Conditional simulations and downscaling of climatic max-stable processes

Aurélien Bechler

supervised by Liliane Bel (AgroParisTech) and Mathieu Vrac (LSCE)

SWGEN, Avignon

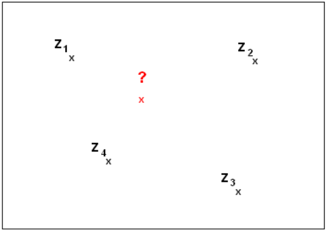
The work is supported by ANR McSim

September 2014

Objectives

Step 1: conditional simulations

Distribution of maxima of precipitation at some new locations given some observations at other locations.



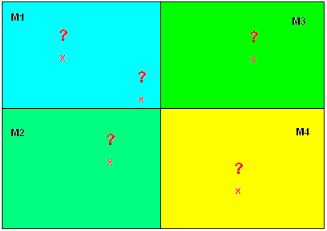
Objectives

Step 1: conditional simulations

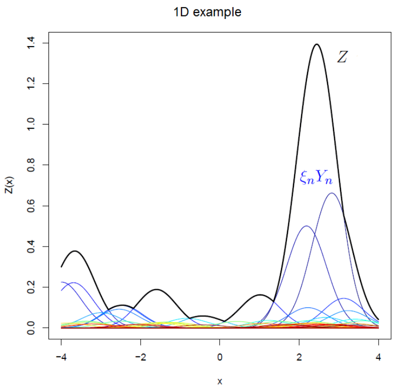
Distribution of maxima of precipitation at some new locations given some observations at other locations.

Step 2: downscaling

Distribution of maxima of precipitation at some new locations given grid-cells information.



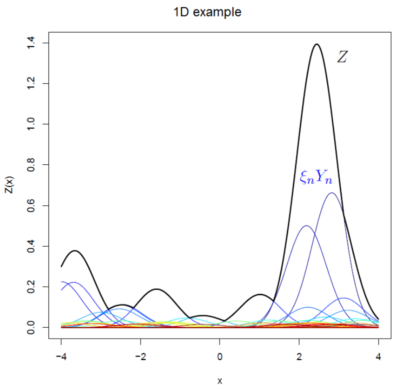
Max-stable processes (De Haan, 1984)



$$\text{Model}: \; Z(x) = \max_{n \geq 1} \{\xi_n Y_n(x)\}$$

where Y_n i.i.d continuous random fields and ξ_n a Poisson process on $\mathbb{R}^+$ with intensity $\alpha t^{-(\alpha+1)} dt$.

Max-stable processes (De Haan, 1984)



Model : $Z(x) = \max_{n \geq 1} \{\xi_n Y_n(x)\}$

where Y_n i.i.d continuous random fields and ξ_n a Poisson process on $\mathbb{R}^+$ with intensity $\alpha t^{-(\alpha+1)} dt$.

- Brown-Resnik processes: Y_n log-Gaussian intrinsic random fields, $\alpha = 1$.
- Schlather processes: Y_n stationary Gaussian fields, $\alpha = 1$.
- Extremal t processes: Y_n stationary Gaussian fields, $\alpha \geq 0$.

Step 1: Conditional simulation

Objective: Simulate $Z(x_0)|Z(x_1) = z_1, \ldots, Z(x_n) = z_n$.

Step 1: Conditional simulation

Objective: Simulate $Z(x_0)|Z(x_1) = z_1, \ldots, Z(x_n) = z_n$.

Dombry et al. (2013) proposed a conditional simulation algorithm for some max-stable processes (Brown-Resnick and Schlather processes).

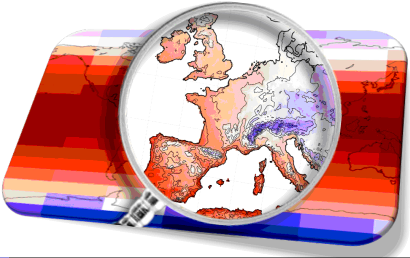
Our contribution:

- Generalization of this algorithm to the extremal t processes family.
- Improvement of the algorithm to allow higher number of conditioning points.

Step 2: downscaling

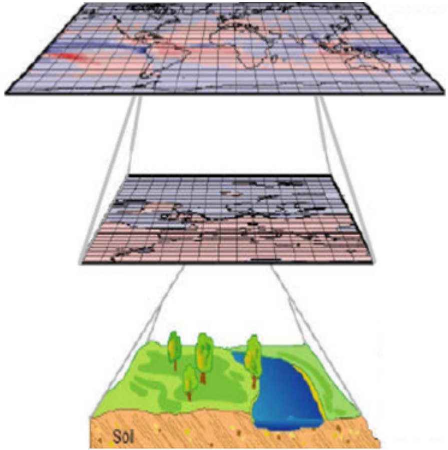
Available tools: climate numerical models at a global or regional scale.

Problems : -Spatial resolution too low to provide good description of the process of interest.
-Poor characterization of extreme behavior.



Step 2: downscaling

Downscaling is the action of generating climatic or meteorological values at a local scale based on information given at a large scale.



Step 2: downscaling

Solution: Downscaling of the regional climate models (RCM) outputs

$$\text{Climate model : cells } M_1, \dots, M_n$$

Objective: Simulate $Z(x_0)|M_1 = m_1, \dots, M_n = m_n$

Step 2: downscaling

Solution: Downscaling of the regional climate models (RCM) outputs

$$\text{Climate model : cells } M_1, \ldots, M_n$$

Objective: Simulate $Z(x_0)|M_1 = m_1, \ldots, M_n = m_n$

Method: use downscaled values as conditioning points.

$\Rightarrow$ Spatial Hybrid Downscaling method (SHD)

SHD : a physical and statistical approach

General Methodology:

- Choose a **given number of points** in the local scale dataset.
- Establish a statistical link (**transfer function**) between the large scale information (from RCM outputs) and these points (calibrated on the past)

SHD : a physical and statistical approach

General Methodology:

- Choose a **given number of points** in the local scale dataset.
- Establish a statistical link (**transfer function**) between the large scale information (from RCM outputs) and these points (calibrated on the past)
- Build with this transfer function some pseudo-observations at these given locations when there is no longer pointwise information available.

SHD : a physical and statistical approach

General Methodology:

- Choose a **given number of points** in the local scale dataset.
- Establish a statistical link (**transfer function**) between the large scale information (from RCM outputs) and these points (calibrated on the past)
- Build with this transfer function some pseudo-observations at these given locations when there is no longer pointwise information available.
- Perform the **conditional simulation algorithm** of max-stable processes with the pseudo-observations as conditioning values.

Local scale: *SAFRAN* data (Quintana-Segui et al., 2008)

Goal: Perform a large number of conditional simulations with the SHD method and check their quality with the test points.

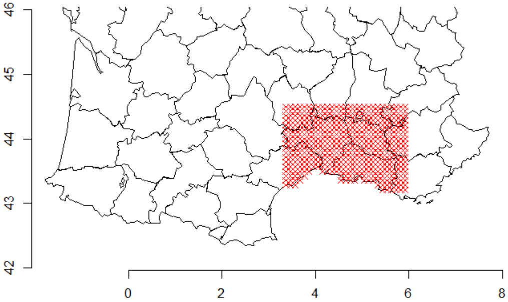


Figure: Study area of *SAFRAN* data subset.

Description:

- Autumnal maximum of daily precipitation for 1960-2007.
- 457 points uniformly distributed.
- Grid data from interpolation.

Large scale: *MEDCORDEX-IPSL-WRF* data

Goal: Perform a large number of conditional simulations with the SHD method and check their quality with the test points.

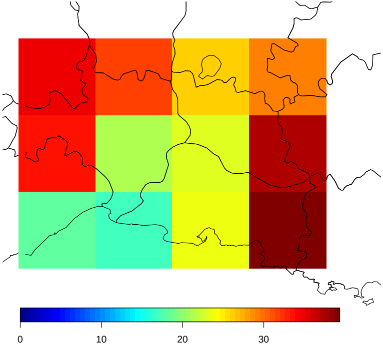


Figure: *MEDCORDEX-IPSL-WRF* maximum precipitation data for the year 1989 (in mm).

Description:

- RCM outputs of autumnal daily precipitation for 1989-2007.
- Grid resolution: about 50km.
- Area of interest is covered by 12 grid-cells.

Transfer functions and methods

Methods	TF	Cond. Sim.	Description
Interpolation	NO	NO	Bilinear interpolation of the RCM outputs values.
Linear downsc.	YES	NO	Linear regression between the local-scale variable and the large-scale variable
Raw	NO	YES	Use directly RCM outputs as conditioning values.
CDF-t	YES	YES	Build a bias correction by comparing the CDF of the large-scale variable and the one of the local-scale variable.
Linear reg.	YES	YES	Same as *Linear downscaling* but only at the conditioning points.
Optimal	NO	YES	Use directly the real observations as conditioning values

Table: Different methods for building the pseudo-observations from the *Medcordex* and *Safran* datasets.

Results

Skill-Score: % of improvement compared to a reference method.

	Models	CRPSS	QSS_{95}	K-S SS	$RMSE_v$ SS
No simulations	Interpolation	-20.1%	-54.8%	-19.8%	-0.4%
	Linear downscaling	0%	0%	0%	0%
Conditional Simulations	Raw	14.0%	44.4%	10.4%	20.2%
	CDF-t	14.1%	**60.6%**	12.3%	**40.4%**
	Linear Regression	**20.1%**	53.4%	**15.0%**	22.5%
	Optimal	24.2%	70.7%	18.7%	52.9%

Table: Skill-Scores (with *Linear downscaling* as reference) of the hybrid algorithm with the different methods for building the pseudo-obs. from the MEDCORDEX dataset.

Results

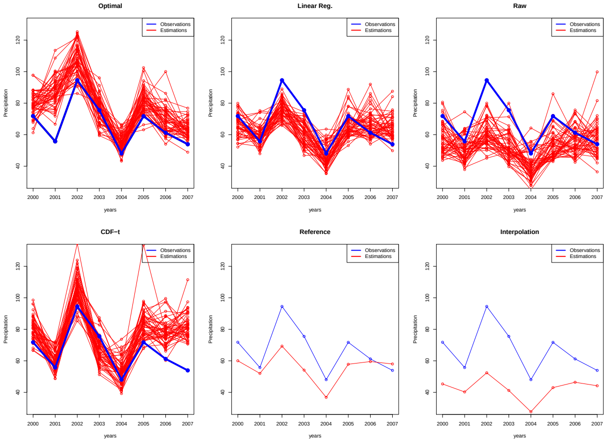


Figure: Annual means obtained by the 6 methods. The observed in blue and each trajectory in red is one conditional simulation.

Conclusions and perspectives

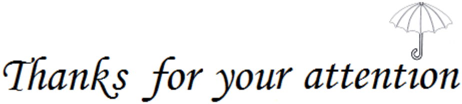
Conclusions:

- Conditional simulation of extremal t processes.
- Use of statistical downscaling and conditional simulations to get distribution of extreme precipitation even when no observation is available.

Perspectives:

- Test on other RCM.
- Choice of the conditioning points in the SHD method (PAM, PAM into a grid-cell,...).
- Describe the future evolution of the extreme precipitation in function of different large-scale scenarios.

Thanks



Brown, B. and Resnick, S. I., 1977, *Extreme Values of Independent Stochastic Processes*, Journal of Applied Probability, Vol. 14 (4), 732-739.

De Haan, L., 1984, *A spectral representation for max-stable processes*, The Annals of Probability, 12 (4), 1194-1204.

Dombry C. and Eyi-Minko F., 2013, *Regular conditional distributions of max infinitely divisible processes*, Electron. J. Probab 18 (7), 1-21.

Dombry C., Ribatet M. and Eyi-Minko F., 2013, *Conditional simulation of max-stable processes*, Biometrika, Vol. 100 (1), 111-124.

Friederichs, P. and Thorarinsdottir, T. L., 2012, *Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction*, Environmetrics, Vol. 23 (7), 579-594.

- Geweke, J., 1991, *Efficient Simulation from the Multivariate Normal and Student-t Distribution Subject to Linear Constraints and the Evaluation of Constraint Probabilities*, Computer Science and Statistics Conference.
- Opitz, T., 2013, *Extremal t processes: Elliptical domain of attraction and a spectral representation*, Journal of Multivariate Analysis, Vol. 122, 409-413.
- Padoan, S. A., 2011, *Multivariate Extreme Models Based on Underlying Skew-t and Skew-Normal Distributions*, Journal of Multivariate Analysis, Vol. 102 (5), 977-991.
- Quintana-Segui, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchistéguy, L., and Morel, S., *Analysis of near surface atmospheric variables: validation of the SAFRAN analysis over France*, 47, 92-107, 2008.

Schlather, M., 2002, *Models for Stationary Max-Stable Random Fields*, Extremes, Vol. 5 (1), 33-44.

Results

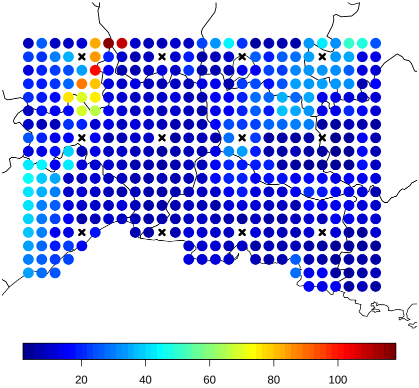


Figure: CRPS (in mm) map for the year 2000.

Comparison scores

Continuous ranking probability score (CRPS):

$$\mathrm{CRPS}(F; y) = \int_{-\infty}^{+\infty} \left[F(t) - H(t-y)\right]^2 dt,$$

with F the empirical cumulative function of the conditional simulations, y the test value and H the Heaviside function.

Continuous ranking probability score (CRPS):

$$\mathrm{CRPS}(F;y) = \int_{-\infty}^{+\infty} \left[F(t) - H(t-y)\right]^2 dt,$$

with F the empirical cumulative function of the conditional simulations, y the test value and H the Heaviside function.

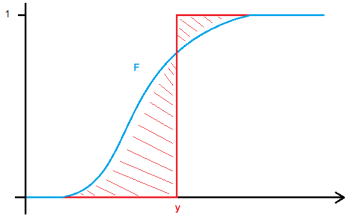


Figure: Illustration of the crps.

Comparison scores

Continuous ranking probability score (CRPS):

$$\mathrm{CRPS}(F;y) = \int_{-\infty}^{+\infty} \left[F(t) - H(t-y)\right]^2 dt,$$

with F the empirical cumulative function of the conditional simulations, y the test value and H the Heaviside function.

The quantile score (QS):

$$\mathrm{QS}_p(F;y) = \nu_p(y - F^{-1}(p))$$

where p is the probability of interest and $\nu_p(u) = pu$ if $u \geq 0$ and $\nu_p(u) = (p-1)u$ otherwise.

Description de l'algorithme

Etape 1 : tirer τ une partition aléatoire de $\{x_1, \ldots, x_k\}$ qui a pour distribution

$$\mathbb{P}[\theta = \tau \mid Z(x) = z] \propto \prod_{j=1}^{|\tau|} \lambda^{\alpha}_{x_{\tau_j}}(z_{\tau_j}) \int_{\{\mathrm{u}_j < z_{\tau_j^c}\}} \lambda^{\alpha}_{x^c_{\tau_j} | x_{\tau_j}, z_{\tau_j}}(\mathrm{u}_j) \mathrm{d}\mathrm{u}_j.$$

Description de l'algorithme

Etape 1 : tirer τ une partition aléatoire de $\{x_1, \ldots, x_k\}$

Etape 2 : sachant $\tau = (\tau_1, \ldots, \tau_l)$, tirer l fonctions extrêmes $\phi_1^+, \ldots, \phi_l^+$ de distribution

$$\mathbb{P}[\phi_j^+(s) \in \mathrm{dv} \mid z, \tau] \propto \left\{ \int 1_{\{\mathrm{u} < z_{\tau_j^c}\}} \lambda^{\alpha}_{(s, x_{\tau_j}^c) | x_{\tau_j}, z_{\tau_j}}(\mathrm{v},\mathrm{u}) \mathrm{du} \right\} \mathrm{dv}.$$

avec s $\in (\mathbb{R}^2)^m$ et $m \in \mathbb{N}$, le nombre de nouveaux points où l'on veut simuler.

Description de l'algorithme

Etape 1 : tirer τ une partition aléatoire de $\{x_1, \ldots, x_k\}$

Etape 2 : sachant $\tau = (\tau_1, \ldots, \tau_l)$, tirer l fonctions extrêmes $\phi_1^+, \ldots, \phi_l^+$

Etape 3 : indépendamment des étapes 1 et 2, simuler $\{\phi_i(\mathbf{x})\}_{i \geq 1}$ sous la contrainte $Z^-(\mathbf{x}) \leq \mathbf{z}$ avec

$$Z^- = \max_{i \geq 1} \{\phi_i^\alpha(x) | \ \phi_i^\alpha(x) \leq z\}$$

Annexe

$$\lambda^{\alpha}_{\mathbf{s}|\mathbf{x}}(u) = \frac{w_{k+\alpha-2}}{w_{k+m+\alpha-2}} \sqrt{|\Sigma_{\mathbf{sx}} \Sigma_{\mathbf{x}}^{-1}|} \frac{[z' \Sigma_{\mathbf{x}}^{-1} z]^{(k+\alpha)/2}}{[(u,z)' \ \Sigma_{\mathbf{s},\mathbf{x}}^{-1} \ (u,z)]^{(k+m+\alpha)/2}}, \quad (1)$$

avec

$$\begin{aligned} k &= \text{nombre de points conditionnants,} \\ m &= \text{nombre de nouveaux points.} \end{aligned}$$

Annexe

$$\begin{aligned}
\mathbb{P}[\theta = \tau \mid Z(\mathbf{x}) = \mathbf{z}] &= \frac{1}{C(\mathbf{x},\mathbf{z})} \prod_{j=1}^{|\tau|} \lambda_{\mathbf{x}_{\tau_j}}(\mathbf{z}_{\tau_j}) \times \\
&\quad \int_{\{\mathbf{u}_j < \mathbf{z}_{\tau_j^c}\}} \lambda_{\mathbf{x}_{\tau_j}^c \mid \mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{u}_j) \mathrm{d}\mathbf{u}_j, \\
&= \frac{\eta(\tau)}{\sum_{\tau \in \mathcal{P}_n} \eta(\tau)}.
\end{aligned}$$

Influency of the partition choice

	Real Partition		Estimated Partition		Random Partition	
α	2	6	2	6	2	6
CRPS	**0.158**	**0.118**	0.237	0.173	0.250	0.175
CRPSS	**36.8%**	**32.8%**	5.1%	1.2%	0	0

Table: Importance of a good estimation of the partition.