Master's thesis

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

August Sørli Mathisen

Inference on extreme hourly precipitation in Norway using INLA

Master's thesis in Applied Physics and Mathematics

Supervisor: Sara Martino

June 2020

NTNU
Norwegian University of Science and Technology

August Sørli Mathisen

Inference on extreme hourly precipitation in Norway using INLA

Master's thesis in Applied Physics and Mathematics
Supervisor: Sara Martino
June 2020

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

NTNU
Norwegian University of
Science and Technology

Abstract

In this thesis a Bayesian hierarchical modelling framework is used to model extreme hourly precipitation in the Oslofjord region. The data considered is hourly precipitation data from 148 weather stations. It is used to study the seasonal and spatial variations in the 99.998% quantile. The model considered has three stages. The first finds thresholds above which the data can be approximated well by extreme value distributions, the second estimates the above threshold probability and the last fits a generalised Pareto distribution to the above threshold observations. Each stage takes the form of a latent Gaussian model. Inference on each model stage is done efficiently by integrated nested Laplace approximations (INLA). The results indicate that the model captures the seasonal and spatial structure well and that INLA is a useful tool for inference on extreme hourly precipitation.

Sammendrag

I denne oppgaven benyttes et hierarkisk Bayesians modellrammeverk for å studere ekstrem timesnedbør i Oslofjordsområdet. Datapunktene består av observert timesnedbør på 148 værstasjoner. Disse brukes for å studere variasjoner i 99.998%-kvantilen fra sted til sted og fra sesong til sesong. Modellen som benyttes, har tre nivå. Det første nivået finner terskelverdier slik at ekstremverdifordelinger er gode tilnærminger for høyere nedbørsmengder enn terskelverdiene. Det andre nivået estimerer sannsynligheten for at en observasjon overskider terskelen, og det tredje tilpasser en generalisert Paretofordeling til terskeloverskridelsene. Alle modellnivå består av en separat latent Gaussisk modell. Inferens på hvert modellnivå blir gjort effektivt med integrerte nestede Laplace approksimasjoner (INLA). Resultatene tilsier at modellen fanger både romlige og sesongvariasjoner godt, og at INLA kan være et nyttig verktøy for inferens på ekstrem timesnedbør.

Preface

This work is conducted for the course TMA4900 and is the final project for my Master of Science at NTNU, Norwegian University of Science and Technology. I would like to convey huge gratitude to my supervisor Sara Martino for great guidance and support over the course of the previous year.

Contents

Chapter 1

Introduction

Precipitation patterns are important for many applications. Fields as different as agriculture, power supply and weather forecasting all depend heavily on good inference on how precipitation varies in time and space. What aspect of the precipitation distribution that is of interest depends on the application. It can range from the average intensity to the number of wet days. However, for many applications, the precipitation extremes are the statistics of interest. This is especially true for infrastructure planning purposes, since any infrastructure should handle not only the average precipitation loads but also the most extreme cases. When dealing with precipitation extremes, aggregation times matter. When studying the reservoir for a hydroelectric power plant, the total yearly or seasonal precipitation amounts matter, and, by extension, the extreme total yearly precipitation amounts. For flood modelling, it is the total weekly or daily precipitation extremes that matters the most. For some applications, even shorter aggregation times are of interest. Examples of this include storm drain and sewage system design as these have to handle the most extreme precipitation amounts at even shorter time scales. However, these shortest time scales are hard to study from a statistics point of view due to a lack of data. Historically, most of the collected precipitation data has been daily precipitation amounts, as many weather stations has had to be inspected manually. In recent years, there has been a large uptick in the amount of hourly precipitation data collected due to improved measurement and data collection equipment. This has lead to about 150 weather stations in the area surrounding the Oslofjord that collect hourly precipitation data. 80% of these started recording after 2009, which means that it is now possible to study the hourly precipitation patterns in Norway to a much larger extent than just a few years ago.

The use of Bayesian hierarchical modelling of precipitation extremes have some precedence. Cooley, Nychka and Naveau (2007) use a spatial model for threshold exceedances to create return level maps of daily precipitation in Colorado. Dyrrdal, Lenkoski, Thorarinsdottir and Stordal (2015) study hourly extreme precipitation in Norway with Bayesian hierarchical models. They create a return level map for the whole of Norway using a spatial model on yearly maxima. In both cases, their inference methods are based on MCMC. Creating a similar map was considered for this project, but it turned out to be difficult with the data at hand.

Recently, integrated nested Laplace approximations (INLA) have become a much used tool in Bayesian hierarchical modelling. It provides fast and accurate approximate inference on a much used class of hierarchical models (Rue, Martino & Chopin, 2009). Recently, there has been some successful applications of INLA for inference on threshold exceedance models. Castro-Camilo, Huser and Rue (2018) use inference by INLA to forecast extreme wind, while Opitz, Huser, Bakka and Rue (2018) use a spatio-temporal model for threshold exceedances to model extreme quantiles for precipitation in the Netherlands, also with INLA. The latter is especially interesting for this project as their model structure is well suited for modelling extreme precipitation within the INLA framework.

For my project thesis, (Mathisen, 2020) a similar setup to Opitz et al. (2018) was used to investigate the temporal structure of the 99.8% quantile for the precipitation at a single weather station located at Blindern in Oslo. This turned out to give sensible estimates for the hourly precipitation quantiles at that station. However, these results do not necessarily extrapolate well across space. In addition the 99.8% quantile does not correspond to a very extreme event in the context of hourly precipitation. It corresponds to precipitation events observed every 21 days which is quite frequent. The goal of this thesis is therefore to expand this method by introducing a spatial component. This strengthens spatial prediction and leads to more available data. This lets us estimate more extreme quantiles. Specifically, we study quantiles at the 99.998% level for hourly precipitation. This means precipitation events that occur every five to six years at

any given station. We study their spatial and seasonal variations using latent Gaussian models to model threshold exceedances. Inference on these models is done with integrated nested Laplace approximations. Chapter 2 contains useful background material including results from extreme value statistics, theory on latent Gaussian models, some details on INLA and a presentation of penalised complexity priors. Next, in chapter 3, the data set is presented. In chapter 4, the model used is presented. Chapter 5 contains a presentation and discussion of the results.

Chapter 2

Background

In this chapter some of the theoretical background for this project will be discussed. First a general introduction to extreme value statistics will be presented with focus on threshold exceedances. Consequently, a brief overview of latent Gaussian models and an introduction to the estimation of latent Gaussian models through integrated nested Laplace approximations is presented. The concept of penalised complexity priors is introduced in the final section.

2.1 Extreme value statistics

As the goal of this project is to estimate extreme quantiles, a brief introduction to extreme value theory is in order. There are two main strategies for modelling extremes. One considers maximum values within blocks of data, for example yearly maxima. The advantage of modeling maxima is that maxima are well understood theoretically (Coles, 2001), and that in many applications the rest of the data is not available. However, if a full set of data points is available, doing inference on only the block maxima disregards most of the observations. A common approach for taking into account a larger proportion of the available data, is to instead model threshold exceedances. That will be the approach used here.

2.1.1 Asymptotic results for threshold exceedances

When modelling extreme threshold exceedances one wants to find the probability

$$P(X > u + y | X > u), \tag{2.1}$$

where X is some random variable and u is a fixed finite threshold. If X follows a distribution $F(x)$, this probability can be expressed as

$$\frac{1 - F(u + y)}{1 - F(u)}. \tag{2.2}$$

However, we do not in general know $F(x)$, and this is a problem as the variance for most approximate distributions will be very high at extremes. Extreme value theory is therefore applied. Assume $X_1, X_2, ..., X_n$ are independently and identically distributed (iid) with cumulative distribution function $F(x)$, and $M_n = \max\{X_1, X_2, ..., X_n\}$. The Fisher-Tippett-Gnedenko theorem (also referred to as the extreme value theorem or extremal types theorem) states that if there exists constants a_n and b_n such that

$$P\left(\frac{M_n - b_n}{a_n} < z\right) \to G(z) \quad \text{as } n \to \infty, \tag{2.3}$$

and G is not degenerate, then

$$G(z; \sigma, \mu, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}. \tag{2.4}$$

This means that if a limiting distribution for the renormalized maximum exists and is not degenerate, the limiting distribution must belong to the specific family of distributions described by equation (2.4). This family is known as the generalized extreme value (GEV) family of distributions and includes the Weibull,

Gumbel and Frechet distributions depending on the parameter ξ (Coles, 2001). Note that probability density functions (pdf) of the Weibull, Frechet and Gumbel distributions are defined on different domains. This means that the domain of the pdf depends on the value of ξ. Specifically, the pdf is defined on the set $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$ where $\sigma > 0$ and μ and ξ are unbounded. This means that if $\xi > 0$, the distribution has a fixed lower bound (Frechet), if $\xi < 0$, the distribution has a fixed upper bound (Weibull), and if $\xi = 0$, the distribution is unbounded (Gumbel). The value of ξ also determines whether the mean and variance of the GEV is finite. If $\xi > 0.5$ then the variance is infinite, and if $\xi > 1$, the mean is infinite. The Fisher-Tippett-Gnedenko theorem is often considered an extreme value analogy to the central limit theorem for sums of random variables, but note that there is an essential difference between the two as this theorem holds only if the limiting distribution exists.

Returning to the problem of threshold exceedances, this means that the distribution F in (2.2), can be approximated by the GEV distribution for large u. If we assume that (2.3) holds, (2.4) says that

$$F^n(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$
$$\ln F(z) = -\frac{1}{n}\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}.$$

If we Taylor expand the left hand side, we get that for large z,

$$1 - F(z) = \frac{1}{n}\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi},$$

and inserting this back into (2.2) yields

$$P(X > u + y | X > u) \approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]^{-1/\xi}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}} = \left[1 + \frac{\xi y}{\sigma + \xi(u-\mu)}\right]^{-1/\xi}.$$

This leads to the Pickands-Balkema-de Haan theorem for threshold exceedances. Given that (2.3) holds, the distribution of $(X - u)$ conditioned on $X > u$ is approximately distributed as

$$H(y; \sigma_u, \xi) = 1 - \left[1 + \frac{\xi y}{\sigma_u}\right]^{-1/\xi}, \tag{2.5}$$

where $y = x - u$ and

$$\sigma_u = \sigma + \xi(u - \mu). \tag{2.6}$$

This is also known as the generalized Pareto (GP) distribution (Coles, 2001).

Because the GP distribution is based on the GEV distribution, the value of ξ will also determine the domain for the GP distribution. Specifically, if $\xi < 0$, the distribution has an upper bound at $u - \sigma_u/\xi$. If $\xi > 0$, then the distribution has no such upper bound. If $\xi = 0$ the distribution is interpreted as

$$\lim_{\xi \to 0} H(y; \sigma_u, \xi) = 1 - \exp\left[\frac{y}{\sigma_u}\right],$$

which in practice means that y follows an exponential distribution with rate parameter σ_u. The GP distribution is therefore unbounded from above also when $\xi = 0$. Whether or not the mean and variance is finite is also in this case determined by the value of ξ. $\xi \geq 0.5$ leads to an infinite variance while $\xi \geq 1$ leads to an infinite mean.

In practice this parmeterisation is problematic because σ_u depends linearly on ξ which leads to very correlated estimates. This can be seen in Figure 2.1 where data was simulated with parameters similar to those encountered in this project. In that experiment the correlation between shape and scale was -0.65. Because of this (2.5) is in this work parameterised in terms of its q-quantile, κ_q. This yields

$$H(y) = 1 - \left[1 + y \cdot \frac{(1-q)^{-\xi} - 1}{\kappa_q}\right]^{-1/\xi}. \tag{2.7}$$

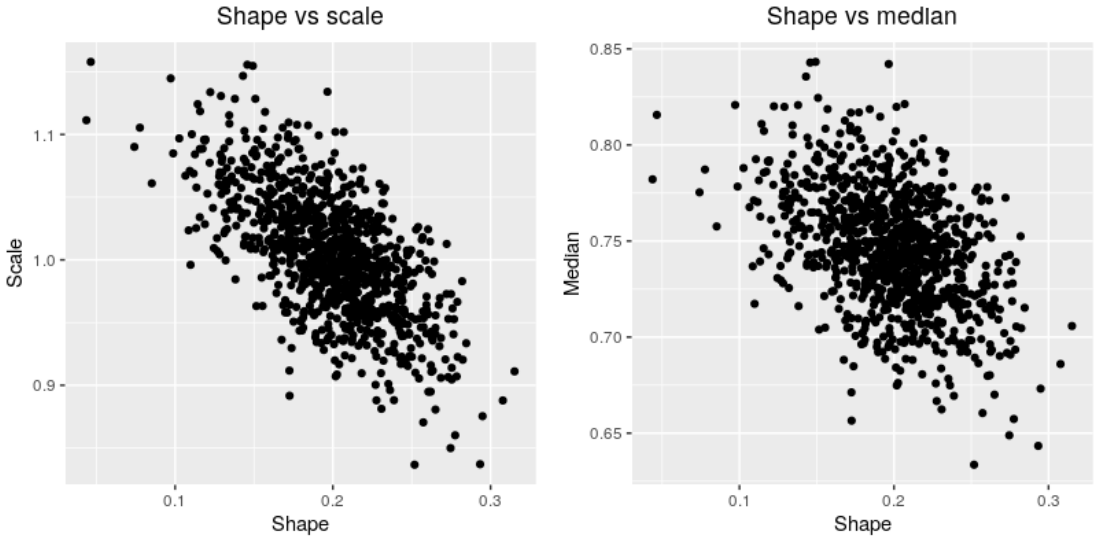


Figure 2.1: Each point is the result of an estimation of shape, scale and median on a simulated sample of 1000 iid GP distributed variables. To the left we see scale plotted against shape while to the right, median against shape. 1000 samples were generated, all with scale 1 and shape 0.2. This translates to a median of 0.75. The correlation coefficient for shape and scale is -0.65, and for shape and median -0.44.

In the same experiment, when parameterising with the median, $\kappa_{0.5}$, the correlation dropped to -0.44. This is still high, but significantly lower than for the scale parameter.

2.1.2 Threshold selection

When modelling threshold exceedances with the GP distribution, threshold selection is essential. Too low thresholds leads to sub-asymptotic behaviour which increases bias. Too high thresholds leads to fewer observations and thereby high variance. Threshold selection therefore involves a crucial bias-variance tradeoff.

There are two properties of the GP distribution that are used for threshold selection. Firstly, the derivation of the GP distribution for threshold exceedances is based on the assumption that above a certain threshold, the GEV distribution is a suitable approximation. If the GEV distribution is a valid approximation for values above a threshold u_0, it should also be so for any $u > u_0$. Therefore the shape parameter, ξ, which comes directly from the GEV distribution, should be constant for for thresholds $u > u_0$. This means that by fitting the model for different thresholds one might get some indication of where the threshold might lie by plotting the shape as a function of threshold and see where it seems to stay constant.

Another common approach is based on so called mean residual life plots. The mean of the GP distribution is $\sigma_u/(1-\xi)$ given that $\xi < 1$ (Coles, 2001). Let $X_1, ..., X_n$ be a series of random variables for which the GP distribution is valid above a threshold u_0. Then

$$E(X_i - u_0 | X_i > u_0) = \frac{\sigma_{u_0}}{1-\xi}.$$

Inserting (2.6), yields

$$E(X_i - u_0 | X_i > u_0) = \frac{\sigma + \xi(u_0 - \mu)}{1-\xi}.$$

If the GP distribution is valid for $X_i > u_0$, it must also be valid for $u > u_0$. In addition, as discussed earlier, ξ must be constant for $u > u_0$. This means that

$$E(X_i - u|X_i > u) = \frac{\sigma_u}{1-\xi} = \frac{\sigma + \xi u - \xi\mu + \xi u_0 - \xi u_0}{1-\xi} = \frac{\sigma_{u_0} + \xi(u-u_0)}{1-\xi}.$$

As ξ is constant for thresholds larger than u_0, this function is linear with respect to u. Another approach to threshold selection is therefore to plot the observed mean of $X_i - u$ for i st. $X_i > u$ for different values of u. Then the threshold can be chosen based on where this plot seems to turn linear with respect to u. These plots are referred to as *mean residual life* (MRL) plots.

These approaches both seem fine in theory, but as Coles (2001) points out, the nice constant or linear plots the theory predicts are rarely seen in practice. In reality they are usually hard to interpret with sudden changes of slope, especially for large thresholds. This means that both the shape parameter plots and the MRL plots to a large extent are subjective methods. These methods are still in practice the most commonly used to determine thresholds.

2.2 Gaussian random fields and covariance functions

A random field $X(\boldsymbol{s})$ on a set D can be defined as a function who's values are random variables for any $\boldsymbol{s} \in D$. Random fields are typically described in terms of their finite dimensional cumulative distribution,

$$F_{\boldsymbol{s}_1,...,\boldsymbol{s}_k}(x_1,...,x_k) = P\big(X(\boldsymbol{s}_1) \leq x_1, ..., X(\boldsymbol{s}_k) \leq x_k\big), \quad \boldsymbol{s}_1, ..., \boldsymbol{s}_k \in D.$$

A *Gaussian* random field is a random field for which the finite dimensional distribution is multivariate normal for any finite subset of D. This means that for any $k < \infty$ and any $\{\boldsymbol{s}_1, ..., \boldsymbol{s}_k\} \subset D$, $F_{\boldsymbol{s}_1,...,\boldsymbol{s}_k}$ is Gaussian.

As the normal distribution is fully described by its mean and covariance, the finite dimensional distribution is fully described by the functions

$$m(\boldsymbol{s}) = E\big(X(\boldsymbol{s})\big)$$
$$C(\boldsymbol{s},\boldsymbol{t}) = \mathrm{Cov}\big(X(\boldsymbol{s}), X(\boldsymbol{t})\big),$$

where $\boldsymbol{s}, \boldsymbol{t} \in D$. $m(\boldsymbol{s})$ can be any function of $\boldsymbol{s}$, but $C(\boldsymbol{s},\boldsymbol{t})$ has to lead to a symmetric semi positive definite covariance matrix. This is because for a multivariate normal vector $\boldsymbol{X}$ and a set of weights c, $\mathrm{Var}(c^T\boldsymbol{X}) = c^T\boldsymbol{\Sigma}c$, where $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{X}$. As a variance is always larger than or equal to zero, $\boldsymbol{\Sigma}$ must be semi positive definite.

If D is a vector space, a much studied group of Gaussian random fields are the stationary fields. Stationarity in the strict sense means that all finite-dimensional distributions are invariant to arbitrary translations. However, the weak formulation of stationarity only assumes translational invariance for the first and second order moments, i.e.

$$m(\boldsymbol{s}) = m, \qquad\qquad C(\boldsymbol{s},\boldsymbol{t}) = C(\boldsymbol{s}-\boldsymbol{t}).$$

This means that the covariance between any two points is only dependent on the distance vector between them (Abrahamsen (1997), Cressie and Wikle (2011)). The most basic example is the covariance function

$$C(\boldsymbol{s},\boldsymbol{t}) = \begin{cases} \sigma^2, & \boldsymbol{t} = \boldsymbol{s} \\ 0, & \boldsymbol{t} \neq \boldsymbol{s} \end{cases}.$$

This always gives the identity matrix as correlation matrix, meaning that any collection of points will be iid. This is referred to as a Gaussian white noise process and corresponds to a highly erratic field. This is rarely used when D is continuous, for example spatial models and continuous time models, but is often used in discrete cases.

If D is also a normed space, like $\mathbb{R}^d$ with the Euclidean norm, the set of isotropic random fields is an even more restricted class of random fields. Isotropic random fields are defined by a covariance function with the property

$$C(\boldsymbol{s},\boldsymbol{t}) = C(h), \quad h = ||\boldsymbol{s}-\boldsymbol{t}||.$$

This means that the covariance between two points depends only on the distance between them. The set of isotropic covariance functions contains many of the most commonly used covariance functions. An example is the Matérn covariance function on $\mathbb{R}^d$. Consider the stochastic partial differential equation

$$(\kappa^2 - \Delta)^{\alpha/2}\big(\tau X(\boldsymbol{s})\big) = \delta(\boldsymbol{s}), \quad \boldsymbol{s} \in D \subset \mathbb{R}^d, \tag{2.8}$$

where Δ is the Laplace operator, $\delta(\boldsymbol{s})$ is a Gaussian white noise process and κ, τ and α are parameters larger than zero. Whittle (1954) showed that the Gaussian random fields that arises as stationary solutions to this stochastic PDE has covariance functions on the form

$$C(\boldsymbol{s}, \boldsymbol{t}) = C(h) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{8\nu} h/r\right)^{\nu} K_\nu \left(\sqrt{8\nu} h/r\right), \quad h = ||\boldsymbol{s} - \boldsymbol{t}||. \tag{2.9}$$

Here, $\sigma^2 = 1/\tau$, $r = \sqrt{8\nu}/\kappa$ and $\nu = \alpha - d/2$, while K_ν is the modified Bessel function of the second kind. It turns out that $C(\boldsymbol{s}, \boldsymbol{s}) = \sigma^2$, so $\text{Var}\big(X(\boldsymbol{s})\big) = \sigma^2$ for any $\boldsymbol{s} \in D$. σ is therefore referred to as the standard deviation. r is referred to as the *range* and marks the distance at which the correlation between two points is approximately 0.1. ν controls the behavior of $C(h)$ as $h \to 0$. Increasing ν leads to higher differentiability of $C(h)$ at $h = 0$ which in turn leads to smoother random fields (Abrahamsen, 1997). It is therefore referred to as the *smoothness* parameter.

2.3 Latent Gaussian models

In this work we assume a Bayesian perspective on inference. This means that the parameters in distributions of random variables can themselves be interpreted as random variables with distributions. Let $Y_1, ..., Y_n$ be a sample of iid random variables from some distribution $p(y; \theta)$. Then assume that θ is a random variable with some distribution $p(\theta)$. This initial distribution is called the *prior* distribution and is used to encode prior knowledge about the given parameter. Inference on the parameters in a Bayesian framework revolves around the *posterior* distribution of θ i.e. $p(\theta|\boldsymbol{Y})$. If this function is known, all characteristics of the parameter like the mean, median, credible intervals etc., can be estimated. The posterior is found by Bayes rule

$$p(\theta|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\theta)p(\theta)}{p(\boldsymbol{Y})}. \tag{2.10}$$

In some rare simple cases, it is possible to find analytic expressions for the posterior, but in general this is unrealistic. Note that the denominator in (2.10) is independent of θ and therefore acts as a normalisation constant. This normalisation constant is often very hard to find as it involves solving very high dimensional integrals. A central component of Bayesian inference is therefore to find ways around this problem.

Hierarchical models are a wide class of models defined in three steps

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{x}, \boldsymbol{\theta} &\sim p(\boldsymbol{Y}|\boldsymbol{x}, \boldsymbol{\theta}) \\ \boldsymbol{x}|\boldsymbol{\theta} &\sim p(\boldsymbol{x}|\boldsymbol{\theta}) \\ \boldsymbol{\theta} &\sim p(\boldsymbol{\theta}). \end{aligned}$$

The first part of this structure is usually referred to as the data layer. This describes what kind of distribution the observed data comes from, and $p(\boldsymbol{Y}|\boldsymbol{x}, \boldsymbol{\theta})$ is often referred to as the likelihood. $\boldsymbol{x}$ is referred to as the latent field, and often describes some real world process that the data is related to. This latent process can describe something very concrete or more abstract depending on the specific problem. $p(\boldsymbol{x}|\boldsymbol{\theta})$ is the prior for the latent field. The final part, $\boldsymbol{\theta}$ is called the hyperparameters, and similarly $p(\boldsymbol{\theta})$ the hyperpriors. These incorporate our best knowledge of how the latent process and likelihood should behave. If much knowledge exists on the behaviour of the latent process, these should leave strict restrictions on what values the hyperparameters can take. This is referred to as an *informative* prior. If no such knowledge is available, the priors should be as little restrictive as possible, also called vague priors. For these kinds of hierarchical models, equation (2.10) takes the form

$$p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{Y})}$$

It is also common to assume that the data is conditionally independent given the latent field. This means that

$$p(\boldsymbol{Y}|\boldsymbol{x},\boldsymbol{\theta}) = \prod_{i=1}^{N} p(Y_i|\boldsymbol{x},\boldsymbol{\theta}),$$

where N is the total number of observations. All information about the dependence structure of the data is thus captured by the latent process. This simplifies the problem both theoretically and practically. It lets the distribution of each data point be a univariate distribution, which is useful as these are usually both better understood theoretically and easier to interpret than their multivariate counterparts. There is also usually a computational advantage to work with conditionally independent models.

LGMs are a commonly used group of models that fit within this framework. These are characterised by that

$$\boldsymbol{x}|\boldsymbol{\theta} \sim N(\boldsymbol{0}, Q(\boldsymbol{\theta})),$$

meaning that the latent field is a Gaussian random field. $Q(\boldsymbol{\theta})$ is here the precision matrix of the normal distribution, i.e. the inverse of the variance matrix.

Another important group of hierarchical models are the generalised additive models where each observation, Y_i, is connected to the latent field through a single parameter η_i. This parameter, called the linear predictor, can be expressed on the form

$$\eta_i = \mu + \boldsymbol{z}_i^T\boldsymbol{\beta} + \sum_k w_k f_{k,j_k(i)} + \varepsilon_i.$$

Here, μ is an intercept, $\boldsymbol{z}_i$ is a covariate vector for observation i and $\boldsymbol{\beta}$ is the coefficients of a linear effect of those covariates. $\boldsymbol{w}$ is a set of known constants and for each k, $\boldsymbol{f}_k$ represents the value some function that influences the latent process. $\boldsymbol{\varepsilon}$ is a noise term with mean zero. The latent field, $\boldsymbol{x}$, then consists of $\boldsymbol{x} = \{\mu, \boldsymbol{\beta}, \boldsymbol{\varepsilon}, \boldsymbol{f}_1, \boldsymbol{f}_2, ...\}$. In the context of this project, it is beneficial to have each observation depend on only one element of the latent field. We therefore let $\boldsymbol{x} = \{\boldsymbol{\eta}, \mu, \boldsymbol{\beta}, \boldsymbol{f}_1, \boldsymbol{f}_2, ...\}$. We assign a Gaussian prior to $\boldsymbol{x}$ so that we are in an additive LGM framework.

2.4 INLA

INLA stands for Integrated Nested Laplace Approximations and is a fairly new method for estimation of additive LGMs (Rue et al., 2009). INLA works by numerical integration of approximated distributions. The joint posterior distribution of an LGM can be expressed as

$$p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{Y}) \propto \exp\left\{\frac{1}{2}\boldsymbol{x}^T Q(\boldsymbol{\theta})\boldsymbol{x}\right\} \prod_i p(y_i|\eta_i,\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

However, the interest usually lies in the posterior marginals

$$p(\theta_i|\boldsymbol{Y}) = \int\int p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{Y})d\boldsymbol{x}d\boldsymbol{\theta}_{-i} = \int p(\boldsymbol{\theta}|\boldsymbol{Y})d\boldsymbol{\theta}_{-i} \tag{2.11}$$

$$p(x_i|\boldsymbol{Y}) = \int\int p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{Y})d\boldsymbol{x}_{-i}d\boldsymbol{\theta} = \int p(x_i|\boldsymbol{\theta},\boldsymbol{Y})p(\boldsymbol{\theta}|\boldsymbol{Y})d\boldsymbol{\theta}. \tag{2.12}$$

Here, $\boldsymbol{x}_{-i}$ is used for the vector $\boldsymbol{x}$ excluding element i. We see that these integrals are solvable without integrating over the latent field if we have approximations to $p(\boldsymbol{\theta}|\boldsymbol{Y})$ and $p(x_i|\boldsymbol{\theta},\boldsymbol{Y})$. This is beneficial because in common situations the space of hyperparameters is smaller than the latent field by several orders of magnitude. The idea behind INLA is therefore clever approximation of these distributions by Laplace approximations.

If we first look at the posterior distribution for the hyperparameters, this takes the form

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{Y})}{p(\boldsymbol{x}|\boldsymbol{Y},\boldsymbol{\theta})} \propto \frac{p(\boldsymbol{Y}|\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{Y},\boldsymbol{\theta})}.$$

The numerator here is the joint posterior. An expression proportional to this can be found directly from the model assumptions. In the denominator, the distribution of the latent field given the data and the hyperparameters is found. This is not trivial to calculate exactly. Because of the prior assumption of normality of the latent field, it is not unreasonable to assume that $p(\boldsymbol{x}|\boldsymbol{Y},\boldsymbol{\theta})$ is close to a multivariate normal distribution. The approach in INLA is therefore to approximate this distribution as a multivariate normal. For a given set of values, $\boldsymbol{\theta}^k$, for the hyperparameters, $p(\boldsymbol{\theta}|\boldsymbol{Y})$ can then be approximated as

$$\tilde{p}(\boldsymbol{\theta}^k|\boldsymbol{Y}) \propto \left.\frac{p(\boldsymbol{Y}|\boldsymbol{x},\boldsymbol{\theta}^k)p(\boldsymbol{x}|\boldsymbol{\theta}^k)p(\boldsymbol{\theta}^k)}{\tilde{p}_G(\boldsymbol{x}|\boldsymbol{Y},\boldsymbol{\theta}^k)}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta}^k)}, \tag{2.13}$$

where $\boldsymbol{x}^*(\boldsymbol{\theta}^k)$ is the mode of $p(\boldsymbol{x}|\boldsymbol{Y},\boldsymbol{\theta}^k)$, and $\tilde{p}_G(\boldsymbol{x}|\boldsymbol{Y},\boldsymbol{\theta}^k)$ is a Gaussian approximation of $p(\boldsymbol{x}|\boldsymbol{Y},\boldsymbol{\theta}^k)$ that matches its mode and the curvature at the mode. This is what is referred to as a Laplace approximation.

From this approximate posterior of $\boldsymbol{\theta}$, the mode $\boldsymbol{\theta}^*$ is calculated by numerical optimisation. Then, based on the curvature of this distribution at $\boldsymbol{\theta}^*$, representative values $\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_K$ for the hyperparameter vector is chosen. Integral (2.11) can then be approximated as

$$\tilde{p}(\theta_i|\boldsymbol{Y}) = \sum_k \tilde{p}(\theta_i,\boldsymbol{\theta}^k_{-i}|\boldsymbol{Y})\Delta_k,$$

where Δ_k is a factor that depends on the location of the support points. The way these points are chosen in practice depends on the dimension of the hyperparameter space. In low dimensions, a full grid is built around the mode in the hyperparameter space. This is very accurate, but the number of grid-points increases exponentially with the dimension of $\boldsymbol{\theta}$ when this approach is applied. Therefore cruder approximations to the hyperparameter space are used when $\dim(\boldsymbol{\theta}) > 3$.

Finding an approximation for $p(x_i|\boldsymbol{\theta},\boldsymbol{Y})$ is more involved, as the approximation has to be done for each element of latent field which can be very large. INLA has three possible solutions to this. The least computationally demanding is to use the marginal distribution of x_i from the Gaussian approximation $\tilde{p}_G(\boldsymbol{x}|\boldsymbol{Y},\boldsymbol{\theta})$. This is not very accurate, and often leads to significant approximation error (Rue et al., 2009). At the other end of the spectrum, the Laplace approximation could be estimated for each element of the latent field. This is very computationally demanding. A third option proposed by Rue et al. (2009), is to use Taylor expansions of the Laplace approximation at its mode to correct the Gaussian approximation for location and skewness. This approach, called simplified Laplace approximations, is relatively fast to compute and fairly accurate. It will therefore be used for the estimation in this project. (2.12) can then be approximated as

$$\tilde{p}(x_i|\boldsymbol{Y}) = \sum_k \tilde{p}(x_i|\boldsymbol{\theta}^k,\boldsymbol{Y})\tilde{p}(\boldsymbol{\theta}^k|\boldsymbol{Y})\Delta_k.$$

In order to evaluate $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{Y})$, a Cholesky decomposition of $Q(\boldsymbol{\theta})$ has to be performed. This turns out to be the most computationally expensive part of the approximation scheme, but it can be greatly sped up if Q is sparse. As a consequence of this, INLA is only efficient runtime wise when the latent field has a sparse precision matrix. The sparseness of the precision matrix turns out to be equivalent to the latent field having some form of Markov structure (Rue, 2005), i.e. that

$$p(x_i|\boldsymbol{x}_{-i}) = p\big(x_i|\{x_j : j \in N(i)\}\big),$$

where $N(i)$ represents some form of neighbourhood around i. In a temporal context, this means that an element corresponding to time i only depends on elements at neighbouring time stages. This class of Gaussian fields with a Markov structure are called Gaussian Markov Random Fields (GMRF). In order for INLA to be an efficient framework for estimating the posterior marginals, the latent field of the model must be a GMRF. Fortunately many of the most commonly used LGMs has GMRF latent fields. Examples of such latent fields include random walk processes, auto-regressive models and some spatial models.

2.5 Penalised complexity priors

In all Bayesian modelling, the choice of priors is central. This is especially true in cases where there is little data available and prior knowledge is scarce. Overfitting can often be an issue with regard to hyperpriors

for LGMs because they often assign zero prior density to lower complexity base models. A prior selection philosophy, proposed by Simpson, Rue, Riebler, Martins and Sørbye (2017), called penalised complexity priors attempts to combat this problem and create a robust framework for prior selection. The idea behind penalized complexity priors is that some distance measure from the model, f, to some lower complexity base model, f_0, is assigned a prior. The posterior distribution can then be shrunk towards this base model with a rate set by the parameters in the prior (Simpson et al., 2017). The distance measure they propose is defined as

$$d\{f, f_0\} = \sqrt{2 \cdot \mathrm{KLD}(f, f_0)}$$

where

$$\mathrm{KLD}(f, f_0) = \int f(x) \ln \frac{f(x)}{f_0(x)} dx$$

is the Kullbach-Leibler divergence. This measure has its origin in information theory and is often used to describe the loss of information from a true distribution to an approximate distribution. This means that if $\mathrm{KLD}(f, f_0)$ is large, then f_0 is a bad approximation of f in the sense that the information loss is large. By penalising deviations from f_0 by the Kullbach-Leibler distance, we thereby say that the model should stay close to the low complexity base model unless that would lead to a significant information loss.

Assume that f depends on some parameter θ that we want a prior for. Simpson et al. (2017) propose to set the prior for θ by penalising deviation from f_0 at a constant rate, λ. This means putting an exponential prior on the Kullbach-Leibler distance,

$$F\big(h(\theta)\big) = 1 - \exp\left[-\lambda \cdot h(\theta)\right], \quad h(\theta) = d\big\{f(x;\theta), f_0(x)\big\}. \tag{2.14}$$

This prior can then be transformed, through a change of variables, to a prior $p(\theta)$. The penalisation rate may be set by the user, which means that the method still remains somewhat subjective. However, it assures that if there exists a θ so that $f(x;\theta) = f_0(x)$, the case $h(\theta) = 0$ is given a positive prior density independently of λ. Also, the priors can be parameterised in a user-friendly way. This removes some of the haphazardness often associated with prior selection in hierarchical Bayesian modelling.

PC priors relevant for this project include a PC prior for the standard deviation of a Gaussian random effect, the range and standard deviation of a Matérn effect and for the tail parameter of a GP distribution. Assume that for some LGM, the linear predictor takes the form

$$\eta_i = \beta_0 + f(g_i), \tag{2.15}$$

where β_0 is an intercept, g is a discrete covariate and $\boldsymbol{f}$ is a Gaussian random effect, meaning that

$$\boldsymbol{f}(\boldsymbol{g}) \sim N_1(0, \sigma^2 R),$$

where R is a matrix with full rank. For the Gaussian random effect, the lower complexity base model is no dependence on covariate g. This can be interpreted as $\sigma = 0$ as this leads to $\eta_i = \beta_0$ for all i. The KLD from any distribution to a constant is undefined. We can therefore not compare a model with $\sigma > 0$ to a model with $\sigma = 0$ directly. However, Simpson et al. (2017) show that for a reference normal distribution $N_0(0, \sigma_0^2 R)$ with $\sigma_0 \ll \sigma$, $d(N_1, N_0) = \sqrt{p\sigma^2/\sigma_0^2}$, where p is the dimension of $\boldsymbol{f}(\boldsymbol{g})$. This means that $d(N_1, N_0) \propto \sigma$. Putting an exponential prior on the standard deviation σ of the Gaussian random effect is thus a way of penalising deviation from a model with much lower standard deviation, which is what we need. This leads to the PC prior

$$p(\sigma; \lambda) = \lambda \exp(-\lambda\sigma). \tag{2.16}$$

In this case we can set

$$\lambda = -\frac{\ln(p_\sigma)}{\sigma_0},$$

in order to let the prior probability of σ exceeding σ_0 be p_σ. Note that in contrast to for example a gamma prior for σ, the PC prior will always have a positive prior density for $\sigma = 0$.

If we instead let g_i from (2.15) be an element of $\mathbb{R}^d$, and $\boldsymbol{f}(\boldsymbol{g})$ have a Matérn covariance function (see section 2.2), the reference model is still a flat effect. However, a flat effect is obtained both by letting

$\sigma \to 0$ and $r \to \infty$. This means that a joint PC prior for σ and r is needed. Fuglstad, Simpson, Lindgren and Rue (2019) show that, with a fixed smoothness parameter, this joint prior takes the form

$$p(r, \sigma; \lambda_1, \lambda_2) = \frac{d}{2}\lambda_1\lambda_2\rho^{-\frac{d}{2}-1}\exp\left\{-\lambda_1\rho^{-\frac{d}{2}} - \lambda_2\sigma\right\}, \quad \sigma > 0,\, r > 0 \tag{2.17}$$

if $d \leq 3$. λ_1 and λ_2 are the rates by which the deviation from a flat effect should be punished with regard to r and σ respectively. (2.17) can easily be decomposed into one prior for the standard deviation and one for the range,

$$p(r; \lambda_1) = \frac{d}{2}\lambda_1\rho^{-\frac{d}{2}-1}\exp\left(-\lambda_1\rho^{-\frac{d}{2}}\right) \tag{2.18}$$

$$p(\sigma; \lambda_2) = \lambda_2\exp\left(-\lambda_2\sigma\right). \tag{2.19}$$

The PC prior for the Matérn standard deviation is thus the same as the PC prior for the standard deviation of a generic Gaussian random effect, as seen in (2.16). λ_1 can in this case be set as

$$\lambda_1 = -\ln(p_r)r_0^{d/2}$$

in order to get prior probability p_r of a shorter range than r_0.

The last relevant PC prior is the PC prior for a GP tail parameter. For threshold exceedance models (see section 2.1.1), penalisation of heavy tailed distributions can be desired. In that regard, Opitz et al. (2018) argue that an exponential distribution is a sensible low complexity reference model to compare GP distribution to. They show that under the assumption that $\xi > 0$, this leads to the following prior

$$p(\xi) = \sqrt{2}\lambda\exp\left(-\sqrt{2}\lambda\frac{\xi}{(1-\xi)^{1/2}}\right)\cdot\left(\frac{1-\xi/2}{(1-\xi)^{3/2}}\right), \quad \xi \in [0, 1), \tag{2.20}$$

where λ is the rate at which we punish deviation from an exponential distribution. We see from (2.20) that the PC prior is well defined only for $\xi < 1$. This is because the Kullback-Leibler divergence of the GP distribution relative to the exponential distribution, tends to infinity as $\xi \to 1$. This means that larger values for ξ will have a prior probability of zero, assuring that the GP distribution will have a well defined mean (see section 2.1).

Chapter 3

Data

The data set used in this project consists of hourly precipitation data from weather stations throughout the Oslofjord area. This chapter is split into two sections, the first giving an overview and the second focusing more on the extremes in particular.

3.1 Overview of the data

The data is provided by the Norwegian Meteorological Institute (MET). The placement of the weather stations can be seen in Figure 3.1. Geographically, the stations are placed between Brekke in Valer municipality to the south and Andelva in Eidsvoll municipality in the north, and between Bø in Midt-Telemark municipality in the west to Tangen in Aurskog-Høland municipality in the east. The stations are not uniformly distributed in that area. The urban areas have higher densities of weather stations. This is especially true in Oslo and Drammen, but more generally stations are placed where people live. This can be seen in Figure 3.1 as clusters of stations to the north-east (Oslo), center west (Drammen) and to the south with the densely populated areas around Skien, in Vestfold and in Østfold.

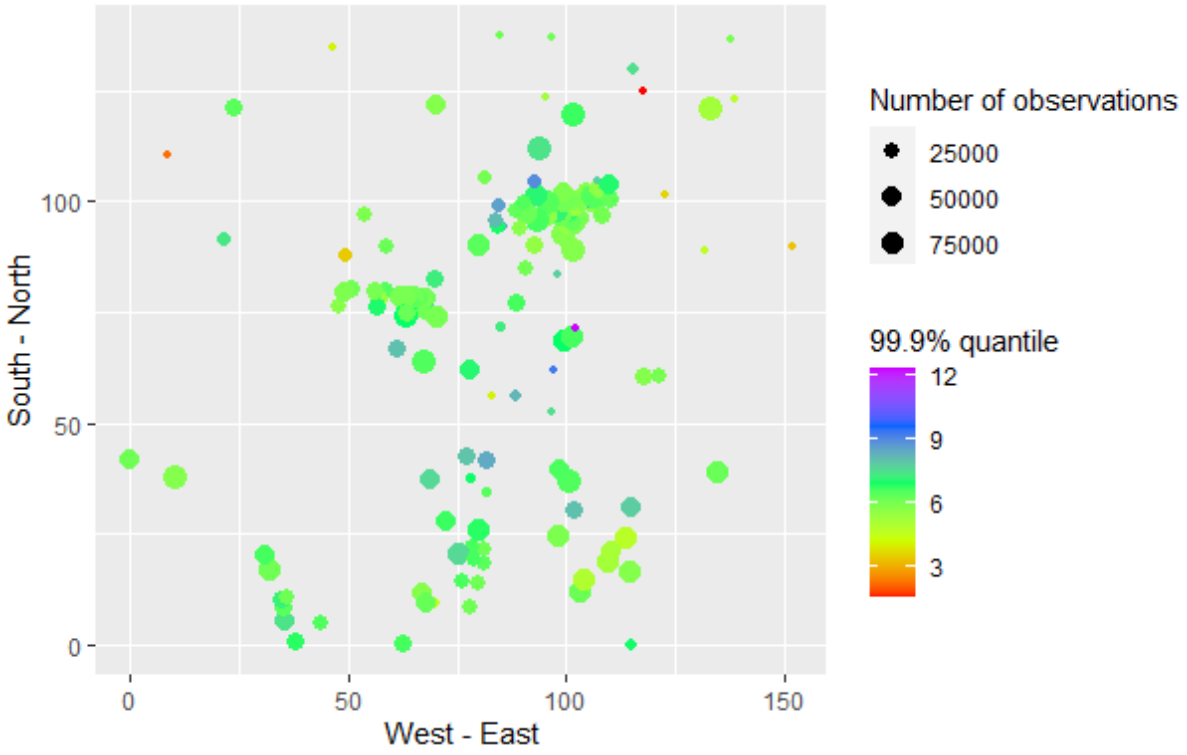


Figure 3.1: The placement of stations throughout the domain of interest. The colour of each dot marks its total mean precipitation and the size marks the number of observations at each station.

There are large variations in the amount of data available at each station. The longest running stations have data as far back as 1967, but this is only the case for very few of them. Only 20% of the stations started recording earlier than 2009, and the data quality was often poor prior to that. The main issue is that many stations, especially in Oslo, recorded missing values instead of zero precipitation. Based on the observations from later years, it is safe to assume that the data from this period also includes many actual missing values, but there is no way of telling for certain what is missing and what is zero. The data prior to 2009 will therefore not be used in this project.

The data recorded since 2009 also has some shortcomings that need to be discussed. These problems varied in character from registering zeros for several years to registering above 100 mm/h for several consecutive hours or again registering missing value instead of zero. Some examples of this can be seen in Figure 3.2. For all stations where such periods were discovered, that period is removed. Sometimes this meant discarding the station as a whole, while other times it meant removing only a few very extreme observation. This resulted in some stations with very short time series. Some have reliable data only from parts of 2019, while others have no observations at all in the relevant time span. The inclusion of stations with very short time series can lead to more noisy inference, especially when working with extremes. Therefore, all observations were grouped by station and calendar week (1 to 53), and if any station had more than three weeks without any observations, the station was discarded.

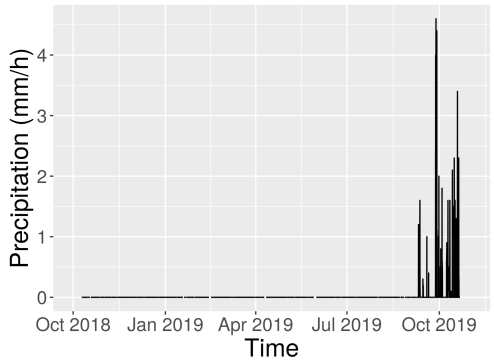
An excerpt of the data can be seen in Figure 3.3. Here we see that even when the shortest time series are removed, the number of observations differ widely between stations. Figure 3.4 shows the mean hourly precipitation for all months on the same subset of stations. It shows that even after filtering, there are still many months with improbable behaviour. A few months have mean hourly precipitation of above 1, which is very unlikely, while others have a mean of zero –meaning no precipitation for the whole month. An even more detailed filtering might therefore be justified, but it is hard to do so without risking the exclusion of important data.

The longer time series in Figure 3.3 and 3.4 give some indication as to what kind of precipitation patterns we expect to find. The data seems to have some seasonality. In Figure 3.3, this can be observed as spikes in the time series with semi-regular occurrences. These spikes are caused by convective precipitation during the Norwegian summers which leads to large precipitation intensities for short timespans. In Figure 3.4, it can be seen as a weak periodic structure, meaning that both the average and the extremes have this behaviour. Figure 3.5 shows that both the mean and the standard deviation of the positive precipitation depends on the time of year. Figure 3.6 also shows the hour to hour and year to year variations across all stations. Afternoon showers are a common phenomenon in Norway. One would therefore expect a tendency towards higher mean positive precipitation during the afternoons. However, Figure 3.6 shows no evidence of this in the data. The year to year variations are expected to be small compared to the seasonal variations when averaged across all stations. That is also what is observed, but relative deviations from the mean is sometimes as large as 15%. Taking into account the amount of data, these fluctuations are significant. This is also the case for the hour to hour variations.

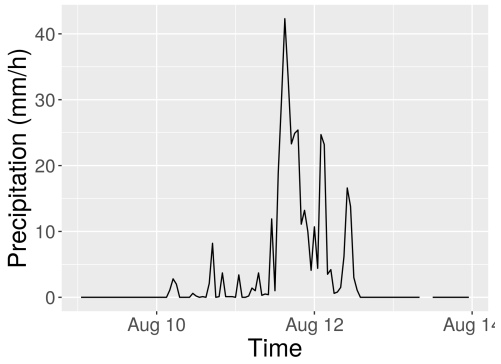
Figure 3.1 shows how the 99.9% quantile for precipitation depends on location. There seems to be some structured variations in extreme precipitation like a low mean to the south-east and a higher to the south-west, but there are too many small scale variations between nearby stations to infer any definitive spatial structure from the plot. Figure 3.7 shows the same quantiles plotted as a function of altitude. There is no apparent structure here apart from a slight increase for the largest altitudes. However there are few stations at altitudes above 200 m which leads to difficulties with estimating any such effect.

In Figure 3.8 we see the small scale behaviour of the data for a subset of the stations. The stations presented there are placed throughout the region, but still the precipitation occurs in clear clusters across space. In addition the precipitation registered at each station is clearly auto-correlated. If rain is observed during one hour, the chance of rain the consecutive hour is increased. The probability of observing zero is also affected by the previous precipitation amount to a large extent.

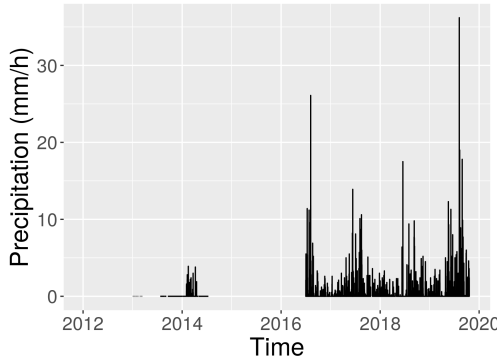
There are many negative values in the data set. Most of these seem to be measurement noise around zero. The proportion of negative observations depends on the station in question. Figure 3.9 shows the amount of negative observations for each station where negative observations occur. In most cases it lies well below 1%, but at some stations it is as high as 5%. A closer inspection of the time series reveals that



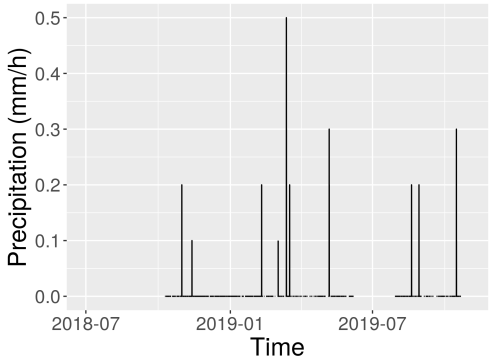
(a) Precipitation data from station SN17640. For almost a year, no precipitation was measured.



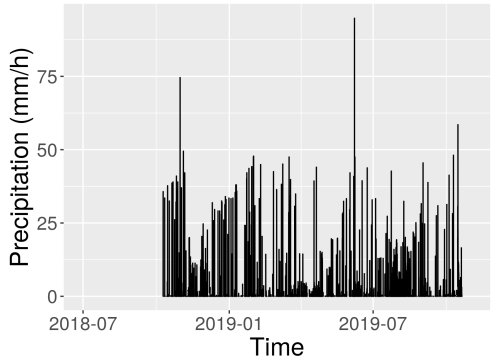
(b) Precipitation data from station SN17875 in the period 09.08 to 14.08 2019. During this short period the cumulative precipitation was comparable to a yearly total.



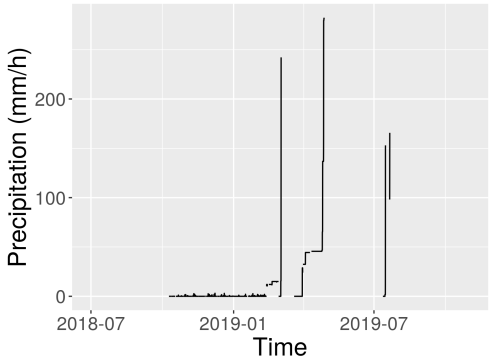
(c) Precipitation data from station SN19490. During 2013 and 2014 an over abundance of zeros were recorded.



(d) Precipitation data from station SN18265. The cumulative precipitation observed for a year of observations was less than 4 mm.



(e) Precipitation data from station SN18690. Precipitation amounts larger than the record precipitation in Oslo are scattered throughout the time series.



(f) Precipitation data from station SN19660. During the spring of 2019, there is a sudden shift from reasonable observations to observations far beyond the record precipitation.

Figure 3.2: Examples of problematic observations in the data set.

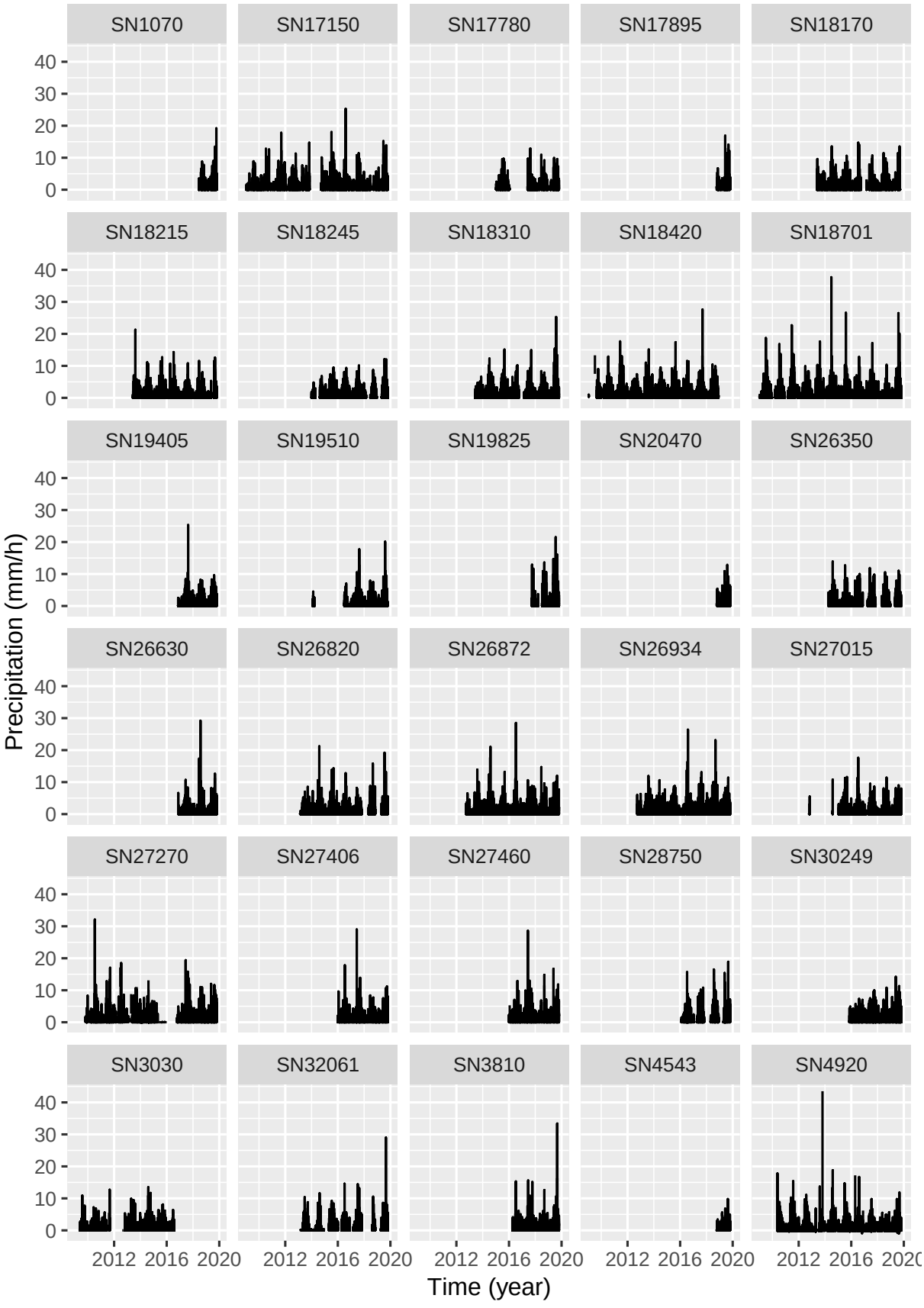


Figure 3.3: Display of the filtered data from 30 randomly selected stations.

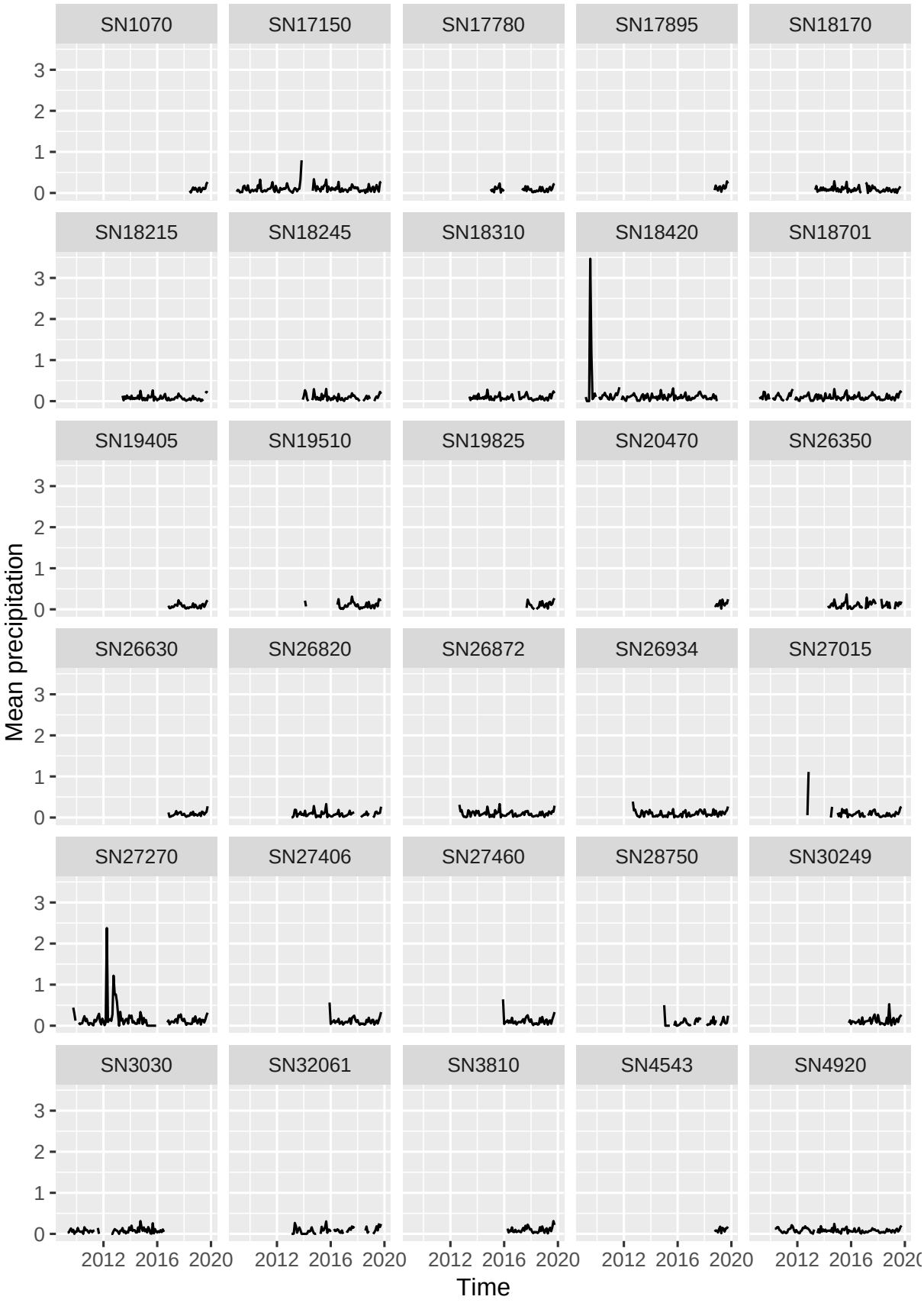


Figure 3.4: Mean hourly precipitation month by month for a subset of stations.

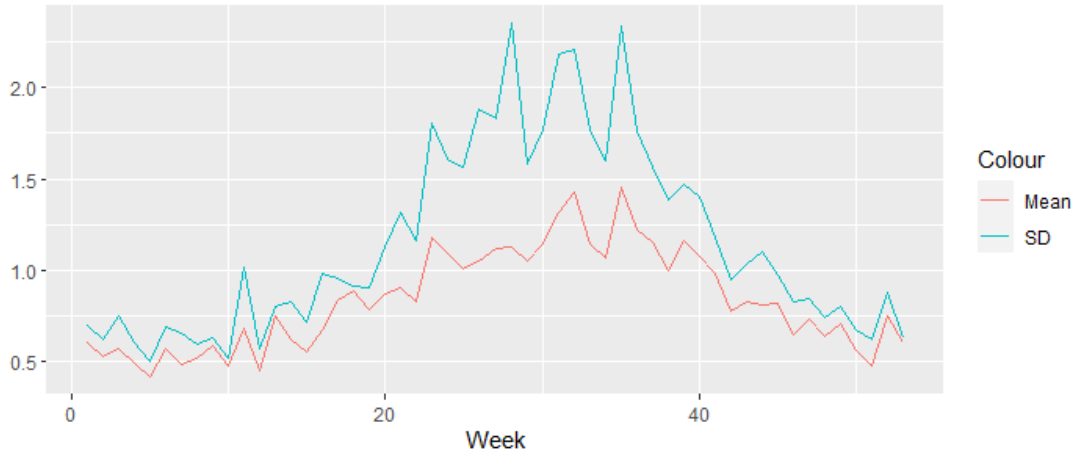


Figure 3.5: The mean and standard deviation of the positive precipitation observed each week across all stations.

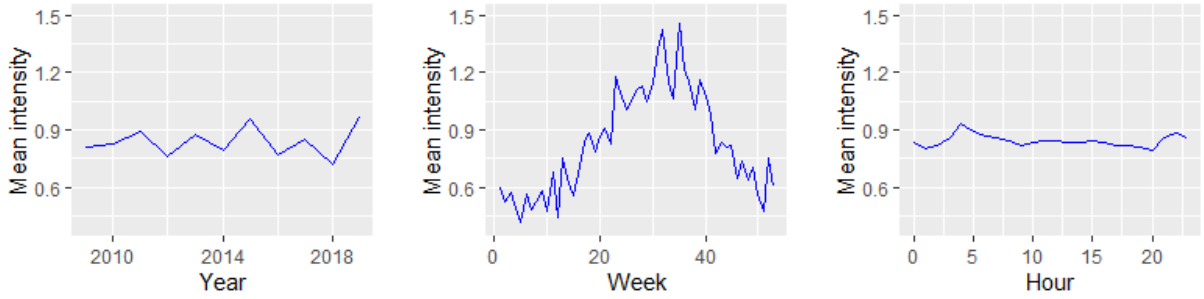


Figure 3.6: The mean positive precipitation plotted as function of different time components.

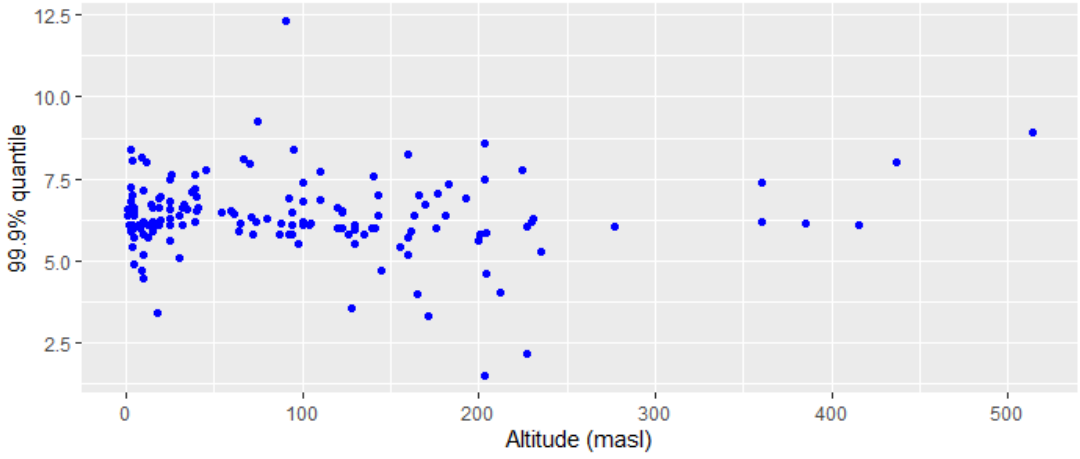


Figure 3.7: The 99.9% quantile in mm/h as a function of altitude in metres above sea level.

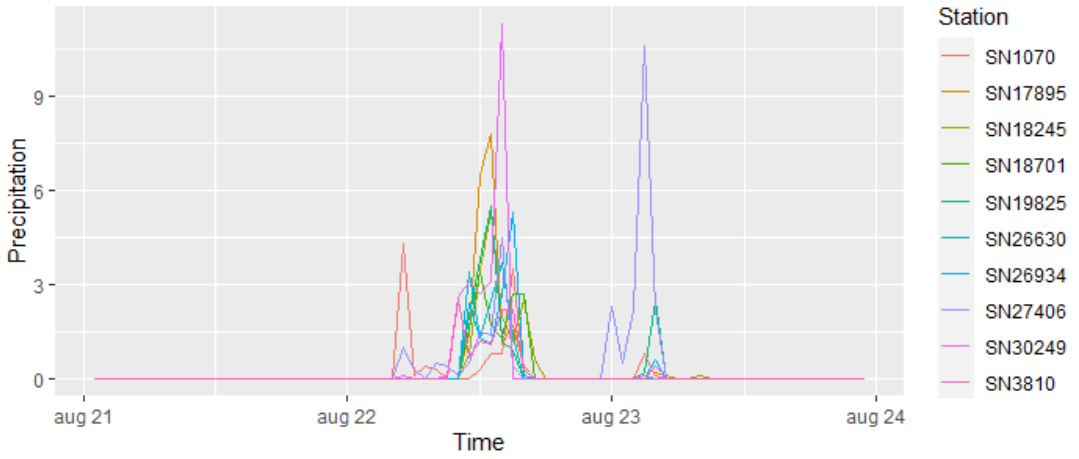


Figure 3.8: The precipitation on a subset of stations from across the region for a few days in August 2019.

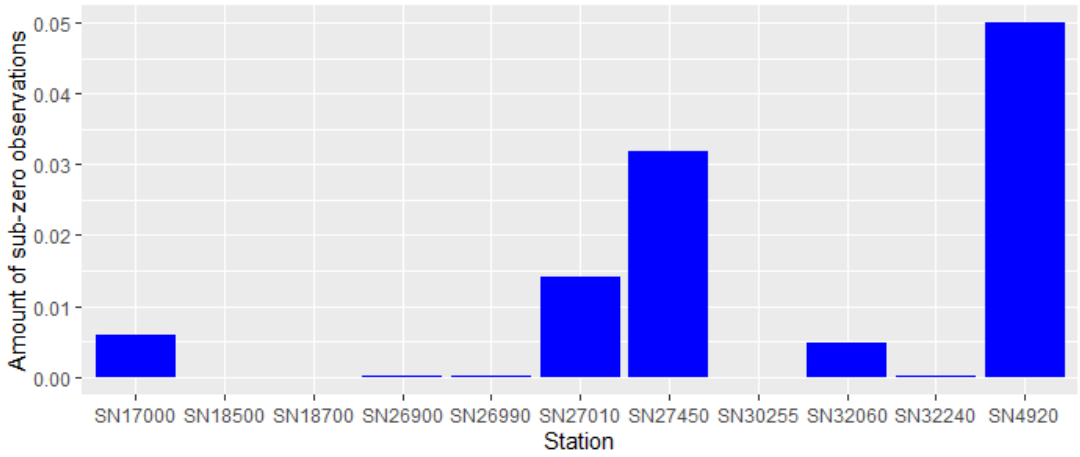


Figure 3.9: Bar chart displaying the amount of negative observations at all stations where such observations occur.

in most cases, the negative observations are measurement noise around zero. In this project, the negative observations will therefore be treated as zeros. The noise that causes these observations usually leads to an equal amount of small positive observations. These should also be treated as zeros. Simply removing the negative observations might therefore lead to inflation of low precipitation observations. This can lead to bias at the stations with very noisy observations. However the measurement errors are assumed to have little relative impact on the extremes, and as that is the focus of this thesis, any inflation of low values is assumed to be handled by the model. Apart from the few stations where the amount exceeds 1% it probably has little effect on the results.

Each station is owned by some institution, and the quantiles seems to depend on which institution records the data. Some of this variability can be attributed to the spatial structure of the data. Many of the stations are owned by local municipalities, and are thus confined to smaller sub-domains with similar precipitation patterns. Of all the station owners, Statens Vegvesen, the institution responsible for the Norwegian road network, is the clearest outlier. The variance of the observed 99.9% quantile between their stations is much larger than for any other institution. Figure 3.10 shows that most outliers, both high and low, are observed at their stations. In their defence, all of their stations have been recording for only a year, and they have larger geographic spread than most other station owners. However, both MET and NIBIO (Norwegian Institute of Bioeconomy Research) have similar geographic spreads and much lower variance. Also, the other stations with time series of similar length have lower variance than those owned

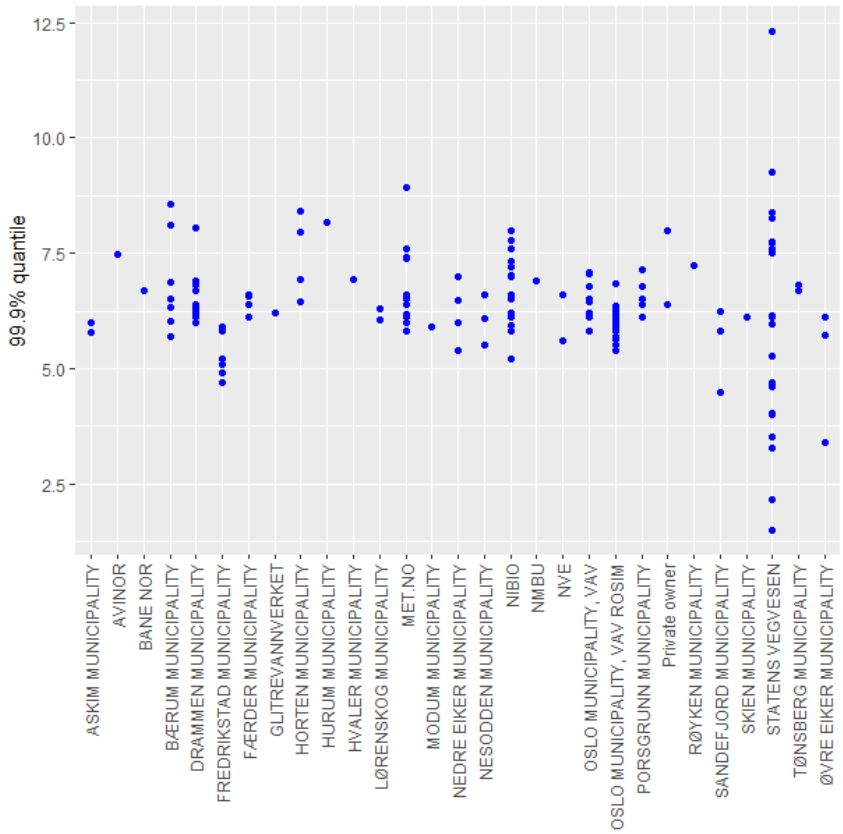


Figure 3.10: The 99.9% quantile for the different stations grouped by station owner.

by Statens Vegvesen. Our best guess is that this is a calibration issue. Zooming in on the time series from these stations reveals that they record precipitation at very sensible times, but that the amounts are consistently either higher or lower than expected depending on the station. We do not have access to information about the type of pluviometer used at the different stations, but the data suggests that Statens Vegvesen is using instruments prone to calibration issues.

3.2 Initial analysis of extremes

In order to see the scale of the 99.998% quantile, maps of the monthly maxima at each station are shown in Figure 3.11. Note that these maxima are lower on average than the 99.998% quantile as the stations with the longest time series only have about 8000 observations for any given month. Still this gives us some impression of how the extremes tend to vary in time and space. It shows that there is very clear temporal dependence in the extremes. In general, July and August has max values around 20 mm/h for any time series of significant length, while in January and February the maxima are well below 10 mm/h. The spatial structure is much harder to infer. The plot also shows one of the problems with the data. Some of the most extreme observations occur at very unlikely times and places. The clearest example is a point to the upper right which shows that a short running station observed about 50 mm/h in March. This is far beyond any other observations in the same period. Something similar can be said about a very high observation at the neighbouring station in October. These kinds of observations might lead to overestimation of the tail behaviour.

Mean residual life (MRL) plots are a useful tool for data exploration in extreme value statistics (see section 2.1.2). In order to draw conclusions from MRL plots, the underlying data has to be independent. Figure 3.8 clearly shows that an assumption of independence does not hold. It shows both that the observations at any single station are correlated in time, and that rain events on any given station increases the probability of rain on any other station. Since the data is clustered in space and time, some form of declustering should be performed in order to assume independent observations.

With regard to temporal clusters, the approach is to choose the maximal observed precipitation within any given extreme rain event. This is a common approach when dealing with threshold exceedances and time series data (Coles, 2001). Assuming that any observation larger than some threshold u is extreme, we define an extreme rain event as a group of above threshold observations where consecutive extreme observations are less than six hours apart. This means that if there is more than six hours from an above threshold observation to the next, the two belong to different extreme events. If not, they belong to the same. Other studies have used dry periods of similar length to separate rain events, see for example De Michele (2003). The declustering was done by extracting the maximum value from all the extreme events for any given station. This will lead to data that is approximately independent for any one station. As seen in Figure 3.8, the spatial structure of the data also causes problems for the assumption of independence. However, there are no widely accepted methods for spatial declustering.

It is of interest to know what kind of precipitation amounts we can expect to be well modelled with extreme value statistics. Here we have looked at MRL plots for a sense of this. As mentioned, threshold selection by MRL plots assumes that the data is independent. Therefore, a declustering as described in the previous paragraph is performed. This declustering has to be done for each threshold when creating the MRL plots. This means that lowering the threshold does not always lead to more observations because it can lead to a larger number of observations being deemed part of the same cluster. With regard to the spatial clusters, the solution used is to make separate MRL plots for each station.

The MRL plots for a subset of stations can be seen in Figure 3.12. According to the results presented in section 2.1.2, we should see a linear relationship for threshold values larger than some u_0. As mentioned in the previous paragraph, the declustering leads to fewer available observations, especially for lower thresholds. The variance is therefore quite high, even for lower thresholds. As is often the case with MRL plots, it is hard to draw any general conclusions from the plots. Some, like SN18420 and SN27270, have positive trends that are close to linear even from quite low thresholds. Others, like SN17780 and SN26350, have trends that go from positive to negative at thresholds of around 5 mm/h, which means that linearity should not be assumed for any lower threshold than 5 mm/h. Station SN32061 even has a large jump for

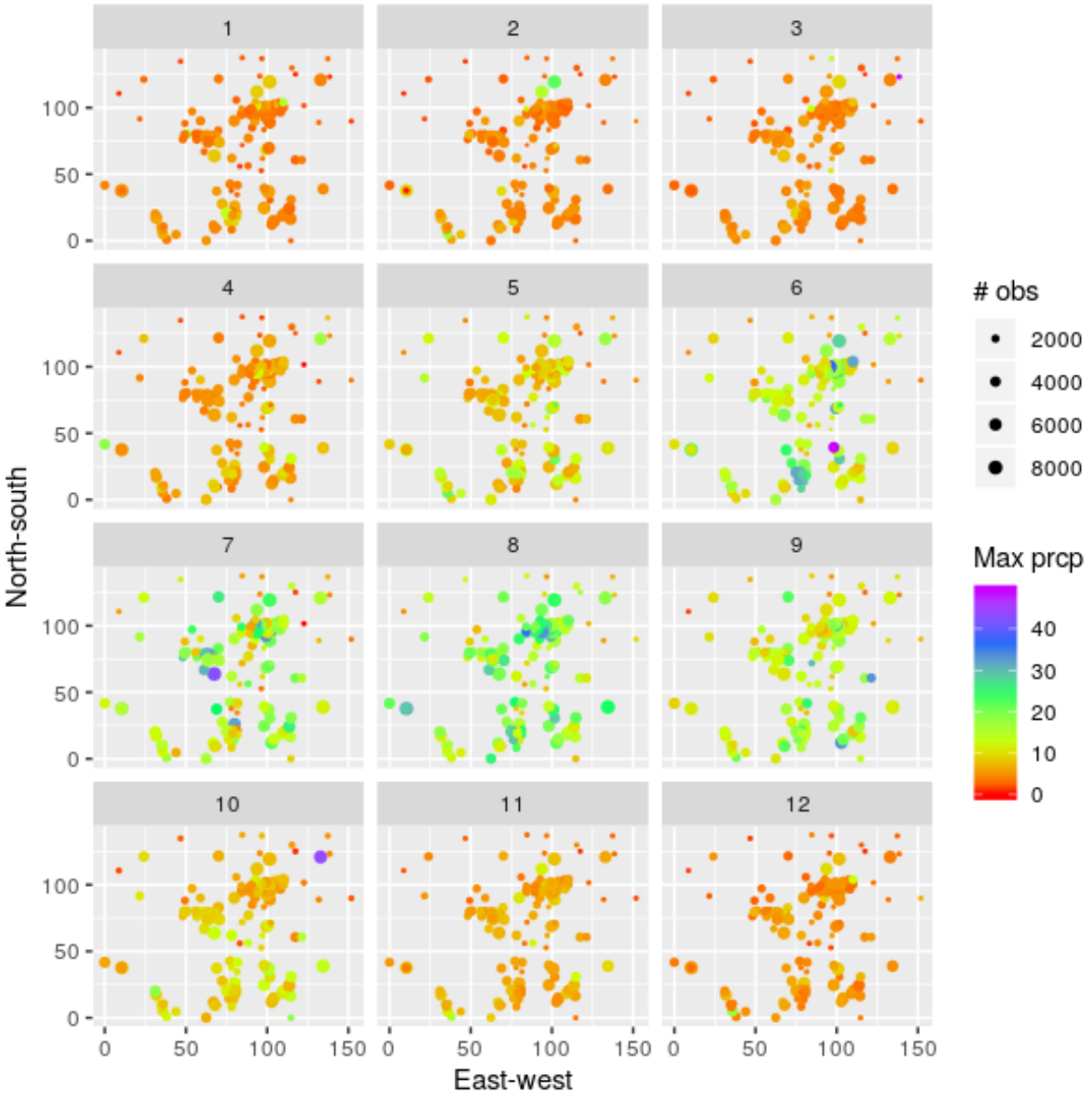


Figure 3.11: The monthly observed maxima at each station plotted in space. Each panel marks one month. The colour corresponds to max value, and the size represents the number of observations.

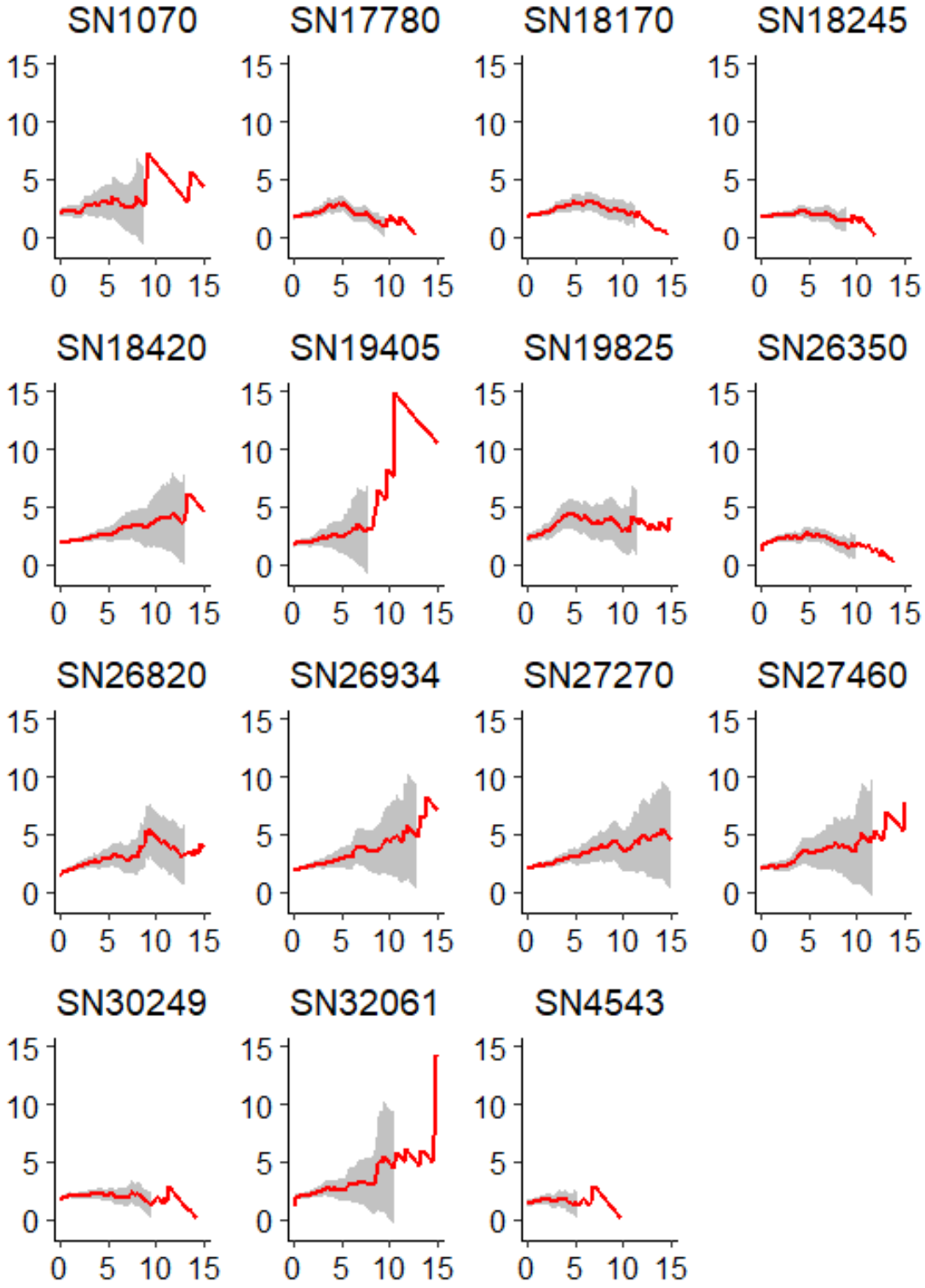


Figure 3.12: Mean residual life plots for a subset of stations. The x-axis marks threshold and y-axis mean residual life. The declustering routine used took the max within any group of threshold exceedances less than six hours apart. Uncertainty is not plotted when there are less than six exceedances.

$u \approx 8$ meaning that there is evidence of nonlinearity for as high thresholds as that. It is generally hard to justify any fixed threshold for this data set based on the MRL plots for each station. Based on the station of interest, thresholds in the range from $u = 2$ to $u = 8$ could be supported. For the model used in this project, u is allowed to vary in time and space. We expect to find thresholds in the range suggested by the MRL plots.

Chapter 4

Model and estimation

In this project the goal is to fit a spatio temporal model to hourly precipitation data within the INLA framework in order to estimate extreme quantiles. The model used is similar to Opitz et al. (2018). First, the general structure of the model is presented, followed by a discussion of the priors used. Then the estimation procedure is presented before a discussion of the model assumptions.

4.1 Model

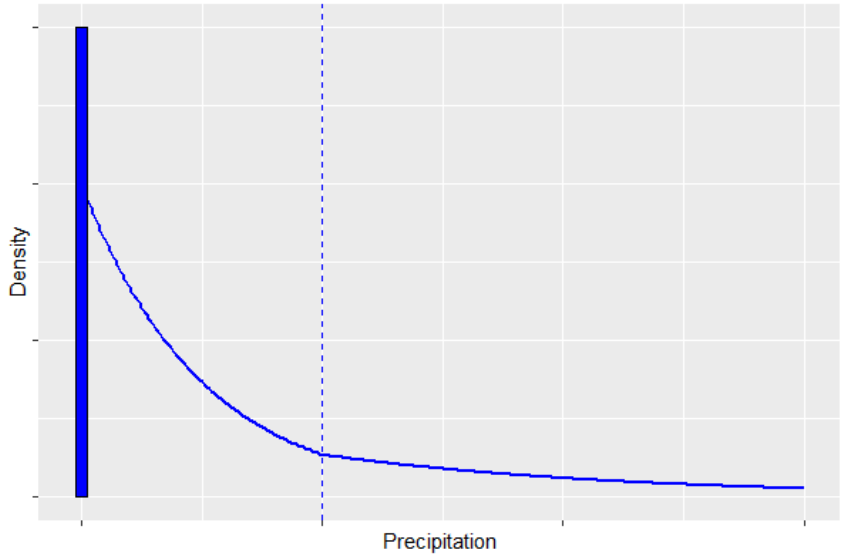


Figure 4.1: An illustration of a precipitation distribution. The column to the left marks a pointmass at zero. The dashed line marks the threshold, u, at which the distribution switches from a gamma to a GP distribution.

In statistical modelling of precipitation, the most commonly used distribution is a mixture of a gamma and a point mass at zero. See Martinez-Villalobos and Neelin (2019) and the references therein for details. This means that the likelihood has two elements, the probability of observing positive precipitation and a gamma distribution for the positive precipitation. However, as Martinez-Villalobos and Neelin (2019) point out, some of the assumptions they use to justify the use of a gamma distribution are invalid for sub-daily time intervals. In practice, the exponential decay of the gamma distribution's tail is often too light for hourly precipitation data. A GP distribution fitted to the extreme observations is therefore usually in order. This leads to a spliced model with a point mass at zero, a gamma distribution for precipitation amounts below some threshold u and a GP distribution for precipitation amounts above u. This is illustrated in Figure 4.1. A central question for such models is then the definition of the threshold u above which the GP distribution holds. It can be argued that the threshold should be allowed to vary in time and space (Opitz et al., 2018). In this work, the bulk of the distribution, i.e. the gamma part of the likelihood, is used to set u as a function of both the spatial and temporal domain.

The model is therefore built in three stages. First assume that the positive precipitation $Y_{s,t}$ at time t and location s follows a gamma distribution,

$$Y_{s,t}|Y_{s,t} > 0 \sim \text{Gamma}(\mu_{s,t}, k), \quad t = 1, ..., T, \quad s = 1, ..., S$$

where $\mu_{s,t}$ is the mean at time t and location s, k is a common shape parameter and S and T are the total number of locations and times respectively. Then define a threshold, $u_{s,t}$, as some function of $\mu_{s,t}$ and k. The second stage models the probability of exceeding $u_{s,t}$. Let

$$I(Y_{s,t} > u_{s,t}) \sim \text{Binomial}(1, p_{s,t}).$$

I is here an indicator function, and $p_{s,t}$ is the probability of exceeding the threshold. At this stage the zeros are reintroduced as the probability of observing zero precipitation may have spatial and temporal structure. The last stage is the GP distribution for the threshold exceedances, meaning

$$Y_{s,t} - u_{s,t}|Y_{s,t} > u_{s,t} \sim \text{GP}(\kappa_{s,t}^{(q)}, \xi).$$

Here, $\kappa_{s,t}^{(q)}$ is the q-quantile of the GP distribution at time t and location s. Based on equation (2.7), the α-quantile of the full distribution at time t and location s, $y_{s,t}^{(\alpha)}$, can be expressed as

$$y_{s,t}^{(\alpha)} = u_{s,t} + \kappa_{s,t}^{(q)} \frac{\left(\frac{1-\alpha}{p_{s,t}}\right)^{-\xi} - 1}{(1-q)^{-\xi} - 1}, \tag{4.1}$$

given that $p_{s,t} \geq 1 - \alpha$.

Each of the three likelihoods each constitute the first stage of an LGM as described in section 2.3. The linear predictor for each likelihood is connected with a separate latent field as follows:

$$\begin{aligned} \ln(\mu_{s,t}) &= \beta_0^{\text{gam}} + \tilde{f}_{s,t}^{\text{gam}}(s, w_t) + \tilde{f}_s^{\text{gam}}(s) \\ \text{logit}(p_{s,t}) &= \beta_0^{\text{bin}} + \tilde{f}_t^{\text{bin}}(w_t) + \tilde{f}_s^{\text{bin}}(s) \\ \ln(\kappa_{s,t}^{(0.5)}) &= \ln(\mu_{s,t}) + \beta_0^{\text{GP}} + \tilde{f}_t^{\text{GP}}(w_t) + \tilde{f}_s^{\text{GP}}(s), \end{aligned} \tag{4.2}$$

where β_0 in each case is an intercept, $\tilde{f}_t(w_t)$ is a random effect associated with the calendar week w_t at time t and $\tilde{f}_s(s)$ is an effect associated with the location s. The element $\tilde{f}_{s,t}(s, w_t)$ in the gamma stage denotes a space-time interaction term. This is included to allow for latent field temporal structures that vary in space. This term has high dimensionality, and it is therefore not included in the latent field for model stages 2 and 3 where there is less data. Note that the log mean from the gamma distribution is included as an offset in the GP stage. This is because the extremes are assumed to generally follow the same patterns as the bulk of the distribution. Including this offset therefore lets the GP stage borrow information from the bulk of the observations.

The spatial effect in all three model stages takes the form

$$\tilde{f}_s(s) = f_s(s) + g_s(s),$$

where $f_{\text{s}}(s)$ is a structured effect while $g_{\text{s}}(s)$ is an unstructured effect. Let $\boldsymbol{s} = \{s_1, s_2, ..., s_S\}$ be the station locations. The structured effect is then defined as

$$\boldsymbol{f}_s \sim N\left(\mathbf{0}, Q^{-1}(\boldsymbol{s})\right),$$

where $Q(\boldsymbol{s})$ is a precision matrix determined by the Matérn covariance function as described in equation (2.9). The unstructured effect, $g_{\text{s}}(s)$, is assigned an iid normal prior meaning that

$$\boldsymbol{g}_s \sim N\left(0, \sigma_{\text{IID}}^2 I_S\right).$$

The temporal effects for the binomial and GP stages are also decomposed into a structured and unstructured part. The structured component is assigned a second order random walk prior. If the process

$\boldsymbol{f} = \{f(1), ..., f(M)\}$ is a second order random walk, its second order increments are normally distributed, i.e.

$$f(i-1) - 2f(i) + f(i+1) \sim N(0, \sigma^2_{\text{RW}}), \quad i = 2, ..., M-1 \tag{4.3}$$

where σ_{RW} is the standard deviation on the normal distribution. In the case of a week to week process, it is also natural to assume that it is cyclic, meaning that (4.3) also holds at the endpoints $i = 1, M$ by letting $f(0) = f(M)$ and $f(M+1) = f(1)$. The unstructured component is also in these cases assigned an iid normal prior.

The space-time interaction term in the gamma stage is assumed to be the sum of two periodic functions,

$$\tilde{f}_{s,t}(s, w_t) = f_{\sin}(s) \cdot \sin(2\pi w_t/53) + f_{\cos}(s) \cdot \cos(2\pi w_t/53),$$

where $\boldsymbol{f}_{\sin}$ and $\boldsymbol{f}_{\cos}$ are assigned Matérn priors. The motivation for this is that the mean positive precipitation in general seems to have a temporal structure that is similar to a periodic function with a period of a year (see Figure 3.5). However, there is reason to believe that there are differences in temporal structure between stations. These variations are assumed to be smooth in space, and the coefficients are therefore assigned Matérn priors. This model remains quite simple, but it is able to capture variations in temporal structure between spatial locations such as differences in magnitude and mode. The tradeoff is that it comes with much stronger assumptions on the general shape of the temporal variations than a random walk effect does. As $u_{s,t}$ is not known until the first stage is fitted, prior knowledge about the temporal structure in model stage 2 and 3 is not available. This is part of the reason why the temporal effects in those stages are assigned random walk priors. Note that the differences in temporal dependence across space are available to the GP stage due to the inclusion of the gamma log-mean as an additive offset to the latent field. There is no unstructured effect included in the space time interaction effect.

All model stages also include an intercept. The prior for these is set to a zero mean normal distribution. For the binomial and GP stages, the precision is set to zero, leading to an improper prior. Such an improper prior for the intercept lead to numerical issues for the gamma stage, and, therefore, a precision of 1 is used. This is assumed to have little influence on the model fit.

In summary this means that when (4.2) is written out, the latent fields are

$$\begin{aligned} \ln(\mu_{s,t}) =& \beta_0^{\text{gam}} + f_s^{\text{gam}}(s) + g_s^{\text{gam}}(s) + \\ & f_{\sin}^{\text{gam}}(s) \cdot \sin(2\pi w_t/53) + f_{\cos}^{\text{gam}}(s) \cdot \cos(2\pi w_t/53) \end{aligned} \tag{4.4}$$

$$\text{logit}(p_{s,t}) = \beta_0^{\text{bin}} + f_t^{\text{bin}}(w_t) + g_t^{\text{bin}}(w_t) + f_s^{\text{bin}}(s) + g_s^{\text{bin}}(s) \tag{4.5}$$

$$\ln\left(\kappa_{s,t}^{(0.5)}\right) = \ln(\mu_{s,t}) + \beta_0^{\text{GP}} + f_t^{\text{GP}}(w_t) + g_t^{\text{GP}}(w_t) + f_s^{\text{GP}}(s) + g_s^{\text{GP}}(s) \tag{4.6}$$

Here, $f_s(s)$, $f_{\cos}(s)$ and $f_{\sin}(s)$ are structured spatial effects with Matérn covariance functions, $f_t(w_t)$ marks second order random walk effects of calendar week, and $g_s(s)$ and $g_t(w_t)$ are iid effects of location and calendar week respectively.

Note that this model demands a method for setting the thresholds, $u_{s,t}$, as a function of the gamma stage parameters. The approach used by Opitz et al. (2018) is to fix a probability p^+ and let $u_{s,t}$ be the p^+-quantile of the estimated gamma distribution,

$$u_{s,t} = F^{-1}(p^+; \mu_{s,t}, k), \tag{4.7}$$

where F^{-1} is the inverse of the estimated cumulative gamma distribution. This will be referred to as the gamma quantile (GQ) approach. Its main drawback is that it assumes that the gamma distribution with the chosen parameterisation is a good approximation of the true distribution around the true value for u, i.e. the point at which the GP distribution becomes a good approximation. We therefore propose the more naive approach of setting the threshold as a fixed multiple of the estimated gamma mean,

$$u_{s,t} = c \cdot \mu_{s,t}. \tag{4.8}$$

This approach will be referred to as the multiple of gamma mean (MGM) approach. Both methods will be explored, and a comparison is found in section 5.1.

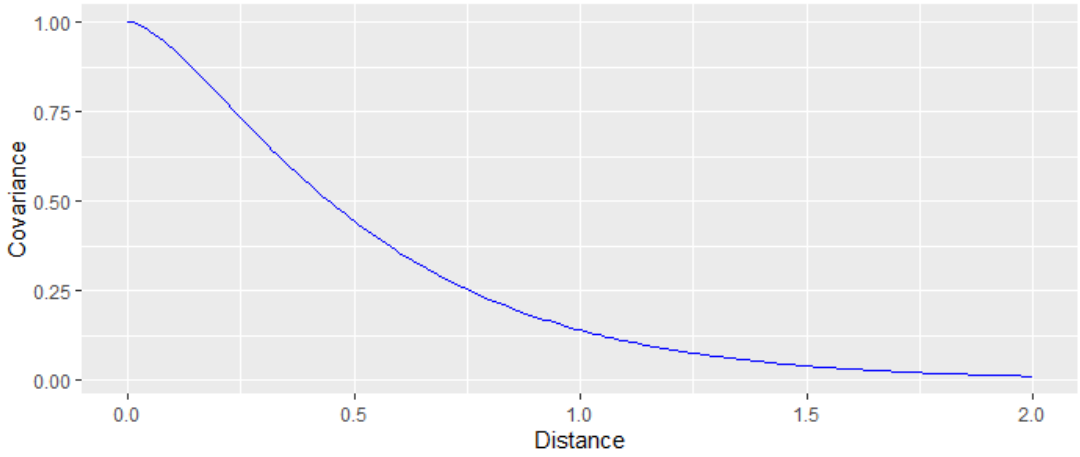


Figure 4.2: The Matérn covariance function with parameters $\sigma = 1$, $r = 1$ and $\nu = 1$.

4.2 Hyperpriors

The hyperparameters of the model stages include the range and standard deviations and roughness parameter of all spatial components, the random walk standard deviations and the iid effect standard deviations. In addition there are two likelihood parameters in total: the gamma shape, k, and the GP tail parameter ξ. Treating the roughness parameter, ν, as a hyperparameter leads to a poorly behaved likelihood. There is also usually little information about it in the data, and as pointed out by Rasmussen (2006), it is hard to justify high values for ν unless prior knowledge about the higher derivatives of the random fields sample paths exists. It will therefore be treated as a fixed parameter with value $\nu = 1$. This value is also used by Opitz et al. (2018), and it has been suggested as the natural value for $\mathbb{R}^2$ (Lindgren & Rue, 2015). An illustration of a Matérn covariance function with $\nu = 1$ can be seen in Figure 4.2.

There are then eight hyperparameters for the gamma stage, five for the binomial stage and six for the GP stage. For all the standard deviations, a PC prior is used (see equation (2.16)). The inclusion of the unstructured effects is assumed to ensure quite smooth structured effects and a quite vague prior with parameters $(p_\sigma, \sigma_0) = (0.1, 0.5)$ is chosen for all model components in order to ensure sufficient flexibility. This goes for all standard deviations in random walk processes, Matérn effects and iid effects. PC priors are also used for all Matérn range parameters (see equation (2.18)). There is no prior information to indicate different ranges for the different spatial components, so the same reference range and probability of $(p_r, r_0) = (0.1, 100)$ is used for all Matérn ranges. 100 km is close to the radius of the domain, which means that we assume some correlation across the domain unless there is evidence in the data for shorter range parameters.

Previous research on the value of ξ in Norway seems to indicate that a value in the range from 0 to 0.5 is reasonable for south-eastern Norway. Dyrrdal et al. (2015) work with annual maxima across Norway and conclude that the 95% credible interval ranged from -0.65 to 0.87 with a mean at 0.11. In their analysis of yearly maxima at SN18701 located at Blindern in Oslo, they observe that a model with a fixed shape parameter of 0.15 gave good results and that the maximum likelihood estimate of ξ for that data was 0.17. Dyrrdal, Skaugen, Stordal and Førland (2016) work with daily areal rainfall, but their analysis indicate that positive values for ξ is the norm for the continental inland, and their estimated values ranged from 0 to 0.3 for south-eastern Norway. Also note that the use of INLA demands a well behaved likelihood. As negative values for ξ leads to a change of domain, the prior probability of a negative shape parameter is set to 0. This means that a PC prior for ξ as shown in equation (2.20) with penalisation rate $\lambda = 4.5$ is a suitable choice. This corresponds to a prior probability of 0.04 for $\xi < 0.4$, which means that there is a low prior probability of a GP likelihood with infinite variance. The shape of the gamma distribution is assigned a gamma prior with shape 1 and rate 0.01.

4.3 Model fitting

The three model stages are fitted individually with the `r-INLA` package available in `R`. A joint estimation of the three stages is not possible due to the fact that the threshold $u_{s,t}$ determines the data for the binomial and GP stage. The Matérn covariance function does not give rise to a GMRF, and therefore leads to a dense precision matrix. This would be a problem if the goal was prediction at many unobserved locations, but for this thesis the main interest is strengthened inference on observation sites with short time series. This, in conjunction with the relatively low number of 148 stations, means that INLA is able to handle the dense precision matrices with high enough efficiency. This leads to an estimation process that takes between one and two hours on eight cores, the vast majority of which is spent on the gamma stage.

As discussed in section 2.1.2, the choice of u involves a crucial bias-variance tradeoff. This manifests in finding suitable values for p^+ and c from equations (4.7) and (4.8). In the case of both p^+ and c, values that are too low lead to sub-asymptotic bias, whilst values that are too high result in little data and high variance. For this project, these parameters are set by comparisons of the quantile loss. The quantile loss function takes the form

$$l_\alpha(Y_{s,t}, \hat{y}_{s,t}^{(\alpha)}) = \begin{cases} \alpha(Y_{s,t} - \hat{y}_{s,t}^{(\alpha)}) & Y_{s,t} > \hat{y}_{s,t}^{(\alpha)} \\ -(1-\alpha)(Y_{s,t} - \hat{y}_{s,t}^{(\alpha)}) & Y_{s,t} \leq \hat{y}_{s,t}^{(\alpha)} \end{cases} \tag{4.9}$$

where $\hat{y}_{s,t}^{(\alpha)}$ is the estimated α-quantile and $Y_{s,t}$ is the observation at time t and location s. By summing the quantile loss for all s and t, an estimate of the total quantile loss is obtained. This can be used as evaluation function in a cross-validation scheme. This is what Opitz et al. (2018) propose. However, their approach is very computationally demanding, and a simple five fold cross validation scheme is implemented instead. Each station is assigned randomly to one of five folds. Then the model is fitted for each combination of four folds and each parameter value of interest with the remaining fold used as validation set. Then the loss across all folds is summed and compared for each value of p^+ and c. This selection of folds is used in order to optimise spatial prediction, as part of our goal is to strengthen inference across space. Note that the values for p^+ and c does not affect the gamma stage fit. We can therefore fit the gamma stage once per fold and then fit the binomial and GP stages for each value of p^+ and c. This speeds up the process significantly, as the gamma stage demands the most computational resources. This, combined with INLA's efficiency, means that the cross validation study can be conducted without access to high performance computing resources.

A subtle question that arises when this model is fitted is how to go from the marginal distributions for the three latent fields to an estimate for the final quantiles with uncertainty. First, note that the unstructured effects are thought to handle any noise the model has not otherwise accounted for and are therefore not included when computing the quantiles. This might introduce some bias towards lower values for the quantiles as the exponentiated iid effects will have a mean larger than zero. Second, the way in which the marginal uncertainty in $u_{s,t}$ influences the result is hard to account for as it determines the data used to fit the two other model stages. Two approaches are used to handle this. For the cross validation study, the median of the sum of the relevant latent field components is used both to set $u_{s,t}$ and the final quantiles. Equation (4.1) gives

$$\hat{y}_{s,t}^{(\alpha)} = \hat{u}_{s,t} + \hat{\kappa}_{s,t}^{(0.5)} \frac{\left(\frac{1-\alpha}{\hat{p}_{s,t}}\right)^{-\hat{\xi}} - 1}{0.5^{-\hat{\xi}} - 1},$$

where $\hat{\xi}$ is the posterior median for ξ and

$$\begin{aligned} \hat{\mu}_{s,t} &= \exp\left(\widehat{\beta_0^{\text{gam}} + f_s^{\text{gam}}(s) + f_{\sin}^{\text{gam}}(s) \cdot \sin(2\pi w_t/53) + f_{\cos}^{\text{gam}}(s) \cdot \cos(2\pi w_t/53)}\right) \\ \hat{p}_{s,t} &= \text{logit}^{-1}\left(\widehat{\beta_0^{\text{bin}} + f_t^{\text{bin}}(w_t) + f_s^{\text{bin}}(s)}\right) \\ \hat{\kappa}_{s,t}^{(0.5)} &= \exp\left(\ln \hat{\mu}_{s,t} + \widehat{\beta_0^{\text{GP}} + f_t^{\text{GP}}(w_t) + f_s^{\text{GP}}(s)}\right). \end{aligned}$$

The hats across a sum of latent field elements denotes the posterior median of the sum. The threshold is set as

$$\hat{u}_{s,t} = F^{-1}(p^+; \hat{\mu}_{s,t}, \hat{k}) \qquad \text{or} \qquad \hat{u}_{s,t} = c \cdot \hat{\mu}_{s,t}$$

depending on the threshold setting approach. $\hat{k}$ is the posterior median of the gamma shape parameter. This does not demand any further computations, and is assumed to give results close to the median quantile even though the quantiles' dependence on $u_{s,t}$ is not necessarily strictly increasing. However, this is not a suitable way to get inference on the uncertainty of the quantiles.

The other approach is to generate posterior samples and estimate the quantiles for each sample. It is possible to generate approximate posterior samples with INLA, so this is feasible. However, this model is not well suited to inference by sampling as the two upper stages has to be fitted for each sample of the gamma stage in order to get a single sample of the final quantile. This is very computationally demanding and thus not feasible to do within a cross validation study. Therefore the approach of plugging in the medians will be used for the cross validation study while posterior sampling will be used for the reported results.

4.4 Discussion of model assumptions

There are some assumptions made in this model that need to be discussed. The assumption of conditional independence is problematic when modelling hourly precipitation. As seen in Figure 3.8, the data clearly has dependent observations within each week. Cooley et al. (2007) use a declustering routine to ensure that their observations are approximately independent. They do not observe significant differences between a model fitted to the declustered data and a model that simply assumed conditional independence. Opitz et al. (2018) also assume conditional independence with seemingly few consequences. However, both cases work with daily precipitation which is significantly less auto-correlated than hourly. As the interest of this project lies on inference on the marginal distributions of the latent field, the problems of dependent observation is less decisive.

The assumption that only the time of year should influence the hourly precipitation amount is also arguable. For example, afternoon showers is a common phenomenon during the Norwegian summers, and the time of day is therefore worth considering. In Mathisen (2020) experimentation with a separate random walk process for the time of day was conducted for station SN18700. This did not show any significant effects of time of day. Figure 3.6 does not suggest any strong effect of time of day for this data set either, and it is therefore not taken into account in this model.

The assumption of spatial stationarity is also questionable. Some stations are separated by hills and some by the fjord, while others have only flat land between them. This can lead to different correlation between stations even though they are similar distances apart. As seen in Figure 3.1, the spatial structure of the data is quite weak. This means that any nonstationarities are hard to detect. We therefore assume that stationarity is a good approximation.

There is also some year to year variations in precipitation patterns. Sorteberg et al. (2018) show that there is a significant trend in yearly maximal hourly precipitation for several of the stations within the domain over a 30 year period. These stations are located in both Ås and Oslo, so they have some geographical spread. However, the effect of long term climate change is suspected to have little influence over the ten year period studied here. Figure 3.6 also suggests some year to year variations in the mean positive precipitation. The inclusion of an unstructured effect of the year might therefore be sensible. Still, the effects is very small compared to the effect of the time of year. It is therefore not included in order to reduce the model complexity.

There are other covariates that could have an impact on the precipitation extremes. Dyrrdal et al. (2015) use altitude, distance to the sea, temperature data, and information about the bulk of the distribution like the mean precipitation and number of wet days as covariates for their model of hourly maxima. In our case, temperature data is not easily available. As this model has the bulk of the data available to it, the covariates for information about these will not be taken into account. As seen in Figure 3.7, the altitude is

also a covariate with a weak effect on the results at best, possibly due to little variation in altitudes among stations. A similar problem occurs with regard to the distance to the sea. Almost all stations are located close to the fjord, and the data contains little information about its effect. Therefore altitude and distance to the sea are not included in this model.

Another assumption made in this model is that both the shape of the gamma distribution and the tail parameter for the GP distribution remain constant over time. It might be sensible to also link these to the latent field. For example both Cooley et al. (2007) and Dyrrdal et al. (2015) let ξ vary in space. However, as these parameters enter the likelihood in a nonlinear way, the current implementation of INLA demands that they stay constant in time and space. In the case of ξ, it is also hard to estimate, and Dyrrdal et al. (2015) conclude that a fixed $\xi = 0.15$ gives about the same predictive power as letting it vary in space. Letting ξ be independent of covariate information is therefore justifiable. A discussion of the consequences of this is found in section 5.4.

Chapter 5

Results and discussion

As mentioned in the introduction, the goal of this thesis is to investigate whether or not INLA can be used to reliably estimate a high quantile for hourly precipitation. More precisely, the α-quantile with $\alpha = 0.99998$ was chosen as a target. This roughly corresponds to events that will happen each 5-6 years at any given station. In this chapter, the extent to which this is accomplished is presented and discussed.

5.1 Threshold selection

As discussed in section 4.3, we want to compare two approaches for setting the threshold $u_{s,t}$ by comparison of the quantile loss (see equation (4.9)). The GQ approach sets the threshold by equation (4.7), i.e. as the p^+-quantile of the estimated gamma distribution. The MGM approach sets the threshold as a fixed factor, c, times the gamma mean (see equation (4.8)). Due to a misunderstanding regarding `r-INLA`'s parameterisation of the gamma distribution, we believed that the GQ approach gave very unreasonable thresholds. The MGM approach was proposed in order to solve that problem. The misunderstanding was later clarified, but that happened very recently. Fitting the whole model is rather cumbersome work, and, as we show in this section, the two approaches turn out to be virtually equivalent for this study.

The results from the five-fold cross validation study with respect to the threshold probability, p^+, can be found in Figure 5.1a. The model prefers very high values for p^+. Low loss is observed in the range $p^+ \in (0.999, 0.9999)$, with a minimum for $p^+ = 0.9994$. As mentioned in section 4.1, the expression we use to estimate the quantiles, equation (4.1), assumes that $p_{s,t} \geq 1 - \alpha$. For $p^+ \geq 0.99999$, we get $p_{s,t} \leq 1 - \alpha$, and consequently the full set of α-quantiles cannot be estimated from the model fit. Opitz et al. (2018) reports an optimum for $p^+ = 0.92$, which is considerably lower than the p^+ suggested here.

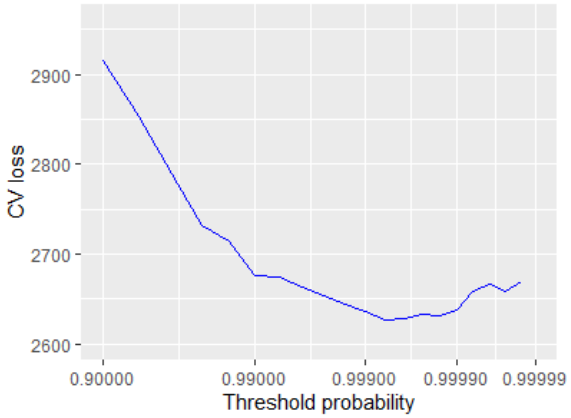
A five-fold cross validation study was also conducted with the threshold set by the MGM approach. This resulted in the plot seen in Figure 5.1b. There is a sharp decrease in cross validation loss as c goes from 2 to around 5. After that, a steady decrease is observed until a minimum at $c = 7.7$ followed by a slight increase for higher values. With $c \geq 11.9$, the above threshold probability $p_{s,t}$ is lower than $1 - \alpha$ for some week at some station, meaning that the α-quantile can no longer be calculated from the model.

The GP tail parameter, ξ, has nice behaviour in both cases. As discussed in section 2.1.2, the tail parameter should stabilise for thresholds where the GEV distribution is a good approximation. This does not generalise to the more complex setting explored here. However, as displayed in Figure 5.1c and 5.1d, the tail parameter still has this behaviour. It is promising to see that, in both cases, the tail parameter stabilises in the range where lowest loss is observed. This suggests that the threshold criteria manage to find suitable thresholds to satisfy a GP distributed tail.

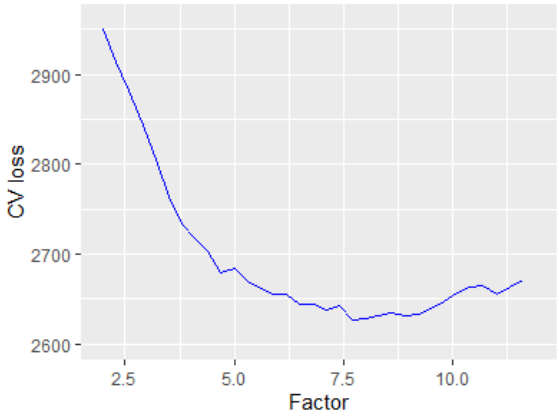
The resulting thresholds for station SN18700 is found in Figure 5.2. The thresholds and their credible intervals align almost perfectly. This is reflected in Figure 5.1, where both the loss plots and the plots of ξ are very similar for the two methods. This is due to an estimated shape parameter k close to 1 for the gamma distribution. $k = 1$ corresponds to the exponential distribution which has the property

$$F^{-1}(p^+; \mu) = -\ln(1 - p^+) \cdot \mu.$$

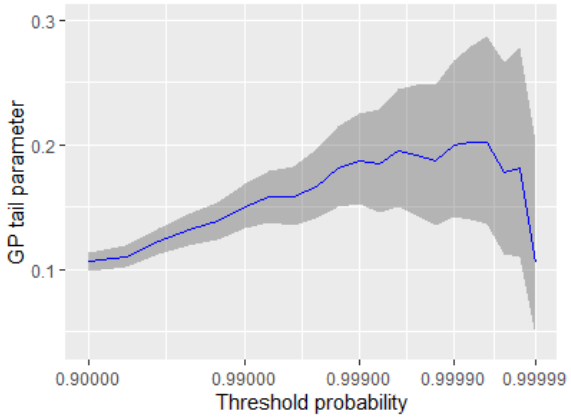
The p^+-quantile of the estimated gamma distribution is therefore very close to a fixed multiple of the mean $\mu_{s,t}$. This means that the GQ and MGM approach are virtually equivalent for our study. We are therefore confident that the remainder of the results would change negligibly by changing threshold criterion. The



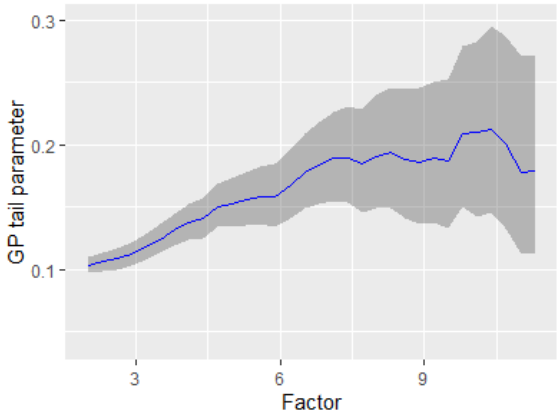
(a) The cross validation loss as a function of the threshold probability p^+.



(b) The cross validation loss as a function of the factor c.



(c) GP tail parameter, ξ, as a function of the threshold probability. The gray band marks the 95% credible interval. The value is based on estimates from the full data set, not the cross validation fits.



(d) GP tail parameter, ξ, as a function of the factor c. The gray area marks the 95% credible interval. The values are computed based on the full data set.

Figure 5.1: Plots of the quantile loss (see equation (4.9)) and corresponding values for the GP tail parameter with two different approaches for setting the threshold $u_{s,t}$. The plots to the left corresponds to setting $u_{s,t}$ as the p^+-quantile of the gamma distribution, while the plots to the right corresponds to setting it as a factor c times the gamma mean. The target quantile is the 0.99998-quantile.

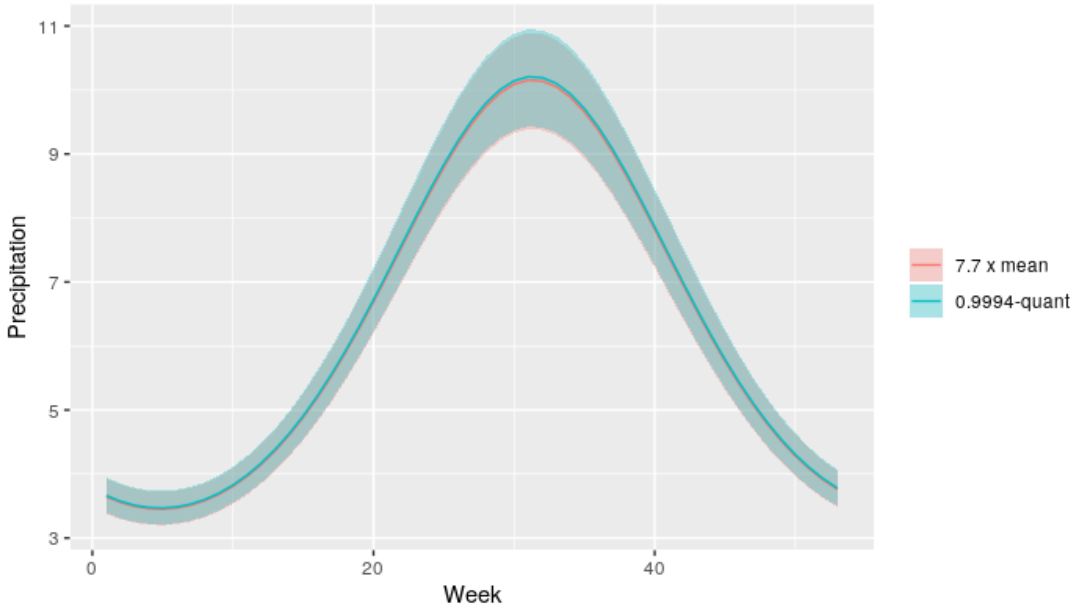


Figure 5.2: The threshold for station SN18700 (Blindern) for both the GQ and MGM approach. The shaded areas marks 95% credible intervals generated by approximate posterior sampling. The two thresholds are almost equivalent.

full set of results was analysed with the MGM method before the mentioned misunderstanding was clarified. It is therefore used for the remainder of this chapter with $c = 7.7$.

With $c = 7.7$, there are a total of 3220 exceedances of the median threshold. As a consequence, only about 0.4% of the positive precipitation observations are used to estimate the tail behaviour. This corresponds to 0.05% of all the full data set. The values for $p_{s,t}$ from the binomial stage are therefore expected to be small. Since a quantile can only be estimated from this model if $p_{s,t} > 1 - \alpha$, the approach used here is only applicable for quite high quantiles. The results from the binomial stage suggests a lower limit of $\alpha = 0.9999$ for this data set with this approach. The 0.9999-quantile roughly corresponds to weather events that happen on average once yearly at each site. If return levels over shorter time spans are of interest, other methods should be used.

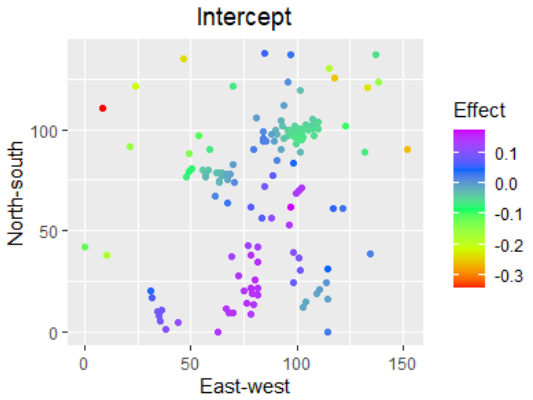
5.2 Analysis of the model stages

In this section, the results for the different model stages are analysed. The results for the gamma stage will be discussed first, then the binomial stage and lastly the GP stage.

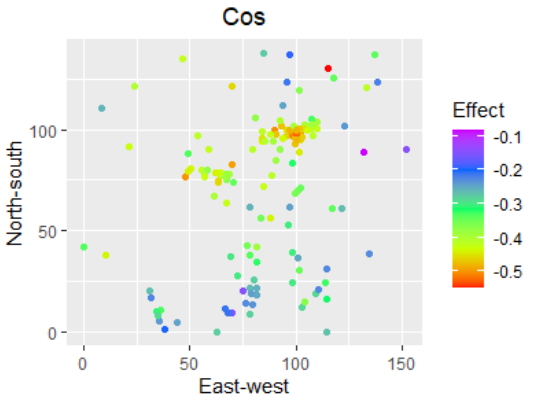
5.2.1 The gamma stage

Figure 5.4 and 5.3 show the latent field for the gamma stage (see equation (4.4)). The Matérn effect of location, $\boldsymbol{f}_s^{\text{gam}}$, is displayed in Figure 5.3a and 5.4a. This seems to indicate a significant spatial structure even though the unstructured effect, $\boldsymbol{g}_s^{\text{gam}}$, displayed in Figure 5.5, accounts for a large proportion of the variability (see Figure 5.7b).

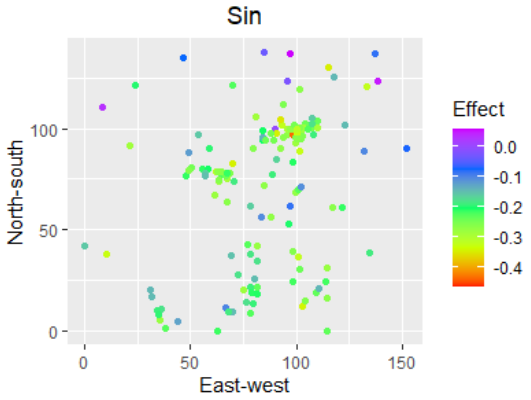
The remaining plots in Figure 5.4 displays the effects $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$ and $\boldsymbol{f}_{\text{cos}}^{\text{gam}}$. $\boldsymbol{f}_{\text{cos}}^{\text{gam}}$ can be interpreted as the difference between winter and summer, while the $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$ controls the difference between spring and fall. $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$ is, as expected, negative for all stations meaning that the summers have a higher average positive hourly precipitation than the winters. There are significant variations with regard to the amplitude of these variations with values in the range from -0.55 to -0.10. The values of $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$ are also mostly negative,



(a) The location of each station, s, plotted with colour corresponding to the spatial effect $f_s^{\text{gam}}(s)$.



(b) The location of each station, s, plotted with colour corresponding to the spatial effect $f_{\cos}^{\text{gam}}(s)$.



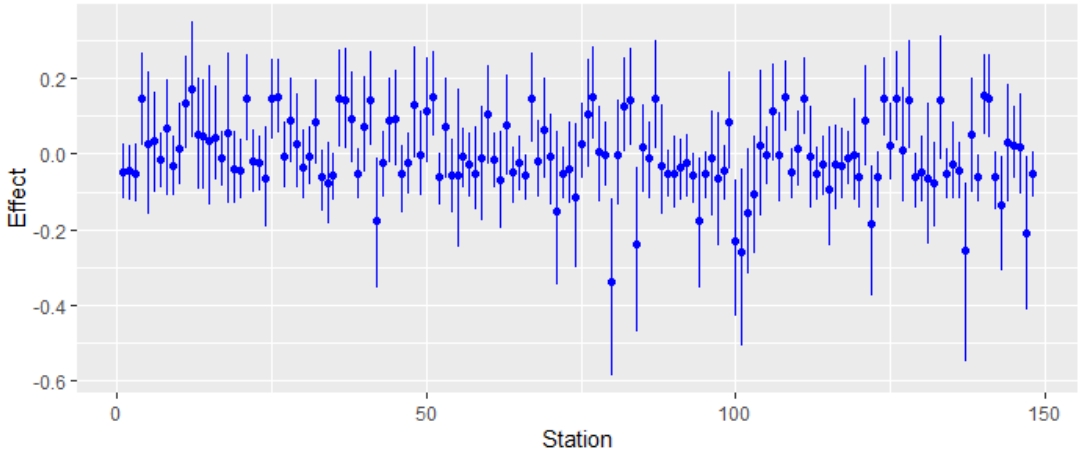
(c) The location of each station, s, plotted with colour corresponding to the spatial effect $f_{\text{sin}}^{\text{gam}}(s)$.

Figure 5.3: Maps of the different spatial effects in the gamma stage latent field.

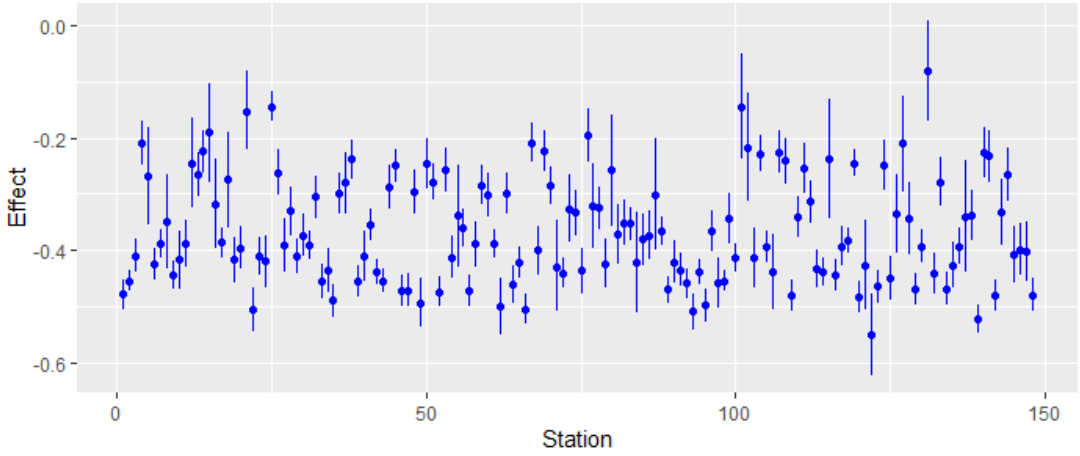
meaning more rain during autumn than spring. In this case, there are some stations s where $f_{\text{sin}}^{\text{gam}}$ is positive, but the 95% credible intervals in those cases always include zero. There are significant variations throughout the domain in this case too.

The estimated gamma mean, $\mu_{s,t}$, for four different stations is shown in Figure 5.6. The two upper plots show stations SN18700 and SN18410, both of which are located in central Oslo. SN18700 has ten years of recordings and therefore low uncertainty. SN18410 has only one year of data, which shows as much larger variations in empirical mean. However, the uncertainty is still low, meaning that the model borrows information from the neighbouring stations. In both cases, the estimated means seem like reasonable approximations of the empirical means. The two remaining stations are both located in Ås, only two kilometres apart. Perhaps due to calibration issues (see chapter 3), the observed mean at SN17875 is generally higher than at SN17870. This results in a slightly too high estimated mean for SN17870, which has a long time series, and a too low estimate at SN17875, which has a short time series. The uncertainty at these stations is larger than for the stations in Oslo.

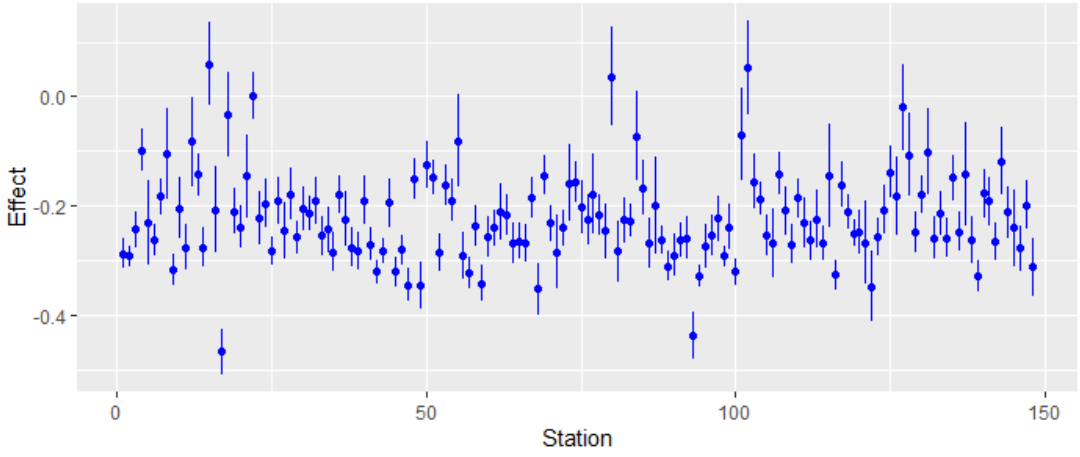
Figure 5.7 shows the posteriors for all eight hyperparameters for the gamma stage. As seen in 5.7a, the shape for the gamma likelihood has a very low variance. As there is a lot of data available at this stage, this is expected. The standard deviation parameters for the structured and unstructured spatial effects can be seen in Figure 5.7b. They both lie in the same range, but the Matérn effect has higher mean and variance. This reflects the results in Figure 5.4 and 5.5, which show that $\boldsymbol{f}_s^{\text{gam}}$ and $\boldsymbol{g}_s^{\text{gam}}$ are of the same magnitude. The standard deviation parameters for $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$ and $\boldsymbol{f}_{\text{cos}}^{\text{gam}}$ are also in the same range (see Figure 5.7c). Both posteriors are clearly different from the prior, as is the case for the standard deviations of



(a) The posterior medians of spatial effect $\boldsymbol{f}_s^{\text{gam}}$, with 95% credible intervals.



(b) The posterior medians of spatial effect $\boldsymbol{f}_{\text{cos}}^{\text{gam}}$, with 95% credible intervals.



(c) The posterior medians of spatial effect $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$, with 95% credible intervals.

Figure 5.4: Structured effects for the gamma stage.

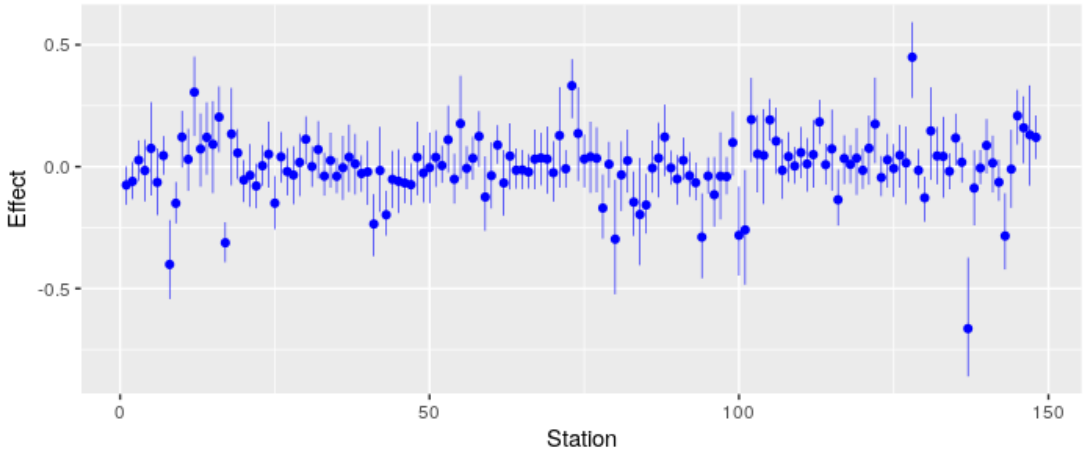
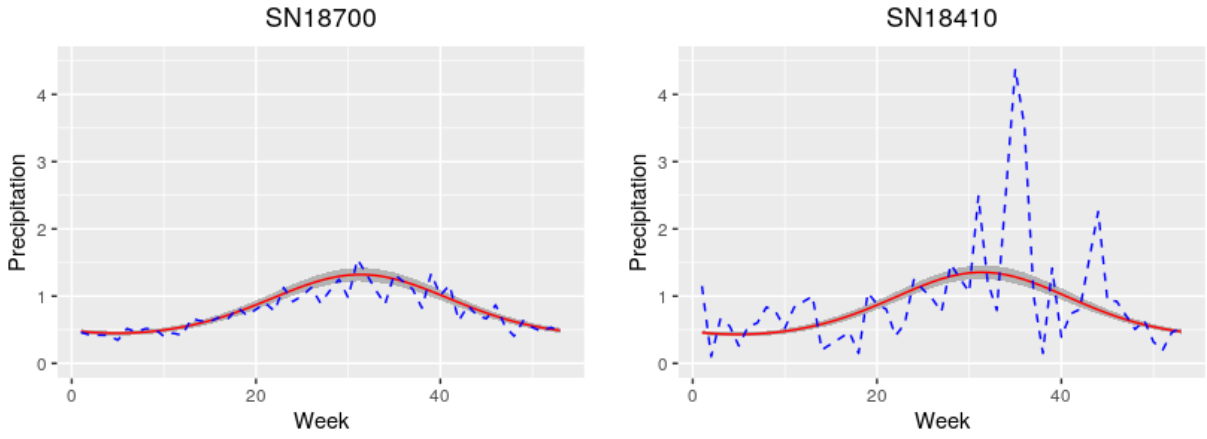


Figure 5.5: Posterior medians with 95% credible intervals for $\boldsymbol{g}_s^{\text{gam}}$, the unstructured spatial effect in the gamma stage.

$\boldsymbol{f}_s^{\text{gam}}$ and $\boldsymbol{g}^{\text{gam}}$. The range parameter posteriors varied to a larger extent between model components. As Figure 5.7d shows, the range for the structured effect, $\boldsymbol{f}_s^{\text{gam}}$, has quite a large median posterior range of 101 km, while $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$ has a short median range of 15 km and $\boldsymbol{f}_{\text{cos}}^{\text{gam}}$ 56 km. The range of $\boldsymbol{f}_s^{\text{gam}}$ in addition has a large variance. This is probably because the model struggles to differentiate between the structured and unstructured effect even though the amount of data is large. Even the range for $\boldsymbol{f}_s^{\text{gam}}$ is much lower than the prior suggests, meaning that the model finds evidence in the data for variations on much smaller scales than the prior radius of the domain. The differences between sine and cosine ranges indicate that the differences between summer and winter are correlated throughout most of the domain while the differences between spring and fall varies more erratically between locations. This is also what Figure 5.3 shows, with rather erratic variations in $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$, less so for $\boldsymbol{f}_{\text{cos}}^{\text{gam}}$ and very large scale correlation structures for $\boldsymbol{f}_s^{\text{gam}}$. It is not straightforward to argue why this difference between $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$ and $\boldsymbol{f}_{\text{cos}}^{\text{gam}}$ is observed.

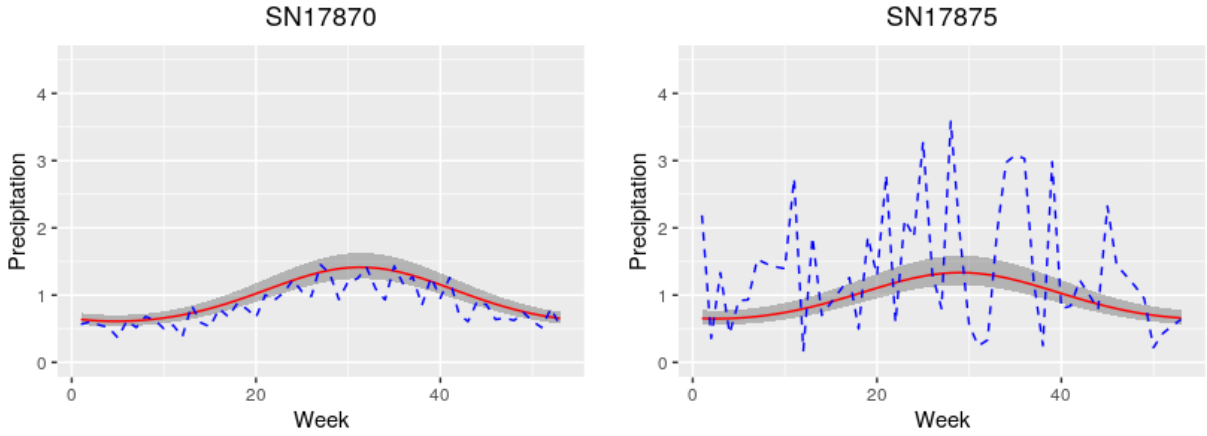
5.2.2 The binomial stage

Figure 5.8 displays the different components of the latent field for the binomial stage (see equation (4.5)). We observe a significant temporal dependence in the above threshold probability as the random walk effect, $\boldsymbol{f}_s^{\text{bin}}$, has clear structure and is significantly different from zero (see Figure 5.8a). We observe a higher probability of exceeding the threshold during summer and early fall, and lower during winter and early spring. This is a reasonable result for the data in question. Nevertheless, the unstructured temporal effect, $\boldsymbol{g}_t^{\text{bin}}$, observed in Figure 5.8b, also seems to have some temporal structure. This suggests that a stronger penalisation of the standard deviation for $\boldsymbol{g}_t^{\text{bin}}$ in the prior is in order for the random walk effect to encapsulate all the temporal structure in the above threshold probability. It is also worth noting that the unstructured temporal effect has a larger variance than the structured effect. This suggests that even though there is clear temporal structure to this model stage, it is also characterised by large unstructured fluctuations. These large unstructured variations suggest that there might be some covariate information we are not taking into account. The structured spatial effect, $\boldsymbol{f}_s^{\text{bin}}$, seen in Figure 5.8c, is not significant in this model stage. The variations are of a much lower magnitude than for the temporal effects, and the credible intervals of the effect at all stations contain zero.



(a) Gamma mean at Blindern.

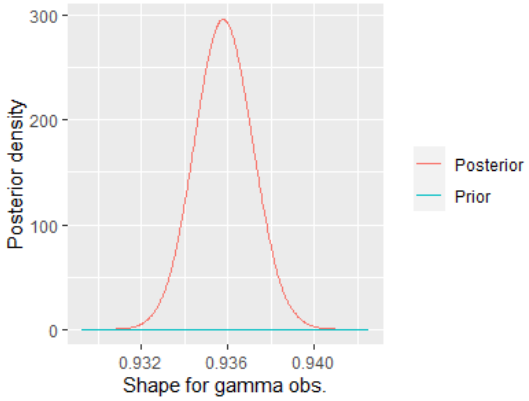
(b) Gamma mean at Aker Sykehus.



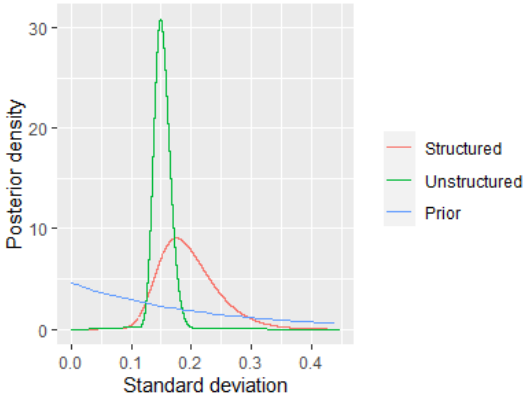
(c) Gamma mean at Ås.

(d) Gamma mean at E18 Holstad.

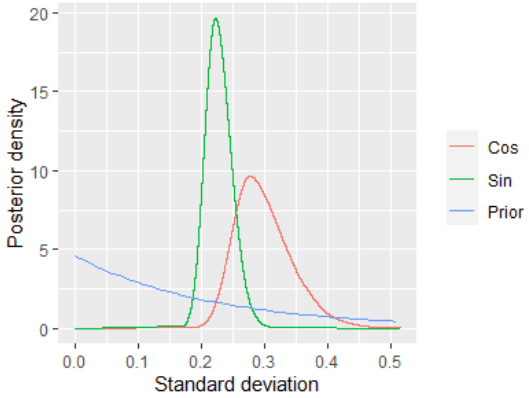
Figure 5.6: Mean positive precipitation as a function of time of year for four different stations. The red lines mark the posterior median and the shaded area the 95% credible interval. The blue lines are the observed means.



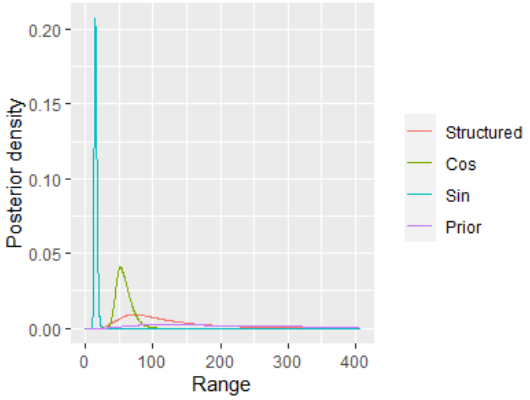
(a) The posterior density for the shape k of the gamma distribution. The blue line marks the prior.



(b) The posterior density for the standard deviation parameter for the structured and unstructured spatial effect in the gamma stage latent field, $\boldsymbol{f}_s^{\text{gam}}$ and $\boldsymbol{g}_s^{\text{gam}}$. The purple line marks the prior.

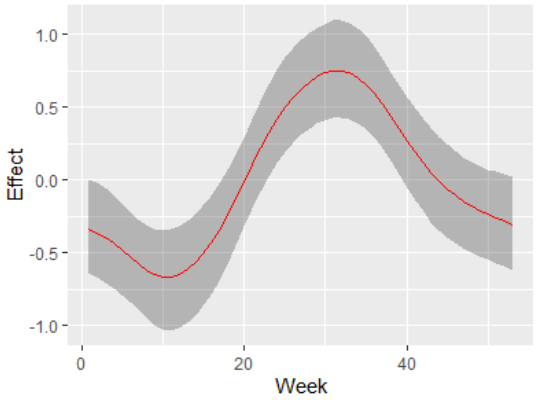


(c) The posterior density for the standard deviation parameter for the sine and cosine coefficient spatial effects in the gamma stage latent field, $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$ and $\boldsymbol{f}_{\text{cos}}^{\text{gam}}$. The purple line marks the prior.

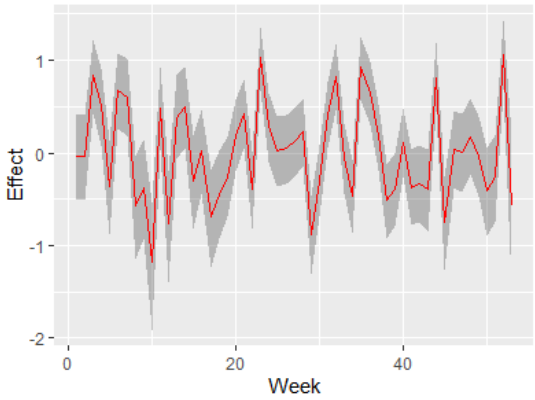


(d) The posterior density for the range parameter for all effects in the gamma stage latent field with Matérn priors. The structured component refers to $\boldsymbol{f}_s^{\text{gam}}$, while the sine and cosine terms refer to $\boldsymbol{f}_{\text{sin}}^{\text{gam}}$ and $\boldsymbol{f}_{\text{cos}}^{\text{gam}}$ respectively. The purple line marks the prior.

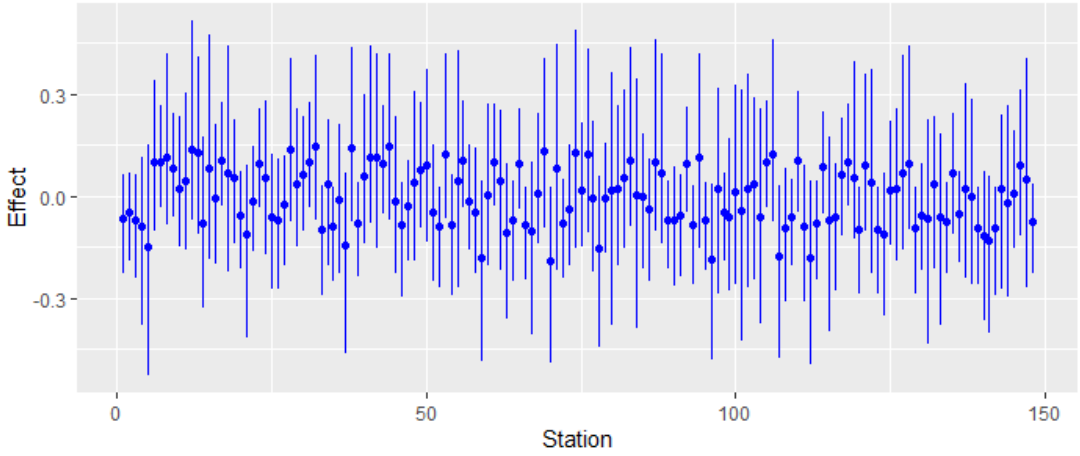
Figure 5.7: Hyperparameter posteriors for the gamma stage compared to their respective priors.



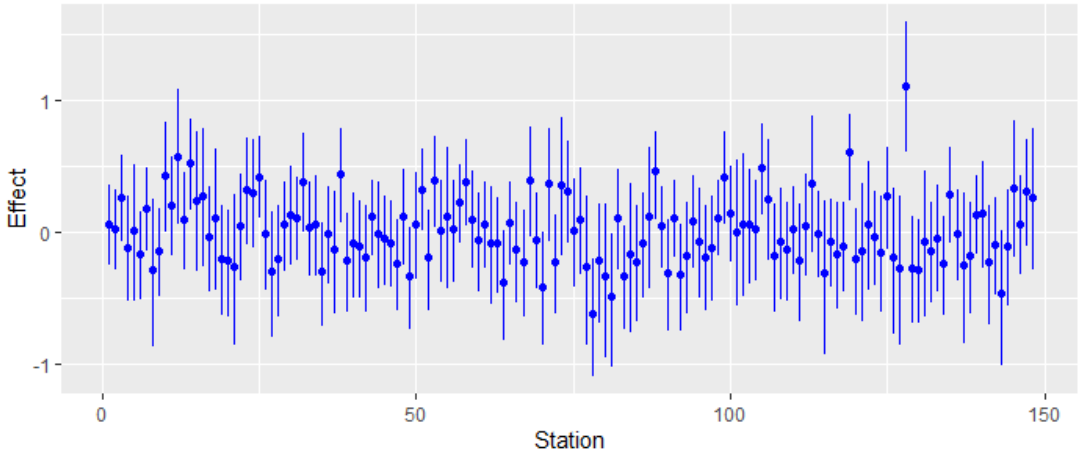
(a) The random walk effect on the latent field for the binomial stage, $\boldsymbol{f}_t^{\text{bin}}$. The red line marks the mean while the gray area marks the 95% credible interval.



(b) The unstructured temporal effect on the latent field for the binomial stage, $\boldsymbol{g}_t^{\text{bin}}$. The red line marks the mean while the gray area marks the 95% credible interval.

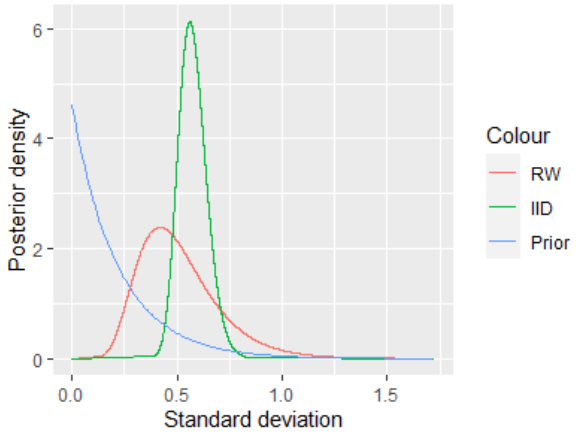


(c) The structured spatial effect of the binomial stage, $\boldsymbol{f}_s^{\text{bin}}$. The ranges marks the 95% credible intervals.

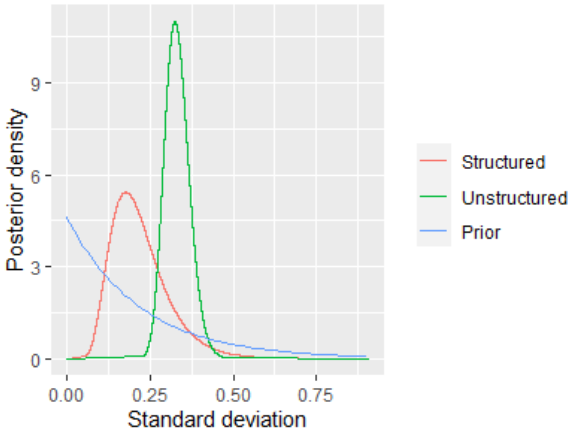


(d) The unstructured spatial effect of the binomial stage, $\boldsymbol{g}_s^{\text{bin}}$. The ranges marks the 95% credible intervals.

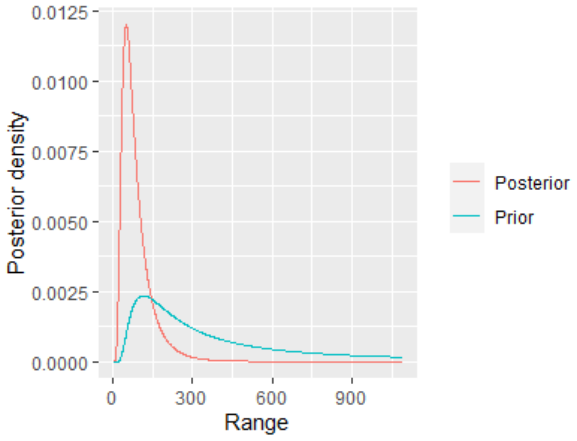
Figure 5.8: All effects on the binomial stage latent field. The two upper plots shows the temporal effects while the lower shows the spatial effects.



(a) The posterior density of the standard deviation parameters for the temporal random walk effect $\boldsymbol{f}_t^{\text{bin}}$ and temporal iid effect, $\boldsymbol{g}_t^{\text{bin}}$ in the binomial stage. The blue line marks the prior.



(b) The posterior density of the standard deviation parameters for the structured spatial effect $\boldsymbol{f}_s^{\text{bin}}$, and unstructured spatial effect, $\boldsymbol{g}_s^{\text{bin}}$, in the binomial stage. The blue line marks the prior.



(c) The posterior density for the range parameter for the structured spatial effect in the binomial stage latent field, $\boldsymbol{f}_s^{\text{bin}}$.

Figure 5.9: Hyperparameter posteriors for the binomial stage.

5.2.3 The generalised Pareto stage

Figure 5.10 shows the posterior for the different components in the latent field for the GP stage (see equation (4.6)). Figure 5.10a shows that also the GP median has a significant temporal structure. $\boldsymbol{f}_t^{\text{GP}}$ has a significant tendency towards higher median threshold exceedance during summer than during fall and early winter. This is interesting as the gamma log mean was included as an additive offset for the linear predictor in the GP stage. The significant temporal structure thus indicates that the temporal structure in the extremes is different from the temporal structure in the bulk of the distribution, with even larger differences between summer and winter than the bulk. For this model stage, the unstructured temporal effect seen in Figure 5.8b is not significant.

The spatial effect for the GP stage is not significant either. This goes both for the structured and unstructured effect (see Figure 5.10c and 5.10d). This is confirmed by the large posterior range seen in Figure 5.11d. This indicates that the spatial variations in the extremes are similar to those for the bulk of the data.

Figure 5.11 shows the posterior distributions for the GP stage hyperparameters. The main takeaway is that the tail parameter, ξ, lies in a sensible range with a posterior median of 0.18 and standard deviation of 0.02. This indicates that the tail is heavier than for an exponential distribution. Figure 5.11b reveals that the standard deviation parameter of $\boldsymbol{f}_t^{\text{GP}}$ is higher than that of $\boldsymbol{g}_t^{\text{GP}}$ for this model stage. This means that a larger proportion of the variability is determined by the structured variations compared to the unstructured. This is especially clear when comparing with Figure 5.9a where the iid effect accounts for more of the variance.

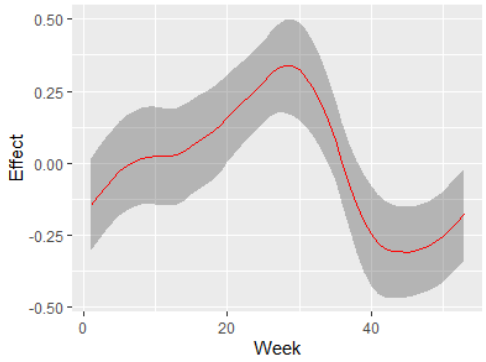
The value of ξ affects the results to a large extent as it is the only hyperparameter directly inserted into (4.1). However, it is also hard to estimate due to the low amounts of data in any extreme value setting. Some prior sensitivity testing is therefore conducted with respect to the penalisation rate in the PC prior for ξ. Figure 5.12 shows the posterior distribution of ξ with different rates in the PC prior. This shows that with $\lambda = 1$, meaning a more or less flat prior on $[0, 1]$, the posterior for ξ is very close to the one obtained by a rate of $\lambda = 4.5$. Similar results are obtained with a doubling of λ. With a penalisation rate of $\lambda = 100$, the posterior is shifted notably to the left compared with the one obtained with a λ determined by prior knowledge. This penalisation rate is very unreasonable, but even with this prior the posterior lies in a sensible range. This indicates that the data contains sufficient information about the tail behaviour. Small variations in the hyperprior has only a weak influence on the results.

5.3 Quantiles

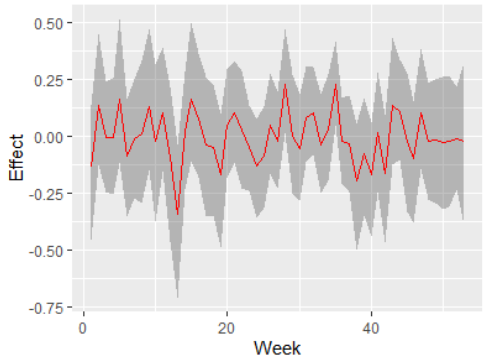
We will now look at the quantiles of interest. We will discuss both the temporal structure at different locations and the spatial structure for different times.

We first consider the temporal structure for different stations. The resulting quantiles from stations SN18700, SN18410 and SN1070 can be seen in Figure 5.13. In all three cases, the plots show significant seasonal variations with higher quantiles during summer than winter. The uncertainty is also larger during the summer than it is during the winter. This is probably a consequence of the higher precipitation amounts during summer, but also the fact that the summers have larger variance in hourly precipitation amounts. Stations SN18410 and SN18700 are located in Oslo, within a five kilometer range of each other and several other stations. Station SN1070 lies in Hvaler more than 15 kilometres from the nearest station, close to the boundary of the domain. We see that despite having less data, SN18410 has lower variance in the quantile than the one in Hvaler. This suggests that the model manages to borrow information across space, especially in clusters with many stations.

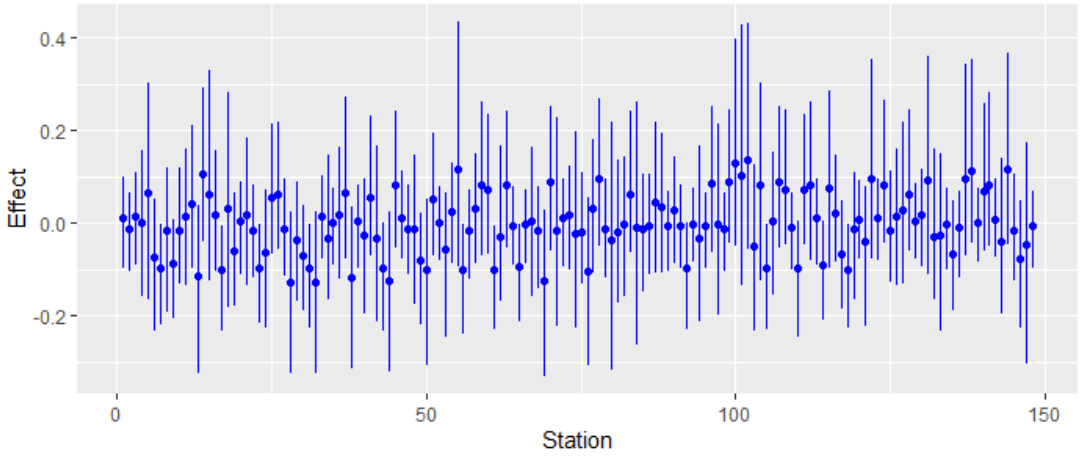
Note that the quantiles estimated by both posterior sampling and insertion of the medians into (4.1) are included in the plot (see section 4.3 for details). This is hard to see because they align more or less fully. Larger deviations were observed at some stations, but the differences are well within the uncertainty at all stations. This shows that the use of quantiles computed by inserting the median effects into 4.1 is a reasonable choice for the cross validation study. Also note that the maximal observations from each week



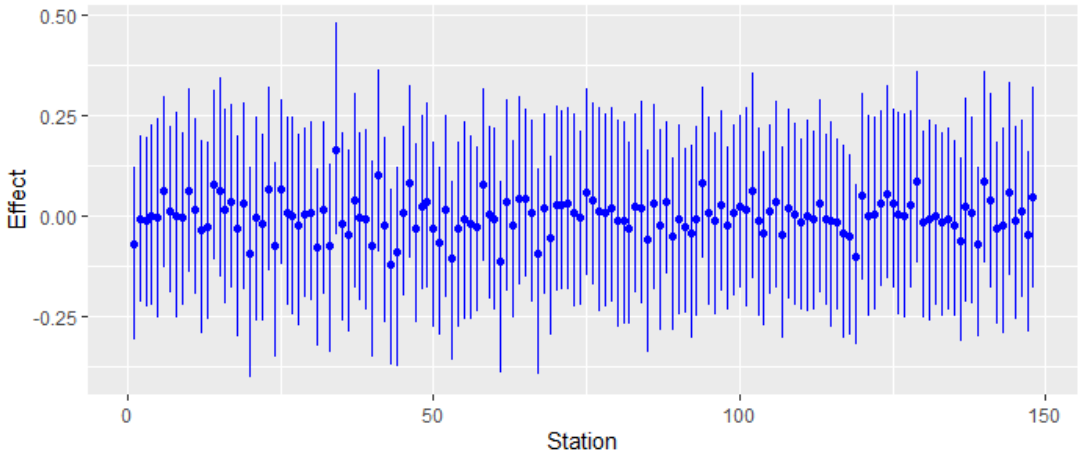
(a) The random walk effect on the latent field for the GP stage, $\boldsymbol{f}_t^{\text{GP}}$. The red line marks the mean while the gray area marks the 95% credible interval.



(b) The temporal iid effect on the latent field for the GP stage, $\boldsymbol{g}_t^{\text{GP}}$. The red line marks the mean while the gray area marks the 95% credible interval.

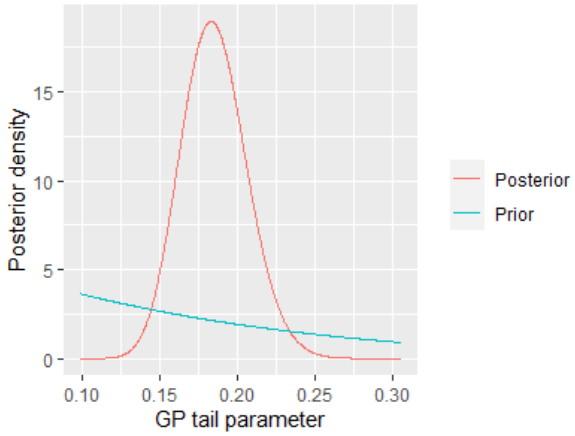


(c) The structured effect of spatial location on the latent field for the GP stage, $\boldsymbol{f}_s^{\text{GP}}$. The ranges marks the 95% credible intervals.

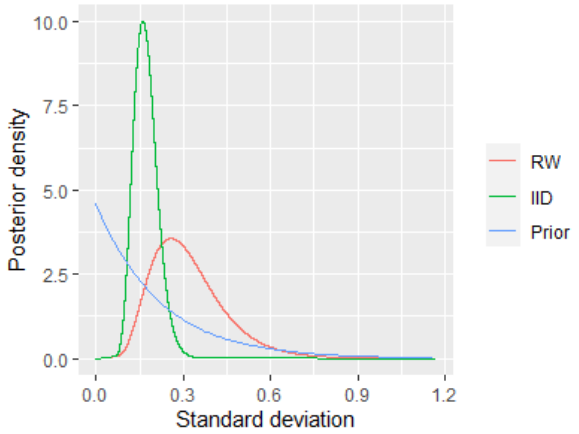


(d) The unstructured effect of spatial location on the GP stage latent field, $\boldsymbol{g}_s^{\text{GP}}$. The ranges marks the 95% credible intervals.

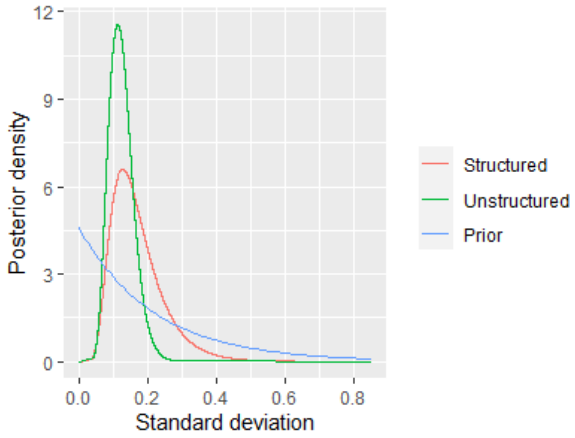
Figure 5.10: Effects in the GP stage latent field.



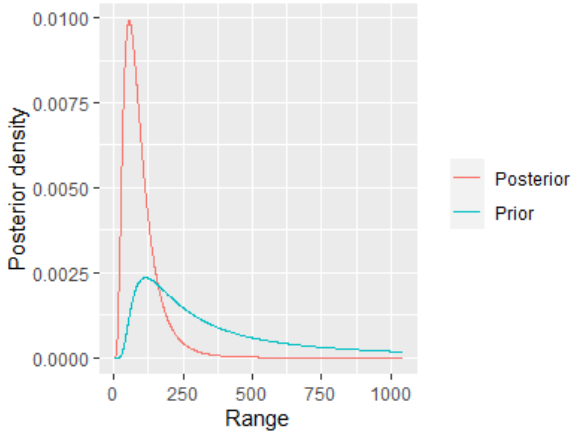
(a) The posterior distribution of the GP tail parameter, ξ. The blue line marks the prior.



(b) The posterior distribution of the standard deviation parameters for the temporal effects in the GP stage, $\boldsymbol{f}_t^{\text{GP}}$ and $\boldsymbol{g}_t^{\text{GP}}$. The orange line marks the structured effect while the green marks the unstructured effect. The blue line is the prior.



(c) The posterior distribution for standard deviation of the structured and unstructured spatial effects of the GP stage, $\boldsymbol{f}_s^{\text{GP}}$ and $\boldsymbol{g}_s^{\text{GP}}$. The blue line marks the prior.



(d) The posterior distribution for the Matérn range parameter in the GP stage.

Figure 5.11: Posterior distributions for the GP stage hyperparameters.

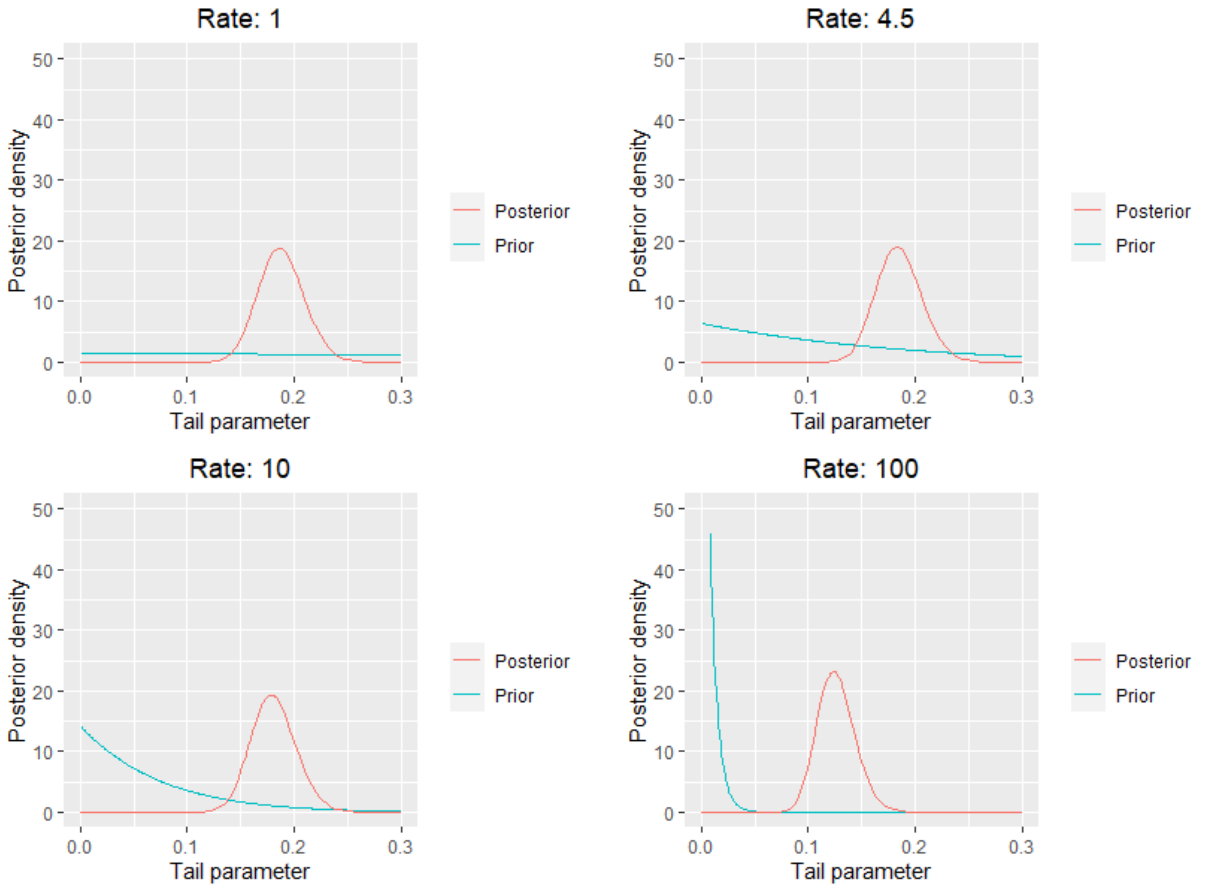


Figure 5.12: Posterior distributions for the GP tail parameter, ξ, for different values of the rate, λ, by which deviation from an exponential distribution is punished in the hyperprior. To the upper left $\lambda = 1$, to the upper right $\lambda = 4.5$, to the lower left $\lambda = 10$ and to the lower right $\lambda = 100$. In all cases the orange line marks the posterior while the blue marks the prior

(the red lines in Figure 5.13) are the values that minimise quantile loss. However, as each week has far from 50000 observations, the max values will be an underestimation of the 0.99998-quantile. The expected number of exceedances of the estimated quantile for a station with ten years of observations is 1.8 meaning that the two observed exceedances for SN18700 is a reasonable amount. Appropriation of the max values as quantile estimate would be a sign of overfitting, and the plots show no evidence of that. Quantiles very close to the max values at the stations with short time series would also be concerning in this regard. Yet, we observe that the quantiles at SN18700 lie closer to the max values than those at the other two stations.

Figure 5.14 shows the spatial variations in quantile for week 4 and 27. For week 27, the variance is high for most stations, and the spatial structure is weak to nonexistent, with the exception of some outliers. For week 4, the spatial structure is significant. These results are representative of the results from other weeks during the same seasons. This shows that there is in fact some space time interaction in the quantiles, as the stations with the highest quantiles during winter are not necessarily the ones with the highest quantile during summer and vice versa. Figure 5.15 shows the spatial structure of the mean for both weeks. This shows that during winter there is a tendency towards lower quantiles in the inner Oslofjord than to the south-west. Undercatchment is a common problem for weather stations where a large proportion of the precipitation comes in the form of snow (Fassnacht, 2004). As we expect more snow in the inner Oslofjord than the outer, the spatial pattern observed for the winters might be linked to this undercatchment. Without data on what precipitation type the stations measure, this is hard to account for. During summer, this spatial structure is no longer there. Instead, the quantiles have a weak tendency towards high values in the center of the domain and lower values further out. This effect is not significant.

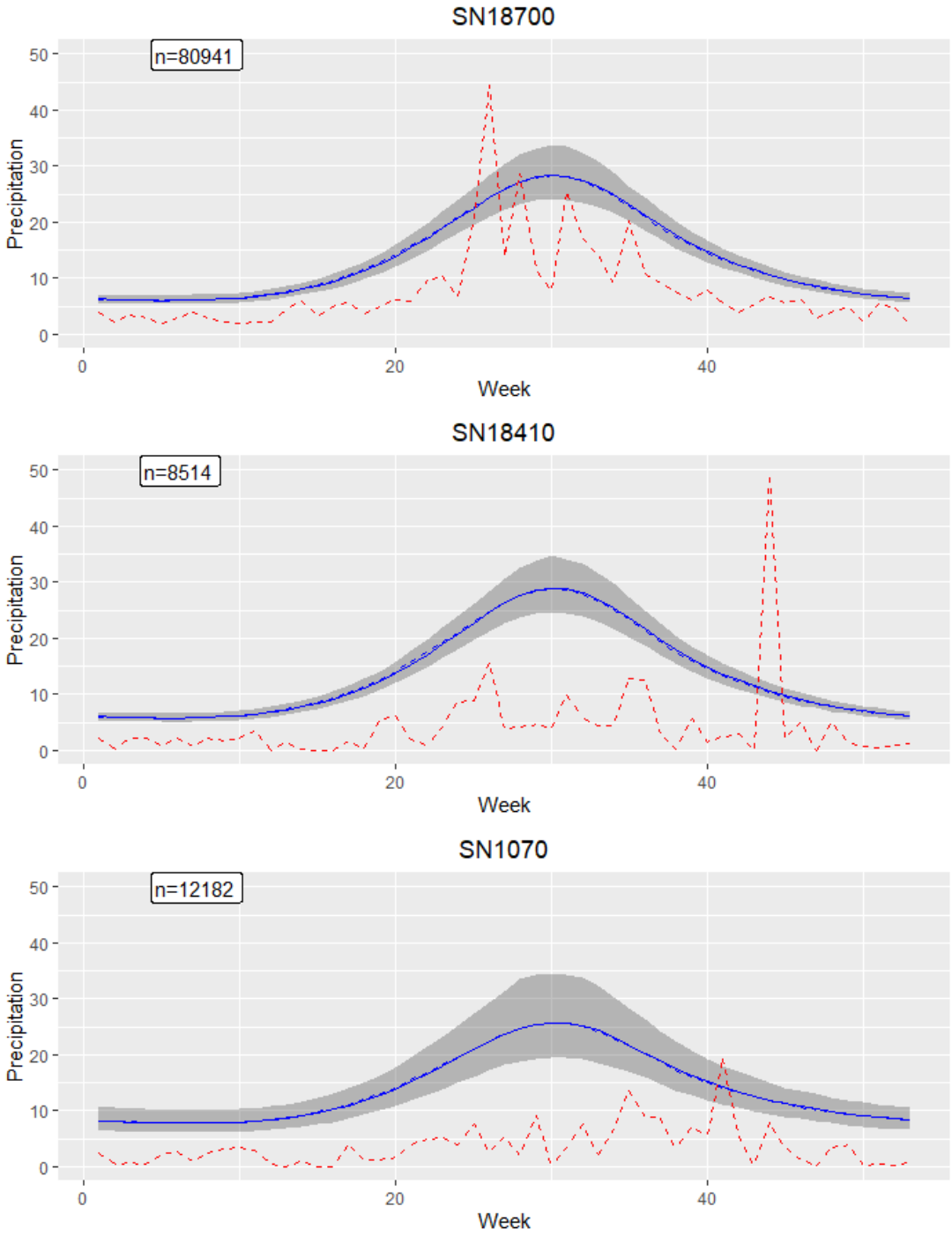
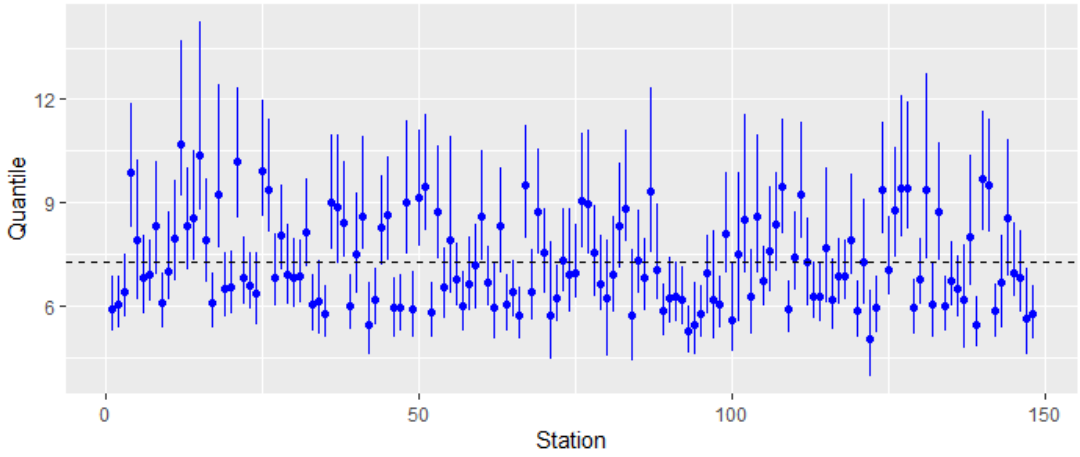
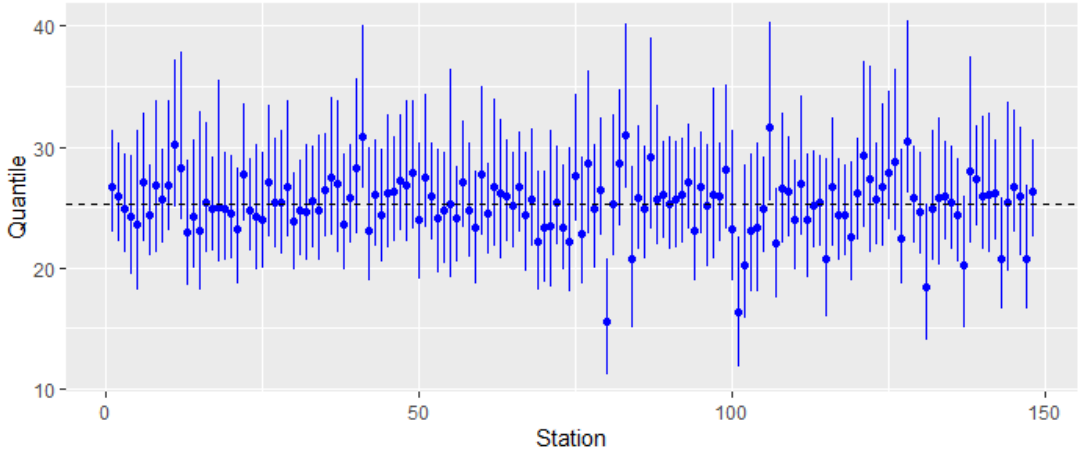


Figure 5.13: The quantiles for stations SN18700, SN18410 and SN1070. The blue lines mark the model estimates of the 0.99998-quantile. The fully drawn line is the mean posterior sample and the dashed is obtained by inserting the posterior medians of each model component into (4.1). The gray area is a 95% credible interval obtained by approximate posterior sampling. The red line marks the observed maximum values for each week. The number n in the upper left corner marks the number of observations at each station.



(a) The quantiles with 95% credible interval for all stations in week 4.



(b) The quantiles with 95% credible interval for all stations in week 27.

Figure 5.14: Effect of spatial location on final quantile for two selected weeks.

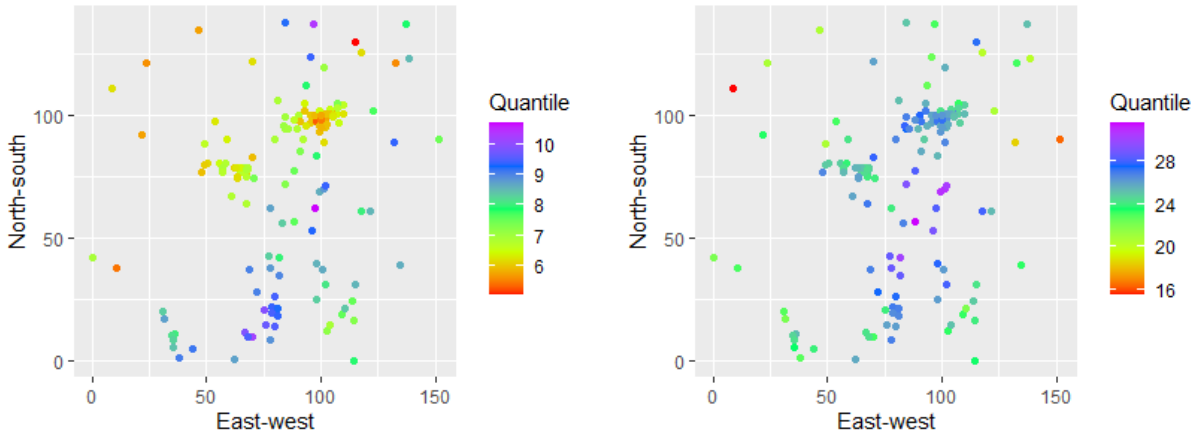
5.4 Discussion of model assumptions

When and where the data exceeds the estimated quantile, is central to the discussion of model fit. Figure 5.16 shows the observed probability of exceeding the estimated quantiles at each station. There are some extreme outliers with exceedance rates as high as ten times the target. These are usually stations with few observations but also one or two extremes. There is no clear spatial structure to the exceedances which indicates that the model captures the relevant spatial dependencies.

For a quantile estimate, $\hat{\boldsymbol{y}}_{s,t}^{(\alpha))}$, the number of exceedances, n can be recorded. Based on this, we can estimate the quantile level

$$\hat{\alpha} = 1 - \frac{n}{m},$$

where m is the total number of observations. By doing this for each posterior sample, we obtain an estimate of what quantile level the posterior quantiles correspond to. Figure 5.17 shows the results of this both globally (the orange line) and on a week to week basis (points with uncertainty). There is some evidence of too many exceedances during summer and too few during late spring and early fall. This signal is not very strong, but suggests that the model smooths the quantiles too much. The main concern observed in Figure 5.17 is that the total number of quantile exceedances is higher than expected. For our quantile estimates, $\hat{\alpha} = 0.999974$ which corresponds to a 30% higher exceedance rate than the target quantile. The



(a) Mean quantiles for week 4 plotted in space.

(b) Mean quantiles for week 27 plotted in space.

Figure 5.15: Mean posterior quantiles for two selected weeks. Note that the color scale is different between the two maps in order to magnify the spatial structure.

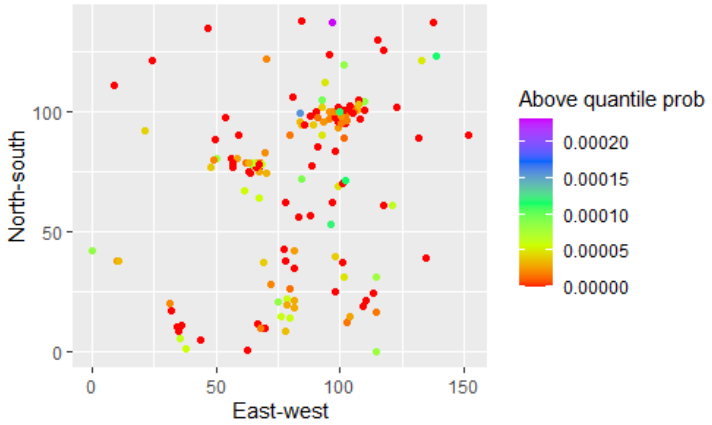


Figure 5.16: The location of each station plotted with colour corresponding to the observed quantile exceedance rate on each station.

95% credible interval for $\hat{\alpha}$ does not contain the target level of 0.99998. There are several possible reasons for this. The target level is very close to the limits of the credible interval, and about 1.6% of the posterior quantile samples correspond to higher levels. This means it could be caused by random noise. However, a similar percentage-wise exceedance rate is observed for $\alpha = 0.9999$, meaning that the problem persists even in a setting with five times more data available. Another possible cause is that $c = 7.7$ is still too low of a value to ensure extreme behaviour, but the lack of data leads to too high variance for higher values.

A third hypothesis for this discrepancy is that the model is biased towards lower values due to the model ignoring the iid effects when calculating the final quantiles. Too large noise terms could also explain the apparent deviation in exceedance rate between summer and spring/fall. Therefore a prior sensitivity test is conducted with regard to the amount of smoothing. Specifically, the upper model stages are fitted with different reference standard deviations, σ_0, for the iid effect hyperpriors. The results can be seen in Figure 5.18. The upper plots show that, for a reference value that is five times lower than what is used in the remainder of the project, the results are more or less equivalent. The model is thus not very sensitive to minor changes in these hyperpriors. However, with $\sigma_0 = 0.001$ corresponding to almost no smoothing, the noise starts to dominate the random walk processes. The final quantiles then tend to overfit. However, for $\sigma_0 = 0.01$, the quantile estimates remain smooth, and, in addition, the observed quantile exceedance rate aligns with the reference rate. This suggests that the inclusion of iid effects for smoothing leads to a

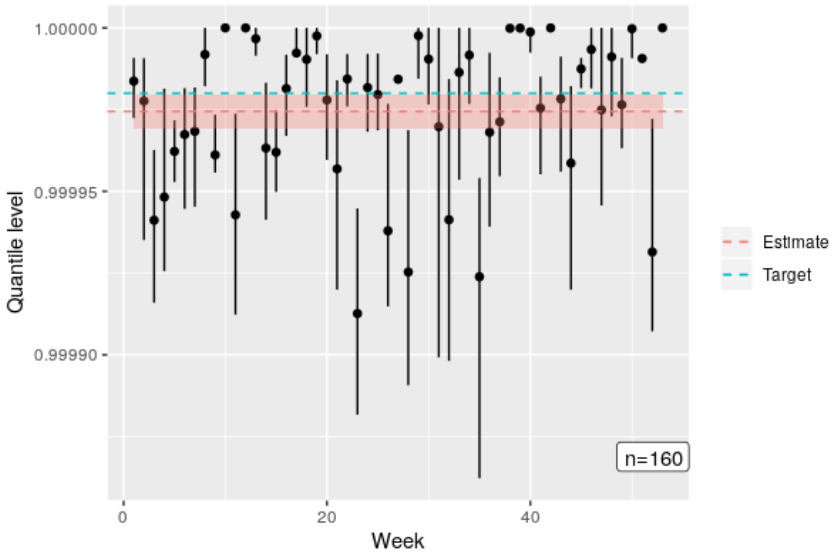


Figure 5.17: Median quantile level for the different weeks with 95% credible intervals. The orange line marks the global quantile level, while the blue line marks the target. The number n marks the median number of above quantile observations. Uncertainty is based on posterior sampling.

bias towards lower quantile values. At SN18700, the resulting quantiles is contained within the uncertainty for our current quantile estimates. This suggests that our estimates are in the right range, but this was not the case at all stations.

A more thorough investigation of the bias-variance tradeoff associated with the inclusion of the iid effects should therefore be conducted in further work. A possibility is to solve this by cross validation. A possible approach is to remove the iid effects from the latent fields and include the reference standard deviation for the structured effects in the cross validation study with regard to the threshold criterion. If a moderate number of penalisation rates are explored, this is feasible without access to high performance computing. A cross validation study to set all major hyperparameters resembling that of Opitz et al. (2018) is also possible in theory, but it demands much more computational power. A more thorough consideration of the impact of the smoothing effects on the quantiles is at least needed.

As mentioned in section 4, there are some model assumptions that should be discussed further. The assumption of constant likelihood parameters across space and time is one. Therefore, a model using only winter data and only summer data is fitted. Note that these models do not have cyclic random walk processes and are thus not directly comparable to the full model. They can still give some indication as to whether or not the hyperparameters depend on the time of year. This leads to hyperparameter posteriors seen in Figure 5.19. The gamma likelihood shape is significantly lower during summer than during winter. However, the tail parameter of the GP likelihood seems to have low variation between seasons. There is some evidence of a slightly higher ξ during winter, but due to larger variance in the winter posterior, the slight shift is not conclusive. This is encouraging as a difference between summer and winter is observed at Blindern in Mathisen (2020). There, the difference was opposite with lower ξ during winter and higher during summer.

A comparison of the resulting quantiles from the seasonal fits and the full model fit is included in Figure 5.20. We see here that for SN18700 the quantiles estimated on the summer data are higher but within the uncertainty of the full model quantile estimates. It resembles the quantiles in Figure 5.18, with a higher prior penalisation of the noise ($\sigma_0 = 0.01$). During winter, the quantiles are lower than the full model suggests. This is surprising, as Figure 5.17 suggests too high exceedance rate during winter. This pattern is not as apparent for other stations, and the temporal patterns for the winter data generally look like overfitting. The seasonal fits still provide some evidence of larger seasonal variations than the model currently accounts for.

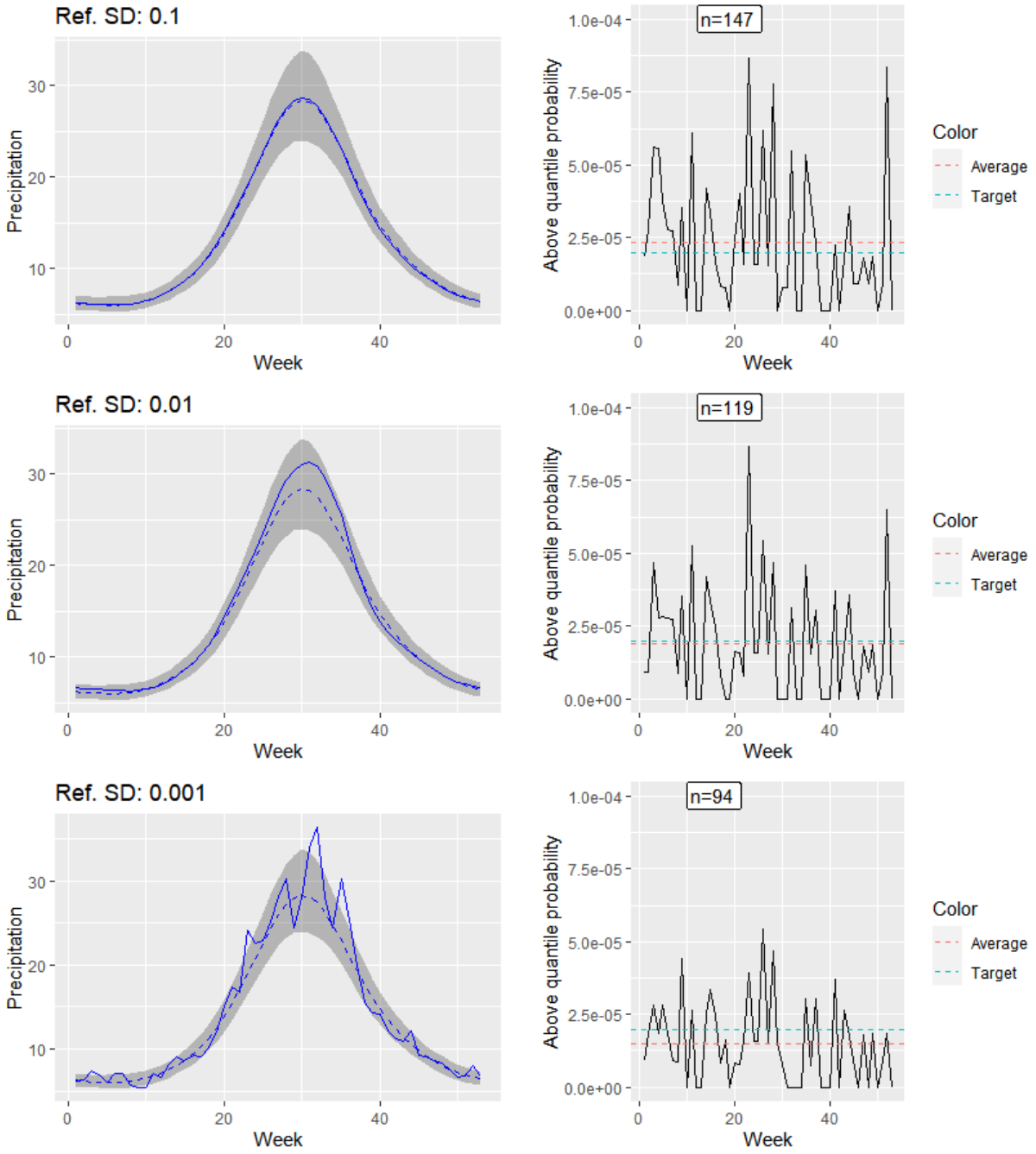
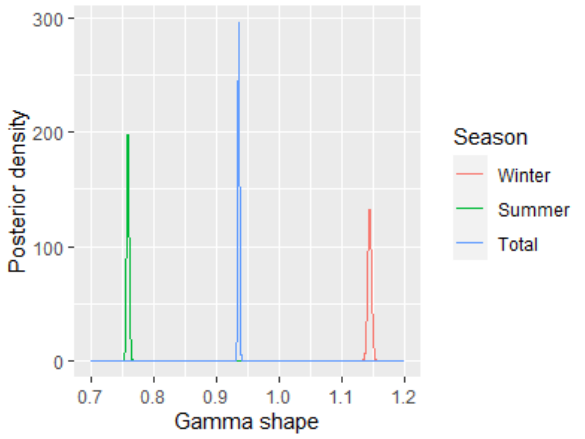
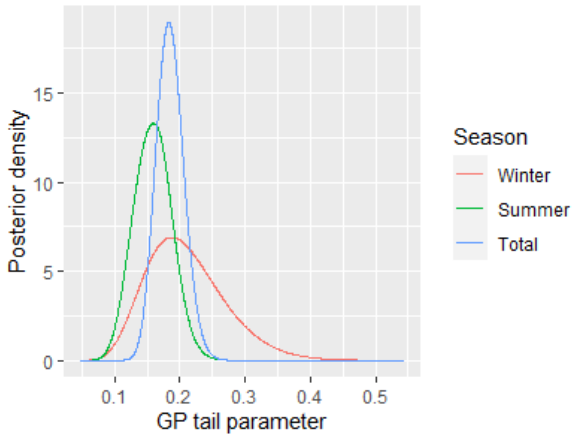


Figure 5.18: Resulting quantiles at SN18700 and above quantile probability for different values of the reference standard deviation, σ_0, in the prior for the iid effect standard deviations in the upper model stages. To the top, $\sigma_0 = 0.1$, in the middle, $\sigma_0 = 0.01$ and at the bottom, $\sigma_0 = 0.001$. For reference, the value used in the remainder of the project is $\sigma_0 = 0.5$. In all cases, $p_\sigma = 0.1$. The quantiles are generated as the posterior medians inserted into (4.1). In the plots to the left, they are compared to the full model quantiles generated by posterior sampling with uncertainty.

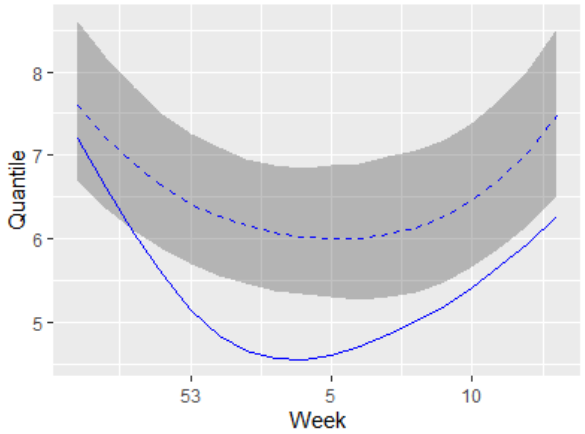


(a) The gamma likelihood shape parameter for models fitted to only summer and winter data compared to the full model.

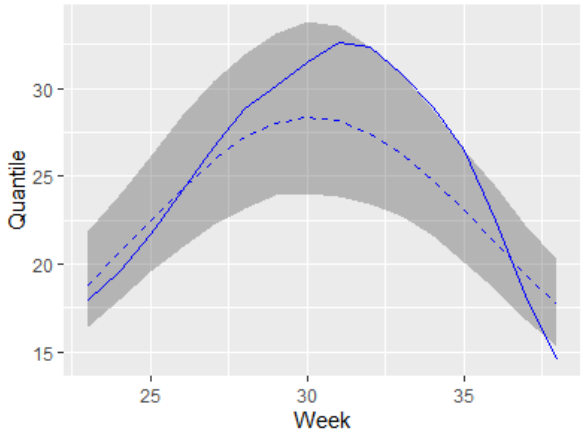


(b) The GP tail parameter for models fitted to only summer and winter data compared to the full model.

Figure 5.19: Comparison of likelihood parameters for models fitted to only summer data and only winter data. The summer data consists of data from week 23-38, while the winder data, week 1-13 and 49-53.



(a) The quantiles for station SN18700 during weeks 1-13 and 49-53. The fully drawn line marks the estimate based on only data from these weeks, while the dashed line marks the estimate from the full model with 95% credible interval.



(b) The quantiles for station SN18700 during weeks 23-38. The fully drawn line marks the estimate based on only data from these weeks, while the dashed marks the estimate from the full model with 95% credible interval.

Figure 5.20: Comparisons of the quantiles from the models for single seasons and the full model at station SN18700.

Chapter 6

Conclusion

In this thesis we study the use of INLA for inference on extreme hourly precipitation. A flexible three stage approach is used. The first model stage finds suitable thresholds above which extreme value distributions are a good approximation. The second models the probability of exceeding that threshold before finally a generalised Pareto distribution is fitted to the threshold exceedances. Each model stage takes the form of a latent additive Gaussian model that accounts for spatial dependencies and dependence on the time of year. The outputs from each model stage are then combined for an estimate of 99.998% quantile and its dependence on time and space. This corresponds to weather events that occur every five to six years at each point in space.

The threshold setting involves a crucial bias-variance tradeoff, and parameters of the threshold setting procedure are therefore set by cross validation. This is feasible without access to high performance computing due to INLA's efficiency. The optimal thresholds are quite high, suggesting that only quantiles at the 99.99% level and higher should be estimated with this method. The final quantile estimates suggests that the model captures the major spatial and temporal structures in the data. This is no small feat as inference on extreme hourly precipitation is a hard task. The model also borrows information from neighbouring stations to reduce variance at stations with short time series. This shows INLA's potential with regard to inference on threshold exceedance models.

There is evidence for issues with the model fit. Models fitted to only summer or winter data yield quantile estimates that are respectively higher and lower than what the model fitted on all the data predicts. This suggests either larger seasonal variations than the model captures or an underestimation of the uncertainty. Smoothing effects are included in each model stage, and there is evidence of bias due to too strong smoothing. Further work should investigate the bias-variance tradeoff associated with smoothing.

A drawback of the approach used is that it only provides inference on the marginals. The model cannot currently provide forecasts or simulations due to lacking information on the joint precipitation distribution of the data. Even though they only work with marginals, Castro-Camilo et al. (2018) demonstrate the possibility for use of INLA for forecasting purposes. Extending the model for use in forecasting and simulation settings is therefore a possible way forward. The model's preference for a very high threshold can also be addressed in further work. With the high threshold used, only 0.4% of the positive precipitation amounts exceeds it. Naveau, Huser, Ribereau and Hannart (2016) propose a method for joint modelling of the bulk and tail of a precipitation distribution without threshold selection and applies it successfully to hourly precipitation data. I would be interesting to see what contributions such a model could give with regard to the Norwegian precipitation data.

References

Abrahamsen, P. (1997). A review of gaussian random fields and correlation functions. Oslo: Norsk regnesentral.

Castro-Camilo, D., Huser, R. & Rue, H. (2018). A spliced gamma-generalized pareto model for short-term extreme wind speed probabilistic forecasting.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer series in statistics. London: Springer.

Cooley, D., Nychka, D. & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, *102*(479), 824–840.

Cressie, N. A. C. & Wikle, C. K. (2011). *Statistics for spatio-temporal data.* Wiley Series in Probability and Statistics. Wiley. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1091896&site=ehost-live

De Michele, C. (2003). A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of geophysical research*, *108*(D2).

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L. & Stordal, F. (2015). Bayesian hierarchical modeling of extreme hourly precipitation in norway. *Environmetrics*, *26*(2), 89–106.

Dyrrdal, A. V., Skaugen, T., Stordal, F. & Førland, E. J. (2016). Estimating extreme areal precipitation in norway from a gridded dataset. *Hydrological Sciences Journal*, *61*(3), 483–494.

Fassnacht, S. (2004). Estimating alter-shielded gauge snowfall undercatch, snowpack sublimation, and blowing snow transport at six sites in the coterminous usa. *Hydrological Processes*, *18*, 3481–3492. doi:10.1002/hyp.5806

Fuglstad, G.-A., Simpson, D., Lindgren, F. & Rue, H. (2019). Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, *114*(525), 445–452. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/01621459.2017.1415907

Lindgren, F. & Rue, H. (2015). Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, *63*(19), 1–25.

Martinez-Villalobos, C. & Neelin, J. (2019). Why do precipitation intensities tend to follow gamma distributions? *Journal of the Atmospheric Sciences*, *76*(11), 3611–3631.

Mathisen, A. (2020). Bayesian inference on extreme hourly precipitation using inla.

Naveau, P., Huser, R., Ribereau, P. & Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, *52*(4), 2753–2769.

Opitz, T., Huser, R., Bakka, H. & Rue, H. (2018). Inla goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, *21*(3), 441–462.

Rasmussen, C. E. (2006). Gaussian processes for machine learning. Cambridge, Mass.: MIT Press.

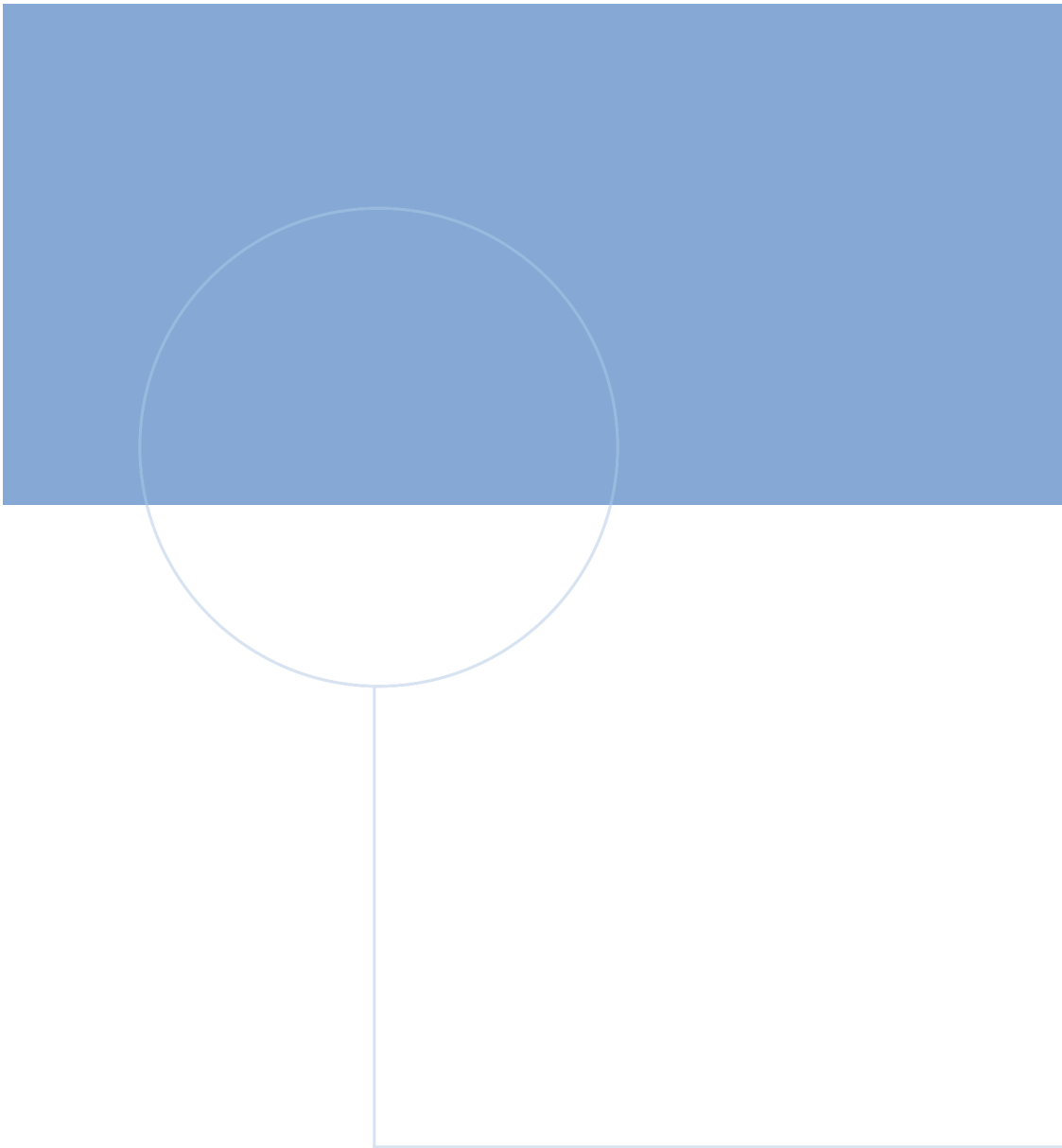
Rue, H. (2005). Gaussian markov random fields : Theory and applications. Boca Raton, Fla: Chapman & Hall/CRC.

Rue, H., Martino, S. & Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392.

Simpson, D., Rue, H., Riebler, A. I., Martins, T. G. & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors.

Sorteberg, A., Lawrence, D., Dyrrdal, A., Mayer, S., Engeland, K., Førland, E., ... Væringstad, T. (2018). Climatic changes in short duration extreme precipitation and rapid onset flooding - implications for design values. Retrieved from https://cms.met.no/site/2/klimaservicesenteret/rapporter-og-publikasjoner/_attachment/13537?_ts=163df95ff7b

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, *41*(3/4), 434–449. Retrieved from http://www.jstor.org/stable/2332724



NTNU

Norwegian University of
Science and Technology