Un Tétris pour générer des précipitations

Rainfall modelling with compound discrete EGPD

Pierre Ailliot[1], Juliette LEGRAND[1], and Carlo GAETAN[2]

[1]Laboratoire de Mathématiques de Bretagne Atlantique, UBO Brest
[2]DAIS, Ca' Foscari University - Venice

25/03/2026

Statistiques au sommet de Rochebrune

Université de Bretagne Occidentale

LMBA

Motivation

Motivation

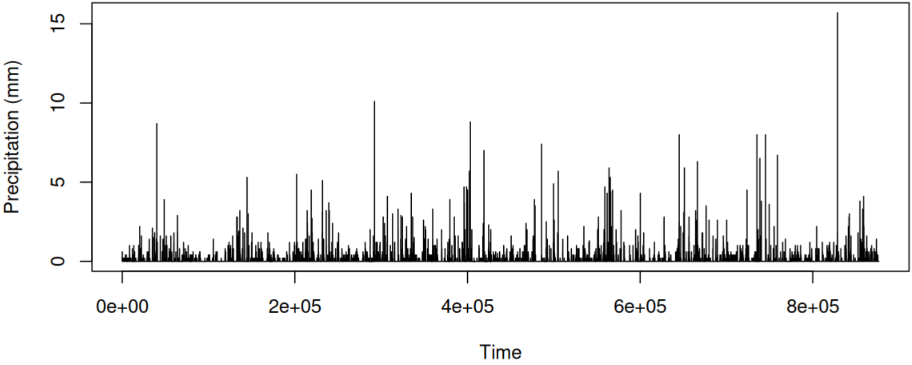
"Fin janvier 2025, la tempête Eowyn, suivie des dépressions Herminia puis Ivo ont apporté des pluies continues. Ces pluies sont tombées sur des sols déjà saturés en eau par un antécédent pluvieux sur l'ensemble du mois, limitant l'infiltration dans le sol, provoquant des débordements de cours d'eau inédits et ralentissant la décrue."

https://www.ille-et-vilaine.gouv.fr

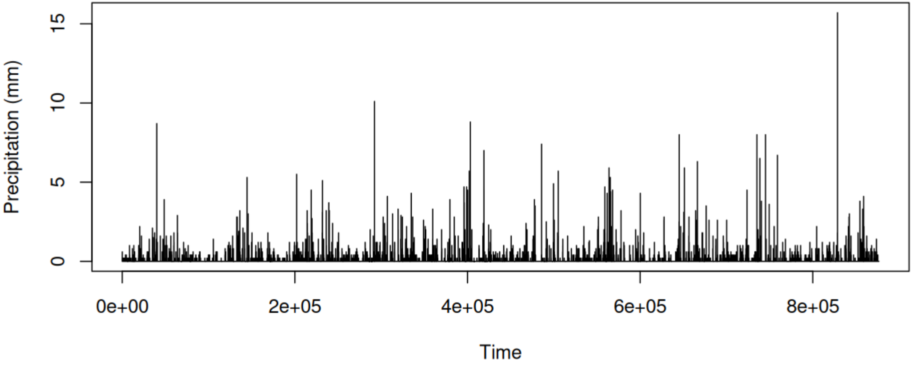
Motivation

https://www.aspas-nature.org

Goals of this project



Goals of this project



- Model precipitation time series by taking into account time dependence and non-stationarities
- Characterize the whole distribution of rainfall, including non-extreme values
- Develop a stochastic precipitation generator for high-resolution time series (6 minutes)

Compound Poisson distribution

Rainfall events:

- $N \sim \mathcal{P}oisson(\lambda)$: number of time it rains in a given period

Compound Poisson distribution

Rainfall events:

- $N \sim \mathcal{P}oisson(\lambda)$: number of time it rains in a given period
- $(E_i)_{i\geq 1}$ *i.i.d.*, nonnegative and indepedent of N: amount of rain falling

Compound Poisson distribution

Rainfall events:

- $N \sim \mathcal{P}oisson(\lambda)$: number of time it rains in a given period
- $(E_i)_{i \geq 1}$ *i.i.d.*, nonnegative and indepedent of N: amount of rain falling

Then, total amount of precipitation in a time period is given by

$$Z := \sum_{i=1}^{N} E_i$$

Compound Poisson distribution

Rainfall events:

- $N \sim \mathcal{P}oisson(\lambda)$: number of time it rains in a given period
- $(E_i)_{i \geq 1}$ *i.i.d.*, nonnegative and indepedent of N: amount of rain falling

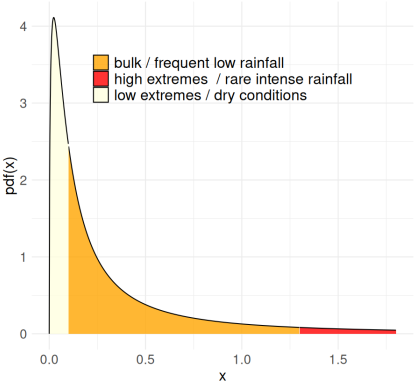
Then, total amount of precipitation in a time period is given by

$$Z := \sum_{i=1}^{N} E_i$$

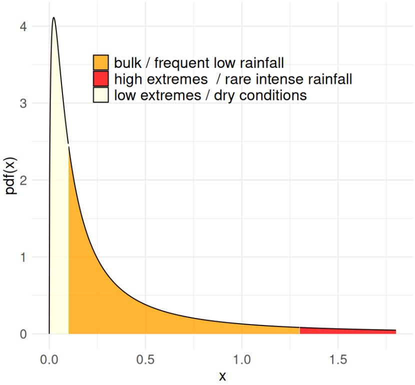
Ex: if $E_i \sim \mathcal{G}amma$, Z is a Poisson-Gamma r.v. (Fisher et al., 1960)

In our case, $E_i \sim$ EGPD (Naveau et al., 2016)

Model for the intensities (or marks)



Model for the intensities (or marks)



Low extremes
κ the GPD parameter of $1/E$

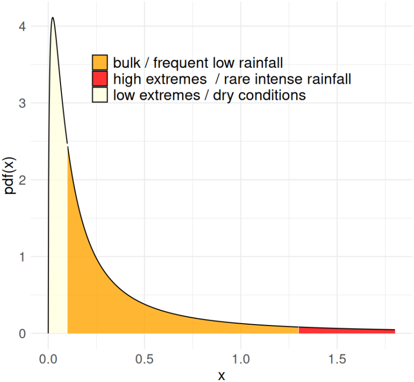
Bulk
B a CDF function on $[0,1]$

High extremes
ξ the GPD parameter of E

$$E = \sigma H_{\xi}^{-1}\left(\left(B^{-1}(U)\right)^{1/\kappa}\right) \sim EGPD(\sigma, \kappa, \xi, B)$$

with $U \sim \text{Uniform}[0,1]$, and typically $B(x) = x$

Model for the intensities (or marks)



Low extremes
κ the GPD parameter of $1/E$

Bulk
B a CDF function on $[0, 1]$

High extremes
ξ the GPD parameter of E

$$E = \sigma H_\xi^{-1}\left(\left(B^{-1}(U)\right)^{1/\kappa}\right) \sim EGPD(\sigma, \kappa, \xi, B)$$

with $U \sim \text{Uniform}[0, 1]$, and typically $B(x) = x$

Sums of *iid* EGPD variables remain EGPD, with same ξ but different B

Adding some temporal dependence: **Trawl processes**[1]

[1] Barndorff-Nielsen, 2011 ; Doukhan et al., 2019

Adding some temporal dependence: **Trawl processes**[1]

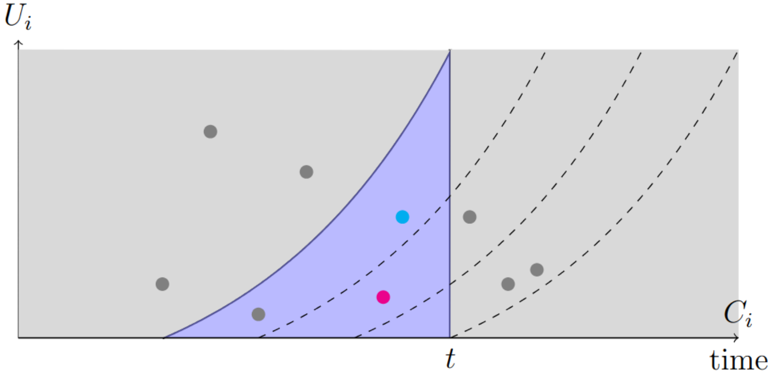
[1] Barndorff-Nielsen, 2011 ; Doukhan et al., 2019

Adding some temporal dependence: **Trawl processes**[1]

Trawl $\neq$ Troll !!

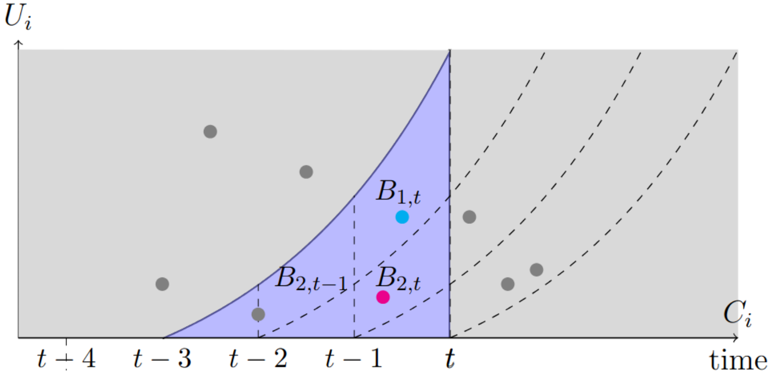
[1] Barndorff-Nielsen, 2011 ; Doukhan et al., 2019

Adding some temporal dependence: **Trawl processes**[1]



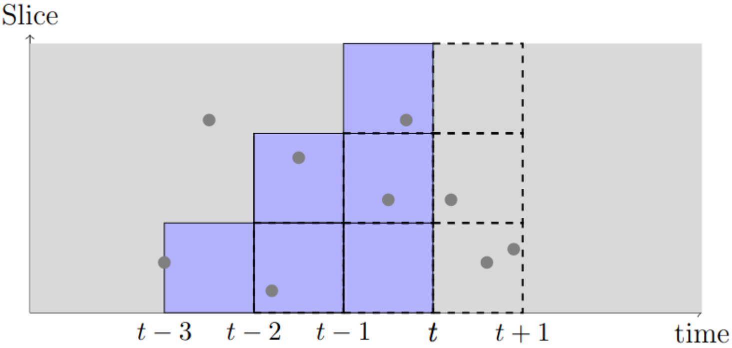
[1] Barndorff-Nielsen, 2011 ; Doukhan et al., 2019

Adding some temporal dependence: **Trawl processes**[1]



[1] Barndorff-Nielsen, 2011 ; Doukhan et al., 2019

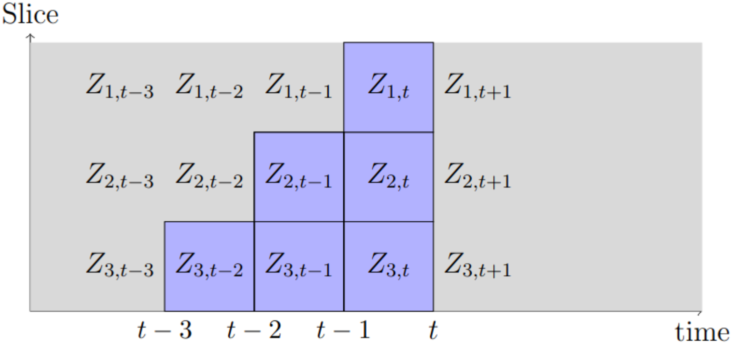
Trawl discretization: **Tetris process**



Trawl discretization: **Tetris process**

In each tetris box:

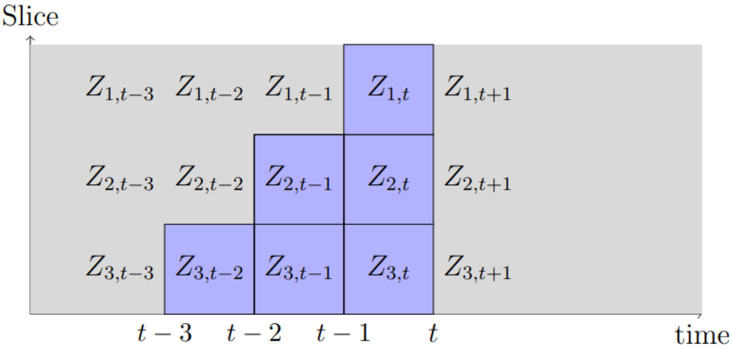
$$Z_{j,t} = \sum_{i=1}^{N_{j,t}} E_i^{j,t},\ j = 1, \ldots, q,\ \ t \in \mathbb{Z}$$



Trawl discretization: **Tetris process**

In each tetris box:

$$Z_{j,t} = \sum_{i=1}^{N_{j,t}} E_i^{j,t}, \; j = 1, \dots, q, \; t \in \mathbb{Z}$$

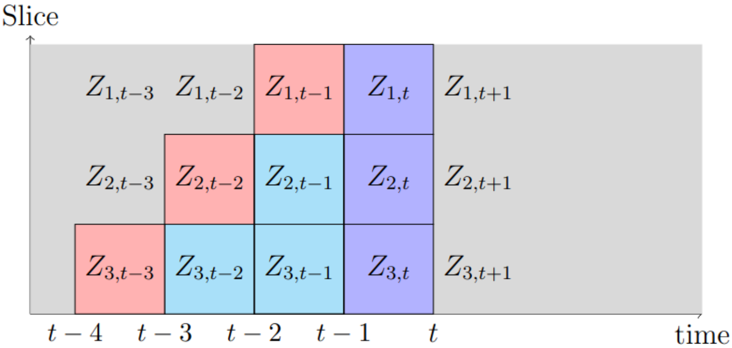


$\Rightarrow$ **q-Tetris process** defined as $Y_t := \sum_{(j,s) \in I_t} Z_{j,s}$

Trawl discretization: **Tetris process**

In each tetris box:

$$Z_{j,t} = \sum_{i=1}^{N_{j,t}} E_i^{j,t}, \; j = 1, \ldots, q, \; t \in \mathbb{Z}$$

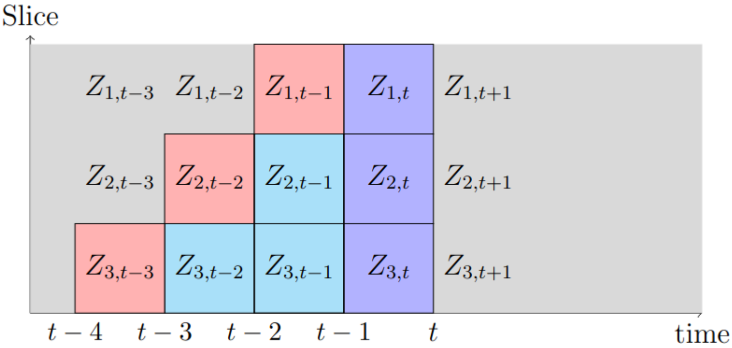


$\Rightarrow$ **q-Tetris process** defined as $Y_t := \sum_{(j,s)\in I_t} Z_{j,s}$

Trawl discretization: **Tetris process**

In each tetris box:

$$Z_{j,t} = \sum_{i=1}^{N_{j,t}} E_i^{j,t}, \; j = 1, \ldots, q, \;\; t \in \mathbb{Z}$$



$\Rightarrow$ **q-Tetris process** defined as $Y_t := \sum_{(j,s)\in I_t} Z_{j,s}$

tetris = tetra + tennis

Trawl discretization: **Tetris process**

In each tetris box:

$N_{j,t}$

$Z_{j,t}$ $\in \mathbb{Z}$

$\Rightarrow$ **q-Tetris process** defined as $Y_t := \sum_{(j,s)\in I_t} Z_{j,s}$

tetris = tetra + tennis

Trawl discretization: **Tetris process**

In each tetris box:

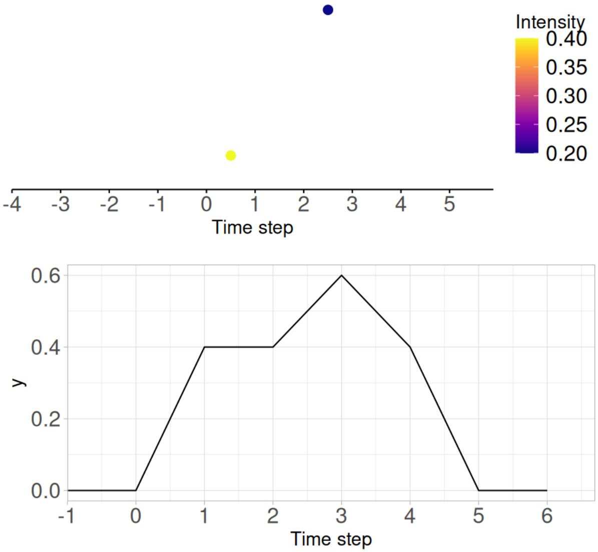
$N_{j,t}$

$Z_{j,t}$  $\in \mathbb{Z}$

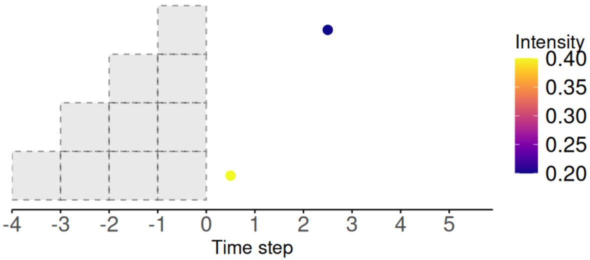
$\Rightarrow$ **q-Tetris process** defined as $Y_t :=$  $+ \cdots +$ 

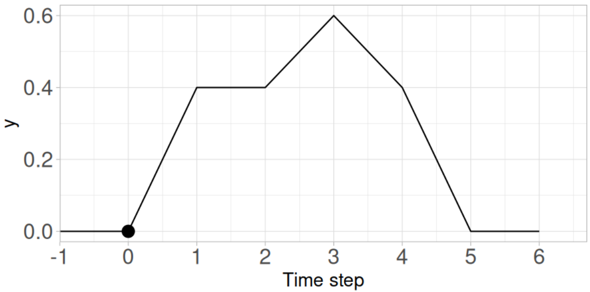
tetris = tetra + tennis

Simulation example of the Tetris process with $q = 4$

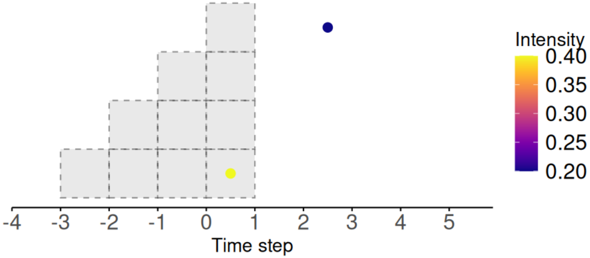


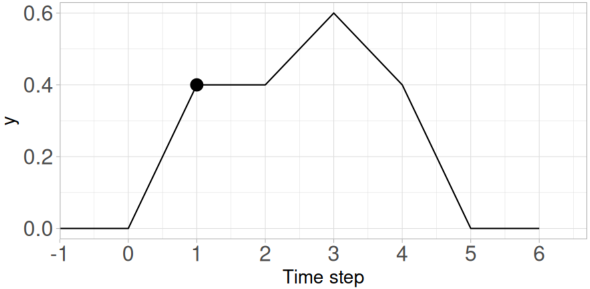
Simulation example of the Tetris process with $q = 4$



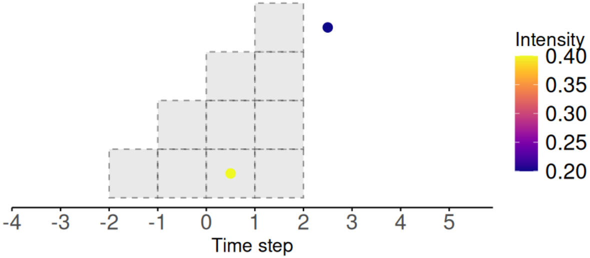


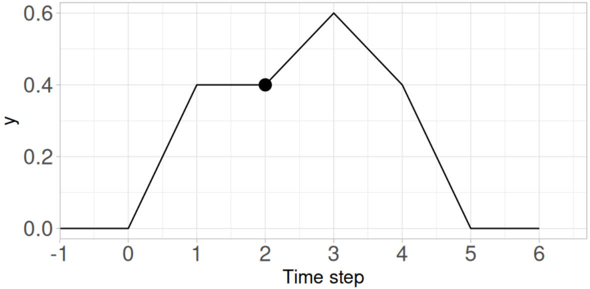
Simulation example of the Tetris process with $q = 4$



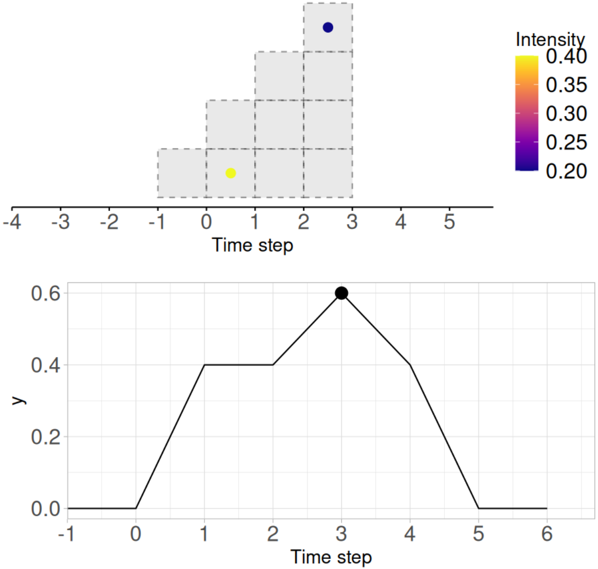


Simulation example of the Tetris process with $q = 4$

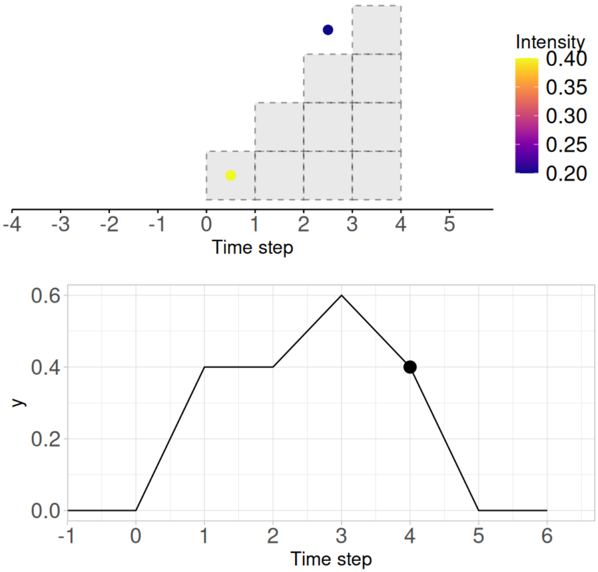




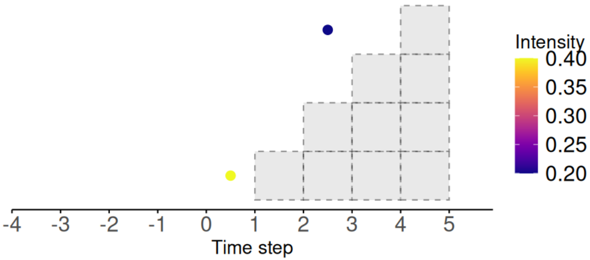
Simulation example of the Tetris process with $q = 4$

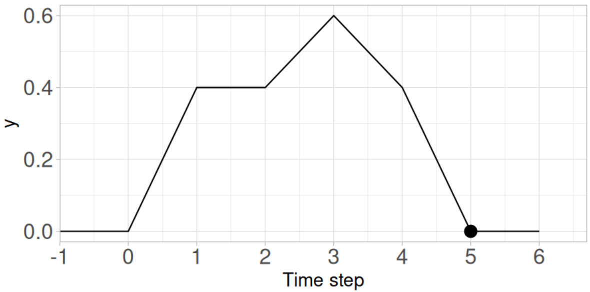


Simulation example of the Tetris process with $q = 4$



Simulation example of the Tetris process with $q = 4$

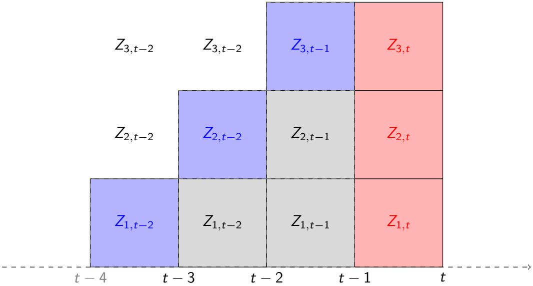




Inference scheme

$$Y_t = \sum_{i \in I_t \setminus I_s} Z_i + \sum_{i \in I_t \cap I_s} Z_i = Y_{(t \setminus s)} + Y_{(t,s)}$$

$$Y_s = \sum_{i \in I_s \setminus I_t} Z_i + \sum_{i \in I_t \cap I_s} Z_i = Y_{(s \setminus t)} + Y_{(t,s)}$$



Inference scheme

$$Y_t = \sum_{i \in I_t \setminus I_s} Z_i + \sum_{i \in I_t \cap I_s} Z_i = Y_{(t \setminus s)} + Y_{(t,s)}$$

$$Y_s = \sum_{i \in I_s \setminus I_t} Z_i + \sum_{i \in I_t \cap I_s} Z_i = Y_{(s \setminus t)} + Y_{(t,s)}$$

Bivariate probability function given by

$$\mathbb{P}(Y_t = y_t, Y_s = y_s) = \sum_{k=0}^{\min(y_t, y_s)} \mathbb{P}(Y_{t \setminus s} = y_t - k)\mathbb{P}(Y_{t,s} = k)\mathbb{P}(Y_{s \setminus t} = y_s - k)$$

[2] Panjer, 1981; Embrechts and Frei, 2009

Inference scheme

$$Y_t = \sum_{i \in I_t \setminus I_s} Z_i + \sum_{i \in I_t \cap I_s} Z_i = Y_{(t \setminus s)} + Y_{(t,s)}$$

$$Y_s = \sum_{i \in I_s \setminus I_t} Z_i + \sum_{i \in I_t \cap I_s} Z_i = Y_{(s \setminus t)} + Y_{(t,s)}$$

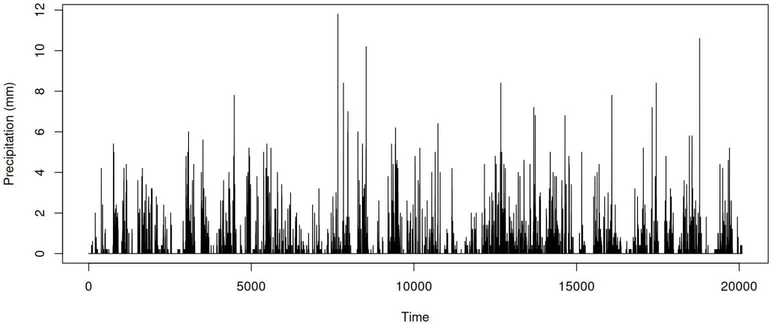
Bivariate probability function given by

$$\mathbb{P}(Y_t = y_t, Y_s = y_s) = \sum_{k=0}^{\min(y_t, y_s)} \mathbb{P}(Y_{t \setminus s} = y_t - k)\mathbb{P}(Y_{t,s} = k)\mathbb{P}(Y_{s \setminus t} = y_s - k)$$

- Inference based on pairwise likelihood and Panjer's algorithm [2]

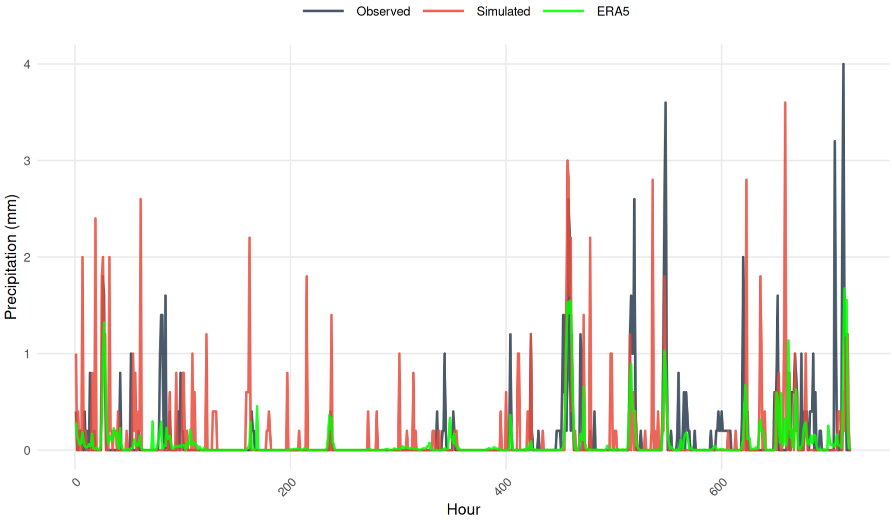
[2] Panjer, 1981; Embrechts and Frei, 2009

Simulation of "realistic" rainfall events

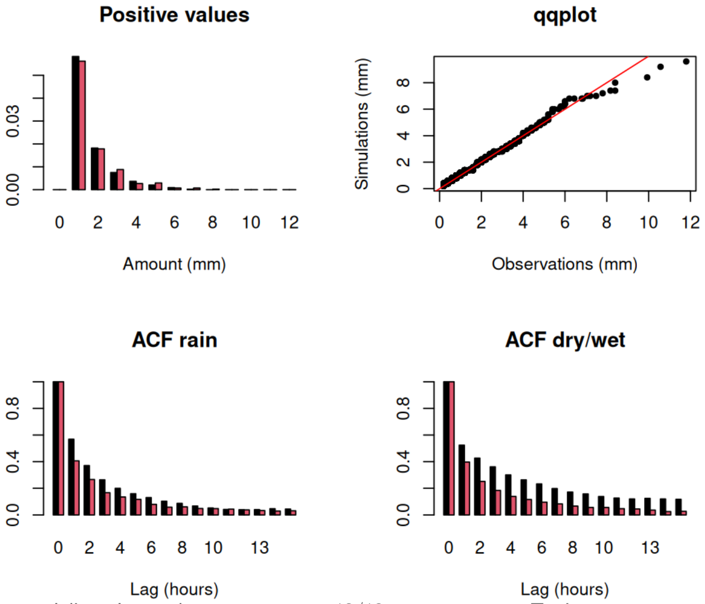


- Precipitation data from Brest-Guipavas station, France
- 1 hour time resolution
- Covariate: ERA5 precipitation at the nearest grid point
- Effect on the λ_t (Poisson parameters) and on the σ (EGPD scale parameter, by tetris-slice)

Simulation of "realistic" rainfall events



Simulation of "realistic" rainfall events ($q = 4$)



Conclusion and perspectives

- Good reproduction of marginal properties and temporal dependence structure
- High flexibility:
  - Covariate effects that can be put on the parameters of the Poisson (occurence) and of the EGPD (marks)
  - Choice for the distribution of the marks

Conclusion and perspectives

- Good reproduction of marginal properties and temporal dependence structure
- High flexibility:
  - Covariate effects that can be put on the parameters of the Poisson (occurence) and of the EGPD (marks)
  - Choice for the distribution of the marks
- A question remains: selection of the Tetris size (*i.e.* the lag)

Conclusion and perspectives

- Good reproduction of marginal properties and temporal dependence structure
- High flexibility:
  - Covariate effects that can be put on the parameters of the Poisson (occurence) and of the EGPD (marks)
  - Choice for the distribution of the marks
- A question remains: selection of the Tetris size (*i.e.* the lag)

$\rightsquigarrow$ Extension to spatial setting (upcoming PhD)