MSTWeatherGen: a multivariate Stochastic Weather Generator

Saïd Obakrim, Lionel Benoit and Denis Allard

Biostatistique et processus Spatiaux (BioSP), MathNum, INRAE
Avignon, France

April 10, 2025

Non extreme climate events, but with huge impact

2016, 2024: around 25% yield loss on wheat in the "Breadbasket" (France)

Succession of adverse conditions at each stage of the wheat season: sowing, growong and harvesting

Stochastic simulation as a tool to address this complexity

INRAE département MathNum Biostatistique BIO/Π & Processus Spatiaux GEOLEARNING

SWGs

Adapted from Yiou (2024)

SWG are tools that generate random series of meteorological variables such as precipitation, temperature, wind speed, etc., with statistics similar to those of recorded data:

- Mean, variance, quantiles, skewness, extremes
- Covariance (dependence) between variables
- Temporal dependence / coherence (persistence)
- Spatial dependence / coherence

- Calibrated on recorded series
- Computational efficiency ⇒ long series and/or large number of realizations

For what purpose?

Used in impact studies

Outputs of SWGs are used as alternative, plausible weather time series in process-based models

- Assessing complex, non linear, responses to climate in agro-ecological systems
- Explore unmeasured climates
- Explore plant / ecosystem models as functions of climate variability
- Optimal decision under uncertainty: simulate up to year $t + k$, optimize decision
- Disaggregating (downscaling) meteorological variables from GCM outputs
- Also used with energy demand models, hydrological models, insurance models, ...

Some challenges that SWGs pose for spatial statistics

- Building models quantifying spatial, temporal and spatio-temporal variations ⇐ **this talk**

- Doing stochastic simulations, both for the bulk and of the tail
- Building models and methods for multivariate, spatio-temporal extreme events
- Devising new approaches for assessing return levels of impactful compound events

Our context; the project BEYOND

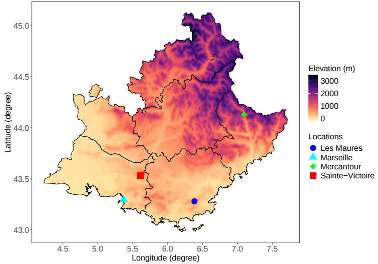
- The **BEYOND** project: towards new tools for epidemiological surveillance (for plants)
- *Xylella Fastidiosa* (Xf) is a plant pathogen propagated by insects
- Major damages: 54,000 ha of dead or uprooted olive orchards in Italy
- Seen in Corsica, Baleares, Tuscany
- Propagation depends on the whole seasonal cycle: not just one "extreme" event

↪ Need for a **stochastic tool able to generate complete cycles over a region**

INRAE département MathNum Biostatistique BIO/Π & Processus Spatiaux GEOLEARNING

Our domain of interest

- Region of interest: PACA, highly non-stationary
- 6 daily variables: precipitation, humidity, radiation, wind, min and max temperature
- SAFRAN reanalysis data from 2012 to 2021



General architecture

Two main assumptions

- Finite number of weather types over the region, $\mathbf{k} = 1, \ldots, K$
- In each weather type k, the weather variables are modeled as transformed latent Gaussian random fields

$$Y_i(\mathbf{s}, t) = \psi_{\mathbf{k},i,\mathbf{s}}(Z_{\mathbf{k},i}(\mathbf{s}, t)), \qquad \mathbf{k} = \mathbf{X(t)}$$

Weather types | Transformation function | Latent Gaussian field

- Weather states are modeled as 1st order Markov chain with transition matrices $\pi(t)$
- We use empirical transformation for $\psi_{k,i}$ with tail adjustments (Peterson and Cavanaugh, 2019)
- Multivariate spatio-temporal GPs for $\boldsymbol{Z} = (Z_i)_{i=1,p}$

Multivariate Gneiting

Proposition (Allard et al., 2022)

$$C_{ij}(\mathbf{h},u) = \frac{\tau_{ij}}{(\eta_{\mathbf{ij}}(u)+b_{ij}^2)^{\tau}} \mathcal{M}\left(\mathbf{h}; \frac{a_{ij}}{\left(\eta_{\mathbf{ij}}(\mathbf{u})+b_{ij}^2\right)^{b/2}}, \nu_{ij}\right),$$

- $[C_{ij}(\mathbf{h},u)]_{i,j=1}^{p}$ is a multivariate space-time covariance under some conditions on the $p \times p$ matrices $\boldsymbol{b}$, $\boldsymbol{a}$, $\boldsymbol{\nu}$ et $\boldsymbol{\tau}$
- $\boldsymbol{\eta}(\mathbf{u})$ is a $p \times p$ matrix-valued unbounded pseudo-variogram on $\mathbb{R}$

- Each variable has its own set of parameters in space and in time
- Illustrated later

Estimation

Use PL on cleverly chosen sets of pairs.

- Partition the spatial domain D into n_A subgroups of minimal area
- For each $\alpha \in A$, build

$$B_\alpha = \left\{ \mathbf{s}_\beta \in D \mid \mathbb{I}\left(U_\beta \leq \frac{b_\alpha}{\|\mathbf{s}_\alpha - \mathbf{s}_\beta\|^2 + 1} \right) \right\}, \quad (1)$$

with b_α such that $\mathbb{E}[|B_\alpha|] = n_B$.

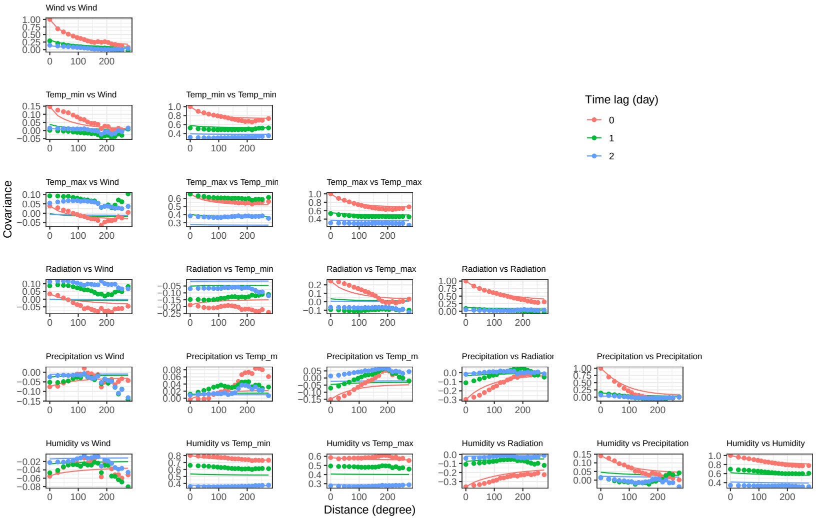
The set of pairs Λ to be used in PL is then defined as

$$\Lambda = \{(\alpha, \beta, \gamma, \delta) \ : \ \alpha \in A; \beta \in B_\alpha; |t_\gamma - t_\delta| \leq t_{max}, \} \quad (2)$$

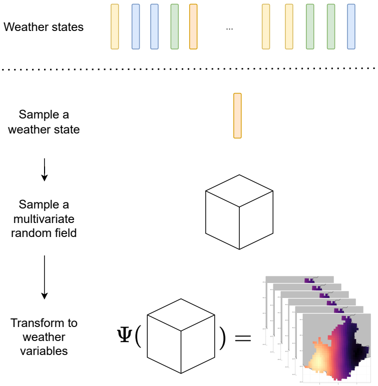
Number of pairs approximately equal to $n_S^2/(n_A n_B) \times n_T / t_{max}$

INRAE département MathNum Biostatistique BIO/Π & Processus Spatiaux GEOLEARNING

Multivariate covariance

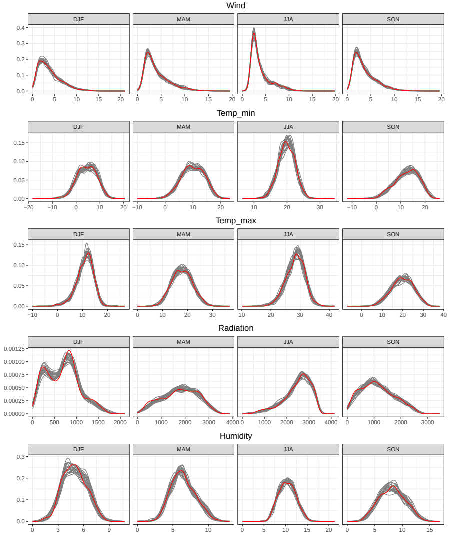


Simulation


Weather states
...
Sample a weather state
Sample a multivariate random field
Transform to weather variables
$\Psi($ $) =$

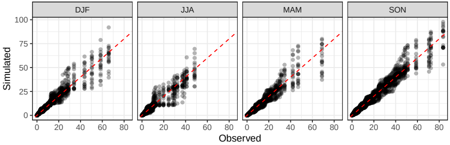
Marginals

All seasons, all continuous variables



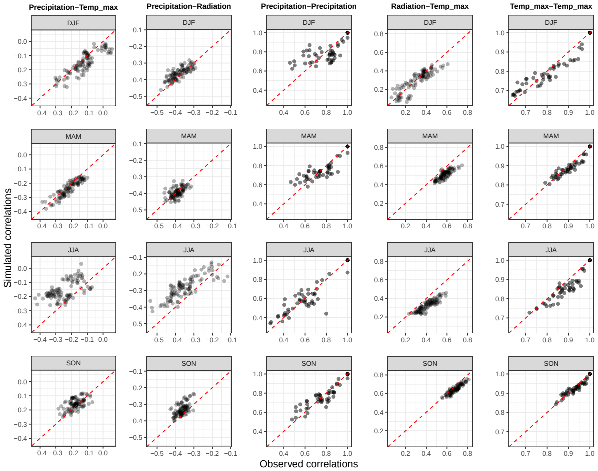
Marginals

All seasons, rainfall



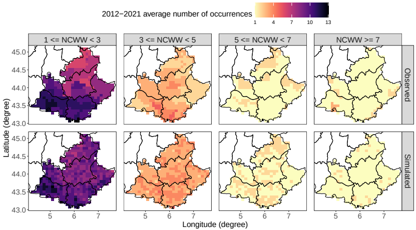
Correlations

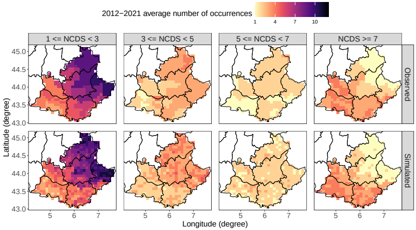
Winter, 10 random locations



Wet and dry spells

Wet winter spells, Dry summer spells

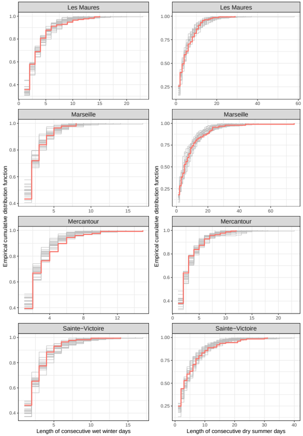




INRAE département MathNum BIO/Π & Processus Spatiaux GEOLEARNING

Empirical cumulative distribution function of spells

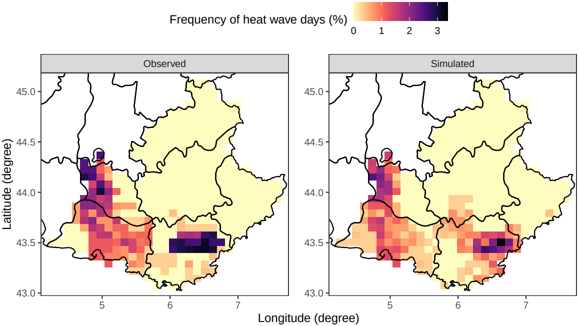
Wet winter spells, Dry summer spells



INRAE département MathNum Biostatistique BIO/Π & Processus Spatiaux GEOLEARNING

Heat Waves

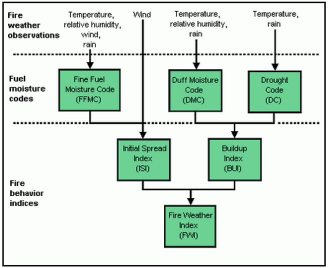
Defined as three successive days with Tmin > 21.5℃ and Tmax >34.5℃



Slight underestimation ⟹ need for better modeling of extremes and their temporal persistence

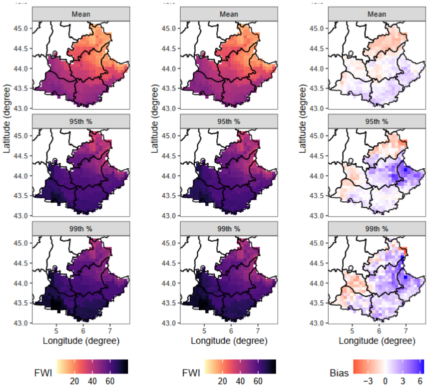
Fire Weather Index

- Developed and maintained by the Candian Forest Service
- consists of six components that account for the effects of fuel moisture and weather conditions on fire behavior.
-
- Updated daily using FWI(day before) and Temp, Rain, Wind, RH
- Used in France for the alert system



Fire Weather Index

Summer average

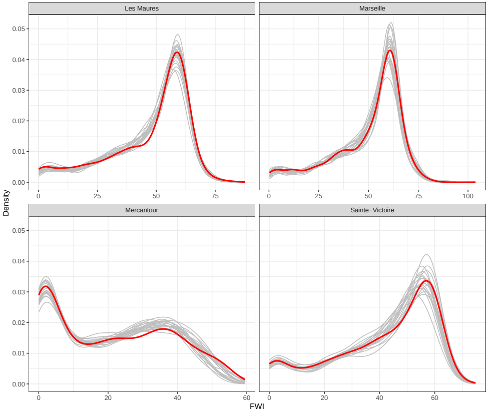


Obs.

Sim.

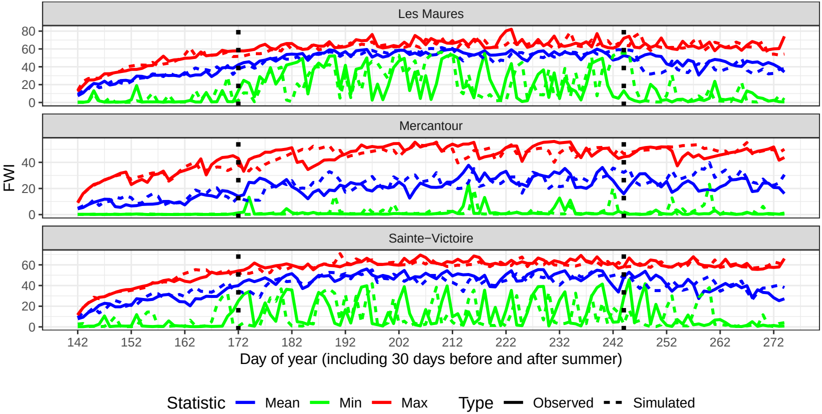
Fire Weather index

Summer density



Fire Weather Index

Summer



INRAE département MathNum BioSP GEOLEARNING

Beyond MSTWeatherGen

MSTWeatherGen

✓Obakrim S., Benoit L., Allard D. (2025) A multivariate and space-time stochastic weather generator using a latent Gaussian framework. *Stochastic Environmental Research and Risk Assessment* (In press).

✓Obakrim S, Benoit L, Allard D, Rey J (2024). MSTWeatherGen: Multivariate Space-Time Weather Generator. R package
`https://sobakrim.github.io/MSTWeatherGen/index.html`

- CDD 2 ans à BioSP financé par la chaire
- Première mission: consolider MSTWeatherGen et l'interface utilisateur
- Au-delà:
  - Intégrer les résultats obtenus par Antoine, Rita, Tiziano etc. (?)
  - Reproduire des séries ou capacité d'interpolation ?
  - Et puis ?

INRAE · département MathNum · Biostatistique BIO/Π & Processus Spatiaux · GEOLEARNING