1 Methods

1.1 Likelihood Review

Let $\mathcal{S}$ denote the spatial region of interest (e.g. Norway) and $s \in \mathcal{S}$ a specific site in this region. Let y_{ts} follow a GEV distribution with spatially dependent parameters,

$$y_{ts} \sim \text{GEV}(\mu_s, \sigma_s, \xi_s).$$

That is, the density of y_{ts} is given by

$$pr(y_{ts}|\mu_s, \kappa_s, \xi_s) = \kappa_s h(y_{ts})^{-(\xi_s+1)/\xi_s} \exp\left(-h(y_{ts})^{-\xi_s^{-1}}\right), \tag{1}$$

for $h(y_{ts}) > 0$ with

$$h(y_{ts}) = 1 + \xi_s \kappa_s (y_{ts} - \mu_s).$$

1.2 The constrained shape hierarchical model

Each location $s \in \mathcal{S}$ has its own set of parameters (μ_s, κ_s, ξ_s). The spatial variability is the result of a number of factors related to the variation in terrain and climate. To capture this information, we collect the additional covariates $\boldsymbol{x}_s$ listed in Table **??** which aim to incorporate these features. The model for e.g. μ_s is then specified as

$$\mu_s = \boldsymbol{x}_s^\top \boldsymbol{\theta}^\mu, \tag{2}$$

and similarly for κ_s and ξ_s. Here, we some elements of $\boldsymbol{\theta}^\mu$ may be structurally fixed at 0 according to a model M^μ for a fixed model M^μ. Overdispersion in the model is handled via the model

$$\mu_s = \boldsymbol{x}_s^\top \boldsymbol{\theta}^\mu + \tau_s^\mu, \tag{3}$$

where τ_S^μ follows a Gaussian Process. In addition, for the shape parameter ξ_s we specific the model

$$\xi_s^* = \boldsymbol{x}_s^\xi \boldsymbol{\theta}^\xi + \tau_s^\xi$$

$$\xi_s = \begin{cases} \xi_u & \text{if } \xi_s^* > \xi_u \\ \xi_s^* & \text{if } \xi_s^* \in [\xi_l, \xi_u] \\ \xi_l & \text{if } \xi_s^* < \xi_l \end{cases}$$

For the range (ξ_l, ξ_u).

This model imposes a conditional independence assumption on the full likelihood. Letting $\mathcal{Y}_o$ denote all observed responses, the likelihood satisfies

$$pr(\mathcal{Y}_o|\{\mu_s, \kappa_s, \xi_s^*\}_{s \in \mathcal{S}_o}) = \prod_{s \in \mathcal{S}_o} \prod_{t=1}^{T_S} pr(y_{ts}|\mu_s, \kappa_s, \xi_s)$$

Inference is performed via Markov chain Monte Carlo (MCMC) under the model in (**??**) where each component of the model is updated in turn. The joint update of the regression parameters $\boldsymbol{\theta}^\nu$ and the

model M_ν for $\nu \in \{\mu, \kappa, \xi\}$ is discussed in the next section. The updates for the Gaussian processes $\{\tau_s^\nu\}_{s\in\mathcal{S}_o}$ and the corresponding hyperparameters α^ν and λ^ν for $\nu \in \{\mu, \kappa, \xi\}$ together with the associated prior assumptions are discussed in the Appendix. Here, we use second-order Taylor expansions of the log-likelihood of the model in (1) to construct Gaussian proposal densities, thereby eliminating the need for user-defined tuning parameters for the proposal distributions, see e.g. Chapter 4.4 of **?**. The only nuance is the relationship between the uncensored ξ_s^* and the likelihood, in particular when $\xi_s \notin [\xi_l, \xi_u]$ then

$$\frac{\partial}{\partial \xi_s^*} pr(Y_s | \mu_s, \kappa_s, \xi_s) = 0.$$

which is an important feature of the updates for the overdispersion factors τ_s^ξ

2 MCMC updates of the random effects and the related hyperparameters

Here, we discuss the MCMC updates of the Gaussian processes τ_s^μ, τ_s^κ and τ_s^ξ for each $s \in \mathcal{S}_o$ as well as the related hyperparameters α_ν and λ_ν for $\nu \in \{\mu, \kappa, \xi\}$. Most of these parameters require a Metropolis-Hastings update and the associated Hastings ratios (e.g. **?**) can be calculated in a straight-forward manner. That is, assume we want to update the parameter η in our model, where η is the current value. We then draw a new value η' from a proposal distribution $pr(\eta'|\eta, \cdot)$ and accept the proposal with probability $\min\{r, 1\}$ where

$$r = \frac{pr(\boldsymbol{y}|\eta', \cdot)pr(\eta'|\cdot)pr(\eta|\eta', \cdot)}{pr(\boldsymbol{y}|\eta, \cdot)pr(\eta|\cdot)pr(\eta'|\eta, \cdot)}.$$

Here, $pr(\boldsymbol{y}|\eta, \cdot)$ denotes the likelihood of our full data set $\boldsymbol{y}$ which depends on η and potentially other parameters which are kept fixed throughout, and $pr(\eta|\cdot)$ is the prior distribution of η which similarly might depend on the other parts of the model. Given the complexity of our model, it is vital to design efficient proposal distributions which return good proposals and are robust in that they do not require fine-tuning for each individual data set.

2.1 Random effects

Under the Gaussian process model in (**??**) and (**??**), the conditional distribution of τ_s (omitting the index ν) conditional on the remaining values $\boldsymbol{\tau}_{\mathcal{S}_o \setminus s} = \{\tau_{s'}\}_{s' \in \mathcal{S}_o \setminus s}$ is given by

$$\tau_s | \boldsymbol{\tau}_{\mathcal{S}_o \setminus s}, \alpha, \lambda \sim \mathcal{N}(\hat{\tau}_s, \varsigma_s), \tag{4}$$

where

$$\hat{\tau}_s = \mathcal{K}_{\alpha,\lambda}(s, \mathcal{S}_o \setminus s) \, \mathcal{K}_{\alpha,\lambda}^{-1}(\mathcal{S}_o \setminus s, \mathcal{S}_o \setminus s) \, \boldsymbol{\tau}_{\mathcal{S}_o \setminus s}$$
$$\varsigma_s = \mathcal{K}_{\alpha,\lambda}(s, s) - \mathcal{K}_{\alpha,\lambda}(s, \mathcal{S}_o \setminus s) \, \mathcal{K}_{\alpha,\lambda}^{-1}(\mathcal{S}_o \setminus s, \mathcal{S}_o \setminus s) \, \mathcal{K}_{\alpha,\lambda}(\mathcal{S}_o \setminus s, s).$$

We use this distribution as the prior distribution for τ_s.

For designing the proposal distribution, we employ the Gaussian approximation discussed, for instance, in Chapter 4.4 of **?**. Assume that the posterior distribution of the parameter τ'_s is written on the form

$$pr(\tau'_s|\cdot) \propto \exp\big(f(\tau'_s)\big),$$

for some function f. A quadratic Taylor expansion of the log-posterior $f(\tau'_s)$ around the value τ_s gives

$$f(\tau'_s) \approx f(\tau_s) + f'(\tau_s)(\tau'_s - \tau_s) + \frac{1}{2}f''(\tau_s)(\tau'_s - \tau_s)^2$$
$$= a + b\tau'_s - \frac{1}{2}c(\tau'_s)^2,$$

where $b = f'(\tau_s) - f''(\tau_s)\tau_s$ and $c = -f''(\tau_s)$. The posterior distribution $pr(\tau'_s|\cdot)$ may now be approximated by

$$\widetilde{pr}(\tau'_s|\cdot) \propto \exp\Big(-\frac{1}{2}c(\tau'_s)^2 + b\tau'_s\Big),$$

the denisty of the Gaussian distribution $\mathcal{N}(b/c, c^{-1})$. We thus choose $\mathcal{N}(b/c, c^{-1})$ as our proposal distribution, where τ_s is the current state of the MCMC chain. From (4) it follows that

$$f'(\tau_s) = \sum_{t=1}^{T_s} \frac{\partial}{\partial \tau_s} \log pr(y_{ts}|\tau_s, \cdot) - \frac{\tau_s - \hat{\tau}_s}{\varsigma_s}$$

$$f''(\tau_s) = \sum_{t=1}^{T_s} \frac{\partial^2}{(\partial \tau_s)^2} \log pr(y_{ts}|\tau_s, \cdot) - \frac{1}{\varsigma_s},$$

However, note that for ξ we have

$$f'(\tau_s) = \sum_{t=1}^{T_s} \frac{\partial}{\partial \tau_s} \log pr(y_{ts}|\tau_s, \cdot)\mathbf{1}\{\xi_s \in [\xi_l, \xi_u]\} - \frac{\tau_s - \hat{\tau}_s}{\varsigma_s}$$

$$f''(\tau_s) = \sum_{t=1}^{T_s} \frac{\partial^2}{(\partial \tau_s)^2} \log pr(y_{ts}|\tau_s, \cdot)\mathbf{1}\{\xi_s \in [\xi_l, \xi_u]\} - \frac{1}{\varsigma_s},$$