Supplementary material for "An Efficient Workflow for Modelling High-Dimensional Spatial Extremes"

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

[1*]Department of Mathematics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway .
[2]Statistics program, CEMSE Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

*Corresponding author(s). E-mail(s): siliusv@gmail.com;
Contributing authors: sara.martino@ntnu.no;
raphael.huser@kaust.edu.sa;

1 Post hoc adjustment toy example

Shaby (2014) proposes an adjustment method that is able to properly recover the correct frequency properties of the posterior distribution when the model is misspecified. The method proposed is a post hoc adjustment of MCMC samples. Here, we show that a simular method also works for adjusting model fits from `R-INLA`. We do so by examining posterior frequency properties of the parameters of a spatial Gaussian random field. The misspecification in our example consists in approximating the Gaussian field with an SPDE approximation of too low rank.

Inside the spatial domain $\mathcal{S} = [0, 25] \times [0, 25]$ we sample n independent realisations of a spatial Gaussian random field with a Matérn covariance function, which we observe at 400 random locations. The Matérn covariance function is

$$\text{Cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}')) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|\boldsymbol{s} - \boldsymbol{s}'\|)^\nu K_\nu(\kappa\|\boldsymbol{s} - \boldsymbol{s}'\|), \tag{1}$$

where σ^2 is the marginal variance, $\nu > 0$ is the smoothness parameter and $\rho = \sqrt{8\nu}/\kappa$ is the range parameter of $Z(\boldsymbol{s})$. Furthermore, K_ν is the modified Bessel function of the second kind and order ν. Our spatial Gaussian random field has variance parameter $\sigma^2 = 1$, range parameter $\rho = 12$ and known smoothness parameter $\nu = 1.5$. We also add a Gaussian nugget effect with precision $\tau = 100$ to the random field. Parameter estimation is then performed using an SPDE approximation of low rank, i.e., based on a coarse triangulated mesh used to discretise the spatial domain. Such low-rank approximations are typically unable to capture all the variability in the data, which means that the nugget effect has to explain a large percentage of the variance, leading to underestimation of the precision τ. Thus, we expect the asymptotic maximum likelihood estimator $\boldsymbol{\theta}^*$ to be different from the true parameters $\boldsymbol{\theta} = (\tau, \rho, \sigma)^T$. To estimate $\boldsymbol{\theta}^*$, we simulate $n = 10^4$ realisations of the Gaussian Matérn field and compute the maximum likelihood estimator for the misspecified SPDE model. This gives

Table 1: Coverage percentages for unadjusted and adjusted credible intervals using the SPDE approach with a coarse mesh.

Aim	τ	τ_{adj}	ρ	ρ_{adj}	σ	σ_{adj}
90%	48%	93%	91%	90%	90%	90%
95%	55%	97%	95%	95%	95%	96%
99%	69%	99%	99%	98%	100%	99%

$\boldsymbol{\theta}^* = (\tau^*, \rho^*, \sigma^*) \approx (13.0, 14.5, 1.2)^T$. As expected, τ is severely underestimated, while ρ and σ are slightly overestimated.

For examination of frequency properties, we then sample $n = 200$ new realisations of the spatial field, and perform Bayesian inference using `R-INLA`. We assign τ a gamma prior with shape 1 and scale 2×10^4, while ρ and σ are given a joint penalised complexity (PC) prior (Fuglstad, Simpson, Lindgren, & Rue, 2019; Simpson, Rue, Riebler, Martins, & Sørbye, 2017), setting $\text{P}(\rho < 12) = 0.5$ and $\text{P}(\sigma > 1) = 0.5$. Inference is performed, the posterior distribution is adjusted as described in Section 4 of the main paper, and credible intervals are created for both the adjusted and the unadjusted model fits. For this simple toy example, we do not focus on adjusting the prior distribution as described in Section 4.1. We repeat this procedure 300 times, each time sampling $n = 200$ new realisations which we observe at the same 400 locations. Coverage frequencies can then be evaluated by examining how many of the 300 credible intervals include $\boldsymbol{\theta}^*$.

Table 1 displays the estimated coverage probabilities detailing how often the parameters of $\boldsymbol{\theta}^*$ are included in their respective credible intervals. The adjustment of the posterior yields a considerable improvement for τ. The unadjusted frequency properties of ρ and σ, however, are already good, and our adjustment method does not deteriorate the credible intervals for these parameters.

2 Simulation study

We now conduct a more complex simulation study to demonstrate our proposed workflow for modelling spatial extremes on synthetic data. Given a set of extreme realisations from simulated data we show how to compute relevant statistics of the data and how to use these for making an informed decision about the appropriate models for the standardising functions $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$. Then, we discuss details on how to define the SPDE mesh and on performing inference with `R-INLA` and the composite likelihood. Finally, we adjust the posterior distribution for possible misspecification and we evaluate the performance of the model fit.

We sample $n = 10^4$ realisations of a spatial Gaussian random field $\mathcal{Y} = \{Y_i(\boldsymbol{s}) : i = 1, \ldots, n, \boldsymbol{s} \in \mathcal{S}\}$, observed on a regular grid $\mathcal{S}$ with resolution 1×1 and size 100×100. The spatial Gaussian random field has a Matérn covariance function (1) with parameters $\sigma^2 = 1$, $\nu = 1$ and $\rho = 40$, and an additional nugget effect with variance 0.1^2. All the samples are created using an SPDE approximation. In order to model threshold exceedances with the spatial conditional extremes model, we transform the observations to have Laplace marginals using the probability integral transform. We then choose a threshold t equal to the 99.9% quantile of the Laplace distribution.

As a first step, we examine extremal dependence in the available data. Similarly to the case study in the main paper, we here (correctly) assume stationarity and isotropy, and we denote the extremal correlation coefficient as $\chi_p(\boldsymbol{s}_1, \boldsymbol{s}_2) \equiv \chi_p(d)$, where $d = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$. We estimate $\chi_p(d)$ empirically using a sliding window approach, i.e., for any value of d, we iterate over all location pairs $(\boldsymbol{s}, \boldsymbol{s}') \in \mathcal{S}^2$ satisfying $|d - \|\boldsymbol{s} - \boldsymbol{s}'\|| < \delta$, for some small tolerance $\delta > 0$, and then we count the number of times that $Y(\boldsymbol{s}) > F^{-1}(p)$ given that $Y(\boldsymbol{s}') > F^{-1}(p)$, where $F^{-1}(\cdot)$ is the quantile function of the Laplace distribution. We here choose $\delta = 0.5$. Estimators for $\chi_p(d)$ are displayed in the top-left subplot of Figure 1. Since the data have a Gaussian copula, we know that $\chi(d) = 0$ for all $d > 0$, meaning that $\chi_p(d)$ is far away from its limit

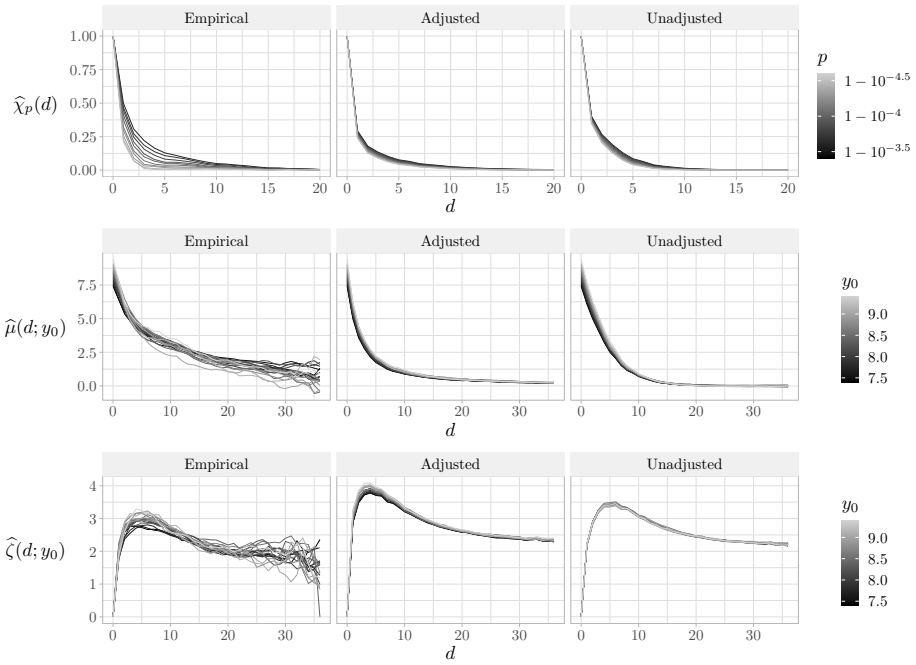

Fig. 1: Empirical estimators of $\chi_p(d)$, $\mu(d; y_0)$ and $\zeta(d; y_0)$ (top to bottom) from three different data sources. The leftmost column displays empirical estimators using the original data, while the two rightmost columns displays empirical estimators using data simulated from the adjusted and the unadjusted model fits, respectively.

$\chi(d)$ at small distances. Even if $\chi(d)$ is unknown in practice, we can here observe a clear trend of weakening dependence at increasing threshold levels, implying that the limit has not yet been reached. This demonstrates the need for a model that allows for flexible modelling of sub-asymptotic dependence, such as the spatial conditional extremes model.

To perform inference with the spatial conditional extremes model we must decide upon models for $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$. The limiting forms of these functions as $t \to \infty$ are already known for a spatial Gaussian random field (Wadsworth & Tawn, 2022). However, we here assume that the distribution of the data is unknown. Additionally, since we have chosen a finite threshold t where $\chi_p(d)$ is far away from its limit $\chi(d)$, other models for $a(\cdot)$ and $b(\cdot)$ may fit the data better than the known limiting

forms. To examine the shape of the standardising functions, we (correctly) assume stationarity in the sense that all model parameters are independent of the choice of conditioning sites, and we assume that $a(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ and $b(\boldsymbol{s}; \boldsymbol{s}_0, y_0)$ only depend on the distance $d = \|\boldsymbol{s} - \boldsymbol{s}_0\|$ and threshold exceedance y_0, meaning that we can define the standardising functions as $a(d; y_0)$ and $b(d; y_0)$ analogously. With these assumptions, we can visualise the forms of $a(d; y_0)$ and $b(d; y_0)$ by empirically computing conditional means and variances of the data. In our model, all random variables with distance d from $\boldsymbol{s}_0$ have conditional mean $\mu(d; y_0) = a(d; y_0)$ and conditional variance $\zeta^2(d; y_0) = \sigma^2(d) b^2(d; y_0) + \tau^{-1}$, where $\sigma^2(d)$ is the variance of the residual field at distance d from the conditioning site, and τ is the precision of the nugget effect. Similarly to $\widehat{\chi}_p(d)$, the empirical conditional moments of the data can be computed using a sliding window approach. However, this time, the window must slide over both values of d and y_0. We choose a rectangular window with a width of 1 in the d-direction and a width of 0.1 in the y_0-direction. The conditional moment estimators are displayed in the leftmost column of Figure 1. The conditional mean, $\widehat{\mu}(d; y_0)$, is equal to y_0 at $d = 0$, and then seems to decay exponentially towards zero as d increases. This fits well with the proposed model by Wadsworth and Tawn (2022),

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

where we set $\Delta = 0$. The conditional variance is zero at $d = 0$, and then it increases as we move away from the conditioning site and towards "the edge of the storm". Here, $\zeta(d; y_0)$ is at its largest, as it is uncertain if observations are "inside the storm", i.e., extreme, or "outside the storm", i.e., non-extreme. This is also where $\zeta(d; y_0)$ varies the most as a function of y_0. Moving further away from the conditioning site, $\zeta(d; y_0)$ decreases to a constant, as we are certainly "outside the storm", so the variance should not depend on y_0 anymore. This fits well together with a model where $b(d; y_0) = y_0^{\beta(d)}$

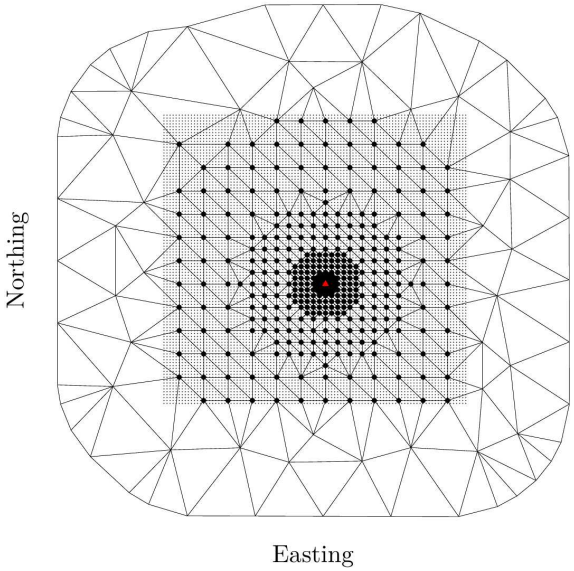

Fig. 2: Given a conditioning site $\boldsymbol{s}_0$ (displayed with (▲)), locations used for inference are displayed as big black dots (●) and locations in $\mathcal{S}$ that are not used for inference are displayed as small dots ($\cdot$). The SPDE mesh is displayed using black lines.

and where $\beta(d)$ decays to zero as the distance increases. We choose to follow Richards, Tawn, and Brown (2022) in assuming that $\beta(d) = \beta_0 \exp(-(d/\lambda_b)^{\kappa_b})$, with $0 < \beta_0 < 1$ and $\lambda_b, \kappa_b > 0$.

As seen in Figure 1, the largest changes in $\mu(d;\cdot)$ and $\zeta(d;\cdot)$ seem to occur when d is small. However, the majority of locations in $\mathcal{S}$ are located far away from $\boldsymbol{s}_0$. To account for this and give more weight to close-by locations, we discard some of the observations far away from $\boldsymbol{s}_0$ during inference, which also leads to increased inference speed. Figure 2 shows an example of the locations used to perform inference for one specific conditioning site. We stress that these locations can vary for each conditioning site used during inference.

The SPDE approach for modelling $Z_b(\cdot)$ requires that we define a triangulated mesh. Our proposed constraining method from Section 3 of the main paper requires that a mesh node is located at each conditioning site used for inference. Furthermore,

Table 2: Prior distributions for all model parameters. $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2. We give τ a penalised complexity (PC) prior such that $\mathrm{P}(\tau^{-1/2} > 1) = 0.95$. Additionally, ρ and σ are given the joint PC prior of Fuglstad et al. (2019) such that $\mathrm{P}(\rho < 60) = 0.95$ and $\mathrm{P}(\sigma > 4) = 0.05$.

$\tau \sim \mathrm{PC}(1, 0.95)$,	$\log(\lambda) \sim \mathcal{N}(3, 4^2)$,	$\log(\kappa) \sim \mathcal{N}(0, 3^2)$,
$\sigma \sim \mathrm{PC}(4, 0.05)$,	$\rho \sim \mathrm{PC}(60, 0.95)$,	$\log(\frac{\beta_0}{1-\beta_0}) \sim \mathcal{N}(0, 2^2)$,
$\log(\lambda_b) \sim \mathcal{N}(3, 4^2)$,	$\log(\kappa_b) \sim \mathcal{N}(0, 3^2)$,	

the mesh should be quite dense close to the conditioning sites to correctly capture the changes in $b(\cdot)$. Therefore, we define the mesh so that a mesh node is placed at each location used for inference. This can be problematic when performing inference with a composite likelihood that depends on multiple conditioning sites, meaning that the mesh has to be dense "everywhere" in $\mathcal{S}$, which leads to computationally demanding inference. Consequently, we choose to model $Z_b(\cdot)$ with a different mesh for each conditioning site used in the composite likelihood. Modelling different realisations of a random field with different mesh designs is not a readily available option in `R-INLA`, but this can be easily implemented using the `rgeneric`/`cgeneric` framework. An example of a mesh design for one specific conditioning site is displayed in Figure 2.

Our chosen models for $a(\cdot)$ and $b(\cdot)$ are implemented using the `cgeneric` framework, and inference is performed with `R-INLA`. The chosen priors for all the model parameters are described in Table 2. The priors are weakly informative, but with quite large variances. Using all locations in $\mathcal{S}$ as conditioning sites in the composite likelihood is computationally demanding, so we define a regular sub-grid $\mathcal{S}_0$ with resolution 6×6 and build the composite likelihood using these $|\mathcal{S}_0| = 256$ conditioning locations. The post-hoc adjustment method is then applied to robustify the model fit. Due to the large amount of available data we do not find it necessary to also adjust the prior distribution.

Figure 3 displays the adjusted and unadjusted posterior distributions of all model parameters. We see that the working assumption of independence in the composite likelihood leads to overconfidence and too focused posterior distributions, and that the

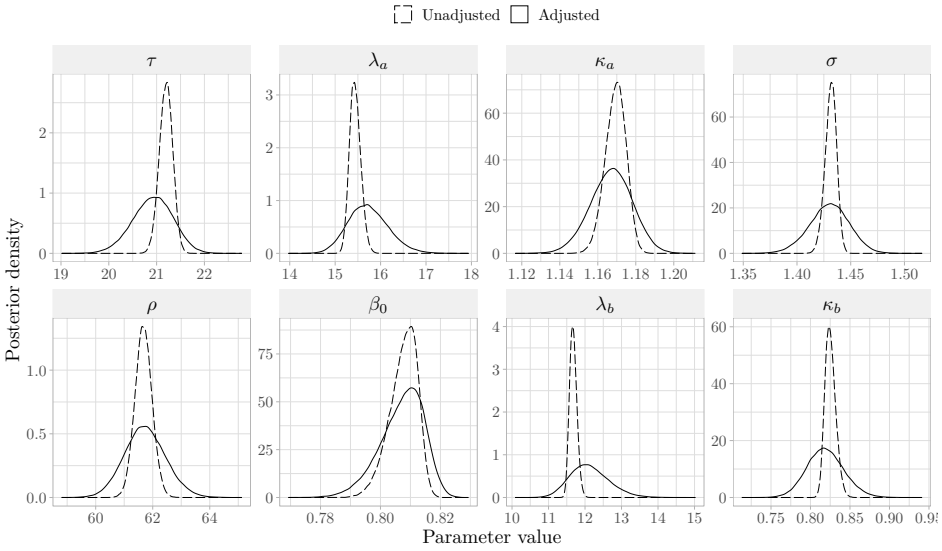

Fig. 3: Posterior distributions for all model parameters from the adjusted (solid) and the unadjusted (dashed) model fits.

adjustment method therefore increases the posterior variance to account for this misspecification. To examine the performance of our model fits, we simulate 10^5 extreme spatial fields from each fitted model, and compute $\widehat{\chi}_p(d)$, $\widehat{\mu}(d; y_0)$ and $\widehat{\zeta}(d; y_0)$ using the simulated extremes. The estimators are displayed in the two rightmost columns of Figure 1. The properties of the model fits are similar to those of the original data. There are some noticeable differences in the estimated conditional variance, which probably stems from a too simple model for $b(d; y_0)$. However, tailoring the perfect model choice for $b(d; y_0)$ is not the focus of this simulation study. Although adjusting posteriors plays a big role in properly quantifying posterior uncertainty, there are no clear differences between the point estimates from the two model fits in Figure 1. This is not very surprising, as these estimators are different types of sample means, that might be less affected by changes in the posterior variances.

Finally, we wish to quantitatively compare the adjusted model fit with the unadjusted model fit, to find out which one performs best. We choose not to compare the fits by evaluating frequency properties, as in the toy example in Section 1, because accurate

estimation of $\boldsymbol{\theta}^*$ and the repetition of the high-dimensional simulation study hundreds of times is too computationally demanding with our computational resources. Additionally, such comparisons are impossible to perform for most real-life applications with finite amounts of available data. Instead, we choose to compare the model fits by computing log-scores (e.g., Gneiting & Raftery, 2007) for a test data set that has not been used during inference. Marginal composite likelihoods may be estimated using Monte Carlo estimation: given n_s samples $\boldsymbol{\theta}_1, \dots \boldsymbol{\theta}_{n_s}$ from the posterior distribution $\pi(\boldsymbol{\theta} \mid \mathcal{Y})$, the marginal composite likelihood for a new set of observations $\mathcal{Y}_0$ is estimated as $\widehat{L}_c(\mathcal{Y}_0) = \frac{1}{n_s} \sum_{i=1}^{n_s} L_c(\boldsymbol{\theta}_i; \mathcal{Y}_0)$, where $L_c(\cdot)$ is the composite likelihood for the spatial conditional extremes model. We then denote $\log(\widehat{L}_c(\mathcal{Y}_0))$ as the estimated log-score. We sample 5×10^4 new realisations of data from the true model and locate all threshold exceedances from the 256 conditioning sites used for performing inference. Log-scores are then estimated using $n_s = 1000$ posterior samples. This results in a log-score of -2502219 for the adjusted model fit, and -2504558 for the unadjusted model fit, meaning that the adjusted model fit attains the highest log-score, with a difference of 2338. Nonparametric bootstrapping of the 5×10^4 realisations of the spatial Gaussian random field is performed to examine if the difference in log-score is significant. Using 5000 bootstrap samples, we find that the adjusted log-score always is larger than the unadjusted log-score, with a difference between 1000 and 4500. We conclude that the adjusted posterior performs better than the unadjusted posterior, even though they both provide good point estimates and reasonable fits to the simulated data.

3 Case study prerequisites

In order to perform inference with the conditional extremes model for a random process $X(\boldsymbol{s})$, one must first standardise it to a random process $Y(\boldsymbol{s})$ with Laplace margins. This is performed using the probability integral transform (Keef, Papastathopoulos,

& Tawn, 2013):

$$Y(\boldsymbol{s}) = \begin{cases} \log\left\{2F_{X(\boldsymbol{s})}(X(\boldsymbol{s}))\right\}, & X(\boldsymbol{s}) < F_{X(\boldsymbol{s})}(1/2) \\ -\log\left\{2\left[1 - F_{X(\boldsymbol{s})}(X(\boldsymbol{s}))\right]\right\}, & X(\boldsymbol{s}) \geq F_{X(\boldsymbol{s})}(1/2), \end{cases}$$

where $F_{X(\boldsymbol{s})}$ is the marginal distribution function of the random variable $X(\boldsymbol{s})$. We estimate the marginal distribution functions as the site-wise empirical distribution function of $X(\boldsymbol{s})$. However, independent standardisation of data at each location can lead to an unrealistic lack of smoothness in the transformed process $Y(\boldsymbol{s})$. Therefore, we apply a sliding window approach for computing the empirical distribution function, where the distribution at location $\boldsymbol{s}$ is estimated as the empirical distribution function of pooled data from all locations $\boldsymbol{s}'$ such that $\|\boldsymbol{s} - \boldsymbol{s}'\| \leq r$ for some radius r. Based on exploratory analysis we find $r = 5$ km to yield a realistic degree of smoothness in the estimated marginal distributions of $X(\boldsymbol{s})$ (results not shown).

A problem when modelling precipitation is that the empirical distribution has a point mass at zero. This leads to $Y(\boldsymbol{s})$ having a truncated Laplace distribution with a point mass, which can cause problems during inference. In order for $Y(\boldsymbol{s})$ to follow a non-truncated Laplace distribution, we choose to remove all zeros from the process $X(\boldsymbol{s})$ and only focus on positive precipitation. This makes us unable to model the absence of precipitation, which can lead to a slight overestimation of return levels for spatially aggregated precipitation. However, applying the fitted model for estimating properties of the untransformed process $X(\boldsymbol{s})$ is outside the scope of this paper. We believe that our choice of removing all zeros and estimating marginals using empirical distribution functions of the positive precipitation values is acceptable given the aim of our paper. In future research, we plan to properly model precipitation intermittence by appropriately accounting for the point mass at zero.

Similarly to the simulation study in Section 2, we examine extremal correlation coefficients and empirical conditional moments of the data in order to propose a good model for the extremes. The spatial domain in the case study is small enough that we can assume stationarity in the data, meaning that we can employ the same estimation methods as in the simulation study. Empirical conditional moments of the data are displayed in Figure 4. These estimators imply that the threshold t must be chosen quite large for performing successful modelling with the spatial conditional extremes model. If the threshold is chosen too low, we experience crossing in the conditional mean, i.e., for $y_1 \neq y_2$, $\widehat{\mu}(d; y_1)$ is both smaller and larger than $\widehat{\mu}(d; y_2)$ depending on the value of d. This means that a model for $a(\cdot)$ on the form $a(d; y_0) = \alpha(d)y_0$ becomes unsuitable. Furthermore, there is a clear change in the shape of the conditional variance as y_0 increases, and the spread in variance at "the edge of the storm" is so large that a model on the form $b(d; y_0) = y_0^{\beta(d)}$ would require $\beta(d) \approx 2$ for small distances d. However, $\beta(d) > 1$ leads to an ill-defined model (Wadsworth & Tawn, 2022). A more flexible model of the form $a(d; y_0) = \alpha(d, y_0)y_0$, that allows crossing, and $b(d; y_0) = y_0^{\beta(d,y_0)}$, that allows $\beta(d, y_0) > 1$ for small values of y_0, would probably fit well to the data, and could easily be implemented within the `rgeneric`/`cgeneric` framework. However, developing complex new variants of the spatial conditional extremes model is outside the scope of this paper. Consequently, we instead choose a large threshold t equal to the 99.97% threshold of the Laplace distribution, which removes the problems of crossing and excessively large values of $\beta(d)$. As we have approximately 4000 positive observations at each location, this corresponds to a mean of 1.2 threshold exceedances at each conditioning site. In practice, it yields between 0 and 5 threshold exceedances at each conditioning site.

References

Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H. (2019). Constructing priors that

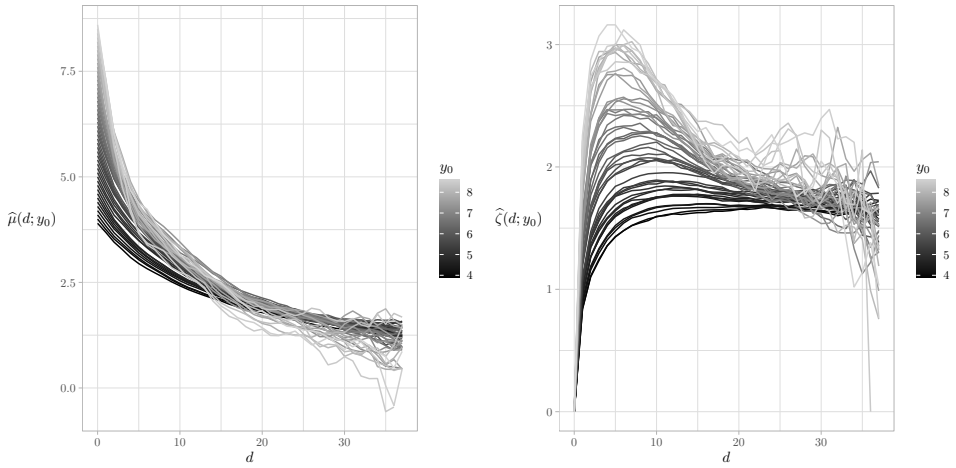

Fig. 4: Empirical estimators for $\mu(d; y_0)$ and $\zeta(d; y_0)$ using the transformed precipitation data for $y_0 > 4$ (99% quantile of the Laplace distribution)

penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, *114*(525), 445-452, https://doi.org/10.1080/01621459.2017.1415907

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

Keef, C., Papastathopoulos, I., Tawn, J.A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, *115*, 396–404, https://doi.org/10.1016/j.jmva.2012.10.012

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

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

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

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