

An Efficient Workflow for Modelling High-Dimensional Spatial Extremes

Silius M. Vandeskog

Department of Mathematics, The Norwegian University of Science and Technology (NTNU)
and

Sara Martino

Department of Mathematics, The Norwegian University of Science and Technology (NTNU)
and

Raphaël Huser

Statistics program, CEMSE Division, King Abdullah University of Science and Technology (KAUST)

December 13, 2022

Abstract

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

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

1 Introduction

The effects of climate change and the increasing availability of large and high-quality data sets has lead to a surge of research on the modelling of spatial extremes (e.g., Simpson and Wadsworth, 2021; Vandeskog et al., 2022; Opitz et al., 2018; Richards et al., 2022; Koch et al., 2021; Castro-Camilo et al., 2019; Shooter et al., 2019; Koh et al., 2021). Modelling spatial extremes is challenging for two main reasons: 1) classical models are often not flexible enough to provide realistic descriptions of extremal dependence, and 2) inference can be computationally demanding or intractable, so modellers must often rely on less efficient inference methods; see Huser and Wadsworth (2022) for a review of these challenges. In this paper, we propose a comprehensive methodological workflow, as well as practical strategies, on how to perform efficient and flexible high-dimensional modelling of spatial extremes.

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

Most classical models for spatial extremes are based on max-stable processes (Davison et al., 2012, 2019). These allow for rich modelling of asymptotic dependence, but are often too rigid in their descriptions of asymptotic independence and sub-asymptotic dependence. Other approaches have been proposed, such as scale-mixture models (Engelke et al., 2019; Huser and Wadsworth, 2019), which allow for rich modelling of both asymptotic dependence

and independence, and a more flexible description of sub-asymptotic dependence. However, these models require that all location pairs share the same asymptotic dependence class, which is problematic as one would expect neighbouring locations to be asymptotically dependent and far-away locations to be asymptotically independent. Max-mixture model (Wadsworth and Tawn, 2012) allow for even more flexible modelling of sub-asymptotic dependence, and for changing the asymptotic dependence class as a function of distance. However, it is often difficult to estimate the key model parameter, which describes the transition between extremal dependence classes. Additionally, these models must often rely on less efficient inference methods. Further improvements are given by the kernel convolution model of Krupskii and Huser (2022), more recent scale-mixture models such as that of Hazra et al. (2021), and the spatial conditional extremes model of Wadsworth and Tawn (2022), which all allows for flexible modelling of different extremal dependence classes as a function of distance. The spatial conditional extremes model allows for a particularly simple way of modelling spatial extremes. It is based on the conditional extremes model of Heffernan and Tawn (2004); Heffernan and Resnick (2007), which describes the behaviour of a random vector conditional on one of its components being extreme, and it can be interpreted as a semi-parametric regression model, which makes it intuitive and simple to tailor or extend. Due to its high flexibility and conceptual simplicity, this is our chosen model for high-dimensional spatial extremes.

To make the spatial conditional extremes model computationally efficient in higher dimensions, Wadsworth and Tawn (2022) propose to model spatial dependence using a residual random process constructed from a Gaussian copula and delta-Laplace marginal distributions. However, inference for Gaussian processes typically requires computing the inverse of the covariance matrix, whose cost scales cubically with the model dimension. Thus,

Simpson et al. (2020), propose to exchange the delta-Laplace process with a Gaussian Markov random field (Rue and Held, 2005) created using the so-called stochastic partial differential equations (SPDE) approach of Lindgren et al. (2011). Furthermore, in order to perform spatial high-dimensional Bayesian inference, Simpson et al. (2020) modify the spatial conditional extremes model into a latent Gaussian model, which allows for performing inference using the integrated nested Laplace approximation (INLA; Rue et al., 2009), implemented in the **R-INLA** software (Rue et al., 2017). This allows for a considerable improvement in the Bayesian modelling of high-dimensional spatial extremes. However, there is still much room for improvement. In this paper, we thus build upon the modelling framework of Simpson et al. (2020) and develop a more general methodology for modelling spatial conditional extremes with **R-INLA**. We also point out a theoretical weakness in the constraining methods proposed by Wadsworth and Tawn (2022) and used by Simpson et al. (2020), and we demonstrate a computationally efficient way of fixing it.

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

As extreme behaviour is, by definition, rare, inference with the conditional extremes model often relies on a composite likelihood that combines data from different conditioning sites under the working assumption of independence (Heffernan and Tawn, 2004; Wadsworth and Tawn, 2022; Richards et al., 2022; Simpson and Wadsworth, 2021). However, composite likelihoods can lead to large amounts of misspecification (Ribatet et al., 2012), and Simpson et al. (2020) thus abstain from using a composite likelihood to avoid the problems that occur when performing Bayesian inference with a composite likelihood using `R-INLA`. We show that the post hoc adjustment method accounts for the misspecification from the composite likelihood, thus allowing for more efficient inference using considerably more data.

To sum up, in this paper we develop a general workflow for performing high-dimensional modelling of spatial extremes using the spatial conditional extremes model. We improve upon the work of Simpson et al. (2020) by developing a more general, flexible and computationally efficient methodology for modelling spatial conditional extremes with `R-INLA` and the SPDE approach. Then, we make inference more robust towards misspecification by extending the post hoc adjustment method of Shaby (2014), and we further apply this adjustment method for more efficient inference by combining information from multiple conditioning sites.

The remainder of the paper is organised as follows: In Section 2, the spatial conditional extremes model is presented as a flexible choice for modelling spatial extremes. Modifications and assumptions that allow for computationally efficient inference with improved data utilisation are also presented. Then, in Section 3, we develop a general methodology for implementing a large variety of spatial conditional extremes models in `R-INLA`. Section 4 examines the problems that can occur when performing Bayesian inference based on a misspecified likelihood, and demonstrates how to perform more robust inference with `R-INLA` by accounting for possible misspecification. In Section 5, a simulation study is presented

where we demonstrate and validate our entire workflow for high-dimensional modelling of spatial extremes. Then, in Section 6, our proposed workflow is applied for modelling extreme hourly precipitation from high-resolution radar data in Norway. Finally, we conclude in Section 7 with some discussion and perspectives on future research.

2 Flexible modelling with spatial conditional extremes

2.1 The spatial conditional extremes model

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

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

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

$$\chi(\mathbf{s}_1, \mathbf{s}_2) = \lim_{p \rightarrow 1} \chi_p(\mathbf{s}_1, \mathbf{s}_2) = \lim_{p \rightarrow 1} \mathbb{P}(Y(\mathbf{s}_1) > F_Y^{-1}(p) \mid Y(\mathbf{s}_2) > F_Y^{-1}(p)),$$

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

Wadsworth and Tawn (2022) provide some guidance on parametric functions for $a(\cdot)$ and $b(\cdot)$ together with parametric distributions for $Z(\cdot)$ that cover a large range of already

existing models. For modelling $a(\cdot)$ they specifically propose the parametric function

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

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

Clearly, the best model for the standardising functions $a(\cdot)$ and $b(\cdot)$ depends on the application. Therefore, in Section 3, we develop a general methodology for implementing the conditional spatial extremes model in R-INLA for any kind of functions $a(\cdot)$ and $b(\cdot)$. In addition, we provide practical guidance and diagnostics for selecting appropriate standardising functions in our simulation study in Section 5 and data application in Section 6.

2.2 Modifications for high-dimensional modelling

To perform high-dimensional inference, Wadsworth and Tawn (2022) propose to model $Z(\cdot)$ as a random process with a Gaussian copula and delta-Laplace marginal distributions. Their proposed model for $Z(\cdot)$ has later seen usage by, e.g., Shooter et al. (2021a,b, 2022) and Richards et al. (2022). However, in order to perform Bayesian inference with R-INLA, Simpson et al. (2020) modify (1) into a latent Gaussian model by adding a Gaussian nugget effect and requiring $Z(\cdot)$ to be a fully Gaussian random field. This gives the model

$$[Y(\mathbf{s}) \mid Y(\mathbf{s}_0) = y_0 > t] \stackrel{d}{=} a(\mathbf{s}; \mathbf{s}_0, y_0) + b(\mathbf{s}; \mathbf{s}_0, y_0)Z(\mathbf{s}; \mathbf{s}_0) + \epsilon(\mathbf{s}; \mathbf{s}_0), \quad (3)$$

where $\epsilon(\mathbf{s}; \mathbf{s}_0)$ is Gaussian white noise with constant variance, satisfying $\epsilon(\mathbf{s}_0; \mathbf{s}_0) = 0$ almost surely. They further assume that $Z(\cdot)$ has zero mean and a Matérn covariance structure, so that it can be approximated using the SPDE approach of Lindgren et al. (2011), which speeds up inference by approximating the precision matrix of $Z(\cdot)$ with a sparse and low-rank matrix. However, making the precision matrix too sparse and/or low-rank leads to some model misspecification, which is further amplified by the fact that $Y(\mathbf{s})$ has Laplace marginal distributions, but is modelled using a fully Gaussian random field. We here nevertheless adopt the modelling assumptions of Simpson et al. (2020), as we find them necessary for performing truly high-dimensional Bayesian inference with R-INLA. However, unlike Simpson et al. (2020), we then account for the possible misspecification of these assumptions using the robustifying approach described in Section 4.

2.3 Efficient data utilisation with a composite likelihood

The spatial conditional extremes model consists in modelling a spatial process conditional on extreme behaviour at a predefined conditioning site. However, inference is often made challenging because the conditioning site contains few observed extremes. To strengthen

inference it is therefore common to assume stationarity, in the sense that all parameters of $a(\cdot)$, $b(\cdot)$, $Z(\cdot)$ and $\epsilon(\cdot)$ are independent of the conditioning site. Under such stationarity, it is possible to combine information from multiple conditioning sites into one global model fit, using the composite likelihood of Heffernan and Tawn (2004) and Wadsworth and Tawn (2022). Given observations $\mathcal{Y} = \{y_i(\mathbf{s}_j) : i = 1, 2, \dots, n, j = 1, 2, \dots, m\}$ from n time points and m locations, the composite log-likelihood may be expressed as

$$\ell_c(\boldsymbol{\theta}; \mathcal{Y}) = \sum_{i=1}^n \sum_{j=1}^m \ell(\boldsymbol{\theta}; \mathbf{y}_{i,-j} \mid y_i(\mathbf{s}_j)) I(y_i(\mathbf{s}_j) > t), \quad (4)$$

where $\ell(\cdot)$ is the log-likelihood of the conditional extremes model, $\mathbf{y}_{i,-j}$ is a $(m - 1)$ -dimensional vector containing observations from time point i , for all locations except \mathbf{s}_j , $I(\cdot)$ is the indicator function and $\boldsymbol{\theta}$ contains all parameters of the spatial conditional extremes model. If m is too large, one may choose to build the composite likelihood using only a subset of the available conditioning sites, and $\mathbf{y}_{i,-j}$ may be modified to contain only a subset of the available observations from time point i , which may vary with both i and j .

The composite likelihood is not a valid likelihood, since multiple of the terms in (4) may contain the same observations. Incorrectly interpreting the composite likelihood as a true likelihood is therefore tantamount to specifying a model in which $[\mathbf{y}_{i,-j} \mid y_i(\mathbf{s}_j)]$ is (wrongly) assumed to be independent from $[\mathbf{y}_{i,-k} \mid y_i(\mathbf{s}_k)]$ for all time points i and locations pairs $(\mathbf{s}_j, \mathbf{s}_k)$ with $y_i(\mathbf{s}_j) > t$ and $y_i(\mathbf{s}_k) > t$. Performing inference with the composite likelihood can therefore lead to considerable misspecification, which should be accounted for before drawing conclusions from the model fit. In Section 4, we therefore show how to robustify inference by accounting for the possible model misspecification. To the best of our knowledge, our paper is the first attempt to perform Bayesian inference for the spatial conditional extremes model based on a composite likelihood.

3 Fast inference using R-INLA

3.1 Latent Gaussian model framework

R-INLA performs inference on latent Gaussian models of the form

$$[y_i | \mathbf{u}, \boldsymbol{\theta}_1] \stackrel{i.i.d.}{\sim} \pi(y_i | \eta_i(\mathbf{u}), \boldsymbol{\theta}_1), \quad i = 1, 2, \dots, n,$$

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

$$(\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top \sim \pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2),$$

where \mathbf{u} is a latent Gaussian field with mean $\boldsymbol{\mu}(\boldsymbol{\theta}_2)$ and precision matrix $\mathbf{Q}(\boldsymbol{\theta}_2)$, and the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ are assigned priors $\pi(\boldsymbol{\theta}_1)$ and $\pi(\boldsymbol{\theta}_2)$. Observations $\mathbf{y} = (y_1, \dots, y_n)^\top$ are linked to the latent field through the linear predictor $\boldsymbol{\eta} = (\eta_1(\mathbf{u}), \dots, \eta_n(\mathbf{u}))^\top = \mathbf{A}\mathbf{u}$, where \mathbf{A} is a known design matrix. This linear predictor defines the location parameter of the likelihood $\pi(\mathbf{y} | \boldsymbol{\eta}, \boldsymbol{\theta}_1)$, via a possibly non-linear link function. All observations are assumed to be conditionally independent given $\boldsymbol{\eta}$ and $\boldsymbol{\theta}_1$, so that $\pi(\mathbf{y} | \boldsymbol{\eta}, \boldsymbol{\theta}_1) = \prod_{i=1}^n \pi(y_i | \eta_i(\mathbf{u}), \boldsymbol{\theta}_1)$. For computational reasons, when using R-INLA, the likelihood must be chosen from a predefined set of likelihood functions. The linear predictor can be decomposed into $N \geq 1$ components, $\boldsymbol{\eta} = \mathbf{A}^{(1)}\mathbf{u}^{(1)} + \dots + \mathbf{A}^{(N)}\mathbf{u}^{(N)}$, where each component represents, e.g., an intercept term, a linear combination of regression coefficients, an SPDE component, etc. All of these components must either be predefined in R-INLA or defined by the user, using the `rgeneric` framework or the recently added `cgeneric` framework.

The spatial conditional extremes model in (3) corresponds to a latent Gaussian model where the likelihood is Gaussian with variance θ_1 , say, and the linear predictor is equal to $a(\mathbf{s}; \mathbf{s}_0, y_0) + b(\mathbf{s}; \mathbf{s}_0, y_0)Z(\mathbf{s}; \mathbf{s}_0)$, with $\boldsymbol{\theta}_2$ containing the parameters of $a(\cdot)$, $b(\cdot)$ and $Z(\cdot)$. For most forms of $a(\cdot)$ and $b(\cdot)$, R-INLA does not contain suitable predefined model components for describing the linear predictor, so we must define these manually. In

order to define a new R-INLA component $\mathbf{u}^{(N+1)}$, with parameters $\boldsymbol{\theta}^{(N+1)}$, using one of the `rgeneric`/`cgeneric` frameworks, one must provide functions written in R or C, respectively, that compute the precision matrix mean and prior density of $\mathbf{u}^{(N+1)}$ for any value of $\boldsymbol{\theta}^{(N+1)}$. The `cgeneric` framework yields considerably faster inference than the `rgeneric` framework, but it requires knowledge of the lower-level C programming language. In this paper, we propose a method for defining model components for $a(\mathbf{s}; \mathbf{s}_0, y_0)$ and $b(\mathbf{s}; \mathbf{s}_0, y_0)Z(\mathbf{s}; \mathbf{s}_0)$ using the `rgeneric`/`cgeneric` frameworks, for any kind of functions $a(\cdot)$ and $b(\cdot)$. In the online supplementary material, we provide the necessary code for defining the models used in Section 5 and 6 with the `cgeneric` framework.

3.2 Defining $b(\mathbf{s}; \mathbf{s}_0, y_0)Z(\mathbf{s}; \mathbf{s}_0)$ in R-INLA

The SPDE approach creates a Gaussian Markov random field $\widehat{Z}(\mathbf{s})$ that is an approximation to a Gaussian random field $Z(\mathbf{s})$ with Matérn covariance function

$$\text{Cov}(Z(\mathbf{s}), Z(\mathbf{s}')) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa\|\mathbf{s} - \mathbf{s}'\|)^{\nu} K_{\nu}(\kappa\|\mathbf{s} - \mathbf{s}'\|), \quad (5)$$

where σ^2 is the marginal variance, $\nu > 0$ is a smoothness parameter, $\rho = \sqrt{8\nu}/\kappa$ is a range parameter and K_{ν} is the modified Bessel function of the second kind and order ν . The smoothness parameter ν is difficult to estimate from data and is therefore often given a fixed value (Lindgren and Rue, 2015). The SPDE approximation $\widehat{Z}(\mathbf{s})$ is constructed as a linear combination of Gaussian Markov random variables on a triangulated mesh, i.e., $\widehat{Z}(\mathbf{s}) = \sum_{i=1}^M \phi_i(\mathbf{s})W_i$, where W_1, \dots, W_M are random variables from a Gaussian Markov random field, and ϕ_1, \dots, ϕ_M are piecewise linear basis functions. In order to approximate the non-stationary Gaussian random field $b(\mathbf{s}; \mathbf{s}_0, y_0)Z(\mathbf{s})$ with the SPDE approach, for any

function $b(\mathbf{s}; \mathbf{s}_0, y_0)$, we modify the weights W_i to get

$$\widehat{Z}_b(\mathbf{s}; \mathbf{s}_0, y_0) = \sum_{i=1}^M \phi_i(\mathbf{s}) b(\mathbf{s}_i; \mathbf{s}_0, y_0) W_i \quad (6)$$

where $\mathbf{s}_1, \dots, \mathbf{s}_M$ are the locations of the M mesh nodes. This shares some similarities with the approach of Ingebrigtsen et al. (2014) for implementing non-stationary SPDE fields.

Since $\widehat{Z}_b(\cdot)$ is a linear combination of Gaussian random variables, its variance equals

$$\text{Var} \left(\widehat{Z}_b(\mathbf{s}; \mathbf{s}_0, y_0) \right) = \sum_{i,j=1}^M \phi_i(\mathbf{s}) \phi_j(\mathbf{s}) b(\mathbf{s}_i; \mathbf{s}_0, y_0) b(\mathbf{s}_j; \mathbf{s}_0, y_0) \text{Cov}(W_i, W_j),$$

which is unequal to $b(\mathbf{s}; \mathbf{s}_0, y_0)^2 \sigma^2$, the variance of $b(\mathbf{s}; \mathbf{s}_0, y_0) Z(\mathbf{s})$. However, if \mathbf{s} coincides with a mesh node, then one of the basis functions equals 1, while the others equal 0, giving $\text{Var} \left(\widehat{Z}_b(\mathbf{s}; \mathbf{s}_0, y_0) \right) = b(\mathbf{s}; \mathbf{s}_0, y_0)^2 \text{Var}(W_i)$, which is much closer to the correct variance. On the contrary, if \mathbf{s} is far away from a mesh node, the variance of $\widehat{Z}_b(\mathbf{s}; \mathbf{s}_0, y_0)$ may be considerably different from $b(\mathbf{s}; \mathbf{s}_0, y_0)^2 \sigma^2$. If possible, it is therefore recommended to use a fine mesh, so all observation locations are close enough to a mesh node.

The process $\widehat{Z}_b(\cdot)$ approximates the unconstrained Gaussian random field $b(\mathbf{s}; \mathbf{s}_0, y_0) Z(\mathbf{s})$. However, in order to define the conditional extremes model in R-INLA, we need to approximate the constrained field $b(\mathbf{s}; \mathbf{s}_0, y_0) Z(\mathbf{s}; \mathbf{s}_0)$, where $Z(\mathbf{s}_0; \mathbf{s}_0) = 0$ almost surely. Wadsworth and Tawn (2022) describe two different methods for turning an unconstrained Gaussian field $Z(\mathbf{s})$ into a constrained field $Z(\mathbf{s}; \mathbf{s}_0)$. The first one is to constrain the field by conditioning, i.e., $Z(\mathbf{s}; \mathbf{s}_0) = [Z(\mathbf{s}) \mid Z(\mathbf{s}) = 0]$; and the second one is to constrain it by subtraction, i.e., $Z(\mathbf{s}; \mathbf{s}_0) = Z(\mathbf{s}) - Z(\mathbf{s}_0)$. In their case studies, Wadsworth and Tawn (2022) use the first method, while Simpson et al. (2020) use the second method. We argue that constraining by subtraction yields unrealistic dependence structures, and should be avoided if other alternatives are available. A quick computation indeed shows that if $Z(\mathbf{s}; \mathbf{s}_0)$ is a stationary random process that has been constrained through subtraction, then the limiting correlation between $Z(c\mathbf{s}; \mathbf{s}_0)$ and $Z(-c\mathbf{s}; \mathbf{s}_0)$, as $c \rightarrow \infty$ equals 1/2. Furthermore, the limiting

correlation of $Z(\mathbf{s}_0 + \Delta\mathbf{s}; \mathbf{s}_0)$ and $Z(\mathbf{s}_0 - \Delta\mathbf{s}; \mathbf{s}_0)$ as $\|\Delta\mathbf{s}\| \rightarrow 0$ is often negative and equals 0 if the unconstrained random field had an exponential correlation function or -1 if the unconstrained field had a Gaussian correlation function. Thus, with the subtraction approach, points that are infinitely far away from each other are strongly correlated while points that are infinitesimally close to each other might be negatively correlated or independent.

R-INLA contains an implementation for constraining a random field by conditioning on linear combinations of itself (Rue et al., 2009). One can therefore easily constrain $\widehat{Z}_b(\cdot)$ by conditioning using the `extraconstr` option in R-INLA. Unfortunately, this conditioning method requires the computation of an $(n \times k)$ -dimensional dense matrix, where n is the number of rows of the precision matrix and k is the number of added constraints. In practice, we therefore experience that constraining by conditioning with R-INLA requires considerably more computational resources, and that it quickly turns intractable for large data sets.

We propose a third method for constraining the residual field with R-INLA. It is known that, for a Gaussian random vector $\mathbf{y} = (\mathbf{y}_1^\top, \mathbf{y}_2^\top)^\top$ with zero mean and precision matrix

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix},$$

the conditional distribution of $[\mathbf{y}_1 \mid \mathbf{y}_2 = \mathbf{0}]$ is Gaussian with zero mean and precision matrix \mathbf{Q}_{11} (Rue and Held, 2005). Thus, if we ensure that a mesh node coincides with \mathbf{s}_0 , we can constrain $Z(\cdot)$ by removing all rows and columns of \mathbf{Q} that correspond to the mesh node at \mathbf{s}_0 . This is easily achievable using the `rgeneric/cgeneric` framework, and it requires no extra computational effort.

3.3 Defining $a(\mathbf{s}; \mathbf{s}_0, y_0)$ in R-INLA

All components of the latent Gaussian field in R-INLA must be Gaussian random variables, but $a(\mathbf{s}; \mathbf{s}_0, y_0)$ is a deterministic function and not a random variable. However, clearly, the deterministic vector \mathbf{a} can be approximated well by the Gaussian random vector $\mathbf{a} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ has zero mean and covariance matrix $\delta^2 \mathbf{I}$, with \mathbf{I} being the identity matrix and δ^2 being a small, fixed marginal variance. Thus, using the `rgeneric/cgeneric` framework, we can approximate any deterministic function $a(\mathbf{s}; \mathbf{s}_0, y_0)$ with a latent Gaussian random field with mean $a(\mathbf{s}; \mathbf{s}_0, y_0)$ and diagonal covariance matrix $\delta^2 \mathbf{I}$. Here, we choose $\delta^2 = \exp(-15)$.

4 Robust inference using post hoc adjustments

4.1 Adjusting posterior samples

It is well known that all models are wrong, in the sense that the data, to a certain extent, always deviate from the model assumptions. This is particularly true when modelling extremes, where most models are based on imposing asymptotically justified assumptions onto finite amounts of data. It is also particularly true when modelling high-dimensional data, because high-dimensional models often are based on strict assumptions of (un)conditional independence and Gaussianity, in order to make inference computationally tractable, and because the amount of misspecification naturally tends to increase with the data size and dimensionality while keeping everything else constant. Accounting for this misspecification should therefore be an important step in any successful modelling strategy for high-dimensional spatial extremes.

Given n independent realisations $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ of a random vector \mathbf{Y} with true distribution G , and a chosen likelihood $L(\boldsymbol{\theta}; \mathcal{Y}) = \prod_{i=1}^n L(\boldsymbol{\theta}; \mathbf{y}_i)$, it is well known that the

maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ is asymptotically Gaussian, i.e.,

$$\mathcal{I}(\boldsymbol{\theta}^*)^{1/2} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right) \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{I}), \text{ as } n \rightarrow \infty,$$

under some weak regularity conditions (White, 1982), where \mathbf{I} is the identity matrix and $\boldsymbol{\theta}^*$ minimises the Kullback-Leibler divergence (KLD; Kullback and Leibler, 1951) between $L(\boldsymbol{\theta}; \cdot)$ and the likelihood of the true distribution G . Furthermore, $\mathcal{I}(\boldsymbol{\theta})$ is the so-called Godambe sandwich information matrix (Godambe, 1960), i.e.,

$$\mathcal{I}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})^{-1}\mathbf{H}(\boldsymbol{\theta}), \tag{7}$$

with $\mathbf{H}(\boldsymbol{\theta}) = -\mathbb{E}[\nabla_{\boldsymbol{\theta}}^2 \ell(\boldsymbol{\theta}; \mathcal{Y})]$ and $\mathbf{J}(\boldsymbol{\theta}) = \text{Cov}(\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \mathcal{Y}))$, where $\ell(\cdot) = \log L(\cdot)$ is the log-likelihood, and all expectations are taken with respect to G . If $L(\boldsymbol{\theta}^*; \cdot)$ is equal to the likelihood of the true distribution G , then $\mathbf{J}(\boldsymbol{\theta}^*) = \mathbf{H}(\boldsymbol{\theta}^*)$, and $\mathcal{I}(\boldsymbol{\theta}^*)$ reduces to $\mathbf{H}(\boldsymbol{\theta}^*)$.

From a Bayesian perspective, given a prior $\pi(\boldsymbol{\theta})$ and appropriate regularity conditions, it is also known that the posterior density, $\pi(\boldsymbol{\theta} \mid \mathcal{Y}) \propto L(\boldsymbol{\theta}; \mathcal{Y})\pi(\boldsymbol{\theta})$, converges asymptotically to a Gaussian density with mean $\boldsymbol{\theta}^*$ and covariance matrix $\mathbf{H}(\boldsymbol{\theta}^*)^{-1}$ (Berk, 1966; Kleijn and van der Vaart, 2012). As the sample size increases and the effect of the prior distribution diminishes, credible intervals and confidence intervals should be expected to coincide. However, if the likelihood is misspecified so that $\mathcal{I}(\boldsymbol{\theta}^*) \neq \mathbf{H}(\boldsymbol{\theta}^*)$, then the resulting asymptotic $(1 - \alpha)$ -credible interval differs from all well-calibrated asymptotic $(1 - \alpha)$ -confidence intervals, and we say that they attain poor frequency properties. Ribatet et al. (2012) illustrate how easily a misspecified likelihood function can lead to misleading inference through posterior intervals with poor frequency properties.

Several approaches have been proposed for robustifying inference under a misspecified likelihood (Chandler and Bate, 2007; Pauli et al., 2011; Ribatet et al., 2012; Syring and Martin, 2018), but all these methods are based on modifying the likelihood function before

inference, which is impossible to do within the R-INLA framework. However, Shaby (2014) proposes a post hoc adjustment method that properly accounts for misspecification in the likelihood by an affine transformation of posterior samples when performing MCMC-based inference. Since this is a post hoc adjustment method, it is possible to extend it for usage with R-INLA. Given a sample $\boldsymbol{\theta}$ from a posterior distribution based on a misspecified likelihood, the adjusted posterior sample is defined as

$$\boldsymbol{\theta}_{\text{adj}} = \boldsymbol{\theta}^* + \mathbf{C}(\boldsymbol{\theta} - \boldsymbol{\theta}^*), \quad (8)$$

where the matrix \mathbf{C} is chosen such that the asymptotic distribution of $\boldsymbol{\theta}_{\text{adj}}$, as $n \rightarrow \infty$, is Gaussian with mean $\boldsymbol{\theta}^*$ and covariance matrix $\mathcal{I}(\boldsymbol{\theta}^*)^{-1}$. This can be achieved by setting $\mathbf{C} = (\mathbf{M}_1^{-1}\mathbf{M}_2)^\top$, where $\mathbf{M}_1^\top\mathbf{M}_1 = \mathbf{H}(\boldsymbol{\theta}^*)^{-1}$ and $\mathbf{M}_2^\top\mathbf{M}_2 = \mathcal{I}(\boldsymbol{\theta}^*)^{-1}$. The matrix square roots \mathbf{M}_1 and \mathbf{M}_2 can be computed using, e.g., singular value decomposition.

A problem with the adjustment method of Shaby (2014) is that it distorts the contributions of the prior distribution. Using the formula for the probability density function of a transformed random variable, one can show that the distribution of the adjusted samples is

$$\pi(\boldsymbol{\theta}_{\text{adj}} \mid \mathcal{Y}) \propto L(\boldsymbol{\theta}^* + \mathbf{C}^{-1}(\boldsymbol{\theta}_{\text{adj}} - \boldsymbol{\theta}^*); \mathcal{Y})\pi(\boldsymbol{\theta}^* + \mathbf{C}^{-1}(\boldsymbol{\theta}_{\text{adj}} - \boldsymbol{\theta}^*)).$$

However, the prior distribution reflects our prior knowledge about $\boldsymbol{\theta}$, and it should ideally not be affected when adjusting for the misspecification in L . If the prior is not overly informative and the sample size is large enough, this may not matter, as the contribution of the prior will be minimal. However, if that is not the case, we propose to additionally adjust the prior distribution before inference as

$$\pi_{\text{adj}}(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}^* + \mathbf{C}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)) \cdot |\mathbf{C}|, \quad (9)$$

such that the adjusted posterior samples have distribution

$$\pi(\boldsymbol{\theta}_{\text{adj}} \mid \mathcal{Y}) \propto L(\boldsymbol{\theta}^* + \mathbf{C}^{-1}(\boldsymbol{\theta}_{\text{adj}} - \boldsymbol{\theta}^*); \mathcal{Y})\pi(\boldsymbol{\theta}_{\text{adj}}).$$

Using the `rgeneric/cgeneric` framework, one can easily define a model in R-INLA with the adjusted prior distribution $\pi_{\text{adj}}(\boldsymbol{\theta})$.

As mentioned in Section 2.3, composite likelihoods are not valid likelihood functions. However, the theory in this section holds true for a wide class of loss functions that includes negative composite log-likelihoods. Thus, we can still perform the adjustment method if we exchange $\ell(\boldsymbol{\theta}; \mathcal{Y})$ with the composite log-likelihood in (4), and $\pi(\boldsymbol{\theta} \mid \mathcal{Y})$ with the pseudo-posterior distribution $\pi_c(\boldsymbol{\theta} \mid \mathcal{Y}) \propto L_c(\boldsymbol{\theta}; \mathcal{Y})\pi(\boldsymbol{\theta})$, where $L_c(\cdot) = \exp(\ell_c(\cdot))$.

4.2 Estimating \mathbf{C} and $\boldsymbol{\theta}^*$

Here, we detail how to estimate the KLD minimiser $\boldsymbol{\theta}^*$ and the matrix \mathbf{C} when inference is based on the composite log-likelihood in (4). Our approach can also be applied in settings where inference is based on valid likelihood functions or other types of loss functions.

The KLD minimiser $\boldsymbol{\theta}^*$ can be estimated by the mode of the posterior $\pi_c(\boldsymbol{\theta} \mid \mathcal{Y})$, denoted $\hat{\boldsymbol{\theta}}^*$. The estimator $\hat{\boldsymbol{\theta}}^*$ indeed provides a good approximation to $\boldsymbol{\theta}^*$ if the prior is not overly informative and the sample size is large enough, but other estimators, such as the maximum likelihood estimator, may be more suitable if this does not hold.

In order to estimate \mathbf{C} one must first estimate $\mathbf{H}(\boldsymbol{\theta}^*)$ and $\mathbf{J}(\boldsymbol{\theta}^*)$. The first step towards estimating $\mathbf{J}(\boldsymbol{\theta}^*)$ is to compute the gradients $\nabla_{\boldsymbol{\theta}} \ell(\hat{\boldsymbol{\theta}}^*; \mathbf{y}_{i,-j} \mid y_i(\mathbf{s}_j))$ for all $j = 1, 2, \dots, m$ and $i = 1, 2, \dots, n$, such that $y_i(\mathbf{s}_j) > t$. These gradients can be computed analytically or estimated using numerical derivation methods. We then estimate $\mathbf{J}(\boldsymbol{\theta}^*)$ with

$$\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}^*) = \sum_{i,j} I(y_i(\mathbf{s}_j) > t) \nabla_{\boldsymbol{\theta}} \ell(\hat{\boldsymbol{\theta}}^*; \mathbf{y}_{i,-j} \mid y_i(\mathbf{s}_j)) \sum_{(i',j') \in \Delta(i,j)} I(y_{i'}(\mathbf{s}_{j'}) > t) \left(\nabla_{\boldsymbol{\theta}} \ell(\hat{\boldsymbol{\theta}}^*; \mathbf{y}_{i',-j'} \mid y_{i'}(\mathbf{s}_{j'})) \right)^\top, \quad (10)$$

where $\Delta(i, j)$ is the set of neighbours of (i, j) , i.e., $(i', j') \in \Delta(i, j)$ if and only if $\nabla_{\boldsymbol{\theta}} \ell(\cdot; \mathbf{y}_{i',-j'} \mid y_{i'}(\mathbf{s}_{j'}))$ is correlated with $\nabla_{\boldsymbol{\theta}} \ell(\cdot; \mathbf{y}_{i,-j} \mid y_i(\mathbf{s}_j))$. Summing over all non-correlated pairs of

tuples introduces unnecessary noise that could cause the estimator to approximately equal zero (Lumley and Heagerty, 1999). In practice one can often compute (10) using a sliding window approach. This is an improvement over the proposed estimation methods of Shaby (2014), which require either that all log-likelihood terms are independent or that it is possible to simulate data from the true and unknown distribution of the data.

Estimation of $\mathbf{H}(\boldsymbol{\theta}^*)$ is often easier than that of $\mathbf{J}(\boldsymbol{\theta}^*)$, since the former is a matrix of expected values, whereas the latter is a covariance matrix. The law of large number implies that, for n and m large enough, a good estimator for $\mathbf{H}(\boldsymbol{\theta}^*)$ is simply $\tilde{\mathbf{H}}(\hat{\boldsymbol{\theta}}^*) = -\nabla_{\boldsymbol{\theta}}^2 \ell_c(\hat{\boldsymbol{\theta}}^*; \mathcal{Y})$. Thus, all we need for estimating $\mathbf{H}(\boldsymbol{\theta}^*)$ is to compute the Hessian of the composite log-likelihood terms at $\hat{\boldsymbol{\theta}}^*$. For users of R-INLA, this is especially simple, since the program returns the inverse of the Hessian matrix

$$\tilde{\mathbf{H}}(\hat{\boldsymbol{\theta}}^*) = -\nabla_{\boldsymbol{\theta}}^2 \pi_c(\hat{\boldsymbol{\theta}}^* | \mathcal{Y}) = -\nabla_{\boldsymbol{\theta}}^2 \ell_c(\hat{\boldsymbol{\theta}}^*; \mathcal{Y}) - \nabla_{\boldsymbol{\theta}}^2 \log \pi(\hat{\boldsymbol{\theta}}^*).$$

Thus, if the prior is not overly informative and the sample size is large enough, we can set estimate $\mathbf{H}(\boldsymbol{\theta}^*)$ with $\tilde{\mathbf{H}}(\hat{\boldsymbol{\theta}}^*)$. If this is not the case, we can still estimate $\mathbf{H}(\boldsymbol{\theta}^*)$ by simply subtracting the contribution of the prior distribution from $\tilde{\mathbf{H}}(\hat{\boldsymbol{\theta}}^*)$.

If we adjust the prior distribution as in (9), it may be necessary to run R-INLA twice: once for estimating $\boldsymbol{\theta}^*$ and \mathbf{C} , and once for performing inference with the adjusted prior.

A small numerical example is shown in the supplementary material, for demonstrating that the adjustment method is able to recover the frequency properties of a posterior distribution that is based on a misspecified likelihood.

5 Simulation study

We now conduct a simulation study to demonstrate our proposed workflow for modelling spatial extremes. Given a set of extreme realisations from simulated data we show how to

compute relevant statistics of the data and how to use these for making an informed decision about the appropriate models for the standardising functions $a(\mathbf{s}; \mathbf{s}_0, y_0)$ and $b(\mathbf{s}; \mathbf{s}_0, y_0)$. 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 $\mathcal{Y} = \{Y_i(\mathbf{s}) : i = 1, \dots, n, \mathbf{s} \in \mathcal{S}\}$, observed on a regular grid \mathcal{S} with resolution 1×1 and size 100×100 . The spatial Gaussian random field has a Matérn covariance function (5) with parameters $\sigma^2 = 1$, $\nu = 1$ and $\rho = 40$, and an additional nugget effect with variance 0.1^2 . All the samples are created using an SPDE approximation. In order to model threshold exceedances with the spatial conditional extremes model, we transform the observations to have Laplace marginals using the probability integral transform. We then choose a threshold t equal to the 99.9% quantile of the Laplace distribution.

As a first step, we examine extremal dependence in the available data. If we (correctly) assume stationarity and isotropy, we can denote the extremal correlation coefficient as $\chi_p(\mathbf{s}_1, \mathbf{s}_2) \equiv \chi_p(d)$, where $d = \|\mathbf{s}_1 - \mathbf{s}_2\|$. We estimate $\chi_p(d)$ empirically using a sliding window approach, i.e., for any value of d , we iterate over all location pairs $(\mathbf{s}, \mathbf{s}') \in \mathcal{S}^2$ satisfying $|d - \|\mathbf{s} - \mathbf{s}'\|| < \delta$, for some small tolerance $\delta > 0$, and then we count the number of times that $Y(\mathbf{s}) > F^{-1}(p)$ given that $Y(\mathbf{s}') > F^{-1}(p)$, where $F^{-1}(\cdot)$ is the quantile function of the Laplace distribution. We here choose $\delta = 0.5$. Estimators for $\chi_p(d)$ are displayed in the top-left subplot of Figure 1. Since the data have a Gaussian copula, we know that $\chi(d) = 0$ for all $d > 0$, meaning that $\chi_p(d)$ is far away from its limit $\chi(d)$ at small distances. Even if $\chi(d)$ is unknown in practice, we can here observe a clear trend of weakening dependence at increasing threshold levels, implying that the limit has not yet

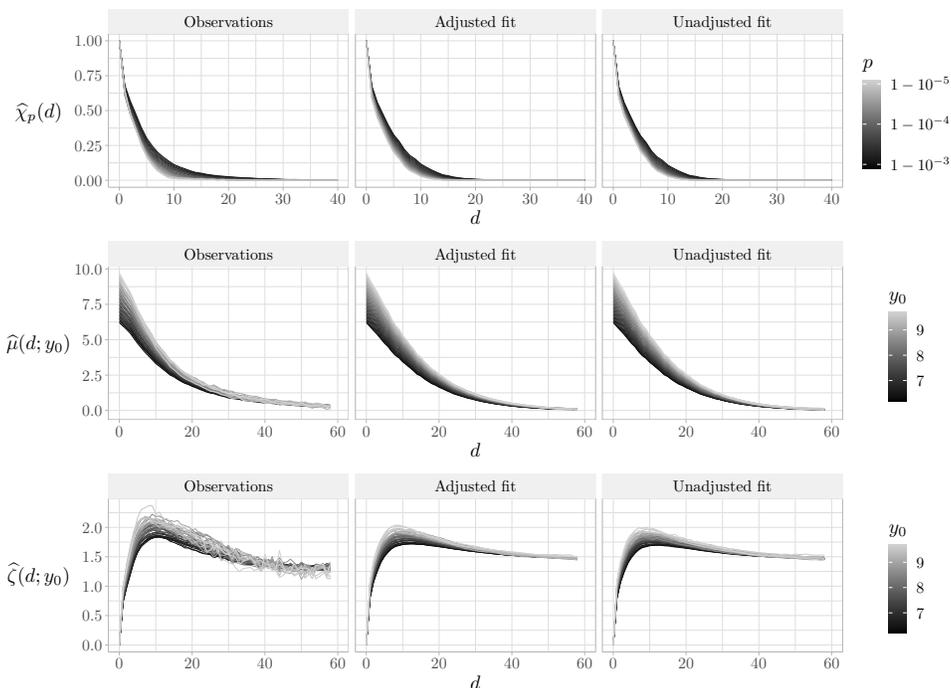


Figure 1: Empirical estimators of $\chi_p(d)$, $\mu(d; y_0)$ and $\zeta(d; y_0)$ (top to bottom) from three different data sources. The leftmost column displays empirical estimators using the original data, while the two rightmost columns displays empirical estimators using data simulated from the adjusted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

been reached. This demonstrates the need for a model that allows for flexible modelling of sub-asymptotic dependence, such as the spatial conditional extremes model.

To perform inference with the spatial conditional extremes model from (3), we must decide upon models for $a(\mathbf{s}; \mathbf{s}_0, y_0)$ and $b(\mathbf{s}; \mathbf{s}_0, y_0)$. The limiting forms of these functions as $t \rightarrow \infty$ are already known for a spatial Gaussian random field (Wadsworth and 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 $a(\mathbf{s}; \mathbf{s}_0, y_0)$ and $b(\mathbf{s}; \mathbf{s}_0, y_0)$ only depend on the distance $d = \|\mathbf{s} - \mathbf{s}_0\|$ 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 \mathbf{s}_0 have conditional mean $\mu(d; y_0) = a(d; y_0)$ and conditional variance $\zeta^2(d; y_0) = \sigma^2(d)b^2(d; y_0) + \tau^{-1}$, where $\sigma^2(d)$ is the variance of the residual field at distance d from the conditioning site, and τ is the precision of the nugget effect. Similarly to $\widehat{\chi}_p(d)$, the empirical conditional moments of the data can be computed using a sliding window approach. However, this time, the window must slide over both values of d and y_0 . We choose a rectangular window with a width of 1 in the d -direction and a width of 0.1 in the y_0 -direction. The conditional moment estimators are displayed in the leftmost column of Figure 1. The conditional mean, $\widehat{\mu}(d; y_0)$, is equal to y_0 at $d = 0$, and then seems to decay exponentially towards zero as d increases. This fits well with the proposed model in (2) if we set $\Delta = 0$. The conditional variance is zero at $d = 0$, and then it increases as we move away from the conditioning site and towards “the edge of the storm”. Here, $\zeta(d; y_0)$ is at its largest, as it is uncertain if observations are “inside the storm”, i.e., extreme, or “outside the storm”, i.e., non-extreme. This is also where $\zeta(d; y_0)$ varies the most as a function of y_0 . Moving further away from the conditioning site, $\zeta(d; y_0)$ decreases to a constant, as we are certainly “outside the storm”, so the variance should not depend on y_0 anymore. This fits well together with a model where $b(d; y_0) = y_0^{\beta(d)}$ and where $\beta(d)$ decays to zero as the distance increases. We choose to follow Richards et al. (2022) in assuming that $\beta(d) = \beta_0 \exp(-(d/\lambda_b)^{\kappa_b})$, with $0 < \beta_0 < 1$ and $\lambda_b, \kappa_b > 0$.

As seen in Figure 1, the largest changes in $\mu(d; \cdot)$ and $\zeta(d; \cdot)$ seem to occur when d is small. However, the majority of locations in \mathcal{S} are located far away from \mathbf{s}_0 . To account for

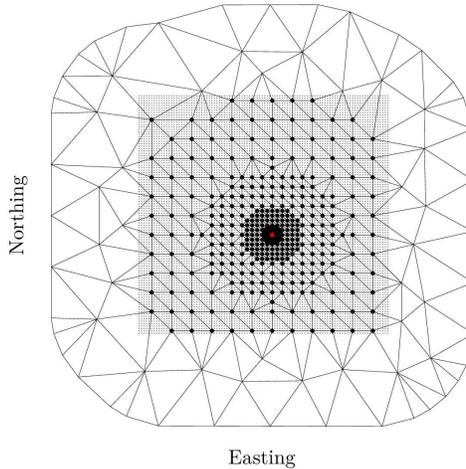


Figure 2: Given a conditioning site \mathbf{s}_0 (displayed with (▲)), locations used for inference are displayed as big black dots (●) and locations in \mathcal{S} that are not used for inference are displayed as small dots (·). The SPDE mesh is displayed using black lines.

this and give more weight to close-by locations, we discard some of the observations far away from \mathbf{s}_0 during inference, which also leads to increased inference speed. Figure 2 shows an example of the locations used to perform inference for one specific conditioning site. We stress that these locations can vary for each conditioning site used during inference.

The SPDE approach for modelling $Z_b(\cdot)$ requires that we define a triangulated mesh. Our proposed constraining method from Section 3.2 requires that a mesh node is located at each conditioning site used for inference. Furthermore, the mesh should be quite dense close to the conditioning sites to correctly capture the changes in $b(\cdot)$. Therefore, we define the mesh so that a mesh node is placed at each location used for inference. This can be problematic when performing inference with a composite likelihood that depends on multiple conditioning sites, meaning that the mesh has to be dense “everywhere” in \mathcal{S} , which leads to computationally demanding inference. Consequently, we choose to model $Z_b(\cdot)$ with a different mesh for each conditioning site used in the composite likelihood. Modelling different realisations of a random field with different mesh designs is not a readily available option in

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

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

R-INLA, but this can be easily implemented using the `rgeneric/cgeneric` framework. An example of a mesh design for one specific conditioning site is displayed in Figure 2.

Our chosen models for $a(\cdot)$ and $b(\cdot)$ are implemented using the `cgeneric` framework, and inference is performed with R-INLA. The chosen priors for all the model parameters are described in Table 1. The priors are weakly informative, but with quite large variances. Using all locations in \mathcal{S} as conditioning sites in the composite likelihood is computationally demanding, so we define a regular sub-grid \mathcal{S}_0 with resolution 6×6 and build the composite likelihood using these $|\mathcal{S}_0| = 256$ conditioning locations. The post-hoc adjustment procedure from Section 4 is then applied to robustify the model fit. Due to the large amount of available data we do not find it necessary to adjust the prior distribution as proposed in (9).

Figure 3 displays the adjusted and unadjusted posterior distributions of all model parameters. We see that the working assumption of independence in the composite likelihood leads to overconfidence and too focused posterior distributions, and that the adjustment method therefore increases the posterior variance to account for this misspecification. To examine the performance of our model fits, we simulate 10^5 extreme spatial fields from each fitted model, and compute $\hat{\chi}_p(d)$, $\hat{\mu}(d; y_0)$ and $\hat{\zeta}(d; y_0)$ using the simulated extremes. The

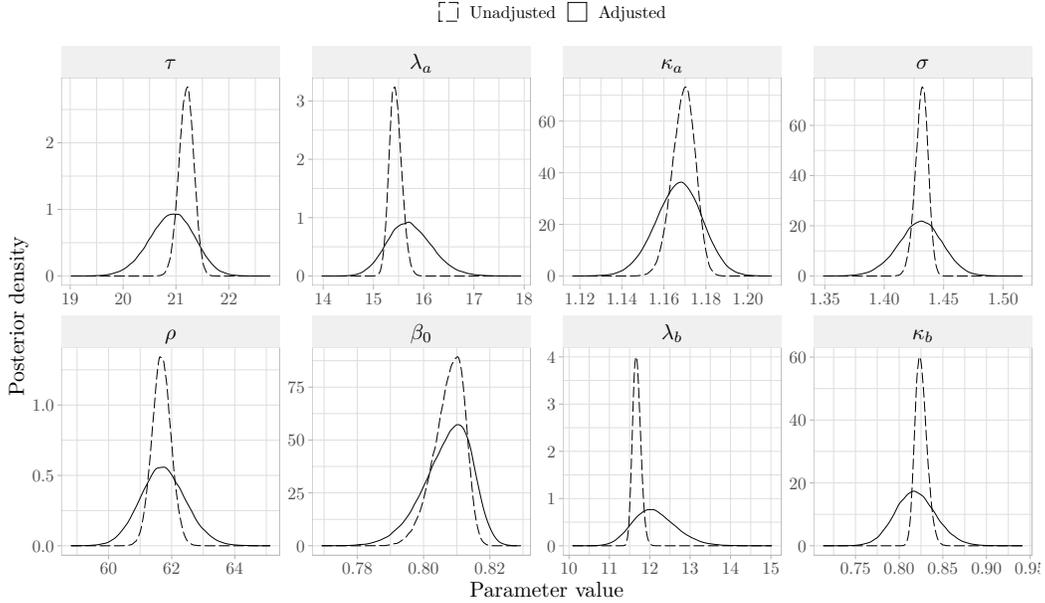


Figure 3: Posterior distributions for all model parameters from the adjusted (solid) and the unadjusted (dashed) model fits.

estimators are displayed in the two rightmost columns of Figure 1. The properties of the model fits are similar to those of the original data. There are some noticeable differences in the estimated conditional variance, which probably stems from a too simple model for $b(d; y_0)$. However, tailoring the perfect model choice for $b(d; y_0)$ is not the focus of this simulation study. Although adjusting posteriors plays a big role in properly quantifying posterior uncertainty, there are no clear differences between the point estimates from the two model fits in Figure 1. This is not very surprising, as these estimators are different types of sample means, that might be less affected by changes in the posterior variances.

Finally, we wish to quantitatively compare the adjusted model fit with the unadjusted model fit, to find out which one performs best. We choose not to compare the fits by evaluating frequency properties, as in the toy example in the supplementary material, because accurate estimation of θ^* 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 and 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.

6 Case study: Extreme precipitation in Norway

We apply our proposed methodology to the modelling of extreme hourly precipitation in Norway. Data are presented in Section 6.1 and the inference is described in Section 6.2. Results are presented and evaluated in Section 6.3.

6.1 Data

We consider $1 \times 1 \text{ km}^2$ maps of mean hourly precipitation, produced by the Norwegian Meteorological Institute by processing raw reflectivity data from the weather radar located in Rissa ($63^\circ 41' 26''\text{N}$, $10^\circ 12' 14''\text{E}$) in central Norway. Such maps are available online (<https://thredds.met.no>), dating back to 1 January 2010. We extract data from a rectangular domain, close to the Rissa radar, of size $31 \times 31 \text{ km}^2$. Denote the set of all grid points in the rectangular domain as \mathcal{S} . We then have $|\mathcal{S}| = 961$ unique locations containing hourly precipitation estimates. A map containing \mathcal{S} and the Rissa radar is displayed in Figure 4. For each $\mathbf{s} \in \mathcal{S}$, we extract all hourly observations from the summer months (June, July and August) for the years 2010–2021. Removal of missing data gives a total of 25,512 observations at each location. The number of observations with positive precipitation amounts at each location varies from 6,000 to 17,000. These large differences are likely numerical artefacts from the processing method of the Norwegian Meteorological Institute. Consequently, we set all observations smaller than 0.1 mm precipitation equal to 0. This gives a total of between 3,500 and 4,500 positive precipitation observations at each location.

6.2 Modelling and inference

The conditional extremes model in (3) is defined for a random process with Laplace margins. Thus, in order to perform inference with the conditional extremes model, we standardise the marginal distributions of the precipitation data using the probability integral transform. This is described in the supplementary materials.

Initial data exploration shows that the threshold t must be very large for a model on the form $a(d; y_0) = \alpha(d)y_0$ and $b(d; y_0) = y_0^{\beta(d)}$ to provide a good fit. Consequently, we choose a threshold equal to the 99.97% quantile of the Laplace distribution, which yields between 0

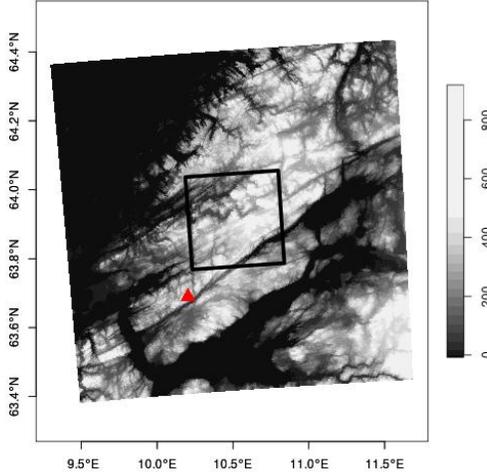


Figure 4: Elevation (m) map, over the Fosen area in central Norway. The study area \mathcal{S} is located inside the black rectangle, and the Rissa radar is displayed using a triangle (\blacktriangle).

and 5 threshold exceedances at each conditioning site. See the supplementary materials for more details and discussion on this choice. Estimators for $\chi_p(d)$, $\mu(d; y_0)$ and $\zeta(d; y_0)$ using this threshold are displayed in the leftmost column of Figure 5. Based on the lack of changes in $\widehat{\zeta}(\cdot; y_0)$ as y_0 varies, we choose to model $b(d; y_0)$ as a function not depending on y_0 . We choose the model $b(d; y_0) \equiv b(d) = 1 + b_0 \exp \{ - (d/\lambda_b)^{\kappa_b} \}$, with positive parameters b_0 , λ_b and κ_b , while we use the model (2) for $a(\cdot)$, just as in Section 5.

With only one or two threshold exceedances at most conditioning sites, separate inference for each conditioning site would be highly challenging. Inference is therefore performed using R-INLA and the composite likelihood, based on every single conditioning site in \mathcal{S} . However, we remove the two last years of the data before performing inference, so these can be used for comparing the performance of the adjusted and the unadjusted model fits. Just as in Section 5, some of the observations far away from the conditioning sites are discarded during inference, and we also define different triangulated meshes for each conditioning site. The Matérn smoothness parameter ν is fixed to a value of 1.5. Prior distributions are set equal to those in Table 1, except that we exchange the parameter β_0 from Section 5

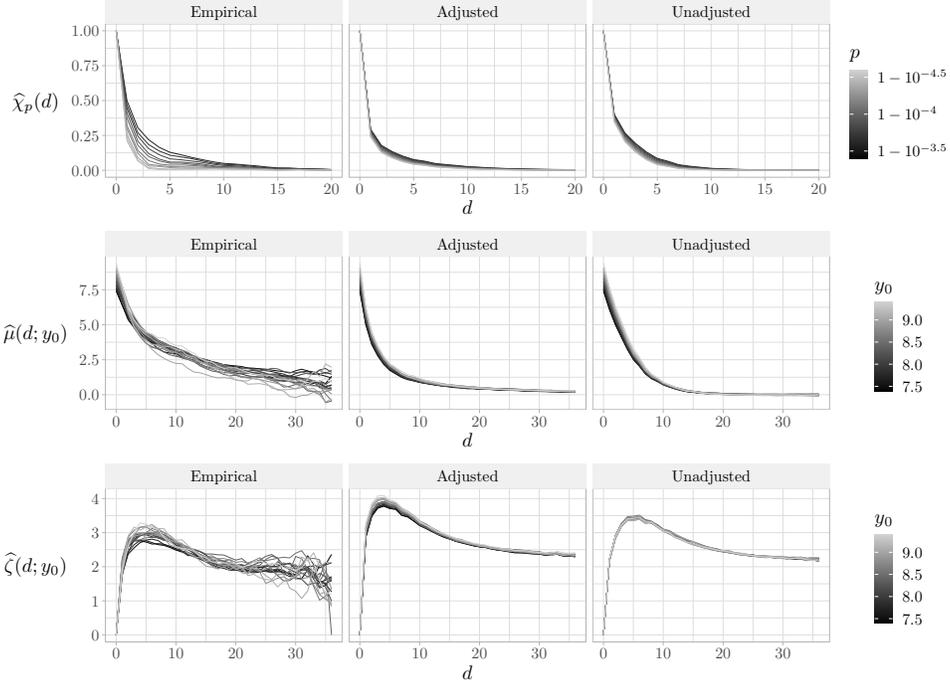


Figure 5: Empirical estimators for $\chi_p(d)$, $\mu(d; y_0)$ and $\zeta(d; y_0)$ from three different data sources. The leftmost column displays empirical estimators using the original data, while the two rightmost columns displays empirical estimators using simulated data from the adjusted and the unadjusted model fits, respectively.

with the parameter b_0 , where we place a Gaussian prior on $\log b_0$ with zero mean and a variance of 4^2 . Finally, the post hoc adjustment method is performed on the output of R-INLA. Once more, we have a large enough sample size that we choose not to adjust the prior distribution as in (9). Estimates for θ^* and $\mathbf{H}(\theta^*)$ are provided directly from R-INLA, while $\mathbf{J}(\theta^*)$ is estimated using (10) with a sliding window that has a width of 10 hours.

6.3 Results

We simulate 10^5 extreme realisations from the adjusted and unadjusted model fits. Statistics of the simulated data are displayed in the two rightmost columns of Figure 5. There are

noticeable differences between the samples from the two model fits. However, both model fits seem to capture a large part of the trends in the transformed precipitation data well. Interestingly, the adjusted conditional second moments $\widehat{\zeta}(d; y_0)$ are more different from those of the original data than those of the unadjusted model fit. However, this is not reflected in the estimated extremal correlation coefficients, $\widehat{\chi}_p(d)$. As our main goal is to capture the trends in $\chi_p(d)$, we see that the adjusted model fit seems to outperform the unadjusted one overall, especially for higher values of p . None of the model fits fully capture the rate of weakening dependence with increasing thresholds, which probably requires a more complex model for $b(d; y_0)$. As discussed in the supplementary materials, this is outside the scope of this paper and it would require further investigation in future research. However, any such model extension can easily be implemented using our proposed R-INLA methodology.

Posterior distributions for all parameters of the two model fits are displayed in Figure 6. There are considerable differences between the adjusted and the unadjusted posterior distributions. The latter one again too focused due to the working assumption of independence in the composite likelihood.

To further compare the two model fits, composite log-scores are computed using the last two years of the available data. These are estimated using $n_s = 1000$ posterior samples. This results in a composite log-score of -96520 for the unadjusted model fit, and -88029 for the adjusted model fit, meaning that the adjusted model fit seems to perform considerably better. Nonparametric bootstrapping of all time points in the test data is performed to examine if the difference is significant. Using 5000 bootstrap samples, we find that the difference in composite log-score is significantly different from zero at a 0.1% significance level, so we conclude that the adjusted model fit outperforms the unadjusted model fit.

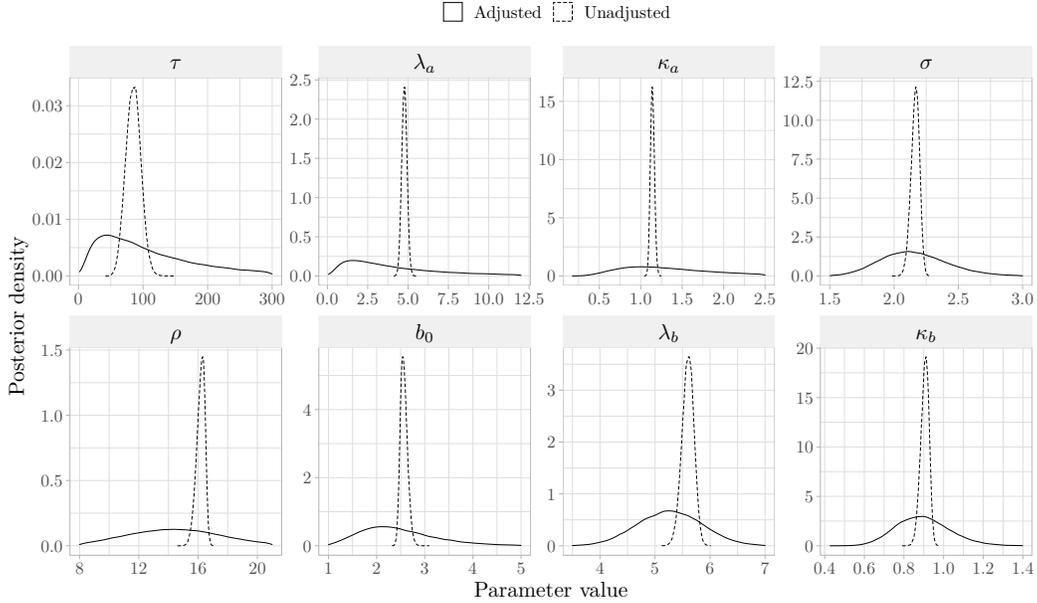


Figure 6: Posterior distributions for all model parameters of the adjusted (solid) and the unadjusted (dashed) model fits.

7 Conclusion

We propose an efficient workflow for robust modelling of spatial high-dimensional extremes using the spatial conditional extremes model with a composite likelihood and R-INLA, and a post hoc adjustment method that corrects for possible model misspecification. The workflow is demonstrated and shown to perform well in a large-scale simulation study, where we also propose a methodology for selecting appropriate forms for the standardising functions $a(\mathbf{s}; \mathbf{s}_0, y_0)$ and $b(\mathbf{s}; \mathbf{s}_0, y_0)$. Finally, the workflow is applied for modelling spatial high-dimensional extremes of Norwegian precipitation data. The methodology performs well, and we are able to capture the main extremal dependence trends in the data.

In developing our workflow, we describe a flaw in previously-used constraining methods for the residual field in the spatial conditional extremes model, and we develop a novel constraining method that is fast and easy to use when performing inference with R-INLA. We

also propose and demonstrate a general methodology for defining and implementing a large variety of spatial conditional extremes models in R-INLA using the `rgeneric/cgeneric` frameworks. Additionally, we propose an improved extension to the post hoc adjustment method that allows for correct model contributions from the prior distribution.

For transforming precipitation data onto Laplace marginals, a nonparametric method is used for estimating the marginal distributions of the original data. This method can be problematic if the aim is to estimate properties of the original and untransformed process. Further work should therefore focus on improving the transformation method when modelling extremes with the spatial conditional extremes model. Additionally, even though the spatial conditional extremes model provides good fits to the data in both the simulation study and the case study, there are still some small differences between properties of the data and properties of the model fits. These differences can probably be reduced by choosing better, possibly more complex, parametric or semiparametric forms for $a(\mathbf{s}; \mathbf{s}_0, y_0)$ and $b(\mathbf{s}; \mathbf{s}_0, y_0)$, such as, e.g., $a(\mathbf{s}; \mathbf{s}_0, y_0) = \alpha(\mathbf{s}; \mathbf{s}_0, y_0)y_0$ or $b(\mathbf{s}; \mathbf{s}_0, y_0) = y_0^{\beta(\mathbf{s}; \mathbf{s}_0, y_0)}$. Further work should therefore focus on the theoretical properties of more complex models for $a(\cdot)$ and $b(\cdot)$, and on how to best perform model selection with the spatial conditional extremes model. Finally, the selection of a threshold t for the spatial conditional extremes model can have great importance for the resulting model fit, as seen in the case study. However, to the best of our knowledge, little attention has so far been given to the problem of threshold selection in this context. Further work should therefore focus on methods for choosing thresholds that are large enough to provide a somewhat correct model fit and small enough to perform inference with low uncertainty.

Acknowledgements

The authors are grateful to Jordan Richards, Håvard Rue and Geir-Arne Fuglstad for many helpful discussions.

Funding Raphaël Huser was partially supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2020-4394.

Conflict of interest The authors report there are no competing interests to declare.

Code and data availability The necessary code and data for achieving these results are available online at <https://github.com/siliusmv/spatialConditionalExtremes>.

References

- Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, 37(1):51–58.
- 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(3):517–534.
- Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183.
- Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.
- Davison, A. C., Huser, R., and Thibaud, E. (2019). Spatial extremes. In Gelfand, A. E., Fuentes, M., Hoeting, J. A., and Smith, R. L., editors, *Handbook of Environmental and Ecological Statistics*, pages 711–744. Chapman and Hall/CRC.

- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.
- Engelke, S., Opitz, T., and Wadsworth, J. (2019). Extremal dependence of random scale constructions. *Extremes*, 22(4):623–666.
- 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(525):445–452.
- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.
- Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211.
- Hazra, A., Huser, R., and Bolin, D. (2021). Realistic and fast modeling of spatial extremes over large geographical domains. arXiv preprint 2112.10248.
- Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.
- Huser, R. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114(525):434–444.
- Huser, R. and Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews (WIREs): Computational Statistics*, 14(1):e1537.
- Ingebrigtsen, R., Lindgren, F., and Steinsland, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8:20–38.
- Kleijn, B. and van der Vaart, A. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electronic Journal of Statistics*, 6:354–381.

- Koch, E., Koh, J., Davison, A. C., Lepore, C., and Tippet, M. K. (2021). Trends in the extremes of environments associated with severe U.S. thunderstorms. *Journal of Climate*, 34(4):1259 – 1272.
- Koh, J., Pimont, F., Dupuy, J.-L., and Opitz, T. (2021). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *Annals of Applied Statistics*, to appear.
- Krupskii, P. and Huser, R. (2022). Modeling spatial tail dependence with Cauchy convolution processes. *Electronic Journal of Statistics*, 16(2):6135–6174.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.
- Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19):1–25.
- 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(4):423–498.
- Lumley, T. and Heagerty, P. (1999). Weighted empirical adaptive variance estimators for correlated data regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):459–477.
- 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(3):441–462.
- Pauli, F., Racugno, W., and Ventura, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, 21(1):149–164.
- Ribatet, M., Cooley, D., and Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22(2):813–845.
- Richards, J., Tawn, J. A., and Brown, S. (2022). Modelling extremes of spatial aggregates of precipitation using conditional methods. *The Annals of Applied Statistics*, 16(4):2693–2713.

- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. CRC press.
- 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(2):319–392.
- Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application*, 4(1):395–421.
- Shaby, B. A. (2014). The open-faced sandwich adjustment for MCMC using estimating functions. *Journal of Computational and Graphical Statistics*, 23(3):853–876.
- Shooter, R., Ross, E., Ribal, A., Young, I. R., and Jonathan, P. (2021a). Spatial dependence of extreme seas in the North East Atlantic from satellite altimeter measurements. *Environmetrics*, 32(4):e2674.
- Shooter, R., Ross, E., Ribal, A., Young, I. R., and Jonathan, P. (2022). Multivariate spatial conditional extremes for extreme ocean environments. *Ocean Engineering*, 247:110647.
- Shooter, R., Ross, E., Tawn, J., and Jonathan, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics*, 30(6):e2562.
- Shooter, R., Tawn, J., Ross, E., and Jonathan, P. (2021b). Basin-wide spatial conditional extremes for severe ocean storms. *Extremes*, 24(2):241–265.
- Sibuya, M. et al. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210.
- Simpson, E. S., Opitz, T., and Wadsworth, J. L. (2020). High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. doi: 10.48550/arxiv.2011.04486.
- Simpson, E. S. and Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, 41:100482.

- Syring, N. and Martin, R. (2018). Calibrating general posterior credible regions. *Biometrika*, 106(2):479–486.
- Vandeskog, S. M., Martino, S., Castro-Camilo, D., and Rue, H. (2022). Modelling sub-daily precipitation extremes with the blended generalised extreme value distribution. *Journal of Agricultural, Biological and Environmental Statistics*, 27(4):598–621.
- Wadsworth, J. and Tawn, J. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51:100677.
- Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.