when an observation was included inside the catchment areas. Further, the results showed that when the degree of non-stationarity in the process is small, a stationary model has good predictive performance.

Sammendrag

I denne oppgaven lager vi romlige statistiske modeller for årlig nedbør i områder hvor klimaet er preget av orografisk nedbør. Modellene er av typen latente Gaussiske modeller. Vi bruker en stokastisk partiell differensialligning (SPDE) som tilnærming til romlig modellering, for å redusere de beregningsmessige kostnadene. Metoden integrert nøstet Laplace approximering (INLA) benyttes til å gjøre statistisk inferens og interpoleringer. Målet med oppgaven er flerfoldig. For det første, har oppgaven anvendte formål. Vi ønsker å lage gode modeller som kan benyttes til interpolering av nedbør i områder preget av orografisk nedbør, samt til predikering av den totale nedbøren i avrenningsområder, arealnedbøren. Det fokuseres på at slike modeller skal klare å estimere usikkerhetene i interpoleringene. Det er også ønskelig å oppnå en bedre forståelse av den fysiske prosessen nedbør i denne typen terreng, for eksempel hvordan nedbør varierer med høyde. For det andre har oppgaven statitiske formål. Vi ønsker å øke kunnskapen innen stasjonær og ikke-stasjonær modellering av romlige prosesser. Spesielt ønsker vi å utforske hvor stor grad av ikke-stasjonæritet det må være i en prosess, for at det skal oppdages og være relevant for prediksjonsevnen til modellene. Vi sammenligner en stasjonær modell og en ikke-stasjonær modell med avhengighetsstruktur som varierer med høyden. Vi bruker enkle eksempler til å utforske konsekvensene av å ha avhengighetsstruktur som varierer med høyden. Vi analysere ekte observasjoner av nedbør i Hordaland, og i tillegg utfører vi et simuleringsstudium. Resultatene viste at en ikke-stasjonær modell har en noe bedre prediksjonsevne og usikkerhetsestimering, når det gjøres interpolering av nedbør i områder preget av orografisk nedbør. Det er likevel fortsatt noen store feil når man bruker den ikke-stasjonære modellen til å predikere arealnedbør. Resultatene viste også at å inkludere en observasjon inni avrenningsområdet hvor arealnedbøren predikeres, fører til bedre prediksjonsevne for begge modellene. Videre viste resultatene av når graden av ikke-stasjonæritet i prosessen er liten, har en stasjonær modell god prediksjonsevne.

Preface

This thesis concludes my master's degree in Industrial Mathematics at the Norwegian University of Science and Technology (NTNU). The work was carried out during the spring of 2017, and builds upon a preliminary project in the course TMA4500, performed during the autumn of 2016.

I would like to thank my supervisor Ingelin Steinsland, for guiding me through the work with the master and the preliminary project with great knowledge and enthusiasm, and for always taking the time to answer all my questions. I would also like to thank PhD student Thea Roksvåg, for patiently helping me out with `R-INLA`, especially during the preliminary project.

Finally, I would like to thank my family, for their endless help and support through the years of school and study. And last, but not least, for being a great inspiration, and for always supporting me, motivating me and cheering me up, I would like to thank my boyfriend Espen.

Jorid Ødegård
Trondheim, 2017

Contents

Chapter 1

Introduction

Statistics is the scientific field of learning from available data (Ingebritsen, 2014). In some sense, one can say that to do statistics is to collect numbers and deduce convenient summaries of these (Ripley, 1981). By applying accessible information about some phenomenon or process, one can build stochastic mathematical models. These models can be used to gain insight about the process through statistical inference and to make predictions. Spatial statistics occurs when the process is spatially referenced. This means that the process is connected to spatial locations, which contributes directly to the stochastic model (Gelfand et al., 2010). Such a stochastic process is called a random field.

One important characteristic of a random field, is its dependency structure, often modelled through a covariance function. A covariance function is a measure of how two stochastic variables vary together. Loosely speaking, a larger covariance between two stochastic variables means that the variables tend to show more similar behaviour. If two variables are independent, the covariance of the variables is zero.

A random field can be stationary or non-stationary. If it is stationary, the mean of the field is the same in the whole domain, and the covariance between two locations depends only on the vector distance between them. For a non-stationary, the mean and the covariance can vary with the spatial locations. In general, stationary fields are easier to model and work with. However, for many processes, a stationary model makes a poor fit, and a non-stationary model might be better and more realistic.

One of the applications of spatial statistics, is spatial interpolation. This means that available observations of a random field at some locations are used to make estimations at locations where the values are unknown.

In this thesis, we investigate the spatial process precipitation in mountainous areas. The benefits of having good models for precipitation are many. One arises by the possibility to interpolate precipitation in catchment areas. By measuring the precipitation at a few weather stations, one can use interpolation to estimate the runoff from large areas. In particular, this is desired by the hydro power industry, which is the main source of electricity in Norway (Statkraft, 2009). It is a great advantage for this industry, to be able to estimate the expected inflow to dams. Another important advantage, is the possibility to make reliable flood warnings (NVE, 2015).

It is challenging to estimate the runoff in mountainous areas. One reason for this is precipitation's dependency on topography. It has been known for a long time that there is a close relation between orography and precipitation (Roe, 2005). Anyone with some experience from mountainous areas has observed the increase in intensity and frequency of precipitation as the mountains get higher and steeper. This can also be seen by the vegetation: On the windward flank of the mountain there is a humid local climate. Plants are affected by the soil water levels, as well as humidity of the air, and these areas have therefore a rich diversity of plants. On the lee side of the mountain, it is much dryer, which gives a shortage of water in the soil, and consequently there is less vegetation. The cause of these differences is the phenomenon orographic precipitation, which arises when moist air moves over a mountain. As the air rises against the mountain side, it gets cooled down, and looses its ability to keep water. The water condensates and falls out as precipitation. A further introduction to orographic precipitation can be found in Roe (2005).

Another complication when predicting runoff, is that the weather stations closest to the catchment areas in general lie at a lower altitude than the lowest points in the catchment area (Ingebritsen, 2014). This is due to easier and cheaper maintenance of the weather stations. Hence, the effect on precipitation by moving towards the higher peaks may not always be detected in the data.

In order to handle the difficulties with orographic precipitation, and the high located catchment areas, it seems reasonable to consider a model that includes elevation. In Ødegård (2017), this was tested by comparing a model including elevation as an explanatory variable with a model without any elevation term. Both models were stationary. The study showed that when doing interpolations at locations with higher elevations than the closest weather stations, a model including elevation gives a better fit. However, the interpolations were still not as good as desired, and needed improvement. In particular, the model seemed to underestimate the runoff from a catchment area considered in the case study. In Ingebritsen (2014), a study of interpolating precipitation in southern Norway is carried out. The study compares a non-stationary model with a stationary model, both including elevation. The study shows that the non-stationary model gives a better fit in mountainous terrain.

This thesis is a continuation of the studies in Ødegård (2017), and the objective is to make a model for interpolating precipitation in areas characterised by orographic precipitation. We apply the results from Ingebritsen (2014) by considering a non-stationary model with dependency structure varying with elevation, and comparing it with a stationary model. We wish to further investigate the differences between a stationary and non-stationary model when the considered domain is smaller, and the whole area is characterised by orographic precipitation. The aim of the study is manifold. Firstly, the study has applied purposes. We aim to make a good model that can be used to interpolate precipitation in areas characterised by orographic precipitation, and to predict the total precipitation in the area. Such a model should be able to quantify the uncertainty of the interpolations, and have a good predictive performance. It is also desired to get a better understanding of the precipitation process in this kind of terrain, e.g., how precipitation varies with elevation. We intent to investigate weather a stationary or non-stationary model, with dependency structure varying with elevation, is preferred in this kind of terrain and climate. Secondly, the study has a statistical purpose. We aim to obtain a better knowledge about non-stationary and stationary modelling of spatial processes. In particular, we wish to explore how large degree of non-stationarity there has to be in a process, before it is detectable and relevant for the model-fitting.

In order to accomplish this, we do a case study, using real observations of precipitation, and a simulation study. In addition, we use simple toy examples to further explore the consequences of having dependency structure that varies with elevation. In the studies, the models are tested by doing interpolations of precipitation to points, based on point observations of precipitation. They are also tested by doing predictions of the total precipitation in catchment areas, based on point observations of precipitation. We refer to the total amount of precipitation as areal precipitation. Our study region is the county Hordaland, which is located south-west in Norway. The south-western and south-eastern parts of Norway are separated by the mountain range Langfjella (Ingebritsen, 2014), and Hordaland is located on the eastern, windward side of the mountains. The climate there is characterised by orographic precipitation, and is therefore well-suited for testing our models.

When making our models, there are certain properties that need to be included. One is a general spatial concept, known as Tobler's first law of geography: "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). This law simply states that random variables that are located spatially closer to each other tend to have a higher dependency than locations further apart. To include this property in our models, we consider random fields with appropriate dependency structures. Another property we need to include, is the dependency of precipitation on elevation. In both models we assume a linear relation, and include elevation as an explanatory variable. In the non-stationary model we also include elevation in the dependency structure.

We build our models as latent Gaussian models, which is a subclass of hierarchical models, within a Bayesian framework. Due to their flexibility, hierarchical models are very applicable and nice to work with (Blangiardo and Cameletti, 2015). The models consist of several components, or levels. At each level, the models can be simple or complex, which give rise to the variety of models in this class. At the first level, we have the data model, which specifies the likelihood of the observation, given some latent (unobserved) variables and some hyperparameters. At the next level, we have the process model, which is a probability model for the spatial process, given some parameters. At the lowest level, we have the parameter model, which specifies the prior distributions of the hyperparameters. As we are using a Bayesian framework, we need to assign prior distributions to all variables, and the aim is typically to make inference about their posterior distributions or do predictions.

We use the stochastic partial differential equation (SPDE) approach to spatial modelling, and the methodology integrated nested Laplace approximation (INLA) to do inference about our models, and to do interpolations. This is a deterministic algorithm for Bayesian inference, specially designed to effectively handle latent Gaussian models. Traditionally, Markov chain Monte Carlo (MCMC) has been used to do inference about Bayesian hierarchical models. The reason for choosing INLA, is that this method is computationally more efficient, with the same accuracy. We use the R-package R-INLA to do all model fitting, inference and interpolation. The package is available at `www.r-inla.org`, and information about the package can be found in Lindgren and Rue (2013).

We now give a short introduction to some of the notation used throughout the thesis. In general, bold symbols are used to denote vectors and matrices. We use $\boldsymbol{s}$ to denote a set of locations, and $\boldsymbol{s}_i$ to denote location i. Every location is represented by two coordinates, i.e., $\boldsymbol{s}_i = (s_{i,1}, s_{i,2})$. The true precipitation at location i in year j is denoted $\eta_j(\boldsymbol{s}_i)$, whereas the observed precipitation is denoted Y_{ij}. We use $\pi(\cdot)$ to denote a probability distribution.

The thesis is outlined as follows: In Chapter 2, we introduce the study region and the observations. In Chapter 3, we introduce some of the background and underlying theory. In Chapter 4, we build our models and choose the prior distributions. In Chapter 5, we present a simple toy example, illustrating the relevance of the dependency structure. In Chapter 6, we perform a case study, using real observations of precipitation. In Chapter 7, we do a simulation study. In Chapter 8, we discuss the results from the studies.

Chapter 2

Study region, data and explanatory analysis

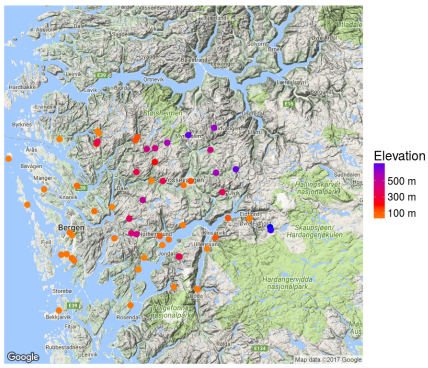
In this chapter, we present our study region and the data set of annual precipitation. Further, we do some simple explanatory analysis.

2.1 Study region and geographical data

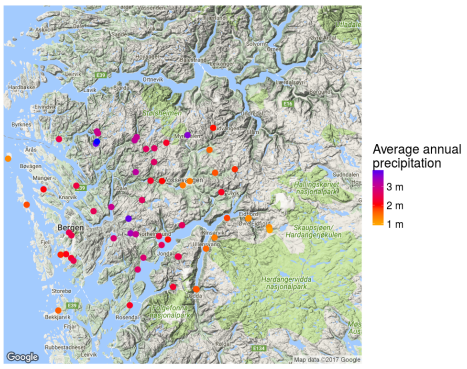
We consider observations of annual precipitation in the county Hordaland in Norway. Hordaland is a mountainous county, mainly located on the western, windward side of Langfjella, where the climate is characterised by orographic precipitation. As the aim of this thesis is to make good models for precipitation in mountainous areas characterised by orographic precipitation, Hordaland is a suited area to perform our investigations. However, there is a watershed on the mountain plateau Hardangervidda, and the areas on the eastern side of the watershed have much drier climatology than the rest of the county. In Ødegård (2017), it was shown that the different climatology at the east side leads to a poor fit for the weather stations located in this area. Parts of the municipality Ulvik lie in this eastern area, and is excluded from this study.

We consider 60 weather stations, located in Hordaland. Information about the weather stations, such as coordinates, elevation and annual precipitation, has been provided by the Norwegian Water Resources and Energy Directorate (NVE). The observations are public accessible at the climate database of the Norwegian Metrological Institute, through the web portal `eKlima.no`. An overview of the weather stations, with name, station number, elevation, coordinates and average annual precipitation, is given in Appendix C. Figure 2.1a displays a map of

Hordaland, including the 60 weather stations represented by coloured dots. The colour scale of the dots gives the elevation in meters above sea level (m.a.s) of the weather stations.



(a) Elevation ([m]) at the weather stations.



(b) Average annual precipitation ([m/years]) at the weather stations.

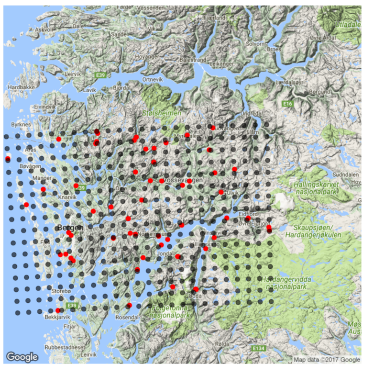
Figure 2.1: Maps of Hordaland, including dots representing the 60 weather stations. The colour scale of the dots, gives the elevation ([m]) and average annual precipitation ([m/year]) at the weather stations.

Because our models include elevation, we need to know the elevation at all locations where we do inference and interpolation. As a part of the SPDE approach, we use a mesh (see Section 3.3), thus we also need to know the elevation at the mesh nodes. The elevation at the weather stations is known. The elevation at other locations is found using Google Maps Elevation API, which is an online service offered by Google. The service provides elevation at every location at the surface of the earth. At locations where Google does not posses the exact elevation, the service interpolates the elevation using the four nearest locations. For more information about Google Maps Elevation API, see Google (2017).

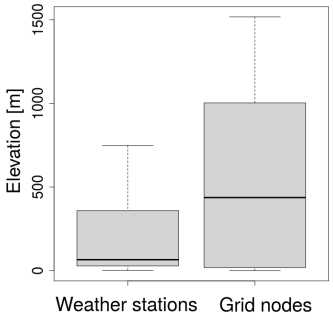
In the study, we consider a grid covering Hordaland. Figure 2.2a presents a map of Hordaland, including the grid. The red dots represent the weather stations. Figure 2.2b displays boxplots of the elevation at the weather stations and at the grid nodes. The plot shows that the mean elevation is much higher at the grid nodes, and there is a larger variance in the elevation there. The elevation is given in m.a.s.

2.2 Annual precipitation data

We use observations of annual precipitation at the weather stations from the period 1981-2015. Annual precipitation is given in [m/year]. We use hydrological



(a) Grid covering Hordaland, including the 60 weather stations (red dots).



(b) Boxplot of the elevation at the weather stations and grid nodes.

Figure 2.2: Illustration of a grid covering Hordaland, and a boxplot of the elevation [m] at the weather stations and at the grid nodes.

years, which go from 1st September to 31st August, such that the year 1981 starts 1st September 1980 and ends 31st August 1981. The data are given in a monthly format, and we have obtained annual precipitation by adding the monthly observations. Most of the weather stations do not have data from all the years, and there are differences in which weather stations that have observations from which years. In the case study, we use all available data each years. If a weather station misses data from one or more months a given year, the whole year is left out. The total number of observations we use is $N = 984$.

Figure 2.1b displays a map of Hordaland, including dots representing the 60 weather stations, where the colour scale of the dots gives the average annual precipitation ([m/year]) at the stations. The average precipitation of all weather stations is 2.4 m/year.

2.3 Catchment areas

We consider three catchment areas in Hordaland. A map of Hordaland, including the catchment areas, is displayed in Figure 2.3. The blue areas represent the catchments, and the red dots represent the weather stations. We consider the total precipitation in the catchments, which we refer to as areal precipitation.

We use observations of annual runoff and evaporation, to obtain observations of areal precipitation. Table 2.1 gives the area [m^2], mean elevation [m], annual runoff [10^8 m^3] per square meter and annual evaporation [10^8 m^3] per square meter of the catchment areas. The mean elevation is found by considering a

grid inside the catchment areas and take the mean of the elevation at the grid nodes. The grid has a resolution of 1 grid node/km^2. The areas, mean annual runoff mean annual evaporation is provided by NVE. We do not consider the measurement error of the observations of runoff and evaporation. This is because these errors are very small, compared to the errors we obtain when estimating the areal precipitation (Roksvåg, 2016).

Table 2.1: Area [m^2], mean elevation [m], mean annual runoff [10^8 m^3] per square meter and mean annual evaporation [10^8 m^3] per square meter of the catchment areas. The mean elevation is found by considering a grid inside the catchment areas and take the mean of the elevation at the grid nodes. The grid has a resolution of 1 grid node/km^2.

Catchment area	Area [m^2]	Mean elevation [m]	Mean annual runoff/m^2	Mean annual evaporation/m^2
1. Fjellanger	12.8	903	0.316	0.20
2. Svartavatn	72.3	715	2.20	0.27
3. Slondalsvatn	41.9	1197	0.99	0.15

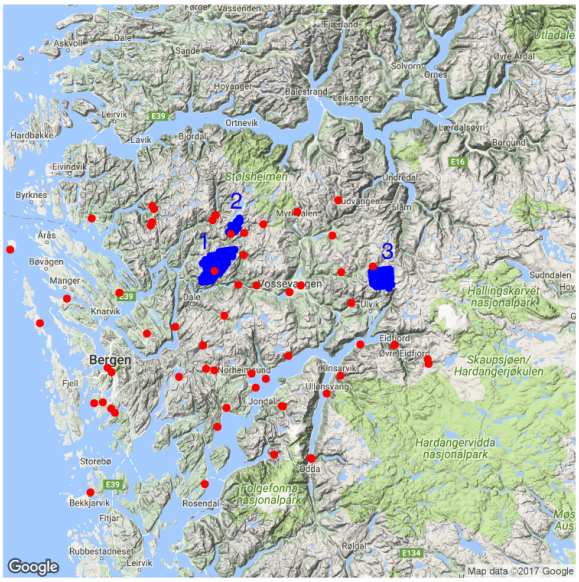


Figure 2.3: A map of Hordaland, including the weather stations (red dots) and catchment areas (blue areas). The numbers correspond to the following catchment areas: 1: Fjellanger, 2: Svartavatn, 3: Slondalsvatn.

2.4 Explanatory analysis

Figure 2.4 displays a boxplot of the annual precipitation at all weather stations for each year in the period 1981-2015. The plot shows that the average annual precipitation of all weather stations varies in the period, and also that the variance in annual precipitation at different weather stations varies.

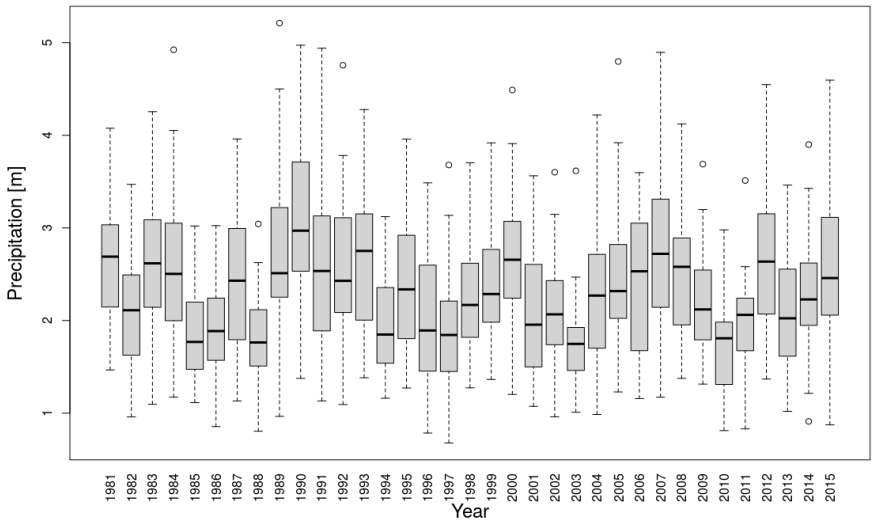


Figure 2.4: Annual precipitation at every weather station for each year in the period 1981-2015.

Figure 2.5 displays a boxplot of the annual precipitation for all years in the period 1981-2015 at each weather station. The plot shows that there are large variations in the average annual precipitation at different weather stations. It also shows that the variance in annual precipitation at different weather stations varies, and, in particular, it seems to increase as the average annual precipitation increases.

Figure 2.6 shows the average annual precipitation at each weather station, plotted against the elevation of the weather station. A simple linear regression is performed on the data set, and the regression line is plotted in red. The line is slightly increasing, i.e., there seem to be a linear increase in annual precipitation with elevation.

Figure 2.7 presents the correlation between the different weather stations, plotted against the distance [km] between them. The plot only shows correlations between weather stations that have observations for more than 10 of the same years. The plot shows that when the distance between weather stations is both small and large, the correlation might be small or large.

Figure 2.8 displays a plot of the annual precipitation [m] at six weather stations, along with a map, showing the locations of the weather stations. The plot illustrates that the weather stations seem to vary in the same way, indicating some correlation.

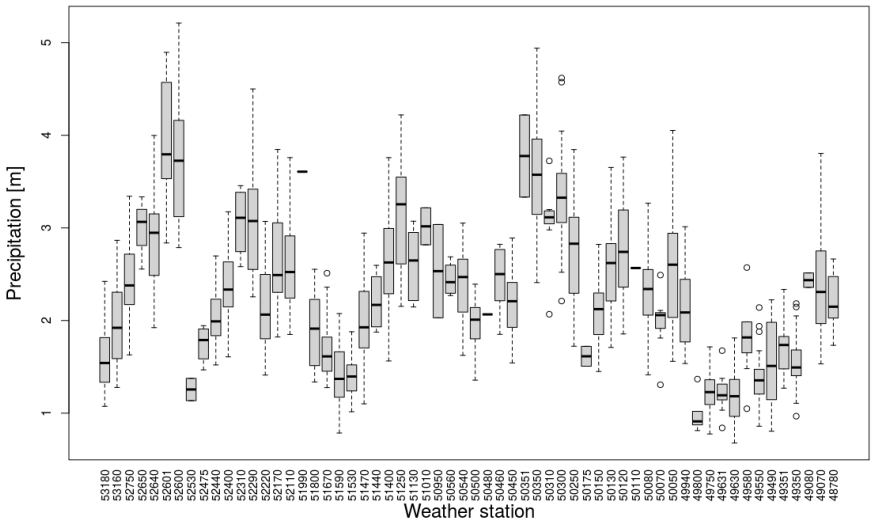


Figure 2.5: Annual precipitation for every year in the period 1981-2015 at each weather station. The weather stations are represented by their station number, and an overview of the weather stations with station number is given in Appendix C.

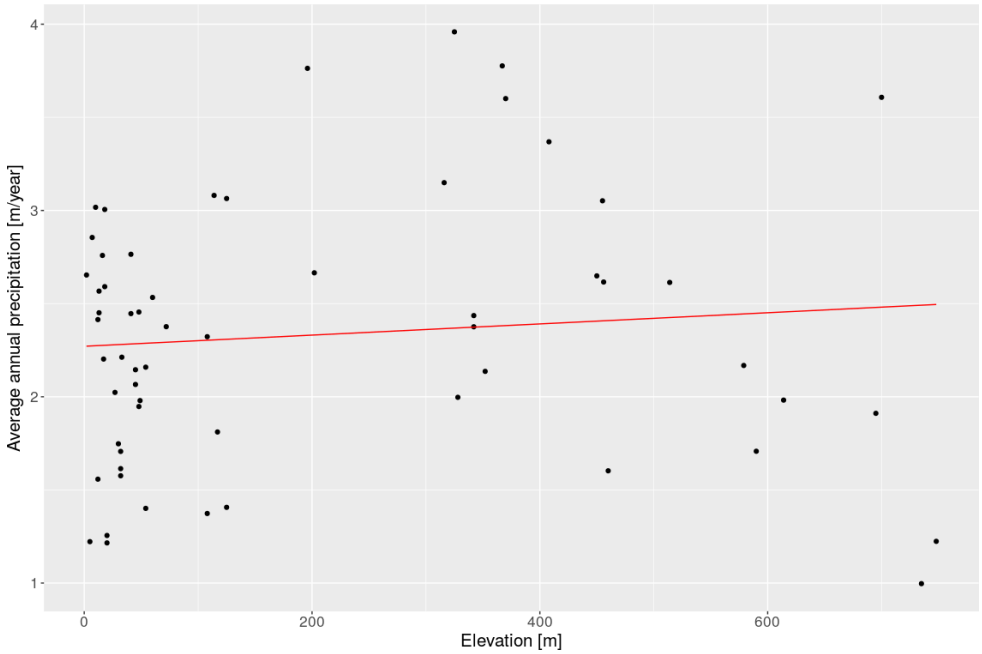


Figure 2.6: Average annual precipitation at each weather station, plotted against the elevation of the weather station. The red line is a linear regression line.

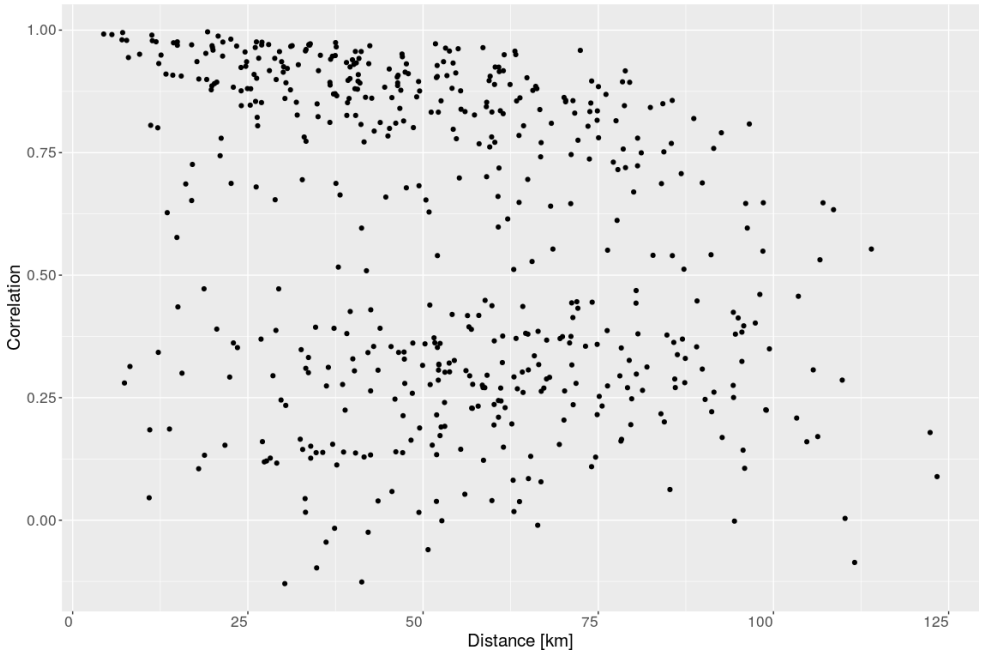
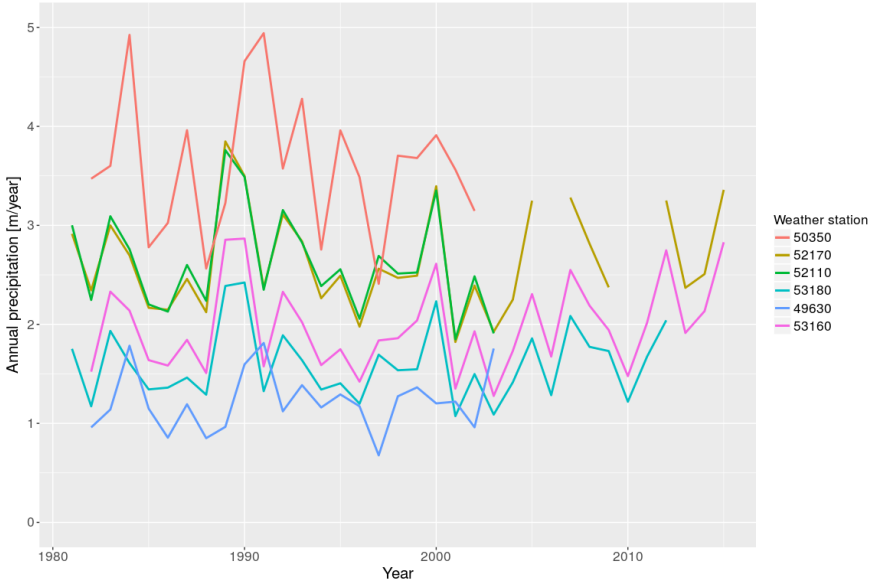


Figure 2.7: Correlation between annual precipitation at the different weather stations, plotted against the distance [km] between them.



(a) Annual precipitation [m] at six weather stations.

(b) A map showing the location of the 6 weather stations.

Figure 2.8: Annual precipitation [m] at six weather stations. The map displays the location of the weather stations.

Chapter 3

Background and underlying theory

In this chapter, we introduce some of the background and underlying theory needed to build our models and do inference about them. For readability, we name references at the beginning of each section, instead of naming them during the text.

3.1 Gaussian random fields

We start out with an introduction to random fields and, in particular, Gaussian random fields. The section is based on Gelfand et al. (2010) and Lindgren et al. (2011).

A random field is a probability model for a stochastic variable that varies over a continuous domain , $\mathcal{D} \subseteq \mathbb{R}^d$. Here, d is the dimension of the domain, which typically is 2 or 3. In this work, we consider a two-dimensional domain, i.e. $d = 2$. At any location $\boldsymbol{s} \in \mathcal{D}$, we consider $Y(\boldsymbol{s})$ a random variable, and at any fixed, finite set of locations $[\boldsymbol{s}_1, ..., \boldsymbol{s_n}] \in \mathcal{D}$ we consider $[Y(\boldsymbol{s}_1), ..., Y(\boldsymbol{s}_n)]$ a random vector, with spatial dependency given by its multivariate distribution. A random field $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D} \subseteq \mathbb{R}^d\}$ is specified by the finite-dimensional joint distributions

$$F(y_1, ..., y_n; \boldsymbol{s}_1, ..., \boldsymbol{s}_n) = P(Y(\boldsymbol{s}_1) \leq y_1, ..., P(Y(\boldsymbol{s}_n) \leq y_n), \tag{3.1}$$

for any n and every finite set of locations $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ in $\mathcal{D}$.

A special case of random fields, is Gaussian random fields, GRFs, which is a popular and common model class for spatial phenomena. For GRFs, the joint probability distributions (3.1) are multivariate normal, i.e., $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D} \subseteq \mathbb{R}^d\}$ is a GRF if

$$[Y(\boldsymbol{s}_1), ..., Y(\boldsymbol{s}_n)] \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix which characterise the GRF. For a definition and some properties of the multivariate normal distribution, see Appendix A.

A random field is said to be strictly stationary if the joint probability distributions (3.1) are invariant under spatial shift, i.e., if

$$F(y_1, ..., y_n; \boldsymbol{s}_1 + \boldsymbol{h}, ..., \boldsymbol{s}_n + \boldsymbol{h}) = F(y_1, ..., y_n; \boldsymbol{s}_1, ..., \boldsymbol{s}_n),$$

for any vectors $\boldsymbol{h}$ in $\mathbb{R}^d$. It is said to be second order stationary, or weakly stationary, if the mean and covariance are invariant to spatial shift. This means that the mean is constant in the whole domain, and the covariance between two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ in the domain only depends on the vector difference $\boldsymbol{s}_i - \boldsymbol{s}_j$, i.e.

$$\mathrm{E}[Y(\boldsymbol{s_i})] = \mathrm{E}[Y(\boldsymbol{s_i} + \boldsymbol{h})] = \mu$$

and

$$\mathrm{Cov}[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)] = \mathrm{Cov}[Y(\boldsymbol{s}_i + \boldsymbol{h}), Y(\boldsymbol{s}_j + \boldsymbol{h})] = C(\boldsymbol{s}_i - \boldsymbol{s}_j),$$

where the function $C(\cdot)$ is a covariance function. A covariance function may be any positive definite function. In the case of GRFs, weak stationarity also implies strict stationarity, due to properties of the Gaussian distribution. This is a special property for GRFs, and it is not true in general. In the following we refer to strictly and weakly stationary GRFs, as simply stationary GRFs.

When the covariance between two locations in the domain only depends on the Eucledian distance between the locations, i.e.,

$$\mathrm{Cov}[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)] = C(||\boldsymbol{s}_i - \boldsymbol{s_j}||),$$

we have an isotropic GRF and an isotropic covariance function $C(\cdot)$.

The most popular and commonly used family of isotropic covariance functions, is the family of Matérn covariance functions. We now give a short introduction to the Matérn covariance functions, due to the application they have in the stochastic partial differential equation (SPDE) approach to spatial statistics, which we introduce in Section 3.3.

Consider two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ in $\mathcal{D}$. The Matérn covariance function between the locations is defined as

$$C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa||\boldsymbol{s}_i - \boldsymbol{s}_j||)^\nu K_\nu(\kappa||\boldsymbol{s}_i - \boldsymbol{s}_j||) \quad \nu > 0, \quad \kappa > 0. \tag{3.2}$$

Here, $||\cdot||$ is the Euclidean distance, $\Gamma(\cdot)$ is the gamma function, K_ν is the modified Bessel function of second kind and order ν, ν is a shape parameter which controls the smoothness of the field, κ is a scale parameter and σ^2 is the marginal variance. The scale parameter κ has a more natural interpretation when considering the range instead. The range is the Euclidean distance between two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ where $Y(\boldsymbol{s}_i)$ and $Y(\boldsymbol{s}_j)$ are almost independent. For the Matérn covariance functions the range is empirically derived to be

$$\rho = \frac{\sqrt{8\nu}}{\kappa}. \tag{3.3}$$

At this distance the correlation is approximately 0.1.

Due to the dependency structure of GRFs, both the covariance matrix, $\boldsymbol{\Sigma}$, and the precision matrix, $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$, are typically dense. This gives matrix operations, such as inversion and factorisation, a computational cost of $O(n^3)$, where n is the dimension of the matrix. As this is computationally expensive, it is desired to reduce the computational cost. This gives the motivation to consider Gaussian Markov random fields, which we in the next section shall see are much faster to work with.

3.2 Gaussian Markov random fields

In this section, we give an introduction to Gaussian Markov random fields, GMRFs, motivated by their computational benefit. The following introduction is based on Rue and Held (2005) and Rue et al. (2009). We have adopted the definition of a GMRF from Rue and Held (2005), and the use of graphs to represent the conditional independence structure in a GMRF. Before we can present GMRFs, we start out with a short introduction to graphs, followed by an explanation of the notation we use.

An undirected graph is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of vertices, or nodes, and $\mathcal{E}$ is a set of edges, $\{i, j\}$, where $i, j \in \mathcal{V}$ and $i \neq j$. There is an edge between nodes i and j in the graph if, and only if, $\{i, j\} \in \mathcal{E}$. If we assume $\mathcal{V} = \{1, ..., n\}$, we have a labelled graph. An example of an undirected, labelled graph can be seen in Figure 3.1. This graph has nodes $\mathcal{V} = \{1, 2, 3, 4\}$ and edges $\mathcal{E} = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 2\}\}$.

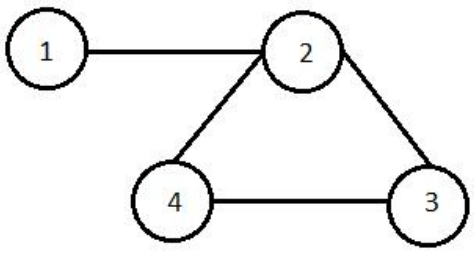


Figure 3.1: An undirected, labelled graph, with nodes $\mathcal{V} = \{1, 2, 3, 4\}$ and edges $\mathcal{E} = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 2\}\}$.

Let $\boldsymbol{Y} = [Y_1, ..., Y_n]$ be a Gaussian distributed random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Denote by $\boldsymbol{Y}_{-ij}$ all the elements of $\boldsymbol{Y}$, except the elements Y_i and Y_j. Furthermore, let $Y_i \perp Y_j | \boldsymbol{Y}_{-ij}$ denote conditional independence between Y_i and Y_j given $\boldsymbol{Y}_{-ij}$. Conditional independence is implied if and only if the following is true:

$$\pi(Y_i, Y_j, \boldsymbol{Y}_{-ij}) = \pi(Y_i, \boldsymbol{Y}_{-ij})\pi(Y_j, \boldsymbol{Y}_{-ij}).$$

Now, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected, labelled graph, with $\mathcal{V} = 1, ..., n$ and with $\mathcal{E}$ such that there is no edge between i and j if, and only if, $Y_i \perp Y_j | \boldsymbol{Y}_{-ij}$. Then we say that $\boldsymbol{Y}$ is a GMRF with respect to $\mathcal{G}$.

An important result about GMRFs is about the precision matrix, $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$. It can be proven that if $\boldsymbol{Y}$ is Gaussian with mean $\boldsymbol{\mu}$ and positive definite precision matrix $\boldsymbol{Q}$, then

$$Y_i \perp Y_j | \boldsymbol{Y}_{-ij} \iff Q_{ij} = 0 \quad \text{for} \quad i \neq j.$$

A proof is given in Rue and Held (2005), and is reproduced in Appendix A. This result simply says that if Y_i and Y_j are conditional independent, then $Q_{ij} = 0$. This brings us to the formal definition of a GMRF:

A random vector $\boldsymbol{Y} = [Y_1, ..., Y_n] \in \mathbb{R}^n$ is called a GMRF with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and positive definite precision matrix $\boldsymbol{Q}$, if, and only if, its density has the form

$$\pi(\boldsymbol{Y}) = (2\pi)^{-n/2} |\boldsymbol{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{Q} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

and

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

The advantage of GMRFs arises as a result of the Markov property. The local

Markov property can be stated as

$$Y_i \perp \boldsymbol{Y}_{-iNe(i)} | \boldsymbol{Y}_{Ne(i)} \quad \text{for every } i \in \mathcal{V}.$$

Here, $Ne(i)$ is the neighbourhood of node i, i.e., all the nodes in $\mathcal{G}$ which have an edge to node i. This property says that $\pi(Y_i|\boldsymbol{Y}_{-i}) = \pi(Y_i|\boldsymbol{Y}_{ne(i)})$, which loosely speaking means that the conditional distribution of Y_i given everything else, only depends on elements in the neighbourhood of i. Due to the non-zero structure of the precision matrix, $\boldsymbol{Q}$, of a GRMF with the Markov property, the precision matrix is sparse. This makes matrix operations, such as factorisation, much faster for GMRFs than for GRFs. For two-dimensional GMRFs the computational cost is $\boldsymbol{O}(n^{3/2})$ and for three-dimensional GMRFs it is $\boldsymbol{O}(n^2)$.

3.3 The stochastic partial differential equation

By now, we have seen how GRFs are preferred when doing spatial modelling, but that GMRFs are computationally faster. In the following, we shall see how the stochastic partial differential equation (SPDE) approach lets us do the modelling using GRFs, but by using a GRMF representation of the GRF, it enables fast computations. The following section is based on Lindgren et al. (2011) and Ingebritsen (2014).

We start out by introducing the SPDE, which has the following form:

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0. \quad (3.4)$$

Here, $\mathcal{W}$ is spatial Gaussian white noise and Δ is the Laplacian, defined by

$$\Delta = \sum_{i=1}^{d} \frac{\partial^d}{\partial x_i^d}.$$

Further, κ controls the spatial range, τ controls the variance, ν and α control the smoothness and d is the dimension of the domain. The stationary solutions to Equation (3.4), are GRFs with Matérn covariance functions. The family of Matérn covariance functions was defined in Equation (3.2) in Section 3.1. The parameters of the Matérn covariance function and the SPDE are the same. The marginal variance is given by

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}, \quad (3.5)$$

where $\Gamma(\cdot)$ is the gamma function.

The link between a GRF and GMRF, proposed by Lindgren et al. (2011), is a finite element representation of the solutions to the SPDE (3.4). The finite element method is a numerical scheme for approximating solutions of partial differential equations. The spatial domain $\mathcal{D}$ is covered by a triangulated mesh with m nodes. The solution to the SPDE (3.4) is then given a finite element representation

$$x(\boldsymbol{s}) = \sum_{i=1}^{m} \psi_i(\boldsymbol{s}) w_i, \tag{3.6}$$

for some basis functions $\{\psi_i\}$ and some Gaussian distributed weights $\boldsymbol{w}$. Lindgren et al. (2011) choose to use basis functions that are piecewise linear in each triangle, i.e. ψ_i has value 1 in node i and 0 in all other nodes. Thus, in the nodes, the value of the field is determined by the weights, whereas in the interior of the triangles, the value of the field is determined by linear interpolation. Figure 3.2 illustrates a triangulation. Figure 3.2a shows a smooth surface, and Figure 3.2b shows a linear approximation of the surface, obtained using a triangular mesh and piecewise linear basis functions. The illustrations are copied from Hu and Steinsland (2016).

We need to choose the weights $\boldsymbol{w}$ so that the distribution of the representation (3.6) approximates the solutions to the SPDE (3.4) on the domain. Let the precision matrix of the Gaussian weights $\boldsymbol{w}$ be denoted $\boldsymbol{Q}_{\alpha,\kappa^2}$, where α and κ are the parameters in the SPDE (3.4), with $\alpha = 1, 2, ...$ By using Neumann boundary conditions, $\boldsymbol{Q}_{\alpha,\kappa^2}$ is given by

$$\begin{cases} \boldsymbol{Q}_{1,\kappa^2} = \boldsymbol{K}_{\kappa^2}, & \text{for } \alpha = 1 \\ \boldsymbol{Q}_{2,\kappa^2} = \boldsymbol{K}_{\kappa^2}\boldsymbol{C}^{-1}\boldsymbol{K}_{\kappa^2}, & \text{for } \alpha = 2 \\ \boldsymbol{Q}_{\alpha,\kappa^2} = \boldsymbol{K}_{\kappa^2}\boldsymbol{C}^{-1}\boldsymbol{Q}_{\alpha-2,\kappa^2}\boldsymbol{C}^{-1}\boldsymbol{K}_{\kappa^2}, & \text{for } \alpha = 3, 4, ... \end{cases} \tag{3.7}$$

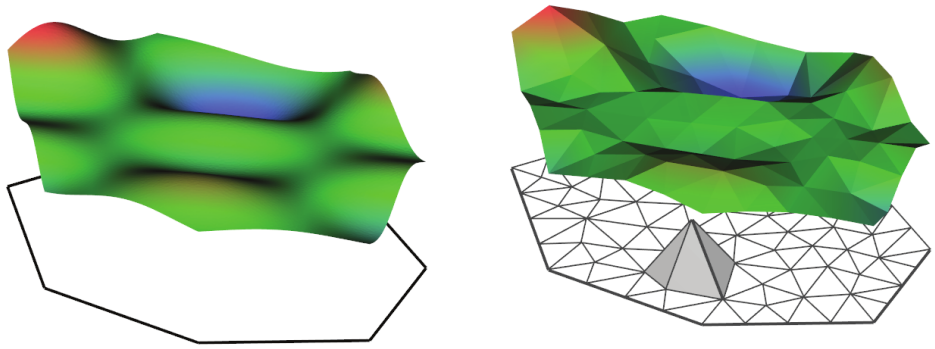
The $m \times m$ matrix $\boldsymbol{K}_{\kappa^2}$ has elements

$$(K_{\kappa^2})_{ij} = \kappa^2 C_{ij} + G_{ij},$$

and the $m \times m$ matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ have elements

$$C_{ij} = \langle \psi_i, \psi_j \rangle, \quad G_{ij} = \langle \nabla\psi_i, \nabla\psi_j \rangle.$$

The matrix $\boldsymbol{C}^{-1}$ is dense, which also makes the precision matrix dense, but $\boldsymbol{C}$ can be approximated by a diagonal matrix $\tilde{\boldsymbol{C}}$, with elements $\tilde{C}_{ii} = \langle \psi_i, 1 \rangle$. By using this approximation, the precision matrix $\boldsymbol{Q}_{\alpha,\kappa^2}$ becomes sparse. This means we have Gaussian distributed weights with sparse precision matrix, and hence the weights $\boldsymbol{w}$ are a GMRF with Markov properties determined by the triangulation.



(a) A smooth surface.

(b) A linear approximation of a smooth surface.

Figure 3.2: Figure 3.2a displays a smooth surface, and Figure 3.2b illustates a linear approximation of the surface, obtained using a triangular mesh and piecewise linear basis functions. The figures are copied from Hu and Steinsland (2016).

In the following we will assume $d = 2$ and $\nu = 1$, as this will be the dimension and shape parameter we use in the studies in this thesis. This implies $\alpha = 2$.

One of the advantages with the SPDE approach, is its flexibility. The properties of the random field, which usually are determined by a covariance function, are now characterised by the SPDE. By modifying the SPDE we can obtain the desired properties of the field. In particular, it is possible to extend the SPDE to model non-stationarity. The SPDE automatically ensures a valid model with positive semi-definite non-stationary covariance. An extended version of the SPDE (3.4) is obtained by allowing the parameter κ, which is linked to the spatial correlation range, and the parameter τ, which controls the marginal variance of the field, to vary in space, i.e.

$$(\kappa(\boldsymbol{s})^2 - \Delta)(\tau(\boldsymbol{s})x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \quad \kappa > 0. \tag{3.8}$$

When one or both parameters are non-constant, we have non-stationarity, because the variance and range vary in space.

The parameters $\tau(\boldsymbol{s})$ and $\kappa(\boldsymbol{s})$ can be represented by log-linear models

$$\log\tau(\boldsymbol{s}) = \theta_{\tau,1} + \sum_j b_{\tau,j}(\boldsymbol{s})\theta_{\tau,j} \quad \text{and} \quad \log\kappa(\boldsymbol{s}) = \theta_{\kappa,1} + \sum_k b_{\kappa,k}(\boldsymbol{s})\theta_{\kappa,k}.$$

Here, $\{b_{\tau,j}(\cdot)\}$ and $\{b_{\kappa,k}(\cdot)\}$ are deterministic basis functions, and the θ's are weights. Spatial explanatory variables can be used as basis functions.

By letting the parameters τ and κ vary in space, the precision matrix of the Gaussian weights $\boldsymbol{w}$ from the finite element representation (3.6) change. For $\alpha = 2$, the precision matrix $\boldsymbol{Q}_{N-S}$ is now given by

$$\boldsymbol{Q}_{N-S} = \boldsymbol{T}(\boldsymbol{K}^2\boldsymbol{C}\boldsymbol{K}^2 + \boldsymbol{K}^2\boldsymbol{G} + \boldsymbol{G}\boldsymbol{K}^2 + \boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G})\boldsymbol{T}. \tag{3.9}$$

The $m \times m$ matrices $\boldsymbol{T}$ and $\boldsymbol{K}$ are diagonal with elements $T_{ii} = \tau(\boldsymbol{s}_i)$ and $K_{ii} = \kappa(\boldsymbol{s}_i)$, the $m \times m$ matrix $\boldsymbol{C}$ is diagonal with elements $C_{ii} = \int \psi_i(\boldsymbol{s}) d\boldsymbol{s}$ and the $m \times m$ matrix $\boldsymbol{G}$ is sparse positive semi-definite, with elements $G_{ij} = \int \nabla\psi_i(\boldsymbol{s}) \nabla\psi_j(\boldsymbol{s}) d\boldsymbol{s}$. Thus, the precision matrix is still sparse, and we still have a GMRF representation.

The marginal variance defined in (3.5), and the corresponding range given in (3.3), are only valid in the stationary case. Nominal approximations of the variance and range are

$$\sigma^2(\boldsymbol{s}) \approx \frac{1}{4\pi\kappa(\boldsymbol{s})^2\tau(\boldsymbol{s})^2} \tag{3.10}$$

and

$$\rho(\boldsymbol{s}) \approx \frac{\sqrt{8}}{\kappa(\boldsymbol{s})}. \tag{3.11}$$

These approximations are valid for slowly varying $\kappa(\boldsymbol{s})$. The approximations help us interpret the parameters.

3.4 Latent Gaussian models

In this thesis, we use models from the class of latent Gaussian models, LGMs, which is a special case of Bayesian hierarchical models. This is a popular and much used class of models, due to its flexibility. We now give an introduction to Bayesian hierarchical models and LGMs, based on Rue et al. (2009), Ingebritsen (2014) and Blangiardo and Cameletti (2015).

The name hierarchical models refers to the composition of the models. The models consist of several components, or levels. At each level the models can be simple or complex, which give rise to the variety of models in this class. The joint distribution of all the levels is found by multiplying the conditional distributions for each level. In this thesis we consider hierarchical models with three levels, thus, for simplicity, we restrict our introduction to only include models of this kind.

At the top level of the hierarchical model is the data model, which specifies the likelihood of the observations, depending on some latent (unobserved) process and some hyperparameters. On the next level, we have the process model, which is a probability model for the latent process, given some parameters. Thus, the latent process is itself a hierarchical model. At the lowest level is the parameter model, i.e., the prior distribution of the hyperparameters. In the case of Bayesian hierarchical models, the parameter model is the joint prior distribution for all the model parameters. The aim is typically to make inference about the latent process and hyperparameters, based on the observations.

We now proceed to define an LGM for n observations $y_1, ..., y_n$ at locations $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$. Assume the observations belong to the exponential family, and that they have the following data model:

$$y_i|\boldsymbol{x}, \boldsymbol{\theta}_1 \sim \pi(y_i|\eta_i, \boldsymbol{\theta}_1).$$

Here, $\boldsymbol{x}$ are some latent (unobserved) parameters and $\boldsymbol{\theta}_1$ are hyperparameters, which not necessarily are Gaussian. We assume that the observations y_i are conditionally independent given $\boldsymbol{x}$ and $\boldsymbol{\theta}_1$. The mean μ_i is defined as a function of the additive predictor η_i through a link function $g(\cdot)$, such that $g(\mu_i) = \eta_i$. The linear predictor η_i is defined as

$$\eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_{ki} \sum_{j=1}^{n_f} f^{(j)}(u_{ji}).$$

Here, α is the intercept, the $\{\beta_k\}$'s are linear effects of the covariates $\boldsymbol{z}$ and the $\{f^{(j)}(\cdot)\}$'s are unknown functions of the covariates $\boldsymbol{u}$.

The latent parameters of the model are $\boldsymbol{x} = [\boldsymbol{\eta}, \alpha, \boldsymbol{\beta}, \boldsymbol{f}]$. As we are assuming a Bayesian model, we assign priors to the latent parameters. By letting the priors be Gaussian, we have an LGM, with a latent Gaussian field $\boldsymbol{x}$, i.e.,

$$\boldsymbol{x}|\boldsymbol{\theta}_2 \sim \mathcal{N}(0, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2)).$$

Here, $\boldsymbol{Q}(\boldsymbol{\theta_2})$ is a precision matrix, depending on some hyperparameters $\boldsymbol{\theta}_2$, which are not necessarily given Gaussian priors. To simplify the notation in the following, we denote $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$. We assume that the latent parameters are independent, hence the precision matrix $\boldsymbol{Q}(\boldsymbol{\theta}_2)$ is sparse, and the latent Gaussian field $\boldsymbol{x}$ is a GMRF.

It remains to assign priors to the hyperparameters $\boldsymbol{\theta}$, i.e.,

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}).$$

The joint posterior distribution of the latent parameters and the hyperparameters, given the observations, is now given as the product of the data likelihood, the density of the latent parameters and the prior of the hyperparameters, i.e.,

$$\begin{aligned}\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta}) \prod_{i=1}^{n} \pi(y_i|x_i, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta})|\boldsymbol{Q}(\boldsymbol{\theta}|^{1/2} \exp\left[-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x} + \sum_{i=1}^{n} \log(\pi(y_i|x_i, \boldsymbol{\theta}))\right]. \qquad (3.12)\end{aligned}$$

3.5 Integrated nested Laplace approximation

In the previous, section we introduced latent Gaussian models (LGMs), with a Bayesian approach. In this section, we introduce the method integrated nested Laplace approximation (INLA), and we show how INLA can be used to make approximate Bayesian inference about LGMs. The following section is based on Rue et al. (2009) and Ingebritsen (2014).

3.5.1 Approximate Bayesian inference with INLA

Consider a latent Gaussian field $\boldsymbol{x}$ and hyperparameters $\boldsymbol{\theta}$ of an LGM. In Section 3.4 the joint posterior distribution of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ was given in Equation (3.12). The aim now, is to approximate their posterior marginal distributions, given the observations $\boldsymbol{y}$. The posterior marginal distribution of the latent variables may be written as

$$\pi(x_i|\boldsymbol{y}) = \int \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})\mathrm{d}\boldsymbol{\theta} \quad i = 1, ..., n, \tag{3.13}$$

and the marginal posterior distribution of the hyperparameters may be written as

$$\pi(\theta_j|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y})\mathrm{d}\boldsymbol{\theta}_{-j} \quad j = 1, ..., m. \tag{3.14}$$

The INLA approach consists of three steps; (i) find an approximation to $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, (ii) find an approximation to $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ and (iii) use a numerical integration scheme to compute the integrals (3.13) and (3.14). To ensure fast computations, there are some requirements that should be fulfilled. The first is that the number of hyperparameters, $\boldsymbol{\theta}$, should not be too large. This is because we perform numerical integration over the parameter space. Another requirement is that the latent Gaussian field, $\boldsymbol{x}$, should be given a GMRF prior with a sparse precision matrix, as sparse matrix algorithms are used. We now give a brief presentation of the three INLA steps listed above. We refer to Rue et al. (2009) for details.

We start out by approximating the marginal posterior distribution of the hyperparameters $\boldsymbol{\theta}$. The following Laplace approximation $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is used:

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}, \tag{3.15}$$

where $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation of the full conditional distribution of $\boldsymbol{x}$, and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional of $\boldsymbol{x}$ for a given $\boldsymbol{\theta}$. The reason for the proportionality sign, is that the normalising constant is unknown. The mode of $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is found by optimising $\log(\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}))$ with respect to

$\boldsymbol{\theta}$, using some numerical optimisation algorithm. When performing the numerical integration, it is sufficient to have some good evaluation points where $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is evaluated. These are found by computing the numerical Hessian, evaluated at the mode, and then the Hessian is used to distribute the evaluation points, $\{\boldsymbol{\theta_k}\}$.

The next step is to approximate $\pi(x_i|\boldsymbol{\theta_k}, \boldsymbol{y})$ for each of the evaluation points in $\{\boldsymbol{\theta}_k\}$. There are several ways to do this. In Rue et al. (2009), three approximations are suggested; Gaussian approximation, Laplace approximation and simplified Laplace approximation. Laplace approximation is a method used to approximate integrals on the form $\int f(x)\mathrm{d}x$, for some function $f(\cdot)$. The idea is to fit a Gaussian distribution at the maximum x^* of $f(x)$, and use the cumulative Gaussian distribution to approximate the integral. The Laplace approximation is the most accurate approximation of the three suggested above, and in general, this one is preferred. The simplified Laplace approximation has a slight loss in accuracy, compared to the Laplace approximation, but it compensates by having a much smaller cost. The Gaussian approximation is simplest and cheapest, and it might give reasonable results, but it might also lead to errors.

Using the Laplace approximation to $\pi(x_i|\boldsymbol{\theta_k}, \boldsymbol{y})$, we obtain

$$\tilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{\pi}_G(x_i|\boldsymbol{x}_{-i}, \boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-i}(x_i,\boldsymbol{\theta})}. \tag{3.16}$$

Here $\tilde{\pi}_G$ is the Gaussian approximation to $x_i|\boldsymbol{x}_{-i}, \boldsymbol{\theta}, \boldsymbol{y}$, and $\boldsymbol{x}^*_{-i}(x_i, \boldsymbol{\theta})$ is its mode for a given x_i and $\boldsymbol{\theta}$.

The final step is to approximate the integrals (3.13) and (3.14) using numerical integration and the approximations (3.15) and (3.16).

When the observation likelihood is Gaussian, which is the case for LGMs, the approximations (3.15) and (3.16) are exact and can therefore be omitted. The accuracy of the resulting posteriors is then only determined by the accuracy of the numerical approximations of the integrals (3.13) and (3.14).

The INLA methodology is available in the R-package `R-INLA`. All inference and interpolations performed in this work are calculated using `R-INLA`, and some examples of INLA code are given in Appendix B.

3.6 Evaluation schemes

So far, we have described an approach for building Bayesian hierarchical models, making inference about them and doing interpolations. What remains is some

schemes that can be used to evaluate and compare the models and their predictive performances. In this section, we give an introduction to the schemes we use; the root-mean-square error (RMSE), the continuous ranked probability score (CRPS), coverage probability and cross-validation.

3.6.1 Root-mean-square error

One of the most used evaluation schemes for the predictive performance of a model, is to measure the average inaccuracy of the estimates produced by the model, i.e., the average residuals (Willmott and Matsuura, 2005). One type of such average measure, is the RMSE. Consider a variable $Y_i, i = 1, ..., n$, which we want to predict. Assume y_i is an observation of the variable, and that $\hat{y}_i$ is a predicted estimate of the variable. The RMSE is then defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}.$$

3.6.2 Continuous ranked probability score

This subsection is based on Hersbach (2000).

Let Y_i denote the variable we wish to predict, with y_i and $\hat{y}_i$, $i = 1, ..., n$, denoting an observation and predicted estimate of the variable, respectively. Further, let $F(\hat{y}_i)$ denote the predictive cumulative probability function of Y_i. Then, the CRPS is defined as

$$\text{CRPS} = \text{CRPS}(F, \hat{y}_i) = \int_{-\infty}^{\infty} [F(\hat{y}_i) - H(\hat{y}_i - y_i)]^2 \mathrm{d}\hat{y}_i,$$

where H is the Heaviside function, defined as

$$H(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0. \end{cases}$$

Thus, the CRPS is a measure of the distance between the predicted and observed cumulative distributions. This means that a better prediction, leads to a smaller value of the CRPS. In the case $\hat{y}_i = y_i$, i.e., a perfect prediction, the CRPS becomes 0. Note that for deterministic predictions, where the uncertainty is zero, the CRPS is the same as the mean absolute error.

An advantage of the CRPS, compared to the RMSE, is that the CRPS takes the whole posterior distribution into account. This is useful, when the aim is to quantify the uncertainty of predictions and interpolations.

In this thesis, the variable that is to be interpolated is annual precipitation, Y_{ij}, for location $i = 1, ..., n$ and year $j = 1, ..., J$. When evaluating the models, we are interested in the mean CRPS at location i over several years, which we denote $\overline{\text{CRPS}}_i$, and the total mean over several locations and several years, which we denote $\overline{\text{CRPS}}$. The total average is calculated as

$$\overline{\text{CRPS}} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{J}\sum_{j=1}^{J}\text{CRPS}(F, \hat{y}_{ij}).$$

In our analysis, we use the R-function `crps()` from the package `verification` to calculate the CRPS. This function assumes that the cumulative distribution is Guassian. As mean and standard deviation we use the posterior mean and posterior standard deviation of precipitation.

3.6.3 Paired samples t-test

When comparing the models, we would like to investigate if there is a significant difference between the CPRS of the models. To do that, we perform a paired samples t-test. In this section, we give a general introduction to paired samples t-test, based on Yeager (2017).

The paried samples t-test is a hypothesis test that tests if the mean of the differences between two related statistics is significantly different from zero. Consider to stochastic variables X_1 and X_2, and denote the difference between them $D = X_1 - X_2$. Assume the difference is normally distributed, such that $D \sim (\mu_D, \sigma_D^2)$. The hypotheses we test are then

$$\begin{gathered}\text{H}_0 : \mu_D = 0 \\ \text{vs} \\ \text{H}_1 : \mu_D \neq 0.\end{gathered}$$

Denote a sample of observations $x_{1,1}, ..., x_{1,n},\ x_{2,1}, ..., x_{2,n},$ with corresponding differences $d_1, ..., d_n$. We use the following test statistic:

$$t = \frac{\overline{d}}{\sqrt{\text{Var}(d)/n}},$$

with $d = \frac{1}{n}\sum_{i=1}^{n} d_i$ and $\text{Var}(d) = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2$.

We have that the test statistic t follows a student t distribution, with $n-1$ degrees of freedom. Thus, the hypothesis is rejected if the value of t is greater than the critical value $t_{1-\alpha,n-1}$, for a chosen confidence level $1-\alpha$.

3.6.4 Coverage probability

The coverage probability of prediction intervals is the proportion of times the true value of the variable occurs within the prediction interval. Thus, the coverage probability give an indication of a model's ability to make prediction, and to estimate the uncertainty of its predictions.

Let Y_i denote the variable we wish to predict, with y_i and $[\hat{y}_{i,L}, \hat{y}_{i,U}]$, $i = 1, ..., n$, denoting an observation and prediction interval of the variable, respectively. The coverage probability is then estimated by

$$\hat{\text{Cover}} = \frac{1}{n}\sum_{i=1}^{n} \text{I}_{[\hat{y}_{i,L}, \hat{y}_{i,U}]}(y_i).$$

Here, $\text{I}_{[\hat{y}_{i,L}, \hat{y}_{i,U}]}(\cdot)$ is the indicator function, defined as

$$\text{I}_{[\hat{y}_{i,L}, \hat{y}_{i,U}]}(y_i) = \begin{cases} 1 & \text{if } y_i \in [\hat{y}_{i,L}, \hat{y}_{i,U}] \\ 0 & \text{if } y_i \notin [\hat{y}_{i,L}, \hat{y}_{i,U}] \end{cases}.$$

3.6.5 Cross-validation

Cross-validation is an old, but useful, method for evaluation of the predictive performance of a model. k-fold cross-validation, is a general case of cross-validation. We now give an introduction to k-fold cross-validation, based on Kohavi (1995).

Let D be a data set, to which we want to fit a model. In k-fold cross-validation, this data set is divided into k non-overlapping subsets, $D_1, ..., D_k$, of approximately equal size. The following procedure is repeated k times: For each time $t \in \{1, ..., k\}$, let D_t be test data, and let the remaining data, $D \setminus D_t$, be training data. Fit the model to the training data, and then use the fitted model to predict the test data based on the training data. The predicted values can now be compared with the true values, and the model is evaluated, using e.g. RMSE or CRPS. After this has been done all k times, the mean RMSE or CRPS of the k times can be used as a measure of the predictive performance of the model.

Leave-one-out cross-validation is a special case of k-fold cross-validation. In this case, k is equal to the number of observations in the data set, hence, we only use one observation as test data each time. This is particularly useful if there are few observations available.

Chapter 4

Models and methods

In this chapter, we propose two models for annual precipitation; a stationary model and a non-stationary model. The models are of the class LGMs, which were presented in Section 3.4, i.e., the spatial fields are assumed to be stationary and non-stationary GRFs. We use the subscripts $N-S$ and S to denote non-stationary and stationary models or model components. When we refer to general cases, i.e., when what we describe is common for both models, we leave out the subscript.

4.1 Annual precipitation model

We start out by introducing a general data model and process model for annual precipitation.

Consider a vector of spatial locations $\boldsymbol{s} \in \mathcal{D}$, where $\mathcal{D}$ is a spatial domain, which in our case is the county Hordaland. Let the spatial process $\{\eta_j(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ denote the true precipitation in year j at locations $\boldsymbol{s}$. Further, let y_{ij} denote the observed precipitation at location $\boldsymbol{s}_i$ in year j, with $i = 1, ..., n$. We assume the following data model for our observations:

$$y_{ij} = \eta_j(\boldsymbol{s}_i) + \epsilon_{ij}. \tag{4.1}$$

Here, ϵ_{ij} is the measurement uncertainty at location $\boldsymbol{s}_i$ in year j. We assume the measurement uncertainties are independent of the true precipitation, and that

they are Gaussian iid, with

$$\boldsymbol{\epsilon}_j = [\epsilon_{1j}, ..., \epsilon_{nj}] \sim \mathcal{N}(0, \tau_\epsilon^{-1}\text{diag}(\boldsymbol{f}_j)).$$

The vectors $\boldsymbol{f_j} = [f_{1j}, ..., f_{nj}]$ are annual fixed scales, enabling the variance to vary at different locations. This is included because the variance is known to increase as the true precipitation increases. In particular, this was illustrated in the explanatory analysis in Chapter 2, in Figure 2.5. A reasonable way to model this, is by using $f_{ij} = (0.1\eta_j(\boldsymbol{s}_i))^2$, and this will be done in the simulation study. In the case study, we don't know the values of $\eta_j(\boldsymbol{s}_i)$, and thus it is not technically possible in INLA to use these scales. In the case study, we therefore use $f_{ij} = (0.1y_{ij})^2$ instead.

The process model is assumed to be compounded by several terms. We assume an annual intercept, α_j, and two spatial processes, modelled as GRFs. One of the spatial processes represents the spatial variations in precipitation that are the same every year, and which are caused by the climatology. We call this process the climatology, denoted by $c(\boldsymbol{s})$. The other spatial process represents the additional annual spatial variations, which are different each year. We call this the annual spatial variability, and denote it $z_j(\boldsymbol{s})$. Finally, we assume there is a linear relation between elevation and precipitation, and we model this as a term $\beta_h h(\boldsymbol{s})$, where β_h is the linear effect of the elevation $h(\boldsymbol{s})$ at location $\boldsymbol{s}$. This yields the following stationary and non-stationary process model for true precipitation:

$$\eta_{j,S}(\boldsymbol{s}) = \alpha_j + c_S(\boldsymbol{s}) + z_{j,S}(\boldsymbol{s}) + \beta_h h(\boldsymbol{s}) \tag{4.2}$$
$$\eta_{j,N-S}(\boldsymbol{s}) = \alpha_j + c_{N-S}(\boldsymbol{s}) + z_{j,N-S}(\boldsymbol{s}) + \beta_h h(\boldsymbol{s}). \tag{4.3}$$

4.2 The SPDE approach to the annual precipitation model

We now use the SPDE approach, introduced in Section 3.3, to model the GRFs $c_S(\boldsymbol{s})$, $c_{N-S}(\boldsymbol{s})$, $z_{j,S}(\boldsymbol{s})$ and $z_{j,N-S}(\boldsymbol{s})$ in (4.2) and (4.3). We construct a triangular mesh with m nodes, which covers our domain, $\mathcal{D}$. We consider the same mesh for all processes and all years. The mesh is displayed in Figure 4.1. The red dots represent the weather stations.

By using the stationary SPDE (3.4) and the non-stationary SPDE (3.8), we obtain

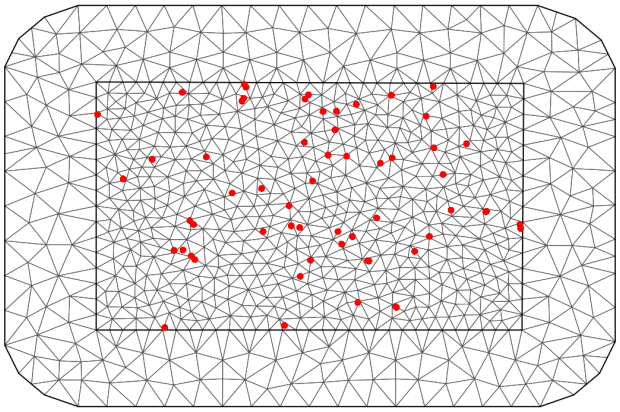


Figure 4.1: A triangular mesh. The red dots represent the weather stations.

the following finite element representations of the GRFs:

$$c_S(\boldsymbol{s}_i) = \sum_{k=1}^{m} \xi_k(\boldsymbol{s}_i) u_{k,S}, \quad c_{N-S}(\boldsymbol{s}_i) = \sum_{k=1}^{m} \xi_k(\boldsymbol{s}_i) u_{k,N-S},$$
$$z_{j,S}(\boldsymbol{s}_i) = \sum_{k=1}^{m} \xi_k(\boldsymbol{s}_i) w_{kj,S}, \quad z_{N-S_j}(\boldsymbol{s}_i) = \sum_{k=1}^{m} \xi_k(\boldsymbol{s}_i) w_{kj,N-S}$$

As explained in Section 3.3, $\{\xi_k(\cdot)\}$ are linear basis function, such that each sum is over three non-zero basis functions. As we are using the same mesh, the basis functions are also the same for all the processes and all years. The u_k's and w_{kj}'s are weights. For the climatology the weights are the same each year, whereas for the annual spatial variation, they are different.

The process models (4.2) and (4.3) can now be written as

$$\eta_{j,S}(\boldsymbol{s}_i) = \alpha_j + \sum_{k=1}^{m} \xi_k(\boldsymbol{s}_i)(u_{k,S} + w_{kj,S}) + \beta_h h(\boldsymbol{s})$$
$$\eta_{j,N-S}(\boldsymbol{s}_i) = \alpha_j + \sum_{k=1}^{m} \xi_k(\boldsymbol{s}_i)(u_{k,N-S} + w_{kj,N-S}) + \beta_h h(\boldsymbol{s}).$$

In vector form, this becomes

$$\eta_{j,S}(\boldsymbol{s}_i) = [1 \; h(\boldsymbol{s}_i) \; \boldsymbol{A}_i][\alpha_j \; \beta_h \; (\boldsymbol{u}_S + \boldsymbol{w}_{j,S})]^T \tag{4.4}$$
$$\eta_{j,N-S}(\boldsymbol{s}_i) = [1 \; h(\boldsymbol{s}_i) \; \boldsymbol{A}_i][\alpha_j \; \beta_h \; (\boldsymbol{u}_{N-S} + \boldsymbol{w}_{j,N-S})]^T, \tag{4.5}$$

where $\boldsymbol{A}$ is an $n \times m$ matrix, with $A_{ik} = \xi_k(\boldsymbol{s}_i)$, and $\boldsymbol{A}_i$ is the i'th row of $\boldsymbol{A}$. $\boldsymbol{A}$ is called a projection matrix, because it projects the values in the mesh nodes to

values in the locations $\boldsymbol{s}_i$. The expressions (4.4) and (4.5) are used as our linear predictors for annual precipitation.

We recall that the SPDE depends on some hyperparameters:

$$\theta_\tau(\boldsymbol{s}) = \theta_{\tau,1} + \sum_j b_{\tau,j}(\boldsymbol{s})\theta_{\tau,j} \quad \text{and} \quad \theta_\kappa(\boldsymbol{s}) = \theta_{\kappa,1} + \sum_k b_{\kappa,k}(\boldsymbol{s})\theta_{\kappa,k}.$$

For the stationary case, the parameters are constants, and we simply denote them $\boldsymbol{\theta}_S^* = [\theta_{\tau,1}, \theta_{\kappa,1}]$. In the non-stationary case, the parameters vary in space. We assume that the parameters are linearly dependent on elevation, which gives them the following form:

$$\theta_{\tau,N-S}(\boldsymbol{s}) = \theta_{\tau,1} + \theta_{\tau,h}h(\boldsymbol{s}) \quad \text{and} \quad \theta_{\kappa,N-S}(\boldsymbol{s}) = \theta_{\kappa,1} + \theta_{\kappa,h}h(\boldsymbol{s}). \tag{4.6}$$

We denote the parameters $\boldsymbol{\theta}_{N-S}^*(\boldsymbol{s}) = [\theta_{\tau,1}, \theta_{\tau,h}, \theta_{\kappa,1}, \theta_{\kappa,h}]$.

When we need to distinguish between the parameters corresponding to the climatology, $c(\boldsymbol{s})$, and annual spatial variability, $z_j(\boldsymbol{s})$, we include the subscript u (representing the climatology) and w (representing the annual spatial variability) in the notation of the parameters. Further, if it is not clear from the context if $\theta_{\tau,1}$ and $\theta_{\kappa,1}$ refers to the stationary or non-stationary model, we add the subscript S or $N-S$, respectively.

4.3 Prior assumptions

As we are using a Bayesian approach, we need to assign prior distributions.

We assign Gaussian priors to the intercept α_j and elevation coefficient, β_h. This implies that the latent variables, $\boldsymbol{x}_j = [\alpha_j, \beta_h, \boldsymbol{u}, \boldsymbol{w}_j]$, are jointly Gaussian distributed, given the SPDE parameters $\boldsymbol{\theta}^*$. They thereby fit into the framework of LGMs, described in Section 3.4. We choose informative priors for the intercept, α_j, and elevation coefficient β_h. We choose the intercept mean to be 2 m/year, which is close to the mean of the observations. The variance is chosen to be 1. When doing the model-fitting in R, we scale elevation, such that it is given in km. For the elevation coefficient we choose the mean to be 1 m/year, i.e., an annual increase of 1 m precipitation, as the elevation increases by 1 km. The variance is chosen to be 1^2. The choice of the prior distribution for the elevation coefficient is based on the results of the investigations in Ødegård (2017). The distributions become

$$\alpha_j \sim \mathcal{N}(2, 1)$$

and

$$\beta_h \sim \mathcal{N}(1, 1).$$

Finally, we assign prior distributions to all the hyperparameters, $\boldsymbol{\theta} = [\tau_\epsilon, \boldsymbol{\theta}^*]$. For the measurement uncertainty precision, τ_ϵ, we choose a gamma distribution,

$$\tau_\epsilon \sim \text{Gamma}(10, 10).$$

This implies

$$\text{E}[\tau_\epsilon] = 1 \quad \text{and} \quad \text{Var}[\tau_\epsilon] = 0.1.$$

When choosing prior distributions for the hyperparameters of the dependency structure, $\boldsymbol{\theta}^*$, we need to be careful. Several studies, e.g., Ingebritsen (2014), have demonstrated that these parameters are very prior sensitive. As the parameters τ and κ don't have any clear physical interpretation, it is difficult to choose informative priors. We follow the procedure demonstrated in Ingebritsen (2014) when choosing the prior distributions. We use the same prior distributions for the hyperparameters of the climatology, $c(\boldsymbol{s})$, and the annual spatial variability, $z(\boldsymbol{s})$.

We start out with choosing prior distributions for the parameters in the stationary case, $\boldsymbol{\theta}^*_S$. We assume that the parameters are Gaussian, i.e., $\theta_{\tau,1,S} \sim \mathcal{N}(\mu_{\tau,1,S}, \sigma^2_{\tau,1,S})$ and $\theta_{\kappa,1,S} \sim \mathcal{N}(\mu_{\kappa,1,S}, \sigma^2_{\kappa,1,S})$, and that $\theta_{\tau,1,S}$ and $\theta_{\kappa,1,S}$ are independent. In order to specify the parameters $\mu_{\tau,1,S}$, $\sigma^2_{\tau,1,S}$, $\mu_{\kappa,1,S}$ and $\sigma^2_{\kappa,1,S}$, we exploit the relationships between the parameters and the marginal deviance of the field,

$$\sigma_S = \frac{1}{\sqrt{4\pi}\tau\kappa},$$

and the spatial range,

$$\rho_S = \frac{\sqrt{8}}{\kappa}.$$

These parameters have a physical interpretation, which can be used to assign them informative prior distribution. We recall that $\theta_\tau = \log(\tau)$ and $\theta_\kappa = \log(\kappa)$. By properties of the log-normal distribution, we have that

$$\rho_S \sim \log\mathcal{N}(\log\sqrt{8} - \mu_{\kappa,1,S}, \sigma^2_{\kappa,1,S})$$

and

$$\sigma_S \sim \log\mathcal{N}(-\log\sqrt{4\pi} - \mu_{\tau,1,S} - \mu_{\kappa,1,S}, \sigma^2_{\tau,1,S} + \sigma^2_{\kappa,1,S}).$$

We can now make use of the lognormal quantile functions, which are given by

$$\rho_S(p) = \sqrt{8}\exp(-\mu_{\kappa,1,S} + \sigma_{\kappa,1,S}\Phi^{-1}(p)) \tag{4.7}$$

for the range, and

$$\sigma_S(p) = \frac{1}{\sqrt{4\pi}}\exp(-\mu_{\tau,1,S} - \mu_{\kappa,1,S} + \sqrt{\sigma^2_{\tau,1,S} + \sigma^2_{\kappa,1,S}}\Phi^{-1}(p)) \tag{4.8}$$

for the marginal deviance. Here $0 < p < 1$ is some quantile, and $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. By assuming values for the spatial range and marginal deviance at two chosen quantiles, we can solve the set of equations (4.7) and (4.8), in order to get values for the unknown parameters $\mu_{\tau,1,S}$, $\sigma^2_{\tau,1,S}$, $\mu_{\kappa,1,S}$ and $\sigma^2_{\kappa,1,S}$.

We assume values for the spatial range and marginal deviance at the 0.5 quantile and 0.9 quantile, when solving the equations (4.7) and (4.8). The domain we consider is approximately 95km × 140km. We choose the 0.5 quantile of the spatial range to be 30 km and the 0.9 quantile to be 95 km. Using these values to solve Equation (4.7), yield $\mu_{\kappa,1,S} = -2.36$ and $\sigma^2_{\kappa,1,S} = 0.81$. For the marginal deviance, we assume the 0.5 quantile to be 0.2 m and the 0.9 quantile to be 2 m. Solving Equation (4.8) for these values yields $\mu_{\tau,1,S} = 2.71$ and $\sigma^2_{\tau,1,S} = 2.42$.

Also in the non-stationary case we assign Gaussian priors to the hyperparameters, $\boldsymbol{\theta}^*_{N-S}$, i.e.,

$$\begin{aligned}\theta_{\tau,1,N-S} &\sim \mathcal{N}(\mu_{\tau,1,N-S}, \sigma^2_{\tau,1,N-S}), \quad \theta_{\tau,h,N-S} \sim \mathcal{N}(\mu_{\tau,h,N-S}, \sigma^2_{\tau,h,N-S})\\ \theta_{\kappa,1,N-S} &\sim \mathcal{N}(\mu_{\kappa,1,N-S}, \sigma^2_{\kappa,1,N-S}), \quad \theta_{\kappa,h,N-S} \sim \mathcal{N}(\mu_{\kappa,h,N-S}, \sigma^2_{\kappa,h,N-S}).\end{aligned}$$

Again, we need to specify the unknown parameters, $\mu_{\tau,1,N-S}$, $\sigma^2_{\tau,1,N-S}$, $\mu_{\tau,h,N-S}$, $\sigma^2_{\tau,h,N-S}$, $\mu_{\kappa,1,N-S}$, $\sigma^2_{\kappa,1,N-S}$, $\mu_{\kappa,h,N-S}$ and $\sigma^2_{\kappa,h,N-S}$. However, in the non-stationary case, we don't have exact expressions for the spatial range and the marginal deviance. Instead, we use the nominal approximations (3.11) and (3.10). The log-normal distributions then become

$$\rho_{N-S}(h) \sim \log\mathcal{N}(\log\sqrt{8} - \mu_{\kappa,1,N-S} - h\mu_{\kappa,h,N-S}, \sigma^2_{\kappa,1,N-S} + h^2\sigma^2_{\kappa,h,N-S})$$

and

$$\begin{aligned}\sigma_{N-S}(h) \sim \log\mathcal{N}(&-\log\sqrt{4\pi} - \mu_{\tau,1,N-S} - \mu_{\kappa,1,N-S} - h(\mu_{\tau,h,N-S} + \mu_{\kappa,h,N-S}),\\ &\sigma^2_{\tau,1,N-S} + \sigma^2_{\kappa,1,N-S} + h^2(\sigma^2_{\tau,h,N-S} + \sigma^2_{\kappa,h,N-S})),\end{aligned}$$

with $h = h(\boldsymbol{s})$. We require that the distributions fulfil the conditions proposed in Ingebritsen (2014). The conditions are:

1. $\rho_{N-S}(0) \overset{d}{=} \rho_S \quad \text{and} \quad \sigma_{N-S}(0) \overset{d}{=} \sigma_S$

2. $\mu_{\tau,h,N-S} = \mu_{\kappa,h,N-S} = 0$

3. Given a reference elevation h_0, c_ρ is the coefficient of variation for the ratio $\rho_{N-S}(h_0)/\rho_{N-S}(0)$, and c_σ is the coefficient of variation for the ratio $\sigma_{N-S}(h_0)/\sigma_{N-S}(0)$

The first condition defines that the non-stationary distributions of the spatial range and the marginal deviance at sea-level are the same as the stationary distributions. The second condition states that the means of the distributions do not depend on elevation, implying that the prior assumption is that the distributions are not affected by elevation. The third condition controls how much the priors vary with elevation, by specifying the coefficients of elevation. We can now use ρ_S, σ_S, h_0, c_ρ and c_σ to define the prior distributions of $\boldsymbol{\theta}^*_{N-S}$ through the following equations:

$$\begin{aligned}
&\mu_{\tau,1,N-S} = \mu_{\tau,1,S}, \quad \sigma^2_{\tau,1,N-S} = \sigma^2_{\tau,1,S},\\
&\mu_{\tau,h,N-S} = 0, \quad \sigma^2_{\tau,h,N-S} = \frac{1}{h_0^2}\log\left(\frac{c_\sigma^2+1}{c_\rho^2+1}\right),\\
&\mu_{\kappa,1,N-S} = \mu_{\kappa,1,S}, \quad \sigma^2_{\kappa,1,N-S} = \sigma^2_{\kappa,1,S},\\
&\mu_{\kappa,h,N-S} = 0, \quad \sigma^2_{\kappa,h,N-S} = \frac{1}{h_0^2}\log(c_\rho^2+1),
\end{aligned}$$

for a given reference elevation h_0, and some coefficients of variation c_ρ and c_σ. In order to ensure positive variance, we must choose coefficients of variations such that $c_\sigma > c_\rho$.

We choose h_0 to be 0.4 km, and we set the coefficients of variation to be $c_\rho = 0.8$ and $c_\sigma = 1.3$. This gives $\sigma^2_{\tau,h,N-S} = 3.09$ and $\sigma^2_{\kappa,h,N-S} = 3.09$.

4.4 Approximate variance of interpolated precipitation

Both when finding the CRPS and when calculating the prediction intervals used to find coverage probability of interpolated precipitation, we need an estimate of the variance of the interpolated precipitation. This is not given by `R-INLA`, so we need to calculate it.

Denote all the observations of precipitation $\boldsymbol{y}_{obs}$, and denote an interpolated estimate of precipitation at one location given all the observations $\hat{y}_{pred}|\boldsymbol{y}_{obs}$. The actual observed precipitation at the location of interpolation is denoted y_{pred}. What we are interested in, is an estimated of the variance of the interpolated precipitation, given all observations, i.e., $\text{Var}[\hat{y}_{pred}|\boldsymbol{y}_{obs}]$. We know from our data model (4.1) that

$$\hat{y}_{pred}|\boldsymbol{y}_{obs} = \hat{\eta}_{pred}|\boldsymbol{y}_{obs} + \hat{\epsilon}_{pred}|\boldsymbol{y}_{obs}.$$

We approximate $\text{Var}[\hat{y}_{pred}|\boldsymbol{y}_{obs}]$ by assuming that $\hat{\eta}_{pred}$ and $\hat{\epsilon}_{pred}$ are conditionally independent, given all observations, and that $\hat{\epsilon}_{pred}|\boldsymbol{y}_{obs}$ follow a normal distribution. This yields

$$\text{Var}[\hat{y}_{pred}|\boldsymbol{y}_{obs}] \approx \text{Var}[\hat{\eta}_{pred}|\boldsymbol{y}_{obs}] + \text{Var}[\hat{\epsilon}_{pred}|\boldsymbol{y}_{obs}] = \text{Var}[\hat{\eta}_{pred}|\boldsymbol{y}_{obs}] + \tau_{\epsilon}^{-1} f_{pred}.$$

The variance of $\hat{\eta}_{pred}|\boldsymbol{y}_{obs}$ is provided by `R-INLA`. The precision, τ_ϵ, is one of the hyperparameters, and we use the posterior mean, estimated by `R-INLA`. As scale, f_{pred}, we use $f_{pred} = (0.1\eta_{pred})^2$, in the simulation study, where η_{pred} is the true precipitation at the location of interpolation. In the case study, we use $f_{pred} = (0.1\hat{y}_{pred}|\boldsymbol{y}_{obs})^2$. Generally, $\hat{\eta}_{pred}$ and $\hat{\epsilon}_{pred}$ are dependent, but due to how the INLA methodology is set up, this dependency is not quantified.

4.5 Interpolation to area

To be able to estimate the runoff from catchment areas, is a reason why we need a good model to interpolate precipitation. Therefore, we consider three catchments, where we estimate the annual areal precipitation. The catchments were presented in Chapter 2. In the following, we describe the procedure used to estimate the annual areal precipitation.

We consider the annual areal precipitation in each catchment D_i, $i = 1, 2, 3$, a stochastic variable, denoted by $A_j(D_i)$ for year j. The annual areal precipitation is modelled as the integral of the true precipitation over the area of the catchments, i.e.,

$$A_j(D_i) = \int_{\boldsymbol{s} \in D_i} \eta_j(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}.$$

We do a numerical approximation to the integral. Inside each catchment, a grid is constructed, denoted $\mathcal{L}_{D_i}$. The true precipitation is assumed known or estimated in all grid nodes. The approximation becomes

$$A_j(D_i) = \Delta_{D_i} \sum_{\boldsymbol{s}_i \in \mathcal{L}_{D_i}} \eta_j(\boldsymbol{s}_i).$$

Here, Δ_{D_i} is an area element in catchment D_i, obtained by divinding the area of D_i by the number of grid nodes in the grid $\mathcal{L}_{D_i}$.

When predicting the annual areal precipitation, the annual precipitation in all the grid nodes in the catchments are interpolated jointly. The grids have an approximate resolution of 1 grid node/km^2. We fit the stationary and non-stationary

models to observations of annual precipitation at the 60 weather stations, and use the models to interpolate precipitation at the grid nodes inside the catchments, based on the weather stations. The predictor becomes

$$\hat{A}_j(D_i) = \sum_{\boldsymbol{s}_i \in \mathcal{L}_{D_i}} \hat{\eta}_j(\boldsymbol{s}_i).$$

Thus, the predictor for areal precipitation is a linear combination of the interpolated precipitation in the grid nodes.

We compare the predicted areal precipitation to the observations presented in Chapter 2. We have observations of runoff and evaporation at the catchments, and obtain observed areal precipitation by adding runoff and evaporation.

Chapter 5

Toy examples: The relevance of the dependency structure

In this chapter, we present some simple toy examples. The purpose of the examples is to further enlighten the consequences of having dependency structure that varies with elevation. We aim to obtain a better understanding of the model and the effect of the hyperparameters, and their physical interpretation. The explorations are carried out by simulating non-stationary data with varying dependency structure parameters, $\boldsymbol{\theta}^*$, and fitting stationary and non-stationary models to the data.

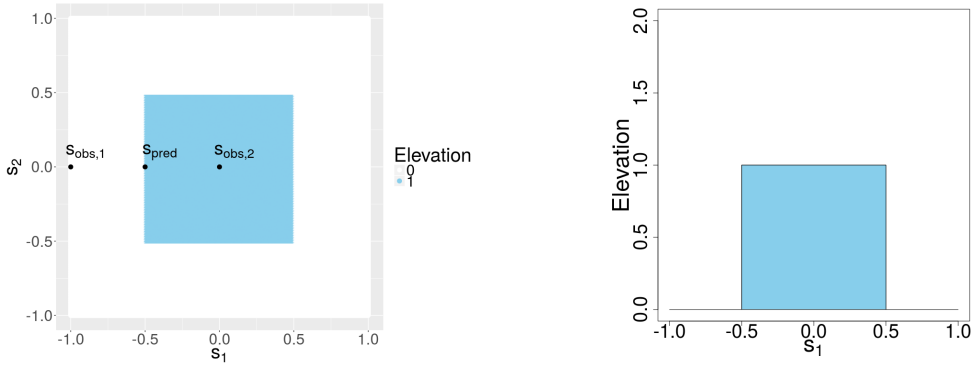
5.1 Toy example design and analysis

Consider the domain $\mathcal{D} : \{\boldsymbol{s} = (s_1, s_2) : |s_1| \leq 1, |s_2| \leq 1\}$, i.e. a square in $\mathbb{R}^2$. Assume there is a table shaped mountain inside the domain, such that

$$h(s_1, s_2) = \begin{cases} 1 & \text{if} \quad |s_1| \leq 0.5, |s_2| \leq 0.5 \\ 0 & \text{else,} \end{cases}$$

where $h(s_1, s_2)$ is the elevation at location $\boldsymbol{s} = (s_1, s_2)$. Illustrations of the domain with the mountain can be seen in Figure 5.1.

Consider two locations, $\boldsymbol{s}_{obs,1}$ and $\boldsymbol{s}_{obs,2}$, in the domain, where precipitation is observed. The locations are given coordinates $\boldsymbol{s}_{obs,1} = (-1, 0)$ and $\boldsymbol{s}_{obs,2} = (0, 0)$. Further, consider a location $\boldsymbol{s}_{pred}$, where precipitation is to be interpolated. This location is given coordinates $\boldsymbol{s}_{pred} = (-0.5, 0)$. All three locations are displayed



(a) The domain, viewed from above. **(b)** The mountain, viewed from the side.

Figure 5.1: Illustrations of the domain, including a table shaped mountain, used in the toy examples. The blue area represents the mountain.

in Figure 5.1a. Assume the precipitation at the three locations follow a non-stationary Gaussian random field with zero mean, where the parameters of the dependency structure are on the form given in Equation (4.6), i.e., linearly dependent on elevation.

In the following, the precipitation at locations $\boldsymbol{s}_{obs,1}$, $\boldsymbol{s}_{obs,2}$ and $\boldsymbol{s}_{pred}$ is denoted $Y_{obs,1}$, $Y_{obs,2}$ and Y_{pred}, respectively. Further, a vector of the observed precipitation is denoted $\boldsymbol{Y}_{obs} = [Y_{obs,1}\ Y_{obs,2}]^T$.

Consider the joint distribution of the variables $Y_{obs,1}$, $Y_{obs,2}$ and Y_{pred}. In this simple example, we have that the variables follow a multivariate normal distribution:

$$\begin{bmatrix} \boldsymbol{Y}_{obs} \\ Y_{pred} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{obs} & \Sigma \\ \Sigma^T & \sigma^2_{pred} \end{bmatrix} \right).$$

Here, Σ_{obs} is the covariance matrix of the observations, σ^2_{pred} is the marginal variance of Y_{pred}, and Σ is the vector of covariances between $\boldsymbol{Y}_{obs}$ and Y_{pred}. By properties of the normal distribution (see Appendix A), the conditional distribution of $Y_{pred}|\boldsymbol{Y}_{obs} = \boldsymbol{y}_{obs}$ is

$$Y_{pred}|\boldsymbol{Y}_{obs} = \boldsymbol{y}_{obs} \sim \mathcal{N}\left(\Sigma\Sigma^{-1}_{obs}\boldsymbol{y}_{obs}, \sigma^2_{pred} - \Sigma\Sigma^{-1}_{obs}\Sigma^T\right).$$

This implies that the interpolated precipitation $\hat{Y}_{pred}$ is a linear combination of the observations, and the weights of the linear combinations depend on the dependency structure variables $\boldsymbol{\theta}^*$, i.e.,

$$\hat{Y}_{pred} = w_1(\boldsymbol{\theta}^*)Y_{obs,1} + w_2(\boldsymbol{\theta}^*)Y_{obs,2}, \tag{5.1}$$

for some weights $w_1(\boldsymbol{\theta}^*)$ and $w_2(\boldsymbol{\theta}^*)$.

Next, we consider the physical interpretation of $\boldsymbol{\theta}^*$. Recall that the range of a non-stationary field is approximately

$$\rho(\boldsymbol{s}) \approx \frac{\sqrt{8}}{\kappa(\boldsymbol{s})},$$

with $\log(\kappa(\boldsymbol{s})) = \theta_{\kappa,1} + h\theta_{\kappa,h}$.

Thus, the range of the model depends on elevation. As an example, assume now that the values of $\theta_{\kappa,1}$ and $\theta_{\kappa,h}$, are such that the range at sea-level is 0.8, whereas at elevation $h = 1$, it is 0.3. Then the correlation between $Y_{obs,1}$ and Y_{pred} is approximately 0.2, whereas the correlation between $Y_{obs,2}$ and Y_{pred} is approximately 0.01. Thus, there is a considerable correlation between $Y_{obs,1}$ and Y_{pred}, whereas there is almost no correlation between $Y_{obs,2}$ and Y_{pred}. Obviously, similar results yields for other values of the parameters $\theta_{\kappa,1}$ and $\theta_{\kappa,h}$, when the range at sea level is higher than the distance between $Y_{obs,1}$ and Y_{pred} and the range at the top of the mountain is smaller than the distance between $Y_{obs,2}$ and Y_{pred}.

Next, we recall that the marginal variance of a non-stationary field is approximately

$$\sigma^2(\boldsymbol{s}) \approx \frac{1}{4\pi\kappa(\boldsymbol{s})^2\tau(\boldsymbol{s})^2},$$

with $\log(\tau(\boldsymbol{s})) = \theta_{\tau,1} + h\theta_{\tau,h}$ and $\log(\kappa(\boldsymbol{s})) = \theta_{\kappa,1} + h\theta_{\kappa,h}$.

This means that the marginal variance of precipitation also depends on the elevation. If the values of the parameters $\boldsymbol{\theta}^*$ are such that the marginal variance is much larger or much smaller at elevation $h = 1$ than at sea level, there might be large differences in the values of $Y_{obs,1}$ and $Y_{obs,2}$, and this amplifies the effect of the range.

These properties are reflected in the weights $w_1(\boldsymbol{\theta}^*)$ and $w_2(\boldsymbol{\theta}^*)$, in the linear interpolation (5.1) of Y_{pred}. If we fit a non-stationary model to observations of a process where the range varies as just described, $Y_{obs,1}$ has a higher weight than $Y_{obs,2}$. For a stationary model, on the other hand, the correlation only depends on the distance between the locations, and thus, the weights w_1 and w_2 are equal.

5.2 Toy example simulations

The non-stationary model given in Equation (4.3) is used to simulate precipitation. The model is simplified by leaving out the climatology, $c(\boldsymbol{s})$, the intercept and the linear effect of elevation, i.e., we simulate from a Gaussian random field, $z(\boldsymbol{s})$.

In the simulations, the constant parts of the dependency structure parameters are given values $\theta_{\tau,1} = -1.8$ and $\theta_{\kappa,1} = 1.4$. This corresponds to range $\rho = 0.70$ and marginal variance $\sigma^2 = 0.18$ at sea level. The explanatory variables in the dependency structure, $\theta_{\tau,h}$ and $\theta_{\kappa,h}$, are varied. We let $\theta_{\tau,h} = -8, ..., -1$ and $\theta_{\kappa,h} = 2, 3, 4$. With these values, we have that at elevation $h = 1$, the range decreases as the value of $\theta_{\kappa,h}$ increases, and it is smaller than at sea level. The marginal variance increases as the value of $\theta_{\kappa,h}$ increases and as the value of $\theta_{\tau,h}$ decreases, and it is higher than at sea level.

As prior distributions for the parameters, we use informative Gaussian prior distributions:

$$\theta_{\tau,1} \sim \mathcal{N}(-1.8, 0.5), \quad \theta_{\tau,h} \sim \mathcal{N}(\mu_{\tau,h}, 0.5)$$
$$\theta_{\kappa,1} \sim \mathcal{N}(1.4, 0.5), \quad \theta_{\kappa,h} \sim \mathcal{N}(\mu_{\kappa,h}, 0.5).$$

with $\mu_{\tau,h}$ and $\mu_{\kappa,h}$ equal to the true values of $\theta_{\tau.h}$ and $\theta_{\kappa,h}$, respectively. The prior distributions of $\theta_{\tau,1}$ and $\theta_{\kappa,1}$ refer to both the stationary model and the non-stationary model.

We consider four different toy example designs. In all examples the domain is the same as described in Section 5.1, i.e., $\mathcal{D} : \{(s_1, s_2) : |s_1| \leq 1, |s_2| \leq 1\}$.

5.2.1 Design 1

In Design 1, we consider the same set up as described in Section 5.1, with the same mountain, the same locations of observations $\boldsymbol{s}_{obs,1}$ and $\boldsymbol{s}_{obs,2}$ and the same location of interpolation, $\boldsymbol{s}_{pred}$. We simulate precipitation in all three locations, and fit a stationary and a non-stationary model to the observations $Y_{obs,1}$ and $Y_{obs,2}$. The fitted models are used to interpolate precipitation at location $\boldsymbol{s}_{pred}$.

We consider two versions of the experimental design: Design 1a and Design 1b. In Design 1a, we consider the hyperparameters $\boldsymbol{\theta}^*$ as known when fitting the models. In this case, we only use 1 replicate.

In Design 1b, we estimate the hyperparameters when fitting the models. In this case, we use 30 replicates, in order to have enough observations to estimate the parameter values.

To evaluate the performance of the two models, we consider the coverage probability of a 95% prediction interval. Furthermore, we consider the estimated posterior mean of the interpolated precipitation Y_{pred}, based on the two models, and compare it with the true value of Y_{pred}. We do $S = 200$ simulations, and we consider the mean of the posterior mean from the simulations. As the mean is expected to be zero for both models and for the true value, we only consider the simulations where $Y_{obs,1} < Y_{obs,2}$. Y_{pred} is more correlated to $Y_{obs,1}$ than $Y_{obs,2}$, thus we expect the mean from the non-stationary model to be close to the true mean, whereas the stationary model, which weights $Y_{obs,1}$ and $Y_{obs,2}$ equally, is expected to have a too high posterior mean.

5.2.2 Design 2

In Design 2, we consider the same mountain as described in Section 5.1. In this example, we estimate the areal precipitation at the mountain. To do so, we interpolate precipitation at some locations on the top of the mountain. Figure 5.2 displays the mountain, including the locations of observations (black labelled dots) and the locations of interpolation (black dots).

We follow the procedure described in Section 4.5 when estimating the areal precipitation. We evaluate the models by considering the coverage probability of a 95% prediction interval and the mean CRPS, $\overline{\text{CRPS}}$, of $S = 100$ simulations. The hyperparameters $\boldsymbol{\theta}^*$ are estimated during the model fitting, and we use $J = 5$ replicates.

5.2.3 Design 3

In Design 3, we consider a different mountain from the previous designs. The mountain is now cone shaped, such that the elevation h is given by

$$h(r(s_1, s_2)) = \begin{cases} 0 & \text{if } r(s_1, s_2) > 0.5 \\ 1 - \frac{r(s_1,s_2)}{0.5} & \text{else} \end{cases}$$

for radius $r(s_1, s_2)$, at location $\boldsymbol{s} = (s_1, s_2)$, where $|s_1| \leq 1$ and $s_2 \leq 1$, as before.

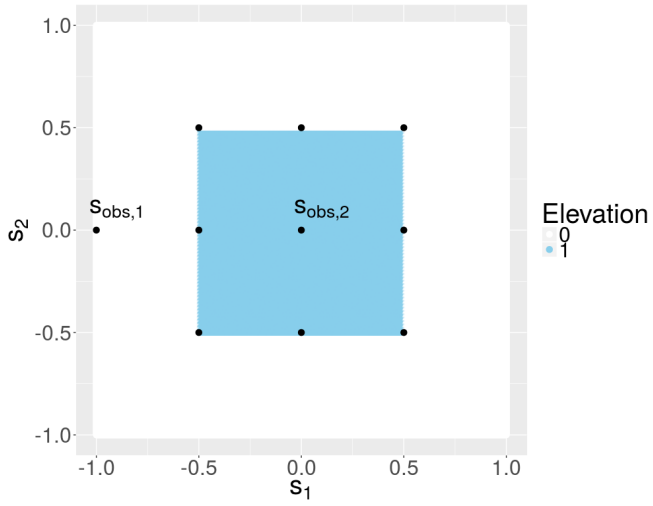


Figure 5.2: Illustration of the domain used in Design 2. The black, labelled dots represent locations of observations and the remaining black dots represents locations of interpolation.

Illustrations of the mountain can be seen in Figure 5.3.

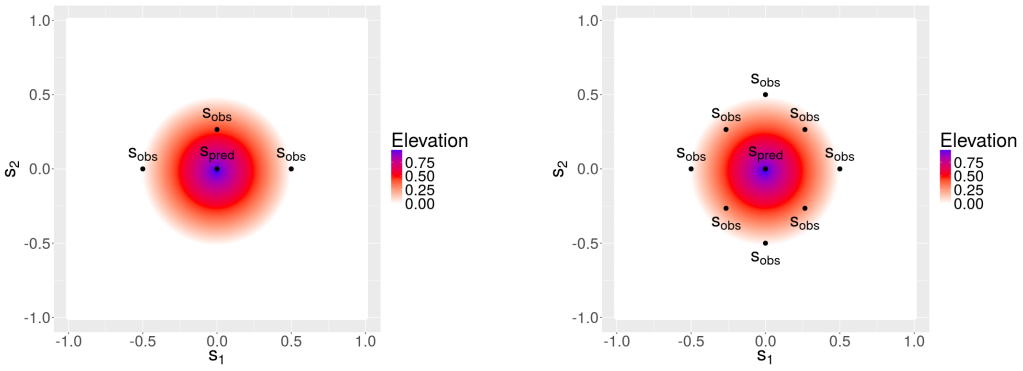
We consider two versions of the design, Design 3a and Design 3b. In both designs, the hyperparameters $\boldsymbol{\theta}^*$ are estimated during the model fitting, and we use $J = 5$ replicates.

In Design 3a, we assume there is three locations where we observe precipitation; two are located at the foot of the mountain, with elevation $h = 0$, such that the mountain is between the two locations. The third is located a bit higher, with elevation $h = 0.25$. Further, we assume there is a location where we interpolate precipitation, with coordinates $\boldsymbol{s}_{pred} = (0, 0)$, i.e. at the top of the mountain. Figure 5.3a displays Design 3a.

In Design 3b, we consider several locations of observations, all located below elevation $h = 0.25$. Further, we consider the same location where we interpolate precipitation as in Design 3a, i.e., at the top of the mountain. Figure 5.3b displays Design 3b.

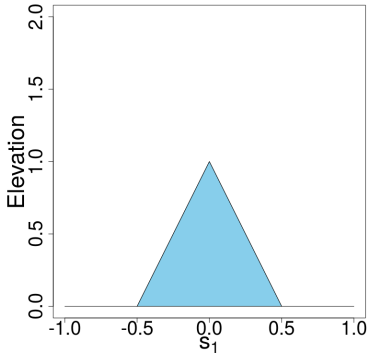
5.2.4 Design 4

In Design 4, we consider the same cone-shaped mountain as in Design 3, with the same locations of observations as in Design 3b. In this design, we estimate the areal precipitation at the mountain where the elevation $h > 0.25$. We estimate the areal precipitation following the procedure described in Section 4.5. In both



(a) Design 3a.

(b) Design 3b.



(c) The mountain viewed from the side.

Figure 5.3: Illustrations of a cone shaped mountain used in Design 3.

designs, the hyperparameters $\boldsymbol{\theta}^*$ are estimated during the model fitting, and we use $J = 5$ replicates.

We consider two versions of Design 4: Design 4a and Design 4b. In design 4a, the observations are the same as in Design 3b. In Design 4b, there is an additional observation with coordinates $\boldsymbol{s}_{obs} = (0, 0)$, i.e., at the top of the mountain, and we investigate how this affects the estimated areal precipitation from the two models.

5.3 Toy example results

5.3.1 Design 1

Figure 5.4 displays plots of the coverage probability of a 95% prediction interval and the estimated posterior mean from Design 1a (crosses) and 1b (circles). The left column displays the coverage probability, with the red marks representing the coverage probability of the non-stationary model and the blue marks representing the coverage probability of the stationary model. The black dashed line shows where 0.95 is, which is where the coverage of a good model should be.

The plots in the right column display the mean of the interpolated posterior means from the two models. Red marks represent the non-stationary model and blue marks represent the stationary model. The mean of the true, simulated precipitation, is included as green marks. The mean is based on simulations where $Y_{obs,1} < Y_{obs,2}$. The mean is plotted against the corresponding value of $\theta_{\tau,h}$, and the different plots are based on different values of $\theta_{\kappa,h}$. Notice the different scales of the vertical axis for the plots.

For Design 1a, where the hyperparameters $\boldsymbol{\theta}^*$ were assumed known, the plots demonstrate clearly that the non-stationary model has a good coverage all the time, and the coverage is close to the 0.95-line in all cases. The stationary model, has a poor coverage when the values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$ decrease, i.e., when the range gets smaller and the marginal variance gets higher on the top of the mountain.

The plots of the posterior mean illustrate what we expected; the posterior mean of the non-stationary model is quite close to the true value, whereas the stationary model differs a lot from the true value when $\theta_{\tau,h}$ and $\theta_{\tau,h}$ decrease.

For Design 1b, where the hyperparameters $\boldsymbol{\theta}^*$ where estimated during the model

fitting, the coverage probability of the non-stationary model is close to the 0.95-line for all values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$, as in Design 1a. This indicates that the model has recovered the values of the hyperparameter $\boldsymbol{\theta}^*$. The coverage probability of the stationary model, is notably better for Design 1b than it was for Design 1a, although it is still not as good as the non-stationary model. This is as expected, because now the values of $\theta_{\tau,1}$ and $\theta_{\tau,1}$ are estimated to fit somewhere in between the values of the true parameters at sea level and at the top of the mountain. However, when the underlying process becomes very non-stationary, the predictive performance of the stationary model is not good enough.

The posterior mean of the non-stationary model is still very close to the true value, as expected. The posterior mean of the stationary model is closer to the true value than in Design 1b. However, it is still not as good as the non-stationary model when the process becomes very non-stationary.

5.3.2 Design 2

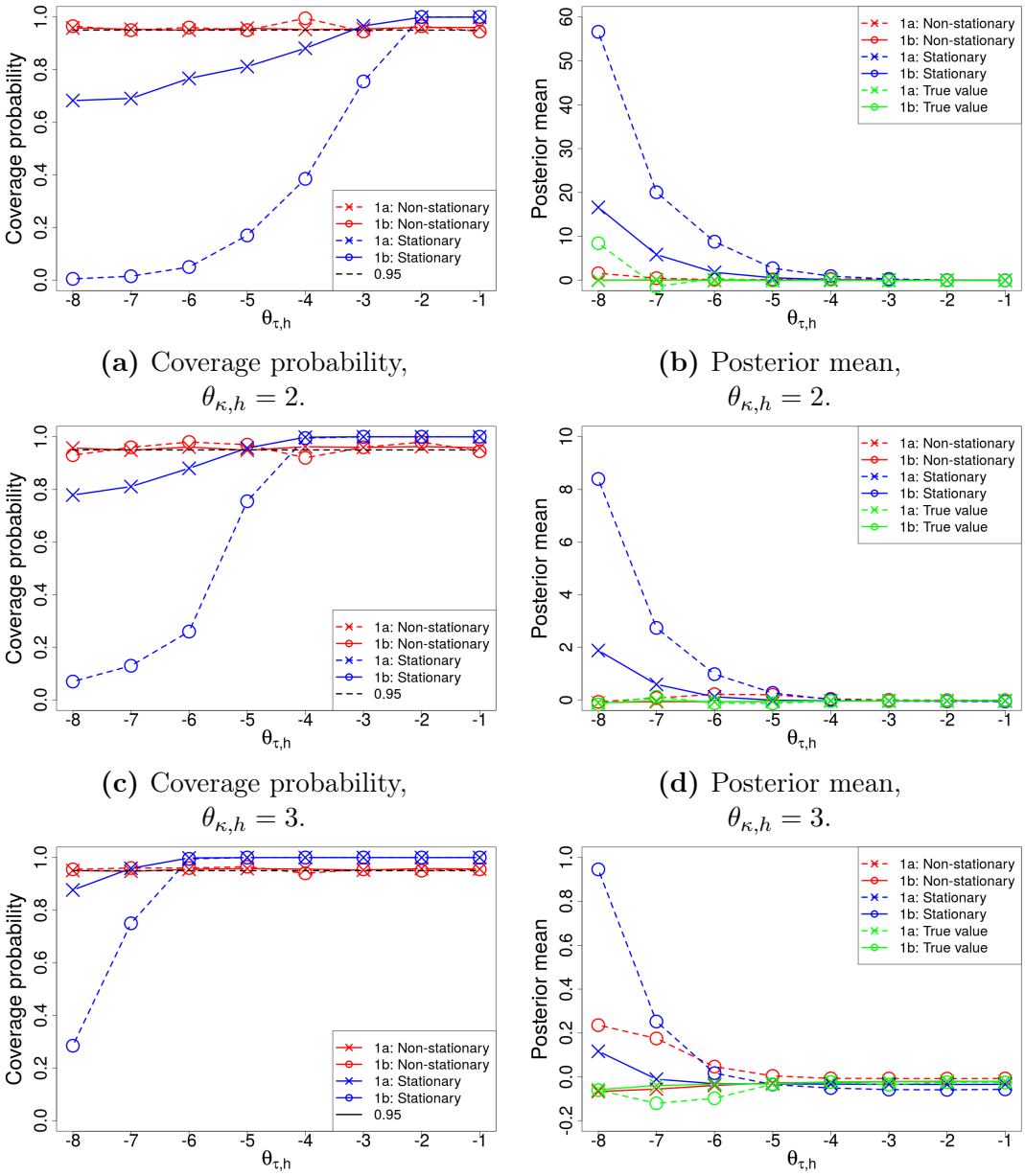
Figure 5.5 displays the coverage probability and mean CRPS, $\overline{\text{CRPS}}$, for the areal precipitation estimated by the two models. The coverage probability and the mean are over $S = 100$ simulations. In all plots, the red circles represent the non-stationary model and the blue circles represent the stationary model. In the plots of the coverage probability, the black dashed line represents 0.95, i.e., where the coverage probability should ideally be. Notice the different vertical axis in the plots of the $\overline{\text{CRPS}}$.

The coverage probability of the non-stationary model is good in all cases; it lies very close to the 0.95-line. The coverage probability of the stationary model is not as good. The coverage is too low when the field is very non-stationary, and it is to high when the field gets closer to stationary.

The $\overline{\text{CRPS}}$ of the non-stationary model is better than the $\overline{\text{CRPS}}$ of the stationary model when the process is very non-stationary. When the process gets closer to stationary, the $\overline{\text{CRPS}}$ of the two models is quite similar. For both the stationary and the non-stationary model, the $\overline{\text{CRPS}}$ increases when the non-stationarity increases.

5.3.3 Design 3

Figure 5.6 and Figure 5.7 display the coverage probability and mean CRPS, $\overline{\text{CRPS}}$, for Design 3a and 3b, respectively. The coverage probability and the



(a) Coverage probability, $\theta_{\kappa,h} = 2$.

(b) Posterior mean, $\theta_{\kappa,h} = 2$.

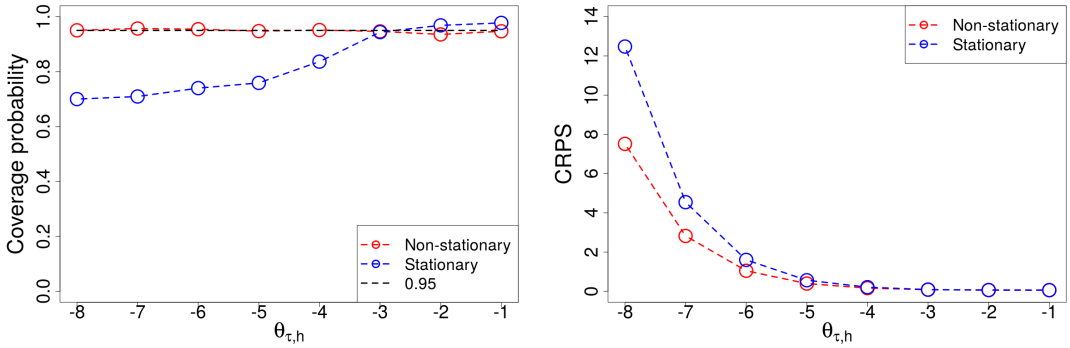
(c) Coverage probability, $\theta_{\kappa,h} = 3$.

(d) Posterior mean, $\theta_{\kappa,h} = 3$.

(e) Coverage probability, $\theta_{\kappa,h} = 4$.

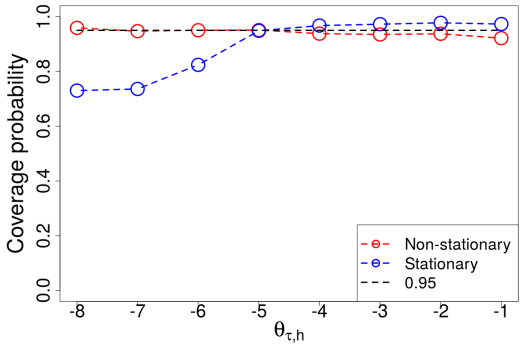
(f) Posterior mean, $\theta_{\kappa,h} = 4$.

Figure 5.4: Design 1a and 1b. Comparison of the predictive performances of the non-stationary (red) and stationary (blue) models in the example. The plots in the left column display the coverage probability of the interpolated precipitation for different values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$. The right columns displays the posterior mean of the interpolated precipitation, for $Y_{obs,1} < Y_{obs,2}$, estimated by the non-stationary and stationary models, along with the true precipitation, for different values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$.

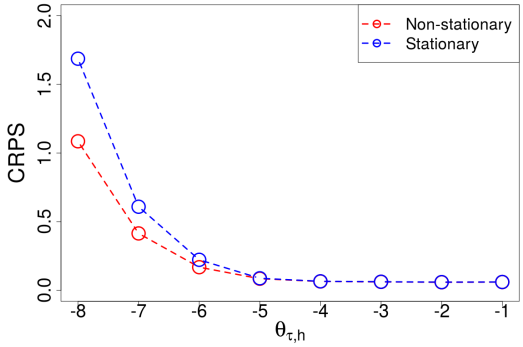


(a) Coverage probability, $\theta_{\kappa,h} = 2$.

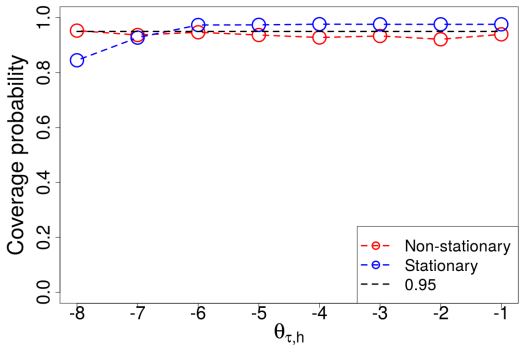
(b) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 2$.



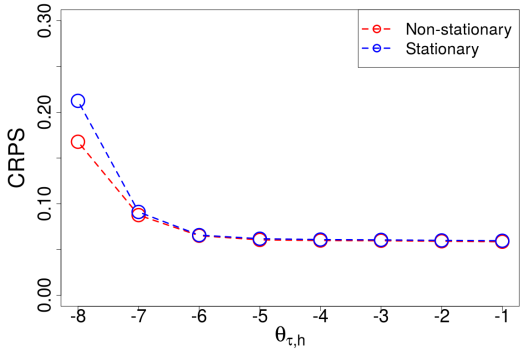
(c) Coverage probability, $\theta_{\kappa,h} = 3$.



(d) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 3$.



(e) Coverage probability, $\theta_{\kappa,h} = 4$.



(f) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 4$.

Figure 5.5: Design 2. Comparison of the predictive performances of the non-stationary (red) and stationary (blue) models from Design 2. The plots in the left column display the coverage probability of the estimated areal precipitation and the plots in the right column displays the mean CRPS, $\overline{\text{CRPS}}$, for different values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$. The coverage probability and the mean are over $S = 100$ simulations.

mean are over $S = 100$ simulations. In all plots, the red circles represent the non-stationary model and the blue circles represent the stationary model. Notice the different vertical axis in the plots of the $\overline{\text{CRPS}}$.

The plots of the coverage probability from Design 3a, show that the non-stationary model has good predictive performance in all cases. The stationary model has a poor coverage when the true process is very non-stationary.

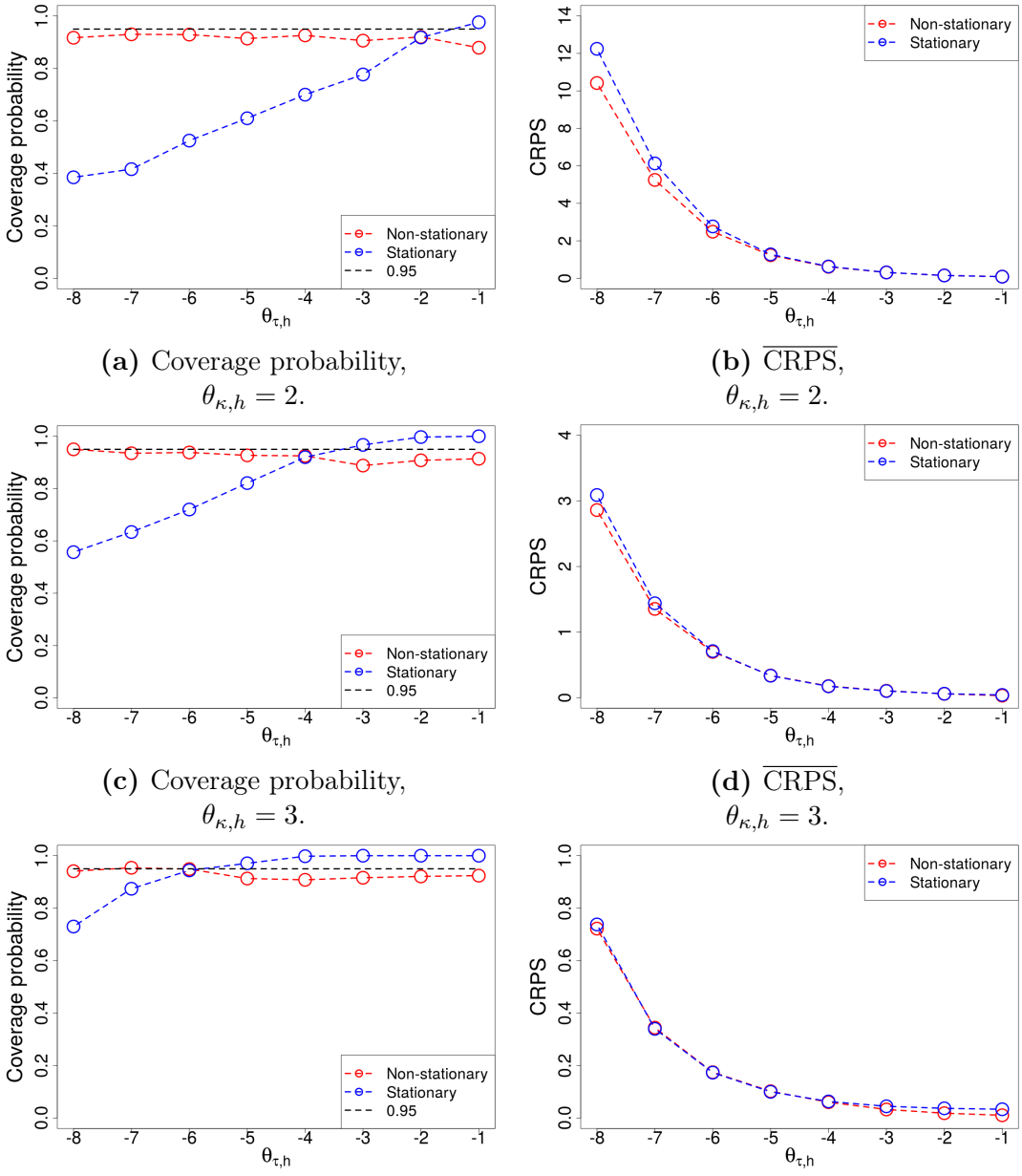
The $\overline{\text{CRPS}}$ of the two models is quite similar for Design 3a. Only when the process is extremely non-stationary, the non-stationary model has a better $\overline{\text{CRPS}}$.

In Design 3b, the difference between the two models is larger. The predictive performance of the non-stationary model is still good in all cases. The predictive performance of the stationary model is worse than in Design 3a. We recall that in Design 3b there are more observations than in Design 3a, and all observations are located below $h = 0.25$. The increase in observations leads the stationary model to be more certain about the predictions, and because it is not able to detect the effect of elevation on the range and the marginal variance, it estimates it to be too low, with too small uncertainty.

Figure 5.8 and Figure 5.9 display the estimated range and marginal variance by the stationary model (blue line) and non-stationary model (red line), along with the true range and marginal variance (green lines) in Design 3b. The estimates are based on the posterior mean of the hyperparameters $\boldsymbol{\theta}^*$ from one simulation. In Figure 5.8 the true values of the hyperparameters $\theta_{\kappa,h}$ and $\theta_{\tau,h}$ are $\theta_{\kappa,h} = 2$ and $\theta_{\tau,h} = -1, -4, -8$. In Figure 5.9 the true values of the hyperparameters $\theta_{\kappa,h}$ and $\theta_{\tau,h}$ are $\theta_{\kappa,h} = 4$ and $\theta_{\tau,h} = -1, -4, -8$.

The plots illustrate that the stationary model, which must estimate a constant range and marginal variance, estimates it to be too low when the process becomes very non-stationary, as expected.

Furthermore, we see that the non-stationary model estimates the range very well. It also estimates the marginal variance quite well. However, when $\theta_{\kappa,h} = 2$, the non-stationary model does not fully manage to recover the parameters. The marginal variance is estimated to be too low, especially at low elevations. For $\theta_{\kappa,h} = 4$, the non-stationary model estimates the marginal variance better. In this case the data is not as non-stationary as for $\theta_{\kappa,h} = 2$.



(a) Coverage probability, $\theta_{\kappa,h} = 2$.

(b) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 2$.

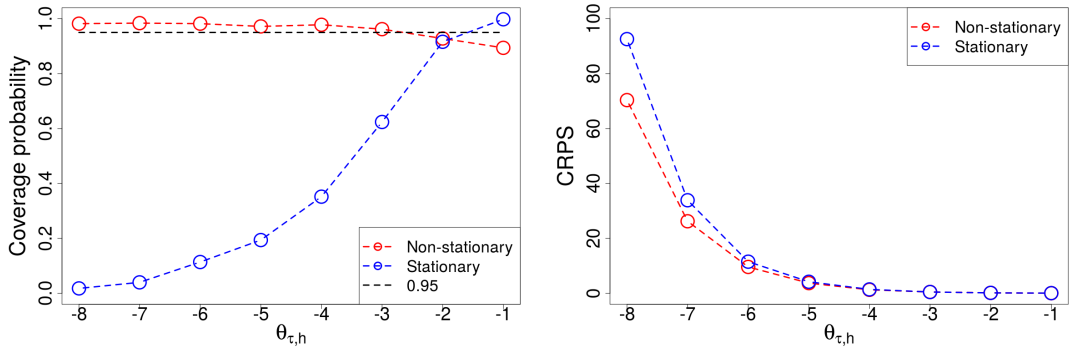
(c) Coverage probability, $\theta_{\kappa,h} = 3$.

(d) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 3$.

(e) Coverage probability, $\theta_{\kappa,h} = 4$.

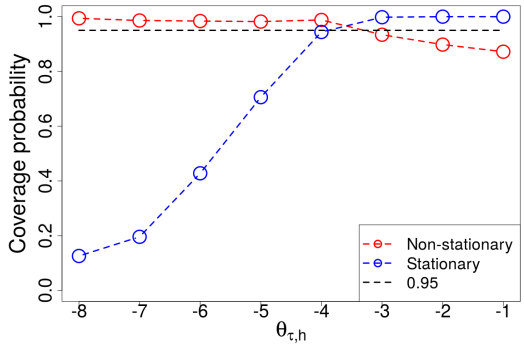
(f) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 4$.

Figure 5.6: Design 3a. Comparison of the predictive performances of the non-stationary (red) and stationary (blue) models from Design 3a. The plots in the left column display the coverage probability and the plots in the right column displays the mean CRPS, $\overline{\text{CRPS}}$, for different values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$. The coverage probability and the mean are over $S = 100$ simulations.

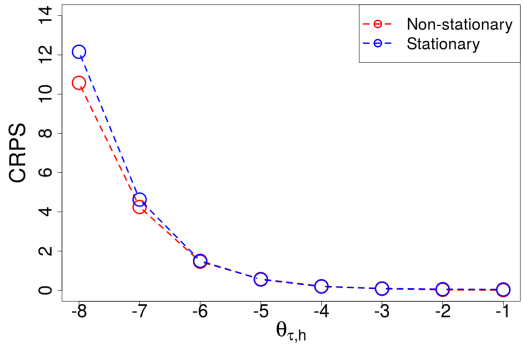


(a) Coverage probability, $\theta_{\kappa,h} = 2$.

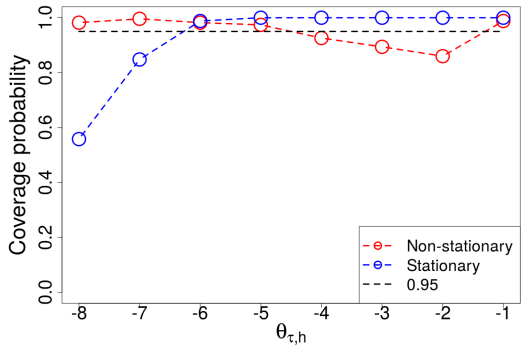
(b) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 2$.



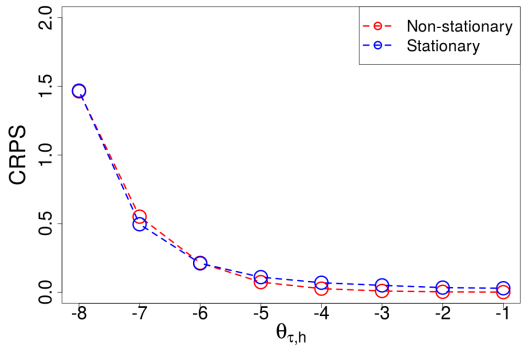
(c) Coverage probability, $\theta_{\kappa,h} = 3$.



(d) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 3$.

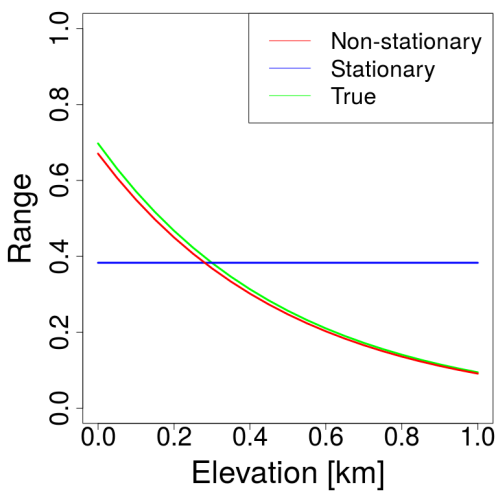


(e) Coverage probability, $\theta_{\kappa,h} = 4$.

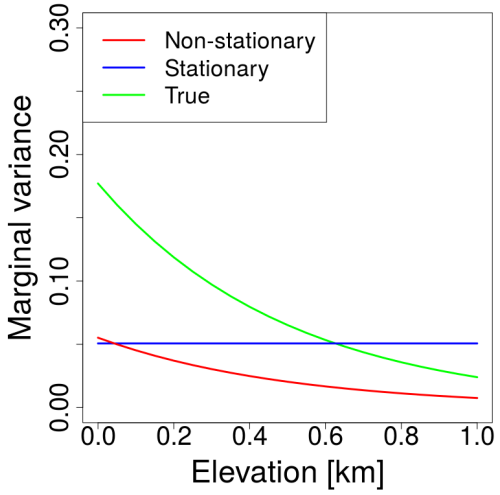


(f) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 4$.

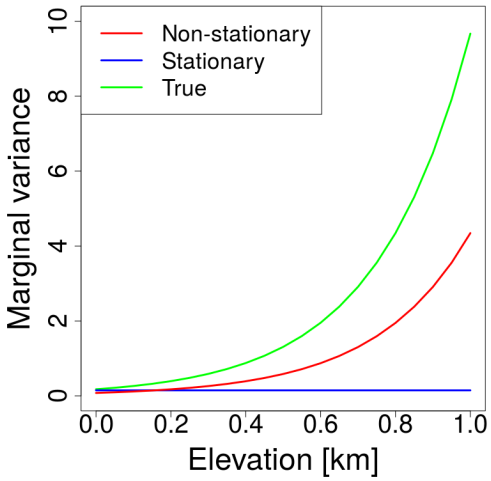
Figure 5.7: Design 3b. Comparison of the predictive performances of the non-stationary (red) and stationary (blue) models from Design 3b. The plots in the left column display the coverage probability and the plots in the right column displays the mean CRPS, $\overline{\text{CRPS}}$, for different values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$. The coverage probability and the mean are over $S = 100$ simulations.



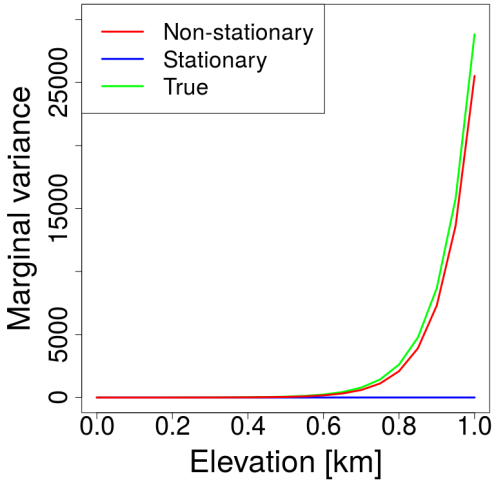
(a) Range, $\theta_{\kappa,h} = 2$.



(b) Marginal variance, $\theta_{\kappa,h} = 2$, $\theta_{\tau,h} = -1$.

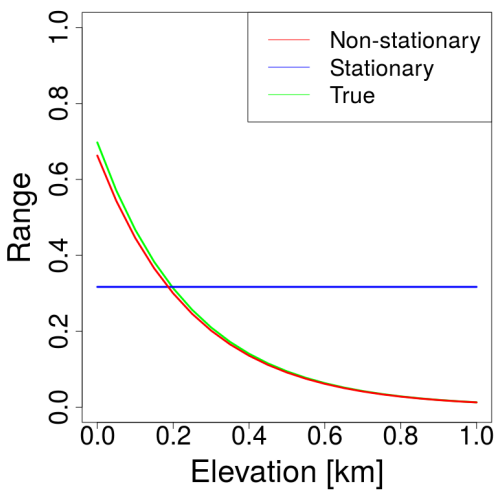


(c) Marginal variance, $\theta_{\kappa,h} = 2$, $\theta_{\tau,h} = -4$.

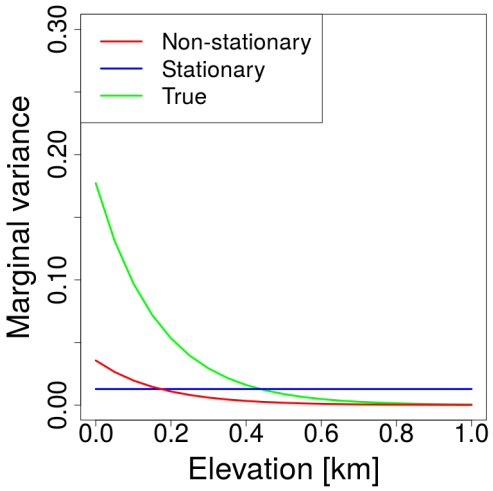


(d) Marginal variance, $\theta_{\kappa,h} = 2$, $\theta_{\tau,h} = -8$.

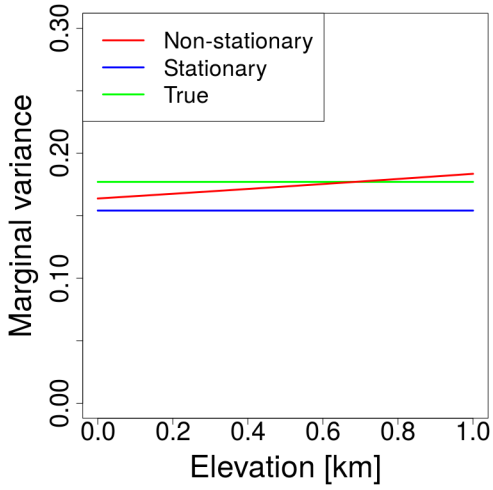
Figure 5.8: Estimated range and marginal variance, based on the posterior mean of the hyperparamters, estimated by the non-stationary model (red) and stationary model (blue). The true values of the hyperparameters $\theta_{\kappa,h}$ and $\theta_{\tau,h}$ are $\theta_{\kappa,h} = 2$ and $\theta_{\tau,h} = -1, -4, -8$. The estimation is done using Design 3b. The green lines represent the true range and marginal variance.



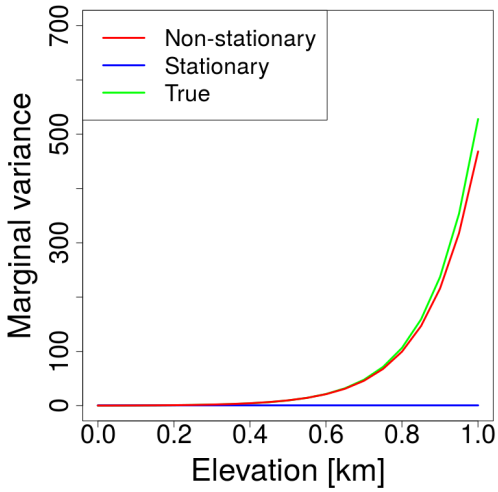
(a) Range, $\theta_{\kappa,h} = 4$.



(b) Marginal variance, $\theta_{\kappa,h} = 4$, $\theta_{\tau,h} = -1$.



(c) Marginal variance, $\theta_{\kappa,h} = 4$, $\theta_{\tau,h} = -4$.



(d) Marginal variance, $\theta_{\kappa,h} = 4$, $\theta_{\tau,h} = -8$.

Figure 5.9: Estimated range and marginal variance, based on the posterior mean of the hyperparamters, estimated by the non-stationary model (red) and stationary model (blue). The true values of the hyperparameters $\theta_{\kappa,h}$ and $\theta_{\tau,h}$ are $\theta_{\kappa,h} = 4$ and $\theta_{\tau,h} = -1, -4, -8$. The estimation is done using Design 3b. The green lines represent the true range and marginal variance.

5.3.4 Design 4

Figure 5.10 and Figure 5.11 display the coverage probability and mean CRPS, $\overline{\text{CRPS}}$, from estimation of areal precipitation using Design 4a and 4b, respectively. The coverage probability and the mean are over $S = 100$ simulations. In all plots, the red circles represent the non-stationary model and the blue circles represent the stationary model. Notice the different vertical axis in the plots of the $\overline{\text{CRPS}}$.

We see that in Design 4a, where all observations are located below $h = 0.25$, the non-stationary model has a good coverage probability in all cases. The coverage probability of the stationary model is poor. When the process is very non-stationary, it goes as low as approximately 0.1.

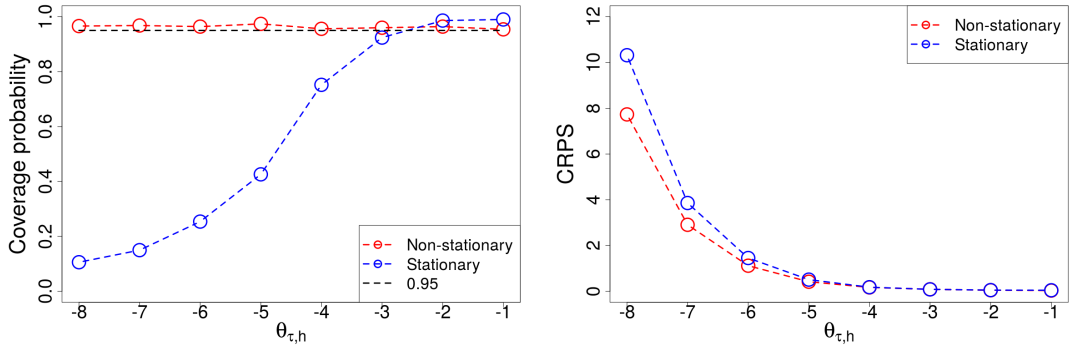
The $\overline{\text{CRPS}}$ of the models are more similar, it is only slightly better for the non-stationary model. For both model, the $\overline{\text{CRPS}}$ gets worse as the non-stationarity of the data increases. This indicates that the non-stationary model does not fully recover the hyperparameters.

There is a large difference in the results from Design 4b, compared to the results from Design 4a. Recall that in Design 4b there is an observation at the top of the mountain. The most notable difference is in the coverage probability of the stationary model. It is much higher with this design, lying around the 0.95-line for many of the cases. The coverage probability of the non-stationary model is good in all cases, as before.

The $\overline{\text{CRPS}}$ of both models is much better in Design 4b than in Design 4a (notice the different axis). It is a larger difference between the $\overline{\text{CRPS}}$ of the non-stationary model and the stationary model in Design 4b, it is notably better for the non-stationary model. Furthermore, for the non-stationary model, there is only a small increase in the $\overline{\text{CRPS}}$ as the non-stationarity of the data increases, indicating the model recovers the parameters.

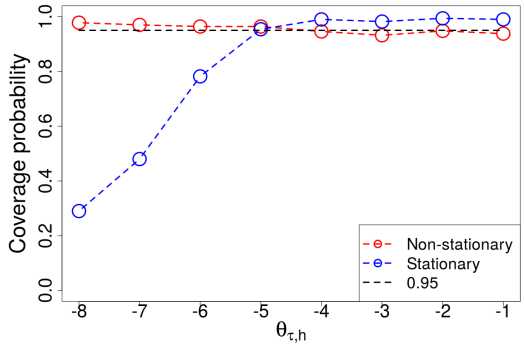
Figure 5.12 and Figure 5.13 display the estimated range and marginal variance by the stationary model (blue line) and non-stationary model (red line), along with the true range and marginal variance (green lines) in Design 4b. The estimates are based on the posterior mean of the hyperparameters $\boldsymbol{\theta}^*$ from one simulation. In Figure 5.12 the true values of the hyperparameters $\theta_{\kappa,h}$ and $\theta_{\tau,h}$ are $\theta_{\kappa,h} = 2$ and $\theta_{\tau,h} = -1, -4, -8$. In Figure 5.13 the true values of the hyperparameters $\theta_{\kappa,h}$ and $\theta_{\tau,h}$ are $\theta_{\kappa,h} = 4$ and $\theta_{\tau,h} = -1, -4, -8$.

The plots illustrate that the non-stationary model estimates the range and marginal variance very well. The estimates of the marginal variance are notably better

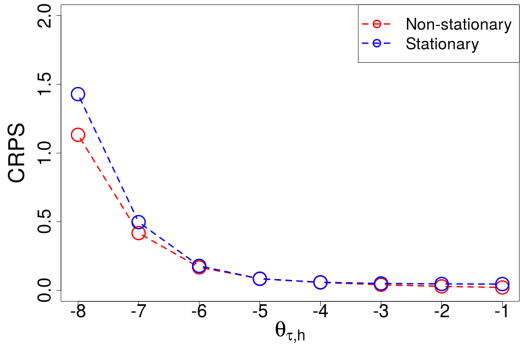


(a) Coverage probability, $\theta_{\kappa,h} = 2$.

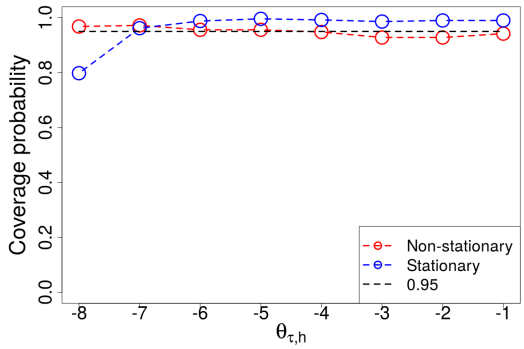
(b) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 2$.



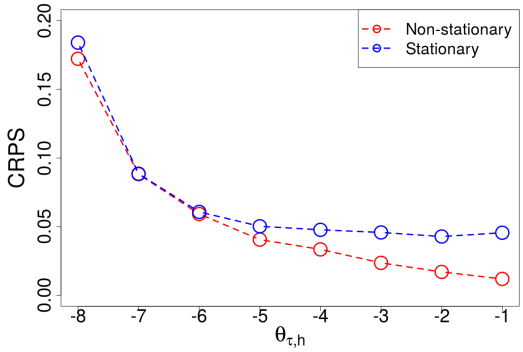
(c) Coverage probability, $\theta_{\kappa,h} = 3$.



(d) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 3$.

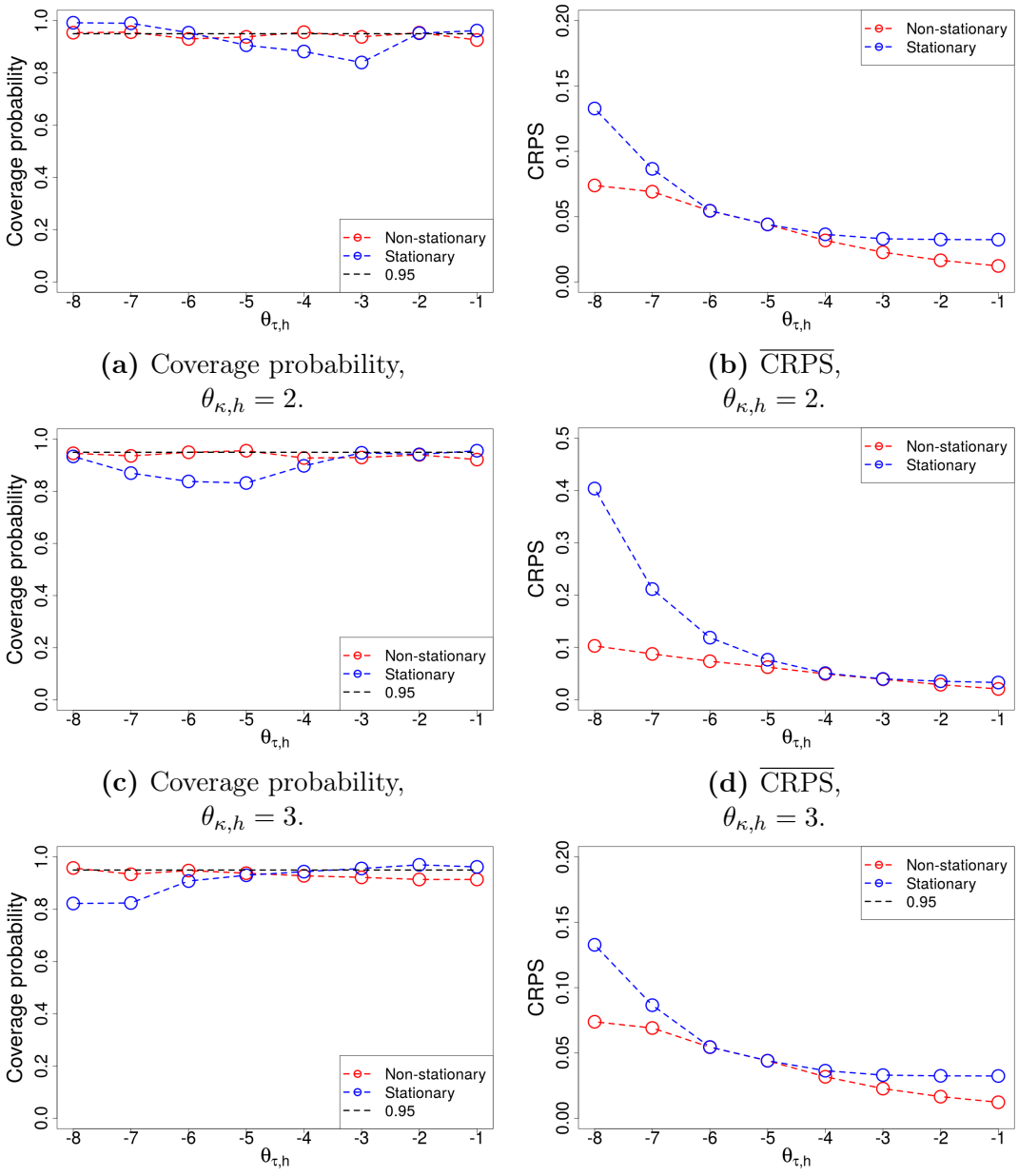


(e) Coverage probability, $\theta_{\kappa,h} = 4$.



(f) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 4$.

Figure 5.10: Design 4a. Comparison of the predictive performances of the non-stationary (red) and stationary (blue) models from Design 4a. The plots in the left column display the coverage probability and the plots in the right column displays the mean CRPS, $\overline{\text{CRPS}}$, for different values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$. The coverage probability and the mean are over $S = 100$ simulations.



(a) Coverage probability, $\theta_{\kappa,h} = 2$.

(b) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 2$.

(c) Coverage probability, $\theta_{\kappa,h} = 3$.

(d) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 3$.

(e) Coverage probability, $\theta_{\kappa,h} = 4$.

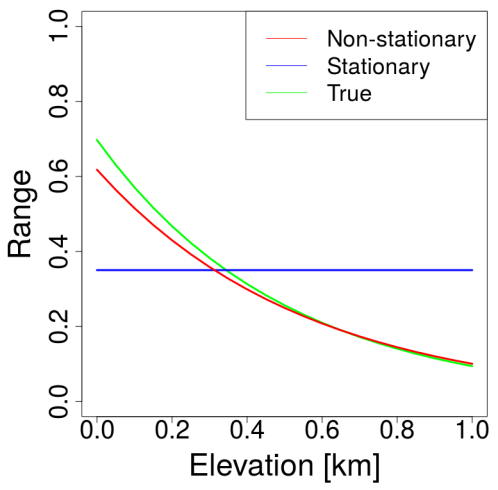
(f) $\overline{\text{CRPS}}$, $\theta_{\kappa,h} = 4$.

Figure 5.11: Design 4b. Comparison of the predictive performances of the non-stationary (red) and stationary (blue) models from Design 4b. The plots in the left column display the coverage probability and the plots in the right column displays the mean CRPS, $\overline{\text{CRPS}}$, for different values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$. The coverage probability and the mean are over $S = 100$ simulations.

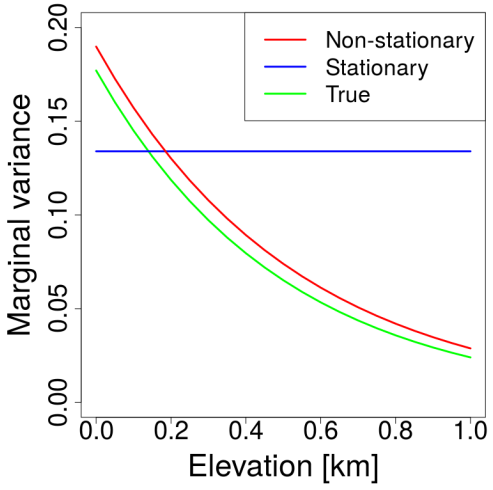
than in Design 3b, which had the same observations, except the one at the top of the mountain. The stationary model estimates the range and the variance quite similar to the estimates in Design 3b.

5.3.5 Summary

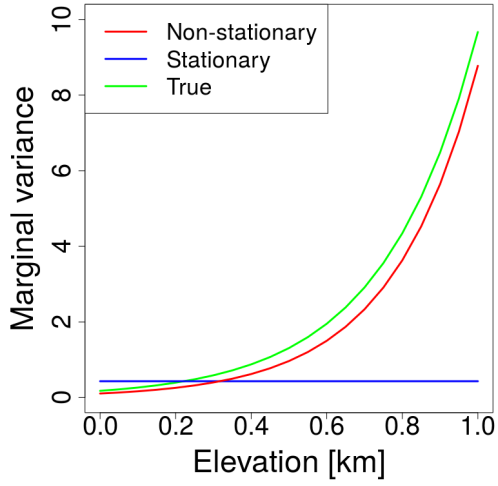
These examples illustrate the effect of increasing the non-stationarity of a process. When the explanatory variables of the dependency structure both are close to zero, both a stationary and non-stationary model have good predictive performance. However, as the explanatory variables get more extreme, and the range and marginal variance differs a lot at different altitudes, a stationary model has a poor predictive performance, and a non-stationary model is necessary. In particular, this is the case when the locations of interpolation have a higher elevation than the observations. In this case, the predictive performance of a stationary model is reduced when the number of observations increases. Furthermore, the examples illustrate the effect of including an observation at a high elevation inside the catchment area. Both for the stationary model and the non-stationary model, this leads to a large improvement in their predictive performance.



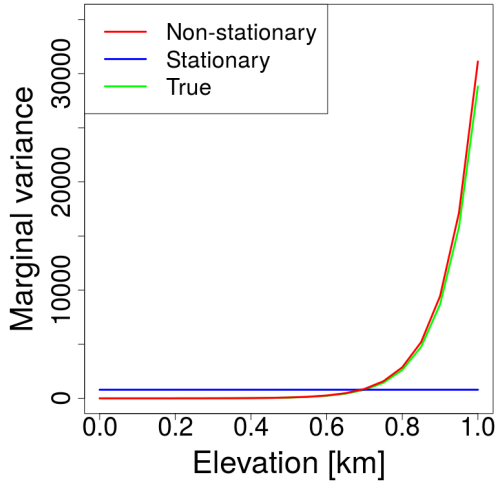
(a) Range, $\theta_{\kappa,h} = 2$.



(b) Marginal variance, $\theta_{\kappa,h} = 2$, $\theta_{\tau,h} = -1$.

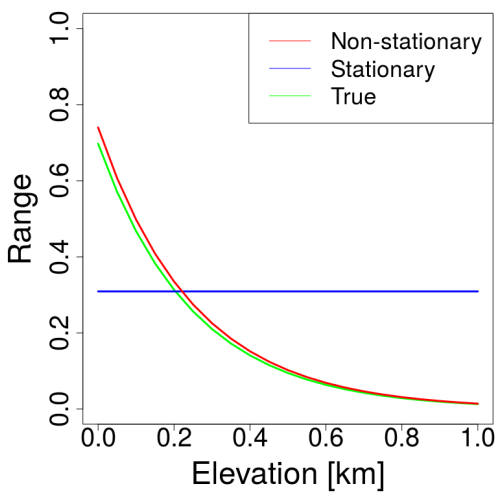


(c) Marginal variance, $\theta_{\kappa,h} = 2$, $\theta_{\tau,h} = -4$.

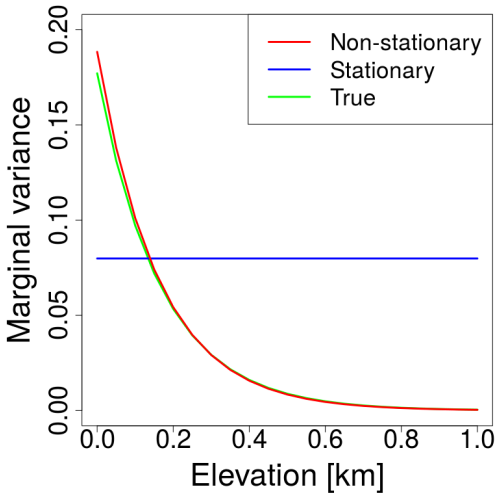


(d) Marginal variance, $\theta_{\kappa,h} = 2$, $\theta_{\tau,h} = -8$.

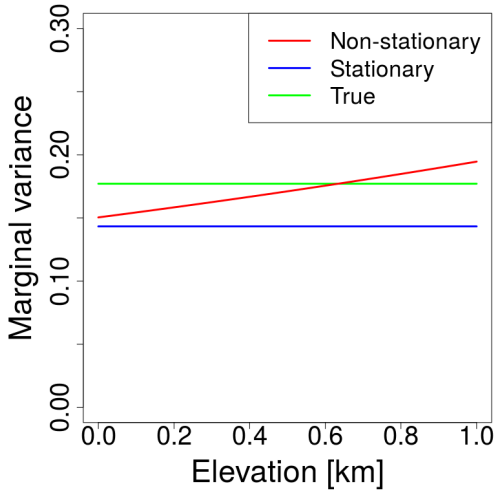
Figure 5.12: Estimated range and marginal variance, based on the posterior mean of the hyperparamters, estimated by the non-stationary model (red) and stationary model (blue). The true values of the hyperparameters $\theta_{\kappa,h}$ and $\theta_{\tau,h}$ are $\theta_{\kappa,h} = 2$ and $\theta_{\tau,h} = -1, -4, -8$. The estimation is done using Design 4b. The green lines represent the true range and marginal variance.



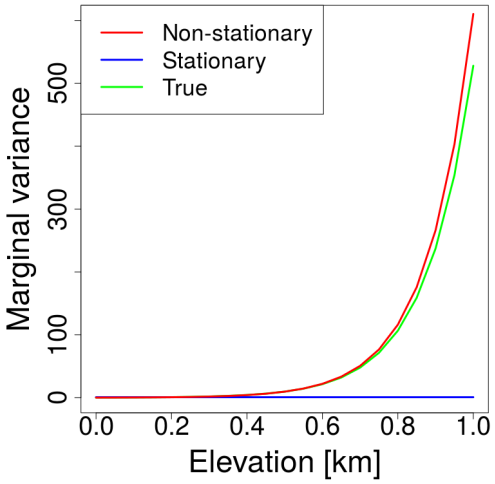
(a) Range, $\theta_{\kappa,h} = 4$.



(b) Marginal variance, $\theta_{\kappa,h} = 4$, $\theta_{\tau,h} = -1$.



(c) Marginal variance, $\theta_{\kappa,h} = 4$, $\theta_{\tau,h} = -4$.



(d) Marginal variance, $\theta_{\kappa,h} = 4$, $\theta_{\tau,h} = -8$.

Figure 5.13: Estimated range and marginal variance, based on the posterior mean of the hyperparamters, estimated by the non-stationary model (red) and stationary model (blue). The true values of the hyperparameters $\theta_{\kappa,h}$ and $\theta_{\tau,h}$ are $\theta_{\kappa,h} = 4$ and $\theta_{\tau,h} = -1, -4, -8$. The estimation is done using Design 4b. The green lines represent the true range and marginal variance.

Chapter 6

Case study

In this chapter, we fit the models (4.2) and (4.3) to the annual observations introduces in Chapter 2. We perform leave-one-out cross-validation, in order to evaluate the predictive performance of the models. Furthermore, we consider three catchment areas, where we estimate the areal precipitation, and compare the estimates to the observations introduced in Chapter 2.

6.1 Estimation and evaluation

We fit the models (4.2) and (4.3) from Chapter 4 to observations of precipitation from 35 years at the 60 weather stations. The total number of observations is $N = 984$. The prior distributions of the parameters were given in Section 4.3.

In order to evaluate and compare the models' ability to interpolate to points, we perform leave-one-out cross-validation, as described in Section 3.6.5. In each iteration, we leave out one weather station, and fit the models to the observations of annual precipitation at the remaining weather stations. The fitted models are used to interpolate precipitation at the left out weather station. As evaluation schemes of the predictive performance of the models, we compare CRPS and coverage probability of a 95% prediction interval. These schemes were introduced in Section 3.6. We consider both the mean CRPS at each weather station i, taken over all years, $\overline{\text{CRPS}}_i$, and the mean CRPS taken over all years and all locations, $\overline{\text{CRPS}}$. Correspondingly, we consider the coverage probability at the same location over all years and the coverage probability over all years and all locations. Further, in order to investigate if the CRPS of the models are significantly different, we do a paired samples t-test to the differences in the CRPS from the two

models. The paired samples t-test was introduced in Section 3.6. The differences are computed as $d_{ij} = \text{CRPS}_{ij,N-S} - \text{CRPS}_{ij,S}$, for year $j = 1, ..., J$ and location $i = 1, ..., n$.

In order to compare the predictive performance of the models when doing interpolation to area, we estimate the areal precipitation of the three catchment areas presented in Chapter 2. We follow the procedure introduced in Section 4.5. When we estimate the variance of the areal precipitation, we leave out the measurement error from observed runoff and evaporation. This is because these errors are very small compared to the errors of the estimated areal precipitation (Roksvåg, 2016).

6.2 Analysis of results

In this section, we present the results from the case study. We start out with presenting the estimated posterior distributions of the hyperparameters $\boldsymbol{\theta}$, followed by the results from the cross-validation and the results from estimation of areal precipitation.

Table 6.1 presents the posterior mean and a 95% credible interval for the parameters in the models, based on fitting the models to observations of annual precipitation at all weather stations.

We see that the posterior mean and the 95% credible interval of the precision τ_ϵ of the measurement error is the same for the stationary and the non-stationary model.

We consider the posterior distributions of the hyperparameters of the annual spatial variation, which we recall have subscript w. From the table, we see that the posterior mean of the stationary parameters $\theta_{\tau,1,S}$ and $\theta_{\kappa,1,S}$ are quite close to the non-stationary parameters, $\theta_{\tau,1,N-S}$ and $\theta_{\kappa,1,N-S}$. Also, the 95% credible intervals are quite similar in the stationary and non-stationary case. Further, the parameters of the explanatory variable of the dependency structure in the non-stationary case, $\theta_{\tau,h,N-S}$ and $\theta_{\kappa,h,N-S}$, have posterior means 0.51 and 0.71, respectively. We notice that their credible intervals are wide, both on the positive and negative side. All these results indicate that the annual spatial variation field is close to stationary.

In Figure 6.1a and 6.1b, the estimated marginal variance and range for the spatial annual variation, based on the posterior mean of the hyperparameters, are plotted against elevation [km]. The red line represents the non-stationary model and the

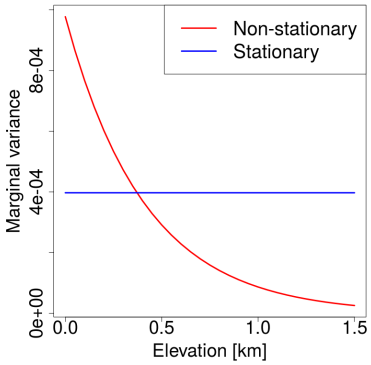
blue line represents the stationary model. We see that the marginal variance is very close to zero in both the stationary and the non-stationary case. The range of the non-stationary field decreases as the elevation increases. At sea-level, it is 15.03 km. The range is the same for the two models at elevation $h = 190$ m.a.s.

Further, we consider the posterior distributions of the hyperparameters of the climatology, which we recall have subscript u. We notice that in this case there is a larger difference between the posterior means and credible intervals of the stationary parameters $\theta_{\tau,1,S}$ and $\theta_{\kappa,1,S}$ and the non-stationary parameters, $\theta_{\tau,1,N-S}$ and $\theta_{\kappa,1,N-S}$. Furthermore, the posterior mean of $\theta_{\tau,h,N-S}$ is -1.36, and most of the credible interval is below zero. The posterior mean of $\theta_{\kappa,h,N-S}$ is 0.88, and most of the credible interval is above zero. These results indicate that there is some non-stationarity in the climatology, and this field is more non-stationary than the annual spatial variation.

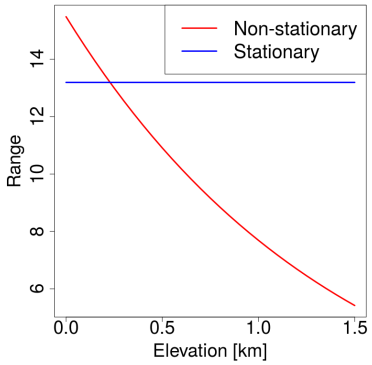
Figure 6.1c and 6.1d display the estimated marginal variance and range for the climatology, based on the posterior mean of the hyperparameters, plotted against elevation [km]. The plots show that for the non-stationary model, the marginal variance increases as the elevation increases. It is 0.19 at sea-level and it is the same for the two models at elevation $h = 500$ m.a.s. The range decreases as the elevation increases. It is 25.5 km at sea level and it is the same for the two models at elevation $h = 720$ m.a.s. For both the marginal variance and the range, there is a quite large difference between sea-level and elevation $h = 1500$ m.a.s. This support the indication of non-stationarity in the climatology.

Table 6.1: Posterior mean and a 95% credible interval of the hyperparameters, estimated by fitting the stationary (S) and non-stationary (N-S) model to all observations of precipitation in Hordaland.

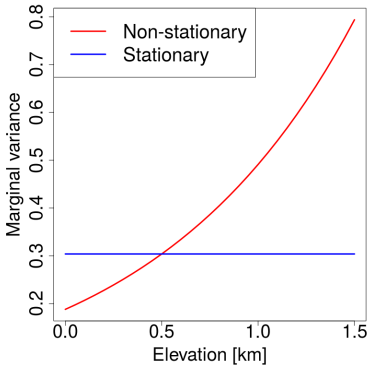
	Posterior mean		95% credible interval	
Parameter	S	N-S	S	N-S
τ	2.72	2.72	(2.48, 2.97)	(2.48,2.97)
$\theta_{\tau,1,w}$	4.19	3.90	(2.20, 6.19)	(1.86,6.02)
$\theta_{\tau,h,w}$		0.51		(-2.61,3.60)
$\theta_{\kappa,1,w}$	-1.54	-1.67	(-3.06,0.12)	(-3.15,-0.03)
$\theta_{\kappa,h,w}$		0.70		(-2.21,3.61)
$\theta_{\tau,1,u}$	0.90	1.77	(0.07,1.58)	(0.65,2.79)
$\theta_{\tau,h,u}$		-1.36		(-2.81,0.19)
$\theta_{\kappa,1,u}$	-1.57	-2.20	(-2.23,-0.76)	(-3.26,-1.03)
$\theta_{\kappa,h,u}$		0.88		(-0.59,2.24)



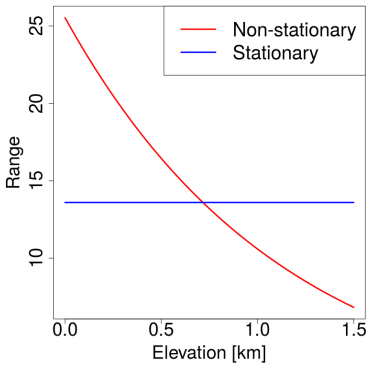
(a) Marginal variance for annual spatial variation.



(b) Range for annual spatial variation.



(c) Marginal variance for climatology.



(d) Range for climatology.

Figure 6.1: Estimated marginal variance and range using a stationary (blue) and non-stationary (red) model. The upper plots correspond to the annual spatial variation field and the lower plots correspond to the climatology field.

6.2.1 Results from interpolation to points

Table 6.2 displays a comparison of the predictive performance of the stationary model (S) and non-stationary model (N-S), based on leave-one-out cross-validaton. The table displays the coverage probability of a 95% prediction interval, mean CPRS ($\overline{\text{CRPS}}$), and mean RMSE ($\overline{\text{RMSE}}$). The coverage probability and the means are over all years and all locations.

We see that the coverage probability is quite good for both models, but it is slightly better for the non-stationary models than for the stationary model. For the non-stationary model it is 0.95, which is what it ideally should be. The $\overline{\text{CRPS}}$ and the $\overline{\text{RMSE}}$ are almost the same for both models, indicating that the predictive performances of the models are very similar.

A paired samples t-test is performed on the difference in the CRPS of the models. The t-value is $t = -5.64$. Because we have as much as $N = 984$ observations, the

t-statistic approximately follows a normal distribution. Thus, we compare it with the critical values for the normal distribution at a 95% confidence level, which are $\pm z_{0.95} = \pm 1.96$. Hence, we have that $t = -5.64 < -z_{0.95}$, and we reject the null hypothesis H_0, that the difference between the CRPS of the two models is zero. Further, the t-value is negative, implying the mean of the differences is negative. This indicates that the CRPS of the non-stationary model is smaller, i.e. better, than the CRPS of the stationary model.

Table 6.2: A comparison of the predictive performance of the stationary (S) and non-stationary (N-S) model. The results are based on leave-one-out cross-validation. The comparison is of the coverage probability of a 95% prediction interval, mean CPRS ($\overline{\text{CRPS}}$), and mean RMSE ($\overline{\text{RMSE}}$). The coverage probability and the means are over all years and all locations.

	Coverage probability	$\overline{\text{CRPS}}$	$\overline{\text{RMSE}}$
S	0.93	0.475	0.712
N-S	0.95	0.468	0.705

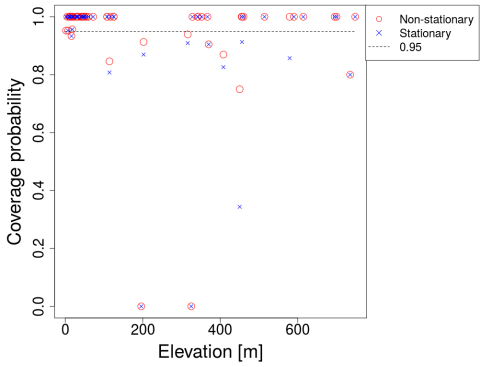
Figure 6.2 displays the coverage probability of a 95% prediction interval and difference in mean CRPS between the non-stationary and stationary model, $\overline{\text{CRPS}}_{i,diff}$, plotted against the elevation of the weather stations. The coverage probability and the mean of the CRPS is over all years at each location. The differences are calculated as $\overline{\text{CRPS}}_{i,\text{diff}} = \overline{\text{CRPS}}_{i,NS} - \overline{\text{CRPS}}_{i,S}$.

Figure 6.2a shows that the coverage probability is quite similar for the two model, but there are some locations where the non-stationary model is better. Figure 6.2b shows that the $\overline{\text{CRPS}}_i$ is quite similar for the two models at low elevations, but as the elevation gets high (higher than 400 m), the $\overline{\text{CRPS}}_i$ of the non-stationary model is better.

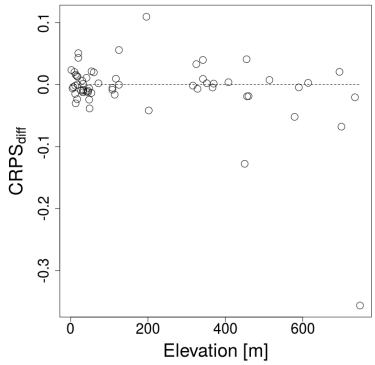
The results from the cross-validation indicate that the predictive performances of the models are quite similar. However, the non-stationary is slightly better as the elevation increases.

6.2.2 Results from interpolation to area

Figure 6.3, 6.4 and 6.5 display the posterior predictions of annual areal precipitation at the three catchment areas Fjellanger, Svartavatn and Slondalsvatn, respectively. The catchments were presented in Chapter 2. The blue lines represent the stationary model and the red lines represent the non-stationary model. The dashed lines correspond to 95% prediction intervals for the stationary model (blue) and the non-stationary model (red). The plots also display the observed annual areal precipitation (green line). The areal precipitation is given in 10^8m^3. Notice that the vertical axis is differently scaled in the three cases.



(a) Coverage probability of a 95% prediction interval, plotted against elevation [m].



(b) Difference in $\overline{\text{CRPS}}_i$ between non-stationary and stationary model, plotted against elevation [m].

Figure 6.2: Coverage probability of a 95% prediction interval and difference in mean CRPS, $\overline{\text{CRPS}}_{i,diff}$, between the non-stationary and stationary model, plotted against elevation [m]. The coverage probability is over all years at each location. The mean of the CRPS is taken over all years at each location, and the difference is calculated as $\overline{\text{CRPS}}_{i,\text{diff}} = \overline{\text{CRPS}}_{i,NS} - \overline{\text{CRPS}}_{i,S}$. The results are based on leave-one-out cross-validation.

The plots show that at all catchment areas, the estimated mean annual areal precipitation is very similar for the two models. The prediction intervals of the models differs; they are wider for the non-stationary model. At Fjellanger and Svartavatn the prediction intervals of neither of the models cover the observed areal precipitation very well. At Slondalsvatn the prediction intervals of the two models are much wider, and they cover all the the observed areal precipitation.

Table 6.3 displays the coverage probability of a 95% prediction interval, mean CRPS, $\overline{\text{CRPS}}$, and mean RMSE, $\overline{\text{RMSE}}$, of the areal precipitation at the three catchment areas, predicted by the two models. The coverage probability and the mean of the CRPS and RMSE are over all years. The table shows that there is very little difference in the predictive performances of the two models. At Fjellanger, the results are identical. At Svartavatn the coverage probability is the same for the two models, whereas the $\overline{\text{CRPS}}$ and $\overline{\text{RMSE}}$ are slightly better for the non-stationary model. At Slondalsvatn, the coverage probability is the same for the two models, the $\overline{\text{CRPS}}$ is slightly better for the stationary model and the $\overline{\text{RMSE}}$ is slightly better for the non-stationary model. Furthermore, vi notice that the coverage probabilities at Fjellanger and Svartavatn are low. In particular, at Svartavatn it is only 0.47. At Slondalsvatn on the other hand, the coverage is 1, which is higher than desired. Considering the coverage probability and the plot in Figure 6.5, it seems the prediction intervals at Slondalsvatn are too wide.

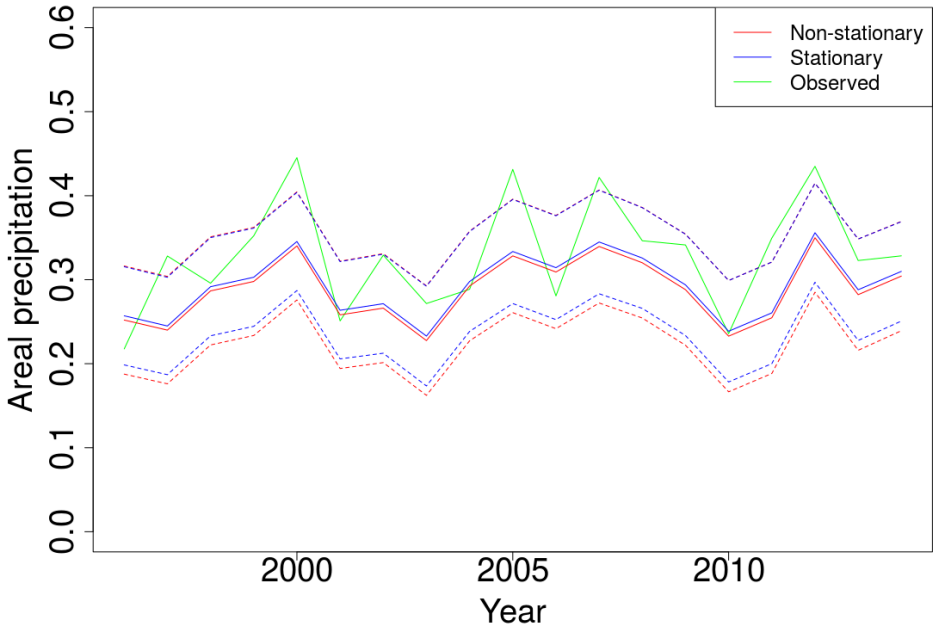


Figure 6.3: Estimated areal precipitation at Fjellanger using stationary model (blue line) and non-stationary model (red line), along with the observed areal precipitation (green line). The dashed lines represent 95% prediction interval for the stationary (blue) and non-stationary (red) model. The areal precipitation is given in $10^8 m^3$.

One reason for the noticeable large confidence intervals at Slondalsvatn, might be that this catchment area is located at higher elevation than the two others (see Table 2.1 in Section 2). This implies large variance for the precipitation here.

Table 6.3: Comparison of the predictive performance of the stationary (S) and non-stationary (N-S) model at the catchment areas. The comparison is of the coverage probability of a 95% prediction interval, mean CPRS ($\overline{\text{CRPS}}$) and mean RMSE ($\overline{\text{RMSE}}$). The predictions are of areal precipitation, and are compared with observed areal precipitation.

	Coverage probability		$\overline{\text{CRPS}}$		$\overline{\text{RMSE}}$	
	S	N-S	S	N-S	S	N-S
Fjellanger	0.68	0.68	0.04	0.04	0.06	0.06
Svartavatn	0.47	0.47	0.35	0.34	0.53	0.51
Slondalsvatn	1.00	1.00	0.08	0.10	0.14	0.12

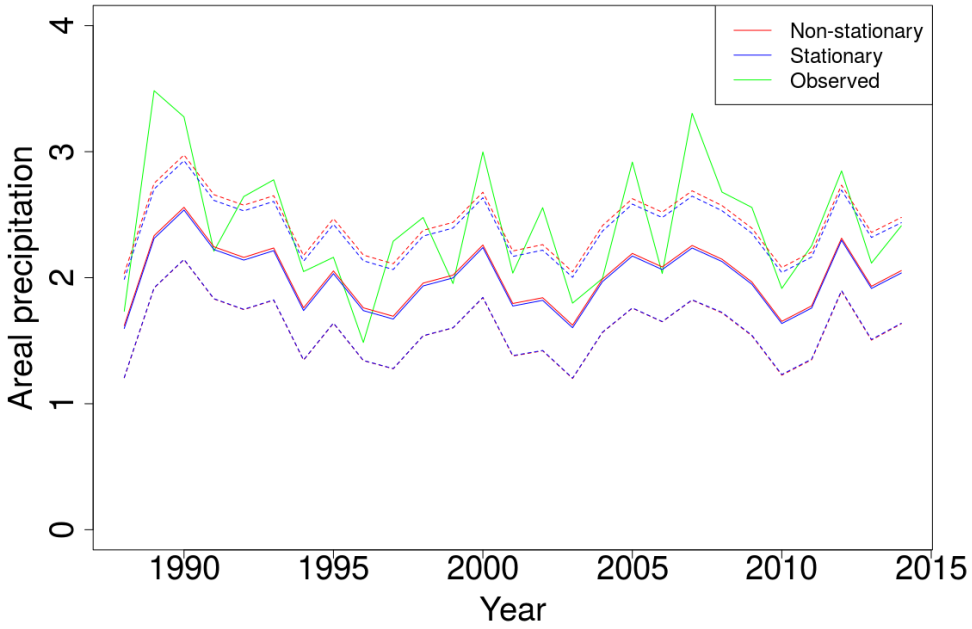


Figure 6.4: Estimated areal precipitation at Svartavatn using stationary model (blue line) and non-stationary model (red line), along with the observed areal precipitation (green line). The dashed lines represent 95% prediction interval for the stationary (blue) and non-stationary (red) model. The areal precipitation is given in 10^8m^3.

6.2.3 Summary

To summarise, the case study has shown indications of non-stationarity in the process. The estimated range and marginal variance of the climatology from the non-stationary model, varies with elevation. When doing interpolations to points, the non-stationary model seems to have a slightly better predictive performance than the stationary model as the elevation gets high. In particular, it is better at quantifying the uncertainty of the interpolations. When doing interpolations to area, neither of the models are very good.

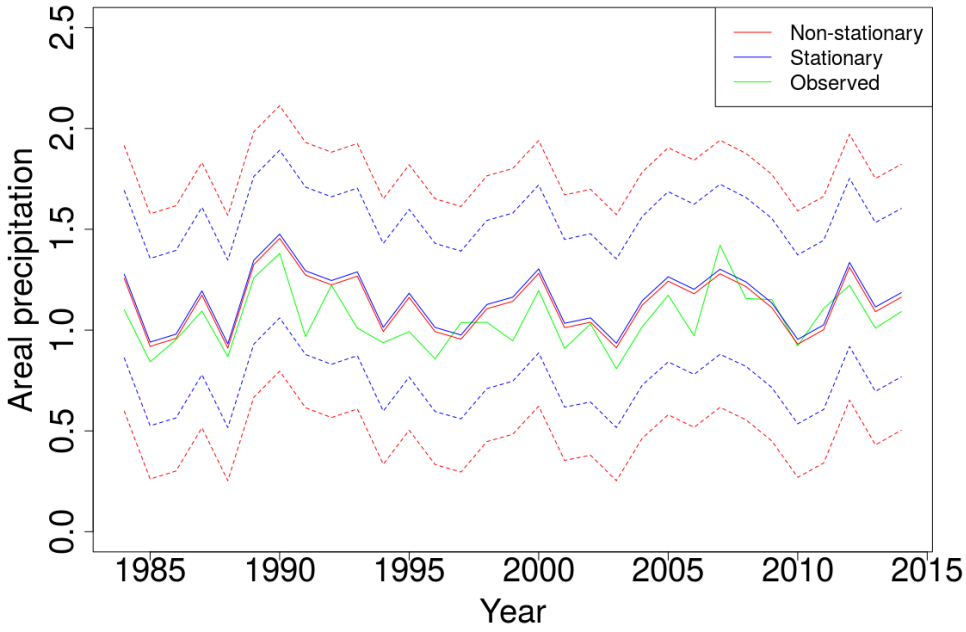


Figure 6.5: Estimated areal precipitation at Slondalsvatn using stationary model (blue line) and non-stationary model (red line), along with the observed areal precipitation (green line). The dashed lines represent 95% prediction interval for the stationary (blue) and non-stationary (red) model. The areal precipitation is given in $10^8 \mathrm{m}^3$.

Chapter 7

Simulation study

In order to get a better understanding of the precipitation models described in Chapter 4, we perform a simulation study. By doing so, we want to further investigate the effect of non-stationarity, and its impact on the predictive performances of the models. An advantage of simulating data, is that the models can be evaluated for more locations and elevations. In particular, the models can be evaluated at higher elevations than the elevation of the weather stations.

7.1 Experimental set up

We consider the 60 weather stations introduced in Chapter 2 as locations where we observe precipitation. Further, we make a grid, covering most of Hordaland, and the grid nodes are locations where precipitation is interpolated. A map of Hordaland, including the weather stations and the grid, can be seen in Figure 7.1. In the figure, the grid nodes are represented by black dots and the weather stations are represented by red dots.

We simulate precipitation from the non-stationary model given in Equation (4.3) at all locations, both the weather stations and the grid nodes. Denote the total number of locations, i.e., both weather stations and grid nodes, n. Further, denote the number of replicates J. The simulation is performed using Algorithm 1, and an example of the R-code used in the simulation can be seen in Appendix

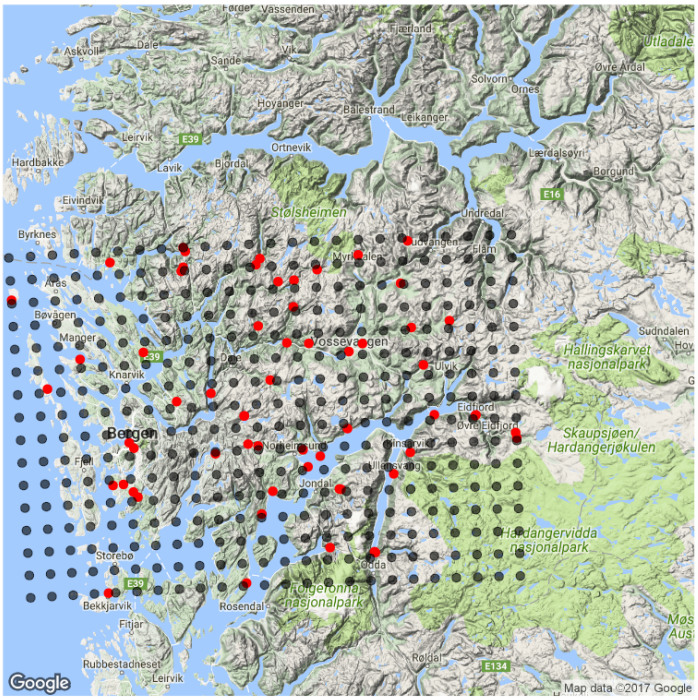


Figure 7.1: A map of Hordaland, including the 60 weather stations, represented by red dots, and a grid used in the simulation study.

B.2.

Initialise α_j, β_h, $\boldsymbol{\theta}^*_{N-S}$;
Make projection matrix $\boldsymbol{A}$;
Make precision matrix for climatology $\boldsymbol{Q}_u$;
Make precision matrix for annual spatial variability $\boldsymbol{Q}_w$;
Sample weights for climatology $\boldsymbol{u} \sim \mathcal{N}_n(0, \boldsymbol{Q}_u^{-1})$ (1 sample, repeated J times);
Sample weights for annual spatial variability $\boldsymbol{w} \sim \mathcal{N}_n(0, \boldsymbol{Q}_w^{-1})$ (J samples);
Calculate true precipitation: $\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{A}(\boldsymbol{u} + \boldsymbol{w}) + \beta_h \boldsymbol{h}$;
Sample measurement error: $\boldsymbol{\epsilon} \sim \mathcal{N}_{n \cdot J}(0, \tau_\epsilon^{-1} 0.1\boldsymbol{\eta})$;
Calculate observed precipitation: $\boldsymbol{y} = \boldsymbol{\eta} + \boldsymbol{\epsilon}$;

Algorithm 1: Algorithm used to simulate precipitation

We use $J = 5$ replicates, i.e., we assume observations from 5 years. We fit the stationary and non-stationary models given in Equation (4.2) and (4.3) to the simulated precipitation at the weather stations, and use the models to interpolate precipitation at the grid nodes, based on the weather stations. This is repeated $S = 100$ times.

Furthermore, we consider the three catchment areas presented in Chapter 2. We simulate annual precipitation from the non-stationary model (4.3) at all weather

stations and at grids inside the catchment areas, using Algorithm 1. We fit the non-stationary model (4.3) and the stationary model (4.2) to the annual precipitation at the weather stations, and use the models to interpolate precipitation at the grids inside the catchment areas, based on the weather stations. We then use the procedure introduced in Section 4.5 to estimate the annual areal precipitation in the catchment areas.

We use four different study designs, where the differences between the designs are the parameter values or the prior distributions of the parameters. In study design 1-3 we use the following values for the intercept, α_j, the elevation coefficient, β_h, and the constant parts of the dependency structure parameters, $\theta_{\tau,1}$ and $\theta_{\kappa,1}$:

$$\alpha_j = 2, \quad \beta_h = 1, \quad \theta_{\tau,1} = 2.42, \quad \theta_{\kappa,1} = -2.36.$$

Here, $\theta_{\tau,1}$ and $\theta_{\kappa,1}$ are the parameters of both the climatology, $c(\boldsymbol{s})$, and the annual spatial variability, $z(\boldsymbol{s})$, i.e., we assume the same values of the parameters for the two fields. These are the same values as discussed in Section 4.3 for the prior means. As prior distribution of these parameters, we also use those suggested in Section 4.3.

In study design 4, we use the values of the posterior means estimated by the non-stationary model in the case study as parameter values and means of the prior distributions for the hyperparameters $\theta_{\tau,1}$ and $\theta_{\kappa,1}$, i.e.,

$$\begin{gathered} \alpha_j = 2, \quad \beta_h = 1, \\ \theta_{\tau,1,w} = 3.90, \quad \theta_{\kappa,1,w} = -1.67 \quad \theta_{\tau,1,u} = 1.77 \quad \theta_{\kappa,1,u} = -2.20. \end{gathered}$$

The variance of the prior distributions is the same as before.

For the coefficients of the explanatory variables in the dependency structure, $\theta_{\tau,h}$ and $\theta_{\kappa,h}$, the parameter specifications of the four designs are given in Table 7.1. For study design 1 we simulate non-stationary data, and the prior mean of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$ are the same as the true parameter values. The range decreases as the elevation increases and the marginal variance increases as the elevation increases. These are the same tendencies as we saw for the estimated posterior range and marginal variance for the climatology in the case study in Chapter 6. The parameter values we use in study design 1 makes the field very non-stationary. As an example, the range is 30 km and the marginal variance is 0.07 at sea level, whereas at elevation $h = 1000$ m, they are 6.7 km and 10.47, respectively. The reason for these excessive values, is to investigate the behaviour of the models when there is a known, large non-stationarity in the process, and to compare it to cases where the non-stationarity is less extreme. In study design 2, the true values of the parameters $\theta_{\tau,h}$ and $\theta_{\kappa,h}$ are the same as in study design 1, but now the prior means are set to zero. By this design, we investigate the prior sensitivity of the non-stationary model, and its ability to recover the parameter values which the data set is based on. In study design 3, the true values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$ are set

to zero, i.e., we simulate stationary data. In study design 4 all the true values of the parameters and prior means of the parameters are the same as the posterior means from the case study. These values make the process non-stationary, but not as much as for study design 1 and 2.

Table 7.1: Specification of parameter values and prior distributions of the parameters in the four different study designs.

	Study design 1	Study design 2	Study design 3	Study design 4
$\theta_{\tau,h,w}$	-4	-4	0	0.51
$\pi(\theta_{\tau,h,w})$	$\mathcal{N}(-4, 1.76)$	$\mathcal{N}(0, 1.76)$	$\mathcal{N}(0, 1.76)$	$\mathcal{N}(0.51, 1.76)$
$\theta_{\kappa,h,w}$	1.5	1.5	0	0.70
$\pi(\theta_{\kappa,h,w})$	$\mathcal{N}(1.5, 1.76)$	$\mathcal{N}(0, 1.76)$	$\mathcal{N}(0, 1.76)$	$\mathcal{N}(0.70, 1.76)$
$\theta_{\tau,h,u}$	-4	-4	0	-1.36
$\pi(\theta_{\tau,h,u})$	$\mathcal{N}(-4, 1.76)$	$\mathcal{N}(0, 1.76)$	$\mathcal{N}(0, 1.76)$	$\mathcal{N}(-1.36, 1.76)$
$\theta_{\kappa,h,u}$	1.5	1.5	0	0.88
$\pi(\theta_{\kappa,h,u})$	$\mathcal{N}(1.5, 1.76)$	$\mathcal{N}(0, 1.76)$	$\mathcal{N}(0, 1.76)$	$\mathcal{N}(0.88, 1.76)$

When evaluating and comparing the predictive performance of the models, we use the mean RMSE ($\overline{\text{RMSE}}$), mean CRPS ($\overline{\text{CRPS}}$), the standard deviation of the CRPS (SD(CRPS)) and the coverage probability of a 95% prediction interval for precipitation over all $S = 100$ simulations. These evaluation schemes were introduced in Section 3.6. For the ($\overline{\text{RMSE}}$), ($\overline{\text{CRPS}}$) and CRPS (SD(CRPS)) we use the mean of all $S = 100$ simulations.

7.2 Analysis of results

7.2.1 Interpolation to points

In this section, we compare the predictive performance of the two models. A summary of the results using the four study designs, is given in Table 7.2. The table displays the coverage probability of a 95% prediction interval, over all $S = 100$ simulations, all grid nodes and all years, the mean CRPS, $\overline{\text{CRPS}}$, the standard deviation of the CRPS, SD(CRPS), and the mean RMSE, $\overline{\text{RMSE}}$. The means are taken over all simulation, all grid nodes and all years.

We see that for study design 1 and 2, the results are very similar. In both cases, the non-stationary model has a better coverage probability, a better $\overline{\text{CRPS}}$ and a better $\overline{\text{RMSE}}$, than the stationary model. Also, the standard deviation of the CRPS is lower for the non-stationary model than the stationary model. Thus, in these cases the non-stationary model has a better predictive performance than the stationary model. The coverage probability of the non-stationary model is

0.95 in both designs, which is just what it should be. For the stationary model, the coverage probability is 0.86, which is too low.

We recall that in study design 1 the prior means of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$ were set equal to the true parameter values, whereas in study design 2 the prior means are set equal to zero. Considering the similar results from the two designs, it seems that the choice of prior means does not have a large impact on the predictive performance of the models.

In study design 3 and 4, we see that the stationary and non-stationary model have the same predictive performance. In all cases, the coverage probability is about 0.95, which is as desired. In study design 3, we fit the models to stationary data. The equally good predictive performance in study design 3, indicates that the non-stationary model detects that the data is stationary.

In study design 4, the prior means of all parameters were set equal to the posterior means from the case study. By the equal predictive performance of the models using this design, it seems that the parameters $\theta_{\tau,h}$ and $\theta_{\kappa,h}$ are so close to zero, that the process is almost stationary. Thus, the estimates of the constant dependency parameters by the stationary model, seem to make a good enough approximation.

Table 7.2: A comparison of the predictive performance of the stationary (S) and non-stationary (N-S) model. The comparison is of the coverage probability of a 95% prediction interval, mean CPRS ($\overline{\text{CRPS}}$), standard deviation of the CRPS (SD(CRPS)) and mean RMSE ($\overline{\text{RMSE}}$). The results are based on $S = 100$ simulations.

	Coverage probability		$\overline{\text{CRPS}}$		SD(CRPS)		$\overline{\text{RMSE}}$	
	S	N-S	S	N-S	S	N-S	S	N-S
Study design 1	0.86	0.95	1.36	1.13	4.15	2.78	6.70	6.52
Study design 2	0.86	0.95	1.36	1.17	4.16	2.86	6.73	6.67
Study design 3	0.94	0.94	0.22	0.22	0.03	0.03	0.87	0.87
Study design 4	0.94	0.95	0.30	0.30	0.06	0.06	1.24	1.24

In Table 7.3, the same comparison of the predictive performance of the two models as in Table 7.2 is displayed, but in this case, the evaluation is only based on locations having a higher elevation than 500 m.

For study design 1 and 2, we have the same tendencies as in Table 7.2, with the results being the same for the two designs, and with the non-stationary model having a better predictive performance than the stationary model. The coverage of the non-stationary model is still 0.95, whereas for the stationary model it has decreased and is about 0.72, which is low. This shows that the stationary model performs poorer as the elevation increases, which is as expected.

For study design 3, the results are very similar as those in Table 7.2. This is as expected, because the increase in elevation does not have any impact on the dependency structure in this design.

For study design 4, we notice that the coverage probability has decreased to 0.91 for the stationary model. In Table 7.2 there was no notable difference between the two models for study design 4, whereas now the non-stationary model has a better coverage probability. This indicates that even though the values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$ are quite close to zero, the non-stationarity of the data has some impact as the elevation increases.

Table 7.3: A comparison of the predictive performance of the stationary (S) and non-stationary (N-S) model at weather stations having an altitude higher than 500 m. The comparison is of the coverage probability of a 95% prediction interval, mean CPRS ($\overline{\text{CRPS}}$), standard deviation of the CRPS (SD(CRPS)) and mean RMSE ($\overline{\text{RMSE}}$). The results are based on $S = 100$ simulations.

	Coverage probability		$\overline{\text{CRPS}}$		SD(CRPS)		$\overline{\text{RMSE}}$	
	S	N-S	S	N-S	S	N-S	S	N-S
Study design 1	0.71	0.95	2.29	2.03	6.91	4.17	9.54	9.31
Study design 2	0.72	0.95	2.29	2.10	6.96	4.20	9.53	9.58
Study design 3	0.94	0.94	0.24	0.24	0.03	0.03	0.96	0.97
Study design 4	0.91	0.95	0.37	0.37	0.09	0.08	1.49	1.49

Figure 7.2 and 7.3 display plots of the coverage probability (left column) and $\overline{\text{CRPS}}_i$ (right column) against elevation, for the two models, using the four study designs. The coverage probability and the mean of the CRPS is over all $S = 100$ simulations and all years at each location i. In all plots, blue circles correspond to the stationary model and red circles correspond to the non-stationary model.

For study design 1 and 2, the plotted coverage probability shows that the non-stationary model has a good predictive performance at all elevations. The stationary model, has too high coverage probability, at low elevations, indicating too large variance. As the elevation increases, the coverage probability decreases, and becomes very low. For study design 3, the coverage probability lies around the 0.95-line for both models and all elevations. For study design 4, the coverage probability of the non-stationary model lies around the 0.95-line, as before. For the stationary model, it is a bit higher for low elevations, and lower for higher elevation. Thus, we see again that the values of $\theta_{\tau,h}$ and $\theta_{\kappa,h}$ has an impact on the coverage probability as the elevation gets high enough.

The $\overline{\text{CRPS}}_i$ increases as the elevation increases for both models and all study designs. This is as expected, because the simulated measurement error increases as elevation increases. Furthermore, the grid nodes at higher elevations have in general a larger distance to the closest weather stations than those with lower

elevations, which also might cause the higher $\overline{\text{CRPS}}_i$ at higher elevations. We see that for study design 1 and 2 the $\overline{\text{CRPS}}_i$ is better for the non-stationary model than the stationary model, and the difference between the two models increases as elevation increases. For study design 3 and 4, the $\overline{\text{CRPS}}_i$ is about the same for both models.

Figure 7.5 and 7.4 present maps of Hordaland where the grid nodes are plotted with a colour scale representing the coverage probability of a 95% prediction interval at the given location. The coverage probability is over all $S = 100$ simulations and all years. The left column displays the coverage probability of the stationary model and the right column displays the coverage probability of the non-stationary model. The same colour scale is used in all plots.

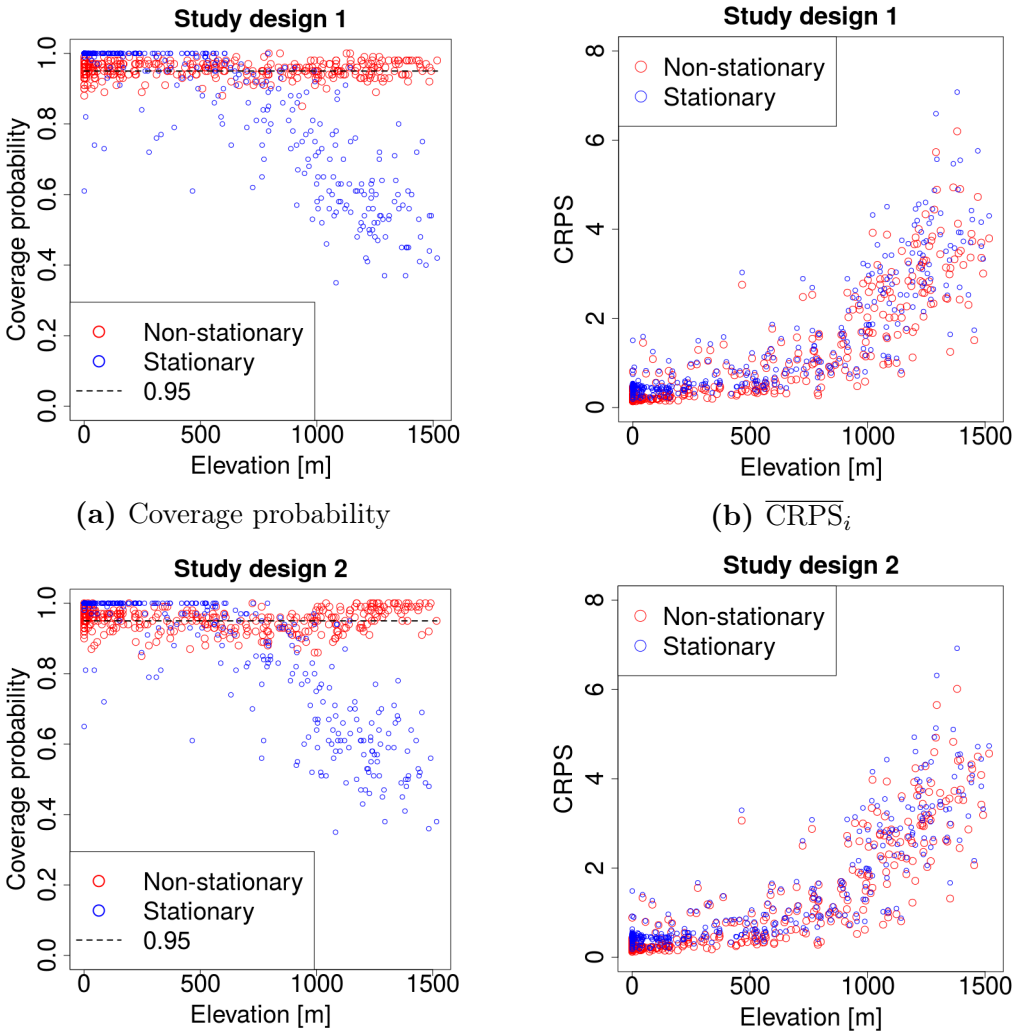
The maps from study design 1 and 2 show that the coverage probability of the stationary model decreases when moving away from the cost, towards the more mountainous areas. The non-stationary model has about the same coverage probability in all the domain. For study design 3, both models have coverage probability close to 0.95 all over the domain. For study design 4, the stationary model has high coverage probability at the flat areas at the cost, but it decreases a bit as we move towards the more mountainous areas. The coverage probability of the non-stationary model is close to 0.95 all over the domain, as before.

7.2.2 Interpolation to area

Table 7.4 displays the coverage probability of a 95% prediction interval and mean CRPS, $\overline{\text{CRPS}}$, for the posterior prediction of areal precipitation at the three catchment areas using the stationary and non-stationary model. The coverage probability and the mean are over $S = 100$ simulations. With this number of simulation, a 95% confidence interval of the coverage probability is [0.91,0.99].

For study design 1 and 2, where the process is very non-stationary, the coverage probability and the $\overline{\text{CRPS}}$ from the non-stationary model is better than from the stationary model. In particular, we notice the catchment area Slondalsvatn. Here, the coverage probability is as low as 0.43 and 0.40 for study design 1 and 2, respectively. For the non-stationary model, it is 0.97 and 0.98, respectively, which both are inside the 95% confidence interval of the coverage probability.

We notice that the coverage probability is better in study design 1 than in study design 2 for all catchment areas, and for both the stationary and non-stationary model. We recall that in study design 1, the prior means of the coefficients of the explanatory variable in the dependency structure, $\theta_{\tau,h}$ and $\theta_{\kappa,h}$, are the same as the true values, whereas in study design 2 they are set to zero. This indicates



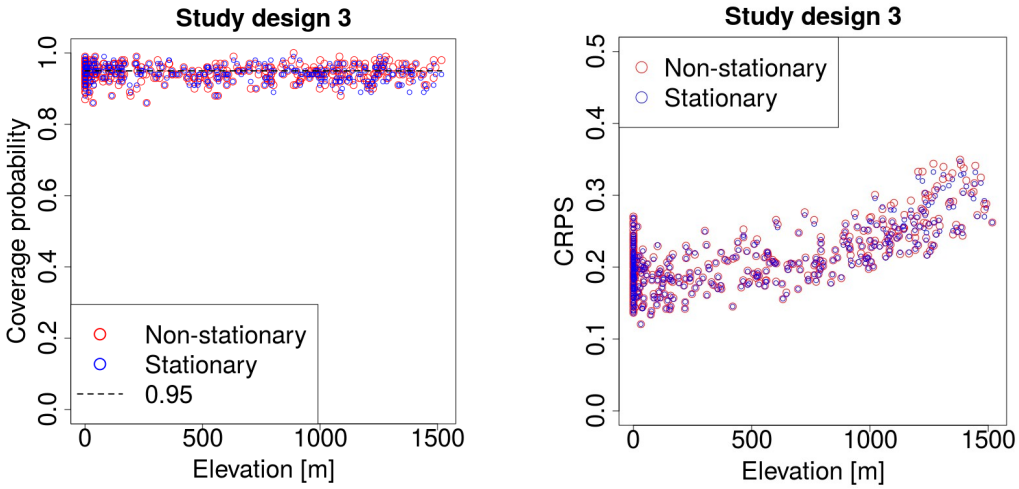
(a) Coverage probability

(b) $\overline{\text{CRPS}}_i$

(c) Coverage probability

(d) $\overline{\text{CRPS}}_i$

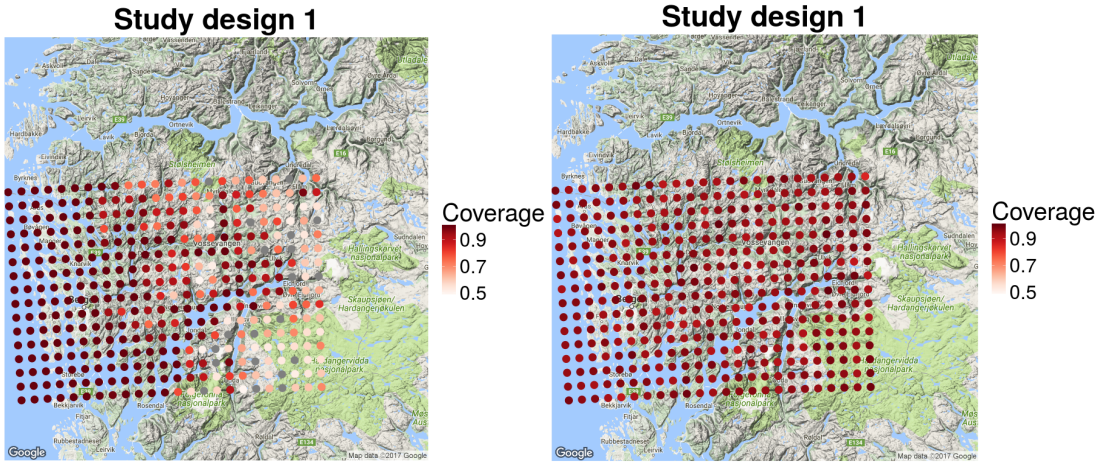
Figure 7.2: A comparison of the predictive performances of the non-stationary (red) and stationary (blue) models for study design 1 and 2. In the left column the coverage probability of a 95% prediction interval for the models is plotted against the elevation. In the right column the mean CRPS, $\overline{\text{CRPS}}_i$, is plotted against the elevation. The coverage probability and the mean of the CRPS is over all $S = 100$ simulations and all years at each location i.



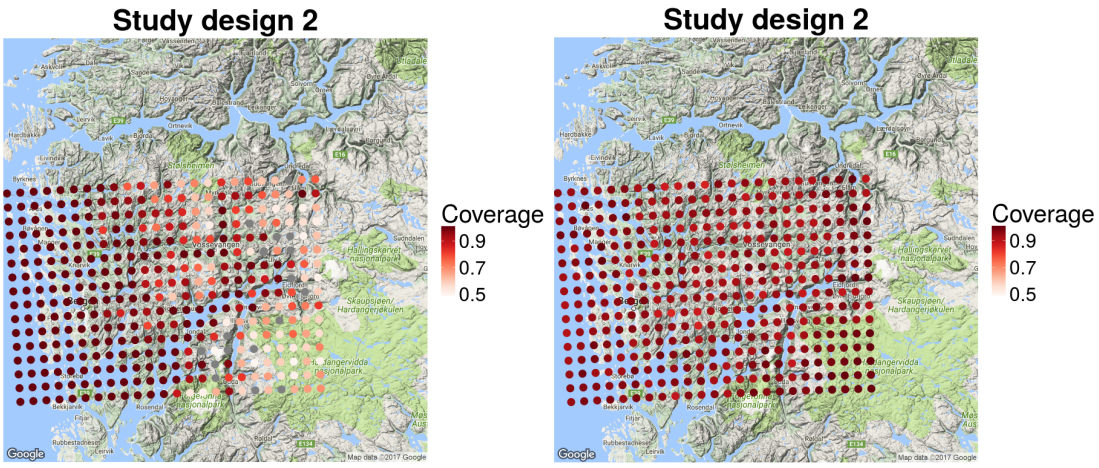
(a) Coverage probability **(b)** $\overline{\text{CRPS}}_i$

(c) Coverage probability **(d)** $\overline{\text{CRPS}}_i$

Figure 7.3: A comparison of the predictive performances of the non-stationary (red) and stationary (blue) models for study design 1 and 2. In the left column the coverage probability of a 95% prediction interval for the models is plotted against the elevation. In the right column the mean CRPS, $\overline{\text{CRPS}}_i$, is plotted against the elevation. The coverage probability and the mean of the CRPS is over all $S = 100$ simulations and all years at each location i.

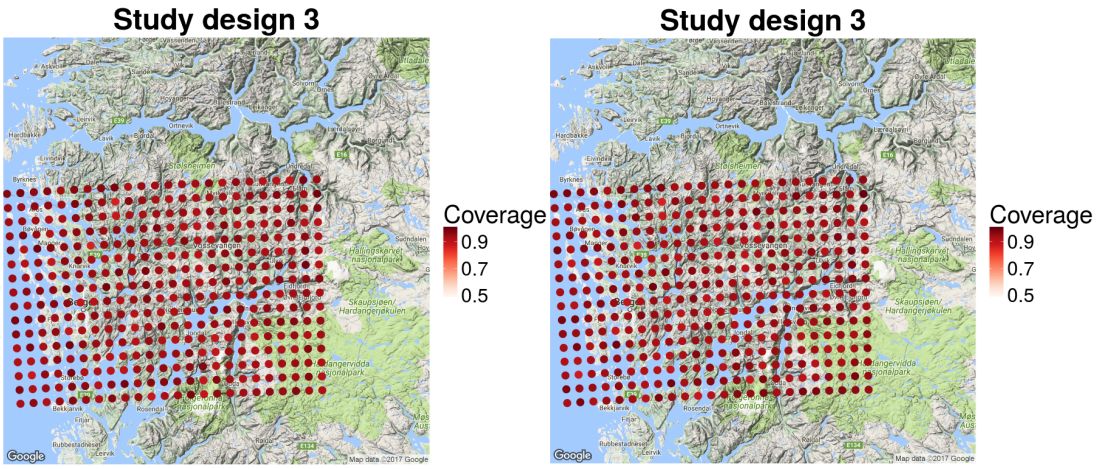


(a) Stationary model **(b)** Non-stationary model

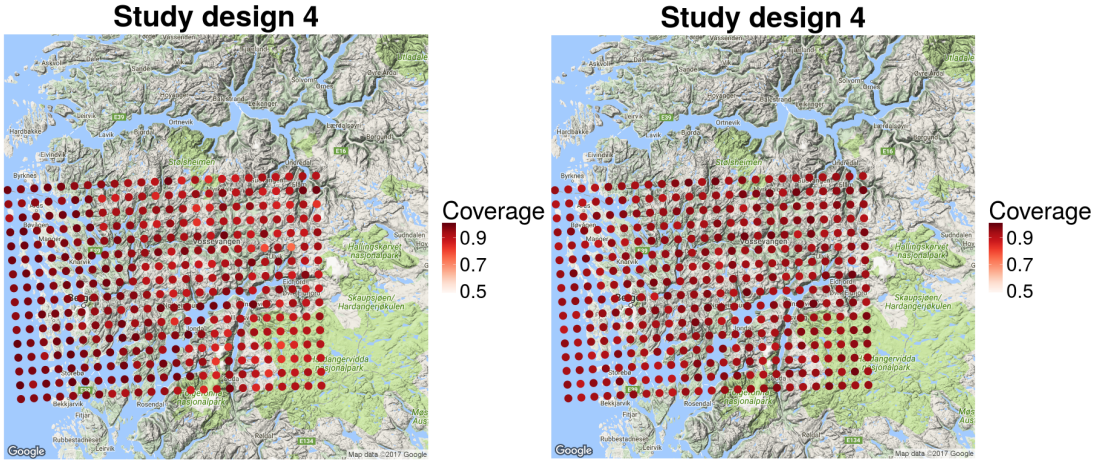


(c) Stationary model **(d)** Non-stationary model

Figure 7.4: Maps of Hordaland, where the grid nodes are plotted with a colour scale, representing their coverage probability of a 95% prediction interval over $S = 100$ simulations. The left column displays the coverage probability of the stationary model, and the right column displays the coverage of the right column. The same colour scales are used for the non-stationary and stationary models.



(a) Stationary model **(b)** Non-stationary model



(c) Stationary model **(d)** Non-stationary model

Figure 7.5: Maps of Hordaland, where the grid nodes are plotted with a colour scale, representing their coverage probability of a 95% prediction interval over $S = 100$ simulations. The left column displays the coverage probability of the stationary model, and the right column displays the coverage of the right column. The same colour scales are used for the non-stationary and stationary models. The plots are based on $S = 100$ simulations.

that when estimating areal precipitation, the choice of prior distributions for the hyperparameters is important.

For study design 3, where the simulated data is stationary, the results from the two models are quite similar, as expected.

For study design 4, where the values of the hyperparameters are the same as the estimated posterior mean in the case study, the non-stationary model is slightly better than the stationary model at Fjellanger and Svartavatn. At Slondalsvatn, the non-stationary model is notably better. Here, the coverage probability of the stationary model is 0.86, which is low, whereas for the non-stationary model, it is 0.97, which is inside the 95% confidence interval.

These results coincide with what we saw in the case study; the predictive performance of the stationary model is not good enough as the elevation increases. At Slondalsvatn, which is located at higher elevation than the other catchment areas, the stationary model has a poor predictive performance. A wider prediction interval is necessary to cover the uncertainty of the data at this elevation. For the non-stationary model, the prediction interval is wider.

Table 7.4: Comparison of the predictive performance of the stationary (S) and non-stationary (N-S) model at the catchment areas. The comparison is of the coverage probability of a 95% prediction interval and mean CPRS ($\overline{\text{CRPS}}$) over $S = 100$ simulations. The predictions are of areal precipitation.

	Catchment area	Coverage probability		$\overline{\text{CRPS}}$	
		S	N-S	S	N-S
Study design 1	Fjellanger	0.83	0.96	0.66	0.52
	Svartavatn	0.83	0.95	0.54	0.45
	Slondalsvatn	0.43	0.97	3.39	2.95
Study design 2	Fjellanger	0.77	0.89	0.69	0.57
	Svartavatn	0.77	0.88	0.65	0.50
	Slondalsvatn	0.40	0.98	3.34	3.19
Study design 3	Fjellanger	0.91	0.92	0.11	0.11
	Svartavatn	0.94	0.94	0.10	0.10
	Slondalsvatn	0.92	0.95	0.16	0.17
Study design 4	Fjellanger	0.89	0.91	0.17	0.18
	Svartavatn	0.91	0.93	0.12	0.12
	Slondalsvatn	0.86	0.97	0.30	0.32

7.2.3 Summary

To summarise, this simulation study has illustrated that when the underlying process is sufficiently non-stationary, i.e., when the varying part of the depen-

dency structure, $\theta_{\tau,h}h$ and $\theta_{\kappa,h}h$, differ enough from zero, a non-stationary model is necessary to make good interpolations. In particular, this is necessary when the elevation increases. When the data is stationary, the predictive performances of a stationary and non-stationary model can be equally good, when the prior means of $\theta_{\kappa,h}$ and $\theta_{\tau,h}$ are set to zero.

Chapter 8

Discussion and conclusion

In this thesis, we have investigated the annual precipitation process in the county Hordaland, Norway. We have fitted a stationary and a non-stationary latent Gaussian model to observations of annual precipitation using the SPDE approach to spatial modelling. We used the methodology INLA to do statistical inference and to make interpolations.

One of the purposes of the thesis was to learn about the precipitation process in areas characterised by orographic precipitation. Furthermore, we aimed to make good statistical models for interpolation of precipitation, and the models should in particular be able to estimate the uncertainty of the interpolations. The results from the studies have shown that such a model should be non-stationary. By fitting a model with dependency structure varying with elevation to annual observations of precipitation, we have seen that the explanatory variables of the dependency structure are significantly different from zero. Further, a non-stationary model has better coverage and better CRPS than a stationary model when interpolating to points. This is particularly the case when the elevation increases, indicating that the process becomes more non-stationary at locations with higher elevation. When doing predictions of areal precipitation for three catchments, the predictive performance of the two models were quite similar. At two of the considered catchment areas, both models had a poor coverage probability of a 95% confidence interval. At a third catchment area, both models made a very wide prediction interval, especially the non-stationary model. Thus, we have seen that even though a non-stationary model, with dependency structure varying with elevation, is slightly better than a stationary model at doing interpolations and quantifying the uncertainty, it still has some large errors. The errors are most noticeable when doing predictions of areal precipitation.

Another purpose of this thesis has been to obtain better knowledge about station-

ary and non-stationary modelling of spatial processes. In particular, we aimed to explore how large degree of non-stationarity there has to be in a process before it is detectable and relevant for the predictive performance. This was done through simple toy examples and through a simulation study. Both the toy examples and the simulation study showed that when the non-stationarity of the process is quite small, the stationary model make constant approximations of the dependency parameters which make quite good interpolations and uncertainty estimations. However, when the non-stationarity of the process gets more extreme, a stationary model does not have a sufficiently good predictive performance.

We also explored how the models perform when there are many observations at low elevations, whereas the locations of interpolations are higher located. The results showed that the predictive performance of the non-stationary model was good in all cases, whereas the stationary model performed poorly. When all observations where located at low elevations, the coverage probability of a 95% prediction interval and the CRPS of the stationary model decreased as the number of observations increased.

Further, we investigated the predictive performance of the models when doing predictions of areal precipitation at catchments located at higher elevation than the observations. We explored how the inclusion of one observation inside the catchment affected the predictive performances of the models. The results showed that the predictive performance of both models became noticeably better when an observation was included inside the catchment areas.

The results from the toy examples and the simulation study indicate that one of the reasons for poor performance of the non-stationary model when doing prediction of areal precipitation at high elevations, might be that almost all the weather stations are located at low elevations. Most of the weather stations are located much lower than the catchment areas. Thus, the non-stationary model might not be able to detect the degree of non-stationarity at higher elevations.

Appendix A

The multivariate normal distribution

Throughout this thesis, we have used the multivariate normal distribution when describing the processes and building our models. Some properties of the normal distribution was introduced in Chapter 3. In this appendix we introduce some additional properties of the normal distribution, which we use in the thesis. We also give a proof to one of the theorems in Section 3.2.

A.1 Definition and properties

In this section we introduce some of the basic properties of the multivariate normal distribution. The section is based on Tong (1990), and the definitions and theorems are taken from this book.

A.1.1 Definition

The multivariate normal distribution is defined by the following:

Let $\boldsymbol{Y}$ be an n-dimensional random vector, with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We say that Y follows a multivariate normal distribution if (i) $\boldsymbol{\Sigma}$ is positive definite and (ii) the density function of $\boldsymbol{Y}$ has the form

$$f(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right).$$

A.1.2 Marginal distribution

We now consider the marginal distributions of a multivariate normal variable. We partition the variable $\boldsymbol{Y}$, and the mean and covariance matrix $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, as

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{\mathbf{22}} \end{bmatrix}. \tag{A.1}$$

Here, $\boldsymbol{Y}_1$ and $\boldsymbol{\mu}_1$ is of dimension k_1, $\boldsymbol{Y}_2$ and $\boldsymbol{\mu}_2$ is of dimension k_2 and $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$ are of dimensions $k_1 \times k_1$, $k_1 \times k_2$, $k_2 \times k_1$ and $k_2 \times k_2$, respectively, and $k_1 + k_2 = n$.

For such a partition, the following theorem can be proved:

If $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then for any fixed $k_1, k_2 < n$ where $k_1 + k_2 = n$, the marginal distributions of $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are

$$\boldsymbol{Y_1} \sim \mathcal{N}_{k_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \quad \textit{and} \quad \boldsymbol{Y_2} \sim \mathcal{N}_{k_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$$

A proof is given in Tong (1990).

A.1.3 Conditional distribution

Consider again the partitioning given in (A.1). We now consider the conditional distribution of $\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{y}_2$. The conditional distribution is given by the following theorem:

Let $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ being positive definite, and partition $\boldsymbol{Y}$ as in (A.1)*, with $k_1 + k_2 = n$. Then the conditional distribution of $\boldsymbol{Y}_1$ given $\boldsymbol{Y}_2 = \boldsymbol{y}_2$ is*

$$\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{y}_2 \sim \mathcal{N}_{k_1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*),$$

with

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2) \quad \textit{and} \quad \boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

A proof is given in Tong (1990).

A.2 Proof of the theorem in Section 3.2

In Section 3.2 we stated the following theorem:

If $\boldsymbol{Y}$ is Gaussian with mean $\boldsymbol{\mu}$ and positive definite precision matrix $\boldsymbol{Q}$, then

$$Y_i \perp Y_j | Y_{-ij} \iff Q_{ij} = 0 \quad \textit{for} \quad i \neq j.$$

We now reproduce the proof provided in Rue and Held (2005) of this theorem. Before we start, we recall another theorem from Section 3.2:

Conditional independence is implied if and only if the following is true:

$$\pi(Y_i, Y_j, Y_{-ij}) = \pi(Y_i, Y_{-ij})\pi(Y_j, Y_{-ij}).$$

This theorem is used in the proof.

Proof: Partition the vector $\boldsymbol{Y}$ as $[Y_i\ Y_j\ \boldsymbol{Y}_{-ij}]^T$. Fix $i \neq j$, and assume $\boldsymbol{\mu} = 0$. This does not lead to any loss of generality. The joint probability of these elements is then

$$\begin{aligned}\pi(Y_i, Y_j, \boldsymbol{Y}_{-ij}) &\propto \exp\left(-\frac{1}{2}\sum_{k,l} Y_k Q_{k,l} Y_l\right) \\ &\propto \exp\left(-\frac{1}{2} Y_i Y_j (Q_{ij} + Q_{ji}) - \frac{1}{2}\sum_{\{k,l\}\neq\{i,j\}} Y_k Q_{k,l} Y_l\right).\end{aligned}$$

We see that only the first term in the second line contains $Y_i Y_j$, and the last sum in the second line is the same as $\pi(Y_i Y_{-ij})\pi(Y_j Y_{-ij})$. Thus, we have that $\pi(Y_i, Y_j, Y_{-ij}) = \pi(Y_i, Y_{-ij})\pi(Y_j, Y_{-ij})$ if and only if $Q_{ij} = 0$, i.e., we have conditional independence if and only if $Q_{ij} = 0$.

Appendix B

INLA-code

In this appendix we present some of the INLA-code used when doing simulations and fitting models in R. The functions are from the R-package `R-INLA`. We refer to `www.r-inla.org` for more information about the package, and Lindgren and Rue (2013) for more examples and demonstrations of the package.

B.1 Making mesh

Both when doing simulations and when fitting models , we use the SPDE-approach, and we need a triangular mesh. The following INLA-code is used to construct a mesh:

```
#Making mesh
max.edge=c(1,14); offset=c(1,30); cutoff=4
mesh=inla.mesh.2d(loc.domain=as.matrix(all.coords) ,cutoff=cutoff ,
    max.edge=max.edge ,offset=offset)
```

Here, `all.coords` is a data frame storing the coordinates of all locations where we observe or interpolate precipitation. The offset variable defines the size of the inner and outer extensions around the locations, and max.edge specifies the maximum length of the triangle edges in the inner and outer extensions. The cutoff variable prevents too many triangles to be built around clustered locations.

B.2 Simulation

The following R-code is used to simulate non-stationary precipitation:

```
#Parameter values
bet.0 <- 2; bet.h <- 1; lambda <- 1
w.theta.tau.1 <- 2.42; w.theta.tau.h <- -4
w.theta.kap.1 <- -2.36; w.theta.kap.h <- 1.5
u.theta.tau.1 <- 2.42; u.theta.tau.h <- -4
u.theta.kap.1 <- -2.36; u.theta.kap.h <- 1.5

#Make non-stationary SPDE-object
spde=inla.spde2.matern(mesh,alpha=2 ,B.tau=cbind(0,1,elev.mesh,0,0),
    B.kappa=cbind(0,0,0,1,elev.mesh))

#Make projection matrix
A = inla.spde.make.A(mesh, loc = as.matrix(coords.all),index = rep
    (1:n.all, times = n.repl),repl = rep(1:n.repl, each = n.all) )

#Make precision matrices for climatology and annual spatial
    variability
Q.w=inla.spde.precision(spde=spde ,theta=c(w.theta.tau.1,w.theta.tau
    .h,w.theta.kap.1,w.theta.kap.h)) #Annual spatial variability
Q.u=inla.spde.precision(spde=spde ,theta=c(u.theta.tau.1,u.theta.tau
    .h,u.theta.kap.1,u.theta.kap.h)) #Climatology

#Sample weights
w=as.vector(inla.qsample(n=n.repl,Q=Q.w))
u=as.vector(inla.qsample(n=1,Q=Q.u))
u<-rep(u,n.repl)

#Calculate true precipitation
precip.all=bet.0+as.vector(A%*%(w+u))+ bet.h*rep(elev.all,n.repl)

#Sample measurement error
epsilon<-rnorm(length(precip.all),mean=0,sd=0.1*abs(precip.all))

#Calculate observed precipitation
precip.all<-precip.all+epsilon
```

Here, `mesh` is a mesh-object constructed using the code in Section B.1, `coords.all` is a matrix with coordinates of all locations where precipitation is to be simulated, `elev.mesh` and `elev.all` are vectors of elevation at the locations of the mesh nodes and the locations where we simulate precipitation, and `n.all` and `n.repl` is the number of locations and replicates.

The function `inla.spde2.matern()` is used to make an SDPE-object. The matrices `B.tau` and `B.kap` store the basis functions of the SPDE (see Section 3.3). The five columns of the matrices are multiplied with the corresponding

element of a parameter vector $[\theta_0\ \theta_1\ \theta_2\ \theta_3\ \theta_4]$, in order to obtain the parameter space. θ_0 is an offset, which we don't use, and we set the first column to zero. The next four columns correspond to our hyperparameters; $\theta_{\tau,1}$, $\theta_{\tau,h}$, $\theta_{\kappa,1}$ and $\theta_{\kappa,h}$ (Ingebritsen, 2014).

In the stationary case, we only need to specify the first matrix row. The following code is used to construct an SPDE-object in the stationary case:

```
spde_S<-inla.spde2.matern(mesh=mesh,alpha=2, B.tau=matrix(c
    (0,1,0),nrow=1,ncol=3),B.kappa=matrix(c(0,0,1),nrow=1,ncol=3))
```

B.3 Model fitting

In this section we present the code used to fit models and interpolate precipitation. We only give the code used to fit non-stationary models.

```
#Making A-matrices
A.obs=inla.spde.make.A(mesh,loc=as.matrix(coords.obs),index = rep(1:
    n.obs, times = n.repl),repl = rep(1:n.repl, each = n.obs) )
A.pred=inla.spde.make.A(mesh,loc=as.matrix(coords.pred),index = rep
    (1:n.pred, times = n.repl),repl = rep(1:n.repl, each = n.pred) )

#Assigning priors to kappa and tau
priors.w<-list(mu.t=w.theta.tau.1, sig.t=2.4190, mu.t.h=w.theta.tau.
    h, sig.t.h=1, mu.k=w.theta.kap.1, sig.k=0.8089, mu.k.h=w.theta.
    kap.h,sig.k_h=1)
priors.u<-list(mu.t=u.theta.tau.1, sig.t=2.4190, mu.t.h=u.theta.tau.
    h, sig.t.h=1, mu.k=u.theta.kap.1, sig.k=0.8089, mu.k.h=u.theta.
    kap.h,sig.k.h=1)


#Making SPDE-object
spde.w<-inla.spde2.matern(mesh=mesh,alpha=2 ,B.tau=cbind(0,1,elev.
    mesh,0,0),B.kappa=cbind(0,0,0,1,elev.mesh),theta.prior.mean = c(
    priors.w$mu.t,priors.w$mu.t.h, priors.w$mu.k, priors.w$mu.k.h),
    theta.prior.prec = c(1/priors.w$sig.t,1/priors.w$sig.t.h, 1/
    priors.w$sig.k, 1/priors.w$sig.k.h))

spde.u<-inla.spde2.matern(mesh=mesh,alpha=2 ,B.tau=cbind(0,1,elev.
    mesh,0,0),B.kappa=cbind(0,0,0,1,elev.mesh),theta.prior.mean = c(
    priors.u$mu.t,priors.u$mu.t.h, priors.u$mu.k, priors.u$mu.k.h),
    theta.prior.prec = c(1/priors.u$sig.t,1/priors.u$sig.t.h, 1/
    priors.u$sig.k, 1/priors.u$sig.k.h))

#Making indexes
s.index.w=inla.spde.make.index(name="field.w",n.spde=spde.w$n.spde,n
    .repl = n.repl)
```

```
s.index.u=inla.spde.make.index(name="field.u",n.spde=spde.u$n.spde)
s.index.u$field.u=s.index.w$field.w
s.index.u$field.u.repl=rep(1,length(s.index.w$field.w.repl))
s.index.u$field.u.group=s.index.u$field.u.repl


#Intercept
intercept=rep(1:n.repl,each=mesh$n)

#Stacks
stack.obs=inla.stack(data=list(precip=precip.obs),effects=list(c(s.
    index.w,s.index.u,intercept=list(1)),list(elevation=elev.obs.repl
    )),A=list(A.obs,1),tag='obs')

stack.pred=inla.stack(data=list(precip=NA),effects=list(c(s.index.w,
    s.index.u,intercept=list(1)),list(elevation=elev.pred.repl)),A=
    list(A.pred,1),tag="pred")

join.stack=inla.stack(stack.obs,stack.pred)


#Priors
tau.precision.prior<-list(theta=list(param=c(shape=1,rate=1)))
precip.scale<-1/(0.1*precip.all)^2
prior.beta0<-c(2,1)
prior.beta1<-c(1,1)


#Formula
formula<-precip~-1+intercept+f(field.w,model=spde.w,replicate =
    field.w.repl)+f(field.u, model=spde.u)+elevation


#Model
model<-inla(formula,family = "gaussian",data=inla.stack.data(join.
    stack),  control.predictor=list(A=inla.stack.A(join.stack),
    compute=T),control.fixed = list(expand.factor.strategy = "inla",
    mean=list(intercept=prior.beta0[1], elevation=prior.beta1[1]),
    prec=list(intercept=prior.beta0[2],elevation=prior.beta1[2])),
    control.family=list(hyper=tau.precision.prior),scale=precip.scale
    )
```

Here, `coords.obs` and `coords.pred` are data frames storing the coordinates of locations where we observe and interpolate precipitation. `n.obs` and `n.pred` are the numbers of locations where we observe and interpolate precipitation, and `n.repl` is the number of replicates. `mesh` is a mesh-object constructed using the code in Section B.1. Again, we construct SPDE-objects, and now we also assign priors to the hyperparameters when making the SPDE-object. The function `inla.spde.make.index()` makes the indexes required by the SPDE model (Blangiardo and Cameletti, 2015).

The function `inla.stack()`, is a tool used to manage the SPDE object, when many random effects are included in the linear predictor.

When we predict the areal precipitation in catchment areas, we are interested in linear combinations of the precipitation. This can be included in the model fitting in R-INLA, using the function `inla.make.lincomb()`. The following code demonstrates how it can be used:

```
1
lc1 = inla.make.lincomb(Predictor = c(rep(NA,n.obs),rep(1,n.pred)))
3
names(lc1) = "lc1"
5
model<-inla(formula,family = "gaussian",data=inla.stack.data(join.
    stack), lincomb=lc1, control.predictor=list(A=inla.stack.A(join.
    stack),compute=T),control.fixed = list(expand.factor.strategy = "
    inla",mean=list(intercept=prior.beta0[1], elevation=prior.beta1
    [1]),prec=list(intercept=prior.beta0[2],elevation=prior.beta1[2])
    ),control.family=list(hyper=tau.precision.prior),scale=precip.
    scale)
```

In the above code, the `Predictor` argument in the `inla.make.lincomb()` function states that the linear predictors should be added. In this code, the linear predictor is a vector, where the `n.obs` first elements are observations, and the next `n.pred` elements are predictions. Thus, the above code states that linear combination should be of all the predictions.

Appendix C

Weather stations

Throughout the thesis we use 60 weather stations located in the county Hordaland. Information about the weather stations is given in Table C.1. The information is provided by the Norwegian Water Resources and Energy Directorate (NVE), and is also available at the web portal eKlima.no, which gives public access to the climate database of the Norwegian Metrological Institute.

Station number	Name	Elevation [m]	Latitude	Longitude	Average annual precipitation [m/year]
48780	Mauranger kraftstasjon	33	60.1322	6.3305	2.21
49070	Kvåle	342	60.2803	6.3778	2.37
49080	Øvre Krossdalen	342	60.2795	6.3857	2.44
49350	Tyssedal I	32	60.1198	6.5608	1.58
49351	Tyssedal IA	32	60.1218	6.5562	1.71
49490	Ullensvang Forsøksgård	12	60.3185	6.6538	1.56
49550	Kinsarvik	108	60.3725	6.7382	1.37
49580	Eidsfjord-Bu	117	60.4672	6.8605	1.81
49630	Eidsfjord	5	60.4668	7.0723	1.22
49631	Eidsfjord II	20	60.4647	7.0685	1.22
49750	Liset	748	60.4226	7.2739	1.22
49800	Fet i Eidfjord	735	60.4085	7.2798	1.00
49940	Granvin	352	60.5918	6.8050	2.14
50050	Nedre Ålvik	18	60.4323	6.4190	2.59
50070	Kvamsøy	49	60.3631	6.2805	1.98
50080	Øystese Borge	108	60.3790	6.1927	2.32
50110	Kvam Aksneset	13	60.3363	6.2182	2.57
50120	Skulafossen kraftstasjon	16	60.2747	6.0392	2.76
50130	Omastrand	2	60.2170	5.9837	2.65
50150	Hatlestrand	45	60.0422	5.9057	2.15

50175	Austevoll	32	60.0167	5.2038	1.61
50250	Tysse	41	60.3690	5.7473	2.76
50300	Kvamskogen	408	60.3933	5.9133	3.37
50310	Kvamskogen Jonshøgdi	455	60.3887	5.9640	3.05
50350	Samnanger	370	60.4640	5.8938	3.60
50351	Samnanger II	367	60.4640	5.8938	3.78
50450	Fana - Stend	54	60.2728	5.3305	2.16
50460	Fana forsøksstasjon	48	60.2600	5.3533	2.45
50480	Bergen - Sandsli	45	60.2918	5.2798	2.07
50500	Flesland	48	60.2892	5.2265	1.95
50540	Bergen Florida	12	60.3830	5.3327	2.41
50560	Bergen Fredriksberg	41	60.3967	5.3086	2.45
50950	Osterøy Gjerstad	60	60.5000	5.5500	2.53
51010	Fossmark	10	60.5205	5.7243	3.02
51130	Kaldestad	514	60.5540	6.0255	2.61
51250	Øystedal	316	60.6887	5.9647	3.15
51400	Brekkhus	202	60.7365	6.1438	2.67
51440	Evanger	17	60.6470	6.1105	2.20
51470	Bulken	328	60.6455	6.2220	2.00
51530	Vossevangen	54	60.6250	6.4262	1.40
51590	Voss - Bø	125	60.6450	6.4950	1.41
51670	Reimegrend	590	60.6850	6.7433	1.71
51800	Mjølfjelluh	695	60.7022	6.9373	1.91
51990	Myrkdalen Vetlebotn	700	60.8655	6.4733	3.61
52110	Fjellanger II	456	60.7992	6.0660	2.62
52170	Eksingedal	450	60.8028	6.1469	2.65
52220	Gullbrå	579	60.8289	6.2645	2.17
52290	Modalen II	114	60.8410	5.9533	3.08
52310	Modalen III	125	60.8562	5.9733	3.06
52400	Eikanger Myr	72	60.6230	5.3808	2.38
52440	Holsnøy Landsvik	27	60.6055	5.0590	2.02
52475	Blomvåg Sele	30	60.5313	4.8933	1.75
52530	Hell i Søyfyr	20	60.7528	4.7106	1.26
52600	Haukeland	196	60.8248	5.5732	3.76
52601	Haukeland Storevatn	325	60.8347	5.5833	3.96
52640	Matre kraftstasjon	7	60.8740	5.5937	2.86
52650	Matre	18	60.8833	5.5833	3.01
52750	Frøyset	13	60.8462	5.2108	2.45
53160	Jordalen Nåsen	614	60.9004	6.7243	1.98
53180	Brandset	460	60.7950	6.6887	1.60

Table C.1: An overview of 60 weather stations in Hordaland.

Bibliography

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

Gelfand, A. E., Diggle, P. J., Fuentes, M., and Guttorp, P. (2010). *Handbook of spatial statistics*. Chapman & Hall/CRC.

Google (2017). *Google Maps API*. `https://developers.google.com/maps/documentation/elevation/int` [Accessed: 23.03.17].

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems.

Hu, X. and Steinsland, I. (2016). Spatial modelling with system of stochastic partial differential equations. *WIREs Computational Statistics*, page 112–125.

Ingebritsen, R. (2014). *Bayesian spatial modelling of non-stationary processes and misaligned data utilising Markov properties for computational efficiency*. PhD thesis, Norwegian University of Science and Technology.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection.

Lindgren, F. and Rue, H. (2013). Bayesian spatial modelling with r-inla. *Journal of Statistical Software*.

Lindgren, F., Rue, H., and Lindstrõm, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach.

NVE (2015). *Statistiske modeller*. `https://www.nve.no/hydrologi/analysemetoder-og-modeller/statistiske-modeller/` [Accessed: 16.12.16].

Ripley, B. D. (1981). *Spatial Statistics*. John Wiley and Sons, Inc.

Roe, G. H. (2005). Orographic precipitation.

Roksvåg, T. J. T. (2016). A bayesian model for area and point predictions. Master's thesis, Norwegian University of Science and Technology.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. Chapman & Hall/CRC.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations.

Statkraft (2009). *Hydropower*. `http://www.statkraft.com/energy-sources/hydropower/` [Accessed: 30.11.16].

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. economic geography 46, 234–240.

Tong, Y. L. (1990). *The multivariate normal distribution*. Springer-Verlag.

Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance.

Yeager, K. (2017). *SPSS Tutorials: Paired samples t test*. Kent state university libraries. `http://libguides.library.kent.edu/SPSS` [Accessed: 08.06.17].

Ødegård, J. (2017). A study of the importance of including elevation as an explanatory variable in bayesian spatial models for annual precipitation.