Génération stochastiques de précipitations spatio-temporelles à l'aide d'un HMM à émissions binaires, d'un champ Gaussien conditionné et de distributions EGPD

Séminaire de statitiques au sommet de Rochebrune - 24 mars 2026

Caroline Cognot, Liliane Bel, David Métivier et Sylvie Parey

AgroParisTech Talents d'une planète soutenable
INRAE
université PARIS-SACLAY
EDF

Goal : sampling from meteorological variables

Physical Climate Models (e.g. GCMs, RCMs)

- Use physical laws: radiation, fluid dynamics, etc.
- Global and regional simulations

+ Physically consistent, multi-variable outputs

− Heavy computation

Goal : sampling from meteorological variables

Physical Climate Models (e.g. GCMs, RCMs)

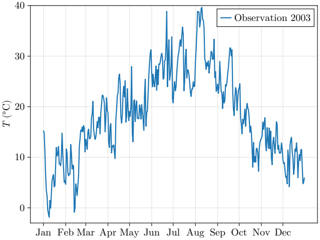
- Use physical laws: radiation, fluid dynamics, etc.
- Global and regional simulations

+ Physically consistent, multi-variable outputs

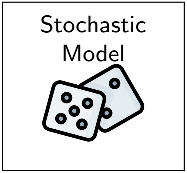
− Heavy computation

Stochastic Weather Generators:

Weather data

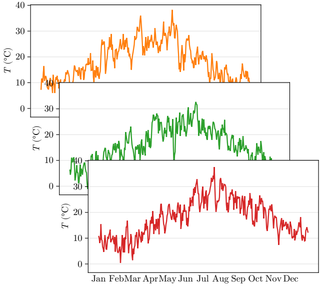


→

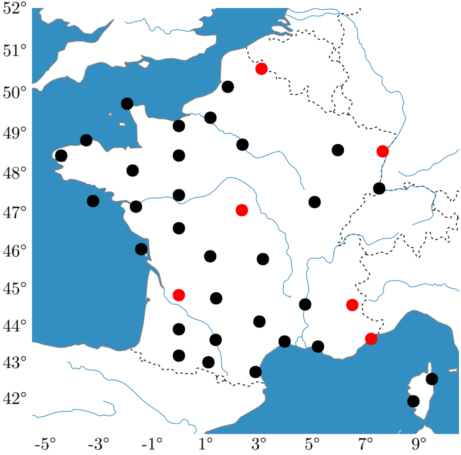


→

Simulations



Objective : precipitation weather generator



$D = 37$ stations, ECA&D - 51 years

Objective: Build a stochastic weather generator

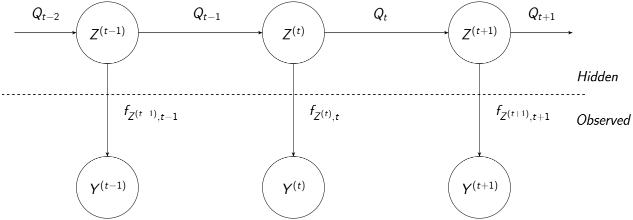
- For the multisite rain occurrence $Y^{(t)} = (Y_1^{(t)}, \ldots, Y_D^{(t)})$
- Then for precipitation intensity $R^{(t)} = (R_1^{(t)}, \ldots, R_D^{(t)})$

→ Reproduce the spatial-temporal structure of the data.

- In particular **large scales dry/wet episodes**

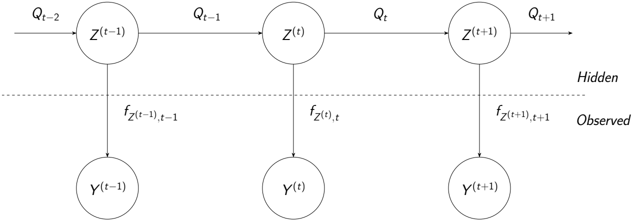
Rain occurrence model (1/3) : Multisite HMM

Rain occurrence $Y^{(t)} = (Y_1^{(t)}, \dots, Y_D^{(t)}) \in \{0,1\}^D$ - Unobserved weather type $Z^{(t)} \in \{1, \dots, K\}$



Rain occurrence model (1/3) : Multisite HMM

Rain occurrence $Y^{(t)} = (Y_1^{(t)}, \dots, Y_D^{(t)}) \in \{0,1\}^D$ - Unobserved weather type $Z^{(t)} \in \{1, \dots, K\}$



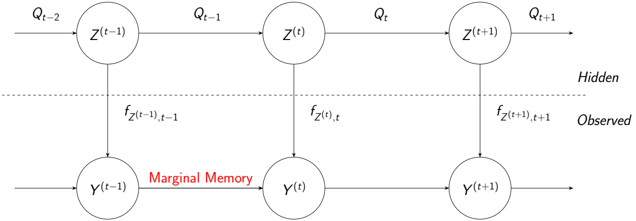
Conditional independence (Zucchini et al. 1991, Gobet et al. 2024)

$$\mathbb{P}\left(Y = y \mid Z = z^{(t)}\right) = f_{z^{(t)},t}(y) = \prod_{s=1}^{S} \left(y_s \lambda_{Z^{(t)},t,s} + (1 - y_s)(1 - \lambda_{Z^{(t)},t,s})\right)$$

With $\lambda_{k,t,i} = \mathbb{P}(Y_i^{(t)} = 1 | Z^{(t)} = k)$ for $i \in 1, \dots, D$

Rain occurrence model (1/3) : Multisite HMM

Rain occurrence $Y^{(t)} = (Y_1^{(t)}, \dots, Y_D^{(t)}) \in \{0,1\}^D$ - Unobserved weather type $Z^{(t)} \in \{1, \dots, K\}$



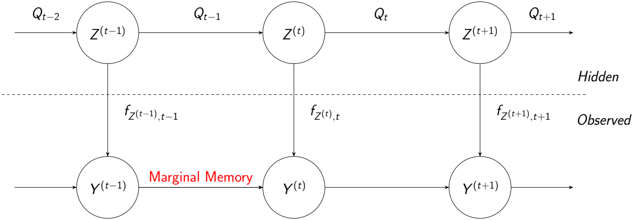
Conditional independence (Zucchini et al. 1991, Gobet et al. 2024)

$$\mathbb{P}\left(Y = y \mid Z = z^{(t)}, Y^{(t-1)} = y^{t-1}\right) = f_{z^{(t)},t}(y) = \prod_{s=1}^{D}\left(y_s \lambda_{Z^{(t)},t,s,y_s^{t-1}} + (1-y_s)(1-\lambda_{Z^{(t)},t,s,y_s^{t-1}})\right)$$

With $\lambda_{k,t,i,y_i^{t-1}} = P(Y_i^{(t)} = 1 | Z^{(t)} = k, Y_i^{(t-1)} = y_i^{t-1})$ for $i \in 1, \dots, D$

Rain occurrence model (1/3) : Multisite HMM

Rain occurrence $Y^{(t)} = (Y_1^{(t)}, \ldots, Y_D^{(t)}) \in \{0,1\}^D$ - Unobserved weather type $Z^{(t)} \in \{1, \ldots, K\}$

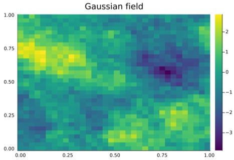


Conditional independence (Zucchini et al. 1991, Gobet et al. 2024)

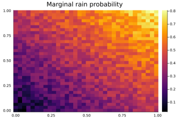
– Stations must be "far apart enough": 10 stations in the paper

+ Correlations between stations are captured by the weather types Z

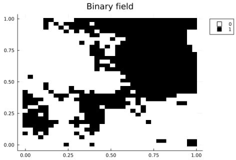
Rain occurrence model (2/3) : generating binary correlated variables



$X \sim \mathcal{N}(0, C_\theta)$: latent

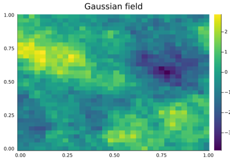


$\xrightarrow{\lambda_s}$

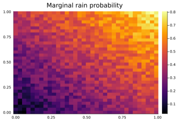


$$\forall s, Y_s = \begin{cases} 1 & \text{if } X_s \leq \Phi^{-1}(\lambda_s) \\ 0 & \text{else} \end{cases}$$

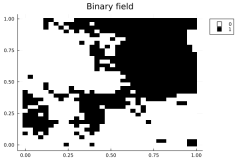
Rain occurrence model (2/3) : generating binary correlated variables



$X \sim \mathcal{N}(0, C_\theta)$: latent



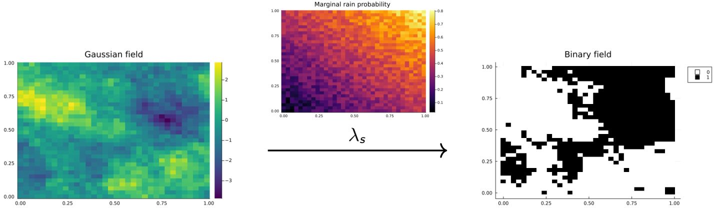
$\xrightarrow{\lambda_s}$



$$\forall s,\ Y_s = \begin{cases} 1 & \text{if } X_s \leq \Phi^{-1}(\lambda_s) \\ 0 & \text{else} \end{cases}$$

Parameters : $\lambda_i = \mathbb{P}(Y_i = 1)$ for $i \in 1, \dots, D$ and θ the parameters of the latent covariance.

Rain occurrence model (2/3) : generating binary correlated variables



$X \sim \mathcal{N}(0, C_\theta)$: latent

$$\forall s, Y_s = \begin{cases} 1 & \text{if } X_s \leq \Phi^{-1}(\lambda_s) \\ 0 & \text{else} \end{cases}$$

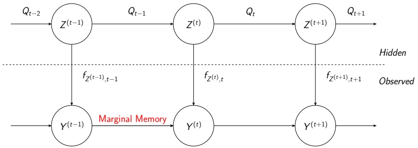
Parameters : $\lambda_i = \mathbb{P}(Y_i = 1)$ for $i \in 1, \ldots, D$ and θ the parameters of the latent covariance.

Associated distribution

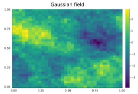
$$f(y, \theta) = \int_{a_1}^{b_1} \cdots \int_{a_D}^{b_D} f_X((x_1, ..., x_D)) d(x_1, ..., x_D)$$

With $a_i = -\infty$ if $Y_i = 1, \Phi^{-1}(\lambda_i)$ else, $b_i = \infty$ if $Y_i = 0, \Phi^{-1}(\lambda_i)$ else

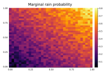
Rain occurrence model (3/3): Multisite HMM



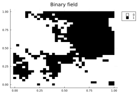
$+$



$X \sim \mathcal{N}(0, C_\theta)$: latent



$\xrightarrow{\lambda_s}$



$$\forall s,\ Y_s = \begin{cases} 1 & \text{if } X_s \leq \Phi^{-1}(\lambda_s) \\ 0 & \text{else} \end{cases}$$

+ No restriction on the stations distance
- The emission is more complicated !

Rain intensity model (1/2) : Marginals

What about **Rain intensity** $(R_1^{(t)}, ..., R_D^{(t)})$?

Rain intensity model (1/2) : Marginals

What about **Rain intensity** $(R_1^{(t)}, ..., R_D^{(t)})$?
Reuse the HMM : $R^{(t)}$ depends on $Z^{(t)}$ and $Y^{(t)}$.

Rain intensity model (1/2) : Marginals

What about **Rain intensity** $(R_1^{(t)}, ..., R_D^{(t)})$?
Reuse the HMM : $R^{(t)}$ depends on $Z^{(t)}$ and $Y^{(t)}$.

Rain intensity model (1/2) : Marginals

What about **Rain intensity** $(R_1^{(t)}, ..., R_D^{(t)})$?
Reuse the HMM : $R^{(t)}$ depends on $Z^{(t)}$ and $Y^{(t)}$.

Extended Generalized Pareto Distribution, Naveau et al. (2016),Gamet et al. (2022)

EGPD models exceedances over a **low** threshold u:

$$K(r; \sigma, \xi) = G(H(r; \sigma, \xi))$$

Rain intensity model (1/2) : Marginals

What about **Rain intensity** $(R_1^{(t)}, ..., R_D^{(t)})$?
Reuse the HMM : $R^{(t)}$ depends on $Z^{(t)}$ and $Y^{(t)}$.

Extended Generalized Pareto Distribution, Naveau et al. (2016),Gamet et al. (2022)

EGPD models exceedances over a **low** threshold u:

$$K(r; \sigma, \xi) = G(H(r; \sigma, \xi))$$

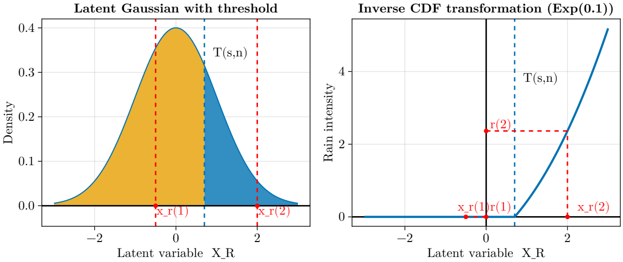
- Need appropriate $G : [0,1] \to [0,1]$: Truncated Beta of Gamet et al. (2022)
- Only need to model the very low values separately

Rain intensity model (2/2) : Dependence

Usual model : Censored, transformed Gaussian

- $F_s^{(n)}$ the marginal cdf
- $T(s,n) = \Phi^{-1}(\mathbb{P}(R_s^{(n)} = 0))$
- $X_R(.,.) \sim \mathcal{N}(0, C_R(.,.))$

$$R_s^{(n)} = \begin{cases} \left(F_s^{(n)}\right)^{(-1)} (\Phi_{T(s,n)}(X_R(s,n))) & \text{if } X_R(s,n) \geq T(s,n) \\ 0 & \text{if } X_R(s,n) < T(s,n) \end{cases}$$

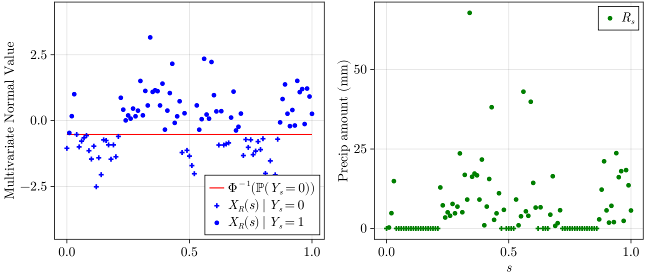


Rain intensity model (2/2) : Dependence

Truncated, transformed Gaussian

- $F^{(n)}_{z^{(n)},s}$ the marginal cdf
- $T(s,n) = \Phi^{-1}(\mathbb{P}(R_s^{(n)} = 0)) = \Phi^{-1}(1 - \lambda^{(n)}_{z^{(n)},s,y_s^{(n-1)}})$
- $X_R(.,n) \mid Y^{(n)}, Z^{(n)} \sim \mathrm{TMVN}(0, C_R, l(.,n), u(.,n))$, $(l(s,n), u(s,n)) = \begin{cases} (-\infty, \; T(s,n)) & \text{if } Y_s^{(n)} = 0, \\ (T(s,n), \; \infty) & \text{if } Y_s^{(n)} = 1. \end{cases}$

$$R_s^{(n)} = \begin{cases} \left(F^{(n)}_{z^{(n)},s}\right)^{(-1)} (\Phi_{T(s,n)}(X_R(s,n))) & \text{if } Y(s,n) = 1 \\ 0 & \text{if } Y(s,n) = 0 \end{cases}$$



Rain parameters - overview

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_{HMM}, \boldsymbol{\theta}_{EGPD}, \boldsymbol{\theta}_{CR})$$

Periodic parameterisation (all but latent Gaussian)

$$\text{For each } \theta, P_\theta(t) = c_0^\theta + \sum_{j=1}^{degP} \left(c_{2j-1}^\theta \cos(2\pi jt/T) + c_{2j}^\theta \sin(2\pi jt/T) \right)$$

$\theta^{(n)} = \psi_\theta(P_\theta(t_n)) = \psi_\theta(P_\theta(n))$ for constraints

Covariance for latent Gaussian : Gneiting-Matérn covariance function

$$C(h, u) = \frac{\sigma^2}{(\alpha u^{2a} + 1)^{b+\delta}} \mathcal{M}\left(\frac{h}{\sqrt{\alpha u^{2a} + 1}^{b}}; r; \nu \right)$$

Inference for Rain Occurrence Model (HMM)

Hidden states $\rightarrow$ Expectation-Maximization (EM).

$$\theta^{(q+1)} \in \arg\max_{\theta} \mathbb{E}_{Z|Y,\theta^{(q)}} \log L(Y, Z|\theta)$$

E-step : computation with classical Forward-Backward of... :

$\pi_{n|N}(k)=\mathbb{P}(Z^{(n)}=k|Y^{(1)},\dots,Y^{(N)})$, $\pi_{n,n+1|N}(k,\ell)=\mathbb{P}(Z^{(n)}=k,Z^{(n+1)}=\ell|Y^{(1)},\dots,Y^{(N)})$

M-step : maximisation of...

$$\mathcal{R}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) = \sum_{k=1}^{K} \pi_{1|N}^{(q)}(k) \log(\eta_k) + \sum_{k,\ell=1}^{K} \sum_{n=1}^{N-1} \pi_{n,n+1|N}^{(q)}(k,\ell) \log(Q^{(n)}(k,\ell))$$
$$+ \sum_{k=1}^{K} \sum_{n=1}^{N} \pi_{n|N}^{(q)}(k) \log L_Y(\boldsymbol{y}^{(n)} \mid \boldsymbol{\lambda}_{k,\boldsymbol{y}^{(n-1)}}^{(n)}, \rho_{CY,k}^{(n)}).$$

Cost: E step: N integrals of the D-variate normal CDF.
M step : maximising sum of N integrals of the D-variate normal CDF.

Inference for Rain Occurrence Model (HMM)

Hidden states + high spatial dimension → Pairwise-Expectation-Maximization (PEM).

$$\theta^{(q+1)} \in \arg\max_{\theta} \Big(\sum_{i,j} w_{ij}\, \mathbb{E}_{(Z \mid \mathbf{Y}_{ij}, \theta^{(q)})} \log\big(L(\mathbf{Y}_{ij}, Z \mid \theta)\big)\Big).$$

E-step : computation with classical Forward-Backward of... (for each pair $\{i,j\}$):

$$\pi_{ij,n|N}(k) = \mathbb{P}(Z^{(n)} = k \mid Y_{ij}^{(1)}, \dots, Y_{ij}^{(N)}), \qquad \pi_{ij,n,n+1|N}(k,\ell) = \mathbb{P}(Z^{(n)} = k, Z^{(n+1)} = \ell \mid Y_{ij}^{(1)}, \dots, Y_{ij}^{(N)})$$

M-step : maximisation of...

$$\mathcal{R}_c(\theta, \theta^{(q)}) = \sum_{\mathbf{i,j}} \mathbf{w_{ij}} \sum_{k=1}^{K} \pi^{(\mathbf{q})}_{\mathbf{ij,1|N}}(\mathbf{k}) \log(\eta_k) + \sum_{\mathbf{i,j}} \mathbf{w_{ij}} \sum_{k,\ell=1}^{K} \sum_{n=1}^{N-1} \pi^{(\mathbf{q})}_{\mathbf{ij,n,n+1|N}}(\mathbf{k}, \ell) \log(Q^{(n)}(k,\ell))$$
$$+ \sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{\mathbf{i,j}} \mathbf{w_{ij}} \pi^{(\mathbf{q})}_{\mathbf{ij,n|N}}(\mathbf{k}) \log \mathbf{L_{Y2}}(\mathbf{y_{ij}^{(n)}} \mid \boldsymbol{\lambda}^{(\mathbf{n})}_{\mathbf{k,ij,y^{(n-1)}}}, \rho^{(\mathbf{n})}_{\mathbf{CY,k}}),$$

Cost: E step : N integrals of the **bivariate** normal CDF **for each pair of locations**.
M step : pooling pairwise information.

Inference for Rain Occurrence Model (HMM)

Hidden states $\rightarrow$ Hybrid Expectation-Maximization (Hybrid EM).

$$\theta^{(q+1)} \in \arg\max_{\theta} \mathbb{E}_{Z|Y,\theta^{(q)}} \log L(Y, Z|\theta)$$

E-step : computation with classical Forward-Backward of... :

$\pi_{n|N}(k)=\mathbb{P}(Z^{(n)}=k|Y^{(1)},\ldots,Y^{(N)})$, $\pi_{n,n+1|N}(k,\ell)=\mathbb{P}(Z^{(n)}=k,Z^{(n+1)}=\ell|Y^{(1)},\ldots,Y^{(N)})$

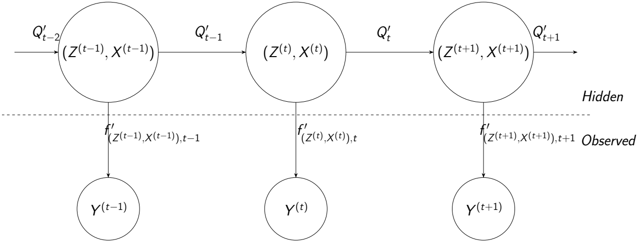
M-step : maximisation of...

$$\mathcal{R}_c'(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) = \sum_{k=1}^{K} \pi_{1|N}^{(q)}(k) \log(\eta_k) + \sum_{k,\ell=1}^{K} \sum_{n=1}^{N-1} \pi_{n,n+1|N}^{(q)}(k,\ell) \log(Q^{(n)}(k,\ell))$$
$$+ \sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{\mathbf{i,j}} \mathbf{w_{ij}} \pi_{\mathbf{n|N}}^{(\mathbf{q})}(\mathbf{k}) \log \mathbf{L_{Y2}}(\mathbf{y_{ij}^{(n)}} \mid \boldsymbol{\lambda}^{(\mathbf{n})}_{\mathbf{k,ij,y^{(n-1)}}}, \rho^{(\mathbf{n})}_{\mathbf{CY,k}})$$

Cost: E step: N integrals of the D-variate normal CDF.
M step: maximising sum of N integrals of the bivariate normal CDF. **Not guaranteed monotone, much faster in practice.**

Alternative HMM modelization

Unobserved : weather type and latent field $(Z^{(t)}, X^{(t)}) \in \{1, \dots, K\} \times \mathbb{R}^D$
Rain occurence $Y_s^{(t)} | (X_s^{(t)}, Z^{(t)}) \sim \mathit{Bernoulli}(\Phi(X_s^{(t)} + \sqrt{2}\Phi^{-1}(\lambda_{Z^{(t)},t,s})))$



+ Same interpretation for the parameters
+ Much simpler M-step : conditional independence of stations
- No longer a discrete HMM : E-step bears all the difficulty
- Need to see if there is difference in simulations before doing the work

SWG Evaluation

Goal: SWG focuses on generation, not prediction, which makes evaluation challenging.

Idea: Compare indicators from the observations and from many simulations. The model is adequate if the indicators on observed values are in the range of the indicators on the simulations.

Marginal indicators: distributional properties (histograms, moments, quantiles, etc.)

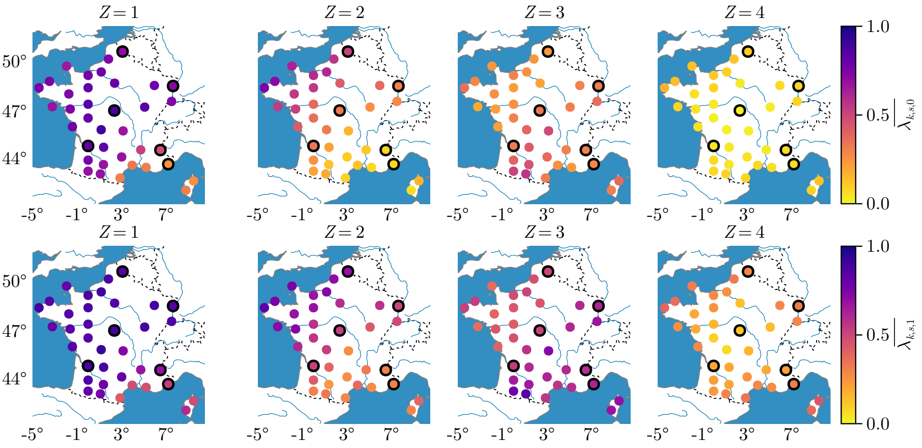
Pairwise indicators:

- Correlations : $\mathrm{Cor}(Z(s,t), Z(s',t))$, $\mathrm{Cor}(Z(s,t), Z(s,t+t'))$, $\mathrm{Cor}(Z(s,t), Z(s',t+t'))$,
- Joint exceedance : $P(X(s_i) > q_\alpha(s_i) | X(s_j) > q_\alpha(s_j))$

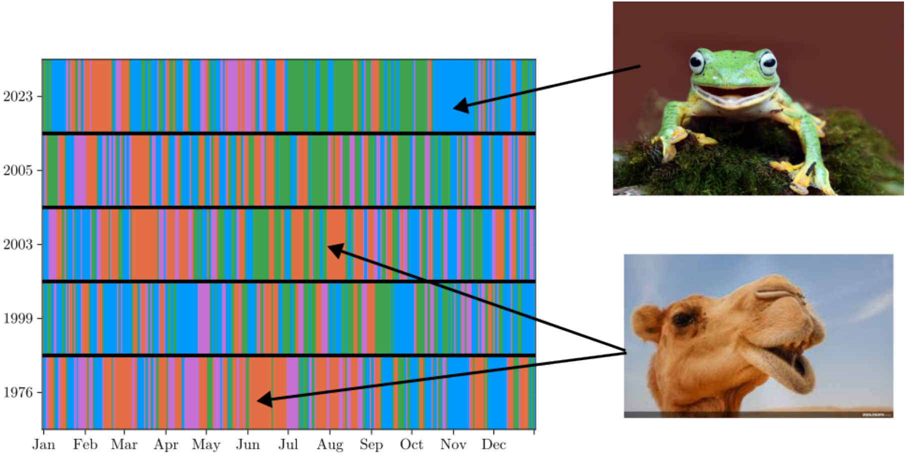
Spatial / Spatio-temporal indicators:

- $\mathrm{QER}(t)$: Fraction of exceedances over spatial domain
- $\mathrm{ROR}(t)$ (for rainfall): Fraction of nonzero precipitation

Rain occurrence - Parameters fitted from real data

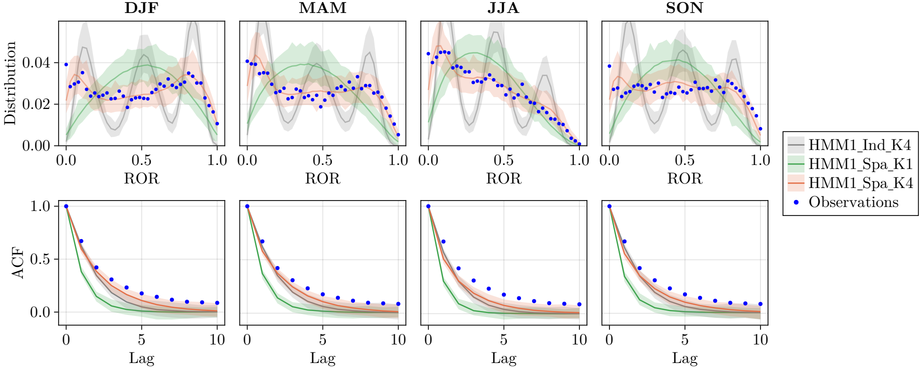


Rain occurrence - Most likely sequence of states : interpretability

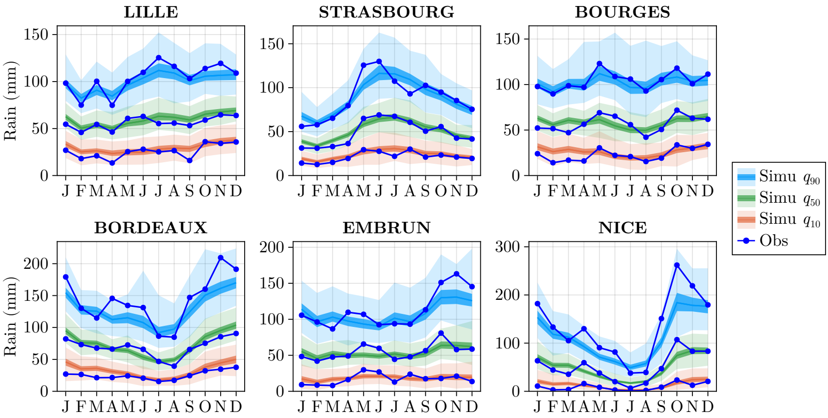


Rain occurrence - Spatiotemporal evaluation

$$\text{Rain Occurrence Rate (ROR)} = \frac{\sum_{s=1}^{D} Y_s^{(t)}}{D}$$



Full rain model - periodic marginals



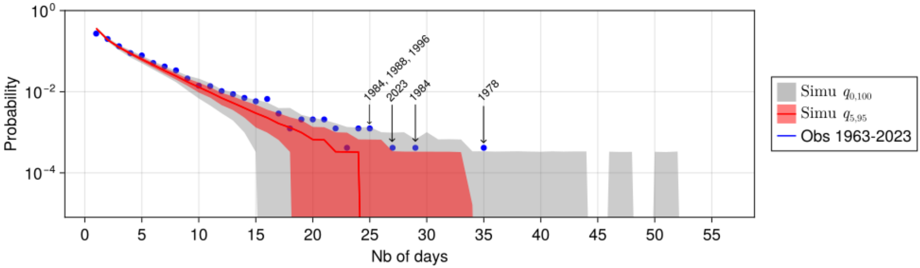
- Cycle of seasonality well-represented
- More complexity can be needed for locations like Nice

Full rain model - wet spells

"2026 had the longest experienced wet spell in France, exceeding the record previously held by 2023 and 1988"

Wet day definition (close to MeteoFrance)

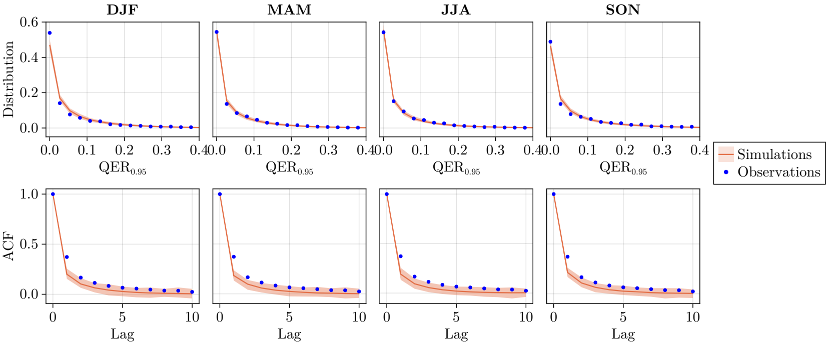
$$t \text{ a wet day, if } \frac{1}{D}\sum_{s=1}^{D} R_s^{(t)} > 1.0\mathrm{mm}$$



Full rain model - Joint extremes

Quantile Exceedance Ratio :

$$QER_\alpha(t) = \frac{\sum_{s=1}^{D} \mathbb{1}_{R_s^{(t)} > q_\alpha(s)}}{D}.$$



Conclusion

Precipitation generator

- HMM model with correlated binary emission, inference using hybrid EM
- Periodic EGPD for rain intensity
- New truncated Gaussian formalism for the dependence of the intensity
- GitHub code available
- Development of `SpatialBernoulli.jl`, contribution to `StochasticWeatherGenerators.jl`
- Preprint on HAL (`https://hal.science/hal-05523768`), submitted (under review).

Perspectives

Temperature/precipitation bivariate model:

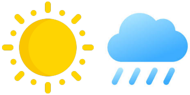
- Use bivariate covariance function for (reduced temperature, normalized precip)
- Potentially make the parameters for temperature also depend on state or rain occurrence

Improve precipitation model :

- Spatialisation : make the rain occurrence model valid in space.

Single-site weather types model

Initial idea of Richardson (1981): 2 weather types.



- The weather states/regimes/types Z_t are a Markov chain of order r :

$$P(Z_t|Z_{t-1}, \dots, Z_{t-r}) \tag{1}$$

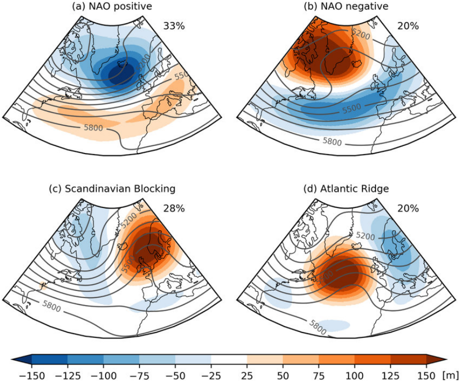
- Observed meteorological variables Y_t are generated conditionally on Z_t using **emission distributions**

$$P(Y_t|Z_t) \tag{2}$$

Separates the complexity of the model into several categories

Spatial weather types

Weather types derived from atmospheric circulation or predefined

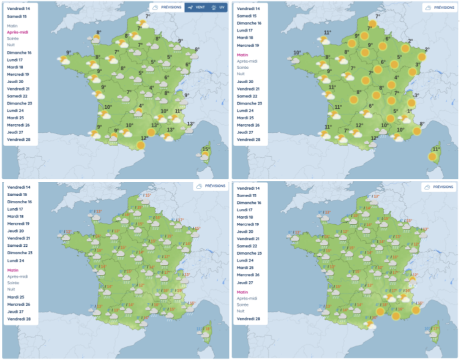


Atlantic weather types

→ Not data-driven

Hidden Markov chain

→ Weather types learned from the data



→ **This work**

Parameterization : periodic parameters

Periodic parameterisation

$$\text{For each } \theta, P_\theta(t) = c_0^\theta + \sum_{j=1}^{degP} \left(c_{2j-1}^\theta \cos(2\pi jt/T) + c_{2j}^\theta \sin(2\pi jt/T) \right)$$

$\theta^{(n)} = \psi_\theta(P_\theta(t_n)) = \psi_\theta(P_\theta(n))$ for constraints

HMM parameters :

$$Q^{(n)}(k,\ell) = \frac{\exp(P_{Q(k,l)}(n))}{1 + \sum_{l=1}^{K-1} \exp(P_{Q(k,l)}(n))} \text{ for } l < K, \qquad Q^{(n)}(k,K) = \frac{1}{1 + \sum_{l=1}^{K-1} \exp(P_{Q(k,l)}(n))},$$

$$\lambda_{k,s,y}^{(n)} = \zeta(P_{\lambda_{k,s,y}}(n)).$$

$$C_{Y,k}^{(n)}(h;\rho) = \exp(-h/\rho_{CY,k}^{(n)}), \qquad \rho_{CY,k}^{(n)} = \exp\left(P_{\rho_{CY,k}}(n)\right).$$

Model selection

2 hyperparameters : deg_P and K

Integrated Complete Likelihood

$$ICL(K, \deg_P) = L(Y, \widehat{Z}, \theta) - \frac{\log(n)}{2}|\theta|$$

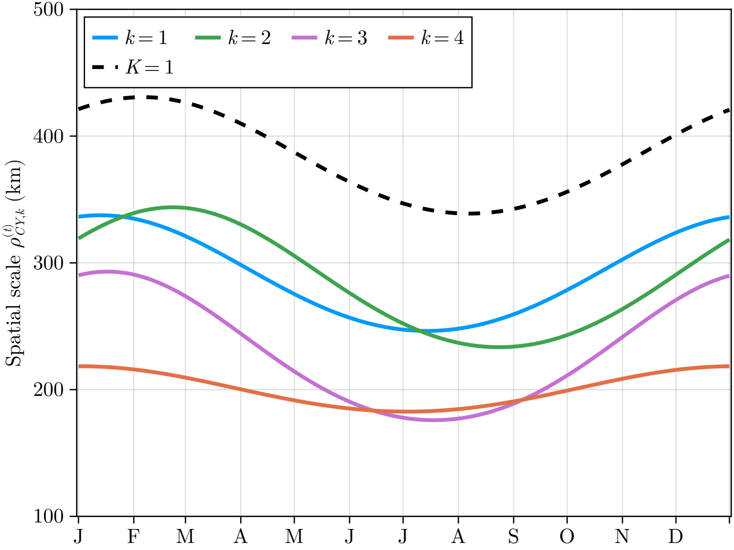
With $\widehat{Z}$ the Viterbi hidden states.

ICL value	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$
$d=0$	-3.12255e5	-2.95841e5	-2.94279e5	*-2.92816e5*	-2.929e5
$d=1$	-3.09932e5	-2.94661e5	-2.93450e5	**-2.92643e5**	-2.93197e5
$d=2$	-3.09927e5	-2.95501e5	*-2.95018e5*	-2.95164e5	-2.96406e5

Table: ICL values for several K, d, models for the rain occurrence fitted on French station data. The best value for each degree d is in italics, and the best overall value is bolded.

Selection for $m = 0$: $K = 4, degP = 1$. Same K as conditionally independent model !

Parameters fitted from real data



Parameters fitted from real data

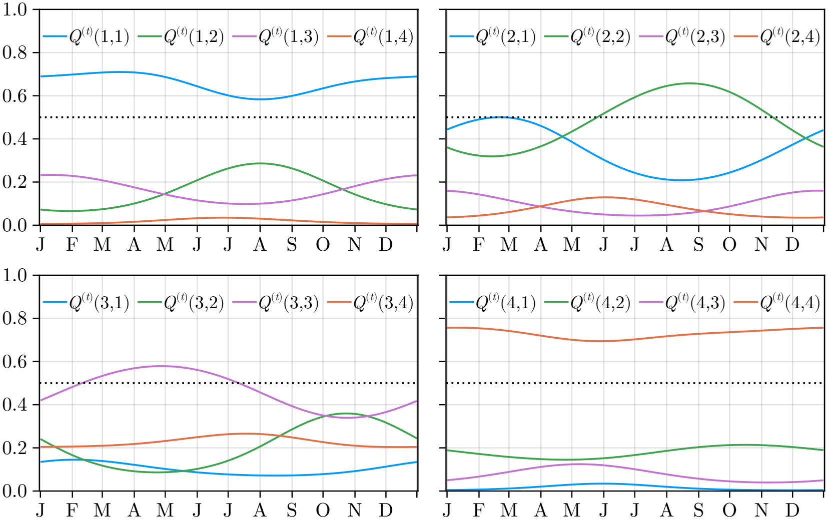


Figure: Transition parameters between states

Large scale weather types - dependent model : $D = 37$ weather stations

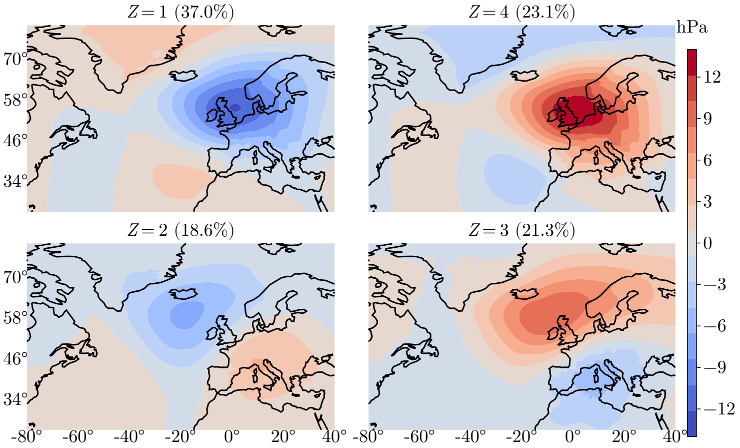


Figure: Pressure map $\Delta P = \mathbb{E}(P \mid Z = k) - \mathbb{E}(P)$ for winter months

Parameters fitted from real data

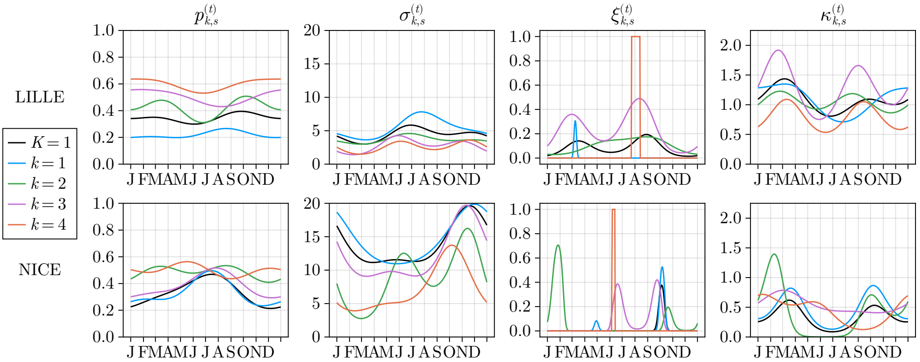


Figure: EGPD parameters

Implementation

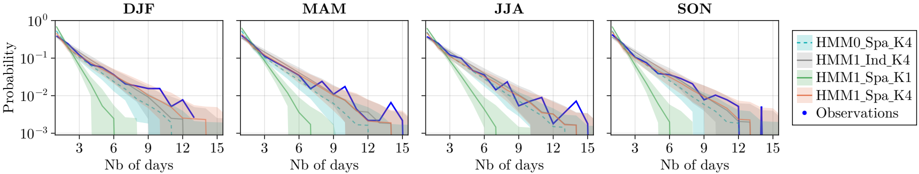
Bottleneck in HMM inference : computing multivariate normal integrals of type

$$Pr(X_1 < x_1, \ldots, X_D < x_D) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_D} f_X((x_1', ..., x_D'))d(x_1', ..., x_D')$$

- Dimension D (in E step): Quasi-Monte-Carlo, julia package MvNormalCDF. Best method, from Genz (1992).
- Other options : approximations, unfortunately bad for highly correlated model.
- Dimension 2 (in M step) : Expression from Tsay (2023)

Rain occurrence - Dry spells

Dry spell of Rain Occurrence Rate (ROR) $P_\ell = \mathbb{P}(ROR_t < 0.2, \dots, ROR_{t+\ell} < 0.2)$

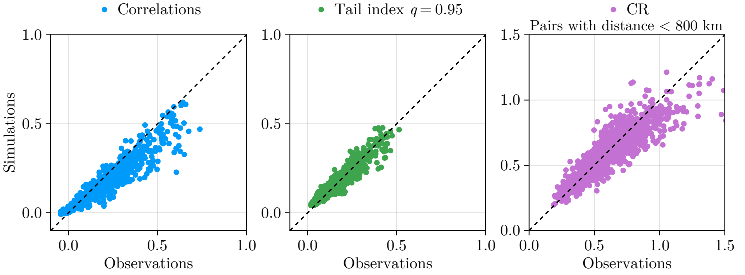


Rain generator - bivariate indicators

Spatial continuity ratio (Wilks, 1998)

For two sites s_k and s_ℓ, the **spatial continuity ratio** is

$$\mathrm{CR}(s_k, s_\ell) = \frac{\mathbb{E}\left[R_{s_\ell,t} > 0 \,\middle|\, R_{s_k,t} = 0\right]}{\mathbb{E}\left[R_{s_\ell,t} > 0 \,\middle|\, R_{s_k,t} > 0\right]}.$$

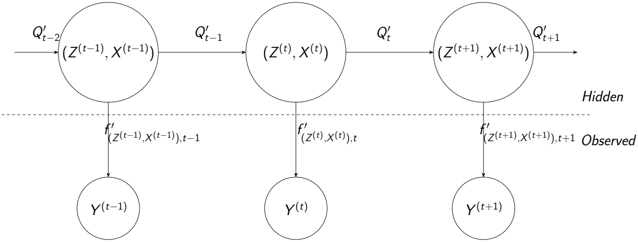


Alternative HMM modelization

Unobserved : weather type and latent field $(Z^{(t)}, X^{(t)}) \in \{1, \dots, K\} \times \mathbb{R}^D$
Rain occurence $Y_s^{(t)} | (X_s^{(t)}, Z^{(t)}) \sim Bernoulli(\Phi(X_s^{(t)} + \sqrt{2}\Phi^{-1}(\lambda_{Z^{(t)},t,s})))$

Alternative HMM modelization

Unobserved : weather type and latent field $(Z^{(t)}, X^{(t)}) \in \{1, \ldots, K\} \times \mathbb{R}^D$
Rain occurence $Y_s^{(t)} | (X_s^{(t)}, Z^{(t)}) \sim \mathit{Bernoulli}(\Phi(X_s^{(t)} + \sqrt{2}\Phi^{-1}(\lambda_{Z^{(t)},t,s})))$



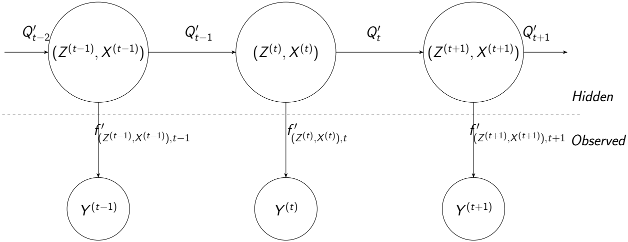
$$Q'_t((z^{(t+1)}, x^{(t+1)}), (z^{(t)}, x^{(t)})) = Q_t(z^{(t+1)}, z^{(t)}) \times f_{z^{(t+1)},t+1}(x^{(t+1)})$$

$$f'_{(z^{(t)},x^{(t)}),t}(y^{(t)}) = \prod_{s=1}^{D} \left[y_s^{(t)} \Phi\big(\xi_{t,s}\big) + (1 - y_s^{(t)})\big(1 - \Phi\big(\xi_{t,s}\big)\big) \right],$$

where $\xi_{t,s} = x_s^{(t)} + \sqrt{2}\, \Phi^{-1}(\lambda_{z^{(t)},t,s})$

Alternative HMM modelization

Unobserved : weather type and latent field $(Z^{(t)}, X^{(t)}) \in \{1, \dots, K\} \times \mathbb{R}^D$
Rain occurence $Y_s^{(t)} | (X_s^{(t)}, Z^{(t)}) \sim \mathit{Bernoulli}(\Phi(X_s^{(t)} + \sqrt{2}\Phi^{-1}(\lambda_{Z^{(t)},t,s})))$



+ Same interpretation for the parameters
+ Much simpler M-step : conditional independence of stations
- No longer a discrete HMM : E-step bears all the difficulty
- Need to see if there is difference in simulations before doing the work