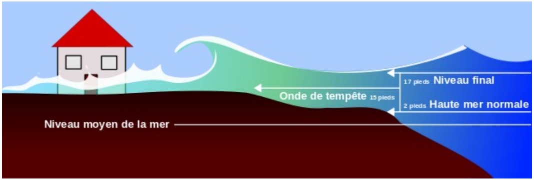
Simulation of Extreme Events in the Presence of Spatial Dependence

Nicholas Beck
Marie-Pier Côté
Martin Durocher
Jonathan Jalbert
Bouchra Nasri
Juliana Schulz
Marie-Hélène Toupin
Véronique Tremblay
Fateh Chebana
Jean-François Plante
Jean-François Quessy
Nouredine Daili

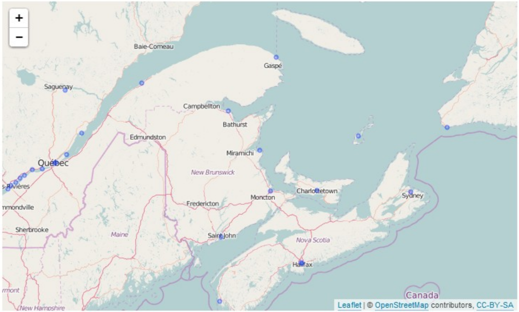
Desjardins General Insurance Group

Problem

- Evaluate the risk of **coastal floods** (from salt water)
- Deterministic model for the propagation of water in land
- Create a model to **simulate water levels** on the coast
- Capture spatial dependence to evaluate risk properly



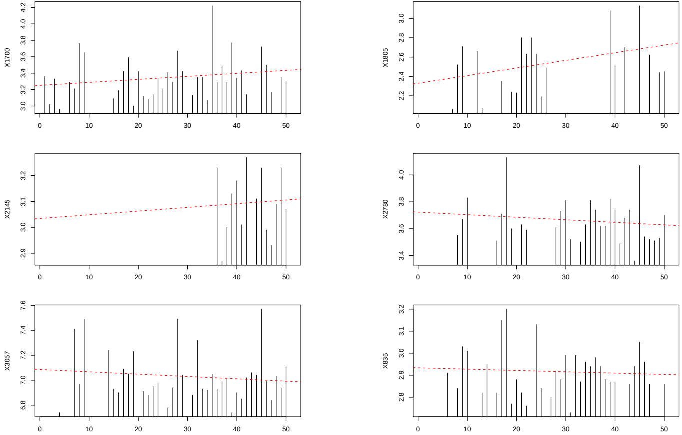
Data and related choices



- Measure from tide gauges
- 21 locations, east of Québec City
- (at most) 50 years of data (much fewer in some locations)
- Interest in Z, the **annual maximum** water level
- L, the normal (theoretical) **level of a high tide** is known
- **Longitude** and **latitude** used for location

Working hypothesis #1

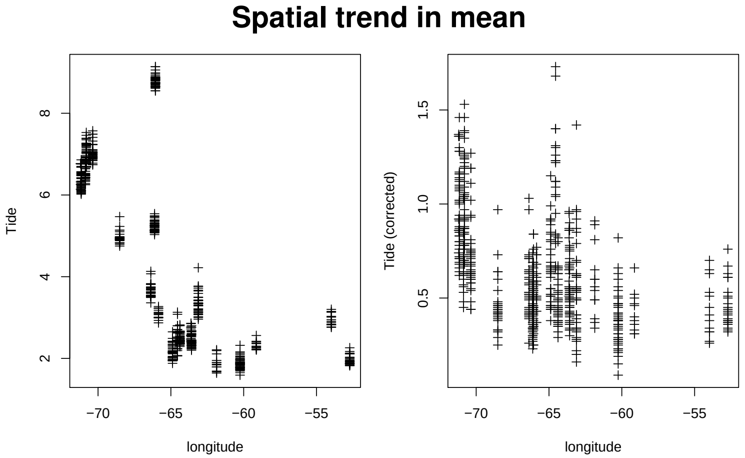
No temporal trends in the annual maxima



Shown above : Annual maximum vs year at 6 different sites.

Climate change ignored : insurance contract are for one year.

Working hypothesis #2



Annual maxima vs longitude.

Left : raw data (Z) ; Right : corrected for tide level ($Z - L$)

Notation

Notation

- Let a random fields $Z(s)$, where $Z(s)$ are annual maxima at location $s \in \mathbb{R}^2$
- Joint distribution

$$H(z_1, \ldots, z_n) = Pr\left\{Z(s_1) \leq z_1, \ldots, Z(s_n) \leq z_n\right\}$$

Looking for a model for H

Plan

- Marginal model — Generalized Extreme Value
- Spatial model — *t* Copula
- Solution #1 : Frequentist approach
- Solution #2 : Bayesian approach
- Conclusion

Extreme value theory

Theorem (Fisher-Tippett)

Let $Y_1, \dots, Y_m$ be an iid sequence of variables and $Z_m = \max\{Y_1, \dots, Y_m\}$. If there exists some $b_m > 0$ and $a_m \in \mathbb{R}$ such that

$$\frac{Z_m - a_m}{b_m} \rightsquigarrow H,$$

for some non-degenerate distribution H, then H has a generalized extreme value (GEV) distribution, that is

$$F(y;\mu,\sigma,\xi) = \begin{cases} \exp\left\{-(1+\xi\frac{y-\mu}{\sigma})^{-1/\xi}\right\}, & \xi \neq 0,\ 1+\xi\frac{y-\mu}{\sigma} > 0 \\ \exp\left\{-\exp\left(-\frac{y-\mu}{\sigma}\right)\right\}, & \xi = 0,\ y \in \mathbb{R}. \end{cases}$$

In particular, when m is large enough,

$$Z_m \approx GEV(\mu, \sigma, \xi)$$

and this approximation holds true even if $Y_1, \ldots, Y_m$ is not independent.

- short/medium range autocorrelation "disappears" when taking maximum (yay !)

Copula approach

Decomposition

$$H(z_1, \ldots, z_{21}) = C_{\Sigma_{\theta,\gamma}} \{F_{\beta_1}(z_1), \ldots, F_{\beta_{21}}(z_{21})\}$$

- $C_{\Sigma_{\theta,\gamma}}$ is a copula and F_β are marginal distributions
- $(\Sigma_{\theta,\gamma})_{ij} = g_\gamma\left(\frac{\delta_{ij}}{\theta}\right)$
- Isotropic random field

Copula approach

Student copula

- Non zero upper tail dependence

$$\lambda_U = \lim_{u \to 1^-} Pr\left\{X_1 > F_1^{-1}(u) \mid X_2 > F_2^{-1}(u)\right\}$$

- Close under margins

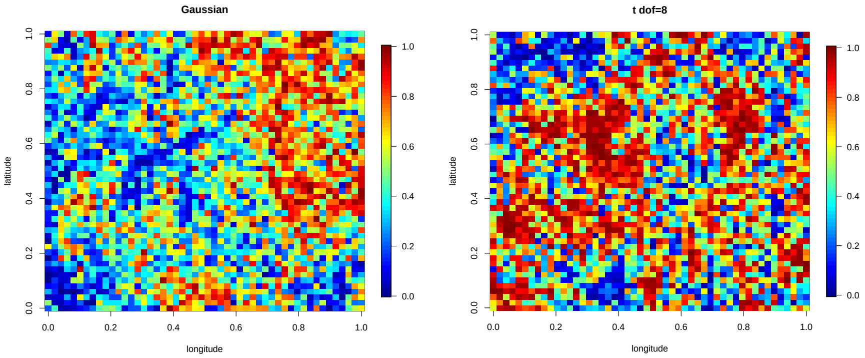
Power exponential model

$$g_\nu(\delta) = \exp\left[-3\left(\frac{\delta}{\theta}\right)^\gamma\right]$$

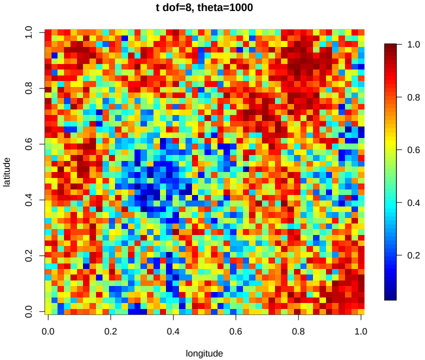
where θ is a range parameter and γ is a smoothness parameter.

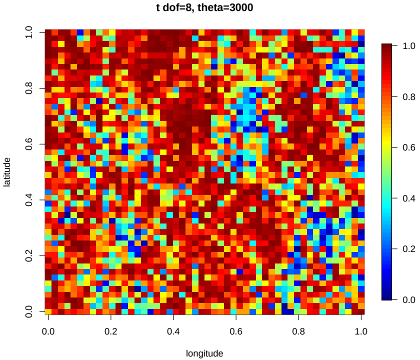
Copula approach

Effect of the tail dependence

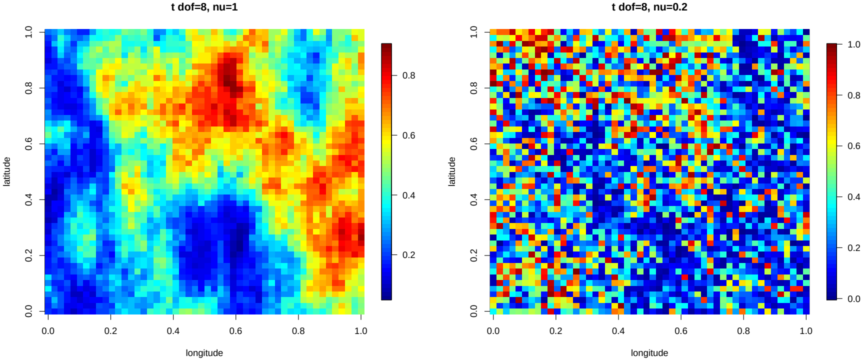


Effect of the range parameter θ





Effect of the smoothness parameter γ



Frequentist model

Marginal GEV model

- Marginal distribution with covariates

$$(Z - L) = Z^* \rightarrow \mathrm{GEV}\left(\mu\left(s_{lon}\right), \sigma, \xi\right)$$

- Location : $\mu\left(s_{lon}\right) = a_0 + a_1\ s_{lon} + a_2\ s_{lon}^2$
- Constant scale σ and shape ξ
- Likelihood for the marginal

$$\mathcal{L}(a_0, a_1, a_2, \sigma, \xi) = \prod_{k=1}^{n^*} f(z_k; a_0, a_1, a_2, \sigma, \xi)$$

where n^* = number of year-site

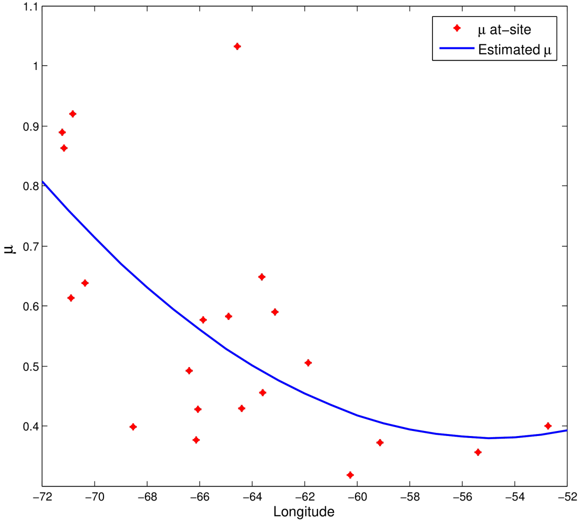
Frequentist model

Marginal GEV model

- $\widehat{\mu} = 4.74 + 0.16 s_{long} + 0.002 s_{long}^2$
- $\widehat{\sigma} = 0.20$
- $\widehat{\xi} = 0.07$

Frequentist model

Estimation of the location parameters μ



Frequentist model

Copula estimation

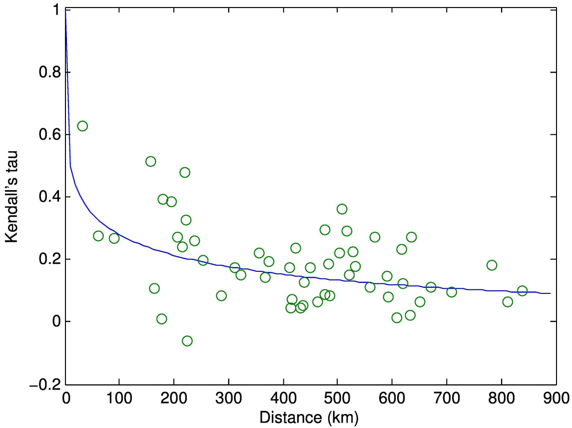
- Marginal parameters are known : $\widehat{\beta}_i = (\widehat{\mu}_i, \widehat{\sigma}, \widehat{\xi})$
- Second step : Pseudo-likelihood

$$\mathcal{L}(\theta, \gamma) = \prod_{i=1}^{n} c^{(d_i)}_{\Sigma_{\theta,\gamma}} \left\{ F_{\widehat{\beta}_1}(Z_{i,1}), \ldots, F_{\widehat{\beta}_{d_i}}(Z_{i,d_i}) \right\},$$

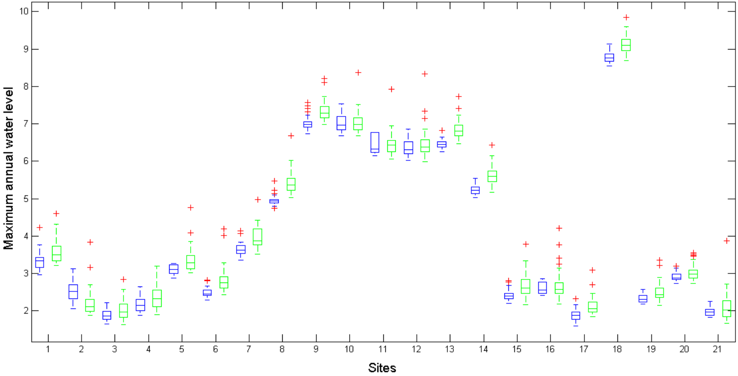
where d_i is the number of non missing sites at time $i = 1, \ldots, n$.

Frequentist model

Evolution of the dependence ($\hat{\theta} = 2850$, $\hat{\gamma} = .37$ $df = 21$)



Boxplot of simulated (green) vs true (blue) values



Hierarchical Bayesian model

1) Data level

$\mathbf{Z}(\mathbf{s})|\theta$ has copula C_θ and

$$Z(s_i)|\mu(s_i), \sigma, \xi \sim GEV(\mu(s_i), \sigma, \xi)$$

2) Process level

Non-informative priors are used for σ, ξ and $\theta = (\theta_1, \theta_2)$ and $(\boldsymbol{\mu}(\mathbf{s})|\boldsymbol{\beta}, \tau, \varrho) \sim \mathcal{N}_d$.

3) Hyper priors

$\boldsymbol{\beta} = (\beta_0, \beta_1)$ and ϱ have non-informative priors and τ has a non-informative prior.

Hierarchical Bayesian model

Given the parameters $\boldsymbol{\mu}, \sigma, \xi$ and $\boldsymbol{\theta}$, the joint distribution of the maxima is

$$F_{\mathbf{Z}|\boldsymbol{\mu},\sigma,\xi,\boldsymbol{\theta}}(\mathbf{z}) = C_{\boldsymbol{\theta}}\left[F\{z(s_1)|\mu_1,\sigma,\xi\},\ldots,F\{z(s_d)|\mu_d,\sigma,\xi\}\right],$$

where

- $C_{\boldsymbol{\theta}}$ is a t copula
- F is the cdf of the GEV distribution.

Fitting is performed by MCMC.

Hierarchical Bayesian model — copula parameters

The degrees of freedom of the t copula are fixed.

The correlation depends on the distance between the stations :

$$\rho_Z(s_i, s_j) = \theta_1^{\mathrm{dist}(s_i,s_j)^{\theta_2}},$$

with f_{θ_1} is $\mathcal{U}(0,1)$ and f_{θ_2} is $\mathcal{U}(0,2)$.

Hierarchical Bayesian model — marginal parameters

At the process level, we use non-informative priors for σ and ξ, while, for $\mathbf{s} = (s_1, \ldots, s_d)$,

$$\boldsymbol{\mu}(\mathbf{s}) \sim \mathcal{N}_d(X(\mathbf{s})\boldsymbol{\beta}, \Sigma(\mathbf{s})),$$

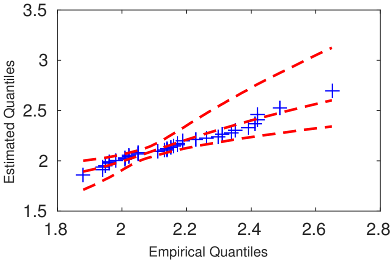
where

- $X(s_i)\boldsymbol{\beta} = \beta_0 + \beta_1 x_i$ and x_i is the deterministic predicted maximum tide level at station s_i,
- $\Sigma(\mathbf{s}) = (\rho_\mu(s_i, s_j))_{d \times d}$, with

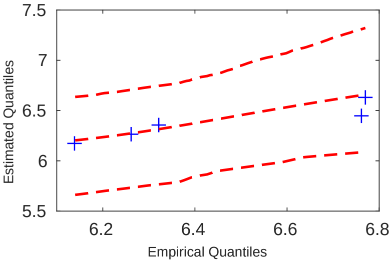
$$\rho_\mu(s_i, s_j) = \tau \exp(-\text{dist}(s_i, s_j)/\varrho).$$

A non-informative prior was used for $\boldsymbol{\beta}$,
f_ϱ is $\mathcal{U}(0, \max_{i,j}\{\text{dist}(s_i, s_j)\})$ and $f_\tau(\tau) \propto 1/\tau$.

Goodness of fit testing



(a) Station 2



(b) Station 11

FIGURE – QQ Plots for GEV margins

Goodness of fit testing (cont.)

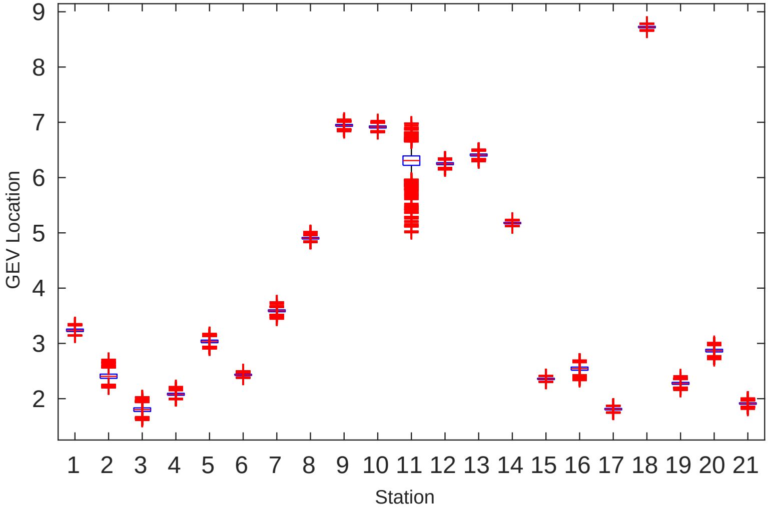


FIGURE – Boxplots for stations' location parameter

Simulation

Simulation from this model can be done with the following (simple !) algorithm :

(1) Generate the parameters $\boldsymbol{\beta}$, τ, ϱ, $\boldsymbol{\mu}(\boldsymbol{s})$, σ, ξ and $\boldsymbol{\theta}$

(2) Generate realization from the copula C_θ, i.e. $\mathbf{u} = (u_1, ..., u_d)$

(3) Using the GEV margins, invert the copula realizations using $x_i = F^{-1}(u_i|\mu(s_i), \sigma, \xi)$, $i = 1, ..., d$.

Can also be used for interpolation between sites as the hierarchical structure accounts for spatiality.

Conclusion

- Applicable model with reasonnable results
- Scalable
- Annual maxima vs multiple claims the same year
- Room to improve the modelling of μ and σ
- Need to check the fit on simultaneous data
- Great team !