A spatial Bayesian hierarchical model for flood frequency analysis

Árni Viðir Jóhannesson

Faculty of Physical Sciences
University of Iceland
2018
A SPATIAL BAYESIAN HIERARCHICAL MODEL
FOR FLOOD FREQUENCY ANALYSIS

Árni Víðir Jóhannesson

60 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Statistics

Advisor
Birgir Hrafnkelsson

Faculty Representative
Anna Helga Jónsdóttir

Examiner
Egil Ferkingstad

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, júní 2018
Abstract

The goal of this thesis is to model data on annual maximum series of peak flow from 553 catchments across the UK. An accurate estimate of extreme floods is of interest in many circumstances, especially concerning vital civil infrastructure. For this purpose, a latent Gaussian model with a multivariate link function for the location, scale and shape parameters of the data density is proposed for flood frequency analysis. The observations are assumed to follow the generalised extreme value (GEV) distribution. The model is inferred using the Bayesian methodology, and due to the large dimension of the data, an approximation is made to make the inference computationally feasible. Structured additive regression models are proposed for the three parameters of the data density. These regression models all contain fixed linear effects for catchment descriptors, e.g., catchment area and average annual rainfall. Two spatial components, for the location and scale parameters, are introduced in order to explain some of the otherwise unexplained variability introduced by the geographical locations of the catchments. Using this model, the quantiles of the data are investigated so that predictions can be made for the return level of a given return period T.

The results show that latent Gaussian models are a viable option for flood frequency analysis. The quantile plots for the return periods of floods show promising results, and a model of this form could prove to be useful. The results show that the spatial components for the location and scale parameter of the data density are both of importance and neither should be ignored.
Ágrip


Niðurstöður ritgerðarinnar benda til þess að líkanið sé góður kostur til þess að gera flóðagreiningar. Myndir af hlutfallsmörkum benda til þess að módelið sé vel til þess fallið að spá fyrir um stærð flóða fyrir gefinn endurkomutíma. Niðurstöður benda einnig til þess að mikilvægt sé að taka tillit til rúmhæðis staðsetningarstikans og skölunarstikans við framkvæmd flóðagreiningar.
## Contents

List of Figures xi
List of Tables xvii
Nomenclature xix
Acknowledgments xxi

1 Introduction 1
   1.1 Motivation and Goals ........................................ 1
   1.2 Literature Review ........................................... 2
   1.3 Structure of the Thesis ..................................... 3

2 Theory and Methods 5
   2.1 Bayesian Inference ......................................... 5
   2.2 Latent Gaussian Models ..................................... 6
   2.3 Integrated Nested Laplace Approximations ................. 6
   2.4 Penalised Complexity Priors ................................. 8
   2.5 Spatial Statistics ........................................... 9
      2.5.1 Gaussian Fields and Gaussian Markov Random Fields  .... 11
      2.5.2 The Stochastic Partial Differential Equation Approach . 12
   2.6 Extreme Value Theory ....................................... 13

3 Extended Latent Gaussian Models 15
   3.1 Inference Scheme for Extended LGMs ....................... 16

4 Application of an Extended LGM on Maximum Flow in Rivers 21
   4.1 Model Description ........................................... 21
      4.1.1 Data Level ............................................ 22
      4.1.2 Latent Level .......................................... 22
      4.1.3 Hyperparameter Level .................................. 23
      4.1.4 Spatial Components .................................... 24
   4.2 Inference ..................................................... 25

5 Data and Preliminary Analysis 29
   5.1 Closer Look at the MLEs .................................... 30
   5.2 A Closer Look at Four Stations ............................. 37
   5.3 Covariates .................................................... 40
## Contents

5.4 Covariate Selection .................................................. 57
  5.4.1 Covariate Selection for $\psi$ ................................ 59
  5.4.2 Covariate Selection for $\tau$ .................................. 61
  5.4.3 Covariate Selection for $\xi$ .................................. 64

6 Results ........................................................................... 67
  6.1 Convergence Assessment .............................................. 67
  6.2 Hyperparameters ......................................................... 72
  6.3 Latent Parameters ....................................................... 77
    6.3.1 The Latent Parameters in the Likelihood .................. 77
    6.3.2 Covariate Coefficients ....................................... 81
    6.3.3 Spatial Components .......................................... 84
    6.3.4 Regression Analysis ........................................... 89
  6.4 Extreme Floods ......................................................... 90
  6.5 Goodness of Fit ......................................................... 92

7 Conclusions .................................................................... 95
  7.1 Future Work ............................................................... 96

References ......................................................................... 97

A. Return Period Plots .................................................... 101
# List of Figures

4.1 Triangulated mesh over the UK. .................................................. 24

5.1 Number of observed stations as a function of year. ....................... 29

5.2 Station locations. ................................................................. 30

5.3 The densities of the MLE estimates. .......................................... 31

5.4 $\log(\hat{\sigma}_i)$ as a function of $\log(\hat{\mu}_i)$. ......................... 32

5.5 $\hat{\tau}_i = \log(\hat{\sigma}_i)$ as a function of $\hat{\psi}_i = \log(\hat{\mu}_i)$. .................................................. 32

5.6 The MLE $\hat{\psi}$ on a map of the UK. ....................................... 33

5.7 The marginal standard deviation of $\hat{\psi}$ (evaluated with the Hessian matrix of the likelihood function) on a map of the UK. ................. 33

5.8 The MLE $\hat{\tau}$ on a map of the UK. ....................................... 34

5.9 The marginal standard deviation of $\hat{\tau}$ (evaluated with the Hessian matrix of the likelihood function) on a map of the UK. ................. 34

5.10 The MLE $\hat{\xi}$ on a map of the UK. ....................................... 35

5.11 The marginal standard deviation of $\hat{\xi}$ (evaluated with the Hessian matrix of the likelihood function) on a map of the UK. ................. 35

5.12 The standard deviation of the MLEs as a function of the number of observations on the log-log scale. ........................................... 36

5.13 Time series for four selected stations. ...................................... 37

5.14 Empirical CDF for four selected stations. .................................. 38

5.15 Boxplots for four selected stations. ........................................ 39
5.16 Correlation plot for the covariates ................................................. 41
5.17 Correlation plot for LDP, DPLBAR and AREA ............................ 42
5.18 Correlation plot for PROPWET, ALTBAR, DPSBAR, SAAR and SPRHOST 42
5.19 The density of catchment area (original scale and log-scale) and the relationship between the logarithm of the catchment area and the MLEs ................. 43
5.20 The density of average annual rainfall (original scale and log-scale) and the relationship between the logarithm of the average annual rainfall and the MLEs ................................................................. 44
5.21 The density of FARL and log(FARL), and the relationship between log(FARL) and the MLEs ................................................................. 45
5.22 The density of BFIHOST and BFIHOST$^2$ and the relationship between BFIHOST$^2$ and the MLEs ................................................................. 46
5.23 The density of FPEXT and log(FPEXT), and the relationship between log(FPEXT) and the MLEs ................................................................. 47
5.24 The density of URBEXT and log(1+URBEXT), and the relationship between log(1+URBEXT) and the MLEs ................................................................. 48
5.25 The density of DPLBAR and log(DPLBAR), and the relationship between log(DPLBAR) and the MLEs ................................................................. 49
5.26 The density of DPSBAR and log(DPSBAR), and the relationship between log(DPSBAR) and the MLEs ................................................................. 50
5.27 The density of LDP and log(LDP), and the relationship between log(LDP) and the MLEs ................................................................. 51
5.28 The density of SPRHOST and log(SPRHOST), and the relationship between SPRHOST and the MLEs ................................................................. 52
5.29 The density of ASPBAR and log(ASPBAR), and the relationship between ASPBAR and the MLEs ................................................................. 53
5.30 The density of ALTBAR and log(ALTBAR), and the relationship between ALTBAR and the MLEs ................................................................. 54
5.31 The density of ASPVAR and log(ASPVAR), and the relationship between log(ASPVAR) and the MLEs ................................................................. 55
5.32 The density of PROPWET and the relationship between PROPWET and the MLEs. ........................................... 56

5.33 The lowest cross-validation error as a function of the number of covariates (including an intercept) for the $\psi$ model. .................. 60

5.34 The lowest cross-validation error as a function of the number of covariates (including an intercept) for the $\tau$ model. .................. 63

5.35 The lowest cross-validation error as a function of the number of covariates (including an intercept) for the $\xi$ model. .................. 65

6.1 Trace plots for the parameters. .................................. 69

6.2 Histogram plots for the parameters. ................................ 69

6.3 Autocorrelation plots for $\psi_8$, $\tau_8$ and $\xi_8$. The four horizontal panels represent each of the four chains. .................. 70

6.4 Autocorrelation plots for $\beta_{\psi,\text{AREA}}$, $\beta_{\tau,\text{SAAR}}$ and $\beta_{\xi,\text{FPEXT}}$. The four horizontal panels represent each of the four chains. .................. 70

6.5 Autocorrelation plots for the hyperparameters $q_\psi$, $s_\psi$ and $\rho_\tau$. The four horizontal panels represent each of the four chains. .................. 71

6.6 The joint posterior distribution of $\rho_\psi$ and $s_\psi$, and the marginal posterior densities of $\rho_\psi$ and $s_\psi$. .................. 73

6.7 The joint posterior distribution of $\rho_\tau$ and $s_\tau$, and the marginal posterior densities of $\rho_\tau$ and $s_\tau$. .................. 74

6.8 The marginal posterior density of $q_\psi$, $q_\tau$ and $q_\xi$ along with the marginal posterior density of $\sigma_{\psi,e}$, $\sigma_{\tau,e}$ and $\sigma_{\xi,e}$. .................. 75

6.9 The standard deviations of the error components and the spatial components for the $\psi$ and $\tau$ models, namely, $\sigma_{\psi,e}$, $s_\psi$, $\sigma_{\tau,e}$ and $s_\tau$. .................. 76

6.10 Posterior mean of $\psi$. .................................. 78

6.11 Posterior standard deviation of $\psi$. .................................. 78

6.12 Posterior mean of $\tau$. .................................. 79

6.13 Posterior standard deviation of $\tau$. .................................. 79
LIST OF FIGURES

6.14 Posterior mean of $\xi$. ................................. 80

6.15 Posterior standard deviation of $\xi$. ....................... 80

6.16 Posterior mean of $A_u \psi$. ............................. 86

6.17 Posterior standard deviation of $A_u \psi$. ............... 86

6.18 Posterior mean of $A_u \tau$. .............................. 87

6.19 Posterior standard deviation of $A_u \tau$. .................. 87

6.20 Posterior mean of $A_u \psi + A_u \tau$. .................... 88

6.21 Posterior standard deviation of $A_u \psi + A_u \tau$. ...... 88

6.22 The three figures on the left show $\hat{\eta}$ as a function of $Z\nu$ and the three figures on the right show a residual plot, $\hat{\eta} - Z\nu$ as a function of $Z\nu$. The three horizontal panels represent $\psi$, $\tau$ and $\xi$, respectively. ......................... 89

6.23 The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels. ....... 91

6.24 The empirical cdf and the theoretical cdf along with a 95% credible interval for four stations. ................................. 91

6.25 The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account). ................................. 92

6.26 P-values of the Anderson-Darling test assuming a GEV distribution with the posterior means of $\psi_k, \tau_k$ and $\xi_k$ for each station $k$. ............... 93

7.1 The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels. ....... 102

7.2 The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account). ................................. 102

7.3 The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels. ....... 103
7.4 The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account). . . . . . . . . . . . . . . . . . . . . . . 103

7.5 The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels. . . . . . . . . 104

7.6 The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account). . . . . . . . . . . . . . . . . . . . . . . 104

7.7 The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels. . . . . . . . . 105

7.8 The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account). . . . . . . . . . . . . . . . . . . . . . . 105

7.9 The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels. . . . . . . . . 106

7.10 The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account). . . . . . . . . . . . . . . . . . . . . . . 106
## List of Tables

5.1 Summary information for the stations. ........................................ 37
5.2 Description of the covariates (catchment descriptors). ................. 40
5.3 Correlation between log(AREA) and the MLEs. ........................... 43
5.4 Correlation between log(SAAR) and the MLEs. ............................. 44
5.5 Correlation between log(FARL) and the MLEs. ............................. 45
5.6 Correlation between BFIHOST$^2$ and the MLEs. .......................... 46
5.7 Correlation between log(FPEXT) and the MLEs. ........................... 47
5.8 Correlation between log(URBEXT$+1$) and the MLEs. ..................... 48
5.9 Correlation between log(DPLBAR) and the MLEs. ......................... 49
5.10 Correlation between log(DPSBAR) and the MLEs. ......................... 50
5.11 Correlation between log(LDP) and the MLEs. ............................. 51
5.12 Correlation between log(SPRHOST) and the MLEs. ....................... 52
5.13 Correlation between ASPBAR and the MLEs. .............................. 53
5.14 Correlation between log(ALTBAR) and the MLEs. ......................... 54
5.15 Correlation between log(ASPVAR) and the MLEs. ......................... 55
5.16 Correlation between log(PROPWET) and the MLEs. ....................... 56
5.17 Covariate selection for $\psi$. ............................................. 59
5.18 Fixed effects for $\psi$. ..................................................... 61
LIST OF TABLES

5.19 Covariate selection for $\tau$ .................................................. 62
5.20 Fixed effects for $\tau$ .............................................................. 64
5.21 Covariate selection for $\xi$ ...................................................... 64
5.22 Fixed effects for $\xi$ .............................................................. 66

6.1 Gelman-Rubin statistic for the parameters ................................. 68
6.2 Summary statistics for the posterior distribution of the hyperparameters ......................................................... 73
6.3 Covariate coefficients for $\psi$, $\tau$ and $\xi$ ................................. 81
6.4 The effect of going from the median to the 1st and 3rd quartile (0.25 quantile and 0.75 quantile), for each covariate (keeping other covariates and random effects fixed), on the location parameter, $\mu$, the scale parameter, $\sigma$, and the shape parameters, $\xi$, of the GEV density. For $\mu$ and $\sigma$ the effect is multiplicative but for $\xi$ the effect is additive ................................................................. 83
### Nomenclature

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>Scalar</td>
</tr>
<tr>
<td>$\mathbf{x}$</td>
<td>Vector</td>
</tr>
<tr>
<td>$x_i$</td>
<td>Entry $i$ in vector $x$</td>
</tr>
<tr>
<td>$\mathbf{x}_{-i}$</td>
<td>$x$ excluding $x_i$</td>
</tr>
<tr>
<td>$A$</td>
<td>Matrix</td>
</tr>
<tr>
<td>$A^{-1}$</td>
<td>Matrix inverse</td>
</tr>
<tr>
<td>$A^\top$</td>
<td>Matrix transpose</td>
</tr>
<tr>
<td>$\text{diag}(A)$</td>
<td>The diagonal elements of $A$</td>
</tr>
<tr>
<td>$\text{bdiag}(A_1, \ldots, A_n)$</td>
<td>Block diagonal matrix with $A_1, \ldots, A_n$ on the diagonal</td>
</tr>
<tr>
<td>$\pi(x)$</td>
<td>Probability density function of the random variable $X$</td>
</tr>
<tr>
<td>$\pi(x</td>
<td>y)$</td>
</tr>
<tr>
<td>$f(x) \propto g(x)$</td>
<td>There exists a constant, $a$, such that $f(x) = a \cdot g(x)$</td>
</tr>
</tbody>
</table>
Acknowledgments

First and foremost, I would like to thank my advisor Birgir Hrafnkelsson for giving me the opportunity to work on this project, which was his idea from the start. I would also like to thank him for all the help during the process of designing, programming and inferring the model used in the thesis. Birgir and I visited the University of Bath during last winter and had insightful talks with Ilaria Prosdocimi and Thomas Kjeldsen. Both Ilaria and Thomas have valuable experience in flood frequency analysis, and they are familiar with the data used in this thesis. Ilaria was also the one that pointed Birgir towards the data, and I thank her for that. I would also like to thank Haakon Bakka from Kaust University for all his help on INLA related matters.

My good friend Ottó Hólm Reynisson, who studied statistics at the same time as I, was invaluable during my graduate studies. Ottó and I took courses, taught classes, and worked at the University of Iceland together during our graduate studies. Furthermore, during the writing of my thesis, he was a valuable asset, we had long, insightful phone conversations which helped develop my thesis both in terms of content and setup.

I would also like to thank my family for their support my girlfriend María Bjarnadóttir, my mother Sólveig Þóra Jónsdóttir, my father Jóhannes Árnason, and my brother Ævar Þórar Ingi Jóhannesson. They took time out of their busy schedule to read some parts of the thesis and fix spelling and grammar mistakes. My good friend Ísak Kári Kárason also read parts of the thesis to help me regarding the correct use of the English language. Without them, the thesis would be as badly written as this chapter.
1 Introduction

1.1 Motivation and Goals

A common goal when modelling physical phenomena is to be able to say something about the return period of an extreme event such as a big flood. For this purpose, streams, wells, lakes, canals, reservoirs and other water bodies are monitored carefully using gauging stations. An accurate estimate of extreme floods is of interest in many circumstances, especially concerning vital civil infrastructure. The management and operations of these infrastructures depend on accurate flood predictions. This includes making predictions for an ungauged catchment which can be estimated from gauged river catchments.

The goal of this thesis is to model data on annual maximum series of peak flow from 553 catchments across the UK. For this purpose, a latent Gaussian model, using the Bayesian framework, with a multivariate link function as discussed in T. G. Martins et al. (2013) is introduced. This kind of model is referred to as an extended LGM in Geirsson (2015). At the latent level a generalised additive model for the location, scale and shape parameters (referred to as GAMLSS in Rigby & Stasinopoulos (2005)) of the data density, which is assumed to be a generalised extreme value (GEV) density, is proposed. The generalised additive model at the latent level contains catchment descriptors (covariates), e.g., catchment area and average annual rainfall, along with spatial components which are modelled using the SPDE approach (Lindgren et al. (2011)) are proposed for the location and scale parameters of the data density. Instead of inferring the extended LGM with the MCMC split sampler in Geirsson, Hrafnkelsson, Simpson, & Sigurðarson (2015), an approximation was made due to the high dimension of the data where the maximum likelihood estimates of the data density where essentially used as the data. Like mentioned before, there is an interest in estimating the return period of extreme events, and in particular, there is a special interest in computing the quantiles of the annual maximum flow, for example, the 0.99 quantile which, when the data is on an annual basis, is the same as the 100-year event. For this purpose, the GEV distribution is practical since the cumulative density function of the GEV distribution is invertible and thus the quantile function has an explicit expression.
1 Introduction

1.2 Literature Review

One of the first approaches in flood frequency analysis was the index flood method introduced in Dalrymple (1960). The index flood method was designed to handle cases where little or no at-site information is available by borrowing strength from similar catchments. This method is based on two steps, regionalization, which identifies geographical and climatological homogeneous regions, and a specification of a regional standardized flood frequency curve for a $T$-year return period. Various developments have been made to the index flood method, for example in J. R. Hosking et al. (1985) and GREHY (1996). Various Bayesian methods have also been introduced, for instance, Cunnane & Nash (1974), Rosbjerg & Madsen (1995), Kuczera (1999) and E. S. Martins & Stedinger (2000), but in most cases, they are derived from the index flood method. Thus, this thesis provides an uncommon approach to flood frequency analysis, where the spatial dependency of the scale parameter is considered. Davidsson (2015) proposed a Bayesian LGM for modelling monthly maxima where the proposed data density is a Gumbel distribution, a special case of the GEV, where the logarithm of the location and shape parameters are then modeled using linear regression at the latent level.

In the UK, flood frequency analysis is commonly modeled using the pooling group method, see Robson & Reed (1999), and later improved in Kjeldsen & Jones (2009b). The pooling group method is a combination of the index flood method based on the use of L-moments and annual maximum peak flow data, see J. R. M. Hosking & Wallis (2005), and the region of influence (ROI) approach, see Burn (1990). The data on annual maximum series of peak flow used in this thesis has been analysed and investigated before, for example in Kjeldsen & Jones (2006), Kjeldsen & Jones (2009a), Kjeldsen (2010) and Kjeldsen et al. (2017). Kjeldsen & Jones (2006) look at the prediction uncertainty in a median-based index flood method using L-moments and the generalised logistic distribution (GLO). Kjeldsen & Jones (2009a) suggest a recursive generalised least squares (GLS) procedure on the median of pooling groups and the parameters of the GLO distribution. Kjeldsen (2010) looked at the impact of urbanization on flood frequency relationships, using the catchment descriptor URBEXT. In this thesis, urbanization is used as a covariate at the latent level. Kjeldsen et al. (2017) proposed using the four-parameter kappa distribution. The generalised logistic (GLO), generalised extreme value (GEV), general Pareto (GPA) and Gumbel models, are all special cases of the four-parameter kappa distribution. The GEV distribution is commonly used in flood frequency analysis, for example, Stedinger & Lu (1995) and Sveinsson et al. (2003).

The extended LGM framework, where latent models are imposed on each parameter of the data density, is in line with the structured additive distributional regression approach proposed in Rigby & Stasinopoulos (2005). The extended LGM proposed in this thesis is motivated by the work of Hrafnkelsson et al. (2012) and Geirsson, Hrafnkelsson, & Simpson (2015). Hrafnkelsson et al. (2012) proposed a Bayesian hierarchical model with spatial components for modelling annual minimum and maximum temperatures in Iceland using the GEV distribution, which is also used in this thesis. In fact, Hrafnkelsson et al.
(2012) also modeled the scale parameter spatially while the shape parameter of the GEV was assumed fixed over all stations after considering a spatial component for it. Geirsson, Hrafnkelsson, & Simpson (2015) used an extended LGM model for annual maximum 24h precipitation where the setup is the same as in this thesis except for the fact that they also modeled the shape parameter of the GEV distribution spatially. The model introduced in this thesis is thus inspired by the model in Geirsson, Hrafnkelsson, & Simpson (2015), which was inferred using the MCMC split sampler presented in Geirsson, Hrafnkelsson, Simpson, & Sigurðarson (2015). Since the dimension of the data is large, an approximation had to be made to make it computationally feasible. One might then look at the work in this thesis as an approximation method for the full Bayesian posterior inference of the extended LGM model introduced in Geirsson, Hrafnkelsson, & Simpson (2015).

1.3 Structure of the Thesis

Chapter 2 describes the theory and methods used in the thesis. In Chapter 3 extended LGMs are introduced, along with the inference scheme used in the thesis. Chapter 4 presents the model that is used for the annual maximum flow data and lays out the inference scheme. In Chapter 5 a closer look is taken at the data. Chapter 6 displays the results of the model, along with a convergence assessment. In Chapter 7 conclusions drawn from this thesis are presented with suggested future steps.
2 Theory and Methods

2.1 Bayesian Inference

The Bayesian statistical framework is a methodology where unknown parameters are interpreted probabilistically, as random variables, and the data is considered fixed. For a detailed overview of Bayesian statistics see Gelman et al. (2014) or Berger (2013). The Bayesian approach differs from the classical approach in that the unknown parameters of the statistical model are treated as random variables instead of unknown constants. A Bayesian statistical model generally consists of a parametric statistical model or a sampling distribution, $\pi(y|\theta)$, to describe the data, and a prior distribution on the parameters, $\pi(\theta)$. Here $y$ represents the data and $\theta$ are the unknown parameters of the model. With the sampling distribution and the prior distribution for the unknown parameters defined, the objective of Bayesian inference is then to infer about the parameters through the posterior distribution, $\pi(\theta|y)$, using Bayes’ theorem:

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

(2.1)

where $\pi(y) = \int \pi(y|\theta)\pi(\theta)d\theta$. The density $\pi(y)$ does not depend on $\theta$ and Equation 2.1 can thus be written

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta).$$

(2.2)

Often it is easier to work with Equation 2.2 since calculating $\pi(y)$ can often be difficult. If $\theta$ is multivariate, the marginal posterior density, $\pi(\theta_j|y)$, for an unknown parameter $\theta_j$, is often of interest, for example, to obtain a credible interval, the Bayesian alternative to confidence intervals. To obtain the marginal posterior distribution, one must integrate out other parameters of the model, $\theta_{-j}$, from the posterior distribution

$$\pi(\theta_j|y) = \int \pi(\theta|y)d\theta_{-j} .$$
2 Theory and Methods

2.2 Latent Gaussian Models

Latent Gaussian Models (LGMs) are a subset of Bayesian hierarchical models in which the parameter \( x \) at the latent level has a Gaussian prior density.

An \( n \)-dimensional random vector \( x \) is said to follow a Gaussian distribution, denoted \( x \sim N(\mu, \Sigma) \), with mean \( E(x) = \mu \in \mathbb{R} \) and a positive definite covariance matrix, \( \Sigma \in \mathbb{R}^{n \times n} \) where \( \text{cov}(x_i, x_j) = \Sigma_{i,j} \) and \( \text{var}(x_i) = \Sigma_{i,i} \), if its density function is

\[
\pi(x) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \exp \left( -\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right).
\]

As discussed in Rue et al. (2009) latent Gaussian models are a subclass of structured additive regression models where the observations, \( y_i \), are assumed to have a density from the exponential family and the mean, \( \mu_i \), is then linked to a structured additive predictor, \( \eta_i \), through a link function, \( g \), such that \( g(\mu_i) = \eta_i \). The structured additive predictor, \( \eta_i \), can then account for covariates and effects in an additive way, namely,

\[
\eta_i = \alpha + \sum_j f^{(j)}(u_{ij}) + \sum_k \beta_k z_{ki} + \epsilon_i,
\]

(2.3)

where \( \{f^{(j)}(.)\} \) are unknown functions of some covariates \( u_{i1}, ..., u_{iJ} \), \( \{\beta_k\} \) are linear effects of covariates \( z_{i1}, ..., z_{iK} \) and \( \epsilon_1, ..., \epsilon_n \) are unstructured model errors. In LGMs the same form as in Equation 2.3 applies but \( \{f^{(j)}(.)\}, \{\beta_k\} \) and \( \{\epsilon_i\} \) all have a Gaussian prior distribution. Let \( x \) contain the latent parameters, namely, \( \{f^{(j)}(.)\}, \{\beta_k\} \) and \( \{\eta_i\} \), i.e., \( \{\eta_i\} \) is included instead of \( \{\epsilon_i\} \). The parameters in \( x \) all have a Gaussian prior density but they can potentially depend on hyperparameters, \( \theta = (\theta_1, ..., \theta_p)^T \), which do not necessarily have a Gaussian prior density.

For the purpose of this thesis, LGMs can be written in the same form as is done in Geirsson (2015), as follows.

**Data level:** The observations \( y \) are assumed to be dependent on the latent parameters \( x \) and have a data density \( \pi(y|x, \theta) \).

**Latent level:** The latent parameters \( x \) have a Gaussian prior density and are potentially dependent on hyperparameters \( \theta \), and have a density

\[
\pi(x|\theta) = \mathcal{N}(x|\mu(\theta), \Sigma(\theta)).
\]

**Hyperparameter level:** The hyperparameters are assigned a prior distribution \( \pi(\theta) \).

2.3 Integrated Nested Laplace Approximations

*Integrated nested Laplace approximations* (INLA) is a method for approximate Bayesian inference for LGMs that was introduced in Rue et al. (2009). Before INLA was introduced, Rue & Martino (2007) posed the question whether MCMC was needed to estimate...
2.3 Integrated Nested Laplace Approximations

marginal densities. Rue & Martino (2007) proposed using Gaussian Markov random field approximation for the inference. Like in Chapter 2.2, \( x \) will denote all parameters at the latent level that have a Gaussian prior density and \( \theta \) will represent the hyperparameters. INLA focuses on Bayesian inference for the posterior marginals of interest, which can be written as

\[
\pi(x_i | y) = \int \pi(x_i | \theta, y) \pi(\theta | y) d\theta ,
\]

\[
\pi(\theta_j | y) = \int \pi(\theta | y) d\theta_{-j} .
\]

The main feature of the INLA approach is to approximate the posterior marginals with

\[
\tilde{\pi}(x_i | y) = \int \tilde{\pi}(x_i | \theta, y) \tilde{\pi}(\theta | y) d\theta ,
\]

\[
\tilde{\pi}(\theta_j | y) = \int \tilde{\pi}(\theta | y) d\theta_{-j} .
\]

Here, \( \tilde{\pi}(\cdot | \cdot) \) represents an approximation of the density of its arguments. The approximations in Equation 2.4 are calculated by approximating \( \pi(\theta | y) \) and \( \pi(x_i | \theta, y) \) and then using numerical integration to integrate out \( \theta \) and \( \theta_{-j} \) from Equation 2.4. The INLA approach is based on approximating the marginal posterior of \( \theta \) with

\[
\tilde{\pi}(\theta | y) \propto \frac{\pi(x, \theta, y)}{\tilde{\pi}_G(x | \theta, y)} \bigg|_{x=x^*(\theta)}
\]

where \( \tilde{\pi}_G(x | \theta, y) \) is a Gaussian approximation of the full conditional of \( x \), and \( x^*(\theta) \) is the mode of the full conditional for a given value of \( \theta \). To approximate the marginals of the latent parameters in Equation 2.4, numerical integration of the form

\[
\tilde{\pi}(x_i | y) = \sum_k \tilde{\pi}(x_i | \theta_k, y) \tilde{\pi}^*(\theta_k | y) \Delta_k
\]

is used and the main use of \( \tilde{\pi}(\theta | y) \) is then to integrate out the uncertainty of \( \theta \) from Equation 2.6. To do this, it is not necessary to represent \( \tilde{\pi}(\theta | y) \) parametrically but rather explore it sufficiently so that one may choose good evaluation points for the numerical integration. In Rue et al. (2009) several steps are listed to explore \( \tilde{\pi}(\theta | y) \) sufficiently.

The simplest way to approximate \( \pi(x_i | \theta, y) \) would be to use a Gaussian approximation. Where the mean \( \mu_i(\theta) \) and the marginal variance \( \sigma_i(\theta) \) can be computed using, for example, the recursive method described in Rue et al. (2009). In fact when computing \( \tilde{\pi}(\theta | y) \) in Equation 2.5, the Gaussian approximation \( \tilde{\pi}_G(x | \theta, y) \) was computed so one only needs to compute the marginal variance.

A way to improve the Gaussian approximation is to use the Laplace approximation:

\[
\tilde{\pi}_{LA}(x_i | \theta, y) \propto \frac{\pi(x, \theta, y)}{\tilde{\pi}_G(x_{-i} | x_i, \theta, y)} \bigg|_{x_{-i}=x^*_{-i}(x_i, \theta)}.
\]
Here $\tilde{\pi}_{GG}(x \mid x_i, \theta, y)$ is the Gaussian approximation of the density $\pi(x \mid x_i, \theta, y)$ and $x^*_i(x, \theta)$ is the modal configuration. Equation 2.7 implies that $\tilde{\pi}_{GG}$ needs to be recomputed for each value of $x_i$ and $\theta$ since its precision matrix depends on both $x_i$ and $\theta$. This approximation can thus be computationally expensive and often infeasible. Rue et al. (2009) propose modifications to make these calculations feasible.

The first modification made in Rue et al. (2009) is to approximate the modal configuration with

$$x^*_i(x, \theta) \approx E_{\tilde{\pi}_G}(x \mid x_i)$$

where $E_{\tilde{\pi}_G}(x \mid x_i)$ is evaluated under the conditional density derived from the Gaussian approximation $\tilde{\pi}_G(x \mid \theta, y)$. This is computationally less expensive, firstly the conditional mean can be computed using a rank 1 update commonly known as conditioning by kriging and this rank 1 update is computed once for each $i$ since it’s linear in $x_i$.

The next modification is based on the idea that only those $x_j$ that are "close" to $x_i$ should have an effect on the marginal of $x_i$. If the dependence between $x_i$ and $x_j$ decays as the distance between nodes $i$ and $j$ increases, only those $x_j$’s that are close to $x_i$ in this sense are used to determine the marginal of $x_i$. The most important computational benefit of this is that now one only needs to factorize an $|R_i(\theta)| \times |R_i(\theta)|$ sparse matrix in the denominator of Equation 2.7, where $R_i(\theta)$ contains those indexes $j$ such that $x_j$ is close to $x_i$ for a given value of $\theta$.

For more information on INLA, the reader is encouraged to read Rue et al. (2009). The R-package R-inla is useful in practice and was used extensively in this thesis. To download and get information on the R-package the reader is referred to the website www.r-inla.org.

### 2.4 Penalised Complexity Priors

Penalised complexity priors (PC-priors) are informative priors introduced in Simpson et al. (2017). The information in these priors is specified according to four principles. **Principle 1: Occam’s razor.** Simpler model formulations should be preferred unless there is enough support for more complex model formulation. Considering, for example, a case

$$\pi(x \mid \xi), \quad \xi \geq 0$$

where $\pi(x \mid \xi = 0)$ is the base model. In this case, the prior for $\xi > 0$ should be penalised for the complexity introduced by $\xi$ and the prior should be decaying with increased complexity. The mode of the prior density should then be at the base model, $\pi(x \mid \xi = 0)$.

**Principle 2: Measure of complexity.** The Kullback-Leibler divergence (KLD) (Kullback & Leibler (1951)) is used to measure model complexity. The KLD,

$$\text{KLD}(f \mid | g) = \int f(x) \log \left( \frac{f(x)}{g(x)} \right) dx,$$
is a measure of information lost when using a base model \( g \) instead of a more flexible model \( f \). In the example described in Principle 1, \( f \) would denote a model with \( \xi > 0 \), and \( g \) would denote the base model with \( \xi = 0 \). For more detail see Simpson et al. (2017).

**Principle 3: Constant rate penalisation.** Using a unidirectional measure, \( d( f \| g ) = \sqrt{2 \text{KLD}( f \| g )} \), penalising the deviation from the base model with a constant decay-rate \( r \) parameterised by the distance \( d \)

\[
\frac{\pi(d + \delta)}{\pi(d)} = r^\delta, \quad d, \delta \geq 0
\]

for a constant \( 0 < r < 1 \). The mode is at the base model corresponding to \( d = 0 \) and a constant rate penalisation implies an exponential prior for \( d \) so for \( r = \exp(-\lambda) \), \( \pi(d) = \lambda \exp(-\lambda d) \). If \( d \) has an upper bound then a truncated exponential prior for \( d \) is used.

**Principle 4: User-defined scaling.** The user has an idea of a sensible size of the parameter of interest, \( \xi \). This principle is formulated with

\[
P(Q(\xi) > U) = \alpha,
\]

where \( Q(.) \) is some interpretable transformation of \( \xi \), \( U \) is a reference value such that the user assumes a priori that \( Q(\xi) \) can be greater than \( U \) with probability \( \alpha \).

In Simpson et al. (2017) a PC-prior for the precision, \( q \), of a Gaussian random effect is derived based on these four principles above. The prior density is

\[
\pi(q) = \frac{\lambda}{2} q^{-3/2} \exp(-\lambda q^{-1/2})
\]

and requires the user to specify \( \alpha \) and \( U \) such that \( P(1/\sqrt{q} > U) = \alpha \) where \( \lambda = -\log(\alpha)/U \). The PC-prior with density defined by Equation 2.8 is used in Chapter 4.1.3.

In Fuglstad et al. (2017) the PC-prior framework is extended to Gaussian random fields (GRFs) and in particular, a PC-prior is introduced for a GRF, \( u \), with a Matérn covariance function with marginal standard deviation \( s \), a range parameter \( \rho \) and a smoothness parameter \( \nu \). The joint prior density of \( s \) and \( \rho \) is

\[
\pi(s, \rho) = \frac{d}{2} \lambda_1 \lambda_2 \rho^{-d/2-1} \exp(-\lambda_1 \rho^{-d/2} - \lambda_2 s), \quad s > 0, \rho > 0,
\]

where \( \lambda_1 = -\log(\alpha_1) \rho_0^{d/2} \), \( \lambda_2 = -\log(\alpha_2)/s_0 \) and \( \alpha_1, \alpha_2, \rho_0 \) and \( s_0 \) are defined by the user with \( P(\rho < \rho_0) = \alpha_1 \) and \( P(s > s_0) = \alpha_2 \). For a two-dimensional GRF, \( d = 2 \) in Equation 2.9. This prior density is used in Chapter 4.1.3.

**2.5 Spatial Statistics**

Research in many fields such as climatology, ecology, environmental health and real estate are often faced with the task of modelling and analysing geographically referenced data,
where the relative position of observations in space has to be taken into account, see Banerjee et al. (2014). When modelling spatially referenced data, it is often inadequate to ignore the position of observations in space because that information will often explain some of the variability that would otherwise be unexplained. When analysing data that is spatially referenced the case can arise, and usually does, where observations close together are highly dependent, and observations farther apart are less dependent. In those cases using the spatial information of an observation could be used to borrow strength to parameters corresponding to nearby locations.

For this thesis, the focus will be on point-level data where the location \( s \in D \) varies continuously over \( D \) where \( D \) can, for example, be \( D = \mathbb{R}^d \) where \( d \in \mathbb{N} \) or a subset of \( \mathbb{R}^d \).

The data can then be thought of, partially or fully, a realization of a spatial random field (random process or stochastic process),

\[
\{u(s) : s \in D\}, \tag{2.10}
\]

see Cressie (1993) for a more detailed definition. For statistical analysis of spatially referenced data with locations \( s_1, \ldots, s_n \in D \) the random vector \( u \) defined as \( u = (u(s_1), \ldots, u(s_n)) \) is used, where \( u(s_k) \) is a random variable corresponding to the location \( s_k \) where \( k \in \{1, \ldots, n\} \). The random vector \( u \) inherits the dependence structure of the spatial field \( \{u(s)\} \). The dependence structure can be complicated and often defined by a function \( C : D \times D \rightarrow \mathbb{R} \) which satisfies

\[
C(s, r) = \text{cov}(u(s), u(r)), \quad \forall s, r \in D,
\]

furthermore, for the random vector \( u \), the covariance matrix can be defined by

\[
\Sigma_{i,j} = C(s_i, s_j), \quad \forall i, j \in \{1, \ldots, n\}.
\]

The function \( C(\cdot, \cdot) \) is then called a covariance function. In this thesis, the focus will be on second-order (or weak) stationary spatial fields which are defined as follows.

**Definition 1.** A spatial field \( \{u(\cdot)\} \) defined on \( D \) with a constant mean

\[
\text{E}(u(s)) = \mu, \quad \forall s \in D,
\]

its variance, \( \text{var}(u(s)) \), exists for all \( s \in D \) and the covariance between two locations \( s \) and \( r \) only depends on \( s - r \), i.e.,

\[
\text{cov}(u(s), u(r)) = C(s - r),
\]

is called a **weak stationary spatial field**.

Furthermore, if the covariance only depends on the distance \( ||s - r|| \), that is

\[
\text{cov}(u(s), u(r)) = C(||s - r||), \quad \forall s, r \in D,
\]

and \( ||\cdot|| \) denotes a Euclidian distance, then the covariance function is referred to as an **isotropic** covariance function. There are some common examples of isotropic covariance functions, e.g., the exponential covariance function defined as

\[
C(s, r) = \sigma^2 \exp(-\phi ||s - r||), \quad \forall s, r \in D, \tag{2.11}
\]
where $\sigma^2 > 0$ is the marginal variance and $\phi > 0$ is a decay parameter, $\phi$ tunes how rapidly the spatial dependency decays as a function of distance. The exponential covariance function is easily interpretable. The Matérn covariance function, see for example Banerjee et al. (2014), can be seen as an extension of the exponential covariance function, (2.11), in that it allows for adjusting both the decay and the smoothness of the spatial field. In this thesis the Matérn covariance function is used, see Chapter 4.1.3, to control the dependency structure in the spatial components. The Matérn covariance function has the form

$$C(s, r) = \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} \kappa^{\nu \theta} K_\nu(\kappa ||s - r||),$$

(2.12)

where $\Gamma$ is the gamma function, $K_\nu$ is the modified Bessel function of the second kind and order $\nu > 0$, $\sigma^2$ is the marginal variance of the field, $\kappa > 0$ is a scaling parameter and $\nu$ is a smoothness parameter. The Matérn covariance function can also be written in terms of the range parameter, $\rho$, where

$$\rho = \sqrt{\frac{8\nu}{\kappa}},$$

and makes $\rho$ the distance at which the correlation is approximately 0.1.

### 2.5.1 Gaussian Fields and Gaussian Markov Random Fields

**Definition 2.** A Gaussian field (GF) is a spatial field, $\{u(s) : s \in \mathcal{D}\}$ with a mean and covariance functions

$$\mu(s) = \mathbb{E}(u(s)), \quad C(s, r) = \text{cov}(u(s), u(r)), \quad \forall s, r \in \mathcal{D}$$

respectively, if for any sites $s_1, ..., s_n \in \mathcal{D}$ the random vector $u = (u(s_1), ..., u(s_n))$ follows a multivariate Gaussian distribution. Denoted

$$u \sim \mathcal{N}(\mu, \Sigma),$$

where $\mu = (\mu(s_1), ..., \mu(s_n))$ and $\Sigma$ is such that $\Sigma_{i,j} = C(s_i, s_j)$.

One of the disadvantages of using covariance functions to define the dependency between two random variables at different locations is known as "the big $n$ problem", see Banerjee et al. (2014). This problem arises when using algebraic operations on a dense covariance matrix, $\Sigma$, for example, the calculations of the determinant and the inverse, by solving a linear $n \times n$ system. For a Gaussian field with $n$ locations the algebraic operations scale as $O(n^3)$. When analysing data of large dimension, the problem can become computationally infeasible. However, approximating a GF with a Gaussian Markov random field (GMRF) can lead to significantly faster computations, see Rue (2001) and Rue & Held (2005).

For a GF, $\{u(s), s \in \mathcal{D}\}$, and locations $s_1, ..., s_n \in \mathcal{D}$ and a vector $u$ where the $i$-th element of $u$ is $u_i = u(s_i)$, a GMRF approximation of that GF can be specified using a neighbourhood structure, $\mathcal{N}(i)$, such that $\mathcal{N}(i)$ contains all indexes $j \in \{1, ..., n\}$ such that the random variables $u_i | u_{-ij}$ and $u_j | u_{-ij}$ are dependent.

**Definition 3.** A Gaussian Markov random field is a Gaussian random field, $u \sim \mathcal{N}(\mu, Q^{-1})$ where $u$ is $n$-dimensional, such that the conditional densities of $u_i$ for all $i \in \{1, ..., n\}$,
are such that
\[ \pi(u_i | \{u_j : j \neq i\}) = \pi(u_i | \{u_j : j \in \mathcal{N}(i)\}) \]
for neighbourhood sets \( \{\mathcal{N}(i) : i \in \{1, ..., n\}\} \).

The reason for using the precision matrix, \( Q \), rather than the covariance matrix, \( \Sigma = Q^{-1} \), in the definition of a GMRF is because the precision matrix for a GMRF is sparse, in particular, the only elements of \( Q \) that are non-zero are the diagonal elements and elements \( Q_{ij} \) where \( i \) and \( j \) are neighbours. This statement can be written as

\[ Q_{ij} \neq 0 \iff j \in \{i, \mathcal{N}(i)\}. \]

This, in turn, makes calculating the determinant and the inverse faster, and it scales like \( O(n^{3/2}) \) see Simpson et al. (2012).

### 2.5.2 The Stochastic Partial Differential Equation Approach

The Stochastic Partial Differential Equation (SPDE) approach is a method developed by Lindgren et al. (2011) to parametrize the precision matrix of a GMRF to get a predefined covariance structure with sparse precision matrices. The SPDE approach is based on the relationship that a GF, \( \{u(s) : s \in \mathbb{R}^d\} \), with the Matérn covariance function, (2.12), is a solution to the stochastic partial differential equation

\[ (\kappa^2 - \Delta)^{\alpha/2}u(s) = \mathcal{W}(s), \quad s \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \nu, \kappa > 0, \quad (2.13) \]

where \( \mathcal{W}(s) \) is a white noise process, and \( \Delta \) is the Laplacian. The marginal variance of the solution to (2.13) is

\[ s^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}}. \]

Lindgren et al. (2011) propose constructing a discrete approximation of the continuous field, \( \{u(s) : s \in \mathbb{R}^d\} \), using basis functions, \( \{\psi_k\} \), and weights, \( \{w_k\} \), on a triangulated mesh with \( n \) vertices such that

\[ u(s) = \sum_{k=1}^n \psi_k(s)w_k \]

and then find the distribution of \( w_k \) by solving (2.13). Lindgren et al. (2011) suggest a piecewise linear basis function, \( \{\psi_k\} \), such that \( \psi_k \) is one at vertex \( k \) but zero at all other vertices. The weights \( \{w_k\} \) then control the size of the field at each vertex. A location inside a triangle is then determined by linear interpolation of the triangle vertices that contains the location. This approach yields a continuously indexed approximate solution to Equation 2.13. This means that for a finite set of locations there exists a linear transformation, \( A \), from the vertices of the triangulated mesh to the locations of interest.
2.6 Extreme Value Theory

To solve Equation 2.13 Lindgren et al. (2011) first consider a weak solution given by

$$\left\{ \langle \phi_j, (\kappa^2 - \Delta)^{\alpha/2} u \rangle \right\}_{j=1}^n = \left\{ \langle \phi_j, W \rangle \right\}_{j=1}^n$$

(2.14)

where $\langle \cdot, \cdot \rangle$ is defined, for $f$ and $g$, as

$$\langle f, g \rangle = \int_{\mathbb{R}^d} f(s)g(s)ds$$

and $\{ \phi_j : j \in \{1, \ldots, n\} \}$ are test functions which can be chosen so that (2.14) holds. By replacing $u$ with $\sum_{k=1}^n \psi_k w_k$ in (2.14), defining the matrices

$$C_{ij} = \langle \psi_i, \psi_j \rangle,$$
$$G_{ij} = \langle \Delta \psi_i, \Delta \psi_j \rangle,$$
$$(K_{\kappa^2})_{ij} = \kappa^2 C_{ij} + G_{ij},$$

(2.15)

selecting the test functions in the same way as in Lindgren et al. (2011) and using Neuman boundary conditions then the precision matrices, $Q_{\alpha,\kappa^2}$, for the Gaussian weights, $w = (w_1, \ldots, w_n)$, for $\alpha = 1, 2, \ldots$ are given by

$$Q_{1,\kappa^2} = K_{\kappa^2}$$
$$Q_{2,\kappa^2} = K_{\kappa^2} C^{-1} K_{\kappa^2}$$
$$Q_{\alpha,\kappa^2} = K_{\kappa^2} C^{-1} Q_{\alpha-2,\kappa^2} C^{-1} K_{\kappa^2}, \quad \alpha = 3, 4, \ldots$$

(2.16)

Now the values of $C$ and $G$ are easy to compute since they are non-zero only for pairs of basis functions that share a common triangle and they are not a function of $\kappa$, thus they don’t need to be recomputed, for example, in each iteration of an MCMC simulation. The matrix $C^{-1}$ can be dense, and thus computationally demanding. However, Lindgren et al. (2011) show that $C$ can be replaced with a diagonal matrix $\tilde{C}$ defined as $\tilde{C}_{ii} = \langle \psi_i, 1 \rangle$, which makes the precision matrices sparse, and makes $w$ a GMRF.

2.6 Extreme Value Theory

A classic statistical problem when considering a random sample $\{X_i, 1 \leq i \leq n\}$ deals with the mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ this problem is often based on the central limit theorem and often returns to the normal distribution as a basis of inference. The classical central limit theorem (CLT) states that the random variable

$$\sqrt{n} \left( \frac{\bar{X}_n - \mathbb{E}(X)}{\sqrt{\text{var}(X)}} \right)$$

converges to a standard normal distribution when $n \to \infty$. More generally the CLT deals with the sum $S_n = X_1 + \ldots + X_n$ with the goal of finding constants $a_n$ and $b_n$ such that the random variable $Y_n = \frac{S_n - b_n}{a_n}$ converges in distribution to a non-degenerate distribution.
A similar limit approach can be taken when looking at the sample maximum instead of the sample mean. Instead of looking at the limit of the sum \( S_n \), the sample maximum will be investigated, \( X_{(n),n} = \max\{X_1, X_2, ..., X_n\} \). It turns out that the limiting distribution of the sample maximum,

\[
P\left( \frac{X_{(n),n} - b_n}{a_n} \leq x \right) \rightarrow G(x),
\]

is

\[
G(x) = \exp(-(1 + \gamma x)^{-\gamma}), \quad \text{for } 1 + \gamma x > 0.
\]

For more details and proofs see Beirlant et al. (2006). In this thesis, a distribution of this form will be used for the likelihood, i.e., the general extreme value distribution (GEV). The cumulative density function of the GEV distribution is

\[
G(x) = \exp\left( -\left( 1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right)
\]

for \( 1 + \xi \frac{x - \mu}{\sigma} > 0 \) else \( G(x) = 0 \). The GEV distribution has three parameters \( \mu, \sigma \) and \( \xi \) which are location, scale, and shape parameters, respectively. The shape parameter, \( \xi \), controls the tail behavior of the GEV. For \( \xi > 0 \) the GEV distribution has a lower bound, with the support \( x > \mu - \sigma / \xi \), for \( \xi < 0 \) it has an upper bound, with support \( x < \mu - \sigma / \xi \).

In the special case of \( \xi = 0 \), Equation 2.17 is not defined and the cumulative density is replaced by the limit of Equation 2.17 as \( \xi \rightarrow 0 \),

\[
G(x) = \exp\left( -\exp\left( \frac{x - \mu}{\sigma} \right) \right).
\]

This is the Gumbel distribution which is defined for all \( x \in \mathbb{R} \).
3 Extended Latent Gaussian Models

Just like the latent Gaussian models discussed in Chapter 2.2 and defined in Rue et al. (2009), extended latent Gaussian models are such that parameters defined at the latent level follow a Gaussian distribution. Extended LGMs are such that within each group the observations are assumed to come from a distribution that contains two or more parameters. These parameters are then modeled at the latent level with a linear model, while in the case of the LGMs defined in Rue et al. (2009) only one parameter of the data distribution is modeled at the latent level. In fact, here the focus is on distributions with three parameters, namely, a location parameter, a scale parameter and a shape parameter in line with the generalised additive model for location, scale and shape (GAMLSS) as discussed in Rigby & Stasinopoulos (2005). Furthermore, it is assumed that these three parameters are mapped to three latent parameters through a multivariate link function. Each of these latent variables is modeled with a linear model. It is assumed that within each group \( i \) there are \( n_i \) observations. The total number of groups is \( J \) and the largest number of observations per group is \( T \). Observations from different groups are assumed to be independent conditional on the latent process. Furthermore, independence is assumed between observations from the same group conditional on the latent parameters of the group. The groups can represent various types of sampling set-ups. Here the focus will be on geographical sites observed over time, i.e., the sites are the groups.

The probability density function of observation \( y_{it} \) from site \( i \) at time \( t \) is denoted by \( \pi(y_{it} | \mu_i, \sigma_i, \xi_i) \) where \( \mu_i, \sigma_i > 0 \) and \( \xi_i \) are location, scale and shape parameters, respectively. Let \( y \) be a vector containing all the observations. Let

\[
\boldsymbol{\mu} = (\mu_1, ..., \mu_J)^T, \quad \boldsymbol{\sigma} = (\sigma_1, ..., \sigma_J)^T, \quad \boldsymbol{\xi} = (\xi_1, ..., \xi_J)^T,
\]

be vectors, containing the \( J \) location, scale and shape parameters, respectively. Conditional on \( \boldsymbol{\mu}, \boldsymbol{\sigma} \) and \( \boldsymbol{\xi} \), the probability density function of \( y \) is given by

\[
\pi(y | \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}) = \prod_{i=1}^{J} \prod_{t \in A_i} \pi(y_{it} | \mu_i, \sigma_i, \xi_i),
\]

where \( A_i \) is an index set for site \( i \). A link function \( g : \mathbb{R}^3 \to \mathbb{R}^3 \) such that \( g(\mu_i, \sigma_i, \xi_i) = (\psi_i, \tau_i, \xi_i) \) is proposed such that \( \psi_i, \tau_i \) and \( \xi_i \) are on the real line. A linear model for the parameters

\[
\boldsymbol{\psi} = (\psi_1, ..., \psi_J)^T, \quad \boldsymbol{\tau} = (\tau_1, ..., \tau_J)^T \text{ and } \boldsymbol{\xi} = (\xi_1, ..., \xi_J)^T
\]
3 Extended Latent Gaussian Models

is specified at the latent level, namely,

\[
\begin{align*}
\psi &= X_\psi \beta_\psi + A_\psi u_\psi + \epsilon_\psi, \\
\tau &= X_\tau \beta_\tau + A_\tau u_\tau + \epsilon_\tau, \\
\xi &= X_\xi \beta_\xi + A_\xi u_\xi + \epsilon_\xi,
\end{align*}
\]

(3.1)

where \( \beta_\psi, \beta_\tau \) and \( \beta_\xi \) are fixed effects, \( X_\psi, X_\tau \) and \( X_\xi \) are the corresponding design matrices containing covariate information, \( u_\psi, u_\tau \) and \( u_\xi \) are random effects and \( A_\psi, A_\tau \) and \( A_\xi \) are their corresponding matrices and \( \epsilon_\psi, \epsilon_\tau \) and \( \epsilon_\xi \) are independent error terms, referred to as model errors.

### 3.1 Inference Scheme for Extended LGMs

Assuming the model described in Chapter 3, let

\[
\eta = (\psi^T, \tau^T, \xi^T)^T,
\]

\[
\nu = (\beta_\psi^T, u_\psi^T, \beta_\tau^T, u_\tau^T, \beta_\xi^T, u_\xi^T)^T,
\]

\[
\epsilon = (\epsilon_\psi^T, \epsilon_\tau^T, \epsilon_\xi^T)^T.
\]

A priori the vectors \( \beta_\psi, u_\psi, \beta_\tau, u_\tau, \beta_\xi \) and \( u_\xi \) are assumed to be independent. Denote the means and precisions matrices of these six vectors by \( \mu_{\beta_\psi}, \mu_{u_\psi}, \mu_{\beta_\tau}, \mu_{u_\tau}, \mu_{\beta_\xi}, \mu_{u_\xi}, Q_{\beta_\psi}, Q_{u_\psi}, Q_{\beta_\tau}, Q_{u_\tau}, Q_{\beta_\xi}, Q_{u_\xi} \). The prior mean of \( \nu \) is

\[
\mu_\nu = (\mu_{\beta_\psi}^T, \mu_{u_\psi}^T, \mu_{\beta_\tau}^T, \mu_{u_\tau}^T, \mu_{\beta_\xi}^T, \mu_{u_\xi}^T)^T.
\]

The covariance matrix and the precision matrix of \( \nu \) are block diagonal matrices of the form

\[
\Sigma_\nu = \text{bdiag}(Q_{\beta_\psi}^{-1}, Q_{u_\psi}^{-1}, Q_{\beta_\tau}^{-1}, Q_{u_\tau}^{-1}, Q_{\beta_\xi}^{-1}, Q_{u_\xi}^{-1}),
\]

\[
Q_\nu = \text{bdiag}(Q_{\beta_\psi}, Q_{u_\psi}, Q_{\beta_\tau}, Q_{u_\tau}, Q_{\beta_\xi}, Q_{u_\xi}).
\]

The precision matrices of \( \epsilon_\psi, \epsilon_\tau \) and \( \epsilon_\xi \) are diagonal matrices that are denoted by \( Q_{\epsilon_\psi}, Q_{\epsilon_\tau} \) and \( Q_{\epsilon_\xi} \). The covariance matrix and the precision matrix of \( \epsilon \) are given by

\[
\Sigma_\epsilon = \text{bdiag}(Q_{\epsilon_\psi}^{-1}, Q_{\epsilon_\tau}^{-1}, Q_{\epsilon_\xi}^{-1}),
\]

\[
Q_\epsilon = \text{bdiag}(Q_{\epsilon_\psi}, Q_{\epsilon_\tau}, Q_{\epsilon_\xi}).
\]

Define the matrix \( Z \) based on \( X_\psi, A_\psi, X_\tau, A_\tau, X_\xi \) and \( A_\xi \) as

\[
Z = \begin{pmatrix}
X_\psi & A_\psi & \cdot & \cdot & \cdot \\
\cdot & \cdot & X_\tau & A_\tau & \cdot \\
\cdot & \cdot & \cdot & \cdot & X_\xi & A_\xi
\end{pmatrix}
\]

where the dots denote zero entries. It is easy to see with simple matrix multiplication that Equation 3.1 can be written as

\[
\eta = Z\nu + \epsilon.
\]
Since the model is an extended LGM then the prior densities of $\nu$ and $\eta|\nu$ are Gaussian. Furthermore, the precision matrices $Q_{\beta\psi}$, $Q_{\beta\tau}$ and $Q_{\beta\xi}$ are assumed to be fixed while hyperparameters govern the precision matrices for the random effects $u_\psi$, $u_\tau$ and $u_\xi$, i.e., $Q_{u\psi}$, $Q_{u\tau}$ and $Q_{u\xi}$. If one of the random effects is, for example, a spatial component defined by using the SPDE approach described in Chapter 2.5.2 then its precision matrix is sparse. The precision matrices for the model errors $\epsilon_\psi$, $\epsilon_\tau$ and $\epsilon_\xi$ are diagonal matrices

$$Q_{\epsilon\psi} = \frac{1}{\sigma_{\epsilon\psi}}I, \quad Q_{\epsilon\tau} = \frac{1}{\sigma_{\epsilon\tau}}I, \quad Q_{\epsilon\xi} = \frac{1}{\sigma_{\epsilon\xi}}I,$$

where $\sigma_{\epsilon\psi}^2$, $\sigma_{\epsilon\tau}^2$ and $\sigma_{\epsilon\xi}^2$ are the variances of $\epsilon_\psi$, $\epsilon_\tau$ and $\epsilon_\xi$, respectively. The posterior density of $(\eta, \nu, \theta)$ is

$$\pi(\eta, \nu, \theta|y) \propto \pi(\theta)\pi(\eta, \nu|\theta)\pi(y|\eta)$$

where $y$ denotes the data. When $\pi(y|\eta)$ is used for inference then it is referred to as the likelihood function and viewed as a function of $\eta$. An approximation for the posterior density of $(\eta, \nu, \theta)$ is introduced, and it is based on approximating the likelihood function with a Gaussian density. Let $L(\eta|y)$ denote the likelihood function, where

$$L(\eta|y) = \pi(y|\eta)$$

and let $\hat{L}$ denote the Gaussian approximation of $L$ that is, $\hat{L}(\eta|y) \approx L(\eta|y)$, where

$$\hat{L}(\eta|y) = \mathcal{N}(\eta|\hat{\eta}, \Sigma_{ny}),$$

where $\hat{\eta}$ is the mode of $\log(L(\eta|y))$, i.e., the maximum likelihood estimate for $\eta$. $H_{ny}$ is the Hessian matrix of $\log(L(\eta|y))$ evaluated at $\eta = \hat{\eta}$ and $\Sigma_{ny} = (-H_{ny})^{-1}$. Then the approximated posterior density is given by

$$\pi(\eta, \nu, \theta|y) \approx \tilde{\pi}(\eta, \nu, \theta|y)$$

$$\propto \pi(\theta)\pi(\eta, \nu|\theta)\tilde{L}(\eta|y)$$

$$\propto \pi(\theta)\pi(\eta, \nu|\theta)\mathcal{N}(\eta|\hat{\eta}, \Sigma_{ny}).$$

Now take a look at a model that is such that $\hat{\eta}$ is treated as the data. The proposed data density is $\mathcal{N}(\hat{\eta}|\eta, Q_{ny}^{-1})$ where $Q_{ny} = \Sigma_{ny}^{-1}$, and $Q_{ny}$ is assumed to be known. This model can be written hierarchically as

$$\pi(\hat{\eta}|\eta, Q_{ny}, \theta) = \mathcal{N}(\hat{\eta}|\eta, Q_{ny}^{-1}),$$

$$\pi(\eta|\nu, \theta) = \mathcal{N}(\eta|\mu_\nu, Q_{\nu}^{-1}),$$

$$\pi(\nu|\theta) = \mathcal{N}(\nu|\mu_\nu, Q_{\nu}^{-1})$$

and the posterior density of this model is given by

$$\pi(\eta, \nu, \theta|\hat{\eta}) \propto \pi(\theta)\pi(\eta, \nu|\theta)\pi(\hat{\eta}|\eta, Q_{ny}, \theta)$$

$$\propto \pi(\theta)\pi(\eta, \nu|\theta)\mathcal{N}(\hat{\eta}|\eta, Q_{ny}^{-1})$$

$$\propto \pi(\theta)\pi(\eta, \nu|\theta)\mathcal{N}(\eta|\hat{\eta}, \Sigma_{ny}).$$

17
3 Extended Latent Gaussian Models

The above posterior density stems from looking at it in terms of \( \eta \) and taking \( \hat{\eta} \) as a fixed quantity, and that gives

\[
N(\hat{\eta}|\eta, Q_{\eta|\eta}^{-1}) = N(\eta|\hat{\eta}, Q_{\eta|\hat{\eta}}^{-1}).
\]

So, the above posterior density is exactly the same as the approximated posterior density of the previous model. The latter model is a Gaussian-Gaussian model and it is preferable to approach the inference for the unknown parameters through this model since the conditional posterior density of \((\eta, \nu)\) given \(\theta\) is Gaussian. To tackle the inference for \((\eta, \nu, \theta)\) the joint prior density of \(x = (\eta^T, \nu^T)^T\) conditioned on \(\theta\) is derived. It is given by

\[
\pi(x|\theta) = N\left(x \left| \begin{bmatrix} Z\mu_\nu \\ \mu_\nu \end{bmatrix}, \begin{bmatrix} Q_\epsilon^{-1} + ZQ_\nu^{-1}Z^T & ZQ_\nu^{-1} \\ ZQ_\nu^{-1} & Q_\nu^{-1} \end{bmatrix} \right. \right).
\]

The covariance matrix of \(x\) given \(\theta\) is

\[
\Sigma_x = Q_x^{-1} = \begin{bmatrix} \Sigma_\eta \Sigma_\nu \\ Q_\nu^{-1}Z^T \\ Q_\nu^{-1} \end{bmatrix}.
\]

Let \(p = \dim(\eta) = 3J\) and \(m = \dim(\nu)\) and \(B = [I_{p \times p} \ 0_{p \times m}]\). Because one has \(\hat{\eta} = Bx + e\) where \(e \sim N(0, Q_{\eta|\eta}^{-1})\), the joint density of \((\hat{\eta}^T, \nu^T)^T = (\hat{\eta}^T, \eta^T, \nu^T)^T\) is given by

\[
\pi\left(\begin{bmatrix} \hat{\eta} \\ x \end{bmatrix}\right) = N\left(\begin{bmatrix} \hat{\eta} \\ x \end{bmatrix} \left| \begin{bmatrix} B\mu_\nu \\ \mu_\nu \end{bmatrix}, \begin{bmatrix} Q_{\eta\gamma}^{-1} + BQ_\nu^{-1}B^T & BQ_\nu^{-1} \\ BQ_\nu^{-1} & Q_\nu^{-1} \end{bmatrix} \right. \right).
\]

The mean vector and the covariance matrix of \((\hat{\eta}, x)\) can be written as

\[
\begin{bmatrix} Z\mu_\nu \\ \mu_\nu \end{bmatrix}, \begin{bmatrix} Q_\nu^{-1} + Q_\nu^{-1}Z^TQ_\nu^{-1} & Q_\nu^{-1}Z^TQ_\nu^{-1} \nu \\ Q_\nu^{-1}Z^TQ_\nu^{-1} & Q_\nu^{-1}Z^TQ_\nu^{-1} \nu \end{bmatrix}.
\]

The precision matrix of \((\hat{\eta}, x)\) is

\[
Q_{\hat{\eta}x} = \Sigma_{\hat{\eta}x}^{-1} = \begin{bmatrix} \Sigma_{\eta\gamma}^{-1} & -Q_{\eta\gamma}B \\ -B^TQ_{\eta\gamma} & Q_x + B^TQ_{\eta\gamma}B \end{bmatrix}
\]

and it can be written as

\[
Q_{\hat{\eta}x} = \begin{bmatrix} \Sigma_{\eta\gamma}^{-1} & -Q_{\eta\gamma}B \\ -Q_{\eta\gamma} & Q_x + B^TQ_{\eta\gamma}B \end{bmatrix}.
\]

The marginal density of \(\hat{\eta}\) (given \(\theta\)) is given by

\[
\pi(\hat{\eta}|\theta) = N(\hat{\eta}|B\mu_\nu, Q_{\eta\gamma}^{-1} + BQ_\nu^{-1}B^T) = N(\hat{\eta}|Z\mu_\nu, Q_{\eta\gamma}^{-1} + Q_\nu^{-1}Z^TQ_\nu^{-1}Z^T).
\]

The marginal posterior density of \(\theta\) given \(\hat{\eta}\) is

\[
\pi(\theta|\hat{\eta}) \propto \pi(\theta)\pi(\hat{\eta}|\theta)
\]
and it can be represented as
\[
\pi(\theta | \hat{\eta}) \propto \pi(\theta) \pi(\hat{\eta} | \theta) \propto \pi(\theta) \frac{\pi(\hat{\eta} | x, \theta) \pi(x | \theta)}{\pi(x | \theta, \hat{\eta})},
\]
where \( \pi(\hat{\eta} | x, \theta) \) and \( \pi(x | \theta) \) have precision matrices \( Q_{xy} \) and \( Q_x \). The precision matrix of \( \pi(x | \theta, \hat{\eta}) \) is given below. The density \( \pi(\theta | \hat{\eta}) \) can be evaluated with any value for \( x \) because it does not depend on \( x \). Thus, \( x \) should be set equal to a convenient value, e.g., \( x = 0 \). The conditional posterior density of \( x \) given \( \hat{\eta} \) and \( \theta \) is
\[
\pi(x | \hat{\eta}, \theta) = \mathcal{N}(x | Q_{x|\hat{\eta}}^{-1}(Q_x \mu_x + B^T Q_{y|\hat{\eta}} \hat{\eta}), Q_x^{-1}),
\]
where
\[
Q_{x|\hat{\eta}} = Q_x + B^T Q_{y|\hat{\eta}} B = \begin{bmatrix}
Q_\epsilon + Q_{y\epsilon} & -Q_\epsilon Z \\
-Z^T Q_\epsilon & Q_\nu + Z^T Q_\epsilon Z
\end{bmatrix}
\]
and
\[
Q_x \mu_x + B^T Q_{y|\hat{\eta}} \hat{\eta} = \begin{bmatrix}
Q_{y\epsilon} \\ Q_{\nu} \mu_\nu
\end{bmatrix}.
\]
To sample from the posterior density of \( (x, \theta) \), first, a sample from \( \pi(\theta | \hat{\eta}) \), which is an approximation of \( \pi(\theta | y) \), is taken. Then a sample from \( \pi(x | \hat{\eta}, \theta) \), which is an approximation of \( \pi(x | y, \theta) \), is taken.

In order to gain insight into how \( \hat{\eta} \) is used to update our prior knowledge of \( \eta \) and \( \nu \) then the conditional distribution of \( \eta \) given \( \hat{\eta} \) is explored as well as the conditional distribution of \( \nu \) given \( \hat{\eta} \). The density of the first distribution can be found by using the fact that
\[
\pi(\eta | \hat{\eta}) \propto \pi(\eta) \pi(\hat{\eta} | \eta)
\]
where \( \pi(\hat{\eta} | \eta) = \mathcal{N}(\hat{\eta} | \eta, Q_{\hat{\eta}|\eta}^{-1}) \) and \( \pi(\eta) = \mathcal{N}(\eta | Z \mu_\nu, Q_\nu^{-1} + Z Q_\nu^{-1} Z^T) \). The result is a Gaussian density with mean \( \mu_{\hat{\eta}|\eta} \) and precision matrix \( Q_{\hat{\eta}|\eta} \), where
\[
\mu_{\hat{\eta}|\eta} = Q_{\hat{\eta}|\eta}^{-1}((Q_\epsilon^{-1} + Z Q_\nu^{-1} Z^T)^{-1} Z \mu_\nu + Q_{y|\hat{\eta}} \hat{\eta}), \quad Q_{\hat{\eta}|\eta} = (Q_\epsilon^{-1} + Z Q_\nu^{-1} Z^T)^{-1} + Q_{y|\hat{\eta}}.
\]
It is clear from the above formula for the posterior mean \( \mu_{\hat{\eta}|\eta} \), that it is a linear combination of the prior mean \( Z \mu_\nu \) and the maximum likelihood estimates in \( \hat{\eta} \). The posterior mean stems from Bayesian regression where the design matrix is an identity matrix, the error precision matrix is \( Q_{y|\hat{\eta}} \), and the prior precision matrix is \( Q_\eta = (Q_\epsilon^{-1} + Z Q_\nu^{-1} Z^T)^{-1} \). The prior precision matrix assigns weights for the prior mean \( Z \mu_\nu \) while the precision matrix \( Q_{y|\hat{\eta}} \) assigns weights to \( \hat{\eta} \) (the larger the weights in \( Q_{y|\hat{\eta}} \) the greater is the influence of \( \hat{\eta} \) on the posterior mean of \( \eta \)). If the prior mean of \( \mu_\nu \) is equal to zero then the posterior mean of \( \eta \) is
\[
\mu_{\hat{\eta}|\eta} = Q_{\hat{\eta}|\eta}^{-1} Q_{y|\hat{\eta}} \hat{\eta} = (Q_{y|\hat{\eta}}^{-1}(Q_\epsilon^{-1} + Z Q_\nu^{-1} Z^T)^{-1} + I)^{-1} \hat{\eta} = (Q_{y|\hat{\eta}}^{-1} Q_\eta + I)^{-1} \hat{\eta}
\]
so, if the elements of \( Q_{y|\hat{\eta}} \) are large with respect to the elements of \( Q_\eta \) the closer \( \mu_{\hat{\eta}|\eta} \) will be to \( \hat{\eta} \).

The conditional density of \( \nu \) given \( \hat{\eta} \) can be found by using the fact that
\[
\pi(\nu | \hat{\eta}) \propto \pi(\nu) \pi(\hat{\eta} | \nu)
\]
where \( \pi(\hat{\eta}|\nu) = \mathcal{N}(\hat{\eta}|Z\nu, Q_{\eta\nu}^{-1} + Q_\epsilon^{-1}) \) and \( \pi(\nu) = \mathcal{N}(\nu|\mu_\nu, Q_\nu^{-1}) \). The result is a Gaussian density with mean \( \mu_{\nu|\hat{\eta}} \) and precision matrix \( Q_{\nu|\hat{\eta}} \) where

\[
\mu_{\nu|\hat{\eta}} = Q_{\nu|\hat{\eta}}^{-1}(Q_\nu \mu_\nu + Z^T(Q_\epsilon^{-1} + Q_{\eta\eta}^{-1})^{-1}\hat{\eta}), \quad Q_{\nu|\hat{\eta}} = Q_\nu + Z^T(Q_\epsilon^{-1} + Q_{\eta\eta}^{-1})^{-1}Z
\]

The formula for \( \mu_{\nu|\hat{\eta}} \) shows that the posterior mean for \( \nu \) is a linear combination of the prior mean \( \mu_\nu \) and the maximum likelihood estimates in \( \hat{\eta} \). This posterior mean is the result of a Bayesian regression with design matrix \( Z \), the prior precision matrix \((Q_\epsilon^{-1} + Q_{\eta\eta}^{-1})^{-1}\) and the prior precision matrix \( Q_\nu \). The larger the weights in \((Q_\epsilon^{-1} + Q_{\eta\eta}^{-1})^{-1}\) the greater is the influence of \( \hat{\eta} \) on the posterior mean \( \mu_{\nu|\hat{\eta}} \). If the prior mean of \( \mu_\nu \) is equal to zero then the posterior mean of \( \nu \) is

\[
\mu_{\nu|\hat{\eta}} = (Q_\nu + Z^T(Q_\epsilon^{-1} + Q_{\eta\eta}^{-1})^{-1}Z)^{-1}Z^T(Q_\epsilon^{-1} + Q_{\eta\eta}^{-1})^{-1}\hat{\eta},
\]

so, in this case the posterior mean \( \mu_{\nu|\hat{\eta}} \) is solely a linear combination of \( \hat{\eta} \).
4 Application of an Extended LGM on Maximum Flow in Rivers

4.1 Model Description

Let $y_{it}$ be the annual maximum flow at station $i$ in year $t$. The data are assumed to follow the generalised extreme value (GEV) density, in particular, $y_{it}$ follows a GEV density with location parameter $\mu_i$, scale parameter $\sigma_i$ and shape parameter $\xi_i$ see Chapter 2.6,

$$y_{it} \sim \text{GEV}(y_{it}|\mu_i,\sigma_i,\xi_i), \quad \mu_i \in \mathbb{R}, \quad \sigma_i > 0, \quad \xi_i \in \mathbb{R}, \quad \{\forall y_{it} : 1 + \xi_i \frac{y_{it} - \mu_i}{\sigma_i} > 0\}$$

where $i \in \{1, \ldots, J\}$ and $t \in \{1, \ldots, n_i\}$. Let $\gamma_i = (\mu_i, \sigma_i, \xi_i)^T$. The data density for $y_i = (y_{i1}, \ldots, y_{in_i})^T$ is then

$$\pi(y_i|\mu_i, \sigma_i, \xi_i) = \prod_{t=1}^{n_i} \frac{1}{\sigma_i} \left( 1 + \xi_i \frac{y_{it} - \mu_i}{\sigma_i} \right)^{-\xi_i + 1} \exp \left( - \left( 1 + \xi_i \frac{y_{it} - \mu_i}{\sigma_i} \right)^{-\xi_i} \right)$$

given that each $y_{it}$ is in the support, $\{\forall y_{it} : 1 + \xi_i \frac{y_{it} - \mu_i}{\sigma_i} > 0\}$.

Now let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be such that $g(x_1, x_2, x_3) = (\log(x_1), \log(x_2/x_1), x_3)^T$. Then $g^{-1}(u_1, u_2, u_3) = (\exp(u_1), \exp(u_1 + u_2), u_3)^T$ and let

$$\eta_i := (\psi_i, \tau_i, \xi_i)^T := g(\gamma_i) = (\log(\mu_i), \log(\sigma_i/\mu_i), \xi_i)^T.$$

Now define $\hat{\gamma}_i := (\hat{\mu}_i, \hat{\sigma}_i, \hat{\xi}_i)^T$ to be the maximum likelihood estimates based on the data density for station $i$. Instead of using $y_i$ as the data then $\hat{\eta}_i = (\hat{\psi}_i, \hat{\tau}_i, \hat{\xi}_i)^T = (\log(\hat{\mu}_i), \log(\hat{\sigma}_i/\hat{\mu}_i), \hat{\xi}_i)^T$ will be used as the data for station $i$ and the likelihood function for $\eta_i$ becomes

$$\pi(y_i|\eta_i) \approx \pi(\hat{\eta}_i|\eta_i) = N(\hat{\eta}_i|\eta_i, (-H_{\eta})^{-1})$$

where

$$H_{\eta_i} = \nabla^2 \log(L(\eta_i|y_i))|_{\eta_i=\hat{\eta}_i}.$$

More details on the approximation of using $\hat{\eta}$ as the data are given in Chapter 3.1. The Hessian matrix $H_{\eta_i} = \nabla^2 \log(L(\gamma_i|y_i))|_{\gamma_i=\hat{\gamma}_i}$ and the mode, $\gamma_i$, can be computed using standard routines in software such as R. The mode for $\eta_i$ is found with $\hat{\eta}_i = g(\hat{\gamma}_i)$ and the Hessian matrix $H_{\eta_i}$, can be written in terms of $H_{\gamma_i}$ as
$H_\eta = D(\hat{\eta})H_\eta D(\hat{\eta})^T$

where

$$D(\eta) = \begin{bmatrix}
\frac{\partial \mu}{\partial \psi}(\eta) & \frac{\partial \mu}{\partial \tau}(\eta) & 0 \\
\frac{\partial \sigma}{\partial \psi}(\eta) & \frac{\partial \sigma}{\partial \tau}(\eta) & 0 \\
0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
\exp(\psi) & \exp(\tau + \psi) & 0 \\
0 & \exp(\tau + \psi) & 0 \\
0 & 0 & 1
\end{bmatrix}.$$

Now let $\eta = (\psi^T, \tau^T, \xi^T)^T$ with $\psi = (\psi_1, ..., \psi_J)$, $\tau = (\tau_1, ..., \tau_J)$ and $\xi = (\xi_1, ..., \xi_J)$ then

$$\pi(\hat{\eta}|\eta) = N(\hat{\eta}|\eta, Q_{\eta y}^{-1})$$

where $\hat{\eta} = (\hat{\psi}^T, \hat{\tau}^T, \hat{\xi}^T)^T$ and since

$$\text{var}(\hat{\eta}^*) = (Q_{\eta y}^*)^{-1} = \text{bdiag}((-H_{\eta 1})^{-1}, ..., (-H_{\eta J})^{-1})$$

with $\hat{\eta}^* = (\hat{\psi}_1, \hat{\tau}_1, \hat{\xi}_1, ..., \hat{\psi}_J, \hat{\tau}_J, \hat{\xi}_J)^T$ is known, one can just rearrange $\hat{\eta}^*$ to get $\hat{\eta}$ and rearrange $(Q_{\eta y}^*)^{-1}$ accordingly to get $Q_{\eta y}^{-1}$.

### 4.1.1 Data Level

Now we think of $\hat{\eta}$ as the data and take $Q_{\eta y}^{-1}$ as fixed. The data distribution is then

$$\pi(\hat{\eta}|\eta, \nu, \theta) = N(\hat{\eta}|\eta, Q_{\eta y}^{-1})$$

where $\nu$ are all unknown parameters that are assigned a Gaussian prior distribution and $\theta$ are all unknown hyperparameters.

### 4.1.2 Latent Level

Three linear models of the form

$$\psi = X_\psi \beta_\psi + A_\psi u_\psi + \epsilon_\psi$$
$$\tau = X_\tau \beta_\tau + A_\tau u_\tau + \epsilon_\tau$$
$$\xi = X_\xi \beta_\xi + \epsilon_\xi.$$

are proposed for $\psi$, $\tau$ and $\xi$ where $A_\psi u_\psi$ and $A_\tau u_\tau$ are approximations of two Gaussian fields at the observation locations based on the SPDE approach from Chapter 2.5.2.
The three vectors $\boldsymbol{e}_\psi$, $\boldsymbol{e}_\tau$ and $\boldsymbol{e}_\xi$ contain unstructured model errors with variances $\sigma^2_{e\psi}$, $\sigma^2_{e\tau}$ and $\sigma^2_{e\xi}$, respectively. The prior densities for $\boldsymbol{e}_\psi$, $\boldsymbol{e}_\tau$ and $\boldsymbol{e}_\xi$ can be written as

$$\begin{align*}
\boldsymbol{e}_\psi & \sim \text{iid } N(0, \sigma^2_{e\psi} I), \\
\boldsymbol{e}_\tau & \sim \text{iid } N(0, \sigma^2_{e\tau} I), \\
\boldsymbol{e}_\xi & \sim \text{iid } N(0, \sigma^2_{e\xi} I).
\end{align*}$$

The variances $\sigma^2_{e\psi}$, $\sigma^2_{e\tau}$ and $\sigma^2_{e\xi}$ are defined at the hyperparameter level. A priori the vectors $\beta_\psi$, $\boldsymbol{u}_\psi$, $\beta_\tau$, $\boldsymbol{u}_\tau$ and $\beta_\xi$ are assumed to be independent and follow a Gaussian density. Denote the means and precision matrices of these five vectors by $\mu_\beta_\psi$, $\mu_{\beta_\psi}$, $\mu_\beta_\tau$, $\mu_{\beta_\tau}$, $\mu_\beta_\xi$, $Q_{\beta_\psi}$, $Q_{\beta_\tau}$, $Q_{\beta_\tau}$ and $Q_{\beta_\xi}$. The parameters $\beta_\psi$, $\beta_\psi$ and $\beta_\psi$ have fixed precision matrices while hyperparameters govern the precision matrices for $\boldsymbol{u}_\psi$ and $\boldsymbol{u}_\tau$, namely, $s_\psi$, $\rho_\psi$, $s_\tau$ and $\rho_\tau$. The prior specification for the covariate coefficients is assumed to be the same for all covariates and parameters, namely,

$$\beta_{k,c} \sim N(0, 100^2),$$

with $k \in \{\psi, \tau, \xi\}$ and here $c$ could be any of the covariates selected in Chapter 5.4.

### 4.1.3 Hyperparameter Level

All parameters that do not have Gaussian prior distributions are defined at the hyperparameter level. To control the spatial fields two hyperparameters are used, namely, range parameters, $\rho_\psi$ and $\rho_\tau$, and the marginal standard deviations of the spatial fields, $s_\psi$ and $s_\tau$. The smoothness parameter of the Matérn covariance function, Equation 2.12, is assumed fixed, $\nu_\psi = \nu_\tau = 1$. For the spatial parameters a PC-prior is used:

$$\pi(s_k, \rho_k) = \lambda_1 \lambda_2 \rho_k^{-2} \exp(-\lambda_1 \rho_k^{-1} - \lambda_2 s_k), s_k > 0, \rho_k > 0,$$

where $P(\rho_k < \rho_{0k}) = \alpha_1$ and $P(s_k > s_{0k}) = \alpha_2$ are achieved by

$$\begin{align*}
\lambda_1 &= -\log(\alpha_1) \rho_{0k} \\
\lambda_2 &= -\frac{\log(\alpha_2)}{s_{0k}},
\end{align*}$$

where $k \in \{\psi, \tau\}$. This prior is defined in Fuglstad et al. (2017) and is based on the PC-prior framework (Simpson et al. (2017)). The prior specifications for $s_\psi$, $\rho_\psi$, $s_\tau$ and $\rho_\tau$ are such that $\rho_{0k} = 0.5$, $\alpha_1 = 0.95$, $s_{0k} = 1$ and $\alpha_2 = 0.01$ for $k \in \{\psi, \tau\}$. The observation locations were scaled such that they were on the interval from zero to one. This explains why $\rho_{0k} = 0.5$ and $\alpha_1 = 0.95$ is reasonable, since the range parameter $\rho_k$ represents the distance at which the correlation is approximately 0.1 and a distance of 0.5 is substantial.

For $\sigma_{e\psi}$, $\sigma_{e\tau}$ and $\sigma_{e\xi}$ the precision parameters

$$\begin{align*}
q_\psi &= \frac{1}{\sigma^2_{e\psi}}, \\
q_\tau &= \frac{1}{\sigma^2_{e\tau}}, \\
q_\xi &= \frac{1}{\sigma^2_{e\xi}}
\end{align*}$$

are given a PC-prior of the form

$$\pi(q_k) = \frac{\lambda}{2} q_k^{-3/2} \exp(-\lambda q_k^{-1/2}), \quad q_k > 0.$$
for \( \lambda > 0 \) where \( \lambda = - \frac{\log(\alpha)}{u} \) and \((u, \alpha)\) are interpreted as

\[
P(1/\sqrt{q_\kappa} > u) = P(\sigma_\kappa > u) = \alpha
\]

where \( k \in \{\psi, \tau, \xi\} \). This prior density is derived in Simpson et al. (2017). The prior specifications for \( q_\psi, q_\tau \) and \( q_\xi \) are \( u = 0.6 \) and \( \alpha = 0.05 \).

### 4.1.4 Spatial Components

Like mentioned before the hyperparameters \( s_\psi, \rho_\psi, s_\tau \) and \( \rho_\tau \) are used to control the Gaussian fields corresponding to \( u_\psi \) and \( u_\tau \). The matrix \( A \) is a fixed projection matrix that describes the linear interpolation of the finite GMRF representation, using the SPDE approach (Lindgren et al. (2011)), at the vertices of the mesh, Figure 4.1, on to the observation locations. The products \( Au_\psi \) and \( Au_\tau \) are thus the effects of the spatial fields at the observation locations while \( u_\psi \) and \( u_\tau \) are the effects of the spatial fields at each vertex of the mesh.

![Triangulated mesh over the UK.](image)
4.2 Inference

The model described in Chapter 4.1 is of the same form as the extended latent Gaussian model described in Chapter 3 and the inference scheme in Chapter 3.1 applies here. To sample from the posterior first a sample from

$$\pi(\theta | \hat{\eta}) \propto \pi(\theta) \pi(\hat{\eta} | \theta) \propto \pi(\theta) \frac{\pi(\hat{\eta} | x, \theta) \pi(x | \theta)}{\pi(x | \theta, \hat{\eta})},$$

is taken and then a sample from the conditional posterior density of $x$ given $\hat{\eta}$ and $\theta$, namely,

$$\pi(x | \hat{\eta}, \theta) = N(x | Q_{x|\hat{\eta}}^{-1}(Q_x \mu_x + B^T Q_{yy} \hat{\eta}), Q_{x|\hat{\eta}}^{-1}),$$

where

$$Q_{x|\hat{\eta}} = Q_x + B^T Q_{yy} B = \begin{bmatrix} Q_x + Q_{yy} & -Q_x \mu \nu \\ -Z^T Q_e & Q_e + Z^T Q_e Z \end{bmatrix}.$$

Here $\theta = (\theta_1, \ldots, \theta_T)^T = (\rho_\psi, s_\psi, q_\psi, \rho_r, s_r, q_r, q_\xi)^T$. Instead of sampling from $\pi(\theta | \hat{\eta})$ a sample from $\pi(\kappa | \hat{\eta})$ is taken where the $j$-th element of $\kappa$ is such that $\kappa_j = \log(\theta_j)$. Samples from $\pi(\kappa | \hat{\eta})$ are taken with a metropolis step where the proposal distribution is

$$q(\kappa^* | \kappa, Q_e^{-1}) = N(\kappa^* | \kappa, Q^{-1}).$$

This transformation, $\kappa = (\log \theta_1, \ldots, \log \theta_T)^T$, means the prior densities of $\theta$ need to be recalculated using $\kappa$. In the case of the spatial parameters $s_k$ and $\rho_k$, $k \in \{\psi, \tau\}$, the transformation is $U_k = \log(s_k)$ and $V_k = \log(\rho_k)$ and the prior density becomes

$$f_{U_k, V_k}(u, v) = f_{s_k, \rho_k}(e^u, e^v) \cdot J$$

where $J$ is the Jacobian

$$J = \begin{vmatrix} \frac{\partial s_k}{\partial u_k} & \frac{\partial s_k}{\partial v_k} \\ \frac{\partial \rho_k}{\partial u_k} & \frac{\partial \rho_k}{\partial v_k} \end{vmatrix} = \begin{vmatrix} e^u & 0 \\ 0 & e^v \end{vmatrix} = e^{u_k} e^{v_k}$$

so the prior density becomes

$$f_{U_k, V_k}(u, v) = f_{s_k, \rho_k}(e^u, e^v) \cdot J = f_{s_k, \rho_k}(e^u, e^v) e^{ue^v} = \lambda_1 \lambda_2 e^{-w} e^{w} \exp(-\lambda_1 e^{-v} - \lambda_2 e^u).$$

There is also a log transformation of $q_\psi$, $q_\tau$ and $q_\xi$ which will be called $w_\psi = \log q_\psi$, $w_\tau = \log q_\tau$ and $w_\xi = \log q_\xi$. The prior for $w_\psi$, $w_\tau$ and $w_\xi$ then becomes

$$f_{w_k}(w) = f_{q_k}(e^w) \cdot \frac{\partial q}{\partial w} = \frac{\lambda}{2} \exp(-\lambda e^{-w/2} - w/2)$$

where $k \in \{\psi, \tau, \xi\}$. The prior distribution of $\kappa$ can now be written as

$$\pi(\kappa) = f_{W_\psi}(w_\psi) f_{W_\tau}(w_\tau) f_{W_\xi}(w_\xi) f_{U_\psi, V_\psi}(u_\psi, v_\psi) f_{U_\tau, V_\tau}(u_\tau, v_\tau)$$

25
and the marginal posterior of $\kappa|\hat{\eta}$ becomes

$$\pi(\kappa|\hat{\eta}) \propto \pi(\kappa)\frac{\pi(\hat{\eta}|x, \theta)\pi(x|\theta)}{\pi(x|\theta, \hat{\eta})}. $$

where

$$\theta = (\rho_\psi, s_\psi, q_\psi, \rho_\tau, s_\tau, q_\tau, q_\xi)^T = (\exp u_\psi, \exp v_\psi, \exp w_\psi, \exp u_\tau, \exp v_\tau, \exp w_\tau, \exp w_\xi)^T.$$

Algorithm 1 shows how samples are taken from $\pi(\kappa|\hat{\eta})$. Calculating $\log(r)$ is the most time consuming part of Algorithm 1. Now $\log(r)$ can be written as

$$\log(r) = \log\frac{\pi(\kappa^*|\hat{\eta})}{\pi(\kappa^k|\hat{\eta})}$$

$$= \log \pi(\kappa^*) - \log \pi(\kappa^k)$$

$$+ \log \pi(\hat{\eta}|x, \theta^*) - \log \pi(\hat{\eta}|x, \theta^k)$$

$$+ \log \pi(x|\theta^*) - \log \pi(x|\theta^k)$$

$$- \log \pi(x|\theta^*, \hat{\eta}) + \log \pi(x|\theta^k, \hat{\eta})$$

$$= \log \pi(\kappa^*) - \log \pi(\kappa^k)$$

$$+ \log \pi(x|\theta^*) - \log \pi(x|\theta^k)$$

$$- \log \pi(x|\theta^*, \hat{\eta}) + \log \pi(x|\theta^k, \hat{\eta})$$

and $\pi(\kappa|\hat{\eta})$ is independent of $x$, so a convenient value of $x$ can be chosen, for example, $x = 0$. The log-density of $x|\theta$ can be written as

$$\log \pi(x|\theta)|_{x=0} = \log \mathcal{N}(x|\mu_x, Q_x^{-1})|_{x=0}$$

$$= \log \left( \frac{\exp(-\frac{1}{2}(0 - \mu_x)^TQ_x(0 - \mu_x))}{(2\pi)^{\dim(x)/2}\det(Q_x^{-1})^{1/2}} \right)$$

$$= \log(\exp(0)) + \frac{1}{2} \log(\det(Q_x)) + C$$

(Since $\mu_x = 0$)

$$= \frac{1}{2} \log(\det(Q_x)) + C$$
4.2 Inference

where $C$ is a constant. The logarithm of the determinant of $Q_x$ can be calculated quickly since $Q_x$ is positive, symmetric and sparse. Similarly, the log-density of $x|\theta, \hat{\eta}$ can be written as

$$
\log \pi(x|\theta, \hat{\eta})|_{x=0} = \log \mathcal{N}(x|\mu_x|\hat{\eta}, Q_{x|\hat{\eta}}^{-1})|_{x=0} \nonumber
$$

$$
= \log \left( \exp\left( -\frac{1}{2} (0 - \mu_{x|\hat{\eta}})^T Q_{x|\hat{\eta}} (0 - \mu_{x|\hat{\eta}}) \right) \right) \nonumber
$$

$$
= \log(\exp(-\frac{1}{2}\mu_{x|\hat{\eta}}^T Q_{x|\hat{\eta}} \mu_{x|\hat{\eta}})) + \frac{1}{2} \log(\det(Q_{x|\hat{\eta}})) + C \nonumber
$$

$$
= -\frac{1}{2} \mu_{x|\hat{\eta}}^T Q_{x|\hat{\eta}} \mu_{x|\hat{\eta}} + \frac{1}{2} \log(\det(Q_{x|\hat{\eta}})) + C
$$

where $\mu_{x|\hat{\eta}} = Q_{x|\hat{\eta}}^{-1}(Q_x \mu_x + B^T Q_{\eta \hat{\eta}})$ can be computed quickly using Cholesky decomposition. The logarithm of the determinant of $Q_{x|\hat{\eta}}$ can be calculated fairly quickly since it’s sparse, symmetric and positive. $Q_{x|\hat{\eta}}$ is not as sparse as $Q_x$ and calculating the determinant of $Q_{x|\hat{\eta}}$ is the most computationally expensive part in calculating $\log(r)$ in Algorithm 1.

To sample $x$ one can simply use the samples from $\pi(\kappa|\hat{\eta})$, taken using Algorithm 1, to sample from $\pi(x|\hat{\eta}, \theta)$. To do that one can follow these steps for each sample of $\theta$:

**Step 1:** Compute $R$, the Cholesky decomposition of $Q_{x|\hat{\eta}}$ such that $Q_{x|\hat{\eta}} = R^T R$.

**Step 2:** Draw $z \sim \mathcal{N}(0, I_{\text{dim}(x)})$.

**Step 3:** Solve $R u = z$ for $u$.

**Step 4:** Solve $R^T v = Q_x \mu_x + Z^T Q_{\eta \hat{\eta}}$ for $v$.

**Step 5:** Solve $R \mu_{x|\hat{\eta}} = v$ for $\mu_{x|\hat{\eta}}$.

**Step 6:** Compute $x = \mu_{x|\hat{\eta}} + u$. 
5 Data and Preliminary Analysis

The data used in this thesis comes from the Center for Ecology and Hydrology in the UK. The data consists of annual maximum flow for 954 gauging stations in the UK. In this thesis, only data from 553 stations will be used, because these stations are marked as more reliable in the data set. Figure 5.1 shows the number of observations for each year in the data set. From the figure, one can see that the observations date back to 1851, but most of the observations are from the years 1970 to 2013. This data has been investigated and modeled before, see, for example, Kjeldsen & Jones (2006), Kjeldsen & Jones (2009a) and Kjeldsen (2010).

![Figure 5.1: Number of observed stations as a function of year.](image)
5 Data and Preliminary Analysis

Figure 5.2 shows the locations of the stations on a map of the UK. The stations are located all over the UK and are of various sizes.

![Station locations](image)

**Figure 5.2: Station locations.**

5.1 Closer Look at the MLEs

In this chapter a closer look will be taken at the MLE estimates defined as $\hat{\eta}_k = g(\hat{\mu}_k, \hat{\sigma}_k, \hat{\xi}_k) = (\log(\hat{\mu}_k), \log(\hat{\sigma}_k/\hat{\mu}_k), \hat{\xi}_k)^T$ for station $k$. The first reason behind the transformation $g$ is that the MLEs are much easier to handle when transformed using $g$. Figure 5.3 shows the densities of the MLEs both before and after the transformation, $g$. The second reason behind this transformation is the relationship between $\log(\sigma_i)$ and $\log(\mu_i)$ which can be seen in Figure 5.4. To model the variability in $\sigma$ that is not explained by $\mu$ the transformation $\log(\sigma_i/\mu_i)$ is used. The relationship between $\tau$ and $\psi$ can be seen in Figure 5.5.

Figures 5.6-5.11 show the MLEs $\hat{\psi}$, $\hat{\tau}$ and $\hat{\xi}$ and the marginal standard deviation of the MLEs, evaluated using the Hessian of the likelihood function, on a map of the UK.
5.1 Closer Look at the MLEs

Figure 5.3: The densities of the MLE estimates.

Figure 5.12 shows the marginal standard deviation of the MLEs (evaluated with the Hessian matrix of the likelihood function) as a function of the number of observations, \( n \). As expected, the size of the marginal standard deviation decreases as a function of \( n \). Plotting this relationship on the log-log scale reveals a linear relationship between the marginal standard deviation and the number of observations, \( n \). This relationship is obvious for \( \hat{\tau} \) and \( \hat{\xi} \) but not as apparent for \( \hat{\psi} \).
Figure 5.4: \( \log(\hat{\sigma}_i) \) as a function of \( \log(\hat{\mu}_i) \).

Figure 5.5: \( \hat{\tau}_i = \log(\hat{\sigma}_i) \) as a function of \( \hat{\psi}_i = \log(\hat{\mu}_i) \).
5.1 Closer Look at the MLEs

Figure 5.6: The MLE $\hat{\psi}$ on a map of the UK.

Figure 5.7: The marginal standard deviation of $\hat{\psi}$ (evaluated with the Hessian matrix of the likelihood function) on a map of the UK.
5 Data and Preliminary Analysis

Figure 5.8: The MLE $\hat{\tau}$ on a map of the UK.

Figure 5.9: The marginal standard deviation of $\hat{\tau}$ (evaluated with the Hessian matrix of the likelihood function) on a map of the UK.
5.1 Closer Look at the MLEs

Figure 5.10: The MLE $\hat{\xi}$ on a map of the UK.

Figure 5.11: The marginal standard deviation of $\hat{\xi}$ (evaluated with the Hessian matrix of the likelihood function) on a map of the UK.
Figure 5.12: The standard deviation of the MLEs as a function of the number of observations on the log-log scale.
5.2 A Closer Look at Four Stations

This chapter takes a closer look at four of the gauging stations in the data set. The four stations were chosen at random, and their summary statistics can be seen in Table 5.1. Table 5.1 shows that the rivers are of various sizes, for example, Station 45816 is measuring a small river, with few observations, compared to station 12001 which is measuring a big river, with observations spanning 77 years.

<table>
<thead>
<tr>
<th>Station</th>
<th>Nr. of obs.</th>
<th>Mean</th>
<th>Sd.</th>
<th>Min.</th>
<th>Med.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>12001</td>
<td>77</td>
<td>427.59</td>
<td>168.25</td>
<td>114.10</td>
<td>436.81</td>
<td>1132.52</td>
</tr>
<tr>
<td>27034</td>
<td>48</td>
<td>248.39</td>
<td>56.79</td>
<td>128.46</td>
<td>242.64</td>
<td>380.34</td>
</tr>
<tr>
<td>33005</td>
<td>28</td>
<td>21.23</td>
<td>6.78</td>
<td>3.40</td>
<td>21.80</td>
<td>35.45</td>
</tr>
<tr>
<td>45816</td>
<td>22</td>
<td>4.22</td>
<td>2.68</td>
<td>1.45</td>
<td>3.49</td>
<td>11.66</td>
</tr>
</tbody>
</table>

Figure 5.13 shows the time series for the chosen stations. It can be seen, from the figure, that the size and pattern of the maximum annual flow for the four rivers are different.

Figure 5.14 shows the empirical cumulative density function (CDF) for the four chosen stations, by looking at the empirical CDFs it can be seen that the tail behavior is different.
between the stations. Figure 5.15 shows box plots for the selected stations, this figure shows clearly what has been mentioned before in this chapter, the size difference between the rivers and the difference in tail behavior.

![Empirical CDF for four selected stations.](image)

*Figure 5.14: Empirical CDF for four selected stations.*
Figure 5.15: Boxplots for four selected stations.
5 Data and Preliminary Analysis

5.3 Covariates

In this chapter, potential covariates for the annual maximum flow model (Chapter 4.1) will be examined. A transformation for each covariate will be chosen by looking at the density and the relationship with the MLEs, \( \hat{\psi} \), \( \hat{\tau} \) and \( \hat{\xi} \). Table 5.2 shows the potential covariates of the model and a short description of each one.

<table>
<thead>
<tr>
<th>Catchment descriptor</th>
<th>Description</th>
<th>Stats.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREA</td>
<td>Catchment drainage area is the total area, in km(^2), where the river collects water [km(^2)].</td>
<td>Min 1.63 Mean 440.65 Max 9930.80</td>
</tr>
<tr>
<td>SAAR</td>
<td>Average annual rainfall [mm].</td>
<td>Min 559.00 Mean 1114.12 Max 2913.00</td>
</tr>
<tr>
<td>FARL</td>
<td>The Flood Attenuation by Reservoirs and Lakes index provides a guide to the degree of flood attenuation attributable to reservoirs and lakes in the catchment.</td>
<td>Min 0.64 Mean 0.96 Max 1.00</td>
</tr>
<tr>
<td>BFIHOST</td>
<td>A measure of catchment responsiveness.</td>
<td>Min 0.17 Mean 0.49 Max 0.97</td>
</tr>
<tr>
<td>FPEXT</td>
<td>Floodplain extent is the fraction of the total area of the catchment that is covered by the 100-year event.</td>
<td>Min 0.00 Mean 0.06 Max 0.30</td>
</tr>
<tr>
<td>URBEXT</td>
<td>Urban extent is an index that depicts the extent of urban/suburban land cover in the catchment area.</td>
<td>Min 0.00 Mean 0.03 Max 0.59</td>
</tr>
<tr>
<td>DPLBAR</td>
<td>Is the mean of the distances between each grid node and the catchment outlet [km].</td>
<td>Min 1.30 Mean 23.05 Max 139.87</td>
</tr>
<tr>
<td>DPSBAR</td>
<td>Mean drainage path slope provides an index of steepness in the catchment, [m/km].</td>
<td>Min 13.30 Mean 100.12 Max 441.80</td>
</tr>
<tr>
<td>LDP</td>
<td>Longest drainage path [km].</td>
<td>Min 2.21 Mean 42.94 Max 286.84</td>
</tr>
<tr>
<td>SPRHOST</td>
<td>Standard percentage runoff associated with each HOST soil class.</td>
<td>Min 4.85 Mean 37.20 Max 59.85</td>
</tr>
<tr>
<td>ASPBAR</td>
<td>The mean aspect or orientation of a catchment in degrees.</td>
<td>Min 1.00 Mean 151.95 Max 358.00</td>
</tr>
<tr>
<td>ALTBAR</td>
<td>The mean elevation of a catchment area [m].</td>
<td>Min 25.00 Mean 220.78 Max 682.00</td>
</tr>
<tr>
<td>ASPVAR</td>
<td>The aspect variance in the catchment in degrees.</td>
<td>Min 0.02 Mean 0.17 Max 0.59</td>
</tr>
<tr>
<td>PROPWET</td>
<td>An index that indicates the proportion of time the soil in the catchment area is wet.</td>
<td>Min 0.23 Mean 0.48 Max 0.85</td>
</tr>
</tbody>
</table>
Figure 5.16 shows a correlation plot of the potential covariates. It can be seen from Figure 5.16 that there are two groups of covariates that are positively correlated, LDP, DPLBAR and AREA in one group and PROPWET, ALTBAR, DPSBAR, SAAR and SPRHOST in the other group. SPRHOST and BFIHOST are also highly negatively correlated. Figures 5.17 and 5.18 take a closer look at these covariates, the figures show that there is a correlation of $> 0.8$ between LDP, AREA and DPLBAR and that there is also a high correlation between most of the catchment descriptors in Figure 5.18. This is something to have in mind when the covariates are selected in Chapter 5.4.

![Correlation plot for the covariates.](image)

Figure 5.16: Correlation plot for the covariates.
Figure 5.17: Correlation plot for LDP, DPLBAR and AREA.

Figure 5.18: Correlation plot for PROPWET, ALTBAR, DPSBAR, SAAR and SPRHOST.
5.3 Covariates

Catchment drainage area (AREA) is the total area, in km$^2$, where the river collects precipitation. This includes all the surface water from rain, snowmelt, nearby streams that run towards the station and groundwater underneath the earth’s surface. Figure 5.19 shows the density of AREA and log(AREA) and that catchment area is much easier to handle on the log-scale. The figure also shows the relationship between catchment area and the MLEs, the figure shows a clear positive relationship for $\hat{\psi}$ and a negative relationship with $\hat{\tau}$. The correlation between the MLEs and log(AREA) are shown in Table 5.3.

![Figure 5.19: The density of catchment area (original scale and log-scale) and the relationship between the logarithm of the catchment area and the MLEs.](image)

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\psi}$</td>
<td>0.722</td>
</tr>
<tr>
<td>$\hat{\tau}$</td>
<td>-0.252</td>
</tr>
<tr>
<td>$\hat{\xi}$</td>
<td>-0.045</td>
</tr>
</tbody>
</table>
**Average annual rainfall (SAAR)** is derived from the monthly catchment rainfall series from the National River Flow Archive. It is computed from the monthly mean rainfalls only using data from years where the monthly rainfall record is complete. In this thesis, SAAR will be used on the log scale. Figure 5.20 shows the difference in density for SAAR on the original- and the log-scale. The figure also shows the relationship between the MLEs and log(SAAR). The correlation between log(SAAR) and the MLEs can be seen in Table 5.4.

![Figure 5.20: The density of average annual rainfall (original scale and log-scale) and the relationship between the logarithm of the average annual rainfall and the MLEs.](image)

**Table 5.4: Correlation between log(SAAR) and the MLEs.**

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\psi})</td>
<td>0.483</td>
</tr>
<tr>
<td>(\hat{\tau})</td>
<td>-0.500</td>
</tr>
<tr>
<td>(\hat{\xi})</td>
<td>-0.017</td>
</tr>
</tbody>
</table>
The Flood Attenuation by Reservoirs and Lakes (FARL) index provides a guide to the degree of flood attenuation attributable to reservoirs and lakes in the catchment above a gauging station. Figure 5.21 shows the density of FARL and log(FARL). In this project FARL is used on the log-scale. The figure also shows the relationship between FARL on the log-scale and the MLEs, the correlation between log(FARL) and the MLEs is shown in Table 5.5.

Figure 5.21: The density of FARL and log(FARL), and the relationship between log(FARL) and the MLEs.

Table 5.5: Correlation between log(FARL) and the MLEs.

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\psi}$</td>
<td>-0.117</td>
</tr>
<tr>
<td>$\hat{\tau}$</td>
<td>0.038</td>
</tr>
<tr>
<td>$\hat{\xi}$</td>
<td>-0.013</td>
</tr>
</tbody>
</table>
Catchment responsiveness (BFIHOST) is a measure of catchment responsiveness. In this thesis, BFIHOST will be transformed to BFIHOST$^2$, as suggested in Kjeldsen & Jones (2009a). Figure 5.22 shows the density of BFIHOST and BFIHOST$^2$ as well as the relationship between BFIHOST$^2$ and the MLEs. The correlation between BFIHOST$^2$ and the MLEs can be seen in Table 5.6.

![Figure 5.22: The density of BFIHOST and BFIHOST$^2$ and the relationship between BFIHOST$^2$ and the MLEs.](image)

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\psi})</td>
<td>-0.583</td>
</tr>
<tr>
<td>(\hat{\tau})</td>
<td>0.276</td>
</tr>
<tr>
<td>(\hat{\xi})</td>
<td>-0.044</td>
</tr>
</tbody>
</table>
Floodplain extent (FPEXT) is the fraction of the total area of the catchment that is estimated to be covered by the 100-year event. Figure 5.23 shows the density of FPEXT on the original and log-scale, it shows that most of the mass is close to zero and that log-transforming the variable helps with that. The figure also shows the relationship between log(FPEXT) and the MLEs. There does not seem to be any obvious relationship between them. The correlation between log(FPEXT) and the MLEs can be seen in Table 5.7.

**Figure 5.23:** The density of FPEXT and log(FPEXT), and the relationship between log(FPEXT) and the MLEs.

**Table 5.7: Correlation between log(FPEXT) and the MLEs.**

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\psi}$</td>
<td>-0.006</td>
</tr>
<tr>
<td>$\hat{\tau}$</td>
<td>-0.083</td>
</tr>
<tr>
<td>$\hat{\xi}$</td>
<td>-0.167</td>
</tr>
</tbody>
</table>
Urban extent (URBEXT) is an index that depicts the extent of urban/suburban land cover in the catchment area. It ranges from 0 to 1. In this thesis, the transformation log(1+URBEXT) is used since there are a lot of zero values. The density of URBEXT can be seen in Figure 5.24 where it can be seen that most of the mass is really close to zero. The figure also shows the relationship between log(1+URBEXT) and the MLEs, visually there are no obvious relationships. The correlation between URBEXT and the MLEs can be seen in Table 5.8.

![Figure 5.24: The density of URBEXT and log(1+URBEXT), and the relationship between log(1+URBEXT) and the MLEs.](image)

**Table 5.8: Correlation between log(URBEXT+1) and the MLEs.**

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\psi})</td>
<td>-0.111</td>
</tr>
<tr>
<td>(\hat{\tau})</td>
<td>-0.010</td>
</tr>
<tr>
<td>(\hat{\xi})</td>
<td>0.048</td>
</tr>
</tbody>
</table>
DPLBAR is the mean of the distances between each grid node and the catchment outlet, in kilometers. As one might suspect this variable is highly correlated with catchment area. Figure 5.25 shows that the density of the log-transformed DPLBAR is less skewed. The figure also shows the relationship between the MLEs and log(DPLBAR), the correlation between log(DPLBAR) and the MLEs can be seen in Table 5.9.

![Figure 5.25: The density of DPLBAR and log(DPLBAR), and the relationship between log(DPLBAR) and the MLEs.](image)

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\psi} )</td>
<td>0.671</td>
</tr>
<tr>
<td>( \hat{\tau} )</td>
<td>-0.233</td>
</tr>
<tr>
<td>( \hat{\xi} )</td>
<td>-0.045</td>
</tr>
</tbody>
</table>
DPSBAR (Mean drainage path slope) mean drainage path slope provides an index of steepness in the catchment with units meters per second. It can be seen in figure 5.26 that DPSBAR is less skewed when log transformed and will thus be used on the log-scale. The figure also shows a positive relationship between log(DPSBAR) and $\hat{\psi}$ and a negative relationship between log(DPSBAR) and $\hat{\tau}$, and Table 5.10 shows the correlation between log(DPSBAR) and the MLEs.

![Graph showing density and relationship between log(DPSBAR) and MLEs](image)

**Figure 5.26: The density of DPSBAR and log(DPSBAR), and the relationship between log(DPSBAR) and the MLEs.**

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\psi}$</td>
<td>0.500</td>
</tr>
<tr>
<td>$\hat{\tau}$</td>
<td>-0.324</td>
</tr>
<tr>
<td>$\hat{\xi}$</td>
<td>0.051</td>
</tr>
</tbody>
</table>

**Table 5.10: Correlation between log(DPSBAR) and the MLEs.**
Longest drainage path (LDP) is the longest path from a catchment node to the defined outlet in kilometers. This variable is very correlated with catchment drainage area which could cause trouble. Figure 5.27 shows that LDP is less skewed on the log scale than the original scale. The figure also shows the relationship between log(LDP) and the MLEs, and there is a positive relationship between log(LDP) and $\hat{\psi}$ and a negative relationship between log(LDP) and $\hat{\tau}$.

Figure 5.27: The density of LDP and log(LDP), and the relationship between log(LDP) and the MLEs.

Table 5.11: Correlation between log(LDP) and the MLEs.

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\psi}$</td>
<td>0.677</td>
</tr>
<tr>
<td>$\hat{\tau}$</td>
<td>-0.232</td>
</tr>
<tr>
<td>$\hat{\xi}$</td>
<td>-0.045</td>
</tr>
</tbody>
</table>
Standard percentage runoff (SPRHOST) associated with each HOST soil class. Figure 5.28 shows how the density of SPRHOST is on both the log-scale and the original scale. The density is less skewed on the original scale and will thus be used on the original scale. The figure also displays the relationship between the MLEs and SPRHOST. Table 5.12 shows the correlation between SPRHOST and the MLEs.

![Figure 5.28: The density of SPRHOST and log(SPRHOST), and the relationship between SPRHOST and the MLEs.](image)

**Table 5.12: Correlation between log(SPRHOST) and the MLEs.**

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\psi}$</td>
<td>0.555</td>
</tr>
<tr>
<td>$\hat{\tau}$</td>
<td>-0.262</td>
</tr>
<tr>
<td>$\hat{\xi}$</td>
<td>0.032</td>
</tr>
</tbody>
</table>
ASPBAR is the mean aspect or orientation of a catchment in degrees. It’s an index that indicates the dominant aspect of a catchments slope. Figure 5.29 shows that the density of ASPBAR on the original scale and the log-scale, in this thesis it will be used as ASPBAR/100. The relationship between ASPBAR and the MLEs can also be seen in Figure 5.29. Table 5.13 shows the correlation between ASPBAR and the MLEs.

![Graph of ASPBAR and log(ASPBAR)](image)

Figure 5.29: The density of ASPBAR and log(ASPBAR), and the relationship between ASPBAR and the MLEs.

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\psi}$</td>
<td>-0.087</td>
</tr>
<tr>
<td>$\hat{\tau}$</td>
<td>-0.042</td>
</tr>
<tr>
<td>$\hat{\xi}$</td>
<td>0.003</td>
</tr>
</tbody>
</table>
**ALTBAR** is the mean elevation of a catchment area in meters. Figure 5.30 shows the density of ALTBAR and log(ALTBAR), and the relationship between log(ALTBAR) and the MLEs. The correlation between log(ALTBAR) and the MLEs can be seen in Table 5.14.

![Graph showing density of ALTBAR and log(ALTBAR), and relationship with MLEs](image)

**Figure 5.30:** The density of ALTBAR and log(ALTBAR), and the relationship between ALTBAR and the MLEs.

**Table 5.14:** Correlation between log(ALTBAR) and the MLEs.

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>\hat{\psi}</td>
<td>0.544</td>
</tr>
<tr>
<td>\hat{\tau}</td>
<td>-0.335</td>
</tr>
<tr>
<td>\hat{\xi}</td>
<td>0.070</td>
</tr>
</tbody>
</table>
ASPVAR is the aspect variance in the catchment in degrees. Figure 5.31 shows the density of ASPVAR and log(ASPVAR), in this thesis, ASPVAR is used on the log scale. The figure also shows the relationship between log(ASPVAR) and the MLEs, Table 5.15 shows the correlation between log(ASPVAR) and the MLEs.

![Figure 5.31: The density of ASPVAR and log(ASPVAR), and the relationship between log(ASPVAR) and the MLEs.](image)

Table 5.15: Correlation between log(ASPVAR) and the MLEs.

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\psi}$</td>
<td>-0.449</td>
</tr>
<tr>
<td>$\hat{\tau}$</td>
<td>0.179</td>
</tr>
<tr>
<td>$\hat{\xi}$</td>
<td>0.018</td>
</tr>
</tbody>
</table>
PROPWET is an index that indicates the proportion of time the soil in the catchment area is wet. Figure 5.32 shows the density of PROPWET and of log(PROPWET), in this thesis, PROPWET is used on the log scale. The figure also shows the relationship between log(PROPWET) and the MLEs. Table 5.16 displays the correlation between log(PROPWET) and the MLEs.

![Figure 5.32: The density of PROPWET and the relationship between PROPWET and the MLEs.](image)

<table>
<thead>
<tr>
<th>MLE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\psi} )</td>
<td>0.582</td>
</tr>
<tr>
<td>( \hat{\tau} )</td>
<td>-0.518</td>
</tr>
<tr>
<td>( \hat{\xi} )</td>
<td>0.016</td>
</tr>
</tbody>
</table>
5.4 Covariate Selection

A forward selection was used to select covariates for the linear models at the latent level, see Equation 4.1. A simplified version of the Bayesian hierarchical model from Chapter 4.1 was used, i.e., for a catchment \(i\) the covariance between \(\hat{\psi}_i\) and \(\hat{\tau}_i\), \(\hat{\psi}_i\) and \(\hat{\xi}_i\), and \(\hat{\tau}_i\) and \(\hat{\xi}_i\), at the data level was set equal to zero, and only the marginal variances were taken into consideration. The data density is modified and is of the form

\[
\pi(\hat{\eta}|\eta, \nu, \theta) = \mathcal{N}(\hat{\eta}|\eta, \text{diag}(Q_{\eta y}^{-1})).
\]

The latent level still has the same form as in Equation 4.1

\[
\psi = X_\psi \beta_\psi + A_\psi u_\psi + \epsilon_\psi
\]
\[
\tau = X_\tau \beta_\tau + A_\tau u_\tau + \epsilon_\tau
\]
\[
\xi = X_\xi \beta_\xi + \epsilon_\xi.
\]

The model for \(\psi\) then takes the form

\[
\pi(\hat{\psi}|\psi, \nu, \theta) = \mathcal{N}(\hat{\psi}|\psi, Q^{-1}_\psi),
\]

where \(Q^{-1}_\psi\) is the sample variance of \(\hat{\psi}\) taken from \(\text{diag}(Q_{\eta y}^{-1})\). Similarly, the models for \(\tau\) and \(\xi\) at the data level are given by

\[
\pi(\hat{\tau}|\tau, \nu, \theta) = \mathcal{N}(\hat{\tau}|\tau, Q^{-1}_\tau) \quad \text{and} \quad \pi(\hat{\xi}|\xi, \nu, \theta) = \mathcal{N}(\hat{\xi}|\xi, Q^{-1}_\xi).
\]

At the latent level \(\psi\) has the same form as above and the models for \(\tau\) and \(\xi\) also have the same form. When modelling \(\psi\), \(\tau\) and \(\xi\) separately it’s possible to use the INLA approach discussed in Chapter 2.3 and introduced in Rue et al. (2009). By using INLA the inference takes much less time to run and it is made possible to select covariates with forward selection within a reasonable time frame. To select the best covariates a forward selection algorithm based on 10-fold cross-validation was used. This cross-validation involves splitting the data into ten randomly chosen equally sized sets. The model was then trained on nine of those sets and tested using the one set left. This step is then repeated until each of the ten sets have been used for testing. Algorithm 2 shows roughly how the forward selection is performed.
Algorithm 2 Forward selection

1: procedure FORWARD_SELECTION(Covariates, Sets)
2: Covariates_Left = Covariates \quad \triangleright \text{Keeps track of covariates not chosen}
3: Covariates_Selected = c() \quad \triangleright \text{Chosen covariates, } c() = \text{empty vector}
4: for i in 1:length(Covariates) do
5: \quad Min_error = \infty
6: \quad for Covariate in Covariates_Left do
7: \quad \quad Error = CALC_CV(c(Covariates_Selected, Covariate), Sets)
8: \quad \quad if Error<Min_error then
9: \quad \quad \quad Min_error = Error
10: \quad \quad \quad Selected = Covariate
11: \quad \quad Remove Selected from covariates_left.
12: \quad \quad Add Selected to covariates_selected.
13: \quad return covariates_selected \quad \triangleright \text{Covariates selected in correct order}

In Algorithm 2, the cross-validation error is calculated for each model and used as criteria to find the best model of each size. Algorithm 3 shows roughly how the 10-fold cross-validation is calculated.

Algorithm 3 Calculate 10-fold cross-validation

1: procedure CALC_CV(covariates, sets)
2: for i in sets do
3: \quad Fit a model with covariates and all sets except for i
4: \quad Calculate predicted values, using the model, for the MLEs, $\hat{\psi}$, $\hat{\tau}$ or $\hat{\xi}$, for the observations in set $i$, denoted $PRED_i$
5: \quad Extract the MLEs, $MLE_i$, from the observations in set $i$.
6: \quad Calculate $RESIDUALS_i = PRED_i - MLE_i$
7: \quad Calculate $MSE_i = \text{sum}(RESIDUALS_i^2)/\text{NROWS}(i)$ for set $i$
8: \quad return sum(MSE_i) \quad \triangleright \text{The sum of the mean square error for each set}
5.4.1 Covariate Selection for $\psi$

Table 5.17 shows the order of the covariates chosen for $\psi$ by Algorithms 2 and 3.

Table 5.17: Covariate selection for $\psi$.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Cross-validation error</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(AREA)</td>
<td>3.5470</td>
</tr>
<tr>
<td>BFIHOST$^2$</td>
<td>1.9799</td>
</tr>
<tr>
<td>log(SAAR)</td>
<td>1.6917</td>
</tr>
<tr>
<td>log(FARL)</td>
<td>1.4002</td>
</tr>
<tr>
<td>log(URBEXT+1)</td>
<td>1.3831</td>
</tr>
<tr>
<td>log(PROPWET)</td>
<td>1.3699</td>
</tr>
<tr>
<td>log(ASPVAR)</td>
<td>1.3710</td>
</tr>
<tr>
<td>log(LDP)$^*$</td>
<td>1.3698</td>
</tr>
<tr>
<td>log(DPSBAR)</td>
<td>1.3719</td>
</tr>
<tr>
<td>log(FPEXT)</td>
<td>1.3702</td>
</tr>
<tr>
<td>log(DPLBAR)</td>
<td>1.3720</td>
</tr>
<tr>
<td>ASPBAR</td>
<td>1.3787</td>
</tr>
<tr>
<td>log(ALTBAR)</td>
<td>1.3887</td>
</tr>
<tr>
<td>log(SPRHOST)</td>
<td>1.4074</td>
</tr>
</tbody>
</table>

* Lowest cross-validation error.

Table 5.17 and Figure 5.33 show that the cross-validation error is at its lowest point with eight covariates but since LDP and AREA have a high positive correlation of 0.96 the chosen covariates will be: AREA, BFIHOST, FARL, SAAR, URBEXT and PROPWET. $X_\psi$ will thus include an intercept and these six covariates.
5 Data and Preliminary Analysis

![Graph showing the lowest cross-validation error as a function of the number of covariates (including an intercept) for the ψ model.](image)

**Figure 5.33:** The lowest cross-validation error as a function of the number of covariates (including an intercept) for the ψ model.

Table 5.18 shows the estimates from the model described in Chapter 5.4 for the coefficients corresponding to the covariates chosen for the ψ model. From Table 5.18 it can be seen that none of the coefficients contain zero in their 95% posterior intervals.
5.4 Covariate Selection

Table 5.18: Fixed effects for $\psi$.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Sd.</th>
<th>0.025%</th>
<th>50%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-10.171</td>
<td>0.967</td>
<td>-12.071</td>
<td>-10.171</td>
<td>-8.274</td>
</tr>
<tr>
<td>log(AREA)</td>
<td>0.898</td>
<td>0.013</td>
<td>0.872</td>
<td>0.898</td>
<td>0.923</td>
</tr>
<tr>
<td>BFIHOST$^2$</td>
<td>-3.145</td>
<td>0.138</td>
<td>-3.417</td>
<td>-3.145</td>
<td>-2.875</td>
</tr>
<tr>
<td>log(SAAR)</td>
<td>1.495</td>
<td>0.125</td>
<td>1.249</td>
<td>1.495</td>
<td>1.739</td>
</tr>
<tr>
<td>log(FARL)</td>
<td>3.735</td>
<td>0.315</td>
<td>3.116</td>
<td>3.736</td>
<td>4.354</td>
</tr>
<tr>
<td>log(URBEXT+1)</td>
<td>1.405</td>
<td>0.383</td>
<td>0.653</td>
<td>1.405</td>
<td>2.157</td>
</tr>
<tr>
<td>log(PROPWET)</td>
<td>0.465</td>
<td>0.157</td>
<td>0.152</td>
<td>0.466</td>
<td>0.771</td>
</tr>
</tbody>
</table>

5.4.2 Covariate Selection for $\tau$.

Table 5.19 shows the order of the covariates chosen by Algorithms 2 and 3 according to $\tau$. The order of the selected covariates according to Tables 5.19 and 5.17 are similar but not the same. For example, FPEXT is the first covariate chosen for $\tau$ while it was the tenth covariate selected for $\psi$. FPEXT must then be a good predictor for the scale parameter $\sigma$ since the definition of $\tau$ is $\tau = \log(\sigma/\mu) = \log(\sigma) - \psi$ and $\sigma$ is such that $\log(\sigma) = \tau + \psi$. Furthermore, for a given site $j$, the model for $\log(\sigma_j)$ is

$$\log(\sigma_j) = x_{\tau,j}\beta_\tau + x_{\psi,j}\beta_\psi + a_j(u_\tau + u_\psi) + \epsilon_{\tau,j} + \epsilon_{\psi,j}$$

where $x_{\tau,j}$, $x_{\psi,j}$ and $a_j$ are the $j$-th rows of $X_\tau$, $X_\psi$ and $A$, respectively, $\epsilon_{\tau,j}$ and $\epsilon_{\psi,j}$ are the $j$-th elements of $\epsilon_\tau$ and $\epsilon_\psi$, respectively. If the $k$-th covariate is the same for $\tau$ and $\psi$ then the one unit change in the covariate $x_{k,j} = x_{\tau,k,j} = x_{\psi,k,j}$ on $\log(\sigma_j)$ is $\beta_{\tau,k} + \beta_{\psi,k}$. This means that $\beta_{\tau,k}$ is the additional effect of one unit change in the covariate $x_k$ on $\log(\sigma)$ after taking $\beta_{\psi,k}$ into account.
Table 5.19: Covariate selection for $\tau$.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Cross-validation error</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(FPEXT)</td>
<td>0.5130</td>
</tr>
<tr>
<td>log(ALTBAR)</td>
<td>0.4995</td>
</tr>
<tr>
<td>log(AREA)</td>
<td>0.4904</td>
</tr>
<tr>
<td>log(FARL)</td>
<td>0.4785</td>
</tr>
<tr>
<td>log(URBEXT+1)</td>
<td>0.4695</td>
</tr>
<tr>
<td>log(SAAR)</td>
<td>0.4641</td>
</tr>
<tr>
<td>log(PROPWET)*</td>
<td>0.4625</td>
</tr>
<tr>
<td>log(ASPVAR)</td>
<td>0.4632</td>
</tr>
<tr>
<td>log(DPSBAR)</td>
<td>0.4642</td>
</tr>
<tr>
<td>BFIHOST$^2$</td>
<td>0.4656</td>
</tr>
<tr>
<td>ASPBAR</td>
<td>0.4676</td>
</tr>
<tr>
<td>log(LDP)</td>
<td>0.4697</td>
</tr>
<tr>
<td>log(DPLBAR)</td>
<td>0.4713</td>
</tr>
<tr>
<td>log(SPRHOST)</td>
<td>0.4737</td>
</tr>
</tbody>
</table>

* Lowest cross-validation error.

Figure 5.33 shows that the cross-validation error is lowest using a model with seven covariates namely: FPEXT, ALTBAR, AREA, FARL, URBEXT, SAAR and PROPWET. The matrix $X_\tau$ will thus contain an intercept as well as these seven covariates.
Figure 5.34: The lowest cross-validation error as a function of the number of covariates (including an intercept) for the $\tau$ model.

Table 5.20 shows the estimates from the model described in Chapter 5.4 for the coefficients corresponding to the covariates chosen for the $\tau$ model. From Table 5.20 it can be seen that only ALTBAR contains zero in its marginal posterior interval.
Table 5.20: Fixed effects for $\tau$.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Sd.</th>
<th>0.025%</th>
<th>50%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.417</td>
<td>0.614</td>
<td>1.211</td>
<td>2.417</td>
<td>3.624</td>
</tr>
<tr>
<td>log(FPEXT)</td>
<td>-0.129</td>
<td>0.026</td>
<td>-0.181</td>
<td>-0.129</td>
<td>-0.078</td>
</tr>
<tr>
<td>log(ALTBAR)</td>
<td>0.012</td>
<td>0.045</td>
<td>-0.077</td>
<td>0.012</td>
<td>0.101</td>
</tr>
<tr>
<td>log(AREA)</td>
<td>-0.050</td>
<td>0.009</td>
<td>-0.068</td>
<td>-0.050</td>
<td>-0.032</td>
</tr>
<tr>
<td>log(FARL)</td>
<td>-1.080</td>
<td>0.195</td>
<td>-1.463</td>
<td>-1.080</td>
<td>-0.697</td>
</tr>
<tr>
<td>log(URBEXT+1)</td>
<td>-0.839</td>
<td>0.226</td>
<td>-1.283</td>
<td>-0.838</td>
<td>-0.395</td>
</tr>
<tr>
<td>log(SAAR)</td>
<td>-0.561</td>
<td>0.092</td>
<td>-0.741</td>
<td>-0.561</td>
<td>-0.381</td>
</tr>
<tr>
<td>log(PROPWET)</td>
<td>-0.240</td>
<td>0.100</td>
<td>-0.436</td>
<td>-0.241</td>
<td>-0.041</td>
</tr>
</tbody>
</table>

5.4.3 Covariate Selection for $\xi$

Table 5.21 shows the order of the covariates chosen for $\xi$ with Algorithm 2 and 3 and it also shows the cross-validation error.

Table 5.21: Covariate selection for $\xi$.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Cross-validation error</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(FPEXT)</td>
<td>0.4507</td>
</tr>
<tr>
<td>log(URBEXT+1)</td>
<td>0.4475</td>
</tr>
<tr>
<td>log(SAAR)</td>
<td>0.4464</td>
</tr>
<tr>
<td>log(PROPWET)</td>
<td>0.4452</td>
</tr>
<tr>
<td>log(FARL)*</td>
<td>0.4449</td>
</tr>
<tr>
<td>ASPBAR</td>
<td>0.4455</td>
</tr>
<tr>
<td>BFIHOST$^2$</td>
<td>0.4461</td>
</tr>
<tr>
<td>log(SPRHOST)</td>
<td>0.4464</td>
</tr>
<tr>
<td>log(ALTBAR)</td>
<td>0.4473</td>
</tr>
<tr>
<td>log(ASPVAR)</td>
<td>0.4483</td>
</tr>
<tr>
<td>log(DPSBAR)</td>
<td>0.4500</td>
</tr>
<tr>
<td>log(AREA)</td>
<td>0.4523</td>
</tr>
<tr>
<td>log(DPLBAR)</td>
<td>0.4537</td>
</tr>
<tr>
<td>log(LDP)</td>
<td>0.4552</td>
</tr>
</tbody>
</table>

* Lowest cross-validation error.

Figure 5.35 shows that the lowest cross-validation error comes using five covariates, FPEXT, URBEXT, SAAR, PROPWET and FARL along with an intercept. The matrix $X_\xi$ will thus contain these five covariates and an intercept.
Figure 5.35: The lowest cross-validation error as a function of the number of covariates (including an intercept) for the $\xi$ model.

Table 5.22 shows that two of the coefficients, FARL and URBEXT, contain zero at the 95% posterior interval.
Table 5.22: Fixed effects for $\xi$.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Sd.</th>
<th>0.025%</th>
<th>50%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.900</td>
<td>0.348</td>
<td>0.216</td>
<td>0.900</td>
<td>1.585</td>
</tr>
<tr>
<td>log(FPEXT)</td>
<td>-0.089</td>
<td>0.015</td>
<td>-0.118</td>
<td>-0.089</td>
<td>-0.059</td>
</tr>
<tr>
<td>log(URBEXT+1)</td>
<td>0.276</td>
<td>0.148</td>
<td>-0.015</td>
<td>0.276</td>
<td>0.567</td>
</tr>
<tr>
<td>log(SAAR)</td>
<td>-0.161</td>
<td>0.047</td>
<td>-0.253</td>
<td>-0.161</td>
<td>-0.068</td>
</tr>
<tr>
<td>log(PROPWET)</td>
<td>0.117</td>
<td>0.051</td>
<td>0.018</td>
<td>0.117</td>
<td>0.217</td>
</tr>
<tr>
<td>log(FARL)</td>
<td>-0.255</td>
<td>0.144</td>
<td>-0.538</td>
<td>-0.255</td>
<td>0.029</td>
</tr>
</tbody>
</table>
6 Results

6.1 Convergence Assessment

Convergence in the simulation for the model described in Chapter 4.2 was explored by applying trace plots, histogram plots, the Gelman-Rubin statistic and plots of the sample autocorrelation to the simulated MCMC samples.

**Trace plots:** Trace plots are a useful way to visualize the MCMC simulations. A trace plot shows the MCMC samples as a function of iterations. By looking closely at trace plots, one can see approximately when the MCMC simulation has converged and how many samples should potentially be discarded. This can be done by looking at the global trend of the trace plot when this trend is approximately zero it suggests the simulation has converged. By visually inspecting trace plots, it is possible to see whether the MCMC simulation gets stuck on a particular value. Trace plots can be seen in Figure 6.1.

**Histogram plots:** Histograms of the MCMC samples can be seen as an approximation of the marginal posterior density of a given parameter when the MCMC simulation has converged. Histogram plots can be seen in Figure 6.2

**The Gelman-Rubin statistic:** the Gelman-Rubin statistic is a way to estimate the factor by which the scale of the current distribution might be reduced if the simulations were continued in the limit $L \rightarrow \infty$ see Gelman & Rubin (1992). A Gelman-Rubin statistic close to 1 indicates that the MCMC simulations are close to the target distribution of the MCMC simulation. In most cases a value of the Gelman-Rubin statistic less than 1 is sufficient. The Gelman-Rubin statistic can be seen in Table 6.1

**Autocorrelation plots:** Plots of the sample autocorrelation indicate the quality of the MCMC simulations by looking at the autocorrelation as a function of lag. The $i$-th lag autocorrelation $r_i$ is the correlation between each $i$-th successive sample of the MCMC simulation. With a computationally efficient MCMC algorithm, the $i$-th lag autocorrelation should decrease rapidly as a function of $i$. Autocorrelation plots can be seen in Figures 6.3, 6.4 and 6.5.

To investigate the MCMC simulation one station was chosen at random, Station 8. For that station the parameters $\psi_8$, $\tau_8$ and $\xi_8$ were chosen. Furthermore, the covariate coefficients $\beta_{\psi,\text{AREA}}$, $\beta_{\tau,\text{SAAR}}$ and $\beta_{\xi,\text{FPEXT}}$ and the hyperparameters $q_\psi$, $s_\psi$ and $\rho_\tau$ where also chosen to investigate the convergence of the MCMC simulation.
Table 6.1: Gelman-Rubin statistic for the parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Point est.</th>
<th>Upper C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi_8$</td>
<td>1.0000</td>
<td>1.0002</td>
</tr>
<tr>
<td>$\beta_{\psi,\text{AREA}}$</td>
<td>0.9999</td>
<td>1.0000</td>
</tr>
<tr>
<td>$q_\psi$</td>
<td>1.0072</td>
<td>1.0149</td>
</tr>
<tr>
<td>$\tau_8$</td>
<td>1.0000</td>
<td>1.0003</td>
</tr>
<tr>
<td>$\beta_{\tau,\text{SAAR}}$</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>$s_\psi$</td>
<td>1.0074</td>
<td>1.0203</td>
</tr>
<tr>
<td>$\xi_8$</td>
<td>1.0000</td>
<td>1.0001</td>
</tr>
<tr>
<td>$\beta_{\xi,\text{FPEXT}}$</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>$\rho_\tau$</td>
<td>1.0017</td>
<td>1.0023</td>
</tr>
</tbody>
</table>

For the MCMC simulation, four chains of 10,000 samples with a burn-in of 2,000 samples were obtained. In this chapter, only the 8,000 samples, that are left after removing the burn-in are, shown. Table 6.1 shows the Gelman-Rubin statistic, the statistic is less than 1.1 for all the parameters tested. The trace plots in Figure 6.1 are all reasonable, i.e., there is no visual global trend, and the chains blend well. Figure 6.2 show the histograms for each MCMC chains overlaid on top of each other, the histograms blend well and distinguishing between the chains is not possible. The plots in Figures 6.3, 6.4 and 6.5 show the autocorrelation of the MCMC simulation for each chain as a function of lag, one can see that the autocorrelation decreases rapidly as a function of lag for the covariate coefficients, $\beta_{\psi,\text{AREA}}$, $\beta_{\tau,\text{SAAR}}$ and $\beta_{\xi,\text{FPEXT}}$, and the parameters $\psi_8$, $\tau_8$ and $\xi_8$. The autocorrelation for the hyperparameters, $q_\psi$, $s_\psi$ and $\rho_\tau$ also decreases as a function of lag but not as rapidly as the others parameters. However, Figure 6.5, which shows the sample autocorrelation of the hyperparameters, indicates good properties of the MCMC algorithm.
6.1 Convergence Assessment

Figure 6.1: Trace plots for the parameters.

Figure 6.2: Histogram plots for the parameters.
Figure 6.3: Autocorrelation plots for $\psi_8$, $\tau_8$ and $\xi_8$. The four horizontal panels represent each of the four chains.

Figure 6.4: Autocorrelation plots for $\beta_{\psi,\text{AREA}}$, $\beta_{\tau,\text{SAAR}}$ and $\beta_{\xi,\text{FPEXT}}$. The four horizontal panels represent each of the four chains.
6.1 Convergence Assessment

Figure 6.5: Autocorrelation plots for the hyperparameters $q_\psi$, $s_\psi$, and $\rho_\tau$. The four horizontal panels represent each of the four chains.
6.2 Hyperparameters

Figures 6.6 and 6.7 show the hyperparameters that govern the spatial fields for $\psi$ and $\tau$, respectively. These parameters had joint prior densities $\pi(s_\psi, \rho_\psi)$ and $\pi(s_\tau, \rho_\tau)$, thus, the joint posterior densities of $(s_\psi, \rho_\psi)$ and $(s_\tau, \rho_\tau)$ along with the marginal densities of each parameter are shown in Figures 6.6 and 6.7.

The posterior densities for the hyperparameters that govern the unstructured (iid.) model errors, $\epsilon_\psi$, $\epsilon_\tau$ and $\epsilon_\xi$, namely, $q_\psi$, $\sigma_{\psi\epsilon}$, $q_\tau$, $\sigma_{\tau\epsilon}$, $q_\xi$ and $\sigma_{\xi\epsilon}$, are shown in Figure 6.8. Note that $q_\psi = 1/\sigma_{\psi\epsilon}^2$, $q_\tau = 1/\sigma_{\tau\epsilon}^2$, and $q_\xi = 1/\sigma_{\xi\epsilon}^2$.

Table 6.2 displays summary statistics for the hyperparameters. It is interesting to compare the model errors with the marginals standard deviation of the spatial component for the $\psi$ and $\tau$ model, in both cases the marginal standard deviation of the spatial component is bigger than the standard deviation of the model errors $\epsilon_\psi$ and $\epsilon_\tau$. Figure 6.9 shows a more visual comparison of these hyperparameters by comparing the posterior means and the 95% credible intervals of $\sigma_{\psi\epsilon}$, $s_\psi$, $\sigma_{\tau\epsilon}$ and $s_\tau$. Fitting simplified models for $\psi$ and $\tau$, similar to the models used for forward selection in Chapter 5.4, without a spatial component using the same covariates as in the full model gave $\sigma_{\psi\epsilon}^* = 0.38$ and $\sigma_{\tau\epsilon}^* = 0.20$. Here $\sigma_{\psi\epsilon}^*$ and $\sigma_{\tau\epsilon}^*$ are the posterior means of the standard deviations of the model errors of the simplified models for $\psi$ and $\tau$ without the spatial component. This result is not surprising since

$$\sigma_{\psi\epsilon}^* = 0.383 \approx \sqrt{s_{\psi,\text{post}}^2 + \sigma_{\psi\epsilon,\text{post}}^2} = \sqrt{0.307^2 + 0.243^2} = 0.392$$

and

$$\sigma_{\tau\epsilon}^* = 0.199 \approx \sqrt{s_{\tau,\text{post}}^2 + \sigma_{\tau\epsilon,\text{post}}^2} = \sqrt{0.173^2 + 0.128^2} = 0.215,$$

where $s_{\psi,\text{post}}^2$, $\sigma_{\psi\epsilon,\text{post}}^2$, $s_{\tau,\text{post}}^2$ and $\sigma_{\tau\epsilon,\text{post}}^2$ are the posterior means of $s_{\psi}^2$, $\sigma_{\psi\epsilon}^2$, $s_{\tau}^2$ and $\sigma_{\tau\epsilon}^2$, respectively. This fact means that the variability that the spatial component is explaining would otherwise be unexplained.
6.2 Hyperparameters

Table 6.2: Summary statistics for the posterior distribution of the hyperparameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Sd</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_\psi)</td>
<td>0.243</td>
<td>0.019</td>
<td>0.208</td>
<td>0.243</td>
<td>0.281</td>
</tr>
<tr>
<td>(s_\psi)</td>
<td>0.307</td>
<td>0.025</td>
<td>0.258</td>
<td>0.308</td>
<td>0.355</td>
</tr>
<tr>
<td>(\rho_\psi)</td>
<td>0.053</td>
<td>0.011</td>
<td>0.037</td>
<td>0.052</td>
<td>0.077</td>
</tr>
</tbody>
</table>

Hyperparameters for \(\tau\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Sd</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_\tau)</td>
<td>0.128</td>
<td>0.012</td>
<td>0.102</td>
<td>0.128</td>
<td>0.151</td>
</tr>
<tr>
<td>(s_\tau)</td>
<td>0.173</td>
<td>0.018</td>
<td>0.141</td>
<td>0.172</td>
<td>0.212</td>
</tr>
<tr>
<td>(\rho_\tau)</td>
<td>0.090</td>
<td>0.022</td>
<td>0.057</td>
<td>0.087</td>
<td>0.142</td>
</tr>
</tbody>
</table>

Hyperparameters for \(\xi\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Sd</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_\xi)</td>
<td>0.152</td>
<td>0.008</td>
<td>0.138</td>
<td>0.152</td>
<td>0.168</td>
</tr>
</tbody>
</table>

Figure 6.6: The joint posterior distribution of \(\rho_\psi\) and \(s_\psi\), and the marginal posterior densities of \(\rho_\psi\) and \(s_\psi\).
Figure 6.7: The joint posterior distribution of $\rho_\tau$ and $s_\tau$, and the marginal posterior densities of $\rho_\tau$ and $s_\tau$. 
6.2 Hyperparameters

Figure 6.8: The marginal posterior density of $q_\psi$, $q_\tau$ and $q_\xi$ along with the marginal posterior density of $\sigma_{\psi\epsilon}$, $\sigma_{\tau\epsilon}$ and $\sigma_{\xi\epsilon}$. 
Figure 6.9: The standard deviations of the error components and the spatial components for the $\psi$ and $\tau$ models, namely, $\sigma_{\psi e}$, $s_\psi$, $\sigma_{\tau e}$ and $s_\tau$. 
6.3 Latent Parameters

In this chapter results based on the posterior density for the latent parameters of the model, \( x = (\eta^T, \nu^T)^T \), are presented. Recall that \( \eta \) contain the latent parameters which are found in the likelihood, that is

\[
\eta = (\psi^T, \tau^T, \xi^T)^T.
\]

The vector \( \nu \) contains the covariate coefficients, \( \beta_\psi, \beta_\tau \) and \( \beta_\xi \), and the spatial components \( u_\psi \) and \( u_\tau \), that is

\[
\nu = (\beta_\psi^T, u_\psi^T, \beta_\tau^T, u_\tau^T, \beta_\xi^T)^T.
\]

6.3.1 The Latent Parameters in the Likelihood

Figures 6.10 and 6.11 show the posterior mean and standard deviation for \( \psi \), Figures 6.12 and 6.13 show the posterior mean and standard deviation for \( \tau \) and Figures 6.14 and 6.15 show the posterior mean and standard deviation for \( \xi \).
6 Results

Figure 6.10: Posterior mean of $\psi$.

Figure 6.11: Posterior standard deviation of $\psi$. 
Figure 6.12: Posterior mean of $\tau$.

Figure 6.13: Posterior standard deviation of $\tau$. 
Figure 6.14: Posterior mean of $\xi$.

Figure 6.15: Posterior standard deviation of $\xi$. 
## 6.3 Latent Parameters

### 6.3.2 Covariate Coefficients

In this chapter, a closer look is taken at the covariate coefficients, \( \beta_\psi, \beta_\tau \) and \( \beta_\xi \), for \( \psi, \tau \) and \( \xi \), respectively. Table 6.3 shows the posterior mean, standard deviation and quantiles for the covariate coefficients, \( \beta_\psi, \beta_\tau \) and \( \beta_\xi \). From the table, one can see that only \( \beta_{\tau,\text{ALTBAR}}, \beta_{\xi,\text{FARL}} \) and \( \beta_{\xi,\text{URBEXT}} \) contain zero on the 95% interval. This is in line with the results of the simplified model used in Chapter 5.4, see Tables 5.18, 5.20 and 5.22.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Mean</th>
<th>Sd.</th>
<th>0.025</th>
<th>0.50</th>
<th>0.975</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Covariate coefficients for ( \psi )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-10.147</td>
<td>0.969</td>
<td>-12.052</td>
<td>-10.149</td>
<td>-8.249</td>
</tr>
<tr>
<td>log(AREA)</td>
<td>0.896</td>
<td>0.013</td>
<td>0.871</td>
<td>0.896</td>
<td>0.922</td>
</tr>
<tr>
<td>log(SAAR)</td>
<td>1.492</td>
<td>0.125</td>
<td>1.246</td>
<td>1.492</td>
<td>1.737</td>
</tr>
<tr>
<td>log(FARL)</td>
<td>3.718</td>
<td>0.315</td>
<td>3.095</td>
<td>3.718</td>
<td>4.333</td>
</tr>
<tr>
<td>BFIHOST(^2)</td>
<td>-3.138</td>
<td>0.137</td>
<td>-3.406</td>
<td>-3.138</td>
<td>-2.869</td>
</tr>
<tr>
<td>log(URBEXT+1)</td>
<td>1.404</td>
<td>0.384</td>
<td>0.655</td>
<td>1.403</td>
<td>2.154</td>
</tr>
<tr>
<td>log(PROPWET)</td>
<td>0.460</td>
<td>0.158</td>
<td>0.149</td>
<td>0.460</td>
<td>0.766</td>
</tr>
<tr>
<td><strong>Covariate coefficients for ( \tau )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.430</td>
<td>0.634</td>
<td>1.196</td>
<td>2.430</td>
<td>3.670</td>
</tr>
<tr>
<td>log(AREA)</td>
<td>-0.048</td>
<td>0.009</td>
<td>-0.066</td>
<td>-0.048</td>
<td>-0.029</td>
</tr>
<tr>
<td>log(SAAR)</td>
<td>-0.573</td>
<td>0.095</td>
<td>-0.759</td>
<td>-0.572</td>
<td>-0.388</td>
</tr>
<tr>
<td>log(FARL)</td>
<td>-1.059</td>
<td>0.202</td>
<td>-1.457</td>
<td>-1.059</td>
<td>-0.662</td>
</tr>
<tr>
<td>log(FPEXT)</td>
<td>-0.125</td>
<td>0.027</td>
<td>-0.177</td>
<td>-0.125</td>
<td>-0.072</td>
</tr>
<tr>
<td>log(URBEXT+1)</td>
<td>-0.797</td>
<td>0.234</td>
<td>-1.253</td>
<td>-0.797</td>
<td>-0.341</td>
</tr>
<tr>
<td>log(PROPWET)</td>
<td>-0.244</td>
<td>0.104</td>
<td>-0.444</td>
<td>-0.245</td>
<td>-0.038</td>
</tr>
<tr>
<td>log(ALTBAR)*</td>
<td>0.023</td>
<td>0.047</td>
<td>-0.069</td>
<td>0.023</td>
<td>0.114</td>
</tr>
<tr>
<td><strong>Covariate coefficients for ( \xi )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.877</td>
<td>0.363</td>
<td>0.171</td>
<td>0.875</td>
<td>1.591</td>
</tr>
<tr>
<td>log(PROPWET)</td>
<td>0.116</td>
<td>0.053</td>
<td>0.014</td>
<td>0.116</td>
<td>0.221</td>
</tr>
<tr>
<td>log(SAAR)</td>
<td>-0.156</td>
<td>0.049</td>
<td>-0.253</td>
<td>-0.156</td>
<td>-0.060</td>
</tr>
<tr>
<td>log(FARL)*</td>
<td>-0.268</td>
<td>0.151</td>
<td>-0.562</td>
<td>-0.269</td>
<td>0.028</td>
</tr>
<tr>
<td>log(FPEXT)</td>
<td>-0.085</td>
<td>0.016</td>
<td>-0.116</td>
<td>-0.085</td>
<td>-0.054</td>
</tr>
<tr>
<td>log(URBEXT+1)*</td>
<td>0.281</td>
<td>0.155</td>
<td>-0.024</td>
<td>0.282</td>
<td>0.585</td>
</tr>
</tbody>
</table>

*Contains zero on the 95% credible interval.

The location and scale parameters, \( \mu \) and \( \sigma \) of the GEV density can be written in terms of \( \psi \) and \( \tau \) for station \( k \) as

\[
\mu_k = \exp(\psi_k) = \exp(x_{k,\psi}\beta_\psi + a_ku_\psi + \epsilon_{\psi,k})
\]
6 Results

and

\[ \sigma_k = \exp(\psi_k + \tau_k) = \exp(x_{k,\psi} \beta_{\psi} + x_{k,\tau} \beta_{\tau} + a_k(u_k + u_{\psi}) + \epsilon_{\psi,k} + \epsilon_{\tau,k}). \]

Using the posterior means of the covariates coefficients from Table 6.3 the location and scale parameters can be estimated as

\[
\mu_k = e^{-10.147(\text{AREA}_k)^{0.896}(\text{SAAR}_k)^{1.492}(\text{FARL}_k)^{3.718}c^{-3.138(\text{BFIHOST}_k)^2}} \\
\times (\text{URBEXT}_k + 1)^{1.404}(\text{PROPWET}_k)^{0.460}e^{a_k(u_k + u_{\psi}) + \epsilon_{\psi,k}}, \tag{6.1}
\]

\[
\sigma_k = e^{-7.717(\text{AREA}_k)^{0.848}(\text{SAAR}_k)^{0.919}(\text{FARL}_k)^{2.659}} \\
\times (\text{URBEXT}_k + 1)^{0.607}(\text{PROPWET}_k)^{0.216}(\text{ALTBAR}_k)^{0.023} \\
\times (\text{FPEXT}_k)^{-0.125}e^{a_k(u_k + u_{\tau}) + \epsilon_{\psi,k} + \epsilon_{\tau,k}}. \tag{6.2}
\]

The effect of the covariates that are used both for the \( \psi \) and the \( \tau \) model on \( \sigma \) are added together, for example, the effect of \( \text{AREA} \) on \( \sigma_k \) is

\[ 0.896 - 0.048 = 0.848 \].

The shape parameter, \( \xi \), has a linear model at the latent level and can be estimated as

\[
\xi_k = 0.877 + 0.116 \log(\text{PROPWET}_k) - 0.156 \log(\text{SAAR}_k) - 0.268 \log(\text{FARL}_k) \\
- 0.085 \log(\text{FPEXT}_k) + 0.281 \log(\text{URBEXT}_k + 1) + \epsilon_{\xi,k}. \tag{6.3}
\]

Equation 6.1 shows that the effect of \( \text{AREA} \) on \( \mu \) is not exactly linear and the slope decreases with increased catchment \( \text{AREA} \). The relationship between \( \text{SAAR} \) and \( \mu \) is not linear, in fact, the slope increases as a function of \( \text{SAAR} \). The estimates for the covariate coefficients \( \text{AREA} \) and \( \text{SAAR} \), for the location parameter, \( \mu \), is not similar to the results in Kjeldsen & Jones (2009a), although, they modeled median annual maximum flow, for each station, instead of the location parameter, \( \mu \), of the GEV density. Equation 6.2 tells the same story for \( \sigma \) with respect to \( \text{AREA} \), i.e., not a linear relationship and the slope decreases as a function of \( \text{AREA} \).

Table 6.4 shows how much \( \mu_k \), \( \sigma_k \) and \( \xi_k \) change when a covariate goes from its median to its third quartile and the change when a covariate goes from its median to its first quartile (keeping other covariates fixed and assuming the same location and model error). For example, if \( \mu_i \) and \( \mu_j \) are such that the catchment area for station \( i \) is the 0.5 quantile (median) of \( \text{AREA} \) and station \( j \) has a catchment area equal to the 0.75 quantile (third quartile) of \( \text{AREA} \), but otherwise these two catchments would have the same covariates, spatial component and model error then \( \mu_j = 2.090\mu_i \) (see the first row in Table 6.4). So, the effect is multiplicative for the location and scale parameters, \( \mu \) and \( \sigma \), while for the shape parameter, \( \xi \), the effect is additive. For example, given stations \( i \) and \( j \) such that the average annual rainfall (\( \text{SAAR} \)) is the 0.75 quantile of \( \text{SAAR} \) for station \( j \) and the average annual rainfall (\( \text{SAAR} \)) is the 0.5 quantile (median) of \( \text{SAAR} \) for station \( i \), but other covariates and random effects are the same, then \( \xi_j = \xi_i - 0.041 \). This is one way of seeing the relative importance of the covariates on the location, scale and shape parameters of the GEV density.
Table 6.4: The effect of going from the median to the 1st and 3rd quartile (0.25 quantile and 0.75 quantile), for each covariate (keeping other covariates and random effects fixed), on the location parameter, $\mu$, the scale parameter, $\sigma$, and the shape parameters, $\xi$, of the GEV density. For $\mu$ and $\sigma$ the effect is multiplicative but for $\xi$ the effect is additive.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Median to 0.75 quantile</th>
<th>Median to 0.25 quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AREA</td>
<td>2.090</td>
<td>0.485</td>
</tr>
<tr>
<td>SAAR</td>
<td>1.475</td>
<td>0.702</td>
</tr>
<tr>
<td>FARL</td>
<td>1.058</td>
<td>0.887</td>
</tr>
<tr>
<td>BFIHOST</td>
<td>0.756</td>
<td>1.177</td>
</tr>
<tr>
<td>URBEXT</td>
<td>1.031</td>
<td>0.992</td>
</tr>
<tr>
<td>PROPWET</td>
<td>1.098</td>
<td>0.857</td>
</tr>
<tr>
<td>$\sigma$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AREA</td>
<td>2.009</td>
<td>0.504</td>
</tr>
<tr>
<td>SAAR</td>
<td>1.271</td>
<td>0.804</td>
</tr>
<tr>
<td>FARL</td>
<td>1.041</td>
<td>0.918</td>
</tr>
<tr>
<td>FPEXT</td>
<td>0.955</td>
<td>1.050</td>
</tr>
<tr>
<td>URBEXT</td>
<td>1.013</td>
<td>0.997</td>
</tr>
<tr>
<td>PROPWET</td>
<td>1.045</td>
<td>0.930</td>
</tr>
<tr>
<td>ALTBAR</td>
<td>1.010</td>
<td>0.989</td>
</tr>
<tr>
<td>$\xi$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROPWET</td>
<td>0.024</td>
<td>-0.039</td>
</tr>
<tr>
<td>SAAR</td>
<td>-0.041</td>
<td>0.037</td>
</tr>
<tr>
<td>FARL</td>
<td>-0.004</td>
<td>0.009</td>
</tr>
<tr>
<td>FPEXT</td>
<td>-0.031</td>
<td>0.033</td>
</tr>
<tr>
<td>URBEXT</td>
<td>0.006</td>
<td>-0.002</td>
</tr>
</tbody>
</table>
6 Results

6.3.3 Spatial Components

Figures 6.16 and 6.17 show the mean and standard deviation of $A\mu$, that is, the spatial component for $\psi$ at each of the observed sites. Figures 6.18 and 6.19 show the mean and standard deviation of $A\tau$, the spatial components of $\tau$ at each observed site. Figures 6.20 and 6.21 show the posterior mean and standard deviation of $A\psi + A\tau$, which represents the spatial component of $\log(\sigma)$ since for a station $k$

$$\tau_k = \log(\sigma_k / \mu_k) = \log(\sigma_k) - \psi_k$$

and thus

$$\log(\sigma_k) = \psi_k + \tau_k = (x_{\psi,k}\beta_{\psi} + x_{\tau,k}\beta_{\tau}) + (a_k \mu_{\psi} + a_k \mu_{\tau}) + (\epsilon_{\psi,k} + \epsilon_{\tau,k}).$$

The spatial components $u_{\psi}$ and $u_{\tau}$ have a substantial effect on $\psi$ and $\tau$, respectively. Under the latent model, (4.1), the expected value of $\psi_k$ and $\tau_k$, for station, $k$, at a given location with covariates $x_{\psi,k}$ and $x_{\tau,k}$ is

$$\psi_k^* = E(\psi_k | \beta_{\psi}, u_{\psi}) = x_{\psi,k}\beta_{\psi} + a_k \mu_{\psi},$$

$$\tau_k^* = E(\tau_k | \beta_{\tau}, u_{\tau}) = x_{\tau,k}\beta_{\tau} + a_k \mu_{\tau}. \quad (6.4)$$

To quantify the effect of the spatial fields on $\mu_k$ and $\sigma_k$, one can imagine a scenario where two catchments have exactly the same covariates but have different locations in the UK. The expected value of $\exp(\psi_k)$ and $\exp(\tau_k)$ under the latent model are given by

$$E(\exp(\psi_k) | \beta_{\psi}, u_{\psi}) = \exp(x_{\psi,k}\beta_{\psi} + a_k \mu_{\psi} + \sigma_{\psi k}^2/2),$$

$$E(\exp(\tau_k) | \beta_{\tau}, u_{\tau}) = \exp(x_{\tau,k}\beta_{\tau} + a_k \mu_{\tau} + \sigma_{\tau k}^2/2). \quad (6.5)$$

Using the expected values in Equation 6.5, the fact that $\mu_k = \exp(\psi_k)$ and $\sigma_k = \exp(\psi_k + \tau_k)$, and that $\psi_k$ and $\tau_k$ are independent at the latent level, one can estimate the location and scale parameters of stations $i$ and $j$ as

$$E(\mu_i | \beta_{\psi}, u_{\psi}) = \mu_i^* = \exp(x_{\psi,i}\beta_{\psi}) \exp(a_i \mu_{\psi}) \exp(2\sigma_{\psi i}^2/2),$$

$$E(\mu_j | \beta_{\psi}, u_{\psi}) = \mu_j^* = \exp(x_{\psi,j}\beta_{\psi}) \exp(a_j \mu_{\psi}) \exp(2\sigma_{\psi j}^2/2),$$

$$E(\sigma_i | \beta_{\psi}, u_{\psi}, \beta_{\tau}, u_{\tau}) = \sigma_i^* = \exp(x_{\psi,i}\beta_{\psi} + x_{\tau,i}\beta_{\tau}) \exp(a_i \mu_{\psi} + a_i \mu_{\tau}) \exp(2\sigma_{\psi i}^2/2 + 2\sigma_{\tau i}^2/2),$$

$$E(\sigma_j | \beta_{\psi}, u_{\psi}, \beta_{\tau}, u_{\tau}) = \sigma_j^* = \exp(x_{\psi,j}\beta_{\psi} + x_{\tau,j}\beta_{\tau}) \exp(a_j \mu_{\psi} + a_j \mu_{\tau}) \exp(2\sigma_{\psi j}^2/2 + 2\sigma_{\tau j}^2/2).$$

Now if the same covariates are assumed for both stations then

$$\frac{\mu_i^*}{\mu_j^*} = \frac{\exp(x_{\psi,i}\beta_{\psi}) \exp(a_i \mu_{\psi}) \exp(2\sigma_{\psi i}^2/2)}{\exp(x_{\psi,j}\beta_{\psi}) \exp(a_j \mu_{\psi}) \exp(2\sigma_{\psi j}^2/2)} = \frac{\exp(a_i \mu_{\psi})}{\exp(a_j \mu_{\psi})},$$

$$\frac{\sigma_i^*}{\sigma_j^*} = \frac{\exp(x_{\psi,i}\beta_{\psi} + x_{\tau,i}\beta_{\tau}) \exp(a_i \mu_{\psi} + a_i \mu_{\tau}) \exp(2\sigma_{\psi i}^2/2 + 2\sigma_{\tau i}^2/2)}{\exp(x_{\psi,j}\beta_{\psi} + x_{\tau,j}\beta_{\tau}) \exp(a_j \mu_{\psi} + a_j \mu_{\tau}) \exp(2\sigma_{\psi j}^2/2 + 2\sigma_{\tau j}^2/2)} = \frac{\exp(a_i \mu_{\psi} + a_i \mu_{\tau})}{\exp(a_j \mu_{\psi} + a_j \mu_{\tau})}.$$

So, $\mu_i^*$ and $\sigma_i^*$ can be written in terms of $\mu_j^*$ and $\sigma_j^*$ as

$$\mu_i^* = \frac{\mu_j^* \mu_i^*}{\mu_j^*} = \frac{\mu_j^* \exp(a_i \mu_{\psi})}{\exp(a_j \mu_{\psi})},$$

$$\sigma_i^* = \frac{\sigma_j^* \sigma_i^*}{\sigma_j^*} = \frac{\sigma_j^* \exp(2\sigma_{\psi i}^2/2 + 2\sigma_{\tau i}^2/2)}{\exp(a_j \mu_{\psi} + a_j \mu_{\tau})}.$$
\[ \sigma_i^* = \sigma_j^* \frac{\sigma_i^*}{\sigma_j^*} = \sigma_j^* \frac{\exp(a_i u_\psi + a_i u_r)}{\exp(a_j u_\psi + a_j u_r)}. \]

The range of the posterior mean for \(Au_\psi\) is \((-1.25, 0.66)\) and if station \(i\) has \(a_i u_\psi = 0.66\) and station \(j\) has \(a_j u_\psi = -1.25\) then

\[ \mu_i^* = \frac{\mu_j^* \exp(0.66)}{\exp(-1.25)} = 6.75 \mu_j^*, \]

so the difference in the location parameter can potentially be substantial for similar catchments located at two different geographical locations. The range of \(Au_\psi + Au_r\) is \((-1.41, 0.84)\) so for stations \(i\) and \(j\) with \(a_i u_\psi + a_i u_r = 0.84\) and \(a_j u_\psi + a_j u_r = -1.41\) then

\[ \sigma_i^* = \sigma_j^* \frac{\exp(0.84)}{\exp(-1.41)} = 9.49 \sigma_j^*. \]

Here the most extreme cases were taken, but this gives an idea of how much effect the geographical location can have. Comparison of the 1st and 3rd quartile of the posterior mean of \(Au_\psi\) and \(Au_\psi + Au_r\) gives

\[ \mu_i^* = \frac{\mu_j^* \exp(0.17)}{\exp(-0.16)} = 1.39 \mu_j^*, \]

and

\[ \sigma_i^* = \frac{\sigma_j^* \exp(0.22)}{\exp(-0.16)} = 1.46 \sigma_j^*. \]
6 Results

Figure 6.16: Posterior mean of $A u_\psi$.

Figure 6.17: Posterior standard deviation of $A u_\psi$. 
6.3 Latent Parameters

Figure 6.18: Posterior mean of $A u_t$.

Figure 6.19: Posterior standard deviation of $A u_t$. 
6 Results

Figure 6.20: Posterior mean of $A\psi + A\tau$.

Figure 6.21: Posterior standard deviation of $A\psi + A\tau$.
6.3.4 Regression Analysis

Here the linear regression at the latent level is investigated. To investigate these models one can look at the distribution of $\hat{\eta}|\nu$.

$$\hat{\eta}|\nu \sim N(\hat{\eta}|Z\nu, Q_{\eta\nu}^{-1} + Q_{\eta y}^{-1})$$

where

$$Z\nu = \begin{bmatrix} X_\psi \beta_\psi + Au_\psi \\ X_\tau \beta_\tau + Au_\tau \\ X_\xi \beta_\xi \end{bmatrix}.$$  

Figure 6.22 shows $\hat{\eta}$ as a function of $Z\nu$ and residual plots based on the residuals, $\hat{\eta} - Z\nu$, as a function of $Z\nu$. $\hat{\eta}$ and $Z\nu$ are both partitioned into three vectors according to $\psi$, $\tau$ and $\xi$.

The residual plots in Figure 6.22 seem reasonable although the residual plot for $\psi$ indicates that the variance of $e_\psi$ decreases as a function of $X_\psi \beta_\psi + Au_\psi$. The three plots in the
left column showing $\hat{\eta}$ as a function of $Z\nu$ indicate a clear linear relationship for the $\psi$ and $\tau$ partitions of $\hat{\eta}$ and $Z\nu$, while the $\xi$ partition is not as clear.

### 6.4 Extreme Floods

To investigate the extreme events, e.g., the 100-year event, the corresponding quantiles need to be estimated. The cumulative density function of the GEV distribution is invertible which means that the quantile function has an explicit expression, the expression for the $p$-th quantile is

$$Q(p|\mu,\sigma,\xi) = \begin{cases} 
\mu + \sigma((-\log(p))^{-\xi} - 1)/\xi & \text{if } \xi \neq 0, p \in (0, 1] \\
\mu - \sigma \log(-\log(p)) & \text{if } \xi = 0, p \in (0, 1). 
\end{cases}$$

Four stations, the same stations as in Chapter 5.2, were chosen to show the estimation of the return levels and cdfs. Figure 6.23 shows the return levels as a function of the return period; the x-axis is on the log-scale. Figure 6.24 shows the empirical cdf of the observations of the four stations and the posterior mean of the cdf along with 95% credible bands. Figure 6.25 shows the ordered observations as a function of the return period, using the uncertainty in the order statistics for each station. In Appendix A one can see both quantile figures, similar to Figure 6.23, and figures showing the ordered data, similar to Figure 6.25, for 20 more stations that were randomly selected.
6.4 Extreme Floods

Figure 6.23: The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels.

Figure 6.24: The empirical cdf and the theoretical cdf along with a 95% credible interval for four stations.
6 Results

![Graphs showing flow vs return period for different stations]

*Figure 6.25: The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account).*

6.5 Goodness of Fit

To test the goodness of fit of the GEV model to the observed flow at each station the Anderson-Darling test from Stephens (1974) was used. The $p$-values of the Anderson-Darling test, using the posterior mean of $\mu$, $\sigma$ and $\xi$ as parameters in the GEV density, is presented in Figure 6.26. The Anderson-Darling test used here is not the Bayesian version but the classical one and is just used to give an idea of how well the GEV density, using the posterior means as parameters, fits the data. The null hypothesis of the Anderson-Darling test is that the data follows the specified distribution. This means that a $p$-value close to zero indicates that the data does not follow the specified distribution. The GEV model fits the data well according to Figure 6.26 since only five stations out of 553 had a $p$-value less than 0.05 and most of the stations had a $p$-value much closer to one.
Figure 6.26: P-values of the Anderson-Darling test assuming a GEV distribution with the posterior means of $\psi_k, \tau_k$ and $\xi_k$ for each station $k$. 
7 Conclusions

Latent Gaussian model (LGM) with a multivariate link function, referred to as an extended LGM, was proposed for the analysis of flood frequency for maximum annual peak data on flow from the UK. An approximation was made for the posterior inference since the dimension of the data was too large for Markov chain Monte Carlo based Bayesian inference scheme. This approximation made the inference computationally feasible. The model proposed is such that the data can be used without any transformation, such as normalisation since data from different catchments is assumed to come from different distributions. Normalisation is required for some commonly used flood frequency models such as the index flood method. The extended LGM framework has the property that it is easy to add covariates and additional random effects to the latent level without too many complications. This property of the model suited the data well since each catchment had a variety of catchment descriptors (covariates) and a geographical location.

PC-prior densities were defined a priori for the hyperparameters of the model. The hyperparameters in the extended LGM setup govern the latent parameters which are a part of an additive regression model, and thus the penalisation of increased complexity becomes essential to avoid overfitting.

Integrated nested Laplace approximation (INLA) allowed for inference of a simplified model where the covariance at the data level was ignored. By using the simplified model, it was possible to get some approximations of the hyperparameters, with the help of INLA, and use those approximations to make an informed user-defined scaling for the PC-prior densities of the hyperparameters.

Using the simplified model, with the help of INLA, a forward selection, with 10-fold cross-validation as criteria, was used to select the covariates for the additive regression model at the latent level. The use of 10-fold cross-validation as criteria helps avoid overfitting since it is based on out-of-sample error. Out of the selected covariates for \( \psi \), \( \tau \) and \( \xi \) only three contained zero on the 95% credible interval, and two of those were not close to being centred around zero. None of the covariates for the location parameter, \( \psi \), contained zero in the 95% credible interval.

The results show that all parameters of the model converged quickly and most parameters are well defined, in particular, the hyperparameters, which all had narrow posterior densities. The results show that the spatial fields for \( \psi \) and \( \tau \) explained much of the otherwise unexplained variability. This was demonstrated by fitting a model without the spatial components and comparing error terms of a model with and without spatial components.
7 Conclusions

The sum of the marginal variances of the spatial components and the variances of the unstructured error terms in the spatial model was approximately equal to the variances of the unstructured error terms in the model that had no spatial components, i.e., the variability explained by the spatial components would be unexplained without them present. The goodness of fit was also investigated by using the Anderson-Darling test for the GEV density. The $p$-values of the Anderson-Darling test showed that the GEV density fits the data well for most stations, only five stations out of 553 had a $p$-value less than 0.05.

The return period of extreme floods was investigated by computing the quantiles of the GEV distribution and taking into account the uncertainty in the order statistics of the observations. By looking at plots of the ordered observations as a function of return period with prediction intervals, one can see that the observations are, in most cases, contained in the prediction interval.

7.1 Future Work

Although the model introduced in this thesis has proven to be a good model for analysing flood frequency of the UK data, there is still room for improvements. When selecting covariates, using 10-fold cross-validation error as criteria, some covariates (catchment descriptors) contained zero on the 95% credible interval. These covariates deserve better analysis, especially the catchment descriptor ALTBAR when used to predict $\tau$ since it had a credible interval centred around zero. The reason for this may be the correlation between ALTBAR, SAAR and PROPWET. Covariate selection in a model with three linear models at the latent level must be done carefully, and improvements can most likely be made regarding this selection.

Modelling the shape parameter, $\xi$, of the GEV distribution caused problems. Attempts were made to use a spatial field, at the latent level, for the shape parameter similar to the spatial fields for the location and scale parameter. These attempts caused complications, and a decision was made to not include a spatial component for it. Modelling the shape parameter spatially is still something worth investigating and can possibly be achieved. A potential improvement would be to include a transformation on the shape parameter, similar to what is done for the location and scale parameter. In particular, it could be useful to constrain the shape parameter to an interval around zero.

One of the goals mentioned in the introduction, was making predictions for ungauged catchments. Even though this was not demonstrated in the thesis the model is still capable of these predictions. In fact, the simplified model, inferred using INLA, made out-of-sample predictions when selecting covariates for the additive regression model at the latent level. Making these predictions is of interest and should be considered in the future.
References


REFERENCES


A. Return Period Plots

In this appendix, 20 random stations were selected to investigate extreme events. The figures on the following five pages show return period plots, identical to Figures 6.23 and 6.25, for the 20 randomly selected stations.
Figure 7.1: The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels.

Figure 7.2: The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account).
Figure 7.3: The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels.

Figure 7.4: The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account).
Figure 7.5: The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels.

Figure 7.6: The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account).
Figure 7.7: The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels.

Figure 7.8: The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account).
Figure 7.9: The return levels as a function of the return period for four stations. The dashed lines show the 95% credible interval for the return levels.

Figure 7.10: The ordered observations as a function of return period along with a 95% prediction interval for the ordered observations (here the variability in the order statistics is taken into account).