Multivariate generator at sub-daily resolution in a semi-arid Tunisian catchment

Farhani N.\textsuperscript{1,3}, Carreau J.\textsuperscript{2}, Boulet G.\textsuperscript{3}, Kassouk Z.\textsuperscript{1} et al.

\textsuperscript{1}LR17AGR01 / GREEN-TEAM (Institut National Agronomique de Tunis / Univ. Carthage), Tunisie - farhani.nesrine@gmail.com

\textsuperscript{2}HydroSciences Montpellier (Univ. Montpellier/CNRS/IRD), France

\textsuperscript{3}Centre d’Etudes Spatiales de la Biosphère (CNRS/UPS/CNES/IRD), France
gestion des ressources en eau dans les milieux ruraux tunisiens

water resources management in Tunisian rural areas

Bilateral research laboratory Tunisia / France

Co-directed: Insaf MEKKI, INRGREF Tunis & Frédéric JACOB, IRD / LISAH Montpellier

Co-funded: MESRS / IRESA / IRD
Study areas

Two watersheds on the Tunisian mountain ridge

**Lebna**
- 200 km²
- sub-humid climate

**Merguellil**
- 1200 km² / 400 km²
- semi-arid climate
Evapotranspiration: key variable to monitor

1. actual water use
2. agricultural drought (hydric stress on plants)

Estimation:

- double source model based on energy balance equations
- relies on remote sensing data and observations
Forcing scenarios

Hydro-meteorological variables measured at sub-daily resolution:

- Tair, Hr, Rg, Wave, Pa & Precip

**Goal**: to provide spatio-temporal scenarios to extend observations in space and in time

**Main issues**: strong spatio-temporal variability, non-gaussianity, scarcely gauged networks

**Project**: Approches stochastiques et seMi-pAramétriques combinées à la télédétection pour l’étude du stress hyDriquE
Merguellil watershed

semi-arid climate: 200 - 650 mm / year

30 min temporal resolution
Three stations in the plains: geographical and meteorological proximity
1st step

- **Imputation**: to fill in missing values at each of the 3 stations
- **Virtual station**: to create a single representative station in the plains

**Stochastic GLM**: adaptation of RGlimClim from R. Chandler
1\textsuperscript{st} step

**Imputation**: to fill in missing values at each of the 3 stations

**Virtual station**: to create a single representative station in the plains

**Stochastic GLM**: adaptation of RGlimClim from R. Chandler

2\textsuperscript{nd} step

**Temporal extension** based on bias correction applied to Era-Interim’s reanalyses

**Spatial extension** based on analogs applied to WRF’s simulations

Could be compared with stochastic GLM approach
Stochastic GLM

- Inter-variable dependency
- GLM families / distributions
- Temporal and spatial effects - covariates
- Model selection
- Model validation
- Spatial dependence
Inter-variable dependency

Product rule: \( Y_i \) hydro-meteorological variable + \textit{transformation}

\[
\mathbb{P}(Y_1, Y_2, \ldots, Y_p) = \mathbb{P}(Y_1) \prod_{i=2}^{p} \mathbb{P}(Y_i | Y_{i-1}, \ldots, Y_1)
\]
GLM families

Normal distribution

\[ Y \mid X = x \sim \mathcal{N}(\beta x, \sigma^2) \iff Y = \beta X + \epsilon \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \]
With a transformation $Y = \tan(\pi \tilde{Y} - 0.5)$
Heteroscedastic Normal distribution

\[ Y | X = x \sim \mathcal{N}(\beta x, \sigma(x)^2) \quad \sigma(x) = \exp(\eta x) \]
Gamma distribution

\[ Y | X = x \sim Gamma(\mu(x), \nu) \quad \mu(x) = \exp(\eta x) \]

\[ Y = \max(\log(\tilde{Y})) - \log(\tilde{Y}) \]
Temporal effects

Seasonal effects

Covariates:
\[
\cos \left( \frac{2\pi d}{k} \right), \quad \sin \left( \frac{2\pi d}{k} \right)
\]
with \(d\) the day of the year and \(k \in \{183, 365.25, \ldots \}\) the period

Diurnal effects

Covariates:
\[
\cos \left( \frac{2\pi h}{k} \right), \quad \sin \left( \frac{2\pi h}{k} \right)
\]
with \(h\) the hour of the day and \(k \in \{12, 24, \ldots \}\) the period
Spatial effects

Covariates:

- geographical coordinates
- landscape variables
Lags of

- the variable itself $Y_{t-k}$
- a moving average

\[
\frac{1}{m} \sum_{i=1}^{m} Y_{t-k-i+1}
\]

- a spatial average

\[
\frac{1}{s} \sum_{j=1}^{s} Y^{(j)}_{t-k}
\]

- a moving spatial average
Model selection

With AIC, BIC and p-values

Number of selected memory effects

<table>
<thead>
<tr>
<th></th>
<th>itself</th>
<th>MA</th>
<th>SA</th>
<th>MSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pa</td>
<td>30</td>
<td>15</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>U</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>V</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Tair</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Hr</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Rg</td>
<td>12</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
Model validation

Comparison of unconditional simulations from fitted models and observations

- general: temporal plots
- intensity: qq-plots
- dependence: pairwise scatter plots and Kendall’s $\tau$
General assessment

Air temperature at Chebika station

![Air temperature graph](image)
Multivariate generator at sub-daily resolution in a semi-arid Tunisian catchment
Intensity assessment

Chebika Station

[Graphs showing intensity assessment at the Chebika station]
Inter-sites dependence

Farhani N., Carreau J., Boulet G., Kassou Multivariate generator at sub-daily reso.
Farhani N., Carreau J., Boulet G., Kassouk Z. et al. (LR17AGR01 / GREEN-TEAM (Institut National Agronomique de Tunis / Univ. ... Montpellier (Univ. Montpellier/CNRS/IRD), France, Centre d'Etudes Spatiales de la Biosphère (CNRS/UPS/CNES/IRD), France)

Multivariate generator at sub-daily resolution in a semi-arid Tunisian catchment
Multivariate generator at sub-daily resolution in a semi-arid Tunisian catchment
Conditional simulation: imputation

Farhani N., Carreau J., Boulet G., Kassou

Multivariate generator at sub-daily resolution in a semi-arid Tunisian catchment
Farhani N., Carreau J., Boulet G., Kassouk Z. et al. (LR17AGR01 / GREEN-TEAM (Institut National Agronomique de Tunis / Univ. Montpellier (Univ. Montpellier/CNRS/IRD), France, Centre d'Etudes Spatiales de la Biosphère (CNRS/UPS/CNES/IRD), France)

Multivariate generator at sub-daily resolution in a semi-arid Tunisian catchment
A careful look at model selection, in particular for memory effects

More validation is needed regarding
- inter-variable dependencies
- temporal dependencies
Summary and outlook

A careful look at model selection, in particular for memory effects

More validation is needed regarding

- inter-variable dependencies
- temporal dependencies

→ Need to introduce spatial dependencies
Summary and outlook

A careful look at model selection, in particular for memory effects

More validation is needed regarding

- inter-variable dependencies
- temporal dependencies

→ Need to introduce spatial dependencies

→ Spatial interpolation thanks to geographical covariates
Summary and outlook

A careful look at model selection, in particular for memory effects

More validation is needed regarding

- inter-variable dependencies
- temporal dependencies

→ Need to introduce spatial dependencies

→ Spatial interpolation thanks to geographical covariates

→ Validation in terms of estimation of evapotranspiration and stress index
ABC for spatial max-stable model with potential asymmetry and anisotropy

Carreau J.¹, Toulemonde G.²

¹HydroSciences Montpellier (Univ. Montpellier/CNRS/IRD), France - julie.carreau@ird.fr

²Institut Montpelliérain Alexander Grothendieck, (Univ. Montpellier/CNRS) & INRIA / LEMON, France - gwladys.toulemonde@umontpellier.fr
Our goal

Model the joint distribution of annual maxima at $d$ sites $Y_1, \ldots, Y_d$
Exploratory analysis of annual maxima

Anisotropy in the strength of the dependence

Correlograms of Kendall’s $\tau$ for classes of distances $\Rightarrow$ along two directions 0 and $\pi/2$

$$R(h; \beta, \alpha) = 2 \exp \left\{ - \left( \frac{h}{\beta} \right)^{\alpha} \right\} - 1$$
Asymmetry in the dependence

For a pair of stations: proportion of values in $[0, 1]^2$ under the diagonal

Asymmogram: Proportions for classes of distances

along two directions 0 and $\pi/2$
Extra-parametrized Gumbel (XGumbel)

The Gumbel copula in $d$-dimension and parameter $\beta \geq 1$ is given as:

$$C_\beta(u) = \exp \left\{ - \left[ \sum_{i=1}^{d} (-\ln u_i)^\beta \right]^{1/\beta} \right\}$$

Let $a = (a_1, \ldots, a_d) \in I^d$ be extra-parameters.

The extra-parametrized Gumbel copula is defined as, with $\psi = (a, \beta_A, \beta_B)$, $\beta_A \leq \beta_B$:

$$C_\psi(u) = C_{\beta_A}(u^a) C_{\beta_B}(u^{1-a})$$

Equivalently, let $U \sim C_{\beta_A}$ and $V \sim C_{\beta_B}$. Then

$$Z = \max(U^{1/a}, V^{1/(1-a)}) \sim C_\psi$$
Extra-parameter mapping

Max-stable spatial process defined by its finite-dimensional distributions

\( \forall (s_1, \ldots, s_d) \) with \( s_i \in S \), let \( x_{s_i} \in \mathbb{R}^p \) be covariates

Define extra-parameters thanks to a mapping with parameters \( \theta \):

\[ a_{s_i} = a(x_{s_i}; \theta) \]

With \( \psi = (a(x_{s_1}; \theta), \ldots, a(x_{s_d}; \theta), \beta_A, \beta_B) \), let

\[ (Z_{s_1}, \ldots, Z_{s_d}) \sim C_\psi \]

Since \( C_\beta(\cdot) \) is max-stable \( \Rightarrow C_\psi(\cdot) \) is max-stable as well
Strength of dependence

\[ \chi(s_1, s_2) = \lim_{u \uparrow 1} \mathbb{P}(Z_{s_2} > u | Z_{s_1} > u) \]

\[ = 2 - [(a_1^{\beta_A} + a_2^{\beta_A})^{1/\beta_A} + ((1 - a_1)^{\beta_B} + (1 - a_2)^{\beta_B})^{1/\beta_B}] \]

\[ a_1 = a_2 = 1 \rightarrow C_\psi(u_1, u_2) = C_{\beta_A}(u_1, u_2) \]

\[ a_1 = a_2 = 0 \rightarrow C_\psi(u_1, u_2) = C_{\beta_B}(u_1, u_2) \]

How to define \( a(\cdot; \theta) \) so that

- dependence strength decreases with distance?
- potential anisotropy?
→ Need of a stochastic mapping changing with time
Asymmetry
Spherical mapping whose center changes each year $t$

\[ a(x; c_t, \sigma) = 1 - \exp \left\{ -\frac{||x - c_t||^2}{2\sigma^2} \right\} \quad c_t \in \mathbb{R}^p, \sigma \in \mathbb{R} \]

The centers $c_t$ are sampled from

\[ \sum_{j=1}^{m} \pi_j \mathcal{N}(\mu_j, \sigma_j I_p) \]

Then $\theta = (\pi_1, \ldots, \pi_m, \mu_1, \ldots, \mu_m, \sigma_1, \ldots, \sigma_m)$
Bivariate density

\[ c_{a_1a_2}(u_1, u_2) = a_1a_2u_1^{a_1-1}u_2^{a_2-1}c_{\beta_B}(u_1^{1-a_1}, u_2^{1-a_2})c_{\beta_A}(u_1^{a_1}, u_2^{a_2}) + (1 - a_1)(1 - a_2)u_1^{-a_1}u_2^{-a_2}c_{\beta_A}(u_1^{a_1}, u_2^{a_2})c_{\beta_B}(u_1^{1-a_1}, u_2^{1-a_2}) - a_1(1 - a_2)u_1^{a_1-1}u_2^{-a_2}\frac{\partial c_{\beta_A}(u_1^{a_1}, u_2^{a_2})}{\partial u_1^{a_1}}\frac{\partial c_{\beta_B}(u_1^{1-a_1}, u_2^{1-a_2})}{\partial u_2^{1-a_2}} + (1 - a_1)a_2u_1^{-a_1}u_2^{a_2-1}\frac{\partial c_{\beta_A}(u_1^{a_1}, u_2^{a_2})}{\partial u_2^{a_2}}\frac{\partial c_{\beta_B}(u_1^{1-a_1}, u_2^{1-a_2})}{\partial u_1^{1-a_1}} \]

→ Pairwise log-likelihood works fine for standard XGumbel
Not possible for latent version
Approximate Bayesian Computing

Let $\mathbf{Z} = (Z_{s_1}, \ldots, Z_{s_d})$ the observed sample
Let $S(\mathbf{Z})$ be a set of statistics

Rejection ABC

1. Draw $(\psi_i, \theta_i)$ from prior
2. Simulate $\mathbf{Z}^{(i)} = (Z_{s_1}^{(i)}, \ldots, Z_{s_d}^{(i)})$ from latent XGumbel with parameters $(\psi_i, \theta_i)$
3. Reject $(\psi_i, \theta_i)$ if $d(S(\mathbf{Z}^{(i)}), S(\mathbf{Z})) > \epsilon$
\[ S(Z) = R(h; \hat{\beta}, \hat{\alpha}) \quad S(Z^{(i)}) = R(h; \hat{\beta}_i, \hat{\alpha}_i) \]

→ the entire \( \rho \)-correlogram curves fitted to the observations / simulations

Define

\[
d(S(Z^{(i)}), S(Z)) = \int |R(h; \hat{\beta}, \hat{\alpha}) - R(h; \hat{\beta}_i, \hat{\alpha}_i)|dh
\]

Anisotropy : distances between directional \( \rho \)-correlogram curves

Erhardt, R.J. & Smith, R.L. (2012) *Approximate Bayesian computing for spatial extremes*, CSDA, 56(6), 1468-1481
First results

Omni-directional $\rho$ correlogram

[Graph showing the correlogram with data points and lines representing different scenarios. The graph has a y-axis labeled $\rho$ ranging from 0.0 to 0.4 and an x-axis labeled $h$ ranging from 0 to 80.]
Bi-directional $\rho$ correlogram
Summary and outlook

Spatial latent XGumbel provides a flexible model for spatial maxima with potential asymmetry and anisotropy

Need to better understand how to design the mapping of the extra-parameters

ABC seems promising to fit models for spatial extremes

Need to assess if asymmetry can be reproduced as a summary statistics

→ Perform a general validation of the spatial latent XGumbel model

Carreau J., Toulemonde G. (HydroSciences Montpellier (Univ. Montpellier/CNRS/IRD), France - julie.carreau@ird.fr, Institut Montpelliérain Alexander Grothendiek, (Univ. Montpellier/CNRS) & INRIA / LEMON, France - gwladys.toulemonde@umontpellier.fr)