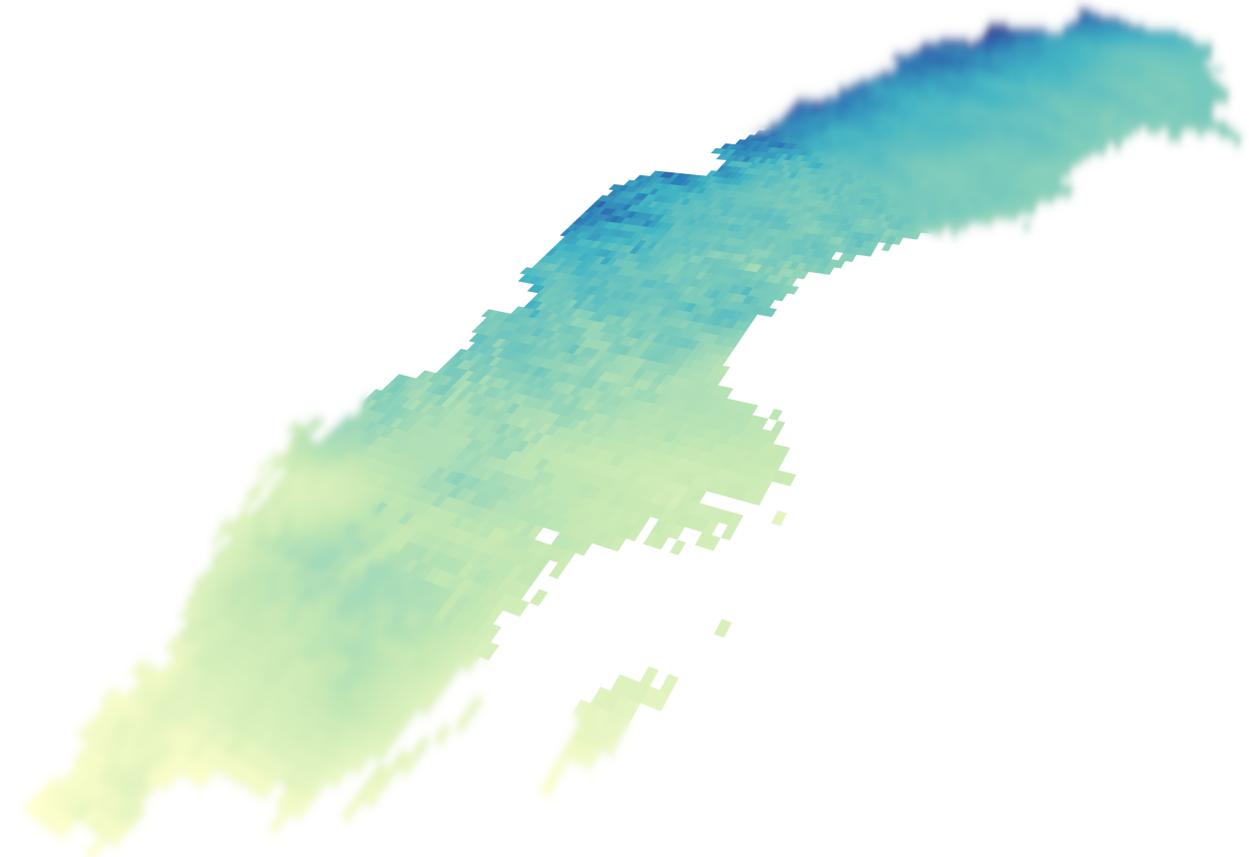




CHALMERS
UNIVERSITY OF TECHNOLOGY



Predicting Severe Snow Loads Using Spatial Extremes

Master's Thesis in Engineering Mathematics and Computational Science

PETER HANSSON
VINCENT SZOLNOKY

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

MASTER'S THESIS 2019

Predicting Severe Snow Loads Using Spatial Extremes

PETER HANSSON
VINCENT SZOLNOKY



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Predicting Severe Snow Loads Using Spatial Extremes
PETER HANSSON
VINCENT SZOLNOKY

© PETER HANSSON 2019.
© VINCENT SZOLNOKY 2019.

Supervisor: David Bolin, Department of Mathematical Sciences
Examiner: David Bolin, Department of Mathematical Sciences

Master's Thesis 2019
Department of Mathematical Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Stylized rendition of 50 year return levels of snow depth in Sweden

Typeset in L^AT_EX
Gothenburg, Sweden 2019

Predicting Severe Snow Loads Using Spatial Extremes

PETER HANSSON

VINCENT SZOLNOKY

Department of Mathematical Sciences Chalmers University of Technology

Abstract

Severe snow load, unlike severe snow fall, happens over a longer period of time for which snow accumulates causing it to produce increasing amounts of down force on the structure below it. When a structure is forced to bear a load it was not designed for, structural wear and damage can take place and in the worst case scenario, total failure in which the structure collapses.

To avoid such occurrences, extreme snow load should be modelled and the potential risks identified so as to improve laws and regulation pertaining to the maximum load a building must handle. This thesis makes use of spatial statistics and extreme value theory to analyse weather data and create models that can aid in predicting extreme snow depth which is directly linked to extreme snow load. Specifically interpolation by Kriging and non-stationary GEV methods are used to obtain predictions between stations. The results are compared to maps already published by the Swedish Building and Housing authority, Boverket, and the discrepancies between them show that certain regions in Sweden are currently being under-estimated. This under-estimation can lead to buildings being constructed to withstand loads less than what is predicted and therefore are at risk of structural fatigue. However in areas of heavy snowfall where the danger is greater, the map by Boverket actually overestimates predicted extreme values. Hence buildings constructed in areas where the risk of very high snow loads is prevalent, should be well future-proofed, and be able to withstand even more load than what is expected.

Keywords: geostatistics, extreme value theory, monte carlo simulation, kriging, weather

Acknowledgements

We would first and foremost thank our supervisor David Bolin for all his help and expertise in this project. Without his valuable input, interesting discussions, recommendations of literature, and correction of this paper, this paper would not have been possible.

We would also like to thank Boverket, SMHI and the Copernicus project for the free and easily accessible data sets.

We would also like to express our gratitude to all the contributors of the R community and its libraries. Special thanks goes to the authors and contributors of the R packages `geoR`, `gstat` and `extRemes`, which were used extensively in this project.

A special thanks goes to all our family and friends for the support during the work of this project.

Contents

List of Figures	xi
List of Tables	xiii
1 Introduction	1
1.1 Problem definition and aim	1
2 Theory	3
2.1 Extreme value theory	3
2.2 Non-stationary GEV Models	5
2.3 Spatial regression model	6
2.4 Estimation of variogram parameters	8
2.5 Kriging	10
2.6 Model validation	11
2.7 Error metric for the GEV parameters	11
2.8 A Monte Carlo method for uncertainty assessment	12
2.8.1 Unconditional Gaussian simulation	13
2.8.2 Conditional Gaussian simulation via Kriging	13
2.8.3 Conditional Gaussian simulation via LU-decomposition	14
3 Analysis of SMHI's meteorological data	15
3.1 The problem of missing observations	16
3.2 Refinement of data into winter months	16
3.3 Discarding years with few daily measurements	17
3.4 Initial fitment of point-wise GEV distributions	19
4 Method	21
4.1 Extreme value theory applied to annual snow depth maxima	21
4.2 Kriging	21
4.3 Choice of Covariates	22
4.4 Estimators of linear coefficients	22
4.5 Non-stationary GEV	25
4.6 Simulations	27
5 Results	31
6 Conclusion	35

List of Figures

1.1	Boverket's snow load map based on 50 year return levels [Olsson, 2011]	2
3.1	Average number of daily measurements per year	15
3.2	Basic monthly statistics for all stations	16
3.3	Month the yearly maximum was measured	17
3.4	Number of stations vs. number of years which they have at least 70% daily measurements	18
3.5	Plots of stations in Sweden. The most-left plot is all stations with at least one year. The middle plot and last plot are stations with at least 20 and 30 years respectively	19
3.6	Plot of each parameter after being estimated for each station individually. From left to right: location, scale and shape.	20
3.7	Plot showing which stations the shape parameter is significantly non-zero.	20
4.1	Kriging interpolation of GEV parameters	24
4.2	50 year return levels of snow depth (in metres) using Kriging interpolated parameters	25
4.3	Density plots of non-stationary Gumbel and GEV distributions respectively.	26
4.4	50 year return levels of snow depth (in metres) using linear models with coefficient estimates from MLE of a Gumbel and GEV model respectively.	27
4.5	Example realisations from the MC simulations of the 50 year return levels of snow depth in metres.	28
4.6	Summary of 1000 Monte Carlo simulations of the 50 year return levels of snow depth in metres. From left to right: The mean of all simulations, the upper limit of the 95% confidence interval, and the width of the 95% confidence interval.	29
5.1	50 year return levels of snow load (in kN/m ²). Figure (a) is with the Kriging prediction and figure (b) is with the upper limit 95% of the confidence interval from the MC simulations.	32

5.2	Discrepancy between estimated 50 year return levels of snow loads and Boverket's recommended snow loads (in kN/m^2) that a building should be able to withstand. Left is with the Kriging prediction and right is with the upper limit 95% of the confidence interval from the MC simulations.	33
-----	--	----

List of Tables

4.1	The optimal trends for each parameter respectively	23
4.2	Coefficients for intercept and 1st degree covariates in the external trend	23
4.3	Coefficients for 2nd degree covariates in the external trend	23
4.4	Coefficients for intercept and 1st degree covariates in the non-stationary GEV model	25
4.5	Coefficients for 2nd degree covariates in the non-stationary GEV model	26
5.1	Table of the error metric values defined in (2.37) for non-stationary GEV model, Kriging without MLE and Kriging with MLE, respec- tively. Leave one out cross validation for all stations was used to calculate the values.	31

1

Introduction

Climate change is an ever encroaching problem that is likely to change known, stable weather patterns making them more unpredictable and extreme. Snowfall is one such pattern that is sensitive to a number of factors impacted by climate change, but also for which the effects from extreme snow fall can be dangerous. It is for this reason extreme snow fall should be modelled and better understood so as to minimise risk and the potential for disastrous outcomes. A major concern with severe snow fall is when it accumulates on buildings, increasing the structural load they must bear. If a building was not designed to handle such an increase in load, it can be fatigued or, in the worst case scenario, collapse. This is for good reason concerning and something that the Swedish housing and building authority, Boverket, needs to have a good grasp of. This is so they can create laws and regulations that enforce a set of standards to ensure that such events are curtailed. This thesis aims to investigate extreme snow fall in Sweden and its impact on snow load so as to help improve the analysis and models Boverket use.

Currently Boverket distributes the map shown in figure 1.1 for deciding the amount of down force, due to snow, a building must be constructed to handle. It currently has a number of unknowns that affect its accuracy and reliability. There is little description of the model used to produce the map nor the data used. According to Boverket the data is from the Swedish Meteorology and Hydrological Institute (SMHI), and consists of “measurements of snow depth at a large number of stations during a long period of time”. As it will be later discussed in the paper however, the data from SMHI is highly irregular which brings into the question the accuracy of the model and whether these irregularities were properly handled. Additionally the map has been simplified into a contour plot for which values have been placed between isolines instead of on the isolines themselves. Similarly no measurement of uncertainty has been given, nor is there any specification for whether the values might already incorporate a measure of uncertainty.

1.1 Problem definition and aim

Although there are a number of uncertainties surrounding the current map that Boverket uses, it is not necessarily incorrect. Therefore the ultimate goal of this thesis is to verify whether this is the case or not. To start with, analysis will be performed on the raw data from SMHI and to try and clean it up and mitigate any faults in it. After which new models will be devised, tested and compared to identify the ones with the best performance and accuracy. Finally the results from

1. Introduction

the models will be used to create updated maps that are comparable to the map by Boverket and any incongruities can then be presented.

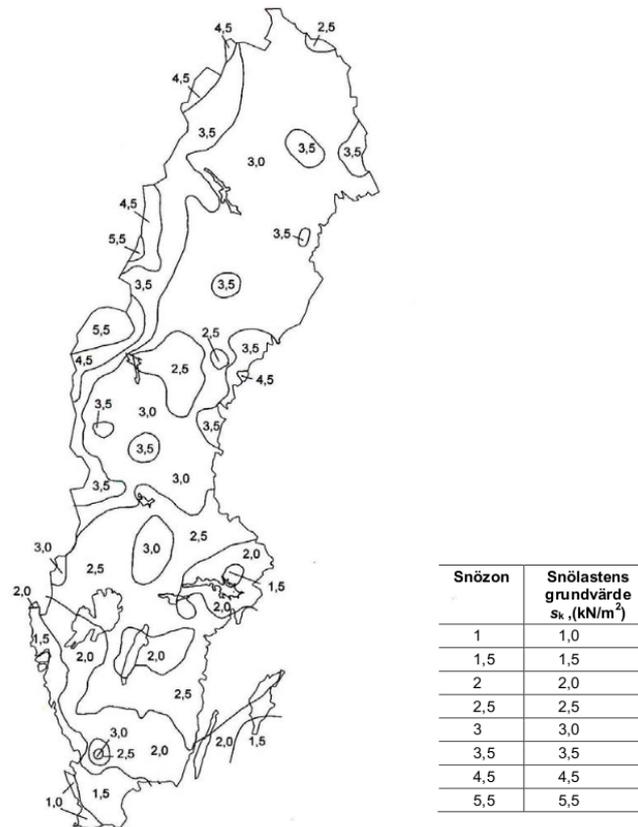


Figure 1.1: Boverket's snow load map based on 50 year return levels [Olsson, 2011]

2

Theory

2.1 Extreme value theory

Let $F(x)$ denote the CDF of the random variable \mathbf{X} . Consider n independent random variables $\mathbf{X}_1, \dots, \mathbf{X}_n$ drawn from this distribution and let $F_M(x)$ denote the distribution of the maximum value of $\mathbf{X}_1, \dots, \mathbf{X}_n$. The distribution of $F_M(x)$ is easily derived by

$$F_M(x) = P(\mathbf{X}_1, \dots, \mathbf{X}_n \leq x) = P(\mathbf{X}_1 \leq x) \cdots P(\mathbf{X}_n \leq x) = F^n(x). \quad (2.1)$$

Thus theoretically, if we want to estimate $F_M(x)$ based on a set of n samples from a random distribution, it is possible to first estimate $F(x)$ and then take the power of n of $F(x)$ to derive $F_M(x)$. However in practice, this method is often not robust because any small estimation error of $F(x)$ is amplified of the power of n , which can have a huge impact on $F_M(x)$, especially when n is large.

Another way, frequently used in practice, it to utilise the following theorem.

Theorem 1 *Let \mathbf{X} be a random variable. If there exist sequences of constants $a_n > 0$ and b_n such that*

$$\lim_{n \rightarrow \infty} P\left(\frac{\mathbf{M}_n - b_n}{a_n} \leq z\right) = G(z) \quad (2.2)$$

where G is a non-degenerate distribution function and \mathbf{M}_n is the maximum value of n independent samples from \mathbf{X} , then G belongs to one of the following families

$$G(z) = \exp\left(-\exp\left(-\frac{z-b}{a}\right)\right), \quad -\infty < z < \infty \quad (2.3)$$

$$G(z) = \begin{cases} 0, & z \leq b \\ \exp\left(-\left(\frac{z-b}{a}\right)^{-\alpha}\right), & z > b \end{cases} \quad (2.4)$$

$$G(z) = \begin{cases} \exp\left(\left(\frac{z-b}{a}\right)^\alpha\right), & z < b \\ 1, & z \geq b \end{cases} \quad (2.5)$$

for parameters $a > 0$, $b \in \mathbb{R}$ and $\alpha > 0$.

The three distribution families defined in (2.3), (2.4) and (2.5) are known as the Gumbel distribution, Fréchet distribution and Weibull distribution, respectively. It can easily be checked that the three families can be merged into a single CDF on the form

$$\text{GEV}(z; \mu, \sigma, \xi) = \exp \left[- \left(1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right)^{-\frac{1}{\xi}} \right], \quad z \in \left\{ z : 1 + \xi \left(\frac{z - \mu}{\sigma} \right) > 0 \right\}, \quad (2.6)$$

where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter and $-\infty < \xi < \infty$ is the shape parameter. The distribution in (2.6) is known as the *generalised extreme value* (GEV) distribution. When $\xi > 0$ the GEV distribution coincides with the Fréchet distribution and when $\xi < 0$ the GEV distribution coincides with the Weibull distribution. In the special case when $\xi = 0$, the GEV distribution as defined in (2.6) can not be used directly. The limit distribution of (2.6) as $\xi \rightarrow 0$ is used instead, which coincides with the distribution in (2.3) i.e. the Gumbel distribution. Theorem 1 and the result that the three families can be merged into one single distribution, is known as the Fisher-Tippett-Gnedenko theorem.

The values of a_n and b_n in (2.1) is rarely known in practice. Instead, the following result is used. If we assume n is large, then by (2.1)

$$P \left(\frac{\mathbf{M}_n - b_n}{a_n} \leq z \right) \approx G(z), \quad (2.7)$$

which is equivalent to,

$$P(\mathbf{M}_n \leq z) \approx G\left(\frac{z - b_n}{a_n}\right) = G'(z). \quad (2.8)$$

From (2.6), is easily to see that $G'(z)$ is also in the GEV family.

An alternative method to the method described earlier to estimate $F_M(x)$, is by dividing a set of samples from \mathbf{X} into equal sized blocks of size n . From the earlier discussion, assuming n is large, the maximum of each block follows approximately a GEV distribution. Hence it is possible to estimate a GEV distribution to the block maxima.

The independence assumption in Theorem 1 can be restrictive in many real applications. Consider, for example, the following situation where a stationary stochastic process $\mathbf{X}(t)$, $t \in \mathbb{R}$ is sampled at discrete time steps t_1, \dots, t_n . If the process is correlated in the sense that it exist t and t' such that $\text{Cov}(t, t') \neq 0$, then theorem 1 can not applied by using the block maxima method for estimating the distribution $\mathbf{X}(t)$. Luckily, there exists a generalisation of this theorem, which states that under some assumptions of the correlation between observations of the process, the maximum of the observations $\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)$ does asymptotically follows a GEV distribution. The correlation assumption implies informally, inter alia, that the dependency between any two samples from the time series with a minimal lag difference l , can be arbitrary small for some choice of l . A formal statement and proof of the theorem can be found in Leadbetter [1983].

For any probability $0 < p < 1$, we define the *return period* as $T = \frac{1}{p}$ and the *return level* as $z_p = Q(1-p)$, where Q is the quantile function corresponding to a given GEV distribution. Consider a stochastic time series which has been partitioned into equal sized blocks and assume that the maximum value follows a GEV distribution. Since the CDF of the GEV distribution is invertible, there is a bijective mapping between the return level and return period. Let the block size, i.e., the time difference between the first and last observed values in the blocks, be one time unit. For such a case, z_p is expected to be exceeded once every T time units. In extreme value theory the return period and return level are important statistics and since we are often interested in extreme events, the value of p is often small since it considers large values of the observations. The return level z_p of a given GEV distribution and return period $\frac{1}{p}$, is given by

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} \left[1 - (-\log(1-p))^{-\xi} \right], & \xi \neq 0 \\ \mu - \sigma \log(-\log(1-p)), & \xi = 0. \end{cases} \quad (2.9)$$

There exists various frameworks to estimate a GEV distribution to a set of observations. Notable methods are maximum likelihood estimation (MLE), generalised maximum likelihood estimation (GMLE), L-moments estimates and Bayesian methods. In this paper we restrict ourselves to using MLE only. One major advantage of using MLE is that the MLE-estimators are asymptotically normally distributed if certain regularity conditions are satisfied for the distribution of the model to be estimated. It can easily be showed that not all regularity conditions hold for the GEV distribution. For example, one of the regularity conditions states that the support must be constant, independent of the parameters. But it is clear from (2.6) that this is not true. Nevertheless, it has been shown that if $\xi > -\frac{1}{2}$, the ML-estimate of the parameters of the GEV distribution is normally distributed [Smith, 1985].

2.2 Non-stationary GEV Models

As derived in (2.6), the generalised extreme value distribution has three parameters μ , σ and ξ which are the location, scale and shape parameters respectively. For the stationary case of the GEV distribution, these parameters are constant and determined when constructing the distribution. This can however be expanded to the non-stationary case which includes trends for each parameter. This means that each parameter can be expressed as a linear function of covariates, such as spatial data or time. Therefore each parameter can be expressed as a sum of covariates and their corresponding coefficients:

$$\begin{cases} \mu_k = \mu(\mathbf{s}_k) = \beta_\mu^0 + \sum_{i=1}^{n_\mu} \beta_\mu^i X_\mu^i(\mathbf{s}_k) = \boldsymbol{\beta}_\mu^T \mathbf{X}_\mu \\ \sigma_k = \sigma(\mathbf{s}_k) = \beta_\sigma^0 + \sum_{i=1}^{n_\sigma} \beta_\sigma^i X_\sigma^i(\mathbf{s}_k) = \boldsymbol{\beta}_\sigma^T \mathbf{X}_\sigma \\ \xi_k = \xi(\mathbf{s}_k) = \beta_\xi^0 + \sum_{i=1}^{n_\xi} \beta_\xi^i X_\xi^i(\mathbf{s}_k) = \boldsymbol{\beta}_\xi^T \mathbf{X}_\xi \end{cases} \quad (2.10)$$

where \mathbf{s}_k is the spatial location with index k , n_μ , n_σ and n_ξ are the number of covariates for each respective parameter, $\boldsymbol{\beta}_\eta = (\beta_\eta^0, \dots, \beta_\eta^{n_\eta})^T$, $\mathbf{X}_\eta = (1, X_\eta^1, \dots, X_\eta^{n_\eta})^T$,

where η is any of the parameters μ , σ or ξ . Similarly β_μ , β_σ and β_ξ are the coefficients along with \mathbf{X}_μ , \mathbf{X}_σ , and \mathbf{X}_ξ being the covariates for μ , σ and ξ respectively. \mathbf{s}_k are the locations in the model for which one has covariate values for. The generalised likelihood function for this model is given by El Adlouni et al. [2007] as

$$L(\beta_\mu, \beta_\sigma, \beta_\xi | \mathbf{x}) = \prod_{k=1}^M \frac{1}{\sigma_k} \exp \left\{ - \left[1 - \xi_k \left(\frac{x_k - \mu_k}{\sigma_k} \right) \right]^{-\frac{1}{\xi_k}} \right\} \cdot \left[1 - \xi_k \left(\frac{x_k - \mu_k}{\sigma_k} \right) \right]^{-\left(1 - \frac{1}{\xi_k}\right)} \cdot \prod_{i=M+1}^N \frac{1}{\sigma_k} \exp \left\{ - \left(\frac{x_k - \mu_k}{\sigma_k} \right) \right\} \cdot \exp \left\{ - \exp \left[- \frac{x_k - \mu_k}{\sigma_k} \right] \right\}. \quad (2.11)$$

Where M is the number of observations where $\xi_k \neq 0$ and N the total number of observations. This likelihood can then be used with MLE to estimate the hyper parameters β_μ , β_σ and β_ξ .

The scale parameter σ is only viable for positive values in the GEV distribution. If the estimated model in (2.10) is used for prediction, there might exist prediction locations where σ_k is non-positive. For this reason, one often uses the following expression for σ_k

$$\sigma_k = \sigma(\mathbf{s}_k) = \exp(\beta_\sigma^T \mathbf{X}_\sigma), \quad (2.12)$$

which guarantees that prediction values for the scale parameter are always positive.

2.3 Spatial regression model

Consider a stochastic process \mathbf{Y} on a domain $\mathcal{D} \subseteq \mathbb{R}^d$. Let $\mathbf{Y}(\mathbf{s})$ denote the random variable at location $\mathbf{s} \in \mathcal{D}$ of the stochastic process \mathbf{Y} . In a *spatial regression model*, we consider stochastic model on the form

$$\mathbf{Y}(\mathbf{s}) = m(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}), \quad (2.13)$$

where $\boldsymbol{\epsilon}$ is a centred process on the domain \mathcal{D} , referred to as the *residual* of \mathbf{Y} and $m(\mathbf{s})$ is a deterministic value at location \mathbf{s} . Since $\boldsymbol{\epsilon}$ is centred, it holds that $E[\mathbf{Y}(\mathbf{s})] = m(\mathbf{s})$. The deterministic part $m(\mathbf{s})$ can be modelled in various ways and henceforth we will consider $m(\mathbf{s})$ to be linear function of some explanatory variables

$$m(\mathbf{s}) = \beta_0 + \sum_{i=1}^p X_{s,i} \beta_i, \quad (2.14)$$

where $X_{1,s}, \dots, X_{p,s}$ are explanatory variables for location \mathbf{s} and β_0, \dots, β_p are the linear coefficients. Equation (2.14) can be written in a vector format

$$m(\mathbf{s}) = \mathbf{X}_s^T \boldsymbol{\beta}, \quad (2.15)$$

where $\mathbf{X}_s = (1, X_{s,1}, \dots, X_{s,p})^T$ and $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T$. Now consider N locations $\mathbf{s}_1, \dots, \mathbf{s}_N$ and let \mathbf{Y}_N denote the joint distribution at the N locations of the stochastic

process \mathbf{Y} . Analogously, let $\boldsymbol{\epsilon}_N$ denote the joint distribution at the N locations of the residual. Then \mathbf{Y}_N can be written as

$$\mathbf{Y}_N = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_N, \quad (2.16)$$

where \mathbf{X} is a $N \times p$ matrix consisting of explanatory variables where $\mathbf{X}_{i,\cdot} = \mathbf{X}_{\mathbf{s}_i}^T$. Assume now that the residual process is a Gaussian process. Let further assume that the process is second order stationary and that the covariance structure of $\boldsymbol{\epsilon}$ is given by

$$\forall \mathbf{s}, \mathbf{t} \in \mathcal{D}, \quad \text{Cov}(\boldsymbol{\epsilon}(\mathbf{s}), \boldsymbol{\epsilon}(\mathbf{t})) = C(\|\mathbf{s} - \mathbf{t}\|) = C(\|\mathbf{h}\|) = C(h), \quad (2.17)$$

where $\|\cdot\|$ can be any norm in \mathbb{R}^d and will henceforth be the Euclidean norm, $\mathbf{h} = \mathbf{s} - \mathbf{t}$, $h = \|\mathbf{h}\|$ and $C(\cdot)$ is a positive definite function $C : \mathbb{R}^+ \rightarrow \mathbb{R}$. A stochastic process which satisfies the relationship in (2.17) is called an *isotropic* process. In contrast, a process which is not isotropic is called *anisotropic*. Since the residual is a centred Gaussian process, the *isotropic covariance* function $C(\cdot)$ completely defines the process. Also, $\boldsymbol{\epsilon}_N$ is distributed as

$$\boldsymbol{\epsilon}_N \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \quad (2.18)$$

where

$$\Sigma_{i,j} = C(\|\mathbf{s}_i - \mathbf{s}_j\|), \quad \forall i, j \in \{1, \dots, N\}. \quad (2.19)$$

Instead of describing the covariance structure of the process with the isotropic covariance function, it is often preferred in spatial statistics to use the *isotropic variogram* function $\gamma(h) = C(0) - C(h)$. The variogram can also be defined for a less restricted model, not assuming isotropy or stationarity

$$2\gamma(\mathbf{s}, \mathbf{t}) = \text{Var}(\mathbf{Y}(\mathbf{s}) - \mathbf{Y}(\mathbf{t})), \quad \gamma : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}^+. \quad (2.20)$$

There exists various models for the variogram function $\gamma(h)$. *Spherical*, *Gaussian*, *Exponential* and *Matérn* are a few variogram models which are commonly used. The Matérn variogram function, which is used in the method of this thesis, is given by

$$\gamma(h) = \sigma^2 \left[1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{h}{\rho} \right)^\nu \mathcal{K}_\nu \left(\frac{h}{\rho} \right) \right], \quad (2.21)$$

where ρ , ν and σ are positive parameters of the Matérn function, \mathcal{K}_ν is the modified Bessel function with parameter ν , and Γ is the gamma function. The mentioned variogram functions are all conditionally negative definite, which means that the matrix defined by

$$A = (a_{i,j})_{i,j=1}^n = C(\|\mathbf{s}_i - \mathbf{s}_j\|), \quad (2.22)$$

is positive semi-definite for all n and $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{D}$. For any stochastic process $\mathbf{X}(\mathbf{s})$, the corresponding matrix A in (2.22) is given by

$$A = (a_{i,j})_{i,j=1}^n = \text{Cov}(\mathbf{X}(\mathbf{s}), \mathbf{X}(\mathbf{s})). \quad (2.23)$$

It can be shown, that for all stochastic processes, the matrix A in (2.23) is positive semi-definite, for all n and $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{D}$. Consequently, the use of the mentioned variogram models does not violate this necessity.

To account for any measurement errors when sampling from \mathbf{Y} , an extra stochastic process $\boldsymbol{\varepsilon}$ can be added to model in (2.13). The stochastic process added for this purpose of incorporating measurement errors is called *nugget effect*. The nugget effect is often modelled as a centred Gaussian model with covariance structure

$$\text{Cov}(\boldsymbol{\varepsilon}_{\mathbf{s}_i}, \boldsymbol{\varepsilon}_{\mathbf{s}_j}) = \begin{cases} \sigma_{\mathbf{s}_i}^2, & \text{if } \mathbf{s}_i = \mathbf{s}_j \\ 0, & \text{if } \mathbf{s}_i \neq \mathbf{s}_j. \end{cases} \quad (2.24)$$

The nugget defined in (2.24) is quite general and the variance is often modelled such that variance is location independent i.e. $\sigma_{\mathbf{s}_i} = \sigma, \forall \mathbf{s}_i \in \mathcal{D}$. The corresponding isotropic variogram function for location independent nugget is given by

$$\gamma(h) = C(0) - C(h) = \begin{cases} 0, & \text{if } h = 0 \\ \sigma^2, & \text{if } h > 0. \end{cases} \quad (2.25)$$

The variogram is additive in the sense that if two independent isotropic processes \mathbf{Y}_1 and \mathbf{Y}_2 with variogram γ_1 and γ_2 , then the process defined by $\mathbf{Y}_1 + \mathbf{Y}_2$ is also isotropic with variogram $\gamma_1 + \gamma_2$. This means that when a nugget effect is added to a variogram model, the variogram function is only elevated by a constant equal to σ^2 for $h > 0$. For many variogram models, the corresponding variogram function is continuous for $h \geq 0$. But when a nugget with a non zero variance is added to the model, the resulting variogram function is discontinuous at $h = 0$.

Analogously to the denotation of the joint distribution in (2.16), the joint distribution of \mathbf{Y}_N in the case of an nugget effect, is given by

$$\mathbf{Y}_N = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_N + \boldsymbol{\varepsilon}_N. \quad (2.26)$$

2.4 Estimation of variogram parameters

In practical applications it is often desired to estimate the parameters of a spatial model from a set of observations on the stochastic process \mathbf{Y} . For this purpose, we will describe two different methods, which often are used in estimating a spatial linear model. Inference on this type of model estimates both the linear covariate coefficients β_0, \dots, β_p as in (2.14) and the parameters of the variogram function, and in the case when nugget is included in the model and is unknown prior to the inference, the nugget variance is also estimated.

Let $\mathbf{Y}' = (Y_1, \dots, Y_N)^T$ denote the N observations of a single outcome of the stochastic process \mathbf{Y} at locations $\mathbf{s}_1, \dots, \mathbf{s}_N$. Let further $\hat{\boldsymbol{\beta}}$ denote the estimate of the vector $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ the vector of parameter values of the variogram model with corresponding estimate $\hat{\boldsymbol{\eta}}$. The first inference method will now be described. A first estimate of $\boldsymbol{\beta}$ is given by ordinary least squares (OLS)

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{Y}'\| = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}'. \quad (2.27)$$

With first estimate we mean in the sense that it can later be heuristically improved, which will be described later. An approximate realisation of the stochastic part ϵ_N in (2.16) is given by $\hat{\epsilon}_N = \mathbf{Y}' - \mathbf{X}\hat{\beta}$. In the case of a nugget effect as in (2.26), $\mathbf{Y}' - \mathbf{X}\hat{\beta}$ is an approximate realisation of $\epsilon_N + \epsilon_N$. An estimate of $\gamma(h)$ is given by

$$\hat{\gamma}(h) = \frac{1}{2\#N(h)} \sum_{i,j \in N(h)} (\hat{\epsilon}_i - \hat{\epsilon}_j)^2, \quad (2.28)$$

where

$$N(h) = \{i, j : \|\mathbf{s}_i - \mathbf{s}_j\| \in [h - \tau, h + \tau], \mathbf{s}_i, \mathbf{s}_j \in \mathcal{D}\}, \quad (2.29)$$

$\#N(h)$ is the number of elements in the set $N(h)$ and τ is a positive tolerance constant. Let now calculate $\hat{\gamma}(h)$ for a set of values h_1, \dots, h_n with corresponding sets $N_i(h_i), \dots, N_n(h_n)$. In practice, the points h_1, \dots, h_n are often equally spaced. Also, the sets $N_i(h_i), \dots, N_n(h_n)$ are mutual exclusive and all distances between the pairwise permutations of locations are contained in the union of the sets. Now the chosen variogram function γ is fitted to the set of points $(h_1, \hat{\gamma}_1), \dots, (h_n, \hat{\gamma}_n)$. This is often done in practice by solving the weighted non-linear least square problem on the form

$$\hat{\boldsymbol{\eta}} = \operatorname{argmin}_{\boldsymbol{\eta}} \sum_{i=1}^n W_i (\hat{\gamma}(h_i) - \gamma(h_i; \boldsymbol{\eta}))^2, \quad (2.30)$$

where the weight W_i for all i is given by $\frac{\#N(h_i)}{\gamma(h_i; \boldsymbol{\eta})}$. This choice for W_i considers to some extent the difference in reliability of the values $\hat{\gamma}_1, \dots, \hat{\gamma}_n$, since the number of observations used in the estimates of (2.29) differs.

At this stage, all the parameters of the model have been estimated. But as was mentioned earlier, the covariates estimates contained in the vector $\hat{\beta}$ can be further improved. This is done by using *generalised least squares* (GLS), which is a generalisation of OLS and new estimate of $\hat{\beta}$ is given by

$$\hat{\beta} = \operatorname{argmin}_{\beta} (\mathbf{Y}' - \mathbf{X}\beta)^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y}' - \mathbf{X}\beta) = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}', \quad (2.31)$$

where $\boldsymbol{\Sigma}$ is the estimated covariance structure of the joint distribution ϵ_N defined by the function $\hat{\gamma}(h; \hat{\boldsymbol{\eta}})$. It is straightforward to derive that the estimator of β in (2.31) is actually the ML-estimator of the parameters for known variogram in (2.16) if the residual is Gaussian. A new realisation of $\hat{\epsilon}_N$ and $\boldsymbol{\eta}$ can now be estimated analogously to how it was estimated previously. This cyclic framework of estimating $\hat{\beta}$ and $\boldsymbol{\eta}$ can be done any number of times and it is not unusual to repeat the process more than twice.

The second used inference method to estimate the parameters of the linear spatial model is a MLE-based approach. The likelihood function is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma_n^2) = \mathcal{N}(\mathbf{Y}'; \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}} + \mathbf{I}\sigma_n^2), \quad (2.32)$$

where \mathcal{N} is the multivariate normal distribution which is given by

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (2.33)$$

where k is the number of elements in \mathbf{x} .

Estimation of the parameters is found simply by maximising (2.32) in regards to the parameters. There exists no closed form expression for maximising (2.32), for any general choice of variogram model and thus a non linear optimisation solver is needed. The MLE-inference based approach is of special interest since it is possible to fix a location dependent nugget i.e. we fix each location \mathbf{s}_i to have a fixed nugget with variance σ_i^2 . Then the following ML function is maximised

$$L(\boldsymbol{\beta}, \boldsymbol{\eta}) = \mathcal{N}(\mathbf{Y}'; \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_\eta + \boldsymbol{\sigma}^2), \quad (2.34)$$

where

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \end{pmatrix}.$$

This model can be used when the variance of the measurement error is known or can be estimated prior to the inference.

2.5 Kriging

Kriging is a geostatistical method used to estimate values at unobserved locations by interpolating values from a number of surrounding observed locations. It does so by modelling the data as a random field together with some predetermined covariances. To obtain the new interpolated values, the *Kriging weights* are calculated and are used to compute a weighted sum of values from nearby locations. *Universal kriging* is one of the Kriging methods, which assumes that the distribution of \mathbf{Y} follows a spatial linear model.

Assume that we have a random field $\mathbf{Y}(\cdot)$ which follows a linear spatial model and an arbitrary unobserved location \mathbf{s}_0 at which we want to estimate the value of \mathbf{Y} at. Universal Kriging relies on that the covariance structure $\boldsymbol{\Sigma}$ is known for the observed locations, which can be estimated from any of the methods discussed in section 2.4 from the estimated variogram function. Let \mathbf{c} denote the vector of covariance structure between the unobserved stations and the observed stations i.e. $\mathbf{c} = (\text{Cov}(Y_{\mathbf{s}_0}, Y_{\mathbf{s}_1}), \dots, \text{Cov}(Y_{\mathbf{s}_0}, Y_{\mathbf{s}_N}))^T$, which is analogously determined from the variogram function. Let further $\mathbf{X}_{\mathbf{s}_0}$ denote the values of the explanatory values, as in section 2.3, for the unobserved location. The Universal Kriging predictor for location \mathbf{s}_0 is calculated from

$$\hat{Y}(\mathbf{s}_0) = \left[\mathbf{c}^T \boldsymbol{\Sigma}^{-1} + (\mathbf{X}_{\mathbf{s}_0} - \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{c})^T (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-T} \mathbf{X} \boldsymbol{\Sigma}^{-1} \right] \mathbf{Y}_N. \quad (2.35)$$

The Kriging estimation is obviously linear, i.e., it can be written on the form $\hat{Y}(\mathbf{s}_0) = \boldsymbol{\lambda}^T \mathbf{Y}_N$. If a Gaussian residual is assumed of the linear spatial model, the estimation

is unbiased ,i.e., $E[\hat{Y}(\mathbf{s}_0)] = E[Y(\mathbf{s}_0)]$. With these two properties, the Kriging predictor is classified as the best linear unbiased predictor (BLUP). This means that it is a linear predictor that minimises the variance of the prediction error while still maintaining an unbiased output. It is also possible to calculate the variance of the Kriging prediction error $\sigma^2(\mathbf{s}_0)$ from

$$\sigma^2(\mathbf{s}_0) = \sigma_0^2 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} + (\mathbf{X}_{\mathbf{s}_0} - \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{c})(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}(\mathbf{X}_{\mathbf{s}_0} - \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{c}), \quad (2.36)$$

where σ_0^2 is the variance of $\mathbf{Y}(\mathbf{s}_0)$.

Unless there is zero nugget in the linear spatial model, Kriging is not a perfect interpolation method in the sense that the interpolant at an observed location is not necessary equal to the observed value at the location. This means that Kriging supports improved measurements at the observed locations in case of any measurement errors included of the observed data.

2.6 Model validation

To be able to compare models without having to compute any complex statistics, a flexible framework is needed that places on assumptions on the models. Cross-validation is a technique that operates purely on the output of the models and assesses their predictive performance.

Consider a model we want to evaluate which can predict unobserved data from observed data from a number of locations $\mathbf{s}_0, \dots, \mathbf{s}_N$. Through an iterative process, we remove a number of locations, the *validation stations*, from the observed ones and assume the observed values for these stations are unknown. The model is then trained on the remaining locations, the *training stations*, and the predictions for the validation stations are obtained. Any error metric e.g. mean-squared error can then be computed between the previously removed known data and the corresponding predicted values for the data. This process is furthermore repeated a number of times where at each new iteration, a different set of observed values are removed. This is to work around any bias the model might have to a certain subset of the data for which it performs better or worse than average. For each iteration, the error is calculated, i.e., the difference between real value and predicted value by the model of the validation stations.

2.7 Error metric for the GEV parameters

Consider a model which predicts the GEV parameters μ , σ and ξ for a set of locations where the parameters are unknown. Using cross validation for validating this model requires a suitable metric for calculating the error between the predicted and real values of the GEV parameters. A possible metric could be any metric which considers only the predicted and the real parameters, e.g., $\sqrt{(\mu - \hat{\mu})^2 + (\sigma - \hat{\sigma})^2 + (\xi - \hat{\xi})^2}$. But it is also possible to include, if any, observed realisations which are assumed to follow a GEV distribution, into the metric. This is especially preferred when the

parameters is not the real parameters of the distribution, but only estimated from observed data. A frequently used method to include the observed data into the metric is to calculate the difference between the empirical quantile function, which are determined from the observed values, and the quantile function of the estimated distribution which is determined by the predicted parameters. This is done by sorting the observed values, where $q_{i,k}$ is the k 'th sorted value for location \mathbf{s}_i . Let K_i be the number of observed values at location \mathbf{s}_i and let $q'_{i,k} = Q(p_{k,i}; \hat{\mu}_i, \hat{\sigma}_i, \hat{\xi}_i)$ where $p_{k,i} = \frac{k-0.5}{K_i}$ and Q is the quantile function of the GEV distribution for the estimated parameters at location \mathbf{s}_i . The following error metrics can be used to validate the estimated parameters

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{1}{K_i} \sum_{k=1}^{K_i} (q'_{i,k} - q_{i,k})^2} \quad (2.37)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \frac{1}{K_i} \sum_{k=1}^{K_i} |q'_{i,k} - q_{i,k}| \quad (2.38)$$

$$\text{MPE} = \max_{\{i=1, \dots, N\}} \max_{\{k=1, \dots, K_i\}} |q'_{i,k} - q_{i,k}| \quad (2.39)$$

$$\text{BIAS} = \frac{1}{N} \sum_{i=1}^N \frac{1}{K_i} \sum_{k=1}^{K_i} (q'_{i,k} - q_{i,k}), \quad (2.40)$$

where N is the number of validation stations.

2.8 A Monte Carlo method for uncertainty assessment

The problem with the previous methods of interpolation is that it is difficult to get a reliable measure of the total uncertainty for the interpolated values. Uncertainty is introduced in multiple steps during the analysis, and there is no simple way to reconcile all of them into a single value for assessment. A technique that is commonly employed in this kind of scenario is the Monte Carlo (MC) method.

MC simulation itself is a very general concept but the driving principle behind MC simulation, is to perturb the inputs to the function one wants to assess and observe the deviation of the output. This is usually done by assigning a distribution to the input parameters from which different input values are sampled from. The corresponding output is then recorded and the process is done many times to evaluate the distribution of the output from which one can draw conclusions from about eventual risk or propagated uncertainties.

A common example used to illustrate how the MC method works is to estimate the mathematical constant π . You first start with a unit square, and inscribe in it a circle with radius 0.5. We know that the area of the circle is πr^2 , where r is the radius being 0.5. Therefore we can calculate the ratio between the area of the circle and the area of the square to be $\frac{\pi \cdot 0.5^2}{1 \cdot 1} = \frac{\pi}{4}$. Now a large number of uniformly distributed points are sampled in the square. The uniform distribution acts as the distribution to sample the inputs from, as described earlier and the function we

want to evaluate is to check if the input is within the circle or not. For each input sampled from the distribution we keep track of the points that fall within the circle or not. If we calculate the ratio between the number of points within the circle and the total number of points, we should get an approximation of the same area ratio from before. Therefore we can estimate π to be $\pi \approx 4 \frac{N_{\text{inner}}}{N_{\text{total}}}$, where N_{inner} are the number of points inside the circle and N_{total} are the total number of points. If we only sample a small number of points, then the approximation will not be very good, but the more points that are sampled the better the approximation becomes. For our specific case, the goal is to create multi-variate distributions for the two GEV parameters location and scale, over the surface of Sweden, from which they can freely be sampled from. For each sample of them, the corresponding return levels are calculated which allows for observations to be made about how they vary as a whole. Because the underlying distribution of return levels is unknown, MC simulations provides a flexible way to analyse them. This way we can get idea of the uncertainty in the return levels themselves and therefore better handle possibilities of higher risk than first anticipated purely by an interpolation method.

2.8.1 Unconditional Gaussian simulation

An unconditional simulation of a random field is the method of obtaining a realisation of that field with its intrinsic characteristics. Assuming one has a covariance matrix \mathbf{C} from the field associated with locations $\mathbf{s}_1, \dots, \mathbf{s}_N$ at which one wants to simulate. One can obtain such a simulation by means of LU-decomposition proposed by Davis [1987]. Using LU-decomposition, \mathbf{C} can be decomposed into the product of a lower and upper triangular matrix as such $\mathbf{C} = \mathbf{L}\mathbf{U}$. If one constructs a new random vector $\mathbf{z} = \mathbf{L}\mathbf{y}$, where each element of the vector \mathbf{y} is $y_n \sim N(0, 1)$, then it can be shown that \mathbf{z} also has covariance \mathbf{C} [Davis, 1987]. This in turn means that \mathbf{z} is in fact a realisation of the random field with covariance matrix \mathbf{C} . It should be noted that because \mathbf{C} is a covariance matrix, it is therefore symmetric and positive semi-definite. Therefore one should use Cholesky decomposition to decompose \mathbf{C} into a lower triangular matrix and its conjugate compose as $\mathbf{C} = \mathbf{L}\mathbf{L}'$. This is because Cholesky decomposition is much more efficient at performing LU-decomposition than the standard algorithm.

2.8.2 Conditional Gaussian simulation via Kriging

An extension of unconditional simulation is to condition it so that locations that lie near ones which already have been observed, have simulated values that more closely resemble the observed values. Formally one wants to simulate a field \mathbf{X} at unobserved locations \mathbf{s}_u given already observed data at locations \mathbf{s}_o to obtain realisations \mathbf{z} such that $\mathbf{z} \sim (\mathbf{X}(\mathbf{s}_u) | \mathbf{X}(\mathbf{s}_o))$.

The simplest method for simulating a conditional random field is Gaussian simulation via Kriging [Emery, 2007]. The steps involve first performing Kriging on the observed data yielding predictions $\hat{\mathbf{z}}$ at the unobserved locations. Then an unconditional simulation is performed, usually using the technique in the previous section 2.8.1, producing a realisation \mathbf{z}_{sim} . With the unconditional simulation, Kriging is

performed together with the data at observed locations from the simulation itself. This gives predictions for the unconditional simulation \hat{z}_{sim} . The error is then calculated between the actual simulation and the prediction of the simulation giving the residuals $z_{\text{sim}} - \hat{z}_{\text{sim}}$. The final conditional simulation is then constructed as

$$\underbrace{z_{\text{condsim}}}_{\text{Conditional simulation}} = \underbrace{\hat{z}}_{\text{Kriged from observed data}} + \underbrace{(z_{\text{sim}}(\mathbf{s}) - \hat{z}_{\text{sim}}(\mathbf{s}))}_{\text{Kriging error from unconditional simulation}}. \quad (2.41)$$

2.8.3 Conditional Gaussian simulation via LU-decomposition

An issue with simulating via Kriging, is that it is very computationally intensive as a Kriging prediction must be done for every new realisation, unless one does optimisations to store the decomposed covariance matrix. Therefore many other methods have been devised to speed up this process. One such method is done via LU-decomposition in a similar way as the unconditional simulation [Davis, 1987]. This method allows quick simulation of many realisations in one single step.

If one goes by the same assumptions as in the unconditional case but this time with the distinction that the locations are divided up into observed and unobserved, $\mathbf{s} = [\mathbf{s}_1; \mathbf{s}_2]$. First, is to construct the covariance matrix \mathbf{C} by using the covariance matrices \mathbf{C}_{11} and \mathbf{C}_{22} which are the auto-covariance matrices of the observed points \mathbf{s}_1 and \mathbf{s}_2 respectively. \mathbf{C}_{12} and \mathbf{C}_{21} are the cross-covariance matrices between the observed and unobserved points and are each other's transpose.

Performing LU-decomposition on this construction yields

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \mathbf{L}\mathbf{U} = \begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ 0 & \mathbf{U}_{22} \end{bmatrix}. \quad (2.42)$$

By using the lower triangular matrix \mathbf{L} from equation (2.42), an unconditional simulation can be obtained as

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}. \quad (2.43)$$

To condition this simulation, \mathbf{y}_1 is replaced with $\mathbf{L}_{11}^{-1}\mathbf{x}_1$ where \mathbf{x}_1 are the values at observed locations \mathbf{s}_1 .

$$\begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{11}^{-1}\mathbf{x}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{L}_{21}\mathbf{L}_{11}^{-1}\mathbf{x}_1 + \mathbf{L}_{22}\mathbf{y}_2 \end{bmatrix} \quad (2.44)$$

In conclusion equation 2.44 will produce the conditional simulation and multiple simulations can be obtained by re-sampling \mathbf{y}_2 from a $N(0, 1)$ multivariate normal distribution. It is also worth mentioning that in the simulated term, $\mathbf{L}_{21}\mathbf{L}_{11}^{-1}\mathbf{x}_1$ can be seen as the conditional term while $\mathbf{L}_{22}\mathbf{y}_2$ is the unconditional term, as these will be referenced later.

One requirement that Davis [1987] states is that the data \mathbf{x}_1 should be normally distributed. This is generally a requirement when doing any form of Gaussian simulation. This can be done by performing a normal score transformation before simulation and then back transforming afterwards.

3

Analysis of SMHI's meteorological data

The data was obtained from the Swedish Meteorological and Hydrological Institute *SMHI* and contained daily measurements of the snow depth at multiple stations around Sweden between 1939 and the 2019.

The data set from SMHI is not complete in the sense that daily snow depth measurements are missing, and for the majority of the data up until 2007, large amounts of measurements are missing. Additionally not all stations have been recording measurements for equal time periods. Figure 3.1 shows the average number of daily measurements for each year. For most of the period the average is well below 365 days. This introduces difficulties when attempting to model the data with extreme value methods that will have to be solved before proceeding.

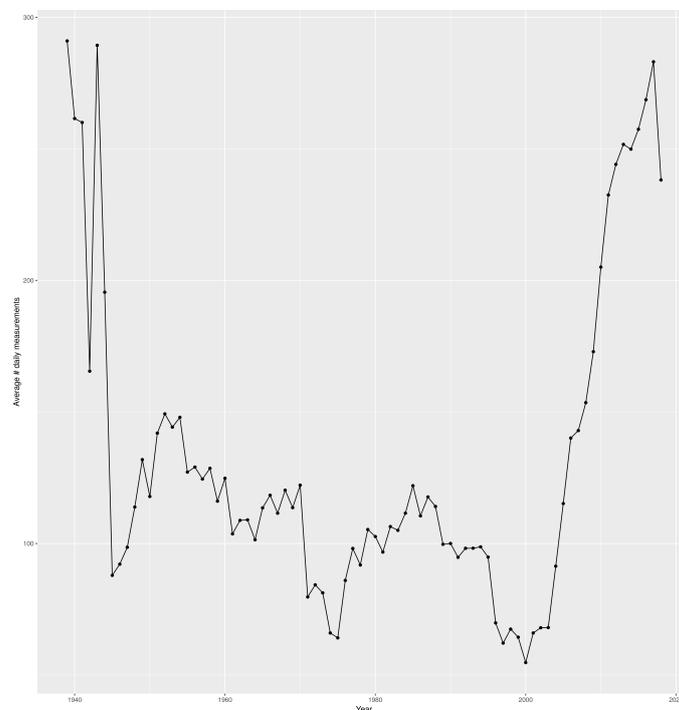


Figure 3.1: Average number of daily measurements per year

3.1 The problem of missing observations

To extract the extreme value data we want to retrieve the maximum value that was observed during each given year. Initially the data was grouped year-wise and the maximum value was calculated for each group. A problem arises in the fact that if daily measurements are missing then there is a chance that the true maximum was not recorded. This issue is amplified when fitting an extreme value model to this data as the extreme values will be underestimated. By underestimating the extreme values, one will underestimate potential risk and therefore not draw correct conclusions which are usually critical when risk is involved. Unsurprisingly, this problem has also been documented by SMHI themselves when extracting record snow depths from daily snow measurements [SMHI, 2016].

3.2 Refinement of data into winter months

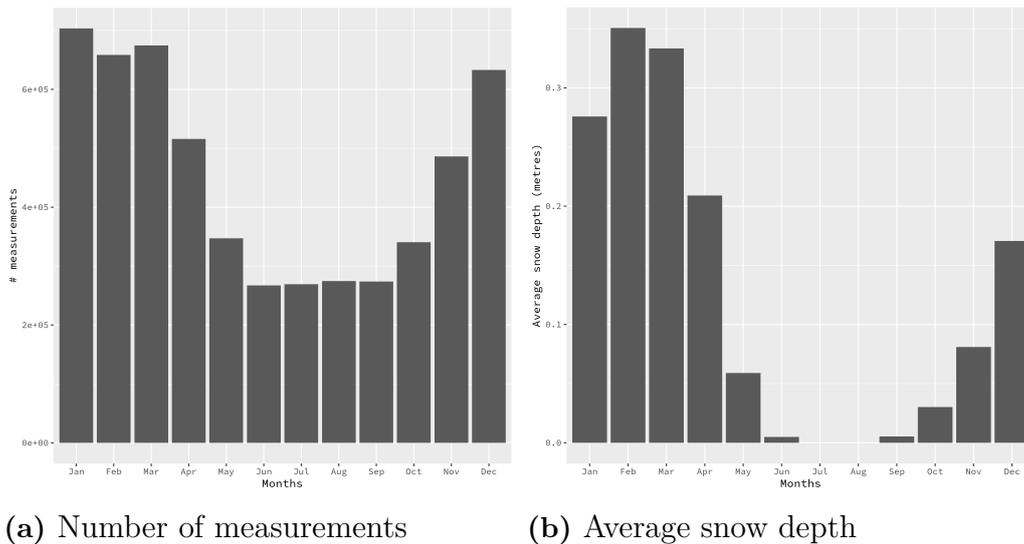


Figure 3.2: Basic monthly statistics for all stations

In figure 3.2a is a histogram of the total number of measurements for each month. There is a steep decline during the summer months which is due to the fact that it is highly unlikely to snow then and therefore unnecessary to perform any measurements. Additionally in figure 3.2a during the same months when fewer measurements are performed the average snow depth is also very low (in most of Sweden it is 0). Using this information we only consider the months during the “winter months”, which will be from December 1st to May 1st.

This is further backed up by figure 3.3 which shows in which month the yearly max was measured. This means that we can focus on extracting stations that have a good number of measurements during these months alone.

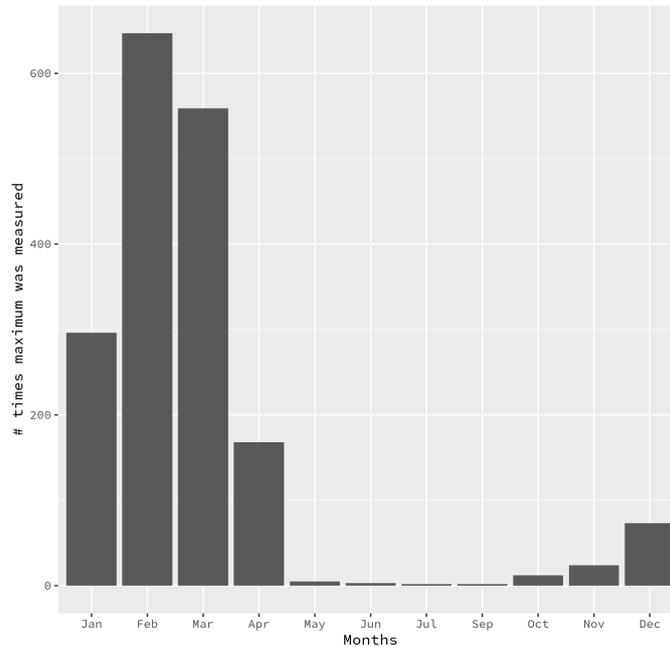


Figure 3.3: Month the yearly maximum was measured

3.3 Discarding years with few daily measurements

Another concession we make to try to improve the quality of the data, is to discard years in which a station does not have enough daily measurements. Ideally the threshold which defines how many measurements a station must have for a given year for it to be included, should be high so as to minimise the risk of missing an extreme observation. Fortunately in the case of snow depth which is an accumulative process, the measurement is unlikely to drastically change between daily measurements. Contrast to, for example, precipitation which can change hourly. This is due to the fact that the snow stays on the ground for long periods of time, hence having a small number missing daily observations is unlikely to miss an extremal value. Even if it is missed, the chance of surrounding observations to be close to that extremal value is still great and should not be far off. In summary this means that in our particular case, we can lower the threshold for the number of needed daily observations, which in turn increases the number of years we can include for a multitude of stations. We approximate a good threshold to be that at least 70% of the days between December 1st and May 1st, for a given year, should have measurements. Such a year we will call an *active year*.

The intuition behind figure 3.4 is that we want to find an ideal trade-off between total number of stations such that we cover enough of Sweden and retain enough spatial data. While at the same time ensuring that each station included has enough active years such that estimating a GEV distribution produces faithful estimation of the parameters. As the number of years on the x-axis decreases, so does the number of stations that have measured **at least** that many years.

Figure 3.4 shows that slightly less than half of the stations haven't even obtained one year with more than 70% daily measurements (number of stations with at least 1

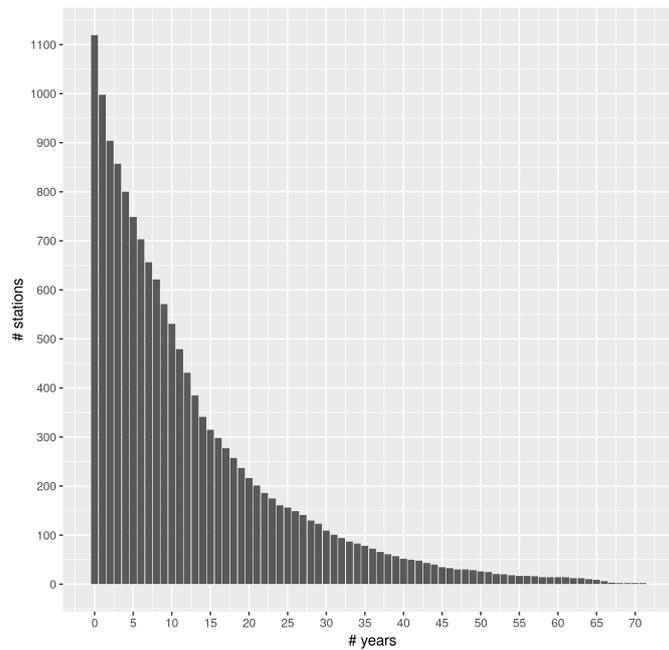


Figure 3.4: Number of stations vs. number of years which they have at least 70% daily measurements

active year is 1100). At 20 years there is a good population of 234 stations, and at 30 years it has further decreased to 134. In the paper by Blanchet and Lehning [2010], the authors determined that using 100 stations for their analysis of maximum snow depth in Switzerland was adequate. The 100 stations that have at least 30 active years do not cover the northern part of Sweden very well so a compromise is made to instead reduce the limit to 20 active years which brings the total back up to 234. 20 measurements is probably the lowest amount of one wants to estimate a GEV from but this is just a lower bound and many stations have more measurements than that.

The plots in figures 3.5 shows where the stations are located after filtering. In total there are 1800 stations that were active at some point, however more than half of these have not even been active for more than a year. Filtering for stations that have at least one year where they were actively conducting measurements shows that there were 759 active stations, displayed in figure 3.5a. Figures 3.5b and 3.5c show the spatial trade-off that occurs when requiring more active years per station.

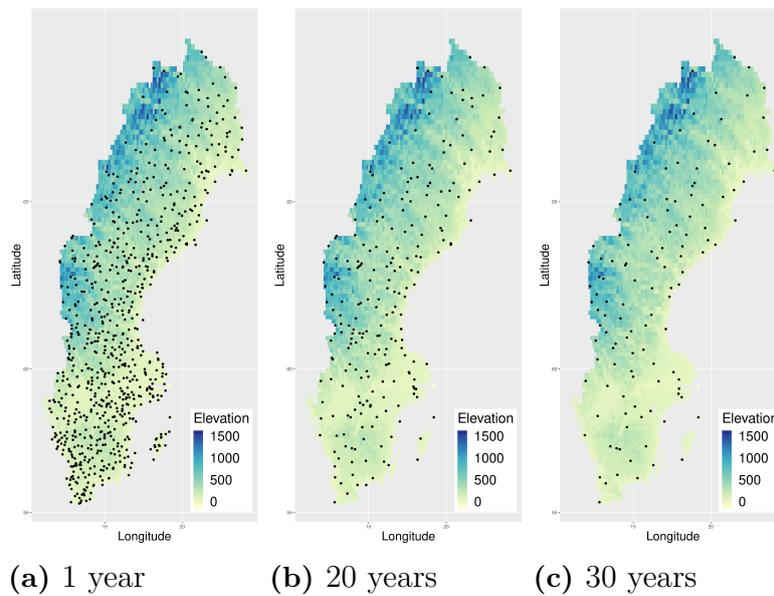


Figure 3.5: Plots of stations in Sweden. The most-left plot is all stations with at least one year. The middle plot and last plot are stations with at least 20 and 30 years respectively

3.4 Initial fitment of point-wise GEV distributions

To get an overview of what the data looks like from an extreme value perspective, we fit a GEV distribution to each station individually. This is using the filtered data from figure 3.5b together with MLE estimation of the parameters from section 2.1. The results are displayed in figure 3.6.

The location parameter, which is akin to the mean of the distribution, is seen to increase with latitude in figure 3.6a. This seems reasonable as it snows more in the north than in the south. The other two parameters in figures 3.6b and 3.6c do not have any clear trends that can initially be seen by looking at the plots. An interesting observation however, is that the shape parameter is mostly negative which has the effect of placing an upper limit on the tail of the GEV distribution. Generally this is not wanted and in other examples of modelling weather patterns the shape parameter is either explicitly set to zero or used only when greater than zero. This is also the case with the map produced by SMHI which uses an explicit Gumbel distribution with the shape parameter set to 0 across Sweden.

To ensure that this is valid, a simple check to see whether the shape parameter is statistically significantly non-zero, is done. Figure 3.7 shows the results and an interesting pattern can be seen that divides Sweden into two distinct sections. A north-western area where it is significantly non-zero indicating a GEV distribution and a south-eastern area where it is the opposite and not significantly non-zero, so a Gumbel distribution seems most appropriate for this area. A possibility is to use two distribution assumptions, GEV for the north-west and Gumbel for the south-east, however this adds an extra layer of complexity. Additionally, the shape

3. Analysis of SMHI's meteorological data

parameter is negative for where it is significantly non-zero which means that there is an upper limit on the tail of the GEV distribution. This is generally uncommon when modelling weather patterns and it is usually set either to 0 or ensured that it is positive. So for this article the same assumption as the one made in the map by SMHI, i.e. a Gumbel distribution for the entire Sweden, will be used.

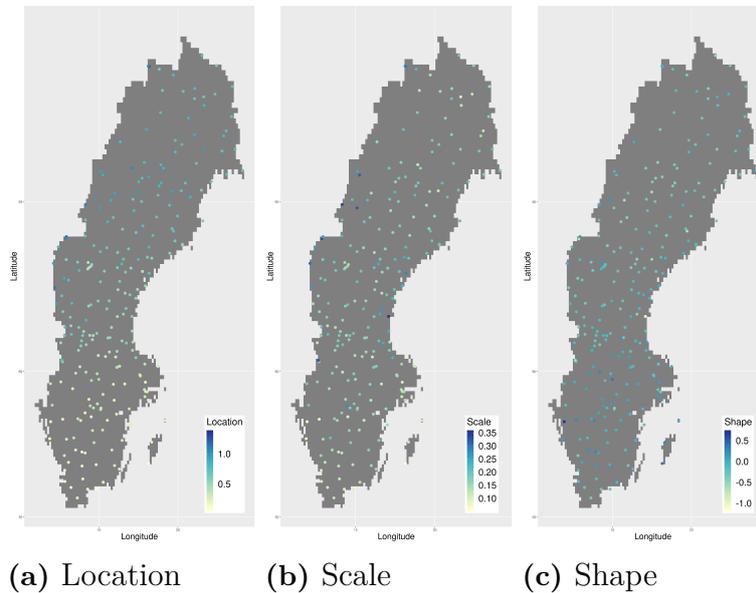


Figure 3.6: Plot of each parameter after being estimated for each station individually. From left to right: location, scale and shape.

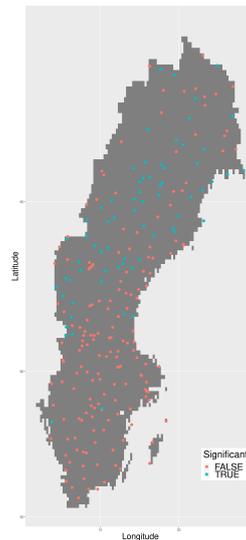


Figure 3.7: Plot showing which stations the shape parameter is significantly non-zero.

4

Method

4.1 Extreme value theory applied to annual snow depth maxima

In the previous section the filtering of the data to obtain annual snow depth maxima for the observed stations was discussed. The premise of this thesis is that the annual maxima for a given location will follow a GEV distribution. The rationale is that the daily observations are blocked into annual maxima and by theorem 1, assuming the assumptions in the theorem holds, will asymptotically follow a GEV distribution. This means that the daily observations must be equally distributed. This is of course by intuition not true, since generally speaking, it will snow less in the summer than in the winter. Nevertheless, it will still be assumed that the annual maxima follows a GEV distribution, which of course for the reasons mentioned, is only an approximation. This is very standard for any GEV analysis. This approximation can at least be a little justified since we are mainly considering snowfall for the months of the year which snowfall is frequent, as discussed in the previous section. The considered months, by intuition, will differ less distribution wise compared, let us say, to a month in the summer and in the winter.

The daily snow depth observations are of course correlated. This again contradicts the independence assumption of theorem 1. But as discussed in section 2.1, there exists a generalisation of the theorem which considers dependency of the observations. This relaxation of the independency assumption states informally, that observations are almost uncorrelated, given that the lag difference between the observations are large. A similar argument can be made for daily observations of snow depth, which justifies the usage of a GEV distribution to model the annual maxima.

4.2 Kriging

From the observed annual block maxima, it is possible to estimate a single GEV distribution to each of the observed locations. There are now two ways of obtaining a continuous surface of return levels. The first way is to interpolate return levels directly from each station while the second way is to interpolate the estimated parameters from each station and then for each new point calculate the return level explicitly from the interpolated parameters. The former is slightly simpler as it only involves one interpolation step and the Kriging variance can be used as an informal uncertainty estimate. The latter however has the capability to integrate

more information from each station as it is interpolating 3 values instead of one, each with their own covariance function and trend. However this also means that each parameter has its own Kriging variance and combining them together to form a singular estimation of uncertainty is not clear. For this another approach is needed, such as the Monte Carlo method in section 2.8.

In section 2.4, two different methods were described to estimate the mean structure and variogram of the linear spatial model prior to the kriging interpolation. Both of these methods will be used and compared.

4.3 Choice of Covariates

All prediction models used in this thesis use covariates. Both Universal Kriging and non-stationary GEV distribution includes covariates in the model. The choice of covariates depends on application and available data. The covariate values must also be known for all prediction locations. In any type of linear regression model, it is possible to include more than one transformation of the same covariate into the model. In practice, this is often utilised by including different polynomials of the covariate.

All covariate values are henceforth assumed to be normalised. Normalisation of covariates makes it possible to compare the influence of different covariates, by comparing the corresponding estimated linear coefficients. The normalisation transformation for given covariate with values $(x_i)_{i=1}^n$ is given by $x'_i = (x_i - x_{min}) / (x_{max} - x_{min})$, where n is the number of covariate values, x_{min} and x_{max} is the minimum and maximum of the covariate values, respectively, and x'_i is the new transformed value. This normalises all covariate values to be in the interval $[0, 1]$.

In geostatistics, the coordinates of the geographical position are often included in the set of covariates, which also is used in our prediction models. The coordinates are available from SMHI's data, which to every observation location has a corresponding set of coordinates in the WGS84 geodetic system. Altitude is also used as a covariate and the data originates from the EU-DEM 1.1 [Copernicus, 2010] data set which includes altitude data over Europe. Finally, the mean snow depth is also used as a covariate. The mean snow depth is determined by calculating the mean of the snow depth observations in the winter months. The mean snow depth values at prediction stations are then determined by inverse distance weighting interpolation.

4.4 Estimators of linear coefficients

To determine the best model for each parameter, polynomial transformations of the covariates is used and studied in the spatial regression model. The best choice of polynomial degrees for each covariate needs to be found with a systematic approach. The approach we used for this purpose, is by utilising cross validation with a RMSE cost function, to compare different choices of polynomial covariate transformations. More specifically, all combinations of the highest polynomial degree for each covariate, with a maximum of degree 2, was tested and compared with cross validation. This means that 64 (4^3) covariate functions were compared. This was done for all

3 parameters of the GEV distribution with Universal Kriging as the predictor. The Matérn covariance function is used along with MLE to estimate its parameters as well as the parameters for the external trend in Universal Kriging. Observe however that the model for the shape parameter is not actually used in any final models due to section 3.4, it's merely included to check any spatial structure that might exist. The optimal trends for each GEV parameter using this procedure is displayed in table 4.1.

Parameter	Trend
Location	elevation + mean + (mean) ²
Shape	longitude + (longitude) ² + elevation + mean + (mean) ²
Scale	latitude + (latitude) ² + mean + (mean) ²

Table 4.1: The optimal trends for each parameter respectively

Parameter	Intercept	Longitude	Latitude	Elevation	Mean
Location	0.0887	-	-	-0.0510	1.3820
Scale	0.0875	-0.2419	0	-0.0904	0.5670
Shape	0.1960	-	0.6283	-	-1.7355

Table 4.2: Coefficients for intercept and 1st degree covariates in the external trend

Parameter	Longitude ²	Latitude ²	Elevation ²	Mean ²
Location	-	-	-	-0.1385
Scale	0.2543	-	-	-0.2953
Shape	-	-0.9087	-	1.4371

Table 4.3: Coefficients for 2nd degree covariates in the external trend

It is interesting to note that shape and scale seem to be somewhat orthogonal to each other, with shape relying on longitude and scale using latitude. The covariates for each variable are displayed in tables 4.2 and 4.3 , each variable has been normalised so the covariates can be reliably compared to each other. For all three trends, the mean is the one with the biggest impact and is found to be the biggest indicator of extreme snow fall. This is to no surprise as locations with a higher mean have more snowfall hence a larger probability of having a greater max depth.

The plots in figure 4.1 show the interpolated parameters using the trends from table 4.1 and Universal Kriging. The shape and scale parameters vary somewhat similarly, as can also be seen in their trends. This could be due to the fact that both have a large effect on the magnitude of the GEV distribution and therefore work in tandem as the amount of snow increases.

The shape parameter on the other hand seems to be trying to utilise the distance from the coast and divides Sweden into two sections as seen in figure 4.1b. The south which has a shape parameter close to zero and surrounded by coastline, while the north has a more negative value and only a small bit of coastline to the east. It

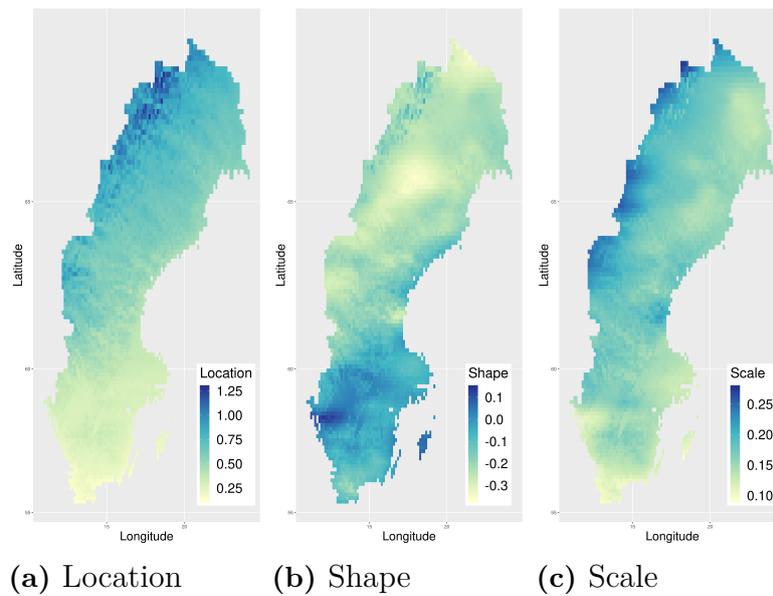


Figure 4.1: Kriging interpolation of GEV parameters

is known that proximity to large bodies of water, greatly affect the weather pattern. As stated, the south of Sweden is surrounded by coast, but also has two large lakes which behave similarly to the ocean. Additionally other phenomena such as snow-cannons on the east coast of Sweden can also explain the two divisions of the shape parameter. More meteorological analysis is needed however, to accurately determine whether such distinct snowfall patterns are actually exhibited.

Using the interpolated parameters a new surface containing the 50 year return levels can be calculated. This is displayed in figure 4.2. When comparing to the elevation maps in figure 3.5, the return levels appear to increase with elevation.

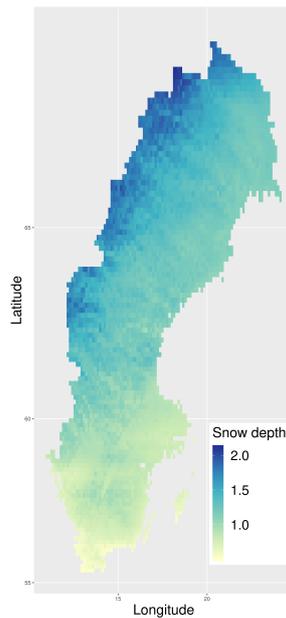


Figure 4.2: 50 year return levels of snow depth (in metres) using Kriging interpolated parameters

4.5 Non-stationary GEV

Instead of doing point-wise estimation of the GEV parameters, one can instead employ MLE-based optimisation together with the non-stationary GEV model from section 2.2 to construct a continuous, spatial GEV that can be used to estimate the parameters at any location. The ML-function can be found in equation (2.10), where x is the vector of observed block maxima. Notice that such a ML-function is only a product of marginal distributions and consequently, spatial dependency is ignored between samples. This approximation is a special case of composite likelihood and has been showed to give satisfying results for many applications [Blanchet and Lehning, 2010].

The trends that were found to be the most optimal for Kriging in table 4.1 are used as the trends in the GEV model as well. One concession is made, which is that MLE does not work very well when optimising a non-stationary shape parameter and is often not recommended [Coles and Dixon, 1999]. Instead one generally uses either a constant shape parameter or simply the Gumbel distribution with the shape parameter set to 0.

Parameter	Intercept	Longitude	Latitude	Elevation	Mean
Location	0.1128	-	-	-0.0269	2.5399
Scale	0.0855	-0.0904	-	0.0953	-0.0160
Shape	-0.0552	-	-	-	-

Table 4.4: Coefficients for intercept and 1st degree covariates in the non-stationary GEV model

4. Method

Parameter	Longitude ²	Latitude ²	Elevation ²	Mean ²
Location	-	-	-	-0.9777
Scale	0.0127	-	-	0.1922
Shape	-	-	-	-

Table 4.5: Coefficients for 2nd degree covariates in the non-stationary GEV model

The interpolation is conducted by constructing a linear model with the trends from table 4.1 and the coefficients from tables 4.4 and 4.5. This is in contrast to Universal Kriging which also incorporates spatial information when interpolating.

The coefficients are estimated using MLE however with a constant trend for shape due to the problems when using a non-stationary trend as mentioned before. When compared to the coefficients that were found when using MLE but for Kriging in tables 4.2 and 4.3, there is not a huge difference between them, and all the parameters have the same sign and similar magnitude.

The density plots of the non-stationary distributions in the case of both Gumbel and GEV are shown in figure 4.3. The empirical density plots are derived by transforming the non-stationary data to the standard Gumbel distribution. This transformation is determined by the estimated parameters of the GEV distribution, which are location specific. In relation to each other they are similar and mainly because the estimated shape parameter from the GEV is estimated to -0.0552 from table 4.4 which mostly only affects the tail behaviour of the distribution.

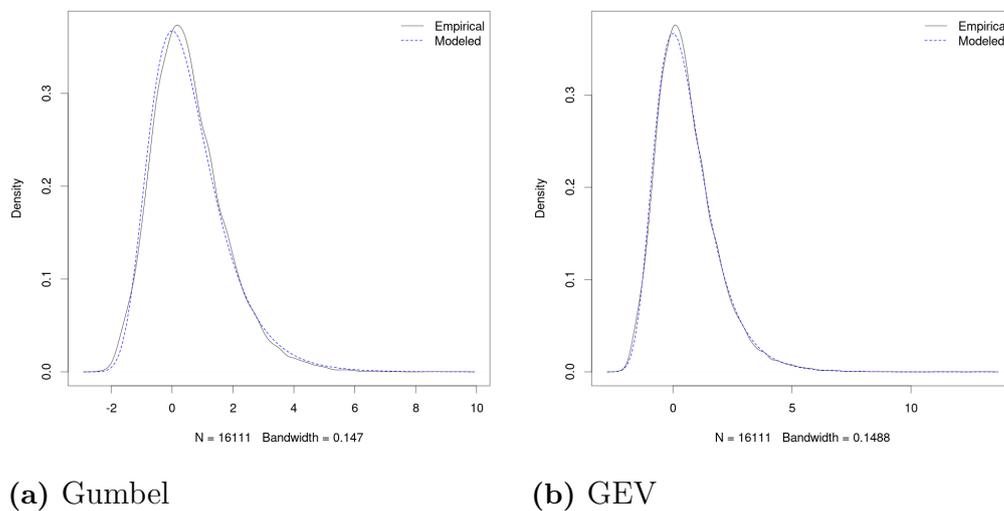


Figure 4.3: Density plots of non-stationary Gumbel and GEV distributions respectively.

The return level plots for both a Gumbel and GEV distribution are shown in figure 4.4 which again are pretty much the same. This supports our initial assumption that a Gumbel distribution should be used throughout Sweden.

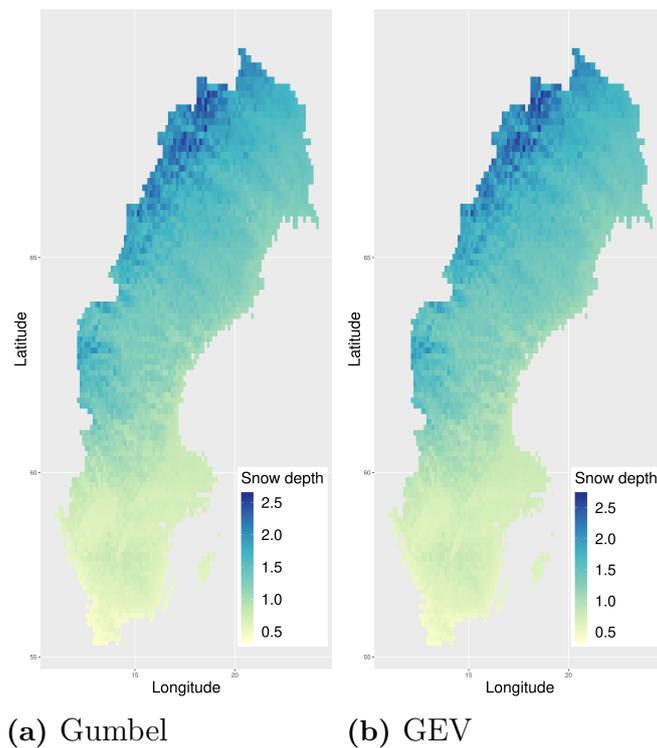


Figure 4.4: 50 year return levels of snow depth (in metres) using linear models with coefficient estimates from MLE of a Gumbel and GEV model respectively.

4.6 Simulations

Because the maps may be used for real-world applications when constructing buildings, it is important that an evaluation of the related uncertainty is also included. The Kriging variance can be used as a measure of uncertainty, however it only accounts for the spatial uncertainty in the model. One possible solution is to include any uncertainty as a nugget effect as mentioned in 2.32. This may still not account for all the error in the system as there are many points of entry where error is introduced into the system. A more comprehensive solution to allow the propagation of error through the model is the Monte Carlo method from section 2.8, specifically the Conditional Gaussian Field simulation by LU-decomposition technique in section 2.8.3 is used.

The main reason that we can not simulate return levels directly by creating a Gaussian Random Field (GRF) of them to simulate from, is because we can not assume that the return levels are Gaussian distributed. However due to the MLE method, the parameter estimates are in fact assumed to follow a Gaussian distribution [Coles, 2001]. Additionally more information might be retained by calculating return levels from Monte Carlo simulations of both location and scale, which individually have different spatial structures, instead of just simulating return levels directly.

So in our case, we wish to simulate the parameters, location and scale individually. Due to the use of MLE, under some regularity conditions, the approximating distribution of the parameters will in fact be multivariate normal such that $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \mathcal{I}^{-1}(\boldsymbol{\theta}))$, where $\mathcal{I}^{-1}(\boldsymbol{\theta})$ is the inverse of the Fisher information matrix

evaluated at θ and θ are the parameters with $\hat{\theta}$ being the parameter estimates. It also follows, under similar regularity conditions, that each parameter is univariate normal as $\hat{\theta}_i \sim N(\theta_i, \phi_{i,i})$ where $\phi_{i,i}$ is the corresponding diagonal element from $\mathcal{I}^{-1}(\theta)$.

Using this we can construct a GRF for each parameter, by using the parameter estimate as the mean and an appropriate covariance matrix between all the points we want to simulate. The covariance matrix is constructed by using the covariance function and the estimated parameters for it, from Kriging interpolation in section 4.4. However an extra modification is done to the algorithm in section 2.8.3, which is to incorporate the standard error estimate $\hat{\phi}_{i,i}$ that was obtained at every observed location during the MLE procedure when estimating the GEV parameters in section 3.4, as a nugget effect on the diagonal of \mathbf{C}_{11} . This is to better represent the actual distribution of the parameter estimates at already observed locations.

One thousand realisations are created for each parameter and after the simulations are complete, 50 year return levels are calculated for each realisation of the parameters. From the simulated return levels we can also calculate the 95% confidence intervals to get an idea of the uncertainty in the maps.

The maps in figure 4.5 show some example realisations from the Monte Carlo simulations. It can be observed, when compared later with the summary statistics in figure 4.6, that the variability is in fact greater in areas with greater uncertainty.

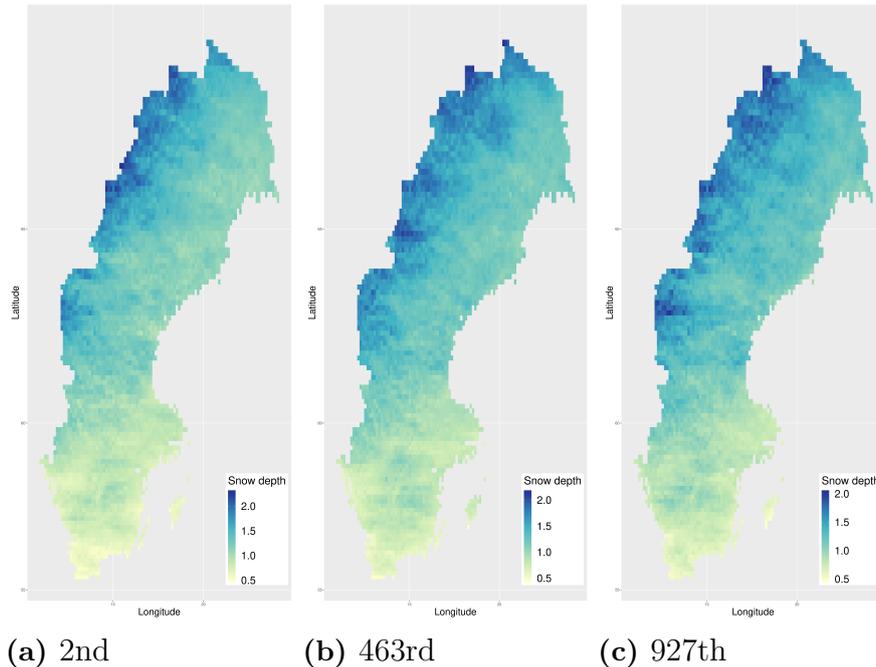


Figure 4.5: Example realisations from the MC simulations of the 50 year return levels of snow depth in metres.

In figure 4.6a is the mean of all simulations, and we see is the same as the Kriging prediction in figure 4.2 which is what was expected from the theory. The variance of the simulations was also confirmed to be the same as the Kriging variance.

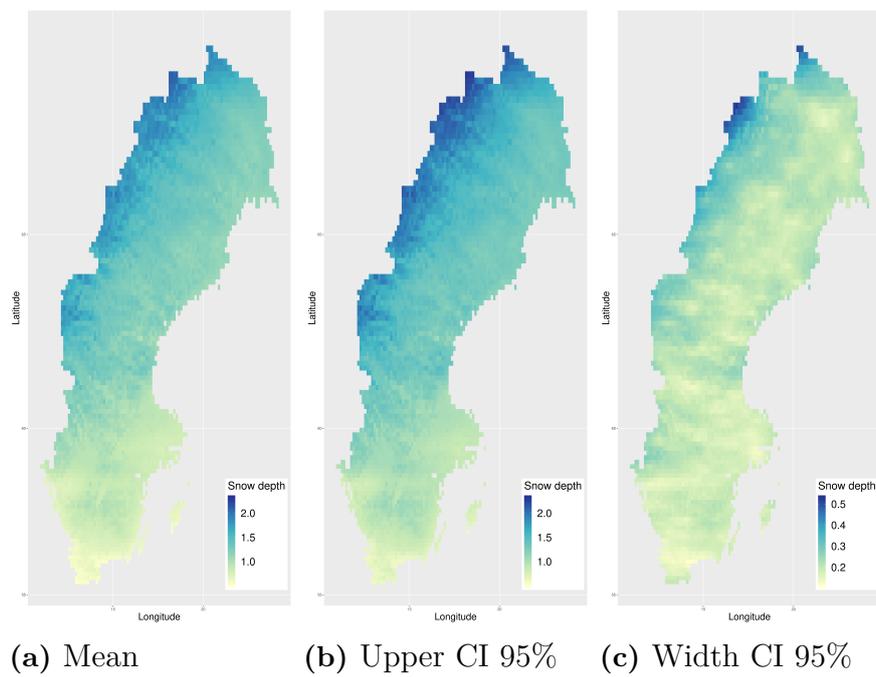


Figure 4.6: Summary of 1000 Monte Carlo simulations of the 50 year return levels of snow depth in metres. From left to right: The mean of all simulations, the upper limit of the 95% confidence interval, and the width of the 95% confidence interval.

5

Results

A table of the error metric values for the model comparisons with leave one out cross validation can be found in table 5.1. It is clear from this table that Kriging with the MLE approach to estimate the variogram performs best. For this reason, all Kriging interpolation maps henceforth is assumed to be interpolated with this method. Using the Kriging prediction from section 4.2 and the upper limit of the 95% CI from section 4.6, we can calculate the snow load from the snow depth maps. Boverket has identified different densities for snow in different regions of Sweden, and are published in a table as part of the Swedish construction regulation. This means that the snow depth maps can not be trivially used to deduce areas of heavy snow load [Olsson, 2011]. Instead the snow load maps must be calculated by applying the different densities for each region and calculating the down force produced (in kN/m^2). The final results are produced in the maps in figure 5.1. These maps can be directly compared to the map that Boverket has given and there are a number of differences between them.

	RMSE	MAE	MPE	BIAS
Non-stationary GEV	0.11	0.09	0.29	0.02
Kriging	0.10	0.09	0.20	0.01
Kriging (MLE)	0.05	0.04	0.15	0.00

Table 5.1: Table of the error metric values defined in (2.37) for non-stationary GEV model, Kriging without MLE and Kriging with MLE, respectively. Leave one out cross validation for all stations was used to calculate the values.

To better analyse the discrepancy between the maps, Boverket has also published the recommended snow load that a building should be able to withstand, for each municipality in Sweden. We can now check if the values given by Boverket over or underestimate potential extreme snow loads and assuming that buildings have been built according to these values, how great the risk is for collapse. The maps which show the discrepancies are displayed in figure 5.2.

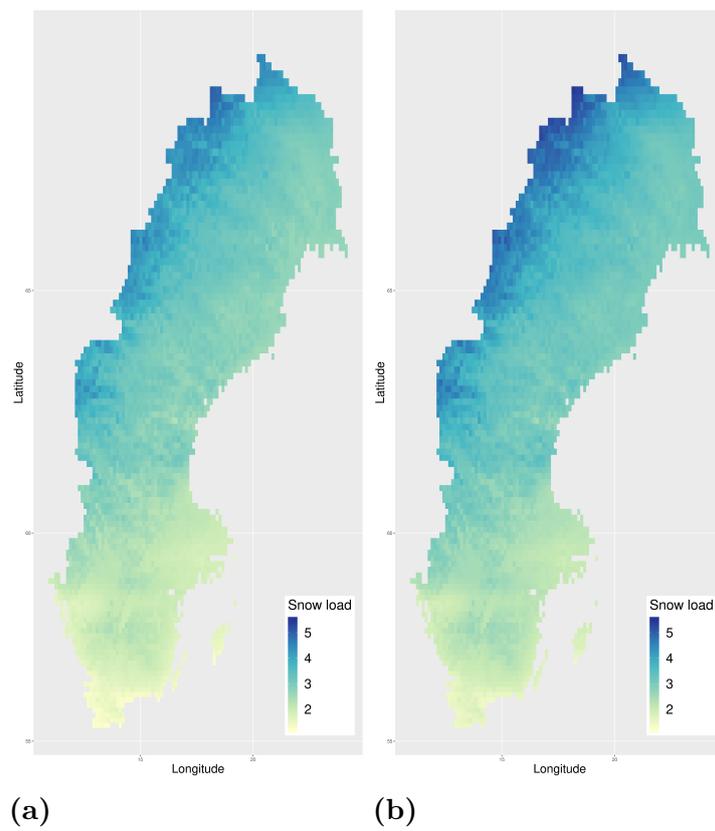


Figure 5.1: 50 year return levels of snow load (in kN/m^2). Figure (a) is with the Kriging prediction and figure (b) is with the upper limit 95% of the confidence interval from the MC simulations.

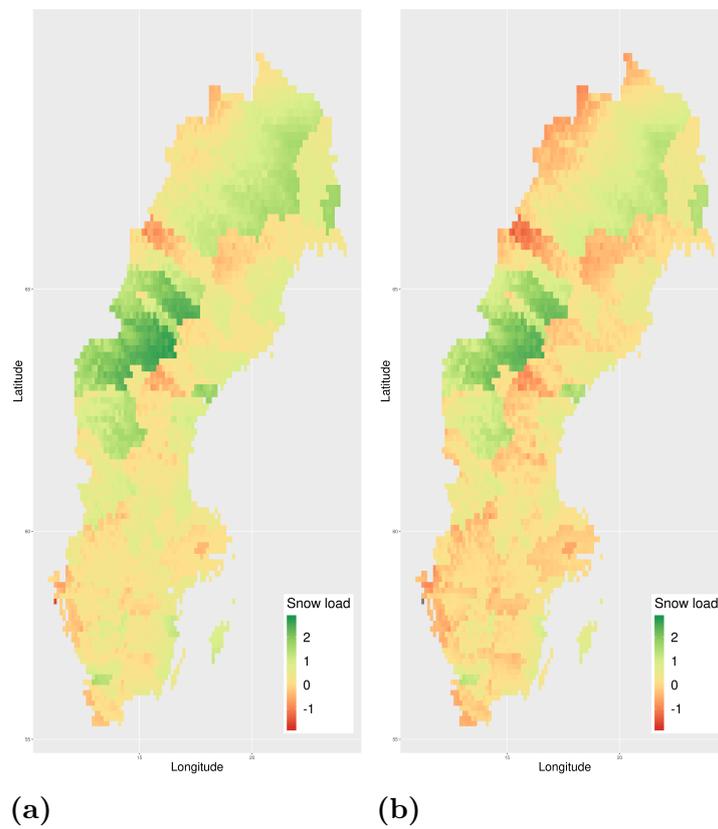


Figure 5.2: Discrepancy between estimated 50 year return levels of snow loads and Boverket's recommended snow loads (in kN/m^2) that a building should be able to withstand. Left is with the Kriging prediction and right is with the upper limit 95% of the confidence interval from the MC simulations.

6

Conclusion

The maps in figure 5.2 show that for a large portion Sweden, the map by Boverket is generally underestimating extreme snow load. This is of concern as it's also happening in areas that are densely populated in the south of Sweden where there are many more buildings. This thesis does not attempt to identify what the cause of this discrepancy is as it could be from a number of reasons, many of them not related to the mathematical modelling such as the density of snow not being accurate. Luckily in areas of very high snow load, such as around Åre, the map by Boverket actually overestimates predicted extreme values. Therefore buildings in these areas, should be well future-proofed and be well within safety margins to handle extreme load. It is obvious from table 5.1 that interpolating by kriging is to be preferred over interpolating by non-stationary GEV. It is hard to conclude why this is the case. It is possible that the real parameters of the GEV distribution varies to much over the spatial region. This means that the low polynomial covariate trends can not be fitted to the real parameter values. In contrast, interpolating by kriging is more flexible, since the residual is modeled by a gaussian random process. The reversed result was found in Blanchet and Lehning [2010], which compared different interpolating methods of snow depths in Switzerland. They showed that interpolating by non-stationary GEV was to be preferred over interpolating the individual parameters by kriging. It is possibly that the contrary results can simply be explained by that the spatial regions of study differs. However, all nugget effects were excluded from their spatial regression models, which can also explain why the kriging interpolation underperformed.

One can argue if the methods used for comparing the two interpolation methods is fair, since the prediction depends on the choice of covariates. It is plausible that the reversed result would hold for some other choice of covariate trends. However, the comparison were executed for many different choices of trends which all resulted that kriging were indisputably to be preferred.

A disadvantage with the methods used in this project for interpolating univariate GEV distributions is that any spatial multivariate analysis is not possible. Consequently, the probability that the annual snow depth exceeds some given value, for two or more locations can not be determined. Determining such probabilities is of importance for applications such as land-use planning. In contrast, the usage of max-stable processes for modelling spatial extremes has the advantage that spatial multivariate analysis is possible. The univariate marginal distribution for any locations of a max-stable process has the max-stable property, i.e., it follows a GEV distribution. The study of max-stable processes is a active topic of research in spatial statistics.

Bibliography

- J. Blanchet and M. Lehning. Mapping snow depth return levels: smooth spatial modeling versus station interpolation. *Hydrology and Earth System Sciences*, 14 (12):2527–2544, 2010.
- Stuart Coles. *Classical Extreme Value Theory and Models*. Springer London, London, 2001.
- Stuart G. Coles and Mark J. Dixon. Likelihood-based inference for extreme value models. *Extremes*, 2(1):5–23, Mar 1999.
- Copernicus. European digital elevation model (eu-dem), version 1.1. <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=download>, 2010. Online; accessed 28th May 2019.
- Michael W. Davis. Production of conditional simulations via the lu triangular decomposition of the covariance matrix. *Mathematical Geology*, 19(2):91–98, 1987.
- Salaheddine El Adlouni, Taha Ouarda, X Zhang, R Roy, and Bernard Bobée. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43, 03 2007.
- Xavier Emery. Conditioning simulations of gaussian random fields by ordinary kriging. *Mathematical Geology*, 39(6):607–623, Aug 2007.
- M. R. Leadbetter. Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(2):291–306, Dec 1983.
- Catarina Olsson. Boverkets föreskrifter och allmänna råd (2011:10) om tillämpning av europeiska konstruktionsstandarder (eurokoder). Technical report, Boverket, 2011.
- SMHI. Svenska snödjupsrekord. <http://www.smhi.se/kunskapsbanken/meteorologi/svenska-snodjupsrekord-1.88638>, 2016. Online; accessed 28th May 2019.
- Richard L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.