Modelling of sparse conditional spatial extremes processes subject to left-censoring

Léo Belzile
joint work with Rishikesh Yadav (HEC Montréal)
and Nicholas Beck (Zelus Analytics)

June 27th, 2023
Extreme Value Analysis, Bocconi University

HEC MONTRĒAL

Motivation: 2021 British Columbia floods

Figure 1: Aerial pictures of Abbotsford and Chilliwack, British Columbia, Nov. 23, 2021. Province of BC, CC license;

- 24 B.C. communities received more than 100mm of rain over three days
- Estimated 252mm precipitation for the town of Hope over that period.

Data for British Columbia lower mainland

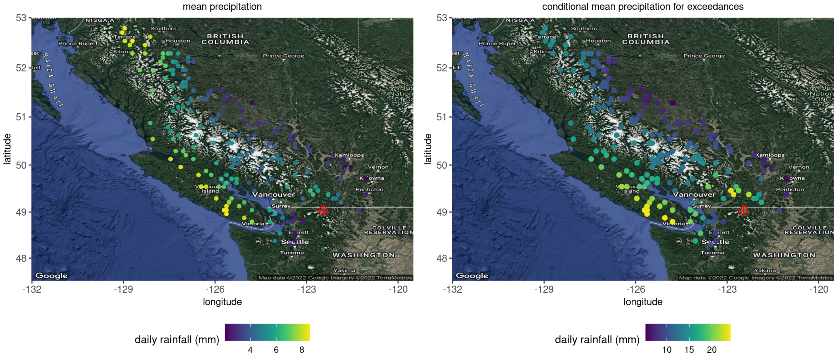


Figure 2: Average daily rainfall and conditional average given large rainfall near Abbotsford (BC). Pacific Climate Impacts Consortium daily gridded meteorological dataset NRCANMET (1950-2012).

Objective and stylized facts

Objective: construct a stochastic generator and create catalogues of daily cumulative rainfall fields given exceedance at one site.

Stylized facts:

- Extreme rainfall events tend to become more spatially localized as their intensity increases.
- Rainfall is zero-inflated (either no rain or positive amount).
- Flooding is caused by either heavy localized rainfall, or large-scale events.

Both marginal and spatio-temporal dependence of interest.

Asymptotic dependence regime

Preliminary checks suggestive of asymptotic independence at intermediate distance.

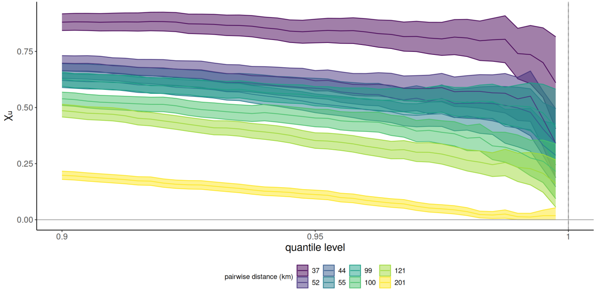


Figure 3: Tail correlation coefficient for selected sites

Conditional spatial extremes model

- We focus on the spatio conditional extremes model of Wadsworth and Tawn (2022).
- This model describes the stochastic behaviour of
  - a stationary spatial process $\{X(\boldsymbol{s}), \boldsymbol{s} \in \mathscr{S}\}$
  - with exponential tails
  - given that the value at a conditioning site s_0 within the spatial domain is large.

Conditional spatial extremes framework

We assume that there exists scaling functions $a : (\mathscr{S}, \mathbb{R}) \mapsto \mathbb{R}$ and $b : (\mathscr{S}, \mathbb{R}) \mapsto \mathbb{R}_+$, such that, for $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d \in \mathscr{S}$, as the threshold $u \to \infty$,

$$\left(X(\boldsymbol{s}_0) - u, \left[\frac{X(\boldsymbol{s}_j) - a\{\boldsymbol{s}_j, X(\boldsymbol{s}_0)\}}{b\{\boldsymbol{s}_j, X(\boldsymbol{s}_0)\}} \right]_{j=1,\ldots,d} \right) \,\middle|\, X(\boldsymbol{s}_0) > u$$

$$\xrightarrow{\mathrm{w}} \left[E, \{Z(\boldsymbol{s}_j)\}_{j=1,\ldots,d} \right].$$

COSPEX model

For sufficiently high threshold u, the standardized conditional field $X_0(\boldsymbol{s}) := X(\boldsymbol{s}) \mid X(\boldsymbol{s}_0) > u$ is of the form

$$X_0(\boldsymbol{s}) \stackrel{d}{\approx} a\{\boldsymbol{s}, X(\boldsymbol{s}_0)\} + b\{\boldsymbol{s}, X(\boldsymbol{s}_0)\}Z(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathscr{S}.$$

- $Z(\boldsymbol{s}_0) = 0$ almost surely (identifiability) and
- $Z(\boldsymbol{s})$ independent of $\lim_{u\to\infty} X(\boldsymbol{s}_0) - u \mid X(\boldsymbol{s}_0) > u \sim \mathsf{Exp}(1)$.

We consider a Gaussian residual random field $Z(\boldsymbol{s})$ to model the spatial dependence.

Normalizing functions

Wadsworth and Tawn (2022) list conditions for scaling functions $a(\cdot)$ and $b(\cdot)$ that guarantee valid limiting models

- for example, $a(\boldsymbol{s}_0, x) = x$ and decreasing with distance.
- taking

$$a(\boldsymbol{s}, x) = x\rho(\mathrm{max}\{0, \|\boldsymbol{s} - \boldsymbol{s}_0\| - \delta\}).$$

  with ρ any correlation function works and yields asymptotic dependence up to distance lag δ.

In the sequel, we consider for simplicity

- $b(\boldsymbol{s}, x) = x^{\beta}$ for $\beta \in (0, 1)$
- an exponential correlation function

$$\rho(d) = \exp(-d/\kappa_a).$$

Inference for COSPEX

- Up until now, only two-stage estimation has been considered.
- Mostly frequentist inference, except Simpson, Opitz, and Wadsworth (2023) and Vandeskog, Martino, and Huser (2022) who use INLA.
- Same computational bottlenecks for left-censoring (e.g., zero-inflated data) as other spatial extreme value models.
  - Need to evaluate numerically high dimensional Gaussian integrals
  - the censoring pattern changes from one observation to the next (repeated matrix decompositions)
- Censoring only considered Richards, Tawn, and Brown (2022) and extension thereof, using composite likelihoods.

Objectives and contribution

We extend previous work on the conditional spatial extremes model to

- Establish a computationally feasible way for handling large-scale inference in the presence of censoring
  - use full likelihoods
  - data augmentation (Bayesian paradigm)
- Simultaneously estimate marginal parameters and dependence structure
  - as there is plenty of uncertainty in the margins,
  - and the Laplace margins are not compatible with the model specification.

Data augmentation and the nugget trick

Zhang, Shaby, and Wadsworth (2022) proposed adding a nugget $\varepsilon \sim \mathsf{No}(0, \tau^2)$ to facilitate data augmentation and account for measurement error

$$X_0(\boldsymbol{s}) \stackrel{d}{\approx} a\{\boldsymbol{s}, X(\boldsymbol{s}_0)\} + b\{\boldsymbol{s}, X(\boldsymbol{s}_0)\}Z(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}).$$

Data augmentation

Benefits: $X_i \equiv X_0(\boldsymbol{s}_i)$ $(i = 1, \ldots, d)$ are conditionally independent given $\boldsymbol{Z}$.

If observations are left-censored at quantile q, the conditional likelihood contribution is

$$p\left(X_j = x_j \mid Z_j; \boldsymbol{\theta}_a, \boldsymbol{\theta}_b, \tau\right) = \begin{cases} \Phi(q; \mu_j, \tau^2), & x_j \leq q \\ \phi(x_j; \mu_j, \tau^2), & x_j > q \end{cases}$$

where $\mu_j = a(\boldsymbol{s}_j, x_0) + b(\boldsymbol{s}_j, x_0) Z_j$.

No free lunch!: This moves the problem to imputation of random effects $\boldsymbol{Z}$, but calculating the likelihood of the latter requires $\mathrm{O}(d^3)$ flops for each time replication …still unscalable.

Sparse residual process

Use the SPDE approach of Lindgren, Rue, and Lindström (2011) for a Matérn field to get a Gaussian Markov random field approximation.

Use this for the residual process, as in Simpson, Opitz, and Wadsworth (2023), where in vector form

$$\boldsymbol{Z} = \sqrt{r_Z}\mathbf{A}\boldsymbol{W} + \sqrt{(1 - r_Z)}\boldsymbol{\epsilon}.$$

- $\boldsymbol{W} \sim \mathsf{No}_K(0_K, \mathbf{Q}^{-1})$ are Gaussian weights,
- the sparse precision matrix $\mathbf{Q}$ depends on the mesh,
- the projector matrix $\mathbf{A}$ relates observations to the mesh
- $\epsilon_Z \sim \mathsf{No}(0, 1)$ are i.i.d. white noise (nugget).

Mesh and conditioning site

As in Vandeskog, Martino, and Huser (2022), we create a mesh with a node at $\boldsymbol{s}_0$, extract the precision matrix and shed it.

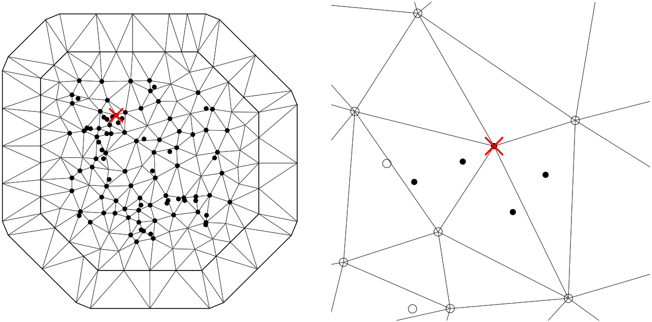


Figure 4: Mesh for SPDE approximation with conditioning site (red cross)

Leveraging the sparsity leads to efficient data augmentation building on Zhang, Shaby, and Wadsworth (2022)

$$\begin{aligned} X_0(\boldsymbol{s}_j) \mid Z_j = z_j &\sim \mathsf{No}\left\{\mathbf{a}_j(x_0) + \mathbf{b}_j(x_0)z_j, \tau^2\right\}, \\ Z_j \mid \boldsymbol{W} = \boldsymbol{w} &\sim \mathsf{No}\left(\sqrt{r_Z}\mathbf{A}_{j,\mathsf{ne}(j)}\boldsymbol{w}_{\mathsf{ne}(j)}, 1 - r_Z\right) \end{aligned}$$

where the mean of Z_j depends only on neighbours as a result of the sparsity of $\mathbf{A}$.

This means that, observations are conditionally independent given random effects.

Hierarchical model formulation

We censor observations $X(\boldsymbol{s}_j)$ below marginal quantile q_j,

$$
\begin{aligned}
p\left(X_j \mid \boldsymbol{Z}, \boldsymbol{\Theta}_a, \boldsymbol{\Theta}_b, \tau\right) &= \begin{cases} \phi\left\{x_j; \mathbf{a}_j(x_0) + \mathbf{b}_j(x_0)z_j, \tau^2\right\}, & x_j > q_j \\ \Phi\left\{q_j; \mathbf{a}_j(x_0) + \mathbf{b}_j(x_0)z_j, \tau^2\right\}, & x_j \leq q_j \end{cases} \\
\boldsymbol{Z} \mid \boldsymbol{W}, r_Z &\sim \mathsf{No}_d\left\{\sqrt{r_Z}\mathbf{A}\boldsymbol{W}, (1 - r_Z)\mathbf{I}_d\right\}; \\
\boldsymbol{W} \mid r_Z, \rho &\sim \mathsf{No}_K(0_K, r_Z\mathbf{Q}^{-1}); \\
\boldsymbol{\Theta} &\sim \pi(\boldsymbol{\Theta}).
\end{aligned}
$$

We use Markov chain Monte Carlo methods to draw posterior samples from the model.

- Gibb's sampling for the weights $\boldsymbol{W}$
- random walk Metropolis–Hastings, MALA and second-order approximations for $\boldsymbol{Z}$ and model parameters $\boldsymbol{\Theta}$.

Convergence diagnostics

There is strong autocorrelation between some of the parameters of the normalizing functions, so block updates are advisable.

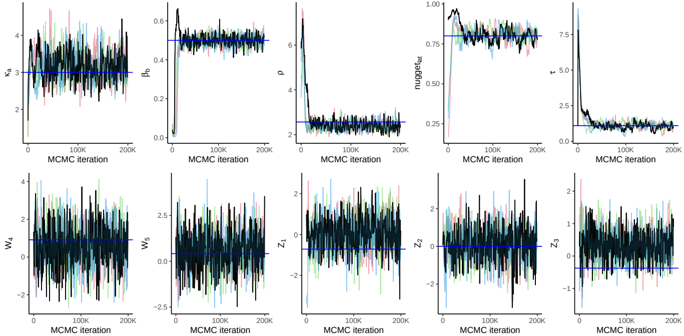


Figure 5: Traceplots of four Markov chains for selected parameters.

Goodness-of-fit for simulated data

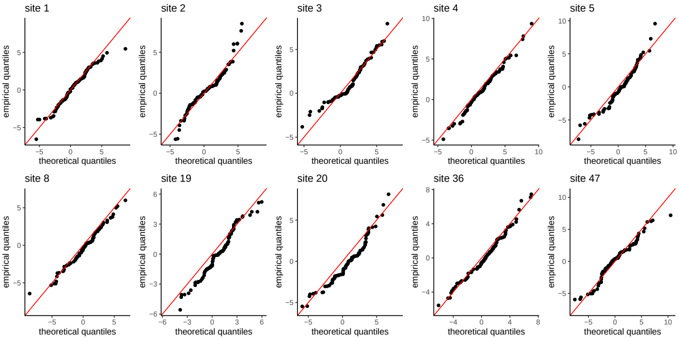


Figure 6: Quantile-quantile plots of marginal standardized observations for holdout data (top) and in-sample sites (bottom) for simulated data.

Joint estimation and marginal transformations

Write G_j and F_j for the distribution functions at site $\boldsymbol{s}_j$

- $Y(\boldsymbol{s}_j) \mid Y(\boldsymbol{s}_0) > G_0^{-1}(q)$ (data scale) and
- $X(\boldsymbol{s}_j) \mid X(\boldsymbol{s}_0) > -\log(-q)$ (standardized)

The likelihood contribution on the standardized scale is

$$p\left(y_j \mid \boldsymbol{z}, \boldsymbol{\Theta}\right) \propto \begin{cases} J_j \phi\left[F_j^{-1}\{G_j(y_j)\}; \mathbf{a}_j(x_0) + \mathbf{b}_j(x_0) z_j, \tau^2\right], & y_j > q_j \\ \Phi\left[F_j^{-1}\{G_j(q_j)\}; \mathbf{a}_j(x_0) + \mathbf{b}_j(x_0) z_j, \tau^2\right], & y_j \leq q_j \end{cases},$$

where J_j is the Jacobian of the marginal transformation.

Marginal model misspecification

Since the theoretical framework requires data to have exponential tails, observations are typically mapped to the unit Laplace scale.

But if we integrate out the random effects, the conditional distribution of the d-vector of observations is

$$\begin{aligned} \boldsymbol{X} \mid X(\boldsymbol{s}_0) = x_0 > u &\sim \mathsf{No}_d\{\boldsymbol{a}(x_0), \mathbf{V}(x_0)\}, \\ X(\boldsymbol{s}_0) - u \mid X(\boldsymbol{s}_0) > u &\sim \mathsf{Exp}(1). \end{aligned}$$

This hierarchical formulation characterizes the marginal distributions of $X_0(\boldsymbol{s})$ – conditional on $X(\boldsymbol{s}_0) > u$.

Marginal transformation

We need to evaluate both

- the quantile function F_j^{-1} and
- the density f_j

of $X(\boldsymbol{s}_j) \mid X(\boldsymbol{s}_0) > u$, pointwise at every site $\boldsymbol{s}_j (j = 1, \ldots, d)$.

Approximating conditional density and quantiles

Interchanging the order of integration,

$$f_j(x) = \int_u^\infty \phi\left\{\frac{x-\mu(x_0)}{\tau}\right\}\exp(-x_0+u)\mathrm{d}x_0$$

which suggests the Monte Carlo estimator

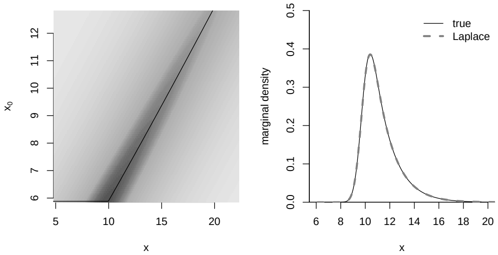
$$\widehat{f}_j(X) \approx \frac{1}{B}\sum_{b=1}^{B}\phi\left\{\frac{x-\mu(X_{0b})}{\tau}\right\}, \qquad X_{0b} \sim \mathsf{Exp}(1)+u$$

For the quantile function, we draw B observations from the joint model and approximate F_j^{-1} using the marginal empirical quantile of X_j.

Laplace approximation to marginal

Approximate the denominator of (1) by a Gaussian distribution (Laplace approximation)[1]

$$p(\boldsymbol{X}_0(s) = x) = \frac{p(\boldsymbol{X}(s) = x, X_0(\boldsymbol{s}_0) = x_0)}{p(X_0(\boldsymbol{s}_0) = x_0 \mid \boldsymbol{X}_0(s) = x)}. \tag{1}$$



[1]Compute the conditional mode $x_0^*(x) = \max_{x_0 \in [u,\infty)} f(x, x_0)$ and replace the denominator by a *truncated* Gaussian distribution above u.

Design of experiments

We need to estimate the quantile surface or density for

- many values of x and
- (arbitrary) parameter combinations

Estimation can be done offline using our emulator!

- links with computer experiments in experimental design (e.g., Latin hypercube sampling)
- we could approximate the surface using a suitable neural network?

Summary and future work

- Use Gaussian Markov random field residual process with data augmentation to effectively deal with left-censoring
- Efficient full-likelihood-based inference based on Markov chain Monte Carlo sampling

Work in progress includes

- efficient proposals when geometric anisotropy is weak,
- application to daily precipitation data from British Columbia,
- more efficient implementation of the joint estimation scheme,
- comparison with the two-stage approach and with INLA.

Funding acknowledgement

CANADA FIRST RESEARCH EXCELLENCE FUND | APOGÉE CANADA FONDS D'EXCELLENCE EN RECHERCHE

NSERC CRSNG

IVADO

Calcul Québec

Thank you for your attention. Questions, comments, suggestions?

References I

- Lindgren, Finn, Håvard Rue, and Johan Lindström (2011). "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4, pp. 423–498. DOI: 10.1111/j.1467-9868.2011.00777.x.
- Richards, Jordan, Jonathan A. Tawn, and Simon Brown (2022). "Modelling extremes of spatial aggregates of precipitation using conditional methods". In: *The Annals of Applied Statistics* 16.4, pp. 2693–2713. DOI: 10.1214/22-AOAS1609.
- Simpson, Emma S, Thomas Opitz, and Jennifer L. Wadsworth (2023). "High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and the SPDE approach". In: *Extremes*.

References II

- Vandeskog, Silius M, Sara Martino, and Raphaël Huser (2022). "An Efficient Workflow for Modelling High-Dimensional Spatial Extremes". In: *arXiv*. DOI: 10.48550/ARXIV.2210.00760.
- Wadsworth, Jennifer L. and Jonathan A. Tawn (2022). "Higher-dimensional spatial extremes via single-site conditioning". In: *Spatial Statistics* 51, p. 100677. DOI: 10.1016/j.spasta.2022.100677.
- Zhang, Likun, Benjamin A. Shaby, and Jennifer L. Wadsworth (2022). "Hierarchical Transformed Scale Mixtures for Flexible Modeling of Spatial Extremes on Datasets With Many Locations". In: *Journal of the American Statistical Association* 117.539, pp. 1357–1369. DOI: 10.1080/01621459.2020.1858838.