Reparameterization of extreme value framework for improved Bayesian workflow

Théo Moins[1]

Joint with Julyan Arbel[1], Stéphane Girard[1], Anne Dutfoy[2]

[1]Statify, Inria Grenoble Rhône-Alpes

[2]EDF R&D dept. Périclès

June 13, 2023

Table of Contents

Motivations

Aim: Understand the risks of hazardous meteorological events.

Inondations : le Lot-et-Garonne touché par la "crue la plus importante depuis quarante ans"
(Source: `lemonde.fr`, Février 2021)

What about unobserved flows?

Motivations

More formally: Let $(X_1, \ldots, X_n)$ be i.i.d. random variables with distribution function $F = 1 - \bar{F}$, and $M_n = \max\{X_1, \ldots, X_n\}$, whose distribution is consequently F^n.

Motivations

More formally: Let $(X_1, \ldots, X_n)$ be i.i.d. random variables with distribution function $F = 1 - \bar{F}$, and $M_n = \max\{X_1, \ldots, X_n\}$, whose distribution is consequently F^n.

Estimation of extreme quantiles: for a given p_n such that $np_n \longrightarrow c < \infty$, we want to estimate $q_{p_n} = F^{-1}(1 - p_n)$.

Motivations

More formally: Let $(X_1, \dots, X_n)$ be i.i.d. random variables with distribution function $F = 1 - \bar{F}$, and $M_n = \max\{X_1, \dots, X_n\}$, whose distribution is consequently F^n.

Estimation of extreme quantiles: for a given p_n such that $np_n \longrightarrow c < \infty$, we want to estimate $q_{p_n} = F^{-1}(1 - p_n)$.

Return level: quantile of order $1 - \frac{1}{T}$ associated with a given return period T.
Example: a millennial return level corresponds to a quantile of order $1 - 0.001$.
$\hookrightarrow$ The return period is the average waiting time before the next occurrence associated to the return level.

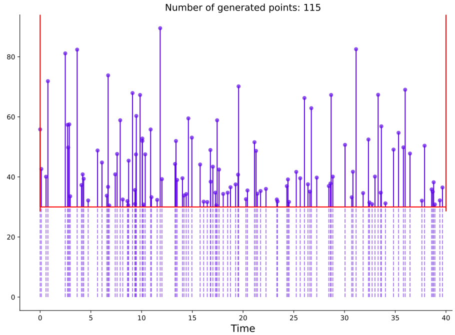
Poisson process characterisation of extremes

Let $(X_1, \dots, X_n)$ be i.i.d. r.v. and the associated point process N_n, evaluated on $I_u = [u, +\infty)$.

Theorem 1 (Coles (2001))

Under mild conditions and for a sufficiently large u, N_n can be approximated by a non-homogeneous Poisson process N of intensity measure Λ with parameters $\boldsymbol{\theta} = (\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}$, such that for all $x > u$,

$$\Lambda(I_x) = \int_x^{+\infty} \lambda(t)dt = \begin{cases} \left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}_+^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{if } \xi = 0. \end{cases}$$



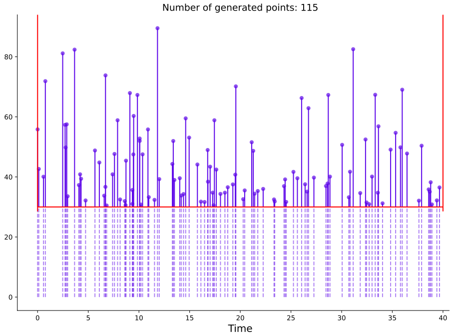
Poisson process characterisation of extremes

Let $(X_1, \ldots, X_n)$ be i.i.d. r.v. and the associated point process N_n, evaluated on $I_u = [u, +\infty)$.

Theorem 1 (Coles (2001))

Under mild conditions and for a sufficiently large u, N_n can be approximated by a non-homogeneous Poisson process N of intensity measure Λ with parameters $\boldsymbol{\theta} = (\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}$, such that for all $x > u$,

$$\Lambda(I_x) = \int_x^{+\infty} \lambda(t)dt = \begin{cases} \left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}_+^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{if } \xi = 0. \end{cases}$$



Inference on parameters (μ, σ, ξ) using Bayesian methods such as MCMC.

Table of Contents

Bayesian paradigm

Data above a threshold u: $\mathbf{x} := x_1, \cdots, x_{n_u}$

Bayesian paradigm

Data above a threshold u: $\mathbf{x} := x_1, \cdots, x_{n_u}$

PP model: $p(\mathbf{x}, n_u \mid \mu, \sigma, \xi) = \exp\left\{-m\left(1+\xi\left(\frac{u-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right\}\sigma^{-n_u}\prod_{i=1}^{n_u}\left(1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}-1}$

Bayesian paradigm

Data above a threshold u: $\mathbf{x} := x_1, \cdots, x_{n_u}$

PP model: $p(\mathbf{x}, n_u \mid \mu, \sigma, \xi) = \exp\left\{-m\left(1+\xi\left(\frac{u-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right\}\sigma^{-n_u}\prod_{i=1}^{n_u}\left(1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}-1}$

Prior uncertainty on $\boldsymbol{\theta} := (\mu, \sigma, \xi)$: $p(\mu, \sigma, \xi)$

Bayesian paradigm

Data above a threshold u: $\mathbf{x} := x_1, \cdots, x_{n_u}$

PP model: $p(\mathbf{x}, n_u \mid \mu, \sigma, \xi) = \exp\left\{-m\left(1+\xi\left(\frac{u-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right\}\sigma^{-n_u}\prod_{i=1}^{n_u}\left(1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}-1}$

Prior uncertainty on $\boldsymbol{\theta} := (\mu, \sigma, \xi)$: $p(\mu, \sigma, \xi)$

$\hookrightarrow$ **Bayesian update:**

$$p(\mu, \sigma, \xi \mid \mathbf{x}) \propto p(\mathbf{x} \mid \mu, \sigma, \xi) p(\mu, \sigma, \xi)$$

Bayesian paradigm

$$p(\mu, \sigma, \xi \mid \mathbf{x}) \propto p(\mathbf{x} \mid \mu, \sigma, \xi) p(\mu, \sigma, \xi)$$

What's next?

Bayesian paradigm

$$p(\mu, \sigma, \xi \mid \mathbf{x}) \propto p(\mathbf{x} \mid \mu, \sigma, \xi) p(\mu, \sigma, \xi)$$

What's next? All computations reduce to posterior means of quantity of interest $f(\boldsymbol{\theta})$:

$$\mathbb{E}_{p(\cdot \mid \mathbf{x})}[f(\boldsymbol{\theta})] = \int f(\boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta}$$

$\hookrightarrow$ Inference using MCMC algorithms.

Here, the quantities of interest is the T-year return level ℓ_T:

$$\ell_T = \mu - \frac{\sigma}{\xi}\left(1 - (-\log(1 - 1/T))^{-\xi}\right)$$

Bayesian paradigm

Main advantages (Coles and Powell, 1996):

- Consideration of expert information with informative prior,
- Can be used in any case, even when the likelihood is not available ($\xi < -1$),
- Access to the posterior predictive distribution:

$$p(\tilde{x} \mid \boldsymbol{x}) = \int \underbrace{p(\tilde{x} \mid \boldsymbol{\theta})}_{\substack{\text{randomness} \\ \text{of future obs.}}} \quad \underbrace{p(\boldsymbol{\theta} \mid \boldsymbol{x})}_{\substack{\text{parameter} \\ \text{uncertainty}}} \, d\boldsymbol{\theta}.$$

Bayesian paradigm

Main advantages (Coles and Powell, 1996):

- Consideration of expert information with informative prior,
- Can be used in any case, even when the likelihood is not available ($\xi < -1$),
- Access to the posterior predictive distribution:

$$p(\tilde{x} \mid \boldsymbol{x}) = \int \underbrace{p(\tilde{x} \mid \boldsymbol{\theta})}_{\substack{\text{randomness} \\ \text{of future obs.}}} \underbrace{p(\boldsymbol{\theta} \mid \boldsymbol{x})}_{\substack{\text{parameter} \\ \text{uncertainty}}} \, d\boldsymbol{\theta}.$$

Challenges:

- Convergence of MCMC algorithms?
- Choice of $p(\mu, \sigma, \xi)$ (in the informative and non-informative cases),

  $\hookrightarrow$ One possible solution: **Reparameterization**

Table of Contents

Reparameterization and MCMC

A reparameterization reshape the geometry of the likelihood.

Reparameterization and MCMC

A reparameterization reshape the geometry of the likelihood.

In particular, the correlation between the coordinates affects the convergence of MCMC algorithms:

Reparameterization and MCMC

A reparameterization reshape the geometry of the likelihood.

In particular, the correlation between the coordinates affects the convergence of MCMC algorithms:

- Gibbs and Metropolis–Hastings (Gilks et al., 1995),
- Hamiltonian Monte Carlo (HMC) (Betancourt, 2019)

REPARAMETERIZATION 91

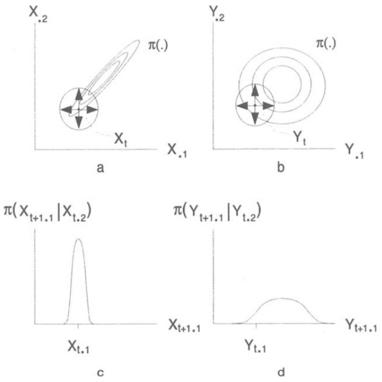


Figure 6.1 *Illustrating Gibbs sampling and Metropolis algorithms for a bivariate target density $\pi(.)$. Contours of $\pi(.)$: (a) before reparameterization; (b) after reparameterization. Full conditional densities at time t: (c) before reparameterization; (d) after reparameterization. See text for explanation.*

From Gilks et al. (1995)

Orthogonal parameters

Denote by $I(\boldsymbol{\theta})$ the Fisher information for $\boldsymbol{\theta}$:

$$I(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log p(\mathbf{x} \mid \theta) \mid \theta\right].$$

Definition of orthogonality according to Jeffreys (1961):

Parameter $\boldsymbol{\theta}$ are orthogonal $\iff$ $I(\boldsymbol{\theta})$ is diagonal.

Orthogonal parameters

Denote by $I(\boldsymbol{\theta})$ the Fisher information for $\boldsymbol{\theta}$:

$$I(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2} \log p(\mathbf{x} \mid \theta) \mid \theta\right].$$

Definition of orthogonality according to Jeffreys (1961):

$$\text{Parameter } \boldsymbol{\theta} \text{ are orthogonal} \iff I(\boldsymbol{\theta}) \text{ is diagonal.}$$

- Null asymptotic covariances between parameters,

Orthogonal parameters

Denote by $I(\boldsymbol{\theta})$ the Fisher information for $\boldsymbol{\theta}$:

$$I(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log p(\mathbf{x} \mid \theta) \mid \theta\right].$$

Definition of orthogonality according to Jeffreys (1961):

$$\text{Parameter } \boldsymbol{\theta} \text{ are orthogonal } \iff I(\boldsymbol{\theta}) \text{ is diagonal.}$$

- Null asymptotic covariances between parameters,
- Leads to asymptotic posterior independence when a Bernstein–von Mises theorem holds,

Orthogonal parameters

Denote by $I(\boldsymbol{\theta})$ the Fisher information for $\boldsymbol{\theta}$:

$$I(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log p(\mathbf{x} \mid \theta) \mid \theta\right].$$

Definition of orthogonality according to Jeffreys (1961):

$$\text{Parameter } \boldsymbol{\theta} \text{ are orthogonal} \iff I(\boldsymbol{\theta}) \text{ is diagonal.}$$

- Null asymptotic covariances between parameters,
- Leads to asymptotic posterior independence when a Bernstein–von Mises theorem holds,
- For more than 2 parameters: no general methods to find an orthogonal parameterization.

Orthogonal parameters

Denote by $I(\boldsymbol{\theta})$ the Fisher information for $\boldsymbol{\theta}$:

$$I(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log p(\mathbf{x} \mid \theta) \mid \theta\right].$$

Definition of orthogonality according to Jeffreys (1961):

$$\text{Parameter } \boldsymbol{\theta} \text{ are orthogonal} \iff I(\boldsymbol{\theta}) \text{ is diagonal.}$$

- Null asymptotic covariances between parameters,
- Leads to asymptotic posterior independence when a Bernstein–von Mises theorem holds,
- For more than 2 parameters: no general methods to find an orthogonal parameterization.

$\hookrightarrow$ No direct link between parameter orthogonality and mixing properties of MCMC chains.

Orthogonal parameters

For Poisson process characterization of extremes (Chavez-Demoulin and Davison, 2005):

Orthogonal parameters

For Poisson process characterization of extremes (Chavez-Demoulin and Davison, 2005):

$$(r, \nu, \xi) = \left(m \left(1 + \xi \left(\frac{u - \mu}{\sigma} \right) \right)^{-1/\xi}, (1 + \xi)(\sigma + \xi(u - \mu)), \xi \right).$$

Orthogonal parameters

For Poisson process characterization of extremes (Chavez-Demoulin and Davison, 2005):

$$(r, \nu, \xi) = \left(m \left(1 + \xi \left(\frac{u - \mu}{\sigma} \right) \right)^{-1/\xi}, (1+\xi)(\sigma + \xi(u - \mu)), \xi \right).$$

Likelihood:

$$p(\boldsymbol{x} \mid r, \nu, \xi) = e^{-r} \left(\frac{r}{m} \right)^{n} \left(\frac{\nu}{1+\xi} \right)^{-n_u} \prod_{i=1}^{n_u} \left(1 + \frac{\xi(1+\xi)}{\nu}(x_i - u) \right)^{-1-1/\xi},$$

Orthogonal parameters

For Poisson process characterization of extremes (Chavez-Demoulin and Davison, 2005):

$$(r, \nu, \xi) = \left(m \left(1 + \xi \left(\frac{u - \mu}{\sigma} \right) \right)^{-1/\xi}, (1+\xi)(\sigma + \xi(u - \mu)), \xi \right).$$

Likelihood:

$$p(\boldsymbol{x} \mid r, \nu, \xi) = e^{-r} \left(\frac{r}{m} \right)^{n} \left(\frac{\nu}{1+\xi} \right)^{-n_u} \prod_{i=1}^{n_u} \left(1 + \frac{\xi(1+\xi)}{\nu} (x_i - u) \right)^{-1-1/\xi},$$

And if $\xi > -1/2$,

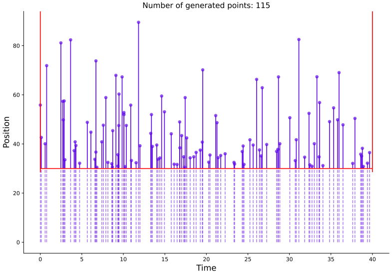
$$\mathcal{I}(r, \nu, \xi) = \mathrm{diag} \left(\frac{1}{r}, \frac{r}{\nu^2(1+2\xi)}, \frac{r}{(1+\xi)^2} \right).$$

SImulations

Simulations of datasets for the three cases for ξ:

SImulations

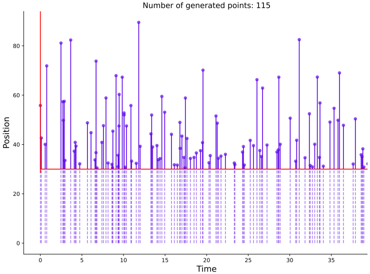
Simulations of datasets for the three cases for ξ:



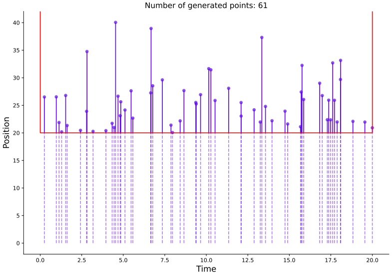
$\xi < 0$

SImulations

Simulations of datasets for the three cases for ξ:

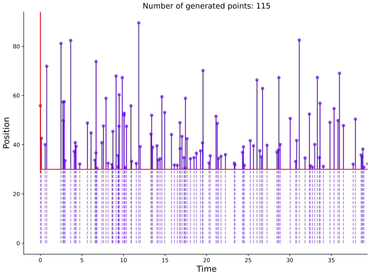


$\xi < 0$

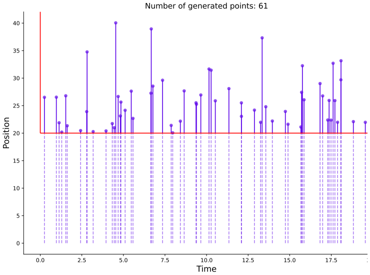


$\xi = 0$

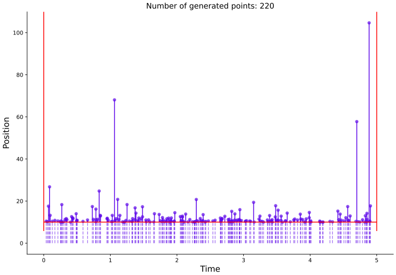
Simulations of datasets for the three cases for ξ:



$\xi < 0$



$\xi = 0$



$\xi > 0$

Convergence Diagnostics

How to discriminate the different parameterizations?

Convergence Diagnostics

How to discriminate the different parameterizations?

- **Autocorrelation:** measure the dependence between the elements of the chain (impact the quality of approximation)

Convergence Diagnostics

How to discriminate the different parameterizations?

- **Autocorrelation:** measure the dependence between the elements of the chain (impact the quality of approximation)
- **Effective Sample Size (ESS):** estimation of the equivalent number of independent draws.

Convergence Diagnostics

How to discriminate the different parameterizations?

- **Autocorrelation:** measure the dependence between the elements of the chain (impact the quality of approximation)
- **Effective Sample Size (ESS):** estimation of the equivalent number of independent draws.
- **Potential Scale Reduction Factor ($\widehat{R}$):** scalar diagnostic based on multiple chains analysis of variance.

Recommendations:

$$\hat{R} \in [1, 1.01] \implies \text{"Chains are mixing well"}.$$
$$ESS > 400 \implies \text{"Enough data for estimation"}.$$

Convergence Diagnostics

How to discriminate the different parameterizations?

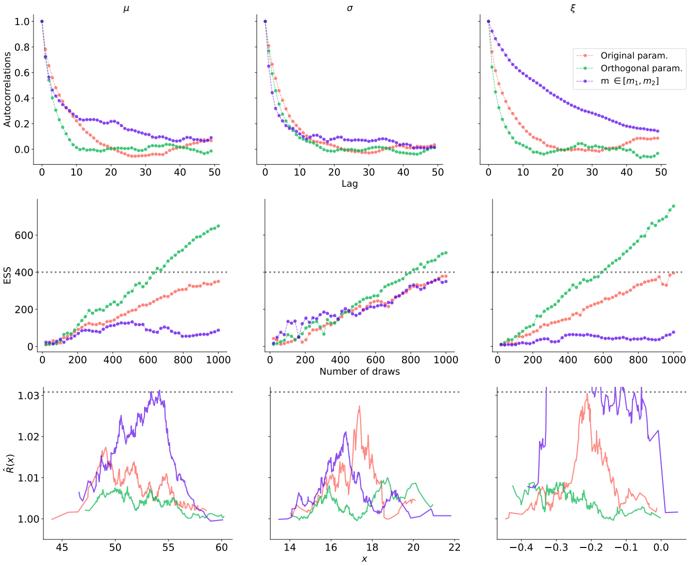
- **Autocorrelation:** measure the dependence between the elements of the chain (impact the quality of approximation)
- **Effective Sample Size (ESS):** estimation of the equivalent number of independent draws.
- **Potential Scale Reduction Factor ($\widehat{R}$):** scalar diagnostic based on multiple chains analysis of variance.

Recommendations:

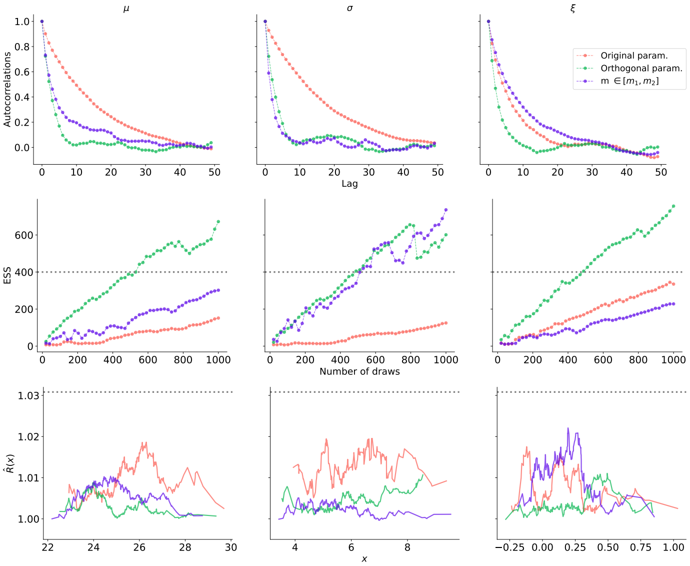
$$\hat{R} \in [1, 1.01] \implies \text{"Chains are mixing well"}.$$
$$ESS > 400 \implies \text{"Enough data for estimation"}.$$

We use here a refinement of $\hat{R}$ named $\hat{R}(x)$ (Moins et al., 2022), which aims at ensuring the convergence at a given quantile x of the distribution.

Results with $\xi < 0$



Results with $\xi = 0$



Results with $\xi > 0$

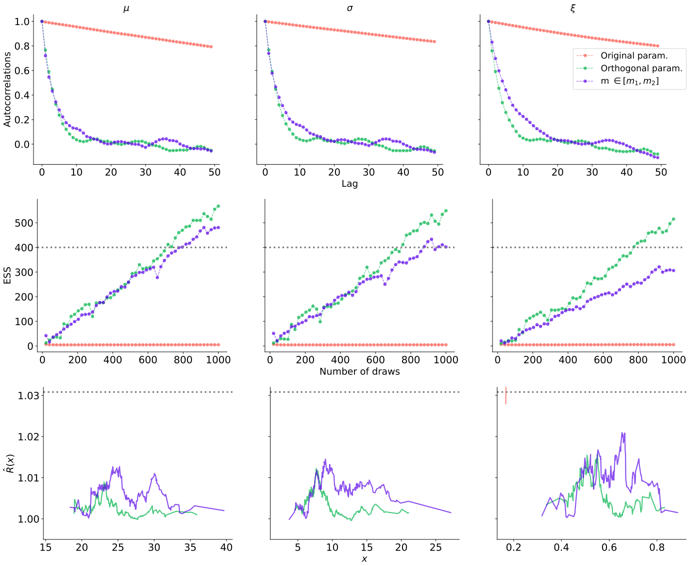


Table of Contents

Uninformative prior - Jeffreys

Jeffreys Prior : $p_J(\boldsymbol{\theta}) \propto \sqrt{\det I(\boldsymbol{\theta})}$.

$\hookrightarrow$ Invariant to reparametrisation: if $\phi = h(\boldsymbol{\theta})$, then $p(\phi) \propto \sqrt{\det I(\phi)}$.

Uninformative prior - Jeffreys

Jeffreys Prior : $p_J(\boldsymbol{\theta}) \propto \sqrt{\det I(\boldsymbol{\theta})}$.

$\hookrightarrow$ Invariant to reparametrisation: if $\phi = h(\boldsymbol{\theta})$, then $p(\phi) \propto \sqrt{\det I(\phi)}$.

Proposition 1 (Moins et al.)

Jeffreys prior associated with a Poisson process for extremes with parameters (r, ν, ξ) exists provided $\xi > -1/2$, and can be written as

$$p_{\mathrm{J}}(r, \nu, \xi) \propto \frac{r^{1/2}}{\nu(1+\xi)(1+2\xi)^{1/2}}.$$

Uninformative prior - Jeffreys

Jeffreys Prior : $p_J(\boldsymbol{\theta}) \propto \sqrt{\det I(\boldsymbol{\theta})}$.
$\hookrightarrow$ Invariant to reparametrisation: if $\phi = h(\boldsymbol{\theta})$, then $p(\phi) \propto \sqrt{\det I(\phi)}$.

Proposition 1 (Moins et al.)

Jeffreys prior associated with a Poisson process for extremes with parameters (r, ν, ξ) exists provided $\xi > -1/2$, and can be written as

$$p_{\mathrm{J}}(r, \nu, \xi) \propto \frac{r^{1/2}}{\nu(1+\xi)(1+2\xi)^{1/2}}.$$

Proposition 2 (Moins et al.)

Jeffreys prior for a Poisson process for extremes yields a proper posterior distribution, as soon as $\xi > -1/2$.

Informative prior on ξ - PC Prior

The shape parameter ξ plays a crucial role for parameter estimation.

Informative prior on ξ - PC Prior

The shape parameter ξ plays a crucial role for parameter estimation.
Even in a non-informative case:

Informative prior on ξ - PC Prior

The shape parameter ξ plays a crucial role for parameter estimation.
Even in a non-informative case:

- The case $\xi = 0$ concentrates an entire domain of attraction,

Informative prior on ξ - PC Prior

The shape parameter ξ plays a crucial role for parameter estimation.
Even in a non-informative case:

- The case $\xi = 0$ concentrates an entire domain of attraction,
- Almost all known distributions have $|\xi| < 1$.

Informative prior on ξ - PC Prior

The shape parameter ξ plays a crucial role for parameter estimation.
Even in a non-informative case:

- The case $\xi = 0$ concentrates an entire domain of attraction,
- Almost all known distributions have $|\xi| < 1$.

PC priors (Simpson et al., 2017): prior that penalizes the distance between a model $p_\xi := p(\cdot \mid \xi)$ with a given ξ and the baseline p_0 with $\xi = 0$:

$$p_{\mathrm{PC}}(\xi \mid \lambda) = \lambda \exp(-\lambda d(\xi)) \left| \frac{\partial d(\xi)}{\partial \xi} \right|,$$

with $\lambda > 0$ and $d(\xi) = \sqrt{2\mathrm{KL}\left(p_\xi || p_0\right)}$.

$\hookrightarrow$ Invariance to reparameterization for ξ.

Informative prior on ξ - PC Prior

The computation with GPD has already been done by Opitz et al. (2018) for the case $\xi \geq 0$.

Informative prior on ξ - PC Prior

The computation with GPD has already been done by Opitz et al. (2018) for the case $\xi \geq 0$.

Proposition 3 (Moins et al.)

PC prior associated with a Poisson process for extremes exists for any $\xi < 1$ and can be written as

$$p_{\mathrm{PC}}(\xi \mid \lambda) = \frac{\lambda}{2}\left(\frac{1-\xi/2}{(1-\xi)^{3/2}}\right)\exp\left(-\lambda\frac{|\xi|}{\sqrt{1-\xi}}\right). \tag{1}$$

Informative prior on ξ - PC Prior

The computation with GPD has already been done by Opitz et al. (2018) for the case $\xi \geq 0$.

Proposition 3 (Moins et al.)

PC prior associated with a Poisson process for extremes exists for any $\xi < 1$ and can be written as

$$p_{\mathrm{PC}}(\xi \mid \lambda) = \frac{\lambda}{2}\left(\frac{1-\xi/2}{(1-\xi)^{3/2}}\right)\exp\left(-\lambda\frac{|\xi|}{\sqrt{1-\xi}}\right). \quad (1)$$

Jeffreys' rule on (r,ν): $p_{\mathrm{J}}(r,\nu) \propto 1/\nu$

Proposition 4 (Moins et al.)

The prior defined as $p(r,\nu,\xi) \propto p_{\mathrm{PC}}(\xi)p_{\mathrm{J}}(r,\nu) \propto p_{\mathrm{PC}}(\xi)/\nu$ for the Poisson process for extremes yields a proper posterior distribution.

Informative prior on ξ - PC Prior

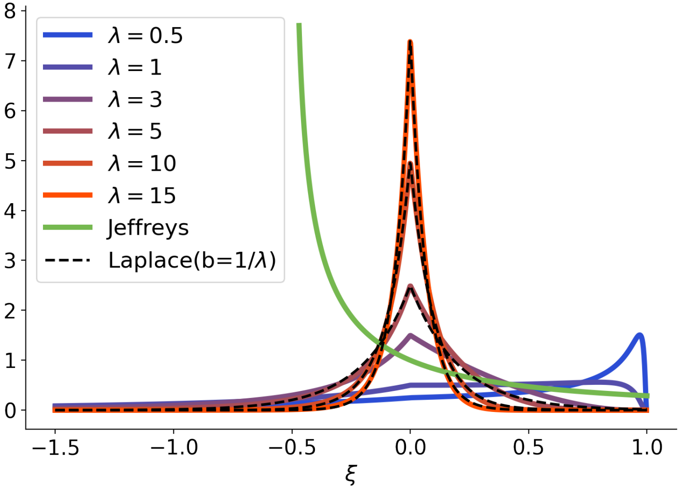


Table of Contents

River flow data - Preprocessing

Daily measurements of the Garonne river flow, from 1915 to 2013 $\Longrightarrow$ 36 160 observations.

Preprocessing steps:

River flow data - Preprocessing

Daily measurements of the Garonne river flow, from 1915 to 2013 $\Longrightarrow$ 36 160 observations.

Preprocessing steps:

- **Seasonality**: conservation of the rainy season from December to May.

River flow data - Preprocessing

Daily measurements of the Garonne river flow, from 1915 to 2013 $\Longrightarrow$ 36 160 observations.

Preprocessing steps:

- **Seasonality**: conservation of the rainy season from December to May.
- **Correlation**: clusters of exceedances of parameters $r = 3$ days.

River flow data - Preprocessing

Daily measurements of the Garonne river flow, from 1915 to 2013 $\Longrightarrow$ 36 160 observations.

Preprocessing steps:

- **Seasonality**: conservation of the rainy season from December to May.
- **Correlation**: clusters of exceedances of parameters $r = 3$ days.
- **Choice of theshold**: elicitation methods leads to $u \approx 2\,000$ m^3/s.

River flow data - Preprocessing

Daily measurements of the Garonne river flow, from 1915 to 2013 $\Longrightarrow$ 36 160 observations.

Preprocessing steps:

- **Seasonality**: conservation of the rainy season from December to May.
- **Correlation**: clusters of exceedances of parameters $r = 3$ days.
- **Choice of theshold**: elicitation methods leads to $u \approx 2\,000$ m^3/s.

$\hookrightarrow$ In the end: $n = 182$ exceedances

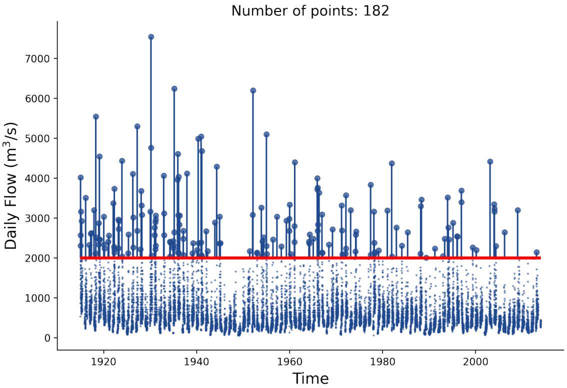
River flow data - Preprocessing

Daily measurements of the Garonne river flow, from 1915 to 2013 $\Longrightarrow$ 36 160 observations.

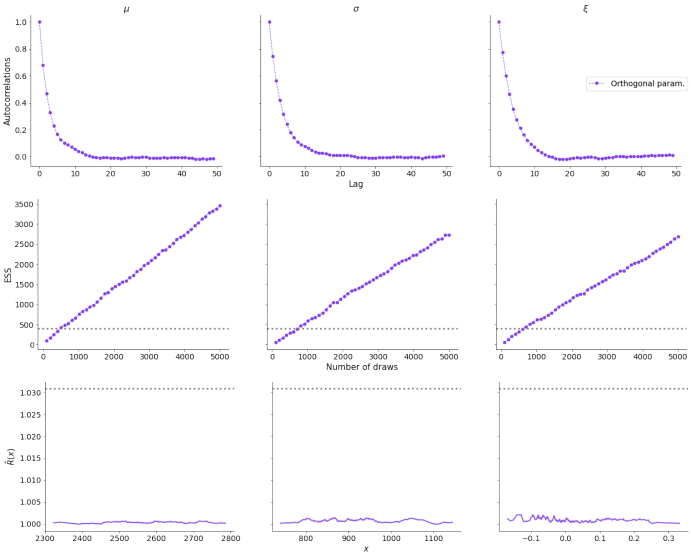
Preprocessing steps:

- **Seasonality**: conservation of the rainy season from December to May.
- **Correlation**: clusters of exceedances of parameters $r = 3$ days.
- **Choice of theshold**: elicitation methods leads to $u \approx 2\,000$ m^3/s.

$\hookrightarrow$ In the end: $n = 182$ exceedances



River flow data - Convergence diagnostics



River flow data - Estimation

	Post. Mean	Post. SD	95%-CI	ESS	$\hat{R}_\infty$
μ	2 560.8	84.1	[2 409.8, 2 724.1]	3 473	≈ 1.0
σ	919.6	73.2	[787.2, 1 063.3]	2 709	≈ 1.0
ξ	0.015	0.077	[−0.120, 0.164]	2 702	≈ 1.0

Posterior summaries (mean, standard deviation (SD), credible interval (CI) at 95%) and convergence diagnostics (ESS and $\hat{R}_\infty$) for (μ, σ, ξ) associated with annual maxima ($m = 99$).

River flow data - Estimation

	Post. Mean	Post. SD	95%-CI	ESS	$\hat{R}_\infty$
μ	2 560.8	84.1	[2 409.8, 2 724.1]	3 473	≈ 1.0
σ	919.6	73.2	[787.2, 1 063.3]	2 709	≈ 1.0
ξ	0.015	0.077	[−0.120, 0.164]	2 702	≈ 1.0

Posterior summaries (mean, standard deviation (SD), credible interval (CI) at 95%) and convergence diagnostics (ESS and $\hat{R}_\infty$) for (μ, σ, ξ) associated with annual maxima ($m = 99$).

$$\xi = 0 \text{ ?}$$

River flow data - Estimation

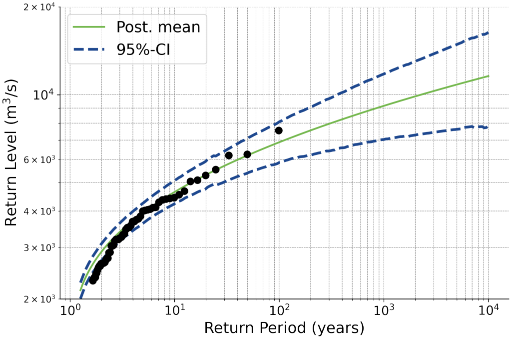
$$\ell_T = \mu - \frac{\sigma}{\xi}\left(1 - (-\log(1 - 1/T))^{-\xi}\right)$$

Mean and 2.5%/97.5% quantiles on the posterior distribution of ℓ_T:

River flow data - Estimation

$$\ell_T = \mu - \frac{\sigma}{\xi}\left(1 - (-\log(1 - 1/T))^{-\xi}\right)$$

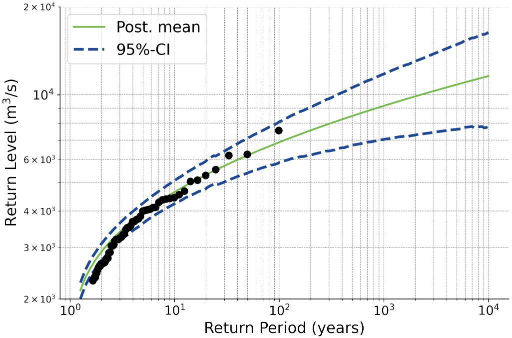
Mean and 2.5%/97.5% quantiles on the posterior distribution of ℓ_T:



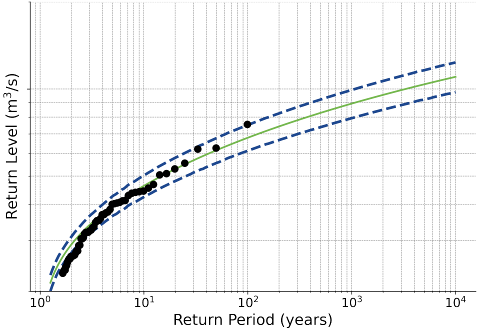
River flow data - Estimation

$$\ell_T = \mu - \frac{\sigma}{\xi}\left(1 - (-\log(1 - 1/T))^{-\xi}\right)$$

Mean and 2.5%/97.5% quantiles on the posterior distribution of ℓ_T:



Estimation of ξ



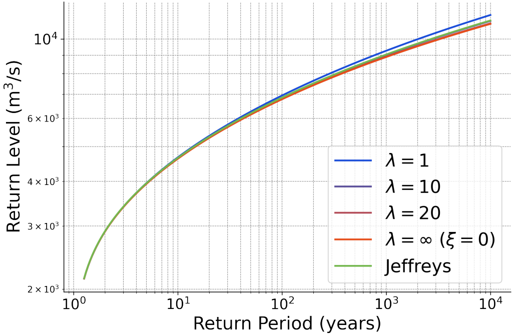
$\xi = 0$

River flow data - Prior influence

PC prior allows to navigate between the two cases, thanks to hyperparameter λ:

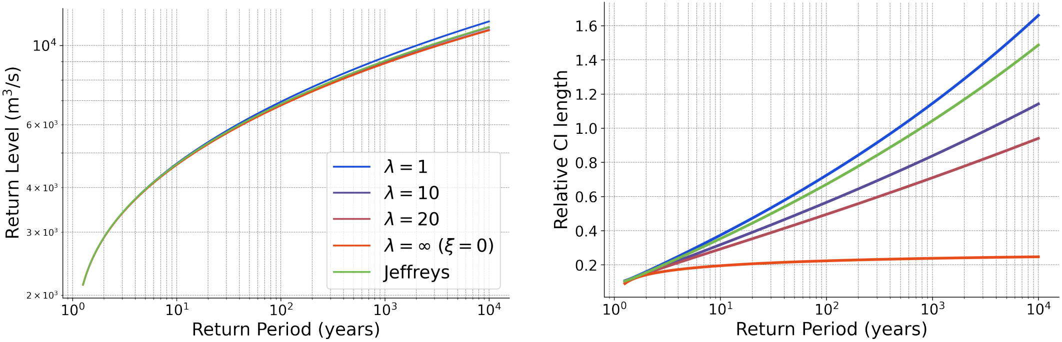
River flow data - Prior influence

PC prior allows to navigate between the two cases, thanks to hyperparameter λ:



River flow data - Prior influence

PC prior allows to navigate between the two cases, thanks to hyperparameter λ:



Comparison of return levels with different priors as functions of return period (log scale). On the left: return levels with posterior mean parameters. On the right: return level credible interval (CI) length relative to the point estimate (in %).

Conclusion

Conclusion

- Orthogonal parameterization facilitate the convergence of MCMC algorithms...

Conclusion

- Orthogonal parameterization facilitate the convergence of MCMC algorithms. . .
- . . . and facilitates the computation of Jeffreys prior (and posterior propriety).

Conclusion

- Orthogonal parameterization facilitate the convergence of MCMC algorithms...
- ...and facilitates the computation of Jeffreys prior (and posterior propriety).
- Posterior uncertainty can be significantly reduced by adding prior information of ξ with PC prior.

Conclusion

- Orthogonal parameterization facilitate the convergence of MCMC algorithms…
- …and facilitates the computation of Jeffreys prior (and posterior propriety).
- Posterior uncertainty can be significantly reduced by adding prior information of ξ with PC prior.
- **Future work**: Study in more details the posterior uncertainty of return levels.

Conclusion

- Orthogonal parameterization facilitate the convergence of MCMC algorithms...
- ...and facilitates the computation of Jeffreys prior (and posterior propriety).
- Posterior uncertainty can be significantly reduced by adding prior information of ξ with PC prior.
- **Future work**: Study in more details the posterior uncertainty of return levels.

Python implementation using PyMC3 (Salvatier et al., 2016):
https://github.com/TheoMoins/ExtremesPyMC

Preprint! 🎉

Reparameterization of extreme value framework for improved Bayesian workflow

Théo Moins * Julyan Arbel * Stéphane Girard * Anne Dutfoy †

October 12, 2022

Abstract

Combining extreme value theory with Bayesian methods offers several advantages, such as a quantification of uncertainty on parameter estimation or the ability to study irregular models that cannot be handled by frequentist statistics. However, it comes with many options that are left to the user concerning model building, computational algorithms, and even inference itself. Among them, the parameterization of the model induces a geometry that can alter the efficiency of computational algorithms, in addition to making calculations involved. We focus on the Poisson process characterization of extremes and outline two key benefits of an orthogonal parameterization addressing both issues. First, several diagnostics show that Markov chain Monte Carlo convergence is improved compared with the original parameterization. Second, orthogonalization also helps deriving Jeffreys and penalized complexity priors, and establishing posterior propriety. The analysis is supported by simulations, and our framework is then applied to extreme level estimation on river flow data.

T. Moins, J. Arbel, A. Dutfoy & S. Girard. (2022+) "*Reparameterization of extreme value framework for improved Bayesian workflow*" https://arxiv.org/abs/2210.05224

References I

Betancourt, M. (2019). Incomplete Reparameterizations and Equivalent Metrics. arXiv:1910.09407.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C*, 54(1):207–222.

Coles, S. (2001). A point process characterization of extremes. In *An Introduction to Statistical Modeling of Extreme Values*, pages 124–141. Springer.

Coles, S. G. and Powell, E. A. (1996). Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review / Revue Internationale de Statistique*, 64(1):119–136.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. CRC press.

Jeffreys, H. (1961). Theory of probability.

References II

Moins, T., Arbel, J., Dutfoy, A., and Girard, S. (2022). On the use of a local $\hat{R}$ to improve MCMC convergence diagnostic. arXiv:2205.06694.

Opitz, T., Huser, R., Bakka, H., and Rue, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, 21(3):441–462.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28.