Dissertation for the degree of doctor of philosophy

Computationally efficient Bayesian statistical modeling and inference for latent Gaussian models with an application to spatial extremes

Óli Páll Geirsson

UNIVERSITY OF ICELAND

School of Engineering and Natural Sciences
Faculty of Physical Sciences
Reykjavík, August 2015

Dissertation submitted in partial fulfilment of a *Philosophiae Doctor* degree in Statistics

Doctoral committee
Associate professor Birgir Hrafnkelsson, advisor
University of Iceland

Professor Gunnar Stefánsson,
University of Iceland

Professor Haraldur Ólafsson
University of Iceland

Opponents

Dr. Thordis L. Thorarinsdottir
Senior Research Scientist, Norwegian Computing Center

Dr. Alex Lenkoski
Senior Research Scientist, Norwegian Computing Center

Computationally efficient Bayesian statistical modeling and inference for latent Gaussian models with an application to spatial extremes

Printed in Iceland by Háskólaprent
ISBN 978-9935-9263-5-7

Abstract

The research project presented in this Ph.D. thesis aims to provide both a general and a computationally efficient approach to Bayesian statistical modeling and Bayesian inferential computations, with the focus on applications to spatial extremes. The research project can be can be split into the three following categories.

First, the author provides his stance towards a general approach to statistical analysis in practice. To that end, it is discussed how various regression models can be viewed as a broader class of a more general and flexible framework for statistical models. Furthermore, it is established that latent Gaussian models (LGMs) provide a flexible and interpretable class of statistical models for various statistical applications, in particular, for the statistical analysis provided in this thesis. A discussion is also given on how LGMs can be viewed as a specific extension of structured additive regression models, in the sense that, the data density function of each data point can depend on more than a single linear link function of the latent parameters.

Second, a novel computationally efficient Markov chain Monte Carlo (MCMC) sampling scheme for LGMs is proposed in this thesis. The sampling scheme is a two block Gibbs sampling scheme designed to exploit the model structure of LGMs. The author refers to the proposed sampling scheme as the MCMC split sampler in this thesis. The principle idea behind the MCMC split sampler is to split the latent Gaussian parameters into two vectors. The former vector consists of latent parameters which appear in the data density function while the latter vector consists of latent parameters which do not appear in it. The former vector is placed in the first block of the proposed sampling scheme and the latter vector is placed in the second block along with any potential hyperparameters. The resulting conditional posterior density functions within the blocks allow the MCMC split sampler to handle, by design, LGMs with latent models imposed on more than just the mean structure of the data density function. The MCMC split sampler is also designed to be applicable for any choice of a parametric data density function. Moreover, it scales well in terms of computational efficiency when the dimension of the latent model increases.

Third, a computationally efficient statistical method is proposed in this thesis to obtain distributional properties of annual maximum 24-hour precipitation on a 1 km by 1 km regular grid over Iceland. A covariate which is based on a local meteorological model that captures information on the physical processes

of precipitation is constructed, providing an additional spatial information on maximum precipitation. An LGM is built which takes into account observed maximum precipitation, the covariate based on the local meteorological model, and spatial variations. The observations are assumed to follow the generalized extreme value distribution, where spatial models based on approximate solutions to stochastic partial differential equations are implemented for the location, scale, and shape parameters of the data density function. The MCMC spilt sampler is applied as inferential algorithm, which in turn exploits the sparse matrices induced by the stochastic partial differential equation modeling, yielding continuous spatial predictions for spatially varying model parameters and quantiles. The construction of the meteorological covariate is extendable to any regions in the world where local meteorological models are available. The proposed modeling strategy is general in the sense that it is extendable to any spatial domain of interest.

Ágrip

Markmið doktorsritgerðarinnar er að setja fram almenna og skilvirka aðferðafræði fyrir bæði Bayesíska tölfræðilega líkanagerð og reiknilega Bayesíska ályktunartölfræði, með áherslu á staðsetningaháð útgildi. Rannsóknunum má skipta niður í þrjá meginflokka.

Fyrst er almenn afstaða tekin til tölfræðilegrar líkanagerðar. Leitt er í ljós að línuleg aðhvarfsgreining sé sértilfelli af almennri tölfræðilegri líkanagerð þar sem gagnaþéttleiki er settur fram til að lýsa gögnum og líkön sett fram til að lýsa stikum gagnaþéttleikans. Enn fremur er sýnt fram á að stigskipt Gaussísk líkön mynda sveigjanlegan og túlkanlegan flokk af tölfræðilíkönum sem hægt er að beita á ýmis tölfræðileg verkefni. Einnig er sýnt fram á að stigskipt Gaussísk líkön framlengja flokk almennra línulegra líkana í þeim skilningi að hægt er að setja fram líkön til að lýsa ekki einungis staðsetningarstika gagnaþéttleikans heldur einnig öðrum stikum gagnaþéttleikans eins og til dæmis skölunarstika hans.

Í öðru lagi er sett fram nýtt almennt Markov keðju Monte Carlo hermunar reiknirit fyrir Bayesíska ályktunartölfræði sem sniðið er að stigskiptum Gaussískum líkönum. Reikniritið er byggt á tveggja blokka Gibbs reikniriti og er hannað til að nýta stærðfræðilegt form stigskiptra Gaussískra líkana. Grunnhugmyndin á bak við reikniritið er að setja stika líkansins sem koma fram í gagnaþéttleikanum og þá stika sem koma ekki fram í honum í tvær aðskildar blokkir. Tilsvarandi stærðfræðilegt form skilyrtra eftiráþéttleika í hvorri blokk fyrir sig má nýta til að sníða reikniritið að stigskiptum Gaussískum líkönum þar sem líkön eru sett fram til að lýsa staðsetningarstika og skölunarstika gagnaþéttleikans sem og öðrum stikum hans. Reikniritið er hannað fyrir öll stigskipt Gaussísk líkön með stikuðum gagnaþéttleikum. Reikniritið heldur reiknilegri skilvirkni sinni með vaxandi fjölda stika í þeim hluta líkansins sem lýst er með Gaussískum þéttleika.

Að lokum er tölfræðilegt líkan sett fram til að lýsa gagnaþéttleika árlegrar hámarks sólarhringsúrkomu í hverjum punkti á þéttu neti yfir Íslandi. Annars vegar er sýnt fram á hvernig megi nýta upplýsingar úr veðurfræðilíkani sem aukalega upplýsingagátt fyrir hegðun hámarksúrkomu í öllum punktum þétta netsins þar sem engar mælingar á úrkomu eru til staðar. Aðferðin felur í sér smíði á skýribreytum í öllum punktum þétta netsins sem byggðar eru á upplýsingum úr veðurfræðilíkaninu. Hins vegar er sett fram stigskipt Gaussískt líkan sem nýtir áðurnefndar upplýsingar í formi skýribreyta og tekur tillit til breytileika úrkomu í rúmi. Breytileika í rúmi er lýst með tölfræðilegu rúmlíkani sem byggir á nálgunarlausn á slembni hlutafleiðujöfnu sem gefur af sér aðferðafræði fyrir hraðvirka

útreikninga. Áðurnefnt reiknirit er svo notað til að meta stika líkansins sem gefur samfellda rúmspá fyrir rúmháða stika í gagnaþéttleika hámarksúrkomu og nýtist við spá um stærðargráðu aftakaúrkomu í sérhverjum punkti þétta netsins yfir Íslandi. Auðsótt er að yfirfæra tölfræðilíkanið og aðferðafræðina á önnur svæði að því gefnu að hægt sé að nálgast upplýsingar úr veðurfræðilíkani af svæðinu.

Contents

List of Figures

List of Tables

List of Algorithms

Publications

This thesis is based on the following three papers. The first paper is a discussion paper to the paper 'Beyond mean regression' and has been published in *Statistical Modelling*. The second paper is a research paper and has been submitted to *Statistics and Computing*. The third paper is a research paper and has been published in *Environmetrics*.

Paper I

Martins T.G., Simpson D., Illian J.B., Rue H. and Geirsson Ó.P. 2013 Discussion of 'Beyond mean regression'. *Statistical Modelling* 13(4):355-361

Paper II

Geirsson Ó.P., Hrafnkelsson B., Simpson D. and Sigurðarson H. 2015 The MCMC split sampler: A block Gibbs sampling scheme for latent Gaussian models. *Submitted to Statistics and Computing*

Paper III

Geirsson Ó.P., Hrafnkelsson B. and Simpson D. 2015 Computationally efficient spatial modeling of annual maximum 24-h precipitation on a fine grid. *Environmetrics* 26(5):339-353, DOI: 10.1002/env.2343

Acknowledgements

It is both rewarding and astonishing to realize that this journey, which has governed almost one sixth of my life, is coming to an end. If I am to summarize my feelings in one word at this very moment it would be "gratitude". I am very grateful for the opportunity that I got to work on this research project and for the learning experience. I am, however, first and foremost ever grateful to all the people who have been with me along the way and made all of this possible. I would like to recognize them here.

First, I would like to express my most sincere gratitude to my principle advisor Assoc. Prof. Birgir Hrafnkelsson, who has guided me with his experience and knowledge through this journey. You have been both a fantastic advisor and good friend. I will be ever thankful for the trust you have placed in me and for giving me the opportunity to explore my interests and my goals. Furthermore, you have shown me patience that I hardly deserve and support through hard times that I am immensely grateful for. I would also like to extend my sincere gratitude to my doctoral committee, Prof. Gunnar Stefánsson and Prof. Haraldur Ólafsson.

Over the years, I have been lucky enough to do my research work among close friends. I will always be grateful for your support and friendship. I would like to give special thanks to Atli Norðmann Sigurðarson, Helgi Sigurðarson, Ólafur Birgir Davíðsson and Guðmundur Einarsson who are all top-notch friends and generally great human beings. I would also like to thank Sölvi Rögnvaldsson, Axel Örn Jansson, Gunnar Daníel Sæmundsson, and Darri Eyþórsson for their help and support, especially at the very end of the research project. I would like to thank Bjarki Þór, Anna Helga, Helga and Þóra for their friendship and help with various issues. Further thanks go to my friends, both at the office and out there in the real life, that have helped and inspired me over the years.

During my Ph.D. studies I have been fortunate enough to spend time at the Department of Mathematical Sciences at the Norwegian University of Science and Technology in Trondheim with an amazing group of people who I am lucky enough to count among my friends. I would like to especially thank Prof. Håvard Rue for his invitation and valuable conversations. A very special thanks goes to my dear friend Dr. Daniel Simpson. You have both been a great friend and co-worker for the past six years. I will be ever grateful for your valuable insights, your great support and your friendship. Further thanks go to Geir-Arne, Erlend Aune, Håvard Granlund, Siri-Malén, Finn Lindgren and Janine for your friendship and making me feel welcome in Trondheim.

I would like to thank the University of Iceland Doctoral Fund, the University of Iceland Research Fund, and Landsvirkjun Energy Research Fund that financially supported the research. I would also like to thank the Icelandic Meteorological Office, in particular, Dr. Philippe Crochet at IMO for providing the data and valuable discussions. Further thanks go to the Nordic Network on Statistical Approaches to Regional Climate Models for Adaptation (SARMA), especially Prof. Peter Guttorp, for providing travel support. I would also like to give special thanks to my opponents, Dr. Þórdís Linda Þórarinsdóttir and Dr. Alex Lenkoski, for their hard working reviewing my thesis.

Ég vil þakka fjölskyldunni minni fyrir allan stuðninginn sem þau hafa veitt mér í gegnum árin. Til að byrja með vil ég fá að þakka foreldrum mínum, Ann Mikkelsen og Geir Björnssyni, sem eru bestu foreldrar sem völ er á. Þau hafa ávallt staðið við bakið á mér í gegnum bæði súrt og sætt. Ég vil líka fá að þakka bróður mínum, Sigurgeiri Birni Geirssyni sem er einnig augljóslega besti bróðir í heimi. Einnig þakka ég fjölskyldunni hans, Karen Ósk Úlfarsdóttur, Úlfari Freyr og Snorra Páli. Stuðningur fjölskyldunnar minnar í gegnum árin er mér ómetanlegur og ég verð ævinlega þakklátur fyrir að eiga ykkur öll að. Mig langar svo að þakka tengdafjölskyldunni minni, Guðjóni Bachmann, Kristínu Önnu, Bjarna, Hjördísi og Jón Gústa fyrir að allan þann stuðning sem þau hafa öll hafa veitt mér.

Mig langar til að þakka sérstaklega frænku minni, Öddu Geirsdóttur, fyrir hennar hlýju, vinskap, stuðning og ótal skemmtilegar samræður. Einnig langar mig til að þakka afa mínu heitnum, honum Birni Bjarnasyni, fyrir að vekja áhuga minn á stærðfræði og vísindum á yngri árum. Afi hefur verið mín fyrirmynd ávallt síðan, ekki bara sem stærðfræðingur heldur líka sem góðhjartaður einstaklingur. Ég vil fá að þakka frænku minni Helgu Björnsdóttur fyrir hennar leiðsögn og hvatningu, sérstaklega þegar ég hóf mitt háskólanám í stærðfræði. Ég vil líka þakka ömmum mínum heitnum, þeim Erlu Geirsdóttur og Ídu Kristjánsdóttur, fyrir þeirra endalausu hlýju.

Síðast en ekki síst vil ég fá að þakka þér Bessý mín. Þú hefur verið mín stoð og stytta í gegnum allt þetta ferli. Þú hefur glaðst með mér og hjálpað mér í gegnum erfiðustu tímana. Þú hefur hlustað á mitt vonleysishjal og hvatt mig áfram. Ég vil þakka þér fyrir Hilmi Bjarna sem kom eins og ljósgeisli inn í líf okkar og gaf því nýjan tilgang. Ekkert gerir mig meira stoltan en að vera pabbi hans Hilmis. Ég veit ekki hvort ég hefði mögulega getað komist í gegnum þetta ferli án ykkar en ég veit að ég hefði ekki viljað gera það án ykkar. Ég er þér ævinlega þakklátur fyrir þína hlýju, styrk og þolinmæði. Mínar dýpstu þakkir fyrir þína umhyggju og ást.

Structure of the thesis

This thesis is divided into two mains parts. Part I, Thesis, serves as an overview of the entire research work carried out in the three papers, along with some additional discussion and motivation. Part I begins with an introduction to the research project, along with a brief overview of the development process of the statistical methodology proposed in this thesis. A background is provided of the statistical methods which form the basis of the research in this thesis. The aims and goals of the thesis are then formally defined and outlined. Part I further provides an overview of the proposed statistical methodology, along with the corresponding main results. Part I concludes with a discussion on future work and improvements. References made to the three papers in Part I are denoted in bold letters. Part II, Papers, consists of the three papers which this thesis is based on, with some minor editing.

Nomenclature

x	Scalar
$\boldsymbol{x}$	Column vector
$\boldsymbol{x}^\mathsf{T}$	Row vector
$\boldsymbol{x}_i$	The i-th subvector of $\boldsymbol{x}$
$\boldsymbol{x}_{-i}$	The subvector of $\boldsymbol{x}$ excluding the i-th subvector of $\boldsymbol{x}$
$\boldsymbol{A}$	Matrix
$\boldsymbol{A}^{-1}$	Matrix inverse
$\boldsymbol{A}^\mathsf{T}$	Matrix transpose
$\pi(x)$	Probability density function of the random variable X
$\pi(x\|y)$	Probability density function of X given Y
$\nabla f(\boldsymbol{x})$	The gradient of the function f evaluated at $\boldsymbol{x}$
$\nabla^2 f(\boldsymbol{x})$	The Hessian of the function f evaluated at $\boldsymbol{x}$
$f_1(x) \propto f_2(x)$	Denotes $f_1(x) = af_2(x)$, where a is a constant
$\operatorname{diag}(\boldsymbol{x})$	Diagonal matrix with the vector $\boldsymbol{x}$ as the diagonal
$\operatorname{bdiag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n)$	Block diagonal matrix of the matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$
$\circ$	Entrywise product (Hadamard product operator)
$\otimes$	Kronecker product operator

Part I

Thesis

1

Introduction

The original goal of the Ph.D. research project presented in this thesis was to develop a statistical method for describing the distributional properties of annual maximum 24-hour precipitation in Iceland. This distributional information is needed at a fine spatial resolution as the characteristics of extreme precipitation events can be local, especially in regions with heterogeneous topography. As a result, statistical spatial modeling for this type of spatial data becomes challenging, especially when observational sites are scarce relative to changes in the topography. The main objective was thus to develop an efficient spatial modeling framework for maximum precipitation on a fine grid.

Early in the development process, it was established that the Bayesian paradigm can provide the necessary modeling flexibility to accomplish this goal. However, it quickly became apparent that the desired Bayesian model structure introduced several challenging computational problems. Established Bayesian inferential algorithms were either too slow to converge to be practical or not applicable to the desired fine scale spatial model structure. The work presented in this thesis aims to provide both a general and a computationally efficient approach to Bayesian statistical modeling and inference, with the focus on applications to spatial extremes, in particular, annual maximum 24-hour precipitation in Iceland.

It is established in this thesis that the class of latent Gaussian models (LGMs) is well suited for the purposes of this research project. LGMs form a flexible subclass of Bayesian hierarchical models and are practical from a statistical modeling point of view and readily interpretable. Consequently, LGMs have become popular in many areas of statistics and various fields of applications. However, the availability of inferential algorithms for LGMs limits the potential scope and flexibility of LGMs in practice. One of the main novel contributions of this Ph.D. research project is to develop an inferential algorithm aimed to address this issue. To that end, a computationally efficient MCMC sampling scheme designed for LGMs is proposed in this thesis. The proposed sampling scheme is designed to be naturally applicable to LGMs where the data density of each data point is

allowed to depend on more than a single linear link function of the latent parameters. Furthermore, the sampling scheme is also designed to be applicable to LGMs with any choice of a parametric data density function. The author refers to the proposed sampling scheme as *the MCMC split sampler*.

By working within the the LGM framework with the MCMC split sampler as an inferential algorithm, the following novel contributions to statistical modeling of maximum precipitation in a spatial setting are presented in this thesis. First, a method is proposed which leverages information on the physical processes of precipitation from a local meteorological model to yield additional spatial information on maximum precipitation. This is achieved by constructing a spatial covariate from the outputs of the local meteorological model. Secondly, an LGM is proposed for observed annual maximum 24-hour precipitation which takes into account both the covariate based on the local meteorological model and spatial variation of the underlying physical processes of precipitation. The observations are assumed to follow the generalized extreme value distribution, where spatial models based on approximate solutions to stochastic partial differential equations are implemented for the location, scale, and shape parameters of the data density.

Part I of this thesis is organized as follows. Chapter 2 provides a background of the statistical methods which form the theoretical basis for the thesis. More specifically, in Section 2.1 a brief introduction to the Bayesian paradigm and Bayesian hierarchical models is given. Section 2.2 provides an overview of inferential algorithms for Bayesian statistical models. A summary of spatial statistics is given in Section 2.3, where spatial random fields and their applications in practice are explored. The chapter concludes with Section 2.4, which provides a brief overview of statistical models for extreme precipitation. In Chapter 3, the aims and goals of this thesis are formally defined and outlined.

Chapter 4 provides an overview of the proposed statistical methodology and novel contributions developed in this thesis and presented in the three papers. In Section 4.1 a summary of the discussion in **Paper I** is given, where the author's view on modern statistical modeling is elaborated on. Section 4.2 provides an overview of the novel inferential algorithm, the MCMC split sampler, which is presented in **Paper II**. The section discusses the derivation, design and the implementation of the MCMC spilt sampler to LGMs. In Section 4.3 a summary, motivation and discussion of the statistical modeling of annual maximum 24-hour precipitation in Iceland is given, which is presented in **Paper III**.

In Chapter 5, the main results from the research project are presented. Section 5.1 provides an evaluation of the computational efficiency of the MCMC split sampler. This is carried out by applying the MCMC split sampler to three different LGMs, one of which, is the proposed model for annual 24-hour maximum precipitations. Section 5.2 gives an overview of the main findings from the statistical modeling of annual maximum 24-hour precipitation in Iceland. Finally, Chapter 6 provides conclusions and reflections regarding future research stemming from the proposed methodology in this thesis.

2

Background

2.1 Bayesian statistical modeling

The Bayesian statistical methodology presents a well-established framework for making inference from observed data for quantities of interest by using an underlying probability model, see Gelman et al. (2014) and Berger (2013) for a comprehensive overview of modern Bayesian statistical analysis. The Bayesian methodology differs from the classical frequentist approach in that all of the unknown parameters in the underlying probability model are treated as random variables, as opposed to unknown constants[1]. As such, the unknown parameters are assigned *prior distributions* which are based on *a priori* subjective beliefs or scientific knowledge about the unknown parameters. In other words, prior distributions serve as probabilistic descriptions of what is known about the unknown parameters before observational data are collected and analyzed. After observing the data, knowledge about the model parameters and quantities of interest are updated by conditioning the underlying probability model on observed data by utilizing Bayes' theorem. This process, referred to as Bayesian inference, yields the *posterior distribution* of the model. The posterior distribution incorporates updated knowledge about the unknown parameters in light of new knowledge obtained from the observed data, as discussed in Berger (2013).

As presented in Gelman et al. (2014), the process of the Bayesian statistical methodology can be categorized into three main steps:

Bayesian statistical modeling: Setting up the underlying probability model, by constructing a joint probability distribution for both the observations and the model parameters. A preferable model, in most situation, is a model which incorporate *a priori* knowledge about the model parameters through prior distributions.

[1] The scope of this thesis is limited to Bayesian parametric models. Non-parametric Bayesian models are not considered.

Bayesian inference: Conditioning on the observed data. This step consists of calculating and interpreting the posterior distribution of the model parameters.

Model evaluation: Evaluating how well the model, which is induced by posterior distribution of the model parameters, describes the observed data.

The main concepts of the Bayesian methodology can be described in mathematical terms as follows. Let $\boldsymbol{y}$ denote a vector of observed data and let $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ denote the data density function conditioned on the model parameters $\boldsymbol{\theta}$. Additionally, let $\pi(\boldsymbol{\theta})$ denote the prior density function of the model parameters and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ denote the corresponding posterior density function of the model parameters conditioned on the data vector $\boldsymbol{y}$. Applying Bayes' theorem yields the following relation

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{\theta})}{\pi(\boldsymbol{y})} \tag{2.1}$$

where $\pi(\boldsymbol{y})$ denotes the marginal density function of the data $\boldsymbol{y}$, that is

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

As the marginal data density function $\pi(\boldsymbol{y})$ is independent of parameters $\boldsymbol{\theta}$, the expression in (2.1) is often summarized as

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{\theta}).$$

Assuming the vector of parameters $\boldsymbol{\theta}$ is multivariate the *marginal posterior density function* of a single parameter θ_i is obtained by integrating out other parameters $\boldsymbol{\theta}_{-i}$ from the joint posterior density function. That is,

$$\pi(\theta_i|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta}_{-i},$$

where the subscript $-i$ denotes all but the i-th entry of vector $\boldsymbol{\theta}$.

2.1.1 Bayesian hierarchical models

In many statistical applications, the observable data are believed, or known *a priori*, to have a certain latent dependence on underlying processes which drives the observed data. For example, in Hrafnkelsson et al. (2012), minimum and maximum temperature in Iceland is analyzed. The observed data, consisting of measurements of temperature, exhibit a latent dependence structure in the sense that the temperature is dependent on where and when it is measured. The location and time of measurements, along with other potential factors, thus drive the observed temperature.

The Bayesian paradigm presents a systematic modeling methodology to capture the latent dependence structure of observed data, referred to as Bayesian

hierarchical modeling (Gelman and Hill 2006). In this framework, a Bayesian model involving multiple parameters is proposed, where the model parameters are related or dependent in a systematic manner. The resulting joint probability model should thus reflect the dependence structure of the parameters. Further, it is considered natural to structure a model of this type hierarchically, with observable data modeled conditionally on a certain set of model parameters, which in turn can be potentially dependent on another set of model parameters. See for example Gelman and Hill (2006), Gelman et al. (2014) and Banerjee et al. (2014), for an overview of Bayesian hierarchical modeling.

The hierarchy of the Bayesian hierarchical modeling adhered to in this thesis, can be idealized in the following three levels.

Data level: A data density is chosen for the observations conditioned on the latent processes and other potential parameters.

Latent level: A probability model is constructed for the latent processes conditioned on other potential parameters. This is attained by selecting prior distribution which should ideally incorporate *a priori* knowledge on the latent processes.

Hyper level: Prior distributions for the parameters of the latent processes are chosen.

In particular, the above modeling hierarchy is designed to combine observed data sources with known scientific knowledge on the latent processes. This is generally achieved by constructing prior distributions at the latent level, which are in line with prior knowledge on the corresponding latent processes. Thus, due to its flexibility, Bayesian hierarchical modeling has become a popular statistical tool for modeling physical phenomena, as well as various other fields of application. See for example, Guttorp and Gneiting (2006), Cooley et al. (2007), Tebaldi and Sansó (2009), Davison et al. (2012), Fuglstad et al. (2013), and Dyrrdal et al. (2015) where Bayesian hierarchical modeling is used to model precipitation and its underlying physical latent processes, briefly discussed in Section 2.4.

The mathematical formulation of Bayesian hierarchical models adhered to in this thesis, can be summarized as follows. Let $\boldsymbol{y}$ denote the data vector and $\boldsymbol{x}$ denote the latent process, refereed to as the *latent parameters*. Additionally, denote the parameters of the probability model for the latent process, referred to as *hyperparameters*, with $\boldsymbol{\theta}$. The hierarchy can then be categorized a follows

Data level: A data density, $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$, is chosen for the data vector $\boldsymbol{y}$ conditioned on the latent parameters $\boldsymbol{x}$ and the hyperparmeters $\boldsymbol{\theta}$,

Latent level: A prior density function, $\pi(\boldsymbol{x}|\boldsymbol{\theta})$, is selected for the latent parameters $\boldsymbol{x}$ given the hyperparameters $\boldsymbol{\theta}$.

Hyper level: A prior density function, $\pi(\boldsymbol{\theta})$, is chosen for the hyperparameters $\boldsymbol{\theta}$.

In order to infer the above Bayesian hierarchical model, the posterior density function $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$ needs to be evaluated. Repeated use of Bayes' theorem yields the following proportional relation

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}, \boldsymbol{\theta}) \propto \pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{2.2}$$

The relation in (2.2) can be utilized to evaluate or approximate the posterior density function $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$.

2.1.2 Latent Gaussian models

An important subclass of Bayesian hierarchical models is the class of latent Gaussian models (LGMs), in which *Gaussian* priors distributions are imposed on the latent parameters $\boldsymbol{x}$ the latent parameters $\boldsymbol{x}$ of the hierarchical model. The Gaussian distribution is arguably the most recognized and widely used probability distribution in statistics, see for instance Johnson et al. (1992) and Tong (2012) for an extensive overview. The multivariate Gaussian distribution is defined as follows.

Definition 1. *An n-dimensional random vector $\boldsymbol{x}$ is said to follow a Gaussian distribution with a mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and a positive definite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$, denoted $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its density function is of the form*

$$\pi(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right), \tag{2.3}$$

where $\mu_i = \mathbb{E}[x_i]$, $\Sigma_{ij} = Cov(x_i, x_j)$ and $\Sigma_{ii} = Var(x_i)$. A density of the form (2.3) is referred to as a Gaussian density and denoted with $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

LGMs have become popular in many areas of statistics and various fields of applications. For example, LGMs are widely used in modern spatial statistics, see for example Cressie (1993), Diggle et al. (1998), Rue and Held (2005), Guttorp and Gneiting (2006) and Chiles and Delfiner (2009). The main reason for the popularity of LGMs is that they are practical from a statistical modeling point of view, readily interpretable, and in most cases it is possible to evaluate the corresponding posterior density. Additionally, LGMs can be viewed as a specific extension of structured additive regression models (Fahrmeir et al. 1994), which is further discussed in Rue et al. (2009). That is, LGMs extend structured additive regression models in the sense that the data density function of each data point can depend on more than a single linear link function of the latent parameters through more than just the mean structure of the data density. This viewpoint is further discussed in **Paper I** and **Paper II**.

A generalized mathematical formulation of LGMs, which is adhered to in this thesis, can be summarized as follows.

Data-level: The observations $\boldsymbol{y}$ depend on the latent field $\boldsymbol{x}$, through some choice of data distribution with a data density function $\pi(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta})$.

Latent level: The prior for the latent field $\boldsymbol{x}$ is Gaussian and is potentially dependent on hyperparameters $\boldsymbol{\theta}$, with a density function

$$\pi(\boldsymbol{x} \mid \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{x} \mid \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}(\boldsymbol{\theta})^{-1}\right).$$

Hyperparameter level: A prior distribution is assigned for the hyperparameters $\boldsymbol{\theta}$, with a density function $\pi(\boldsymbol{\theta})$.

2.2 Computations for Bayesian inference

Once a Bayesian model has been proposed for a certain statistical application, the next step is to assess how the corresponding posterior distribution is to be evaluated. In a few special cases the posterior density function can be evaluated analytically. However, for involved Bayesian model structures, such as in the case of most Bayesian hierarchical models, the full joint posterior in only known up to a normalizing constant. In these cases, the posterior density function needs to be approximated numerically. Most of the commonly used numerical methods use stochastic simulations, via Markov chain Monte Carlo (MCMC) simulations, to sample from the posterior distribution. The samples from the stochastic simulations yield a numerical approximation of the posterior density function, which in turn can be used for Bayesian inference purposes.

2.2.1 Fundamentals of Markov chain Monte Carlo simulations

Markov chain Monte Carlo (MCMC) methods form the backbone of modern numerical methods for Bayesian posterior inference, as they are essentially applicable to almost any Bayesian model. See for example Gelman and Rubin (1992), Smith and Roberts (1993), Gilks (2005) and Robert and Casella (2013) for a detailed overview of the subject. MCMC methods consist of general algorithms for simulating Markov chains which have a desired target density as a stationary distribution[2]. In the Bayesian paradigm, the posterior density of a proposed Bayesian model is the desired target density. After a certain *burn in* period, that is, after a certain number of iterations in the Markov chain, the sequence of simulations from the Markov chain serve as samples from the posterior distribution of the proposed model. This procedure is often carried out by the use of the Gibbs sampling algorithm (Geman and Geman 1984, Casella and George 1992) and the Metropolis–Hastings algorithm (Metropolis et al. 1953, Hastings 1970), which are both outlined in the subsequent sections.

2.2.2 The Gibbs sampler

The Gibbs sampler is an iterative MCMC sampling algorithm designed to obtain samples from a joint distribution of multidimensional random variable. The Gibbs sampler is, therefore, well suited for obtaining samples from the posterior

[2]The Markov chain's stationary distribution exist under the assumptions of aperiodicity and irreducibility.

distribution of a given multi-parameter Bayesian model. The mathematical formulation of the Gibbs sampler, as designed for Bayesian inference, is as follows.

Let $\boldsymbol{\theta}$ denote a vector of model parameters of a given multi-parameter Bayesian model and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ denote the corresponding joint posterior density function. Further assume that the vector $\boldsymbol{\theta}$ can be decomposed into J subvectors, that is,

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\mathsf{T}, \ldots, \boldsymbol{\theta}_J^\mathsf{T})^\mathsf{T}.$$

In every iteration, the Gibbs sampler cycles through the subvectors of $\boldsymbol{\theta}$ and draws samples from the conditional posterior distribution of the subvectors of $\boldsymbol{\theta}$ conditioned on latest values of other subvectors of $\boldsymbol{\theta}$. This procedure generates a Markov chain consisting of samples of $\boldsymbol{\theta}$ obtained in every iteration, that can be shown to converge to the target distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, see for example Gilks (2005).

The algorithm for the Gibbs sampler, based on K iterations and with $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ as a target density, is outlined in Algorithm 1.

Algorithm 1 The Gibbs sampler, based on K iterations with a target density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$.

1: Initialize the Markov chain, by choosing an arbitrary starting value $\boldsymbol{\theta}^0$ such that $\pi(\boldsymbol{\theta}^0|\boldsymbol{y}) > 0$.
2: **for** $k = 1, \ldots, K-1$
3: **for** $j = 1, \ldots, J$
4: Sample $\boldsymbol{\theta}_j^{(k+1)}$ from $\pi(\boldsymbol{\theta}_j \mid \boldsymbol{y}, \boldsymbol{\theta}_1^{k+1}, \ldots, \boldsymbol{\theta}_{j-1}^{k+1}, \boldsymbol{\theta}_{j+1}^{k}, \ldots, \boldsymbol{\theta}_J^{k})$
5: **end for**
6: **end for**

2.2.3 Metropolis–Hastings Algorithm

The Gibbs sampler, in the form presented in Section 2.2.2, is only applicable to a multivariate Bayesian model if the conditional posterior distributions $\pi(\boldsymbol{\theta}_j|\boldsymbol{\theta}_{-j}, \boldsymbol{y})$ are known for all subvectors $\boldsymbol{\theta}_j$. However, in many statistical applications the conditional posterior distributions $\pi(\boldsymbol{\theta}_j|\boldsymbol{y}, \boldsymbol{\theta}_{-j})$ are only known up to a normalizing constant. The Metropolis–Hastings algorithm, which is arguably one of the most successful and widely used MCMC algorithms (Beichl and Sullivan 2000), is a class of algorithms which can be considered as an extension to the Gibbs sampler, which serves to address this issue. That is, the Metropolis–Hastings algorithm is applicable for drawing samples from a desired target density if the target density is known up to a normalizing constant. As such, the algorithm can be implemented within the Gibbs sampling scheme when the conditional posterior distributions $\pi(\boldsymbol{\theta}_j|\boldsymbol{\theta}_{-j}, \boldsymbol{y})$ are only known up to a normalizing constant.

The Metropolis–Hastings algorithm generates a Markov chain that converges to the desired target density, which is the posterior density in the Bayesian setting, under mild regularity conditions (Chib and Greenberg 1995). For demonstrative

purposes assume that $\boldsymbol{\theta}$ is a random vector with a target density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. The Metropolis–Hastings algorithm for obtaining K samples from the target density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is summarized in Algorithm 2.

Algorithm 2 The Metropolis–Hastings algorithm, based on K number of iterations with a target density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$.

1: Initialize the Markov chain, by choosing an arbitrary starting value $\boldsymbol{\theta}^0$ such that $\pi(\boldsymbol{\theta}^0|\boldsymbol{y}) > 0$.
2: **for** $k = 0, \ldots, K-1$
3: Sample a proposal value $\boldsymbol{\theta}^*$ from a given proposal density $q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)$
4: Calculate

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{y})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{y})} \frac{q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)}\right\} \tag{2.4}$$

5: Sample $u \sim \mathcal{U}(0,1)$
6: **if** $\alpha > u$
7: $\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^*$
8: **else**
9: $\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k$
10: **end if**
11: **end for**

The probability in (2.4) is often referred to as acceptance probability. Additionally, the ratio in (2.4) is referred to as the acceptance ratio in this thesis.

One of the advantages presented by the Metropolis–Hastings algorithm is that the algorithm converges to the target density for almost any choice of a proposal density function[3], see Rue and Held (2005) for discussion on this point. Furthermore, different choice of proposal densities result in different classes of Metropolis–Hastings algorithms. For example, consider the special case when the target density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is known. By selecting the target density $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ as proposal density in Metropolis–Hastings algorithm, that is $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^k) = \pi(\boldsymbol{\theta}|\boldsymbol{y})$, then the acceptance probability in (2.4) becomes equal to one. This version of the Metropolis–Hastings algorithm is the Gibbs sampler.

Common classes of Metropolis–Hastings algorithms, discussed in this thesis, are based on the following types of proposal densities.

Symmetric random walk proposal: If the proposal density is symmetric, such that, the following relation holds

$$q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*) = q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k),$$

then the acceptance probability in (2.4) simplifies to

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{y})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{y})}\right\}.$$

[3]Under mild regularity conditions. One which is that the support of the proposal density function must contain the support of the target density function.

The most common symmetric random walk proposal density, is a Gaussian density centered at the current value of $\boldsymbol{\theta}$ with a predetermined variance or covariance matrix.

Independence proposal: A proposal density referred to as an independence proposal density if it is independent of the current value of $\boldsymbol{\theta}$ in MCMC chain, that is, if $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^k) = \boldsymbol{q}(\boldsymbol{\theta}^*)$. In this case, the acceptance probability in (2.4) simplifies to

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{y})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{y})} \frac{q(\boldsymbol{\theta}^k)}{q(\boldsymbol{\theta}^*)}\right\}.$$

2.2.4 Assessing computational efficiency

Although Metropolis–Hastings algorithms are guaranteed to converge to the target density, the rate of convergence and the degree of dependence between successive samples can vary significantly between particular classes of Metropolis–Hastings algorithms and how they are implemented. In general, fast convergence and low dependence between successive samples yield higher quality of MCMC chains. The quality of a simulated MCMC chain can be assessed through its effective sample size, which is an estimate of the equivalent number of independent iterations in the simulated MCMC chain, see Gelman et al. (2014). In this thesis, MCMC algorithms which yield high quality MCMC chains are referred to as being computationally efficient, or just efficient for short[4].

The convergence diagnostics tools which are normally used for assessing computational efficiency of MCMC chains are the following.

Trace plots: A traceplot is a plot of an MCMC simulation, in which the value of the MCMC chain is plotted as a function of iterations. By visual inspection of the trace plots, it is possible to identify if and where the MCMC chain gets stuck in the same value for many consecutive iterations. If the MCMC chain does get stuck, that indicates low computational efficiency.

Running mean plots: A running mean plot is a plot, in which the mean values of the draws in the MCMC chain up to the current iterations are plotted as a functions of iterations. The running plots serve as a visual inspection tool for assessing the rate convergence in the mean of the MCMC chain.

Autocorrelation: The dependence between successive samples of the Markov chain is evaluated with its autocorrelation which is estimated with the sample correlation. The j-th lag autocorrelation ρ_j is defined as the correlation between every j successive draws. The j-th lag autocorrelation of a MCMC chain $\{\theta_k\}_{k=1}^K$ can be estimated with

$$\hat{\rho}_j = \frac{\sum_{k=1}^{K-j} \left(\theta_k - \bar{\theta}\right)\left(\theta_{k+j} - \bar{\theta}\right)}{\sum_{k=1}^{K} \left(\theta_k - \bar{\theta}\right)^2},$$

[4]Also referred to as having good mixing and convergence properties by many authors.

where $\bar{\theta} = K^{-1} \sum_{k=1}^{K} \theta_k$. How the j-th lag autocorrelation decreases as function of lag k yields insight into the computational efficiency of the MCMC sampler. That is, the j-th lag autocorrelation decreases rapidly if the MCMC algorithm is computationally efficient. However, high j-th lag autocorrelation for relatively high values of j indicates poor computational efficiency.

Autocorrelation plots, which are plots showing the j-th lag autocorrelation as a function of lag j, are useful visual diagnostics tools for assessing the behavior of the autocorrelation.

Gelman–Rubin statistic: The Gelman–Rubin statistic[5], proposed by Gelman and Rubin (1992) is a metric for assessing convergence of iterative MCMC simulations. The Gelman–Rubin statistic is evaluated from the m simulated MCMC chains, which have different initial values and have been simulated independently of each other. The algorithm for calculating the Gelman–Rubin statistic is thoroughly outlined in Brooks and Gelman (1998). The Gelman–Rubin statistic can be interpreted as follows. A Gelman–Rubin statistic close to 1 suggests that the MCMC simulations are close to the target distribution. In most practical cases, values below 1.05 are acceptable. However large values of the Gelman–Rubin statistic, typically greater than 1.1, indicate that the simulations have not converged to the target density.

Gelman–Rubin plots are plots where the Gelman–Rubin statistic is plotted as a function of iteration. These plots can be used as a visitation tool for assessing the rate of convergence of the given MCMC chain.

2.2.5 Proposal densities and efficiency

The computational efficiency of a Metropolis–Hasting algorithm is highly influenced by the choice of a proposal density. Utilizing a proper density, such that the MCMC chain converges in theoretical terms, but a poorly designed one may lead to an MCMC chain that converges too slowly to be considered practical. Thus, designing an efficient proposal density is vital when implementing the Metropolis–Hasting algorithm in practice. In this section a few common guidelines and known results are given for designing an efficient proposal density.

The computational efficiency of the Metropolis–Hasting algorithm can often be controlled through the acceptance probability α, or the expected acceptance rate $\mathbb{E}[\alpha]$ of the MCMC chain in equilibrium. Depending on the class of Metropolis–Hastings algorithms, different behavior of the acceptance rate $\mathbb{E}[\alpha]$ is desired for optimal computational efficiency. This in turn, leads to guidelines for tuning the proposal densities for optimal efficiency.

Random walk proposals: If the acceptance rate $\mathbb{E}[\alpha]$ is too high, the proposed values tend to be close to the current position of the chain. In other words, the

[5] Also referred to as potential scale reduction factor.

main bulk of the proposal's probability density is too narrow to properly explore the space of the posterior density. Therefore, the Markov chain will take too small steps in every iteration. This results in an MCMC chain that exhibits a high autocorrelation, and thus reduces computational efficiency. However, when the acceptance rate is too low, the proposed draws take large steps from the current positions in the chain but are frequently rejected by the Metropolis–Hastings algorithm. This results in the chain getting stuck in the same state for many iterations, which reduces computational efficiency. Roberts et al. (1997) show that the optimal acceptance rate for random walk proposals is 44% when the θ is a scalar and 23% when θ is a multidimensional.

In practice random walk proposal densities are tuned to achieve the desired acceptance rate for computational efficiency. An example of a random walk proposal for a target density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, which is tuned for computational efficiency, is given in Roberts et al. (1997). That is, a proposal density based on the Gaussian distribution centered on the last draw of $\boldsymbol{\theta}$ with a precision matrix $-c\boldsymbol{H}$ where $\boldsymbol{H}$ is the of the Hessian matrix of $\log \pi(\boldsymbol{\theta}|\boldsymbol{y})$ evaluated at the mode $\boldsymbol{\theta}_0$. That is,

$$\boldsymbol{H} = \nabla^2 \log \pi(\boldsymbol{\theta}|\boldsymbol{y})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{2.5}$$

Further, the scaling parameters c serves as a tuning parameter for the proposal density. However, the scaling $c = 2.382/\dim(\boldsymbol{\theta})$ can be shown to yield optimal acceptance rates in a particular large dimension scenario, see Roberts et al. (1997). The resulting proposal density therefore becomes

$$q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^k) = \mathcal{N}\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k, (-c\boldsymbol{H})^{-1}\right).$$

Independence proposals: Consider again the special case when the target density is known and chosen as a proposal density. In the role as a proposal density, the target density is an independence proposal, as it is independent of the current state in the chain, and has the acceptance rate 1. As the target density is trivially the optimal proposal density the following can be stated. In the case of independence proposal when the target density is unknown, choosing a proposal density which approximates the target density well yields higher computational efficiency. Further, high values of expected acceptance rate $\mathbb{E}[\alpha]$ indicates that the proposal density approximates the target density well. Therefore, higher value of $\mathbb{E}[\alpha]$ are desired for independence proposals.

An example of an independence proposal, is a proposal designed as a Gaussian approximation of the target density evaluated at its mode. That is, assume that $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ is the target density with the mode θ_0. The resulting proposal density becomes

$$q(\boldsymbol{\theta}^*) = \mathcal{N}\left(\boldsymbol{\eta} \mid \boldsymbol{\theta}^0, (-\boldsymbol{H})^{-1}\right) \tag{2.6}$$

where $\boldsymbol{H}$ is the Hessian of the logarithm of conditional posterior evaluated at the mode, given in equation (2.5).

Mengersen et al. (1996) show that, in some cases, using independence proposals may result in the MCMC chain getting stuck in regions of low target density.

Further, Mengersen et al. (1996) give a criterion for independence Metropolis–Hasting algorithms such that they do not exhibit this behavior, that is, they show that the resulting MCMC chains is geometrically convergent as a function of iterations if and only if the proposal density is bounded below by a multiple of the target density. Their analytic results can be interpreted as follows. If the tail of the proposal density is "thick enough" with respect to target density, the resulting algorithm does not stick and exhibits fast convergence.

The Metropolis Adjusted Langevin algorithm: A proposal density can be derived from a discretized Langevin diffusion, see Roberts and Rosenthal (1998) referred to as the Metropolis Adjusted Langevin algorithm (MALA). The MALA algorithm aims to provide a high computational efficiency, such that both large proposal steps are proposed that are also accepted with high probability. To that extent, the MALA algorithm uses the gradient of the target density to drive the proposal density toward regions of higher posterior density. That is, assume that the target density, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is known up to a constant, then the MALA proposal density is given by

$$q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^k) = \mathcal{N}\left(\boldsymbol{\theta}^* \mid \boldsymbol{\mu}(\boldsymbol{\theta}^k, \varepsilon), \varepsilon^2 \boldsymbol{I}\right)$$

where

$$\boldsymbol{\mu}(\boldsymbol{\theta}^k, \varepsilon) = \nabla \log \pi(\boldsymbol{\theta}|\boldsymbol{y})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^k}$$

and ε is a tuning parameter. As the proposal density is not symmetric, the acceptance probability for the MALA class Metropolis–Hastings algorithm is of the same form as in equation (2.4). In order to circumvent computationally costly tuning of the MALA algorithm, Girolami and Calderhead (2011) suggested an automated adaptation mechanisms for ε for optimal computational efficiency. Their methods exploit the Riemannian geometry, see Boothby (2003), of the parameter space of statistical models which provide highly efficient convergence and exploration of the target density.

2.2.6 Block sampling

A popular inference scheme for sampling from a multidimensional target density, which is the posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ for Bayesian hierarchical models, is the *single-site* updating scheme. The single-site updating scheme is a Gibbs-sampler where every one-dimensional element θ_j of the vector $\boldsymbol{\theta}$ is updated, using a Metropolis–Hastings algorithm if necessary, conditioned on $\boldsymbol{\theta}_{-j}$. However, in the case of most Bayesian hierarchical models, parameters can become highly dependent in the posterior, see for example Murray and Adams (2010). In this case, Knorr-Held and Rue (2002) demonstrate that the resulting MCMC chain transits slowly around the target density space, which in turn leads to poor computational efficiency. Further, Knorr-Held and Rue (2002) suggest a general approach to circumvent this issue by updating subvectors of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_j$, which are highly dependent in the target density together in one *block*. Ideally, the vector $\boldsymbol{\theta}$ should be partitioned into blocks that have high dependence within the blocks but low dependence between them. This general scheme is referred to as *block*

sampling scheme in this thesis. In Rue and Held (2005) a thorough analyzes of block sampling is given, along with various blocking strategies for various types of Bayesian models.

Of particular interest in this thesis are sampling schemes for LGMs. Filippone et al. (2013) conducted a detailed comparison of modern sampling schemes[6] for LGMs and found that the *one-block* updating scheme of Knorr-Held and Rue (2002) performed best, in terms of efficiency, when applicable. The one-block sampling strategy is as follows.

Assume the LGMs setup in Section 2.1.2. The posterior density, $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$, of the model is proportional to

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

Additionally, assume that the form of the conditional posterior $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ is a known Gaussian density. The one-block updating scheme is a Metrpolis–Hastings algorithm with the following proposal density

$$\begin{aligned} \boldsymbol{\theta}^* &\sim q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k) \\ \boldsymbol{x}^* &\sim \pi(\boldsymbol{x}^* \mid \boldsymbol{y}^{k+1}, \boldsymbol{\theta}^*). \end{aligned} \tag{2.7}$$

Denote the proposal density implied by (2.7) with $q(\boldsymbol{x}^*, \boldsymbol{\theta}^* \mid \boldsymbol{x}^k, \boldsymbol{\theta}^k)$. A proposed value $(\boldsymbol{x}^*, \boldsymbol{\theta}^*)$ in the $(k+1)$-th iteration is then accepted jointly with acceptance probability

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{x}^*, \boldsymbol{\theta}^* \mid \boldsymbol{y},)}{\pi(\boldsymbol{x}^k, \boldsymbol{\theta}^k \mid \boldsymbol{y})} \frac{q(\boldsymbol{x}^k, \boldsymbol{\theta}^k \mid \boldsymbol{x}^*, \boldsymbol{\theta}^*)}{q(\boldsymbol{x}^*, \boldsymbol{\theta}^* \mid \boldsymbol{x}^k, \boldsymbol{\theta}^k)}\right\}. \tag{2.8}$$

Further, it can be shown that the acceptance probability in (2.8) can be written as

$$\alpha = \left\{1, \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{y})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{y})} \frac{q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)}\right\} \tag{2.9}$$

The acceptance ratio in (2.9) shows that the expected acceptance rate for the one-block sampler is only dependent on $\boldsymbol{\theta}$. Further, the result in (2.9) indicate that by sampling $\boldsymbol{x}$ from the known Gaussian density $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$, the vector $\boldsymbol{x}$ has been essentially integrated out from the joint posterior $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$. The computational efficiency of the one-block sampler is thus only dependent of the choice of proposal density $q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)$ for $\boldsymbol{\theta}$. As $\boldsymbol{\theta}$ is normally of low dimension, implementing an efficient proposal density for $\boldsymbol{\theta}$ is often readily achievable. For example, the proposal densities from Section 2.2.5 are applicable for $\boldsymbol{\theta}$ in this setup.

However, if $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ is not know, which is often case for LGM with a non-Gaussian data-density, the one-block scheme is not applicable as presented above. In **Paper II** a more general sampling scheme for LGMs, referred to a the MCMC split sampler, is presented which can be viewed as an extension to the one-block sampler.

[6] Including, the Riemannian geometry manifold MALA based Metropolis–Hastings algorithm.

2.2.7 Approximate inference

Alternative to MCMC methods are deterministic approximate posterior inference methods, such as the Integrated nested Laplace approximation (INLA) proposed by Rue et al. (2009). INLA is a fast approximate inference method for LGMs in which the data density of each data point only depends on a single linear functional of the latent field. INLA makes use of deterministic nested Laplace approximations which provides a fast and accurate alternative to MCMC methods[7].

2.3 Spatial statistics

Physical modeling of spatial phenomena without stochastic components may not adequately capture the complex processes of nature. However, spatial physical phenomena which are located close together is space tend to exhibit a high degree of dependency, while physical phenomena located far apart tend to show low levels of dependency. On that point, Cressie and Wikle (2011) comment "*[...] there is a flux of causal relationships in the space-time continuum that, when integrated out over time or captured in a micro-instant of time, shows neighboring values to be more highly correlated*". The dependence between neighboring locations can be utilized to *borrow information* for statistical modeling of spatial data.

Spatial statistics is a branch of statistics which involves the quantitative analysis of spatially referenced data. In other words, spatial statistics can be described as the statistical modeling of spatial variability and uncertainty. In recent decades, great advances have been made in the field of spatial statistics, both in terms of flexible statistical modeling and computational efficiency, see for example Besag (1974), Grondona and Cressie (1991), Rue and Held (2005), Cressie and Wikle (2011), Lindgren et al. (2011) and Banerjee et al. (2014). In this section, a brief overview of spatial statistics and computationally efficient spatial modeling is given.

2.3.1 Spatial random fields

To motivate the fundamental mathematical concepts of spatial statistics, consider a Euclidean spatial domain of interest and denote it with $\mathcal{D}$. The spatial domain $\mathcal{D}$ can be, for instance, a flat surface such that $\mathcal{D} \subset \mathbb{R}^2$. Although the theory of spatial statistics extends to non-Euclidean spatial domains, such as a spheroid like the Earth, only Euclidean spatial domains, such that $\mathcal{D} \subset \mathbb{R}^d$, are considered in this thesis. Serving as an analogy, a spatial field can be viewed as a temporal snapshot of a physical phenomena over the domain $\mathcal{D}$, aggregated over time or a frozen temporal space-time process (Cressie and Wikle 2011).

To illustrate further, assume that a potential datum $u(\boldsymbol{s})$ for any given location $\boldsymbol{s} \in \mathcal{D}$ is a random variable. A *spatial random field* is generated by varying

[7] Other deterministic methods include for example the expectation–maximization (EM) algorithm, which is an iterative method to finding the mode of a posterior density, see McLachlan and Krishnan (2007) for an overivew.

the spatial index over the spatial domain $\boldsymbol{s} \in \mathcal{D}$ yielding

$$\{u(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}, \tag{2.10}$$

which is denoted with $\{u(\boldsymbol{s})\}$ for short, see Cressie (1993) for a more rigorous definition[8]. In this thesis, the spatial index is assumed continuous, in the sense that the spatial index $\boldsymbol{s}$ in (2.10) is assumed to vary continuously over the spatial domain $\mathcal{D}$. Continuing the aforementioned analogy, any realization of the random field can be viewed as a snapshot of a spatially referenced physical phenomenon.

For statistical modeling of spatial fields, a set of n locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$ is selected. Let $\boldsymbol{u}$ denote a spatially referenced random vector based on the spatial field in (2.10) at those locations, that is, $\boldsymbol{u} = (u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_n))$. The i-th element of $\boldsymbol{u}$ corresponds to the random variable at the location $\boldsymbol{s}_i$, for all $i = 1, \ldots, n$. Observing the values of $\boldsymbol{u}$ at every location $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$ yields a spatially referenced observational data, which in turn can be used for statistical analysis. Furthermore, the dependence structure of the random vector $\boldsymbol{u}$ can reflect on the dependence of the spatial field $\{u(\boldsymbol{s})\}$ between locations in $\mathcal{D}$. The dependence structure is often described with covariance functions, which are defined as follows.

Definition 2. *Let $\{u(\boldsymbol{s})\}$ be a given random field over a domain $\mathcal{D}$. The covariance function of $\{u(\boldsymbol{s})\}$ is a function $C : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ which satisfies*

$$C(\boldsymbol{s}, \boldsymbol{t}) = Cov(u(\boldsymbol{s}), u(\boldsymbol{t})), \quad \text{for all } \boldsymbol{s}, \boldsymbol{t} \in \mathcal{D}.$$

Furthermore, for a given set of locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$ let $\boldsymbol{u} = (u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_n))$. The covariance matrix of $\boldsymbol{u}$ is then given as

$$\Sigma_{ij} = C(\boldsymbol{s}_i, \boldsymbol{s}_j) \quad \text{for all } i, j = 1, \ldots, n.$$

A widely used subclass in the statistical analysis of spatial data are stationary spatial fields. The reason for this is that the dependence structure of stationary spatial fields is only dependent on the difference between locations in the domain. In other words, the dependence structure of stationary spatial fields is invariant to translations in the spatial domain. Therefore, the parameterization of stationary fields is more accessible than in the case of non-stationary spatial fields. The definition of a stationary spatial field is as follows.

Definition 3. *Let $\{u(\boldsymbol{s})\}$ be a given random field over a domain $\mathcal{D}$. The spatial field $\{u(\boldsymbol{s})\}$ is said to be* strongly stationary *if for every set of n points $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$ the following holds*

$$(u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_n)) \overset{D}{=} (u(\boldsymbol{s}_1 + \boldsymbol{h}), \ldots, u(\boldsymbol{s}_n + \boldsymbol{h})) \quad \text{for any } \boldsymbol{h} \text{ in } \mathbb{R}^d,$$

where $\overset{D}{=}$ denotes equality in distribution.

[8] That is, equation (2.10) is essentially $\{u(\boldsymbol{s}, \omega) : \boldsymbol{s} \in \mathcal{D}, \omega \in \Omega\}$ where $\{\Omega, \mathcal{F}, P\}$ is a probability space.

The spatial field $\{u(\boldsymbol{s})\}$ *is said to be* weakly stationary *if it has a constant mean function, that is*

$$E\left[Y(\boldsymbol{s})\right] = \mu \quad \textit{for all } \boldsymbol{s} \in \mathcal{D},$$

and its covariance functions depends on the distance between two locations $\boldsymbol{s}$ *and* $\boldsymbol{t} = \boldsymbol{s} + \boldsymbol{h}$ *such that,*

$$C(\boldsymbol{s}, \boldsymbol{s} + \boldsymbol{h}) = C(\boldsymbol{h}) \quad \textit{for all } \boldsymbol{s} \in \mathcal{D}, \tag{2.11}$$

A covariance function which satisfies the condition in (2.11) is said to be stationary.

It can be shown that strong stationary implies weak stationary, see for example Grimmett and Stirzaker (2001). For the rest of this thesis, weakly stationary spatial fields are referred to as stationary spatial fields. A stationary covariance function $C(\cdot, \cdot)$ that only depends on the distance between two locations $\boldsymbol{s}, \boldsymbol{t} \in \mathcal{D}$, that is

$$C(\boldsymbol{s}, \boldsymbol{t}) = C(||\boldsymbol{s} - \boldsymbol{t}||) \quad \text{for all } \boldsymbol{s}, \boldsymbol{t} \in \mathcal{D} \tag{2.12}$$

where $||\cdot||$ denotes some a Euclidean distance, is referred to as an *isotropic* covariance function. A simple example of an isotropic covariance function for describing dependence of spatial data, is the *exponential* covariance function, which is defined as follows.

Definition 4. *An isotropic covariance function of a spatial field* $\{u(\boldsymbol{s})\}$ *over a domain* $\mathcal{D}$ *is called an exponential covariance function if*

$$C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \sigma^2 \exp\left(-\phi||\boldsymbol{s}_i - \boldsymbol{s}_j||\right), \quad \textit{for all } \boldsymbol{s}, \boldsymbol{t} \in \mathcal{D},$$

where σ^2 *is a marginal variance,* $\phi > 0$ *is a decay parameter.*

The marginal variance parameter σ^2 is the variance of the spatial field for a given location, and the decay parameter ϕ describes how rapidly the spatial correlation decays with increasing distance. Although the exponential covariance function is readily interpretable, it lacks the means to adapt to the *smoothness* of the spatial field.

The Matérn covariance function, see for instance Cressie and Huang (1999) and Stein (1999), is a class of covariance functions which can be viewed as an extension to the exponential covariance function. That is, Matérn covariance function can describe the decay rate of the spatial correlation and the smoothness of the spatial field. Due to this flexibility, the Matérn covariance function has become popular in recent years for modeling physical processes in a spatial setting, see Guttorp and Gneiting (2006). The Matérn covariance function is defined as follows.

Definition 5. *An isotropic covariance function of a spatial field* $\{u(\boldsymbol{s})\}$ *over a domain* $\mathcal{D}$ *is referred to as the Matérn covariance function if*

$$C(\boldsymbol{s},\boldsymbol{t}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\kappa\|\boldsymbol{s}-\boldsymbol{t}\|\right)^{\nu} K_{\nu}\left(\kappa\|\boldsymbol{s}-\boldsymbol{t}\|\right) \tag{2.13}$$

where Γ is the gamma function; K_ν is the modified Bessel function of the second kind and order $\nu > 0$ (Abramowitz and Stegun 1964); $\kappa > 0$ is a scaling parameter; and σ^2 denotes the marginal variance of the Matérn field.

The parameters κ and ν are referred to as the scaling and smoothness parameters, respectively, of the spatial field. For interpretation, the Matérn covariance function is often parameterized with

$$C(\boldsymbol{s},\boldsymbol{t}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\|\boldsymbol{s}-\boldsymbol{t}\|}{\phi}\right)^{\nu} K_{\nu}\left(\frac{\|\boldsymbol{s}-\boldsymbol{t}\|}{\phi}\right)$$

where ϕ acts as a decay rate of the spatial correlation, see Diggle et al. (1998). The scaling parameter κ is related to the decay rate of the spatial correlation ϕ, through the relation $\phi = 1/\kappa$. The form in equation (2.13) is adhered to in this thesis for the Matérn covariance function. The integer value of ν determines the mean-square differentiability of the latent field, which in turn, determines the smoothness of the spatial field.

2.3.2 Gaussian fields

Gaussian fields (GFs) are spatial fields which build upon the Gaussian distribution. As GFs share the analytical properties of the Gaussian distribution, they form a flexible and practical class of spatial models. Further, GFs are one of a few multivariate models with an explicit and computable normalizing constant. Due to these properites, GFs play a dominant role in the field of spatial statistics, see for example Cressie (1993), Diggle et al. (1998) and Stein (2012). The definition of a GF a over spatial domain $\mathcal{D}$ is as follows.

Definition 6. *Let $\{u(\boldsymbol{s})\}$ be a spatial field over a spatial domain $\mathcal{D}$. The field is called a Gaussian field with a mean function and a covariance function*

$$\mu(\boldsymbol{s}) = \mathbb{E}\left[u(\boldsymbol{s})\right], \quad C(\boldsymbol{s},\boldsymbol{t}) = Cov(u(\boldsymbol{s}), u(\boldsymbol{t})), \quad \text{for all } \boldsymbol{s},\boldsymbol{t} \in \mathcal{D}$$

if for any $n \geq 1$ and any finite collection of locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$ the random vector $\boldsymbol{u} = (u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_n))^{\mathsf{T}}$ follows a multivariate Gaussian distribution with a mean vector $\boldsymbol{\mu} = (\mu(\boldsymbol{s}_1), \ldots, \mu(\boldsymbol{s}_n))^{\mathsf{T}}$ and a covariance matrix $\boldsymbol{\Sigma}_{ij} = C(\boldsymbol{s}_i, \boldsymbol{s}_j)$, that is

$$\boldsymbol{u} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right). \tag{2.14}$$

Gaussian fields are thus uniquely determined by the mean function and the covariance function. The Gaussian vector in (2.14) serves as finite dimensional representation of the GF $\{u(\boldsymbol{s})\}$. The following theorem, see Rue and Held (2005) for proof, demonstrates analytical closure properties of the Gaussian distribution which are particularly useful in a spatial setting.

Theorem 1. *Let $\boldsymbol{u}$ be an n-dimensional random vector such that $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Further, split $\boldsymbol{u}$ into two parts, $\boldsymbol{u} = (\boldsymbol{u}_A^\mathsf{T}, \boldsymbol{u}_B^\mathsf{T})^\mathsf{T}$ and split the mean vector $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ accordingly, that is,*

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}.$$

Then the following holds

a) $\boldsymbol{u}_A \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA})$

b) $\boldsymbol{\Sigma}_{AB} = 0$ if and only if $\boldsymbol{u}_A$ *and* $\boldsymbol{u}_B$ *are independent*

c) The conditional density of $\boldsymbol{u}_A$ *conditioned on* $\boldsymbol{u}_B$ *is*

$$\pi(\boldsymbol{u}_A|\boldsymbol{u}_B) = \mathcal{N}\left(\boldsymbol{u}_A|\boldsymbol{\mu}_{A|B}, \boldsymbol{\Sigma}_{A|B}\right) \tag{2.15}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{A|B} &= \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\boldsymbol{u}_B - \boldsymbol{\mu}_B) \\ \Sigma_{A|B} &= \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA}. \end{aligned} \tag{2.16}$$

d) Assume that $\boldsymbol{Z} \in \mathbb{R}^{k\times n}$*, where* $k \leq n$*, is some fixed matrix, then*

$$\pi(\boldsymbol{Z}\boldsymbol{u}) = \mathcal{N}\left(\boldsymbol{Z}\boldsymbol{u}|\boldsymbol{Z}\boldsymbol{\mu}, \boldsymbol{Z}\boldsymbol{\Sigma}\boldsymbol{Z}^\mathsf{T}\right).$$

To demonstrate how the properties in Theorem 1 can be utilized for modeling spatial data in a domain $\mathcal{D}$ with a GF $\{u(\boldsymbol{s})\}$ in the Bayesian paradigm, consider the following. Let $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$ be any locations in the spatial domain selected for observing the spatial physical process. Impose a Gaussian prior density on the random vector $\boldsymbol{u} = (u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_n))^\mathsf{T}$, that is, the prior density of $\boldsymbol{u}$ is

$$\pi(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{u}|\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

and $\boldsymbol{u}$ is assumed to be observable at the locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$. Further assume that both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ have been estimated based on the observations[9]. Then, estimates of any subvector $\boldsymbol{u}_A \subset \boldsymbol{u}$ follow immediately, that is without any additional estimation calculations, from the estimation of $\boldsymbol{u}$, by property *a)* in Theorem 1. This is particularly useful from a computational point of view, as estimates of the covariance structure of any $\boldsymbol{u}_A \subset \boldsymbol{u}$ is readily available. Further on that point, by property *b)* in Theorem 1, the marginal dependence structure of $\boldsymbol{u}$ is also available without the need of any extra calculations.

Next, assume that a set of locations $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_m \in \mathcal{D}$ are given where the values of the GF are unobserved, and denote the unobserved values with $\boldsymbol{u}_{un} =$

[9] In Bayesian hierarchical modeling, the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ are allowed to depend on any potential hyperparameters $\boldsymbol{\theta}$. The estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ then involves evaluating the joint posterior distribution of the model parameters.

$(u(\boldsymbol{t}_1), \ldots, u(\boldsymbol{t}_m))^\mathsf{T}$. Property *c)* in Theorem 1 can be utilized to make statements about the values at the unobserved locations. That is, by the model assumptions and property *c)* in Theorem 1 the conditional distribution of $\boldsymbol{u}_{un}$ conditioned on the observed values $\boldsymbol{u}$ is a Gaussian distribution given by equations (2.15) and (2.16), with $\boldsymbol{u}$ and $\boldsymbol{u}_{un}$ in the role of $\boldsymbol{u}_A$ and $\boldsymbol{u}_B$, respectively. This procedure thus yields spatial predictions of the unobserved values. The spatial predictions are, in a sense, a best linear prediction given the model values and referred to as *Kriging*, see Stein (1999) for further discussion.

Property *d)* in Theorem 1 states that any linear combination of a Gaussian vector $\boldsymbol{u}$ is also a Gaussian vector. This property will become particularly useful for approximating GFs with Gaussian Markov random fields. A further discussion on the subject can be found in Section 2.3.3 and Section 2.3.4.

2.3.3 Gaussian Markov random fields

Although GFs are analytically convenient from both an interpretational and implementational standpoint, they become increasingly computationally demanding to evaluate as the dimensions of the corresponding finite Gaussian vectors increase, for example when the number of observational points increase. Consider, from a computational point of view, an n-dimensional Gaussian vector $\boldsymbol{u}$ from a GF, such that $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In the calculations of, for instance, the log-likelihood of $\boldsymbol{u}$, which is given by

$$\log \pi(\boldsymbol{u}) = -\frac{1}{2} \log \det \boldsymbol{\Sigma} - \frac{1}{2}(\boldsymbol{u} - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\boldsymbol{u} - \boldsymbol{\mu}) + \text{constant}$$

both determinant and inverse calculations[10] of the $n \times n$ covariance matrix $\boldsymbol{\Sigma}$ are involved. As $n \times n$ covariance matrices have $\mathcal{O}(n^2)$ unique elements, both the determinant and inverse calculations scale as $\mathcal{O}(n^3)$. However, GFs can be approximated with Gaussian Markov random fields (GMRFs), which in turn can increase the speed of computation significantly, see Rue (2001) and Rue and Held (2005).

A fundamental concept for both the interpretation and the structure of GMRFs is conditional independence, which is defined as follows.

Definition 7. *Let $\boldsymbol{u}$ be a finite random vector. Two elements u_i and u_j, such that $i \neq j$ of $\boldsymbol{u}$ are said to be conditionally independent given $\boldsymbol{u}_{-ij}$ if the random variables $u_i|\boldsymbol{u}_{-ij}$ and $u_j|\boldsymbol{u}_{-ij}$ are independent, denoted $u_i \perp u_j|\boldsymbol{u}_{-ij}$.*

A GMRF is essentially an n-dimensional Gaussian vector parameterized with a *precision matrix* $\boldsymbol{Q}$, which is the inverse of the covariance matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$. Further, a GMRF is defined with respect to certain *neighbor index sets*[11]. That is, for given n locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$, let $N(i)$ denote the neighbor index set of locations which are "close" to the location $\boldsymbol{s}_i$, in the sense that,

$$N(i) = \Big\{ j = 1, \ldots, n \mid j \neq i,\ u_i \text{ and } u_j \text{ are conditionally dependent given } \boldsymbol{u}_{-ij} \Big\},$$

[10] In form of solving an $n \times n$ linear system.

[11] GRMFs are alternatively defined with respect to a *directed graph*, see Rue and Held (2005).

for all $i = 1, \ldots, n$, and denote the collection of neighbor index sets as $\{N(i)\}_{i=1}^{n}$. A Gaussian Markov random field can then be defined with respect to the neighbor index sets as follows.

Definition 8. *An n-dimensional Gaussian vector* $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ *which satisfies*

$$u_i|\boldsymbol{u}_{-i} \stackrel{D}{=} u_i|\{u_j : j \in N(i)\}, \quad \textit{for all } i = 1, \ldots, n$$

is referred to as a Gaussian Markov random field with respect to the collection of neighbor index sets $\{N(i)\}_{i=1}^{n}$.

To compare the interpretation of the elements of a precision matrix $\boldsymbol{Q}$ to the elements of the corresponding covariance matrix $\boldsymbol{\Sigma}$, consider the following theorem.

Theorem 2. *Let* $\boldsymbol{u}$ *be a GMRF as in Definition 8. Then the following holds*

$$\begin{aligned}
\mathbb{E}\left[u_i|\boldsymbol{u}_{-i}\right] &= \mu_i - \frac{1}{Q_{ii}} \sum_{j \in N(i)} Q_{ij}(x_j - \mu_j), \\
Prec\left(u_i|\boldsymbol{u}_{-i}\right) &= Q_{ii}, \\
Corr\left(u_i, u_j|\boldsymbol{u}_{-i,j}\right) &= -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}, \quad \textit{for } i \neq j.
\end{aligned}$$

See Rue and Held (2005) for proof. Theorem 2 states that the diagonal elements of $\boldsymbol{Q}$ consist of the conditional precisions of u_i conditioned on $\boldsymbol{u}_{-i}$, for all i, and the off-diagonals provide insight into the conditional correlation between u_i and u_j conditioned on $\boldsymbol{u}_{-ij}$. On the other hand, the covariance matrix yields information on the marginal variances for all u_i and the marginal correlation between every pairs of u_i of u_j. Therefore, interpretation of elements in precision matrices is thus arguably not as intuitive as the interpretation of the elements in the covariance matrix. In terms of interpretation, Rue and Held (2005) comment that "*The interpretation provided by* $\boldsymbol{Q}$ *is hard (or nearly impossible) to interpret marginally [...]*".

However, the parameterization of GRMF through precision matrices, as opposed to covariance matrices, reveals a connection between the Markov property and the neighbor index sets of a GRMF, as illustrated in the following theorem.

Theorem 3. *Let* $\boldsymbol{u}$ *be a GMRF as in Definition 8. Then*

$$u_i \perp u_j|\boldsymbol{u}_{-ij} \quad \textit{if and only if} \quad Q_{ij} = 0.$$

In other words, Theorem 3 demonstrates that the nonzero pattern of a precision matrix $\boldsymbol{Q}$ determines the conditional dependence structure of neighbor index sets. Reversely, the conditional independence structure of the neighbor index sets induces the zero pattern in $\boldsymbol{Q}$. As a consequence, the Markov property thus generates sparse precision matrices for GMRFs in a spatial setting. This fact leads to fast computation of samples, likelihoods and other quantities of interest of a

GMRF in a spatial setting[12]. For example, in a one-dimensional spatial setting, where $\mathcal{D} \subset \mathbb{R}$, these quantities can be computed using $\mathcal{O}(n)$ operations. In a two dimensional spatial setting, where $\mathcal{D} \subset \mathbb{R}^2$, these quantities can be computed using $\mathcal{O}(n^{3/2})$ operations, see Simpson et al. (2012). In both of the these spatial settings, which are the primary spatial settings in this thesis, using GMRFs as opposed to GFs for modeling spatial data offers significant advantages to computational speed.

Another important result of GMRFs, from a computational point of view, is stated in the following theorem.

Theorem 4. *Let $\boldsymbol{u}$ be a GMRF as in Definition 8. Split the vector $\boldsymbol{u}$ into two parts, $\boldsymbol{u} = (\boldsymbol{u}_A^\mathsf{T}, \boldsymbol{u}_B^\mathsf{T})^\mathsf{T}$ and split the mean vector $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ accordingly, that is,*

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \quad \boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{AA} & \boldsymbol{Q}_{AB} \\ \boldsymbol{Q}_{BA} & \boldsymbol{Q}_{BB} \end{pmatrix}.$$

The vector $\boldsymbol{u}_A|\boldsymbol{u}_B$ is then a GMRF with respect to sub-neighbor index set $\{N_A(i)\}_{i\in A}$, where

$$N_A(i) = \Big\{ j \in A \mid j \neq i,\ u_i \text{ and } u_j \text{ are conditionally dependent given } \boldsymbol{u}_{-ij} \Big\},$$

for all $i \in A$. The conditional density function $\boldsymbol{u}_A$ given $\boldsymbol{u}_B$ thus becomes

$$\pi(\boldsymbol{u}_A|\boldsymbol{u}_B) = \mathcal{N}\left(\boldsymbol{u}_A|\boldsymbol{\mu}_{A|B}, \boldsymbol{Q}_{A|B}\right) \tag{2.17}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{A|B} &= \boldsymbol{\mu}_A + \boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}(\boldsymbol{u}_B - \boldsymbol{\mu}_B) \\ \boldsymbol{Q}_{A|B} &= \boldsymbol{Q}_{AA} \end{aligned} \tag{2.18}$$

The computational importance of Theorem 4 stems from the fact that if $\boldsymbol{Q}$ is sparse, then every precision matrix $\boldsymbol{Q}_{A|B}$ of the corresponding conditional, as implied by equations (2.17), is also sparse, as demonstrated by equation (2.18). In other words, the sparsity of the precision matrix $\boldsymbol{Q}$ is preserved under any conditioning of any subvectors of $\boldsymbol{u}$. This property is referred to as *preserving the sparse GMRF precision structure* in this thesis. Comparing equations (2.18) in Theorem 4 to corresponding general result (2.16) in Theorem 1 reveals the computational advantages presented by the GMRF structure under conditioning. This result is utilized extensively in both **Paper II** and **Paper III**.

2.3.4 SPDE spatial models

GMRFs can offer significant computational advantages over GFs in spatial setting, as seen in Section 2.3.3, when their precision structures are sparse. To that

[12]The key factor in the fast computations of GMRFs, is the Cholesky decomposition of the sparse precision matrix. That is, $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^\mathsf{T}$, where $\boldsymbol{L}$ is a lower triangular matrix. Rue and Held (2005) provide an extensive overview of fast algorithms for GMRFs.

extent, Lindgren et al. (2011) proposed a method to parametrize the precision matrix of a GMRF to achieve a predefined spatial covariance which yields sparse precision structure. The method computes a numerical approximation to a GF on a triangulated mesh based on a stochastic partial differential equation (SPDE) representation, which is referred to a SPDE spatial models in this thesis.

The theory on SPDE spatial models is built upon the following theorem.

Theorem 5. *The solution* $\{u(\boldsymbol{s})\}$ *to the following fractional SPDE*

$$(\kappa^2 - \Delta)^{\alpha/2} u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{2.19}$$

is an isotropic mean zero GF with a Matérn covariance function as given in equation (2.13).

The process $\mathcal{W}(\boldsymbol{s})$ in Theorem 5 is an isotropic zero mean GF over $\mathbb{R}^d$ with a unit variance[13]; the term $(\kappa^2 - \Delta)^{\alpha/2}$ serves as a pseudo-differential operator[14] ; and Δ is the Laplacian. Further, the marginal variance of the solution $\{u(\boldsymbol{s})\}$ is

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}}.$$

Lindgren et al. (2011) proposed constructing a finite dimensional representations of the solution to the SPDE in (2.19) with a sparse GRMF precision structure. To elaborate, first consider the weak solution form of the SPDE in (2.19), which given by

$$\left\{\langle \phi_j, (\kappa^2 - \Delta)^{\alpha/2} u\rangle\right\}_{j=1}^{m} \overset{D}{=} \left\{\langle \phi_j, \mathcal{W}\rangle\right\}_{j=1}^{m}, \tag{2.20}$$

for a set of test functions $\{\phi_j(\boldsymbol{s})\}_{j=1}^{m}$, where the inner product is defined as

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

The spatial domain is then subdivided into a set of non-intersecting triangles, referred to as a triangulated mesh. For every point $\boldsymbol{s}$ in the spatial domain, a finite element representation (Brenner and Scott 2008) of the solution $u(\boldsymbol{s})$ is constructed. The finite element representation is of the form

$$u(\boldsymbol{s}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{s}) w_k \tag{2.21}$$

where n is the number of vertices in the mesh, ψ_k are a set of basis functions, and w_k are Gaussian weights at the vertices of the triangles in the mesh. Lindgren et al. (2011) selected the functions in $\{\psi_k\}_{k=1}^{n}$ in (2.21) as a piecewise linear basis

[13] An isotropic mean-zero Gaussian field with a unit variance is referred to as a white noise by many authors.

[14] A general pseudo-differential operator is defined by the Fourier inversion formula.

functions within each triangle, such that ψ_k is 1 at vertex k and 0 at other vertices. The Gaussian weights in the finite element representation (2.21) thus determine the values at the vertices of the mesh. Moreover, the values in the interior of the triangles are determined by linear interpolation of the surrounding vertices[15]. As a result, the methods yields a continuously indexed approximate solution to the SPDE in (2.19). In practice this mean that the approximate solution can be projected from the vertices of the mesh with a linear transformation onto every finite set of points, denoted with $\mathcal{S}$, within the spatial domain. Such a linear projection matrix is denoted with $\boldsymbol{A}_{\mathcal{S}}$ in this thesis.

In order to construct a finite dimensional representations of the solution, the solution $u(\boldsymbol{s})$ in equation (2.20) is replaced with the finite element representation in (2.21). The distribution of Gaussian the weights $\boldsymbol{w} = (w_1, \ldots, w_n)^\mathsf{T}$ are then found such that (2.20) holds, for a specific set of test functions $\{\phi_k\}_{k=1}^m$, with $m = n$. The structure of the test functions affect the structural properties of the corresponding the finite dimensional representations. For further details on the choice of test functions see Lindgren et al. (2011). Finally, by defining the following precision matrices

$$\begin{aligned} C_{ij} &= \langle \psi_i, \psi_j \rangle \\ G_{ij} &= \langle \nabla\psi_i, \nabla\psi_j \rangle \\ K_{ij} &= \kappa^2 C_{ij} + G_{ij}, \end{aligned}$$

and using Neumann boundary condition on the boundary of the spatial domain, the following be shown.

Theorem 6. *Let $\boldsymbol{Q}_\alpha(\kappa^2)$ denote the precision matrix for the Gaussian weights $\boldsymbol{w}$ in the finite element representations in (2.21), for $\alpha = 1, 2, \ldots$ as a function of κ^2. Then the finite dimensional representations of the solution to the SPDE in (2.19)*

$$\begin{aligned} \boldsymbol{Q}_1(\kappa^2) &= \boldsymbol{K}(\kappa^2) \\ \boldsymbol{Q}_2(\kappa^2) &= \boldsymbol{K}(\kappa^2)\boldsymbol{C}^{-1}\boldsymbol{K}(\kappa^2) \\ \boldsymbol{Q}_\alpha(\kappa^2) &= \boldsymbol{K}(\kappa^2)\boldsymbol{C}^{-1}\boldsymbol{Q}_{\alpha-2}\boldsymbol{C}^{-1}\boldsymbol{K}(\kappa^2) \quad \alpha \geq 3 \end{aligned}$$

From a computational point of view, consider the following. The elements in $\boldsymbol{C}$ and $\boldsymbol{G}$ are non-zero only for pairs of basis functions which share common triangles. This means $\boldsymbol{C}$ and $\boldsymbol{G}$ are both sparse and easily computable. Moreover, the values of the elements of $\boldsymbol{C}$ and $\boldsymbol{G}$ are independent of the parameter κ. From a Bayesian computational standpoint, this means that the $\boldsymbol{C}$ and $\boldsymbol{G}$ matrices can be computed invariantly of Bayesian inference computations. For example, in the MCMC setting, the $\boldsymbol{C}$ and $\boldsymbol{G}$ do not have to be recalculated in every iterations, which reduces the computational cost of every iteration significantly.

Lindgren et al. (2011) further demonstrate that, although the matrix $\boldsymbol{C}^{-1}$ is dense, it can be replaced by a diagonal matrix $\widetilde{\boldsymbol{C}}$, such that, $\widetilde{C}_{ii} = \langle \psi_i, 1 \rangle$.

[15]The linear interpolation into the triangles is essentially as convex linear combination of Gaussian weights at the surrounding vertices.

This in turn makes all the precision matrices in Theorem 6 sparse. As a result, Theorem 6 yields a way to construct a GMRF representations with a sparse precision structure of a Gaussian field with a Matérn covariance function.

2.4 Extreme precipitation

Understanding the spatial behavior of extreme precipitation, and the frequency and intensity of extreme events is important for public safety and various types of long term agricultural, industrial and urban planning. Extreme precipitation can be a major trigger of floods in urban areas resulting in large economic loss every year, and can be a threat to life and property. For example, extreme rainfall in Reykjavík on the 16th of August 1991 resulted in overloaded local drainage systems causing severe damage to industrial buildings, houses and apartments. As an another example, the Norwegian Water Resources and Energy Directorate (NVE) has estimated the average annual cost of flood damage in Norway to be about 200 million NOK, see Dyrrdal (2012). Moreover, the information on extreme precipitation is needed at a fine resolution as the characteristics of extreme precipitation events can be local, especially in regions with heterogeneous topography.

Meteorological models, of various complexities, have been proposed to model precipitation spatially. Although meteorological models are well suited to give insight into the spatial mean behavior of precipitation, they tend to deviate significantly from observations when predicting extreme precipitation, see for example the analyzes given in Section 4.3.2. Thus, there is an additional need for statistical methods which take into account observations and spatial variation of extreme precipitation events. Reliable fine scale probabilistic spatial predictions of extremes are of importance as they can provide climatologists with more informative maps on the severity and variability in extremes. Such predictions are also important for mapping risk to human health and useful for engineering purposes, for instance, for design and safety issues. A map showing the severity for locations on a fine grid makes it possible to identify which areas are subject to greater risk than others

In this section, a brief overview of both modern meteorological precipitation models and statistical models for extreme precipitation is given.

2.4.1 Meteorological models for precipitation

Meteorological models, of various complexities, have been proposed to model precipitation on a fine grid. Traditional meteorological models implement interpolation methods to estimate area-averaged precipitation or to produce a gridded datasets based on observations from irregularly spaced network of observational sites, see for example Creutin and Obled (1982). However, such methods are inadequate for mountainous regions with large spatial gradients from both a theoretical and practical standpoint. In order to tackle this problem various meteorological models have been proposed, such as regression models with topographical and atmospherical covariates (Basist et al. 1994, Daly et al. 1994, Wotling et al.

2000). These methods, however, rely heavily on observational coverage and can fail to provide good estimates of precipitation in the absence of a dense network of observational sites. To address these issues, Smith and Barstad (2004) proposed a quasi-analytic linear orographic method to simulate precipitation over a spatial domain that takes into account the effects of topography as well as atmospheric information on orographic precipitation. Furthermore, Crochet et al. (2007) have adopted the method to precipitation in Iceland. The resulting model simulates daily precipitation on a 1 km by 1 km regular grid across Iceland over the years 1958-2002. Crochet et al. (2007) showed that this model is well suited to give insight into the observed spatial mean behavior of precipitation.

2.4.2 Statistical models for extreme precipitation

In recent years, several statistical methods capable of better quantifying the probability of extreme events have grown in popularity in environmental sciences, see for example Coles and Tawn (1996) and Sang and Gelfand (2009). The standard likelihood based approach in the literature is modeling extreme events with the generalized extreme value (g.e.v.) distribution[16] , which has a cumulative density function of the form

$$F(y) = \exp\left\{-\left(1+\xi\left(\frac{y-\mu}{\sigma}\right)\right)^{-1/\xi}\right\} \tag{2.22}$$

if $1+\xi(x-\mu)/\sigma > 0$, $F(y) = 0$ otherwise. The parameters μ, σ and ξ serve as location, scale and shape parameters, respectively. The shape parameter ξ governs the tail behavior of the distribution. That is, if $\xi > 0$ the g.e.v. distribution has a lower bound, such that the support is $y > \mu - \sigma/\xi$, and belongs to the Fréchet family of distributions. For the values $\xi < 0$ then the g.e.v. distribution has an upper bound, with support $y < \mu - \sigma/\xi$, and belongs to the negative Weibull family of distributions. For the value $\xi = 0$, the expression in (2.22) is undefined. In this case, the expression in (2.22) is replaced by the limiting distributional form when $\xi \to 0$, that is

$$F(y) = \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\},$$

which is the Gumbel type distribution, which in turn is defined for all $y \in \mathbb{R}$.

Data on extreme precipitation in Iceland have been analyzed before using statistical methods, see Elíasson (1994). Spatial predictions of extreme precipitation in Iceland were given in Eliasson et al. 2009 where precipitation simulations on an 8 km by 8 km grid based on the PSU/NCAR Mesoscale Model MM5 were used to predict annual extreme 24-hour precipitation with 5 years return period. Two of the suggestions in Eliasson et al. (2009) were to check the assumptions of the Gumbel distribution at the data level and to move to a finer grid

As previously discussed, the Bayesian hierarchical model offers a flexible probabilistic framework for spatial modeling of physical phenomena and quantifying

[16] Alternative modeling approaches have been explored, such as the peaks over threshold methods with the generalized Pareto distribution (Cooley et al. 2007).

uncertainty of the latent physical processes; see for example in Berliner (1996), Berliner (2003), and Tebaldi and Sansó (2009). Bayesian hierarchal modeling is also well suited to leverage external information on physical phenomena into the statistical model. For example, Cooley et al. (2007) demonstrated that covariates which assimilate available spatial information, such as the topography of the domain and the underlying physical processes of precipitation, are reasonable covariates for extreme precipitation.

As an example of Bayesian hierarchal modeling of extreme precipitation, Sang and Gelfand (2009) implemented a Bayesian hierarchical model with a g.e.v. data density function to model extreme rainfall for the Cape Floristic Region of South Africa. Spatial autoregressive models[17] were imposed on locations and scaling parameters to model the spatial variability. Another comparable spatial modeling is presented by Cooley and Sain (2010), where extreme precipitation simulated by a regional climate model is characterized with a Bayesian hierarchical model, where all the data density parameters are allowed to vary in space. However, using autoregressive models to the spatial variation is only feasible if the data are observed on a regular grid, which is rarely the case, and in particular, not the case for observational sites in Iceland.

Alternative approaches to modeling extreme precipitation is, for example, the peaks over threshold method with the generalized Pareto distribution as presented in (Cooley et al. 2007). The advantage of that approach is that more of the data can be incorporated into the modeling, depending on the choice of threshold. Davison et al. (2012) give another alternative approach to modeling of maximum precipitation, where a t-copula with g.e.v. marginals is utilized. The method presents an appealing choice to model the probabilistic dependence of the observations and for simulating spatial realizations of the extreme precipitation for the next year or unobserved sites for an observed year.

Guttorp and Gneiting (2006) demonstrated that Gaussian fields with a Matérn covariance function present a flexible and interpretable way of modeling underlying physical processes of natural phenomena in a spatial setting. This approach does not require the observational sites to be located an a regular grid. For example, Hrafnkelsson et al. (2012) implemented a Bayesian hierarchical model with a g.e.v. data density to model extreme temperatures in Iceland. Spatial models based on Gaussian fields with a Matérn covariance function were imposed on all of the parameters of the data density, to model the spatial variation of the underlying processes. However, as discussed in Section 2.3 posterior inference for Gaussian fields becomes increasingly computationally demanding as number of observational points increase. As an example of this can be seen in Schliep et al. (2010), where a spatial model with a GF and Matérn covariance structure is found to be too computationally challenging for The North American Regional Climate Change Assessment (NARCCAP) due to the high dimension of the underlying GF over the domain of NARCCAP. **Paper III** addresses the problem of high dimensional GFs by implementing SPDE spatial models.

[17] which are essentially GMRFs with a neighbor index sets defined on a regular grid

3

Aim

The main goals and objectives of this research in this Ph.D. research project are essentially twofold: Firstly, to propose a computationally efficient spatial modeling framework for maximum precipitation on a fine grid. Secondly, to develop a computationally efficient and general inference scheme for latent Gaussian models which is, in particular, applicable to computationally challenging spatial modeling problems. These objective are addressed in the three papers presented in this thesis. The following summarizes the novel contributions presented in each paper.

Paper I: The main objective of the discussion in **Paper I** is to demonstrate that LGMs present a flexible modeling class for various statistical problems. That is, the authors establish that LGMs can be viewed as a specific extension of structured additive regression models Fahrmeir et al. (1994), in the sense that, the data density function of each data point can depend on more than a single linear link function of the latent parameters through more than just the mean structure of the data density.

Paper II: The aim of the research presented in **Paper II** is to develop a novel computationally efficient MCMC sampling scheme for LGMs. The authors refer to the sampling scheme as *the MCMC split sampler*. The proposed sampling is designed to be naturally applicable to LGMs with latent models imposed on more than just the mean structure of the data density function; designed for any choice of a parametric data density function; and designed to scale well in terms of computational efficiency when the dimension of the latent model increase.

In the paper a theoretical basis is established for the development of MCMC split sampler, along with necessary proofs. Furthermore, two examples are given where the MCMC split sampler is used as an inferential algorithm for LGMs. Both the examples demonstrate that the MCMC split sampler is

computationally efficient in terms good mixing properties of the posterior samples, and it scales well as the dimensions of the latent fields increase.

Paper III: The main goal in **Paper III** is to propose a computationally efficient spatial model for annual maximum 24-h precipitation on a fine grid. To achieve this goal, the two following novel contributions were are proposed.

First, a the method is proposed which leverages information on the physical processes of precipitation from a local meteorological model, in order to yield additional spatial information on maximum precipitation. The proposed method is extendable to any regions in the world where local meteorological models are available.

Secondly, a LGM is constructed with a data density based on the g.e.v. distribution, where the spatial variation of the underlying physical processes of extreme precipitation is modeled with the SPDE approach. To that extend, SPDE spatial models are imposed on the location, scale and shape parameters of the g.e.v. distribution. Although LGMs with a GMRF structure have been proposed to model extreme precipitation, LGMs with spatial models based on the SPDE approach have not been proposed in the literature on spatial extreme precipitation before. The proposed modeling strategy is general in the sense that it is extendable to any spatial domain of interest. The proposed LGM is applied to data on annual maximum 24-hour precipitation in Iceland.

4

Proposed statistical methodology

In this chapter, an overview of the proposed statistical methodology developed in this thesis is given. The discussion in this chapter cycles through the proposed statistical methodology developed and implemented in the thesis and is in accordance with the objectives outlined in Chapter 3, with appropriate references to the three papers. Additional discussion and motivation to the material in the three papers is also given when relevant.

4.1 Towards applicable, flexible and interpretable statistical modeling

In **Paper I**, the authors participate in a research discussion where they share their view on how modern statistical analysis of involved problems should be thought of and practiced. The authors' stance towards a general approach to statistical analysis in practice can be summarized in the following three steps.

1. Design a sensible statistical model for a specific problem. This can be accomplished through the probabilistic framework offered by Bayesian statistical modeling, as discussed in Section 2.1.

2. Implement, or develop, a inferential algorithm which can provide the posterior distribution of the proposed model. Section 2.2 offers a brief overview of the building blocks for Bayesian inferential algorithms.

3. Based on samples from the posterior distribution obtained by applying the inferential algorithm, compute quantities of interest such as posterior means, posterior intervals and posterior quantiles.

In the authors' view, "modeling" and "inferential algorithms" should be considered as two separate but not entirely unrelated entities. That is, in some sense, as two opposite sides of the same coin. As such, the design of a sensible statistical model for a specific problem should, in an ideal situation, be accomplished without the concern for the availability of fast inferential algorithms for it[1]. Once a model has been proposed and put forth, then a suitable inferential algorithm should be implemented if available, or developed if not available[2].

In the design of a sensible statistical model the focus should, from the authors' perspective, be on statistical modeling as opposed to *regression modeling*. In order to clarify the authors' viewpoint, consider a Gaussian response with one covariate z, where the following statistical model is imposed

$$y_i \mid \ldots \sim \mathcal{N}(\beta_0 + \beta_1 z_i, \sigma^2).$$

This statistical model does in fact encompass a linear mean "regression" model of the type

$$y_i = \beta_0 + \beta_1 z_i + \epsilon_i, \quad \text{where} \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2),$$

but that is due to the fact that the mean structure is a natural parameter in the Gaussian distribution. Further, consider a regression model that "goes beyond mean regression". For example, using the same covariate z, a regression model that regresses both the mean and the log variance parameter on the data y. This general regression modeling[3] can also be achieved through the use of the following probabilistic model

$$y_i \mid \ldots \sim \mathcal{N}(\beta_0 + \beta_1 z_i, \exp(b_0 + b_1 z_i)).$$

To allow for more flexibility the linear dependence on the covariate z may be relaxed

$$y_i \mid \ldots \sim \mathcal{N}(f(z_i), \exp(g(z_i))),$$

where, $f(\cdot)$ and $g(\cdot)$ are two "smooth" unknown functions of the covariate z, typically some finite-dimensional GF representations. Assuming a Bayesian inferential algorithm is available, or can be developed, inference for any quantities of interest can be extracted from the posterior distribution alone. From the authors' point of view, this way of thinking about statistical analysis moves the focus from regression modeling towards statistical modeling.

With this discussion in mind, the question remains whether a common, flexible and interpretable class of Bayesian models exists for which computationally efficient and fast inferential algorithms are available or can be developed? To that end, the class of latent Gaussian models (LGMs), introduced in Section 2.1.2, potentially provides such a statistical framework. LGMs serve as a flexible and interpretable model class which stems from the mathematical properties of

[1]The authors' recognize that this is not always the case, and from a practical point of view the authors further agree on that "A man's got to do what a man's got to do".

[2]Which is exactly the story in this thesis and how the MCMC split sampler came into being.

[3]This class of regression models is called "Generalized Additive Models for Location, Scale and Shape", abbreviated with GAMLSS, see Rigby and Stasinopoulos (2005) for further details.

the Gaussian distribution. While this property of LGMs is true in general for various statistical modeling problems, this is particularly useful in spatial statistics as discussed in Section 2.3.2, Section 2.3.3 and Section 2.3.4.

To add a viewpoint to the previous discussion on the flexibility of LGMs, consider several linear predictors, say one for each natural parameter of a given data density with k parameters, that is

$$\boldsymbol{\eta}^{(1)}, \ldots, \boldsymbol{\eta}^{(k)}.$$

In the LGM framework, the joint prior distribution of $(\boldsymbol{\eta}^{(1)}, \ldots, \boldsymbol{\eta}^{(k)})$ conditioned on any potential hyperparameters $\boldsymbol{\theta}$ is Gaussian. Assuming the observations in $\boldsymbol{y}$ are conditionally independent over i, and depended on all the linear predictors in *some* form, the following statistical model can be constructed

$$y_i \mid \ldots \sim \pi(y_i \mid \eta_i^{(1)}, \ldots, \eta_i^{(k)}). \tag{4.1}$$

The model is completed with a prior density on the non-Gaussian components in the vector $\boldsymbol{\theta}$, denoted by $\pi(\boldsymbol{\theta})$. This model formulation is further generalized in the hierarchical form given below, which is an alternative yet an equivalent expression of the formulation in Section 2.1.2. The authors refer to the following LGM setup as the extended LGM framework[4], when the need for such clarity is necessary.

Stage 1: The observations $\boldsymbol{y}$ depend on the latent field $\boldsymbol{x}$ and potential hyperparameters $\boldsymbol{\theta}$, through

$$y_i \mid \boldsymbol{x}, \boldsymbol{\theta} \sim \pi(y_i \mid \{x_j : j \in \mathcal{I}_i\}, \boldsymbol{\theta})$$

for some set $\mathcal{I}_i$.

Stage 2 The latent field $\boldsymbol{x}$ is Gaussian with hyperparameters $\boldsymbol{\theta}$,

$$\boldsymbol{x} \mid \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}(\boldsymbol{\theta})^{-1}\right)$$

Stage 3 The hyperparameters are given a prior density $\pi(\boldsymbol{\theta})$.

To emphasize, in the extended LGM framework the linear predictors $\boldsymbol{\eta}^{(1)}, \ldots, \boldsymbol{\eta}^{(k)}$ can be associated with any parameters of the data density, not only the location parameter. In this thesis, this property of the extended LGM framework is often phrased as "imposing latent models on more than just the mean structure of the data density".

From an inferential perspective, the mathematical structure of LGMs is well suited for the development of efficient inferential algorithms. Of particular notice is the integrated nested Laplace approximation (INLA) proposed by Rue et al. (2009). INLA is an inferential algorithm which performs approximate Bayesian inference on a subclass of LGMs, through use of deterministic nested Laplace approximations. INLA has been shown to give fast and accurate estimates of

[4]This is referred to as "LGM-2.0" in **Paper I**

posterior marginals, and has also been shown to be a valuable tool in practice via the R-package `R-INLA`; see web-site `www.r-inla.org` for available software. However, `R-INLA` only provides support for LGMs in which the data density of each data point only depends on a single linear predictor η_i of the latent parameters, where η_i is limited to the location parameter, such that

$$y_i \mid \ldots \sim \pi(y_i \mid \eta_i).$$

In **Paper II** a computationally efficient simulation based inferential algorithm is proposed that is designed, in particular, to be applicable to statistical models from the extended LGM framework.

The concept of LGMs has proven to be very successful for doing Bayesian inference, and a wealth of models may be presented as LGMs, which in turn are amenable to the Bayesian inference approaches developed specifically for LGMs; see Martins et al. (2013b) for various applications and further references.

4.1.1 A demonstration of an extended LGM setup

Within the extended LGM framework, latent models can be imposed on any parameter of a given data density function. In **Paper II**, a demonstration of such an LGM setup is given as a motivational example. In this section, this demonstration is given to highlight how LGMs can be expressed in order to allow latent models to enter more than just the mean structure of a given data density.

Consider, without loss of generality, a data density function $\pi(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\tau})$ where $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$ are vectors of location and log-scale parameters, respectively. Next, impose latent Gaussian models on both $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$, for example, the following additive model structure

$$\boldsymbol{\mu} = \boldsymbol{X}_\mu \boldsymbol{\beta}_\mu + \boldsymbol{A}_\mu \boldsymbol{u}_\mu + \boldsymbol{\epsilon}_\mu \quad \text{and} \quad \boldsymbol{\tau} = \boldsymbol{X}_\tau \boldsymbol{\beta}_\tau + \boldsymbol{A}_\tau \boldsymbol{u}_\tau + \boldsymbol{\epsilon}_\tau, \tag{4.2}$$

where $\boldsymbol{X}_\mu$ and $\boldsymbol{X}_\tau$ are fixed design matrices; $\boldsymbol{\beta}_\mu$ and $\boldsymbol{\beta}_\tau$ are the corresponding weights; $\boldsymbol{A}_\mu$ and $\boldsymbol{A}_\tau$ are fixed matrices[5]; $\boldsymbol{u}_\mu$ and $\boldsymbol{u}_\tau$ are structured random effects; and $\boldsymbol{\epsilon}_\mu$ and $\boldsymbol{\epsilon}_\tau$ are unstructured random effects which serve as model errors terms[6].

Working within the LGM framework, assign the following Gaussian prior density functions to the latent model parameters

$$\begin{aligned}
\pi(\boldsymbol{\beta}_\mu) &= \mathcal{N}(\boldsymbol{\beta}_\mu \mid \boldsymbol{\mu}_{\beta\mu}, \boldsymbol{Q}_{\beta\mu}^{-1}), \quad & \pi(\boldsymbol{\beta}_\tau) &= \mathcal{N}(\boldsymbol{\beta}_\tau \mid \boldsymbol{\mu}_{\beta\tau}, \boldsymbol{Q}_{\beta\tau}^{-1}) \\
\pi(\boldsymbol{u}_\mu) &= \mathcal{N}(\boldsymbol{u}_\mu \mid \boldsymbol{\mu}_{u\mu}, \boldsymbol{Q}_{u\mu}^{-1}), \quad & \pi(\boldsymbol{u}_\tau) &= \mathcal{N}(\boldsymbol{u}_\tau \mid \boldsymbol{\mu}_{u\tau}, \boldsymbol{Q}_{u\tau}^{-1}) \\
\pi(\boldsymbol{\epsilon}_\mu) &= \mathcal{N}(\boldsymbol{\epsilon}_\mu \mid \boldsymbol{0}, \boldsymbol{Q}_{\epsilon\mu}^{-1}), & \pi(\boldsymbol{\epsilon}_\tau) &= \mathcal{N}(\boldsymbol{\epsilon}_\tau \mid \boldsymbol{0}, \boldsymbol{Q}_{\epsilon\tau}^{-1})
\end{aligned} \tag{4.3}$$

[5] For example, spline matrices or projection matrices.

[6] Adding the unstructured random effects is reasonable in many cases from a statistical modeling point of view as they serve as error terms for the latent model. Small variances can be imposed *a priori* on the unstructured random effects $\boldsymbol{\epsilon}_\mu$ and $\boldsymbol{\epsilon}_\tau$ if they are not desired in the model. See **Paper II** for further discussion.

where the parameters of the prior density functions can potentially depend on a set of hyperparameters $\boldsymbol{\theta}$, and $\boldsymbol{Q}_{\epsilon\mu}^{-1}$ and $\boldsymbol{Q}_{\epsilon\tau}^{-1}$ are diagonal matrices.

The latent model structure in equation (4.2) and the prior distributions given by the equations in (4.3) can be written in an equivalent joint matrix form, which is arguably more convenient to work with from a mathematical standpoint[7]. Thus, define the following vectors and matrices

$$\boldsymbol{\eta} = \left(\boldsymbol{\mu}^{\mathsf{T}}, \boldsymbol{\tau}^{\mathsf{T}}\right)^{\mathsf{T}}, \quad \boldsymbol{\nu} = \left(\boldsymbol{\beta}_{\mu}^{\mathsf{T}}, \boldsymbol{u}_{\mu}^{\mathsf{T}}, \boldsymbol{\beta}_{\tau}^{\mathsf{T}}, \boldsymbol{u}_{\tau}^{\mathsf{T}}\right)^{\mathsf{T}}, \quad \boldsymbol{\epsilon} = \left(\boldsymbol{\epsilon}_{\mu}^{\mathsf{T}}, \boldsymbol{\epsilon}_{\tau}^{\mathsf{T}}\right)^{\mathsf{T}}$$

and

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X}_{\mu} & \boldsymbol{A}_{\mu} & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{X}_{\tau} & \boldsymbol{A}_{\tau} \end{pmatrix}, \quad \boldsymbol{Q}_{\epsilon} = \begin{pmatrix} \boldsymbol{Q}_{\epsilon\mu} & \cdot \\ \cdot & \boldsymbol{Q}_{\epsilon\tau} \end{pmatrix}.$$

Next, group the following parameters and matrices together and define

$$\boldsymbol{\mu}_{\nu} = \begin{pmatrix} \boldsymbol{\mu}_{\beta\mu} \\ \boldsymbol{\mu}_{u\mu} \\ \boldsymbol{\mu}_{\beta\tau} \\ \boldsymbol{\mu}_{u\tau} \end{pmatrix}, \quad \boldsymbol{Q}_{\nu} = \begin{pmatrix} \boldsymbol{Q}_{\mu\beta} & \cdot & \cdot & \cdot \\ \cdot & \boldsymbol{Q}_{u\mu} & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{Q}_{\tau\beta} & \cdot \\ \cdot & \cdot & \cdot & \boldsymbol{Q}_{u\tau} \end{pmatrix}$$

where the dotted entries denote zero entries. The additive model structure implied by (4.2) for both the location and log-scale parameter can then be jointly expressed in the following equivalent matrix form

$$\boldsymbol{\eta} = \boldsymbol{Z}\boldsymbol{\nu} + \boldsymbol{\epsilon}, \tag{4.4}$$

and the Gaussian prior assumptions in (4.3) can be jointly expressed in following equivalent joint distributional form

$$\begin{aligned} \pi\left(\boldsymbol{\eta} \mid \boldsymbol{\nu}\right) &= \mathcal{N}\left(\boldsymbol{\eta} \mid \boldsymbol{Z}\boldsymbol{\nu}, \boldsymbol{Q}_{\epsilon}^{-1}\right), \\ \pi\left(\boldsymbol{\nu}\right) &= \mathcal{N}\left(\boldsymbol{\nu} \mid \boldsymbol{\mu}_{\nu}, \boldsymbol{Q}_{\nu}^{-1}\right). \end{aligned} \tag{4.5}$$

As the data density function and the corresponding parameters are arbitrarily chosen, analogous model setup can be carried out for any parametric data density function and any of its parameters. For example, in addition to imposing latent Gaussian models on the location and log-scale parameters of the g.e.v. distribution, a latent Gaussian model can also be imposed on the shape parameter, see both **Paper II** and **Paper III** for such model setups. Therefore, equations (4.4) and (4.5) are general in the sense that most of LGMs used in practice can be expressed in the same form. The model setup implied by equations (4.4) and (4.5) is therefore adopted as a general setup for the latent model structures for LGM in this thesis.

To see how the LGMs setup implied by (4.4) and (4.5) relates to the general LGM setup in Section 2.1.2, consider the following. Assume the vector $\boldsymbol{\nu}$ depends

[7] In fact, this joint matrix form will become essential for exploring some of the mathematical properties offered by the extended LGM framework. This is explored in great details in **Paper II** in the derivation of the MCMC split sampler.

on a vector of hyperparameters $\boldsymbol{\theta}$. By using the Gaussian prior assumptions in (4.5), it can be shown that the joint distribution of $\boldsymbol{\eta}, \boldsymbol{\nu}|\boldsymbol{\theta}$, becomes

$$\pi\left(\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix}\middle|\boldsymbol{\theta}\right) = \mathcal{N}\left(\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix}\middle|\begin{pmatrix}\boldsymbol{Z}\boldsymbol{\mu}_\nu\\ \boldsymbol{\mu}_\nu\end{pmatrix}, \begin{pmatrix}\boldsymbol{Q}_\epsilon & -\boldsymbol{Q}_\epsilon\boldsymbol{Z}\\ -\boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon & \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}\end{pmatrix}^{-1}\right) \tag{4.6}$$

see Appendix II.B in **Paper II** for proof. As the vector $\boldsymbol{x} = (\boldsymbol{\eta}^\mathsf{T}, \boldsymbol{\nu}^\mathsf{T})^\mathsf{T}$ is Gaussian, which follows immediately from (4.6), it can be viewed as the latent parameter $\boldsymbol{x}$ in the general LGM setup in Section 2.1.2.

4.2 The MCMC split sampler

As discussed in Section 4.1 and in **Paper I**, the extended LGM framework is a flexible modeling class for various statistical applications. However, LGMs become increasingly more computationally challenging to infer in practice when latent models are desired for more than just the mean structure of the data density function; when the number of parameters associated with the latent model increase; or when the data density function is non-Gaussian.

One of the main goals of this thesis is to propose a novel computationally efficient sampling based inference scheme for LGMs that addresses the aforementioned computational issues, which is the main research objective of **Paper II**. The proposed sampling scheme in **Paper II** is an MCMC strategy which is referred to as *the MCMC split sampler* in this thesis. The MCMC split sampler is designed to handle LGMs where latent models are imposed on more than just the mean structure of the data density; to scale well in terms of computational efficiency when the dimensions of the latent models increase; and to be applicable for any choice of a parametric data density function. The main novelty of the MCMC split sampler lies in how the model parameters of a LGM are split into two blocks, such that one of the blocks exploits the latent Gaussian structure in a natural way and becomes invariant of the data density function. A brief overview of the MCMC split sampler is given in this section.

4.2.1 The block design

The MCMC split sampler is a two block Gibbs sampling scheme, which is designed to exploit the model structure of LGMs. The principle idea behind the MCMC split sampler is to split the latent Gaussian parameters $\boldsymbol{x}$ into two vectors, such that $\boldsymbol{x} = (\boldsymbol{\eta}^\mathsf{T}, \boldsymbol{\nu}^\mathsf{T})^\mathsf{T}$, where $\boldsymbol{\eta}$ consists of elements that appear in the data density function and $\boldsymbol{\nu}$ consists of elements that do not appear in it. Thus, the data $\boldsymbol{y}$ become conditionally independent of the parameters $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ conditioned on the parameter $\boldsymbol{\eta}$, that is,

$$\pi(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) = \pi(\boldsymbol{y} \mid \boldsymbol{\eta}). \tag{4.7}$$

In the LGM setup implied by equations (4.4) and (4.5), which is hereby adopted as the setup for the extended LGM framework in this thesis, both $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$ are

designed such that the equality in equation (4.7) holds[8]. The parameters $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$ are referred to as the data-rich and data-poor components of the latent field, respectively, in this thesis.

The block sampling scheme of the MCMC split sampler consists of grouping the model parameters $\boldsymbol{\eta}$, $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ into two blocks. That is, $\boldsymbol{\eta}$ is placed in a block referred to as the *data-rich* block in this thesis, while both $\boldsymbol{\nu}$ and the hyperparameters $\boldsymbol{\theta}$ are placed in another block referred to as the *data-poor* block. A Gibbs sampling strategy is then implemented for each block, conditioned on the other block. That is, the $(k+1)$-th MCMC sample from the posterior density $\pi(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y})$ is obtained by using the following two block Gibbs sampling scheme

Data-rich block: sample $\boldsymbol{\eta}^{k+1}$ from $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)$

Data-poor block: sample $(\boldsymbol{\nu}^{k+1}, \boldsymbol{\theta}^{k+1})$ jointly from $\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})$

To elaborate on the potential computational efficiency gain offered by this blocking scheme, consider the following. In many practical applications of LGMs with a non-Gaussian data density function, especially in spatial statistics, the data-rich component of the latent field $\boldsymbol{\eta}$ is of relatively low dimension compared to the dimension of the data-poor component of the latent field $\boldsymbol{\nu}$. However, due to the non-Gaussianity of the data density function, the conditional posterior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ in the data-rich block can become computationally challenging in posterior simulations. By design, the data-rich block contains a minimum number of parameters which are needed for the evaluation of the potentially computationally demanding conditional posterior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$. In other words, the dimensions of the potentially computationally demanding data-rich block is designed to be as low as possible.

An additional key part of the design is the following. Due to the splitting of the latent field expressed in equation (4.7) the parameters $(\boldsymbol{\nu}, \boldsymbol{\theta})$ in the data-poor block become conditionally independent of $\boldsymbol{y}$ conditioned on the vector $\boldsymbol{\eta}$ from the data-rich block, that is,

$$\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta}) = \pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{\eta}).$$

In other words, the form of the conditional posterior $\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta})$ becomes invariant of the data density function. The resulting conditional posterior density function in the data-poor block is thus proportional to

$$\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta}). \tag{4.8}$$

Furthermore, the conditional posterior density of $\boldsymbol{\nu}$ conditioned on the parameter vector $\boldsymbol{\eta}$ and the hyperparameters $\boldsymbol{\theta}$ becomes Gaussian and is independent of the data $\boldsymbol{y}$, that is,

$$\begin{aligned} \pi(\boldsymbol{\nu} \mid \boldsymbol{y}, \boldsymbol{\eta}, \boldsymbol{\theta}) &= \pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta}) \\ &= \mathcal{N}\left(\boldsymbol{\nu} \middle| \boldsymbol{Q}_{\nu|\eta}^{-1}\left(\boldsymbol{Q}_\nu \boldsymbol{\mu}_\nu + \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\epsilon \boldsymbol{\eta}\right), \boldsymbol{Q}_{\nu|\eta}^{-1}\right) \end{aligned} \tag{4.9}$$

[8] To give a demonstration, in the context of the example setup in Section 4.1.1, the latent parameter $\boldsymbol{x}$ is then split into $\boldsymbol{x} = (\boldsymbol{\eta}, \boldsymbol{\nu})$ with $\boldsymbol{\eta} = \left(\boldsymbol{\mu}^\mathsf{T}, \boldsymbol{\tau}^\mathsf{T}\right)^\mathsf{T}$ and $\boldsymbol{\nu} = \left(\boldsymbol{\beta}_\mu^\mathsf{T}, \boldsymbol{u}_\mu^\mathsf{T}, \boldsymbol{\beta}_\tau^\mathsf{T}, \boldsymbol{u}_\tau^\mathsf{T}\right)^\mathsf{T}$.

where $\boldsymbol{Q}_{\nu|\eta} = \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}$. See Appendix II.B in **Paper II** for proof. Due to the Gaussianity offered by equation (4.9), it is possible to obtain exact samples from $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$ in every MCMC iterations. Consequentially, the proposed sampling scheme scales well, in terms of computational efficiency, when the dimension of $\boldsymbol{\nu}$ increases.

The MCMC split sampler is designed such that, in principle, any efficient MCMC sampler can be implemented for each block. In this thesis, computationally efficient sampling strategies tailored to the particular conditional model structure of each block are proposed. These strategies are described in the subsequent sections.

4.2.2 Data-rich block

The conditional posterior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ in the data-rich block is often only known up to a constant. Thus, a Metropolis–Hastings algorithm with an independence proposal density[9] is proposed for the data-rich block. An efficient proposal density is constructed in this section by approximating the conditional posterior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ with a Gaussian approximation evaluated at the mode of conditional posterior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$, denoted with $\boldsymbol{\eta}^0$. It can be shown that the Gaussian approximation of the conditional posterior becomes

$$\tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{\eta} \mid \boldsymbol{\eta}^0, (\boldsymbol{Q}_\epsilon - \boldsymbol{H})^{-1}\right) \tag{4.10}$$

where $\boldsymbol{H} = \nabla^2 \log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$. Due to the potentially low dimension of the data-rich block, the calculation of the Gaussian approximation in equation (4.10) can be reasonably fast within every MCMC iteration.

Once a new $\boldsymbol{\eta}^*$ is proposed in the $(k+1)$-th iteration of the MCMC chain with the independence proposal density $q(\boldsymbol{\eta}^*) = \tilde{\pi}(\boldsymbol{\eta}^* \mid \boldsymbol{y}, \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)$, it is accepted with the following acceptance probability

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{\eta}^* \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})}{\pi(\boldsymbol{\eta}^k \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})} \cdot \frac{q(\boldsymbol{\eta}^k)}{q(\boldsymbol{\eta}^*)}\right\}. \tag{4.11}$$

The logarithm of the acceptance ratio given in (4.11) can be simplified to

$$r = f(\boldsymbol{\eta}^*) - \left(\frac{1}{2}(\boldsymbol{\eta}^*)^\mathsf{T}\boldsymbol{H} + \boldsymbol{b}^\mathsf{T}\right)\boldsymbol{\eta}^* - f(\boldsymbol{\eta}^k) + \left(\frac{1}{2}(\boldsymbol{\eta}^k)^\mathsf{T}\boldsymbol{H} + \boldsymbol{b}^\mathsf{T}\right)\boldsymbol{\eta}^k \tag{4.12}$$

where $\boldsymbol{b} = \nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H}\boldsymbol{\eta}^0$, see Appendix II.B in **Paper II**. As the gradient $\nabla f(\boldsymbol{\eta}^0)$ and Hessian $\boldsymbol{H}$ have already been calculated to obtain (4.10), the expression in (4.11) can be computed with low computational cost in every iteration of the MCMC chain.

In many applications conditional independence assumptions are imposed on the data density function. That is, there exists a partition of $\boldsymbol{\eta}$ into subvectors $\boldsymbol{\eta}_i$, such that

$$\pi(\boldsymbol{y} \mid \boldsymbol{\eta}) = \prod_i \pi_i(\boldsymbol{y}_i \mid \boldsymbol{\eta}_i).$$

[9] That is, in the $(k+1)$-th iteration the proposal density is invariant of $\boldsymbol{\eta}^k$, such that $q(\boldsymbol{\eta}^* \mid \boldsymbol{\eta}^k) = q(\boldsymbol{\eta}^*)$.

In some cases, a proposal density based on the Gaussian approximation in (4.10) can be a poor approximation of the conditional posterior density in some partition of $\boldsymbol{\eta}$. Updating the whole vector $\boldsymbol{\eta}$ in one block may then result in the MCMC chain getting stuck, thus leading to lower computational efficiency. In order to circumvent this issue and to retain the computational speed gained by using the Gaussian approximation in (4.10) as a proposal density, a modification can be made to the sampling scheme which utilizes the conditional independence of the partitions within the data-rich block. The details on the modification can be seen in Appendix II.A in **Paper II**.

The proposed sampling scheme for the data-poor block is outlined in Algorithm 3 in **Paper II**. Note that by choosing $I = 1$ in Algorithm 3, the above sampling scheme without the conditional independence assumptions on the likelihood is obtained, while selecting $I \geq 2$ in Algorithm 3 assumes the aforementioned partitioning of $\boldsymbol{\eta}$ and that each $\boldsymbol{\eta}_i$ is accepted or rejected separately.

4.2.3 Data-poor block

The relation in (4.8) and the Gaussianity of the conditional posterior density $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$ in (4.9) demonstrate that the one-block updating scheme of Knorr-Held and Rue (2002), discussed in Section 2.2.6, is applicable within the data-poor block with some modifications[10]. As discussed in Section 2.2.6, the one-block updating scheme is known to outperform most modern sampling algorithms in terms of efficiency when applicable. Therefore, the following modified version of the one-block updating scheme, adapted for the data-poor block of the MCMC split sampler, is proposed.

For some proposal density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^k)$ for the hyperparameters $\boldsymbol{\theta}$, a new proposed value $(\boldsymbol{\nu}^*, \boldsymbol{\theta}^*)$ is generated jointly as follows:

$$\begin{aligned} \boldsymbol{\theta}^* &\sim q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k) \\ \boldsymbol{\nu}^* &\sim \pi(\boldsymbol{\nu}^*|\boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^*). \end{aligned} \tag{4.13}$$

Denote the proposal density implied by (4.13) with $q(\boldsymbol{\nu}^*, \boldsymbol{\theta}^*|\boldsymbol{\nu}^k, \boldsymbol{\theta}^k)$. The proposed value $(\boldsymbol{\nu}^*, \boldsymbol{\theta}^*)$ is then accepted jointly with acceptance probability

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})} \frac{q(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{\nu}^*, \boldsymbol{\theta}^*)}{q(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)}\right\}. \tag{4.14}$$

It can be shown, see **Paper II**, that the acceptance ratio in (4.14) can be simplified to

$$\frac{\pi(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})} \frac{q(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{\nu}^*, \boldsymbol{\theta}^*)}{q(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)} = \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^{k+1})} \frac{q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)} \tag{4.15}$$

[10]The author would like to add that this fact is the original motivation of the MCMC split sampler. That is, to find a subset of the model parameter space for which the one-block sampling scheme is applicable.

The relation in (4.15) shows that the acceptance ratio is only dependent on the acceptance ratio for $\boldsymbol{\theta}$. Further, since the conditional posterior $\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta})$ is a known Gaussian the proposed sampling strategy scales well in terms of computational efficiency as the dimensions of the data-poor component of the latent field, that is $\boldsymbol{\nu}$, increases.

If the Gaussian models in the prior assumptions (4.5) are GMRF density functions with a sparse precision structure, the ratio in (4.15) is computationally costly to calculate directly, since $\pi(\boldsymbol{\theta} \mid \boldsymbol{\eta}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta})$ and $\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta})$ does not necessarily preserve the sparse GMRF structure. However, as the ratio in (4.15) is only dependent on the acceptance ratio for $\boldsymbol{\theta}$ it can be shown that the ratio in (4.15) can be rewritten in order to preserve the sparse GMRF precision structure. That is, in **Paper II** it is shown that the term $\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^{k+1})/\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^{k+1})$ in equation (4.15) equates to

$$\frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^{k+1})} = \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^k)} \cdot \frac{\pi(\boldsymbol{\eta}^{k+1} \mid \mathbf{0}, \boldsymbol{\theta}^*)\pi(\mathbf{0} \mid \boldsymbol{\theta}^*)}{\pi(\mathbf{0} \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^*)} \cdot \frac{\pi(\mathbf{0} \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^k)}{\pi(\boldsymbol{\eta}^{k+1} \mid \mathbf{0}, \boldsymbol{\theta}^k)\pi(\mathbf{0} \mid \boldsymbol{\theta}^k)} \tag{4.16}$$

where $\boldsymbol{\nu}$ is set as the $\mathbf{0}$ vector for computational reasons and as the ratio in (4.16) is independent of $\boldsymbol{\nu}$, see Theorem II.5 in Paper **Paper II** for further details. By Theorem 4 in Section 2.3.3, all of the conditional posterior density function on the right hand side of equation (4.16) are GMRF density function with a sparse precision structure, which is discussed in Section 2.3.3. Additionally, Theorem II.5 in **Paper II** shows the form of the conditional posterior densities on the right hand side in (4.16) on a logarithmic scale.

The results in (4.15) and (4.16) show how the ratio in (4.14) can be calculated with low computational cost. This is a key result for the implementation of the proposed sampling scheme in the data-poor block for GMRFs with sparse precision structures. The algorithm for the sampling scheme in the data-poor block is summarized in Algorithm 4 in **Paper II**.

4.2.4 Examples

Two examples are presented in Section II.3 in **Paper II**, where the MCMC split sampler is applied as an inferential algorithm to two different LGMs with different model structures. In the former example, a data set on annual precipitation in Iceland is analyzed. The latter example is on simulated extreme flood events. To emphasize, the aim of the examples in **Paper II** is to present some of the possibilities offered by the MCMC split sampler as an flexible and efficient inferential algorithm, rather than to claim which model is the best for each data set. The main purpose of the first example, see Section II.3.1 in **Paper II**, is to demonstrate that the MCMC split sampler is well suited to infer LGMs with spatial models imposed on both location and scale parameters of the data density function. Furthermore, the goal is to demonstrate that the computational efficiency of the MCMC split sampler scales well as the dimension of the spatial model, which belongs to the data-poor component of the latent field, increases. The main goal of the latter example, see Section II.3.2 in **Paper II**, is to show

that the MCMC split sampler is well suited to infer LGMs with a non-Gaussian, three parameter data density function, where all the three parameters of the data density are modeled with latent Gaussian models.

4.3 Spatial modeling of extreme precipitation on a fine grid

One of the main goals in this research project, outlined in Chapter 3, is to present a flexible modeling framework for statistical spatial analysis of extreme precipitation. As previously discussed in Section 2.4, there is a need for a statistical approach for modeling observed extreme precipitation as meteorological models for precipitation tend to underestimate observed extreme precipitation. An example of this underestimation is given in Section 4.3.2, where observations on 24-hour extreme precipitation in Iceland are compared to the simulated values of 24-hour extreme precipitation based on the model of (Crochet et al. 2007). Furthermore, the information on the distributional properties of extreme precipitation is needed at a fine spatial resolution as the characteristics of extreme precipitation events can be local, especially in regions with heterogeneous topography.

The research goal of **Paper III** is thus to develop a statistical methodology to model the distributional properties of extreme precipitation on a fine grid. The statistical model in **Paper III**, is developed by using the statistical modeling methodology introduced in the previous sections of this thesis. To elaborate on this point, the modeling approach in **Paper III** is line with the authors' stance towards statistical modeling presented in Section 4.1. That is, an LGM with a g.e.v data density, discussed in Section 2.4.2, is constructed where latent models are imposed on the location, scale and shape parameter, as discussed in Section 4.1.1. Furthermore, the underlying spatial variation is modeled with a GF with a Matérn covariance function which is approximated with a finite dimensional GMRF representation constructed with the SPDE approach[11], as presented in Section 2.3. Finally, the MCMC split sampler is used as an inferential algorithm, which was introduced in Section 4.2. The proposed modeling strategy is general in the sense that it is extendable to any spatial domain of interest.

Furthermore, the aim of the research is also to provide information about the characteristics of extreme precipitation at a fine spatial resolution, in particular, where limited spatial information is available due to a relatively sparse network of observational sites. To address the sparsity of the observational sites, a method is proposed in **Paper III** that leverages information from an external local meteorological model that provides spatial information on mean precipitation on a fine grid. The leveraged information is thus based on scientific knowledge on the physical processes of precipitation and not directly on observations on extreme precipitation. This is a novel extension of the concept presented in Benestad et al.

[11] Although LGMs with a GMRF structure have been proposed to model extreme precipitation (Schliep et al. 2010, Cooley and Sain 2010), LGMs with spatial models based on the SPDE approach have not been proposed in the literature on spatial extreme precipitation before.

(2012), which states that observed mean precipitation provides information on observed extreme precipitation. To demonstrate the use of the proposed method for a particular region, which is Iceland in the analysis in this thesis, information is extracted from a meteorological model proposed by Smith and Barstad (2004) which Crochet et al. (2007) adapted to a 1 km by 1 km grid over Iceland. The proposed method is extendable to any regions in the world where outputs from a local meteorological models are available

4.3.1 The data and visual inspection

The observed data on precipitation were provided by the Icelandic Meteorological Office (IMO). The data set contains observations on accumulated 24-hour annual maximum precipitation from 40 observational sites in Iceland over the years 1958 to 2006, as seen in (Crochet et al. 2007). The locations of the observational sites can be seen in Figure 4.1. The observations have been corrected according to a dynamic correction method proposed by Førland and Hanssen-Bauer (2000), which Crochet et al. (2007) adapted to the Icelandic data set. The correction method accounts for trace; wetting and evaporation losses; and for the catch deficiencies due to aerodynamic effects and is applied to daily precipitation observations from every observational site.

Time series from four observational sites, Reykjavík, Æðey, Akureyri, and Kvísker, can be seen in Figure III.2 in **Paper III**. Reykjavík and Akureyri were chosen because they are the most populated areas in Iceland. Æðey was chosen due to its geological position. Finally, Kvísker was chosen as it has the highest observed precipitation. Furthermore, the geographical locations of the observational sites in Reykjavík, Æðey, Akureyri, and Kvísker can also be seen in **Paper III**.

Boxplots of the observed 24-hour annual maximum precipitation is given in Figure 4.2 for all of the observational sites. The observational sites are placed on the x-axis as follows. The leftmost site on the x-axis is in Reykjavík. The rest of sites are placed on the x-axis corresponding to a clockwise labeling across the sites shown in Figure 4.2. For further clarification of the arrangement, corresponding observational sites in Figure 4.1 and Figure 4.2 share the same color labeling. Visual inspection of the boxplots in Figure 4.2 reveals that observational sites which are located close together in space exhibit similar observed distributional properties. That is, both the observed sample median and interquartile range of the 24-hour annual maximum precipitation, show evidence of spatial dependence between neighboring observational site. This observational fact is used as a motivation for the spatial statistical modeling presented in Section 4.3.4.

4.3.2 The meteorological model

The meteorological model which is used in this thesis to leverage spatial information on precipitation, is based on a linear orographic precipitation model proposed by Smith and Barstad (2004). Crochet et al. (2007) have adapted the method to precipitation in Iceland. The model is driven by coarse resolution

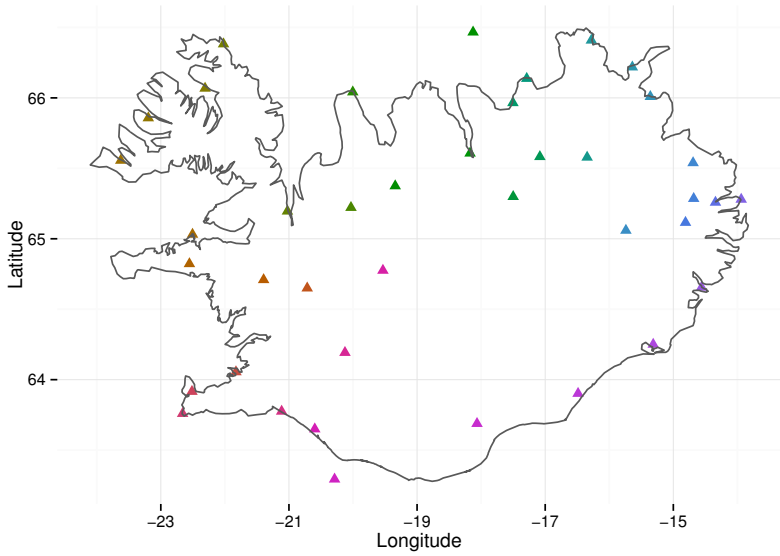

Figure 4.1: The figure shows the locations of the observational sites in Iceland.

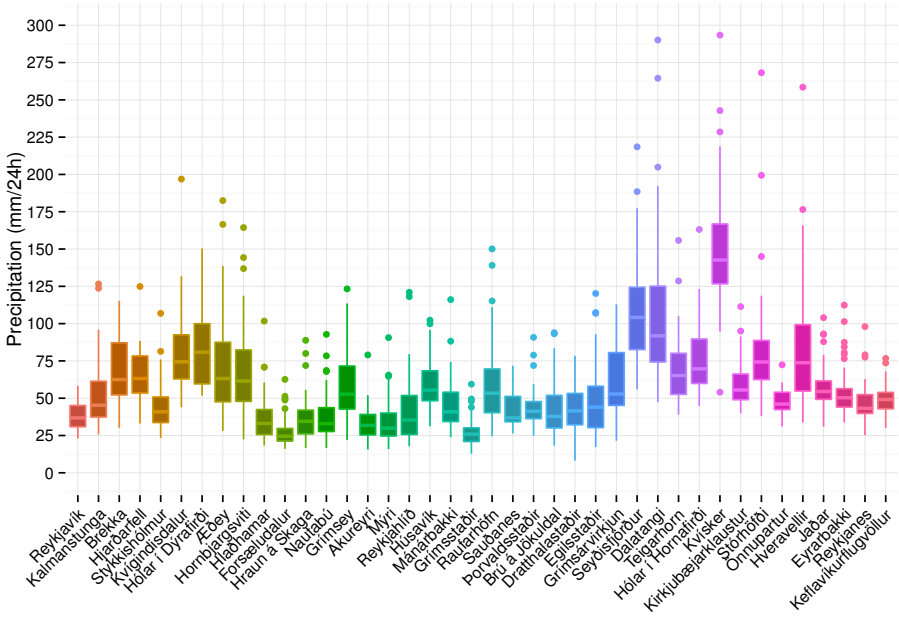

Figure 4.2: The figure shows boxplots of observed 24-hour annual maximum precipitation at every observational site.

precipitation, wind and temperature data obtained from re-analyses (1958-2001) (ERA-40) (Uppala et al. 2005) and analyses made by the European Center for Medium Range Weather Forecast (2002-2004). The model takes into account the topography of the spatial domain; airflow dynamics; condensed water advection; and downslope evaporation. This means that outputs from the model contain information about the underlying physical processes of precipitation.

The meteorological model of Crochet et al. (2007), hereby referred to as the meteorological model, is a physical model mainly driven by the scientific knowledge on the physical processes of precipitation. However, the meteorological model has three free parameters that need to be calibrated. The calibration made by Crochet et al. (2007) is mainly based on an extensive data set on glacier mass balance observations in Iceland at the glaciers Vatnajökull, Hofsjökull and Langjökull (Sigurdsson et al. 2004). The data on glacier mass balance are not used in the statistical modeling of annual maximal precipitation proposed in this thesis, but are only used to calibrate the three free parameters of the meteorological model. Furthermore, Crochet et al. (2007) also used daily observations of precipitation over the period 1995-2000 as complementary source of information for the calibration of the free parameters. The resulting model simulates daily precipitation on a 1 km by 1 km regular grid, of the size 521 km $\times$ 361 km, across Iceland for the years 1958-2002.

Although the meteorological model has been shown to adequately capture the observed mean behavior of precipitation across all observational sites, see the Crochet et al. (2007) for results, the model tends to underestimate extreme precipitation, in particular, 24-hour annual maximum precipitation. The underestimation between the observational sites ranges from being relatively minor to significantly large. As a demonstration of this underestimation, Figure 4.3 shows the empirical cumulative distributions of both the simulated maximum precipitation based on the meteorological model, and of the observed maximum precipitation at the same locations. The comparison is made for the observational sites at Reykjavík, Æðey, Akureyri and Kvísker. Figure 4.3 demonstrates that the meteorological model yields moderate underestimates of maximum precipitation in Reykjavík and Akureyri. However, Figure 4.3 shows significant discrepancies between the model values and the observed values at both Æðey and Kvísker. As Kvísker is known to have the highest observed maximum precipitation in Iceland, see for example the boxplot of the observations in Figure 4.2, it becomes of particular interest for the statistical modeling of maximum precipitation in Iceland.

4.3.3 Leveraging spatial information on precipitation

Covariates which assimilate available spatial information, such as the topography of the domain and the underlying physical processes of precipitation, have been shown to be reasonable for statistical modeling of extreme precipitation, see Cooley et al. (2007). By constructing covariates based on the outputs from the meteorological model, the information about the above factors can be assimilated. Moreover, Benestad et al. (2012) suggested that observed mean values

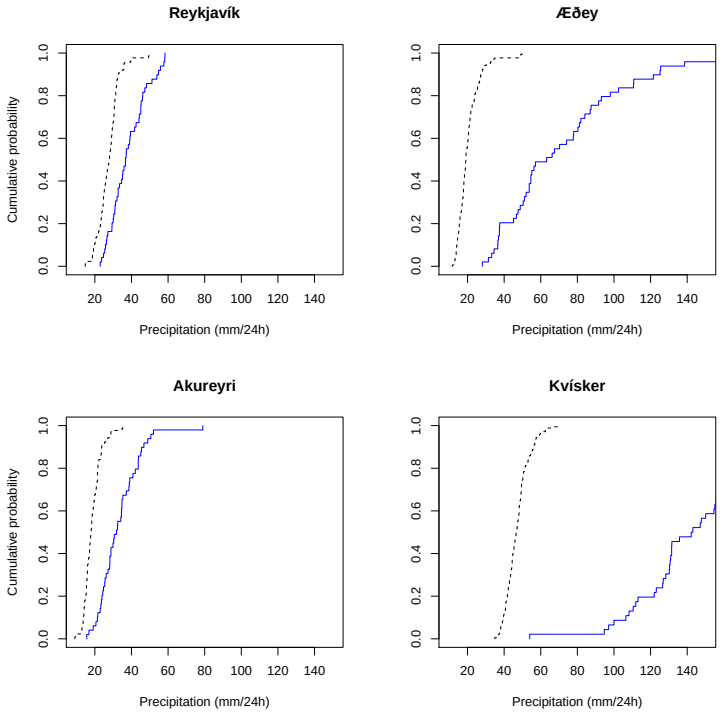

Figure 4.3: The black dashed curves show the empirical cumulative distribution of the simulated values annual maximum 24-hour precipitation based on the meteorological model. The blue solid curves show the empirical cumulative distribution of the observed annual maximum 24-hour precipitation at the corresponding observational sites.

of precipitation have high predictive power for maximum precipitation. As the meteorological model has been shown to capture the observed mean behavior of precipitation across all observational sites well, see Crochet et al. (2007), the argument made in Benestad et al. (2012) can be extended to the outputs of the meteorological model, assuming the outputs describe the mean behavior of precipitation reasonably well in every regular grid point. As such, the authors propose that calculated sample mean values based the outputs of the meteorological model serve as reliable predictors for extreme precipitation at every regular grid point. The sample means based on the meteorological model are referred to hereafter as the meteorological covariates[12]. Furthermore, for notational convenience let $\mathcal{G}$ denote the set of every regular grid point in the 1 km by 1 km grid and let $\mathcal{S}$ denote the set of the observational sites.

To emphasize, the information stemming from the meteorological model is

[12]However, as observational sites are not necessarily at the regular grid points a spatial smother was implemented in order construct the covariates at the observational sites. See Section III.A.1 in **Paper III** for details on the implementation of the spatial smoother.

primarily based on scientific knowledge on the physical process of precipitation. Also note that the meteorological model of Crochet et al. (2007) used six years worth of data on observed daily precipitation and observed monthly means, not observed annual 24 hour maximum precipitation. The resulting constructed covariate is thus related to the observed mean of the time period. Based on the argument made by Benestad et al. (2012) and the quality of the outputs from the meteorological model in terms of mean precipitation, the constructed covariate serves to describe the spatial variation of climate, in terms of precipitation, which is then relevant for describing extreme precipitation.

4.3.4 Statistical model

The formulation of the proposed LGM for the statistical modeling annual maximum 24-hour precipitation is as follows.

The data level: As the observed data on annual maximum 24 hour precipitation are based on extreme precipitation events, the generalized extreme value distribution, discussed in Section 2.4.1, is implemented where the observations are assumed to be conditionally independent. That is, let y_{it} denote the annual maximum 24 hour precipitation at station i at year t, with a cumulative density function of the form

$$F(y_{it}) = \exp\left\{-\left(1+\xi_i\left(\frac{y_{it}-\mu_i}{\sigma_i}\right)\right)^{-1/\xi_i}\right\}, \quad i=1,\ldots,I, \quad t=1,\ldots,T$$

if $1+\xi_i(x-\mu_i)/\sigma_i > 0$, $F(y_{it}) = 0$ otherwise. The parameters μ_i, σ_i and ξ_i are location, scale and shape parameters; I is the number of observational sites; and T is the number of years.

The latent level: The motivation for the proposed latent model for the location parameter is as follows. The boxplots in Figure 4.2 demonstrate that observed maximum precipitation at observational sites which are close together in space exhibit similar observed distributional properties. In particular, this observational fact yields evidence for spatially dependent medians of observed maximum precipitation. Furthermore, preliminary statistical analysis revealed a linear relationship between the ML estimates of the location parameter μ and the meteorological covariate at the observational sites, see Figure III.4 in **Paper II**.

Therefore, the following model structure is implemented for the location parameter $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_I)^\mathsf{T}$ at the latent level of the model,

$$\boldsymbol{\mu} = \boldsymbol{X}_\mu\boldsymbol{\beta}_\mu + \boldsymbol{A}_\mathcal{S}\boldsymbol{u}_\mu + \boldsymbol{\epsilon}_\mu. \tag{4.17}$$

The matrix $\boldsymbol{X}_\mu$ in (4.17) is a design matrix consisting of a vector of ones and the meteorological covariate, where $\boldsymbol{\beta}_\mu$ denotes the corresponding weights. Furthermore, $\boldsymbol{u}_\mu$ in (4.17) denotes a finite GMRF representation of a Gaussian field with a Matérn covariance function, constructed with the SPDE approach in Theorem 5, on a triangulated mesh over the spatial domain, shown in Figure 4.4. The

precision matrix of $\boldsymbol{u}_\mu$ given in Theorem 5, denoted with $\boldsymbol{Q}_{u\mu}$ in the proposed LGM, has two parameters, $\kappa_{u\mu}$ and $\omega_{u\mu}$, which enter the proposed LGM as hyperparameters. The hyperparameter $\kappa_{u\mu}$ is inversely proportional to the range of the Gaussian field corresponding to $\boldsymbol{u}_\mu$ and the hyperparameter $\omega_{u\mu}$ is related to the marginal variance of the spatial effect $\boldsymbol{u}_\mu$, see Appendix III.A.2 of **Paper III** for details. The parameter α in Theorem 5 is chosen as one, which corresponds to an almost once differentiable Gaussian field with a Matérn covariance function. The matrix $\boldsymbol{A}_\mathcal{S}$ in (4.17) is a fixed projection matrix which describes the convex linear interpolation of the finite GMRF representation located at the vertices of the mesh onto the location of the observational sites. As each observational site is located in the interior of only one triangle, the number of non-zero entries in each row of $\boldsymbol{A}_\mathcal{S}$ is three and the sum of every row is one. The matrix product $\boldsymbol{A}_\mathcal{S}\boldsymbol{u}_\mu$ in (4.17) thus denotes the spatial effect at the observational sites, which serves to captures the spatial variation in the data that is unexplained by the meteorological covariate. Finally, the parameter $\boldsymbol{\epsilon}_\mu$ in (4.17), which serves as a model error term, is an unstructured random effect with a variance $\sigma^2_{\epsilon\mu}$. The variance $\sigma^2_{\epsilon\mu}$ enters the proposed LGM as a hyperparameter.

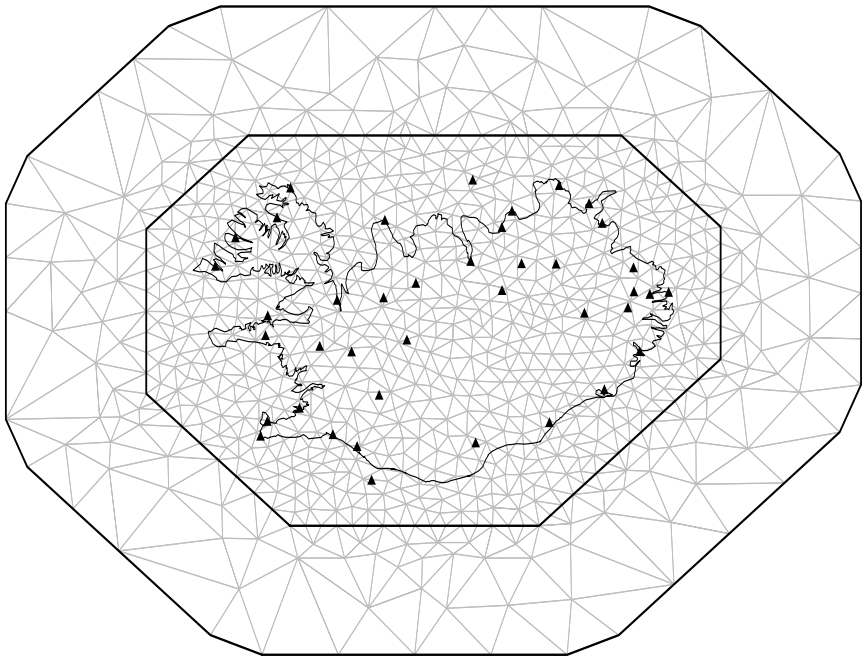

Figure 4.4: Triangulated mesh over Iceland.

Similar to the observed medians, the boxplots in Figure 4.2 show evidence for spatially dependent scale structure of the observed maximum precipitation. Additionally, the preliminary statistical analysis also revealed a linear relationship between the ML estimates of the scale parameter σ and the meteorological covariate on a logarithmic scale, see Figure III.4 in **Paper III**. Therefore, an analogous

model structure was implemented for the scale parameter on a logarithmic scale. That is, let $\tau_i = \log \sigma_i$ and then model $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_I)^\mathsf{T}$ as

$$\boldsymbol{\tau} = \boldsymbol{X}_\tau \boldsymbol{\beta}_\tau + \boldsymbol{A}_\mathcal{S} \boldsymbol{u}_\tau + \boldsymbol{\epsilon}_\tau.$$

The design matrix $\boldsymbol{X}_\tau$ consists of a vector of ones and the meteorological covariate on a logarithmic scale; $\boldsymbol{u}_\tau$ is a spatial random effect with the same SPDE structure as $\boldsymbol{u}_\mu$ and has parameters $\kappa_{u\tau}$ and $\omega_{u\tau}$; and $\boldsymbol{\epsilon}_\tau$ is an unstructured random effect with a variance $\sigma^2_{\epsilon\tau}$.

The boxplots in Figure 4.2 show that the right tail behavior of maximum precipitation is similar between observational sites close together in space. Additionally, the preliminary statistical analysis revealed that ML estimates for the shape parameter ξ vary among the observational sites. This evidence might indicate spatially dependent tail structure of maximum precipitation. However, no evidence was seen for a linear relationship between the ML estimates of ξ and the meteorological covariate or other available geological covariates ,see Figure 4.2 in **Paper III**. Thus, the following spatial model was implemented for the shape parameter

$$\boldsymbol{\xi} = \boldsymbol{X}_\xi \beta_\xi + \boldsymbol{A}_\mathcal{S} \boldsymbol{u}_\xi + \boldsymbol{\epsilon}_\xi$$

where $\boldsymbol{X}_\xi$ is vector of ones; the weight β_ξ serves as an overall effect for the shape parameter over the spatial domain; $\boldsymbol{u}_\xi$ is a spatial random effect that captures the spatial deviation from the overall effect with the same SPDE structure as $\boldsymbol{u}_\mu$ with parameters $\kappa_{u\xi}$ and $\omega_{u\xi}$; and $\boldsymbol{\epsilon}_\xi$ is an unstructured random effect with variance $\sigma^2_{\epsilon\xi}$.

Prior density functions: Working within the LGM setup, the following prior density functions were assigned to parameters at the latent level of the proposed model.

$$\begin{aligned}
&\pi(\boldsymbol{\beta}_\mu) = \mathcal{N}(\boldsymbol{\beta}_\mu \mid \mathbf{0}, \sigma^2_{\beta\mu}\boldsymbol{I}), \quad \pi(\boldsymbol{\beta}_\tau) = \mathcal{N}(\boldsymbol{\beta}_\tau \mid \mathbf{0}, \sigma^2_{\beta\tau}\boldsymbol{I}), \quad \pi(\beta_\xi) = \mathcal{N}(\beta_\xi \mid \mathbf{0}, \sigma^2_{\beta\xi}),\\
&\pi(\boldsymbol{u}_\mu) = \mathcal{N}(\boldsymbol{u}_\mu \mid \mathbf{0}, \boldsymbol{Q}^{-1}_{u\mu}), \quad \pi(\boldsymbol{u}_\tau) = \mathcal{N}(\boldsymbol{u}_\tau \mid \mathbf{0}, \boldsymbol{Q}^{-1}_{u\tau}), \quad \pi(\boldsymbol{u}_\xi) = \mathcal{N}(\boldsymbol{u}_\xi \mid \mathbf{0}, \boldsymbol{Q}^{-1}_{u\xi})\\
&\pi(\boldsymbol{\epsilon}_\mu) = \mathcal{N}(\boldsymbol{\epsilon}_\mu \mid \mathbf{0}, \sigma^2_{\epsilon\mu}\boldsymbol{I}), \quad \pi(\boldsymbol{\epsilon}_\tau) = \mathcal{N}(\boldsymbol{\epsilon}_\tau \mid \mathbf{0}, \sigma^2_{\epsilon\tau}\boldsymbol{I}), \quad \pi(\boldsymbol{\epsilon}_\xi) = \mathcal{N}(\boldsymbol{\epsilon}_\xi \mid \mathbf{0}, \sigma^2_{\epsilon\xi}\boldsymbol{I})
\end{aligned}$$

The parameters $\boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\tau$ and β_ξ are assumed *a priori* to have a low precision on their native scales in order to let the data play the dominate role in their inference. Thus, the parameter values $\sigma_{\beta\mu} = 20$, $\sigma_{\beta\tau} = 4$ and $\sigma_{\beta\xi} = 2$ were chosen for the prior distributions.

The hyper level: Let $\boldsymbol{\theta}$ denote all the hyperparameters of the model that are not fixed, that is

$$\boldsymbol{\theta} = (\kappa_{u\mu}, \omega_{u\mu}, \sigma_{\epsilon\mu}, \kappa_{u\tau}, \omega_{u\tau}, \sigma_{\epsilon\tau}, \kappa_{u\xi}, \omega_{u\xi}, \sigma_{\epsilon\xi}). \tag{4.18}$$

Lognormal prior distributions with fixed parameters were assigned to the hyperparmeters in $\boldsymbol{\theta}$. The selection of the prior distributions is further discussed in Section III.3.2 and Appendix III.A.2 in **Paper III**.

4.3.5 Model setup for the MCMC split sampler

In order to implement the MCMC split sampler as an inferential algorithm for the proposed LGM in Section 4.3.4, the model parameters are first assigned into the two block design of the MCMC split sampler. The data-rich component of latent field, which is $\boldsymbol{\eta} = (\boldsymbol{\mu}^\mathsf{T}, \boldsymbol{\tau}^\mathsf{T}, \boldsymbol{\xi}^\mathsf{T})^\mathsf{T}$, is assigned into the data-rich block. The data-poor component of the latent field, which consists of $\boldsymbol{\nu} = (\boldsymbol{\beta}_\mu^\mathsf{T}, \boldsymbol{u}_\mu^\mathsf{T}, \boldsymbol{\beta}_\tau^\mathsf{T}, \boldsymbol{u}_\tau^\mathsf{T}, \beta_\xi^\mathsf{T}, \boldsymbol{u}_\xi^\mathsf{T})^\mathsf{T}$, is are assigned to data-poor block along with the hyperparameter vector $\boldsymbol{\theta}$, which is given in equation (4.18).

To adapt the formulation of the LGM in Section 4.3.4 to the model setup of the MCMC split sampler, define the following sparse matrices

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X}_\mu & \boldsymbol{A}_\mathcal{S} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{X}_\tau & \boldsymbol{A}_\mathcal{S} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \boldsymbol{X}_\xi & \boldsymbol{A}_\mathcal{S} \end{pmatrix}, \quad \boldsymbol{Q}_\epsilon = \begin{pmatrix} \sigma_{\epsilon\mu}^{-2}\boldsymbol{I} & \cdot & \cdot \\ \cdot & \sigma_{\epsilon\tau}^{-2}\boldsymbol{I} & \cdot \\ \cdot & \cdot & \sigma_{\epsilon\xi}^{-2}\boldsymbol{I} \end{pmatrix}.$$

Furthermore let prior mean of $\boldsymbol{\nu}$, denoted with $\boldsymbol{\mu}_\nu$, be such that $\boldsymbol{\mu}_\nu = \boldsymbol{0}^\mathsf{T}$ and

$$\boldsymbol{Q}_\nu = \begin{pmatrix} \kappa_{\beta\mu}\boldsymbol{I} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \boldsymbol{Q}_{u\mu} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \kappa_{\beta\tau}\boldsymbol{I} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \boldsymbol{Q}_{u\tau} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \kappa_{\beta\xi}\boldsymbol{I} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \boldsymbol{Q}_{u\tau} \end{pmatrix}.$$

The latent model of the LGM in Section 4.3.4 can thus be expressed equivalently with equations (4.4) and (4.5), that is

$$\boldsymbol{\eta} = \boldsymbol{Z}\boldsymbol{\nu} + \boldsymbol{\epsilon},$$

with the prior densities written as

$$\begin{aligned} \pi(\boldsymbol{\eta} \mid \boldsymbol{\nu}) &= \mathcal{N}\left(\boldsymbol{\eta} \mid \boldsymbol{Z}\boldsymbol{\nu}, \boldsymbol{Q}_\epsilon^{-1}\right), \\ \pi(\boldsymbol{\nu}) &= \mathcal{N}\left(\boldsymbol{\nu} \mid \boldsymbol{\mu}_\nu, \boldsymbol{Q}_\nu^{-1}\right). \end{aligned}$$

For further details on the implementation of the MCMC split sampler to the proposed model, see Appendix III.A.3 in **Paper III**.

4.3.6 Continuous spatial predictions

The SPDE approach discussed in Section 2.3.4 yields continuously indexed approximate solutions to the underlying GF, which is achieved through the use of the piecewise linear basis functions in the finite element representation given in (2.21). To briefly elaborate on how this property of the SPDE approach is implemented for continuous spatial predictions in this analysis, assume that the posterior distribution of the spatially varying model parameters of the proposed LGM has been obtained. Note that, obtaining the posterior distribution of, for example the spatially varying model parameter $\boldsymbol{u}_\mu$, equates to obtaining the posterior distribution of the Gaussian weights in equation (2.21) for the finite GMRF

representation of the GF for the location parameter. By construction, every regular grid points in $\mathcal{G}$ belongs to the interior of only one triangle in the mesh. Consequently, by using the piecewise linear basis functions in (2.21), MCMC samples of the spatially varying model parameters at the regular grid point in $\mathcal{G}$ can be obtained by a linear interpolation of the MCMC samples of the spatially varying model parameters located at vertices of the surrounding triangle. The linear interpolation of the k-th MCMC sample of the spatial varying model parameters is thus calculated with

$$\boldsymbol{A}_{\mathcal{G}}\boldsymbol{u}^{[k]} = \boldsymbol{u}_{\mathcal{G}}^{[k]}, \tag{4.19}$$

where the matrix $\boldsymbol{A}_{\mathcal{G}}$ describes the linear interpolation from the vertices of the mesh onto the regular grid points in $\mathcal{G}$; $\boldsymbol{u}^{[k]}$ denotes the k-th posterior MCMC sample of the spatial effects for the location, scale or shape parameters at the vertices of the mesh; and $\boldsymbol{u}_{\mathcal{G}}^{[k]}$ denotes the k-th posterior MCMC sample of the spatial effects at the regular grid points in $\mathcal{G}$. The calculations in equation (4.19) can be carried out in post calculations after the MCMC simulations.

Furthermore, as the meteorological covariates are available by construction at every regular grid point in $\mathcal{G}$, the above procedure can be implemented to obtain MCMC samples of the location, scale, and shape parameters of the data density at every grid point in $\mathcal{G}$. Consequently, the posterior distribution of every quantity of interest of the generalized extreme value can be evaluated at every regular grid point in $\mathcal{G}$. In particular, the posterior distribution of the p-th quantile of the generalized extreme value distribution can be calculated at every regular grid point $\mathcal{G}$. See Section III.3.4 in **Paper III** for further details on the corresponding computations.

5

Results

A concise overview of the main results of this Ph.D. research project is given in this chapter, with appropriate references to the papers. Figures and tables are not repeated from the papers, however, additional figures are shown to broaden the overview of the results when relevant. The results in this chapter are divided into two main categories. Firstly, in Section 5.1, the computational efficiency of the MCMC split sampler is evaluated, based on result from MCMC simulations from both **Paper II** and **Paper III**. Secondly, the results of the statistical analysis of 24-hour annual maximum precipitation carried out in **Paper III** are briefly summarized in Section 5.2.

5.1 MCMC split sampler: Evaluation of computational efficiency

The computational efficiency of the MCMC split sampler in practice is explored in this section. This is achieved by applying the MCMC split sampler as an inferential algorithm to three different LGMs. The quality of the resulting simulated MCMC chains from each model is assessed by using the posterior diagnostics tools introduced in Section 2.2.4. The first two models which are used for the evaluation of the MCMC split sampler are presented and discussed in details in Section II.3 in Paper II. The goal of presenting the first two models is to show some of the possibilities offered by the MCMC split sampler rather than to claim which model is the best for each data set. The third model in this section is the proposed statistical model for 24-hour maximum precipitation, which is presented in Section 4.3.4.

All of the following convergence diagnostics are based on four MCMC chains sampled in parallel with the MCMC split sampler as a sampling scheme. Every chain was simulated with 50000 iterations where 10000 iterations were burned in. All calculations were carried out using `R` on a modern desktop (Ivy Bridge Intel

Core i7-3770K, 16GB RAM and a solid state hard drive).

5.1.1 Annual mean precipitation in Iceland

In this example, a data set on annual mean precipitation in Iceland is modeled with an LGM that has a SPDE spatial model structure at the latent level. The data density is assumed Gaussian, where latent models are imposed both on the mean and log-variance parameters of the data density. For detailed description of the data and the model structure, see Section II.3.1 in **Paper II**. The main purpose of this example is to demonstrate that the MCMC split sampler is well suited to infer LGMs with spatial models on both location and log-variance parameters of the data density function, and that the computational efficiency of the sampler scales well as the number of vertices in the mesh increases. To that extend, the MCMC split sampler is applied as an inferential algorithm to three versions of the proposed LGM, each with spatial model based on a different mesh resolution. That is, a coarse resolution based on 411 mesh points; a medium resolution based on 858 mesh points; and a dense resolution based 1752 mesh points. In Figure II.3 in **Paper II** the three different meshes are presented on a same scale. The top, middle and bottom panels in Figure II.3 show the coarse resolution, medium resolution and dense resolution meshes, respectively

Gelman–Rubin plots, based on the MCMC simulations, of the the log-variance parameter τ at Kvísker; the covariate coefficient $\beta_{\tau 2}$; and the standard deviation of the unstructured random effect $\sigma_{\epsilon\tau}$ are shown in the first, second and third column, respectively, in Figure 5.1. The results based on the coarse resolution, medium resolution and dense resolution meshes for the aforementioned parameters are shown in the first, second and third row, respectively, in Figure 5.1. A comparison of the results in Figure 5.1 between the different mesh resolutions reveals that the convergence of the Gelman–Rubin statistic is achieved for the the three parameters at a similar rate. Furthermore, the Gelman–Rubin plots in Figure 5.1 show that the sampler has converged in the mean after roughly 7500 iterations for all mesh resolutions. Similar results hold for the mean parameter μ in Reykjavík; the covariate coefficient $\beta_{\mu 2}$; and the marginal standard deviation for the spatial field $\sigma_{u\mu}$, see Figure II.4 in **Paper II**. Comparable results also hold for all other model parameters (results not shown).

Autocorrelation plots for the same set of parameters and arranged identically as in Figure 5.1 are shown in Figure 5.2. The results demonstrate that the MCMC chains for the log-variance parameter τ in Kvísker and the covariate coefficient $\beta_{\tau 2}$ exhibit a negligible autocorrelation after lag 10. The MCMC samples of the hyperparameter $\sigma_{\epsilon\tau}$ show autocorrelation around 0.3 at lag 50. Furthermore, the autocorrelation plots of μ in Reykjavík, $\beta_{\mu 2}$ and $\sigma_{u\mu}$ given in Figure II.5 in **Paper II**, reveal a similar behavior. Comparable results hold for all the other model parameters (results not shown).

Additionally, trace plots and running mean plots for the same set of parameters which are arranged identically as in Figure 5.1 are shown in Figure 5.3 and Figure 5.4, respectively. Figure 5.3 shows that the MCMC chain of the parameters does not get stuck and mixes well for all resolutions. A comparison of

the results in Figure 5.4 between the different mesh resolutions reveals that the convergence in the mean is achieved for the three parameters at a similar rate. Comparable results hold for all the other model parameters (results not shown).

Relying on the Gelman–Rubin statistics, the autocorrelation plots, the trace plots and running mean plots, the MCMC chains exhibit all signs of having converged. Moreover, the autocorrelation plots in Figure 5.2 reveal that the autocorrelation in the MCMC chains does not increase with number of vertices in the mesh, which in turn indicates that the autocorrelation in the MCMC chains is invariant of the dimension of $\boldsymbol{\nu}$, that is, the data-poor part of the latent field. These results demonstrate that the MCMC split sampler retains its computational efficiency when the number of mesh points increases.

The runtime of this model was approximately 6, 6.5 and 7 hours for the coarse, medium and dense mesh resolution, respectively. For further discussion on the results see Section II.3 in **Paper II**.

5.1.2 Flood analysis

This example is based on a simulation study on extreme flood events in Iceland. An LGM is constructed with a g.e.v. data density to describe monthly maximum instantaneous flow from 10 rivers. Latent seasonal models are imposed on the log-location, log-scale and shape parameters of the g.e.v. data density. See Section II.3.2 in **Paper II** for detailed data and model descriptions. The goal of this example is to show that the MCMC split sampler is designed to infer LGMs with a non-Gaussian three parameter data density function, where all the three parameters are modeled with latent Gaussian models.

Gelman–Rubin plots, autocorrelation plots, trace plots and running mean plots of nine model parameters based on the MCMC run are shown in Figures 5.5, Figure 5.6, Figure 5.7 and Figure 5.8, respectively. All the plots are based on the same set of parameters and arranged identically. Three parameters were chosen from the location, scale and shape structures of the proposed model which were placed in the first, second and third rows, respectively, in each figure. The first columns are based on parameters from the data-rich component of the latent field; the second columns are based on parameters from the data-poor component of the latent field; and the third column is based on hyperparameters.

The Gelman–Rubin plots in Figure 5.5 show that the sampler has converged after roughly 10.000 iteration. The autocorrelation plots in Figure 5.6 demonstrate that the MCMC chains of all nine parameters, exhibit an autocorrelation around 0.1 after lag 40. The low autocorrelation of the hyperparameters is of particular note, as the hyperparameters tend to be hard to infer in the LGM setup due to their posterior dependence on other model parameters. Additionally, Gelman–Rubin and autocorrelation plots for another set of nine parameter are shown in Figure II.8 and Figures II.9, respectively, in **Paper II**, which in turn exhibit similar behavior. The trace plots in Figure 5.7 show that the MCMC chains of all the nine parameters mix well. Figure 5.8 further demonstrates that the MCMC samples of the nine parameters have converged in the mean. Simlar results hold for all the other model parameters (results not shown).

Relying on these results, the MCMC chains exhibit all signs of having converged. Moreover, these results further indicate high efficiency at all model levels. Of particular note is that highly efficient MCMC samples of the hyperparameters are obtained.[1] Furthermore, the runtime of this model was approximately 7 hours. For more detailed discussion on the results see Section II.3.2 in **Paper II**.

5.1.3 Annual maximum precipitation in Iceland

The MCMC split sampler is applied as an inferential algorithm to the proposed model for annual maximum 24-hour precipitation, presented in Section 4.3.4. See Section 4.3.5 and the Appendix III.A.3 **Paper III** for implementation of the MCMC split sampler for this model.

Gelman–Rubin plots, autocorrelation plots, trace plots and running mean plots of nine model parameters based on the MCMC run are shown in Figures 5.9, Figure 5.10, Figure 5.11 and Figure 5.12, respectively. These four plots are all based on the same parameters and are arranged identically to the figures in Section 5.1.2. That is, the first columns are based on parameters from the data-rich component of the latent field; the second columns are based on parameters from the data-poor component of the latent field; and the third column is based on hyperparameters. As the observational site Kvísker is of particular interest in this analysis, the parameters μ, τ, and ξ at Kvísker were chosen from the data-rich component of the latent field. Furthermore, the covariate coefficients $\beta_{\mu 2}$, $\beta_{\tau 2}$ and the intercept coefficient β_{ξ} were chosen from the data-poor component of the latent field. Finally, the range parameters of the SPDE spatial models, that is $\omega_{u\mu}$, $\omega_{u\tau}$ and $\omega_{u\xi}$, were chosen from the hyperparameters.

The Gelman–Rubin plots in Figure 5.9 show that the sampler has converged after roughly 15.000 iteration. The autocorrelation plots in Figure 5.6 demonstrate that the MCMC chains for the parameters from both the data-rich and the data-poor parts of the latent field, exhibit an autocorrelation less than 0.10 after lag 15. The MCMC samples of the hyperparameter $\omega_{u\mu}$ demonstrate autocorrelation around 0.25 at lag 50, which is within an acceptable range. However, the autocorrelation of both $\omega_{u\tau}$ and $\omega_{u\xi}$ are negligible after lag 50, which demonstrates high efficiency[2]. The trace plots in Figure 5.11 demonstrate that the MCMC chains of all the nine parameters mixes reasonably well. Finally, Figure 5.8 shows that the MCMC samples of the nine parameters have converged in the mean. Comparable results hold for all the other model parameters (results not shown). Relying on these results, the MCMC chains exhibit all signs of having converged and the MCMC split sampler shows high computational efficiency. The runtime of this model was approximately 7 hours.

[1] In fact, a proposal density of Roberts et al. (1997) implied by equation (II.27) in **Paper II** is implemented in the data-poor block in this example, as opposed the proposal density (II.20) in **Paper II**.

[2] In this analysis, the proposal density of (II.26) in **Paper II** is used in the data-poor block due to its implementational simplicity. However, preliminary computational analysis reveals that using the proposal density of Roberts et al. (1997), as implemented for the model in Section 5.1.2, further improves the efficiency of the MCMC split sampler for inferring the hyperparameters of this model

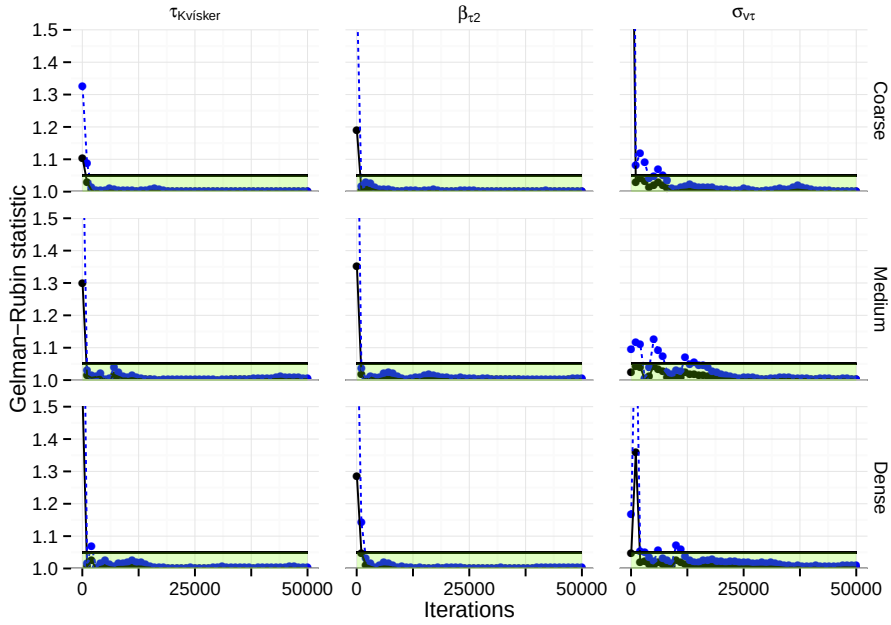

Figure 5.1: Gelman–Rubin plots based on the MCMC simulations in Section 5.1.1. The black solid curve denotes the median of the Gelman–Rubin statistics, and the blue dashed curve denotes the upper limit of the 95% confidence interval for the Gelman–Rubin statistics.

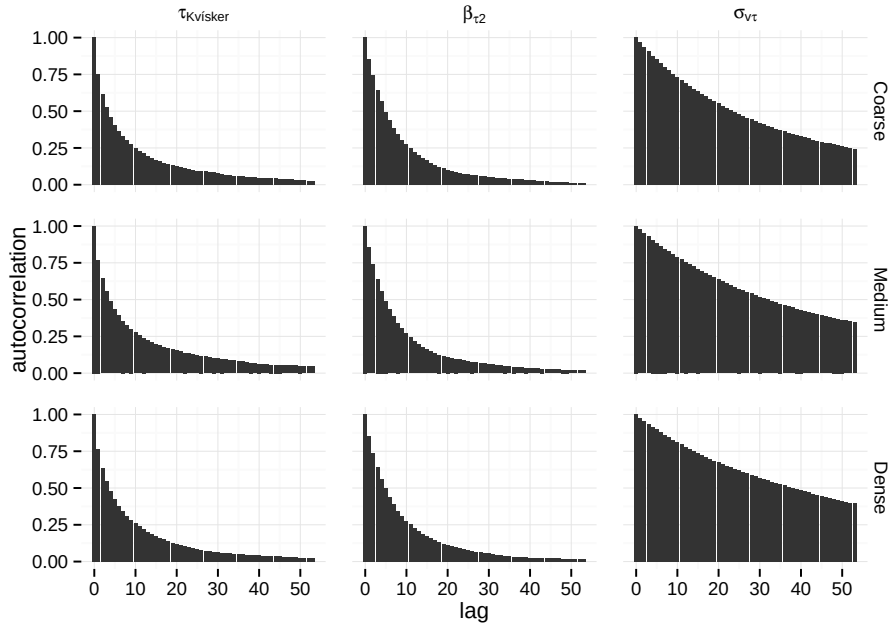

Figure 5.2: Autocorrelation plots based on the MCMC simulations in Section 5.1.1

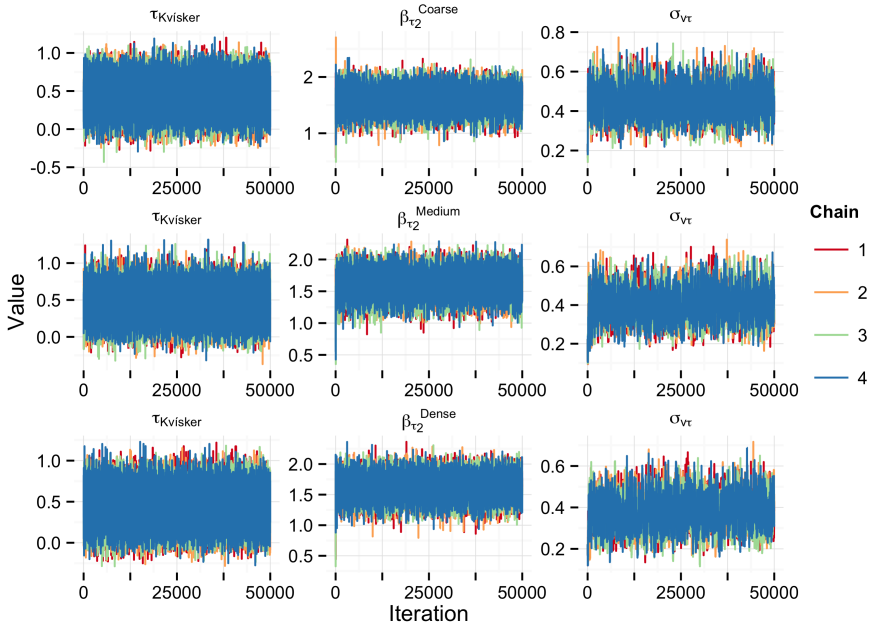

Figure 5.3: Trace plots based on the MCMC simulations in Section 5.1.1

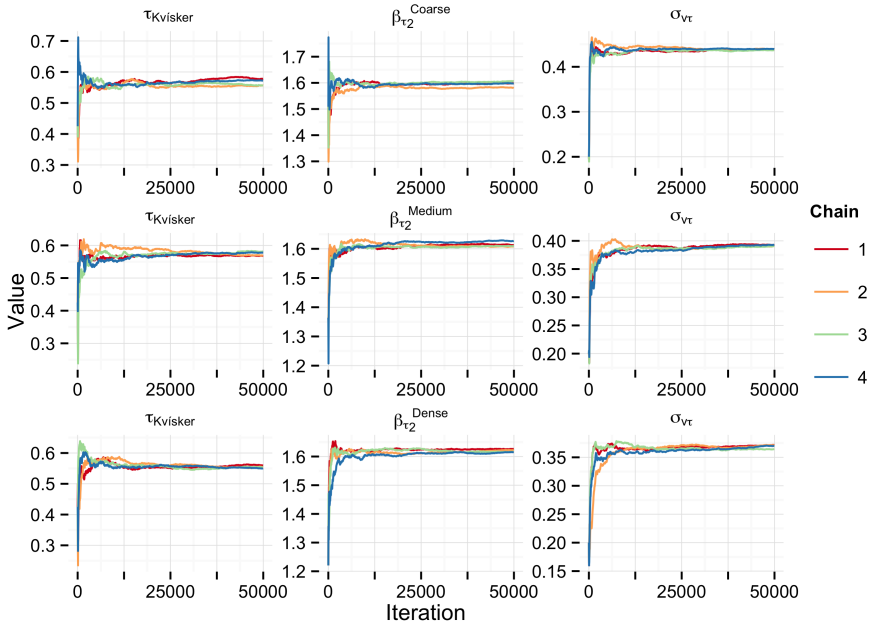

Figure 5.4: Running mean plots based on the MCMC simulations in Section 5.1.1

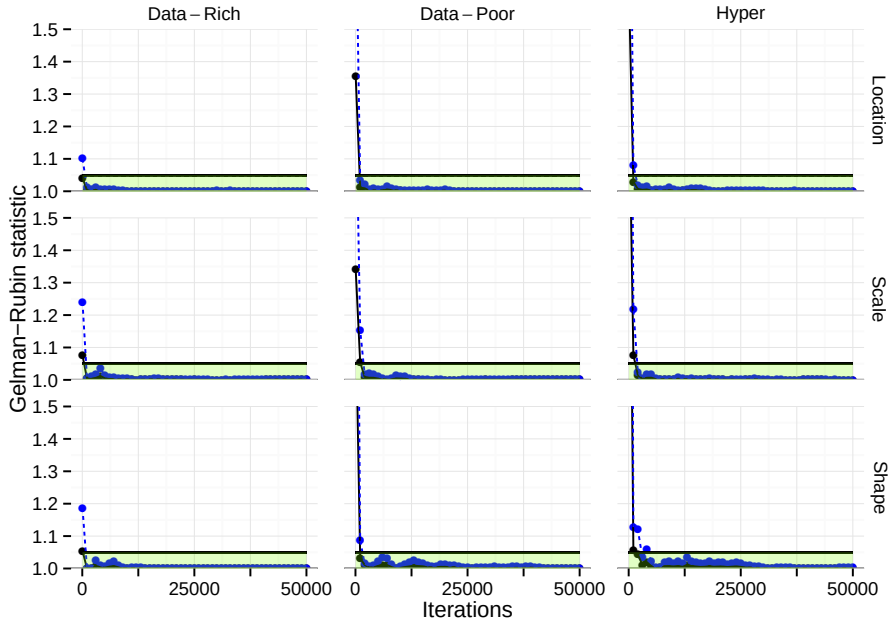

Figure 5.5: Gelman–Rubin plots based on the MCMC simulations in Section 5.1.2. The black solid curve denotes the median of the Gelman–Rubin statistics, and the blue dashed curve denotes the upper limit of the 95% confidence interval for the Gelman–Rubin statistics.

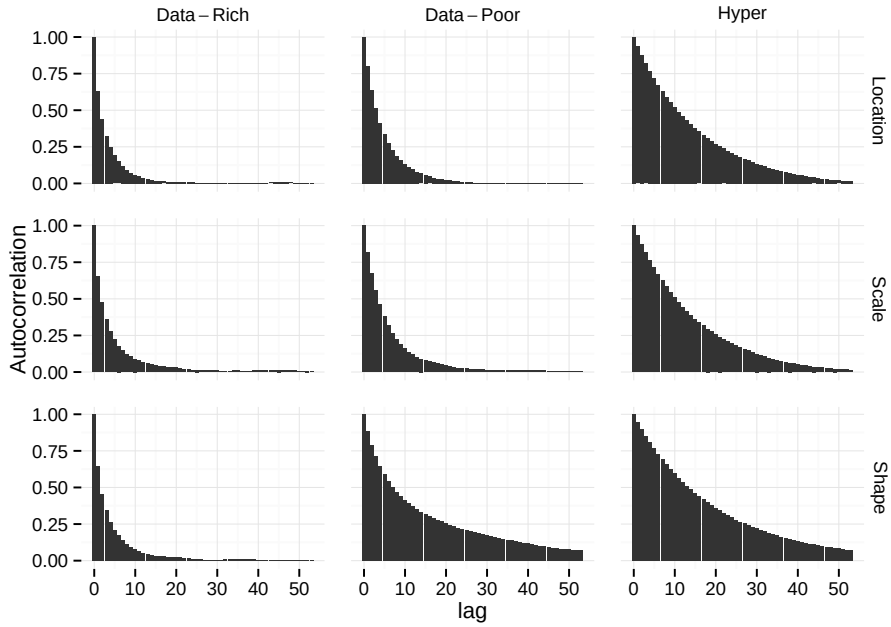

Figure 5.6: Autocorrelation plots based on the MCMC simulations in Section 5.1.2

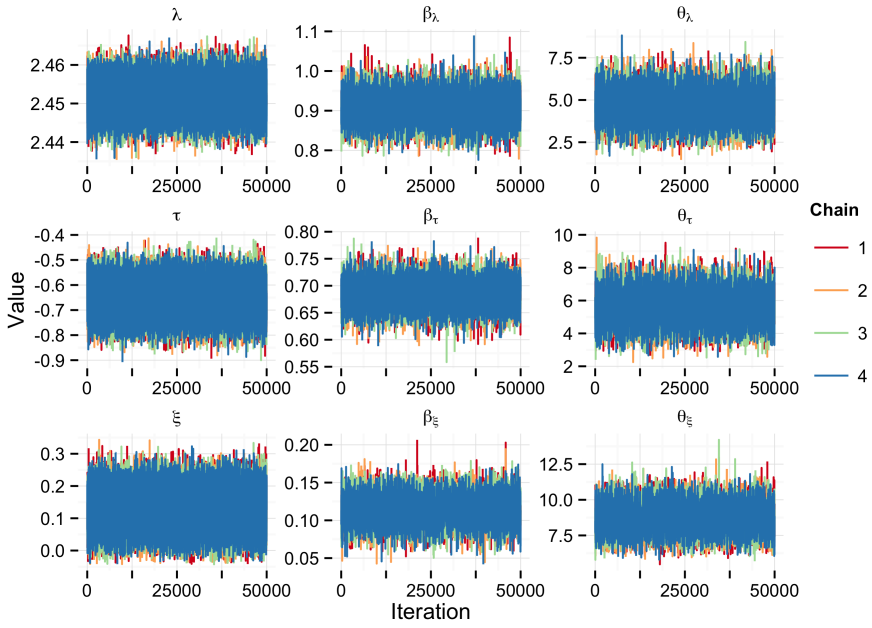

Figure 5.7: Trace plots based on the MCMC simulations in Section 5.1.2

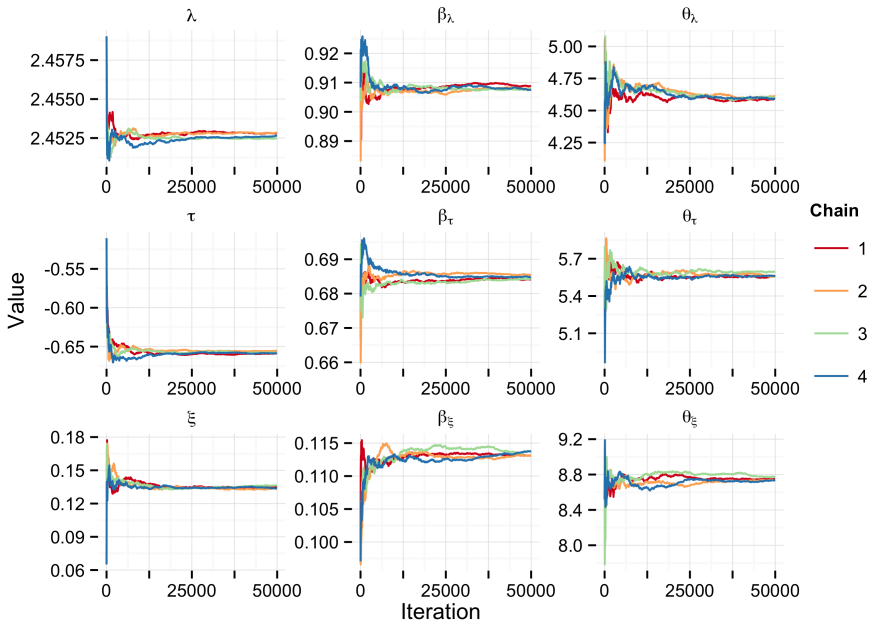

Figure 5.8: Running mean plots based on the MCMC simulations in Section 5.1.2

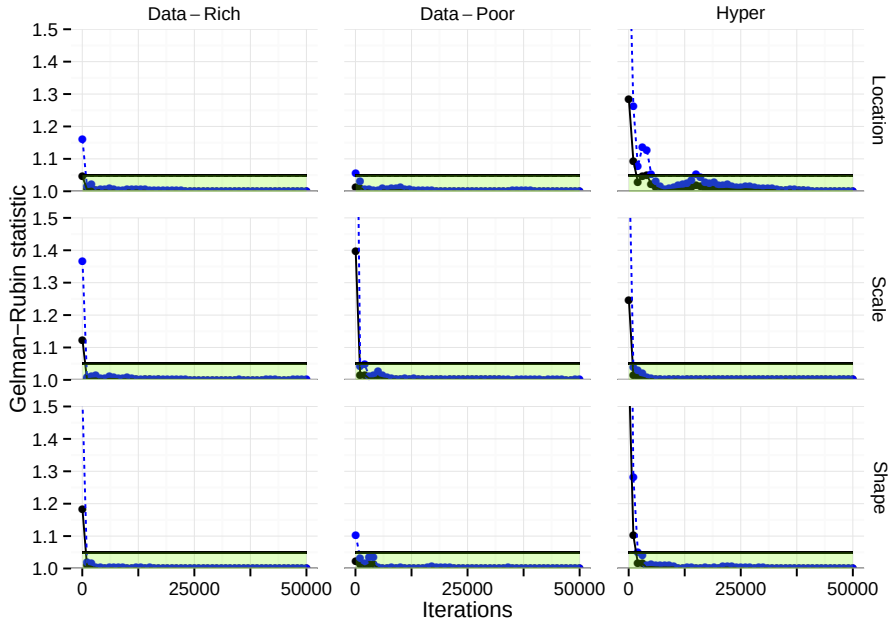

Figure 5.9: Gelman–Rubin plots based on the MCMC simulations in Section 5.1.3. The black solid curve denotes the median of the Gelman–Rubin statistics, and the blue dashed curve denotes the upper limit of the 95% confidence interval for the Gelman–Rubin statistics.

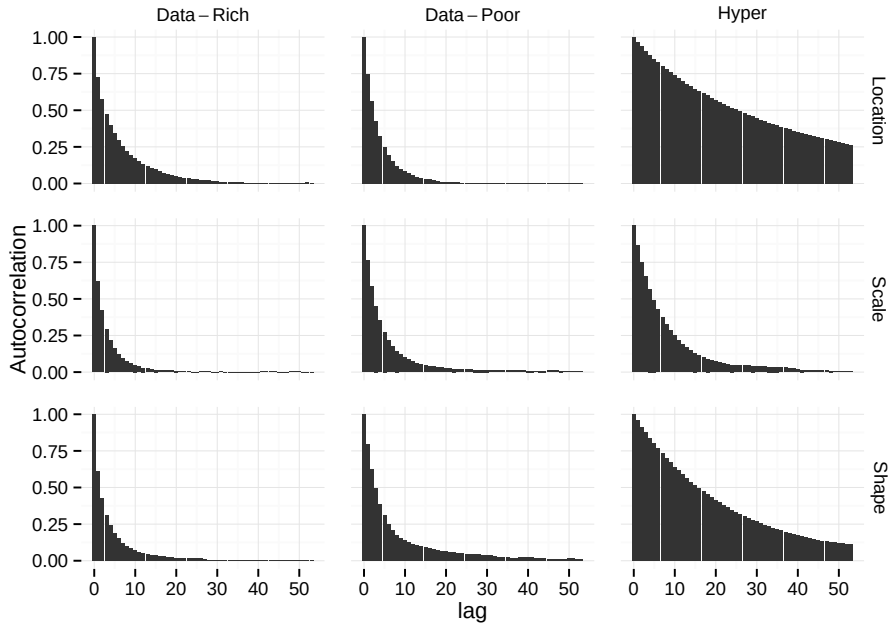

Figure 5.10: Autocorrelation plots based on the MCMC simulations in Section 5.1.3

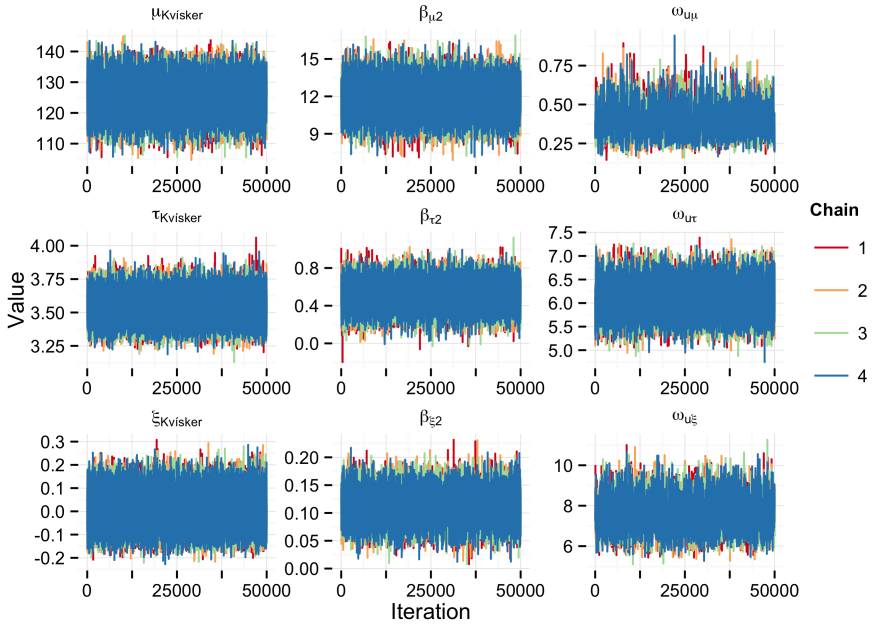

Figure 5.11: Trace plots based on the MCMC simulations in Section 5.1.3

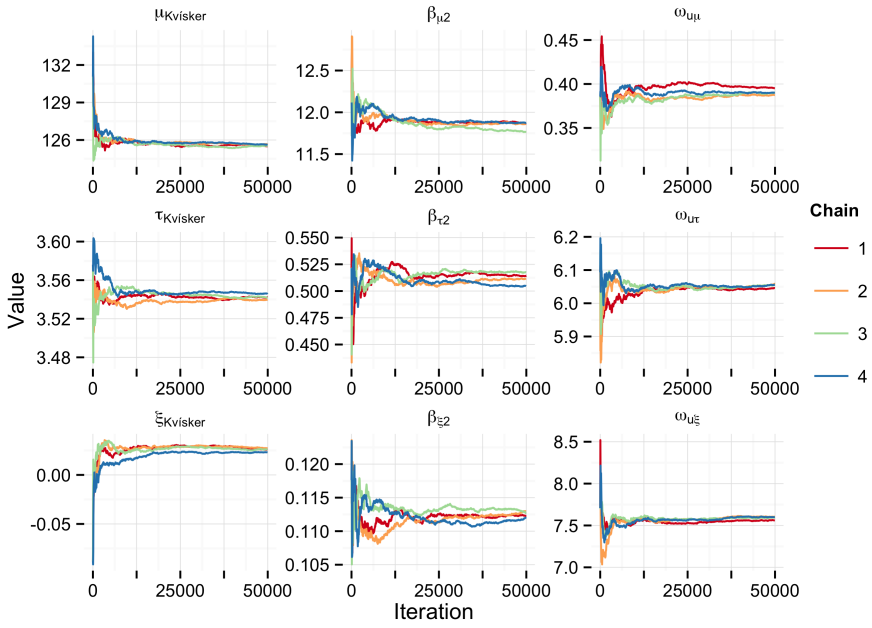

Figure 5.12: Running mean plots based on the MCMC simulations in Section 5.1.3

5.2 Annual maximum 24-hour precipitation in Iceland

A brief summary of the results from the statistical analysis of annual maximum 24-hour precipitation in Iceland based on the statistical model presented in Section 4.3 and **Paper III**, is given in this section. That is, posterior estimates of the model parameters, which are based one the four MCMC chains examined in Section 5.1.3, are evaluated and discussed. The spatial predictions of the location, the log-scale and shape parameters are then considered. Lastly, the spatial maps of the 0.90 and 0.99 quantiles based on the proposed LGM presented in Section 4.3 for annual maximum 24-hour precipitation in Iceland are examined. For a more detailed discussion on the results, see Section III.4 in **Paper III**.

5.2.1 Posterior estimates

Figure III.6 in **Paper III** shows the posterior mean, along with 95% posterior intervals, of the location, log-scale and shape parameters of the proposed model for every observational site. The top left and right panels in Figure III.6 in **Paper III** show that both the location and the scale parameters of the generalized extreme values distribution have the highest posterior estimates in the south-eastern part in Iceland, especially on the southern side of Vatnajökull Glacier. The lowest posterior estimates of the location and scale parameters are found in the northern part, and in the vicinity of Akureyri. Furthermore, the bottom left panel in Figure III.6 indicates that the posterior mean of the shape parameter is positive at all the observational sites, indicating an unbounded right tail behavior of observed maximum precipitation. However, some of the 95% posterior intervals for the shape parameter include zero.

Table III.1 in **Paper III** shows the posterior mean, posterior standard deviation and posterior 0.025 and 0.975 quantile estimates for the non-spatially varying model parameters. The results shown in Table III.1 indicate that the posterior means of $\beta_{\mu 1}$ and $\beta_{\mu 2}$ are 10.6 and 11.9, respectively, and that both parameters are most likely positive as their corresponding 95% posterior intervals are both above zero. These results indicate that the location parameter describing the extreme precipitation is roughly 12 times the value of the corresponding meteorological covariate, see **Paper III** for further discussion. Furthermore, the posterior means of $\beta_{\tau 1}$ and $\beta_{\tau 2}$ are close to 2.05 and 0.5, respectively, and both parameters are positive as their 95% posterior intervals are above zero. That means that the relationship between the meteorological covariate and the scale parameter of the generalized extreme value distribution for some observational site i can be roughly summarized as

$$\hat{\sigma}_i = e^{\hat{\beta}_{\tau 1}} \bar{x}_i^{*\hat{\beta}_{\tau 2}} \approx 7.8 \cdot \bar{x}_i^{*0.5}$$

where $\hat{\sigma}_i$ denotes a rough estimate of the scale parameter at observational site i and $\bar{x}_i^*$ denotes the value of the meteorological covariate at observational site i. The posterior mean of β_ξ is 0.12 and its posterior interval is $(0.07, 0.17)$. That

indicates that the field describing the shape parameter has mean greater than zero. This in turn indicates that the overall mean tail behavior of 24-hour annual maximum precipitation has a lower bound and is unbounded above.

The posterior mean of the parameters $\kappa_{u\mu}$, $\kappa_{u\tau}$ and $\kappa_{u\xi}$, suggest that the correlation between two points in space is near 0.1 at a roughly 40 km distance for the location parameter; roughly 25 km distance for the scale parameter; and almost 7 km for the shape parameter, which is short compared to the range of μ and τ and the scale of the spatial domain. The posterior mean of the parameters $\omega_{u\mu}$, $\omega_{u\tau}$ and $\omega_{u\xi}$ indicate that the marginal standard deviation of $\boldsymbol{u}_\mu$ is approximately 8; roughly 0.45 for $\boldsymbol{u}_\tau$; and almost 0.1 for $\boldsymbol{u}_\xi$.

For further analysis of the posterior distribution of $\sigma_{v\mu}$, $\sigma_{v\tau}$ and $\sigma_{v\mu}$, see Section III.4.2 in **Paper III**.

5.2.2 Spatial predictions

Figures III.9, III.10 and III.11 in **Paper III** show the spatial predictions for the spatially varying model parameters, namely, the the location parameter μ; the log-scale parameter τ; and the shape parameter ξ and their corresponding spatial effects. The spatial predictions are based on the method outlined in Section 4.3.6 and Section III.3.4 in **Paper III**.

The top left panel of Figure III.9 shows the spatial predictions for the spatial random effect $\boldsymbol{u}_\mu$ based on the posterior mean of (4.19) on the regular grid $\mathcal{G}$. The figure shows where the spatial random effect lowers and raises the prediction surface, therefore revealing the areas where the effect of the meteorological covariate overestimates and underestimates the extreme precipitation, respectively. The spatial prediction surface is lowered in most of the south-western part of Iceland, but is raised in the north-western and south-eastern part. The spatial random effect yields high positive values for Kvísker, which is known to have the highest observed precipitation in Iceland; negative value in the south-eastern part close to Reykjavík and values close to zero in the interior of Iceland where there are no observational sites. The top right panel of Figure III.9 shows the spatial prediction for the standard deviation of the spatial random effect $\boldsymbol{u}_\mu$ on the regular grid $\mathcal{G}$. As expected, the standard deviation increases at points further away from the observational sites, forming sinks in the standard deviation near the observational sites.

The bottom left panel of Figure III.9 shows the spatial prediction of the location parameter $\boldsymbol{\mu}_\mathcal{G}$, based on (III.6) in **Paper III**, on the regular grid $\mathcal{G}$. The figure shows that the location parameter is at its highest in the south-eastern parts of Iceland, in the vicinity of the southern side of Vatnajökull Glacier. This is to be expected, as the spatial gradient at the southern side of Vatnajökull increases rapidly moving from the nearby coastline to the top of the glacier. Due to these topographical properties and the fact that humid air blows in from the southern shoreline towards the roots of the glacier, the physical law of orographic precipitation predicts high precipitation. However, the areas north of Vatnajökull Glacier and in the middle of the country are known to be in a rain shadow by the same meteorological law. This is in line with the results seen in the bottom

left panel of Figure III.9.

Figure III.10 shows the spatial prediction for the scale parameter on a logarithmic scale. The figure is arranged in the same manner as Figure III.9. In the top left panel of Figure III.10 it can been seen that the spatial random effect $\boldsymbol{u}_\tau$ raises the spatial prediction surface in the eastern, south-eastern and north-western parts and lowers it in the south-western part. In the top right panel of Figure III.10, similar results appear for the standard deviation as for the location parameter, that is, sinks in the standard deviation form close to the observational sites. Moreover, the figure demonstrates that the spatial range for τ is shorter than for μ, as discussed above. On the bottom left panel of Figure III.10, it can be seen that the estimates for the scale parameter are highest along the south-eastern coastline.

Figure III.11 shows the spatial prediction for the shape parameter, arranged in the same manner as Figures III.9 and III.10. The top left panel of Figure III.11 shows the posterior mean of $\boldsymbol{u}_\xi$ on the regular grid $\mathcal{G}$. The magnitude of $\boldsymbol{u}_\xi$ is only 0.015 which is smaller than the variation in the ML estimates (results not shown). Furthermore, the top right panel in Figure III.11 shows that the sinks in the standard deviation of the spatial field for ξ are almost not visible. This is to be expected as the range of the spatial field for ξ is only 7 km, as discussed above. The bottom left panel of Figure III.11 demonstrates that the predicted posterior mean values of ξ are close to the posterior mean of the overall shape effect β_ξ, which is 0.117. These results indicate that a data set based on a denser network of observational sites is required to give better spatial predictions for ξ.

Spatial predictions of the 0.90 quantile and the 0.99 quantiles of 24-hour maximum precipitation in Iceland are shown in Figure 5.13. The predictions are based on the posterior mean of the 0.90 and 0.99 quantile of the g.e.v. distribution. The predictions are based on the methods discussed in Section 4.3.6 and Section III.3.4 in **Paper III**. The 0.90 and 0.99 quantile can be interpreted as the 10-year and 100-year precipitation events, respectively. The results reflect the previous posterior results about the location and scale parameters. For example, the highest predicted 10-year and 100-year precipitation events are along the south and south-eastern coastlines, while the lowest predicted events are in the interior of Iceland. The maximum of the predicted values are located at the southern side of Hvannadalshnjúkur mountain and are approximately 400 mm per 24 hours for the 10-year precipitation event and 570 mm per 24 hours for the 100-year precipitation event. See Section III.4.4 for further discussion on the spatial predictions of annual maximum 24-hour precipitation in Iceland. An informative spatial covariate for ξ, if it exists, would also potentially improve the spatial predictions of the shape parameter ξ.

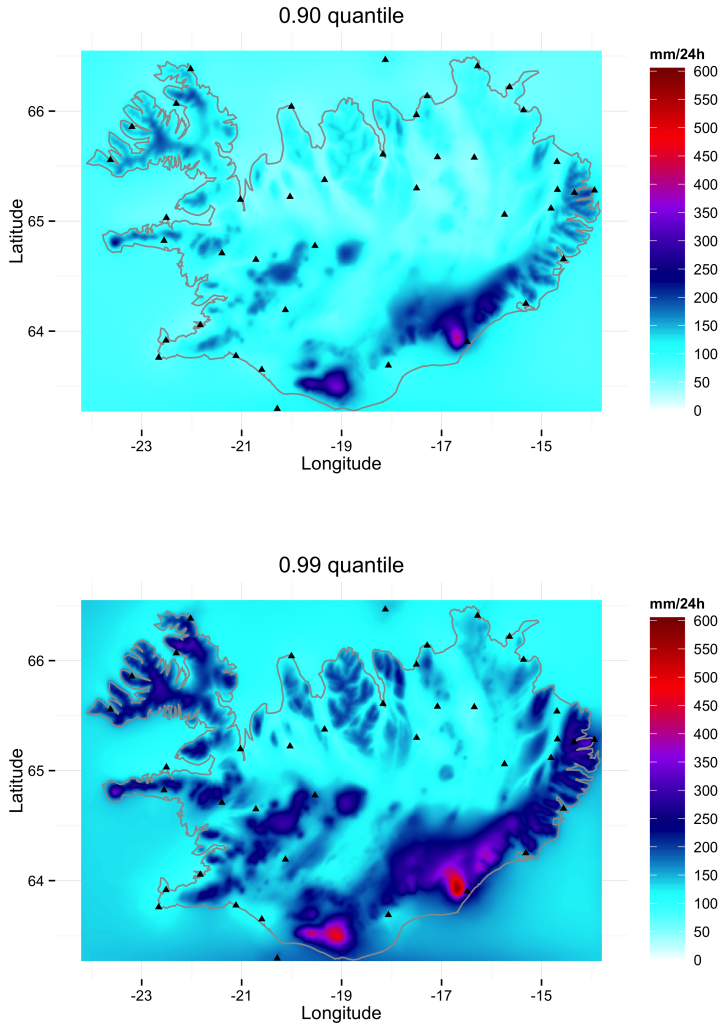

Figure 5.13: The posterior mean of the 0.90 quantile (top) and 0.99 quantile of annual maximum 24-hour precipitation assuming the generalized extreme value distribution

6

Conclusions and future perspective

Various aspects of computationally efficient Bayesian statistical methodology, with focus on applications to spatial extremes, have been explored in this thesis. To summarize the proposed methods and the main novel contributions in this thesis, the conclusions of the research project are divided into the following categories. In Section 6.1, a brief overview is given of the efficient and interpretable modeling approach offered by the extended LGM framework[1]. Section 6.2 discusses the novel inferential algorithm, the MCMC split sampler, developed in this thesis for the extended LGM framework. Furthermore, Section 6.2 discusses potential future improvements and extensions to the MCMC split sampler. Section 6.3 examines the proposed method that leverages spatial information on precipitation by using an external meteorological model as an additional source of information. Future improvements and extension to the proposed methodology are also discussed. In Section 6.4, the advantages of using the SPDE spatial models in the statistical modeling of spatial extremes are discussed and possible extensions to the proposed modeling framework. The chapter concludes in Section 6.5, where a discussion is given on how the combinations of the proposed methodology can be extended to other areas of future research.

6.1 Computationally efficient Bayesian modeling

The statistical model constructed in Section 4.3.4, and the statistical models presented in Section II.3 in **Paper II**, demonstrate that the LGM framework offers an interpretable, flexible and consistent approach to constructing statistical models. To elaborate on this point, the interpretability of LGMs stems mainly from

[1] Referred to as LGM-2.0 in **Paper I**

the analytical properties of the of the Gaussian distribution. This is particularly true in a spatial setting, where the latent Gaussian parameters of an LGM can be thought of as a finite representation of an underlying GF with a certain predefined covariance structure. For instance, in the model in Section 4.3.4, both the range and the marginal variance of the underlying GF serve as hyperparameters within the LGM setup and have thus a predefined interpretation.

To give examples of the flexibility offered by LGMs, consider the following two properties of LGMs. First, in a spatial setting, LGMs can be used to give spatial predictions at unobserved locations by utilizing the closure property of the Gaussian distribution in Theorem 1, as discussed in detail in Section 2.3.2. Second, consider the boxplots given in Figure 4.2. The boxplots indicate that both the location and scale structure of the data distribution vary among the observational sites. This behavior can be modeled within the flexible LGM framework by imposing latent Gaussian models on both the location and scale parameters. As such, LGMs of this type can be inferred using the MCMC split sampler, introduced in Section 4.2. See Section 6.2 for further discussion on this point.

The demonstration in Section 4.1.1 shows how LGMs can be constructed in a consistent manner. This general approach to constructing an LGM is further demonstrated by the statistical model for annual maximum 24-hour precipitation in Iceland, as derived in Section 4.3.4. The LGMs in Section II.3.1 and Section II.3.2 in **Paper II** serve as further demonstration of this general consistent approach to constructing LGMs. The LGM framework offers a modular approach for statistical modeling. That is, first a data density is selected, then the latent behavior of the data density parameters is modeled with Gaussian models. To allow for more flexibility of the overall model, hyperparameters are included in the model structure.

To reduce the computational cost of LGMs with a GF at the latent level, the GF can be represented with a GMRF. As discussed in Section 2.3.3, the GMRF parameterization can offer significant computational advantage over the standard GF parameterization. To further increase the practicality of the GMRFs in the LGM framework, especially in a spatial setting, the newly developed SPDE approach, of Lindgren et al. (2011) introduced in Section 2.3.4, provides a consistent method to utilize the computational advantages offered by GMRFs, without sacrificing flexibility of a predefined covariance structure. Serving as an example of model structure, is the model introduced in Section 4.3.4 and **Paper III**, which demonstrates that this framework is well suited for computationally efficient modeling of spatial extremes.

6.2 Computationally efficient inferential algorithm for LGMs

As previously discussed in Section 2.2.6, the one block updating scheme of Knorr-Held and Rue (2002) is known to be a highly computationally efficient simulation based inferential algorithm for LGMs. Moreover, the computational cost of every iteration of the one block updating scheme can be further reduced if the LGM is

parameterized with a GMRF, by using fast sampling algorithms designed for GMRFs, see Rue (2001) and Rue and Held (2005). However, the one-block updating scheme is only applicable to LGMs where the data density is assumed Gaussian. Different versions of the one block sampling scheme have been proposed, see Rue and Held (2005), however in the authors experience those algorithms often result in the MCMC chain getting stuck, as discussed briefly in **Paper II**. Furthermore, although INLA offers a very fast and accurate approach to posterior inference of LGMs, it is only applicable[2] if the data density of each data point only depends on a single linear predictor, as discussed in Section 4.1.

The MCMC split sampler is a novel sampling based inferential algorithm which is built upon the one block updating scheme of Knorr-Held and Rue (2002) and addresses the aforementioned issues. That is, the MCMC split sampler is designed to be applicable to LGMs with any parametric data density, and is also designed for LGM where latent models are imposed on more than just the mean structure of the data density. This is achieved by finding a subset of the model parameter space of the extended LGM framework, for which the one-block sampling scheme is applicable. Namely, by separating the parameters in the data-rich block from the parameters in the data-poor block, a modified version of the one-block updating scheme becomes applicable to the data-poor block. From the author's perspective, the MCMC split sampler is therefore a natural extension of the one block updating scheme for the extended LGM framework.

The MCMC split sampler is, in principle, designed as a modular sampling scheme in the sense that any MCMC sampling scheme can be implemented within the blocks. Therefore, potential new sampling schemes for either block can be developed independently of the other block. To elaborate on some of the potential extensions to the proposed sampling schemes within both blocks, consider the following.

The proposed sampler for the data-rich block, discussed in Section 4.2.2 exhibits high computational efficiency, in terms of good mixing properties, as demonstrated by the results in Section 5.1. However, the sampler is only applicable in practice if the mode of the conditional posterior density function can be found, and can be calculated reasonably fast in every iteration of the MCMC simulations. For example, in the case of models where each observed data point has more than one unique data density parameters associated with it, say of the type $y_i \sim \pi(y_i|\mu_i, \sigma_i)$ for every measurement i, finding the mode of the conditional posterior $\pi(\mu_i, \sigma_i|y_i)$ becomes computationally impractical in some cases. Models of this type include, for example, certain spatio-temporal models (Hrafnkelsson et al. 2012). Similar computational issues also arise if data dependence at the data-level of an LGM is desired, see for example Davison et al. (2012) where t-copulas are implemented with g.e.v. marginal density functions at the data level as a model for spatial extremes.

In the aforementioned cases, the two block design of the MCMC split sampler offers a platform for the development of a potentially efficient inferential algorithm. Consider, for example, an LGM with with a t-copula based data density function. By using the two block design of the MCMC split sampler, the computa-

[2]In its current state.

tionally costly data-rich part, which includes calculations of the computationally costly t-copula based data density, is separated from the data-poor part. To tackle the data-rich block, a Riemann manifold MALA based Metropolis–Hastings algorithm is, for example, applicable. This can be achieved without changing the structure of the data-poor block, which in turn means that the modified version of the one-block sampler can still be used in the data-poor block in this case.

In the data-poor block, it is shown in Section 4.2.3 that the conditional posterior density $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$ is a known Gaussian density and the acceptance probability of the proposed sampling scheme for the data-poor block in Section II.2.4 is only dependent on the hyperparameters. Consequently, choosing a computationally efficient sampling scheme for the hyperparameters can thus increase the computational efficiency of the overall sampling scheme within the data-poor block. In Section II.3.2 in Paper II, this aspect of the MCMC split sampler is further explored by using two different proposal densities for the hyperparameters. Developing more efficient proposal density for the hyperparmeters in the data-poor block can further increase the computational efficiency of the MCMC split sampler.

Furthermore, as the proposed sampler in the data-poor block is a modified version of the one block sampling scheme, linear constraints can be imposed on the data-poor component of the latent field. That is, MCMC samples obtained from the proposed sampler in Section 4.2.3, can be corrected for given linear constraints by conditioning by kriging, see for example Cressie (1993) and Rue and Held (2005). The development of conditioning by kriging for the MCMC split sampler is thus natural extension to the proposed inferential algorithm.

Due to the modularity of the MCMC split sampler, sampling schemes for the data-rich block can be developed and improved independently of the sampler in the data-poor block, and vice versa. Additionally, as the conditional posterior density $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$ becomes invariant of the data in the data-poor block, the computational advantages introduced by the conditional posterior structure in the data-poor block hold for all LGMs. Thus, in the author's view, further developing and improving sampling schemes that utilize the computational advantages introduced by the MCMC spilt sampler presents a promising area of future research.

6.3 Exracting external information from physical models

A novel method is proposed in this thesis that leverages information on the physical processes of precipitation from a local meteorological model, in order to yield an additional external source of spatial information on the physical processes of precipitation. The method is, in particular, aimed to address the limited amount of observational data from the sparse network of observational sites in Iceland. By extending the argument made in Benestad et al. (2012), the proposed method extracts spatial information on the physical processes of maximum precipitation by calculating sample mean values from outputs of the local meteorological model

of Crochet et al. (2007) on a 1 km by 1 km grid over the Icelandic spatial domain. This in turn yields a spatially referenced covariate, which aims to describe the spatial variation of average climate, in terms of precipitation, at every grid point. Although the proposed method is presented for an Icelandic data set, the method is extendable to any other location in the world where outputs from local meteorological models are available on a fine grid.

In addition to providing spatial information on extreme precipitation, the method can be extended to other response variables. For example, the method is likely to be well suited to resolve spatial information on annual or monthly precipitation. This is briefly explored in the example presented in Section II.3.1 in **Paper II**, which shows promising initial results. Additionally, the method can be utilized to extract any desired sample statistics from meteorological models in order to leverage spatial information. For example, instead of calculating the mean values from the outputs of the meteorological model at every regular grid point, standard deviation or quantiles could be extracted. To conclude, leveraging information from meteorological models for different regions, for other response variables and basing the information on various extracted sample statistics presents a potential venue for future research.

6.4 The statistical modeling of spatial extremes

The statistical analysis of annual maximum 24-hour precipitation in Iceland presented in this thesis is in line with the general approach to statistical modeling discussed in Section 4.1. Furthermore, the analysis demonstrates how an LGM with latent spatial models imposed on the location, scale and shape parameter of the g.e.v. distribution can be applied to the statistical modeling of spatial extremes. Moreover, the proposed spatial models are constructed with the SPDE method, which is a novel approach to the statistical modeling of spatial extremes, to the best of the author's knowledge. The main advantage of using spatial models based on the SPDE method is that they can be implemented with low computational cost. As such, SPDE spatial models scale well, in terms of computational cost, as the number of vertices in the mesh increases as discussed in Section 2.3.3 and Section 2.3.4. Therefore, the entire finite dimensional representation of the underlying GF can be sampled in each MCMC iteration with low computational cost[3]. In other words, the SPDE spatial models yield spatial predictions alongside the MCMC simulations. Further, by using linear interpolation in post calculation as discussed in Section 2.3.4 and Section 4.3.6, MCMC samples of all spatially varying model parameters are obtainable at every location in the spatial domain. Therefore, in this sense, the SPDE approach yields continuous spatial prediction of the spatially varying model parameters. As a consequence of this property, the posterior distribution of every p-th quantile of the g.e.v. distribution is also obtainable at every location in the spatial domain. For example, in Figure 5.13, the posterior mean of the 0.9 and the 0.99 quantiles are given at every regular point in the regular grid $\mathcal{G}$. Furthermore, the proposed LGM structure is extend-

[3]Especially, if the MCMC split sampler is used as an inferential algorithm

able to any regions in world, to any spatial domains of interest, and for any other response variable.

Furthermore, due to the modular property of LGMs, a different data density can be chosen in the proposed model setup without the need to change the structure of the SPDE spatial models at the latent level. As an example of a different choice of a data density for extreme precipitation is a t-copula with g.e.v marginals, as explored for example by Davison et al. (2012). The t-copula based data density presents an appealing choice to model the probabilistic dependence of the observations. The approach is well suited for simulating spatial realizations of the extreme precipitation for the next year or unobserved sites for an observed year[4]. Furthermore, the t-copula based data density can be implemented within the presented modeling framework, and presents possible future extensions to the proposed model

Stationary SPDE spatial models were implemented at the latent level of the proposed model, that is, the underlying GF is assumed stationary. However, the results in this thesis show evidence of a non-stationary behavior in some of the mountainous regions, for example near Kvísker, as discussed in Section III.4.2 in **Paper III**. Non-stationary SPDE spatial models have been proposed recently in the literature, see for example Fuglstad et al. (2013) and Ingebrigtsen et al. (2014). Implementing non-stationary SPDE models was beyond the scope of the research in this thesis. However, these models yield an appealing modeling option for non-stationary spatial fields. Future research could involve using spatial covariates in the dependence structure of the spatial model in a similar fashion as in Ingebrigtsen et al. (2014). For example, new spatial covariates based on outputs from a local meteorological model, as suggested here, could be designed to provide useful information for the dependence structure of non-stationary SPDE spatial models. This is, in the author's view, an interesting example of how the proposed method can be extended to other areas of future research.

6.5 Final words

This thesis has presented a general and computationally efficient approach to Bayesian statistical modeling and posterior inference. Through the use of LGMs and GMRF representations, it has been shown in this thesis that flexible Bayesian hierarchical models can be efficiently adapted to model involved and computationally challenging statistical problems. Furthermore, with the introduction of the MCMC split sampler to the extended LGM framework, a wealth of flexible statistical models become applicable in practice, from a computational standpoint, to various statistical problems. Therefore, from the author's point of view, the extended LGMs framework armed with the MCMC split sampler as

[4]There are two main reasons why this approach was not chosen in the analysis in this thesis. First, the main purpose is to give spatial predictions of marginal quantiles using the SPDE approach but not to simulate spatial realizations of the extremal surfaces. Secondly, many of the observational sites had missing observations, which presents further computational difficulties with the copula based data density.

an inferential algorithm provides a sizable venue of interesting future research of computationally efficient statistical research.

The efficiency of this general approach becomes particularly beneficial in the field of spatial statistics with the use of SPDE spatial models. By using the sparse precision structures of the GMRF representations, constructed with the SPDE approach, high dimensional representations of underlying spatial fields can be implemented efficiently within the extended LGM framework to various spatially referenced data. As demonstrated in this thesis, the extended LGM framework with SPDE spatial models is well suited for statistical modeling of spatial extremes and any statistical problem involving spatial modeling. Moreover, as a result of modularity of the MCMC split sampler, future improvements can be made separately to each of the two blocks of the model. Therefore, in the author's opinion, implementation of the proposed statistical model for spatial extremes to other areas in the world along with future improvements of the MCMC split sampler, presents an interesting and vivid area of future research.

Part II

Papers

I

Paper I

Discussion of 'Beyond mean regression'

Thiago G. Martins, Daniel Simpson, Janine B. Illian, Håvard Rue &
Óli Páll Geirsson

Martins T.G., Simpson D., Illian J.B., Rue H. and Geirsson Ó.P. 2013
Discussion of 'Beyond mean regression'. *Statistical Modelling* 13(4):355-361

I.1 Statistical modeling

The author of Kneib (2013) correctly points out that there is much more out there than "regression models for the mean", and discusses quantile regression in particular. While we agree with the first statement, from an academic point of view we do not share the enthusiasm about replacing the likelihood model with the generic quantile model. In a Bayesian context this leads to results that far from easy to interpret (see Yue and Rue (2011)). In addition, we do not advocate inferential schemes that do not yield good/reasonable estimates for the uncertainty involved. However, we agree that from a practical point of view "A man's got to do what a man's got to do", and sometimes just providing a (quantile) estimate for a specific dataset by using any reasonable approach with appropriate software, is the only thing one can do. Since this is a research discussion, however, the following focuses on how we think things should be done instead of how things are currently done in practice by applied researchers.

Since the discussion in Kneib (2013) does not explicitly distinguish between "modelling" and "inference", it is not clear to us if the author favours the non-parametric quantile approach due to its non-parametric nature or due to the availability of fast inferential algorithms for it. Our general approach to modelling complex problems is to design a sensible statistical model for a specific problem, to then develop a formal Bayesian machinery that is able to provide posterior distributions, and to then compute all the quantities of interest, such as quantiles, means, credible intervals, etc., from these posteriors.

Regarding the design of a sensible statistical model, we think that the focus of the discussion in Kneib (2013) should not only be on going "beyond *mean* regression", also beyond "beyond mean *regression*", i.e. with emphasis on regression, in the sense that it is perhaps time to think about statistical modelling rather than about "regression models". With a Gaussian response and with one covariate z we might then postulate that

$$y \mid \ldots \sim \mathcal{N}(\beta_0 + \beta_1 z, \exp(b_0)).$$

Although this model also models the mean structure it does this because the mean is a natural parameter in the Gaussian distribution. If we want to go "beyond mean regression" in this example, we could, as mentioned in Kneib (2013), model the log-variance as well, using the same covariate

$$y \mid \ldots \sim \mathcal{N}(\beta_0 + \beta_1 z, \exp(b_0 + b_1 z)).$$

This is an example of a GAMLSS model that results naturally from allowing the canonical parameters in the likelihood function to depend on various covariates. To allow for more flexibility the linear dependence on the covariate z may be relaxed

$$y \mid \ldots \sim \mathcal{N}(f(z), \exp(g(z))),$$

where, $f(\cdot)$ and $g(\cdot)$ are two "smooth" unknown functions of the covariate z, typically some finite-dimensional Gaussian process models. Assuming that we have access to a Bayesian inference machine, inference for the quantiles could be

extracted from the posterior distribution alone. In our view, it is not natural to use the "quantile-likelihood", as discussed in Kneib (2013), since the likelihood itself is often better understood than how various covariates enter the model. And clearly, with this way of thinking we are already moving away from regression modelling and towards statistical modelling.

More generally, we could proceed in a similar way and include a linear predictor or functions $f(\cdot)$ and $g(\cdot)$, or combinations of these, possibly sharing effects, in various places in the likelihood depending on the specific application. (We ignore the issue of parameterizing likelihood models here. Note that the GAMLSS models follow this route for location, scale and shape.) If we agree that this is a useful approach, then the question remains as to how to perform (approximate) Bayesian inference for such models, without using simulation based inference, which in our opinion is still too slow and too unreliable for routine usage. We return to this question below.

Another useful extension are models with more than one type of response. For instance, to expand on the model of the "Munich rental data" used by the author in Kneib (2013), we might be interested in jointly analysing a certain quantile of the rent per unit with, say, for the sake of argument, some longitudinal data on the people renting the apartments. It is reasonable to assume that these two responses share common covariates or "random"-effects, hence it makes statistical sense to do a joint analysis of these data. With a parametric form for the longitudinal data, it is not clear to us what scaling the quantile-likelihood, as presented in Kneib (2013), should have in such a joint formulation and if that approach is useful in this extended setting at all. Of course, a Bayesian approach will automatically make sense and provide a reasonable measure of the uncertainty in the estimates – *if* we can do the computations.

With these extensions in mind, is there any common model structure for which we can build a general approximate Bayesian inferential machine? In this context, we notice that, although implicitly mentioned, the role of "Gaussian Models" is not highlighted by the author of Kneib (2013). However, all the semiparametric models in Section 2 in Kneib (2013) are Gaussian models, conditional on the non-Gaussian variables $\boldsymbol{\theta}$. As sums of Gaussians are Gaussian, the joint conditional distribution of $(\boldsymbol{\eta}, \beta_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p) \mid \boldsymbol{\theta}$ is also Gaussian. With several linear predictors, say one for each natural parameter of a given likelihood with k parameters

$$\boldsymbol{\eta}^{(1)}, \ldots, \boldsymbol{\eta}^{(k)},$$

their joint conditional distribution conditional on $\boldsymbol{\theta}$ is also Gaussian. With (conditional independent) observations $\boldsymbol{y}$ depending on all the linear predictors in *some* form, we have

$$y_i \mid \ldots \sim \pi(y_i \mid \eta_i^{(1)}, \ldots, \eta_i^{(k)}).$$

The model is completed with a prior on the non-Gaussian components $\pi(\boldsymbol{\theta})$. To further generalise the model formulation, we may then write it in a hierarchical form, as outlined below.

Stage 1: The observations $\boldsymbol{y}$ depend on the latent field $\boldsymbol{x}$ and potential hyper-

parameters $\boldsymbol{\theta}$, through

$$y_i \mid \boldsymbol{x}, \boldsymbol{\theta} \;\sim\; \pi(y_i \mid \{x_j : j \in \mathcal{I}_i\}, \boldsymbol{\theta})$$

for some set $\mathcal{I}_i$.

Stage 2 The latent field $\boldsymbol{x}$ is Gaussian with hyperparameters $\boldsymbol{\theta}$,

$$\boldsymbol{x} \mid \boldsymbol{\theta} \;\sim\; \mathcal{N}\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}(\boldsymbol{\theta})^{-1}\right)$$

Stage 3 The hyperparameters have prior $\pi(\boldsymbol{\theta})$.

A reader familiar with the INLA approach to approximate Bayesian inference for latent Gaussian models (LGMs), will recognise the similarity of the definition of a LGM in Rue et al. (2009) and this more general one defined here, where each data point can be connected with more than one latent component in an arbitrary way, see Rue et al. (2009), Martins et al. (2013b) and the web-site `www.r-inla.org` for available software. In particular, a more restrictive definition is obtained with $|\mathcal{I}_i| \in \{0, 1\}$ for all i.

The concept of LGMs has proved very successful for doing Bayesian inference, and the wealth of models that may be interpreted as LGMs, and hence are amenable to the Bayesian inference approach developed specifically for LGMs, is simply impressive; see Martins et al. (2013b) for various applications and further references. Needless to say, the more general setting outlines above, which we might call "LGM–2.0" for the moment, can do even more. In fact, much much more, and it is to our view *the* most commonly used (Bayesian) model construction in statistics. However, only the case where $|\mathcal{I}_i| \in \{0, 1\}$ for all i is supported by the current INLA approach, and the `R`-package `R-INLA` provides support for the most common models, although the INLA idea has already been re-implemented in a special case of a LGM–2.0 (Ferkingstad et al. 2008). The extension in LGM–2.0 might seem trivial but it is not. However, it is solvable. Due to the generality of LGM–2.0, a further challenge is to implement a general interface within `R`.

I.2 Example: Modeling maximum precipitation

We present an example where one should consider to go beyond a simple mean regression scheme and think more in terms of statistical modelling. Within this framework the usefulness of LGMs is obvious. The example is based on observations of annual maximum 24 hour precipitation from 86 observation sites in Iceland, as seen in Figure I.1, and outputs from a local climate model on a 1 km regular grid as covariates. We are interested in making quantile predictions of maximum precipitation on the 1 km grid. Thus, in the statistical modelling sense, the aim is to obtain distributional properties of maximum precipitation on the grid.

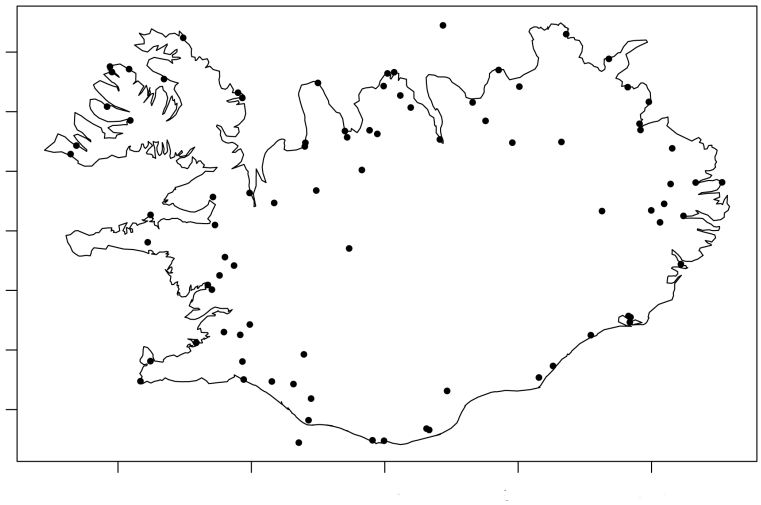

Figure I.1: The location of observation sites

Following the modelling formulation above, we propose the following model.

Stage 1 The data are modelled with a Bayesian hierarchical model assuming a generalized extreme value distribution for the observations. That is, let y_{it} denote the annual maximum 24 hour precipitation at station i at year t, with a cumulative distribution function of the form

$$F(y_{it}) = \exp\left\{-\left(1+\xi\left(\frac{y_{it}-\eta_i}{\sigma_i}\right)\right)^{-1/\xi}\right\},$$

where η_i, σ_i and ξ are location, scale and shape parameters, respectively. Observations are assumed to be independent at the data level.

Stage 2 Working within the LGM framework, we decompose $\boldsymbol{\eta}$ into

$$\boldsymbol{\eta} = \boldsymbol{X}_\eta \beta_\eta + \boldsymbol{u}_\eta + \boldsymbol{v}_\eta,$$

where $\boldsymbol{X}$ is a design matrix based on outputs from the local climate model, β_η are the corresponding weights, $\boldsymbol{u}_\eta$ denotes a Matérn type spatial field constructed with the SPDE approach of Lindgren et al. (2011) and $\boldsymbol{v}_\eta$ is an unstructured random effect. We assign the following priors

$$\begin{aligned}
\beta &\sim \mathcal{N}(0, \kappa_\beta^{-1}) \\
\boldsymbol{u} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_u(\theta_u)^{-1}) \\
\boldsymbol{v} &\sim \mathcal{N}(\boldsymbol{0}, \kappa_v^{-1}\boldsymbol{I}).
\end{aligned}$$

The joint distribution of $(\boldsymbol{\eta}, \boldsymbol{u}, \beta)$ is a zero mean Gaussian with precision

matrix

$$\boldsymbol{Q}_\eta = \begin{pmatrix} \kappa_v \boldsymbol{I} & -\kappa_v \boldsymbol{I} & -\kappa_v \boldsymbol{X}_\eta \\ -\kappa_v \boldsymbol{I} & \kappa_v \boldsymbol{I} + \boldsymbol{Q}_u & \kappa_v \boldsymbol{X}_\eta \\ -\kappa_v \boldsymbol{X}_\eta^T & \kappa_v \boldsymbol{X}_\eta^T & \kappa_v \boldsymbol{X}_\eta^T \boldsymbol{X}_\eta + \kappa_\beta. \end{pmatrix}$$

Let $\tau_i = \log \sigma_i$, then a similar structure can be implemented for $\boldsymbol{\tau}$. Let

$$\boldsymbol{x} = (\boldsymbol{\eta}, \boldsymbol{u}_\eta, \beta_\eta, \boldsymbol{\tau}, \boldsymbol{u}_\tau, \beta_\tau).$$

The joint distribution of $\boldsymbol{x}$ is then a zero mean Gaussian with precision matrix

$$\boldsymbol{Q}_x = \begin{pmatrix} \boldsymbol{Q}_\eta & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_\tau. \end{pmatrix}$$

Note that the regression aspect of this modelling scheme appears naturally in the parameterization of the Gaussian distribution.

Stage 3 Suitable prior distributions are chosen for the hyper parameters $\boldsymbol{\theta}$.

Successfully inferring the parameters of the model yields posterior estimates for the distribution of maximum annual 24 hour precipitation in every grid point. In particular, an estimate of the p-th quantile can be given in every grid point. The p-th quantile function of the likelihood is of the form

$$q_p(\eta, \tau, \xi) = \eta + \frac{\exp(\tau)}{\xi} \left(-\log(p)^{-\xi} - 1 \right).$$

So a posterior estimate for the p-th quantile can be found with

$$E\left[q_p(\eta, \tau, \xi) | \, \boldsymbol{y} \right] = \int q_p(\eta, \tau, \xi) \pi(\eta, \tau, \xi | \boldsymbol{y}) d\eta d\tau d\xi.$$

This estimate can be given in each grid point, as can be seen in Figure I.2, where estimates for the 95th quantile are given.

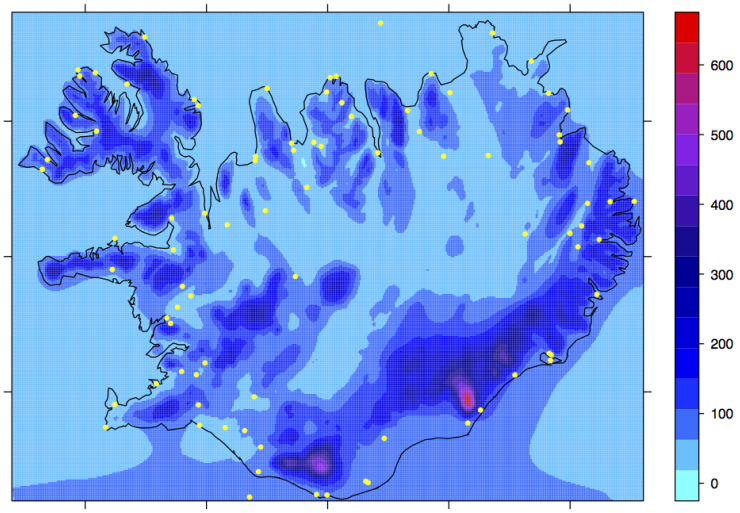

Figure I.2: Estimates for the 95th quantile of maximum precipitation.

II

Paper II

The MCMC split sampler: A block Gibbs sampling scheme for latent Gaussian models

Óli Páll Geirsson, Birgir Hrafnkelsson, Daniel Simpson &
Helgi Sigurðarson

Geirsson Ó.P., Hrafnkelsson B., Simpson D. and Sigurðarson H. 2015 The MCMC split sampler: A block Gibbs sampling scheme for latent Gaussian models. *Submitted to Statistics and Computing*

Abstract

A novel computationally efficient Markov chain Monte Carlo (MCMC) scheme for latent Gaussian models (LGMs) is proposed in this paper. The sampling scheme is a two block Gibbs sampling scheme designed to exploit the model structure of LGMs. We refer to the proposed sampling scheme as the MCMC split sampler. The principle idea behind the MCMC split sampler is to split the latent Gaussian parameters into two vectors. The former vector consists of latent parameters which appear in the data density function, while the latter vector consists of latent parameters which do not appear in it. The former vector is placed in the first block of the proposed sampling scheme and the latter vector is placed in the second block along with any potential hyperparameters. The resulting conditional posterior density functions within the blocks allow the MCMC split sampler to handle, by design, LGMs with latent models imposed on more than just the mean structure of the data density function. The MCMC split sampler is also designed to be applicable for any choice of a parametric data density function. Moreover, it scales well in terms of computational efficiency when the dimension of the latent model increase.

II.1 Introduction

Latent Gaussian models (LGMs) form a flexible subclass of Bayesian hierarchical models and have become popular in many areas of statistics and various fields of applications, as LGMs are practical from a statistical modeling point of view and readily interpretable. For example, LGMs play an important role in spatial statistics, see Cressie (1993), Diggle et al. (1998), Chiles and Delfiner (2009); statistical climatology (Cooley et al. 2007, Guttorp and Gneiting 2006); disease mapping (Pettitt et al. 2002, Lawson 2013); stochastic volatility models (Martino et al. 2011); and hydrology (Schaefli et al. 2007), to name a few. Moreover, LGMs can be viewed as a specific extension of structured additive regression models (Fahrmeir et al. 1994, Rue et al. 2009), in the sense that, the data density function of each data point can depend on more than a single linear functional of the latent field through more than just the mean structure, as discussed in Martins et al. (2013a).

Although LGMs are well suited from a statistical modeling point of view their posterior inference becomes computationally challenging when latent models are desired for more than just the mean structure of the data density function (Martins et al. 2013a), when the number of parameters associated with the latent model increase; or when the data density function is non-Gaussian. The aim of this paper is to propose a novel computationally efficient Markov chain Monte Carlo (MCMC) scheme which serves to address these computational issues. The proposed sampling scheme is referred to as the *MCMC split sampler* in this paper. It is designed to handle LGMs where latent models are imposed on more than just the mean structure of the likelihood. It scales well in terms of computational efficiency when the dimensions of the latent models increase and it is applicable for any choice of a parametric data density function. The main novelty of the

MCMC split sampler lies in how the model parameters of a LGM are split into two blocks. As a result of the proposed blocking scheme, one of the blocks exploits the latent Gaussian structure in a natural way and becomes invariant of the data density function.

Markov chain Monte Carlo (MCMC) methods form the backbone of modern Bayesian posterior inference and are, in principle, applicable to almost any Bayesian model. However, the mixing and convergence properties of the MCMC chains can be poor for involved models structures and large data sets if parameters that are dependent in the posterior are not dealt with properly, see for example Murray and Adams (2010). In particular, the mixing and convergence properties of the popular single site updating strategy can be extremely poor due to strong dependencies of parameters in the posterior distribution as discussed in Knorr-Held and Rue (2002). Several MCMC sampling strategies have been suggested for Bayesian hierarchical models to improve the mixing properties of MCMC algorithms. For example, methods based on approximate diffusions such as the Metropolis Adjusted Langevin algorithm (MALA), see Roberts and Rosenthal (1998); methods based on Hamiltonian mechanics (HMC), suggested by Neal (1993), which use the gradient of the target density to drive the proposal mechanism toward regions of higher posterior density; manifold methods, proposed by Girolami and Calderhead (2011), which provide a systematic way of designing proposal densities for MALA and HMC by making use of the gradient and curvature information of the target density; and various block sampling strategies such as the one block updating strategy of Knorr-Held and Rue (2002). Filippone et al. (2013) conducted a detailed comparison of these methods for LGMs and found that the single block strategy of Knorr-Held and Rue (2002), in which the latent field and its corresponding hyperparameters are updated jointly in a single block, performed best in most situations. Furthermore, by using numerical methods for fast sampling of Gaussian Markov random fields (GMRFs) (Rue 2001), the single block sampler can be implemented with a low computational cost for LGMs. However, using only a single block sampler for LGMs with a non-Gaussian likelihood and high dimensional latent fields can be problematic, as parameters accepted in regions of low posterior probability can cause the MCMC chain to get stuck.

Alternative to MCMC methods are deterministic approximate posterior inference methods, such as the Integrated nested Laplace approximation (INLA) (Rue et al. 2009). INLA is a fast approximate inference method for LGMs in which the data density of each data point only depends on a single linear functional of the latent field. While this assumption holds in many practical cases, there are many models in which we want the latent field to enter the data density of a single observation through two or more parameters, see Kneib (2013) and Martins et al. (2013a) for further discussion. For example in Hrafnkelsson et al. (2012) and Geirsson et al. (2015), latent Gaussian spatial models were imposed on the location, scale and shape structure of the data density function.

The MCMC split sampler is a two block Gibbs sampling scheme (Geman and Geman 1984, Casella and George 1992) designed for LGMs, which addresses the aforementioned inference problems. The MCMC split sampler is based on the

following model setup for LGMs, to which we adhere to in this paper.

Data-level: The observations $\boldsymbol{y}$ depend on the latent field $\boldsymbol{x}$, through some choice of data distribution with a data density function $\pi(\boldsymbol{y} \mid \boldsymbol{x})$.

Latent level: The prior for the latent field $\boldsymbol{x}$ is Gaussian and is potentially dependent on hyperparameters $\boldsymbol{\theta}$, with a density function

$$\pi(\boldsymbol{x} \mid \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{x} \mid \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}(\boldsymbol{\theta})^{-1}\right).$$

Hyperparameter level: A prior distribution is assigned for the hyperparameters $\boldsymbol{\theta}$, with a density function $\pi(\boldsymbol{\theta})$.

The principle idea behind the MCMC split sampler is to split the latent Gaussian parameters $\boldsymbol{x}$ into two vectors, $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$, where $\boldsymbol{\eta}$ consists of elements that appear in the data density function and $\boldsymbol{\nu}$ consists of elements that do not appear in it. Thus, the data $\boldsymbol{y}$ become conditionally independent of $(\boldsymbol{\nu}, \boldsymbol{\theta})$ conditioned on $\boldsymbol{\eta}$, that is, $\pi(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) = \pi(\boldsymbol{y} \mid \boldsymbol{\eta})$. For the posterior inference, all the model parameters are grouped into two blocks. That is, $\boldsymbol{\eta}$ is placed in a block we refer to as the *data-rich* block in this paper, while both $\boldsymbol{\nu}$ and the hyperparameters $\boldsymbol{\theta}$ are placed in another block referred to as the *data-poor* block. A Gibbs sampling strategy is then implemented for each block, conditioned on the other block.

In many practical applications with non-Gaussian data density functions, especially in the field of spatial statistics, the vector $\boldsymbol{\eta}$ in the data-rich block has a complicated but low-dimensional conditional posterior structure, while the parameter vector $\boldsymbol{\nu}$ in the data-poor block is often of much higher dimension than that of the parameters in the data-poor block (Hrafnkelsson et al. 2012). Therefore, by using the proposed blocking scheme the potentially computationally demanding conditional posterior density in the data-rich block contains a minimum number of necessary parameters. Furthermore, the conditional posterior density of $\boldsymbol{\nu}$ becomes conditionally Gaussian conditioned on the parameter vector $\boldsymbol{\eta}$ and the hyperparameters $\boldsymbol{\theta}$. This designed structure, that is, the minimal dimension of the parameters in the data-rich block and the conditional Gaussian posterior structure within the data-poor block, is exploited to implement computationally efficient sampling schemes within the blocks. Further, the proposed scheme scales well when the dimension of $\boldsymbol{\nu}$ increases, as discussed in the ensuing paragraph.

The MCMC split sampler is modular by design such that, in principle, any efficient MCMC sampler can be implemented for each block. In this paper, we propose computationally efficient sampling strategies that are tailored to the particular conditional model structure of each block. Within the data-rich block we present a strategy based on the gradient and curvature information of the target density that results in an independence proposal mechanism as discussed in Rue and Held (2005). Conditional independence resulting from the model structure within the block can be utilized in some cases to increase acceptance in the Metropolis-Hastings algorithm (Metropolis et al. 1953, Hastings 1970). In order to update the data-poor block, a modified version of the fast single block updater of Knorr-Held and Rue (2002) is proposed, which exploits the fact that

the conditional posterior of the vector $\boldsymbol{\nu}$ is Gaussian conditioned on $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$. This step is invariant of the choice of a data density function as the data-poor block is updated conditioned on the data-rich block. Moreover, if the latent field $\boldsymbol{x}$ is a GMRF with a sparse precision structure, the sampling strategy for the data-poor block is shown to conserve the sparse GMRF precision structure, allowing for fast sampling of the corresponding GMRF.

The paper is organized as follows. Section 2.1 and Section 2.2 are devoted to the motivation, introduction and the setup of the MCMC split sampler. The proposed sampling schemes within the data-rich and data-poor blocks are presented in Sections 2.3 and Section 2.4, respectively. Examples on the implementation of the MCMC split sampler are given in Section 3. In Section 3.1 we present a LGM with a latent spatial model structure on mean and log-variance parameters, and show how the MCMC split sampler scales well as the dimensions of the latent parameters in the data-poor block increase. An example on extremes based on a simulated data set is given in Section 3.2, where the focus is on a LGM where latent models are imposed on all three parameters of the generalized extreme value distribution. In Appendix A, we give an extension to the sampling scheme proposed in Section 2.3, which is applicable if conditional independence assumptions are imposed on the data data density function. Lastly, in Appendix B, we show the necessary proofs for the main results in the paper.

II.2 The MCMC split sampler

II.2.1 Motivation and model setup

Consider, as a motivation and without loss of generality, a data density function $\pi(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\tau})$ where $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$ are vectors of location and log-scale parameters, respectively, which are modeled with latent Gaussian fields. That is, assume the following additive model structure

$$\boldsymbol{\mu} = \boldsymbol{X}_\mu \boldsymbol{\beta}_\mu + \boldsymbol{A}_\mu \boldsymbol{u}_\mu, \quad \boldsymbol{\tau} = \boldsymbol{X}_\tau \boldsymbol{\beta}_\tau + \boldsymbol{A}_\tau \boldsymbol{u}_\tau$$

where $\boldsymbol{X}_\mu$ and $\boldsymbol{X}_\tau$ are fixed design matrices; $\boldsymbol{\beta}_\mu$ and $\boldsymbol{\beta}_\tau$ are the corresponding weights; $\boldsymbol{A}_\mu$ and $\boldsymbol{A}_\tau$ are fixed matrices; and $\boldsymbol{u}_\mu$ and $\boldsymbol{u}_\tau$ are structured random effects. In order to increase computational stability in the posterior inference calculations and for mathematical derivations of the MCMC split sampler, we introduce unstructured random effects, $\boldsymbol{\epsilon}_\mu$ and $\boldsymbol{\epsilon}_\tau$, to the model. That is,

$$\boldsymbol{\mu} = \boldsymbol{X}_\mu \boldsymbol{\beta}_\mu + \boldsymbol{A}_\mu \boldsymbol{u}_\mu + \boldsymbol{\epsilon}_\mu, \quad \boldsymbol{\tau} = \boldsymbol{X}_\tau \boldsymbol{\beta}_\tau + \boldsymbol{A}_\tau \boldsymbol{u}_\tau + \boldsymbol{\epsilon}_\tau. \tag{II.1}$$

Small variances can be imposed *a priori* on the unstructured random effects $\boldsymbol{\epsilon}_\mu$ and $\boldsymbol{\epsilon}_\tau$ if they are not desired in the model. However, adding the unstructured random effects is reasonable in many cases from a statistical modeling point of view as they serve as error terms for the latent models. Furthermore, adding the $\boldsymbol{\epsilon}_\mu$ and $\boldsymbol{\epsilon}_\tau$ terms to the latent models yields an analogous latent model structure as implied by the structured additive regression in Fahrmeir et al. (1994) where only the mean is linked to a structured additive predictor through a link function.

Assign the following Gaussian prior density functions to the latent model parameters

$$\begin{aligned}
\pi(\boldsymbol{\beta}_\mu) &= \mathcal{N}(\boldsymbol{\beta}_\mu \mid \boldsymbol{\mu}_{\mu\beta}, \boldsymbol{Q}_{\mu\beta}^{-1}), \quad & \pi(\boldsymbol{\beta}_\tau) &= \mathcal{N}(\boldsymbol{\beta}_\tau \mid \boldsymbol{\mu}_{\tau\beta}, \boldsymbol{Q}_{\tau\beta}^{-1}) \\
\pi(\boldsymbol{u}_\mu) &= \mathcal{N}(\boldsymbol{u}_\mu \mid \boldsymbol{\mu}_{\mu u}, \boldsymbol{Q}_{\mu u}^{-1}), \quad & \pi(\boldsymbol{u}_\tau) &= \mathcal{N}(\boldsymbol{u}_\tau \mid \boldsymbol{\mu}_{\tau u}, \boldsymbol{Q}_{\tau u}^{-1}) \\
\pi(\boldsymbol{\epsilon}_\mu) &= \mathcal{N}(\boldsymbol{\epsilon}_\mu \mid \mathbf{0}, \boldsymbol{Q}_{\mu\epsilon}^{-1}), \quad & \pi(\boldsymbol{\epsilon}_\tau) &= \mathcal{N}(\boldsymbol{\epsilon}_\tau \mid \mathbf{0}, \boldsymbol{Q}_{\tau\epsilon}^{-1})
\end{aligned} \tag{II.2}$$

where parameters of the prior density functions can potentially depend of a set of hyperparameters $\boldsymbol{\theta}$, and $\boldsymbol{Q}_{\mu\epsilon}^{-1}$ and $\boldsymbol{Q}_{\tau\epsilon}^{-1}$ are diagonal matrices.

As the vector $\boldsymbol{\mu}$ in equation (II.1) is a linear combination of $\boldsymbol{\beta}_\mu$, $\boldsymbol{u}_\mu$ and $\boldsymbol{\epsilon}_\mu$, it is equivalent to obtain MCMC samples from the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\beta}_\mu, \boldsymbol{u}_\mu)$ and from the posterior distribution of $(\boldsymbol{\beta}_\mu, \boldsymbol{u}_\mu, \boldsymbol{\epsilon}_\mu)$. Analogous argument holds for the log-scale parameters. The MCMC split sampler is designed to obtain MCMC samples from the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\tau, \boldsymbol{u}_\mu, \boldsymbol{u}_\tau)$ as opposed to $(\boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\tau, \boldsymbol{u}_\mu, \boldsymbol{u}_\tau, \boldsymbol{\epsilon}_\mu, \boldsymbol{\epsilon}_\tau)$ as in the former parameterization only the vector $(\boldsymbol{\mu}, \boldsymbol{\tau})$ enters the data density function, while all the elements of the latter vector enter the data density function in the latter parameterization. This parameterization for posterior inference is along the lines of the posterior inference scheme proposed in Rue et al. (2009). Thus, define

$$\boldsymbol{\eta} = \left(\boldsymbol{\mu}^\mathsf{T}, \boldsymbol{\tau}^\mathsf{T}\right)^\mathsf{T}, \quad \boldsymbol{\nu} = \left(\boldsymbol{\beta}_\mu^\mathsf{T}, \boldsymbol{u}_\mu^\mathsf{T}, \boldsymbol{\beta}_\tau^\mathsf{T}, \boldsymbol{u}_\tau^\mathsf{T}\right)^\mathsf{T}$$

which will act as the splitting of the parameters of the latent field.

The latent model structure in (II.1) and the prior distributions in (II.2) can be written in a joint matrix form, which forms the basis for the derivation of the MCMC split sampler. Define the following matrices and vectors

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X}_\mu & \boldsymbol{A}_\mu & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{X}_\tau & \boldsymbol{A}_\tau \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_\mu \\ \boldsymbol{\epsilon}_\tau \end{pmatrix}, \quad \boldsymbol{Q}_\epsilon = \begin{pmatrix} \boldsymbol{Q}_{\mu\epsilon} & \cdot \\ \cdot & \boldsymbol{Q}_{\tau\epsilon} \end{pmatrix}$$

and group the following parameters and matrices together

$$\boldsymbol{\mu}_\nu = \begin{pmatrix} \boldsymbol{\mu}_{\mu\beta} \\ \boldsymbol{\mu}_{\mu u} \\ \boldsymbol{\mu}_{\tau\beta} \\ \boldsymbol{\mu}_{\tau u} \end{pmatrix}, \quad \boldsymbol{Q}_\nu = \begin{pmatrix} \boldsymbol{Q}_{\mu\beta} & \cdot & \cdot & \cdot \\ \cdot & \boldsymbol{Q}_{\mu u} & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{Q}_{\tau\beta} & \cdot \\ \cdot & \cdot & \cdot & \boldsymbol{Q}_{\tau u} \end{pmatrix}$$

where the dotted entries denote zero entries. The additive model structure implied by (II.1) for both latent parameters is thus equivalent to the matrix form

$$\boldsymbol{\eta} = \boldsymbol{Z}\boldsymbol{\nu} + \boldsymbol{\epsilon} \tag{II.3}$$

and the Gaussian prior assumptions in (II.2) are equivalent to

$$\pi(\boldsymbol{\eta} \mid \boldsymbol{\nu}) = \mathcal{N}\left(\boldsymbol{\eta} \mid \boldsymbol{Z}\boldsymbol{\nu}, \boldsymbol{Q}_\epsilon^{-1}\right), \qquad \text{(II.4)}$$
$$\pi(\boldsymbol{\nu}) = \mathcal{N}\left(\boldsymbol{\nu} \mid \boldsymbol{\mu}_\nu, \boldsymbol{Q}_\nu^{-1}\right).$$

As the data density function and the corresponding parameters were arbitrarily chosen above, analogous derivations can be carried out for any parametric data density function and any of its parameters. For example, in addition to imposing latent Gaussian models on the location and log-scale parameters of the generalized extreme value distribution a latent Gaussian model can also be imposed on the shape parameter, see Section II.3.2 for details. Therefore, equations (II.3) and (II.4) are general in the sense that most of LGMs used in practice can be expressed in the same form. We will thus adapt equations (II.3) and (II.4) as a general setup for the latent model structures for LGMs henceforth in this paper.

The following lemma, based on known results, plays a vital role in the implementation of the MCMC split sampler.

Lemma II.1. *Assume the distribution assumptions given in (II.4), for any mean vector $\boldsymbol{\mu}_\nu$, fixed matrix $\boldsymbol{Z}$, and precision matrices $\boldsymbol{Q}_\epsilon$ and $\boldsymbol{Q}_\nu$. The joint prior density function of $(\boldsymbol{\eta}, \boldsymbol{\nu})$ is then Gaussian of the form*

$$\pi\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix} = \mathcal{N}\left(\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix}\middle|\, \begin{pmatrix}\boldsymbol{Z}\boldsymbol{\mu}_\nu\\ \boldsymbol{\mu}_\nu\end{pmatrix}, \begin{pmatrix}\boldsymbol{Q}_\epsilon & -\boldsymbol{Q}_\epsilon\boldsymbol{Z}\\ -\boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon & \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}\end{pmatrix}^{-1}\right) \qquad \text{(II.5)}$$

and the conditional density function of $\boldsymbol{\nu}$ conditioned on $\boldsymbol{\eta}$ becomes

$$\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}) = \mathcal{N}\left(\boldsymbol{\nu}\middle|\boldsymbol{Q}_{\nu|\eta}^{-1}(\boldsymbol{Q}_\nu\boldsymbol{\mu}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{\eta}), \boldsymbol{Q}_{\nu|\eta}^{-1}\right) \qquad \text{(II.6)}$$

where $\boldsymbol{Q}_{\nu|\eta} = \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}$.

See Appendix II.B.1 for proof. Note that, as the vector $(\boldsymbol{\eta}^\mathsf{T}, \boldsymbol{\nu}^\mathsf{T})^\mathsf{T}$ is jointly Gaussian it can be viewed as the latent Gaussian vector $\boldsymbol{x}$ in the LGM setup.

II.2.2 The sampling scheme

The vector $\boldsymbol{\eta}$ in (II.3) consists of the parameters of the latent field that explicitly enter the likelihood function while the vector $\boldsymbol{\nu}$ consists of the parameters of the latent field which do not enter it. Therefore, the data vector $\boldsymbol{y}$ is conditionally independent of $\boldsymbol{\nu}$ conditioned on $\boldsymbol{\eta}$, that is $\pi(\boldsymbol{y} \mid \boldsymbol{\eta}, \boldsymbol{\nu}) = \pi(\boldsymbol{y} \mid \boldsymbol{\eta})$. The parameters $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$ are referred to as the data-rich and data-poor components of the latent field, respectively, in this paper. The corresponding posterior distribution, where the data-poor components of the latent field are potentially dependent on a vector of hyperparameters $\boldsymbol{\theta}$, is thus proportional to

$$\pi(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}) \propto \pi(\boldsymbol{y} \mid \boldsymbol{\eta})\pi(\boldsymbol{\eta}, \boldsymbol{\nu} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}). \qquad \text{(II.7)}$$

Using the relationship in (II.7), we propose the following two block MCMC sampling scheme to obtain MCMC samples from the posterior density $\pi(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y})$. The vector $\boldsymbol{\eta}$ is placed in the data-rich block, and the vectors $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ are grouped together in the data-poor block. The MCMC split sampler obtains a sample from the posterior density $\pi(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y})$ by sampling from one of the blocks conditioned on the other in a Gibbs sampling setting. That is, the $(k+1)$-th MCMC sample from the posterior density $\pi(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y})$ is obtained by using the following two block Gibbs sampling scheme

Data-rich block: sample $\boldsymbol{\eta}^{k+1}$ from $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)$

Data-poor block: sample $(\boldsymbol{\nu}^{k+1}, \boldsymbol{\theta}^{k+1})$ jointly from $\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})$

This scheme forms the basis of the MCMC split sampler. The potentially involved but often low-dimensional structure of the data-rich block is separated from the parameters in the data-poor block. By separating the two blocks, MCMC sampling strategies which exploit the conditional model structures can be implemented within each block in order to increase computational efficiency. Although any computationally efficient MCMC samplers are applicable within the blocks, we propose the following sampling schemes which are tailored for the conditional model structures of the blocks. The details of the proposed samplers for each block are summarized in Section II.2.3 and Section II.2.4.

II.2.3 Sampler for the data-rich block

The conditional posterior density function $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ in the data-rich block is intractable in most applications. In order to obtain MCMC samples from the conditional posterior density function we propose the following Metropolis–Hasting type MCMC algorithm with a tailored independence proposal density (Rue and Held 2005).

To construct a computationally efficient independence proposal density, we approximate the conditional posterior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ with a Gaussian approximation evaluated at the mode of conditional posterior density. Using the logarithm of the conditional posterior, that is

$$\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = f(\boldsymbol{\eta}) - \frac{1}{2}\boldsymbol{\eta}^{\mathsf{T}} \boldsymbol{Q}_{\epsilon} \boldsymbol{\eta} + (\boldsymbol{Q}_{\epsilon} \boldsymbol{Z} \boldsymbol{\nu})^{\mathsf{T}} \boldsymbol{\eta} + \text{const} \tag{II.8}$$

where $f(\boldsymbol{\eta}) = \log \pi(\boldsymbol{y} \mid \boldsymbol{\eta})$ for notational convenience, the following can be shown.

Theorem II.2. *The Gaussian approximation of the conditional posterior density* $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ *is given by*

$$\tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{\eta} \mid \boldsymbol{\eta}^0, (\boldsymbol{Q}_{\epsilon} - \boldsymbol{H})^{-1}\right) \tag{II.9}$$

where $\boldsymbol{\eta}^0$ *is the mode of the conditional posterior density* $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ *and* $\boldsymbol{H}$ *is the Hessian of the logarithm of conditional posterior evaluated at the mode,* $\boldsymbol{H} = \nabla^2 f(\boldsymbol{\eta}^0)$.

See Appendix II.B.1 for proof. Note that, adding the additive unstructured error term $\boldsymbol{\epsilon}$ to the model in (II.2) prevents the precision matrix in (II.9) from being singular and thus ensures numerical stability.

As the Gaussian approximation in (II.9) is constructed at the conditional posterior mode $\boldsymbol{\eta}^0$, a proposal density q for $\boldsymbol{\eta}$ based on (II.9) thus becomes invariant of the current position of $\boldsymbol{\eta}$ in the MCMC iteration. Therefore, the proposal density q is an independence proposal density (Chib and Greenberg 1995, Rue and Held 2005). That is, in the $(k+1)$-th iteration the proposal density is invariant of $\boldsymbol{\eta}^k$, that is $q(\boldsymbol{\eta}^* \mid \boldsymbol{\eta}^k) = q(\boldsymbol{\eta}^*)$.

When a new $\boldsymbol{\eta}^*$ is proposed with the independence proposal density in (II.9) in the $(k+1)$-th iteration, it is accepted with probability

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{\eta}^* \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})}{\pi(\boldsymbol{\eta}^k \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})} \cdot \frac{q(\boldsymbol{\eta}^k)}{q(\boldsymbol{\eta}^*)}\right\}. \tag{II.10}$$

The logarithm of the ratio in (II.10) can be simplified, as stated in Lemma II.3, in order to reduce computational cost.

Lemma II.3. *Assume the proposal density q implied by the Gaussian approximation in (II.9) for the data-rich block. The logarithm of the acceptance ratio given in (II.10) can be simplified to*

$$r = f(\boldsymbol{\eta}^*) - \left(\frac{1}{2}(\boldsymbol{\eta}^*)^\mathsf{T}\boldsymbol{H} + \boldsymbol{b}^\mathsf{T}\right)\boldsymbol{\eta}^* - f(\boldsymbol{\eta}^k) + \left(\frac{1}{2}(\boldsymbol{\eta}^k)^\mathsf{T}\boldsymbol{H} + \boldsymbol{b}^\mathsf{T}\right)\boldsymbol{\eta}^k \tag{II.11}$$

where $\boldsymbol{b} = \nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H}\boldsymbol{\eta}^0$.

See Appendix II.B.2 for proof. As the gradient $\nabla f(\boldsymbol{\eta}^0)$ and Hessian $\boldsymbol{H}$ have already been calculated to obtain (II.9), the expression in (II.11) is computationally efficient to calculate.

In many applications conditional independence assumptions are imposed on the data density function. That is, there exists a partition of $\boldsymbol{\eta}$ into subvectors $\boldsymbol{\eta}_i$, such that $\pi(\boldsymbol{y} \mid \boldsymbol{\eta}) = \prod_i \pi_i(\boldsymbol{y}_i \mid \boldsymbol{\eta}_i)$, which is turn implies $f(\boldsymbol{\eta}) = \sum_i f_i(\boldsymbol{\eta}_i)$, where f_i is the logarithm of the marginal data density function in the i-th partition. In some cases, a proposal density based on the Gaussian approximation in (II.9) can be a poor approximation of the conditional posterior density in some partition of $\boldsymbol{\eta}$. Updating the whole vector $\boldsymbol{\eta}$ in one block may then result in the MCMC chain getting stuck. As a result the computational efficiency of the sampler is reduced. In order to circumvent this issue and to retain the computational speed gained by using the Gaussian approximation in (II.9) as a proposal density, a modification can be made to the sampling scheme which utilizes the conditional independence of the partitions within the data-rich block. The details on the modification can be seen in Appendix II.A. The resulting sampling scheme is outlined in Algorithm 3. Note that by choosing $I = 1$ in Algorithm 3, the above sampling scheme without the conditional independence assumptions on the likelihood is obtained, while selecting $I \geq 2$ in Algorithm 3 assumes the aforementioned partitioning of $\boldsymbol{\eta}$ and that each $\boldsymbol{\eta}_i$ is accepted or rejected separately.

Algorithm 3 The proposed algorithm for obtaining the $(k+1)$-th sample from $\pi(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ in the data-rich block. By choosing $I = 1$, the sampling scheme introduced in Section II.2.3 is obtained. For $I \geq 2$ the modified sampling scheme, which is derived in Appendix II.A, is obtained for the partitions.

Input: $(\boldsymbol{\eta}^k, \boldsymbol{\nu}^k)$

Find the mode $\boldsymbol{\eta}^0 = \arg\max_{\boldsymbol{\eta}} \log \pi(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{\nu}^k, \boldsymbol{\theta})$

Calculate $\boldsymbol{H} = \nabla^2 f(\boldsymbol{\eta}^0)$ and $\boldsymbol{b} = \nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H}\boldsymbol{\eta}^0$

Sample $\boldsymbol{\eta}^* \sim \mathcal{N}\left(\boldsymbol{\eta}^0, (\boldsymbol{Q}_\epsilon - \boldsymbol{H})^{-1}\right)$

Calculate $\boldsymbol{\rho}(\boldsymbol{\eta}^k)$ and $\boldsymbol{\rho}(\boldsymbol{\eta}^*)$, where

$$\boldsymbol{\rho}(\boldsymbol{\eta}) = \left(-\frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{H} - \boldsymbol{b}^\mathsf{T}\right) \circ \boldsymbol{\eta}$$

and $\circ$ denotes an entrywise product

for $i = 1, \ldots, I$

  Calculate $r_i = f_i(\boldsymbol{\eta}_i^*) + \boldsymbol{\rho}(\boldsymbol{\eta}^*)_i^\mathsf{T}\mathbf{1} - \left(f_i(\boldsymbol{\eta}_i^k) + \boldsymbol{\rho}(\boldsymbol{\eta}^k)_i^\mathsf{T}\mathbf{1}\right)$

  Calculate $\alpha_i = \min\{1, \exp r_i\}$

  Sample $u_i \sim \mathcal{U}(0,1)$

  if $\alpha_i > u_i$

    $\boldsymbol{\eta}_i^{k+1} = \boldsymbol{\eta}_i^*$

  else if $\alpha_i < u_i$

    $\boldsymbol{\eta}_i^{k+1} = \boldsymbol{\eta}_i^k$

  end if

end for

Output: $\boldsymbol{\eta}^{k+1}$

II.2.4 Sampler for the data-poor block

The parameters $(\boldsymbol{\nu}, \boldsymbol{\theta})$ in the data-poor block are, by construction, conditionally independent of $\boldsymbol{y}$ conditioned on the vector $\boldsymbol{\eta}$ from the data-rich block, that is,

$$\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta}) = \pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{\eta}).$$

The conditional posterior density function of the data-poor block is therefore invariant of the choice of likelihood function and proportional to

$$\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta}) \propto \ \pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{II.12}$$

where the conditional density function $\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta})$ is a Gaussian density of the form given in equation (II.6). Moreover, if the Gaussian density functions in the prior assumptions in (II.4) are GMRFs with sparse precision structures then the Gaussian density function in (II.6) retains the sparse GMRF structure induced by the prior assumption, by known results about conditioning of GRMFs (Rue and Held 2005). Fast sampling algorithms for GMRFs can thus be implemented to obtain samples from the Gaussian density function in (II.6), as discussed in Rue (2001).

The relation in (II.12) and the Gaussianity of $\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta})$ in (II.6) motivate the following Metropolis–Hastings based sampling algorithm, which is a modified version of the one block sampler of Knorr-Held and Rue (2002). For some proposal density $q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)$ for the hyperparameters $\boldsymbol{\theta}$, a new proposed value $(\boldsymbol{\nu}^*, \boldsymbol{\theta}^*)$ is generated jointly as follows:

$$\begin{aligned} \boldsymbol{\theta}^* &\sim q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k) \\ \boldsymbol{\nu}^* &\sim \pi(\boldsymbol{\nu}^* \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^*). \end{aligned} \tag{II.13}$$

Denote the proposal density implied by (II.13) with $q(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)$. The proposed value $(\boldsymbol{\nu}^*, \boldsymbol{\theta}^*)$ is then accepted jointly with acceptance probability

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})} \frac{q(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{\nu}^*, \boldsymbol{\theta}^*)}{q(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)}\right\}. \tag{II.14}$$

In Lemma II.4 we show how the acceptance ratio in (II.14) can be simplified, which is modified version of the results shown in Knorr-Held and Rue (2002).

Lemma II.4. *Assume the proposal density implied by (II.13) for the data-poor block, and denote the proposal density with $q(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)$. The corresponding acceptance ratio in (II.14), can be simplified to*

$$\frac{\pi(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})} \frac{q(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{\nu}^*, \boldsymbol{\theta}^*)}{q(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)} = \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^{k+1})} \frac{q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)} \tag{II.15}$$

and is therefore independent of the value of $\boldsymbol{\nu}$.

In other words, the acceptance ratio in (II.15) is only dependent on the acceptance ratio for $\boldsymbol{\theta}$. Further, since the conditional posterior $\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta})$ is a known Gaussian the proposed sampling strategy scales well in terms of computational efficiency as the dimensions of the data-poor component of the latent field $\boldsymbol{\nu}$ increases.

When the Gaussian models in the prior assumptions (II.4) are GMRF density functions with a sparse precision structure, the ratio in (II.15) is computationally costly to calculate directly, since $\pi(\boldsymbol{\theta} \mid \boldsymbol{\eta}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta})$ and $\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta})$ does not necessarily preserve the sparse GMRF structure. However, as the ratio in (II.15) is only dependent on the acceptance ratio for $\boldsymbol{\theta}$ it can be shown that the ratio in (II.15) can be rewritten in order to preserve the sparse GMRF precision structure, as stated in the Theorem II.5.

Theorem II.5. *The term $\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^{k+1})/\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^{k+1})$ in (II.15) can be rewritten as*

$$\frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^{k+1})} = \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^k)} \cdot \frac{\pi(\boldsymbol{\eta}^{k+1} \mid \mathbf{0}, \boldsymbol{\theta}^*)\pi(\mathbf{0} \mid \boldsymbol{\theta}^*)}{\pi(\mathbf{0} \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^*)} \cdot \frac{\pi(\mathbf{0} \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^k)}{\pi(\boldsymbol{\eta}^{k+1} \mid \mathbf{0}, \boldsymbol{\theta}^k)\pi(\mathbf{0} \mid \boldsymbol{\theta}^k)} \tag{II.16}$$

Additionally, the conditional density functions on the right hand side in (II.16) on a logarithmic scale are

$$\begin{aligned}
\log \pi(\boldsymbol{\eta}|\mathbf{0}, \boldsymbol{\theta}) &= \frac{1}{2} \log \det \boldsymbol{Q}_\epsilon - \frac{1}{2} \boldsymbol{\eta}^\mathsf{T} \boldsymbol{Q}_\epsilon \boldsymbol{\eta} + const \\
\log \pi(\mathbf{0}|\boldsymbol{\theta}) &= \frac{1}{2} \log \det \boldsymbol{Q}_\nu + const \\
\log \pi(\mathbf{0}|\boldsymbol{\eta}, \boldsymbol{\theta}) &= \frac{1}{2} \log \det \left(\boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\epsilon \boldsymbol{Z} \right) + \\
&\quad \left(\left(\boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\epsilon \boldsymbol{Z} \right)^{-1} \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\epsilon \boldsymbol{\eta} \right) \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\epsilon \boldsymbol{\eta} + const
\end{aligned} \tag{II.17}$$

Moreover, if the Gaussian prior density functions in (II.4) are GMRFs with sparse precision structures, then all of the conditional density functions on the right hand side of (II.16) are GMRFs with sparse precision structures.

Theorem II.5 shows how the ratio in (II.14) can be calculated with low computational cost by using the results in (II.15), (II.16) and (II.17) in case of GMRFs with sparse precision structures. This is a key result for the implementation of the proposed sampling scheme in the data-poor block for GMRFs with sparse precision structures. The algorithm for the sampling scheme in the data-poor block is is summarized in Algorithm 4.

Algorithm 4 The proposed algorithm for obtaining the $(k+1)$-th sample from $\pi(\boldsymbol{\nu}, \boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\eta})$ in the data-poor block.

Input: $(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k, \boldsymbol{\eta}^{k+1})$

Sample each element of $\boldsymbol{\theta}^*$ from a proposal density $q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)$

Calculate

$$r = \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^k)} \cdot \frac{\pi(\boldsymbol{\eta}^{k+1}|\mathbf{0}, \boldsymbol{\theta}^*)\pi(\mathbf{0}|\boldsymbol{\theta}^*)}{\pi(\mathbf{0}|\boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^*)} \frac{\pi(\mathbf{0}|\boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^k)}{\pi(\boldsymbol{\eta}^{k+1}|\mathbf{0}, \boldsymbol{\theta}^k)\pi(\mathbf{0}|\boldsymbol{\theta}^k)} \cdot \frac{q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)}$$

on a logarithmic scale, using the equations in (II.17) for the conditional posterior densities functions

Calculate $\alpha = \min\{1, r\}$

Sample $u \sim \mathcal{U}(0, 1)$

if $\alpha > u$

  Calcualte $\boldsymbol{Q}_{\nu|\eta} = \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}$

  Sample $\boldsymbol{\nu}^*$ from

$$\boldsymbol{\nu}^*|\boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^* \sim \mathcal{N}\left(\boldsymbol{\nu}^* \middle| \boldsymbol{Q}_{\nu|\eta}^{-1}(\boldsymbol{Q}_\nu\boldsymbol{\mu}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{\eta}^{k+1}), \boldsymbol{Q}_{\nu|\eta}^{-1}\right)$$

  $(\boldsymbol{\nu}^{k+1}, \boldsymbol{\theta}^{k+1}) = (\boldsymbol{\nu}^*, \boldsymbol{\theta}^*)$

else if $\alpha < u$

  $(\boldsymbol{\nu}^{k+1}, \boldsymbol{\theta}^{k+1}) = (\boldsymbol{\nu}^k, \boldsymbol{\theta}^k)$

end if

Output: $(\boldsymbol{\nu}^{k+1}, \boldsymbol{\theta}^{k+1})$

II.3 Examples

Two examples are presented in this section where the MCMC split sampler is applied to obtain posterior samples from the proposed models. In the former example, a data set on annual mean precipitation in Iceland is modeled with a LGM that has a spatial model structure at the latent level. The latter example is on extreme flood events.

We will emphasize that the aim of this section is to present some of the possibilities offered by the MCMC split sampler rather than to claim which model is the best for each data set. The main purpose of the first example is to demonstrate that the MCMC split sampler is well suited to infer LGMs with a spatial models on both location and scale parameters of the data density function, and that the computational efficiency of the sampler scales well as the number of unobserved spatial grid points increases. The main goal of the latter example is show that the MCMC split sampler is designed to infer LGMs with a non-Gaussian three parameter data density function, where all the three parameters are modeled with latent Gaussian models.

II.3.1 Annual mean precipitation in Iceland

The data set analyzed in this section is on observations on annual precipitation from 86 observational sites across Iceland, see Figure II.1, over the years 1962 to 2006. Times series on annual precipitation from the observational sites Reykjavík, Æðey, Akureyri and Kvísker are shown in Figure II.2. The data was provided by the Icelandic Meteorological Office (IMO).

A LGM with a SPDE spatial model structure (Lindgren et al. 2011) at the latent level is presented to obtain the spatially varying distributional properties of annual precipitation over the domain. We will demonstrate that the computational efficiency of the MCMC split sampler scales well as the number of grid points in the mesh in the SPDE approach increase.

Model setup

The data level: The data were modeled with a LGM assuming the Gaussian distribution for the observations and conditional independence over the observational sites. That is, let y_{it} denote the annual precipitation at observational site i at year t then the data density function becomes

$$\begin{aligned}&\pi(y_{it}|\mu_i, \tau_i) = \\ &\mathcal{N}\left(y_{it} \mid \mu_i, \exp(\tau_i)\right), \quad i = 1, \ldots, I, \quad t = 1, \ldots, T\end{aligned}$$

where I is the number of sites, T is the number of years; μ_i and τ_i are mean and log-variance parameters, respectively, which are both allowed to vary spatially.

The latent level: The following model structure was implemented for the mean parameter $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_I)^\mathsf{T}$ at the latent level of the model,

$$\boldsymbol{\mu} = \boldsymbol{X}_\mu \boldsymbol{\beta}_\mu + \boldsymbol{A}_\mu \boldsymbol{u}_\mu + \boldsymbol{\epsilon}_\mu,$$

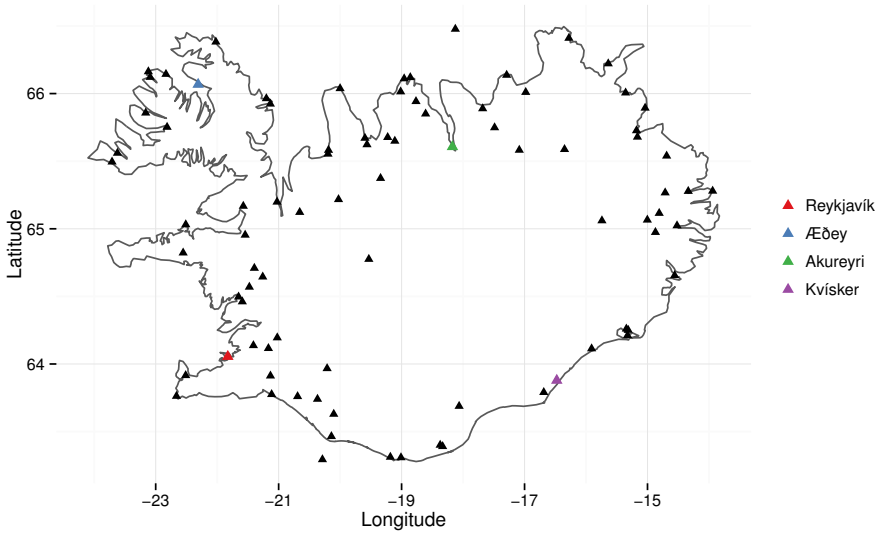

Figure II.1: The $I = 86$ observational sites in Iceland. Reykjavík is marked with red, Æðey is marked with blue, Akureyri is marked with green and Kvísker is marked with purple.

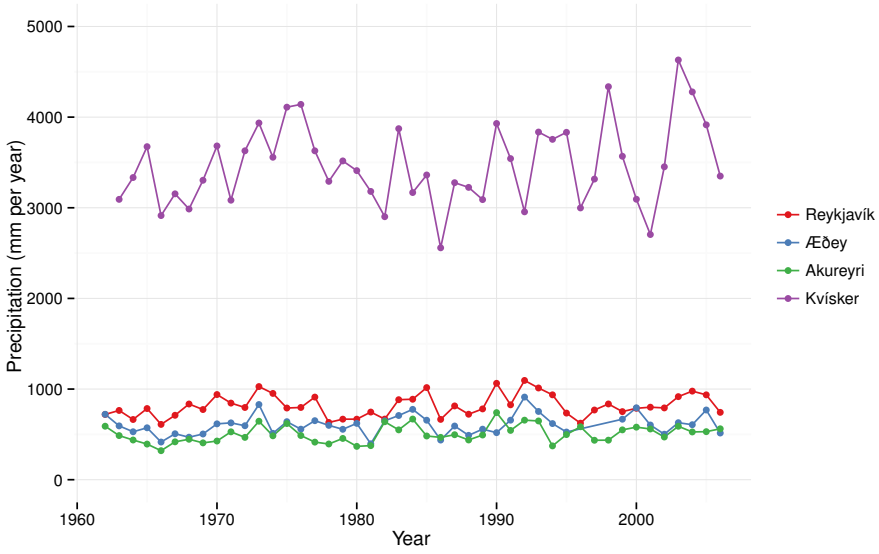

Figure II.2: Times series over the years 1962 to 2006 on annual precipitation. The time series are based on observations from Reykjavík (red curve), Æðey (blue curve), Akueyri (green curve) and Kvísker (purple curve).

where $\boldsymbol{X}_\mu$ is a design matrix consisting of a vector of ones and covariates that are based on the meteorological model of Crochet et al. (2007), see Geirsson et al. (2015) for details; $\boldsymbol{\beta}_\mu$ are the corresponding weights; $\boldsymbol{u}_\mu$ denotes a Matérn type spatial field constructed with the SPDE approach (Lindgren et al. 2011) on a triangulated mesh over the spatial domain with a smoothness parameters chosen as one, which corresponds to an almost once differentiable Matérn field and $\alpha = 2$ in the SPDE method; $\boldsymbol{A}_\mathcal{S}$ is a known projection matrix; the matrix product $\boldsymbol{A}_\mathcal{S}\boldsymbol{u}_\mu$ then denotes the spatial effect at the observational sites, which captures the spatial variation in the data that is unexplained by the covariate and $\boldsymbol{\epsilon}_\mu$ is an unstructured random effect. Analogous model structure is also implemented for the log-variance parameter, that is

$$\boldsymbol{\tau} = \boldsymbol{X}_\tau\boldsymbol{\beta}_\tau + \boldsymbol{A}_\mathcal{S}\boldsymbol{u}_\tau + \boldsymbol{\epsilon}_\tau.$$

where $\boldsymbol{X}_\tau$ is a design matrix consisting of a vector of ones and the aforementioned meteorological covariate on a logarithmic scale.

Working within the LGM setup, the following prior density functions were assigned to parameters at the latent level of the model.

$$\begin{aligned}
\pi(\boldsymbol{\beta}_\mu) &= \mathcal{N}(\boldsymbol{\beta}_\mu \mid \boldsymbol{0}, \kappa_{\beta\mu}^{-1}\boldsymbol{I}), & \pi(\boldsymbol{\beta}_\tau) &= \mathcal{N}(\boldsymbol{\beta}_\tau \mid \boldsymbol{0}, \kappa_{\beta\tau}^{-1}\boldsymbol{I}),\\
\pi(\boldsymbol{u}_\mu) &= \mathcal{N}(\boldsymbol{u}_\mu \mid \boldsymbol{0}, \boldsymbol{Q}_{u\mu}^{-1}), & \pi(\boldsymbol{u}_\tau) &= \mathcal{N}(\boldsymbol{u}_\tau \mid \boldsymbol{0}, \boldsymbol{Q}_{u\tau}^{-1}),\\
\pi(\boldsymbol{\epsilon}_\mu) &= \mathcal{N}(\boldsymbol{\epsilon}_\mu \mid \boldsymbol{0}, \sigma_{\epsilon\mu}^2\boldsymbol{I}), & \pi(\boldsymbol{\epsilon}_\tau) &= \mathcal{N}(\boldsymbol{\epsilon}_\tau \mid \boldsymbol{0}, \sigma_{\epsilon\tau}^2\boldsymbol{I}).
\end{aligned}$$

The parameter values $\kappa_{\beta\mu} = 0.0025$ and $\kappa_{\beta\tau} = 0.25$ were fixed in the prior distributions for $\boldsymbol{\beta}_\mu$ and $\boldsymbol{\beta}_\tau$. The precision matrices $\boldsymbol{Q}_{u\mu}$ and $\boldsymbol{Q}_{u\mu}$ are constructed with SPDE approach, and have sparse GMRF precision structures. Further, the precision matrix $\boldsymbol{Q}_{u\mu}$ has two parameters, $\sigma_{u\mu}$ and $\kappa_{u\mu}$, which serve as hyperparameters of the spatial model for μ. The hyperparameters $\sigma_{u\mu}$ and $\kappa_{u\mu}$ are related to the marginal variance and range of the spatial field, respectively. Analogous structure holds for $\boldsymbol{Q}_{u\tau}$. The parameters $\sigma_{\epsilon\mu}^2$ and $\sigma_{\epsilon\tau}^2$ are unknown variance parameters for the unstructured random effects.

The hyper level: Let $\boldsymbol{\theta}$ denote all the hyper parameters of the model that are not fixed, that is

$$\boldsymbol{\theta} = (\sigma_{u\mu}, \kappa_{u\mu}, \sigma_{\epsilon\mu}, \sigma_{u\tau}, \kappa_{u\tau}, \sigma_{\epsilon\tau})^\mathsf{T}.$$

Lognormal prior distributions with fixed parameters were assigned to the hyperparmeters in $\boldsymbol{\theta}$.

Posterior inference

In order to apply the MCMC split sampler to the aforementioned model, the model parameters are assigned to the data-rich block which includes $\boldsymbol{\eta} = (\boldsymbol{\mu}^\mathsf{T}, \boldsymbol{\tau}^\mathsf{T})^\mathsf{T}$ and the data-poor block which consists of $\boldsymbol{\nu} = (\boldsymbol{\beta}_\mu^\mathsf{T}, \boldsymbol{u}_\mu^\mathsf{T}, \boldsymbol{\beta}_\tau^\mathsf{T}, \boldsymbol{u}_\tau^\mathsf{T})^\mathsf{T}$ and the hyperparameters $\boldsymbol{\theta}$.

The aforementioned model setup and prior assumptions are equivalent to the setup implied in equations (II.3) and (II.4) with

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X}_\mu & \boldsymbol{A}_\mu & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{X}_\tau & \boldsymbol{A}_\tau \end{pmatrix}, \quad \boldsymbol{Q}_\epsilon = \begin{pmatrix} \sigma_{\epsilon\mu}^{-2}\boldsymbol{I} & \cdot \\ \cdot & \sigma_{\epsilon\tau}^{-2}\boldsymbol{I} \end{pmatrix}$$
$$\boldsymbol{Q}_\nu = \begin{pmatrix} \kappa_{\beta\mu}\boldsymbol{I} & \cdot & \cdot & \cdot \\ \cdot & \boldsymbol{Q}_{u\mu} & \cdot & \cdot \\ \cdot & \cdot & \kappa_{\beta\tau}\boldsymbol{I} & \cdot \\ \cdot & \cdot & \cdot & \boldsymbol{Q}_{u\tau} \end{pmatrix}. \tag{II.18}$$

Data-rich block: The conditional posterior $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ in the data-rich block is intractable. However, the logarithm of the conditional posterior of the data-rich block is of the same form as in equation (II.8), with $\boldsymbol{Z}$ and $\boldsymbol{Q}_\epsilon$ defined in equation (II.18) and

$$f(\boldsymbol{\eta}) = \sum_{i=1}^{I} f_i(\boldsymbol{\eta}_i) = \sum_{i=1}^{I} \sum_{t \in \mathcal{A}_i} \log \mathcal{N}(y_{it} | \mu_i, \exp \tau_i), \tag{II.19}$$

where the set $\mathcal{A}_i$ contains the indices of the years t observed at site i. By model assumptions, the vectors $\boldsymbol{\eta}_i = (\mu_i, \tau_i)^\mathsf{T}$ become conditionally independent in the conditional posterior $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ over observational sites i. This demonstrates that the modification of the sampling scheme in Section II.2.3, outlined in Appendix II.A, is applicable. Therefore Algorithm 3 was used to obtain MCMC samples from the conditional posterior $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ with $I = 86$, $\boldsymbol{Q}_\epsilon$ as in (II.18) and $f(\boldsymbol{\eta})$ as in (II.19).

Data-poor block: In order to implement the sampling strategy outlined in Section II.2.4 and to obtain MCMC samples from the conditional posterior $\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta})$, a proposal density q for the hyperparameters $\boldsymbol{\theta}$ must be chosen. In this example, the proposal strategy suggested in (Knorr-Held and Rue 2002) is used for each element of $\boldsymbol{\theta}$. That is, let $\theta_i^* = f\theta_i^k$ where the scaling factor f has the density

$$\pi(f) \propto 1 + 1/f \quad \text{for } f \in [1/F, F] \tag{II.20}$$

where $F > 1$ is a tuning parameter. Knorr-Held and Rue (2002) show that this is a symmetric proposal density in the sense that $q(\theta_i^* | \theta_i^k) = q(\theta_i^k | \theta_i^*)$. Therefore, by using this proposal density, the acceptance probability in equation (II.14) simplifies to

$$\alpha = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^{k+1})} \right\}$$

Moreover, the ratio in (II.16) in Theorem II.5 was used to calculate the acceptance probability, which preserves the sparse GMRF precision structure induced by the SPDE approach. Thus, Algorithm 4 is was implemented to obtain MCMC samples from the conditional posterior $\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta})$, with $\boldsymbol{Z}$, $\boldsymbol{Q}_\epsilon$ and $\boldsymbol{Q}_\nu$ defined in (II.18) and the proposal density in (II.20).

Convergence diagnostics

The following convergence diagnostics are based on four MCMC chains sampled in parallel with the MCMC split sampler from the proposed model. Each chain was calculated with 50000 iterations where 10000 iterations were burned in. The posterior inference was carried out separately for three different mesh resolutions. That is, a coarse resolution based on 411 mesh points; a medium resolution based on 858 mesh points; and a dense resolution based 1752 mesh points. In Figure II.3 the three different meshes are presented on a same scale. The top, middle and bottom panels in Figure II.3 show the coarse resolution, medium resolution and dense resolution meshes, respectively. Runtime, on a modern desktop (Ivy Bridge Intel Core i7-3770K, 16GB RAM and a solid state hard drive), was approximately 6, 6.5 and 7 hours for the coarse, medium and dense mesh resolution, respectively. All calculations were carried out using `R`.

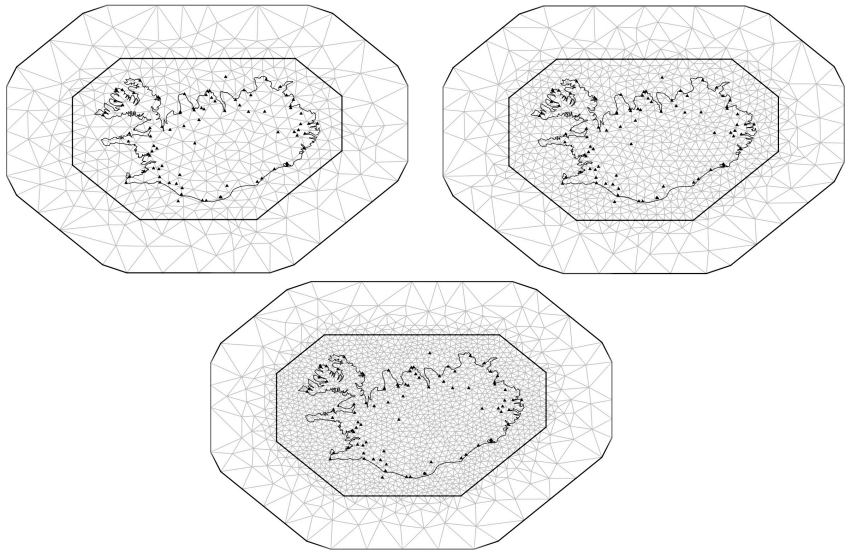

Figure II.3: The triangulated meshes over the spatial domain, based on the coarse mesh (top left), medium mesh (top right) and dense mesh (bottom).

Gelman–Rubin plots, based on the MCMC runs, of the mean parameter μ in Reykjavík; the covariate coefficient $\beta_{\mu 2}$; and the marginal standard deviation for the spatial field $\sigma_{u\mu}$ are shown in the first, second and third column, respectively, in Figure II.4. The results based on the coarse resolution, medium resolution and dense resolution meshes for the aforementioned parameters are shown in the first, second and third row, respectively, in Figure II.4. A comparison of the results in Figure II.4 between the different mesh resolutions reveals that the convergence in the mean is achieved for the three parameters at a similar rate. Furthermore, the Gelman–Rubin plots in Figure II.4 show that the sampler has converged in the mean after roughly 7500 iterations for all mesh resolutions. Similar results hold for all of the other model parameters (results not shown).

Autocorrelation plots for the same set of parameters and arranged identically as in Figure II.4 are shown in Figure II.5. The results demonstrate that the MCMC chains for the mean parameter μ in Reykjavík and the covariate coefficient $\beta_{\mu 2}$ exhibit a negligible autocorrelation after lag 10. The MCMC samples of the hyperparameter $\sigma_{u\mu}$ show autocorrelation around 0.3 at lag 50. Similar results hold for all the other model parameters (results not shown).

Relying on the Gelman-Rubin statistics and the autocorrelation plots, the MCMC chains exhibit all signs of having converged. Moreover, the autocorrelation plots in Figure II.5 reveal that the autocorrelation in the MCMC chains does not increase with number of mesh points, which in turn indicates that the autocorrelation in the MCMC chains is invariant of the dimensions of the data-poor part of the latent field $\boldsymbol{\nu}$. These results demonstrate that the MCMC split sampler retains its computational efficiency when the number of mesh points increases, which is to be expected as the acceptance probability in (II.15) in the data-poor block in independent of the $\boldsymbol{\nu}$.

Furthermore, as the acceptance probability in (II.15) within the data-poor block is only dependent on the hyperparameters, the autocorrelation seen in MCMC chains for the hyperparameter $\sigma_{u\mu}$ in Figure II.5 is mainly affected by the choice of proposal density for $\boldsymbol{\theta}$, which is in this example the sampler in (II.20). In Section II.3.2 we will demonstrate the modularity of the MCMC split sampler, by choosing another proposal density for the hyperparmeters $\boldsymbol{\theta}$ in Algorithm 4 which significantly reduces the autocorrelation in MCMC chains for $\boldsymbol{\theta}$.

II.3.2 Flood analyzis

In this section, we present a simulation study on extreme events. The data set consists of simulations of monthly maximum instantaneous flow based on characteristics of ten river catchments around Iceland. The characterizing features that were used to simulate the data for each river were chosen as river catchment area and maximum daily precipitation, as both river catchment area and maximum daily precipitation are known to be positively correlated with maximum instantaneous flow, see Davíðsson (2015) and Crochet et al. (2012). The simulated time series were chosen to be 150 years.

Model setup

The data level: The data were modeled with a LGM assuming the generalized extreme value distribution (g.e.v.) for the observations. To that extend, let $y_{mj,t}$ denote the value from river j at month m and year t, with a cumulative density function of the form

$$F(y_{mj,t}) = \exp\left\{-\left(1+\xi_{mj}\left(\frac{y_{mj,t}-\mu_{mj}}{\sigma_{mj}}\right)\right)^{-1/\xi_{mj}}\right\}$$

if $1+\xi_{mj}(x-\mu_{mj})/\sigma_i > 0$, $F(y_{it}) = 0$ otherwise. The parameters μ_{mj}, σ_{mj} and ξ_{mj} are the location, scale and shape parameters of the g.e.v. distribution for river j in month m. Additionally, J is the number of rivers and T is the

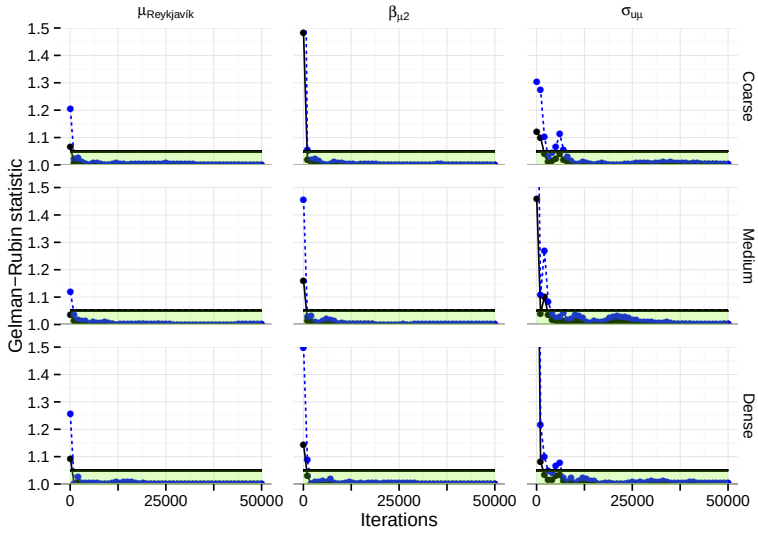

Figure II.4: Gelman–Rubin plots for μ, $\beta_{\mu 2}$, and $\sigma_{u\mu}$ for three different mesh resolutions. The black solid curve denotes the median of the Gelman–Rubin statistics, and the blue dashed curve denotes the upper limit of the 95% confidence interval for the Gelman–Rubin statistics. The first row is based on a coarse resolution (411 mesh points). The second row is based on a medium resolution (858 mesh points). The third row is based on a dense resolution (1752 mesh points). The results demonstrate the MCMC split sampler has converged in the mean after roughly 7500 iterations.

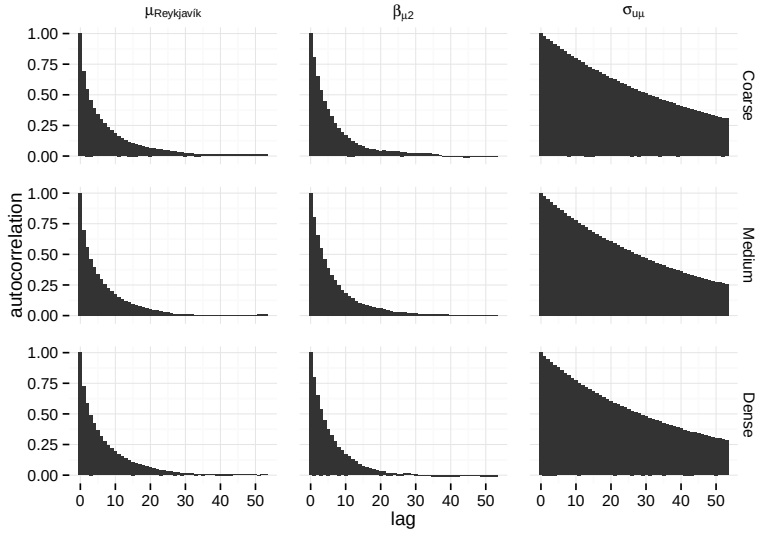

Figure II.5: Autocorrelation plots for μ, $\beta_{\mu 2}$, and $\sigma_{u\mu}$ for three different mesh sizes. The first row is based on a coarse resolution (411 mesh points). The second row is based on a medium resolution (858 mesh points). The third row is based on a dense resolution (1752 mesh points). The results demonstrate that the autocorrelation in the MCMC chains decays rapidly and in invariant of the number of points in the mesh.

number of years. Further, the data is assumed independent between both rivers and between months. These assumptions were made for demonstrative purposes.

The latent level: The location and scale parameters are modeled on a logarithmic scale at the latent level, which is modelling setup along the lines presented in Cunnane and Nash (1971) and GREHY (1996). Thus, define $\lambda_{mj} = \log \mu_{mj}$ and $\tau_{mj} = \log \sigma_{mj}$. The shape parameter is modeled on its native scale.

As discussed in Davíðsson (2015), the underlying processes of monthly maximum instantaneous flow exhibit a seasonal behavior. Therefore, the following seasonal model is proposed for the location parameter on a logarithmic scale. That is,

$$\begin{aligned}\lambda_{mj} = \beta_{0,\lambda} + u_{0,m,\lambda} + x_{1,mj}(\beta_{1,\lambda} + u_{1,m,\lambda}) \\ + \ldots + x_{p,m,j}(\beta_{p,\lambda} + u_{p,m,\lambda}) + \epsilon_{mj,\lambda}\end{aligned} \tag{II.21}$$

where $\beta_{0,\lambda}$ denotes an overall intercept term; $x_{i,mj}$ denotes the i-th covariate in month m at the j-th river; $\beta_{i,\lambda}$ denotes the weight of the i-th covariate for $i = 1, \ldots, p$; $u_{0,m,\lambda}$ denotes the seasonal random effect of the m-th month; $u_{i,m,\lambda}$ denotes the seasonal additional weight of the i-th covariate within month m; and $\epsilon_{mj,\lambda}$ denotes an unstructured random effect.

In order to write the model in a matrix form for the implementation of the MCMC split sampler, combine the location parameters for river j over months. That is,

$$\boldsymbol{\lambda}_j = (\lambda_{1j}, \ldots, \lambda_{12j})^\mathsf{T}, \quad j = 1, \ldots, J$$

and define the following

$$\boldsymbol{u}_{i,\lambda} = (u_{i,1,\lambda}, \ldots, u_{i,12,\lambda})^\mathsf{T}, \boldsymbol{A}_{i,j} = \operatorname{diag}(x_{i,j1}, \ldots, x_{i,j12}),$$

where $i = 0, \ldots, p$ and $x_{0,jm} = 1$ denotes the intercept term for river j and month m. Additionally, define

$$\boldsymbol{X}_j = \begin{pmatrix} 1 & x_{1,j,1} & \ldots & x_{p,j,1} \\ 1 & x_{1,j,2} & \ldots & x_{p,1,2} \\ \vdots & & \vdots & \\ 1 & x_{1,j,12} & \ldots & x_{p,1,12} \end{pmatrix}, \quad \boldsymbol{A}_j = \left(\boldsymbol{A}_{0,j}, \ldots, \boldsymbol{A}_{p,j}\right).$$

The seasonal model presented in (II.21) for the log-location parameter for river j can be written in matrix form as

$$\boldsymbol{\lambda}_j = \boldsymbol{X}_j \boldsymbol{\beta}_\lambda + \boldsymbol{A}_j \boldsymbol{u}_\lambda + \boldsymbol{\epsilon}_{j,\lambda}$$

where $\boldsymbol{\beta}_\lambda = (\beta_{0,\lambda}, \ldots, \beta_{p,\lambda})^\mathsf{T}$, $\boldsymbol{u}_\lambda = (\boldsymbol{u}_{0,\lambda}^\mathsf{T}, \ldots, \boldsymbol{u}_{p,\lambda}^\mathsf{T})^\mathsf{T}$, and $\boldsymbol{\epsilon}_{j,\lambda} = (\epsilon_{1j,\lambda}, \ldots, \epsilon_{12j,\lambda})^\mathsf{T}$. By combing the seasonal model over rivers, the following holds

$$\boldsymbol{\lambda} = \boldsymbol{X}\boldsymbol{\beta}_\lambda + \boldsymbol{A}\boldsymbol{u}_\lambda + \boldsymbol{\epsilon}_\lambda$$

where

$$\boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \vdots \\ \boldsymbol{\lambda}_J \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_J \end{pmatrix}, \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_J \end{pmatrix}, \boldsymbol{\epsilon}_\lambda = \begin{pmatrix} \boldsymbol{\epsilon}_{1,\lambda} \\ \vdots \\ \boldsymbol{\epsilon}_{J,\lambda} \end{pmatrix}.$$

Analogous model structure was also implemented for the log-scale parameter. That is,

$$\boldsymbol{\tau} = \boldsymbol{X}\boldsymbol{\beta}_\tau + \boldsymbol{A}\boldsymbol{u}_\tau + \boldsymbol{\epsilon}_\tau.$$

A reduced model with a similar structure was implemented for the shape parameter ξ. That is

$$\xi_{mj} = \beta_{0,\xi} + u_{0,m,\xi} + \epsilon_{mj,\xi} \tag{II.22}$$

where $\beta_{0,\xi}$ denotes an overall intercept term; $u_{0,m,\xi}$ denotes the seasonal random effect of the m-th month; and $\epsilon_{mj,\xi}$ denotes an unstructured random effect. The full matrix model for ξ becomes

$$\boldsymbol{\xi} = \mathbf{1}_{12J}\beta_{0,\xi} + (\mathbf{1}_J \otimes \boldsymbol{I}_{12})\,\boldsymbol{u}_\xi + \boldsymbol{\epsilon}_\xi$$

where $\mathbf{1}_n$ denotes an n-dimensional vector of ones.

Working within the LGM framework, the following prior density functions were assigned to the latent parameters. First assign,

$$\pi(\boldsymbol{\beta}_\lambda) = \mathcal{N}(\boldsymbol{\beta}_\lambda \mid \mathbf{0}, \sigma^2_{\beta\lambda}\boldsymbol{I}), \quad \pi(\boldsymbol{\beta}_\tau) = \mathcal{N}(\boldsymbol{\beta}_\tau \mid \mathbf{0}, \sigma^2_{\beta\tau}\boldsymbol{I}), \quad \pi(\beta_\xi) = \mathcal{N}(\beta_\xi \mid 0, \sigma^2_{\beta\xi}).$$

The parameters $\boldsymbol{\beta}_\lambda, \boldsymbol{\beta}_\tau$ and β_ξ are assumed *a priori* to have a low precision on their native scales in order to let the data play the dominate role in their inference. Thus, the parameter values $\sigma_{\beta\lambda} = 4$, $\sigma_{\beta\tau} = 4$ and $\sigma_{\beta\xi} = 2$ were chosen for the prior density functions.

Secondly, the selection of prior density functions for the seasonal random effects needs to incorporate a correlation structure that induces a strong correlation between neighbouring months. This is achieved by assigning the following prior density functions

$$\begin{aligned} \pi(\boldsymbol{u}_\lambda) &= \mathcal{N}(\boldsymbol{u}_\lambda \mid \mathbf{0}, \operatorname{diag}(\boldsymbol{\psi}_\lambda) \otimes \boldsymbol{Q}_u^{-1}), \\ \pi(\boldsymbol{u}_\tau) &= \mathcal{N}(\boldsymbol{u}_\tau \mid \mathbf{0}, \operatorname{diag}(\boldsymbol{\psi}_\tau) \otimes \boldsymbol{Q}_u^{-1}), \\ \pi(\boldsymbol{u}_\xi) &= \mathcal{N}(\boldsymbol{u}_\xi \mid \mathbf{0}, \psi_\xi \boldsymbol{Q}_u^{-1}) \end{aligned}$$

where $\boldsymbol{\psi}_\lambda = (\psi_{0,\lambda}, \ldots, \psi_{p,\lambda})^\mathsf{T}$, $\boldsymbol{\psi}_\tau = (\psi_{0,\tau}, \ldots, \psi_{p,\tau})^\mathsf{T}$ and ψ_ξ serve as scaling parameters for the monthly random effects corresponding to the three intercepts and the covariates; and the $\boldsymbol{Q}_u(\kappa)$ is a 12×12 circular band precision matrix that has the vector

$$[1 \quad -2(\kappa^2+2) \quad \kappa^4 + 4\kappa^2 + 6 \quad -2(\kappa^2+2) \quad 1]$$

on the diagonal band, as discussed in Lindgren et al. (2011), which capture the autocorrelation between months. In this example, the decay parameters was fixed to simplify the inference and set equal to $\kappa = 1$. Further, this value of κ induces an autocorrelation *a priori* between consecutive months. Third, for the unstructured random effects, the following priors were chosen.

$$\pi(\boldsymbol{\epsilon}_\lambda) = \mathcal{N}(\boldsymbol{\epsilon}_\lambda \mid \mathbf{0}, \sigma^2_{\epsilon\lambda}\boldsymbol{I}), \quad \pi(\boldsymbol{\epsilon}_\tau) = \mathcal{N}(\boldsymbol{\epsilon}_\tau \mid \mathbf{0}, \sigma^2_{\epsilon\tau}\boldsymbol{I}), \quad \pi(\boldsymbol{\epsilon}_\xi) = \mathcal{N}(\boldsymbol{\epsilon}_\xi \mid \mathbf{0}, \sigma^2_{\epsilon\xi}\boldsymbol{I}).$$

The hyper level: Let $\boldsymbol{\theta}$ denote all the hyperparameters of the model that are not fixed on a logarithmic scale for computational purposes. That is,

$$\boldsymbol{\theta} = (\log\psi_{0,\lambda}, \ldots, \log\psi_{p,\lambda}, \log\psi_{0,\tau}, \ldots, \log\psi_{p,\tau}, \log\psi_\xi, \log\sigma^2_{\epsilon\lambda}, \log\sigma^2_{\epsilon\tau}, \log\sigma^2_{\epsilon\xi})$$

Gaussian prior distributions with fixed parameters were assigned to the hyperparmeters in $\boldsymbol{\theta}$.

Posterior inference

The data-rich block includes $\boldsymbol{\eta}^\mathsf{T} = (\boldsymbol{\mu}^\mathsf{T}, \boldsymbol{\tau}^\mathsf{T}, \boldsymbol{\xi}^\mathsf{T})^\mathsf{T}$ and the data-poor block consists of $\boldsymbol{\nu}^\mathsf{T} = (\boldsymbol{\beta}^\mathsf{T}_\mu, \boldsymbol{u}^\mathsf{T}_\mu, \boldsymbol{\beta}^\mathsf{T}_\tau, \boldsymbol{u}^\mathsf{T}_\tau, \beta^\mathsf{T}_\xi, \boldsymbol{u}^\mathsf{T}_\xi)^\mathsf{T}$ and the hyperparameters $\boldsymbol{\theta}$. For the implementations of the MCMC split sampler, define the following sparse matrices

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{A} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{X} & \boldsymbol{A} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \mathbf{1}_{12J} & \boldsymbol{I}_{12J} \end{pmatrix}, \boldsymbol{Q}_\epsilon = \begin{pmatrix} \sigma^{-2}_{\epsilon\lambda}\boldsymbol{I} & \cdot & \cdot \\ \cdot & \sigma^{-2}_{\epsilon\tau}\boldsymbol{I} & \cdot \\ \cdot & \cdot & \sigma^{-2}_{\epsilon\xi}\boldsymbol{I} \end{pmatrix} \tag{II.23}$$

and

$$\boldsymbol{Q}_\nu = \text{bdiag}\big(\sigma^{-2}_{\beta\lambda}\boldsymbol{I}, \text{diag}(\boldsymbol{\psi}_\lambda) \otimes \boldsymbol{Q}_u^{-1}, \sigma^{-2}_{\beta\tau}\boldsymbol{I}, \text{diag}(\boldsymbol{\psi}_\tau) \otimes \boldsymbol{Q}_u^{-1}, \sigma^{-2}_{\beta\xi}, \psi_\xi \boldsymbol{Q}_u^{-1}\big) \tag{II.24}$$

where bdiag denotes a block diagonal matrix.

Data-rich block: The modified version of the sampling scheme in Section II.2.3, outlined in Appendix II.A, was used to obtain MCMC samples from the conditional posterior $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$. The logarithm of the conditional posterior is of the same form as in equation (II.8), with $\boldsymbol{Z}$ and $\boldsymbol{Q}_\epsilon$ defined in equation (II.23) and

$$f(\boldsymbol{\eta}) = \sum_{m=1}^{12}\sum_{j=1}^{J} f_i(\boldsymbol{\eta}_{mj}) = \sum_{m=1}^{12}\sum_{j=1}^{J}\sum_{t=1}^{T} \log\pi_{\text{gev}}\left(y_{mj,t} \mid \exp\lambda_{mj}, \exp\tau_{mj}, \xi_{mj}\right) \tag{II.25}$$

where π_{gev} denotes the density function of the generalized extreme value distribution. Therefore, Algorithm 3 was used to obtain MCMC samples from the conditional posterior from the data-rich block, with $I = J \cdot 12 = 120$, $\boldsymbol{Q}_\epsilon$ as in (II.23) and $f(\boldsymbol{\eta})$ as in (II.25).

Data-poor block: The sampling scheme outlined in Section II.2.4 was used to obtain MCMC samples from the conditional posterior $\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta})$ in the data-poor block. A proposal density based on the normal distribution centered on the last draw of $\boldsymbol{\theta}$, as discussed in Roberts et al. (1997), was selected for Algorithm 4, with a precision matrix $-c\boldsymbol{H}$ where $\boldsymbol{H}$ is a finite difference estimate of the Hessian matrix of $\log \pi(\boldsymbol{\theta}|\hat{\boldsymbol{\eta}})$ evaluated at the mode. That is,

$$\boldsymbol{H} \approx \nabla^2 \log \pi(\boldsymbol{\theta}|\hat{\boldsymbol{\eta}})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{II.26}$$

where $\hat{\boldsymbol{\eta}}$ is the maximum likelihood estimate of $\boldsymbol{\eta}$ for each river and month; $\boldsymbol{\theta}_0$ is the mode of $\log \pi(\log \boldsymbol{\theta}|\hat{\boldsymbol{\eta}})$; and c is a scaling constant. Conditioning on $\hat{\boldsymbol{\eta}}$, as opposed of $\boldsymbol{\eta}^{k+1}$ for example, removes the necessity to estimate $\boldsymbol{H}$ in every iteration. Moreover, setting a specific scaling c removes the need for tuning. The scaling $c = 2.382/\dim(\boldsymbol{\theta})$ was implemented, as it is optimal in a particular large dimension scenario, see Roberts et al. (1997). The resulting proposal density therefore becomes

$$q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^k) = \mathcal{N}\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k, (-c\boldsymbol{H})^{-1}\right). \tag{II.27}$$

Algorithm 4 was thus implemented to obtain MCMC sampled from the conditional posterior within the data-poor block, with $\boldsymbol{Z}$, $\boldsymbol{Q}_\epsilon$ as in equation (II.23); $\boldsymbol{Q}_\nu$ as in equation (II.24); and the proposal density in (II.27).

Convergence diagnostics

As in Section II.3.1, the following convergence diagnostics are based on four MCMC chains sampled in parallel with the MCMC split sampler. Each chain was calculated with 50000 iterations where 10000 iterations were burned in. Runtime on the same desktop as in Section II.3.1 was approximately 7 hours.

The left panel in Figure II.6 compares the empirical cumulative distribution from river $j = 1$ in January with its posterior cumulative distribution functions based on the MCMC runs. The right panel shows the corresponding probability - probability plots. These result indicate that the model describes the data well. Which in turn demonstrates that the MCMC split sampler recaptures the known underlying model, which was used to generate the simulated data. Analogous results hold across all rivers and months (results not shown).

Furthermore, as the data was generated from a known model setup, the results of the inference based on the MCMC runs can be compared to the known values of the model parameters. In Figure II.7, the known values of the seasonal random effects are shown along with the corresponding 95% posterior intervals. The top panel in Figure II.7 shows this comparison for the seasonal random effect $u_{0,m,\lambda}$ for the log-location parameter λ as a function of months. The middle and the bottom panels in Figure II.7 show the same comparison for $u_{0,m,\tau}$ for the log-scale parameter and $u_{0,m,\xi}$ for the shape parameter ξ, respectively. The results reveal that the 95% posterior intervals for the seasonal random effects contain their known values. These results demonstrate that the MCMC split sampler recaptures the known seasonal random effects. Similar results hold for all other model parameters (results now shown).

Gelman–Rubin plots and auto-correlation plots for nine model parameters based on the MCMC run are shown in Figures II.8 and II.9, respectively. Both plots are based on the same set of parameters and arranged identically. Three parameters were chosen from the location, scale and shape structures of the proposed model which were placed in the first, second and third rows of Figures II.8 and II.9, respectively. The first columns are based on parameters from the data-rich part of the latent field; the second columns are based on parameters from the data-poor part of the latent field; and the third column is based on hyperparamters.

The Gelman–Rubin plots in Figure II.8 show that the sampler has converged in the mean after roughly 10.000 iteration. Similar results hold for all the model parameters (results not shown). Furthermore, the autocorrelation plots in Figure II.9 demonstrate that the MCMC chains for the parameters from both the data-rich and the data-poor parts of the latent field, exhibit a negligible autocorrelation after lag 10. The hyperparameters show a negligible autocorrelation after lag 30. Relying on these results, the MCMC chains exhibit all signs of having converged. Moreover, these results further indicate that the MCMC split sampler, with the modified proposal density of Roberts et al. (1997) implied by equation (II.26) for the hyperparameters, is highly computationally efficient in both the data-rich and data-poor blocks.

II.4 Discussion

The main or novelty of the MCMC split sampler lies in how the proposed blocking scheme leads to conditional posterior structures within in the two blocks, which can be exploited in order to construct computationally efficient sampling schemes for each block. Additionally, the MCMC split sampler is, in principle, designed as a modular sampling scheme in the sense that any MCMC sampling scheme can be implemented within the blocks. Therefore, in the authors' view, the MCMC split sampler presents an interesting area of future research as new sampling schemes for either block can be developed independently of the other block.

In the data-rich block, we proposed a Metropolis–Hastings algorithm with an independence proposal density which was constructed with a Gaussian approximation of the conditional posterior density evaluated at its mode. Furthermore, we proposed a modification the sampler which is applicable if conditional independence assumptions are imposed on the data density function. The modification can potentially increase the computational efficiency of the sampler, as discussed in Appendix II.A.

Although the proposed sampler in the data-rich block is computationally efficient, it is only applicable in practice if the mode of conditional posterior density function can be found, and can be calculated reasonably fast. For example, in the case of models where each observed data point has more than one unique data density parameters associated with it, say of the type $y_i \sim \pi(y_i|\mu_i, \sigma_i)$ for every measurement i, finding the mode of the conditional posterior $\pi(\mu_i, \sigma_i|y_i)$ becomes computationally impractical in some cases. Models of this type include,

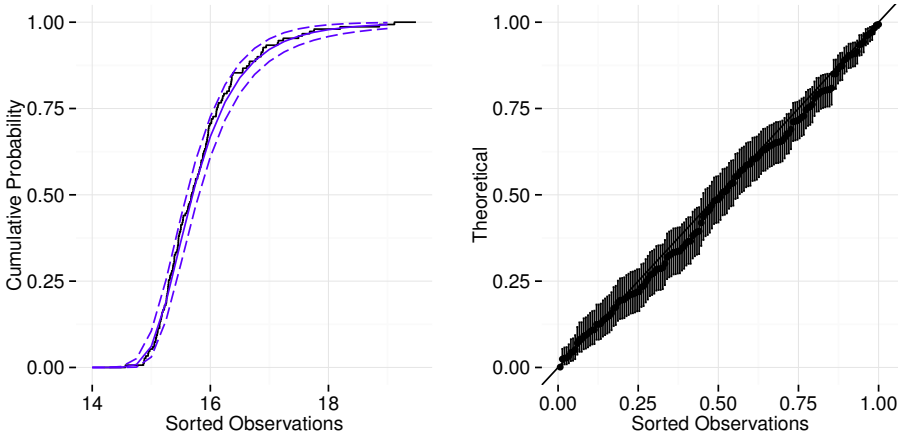

Figure II.6: The left panel shows the empirical cumulative distribution of maximum instantaneous flow from river $j = 1$ in January (black solid curve) and the posterior mean of the corresponding posterior cumulative distribution functions (blue solid curve) and corresponding 95% posterior interval (blue dashed curve). The right panel shows a probability-probability plot of maximum instantaneous flow from river $j = 1$ in January, along with 95% posterior intervals.

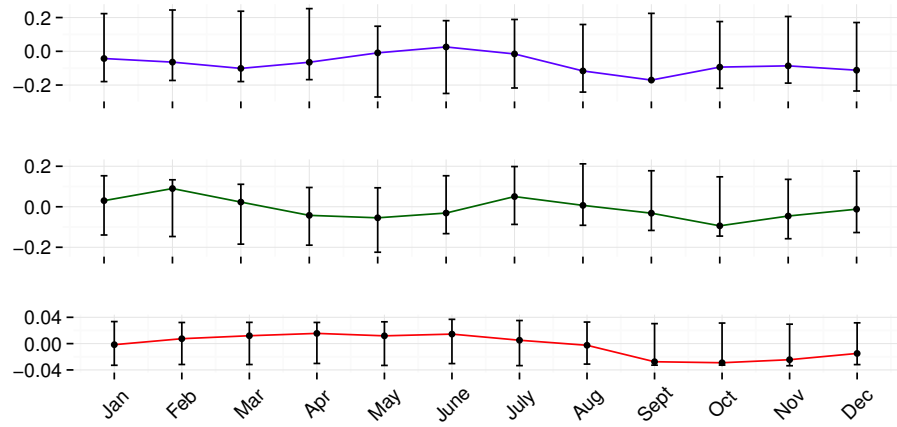

Figure II.7: The top panel in shows the known value (denoted with the blue entries) of the seasonal random effect $u_{0,m,\lambda}$ for the log-location parameter λ as function of months m. The middle panel shows the known value (denoted with the green entries) of the seasonal random effect $u_{0,m,\tau}$ for log-scale parameter τ. The bottom panel shows the known value (denoted with the red entries) of the seasonal random effect $u_{0,m,\xi}$ for the shape parameter ξ. The errors bars in all panels represent the corresponding 95% posterior intervals based on the MCMC-runs.

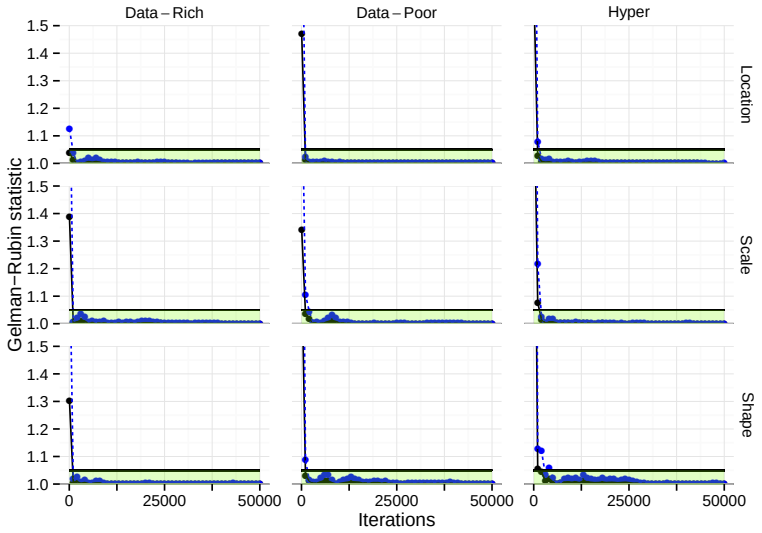

Figure II.8: The figure shows Gelman–Rubin plots based on the MCMC run. The black solid curve denotes the median of the Gelman–Rubin statistics, and the blue dashed curve denotes the upper limit of the 95% confidence interval for the Gelman–Rubin statistics. The first, second and third rows in the first column are based on a randomly chosen log-location parameter λ; covariate coefficient β_λ; and hyperparameter ψ_λ, respectively. The second and third row show an analogous set of parameters based on the log-scale and shape structures, respectively, of the likelihood.

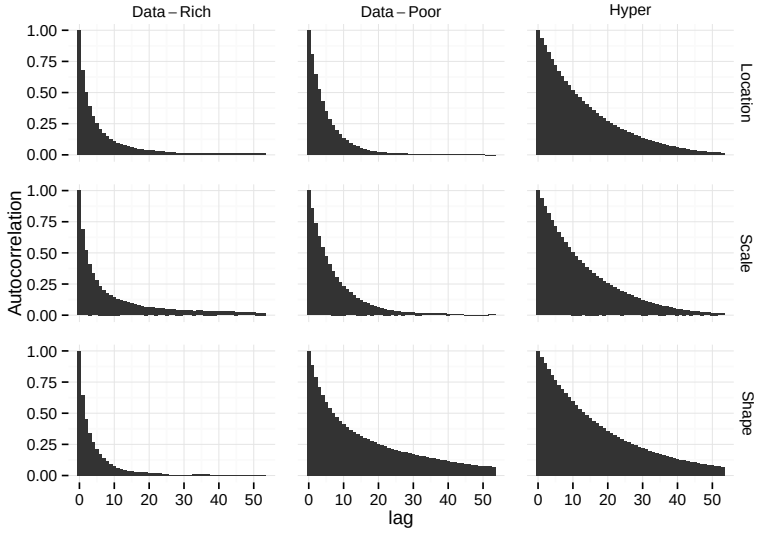

Figure II.9: The figure shows auto-correlation plots based on the MCMC run. The first, second and third rows in the first column are based on a randomly chosen log-location parameter λ; covariate coefficient β_λ; and hyperparameter ψ_λ, respectively. The second and third row show an analogous set of parameters based on the log-scale and shape structures, respectively, of the likelihood.

for example, certain spatial temporal models (Hrafnkelsson et al. 2012). Similar computational issues also arise if data dependence at the data-level of a LGM is desired, see for example Davison et al. (2012) where t-copulas are implemented with g.e.v. marginal density functions at the data level as a model for spatial extremes. However, in both of the aforementioned cases different sampling scheme for the data-rich block can be implemented without changing the sampling scheme of choice in the data-poor block. For example, sampling scheme based on MALA and HMC type algorithm are well suited for the structure of the data-rich block in both cases.

In the data-poor block, the conditional posterior density $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$ is a Gaussian of the form (II.6) and invariant of the data density function. These results serves as one of the main computational advantage introduced by the MCMC split sampler due to the following reasons. First, as the conditional posterior $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$ is Gaussian, a modified version of the one block sampler of Knorr-Held and Rue (2002), which is known to be a highly efficient sampling scheme when applicable (Filippone et al. 2013), becomes applicable within the data-poor block regardless of the data-density function at the data level. As a consequence, computationally efficient sampling algorithms can be used to sample from the exact conditional posterior Gaussian density $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$. Further, if the prior density functions in (II.4) have a sparse GMRF precision structure, then $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$ preserves the sparse structure as discussed in Section II.2.4, which in turn allows for highly efficient sampling algorithms for the Gaussian density $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$. Therefore, the proposed sampling scheme in Section II.2.4 for the data-poor blocks scales well in terms of computational speed and efficiency with increasing dimensions of the data-poor part of the latent field, which is of great importance to achieve, especially in the field of spatial statistics.

Second, as the conditional posterior density $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$ is a known Gaussian and the acceptance rate in the sampling scheme for the data-poor block in Section II.2.4 is only dependent on hyperparameters, the computational efficiency of the proposed sampling scheme for the data-poor block is only dependent of the sampling scheme used for the hyperparameters. In this sense, the sampling scheme in Section II.2.4 is in itself modular, that is, any proposal density for the hyperparmeters is applicable. Choosing a computationally efficient sampling scheme for the hyperparameters can thus increase the computational efficiency of the overall sampling scheme within the data-poor block, as demonstrated in the examples in Section II.3. That is, in Section II.3.1 we demonstrated how a proposal density implied by equation (II.20) may be implemented due to its simplicity. However, in Section II.3.2 we proposed a modified version of the sampling scheme of Roberts et al. (1997) implied by equation (II.26), which reduces the autocorrelation in the MCMC chains. In practical terms, the proposal density implied by (II.26) can also be implemented Section II.3.1, which in turn reduces the autocorrelation in the MCMC chains for the hyperparameters (results omitted).

Due to the modularity of the MCMC split sampler, sampling schemes for the data-rich block can be developed and improved independently of the sampler in the data-poor block, and vice versa. Additionally, as the conditional posterior density $\pi(\boldsymbol{\nu}|\boldsymbol{\eta}, \boldsymbol{\theta})$ becomes invariant of the data in the data-poor block, the com-

putational advantages introduced by the conditional posterior structure in the data-poor block hold for all LGMs. Moreover, the MCMC split sampler can be applied to various LGMs as it is designed to handle LGMs where latent models Gaussian models are imposed on more than just the mean structure of the data density function. Thus, in our view, further developing and improving sampling schemes that utilize the computational advantages introduced by the MCMC spilt sampler presents an interesting area of future research.

Acknowledgements

The authors would like to thank the University of Iceland Doctoral Fund and University of Iceland Research Fund which supported the research. The authors would also like to thank the Icelandic Meteorological Office for providing the data. The authors give their thanks to the Nordic Network on Statistical Approaches to Regional Climate Models for Adaptation (SARMA), especially Prof. Peter Guttorp, for providing travel support. Furthermore, the authors give their thanks to the Department of Mathematical Sciences at the Norwegian University of Science and Technology for hosting Óli Páll Geirsson several times, and special gratitude to Prof. Håvard Rue for his invitation and valuable conversations.

II.A Appendix A: Conditionally independent data density functions

In many applications of LGMs, conditional independence assumptions are imposed on the likelihood function, see for example the discussion in Rue et al. (2009). That is, by the model assumptions there exists a partition of $\boldsymbol{\eta}$ into subvectors $\boldsymbol{\eta}_i$, $i = 1, \ldots, I$, such that

$$\pi(\boldsymbol{y} \mid \boldsymbol{\eta}) = \prod_{i=1}^{I} \pi_i(\boldsymbol{y}_i \mid \boldsymbol{\eta}_i) \tag{II.28}$$

where $\pi_i(\boldsymbol{y}_i|\boldsymbol{\eta}_i)$ denotes the marginal data density functions of the i-th partition. The conditional independence assumptions in (II.28) imply $f(\boldsymbol{\eta}) = \sum_i f_i(\boldsymbol{\eta}_i)$, where f_i is the logarithm of the marginal data density function $\pi_i(\boldsymbol{y}_i|\boldsymbol{\eta}_i)$. As discussed in Section II.2.3, the Gaussian approximation in (II.9) can be, in some scenarios, a poor approximation of the conditional posterior density in some partition i, which in turn may cause the sampler to get stuck and thus lose its efficiency. To address this issue, we suggest a modified version of the sampler proposed in Section II.2.3 which retains the computational speed gained by using a Gaussian approximation as a proposal density and is applicable if conditional independence assumptions are imposed on the data density function. Before we introduce the modifications to the sampling scheme in Section II.2.3, we give a few essential technical results for the modifications, which are as follows.

Lemma II.6. *Assuming the conditional independence assumptions in (II.28). Then, every pair of vectors $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_{i'}$, such that $i \neq i'$, become conditionally independent in the posterior given $(\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$. In other words, the following relation holds*

$$\pi(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\eta}_{-i}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \pi(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) \tag{II.29}$$

for all partitions i. Furthermore, the conditional posterior density of $\boldsymbol{\eta}_i$ is independent of $\boldsymbol{y}_{-i}$, which denotes the subvector of the data vector $\boldsymbol{y}$ that is not observed in partition i. That is, the following relationship holds

$$\pi(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \pi(\boldsymbol{\eta}_i \mid \boldsymbol{y}_i, \boldsymbol{\nu}, \boldsymbol{\theta}) \tag{II.30}$$

for all partitions i

For notational simplicity and based on the relation in (II.29) and (II.30), we denote the conditional posterior density of $\boldsymbol{\eta}_i$ with $\pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ henceforth. The following corollary yields a useful identiy for the conditional posterior $\boldsymbol{\eta}_i$ with $\pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$.

Corollary II.7. *Assume the conditional independence assumptions in (II.28). The logaritym of the conditional posterior density $\pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ is*

$$\log \pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = f_i(\boldsymbol{\eta}_i) - \frac{1}{2}\boldsymbol{\eta}_i^{\mathsf{T}} \boldsymbol{Q}_{\epsilon,(i,i)} \boldsymbol{\eta}_i + \left(\boldsymbol{Q}_{\epsilon,(i,i)} (\boldsymbol{Z}\boldsymbol{\nu})_i\right)^{\mathsf{T}} \boldsymbol{\eta}_i + K \tag{II.31}$$

where K is a constant.

The subsequent corollary is the immediate from Theorem II.2, Lemma II.6 and the relation in (II.31).

Corollary II.8. *Assuming the conditional independence assumptions in (II.28), the Gaussian approximation of the conditional posterior* $\pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ *within each partition* i *is*

$$\tilde{\pi}_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\nu}) = \mathcal{N}\left(\boldsymbol{\eta}_i \mid \boldsymbol{\eta}_i^0, (\boldsymbol{Q}_\epsilon - \boldsymbol{H})_{(i,i)}\right) \tag{II.32}$$

where $\boldsymbol{\eta}_i^0$ *denotes the mode of the marginal posterior density function* $\pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ *and* $(\boldsymbol{Q}_\epsilon - \boldsymbol{H})_{(i,i)}$ *denotes the submatrix of* $(\boldsymbol{Q}_\epsilon - \boldsymbol{H})$ *belonging to partition* i*.*

The next corollary shows the relation between the Gaussian approximation density functions in (II.9) and (II.32).

Corollary II.9. *Assume the conditional independence assumptions in (II.28). Furthermore, let* $\tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ *denote the Gaussian approximation of the conditional posterior density* $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$*, and* $\tilde{\pi}_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ *denote the Gaussian approximation of the conditional posterior* $\pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ *within partition* i*. Then the following relation holds*

$$\tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \prod_i^I \tilde{\pi}_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}). \tag{II.33}$$

The modifications to the sampler proposed in Section II.2.3 are based on the following observations. From the conditional posterior independence relation in (II.29) in Lemma II.6 follows directly

$$\pi(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\eta}_{-i}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \pi(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$$

for all partitions i. It is therefore equivalent to update $\boldsymbol{\eta}_i|\boldsymbol{y}, \boldsymbol{\eta}_{-i}, \boldsymbol{\nu}, \boldsymbol{\theta}$ iteratively over partitions i, with a Gibbs sampling approach using (II.32) as a proposal density and to update $\boldsymbol{\eta}_i|\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}$ separately over partitions i. By updating separately as opposed to iteratively, the number of functions calls is reduced, which in turn reduces computational cost.

However, in practical terms it is faster to compute the mode once by computing the maximum of $\log \pi(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ than computing the maximum of $\log \pi_i(\boldsymbol{\eta}_i|\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ separately in every partition i. This is due to the fact that number of function calls increases as I increases in the numerical optimizing methods when finding the mode of $\log \pi_i(\boldsymbol{\eta}_i|\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ within every partition i. Furthermore, the computational cost of calculating the function $\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ in (II.8) is minimal and scales well as the dimension of the data-poor block increases, as the $\boldsymbol{Q}_\epsilon$ is a diagonal matrix and $\boldsymbol{Z}$ is a fixed sparse matrix. Thus, calculating the gradient and the Hessian matrix of the conditional posterior is also computationally fesiable in many cases.

The relation in (II.33) demonstrates that it is equivalent to propose a new vector $\boldsymbol{\eta}^*$ from the normal approximation $\tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ and to propose new

vectors $\boldsymbol{\eta}_i^*$ separately from $\tilde{\pi}_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ for every i. Therefore, to reduce computational cost and to increase computational efficiency, we propose the following modifications the sampling scheme in Section II.2.3. That is, propose a new vector $\boldsymbol{\eta}^*$ from $q(\boldsymbol{\eta}) = \tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ and accept and reject $\boldsymbol{\eta}_i^*$ within each partition separately with the probability

$$\alpha_i = \min\left\{1, \frac{\pi(\boldsymbol{\eta}_i^* \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})}{\tilde{\pi}(\boldsymbol{\eta}_i^* \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})} \Big/ \frac{\pi(\boldsymbol{\eta}_i^k \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})}{\tilde{\pi}(\boldsymbol{\eta}_i^k \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})}\right\}. \tag{II.34}$$

The acceptance ratio in (II.34) can be calculated separately over partitions i with a low computational cost, as demonstrated in the following corollary.

Corollary II.10. *Assume the conditional independence assumptions in (II.28) and adopt the same notation as in Lemma II.3. The logarithm of the acceptance ratio in (II.34), that is*

$$r_i = \log\left(\frac{\pi(\boldsymbol{\eta}_i^* \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})}{\tilde{\pi}(\boldsymbol{\eta}_i^* \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})} \Big/ \frac{\pi(\boldsymbol{\eta}_i^k \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})}{\tilde{\pi}(\boldsymbol{\eta}_i^k \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})}\right)$$

where $\tilde{\pi}_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ denotes the Gaussian approximation of the conditional posterior $\pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ within partition i. Then r_i can be simplified to

$$r_i = f_i(\boldsymbol{\eta}_i^*) + \boldsymbol{\rho}(\boldsymbol{\eta}^*)_i^\mathsf{T}\mathbf{1} - (f_i(\boldsymbol{\eta}_i^k) + \boldsymbol{\rho}(\boldsymbol{\eta}^k)_i^\mathsf{T}\mathbf{1})$$

for all i, where f_i is the logarithm of the marginal data density function in partition i and

$$\boldsymbol{\rho}(\boldsymbol{\eta}) = \left(\frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{H} + \boldsymbol{b}^\mathsf{T}\right) \circ \boldsymbol{\eta}$$

for notational simplicity, where $\circ$ denotes an entrywise multiplication. The corresponding acceptance probability in partition i is thus

$$\alpha_i = \min\{1, \exp r_i\}.$$

The sampling scheme for the data-rich block with the aforementioned modifications is summarised in Algorithm 4.

II.B Appendix B: Proofs

II.B.1 Proof of Lemma II.1

Proof. Assume the following prior assumptions

$$\begin{aligned}\boldsymbol{\eta} \mid \boldsymbol{\nu} &\sim \mathcal{N}\left(\boldsymbol{Z}\boldsymbol{\nu}, \boldsymbol{Q}_\epsilon^{-1}\right),\\ \boldsymbol{\nu} &\sim \mathcal{N}\left(\boldsymbol{\mu}_\nu, \boldsymbol{Q}_\nu^{-1}\right).\end{aligned}$$

analogous to the setup in (II.4). Since

$$\pi(\boldsymbol{\eta}, \boldsymbol{\nu}) = \pi(\boldsymbol{\eta} \mid \boldsymbol{\nu})\pi(\boldsymbol{\nu})$$

it is known that $(\boldsymbol{\eta}^\mathsf{T}, \boldsymbol{\nu}^\mathsf{T})^\mathsf{T}$ is a Gaussian random vector. In order to find the mean vector and covariance matrix, note that

$$\begin{aligned}
\pi(\boldsymbol{\eta}, \boldsymbol{\nu}) &= \pi(\boldsymbol{\eta} \mid \boldsymbol{\nu})\pi(\boldsymbol{\nu}) \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{Z}\boldsymbol{\nu})^\mathsf{T}\boldsymbol{Q}_\epsilon(\boldsymbol{\eta} - \boldsymbol{Z}\boldsymbol{\nu}) - \frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\mu}_\nu)^\mathsf{T}\boldsymbol{Q}_\nu(\boldsymbol{\nu} - \boldsymbol{\mu}_\nu)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\eta}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{\eta} + \boldsymbol{\nu}^\mathsf{T}\boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}\boldsymbol{\nu} + \boldsymbol{\nu}^\mathsf{T}\boldsymbol{Q}_\nu\boldsymbol{\nu}\right) + \boldsymbol{\eta}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}\boldsymbol{\nu} + (\boldsymbol{Q}_\nu\boldsymbol{\mu}_\nu)^\mathsf{T}\boldsymbol{\nu}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix}^\mathsf{T}\begin{pmatrix}\boldsymbol{Q}_\epsilon & -\boldsymbol{Q}_\epsilon\boldsymbol{Z}\\ -\boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon & \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}\end{pmatrix}\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix} + \begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{Q}_\nu\boldsymbol{\mu}_\nu\end{pmatrix}^\mathsf{T}\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix}\right\}.
\end{aligned}$$

By known results about block inverses it is clear that

$$\begin{pmatrix}\boldsymbol{Q}_\epsilon & -\boldsymbol{Q}_\epsilon\boldsymbol{Z}\\ -\boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon & \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}\end{pmatrix}^{-1} = \begin{pmatrix}\boldsymbol{Q}_\epsilon^{-1} + \boldsymbol{Z}\boldsymbol{Q}_\nu^{-1}\boldsymbol{Z}^\mathsf{T} & \boldsymbol{Z}\boldsymbol{Q}_\nu^{-1}\\ \boldsymbol{Q}_\nu^{-1}\boldsymbol{Z}^\mathsf{T} & \boldsymbol{Q}_\nu^{-1}\end{pmatrix}.$$

that serves as the covariance matrix of the joint prior distribution. The mean vector of the joint prior distribution is therefore

$$\begin{pmatrix}\boldsymbol{Q}_\epsilon^{-1} + \boldsymbol{Z}\boldsymbol{Q}_\nu^{-1}\boldsymbol{Z}^\mathsf{T} & \boldsymbol{Z}\boldsymbol{Q}_\nu^{-1}\\ \boldsymbol{Q}_\nu^{-1}\boldsymbol{Z}^\mathsf{T} & \boldsymbol{Q}_\nu^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{Q}_\nu\boldsymbol{\mu}_\nu\end{pmatrix} = \begin{pmatrix}\boldsymbol{Z}\boldsymbol{\mu}_\nu\\ \boldsymbol{\mu}_\nu\end{pmatrix}$$

The joint posterior distribution of $(\boldsymbol{\eta}, \boldsymbol{\nu})$ is thus

$$\begin{aligned}
\pi\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix} &= \mathcal{N}\left(\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix}\middle|\begin{pmatrix}\boldsymbol{Z}\boldsymbol{\mu}_\nu\\ \boldsymbol{\mu}_\nu\end{pmatrix}, \begin{pmatrix}\boldsymbol{Q}_\epsilon & -\boldsymbol{Q}_\epsilon\boldsymbol{Z}\\ -\boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon & \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}\end{pmatrix}^{-1}\right) \\
&= \mathcal{N}\left(\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\nu}\end{pmatrix}\middle|\begin{pmatrix}\boldsymbol{Z}\boldsymbol{\mu}_\nu\\ \boldsymbol{\mu}_\nu\end{pmatrix}, \begin{pmatrix}\boldsymbol{Q}_\epsilon^{-1} + \boldsymbol{Z}\boldsymbol{Q}_\nu^{-1}\boldsymbol{Z}^\mathsf{T} & \boldsymbol{Z}\boldsymbol{Q}_\nu^{-1}\\ \boldsymbol{Q}_\nu^{-1}\boldsymbol{Z}^\mathsf{T} & \boldsymbol{Q}_\nu^{-1}\end{pmatrix}\right)
\end{aligned}$$

The conditional distribution $\boldsymbol{\nu}$ conditioned on $\boldsymbol{\eta}$ follows directly from Lemma 2.1 in (Rue and Held 2005), that is,

$$\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}) = \mathcal{N}\left(\boldsymbol{\nu}\middle|\boldsymbol{Q}_{\nu|\eta}^{-1}(\boldsymbol{Q}_\nu\boldsymbol{\mu}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{\eta}), \boldsymbol{Q}_{\nu|\eta}^{-1}\right)$$

where $\boldsymbol{Q}_{\nu|\eta} = \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{Z}$. □

II.B.2 Proof of Theorem II.2

Proof. The conditional posterior density function $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ is proportional to the product of the data density function and the conditional Gaussian prior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{\nu}, \boldsymbol{\theta})$ given by (II.4), that is

$$\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) \propto \pi(\boldsymbol{y} \mid \boldsymbol{\eta})\pi(\boldsymbol{\eta} \mid \boldsymbol{\nu}, \boldsymbol{\theta}).$$

Thus, the logarithm of the conditional posterior density function is given by

$$\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = f(\boldsymbol{\eta}) - \frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{\eta} + (\boldsymbol{Q}_\epsilon\boldsymbol{Z}\boldsymbol{\nu})^\mathsf{T}\boldsymbol{\eta} + \text{const}$$

where $f(\boldsymbol{\eta}) = \log \pi(\boldsymbol{y} \mid \boldsymbol{\eta})$ for notational convenience. The second order Taylor approximation of $f(\boldsymbol{\eta})$ expanded around the mode $\boldsymbol{\eta}^0$ of the conditional posterior $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ is

$$\begin{aligned} f(\boldsymbol{\eta}) &\approx f(\boldsymbol{\eta}^0) + \nabla f(\boldsymbol{\eta}^0)^\intercal (\boldsymbol{\eta} - \boldsymbol{\eta}^0) + \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}^0)^\intercal \boldsymbol{H} (\boldsymbol{\eta} - \boldsymbol{\eta}^0) \\ &= \frac{1}{2}\boldsymbol{\eta}^\intercal \boldsymbol{H} \boldsymbol{\eta} + (\nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H}\boldsymbol{\eta}^0)^\intercal \boldsymbol{\eta} + \text{const.} \end{aligned}$$

Consequently, the second order Taylor approximation of $\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ expanded around $\boldsymbol{\eta}^0$ becomes

$$\begin{aligned} \log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) &\approx \frac{1}{2}\boldsymbol{\eta}^\intercal \boldsymbol{H} \boldsymbol{\eta} + (\nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H}\boldsymbol{\eta}^0)^\intercal \boldsymbol{\eta} - \frac{1}{2}\boldsymbol{\eta}^\intercal \boldsymbol{Q}_\epsilon \boldsymbol{\eta} + (\boldsymbol{Q}_\epsilon \boldsymbol{Z}\boldsymbol{\nu})^\intercal \boldsymbol{\eta} + \text{const} \\ &= -\frac{1}{2}\boldsymbol{\eta}^\intercal \left(\boldsymbol{Q}_\epsilon - \boldsymbol{H}\right) \boldsymbol{\eta} + (\boldsymbol{Q}_\epsilon \boldsymbol{Z}\boldsymbol{\nu} + \boldsymbol{b})^\intercal \boldsymbol{\eta} + \text{const}, \end{aligned}$$

where $\boldsymbol{b} = (\nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H}\boldsymbol{\eta}^0)$. This derivation yields a Gaussian approximation with a mean vector

$$\left(\boldsymbol{Q}_\epsilon - \boldsymbol{H}\right)^{-1} (\boldsymbol{Q}_\epsilon \boldsymbol{Z}\boldsymbol{\nu} + \boldsymbol{b})$$

and covariance matrix $\left(\boldsymbol{Q}_\epsilon - \boldsymbol{H}\right)^{-1}$. However, as the vector $\boldsymbol{\eta}^0$ is the mode of the conditional posterior function $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ the following relation holds

$$\nabla \log \pi(\boldsymbol{\eta}^0 \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \nabla f(\boldsymbol{\eta}^0) - \boldsymbol{Q}_\epsilon \boldsymbol{\eta}_0 + (\boldsymbol{Q}_\epsilon \boldsymbol{Z}\boldsymbol{\nu})^\intercal = \boldsymbol{0}.$$

The mean of the Gaussian approximations becomes

$$\begin{aligned} &\left(\boldsymbol{Q}_\epsilon - \boldsymbol{H}\right)^{-1} (\boldsymbol{Q}_\epsilon \boldsymbol{Z}\boldsymbol{\nu} + \boldsymbol{b}) \\ &= \left(\boldsymbol{Q}_\epsilon - \boldsymbol{H}\right)^{-1} (\boldsymbol{Q}_\epsilon \boldsymbol{Z}\boldsymbol{\nu} + \nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H}\boldsymbol{\eta}^0) \\ &= \left(\boldsymbol{Q}_\epsilon - \boldsymbol{H}\right)^{-1} (\boldsymbol{Q}_\epsilon \boldsymbol{\eta}^0 - \boldsymbol{H}\boldsymbol{\eta}^0) \\ &= \left(\boldsymbol{Q}_\epsilon - \boldsymbol{H}\right)^{-1} (\boldsymbol{Q}_\epsilon - \boldsymbol{H})\boldsymbol{\eta}^0 = \boldsymbol{\eta}^0 \end{aligned}$$

Thus, a Gaussian approximation of the conditional posterior density function $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ evaluated at the mode $\boldsymbol{\eta}^0$ is given by

$$\tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{\eta} \mid \boldsymbol{\eta}^0, \left(\boldsymbol{Q}_\epsilon - \boldsymbol{H}\right)^{-1}\right).$$

□

II.B.3 Proof of Lemma II.3

Proof. The logarithm of the acceptance ratio given in (II.10) is

$$r = \log \frac{\pi(\boldsymbol{\eta}^* \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) q(\boldsymbol{\eta}^k)}{\pi(\boldsymbol{\eta}^k \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) q(\boldsymbol{\eta}^*)} \tag{II.35}$$

where $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ is the conditional posterior density function given in (II.4) and $q(\boldsymbol{\eta})$ is the proposal density based on the Gaussian approximation in (II.9). The right hand side term in (II.35) can be written as

$$\begin{aligned}&\log \pi(\boldsymbol{\eta}^* \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) - \log q(\boldsymbol{\eta}^*) \\ &\quad - \left(\log \pi(\boldsymbol{\eta}^k \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) + \log q(\boldsymbol{\eta}^k)\right)\end{aligned}$$

Since the proposal density q is based on the Gaussian approximation in (II.9) the following holds

$$\begin{aligned}\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) - \log q(\boldsymbol{\eta}) &= f(\boldsymbol{\eta}) - \frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{\eta} + (\boldsymbol{Q}_\epsilon\boldsymbol{Z}\boldsymbol{\nu})^\mathsf{T}\boldsymbol{\eta} \\ &\quad - \left(\frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{H}\boldsymbol{\eta} + \boldsymbol{b}^\mathsf{T}\boldsymbol{\eta} - \frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{\eta} + (\boldsymbol{Q}_\epsilon\boldsymbol{Z}\boldsymbol{\nu})^\mathsf{T}\boldsymbol{\eta}\right) + \text{const} \\ &= f(\boldsymbol{\eta}) - \left(\frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{H}\boldsymbol{\eta} + \boldsymbol{b}^\mathsf{T}\boldsymbol{\eta}\right) + \text{const}\end{aligned}$$

which yields the results in (II.11). □

II.B.4 Proof of Lemma II.4

Proof. By definition of the proposal density in (II.13) the following holds

$$\frac{q(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{\nu}^*, \boldsymbol{\theta}^*)}{q(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)} = \frac{\pi(\boldsymbol{\nu}^k \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^k) q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{\pi(\boldsymbol{\nu}^* \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^*) q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)},$$

where $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^k)$ is some proposal density for $\boldsymbol{\theta}$ and $\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta})$ is the conditional Gaussian density function in (II.6) in Lemma II.1. Therefore, the acceptance ratio in (II.14) can be written as

$$\frac{\pi(\boldsymbol{\nu}^*, \boldsymbol{\theta}^* \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\nu}^k, \boldsymbol{\theta}^k \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})} \frac{\pi(\boldsymbol{\nu}^k \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^k)}{\pi(\boldsymbol{\nu}^* \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^*)} \frac{q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)} = \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^{k+1})} \frac{q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)} \tag{II.36}$$

since $\pi(\boldsymbol{\nu}, \boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\eta}) = \pi(\boldsymbol{\nu}, \boldsymbol{\theta}|\boldsymbol{\eta})$, as discussed in Section II.2.4, and $\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{\eta})/\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta}) = \pi(\boldsymbol{\theta} \mid \boldsymbol{\eta})$ for any $\boldsymbol{\eta}, \boldsymbol{\nu}$ and $\boldsymbol{\theta}$. The result in (II.36) demonstrates that the acceptance ratio in (II.14) is only dependant on $\boldsymbol{\theta}$ within in the proposed setup. In other words, the acceptance ratio in (II.14) becomes independent of the value of $\boldsymbol{\nu}$. □

II.B.5 Proof of Theorem II.5

Proof. In order to rewrite $\pi(\boldsymbol{\theta} \mid \boldsymbol{\eta})$ in (II.15) we use the relation

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{\eta}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta}). \tag{II.37}$$

Further, by the law of conditional probability, the following holds

$$\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\eta}, \boldsymbol{\nu} \mid \boldsymbol{\theta})}{\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta})} = \frac{\pi(\boldsymbol{\eta} \mid \boldsymbol{\nu}, \boldsymbol{\theta})\pi(\boldsymbol{\nu} \mid \boldsymbol{\theta})}{\pi(\boldsymbol{\nu} \mid \boldsymbol{\eta}, \boldsymbol{\theta})} \tag{II.38}$$

As $\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta})$ is independent of the value of $\boldsymbol{\nu}$, it follows that the two ratios in (II.38) are invariant of the choice of $\boldsymbol{\nu}$. In particular, the following holds

$$\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\eta} \mid \mathbf{0}, \boldsymbol{\theta})\pi(\mathbf{0} \mid \boldsymbol{\theta})}{\pi(\mathbf{0} \mid \boldsymbol{\eta}, \boldsymbol{\theta})} \tag{II.39}$$

by choosing the value $\boldsymbol{\nu} = \mathbf{0}$. Combining (II.37) and (II.39) yields

$$\begin{aligned}
&\frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^{k+1})}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^{k+1})} = \frac{\pi(\boldsymbol{\theta}^*)\pi(\boldsymbol{\eta}^{k+1} \mid \boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^k)\pi(\boldsymbol{\eta}^{k+1} \mid \boldsymbol{\theta}^k)} \\
&= \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^k)} \times \frac{\pi(\boldsymbol{\eta}^{k+1} \mid \mathbf{0}, \boldsymbol{\theta}^*)\pi(\mathbf{0} \mid \boldsymbol{\theta}^*)}{\pi(\mathbf{0} \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^*)} \times \frac{\pi(\mathbf{0} \mid \boldsymbol{\eta}^{k+1}, \boldsymbol{\theta}^k)}{\pi(\boldsymbol{\eta}^{k+1} \mid \mathbf{0}, \boldsymbol{\theta}^k)\pi(\mathbf{0} \mid \boldsymbol{\theta}^k)}.
\end{aligned}$$

Moreover, if the Gaussian prior density functions in (II.4) are GMRFs with sparse precision structures, then all of the conditional density functions on the rightmost side of (II.38) are GMRFs with sparse precision structures, by known results about conditioning on subvectors as demonstrated in Theorem 2.5 in Rue and Held (2005).

□

II.B.6 Proof of Lemma II.6

Proof. As the matrix $\boldsymbol{Q}_\epsilon$ is a diagonal matrix the following holds

$$\begin{aligned}
\pi(\boldsymbol{\eta} \mid \boldsymbol{\nu}, \boldsymbol{\theta}) &= \mathcal{N}\left(\boldsymbol{\eta} \middle| \boldsymbol{Z}\boldsymbol{\nu}, \boldsymbol{Q}_\epsilon^{-1}\right) \\
&= \prod_{i=1}^{I} \mathcal{N}\left(\boldsymbol{\eta}_i \middle| (\boldsymbol{Z}\boldsymbol{\nu})_i, \boldsymbol{Q}_{\epsilon,(i,i)}^{-1}\right)
\end{aligned} \tag{II.40}$$

where $\boldsymbol{Q}_{\epsilon,(i,i)}$ denotes the submatrix of $\boldsymbol{Q}_\epsilon$ belonging to partition i. Let

$$\pi_i(\boldsymbol{\eta}_i | \boldsymbol{\nu}, \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{\eta}_i \middle| (\boldsymbol{Z}\boldsymbol{\nu})_i, \boldsymbol{Q}_{\epsilon,(i,i)}^{-1}\right)$$

which serves as the conditional prior density function for the data-rich part of the latent field belonging to partition i. The relation in (II.40) along with the conditional independence assumptions in (II.28) yield

$$\begin{aligned}
\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) &\propto \pi(\boldsymbol{y} \mid \boldsymbol{\eta})\pi(\boldsymbol{\eta} \mid \boldsymbol{\nu}, \boldsymbol{\theta}) \\
&= \prod_{i=1}^{I} \pi_i(\boldsymbol{y}_i \mid \boldsymbol{\eta}_i)\pi_i(\boldsymbol{\eta}_i | \boldsymbol{\nu}, \boldsymbol{\theta})
\end{aligned}$$

which demonstrates that the vectors $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_{i'}$ are conditionally independent in the conditional posterior given $(\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$, for all $i \neq i'$.

Furthermore, the conditional independence assumptions in (II.28) also yield

$$\begin{aligned}\pi(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) &\propto \pi(\boldsymbol{y}_i \mid \boldsymbol{\eta}_i)\pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{\nu}, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\eta}_i \mid \boldsymbol{y}_i, \boldsymbol{\nu}, \boldsymbol{\theta}) \end{aligned} \tag{II.41}$$

which demonstrates that $\pi(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ is independent of $\boldsymbol{y}_{-i}$. □

II.B.7 Proof of Corollary II.7

Proof. The conditional independence assumptions in (II.28), the relation in (II.40) and the relation in (II.41) yield

$$\begin{aligned}\log \pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}_i, \boldsymbol{\nu}, \boldsymbol{\theta}) &= \log \pi(\boldsymbol{y}_i \mid \boldsymbol{\eta}_i) + \log \pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{\nu}, \boldsymbol{\theta}) + K \\ &= f_i(\boldsymbol{\eta}_i) - \frac{1}{2}\boldsymbol{\eta}_i^\mathsf{T} \boldsymbol{Q}_{\epsilon,(i,i)}\boldsymbol{\eta}_i + \left(\boldsymbol{Q}_{\epsilon,(i,i)}(\boldsymbol{Z}\boldsymbol{\nu})_i\right)^\mathsf{T} \boldsymbol{\eta}_i + K\end{aligned}$$

for every partition i. □

II.B.8 Proof of Corollary II.8

Proof. The result follows by using Theorem II.2 on the $\log \pi_i(\boldsymbol{\eta}_i \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$ given in (II.31) instead of $\log \pi(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$. □

II.B.9 Proof of Corollary II.9

Proof. The relation in (II.29) and (II.30) in Lemma (II.6) implies that the elements of mode $\boldsymbol{\eta}^0$ belonging to partition i are also the mode of $\pi_i(\boldsymbol{\eta}_i|\boldsymbol{y}_i, \boldsymbol{\nu}, \boldsymbol{\theta})$ in every partition i. The results then follows from Lemma II.6, Theorem II.2 and Corollary II.8. □

II.B.10 Proof of Corollary II.10

Proof. As conditional independence are imposed over partitions i, Lemma II.6 yields the following

$$\begin{aligned}\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\nu}) - \log \tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) &= f(\boldsymbol{\eta}) - \frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{H}\boldsymbol{\eta} - \boldsymbol{b}^\mathsf{T}\boldsymbol{\eta} + \text{const} \\ &= \sum_i \left(f_i(\boldsymbol{\eta}_i) - \frac{1}{2}\boldsymbol{\eta}_i^\mathsf{T}\boldsymbol{H}_{(i,i)}\boldsymbol{\eta}_i - \boldsymbol{b}_i^\mathsf{T}\boldsymbol{\eta}_i \right) + \text{const}.\end{aligned}$$

The results follows by similar derivations as in the proof of Lemma II.3. □

III

Paper III

Computationally efficient spatial modeling of annual maximum 24-h precipitation on a fine grid

Óli Páll Geirsson, Birgir Hrafnkelsson & Daniel Simpson

Geirsson Ó.P., Hrafnkelsson B. and Simpson D. 2015 Computationally efficient spatial modeling of annual maximum 24-h precipitation on a fine grid. *Environmetrics*, 2015, DOI: 10.1002/env.2343

Abstract

A computationally efficient statistical method is proposed to obtain distributional properties of annual maximum 24 hour precipitation on a 1 km by 1 km regular grid over Iceland. A covariate based on a local meteorological model which captures information on the physical processes of precipitation is constructed, providing an additional spatial information on maximum precipitation. A latent Gaussian model is built which takes into account observed maximum precipitation, the covariate based the local meteorological model and spatial variations. The observations are assumed to follow the generalized extreme value distribution, where spatial models based on approximate solutions to stochastic partial differential equations (SPDE) are implemented for the location, scale and shape parameters of the likelihood. An efficient MCMC sampler which exploits the sparse matrices induced by the SPDE modeling is implemented, yielding continuous spatial predictions for spatially varying model parameters and quantiles.

III.1 Introduction

Obtaining distributional properties of maximum precipitation is important in order to properly plan for extreme precipitation events. Furthermore, this information is needed at a fine resolution as the characteristics of extreme precipitation events can be local, especially in regions with heterogeneous topography. Therefore, statistical spatial modeling for this type of spatial data becomes challenging when observational sites are scarce relative to changes in the topography, and because uncertainties associated with parameter estimates become large when the observational data set is relatively small.

The aim of this paper is to provide information about the characteristics of extreme precipitation on a local scale, in particular, where limited spatial information from relatively sparse observational sites is available. To address the sparsity of the observational sites, a method is proposed which leverages information from an external meteorological model which provides spatial information about mean precipitation on a fine grid. The leveraged information is thus based on scientific knowledge on the physical processes of precipitation and not directly on observations on extreme precipitation. This is a novel extension on the concept presented in Benestad et al. (2012), which states that observed mean precipitation provides information on observed extreme precipitation. To demonstrate the use of the proposed method for a particular region, which is Iceland in this paper, information is extracted from a meteorological model proposed by Smith and Barstad (2004) which Crochet et al. (2007) have adapted to an Icelandic data set on a 1 km by 1 km grid. A covariate based on the meteorological model of Crochet et al. (2007) is constructed at each grid point in order to assimilate much of the scientific knowledge about mean precipitation in this region into a statistical model for maximum precipitation.

Various statistical models have been proposed for modeling extreme precipitation, where models based on the generalized extreme value distribution are standard in the literature, see for example Sang and Gelfand (2009). Furthermore,

the Bayesian approach is well suited to quantify uncertainty of the underlying physical processes as further argued in Tebaldi and Sansó (2009). In particular, the spatial variation can be modeled through the likelihood parameters as presented in e.g. Davison et al. (2012) and Hrafnkelsson et al. (2012). Alternative modeling approaches have been explored, such as the peaks over threshold methods with the generalized Pareto distribution (Cooley et al. 2007).

Latent Gaussian models (LGMs) (Rue and Held 2005), which form a flexible and practical subclass of Bayesian hierarchical models, play a dominant role in the vast literature on spatial statistics (Chiles and Delfiner 2009, Diggle et al. 1998, Guttorp and Gneiting 2006). In the LGM framework, Gaussian fields appear at the latent level of the hierarchical model. However, posterior inference for Gaussian fields becomes increasingly computationally demanding as data sets get larger. Gaussian fields can be approximated by Gaussian Markov random fields (GMRFs), which increase the speed of computation significantly (Rue 2001, Rue and Held 2005). LGMs with a GMRF structure are computationally beneficial modeling options in a spatial setting, see for example Schliep et al. (2010), where a latent Gaussian model with a GMRF structure is implemented in order to model spatial extremes.

Although GMRFs are computationally efficient they become difficult to parameterize in a spatial setting. In Lindgren et al. (2011) numerical approximation solutions to Gaussian fields with a Matérn covariance structure are presented. The method constructs an approximation solution of a stochastic partial differential equation (SPDE) on a triangulated mesh. The approximate solutions can then be used to construct a GMRF representation of the desired Gaussian field on the mesh. This allows for continuous spatial predictions.

A computationally efficient Bayesian hierarchical spatial model for maximum precipitation is presented in this paper which integrates the covariates based the meteorological model of Crochet et al. (2007); observations on annual maximum precipitation; and the spatial variation of precipitation in order to obtain distributional properties of extreme precipitation on a fine grid. To that extent, a LGM is proposed where the observations are assumed to follow the generalized extreme value distribution. The spatial variations are modeled through the location, scale and shape parameters of the likelihood with SPDE spatial models. The SPDE spatial models are implemented on a triangulated grid over the spatial domain. An MCMC split sampler, see **Paper II**, is applied to the model structure, yielding an efficient inference scheme. Furthermore, the model structure can also be used to make spatial predictions for the model parameters and quantiles of extreme precipitation on the fine grid.

The following summarizes the two main novel c ontributions of this paper. First, there is the method of leveraging information on the physical processes of precipitation from the local meteorological model, in order to yield additional spatial information on maximum precipitation. The proposed method and the covariate construction is extendable to any regions in the world where local meteorological models are available. Secondly, although LGMs with a GMRF structure have been proposed to model extreme precipitation (Schliep et al. 2010, Cooley and Sain 2010), LGMs with spatial models based on the SPDE approach have

not been proposed in the literature on spatial extreme precipitation before. The proposed modeling strategy is general in the sense that it is extendable to any spatial domain of interest.

The paper is organized as follows. The data and the meteorological model are presented in Section 2, where the method for constructing covariates from outputs of the meteorological model is also outlined. Detailed description of the model structure, the inference method and spatial prediction is given in Section 3. Posterior results are presented and discussed in Section 4. The paper concludes with a discussion in Section 5.

III.2 The data

III.2.1 Observations

The observed data on precipitation were provided by the Icelandic Meteorological Office (IMO). The dataset contains observations of 24 hour annual maximum precipitation from 40 observational sites in Iceland over the years 1958 to 2006, as seen in (Crochet et al. 2007). The locations of the observational sites can be seen in Figure III.1. The observations have been corrected according to a dynamic correction method proposed by (Førland and Hanssen-Bauer 2000), which Crochet et al. (2007) adapted to the Icelandic dataset. The correction method accounts for trace; wetting and evaporation losses; and for the catch deficiencies due to aerodynamic effects and is applied to daily precipitation observations from every observational site. Time series from four observational sites, Reykjavík, Æðey, Akureyri and Kvísker, are shown in Figure III.2. Reykjavík and Akureyri were chosen because they are the most populated areas in Iceland. Æðey was chosen due to its geological position. Finally, Kvísker was chosen as it has the highest observed precipitation.

III.2.2 The meteorological model

The meteorological model, that is used in this paper to leverage spatial information on precipitation, is based on a linear orographic precipitation model proposed by Smith and Barstad (2004). Crochet et al. (2007) have adapted the method to precipitation in Iceland. The model is driven by coarse resolution precipitation, wind and temperature data obtained from re-analyses (1958-2001) (ERA-40) (Uppala et al. 2005) and analyses made by the European Center for Medium Range Weather Forecast (2002-2004). The model takes into account the topography of the spatial domain; airflow dynamics; condensed water advection; and downslope evaporation. This means that outputs from the model contain information about the underlying physical processes of precipitation. The resulting model, herby referred to as the meteorological model, simulates daily precipitation on a 1 km by 1 km regular grid, of the size 521 km $\times$ 361 km, across Iceland for the years 1958-2002.

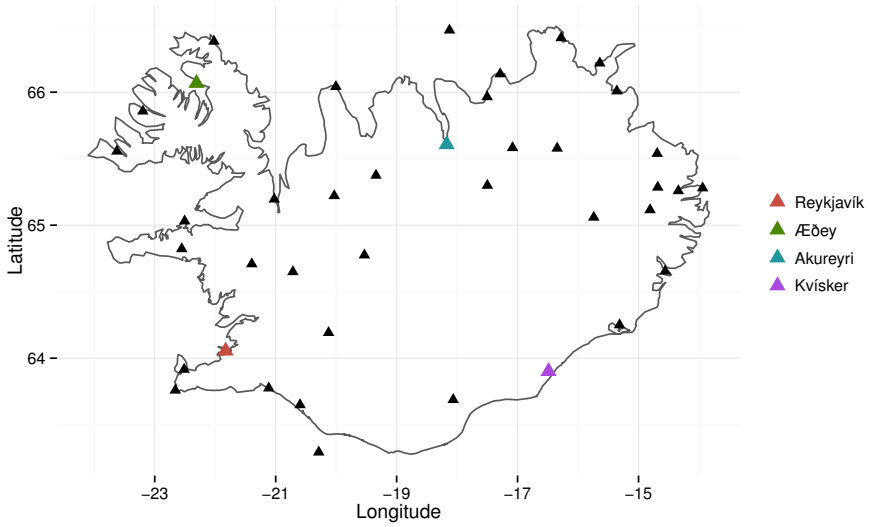

Figure III.1: The location of the observational sites. The observational sites in Reykjavík, Æðey, Akureyri and Kvísker have been highlighted.

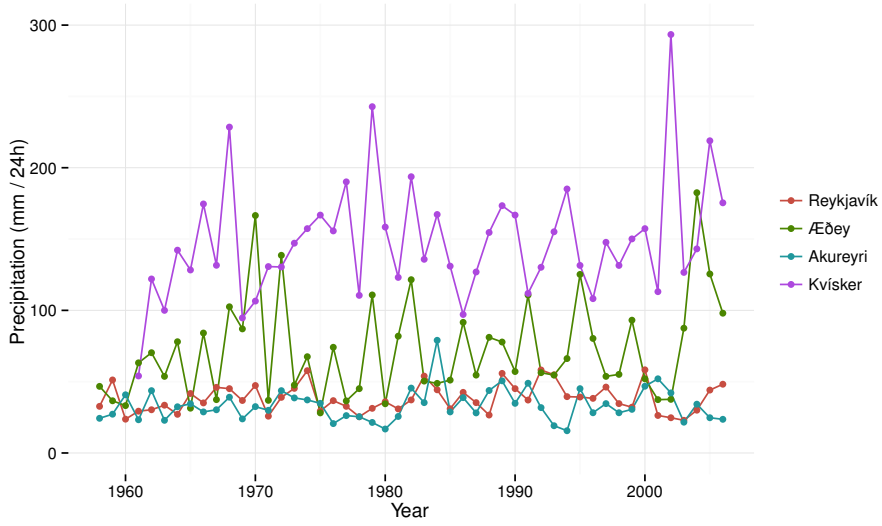

Figure III.2: Observed annual maximum precipitation at the observational sites in Reykjavík, Æðey, Akureyri and Kvísker as a function of years.

III.2.3 Covariates

Cooley et al. (2007) demonstrated that covariates which assimilate available spatial information, such as the topography of the domain and the underlying physical processes of precipitation, are reasonable covariates for extreme precipitation. By constructing covariates based on the outputs from the meteorological model, the information about the above factors can be assimilated. Furthermore, Benestad et al. (2012) suggested that observed mean values of precipitation have high predictive power for maximum precipitation. Extending their argument, we propose that calculated sample mean values from the outputs of the meteorological model serve as reliable predictors for extreme precipitation. This is a novel extension of the concept of Benestad et al. (2012) to spatial prediction of extreme precipitation. The sample means based on the meteorological model are referred to hereafter as the simulated means or the meteorological covariate. However, as observational sites are not necessarily at the regular grid points a spatial smother was implemented in order construct the covariates at the observational sites, see Section S1.1 in the supplementary material for details on the implementation of the smoother. Furthermore, for notational convenience let $\mathcal{G}$ denote the set of every regular grid point and let $\mathcal{S}$ denote the set of the observational sites.

The simulated means are calculated over the entire time period of the meteorological model at every grid point in $\mathcal{G}$, yielding one simulated mean value for each grid point. The simulated means serve as a non-time varying spatial covariate for the average climate, in terms of precipitation, at every grid point in $\mathcal{G}$. As observations of precipitation demonstrate temporal stationary over the time period 1958-2006 (results not shown), it is reasonable to assume that the constructed covariate is also representative for the observed years 2003-2006.

To emphasize, the information stemming from the meteorological model is primarily based on scientific knowledge on the physical process of precipitation. For further calibrations of the meteorological model, Crochet et al. (2007) used six years worth of data on observed daily precipitation and observed monthly means, not observed annual 24 hour maximum precipitation. The resulting constructed covariate is thus related to the observed mean of the time period. Extending the argument made by Benestad et al. (2012), the constructed covariate thus serves to describe the spatial variation of climate, in terms of precipitation, which is then relevant for describing extreme precipitation.

III.3 The model and inference

Guttorp and Gneiting (2006) proposed that using Gaussian fields with a Matérn covariance structure, called Matérn fields, is a flexible and interpretable way of modeling underlying physical processes of natural phenomena in a spatial setting. The Matérn covariance function between locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ in $\mathbb{R}^d$ is defined as

$$c(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\kappa \|\boldsymbol{s_1} - \boldsymbol{s_1}\|\right)^{\nu} K_{\nu} \left(\kappa \|\boldsymbol{s_1} - \boldsymbol{s_2}\|\right) \tag{III.1}$$

where Γ is the Gamma function; K_ν is the modified Bessel function of the second kind of order $\nu > 0$; $\kappa > 0$ is a scaling parameter; and σ^2 denotes the marginal variance of the Matérn field (Stein 1999). However, for any set of n locations the resulting covariance matrix is dense and thus becomes computationally demanding in posterior inference and spatial predictions as datasets get larger.

Approximating Gaussian fields as Gaussian Markov random fields (GMRF), see for example (Knorr-Held and Rue 2002), provides an efficient framework for computationally efficient Gaussian models. GMRFs are parameterized with precision matrices, which are defined as inverses of covariance matrices. Although GMRFs have very good computational properties, there was no standard way to parametrize the precision matrix of a GMRF to achieve a predefined spatial covariance structure until Lindgren et al. (2011) addressed the issue with the introduction of spatial models based on approximate solutions to stochastic partial differential equations (SPDE). The SPDE spatial models are briefly are summarized below.

A Gaussian field $x(\boldsymbol{s})$ with a Matérn covariance function, as in equation (III.1), is a solution to the stochastic partial differential equation

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

The operator $(\kappa^2 - \Delta)^{\alpha/2}$ serves as a pseudo-differential operator; $\mathcal{W}$ is a spatial Gaussian white noise with a unit variance; and Δ is the Laplacian. Lindgren et al. (2011) proposed constructing a finite element representation of the solution to the SPDE in (III.2) on a triangulated mesh over a spatial domain of interest. The proposed approximate solution $u(\boldsymbol{s})$ is, for each point $\boldsymbol{s}$ in the spatial domain, on the form

$$u(\boldsymbol{s}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{s}_k) w_k \tag{III.3}$$

where n is the number of vertices in the mesh, ψ_k are piecewise linear basis functions; and w_k are Gaussian weights at the vertices of the triangles in the mesh. In Lindgren et al. (2011), the authors show that approximate solutions on the form (III.3) can be constructed such that the Gaussian weights w_k have sparse precision matrix structures, and thus making the approximate solutions computationally beneficial to work with. Moreover, approximate solutions can be obtained at each point within every triangle using the linear basis function and the surrounding Gaussian weights. The methods thus forms a GMRF representation of the Matérn field at every point within the mesh. A description of the implementation of the SPDE method is outlined below in Section III.3.1.

In order to model the behavior of the underlying physical processes of extreme precipitation, the location, scale and shape parameters in the likelihood are allowed to vary in space. To that extent, SPDE spatial models are used at the latent level of the proposed hierarchical model to describe continuously the spatial variation of these three latent parameters. Posterior inference and spatial predictions based on this approach also are presented below.

III.3.1 Model structure

The data are modeled with a LGM assuming the generalized extreme value distribution for the observations, where the observations are assumed to be conditionally independent. That is, let y_{it} denote the annual maximum 24 hour precipitation at station i at year t, with a cumulative density function of the form

$$F(y_{it}) = \exp\left\{-\left(1+\xi_i\left(\frac{y_{it}-\mu_i}{\sigma_i}\right)\right)^{-1/\xi_i}\right\}, \quad i=1,\ldots,J, \quad t=1,\ldots,T$$

if $1+\xi_i(x-\mu_i)/\sigma_i > 0$, $F(y_{it}) = 0$ otherwise. The parameters μ_i, σ_i and ξ_i are location, scale and shape parameters; J is the number of observational sites; and T is the number of years. These distributional assumptions are reasonable as the generalized extreme value distribution belongs to the family of extremal distributions and have desired asymptotic properties (Coles et al. 2001).

A preliminary statistical analysis revealed a linear relationship between the ML estimates of the location parameter μ and the meteorological covariate at the observational sites, see top left panel in Figure III.4. Therefore, the following spatial model structure was implemented for the location parameter $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_J)^{\mathsf{T}}$,

$$\boldsymbol{\mu} = \boldsymbol{X}_\mu\boldsymbol{\beta}_{\boldsymbol{\mu}} + \boldsymbol{A}_{\mathcal{S}}\boldsymbol{u}_\mu + \boldsymbol{v}_\mu, \tag{III.4}$$

where $\boldsymbol{X}_\mu$ is a design matrix consisting of a vector of ones and the meteorological covariate; $\boldsymbol{\beta}_\mu$ are the corresponding weights; $\boldsymbol{u}_\mu$ is a spatial effect on a triangulated mesh; $\boldsymbol{A}_{\mathcal{S}}$ is a known projection matrix from the vertices of the mesh onto the observational sites; the matrix product $\boldsymbol{A}_{\mathcal{S}}\boldsymbol{u}_\mu$ denotes the spatial effect at the observational sites which captures the spatial variation in the data that is unexplained by the covariate; and $\boldsymbol{v}_\mu$ is an unstructured random effect with a variance $\sigma^2_{v\mu}$.

The structure of the spatial effect $\boldsymbol{u}_\mu$ is based on the SPDE approach. In order to obtain the SPDE structure for $\boldsymbol{u}_\mu$, the spatial domain is subdivided into a mesh of non-intersecting triangles, see Figure III.3. The GMRF representation is then constructed for the weights of the basis functions in (III.3), which are denoted with $\boldsymbol{u}_\mu$ in (III.4), with a precision matrix $\boldsymbol{Q}_{u\mu}$ with the SPDE approach of Lindgren et al. (2011). The precision matrix is based on the geometry of the mesh and is, by construction, a sparse matrix. In this paper, we chose the smoothness parameter to be $\nu = 1$, which corresponds to an almost once differentiable Matérn field and $\alpha = 2$ in the SPDE method. The structure of the resulting precision matrix $\boldsymbol{Q}_{u\mu}$ can be seen in Lindgren et al. (2011). The matrix has two parameters, $\kappa_{u\mu}$ and $\omega_{u\mu}$, which enter the model as hyperparameters. The hyperparameter $\kappa_{u\mu}$ is inversely proportional to the range of the approximate Matérn field and the hyperparameter $\omega_{u\mu}$ is related to the marginal variance of the spatial effect $\boldsymbol{u}_\mu$.

The spatial locations of $\boldsymbol{u}_\mu$ are on the vertices of the triangles in the mesh, which are not necessarily at the observational sites. However, since the approximate representation (III.3) is assumed to have piecewise linear basis functions, a

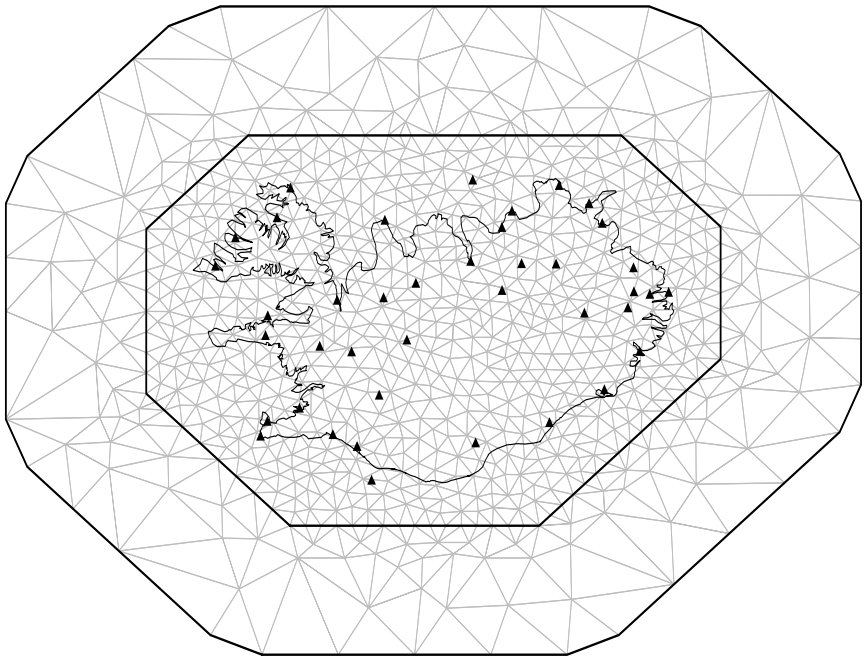

Figure III.3: Triangulated mesh over Iceland

corresponding approximate solution can also be found within each triangle as a convex linear combination of the approximate solutions at the surrounding vertices. The spatial effect $\boldsymbol{u}_\mu$ can be projected linearly from the vertices of the mesh onto every point within every triangle of the mesh. The matrix $\boldsymbol{A}_\mathcal{S}$ denotes the linear projection onto the observational sites. Note that the number of non-zero entries in each line of $\boldsymbol{A}_\mathcal{S}$ is three and the sum of every line is one.

The preliminary statistical analysis also revealed a linear relationship between the ML estimates of the scale parameter σ and the meteorological covariate on a logarithmic scale, see top right panel in Figure III.4. Therefore, a similar model structure was implemented for the scale parameter on a logarithmic scale. That is, let $\tau_i = \log \sigma_i$ and then model $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_J)^\mathsf{T}$ as

$$\boldsymbol{\tau} = \boldsymbol{X}_\tau \boldsymbol{\beta}_\tau + \boldsymbol{A}_\mathcal{S} \boldsymbol{u}_\tau + \boldsymbol{v}_\tau.$$

The design matrix $\boldsymbol{X}_\tau$ consists of a vector of ones and the meteorological covariate on a logarithmic scale; $\boldsymbol{u}_\tau$ is a spatial random effect with the same structure as $\boldsymbol{u}_\mu$ and has parameters $\kappa_{u\tau}$ and $\omega_{u\tau}$; and $\boldsymbol{v}_\tau$ is an unstructured random effect with a variance $\sigma^2_{v\tau}$.

The preliminary statistical analysis revealed that ML estimates for the shape parameter ξ vary among the observational sites. However, no evidence was seen for a linear relationship between the ML estimates of ξ and the meteorological covariate, height above sea level, see Figure III.4 bottom row, or other available geological covariates (results now shown). Thus, the following spatial model was

implemented for the shape parameter

$$\boldsymbol{\xi} = \boldsymbol{X}_\xi \beta_\xi + \boldsymbol{A}_\mathcal{S} \boldsymbol{u}_\xi + \boldsymbol{v}_\xi$$

where $\boldsymbol{X}_\xi$ is vector of ones; the weight β_ξ serves as an overall effect for the shape parameter over the spatial domain; $\boldsymbol{u}_\xi$ is a spatial random effect that captures the spatial deviation from the overall effect with the same structures as $\boldsymbol{u}_\mu$ with parameters $\kappa_{u\xi}$ and $\omega_{u\xi}$; and $\boldsymbol{v}_\xi$ is an unstructured random effect with variance a $\sigma^2_{v\xi}$.

To summarize the model structure, a directed acyclic graph of the above model components can be seen in Figure III.5.

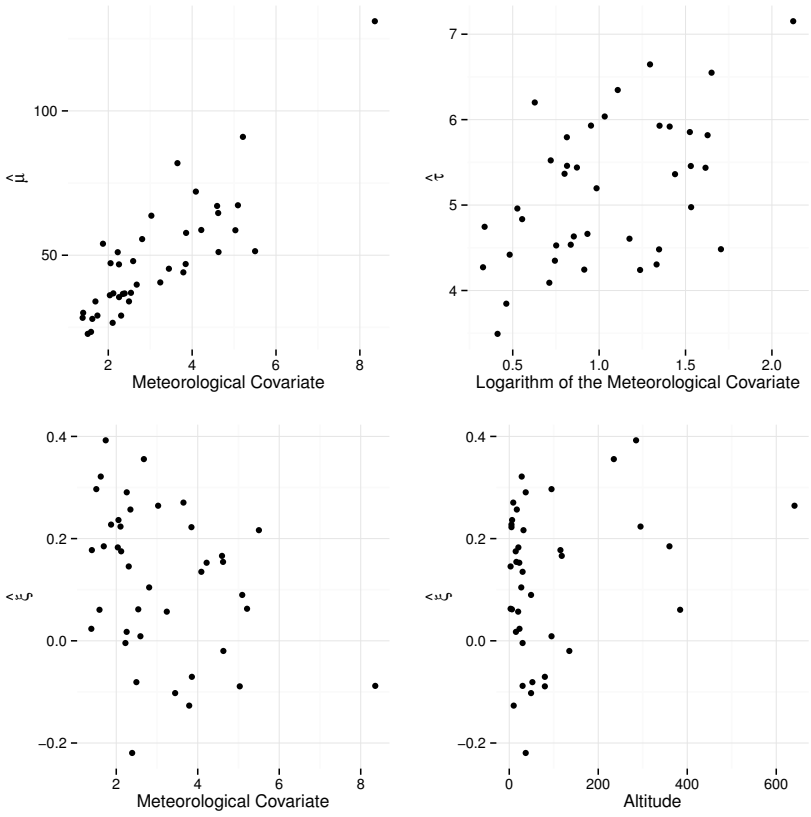

Figure III.4: The top row shows the ML estimates $\hat{\mu}$ and $\hat{\tau}$ as functions of the meteorological covariate and the logarithm of the meteorological covariate, respectively. The bottom row shows the ML estimate $\hat{\xi}$ as a function of the meteorological covariate and altitude over sea level, respectively.

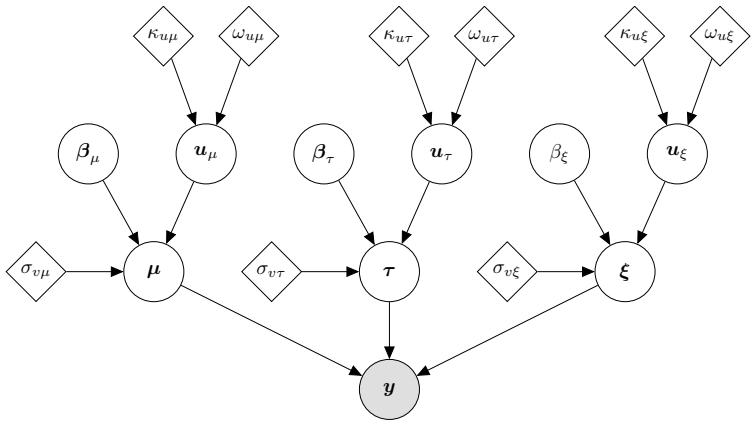

Figure III.5: Directed acyclic graph of the stochastic model components of proposed model. The grey circled note denotes the data vector; the white circled notes denote the model parameters; and the rhombus shaped notes denote the hyperparameters.

III.3.2 Prior selection

The following prior distributions were assigned,

$$\begin{aligned}
&\boldsymbol{\beta}_\mu \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\beta\mu}\boldsymbol{I}), && \boldsymbol{\beta}_\tau \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\beta\tau}\boldsymbol{I}), && \beta_\xi \sim \mathcal{N}(0, \sigma^2_{\beta\xi})\\
&\boldsymbol{u}_\mu \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}^{-1}_{u\mu}), && \boldsymbol{u}_\tau \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}^{-1}_{u\tau}), && \boldsymbol{u}_\xi \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}^{-1}_{u\xi}),\\
&\boldsymbol{v}_\mu \sim \mathcal{N}(\mathbf{0}, \sigma^2_{v\mu}\boldsymbol{I}), && \boldsymbol{v}_\tau \sim \mathcal{N}(\mathbf{0}, \sigma^2_{v\tau}\boldsymbol{I}) && \boldsymbol{v}_\xi \sim \mathcal{N}(\mathbf{0}, \sigma^2_{v\xi}\boldsymbol{I}).
\end{aligned}$$

The parameters $\boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\tau$ and β_ξ are assumed *a priori* to have a low precision on their native scales in order to let the data play the dominate role in their inference. Thus, the parameter values $\sigma_{\beta\mu} = 20$, $\sigma_{\beta\tau} = 4$ and $\sigma_{\beta\xi} = 2$ were chosen for the prior distributions.

Lognormal prior distributions with fixed parameters were assigned to all the remaining hyperparmeters. That is,

$$\begin{aligned}
&\kappa_{u\mu} \sim \mathcal{LN}\left(-2.50, 0.5^2\right), && \kappa_{u\tau} \sim \mathcal{LN}\left(-1.85, 0.5^2\right), && \kappa_{u\xi} \sim \mathcal{LN}\left(-1.85, 0.5^2\right)\\
&\omega_{u\mu} \sim \mathcal{LN}\left(-1.00, 0.25^2\right), && \omega_{u\tau} \sim \mathcal{LN}\left(2.00, 0.1^2\right), && \omega_{u\xi} \sim \mathcal{LN}\left(2.00, 0.1^2\right)\\
&\sigma_{v\mu} \sim \mathcal{LN}\left(1.85, 0.45^2\right), && \sigma_{v\tau} \sim \mathcal{LN}\left(-1.75, 0.30^2\right), && \sigma_{v\xi} \sim \mathcal{LN}\left(-2.00, 0.30^2\right)
\end{aligned}$$

where $\mathcal{LN}$ denotes the lognormal distribution. The selection of prior distributions is further discussed in Section S1.2 in the supplementary material.

III.3.3 Posterior inference

Due to the proposed model structure, in particular, the spatial model for the scale and shape parameters, MCMC methods were necessary to make posterior inference as opposed to approximation methods such as INLA (Rue et al. 2009). However, standard MCMC methods like single site updating converged slowly and

mixed poorly as many model parameters were heavily correlated in the posterior. To address this issue, posterior inference was done by using the MCMC split sampler in **Paper II**.

The MCMC split sampler is a block Gibbs sampling strategy, where Metropolis–Hastings steps are implemented within each block. This sampling strategy is well suited for inferring models that have spatial models entering the likelihood not only through the location parameter, but also through the scale and shape parameters of the likelihood, like the proposed model in Section III.3.1. The details on the implementation of the sampling strategy are summarized in Section S2 in the supplementary material. Further details on the MCMC split sampler can be seen in in **Paper II**.

III.3.4 Spatial prediction

By using the SPDE spatial model structure, the posterior distribution of all the spatially varying model parameters can be obtained at every regular grid point in $\mathcal{G}$. The details are as follows.

Let $\boldsymbol{u}^{[k]}$ be the k-th posterior MCMC sample of the spatial effects for the location, scale or shape parameters. The spatial effects $\boldsymbol{u}^{[k]}$ are located at the vertices of the triangles in the mesh, which do not necessarily coincide with the regular grid points in $\mathcal{G}$. However, since every regular grid point belongs to some triangle in the mesh, the k-th posterior MCMC sample for the spatial effects at the regular grid points can be obtained with a convex linear combination of $\boldsymbol{u}^{[k]}$, that is,

$$\boldsymbol{u}_{\mathcal{G}}^{[k]} = \boldsymbol{A}_{\mathcal{G}}\boldsymbol{u}^{[k]}. \tag{III.5}$$

where the matrix $\boldsymbol{A}_{\mathcal{G}}$ denotes the linear projection from the vertices of the mesh onto the regular grid points in $\mathcal{G}$. The term $\boldsymbol{u}_{\mathcal{G}}^{[k]}$ then serves as the k-th posterior MCMC sample for the spatial effects on the regular grid $\mathcal{G}$, which is calculated in post calculations after the MCMC run. Therefore, after calculating $\boldsymbol{u}_{\mathcal{G}}^{[k]}$ for every iteration k, posterior statistics for the spatial effects can be obtained for every point in $\mathcal{G}$, in particular posterior means and standard deviations.

The covariates are available by construction at every regular grid point. Thus, the k-th posterior MCMC sample for the location parameter, denoted with $\boldsymbol{\mu}_{\mathcal{G}}^{[k]}$, can be calculated at every regular grid point with

$$\boldsymbol{\mu}_{\mathcal{G}}^{[k]} = \boldsymbol{X}_{\mu\mathcal{G}}\boldsymbol{\beta}_{\mu}^{[k]} + \boldsymbol{A}_{\mathcal{G}}\boldsymbol{u}_{\mu}^{[k]} + \boldsymbol{v}_{\mu}^{[k]} \tag{III.6}$$

where $\boldsymbol{X}_{\mu\mathcal{G}}$ is a design matrix consisting of ones and the meteorological covariate at the regular grid points; the vector $\boldsymbol{\beta}_{\mu}^{[k]}$ is the k-th posterior MCMC sample of the coefficients of the covariates; and $\boldsymbol{v}_{\mu}^{[k]}$ is a vector sampled from a mean zero normal distribution with independent elements each with standard deviation $\sigma_{v\mu}^{[k]}$. Analogous results hold for the scale and shape parameters. Consequently, the posterior distribution of the p-th quantile of the generalized extreme value distribution can be calculated at every regular grid point. The p-th quantile

function of the generalized extreme value distribution is

$$q_p(\mu, \tau, \xi) = \mu + \frac{\exp(\tau)}{\xi}\left(-\log(p)^{-\xi} - 1\right). \tag{III.7}$$

The k-th posterior MCMC sample for the p-th quantile is then obtained by plugging in the k-th posterior MCMC samples for the location, scale and shape parameters. Thus, posterior samples and posterior statistics for the p-th quantile are obtained for every point in $\mathcal{G}$.

III.4 Results

The main objective of this analysis is to obtain spatial predictions for the location, the log-scale and shape parameters and to evaluate the posterior distribution of the 0.95 quantile of maximum precipitation across Iceland on the regular grid $\mathcal{G}$. In Section 4.1, results from the MCMC convergence diagnostics are briefly summarized while in Section 4.2 tables and figures of posterior estimates of the model parameters are given and discussed. In Section 4.3 figures of spatial predictions are shown and discussed.

III.4.1 Convergence diagnostics

The following results are based on four MCMC chains from the sampler in Section III.3.3. Each chain is calculated with 50000 iterations where 10000 iterations were burned in. Runtime, on a modern desktop (Ivy Bridge Intel Core i7-3770K, 16GB RAM and a solid state hard drive), was approximately seven hours. All the calculations were done using `R`.

Gelman–Rubin statistics (Brooks and Gelman 1998) were calculated for all model parameters based on the four MCMC chains for both datasets. The Gelman–Rubin statistics for all the model parameters were evaluated as approximately 1, indicating that all the MCMC chains converged in the mean. Moreover, Gelman–Rubin plots (results not shown) indicate that the sampler converges in the mean after 7500 iterations.

The MCMC chains for $\boldsymbol{\mu}$, $\boldsymbol{\tau}$ and $\boldsymbol{\xi}$ and the coefficients of the covariates exhibit a neglectable autocorrelation after lag 10 (results not shown). Samples of the hyperparameters show some autocorrelation as they are highly correlated in the posterior, but within an acceptable range.

III.4.2 Posterior estimates

The following posterior estimates are based on the four MCMC chains after burn-in. Figure III.6 shows the posterior mean with 95% posterior intervals for the location, log-scale and shape parameters for each observational site. The observational sites are placed on the x-axis as follows. The leftmost site on the x-axis is in Reykjavík. The rest of sites are places on the x-axis corresponding to a clockwise labeling across the sites shown in Figure III.1. The top left and right

panels in Figure III.6 show that both the location and the scale parameters of the generalized extreme values distribution have the highest posterior estimates in the south-eastern part in Iceland, and the lowest posterior estimates in the northern part. Furthermore, the bottom left panel in Figure III.6 indicates that the posterior mean of the shape parameter is positive at all the observational sites. However, some of the 95% posterior interval for the shape parameter include zero.

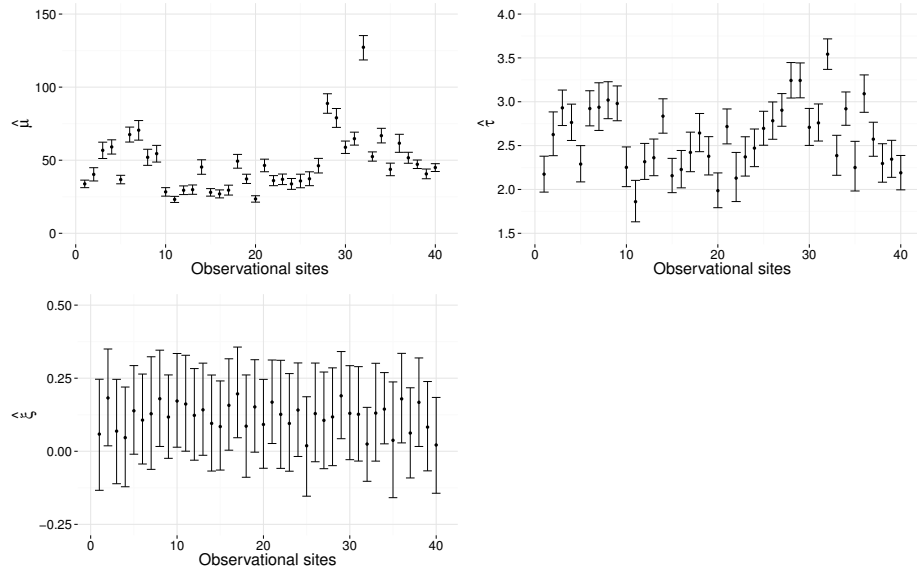

Figure III.6: The first row shows the posterior mean and 0.95 posterior interval for the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$, respectively, for each observational site. The second rows shows the same posterior estimates for $\boldsymbol{\xi}$.

Table III.1 shows the posterior mean, posterior standard deviation and posterior 0.025 and 0.975 quantile estimates for the non-spatially varying model parameters.

The results shown in Table III.1 indicate that the posterior means of $\beta_{\mu 1}$ and $\beta_{\mu 2}$ are 10.6 and 11.9, respectively, and that both parameters are positive as their corresponding 95% posterior intervals are both above zero. These results indicate that the location parameter describing the extreme precipitation is roughly 12 times the corresponding simulated mean value.

Furthermore, the posterior means of $\beta_{\tau 1}$ and $\beta_{\tau 2}$ are close to 2.05 and 0.5, respectively, and both parameters are positive as their 95% posterior intervals are above zero. That means that the relationship between the simulated means and the scale parameter of the generalized extreme value distribution can be roughly summarized as

$$\hat{\sigma}_i = e^{\hat{\beta}_{\tau 1}} \bar{x}_i^{*\hat{\beta}_{\tau 2}} \approx 7.8 \cdot \bar{x}_i^{*0.5}$$

The posterior mean of β_ξ is 0.12 and its posterior interval is $(0.07, 0.17)$. That indicates that the field describing the shape parameter has mean greater

	2.5%	mean	97.5%	sd
$\beta_{\mu 1}$	2.270	10.615	18.791	4.205
$\beta_{\mu 2}$	9.585	11.922	14.296	1.192
$\beta_{\tau 1}$	1.777	2.054	2.332	0.140
$\beta_{\tau 2}$	0.272	0.513	0.754	0.123
β_{ξ}	0.069	0.117	0.165	0.025
$\omega_{u\mu}$	0.244	0.393	0.596	0.091
$\kappa_{u\mu}$	0.045	0.074	0.115	0.018
$\sigma_{v\mu}$	2.532	4.997	8.658	1.566
$\omega_{u\tau}$	5.473	6.061	6.690	0.312
$\kappa_{u\tau}$	0.095	0.123	0.167	0.021
$\sigma_{v\tau}$	0.075	0.149	0.264	0.048
$\omega_{u\xi}$	6.160	7.594	9.192	0.770
$\kappa_{u\xi}$	0.224	0.424	0.861	0.166
$\sigma_{v\xi}$	0.072	0.101	0.138	0.017

Table III.1: Posterior 0.025 quantile, mean, 0.975 quantile and standard deviation of the non-spatially varying model parameters

than zero.

The posterior mean of the parameters $\kappa_{u\mu}$, $\kappa_{u\tau}$ and $\kappa_{u\xi}$, suggest that the correlation between two points in space is near 0.1 at a roughly 40 km distance for the location parameter; roughly 25 km distance for the scale parameter; and almost 7 km for the shape parameter, which is short compared to the range of μ and τ and the scale of the spatial domain. The posterior mean of the parameters $\omega_{u\mu}$, $\omega_{u\tau}$ and $\omega_{u\xi}$ indicate that the marginal standard deviation of $\boldsymbol{u}_{\mu}$ is approximately 8; roughly 0.45 for $\boldsymbol{u}_{\tau}$; and almost 0.1 for $\boldsymbol{u}_{\xi}$.

The posterior mean of $\sigma_{v\mu}$ indicate that the standard deviation of the unstructured random effect $\boldsymbol{v}_{\mu}$ is approximately 5. Which in turn suggests that there is some variation left in the data that is unexplained by the covariates and the spatial model, however, a standard deviation $\sigma_{v\mu}$ of size 5 is small relative the scale of μ and the standard deviation of the spatial field for μ. Similar behavior is observed for the scale parameter, as the posterior mean of $\sigma_{v\tau}$ indicate that the standard deviation of the unstructured random effect $\boldsymbol{v}_{\tau}$ is approximately 0.15 on a logarithmic scale. However, the posterior mean of $\sigma_{v\xi}$ is 0.1, which is a similar value is the posterior mean of the marginal standard deviation of $\boldsymbol{u}_{\xi}$. This is to be expected, as the the range of the spatial model for the shape parameter ξ is short compared to the scale of the spatial domain. These results for the shape parameter indicate that some of the variation left in the data is unexplained by the overall effect β_{ξ} and the spatial field. The cause of this is likely to be a lack of data, that is, observations from more observational sites that are closer together are needed in order to capture the spatial behavior of ξ. Furthermore, useful covariates explaining the spatial variation in ξ could not be found.

In Figure III.7 the prior distributions of all the hyperparameters are shown, along with the corresponding posterior distributions. The panels in the first and

second columns of Figure III.7 are shown on transformed scales for interpretability.

III.4.3 Model evaluation

In Figure III.8 empirical cumulative distributions for the observational sites Reykjavík, Æðey, Akureyri and Kvísker are compared to corresponding posterior cumulative distributions functions. The results indicate that the model describes the data well.

An overall time effect for location parameter was evaluated based on the fitted values at each observational site. That is, let $m_{it} = y_{it} - \mathrm{E}(y_{it})$, where $\mathrm{E}(y_{it})$ is the excepted value of the generalized extreme value distribution with parameter values based on posterior means. A likelihood ratio test was applied to a model for the m's with an overall time effect against a model for the m's without a time effect. The results indicated no significant difference between the two models (significance level 0.05). Thus, there was no reason to add an overall time effect to the proposed model.

III.4.4 Spatial predictions

Figures III.9, III.10 and III.11 show the spatial predictions for the spatially varying model parameters, based on the methods from Section III.3.4.

The middle panel of Figure III.9 shows the spatial predictions for the spatial random effect $\boldsymbol{u}_\mu$ based on the posterior mean of (III.5) on the regular grid $\mathcal{G}$. The figure shows where the spatial random effect lowers and raises the prediction surface, therefore revealing the ares where the effect of the meteorological covariate overestimates and underestimates the extreme precipitation, respectively. The spatial prediction surface is lowered in most of the south-western part of Iceland, but is raised in the south-eastern part. The spatial random effect yields high positive values for Kvísker, which is known to have the highest observed precipitation in Iceland; negative value in the south-eastern part close to Reykjavík and values close to zero in the interior of Iceland where there are no observational sites.

The bottom panel of Figure III.9 shows the spatial prediction for the standard deviation of the spatial random effect $\boldsymbol{u}_\mu$ on the regular grid $\mathcal{G}$. As expected, the standard deviation increases at points further away from the observational sites, forming sinks in the standard deviation near the observational sites. In the areas where there are no data the estimates for the standard deviation for points inside the triangles are lower than for the corresponding edges or vertices. This happens due to the linear basis functions in the SPDE approach and because the standard deviation is estimated based on (III.5).

The top panel of Figure III.9 shows the spatial prediction of the location parameter $\boldsymbol{\mu}_\mathcal{G}$, based on (III.6), on the regular grid $\mathcal{G}$. The figure shows that the location parameter is at its highest in the south-eastern parts of Iceland, in the vicinity of the southern side of Vatnajökull Glacier. This is to be expected, as the spatial gradient at the southern side of Vatnajökull increases rapidly moving

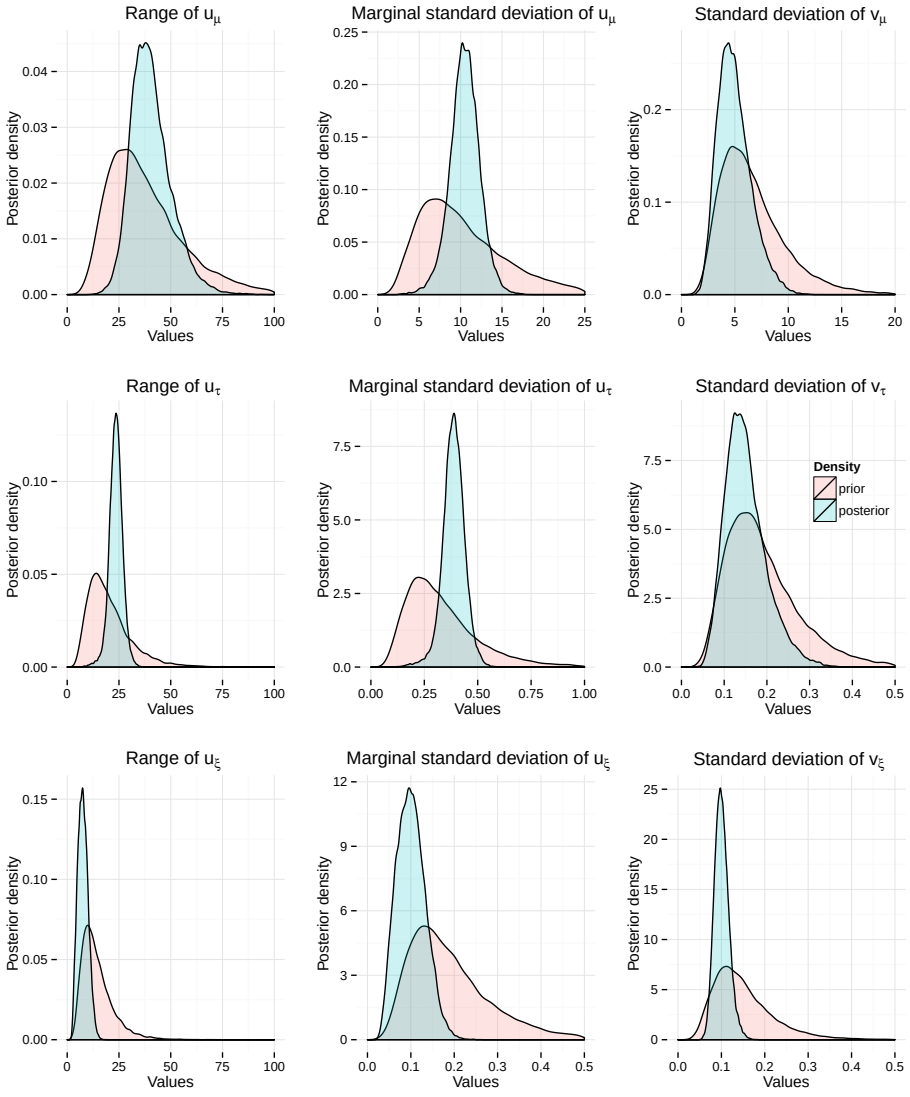

Figure III.7: Prior distribution for all the hyperparameters and corresponding posterior distributions.

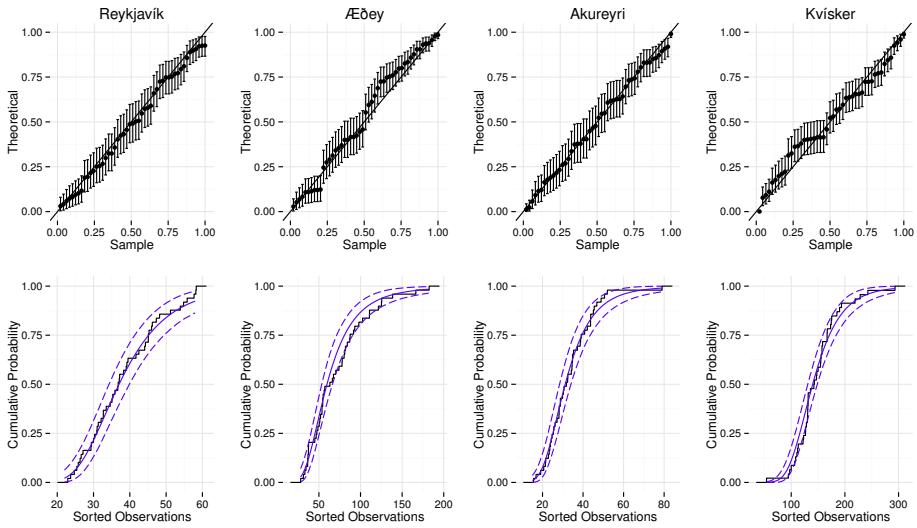

Figure III.8: The top row shows probability-probability plots based on the proposed model for Reykjavík, Æðey, Akureyri and Kvísker. The bottom row shows the empirical cummulative density function (the black curve) and the model cumulative density function (blue curve) along with 95% posterior intervals (dashed blue curves) for the same observational sites.

from the nearby coastline to the top of the glacier. Due to these topographical properties and the fact that humid air blows in from the southern shoreline towards the roots of the glacier, the physical law of orographic precipitation predicts high precipitation. However, the areas north of Vatnajökull Glacier and in the middle of the country are known to be in a rain shadow by the same meteorological law. This is in line with the results seen in the bottom left panel of Figure III.9.

Figure III.10 shows the spatial prediction for the scale parameter on a logarithmic scale. The figure is arranged in the same manner as Figure III.9. In the middle panel of Figure III.10 it can been seen that the spatial random effect $\boldsymbol{u}_\tau$ raises the spatial prediction surface in the eastern, south-eastern and north-western parts and lowers it in the south-western part. In the bottom panel of Figure III.10, similar results appear for the standard deviation as for the location parameter, that is, sinks in the standard deviation form close to the observational sites. Moreover, the figure demonstrates that the spatial range for τ is shorter than for μ, as discussed above. On the top panel of Figure III.10, it can be seen that the estimates for the scale parameter are highest along the south-eastern coastline.

Figure III.11 shows the spatial prediction for the shape parameter, arranged in the ame manner as Figures III.9 and III.10. The middle panel of Figure III.11 shows the posterior mean of $\boldsymbol{u}_\xi$ on the regular grid $\mathcal{G}$. The magnitude of $\boldsymbol{u}_\xi$ is only 0.015 which is smaller that the variation in the ML estimates (results not

shown). Furthermore, the bottom panel in Figure III.11 shows that the sinks in the standard deviation of the spatial field for ξ are almost not visible. This is to be expected as the range of the spatial field for ξ is only 7 km, as discussed above. The top panel of Figure III.11 demonstrates that the predicted posterior mean values of ξ are close to the posterior mean of the overall shape effect β_ξ, which is 0.117. These results indicate that a dataset based on a denser network of observational sites is required to give better spatial predictions for ξ.

Spatial predictions for the posterior mean of the 0.95 quantile of the generalized extreme value distribution can be seen in Figure III.12. The predictions are based on the methods discussed in Section III.7 and can be interpreted as the 20-year precipitation event. The results reflect the previous posterior results about the location and scale parameters. For example, the highest predicted 20-year precipitation events are along the south and south-eastern coastlines, while the lowest predicted events are in the interior of Iceland. The maximum of the predicted values is at the southern side of Hvannadalshnjúkur and is approximately 450 mm per 24 hours. The peak of Hvannadalshnjúkur, which is a part Vatnajökull, is the highest point of Iceland and is approximately 10 km north west of the observational site Kvísker. These results demonstrate the effects of orographic precipitation as Kvísker is around 30 m above sea level and the highest point of Hvannadalshnjúkur is roughly 2110 m above sea level. Furthermore, the predicted 20-year precipitation event in Reykjavík and surroundings is predicted approximately 60 mm per 24 hours and in the vicinity of Akureyri it is around 55 mm per 24 hours.

III.5 Discussion

The proposed method which leverages spatial information on extreme precipitation by calculating sample mean values from outputs of a local meteorological model is extendable to any other location in the world where outputs from local meteorological models are available on a fine grid. Furthermore, the method can also be extended to other response variables. For example, the method is likely to be well suited to resolve spatial information on annual or monthly precipitation. Additionally, the method can be utilized to extract any desired sample statistics from meteorological models in order to leverage spatial information. In the authors' view, leveraging information from meteorological models for different regions, for other response variables and basing the information on various extracted sample statistics presents an interesting area of future research.

The proposed LGM for extreme events with the SPDE spatial models structure is extendable to any regions in world or spatial domains of interest, and for other response variables than precipitation. Furthermore, as the proposed statistical modeling framework is modular by design, other likelihoods can be chosen for the observations without the need to change the structure of the SPDE spatial model at the latent level.

A different likelihood choice for extreme precipitation is, for example, the peaks over threshold methods with the generalized Pareto distribution as pre-

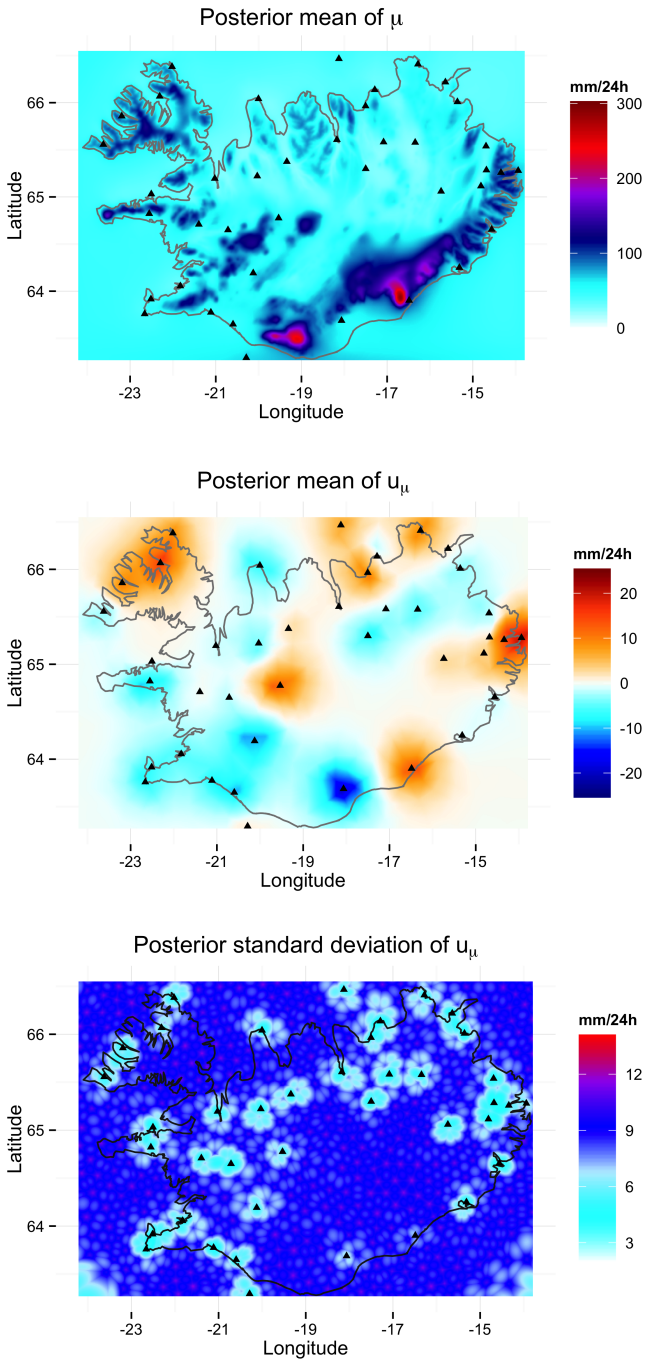

Figure III.9: The figure shows spatial predictions for the location parameter on the regular grid $\mathcal{G}$. The middle panel shows the posterior mean of the spatial random effect $\boldsymbol{u}_\mu$; the bottom panel shows the posterior standard deviation of the spatial random effect $\boldsymbol{u}_\mu$ and the top panel shows posterior mean of the location parameter $\boldsymbol{\mu}_\mathcal{G}$

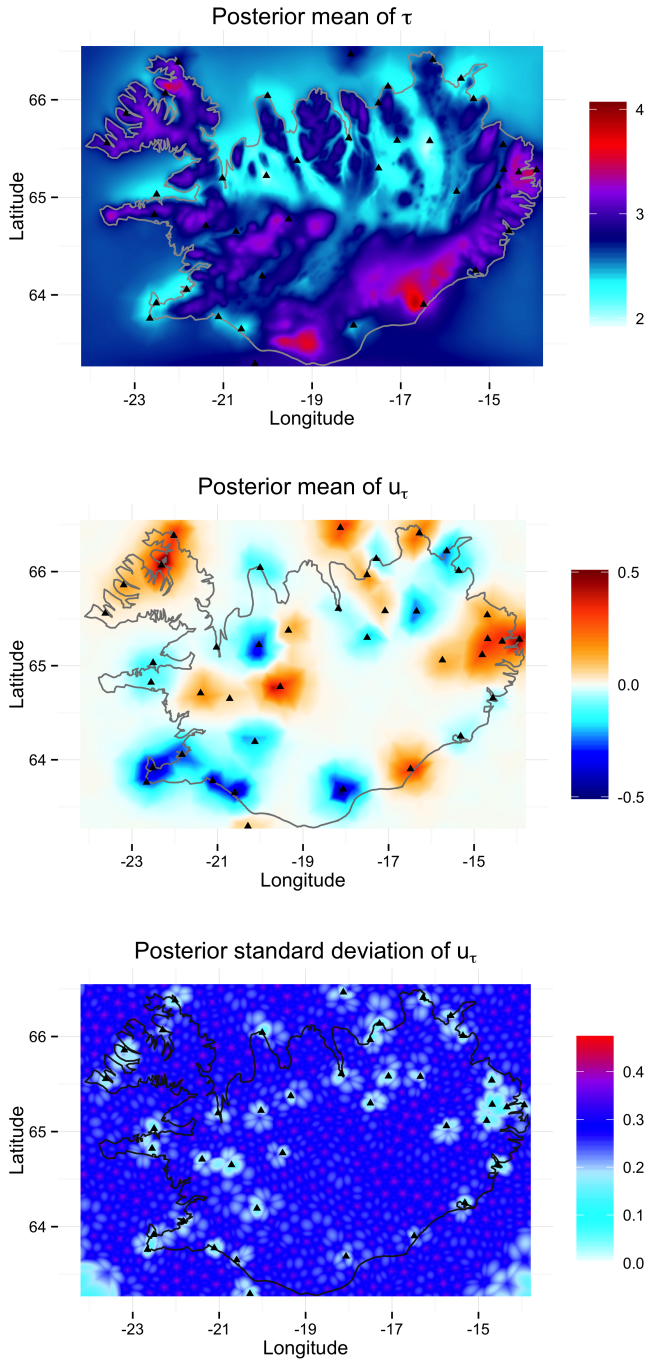

Figure III.10: The figure shows spatial predictions for the scale parameter on a logarithmic scale on the regular grid $\mathcal{G}$. The middle panel shows the posterior mean of the spatial random effect $\boldsymbol{u}_\tau$; the bottom panel shows the posterior standard deviation of the spatial random effect $\boldsymbol{u}_\tau$ and the top panel shows posterior mean of the scale parameter $\boldsymbol{\tau}_\mathcal{G}$ on a logarithmic scale.

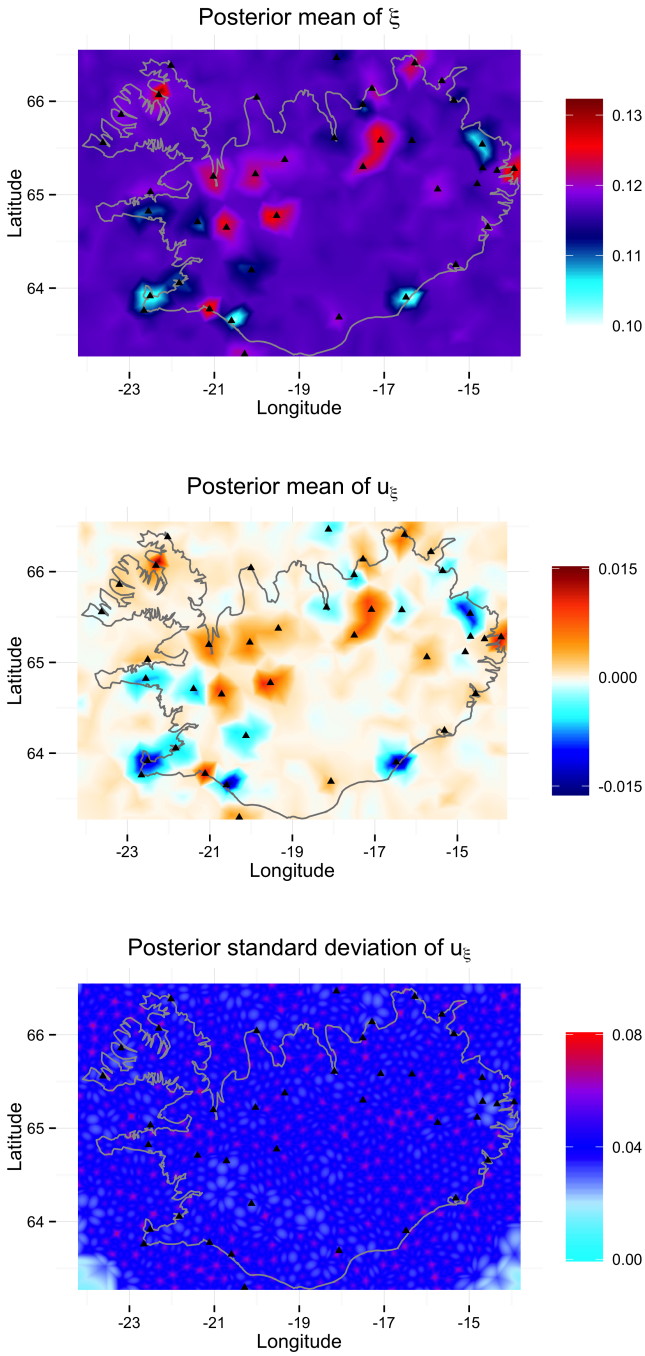

Figure III.11: The figure shows spatial predictions for the shape parameter on the regular grid $\mathcal{G}$. The middle panel shows the posterior mean of the spatial random effect $\boldsymbol{u}_\xi$; the bottom panel shows the posterior standard deviation of the spatial random effect $\boldsymbol{u}_\xi$ and the top panel shows posterior mean of the shape parameter $\boldsymbol{\xi}_\mathcal{G}$

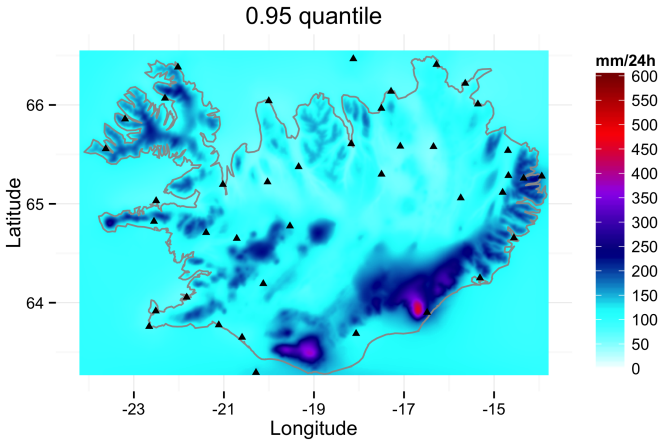

Figure III.12: The posterior mean of the 0.95 quantile of annual maximum 24 hour precipitation assuming the generalized extreme value distribution.

sented in (Cooley et al. 2007). The advantage of thatapproach is that more of the data can be incorporated into the modeling, depending on the choice of threshold. However, this approach was not chosen for the modeling in this paper, as choosing a common threshold for each station is somewhat unrealistic for the datasets explored in this paper. For example, in Figure III.2, it can be seen that highest observed value in Reykjavík is lower than the lowest observed value in Kvísker. Further, selecting a spatially varying threshold requires more research.

Other likelihood approaches are copula likelihoods with marginals based on the generalized extreme value distribution. Likelihoods constructed with the t-copula have been explored for example by Davison et al. (2012) to model extreme precipitation and present an appealing choice to model the probabilistic dependence of the observations. The approach is well suited for simulating spatial realizations of the extreme precipitation for the next year or unobserved sites for an observed year. Although the copula likelihood can be implemented within the presented modeling framework, there were two main reasons why it was not chosen in this paper. First, the main purpose is to give spatial predictions of marginal quantiles using the SPDE approach but not to simulate spatial realizations of the extremal surfaces. Secondly, many of the observational sites had missing observations, which presents further computational difficulties with the copula based likelihood.

Stationary SPDE spatial models were implemented at the latent level of the proposed model. However, the results indicate a non-stationary behavior in some of the mountainous regions, for example near Kvísker, as discussed in Section III.4.2. Non-stationary SPDE spatial models have been proposed, see for example Fuglstad et al. (2013) and Ingebrigtsen et al. (2014). Implementing these models here was beyond the scope of this paper. However, these models yield an appealing modeling option for non-stationary spatial fields. Future research

could involve using spatial covariates in the dependence structure of the spatial model in a similar fashion as in Ingebrigtsen et al. (2014). For example, new spatial covariates based on outputs from a local meteorological model, as suggested here, could be designed to provide useful information for the dependence structure. This is, in the authors' view, an interesting example how the proposed method can be extended to other areas of future research.

Acknowledgements

The authors would like to thank the University of Iceland Doctoral Fund, University of Iceland Research Fund and Landsvirkjun Energy Research Fund which supported the research. The authors would also like to thank the Icelandic Meteorological Office, in particular, Dr. Philippe Crochet of IMO for providing the data and valuable discussions. The authors give their thanks to the Nordic Network on Statistical Approaches to Regional Climate Models for Adaptation (SARMA), especially Prof. Peter Guttorp, for providing travel support. Furthermore, the authors give their thanks to the Department of Mathematical Sciences at the Norwegian University of Science and Technology for hosting Óli Páll Geirsson several times, and special gratitude to Prof. Håvard Rue for his invitation and valuable conversations. The authors would also like to thank the two reviewers of this paper and the associate editor for a thorough review and constructive comments and criticism.

III.A Appendix

III.A.1 Spatial smoother

Let $\mathcal{G}$ denote the set of every regular grid point and let $\mathcal{S}$ denote the set of the observational sites. Furthermore, let $\bar{x}_j^*$ denote the simulated means at every grid point $j \in \mathcal{G}$ and $\bar{x}_i$ denote observed means at each observational site $i \in \mathcal{S}$. For a given tuning distance $r > 0$, find every grid point $j \in \mathcal{G}$ that is within distance r of point $i \in \mathcal{S}$ and denote that index set with $D(r, i)$.

The goal is to construct a covariate at site i that uses information from the grid points in $D(r, i)$. Assign the following decay function

$$w_{ij} = 1 - F_\alpha\left(\frac{d_{ij}}{r}\right) \qquad \text{for all } j \in D(r, i)$$

where F_α denotes the cumulative density function of the beta distribution with parameters α and α, and d_{ij} denotes the Euclidean distance from point i to point j. The two parameters of the beta distribution are chosen to be the same in order to avoid interactions between the tuning distance r and the shape of the decay function. The w_{ij}s describe the relative weights of the neighboring grid points in $D(r, i)$. The covariate at observation point i can then be constructed as the

weighted mean of the surrounding simulated means, that is

$$\bar{x}_i^*(r,\alpha) = \frac{\sum_{j \in D(r,i)} w_{ij} \bar{x}_j^*}{\sum_{j \in D(r,i)} w_{ij}} \quad \text{for every } i \text{ in } \mathcal{S}. \tag{III.8}$$

Following the (Benestad et al. 2012) argument, the parameters in the spatial smoother are tuned such that the simulated means at the observational sites are close to the observed means. We chose to measure the distance between the simulated means, which are calculated with equation (III.8), and the observed means by using mean square distance. Then the optimal r and α can be found in the mean square sense by calculating the following

$$(r_0, \alpha_0) = \arg\min_{r,\alpha} \sum_{i \in \mathcal{S}} ||\bar{x}_i^*(r,\alpha) - \bar{x}_i||^2.$$

The covariates that will be used in this paper are then $\bar{x}_i^* = \bar{x}_i^*(r_0, \alpha_0)$ for all observational sites $i \in \mathcal{S}$.

III.A.2 Prior selection: Details

The fixed parameters in the priors for $\kappa_{u\mu}$, $\kappa_{u\tau}$ and $\kappa_{u\xi}$ were chosen in a weakly informative manner to ensure that the range of the spatial effects is sensible relative to the size of domain of interest. The relationshiphip between the parameters $\kappa_{u\mu}$, $\kappa_{u\tau}$ and $\kappa_{u\xi}$ and the spatial range of the approximated Matérn field used in this paper is $\rho = \sqrt{8\nu}/\kappa$, which corresponds to correlation near 0.1 at a distance ρ. Consequently, the prior distributions for $\kappa_{u\mu}$, $\kappa_{u\tau}$ and $\kappa_{u\xi}$ were chosen such that the corresponding ρ's range up to 150 km, which is approximately half of the length of the spatial domain in this paper.

Exploratory data analysis revealed that the marginal standard deviations for $\boldsymbol{u}_\mu$, $\boldsymbol{u}_\tau$ and $\boldsymbol{u}_\xi$ should exceed 50, 5 and 5, respectively, with low probability on their native scales. The relationship between the parameters $(\kappa_{u\mu}, \omega_{u\mu})$, $(\kappa_{u\tau}, \omega_{u\tau})$ and $(\kappa_{u\xi}, \omega_{u\xi})$ and the marginal standard deviation of the Matérn fields is approximately $1/(\sqrt{4\pi}\kappa\omega)$.

The unstructured random effects $\boldsymbol{v}_\mu$, $\boldsymbol{v}_\tau$ and $\boldsymbol{v}_\xi$ capture the variation in the data which is unexplained by the covariates and the spatial random effects. The standard deviations of $\boldsymbol{v}_\mu$, $\boldsymbol{v}_\tau$ and $\boldsymbol{v}_\xi$ are believed to be mainly between 2 and 20; 0.05 and 1; and 0.05 and 0.75 respectively. This is based on an exploratory data analysis and is in line with the priors for the marginal standard deviation for the Matérn fields.

III.A.3 Implementation of the MCMC split sampler

In order to implement the MCMC split sampling strategy, the model parameters are split into two main blocks. Namely, the *data-rich block* which includes the following $\boldsymbol{\eta} = (\boldsymbol{\mu}^\mathsf{T}, \boldsymbol{\tau}^\mathsf{T}, \boldsymbol{\xi}^\mathsf{T})^\mathsf{T}$ and the *data-poor block* which consists

of $\boldsymbol{\nu} = (\boldsymbol{\beta}_\mu^\mathsf{T}, \boldsymbol{u}_\mu^\mathsf{T}, \boldsymbol{\beta}_\tau^\mathsf{T}, \boldsymbol{u}_\tau^\mathsf{T}, \beta_\xi^\mathsf{T}, \boldsymbol{u}_\xi^\mathsf{T})^\mathsf{T}$ and the hyperparameters

$$\boldsymbol{\theta} = (\kappa_{u\mu}, \omega_{u\mu}, \sigma_{v\mu}, \kappa_{u\tau}, \omega_{u\tau}, \sigma_{v\tau}, \kappa_{u\xi}, \omega_{u\xi}, \sigma_{v\xi})^\mathsf{T}.$$

Then define the following sparse matrices

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X}_\mu & \boldsymbol{A}_\mathcal{S} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{X}_\tau & \boldsymbol{A}_\mathcal{S} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \boldsymbol{X}_\xi & \boldsymbol{A}_\mathcal{S} \end{pmatrix}, \quad \boldsymbol{Q}_\epsilon = \begin{pmatrix} \sigma_{v\mu}^{-2}\boldsymbol{I} & \cdot & \cdot \\ \cdot & \sigma_{v\tau}^{-2}\boldsymbol{I} & \cdot \\ \cdot & \cdot & \sigma_{v\xi}^{-2}\boldsymbol{I} \end{pmatrix}$$

and

$$\boldsymbol{Q}_\nu = \begin{pmatrix} \kappa_{\beta\mu}\boldsymbol{I} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \boldsymbol{Q}_{u\mu}(\boldsymbol{\psi}_{u\mu}) & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \kappa_{\beta\tau}\boldsymbol{I} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \boldsymbol{Q}_{u\tau}(\boldsymbol{\psi}_{u\tau}) & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \kappa_{\beta\xi}\boldsymbol{I} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \boldsymbol{Q}_{u\tau}(\boldsymbol{\psi}_{u\xi}) \end{pmatrix}.$$

The posterior density function is

$$\pi(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\eta})\pi(\boldsymbol{\eta}, \boldsymbol{\nu}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

therefore, the $(k+1)$-th MCMC sample from the posterior density $\pi(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta}|\boldsymbol{y})$ can be obtained by

Data-rich step: sampling $\boldsymbol{\eta}^{k+1}$ from $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\nu}^k, \boldsymbol{\theta}^k)$ and

Data-poor step: sampling $(\boldsymbol{\nu}^{k+1}, \boldsymbol{\theta}^{k+1})$ jointly from $\pi(\boldsymbol{\nu}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta}^{k+1})$.

The details of sampling strategies within each block are as follows.

Data-rich block The structure of the data-rich part is as follows. Let and $\boldsymbol{\eta}_i = (\mu_i, \tau_i, \xi_i)^\mathsf{T}$ for $i = 1, \ldots, J$. The logarithm of the conditional posterior becomes

$$\log \pi(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{Q}_\epsilon\boldsymbol{\eta} + (\boldsymbol{Q}_\epsilon\boldsymbol{Z}\boldsymbol{\nu})^\mathsf{T}\boldsymbol{\eta} + f(\boldsymbol{\eta}) + \text{constant}$$

where

$$f(\boldsymbol{\eta}) = \sum_{i=1}^{J} f_i(\boldsymbol{\eta}_i) = \sum_{i=1}^{J} \sum_{t \in \mathcal{A}_i} \log \pi_{\text{gev}}(y_{it}|\mu_i, \exp\tau_i, \xi_i),$$

where π_{gev} denotes the density of the generalized extreme value distribution and the set $\mathcal{A}_i$ contains the indices of the years t observed at site i.

A proposal density q based on a Gaussian approximation was implemented at the mode of the logarithm of conditional posterior density function, i.e. $\log \pi(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\theta})$. Then q is given by

$$q(\boldsymbol{\eta}^*) = \mathcal{N}\left(\boldsymbol{\eta}^* \mid \boldsymbol{\eta}_0, (\boldsymbol{Q}_\epsilon - \boldsymbol{H})^{-1}\right), \tag{III.9}$$

where $\boldsymbol{\eta}_0$ denotes the mode of the conditional posterior density and $\boldsymbol{H} = \nabla^2 f(\boldsymbol{\eta}_0)$. The proposal density q in (III.9) is therefore an independence proposal density, see (Chib and Greenberg 1995) and (Rue and Held 2005).

In order to the calculate the mode $\boldsymbol{\eta}_0$, numerical optimizing methods are necessary. In this paper, the Newton–Raphson algorithm was implemented with fixed starting values in each MCMC iteration. The calculated mode, denoted by $\hat{\boldsymbol{\eta}}_0$, is not neccesarly the true mode but a value close to it. However, as fixed starting values are used in the Newton–Raphson algorithm the calculated mode $\hat{\boldsymbol{\eta}}_0$ is found independently of $\boldsymbol{\eta}^k$ in every k-th iteration. Thus, the proposal density q in (III.9) is ensured to be an independence proposal density. The steps of the sampler are outlined in Algorithm 3 in **Paper II**.

Data-poor part The proposal strategy suggested in is used for each element of $\boldsymbol{\theta}$. That is, let $\theta_i^* = f\theta_i^k$ where the scaling factor f has the density

$$\pi(f) \propto 1 + 1/f \quad \text{for } f \in [1/F, F] \tag{III.10}$$

where $F > 1$ is a tuning parameter. It can be shown that this is a symmetric proposal density in the sense that

$$q(\theta_i^* | \theta_i^k) = q(\theta_i^k | \theta_i^*)$$

A joint proposal strategy for $(\boldsymbol{\nu}, \boldsymbol{\theta})$ is implemented as seen in **Paper II**. The steps of the sampler are outlined in Algorithm 4 in **Paper II**.

Bibliography

Abramowitz, M. and Stegun, I. A. 1964. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Number 55. Courier Corporation.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. 2014. *Hierarchical modeling and analysis for spatial data*. Crc Press.

Basist, A., Bell, G. D., and Meentemeyer, V. 1994. Statistical relationships between topography and precipitation patterns. *Journal of climate*, 7(9):1305–1315.

Beichl, I. and Sullivan, F. 2000. The metropolis algorithm. *Computing in Science & Engineering*, 2(1):65–69.

Benestad, R., Nychka, D., and Mearns, L. 2012. Spatially and temporally consistent prediction of heavy precipitation from mean values. *Nature Climate Change*, 2(7):544–547.

Berger, J. O. 2013. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Berliner, L. M. 1996. Hierarchical bayesian time series models. In *Maximum entropy and Bayesian methods*, pages 15–22. Springer.

Berliner, L. M. 2003. Physical-statistical modeling in geophysics. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 108(D24).

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.

Boothby, W. M. 2003. *An introduction to differentiable manifolds and Riemannian geometry*, volume 120. Gulf Professional Publishing.

Brenner, S. C. and Scott, R. 2008. *The mathematical theory of finite element methods*, volume 15. Springer Science & Business Media.

Brooks, S. P. and Gelman, A. 1998. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455.

Casella, G. and George, E. I. 1992. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.

Chib, S. and Greenberg, E. 1995. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335.

Chiles, J.-P. and Delfiner, P. 2009. *Geostatistics: modeling spatial uncertainty*, volume 497. John Wiley & Sons.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. 2001. *An introduction to statistical modeling of extreme values*, volume 208. Springer.

Coles, S. G. and Tawn, J. A. 1996. A bayesian analysis of extreme rainfall data. *Applied statistics*, pages 463–478.

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

Cooley, D. and Sain, S. R. 2010. Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of agricultural, biological, and environmental statistics*, 15(3):381–402.

Cressie, N. 1993. *Statistics for Spatial Data*. Wiley, New York, NY, USA.

Cressie, N. and Huang, H.-C. 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1339.

Cressie, N. and Wikle, C. K. 2011. *Statistics for spatio-temporal data*. John Wiley & Sons.

Creutin, J. and Obled, C. 1982. Objective analyses and mapping techniques for rainfall fields: an objective comparison. *Water resources research*, 18(2):413–431.

Crochet, P. et al. 2012. Estimating the flood frequency distribution for ungauged catchments using an index flood procedure. application to ten catchments in northern iceland.

Crochet, P., Jóhannesson, T., Jónsson, T., Sigurðsson, O., Björnsson, H., Pálsson, F., and Barstad, I. 2007. Estimating the spatial distribution of precipitation in iceland using a linear model of orographic precipitation. *Journal of Hydrometeorology*, 8(6):1285–1306.

Cunnane, C. and Nash, J. 1971. Bayesian estimation of frequency of hydrological events. *Proceedings of the Warsaw symposium. Mathematical models in hydrology*, 1(100):47–55.

Daly, C., Neilson, R. P., and Phillips, D. L. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of applied meteorology*, 33(2):140–158.

Davison, A. C., Padoan, S., and Ribatet, M. 2012. Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

Davíðsson, Ó. B. 2015. Bayesian flood frequency analysis using monthly maxima.

Diggle, P. J., Tawn, J., and Moyeed, R. 1998. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

Dyrrdal, A. V. 2012. Estimation of extreme precipitation in norway and a summary of the state-of-the-art.

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

Elíasson, J. 1994. Statistical estimates of pmp values. *Nordic hydrology*, 25(4):301–312.

Eliasson, J., Rögnvaldsson, Ó., and Jonsson, T. 2009. Extracting statistical parameters of extreme precipitation from a nwp model. *Numerical simulations of surface winds and precipitation in Iceland*.

Fahrmeir, L., Tutz, G., Hennevogl, W., and Salem, E. 1994. *Multivariate statistical modelling based on generalized linear models*, volume 2. Springer New York.

Ferkingstad, E., Frigessi, A., Rue, H., Thorleifsson, G., and Kong, A. 2008. Unsupervised empirical bayesian multiple testing with external covariates. *The Annals of Applied Statistics*, pages 714–735.

Filippone, M., Zhong, M., and Girolami, M. 2013. A comparative evaluation of stochastic-based inference methods for gaussian process models. *Machine Learning*.

Førland, E. J. and Hanssen-Bauer, I. 2000. Increased precipitation in the norwegian arctic: true or false? *Climatic Change*, 46(4):485–509.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. 2013. Non-stationary spatial modelling with applications to spatial prediction of precipitation. *arXiv preprint arXiv:1306.0408*.

Geirsson, Ó. P., Hrafnkelsson, B., and Simpson, D. 2015. Computationally efficient spatial modeling of annual maximum 24-h precipitation on a fine grid. *Environmetrics*.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2014. *Bayesian data analysis*, volume 2. Taylor & Francis.

Gelman, A. and Hill, J. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A. and Rubin, D. B. 1992. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

Geman, S. and Geman, D. 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741.

Gilks, W. R. 2005. *Markov chain monte carlo*. Wiley Online Library.

Girolami, M. and Calderhead, B. 2011. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214.

GREHY, G. D. R. E. S. 1996. Presentation and review of some methods for regional flood frequency analysis. *Journal of hydrology(Amsterdam)*, 186(1-4):63–84.

Grimmett, G. and Stirzaker, D. 2001. *Probability and random processes*. Oxford university press.

Grondona, M. O. and Cressie, N. 1991. Using spatial considerations in the analysis of experiments. *Technometrics*, 33(4):381–392.

Guttorp, P. and Gneiting, T. 2006. Studies in the history of probability and statistics xlix on the matern correlation family. *Biometrika*, 93(4):989–995.

Hastings, W. K. 1970. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Hrafnkelsson, B., Morris, J. S., and Baladandayuthapani, V. 2012. Spatial modeling of annual minimum and maximum temperatures in iceland. *Meteorology and Atmospheric Physics*, 116(1-2):43–61.

Ingebrigtsen, R., Lindgren, F., and Steinsland, I. 2014. Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8:20–38.

Johnson, R. A., Wichern, D. W., et al. 1992. *Applied multivariate statistical analysis*, volume 4. Prentice hall Englewood Cliffs, NJ.

Kneib, T. 2013. Beyond mean regression. *Statistical Modelling*, 13(4):275–303.

Knorr-Held, L. and Rue, H. 2002. On block updating in markov random field models for disease mapping. *Scandinavian Journal of Statistics*, 29(4):597–614.

Lawson, A. B. 2013. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC Press.

Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Martino, S., Aas, K., Lindqvist, O., Neef, L. R., and Rue, H. 2011. Estimating stochastic volatility models using integrated nested laplace approximations. *The European Journal of Finance*, 17(7):487–503.

Martins, T. G., Simpson, D., Illian, J. B., Rue, H., and Geirsson, Ó. P. 2013a. Discussion of 'beyond mean regression'. *Statistical Modelling*, 13(4):355–361.

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. 2013b. Bayesian computing with inla: new features. *Computational Statistics & Data Analysis*, 67:68–83.

McLachlan, G. and Krishnan, T. 2007. *The EM algorithm and extensions*, volume 382. John Wiley & Sons.

Mengersen, K. L., Tweedie, R. L., et al. 1996. Rates of convergence of the hastings and metropolis algorithms. *The Annals of Statistics*, 24(1):101–121.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. 1953. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21:1087.

Murray, I. and Adams, R. P. 2010. Slice sampling covariance hyperparameters of latent gaussian models. *arXiv preprint arXiv:1006.0868*.

Neal, R. M. 1993. Probabilistic inference using markov chain monte carlo methods.

Pettitt, A. N., Weir, I., and Hart, A. 2002. A conditional autoregressive gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing*, 12(4):353–367.

Rigby, R. A. and Stasinopoulos, D. M. 2005. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554.

Robert, C. and Casella, G. 2013. *Monte Carlo statistical methods*. Springer Science & Business Media.

Roberts, G. O., Gelman, A., Gilks, W. R., et al. 1997. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120.

Roberts, G. O. and Rosenthal, J. S. 1998. Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268.

Rue, H. 2001. Fast sampling of gaussian markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):325–338.

Rue, H. and Held, L. 2005. *Gaussian Markov random fields: theory and applications*. CRC Press.

Rue, H., Martino, S., and Chopin, N. 2009. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.

Sang, H. and Gelfand, A. E. 2009. Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16(3):407–426.

Schaefli, B., Talamba, D. B., and Musy, A. 2007. Quantifying hydrological modeling errors through a mixture of normal distributions. *Journal of Hydrology*, 332(3):303–315.

Schliep, E. M., Cooley, D., Sain, S. R., and Hoeting, J. A. 2010. A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13(2):219–239.

Sigurdsson, O., Thorsteinsson, T., Agustsson, S., and Einarsson, B. 2004. Mass balance of hofsjokull. *National Energy Authority*, 2004(029):54.

Simpson, D., Lindgren, F., and Rue, H. 2012. Think continuous: Markovian gaussian models in spatial statistics. *Spatial Statistics*, 1:16–29.

Smith, A. F. and Roberts, G. O. 1993. Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–23.

Smith, R. B. and Barstad, I. 2004. A linear theory of orographic precipitation. *Journal of the Atmospheric Sciences*, 61(12).

Stein, M. L. 1999. *Interpolation of spatial data: some theory for kriging*. Springer, New York.

Stein, M. L. 2012. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

Tebaldi, C. and Sansó, B. 2009. Joint projections of temperature and precipitation change from multiple climate models: a hierarchical bayesian approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1):83–106.

Tong, Y. L. 2012. *The multivariate normal distribution*. Springer Science & Business Media.

Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., et al. 2005. The era-40 reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012.

Wotling, G., Bouvier, C., Danloux, J., and Fritsch, J.-M. 2000. Regionalization of extreme precipitation distribution using the principal components of the topographical environment. *Journal of hydrology*, 233(1):86–101.

Yue, Y. R. and Rue, H. 2011. Bayesian inference for additive mixed quantile regression models. *Computational Statistics & Data Analysis*, 55(1):84–96.