Modeling of daily precipitation data, with heavy rainfall and long periods of drought

Antoine Doizé

Advisors : D. Allard, P. Naveau, O. Wintenberger

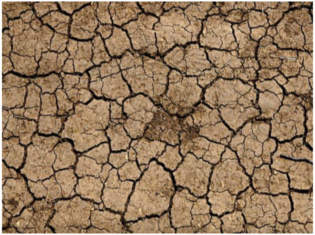
Sorbonne-Université - Financé par la chaire geolearning

01/04/25

Drought and floods impacts

Direct impact: floods and agriculture

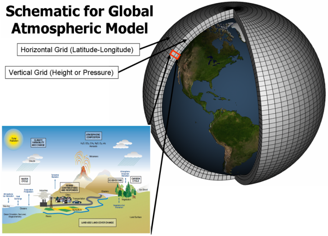
Shrink-swell of clays



Forest wildfires, nuclear plant cooling system

Stochastic Weather Generator approach

Climate model



Driven by physics: solving physics equations

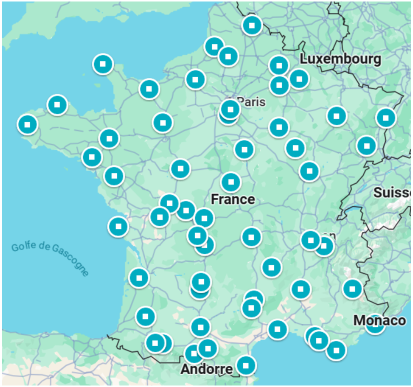
- \+ Physics consistency of weather variables
- \- Simulation computationally expensive
- \- Producing several scenarios is costly.

Stochastic Weather Generator

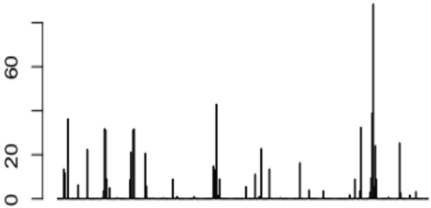
Driven by data: finding pattern in data and emulate it

- \- Local predictions. Challenges for numerous weather variables/
- \+ Simulation computationally cheaper
- \+ Produce numerous long scenarios

Precipitation data



ECAD (Europe Climate Assessment and Dataset) precipitation data in France



- Daily aggregated data : 50-80 years
- 50 stations over France
- Modelling site by site (spatialization in future work)

Critical points of our Stochastic Rain Generator

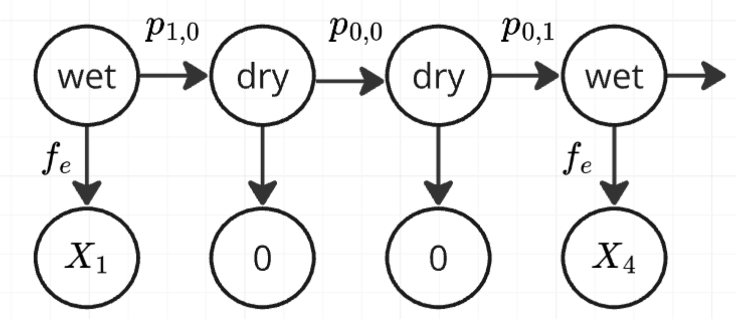
1. Produce series of "0" and rain intensities
2. Control over rain intensity extreme values
3. Control over long periods of dry days or rain days

Quick comparison of Stochastic Weather Generators

Method	Simulation	Spatialization	Rainfall Extremes	Dry Spells Extremes
Resampling	+	+	-	-
Semi-Markov	+	-	+	+
Truncated GP	+	+	+	-
Markov models	+	+	?	? / -

Focus: Markov model for rainfall modeling[1]

- First: Rain occurrence
- Then: Derive rain intensities



[1] Richardson, C. "Stochastic simulation of daily precipitation, temperature, and solar radiation", Water Resources Research, 1981

Reminder

Geometric sejourn times

In a Markov chain with discrete state space, the time spent in a given state follows a geometric law.

Set $\mathcal{S} := \{1, 0\}$ and define

$$p_{0,0} := \mathbb{P}(X_{n+1} = 0 | X_n = 0).$$

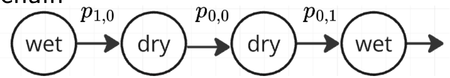
Then,

$$\mathbb{P}_{X_0=0}(X_1 = 0, \ldots, X_{n-1} = 0, X_n = 1)) = (p_{0,0})^{n-1}(1 - p_{0,0}), \quad n \geq 1$$

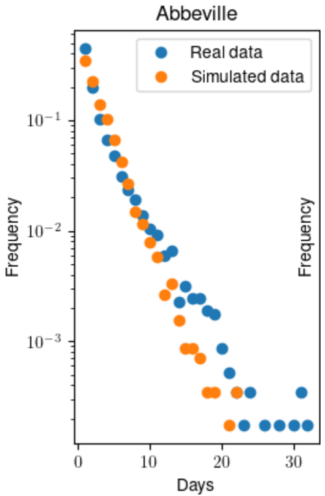
Thus, a 2-states Markov chain will leads to "geometric" dry spell duration.

Distribution of dry spell duration (1/2)

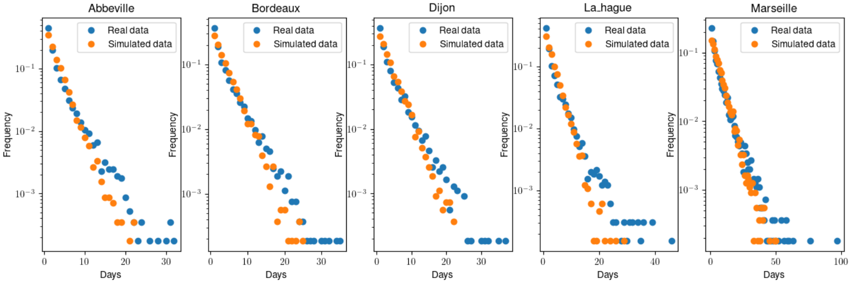
1. Fit first-order two-step Markov chain



2. Compare real dry days durations vs simulated dry days duration
3. Frequency in log scale : straight line means geometric distribution



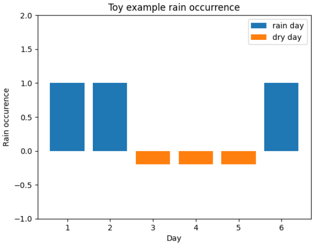
Distribution of dry spell duration (2/2)



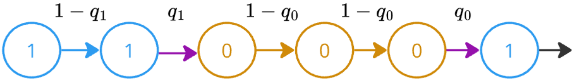
High quantiles are underestimated.

Modeling rainfall occurrence: intuition

Rainfall occurrence toy data for 6 days:



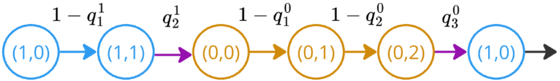
Let A be the toy data realization.



Simple Markov model $(X_n)_{n=0\dots}$

Simple Markov model:

$$\mathbb{P}(A) = \mathbb{P}_{(q_0,q_1)}(A)$$



Enlarged Markov model $(X_n, T_n)_{n=0\dots}$

Enlarged Markov model:

$$\mathbb{P}(A) = \mathbb{P}_{(q_i^0,q_i^1),i=0\dots}(A)$$

Waiting Time Representation of a discrete distribution

Proposition 2.1 (from (Kozubowski)[2])

Let N be a discrete random variable on $\mathbb{N}$. Let $\{B_n, n \in \mathbb{N}\}$ be a sequence of independent Bernoulli trials with success probabilities:

$$q_n = \mathbb{P}(B_n = 1) = \frac{\mathbb{P}(N = n)}{\mathbb{P}(N \geq n)} = \mathbb{P}(N = n \mid N \geq n),$$

whenever $\mathbb{P}(N \geq n) > 0$, and $q_n = 1$ when $\mathbb{P}(N \geq n) = 0$. Then:

$$N \stackrel{d}{=} \inf\{n \in \mathbb{N} : B_n = 1\}.$$

[2] Tomasz J. Kozubowski, Dorota Młynarczyk, Anna K. Panorska "Waiting time representation of discrete distributions", Statistics and Probability Letters, 2025

Consequence of waiting time representation

Using enlarged state space (X_n, T_n), let us have i.i.d. sojourn times having any chosen discrete distribution.

This let us design a flexible rainfall occurrence Markov model.

Examples of sojourn times choices

1. $q_n := q \in (0,1)$ leads to geometric sojourn times
2. $q_n := 1 - \exp(-\lambda(n+1)^\beta - n^\beta)$ leads to discrete Weibull sojourn time
3. $q_n := 1 - (\frac{1+\sigma\alpha n}{1+\sigma\alpha(n+1)})^{1/\alpha}$ leads to discrete Pareto sojourn time
4. $q_n := \frac{G(H(\frac{n+1}{\sigma})) - G(H(\frac{n}{\sigma}))}{1 - G(H(\frac{n-1}{\sigma}))}$ leads to ext-GPD distribution (details on G and H next slides)

Flexible rainfall occurrence Markov model

Model definition

Let us have $(q_n^0)_{n\in\mathbb{N}}$, $(q_n^1)_{n\in\mathbb{N}}$,sequences in $(0,1)$. $(U_n)_{n\in\mathbb{N}}$ i.i.d. uniform random variables.
For given initial values $r_0 \in \{0,1\}$ and $t_0 \in \mathbb{N}$, set $(R_0, T_0) = (r_0, t_0)$, and define recursively, for all $n \in \mathbb{N}$:

$$(R_{n+1}, T_{n+1}) =$$

$$\begin{cases} \begin{cases} (1, 1+T_n), & \text{if } U_{n+1} > q^1_{1+T_n}, \\ (0,0), & \text{if } U_{n+1} \leq q^1_{1+T_n}, \end{cases} & \text{if } R_n > 0, \\ \begin{cases} (0, 1+T_n), & \text{if } U_{n+1} > q^0_{1+T_n}, \\ (1,0), & \text{if } U_{n+1} \leq q^0_{1+T_n}, \end{cases} & \text{if } R_n = 0. \end{cases}$$

- Stay in rain period
- Stay in dry spell
- Switch from rain to dry spell or from dry spell to rain

Reminder: extended Generalized Pareto distribution

Generalized Pareto distribution:

$$H_\xi(z) = \begin{cases} 1 - (1+\xi z)_+^{-1/\xi} & \text{for } \xi \neq 0 \\ 1 - \exp(-z) & \text{for } \xi = 0 \end{cases}$$

- $\xi > 0$ means heavy tailed distribution
- $\xi = 0$ means exponential distribution
- $\xi < 0$ means bounded distribution

X has GPD distribution of parameters ξ, σ if:

$$X = \sigma H_\xi^{-1}(U)$$

Let G be a continuous c.d.f. Y has an extended GPD distribution[3] if:

$$Y = \sigma H_\xi^{-1}(G^{-1}(U))$$

In particular it follows ext-GPD of type 1 if $G(u) = G_\kappa(u) = u^\kappa$

[3] Naveau, P., R. Huser, P. Ribereau, and A. Hannart (2016), Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, Water Resour. Res., 52, 2753-2769

Model with independent intensities (1/2)

Let us have:

To control rain occurrence

- $(q_n^0)_{n\in\mathbb{N}}$, $(q_n^1)_{n\in\mathbb{N}}$ so that waiting-times follow ext-GPD1 distributions.
- $(U_n)_{n\in\mathbb{N}}$ i.i.d. standard uniform random variables.

Random variables controlling rain intensity

- $(I_n)_{n\in\mathbb{N}}$ i.i.d. positive random variable with ext-GPD1 distribution.

Model with independent intensities (2/2)

Rainfall model with independent intensities

Let $x_0 \in \mathbb{R}^+$, and $t_0 \in \mathbb{N}^*$.

$$(X_0, T_0) = (x_0, t_0),$$

and recursively $\forall n \in \mathbb{N}$:

$$(X_{n+1}, T_{n+1}) = \begin{cases} \begin{cases} (I_{n+1}, 1 + T_n), & \text{if } U_{n+1} > q^1_{1+T_n}, \\ (0, 0), & \text{if } U_{n+1} \leq q^1_{1+T_n}, \end{cases} & \text{if } X_n > 0, \\ \begin{cases} (0, 1 + T_n), & \text{if } U_{n+1} > q^0_{1+T_n}, \\ (I_{n+1}, 0), & \text{if } U_{n+1} \leq q^0_{1+T_n}, \end{cases} & \text{if } X_n = 0. \end{cases}$$

- Stay in rain period
- Stay in dry spell
- Switch from rain to dry spell or from dry spell to rain

Model with gaussian copula (1/2)

Let us have:

To control rain occurrence

- $(q_n^0)_{n\in\mathbb{N}}$, $(q_n^1)_{n\in\mathbb{N}}$ so that waiting-times follow ext-GPD1 distributions.
- $(U_n)_{n\in\mathbb{N}}$ i.i.d. standard uniform random variables.

Random variables controlling rain intensity

- $\xi > 0, \sigma > 0, \kappa > 0$ parameters of an extended GPD of type 1
- $(N_n)_{n\in\mathbb{N}}$ i.i.d. standard normal random variables, $\rho \in (0,1)$.
- Denote $F_{\xi,\sigma,\kappa}$ c.d.f. of a extended GPD of type 1, Φ c.d.f. of a standard normal distribution

Model with gaussian copula (2/2)

Rainfall model with independent intensities

Let $x_0 \in \mathbb{R}^+$, and $t_0 \in \mathbb{N}^*$.

$$(X_0, T_0) = (x_0, t_0),$$

and recursively $\forall n \in \mathbb{N}$:

$$(X_{n+1}, T_{n+1}) =$$

$$\begin{cases} \begin{cases} (F^{-1}_{\xi,\sigma,\kappa}(\Phi(\rho\Phi^{-1}(F_{\xi,\sigma,\kappa}(X_n)) + \sqrt{1-\rho^2}N_{n+1})), 1 + T_n), & \text{if } U_{n+1} > q^1_{1+T_n}, \\ (0, 0), & \text{if } U_{n+1} \leq q^1_{1+T_n}, \end{cases} & \text{if } X_n > 0, \\ \begin{cases} (0, 1 + T_n), & \text{if } U_{n+1} > q^0_{1+T_n}, \\ (F^{-1}_{\xi,\sigma,\kappa}(\Phi(N_{n+1})), 0), & \text{if } U_{n+1} \leq q^0_{1+T_n}, \end{cases} & \text{if } X_n = 0. \end{cases}$$

- Stay in rain period
- Stay in dry spell
- Switch from rain to dry spell or from dry spell to rain

Purpose of the study

Quick overview of SWG

Rainfall occurrence

Rainfall intensities

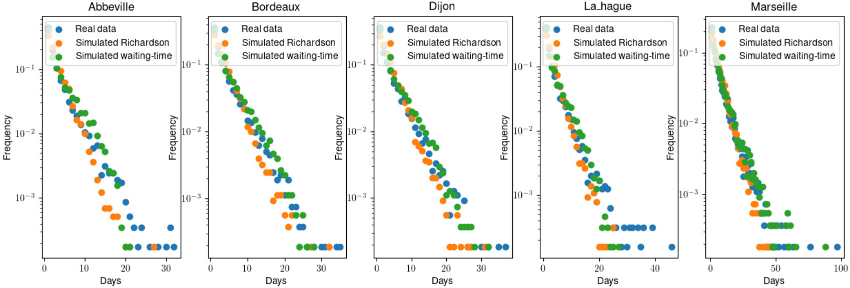
Results and next steps

Appendix

Dry spell distribution

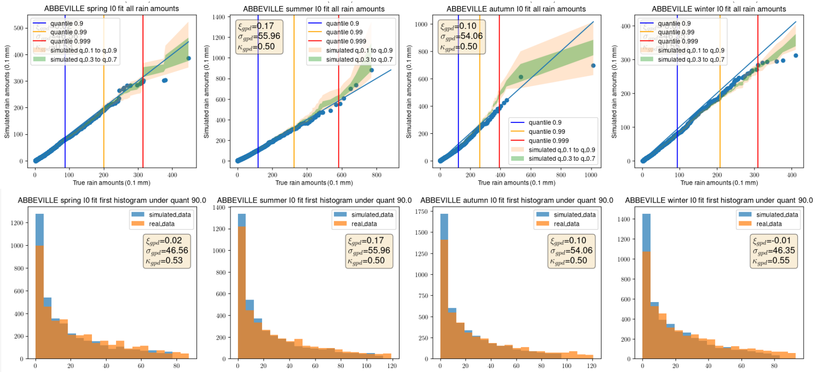
Simulated dry spell and rain spell duration distribution follow ext-GPD.

Fitted with PWM method.

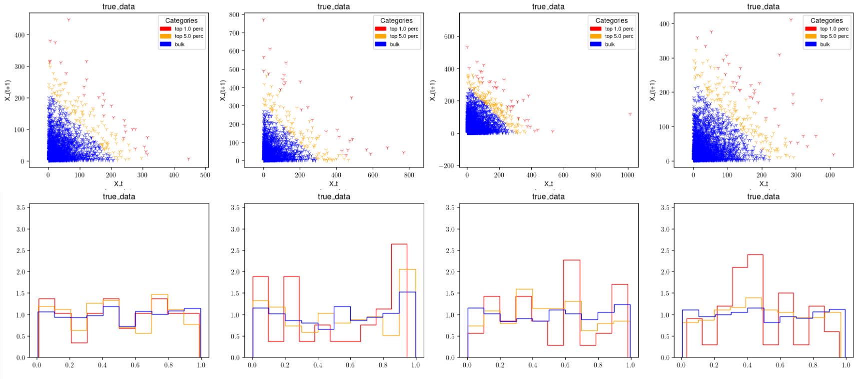


Fit i.i.d. ext-GPD distributions to marginal intensities

- ext-GPD of type 1 (better than type 2-3-4)
- Fit by season
- PWM fit
- Fit on one rain intensity value among each rain periods (for declustering)



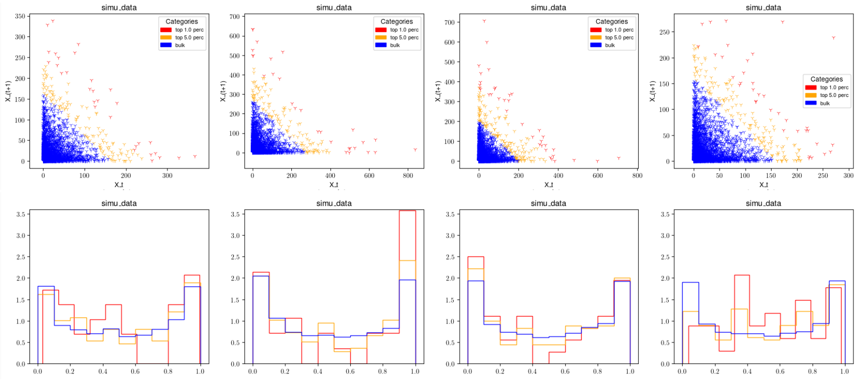
Rain intensity dependency: Abbeville true data



Check angle distribution of consecutive intensities $\frac{X_t}{X_t+X_{t+1}}$

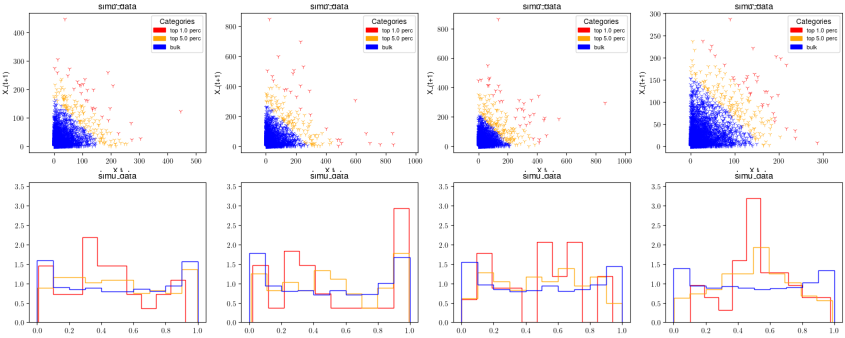
For some stations X season (here in summer): spikes at the edges (independent extremes).

Abbeville simulated data, no dependence



Without dependence: too spiky at the edges

Abbeville simulated data, gaussian copula



We have a little less spikes at the edges (comparison in appendix)

Conclusion and next steps

Done

1. Design of a Stochastic Rain generator
2. Control on dry spell distribution and rain spell distribution
3. Control on extreme rainfall intensity

Next steps

1. Finalize intensity marginals fit and waiting time fit (Bernstein extended-GPD)
2. Work on rainfall intensities dependency (gaussian copula, Stochastic Reccurrent Equation)
3. Theoretical analysis of rainfall occurrence model.
4. Spatialization model

Thank you for your listening !

[1] Richardson, C. "Stochastic simulation of daily precipitation, temperature, and solar radiation", Water Resources Research, 1981

[2] Tomasz J. Kozubowski, Dorota Młynarczyk, Anna K. Panorska "Waiting time representation of discrete distributions", Statistics and Probability Letters, 2025

[3] Naveau, P., R. Huser, P. Ribereau, and A. Hannart, Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, Water Resour. Res., 52, 2753-2769, 2016

[4] Ailliot, P., Allard, D., et al. "Stochastic weather generators: an overview of weather type models". Journal de la société française de statistique, 156(1), 101-113, 2015

SRE model

$(I_n)_{n\in\mathbb{N}}$ i.i.d. positive random variables (we use extended GPD distribution)
$(A_n)_{n\in\mathbb{N}}$ i.i.d. positive random variables (to adapt)

define recursively, for all $n \in \mathbb{N}$:

$$(X_{n+1}, T_{n+1}) =$$

$$\begin{cases} \begin{cases} (A_{n+1}X_n + I_{n+1}, 1 + T_n), & \text{if } U_{n+1} > p^1_{1+T_n}, \\ (0,0), & \text{if } U_{n+1} \le p^1_{1+T_n}, \end{cases} & \text{if } X_n > 0, \\ \begin{cases} (0, 1 + T_n), & \text{if } U_{n+1} > p^0_{1+T_n}, \\ (I_{n+1}, 0), & \text{if } U_{n+1} \le p^0_{1+T_n}, \end{cases} & \text{if } X_n = 0. \end{cases}$$

Rainfall dependency true vs independent vs gaussian copula