Doctoral theses at NTNU, 2023:379

Silius Mortensønn Vandeskog

Efficient high-dimensional modelling of temperature and extreme precipitation

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of Science and Technology

Silius Mortensønn Vandeskog

Efficient high-dimensional modelling of temperature and extreme precipitation

Thesis for the Degree of Philosophiae Doctor

Trondheim, December 2023

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

NTNU

Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-326-7458-9 (printed ver.)
ISBN 978-82-326-7457-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:379

Printed by NTNU Grafisk senter

Preface

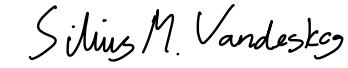
You are now reading my Philosophiae Doctor (PhD) thesis in statistics, which concludes my last four years of work at the Department of Mathematical Sciences of the Norwegian University of Science and Technology (NTNU).

In writing this thesis, I have relied on the help and support of many others, whom I would like to thank. First of all I would like to thank my supervisor, Sara Martino, for all the guidance she has given me over these last four years. She has always had an open door for me and has taken the time needed to help me with my problems whenever it was necessary. She has also urged me to participate in international workshops and conferences, and to go on a longer research stay in another country, which have resulted in many great experiences for me. I would also like to thank my co-supervisors, Ingelin Steinsland and Oddbjørn Bruland, who have always been ready and willing to help me. Ingelin was also my supervisor during my master's thesis, and she encouraged me to apply for this PhD position. I would also like to thank my other master's supervisor, Thordis Thorarinsdottir, who initially lead me on to this path by asking if I wanted to write my master's thesis for her on modelling temperature. I would like to thank Raphaël Huser for inviting me to a three month stay at KAUST and for taking good care of me while I was there. I learned much during my stay, and I got to know a lot of nice people there.

I have also had the pleasure of writing and publishing papers together with many brilliant co-authors, without whom there would probably be no papers and no thesis. In no particular order, these are Thordis Thorarinsdottir, Ingelin Steinsland, Finn Lindgren, Sara Martino, Daniela Castro-Camilo, Håvard Rue and Raphaël Huser. I am grateful for all the help they have given me, and for the good discussions I have had with them.

Finally, I would like to thank my friends and family for all of their support. I would especially like to thank Camilla for always being there for me and supporting me whenever I was feeling down, stressing about presentations and overhanging deadlines, ranting about all of my annoying problems, or being denied entry at German hotels due to confusing and quickly changing COVID rules. Thank you for being you.

In the event that I have forgotten anyone, I thank them both for their help, and for forgiving me of my forgetfulness.



Silius Mortensønn Vandeskog
Oslo, August 2023

Contents

Part I

Background

1 Introduction

The rapid climate changes that have occurred in recent decades are leading to an increase in natural hazards (IPCC, 2023), with flooding being among the most common and costly of these (Razavi et al., 2020). As a very relevant example, while I have been putting the finishing touches on this thesis, the storm "Hans" has ravaged the eastern and southern parts of Norway, leading to large and dangerous floodings over considerable parts of the country. The floods have lead to the evacuation of thousands of people, a partial collapse of a river dam, a closing of the main roads between the south and north of Norway, and overall damages that are projected to cost well over a billion Norwegian kroners to repair. Still, floods are projected to become even more frequent and damaging as climate change continues (Yin et al., 2018; Allan et al., 2020). Improved flood mitigation measures might therefore help in avoiding human casualties and devastating economic losses to our society (Jongman, 2018). However, in order to improve upon the already existing flood mitigation tools, a better understanding of the climatic processes behind the occurring floods might be necessary (Brunner et al., 2021). The overall aim of this thesis is therefore to gain a better understanding of these climatic processes, which might later help in the development of improved flood mitigation tools.

The two main drivers of river flooding are precipitation and snowmelt (Hanssen-Bauer et al., 2015; Berghuijs et al., 2019). However, snowmelt is mainly driven by temperature and precipitation. Furthermore, the amount of precipitation is directly affected by the temperature, as higher temperatures allow the air to hold more moisture, which makes it possible for more precipitation to fall over an area within a short period of time (e.g., Westra et al., 2014). We can therefore say that the two main atmospheric drivers behind river floods are precipitation and temperature. An improved understanding of these two variables can thus lead to an improved understanding of the dangers and possibilities of future floods.

With this as background, the main focus of my thesis has been to develop and improve upon statistical models for temperature and extreme precipitation, in order to provide a better understanding of the underlying processes that can cause floods. I have not attempted to project how temperature and extreme precipitation might be affected by future climate changes, but rather to describe how they behave in the present climate. However, this can also be useful for assessing future climate changes, as a better understanding of today's climate can lead to improvements in the evaluation, development and statistical downscaling of climate models used for projecting future climate change (Maraun & Widmann, 2018). The thesis consists of four papers (see Section 6 for a complete overview). The first paper covers statistical modelling of temperature, while the remaining

three focus on statistical modelling of extreme precipitation.

In Paper 1, we develop a model for diurnal temperature range, which is the difference between daily maximum and daily minimum temperature. Hydrological impact models often require both daily minimum and maximum temperature as input, as their values are important for accurately describing processes like the formation and melting of snow and ice. However, due to computational constraints, many interpolated data sets, statistical downscaling methods and bias correction methods treat daily minimum and maximum temperature as independent variables. This is problematic, as it can result in highly unrealistic estimates of the temperature variability within a given day. It can even lead to crossing behaviour, i.e., estimated daily minimum temperatures that are larger than their corresponding estimated daily maximum temperatures. The Nordic Gridded Climate Data Set Version 2 (Lussana, Saloranta, et al., 2018; Lussana, Tveito, & Uboldi, 2018) is an example of such an interpolated data set where the variables are treated independently, and where temperature crossing occurs for a considerable amount of the data. The direct modelling of diurnal temperature range solves this problem by ensuring that no min/max crossing can occur.

Much work has been put into modelling daily minimum, maximum and mean temperatures, but to the best of our knowledge, a statistical model for the marginal distribution of diurnal temperature range has never previously been proposed in the literature. The first step of our modelling is therefore to propose a marginal distribution for diurnal temperature range. We choose the somewhat obscure five-parameter lambda distribution (Gilchrist, 2000) as our marginal model. To the best of our knowledge, this distribution has never previously been applied for modelling a climate variable. In the paper, we argue for why this distribution is an appropriate choice for modelling differences between extreme quantiles, such as the daily maximum and minimum temperature. We also propose a distributional regression framework for estimating the marginal distribution of diurnal temperature range as a function of explanatory variables. The two proposed models are applied for modelling diurnal temperature range in a case study using data from a set of weather stations in the south of Norway. We find that the five-parameter lambda distribution is a flexible model that provides a good fit to the complex and somewhat unusual marginal distributions of diurnal temperature range. The distributional regression model allows for interpolation of the marginal distributions of diurnal temperature range, and it is found to perform well in a spatial cross-validation study.

The remaining papers focus on statistical modelling of extreme precipitation. This is a complex area of research that contains a large variety of interesting aspects and problems. We have concentrated on providing new insight into two of these problems. The first problem is: what is a realistic estimate for a precipitation value that is so extreme that the probability of

exceeding it within a year is $1/n$, for any integer $n > 1$? This precipitation amount is known as the n-year return level, as it on average will be exceeded once every n years. The concept of return levels is heavily used when designing and building infrastructure, and the Norwegian Energy Regulatory Authority (NVE) therefore produces maps of precipitation return level estimates over all of Norway. Additionally, all dams in Norway are required by law to withstand large precipitation amounts such as the 500-year or the 1000-year return level. The second problem we provide new insight into is: if we today were to observe extreme precipitation somewhere in space, what would the spatial distribution of that precipitation event be? Answers to this problem can tell us how the distribution of extreme precipitation over a water catchment might look like, and thus how focused a rainstorm might be or how much of the precipitation that might end up in the same parts of a river at the same time. Paper 2 provides insight to the first problem, while Paper 3 and 4 provide insight to the second problem.

A reasonable answer to the first problem can often be produced using only properties of the marginal distributions of observed precipitation, while a reasonable answer to the second problem relies heavily on the dependence structure between precipitation observations at multiple locations in space. This means that more complex statistical models become necessary.

To provide insight into the two mentioned problems, we lean heavily upon the field of extreme value theory, which provides a nice mathematical description of how the tails of a probability distribution behave as we approach their limits. Extreme value theory has shown great success in modelling environmental extremes, and it has been receiving increased amounts of interest in recent years, in part due to the impending climate crisis (e.g., Opitz et al., 2018; Shooter et al., 2019; Reich & Shaby, 2019; Castro-Camilo et al., 2019; Bacro et al., 2020; Bopp et al., 2020; Simpson & Wadsworth, 2021; Koch et al., 2021; Richards et al., 2022; Koh et al., 2023).

Even though extreme value theory is a relatively old field, with roots going back to the work of Fisher and Tippett from 1928 (Fisher & Tippett, 1928), the field is still evolving, and there are many important problems left to be solved. In my opinion, an important remaining problem is that of developing extreme value models that have both nice theoretical properties and that are highly usable in practice. Until recent years, the field has mostly focused on models with nice theoretical properties, while less effort has been put into models that work well in practice, in the sense that they are flexible enough to describe real-life problems of interest while also allowing for robust and computationally efficient inference. Particularly when modelling spatial or spatio-temporal extremes, most of the classical extreme value models focus on explaining "the asymptotics", which might often be so far into the distributional tails that it has little practical relevance for modelling climate data with somewhere between 10 and 100 years of

available observations. At the same time, the models fail to provide accurate descriptions of the "subasymptotic" extreme events, which might often be the ones we actually care about and have relevant data for (Huser & Wadsworth, 2022). Inference for most classical models from spatial extreme value theory is also computationally demanding, and scales poorly with an increase in model dimension, which makes the models intractable for describing complex and high-dimensional spatial data sets.

In the three papers about extreme precipitation, we have therefore attempted to provide new insight into two more theoretical/methodological problems, in addition to providing insight into the behaviour of extreme precipitation. The first problem concerns how to develop models for (univariate and spatial) extremes that are flexible enough to provide accurate descriptions of both subasymptotic and asymptotic extreme behaviour, while the second problem concerns how to make modelling assumptions that are as true to the extreme value theory as possible, while also allowing for more robust and more computationally efficient inference.

The remainder of Part I of this thesis is organised as follows. In Section 2, an introduction to the field of extreme value theory is provided, briefly covering the topics of univariate, multivariate and spatial extreme value theory, extremal dependence structures, and asymptotic and subasymptotic models for spatial extremes. Then, in Section 3, a short summary of spatial statistics, with a focus on computationally efficient modelling, is presented. This briefly covers the topics of Gaussian random fields, spatial covariance functions, independence assumptions leading to covariance tapering and low-rank estimation methods, and finally Gaussian Markov random fields and the so-called stochastic partial differential equation (SPDE) approximation. Section 4 presents the framework of latent Gaussian models, and how these models allow for both flexible modelling and computationally efficient inference using the method of integrated nested Laplace approximations (INLA). In Section 5, I present some final thoughts about the thesis, and about possible future extensions of our work. Finally, in Section 6, a short summary of each of the four papers of the thesis are presented.

2 Extreme value theory

2.1 Univariate extreme value theory

Extreme value theory is a field within probability theory and statistics that focuses on estimating properties of the tails of probability distributions (see, e.g., Coles (2001) or Beirlant et al. (2004) for a thorough introduction). The field originates from the problem of finding the distribution of the maximum of a large number of random variables with unknown distribution functions. In their seminal work, Fisher and Tippett (1928) show that,

for the maximum M_n, of n random variables $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$, with unknown distribution function F, if there exists two sequences of standardising constants $\{a_n > 0\}$ and $\{b_n\}$ such that the limiting distribution of the standardised maximum $M_n^* = (M_n - b_n)/a_n$ is non-degenerate as $n \to \infty$, then the limiting distribution of M_n^* must have one of three possible forms, which can be merged together to yield the generalised extreme value (GEV) distribution

$$\lim_{n\to\infty} \mathrm{P}(M_n^* \leqslant x) = G(x) = \begin{cases} \exp\left[-\{1+\xi(x-\mu)/\sigma\}_+^{-1/\xi}\right], & \xi \neq 0, \\ \exp\left[-\exp\left\{-(x-\mu)/\sigma\right\}\right], & \xi = 0, \end{cases} \tag{1}$$

where $a_+ = \max(a, 0)$, μ is a location parameter, $\sigma > 0$ is a scale parameter and ξ is best known as the shape or the tail parameter. If $\xi > 0$, the GEV distribution has a heavy upper tail and is known as a Fréchet distribution. If $\xi = 0$, the GEV distribution has an exponential upper tail and is known as a Gumbel distribution. If $\xi < 0$, the GEV distribution has a finite upper tail and is known as the reverse Weibull distribution. The Gumbel distribution has support on the entire real line, whereas the supports of the Fréchet and reverse Weibull distributions are bounded from below and above, respectively. Figure 1 displays probability density functions of these three sub-classes of the GEV distribution. The result in (1) can also be shown to hold if the n random variables are dependent and come from a stationary time series, although with slightly different location and scale parameters (e.g., Coles, 2001).

GEV distributions are max-stable distributions, meaning that for every natural number $n \in \mathbb{N}$, there exists constants $\alpha_n > 0$ and β_n such that

$$G^n(\alpha_n x + \beta_n) = G(x).$$

Furthermore, any random variable that is max-stable must follow the GEV distribution.

Knowledge about the limiting distribution G of the standardised maximum can be further used for learning about other properties of the upper tail of the unknown distribution F. One can, e.g., show that, if (1) holds for the standardised maximum over n independent realisations of $X \sim F$, then, for a large enough threshold u, the distribution of a threshold exceedance $[X - u \mid X > u]$ can be approximated well by the generalised Pareto (GP) distribution (Pickands III, 1975; Davison & Smith, 1990),

$$\mathrm{P}(X - u \leqslant x \mid X > u) \approx \begin{cases} 1 - (1+\xi x/\sigma_u)^{-1/\xi}, & \xi > 0, x \geqslant 0 \\ 1 - (1+\xi x/\sigma_u)^{-1/\xi}, & \xi < 0, 0 \leqslant x \leqslant -\sigma_u/\xi \\ 1 - \exp\left(-x/\sigma_u\right), & \xi = 0, x \geqslant 0 \end{cases}, \tag{2}$$

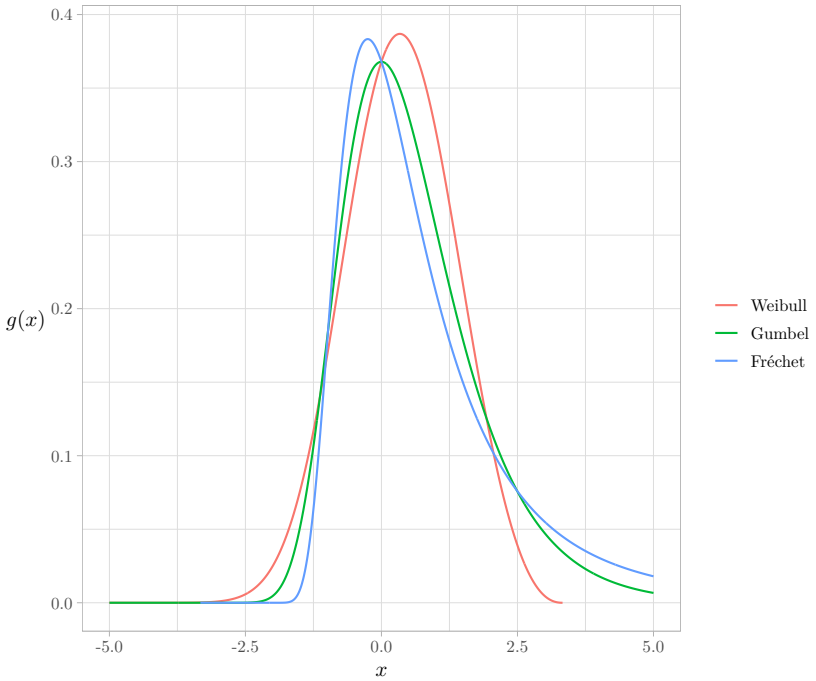


Figure 1: *Probability density function, $g(x)$, for the GEV distribution with location parameter $\mu = 0$ and scale parameter $\sigma = 1$. The tail parameter ξ takes the values -0.3, 0 and 0.3 for the Weibull, Gumbel and Fréchet distributions, respectively.*

where the scale parameter σ_u is a function of the threshold u, and the tail parameter ξ is the same as in (1). This is an incredible result, which implies that any probability distribution with certain tail regularity properties can be approximated well by the GP distribution if one moves far enough into its tails, since, when $x > u$, the probability $\mathrm{P}(X \leqslant x)$ can be decomposed into

$$\mathrm{P}(X \leqslant x) = \mathrm{P}(X \leqslant u) + \mathrm{P}(X > u)P(X \leqslant x \mid X > u).$$

Together, the results in (1) and (2) make it possible to estimate many important properties of the tails of statistical distributions, which would be difficult to estimate using other statistical methods. As an example, the most common methods for estimating a quantile is to either use an empirical estimator, or to fit a probability distribution to the available data and then compute the corresponding quantile of that fitted distribution. However, if we have, e.g., n available observations, and we wish to estimate the $1 - 1/(2n)$ quantile, it is clear that an empirical estimator might perform poorly, and that a probability distribution fitted to the n observations might focus most on the bulk of the data, which can lead to poor performance in the tails. Using extreme value theory, one can fit the GP distribution

directly to only the most extreme observations, i.e., those that are larger than some threshold u, which often yields considerably better estimates for extreme properties such as the $1 - 1/(2n)$ quantile. Similarly, if one is, e.g., designing a dam that needs to withstand the n year return level of river runoff, but only $m < n$ years of data are available, extreme value theory makes it possible to provide an estimate for the n year return level by dividing the river runoff observations into large and disjoint blocks of size k, typically with k equal to one year, and then fitting the GEV distribution to the available block maxima. Then one can estimate the n year return level as the $1 - 1/n$ quantile of the fitted GEV distribution, or some other quantile if k is different from one year (Coles, 2001).

An important challenge when modelling extremes with the GEV or the GP distribution is the choice of the block size k and the threshold u (e.g. Wadsworth & Tawn, 2012a; Langousis et al., 2016; Bader et al., 2018; Bücher & Zhou, 2021). The results in (1) and (2) only hold exactly in the limit where $k \to \infty$ and where $u \to x^*$, where x^* is the upper endpoint of F. Thus, there will always be some model misspecification present for finite k or $u < x^*$, and this misspecification decreases as we increase k or u. However, the number of data points available for inference decreases when k or u increases. This means that the selection problem for k and u can be considered as a type of bias/variance tradeoff.

Another challenge is that both the GP and the GEV distributions have supports that depend on the values of their model parameters. This can cause considerable complications and lead to less robust inference, as small changes in the parameter values can cause observations to fall outside the support, thus lending the model fit unsuitable for the given data (e.g. Smith, 1985; Bücher & Segers, 2017). This lack of robustness can become particularly challenging in settings with few available observations, which is often the case since k and u must be chosen large to minimise model misspecification. It also severely complicates regression-based modelling where one or more of the parameters are modelled as a function of some explanatory variables, as this can lead to different supports for each combination of explanatory variables. In such settings, simply finding a legal set of initial variables for estimating the model parameters can be a considerable challenge. This problem is more severe for the GEV distribution, as the GP distribution only has a support that changes as a function of the parameters values for light-tailed distributions ($\xi < 0$), while the GEV support is bounded and depends on its parameter values for all $\xi \neq 0$.

For these reasons, in the setting of modelling extremes of heavy-tailed phenomena, Castro-Camilo et al. (2022) propose to model block maxima using a modified version of the GEV distribution, denoted the blended GEV (bGEV) distribution. This distribution is equal to a Gumbel distribution ($\xi = 0$) in its lower tail, and equal to a Fréchet distribution ($\xi > 0$) in its

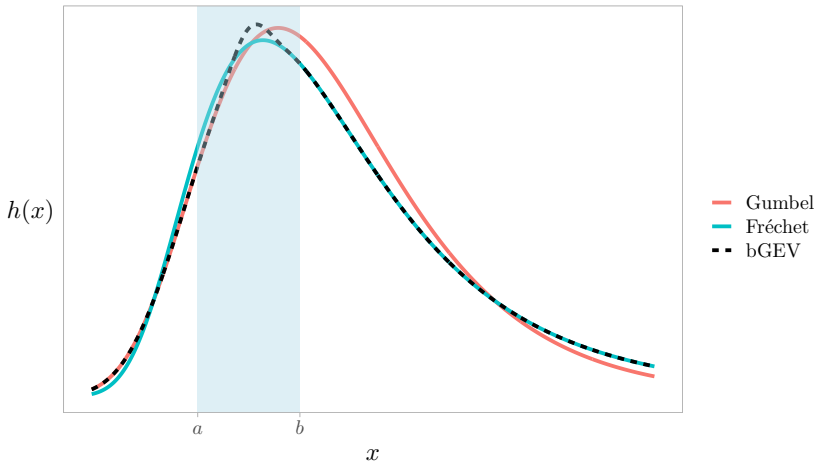


Figure 2: *The probability density function $h(x)$ of the bGEV distribution together with the Fréchet and Gumbel distributions in its upper and lower tail, respectively.*

upper tail, which means that it always has an unbounded support, and that it is able to describe both light-tailed and heavy-tailed phenomena. The cumulative distribution function (CDF) of the bGEV is defined as

$$H(x;\mu,\sigma,\xi,a,b) = F(x;\mu,\sigma,\xi)^{p(x;a,b)} G(x;\tilde{\mu},\tilde{\sigma})^{1-p(x;a,b)},$$

where F is the CDF of a Fréchet distribution, G is the CDF of a Gumbel distribution and $p(x;a,b)$ is a weight function that is zero for $x \leqslant a$, one for $x \geqslant b$, and increasing as a function of x for $a < x < b$. The parameters $(\tilde{\mu},\tilde{\sigma})$ of the Gumbel CDF are functions of (μ,σ,ξ,a,b) for ensuring that H is continuous. The probability density function of the bGEV distribution is displayed in Figure 2. This distribution provides considerably more robust inference than the GEV distribution, and it makes regression modelling easier, faster and more stable. In Paper 2 we apply the bGEV distribution for performing spatial regression modelling of the yearly maxima of sub-daily precipitation data in Southern Norway, in order to estimate large precipitation return levels. Precipitation is well known to be a heavy-tailed phenomenon (Koutsoyiannis et al., 1998; Cooley et al., 2007; van de Vyver, 2012; Papalexiou & Koutsoyiannis, 2013; Huser & Davison, 2014), which makes the bGEV distribution a perfect choice for such a task. Our paper is the first to apply the bGEV distribution for modelling in a complex large-scale case study with real data. In the paper, we also demonstrate in a simulation study how the bGEV distribution can be much more robust and perform considerably better than the GEV distribution when estimating return levels for heavy-tailed problems.

Asymptotically, the GEV distribution is known to be the only correct limiting distribution, and many might therefore argue that the bGEV

distribution is unsuitable for modelling extremes. However, in practice, one will never obtain an infinite amount of data, and the GEV distribution will therefore also always be "wrong", in the sense that the limit in (1) is never fully reached. Thus, with finite amounts of data, there is no way of knowing if the GEV distribution is more or less wrong than the bGEV distribution. The only thing we know is that, if the data are heavy-tailed, and we choose a large enough block size, then the limit in (1) is probably close enough to the truth so that both the GEV distribution and the bGEV distribution can provide good model fits. Then, the only relevant difference between the two is that the bGEV distribution is more robust and easier to perform inference with than the GEV distribution. Approximations or modifications in a similar spirit, for performing more robust and/or more computationally efficient inference, are more common in other sub-fields of statistics that have been more focused on applications. For example, in spatial and spatio-temporal modelling, it is common to model almost all parts of the dependence structures using transformations of Gaussian random fields, even though it is known that other dependence structures can be more appropriate, because the Gaussian copula has nice theoretical properties and can cause significant improvements in the inference speed (see Section 3 for more details). As the field of extreme value theory is becoming more applied, modelling assumptions such as the bGEV, which allow for faster, better and more robust modelling of the subasymptotics, by focusing slightly less on the asymptotics, are starting to become more accepted (Papastathopoulos & Tawn, 2013; Naveau et al., 2016; Stein, 2021b, 2021a; Yadav et al., 2021; Richards & Huser, 2022; Gamet & Jalbert, 2022). In my opinion, modellers should always consider choosing more robust modifications with more focus on the subasymptotics, instead of choosing the less robust, but asymptotically correct, GEV and GP distributions.

2.2 Multivariate extreme value theory

Many of the most impactful extreme events in the real world are multivariate, spatial, temporal or spatio-temporal in nature. For example, extreme negative stock results for a single company is problematic, but simultaneous extreme negative stock results for an entire financial sector can cause severe damage to human society. Extreme levels of one specific air pollutant can be dangerous, but simultaneous extreme levels of multiple different air pollutants might lead to a cocktail-effect that is much more damaging. Extreme heat can be deadly, but extreme heat for a long time and over an entire continent can cause serious damage to the food-safety of millions of people.

There exists a multivariate version of the generalised extreme value theorem (e.g. Coles, 2001). To describe the multivariate version, we require n

realisations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of a d-dimensional random vector $\boldsymbol{X} \sim F$, with unknown distribution function F, and we denote $\boldsymbol{X}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,d})^\top$ for all $i = 1, 2, \ldots, n$. Then, we define the vector of component-wise maxima as $\boldsymbol{M}_n = (M_{n,1}, M_{n,2}, \ldots, M_{n,d})^\top$, where $M_{n,j}$ is the maximum over the jth component of the n random vectors, i.e., $M_{n,j} = \max(X_{1,j}, X_{2,j}, \ldots, X_{n,j})$. The multivariate generalised extreme value theorem now states that, if there exists two standardising sequences of d-dimensional vectors $\{\boldsymbol{a}_n > \boldsymbol{0}\}$ and $\{\boldsymbol{b}_n\}$, with $\boldsymbol{a}_n = (a_{n,1}, \ldots, a_{n,d})^\top$ and $\boldsymbol{b}_n = (b_{n,1}, \ldots, b_{n,d})^\top$, such that the limiting distribution

$$G(\boldsymbol{x}) = \lim_{n \to \infty} P(\boldsymbol{M}_n^* \leqslant \boldsymbol{x})$$

of the standardised maximum

$$\boldsymbol{M}_n^* = \left(\frac{M_{n,1} - b_{n,1}}{a_{n,1}}, \frac{M_{n,2} - b_{n,2}}{a_{n,2}}, \ldots, \frac{M_{n,d} - b_{n,d}}{a_{n,d}} \right)^\top$$

exists and is non-degenerate, then the limiting distribution of $\boldsymbol{M}_n^*$ must be the multivariate GEV distribution. This distribution is nonparametric and quite complex. However, it must satisfy a set of certain properties, such as max-stability and the fact that its marginal distributions must be univariate GEV distributions. This makes it possible to identify and construct parametric families of probability distributions that are proper subsets of the family of multivariate GEV distributions, such as the logistic model (Gumbel, 1960), the asymmetric logistic model (Tawn, 1990) and the Dirichlet model (Coles & Tawn, 1991). Multivariate max-stability means that, for every $n \in \mathbb{N}$, there exist vector constants $\boldsymbol{\alpha}_n > \boldsymbol{0}$ and $\boldsymbol{\beta}_n$ such that

$$G^n(\boldsymbol{\alpha}_n \odot \boldsymbol{x} + \boldsymbol{\beta}_n) = G(\boldsymbol{x}),$$

where the Hadamard product $\odot$ denotes element-wise multiplication.

The multivariate version of the extreme value theorem is not directly applicable in the same way as the univariate extreme value theory, since the different component-wise maxima in $\boldsymbol{M}_n$ may occur at completely different times, meaning that $\boldsymbol{M}_n$ may be different from all the individual observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. Thus, $\boldsymbol{M}_n$ often contains synthetic data that it is difficult to use for describing actual multivariate extreme events. However, just as in the univariate setting, the knowledge about the limiting distribution G makes it possible to say more about the distributional tails of F directly. If the limiting distribution G is nondegenerate, then the conditional distribution of $\boldsymbol{X}$, given that at least one component of $\boldsymbol{X}$ exceeds some large enough threshold, is approximately equal to the multivariate GP distribution (Rootzén, Segers, & Wadsworth, 2018; Rootzén, Segers, & L. Wadsworth, 2018). Furthermore, if G is nondegenerate, it

is also possible to describe threshold exceedances of $\boldsymbol{X}$ as a point-process following a non-homogeneous Poisson distribution (e.g. Davison & Huser, 2015). Thus, extreme value theory produces a large array of usable models for describing multivariate tail behaviour. However, these models are much more mathematically involved than the univariate tail models in Section 2.1. Since they have not been directly used in this thesis, I will not go more into the mathematical details. Multivariate extreme value theory can also be generalised into the functional setting, e.g., for modelling the extremes of a spatial random field. This is described in Section 2.3.

2.3 Spatial extreme value theory

A stochastic process $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ with index set $\mathcal{S}$ is denoted a max-stable process if and only if it has the property that, for every $n \in \mathbb{N}$, there exists two functions $\alpha_n(\boldsymbol{s}) > 0$ and $\beta_n(\boldsymbol{s})$ such that

$$P(X(\boldsymbol{s}) \leqslant \alpha_n(\boldsymbol{s})x + \beta_n(\boldsymbol{s}))^n = P(X(\boldsymbol{s}) \leqslant x).$$

Once more, the marginal distributions of a max-stable process must themselves be max-stable distributions, i.e., GEV distributions. This is therefore the natural generalisation of the GEV distribution into the functional setting. Most classical models for spatial extremes are based on max-stable processes, or variants thereof (e.g. Davison et al., 2019; Huser & Wadsworth, 2022).

De Haan (1984) shows that any max-stable process $X(\boldsymbol{s})$ can be represented as the pointwise maxima over an infinite amount of random processes,

$$X(\boldsymbol{s}) = \sup_{i=1,2,\ldots} R_i W_i(\boldsymbol{s}),$$

where $R_1, R_2, \ldots$ are points of a Poisson process on $[0, \infty)$ with intensity $r^{-2}\mathrm{d}r$, and $W_1(\boldsymbol{s}), W_2(\boldsymbol{s}), \ldots$ are independent realisations of a non-negative stochastic process $W(\boldsymbol{s})$ with mean $\mathrm{E}[W(\boldsymbol{s})] = 1$. This so-called spectral representation allows for an efficient method of simulating approximate max-stable processes, by sampling $R_1 > R_2 > \ldots$ sequentially, and then stopping at some sample nr. n such that R_n is so small that

$$\mathrm{P}\left(\sup_{i=n+1,n+2,\ldots} R_i W_i(\boldsymbol{s}) > \max_{i=1,2,\ldots,n} R_i W_i(\boldsymbol{s})\right) \approx 0,$$

which means that $X(\boldsymbol{s}) \overset{d}{\approx} \max_{i=1,\ldots,n} R_i W_i(\boldsymbol{s})$. This is illustrated in Figure 3. The spectral representation also allows for an easy way of developing models for max-stable processes, as it is possible to define a max-stable process simply by defining a distribution for the spectral process $W(\boldsymbol{s})$ that is nonnegative and has mean 1. This approach has given rise to a large variety of models for spatial extremes, such as the Schlater process (Schlather,

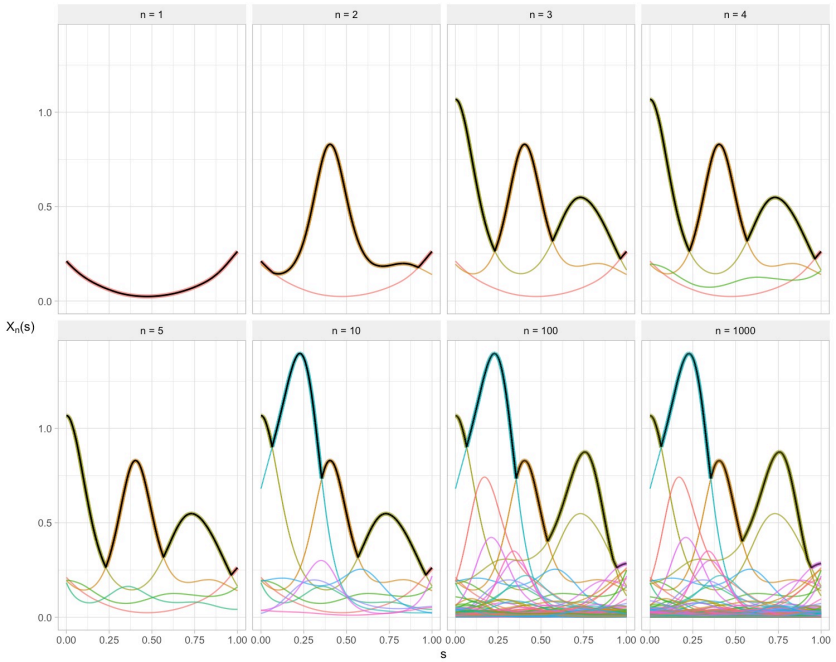


Figure 3: *Illustration of how to simulate max-stable processes using the spectral representation. The different plots display realisations of* $\{R_i W_i(s)\}_{i=1}^n$ *for different values of* n*, along with the pointwise maxima* $X_n(s) = \max\{R_i W_i(s)\}_{i=1}^n$.

2002), the Smith process (Smith, 1990), the Brown-Resnick process (Brown & Resnick, 1977) and the extremal-t process (Opitz, 2013).

Max-stable processes have seen considerable use in the literature for modelling spatial extremes of, e.g., precipitation (Padoan et al., 2010; Davison et al., 2012; Forster & Oesting, 2022), temperature (Davison & Gholamrezaee, 2012; Thibaud et al., 2016; Davison et al., 2019), river discharge (Asadi et al., 2015; Engelke & Hitz, 2020) and air pollution (Vettori et al., 2019).

Once more, a spatial equivalent of the univariate and multivariate threshold exceedance approaches exists, which gives rise to the so-called generalised Pareto and r-Pareto processes (Dombry & Ribatet, 2015; Ferreira & de Haan, 2014; De Fondeville & Davison, 2018). However, these are considerably more mathematically involved, and have only recently started to gain traction in the literature.

A considerable problem with max-stable processes is that the log-likelihood of a d-dimensional observation contains the same number of terms as the dth Bell number (Davison & Huser, 2015). The Bell numbers grow superexponentially towards infinity, as shown in Table 1, which means that likelihood-based inference becomes computationally intractable for

Table 1: *The nth Bell number, B_n, for different values of n (OEIS Foundation, 2023).*

n	B_n
1	1
5	52
10	115975
15	1382958545
20	51724158235372

most values of d larger than approximately 10. This problem can be somewhat overcome by performing inference using lower-dimensional composite likelihoods such as the pairwise likelihood (Padoan et al., 2010). However, this provides less efficient utilisation of the available data, and can easily become computationally demanding due to combinatorial difficulties as the dimension d grows.

Very recently, several likelihood-free inference methods, also known as simulation-based inference methods, have emerged as promising alternatives for performing high-dimensional inference with models that have intractable likelihoods (Lenzi et al., 2023; Sainsbury-Dale et al., 2023; Richards, Sainsbury-Dale, et al., 2023; Walchessen et al., 2023; Rai et al., 2023; Majumder & Reich, 2023). These methods allow for performing inference for essentially any statistical model which one can simulate from. Inference is performed by first simulating a large set of model parameters from a chosen prior distribution. Then, data is simulated from the model for each of the sampled parameter vectors. A machine learning method, such as a deep neural network, is then fitted to the large collection of data and model parameters, with the aim of estimating model parameters given a set of simulated data. Finally, the fitted machine learning method can be applied for performing inference by estimating model parameters for observational data sets. This methodology is promising, and has been shown to allow for fast inference for models that would previously have been considered intractable, such as a high-dimensional max-stable model. Still, the methodology is very new, and there are many challenges and unsolved problems remaining before it can be adopted for widespread use.

2.4 Extremal dependence

An important property of multivariate or spatial extreme value theory is the extremal dependence structure, which describes whether multiple extremes tend to occur simultaneously or not. A much used quantity for describing

extremal dependence is the conditional threshold probability

$$\chi_p(X,Y) = P\left(X > F_X^{-1}(p) \mid Y > F_Y^{-1}(p)\right),$$

where F_X^{-1} and F_Y^{-1} are quantile functions for the random variables X and Y, respectively. In the limit, as $p \to 1$, the conditional threshold probability converges to the tail correlation coefficient

$$\chi(X,Y) = \lim_{p\to 1} \chi_p(X,Y).$$

We say that X and Y are asymptotically dependent if $\chi(X,Y) > 0$, meaning that their extremes tend to occur simultaneously, while they are asymptotically independent if $\chi(X,Y) = 0$. Interestingly, X and Y are asymptotically independent if and only if their limiting block maxima are perfectly independent, meaning that the limiting bivariate GEV distribution reduces to $G(x,y) = G(x)G(y)$.

Max-stable processes provide a thorough and successful description of the asymptotic behaviour of random processes, which typically is described by the tail correlation coefficient χ. However, they often fail to describe the subasymptotic behaviour of such processes, i.e., how the distributional tails behave as we move towards the asymptotics, which can be represented by how χ_p converges towards χ. Instead, the models essentially assume that χ_p is constant and equal to χ for all large enough values of p.

As an illustrating example, let the random vector (X,Y) follow a bivariate Gaussian distribution with correlation ρ. Then, for all $|\rho| < 1$, X and Y are asymptotically independent (Sibuya et al., 1960). The values of χ_p are heavily affected by the value of ρ, with $\rho = 0$ producing small values of χ_p that converge quickly towards 0, while $\rho = 0.99$ produces much larger values of χ_p, that converge more slowly towards 0. This is displayed in Figure 4. However, since χ_p converges to 0 for all $|\rho| < 1$, the standardised component-wise maxima over infinitely many replications of (X,Y) attain identical limiting distributions, meaning that $\rho = 0$ and $\rho = 0.99$ is "the same thing" in the asymptotics. In practice, though, one should clearly care more about, e.g., precipitation amounts that are 99% correlated over a large area, than precipitation amounts that are perfectly independent for each location in space, as the former has a much larger probability of creating dangerous floods.

This demonstrates a problem with the classical models of multivariate and spatial extreme value theory. The models focus on describing the asymptotics, here represented by χ, but the data sets they are fitted to are seldom large enough to be considered truly asymptotic. The classical models may therefore provide poor fits to "subasymptotic data sets", because these are not large enough for the model assumptions to hold. Thus, if we fit a bivariate max-stable distribution to some tens of thousands of realisations

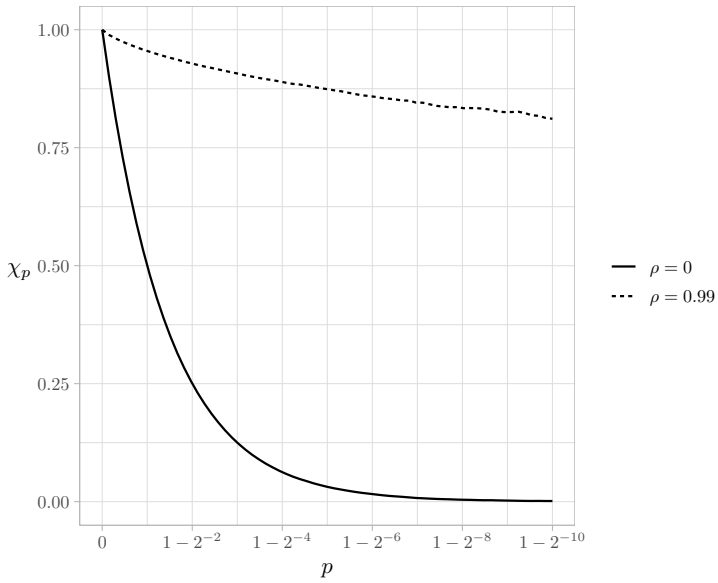


Figure 4: *Empirical estimates of χ_p for a bivariate Gaussian distribution with correlation coefficient ρ, created using Monte Carlo estimation.*

of the bivariate Gaussian distribution with correlation $\rho = 0.99$, we will, as shown in Figure 4, produce an estimate for the tail correlation coefficient of $\chi \approx 0.8$, meaning that we heavily overestimate extremes in the asymptotics. However, if we instead were able to model how χ_p converges towards χ, then we would capture the fact that χ_p is still decreasing, which might produce a better estimate for both the limiting value of χ and for the subasymptotic extremal dependence, represented through χ_p. For this reason, Ledford and Tawn (1996, 1997) developed extremal models with $\chi = 0$ that allow for a large spread of dependence structures at finite levels, and Coles et al. (1999) introduced an alternative measure, $\bar{\chi}$, which provides some information about the subasymptotic dependence strength under asymptotic independence. In recent years, the subasymptotic dependence structure has gained large amounts of interest, and there has been a flourishing of new types of extremal models that aim to describe both asymptotic and subasymptotic dependence. Some of these models are described in Section 2.5.

2.5 Subasymptotic models for spatial extremes

Realistic models for spatial extremes should be able to describe both asymptotic and subasymptotic behaviour, and should therefore capture both the value of χ and the behaviour of χ_p as it converges towards χ. Additionally, in the spatial setting, it is reasonable to assume that neighbouring locations display asymptotic dependence, while far-away locations display asymptotic

independence. Realistic models for spatial extremes should therefore allow for random processes whose asymptotic structure can vary from non-trivial dependence to non-trivial independence. By trivial asymptotic dependence, we here mean complete dependence, while trivial asymptotic independence means complete independence. Finally, to be more usable, such models should allow for computationally efficient inference in high-dimensional settings. Models based on max-stable processes satisfy none of these requirements, but much effort has recently been put into developing new subasymptotic models with more flexible dependence structures and/or less demanding inference. These new model types include the classes of max-infinitely divisible (max-id) processes, inverted max-stable processes, scale-mixture models, max-mixture models and conditional extremes models, among others (Huser & Wadsworth, 2022). Max-id processes are used for describing the distribution of component-wise maxima, while the remaining subasymptotic models are used for describing individual events that exceed some type of extreme threshold.

Max-infinitely divisible (max-id) processes were first characterised by Balkema and Resnick (1977), and more theory for the processes have later been developed (e.g. Giné et al., 1990; Resnick, 2008; Kabluchko & Schlather, 2010). A max-id process, $X(\boldsymbol{s})$, is a random process that can be defined as the point-wise maximum of n i.i.d. realisations of some random process $Y_n(\boldsymbol{s})$, for any $n \in \mathbb{N}$. This makes max-id processes a natural extension of max-stable processes. To be more exact, a finite-dimensional probability distribution F is max-id if and only if F^t is a valid probability distribution for all real $t > 0$, and a random process is max-id if all its finite-dimensional distributions are max-id (Bopp et al., 2021). The class of max-id processes is much larger than the class of max-stable processes, but it is possible to develop constrained sub-classes of max-id processes that are in some sense close enough to a max-stable process to be of use. Such sub-classes of max-id processes have been developed and applied with some success by Padoan (2013), Bopp et al. (2021), Huser et al. (2021), and Zhong et al. (2022). Inference with a max-id process is comparable in complexity to inference with a max-stable process. Lower-dimensional composite likelihoods, such as the pairwise likelihood, are therefore often necessary to achieve computationally tractable inference. The exception is Bopp et al. (2021), who develop a Bayesian hierarchical max-id process with conditional independence assumptions which lead to computationally tractable inference using Markov chain Monte Carlo (MCMC) methods.

Inverted max-stable processes, as proposed by Wadsworth and Tawn (2012b), are reciprocals of max-stable processes. Thus, for any max-stable process $X(\boldsymbol{s})$, the process $\tilde{X}(\boldsymbol{s}) = 1/X(\boldsymbol{s})$ is an inverted max-stable process. An asymptotically dependent max-stable process gives rise to an asymptotically independent inverted max-stable process, thus resulting in a large

class of models for asymptotic independence with different subasymptotic properties. However, inverted max-stable processes are tightly connected to max-stable processes, and likelihood-based inference is therefore computationally intractable for large dimensions without lower-dimensional composite likelihood methods. Additionally, the processes are unable to vary from non-trivial asymptotic dependence to non-trivial asymptotic independence in space. The inverted max-stable models are close to GP threshold exceedance models in that they describe actual events of a random process, instead of describing synthetic data like the vector of component-wise maxima. This makes it possible to provide more realistic descriptions of the extremal dependence structure. However, since the models are asymptotically motivated, data with simultaneous extreme and non-extreme observations can lead to biased inference (Ledford & Tawn, 1996; Huser et al., 2016). It is therefore common to employ censoring of small observations, such that only contributions from large enough observations are fully included into the (composite) likelihood. This leads to more computationally demanding inference because it becomes necessary to integrate out all observations that are too small. This is not a considerate problem if one is already performing inference using a lower-dimensional composite likelihood, but it can be a considerable problem when relying on higher-dimensional likelihoods for performing inference. Some applications of inverted max-stable processes are found in Davison et al. (2013), Thibaud et al. (2013), and Kereszturi et al. (2016).

To address the need for non-trivial descriptions of both asymptotic dependence classes, Wadsworth and Tawn (2012b) developed the class of max-mixture models. The random process

$$X(\boldsymbol{s}) = \max(cX_1(\boldsymbol{s}), (1-c)X_2(\boldsymbol{s}))$$

is a max-mixture process if $X_1(\boldsymbol{s})$ is a max-stable process and $X_2(\boldsymbol{s})$ is an inverted max-stable process that is independent of $X_1(\boldsymbol{s})$. The parameter $c \in [0,1]$ is a mixing parameter. This model allows for simultaneous non-trivial descriptions of both asymptotic dependence classes. However, the mixing parameter c is often difficult to estimate, and it is not so clear how the parametrisations of $X_1(\boldsymbol{s})$ and $X_2(\boldsymbol{s})$ affect the properties of $X(\boldsymbol{s})$. Additionally, likelihood-based inference is even more challenging for the max-mixture processes than for the max-stable and inverted max-stable processes, and censoring is also necessary here. Some applications of max-mixture models are found in (Bacro et al., 2016; Ahmed et al., 2020; Abu-Awwad et al., 2020).

Another proposed solution is that of scale-mixture modelling (Opitz, 2016; Huser et al., 2017; Huser & Wadsworth, 2019; Engelke et al., 2019; Huser & Wadsworth, 2019; Zhang, Risser, et al., 2022; Zhang, Shaby, & Wadsworth, 2022). Inspired by the spectral representation from Section 2.2,

scale-mixture models represent random processes as $X(\boldsymbol{s}) = RW(\boldsymbol{s})$, where $W(\boldsymbol{s})$ can be any kind of stochastic process, while R is a scale random variable that "lifts" the process $W(\boldsymbol{s})$ to the correct level. Typically, the process $W(\boldsymbol{s})$ displays asymptotic independence, such that the process $X(\boldsymbol{s})$ displays asymptotic independence if $W(\boldsymbol{s})$ is more heavy-tailed than R, and asymptotic dependence if R is more heavy-tailed than $W(\boldsymbol{s})$. As an example, Huser and Wadsworth (2019) develop a scale-mixture model on the form $X(\boldsymbol{s}) = R^{\delta}W(\boldsymbol{s})^{1-\delta}$, with $\delta \in [0, 1]$, such that $X(\boldsymbol{s})$ displays asymptotic independence when $\delta > 1/2$ and asymptotic independence when $\delta < 1/2$. The freedom of choice for the distributions of R and $W(\boldsymbol{s})$ leads to large amounts of flexibility for describing the subasymptotic dependence structure, and the simple form of the scale-mixture model leads to more simple inference. However, it is still necessary to perform censoring to avoid bias from non-extreme observations, meaning that inference is computationally demanding and that one often needs to rely on less efficient inference methods such as low-dimensional composite likelihoods. Additionally, the form of $X(\boldsymbol{s})$ makes it impossible to describe non-trivial changes in the asymptotic dependence class as a function of distance without further modifications.

The spatial conditional extremes model (Wadsworth & Tawn, 2022) is the spatial extension of the conditional extremes model (Heffernan & Tawn, 2004; Heffernan & Resnick, 2007), which describes the distribution of a random vector conditional on one of its components exceeding an extreme threshold. The model is quickly gaining popularity in the field of extreme value theory, and has already seen much usage (Shooter et al., 2019; Simpson et al., 2023; Simpson & Wadsworth, 2021; Richards et al., 2022; Shooter et al., 2022; Richards, Tawn, & Brown, 2023). It is also the model we have chosen for performing spatial modelling of extreme precipitation in Paper 3 and Paper 4. Details about the model are given in Section 2.6.

2.6 The spatial conditional extremes model

The spatial conditional extremes model describes the conditional distribution of a spatial random process, $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ with $\mathcal{S} \subseteq \mathbb{R}^2$, given that it exceeds some large threshold τ at a prespecified conditioning site $\boldsymbol{s}_0 \in \mathcal{S}$. The model is easiest described if $X(\boldsymbol{s})$ follows the standardised Laplace marginal distribution, everywhere in space (Keef et al., 2013). The Laplace distribution is also known as the double exponential distribution, and it has probability density function $f(x) = \exp(-|x|)$. Given Laplace marginals, the model assumes that there exist two standardising functions, $a(\boldsymbol{s}; \boldsymbol{s}_0, x_0)$, with $a(\boldsymbol{s}_0; \boldsymbol{s}_0, x_0) = x_0$, and $b(\boldsymbol{s}; \boldsymbol{s}_0, x_0) > 0$, such that the conditional process

$$Z(\boldsymbol{s}; \boldsymbol{s}_0, x_0) = \left[\frac{X(\boldsymbol{s}) - a(\boldsymbol{s}; \boldsymbol{s}_0, x_0)}{b(\boldsymbol{s}; \boldsymbol{s}_0, x_0)} \middle| X(\boldsymbol{s}_0) = x_0 \right]$$

converges in distribution to a residual process $Z(\boldsymbol{s}; \boldsymbol{s}_0)$, which is independent of x_0, in the limit where $x_0 \to \infty$. The residual process has a nondegenerate distribution and satisfies $Z(\boldsymbol{s}_0; \boldsymbol{s}_0) = 0$ almost surely. To use the spatial conditional extremes model in practice, one assumes that the limiting distribution holds exactly when x_0 exceeds a large enough threshold τ, i.e.,

$$Z(\boldsymbol{s}; \boldsymbol{s}_0, x_0) \overset{d}{=} Z(\boldsymbol{s}; \boldsymbol{s}_0) \; \forall \; x_0 > \tau.$$

The requirements on the standardising functions $a(\cdot)$ and $b(\cdot)$ and on the residual process $Z(\cdot)$ are quite weak, which makes the spatial conditional extremes model highly flexible and suitable for modelling a large variety of different phenomena. The model also allows for non-trivial descriptions of both asymptotic dependence and independence, and for non-trivial changes in the asymptotic dependence class as a function of the distance between different locations. Additionally, since the model is designed to describe both extreme and non-extreme observations, conditional on $X(\boldsymbol{s}_0) > \tau$, it becomes possible to perform inference without the censoring schemes that are necessary for all the previously described subasymptotic extreme value models. Thus, if the distribution of the residual process $Z(\cdot)$ is computationally tractable, it is possible to perform efficient likelihood-based inference without any form of lower-dimensional composite likelihood method. In order to model high-dimensional spatial extremes with the spatial conditional extremes model, it is common to use a Gaussian copula for the residual process $Z(\cdot)$, since most other copulas tend to result in computationally intractable inference. Constraining the copula to be Gaussian leads to a loss in model flexibility, but the spatial conditional extremes model can still capture a large variety of different extremal dependence structures, and encompasses several of the classical extremal models based on max-stable processes (Wadsworth & Tawn, 2022).

In Paper 3 and Paper 4, we build upon a variant of the spatial conditional extremes model, proposed by Simpson et al. (2023). In this model variant, the residual process is modelled as a Gaussian Markov random field, created using the so-called stochastic partial differential equation (SPDE) approach of Lindgren et al. (2011), and a nugget effect is added to the model, which turns it into a latent Gaussian model and allows for computationally efficient inference using the method of integrated nested Laplace approximations (INLA; Rue et al., 2009). See Section 3 and Section 4 for more details about the SPDE approach and INLA, respectively.

Unfortunately, the flexibility and computational efficiency of the conditional extremes model comes at a cost. The spatial conditional extremes model depends on choosing a conditioning site $\boldsymbol{s}_0 \in \mathcal{S}$ before performing modelling and inference. Therefore, the model is only a local model, in the sense that it can only perform inference using data from time points where

the threshold is exceeded at $\boldsymbol{s}_0$. In comparison, the previously described subasymptotic models are global models, in the sense that they can use all of the available data for inference, although not without censoring of non-extreme observations. This means that only a small fraction of the data are available for inference when performing high-dimensional modelling with the spatial conditional extremes model. To account for this, Heffernan and Tawn (2004) and Wadsworth and Tawn (2022) develop composite likelihood methods that make it possible to combine information from multiple conditioning sites by creating a global model fit that on average provides the best fit to the data. However, these composite likelihoods wrongfully assume independence between dependent random variables, which often leads to overconfident uncertainty estimates that can be particularly problematic when performing Bayesian inference (e.g. Ribatet et al., 2012). In Paper 3, we show how to account for the misspecification of the global model fit to produce more reliable Bayesian inference using more of the available data.

Another complication of the spatial conditional extremes model is that it lacks self-consistency, in the sense that models based on different conditioning sites can produces different estimates for the same events. This is a well known problem that can occur when defining a joint distribution using a set of conditional distributions (Besag, 1974). We illustrate the problem by assuming that some bivariate random variable (X, Y) has a fully symmetric distribution with Laplace marginal distributions, and that its extremes follow the conditional extremes model, with

$$[X \mid Y = t] \overset{d}{=} [Y \mid X = t] \overset{d}{=} a(t) + b(t)Z, \quad t > \tau,$$

with threshold $\tau = 3$, standardising functions $a(t) = 0.8t$ and $b(t) = t^{0.4}$ and the residual term $Z \sim \mathcal{N}(0, 0.2^2)$. Probability density functions for $[X \mid Y = y]$ and $[Y \mid X = x]$ with $x, y > \tau$ are visualised in the left and right subplots of Figure 5, respectively. The probability that $(X, Y) \in A$ can be computed as

$$\begin{aligned} \mathrm{P}\{(X, Y) \in A\} &= \int \mathrm{P}\{(X, Y) \in A \mid Y = y\}\, \pi(y) \mathrm{d}y \\ &= \int \mathrm{P}\{(X, Y) \in A \mid X = x\}\, \pi(x) \mathrm{d}x, \end{aligned}$$

where $\pi(y)$ and $\pi(x)$ are the probability density functions of Y and X, respectively. From Figure 5 we see that these two integration paths can provide completely different answers. Similarly, multiple probabilities can be computed for the event B, and for any other event where both X and Y exceed the threshold τ. The inconsistency problem of the conditional extremes model is examined by Heffernan and Tawn (2004) and Liu and Tawn (2014), and it is also examined in an unpublished note located right

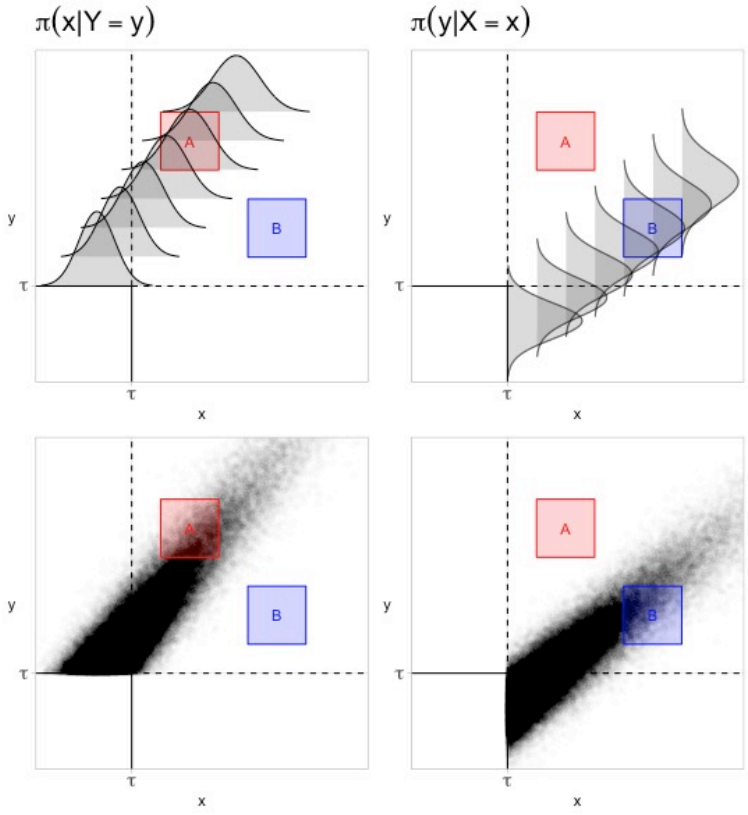


Figure 5: *Visualisation of the self-inconsistency of the conditional extremes model. The upper subplots display probability density functions for* $[X \mid Y = y]$ *with* $y > \tau$ *(left) and* $[Y \mid X = x]$ *with* $x > \tau$ *(right). The lower subplots display Monte Carlo simulations used for estimating probabilities on the form* $\int P\{(X,Y) \in \cdot \mid Y = y\}\, \pi(y)dy$ *(left) and* $\int P\{(X,Y) \in \cdot \mid X = x\}\, \pi(x)dx$ *(right). The red and blue boxes display two events,* A *and* B*, where both* $X > \tau$ *and* $Y > \tau$*, which can have multiple different probabilities, depending on whether we first condition on* $X > \tau$ *or* $Y > \tau$ *when we compute them.*

after Paper 3 in the thesis, but no good solutions have been proposed on how to ensure self-consistency.

The lack of self-consistency and of a global model definition can make the conditional extremes model seem problematic to use for some. However, as we show in Paper 3 and Paper 4, it is possible to account for these problems if one is aware of them, and thus to perform successful modelling of high-dimensional spatial extremes with the spatial conditional extremes model. In Paper 3 we apply a post hoc transformation method that transforms samples of the posterior distribution to achieve Bayesian inference that is more robust towards misspecification, and we show how this leads to more successful global modelling with a composite likelihood. Then, in Paper 4, we propose new forms for the standardising functions $a(\cdot)$ and $b(\cdot)$ of the spatial conditional extremes model, that make it possible to lower the thresholds τ to perform inference with more data for each of the conditioning sites. In some sense, we can say that the spatial conditional extremes model trades away nice theoretical probabilities, such as self-consistency, to achieve good properties for practical modelling of the subasymptotics. Some might find this tradeoff to be problematic, but I believe that it can be a good choice for achieving successful modelling of high-dimensional spatial extremes, as long as one is aware of the pitfalls and the limitations of the model.

3 Computationally efficient spatial modelling

Most of the models in Paper 2, 3 and 4 contain spatial random fields, which can improve model performance by capturing spatial dependence structures and by making it possible to share information between neighbouring locations. For a review of spatial statistics, see, e.g., Diggle et al. (1998), Gelfand et al. (2010), Banerjee et al. (2014), and Cressie and Wikle (2015). A problem with the modelling of complex spatial dependence structures is that it quickly becomes computationally demanding or intractable as the model dimension grows. To circumvent this, statisticians tend to add assumptions to their models that lead to more computationally efficient modelling, typically at the cost of decreased model flexibility or correctness, i.e., a type of bias/speed tradeoff, instead of the more commonly known bias/variance tradeoff.

The by far most common modelling assumption in spatial statistics is that most, or all, of the spatial dependence structure can be described well by a (possibly transformed) Gaussian process. This is a common approach because Gaussian processes have many nice theoretical processes and are relatively flexible, while also being considerably more computationally efficient than most other reasonable choices for describing spatial dependence. It is well understood that the entire world is not perfectly Gaussian, but

the Gaussianity assumption tends to produce small amounts of bias while leading to considerable speed improvements.

For a d-dimensional Gaussian random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$, the covariance function can be used to define the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_d) \\ \mathrm{Cov}(X_2, X_1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathrm{Cov}(X_d, X_1) & \cdots & \cdots & \mathrm{Cov}(X_d, X_d) \end{pmatrix}.$$

This matrix must be positive semi-definite, i.e., $\boldsymbol{c}^\top \boldsymbol{\Sigma} \boldsymbol{c} \geqslant 0 \; \forall \; \boldsymbol{c} \in \mathbb{R}^d$. If $\boldsymbol{\Sigma}$ is positive definite, then the distribution of $\boldsymbol{X}$ is non-degenerate, and it has a probability density function. This probability density function is

$$\pi(\boldsymbol{x}) = (2\pi)^{-d/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}, \tag{3}$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ is the mean vector of $\boldsymbol{X}$ and $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. Thus, an important part of spatial statistics is to propose parametric covariance functions that are flexible and suitable for the application at hand, while also ensuring a positive semi-definite covariance matrix for any given set of locations and model parameters.

Some nonparametric covariance functions have been developed (e.g. Yin et al., 2010; Fox & Dunson, 2015). However, since it can be difficult to guarantee positive semi-definite covariance matrices when using nonparametric covariance functions, the main focus within spatial statistics is on parametric covariance functions. When choosing a parametric covariance function, it is common to assume weak stationarity of the random field, which means that the covariance between two random variables of the random field only depends on the distance between them. Then, given a suitable distance function, we can describe the correlation between two random variables with distance h, using the autocorrelation function $\gamma(h)$. Any positive definite function $\gamma(\cdot)$ guarantees a positive definite covariance matrix (e.g., Gelfand et al., 2010).

Of all possible autocorrelation functions, the Matérn autocorrelation function (Matern, 1986; Guttorp & Gneiting, 2006)

$$\gamma(h) = \frac{1}{2^{\nu-1}\Gamma(\nu)} (\kappa h)^\nu K_\nu(\kappa h), \tag{4}$$

where $\nu > 0$ is a smoothness parameter, $\rho = \sqrt{8\nu}/\kappa$ is a range parameter and K_ν is the modified Bessel function of the second kind and order ν, is one of the most popular choices (Stein, 1999). Other popular choices, due to their simplicity, are the exponential autocorrelation function $\gamma(h) = \exp(-h/\rho)$

and the Gaussian autocorrelation function $\gamma(h) = \exp(-h^2/(2\rho^2))$, where ρ is a range parameter. However, these functions are special cases of the Matérn autocorrelation function, when the smoothness parameter equals $\nu = 1/2$ or as $\nu \to \infty$, respectively.

Even though Gaussian spatial processes lead to computationally efficient inference, the computation of high-dimensional likelihoods can still be demanding, as computing the probability density function in (3) requires computing $\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ and $|\boldsymbol{\Sigma}|$, which typically has a cost that scales cubicly with the model dimension d. Further steps must therefore be taken to achieve computationally tractable inference for large model dimensions. The most common methods for achieving this are either based on representing or approximating the Gaussian likelihood as a product of lower-dimensional conditional or unconditional composite likelihoods (Besag, 1975; Vecchia, 1988; Stein et al., 2004; Varin et al., 2011), or on adding conditional or unconditional independence assumptions to the model, which lead to sparse or low-rank covariance matrices, or to covariance matrices with sparse inverses, and therefore reduce the cost of computing the full likelihood. The most common independence assumptions are those of covariance tapering methods (Furrer et al., 2006; Kaufman et al., 2008) and low-rank estimation methods such as process convolutions (Higdon, 2002), fixed rank Kriging (Cressie & Johannesson, 2008), predictive processes (Banerjee et al., 2008), lattice Kriging (Nychka et al., 2015) and the SPDE approach (Lindgren et al., 2011).

Composite likelihoods have already been mentioned several times in Section 2 due to their usefulness. A composite likelihood for the random vector $\boldsymbol{X}$ is a function that is the product of one or more conditional or unconditional likelihood functions for subsets of $\boldsymbol{X}$ (Lindsay, 1988). An early version of a composite likelihood is the pseudolikelihood of Besag (1974, 1975), which is the product of conditional likelihoods based on neighbouring random variables,

$$\pi_{\text{Besag}}(\boldsymbol{x}) = \prod_{i=1}^{d} \pi(x_i \mid \{x_j : j \in n(i)\}),$$

where $n(i) \subset \{1, 2, \ldots, d\}$ is the set of indices for all neighbours of the ith component of $\boldsymbol{X}$, where the definition of a "neighbour" typically depends on the application. Later, Vecchia (1988) takes advantage of the fact that any distribution can be decomposed into

$$\pi(\boldsymbol{x}) = \pi(x_1) \prod_{i=2}^{d} \pi(x_i \mid x_1, \ldots, x_{i-1}),$$

and develops the Vecchia approximation

$$\pi_{\text{Vecchia}}(\boldsymbol{x}) = \pi(x_1) \prod_{i=2}^{d} \pi(x_i \mid \{x_j : j \in n(i)\}),$$

where, this time, $n(i) \subseteq \{1, 2, \ldots, i-1\}$ is the set of indices that is smaller than i, for all neighbours of the ith component of $\boldsymbol{X}$. Note that there exists generalisations of the Vecchia approximation that cannot be classified as composite likelihoods (Katzfuss & Guinness, 2021). Other popular composite likelihoods include: the independence likelihood (Chandler & Bate, 2007), where all components of $\boldsymbol{X}$ are treated as independent; the pairwise likelihood (Cox & Reid, 2004), where the composite likelihood is built using a subset of the $\binom{d}{2}$ possible joint pairwise distributions for components of $\boldsymbol{X}$; and the block likelihood (Eidsvik et al., 2014), which is a generalisation of the pairwise likelihood, in which the composite likelihood is built using joint distributions over larger (possibly overlapping) blocks of components of $\boldsymbol{X}$. These composite likelihood methods can be very powerful, and can lead to computationally tractable inference in almost any setting, such as inference for max-stable processes (Padoan et al., 2010; Davison et al., 2012; Richards et al., 2022). However, composite likelihood methods can often be less efficient, as they essentially discard all information that can be gained from observations that are far away from each other or that can be gained from interactions between a large group of random variables. Additionally, composite likelihoods may not be proper likelihoods, in the sense that they might not integrate to 1 when integrating out the data. Therefore, it is hard to interpret how the definition of a neighbour or the number of likelihood factors might affect the final inferences.

Covariance tapering methods are based on the fact that the Hadamard product, i.e., element-wise multiplication, of two positive definite matrices is positive definite (Styan, 1973). This makes it possible to produce a sparse, valid covariance matrix, by computing the Hadamard product of the original dense covariance matrix and a sparse covariance matrix that equals zero for all location pairs that are far enough away from each other,

$$\boldsymbol{\Sigma}_{\text{tapered}} = \boldsymbol{\Sigma}_{\text{original}} \odot \boldsymbol{\Sigma}_{\text{sparse}},$$

where $\odot$ denotes the Hadamard product. Low-rank estimation, on the other hand, is based on the assumption that a spatial Gaussian random field can be represented as a linear combination of a lower-dimensional multivariate Gaussian basis vector, i.e.,

$$\boldsymbol{X}(\boldsymbol{s}) = \sum_{i=1}^{k} a_i(\boldsymbol{s}) W_i,$$

where $X(\boldsymbol{s})$ is our spatial Gaussian random field, recorded at some location $\boldsymbol{s} \in \mathcal{S}$, inside the spatial domain $\mathcal{S} \subseteq \mathbb{R}^2$, $\boldsymbol{W} = (W_1, \ldots, W_k)^\top$ is a k-dimensional Gaussian random vector, and $\{a_i\}_{i=1}^k$ is a collection of k spatial basis functions. The d-dimensional random vector $\boldsymbol{X} = (X(\boldsymbol{s}_1), X(\boldsymbol{s}_2), \ldots, X(\boldsymbol{s}_d))^\top$, with $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d \in \mathcal{S}$, can then be represented as

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{W},$$

where $\boldsymbol{A}$ is a $(d \times k)$-dimensional projection matrix, containing the values of the k basis functions at all d locations. In high-dimensional settings, the dimension k of the underlying random vector $\boldsymbol{W}$ is typically much smaller than the dimension d of $\boldsymbol{X}$, which one can take advantage of to perform computationally efficient inference. Additionally, the basis functions tend to have a compact support, which can make the projection matrix $\boldsymbol{A}$ sparse, and therefore speed up computations even further.

A problem with the methods that produce sparse covariance matrices is that the element Σ_{ij} at the ith row and jth column of $\boldsymbol{\Sigma}$ is zero only if X_i and X_j are completely independent. This is a very strong assumption that seldom holds true in practice. For this reason, Rue and Held (2005) propose to reparametrise the Gaussian random field from its covariance matrix $\boldsymbol{\Sigma}$ to its precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$. For an element Q_{ij} of the precision matrix to be zero, it is only required that X_i and X_j are conditionally independent given all other variables, $\boldsymbol{X}_{-ij}$, in $\boldsymbol{X}$. This is a much weaker assumption that more easily holds true. Therefore, one can often achieve more sparse precision matrices using weaker model assumptions. However, there is no such thing as a free lunch, and the elements of $\boldsymbol{Q}$ are much harder to interpret than the elements of $\boldsymbol{\Sigma}$, because they represent global properties of the dependence structure (Rue & Held, 2005). This makes it difficult to create suitable parametrisations of the precision matrix that are both positive definite and that represents realistic dependence structures. Gaussian random fields with sparse precision matrices are known as Gaussian Markov random fields.

In their seminal work, Lindgren et al. (2011) develop a parametric formula for creating valid sparse precision matrices that approximate the precision matrices of Gaussian random fields with Matérn covariance functions. Their method is known as the stochastic partial differential equation (SPDE) approach, and it is based on the fact that all Gaussian random fields $X(\boldsymbol{s})$ with Matérn covariance functions are solutions to the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} X(\boldsymbol{s}) = W(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2,$$

where κ and ν are the parameters of the Matérn covariance function in (4), $W(\boldsymbol{s})$ is Gaussian white noise and Δ is the Laplacian operator (Whittle, 1954). By numerically approximating a solution of this SPDE with the

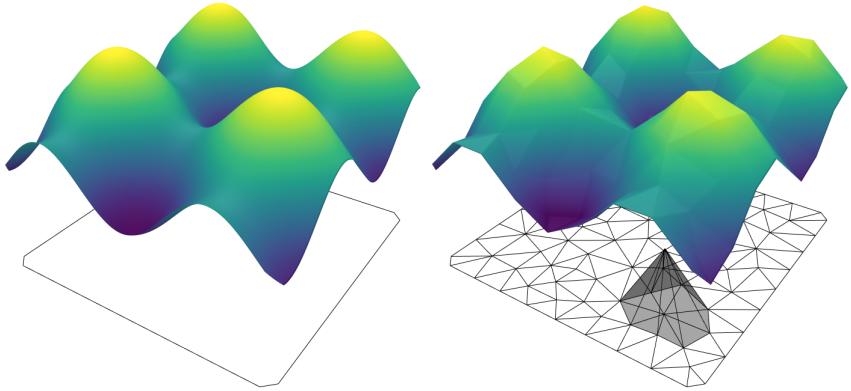


Figure 6: *Illustration of how the finite element method approximates a smooth spatial field (left plot) with a spline function based on piecewise linear basis functions, created using a triangulated mesh (right plot). The spline consists of one piecewise linear basis function for each node in the mesh, and one of these basis functions are displayed right on top of the mesh. This illustration is inspired by the illustration of Cameletti et al. (2013).*

finite element method, they create a parametrised class of Gaussian Markov random fields with precision matrices that are very close to those of the Matérn covariance function. Explained in a simplistic manner, the finite element method approximates the solutions of partial differential equations using a spline function that is based on piecewise linear basis functions that are defined using some triangulated mesh. See, e.g., Bakka (2022) for a simple and non-extensive introduction to the finite element method for statisticians. This is visualised in Figure 6, where the finite element method is used for approximating the smooth function in the left subplot using a spline function, displayed in the right subplot, based on piecewise linear basis functions from a triangulated mesh.

The SPDE approach is heavily applied in Papers 2, 3 and 4, as we find that it allows for flexible and computationally efficient inference for high-dimensional problems, while being based on reasonable assumptions of conditional independence as well as a covariance structure that is close to the Matérn covariance. The SPDE approach has also been found to perform well in comparison with other methods for fast high-dimensional inference (Heaton et al., 2019). Additionally, the SPDE approach performs particularly well in conjunction with the INLA inference method (Rue et al., 2009), which is based upon performing all necessary computations for inference using only the precision matrix. This is further described in Section 4.

As previously mentioned, assumptions of Gaussianity and (possibly conditional) independence are often added to a model for computational reasons,

and not necessarily because we believe them to be completely true. This can lead to model misspecification, which can complicate the interpretation of inferred parameter estimates and uncertainty intervals, and which can be particularly challenging when performing Bayesian inference (Ribatet et al., 2012; Kleijn & van der Vaart, 2012; Shaby, 2014). As an example that is somewhat explored in Papers 2, 3 and 4, the SPDE approximation can lead to large degrees of model misspecification if it is based on too strong smoothness assumptions, which leads to overestimation of the variance parameters in the model. Additionally, if too strong independence assumptions are added to a model, the effective sample size will be overestimated, which in turn can lead to underestimation of the model uncertainty. However, misspecification is not something that one should avoid at all costs, as it might often be necessary to achieve computationally tractable inference. Additionally, it is impossible to completely avoid misspecification, since all models are wrong (Box, 1976). Therefore, as long as one is aware of the consequences and potential pitfalls, the addition of misspecification through strong model assumptions can be an important and successful tool. In Paper 3, we also examine different methods that can achieve more robust inference by accounting for the discrepancies between model assumptions and the available data, thus allowing us to enjoy the benefits of strong modelling assumptions while also attaining more reasonable uncertainty estimates.

4 Latent Gaussian models

The model type that is used the most in Papers 2, 3 and 4 is the latent Gaussian model. Latent Gaussian models are both highly flexible and mathematically quite simple, which has made them into popular and successful tools for modelling complex phenomena in a large range of fields and applications, including spatial and spatio-temporal modelling (Gelman et al., 2013; Banerjee et al., 2014).

The latent Gaussian model assumes that all observations are conditionally independent of each other given a latent Gaussian random process that describes the entire dependence structure of the data. This assumption leads to a flexible model framework that is both elegant and simple to describe. To be more exact, we assume that the n response variables $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ are conditionally independent given an m-dimensional latent field, $\boldsymbol{x} = (x_1, \ldots, x_m)^\top$, and a vector of hyperparameters, $\boldsymbol{\theta}_1$, such that the conditional probability density function of the response variables, commonly referred to as the likelihood, can be written as

$$\pi(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}_1) = \prod_{i=1}^{n} \pi(y_i \mid \eta_i(\boldsymbol{x}), \boldsymbol{\theta}_1),$$

where all the marginal likelihoods, $\{\pi(y_i \mid \eta_i(\boldsymbol{x}), \boldsymbol{\theta}_1)\}_{i=1}^n$, belong to the same family of probability distributions, parametrised by $\eta_i(\boldsymbol{x})$ and $\boldsymbol{\theta}_1$, and $\eta_i(\boldsymbol{x})$ is a linear combination of the variables in $\boldsymbol{x}$, known as the linear predictor. The likelihood can be either a probability density function or a probability mass function for $[\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}_1]$. Furthermore, the latent field $\boldsymbol{x}$ is assumed to be conditionally Gaussian given a vector of hyperparameters $\boldsymbol{\theta}_2$, i.e.,

$$[\boldsymbol{x} \mid \boldsymbol{\theta}_2] \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}_2), \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2)),$$

where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ denotes a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q}$. It is common to perform Bayesian inference for latent Gaussian models, and in that setting the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^\top$ are assumed to follow a prior distribution with a probability density function that tends to be factorised into

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2).$$

Since $\boldsymbol{\eta}_i(\boldsymbol{x})$ is a linear combination of $\boldsymbol{x}$, we can write $\boldsymbol{\eta}_i(\boldsymbol{x}) = \boldsymbol{A}_i^\top \boldsymbol{x}$, where $\boldsymbol{A}_i$ is an m-dimensional column-vector of fixed coefficients known as a design vector, or a projection vector. This leads to the formulation

$$\boldsymbol{\eta}(\boldsymbol{x}) = (\eta_1(\boldsymbol{x}), \eta_2(\boldsymbol{x}), \ldots, \eta_n(\boldsymbol{x}))^\top = (\boldsymbol{A}_1^\top \boldsymbol{x}, \boldsymbol{A}_2^\top \boldsymbol{x}, \ldots, \boldsymbol{A}_n^\top \boldsymbol{x})^\top = \boldsymbol{A}\boldsymbol{x},$$

with the $(n \times m)$-dimensional design matrix $\boldsymbol{A}$. This matrix formulation, together with the Gaussianity of $\boldsymbol{x}$, makes it simple to incorporate a large variety of non-trivial trends and dependence structures into the model by setting

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{(1)}\boldsymbol{x}^{(1)} + \boldsymbol{A}^{(2)}\boldsymbol{x}^{(2)} + \cdots + \boldsymbol{A}^{(j)}\boldsymbol{x}^{(j)},$$

where the vectors $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(j)}$ partition the latent field $\boldsymbol{x}$, and the matrices $\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \ldots, \boldsymbol{A}^{(j)}$ are lower-dimensional design matrices. Here, $\boldsymbol{A}^{(1)}\boldsymbol{x}^{(1)}$ might, e.g., be a set of regression coefficients, $\boldsymbol{x}^{(1)}$, multiplied with a set of covariates, $\boldsymbol{A}^{(1)}$, while $\boldsymbol{A}^{(2)}\boldsymbol{x}^{(2)}$ might represent a Gaussian spline function, $\boldsymbol{A}^{(3)}\boldsymbol{x}^{(3)}$ might be a spatial Gaussian random field used for describing the spatial dependence structure, and so on. To extend a latent Gaussian model by adding a new component, $\boldsymbol{A}^{(j+1)}\boldsymbol{x}^{(j+1)}$, that is independent of all the other components of the latent field, we can simply set

$$\boldsymbol{A}_{\text{new}} = (\boldsymbol{A}, \boldsymbol{A}^{(j+1)}), \qquad \boldsymbol{x}_{\text{new}} = (\boldsymbol{x}^\top, \boldsymbol{x}^{(j+1)^\top})^\top,$$

where the extended latent field $\boldsymbol{x}_{\text{new}}$ has mean and precision matrix

$$\boldsymbol{\mu}_{\text{new}} = (\boldsymbol{\mu}^\top, \boldsymbol{\mu}^{(j+1)^\top})^\top, \qquad \boldsymbol{Q}_{\text{new}} = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}^{(j+1)} \end{pmatrix},$$

respectively, where the added component $\boldsymbol{x}^{(j+1)}$ is a Gaussian random vector with mean $\boldsymbol{\mu}^{(j+1)}$ and precision matrix $\boldsymbol{Q}^{(j+1)}$.

In a latent Gaussian model, both the hyperparameters $\boldsymbol{\theta}$ and the latent field $\boldsymbol{x}$ represent unobservable model parameters. The dimension m of the latent field can be of great magnitude, which makes inference methods like maximum likelihood estimation (MLE) and Markov chain Monte Carlo (MCMC) challenging and computationally demanding. It is possible to reduce the dimension of the parameter space by considering parts of $\boldsymbol{x}$ as nuisance parameters that can be integrated out. However, this requires computing integrals on the form

$$\int_{\mathcal{X}} \pi(y \mid \eta(\boldsymbol{x}), \boldsymbol{\theta}_1) \mathrm{d}\boldsymbol{x},$$

where $\mathcal{X} \subseteq \mathbb{R}^m$. This is possible if the likelihood is Gaussian, but it is typically computationally intractable for most other likelihood functions. However, in the groundbreaking paper of Rue et al. (2009), a novel inference method is proposed that makes it possible to perform fast approximate inference for latent Gaussian random fields with truly high-dimensional latent fields. Their method is based on integrated nested Laplace approximations (INLA) which make it possible to quickly compute numerical approximations to the joint posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$, and to the marginal posterior distributions $\pi(\theta_i \mid \boldsymbol{y})$ and $\pi(x_j \mid \boldsymbol{y})$, with $i = 1, 2, \ldots, n_{\boldsymbol{\theta}}$ and $j = 1, 2, \ldots, m$, where θ_i is the ith element of $\boldsymbol{\theta}$, and $n_{\boldsymbol{\theta}}$ is the cardinality of $\boldsymbol{\theta}$.

The joint posterior distribution of $\boldsymbol{\theta}$ is approximated as

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) = C \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})} \Big|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})} \approx C \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{\pi}_{\mathrm{G}}(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})} \Big|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})},$$

where $\tilde{\pi}_{\mathrm{G}}(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation of the full conditional distribution of $\boldsymbol{x}$, $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the maximiser of $\tilde{\pi}_{\mathrm{G}}(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$ and C is a normalising constant. C can be approximated by numerically integrating over $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ and setting the result equal to 1. This is not too computationally demanding if the dimension of $\boldsymbol{\theta}$ is small. Given the approximation for the joint posterior distribution of $\boldsymbol{\theta}$, the marginal posterior distributions of $\boldsymbol{\theta}$ can be approximated by numerical integration over $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$. Finally, Rue et al. (2009) propose multiple different methods for approximating the marginal posteriors of $\boldsymbol{x}$ with varying degrees of correctness and computational cost. The simplest method is to integrate out $\boldsymbol{\theta}$ from the Gaussian approximation,

$$\pi(x_j \mid \boldsymbol{y}) = \int \pi(x_j \mid \boldsymbol{\theta}, \boldsymbol{y}) \mathrm{d}\boldsymbol{\theta} \approx \int \tilde{\pi}_{\mathrm{G}}(x_j \mid \boldsymbol{\theta}, \boldsymbol{y}) \mathrm{d}\boldsymbol{\theta}, \tag{5}$$

where $\tilde{\pi}_{\mathrm{G}}(x_j \mid \boldsymbol{\theta}, \boldsymbol{y})$ is the marginal distribution of x_j from $\tilde{\pi}_{\mathrm{G}}(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$. However, this method can cause considerable bias in the approximate posterior of x_j (Rue & Martino, 2007). The other alternatives are based on approximating $\pi(x_j \mid \boldsymbol{\theta}, \boldsymbol{y})$ using another layer of Laplace approximations,

or simplified versions thereof. These methods are more computationally demanding, but are found to outperform the Gaussian approximation.

The INLA method is designed to be really fast if the latent Gaussian model has a sparse precision matrix. The Gaussian approximation, $\tilde{\pi}_{\text{G}}(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$, of the full conditional distribution of $\boldsymbol{x}$ has precision matrix $\boldsymbol{Q}+\text{diag}(\boldsymbol{c})$ and mean vector given as the solution of the equation $(\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))\boldsymbol{\mu} = \boldsymbol{b}$, where $\text{diag}(\boldsymbol{c})$ is a diagonal matrix containing the elements of $\boldsymbol{c}$, and $\boldsymbol{b}$ and $\boldsymbol{c}$ are vectors that depend on the likelihood $\pi(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta})$. All necessary computations for approximating the posterior $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ are therefore based on computing determinants for matrices on the form $\boldsymbol{Q} + \text{diag}(\boldsymbol{c})$, and on solving linear systems involving $\boldsymbol{Q}$. These computations become much faster if $\boldsymbol{Q}$ is sparse. This is why, as previously mentioned, the SPDE approximation, which produces realistic and sparse precision matrices for Gaussian random fields, works particularly well in combination with INLA.

The INLA method has been implemented in the `R` package `R-INLA` (Lindgren & Rue, 2015; Rue et al., 2017; Bakka et al., 2018; van Niekerk et al., 2021, 2023), which allows for a fast and easy way of performing inference. The package ships with a large selection of readily implemented likelihood functions. It also provides implementations of design matrices and latent Gaussian fields that can describe different types of regression coefficients, splines, block effects, spatial effects, temporal effects and spatio-temporal effects, all using sparse precision matrices. Recently, van Niekerk and Rue (2021) proposed a new method for approximating the marginal posterior distributions of the latent field $\boldsymbol{x}$. Their method has far-reaching consequences that allow for considerable improvements on how to describe the latent field when using INLA and that yields faster computations of both the joint posterior $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ and the marginal posteriors $\pi(x_j \mid \boldsymbol{y})$, for $j = 1, \ldots, m$. This INLA formulation has been implemented within `R-INLA` and is expected to become the default method in the near future (van Niekerk et al., 2023).

In Papers 2, 3 and 4, we use `R-INLA` for performing inference with our latent Gaussian models. In Papers 3 and 4, we also demonstrate how to implement specialised model components in `R-INLA` that are tailored for specific applications. Originally, the likelihood of the latent Gaussian model should be log-concave for INLA to perform well, but as we show in Paper 2, it is also possible to perform successful inference with probability distributions that are not log-concave, although this requires care to work well.

5 Discussion and conclusion

In the four papers in this thesis, we have developed statistical models for diurnal temperature range and extreme precipitation, with the overall aim

of aiding in the development of improved flood risk assessment methods. The first paper targets modelling of diurnal temperature range, the second paper deals with estimation of large return levels for sub-daily precipitation, and the two last papers focus on modelling the spatial behaviour of hourly precipitation extremes. Methodologically, all four papers focus on statistical modelling of complex phenomena that have unusual and intricate properties, using smart reparametrisations, robust estimators, computationally efficient inference methods and modelling assumptions with various degrees of credibility, in order to achieve fast and robust inference for the properties that we care the most about.

Much of our work has taken place within the field of extreme value theory, where we have attempted to develop models that can bridge the gap between the asymptotic and the subasymptotic, and that perform well in high-dimensional settings. We have focused on models that allow for more robust and computationally efficient inference, two properties we believe to be important in practice, by sacrificing certain theoretical properties, which we believe to be important, but slightly less so than the robustness and computational efficiency. In this way, we believe that we have contributed to the development of statistical models that can be more useful in practice, while staying as true to the theory as possible.

We have developed models for temperature and precipitation to achieve a better understanding of floods. However, it is not temperature or precipitation alone that cause floods, but a combination of the two. Thus, given more time, I would want to build upon all of our work, and use it to examine how precipitation and temperature affect each other, and how they affect river flow and the risk of floods.

In our work, we have mainly focused on spatial modelling, and we have put less effort into describing the temporal properties of our data. However, the temporal aspect can provide important information about the data, and future work should therefore go into performing spatio-temporal modelling of both temperature and extreme precipitation. Especially in Paper 4, I believe that more temporal components, to describe the temporal extent of a storm as well as its spatial extent, could have provided considerable improvements to our model framework.

In Paper 3, we apply a post-hoc transformation of the posterior distributions to account for possible model misspecification. The transformation is fast and works well, but it requires computing gradients of the marginal log-likelihood with respect to the model hyperparameters. For a latent Gaussian model, this is only computationally tractable if the latent field is low-dimensional, or if the likelihood is Gaussian. For this reason, the same post-hoc transformation is not applied in Paper 4, where we apply different latent Gaussian models with high-dimensional latent fields and Bernoulli, gamma and generalised Pareto likelihoods. However, similarly to how INLA

makes inference computationally tractable for these models, I believe that the gradients of these marginal log-likelihoods might be approximated well using nested Laplace approximations. This would be an interesting topic for future research, which might have the potential to produce an automated method of accounting for model misspecification within the `R-INLA` package, or a similar framework.

All of the subasymptotic extreme value models mentioned in Section 2 rely on performing inference using only the data that are extreme in some sense. Thus, inference always relies on some kind of threshold, either for removing or censoring contributions from non-extreme observations (Huser et al., 2016), or as the conditioning threshold of the conditional extremes model. Inference efficiency therefore relies heavily on the threshold choice, with lower thresholds yielding more data but also possibly more model misspecification. It might be that some of the extreme value models have better "threshold convergence" properties than the others, in the sense that a specific low threshold τ_0 might yield less misspecification for, e.g., the scale-mixture models, than for the other models. As an example of this, in the case study of Paper 3 we are forced to choose a very large threshold for the spatial conditional extremes model, which results in few available observations at each conditioning site. Therefore, in Paper 4, we propose a novel form for the standardising functions of the model, which lets us considerably lower the conditioning threshold and perform more efficient inference. Thus, there can clearly be large differences in threshold convergence properties, also within a certain class of extreme value models. To the best of my knowledge, the threshold convergence properties of the aforementioned extreme value models have never been examined in detail or compared against each other. This would be an interesting topic for further research.

6 Paper summaries

Paper 1

We propose the five-parameter lambda distribution as a statistical model for the marginal distribution of diurnal temperature range. Our model takes inspiration from extreme value theory and is able to describe both the bulk and tails of the data. To achieve faster and more robust inference, we propose a reparametrisation of the five-parameter lambda distribution that makes it easier to choose appropriate initial values, and that constrains the possible parameter values in a way that is suitable for our specific application. Three competing inference methods are proposed and evaluated in a simulation study and in a case study. Of these, we prefer the method of quantiles, which estimates parameters by minimising the distance between empirical quantiles and model quantiles, due to its speed, simplicity and good performance. We also propose a distributional quantile regression model for describing the marginal distribution of diurnal temperature range using a set of explanatory variables. The model is divided into two steps: First, we model multiple different distributional quantiles with quantile regression and a set of relevant explanatory variables. Then, at each location of interest, the full marginal distribution of diurnal temperature range is estimated by fitting the five-parameter lambda distribution to the set of estimated quantiles. This "smooths" the quantile regression estimates and avoids the common problem of quantile crossing, which occurs when multiple different quantiles are estimated using independent quantile regression models. The distributional quantile regression model is developed in a way that allows for fast and easy inference using the method of quantiles. We apply our two proposed models for modelling diurnal temperature range from a set of weather stations in the south of Norway, and compare the five-parameter lambda distribution with several competing statistical distributions. The five-parameter lambda distribution provides a good fit to the available data, and is found to outperform the competing distributions. The model fits are evaluated using the probability integral transform (PIT) and the continuous ranked probability score (CRPS), and we find that they provide good fits to the data.

Paper 2

We propose a latent Gaussian model with a bGEV likelihood, for the distribution of the yearly maxima of sub-daily precipitation observations. Fast approximate inference is achieved using INLA. The bGEV likelihood allows for more robust and more numerically stable inference than a GEV likelihood, which we demonstrate with a small simulation study. This is especially true when modelling block maxima in space using explanatory variables. Neither the GEV distribution nor the bGEV distribution are log-concave, which is a requirement for using INLA, but we develop a standardisation technique which we find to yield numerically stable inference with the bGEV distribution. To strengthen inference, by borrowing information from neighbouring locations, we include a spatial Gaussian random field in the latent field of the model. This Gaussian random is field created using the SPDE approach, to achieve faster inference. In practical regression settings, we find that inference using small amounts of data can make it difficult to model both the location and scale parameters of the bGEV as varying in space. We therefore propose a novel two-step estimation procedure that allows for less wasteful and more stable inference, and that makes it easier to let both parameters vary in space. The two-step method extracts additional information from threshold exceedance data when modelling block maxima data. This is achieved by first modelling the standard deviation of threshold exceedance data, and then standardising the block maxima data using the estimated standard deviations. The standardised block maxima are then modelled with the bGEV distribution. Proper uncertainty estimation is achieved by propagating uncertainty from the threshold exceedance step into the block maxima step using bootstrapping. We evaluate model fits using a novel scaled and threshold weighted version of the CRPS, which mainly focuses on model performance in the upper distributional tails. Our two models are applied for estimating return levels of sub-daily precipitation data from a set of weather stations in the south of Norway, and we find the resulting model fits to provide promising results. Model fits based on the two-step model always outperform model fits where trends in both the location and the scale parameters are estimated simultaneously.

Paper 3

We develop a general statistical methodology for modelling high-dimensional spatial extremes using the spatial conditional extremes model, while performing robust and computationally efficient inference with INLA. We build upon a latent Gaussian model variant of the spatial conditional extremes model proposed by Simpson et al. (2023), which allows for fast and easy modelling, using INLA and the SPDE approximation, with the `R-INLA` package. Using the recently developed `cgeneric` component of the `R-INLA` package, we create a general methodology for implementing specialised variants of the spatial conditional extremes model with `R-INLA`. To achieve this, we also propose a nonstationary variation of the SPDE approximation that is more suitable for the spatial conditional extremes model. Extreme value models are typically based on strong asymptotically motivated assumptions, which can lead to a certain degree of model misspecification when describing finite amounts of data. This can be particularly troublesome for Bayesian inference methods, which are ill-suited for tackling model misspecification. We therefore extend upon the post hoc transformation method of Shaby (2014), which yields more robust Bayesian inference by affinely transforming samples from the posterior distribution to get a more reasonable variance. We propose new methods for estimating the scaling factors of the post hoc transformation method, and we also propose an improvement of the method, based on transforming the prior distribution before performing inference. The spatial conditional extremes model can produce considerably more efficient inference by employing the composite likelihood method of Wadsworth and Tawn (2022). However, this leads to considerable model misspecification and has therefore, to the best of our knowledge, never previously been attempted while performing Bayesian inference. We show how the post hoc transformation accounts for the misspecification from the composite likelihood, and thus allows for more efficient inference. Our methodology is applied for modelling both high-dimensional spatial extremes, in a simulation study, and gridded high-density observations of hourly precipitation from a weather radar in Central Norway, in a case study. The model fits are evaluated visually and using the logarithmic score, and found to provide good fits to the available data. Furthermore, the post hoc adjusted model fits are always found to outperform their unadjusted counterparts.

In working on this paper, we developed an unpublished note about the lack of self-consistency of the conditional extremes model, which we later decided to not include into the paper, to keep it more streamlined and focused. However, the note is still quite interesting, in my opinion, so I wanted to include it in the thesis. The note is found right after the contents of Paper 3.

Paper 4

We build upon our work from Paper 3, for developing a complete framework for modelling and simulating high-dimensional spatial precipitation extremes with the spatial conditional extremes model. In Paper 3, a simplified procedure is applied for standardising the marginals to Laplace distributions. Additionally, in the case study of Paper 3 we only focus on modelling nonnegative precipitation, as there is a bulk at zero in the marginal distribution of hourly precipitation. In Paper 4, however, we develop a flexible and computationally efficient framework for spatio-temporal modelling of the marginal distributions of precipitation data to ensure a proper standardisation to Laplace marginals. The framework is based upon a combination of two different latent Gaussian models with a gamma likelihood and a generalised Pareto likelihood, for accurate modelling of the bulk and the upper tail of the marginal distributions of nonnegative precipitation observations, respectively. For modelling the presence or absence of precipitation, we propose multiple competing binary spatial models, all aimed at capturing different properties of the presence/absence process. In the case study of Paper 3, we struggle with a model choice that makes it necessary to choose a very high threshold for the conditional extremes model, thus leading to less efficient inference. In Paper 4, we develop new empirical diagnostics for selecting a suitable model, and we use these diagnostics for proposing new parametric forms for the standardisation functions of the spatial conditional extremes model, that allow us to considerably lower the extreme threshold and perform more efficient inference. The marginal models, presence/absence models and the spatial conditional extremes model are all combined in a case study for simulating from the high-density spatial distribution of extreme hourly precipitation observation over a water catchment in Central Norway, using data from the same weather radar as in Paper 3. The resulting simulations display much promise, and capture many important properties of the data with high performance.

References

Abu-Awwad, A., Maume-Deschamps, V., & Ribereau, P. (2020). Fitting spatial max-mixture processes with unknown extremal dependence class: An exploratory analysis tool. *TEST*, *29*(2), 479–522. https://doi.org/10.1007/s11749-019-00663-5

Ahmed, M., Maume-Deschamps, V., Ribereau, P., & Vial, C. (2020). Spatial risk measures for max-stable and max-mixture processes. *Stochastics*, *92*(7), 1005–1020. https://doi.org/10.1080/17442508.2019.1687703

Allan, R. P., Barlow, M., Byrne, M. P., Cherchi, A., Douville, H., Fowler, H. J., Gan, T. Y., Pendergrass, A. G., Rosenfeld, D., Swann, A. L. S., Wilcox, L. J., & Zolina, O. (2020). Advances in understanding large-scale responses of the water cycle to climate change. *Annals of the New York Academy of Sciences*, *1472*(1), 49–75. https://doi.org/10.1111/nyas.14337

Asadi, P., Davison, A. C., & Engelke, S. (2015). Extremes on river networks. *The Annals of Applied Statistics*, *9*(4), 2023–2050. https://doi.org/10.1214/15-AOAS863

Bacro, J.-N., Gaetan, C., & Toulemonde, G. (2016). A flexible dependence model for spatial extremes. *Journal of Statistical Planning and Inference*, *172*, 36–52. https://doi.org/10.1016/j.jspi.2015.12.002

Bacro, J.-N., Gaetan, C., Opitz, T., & Toulemonde, G. (2020). Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *Journal of the American Statistical Association*, *115*(530), 555–569. https://doi.org/10.1080/01621459.2019.1617152

Bader, B., Yan, J., & Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics*, *12*(1), 310–329. https://doi.org/10.1214/17-AOAS1092

Bakka, H. (2022). How to solve the stochastic partial differential equation that gives a Matérn random field using the finite element method. https://doi.org/10.48550/arXiv.1803.03765

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., & Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *WIREs Computational Statistics*, *10*(6), e1443. https://doi.org/10.1002/wics.1443

Balkema, A. A., & Resnick, S. I. (1977). Max-infinite divisibility. *Journal of Applied Probability*, *14*(2), 309–319. https://doi.org/10.2307/3213001

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC. https://doi.org/10.1201/b17115

Banerjee, S., Gelfand, A. E., Finley, A. O., & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(4), 825–848. https://doi.org/10.1111/j.1467-9868.2008.00663.x

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. L. (2004). *Statistics of extremes: Theory and applications*. John Wiley & Sons.

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The relative importance of different flood-generating mechanisms across

Europe. *Water Resources Research*, *55*(6), 4582–4593. https://doi.org/10.1029/2019WR024841

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 192–225. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x

Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *24*(3), 179–195. https://doi.org/10.2307/2987782

Bopp, G. P., Shaby, B. A., Forest, C. E., & Mejía, A. (2020). Projecting flood-inducing precipitation with a Bayesian analogue model. *Journal of Agricultural, Biological and Environmental Statistics*, *25*(2), 229–249. https://doi.org/10.1007/s13253-020-00391-6

Bopp, G. P., Shaby, B. A., & Huser, R. (2021). A hierarchical max-infinitely divisible spatial model for extreme precipitation. *Journal of the American Statistical Association*, *116*(533), 93–106. https://doi.org/10.1080/01621459.2020.1750414

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. http://www.jstor.org/stable/2286841

Brown, B. M., & Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, *14*(4), 732–739. https://doi.org/10.2307/3213346

Brunner, M. I., Slater, L., Tallaksen, L. M., & Clark, M. (2021). Challenges in modeling and predicting floods and droughts: A review. *WIREs Water*, *8*(3), e1520. https://doi.org/10.1002/wat2.1520

Bücher, A., & Segers, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, *20*(4), 839–872. https://doi.org/10.1007/s10687-017-0292-6

Bücher, A., & Zhou, C. (2021). A Horse Race Between the Block Maxima Method and the Peak-Over-Threshold Approach. *Statistical Science*, *36*(3), 360–378. https://doi.org/10.1214/20-STS795

Cameletti, M., Lindgren, F., Simpson, D., & Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, *97*(2), 109–131. https://doi.org/10.1007/s10182-012-0196-3

Castro-Camilo, D., Huser, R., & Rue, H. (2019). A spliced gamma-generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(3), 517–534. https://doi.org/10.1007/s13253-019-00369-z

Castro-Camilo, D., Huser, R., & Rue, H. (2022). Practical strategies for generalized extreme value-based regression models for extremes. *Environmetrics*, *33*(6), e2742. https://doi.org/10.1002/env.2742

Chandler, R. E., & Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, *94*(1), 167–183. https://doi.org/10.1093/biomet/asm015

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer, London. https://doi.org/10.1007/978-1-4471-3675-0

Coles, S., Heffernan, J., & Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, *2*(4), 339–365. https://doi.org/10.1023/A:1009963131610

Coles, S., & Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(2), 377–392. https://doi.org/10.1111/j.2517-6161.1991.tb01830.x

Cooley, D., Nychka, D., & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, *102*(479), 824–840. https://doi.org/10.1198/016214506000000780

Cox, D. R., & Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, *91*(3), 729–737. https://doi.org/10.1093/biomet/91.3.729

Cressie, N., & Johannesson, G. (2008). Fixed rank Kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 209–226. https://doi.org/10.1111/j.1467-9868.2007.00633.x

Cressie, N., & Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Davison, A. C., & Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application*, *2*(1), 203–235. https://doi.org/10.1146/annurev-statistics-010814-020133

Davison, A. C., Huser, R., & Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, *45*(5), 511–529. https://doi.org/10.1007/s11004-013-9469-y

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, *27*(2), 161–186. https://doi.org/10.1214/11-STS376

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(3), 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

Davison, A. C., & Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *468*(2138), 581–608. https://doi.org/10.1098/rspa.2011.0412

Davison, A. C., Huser, R., & Thibaud, E. (2019). Spatial extremes. In A. E. Gelfand, M. Fuentes, J. A. Hoeting, & R. L. Smith (Eds.), *Handbook of environmental and ecological statistics* (pp. 711–744). Chapman; Hall/CRC.

De Fondeville, R., & Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, *105*(3), 575–592. https://doi.org/10.1093/biomet/asy026

De Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, *12*(4), 1194–1204. https://doi.org/10.1214/aop/1176993148

Diggle, P. J., Tawn, J. A., & Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *47*(3), 299–350. https://doi.org/10.1111/1467-9876.00113

Dombry, C., & Ribatet, M. (2015). Functional regular variations, pareto processes and peaks over threshold. *Statistics and its Interface*, *8*(1), 9–17. https://doi.org/10.4310/SII.2015.v8.n1.a2

Eidsvik, J., Shaby, B. A., Reich, B. J., Wheeler, M., & Niemi, J. (2014). Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, *23*(2), 295–315. https://doi.org/10.1080/10618600.2012.760460

Engelke, S., & Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. https://doi.org/10.1111/rssb.12355

Engelke, S., Opitz, T., & Wadsworth, J. (2019). Extremal dependence of random scale constructions. *Extremes*, *22*(4), 623–666. https://doi.org/10.1007/s10687-019-00353-3

Ferreira, A., & de Haan, L. (2014). The generalized pareto process; with a view towards application and simulation. *Bernoulli*, *20*(4), 1717–1737. https://doi.org/10.3150/13-BEJ538

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190. https://doi.org/10.1017/S0305004100015681

Forster, C., & Oesting, M. (2022). Non-stationary max-stable models with an application to heavy rainfall data. https://doi.org/10.48550/arXiv.2212.11598

Fox, E. B., & Dunson, D. B. (2015). Bayesian nonparametric covariance regression. *Journal of Machine Learning Research*, *16*(77), 2501–2542.

Furrer, R., Genton, M. G., & Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, *15*(3), 502–523. https://doi.org/10.1198/106186006X132178

Gamet, P., & Jalbert, J. (2022). A flexible extended generalized Pareto distribution for tail estimation. *Environmetrics*, *33*(6), e2744. https://doi.org/10.1002/env.2744

Gelfand, A. E., Diggle, P. J., Fuentes, M., & Guttorp, P. (2010). *Handbook of spatial statistics*. CRC press. https://doi.org/10.1201/9781420072884

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman; Hall/CRC. https://doi.org/10.1201/b16018

Gilchrist, W. (2000). *Statistical modelling with quantile functions*. Chapman; Hall/CRC. https://doi.org/10.1201/9781420035919

Giné, E., Hahn, M. G., & Vatan, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probability Theory and Related Fields*, *87*(2), 139–165. https://doi.org/10.1007/BF01198427

Gumbel, E. J. (1960). Distributions des valeurs extremes en plusiers dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, *9*, 171–173.

Guttorp, P., & Gneiting, T. (2006). Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, *93*(4), 989–995. https://doi.org/10.1093/biomet/93.4.989

Hanssen-Bauer, I., Drange, H., Førland, E., Roald, L., Børsheim, K., Hisdal, H., Lawrence, D., Nesje, A., Sandven, S., Sorteberg, A., et al. (2015). Klima i Norge 2100 [Kunnskapsgrunnlag for klimatilpasning oppdatert i 2015, Norsk klimasenter, Oslo, Norway].

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F.,

Nychka, D. W., Sun, F., & Zammit-Mangion, A. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(3), 398–425. https://doi.org/10.1007/s13253-018-00348-w

Heffernan, J. E., & Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, *17*(2), 537–571. https://doi.org/10.1214/105051606000000835

Heffernan, J. E., & Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 497–546. https://doi.org/10.1111/j.1467-9868.2004.02050.x

Higdon, D. (2002). Space and space-time modeling using process convolutions. In C. W. Anderson, V. Barnett, P. C. Chatwin, & A. H. El-Shaarawi (Eds.), *Quantitative methods for current environmental issues* (pp. 37–56). Springer London.

Huser, R., & Davison, A. C. (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(2), 439–461. https://doi.org/10.1111/rssb.12035

Huser, R., Davison, A. C., & Genton, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes*, *19*(1), 79–103. https://doi.org/10.1007/s10687-015-0230-4

Huser, R., Opitz, T., & Thibaud, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics*, *21*, 166–186. https://doi.org/10.1016/j.spasta.2017.06.004

Huser, R., Opitz, T., & Thibaud, E. (2021). Max-infinitely divisible models and inference for spatial extremes. *Scandinavian Journal of Statistics*, *48*(1), 321–348. https://doi.org/10.1111/sjos.12491

Huser, R., & Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, *114*(525), 434–444. https://doi.org/10.1080/01621459.2017.1411813

Huser, R., & Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews (WIREs): Computational Statistics*, *14*(1), e1537. https://doi.org/10.1002/wics.1537

IPCC. (2023). *Climate change 2023: Synthesis report. a report of the intergovernmental panel on climate change. contribution of working groups I, II and III to the sixth assessment report of the Intergovernmental Panel on Climate Change* (Core Writing Team, H. Lee, & J. Romero, Eds.).

Jongman, B. (2018). Effective adaptation to rising flood risk. *Nature Communications*, *9*(1), 1986. https://doi.org/10.1038/s41467-018-04396-1

Kabluchko, Z., & Schlather, M. (2010). Ergodic properties of max-infinitely divisible processes. *Stochastic Processes and their Applications*, *120*(3), 281–295. https://doi.org/10.1016/j.spa.2009.12.002

Katzfuss, M., & Guinness, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statistical Science*, *36*(1), 124–141. https://doi.org/10.1214/19-STS755

Kaufman, C. G., Schervish, M. J., & Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the*

American Statistical Association, *103*(484), 1545–1555. https://doi.org/10.1198/016214508000000959

Keef, C., Tawn, J. A., & Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, *24*(1), 13–21. https://doi.org/10.1002/env.2190

Kereszturi, M., Tawn, J. A., & Jonathan, P. (2016). Assessing extremal dependence of North Sea storm severity. *Ocean Engineering*, *118*, 242–259. https://doi.org/10.1016/j.oceaneng.2016.04.013

Kleijn, B., & van der Vaart, A. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electronic Journal of Statistics*, *6*, 354–381. https://doi.org/10.1214/12-EJS675

Koch, E., Koh, J., Davison, A. C., Lepore, C., & Tippett, M. K. (2021). Trends in the extremes of environments associated with severe U.S. thunderstorms. *Journal of Climate*, *34*(4), 1259–1272. https://doi.org/10.1175/JCLI-D-19-0826.1

Koh, J., Pimont, F., Dupuy, J.-L., & Opitz, T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *The Annals of Applied Statistics*, *17*(1), 560–582. https://doi.org/10.1214/22-AOAS1642

Koutsoyiannis, D., Kozonis, D., & Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, *206*(1), 118–135. https://doi.org/10.1016/S0022-1694(98)00097-3

Langousis, A., Mamalakis, A., Puliga, M., & Deidda, R. (2016). Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research*, *52*(4), 2659–2681. https://doi.org/10.1002/2015WR018502

Ledford, A. W., & Tawn, J. A. (1996). Statistics for Near Independence in Multivariate Extreme Values. *Biometrika*, *83*(1), 169–187. https://doi.org/10.1093/biomet/83.1.169

Ledford, A. W., & Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*(2), 475–499. https://doi.org/10.1111/1467-9868.00080

Lenzi, A., Bessac, J., Rudi, J., & Stein, M. L. (2023). Neural networks for parameter estimation in intractable models. *Computational Statistics & Data Analysis*, *185*, 107762. https://doi.org/10.1016/j.csda.2023.107762

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1–25. https://doi.org/10.18637/jss.v063.i19

Lindgren, F., Rue, H., & 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. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary mathematics*, *80*(1), 221–239.

Liu, Y., & Tawn, J. A. (2014). Self-consistent estimation of conditional multivariate extreme value distributions. *Journal of Multivariate Analysis*, *127*, 19–35. https://doi.org/10.1016/j.jmva.2014.02.003

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., & Andersen, J. (2018). SeNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. *Earth System Science Data*, *10*(1), 235. https://doi.org/10.5194/essd-10-235-2018

Lussana, C., Tveito, O. E., & Uboldi, F. (2018). Three-dimensional spatial interpolation of 2 m temperature over Norway. *Quarterly Journal of the Royal Meteorological Society*, *144*(711), 344–364. https://doi.org/10.1002/qj.3208

Majumder, R., & Reich, B. J. (2023). A deep learning synthetic likelihood approximation of a non-stationary spatial model for extreme streamflow forecasting. *Spatial Statistics*, *55*, 100755. https://doi.org/10.1016/j.spasta.2023.100755

Maraun, D., & Widmann, M. (2018). *Statistical downscaling and bias correction for climate research*. Cambridge University Press. https://doi.org/10.1017/9781107588783

Matern, B. (1986). *Spatial variation* (2nd ed., Vol. 36). Springer New York. https://doi.org/10.1007/978-1-4615-7892-5

Naveau, P., Huser, R., Ribereau, P., & Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, *52*(4), 2753–2769. https://doi.org/10.1002/2015WR018552

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., & Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, *24*(2), 579–599. https://doi.org/10.1080/10618600.2014.914946

OEIS Foundation. (2023). *The on-line encyclopedia of integer sequences*. https://oeis.org/

Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, *122*, 409–413. https://doi.org/10.1016/j.jmva.2013.08.008

Opitz, T. (2016). Modeling asymptotically independent spatial extremes based on Laplace random fields. *Spatial Statistics*, *16*, 1–18. https://doi.org/10.1016/j.spasta.2016.01.001

Opitz, T., Huser, R., Bakka, H., & Rue, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, *21*(3), 441–462. https://doi.org/10.1007/s10687-018-0324-x

Padoan, S. A., Ribatet, M., & Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, *105*(489), 263–277. https://doi.org/10.1198/jasa.2009.tm08577

Padoan, S. A. (2013). Extreme dependence models based on event magnitude. *Journal of Multivariate Analysis*, *122*, 1–19. https://doi.org/10.1016/j.jmva.2013.07.009

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, *49*(1), 187–201. https://doi.org/10.1029/2012WR012557

Papastathopoulos, I., & Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, *143*(1), 131–143. https://doi.org/10.1016/j.jspi.2012.07.001

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131. https://doi.org/10.1214/aos/1176343003

Rai, S., Hoffman, A., Lahiri, S., Nychka, D. W., Sain, S. R., & Bandyopadhyay, S. (2023). Fast parameter estimation of generalized extreme value distribution using neural networks. https://doi.org/10.48550/arXiv.2305.04341

Razavi, S., Gober, P., Maier, H. R., Brouwer, R., & Wheater, H. (2020). Anthropocene flooding: Challenges for science and society. *Hydrological Processes*, *34*(8), 1996–2000. https://doi.org/10.1002/hyp.13723

Reich, B. J., & Shaby, B. A. (2019). A spatial Markov model for climate extremes. *Journal of Computational and Graphical Statistics*, *28*(1), 117–126. https://doi.org/10.1080/10618600.2018.1482764

Resnick, S. I. (2008). *Extreme values, regular variation, and point processes* (Vol. 4). Springer Science & Business Media. https://doi.org/10.1007/978-0-387-75953-1

Ribatet, M., Cooley, D., & Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, *22*(2), 813–845. http://www.jstor.org/stable/24310036

Richards, J., & Huser, R. (2022). Regression modelling of spatiotemporal extreme U.S. wildfires via partially-interpretable neural networks. https://doi.org/10.48550/arXiv.2208.07581

Richards, J., Sainsbury-Dale, M., Zammit-Mangion, A., & Huser, R. (2023). Likelihood-free neural Bayes estimators for censored inference with peaks-over-threshold models. https://doi.org/10.48550/arXiv.2306.15642

Richards, J., Tawn, J. A., & Brown, S. (2022). Modelling extremes of spatial aggregates of precipitation using conditional methods. *The Annals of Applied Statistics*, *16*(4), 2693–2713. https://doi.org/10.1214/22-AOAS1609

Richards, J., Tawn, J. A., & Brown, S. (2023). Joint estimation of extreme spatially aggregated precipitation at different scales through mixture modelling. *Spatial Statistics*, *53*, 100725. https://doi.org/10.1016/j.spasta.2022.100725

Rootzén, H., Segers, J., & L. Wadsworth, J. (2018). Multivariate peaks over thresholds models. *Extremes*, *21*(1), 115–145. https://doi.org/10.1007/s10687-017-0294-4

Rootzén, H., Segers, J., & Wadsworth, J. L. (2018). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, *165*, 117–131. https://doi.org/10.1016/j.jmva.2017.12.003

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

Rue, H., & Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models [Special Issue: Bayesian Inference for Stochastic Processes]. *Journal of Statistical Planning and Inference*, *137*(10), 3177–3192. https://doi.org/10.1016/j.jspi.2006.07.016

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.

Journal of the Royal Statistical Society: Series B (Statistical Methodology), *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, *4*(1), 395–421. https://doi.org/10.1146/annurev-statistics-060116-054045

Sainsbury-Dale, M., Zammit-Mangion, A., & Huser, R. (2023). Neural point estimation for fast optimal likelihood-free inference. https://doi.org/10.48550/arXiv.2208.12942

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, *5*(1), 33–44. https://doi.org/10.1023/A:1020977924878

Shaby, B. A. (2014). The open-faced sandwich adjustment for MCMC using estimating functions. *Journal of Computational and Graphical Statistics*, *23*(3), 853–876. https://doi.org/10.1080/10618600.2013.842174

Shooter, R., Ross, E., Tawn, J., & Jonathan, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics*, *30*(6), e2562. https://doi.org/10.1002/env.2562

Shooter, R., Ross, E., Ribal, A., Young, I. R., & Jonathan, P. (2022). Multivariate spatial conditional extremes for extreme ocean environments. *Ocean Engineering*, *247*, 110647. https://doi.org/10.1016/j.oceaneng.2022.110647

Sibuya, M., et al. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, *11*(2), 195–210.

Simpson, E. S., Opitz, T., & Wadsworth, J. L. (2023). High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. *Extremes*. https://doi.org/10.1007/s10687-023-00468-8

Simpson, E. S., & Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, *41*, 100482. https://doi.org/10.1016/j.spasta.2020.100482

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, *72*(1), 67–90. https://doi.org/10.1093/biomet/72.1.67

Smith, R. L. (1990). Max-stable processes and spatial extremes [Unpublished manuscript].

Stein, M. L. (1999). *Interpolation of spatial data: Some theory on Kriging*. Springer New York. https://doi.org/10.1007/978-1-4612-1494-6

Stein, M. L. (2021a). A parametric model for distributions with flexible behavior in both tails. *Environmetrics*, *32*(2), e2658. https://doi.org/10.1002/env.2658

Stein, M. L. (2021b). Parametric models for distributions when interest is in extremes with an application to daily temperature. *Extremes*, *24*(2), 293–323. https://doi.org/10.1007/s10687-020-00378-z

Stein, M. L., Chi, Z., & Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *66*(2), 275–296. https://doi.org/10.1046/j.1369-7412.2003.05512.x

Styan, G. P. (1973). Hadamard products and multivariate statistical analysis. *Linear Algebra and its Applications*, *6*, 217–240. https://doi.org/10.1016/0024-3795(73)90023-2

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, *77*(2), 245–253. https://doi.org/10.1093/biomet/77.2.245

Thibaud, E., Mutzner, R., & Davison, A. C. (2013). Threshold modeling of extreme spatial rainfall. *Water Resources Research*, *49*(8), 4633–4644. https://doi.org/10.1002/wrcr.20329

Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., & Heikkinen, J. (2016). Bayesian inference for the Brown-Resnick process, with an application to extreme low temperatures. *The Annals of Applied Statistics*, *10*(4), 2303–2324. https://doi.org/10.1214/16-AOAS980

van de Vyver, H. (2012). Spatial regression models for extreme precipitation in Belgium. *Water Resources Research*, *48*(9). https://doi.org/10.1029/2011WR011707

van Niekerk, J., Bakka, H., Rue, H., & Schenk, O. (2021). New frontiers in Bayesian modeling using the INLA package in R. *Journal of Statistical Software*, *100*(2), 1–28. https://doi.org/10.18637/jss.v100.i02

van Niekerk, J., Krainski, E., Rustand, D., & Rue, H. (2023). A new avenue for Bayesian inference with INLA. *Computational Statistics & Data Analysis*, *181*, 107692. https://doi.org/10.1016/j.csda.2023.107692

van Niekerk, J., & Rue, H. (2021). Correcting the Laplace method with variational bayes. https://doi.org/10.48550/arXiv.2111.12945

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*(1), 5–42. http://www.jstor.org/stable/24309261

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *50*(2), 297–312. https://doi.org/10.1111/j.2517-6161.1988.tb01729.x

Vettori, S., Huser, R., & Genton, M. G. (2019). Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics*, *75*(3), 831–841. https://doi.org/10.1111/biom.13051

Wadsworth, J. L., & Tawn, J. A. (2012a). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *74*(3), 543–567. https://doi.org/10.1111/j.1467-9868.2011.01017.x

Wadsworth, J. L., & Tawn, J. A. (2012b). Dependence modelling for spatial extremes. *Biometrika*, *99*(2), 253–272. https://doi.org/10.1093/biomet/asr080

Wadsworth, J. L., & Tawn, J. A. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, *51*, 100677. https://doi.org/10.1016/j.spasta.2022.100677

Walchessen, J., Lenzi, A., & Kuusela, M. (2023). Neural likelihood surfaces for spatial processes with computationally intensive or intractable likelihoods. https://doi.org/10.48550/arXiv.2305.04634

Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., & Roberts, N. M. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, *52*(3), 522–555. https://doi.org/10.1002/2014RG000464

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

Yadav, R., Huser, R., & Opitz, T. (2021). Spatial hierarchical modeling of threshold exceedances using rate mixtures. *Environmetrics*, *32*(3), e2662. https://doi.org/10.1002/env.2662

Yin, J., Gentine, P., Zhou, S., Sullivan, S. C., Wang, R., Zhang, Y., & Guo, S. (2018). Large increase in global storm runoff extremes driven by climate and anthropogenic changes. *Nature Communications*, *9*(1), 4389. https://doi.org/10.1038/s41467-018-06765-2

Yin, J., Geng, Z., Li, R., & Wang, H. (2010). Nonparametric covariance model. *Statistica Sinica*, *20*(1), 469–479.

Zhang, L., Risser, M. D., Molter, E. M., Wehner, M. F., & O'Brien, T. A. (2022). Accounting for the spatial structure of weather systems in detected changes in precipitation extremes. *Weather and Climate Extremes*, *38*, 100499. https://doi.org/10.1016/j.wace.2022.100499

Zhang, L., Shaby, B. A., & Wadsworth, J. L. (2022). Hierarchical transformed scale mixtures for flexible modeling of spatial extremes on datasets with many locations. *Journal of the American Statistical Association*, *117*(539), 1357–1369. https://doi.org/10.1080/01621459.2020.1858838

Zhong, P., Huser, R., & Opitz, T. (2022). Modeling nonstationary temperature maxima based on extremal dependence changing with event magnitude. *The Annals of Applied Statistics*, *16*(1), 272–299. https://doi.org/10.1214/21-AOAS1504

Part II

Research Papers

— 1 —

Quantile based modelling of diurnal temperature range with the five-parameter lambda distribution

Silius M. Vandeskog, Thordis L. Thorarinsdottir, Ingelin Steinsland and Finn Lindgren

Environmetrics, 33(4), e2719

Quantile based modelling of diurnal temperature range with the five-parameter lambda distribution

Silius M. Vandeskog[1] Thordis L. Thorarinsdottir[2]
Ingelin Steinsland[1] Finn Lindgren[3]

[1] Norwegian University of Science and Technology (NTNU)
[2] Norwegian Computing Centre
[3] The University of Edinburgh

Abstract

Diurnal temperature range is an important variable in climate science that can provide information regarding climate variability and climate change. Changes in diurnal temperature range can have implications for hydrology, human health and ecology, among others. Yet, the statistical literature on modelling diurnal temperature range is lacking. In this paper we propose to model the distribution of diurnal temperature range using the five-parameter lambda (FPL) distribution. Additionally, in order to model diurnal temperature range with explanatory variables, we propose a distributional quantile regression model that combines quantile regression with marginal modelling using the FPL distribution. Inference is performed using the method of quantiles. The models are fitted to 30 years of daily observations of diurnal temperature range from 112 weather stations in the southern part of Norway. The flexible FPL distribution shows great promise as a model for diurnal temperature range, and performs well against competing models. The distributional quantile regression model is fitted to diurnal temperature range data using geographic, orographic and climatological explanatory variables. It performs well and captures much of the spatial variation in the distribution of diurnal temperature range in Norway.

1 Introduction

In this paper we develop distributional models for diurnal temperature range, which is the difference between daily maximum and minimum temperature. The fourth IPCC assessment report identified diurnal temperature range as a key uncertainty factor (IPCC, 2007), and the fifth report described a substantial knowledge gap surrounding this climate variable (IPCC, 2013).

While more effort has since been made to understand diurnal temperature range, the literature on it is still lacking (Thorne, Donat, et al., 2016; Thorne, Menne, et al., 2016; Sun et al., 2019; IPCC, 2021). Diurnal temperature range can be used as an index of radiative forced climate change, and as a useful index for assessing the output of general circulation models (Braganza et al., 2004). Additionally, it has been shown that diurnal temperature range is linked to human health conditions such as the risk of influenza (Park et al., 2020), risk of stroke (Vered et al., 2020), and overall health and mortality (Cheng et al., 2014; Lim et al., 2015). Changes in diurnal temperature range can also be of large importance within ecology (Henry, 2007; Peng et al., 2013; Vasseur et al., 2014; Kovi et al., 2016) and hydrology (Hanssen-Bauer et al., 2016), among others. Despite these areas of usage, to the best of our knowledge, there do not exist any attempts at statistical modelling of the distribution of diurnal temperature range in the literature. In their recent work of reassessing changes in diurnal temperature range worldwide, Thorne, Donat, et al. (2016) conclude that there is "only *medium confidence* in the magnitude of reductions in diurnal temperature range since 1950" and that "there is *low confidence* in trends and multidecadal variability in diurnal temperature range prior to 1950". Thus, more knowledge about diurnal temperature range is needed.

In this paper we propose to model the marginal distribution of diurnal temperature range with the five-parameter lambda (FPL) distribution (Gilchrist, 2000). The FPL distribution is an extension of the four-parameter generalised lambda distribution (Ramberg & Schmeiser, 1974), itself an extension of Tukey's three-parameter lambda distribution (Tukey, 1962). This family of distributions has seen infrequent use within the statistical literature. Some areas of usage for the FPL distribution have been income modelling (Tarsitano, 2004) and reliability analysis (Ahmadabadi et al., 2012; Nair et al., 2013). The FPL distribution is tightly linked to the generalised Pareto distribution (e.g. Coles, 2001), as its quantile function is equal to the difference between two generalised Pareto quantile functions (see Section 3.1). The generalised Pareto distribution is often used for estimating extremely large or small quantiles, and has been much used for modelling both the upper and lower tails of temperature distributions (e.g. Stein, 2021a; Rohrbeck et al., 2021). Thus, the FPL distribution is a natural choice for modelling the difference between daily maximum and daily minimum temperature if one considers these as extreme upper and lower quantiles in the daily temperature distribution.

The FPL distribution can be used for modelling the distribution of diurnal temperature range in locations with available observations. However, most locations do not contain any available temperature data. Thus, it is also of interest to model diurnal temperature range in locations without daily temperature observations, using a regression model. Most classical

regression models focus on estimating the conditional mean of a distribution, given a set of explanatory variables. However, the distribution of diurnal temperature range is complex, and its variance, skewness and kurtosis vary in space (see Section 2). Thus, a regression model for the mean would not provide enough information about diurnal temperature range to be of much use. An alternative to regression on the mean is quantile regression (Koenker, 2005), where one models a set of conditional quantiles given some explanatory variables. Quantile regression is based on fewer assumptions than mean regression, and it allows for more flexible modelling of complex distributions. However, it can often lead to quantile crossing, meaning that certain combinations of the explanatory variables lead to non-monotonic quantile functions (Bondell et al., 2010; Cannon, 2011; Rodrigues & Fan, 2017). Furthermore, quantile regression can only estimate a finite set of quantiles, and it does not lend an easy way of estimating distributional properties like moments. An alternative that has gained more popularity in recent years is distributional regression, where one attempts to model the entire conditional distribution given a set of explanatory variables (e.g. Klein et al., 2015; Schlosser et al., 2019; Henzi et al., 2021). As stated by Hothorn et al. (2014), this should be the ultimate goal of any regression analysis. However, most distributional regression models can be somewhat complex and computationally demanding.

Here, we propose a conceptually simple and highly parallelisable distributional regression model, based on a combination of quantile regression and marginal modelling with the FPL distribution. Parameter estimation for the FPL distribution is performed using the method of quantiles (e.g. Koenker, 2005), which is based on minimising the distance between a quantile function and a set of estimated quantiles. A thorough description of this estimation method is presented, and it is compared to competing estimation methods for the FPL distribution, both in a simulation study, and using real temperature observations. In order to ease parameter interpretation and numerical inference, a novel reparametrisation of the FPL distribution is developed. The marginal FPL model and the distributional quantile regression model are fitted to thirty years of daily temperature observations from 112 weather stations in southern Norway, including both coastal and inland stations over a large range of altitudes. In order to properly evaluate the model fits, a closed-form expression for the continuous ranked probability score (CRPS; Matheson & Winkler, 1976) with an FPL forecast distribution is developed. This new modelling framework provides a rigorous alternative to analyse diurnal temperature range and its observed variation in time and space compared to current empirical approaches (e.g. Wang et al., 2017; Vinnarasi et al., 2017; Shelton et al., 2021; Sun et al., 2021).

The remainder of the paper is organised as follows. Section 2 introduces daily temperature data from the southern part of Norway, and associated

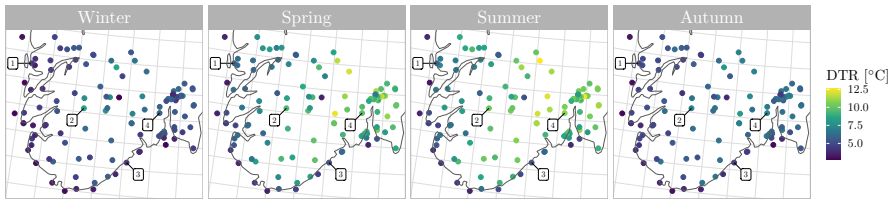


Figure 1: *Empirical season medians of diurnal temperature range (DTR) at the 112 weather stations in our data set. The locations of the four stations from Figures 2, 6 and 8 are presented in all the plots. These are: (1) Flesland, (2) Hovden – Lundane, (3) Lyngør fyr and (4) Sande – Galleberg.*

explanatory variables. Section 3 provides a motivation for the choice of modelling diurnal temperature range with the FPL distribution, and presents some of the properties of the distribution. The distributional quantile regression model is also developed here. In Section 4, the method of quantiles and two other competing inference methods are described, and a closed-form expression for the CRPS with an FPL forecast distribution is developed. In Section 5, a simulation study is performed, where we compare the method of quantiles with two competing inference methods. Finally, we apply our models to Norwegian diurnal temperature range data and evaluate the model fits in Section 6. The paper concludes with a short discussion in Section 7.

2 Data

The analysis in this paper is based on daily time series of air temperature observations from a set of weather stations in southern Norway. The data are openly available from Norwegian Meteorological Institute (2019). For each weather station, daily minimum and maximum temperatures between 18–18 UTC are used to find time series of diurnal temperature range. Data is downloaded from the thirty year time period from 1 January 1989 to 31 December 2018. Two thirds of the weather stations were already established in 1989, and the ages of the remaining stations are almost uniformly distributed between one and thirty years. Some stations are too recently established to be useful for our purposes, and others contain large amounts of missing data. Data cleaning is therefore performed by removing all weather stations that contains less than 180 observations from any of the four seasons of the year (winter: December-February; spring: March-May; summer: June-August; autumn: September-November). By cleaning the data we reduce the number of weather stations from an initial 133 to a

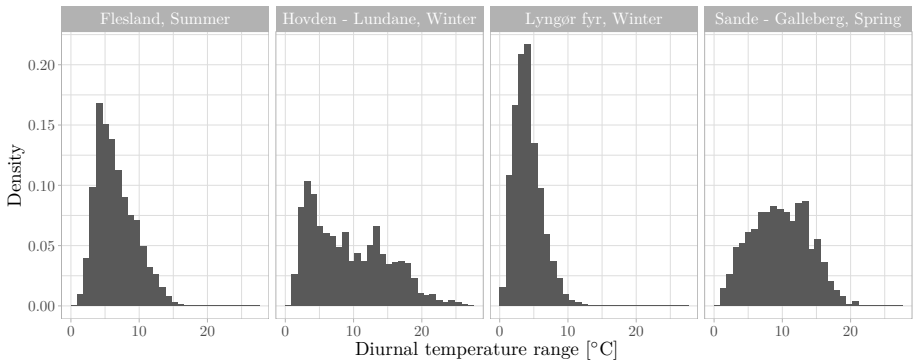


Figure 2: *Histograms of diurnal temperature range for four selected weather stations and seasons.*

new value of 112 stations. The locations of these are displayed in Figure 1 together with median diurnal temperature range for each season. These 112 stations span altitudes from 0 to 1900 meters above sea level, and contains observations from tundra, subarctic and oceanic climates (Kottek et al., 2006).

Our modelling framework does not specifically account for measurement and round-off errors. Obvious errors, resulting in negative diurnal temperature range, are removed from the data. Due to the existence of negative data, it is expected that there also are erroneous data among the positive range values. However, accounting for these errors is outside the scope of this paper.

Six explanatory variables are used for modelling diurnal temperature range: easting, northing, distance to the open sea and altitude, in addition to the historical mean and variance of daily mean temperature at each location. This is estimated using the records of daily temperature observations from each station. Exploratory analysis finds evidence that the marginal distribution of diurnal temperature range can be approximated as being constant within each season (results not shown), but that it varies between the seasons. Historical mean and variance of the daily mean temperature are therefore computed for each season separately. The distance to the open sea is derived from a digital elevation model of Norway, with resolution 50×50 m^2, published by the Norwegian Mapping Authority (https://hoydedata.no). Time series of daily mean temperature, along with longitude, latitude and altitude are freely available from Norwegian Meteorological Institute (2019). Easting and northing are based on UTM 32 coordinates.

Figure 1 shows the median of diurnal temperature range for each season and each weather station. It reveals clear seasonal and spatial patterns. In particular, the values appear higher during spring and summer than during winter and autumn. Similar patterns are also found when examining

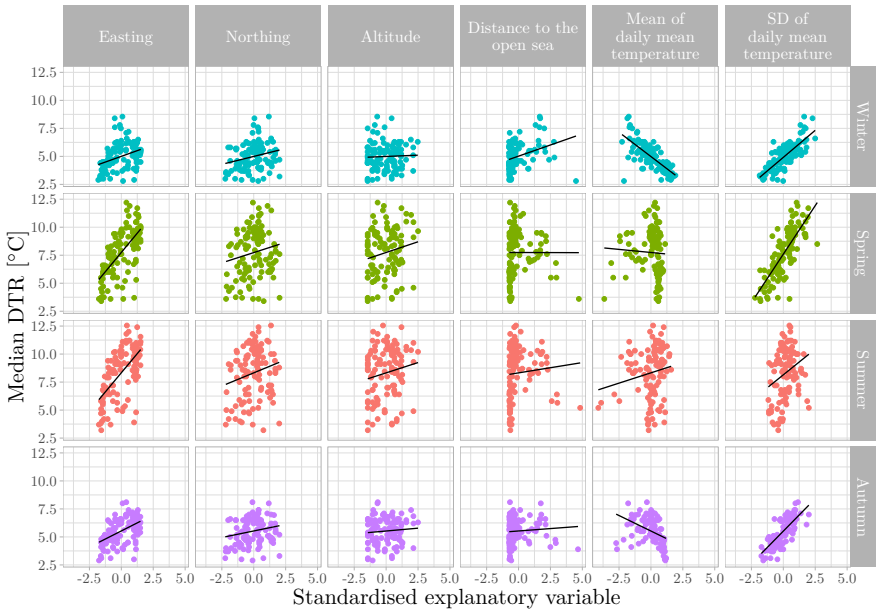


Figure 3: *Linear relationships between the six explanatory variables and the median of diurnal temperature range (DTR) at the 112 weather stations in our data set. All explanatory variables are standardised to have zero mean and a standard deviation of one.*

other quantiles. Histograms of diurnal temperature range for four selected stations and seasons are presented in Figure 2. We observe considerable differences in the shapes of the histograms.

To explore the relation between the explanatory variables and diurnal temperature range, we fit a simple linear model with different quantiles of diurnal temperature range against standardised versions of the explanatory variables. These linear model fits are presented in Figure 3 for the median of diurnal temperature range. Most of the estimated trends are significant at the 5%-level. Especially for the mean and variance of historical daily mean temperature, there is a strong linear relationship during winter and autumn. Similar trends are found for all other examined quantiles.

3 Models

3.1 Marginal modelling with the FPL distribution

We assume that daily minimum and maximum temperature can be described well by the generalised Pareto distribution. The Generalised Pareto distribution is a common model for extreme observations (e.g. Coles, 2001), and it is often used for modelling extreme temperature (e.g. Davison et al.,

2019; Castro-Camilo et al., 2021). Additionally, Rohrbeck et al. (2021) model daily temperature in the Red Sea by assuming that both the upper and lower tails of daily temperature can be modelled using the generalised Pareto distribution, and Stein (2021b) proposes to model climatological phenomena using parametric distributions that behaves like the generalised Pareto distribution in both tails, and uses this approach to model daily average temperature near Calgary during winter. The generalised Pareto distribution can be described through its quantile function (e.g. Hosking & Wallis, 1987)

$$\begin{aligned} Q(p;\mu,\eta,\xi) &= \inf\{x \in \mathbb{R} : p \leq F(x;\mu,\sigma,\xi)\} \\ &= \mu + \eta \begin{cases} \left(1-(1-p)^{\xi}\right)/\xi, & \xi \neq 0 \\ -\log(1-p), & \xi = 0, \end{cases} \end{aligned}$$

where F is the cumulative distribution function, and μ, η and ξ acts as the location, scale and shape parameters, respectively. Diurnal temperature range is equal to the difference between daily maximum and daily minimum temperature, and can therefore be modelled as the difference between the quantiles of two generalised Pareto distributions. We model maximum daily temperature as some quantile of a generalised Pareto distribution, while minimum daily temperature is modelled as some quantile of a reflected generalised Pareto distribution. If the random variable X has a quantile function $Q_X(p)$, then $-X$ has the quantile function $Q_{-X}(p) = -Q_X(1-p)$. This results in an expression for the diurnal temperature range:

$$\begin{aligned} Q_{\text{range}} &= Q_{\max}(p_1;\mu_1,\eta_1,\xi_1) - Q_{\min}(p_2;\mu_2,\eta_2,\xi_2) \\ &= \mu_1 + \frac{\eta_1}{\xi_1}(1-(1-p_1)^{\xi_1}) + \mu_2 + \frac{\eta_2}{\xi_2}(1-p_2^{\xi_2}). \end{aligned}$$

In the case of $\xi_1 = 0$ or $\xi_2 = 0$ the expression is simplified, since

$$\lim_{\lambda \to 0}(p^{\lambda}-1)/\lambda = \log p.$$

We reparametrise by setting $p_1 = p$, $p_2 = p^a$, $\xi_2^* = a\xi_2$ and $\eta_2^* = a\eta_2$, where $a = \log p_2/\log p$. This gives

$$Q_{\text{range}}(p) = \mu_1 + \frac{\eta_1}{\xi_1}(1-(1-p)^{\xi_1}) + \mu_2 + \frac{\eta_2^*}{\xi_2^*}(1-p^{\xi_2^*}).$$

Further manipulation of the expression yields

$$\begin{aligned} Q_{\text{range}}(p) &= (\mu_1+\mu_2) \\ &+ \frac{\eta_1-\eta_2^*}{2}\left\{\left(1-\frac{\eta_1+\eta_2^*}{\eta_1-\eta_2^*}\right)\frac{p^{\xi_2^*}-1}{\xi_2^*} - \left(1+\frac{\eta_1+\eta_2^*}{\eta_1-\eta_2^*}\right)\frac{(1-p)^{\xi_1}-1}{\xi_1}\right\}, \end{aligned}$$

which is equal to the quantile function of a five-parameter lambda (FPL) distribution (Gilchrist, 2000),

$$Q(p;\boldsymbol{\lambda}) = \lambda_1 + \frac{\lambda_2}{2}\left\{(1-\lambda_3)\frac{p^{\lambda_4}-1}{\lambda_4} - (1+\lambda_3)\frac{(1-p)^{\lambda_5}-1}{\lambda_5}\right\}, \quad \lambda_2 > 0, \lambda_3 \in [-1,1], \tag{1}$$

Consequently, we expect the FPL distribution to be a suitable model for diurnal temperature range.

No analytic expressions for the probability density function or cumulative distribution function of the FPL distribution exist, although the density for a given p can be obtained as the reciprocal of the quantile derivative, $(\mathrm{d}Q(p;\boldsymbol{\lambda})/\mathrm{d}p)^{-1}$. From the quantile function (1) of the FPL distribution it is clear that λ_1 acts as a location parameter and λ_2 as a scale parameter of the distribution. We notice that $\lambda_4 = \xi_2^*$ and $\lambda_5 = \xi_1$, which means that these two parameters control the behaviour of the left and right tails, respectively. The final parameter λ_3 acts as a weight between the two tails.

The support of the FPL distribution can be both finite and infinite. This makes the distribution flexible for modelling a variety of different phenomena. The support is given by

$$[Q(0;\boldsymbol{\lambda}), Q(1,\boldsymbol{\lambda})] = \lambda_1 + \frac{\lambda_2}{2}\begin{cases}[-\frac{1-\lambda_3}{\lambda_4}, \frac{1+\lambda_3}{\lambda_5}], & \lambda_4, \lambda_5 > 0\\ [-\frac{1-\lambda_3}{\lambda_4}, \infty), & \lambda_4 > 0, \lambda_5 \leqslant 0\\ (-\infty, \frac{1+\lambda_3}{\lambda_5}], & \lambda_4 \leqslant 0, \lambda_5 > 0\end{cases}.$$

Diurnal temperature range is always positive. In order to ensure a positive support for the FPL distribution, one must enforce the inequality-constraints

$$\lambda_1 - \frac{\lambda_2(1-\lambda_3)}{2\lambda_4} > 0, \qquad \lambda_4 > 0. \tag{2}$$

The parametrisation in (1) is intuitive in the sense that it stems from the combination of two generalised Pareto distributions. However, for performing numerical parameter estimation, other representations are more appropriate. The location λ_1 and scale λ_2 are not clearly linked to any central moments or quantiles of the FPL distribution. We thus propose a reparametrisation scheme with a new location parameter that is equal to the median of the FPL distribution, and a new scale parameter that is equal to the inter-quartile range of the FPL distribution,

$$\begin{aligned}\lambda_1^* &= Q(0.5;\boldsymbol{\lambda})\\ &= \lambda_1 + \frac{\lambda_2}{2}\left\{(1-\lambda_3)\frac{0.5^{\lambda_4}-1}{\lambda_4} - (1+\lambda_3)\frac{0.5^{\lambda_5}-1}{\lambda_5}\right\},\\ \lambda_2^* &= Q(0.75;\boldsymbol{\lambda}) - Q(0.25;\boldsymbol{\lambda})\\ &= \frac{\lambda_2}{2}\left\{\frac{1-\lambda_3}{\lambda_4}\left(0.75^{\lambda_4} - 0.25^{\lambda_4}\right) + \frac{1+\lambda_3}{\lambda_5}\left(0.75^{\lambda_5} - 0.25^{\lambda_5}\right)\right\}.\end{aligned} \tag{3}$$

The new parameter vector is denoted $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \lambda_3, \lambda_4, \lambda_5)$. In order to simplify parameter constraints during any numerical estimation procedures, we further introduce the reparametrisation

$$\begin{aligned} \tilde{\lambda}_1 = \lambda_1^*, \quad \tilde{\lambda}_2 = \log\left(e^{\lambda_2^*} - 1\right), \quad \tilde{\lambda}_3 = \log\left(\frac{1-\lambda_3}{1+\lambda_3}\right), \\ \tilde{\lambda}_4 = \log\left(e^{\lambda_4} - 1\right), \quad \tilde{\lambda}_5 = \log\left(e^{\lambda_5 + 0.5} - 1\right), \end{aligned} \tag{4}$$

This results in an unconstrained parameter vector $\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^5$ and guarantees that $\lambda_2 > 0$, $\lambda_3 \in (-1, 1)$ and $\lambda_4 > 0$. We also restrict λ_5 to the interval $(-0.5, \infty)$, as this guarantees a finite mean and variance for the FPL distribution (e.g. Coles, 2001; Tarsitano, 2010). Exploratory data analysis (results not shown) finds that the right tail parameter in the diurnal temperature range distribution tends to be considerably larger than -0.5. Davison et al. (2019) model extreme temperatures in Spain with the generalised Pareto distribution and find that the tail parameter, which we denote by λ_5, is approximately equal to 0.4. Castro-Camilo et al. (2021) and Rohrbeck et al. (2021) model Red Sea temperatures with the generalised Pareto distribution and find that the tail parameter is larger than -0.1. O'Sullivan et al. (2020) model temperature extremes in Dublin and find that the posterior median of the tail parameter is larger than 0.1. Based on these results and our exploratory data analysis, we are confident that the restriction of $\lambda_5 > -0.5$ should not lead to any loss in model performance.

The standard way of reparametrising a parameter θ that is bounded away from zero is to set $\tilde{\theta} = \log \theta$. However, if θ attains a large value, a small error in the estimate for $\tilde{\theta}$ leads to a considerable error in the estimate for θ. The function $g(x) = \log(e^x - 1)$ has the property that $g(x) \approx \log x$ for small x and $g(x) \approx x$ for large x. This allows us to constrain the FPL parameters without risking large reparametrisation instability because of an exponential relation between $\boldsymbol{\lambda}$ and $\tilde{\boldsymbol{\lambda}}$.

The reparametrisation to $\tilde{\boldsymbol{\lambda}}$ eases numerical inference methods, and is used whenever we perform parameter estimation. However, the $\boldsymbol{\lambda}^*$ parametrisation is more intuitive, and is therefore primarily used when describing our methods.

3.2 Distributional quantile regression

We wish to model the marginal distribution of diurnal temperature range at locations without available temperature observations, using a regression model with explanatory variables. As described in Section 2, the distribution of diurnal temperature range is very rich. Many of its distributional properties seem to vary in space, and between seasons. Thus, it does not seem good enough to use e.g. a generalised linear model (GLM) where we

only allow the mean and variance of diurnal temperature range to vary in space. We suggest that one should apply a distributional regression model, where the entire distribution function is allowed to vary in space, (e.g. Klein et al., 2015; Schlosser et al., 2019; Henzi et al., 2021) for modelling diurnal temperature range. One way of performing distributional regression is to apply a latent Gaussian model with an FPL likelihood, such that all five FPL parameters are modelled as a linear combination of explanatory variables and Gaussian white noise. However, this leads to an unnecessarily complex model. Additionally, as described in Section 5, the flexibility of the five FPL parameters might lead to something similar to identifiability problems, that can be problematic when we perform regression directly on the parameters and not on the distribution itself. Here, we propose a novel distributional regression model for diurnal temperature range which is based on combining quantile regression and marginal modelling with the FPL distribution, and we use this for modelling the marginal distribution of diurnal temperature range at locations with no temperature observations.

We first assume that any quantile in the distribution of diurnal temperature range can be modelled as a linear combination of explanatory variables. Let $y_i(\boldsymbol{s})$ be observation i of diurnal temperature range at location $\boldsymbol{s} \in \mathcal{S}$, where $\mathcal{S}$ is the given study area. For any probability $p \in (0, 1)$, we assume that the diurnal temperature range can be modelled as

$$y_i(\boldsymbol{s}) = \mathbf{x}(\boldsymbol{s})^T \boldsymbol{\beta}_p + \epsilon_{i,p}(\boldsymbol{s}), \quad i = 1, \ldots, n(\boldsymbol{s}), \tag{5}$$

with explanatory variables $\mathbf{x}(\boldsymbol{s})$ and regression coefficients $\boldsymbol{\beta}_p$. The error terms $\epsilon_{i,p}(\boldsymbol{s})$ are assumed to be independent and distributed such that $P(\epsilon_{i,p}(\boldsymbol{s}) \leq 0) = p$ for all $\boldsymbol{s} \in \mathcal{S}$ and $i = 1, \ldots, n(\boldsymbol{s})$ (Koenker, 2005). We are now able to estimate quantiles $q_p(\boldsymbol{s}) = \mathbf{x}(\boldsymbol{s})^T \boldsymbol{\beta}_p$ of diurnal temperature range for any $p \in (0, 1)$, at any location $\boldsymbol{s}$ with available explanatory variables $\mathbf{x}(\boldsymbol{s})$.

In order to turn this into a distributional regression model, we further propose to treat all $q_p(\boldsymbol{s})$ as quantiles of the FPL distribution at location $\boldsymbol{s}$. The only necessary assumption for performing quantile regression is that $P(y_i(\boldsymbol{s}) \leq \mathbf{x}(\boldsymbol{s})^T \boldsymbol{\beta}_p) = p$, and there need not be any disagreements between this and the assumption that the marginal distribution at any location $\boldsymbol{s}$ is the FPL distribution. Consequently, we model the distribution of diurnal temperature range at location $\boldsymbol{s}$ using the FPL distribution with the parameters $\boldsymbol{\lambda}^*$ that minimise the distance between $q_p(\boldsymbol{s})$ and the quantile function $Q(p; \boldsymbol{\lambda}^*)$ of the FPL distribution (1). With this approach, we are able to describe the distribution of diurnal temperature range everywhere, using a parametric model. This makes it easier to interpret the distributional properties of diurnal temperature range than when we only use the semi-parametric quantile regression model.

A common problem with quantile regression is that the different estimated quantile models may cross, such that the estimator for $q_{p_i}(\boldsymbol{s})$ is larger than the estimator for $q_{p_j}(\boldsymbol{s})$ for $p_j > p_i$ (Bondell et al., 2010; Rodrigues & Fan, 2017; Cannon, 2018). However, by first performing quantile regression and then fitting the FPL distribution to the regression quantiles, this problem is easily fixed, as the FPL quantile function always is monotonic increasing. Consequently, there is no need to implement complicated quantile regression methods that ensure non-crossing quantiles, and we can base our modelling on the fast and simple regression model where each quantile is modelled separately.

For simplicity, the distributional quantile regression model is referred to as the regression model for the remainder of the paper.

4 Inference

4.1 Parameter estimation for the FPL distribution

We present three marginal parameter estimation methods for the FPL distribution: the method of quantiles, maximum likelihood estimation and the starship method.

The method of quantiles

The method of quantiles is an estimation method similar to the better known method of moments, in which the distance between quantiles of a parametric distribution and empirical quantiles from observed data is minimised. Let $y_1, \ldots, y_n$ be independent and identically FPL distributed random variables with quantile function $Q(p; \boldsymbol{\lambda}^*)$ and parameters $\boldsymbol{\lambda}^*$. A set of m empirical quantiles $\widehat{Q}(p_i) = \widehat{q}_{p_i},\ i = 1, 2, \ldots, m$, are constructed from the observations $\boldsymbol{y}$. The method of quantiles estimator for $\boldsymbol{\lambda}^*$, is found by minimising the absolute distance

$$L(\boldsymbol{\lambda}^*) = \sum_{i=1}^{m} \left| \widehat{q}_{p_i} - Q(p_i; \boldsymbol{\lambda}^*) \right|. \tag{6}$$

There do not exist any straightforward expressions for the probability density or the cumulative probability function of the FPL distribution. However, a simple and closed-form expression exists for its quantile function. This can make it more natural to perform parameter estimation based on quantile matching instead of e.g. likelihood-based estimation methods. In addition, Bignozzi et al. (2018) state that parameter estimation methods based on quantile matching can be preferable when distributions are heavy-tailed or their support varies with the parameters. Both of these conditions hold for the FPL distribution. Additionally, Bhatti et al. (2018) find that the

method of quantiles outperforms both the method of moments and maximum likelihood estimation for parameter estimation under the Pareto distribution. As the FPL distribution can be described as the difference between two Pareto distributions, it should share some of the same properties. Tarsitano (2005) applies the method of quantiles for parameter estimation with the FPL distribution, using only five quantiles. He concludes that the method has several advantages, while a theoretical justification for the choice of quantiles is lacking. For a large set of observations $y_{(1)} \leqslant y_{(2)} \leqslant \dots y_{(n)}$, the distribution function of $y_{(i)}$ is close to the empirical distribution function $\widehat{F}(y_{(i)}) = (i - 0.5)/n$. Consequently, for a large set of n observations, we perform the method of quantiles by setting $p_i = (i - 0.5)/n$ and $\widehat{q}_{p_i} = y_{(i)}$ for $i = 1, 2, \dots, n$, thus avoiding the issue of which quantiles to select. This is somewhat similar to the method of least absolute deviations by Tarsitano (2010). Koenker (2005) has shown that the method of quantiles estimator is consistent, provided that all estimated quantiles $\widehat{q}$ are consistent. Note that the order statistics of $\boldsymbol{y}$ are dependent. This must be taken into account if one attempts to compute the variance of the estimator for $\boldsymbol{\lambda}^*$.

A weakness of the method of quantiles is that it might return a parameter estimator $\widehat{\boldsymbol{\lambda}^*}$ such that $Q(0; \widehat{\boldsymbol{\lambda}}^*) > y_{(1)}$ or $Q(1; \widehat{\boldsymbol{\lambda}}^*) < y_{(n)}$. This problem is addressed by introducing inequality constraints when minimising the loss function in (6), demanding that $Q(0; \boldsymbol{\lambda}^*) < y_{(1)}$ and $y_{(n)} < Q(1; \boldsymbol{\lambda}^*)$. In order to guarantee a positive support, the inequality constraints from (2) can also be enforced. Note that $\lambda_4 > 0$ is automatically enforced by optimising over the $\tilde{\boldsymbol{\lambda}}$ parametrisation. The positive support constraint is slightly relaxed by only demanding that $Q(10^{-4}; \boldsymbol{\lambda}^*) > 0$, as we find that this can considerably improve the model fit in certain cases. These constraints are enforced by performing numerical optimisation using an augmented Lagrangian formulation (e.g. Nocedal & Wright, 2006), implemented within the `R` package `nloptr` (Birgin & Martínez, 2008; Johnson, 2020). Closed-form expressions are available both for the quantile loss function and for its gradient. However, in practice we find that the method of quantiles performs better when not including gradient information in the optimiser. Consequently, we minimise the augmented Lagrangian using the derivative-free Nelder-Mead algorithm (Nelder & Mead, 1965), also implemented in the `nloptr` package.

Due to the flexibility of the FPL distribution, the quantile loss (6) proves to be difficult to minimise without a good initial value for the FPL parameters. We utilise the connection between $(\lambda_1^*, \lambda_2^*)$ and the quantiles of the FPL distribution by setting the initial values equal to the empirical median and inter-quartile range of $\boldsymbol{y}$, respectively. Initial values for the remaining three parameters are selected using a quick grid search. We compute the quantile loss function for all combinations of $\lambda_3 \in \{-0.5, -0.25, 0, 0.25, 0.5\}$,

$\lambda_4 \in \{0.1, 0.2, 0.4, 0.8, 1, 1.5\}$ and $\lambda_5 \in \{-0.4, -0.1, 0.1, 0.2, 0.4, 0.8, 1, 1.5\}$ and select the combination of parameters that minimises it as initial values.

Maximum likelihood estimation

No closed-form expression exists for the cumulative distribution function of the FPL distribution. However, as mentioned in Section 3.1, for a given probability p, the probability density function of the FPL distribution can be obtained as the reciprocal of the quantile derivative

$$f(y; \boldsymbol{\lambda}) = \frac{2}{\lambda_2}\Big\{(1-\lambda_3)p(y)^{\lambda_4-1} + (1+\lambda_3)(1-p(y))^{\lambda_5-1}\Big\}^{-1},$$

with $p(y) = F(y; \boldsymbol{\lambda})$. A numerical approximation for the cumulative distribution function of the FPL distribution is available from the `R` package `gld` (King et al., 2020). For $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ this gives rise to the log-likelihood

$$\ell(\boldsymbol{\lambda}; \boldsymbol{y}) = -n \log \frac{\lambda_2}{2} - \sum_{i=1}^{n} \log \Big\{(1-\lambda_3)p(y_i)^{\lambda_4-1} + (1+\lambda_3)(1-p(y_i))^{\lambda_5-1}\Big\}.$$

A straightforward expression for the gradient of the log-likelihood cannot be provided, as it requires computing the derivative of $F(y; \boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$. Maximisation of the log-likelihood is performed using the same grid-search and augmented Lagrangian as for the method of quantiles, where we include the same inequality constraint to guarantee a positive support.

The starship method

A straightforward way of modelling with the FPL distribution is to use the already implemented functions in the `R` package `gld` (King et al., 2020). This package is mostly focused on the four-parameter generalised lambda distribution, but it also includes one inference method for the FPL distribution, namely the starship method (Owen, 1988; King & MacGillivray, 1999). The starship method is based on the fact that if $\boldsymbol{y}$ has distribution function $F(\cdot; \boldsymbol{\lambda}^*)$, the transformed variable $\boldsymbol{u} = F(\boldsymbol{y}; \boldsymbol{\lambda}^*)$ has a uniform distribution. Parameters can therefore be estimated by minimising any goodness-of-fit statistic between the uniform distribution and $F(\boldsymbol{y}; \boldsymbol{\lambda}^*)$.

The starship suffers from the same problems as the maximum likelihood estimator, namely that no closed-form expression exists for the cumulative distribution function of the FPL distribution, which therefore has to be numerically approximated. The implemented method in the `gld` package performs minimisation using the Anderson-Darling statistic. However, the `gld` implementation of the starship method is very computationally inefficient, so in order to compare it with the previously described methods for large amounts of data, we implement our own version of the starship, based

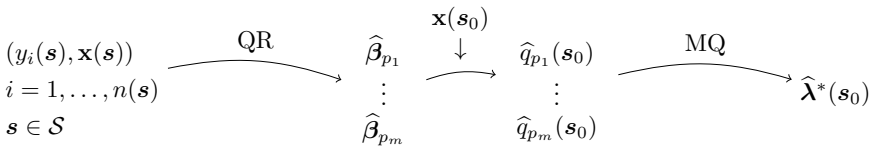


Figure 4: *Diagram of the regression model. First, quantile regression (QR) is performed for modelling different quantiles of diurnal temperature range $y_i(\boldsymbol{s})$ given the explanatory variables $\boldsymbol{x}(\boldsymbol{s})$. Using the quantile regression model, the conditional quantiles $\widehat{q}_{p_1}(\boldsymbol{s}_0), \widehat{q}_{p_2}(\boldsymbol{s}_0) \ldots, \widehat{q}_{p_m}(\boldsymbol{s}_0)$ are estimated, given the explanatory variables $\boldsymbol{x}(\boldsymbol{s}_0)$ at a specific location. Then, using the method of quantiles (MQ), the FPL distribution is fitted to the conditional quantiles, resulting in the estimator $\widehat{\boldsymbol{\lambda}}^*(\boldsymbol{s}_0)$ for the FPL parameters.*

on the `gld` implementation. This implementation performs minimisation of the Anderson-Darling statistic using the same optimisation approach as in the precious methods. On a sample of 10^4 observations, our version of the starship estimates $\boldsymbol{\lambda}^*$ in approximately 50 seconds, while the `gld` implementation uses approximately 330 seconds. The two implementations seem to perform equally well numerically.

4.2 Parameter estimation in the regression model

Inference for the regression model is divided into two steps. The estimation procedure is illustrated in Figure 4. First, quantile regression is performed using the `R` package `quantreg` (Koenker, 2018), separately for each of the probabilities $p_i = i/100,\ i = 1, 2, \ldots, 99$. Then, at any location $\boldsymbol{s}_0$ with available explanatory variables where we wish to model diurnal temperature range, we estimate the conditional quantiles $\widehat{q}_{p_i}(\boldsymbol{s}_0)$. The quantile function of the FPL distribution is then fitted to the 99 estimated quantiles $\widehat{q}_{p_1}(\boldsymbol{s}_0), \widehat{q}_{p_2}(\boldsymbol{s}_0) \ldots, \widehat{q}_{p_{99}}(\boldsymbol{s}_0)$ using a marginal parameter estimation method. There are no good ways of extending the maximum likelihood or starship method to fit a quantile function to a set of quantiles. However, the method of quantiles is perfect for this kind of parameter estimation problem. Consequently, the method of quantiles is used for fitting the FPL distribution to the 99 estimated quantiles, resulting in an estimator $\widehat{\boldsymbol{\lambda}}^*(\boldsymbol{s}_0)$ for the FPL parameters at $\boldsymbol{s}_0$. All 99 quantiles are modelled independently of each other, meaning that numerical inference can be executed in parallel. Fitting the FPL distribution to estimated quantiles at different locations is also an independent operation that can be performed in parallel. The proposed regression model is therefore highly parallelisable.

4.3 Model evaluation

Model performance is evaluated using the continuous ranked probability score (CRPS; Matheson & Winkler, 1976; Gneiting & Raftery, 2007). Given a forecast distribution F and an observation y, the CRPS is equal to

$$S(F, y) = \int_{-\infty}^{\infty} (F(t) - I(t \geq y))^2 \, \mathrm{d}t = 2 \int_0^1 \rho_p(y - F^{-1}(p)) \, \mathrm{d}p, \quad (7)$$

with $\rho_p(u) = pu - I(u < 0)u$, where $I(\cdot)$ is an indicator function. This is a strictly proper scoring rule, meaning that if y has distribution function G, then $E[S(G, y)] \leq E[S(F, y)]$ for all forecast distributions F, with equality only if $F = G$. Due to scarce usage of the FPL distribution in the literature, to the best of our knowledge, a closed-form expression for the CRPS with an FPL forecast distribution has not yet been provided. Consequently, we derive the expression for the CRPS with an FPL forecast distribution. When modelling diurnal temperature range with the FPL distribution, the quantile formulation of the CRPS is especially useful. Given an FPL forecast distribution F with quantile function Q, the CRPS can be expressed as

$$\begin{aligned} S(F, y) &= 2 \int_0^1 \rho_p(y - Q(p)) \, \mathrm{d}p \\ &= y(2F(y) - 1) - 2 \int_0^1 pQ(p) \, \mathrm{d}p + 2 \int_{F(y)}^1 Q(p) \, \mathrm{d}p. \end{aligned}$$

Both of these integrals are fairly straight-forward to compute with the FPL quantile function, as they are simply polynomials in p. Solving the first integral yields

$$\begin{aligned} \int_{F(y)}^1 Q(p) \, \mathrm{d}p =& (1 - F(y))\lambda_1 + \frac{\lambda_2}{2} \left\{ (1 - \lambda_3) \left(\frac{1 - F(y)^{\lambda_4 + 1}}{\lambda_4(\lambda_4 + 1)} - \frac{1 - F(y)}{\lambda_4} \right) \right. \\ & \left. - (1 + \lambda_3) \left(\frac{(1 - F(y))^{\lambda_5 + 1}}{\lambda_5(\lambda_5 + 1)} - \frac{1 - F(y)}{\lambda_5} \right) \right\}, \end{aligned}$$

while solving the second integral yields

$$\int_0^1 pQ(p) \, \mathrm{d}p = \frac{\lambda_1}{2} + \frac{\lambda_2}{2} \left(-(1 - \lambda_3) \frac{1}{2(\lambda_4 + 2)} + (1 + \lambda_3) \frac{\lambda_5 + 3}{2(\lambda_5 + 1)(\lambda_5 + 2)} \right).$$

Thus, by numerically approximating $F(y)$ we are able to estimate the CRPS of the FPL distribution. In the special case of $\lambda_4 = 0$ or $\lambda_5 = 0$, the integrals

are solved by using that

$$\int \log p \, \mathrm{d}p \propto p(\log p - 1),$$
$$\int \log(1-p) \, \mathrm{d}p \propto (p-1)\log(1-p) - p,$$
$$\int p \log p \, \mathrm{d}p \propto \frac{1}{2}p^2(2\log p - 1),$$
$$\int p \log(1-p) \, \mathrm{d}p \propto \frac{1}{4}\left(2(p^2-1)\log(1-p) - p(p+2)\right).$$

The CRPS can be used for comparing competing forecasts. Given two forecasts F_1 and F_2, and observations $\boldsymbol{y} = (y_1, \dots y_n)^T$, we can compute the mean CRPS

$$S(F, \boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} S(F, y_i),$$

and choose the forecast with the lowest mean CRPS. However, the mean CRPS in itself does not provide much information about the goodness of fit of a forecast. Given a forecast F and observations $\boldsymbol{y}$ with unknown distribution function G, there is no way of knowing if there is a large difference between $S(F, \boldsymbol{y})$ and $S(G, \boldsymbol{y})$, since G is unknown. Thus, we also evaluate model performance by studying quantile-quantile-plots (QQ-plots) and the probability integral transform (PIT).

If a random variable y has distribution function F, then the transformed variable $u = F(y)$ is uniformly distributed between zero and one. Given a forecast distribution F and observations $\boldsymbol{y}$ one can therefore examine deviations between the distribution of $\boldsymbol{u} = F(\boldsymbol{y})$ and the standard uniform distribution. Heinrich et al. (2020) propose to evaluate model fit by examining the first two moments of $\boldsymbol{u}$. Denote the error in the first moment as $e_\mu = E(\boldsymbol{u}) - 0.5$ and the error in the second central moment as $e_\sigma = \mathrm{SD}(\boldsymbol{u}) - 1/\sqrt{12}$. It follows that $e_\mu < 0$ indicates a positive bias and $e_\mu > 0$ indicates a negative bias. If $e_\sigma < 0$, the forecast distribution F is overdispersive, and if $e_\sigma > 0$, it is underdispersive.

5 Simulation study

5.1 Setup

Simulation studies are performed to compare the different parameter estimation methods for the FPL distribution. We draw 500 random sets of FPL parameters $\boldsymbol{\lambda}^*$. These are sampled such that they are of approximately the same magnitude as the estimated FPL parameters for diurnal temperature range in Section 6 (see Table 4), while also ensuring that we have a

positive and wide enough support. The exact sampling scheme is given in Algorithm 1, with $\mathcal{N}(\mu, \sigma^2)$ denoting a Gaussian distribution with mean μ and variance σ^2 and $U(a, b)$ denoting a uniform distribution with limits a and b. For each set of FPL parameters we then sample n realisations from the FPL($\boldsymbol{\lambda}^*$) distribution with $n = 2^i$ for $i = 7, 8, \ldots, 14$. The parameter estimation methods described in Section 4.1 are then applied for estimating $\boldsymbol{\lambda}^*$. We evaluate the overall fit to data and the ability of recovering the true parameter values. Overall model fit to data is evaluated using the CRPS. Since the true value of $\boldsymbol{\lambda}^*$ is known, we can compute the skill score $1 - \text{CRPS}(F, G)/\text{CRPS}(F, y)$, where F is our estimated distribution, G is the correct distribution and $\text{CRPS}(F, G)$ is the expected value of the CRPS with respect to G. Thus, a perfect forecast gives a skill score of zero, while all other forecasts give a skill score larger than zero and smaller than one. The ability to recovery the true values of $\boldsymbol{\lambda}^*$ is evaluated using the mean square error (MSE) between the true parameters $\boldsymbol{\lambda}^*$ and the estimated parameters $\hat{\boldsymbol{\lambda}}^*$, over all 500 repetitions,

$$\text{MSE}(\boldsymbol{\lambda}^*, \boldsymbol{\lambda}) = \frac{1}{500} \sum_{i=1}^{500} \frac{1}{5} \sum_{j=1}^{5} \left(\lambda_j^{*(i)} - \lambda_j^{(i)} \right)^2,$$

where $\boldsymbol{\lambda}^{(i)} = (\lambda_1^{(i)}, \ldots \lambda_5^{(i)})^T$ are the true FPL parameters in simulation number i out of 500, and $\boldsymbol{\lambda}^{*(i)}$ are the corresponding estimated parameters.

Algorithm 1 Sampling $\boldsymbol{\lambda}^*$

```
while TRUE do
    Sample λ*_1 ~ N(5, 3^2)
    Sample λ*_2 ~ U(1.5, 8)
    Sample λ_3 ~ U(−0.9, 0.9)
    Sample λ_4 ~ U(0.01, 0.9)
    Sample λ_5 ~ U(−0.3, 0.7)
    if Q(0, λ*) > 0 then                          ▷ Ensure a positive support
        if Q(1, λ*) − Q(0, λ*) > 1 then           ▷ Ensure a wide enough support
            break
        end if
    end if
end while
return λ*
```

5.2 Results

Table 1 displays the skill score, MSE and computation time for all methods. As n grows, the computation times for the maximum likelihood and starship

Table 1: *Mean skill scores, MSE and computation times for the method of quantiles (MQ), maximum likelihood (ML) and starship method, when performing parameter estimation 500 times on n samples drawn from an FPL distribution. The skill scores are multiplied by* 10^3 *to get more readable results. Computation times are reported on a 2.4 gHz computation server.*

	Method	$n = 2^7$	$n = 2^8$	$n = 2^9$	$n = 2^{10}$	$n = 2^{11}$	$n = 2^{12}$	$n = 2^{13}$	$n = 2^{14}$
Skill score ($\cdot 10^3$)	ML	6.12	3.18	1.48	0.73	0.40	0.21	0.11	0.09
	MQ	6.42	3.33	1.58	0.81	0.43	0.21	0.10	0.06
	starship	6.12	3.18	1.49	0.75	0.40	0.26	0.20	0.21
MSE	ML	0.17	0.14	0.06	0.03	0.02	0.02	0.02	0.01
	MQ	14.01	207.16	201.66	6.87	32.20	3.85	5.04	0.47
	starship	1.27	0.35	0.28	0.26	0.15	0.02	0.02	0.03
Time [s]	ML	0.6	0.9	1.5	3.0	5.6	11.0	22.6	46.4
	MQ	0.2	0.2	0.3	0.3	0.4	0.6	1.1	1.9
	starship	0.8	1.1	1.8	3.3	6.5	14.1	30.4	63.9

methods grow considerably faster than the time for the method of quantiles. This happens because the method of quantiles is based solely on analytical expressions, while the other two methods require numerical estimation of the likelihood or distribution function of the FPL distribution. The method of quantiles has a worse skill score for small to medium sample sizes and a slightly better skill score for large sample sizes, whereas the starship method attains the worst skill score for large sample sizes. Interestingly, the method of quantiles fails to recover the correct FPL parameters, and has a much larger MSE than the other two methods. This demonstrates the flexibility of the FPL distribution, as we are able to achieve a better CRPS using the "wrong" parameter estimates. A closer examination of the estimated parameters finds that the large increase in MSE is caused almost solely by too large estimates of λ_4 and λ_5. In some situations it seems that a large increase in λ_4 combined with a decrease in λ_3 yields almost no change in the overall shape of the FPL distribution. This makes sense, as increasing λ_4 leads to a thinner left tail, while decreasing λ_3 places more weight on the left tail. Similarly, a large increase in λ_5 can be mitigated by increasing λ_3. When modelling diurnal temperature range, model fit is much more important than parameter recovery, as the FPL model is merely an assumption, and possibly not the true underlying distribution of diurnal temperature range. Thus, the low skill score and fast computation times of the method of quantiles for large sample sizes make up for the fact that we seem to lose the ability to always recover the true parameters.

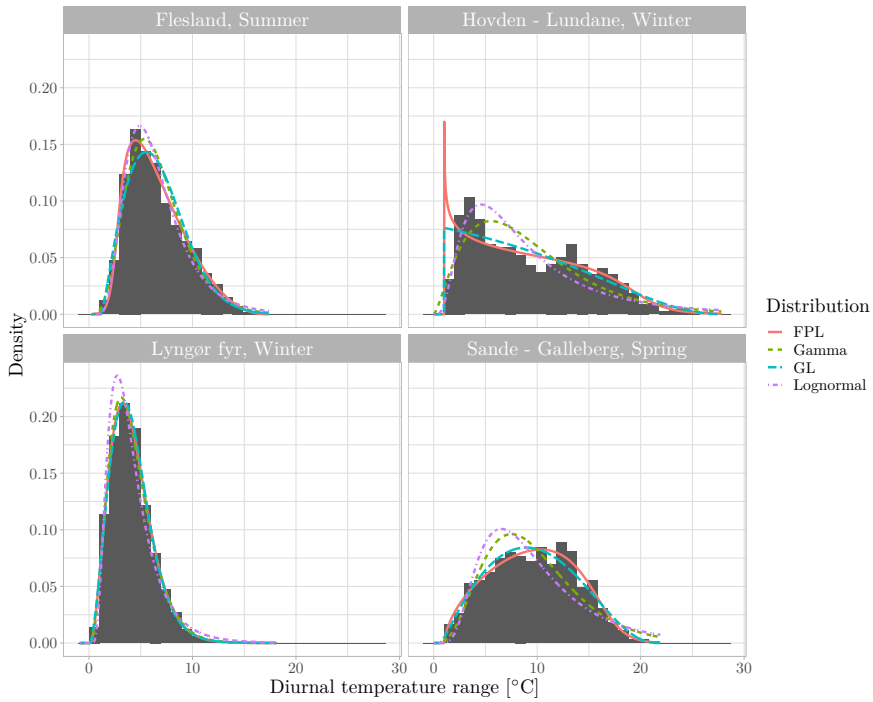


Figure 5: *Estimated probability density functions using four different parametric distribution for modelling diurnal temperature range. The observations of diurnal temperature range are displayed using histograms.*

6 Modelling diurnal temperature range in southern Norway

Diurnal temperature range in southern Norway is modelled separately for all seasons, using our two proposed models. Model calibration is evaluated using QQ-plots and the PIT. The marginal model is compared against several competing models using the CRPS. The regression model is tested in a leave-one-out cross-validation study, and the results are compared with the marginal model fits.

6.1 Marginal modelling

In order to evaluate the model performance of the FPL distribution for diurnal temperature range, the distribution is used for modelling data from southern Norway. Diurnal temperature range is modelled separately for all seasons and weather stations, using the FPL distribution and three other competing parametric distributions: the gamma distribution, lognormal distribution and the generalised lambda (GL) distribution. The gamma- and

Table 2: *Mean CRPS over all 112 weather stations during each season. Four different distributions are fitted to diurnal temperature range data. The FPL distribution are fitted to data using the method of quantiles (MQ), maximum likelihood (ML) estimation, and the starship method. The best mean CRPS for each season is written in* ***bold****.*

Season	FPL (MQ)	FPL (ML)	FPL (starship)	GL	Gamma	Lognormal
Winter	**1.577**	1.577	1.577	1.578	1.580	1.589
Spring	2.109	2.109	**2.108**	2.111	2.122	2.141
Summer	1.851	1.849	**1.849**	1.851	1.857	1.867
Autumn	**1.715**	1.715	1.715	1.717	1.720	1.732

lognormal distributions are fitted to data using maximum likelihood estimation, implemented in the `R` package `MASS` (Venables & Ripley, 2002). The GL distribution is a specialisation of the FPL distribution, parametrised with four parameters, and it is fitted using maximum likelihood, implemented in the `gld` package. The FPL distribution is fitted to diurnal temperature range using all three inference methods described in Section 4. In-sample model comparison is performed using the CRPS. This is computed numerically for the GL distribution, using (7). For the gamma- and lognormal distributions, CRPS is computed using the `scoringRules` package (Jordan et al., 2019).

Table 2 displays the mean CRPS over all 112 weather stations for each season and all our chosen models. Apart from the FPL results during summer, all three model fits with the FPL distribution attain a lower mean CRPS than the competing models. The differences in CRPS might seem small, but a simple permutation test shows that there is a statistically significant difference between the scores of the FPL and GL distributions, except during summer when inference is performed with the method of quantiles. The same permutation test finds no evidence that there is a difference in CRPS when using the method of quantiles and the starship method, but there is some evidence that both the starship and the method of quantiles attains better model fits than the maximum likelihood estimation. The mean computation time for estimating the FPL parameters at a single location is approximately 0.3 seconds with the method of quantiles, 2.7 seconds with maximum likelihood estimation and 4.0 seconds with the starship method, meaning that it takes two minutes to estimate parameters for all stations and seasons with the method of quantiles, and thirty minutes with the starship method. Figure 5 displays the fitted probability density functions of all models at the four stations from Figure 2. The FPL distribution is fitted to data using the method of quantiles. It seems that

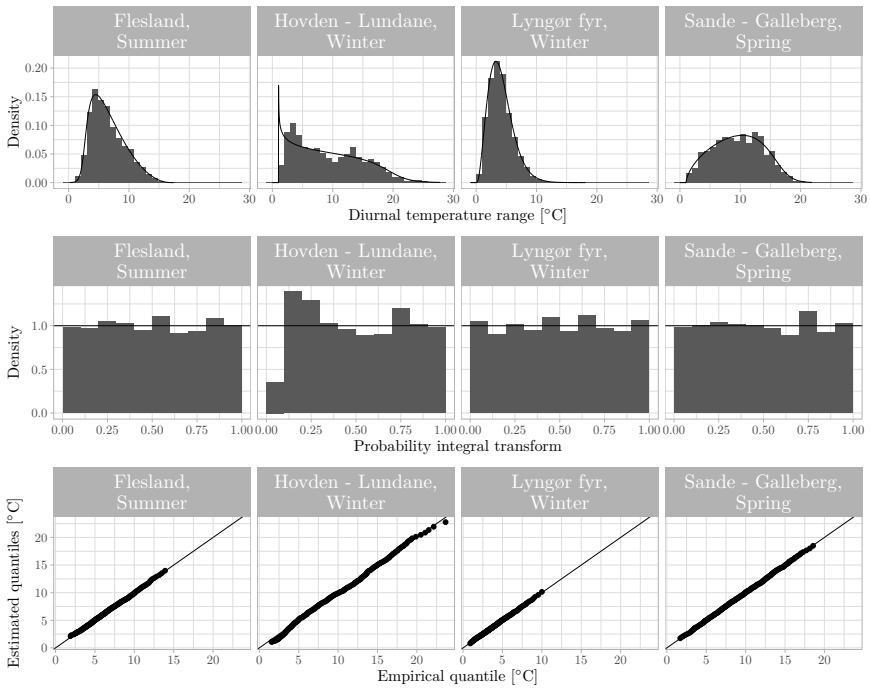


Figure 6: *Upper plots: Histograms displaying observed diurnal temperature range are plotted along with the probability density functions of the FPL distribution. Middle plots: Histograms displaying the PIT of diurnal temperature range with the estimated FPL parameters. Lower plots: QQ-plots displaying sample quantiles of diurnal temperature range against distributional quantiles of the estimated FPL distributions.*

the flexibility of the FPL distribution makes it able to model the many shapes of diurnal temperature range better than the competing models. Especially in the lower right plot one can see how the added flexibility of the FPL distribution allows it to provide a slightly better fit to data than the GL distribution, which clearly is the strongest competitor. The FPL distribution attains a lower CRPS than the competing models in all four sub-plots.

The FPL model fits are further evaluated to assess absolute performance. For the remainder of the paper we choose to use the method of quantiles for fitting the FPL distribution to diurnal temperature range data, as all three inference methods perform almost equally well, and the method of quantiles is necessary for the regression model described in Section 3.2. Figure 6 displays different properties of the model fit of the FPL distribution at the four stations from Figure 5. It is evident that the FPL distribution provides a good model fit to observed diurnal temperature range data. The model struggles slightly with the bimodal distribution at Hovden – Lundane,

Table 3: *Mean CRPS and PIT errors e_μ and e_σ over all 112 weather stations, computed for the marginal model and the regression model. The regression model is fitted to data both in-sample, and out-of-sample using leave-one-out cross-validation.*

Season	Model	CRPS	e_μ	e_σ
Winter	Marginal	1.58	$0.00 \cdot 10^{-2}$	$0.00 \cdot 10^{-2}$
	Regression, in-sample	1.63	$1.46 \cdot 10^{-2}$	$-0.30 \cdot 10^{-2}$
	Regression, out-of-sample	1.63	$1.44 \cdot 10^{-2}$	$-0.35 \cdot 10^{-2}$
Spring	Marginal	2.11	$0.06 \cdot 10^{-2}$	$-0.05 \cdot 10^{-2}$
	Regression, in-sample	2.29	$1.10 \cdot 10^{-2}$	$-1.35 \cdot 10^{-2}$
	Regression, out-of-sample	2.31	$1.08 \cdot 10^{-2}$	$-1.49 \cdot 10^{-2}$
Summer	Marginal	1.85	$0.01 \cdot 10^{-2}$	$0.01 \cdot 10^{-2}$
	Regression, in-sample	2.12	$1.70 \cdot 10^{-2}$	$-3.51 \cdot 10^{-2}$
	Regression, out-of-sample	2.16	$1.58 \cdot 10^{-2}$	$-3.74 \cdot 10^{-2}$
Autumn	Marginal	1.71	$0.03 \cdot 10^{-2}$	$-0.05 \cdot 10^{-2}$
	Regression, in-sample	1.76	$0.97 \cdot 10^{-2}$	$-0.54 \cdot 10^{-2}$
	Regression, out-of-sample	1.77	$0.94 \cdot 10^{-2}$	$-0.60 \cdot 10^{-2}$

but has an excellent fit to the data at the other three stations. A visual assessment of the model fit for all other stations finds that the results in Figure 6 are representative for most of the available data. The mean PIT errors e_μ and e_σ are displayed in Table 3. Both are of magnitude 10^{-4} for the marginal FPL model, implying high overall performance.

An overview of the estimated FPL parameters is given in Table 4. The location and scale parameters are largest during summer and spring, and the tail parameters λ_4 and λ_5 seem to take approximately the same values for all seasons. The estimator for λ_5 is mostly far away from -0.5, indicating that the restriction of $\lambda_5 > -0.5$ has not lead to a decrease in model performance. The tail weight λ_3 seems to be almost evenly distributed between -1 and 1, but its distribution is clearly most focused on the positive side, where it lends most weight to the right tail of the FPL distribution. When examining the marginal parameter estimates in a map (results not shown) we find that both $\hat{\lambda}_1^*$ and $\hat{\lambda}_4$ increase when moving eastwards. The opposite is found for $\hat{\lambda}_3$, which attains its largest values to the west. $\hat{\lambda}_2^*$ and $\hat{\lambda}_5$ take on low values along the coast, and increase as we move further away from the sea,

Table 4: *The FPL parameters are estimated at all weather stations, using the marginal model and the out-of-sample regression model. Median parameter estimates over all 112 locations are displayed, along with the* 2.5% *and the* 97.5% *quantiles.*

		Winter			Spring			Summer			Autumn		
Model	$\hat{\boldsymbol{\lambda}}$	2.5%	50.0%	97.5%	2.5%	50.0%	97.5%	2.5%	50.0%	97.5%	2.5%	50.0%	97.5%
Marginal	$\hat{\lambda}_1$	3.0	5.0	7.7	3.6	8.0	11.5	3.7	8.8	11.8	3.1	5.6	7.4
	$\hat{\lambda}_2$	1.9	3.7	7.0	2.3	5.7	8.1	1.9	4.9	7.1	1.9	4.4	6.9
	$\hat{\lambda}_3$	−0.6	0.2	1.0	−0.9	0.1	1.0	−0.6	0.4	1.0	−0.7	0.6	1.0
	$\hat{\lambda}_4$	0.0	0.4	0.8	0.0	0.4	1.2	0.0	0.2	0.5	0.0	0.2	1.1
	$\hat{\lambda}_5$	−0.1	0.1	0.3	−0.2	0.2	0.5	−0.1	0.3	0.6	0.0	0.2	0.5
Regression	$\hat{\lambda}_1$	2.9	4.8	7.1	3.4	7.8	12.5	4.6	8.3	11.2	3.3	5.5	7.6
	$\hat{\lambda}_2$	1.5	3.8	6.2	2.3	5.7	8.7	3.6	5.2	7.1	1.9	4.6	6.7
	$\hat{\lambda}_3$	−0.5	0.2	0.9	−0.5	0.5	1.0	0.0	0.6	0.9	−0.3	0.5	0.8
	$\hat{\lambda}_4$	0.0	0.4	0.6	0.0	0.3	0.7	0.0	0.1	0.5	0.0	0.4	0.8
	$\hat{\lambda}_5$	−0.3	0.0	0.3	−0.3	0.2	0.7	0.0	0.3	0.7	−0.2	0.2	0.4

and further to the east. There are some locations where the estimates for λ_4 and λ_5 are much larger than 1. This is most likely caused by the problems discussed in Section 5, where a large change in a tail parameter combined with a change in λ_3 results in little change in the overall shape of the FPL distribution. The estimators for λ_2^* and λ_3 are also showing unusual values at these locations.

6.2 Regression model

The regression model is applied for modelling diurnal temperature range for each season separately, using all the explanatory variables introduced in Section 2. Before we apply the regression model, each explanatory covariate is standardised to have zero mean and a standard deviation of one. Modelling is performed in-sample using all available data, and out-of-sample in a leave-one-out cross-validation study. Thus, in the cross-validation study, the regression coefficients $\boldsymbol{\beta}_{p_i}$, $i = 1, 2, \ldots, 99$, are estimated 112 times for each season, each time by leaving one station out of the training data. The FPL parameters are then estimated at the one station that was not included in training the quantile regression models. Table 3 shows little difference in performance between in-sample and out-of-sample estimation, indicating that our model does not overfit to the data. For winter and autumn data, the differences between the CRPS of the marginal model and the regression model are small. However, during summer and spring, there is a considerable difference in performance between the two models. The calibration of the regression model is clearly worse than that of the marginal model for all

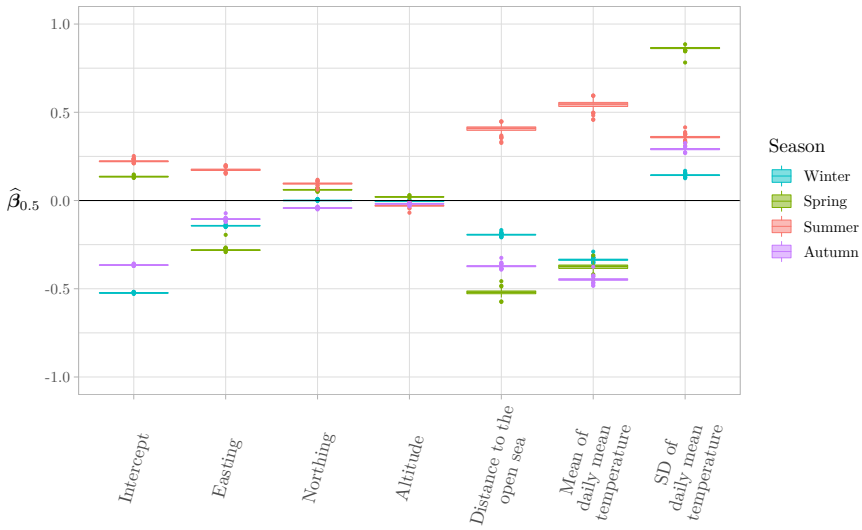


Figure 7: *Estimated regression coefficients* $\widehat{\boldsymbol{\beta}}_{0.5}$ *for the median regression. The parameter estimation is performed out-of-sample, resulting in 112 parameter estimates within each season. The whiskers in the box-plot have a maximum length of* 1.5 *times the interquartile range. The y-axis has unit "standard deviations of* $\boldsymbol{y}$*", meaning that if a regression coefficient is e.g. 0.4, then a change of one standard deviation for the corresponding explanatory variable causes a change of* $0.4 \cdot SD(\boldsymbol{y})$ *in the median of* $\boldsymbol{y}$*.*

seasons. However, PIT errors with a magnitude of 10^{-2} is still good, even though it is worse than a magnitude of 10^{-4}. All estimated regression coefficients for the 112 out-of-sample median regressions are displayed in Figure 7. The variability in $\widehat{\boldsymbol{\beta}}_{0.5}$ within each season is small, indicating that the estimation procedure is robust against minor changes in the training data. We see that the most influential explanatory variables across all seasons are the distance to the open sea and the historical temperature observation. There is a large difference between summer, spring and the other two seasons for these explanatory variables. For example, a larger mean temperature will lead to larger median range during summer, but lower median range during all other seasons. This might be connected to the difference in model performance during summer and spring. Similar trends are found in the $\widehat{\boldsymbol{\beta}}_p$ for all other probabilities p.

Figure 8 displays the out-of-sample estimation results for the same stations and seasons as in Figure 6. While the regression model is able to capture the overall shapes at each location, some deviations are noticeable. In particular, the estimated right tail is too heavy at all stations but Hovden – Lundane, where it is too light. Apart from this, the model fits seem

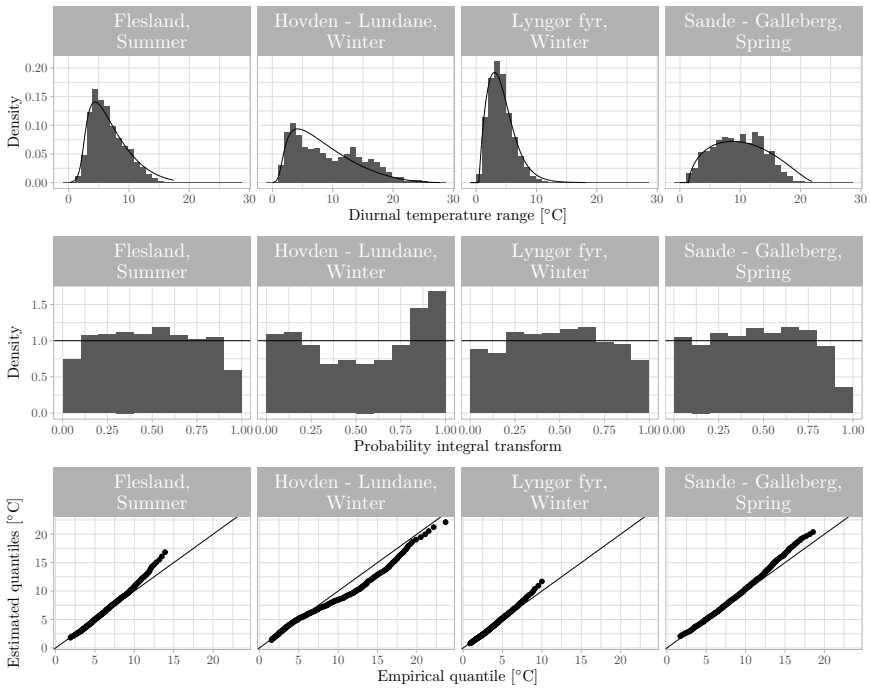


Figure 8: *Upper plots: Histograms displaying observed diurnal temperature range are plotted along with the probability density functions of the FPL distribution from the out-of-sample regression model. Middle plots: Histograms displaying the PIT of diurnal temperature range with the estimated FPL parameters. Lower plots: QQ-plots displaying sample quantiles of diurnal temperature range against distributional quantiles of the estimated FPL distributions.*

adequate for the bulk of the data. As seen in Table 4, the estimated FPL parameters from the out-of-sample regression model shares many similarities with the parameter estimates from the marginal model. Most of the spatial trends from the marginal model fits are also found when examining the estimates from the regional model.

Figure 9 summarises the calibration of both the marginal model fits and the out-of-sample regression model fits for summer and winter seasons by displaying the PIT errors e_μ and e_σ at all locations. For the out-of-sample assessment of the regression model, the calibration of the estimated distributions varies substantially between the two seasons. In winter, the calibration is, expectedly, somewhat worse than that of the marginal model. However, only a few stations show considerable lack of calibration. Both positive and negative values of e_μ are observed, indicating both negative and positive biases. However, the values of e_σ are rather negative than positive, indicating a slight tendency towards overdispersion or too large spread. Similar patterns are observed for autumn (results not shown). The

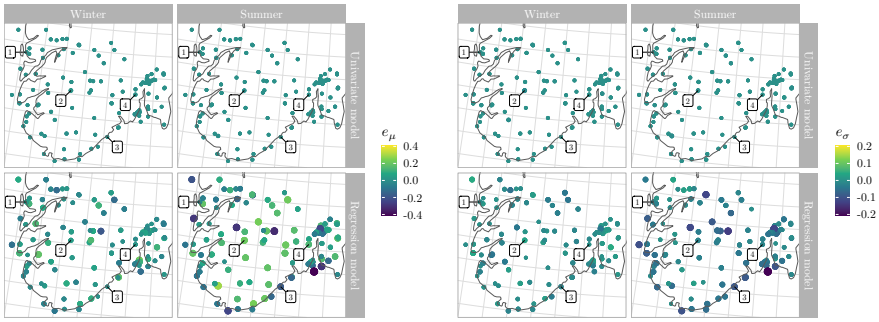


Figure 9: *PIT mean errors e_μ and standard deviation errors e_σ are displayed for the regression model and the marginal model, during summer and winter. The magnitude of each error is represented by the radius of each dot. Values of each error are represented by the colour of each dot. The locations of the four stations from Figures 2, 6 and 8 are presented in all the plots. These are: 1) Flesland, 2) Hovden – Lundane, 3) Lyngør fyr and 4) Sande – Galleberg.*

performance of the regression model in summer is considerably worse than that in winter. There is a distinct jump in the values of e_μ and e_σ as the distance from the sea increases. The jump in e_μ implies that we observe mostly positive biases along the coast and negative biases at inland locations. Figure 1 shows that there is a similar pattern of change in the median of diurnal temperature range along the coast and further inland during summer and spring. This indicates that the regression model is too smooth, and fails to model the transition from coastal to inland climate. A similar jump in e_μ and e_σ can be seen for spring data, although the difference is not as considerable as for the summer data.

Figure 3 hints that there is more information to gain from the climatological explanatory variables describing the distribution of mean temperature in winter and autumn than in summer and spring. In order to investigate whether the seasonal difference in skill is due to this effect, we repeated the regression analysis using only the four geographic and orographic explanatory variables. Under this model, the performance for winter and autumn data is considerably worse, while it stays almost unchanged for spring and summer data (results not shown). The magnitude of the PIT errors during winter and autumn with this model is also comparable with those for spring and summer data in the original model.

7 Conclusion and discussion

This paper proposes to use the five-parameter lambda (FPL) distribution to model the distribution of diurnal temperature range. A distributional

quantile regression model is also proposed, where diurnal temperature range is modelled using a combination of quantile regression and marginal modelling with the FPL distribution. Parameter estimation is performed using the method of quantiles, which is a fast inference method that compares well with the competing inference methods of maximum likelihood estimation and the starship method. Diurnal temperature range from southern Norway is modelled using the marginal FPL model and the regression model. The marginal FPL model provides a good fit to diurnal temperature range data, and the regression model shows promise and is able to capture much, but not all, of the spatial trends in the distribution of diurnal temperature range.

We propose a reparametrisation of the FPL distribution that allows for easier inference by directly connecting the location λ_1 and scale λ_2 to quantiles of the distribution. This allows us to select better initial values for parameter estimation of the FPL parameters. The reparametrisation also limits the parameter space of $\boldsymbol{\lambda}$ so that $\lambda_5 > -1/2$. This does not appear to influence our results.

Although the method of quantiles shows great success in fitting the FPL distribution to data, we experience situations where the estimators for λ_4 and λ_5 become too large with respect to the true values, which the method accounts for by providing biased estimates for λ_3. For large sample sizes this does not seem to affect the model fit much, but for smaller sample sizes it seems to negatively affect the performance. Further work should be put into understanding the effect of the parameters of the FPL distribution, and why the method of quantiles overestimates tail parameters in certain situations.

Model evaluation is mainly performed using the probability integral transform (PIT) and the continuous ranked probability score (CRPS). A novel closed-form expression for the CRPS with an FPL forecast distribution is developed, which makes model evaluation faster and simpler. Other scoring rules may be more appropriate for evaluating model fit in larger and more inhomogeneous regions (Bolin & Wallin, 2023), but we believe the CRPS to be a good choice for model evaluation in the current setting.

While the regression model is not able to fully capture the spatial patterns in diurnal temperature range for spring and summer, its performance is promising, especially for winter and autumn data. It is our belief that one can achieve better results with an improved selection of explanatory variables and, potentially, further development of the distributional quantile regression method. As an example, many of the weather stations with high PIT errors e_μ and e_σ are located along the coast. Consequently, it might be reasonable to better distinguish between coastal observations and inland observations, e.g., through the introduction of a binary explanatory variable. A transformation of variables could also improve our models. The

relationships between the median of diurnal temperature range and the available explanatory variables shown in Figure 3 do not seem linear for the geographical information. One might find possible transformations of the explanatory variables which are able to improve the linear relationships between the quantiles of diurnal temperature range and the available explanatory variables. In addition to the explanatory variables used in the current study, one might find important dependencies between diurnal temperature range and other climate variables, such as daily precipitation, wind speed and the degree of cloud cover. Especially precipitation and cloud cover have been found to be highly negatively correlated with diurnal temperature range (Zhou et al., 2009; Waqas & Athar, 2018). It is not obvious how such explanatory variables should be incorporated in our model, though. The regression model for diurnal temperature range only includes spatial fixed effects. The inclusion of spatial random effects might improve model performance (Lum & Gelfand, 2012; Self et al., 2021). Additionally, attempts to estimate parameter uncertainty will be affected by the temporal autocorrelation of diurnal temperature range, which is statistically significant for lags up to approximately one week. Consequently, further modelling attempts should also aspire to include a temporal framework. A comprehensive modelling framework for temperature range should furthermore include modelling components to account for data issues such as data inhomogeneities and measurement errors, including recording precision. This is particularly important in settings where long data series from various data sources are combined into a single analysis, like in the unified Bayesian approach that was introduced in the EUSTACE project (Rayner et al., 2020). In this project temporal and spatial autocorrelations were handled by latent random field components, and the data inhomogeneities were directly estimated via independent random effect variables. That approach allowed the propagation of uncertainty through the entire analysis system, in contrast to more traditional data homogenisation methods that handle this as a separate pre-processing step. In principle, the FPL model for diurnal temperature range can be incorporated into the observation level of such hierarchical models.

Two of the explanatory variables in the regression model are based on historical daily mean temperature. For spatial interpolation one might argue that observations of daily mean temperature are unavailable at most locations where there are no observations of diurnal temperature range. However, while the literature on modelling diurnal temperature range is lacking, much effort and success has been put into the modelling of mean temperature (e.g. Haylock et al., 2008; Maraun & Widmann, 2018). We assume that there already exist satisfactory spatial and temporal models for Norway, which are able to describe the historical mean and variance of daily mean temperature between 1989 and 2018 with a high performance.

The Nordic Gridded Climate Data Set version 2 (Lussana et al., 2018), e.g., models daily mean temperature with high performance everywhere in Norway, Finland and Sweden. Accordingly, all explanatory variables can be provided at any location in Norway.

Better models for diurnal temperature range may be important for improving interpolation and statistical downscaling of temperature projections from climate models (e.g. Maraun & Widmann, 2018). The common approach today is to perform separate modelling of daily maximum, minimum and mean temperature. However, this can lead to inconsistencies such as predictions where the daily mean is larger than the daily maximum temperature (e.g. Lussana et al., 2019). In addition, the three temperature variables are heavily dependent and should be modelled jointly, but multivariate modelling is often too challenging and computationally demanding. An alternative method for modelling these three temperature variables is to transform minimum, maximum and mean temperature to diurnal temperature range, mean temperature, and the location of the daily mean inside the temperature range. This would remove all ordering inconsistencies between minimum, mean and maximum temperature, and it would considerably reduce correlations between the three variables. Analysis of the data used in this paper finds that the absolute values of pairwise sample correlations between daily minimum, maximum and mean temperature mainly lie in the interval $[0.9, 1]$. Sample correlations between diurnal temperature range, daily mean temperature and the location of the mean inside the range, on the other hand, are always below 0.5. Thus, appropriate statistical models for diurnal temperature range and its relationship with daily mean temperature can be of great interest as model components for multivariate temperature modelling approaches. Further work should be conducted to examine this approach.

Acknowledgments

We thank Sara Martino, Cristian Lussana, Irene Brox Nilsen and Erik Kjellström for helpful discussions. Thordis L. Thorarinsdottir acknowledges the support of the Research Council of Norway through project nr. 255517 "Post-processing Climate Projection Output for Key Users in Norway". As part of the EUSTACE project, Finn Lindgren received funding from the European Union's "Horizon 2020 Programme for Research and Innovation", under Grant Agreement no 640171. The marginal FPL model for diurnal temperature range was first developed within the EUSTACE research project for historical global daily temperature reconstruction (Rayner et al., 2020), but was not used in their final analysis method due to constraints in the time frame of the project. The code and data used in this paper are available

at https://github.com/siliusmv/FPLD. This article is an extension of the Master's thesis of Silius M. Vandeskog (Vandeskog et al., 2019).

Conflict of interest

The authors declare no potential conflict of interests.

References

Ahmadabadi, M. N., Farjami, Y., & Moghadam, M. B. (2012). A process control method based on five-parameter generalized lambda distribution. *Quality & Quantity*, *46*(4), 1097–1111. https://doi.org/10.1007/s11135-011-9550-x

Bhatti, S. H., Hussain, S., Ahmad, T., Aslam, M., Aftab, M., & Raza, M. A. (2018). Efficient estimation of Pareto model: Some modified percentile estimators. *PloS one*, *13*(5), e0196456. https://doi.org/10.1371/journal.pone.0196456

Bignozzi, V., Macci, C., & Petrella, L. (2018). Large deviations for method-of-quantiles estimators of one-dimensional parameters. *Communications in Statistics-Theory and Methods*, 1–26. https://doi.org/10.1080/03610926.2018.1554134

Birgin, E. G., & Martínez, J. M. (2008). Improving ultimate convergence of an augmented Lagrangian method. *Optimization Methods and Software*, *23*(2), 177–195. https://doi.org/10.1080/10556780701577730

Bolin, D., & Wallin, J. (2023). Local scale invariance and robustness of proper scoring rules. *Statistical Science*, *38*(1), 140–159. https://doi.org/10.1214/22-STS864

Bondell, H. D., Reich, B. J., & Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, *97*(4), 825–838. https://doi.org/10.1093/biomet/asq048

Braganza, K., Karoly, D. J., & Arblaster, J. M. (2004). Diurnal temperature range as an index of global climate change during the twentieth century. *Geophysical Research Letters*, *31*(13). https://doi.org/10.1029/2004GL019998

Cannon, A. J. (2011). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & geosciences*, *37*(9), 1277–1284.

Cannon, A. J. (2018). Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. *Stochastic Environmental Research and Risk Assessment*, *32*(11), 3207–3225. https://doi.org/10.1007/s00477-018-1573-6

Castro-Camilo, D., Mhalla, L., & Opitz, T. (2021). Bayesian space-time gap filling for inference on extreme hot-spots: An application to Red Sea surface temperatures. *Extremes*, *24*(1), 105–128. https://doi.org/10.1007/s10687-020-00394-z

Cheng, J., Xu, Z., Zhu, R., Wang, X., Jin, L., Song, J., & Su, H. (2014). Impact of diurnal temperature range on human health: A systematic review. *International Journal of Biometeorology*, *58*(9), 2011–2024. https://doi.org/10.1007/s00484-014-0797-5

Coles, S. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). Springer, London. https://doi.org/10.1007/978-1-4471-3675-0

Davison, A. C., Huser, R., & Thibaud, E. (2019). Spatial extremes. In A. E. Gelfand, M. Fuentes, J. A. Hoeting, & R. L. Smith (Eds.), *Handbook of environmental and ecological statistics* (pp. 711–744). Chapman; Hall/CRC.

Gilchrist, W. (2000). *Statistical modelling with quantile functions*. Chapman; Hall/CRC.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Hanssen-Bauer, I., Førland, E., Haddeland, I., Hisdal, H., Mayer, S., Nesje, A., Nilsen, J., Sandven, S., Sandø, A., Sorteberg, A., & Ådlandsvik, B. (2016). Klima i Norge 2100. *Bakgrunnsmateriale til NOU Klimatilpassing., Norsk klimasenter, Oslo, Norway.*

Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., & New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres*, *113*(D20). https://doi.org/10.1029/2008JD010201

Heinrich, C., Hellton, K. H., Lenkoski, A., & Thorarinsdottir, T. L. (2020). Multivariate postprocessing methods for high-dimensional seasonal weather forecasts. *Journal of the American Statistical Association*, 1–12. https://doi.org/10.1080/01621459.2020.1769634

Henry, H. A. L. (2007). Climate change and soil freezing dynamics: Historical trends and projected changes. *Climatic Change*, *87*(3-4), 421–434. https://doi.org/10.1007/s10584-007-9322-8

Henzi, A., Ziegel, J. F., & Gneiting, T. (2021). Isotonic distributional regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *83*(5), 963–993. https://doi.org/10.1111/rssb.12450

Hosking, J. R., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, *29*(3), 339–349. https://doi.org/10.1080/00401706.1987.10488243

Hothorn, T., Kneib, T., & Bühlmann, P. (2014). Conditional transformation models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 3–27. https://doi.org/10.1111/rssb.12017

IPCC. (2007). Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007 [[S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller (eds.)]].

IPCC. (2013). Climate change 2013: The physical science basis. Contribution of working group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [[T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (eds.)]].

IPCC. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [[Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis,

M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]].

Johnson, S. G. (2020). *The NLopt nonlinear-optimization package*. http://github.com/stevengj/nlopt

Jordan, A., Krüger, F., & Lerch, S. (2019). Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software*, *90*(12), 1–37. https://doi.org/10.18637/jss.v090.i12

King, R. A. R., Dean, B., Klinke, S., & van Staden, P. (2020). *Gld: Estimation and use of the Generalised (Tukey) Lambda distribution* [R package version 2.6.2].

King, R. A. R., & MacGillivray, H. L. (1999). Theory & methods: A starship estimation method for the generalized λ distributions. *Australian & New Zealand Journal of Statistics*, *41*(3), 353–374. https://doi.org/10.1111/1467-842X.00089

Klein, N., Kneib, T., Lang, S., & Sohn, A. (2015). Bayesian Structured Additive Distributional Regression With an Application To Regional Income Inequality in Germany. *The Annals of Applied Statistics*, *9*(2), 1024–1052. https://doi.org/10.1214/15-AOAS823

Koenker, R. (2005). *Quantile regression* (Vol. 38). https://doi.org/10.1017/CBO9780511754098

Koenker, R. (2018). *Quantreg: Quantile regression* [R package version 5.35].

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Kovi, M. R., Ergon, Å., & Rognli, O. A. (2016). Freezing tolerance revisited — effects of variable temperatures on gene regulation in temperate grasses and legumes. *Current Opinion in Plant Biology*, *33*, 140–146. https://doi.org/10.1016/j.pbi.2016.07.006

Lim, Y.-H., Reid, C. E., Mann, J. K., Jerrett, M., & Kim, H. (2015). Diurnal temperature range and short-term mortality in large US communities. *International Journal of Biometeorology*, *59*(9), 1311–1319. https://doi.org/10.1007/s00484-014-0941-2

Lum, K., & Gelfand, A. E. (2012). Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Anal.*, *7*(2), 235–258. https://doi.org/10.1214/12-BA708

Lussana, C., Tveito, O. E., Dobler, A., & Tunheim, K. (2019). Senorge_2018, daily precipitation, and temperature datasets over Norway. *Earth System Science Data*, *11*(4), 1531–1551. https://doi.org/10.5194/essd-11-1531-2019

Lussana, C., Tveito, O., & Uboldi, F. (2018). Three-dimensional spatial interpolation of 2 m temperature over Norway. *Quarterly Journal of the Royal Meteorological Society*, *144*(711), 344–364. https://doi.org/10.1002/qj.3208

Maraun, D., & Widmann, M. (2018). *Statistical downscaling and bias correction for climate research*. Cambridge University Press. https://doi.org/10.1017/9781107588783

Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*(10), 1087–1096. https://doi.org/10.1287/mnsc.22.10.1087

Nair, N. U., Sankaran, P., & Balakrishnan, N. (2013). *Quantile-based reliability analysis*. Springer. https://doi.org/0.1007/978-0-8176-8361-0

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*(4), 308–313. https://doi.org/10.1093/comjnl/7.4.308

Nocedal, J., & Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

Norwegian Meteorological Institute. (2019). EKlima: Free access to weather- and climate data from Norwegian Meteorological Institute from historical data to real time observations [Accessed: 2019-01-15].

O'Sullivan, J., Sweeney, C., & Parnell, A. C. (2020). Bayesian spatial extreme value analysis of maximum temperatures in County Dublin, Ireland. *Environmetrics*, *31*(5), e2621. https://doi.org/10.1002/env.2621

Owen, D. B. (1988). The starship. *Communications in Statistics - Simulation and Computation*, *17*(2), 315–323. https://doi.org/10.1080/03610918808812665

Park, J.-E., Son, W.-S., Ryu, Y., Choi, S. B., Kwon, O., & Ahn, I. (2020). Effects of temperature, humidity, and diurnal temperature range on influenza incidence in a temperate region. *Influenza and Other Respiratory Viruses*, *14*(1), 11–18. https://doi.org/10.1111/irv.12682

Peng, S., Piao, S., Ciais, P., Myneni, R. B., Chen, A., Chevallier, F., Dolman, A. J., Janssens, I. A., Peñuelas, J., Zhang, G., Vicca, S., Wan, S., Wang, S., & Zeng, H. (2013). Asymmetric effects of daytime and night-time warming on northern hemisphere vegetation. *Nature*, *501*(7465), 88–92. https://doi.org/10.1038/nature12434

Ramberg, J. S., & Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Communications of the ACM*, *17*(2), 78–82. https://doi.org/10.1145/360827.360840

Rayner, N. A., Auchmann, R., Bessembinder, J., Brönnimann, S., Brugnara, Y., Capponi, F., Carrea, L., Dodd, E. M. A., Ghent, D., Good, E., Høyer, J. L., Kennedy, J. J., Kent, E. C., Killick, R. E., van der Linden, P., Lindgren, F., Madsen, K. S., Merchant, C. J., Mitchelson, J. R., ... Woolway, R. I. (2020). The EUSTACE project: Delivering global, daily information on surface air temperature. *Bulletin of the American Meteorological Society*, *101*(11), E1924–E1947. https://doi.org/10.1175/BAMS-D-19-0095.1

Rodrigues, T., & Fan, Y. (2017). Regression Adjustment for Noncrossing Bayesian Quantile Regression. *Journal of Computational and Graphical Statistics*, *26*(2), 275–284. https://doi.org/10.1080/10618600.2016.1172016

Rohrbeck, C., Simpson, E. S., & Towe, R. P. (2021). A spatio-temporal model for Red Sea surface temperature anomalies. *Extremes*, *24*(1), 129–144. https://doi.org/10.1007/s10687-020-00383-2

Schlosser, L., Hothorn, T., Stauffer, R., & Zeileis, A. (2019). Distributional Regression Forests for Probabilistic Precipitation Forecasting in Complex terrain. *The Annals of Applied Statistics*, *13*(3), 1564–1589. https://doi.org/10.1214/19-AOAS1247

Self, S. W., McMahan, C. S., & Russell, B. T. (2021). Identifying meteorological drivers of PM2.5 levels via a Bayesian spatial quantile regression. *Environmetrics*, 1–20. https://doi.org/10.1002/env.2669

Shelton, S., Pushpawela, B., & Liyanage, G. (2021). The long-term trend in the diurnal temperature range over Sri Lanka from 1985 to 2017 and its

association with total cloud cover and rainfall. *Journal of Atmospheric and Solar-Terrestrial Physics*, *227*, 105810. https://doi.org/10.1016/j.jastp.2021.105810

Stein, M. L. (2021a). A parametric model for distributions with flexible behavior in both tails. *Environmetrics*, *32*(2), e2658. https://doi.org/10.1002/env.2658

Stein, M. L. (2021b). Parametric models for distributions when interest is in extremes with an application to daily temperature. *Extremes*, *24*(2), 293–323. https://doi.org/10.1007/s10687-020-00378-z

Sun, X., Ren, G., You, Q., Ren, Y., Xu, W., Xue, X., Zhan, Y., Zhang, S., & Zhang, P. (2019). Global diurnal temperature range (DTR) changes since 1901. *Climate Dynamics*, *52*(5), 3343–3356. https://doi.org/10.1007/s00382-018-4329-6

Sun, X., Wang, C., & Ren, G. (2021). Changes in the diurnal temperature range over east Asia from 1901 to 2018 and its relationship with precipitation. *Climatic Change*, *166*(3), 1–17. https://doi.org/10.1007/s10584-021-03120-1

Tarsitano, A. (2004). Fitting the generalized lambda distribution to income data. *COMPSTAT'2004 Symposium*, 1861–1867.

Tarsitano, A. (2005). Estimation of the Generalized Lambda Distribution parameters for grouped data. *Communications in Statistics - Theory and Methods*, *34*(8), 1689–1709. https://doi.org/10.1081/STA-200066334

Tarsitano, A. (2010). Comparing estimation methods for the FPLD. *Journal of Probability and Statistics*, *2010*. https://doi.org/10.1155/2010/295042

Thorne, P. W., Donat, M. G., Dunn, R. J. H., Williams, C. N., Alexander, L. V., Caesar, J., Durre, I., Harris, I., Hausfather, Z., Jones, P. D., Menne, M. J., Rohde, R., Vose, R. S., Davy, R., Klein-Tank, A. M. G., Lawrimore, J. H., Peterson, T. C., & Rennie, J. J. (2016). Reassessing changes in diurnal temperature range: Intercomparison and evaluation of existing global data set estimates. *Journal of Geophysical Research: Atmospheres*, *121*(10), 5138–5158. https://doi.org/10.1002/2015JD024584

Thorne, P. W., Menne, M. J., Williams, C. N., Rennie, J. J., Lawrimore, J. H., Vose, R. S., Peterson, T. C., Durre, I., Davy, R., Esau, I., Klein-Tank, A. M. G., & Merlone, A. (2016). Reassessing changes in diurnal temperature range: A new data set and characterization of data biases. *Journal of Geophysical Research: Atmospheres*, *121*(10), 5115–5137. https://doi.org/10.1002/2015JD024583

Tukey, J. W. (1962). The future of data analysis. *The annals of mathematical statistics*, *33*(1), 1–67.

Vandeskog, S. M., Thorarinsdottir, L., Thordis, & Steinsland, I. (2019). *Modelling diurnal temperature range in Norway: Statistical modelling and spatial interpolation with the Five-Parameter Lambda Distribution* [Master's thesis, NTNU].

Vasseur, D. A., DeLong, J. P., Gilbert, B., Greig, H. S., Harley, C. D. G., McCann, K. S., Savage, V., Tunney, T. D., & O'Connor, M. I. (2014). Increased temperature variation poses a greater risk to species than climate warming. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1779), 20132612. https://doi.org/10.1098/rspb.2013.2612

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth). Springer.

Vered, S., Paz, S., Negev, M., Tanne, D., Zucker, I., & Weinstein, G. (2020). High ambient temperature in summer and risk of stroke or transient ischemic attack: A national study in Israel. *Environmental Research*, *187*, 109678. https://doi.org/10.1016/j.envres.2020.109678

Vinnarasi, R., Dhanya, C., Chakravorty, A., & AghaKouchak, A. (2017). Unravelling diurnal asymmetry of surface temperature in different climate zones. *Scientific reports*, *7*(1), 1–8. https://doi.org/10.1038/s41598-017-07627-5

Wang, K., Li, Y., Wang, Y., & Yang, X. (2017). On the asymmetry of the urban daily air temperature cycle. *Journal of Geophysical Research: Atmospheres*, *122*(11), 5625–5635. https://doi.org/10.1002/2017JD026589

Waqas, A., & Athar, H. (2018). Observed diurnal temperature range variations and its association with observed cloud cover in northern Pakistan. *International Journal of Climatology*, *38*(8), 3323–3336. https://doi.org/10.1002/joc.5503

Zhou, L., Dai, A., Dai, Y., Vose, R. S., Zou, C.-Z., Tian, Y., & Chen, H. (2009). Spatial dependence of diurnal temperature range trends on precipitation from 1950 to 2004. *Climate Dynamics*, *32*, 429–440. https://doi.org/10.1007/s00382-008-0387-5

— 2 —

Modelling sub-daily precipitation extremes with the blended generalised extreme value distribution

Silius M. Vandeskog, Sara Martino, Daniela Castro-Camilo and Håvard Rue

Journal of Agricultural, Biological and Environmental Statistics 27, 598–621 (2022)

Modelling sub-daily precipitation extremes with the blended generalised extreme value distribution

Silius M. Vandeskog[1] Sara Martino[1]
Daniela Castro-Camilo[2] Håvard Rue[3]

[1] Norwegian University of Science and Technology (NTNU)
[2] University of Glasgow
[3] King Abdullah University of Science and Technology (KAUST)

Abstract

A new method is proposed for modelling the yearly maxima of sub-daily precipitation, with the aim of producing spatial maps of return level estimates. Yearly precipitation maxima are modelled using a Bayesian hierarchical model with a latent Gaussian field, with the blended generalised extreme value (bGEV) distribution used as a substitute for the more standard generalised extreme value (GEV) distribution. Inference is made less wasteful with a novel two-step procedure that performs separate modelling of the scale parameter of the bGEV distribution using peaks over threshold data. Fast inference is performed using integrated nested Laplace approximations (INLA) together with the stochastic partial differential equation (SPDE) approach, both implemented in `R-INLA`. Heuristics for improving the numerical stability of `R-INLA` with the GEV and bGEV distributions are also presented. The model is fitted to yearly maxima of sub-daily precipitation from the south of Norway, and is able to quickly produce high-resolution return level maps with uncertainty. The proposed two-step procedure provides an improved model fit over standard inference techniques when modelling the yearly maxima of sub-daily precipitation with the bGEV distribution.

1 Introduction

Heavy rainfall over short periods of time can cause flash floods, large economic losses and immense damage to infrastructure. The World Economic Forum states that climate action failure and extreme weather events are perceived among the most likely and most impactful global risks in 2021 (World Economic Forum, 2021). Therefore, a better understanding of heavy rainfall can be of utmost importance for many decision-makers, e.g.,

those that are planning the construction or maintenance of important infrastructure. In this paper, we create spatial maps with estimates of large return levels for sub-daily precipitation in Norway. Estimation of return levels is best described within the framework of extreme value theory, where the most common methods are the block maxima and the peaks over threshold (e.g. Coles, 2001; Davison & Huser, 2015). Due to low data quality (see Section 2 for more details) and the difficulty of selecting high-dimensional thresholds, we choose to use the block maxima method for estimating the precipitation return levels. This method is based on modelling the maximum of a large block of random variables with the generalised extreme value (GEV) distribution, which is the only non-degenerate limit distribution for a standardised block maximum (Fisher & Tippett, 1928). When working with environmental data, blocks are typically chosen to have a size of one year (Coles, 2001). Inference with the GEV distribution is difficult, partially because its support depends on its parameter values. Castro-Camilo et al. (2022) propose to ease inference by substituting the GEV distribution with the blended generalised extreme value (bGEV) distribution, which has the right tail of a Fréchet distribution and the left tail of a Gumbel distribution, resulting in a heavy-tailed distribution with a parameter-free support. Both Castro-Camilo et al. (2022) and Vandeskog et al. (2021) demonstrate with simulation studies that the bGEV distribution performs well as a substitute for the GEV distribution when estimating properties of the right tail. Additionally, in this paper we develop a simulation study that shows how the parameter-dependent support of the GEV distribution can lead to numerical problems during inference, while inference with the bGEV distribution is more robust. This can be of crucial importance in complex and high-dimensional settings, and consequently we choose to model the yearly maxima of sub-daily precipitation using the bGEV distribution.

Modelling of extreme daily precipitation has been given much attention in the literature, and it is well established that precipitation is a heavy-tailed phenomenon (e.g. Katz et al., 2002; Wilson & Toumi, 2005; Papalexiou & Koutsoyiannis, 2013), which makes the bGEV distribution a possible model for yearly precipitation maxima. Spatial modelling of extreme daily precipitation has also received a great amount of interest. Cooley et al. (2007) combine Bayesian hierarchical modelling with a generalised Pareto likelihood for estimating large return values for daily precipitation. Similar methods are also applied by Sang and Gelfand (2009), Davison et al. (2012), Geirsson et al. (2015), and Opitz et al. (2018), using either the block maxima or the peaks over threshold approach. Using a multivariate peaks over threshold approach, Castro-Camilo and Huser (2020) propose local likelihood inference for a specific factor copula model to deal with complex non-stationary dependence structures of precipitation over the contiguous U.S. Spatial modelling of extreme sub-daily precipitation is more difficult,

due to less available data sources. Consequently, this is often performed using intensity-duration-frequency relationships where one pools together information from multiple aggregation times in order to estimate return levels (Koutsoyiannis et al., 1998; Lehmann et al., 2016; Wang & So, 2016; Ulrich et al., 2020). Spatial modelling of extreme hourly precipitation in Norway has previously been performed by Dyrrdal et al. (2015). After their work was published, the number of observational sites for hourly precipitation in Norway has greatly increased. We aim to improve their return level estimates by including all the new data that have emerged over the last years. We model sub-daily precipitation using a spatial Bayesian hierarchical model with a bGEV likelihood and a latent Gaussian field. In order to keep our model simple, we do not pool together information from multiple aggregation times, making our model purely spatial. The model assumes conditional independence between observations, which makes it able to estimate the marginal distribution of extreme sub-daily precipitation at any location, but unable to successfully estimate joint distributions over multiple locations. In the case of hydrological processes such as precipitation, ignoring dependence might lead to an underestimation of the risk of flooding. However, Davison et al. (2012) find that models where the response variables are independent given some latent process can be a good choice when the aim is to estimate a spatial map of marginal return levels.

High-resolution spatial modelling can demand a lot of computational resources and be highly time-consuming. The framework of integrated nested Laplace approximations (INLA; Rue et al., 2009) allows for a considerable speed-up by using numerical approximations instead of sampling-based inference methods like Markov chain Monte Carlo (MCMC). Inference with a spatial Gaussian latent field can be even further sped up with the so-called stochastic partial differential equation (SPDE; Lindgren et al., 2011) approach of representing a Gaussian random field using a Gaussian Markov random field that is the approximate solution of a specific SPDE. Both INLA and the SPDE approach have been implemented in the `R-INLA` library, which is used for performing inference with our model (Bivand et al., 2015; Rue et al., 2017; Bakka et al., 2018). `R-INLA` requires a log-concave likelihood to ensure numerical stability during inference. However, neither the GEV likelihood nor the bGEV likelihood are log-concave, which can cause inferential issues. We present heuristics for mitigating the risk of numerical instability caused by a lack of log-concavity.

A downside of the block maxima method is that inference can be somewhat wasteful compared to the peaks over threshold method. Additionally, most of the available weather stations in Norway that measure hourly precipitation are young and contain quite short time series. This data sparsity makes it challenging to place complex models on the parameters of the bGEV distribution in the hierarchical model. A promising method of accounting

for data-sparsity is the recently developed sliding block estimator, which allows for better data utilisation by not requiring that the block maxima used for inference come from disjoint blocks (Bücher & Segers, 2018; Zou et al., 2021). However, to the best of our knowledge, no theory has yet been developed for using the disjoint block estimator on non-stationary time series, or for performing Bayesian inference with the disjoint block estimator. Vandeskog et al. (2021) propose a new two-step procedure that allows for less wasteful and more stable inference with the block maxima method by separately modelling the scale parameter of the bGEV distribution using peaks over threshold data. Having modelled the scale parameter, one can standardise the block maxima so the scale parameter can be considered as a constant, and then estimate the remaining bGEV parameters. Bücher and Zhou (2021) suggests that, when modelling stationary time series, the peaks over threshold technique is preferable over block maxima if the interest lies in estimating large quantiles of the stationary distribution of the times series. The opposite holds if the interest lies in estimating return levels, i.e. quantiles of the distribution of the block maxima. Thus, both methods have different strengths, and by using this two-step procedure, one can take advantage of the merits and improve the pitfalls of both methods. We apply the two-step procedure for modelling sub-daily precipitation and compare the performance with that of a standard block maxima model where all the bGEV parameters are estimated jointly.

The remainder of the paper is organised as follows. Section 2 introduces the hourly precipitation data and all explanatory variables used for modelling. Section 3 presents the bGEV distribution and describes the Bayesian hierarchical model along with the two-step modelling procedure. Additionally, heuristics for improving the numerical stability of `R-INLA` are proposed, and a score function for evaluating model performance is presented. In Section 4 we perform modelling of the yearly precipitation maxima in Norway. A cross-validation study is performed for evaluating the model fit, and a map of return levels is estimated. Conclusions are presented in Section 5.

2 Data

2.1 Hourly precipitation data

Observations of hourly aggregated precipitation from a total of 380 weather stations in the south of Norway are downloaded from an open archive of historical weather data from MET Norway (https://frost.met.no). The oldest weather stations contain observations from 1967, but approximately 90 percent of the available weather stations are established after 2000. Each observation comes with a quality code, but almost all observations

from before 2005 are of unknown quality. An inspection of the time series with unknown quality detects unrealistic precipitation observation ranging from −300 mm/hour to 400 mm/hour. Other unrealistic patterns, like 50 mm/hour precipitation for more than three hours in a row, or no precipitation during more than half a year, are also detected. The data set contains large amounts of missing data, but these are often recorded as 0 mm/hour, instead of being recorded as missing. Thus, there is no way of knowing which of the zeros represent missing data and which represent an hour without precipitation. Having detected all of this, we decide to remove all observations with unknown or bad quality flags, which accounts for approximately 14% of the total number of observations. Additionally, we clean the data by removing all observations from years with more than 30% missing data and from years where more than two months contain less than 20% of the possible observations. This data cleaning is performed to increase the probability that our observed yearly maxima are close or equal to the true yearly maxima. Having cleaned the data, we are left with 72% of the original observations, distributed over 341 weather stations and spanning the years 2000 to 2020. The total number of usable yearly maxima is approximately 1900. Figure 1 displays the distribution of the number of usable yearly precipitation maxima per weather station. The majority of the weather stations contain five or less usable yearly maxima, and approximately 50 stations have more than 10 usable maxima. Figure 1 also displays the location of all the weather stations. A large amount of the stations are located close to each other, in the southeast of Norway. Such spatial clustering can be an indicator for preferential sampling. However, we do not believe that preferential sampling is an issue for our data. The weather stations are mostly placed in locations with high population densities, and to the best of our knowledge there is no strong dependency between population density and extreme precipitation in Norway, as there are large cities located both in dry and wet areas of the country. Even though most stations are located in areas with high population densities, there is still a good spatial coverage of the entire area of interest, also for areas with low population densities.

The yearly maxima of precipitation accumulated over $1, 2, \ldots, 24$ hours are computed for all locations and available years. A rolling window approach with a step size of 1 hour is used for locating the precipitation maxima. As noted by Robinson and Tawn (2000), a sampling frequency of one hour is not enough to observe the exact yearly maximum of hourly precipitation. With this sampling frequency, one only observes the precipitation during the periods 00:00–01:00, 01:00–02:00, etc., whereas the maximum precipitation might occur e.g. during the period 14:23–15:23. Approximately half of the available weather stations have a sampling frequency of one minute, while the other half only contain hourly observations. We therefore use a sampling

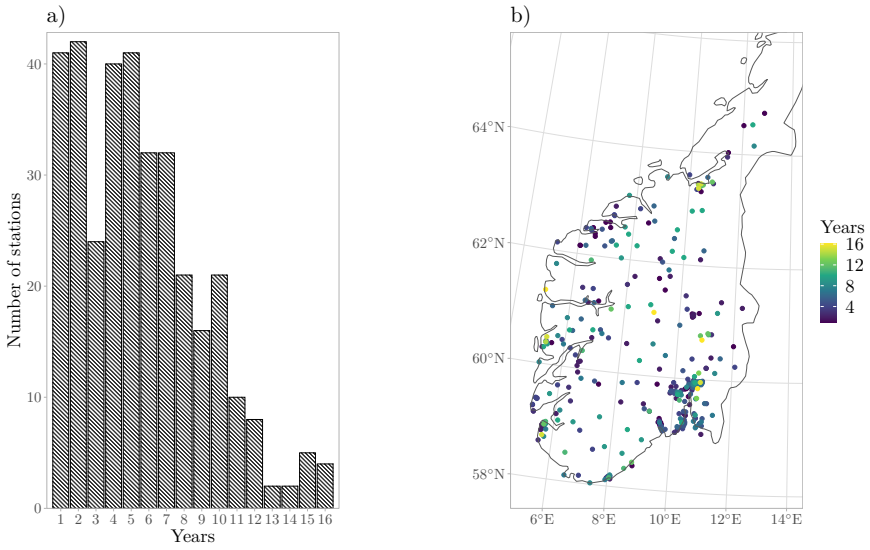


Figure 1: *a) A histogram displaying the number of usable yearly precipitation maxima for all the weather stations used in this paper. b) The location of the 341 weather stations. The number of usable yearly precipitation maxima from each station are displayed using different colours. Note that some points overlap in areas with high station densities.*

frequency of one hour for all weather stations, as this allows us to use all the 341 weather stations without having to account for varying degrees of sampling frequency bias in our model.

Dyrrdal et al. (2015) used the same data source for estimating return levels of hourly precipitation. They fitted their models to hourly precipitation maxima using only 69 weather stations from all over Norway. However, they received a cleaned data set from the Norwegian Meteorological Institute, resulting in time series with lengths up to 45 years. Our data cleaning approach is more strict than that of Dyrrdal et al. (2015) in the sense that it results in shorter time series by removing all data of uncertain quality. On the other hand, we include more locations and get a considerably better spatial coverage, by keeping all time series with at least one good year of observations.

The main focus of this paper is the novel methodology for fast and accurate estimation of return levels, and we believe that we have prepared the data well enough to give a good demonstration of our proposed model and to achieve reliable return level estimates for sub-daily precipitation. It is trivial to add more, or differently cleaned data, to improve the return level estimates at a later time.

Table 1: *Explanatory variables used for modelling sub-daily precipitation extremes. The two rightmost columns show which explanatory variables are used for modelling which parameters of the bGEV distribution for yearly precipitation (see Section 3.2).*

Explanatory variable	Description	Unit	$\mathbf{x}_\mu$	$\mathbf{x}_\sigma$
Mean annual precipitation	Mean annual precipitation for the years 1981–2010	mm	✓	
Easting	Eastern coordinate (UTM 32)	km	✓	✓
Northing	Northern coordinate (UTM 32)	km	✓	✓
Altitude	Height above sea level	m	✓	
Distance to the open sea	Shortest distance to the open sea	km	✓	✓

2.2 Explanatory variables

We use one climate-based and four orographic explanatory variables. These are displayed in Table 1. Altitude is extracted from a digital elevation model of resolution 50×50 m^2, from the Norwegian Mapping Authority (https://hoydedata.no). The distance to the open sea is computed using the digital elevation model. Precipitation climatologies for the period 1981–2010 are modelled by Crespi et al. (2019). The climatologies do not cover the years 2011–2020, from which most of the observations come. We assume that the precipitation patterns have not changed overly much and that they are still representative for the years 2011–2020. Hanssen-Bauer and Førland (1998) find that, in most southern regions of Norway, the only season with a significant increase in precipitation is Autumn. This strengthens our assumption that the change in precipitation patterns is slow enough to not be problematic for us.

Dyrrdal et al. (2015) include additional explanatory variables in their model, such as temperature, summer precipitation and the yearly number of wet days. They find mean summer precipitation to be one of the most important explanatory variables. We compute these explanatory variables at all station locations using the gridded seNorge2 data product (Lussana, Saloranta, et al., 2018; Lussana, Tveito, & Uboldi, 2018). Our examination finds that yearly precipitation, summer precipitation and the yearly number of wet days are close to 90% correlated with each other. There is also a negative correlation between temperature and altitude of around -85%. Consequently, we choose to not use any more explanatory variables for modelling, as highly correlated variables might lead to identifiability issues during parameter estimation.

3 Methods

3.1 The bGEV distribution

Extreme value theory concerns the statistical behaviour of extreme events, possibly larger than anything ever observed. It provides a framework where probabilities associated with these events can be estimated by extrapolating into the tail of the distribution. This can be used for e.g. estimating large quantiles, which is the aim of this work (e.g. Coles, 2001; Davison & Huser, 2015). A common approach in extreme value theory is the block maxima method. Assume that the limiting distribution of the standardised block maximum $(Y_k - b_k)/a_k$ is non-degenerate, where $Y_k = \max\{X_1, X_2, \ldots, X_k\}$ is the maximum over k random variables from a stationary stochastic process, and $\{b_k\}$ and $\{a_k > 0\}$ are some appropriate sequences of standardising constants. Then, for large enough block sizes k, the distribution of the block maximum Y_k is approximately equal to the GEV distribution with (Fisher & Tippett, 1928; Coles, 2001)

$$P(Y_k \leq y) \approx \begin{cases} \exp\left\{-[1 + \xi(y - \mu_k)/\sigma_k]_+^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-\exp[-(y - \mu_k)/\sigma_k]\right\}, & \xi = 0, \end{cases} \tag{1}$$

where $(a)_+ = \max\{a, 0\}$, $\sigma_k > 0$ and $\mu_k, \xi \in \mathbb{R}$. In most settings, k is fixed, so we denote $\sigma = \sigma_k$ and $\mu = \mu_k$. A challenge with the GEV distribution is that its support depends on its parameters. This complicates inference procedures such as maximum likelihood estimation (e.g. Smith, 1985; Bücher & Segers, 2017), and can be particularly problematic in a covariate-dependent setting with spatially varying parameters, as it might also introduce artificial boundary restrictions such as an unnaturally large lower bound for yearly maximum precipitation. Castro-Camilo et al. (2022) propose the bGEV distribution as an alternative to the GEV distribution in settings where the tail parameter ξ is non-negative. The support of the bGEV distribution is parameter-free and infinite. This allows for more numerically stable inference, while also avoiding the possibility of estimated lower bounds that are larger than future observations. The bGEV distribution function is

$$H(y; \mu, \sigma, \xi, a, b) = F(y; \mu, \sigma, \xi)^{v(y;a,b)} G(y; \tilde{\mu}, \tilde{\sigma})^{1 - v(y;a,b)}, \tag{2}$$

where F is a GEV distribution with $\xi \geq 0$ and G is a Gumbel distribution. The weight function is equal to

$$v(y; a, b) = F_\beta\left(\frac{y - a}{b - a}; c_1, c_2\right),$$

where $F_\beta(\cdot; c_1, c_2)$ is the distribution function of a beta distribution with parameters $c_1 = c_2 = 5$, which leads to a symmetric and computationally

efficient weight function. The weight $v(y; a, b)$ is zero for $y \leq a$ and one for $y \geq b$, meaning that the left tail of the bGEV distribution is equal to the left tail in G, while the right tail is equal to the right tail in F. The choice of the weight $v(y; a, b)$ should not considerably affect inference if we let the difference between a and b be small. The parameters $\tilde{\mu}$ and $\tilde{\sigma}$ are injective functions of (μ, σ, ξ) such that the bGEV distribution function is continuous and $F(y; \mu, \sigma, \xi) = G(y; \tilde{\mu}, \tilde{\sigma})$ for $y \in \{a, b\}$. Setting $a = F^{-1}(p_a)$ and $b = F^{-1}(p_b)$ with small probabilities $p_a = 0.1,\ p_b = 0.2$ makes it possible to model the right tail of the GEV distribution without any of the problems caused by a finite left tail. See Castro-Camilo et al. (2022) for guidelines on how to choose c_1, c_2, p_a and p_b.

In the supplementary material we present a simulation study where both the GEV distribution and the bGEV distribution are fitted to univariate samples from a GEV distribution. We demonstrate how a small change in initial values can cause large numerical problems for inference with the GEV distribution, and no noticeable difference for inference with the bGEV distribution. The fact that considerable numerical problems can arise for the GEV distribution in a univariate setting with large sample sizes and perfectly GEV-distributed data strongly indicates that the GEV distribution is not robust enough to be used reliably in complex, high-dimensional problems with noisy data. The bGEV distribution is more robust than the GEV distribution, and we therefore prefer it over the GEV distribution for modelling precipitation maxima in Norway.

Although the bGEV distribution is more robust than the GEV distribution, it might still seem unnatural to model block maxima using the bGEV distribution, when it is known that the correct limiting distribution is the GEV distribution. However, we argue that the bGEV would be a good choice for modelling heavy tailed block maxima even if it had not been more robust than the GEV distribution. In multivariate extreme value theory it is common to assume that the tail parameter ξ of the GEV distribution is constant in time and/or space (e.g. Koutsoyiannis et al., 1998; Sang & Gelfand, 2010; Opitz et al., 2018; Castro-Camilo et al., 2019). This assumption is often made, not because one truly believes that it should be constant, but because estimation of ξ is difficult, and models with a constant ξ often are "good enough". The tail parameter is incredibly important for the shape of the GEV distribution, and small changes in ξ can lead to large changes in return levels, and even affects the existence of distributional moments. A model where ξ varies in space can therefore e.g. provide model fits with a finite mean in one location and an infinite mean in the neighbouring location. Such a model can also give scenarios where a new observation at one location can change the existence of moments in other, possibly far away, locations. Thus, even though it might seem unnatural to use a constant tail parameter, these models often provide more natural fits to data than the models that

allow ξ to vary in space. We claim that the bGEV distribution fulfils a similar role as a model with constant ξ, but for the model support instead of the moments. When ξ is positive, the support of the GEV distribution varies with its parameter values. In regression settings with covariates and finite amounts of data one can therefore experience unnatural lower bounds that are known to be wrong. Furthermore, if only one new observation is smaller than the estimated lower limit, the entire model fit will be invalidated. We therefore prefer the bGEV, which completely removes the lower bound while still having the right tail of the GEV distribution, thus yielding a model that is "good enough" for estimating return levels, but without the unwanted model properties in the left tail of the GEV distribution.

Naturally, the bGEV distribution can only be applied for modelling exponential- or heavy-tailed phenomena ($\xi \geq 0$). However, it is well established that extreme precipitation should be modelled with a non-negative tail parameter. Cooley et al. (2007) performs Bayesian spatial modelling of extreme daily precipitation in Colorado and find that the tail parameter is positive and less than 0.15. Papalexiou and Koutsoyiannis (2013) examine more than 15000 records of daily precipitation worldwide and conclude that the Fréchet distribution performs the best. They propose that even when the data suggests a negative tail parameter, it is more reasonable to use a Gumbel or Fréchet distribution. Less information is available concerning the distribution of extreme sub-daily precipitation. However, Koutsoyiannis et al. (1998) argues that the distribution of precipitation should not have an upper bound for any aggregation period, so ξ must be non-negative. Van de Vyver (2012) estimate the distribution of yearly precipitation maxima in Belgium for aggregation times down to 1 minute, and find that the estimates of ξ increase as the aggregation times decreases, meaning that the tail parameter for sub-daily precipitation should be larger than for daily precipitation. Dyrrdal et al. (2016) estimate ξ for daily precipitation in Norway from the seNorge1 data product (Tveito et al., 2005; Mohr, 2009) and conclude that the tail parameter estimates are non-constant in space and often negative. However, the authors do not provide confidence intervals or p-values and do not state whether the estimates are significantly different from zero. Based on our own exploratory analysis (results not shown) and the overwhelming evidence in the literature, we assume that sub-daily precipitation is a heavy-tailed phenomenon.

Following Castro-Camilo et al. (2022), we reparametrise the bGEV distribution from (μ, σ, ξ) to $(\mu_\alpha, \sigma_\beta, \xi)$, where the location parameter μ_α is equal to the α quantile of the bGEV distribution if $\alpha \geq p_b$. The scale parameter σ_β, hereby denoted the spread parameter, is equal to the difference between the $1 - \beta/2$ quantile and the $\beta/2$ quantile of the bGEV distribution if $\beta/2 \geq p_b$. There is a one to one relationship between the new and the old parameters. The new parametrisation is advantageous as it is considerably

easier to interpret than the old parametrisation. The parameters μ_α and σ_β are directly connected to the quantiles of the bGEV distribution, whereas μ and σ have no simple connections with any kind of moments or quantiles. Consequently, it is much easier to choose informative priors for μ_α and σ_β. Based on preliminary experiments, we find that $\alpha = 0.5$ and $\beta = 0.8$ are good choices that makes it easy to select informative priors. This is because the empirical quantiles close to the median have less variance. We have also experienced that `R-INLA` is more numerically stable when the spread is small, i.e. β is large.

3.2 Models

Let $y_t(\boldsymbol{s})$ denote the maximum precipitation at location $\boldsymbol{s} \in \mathcal{S}$ during year $t \in \mathcal{T}$, where $\mathcal{S}$ is the study area and $\mathcal{T}$ is the time period in focus. We assume a bGEV distribution for the yearly precipitation maxima,

$$[y_t(\boldsymbol{s}) \mid \mu_\alpha(\boldsymbol{s}), \sigma_\beta(\boldsymbol{s}), \xi(\boldsymbol{s})] \sim \text{bGEV}(\mu_\alpha(\boldsymbol{s}), \sigma_\beta(\boldsymbol{s}), \xi(\boldsymbol{s})),$$

where all observations are assumed to be conditionally independent given the parameters $\mu_\alpha(\boldsymbol{s})$, $\sigma_\beta(\boldsymbol{s})$ and $\xi(\boldsymbol{s})$. Correct estimation of the tail parameter is a difficult problem which highly affects estimates of large quantiles. The tail parameter is assumed to be constant, i.e. $\xi(\boldsymbol{s}) = \xi$. As discussed in Section 3.1, this is a common procedure, as inference for ξ is difficult with little data. The tail parameter is further restricted such that $\xi < 0.5$, resulting in a finite mean and variance for the yearly maxima. This restriction makes inference easier and more numerically stable. Exploratory analysis of our data supports the hypothesis of a spatially constant $\xi < 0.5$ and spatially varying $\mu_\alpha(\boldsymbol{s})$ and $\sigma_\beta(\boldsymbol{s})$ (results not shown). Two competing models are constructed for describing the spatial structure of $\mu_\alpha(\boldsymbol{s})$ and $\sigma_\beta(\boldsymbol{s})$.

The joint model

In the first model, denoted the joint model, both parameters are modelled using linear combinations of explanatory variables. Additionally, to draw strength from neighbouring stations, a spatial Gaussian random field is added to the location parameter. This gives the model

$$\begin{aligned} [y_t(\boldsymbol{s}) \mid \mu_\alpha(\boldsymbol{s}), \sigma_\beta(\boldsymbol{s}), \xi] &\sim \text{bGEV}(\mu_\alpha(\boldsymbol{s}), \sigma_\beta(\boldsymbol{s}), \xi), \\ \mu_\alpha(\boldsymbol{s}) &= \mathbf{x}_\mu(\boldsymbol{s})^T \boldsymbol{\beta}_\mu + u_\mu(\boldsymbol{s}), \\ \log\left(\sigma_\beta(\boldsymbol{s})\right) &= \mathbf{x}_\sigma(\boldsymbol{s})^T \boldsymbol{\beta}_\sigma, \end{aligned} \tag{3}$$

where $\mathbf{x}_\mu(\boldsymbol{s})$ and $\mathbf{x}_\sigma(\boldsymbol{s})$ are vectors containing an intercept plus the explanatory variables described in Table 1, and $\boldsymbol{\beta}_\mu$ and $\boldsymbol{\beta}_\sigma$ are vectors of regression

coefficients. The term $u_\mu(\boldsymbol{s})$ is a zero-mean Gaussian field with Matérn correlation function, i.e.,

$$\text{Corr}(u_\mu(\boldsymbol{s}_i), u_\mu(\boldsymbol{s}_j)) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\sqrt{8\nu} \frac{d(\boldsymbol{s}_i, \boldsymbol{s}_j)}{\rho} \right)^\nu K_\nu \left(\sqrt{8\nu} \frac{d(\boldsymbol{s}_i, \boldsymbol{s}_j)}{\rho} \right).$$

Here, $d(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is the Euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, $\rho > 0$ is the range parameter and $\nu > 0$ is the smoothness parameter. The function K_ν is the modified Bessel function of the second kind and order ν. The Matérn family is a widely used class of covariance functions in spatial statistics due to its flexible local behaviour and attractive theoretical properties (Matern, 1986; Stein, 1999; Guttorp & Gneiting, 2006). Its form also naturally appears as the covariance function of some models for the spatial structure of point rain rates (Sun et al., 2015). Efficient inference for high-dimensional Gaussian random fields can be achieved using the SPDE approach of Lindgren et al. (2011), which is implemented in `R-INLA`. It is common to fix the smoothness parameter ν instead of estimating it, as the parameter is difficult to identify from data. The SPDE approximation in `R-INLA` allows for $0 < \nu \leq 1$. We choose $\nu = 1$ as this reflects our beliefs about the smoothness of the underlying physical process. Additionally, Whittle (1954) argues that $\nu = 1$ is a more natural choice for spatial models than the less smooth exponential correlation function ($\nu = 1/2$), and $\nu = 1$ is also the most extensively tested value when using `R-INLA` with the SPDE approach (Lindgren & Rue, 2015).

The joint model is similar to the models of Davison et al. (2012), Geirsson et al. (2015), and Dyrrdal et al. (2015). However, they all place a Gaussian random field in the linear predictor for the log-scale and for the tail parameter. Within the `R-INLA` framework, it is not possible to model the spread or the tail using Gaussian random fields. Based on the amount of available data and the difficulty of estimating the spread and tail parameters, we also believe that the addition of a spatial Gaussian field in either parameter would simply complicate parameter estimation without any considerable contributions to model performance. Consequently, we do not include any Gaussian random field in the spread or tail of the bGEV distribution.

The two-step model

The second model is specifically tailored for sparse data with large block sizes. In such data-sparse situations, a large observation at a single location can be explained by a large tail parameter or a large spread parameter. In practice this might cause identifiability issues between $\sigma_\beta(\boldsymbol{s})$ and ξ, even though the parameters are identifiable in theory. In order to put a flexible model on the spread while avoiding such issues, Vandeskog et al. (2021) propose a model which borrows strength from the peaks over threshold method for separate modelling of $\sigma_\beta(\boldsymbol{s})$.

For some large enough threshold $x_{thr}(\boldsymbol{s})$, the distribution of sub-daily precipitation $X(\boldsymbol{s})$ larger than $x_{thr}(\boldsymbol{s})$ is assumed to follow a generalised Pareto distribution (Davison & Smith, 1990)

$$P(X(\boldsymbol{s}) > x_{thr}(\boldsymbol{s}) + x \mid X(\boldsymbol{s}) > x_{thr}(\boldsymbol{s})) = \left(1 + \frac{\xi x}{\zeta(\boldsymbol{s})}\right)^{-1/\xi},$$

with tail parameter ξ and scale parameter $\zeta(\boldsymbol{s}) = \sigma(\boldsymbol{s}) + \xi(x_{thr}(\boldsymbol{s}) - \mu(\boldsymbol{s}))$, where $\mu(\boldsymbol{s})$ and $\sigma(\boldsymbol{s})$ are the original GEV parameters from (1). Since ξ is assumed to be constant in space, all spatial variations in the bGEV distribution must stem from $\mu(\boldsymbol{s})$ or $\sigma(\boldsymbol{s})$. We therefore assume that the difference $x_{thr}(\boldsymbol{s}) - \mu(\boldsymbol{s})$ between the threshold and the location parameter is proportional to the scale parameter $\sigma(\boldsymbol{s})$. This assumption leads to the spread $\sigma_\beta(\boldsymbol{s})$ being proportional to the standard deviation of all observations larger than the threshold $x_{thr}(\boldsymbol{s})$. Based on this assumption, it is possible to model the spatial structure of the spread parameter independently of the location and tail parameter. Denote

$$\sigma_\beta(\boldsymbol{s}) = \sigma_\beta^* \cdot \sigma^*(\boldsymbol{s}),$$

with σ_β^* a standardising constant and $\sigma^*(\boldsymbol{s})$ the standard deviation of all observations larger than $x_{thr}(\boldsymbol{s})$ at location $\boldsymbol{s}$. Conditional on $\sigma^*(\boldsymbol{s})$, the block maxima can be standardised as

$$y_t^*(\boldsymbol{s}) = y_t(\boldsymbol{s})/\sigma^*(\boldsymbol{s}).$$

The standardised block maxima have a bGEV distribution with a constant spread parameter,

$$\left[y_t^*(\boldsymbol{s}) \mid \mu_\alpha^*(\boldsymbol{s}), \sigma_\beta^*, \xi\right] \sim \text{bGEV}(\mu_\alpha^*(\boldsymbol{s}), \sigma_\beta^*, \xi),$$

where $\mu_\alpha^*(\boldsymbol{s}) = \mu_\alpha(\boldsymbol{s})/\sigma^*(\boldsymbol{s})$. Consequently, the second model is divided into two steps. First, we model the standard deviation of large observations at all locations. Second, we standardise the block maxima observations and model the remaining parameters of the bGEV distribution. We denote this as the two-step model. The two-step model shares some similarities with regional frequency analysis (Dalrymple, 1960; Hosking & Wallis, 1997; Naveau et al., 2014; Carreau et al., 2016), which is a multi-step procedure where the data are standardised and pooled together inside homogeneous regions. However, we standardise the data differently and do not pool together data from different locations. Instead, we borrow strength from nearby locations by adding a spatial Gaussian random fields to our model and by keeping ξ constant for all locations.

The location parameter $\mu_\alpha^*(\boldsymbol{s})$ is modelled as a linear combination of explanatory variables $\mathbf{x}_\mu(\boldsymbol{s})$ and a Gaussian random field $u_\mu(\boldsymbol{s})$, just as

$\mu_\alpha(\boldsymbol{s})$ in the joint model (3). For estimation of $\sigma^*(\boldsymbol{s})$, the threshold $x_{thr}(\boldsymbol{s})$ is chosen as the 99% quantile of all observed precipitation at location $\boldsymbol{s}$. The precipitation observations larger than $x_{thr}(\boldsymbol{s})$ are declustered to account for temporal dependence, and only the cluster maximum of an exceedance is used for estimating $\sigma^*(\boldsymbol{s})$. This might sound counter-intuitive, as the aim of the two-step model is to use more data to simplify inference. However, even when only using the cluster maxima, inference is less wasteful than for the joint model. By using all threshold exceedances for estimating $\sigma^*(\boldsymbol{s})$, we would need to account for the dependence within exceedance clusters, which would add another layer of complexity to the modelling procedure. Consequently, we have chosen to not model the temporal dependence and only use the cluster-maxima for inference in this paper. To avoid high uncertainty from locations with few observations, $\sigma^*(\boldsymbol{s})$ is only computed at stations with more than three years of data. In order to estimate $\sigma^*(\boldsymbol{s})$ at locations with little or no observations, a linear regression model is used, where the logarithm of $\sigma^*(\boldsymbol{s})$ is assumed to have a Gaussian distribution,

$$[\log(\sigma^*(\boldsymbol{s})) \mid \eta(\boldsymbol{s}), \tau] \sim \mathcal{N}(\eta(\boldsymbol{s}), \tau^{-1}),$$

with precision τ and mean $\eta(\boldsymbol{s}) = \mathbf{x}_\sigma(\boldsymbol{s})^T \boldsymbol{\beta}_\sigma$. The estimated posterior mean from the regression model is then used as an estimator for $\sigma^*(\boldsymbol{s})$ at all locations. Consequently, the complete two-step model is given as

$$\begin{aligned}
[\log(\sigma^*(\boldsymbol{s})) \mid \eta(\boldsymbol{s}), \tau] &\sim \mathcal{N}(\eta(\boldsymbol{s}), \tau^{-1}), \\
\eta(\boldsymbol{s}) &= \mathbf{x}_\sigma(\boldsymbol{s})^T \boldsymbol{\beta}_\sigma, \\
\left[y_t^*(\boldsymbol{s}) \mid \mu_\alpha^*(\boldsymbol{s}), \sigma_\beta^*, \xi\right] &\sim \text{bGEV}(\mu_\alpha^*(\boldsymbol{s}), \sigma_\beta^*, \xi), \\
y_t^*(\boldsymbol{s}) &= y_t(\boldsymbol{s})/\sigma^*(\boldsymbol{s}), \\
\mu_\alpha^*(\boldsymbol{s}) &= \mathbf{x}_\mu(\boldsymbol{s})^T \boldsymbol{\beta}_\mu + u_\mu(\boldsymbol{s}).
\end{aligned} \tag{4}$$

Notice that the formulation of the two-step model makes it trivial to add more complex components for modelling the spread. One can, therefore, easily add a spatial Gaussian random field to the linear predictor of $\log(\sigma^*(\boldsymbol{s}))$ while still using the `R-INLA` framework for inference, which is not possible with the joint model. In Section 4 we perform modelling both with and without a Gaussian random field in the spread to test how it affects model performance.

The uncertainty in the estimator for $\sigma^*(\boldsymbol{s})$ is not propagated into the bGEV model for the standardised response, meaning that the estimated uncertainties from the two-step model are likely to be too small. This can be corrected with a bootstrapping procedure, where we draw B samples from the posterior of $\log(\sigma^*(\boldsymbol{s}))$ and estimate $(\mu_\alpha^*(\boldsymbol{s}), \sigma_\beta^*, \xi)$ for each of the B samples. Vandeskog et al. (2021) show that the two-step model with 100 bootstrap samples is able to outperform the joint model in a simple setting.

It might seem contradictory to employ a model based on exceedances in our setting, since we claim that the data quality is too bad to use the peaks over threshold model for estimating return levels. However, merely estimating the standard deviation of all threshold exceedances is a much simpler task than to estimate spatially varying parameters of the generalised Pareto distribution, including the tail parameter ξ. Thus, while we claim that the available data is not of good enough quality to estimate return levels in a similar fashion to Opitz et al. (2018), we also claim that it is of good enough quality to perform the simple task of estimating the trends in the spread parameter. The estimation of all remaining parameters, including ξ, is performed using block maxima data, which we believe to be of better quality.

3.3 INLA

By placing a Gaussian prior on $\boldsymbol{\beta}_\mu$, both the joint and the two-step models fall into the class of latent Gaussian models. This is advantageous as it allows for inference using INLA with the `R-INLA` library (Rue et al., 2009; Bivand et al., 2015; Rue et al., 2017). The extreme value framework is quite new to the `R-INLA` package. Still, in recent years, some papers have started to appear where it is used for modelling extremes with INLA (e.g. Opitz et al., 2018; Castro-Camilo et al., 2019). `R-INLA` includes an implementation of the SPDE approximation for Gaussian random fields with a Matérn correlation function, which is used on the random field $u_\mu(\boldsymbol{s})$ for a considerable improvement in inference speed.

A requirement for using INLA is that the model likelihood is log-concave. Unfortunately, neither the GEV distribution nor the bGEV distribution have log-concave likelihoods when $\xi > 0$. This can cause severe problems for model inference. However, we find that these problems are mitigated by choosing slightly informative priors for the model parameters, which is possible because of the reparametrisation described in Section 3.1. Additionally, we find that `R-INLA` is more stable when given a response that is standardised such that the difference between its 95% quantile and its 5% quantile is equal to 1. Based on the authors' experience, similar standardisation of the response is also a common procedure when using INLA for estimating the Weibull distribution parameters within the field of survival analysis. We believe that the combination of slightly informative priors and standardisation of the response is enough to fix the problems of non-concavity and ensure that `R-INLA` is working well with the bGEV distribution.

3.4 Evaluation

Model performance can be evaluated using the continuous ranked probability score (CRPS; Matheson & Winkler, 1976; Gneiting & Raftery, 2007; Friederichs & Thorarinsdottir, 2012),

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(t) - I(t \geq y))^2 \mathrm{d}t = 2 \int_0^1 \ell_p \left(y - F^{-1}(p)\right) \mathrm{d}p, \quad (5)$$

where F is the forecast distribution, y is an observation, $I(\cdot)$ is an indicator function and $\ell_p(x) = x(p - I(x < 0))$ is the quantile loss function. The CRPS is a strictly proper scoring rule, meaning that the expected value of $\text{CRPS}(F, y)$ is minimised for $G = F$ if and only if $y \sim G$. The importance of proper scoring rules when forecasting extremes is discussed by Lerch et al. (2017). From (5), one can see that the CRPS is equal to the integral over the quantile loss function for all possible quantiles. However, we are only interested in predicting large quantiles, and the model performance for small quantiles is of little importance to us. The threshold weighted CRPS (twCRPS; Gneiting & Ranjan, 2011) is a modification of the CRPS that allows for emphasis on specific areas of the forecast distribution,

$$\text{twCRPS}(F, y) = 2 \int_0^1 \ell_p \left(y - F^{-1}(p)\right) w(p) \mathrm{d}p, \quad (6)$$

where $w(p)$ is a non-negative weight function. A possible choice of $w(p)$ for focusing on the right tail is the indicator function $w(p) = I(p > p_0)$. As described by Bolin and Wallin (2023), the mean twCRPS is not robust to outliers and it gives more weight to forecast distributions with large variances, i.e. at locations far away from any weather station. A scaled version of the twCRPS, denoted the StwCRPS, is created using theorem 5 of Bolin and Wallin (2023):

$$S_{\text{scaled}}(F, y) = \frac{S(F, y)}{|S(F, F)|} + \log\left(|S(F, F)|\right), \quad (7)$$

where $S(F, y)$ is the twCRPS and $S(F, F)$ is its expected value with respect to the forecast distribution,

$$S(F, F) = \int S(F, y) \mathrm{d}F(y).$$

The mean StwCRPS is more robust to outliers and varying degrees of uncertainty in forecast distributions, while still being a proper scoring rule (Bolin & Wallin, 2023).

Using `R-INLA` we are able to sample from the posterior distribution of the bGEV parameters at any location $\boldsymbol{s}$. The forecast distribution at

location $\boldsymbol{s}$ is therefore given as

$$\widehat{F}_{\boldsymbol{s}}(\cdot) = \frac{1}{m}\sum_{i=1}^{m} F(\cdot; \mu_{\alpha}^{(i)}(\boldsymbol{s}), \sigma_{\beta}^{(i)}(\boldsymbol{s}), \xi^{(i)}), \tag{8}$$

where F is the distribution function of the bGEV distribution and the samples $(\mu_{\alpha}^{(i)}(\boldsymbol{s}), \sigma_{\beta}^{(i)}(\boldsymbol{s}), \xi^{(i)})$ are drawn from the posterior distribution of the bGEV parameters for $i = 1, \ldots, m$, where m is a multiple of the number B of bootstrap samples. A closed-form expression is not available for the twCRPS when using the forecast distribution from (8). Consequently, we evaluate the twCRPS and StwCRPS using numerical integration.

4 Modelling sub-daily precipitation extremes in Norway

The models from Section 3 are applied for estimating return levels in the south of Norway. Table 1 shows which explanatory variables are used for modelling the location and spread parameters in both models. All explanatory variables are standardised to have zero mean and a standard deviation of 1, before being applied for modelling. Inference for the two-step model is performed both with and without propagation of the uncertainty in $\sigma^*(\boldsymbol{s})$. The uncertainty propagation is achieved using 100 bootstrap samples, as described in Section 3.2. Additionally, we modify the two-step model and add a random Gaussian field $u_\sigma(\boldsymbol{s})$ to the linear predictor of the log-spread, to test if this can yield any considerable improvement in model performance. Just as $u_\mu(\boldsymbol{s})$, $u_\sigma(\boldsymbol{s})$ has zero mean and a Matérn covariance function.

4.1 Prior selection

Priors must be specified before we can model the precipitation extremes. From construction, the location parameter μ_α is equal to the α quantile of the bGEV distribution. This allows us to place a slightly informative prior on $\boldsymbol{\beta}_\mu$, using quantile regression on $y^*(\boldsymbol{s})$ (Koenker, 2005, 2018). We choose a Gaussian prior for $\boldsymbol{\beta}_\mu$, centred at the α quantile regression estimates and with a precision of 10. There is no unit on the precision in $\boldsymbol{\beta}_\mu$ because the block maxima have been standardised, as described in Section 3.3. The regression coefficients $\boldsymbol{\beta}_\sigma$ differ between the two-step and joint models. In the joint model, all the coefficients in $\boldsymbol{\beta}_\sigma$, minus the intercept coefficient, are given Gaussian priors with zero mean and a precision of 10^{-3}. The intercept coefficient, here denoted $\beta_{0,\sigma}$, is given a log-gamma prior with parameters such that $\exp(\beta_{0,\sigma})$ has a gamma prior with mean equal to the empirical difference between the $1-\beta/2$ quantile and the $\beta/2$ quantile of the standardised block maxima. The precision of the gamma prior is 10.

In the two-step model, all coefficients of $\boldsymbol{\beta}_\sigma$ are given Gaussian priors with zero mean and a precision of 10^{-3}, while the logarithm of σ^*_β is given the same log-gamma prior as the intercept coefficient in the joint model.

The parameters of the Gaussian random fields u_μ and u_σ are given penalised complexity (PC) priors. The PC prior is a weakly informative prior distribution, designed to punish model complexity by placing an exponential prior on the distance from some base model (Simpson et al., 2017). Fuglstad et al. (2019) develop a joint PC prior for the range $\rho > 0$ and standard deviation $\zeta > 0$ of a Gaussian random field, where the base model is defined to have infinite range and zero variance. The prior contains two penalty parameters, that can be decided by specifying the four parameters ρ_0, α_1, ζ_0 and α_2 such that $P(\rho < \rho_0) = \alpha_1$ and $P(\zeta > \zeta_0) = \alpha_2$. We choose $\alpha_1 = \alpha_2 = 0.05$. ρ_0 is given a value of 75 km for both the random fields, meaning that we place a 95% probability on the range being larger than 75 km. To put this range into context, the study area has a dimension of approximately 730×460 km^2, and the mean distance from one station to its closest neighbour is 10 km. ζ_0 is given a value of 0.5 mm for u_σ, meaning that we place a 95% probability on the standard deviation being smaller than 0.5 mm. This seems to be a reasonable value because the estimated logarithm of $\sigma^*(\boldsymbol{s})$ lies in the range between 0.1 mm and 3.5 mm for all available weather stations and all examined aggregation times. For u_μ we set $\zeta_0 = 0.5$, which is a reasonable value because of the standardisation of the response described in Section 3.3.

A PC prior is also placed on the tail parameter ξ. Opitz et al. (2018) develop a PC prior for the tail parameter of the generalised Pareto distribution, which is the default prior for ξ in `R-INLA` when modelling with the bGEV distribution. However, to the best of our knowledge, expressions for the PC priors for ξ in the GEV or bGEV distributions are not previously available in the literature. In the supplementary material we develop expressions for the PC prior of $\xi \in [0, 1)$ with base model $\xi = 0$ for the GEV distribution and the bGEV distribution. Closed-form expressions do not exist, but the priors can be approximated numerically. Having computed the PC priors for the GEV distribution and the bGEV distribution, we find that they are similar to the PC prior of the generalised Pareto distribution, which has a closed-form expression and is already implemented in `R-INLA`. Consequently, we choose to model the tail parameter of the bGEV distribution with the PC prior for the generalised Pareto distribution (Opitz et al., 2018):

$$\pi(\xi) = \frac{\lambda}{\sqrt{2}} \exp\left(-\frac{\lambda}{\sqrt{2}} \frac{\xi}{(1-\xi)^{1/2}}\right) \left(\frac{1-\xi/2}{(1-\xi)^{3/2}}\right),$$

with $0 \leq \xi < 1$ and penalty parameter λ. Even though the prior is defined for values of ξ up to 1, a reparametrisation is performed within `R-INLA`

such that $0 \leq \xi < 0.5$. Since the base model has $\xi = 0$, the prior places more weight on small values of ξ when λ increases. Based on the plots in Figure S2.1 in the supplementary material, we find a value of $\lambda = 7$ to give a good description of our prior beliefs, as we expect ξ to be positive but small.

4.2 Cross-validation

Model performance is evaluated using five-fold cross-validation with the StwCRPS. The StwCRPS weight function is chosen as $w(p) = I(p > 0.9)$. Both in-sample and out-of-sample performance are evaluated. The mean StwCRPS over all five folds are displayed in Table 2. The two-step model outperforms the joint model for all aggregation times. This implies that information about threshold exceedances can provide valuable information when modelling block maxima. When performing in-sample estimation, the variant of the two-step model with a Gaussian field and without bootstrapping always outperforms the other contestants. However, during out-of-sample estimation, the model performs worse than its competitors. This indicates a tendency to overfit when not using bootstrapping to propagate uncertainty in $\sigma^*(\boldsymbol{s})$ into the estimation of $(\mu^*_\alpha(\boldsymbol{s}), \sigma^*_\beta, \xi)$. The two variants of the two-step model that use bootstrapping perform best during out-of-sample estimation. While their model fits yield similar scores, their difference in complexity is quite considerable, as one model contains two spatial random fields, and the other only contains one. This shows that there is little need for placing an overly complex model on the spread parameter. Consequently, for estimation of the bGEV parameters and return levels, we choose to use the two-step model with bootstrapping and without a spatial Gaussian random field in the spread.

4.3 Parameter estimates

The parameters of the two-step model are estimated for different aggregation times between 1 and 24 hours. Uncertainty is propagated using $B = 100$ bootstrap samples. Estimation of the posterior of $(\mu^*_\alpha(\boldsymbol{s}), \sigma^*_\beta, \xi)$ given some value of $\sigma^*(\boldsymbol{s})$ takes less than 2 minutes on a 2.4 gHz laptop with 16 GB RAM, and the 100 bootstraps can be computed in parallel. On a moderately sized computational server, inference can thus be performed in well under 10 minutes.

The estimated values of the regression coefficients $\boldsymbol{\beta}_\mu$ and $\boldsymbol{\beta}_\sigma$, the spread σ^*_β and the standard deviation of the Gaussian field $u_\mu(\boldsymbol{s})$ for the standardised precipitation maxima, are displayed in Table 3 for some selected temporal aggregations. These estimates are computed by drawing 20 samples from each of the 100 posterior distributions. The empirical mean,

Table 2: *Mean StwCRPS with weight function* $w(p) = I(p > 0.9)$ *for 5-fold cross-validation performed using out-of-sample estimation and in-sample estimation. The two-step method is tested with and without a Gaussian field in the spread and bootstrapping for propagation of uncertainty. Cross-validation is performed for precipitation aggregated over periods of 1 hour, 3 hours, 6 hours, 12 hours and 24 hours. The best scores are written in* ***bold****.*

	Model	$u_\sigma(\boldsymbol{s})$	Boot-strap	1 h	3 h	6 h	12 h	24 h
Out-of-sample	Joint			−0.872	−0.597	−0.412	−0.149	0.0500
	Two-step	✓	✓	**−0.905**	**−0.603**	−0.427	−0.207	**0.0425**
	Two-step	✓		−0.893	−0.585	−0.414	−0.197	0.0635
	Two-step		✓	−0.876	−0.594	**−0.429**	**−0.211**	0.0456
	Two-step			−0.872	−0.584	−0.417	−0.196	0.0674
In-sample	Joint			−0.876	−0.608	−0.445	−0.230	0.0206
	Two-step	✓	✓	−1.004	−0.713	−0.564	−0.328	−0.1066
	Two-step	✓		**−1.012**	**−0.721**	**−0.577**	**−0.333**	**−0.1161**
	Two-step		✓	−0.889	−0.607	−0.453	−0.247	−0.0244
	Two-step			−0.886	−0.607	−0.454	−0.243	−0.0182

standard deviation and quantiles of these 2000 samples are then reported. There is strong evidence that all the explanatory variables in $\mathbf{x}_\sigma(\boldsymbol{s})$ are affecting the spread, with the northing being the most important explanatory variable. There is considerably less evidence that all our chosen explanatory variables have an effect on the location parameter. However, as the posterior distribution of $\boldsymbol{\beta}_\mu$ is estimated using 100 different samples from the posterior of $\sigma^*(\boldsymbol{s})$, it might be that the different regression coefficients are more significant for some of the standardisations, and less significant for others. The explanatory variable that has the greatest effect on the location parameter seems to be the mean annual precipitation. Thus, at locations with large amounts of precipitation, we expect the extreme precipitation to be heavier than at locations with little precipitation. From the estimates for $\boldsymbol{\beta}_\sigma$, we also expect more variance in the distribution of extreme precipitation in the south. The standard deviation of $u_\mu(\boldsymbol{s})$ is of approximately the same magnitude as most of the regression coefficients in $\boldsymbol{\beta}_\mu$.

Table 4 displays the posterior range of the $u_\mu(\boldsymbol{s})$. For the available data, the median number of neighbours within a radius of 50 km is 17, and the median number of neighbours within a radius of 100 km is 36. Based on these numbers, one can see that the Gaussian field is able to introduce spatial correlation between a large number of different stations. The range of the Gaussian field is considerably reduced as the temporal aggregation

Table 3: *Estimated regression coefficients $\boldsymbol{\beta}_\mu$, $\boldsymbol{\beta}_\sigma$ and estimated standard deviation $SD(u_\mu)$ of the Gaussian field $u_\mu(\boldsymbol{s})$ in the two-step model for yearly maximum precipitation at different temporal aggregations.*

Temporal aggregation	Parameter	Explanatory variable	Mean	SD	2.5% quantile	50% quantile	97.5% quantile
1 hour	$\boldsymbol{\beta}_\mu$	Intercept	0.669	0.066	0.518	0.675	0.794
		Mean annual precipitation	0.071	0.012	0.044	0.072	0.091
		Altitude	−0.006	0.011	−0.026	−0.006	0.013
		Easting	−0.044	0.035	−0.118	−0.042	0.017
		Northing	0.036	0.038	−0.036	0.036	0.111
		Distance to the open sea	−0.001	0.033	−0.072	0.002	0.055
	SD(u_μ)		0.084	0.023	0.047	0.085	0.127
	σ^*_β		0.118	0.002	0.113	0.118	0.123
	$\boldsymbol{\beta}_\sigma$	Intercept	2.836	0.007	2.822	2.831	2.835
		Easting	0.088	0.015	0.058	0.078	0.087
		Northing	−0.150	0.015	−0.177	−0.160	−0.151
		Distance to the open sea	−0.050	0.017	−0.080	−0.062	−0.051
3 hours	$\boldsymbol{\beta}_\mu$	Intercept	0.847	0.025	0.800	0.846	0.895
		Mean annual precipitation	0.120	0.012	0.096	0.122	0.139
		Altitude	−0.009	0.009	−0.027	−0.010	0.007
		Easting	0.016	0.021	−0.025	0.017	0.057
		Northing	0.022	0.018	−0.015	0.024	0.054
		Distance to the open sea	0.017	0.018	−0.021	0.017	0.052
	SD(u_μ)		0.062	0.013	0.045	0.060	0.085
	σ^*_β		0.123	0.003	0.117	0.123	0.129
	$\boldsymbol{\beta}_\sigma$	Intercept	3.207	0.017	3.177	3.190	3.208
		Easting	0.031	0.014	0.003	0.021	0.030
		Northing	−0.138	0.014	−0.163	−0.147	−0.138
		Distance to the open sea	−0.057	0.016	−0.083	−0.068	−0.058
6 hours	$\boldsymbol{\beta}_\mu$	Intercept	0.911	0.024	0.868	0.911	0.962
		Mean annual precipitation	0.148	0.013	0.123	0.150	0.168
		Altitude	−0.012	0.009	−0.029	−0.011	0.005
		Easting	0.040	0.021	−0.001	0.040	0.080
		Northing	0.008	0.019	−0.029	0.009	0.042
		Distance to the open sea	0.030	0.020	−0.011	0.031	0.067
	SD(u_μ)		0.067	0.011	0.051	0.065	0.090
	σ^*_β		0.125	0.004	0.118	0.125	0.133
	$\boldsymbol{\beta}_\sigma$	Intercept	3.470	0.020	3.435	3.456	3.469
		Easting	−0.032	0.015	−0.062	−0.042	−0.033
		Northing	−0.095	0.014	−0.122	−0.105	−0.096
		Distance to the open sea	−0.065	0.016	−0.093	−0.077	−0.066

Table 4: *Estimated posterior mean and quantiles for the range ρ of the Gaussian field $u_\mu(\boldsymbol{s})$ and the tail parameter ξ in the two-step model for yearly maximum precipitation at different temporal aggregations.*

Parameter	Temporal aggregation	Mean	2.5% quantile	50% quantile	97.5% quantile
ρ	1 hour	235	34	255	478
	3 hours	78	39	75	147
	6 hours	60	32	57	104
	12 hours	84	31	83	145
	24 hours	55	32	50	105
ξ	1 hour	0.178	0.136	0.179	0.211
	3 hours	0.090	0.057	0.089	0.120
	6 hours	0.047	0.028	0.046	0.072
	12 hours	0.032	0.010	0.031	0.048
	24 hours	0.029	0.006	0.029	0.051

increases. It seems that, for 1 hour precipitation, the regression coefficients are unable to explain some kind of large-scale phenomenon that considerably affects the location parameter $\mu_\alpha(\boldsymbol{s})$. To correct this, the range of $u_\mu(\boldsymbol{s})$ has to be large. For longer aggregation periods, this phenomenon is not as important anymore, and the regression coefficients are able to explain most of the large-scale trends. Consequently, the range of $u_\mu(\boldsymbol{s})$ is decreased. The posterior means of $u_\mu(\boldsymbol{s})$ for three different temporal aggregations are displayed over a 1×1 km^2 gridded map in Figure 2. It is known that extreme precipitation dominates in the southeast of Norway for short aggregation times because of its large amount of convective precipitation (see, e.g, Dyrrdal et al., 2015). Based on Figure 2 it becomes evident that our explanatory variables are unable to describe this regional difference when modelling hourly precipitation, and $u_\mu(\boldsymbol{s})$ has to do the job of separating between the east and the west. As the temporal aggregations increase from one hour to three and six hours, the difference between east and west diminishes, and it seems that the explanatory variables do a better job of explaining the trends in the location parameter $\mu_\alpha(\boldsymbol{s})$.

The posterior distribution of ξ is also described in Table 4. The tail parameter seems to decrease quickly as the aggregation time increases, and it is practically constant for precipitation over longer periods than 12 hours.

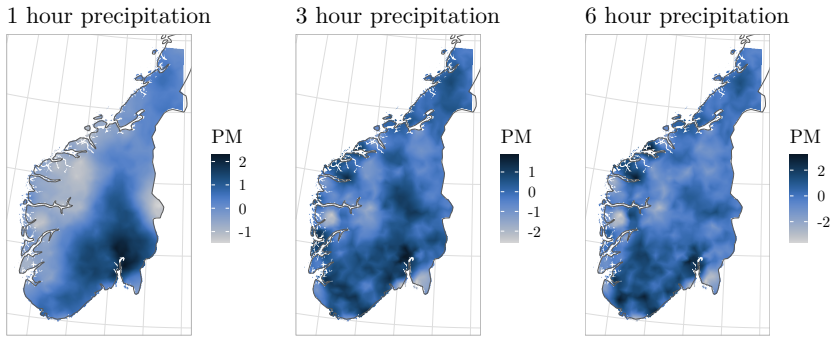


Figure 2: *Estimated posterior mean (PM) of the Gaussian field* $u_\mu(\boldsymbol{s})$ *for three different temporal aggregations of precipitation, with unit mm.*

This makes sense given the observation of Barbero et al. (2019) that most 24 hour annual maximum precipitation comes from rainstorms with lengths of less than 15 hours. Thus, the tail parameter for 24 hour precipitation should be close to the tail parameter for 12 hour precipitation. For 12 hours and up, the tail parameter is so small that one may wonder if a Gumbel distribution would not have given a better fit to the data. However, this is not the case for the shorter aggregation times, where the tail parameter is considerably larger.

4.4 Return levels

We use the two-step model for estimating large return levels for the yearly precipitation maxima. Posterior distributions of the 20 year return levels are estimated on a grid with resolution 1×1 km^2. The posterior means and the widths of the 95% credible intervals are displayed in Figure 3. For a period of 1 hour the most extreme precipitation is located southeast in Norway, while for longer periods, the extreme precipitation is moving over to the west coast. These results are expected since we know that the convective precipitation of the southeast dominates for short aggregation periods. At the same time, the southwest of Norway generally has more precipitation, making it the dominant region for longer aggregation times. The spatial structure and magnitude of the 20 year return levels for hourly precipitation are similar to the estimates of Dyrrdal et al. (2015), but with considerably thinner credible intervals. This makes sense as more data are available, and the two-step model is able to perform less wasteful inference. In addition, our model is much more simple, as they include a random Gaussian field in all three parameters, while we only include a random Gaussian field in

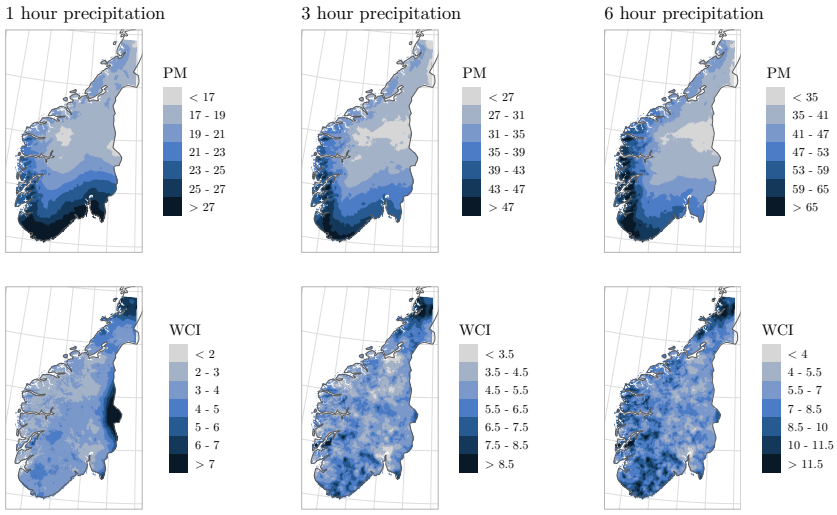


Figure 3: *Estimated posterior mean (PM) and width of the* 95% *credible intervals (WCI) for the 20 year return levels of sub-daily precipitation. Different aggregation times are displayed in different columns. All numbers are given with unit mm.*

the location parameter. This can also lead to less uncertainty in the return level estimates.

5 Conclusion

The blended generalised extreme value (bGEV) distribution is applied as a substitute for the generalised extreme value (GEV) distribution for estimation of the return levels of sub-daily precipitation in the south of Norway. The bGEV distribution simplifies inference by introducing a parameter-free support, but can only be applied for modelling of heavy-tailed phenomena. Sub-daily precipitation maxima are modelled using a spatial Bayesian hierarchical model with a latent Gaussian field. This is implemented using both integrated nested Laplace approximations (INLA) and the stochastic partial differential equation (SPDE) approach, for fast inference. Inference is also made more stable and less wasteful by our novel two-step modelling procedure that borrows strength from the peaks over threshold method when modelling block maxima. Like the GEV distribution, the bGEV distribution suffers from a lack of log-concavity, which can cause problems when using INLA. We are able to mitigate any problems caused by a lack of log-concavity by choosing slightly informative priors and standardising the data. We find that the bGEV distribution performs well as a model for extreme precipitation. The two-step model

successfully utilises the additional information provided by the peaks over threshold data and is able to outperform models that only use block maxima data for inference.

Acknowledgements

We thank Thordis L. Thorarinsdottir and Geir-Arne Fuglstad for helpful discussions.

Funding

This publication is part of the World of Wild Waters (WoWW) project, which falls under the umbrella of Norwegian University of Science and Technology (NTNU)'s Digital Transformation initiative.

Code and data availability

The necessary code and data for achieving these results are available online at https://github.com/siliusmv/inlaBGEV. The unprocessed data are freely available online, as described in Section 2.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Supplementary material

S1 Simulation study

A simulation study is conducted for comparing the performance of the generalised extreme value (GEV) distribution and blended generalised extreme value (bGEV) distribution in a univariate setting. Our penultimate goal when modelling block maxima is to estimate return levels. Thus, the two distributions are compared by evaluating the performance of their return level estimators.

We sample $n \in \{25, 50, 100, 500, 1000\}$ i.i.d. "block maxima" from a GEV distribution resembling the estimated distribution of the yearly maxima of hourly precipitation. From Table 3 and Table 4 in "Modelling sub-daily precipitation extremes with the blended generalised extreme value distribution" we find that the posterior mean of the intercept in $\boldsymbol{\beta}_\mu$ is

$E[\beta_{\mu,0}] = 0.661$, and the posterior mean of the intercept in $\boldsymbol{\beta}_\sigma$ is $E[\beta_{\sigma,0}] = 2.835$. Consequently, we set $\alpha = 0.5$, $\beta = 0.8$, $\mu_\alpha = E[\beta_{\mu,0}] \cdot \exp(E[\beta_{\sigma,0}]) \approx 11.26$, and $\sigma_\beta = E[\sigma_\beta^*] \cdot \exp(E[\beta_{\sigma,0}]) \approx 2.01$. The tail parameter is set equal to the posterior mean from Table 4: $\xi = 0.178$. Using the n block maxima, we compute maximum likelihood estimators for the parameters of the GEV and bGEV distributions, which are further used for computing return level estimators for periods of length 25, 50, 100, 250 and 500. The actual maximum likelihood estimation is performed on the (μ, σ, ξ) parametrisation, in which $\mu \approx 10.05$ and $\sigma \approx 3.21$. Optimisation is performed using the `optim` function in `R`, where the true values of (μ, σ, ξ) are used as initial values. The entire procedure is repeated 500 times for each value of n. This allows us to estimate the distribution of the maximum likelihood estimators for all return levels in question.

Figure S1 displays the distribution of all return level estimators using both the GEV and the bGEV distributions. The difference between the distributions of the GEV estimators and the bGEV estimators are negligible for small values of n. For large return periods and values of n some slight differences are appearing. However, in practical situations one rarely encounters as much as 1000 block maxima, and the blocks are almost never large enough for the limit distribution to hold exactly. Thus, the differences in distribution can be considered negligible for large values of n for almost all real-life settings. Larger differences in model fit could theoretically occur for small blocks where the block maxima deviate further from the GEV distribution. However, in such settings, both the GEV and the bGEV distribution would be misspecified, and we have no reason to believe that the misspecification in the GEV distribution would be strictly better or worse than the misspecification in the bGEV distribution. Therefore, our results imply that we do not lose anything in performance by modelling GEV data with the bGEV distribution instead of the GEV distribution.

In most real-life settings we do not have as much prior knowledge about the true underlying distribution, and we will likely choose less optimal initial values and/or prior distributions. To examine the effects of choosing less optimal initial values, we repeat the maximum likelihood estimation on the exact same data used for Figure S1, using different initial values. Figure S2 displays the distributions of the return level estimators when we use an initial value of 0.9 instead of 3.2 for σ, without changing the initial values for μ or ξ. For small values of n the two distributions yield comparable results. However, for large values of n, the performance of the GEV distribution quickly deteriorates, and the return level estimators are shifted to the left. The bGEV estimators, however, seem to have identical distributions in Figures S1 and S2, and are not affected by the slight change of initial values. This demonstrates how inference with the bGEV distribution is more robust than inference with the GEV distribution. The parameter-

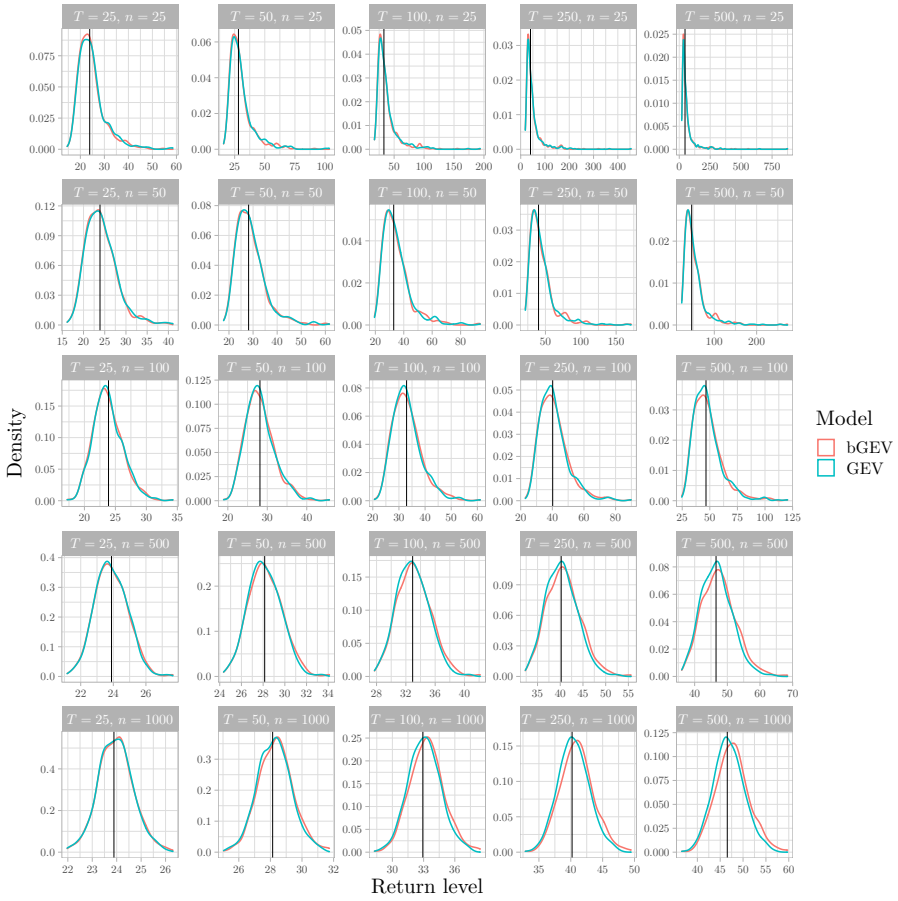


Figure S1: *Distributions of return level estimators computed from model fits with the GEV and bGEV distributions to simulated data from a GEV distribution. The dark vertical lines display the true return levels. Return periods are denoted as T, with unit "block sizes", and the number of block maxima used for fitting the GEV and bGEV distributions is denoted as n.*

dependent support of the GEV complicates likelihood-based inference and makes it more difficult to locate the global maximum of the likelihood, even in simple and univariate settings with unrealistically large amounts of data that are perfectly GEV-distributed.

In such a simple setting, it is relatively easy to see that we chose bad initial values for (μ, σ, ξ). However, when modelling block maxima with a large number of covariates and in high-dimensional settings, it becomes considerably more difficult to choose good enough initial values and/or prior distributions. Additionally, it is not trivial to examine a maximum likelihood estimator or posterior distribution and find out if the inference procedure was successful or not. The GEV return level estimators in Figure S2 take on reasonable values and are large enough to seem realistic. Without knowing the truth, it would not be trivial for a practitioner to detect that they are faulty, and caused by numerical problems. Thus, if one performs likelihood-based inference using the GEV distribution, one should always perform extensive checks to evaluate whether the fit is reasonable, or if it is caused by numerical problems. The increased robustness of the bGEV distribution provides more safety, and most practitioners can be more confident in their model fits after fitting the bGEV distribution to block maxima data instead of the GEV distribution.

Similarly, while attempting to model return levels for precipitation in Norway using the `gev` family in `R-INLA`, we experienced that e.g. adding or removing one covariate from the model could completely alter the model fits if our priors were too wide or the data were not correctly standardised. Examining which of the fits were the best required comprehensive and time-demanding cross-validation studies, and it was difficult to find out if the best model fit was "good enough", or if a slight change in priors or covariates would yield considerably better model fits. These problems were considerably mitigated by switching to the `bgev` family, which resulted in more stable inference using less informative priors, and less optimal initial values.

S2 PC prior for the tail parameter

In order to compute the PC prior for the GEV and bGEV distribution with $\xi \geq 0$ and base model $\xi = 0$, one must first compute $\mathrm{KLD}(\pi_\xi, \pi_0)$ for the GEV distribution and the bGEV distribution. Notice that the base model is identical for both distributions. Writing the GEV distribution function as $F(x) = \exp(-h(x))$ gives the probability density function

$$\pi_\xi(x) = -\exp(-h(x))h'(x), \quad h(x) = (1 + \xi(x-\mu)/\sigma)_+^{-1/\xi}.$$

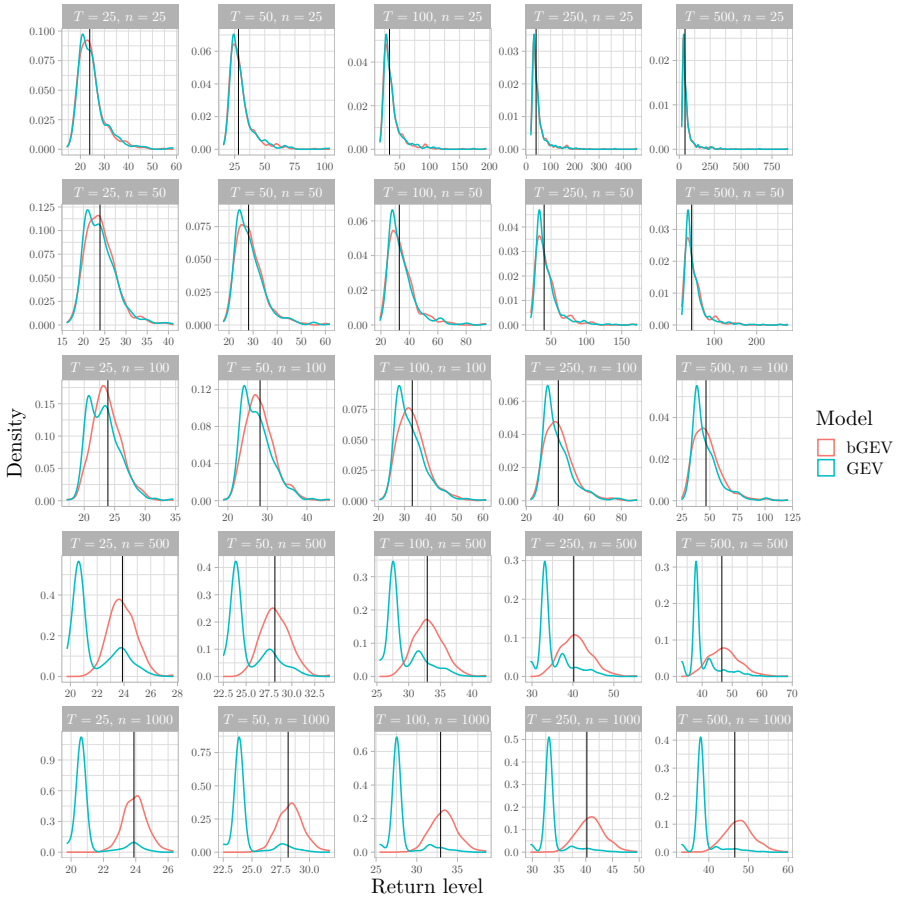


Figure S2: *Distributions of return level estimators computed from model fits with the GEV and bGEV distributions to simulated data from a GEV distribution, when using bad initial values. The dark vertical lines display the true return levels. Return periods are denoted as T, with unit "block sizes", and the number of block maxima used for fitting the GEV and bGEV distributions is denoted as n.*

The KLD for the GEV distribution is equal to

$$\begin{aligned}\mathrm{KLD}(\pi_\xi, \pi_0) &= \int \pi_\xi(x) \log\left(\frac{\pi_\xi(x)}{\pi_0(x)}\right) \mathrm{d}x \\ &= \int -e^{-h(x)}h'(x)\left(-h(x) + \log(-\sigma h'(x)) + \exp\left(-\frac{x-\mu}{\sigma}\right) + \frac{x-\mu}{\sigma}\right) \mathrm{d}x.\end{aligned}$$

The fourth term of the KLD is simply equal to the expectation of $(x-\mu)/\sigma$, which is known. The first term is easily solvable with the substitution $u = h(x)$. Using the same substitution for the second term with the knowledge that $\log(-\sigma h'(x)) = (1+\xi)\log(h(x))$ gives the integral

$$\int -e^{-h(x)}h'(x)\log(-\sigma h'(x)) = \int_0^\infty e^{-u}(1+\xi)\log(u)\ \mathrm{d}u = -(1+\xi)\gamma,$$

where γ is the Euler-Mascheroni constant. We are unable to find a closed-form expression for the third term. However, using substitution with $u = h(x)$ gives

$$\int -e^{-h(x)}h'(x)\exp\left(-\frac{x-\mu}{\sigma}\right)\mathrm{d}x = \int_0^\infty \exp\left(\frac{1}{\xi} - u - \frac{u^{-\xi}}{\xi}\right)\mathrm{d}u,$$

which is finite for $\xi > 0$. With a change of variables, this integral is transformed to have finite limits,

$$\int_0^\infty \exp\left(\frac{1}{\xi} - \left(u + u^{-\xi}/\xi\right)\right)\mathrm{d}u = \int_0^1 \exp\left(\frac{1}{\xi}\left(1 - (-\log v)^{-\xi}\right)\right)\mathrm{d}v.$$

This can easily be numerically approximated. Summarising, the KLD for the GEV is finite for $0 \le \xi < 1$ and equal to

$$\begin{aligned}\mathrm{KLD}(\pi_\xi, \pi_0) = &-1 - (1+\xi)\gamma + \xi^{-1}(\Gamma(1-\xi) - 1) \\ &+ \int_0^1 \exp\left(\frac{1}{\xi}\left(1 - (-\log v)^{-\xi}\right)\right)\mathrm{d}v.\end{aligned} \tag{S1}$$

The PC prior has probability density function $\pi(d) = \lambda\exp(-\lambda d)$ with $d = \sqrt{2\mathrm{KLD}(\pi_\xi, \pi_0)}$. Transforming the PC prior from a distribution on d to a distribution on ξ gives

$$\begin{aligned}\pi(\xi) &= \lambda\exp\left(-\lambda d(\xi)\right)\left|\frac{\partial d(\xi)}{\partial \xi}\right| \\ &= \frac{\lambda}{d(\xi)}\exp\left(-\lambda d(\xi)\right)\left|\frac{\partial}{\partial \xi}\mathrm{KLD}(\pi_\xi, \pi_0)\right|\end{aligned} \tag{S2}$$

Consequently, the derivative of the KLD must also be computed. Using derivation under the integral sign gives

$$\frac{\partial}{\partial \xi}\mathrm{KLD}(\pi_\xi, \pi_0) = -\gamma - \frac{\Gamma(1-\xi)\Psi(1-\xi)}{\xi} - \frac{\Gamma(1-\xi)-1}{\xi^2} + \int_0^1 g(v;\xi)\mathrm{d}v,$$

where $\Psi(\cdot)$ is the digamma function, and

$$g(v;\xi) = \exp\left(\frac{1-(-\log v)^{-\xi}}{\xi}\right)\frac{-1+(-\log v)^{-\xi}\,(1+\xi\log(-\log v))}{\xi^2}.$$

This integral must also be numerically approximated.

When computing the KLD for the bGEV distribution, the integral for the KLD is divided into three parts:

$$\begin{aligned}\mathrm{KLD}(\pi_\xi,\pi_0) &= \int_{-\infty}^{p_a}\pi_\xi(x)\log\left(\frac{\pi_\xi(x)}{\pi_0(x)}\right)\mathrm{d}x\\ &\quad+\int_{p_a}^{p_b}\pi_\xi(x)\log\left(\frac{\pi_\xi(x)}{\pi_0(x)}\right)\mathrm{d}x+\int_{p_b}^{\infty}\pi_\xi(x)\log\left(\frac{\pi_\xi(x)}{\pi_0(x)}\right)\mathrm{d}x\\ &= \mathrm{KLD}_{\mathrm{Gumbel}}(\pi_\xi,\pi_0)+\mathrm{KLD}_{\mathrm{Blending}}(\pi_\xi,\pi_0)+\mathrm{KLD}_{\mathrm{Fréchet}}(\pi_\xi,\pi_0).\end{aligned}$$

A closed-form expression can be found for the Gumbel part of the KLD, where we express the Gumbel distribution function as $G(x)=\exp(-h_2(x))$ and use the same substitution techniques as for the KLD computations with the GEV distribution. We get

$$\begin{aligned}\mathrm{KLD}_{\mathrm{Gumbel}}(\pi_\xi,\pi_0) &= -\Gamma_u(-\log p_a;2)+p_a\log(-\log p_a)-E_i(\log p_a)\\ &\quad+\Gamma_u(-\log p_a;C_1+1)e^{-C_2}-C_1\left(p_a\log(-\log p_a)-E_i(\log p_a)\right)\\ &\quad+p_a\left(C_2+\log\xi+\log\left(\log\left(\frac{\log p_a}{\log p_b}\right)\right)\right)\\ &\quad-p_a\log\left((-\log p_b)^{-\xi}-(-\log p_a)^{-\xi}\right),\end{aligned}$$

where

$$C_1=\frac{1}{\xi}\frac{(-\log p_b)^{-\xi}-(-\log p_a)^{-\xi}}{\log\left(\frac{\log p_a}{\log p_b}\right)},$$

and

$$C_2=C_1\log(-\log p_a)+\frac{1}{\xi}\left((-\log p_a)^{-\xi}-1\right).$$

Here, $\Gamma_u(x;\alpha)=\int_x^\infty t^{\alpha-1}e^{-t}\mathrm{d}t$ is the upper incomplete gamma function, and $E_i(x)=\int_{-\infty}^{x}(e^t/t)\mathrm{d}t$ is the exponential integral, which can be evaluated using the GNU Scientific Library (Hankin, 2006; Galassi et al., 2007). The Fréchet part of the KLD is found in the same way as (S1), only using different integration limits. For $0\le\xi<1$ we get

$$\begin{aligned}\mathrm{KLD}_{\mathrm{Fréchet}}(\pi_\xi,\pi_0) &= -\Gamma_l(-\log p_b;2)+(\xi+1)\left(-p_b\log(-\log p_b)\right.\\ &\quad\left.+E_i(\log p_b)-\gamma\right)+\frac{1}{\xi}\left(\Gamma_l(-\log p_b;1-\xi)-(1-p_b)\right)\\ &\quad+\int_{p_b}^{1}\exp\left(\frac{1}{\xi}\left(1-(-\log u)^{-\xi}\right)\right)\mathrm{d}u,\end{aligned}$$

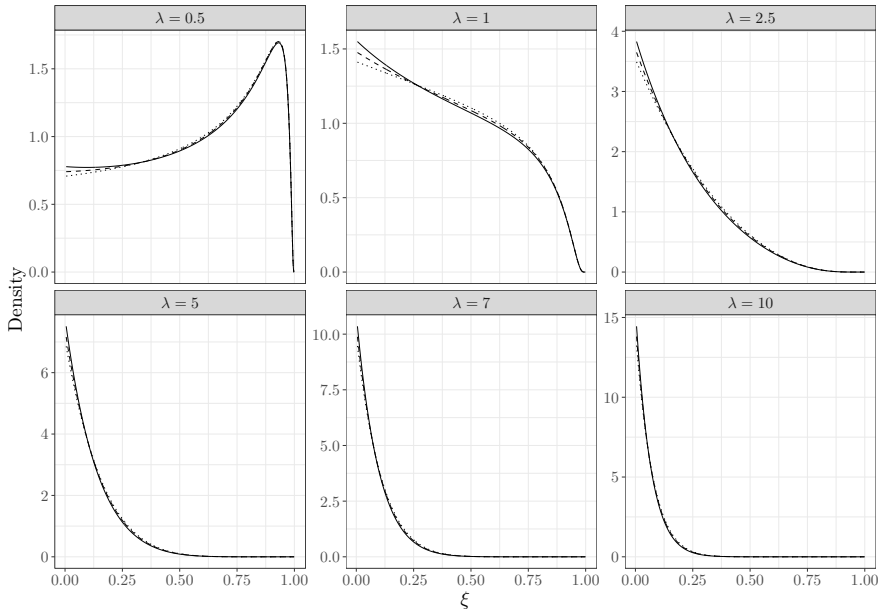


Figure S3: *PC priors with base models $\xi = 0$ and different penalty parameters λ for the GEV (solid line), bGEV (dashed line) and generalised Pareto distribution (dotted line).*

where $\Gamma_l(x;\alpha) = \Gamma(\alpha) - \Gamma_u(x;\alpha)$ is the lower incomplete gamma function. The integral in the expression above must once again be evaluated numerically. We are unable to find an expression for the KLD integral between the p_a quantile and the p_b quantile. Thus, we use numerical integration with $\mu = 0$ and $\sigma = 1$ to compute $\text{KLD}_{\text{Blending}}(\pi_\xi, \pi_0)$. The value of the integral does not depend on the values of μ and σ, but we are unable to compute it without choosing some value for them. Being unable to find an expression for the KLD of the bGEV distribution, we must also use numerical derivation to estimate the derivative of the KLD, needed in (S2).

The PC priors for the GEV, bGEV and generalised Pareto distributions are displayed in Figure S3 for different values of the penalty parameter λ. For all the chosen values of λ the three distributions are so similar that their effect on a posterior distribution probably will be close to identical. The PC prior for the generalised Pareto distribution exists in closed-form and is already implemented in `R-INLA`, while the other PC priors must be computed numerically and are not implemented in `R-INLA`. Consequently, we choose to use the PC prior of the generalised Pareto distribution for modelling the tail parameter of the bGEV distribution.

References

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., & Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *WIREs Computational Statistics*, *10*(6), e1443. https://doi.org/10.1002/wics.1443

Barbero, R., Fowler, H. J., Blenkinsop, S., Westra, S., Moron, V., Lewis, E., Chan, S., Lenderink, G., Kendon, E., Guerreiro, S., Li, X.-F., Villalobos, R., Ali, H., & Mishra, V. (2019). A synthesis of hourly and daily precipitation extremes in different climatic regions. *Weather and Climate Extremes*, *26*, 100219. https://doi.org/10.1016/j.wace.2019.100219

Bivand, R., Gómez-Rubio, V., & Rue, H. (2015). Spatial data analysis with R-INLA with some extensions. *Journal of Statistical Software*, *63*(20), 1–31. https://doi.org/10.18637/jss.v063.i20

Bolin, D., & Wallin, J. (2023). Local scale invariance and robustness of proper scoring rules. *Statistical Science*, *38*(1), 140–159. https://doi.org/10.1214/22-STS864

Bücher, A., & Segers, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, *20*(4), 839–872. https://doi.org/10.1007/s10687-017-0292-6

Bücher, A., & Segers, J. (2018). Inference for heavy tailed stationary time series based on sliding blocks. *Electronic Journal of Statistics*, *12*(1), 1098–1125. https://doi.org/10.1214/18-EJS1415

Bücher, A., & Zhou, C. (2021). A Horse Race Between the Block Maxima Method and the Peak-Over-Threshold Approach. *Statistical Science*, *36*(3), 360–378. https://doi.org/10.1214/20-STS795

Carreau, J., Naveau, P., & Neppel, L. (2016). Characterization of homogeneous regions for regional peaks-over-threshold modeling of heavy precipitation [working paper]. https://hal.ird.fr/ird-01331374

Castro-Camilo, D., & Huser, R. (2020). Local likelihood estimation of complex tail dependence structures, applied to U.S. precipitation extremes. *Journal of the American Statistical Association*, *115*(531), 1037–1054. https://doi.org/10.1080/01621459.2019.1647842

Castro-Camilo, D., Huser, R., & Rue, H. (2019). A spliced gamma-generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(3), 517–534. https://doi.org/10.1007/s13253-019-00369-z

Castro-Camilo, D., Huser, R., & Rue, H. (2022). Practical strategies for generalized extreme value-based regression models for extremes. *Environmetrics*, *33*(6), e2742. https://doi.org/10.1002/env.2742

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer, London. https://doi.org/10.1007/978-1-4471-3675-0

Cooley, D., Nychka, D., & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, *102*(479), 824–840. https://doi.org/10.1198/016214506000000780

Crespi, A., Lussana, C., Brunetti, M., Dobler, A., Maugeri, M., & Tveito, O. E. (2019). High-resolution monthly precipitation climatologies over Norway (1981-2010): Joining numerical model data sets and in situ observations.

International Journal of Climatology, *39*(4), 2057–2070. https://doi.org/10.1002/joc.5933

Dalrymple, T. (1960). *Flood-frequency analyses*. U.S. Government Printing Office.

Davison, A. C., & Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application*, *2*(1), 203–235. https://doi.org/10.1146/annurev-statistics-010814-020133

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, *27*(2), 161–186. https://doi.org/10.1214/11-STS376

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(3), 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L., & Stordal, F. (2015). Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics*, *26*(2), 89–106. https://doi.org/10.1002/env.2301

Dyrrdal, A. V., Skaugen, T., Stordal, F., & Førland, E. J. (2016). Estimating extreme areal precipitation in Norway from a gridded dataset. *Hydrological Sciences Journal*, *61*(3), 483–494. https://doi.org/10.1080/02626667.2014.947289

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190. https://doi.org/10.1017/S0305004100015681

Friederichs, P., & Thorarinsdottir, T. L. (2012). Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. *Environmetrics*, *23*(7), 579–594. https://doi.org/10.1002/env.2176

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, *114*(525), 445–452. https://doi.org/10.1080/01621459.2017.1415907

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M., Rossi, F., & Ulerich, R. (2007). The GNU scientific library reference manual, 2007.

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

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, *29*(3), 411–422. https://doi.org/10.1198/jbes.2010.08110

Guttorp, P., & Gneiting, T. (2006). Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, *93*(4), 989–995. https://doi.org/10.1093/biomet/93.4.989

Hankin, R. K. S. (2006). Special functions in R: Introducing the Gsl package. *R News*, *6*.

Hanssen-Bauer, I., & Førland, E. J. (1998). Annual and seasonal precipitation variations in Norway 1896-1997. *DNMI KLIMA Report 27/98*.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis*.

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Research*, *25*(8), 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8

Koenker, R. (2005). *Quantile regression* (Vol. 38). https://doi.org/10.1017/CBO9780511754098

Koenker, R. (2018). *Quantreg: Quantile regression* [R package version 5.35].

Koutsoyiannis, D., Kozonis, D., & Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, *206*(1), 118–135. https://doi.org/10.1016/S0022-1694(98)00097-3

Lehmann, E. A., Phatak, A., Stephenson, A., & Lau, R. (2016). Spatial modelling framework for the characterisation of rainfall extremes at different durations and under climate change. *Environmetrics*, *27*(4), 239–251. https://doi.org/10.1002/env.2389

Lerch, S., Thorarinsdottir, T. L., Ravazzolo, F., & Gneiting, T. (2017). Forecaster's dilemma: Extreme events and forecast evaluation. *Statistical Science*, *32*(1), 106–127. https://doi.org/10.1214/16-STS588

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19). https://doi.org/10.18637/jss.v063.i19

Lindgren, F., Rue, H., & 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. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., & Andersen, J. (2018). SeNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. *Earth System Science Data*, *10*(1), 235. https://doi.org/10.5194/essd-10-235-2018

Lussana, C., Tveito, O. E., & Uboldi, F. (2018). Three-dimensional spatial interpolation of 2 m temperature over Norway. *Quarterly Journal of the Royal Meteorological Society*, *144*(711), 344–364. https://doi.org/10.1002/qj.3208

Matern, B. (1986). *Spatial variation* (2nd ed., Vol. 36). Springer New York. https://doi.org/10.1007/978-1-4615-7892-5

Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*(10), 1087–1096. https://doi.org/10.1287/mnsc.22.10.1087

Mohr, M. (2009). Comparison of versions 1.1 and 1.0 of gridded temperature and precipitation data for Norway. *met.no note*, *19*.

Naveau, P., Toreti, A., Smith, I., & Xoplaki, E. (2014). A fast nonparametric spatio-temporal regression scheme for generalized Pareto distributed heavy precipitation. *Water Resources Research*, *50*(5), 4011–4017. https://doi.org/10.1002/2014WR015431

Opitz, T., Huser, R., Bakka, H., & Rue, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, *21*(3), 441–462. https://doi.org/10.1007/s10687-018-0324-x

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, *49*(1), 187–201. https://doi.org/10.1029/2012WR012557

Robinson, M. E., & Tawn, J. A. (2000). Extremal analysis of processes sampled at different frequencies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(1), 117–135. https://doi.org/10.1111/1467-9868.00223

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, *4*(1), 395–421. https://doi.org/10.1146/annurev-statistics-060116-054045

Sang, H., & Gelfand, A. (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological and Environmental Statistics*, *15*, 49–65. https://doi.org/10.1007/s13253-009-0010-1

Sang, H., & Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, *16*(3), 407–426. https://doi.org/10.1007/s10651-007-0078-0

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*(1), 1–28. https://doi.org/10.1214/16-STS576

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, *72*(1), 67–90. https://doi.org/10.1093/biomet/72.1.67

Stein, M. L. (1999). *Interpolation of spatial data : Some theory for Kriging*. Springer.

Sun, Y., Bowman, K. P., Genton, M. G., & Tokay, A. (2015). A Matérn model of the spatial covariance structure of point rain rates. *Stochastic Environmental Research and Risk Assessment*, *29*(2), 411–416. https://doi.org/10.1007/s00477-014-0923-2

Tveito, O. E., Bjørdal, I., Skjelvåg, A. O., & Aune, B. (2005). A GIS-based agro-ecological decision system based on gridded climatology. *Meteorological Applications*, *12*(1), 57–68. https://doi.org/10.1017/S1350482705001490

Ulrich, J., Jurado, O. E., Peter, M., Scheibel, M., & Rust, H. W. (2020). Estimating IDF curves consistently over durations with spatial covariates. *Water*, *12*(11). https://doi.org/10.3390/w12113119

Van de Vyver, H. (2012). Spatial regression models for extreme precipitation in Belgium. *Water Resources Research*, *48*(9). https://doi.org/10.1029/2011WR011707

Vandeskog, S. M., Martino, S., & Castro-Camilo, D. (2021). Modelling block maxima with the blended generalised extreme value distribution. *22nd European Young Statisticians Meeting - Proceedings*.

Wang, Y., & So, M. K. (2016). A Bayesian hierarchical model for spatial extremes with multiple durations. *Computational Statistics & Data Analysis*, *95*, 39–56. https://doi.org/10.1016/j.csda.2015.09.001

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

Wilson, P. S., & Toumi, R. (2005). A fundamental probability distribution for heavy rainfall. *Geophysical Research Letters*, *32*(14). https://doi.org/10.1029/2005GL022465

World Economic Forum. (2021). *The global risks report 2021*. http://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf

Zou, N., Volgushev, S., & Bücher, A. (2021). Multiple block sizes and overlapping blocks for multivariate time series extremes. *The Annals of Statistics*, *49*(1), 295–320. https://doi.org/10.1214/20-AOS1957

— 3 —

An efficient workflow for modelling high-dimensional spatial extremes

Silius M. Vandeskog, Sara Martino and Raphaël Huser

In review, but published on arXiv with reference 2210.00760

An efficient workflow for modelling high-dimensional spatial extremes

Silius M. Vandeskog[1] Sara Martino[1] Raphaël Huser[2]

[1] Norwegian University of Science and Technology (NTNU)
[2] King Abdullah University of Science and Technology (KAUST)

Abstract

A successful model for high-dimensional spatial extremes should, in principle, be able to describe both weakening extremal dependence at increasing levels and changes in the type of extremal dependence class as a function of the distance between locations. Furthermore, the model should allow for computationally tractable inference using inference methods that efficiently extract information from data and that are robust to model misspecification. In this paper, we demonstrate how to fulfil all these requirements by developing a comprehensive methodological workflow for efficient Bayesian modelling of high-dimensional spatial extremes using the spatial conditional extremes model while performing fast inference with `R-INLA`. We then propose a post hoc adjustment method that results in more robust inference by properly accounting for possible model misspecification. The developed methodology is applied for modelling extreme hourly precipitation from high-resolution radar data in Norway. Inference is computationally efficient, and the resulting model fit successfully captures the main trends in the extremal dependence structure of the data. Robustifying the model fit by adjusting for possible misspecification further improves model performance.

Keywords: Spatial conditional extremes, Robust Bayesian inference, Computational statistics, `R-INLA`

1 Introduction

The effects of climate change and the increasing availability of large and high-quality data sets has lead to a surge of research on the modelling of spatial extremes (e.g., Opitz et al., 2018; Shooter et al., 2019; Castro-Camilo et al., 2019; Koch et al., 2021; Simpson & Wadsworth, 2021; Vandeskog et al., 2022; Richards et al., 2022; Koh et al., 2023). Modelling spatial extremes is challenging for two main reasons: 1) classical models are often

not flexible enough to provide realistic descriptions of extremal dependence, and 2) inference can be computationally demanding or intractable, so modellers must often rely on less efficient inference methods; see Huser and Wadsworth (2022) for a review of these challenges. In this paper, we propose a comprehensive methodological workflow, as well as practical strategies, on how to perform efficient and flexible high-dimensional modelling of spatial extremes.

An important component of spatial extreme value theory is the characterisation of a spatial process' asymptotic dependence properties (e.g., Coles et al., 1999). Two random variables with a positive limiting probability to experience their extremes simultaneously are denoted asymptotically dependent. Otherwise, they are denoted asymptotically independent. As demonstrated by Sibuya et al. (1960), two asymptotically independent random variables may still be highly correlated and thus exhibit large amounts of so-called sub-asymptotic dependence. Thus, correct estimation of both asymptotic and sub-asymptotic dependence properties is of utmost importance when assessing the risks of spatial extremes.

Most classical models for spatial extremes are based on max-stable processes (Davison et al., 2012; Davison et al., 2019). These allow for rich modelling of asymptotic dependence, but are often too rigid in their descriptions of asymptotic independence and sub-asymptotic dependence. Other approaches have been proposed, such as scale-mixture models (Engelke et al., 2019; Huser & Wadsworth, 2019), which allow for rich modelling of both asymptotic dependence and independence, and a more flexible description of sub-asymptotic dependence. However, these models require that all location pairs share the same asymptotic dependence class, which is problematic as one would expect neighbouring locations to be asymptotically dependent and far-away locations to be asymptotically independent. Max-mixture model (Wadsworth & Tawn, 2012) allow for even more flexible modelling of sub-asymptotic dependence, and for changing the asymptotic dependence class as a function of distance. However, it is often difficult to estimate the key model parameter, which describes the transition between extremal dependence classes. Additionally, these models must often rely on less efficient inference methods. Further improvements are given by the kernel convolution model of Krupskii and Huser (2022), more recent scale-mixture models such as that of Hazra et al. (2021), and the spatial conditional extremes model of Wadsworth and Tawn (2022), which all allows for flexible modelling of different extremal dependence classes as a function of distance. The spatial conditional extremes model allows for a particularly simple way of modelling spatial extremes. It is based on the conditional extremes model of Heffernan and Tawn (2004) and Heffernan and Resnick (2007), which describes the behaviour of a random vector conditional on one of its components being extreme, and it can be interpreted as a semi-

parametric regression model, which makes it intuitive and simple to tailor or extend. Due to its high flexibility and conceptual simplicity, this is our chosen model for high-dimensional spatial extremes.

To make the spatial conditional extremes model computationally efficient in higher dimensions, Wadsworth and Tawn (2022) propose to model spatial dependence using a residual random process constructed from a Gaussian copula and delta-Laplace marginal distributions. However, inference for Gaussian processes typically requires computing the inverse of the covariance matrix, whose cost scales cubicly with the model dimension. Thus, Simpson et al. (2023), propose to exchange the delta-Laplace process with a Gaussian Markov random field (Rue & Held, 2005) created using the so-called stochastic partial differential equations (SPDE) approach of Lindgren et al. (2011). Furthermore, in order to perform spatial high-dimensional Bayesian inference, Simpson et al. (2023) modify the spatial conditional extremes model into a latent Gaussian model, which allows for performing inference using the integrated nested Laplace approximation (INLA; Rue et al., 2009), implemented in the `R-INLA` software (Rue et al., 2017). This allows for a considerable improvement in the Bayesian modelling of high-dimensional spatial extremes. However, there is still much room for improvement. In this paper, we thus build upon the modelling framework of Simpson et al. (2023) and develop a more general methodology for modelling spatial conditional extremes with `R-INLA`. We also point out a theoretical weakness in the constraining methods proposed by Wadsworth and Tawn (2022) and used by Simpson et al. (2023), and we demonstrate a computationally efficient way of fixing it.

As most statistical models for extremes are based on asymptotic arguments and assumptions, a certain degree of misspecification will always be present when modelling finite amounts of data. Additionally, model choices made for reasons of computational efficiency, such as adding Markov assumptions to a spatial random field, may lead to further misspecification. This complicates Bayesian inference and can result in misleading posterior distributions (Ribatet et al., 2012; Kleijn & van der Vaart, 2012). One should therefore strive to make inference more robust towards misspecification when modelling high-dimensional spatial extremes. Shaby (2014) proposes a method for more robust inference through a post hoc transformation of posterior samples created using Markov chain Monte Carlo (MCMC) methods. Here, we develop a refined version of his adjustment method, and we use it for performing more robust inference with `R-INLA`.

As extreme behaviour is, by definition, rare, inference with the conditional extremes model often relies on a composite likelihood that combines data from different conditioning sites under the working assumption of independence (Heffernan & Tawn, 2004; Simpson & Wadsworth, 2021; Wadsworth & Tawn, 2022; Richards et al., 2022). However, composite likeli-

hoods can lead to large amounts of misspecification (Ribatet et al., 2012), and Simpson et al. (2023) thus abstain from using a composite likelihood to avoid the problems that occur when performing Bayesian inference with a composite likelihood using `R-INLA`. We show that the post hoc adjustment method accounts for the misspecification from the composite likelihood, thus allowing for more efficient inference using considerably more data.

To sum up, in this paper we develop a general workflow for performing high-dimensional modelling of spatial extremes using the spatial conditional extremes model. We improve upon the work of Simpson et al. (2023) by developing a more general, flexible and computationally efficient methodology for modelling spatial conditional extremes with `R-INLA` and the SPDE approach. Then, we make inference more robust towards misspecification by extending the post hoc adjustment method of Shaby (2014), and we further apply this adjustment method for more efficient inference by combining information from multiple conditioning sites.

The remainder of the paper is organised as follows: In Section 2, the spatial conditional extremes model is presented as a flexible choice for modelling spatial extremes. Modifications and assumptions that allow for computationally efficient inference with improved data utilisation are also presented. Then, in Section 3, we develop a general methodology for implementing a large variety of spatial conditional extremes models in `R-INLA`. Section 4 examines the problems that can occur when performing Bayesian inference based on a misspecified likelihood, and demonstrates how to perform more robust inference with `R-INLA` by accounting for possible misspecification. In Section 5, a simulation study is presented where we demonstrate and validate our entire workflow for high-dimensional modelling of spatial extremes. Then, in Section 6, our proposed workflow is applied for modelling extreme hourly precipitation from high-resolution radar data in Norway. Finally, we conclude in Section 7 with some discussion and perspectives on future research.

2 Flexible modelling with spatial conditional extremes

2.1 The spatial conditional extremes model

Let $Y(\boldsymbol{s})$ be a random process defined over space ($\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2$) with Laplace margins. For this random process, Wadsworth and Tawn (2022) assume the existence of standardising functions $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ such that, for a large enough threshold t,

$$[Y(\boldsymbol{s}) \mid Y(\boldsymbol{s}_0) = y_0 > t] \overset{d}{=} a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) + b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) Z(\boldsymbol{s}; \boldsymbol{s}_0), \quad \boldsymbol{s}, \boldsymbol{s}_0 \in \mathcal{S}, \tag{1}$$

where $Z(\boldsymbol{s}; \boldsymbol{s}_0)$ is a random process satisfying $Z(\boldsymbol{s}_0; \boldsymbol{s}_0) = 0$ almost surely, and $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) \leq y_0$, with equality when $\boldsymbol{s} = \boldsymbol{s}_0$. The degree of asymptotic dependence may be measured through the extremal correlation coefficient

$$\chi(\boldsymbol{s}_1, \boldsymbol{s}_2) = \lim_{p \to 1} \chi_p(\boldsymbol{s}_1, \boldsymbol{s}_2) = \lim_{p \to 1} \mathrm{P}(Y(\boldsymbol{s}_1) > F_Y^{-1}(p) \mid Y(\boldsymbol{s}_2) > F_Y^{-1}(p)),$$

where $F_Y^{-1}(p)$ is the marginal quantile function of the process $Y(\boldsymbol{s})$. If $\chi(\boldsymbol{s}_1, \boldsymbol{s}_2) > 0$, then $Y(\boldsymbol{s}_1)$ and $Y(\boldsymbol{s}_2)$ are asymptotically dependent, whereas if $\chi(\boldsymbol{s}_1, \boldsymbol{s}_2) = 0$, they are asymptotically independent. It is known that $Y(\boldsymbol{s})$ and $Y(\boldsymbol{s}_0)$, defined in (1), are asymptotically dependent when $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = y_0$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = 1$, while they are asymptotically independent when $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) < y_0$ (Heffernan & Tawn, 2004). However, under asymptotic independence, the convergence of $\chi_p(\cdot)$ to $\chi(\cdot)$ is slower for larger values of $a(\cdot)$ and $b(\cdot)$.

Wadsworth and Tawn (2022) provide some guidance on parametric functions for $a(\cdot)$ and $b(\cdot)$ together with parametric distributions for $Z(\cdot)$ that cover a large range of already existing models. For modelling $a(\cdot)$ they specifically propose the parametric function

$$a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = y_0 \alpha(\|\boldsymbol{s} - \boldsymbol{s}_0\|) = y_0 \exp\left\{-[\max(0, \|\boldsymbol{s} - \boldsymbol{s}_0\| - \Delta)/\lambda_a]^{\kappa_a}\right\}, \quad (2)$$

with parameters $\Delta \geq 0$ and $\lambda_a, \kappa_a > 0$. This function yields a model with asymptotic dependence for locations closer to the conditioning site than a distance Δ, while displaying asymptotic independence for distances larger than Δ, with a weakening sub-asymptotic dependence as we move further away from $\boldsymbol{s}_0$. To the best of our knowledge, this model (and its sub-models) has been adopted by a majority of spatial conditional extremes modellers. Several forms are proposed for $b(\cdot)$, including the form $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = y_0^{\beta}$, when $\Delta = 0$. This allows for modelling asymptotic independence with positive dependence, with the β parameter helping to control the speed of convergence of $\chi_p(\boldsymbol{s}_1, \boldsymbol{s}_2)$ to $\chi(\boldsymbol{s}_1, \boldsymbol{s}_2)$. A weakness of this form is that it enforces the same positive dependence for all distances, including large distances where the observations should be independent of $Y(\boldsymbol{s}_0)$. To remedy this issue, Wadsworth and Tawn (2022) also propose the model $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = 1 + a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)^{\beta}$, which converges to one as the distance increases. Alternatively, Shooter, Tawn, et al. (2021) and Richards et al. (2022) have proposed different models on the form $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = y_0^{\beta(\|\boldsymbol{s} - \boldsymbol{s}_0\|)}$, where they let the function $\beta(d)$ converge to zero as the distance $d \to \infty$.

Clearly, the best model for the standardising functions $a(\cdot)$ and $b(\cdot)$ depends on the application. Therefore, in Section 3, we develop a general methodology for implementing the conditional spatial extremes model in `R-INLA` for any kind of functions $a(\cdot)$ and $b(\cdot)$. In addition, we provide practical guidance and diagnostics for selecting appropriate standardising

functions in our simulation study in Section 5 and data application in Section 6.

2.2 Modifications for high-dimensional modelling

To perform high-dimensional inference, Wadsworth and Tawn (2022) propose to model $Z(\cdot)$ as a random process with a Gaussian copula and delta-Laplace marginal distributions. Their proposed model for $Z(\cdot)$ has later seen usage by, e.g., Shooter, Ross, Ribal, et al. (2021), Shooter, Tawn, et al. (2021), and Shooter et al. (2022) and Richards et al. (2022). However, in order to perform Bayesian inference with `R-INLA`, Simpson et al. (2023) modify (1) into a latent Gaussian model by adding a Gaussian nugget effect and requiring $Z(\cdot)$ to be a fully Gaussian random field. This gives the model

$$[Y(\boldsymbol{s}) \mid Y(\boldsymbol{s}_0) = y_0 > t] \overset{d}{=} a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) + b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) Z(\boldsymbol{s}; \boldsymbol{s}_0) + \epsilon(\boldsymbol{s}; \boldsymbol{s}_0), \quad (3)$$

where $\epsilon(\boldsymbol{s}; \boldsymbol{s}_0)$ is Gaussian white noise with constant variance, satisfying $\epsilon(\boldsymbol{s}_0; \boldsymbol{s}_0) = 0$ almost surely. They further assume that $Z(\cdot)$ has zero mean and a Matérn covariance structure, so that it can be approximated using the SPDE approach of Lindgren et al. (2011), which speeds up inference by approximating the precision matrix of $Z(\cdot)$ with a sparse and low-rank matrix. However, making the precision matrix too sparse and/or low-rank leads to some model misspecification, which is further amplified by the fact that $Y(\boldsymbol{s})$ has Laplace marginal distributions, but is modelled using a fully Gaussian random field. We here nevertheless adopt the modelling assumptions of Simpson et al. (2023), as we find them necessary for performing truly high-dimensional Bayesian inference with `R-INLA`. However, unlike Simpson et al. (2023), we then account for the possible misspecification of these assumptions using the robustifying approach described in Section 4.

2.3 Efficient data utilisation with a composite likelihood

The spatial conditional extremes model consists in modelling a spatial process conditional on extreme behaviour at a predefined conditioning site. However, inference is often made challenging because the conditioning site contains few observed extremes. To strengthen inference it is therefore common to assume stationarity, in the sense that all parameters of $a(\cdot)$, $b(\cdot)$, $Z(\cdot)$ and $\epsilon(\cdot)$ are independent of the conditioning site. Under such stationarity, it is possible to combine information from multiple conditioning sites into one global model fit, using the composite likelihood of Heffernan and Tawn (2004) and Wadsworth and Tawn (2022). Given observations $\mathcal{Y} = \{y_i(\boldsymbol{s}_j) : i = 1, 2, \ldots, n, j = 1, 2, \ldots, m\}$ from n time points and m

locations, the composite log-likelihood may be expressed as

$$\ell_c(\boldsymbol{\theta}; \mathcal{Y}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \ell(\boldsymbol{\theta}; \boldsymbol{y}_{i,-j} \mid y_i(\boldsymbol{s}_j)) I(y_i(\boldsymbol{s}_j) > t), \tag{4}$$

where $\ell(\cdot)$ is the log-likelihood of the conditional extremes model, $\boldsymbol{y}_{i,-j}$ is an $(m-1)$-dimensional vector containing observations from time point i, for all locations except $\boldsymbol{s}_j$, $I(\cdot)$ is the indicator function and $\boldsymbol{\theta}$ contains all parameters of the spatial conditional extremes model. If m is too large, one may choose to build the composite likelihood using only a subset of the available conditioning sites, and $\boldsymbol{y}_{i,-j}$ may be modified to contain only a subset of the available observations from time point i, which may vary with both i and j.

The composite likelihood is not a valid likelihood, since multiple of the terms in (4) may contain the same observations. Incorrectly interpreting the composite likelihood as a true likelihood is therefore tantamount to specifying a model in which $[\boldsymbol{y}_{i,-j} \mid y_i(\boldsymbol{s}_j)]$ is (wrongly) assumed to be independent from $[\boldsymbol{y}_{i,-k} \mid y_i(\boldsymbol{s}_k)]$ for all time points i and locations pairs $(\boldsymbol{s}_j, \boldsymbol{s}_k)$ with $y_i(\boldsymbol{s}_j) > t$ and $y_i(\boldsymbol{s}_k) > t$. Performing inference with the composite likelihood can therefore lead to considerable misspecification, which should be accounted for before drawing conclusions from the model fit. In Section 4, we therefore show how to robustify inference by accounting for the possible model misspecification. To the best of our knowledge, our paper is the first attempt to perform Bayesian inference for the spatial conditional extremes model based on a composite likelihood.

3 Fast inference using R-INLA

3.1 Latent Gaussian model framework

R-INLA performs inference on latent Gaussian models of the form

$$\begin{aligned}
[y_i \mid \boldsymbol{u}, \boldsymbol{\theta}_1] &\overset{i.i.d.}{\sim} \pi(y_i \mid \eta_i(\boldsymbol{u}), \boldsymbol{\theta}_1), \ i = 1, 2, \ldots, n, \\
[\boldsymbol{u} \mid \boldsymbol{\theta}_2] &\sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}_2), \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2)), \\
\left(\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top\right)^\top &\sim \pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2),
\end{aligned}$$

where $\boldsymbol{u}$ is a latent Gaussian field with mean $\boldsymbol{\mu}(\boldsymbol{\theta}_2)$ and precision matrix $\boldsymbol{Q}(\boldsymbol{\theta}_2)$, and the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ are assigned priors $\pi(\boldsymbol{\theta}_1)$ and $\pi(\boldsymbol{\theta}_2)$. Observations $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ are linked to the latent field through the linear predictor $\boldsymbol{\eta} = (\eta_1(\boldsymbol{u}), \ldots, \eta_n(\boldsymbol{u}))^\top = \boldsymbol{A}\boldsymbol{u}$, where $\boldsymbol{A}$ is a known design matrix. This linear predictor defines the location parameter of the likelihood $\pi(\boldsymbol{y} \mid \boldsymbol{\eta}, \boldsymbol{\theta}_1)$, via a possibly non-linear link function. All

observations are assumed to be conditionally independent given $\boldsymbol{\eta}$ and $\boldsymbol{\theta}_1$, so that $\pi(\boldsymbol{y} \mid \boldsymbol{\eta}, \boldsymbol{\theta}_1) = \prod_{i=1}^{n} \pi(y_i \mid \eta_i(\boldsymbol{u}), \boldsymbol{\theta}_1)$. For computational reasons, when using R-INLA, the likelihood must be chosen from a predefined set of likelihood functions. The linear predictor can be decomposed into $N \geq 1$ components, $\boldsymbol{\eta} = \boldsymbol{A}^{(1)}\boldsymbol{u}^{(1)} + \cdots + \boldsymbol{A}^{(N)}\boldsymbol{u}^{(N)}$, where each component represents, e.g., an intercept term, a linear combination of regression coefficients, an SPDE component, etc. All of these components must either be predefined in R-INLA or defined by the user, using the `rgeneric` framework or the recently added `cgeneric` framework.

The spatial conditional extremes model in (3) corresponds to a latent Gaussian model where the likelihood is Gaussian with variance θ_1, say, and the linear predictor is equal to $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) + b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)Z(\boldsymbol{s}; \boldsymbol{s}_0)$, with $\boldsymbol{\theta}_2$ containing the parameters of $a(\cdot)$, $b(\cdot)$ and $Z(\cdot)$. For most forms of $a(\cdot)$ and $b(\cdot)$, R-INLA does not contain suitable predefined model components for describing the linear predictor, so we must define these manually. In order to define a new R-INLA component $\boldsymbol{u}^{(N+1)}$, with parameters $\boldsymbol{\theta}^{(N+1)}$, using one of the `rgeneric`/`cgeneric` frameworks, one must provide functions written in `R` of `C`, respectively, that compute the precision matrix mean and prior density of $\boldsymbol{u}^{(N+1)}$ for any value of $\boldsymbol{\theta}^{(N+1)}$. The `cgeneric` framework yields considerably faster inference than the `rgeneric` framework, but it requires knowledge of the lower-level `C` programming language. In this paper, we propose a method for defining model components for $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)Z(\boldsymbol{s}; \boldsymbol{s}_0)$ using the `rgeneric`/`cgeneric` frameworks, for any kind of functions $a(\cdot)$ and $b(\cdot)$. In the online supplementary material, we provide the necessary code for defining the models used in Section 5 and 6 with the `cgeneric` framework.

3.2 Defining $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)Z(\boldsymbol{s}; \boldsymbol{s}_0)$ in R-INLA

The SPDE approach creates a Gaussian Markov random field $\widehat{Z}(\boldsymbol{s})$ that is an approximation to a Gaussian random field $Z(\boldsymbol{s})$ with Matérn covariance function

$$\operatorname{Cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}')) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|\boldsymbol{s} - \boldsymbol{s}'\|)^{\nu} K_{\nu}(\kappa\|\boldsymbol{s} - \boldsymbol{s}'\|), \tag{5}$$

where σ^2 is the marginal variance, $\nu > 0$ is a smoothness parameter, $\rho = \sqrt{8\nu}/\kappa$ is a range parameter and K_{ν} is the modified Bessel function of the second kind and order ν. The smoothness parameter ν is difficult to estimate from data and is therefore often given a fixed value (Lindgren & Rue, 2015). The SPDE approximation $\widehat{Z}(\boldsymbol{s})$ is constructed as a linear combination of Gaussian Markov random variables on a triangulated mesh, i.e., $\widehat{Z}(\boldsymbol{s}) = \sum_{i=1}^{M} \phi_i(\boldsymbol{s})W_i$, where $W_1, \ldots, W_M$ are random variables from a Gaussian Markov random field, and $\phi_i, \ldots, \phi_M$ are piecewise linear basis

functions. In order to approximate the non-stationary Gaussian random field $b(\boldsymbol{s};\boldsymbol{s}_0,y_0)Z(\boldsymbol{s})$ with the SPDE approach, for any function $b(\boldsymbol{s};\boldsymbol{s}_0,y_0)$, we modify the weights W_i to get

$$\widehat{Z}_b(\boldsymbol{s};\boldsymbol{s}_0,y_0)=\sum_{i=1}^{M}\phi_i(\boldsymbol{s})b(\boldsymbol{s}_i;\boldsymbol{s}_0,y_0)W_i \tag{6}$$

where $\boldsymbol{s}_1,\dots\boldsymbol{s}_M$ are the locations of the M mesh nodes. This shares some similarities with the approach of Ingebrigtsen et al. (2014) for implementing non-stationary SPDE fields. Since $\widehat{Z}_b(\cdot)$ is a linear combination of Gaussian random variables, its variance equals

$$\operatorname{Var}\left(\widehat{Z}_b(\boldsymbol{s};\boldsymbol{s}_0,y_0)\right)=\sum_{i,j=1}^{M}\phi_i(\boldsymbol{s})\phi_j(\boldsymbol{s})b(\boldsymbol{s}_i;\boldsymbol{s}_0,y_0)b(\boldsymbol{s}_j;\boldsymbol{s}_0,y_0)\operatorname{Cov}\left(W_i,W_j\right),$$

which is unequal to $b(\boldsymbol{s};\boldsymbol{s}_0,y_0)^2\sigma^2$, the variance of $b(\boldsymbol{s};\boldsymbol{s}_0,y_0)Z(\boldsymbol{s})$. However, if $\boldsymbol{s}$ coincides with a mesh node, then one of the basis functions equals 1, while the others equal 0, giving $\operatorname{Var}\left(\widehat{Z}_b(\boldsymbol{s};\boldsymbol{s}_0,y_0)\right)=b(\boldsymbol{s};\boldsymbol{s}_0,y_0)^2\operatorname{Var}(W_i)$, which is much closer to the correct variance. On the contrary, if $\boldsymbol{s}$ is far away from a mesh node, the variance of $\widehat{Z}_b(\boldsymbol{s};\boldsymbol{s}_0,y_0)$ may be considerably different from $b(\boldsymbol{s};\boldsymbol{s}_0,y_0)^2\sigma^2$. If possible, it is therefore recommended to use a fine mesh, so all observation locations are close enough to a mesh node.

The process $\widehat{Z}_b(\cdot)$ approximates the unconstrained Gaussian random field $b(\boldsymbol{s};\boldsymbol{s}_0,y_0)Z(\boldsymbol{s})$. However, in order to define the conditional extremes model in `R-INLA`, we need to approximate the constrained field $b(\boldsymbol{s};\boldsymbol{s}_0,y_0)Z(\boldsymbol{s};\boldsymbol{s}_0)$, where $Z(\boldsymbol{s}_0;\boldsymbol{s}_0)=0$ almost surely. Wadsworth and Tawn (2022) describe two different methods for turning an unconstrained Gaussian field $Z(\boldsymbol{s})$ into a constrained field $Z(\boldsymbol{s};\boldsymbol{s}_0)$. The first one is to constrain the field by conditioning, i.e., $Z(\boldsymbol{s};\boldsymbol{s}_0)=[Z(\boldsymbol{s})\mid Z(\boldsymbol{s})=0]$; and the second one is to constrain it by subtraction, i.e., $Z(\boldsymbol{s};\boldsymbol{s}_0)=Z(\boldsymbol{s})-Z(\boldsymbol{s}_0)$. In their case studies, Wadsworth and Tawn (2022) use the first method, while Simpson et al. (2023) use the second method. We argue that constraining by subtraction yields unrealistic dependence structures, and should be avoided if other alternatives are available. A quick computation indeed shows that if $Z(\boldsymbol{s};\boldsymbol{s}_0)$ is a stationary random process that has been constrained through subtraction, then the limiting correlation between $Z(c\boldsymbol{s};\boldsymbol{s}_0)$ and $Z(-c\boldsymbol{s};\boldsymbol{s}_0)$, as $c\to\infty$ equals $1/2$. Furthermore, the limiting correlation of $Z(\boldsymbol{s}_0+\Delta\boldsymbol{s};\boldsymbol{s}_0)$ and $Z(\boldsymbol{s}_0-\Delta\boldsymbol{s};\boldsymbol{s}_0)$ as $\|\Delta\boldsymbol{s}\|\to 0$ is often negative and equals 0 if the unconstrained random field had an exponential correlation function or -1 if the unconstrained field had a Gaussian correlation function. Thus, with the subtraction approach, points that are infinitely far away from each other are strongly correlated while points that are infinitesimally close to each other might be negatively correlated or independent.

R-INLA contains an implementation for constraining a random field by conditioning on linear combinations of itself (Rue et al., 2009). One can therefore easily constrain $\widehat{Z}_b(\cdot)$ by conditioning using the extraconstr option in R-INLA. Unfortunately, this conditioning method requires the computation of an $(n \times k)$-dimensional dense matrix, where n is the number of rows of the precision matrix and k is the number of added constraints. In practice, we therefore experience that constraining by conditioning with R-INLA requires considerably more computational resources, and that it quickly turns intractable for large data sets.

We propose a third method for constraining the residual field with R-INLA. It is known that, for a Gaussian random vector $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top)^\top$ with zero mean and precision matrix

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{pmatrix},$$

the conditional distribution of $[\boldsymbol{y}_1 \mid \boldsymbol{y}_2 = \boldsymbol{0}]$ is Gaussian with zero mean and precision matrix $\boldsymbol{Q}_{11}$ (Rue & Held, 2005). Thus, if we ensure that a mesh node coincides with $\boldsymbol{s}_0$, we can constrain $Z(\cdot)$ by removing all rows and columns of $\boldsymbol{Q}$ that correspond to the mesh node at $\boldsymbol{s}_0$. This is easily achievable using the rgeneric/cgeneric framework, and it requires no extra computational effort.

3.3 Defining $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ in R-INLA

All components of the latent Gaussian field in R-INLA must be Gaussian random variables, but $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ is a deterministic function and not a random variable. However, clearly, the deterministic vector $\boldsymbol{a}$ can be approximated well by the Gaussian random vector $\boldsymbol{a} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ has zero mean and covariance matrix $\delta^2 \boldsymbol{I}$, with $\boldsymbol{I}$ being the identity matrix and δ^2 being a small, fixed marginal variance. Thus, using the rgeneric/cgeneric framework, we can approximate any deterministic function $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ with a latent Gaussian random field with mean $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and diagonal covariance matrix $\delta^2 \boldsymbol{I}$. Here, we choose $\delta^2 = \exp(-15)$.

4 Robust inference using post hoc adjustments

4.1 Adjusting posterior samples

It is well known that all models are wrong, in the sense that the data, to a certain extent, always deviate from the model assumptions. This is particularly true when modelling extremes, where most models are based on imposing asymptotically justified assumptions onto finite amounts of data. It

is also particularly true when modelling high-dimensional data, because high-dimensional models often are based on strict assumptions of (un)conditional independence and Gaussianity, in order to make inference computationally tractable, and because the amount of misspecification naturally tends to increase with the data size and dimensionality while keeping everything else constant. Accounting for this misspecification should therefore be an important step in any successful modelling strategy for high-dimensional spatial extremes.

Given n independent realisations $\mathcal{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ of a random vector $\boldsymbol{Y}$ with true distribution G, and a chosen likelihood $L(\boldsymbol{\theta}; \mathcal{Y}) = \prod_{i=1}^{n} L(\boldsymbol{\theta}; \boldsymbol{y}_i)$, it is well known that the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ is asymptotically Gaussian, i.e.,

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^*)^{1/2} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right) \rightsquigarrow \mathcal{N}(\mathbf{0}, \boldsymbol{I}), \text{ as } n \to \infty,$$

under some weak regularity conditions (White, 1982), where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{\theta}^*$ minimises the Kullback-Leibler divergence (KLD; Kullback & Leibler, 1951) between $L(\boldsymbol{\theta}; \cdot)$ and the likelihood of the true distribution G. Furthermore, $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ is the so-called Godambe sandwich information matrix (Godambe, 1960), i.e.,

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \boldsymbol{H}(\boldsymbol{\theta}) \boldsymbol{J}(\boldsymbol{\theta})^{-1} \boldsymbol{H}(\boldsymbol{\theta}), \tag{7}$$

with $\boldsymbol{H}(\boldsymbol{\theta}) = -\mathrm{E}\left[\nabla^2_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \mathcal{Y})\right]$ and $\boldsymbol{J}(\boldsymbol{\theta}) = \mathrm{Cov}\left(\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \mathcal{Y})\right)$, where $\ell(\cdot) = \log(L(\cdot))$ is the log-likelihood, and all expectations are taken with respect to G. If $L(\boldsymbol{\theta}^*; \cdot)$ is equal to the likelihood of the true distribution G, then $\boldsymbol{J}(\boldsymbol{\theta}^*) = \boldsymbol{H}(\boldsymbol{\theta}^*)$, and $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^*)$ reduces to $\boldsymbol{H}(\boldsymbol{\theta}^*)$.

From a Bayesian perspective, given a prior $\pi(\boldsymbol{\theta})$ and appropriate regularity conditions, it is also known that the posterior density, $\pi(\boldsymbol{\theta} \mid \mathcal{Y}) \propto L(\boldsymbol{\theta}; \mathcal{Y}) \pi(\boldsymbol{\theta})$, converges asymptotically to a Gaussian density with mean $\boldsymbol{\theta}^*$ and covariance matrix $\boldsymbol{H}(\boldsymbol{\theta}^*)^{-1}$ (Berk, 1966; Kleijn & van der Vaart, 2012). As the sample size increases and the effect of the prior distribution diminishes, credible intervals and confidence intervals should be expected to coincide. However, if the likelihood is misspecified so that $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^*) \neq \boldsymbol{H}(\boldsymbol{\theta}^*)$, then the resulting asymptotic $(1 - \alpha)$-credible interval differs from all well-calibrated asymptotic $(1 - \alpha)$-confidence intervals, and we say that they attain poor frequency properties. Ribatet et al. (2012) illustrate how easily a misspecified likelihood function can lead to misleading inference through posterior intervals with poor frequency properties.

Several approaches have been proposed for robustifying inference under a misspecified likelihood (Chandler & Bate, 2007; Pauli et al., 2011; Ribatet et al., 2012; Syring & Martin, 2018), but all these methods are based on modifying the likelihood function before inference, which is impossible to do within the `R-INLA` framework. However, Shaby (2014) proposes a post

hoc adjustment method that properly accounts for misspecification in the likelihood by an affine transformation of posterior samples when performing MCMC-based inference. Since this is a post hoc adjustment method, it is possible to extend it for usage with `R-INLA`. Given a sample $\boldsymbol{\theta}$ from a posterior distribution based on a misspecified likelihood, the adjusted posterior sample is defined as

$$\boldsymbol{\theta}_{\text{adj}} = \boldsymbol{\theta}^* + \boldsymbol{C}(\boldsymbol{\theta} - \boldsymbol{\theta}^*), \tag{8}$$

where the matrix $\boldsymbol{C}$ is chosen such that the asymptotic distribution of $\boldsymbol{\theta}_{\text{adj}}$, as $n \to \infty$, is Gaussian with mean $\boldsymbol{\theta}^*$ and covariance matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^*)^{-1}$. This can be achieved by setting $\boldsymbol{C} = \left(\boldsymbol{M}_1^{-1}\boldsymbol{M}_2\right)^\top$, where $\boldsymbol{M}_1^\top \boldsymbol{M}_1 = \boldsymbol{H}(\boldsymbol{\theta}^*)^{-1}$ and $\boldsymbol{M}_2^\top \boldsymbol{M}_2 = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^*)^{-1}$. The matrix square roots $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ can be computed using, e.g., singular value decomposition.

A problem with the adjustment method of Shaby (2014) is that it distorts the contributions of the prior distribution. Using the formula for the probability density function of a transformed random variable, one can show that the distribution of the adjusted samples is

$$\pi(\boldsymbol{\theta}_{\text{adj}} \mid \mathcal{Y}) \propto L(\boldsymbol{\theta}^* + \boldsymbol{C}^{-1}(\boldsymbol{\theta}_{\text{adj}} - \boldsymbol{\theta}^*); \mathcal{Y})\pi(\boldsymbol{\theta}^* + \boldsymbol{C}^{-1}(\boldsymbol{\theta}_{\text{adj}} - \boldsymbol{\theta}^*)).$$

However, the prior distribution reflects our prior knowledge about $\boldsymbol{\theta}$, and it should ideally not be affected when adjusting for the misspecification in L. If the prior is not overly informative and the sample size is large enough, this may not matter, as the contribution of the prior will be minimal. However, if that is not the case, we propose to additionally adjust the prior distribution before inference as

$$\pi_{\text{adj}}(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}^* + \boldsymbol{C}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)) \cdot |\boldsymbol{C}|, \tag{9}$$

such that the adjusted posterior samples have distribution

$$\pi(\boldsymbol{\theta}_{\text{adj}} \mid \mathcal{Y}) \propto L(\boldsymbol{\theta}^* + \boldsymbol{C}^{-1}(\boldsymbol{\theta}_{\text{adj}} - \boldsymbol{\theta}^*); \mathcal{Y})\pi(\boldsymbol{\theta}_{\text{adj}}).$$

Using the `rgeneric`/`cgeneric` framework, one can easily define a model in `R-INLA` with the adjusted prior distribution $\pi_{\text{adj}}(\boldsymbol{\theta})$.

As mentioned in Section 2.3, composite likelihoods are not valid likelihood functions. However, the theory in this section holds true for a wide class of loss functions that includes negative composite log-likelihoods. Thus, we can still perform the adjustment method if we exchange $\ell(\boldsymbol{\theta}; \mathcal{Y})$ with the composite log-likelihood in (4), and $\pi(\boldsymbol{\theta} \mid \mathcal{Y})$ with the pseudo-posterior distribution $\pi_c(\boldsymbol{\theta} \mid \mathcal{Y}) \propto L_c(\boldsymbol{\theta}; \mathcal{Y})\pi(\boldsymbol{\theta})$, where $L_c(\cdot) = \exp(\ell_c(\cdot))$.

4.2 Estimating $\boldsymbol{C}$ and $\boldsymbol{\theta}^*$

Here, we detail how to estimate the KLD minimiser $\boldsymbol{\theta}^*$ and the matrix $\boldsymbol{C}$ when inference is based on the composite log-likelihood in (4). Our approach can also be applied in settings where inference is based on valid likelihood functions or other types of loss functions.

The KLD minimiser $\boldsymbol{\theta}^*$ can be estimated by the mode of the posterior $\pi_c(\boldsymbol{\theta} \mid \mathcal{Y})$, denoted $\hat{\boldsymbol{\theta}}^*$. The estimator $\hat{\boldsymbol{\theta}}^*$ indeed provides a good approximation to $\boldsymbol{\theta}^*$ if the prior is not overly informative and the sample size is large enough, but other estimators, such as the maximum likelihood estimator, may be more suitable if this does not hold.

In order to estimate $\boldsymbol{C}$ one must first estimate $\boldsymbol{H}(\boldsymbol{\theta}^*)$ and $\boldsymbol{J}(\boldsymbol{\theta}^*)$. The first step towards estimating $\boldsymbol{J}(\boldsymbol{\theta}^*)$ is to compute the log-likelihood gradients $\nabla_{\boldsymbol{\theta}} \ell(\hat{\boldsymbol{\theta}}^*; \boldsymbol{y}_{i,-j} \mid y_i(\boldsymbol{s}_j))$ for all combinations of $j = 1, 2, \ldots, m$ and $i = 1, 2, \ldots, n$ that give $y_i(\boldsymbol{s}_j) > t$. These gradients can be computed analytically or estimated using numerical derivation methods. We then estimate $\boldsymbol{J}(\boldsymbol{\theta}^*)$ with

$$\begin{aligned} \hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}}^*) = \sum_{i,j} \sum_{(i',j') \in \Delta(i,j)} & \Big\{ I(y_i(\boldsymbol{s}_j) > t) I(y_{i'}(\boldsymbol{s}_{j'}) > t) \\ & \times \nabla_{\boldsymbol{\theta}} \ell(\hat{\boldsymbol{\theta}}^*; \boldsymbol{y}_{i,-j} \mid y_i(\boldsymbol{s}_j)) \Big(\nabla_{\boldsymbol{\theta}} \ell(\hat{\boldsymbol{\theta}}^*; \boldsymbol{y}_{i',-j'} \mid y_{i'}(\boldsymbol{s}_{j'})) \Big)^\top \Big\}, \end{aligned} \tag{10}$$

where $\Delta(i, j)$ is the set of neighbours of (i, j), i.e., $(i', j') \in \Delta(i, j)$ if and only if $\nabla_{\boldsymbol{\theta}} \ell(\cdot; \boldsymbol{y}_{i',-j'} \mid y_{i'}(\boldsymbol{s}_{j'}))$ is correlated with $\nabla_{\boldsymbol{\theta}} \ell(\cdot; \boldsymbol{y}_{i,-j} \mid y_i(\boldsymbol{s}_j))$. Summing over all non-correlated pairs of tuples introduces unnecessary noise that could cause the estimator to approximately equal zero (Lumley & Heagerty, 1999). In practice one can often compute (10) using a sliding window approach. This is an improvement over the proposed estimation methods of Shaby (2014), which require either that all log-likelihood terms are independent or that it is possible to simulate data from the true and unknown distribution of the data.

Estimation of $\boldsymbol{H}(\boldsymbol{\theta}^*)$ is often easier than that of $\boldsymbol{J}(\boldsymbol{\theta}^*)$, since the former is a matrix of expected values, whereas the latter is a covariance matrix. The law of large number implies that, for n and m large enough, a good estimator for $\boldsymbol{H}(\boldsymbol{\theta}^*)$ is simply $\hat{\boldsymbol{H}}(\hat{\boldsymbol{\theta}}^*) = -\nabla^2_{\boldsymbol{\theta}} \ell_c(\hat{\boldsymbol{\theta}}^*; \mathcal{Y})$. Thus, all we need for estimating $\boldsymbol{H}(\boldsymbol{\theta}^*)$ is to compute the Hessian of the composite log-likelihood terms at $\hat{\boldsymbol{\theta}}^*$. For users of `R-INLA`, this is especially simple, since the program returns the inverse of the Hessian matrix

$$\tilde{\boldsymbol{H}}(\hat{\boldsymbol{\theta}}^*) = -\nabla^2_{\boldsymbol{\theta}} \pi_c(\hat{\boldsymbol{\theta}}^* \mid \mathcal{Y}) = -\nabla^2_{\boldsymbol{\theta}} \ell_c(\hat{\boldsymbol{\theta}}^*; \mathcal{Y}) - \nabla^2_{\boldsymbol{\theta}} \log \pi(\hat{\boldsymbol{\theta}}^*).$$

Thus, if the prior is not overly informative and the sample size is large enough, we can set estimate $\boldsymbol{H}(\boldsymbol{\theta}^*)$ with $\tilde{\boldsymbol{H}}(\hat{\boldsymbol{\theta}}^*)$. If this is not the case, we

can still estimate $\boldsymbol{H}(\boldsymbol{\theta}^*)$ by simply subtracting the contribution of the prior distribution from $\tilde{\boldsymbol{H}}(\hat{\boldsymbol{\theta}}^*)$.

If we adjust the prior distribution as in (9), it may be necessary to run `R-INLA` twice: once for estimating $\boldsymbol{\theta}^*$ and $\boldsymbol{C}$, and once for performing inference with the adjusted prior.

A small numerical example is shown in the supplementary material, for demonstrating that the adjustment method is able to recover the frequency properties of a posterior distribution that is based on a misspecified likelihood.

5 Simulation study

We now conduct a simulation study to demonstrate our proposed workflow for modelling spatial extremes. Given a set of extreme realisations from simulated data we show how to compute relevant statistics of the data and how to use these for making an informed decision about the appropriate models for the standardising functions $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$. Then, we discuss details on how to define the SPDE mesh and on performing inference with `R-INLA` and the composite likelihood. Finally, we adjust the posterior distribution for possible misspecification and we evaluate the performance of the model fit.

We sample $n = 10^4$ realisations of a spatial Gaussian random field $\mathcal{Y} = \{Y_i(\boldsymbol{s}) : i = 1, \ldots, n, \boldsymbol{s} \in \mathcal{S}\}$, observed on a regular grid $\mathcal{S}$ with resolution 1×1 and size 100×100. The spatial Gaussian random field has a Matérn covariance function (5) with parameters $\sigma^2 = 1$, $\nu = 1$ and $\rho = 40$, and an additional nugget effect with variance 0.1^2. All the samples are created using an SPDE approximation. In order to model threshold exceedances with the spatial conditional extremes model, we transform the observations to have Laplace marginals using the probability integral transform. We then choose a threshold t equal to the 99.9% quantile of the Laplace distribution.

As a first step, we examine extremal dependence in the available data. If we (correctly) assume stationarity and isotropy, we can denote the extremal correlation coefficient as $\chi_p(\boldsymbol{s}_1, \boldsymbol{s}_2) \equiv \chi_p(d)$, where $d = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$. We estimate $\chi_p(d)$ empirically using a sliding window approach, i.e., for any value of d, we iterate over all location pairs $(\boldsymbol{s}, \boldsymbol{s}') \in \mathcal{S}^2$ satisfying $|d - \|\boldsymbol{s} - \boldsymbol{s}'\|| < \delta$, for some small tolerance $\delta > 0$, and then we count the number of times that $Y(\boldsymbol{s}) > F^{-1}(p)$ given that $Y(\boldsymbol{s}') > F^{-1}(p)$, where $F^{-1}(\cdot)$ is the quantile function of the Laplace distribution. We here choose $\delta = 0.5$. Estimators for $\chi_p(d)$ are displayed in the top-left subplot of Figure 1. Since the data have a Gaussian copula, we know that $\chi(d) = 0$ for all $d > 0$, meaning that $\chi_p(d)$ is far away from its limit $\chi(d)$ at small distances. Even if $\chi(d)$ is unknown

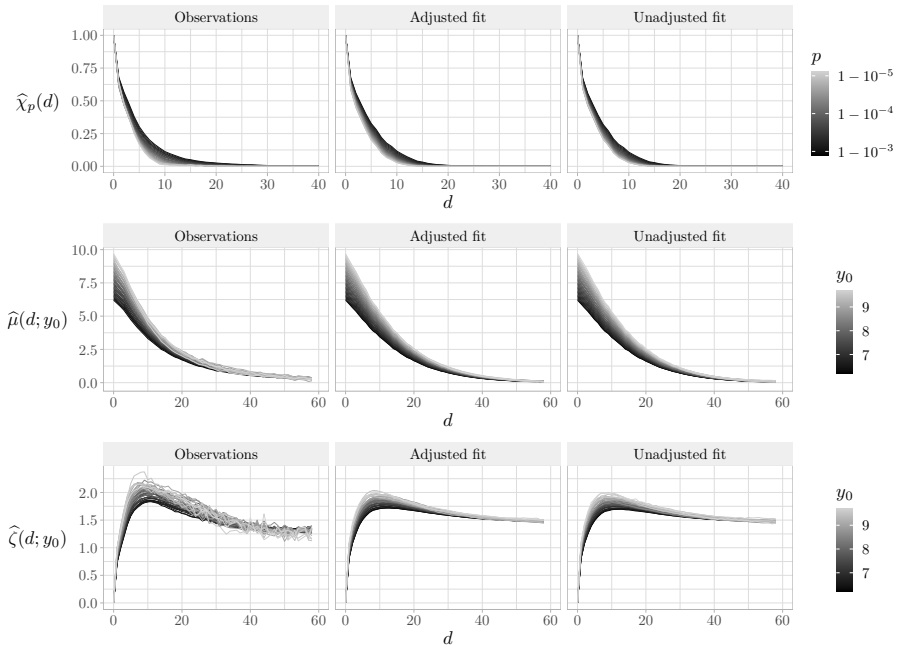


Figure 1: *Empirical estimators of* $\chi_p(d)$, $\mu(d; y_0)$ *and* $\zeta(d; y_0)$ *(top to bottom) from three different data sources. The leftmost column displays empirical estimators using the original data, while the two rightmost columns displays empirical estimators using data simulated from the adjusted and the unadjusted model fits, respectively.*

in practice, we can here observe a clear trend of weakening dependence at increasing threshold levels, implying that the limit has not yet been reached. This demonstrates the need for a model that allows for flexible modelling of sub-asymptotic dependence, such as the spatial conditional extremes model.

To perform inference with the spatial conditional extremes model from (3), we must decide upon models for $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$. The limiting forms of these functions as $t \to \infty$ are already known for a spatial Gaussian random field (Wadsworth & Tawn, 2022). However, we here assume that the distribution of the data is unknown. Additionally, since we have chosen a finite threshold t where $\chi_p(d)$ is far away from its limit $\chi(d)$, other models for $a(\cdot)$ and $b(\cdot)$ may fit the data better than the known limiting forms. To examine the shape of the standardising functions, we (correctly) assume stationarity in the sense that all model parameters are independent of the choice of conditioning sites, and we assume that $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ only depend on the distance $d = \|\boldsymbol{s} - \boldsymbol{s}_0\|$ and threshold exceedance y_0, meaning that we can define the standardising functions as $a(d; y_0)$ and $b(d; y_0)$ analogously. With these assumptions, we can visualise the forms of $a(d; y_0)$ and $b(d; y_0)$ by empirically computing conditional means and

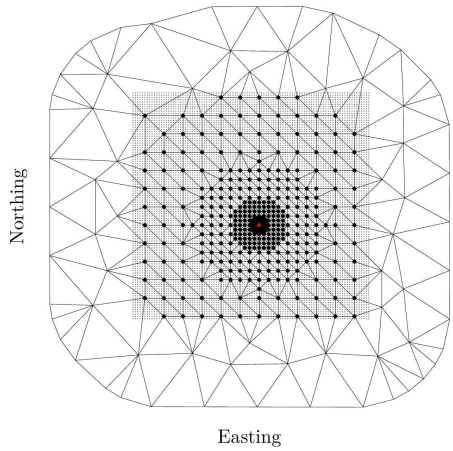


Figure 2: *Given a conditioning site* $\boldsymbol{s}_0$ *(displayed with (▲)), locations used for inference are displayed as big black dots* (•) *and locations in* $\mathcal{S}$ *that are not used for inference are displayed as small dots* (·)*. The SPDE mesh is displayed using black lines.*

variances of the data. In our model, all random variables with distance d from $\boldsymbol{s}_0$ have conditional mean $\mu(d; y_0) = a(d; y_0)$ and conditional variance $\zeta^2(d; y_0) = \sigma^2(d)b^2(d; y_0) + \tau^{-1}$, where $\sigma^2(d)$ is the variance of the residual field at distance d from the conditioning site, and τ is the precision of the nugget effect. Similarly to $\widehat{\chi}_p(d)$, the empirical conditional moments of the data can be computed using a sliding window approach. However, this time, the window must slide over both values of d and y_0. We choose a rectangular window with a width of 1 in the d-direction and a width of 0.1 in the y_0-direction. The conditional moment estimators are displayed in the leftmost column of Figure 1. The conditional mean, $\widehat{\mu}(d; y_0)$, is equal to y_0 at $d = 0$, and then seems to decay exponentially towards zero as d increases. This fits well with the proposed model in (2) if we set $\Delta = 0$. The conditional variance is zero at $d = 0$, and then it increases as we move away from the conditioning site and towards "the edge of the storm". Here, $\zeta(d; y_0)$ is at its largest, as it is uncertain if observations are "inside the storm", i.e., extreme, or "outside the storm", i.e., non-extreme. This is also where $\zeta(d; y_0)$ varies the most as a function of y_0. Moving further away from the conditioning site, $\zeta(d; y_0)$ decreases to a constant, as we are certainly "outside the storm", so the variance should not depend on y_0 anymore. This fits well together with a model where $b(d; y_0) = y_0^{\beta(d)}$ and where $\beta(d)$ decays to zero as the distance increases. We choose to follow Richards et al. (2022) in assuming that $\beta(d) = \beta_0 \exp(-(d/\lambda_b)^{\kappa_b})$, with $0 < \beta_0 < 1$ and $\lambda_b, \kappa_b > 0$.

As seen in Figure 1, the largest changes in $\mu(d; \cdot)$ and $\zeta(d; \cdot)$ seem to occur when d is small. However, the majority of locations in $\mathcal{S}$ are located far

Table 1: *Prior distributions for all model parameters.* $\mathcal{N}(\mu, \sigma^2)$ *denotes the Gaussian distribution with mean* μ *and variance* σ^2*. We give* τ *a penalised complexity (PC) prior such that* $P(\tau^{-1/2} > 1) = 0.95$*. Additionally,* ρ *and* σ *are given the joint PC prior of Fuglstad et al. (2019) such that* $P(\rho < 60) = 0.95$ *and* $P(\sigma > 4) = 0.05$*.*

$\tau \sim \text{PC}(1, 0.95)$,	$\log(\lambda) \sim \mathcal{N}(3, 4^2)$,	$\log(\kappa) \sim \mathcal{N}(0, 3^2)$,
$\sigma \sim \text{PC}(4, 0.05)$,	$\rho \sim \text{PC}(60, 0.95)$,	$\log(\frac{\beta_0}{1-\beta_0}) \sim \mathcal{N}(0, 2^2)$,
$\log(\lambda_b) \sim \mathcal{N}(3, 4^2)$,	$\log(\kappa_b) \sim \mathcal{N}(0, 3^2)$,	

away from $\boldsymbol{s}_0$. To account for this and give more weight to close-by locations, we discard some of the observations far away from $\boldsymbol{s}_0$ during inference, which also leads to increased inference speed. Figure 2 shows an example of the locations used to perform inference for one specific conditioning site. We stress that these locations can vary for each conditioning site used during inference.

The SPDE approach for modelling $Z_b(\cdot)$ requires that we define a triangulated mesh. Our proposed constraining method from Section 3.2 requires that a mesh node is located at each conditioning site used for inference. Furthermore, the mesh should be quite dense close to the conditioning sites to correctly capture the changes in $b(\cdot)$. Therefore, we define the mesh so that a mesh node is placed at each location used for inference. This can be problematic when performing inference with a composite likelihood that depends on multiple conditioning sites, meaning that the mesh has to be dense "everywhere" in $\mathcal{S}$, which leads to computationally demanding inference. Consequently, we choose to model $Z_b(\cdot)$ with a different mesh for each conditioning site used in the composite likelihood. Modelling different realisations of a random field with different mesh designs is not a readily available option in `R-INLA`, but this can be easily implemented using the `rgeneric`/`cgeneric` framework. An example of a mesh design for one specific conditioning site is displayed in Figure 2.

Our chosen models for $a(\cdot)$ and $b(\cdot)$ are implemented using the `cgeneric` framework, and inference is performed with `R-INLA`. The chosen priors for all the model parameters are described in Table 1. The priors are weakly informative, but with quite large variances. Using all locations in $\mathcal{S}$ as conditioning sites in the composite likelihood is computationally demanding, so we define a regular sub-grid $\mathcal{S}_0$ with resolution 6×6 and build the composite likelihood using these $|\mathcal{S}_0| = 256$ conditioning locations. The post-hoc adjustment procedure from Section 4 is then applied to robustify the model fit. Due to the large amount of available data we do not find it necessary to adjust the prior distribution as proposed in (9).

Figure 3 displays the adjusted and unadjusted posterior distributions of all model parameters. We see that the working assumption of indepen-

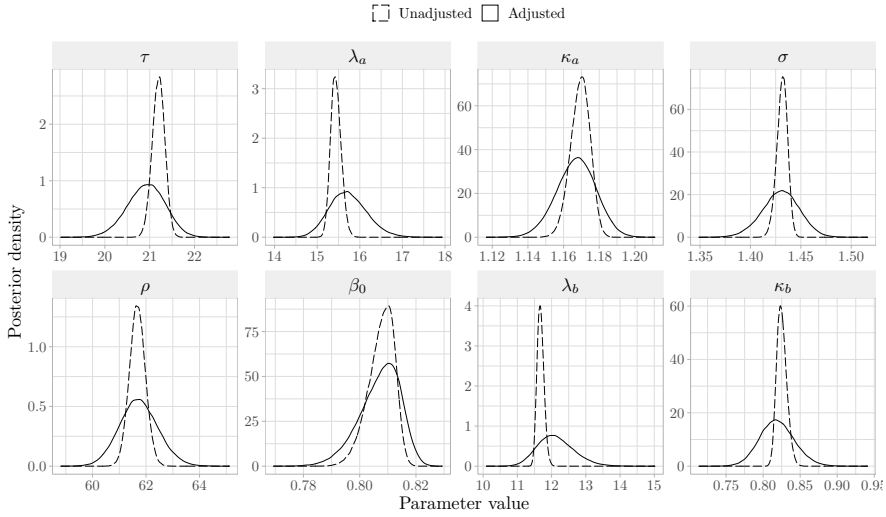


Figure 3: *Posterior distributions for all model parameters from the adjusted (solid) and the unadjusted (dashed) model fits.*

dence in the composite likelihood leads to overconfidence and too focused posterior distributions, and that the adjustment method therefore increases the posterior variance to account for this misspecification. To examine the performance of our model fits, we simulate 10^5 extreme spatial fields from each fitted model, and compute $\widehat{\chi}_p(d)$, $\widehat{\mu}(d; y_0)$ and $\widehat{\zeta}(d; y_0)$ using the simulated extremes. The estimators are displayed in the two rightmost columns of Figure 1. The properties of the model fits are similar to those of the original data. There are some noticeable differences in the estimated conditional variance, which probably stems from a too simple model for $b(d; y_0)$. However, tailoring the perfect model choice for $b(d; y_0)$ is not the focus of this simulation study. Although adjusting posteriors plays a big role in properly quantifying posterior uncertainty, there are no clear differences between the point estimates from the two model fits in Figure 1. This is not very surprising, as these estimators are different types of sample means, that might be less affected by changes in the posterior variances.

Finally, we wish to quantitatively compare the adjusted model fit with the unadjusted model fit, to find out which one performs best. We choose not to compare the fits by evaluating frequency properties, as in the toy example in the supplementary material, because accurate estimation of $\boldsymbol{\theta}^*$ and the repetition of the high-dimensional simulation study hundreds of times is too computationally demanding with our computational resources. Additionally, such comparisons are impossible to perform for most real-life applications with finite amounts of available data. Instead, we choose to compare the model fits by computing log-scores (e.g., Gneiting & Raftery,

2007) for a test data set that has not been used during inference. Marginal composite likelihoods may be estimated using Monte Carlo estimation: given n_s samples $\boldsymbol{\theta}_1, \ldots \boldsymbol{\theta}_{n_s}$ from the posterior distribution $\pi(\boldsymbol{\theta} \mid \mathcal{Y})$, the marginal composite likelihood for a new set of observations $\mathcal{Y}_0$ is estimated as $\widehat{L}_c(\mathcal{Y}_0) = \frac{1}{n_s} \sum_{i=1}^{n_s} L_c(\boldsymbol{\theta}_i; \mathcal{Y}_0)$, where $L_c(\cdot)$ is the composite likelihood for the spatial conditional extremes model. We then denote $\log(\widehat{L}_c(\mathcal{Y}_0))$ as the estimated log-score. We sample 5×10^4 new realisations of data from the true model and locate all threshold exceedances from the 256 conditioning sites used for performing inference. Log-scores are then estimated using $n_s = 1000$ posterior samples. This results in a log-score of -2502219 for the adjusted model fit, and -2504558 for the unadjusted model fit, meaning that the adjusted model fit attains the highest log-score, with a difference of 2338. Nonparametric bootstrapping of the 5×10^4 realisations of the spatial Gaussian random field is performed to examine if the difference in log-score is significant. Using 5000 bootstrap samples, we find that the adjusted log-score always is larger than the unadjusted log-score, with a difference between 1000 and 4500. We conclude that the adjusted posterior performs better than the unadjusted posterior, even though they both provide good point estimates and reasonable fits to the simulated data.

6 Case study: Extreme precipitation in Norway

We apply our proposed methodology to the modelling of extreme hourly precipitation in Norway. Data are presented in Section 6.1 and the inference is described in Section 6.2. Results are presented and evaluated in Section 6.3.

6.1 Data

We consider 1×1 km^2 maps of mean hourly precipitation, produced by the Norwegian Meteorological Institute by processing raw reflectivity data from the weather radar located in Rissa (63°41′26″N, 10°12′14″E) in central Norway. Such maps are available online (https://thredds.met.no), dating back to 1 January 2010. We extract data from a rectangular domain, close to the Rissa radar, of size 31×31 km^2. Denote the set of all grid points in the rectangular domain as $\mathcal{S}$. We then have $|\mathcal{S}| = 961$ unique locations containing hourly precipitation estimates. A map containing $\mathcal{S}$ and the Rissa radar is displayed in Figure 4. For each $\boldsymbol{s} \in \mathcal{S}$, we extract all hourly observations from the summer months (June, July and August) for the years 2010–2021. Removal of missing data gives a total of 25,512 observations at each location. The number of observations with positive precipitation amounts at each location varies from 6,000 to 17,000. These large differences are likely numerical artefacts from the processing method of the Norwegian

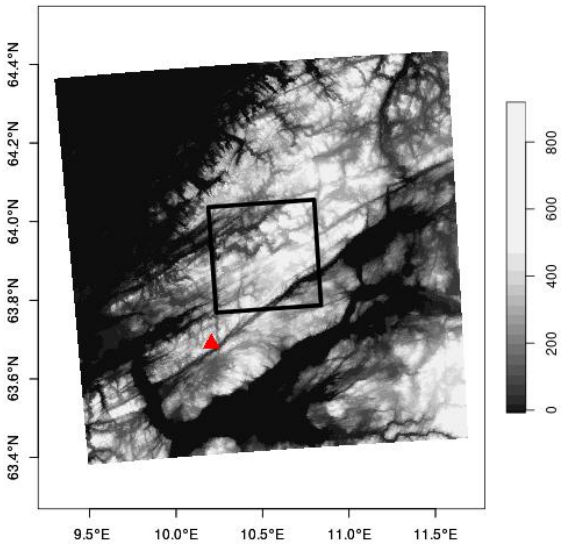


Figure 4: *Elevation (m) map, over the Fosen area in central Norway. The study area $\mathcal{S}$ is located inside the black rectangle, and the Rissa radar is displayed using a triangle (▲).*

Meteorological Institute. Consequently, we set all observations smaller than 0.1 mm precipitation equal to 0. This gives a total of between 3,500 and 4,500 positive precipitation observations at each location.

6.2 Modelling and inference

The conditional extremes model in (3) is defined for a random process with Laplace margins. Thus, in order to perform inference with the conditional extremes model, we standardise the marginal distributions of the precipitation data using the probability integral transform. This is described in the supplementary materials.

Initial data exploration shows that the threshold t must be very large for a model on the form $a(d; y_0) = \alpha(d)y_0$ and $b(d; y_0) = y_0^{\beta(d)}$ to provide a good fit. Consequently, we choose a threshold equal to the 99.97% quantile of the Laplace distribution, which yields between 0 and 5 threshold exceedances at each conditioning site. See the supplementary materials for more details and discussion on this choice. Estimators for $\chi_p(d)$, $\mu(d; y_0)$ and $\zeta(d; y_0)$ using this threshold are displayed in the leftmost column of Figure 5. Based on the lack of changes in $\widehat{\zeta}(\cdot; y_0)$ as y_0 varies, we choose to model $b(d; y_0)$ as a function not depending on y_0. We choose the model

$$b(d; y_0) \equiv b(d) = 1 + b_0 \exp\left\{-(d/\lambda_b)^{\kappa_b}\right\},$$

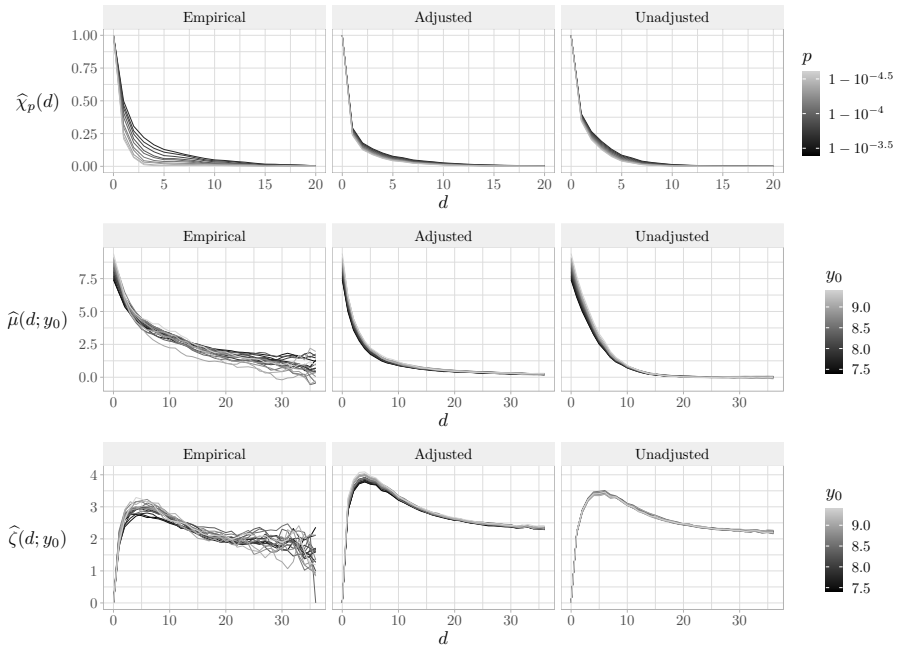


Figure 5: *Empirical estimators for* $\chi_p(d)$, $\mu(d; y_0)$ *and* $\zeta(d; y_0)$ *from three different data sources. The leftmost column displays empirical estimators using the original data, while the two rightmost columns displays empirical estimators using simulated data from the adjusted and the unadjusted model fits, respectively.*

with positive parameters b_0, λ_b and κ_b, while we use the model (2) for $a(\cdot)$, just as in Section 5.

With only one or two threshold exceedances at most conditioning sites, separate inference for each conditioning site would be highly challenging. Inference is therefore performed using `R-INLA` and the composite likelihood, based on every single conditioning site in $\mathcal{S}$. However, we remove the two last years of the data before performing inference, so these can be used for comparing the performance of the adjusted and the unadjusted model fits. Just as in Section 5, some of the observations far away from the conditioning sites are discarded during inference, and we also define different triangulated meshes for each conditioning site. The Matérn smoothness parameter ν is fixed to a value of 1.5. Prior distributions are set equal to those in Table 1, except that we exchange the parameter β_0 from Section 5 with the parameter b_0, where we place a Gaussian prior on $\log(b_0)$ with zero mean and a variance of 4^2. Finally, the post hoc adjustment method is performed on the output of `R-INLA`. Once more, we have a large enough sample size that we choose not to adjust the prior distribution as in (9). Estimates for $\boldsymbol{\theta}^*$ and $\boldsymbol{H}(\boldsymbol{\theta}^*)$ are provided directly from `R-INLA`, while $\boldsymbol{J}(\boldsymbol{\theta}^*)$ is estimated

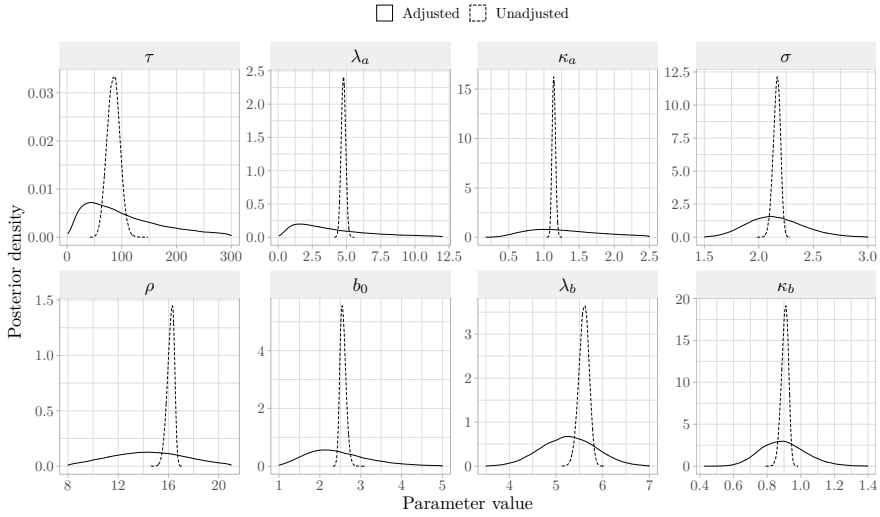


Figure 6: *Posterior distributions for all model parameters of the adjusted (solid) and the unadjusted (dashed) model fits.*

using (10) with a sliding window that has a width of 10 hours.

6.3 Results

We simulate 10^5 extreme realisations from the adjusted and unadjusted model fits. Statistics of the simulated data are displayed in the two rightmost columns of Figure 5. There are noticeable differences between the samples from the two model fits. However, both model fits seem to capture a large part of the trends in the transformed precipitation data well. Interestingly, the adjusted conditional second moments $\widehat{\zeta}(d; y_0)$ are more different from those of the original data than those of the unadjusted model fit. However, this is not reflected in the estimated extremal correlation coefficients, $\widehat{\chi}_p(d)$. As our main goal is to capture the trends in $\chi_p(d)$, we see that the adjusted model fit seems to outperform the unadjusted one overall, especially for higher values of p. None of the model fits fully capture the rate of weakening dependence with increasing thresholds, which probably requires a more complex model for $b(d; y_0)$. As discussed in the supplementary materials, this is outside the scope of this paper and it would require further investigation in future research. However, any such model extension can easily be implemented using our proposed `R-INLA` methodology.

Posterior distributions for all parameters of the two model fits are displayed in Figure 6. There are considerable differences between the adjusted and the unadjusted posterior distributions. The latter one again too focused due to the working assumption of independence in the composite

likelihood.

To further compare the two model fits, composite log-scores are computed using the last two years of the available data. These are estimated using $n_s = 1000$ posterior samples. This results in a composite log-score of -96520 for the unadjusted model fit, and -88029 for the adjusted model fit, meaning that the adjusted model fit seems to perform considerably better. Nonparametric bootstrapping of all time points in the test data is performed to examine if the difference is significant. Using 5000 bootstrap samples, we find that the difference in composite log-score is significantly different from zero at a 0.1% significance level, so we conclude that the adjusted model fit outperforms the unadjusted model fit.

7 Conclusion

We propose an efficient workflow for robust modelling of spatial high-dimensional extremes using the spatial conditional extremes model with a composite likelihood and `R-INLA`, and a post hoc adjustment method that corrects for possible model misspecification. The workflow is demonstrated and shown to perform well in a large-scale simulation study, where we also propose a methodology for selecting appropriate forms for the standardising functions $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$. Finally, the workflow is applied for modelling spatial high-dimensional extremes of Norwegian precipitation data. The methodology performs well, and we are able to capture the main extremal dependence trends in the data.

In developing our workflow, we describe a flaw in previously-used constraining methods for the residual field in the spatial conditional extremes model, and we develop a novel constraining method that is fast and easy to use when performing inference with `R-INLA`. We also propose and demonstrate a general methodology for defining and implementing a large variety of spatial conditional extremes models in `R-INLA` using the `rgeneric`/`cgeneric` frameworks. Additionally, we propose an improved extension to the post hoc adjustment method that allows for correct model contributions from the prior distribution.

For transforming precipitation data onto Laplace marginals, a nonparametric method is used for estimating the marginal distributions of the original data. This method can be problematic if the aim is to estimate properties of the original and untransformed process. Further work should therefore focus on improving the transformation method when modelling extremes with the spatial conditional extremes model. Additionally, even though the spatial conditional extremes model provides good fits to the data in both the simulation study and the case study, there are still some small differences between properties of the data and properties of the model fits.

These differences can probably be reduced by choosing better, possibly more complex, parametric or semiparametric forms for $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$, such as, e.g., $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = \alpha(\boldsymbol{s}; \boldsymbol{s}_0, y_0) y_0$ or $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = y_0^{\beta(\boldsymbol{s}; \boldsymbol{s}_0, y_0)}$. Further work should therefore focus on the theoretical properties of more complex models for $a(\cdot)$ and $b(\cdot)$, and on how to best perform model selection with the spatial conditional extremes model. Finally, the selection of a threshold t for the spatial conditional extremes model can have great importance for the resulting model fit, as seen in the case study. However, to the best of our knowledge, little attention has so far been given to the problem of threshold selection in this context. Further work should therefore focus on methods for choosing thresholds that are large enough to provide a somewhat correct model fit and small enough to perform inference with low uncertainty.

Acknowledgements

The authors are grateful to Jordan Richards, Håvard Rue and Geir-Arne Fuglstad for many helpful discussions.

Funding Raphaël Huser was partially supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR- CRG2020-4394.
Conflict of interest The authors report there are no competing interests to declare.
Code and data availability The necessary code and data for achieving these results are available online at
https://github.com/siliusmv/spatialConditionalExtremes.

Supplementary material

S1 Post hoc adjustment toy example

Shaby (2014) demonstrates that his proposed adjustment method is able to properly recover the correct frequency properties of the posterior distribution. Here, we show that the same holds when extending the post hoc method for adjusting model fits from `R-INLA`, by examining posterior frequency properties after modelling a spatial Gaussian random field with an SPDE approximation of low rank.

Inside the spatial domain $\mathcal{S} = [0, 25] \times [0, 25]$ we sample n independent realisations of a spatial Gaussian random field with a Matérn covariance function, which we observe at 400 random locations. The Matérn covariance

Table S1: *Coverage percentages for unadjusted and adjusted credible intervals using the SPDE approach with a coarse mesh.*

Aim	τ	τ_{adj}	ρ	ρ_{adj}	σ	σ_{adj}
90%	48%	93%	91%	90%	90%	90%
95%	55%	97%	95%	95%	95%	96%
99%	69%	99%	99%	98%	100%	99%

function is

$$\text{Cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}')) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|\boldsymbol{s}-\boldsymbol{s}'\|)^{\nu}K_{\nu}(\kappa\|\boldsymbol{s}-\boldsymbol{s}'\|), \tag{S1}$$

where σ^2 is the marginal variance, $\nu > 0$ is the smoothness parameter and $\rho = \sqrt{8\nu}/\kappa$ is the range parameter of $Z(\boldsymbol{s})$. Furthermore, K_{ν} is the modified Bessel function of the second kind and order ν. Our spatial Gaussian random field has variance parameter $\sigma^2 = 1$, range parameter $\rho = 12$ and known smoothness parameter $\nu = 1.5$. We also add a Gaussian nugget effect with a precision of $\tau = 100$ to the random field. Parameter estimation is then performed using an SPDE approximation of low rank, i.e., based on a coarse triangulated mesh used to discretise the spatial domain. Such low-rank approximations are typically unable to capture all the variability in the data, which means that the nugget effect has to explain a large percentage of the variance, leading to underestimation of the precision τ. Thus, we expect the KLD minimiser $\boldsymbol{\theta}^*$ to be different from the true parameters $\boldsymbol{\theta} = (\tau, \rho, \sigma)^T$. To estimate the unknown KLD minimiser $\boldsymbol{\theta}^*$, we simulate $n = 10^4$ realisations of the Gaussian Matérn field and compute the maximum likelihood estimator for the misspecified SPDE model. This gives $\boldsymbol{\theta}^* = (\tau^*, \rho^*, \sigma^*) \approx (13.0, 14.5, 1.2)^T$. As expected, τ is severely underestimated, while ρ and σ are slightly overestimated.

For examination of frequency properties, we then sample $n = 200$ new realisations of the spatial field, and perform Bayesian inference using `R-INLA`. We assign τ a gamma prior with shape 1 and scale 2×10^4, while ρ and σ are given a joint penalised complexity (PC) prior (Simpson et al., 2017; Fuglstad et al., 2019), setting $\text{P}(\rho < 12) = 0.5$ and $\text{P}(\sigma > 1) = 0.5$. Inference is performed, the posterior distribution is adjusted as described in Section 4.2 of the main paper, and credible intervals are created for both the adjusted and the unadjusted model fits. For this simple toy example, we do not focus on adjusting the prior distribution as described in Section 4.1. We repeat this procedure 300 times, each time sampling $n = 200$ new realisations which we observe at the same 400 locations. Coverage frequencies can then be evaluated by examining how many of the 300 credible intervals include the KLD minimiser $\boldsymbol{\theta}^*$.

Table S1 displays the estimated coverage probabilities detailing how often the parameters of $\boldsymbol{\theta}^*$ are included in their respective credible intervals. The adjustment of the posterior yields a considerable improvement for τ. The unadjusted frequency properties of ρ and σ, however, are already good, and our adjustment method does not deteriorate the credible intervals for these parameters.

S2 Case study prerequisites

In order to perform inference with the conditional extremes model for a random process $X(\boldsymbol{s})$, one must first standardise it to a random process $Y(\boldsymbol{s})$ with Laplace margins. This is performed using the probability integral transform (Keef et al., 2013):

$$Y(\boldsymbol{s}) = \begin{cases} \log\left\{2F_{X(\boldsymbol{s})}(X(\boldsymbol{s}))\right\}, & X(\boldsymbol{s}) < F_{X(\boldsymbol{s})}(1/2) \\ -\log\left\{2\left[1 - F_{X(\boldsymbol{s})}(X(\boldsymbol{s}))\right]\right\}, & X(\boldsymbol{s}) \geq F_{X(\boldsymbol{s})}(1/2), \end{cases}$$

where $F_{X(\boldsymbol{s})}$ is the marginal distribution function of the random variable $X(\boldsymbol{s})$. We estimate the marginal distribution functions as the site-wise empirical distribution function of $X(\boldsymbol{s})$. However, independent standardisation of data at each location can lead to an unrealistic lack of smoothness in the transformed process $Y(\boldsymbol{s})$. Therefore, we apply a sliding window approach for computing the empirical distribution function, where the distribution at location $\boldsymbol{s}$ is estimated as the empirical distribution function of pooled data from all locations $\boldsymbol{s}'$ such that $\|\boldsymbol{s} - \boldsymbol{s}'\| \leq r$ for some radius r. Based on exploratory analysis we find $r = 5$ km to yield a realistic degree of smoothness in the estimated marginal distributions of $X(\boldsymbol{s})$ (results not shown).

A problem when modelling precipitation is that the empirical distribution has a point mass at zero. This leads to $Y(\boldsymbol{s})$ having a truncated Laplace distribution with a point mass, which can cause problems during inference. In order for $Y(\boldsymbol{s})$ to follow a non-truncated Laplace distribution, we choose to remove all zeros from the process $X(\boldsymbol{s})$ and only focus on positive precipitation. This makes us unable to model the absence of precipitation, which can lead to a slight overestimation of return levels for spatially aggregated precipitation. However, applying the fitted model for estimating properties of the untransformed process $X(\boldsymbol{s})$ is outside the scope of this paper. We believe that our choice of removing all zeros and estimating marginals using empirical distribution functions of the positive precipitation values is acceptable given the aim of our paper. In future research, we plan to properly model precipitation intermittence by appropriately accounting for the point mass at zero.

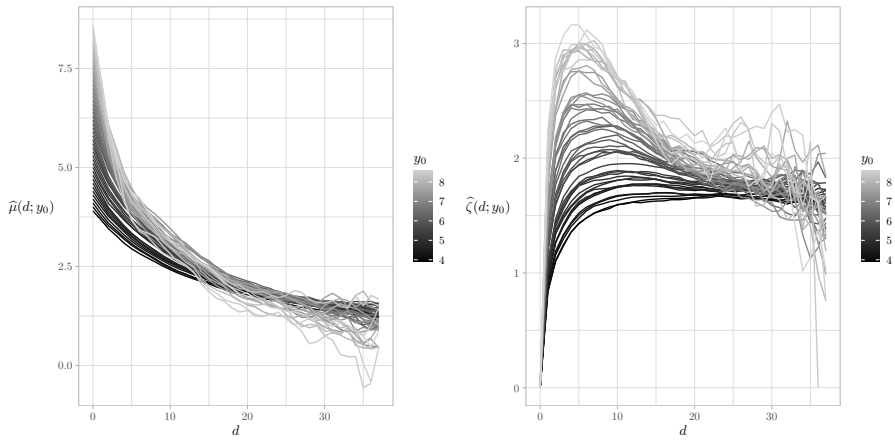


Figure S1: *Empirical estimators for* $\mu(d; y_0)$ *and* $\zeta(d; y_0)$ *using the transformed precipitation data for* $y_0 > 4$ *(99% quantile of the Laplace distribution)*

Similarly to the simulation study in Section 5 of the main paper, we examine extremal correlation coefficients and empirical conditional moments of the data in order to propose a good model for the extremes. The spatial domain in the case study is small enough that we can assume stationarity in the data, meaning that we can employ the same estimation methods as in the simulation study. Empirical conditional moments of the data are displayed in Figure S1. These estimators imply that the threshold t must be chosen quite large for performing successful modelling with the spatial conditional extremes model. If the threshold is chosen too low, we experience crossing in the conditional mean, i.e., for $y_1 \neq y_2$, $\widehat{\mu}(d; y_1)$ is both smaller and larger than $\widehat{\mu}(d; y_2)$ depending on the value of d. This means that a model for $a(\cdot)$ on the form $a(d; y_0) = \alpha(d)y_0$ becomes unsuitable. Furthermore, there is a clear change in the shape of the conditional variance as y_0 increases, and the spread in variance at "the edge of the storm" is so large that a model on the form $b(d; y_0) = y_0^{\beta(d)}$ would require $\beta(d) \approx 2$ for small distances d. However, $\beta(d) > 1$ leads to an ill-defined model (Wadsworth & Tawn, 2022). A more flexible model of the form $a(d; y_0) = \alpha(d, y_0)y_0$, that allows crossing, and $b(d; y_0) = y_0^{\beta(d,y_0)}$, that allows $\beta(d, y_0) > 1$ for small values of y_0, would probably fit well to the data, and could easily be implemented within the `rgeneric`/`cgeneric` framework. However, developing complex new variants of the spatial conditional extremes model is outside the scope of this paper. Consequently, we instead choose a large threshold t equal to the 99.97% threshold of the Laplace distribution, which removes the problems of crossing and excessively large values of $\beta(d)$. As we have approximately 4000 positive observations at each location, this corresponds to a mean of 1.2 threshold exceedances at each conditioning site. In practice, it yields

between 0 and 5 threshold exceedances at each conditioning site.

References

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, *37*(1), 51–58. https://doi.org/10.1214/aoms/1177699597

Castro-Camilo, D., Huser, R., & Rue, H. (2019). A spliced gamma-generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(3), 517–534. https://doi.org/10.1007/s13253-019-00369-z

Chandler, R. E., & Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, *94*(1), 167–183. https://doi.org/10.1093/biomet/asm015

Coles, S., Heffernan, J., & Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, *2*(4), 339–365. https://doi.org/10.1023/A:1009963131610

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, *27*(2), 161–186. https://doi.org/10.1214/11-STS376

Davison, A. C., Huser, R., & Thibaud, E. (2019). Spatial extremes. In A. E. Gelfand, M. Fuentes, J. A. Hoeting, & R. L. Smith (Eds.), *Handbook of environmental and ecological statistics* (pp. 711–744). Chapman; Hall/CRC.

Engelke, S., Opitz, T., & Wadsworth, J. (2019). Extremal dependence of random scale constructions. *Extremes*, *22*(4), 623–666. https://doi.org/10.1007/s10687-019-00353-3

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, *114*(525), 445–452. https://doi.org/10.1080/01621459.2017.1415907

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, *31*(4), 1208–1211.

Hazra, A., Huser, R., & Bolin, D. (2021). Realistic and fast modeling of spatial extremes over large geographical domains. https://doi.org/10.48550/ARXIV.2112.10248

Heffernan, J. E., & Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, *17*(2), 537–571. https://doi.org/10.1214/105051606000000835

Heffernan, J. E., & Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 497–546. https://doi.org/10.1111/j.1467-9868.2004.02050.x

Huser, R., & Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, *114*(525), 434–444. https://doi.org/10.1080/01621459.2017.1411813

Huser, R., & Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews (WIREs): Computational Statistics*, *14*(1), e1537. https://doi.org/10.1002/wics.1537

Ingebrigtsen, R., Lindgren, F., & Steinsland, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, *8*, 20–38. https://doi.org/10.1016/j.spasta.2013.06.002

Keef, C., Papastathopoulos, I., & Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, *115*, 396–404. https://doi.org/10.1016/j.jmva.2012.10.012

Kleijn, B., & van der Vaart, A. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electronic Journal of Statistics*, *6*, 354–381. https://doi.org/10.1214/12-EJS675

Koch, E., Koh, J., Davison, A. C., Lepore, C., & Tippett, M. K. (2021). Trends in the extremes of environments associated with severe U.S. thunderstorms. *Journal of Climate*, *34*(4), 1259–1272. https://doi.org/10.1175/JCLI-D-19-0826.1

Koh, J., Pimont, F., Dupuy, J.-L., & Opitz, T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *The Annals of Applied Statistics*, *17*(1), 560–582. https://doi.org/10.1214/22-AOAS1642

Krupskii, P., & Huser, R. (2022). Modeling spatial tail dependence with Cauchy convolution processes. *Electronic Journal of Statistics*, *16*(2), 6135–6174. https://doi.org/10.1214/22-EJS2081

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1–25. https://doi.org/10.18637/jss.v063.i19

Lindgren, F., Rue, H., & 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. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Lumley, T., & Heagerty, P. (1999). Weighted empirical adaptive variance estimators for correlated data regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(2), 459–477. https://doi.org/10.1111/1467-9868.00187

Opitz, T., Huser, R., Bakka, H., & Rue, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, *21*(3), 441–462. https://doi.org/10.1007/s10687-018-0324-x

Pauli, F., Racugno, W., & Ventura, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, *21*(1), 149–164.

Ribatet, M., Cooley, D., & Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, *22*(2), 813–845. http://www.jstor.org/stable/24310036

Richards, J., Tawn, J. A., & Brown, S. (2022). Modelling extremes of spatial aggregates of precipitation using conditional methods. *The Annals of Applied Statistics*, *16*(4), 2693–2713. https://doi.org/10.1214/22-AOAS1609

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

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, *4*(1), 395–421. https://doi.org/10.1146/annurev-statistics-060116-054045

Shaby, B. A. (2014). The open-faced sandwich adjustment for MCMC using estimating functions. *Journal of Computational and Graphical Statistics*, *23*(3), 853–876. https://doi.org/10.1080/10618600.2013.842174

Shooter, R., Ross, E., Ribal, A., Young, I. R., & Jonathan, P. (2021). Spatial dependence of extreme seas in the North East Atlantic from satellite altimeter measurements. *Environmetrics*, *32*(4), e2674. https://doi.org/10.1002/env.2674

Shooter, R., Ross, E., Tawn, J., & Jonathan, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics*, *30*(6), e2562. https://doi.org/10.1002/env.2562

Shooter, R., Ross, E., Ribal, A., Young, I. R., & Jonathan, P. (2022). Multivariate spatial conditional extremes for extreme ocean environments. *Ocean Engineering*, *247*, 110647. https://doi.org/10.1016/j.oceaneng.2022.110647

Shooter, R., Tawn, J., Ross, E., & Jonathan, P. (2021). Basin-wide spatial conditional extremes for severe ocean storms. *Extremes*, *24*(2), 241–265. https://doi.org/10.1007/s10687-020-00389-w

Sibuya, M., et al. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, *11*(2), 195–210.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*(1), 1–28. https://doi.org/10.1214/16-STS576

Simpson, E. S., Opitz, T., & Wadsworth, J. L. (2023). High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. *Extremes*. https://doi.org/10.1007/s10687-023-00468-8

Simpson, E. S., & Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, *41*, 100482. https://doi.org/10.1016/j.spasta.2020.100482

Syring, N., & Martin, R. (2018). Calibrating general posterior credible regions. *Biometrika*, *106*(2), 479–486. https://doi.org/10.1093/biomet/asy054

Vandeskog, S. M., Martino, S., Castro-Camilo, D., & Rue, H. (2022). Modelling sub-daily precipitation extremes with the blended generalised extreme value distribution. *Journal of Agricultural, Biological and Environmental Statistics*, *27*(4), 598–621. https://doi.org/10.1007/s13253-022-00500-7

Wadsworth, J. L., & Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, *99*(2), 253–272. https://doi.org/10.1093/biomet/asr080

Wadsworth, J. L., & Tawn, J. A. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, *51*, 100677. https://doi.org/10.1016/j.spasta.2022.100677

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*(1), 1–25.

Unpublished note on the self-inconsistency of the conditional extremes model

Silius M. Vandeskog[1] Sara Martino[1] Raphaël Huser[2]

[1] Norwegian University of Science and Technology (NTNU)
[2] King Abdullah University of Science and Technology (KAUST)

This note uses the same notation as Paper 3.

As discussed by Besag (1974), if one defines a joint distribution by specifying a set of conditional distributions, the resulting joint distribution may become self-inconsistent. This means that joint probabilities may not be well defined, because different ways of integrating over the conditional distributions can yield different answers. Since the conditional extremes distribution is defined through a set of conditional distributions, its joint distribution may therefore become self-inconsistent. This self-inconsistency is examined by Heffernan and Tawn (2004) and Liu and Tawn (2014), who show that it is impossible to achieve self-consistency for finite thresholds t, given extremal independence, i.e., $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) < y_0$, when the distribution of the residual field $Z(\cdot)$ is smooth. The consequence of this self-inconsistency is that probabilities for events where more than one random variable is extreme are ill-defined. As an example, assume that the bivariate vector $(X, Y)^\top$ can be described by the conditional extremes model with threshold t. The probability $\mathrm{P}(X > t, Y > t)$ that both components exceed the threshold can, e.g., be computed as

$$\int_t^\infty \int_t^\infty \pi(x \mid y)\pi(y)\mathrm{d}y\mathrm{d}x, \quad \text{or} \quad \int_t^\infty \int_t^\infty \pi(y \mid x)\pi(x)\mathrm{d}x\mathrm{d}y,$$

where $\pi(\cdot)$ denotes different kinds of probability density functions. Due to the lack of self-consistency between conditional models for $\pi(x \mid y)$ and $\pi(y \mid x)$, these two integration paths might provide different answers.

The lack of self-consistency causes problems when one attempts to simulate from the global conditional extremes distribution, which in the

bivariate case is the joint distribution of $[(X, Y) \mid \max(X, Y) > t]$. In order to sample from this distribution, one must first choose one specific integration path and define it as the "correct one". Then, one must sample in a way that corresponds to this integration path. Consequently, two different simulation algorithms for the global conditional extremes distribution may produce samples with considerably different properties. As an example, for a d-dimensional random vector $\boldsymbol{X}$, the simulation algorithm of Keef et al. (2013) is based on estimating probabilities as

$$\mathrm{P}(\boldsymbol{X} \in \cdot \mid \max(\boldsymbol{X}) > t) = \sum_{i=1}^{d} \pi_i^{(1)} \frac{\mathrm{P}(\boldsymbol{X} \in \cdot, X_i = \max(\boldsymbol{X}) \mid X_i > t)}{\mathrm{P}(X_i = \max(\boldsymbol{X}) \mid X_i > t)}, \quad (2)$$

where $\pi_i^{(1)} = \mathrm{P}(X_i = \max(\boldsymbol{X}) \mid \max(\boldsymbol{X}) > t)$, while the Wadsworth and Tawn (2022) algorithm is based on estimating probabilities as

$$\mathrm{P}(\boldsymbol{X} \in \cdot \mid \max(\boldsymbol{X}) > t) = \frac{\sum_K \sum_{i=1}^{d} \pi_i^{(2)} \mathrm{P}(\boldsymbol{X} \in \cdot \cap R_K \mid X_i > t)/|K|}{\sum_K \sum_{i=1}^{d} \pi_i^{(2)} \mathrm{P}(\boldsymbol{X} \in R_K \mid X_i > t)/|K|}, \quad (3)$$

where $\pi_i^{(2)} = \mathrm{P}(X_i > t)/\sum_{j=1}^{d} \mathrm{P}(X_j > t)$, K is an element in the power set of $\{1, 2, \ldots, d\}$, and

$$R_K = \{\boldsymbol{x} = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d : x_i > t \; \forall \; i \in K, x_j \leq t \; \forall \; j \notin K\}.$$

These two algorithms clearly correspond to two different integration paths, and may therefore provide different results, when combined with different conditional extremes models.

We compute $\mathrm{P}(X > t, Y > t \mid \max(X, Y) > t)$ using the two integration paths (2) and (3), for the symmetrical conditional extremes model where X and Y have exponential marginals and the conditional distribution of $[X \mid Y > t]$ is Gaussian with mean $\mu + \alpha Y$ and variance $\sigma^2 Y^{2\beta}$, and vice versa for $[Y \mid X > t]$. If $\mu = 0$, $\sigma = 1$, $\alpha = 0.9$, $\beta = 0.8$ and $t = 4$, then equation (2) gives a probability of 0.17, whereas equation (3) gives a probability of 0.37. We also estimate the probability using Monte Carlo simulation and find, as expected, that the Keef et al. (2013) algorithm agrees with (2), while the Wadsworth and Tawn (2022) algorithm agrees with (3). This massive difference in computed probabilities could potentially have severe consequences for, e.g., infrastructure built to withstand a certain return level of hourly precipitation. As none of the possible integration paths are neither more nor less correct than the others, it is our opinion that one should be extremely careful when estimating return levels by sampling from the conditional extremes model where the conditioning variable is unknown, i.e., $[\boldsymbol{X} \mid \max(X) > t]$. Instead, we believe that one should prefer to estimate return levels by first choosing a specific conditioning

site/variable, e.g., X_i, and then sampling from the conditional extremes distribution of $[\boldsymbol{X} \mid X_i > t]$, which is both proper and self-consistent.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 192–225. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x

Heffernan, J. E., & Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 497–546. https://doi.org/10.1111/j.1467-9868.2004.02050.x

Keef, C., Tawn, J. A., & Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, *24*(1), 13–21. https://doi.org/10.1002/env.2190

Liu, Y., & Tawn, J. A. (2014). Self-consistent estimation of conditional multivariate extreme value distributions. *Journal of Multivariate Analysis*, *127*, 19–35. https://doi.org/10.1016/j.jmva.2014.02.003

Wadsworth, J. L., & Tawn, J. A. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, *51*, 100677. https://doi.org/10.1016/j.spasta.2022.100677

— 4 —

Fast spatial simulation of extreme high-resolution radar precipitation data using INLA

Silius M. Vandeskog, Sara Martino and Raphaël Huser

In review, but published on arXiv with reference 2307.11390

Fast spatial simulation of extreme high-resolution radar precipitation data using INLA

Silius M. Vandeskog[1] Sara Martino[1] Raphaël Huser[2]

[1] Norwegian University of Science and Technology (NTNU)
[2] King Abdullah University of Science and Technology (KAUST)

Abstract

We develop a methodology for modelling and simulating high-dimensional spatial precipitation extremes, using a combination of the spatial conditional extremes model, latent Gaussian models and integrated nested Laplace approximations (INLA). The spatial conditional extremes model requires data with Laplace marginal distributions, but precipitation distributions contain point masses at zero that complicate necessary standardisation procedures. We propose to model conditional extremes of nonzero precipitation only, while separately modelling precipitation occurrences. The two models are then combined to create a complete model for extreme precipitation. Nonzero precipitation marginals are modelled using a combination of latent Gaussian models with gamma and generalised Pareto likelihoods. Four different models for precipitation occurrence are investigated. New empirical diagnostics and parametric models are developed for describing components of the spatial conditional extremes model. We apply our framework to simulate spatial precipitation extremes over a water catchment in Central Norway, using high-density radar data. Inference on a 6000-dimensional data set is performed within hours, and the simulated data capture the main trends of the observed data well.

Keywords: Extreme precipitation, Spatial conditional extremes, INLA, Computational statistics

1 Introduction

Europe is currently experiencing one of its most flood-intense periods within the last 500 years (Blöschl et al., 2020), and floods are projected to become more frequent and damaging in the future due to ongoing climate changes (Yin et al., 2018; Allan et al., 2020). Thus, flood mitigation has the potential of avoiding numerous fatalities and large economical losses (Jongman, 2018). Flood impacts are often assessed using hydrological impact studies that rely on climate variables such as temperature and precipitation as input, typically provided from interpolated observational data sets or climate projections from general circulation models and regional climate models (Hanssen-Bauer et al., 2015; Giorgi, 2019). However, precipitation is a localised phenomenon with much space-time variability, which the observational data sets and climate projections are unable to capture due to computational constraints and sparsity of observations in space and time (Westra et al., 2014; Lopez-Cantu et al., 2020). The observational data sets may also be too short in time to fully capture the risks and consequences of floods, as the most devastating extreme weather events with high flood risk may simply have not happened yet. Stochastic weather generators that simulate realistic climate data have therefore become an important climate impact assessment method, which allows for better exploration of complex weather phenomena by providing longer time series, or by capturing important small-scale spatio-temporal variability that happens "inside the grids" of too coarse climate projections and interpolated data sets (Ailliot et al., 2015; Maraun & Widmann, 2018).

The spatial distribution of precipitation is important for assessing flood risk, but most stochastic weather generators are purely temporal, and the spatial stochastic weather generators tend to focus on simulating "non-extreme" precipitation. Thus, little focus has been given to the simulation of extreme high-resolution precipitation data in space or space-time. Palacios-Rodríguez et al. (2020) simulate high-resolution spatio-temporal precipitation extremes, but their method is based on resampling transformations of observed extreme events, which makes it impossible to generate events with completely new behaviour. Richards et al. (2022, 2023) develop promising spatial simulations of extreme hourly precipitation, but their method is based on an inefficient inference scheme that becomes troublesome for higher-dimensional problems. In this paper we develop a framework for high-dimensional spatial modelling and simulation of extreme precipitation, which we apply to a data set of high-resolution hourly precipitation data from a weather radar in Norway.

Weather radars observe precipitation by sending out radio signals and measuring how much of the signal is reflected back. This makes it possible to create high-resolution spatio-temporal precipitation data sets. The data sets do not capture marginal distributions as well as, e.g., rain gauge data, but

they provide reliable descriptions of spatio-temporal dependence (Bournas & Baltas, 2022). To the best of our knowledge, radar data are currently among the best available products for capturing the small-scale spatio-temporal dependence structure of precipitation. Yet, not many have taken advantage of this, and we are not aware of any previous attempts in the literature of spatial or spatio-temporal modelling of precipitation extremes based on radar data. We believe that radar data have been under-used in the literature, and in this paper we attempt to demonstrate the potential of radar data by using them for producing high-resolution simulations of extreme hourly precipitation.

Our method builds upon extreme value theory (Davison & Huser, 2015), which has shown great success at modelling and assessing environmental risks such as extreme temperature (Castro-Camilo et al., 2021; Simpson & Wadsworth, 2021), precipitation (Huser & Davison, 2014; Opitz et al., 2018; Richards et al., 2022) and wind (Castro-Camilo et al., 2019). An important part of extreme value theory is the modelling of extremal dependence, often described by conditional exceedance probabilities. Given a spatial random field, $X(\boldsymbol{s})$ with $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2$, we define the conditional exceedance probability

$$\chi_p(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathrm{P}(X(\boldsymbol{s}_1) > F_{\boldsymbol{s}_1}^{-1}(p) \mid X(\boldsymbol{s}_2) > F_{\boldsymbol{s}_2}^{-1}(p)),$$

where $F_{\boldsymbol{s}}^{-1}$ is the quantile function of $X(\boldsymbol{s})$, and the tail correlation coefficient $\chi(\boldsymbol{s}_1, \boldsymbol{s}_2) = \lim_{p \to 1} \chi_p(\boldsymbol{s}_1, \boldsymbol{s}_2)$. Two random variables, $X(\boldsymbol{s}_1)$ and $X(\boldsymbol{s}_2)$, are called asymptotically dependent if $\chi(\boldsymbol{s}_1, \boldsymbol{s}_2) > 0$, and asymptotically independent otherwise. Classical models for spatial extremes are based on max-stable processes (Davison et al., 2019), which focus on modelling pointwise maxima and assume that χ is positive while χ_p is nearly constant with level $p \approx 1$. However, experience has shown that environmental data often exhibit weakening dependence as events become more extreme (Huser et al., 2021), i.e., χ_p continuously decreases as $p \to 1$. Thus, alternative models that focus on capturing the so-called subasymptotic dependence structure explained by χ_p are crucial for correctly assessing the risks of spatial extremes in environmental data. A spatial process used for modelling climate data should also be able to exhibit both asymptotic dependence at short distances and asymptotic independence at large distances, but most classical extreme models are unable to describe nontrivial changes in the asymptotic dependence class as a function of distance (Huser & Wadsworth, 2022). This has led to a surge of new models for spatial extremes with more flexible subasymptotic and asymptotic dependence structures, including inverted max-stable processes and max-mixture models, (Wadsworth & Tawn, 2012), max-infinitely divisible processes (Huser et al., 2021), scale-mixture models (Huser et al., 2017; Engelke et al., 2019; Huser & Wadsworth,

2019), kernel convolution models (Krupskii & Huser, 2022) and the spatial conditional extremes model (Wadsworth & Tawn, 2022).

Statistical modelling of spatial dependence often leads to computationally demanding inference. This is particularly true for spatial extreme value models, where a majority of the most popular models have to rely on low-dimensional composite likelihood methods for achieving computationally tractable inference (Padoan et al., 2010; Castruccio et al., 2016). The Gaussian random field is popular in traditional spatial and spatio-temporal statistics, as it has nice theoretical properties while allowing for fast and realistic modelling of complex processes (Gelfand et al., 2010). In particular, the latent Gaussian modelling framework has shown great success within a large range of applications (Banerjee et al., 2014), by yielding flexible and realistic models that utilise assumptions of Gaussianity and conditional independence for performing fast inference using integrated nested Laplace approximations (INLA; Rue et al., 2009). Yet, latent Gaussian models have not achieved similar success for modelling spatial extremes, as their dependence structures are unsuitable for most classical spatial extreme value models (Davison et al., 2012). However, Gaussian dependence structures are becoming more suitable for some newer breeds of spatial extreme value models, such as the spatial conditional extremes model (Wadsworth & Tawn, 2022). Indeed, the model only requires a few minor alterations to become a latent Gaussian model, which makes it possible to perform fast high-dimensional inference with INLA (Simpson et al., 2023; Vandeskog, Martino, & Huser, 2022). In this paper, we build upon the work of Vandeskog, Martino, and Huser (2022) and we develop new empirical diagnostics and parametric models for describing components of the spatial conditional extremes model, as well as improved models for the marginal distributions and a new methodology for describing precipitation zeros.

The spatial conditional extremes model describes the distribution of a spatial random field $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}\subset\mathbb{R}^2}$, with Laplace marginal distributions, given that it exceeds a large threshold τ at some preselected location $\boldsymbol{s}_0 \in \mathcal{S}$. The model states that, for τ large enough, the process $[Y(\boldsymbol{s}) \mid Y(\boldsymbol{s}_0) = y_0]$, with $y_0 > \tau$, is approximately equal in distribution to a spatial random field that only depends on y_0 through a location parameter $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and a scale parameter $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) > 0$. An important part of the modelling process is therefore to choose a suitable class of functions for $a(\cdot)$ and $b(\cdot)$, and to choose a threshold τ that is high enough to yield little model bias, but also small enough to efficiently utilise the data. To the best of our knowledge, this threshold selection problem has not yet attracted much focus in the literature. In this paper we develop new empirical diagnostics for finding reasonable values of the threshold τ and the forms of $a(\cdot)$ and $b(\cdot)$, and we propose a new class of parametric functions for $a(\cdot)$ and $b(\cdot)$ that can provide suitable fits to data at much lower thresholds than used previously

(e.g., Vandeskog, Martino, & Huser, 2022), thus allowing us to utilise more of the data for more efficient inference, without too much model bias.

To fit the spatial conditional extremes model, one must first standardise the data to have Laplace marginal distributions. However, the marginal distributions of hourly precipitation contain a point mass at zero, which makes it impossible to directly transform them to the Laplace scale using the probability integral transform. Richards et al. (2022, 2023) solve this problem by censoring all zeros, but this leads to less efficient inference techniques such as low-dimensional composite likelihoods, and it cannot easily be combined with the INLA framework. Inspired by the so-called Richardson-type stochastic weather generators (Richardson, 1981), we instead propose to model the conditional extremes of nonzero precipitation intensity, while separately modelling the distribution of precipitation occurrences in space. We then combine the two models to describe the full distribution of spatial conditional precipitation extremes.

To transform nonzero precipitation data onto the Laplace scale, we must first estimate their marginal distributions in space and time. In the spatial conditional extremes literature, this is commonly achieved using the empirical distribution functions at each location, possibly combined with a generalised Pareto (GP) distribution for describing the upper tails (Simpson & Wadsworth, 2021; Richards et al., 2022; Wadsworth & Tawn, 2022; Shooter et al., 2022). However, empirical distribution functions can be unsuitable if the marginal distributions of the data vary in space and time, which is often the case for precipitation and other climate variables. Since both the total amount and the spatial distribution of precipitation are important properties for assessing flood risk, we here focus equally much on describing properties of the marginal and the spatial precipitation distribution. Therefore, following Opitz et al. (2018) and Castro-Camilo et al. (2019), we apply a complex spatio-temporal model based on two different latent Gaussian models for describing the marginal distributions. The first model describes the bulk of the data using a gamma likelihood, while the second model describes the upper tails using a GP likelihood.

To the best of our knowledge, there have not been any previous attempts to model precipitation occurrences in space given that extreme precipitation has been observed at a chosen conditioning site. We here propose multiple competing models for describing conditional precipitation occurrences. The probit model is a common regression model for describing binary data (Fahrmeir et al., 2013; Verdin et al., 2015), and we propose to model precipitation occurrences using both the standard probit model and a spatial version of it. We show that both probit models are latent Gaussian models and perform fast inference for them using INLA. However, our probit models produce occurrence processes that are independent of the precipitation intensity process, which is unrealistic. The probit model also

struggles to capture some other important spatio-temporal properties of smooth high-resolution precipitation data. Thus, we propose an additional third model, denoted the threshold model, which is designed to capture the dependence on the precipitation intensity process and to better capture the spatial smoothness properties of precipitation occurrences. For better baseline comparisons, we also propose an occurrence model in which "no precipitation" is interpreted as a tiny but positive amount of precipitation, or in other words, that it always rains.

To sum up, in this paper we develop a framework for modelling and simulating extreme precipitation in space, based on latent Gaussian models and the spatial conditional extremes model. This is applied for simulating precipitation extremes using a high-resolution data set of hourly precipitation from a weather radar in Norway. We separately model precipitation occurrences and intensities to avoid problems with the point mass at zero precipitation. The spatial distribution of precipitation occurrences is described using four competing models, while the marginal distributions of nonzero precipitation are modelled by merging two latent Gaussian models, with a gamma likelihood and a GP likelihood, respectively. We employ a latent Gaussian model version of the spatial conditional extremes model for describing the extremal dependence of the nonzero precipitation. We also develop new empirical diagnostics for choosing model components of the spatial conditional extremes model, and we use these for proposing new parametric functions for the model components, which allow for better data utilisation through a lower threshold. The remainder of the paper is organised as follows: The weather radar data are presented in Section 2. Then, Section 3 describes our general framework for modelling spatial extreme precipitation, and in Section 4, we apply our framework for modelling and simulating extreme precipitation using the chosen radar data. The paper concludes with a final discussion in Section 5.

2 Data

The Rissa radar is located in the Fosen region in Central Norway. It scans the surrounding area multiple times each hour by sending out radio waves and measuring how much is reflected back. The Norwegian Meteorological Institute processes the observed reflectivity data and uses them to create gridded 1×1 km^2 resolution maps of estimated hourly precipitation, measured in mm/h (Elo, 2012). These precipitation maps are freely available, dating back to 1 January 2010, from an online weather data archive (https://thredds.met.no).

We use the radar precipitation maps for modelling and simulating extreme hourly precipitation over the Stordalselva catchment, located close

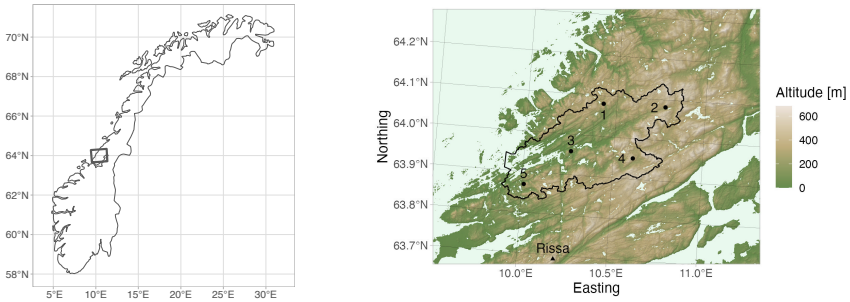


Figure 1: *Left plot: Map of Norway with a square defining the spatial domain $\mathcal{S}$. Right plot: Elevation map of $\mathcal{S}$. The Rissa radar is displayed at the bottom of the map, and the five conditioning sites used in Section 4 are enumerated and marked by black circles. The Stordalselva catchment is displayed using a black polygon.*

to the Rissa radar. To achieve this, we download all data for 2010–2022 inside a spatial domain $\mathcal{S}$ of size $91 \times 71 = 6461$ km^2, centred around the Stordalselva catchment. Figure 1 displays the domain $\mathcal{S}$ and the locations of the Rissa radar and the Stordalselva catchment. The spatial conditional extremes model allows one to model and simulate extremes occurring at any site of interest, by conditioning on that site experiencing extreme behaviour (see Section 3 for more details). For the sake of illustration, we choose five such conditioning sites, somewhat equally spaced throughout the water catchment, for modelling and simulating extreme precipitation in Section 4. These sites are also displayed in Figure 1. There are considerable differences between extreme summer precipitation and extreme winter precipitation in Norway (Dyrrdal et al., 2015), and we therefore choose to only model summer precipitation from June, July and August, which is when most of the intense precipitation events occur in Norway. There are some distortions in the data close to the Rissa radar, so we remove all observations from locations that are within 5 km from the radar.

It can be difficult to distinguish between little and no precipitation using reflectivity data, and the estimated precipitation data contain both exact zeros and values with magnitudes as small as 10^{-5} mm/h. In the Supplementary Material, we show that there are large differences in the proportions of exact zeros for different times, due to an upgrade of the weather radar in 2018, but that the proportion of observations smaller than 0.1 mm/h is approximately constant in time. We therefore round every observation smaller than 0.1 mm/h down to zero.

3 Model framework

3.1 Model overview

We model the spatial extremal dependence structure of the hourly precipitation process, $X_t(\boldsymbol{s})$, at location $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2$ and time $t \in \mathcal{T} \subset \mathbb{N}$, using the spatial conditional extremes model (Wadsworth & Tawn, 2022). The model describes the conditional distribution of $X_t(\boldsymbol{s})$ given that $X_t(\boldsymbol{s}_0) > \tau_t(\boldsymbol{s}_0)$, where $\boldsymbol{s}_0$ is some chosen conditioning site and $\tau_t(\boldsymbol{s})$ is a large threshold that may vary in space and time. The conditioning site can be located anywhere in $\mathcal{S}$, which makes us free to place $\boldsymbol{s}_0$ at a specific location of interest, for modelling and simulating only the extremes that we care about.

To use the spatial conditional extremes model, we must first transform $X_t(\boldsymbol{s})$ into a standardised process $\tilde{X}_t(\boldsymbol{s})$, with Laplace marginal distributions at all locations and time points in $\mathcal{S} \times \mathcal{T}$, using the probability integral transform $\tilde{X}_t(\boldsymbol{s}) = F^{-1}[F_{\boldsymbol{s},t}(X_t(\boldsymbol{s}))]$, where F^{-1} is the quantile function of the Laplace distribution and $F_{\boldsymbol{s},t}$ is the cumulative distribution function of $X_t(\boldsymbol{s})$ (Keef et al., 2013). However, the marginal distribution of hourly precipitation contains a point mass at zero, which means that the marginal distribution of $\tilde{X}_t(\boldsymbol{s})$ also contains a point mass, making it different from the Laplace distribution in its lower tail. Richards et al. (2022, 2023) tackle this problem by left-censoring all zeros. This yields promising results, but the censoring makes high-dimensional inference computationally intractable without the use of low-dimensional composite likelihoods. Moreover, this approach is still computationally demanding as it relies on evaluating bivariate Gaussian distributions many times. We therefore propose another approach for modelling $X_t(\boldsymbol{s})$ with the spatial conditional extremes model. Assume that hourly precipitation can be represented as $X_t(\boldsymbol{s}) = X_t^+(\boldsymbol{s})I_t(\boldsymbol{s})$, where $X_t^+(\boldsymbol{s}) = [X_t(\boldsymbol{s}) \mid X_t(\boldsymbol{s}) > 0]$ represents precipitation intensity and $I_t(\boldsymbol{s})$ is a binary random process that equals 1 when $X_t(\boldsymbol{s}) > 0$ and 0 when $X_t(\boldsymbol{s}) = 0$. So-called Richardson-type stochastic weather generators depend on this formulation by first simulating the precipitation occurrence $I_t(\boldsymbol{s})$ and then simulating the precipitation intensity $X_t^+(\boldsymbol{s})$ if $I_t(\boldsymbol{s}) = 1$ (Richardson, 1981). We build upon this approach by, instead of modelling $[X_t(\boldsymbol{s}) \mid X_t(\boldsymbol{s}) > \tau_t(\boldsymbol{s})]$, performing separate modelling of the conditional intensity process $[X_t^+(\boldsymbol{s}) \mid X_t(\boldsymbol{s}) > \tau_t(\boldsymbol{s})]$ and the conditional occurrence process $[I_t(\boldsymbol{s}) \mid X_t(\boldsymbol{s}) > \tau_t(\boldsymbol{s})]$, and then setting

$$[X_t(\boldsymbol{s}) \mid X_t(\boldsymbol{s}) > \tau_t(\boldsymbol{s})] = \left[X_t^+(\boldsymbol{s})I_t(\boldsymbol{s}) \mid X_t(\boldsymbol{s}) > \tau_t(\boldsymbol{s})\right]. \tag{1}$$

The marginal distribution $F_{\boldsymbol{s},t}^+$, of $X_t^+(\boldsymbol{s})$, does not contain a point mass, so it can more easily be transformed into the Laplace distribution. Thus, we describe the conditional intensity process with the spatial conditional extremes model, while the conditional occurrence process is described with a

suitable binary model. Our model for $F^+_{\boldsymbol{s},t}$ is described in Section 3.3. Then, our model for the conditional intensity process is described is Section 3.4, and our model for the conditional occurrence process is described in Section 3.5. Most of our models fall within the framework of latent Gaussian models, which are introduced in Section 3.2.

3.2 Latent Gaussian Models

A latent Gaussian model assumes that the observations $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ are conditionally independent given a latent Gaussian random field $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^\top$ and a set of hyperparameters $\boldsymbol{\theta}_1$, namely

$$[\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}_1] \sim \prod_{i=1}^{n} \pi(y_i \mid \eta_i(\boldsymbol{x}), \boldsymbol{\theta}_1), \qquad [\boldsymbol{x} \mid \boldsymbol{\theta}_2] \sim \mathcal{N}(\mu(\boldsymbol{\theta}_2), \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2)),$$

where the likelihood $\prod_{i=1}^{n} \pi(y_i \mid \eta_i(\boldsymbol{x}), \boldsymbol{\theta}_1)$ is a parametric distribution with parameters $\eta_i(\boldsymbol{x})$ and $\boldsymbol{\theta}_1$, the linear predictor $\eta_i(\boldsymbol{x})$ is a linear combination of the elements in $\boldsymbol{x}$ and the latent field $\boldsymbol{x}$ is conditionally Gaussian with mean vector $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q}$, given the hyperparameters $\boldsymbol{\theta}_2$. The prior distributions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are $\pi(\boldsymbol{\theta}_1)$ and $\pi(\boldsymbol{\theta}_2)$, respectively.

The latent Gaussian modelling framework is highly flexible, as the likelihood can stem from an essentially arbitrary parametric distribution, while information from explanatory variables and a large variety of dependency structures can be incorporated into the linear predictor $\eta_i(\boldsymbol{x})$. Additionally, non-Gaussian structures can be incorporated into the model through the likelihood and the hyperparameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, which can be given any kind of prior distributions. Another advantage of the latent Gaussian model framework is that it allows for fast approximate inference using INLA, which is implemented in the `R-INLA` package (van Niekerk et al., 2021, 2023). The package contains a large range of pre-implemented model components for the linear predictor, including splines, AR-models, random walk models and the so-called stochastic partial differential equation (SPDE) model of Lindgren et al. (2011), which produces sparse approximations of Gaussian random fields with Matérn autocorrelation function

$$\gamma(d) = \frac{1}{2^{\nu-1}\Gamma(\nu)}(\kappa d)^\nu K_\nu(\kappa d), \tag{2}$$

where d is the distance between two locations, $\nu > 0$ is the smoothness parameter, $\rho = \sqrt{8\nu}/\kappa$ is the range parameter and K_ν is the modified Bessel function of the second kind and order ν. Thus, `R-INLA` and the latent Gaussian model framework make it easy to quickly develop and perform inference with complex models for a large variety of applications.

3.3 Modelling the marginals

We model marginal distributions of the intensity process $X_t^+(\boldsymbol{s})$ by first modelling the bulk of the data with the gamma distribution, and then the upper tail with the GP distribution. Specifically, we model margins as

$$F_{\boldsymbol{s},t}^+(x) = \begin{cases} G_{\boldsymbol{s},t}(x) & x \leqslant u_t(\boldsymbol{s}), \\ G_{\boldsymbol{s},t}(u_t(\boldsymbol{s})) + (1 - G_{\boldsymbol{s},t}(u_t(\boldsymbol{s})))H_{\boldsymbol{s},t}(x - u_t(\boldsymbol{s})) & x > u_t(\boldsymbol{s}), \end{cases} \tag{3}$$

where $G_{\boldsymbol{s},t}$ and $H_{\boldsymbol{s},t}$ are cumulative distribution functions of the gamma and GP distributions respectively, both with parameters that might vary in space and time, while $u_t(\boldsymbol{s})$ is the p_u-quantile of $G_{\boldsymbol{s},t}$, for some large probability p_u. This "split-modelling" approach is a common choice for modelling precipitation when aiming to describe both the bulk and the upper tail of the distribution (e.g., Opitz et al., 2018). We use the gamma and GP parametrisations of Castro-Camilo et al. (2019), which give a probability density function for the gamma distribution on the form

$$g(x) = \frac{x^{\kappa-1}}{\Gamma(\kappa)} \left(\frac{G^{-1}(\alpha;\kappa,1)}{\psi_\alpha} \right)^\kappa \exp\left(\frac{G^{-1}(\alpha;\kappa,1)}{\psi_\alpha} \right), \quad x > 0, \quad \kappa, \psi_\alpha > 0$$

where κ is the standard shape parameter, α is a fixed probability, $G^{-1}(\alpha;\kappa,1)$ is the α-quantile of a gamma distribution with shape κ and scale 1, and ψ_α is the α-quantile of the gamma distribution. Using this parametrisation for the likelihood of our latent Gaussian model lets us directly estimate the α-quantile of the data through the parameter ψ_α. The GP distribution has cumulative distribution function

$$H(x) = \begin{cases} 1 - \left(1 + \left\{(1-\beta)^{-\xi} - 1\right\} \frac{x}{\phi_\beta}\right)_+^{-1/\xi}, & \xi \neq 0, \\ 1 - (1-\beta)^{x/\phi_\beta}, & \xi = 0, \end{cases} \qquad \phi_\beta > 0, \xi \in \mathbb{R},$$

with $(a)_+ = \max(0, a)$, where ξ is the tail parameter of the GP distribution, β is a fixed probability and the parameter ϕ_β equals the β-quantile of the GP distribution. The support of the GP distribution is $(0, \infty)$ for $\xi \geqslant 0$, while it is $\left(0, \phi_\beta \left\{1 - (1-\beta)^{-\xi}\right\}\right)$ for $\xi < 0$.

We estimate the parameters of $G_{\boldsymbol{s},t}$ and $H_{\boldsymbol{s},t}$ separately, with two different latent Gaussian models. For the parameters of $G_{\boldsymbol{s},t}$, we use a latent Gaussian model with a gamma likelihood, where the shape parameter κ is a hyperparameter that is constant in space and time, while ψ_α is allowed to vary in space and time through the linear predictor $\eta = \log \psi_\alpha$. Setting $\alpha = p_u$ lets us directly estimate the threshold $u_t(\boldsymbol{s})$, while estimating the parameters of $G_{\boldsymbol{s},t}$. The value of p_u, the components of η and the priors for $\boldsymbol{\theta}$ vary depending on the application, and are therefore described in

Section 4.1. After having performed inference with `R-INLA`, we estimate the parameters of $G_{\boldsymbol{s},t}$ as the posteriors means of κ and ψ_α.

Having estimated the threshold $u_t(\boldsymbol{s})$, we then model the distribution $H_{\boldsymbol{s},t}$ of the threshold exceedances $[X_t^+(\boldsymbol{s}) - u_t(\boldsymbol{s}) \mid X_t^+(\boldsymbol{s}) > u_t(\boldsymbol{s})]$ with the GP distribution. Here, we apply a latent Gaussian model with a GP likelihood, where the tail parameter ξ is a hyperparameter that is constant in space and time, while the linear predictor is $\eta = \log \phi_\beta$, where we set $\beta = 0.5$, so that ϕ_β is the GP median. Once more, the parameters of $H_{\boldsymbol{s},t}$ are estimated as the posterior means of ξ and ϕ_β. Note that the GP likelihood within `R-INLA` only allows for modelling $\xi > 0$. However, this should not be too problematic when modelling hourly precipitation data, as there is considerable evidence in the literature that precipitation is heavy-tailed, and thus it should be modelled with a non-negative tail parameter, especially for short temporal aggregation times (Cooley et al., 2007; Van de Vyver, 2012; Papalexiou & Koutsoyiannis, 2013; Huser & Davison, 2014).

3.4 Spatial modelling of the conditional intensity process

We transform the precipitation intensities $X_t^+(\boldsymbol{s})$ into the standardised process $Y_t(\boldsymbol{s})$, with Laplace marginal distributions, using the probability integral transform, $Y_t(\boldsymbol{s}) = F^{-1}[F^+_{\boldsymbol{s},t}(X_t^+(\boldsymbol{s}))]$. Given a conditioning site $\boldsymbol{s}_0$ and a threshold $\tau_t(\boldsymbol{s}_0)$, we then model the spatial distribution of $Y_t(\boldsymbol{s})$ given that $X_t(\boldsymbol{s}_0) > \tau_t(\boldsymbol{s}_0)$, which is the same conditioning event as $Y_t(\boldsymbol{s}_0) > F^{-1}[F^+_{\boldsymbol{s}_0,t}(\tau_t(\boldsymbol{s}_0))]$. We assume that the notion of "extremes" can vary across time and space on the original precipitation scale, but not on the transformed Laplace scale. We therefore set $\tau_t(\boldsymbol{s}_0)$ equal to a chosen quantile of $F^+_{\boldsymbol{s}_0,t}$, which gives the constant threshold $\tau = F^{-1}[F^+_{\boldsymbol{s}_0,t}(\tau_t(\boldsymbol{s}_0))]$ on the Laplace scale. The spatial conditional extremes model of Simpson et al. (2023) now states that, for τ large enough, the conditional process $[Y_t(\boldsymbol{s}) \mid Y_t(\boldsymbol{s}_0) = y_0 > \tau]$ is Gaussian,

$$[Y_t(\boldsymbol{s}) \mid Y_t(\boldsymbol{s}_0) = y_0 > \tau] \overset{d}{=} a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) + b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) Z_t(\boldsymbol{s}; \boldsymbol{s}_0) + \varepsilon_t(\boldsymbol{s}; \boldsymbol{s}_0), \quad (4)$$

where $a(\cdot)$ and $b(\cdot)$ are two standardising functions, $Z_t(\boldsymbol{s}; \boldsymbol{s}_0)$ is a spatial Gaussian random field with $Z_t(\boldsymbol{s}_0; \boldsymbol{s}_0) = 0$ almost surely, and $\varepsilon_t(\boldsymbol{s}; \boldsymbol{s}_0)$ is Gaussian white noise with $\varepsilon_t(\boldsymbol{s}_0; \boldsymbol{s}_0) = 0$ almost surely. This is the same as a latent Gaussian model with a Gaussian likelihood and latent field $a(\cdot) + b(\cdot)Z(\cdot)$, which means that computationally efficient approximate inference can be performed using INLA. Simpson et al. (2023) demonstrate how to perform efficient high-dimensional inference by using `R-INLA` and modelling $Z_t(\boldsymbol{s}; \boldsymbol{s}_0)$ with the SPDE approximation. Vandeskog, Martino, and Huser (2022) build upon their work and develop a methodology for implementing computationally efficient parametric models for $a(\boldsymbol{s}; \boldsymbol{s}_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0)$ in `R-INLA` and a method for efficient constraining of $Z_t(\boldsymbol{s}; \boldsymbol{s}_0)$ such

that $Z_t(\boldsymbol{s}_0; \boldsymbol{s}_0) = 0$ almost surely. We apply their methodology for modelling the spatial distribution of conditional precipitation extremes, while developing new diagnostics and models for the standardising functions $a(\cdot)$ and $b(\cdot)$.

3.5 Spatial modelling of the conditional occurrence process

Four competing models are applied to describe the spatial distribution of conditional precipitation occurrences, $[I_t(\boldsymbol{s}) \mid X_t(\boldsymbol{s}_0) > \tau_t(\boldsymbol{s}_0)]$. One of these is the relatively common spatial probit model, which assumes that $I_t(\boldsymbol{s})$ depends on an underlying latent Gaussian process $Z_t(\boldsymbol{s})$ such that $I_t(\boldsymbol{s}) = 1$ when $Z_t(\boldsymbol{s}) + \varepsilon_t(\boldsymbol{s}) \geqslant 0$ and $I_t(\boldsymbol{s}) = 0$ otherwise, where $\varepsilon_t(\boldsymbol{s})$ is zero-mean Gaussian white noise. Thus, the process $I_t(\boldsymbol{s})$ is conditionally independent given $Z_t(\boldsymbol{s})$, with success probability

$$\mathrm{P}\left[I_t(\boldsymbol{s}) = 1 \mid Z_t(\boldsymbol{s})\right] = \Phi(Z_t(\boldsymbol{s})/\sigma),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and σ^2 is the variance of $\varepsilon_t(\boldsymbol{s})$. This means that the probit model is in fact a latent Gaussian model with a Bernoulli likelihood, and that we can perform fast inference using INLA. Within `R-INLA`, we decompose $Z_t(\boldsymbol{s})$ into $Z_t(\boldsymbol{s}) = \mu(\boldsymbol{s}) + Z_t^*(\boldsymbol{s})$, where $Z_t^*(\boldsymbol{s})$ is a zero-mean Gaussian random field, and $\mu(\boldsymbol{s})$ describes the mean of $Z_t(\boldsymbol{s})$. For faster inference, we model $Z_t^*(\boldsymbol{s})$ with the SPDE approximation. Assuming stationarity, we enforce $I_t(\boldsymbol{s}_0) = 1$ by modelling $\mu(\boldsymbol{s})$ as a function of the distance d between $\boldsymbol{s}$ and $\boldsymbol{s}_0$, i.e., $\mu(d) \equiv \mu(\boldsymbol{s})$ with $d = \|\boldsymbol{s} - \boldsymbol{s}_0\|$. Then we ensure that $\mu(0)$ is positive and large, while also enforcing that $Z_t^*(\boldsymbol{s}_0) = 0$ almost surely, using the constraining method of Vandeskog, Martino, and Huser (2022). This does not guarantee that $\mathrm{P}(I_t(\boldsymbol{s}_0) = 1) = 1$ exactly, but if $\mu(0)$ is large enough, then $\mathrm{P}(I_t(\boldsymbol{s}_0) = 1) \approx 1$ for most practical purposes. The exact structure of $\mu(\boldsymbol{s})$ varies depending on the application in question.

The spatial probit model can produce realistic realisations of the spatial binary process, but it can also struggle in situations where the binary field is smooth, in the sense that the variance of $\varepsilon_t(\boldsymbol{s})$ is considerably smaller than the variance of $Z_t(\boldsymbol{s})$. To ensure smooth model realisations, the variance of $Z_t(\boldsymbol{s})/\sigma$ must become so large that the probability $\Phi(Z_t(\boldsymbol{s})/\sigma)$ always is close to either 0 or 1, and this large variance can make it difficult to reliably estimate trends in the mean $\mu(\boldsymbol{s})$. For this reason, we also attempt to model the conditional occurrence process using a probit model without any spatial effects, i.e., where we remove $Z_t^*(\boldsymbol{s})$. This model typically fails at providing realistic-looking realisations of smooth binary processes, but it can perform considerably better at capturing trends in the mean structure.

Our probit models are independent of the conditional intensity model, so it is possible for the simulated occurrence samples to create highly non-

smooth precipitation realisations where areas with large precipitation values suddenly contain a "hole" of zeros close to the most extreme observations. This is an unrealistic behaviour that we wish to avoid. Our third modelling strategy is therefore based on the assumption that the occurrence process is dependent upon the intensity process such that only the smallest values of $X_t^+(\boldsymbol{s})$ gets turned into zeros. Thus, the third approach, denoted the threshold model, estimates the overall probability p of observing zeros in the data, and then set $I_t(\boldsymbol{s})$ equal to zero whenever $X_t^*(\boldsymbol{s})$ is smaller than its estimated p-quantile. Lastly, for improved base-line comparisons, we add a fourth occurrence model, which interprets "no precipitation" as a tiny but positive amount of precipitation, i.e., $I_t(\boldsymbol{s}) \equiv 1$. We denote this the nonzero model.

4 Simulating extreme hourly precipitation

We apply the models from Section 3 to the data from Section 2 for modelling and simulating spatial realisations of extreme hourly precipitation. In Section 4.1 we model and standardise the marginal distributions of the data. Then, in Section 4.2 and Section 4.3, we model the conditional intensities and occurrences of extreme precipitation, respectively. Finally, in Section 4.4, we combine all the model fits for simulating spatial realisations of extreme hourly precipitation.

4.1 Modelling the marginals

We model the marginal distribution $F_{\boldsymbol{s},t}^+$ of nonzero precipitation using the gamma-GP split-model from Section 3.3, where we choose $\alpha = p_u = 0.95$. We attempt to model the linear predictor using a separable space-time model where the spatial effect is modelled with a spatial Gaussian random field, described using the SPDE approximation, and the temporal effect is modelled with a Gaussian smoothing spline, also described using the SPDE approximation. However, we find that the spatial effect is estimated to be almost constant, and that a purely temporal model using only the Gaussian smoothing spline performs better. We therefore decide to use the purely temporal model, where the linear predictor is equal to a temporal Gaussian smoothing spline. A model based on splines is unsuitable for prediction outside the observed spatio-temporal domain, but the aim of this paper is modelling, and not forecasting, so we find it to be a good model choice.

We place the weakly informative penalised complexity (PC) prior (Simpson et al., 2017) of Fuglstad et al. (2019) on the range $\rho = \sqrt{8\nu}/\kappa$ and variance σ^2 such that the prior probability that ρ exceeds 28 days is 5% and the prior probability that σ exceeds a value of 3 is 5%. The smoothness parameter ν can be difficult to estimate (Lindgren & Rue, 2015), so we fix

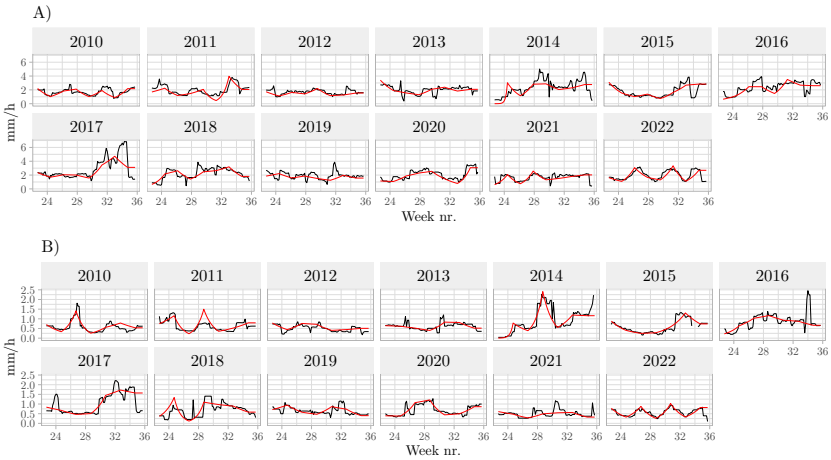


Figure 2: *Estimators for the* 95%*-quantile of hourly precipitation intensity (A) and the median of precipitation threshold exceedances (B) for each summer day during 2010–2022. In black: empirical estimators created using a sliding window of width of one week. In red: posterior means estimated using the latent Gaussian models with a gamma likelihood (A) and a GP likelihood (B).*

it to $\nu = 1.5$. Due to the large amounts of data, we speed up inference by only using observations from a spatial subgrid of the data with resolution 2×2 km^2.

Inference is performed with `R-INLA`, in approximately half an hour, when using only one core on a 2.6 gHz Linux server. The posterior mean of the shape parameter κ equals 0.69, and the posterior mean of the threshold $u_t(\boldsymbol{s})$ is displayed in Subplot A of Figure 2, together with the empirical 95%-quantiles of the precipitation intensities, pooled across space. The smoothing spline seems to capture the temporal trends of the data well. We evaluate the model fit using quantile-quantile (QQ) plots. These are displayed in the Supplementary Material, and they show an almost perfect correspondence between model quantiles and empirical quantiles. We conclude that the model fit is satisfactory.

Having estimated $u_t(\boldsymbol{s})$, we then model the threshold exceedances $[X_t^+(\boldsymbol{s}) - u_t(\boldsymbol{s}) \mid X_t^+(\boldsymbol{s}) > u_t(\boldsymbol{s})]$ with the GP distribution, as described in Section 3.3. Once more, we find that a purely temporal model for the linear predictor performs better than a separable space-time model. We therefore model the linear predictor using a similar spline model as in the gamma model, with the same prior distributions. The tail parameter ξ is given the PC prior of Opitz et al. (2018),

$$\pi(\xi) = \lambda(1 - \xi/2)(1 - \xi)^{-3/2}2^{-1/2} \exp\left(-\lambda\xi/\sqrt{2(1 - \xi)}\right), \quad 0 \leqslant \xi < 1,$$

with penalty parameter λ. The GP distribution has infinite mean for

$\xi \geqslant 1$ and infinite variance for $\xi \geqslant 1/2$, and since it is well established in the literature that ξ tends to be in the range between 0.05 and 0.3 for precipitation data (e.g., Cooley et al., 2007; Van de Vyver, 2012; Papalexiou & Koutsoyiannis, 2013), we enforce $\xi \leqslant 1/2$ to ease parameter estimation. Then, we choose $\lambda = 7$, which gives the prior probability $\mathrm{P}(\xi \leqslant 0.4) \approx 95\%$.

Inference is performed with `R-INLA`, using all the spatial locations, in approximately 2 minutes. The posterior mean of ξ is 0.145, which is far away from the upper bound of 1/2, and corresponds well with the results of Vandeskog, Martino, Castro-Camilo, and Rue (2022), who estimated $\xi = 0.18$ with a 95% credible interval of $(0.14, 0.21)$ when modelling the yearly maxima of hourly precipitation using rain gauge data from a spatial domain that covers $\mathcal{S}$. Subplot B of Figure 2 displays the empirical median of the threshold exceedances, pooled in space, along with the posterior means of the threshold exceedance medians, which seem to agree well with the main temporal trends of the data. We evaluate the model fit using QQ plots, displayed in the Supplementary Material. These demonstrate a good correspondence between model quantiles and empirical quantiles. We once more conclude that our model provides a satisfactory fit to the data.

4.2 Modelling the conditional intensity process

We standardise the precipitation intensities to have Laplace marginal distributions. Then, following Keef et al. (2013), we choose the functions

$$a(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = \alpha(\boldsymbol{s}; \boldsymbol{s}_0) y_0, \qquad b(\boldsymbol{s}; \boldsymbol{s}_0, y_0) = y_0^{\beta(\boldsymbol{s}; \boldsymbol{s}_0)}$$

for the spatial conditional extremes model (4), which, they claim, can cover a large range of dependence structures, including all the standard copulas studied by Joe (1997) and Nelsen (2006). Building upon the work of Vandeskog, Martino, and Huser (2022), we develop new empirical diagnostics for making informed decisions about the value of the threshold τ and the forms of $\alpha(\boldsymbol{s}; \boldsymbol{s}_0)$ and $\beta(\boldsymbol{s}; \boldsymbol{s}_0)$.

We assume that the standardising functions only depend on the Euclidean distance to s_0 and define $\alpha(d) \equiv \alpha(\boldsymbol{s}; \boldsymbol{s}_0)$ with $d = \|\boldsymbol{s} - \boldsymbol{s}_0\|$, and similarly for $\beta(d)$. Assuming that the residual field $Z_t(\boldsymbol{s}; \boldsymbol{s}_0)$ is isotropic, we denote the mean and variance of $[Y_t(\boldsymbol{s}) \mid Y_t(\boldsymbol{s}_0) = y_0]$ as $\mu(d; y_0)$ and $\zeta(d; y_0)^2$, respectively. Under the spatial conditional extremes model (4), these equal

$$\mu(d; y_0) = \alpha(d) y_0, \qquad \zeta(d; y_0)^2 = \sigma(d)^2 y_0^{2\beta(d)} + \sigma_\varepsilon^2,$$

where σ_ε^2 is the variance of the nugget term $\varepsilon_t(\boldsymbol{s}; \boldsymbol{s}_0)$ and $\sigma(d)^2$ is the variance of $Z_t(\boldsymbol{s}; \boldsymbol{s}_0)$ when $\|\boldsymbol{s} - \boldsymbol{s}_0\| = d$. By computing the empirical mean, $\hat{\mu}(d; y_0)$, of the conditional precipitation intensity, we can estimate $\alpha(d)$

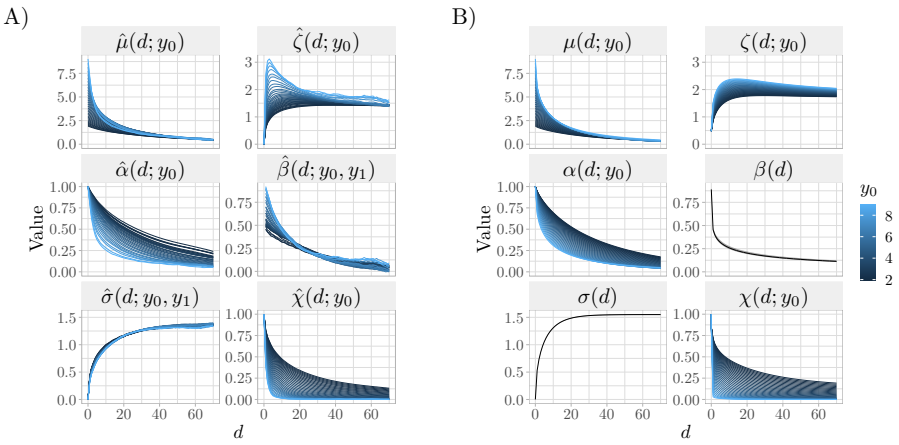


Figure 3: *A) Empirical moments and estimators for $\chi(d; y_0)$ and components of the conditional extremes model (4). B) Posterior means of the same variables from the model fit based on conditioning site nr. 2.*

as $\hat{\alpha}(d; y_0) = \hat{\mu}(d; y_0)/y_0$. Similarly, by assuming that σ_ε is small, we can estimate $\beta(d)$ using empirical conditional variances as

$$\hat{\beta}(d; y_0, y_1) = \frac{\log\left(\hat{\zeta}(d; y_0)/\hat{\zeta}(d; y_1)\right)}{\log\left(y_0/y_1\right)},$$

where $\hat{\zeta}(d; y)$ is the empirical standard deviance of all observations with distance d to a conditioning site with threshold exceedance y, and $y_0 \neq y_1$ are any two threshold exceedances $y_0, y_1 > \tau$. This allows us to create multiple different estimators $\hat{\alpha}(d; y_0)$ and $\hat{\beta}(d; y_0, y_1)$ for $\alpha(d)$ and $\beta(d)$, respectively, by varying the values of y_0 and y_1. This provides a diagnostic for estimating the threshold τ, since the spatial conditional extremes model assumes that $\alpha(d)$ and $\beta(d)$ are constant for all threshold exceedances larger than τ. Thus, we compute $\hat{\alpha}(d; y_0)$ and $\hat{\beta}(d; y_0, y_1)$ for a large range of values of y_0 and y_1, and we set τ equal to the smallest value such that the estimators are approximately constant for all $y_0, y_1 > \tau$. Then, we propose parametric functions for $\alpha(d)$ and $\beta(d)$ that can fit well to the patterns that we find in the empirical estimators. Finally, we also compute $\hat{\sigma}(d; y_0, y_1) = \hat{\zeta}(d; y_0) y_0^{-\hat{\beta}(d; y_0, y_1)}$ to get an idea about the marginal variances of the residual process $Z(\boldsymbol{s}; \boldsymbol{s}_0)$.

Exploratory analysis hints at some weak anisotropy in the precipitation data. However, we do not believe the lack of isotropy is strong enough to cause considerable problems, and the development of a suitable anisotropic model is outside the scope of this paper. We therefore assume an isotropic model. We compute $\hat{\mu}(d; y_0)$ and $\hat{\zeta}(d; y_0)$ with a sliding window approach.

For any value of d and y_0, the moments are estimated using all observations within a distance $d \pm 0.5$ km from a location where a value of $\log(y_0) \pm 0.025$ is observed. We then compute $\hat{\alpha}(d; y_0)$, $\hat{\beta}(d; y_0, y_1)$ and $\hat{\sigma}(d; y_0, y_1)$ as previously described, where we fix y_1 to the 90%-quantile of the Laplace distribution. We also estimate empirical conditional exceedance probabilities $\chi(d; y_0) \equiv \chi_p(\boldsymbol{s}, \boldsymbol{s}_0)$, where $d = \|\boldsymbol{s} - \boldsymbol{s}_0\|$ and $p = F^{-1}(y_0)$ with F^{-1} the quantile function of the Laplace distribution, using a similar sliding window approach. All the estimators are displayed in subplot A of Figure 3. The estimated shape of $\sigma(d)$ corresponds well with the standard deviation of a random field with constant variance, constrained to be zero at $\boldsymbol{s}_0$. The estimators for $\alpha(d)$ seem to behave like functions with exponential-like decay towards zero as d increases. However, the decay occurs at an increasing rate as y_0 increases, and it never seems to stabilise as a function of y_0. This indicates that, with the available amounts of data, we cannot choose a large enough threshold τ such that the function $a(d; y_0) = \alpha(d) y_0$ provides a good fit for all $y_0 > \tau$. Therefore, we instead propose to model the mean as $a(d; y_0) = \alpha(d; y_0) y_0$, where the function $\alpha(\cdot)$ depends on both distance d and intensity level y_0, and we choose a relatively low threshold, $\tau = y_1$.

A common model for $\alpha(d)$ is $\alpha(d) = \exp\{-(d/\lambda_a)^{\kappa_a}\}$, with $\lambda_a, \kappa_a > 0$, see, e.g, Wadsworth and Tawn (2022), Richards et al. (2022) and Simpson and Wadsworth (2021). We therefore examine if the model

$$\alpha(d; y_0) = \exp[-\{d/\lambda_a(y_0)\}^{\kappa_a(y_0)}]$$

can provide a good fit to our data, where $\lambda_a(y_0)$ and $\kappa_a(y_0)$ are parametric functions of y_0. Using a sliding window approach, we estimate $\lambda_a(y_0)$ and $\kappa_a(y_0)$ by minimising the sum of squared errors between $[Y(\boldsymbol{s}) \mid Y(\boldsymbol{s}_0) = y_0]$ and $a(\|\boldsymbol{s} - \boldsymbol{s}_0\|; y_0)$ for different values of y_0. These least squares estimators are displayed in the Supplementary Material. Based on our findings we propose to model $\lambda_a(y_0)$ and $\kappa_a(y_0)$ as

$$\lambda_a(y_0) = \lambda_{a0} \exp(-(y_0 - \tau)/\Lambda_\lambda), \quad \kappa_a(y_0) = \kappa_{a0} \exp(-((y_0 - \tau)/\Lambda_\kappa)^{\varkappa}),$$

with $\lambda_{a0}, \kappa_{a0}, \Lambda_\lambda, \Lambda_\kappa, \varkappa > 0$.

The estimators for $\beta(d)$ in subplot A of Figure 3 have a clear dependence on y_0 at short distances d. However, $\hat{\beta}(d; y_0, y_1)$ seems to be independent of y_0 for longer distances, and the changes as a function of y_0 are much less severe than those in $\hat{\alpha}(d; y_0)$. We therefore stick to a model on the form $b(d; y_0) = y_0^{\beta(d)}$. Based on the estimators in subplot A of Figure 3, it seems that $\beta(d)$ should be modelled with a function that decays exponentially towards zero. We therefore choose the model

$$\beta(d) = \beta_0 \exp(-(d/\lambda_b)^{\kappa_b}), \quad \lambda_b, \kappa_b > 0, \ \beta_0 \in [0, 1).$$

Having chosen the threshold τ and parametric forms for the functions $a(d; y_0)$ and $b(d; y_0)$, we then apply the method of Vandeskog, Martino, and Huser (2022) for defining a nonstationary and constrained SPDE approximation to the spatial Gaussian random field in (4) within `R-INLA`. This SPDE model approximates spatial Gaussian random fields on the form $b(\boldsymbol{s}; \cdot)Z(\boldsymbol{s})$ as a linear combination of m Gaussian mesh nodes, $\hat{Z}_b(\boldsymbol{s}) = \sum_{i=1}^{m} \phi_i(\boldsymbol{s}) b_i W_i$, where $b_1, b_2, \ldots, b_m$ are the values of the function $b(\boldsymbol{s})$ at the location of the m mesh nodes, and ϕ_i and W_i are basis functions and Gaussian mesh nodes from the "standard" SPDE approximation, $\hat{Z}(\boldsymbol{s}) = \sum_{i=1}^{m} \phi_i(\boldsymbol{s}) W_i$, of Lindgren et al. (2011). The nonstationary SPDE approximation is then constrained at $\boldsymbol{s}_0$ by placing one of the mesh nodes at $\boldsymbol{s}_0$, and constraining it to be exactly zero.

For each of the five chosen conditioning sites, we perform inference with `R-INLA`, using data from all time points where τ is exceeded at $\boldsymbol{s}_0$ and all 6404 locations in $\mathcal{S}$. In an empirical Bayes like approach, we place Gaussian priors on the logarithms of the parameters in $a(\cdot)$, with variance 5^2 and means equal to their least squares estimators. For the parameters of $b(\cdot)$, we choose Gaussian priors with variance 5^2 for $\log(\lambda_b)$, $\log(\kappa_b)$ and $\log(\beta_0/(1-\beta_0))$, which ensures $\lambda_b, \kappa_b > 0$ and $\beta_0 \in (0, 1)$. The prior means are chosen based on the diagnostics in Figure 3. We set them equal to $\log(8.5)$, $\log(0.5)$ and $\log(0.65/(1-0.65))$, respectively. The parameters of $Z_t(\boldsymbol{s})$ are given the PC prior of Fuglstad et al. (2019), such that the prior probability that the range ρ exceeds 60 km is 5% and the prior probability that the standard deviation σ exceeds 4 is 5%. We fix the smoothness parameter to $\nu = 0.5$, to represent our belief about the smoothness properties of extreme precipitation fields.

Inference with `R-INLA` is performed within 1–4 hours for each conditioning site, using only one core on the same Linux server as before. We evaluate the five model fits by estimating posterior means of the same variables as in subplot A of Figure 3, using 1000 posterior samples of $\boldsymbol{\theta}$. Subplot B displays these posterior means from the model fit based on conditioning site nr. 2. Although there are some differences between subplot A and B, the patterns of the estimated curves are in general agreement, indicating a satisfactory model fit overall. The posterior mean of $\mu(d; y_0)$ is similar to that of the data, with some slight underestimation for large values of d and y_0. The standard deviation $\zeta(d; y_0)$ is slightly underestimated for small d, and overestimated for large d. For the values of $\chi(d; y_0)$, which we care the most about, this results in a weak underestimation for small d and large y_0, and overestimation for large d and small y_0. We believe that more complex models for $a(\cdot)$ and $b(\cdot)$, e.g., with $\beta(\cdot)$ being a function of y_0 at small d, would be able to further reduce the differences seen in Figure 3. However, for the scope of this paper, we deem that the current fit is good enough. We also believe that the combination of minor overestimation and underestimation of $\chi(\cdot)$ might somewhat cancel each other out. Estimators

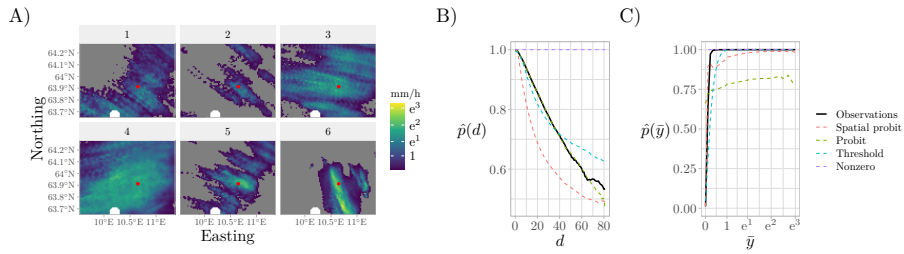


Figure 4: *A) Visualisation of observed precipitation for six time points where a threshold exceedance is observed at conditioning site nr. 4 (marked by a red dot). Grey denotes zero precipitation, while white denotes observations close to the Rissa radar that have been removed. The rightmost plots display empirical estimates for $p(d)$ (B) and $p(\bar{y})$ (C), created using both observed data and simulated data from the four different occurrence models.*

based on all five model fits are displayed in the Supplementary Material, and they all seem to capture the major trends of the data.

4.3 Modelling the conditional occurrence process

To model conditional precipitation occurrences, we first search for patterns in the observed data. Since we model $[I_t(\boldsymbol{s}) \mid Y(\boldsymbol{s}_0) > \tau]$, we expect the occurrence probability to be higher as we move closer to $\boldsymbol{s}_0$. We therefore compute empirical occurrence probabilities $\hat{p}(d)$ at different distances d from $\boldsymbol{s}_0$, using a sliding window of width 1 km. These are displayed by the black line in subplot B of Figure 4. As expected, $\hat{p}(d)$ decreases as d increases, with an almost linear decline. Subplot A of Figure 4 displays six realisations of the precipitation data. The distribution of precipitation occurrence appears to be smooth in space, in the sense that zeros cluster together. Thus, the non-spatial probit model is unable to produce realistic looking simulations. The precipitation intensities also appear to be smooth in space, in the sense that we never observe big jumps in the precipitation values, and that zeros only occur next to other zeros or small precipitation values. To check if this is true for all the available data, we estimate the probability $p(\bar{y})$ of observing precipitation as a function of the mean observed precipitation $\bar{y}$ at the four closest spatial locations. The empirical estimator is displayed using the black line in subplot C of Figure 4. It seems that the probability of observing precipitation is close to zero if $\bar{y} = 0$, and that it increases as a function of $\bar{y}$, and is almost exactly 1 if $\bar{y} > 0.2$ mm/h. This implies that our probit models might produce unrealistic simulations, as they are independent of the intensity model and might produce zeros close to large precipitation values.

Based on the exploratory analysis, we model the mean $\mu(d)$, of the two

probit models, using a spline. More specifically, we model $\partial\mu(d)/\partial d$ as a spline function based on 0-degree B-splines, where we place Gaussian zero-mean priors with a standard deviation of 10 on all spline coefficients. Additionally, in the spatial probit model, we place PC priors on the SPDE parameters (Fuglstad et al., 2019) such that the prior probability that the range parameter ρ exceeds 70 km is 5% and the prior probability that the standard deviation of $Z_t^*(\boldsymbol{s})$ exceeds 5 is 1%. The smoothness parameter is fixed to $\nu = 0.5$. We then perform inference separately for each of the five chosen conditioning sites and the two probit models. Inference with `R-INLA` takes between 10–15 minutes for the spatial probit models, and 2–3 minutes for the non-spatial probit models. For the threshold model, we estimate the threshold by computing the empirical probabilities of observing precipitation inside the Stordalselva catchment given extremes at each of the five conditioning sites.

We evaluate model performance by comparing properties of observed and simulated data. The threshold model depends on the intensity process, so we first simulate conditional intensities, by sampling $\boldsymbol{\theta}$ from its posterior distribution and then sampling both $[Y_t(\boldsymbol{s}_0) \mid Y_t(\boldsymbol{s}_0) > \tau]$ and $\{Y_t(\boldsymbol{s}) \mid Y_t(\boldsymbol{s}_0), \boldsymbol{\theta} : \boldsymbol{s} \in \mathcal{S}\}$ using (4). Then, we "simulate" occurrences with the threshold model by rounding all small enough precipitation intensities down to zero. Figure 4 displays empirical estimates for the probability of observing precipitation as a function of the distance d to $\boldsymbol{s}_0$ and as a function of the mean $\bar{y}$ of the four nearest neighbours, for observed and simulated precipitation data. Clearly, the nonzero model fails to capture the probability of precipitation occurrences. From subplot B we find that the spatial probit model heavily underestimates the probability of observing precipitation for most distances d. The threshold model performs better than the spatial probit model, but it slightly underestimates $p(d)$ for small d and overestimates it for large d. The non-spatial probit simulations, however, seem to agree well with the observed data for all values of d. From subplot C of Figure 4, we see that both probit models simulate zeros right next to large precipitation observations, resulting in an underestimation of $p(\bar{y})$. The spatial probit model performs better than the classical independent one, but it still does not completely solve this misfit. The threshold occurrence simulations, however, seem to agree well with the observations by placing its zeros close to other zeros or small precipitation values. Overall, the threshold model seems to be the best at estimating occurrence probabilities. The classical probit model is considerably better at estimating $p(d)$, but it completely fails at estimating $p(\bar{y})$.

4.4 Simulating spatial precipitation extremes

We combine all of the fitted models to simulate extreme precipitation over the Stordalselva catchment. For each of the five conditioning sites, extreme precipitation is simulated using Algorithm 1, with $N = 10^3$ samples, where Exp(1) denotes the exponential distribution with unit scale, $\mathcal{S}_1 \subset \mathbb{R}^2$ denotes all locations where we simulate extreme precipitation, $\hat{F}_t^{-1}$ denotes the estimated marginal quantile function of positive hourly precipitation at time t, and F denotes the cumulative distribution function of the Laplace distribution. Recall that $\mathcal{T}$ is the set of all time points in our data.

Algorithm 1 Simulating spatial extreme precipitation with conditioning site $\boldsymbol{s}_0$.

Sample N time points $t_1, \ldots, t_N$ uniformly with replacement from $\mathcal{T}$.
for $i = 1, 2, \ldots, N$ **do**
  Sample a threshold exceedance $[Y_i(\boldsymbol{s}_0) \mid Y_i(\boldsymbol{s}_0) > \tau] \sim \tau + \text{Exp}(1)$
  Sample a realisation of the conditional intensity process $\{Y_i(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}_1\} \mid Y_i(\boldsymbol{s}_0)$
  Sample a realisation of the conditional occurrence process $\{I_i(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}_1\} \mid Y_i(\boldsymbol{s}_0)$
  Transform back to the precipitation scale: $X_i^+(\boldsymbol{s}) = \hat{F}_{t_i}^{-1}[F(Y_i(\boldsymbol{s}))]$
  Add zeros to the samples: $X_i(\boldsymbol{s}) = X_i^+(\boldsymbol{s}) I_i(\boldsymbol{s})$
end for

Figure 5 displays observed and simulated realisations of extreme precipitation over the Stordalselva catchment. Simulations from the classical probit model and the nonzero model are not capturing the spatial structure of precipitation occurrence in the observed data, while simulations from the spatial probit model and the threshold model look more realistic. However, unlike the threshold model simulations, many of the spatial probit simulations contain large precipitation intensities right next to zeros, which is unrealistic.

As discussed in Section 1, both the amount of precipitation and its spatial distribution are important features for assessing flood risk. Thus, to further evaluate the simulations, we compare conditional exceedance probabilities and precipitation sums over different areas between the observed and simulated data. Estimators for $\hat{\chi}_p(d)$ are computed using the same sliding window approach as in Section 4.2. Figure 6 displays the estimators from conditioning site nr. 1. The simulations seem to capture $\chi_p(d)$ well, even though, just as in Section 4.2, $\chi_p(d)$ is somewhat underestimated for small d and overestimated for large d. The probit models seem to overestimate $\chi_p(d)$ less than the non-probit models at large distances, which makes sense, because the models are independent of the intensity process and thus can set large intensity values equal to zero. The estimators for $\chi_p(d)$ seem almost identical for the threshold model simulations and the nonzero model simulations. This also makes sense, since the small values that are rounded down to zero by the threshold model are too small to

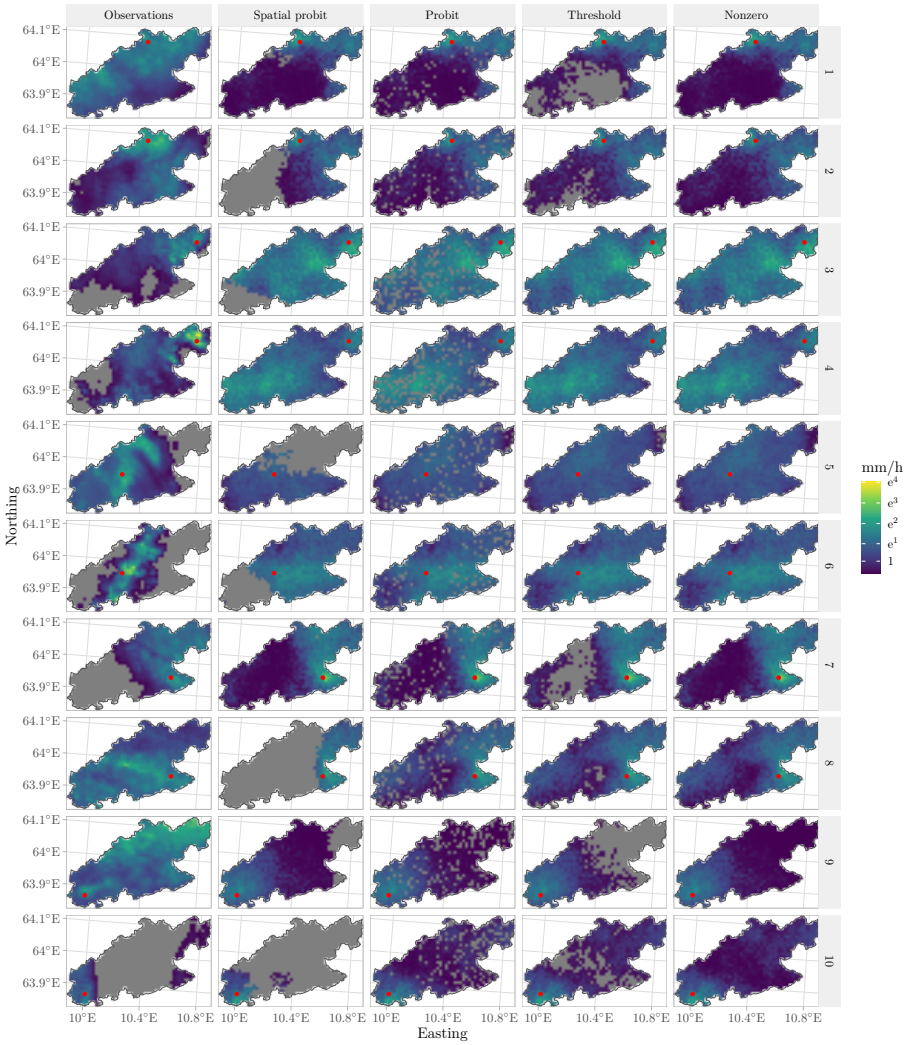


Figure 5: *Realisations of conditional extreme precipitation, from observed and simulated data. The red dots display the locations of the chosen conditioning site for each subplot.*

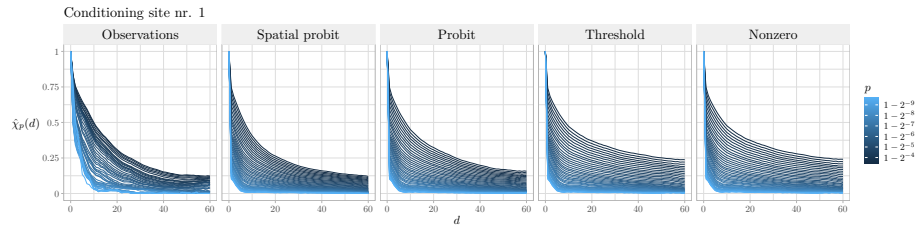


Figure 6: *Estimators of $\chi_p(d)$ for observed and simulated data, using conditioning site nr. 1.*

considerably affect the values of $\chi_p(d)$. Similar patterns are found for the other four conditioning sites. Estimators for $\chi_p(d)$ from all five conditioning sites are displayed in the Supplementary Material. They all display similar patterns.

We also compare aggregated simulated and observed precipitation amounts over the Stordalselva catchment to evaluate the simulations. For each conditioning site, we compute precipitation sums inside $\mathcal{B}_d(\boldsymbol{s}_0) \cap \mathcal{S}_1$, where $\mathcal{B}_d(\boldsymbol{s}_0)$ is a ball of radius d km, centred at $\boldsymbol{s}_0$, and $\mathcal{S}_1$ denotes the catchment of interest. We then compare observed and simulated precipitation sums using QQ plots. Figure 7 display these plots for conditioning site nr. 4. For small d, all the simulations produce similar precipitation amounts, which are close to the observed data, although slightly smaller. As d increases, the underestimation increases somewhat for the probit models, while it decreases for the non-probit models, which seem to agree well with the observed data. Quantiles of the threshold model and the nonzero model are almost identical and can be hard to distinguish. QQ plots for all five conditioning sites are displayed in the Supplementary Material. They all display similar patterns.

5 Discussion and conclusion

We propose a framework for modelling and simulating high-dimensional spatial precipitation extremes, using the spatial conditional extremes model and latent Gaussian models. We model the marginal distributions of nonzero precipitation using a mixture of two latent Gaussian models with a gamma likelihood and a generalised Pareto (GP) likelihood, while separately modelling the extremal dependence of nonzero precipitation extremes with a spatial conditional extremes model formulated as a latent Gaussian model. Precipitation occurrences are modelled using four different competing binary models. Fast approximate inference is achieved using integrated nested Laplace approximations (INLA). We develop new empirical diagnostics for the threshold and standardising functions of the spatial conditional extremes

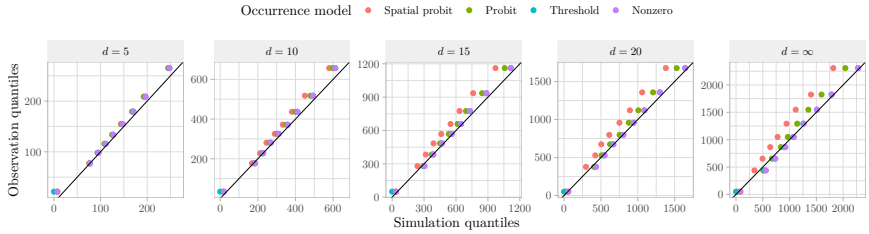


Figure 7: *QQ plots for the sum of aggregated precipitation inside* $\mathcal{B}_d(\boldsymbol{s}_0) \cap \mathcal{S}_1$*, based on conditioning site nr. 4. Quantiles from the four different simulated data sets are displayed using different colours.*

model, and we use these to propose a new class of parametric forms for the standardising functions. The developed framework is applied for simulating high-resolution precipitation extremes over a water catchment in Central Norway, and is found to perform well.

The threshold occurrence model appears to outperform the other occurrence models, as it captures both spatial and marginal properties of the data well, whereas the two probit models fail to capture either the marginal or the spatial properties. However, since the precipitation simulations stem from a combination of intensity samples and occurrence samples, it is nontrivial to conclude that one occurrence model significantly outperforms the others, as a change in the intensity model might potentially cause another occurrence model to produce the best precipitation simulations.

Compared to the probit models, the threshold model lacks some flexibility in the sense that it produces deterministic simulations given the intensity samples and a threshold. However, the threshold model interacts with the intensity model, which the probit models are unable to. This interaction is clearly crucial, as the threshold model ends up being our most successful. For future work, it might prove fruitful to develop probit models that interact more with the intensity model. A common approach for creating such interactions is to perform joint inference, where the two models share some latent model components, similar to Bacro et al. (2020) and Gelfand (2021). This might be challenging for the intensity and occurrence models, as their latent fields have different interpretations and scales, but it might still be possible to create some meaningful link between the two.

The spatial probit model underestimates occurrence probabilities almost everywhere in space. We believe this happens because the spatial clustering of zeros and ones in the data forces the nugget effect to be small, which, in the latent Gaussian model formulation, causes the latent field to have a large enough variance to absorb most of the mean trend. The symmetry of the latent field thus makes all marginal probabilities tend towards 50%, meaning that it underestimates large probabilities and overestimates small

probabilities. One might fix this by removing the conditional independence assumption, i.e., discarding the nugget effect. However, this makes inference with INLA impossible. Alternatively, one might use an asymmetric latent random field, with a skewness that varies in space, so that latent variables close to the conditioning site are right skewed, while latent variables further away from the conditioning site are more left skewed. Cabral et al. (2022) show that inference with INLA can be possible for non-Gaussian latent fields, meaning that `R-INLA` might work with such a model. Future work should attempt to add changes in skewness to the latent field of the spatial probit model.

As discussed in Section 1, radar data are great for capturing the small-scale spatio-temporal variability of precipitation, but not many have used them for modelling extreme precipitation. We show that the radar data let us capture small-scale extremal dependence structures with high precision. However, it weather radars are known to struggle with capturing the exact precipitation amounts, i.e., marginal distributions, well. This might negatively affect our estimates of aggregated precipitation amounts. Thus, to create reliable simulations of extreme precipitation, future work should attempt to combine information from multiple precipitation data sets, by, e.g., modelling extremal dependence using high-density radar data, while modelling marginal distributions using rain gauge observations, which are better at capturing the marginal distributions of precipitation, but that are too sparse in space to provide successful estimates of the extremal dependence structure.

Our models assume isotropy, but the observed data in Figures 4 and 5 display some indications of anisotropy. This does not seem to affect our results much, as the simulated data capture the main trends of the observed data well. However, future work should attempt to add features of anisotropy and/or nonstationarity into the model framework, which is possible within the SPDE and `R-INLA` frameworks (Lindgren et al., 2023).

It is known that `R-INLA` can struggle to approximate the posterior distribution if given suboptimal initial values, or if some parameters are not well identifiable in practice. Our chosen model for the conditional intensity process is highly flexible, and different combinations of the model parameters may sometimes produce similar likelihood values. In practice, we have seen that small changes in the model formulation or initial values can lead to large changes in the estimated parameters, and care should therefore be taken when applying this methodology in other settings. However, since these different parameters produce similar likelihood values, they all seem to perform equally well, when considering the QQ plots and estimates of $\chi_p(d)$ in Section 4.4. We have never observed a small change in model formulation or initial values that leads to a noticeably worse model fit overall.

Parameters of the marginal precipitation distributions are estimated

using latent Gaussian models with conditional independence assumptions given a relatively simple latent field. Such assumptions might fail to account for the complex spatio-temporal dependence structures of precipitation data and might therefore produce too small uncertainty estimates, due to an overestimation of the effective sample size. However, to the best of our knowledge, no computationally tractable methods exist that can accounting well for such complex spatio-temporal dependence in such large and high-dimensional data sets. Additionally, an underestimation of the uncertainty is not too problematic when we only use point estimates of the parameters for modelling the marginal distributions. Also, the reasonable parameter estimates displayed in Section 4.1 and the almost perfect QQ plots in the Supplementary Material imply that our marginal models perform well, even though they are based on oversimplified conditional independence assumption.

Similarly, the conditional intensity and occurrence models are purely spatial, and they assume that observations from different time points are independent, which can lead to too small uncertainty estimates. Future work should focus on the inclusion of a temporal component in all our models, to allow for better uncertainty quantification. Temporal components are also important for creating reliable simulations for hydrological models, as this allows for descriptions of the durations and movements of extreme precipitation events. We do not believe that this will entail too much work, as extensions from space to space-time can be relatively simple to achieve within the `R-INLA` framework. As an example, Simpson et al. (2023) successfully perform both spatial and spatio-temporal modelling with the spatial conditional extremes model, and it should be possible to extend most of their changes for space-time modelling into our developed framework.

Acknowledgements

Funding Raphaël Huser was partially supported by funding from King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2020-4394

Conflict of interest The authors report there are no competing interests to declare.

Supplementary material

S1 Data exploration

We compute the proportions of exact zeros in the precipitation data, pooled over space, for different times. Figure S1 displays temporal distributions of the proportions of observations that are less than or equal to a threshold τ_0, for different values of τ_0. The Rissa radar was upgraded in 2018 (personal communication, 2023), and this is clearly visible from the lower left subplot, as the proportion of exact zeros changed considerably in that year. In order to remove these changing zero-patterns, we post-process the data by rounding every observation smaller than 0.1 mm/h down to zero, as this seems to give a somewhat equal proportion of zeros everywhere in space and time.

S2 Evaluation of fitted marginal distributions

We evaluate goodness of fit for the two marginal models by creating quantile-quantile (QQ) plots. It is not straightforward to compare empirical quantiles with our model quantiles, as the models assume that each day comes with a different distribution, and thus a different set of quantiles. To create QQ plots for the gamma model, we therefore standardise the data by dividing each observation on its estimated scale parameter. Then we compare empirical quantiles of the standardised data with quantiles from a gamma distribution with a scale parameter of 1 and the estimated shape parameter. The resulting QQ plots are displayed in the upper row of Figure S2. Similarly, to create QQ plots for the generalised Pareto (GP) model, we standardise the observed threshold exceedances by dividing on their estimated scale parameters, and then we compare empirical quantiles with those of a GP distribution with a scale of 1 and the estimated tail parameter. The QQ plots are displayed in the lower row of Figure S2. Both of the marginal models seem to perform well. The GP model slightly underestimates the largest quantiles, but we here note that a value of 10 mm/h is so large that it corresponds to the empirical 99.3% quantile of the standardised threshold exceedances, which is approximately the same as the 99.97% quantile of the nonzero precipitation observations, and approximately the same as the $1-5\times10^{-5}$ quantile of the hourly summer precipitation observations, which again is slightly more than the 9 year return level for summer precipitation under the (unlikely) assumption that all observations are i.i.d.

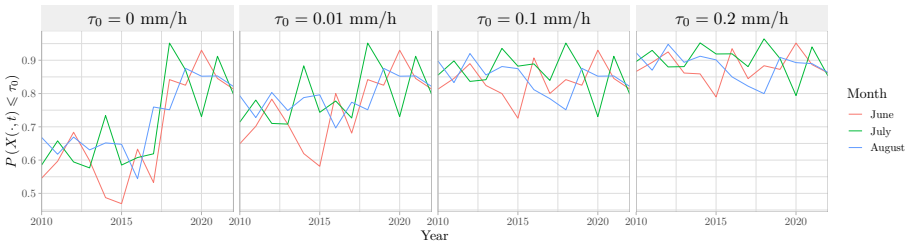


Figure S1: *The temporal distribution of the proportion of observations, pooled in space, that are less than or equal to τ_0 for June, July and August, respectively.*

S3 Modelling the conditional intensity process

In the main paper, when modelling the conditional intensity process, we discover that the empirical estimators $\hat{\alpha}(d; y_0)$ for $\alpha(d)$ seem to depend on y_0 for all values of y_0. We therefore propose to model $\alpha(\cdot)$ as a function $\alpha(d; y_0)$ that depends on y_0, and we propose the form

$$\alpha(d; y_0) = \exp\left[-(d/\lambda_a(y_0))^{\kappa_a(y_0)}\right],$$

where $\lambda_a(y_0), \kappa_a(y_0) > 0$ are parametric functions of y_0.

Since the mean of our spatial conditional extremes model is $a(d; y_0) = y_0\alpha(d; y_0)$, we can easily estimate $\lambda_a(y_0)$ and $\kappa_a(y_0)$ for any fixed value of y_0, by minimising the sum of squared errors between $[Y(\boldsymbol{s}) \mid Y(\boldsymbol{s}_0) = y_0]$ and $a(\|\boldsymbol{s} - \boldsymbol{s}_0\|; y_0)$. Thus, using a sliding window estimator over y_0, with a window size of 0.2, we estimate $\lambda_a(y_0)$ and $\kappa_a(y_0)$ for a set of different threshold exceedances $y_0 > \tau$. The estimators are displayed in the two leftmost plots of Figure S3. We see that the estimators for $\lambda_a(y_0)$ seem to decay exponentially towards zero as y_0 increases, while the estimators for $\kappa_a(y_0)$ look more like they follow the density function of a half-Gaussian distribution. We therefore propose the models

$$\lambda_a(y_0) = \lambda_{a0}\exp(-(y_0 - \tau)/\Lambda), \quad \kappa_a(y_0) = \kappa_{a0}\exp(-((y_0 - \tau)/\Lambda_\kappa)^{\varkappa}), \tag{S1}$$

where the parameters λ_{a0}, Λ, κ_{a0}, Λ_κ and $\varkappa$ all are required to be positive.

We fit the models in (S1) to the data by once more minimising the sum of squared errors. The resulting estimators for $\lambda_a(y_0)$ and $\kappa_a(y_0)$ are displayed as black lines in the two leftmost plots of Figure S3, and the resulting estimators for $\alpha(d; y_0)$ and $a(d; y_0)$ are displayed in the two rightmost plots of Figure S3. The fitted functions seem similar to those displayed in subplot A of Figure 3 in the main paper, and we therefore conclude that our chosen model for $\alpha(d; y_0)$ provides an adequate fit to the data.

Having performed inference with `R-INLA` for the conditional intensity process using each of the five chosen conditioning sites, we evaluate model

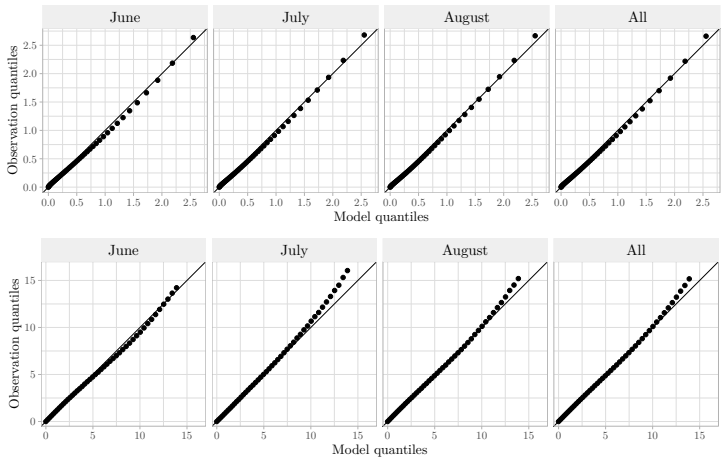


Figure S2: *Upper row: QQ plots comparing empirical quantiles of the standardised observations with quantiles of a gamma distribution with a scale of 1 and the estimated shape parameter. Lower row: QQ plots comparing empirical quantiles of the standardised threshold exceedance observations with quantiles of a GP distribution with a scale of 1 and the estimated tail parameter. The four columns display empirical quantiles for June, July, August and for all the months together, respectively.*

performance by simulating out-of-sample data and computing empirical estimators for $\mu(d; y_0)$, $\zeta(d; y_0)$, $\alpha(d)$, $\beta(d)$, $\sigma(d)$ and $\chi(d; y_0)$, just as in Figure 3 in the main paper. Figure S4 displays these estimators for simulated data based on each of the five conditioning sites. It also displays the estimators from subplot A in Figure 3 of the main paper, which were computed "globally", by using every single location $\mathcal{S}$ as a possible conditioning site. The estimators for $\beta(d)$ vary slightly between the different model fits, and this difference leads to considerable changes in the standard deviation $\zeta(d; y_0)$ of the model fits. These differences might be caused by the fact that, as discussed in the main paper, our chosen model for $\beta(d)$ is somewhat simple, in that it does not account for the fact that the empirical estimators change as a function of y_0 for small values of d. Thus, for some model fits, $\beta(d)$ is given a value that captures the sharp spike of $\hat{\zeta}(d; y_0)$ that occurs at small d with large values of y_0, while for other fits, $\beta(d)$ is given a value that better captures the more smooth values of $\hat{\zeta}(d; y_0)$ when y_0 is small. We believe that a more complex model for $\beta(d)$, possibly that changes as a function of y_0, would be able to capture both of these characteristics better. As shown in Section S4, the model fits with both of the different forms of $\beta(d)$ perform well and produce simulated data that closely capture important properties of the observed data such as its aggregated precipitation sums and extremal dependence structure.

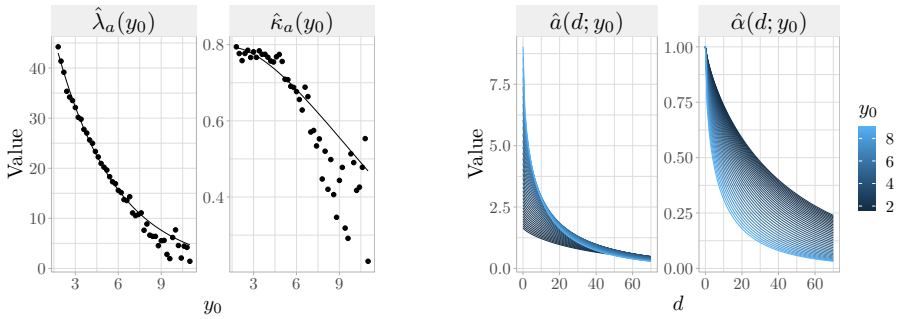


Figure S3: *Leftmost plots: Points displaying least squares estimators of* $\lambda_a(y_0)$ *and* $\kappa_a(y_0)$*, computed using a sliding window over* y_0 *with a width of* 0.2*. The black lines display the least squares estimators of* $\lambda_a(y_0)$ *and* $\kappa_a(y_0)$ *under the model in* (S1)*, computed without the sliding window approach, i.e., using all possible threshold exceedances. Rightmost plots: The estimated functions for* $\alpha(d; y_0)$ *and* $a(d; y_0)$*, created using the least squares estimators of* $\lambda_a(y_0)$ *and* $\kappa_a(y_0)$*.*

S4 Evaluating the final precipitation simulations

We evaluate the precipitation simulations by computing conditional exceedance probability estimators and by creating QQ plots for the sums of aggregated precipitation over different areas inside the Stordalselva catchment. Figure S5 displays empirical estimators for $\chi_p(d)$ for both the observed and the simulated data, using each of the five conditioning sites. The simulated estimators correspond well to the observation estimators, overall. Yet, they tend to be smaller than those of the observed data for large p and small d, and they tend to be larger than those of the observed data for small large d. This is further discussed in the main paper. Figure S6 display QQ plots for the sum of precipitation inside a ball of radius d, centred at $\boldsymbol{s}_0$, for each of the five conditioning sites, between the observed data and simulations from the four different model fits. The QQ plots show that there is a good correspondence between observed and simulated data, but that the simulated data tend to slightly underestimate aggregated precipitation amounts.

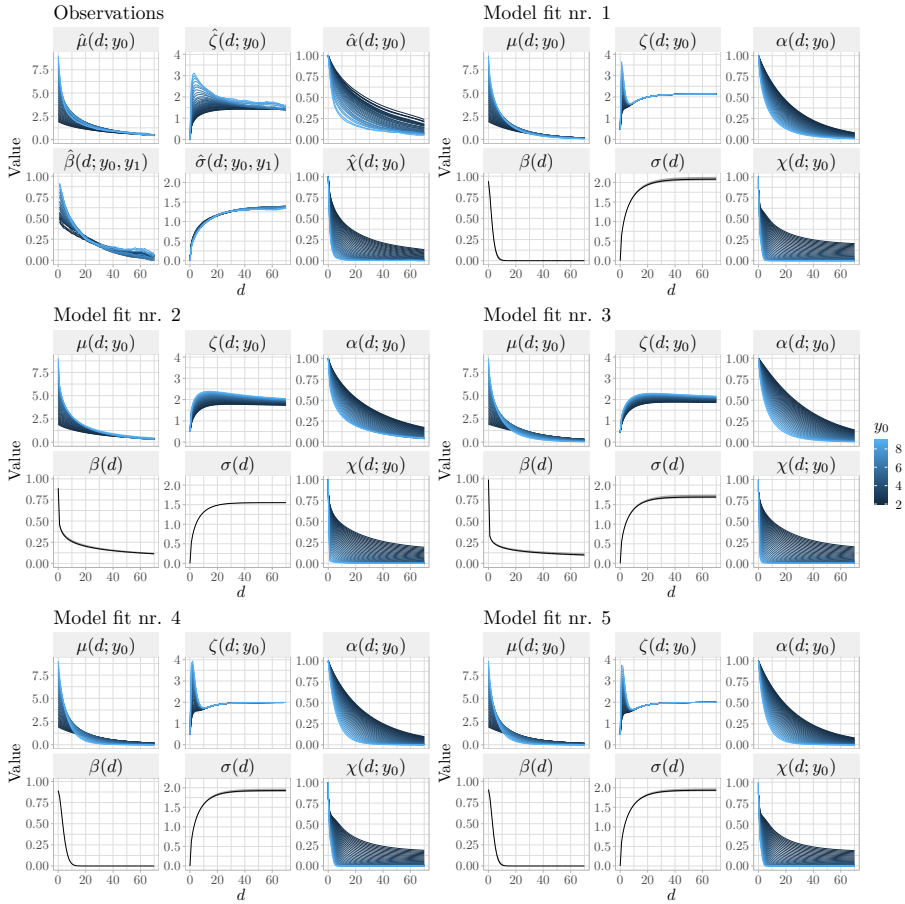


Figure S4: *Estimators for $\mu(d; y_0)$, $\zeta(d; y_0)$, $\alpha(d)$, $\beta(d)$, $\sigma(d)$ and $\chi(d; y_0)$ for observed data and for simulated data using the five chosen conditioning sites. The estimators are computed using the same sliding window approach as in the main paper.*

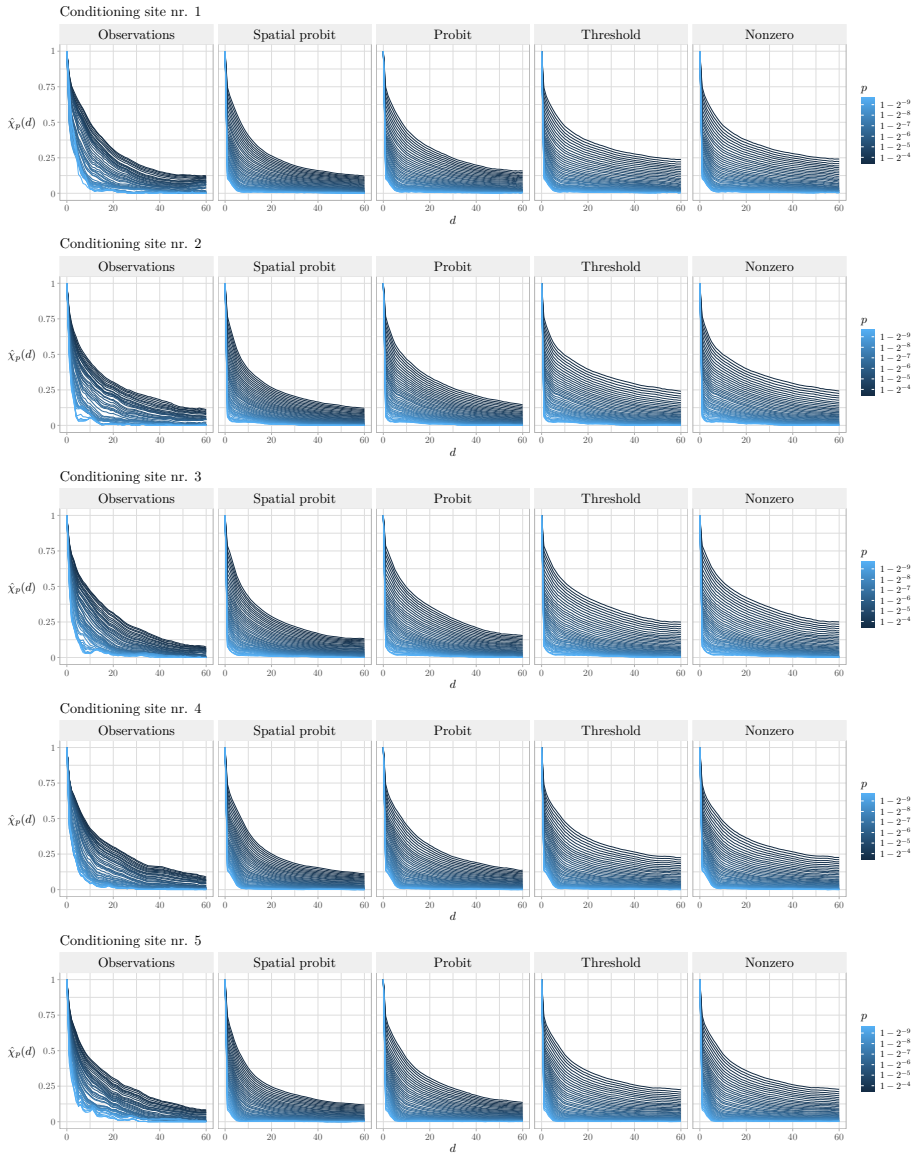


Figure S5: *Estimators of $\chi_p(d)$ for observed and simulated data, using all five conditioning sites.*

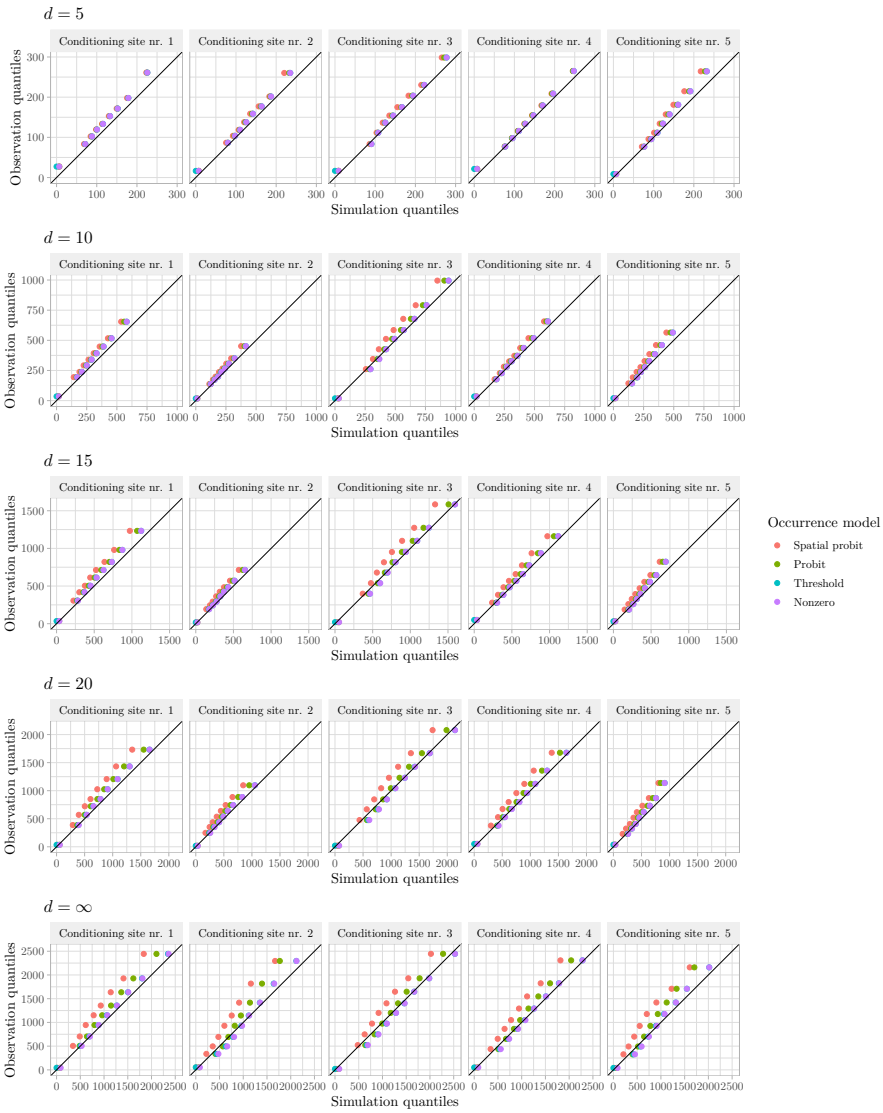


Figure S6: *QQ plots for the sum of aggregated precipitation over the intersection of the Stordalselva catchment and a ball of radius d, centred at $\boldsymbol{s}_0$, for each of the five conditioning sites. The four different simulations are displayed using different colours.*

References

Ailliot, P., Allard, D., Monbet, V., & Naveau, P. (2015). Stochastic weather generators: An overview of weather type models. *Journal de la Société Française de Statistique*, *156*(1), 101–113. http://www.numdam.org/item/JSFS_2015__156_1_101_0/

Allan, R. P., Barlow, M., Byrne, M. P., Cherchi, A., Douville, H., Fowler, H. J., Gan, T. Y., Pendergrass, A. G., Rosenfeld, D., Swann, A. L. S., Wilcox, L. J., & Zolina, O. (2020). Advances in understanding large-scale responses of the water cycle to climate change. *Annals of the New York Academy of Sciences*, *1472*(1), 49–75. https://doi.org/10.1111/nyas.14337

Bacro, J.-N., Gaetan, C., Opitz, T., & Toulemonde, G. (2020). Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *Journal of the American Statistical Association*, *115*(530), 555–569. https://doi.org/10.1080/01621459.2019.1617152

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC. https://doi.org/10.1201/b17115

Blöschl, G., Kiss, A., Viglione, A., Barriendos, M., Böhm, O., Brázdil, R., Coeur, D., Demarée, G., Llasat, M. C., Macdonald, N., Retsö, D., Roald, L., Schmocker-Fackel, P., Amorim, I., Bělínová, M., Benito, G., Bertolin, C., Camuffo, D., Cornel, D., … Wetter, O. (2020). Current European flood-rich period exceptional compared with past 500 years. *Nature*, *583*(7817), 560–566. https://doi.org/10.1038/s41586-020-2478-3

Bournas, A., & Baltas, E. (2022). Determination of the Z-R relationship through spatial analysis of X-band weather radar and rain gauge data. *Hydrology*, *9*(8). https://doi.org/10.3390/hydrology9080137

Cabral, R., Bolin, D., & Rue, H. (2022). Fitting latent non-Gaussian models using variational Bayes and Laplace approximations. https://doi.org/10.48550/arxiv.2211.11050

Castro-Camilo, D., Huser, R., & Rue, H. (2019). A spliced gamma-generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(3), 517–534. https://doi.org/10.1007/s13253-019-00369-z

Castro-Camilo, D., Mhalla, L., & Opitz, T. (2021). Bayesian space-time gap filling for inference on extreme hot-spots: An application to Red Sea surface temperatures. *Extremes*, *24*(1), 105–128. https://doi.org/10.1007/s10687-020-00394-z

Castruccio, S., Huser, R., & Genton, M. G. (2016). High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics*, *25*(4), 1212–1229. https://doi.org/10.1080/10618600.2015.1086656

Cooley, D., Nychka, D., & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, *102*(479), 824–840. https://doi.org/10.1198/016214506000000780

Davison, A. C., & Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application*, *2*(1), 203–235. https://doi.org/10.1146/annurev-statistics-010814-020133

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, *27*(2), 161–186. https://doi.org/10.1214/11-STS376

Davison, A. C., Huser, R., & Thibaud, E. (2019). Spatial extremes. In A. E. Gelfand, M. Fuentes, J. A. Hoeting, & R. L. Smith (Eds.), *Handbook of environmental and ecological statistics* (pp. 711–744). Chapman; Hall/CRC.

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L., & Stordal, F. (2015). Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics*, *26*(2), 89–106. https://doi.org/10.1002/env.2301

Elo, C. A. (2012). Correcting and quantifying radar data [MET Norway report no. 2/2012].

Engelke, S., Opitz, T., & Wadsworth, J. (2019). Extremal dependence of random scale constructions. *Extremes*, *22*(4), 623–666. https://doi.org/10.1007/s10687-019-00353-3

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression*. Springer. https://doi.org/10.1007/978-3-642-34333-9

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, *114*(525), 445–452. https://doi.org/10.1080/01621459.2017.1415907

Gelfand, A. E. (2021). Multivariate spatial process models. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of regional science* (pp. 1985–2016). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-60723-7_120

Gelfand, A. E., Diggle, P. J., Fuentes, M., & Guttorp, P. (2010). *Handbook of spatial statistics*. CRC press. https://doi.org/10.1201/9781420072884

Giorgi, F. (2019). Thirty years of regional climate modeling: Where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, *124*(11), 5696–5723. https://doi.org/10.1029/2018JD030094

Hanssen-Bauer, I., Drange, H., Førland, E., Roald, L., Børsheim, K., Hisdal, H., Lawrence, D., Nesje, A., Sandven, S., Sorteberg, A., et al. (2015). Klima i Norge 2100 [Kunnskapsgrunnlag for klimatilpasning oppdatert i 2015, Norsk klimasenter, Oslo, Norway].

Huser, R., & Davison, A. C. (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(2), 439–461. https://doi.org/10.1111/rssb.12035

Huser, R., Opitz, T., & Thibaud, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics*, *21*, 166–186. https://doi.org/10.1016/j.spasta.2017.06.004

Huser, R., Opitz, T., & Thibaud, E. (2021). Max-infinitely divisible models and inference for spatial extremes. *Scandinavian Journal of Statistics*, *48*(1), 321–348. https://doi.org/10.1111/sjos.12491

Huser, R., & Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, *114*(525), 434–444. https://doi.org/10.1080/01621459.2017.1411813

Huser, R., & Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews (WIREs): Computational Statistics*, *14*(1), e1537. https://doi.org/10.1002/wics.1537

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. Chapman; Hall/CRC. https://doi.org/10.1201/9780367803896

Jongman, B. (2018). Effective adaptation to rising flood risk. *Nature Communications*, *9*(1), 1986. https://doi.org/10.1038/s41467-018-04396-1

Keef, C., Papastathopoulos, I., & Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, *115*, 396–404. https://doi.org/10.1016/j.jmva.2012.10.012

Krupskii, P., & Huser, R. (2022). Modeling spatial tail dependence with Cauchy convolution processes. *Electronic Journal of Statistics*, *16*(2), 6135–6174. https://doi.org/10.1214/22-EJS2081

Lindgren, F., Bakka, H., Bolin, D., Krainski, E., & Rue, H. (2023). A diffusion-based spatio-temporal extension of Gaussian Matérn fields. https://doi.org/10.48550/arXiv.2006.04917

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1–25. https://doi.org/10.18637/jss.v063.i19

Lindgren, F., Rue, H., & 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. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Lopez-Cantu, T., Prein, A. F., & Samaras, C. (2020). Uncertainties in future U.S. extreme precipitation from downscaled climate projections. *Geophysical Research Letters*, *47*(9), e2019GL086797. https://doi.org/10.1029/2019GL086797

Maraun, D., & Widmann, M. (2018). *Statistical downscaling and bias correction for climate research*. Cambridge University Press. https://doi.org/10.1017/9781107588783

Nelsen, R. B. (2006). *An introduction to copulas*. Springer New York. https://doi.org/10.1007/0-387-28678-0

Opitz, T., Huser, R., Bakka, H., & Rue, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, *21*(3), 441–462. https://doi.org/10.1007/s10687-018-0324-x

Padoan, S. A., Ribatet, M., & Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, *105*(489), 263–277. https://doi.org/10.1198/jasa.2009.tm08577

Palacios-Rodríguez, F., Toulemonde, G., Carreau, J., & Opitz, T. (2020). Generalized Pareto processes for simulating space-time extreme events: An application to precipitation reanalyses. *Stochastic Environmental Research and Risk Assessment*, *34*(12), 2033–2052. https://doi.org/10.1007/s00477-020-01895-w

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, *49*(1), 187–201. https://doi.org/10.1029/2012WR012557

Richards, J., Tawn, J. A., & Brown, S. (2022). Modelling extremes of spatial aggregates of precipitation using conditional methods. *The Annals of Applied Statistics*, *16*(4), 2693–2713. https://doi.org/10.1214/22-AOAS1609

Richards, J., Tawn, J. A., & Brown, S. (2023). Joint estimation of extreme spatially aggregated precipitation at different scales through mixture modelling. *Spatial Statistics*, *53*, 100725. https://doi.org/10.1016/j.spasta.2022.100725

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, *17*(1), 182–190. https://doi.org/10.1029/WR017i001p00182

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Shooter, R., Ross, E., Ribal, A., Young, I. R., & Jonathan, P. (2022). Multivariate spatial conditional extremes for extreme ocean environments. *Ocean Engineering*, *247*, 110647. https://doi.org/10.1016/j.oceaneng.2022.110647

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*(1), 1–28. https://doi.org/10.1214/16-STS576

Simpson, E. S., Opitz, T., & Wadsworth, J. L. (2023). High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. *Extremes*. https://doi.org/10.1007/s10687-023-00468-8

Simpson, E. S., & Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, *41*, 100482. https://doi.org/10.1016/j.spasta.2020.100482

Van de Vyver, H. (2012). Spatial regression models for extreme precipitation in Belgium. *Water Resources Research*, *48*(9). https://doi.org/10.1029/2011WR011707

Vandeskog, S. M., Martino, S., Castro-Camilo, D., & Rue, H. (2022). Modelling sub-daily precipitation extremes with the blended generalised extreme value distribution. *Journal of Agricultural, Biological and Environmental Statistics*, *27*(4), 598–621. https://doi.org/10.1007/s13253-022-00500-7

Vandeskog, S. M., Martino, S., & Huser, R. (2022). An efficient workflow for modelling high-dimensional spatial extremes. https://doi.org/10.48550/arxiv.2210.00760

van Niekerk, J., Bakka, H., Rue, H., & Schenk, O. (2021). New frontiers in Bayesian modeling using the INLA package in R. *Journal of Statistical Software*, *100*(2), 1–28. https://doi.org/10.18637/jss.v100.i02

van Niekerk, J., Krainski, E., Rustand, D., & Rue, H. (2023). A new avenue for Bayesian inference with INLA. *Computational Statistics & Data Analysis*, *181*, 107692. https://doi.org/10.1016/j.csda.2023.107692

Verdin, A., Rajagopalan, B., Kleiber, W., & Katz, R. W. (2015). Coupled stochastic weather generation using spatial and generalized linear models. *Stochastic Environmental Research and Risk Assessment*, *29*(2), 347–356. https://doi.org/10.1007/s00477-014-0911-6

Wadsworth, J. L., & Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, *99*(2), 253–272. https://doi.org/10.1093/biomet/asr080

Wadsworth, J. L., & Tawn, J. A. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, *51*, 100677. https://doi.org/10.1016/j.spasta.2022.100677

Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., & Roberts, N. M. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, *52*(3), 522–555. https://doi.org/10.1002/2014RG000464

Yin, J., Gentine, P., Zhou, S., Sullivan, S. C., Wang, R., Zhang, Y., & Guo, S. (2018). Large increase in global storm runoff extremes driven by climate and anthropogenic changes. *Nature Communications*, *9*(1), 4389. https://doi.org/10.1038/s41467-018-06765-2

ISBN 978-82-326-7458-9 (printed ver.)
ISBN 978-82-326-7457-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology