Modelling Sub-daily Precipitation Extremes with the Blended Generalised Extreme Value Distribution

Silius M. VANDESKOG, Sara MARTINO, Daniela CASTRO- CAMILO, and Håvard RUE

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 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.

Supplementary materials accompanying this paper appear on-line.

Key Words: bGEV distribution; Block maxima modelling; INLA; Spatial statistics.

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

S. M. Vandeskog (✉) · S. Martino, Norwegian University of Science and Technology (NTNU), Trondheim, Norway (E-mail: *silius.m.vandeskog@ntnu.no*).

D. Castro-Camilo, University of Glasgow, Glasgow, UK.

H. Rue, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

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. Davison et al. 2015; Coles 2001). Due to low data quality (see Sect. 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 and 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. (2021) 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. (2021) 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. Wilson and Toumi 2005; Katz et al. 2002; Papalexiou and 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); Geirsson et al. (2015); Davison et al. (2012); 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 US 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; Ulrich et al. 2020; Lehmann et al. 2016; Wang and So 2016). 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 speed-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 and Segers 2018; Zou et al. 2019). 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 Sect. 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 Sect. 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 to 400 mm/h. Other unrealistic patterns, like 50 mm/hprecipitation 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/h, 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 2 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.
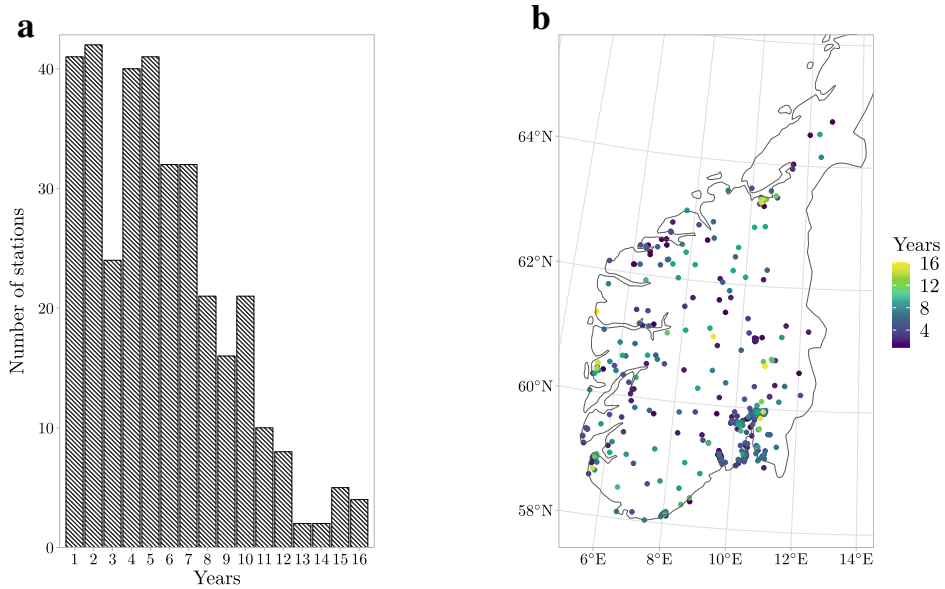
a

b

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 is displayed using different colours. Note that some points overlap in areas with high station densities.

The yearly maxima of precipitation accumulated over 1, 2, . . . , 24 h are computed for all locations and available years. A rolling window approach with a step size of 1 h is used for locating the precipitation maxima. As noted by Robinson and Tawn (2000), a sampling frequency of 1 h 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–SS02:00, 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 1 min, while the other half only contain hourly observations. We therefore use a sampling frequency of 1 h 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-

Table 1. Explanatory variables used for modelling sub-daily precipitation extremes

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	✓	✓

The two rightmost columns show which explanatory variables are used for modelling which parameters of the bGEV distribution for yearly precipitation (see Sect. 3.2)

daily precipitation. It is trivial to add more, or differently cleaned data, to improve the return level estimates at a later time.

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 $50 \times 50 \, \text{m}^2$, 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. (2018). 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 et al. 2018a,b). 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. Davison et al. 2015; Coles 2001). 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 and Tippett 1928; Coles 2001)

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

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. Bücher and Segers 2017; Smith and [Richard L.], 1985) 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. (2021) 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. (2021) 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. Opitz et al. 2018; Koutsoyiannis et al. 1998; Castro-Camilo et al. 2019; Sang and Gelfand 2010). 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) perform 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 suggest 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) argue 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 min, 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 and [Ole Einar], Bjørdal I, Skjelvåag AO, Aune B, 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. (2021), 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 $y_t(s)$ denote the maximum precipitation at location $s \in \mathcal{S}$ 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,

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

where all observations are assumed to be conditionally independent given the parameters $\mu_\alpha(s)$, $\sigma_\beta(s)$ and $\xi(s)$. 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. $\xi(s) = \xi$. As discussed in Sect. 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 $\mu_\alpha(s)$ and $\sigma_\beta(s)$ (results not shown). Two competing models are constructed for describing the spatial structure of $\mu_\alpha(s)$ and $\sigma_\beta(s)$.

3.2.1. 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

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

where $\mathbf{x}_\mu(s)$ and $\mathbf{x}_\sigma(s)$ 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 $u_\mu(s)$ is a zero-mean Gaussian field with Matérn correlation function, i.e.

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

Here, $d(s_i, s_j)$ is the Euclidean distance between s_i and 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 (Stein 1999; Matern 1986; Guttorp and 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 and Rue 2015).

The joint model is similar to the models of Geirsson et al. (2015); Dyrrdal et al. (2015); Davison et al. (2012). 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.

3.2.2. 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 $\sigma_\beta(s)$ 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 $\sigma_\beta(s)$.

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

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

with tail parameter ξ and scale parameter $\zeta(s) = \sigma(s) + \xi(x_{thr}(s) - \mu(s))$, where $\mu(s)$ and $\sigma(s)$ 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 $\mu(s)$ or $\sigma(s)$. We therefore assume that the difference $x_{thr}(s) - \mu(s)$ between the threshold and the location parameter is proportional to the scale parameter $\sigma(s)$. This assumption leads to the spread $\sigma_\beta(s)$ being proportional to the standard deviation of all observations larger than the threshold $x_{thr}(s)$. Based on this assumption, it is possible to model the spatial structure of the spread parameter independently of the location and tail parameter. Denote

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

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

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

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

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

where $\mu_\alpha^*(s) = \mu_\alpha(s)/\sigma^*(s)$. 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 and 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 $\mu_\alpha^*(s)$ is modelled as a linear combination of explanatory variables $\mathbf{x}_\mu(s)$ and a Gaussian random field $u_\mu(s)$, just as $\mu_\alpha(s)$ in the joint model (3). For estimation of $\sigma^*(s)$, the threshold $x_{thr}(s)$ is chosen as the 99% quantile of all observed precipitation at

location s. The precipitation observations larger than $x_{thr}(s)$ are declustered to account for temporal dependence, and only the cluster maximum of an exceedance is used for estimating $\sigma^*(s)$. 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 $\sigma^*(s)$, 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, $\sigma^*(s)$ is only computed at stations with more than 3 years of data. In order to estimate $\sigma^*(s)$ at locations with little or no observations, a linear regression model is used, where the logarithm of $\sigma^*(s)$ is assumed to have a Gaussian distribution,

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

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

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

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 $\log(\sigma^*(s))$ while still using the R-INLA framework for inference, which is not possible with the joint model. In Sect. 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 $\sigma^*(s)$ 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 $\log(\sigma^*(s))$ and estimate $(\mu_\alpha^*(s),\sigma_\beta^*,\xi)$ 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, 2017; Bivand et al. 2015). 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 Sect. 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 and Winkler 1976; Gneiting and Raftery 2007; Friederichs and Thorarinsdottir 2012),

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

where F is the forecast distribution, y is an observation, $\ell_p(x) = x(p - I(x < 0))$ is the quantile loss function and $I(\cdot)$ is an indicator 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 and Ranjan 2011) is a modification of

the CRPS that allows for emphasis on specific areas of the forecast distribution,

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

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 (2019), 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 (2019):

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

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 and Wallin 2019).

Using R-INLA we are able to sample from the posterior distribution of the bGEV parameters at any location s. The forecast distribution at location s is therefore given as

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

where F is the distribution function of the bGEV distribution and $\left(\mu_\alpha^{(i)}(s), \sigma_\beta^{(i)}(s), \xi^{(i)}\right)$ 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 Sect. 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 $\sigma^*(s)$. The uncertainty propagation is achieved using 100 bootstrap samples, as described in Sect. 3.2.2. Additionally, we modify the two-step model and add a random Gaussian field $u_\sigma(s)$ to the

linear predictor of the log-spread, to test if this can yield any considerable improvement in model performance. Just as $u_\mu(s)$, $u_\sigma(s)$ 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 $y^*(s)$ (Koenker 2005, 2020). 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 Sect. 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, which 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 $\sigma^*(s)$ 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 Sect. 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 $0 \le \xi < 1$ 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 $0 \le \xi < 0.5$. 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 fivefold 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 $\sigma^*(s)$ into the estimation of $(\mu_\alpha^*(s), \sigma_\beta^*, \xi)$. 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 h. Uncertainty is propagated using $B = 100$ bootstrap samples. Estimation of the posterior of $(\mu_\alpha^*(s), \sigma_\beta^*, \xi)$ given some value of $\sigma^*(s)$ takes less than 2 min 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 min.

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 $u_\mu(s)$ for the standardised precipitation maxima, are displayed in Table 3 for some selected temporal aggregations. These estimates are computed

Table 2. Mean StwCRPS with weight function $w(p) = I(p > 0.9)$ for fivefold cross-validation performed using out-of-sample estimation and in-sample estimation

	Model	$u_\sigma(s)$	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

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 h, 3 h, 6 h, 12 hand 24 h. The best scores are written in bold

by drawing 20 samples from each of the 100 posterior distributions. The empirical mean, standard deviation and quantiles of these 2000 samples are then reported.

There is strong evidence that all the explanatory variables in $\mathbf{x}_\sigma(s)$ 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 $\sigma^*(s)$, 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 $u_\mu(s)$ is of approximately the same magnitude as most of the regression coefficients in $\boldsymbol{\beta}_\mu$.

Table 4 displays the posterior range of the $u_\mu(s)$. 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 increases. It seems that, for 1 h precipitation, the regression coefficients are unable to explain some kind of large-scale phenomenon that considerably affects the location parameter $\mu_\alpha(s)$. To correct this, the range of $u_\mu(s)$ 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 $u_\mu(s)$ is decreased. The posterior means of $u_\mu(s)$ for three different temporal aggregations are displayed over a $1 \times 1 \text{ km}^2$ gridded map in Fig. 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.

Table 3. Estimated regression coefficients β_μ, β_σ and estimated standard deviation $SD(u_\mu)$ of the Gaussian field $u_\mu(s)$ 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 h	β_μ	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
	$SD(u_\mu)$		0.084	0.023	0.047	0.085	0.127
	σ_β^*		0.118	0.002	0.113	0.118	0.123
	β_σ	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 h	β_μ	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
	$SD(u_\mu)$		0.062	0.013	0.045	0.060	0.085
	σ_β^*		0.123	0.003	0.117	0.123	0.129
	β_σ	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 h	β_μ	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
	$SD(u_\mu)$		0.067	0.011	0.051	0.065	0.090
	σ_β^*		0.125	0.004	0.118	0.125	0.133
	β_σ	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

Dyrrdal et al. (2015)). Based on Fig. 2, it becomes evident that our explanatory variables are unable to describe this regional difference when modelling hourly precipitation, and $u_\mu(s)$ has to do the job of separating between the east and the west. As the temporal aggregations increase from 1 h to 3 and 6 h, 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 $\mu_\alpha(s)$.

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 h. This makes sense given the observation of

Table 4. Estimated posterior mean and quantiles for the range ρ of the Gaussian field $u_\mu(s)$ 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 h	235	34	255	478
	3 h	78	39	75	147
	6 h	60	32	57	104
	12 h	84	31	83	145
	24 h	55	32	50	105
ξ	1 h	0.178	0.136	0.179	0.211
	3 h	0.090	0.057	0.089	0.120
	6 h	0.047	0.028	0.046	0.072
	12 h	0.032	0.010	0.031	0.048
	24 h	0.029	0.006	0.029	0.051
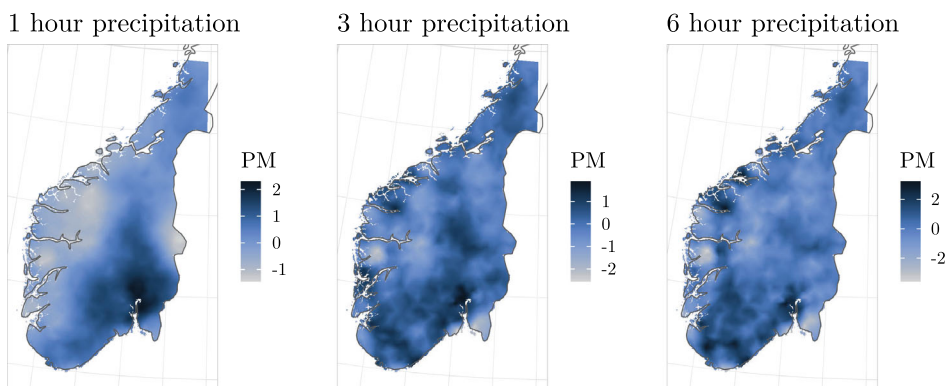
Figure 2. Estimated posterior mean (PM) of the Gaussian field $u_\mu(s)$ for three different temporal aggregations of precipitation, with unit mm.

Barbero et al. (2019) that most 24 h annual maximum precipitation comes from rainstorms with lengths of less than 15 h. Thus, the tail parameter for 24 h precipitation should be close to the tail parameter for 12 h precipitation. For 12 h 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 $1 \times 1 \text{ km}^2$. The posterior means and the widths of the 95% credible intervals are displayed in Fig. 3. For a period of 1 h 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
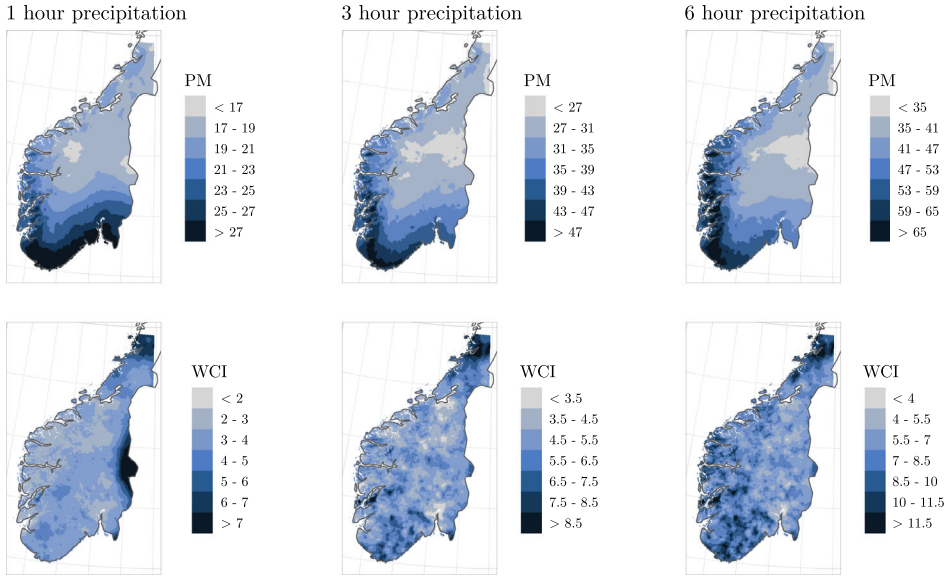
Figure 3. Estimated posterior mean (PM) and width of the 95% credible intervals (WCI) for the 20 year return levels of sub-daily precipitation. Different aggregation times are displayed in different columns. All numbers are given with unit mm.

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 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

REFERENCES

Bakka H, Rue H, Fuglstad G-A, Riebler A, Bolin D, Illian J, Lindgren F (2018) Spatial modeling with R-INLA: a review. WIREs Comput Stat 10(6):e1443. https://doi.org/10.1002/wics.1443

Barbero R, Fowler HJ, Blenkinsop S, Westra S, Moron V, Lewis E, Mishra V (2019) A synthesis of hourly and daily precipitation extremes in different climatic regions. Weather Clim 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. J Statist Softw 63(1):1–31

Bolin D, Wallin J (2019) Scale dependence: Why the average CRPS often is inappropriate for ranking probabilistic forecasts. arXiv:1912.05642

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. Electron J Statist 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. Stat Sci 36(3):360–378. https://doi.org/10.1214/20-STS795

Carreau J, Naveau P, Neppel L (2016) Characterization of homogeneous regions for regional peaksover- threshold modeling of heavy precipitation. working paper or preprint. Retrieved from 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. J Amer Statist Assoc 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. JABES 24(3):517–534. https://doi.org/10.1007/s13253-019-00369-z

Castro-Camilo D, Huser R, Rue H (2021) Practical strategies for GEV-based regression models for extremes. https://doi.org/10.48550/ARXIV.2106.13110

Coles S (2001). An introduction to statistical modeling of extreme values. 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. J Amer Statist Assoc 102(479):824–840. https://doi.org/10.1198/016214506000000780

Crespi A, Lussana C, Brunetti M, Dobler A, Maugeri M, Tveito OE (2018) High-resolution monthly precipitationclimatologies over Norway: assessment of spatial interpolation methods. arXiv:1804.04867

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

Davison AC, Huser R (2015) Statistics of extremes. Annu Rev Stat Appl 2(1):203–235. https://doi.org/10.1146/annurev-statistics-010814-020133

Davison AC, Smith RL (1990) Models for exceedances over high thresholds. J R Stat Soc B 52(3):393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

Davison AC, Padoan SA, Ribatet M (2012) Statistical modeling of spatial extremes. Statist Sci 27(2):161–186. https://doi.org/10.1214/11-STS376

Dyrrdal AV, Lenkoski A, Thorarinsdottir TL, 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 AV, Skaugen T, Stordal F, Førland EJ (2016) Estimating extreme areal precipitation in Norway from a gridded dataset. Hydrol Sci J 61(3):483–494. https://doi.org/10.1080/02626667.2014.947289

Fisher RA, Tippett LHC (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. Math Proc Cambridge Philos Soc 24(2):180–190. https://doi.org/10.1017/S0305004100015681

Friederichs P, Thorarinsdottir TL (2012) Forecast verifification 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. J Amer Statist Assoc 114(525):445–452. https://doi.org/10.1080/01621459.2017.1415907

Geirsson ÓP, 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 AE (2007) Strictly proper scoring rules, prediction, and estimation. J Amer Statist Assoc 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. J Bus Econ Stat 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

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

Hosking JRM, Wallis JR (1997) Regional frequency analysis

Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. Adv Water Res 25(8):1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8

Koenker R (2020) Quantreg: quantile regression. R package version 5.75. Retrieved from https://CRAN.Rproject.org/package=quantreg

Koenker R (2005) Quantile regression. Econometr Soc Monogr. https://doi.org/10.1017/CBO9780511754098

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

Lehmann EA, 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 TL, Ravazzolo F, Gneiting T (2017) Forecaster's dilemma: extreme events and forecast evaluation. Statist Sci 32(1):106–127. https://doi.org/10.1214/16-STS588

Lindgren F, Rue H (2015) Bayesian spatial modelling with R-INLA. J Stat Softw 63(19):1–25

Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J R Stat Soc B 73(4):423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

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

Lussana C, Tveito O, Uboldi F (2018) Three-dimensional spatial interpolation of 2 m temperature over Norway. Q J R Meteorolog Soc 144(711):344–364. https://doi.org/10.1002/qj.3208

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

Matheson JE, Winkler RL (1976) Scoring rules for continuous probability distributions. Manage Sci 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. Technical report. Retrieved from https://www.researchgate.net/profile/Matthias-Mohr/publication/265876264_Comparison_of_Versions_11_and_10_of_Gridded_Temperature_and_Precipitation_Data_for_Norway/links/56c5a19808ae736e7048bd0f/Comparison-of-Versions-11-and-10-of-Gridded-Temperature-and-Precipitation-Data-for-Norway.pdf

Naveau P, Toreti A, Smith I, Xoplaki E (2014) A fast nonparametric spatio-temporal regression scheme for generalized Pareto distributed heavy precipitation. Water Resour Res 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 SM, Koutsoyiannis D (2013) Battle of extreme value distributions: a global survey on extreme daily rainfall. Water Resour Res 49(1):187–201. https://doi.org/10.1029/2012WR012557

Robinson ME, Tawn JA (2000) Extremal analysis of processes sampled at different frequencies. J R Stat Soc B 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. J R Stat Soc B 71(2):319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK (2017) Bayesian computing with INLA: a review. Annu Rev Stat Appl 4(1):395–421

Sang H, Gelfand AE (2009) Hierarchical modeling for extreme values observed over space and time. Environ Ecol Stat 16(3):407–426

Sang H, Gelfand A (2010) Continuous spatial process models for spatial extreme values. J Agric Biol Environ Stat 15:49–65. https://doi.org/10.1007/s13253-009-0010-1

Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH (2017) Penalising model component complexity: a principled, practical approach to constructing priors. Statist Sci 32(1):1–28

Smith RL (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 ML (1999) Interpolation of spatial data?: some theory for Kriging. Springer series in statistics. Springer, New York

Sun Y, Bowman KP, Genton MG, Tokay A (2015) A Matérn model of the spatial covariance structure of point rain rates. Stoch Env Res Risk A 29(2):411–416. https://doi.org/10.1007/s00477-014-0923-2

Tveito OE, Bjørdal I, Skjelvåag AO, Aune B (2005) A GIS-based agro-ecological decision system based on gridded climatology. Meteorol Appl 12(1):57–68

Ulrich J, Jurado OE, Peter M, Scheibel M, Rust HW (2020) Estimating IDF curves consistently over durations with spatial covariates. Water 12(11):3119. https://doi.org/10.3390/w12113119

Van de Vyver H (2012) Spatial regression models for extreme precipitation in Belgium. Water Resour Res. https://doi.org/10.1029/2011WR011707

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

Wang Y, So MK (2016) A Bayesian hierarchical model for spatial extremes with multiple durations. Comput Statist Data Anal 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. Retrieved from http://www.jstor.org/stable/2332724

Wilson PS, Toumi R (2005) A fundamental probability distribution for heavy rainfall. Geophys Res Lett. https://doi.org/10.1029/2005GL022465

World Economic Forum. (2021). The global risks report 2021. Retrieved from http://www3.weforum.org/docs/WEFTheGlobalRisks Report 2021.pdf

Zou N, Volgushev S, Bücher A (2019) Multiple block sizes and overlapping blocks for multivariate time series extremes. arXiv:1907.09477 [math.ST]