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)

30 mai 2023, LSCE, Paris–Saclay

HEC MONTRÉAL

Motivation: 2021 British Columbia floods

Figure 1: Aerial pictures of flooding in Abbotsford and Chilliwack, British Columbia, November 23, 2021. Province of British Columbia, CC license

Studying rare events

CBC reports on flooding caused by an atmospheric river in November 2021:

- many officials called the storm that hit the province a once-in-a-century event.
- 24 B.C. communities received more than 100mm of rain from Saturday to Monday.
- The town of Hope led the way with 252mm over that time period.

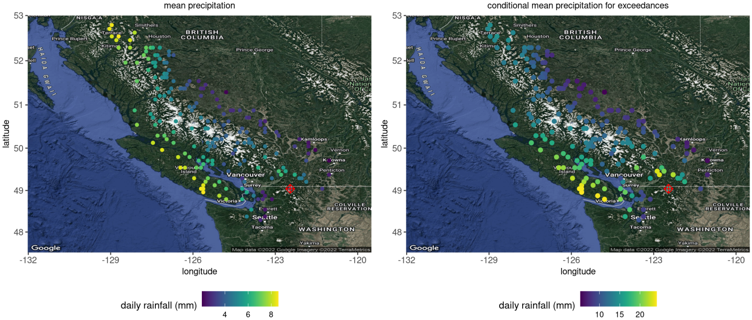


Figure 2: Average daily rainfall and conditional average given large rainfall near Abbotsford (BC). Pacific Climate Impacts Consortium daily gridded meteorological dataset NRCANMET (1950-2012).

Justification for modelling spatial extremes

Flooding can be caused by either

- heavy localized rainfalls
- large-scale events

Both

- the site-wise behaviour of rainfall and
- the spatio-temporal dependence

are of interest.

Objective and stylized facts

The objective of this work is to construct a stochastic generator and create catalogues of daily cumulative rainfall fields given exceedance at one site.

Asymptotic (in)dependence

Correctly accounting for the dependence regime at large levels is crucial. Consider the marginal quantile function F_j^{-1} at site $\boldsymbol{s}_j$.

The tail correlation coefficient is

$$\chi_u(\boldsymbol{s}_i, \boldsymbol{s}_0) = \Pr\left\{X(\boldsymbol{s}_i) > F_i^{-1}(u) \mid X(\boldsymbol{s}_0) > F_0^{-1}(u)\right\}, \qquad u \in [0, 1].$$

Asymptotic (in)dependence

Correctly accounting for the dependence regime at large levels is crucial. Consider the marginal quantile function F_j^{-1} at site $\boldsymbol{s}_j$.

The tail correlation coefficient is

$$\chi_u(\boldsymbol{s}_i, \boldsymbol{s}_0) = \Pr\left\{X(\boldsymbol{s}_i) > F_i^{-1}(u) \mid X(\boldsymbol{s}_0) > F_0^{-1}(u)\right\}, \qquad u \in [0,1].$$

We broadly characterize processes based on tail correlation $\chi = \lim_{u\to 1} \chi(u)$, with

- $\chi > 0$ (asymptotic dependence)
- $\chi = 0$ (asymptotic independence).

Asymptotic dependence regime

Extreme rainfall events tend to become more spatially localized as their intensity increases.

Preliminary checks suggestive of asymptotic independence with strong anisotropy.

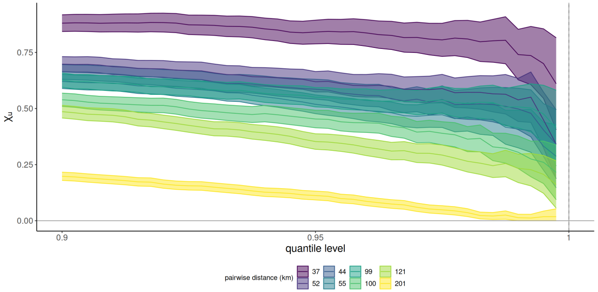


Figure 3: Tail correlation coefficient for selected sites

Conditional spatial extremes model

- We focus on the spatio-temporal conditional extremes model of Wadsworth and Tawn (2022).
- This model describes the limiting behaviour of a stochastic process $X(\boldsymbol{s})$ with exponential tails given that the value at a particular site within the spatial domain is large.
- Why? the model can capture both asymptotic dependence (at short range) and asymptotic dependence.

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

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

To handle spatial dependence, we consider a Gaussian residual random field $Z(\boldsymbol{s})$.

COSPEX model

- Conditional on the random field being extreme at site $\boldsymbol{s}_0$, we assume that the suitably renormalized process $X(\boldsymbol{s})$ converges in distribution to a non-degenerate spatial process $Z(\boldsymbol{s})$ satisfying
  - $Z(\boldsymbol{s}_0) = 0$ almost surely and
  - $Z(\boldsymbol{s})$ independent of $\lim_{u\to\infty} X(\boldsymbol{s}_0) - u \mid X(\boldsymbol{s}_0) > u \sim \mathsf{Exp}(1)$.

Normalizing functions

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

There are many possible choices of normalizing functions: for example, 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 to now, only two-stage estimation (margins estimated first, then transformed to standard Laplace scale) has been considered.
  - Can use a parametric model (spatial pooling) or empirical distribution for the marginals.
- Under stationarity assumption, parameters are the same regardless of the conditioning site. Can use composite likelihood to pool exceedances (repeated observations).
- Mostly frequentist approach for the inference, except Simpson, Opitz, and Wadsworth (2023) and Vandeskog, Martino, and Huser (2022) who use INLA.

Shortcomings of COSPEX

- This model shares the same computational bottlenecks for left-censoring (e.g., zero-inflated data) as other spatial extreme value models, although applicable to lower levels.
- Censoring is only considered in Richards, Tawn, and Brown (2022) and extension thereof, using triplewise composite likelihood.

Computational challenges due to censoring

When we censor observations that are too low, the likelihood contains high dimensional Gaussian integrals.

Since the censoring pattern changes from one observation to the next, we need to do repeated matrix decomposition to extract the marginal and conditionals.

Computational bottleneck: inference is limited to ~ 30 sites with censoring.

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

The benefit of this approach is that $X_i \equiv X_0(\boldsymbol{s}_i)$ $(i = 1, \ldots, d)$ are conditionally independent given $\boldsymbol{Z}$ so, 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 $\mu_j = a(\boldsymbol{s}_j, x_0) + b(\boldsymbol{s}_j, x_0)Z_j$.

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.

Objectives

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 with 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.

Part 1 of the agenda

For computational feasibility, we need to reduce the costs associated with the spatial random effects.

Ideally, we would like $Z(\boldsymbol{s})$ to be a Gaussian Markov random field.

Why? In this case, the precision matrix will be sparse (as zero entries encode conditional independence).

Then, the conditional distribution of $Z_j \mid \boldsymbol{Z}_{-j}$ depends only on few neighbours (much lower dimension).

Sparse residual process

Expanding on the work of Simpson, Opitz, and Wadsworth (2023) (previous talk), we consider the SPDE approximation of the Matérn field Lindgren, Rue, and Lindström (2011)

$$Z(\boldsymbol{s}) = \sqrt{r_Z}\sum_{k=1}^{K} \phi_k(\boldsymbol{s}) W_k + \sqrt{(1-r_Z)}\epsilon_Z,$$

or in vector form, $\boldsymbol{Z} = \sqrt{r_Z}\mathbf{A}\boldsymbol{W} + \sqrt{(1-r_Z)}\boldsymbol{\epsilon}$.

Sparse residual process

Expanding on the work of Simpson, Opitz, and Wadsworth (2023) (previous talk), we consider the SPDE approximation of the Matérn field Lindgren, Rue, and Lindström (2011)

$$Z(\boldsymbol{s}) = \sqrt{r_Z} \sum_{k=1}^{K} \phi_k(\boldsymbol{s}) W_k + \sqrt{(1 - r_Z)} \epsilon_Z,$$

or in vector form, $\boldsymbol{Z} = \sqrt{r_Z} \mathbf{A} \boldsymbol{W} + \sqrt{(1 - r_Z)} \boldsymbol{\epsilon}$.

The resulting discrete approximation is a Gaussian Markov random field.

- $\boldsymbol{W} \sim \mathsf{No}_K(0_K, \mathbf{Q}^{-1})$ are Gaussian weights,
- the sparse precision matrix $\mathbf{Q}$ depends on the mesh,
- $\{\phi_k\}$ are (compactly-supported) piecewise linear basis functions with associated projector matrix $\mathbf{A}$,
- $\epsilon_Z \sim \mathsf{No}(0, 1)$ are i.i.d. white noise (nugget).

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.

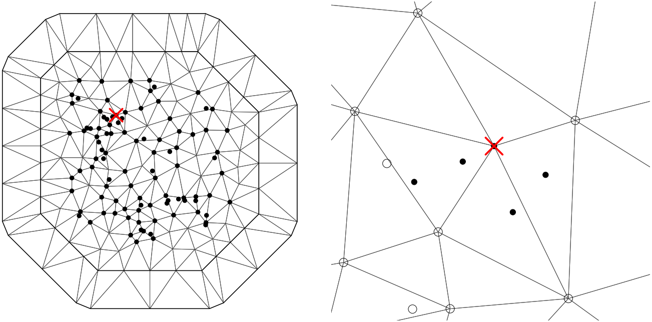


Figure 4: Mesh for SPDE approximation with conditioning site (red cross)

In sparsity we trust

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

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

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, giving

$$p\left(X_j \mid \boldsymbol{Z}, \boldsymbol{\Theta}_a, \boldsymbol{\Theta}_b, \tau\right) = \begin{cases} \phi\left\{x_j; \mathbf{a}_j(x_0) + \mathbf{b}_j(x_0)z_j, \tau^2\right\}, & x_j > q_j \\ \Phi\left\{q_j; \mathbf{a}_j(x_0) + \mathbf{b}_j(x_0)z_j, \tau^2\right\}, & x_j \leq q_j \end{cases}$$

$$\boldsymbol{Z} \mid \boldsymbol{W}, r_Z \sim \mathsf{No}_d\left\{\sqrt{r_Z}\mathbf{A}\boldsymbol{W}, (1 - r_Z)\mathbf{I}_d\right\};$$

$$\boldsymbol{W} \mid r_Z, \rho \sim \mathsf{No}_K(0_K, r_Z\mathbf{Q}^{-1});$$

$$\boldsymbol{\Theta} \sim \pi(\boldsymbol{\Theta}).$$

We can also marginalize over weights $\boldsymbol{W}$ and compute the unconditional precision of $\boldsymbol{Z}$ efficiently using Sherman–Morrisson–Woodbury formula (Nychka et al., 2015).

Parameter estimation

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 $\boldsymbol{\Theta}$.

For bounded parameters in $[a, b]$, we use truncated Gaussian proposals or work on a transformed scale.

Proposal variance are tuned during burn-in period.

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

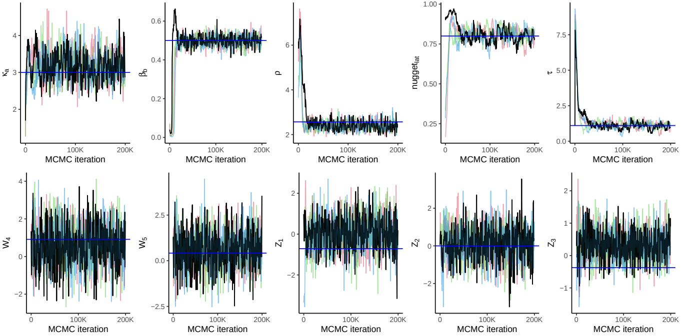


Figure 5: Traceplots of four Markov chains for selected parameters.

Goodness-of-fit for simulated data

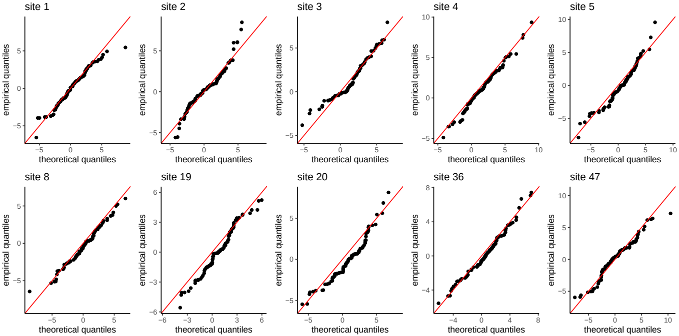


Figure 6: Quantile-quantile plots of marginal standardized observations for holdout data (top) and in-sample sites (bottom) for simulated data.

Marginal model misspecification

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

The problem is that this is not compatible with the model given an exceedance of the conditioning site.

Incorporating 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 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 $\boldsymbol{s}_j (j = 1, \ldots, d)$.

Marginal distribution of standardized variables

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 $X_0(\boldsymbol{s})$ (recall, conditional on $X(\boldsymbol{s}_0) > u$).

Approximation of the conditional density and quantile

Interchanging the order of integration,

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

where $\mu(X_0) = a(X_0) + b(X_0)Z_j$, which suggests the Monte Carlo estimator

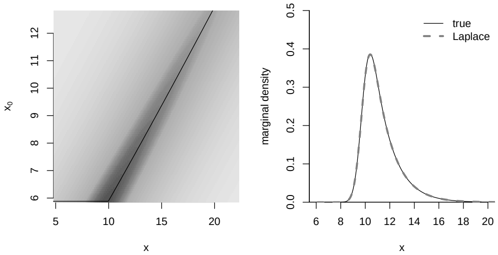
$$f_j \approx \frac{1}{B}\sum_{b=1}^{B}\phi\left\{\frac{x-\mu(X_{0b})}{\tau}\right\}, \qquad X_{0b} \sim \mathsf{Exp}(1) + u$$

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]

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



[1] For given x, we 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.

Computational bottlenecks

- In the simulation, we tackle problems with 1000 sites and 100 replicates. Using only the dependence part, sampling 500K samples from the posterior takes about 10 hours.
- When we add margins, evaluation of the marginal quantile and density functions increase the time per likelihood evaluation by about 20 seconds...
- Possible remedies: parallelization and C++ (work in progress)

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

- application to daily precipitation data from British Columbia
- more efficient implementation of the joint estimation scheme.
- comparison with the two-stage approach and INLA

Funding acknowledgement

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.

Nychka, Douglas et al. (2015). "A multiresolution Gaussian process model for the analysis of large spatial datasets". In: Journal of Computational and Graphical Statistics 24.2, pp. 579–599. DOI: 10.1080/10618600.2014.914946.

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.

References II

- 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.
- 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.