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

Léo Belzile, Rishikesh Yadav, Nicholas Beck

HEC Montréal

May 11, 2023
IMAG-BIRS workshop on Modern Statistical and Machine Learning Approaches for High-Dimensional Compound Spatial Extremes

HEC MONTRĒAL

Motivation

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

2021 British Columbia floods

CBC reports on a storm caused by an atmospheric river in November 2021

- many officials calling 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.

Data for BC lower mainland

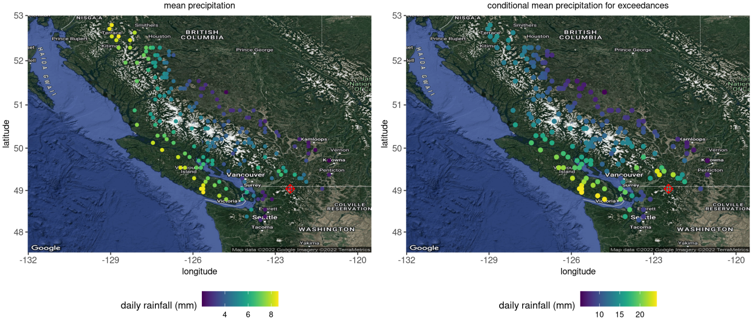


Figure 2: Average daily rainfall and conditional average given large rainfall near Abbotsford (BC). PCIC daily gridded meteorological dataset NRCANMET (1950-2012).

Asymptotic dependence regime

Preliminary checks suggestive of asymptotic independence with strong anisotropy

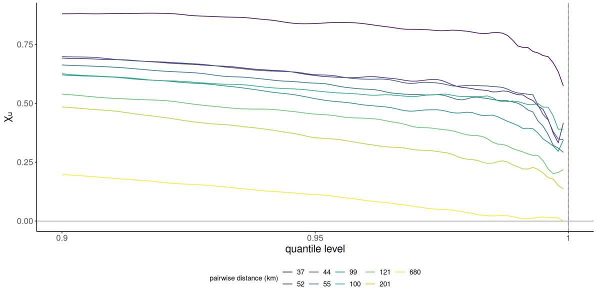


Figure 3: Tail correlation coefficient for selected sites

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

Spatial model

- We focus on the spatio-temporal conditional extremes model of Wadsworth and Tawn (2022).
- This model describes the stochastic behaviour of a spatial process given the value at a particular site within the spatial domain is large.

Model

- Consider a stationary spatial process $\{X(\boldsymbol{s}), \boldsymbol{s} \in \mathscr{S}\}$ with exponential tails
- 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
  - $Z(\boldsymbol{s}_0)$ independent of $\lim_{u\to\infty} X(\boldsymbol{s}_0) - u \mid X(\boldsymbol{s}_0) > u \sim \mathsf{Exp}(1)$.

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

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

We add a nugget term $\varepsilon \sim \mathsf{No}(0, \tau^2)$ to facilitate data augmentation (Zhang, Shaby, and Wadsworth, 2022)

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

Normalizing functions

Wadsworth and Tawn (2022) list conditions for $a(\cdot)$ and $b(\cdot)$ that guarantee valid limiting models, e.g., taking

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

with ρ any correlation function.

In the sequel, we consider

- $b(\boldsymbol{s}, x) = x^\beta$ for $\beta \in (0, 1)$ (doesn't decay with distance)
- an exponential correlation function with geometric anisotropy,

$$\rho(d) = \exp(-d/\kappa_a), \qquad d = \|\mathbf{A}\boldsymbol{h}\|.$$

Recent advances and shortcomings of COSPEX

- Two-stage estimation (margins to standard Laplace)
- Mostly frequentist approach for the inference, except Simpson, Opitz, and Wadsworth (2023) (INLA) and Vandeskog, Martino, and Huser (2022)
- Same computational bottlenecks for left-censoring (e.g., zero-inflated data) as other spatial extreme value models. Censoring only considered Richards, Tawn, and Brown (2022) and extension thereof.

Framework and objectives

- Establish a computationally feasible way for handling large-scale inference in the presence of censoring
- Use full likelihoods with data augmentation (Bayesian paradigm)
- Simultaneous estimation of marginal parameters and dependence
  - lot's of uncertainty in the margins
  - reduce model misspecification

Sparse residual process

Expanding on the work of Simpson, Opitz, and Wadsworth (2023) (previous talk), we consider the SPDE approximation (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 $\boldsymbol{Z} = \sqrt{r_Z}\mathbf{A}\boldsymbol{W} + \sqrt{(1 - r_Z)}\boldsymbol{\epsilon}$

- $\boldsymbol{W} \sim \mathsf{No}_K(0_K, \mathbf{Q}^{-1})$ 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. (nugget)

Mesh

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.

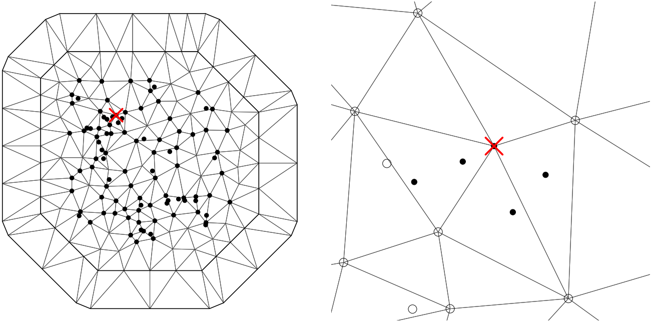


Figure 4: Mesh for SPDE approximation with conditioning site (red cross)

Believe in sparsity

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

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

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

The conditional distribution of basis weight W_k given the others is

$$
W_k \mid \{\boldsymbol{W}_{-k} = \boldsymbol{w}_{-k}\} \sim \mathsf{No}\left(-\mathbf{Q}_{kk}^{-1}\mathbf{Q}_{k,-k}\boldsymbol{w}_{-k}, \mathbf{Q}_{kk}^{-1}\right), \qquad (1)
$$

Can also marginalize over weights $\boldsymbol{W}$ (Nychka et al., 2015).

Hierarchical model formulation

Observations are conditionally independent given $\boldsymbol{Z}$

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

We censor observations below quantiles q_j.

Simulation study and estimation

We generate data from the model and censor observations below their 75 percentile.

We use Markov chain Monte Carlo methods to draw posterior samples

- Gibb's sampling for the weights $\boldsymbol{W}$
- random walk Metropolis–Hastings, MALA and second-order approximations for $\boldsymbol{Z}$
- (transformed scale/truncated normal proposals, block updates)

Results shown next don't include marginal transformations or anisotropy

Convergence diagnostics

There is strong autocorrelation between the parameters of the normalizing functions.

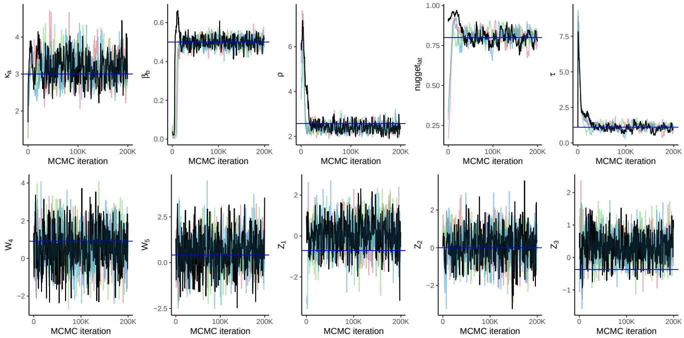


Figure 5: Traceplots of Markov chains for selected parameters.

Goodness-of-fit for simulated data

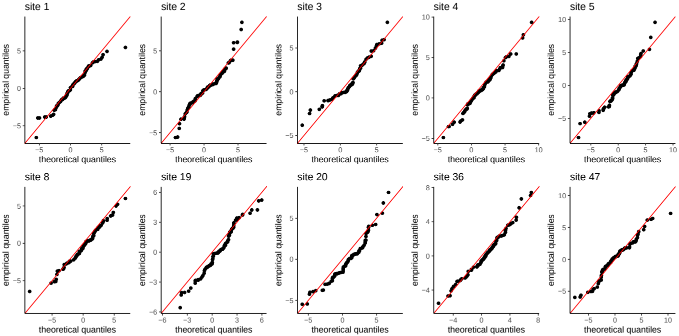


Figure 6: Quantile-quantile plots of marginal standardized observations for holdout data (top) and in-sample sites (bottom) for simulated data.

Marginal distribution of standardized variables

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

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

where

$$\mathbf{V}(x_0) = \mathrm{diag}\{\mathbf{b}(x_0)\}\{r_Z\mathbf{A}\mathbf{Q}_{\rho}^{-1}\mathbf{A}^\top + (1-r_Z)\mathbf{I}_d\}\mathrm{diag}\{\mathbf{b}(x_0)\} + \tau\mathbf{I}_d$$

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

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

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

where J_j is the Jacobian of the marginal transformation.

Marginal transformation

We need to evaluate the quantile function F_j^{-1} and the density f_j pointwise for every site $\boldsymbol{s}_j$ (bottleneck).

Interchanging the order of integration, evaluate using Monte Carlo

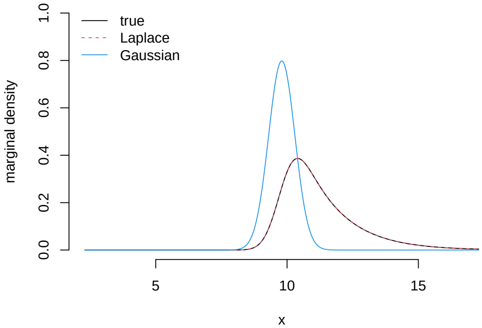
$$f_j \approx \frac{1}{B}\sum_{b=1}^{B} \phi\left\{\frac{x-\mu(X_{0b})}{\tau}\right\}, \qquad X_{0b} \sim \mathsf{Exp}(1) + u$$

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

Laplace approximation to marginal

Approximate the denominator of (2) by a Gaussian distribution (Laplace approximation)[1]

$$p(\boldsymbol{X}_0(s) = x) = \frac{p(\boldsymbol{X}(s) = x, X_0(\boldsymbol{s}_0) = x_0)}{p(X_0(\boldsymbol{s}_0) = x_0 \mid \boldsymbol{X}_0(s) = x)}. \qquad (2)$$



[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

- Using only the dependence part, sampling 500K samples from the posterior takes about 10 hours
- With $d = 1000$ sites and $n = 100$ time points, evaluation of the marginal quantile and density increase the time per iteration 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 include;

- solve identifiability issue when geometric anisotropy in a is weak
- application to daily precipitation data from British Columbia accounting for the non-stationarity and seasonality
- more efficient implementation of the joint estimation scheme.
- comparison with the two-stage approach

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.