Realistic and Fast Modeling of Spatial Extremes over Large Geographical Domains

Raphaël Huser
(joint work with Arnab Hazra and David Bolin)

King Abdullah University of Science and Technology (KAUST)

raphael.huser@kaust.edu.sa

May 25, 2022

Scientific question and Bangladesh rainfall data

The New York Times

Monsoon Rains Pummel South Asia, Displacing Millions

Flooding in Bangladesh, Bhutan, India, Myanmar and Nepal has killed scores of people, destroyed homes and structures, drowned entire villages, and forced many to crouch on rooftops hoping for rescue.

ALJAZEERA News Coronavirus Features Economy Opinion

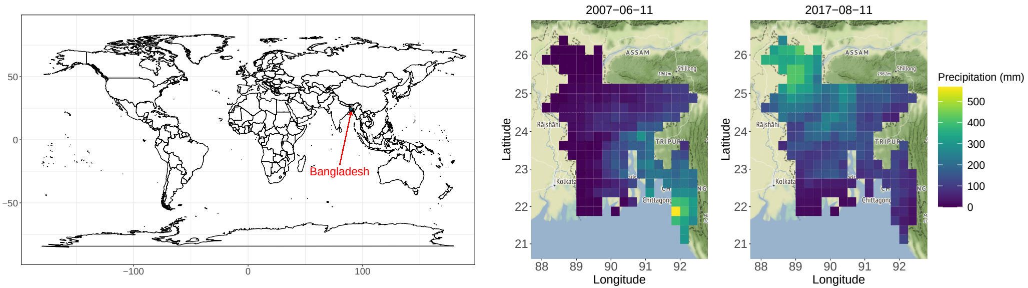
News | Climate Change

'A third of Bangladesh underwater' after heavy rains, floods

At least 1.5 million people are affected, as rivers threaten to burst their banks, officials say.

- Extreme rainfall events frequently lead to flooding and crop damage in Bengladesh.
- If spatial extreme dependence is strong, the flooding risk is amplified due to runoff accumulation.
- Biased estimation of the strength of this dependence can lead to erroneous inferences.

Scientific question and Bangladesh rainfall data



- We consider the daily rainfall data from TRMM Version 7, available from March, 2000 to December, 2019, at a resolution of $0.25^\circ \times 0.25^\circ$.
- In order to study the heavy monsoon rainfall that affects monsoon crops, we consider data only for the months of June to September.
- Total number of grid points $= 195$, and total number of days $= 2440$.

Exploratory analysis (margins)

- Marginally, it is common to model high threshold exceedances with the GPD distribution, i.e.,

$$Y - u \mid Y > u \sim \mathrm{GPD}(\sigma, \xi), \qquad \sigma > 0, \xi \in \mathbb{R}.$$

- In the heavy-tailed case (i.e., with shape $\xi > 0$), this implies that, for large $y \to \infty$,

$$\Pr(Y > y) \sim K y^{-1/\xi}, \quad K > 0.$$

- Thanks to Breiman's Lemma, a similar Pareto-like tail behavior can be obtained with the following Gaussian scale mixture model: $Y \stackrel{\mathrm{D}}{=} \mu + \tau^{-1/2} R W$, where $R \sim \mathrm{Pareto}(\gamma) \perp W \sim \mathrm{Normal}(0, 1)$. We call it the (here, univariate) HOT model (Huser et al., 2017).
- In our application, we compare the empirical and fitted quantiles for both GPD and HOT models, and observe that both models perform comparably. There is only a negligible difference between the (heavy-tailed) GPD model and the HOT model
- In our work, we prefer the HOT model, which naturally unifies the treatment of margins and dependence, while avoiding the artificial use of copulas.

Tail dependence (between pairs)

- Let C and $\bar{C}$ denote the underlying copula and the survival copula, respectively, of some random vector $(Y_1, Y_2)^T$. Two tail dependence measures are $\chi = \lim_{u \to 1} \chi_u$ and $\bar{\chi} = \lim_{u \to 1} \bar{\chi}_u$, where

$$\chi_u = 2 - \frac{\log C(u,u)}{\log(u)} \quad \text{and} \quad \bar{\chi}_u = 2\frac{\log(1-u)}{\log \bar{C}(u,u)} - 1.$$

- For large thresholds $u \to 1$, one has, for $(U_1, U_2) \sim C$:
$\chi_u \sim \Pr\{Y_1 > F_1^{-1}(u) \mid Y_2 > F_2^{-1}(u)\} = \Pr(U_1 > u \mid U_2 > u)$; and
$\Pr\{Y_1 > F_1^{-1}(u) \mid Y_2 > F_2^{-1}(u)\} = \Pr(U_1 > u \mid U_2 > u) \sim \ell\{(1-u)^{-1}\}(1-u)^{\frac{1-\bar{\chi}_u}{1+\bar{\chi}_u}}$.

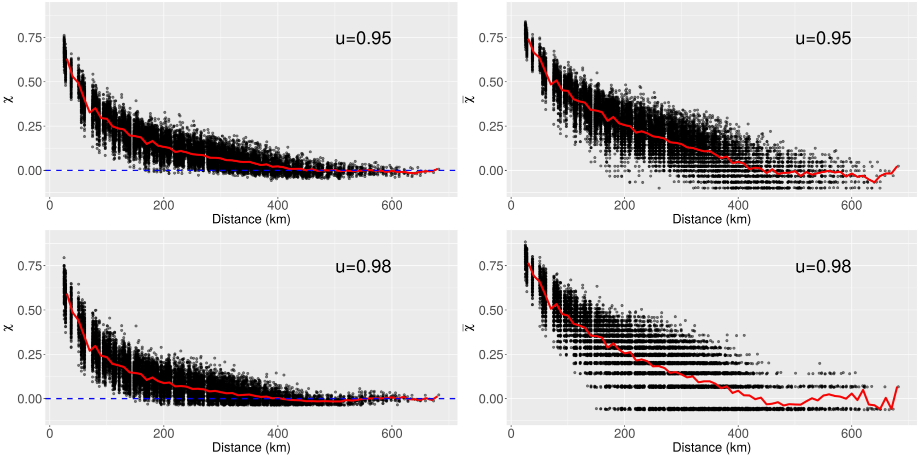
- We have asymptotic (tail) dependence if $\chi \in (0,1]$ and $\bar{\chi} = 1$, while we have asymptotic (tail) independence if $\chi = 0$ and $\bar{\chi} \in [-1, 1)$.

- For a spatial process, we can define the limit measures

$$\chi(s_1, s_2) = \lim_{u \to 1} \chi_u(s_1, s_2) \quad \text{and} \quad \bar{\chi}(s_1, s_2) = \lim_{u \to 1} \bar{\chi}_u(s_1, s_2).$$

- To obtain empirical estimates, we set u to be large and close to one ($u = 0.95, 0.98$, for example).

Exploratory analysis (tail dependence, $u = 0.95, 0.98$, including zeroes)



- The data suggest relatively strong short-range tail dependence, but long-range tail independence.

Objectives

We want a model with the following features:

- Tails of the marginal distributions are comparable with heavy-tailed GPD.
- A unified approach for margins and dependence that avoids the artificial use of copulas.
- Short-range tail dependence and long-range tail independence + possibility of full independence.
- Fast computation in high dimensions, possibly exploiting sparsity, that allows for censoring of non-extreme observations.
- Fast and straightforward simulation.

HOT model

- Huser et al. (2017) propose a Gaussian scale mixture copula model (the so-called HOT model) as $W(\boldsymbol{s}) = RZ(\boldsymbol{s})$, where $Z(\cdot) \sim \mathrm{GaussProc}(\rho)$, R and $Z(\cdot)$ are indep. and $R \sim F_{\beta,\gamma}$,

$$F_{\beta,\gamma}(r) = \begin{cases} 1 - \exp\{-\gamma(r^{\beta} - 1)/\beta\} & \text{if } \beta > 0, \\ 1 - r^{-\gamma} & \text{if } \beta = 0, \end{cases} \qquad r \in [1, \infty).$$

- This model allows either extremal dependence ($\beta = 0$) for all pairs of sites, or, extremal independence ($\beta > 0$) for all pairs of sites, but does not capture complete independence at large distances. Thus, the model may be too rigid for the Bangladesh rainfall data.

- Instead of fitting the HOT model copula, we fit the model directly to the full dataset, after allowing for location and scale parameters, which allows unified inference.

SHOT: Spatial-HOT model

- Replacing the single R with a spatial process $R(\cdot)$ can allow local shocks.

- Upon censoring of low values, we model high threshold exceedances as

$$Y_t(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \tau^{-1/2} X_t(\boldsymbol{s}), \quad \text{where} \quad X_t(\cdot) \overset{\text{IID}}{\sim} X(\cdot),$$

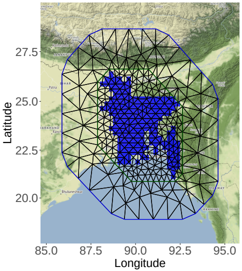
  where

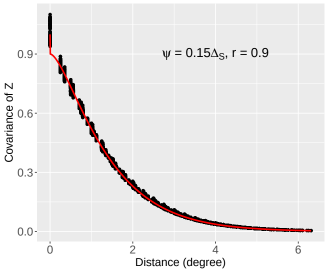
  - We model the mean as $\mu(\boldsymbol{s}) = D(\boldsymbol{s})'\boldsymbol{\theta} + \varepsilon_\mu(\boldsymbol{s})$, where $D(\boldsymbol{s}) = [1, \mathrm{lon}(\boldsymbol{s}), \mathrm{lat}(\boldsymbol{s}), \mathrm{elev}(\boldsymbol{s})]'$ and $\varepsilon_\mu(\boldsymbol{s}) \overset{\text{IID}}{\sim} \mathrm{N}(0, \tau_\mu^{-1})$. The hyperprior choices are $\boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{0}, 100^2 \boldsymbol{I}_4)$ and $\tau_\mu \sim \mathrm{Gamma}(0.1, 0.1)$.
  - The non-informative prior for the spatially-invariant fixed scale term is $\tau \sim \mathrm{Gamma}(0.1, 0.1)$.
  - We model the spatial process as $X(\boldsymbol{s}) = R(\boldsymbol{s})Z(\boldsymbol{s}), \ \boldsymbol{s} \in \mathcal{S}$, where both $R(\cdot)$ and $Z(\cdot)$ are independent spatial processes. We thus call it the SHOT process (Spatial-HOT model).
  - While the spatial variability in the tail is modeled through $\mu(\boldsymbol{s})$, our goal is to allow stationary marginal tails for $X(\boldsymbol{s})$ and the spatial extremal dependence to be approximately stationary.

Sparse construction for $Z(\cdot)$ using GMRFs

- $Z(\cdot)$ is a stationary standard Gaussian process with Matérn correlation $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2)$ that we approximate using a computationally efficient GMRF representation GMRF ε^*.

- ε^* is defined on a mesh $\mathcal{S}^*$ constructed using the SPDE approach (Lindgren et al., 2011) and $\boldsymbol{A} : \mathcal{S}^* \to \mathcal{S}$ is the projection matrix.





- $\boldsymbol{Z} = \sqrt{r}\boldsymbol{A}\varepsilon^* + \sqrt{1-r}\boldsymbol{\eta}, \varepsilon^* \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_\psi^{-1}), \boldsymbol{\eta} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I})$. Smoothness parameter is fixed to one.

- $\boldsymbol{Q}_\psi$ is sparse and obtained by discretizing the underlying SPDE of order 2.

Low-rank construction of $R(\cdot)$ using basis functions

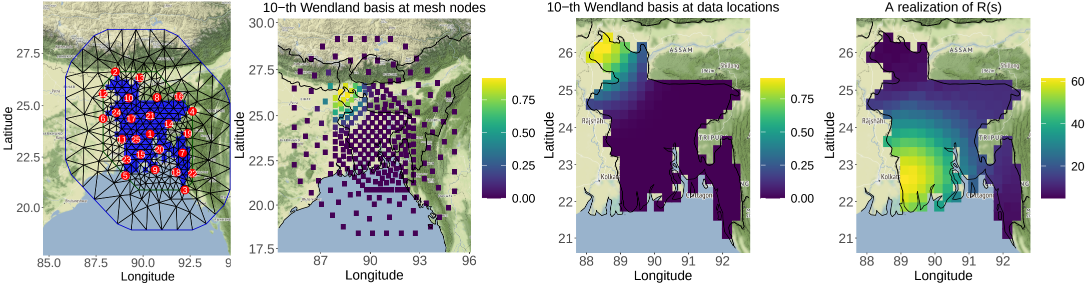
- $R(\cdot)$ is a low-rank process defined as

$$R(\boldsymbol{s}) = \left[1 + \beta \log\left\{\sum_{k=1}^{K} B_k^{1/\gamma}(\boldsymbol{s};\phi)\exp\left(\frac{R_k^{\beta}-1}{\beta}\right)\right\}\right]^{1/\beta}, \quad \beta > 0,$$

where $R_k \overset{\text{IID}}{\sim} F_{\beta,\gamma}$ and $B_k(\cdot)$'s are fixed spatial basis functions, with $\sum_{k=1}^{K} B_k(\boldsymbol{s};\phi) = 1$ for each $\boldsymbol{s}$.

- For $\beta = 0$, $R(\boldsymbol{s})$ is defined as a limit by $R(\boldsymbol{s}) = \sum_{k=1}^{K} B_k^{1/\gamma}(\boldsymbol{s};\phi) R_k$, where $R_k \overset{\text{IID}}{\sim} F_{0,\gamma}$.
- For general β, we have $\Pr\{R(\boldsymbol{s}) > r\} \sim 1 - F_{\beta,\gamma}(r)$ as $r \to \infty$.
- For $B_k(\cdot)$'s, we consider compactly-supported Wendland basis functions defined on the same mesh (excluding certain furthest nodes) as for $Z(\cdot)$ and project to data locations using $\boldsymbol{A}$.
- Specifically, the basis functions are defined as $B_k(\boldsymbol{s};\phi) = C_k(\boldsymbol{s};\phi)/\sum_{k=1}^{K} C_k(\boldsymbol{s};\phi)$, where $C_k(\boldsymbol{s};\phi) = \left(1 - \frac{\|\boldsymbol{s}-\tilde{\boldsymbol{s}}_k^*\|}{\phi}\right)^6 \left(35\frac{\|\boldsymbol{s}-\tilde{\boldsymbol{s}}_k^*\|^2}{\phi^2} + 18\frac{\|\boldsymbol{s}-\tilde{\boldsymbol{s}}_k^*\|}{\phi} + 3\right)\mathbb{I}(\|\boldsymbol{s}-\tilde{\boldsymbol{s}}_k^*\| < \phi)$, with $\tilde{\mathcal{S}} = \{\tilde{\boldsymbol{s}}_1^*, \ldots, \tilde{\boldsymbol{s}}_K^*\}$ a set of K spatial knots.

Basis functions and a realization of $R(\cdot)$



- For illustration, we fix $K = 25$ knots and $\phi = 3$. The nodes are selected according to max-min ordering (Guinness, 2018); we select the j-th mesh node as the k-th knot location if

$$j = \underset{\{j_0 \mid \mathbf{s}^*_{j_0} \in \mathcal{S}^{*(c)} \setminus \{\tilde{\mathbf{s}}^*_1, \ldots, \tilde{\mathbf{s}}^*_{k-1}\}\}}{\arg\max} \quad \min_{k_0 \in \{1,\ldots,k-1\}} d_{j_0 k_0}, \qquad \mathcal{S}^{*(c)} \subset \mathcal{S}^*,$$

where $d_{j_0 k_0}$ denotes the Euclidean distance between the j_0-th node and the k_0-th knot.

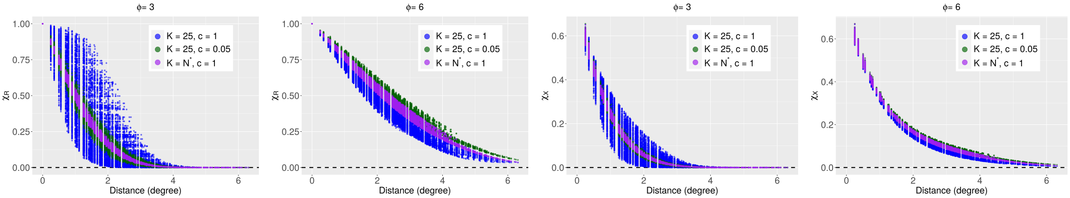
- In the above simulation, $\gamma = 2.5$ (i.e., with marginal tail index 0.4).

Properties of $X(\cdot)$

- The process $X(\cdot)$ is (approximately) tail-stationary.
- For $\beta = 0$, the extremal dependence coefficient of $X(\cdot)$ is

$$\chi_X(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sum_{k:\{B_k(\boldsymbol{s}_1;\phi)>0, B_k(\boldsymbol{s}_2;\phi)>0\}} \Bigg[B_k(\boldsymbol{s}_1;\phi)\,\bar{T}_{\gamma+1}\left(\sqrt{\gamma+1}\,\frac{B_k(\boldsymbol{s}_1;\phi)^{1/\gamma} B_k(\boldsymbol{s}_2;\phi)^{-1/\gamma} - \rho(\boldsymbol{s}_1,\boldsymbol{s}_2)}{\sqrt{1-\rho(\boldsymbol{s}_1,\boldsymbol{s}_2)^2}}\right) + B_k(\boldsymbol{s}_2;\phi)\,\bar{T}_{\gamma+1}\left(\sqrt{\gamma+1}\,\frac{B_k(\boldsymbol{s}_2;\phi)^{1/\gamma} B_k(\boldsymbol{s}_1;\phi)^{-1/\gamma} - \rho(\boldsymbol{s}_1,\boldsymbol{s}_2)}{\sqrt{1-\rho(\boldsymbol{s}_1,\boldsymbol{s}_2)^2}}\right)\Bigg].$$

- Here, $\chi_X(\boldsymbol{s}, \boldsymbol{s}) = 1$, and if no basis function $B_k(\cdot;\phi)$ "covers" both $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, then $\chi_X(\boldsymbol{s}_1, \boldsymbol{s}_2) = 0$.



Priors and computation

- We fix $\beta = 0$ to make sure we capture asymptotic dependence at short distances.
- We also fix $K = 3^2, 4^2, 5^2, 6^2$ basis functions, and choose three different choices of the Wendland range parameter $\phi = \frac{3}{4}\phi_{\text{min}} + \frac{1}{4}\phi_{\text{max}}, \frac{1}{2}\phi_{\text{min}} + \frac{1}{2}\phi_{\text{max}}, \frac{1}{4}\phi_{\text{min}} + \frac{3}{4}\phi_{\text{max}}$, where $\phi_{\min} = \max_i \min_k d(\boldsymbol{s}_i, \tilde{\boldsymbol{s}}_k^*)$ and $\phi_{\max} = \min_k \max_i d(\boldsymbol{s}_i, \tilde{\boldsymbol{s}}_k^*)$.
- For additional parameters, we consider $r \sim \mathrm{Uniform}(0,1)$, $\gamma \sim \mathrm{Uniform}(0,50)$, $\rho \sim \mathrm{Uniform}(0, 2\Delta_{\mathcal{S}})$, where $\Delta_{\mathcal{S}} = \mathrm{diameter}(\mathcal{S})$.
- We perform MCMC where we update $\mu(\boldsymbol{s}_i)$, $\boldsymbol{\theta}$, τ_μ^2, τ^2 and $\boldsymbol{Z}^*$ using Gibbs sampling, and R_k's, γ, r, and ρ using adaptive Metropolis-Hastings algorithm.
- The censored observations below the thresholds are updated using Gibbs sampling; this is possible thanks to the nugget effect that allows fast univariate updating.
- Computation time (100,000 MCMC iterations): 9.66 hours for the GMRF model, 10.40 hours for the HOT-GMRF model, and about 11.96 hours for the SHOT-GMRF model with $K = 5^2$ knots.

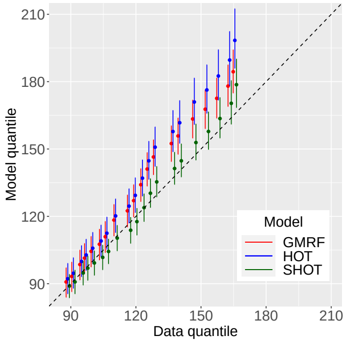
Model Comparison

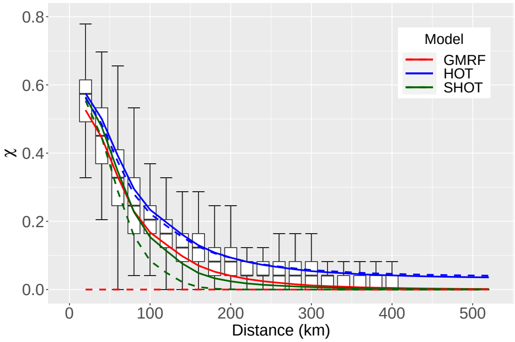
- We compare the models (a smaller value indicates a superior model) based on the DIC
  - **Deviance information criterion (DIC)**: We define deviance as $D(\theta) = -2\log(p(\boldsymbol{y}|\boldsymbol{\theta})) + C$, where $\boldsymbol{y}$ are the data, $\boldsymbol{\theta}$ are the parameters, and $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the data likelihood. Based on the MCMC samples from $\boldsymbol{\theta}$, we calculate $p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$, and further, $\mathrm{DIC} = p_D + \overline{D(\boldsymbol{\theta})}$.
  - For the SHOT model, we get the following DIC values:

Choice of K	$\phi = \frac{3}{4}\phi_{\text{min}} + \frac{1}{4}\phi_{\text{max}}$	$\phi = \frac{1}{2}\phi_{\text{min}} + \frac{1}{2}\phi_{\text{max}}$	$\phi = \frac{1}{4}\phi_{\text{min}} + \frac{3}{4}\phi_{\text{max}}$
$K = 3^2$	-7.583	-7.394	-7.875
$K = 4^2$	-8.098	-8.502	-8.014
$K = 5^2$	**-9.276**	-8.746	-8.552
$K = 6^2$	-8.710	-8.759	-8.168

  - For comparison, the GMRF and HOT models have DIC values -6.560 and -7.644, respectively. The SHOT model provides a clear improvement.

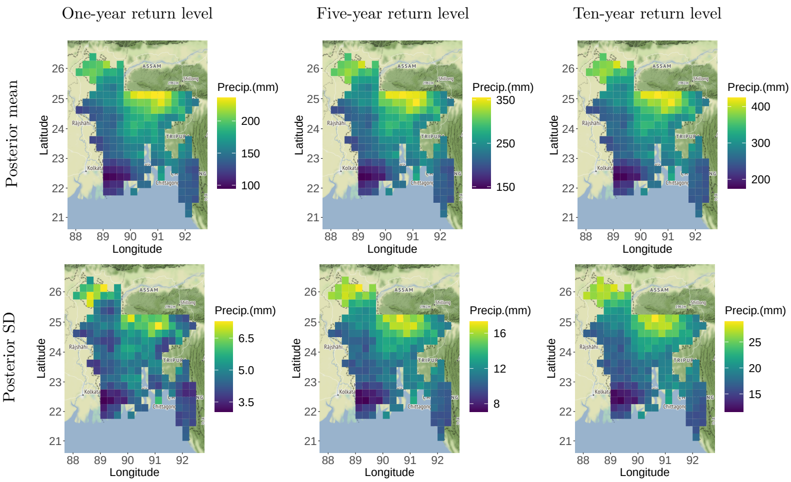
Marginal and extremal dependence structure fits





- The SHOT model provides good marginal fits, while the GMRF and HOT models overestimate high quantiles.
- All models seem to provide reasonable dependence fits at short and moderate distances, but the HOT model clearly overestimates the dependence strength at large distances.

Return level maps based on SHOT-GMRF model



- The higher values observed in the Himalayan foothills are realistic.

Conclusions

- We propose a hierarchical Bayesian spatial model that allows different extremal dependence types as a function of distance, which is realistic for rainfall and other types of environmental processes.
- The model has a sparse precision matrix of the latent Gaussian process using an SPDE-based construction, as well as a low-rank random effect structure smoothed with compactly supported basis functions, designed for fast computation.
- Extremal dependence (χ) is relatively high ($\approx$0.5) for nearby grid cells ($<$ 50km distance) and thus, there is a high chance of strong runoff accumulation.

Thank you!

arXiv preprint 2112.10248

References I

Guinness, J. (2018). Permutation and grouping methods for sharpening gaussian process approximations. *Technometrics*, 60(4):415–429.

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

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.