Modeling daily precipitation occurrence, with long periods of drought

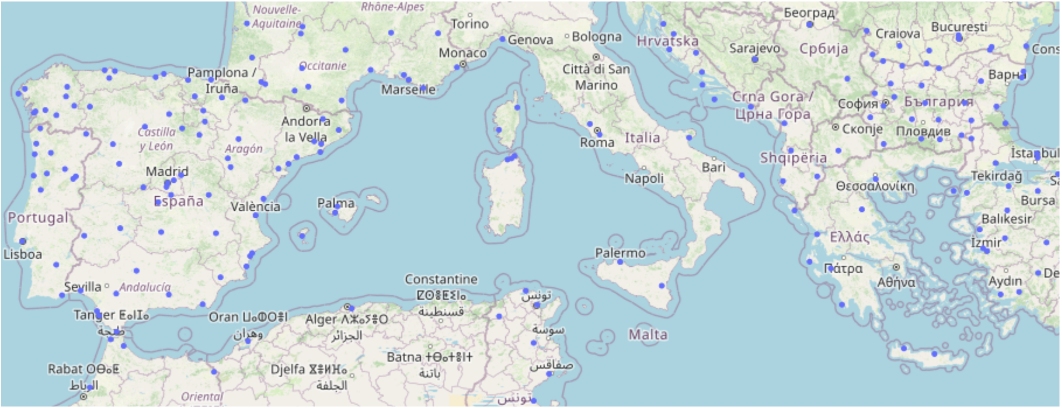
Antoine Doizé

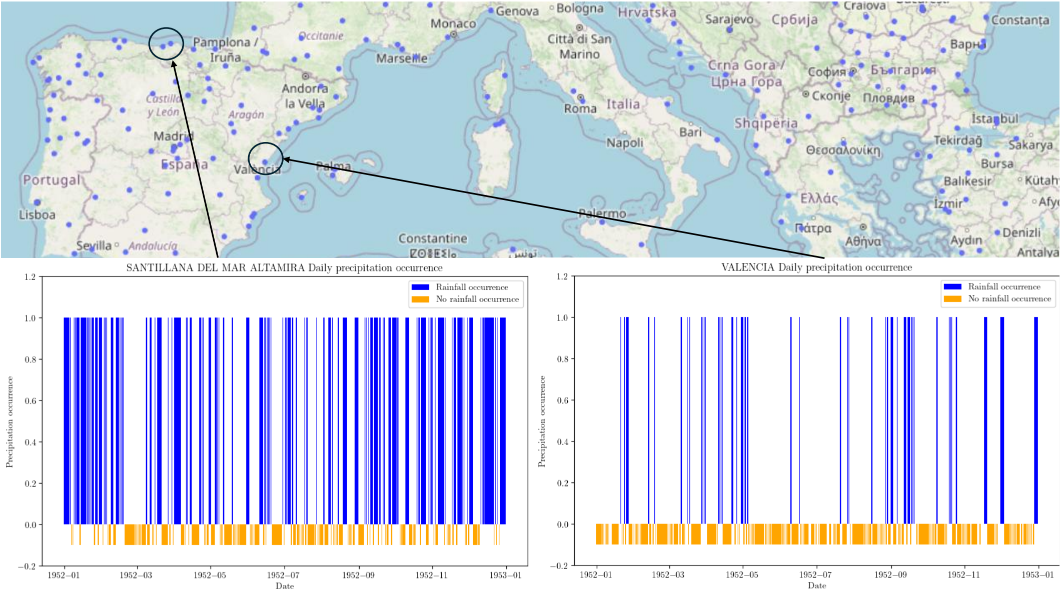
Advisors : D. Allard, P. Naveau, O. Wintenberger

Sorbonne-Université - Funded by Geolearning Chair

April 8, 2026

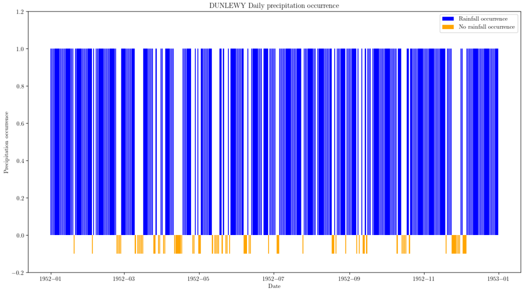
European Climate Assessment & Dataset




Genova
Bologna
Hrvatska
Sarajevo
Србија
Craiova
Constanța
Monaco
Pamplona / Iruña
Occitanie
Marseille
Città di San Marino
Варна
Crna Gora / Црна Гора
София
България
Andorra la Vella
Castilla y León
Aragón
Roma
Italia
Скопje
Пловдив
İstanbul
Shqipëria
Bari
Madrid
Napoli
Θεσσαλονίκη
Tekirdağ
Sakarya
España
Valencia
Palma
Bursa
Portugal
Balıkesir
Kütahya
Ελλάς
İzmir
Lisboa
Palermo
Πάτρα
Αθήνα
Aydın
Denizli
Sevilla
Andalucía
Constantine
Antalya
SANTILLANA DEL MAR ALTAMIRA Daily precipitation occurrence
Rainfall occurrence
No rainfall occurrence
Precipitation occurrence
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
1952−01
1952−03
1952−05
1952−07
1952−09
1952−11
1953−01
Date
VALENCIA Daily precipitation occurrence
Rainfall occurrence
No rainfall occurrence
Precipitation occurrence
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
1952−01
1952−03
1952−05
1952−07
1952−09
1952−11
1953−01
Date

Risk assessment requires several long scenarios

Recorded data



- Limitated to recorded duration ($\sim$ 10-50 years).
- No internal variability.
- No extrapolation beyond recorded values.

Stochastic Weather Generator

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

Markov chain point of view:

$$(R_n)_{n=1,2\dots},$$

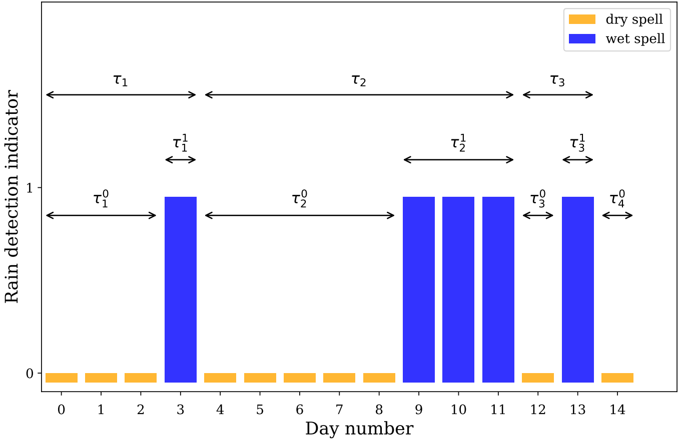
$R_n := \mathbb{1}_{\{\text{Rain has been recorded on day } n\}}$

Alternating Renewal model point of view:

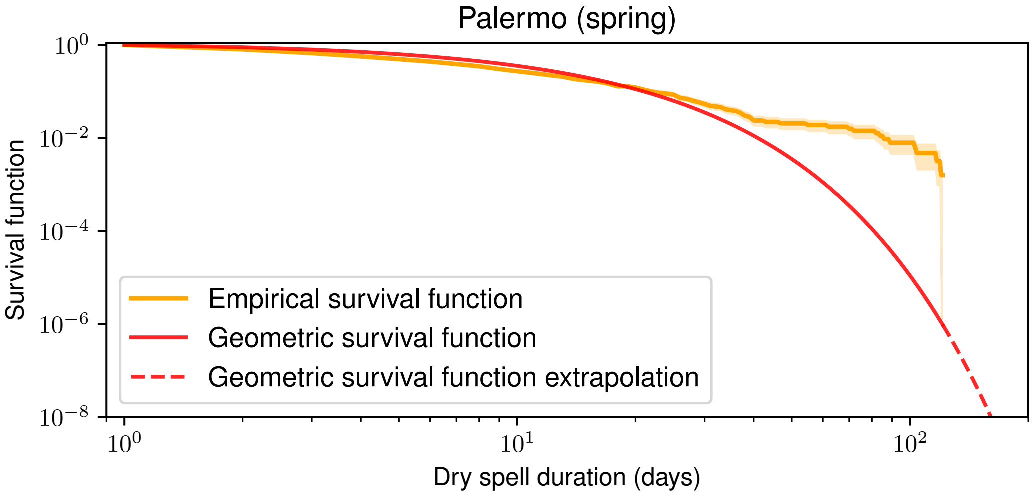
$$(\tau_k^{(0)}, \quad \tau_k^{(1)})_{k=1,2\dots},$$

$\tau_k^{(0)}$: duration of k^{th} dry spell, $\tau_k^{(1)}$: duration of k^{th} wet spell.

Binary Markov Chain with Duration is compatible with both points of view.

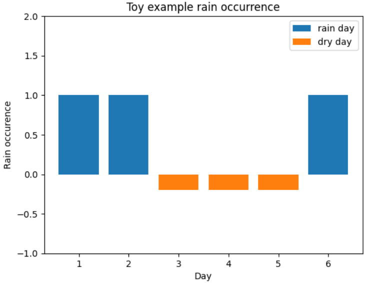


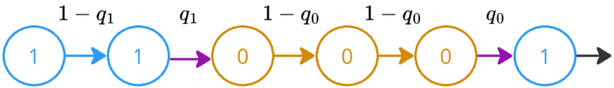
Survival function of $\tau^{(0)}$: empirical vs geometric fit



Modeling rainfall occurrence: intuition

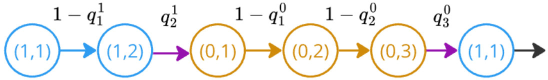
Rainfall occurrence toy data for 6 days:





two-states first-order Markov model $(R_n)_{n=0\dots}$

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



Binary Markov Chain with Duration $(R_n, D_n)_{n=0\dots}$

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

Binary Markov Chain with Duration (BMCD)

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 $n \in \mathbb{N}$:

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

Link between BMCD and spell duration

Proposition (adapted from (Kozubowski)[1])

Let us have a distribution on $\tau^{(r)}$. Let us have a BMCD with parameters $\{q_d^{(r)}\}_{d\geq 1}$ given by:

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

This model has spell duration distributed as $\tau^{(r)}$.

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

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

$\tau^{(r)}$ distribution

$\tau^{(0)}$ **"hdeGP" distribution:** $\mathbb{P}_{f_1,\kappa,\sigma,\xi}(\tau^{(0)} = d) = \begin{cases} f_1, & d = 1, \\ (1 - f_1)\, F_{\kappa,\sigma,\xi}(d-1), & d \geq 2, \end{cases}$

with $F_{\kappa,\sigma,\xi}$ a discretized [2]type 1 extended Generalized Pareto Distribution (eGPD) probability mass function: TL; DR:

1. $\xi > 0$: heavy-tailed
2. $\xi \simeq 0$: exponential tail
3. $\xi < 0$: right-bounded tail

$\tau^{(1)}$ **mixt. geom. distribution:** $\mathbb{P}_{\pi,p_1,p_2}(\tau^{(1)} = d) = \pi\, p_1(1-p_1)^{d-1} + (1-\pi)\, p_2(1-p_2)^{d-1}.$

[2] P. Naveau, R. Huser, P. Ribereau, and A. Hannart. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. Water Resources Research, 52(4):2753–2769, 2016

Parameter estimation

$\tau^{(0)}$ parameters: $(f_1, \kappa, \sigma, \xi)$

$\widehat{f_1} = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}\{\tau_k^{(0)} = 1\}$.

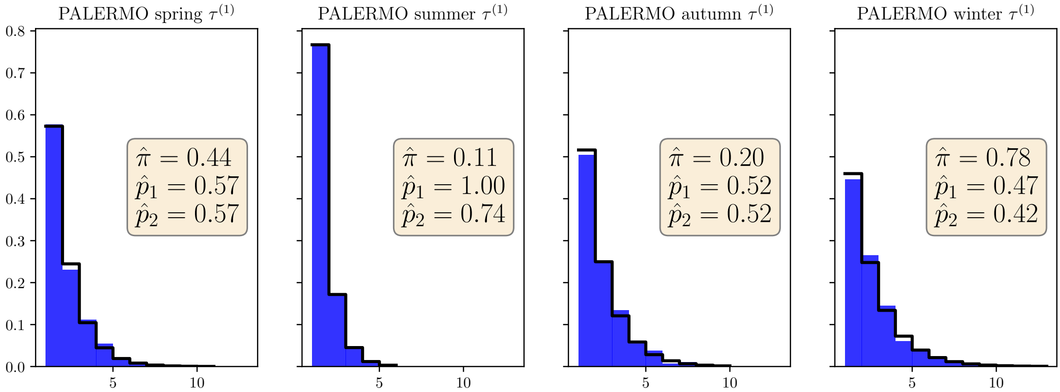
(κ, σ, ξ) estimated by [2]Probability Weighted Moments method

$\tau^{(1)}$ parameters: (π, p_1, p_2)

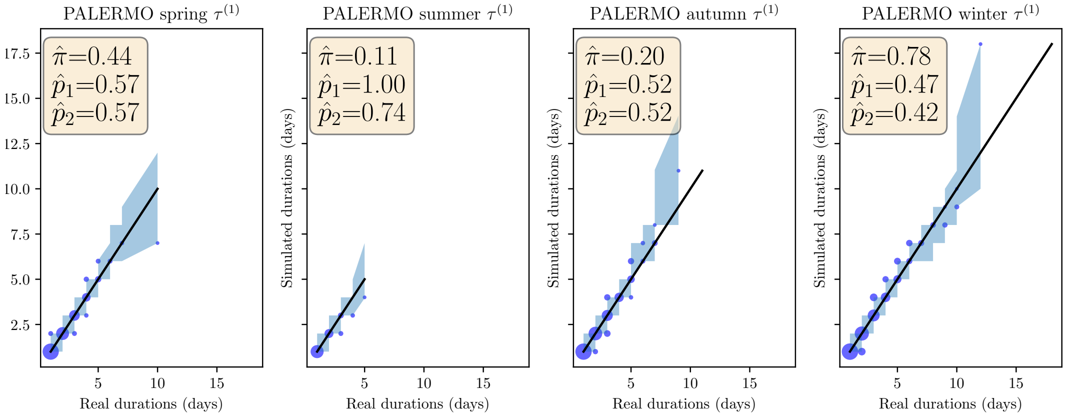
Estimated by Expectation-Maximization algorithm.

[2] P. Naveau, R. Huser, P. Ribereau, and A. Hannart. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. Water Resources Research, 52(4):2753–2769, 2016

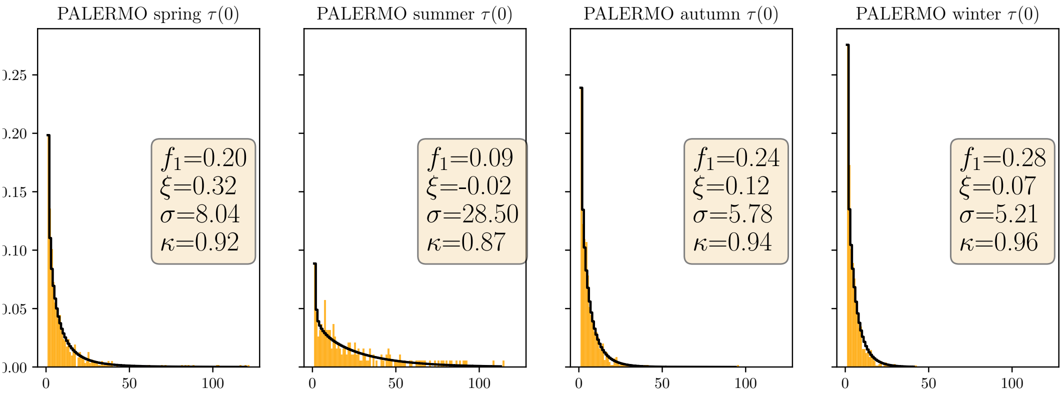
Histogram wet spell duration



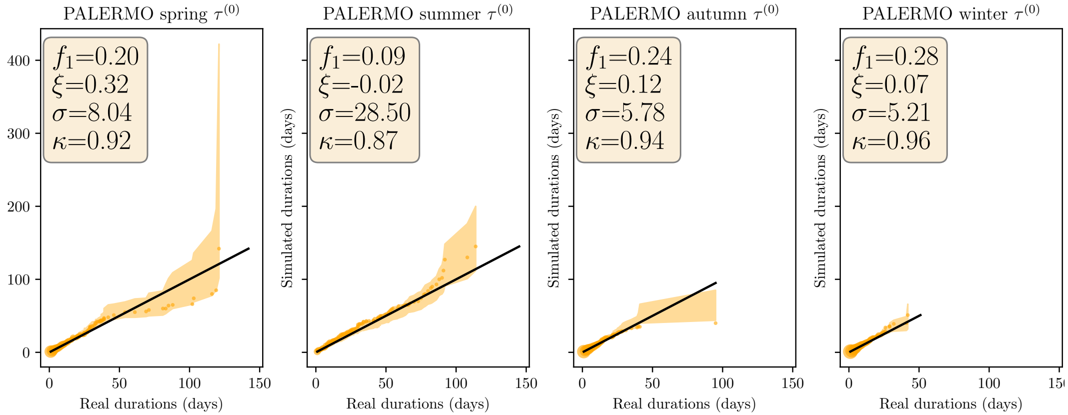
QQplots wet spell duration



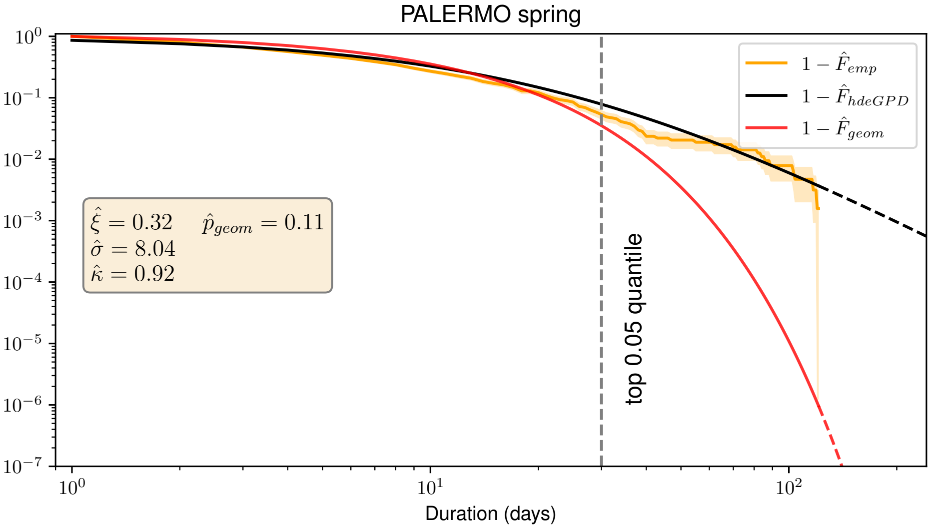
Histogram dry spell duration



QQplots dry spell duration



Survival function of dry spell duration for Palermo spring ($\hat{\xi} = 0.32$)



Consequence: Share of Time Spent in Severe Dry Spells

Main result

Our representation gives asymptotic results on quantities such as:

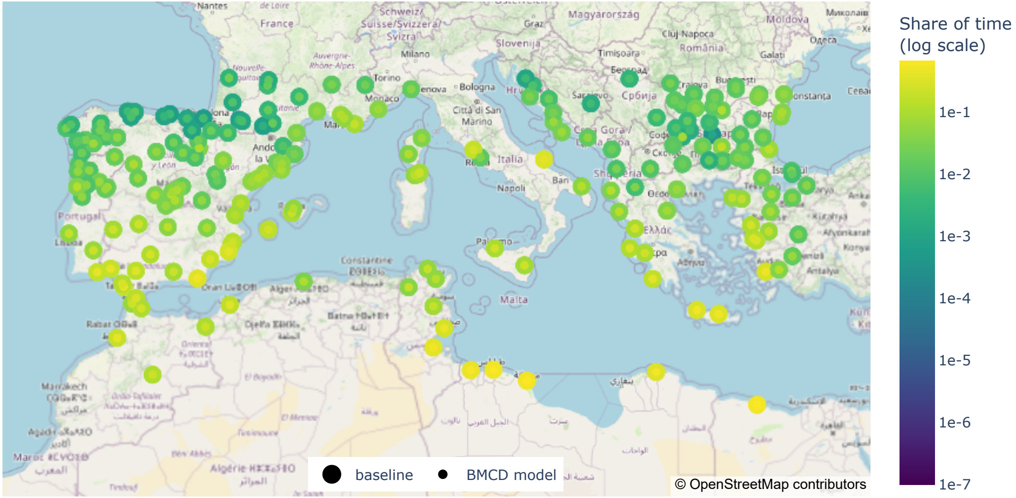
$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n} \mathbb{1}_{\{R_k=0,\ D_k\geq d\}} = \frac{\mathbb{E}\left[\max\left(0, \tau^{(0)} - d\right)\right]}{\mathbb{E}[\tau]} \qquad \text{a.s.}$$

Interpretation: this is the long-run share of time spent in dry spells of length at least d.

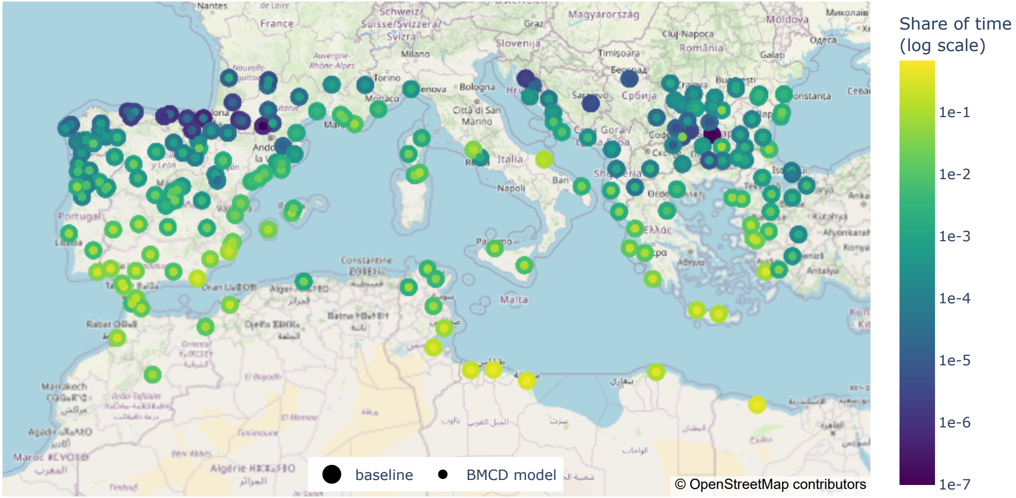
Comparison:

- geometric model,
- our model.

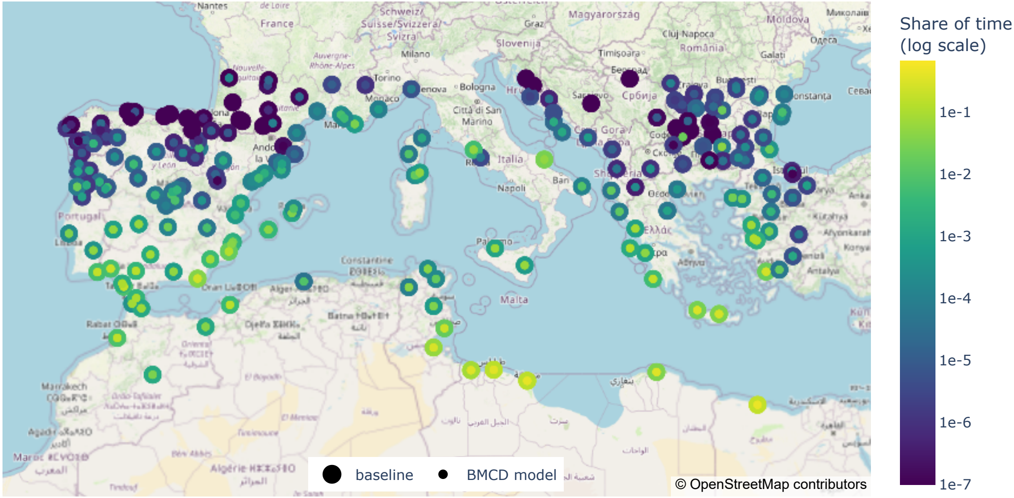
Dry spell share ≥ 20 days



Dry spell share ≥ 40 days



Dry spell share ≥ 60 days



Spatial extension of the model [3]

Single site model:

$$(R_{n+1}, D_{n+1}) = \begin{cases} (1 - R_n, 1), & \text{w. prob. } q^{(R_n)}(D_n), \\ (R_n, D_n + 1), & \text{w. prob. } 1 - q^{(R_n)}(D_n). \end{cases}$$

Spatial extension, define for any $\mathbf{s}$ in a domain $\mathcal{D}$:

$$Z_n(\mathbf{s}) = (-1)^{R_n(\mathbf{s})}\, Y_n(\mathbf{s}), \quad z_n(\mathbf{s}) = \Phi^{-1}\left[q^{(R_n(\mathbf{s}))}(D_n(\mathbf{s}))\right],$$

with $(Y_n)_{n\geq 1}$ i.i.d. Gaussian processes on $\mathcal{D}$, having covariance functions $(C_n)_{n\geq 1}$, and Φ the c.d.f. of a standard Gaussian.

$$(R_{n+1}(\mathbf{s}), D_{n+1}(\mathbf{s})) = \begin{cases} (1 - R_n(\mathbf{s}),\ 1), & \text{if } Z_n(\mathbf{s}) \leq z_n(\mathbf{s}), \\ (R_n(\mathbf{s}),\ D_n(\mathbf{s}) + 1), & \text{if } Z_n(\mathbf{s}) > z_n(\mathbf{s}). \end{cases}$$

[3] D. S. Wilks. Multisite generalization of a daily stochastic precipitation generation model. Journal of Hydrology, 210(1-4): 178–191, 1998. doi: 10.1016/S0022-1694(98)00186-3.

Conclusion

1. Binary Markov Chain with Duration: (Markov chain / alternating renewal model).
2. Extreme value theory distributions: severe dry spells.
3. "Easy" spatialization.

Thank you for your listening !

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

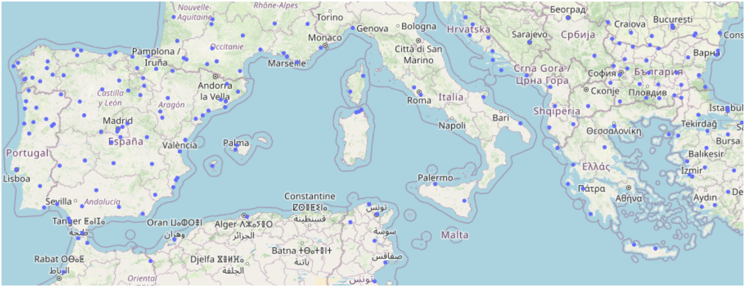
[2] Tomasz J. Kozubowski, Dorota Młynarczyk, Anna K. Panorska "Waiting time representation of discrete distributions", Statistics and Probability Letters, 2025

[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.

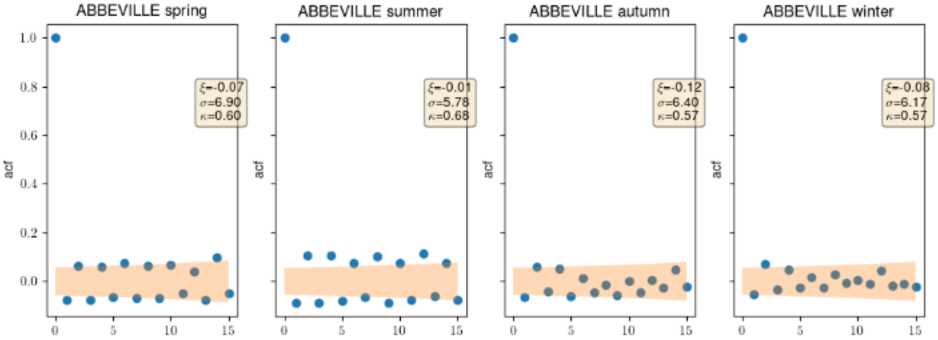
European Climate Assessment & Dataset: data processing



1. South Europe (*lat* < 45) stations daily rainfall records.
2. Only keep stations with more than 50 years of data since 1945, less than 5% missing values.
3. Dropped incomplete spells when missing values.
4. Daily rainfall recordings ≤ 0.6mm are considered as dry.
5. Filter out stations to have a constant "station density".
6. $\sim$200 stations are considered.

Check modeling hypothesis

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



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

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

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.

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).

Coupling dry spell wet spell spatial