Modeling daily precipitation occurrence, with long periods of drought

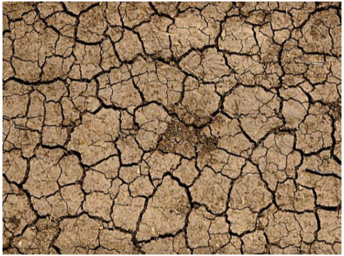
Antoine Doizé

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

Sorbonne-Université - Funded by Geolearning Chair

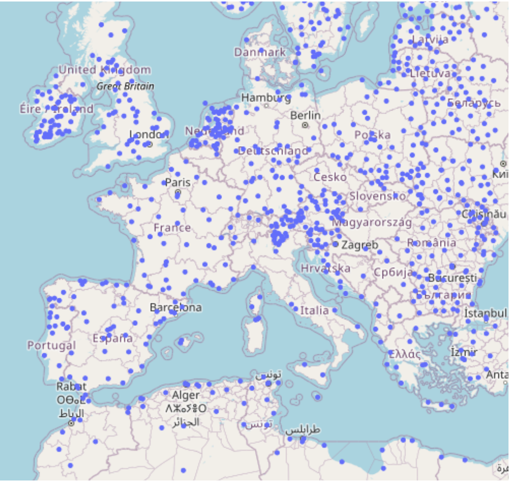
December 03, 2025

Droughts and floods have direct and indirect effects

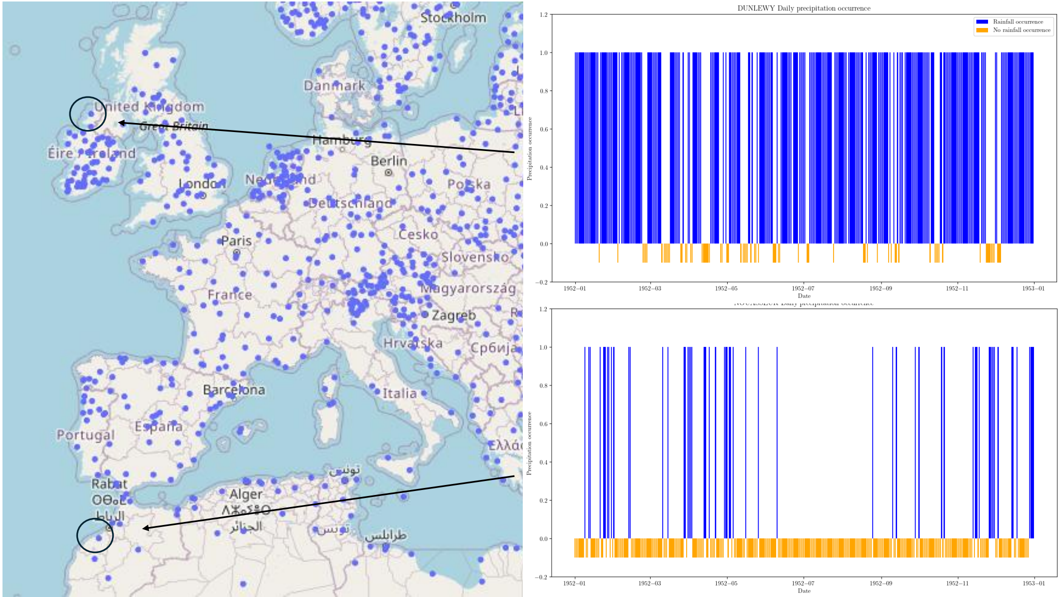


European Climate Assessment & Dataset



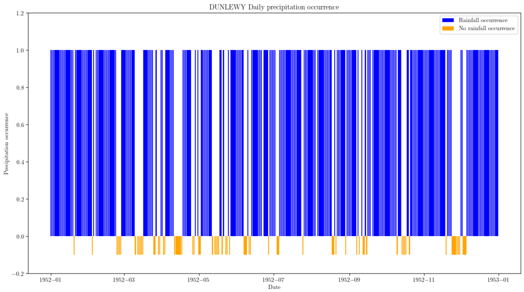
Data processing

1. Europe stations recording daily rainfall.
2. 50 stations by country, more than 50 years of data since 1945, less than 5% missing values.
3. Dropped incomplete spells when missing values.
4. Daily rainfall recordings $\leq$ 0.6mm are considered as dry.
5. $\sim$1000 stations considered.


Stockholm
Danmark
United Kingdom
Great Britain
Éire / Ireland
Hamburg
Berlin
London
Nederland
Deutschland
Polska
Paris
Česko
Slovensko
France
Magyarország
Zagreb
Hrvatska
Србија
Barcelona
Italia
Portugal
España
Rabat
Alger
تونس
الجزائر
طرابلس
DUNLEWY Daily precipitation occurrence
Rainfall occurrence
No rainfall occurrence
Precipitation occurrence
Date
1952–01
1952–03
1952–05
1952–07
1952–09
1952–11
1953–01

Scenarios used for risk assessment

Recorded data



- Limitated to recorded duration ($\sim$ 50 years).
- No internal variability.
- No extrapolation beyond recorded values.

Stochastic Weather Generator

- Unlimited scenario length.
- Produce several scenarios (internal variability).
- Potential extreme extrapolation.

Spell lengths distribution

We focus on rainfall occurrence on each day n,

$$R_n := \mathbb{1}_{\{\text{Rain has been recorded on day } n\}}.$$

In an homogeneous Markov chain, the spell lengths are i.i.d. (strong Markov property), so we can study the spell length distribution, denoted

$$\tau^{(r)},$$

$r = 0$ for the distribution of a dry spell, $r = 1$ for the distribution of a rain spell.

Focus: Markov models for rainfall occurrence modeling[1]

Reminder: Geometric sojourn times

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

Denoting R_n the random variable denoting rainfall occurrence,

$$\mathbb{P}_{R_0=0}\left(R_1 = 0, \ldots, R_n = 0\right) = \left(p_{0,0}\right)^n = \exp\left(-n(-\ln(p_{0,0}))\right)$$

We could derive similar exponentially tailed distributions with:

1. Higher order finite state space Markov chains.
2. Hidden state space Markov chains.

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

Survival function of $\tau^{(0)}$

At a given station let us consider

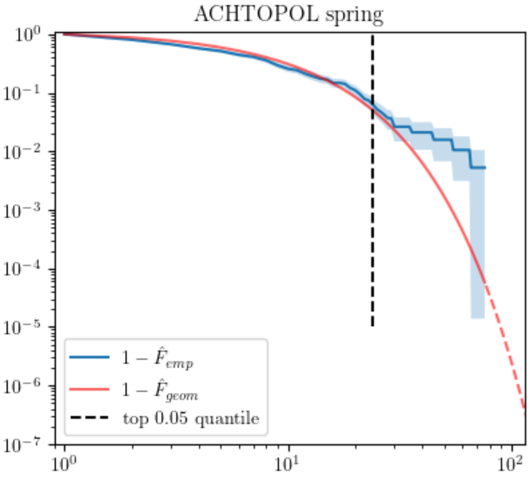
Empirical survival function

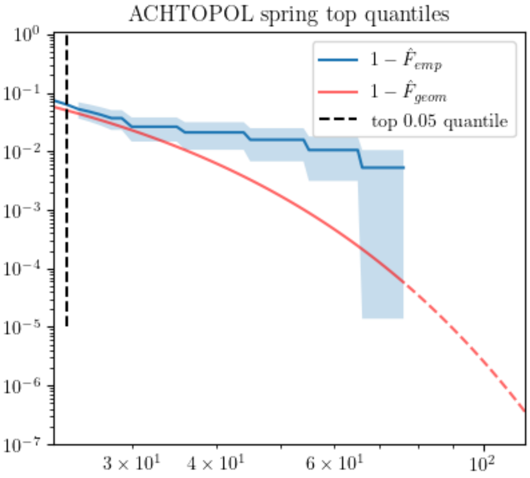
$$1 - \hat{F}_{emp}(d) = K^{-1} \sum_{k=1}^{K} \mathbf{1}\{\tau_k^{(0)} > d\},$$

Geometric survival function

$$1 - \hat{F}_{\mathrm{geom}}(d) = (1 - \hat{p})^d.$$

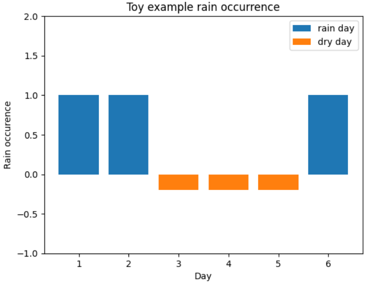
Survival function of $\tau^{(0)}$ in log-log scale

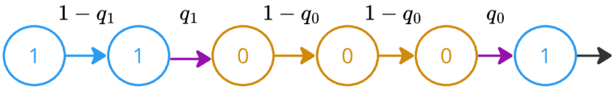




Modeling rainfall occurrence: intuition

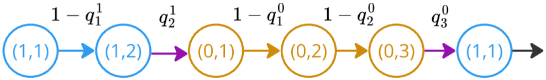
Rainfall occurrence toy data for 6 days:





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

$$\mathbb{P}(\cdot) = \mathbb{P}_{(q_0, q_1)}(\cdot), \quad \mathbb{P}(\tau^{(0)} = d) = (1 - q_0)^{d-1}\, q_0$$



Enlarged Markov model $(R_n, D_n)_{n=0\dots}$

$$\mathbb{P}(\cdot) = \mathbb{P}_{(q_d^{(0)}, q_d^{(1)}),\, d=1,2\dots}(\cdot), \quad \mathbb{P}(\tau^{(0)} = d) = \left(\prod_{k=1}^{d-1} 1 - q_k^{(0)}\right) q_d^{(0)}$$

Waiting Time Representation of a discrete distribution

Proposition (adapted from (Kozubowski)[2])

The distribution of $\tau^{(r)}$ is in one-to-one correspondence with the sequence $\{q_d^{(r)}\}_{d\geq 1}$ defined by

$$q_d^{(r)} = \begin{cases} \mathbb{P}(\tau^{(r)} = d \mid \tau^{(r)} \geq d), & \text{if } \mathbb{P}(\tau^{(r)} \geq d) > 0, \\ 1, & \text{otherwise.} \end{cases}, \forall\, d \geq 1 \quad (1)$$

- Choosing $\tau^{(r)}$ distribution (alternating renewal chain) uniquely determines $(q_d^{(r)})_{d=1\dots}$.
- Choosing $(q_d^{(r)})_{d=1\dots}$ (Markov chain) uniquely determines $\tau^{(r)}$ distribution (we have a condition on the sequence $(q_d^{(r)})_{d=1\dots}$, see Appendix for details).

[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 (R_n, D_n), let us control spell length distribution.

1. If $\tau^{(r)}$ has geometric distribution, $q_d^{(r)} := q \in (0,1)$.
2. If $\tau^{(r)}$ has discrete Weibull distribution, $q_d^{(r)} := 1 - \exp(-\lambda(d+1)^{\beta} - d^{\beta})$.
3. If $\tau^{(r)}$ has discrete Pareto distribution, $q_d^{(r)} := 1 - (\frac{1+\sigma\alpha d}{1+\sigma\alpha(d+1)})^{1/\alpha}$.
4. If $\tau^{(r)}$ has discrete extended-GPD distribution, $q_d^{(r)} := \frac{G(H(\frac{d+1}{\sigma})) - G(H(\frac{d}{\sigma}))}{1-G(H(\frac{d}{\sigma}))}$ (with given G and H, details in appendix).

Flexible rainfall occurrence Markov model

Model definition

Let us have $\{q_d^{(r)}\}_{d\geq 1}$, $r = 0, 1$, sequences in $(0, 1)$.
For given initial values $r_0 \in \{0, 1\}$ and $d_0 \in \mathbb{N}$, set $(R_0, D_0) = (r_0, d_0)$, and for all $n \in \mathbb{N}$:

$$(R_{n+1}, D_{n+1}) = \begin{cases} (1 - R_n, 1), & \text{with probability } q_{D_n}^{(R_n)}, \\ (R_n, D_n + 1), & \text{with probability } 1 - q_{D_n}^{(R_n)}. \end{cases}$$

Parameter estimation:

- Choose a parametric spell duration distribution $\tau^{(0)}, \tau^{(1)}$.
- Estimate the parameters on $(\tau_k)_{k=1...K}$.
- Retrieve the sequences $\{q_d^{(r)}\}_{d\geq 1}$.

Focus on dry spell duration distribution

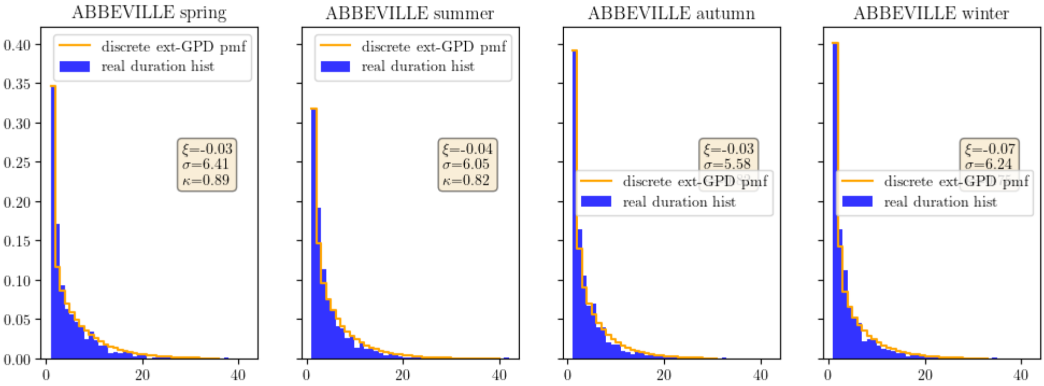
Let $\tau^{(0)}$ follow a degenerate mixture of:

1. A mode in 1,
2. A discretized extended-Generalized Pareto distribution (deGPD) of type 1.

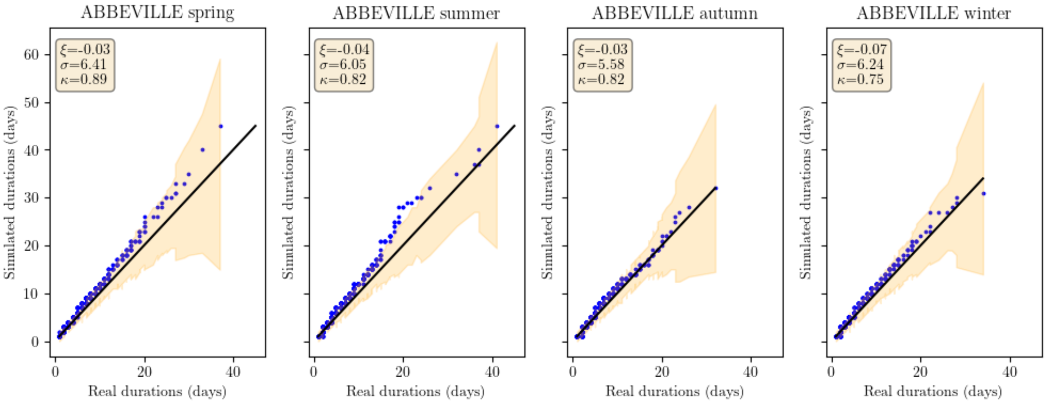
$$\mathbb{P}(\tau^{(0)} = d) = \mathbf{1}_{d=1} f_1 + \mathbf{1}_{d \geq 2} \mathbb{P}\left(D_{\kappa,\sigma,\xi} = d - 2\right),$$

with $\mathbb{P}\left(D_{\hat{\kappa},\hat{\sigma},\hat{\xi}} \leq d\right) = \left(1 - (1 + \hat{\xi} d / \hat{\sigma})^{-1/\xi}\right)^{\hat{\kappa}}$, for any discrete $d \geq 0$.

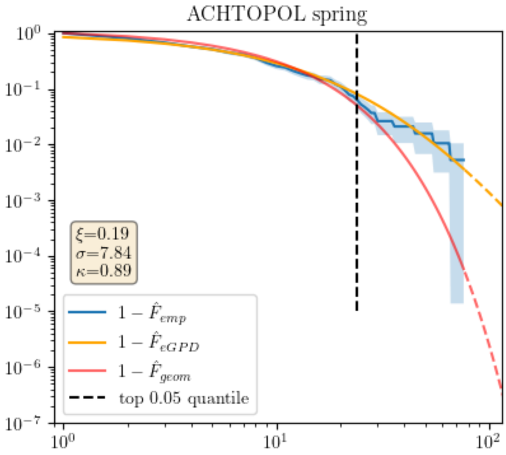
Histogram

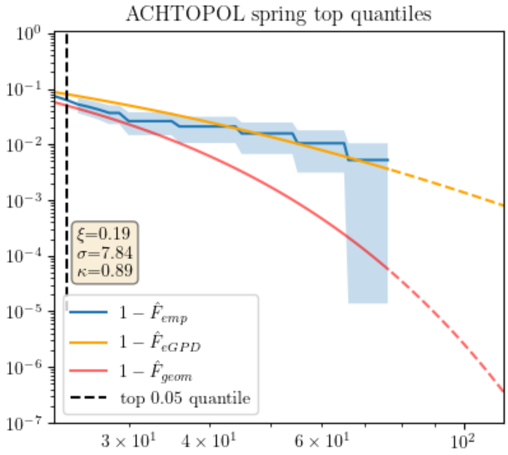

ABBEVILLE spring
discrete ext-GPD pmf
real duration hist
ξ=-0.03
σ=6.41
κ=0.89
ABBEVILLE summer
discrete ext-GPD pmf
real duration hist
ξ=-0.04
σ=6.05
κ=0.82
ABBEVILLE autumn
ξ=-0.03
σ=5.58
discrete ext-GPD pmf
real duration hist
ABBEVILLE winter
ξ=-0.07
σ=6.24
discrete ext-GPD pmf
real duration hist
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0
20
40

QQplots


ABBEVILLE spring
ξ=-0.03
σ=6.41
κ=0.89
ABBEVILLE summer
ξ=-0.04
σ=6.05
κ=0.82
ABBEVILLE autumn
ξ=-0.03
σ=5.58
κ=0.82
ABBEVILLE winter
ξ=-0.07
σ=6.24
κ=0.75
Simulated durations (days)
Real durations (days)
0
10
20
30
40
50
60

Survival function of dry spell duration when $\hat{\xi} > 0$





Probability of ending spell

$$(R_{n+1}, D_{n+1}) = \begin{cases} (1 - R_n, 1), & \text{with probability } q_{D_n}^{(R_n)}, \\ (R_n, D_n + 1), & \text{with probability } 1 - q_{D_n}^{(R_n)}. \end{cases}$$

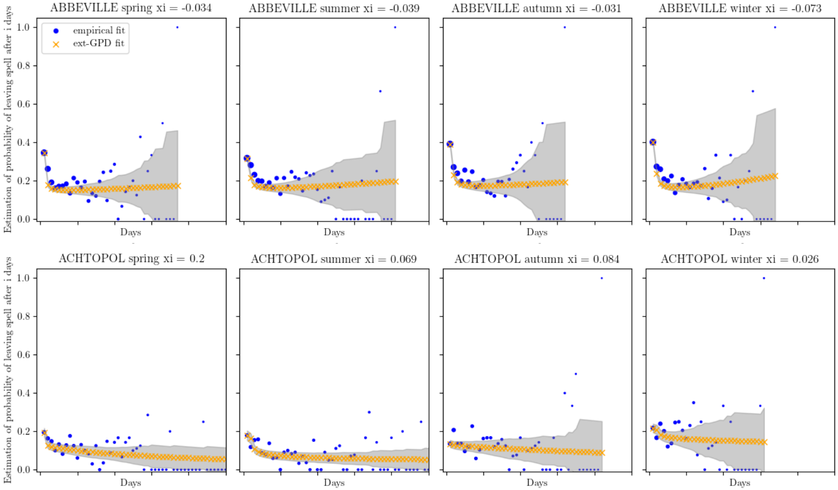
We can compare $\hat{q}_{d,eGPD}$, from the fitted $\tau^{(0)}$ with relation :

$$q_d^{(r)} = \begin{cases} \mathbb{P}(\tau^{(r)} = d \mid \tau^{(r)} \geq d), & \text{if } \mathbb{P}(\tau^{(r)} \geq d) > 0, \\ 1, & \text{otherwise.} \end{cases}, \ \forall\, d \geq 1$$

to an "empirical" fit on the recorded spell durations:

$$\hat{q}_{d,emp}^{(0)} := \frac{\text{Number of spells of length d}}{\text{Number of spells longer than d}}$$

Probability of ending spell



- Smooth fit of ending spell probability.
- We catch the persistence (decreasing probabilities especially in the first days).
- Asymptotic slope linked to sign of ξ.

Map of $\hat{\xi}$ in Spring

Conclusion and extensions

Done

1. Flexible rainfall occurrence model.
2. Control on dry spell tail distribution.
3. Better persistence modeling.

Next steps

1. Spatialization model of the model.

$$(R_{n+1}^{(j)}, D_{n+1}^{(j)}) = \begin{cases} (1 - R_n^{(j)}, 1), & \text{if } U_{R_n^{(j)}}(j) < q_{R_n,D_n}^{(j)}, \\ (R_n^{(j)}, D_n^{(j)} + 1), & \text{if } U_{R_n^{(j)}} > q_{R_n,D_n}^{(j)}. \end{cases}$$

   with $U \sim C(\cdot)$ a chosen copula.

2. Rainfall intensity modeling.

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

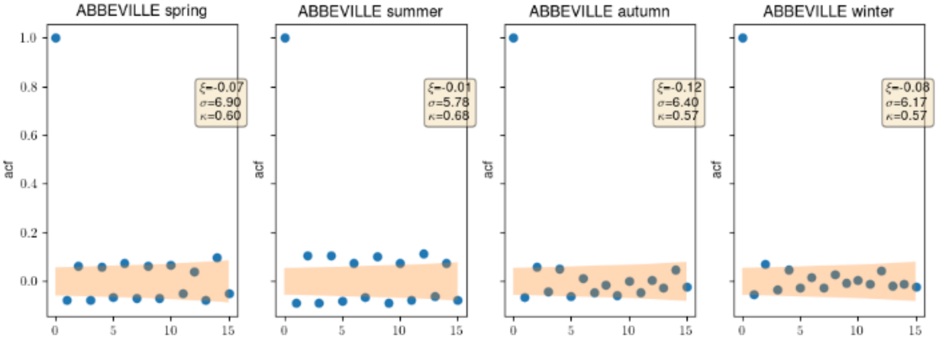
[3] Naveau, P., R. Huser, P. Ribereau, and A. Hannart, Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, Water Resour. Res., 52, 2753-2769, 2016

[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

[5] S. I. Resnick. "Adventures in Stochastic Processes". Birkhäuser Boston, 1992.

Check modeling hypothesis

In an altenating renewal model, we suppose mutual independance of the $(\tau_k^{(r)})_{k=1...}$. We check the autocorrelation.



Condition on the sequence $(q_d^{(r)})_{d=1...}$

$$\tau_1^{(0)} < \infty \;\; a.s.$$

if and only if

$$\sum_{d=1}^{\infty} q_d^{(0)} = \infty.$$

We consider this condition in order for the alternating renewal chain modeling to be relevant.