

Climate Model Output of Extreme Rainfall

Léo Belzile

joint work with Rishikesh Yadav, Nicholas Beck, Jordan Richards

HEC Montréal

June 4, 2024

HEC MONTRÉAL



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

CBC reports on a storm 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.

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

Stylized facts:

- Extreme rainfall events tend to become more spatially localized as their intensity increases (asymptotic independence).
- Rainfall is zero-inflated, with either no rain recorded or a positive amount.

Regional climate model (RCM) output

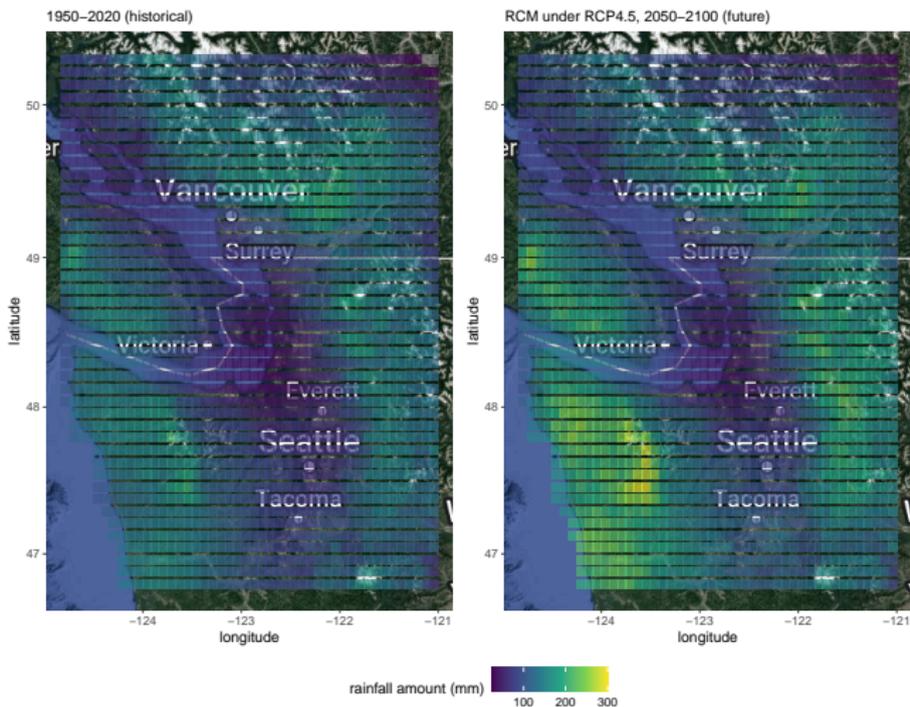
- RCM data from the Pacific Climate Impact Consortium Canadian Downscaled Climate Scenarios - Multivariate (CMIP6)
- The downscaled data from one ensemble member includes data for representative concentration pathway (RCP) 4.5
 - the historical period (1950–2020) and
 - future (until 2100).
- Data include multivariate quantile correction

Environment Canada and NOAA stations have sparse coverage, with less than 100 stations for the region of interest, and uneven temporal coverage and missing values.

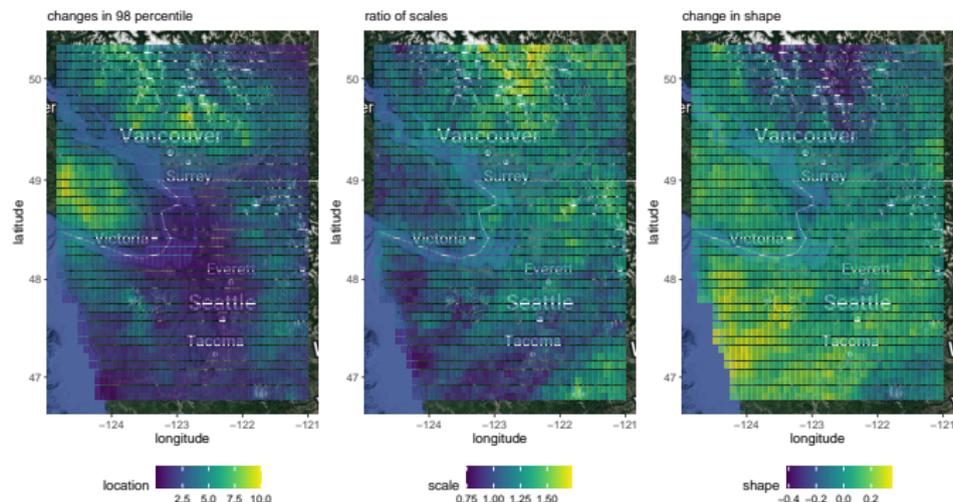
Return level maps based on RCM data

We restrict attention to October–February rainfalls.

Cumulated daily rainfall 50 years return levels



Comparing return levels for historical (1950-2020) versus future data from RCP4.5 (2050-2100) suggest an increase in very extreme events.



Extremal dependence - tail correlation

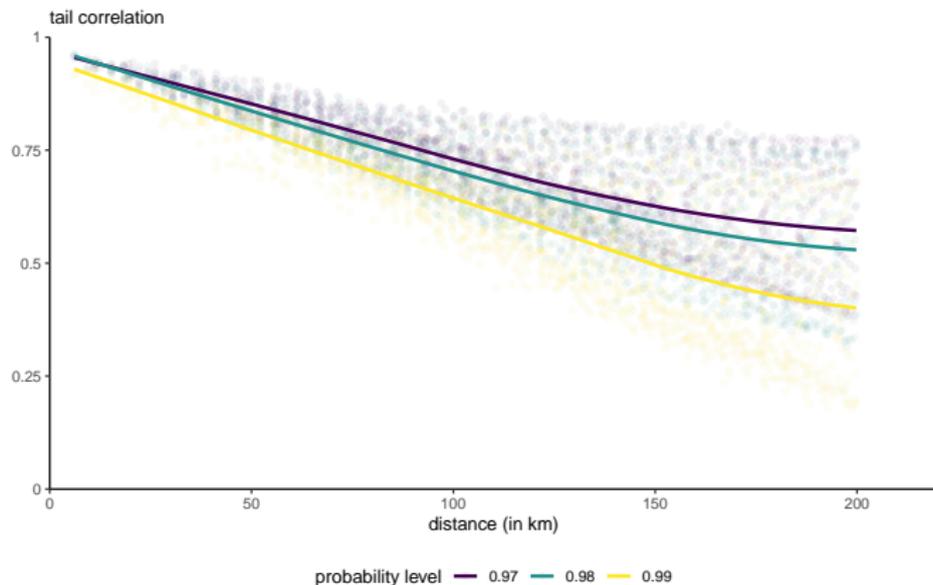


Figure 3: Conditional pairwise tail correlation, relative to grid point closest to Abbotsford. There is strong evidence of spatial heterogeneity.

RCM data exhibit asymptotic independence at intermediate distances.

RCM data are, by construction

- non-stationary
- smoother than real cumulated daily rainfall fields
- exhibit more extremal dependence than station data (?)

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.

- We focus on the conditional spatial extremes (COSPEX) model of **Wadsworth.Tawn:2022**.
- The COSPEX model describes the stochastic behaviour of a spatial process given the value at a particular site s_0 within the spatial domain is large.

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

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

$$X_0(\mathbf{s}) \stackrel{d}{\approx} a\{\mathbf{s}, X(\mathbf{s}_0)\} + b\{\mathbf{s}, X(\mathbf{s}_0)\}Z(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S}.$$

Location function $a(\cdot)$ and scale function $b(\cdot)$ depends on

- location (distance to conditioning site s_0)
- intensity of event $X(s_0)$

- High spatial resolution of gridded data:
 - around 2K locations at 6km distance
- Small temporal resolution
 - roughly 200 temporal replications (3 per year) above 0.98 quantile at s_0

For extremes, this is a high-dimensional problem.

Rainfall can be zero,

- but model doesn't allow for it! (inferential left-censoring)
- Same computational bottlenecks 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).

Add a nugget term to facilitate data augmentation
(**Zhang.Shaby.Wadsworth:2022**),

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

where

$$\varepsilon \sim \text{Normal}(0, \tau^2)$$

- Two-stage estimation (margins mapped to standard Laplace scale).
- Mostly frequentist approach for the inference, except **Simpson.Opitz.Wadsworth:2023** and **Vandeskog.Martino.Huser:2022**, who use INLA.
- Censoring only considered **Richards.Tawn.Brown:2022** and extension thereof, using composite triplewise likelihoods.

- 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.

Expanding on the work of **Simpson.Opitz.Wadsworth:2023**, we consider the SPDE approximation (**Lindgren.Rue.Lindstrom:2011**)

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

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

- $\mathbf{W} \sim \text{Normal}_K(\mathbf{0}_K, \mathbf{Q}^{-1})$ 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} ,
- $\epsilon_Z \sim \text{Normal}(\mathbf{0}, 1)$ are i.i.d. (nugget).

Piecewise linear approximation

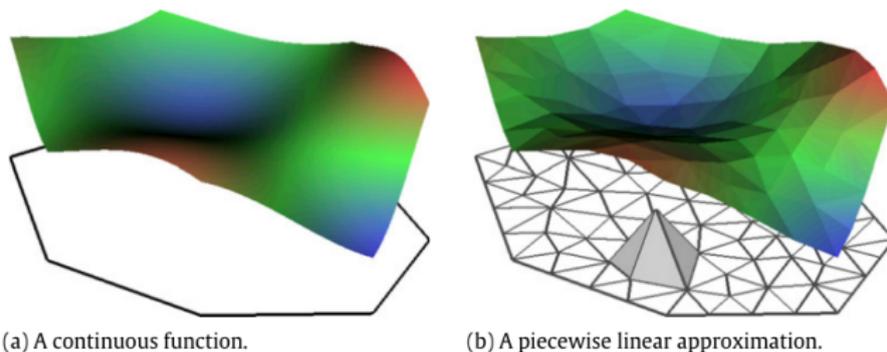


Fig. 2. Piecewise linear approximation of a function over a triangulated mesh.

Figure 4: Piecewise linear approximation to SPDE (image courtesy of Finn Lindgren)

As in **Vandeskog.Martino.Huser:2022**, we create a mesh with a node at s_0 , extract the precision matrix and shed it.

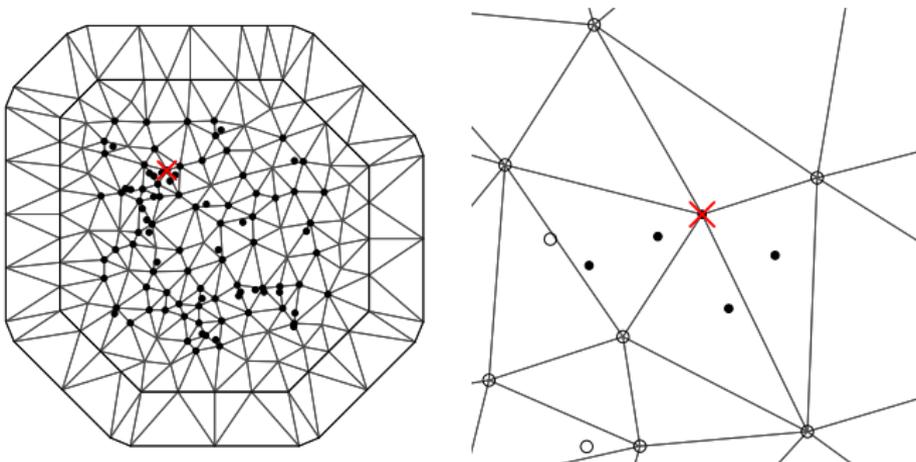


Figure 5: Mesh for SPDE approximation with conditioning site (red cross)

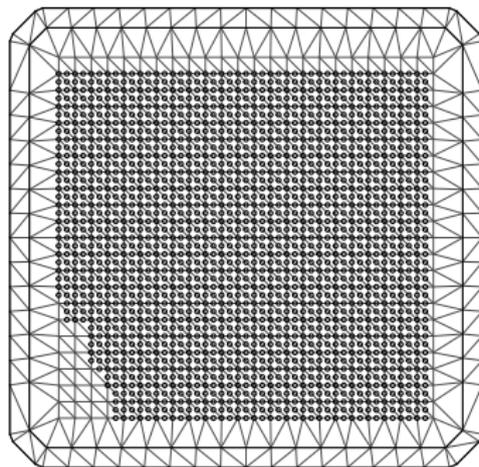


Figure 6: Mesh for RCA data

Reduces number of "locations" from 1910 sites to 688 nodes.

Leveraging the sparsity leads to efficient data augmentation building on **Zhang,Shaby,Wadsworth:2022**

$$\begin{aligned} X_0(\mathbf{s}_j) \mid Z_j &\sim \text{Normal} \left\{ \mathbf{a}_j(x_0) + \mathbf{b}_j(x_0)Z_j, \tau^2 \right\}, \\ Z_j &\sim \text{Normal}(\sqrt{r_Z} \mathbf{A}_{j, \text{ne}(j)} \mathbf{w}_{\text{ne}(j)}, 1 - r_Z) \end{aligned}$$

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

$$W_k \mid \{\mathbf{W}_{-k} = \mathbf{w}_{-k}\} \sim \text{Normal} \left(-\mathbf{Q}_{kk}^{-1} \mathbf{Q}_{k,-k} \mathbf{w}_{-k}, \mathbf{Q}_{kk}^{-1} \right), \quad (1)$$

Can also marginalize over weights \mathbf{W} (**Nychka:2015**).

Observations are conditionally independent given \mathbf{Z}

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

$$\mathbf{Z} | \mathbf{W}, r_Z \sim \text{Normal}_d \left\{ \sqrt{r_Z} \mathbf{A} \mathbf{W}, (1 - r_Z) \mathbf{I}_d \right\};$$

$$\mathbf{W} | r_Z, \rho \sim \text{Normal}_K(0_K, r_Z \mathbf{Q}^{-1});$$

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

We censor observations below quantiles q_j .

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 \mathbf{W}
- random walk Metropolis–Hastings, MALA and second-order approximations for \mathbf{Z}
- (transformed scale/truncated normal proposals, block updates)

Convergence diagnostics

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

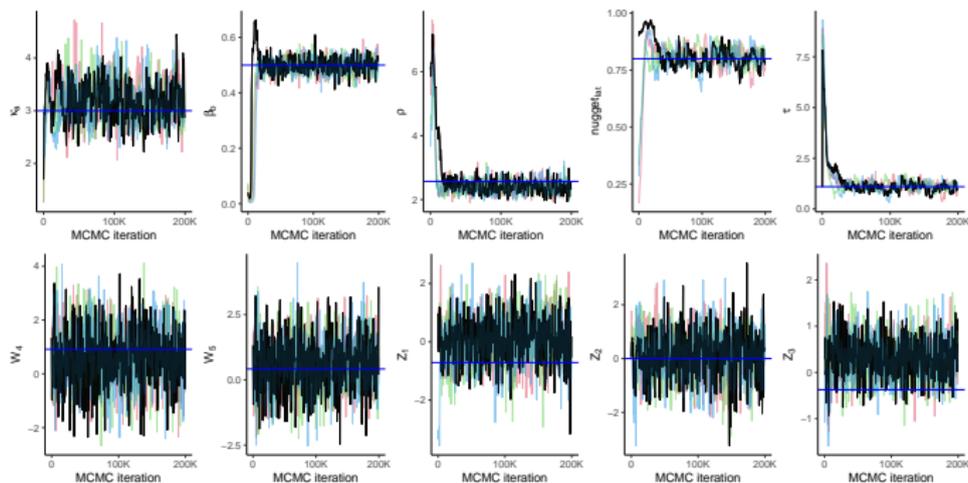


Figure 7: Traceplots of Markov chains for dependence parameters.

Sampling 500K samples from the posterior takes about 10 hours with 200 sites and 100 time replications, but mixing is slow.

Goodness-of-fit for simulated data

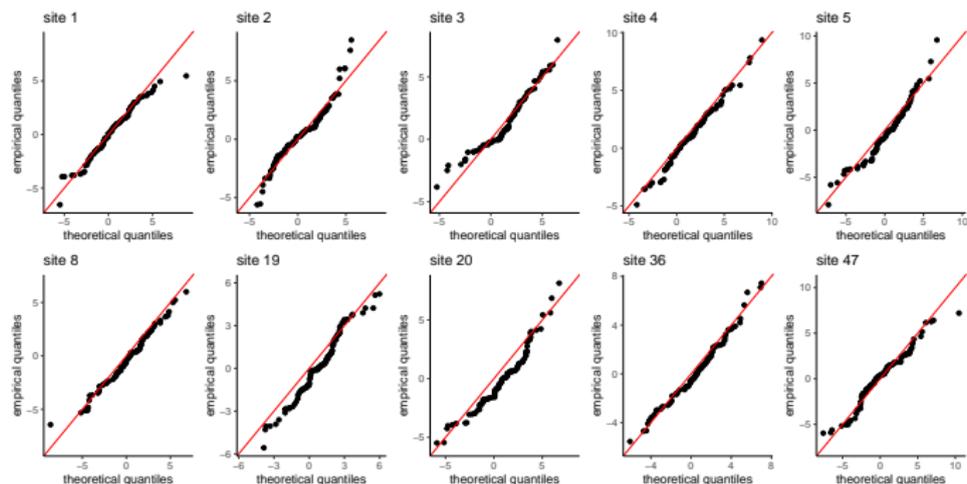


Figure 8: Quantile-quantile plots of marginal standardized observations for holdout data (top) and in-sample sites (bottom) for simulated data.

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

$$\begin{aligned} \mathbf{X} \mid X(\mathbf{s}_0) = x_0 > u &\sim \text{Normal}_d\{\mathbf{a}(x_0), \mathbf{V}(x_0)\}, \\ X(\mathbf{s}_0) \mid X(\mathbf{s}_0) > u &\sim \text{Exp}(1) \end{aligned}$$

where

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

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

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

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

The likelihood contribution on the standardized scale is

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

where J_j is the Jacobian of the marginal transformation.

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

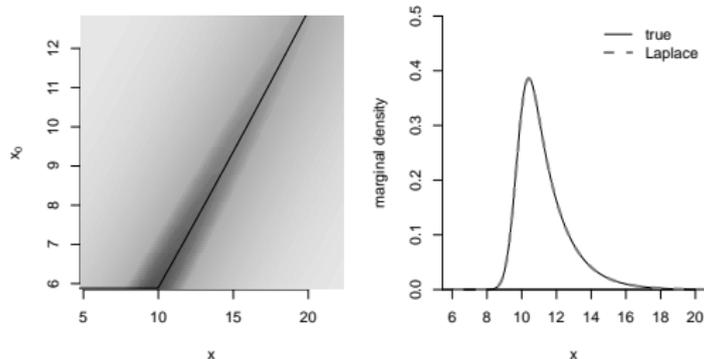
Interchanging the order of integration, evaluate using Monte Carlo

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

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

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

$$p(\mathbf{X}_0(s) = x) = \frac{p(\mathbf{X}(s) = x, X_0(\mathbf{s}_0) = x_0)}{p(X_0(\mathbf{s}_0) = x_0 | \mathbf{X}_0(s) = x)}. \quad (2)$$



¹For 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 .

- 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...
 - We approximate marginal distribution by skewed-unified normal (SUN) that minimizes KL divergence with Laplace approximation
 - Train a neural network to learn parameters for different combinations of parameters Θ , enforcing monotonicity constraints

- 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 \boldsymbol{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



**CANADA
FIRST**
RESEARCH
EXCELLENCE
FUND

**APOGÉE
CANADA**
FONDS
D'EXCELLENCE
EN RECHERCHE



IVADO



Calcul Québec

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

References I