Latent Gaussian Models for High-Dimensional Spatial Extremes

Arnab Hazra¹, Raphaël Huser¹, and Árni V. Jóhannesson²

October 7, 2021

Abstract

In this chapter, we show how to efficiently model high-dimensional extreme peaks-over-threshold events over space in complex non-stationary settings, using extended latent Gaussian Models (LGMs), and how to exploit the fitted model in practice for the computation of long-term return levels. The extended LGM framework assumes that the data follow a specific parametric distribution, whose unknown parameters are transformed using a multivariate link function and are then further modeled at the latent level in terms of fixed and random effects that have a joint Gaussian distribution. In the extremal context, we here assume that the data level distribution is described in terms of a Poisson point process likelihood, motivated by asymptotic extreme-value theory, and which conveniently exploits information from all threshold exceedances. This contrasts with the more common data-wasteful approach based on block maxima, which are typically modeled with the generalized extreme-value (GEV) distribution. When conditional independence can be assumed at the data level and latent random effects have a sparse probabilistic structure, fast approximate Bayesian inference becomes possible in very high dimensions, and we here present the recently proposed inference approach called “Max-and-Smooth”, which provides exceptional speed-up compared to alternative methods. The proposed methodology is illustrated by application to satellite-derived precipitation data over Saudi Arabia, obtained from the Tropical Rainfall Measuring Mission, with 2738 grid cells and about 20 million spatio-temporal observations in total. Our fitted model captures the spatial variability of extreme precipitation satisfactorily and our results show that the most intense precipitation events are expected near the south-western part of Saudi Arabia, along the Red Sea coastline.

1 Introduction

Extreme-value theory (Davison and Huser, 2015) has become the standard probabilistic tool to build statistical models and make inference for high-impact extreme events occurring in a wide range of geo-environmental applications (e.g., Davison and Gholamrezaee, 2012; Reich and Shaby, 2012; Huser and Davison, 2014; Jonathan et al., 2014; Asadi et al., 2015; Jalbert et al., 2017; Vettori et al., 2019; Engelke and Hitz, 2020). Classical extreme-value models rely on asymptotic arguments for block maxima or high threshold exceedances, when the
block size or the threshold, respectively, increases arbitrarily. On one hand, the celebrated Fisher–Tippett Theorem states that, in the univariate case, the only possible non-degenerate limit distribution for renormalized block maxima is the generalized extreme-value (GEV) distribution, and this motivates its use in practice for modeling maxima with large but finite block sizes—typically, yearly blocks in environmental applications. On the other hand, the Pickands–Balkema–de Haan Theorem states that the only possible non-degenerate limit distribution for high threshold exceedances is the generalized Pareto (GP) distribution, and this motivates its use in practice for modeling peaks over high but finite thresholds—often taken as the empirical 95%-threshold. These two seemingly different modeling techniques are in fact intrinsically related to each other through a Poisson point process representation that provides a unified description of the asymptotic upper tail; see Coles (2001) and Davison and Huser (2015) for more details. While these approaches are theoretically equivalent, they all have their pros and cons in practice. The block maximum approach avoids complications with intra-block seasonality and temporal dependence, but it has been criticized for being wasteful of data given that it only uses one value per block and ignores other smaller but potentially large events. By contrast, the threshold exceedance approach uses all extreme observations and often requires a detailed modeling of temporal dependence, which may either be a desideratum or a nuisance. While the “horse-racing” between the block maximum and threshold exceedance approaches is still debated (Bücher and Zhou, 2021), it is often believed that the threshold exceedance approach offers a richer, more convenient, and easily interpretable modeling framework, especially in the spatial context. In this work, we build Bayesian hierarchical models for high-resolution spatial precipitation extremes over Saudi Arabia, by exploiting the Poisson point process representation based on peaks-over-threshold.

Several approaches have been proposed for modeling spatial extremes; see Davison et al. (2012), Davison et al. (2019), and Huser and Wadsworth (2020) for comprehensive reviews on this topic. One possibility is to generalize univariate asymptotic models to the spatial setting, in such a way to model not only the marginal distribution of extremes accurately but also to capture their potentially strong dependencies using models justified by extreme-value theory. This leads to the class of max-stable processes for spatially-indexed block maxima (Padoan et al., 2010), and to generalized Pareto processes for spatial threshold exceedances defined in terms of a certain risk functional (Thibaud and Opitz, 2015). While these asymptotic models are backed up by strong theoretical arguments, their complicated probabilistic structure leads to awkward likelihood functions, which are computationally demanding to evaluate, thus limiting likelihood-based inference to relatively small dimensions (Huser and Davison, 2013; Castruccio et al., 2016; de Fondeville and Davison, 2018; Huser et al., 2019). Various types of “sub-asymptotic” extreme models that circumvent some of the issues of asymptotic models have also been proposed (Wadsworth and Tawn, 2012, 2019; Huser et al., 2017, 2021; Huser and Wadsworth, 2019; Zhong et al., 2021), though they usually still remain difficult
to fit in relatively high dimensions. Alternatively, when the main focus is to obtain accurate marginal return level estimates at observed and unobserved locations, while capturing data-level dependence is of secondary importance, it is convenient to rely on latent Gaussian models (LGMs). These models often assume conditional independence among observations and can thus be fitted more efficiently in high dimensions, while performing well for spatial smoothing and prediction by borrowing strength across locations when spatial effects are embedded within model parameters. LGMs have been successfully applied in a wide range of applications (see, e.g., Rue et al., 2009, and references therein), and belong to the broader class of Bayesian hierarchical models (Banerjee et al., 2003; Diggle and Ribeiro, 2007), whose specification is described at three levels: (i) the data level, specifying a parametric distribution for the observed data; (ii) the latent level, modeling unknown parameters, potentially transformed with a link function, using fixed and random effects; and (iii) the hyperparameter level, specifying prior distributions for hyperparameters. In the case of LGMs, all latent variables are specified with a joint Gaussian distribution. When a multivariate link function is used to transform parameters jointly such that fixed/random effects are embedded within multiple linear predictors at the latent level, then we usually refer to these models as extended LGMs (Geirsson et al., 2020; Hrafnkelsson et al., 2021; Johannesson et al., 2021).

While the data level is designed to capture the marginal stochastic variability using an appropriate probability distribution family, but often assumes conditional independence (of the data given the parameters) for computational convenience, the latent level is designed to capture non-stationary spatio-temporal variation, trends, and dependencies through random and covariate effects. When the goal is to accurately predict the probability of extremes at unobserved locations and to smooth return level estimates across space, it is key to incorporate spatial dependence at the latent level to borrow strength across locations. However, the exact form of dependence assumed at the latent level is not crucial for spatial prediction, and so the multivariate Gaussianity assumption of LGMs is not a major limitation. The hyperparameter level is used to sufficiently, but not overly, constrain model parameters with prior distributions, in order to stabilize estimation of all parameters and latent variables, and/or incorporate expert knowledge if desired.

Several types of extended LGMs have already been used in the extremes literature. Key differences between the proposed models are mainly with respect to the actual data likelihood being used at the data level (e.g., based either on block maxima or peaks-over-threshold), the detailed latent level specification (e.g., whether or not random effects are specified as Gaussian Markov random fields with a sparse precision matrix), and the actual method of inference (e.g., “exact” Markov chain Monte Carlo algorithms, or approximate Bayesian inference methods). While all these differences may appear to be relatively minor at first sight, they have in fact important implications in terms of the methodology’s scalability to high dimensions and the accuracy of the final return level estimates. The first attempt to
model extremes with an LGM in the literature is the paper by Coles and Casson (1998), who analyzed data generated from a climatological model at 55 equally spaced locations to characterize hurricane risk along the eastern coastlines of the United States. A Poisson point process likelihood for high threshold exceedances was used, whereby marginal location, scale and shape parameters were further modeled using a relatively simple representation in terms of mutually independent Gaussian spatial processes, and the inference was performed using a basic Metropolis–Hastings MCMC algorithm with random walk proposals. Later, Cooley et al. (2007) built an LGM designed for the spatial interpolation of extreme return levels, using the GP distribution for high threshold exceedances, and the inference was performed by MCMC based on a precipitation dataset available at 56 weather stations. Huerta and Sansó (2007) proposed a similar model for spatio-temporal extremes, fitted by a customized MCMC algorithm, based on the GEV distribution for block maxima (with replicated observations at 19 stations). Other similar LGMs were also proposed by Davison et al. (2012), Hrafnkelsson et al. (2012), Geirsson et al. (2015), and Dyrrdal et al. (2015), but these models were all applied to block maxima data in relatively small dimensions, due to the computational burden and difficulties related to the convergence of Markov chains in MCMC algorithms. Rue et al. (2009) developed the integrated nested Laplace approximation (INLA), a fast and accurate approximate Bayesian solution for estimating generic LGMs, and it was later exploited by Opitz et al. (2018) and Castro-Camilo et al. (2019) to fit relatively simple spatio-temporal extreme-value models based on the GP distribution, but unfortunately, the R-INLA software currently does not support extended LGMs, where distinct random effects control the behavior of multiple parameters at the latent level. This is a major limitation when the data likelihood function has several parameters (e.g., the location and scale) that display a complex spatially-varying behavior. Geirsson et al. (2020) later developed the LGM split sampler, which provides significant improvements in the mixing of Markov chains for extended LGMs, and thus reduces the overall computational burden, but they illustrated their algorithm on a relatively small extremes example. With gridded or areal data, it is also possible to model random effects using Gaussian Markov random fields (GMRFs), which have a sparse precision matrix. This offers major computational gains by improving the efficiency of random effects’ updates in large dimensions (see, e.g., Sang and Gelfand, 2009, 2010; Cooley and Sain, 2010; Jalbert et al., 2017, for extreme-value data examples). More recently, Johannesson et al. (2021) modeled a complex spatio-temporal dataset of yearly river flow maxima available at several hundreds of irregularly-spaced stations over the United Kingdom using an extended LGM based on the GEV distribution, that embeds multiple latent spatial random effects defined in terms of stochastic partial differential equations (SPDEs). The SPDE approach is intrinsically linked to GMRFs (Lindgren et al., 2011) and this yields fast inference thanks to the sparsity of precision matrices. Moreover, instead of using an “exact” MCMC algorithm, Johannesson et al. (2021) leveraged Max-and-Smooth, a two-step approximate
Bayesian MCMC-based inference scheme recently proposed by Hrafnkelsson et al. (2021), which shares some similarities with INLA to some degree, and achieves exceptional speed and accuracy for fitting extended LGMs with replicates. In spatial applications, this method essentially consists of the two following consecutive steps: first, in the “Max” step, maximum likelihood estimates of model parameters and their observed information matrices are computed at each site separately; then, in the “Smooth” step, these parameter estimates are smoothed jointly using an approximate LGM where the likelihood function has been approximated with a Gaussian density function, while properly accounting for the parameter uncertainty from the first step. Unlike other MCMC schemes for extremes, the Gaussian–Gaussian conjugacy of this approximate LGM can be exploited to improve the mixing of Markov chains and reduce the computational burden drastically.

In this book chapter, we showcase the modeling and inference approach proposed by Johannesson et al. (2021) on a new high-resolution precipitation dataset, but we make a crucial modification: instead of modeling block maxima with the GEV distribution, we here exploit the more convenient and informative framework of threshold exceedances. This allows us to exploit all information available from the upper tail for inference, thus reducing the overall estimation uncertainty. Moreover, similarly to Coles and Casson (1998) but unlike Cooley et al. (2007), we here use the Poisson point process likelihood instead of the GP likelihood, which allows us to model the three marginal parameters jointly (rather than having to fit two separate models to estimate the GP parameters and the threshold exceedance probability), and to avoid an overly strong influence of the threshold choice on the results. We also investigate Max-and-Smooth and the accuracy of the Gaussian likelihood approximation in this context, and demonstrate its usefulness in practice.

For illustration, we consider daily precipitation (mm) data, obtained from the Tropical Rainfall Measuring Mission (TRMM, Version 7) available over the period 2000–2019 without missing values, at a spatial resolution of 0.25° × 0.25° over Saudi Arabia; the dataset is available at https://gpm.nasa.gov/data-access/downloads/trmm. Saudi Arabia has a diverse geography (Arabian desert, steppes, mountain ranges, volcanic lava fields, etc.) and has, for the most part, a hot desert climate with very high daytime temperatures during the summer and a sharp temperature drop at night, with the exception of the south-western region, which features a semi-arid climate. Although the annual precipitation is very low overall, certain regions of Saudi Arabia have been regularly affected over the last two decades by short but intense convective storms, causing devastating flash-floods, extensive damage and fatalities (Deng et al., 2015; Yesubabu et al., 2016). Considering the whole of Saudi Arabia, our dataset comprises 2738 grid cells, resulting in approximately 20 million spatio-temporal observations in total. Figure 1 shows spatial maps of the observations for two extreme days: the day with the highest spatial average precipitation, and the day with the highest daily precipitation localized on a single grid cell. In both cases, these days fall within
Figure 1: Spatial maps of daily precipitation (mm) for the day with the highest spatial average (left), and the day with the highest daily precipitation at a single grid cell (right).

the month of April. In the left panel, the grid cells with higher daily precipitation amounts are mainly stretched between the latitudes 18°N and 26°N. In the right panel, we observe that the grid cells with high precipitation amounts are localized near the Asir mountains along the south-western coastline of the Red Sea, although the precipitation amount drops quickly away from the “wettest” pixel, indicating that tail dependence is relatively short-range. Further exploratory analysis (not shown) reveals that there is no discernible long-term trend across years, but there is a clear intra-year seasonal pattern that varies across space. While most of the grid cells receive the highest precipitation amounts in April, a large fraction of the northern part of the country, and the eastern coastline near the Persian Gulf receive the highest precipitation amounts in November. For simplicity, in this work, we ignore any temporal trend and dependence, and assume that the observations at each grid cell are independent and identically distributed across time. Modeling spatially-varying seasonality patterns and time dependence is out of the scope of this chapter, but is of practical relevance and deserves further investigation in future research. Our main goal here is to illustrate the usefulness of extended LGMs for extreme-value analysis in a high-resolution peaks-over-threshold context, when combined with the powerful Max-and-Smooth inference method.

The rest of this chapter is organized as follows. In Section 2, we provide some background information on univariate extreme-value theory and the Poisson point process formulation of extremes. In Section 3, we specify our proposed spatial extended LGM in detail and describe the SPDE approach for modeling spatial random effects. In Section 4, we describe the Max-and-Smooth method of inference and demonstrate its accuracy in our peaks-over-threshold context. In Section 5, we present the results from our Saudi Arabian precipitation application. We finally conclude in Section 6 with some discussion and perspectives on future research.
2 Univariate extreme-value theory background

Assume that \( Y_1, Y_2, \ldots, \overset{\text{iid}}{\sim} F_Y \) is a sequence of independent and identically distributed (iid) random variables with distribution \( F_Y \), and let \( M_n = \max\{Y_1, \ldots, Y_n\} \). The variables \( Y_i \) can be thought of as daily precipitation measurements observed over time at a single site, while \( M_n \) may represent the annual maximum precipitation amount. According to Fisher–Tippett Theorem, if there exist sequences of constants \( a_n > 0 \) and \( b_n \in \mathbb{R} \) such that, as \( n \to \infty \),

\[
\Pr\{(M_n - b_n)/a_n \leq z\} \to G(z),
\]

for some non-degenerate cumulative distribution function \( G \), then \( G \) is necessarily a generalized extreme-value (GEV) distribution, which may be expressed as

\[
G(z) = \begin{cases} 
\exp \left[ -\left\{ 1 + \xi (z - \mu)/\sigma \right\}^{-1/\xi} \right], & \text{if } \xi \neq 0, \\
\exp \left[ -\exp \left\{ -(z - \mu)/\sigma \right\} \right], & \text{if } \xi = 0,
\end{cases}
\]

defined on \( \{ z \in \mathbb{R} : 1 + \xi (z - \mu)/\sigma > 0 \} \), where \( \mu \in \mathbb{R}, \sigma > 0, \) and \( \xi \in \mathbb{R}, \) are location, scale, and shape parameters, respectively, and \( a_+ = \max\{a, 0\} \). Return levels are then simply defined as high quantiles of \( G \), i.e.,

\[
G^{-1}(1-p) = \begin{cases} 
\mu - \sigma [1 - \left\{ -\log(1-p) \right\}^{-\xi}] / \xi, & \text{if } \xi \neq 0, \\
\mu - \sigma \log\left\{ -\log(1-p) \right\}, & \text{if } \xi = 0,
\end{cases}
\]

for small probabilities \( p \). If the GEV distribution is used as a model for yearly block maxima, then the \( M \)-year return level \( z_M \) is obtained by setting \( p = 1/M \) in (3), i.e., \( z_M = G^{-1}(1 - 1/M) \). In other words, the \( M \)-year return level is the value that is expected to be exceeded once every \( M \) years on average, under temporal stationarity. The most important model parameter in (2), as far as the estimation of return levels is concerned, is thus the shape parameter \( \xi \). When \( \xi < 0 \), the GEV distribution \( G \) is the reverse Weibull distribution with a bounded upper tail. When \( \xi = 0 \), \( G \) is the Gumbel distribution with a light tail. When \( \xi > 0 \), \( G \) is the Fréchet distribution with a heavy tail. Therefore, it is crucial to accurately estimate \( \xi \) when extreme quantiles need to be computed, and in the spatial context, it is important to borrow strength across neighboring locations to reduce the estimation uncertainty and improve estimates of \( \xi \) and high quantiles.

The convergence in (1) is the theoretical justification for using the GEV distribution to model block maxima in practice. Replacing the limit in (1) by an equality, the normalizing constants \( a_n \) and \( b_n \) may then be absorbed into the location and scale parameters \( \mu \) and \( \sigma \) of the limiting distribution \( G \), and the GEV distribution (2) can then be fitted directly to block maxima \( M_n \), for large block sizes \( n \). However, in practice, this approach may be wasteful of data since other extreme observations may be disregarded if they are smaller than the block maximum. In geo-environmental applications, the block size \( n \) is often naturally considered.
to be one year to avoid the intricate modeling of seasonality. Choosing \( n \) to be one month may also be an option, though seasonality then needs to be modeled and this block size may not be large enough to justify using an asymptotic extreme-value model. A convenient alternative approach is to rely on asymptotic models for high threshold exceedances.

Under the assumption that (1) holds, Pickands–Balkema–de Haan Theorem implies that the distribution of \( Y - u \mid Y > u \) may be approximated, for large thresholds \( u \), by the generalized Pareto (GP) distribution \( H(y) \), defined as

\[
H(y) = \begin{cases} 
1 - (1 + \frac{\xi y}{\kappa_u})^{-1/\xi} & \text{if } \xi \neq 0, \\
1 - \exp\left(-\frac{y}{\kappa_u}\right) & \text{if } \xi = 0, 
\end{cases}
\] (4)

where \( \kappa_u > 0 \) is a threshold-dependent scale parameter and \( \xi \in \mathbb{R} \) is the same shape parameter as above. In practice, this result justifies fitting the GP distribution (4) to extreme observations that exceed a high pre-determined threshold \( u \). However, to get a complete description of the tail behavior, this model needs to be supplemented by a Bernoulli model for the threshold exceedance indicators, i.e., \( \mathbb{I}(Y_i > u) \), to estimate the threshold exceedance probability \( \zeta_u = \Pr(Y > u) \). Estimating the model parameters \( \kappa_u, \xi \) and \( \zeta_u \) with two separate models for \( Y_i - u \mid Y_i > u \) and \( \mathbb{I}(Y_i > u) \) may sometimes be inconvenient. Moreover, it can be shown that the choice of threshold \( u \) has an effect on the value of the scale parameter \( \kappa_u \) in (4), which makes it less easily interpretable, especially in presence of covariates.

To circumvent these issues, it is convenient to rely on the Poisson point process representation of extremes (Davison and Smith, 1990), which naturally connects the GEV and GP asymptotic distributions. Consider the two-dimensional point process

\[
\left\{ \left( \frac{i}{n + 1}, Y_i - b_n \right) \cdot \right; i = 1, \ldots, n \right\},
\] (5)

where \( a_n \) and \( b_n \) may be chosen as in (2). Then, assuming (1) holds, the point process (5) converges, in the sense of finite-dimensional distributions, to a Poisson point process with mean measure

\[
\Lambda([t_1, t_2] \times (y, \infty)) = (t_2 - t_1)\left\{ 1 + \frac{\xi(y - \mu)}{\sigma} \right\}^{-1/\xi},
\] (6)

for suitable regions of the form \( A = [t_1, t_2] \times (y, \infty) \) with \( 0 \leq t_1 < t_2 \leq 1 \) and \( y > u_L \) for some limiting lower bound \( u_L \), and where \( \mu, \sigma \) and \( \xi \) are as in the GEV parametrization (2). It can indeed be shown that the GEV representation (2) and GP representation (4) can both be seen as a consequence of the Poisson point process convergence result. The intensity function corresponding to the mean measure (6) can be obtained by differentiation, i.e., \( \lambda(t, y) = -\frac{d^2}{dy^2} \Lambda([t_0, t] \times [y, \infty)) = \sigma^{-1}\left\{ 1 + \frac{\xi(y - \mu)}{\sigma} \right\}^{-1/\xi - 1} \), for \( 0 \leq t_0 < t \leq 1 \). In practice, this limiting Poisson process can be fitted to the original, non-normalized points \( \{i/(n+1), Y_i\}; i = 1, \ldots, n \} \) since the constants \( a_n \) and \( b_n \) may be absorbed into the location
and scale parameters. The corresponding likelihood function for fitting this process to data observed in the region $A_u = [0, 1] \times (u, \infty)$, for some high threshold $u$, is

$$L(\mu, \sigma, \xi; A_u) = \exp \{ -\Lambda(A_u) \} \prod_{i=1}^{N_u} \lambda \{ i/(n + 1), Y_{(n-i+1)} \}$$

$$\propto \exp \left\{ -n_{\text{block}} \left( 1 + \frac{\xi}{\sigma} \frac{u - \mu}{\sigma} \right)^{-1/\xi} \right\} \prod_{i=1}^{N_u} \frac{1}{\sigma} \left( 1 + \xi \frac{Y_{(n-i+1)} - \mu}{\sigma} \right)^{-1/\xi - 1},$$

where $Y_{(1)} < \cdots < Y_{(n)}$ are the order statistics and $N_u$ is the (random) number of observations $Y_i$ exceeding the threshold $u$, the parameters $\mu$, $\sigma$ and $\xi$ are as before, and $n_{\text{block}}$ may be chosen to rescale the intensity function in order that the interpretation of the parameters $\{\mu, \sigma, \xi\}$ estimated from (7) matches that obtained from block maxima in (2) for some desired block size; see Coles (2001). For example, if $n_{\text{block}}$ is equal to the number of years of observations, then the parameters $\{\mu, \sigma, \xi\}$ estimated from (7) will theoretically correspond to those that would be obtained by fitting the GEV distribution to yearly maxima. The benefits of the likelihood inference approach based on (7) is that, unlike the likelihood based on the GP distribution (4), all three marginal parameters $\{\mu, \sigma, \xi\}$ are modeled at once, thus facilitating uncertainty assessment of return levels (which are a function of all three parameters), and that parameter estimates are (relatively) invariant to the chosen threshold $u$. Moreover, there is a direct correspondence between the GEV and Poisson point process parameterizations, which eases interpretation, and return levels can readily be obtained using (3).

In the frequentist framework, we can estimate the model parameters from an iid dataset $Y_1, \ldots, Y_n$, by directly maximizing the log-likelihood corresponding to (7) numerically, thus obtaining maximum likelihood estimates (MLEs) of $\mu$, $\sigma$, and $\xi$. Alternatively, to stabilize the estimation of the shape parameter $\xi$, which is often subject to high uncertainty, and thus to avoid unrealistic estimates, we may penalize certain values of $\xi$ through an additional prior $\pi(\xi)$. In the context of flood data modeling, Martins and Stedinger (2000) proposed using a beta density shifted to the interval $(-0.5, 0.5)$, with mean 0.10 and standard deviation 0.122. Johannesson et al. (2021) instead used a symmetric beta prior, also shifted to the interval $(-0.5, 0.5)$, but with mean zero and standard deviation 0.167. Once the prior $\pi(\xi)$ is specified, we can maximize the “generalized likelihood function”, defined as the product of the actual likelihood function (7) and the prior $\pi(\xi)$, thus providing robust parameter estimates. In our Saudi Arabian precipitation application, we consider the same beta prior for $\xi$ as in Johannesson et al. (2021), i.e., we take it to be a symmetric Beta(4, 4) density over $(-0.5, 0.5)$. The main reason for choosing this prior is that it avoids excessively small or large estimates of $\xi$ (thus preventing overly short tails when $\xi < 0.5$, as well as infinite-variance models when $\xi > 0.5$), even when there are only a limited number of threshold exceedances. Henceforth, estimates obtained by maximizing the generalized likelihood function will simply be referred to as MLEs. As explained in Section 4, obtaining MLEs and a reliable estimate
Figure 2: Maximum likelihood estimates of the spatial Poisson point process parameters $\mu_i$, $\sigma_i$, and $\xi_i$ (left), and their transformations $\psi_i$, $\tau_i$, and $\phi_i$ (right), respectively.

of their observed information matrix, at each site separately, is required to perform fully Bayesian inference with Max-and-Smooth. The left column of Figure 2 shows MLEs (using the additional prior $\pi(\xi)$) obtained by fitting, separately at each location, the limiting Poisson point process model to extreme precipitation peaks exceeding the site-specific 75% empirical quantile of positive precipitation intensities (i.e., excluding the zeros). When accounting for zero precipitation values, this threshold varies spatially and corresponds to the 90.16%–99.74% marginal quantile level depending on each location, which gives an average of 66
threshold exceedances per location (minimum 19, interquartile range 40–70, maximum 713 threshold exceedances). As expected, values of the location parameter \( \mu \) and the scale \( \sigma \) are higher in the south-western part of Saudi Arabia. Estimates of \( \xi \) tend to be slightly positive overall, but the spatial pattern is much more chaotic and noisy, with values of \( \xi \) ranging from \(-0.2\) to \(0.4\). From these plots, it is clear that a spatial model appropriately smoothing values of \( \xi \) (and \( \mu \) and \( \sigma \), as well) would give more sensible results and improve return level estimation. In the next section, we describe our spatial LGM framework to achieve this.

3 Latent Gaussian modeling framework

We now explain the general framework of latent Gaussian models (LGMs) and illustrate it here in the spatial setting of our Saudi Arabian precipitation application, although LGMs can be applied much more generally in other contexts (see, e.g., Rue et al., 2009; Hrafnkelsson et al., 2021). As mentioned in the introduction, an LGM consists of three hierarchical levels (the data level, the latent level, and the hyperparameter level), which are detailed below.

3.1 Data level specification

For simplicity, we assume that there are no missing values and that the number of observations is the same at each location. Let \( Y_t(s) \) denote the daily precipitation process over Saudi Arabia, and we denote by \( y_{it} \) the observed amount on day \( t \in \{1, \ldots, T\} \) at spatial location \( s_i \in \mathcal{D} = \{s_1, \ldots, s_N\} \), where \( T \) and \( N \) denote the total number of temporal replicates per location and the number of locations, respectively. Let \( y_i = (y_{i1}, \ldots, y_{iT})^\top \) be the vector containing all observations at location \( s_i \), and let \( y = (y_1^\top, \ldots, y_N^\top)^\top \) be the combined vector of all observations. At the data level of the Bayesian hierarchical model, we describe stochastic fluctuations using a parametric family \( \pi(y_{|\varphi_1, \ldots, \varphi_K}) \) that depends on \( K \) parameters. Assuming that the observations \( y_{it} \) have density \( \pi(\cdot | \varphi_{1i}, \ldots, \varphi_{Ki}) \), and are conditionally independent (across both space and time) given spatially-varying parameters \( \varphi_1 = (\varphi_{11}, \ldots, \varphi_{1N})^\top, \ldots, \varphi_K = (\varphi_{K1}, \ldots, \varphi_{KN})^\top \), we can then write the probability density function of \( y \), conditional on \( \varphi_1, \ldots, \varphi_K \), as

\[
\pi(y \mid \varphi_1, \ldots, \varphi_K) = \prod_{i=1}^{N} \prod_{t=1}^{T} \pi(y_{it} \mid \varphi_{1i}, \ldots, \varphi_{Ki}).
\tag{8}
\]

Hereafter, we refer to the density \( \pi(y \mid \varphi_1, \ldots, \varphi_K) \) as the likelihood function, when viewed as a function of the parameters \( \varphi_1, \ldots, \varphi_K \). In our spatial extremes context based on peaks-over-threshold, the likelihood in (8) is simply constructed by multiplying sitewise Poisson point process likelihoods of the form (7), with \( K = 3 \) and \( \varphi_1 \equiv \mu \), \( \varphi_2 \equiv \sigma \), and \( \varphi_3 \equiv \xi \). Thus, we assume that spatial variability among marginal extremes may be captured through the underlying spatially-varying Poisson point process parameters \( \varphi_1 \equiv \mu = (\mu_1, \ldots, \mu_N)^\top \),
\( \varphi_2 \equiv \sigma = (\sigma_1, \ldots, \sigma_N)^\top \), and \( \varphi_3 \equiv \xi = (\xi_1, \ldots, \xi_N)^\top \), though in practice some of these parameters may be assumed to have a more parsimonious formulation. The spatial structure of parameters is specified at the latent level.

### 3.2 Latent level specification, and multivariate link function

At the latent level, we first suitably transform parameters and then model them through spatially-structured and/or unstructured Gaussian model components. To be more precise, let \( g : \Omega \to \mathbb{R}^K \) be a \( K \)-variate bijective link function that transforms the original parameters (with support in \( \Omega \subset \mathbb{R}^K \)) as \( g(\varphi_1, \ldots, \varphi_K) = (\eta_1, \ldots, \eta_K)^\top \), in such a way that the domain of each transformed parameter \( \eta_j, j = 1, \ldots, K \), is the whole real line. At each location, the original parameters \( (\varphi_{1i}, \ldots, \varphi_{Ki})^\top \) are thus transformed through \( g \) as \( (\eta_{1i}, \ldots, \eta_{Ki})^\top = g(\varphi_{1i}, \ldots, \varphi_{Ki}), i = 1, \ldots, N \), and we can then combine them across locations like above as \( \eta_1 = (\eta_{11}, \ldots, \eta_{1N})^\top, \ldots, \eta_K = (\eta_{K1}, \ldots, \eta_{KN})^\top \). The link function can be chosen to “Gaussianize” the behavior of parameters, or to ensure they reflect certain desired properties (e.g., having their support on the whole real line), or to reduce confounding issues between latent parameters if their estimates appear to be strongly correlated. Once the link function is specified, the general formulation of the latent model, which consists in modeling the transformed parameters using fixed and random effects, may be expressed as

\[
\begin{align*}
\eta_1 &= X_1 \beta_1 + u_1 + \varepsilon_1, \\
\eta_2 &= X_2 \beta_2 + u_2 + \varepsilon_2, \\
&\vdots \\
\eta_K &= X_K \beta_K + u_K + \varepsilon_K,
\end{align*}
\]

where \( \beta_1, \ldots, \beta_K \) are fixed (covariate) effects specified with independent Gaussian priors (see further details below), \( X_1, \ldots, X_K \) are the corresponding design matrices comprising observed covariates, \( u_1, \ldots, u_K \) are zero-mean spatially-structured Gaussian random effects, and \( \varepsilon_1, \ldots, \varepsilon_K \) are zero-mean unstructured Gaussian “noise” (i.e., everywhere-independent) effects capturing small-scale variability. All the fixed and random effects in (9) are assumed to be mutually independent. Moreover, although there are other possibilities, we shall here define the spatially-structured random effects \( u_j, j = 1, \ldots, K \), by discretizing a stochastic partial differential equation (SPDE) approximating a Matérn Gaussian field, which yields sparse precision matrices and thus fast inference.

In our spatial extremes context, we transform the parameters \( (\varphi_1, \varphi_2, \varphi_3)^\top \equiv (\mu, \sigma, \xi)^\top \) jointly using a multivariate link function. We therefore obtain three jointly transformed parameter vectors at the latent level, namely \( \psi = (\psi_1, \ldots, \psi_N)^\top \), \( \tau = (\tau_1, \ldots, \tau_N)^\top \), and \( \phi = (\phi_1, \ldots, \phi_N)^\top \), where \( (\eta_{1i}, \eta_{2i}, \eta_{3i})^\top \equiv (\psi_i, \tau_i, \phi_i)^\top = g(\mu_i, \sigma_i, \xi_i) \). Similar to Johannesson et al. (2021) our choice of link function is justified by the following arguments. Observing that
Figure 3: The transformation $h : (-0.5, 0.5) \rightarrow \mathbb{R}$ (continuous black curve). The reference line with intercept zero and slope one is shown as a dashed line.

the estimated location parameters $\hat{\mu}_i$ are all positive and right-skewed (see the top left panel of Figure 2), and that the location and scale parameters, $\hat{\mu}_i$ and $\hat{\sigma}_i$, are strongly linearly correlated, we transform them jointly as $\psi_i = \log(\mu_i)$, and $\tau_i = \log(\sigma_i/\mu_i)$, $i = 1, \ldots, N$. Moreover, as the estimation of the shape parameter is generally subject to high uncertainty, we follow Johannesson et al. (2021) and consider the transformation $\phi_i = h(\xi_i)$, where

$$\phi = h(\xi) = a_\phi + b_\phi \log[\frac{1}{\log(\frac{1}{\log(1 - (\xi + 0.5)c_\phi})}]$$

with $c_\phi = 0.8$, $b_\phi = -c_\phi^{-1}\log\{1 - 0.5c_\phi\}\{1 - 0.5c_\phi\}2^{-1} = 0.39563$, $a_\phi = -b_\phi\log[\frac{1}{\log(1 - 0.5c_\phi})] = 0.062376$. The corresponding inverse transformation can be easily computed as

$$\xi = h^{-1}(\phi) = (1 - \exp[-\exp\{(\phi - a_\phi)/b_\phi\}])^{1/c_\phi} - 0.5.$$

This specific transformation, displayed in Figure 3, prevents overly small and large values of $\xi$ by restricting its domain to the interval $(-0.5, 0.5)$; it conveniently ensures that $h(\xi) \approx \xi$ for $\xi \approx 0$, which implies that the transformed shape parameter $\phi$ can be interpreted similarly to $\xi$ in case of light or moderately heavy tails; and it also makes sure that the asymptotic variance of the MLE behaves reasonably; see Johannesson et al. (2021) for more details. Therefore, overall, we can write the chosen multivariate link function as $g : (0, \infty)^2 \times (-0.5, 0.5) \rightarrow \mathbb{R}^3$ with $g(\mu, \sigma, \xi) = (\log(\mu), \log(\sigma/\mu), h(\xi))^\top$. MLEs for the transformed parameters $(\psi_i, \tau_i, \phi_i)^\top = g(\mu_i, \sigma_i, \xi_i)$, $i = 1, \ldots, N$, are shown in the right column of Figure 2. We can see that the distribution of transformed parameters looks reasonably symmetric, spatially well-behaved, and that the very strong correlation between $\mu$ and $\sigma$ has been somewhat reduced in the transformed parameters. We can also see that there seems to be a relatively smooth underlying signal in the transformed parameters, but that the sitewise
estimates are somewhat noisy, especially as far as the shape parameter is concerned. This justifies modeling transformed parameters using a spatial latent model of the form (9).

We now detail more specifically each of the model components involved in the latent structure (9), in the context of our Saudi Arabian precipitation application. More general structures might be considered in other applications. In case relevant covariate information is available, it can be easily incorporated into the modeling of the transformed parameters $\psi$, $\tau$, and $\phi$, but we here do not have any meaningful covariates to use, and thus, we only consider an intercept with some additional spatial effect terms. Moreover, while the spatially-structured effects provide great flexibility in modeling latent parameters, we also need to be careful not to construct an over-complicated model that would be computationally demanding and unstable to fit. Hence, we first investigate whether or not the spatial effects $u_j$ in (9) are truly required to model spatial variability in each of the transformed parameters $\psi$, $\tau$, and $\phi$. To assess this, we compute binned empirical variograms for each of the spatial fields. Figure 4 shows the results for each transformed parameter. We can see that the variograms of $\psi$ and $\tau$ indicate long-range spatial dependence, while the variogram of $\phi$ does not give as much evidence of a strong spatial dependence, but rather seems to indicate that about 50–75% of the marginal variability is due to unstructured noise. This corroborates the initial impression given by Figure 2. Thus, here, we construct a relatively parsimonious spatial model, where both $\psi$ and $\tau$ are modeled using an intercept, a spatially-structured term, and a spatially-unstructured noise term, while the transformed shape parameter $\phi$ is modeled solely using an intercept and a spatially-unstructured noise term, but does not involve any spatially-structured term.

To model spatially-structured effects, we here consider the class of Gaussian processes driven by a Matérn correlation structure, which may be approximated based on finite-dimensional Gaussian Markov random fields (GMRFs). GMRFs are structured in terms of conditional independence relationships, which lead to sparse precision matrices that can be conveniently summarized with a graphical representation, and this can be exploited in practice to speed up computations; see Rue and Held (2005) for more details on GMRFs.
A direct link between continuous-space Gaussian processes with a dense Matérn correlation structure, and GMRFs with a sparse precision matrix, may indeed be established theoretically through the stochastic partial differential equation (SPDE) approach (Lindgren et al., 2011), and this is the model structure that we shall use here.

Therefore, in our application, we can rewrite the latent level (9) more specifically as

\begin{align}
\psi &= \beta_{\psi} \mathbf{1}_N + \mathbf{A} \mathbf{u}_{\psi}^* + \mathbf{e}_{\psi}, \\
\tau &= \beta_{\tau} \mathbf{1}_N + \mathbf{A} \mathbf{u}_{\tau}^* + \mathbf{e}_{\tau}, \\
\phi &= \beta_{\phi} \mathbf{1}_N + \mathbf{e}_{\phi},
\end{align}

(12)

where \( \mathbf{1}_N \) is an \( N \)-dimensional vector of ones representing the intercept, and \( \beta_{\psi}, \beta_{\tau}, \) and \( \beta_{\phi} \) are the corresponding coefficients, the vectors \( \mathbf{u}_{\psi}^* \) and \( \mathbf{u}_{\tau}^* \) are independent spatially-structured random effects defined on a triangulated mesh \( D^* \) covering the region of interest, the matrix \( \mathbf{A} \) in (12) is a projection matrix from \( D^* \) (mesh nodes) to \( D \) (data locations), and the vectors \( \mathbf{e}_{\psi}, \mathbf{e}_{\tau}, \) and \( \mathbf{e}_{\phi} \) are spatially-unstructured noise terms, i.e., nugget effects. To be more precise, the mesh \( D^* \) should be a relatively fine spatial discretization of the domain of interest, that we can construct using the function \texttt{inla.mesh.2d} from the package \texttt{INLA} (www.r-inla.org). Note that the number of mesh nodes and the mesh itself may be defined independently of the data locations, and its resolution should mainly depend on the effective correlation range of the process of interest. Typically, longer-range processes can be approximated using coarser meshes. Figure 5 displays the mesh that we used in our Saudi Arabian precipitation application. It contains 1196 mesh nodes in total, which is much less than the total number of grid cells (2738), thus saving computational time and memory, and it was found to yield satisfactory results. Having constructed the mesh, we can then define the spatially-structured random effects in (12) as \( \mathbf{u}_{\psi}^* \sim \text{Normal}_{|D^*|}(0, s_{\psi}^2 Q_{\rho_{\psi}}^{-1}) \) and \( \mathbf{u}_{\tau}^* \sim \text{Normal}_{|D^*|}(0, s_{\tau}^2 Q_{\rho_{\tau}}^{-1}) \), where \( |D^*| \) denotes the number of mesh nodes in \( D^* \), \( \text{Normal}_{|D^*|} \) denotes the \( |D^*| \)-variate normal distribution, \( s_{\psi}^2 \) and \( s_{\tau}^2 \) denote the marginal variances of \( \mathbf{u}_{\psi}^* \) and \( \mathbf{u}_{\tau}^* \), respectively, and \( Q_{\rho_{\psi}} \) and \( Q_{\rho_{\tau}} \) are sparse precision matrices defined through the Matérn SPDE–GMRF relationship in terms of some range parameters \( \rho_{\psi} \) and \( \rho_{\tau} \); see Lindgren et al. (2011) for mathematical details. Thus, the final correlation matrices of \( \mathbf{u}_{\psi} = \mathbf{A} \mathbf{u}_{\psi}^* \) and \( \mathbf{u}_{\tau} = \mathbf{A} \mathbf{u}_{\tau}^* \) are \( \mathbf{A}^T Q_{\rho_{\psi}}^{-1} \mathbf{A} \) and \( \mathbf{A}^T Q_{\rho_{\tau}}^{-1} \mathbf{A} \), respectively, and they are approximate Matérn correlation matrices with range parameters \( \rho_{\psi} \) and \( \rho_{\tau} \), respectively. Here, we fix the Matérn smoothness parameter to one, which produces reasonably smooth realizations justified by the variograms in Figure 4. As for the nugget effects \( \mathbf{e}_{\psi}, \mathbf{e}_{\tau}, \) and \( \mathbf{e}_{\phi} \) in (12), they are defined as zero-mean white Gaussian noises with marginal variance \( \sigma_{\psi}^2, \sigma_{\tau}^2, \) and \( \sigma_{\phi}^2 \), respectively, i.e.,

\begin{align}
\mathbf{e}_{\psi} &\sim \text{Normal}_N(0, \sigma_{\psi}^2 \mathbf{I}_N), \\
\mathbf{e}_{\tau} &\sim \text{Normal}_N(0, \sigma_{\tau}^2 \mathbf{I}_N), \\
\mathbf{e}_{\phi} &\sim \text{Normal}_N(0, \sigma_{\phi}^2 \mathbf{I}_N),
\end{align}

where \( \mathbf{I}_N \) is the identity matrix.
We can then rewrite the three equations in (12) more compactly in matrix form as

$$\eta = Z\nu + \epsilon,$$

where

$$\eta = (\psi^T, \tau^T, \phi^T)^T, \quad \nu = (\beta_\psi, u_\psi^T, \beta_\tau, u_\tau^T, \beta_\phi)^T, \quad \epsilon = (\epsilon_\psi, \epsilon_\tau, \epsilon_\phi)^T,$$

$$Z = \begin{pmatrix} 1_N & A & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1_N & A & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1_N \end{pmatrix},$$

where the dots denote null vectors/matrices of appropriate dimension. Based on this latent model specification, we write $\pi(\eta | \nu, \theta)$ for the multivariate Gaussian density of the transformed model parameters $\eta$ given the fixed and random effects $\nu$ and some hyperparameter vector $\theta$ (here, comprising the marginal standard deviations and range parameters of random effects). It follows that $\pi(\eta | \nu, \theta)$ is the density of a $\text{Normal}_{3N}(0, \text{bdiag}(\sigma^2_\psi I_N, \sigma^2_\tau I_N, \sigma^2_\phi I_N))$ distribution, where “bdiag” refers to a block diagonal matrix. Moreover, we write $\pi(\nu | \theta)$ for the multivariate Gaussian density of $\nu$ given $\theta$. Since the fixed and random effects are assumed to have independent zero-mean Gaussian priors, $\pi(\nu | \theta)$ is the density of a $\text{Normal}_{3+2|D^*}(0, \text{bdiag}(\sigma^2_{\beta_\psi}, s^2_\psi Q_{\beta_\psi}^{-1}, \sigma^2_{\beta_\tau}, s^2_\tau Q_{\beta_\tau}^{-1}, \sigma^2_{\beta_\phi}))$ distribution, where $\sigma^2_{\beta_\psi}, \sigma^2_{\beta_\tau}, \sigma^2_{\beta_\phi}$ are user-specified prior variances for intercept coefficients. Specifically, we here choose independent vague normal priors in our application, i.e., $\beta_l \sim \text{Normal}(0, \sigma^2_{\beta_l})$ with $\sigma^2_{\beta_l} = 100^2$, for each
\( l \in \{\psi, \tau, \phi\} \). The hyperparameter vector \( \theta \) is crucial in controlling the behavior of latent effects, in order to stabilize estimation while avoiding overly restrictive behaviors. This is controlled at the hyperparameter level by assigning a suitable prior density \( \pi(\theta) \) to \( \theta \). Details are included in the next section.

### 3.3 Hyperparameter level specification

At the hyperparameter level, we finally assign prior distributions to all hyperparameters (i.e., here, the vector \( \theta = (\sigma_\psi, s_\psi, \rho_\psi, \sigma_\tau, s_\tau, \rho_\tau, \sigma_\phi)^T \)). Additionally, we set a prior on the shape parameter \( \xi \) to stabilize its estimation, similar to Martins and Stedinger (2000) and Johannesson et al. (2021). In our precipitation application, we use the following specification:

- For the nugget standard deviations \( \sigma_\psi, \sigma_\tau, \sigma_\phi \), we use penalized-complexity (PC) priors (Simpson et al., 2017) that are parametrization-invariant prior distributions designed to prevent overfitting, by shrinking hyperparameters towards a basic reference model—here, a model without nugget effects. We can show that, in this case, the PC prior for \( \sigma_l \), \( l \in \{\psi, \tau, \phi\} \), is an exponential distribution with some user-defined rate parameter controlling how concentrated the prior for \( \sigma_l \) (and, thus, also the prior for the \( \epsilon_i \) noise term) is around zero; see Theorem 2.1 of Fuglstad et al. (2019). In other words, we set

\[
\pi(\sigma_l) = \lambda_{\sigma_l} \exp(-\lambda_{\sigma_l}\sigma_l), \quad \sigma_l > 0, \quad \text{for some rate parameter} \quad \lambda_{\sigma_l} > 0, \quad l \in \{\psi, \tau, \phi\}.
\]

- For the parameter vectors \((s_\psi, \rho_\psi)\)' and \((s_\tau, \rho_\tau)\)' of the spatially-structured Matérn random effects, we use the same PC prior as in Fuglstad et al. (2019), defined as

\[
\pi(s_l, \rho_l) = \lambda_{\rho_l} \lambda_{s_l} \rho_l^{-2} \exp(-\lambda_{s_l}s_l - \lambda_{\rho_l}\rho_l^{-1}), \quad s_l > 0, \quad \rho_l > 0, \quad l \in \{\psi, \tau\},
\]

where \( \lambda_{\rho_l}, \lambda_{s_l} > 0 \) are user-defined rate parameters, selected according to Theorem 2.6 of Fuglstad et al. (2019).

- For the shape parameter \( \xi = (\xi_1, \ldots, \xi_N)^T \), we consider independent Beta(4, 4) priors shifted to the interval \((-0.5, 0.5)\), which then induces a prior on the transformed shape parameter \( \phi = (\phi_1, \ldots, \phi_N)^T \). Given the transformation \( \phi = h(\xi) \) in (10) (with inverse transformation \( \xi = h^{-1}(\phi) \) in (11)), the prior density for each \( \phi_i \) may thus be expressed as

\[
\pi(\phi_i) = \frac{1}{B(4, 4)b_\phi c_\phi} \left\{ h^{-1}(\phi_i) + \frac{1}{2} \right\}^{4-c_\phi} \left\{ \frac{1}{2} - h^{-1}(\phi_i) \right\}^3 \exp \left\{ \frac{\phi_i - a_\phi}{b_\phi} - \exp \left\{ \frac{\phi_i - a_\phi}{b_\phi} \right\} \right\},
\]

where \( B(\cdot, \cdot) \) is the beta function, and \( a_\phi, b_\phi, c_\phi \) are specified in (10). Hence, we have

\[
\pi(\phi) = \prod_{i=1}^N \pi(\phi_i), \quad \text{with} \quad \pi(\phi_i) \quad \text{defined as above, i.e., a transformed Beta(4,4) density.}
\]

We then assume that the priors for \((s_\psi, \rho_\psi)^T\), \((s_\tau, \rho_\tau)^T\), \(\sigma_\psi, \sigma_\tau, \sigma_\phi\), and \(\phi\), are mutually independent. In particular, the prior for hyperparameters can be written as

\[
\pi(\theta) = \pi(\sigma_\psi) \times \pi(s_\psi, \rho_\psi) \times \pi(\sigma_\tau) \times \pi(s_\tau, \rho_\tau) \times \pi(\sigma_\phi).
\]
3.4 Summarized full model specification

In summary, the model structure is specified at three levels: the data level is specified by the density $\pi(y \mid \eta)$, which is the same as the likelihood function (8), but now expressed in terms of the transformed parameters $\eta = (\psi^T, \tau^T, \phi^T)^T$. Taking the additional transformed beta prior $\pi(\phi)$ into consideration, we obtain the generalized likelihood function

$$L(y \mid \eta) = \pi(y \mid \eta) \times \pi(\phi).$$

(13)

The latent level is specified by the multivariate normal density $\pi(\eta, \nu \mid \eta) = \pi(\eta \mid \nu, \theta) \times \pi(\nu \mid \theta)$, where each term $\pi(\eta \mid \nu, \theta)$ and $\pi(\nu \mid \theta)$ is a multivariate Gaussian density itself as made precise in Section 3.2. Finally, the hyperparameter level is specified by the prior density $\pi(\theta)$ detailed in Section 3.3.

Thus, by exploiting this hierarchical representation, the posterior density is

$$\pi(\eta, \nu, \theta \mid y) \propto L(\eta \mid y)\pi(\eta, \nu \mid \theta)\pi(\theta) = L(\eta \mid y)\pi(\eta \mid \nu, \theta)\pi(\nu \mid \theta)\pi(\theta).$$

(14)

In order to perform Bayesian inference, sampling or numerical approximation of the posterior density (14) is required. Usually, Markov chain Monte Carlo algorithms can be used to generate approximate samples from (14). However, the complicated form of the generalized likelihood function, expressed in terms of the Poisson point process likelihood (7), prevents Gibbs sampling to update model parameters, and the presence of multiple latent high-dimensional spatial random effects makes it computationally impractical. Alternative computational solutions need to be found. The integrated nested Laplace approximation (INLA) is a highly popular approximate Bayesian inference technique, but unfortunately, it does not apply, at least in its current implementation in R, to extended LGMs with a multivariate link function. In the next section, we describe Max-and-Smooth, which consists in approximating the generalized likelihood function $L(y \mid \eta)$ in (13) with a Gaussian density, to exploit the conjugacy of Gaussian–Gaussian LGMs for fast MCMC-based inference.

4 Approximate Bayesian inference with Max-and-Smooth

Max-and-Smooth is a fully Bayesian inference scheme designed to fit extended LGMs, and relies on two successive steps: 1) “Max” step: In the spatial setting, we first obtain parameter estimates independently at each site, as well as a robust estimate of their observed information matrix, and use them to suitably approximate the generalized likelihood function $L(\eta \mid y)$ in (14) with a Gaussian density; and 2) “Smooth” step: exploiting this likelihood approximation, we then smooth parameter estimates by fitting a Gaussian–Gaussian model that is a pseudo model for the exact extended LGM, using a straightforward MCMC algorithm, and treating the parameter estimates as “noisy data” with a known covariance
structure (obtained from the “Max” step). This two-step inference approach can be shown to be equivalent (up to the likelihood approximation) to fitting the original LGM in one single step. The quality of the Gaussian approximation will thus determine the method’s accuracy, and this implies that Max-and-Smooth requires to have enough temporal replicates at each site. Significant computational gains can be obtained with Max-and-Smooth thanks to the conjugacy of the Gaussian–Gaussian pseudo model, and also because the computational burden of the “Smooth” step is insensitive to the number of temporal replicates, unlike other ordinary MCMC or INLA methods. We now describe each step in our peaks-over-threshold extreme-value setting, and study the approximation’s accuracy in this context.

4.1 “Max” step: Computing MLEs and likelihood approximation

The first step of Max-and-Smooth is to obtain MLEs and to approximate the generalized likelihood function with a (rescaled) Gaussian likelihood. From (8) and (13), the generalized likelihood function can be written, thanks to the conditional independence assumption, as the product $L(\eta \mid y) = \prod_{i=1}^{N} L(\psi_i, \tau_i, \phi_i \mid y_i)$, where $L(\psi_i, \tau_i, \phi_i \mid y_i) = \pi(y_i \mid \psi_i, \tau_i, \phi_i)\pi(\phi_i)$ are sitewise generalized likelihoods and $y = (y_1, \ldots, y_N)$, $y_i = (y_{i1}, \ldots, y_{iT})^\top$, $i = 1, \ldots, N$. For each site $i = 1, \ldots, N$, we compute the MLE as $(\hat{\psi}_i, \hat{\tau}_i, \hat{\phi}_i)^\top = \arg\max_{(\psi_i, \tau_i, \phi_i)} L(\psi, \tau, \phi \mid y_i)$, and then approximate the $i$-th generalized likelihood function $L(\psi_i, \tau_i, \phi_i \mid y_i)$ by a possibly rescaled Gaussian density with mean $(\hat{\psi}_i, \hat{\tau}_i, \hat{\phi}_i)^\top$ and covariance matrix $\Sigma_{\eta,y_i} = Q_{\eta,y_i}^{-1}$, where $Q_{\eta,y_i}$ denotes the observed information matrix, i.e., the negative Hessian matrix of $\log L(\psi_i, \tau_i, \phi_i \mid y_i)$ evaluated at $(\hat{\psi}_i, \hat{\tau}_i, \hat{\phi}_i)^\top$. Therefore, $L(\psi_i, \tau_i, \phi_i \mid y_i) \approx c_i L(\psi_i, \tau_i, \phi_i \mid y_i)$, where $L(\psi_i, \tau_i, \phi_i \mid y_i)$ denotes the density of Normal$((\hat{\psi}_i, \hat{\tau}_i, \hat{\phi}_i)^\top, \Sigma_{\eta,y_i})$ and $c_i \geq 0$ is some nonnegative constant independent of $(\psi_i, \tau_i, \phi_i)^\top$. Combining all locations together, we can thus approximate the overall generalized likelihood function as $L(\eta \mid y) = \prod_{i=1}^{N} L(\psi_i, \tau_i, \phi_i \mid y_i) \approx \prod_{i=1}^{N} c_i L(\psi_i, \tau_i, \phi_i \mid y_i) \propto \prod_{i=1}^{N} L(\psi_i, \tau_i, \phi_i \mid y_i) := \hat{L}(\eta \mid y)$.

To clarify this approximation with a simple example, let us consider the setting where $W_1, \ldots, W_m \overset{\text{iid}}{\sim} \text{Normal}(\gamma, 1)$. The likelihood function for $\gamma$ is $L(\gamma \mid W_1, \ldots, W_m) = \prod_{k=1}^{m}[(2\pi)^{-1/2}\exp\{-\frac{1}{2}(W_k - \gamma)^2\}]$, which may be rewritten as a product of two terms, namely $c = m^{-1/2}(2\pi)^{-(m-1)/2}\exp\{-\frac{1}{2}(m^{-1}\sum_{k=1}^{m} W_k^2 - \overline{W}^2)\}$ and $m^{1/2}(2\pi)^{-1/2}\exp\{-\frac{1}{2}m(\gamma - \overline{W})^2\}$, where $\overline{W} = m^{-1}\sum_{k=1}^{m} W_k$. Here, the first term $c$ is a constant with respect to $\gamma$, whereas the second term is the normal density with mean $\overline{W}$, the MLE for $\gamma$, and variance $m^{-1}$; we can easily show that the negative Hessian of $\log L(\gamma \mid W_1, \ldots, W_m)$ is indeed $m$. In this toy example, we consider the data distribution to be normal, and hence, the normal approximation holds as an equality. In more general settings with alternative non-Gaussian likelihoods, the normal approximation is justified thanks to the large-sample properties of the MLE; see Schervish (1995) for example. Given that the normal density is the asymptotic form of the likelihood under mild regularity conditions, this ensures that this likelihood approximation will be accurate, provided the number of temporal replicates is large enough.
In our extreme-value context, \( \hat{L}(\eta \mid y) \) is the density of a \( 3N \)-dimensional normal distribution with mean \( \hat{\eta} = (\hat{\psi}^T, \hat{\tau}^T, \hat{\phi}^T)^T \), where \( \hat{\psi} = (\hat{\psi}_1, \ldots, \hat{\psi}_N)^T \), \( \hat{\tau} = (\hat{\tau}_1, \ldots, \hat{\tau}_N)^T \) and \( \hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_N)^T \), and covariance matrix \( \Sigma_{\eta,y} = Q_{\eta,y}^{-1} \) constructed by properly stacking the entries of \( \Sigma_{\eta,y}, \ldots, \Sigma_{\eta,y} \). The full posterior density (14) may then be approximated based on \( \hat{L}(\eta \mid y) \) as

\[
\pi(\eta, \nu, \theta \mid y) \approx \pi(\eta, \nu, \theta \mid y) \propto \hat{L}(\eta \mid y)\pi(\eta)\pi(\nu)\pi(\theta).
\]

(15)

Now, consider a pseudo LGM where the sitewise MLEs \( \hat{\eta} \) are treated as the data, with distribution \( \hat{\eta} \sim \text{Normal}_{3N} (\eta, Q_{\eta,y}^{-1}) \), where \( Q_{\eta,y} \) is defined as above and assumed to be known. Through the lenses of this pseudo model, we can interpret the vector \( \hat{\eta} \) as a noisy measurement of the true unknown parameter vector \( \eta \). While the likelihood of the pseudo model is different from the likelihood of the original model, the latent and hyperparameter levels are kept the same. Writing the likelihood function of the pseudo model as \( L^*(\eta \mid \hat{\eta}) = \pi(\hat{\eta} \mid \eta) \), the corresponding posterior density may be written as

\[
\pi(\eta, \nu, \theta \mid \hat{\eta}) \propto \pi(\hat{\eta} \mid \eta)\pi(\eta \mid \nu, \theta)\pi(\nu)\pi(\theta) = L^*(\eta \mid \hat{\eta})\pi(\eta \mid \nu, \theta)\pi(\nu \mid \theta)\pi(\theta),
\]

(16)

where \( L^*(\eta \mid \hat{\eta}) = \hat{L}(\eta \mid y) \). Hence, the posterior density for the pseudo model in (16) is the same as the approximated posterior density in (15) that provides a good approximation to the exact posterior density in (14). This justifies using the pseudo model to make inference, instead of relying on the exact but more complicated posterior density.

To verify the accuracy of the Gaussian likelihood approximation in our peaks-over-threshold extreme-value context, Figure 6 displays the true normalized likelihood together with the Gaussian approximation at a representative grid cell (site 208) with 67 threshold exceedances (out of 268 positive precipitation intensities, for a total of 7245 temporal replicates). The likelihood approximation is extremely accurate for \( \psi_i \), very accurate for the transformed shape parameter \( \phi_i \), and reasonably accurate for \( \tau_i \), though for this parameter the approximate likelihood is slightly wider than the true likelihood. The results are expected
to improve quickly when a larger number of threshold exceedances become available, similarly to the findings of Johannesson et al. (2021) in the case of the GEV distribution, which is in line with asymptotic results (Schervish, 1995).

4.2 “Smooth” step: Fitting the Gaussian–Gaussian pseudo model

The second step of Max-and-Smooth is to fit the approximate pseudo model with (16) as its posterior, based on the sitewise MLEs \( \hat{\eta} \) (and their precision matrices \( Q_{\eta,y}^{-1} \)) pre-computed in the “Max” step. We here briefly describe some key aspects of MCMC-based inference for the pseudo model. The hierarchical representation of the pseudo model is

\[
\hat{\eta} | \eta \sim \text{Normal}_{3N}(\eta, Q_{\eta,y}^{-1}),
\]

\[
\eta | \nu, \theta \sim \text{Normal}_{3N}(Z\nu, \text{bdiag}(\sigma_{\psi}^2 I_N, \sigma_{\tau}^2 I_N, \sigma_{\phi}^2 I_N)),
\]

\[
\nu | \theta \sim \pi(\nu | \theta) = \pi(\beta_{\psi}) \times \pi(\beta_{\tau}) \times \pi(u_{\psi} | \theta) \times \pi(u_{\tau} | \theta),
\]

\[
\theta \sim \pi(\theta) = \pi(\sigma_{\psi}) \times \pi(s_{\psi}, \rho_{\psi}) \times \pi(\sigma_{\tau}) \times \pi(s_{\tau}, \rho_{\tau}) \times \pi(\sigma_{\phi}),
\]

where the terms involved at the latent and hyperparameter levels are detailed in Sections 3.2 and 3.3, respectively, and the notation is the same as earlier.

Thanks to the conjugate structure of the Gaussian–Gaussian pseudo LGM, it is easy to verify that the full conditional density of \((\eta^\top, \nu^\top)\)^\top, i.e., \(\pi(\eta, \nu | \hat{\eta}, \theta)\), is multivariate Gaussian. While this density is \((5N+3)\)-dimensional, we can exploit the sparsity of its precision matrix (defined in terms of \(Q_{\eta,y}, Q_{\rho_{\psi}}\) and \(Q_{\rho_{\tau}}\)) for fast sampling. Notice that Gibbs sampling updates for latent parameters would not be possible for the original LGM based on the exact Poisson point process likelihood, and that Metropolis–Hastings updates (or variants thereof) would be required for all of the \(5N+3\) variables. The great benefit of the pseudo model is that latent variables can be updated simultaneously by Gibbs sampling, which drastically reduces the computational burden and greatly improves the mixing and convergence of Markov chains. As for the hyperparameter vector \(\theta\), its (approximate) marginal posterior density, i.e., \(\pi(\theta | \hat{\eta})\), is known up to a constant. Although \(\pi(\theta | \hat{\eta})\) does not have a closed form, posterior samples of \(\theta\) can be easily obtained using the Metropolis–Hastings algorithm or grid sampling, for example. The fact that the exact marginal posterior density of \(\theta\), i.e., \(\pi(\theta | y)\), can be approximated with the marginal posterior of \(\theta\) under the pseudo model is another important aspect of Max-and-Smooth that greatly improves posterior sampling.

A crucial point to note is that once sitewise MLEs \(\hat{\eta}\) are obtained, the rest of the computational time, required to fit the pseudo model in the “Smooth” step, does not depend on the temporal dimension anymore. Moreover, obtaining MLEs in the “Max” step is often quite fast, and can be performed in parallel across spatial locations. As an illustration, the “Max” step takes only a few seconds for our Saudi Arabian precipitation dataset with 2738 grid cells in total. Thus, Max-and-Smooth is doubly beneficial for spatio-temporal datasets with large
Table 1: Posterior summary statistics for the estimated hyperparameters.

| Hyperparameter | Mean | SD  | 2.5% | 50% | 97.5% |
|----------------|------|-----|------|-----|-------|
| $\beta_\psi$   | 2.6545 | 0.3454 | 1.9755 | 2.6499 | 3.3576 |
| $\sigma_\psi$  | 0.0442 | 0.0025 | 0.0392 | 0.0442 | 0.0491 |
| $s_\psi$       | 0.5956 | 0.1347 | 0.4125 | 0.5690 | 0.9562 |
| $\rho_\psi$    | 8.1123 | 2.0028 | 5.3706 | 7.7501 | 13.4077 |
| $\beta_\tau$   | -0.5519 | 0.2340 | -1.0185 | -0.5543 | -0.0745 |
| $\sigma_\tau$  | 0.0028 | 0.0025 | 0.0001 | 0.0020 | 0.0095 |
| $s_\tau$       | 0.3795 | 0.1007 | 0.2546 | 0.3587 | 0.6380 |
| $\rho_\tau$    | 8.3504 | 2.4316 | 5.2165 | 7.8593 | 14.3888 |
| $\beta_\phi$   | 0.0973 | 0.0020 | 0.0932 | 0.0973 | 0.1012 |
| $\sigma_\phi$  | 0.0593 | 0.0017 | 0.0559 | 0.0593 | 0.0626 |

temporal dimensions: first, the Gaussian approximation becomes very accurate; second, the relative computational cost with respect to an “exact” MCMC inference scheme becomes negligible. We also note that although the inference is done in two steps, the uncertainty involved in the “Max” step is properly propagated into the “Smooth” step in a way that provides a valid approximation to the (exact) full posterior distribution.

5 Saudi Arabian precipitation extremes application

In this section, we fit the latent Gaussian model of Section 3 to the Saudi Arabian TRMM daily precipitation dataset and draw approximate Bayesian inference using the Max-and-Smooth approach presented in Section 4.

The model is fitted to high threshold exceedances, based on the Poisson point process likelihood, using site-specific thresholds taken as the 75% empirical quantile of positive precipitation intensities. When taking zero precipitation values into account, the unconditional threshold probability level varies spatially (from about 90.16% at a few locations to about 99.74%), and it is often higher than the 99% quantile. On the scale of daily precipitation, the chosen thresholds range from about 1mm to 15mm, depending on the location. Fine-tuning the threshold at each site would be tedious with such a large spatial dimension, and so we opted to choose the threshold in a pragmatic way to keep a reasonable number of threshold exceedances at each site (from 19 to 713 for an average of 66 threshold exceedances, i.e., about 3.3 per year on average), while providing a suitable bias-variance trade-off, and making sure that the MLEs and their covariance matrices obtained in the “Max” step are reliable.

In the “Smooth” step, we run the Gaussian–Gaussian pseudo LGM for 10000 iterations, and remove 2000 burn-in iterations. Some posterior summary statistics for the estimated hyperparameters are presented in Table 1. All the intercept coefficients $\beta_\psi$, $\beta_\tau$ and $\beta_\phi$ are significantly different from zero based on 95% credible intervals, and are positive for $\psi$ and
Figure 7: Posterior means of the spatially-structured SPDE effects, $u^*_\psi$ (left) and $u^*_\tau$ (right), included in the latent structure for $\psi$ and $\tau$, respectively, projected to a fine grid covering the mesh in Figure 5.

$\phi$ but negative for $\tau$. The value of $\beta_\phi$ is about 0.1, and since the standard deviation $\sigma_\phi$ of the corresponding nugget effect is about 0.06, the estimated transformed shape parameters $\phi_i$ are usually within the range $[-0.02, 0.22]$. This implies that the precipitation distribution is moderately heavy-tailed at most sites. The marginal standard deviations of the spatially-structured SPDE effects, $s_\psi$ and $s_\tau$, are significantly larger than the respective standard deviations of the spatially-unstructured nugget effects, $\sigma_\psi$ and $\sigma_\tau$. This indicates that these transformed location and scale parameters vary quite smoothly over space, and thus, that it is important to include spatially-structured effects in these parameters at the latent level. Both range parameters $\rho_\psi$ and $\rho_\tau$ are fairly large and similar to each other, which indicates long-range spatial dependence and corroborates the empirical variograms in Figure 4.

Figure 7 displays the posterior mean of the spatially-structured SPDE random effects $u^*_\psi$ and $u^*_\tau$, included in the latent structure for $\psi$ and $\tau$, respectively, projected onto a fine grid covering the mesh shown in Figure 5. The spatial patterns in the estimates of $u^*_\psi$ and $u^*_\tau$ are similar to those observed in the maps of the MLEs, $\hat{\psi}_i$ and $\hat{\tau}_i$, plotted in Figure 2. Higher values of the posterior means of $u^*_\psi$ and $u^*_\tau$ are indeed observed near the south-western and south-eastern corners of the region of study, respectively. Since $u^*_\psi$ is involved in the linear model specification for $\psi_i$ and varies approximately between $-1$ and 1.5, the latent spatially-structured spatial effect scales the original location parameter $\mu_i = \exp(\psi_i)$ by a factor ranging from about 0.37 to 4.48. Similarly, the posterior mean of $u^*_\tau$ varies approximately between $-0.6$ and 0.6, which translates into a multiplicative factor for the original scale parameter $\sigma_i$ that ranges from about 0.55 to 1.82, after taking $\mu_i$ into account.

Figure 8 shows scatterplots of model-based estimates (i.e., posterior means) of the original and transformed model parameters plotted against their preliminary sitewise MLEs. We can see that the estimated location parameters $\hat{\mu}_i$ vary highly across space, with values rang-
Figure 8: Scatterplots of model-based posterior estimates of the original Poisson point process parameters (top), $\mu_i$, $\sigma_i$ and $\xi_i$ (left to right), and transformed parameters (bottom), $\psi_i$, $\tau_i$, and $\phi_i$ (left to right), plotted with respect to their corresponding preliminary sitewise MLEs. The main diagonal (with intercept zero and unit slope) is shown in red.

ing from about 5 to 60 depending on the location. Despite this high spatial variation, the model-based posterior estimates appear to be consistent with preliminary sitewise estimates of $\mu_i$ throughout the domain. This shows that the flexible LGM structure that we fitted is able to accurately capture complex spatial patterns in the location parameter, thanks to the latent spatially-structured and unstructured effects, while reducing the posterior uncertainty by pooling information across locations. Posterior estimates of the scale parameter $\sigma_i$ also appear to be generally consistent with preliminary sitewise estimates, although the points in the scatterplots are a bit more spread out around the main diagonal. This slightly larger variability is not an indication of a lack of fit, but it reflects the fact that sitewise estimates are quite noisy (mostly due to maximum likelihood estimation uncertainty), while the model-based posterior estimates are more smooth over space (thanks to the shrinkage induced by the spatial structure of the LGM). As for the shape parameter $\xi_i$, there are significant differences between the model-based posterior estimates (which vary roughly between 0 and 0.25) and the preliminary sitewise MLEs (which vary roughly between $-0.25$ and 0.40). These huge differences are due to two main reasons: first, the latent structure of $\phi$ in (12) does not involve a spatially-structured random effect, which makes it a bit less flexible than the
models for $\psi$ and $\tau$; second, the use of a PC prior for the standard deviation parameter $\sigma_\phi$ of the latent nugget effect induces relatively strong shrinkage towards a spatially constant shape parameter. Nevertheless, we believe that the posterior estimates of $\xi_i$ are much more reasonable than the preliminary sitewise MLEs, which have large uncertainties and yield unrealistic tail behaviors, from bounded upper tails when $\xi_i < 0$ to very heavy-tailed when $\xi_i \approx 0.4$. By contrast, our Bayesian LGM framework succeeds in reducing posterior uncertainty by shrinkage and spatial pooling of information, in order to obtain satisfactory results.

We then illustrate the practical benefit of the proposed LGM framework by computing return levels and posterior predictive densities. Return levels can be estimated by plugging model-based posterior estimates $\hat{\mu}_i, \hat{\sigma}_i$ and $\hat{\xi}_i$ into (3), whereas the posterior predictive density at a spatial location $s_i$ may be approximated as

$$
\pi_i^*(\tilde{y} | y) \approx \pi_i^*(\tilde{y} | \tilde{\eta}) = \int \int \int \pi_u(\tilde{y} | \eta_i) \pi(\eta, \nu, \theta | \tilde{\eta}) d\eta d\nu d\theta,
$$

where $\pi_u(\tilde{y} | \eta_i)$ is the conditional density of $\tilde{Y}$ given $\tilde{Y} > u_i$, with $\tilde{Y} \sim G_{\text{GEV}}^{1/B}(\mu_i, \sigma_i, \xi_i)$, and $G_{\text{GEV}}(\mu_i, \sigma_i, \xi_i)$ is the GEV distribution function with parameters $\mu_i$, $\sigma_i$, and $\xi_i$, $B$ is the block size (here, taken as $B = 365.25$ to represent yearly blocks), and $u_i$ is the threshold chosen for the $i$-th location. Figure 9 shows within-sample posterior predictive densities (top panels) and estimated return level plots (bottom panels) for three representative sites (site 208, 1716, and 1905) taken from three spatially distant parts of the domain. The top panels display kernel density estimates of posterior predictive samples for each selected site, obtained by sampling from (17). We observe that the densities are right-skewed for all three cases, and are relatively well calibrated with the observations. The bottom panels display estimated return levels plotted as a function of the return period, with associated 95% pointwise credible intervals, as well as the order statistics at each site. We can see that the observations are generally contained within, or just slightly outside, the credible bands, indicating that the proposed LGM fits the data reasonably well.

Finally, Figure 10 displays maps of $M$-year return level estimates corresponding to the return periods of $M = 20$, 50, and 100 years, as well as their respective sitewise posterior standard deviations. As expected, return level estimates are realistically higher near the coasts of the Red Sea and the Persian Gulf. They range from 16.7mm to 140.7mm for $M = 20$ years, from 21.5mm to 192.9mm for $M = 50$ years, and from 25.2mm to 242.9mm for $M = 100$ years, with the highest precipitation amount expected at the grid cell with coordinate ($39.875^\circ$E, $20.875^\circ$N), close to the major cities of Jeddah and Makkah. Furthermore, while the standard deviations are quite high near the coast of the Red Sea due to the large spread of the precipitation distribution and the high variability of threshold exceedances in this region, they are also quite high near the arid south-eastern region characterized by a drier climate, which leads to a smaller number of threshold exceedances available to fit the model in this
Figure 9: Within-sample posterior predictive densities at three representative grid cells (top), and return level estimates plotted as a function of the return period on a logarithmic scale (bottom). In the return level plots, black lines are the posterior means, blue bands are 95% pointwise credible intervals, and red dots are the ordered observations at each site.

area. Comparable results were obtained by Davison et al. (2019) after adjustment (as, in that paper, the authors mistakenly forgot a multiplication factor 3 when aggregating three-hourly precipitation rates (mm/h) to daily precipitation (mm/day)) for a smaller region around Jeddah and Makkah, based on max-stable processes fitted to annual precipitation maxima.

6 Discussion

In this chapter, we have shown how complex latent Gaussian models (LGMs) for extremes, based on the convenient and informative Poisson point process representation for peaks-over-threshold events, can be suitably constructed and efficiently fitted using Max-and-Smooth to massive spatio-temporal datasets. Our proposed modeling framework assumes that Gaussian fixed and spatial random effects are embedded within model parameters at the latent level, in order to flexibly capture spatial variability, non-stationary patterns, and dependencies. Our model relies on the stochastic partial differential equation (SPDE) representation of Matérn random fields, whose discretization leads to Gaussian Markov random fields (GMRFs) with
Figure 10: Spatial maps of $M$-year return level estimates (top), with $M = 20, 50, 100$ (left to right), and their corresponding posterior standard deviations (bottom).

sparse precision matrices. Exploiting GMRFs at the latent level, combined with an efficient Bayesian inference scheme makes it possible to tackle complicated problems in very high dimensions. The Saudi Arabian precipitation dataset that we analyzed in this chapter is proof of this, since it comprises 2738 grid cells and numerous temporal replicates at each site, and we did not reach our computational limits by any measure. Our proposed methodology indeed scales up to even bigger and higher-resolution datasets. Importantly, using latent SPDE random effects combined with Max-and-Smooth implies that the computational burden is only moderately impacted by the spatial and temporal dimensions. Computationally speaking, it is indeed crucial to realize that what matters the most is the number of mesh nodes used to discretize the SPDE, which should depend on the spatial effective correlation range and the size of the domain, rather than the actual number of observation locations. Moreover, the great benefit of Max-and-Smooth is to perform inference in two separate steps. The first step is typically very fast and consists in computing sitewise maximum likelihood estimates (MLEs), which can be done in parallel. The second step consists in fitting an approximate Gaussian–Gaussian pseudo LGM to the MLEs, so that the number of temporal replicates only \textit{indirectly} affects the model fit through the variability of MLEs but is irrelevant in terms of the computational time. Therefore, leveraging both Max-and-Smooth and SPDE random effects provides a powerful toolbox for analyzing massive and complex datasets using extended LGMs. We also stress that although we studied extreme precipitation data in this work, this methodology applies more generally and, if suitably adapted, could potentially be
used in other contexts and a variety of statistical applications (see Hrafnkelsson et al., 2021).

The use of high threshold exceedances, which contrasts with the block maximum approach advocated by Johannesson et al. (2021) and many other authors, allows us to exploit information from all extreme peaks-over-threshold events (on a daily scale), potentially leading to a drastic uncertainty reduction, and opens the door to more detailed modeling of intra-year seasonal variability, as well as extremal clusters that are formed due to temporal dependence. In this work, we have decided to ignore these important issues for the sake of simplicity, by assuming that the data were iid in time, but it would be interesting in future research to extend our spatial LGM framework by including temporally-structured latent random effects, that capture temporal non-stationary patterns and long-term time trends. We also stress here that, unlike the model of Cooley et al. (2007), our approach relies on the Poisson point process representation of extremes, which allows us to conveniently describe the data’s tail behavior using a single LGM (rather than two separate models for threshold exceedances, and threshold exceedance occurrences), which gives a combined unified treatment of uncertainty. Moreover, model parameters are relatively insensitive to the threshold choice (unlike LGMs based on the generalized Pareto distribution), and have a one-to-one correspondence with the block maximum approach based on the GEV distribution, which facilitates interpretation.

Beyond the distributional assumptions at the data level and the specific latent model structure, an important modeling aspect with extended LGMs is the specification of a suitable multivariate link function. In our extreme-value context, we transformed the location and scale parameters jointly to avoid overly strong correlations between latent variables, and we used a rather peculiar transformation for the shape parameter that facilitates interpretation, while avoiding pathological behaviors. To prevent estimating models with overly short or heavy tails, we made the a priori choice of restricting the shape parameter to the interval \((-0.5, 0.5)\), thus ensuring finite variance. Our approach has, thus, some links with the recently-introduced concept of property-preserving penalized complexity (P³C) priors (Castro-Camilo et al., 2021).

Overall, this chapter demonstrates the effectiveness of LGMs in estimating spatial return level surfaces with a real, large-scale, geo-environmental application (i.e., precisely, the modeling of precipitation extremes over the whole Saudi Arabian territory). However, the conditional independence assumption at the data level is clearly a limitation when the data themselves exhibit strong spatial dependence. Ignoring this issue might lead to underestimating the uncertainty of estimated parameters and return levels for univariate or spatial quantities (Johannesson et al., 2021). To circumvent this issue, a possibility involves post-processing posterior predictive samples simulated from the model at the observed sites by modifying their ranks in such a way to match the data’s empirical copula, while keeping the same marginal distributions (Johannesson et al., 2021). Another possibility might be to explicitly model data-level dependence similar to Sang and Gelfand (2010), who proposed an
LGM characterized by a Gaussian copula at the data level. However, while this approach corrects the unrealistic conditional independence assumption to some extent, it leads to heavier computations. Furthermore, the Gaussian copula is quite rigid in its joint tail and does not comply with classical extreme-value theory. In the same vein, Hazra and Huser (2021) recently fitted a Bayesian hierarchical model, constructed from a Dirichlet process mixture of low-rank Student’s t processes, to model sea surface temperature data in high dimensions (all the way from low to high quantiles). Fast Bayesian inference was made possible thanks to the availability of Gibbs updates, and to the low-rank structure stemming from the sparse set of suitably chosen spatial basis functions. However, it is not clear how to adapt their methodology to the case where interest lies in making inference for extremes only, modeled with a Poisson point process or GEV likelihood function, thus preventing observations in the bulk of the distribution from affecting the estimation of the tail structure. In future research, it would be interesting to extend Max-and-Smooth to the case of extended LGMs with data-level dependence characterized by various types of copula models.

Acknowledgments

The three authors contributed equally to this work. We thank Birgir Hrafnkelsson for inviting us to write this book chapter, and for helpful discussions and feedback on this paper. We also thank anonymous reviewers for additional comments that helped improve the paper further. This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2020-4394.

References

Asadi, P., Davison, A. C. and Engelke, S. (2015) Extremes on river networks. *Annals of Applied Statistics, 9*, 2023–2050.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2003) *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman & Hall.

Bücher, A. and Zhou, C. (2021) A horse racing between the block maxima method and the peak-over-threshold approach. *Statistical Science, 36*, 360–378.

Castro-Camilo, D., Huser, R. and 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*, 517–534.

— (2021) Practical strategies for GEV-based regression models for extremes. arXiv preprint 2106.13110.
Castruccio, S., Huser, R. and Genton, M. G. (2016) High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics*, 25, 1212–1229.

Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Coles, S. and Casson, E. (1998) Extreme value modelling of hurricane wind speeds. *Structural Safety*, 20, 283–296.

Cooley, D. and Sain, S. R. (2010) Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of Agricultural, Biological and Environmental Statistics*, 15, 381–402.

Cooley, D. S., Naveau, P. and Nychka, D. (2007) Bayesian spatial modeling of extreme precipitation return levels. *Journal of American Statistical Association*, 102, 824–840.

Davison, A. C. and Gholamrezaee, M. M. (2012) Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 468, 581–608.

Davison, A. C. and Huser, R. (2015) Statistics of extremes. *Annual Review of Statistics and its Application*, 2, 203–235.

Davison, A. C., Huser, R. and Thibaud, E. (2019) Spatial Extremes. In *Handbook of Environmental and Ecological Statistics* (eds. A. E. Gelfand, M. Fuentes, J. A. Hoeting and R. L. Smith), 711–744. CRC Press.

Davison, A. C., Padoan, S. A. and Ribatet, M. (2012) Statistical modeling of spatial extremes. *Statistical Science*, 27, 161–186.

Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds (with Discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 52, 393–442.

Deng, L., McCabe, M. F., Stenchikov, G., Evans, J. P. and Kucera, P. A. (2015) Simulation of flash-flood-producing storm events in Saudi Arabia using the weather research and forecasting model. *Journal of Hydrometeorology*, 16, 615–630.

Diggle, P. J. and Ribeiro, P. J. (2007) *Model-based Geostatistics*. New York: Springer.

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L. and Stordal, F. (2015) Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics*, 26, 89–106.

Engelke, S. and Hitz, A. S. (2020) Graphical models for multivariate extremes (with Discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82, 871–932.

de Fondeville, R. and Davison, A. C. (2018) High-dimensional peaks-over-threshold inference. *Biometrika*, 105, 575–592.

Fuglstad, G.-A., Simpson, D., Lindgren, F. and Rue, H. (2019) Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, 114, 445–452.
Geirsson, Ó. P., Hrafnkelsson, B. and Simpson, D. (2015) Computationally efficient spatial modeling of annual maximum 24-h precipitation on a fine grid. *Environmetrics*, **26**, 339–353.

Geirsson, Ó. P., Hrafnkelsson, B., Simpson, D. and Sigurdarson, H. (2020) LGM split sampler: An efficient MCMC sampling scheme for latent Gaussian models. *Statistical Science*, **35**, 218–233.

Hazra, A. and Huser, R. (2021) Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model. *Annals of Applied Statistics*, **15**, 572–596.

Hrafnkelsson, B., Morris, J. S. and Baladandayuthapani, V. (2012) Spatial modeling of annual minimum and maximum temperatures in Iceland. *Meteorology and Atmospheric Physics*, **116**, 43–61.

Hrafnkelsson, B., Siegert, S., Huser, R., Bakka, H., Jóhannesson, Á. V. et al. (2021) Max-and-Smooth: a two-step approach for approximate Bayesian inference in latent Gaussian models. *Bayesian Analysis*, **16**, 611–638.

Huerta, G. and Sansó, B. (2007) Time-varying models for extreme values. *Environmental and Ecological Statistics*, **14**, 285–299.

Huser, R. and Davison, A. C. (2013) Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, **100**, 511–518.

— (2014) Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 439–461.

Huser, R., Dombry, C., Ribatet, M. and Genton, M. G. (2019) Full likelihood inference for max-stable data. *Stat*, **8**, e218.

Huser, R., Opitz, T. and Thibaud, E. (2017) Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics*, **21**, 166–186.

— (2021) Max-infinitely divisible models and inference for spatial extremes. *Scandinavian Journal of Statistics*, **48**, 321–348.

Huser, R. and Wadsworth, J. L. (2019) Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, **114**, 434–444.

— (2020) Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics*, e1537. To appear.

Jalbert, J., Favre, A.-C., Bélisle, C. and Angers, J.-F. (2017) A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **66**, 941–962.
Johannesson, Á. V., Sieger, S., Huser, R., Bakka, H. and Hrafnkelsson, B. (2021) Approximate Bayesian inference for analysis of spatio-temporal flood frequency data. *Annals of Applied Statistics*. To appear.

Jonathan, P., Ewans, K. and Randell, D. (2014) Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, **25**, 172–188.

Lindgren, F., Rue, H. and Lindström, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**, 423–498.

Martins, E. S. and Stedinger, J. R. (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, **36**, 737–744.

Opitz, T., Huser, R., Bakka, H. and Rue, H. (2018) INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, **21**, 441–462.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, **105**, 263–277.

Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *Annals of Applied Statistics*, **6**, 1430.

Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.

Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 319–392.

Sang, H. and Gelfand, A. E. (2009) Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, **16**, 407–426.

— (2010) Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, **15**, 49–65.

Schervish, M. J. (1995) *Theory of Statistics*. Springer Science & Business Media.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H. et al. (2017) Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, **32**, 1–28.

Thibaud, E. and Opitz, T. (2015) Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, **102**, 855–870.

Vettori, S., Huser, R. and Genton, M. G. (2019) Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics*, **75**, 831–841.

Wadsworth, J. L. and Tawn, J. A. (2012) Dependence modelling for spatial extremes. *Biometrika*, **99**, 253–272.
— (2019) Higher-dimensional spatial extremes via single-site conditioning. arXiv preprint 1912.06560.

Yesubabu, V., Venkata Srinivas, C., Langodan, S. and Hoteit, I. (2016) Predicting extreme rainfall events over Jeddah, Saudi Arabia: Impact of data assimilation with conventional and satellite observations. *Quarterly Journal of the Royal Meteorological Society, 142*, 327–348.

Zhong, P., Huser, R. and Opitz, T. (2021) Modeling non-stationary temperature maxima based on extremal dependence changing with event magnitude. *Annals of Applied Statistics*. To appear.