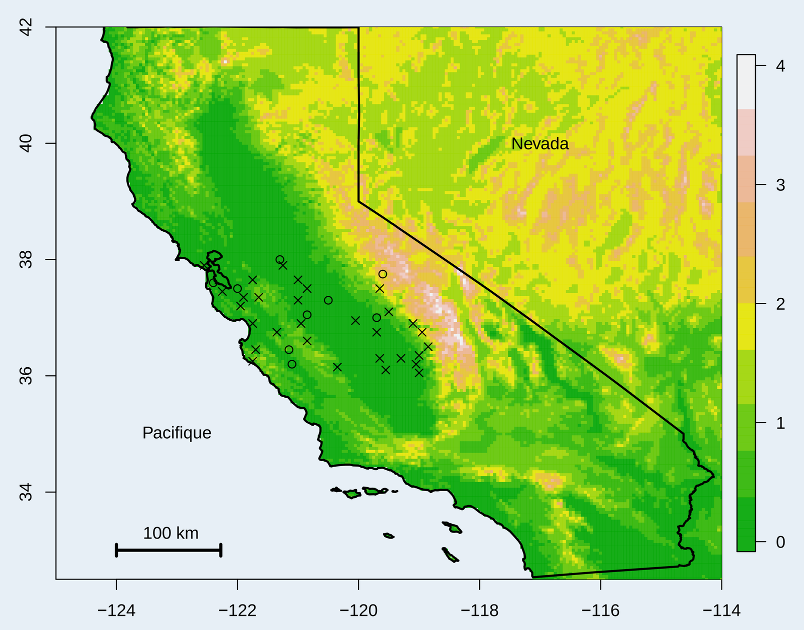
Approximate Bayesian Computing to the estimation of max-stable processes with application to extremes rainfall in a small region of California

Martin Durocher, Mathieu Ribatet et Taha Ouarda

INRS Université d'avant-garde

Introduction

- **Objective** : Compute return level maps
- **Subject** : Extreme rainfalls (annual maximums)
- **Study region** : California - elevation (km)



X : Estimation (31 sites), O : Validation (8 sites)

- **Challenge** : Spatial dependency among sites
- **Solution** : Spatial model from max-stable processes

Max-stable processes

- **Objective** : Generalizing univariate extreme value theory
- **Definition**: Y is a **max-stable process** if there exist functions $a_n(x)$ et $b_n(x) > 0$ such as Y is equal in distribution to

$$\frac{Z_n(x) - a_n(x)}{b_n(x)}, \qquad \forall n$$

with

$$Z_n(x) = \max_{i=1,\dots,n} Y_i(x)$$

- **Representation** :

$$Z(x) = \max_{i=1,2,\dots} \lambda_i Y_i(x)$$

λ_i : Poisson process

Y_i : Underlying stochastic processes

- **Interpretation** : Maximum over an infinite number of storms (Y_i) occurring as a Poisson process
- **Challenge** : Full likelihood is unknown

Approximate Bayesian Computing (ABC)

- **Objective** : Sampling from posterior distribution

$$\pi(\psi \mid T(y)) \approx \pi(\psi, y^* \mid d(T(y), d(T(y^*)) < \epsilon)$$

y: Observations

T: Summary variables

d: Proxy distance

ϵ: Tolerance value

- **Remark** : Classic Bayesian framework :

$$\epsilon = 0, y = T(y)$$

- **Approximate rejection algorithm**

1. Sample ψ_i from prior distribution $\pi(\psi)$
2. Simulate data $y_i = \eta(\psi_i)$
3. Accept ψ_i if $d\,(T(y), T(y_i)) < \epsilon$

- **Post-treatment** :

Regression model : $s_i = T(y_i)$

$$\psi_i = m(s_i) + v(s_i)e_i$$

m : Conditional mean

v^2 : Conditional variance

$m(s), v(s)$ are true posterior moments when $s = T(y)$

Adjusting the parameter :

$$\psi_i^* = m(s) + v(s)\frac{\psi_i - m(s_i)}{v(s_i)}$$

- **Pros** : Bayesian framework, robust
- **Cons** : Time consuming, Partial information

Pairwise likelihood (PL)

- **Definition**: **Pairewise log-likelihood**

$$l_p(\psi \mid \mathbf{z}) = \sum_{i=1}^{n} \sum_{j>k} \log f(z_{i,j}, z_{i,k} \mid \psi)$$

$\mathbf{z} = \{z_{i,j}\}$: observation year i, site j

f : Density function

ψ : Parameters

- **Properties**:

(i) Unbiased and Normally distributed (asymptotic)

(iii) Covariance matrix : $\hat{V} = H^{-1}JH^{-1}$

$$\text{Hessian} : H = -\mathbb{E}(\nabla^2 l_p(\hat{\psi}))$$

$$\text{Score} : J = \mathrm{Var}(\nabla l_p(\hat{\psi}))$$

- **Jackknife** : $\hat{\psi}_{-i}$ estimate without year i

$$\hat{V}_{jack} = \sum_{i=1}^{n} (\hat{\psi}_{-i} - \hat{\psi})(\hat{\psi}_{-i} - \hat{\psi})^T$$

- **Pros** : Fast, efficient
- **Cons** : Optimization difficult

Extremal coefficients

- **Objective** : Describe pairwise spatial dependency
- **Definition** : **Extremal coefficients** $\theta(h)$ respect

$$p\,(Z(x) < u, Z(x+h) < u) = \exp\,(-\theta(h)/u)$$

- **Interpretation**:

$\theta = 1$: Totally dependant pair of sites

$\theta = 2$: Independent pair of sites

- **Definition : F-madogram**

$$2\nu(h) = \mathbb{E}|F_x\,(Z(x)) - F_{x+h}\,(Z(x+h))\,|$$

F_x :Marginal distribution function of $Z(x)$

- **Property** :

$$\theta(h) = \frac{1 + 2\nu(h)}{1 - 2\nu(h)}$$

- **Remark** : F-madgram is used to estimate θ

Schlather model

- **Definition** : Max-stable process with Y_i as underlying Gaussian random fields.
- **Property** : Distribution function : $F(u, v) = e^{-V(u,v)}$

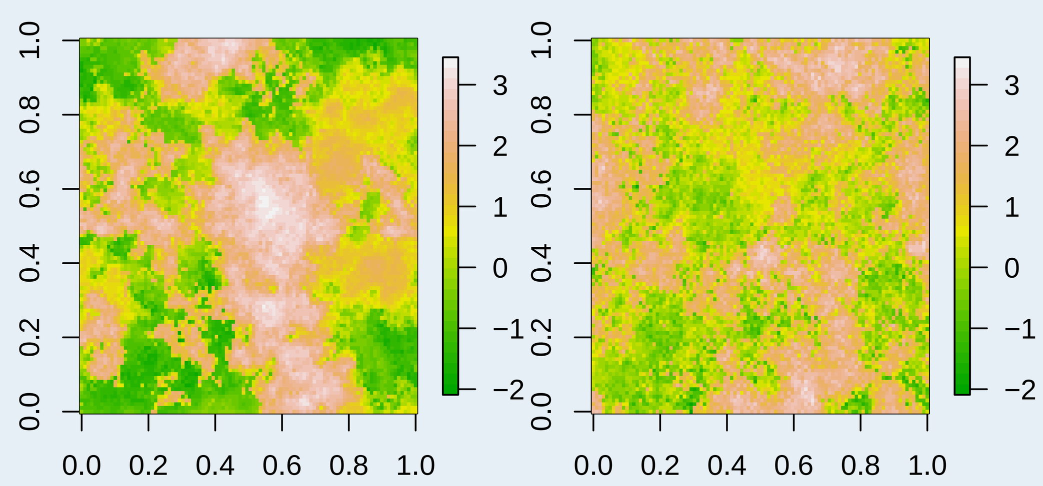
$$V(u, v) = \frac{1}{2}\left(\frac{1}{u} + \frac{1}{v}\right)\left(1 + \sqrt{1 - 2\,(\rho(h) + 1)\frac{uv}{(u+v)^2}}\right)$$

- **Remark** : $\theta(h) = 1 + \sqrt{\frac{1-\rho(h)}{2}}$

Does not reach complete independence. If $\rho(h) > 0$ and monotonously increasing

$$\theta(\infty) = 1 + \sqrt{1/2} \approx 1.7$$

- **Examples** :



Reference

[1] Michael Blum and Olivier Francois.
Non-linear regression models for approximate bayesian computation.
Statistics and Computing, 20(1):63–73, 2010.

[2] Dan Cooley, Philippe Naveau, and Paul Poncet.
Variograms for spatial max-stable random fields.
In *Dependence in Probability and Statistics*, volume 187 of *Lecture Notes in Statistics*, pages 373–390. Springer New York, 2006.

[3] S.A. Padoan, M. Ribatet, and S.A. Sisson.
Likelihood-based inference for max-stable processes.
Journal of the American Statistical Association, 105:263–277, 2010.

[4] M. Schlather.
Models for stationary max-stable random fields.
Extremes, 5(1):33–44, 2002.

Case study

- **Modèle** : Schlather with $\rho(h) = e^{-|h|/\alpha}$

GEV trend surface :

$$\mu(x) = \beta_{0,\mu} + \beta_{1,\mu}x_1 + \beta_{2,\mu}x_1^2 + \beta_{3,\mu}x_2 + \beta_{4,\mu}x_2^2 + \beta_{5,\mu}x_3$$

$$\sigma(x) = \beta_\sigma\,\mu(x)$$

$$\xi(x) = \beta_\xi$$

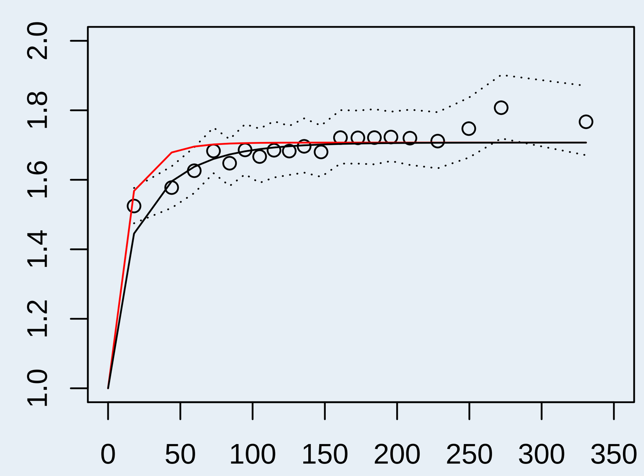
with (x_1, x_2, x_3) = (longitude, latitude, altitude)

- **ABC setting** :

T : F-madogram + L-moments

Prior : Uniform

- **F-madogram** : ABC(black), PL(red)

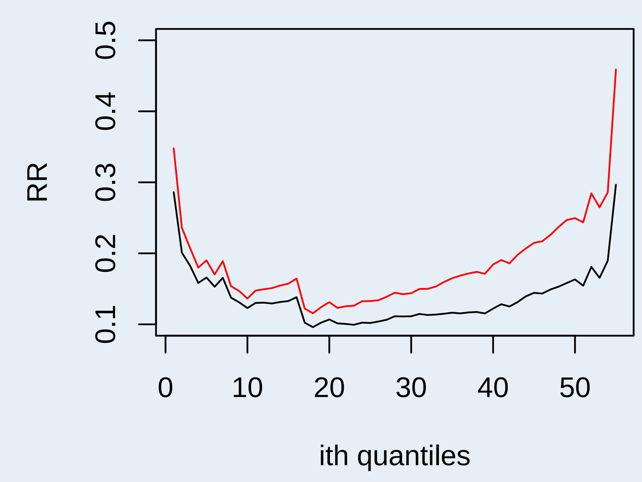


- **Evaluating Criteria**

$$RR^2 = E\left[\left(\frac{Q_i - \hat{Q}_i}{Q_i}\right)^2\right]$$

Q_i : ith sample quantile

$\hat{Q}_i$: ith theoretical quantile

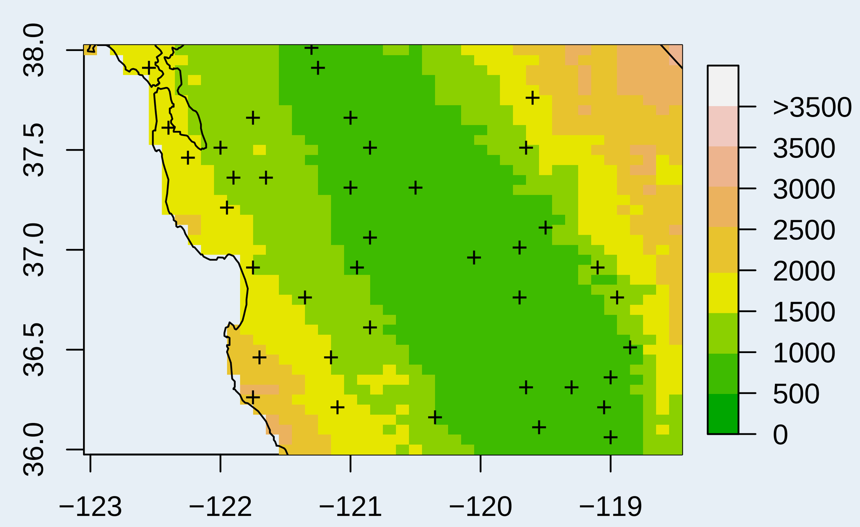


- : **Preferred method of estimation** : ABC

(i) Under estimation of close range dependency for PL estimate

(ii) Better RR for ABC with longer return period.

- **Return levels map** (ABC)



Return levels of 20 years (mm)

Conclusion

- Max-stable processes are realistic
- ABC is a good alternative to PL
- ABC allows complex non linear trend

Future work:

- Optimize simulation of max-stable process
- Find more informative summary statistics T

Contact information

Martin Durocher, Ph. D student
INRS-ETE, Université du Québec
490 de la couronne, Québec (Qc)
G1K 9A9 CANADA
Email : martin.durocher@ete.inrs.ca