# Autoregressive models for maxima with atmospheric applications

**Gwladys Toulemonde**
*Université Paris VI, France*

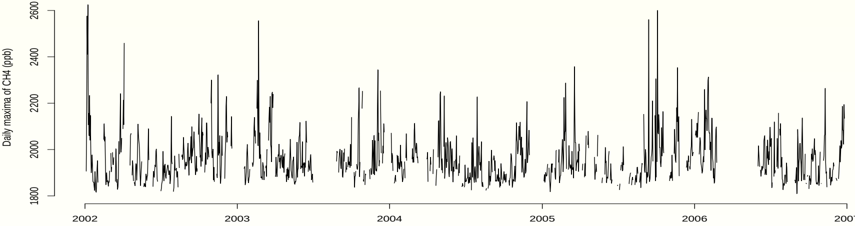
**AssimilEx-ANR project members**

Journées MAS de la SMAI.

27/08/2008

# Outline

# Series of daily maxima of CH4 in Gif-sur-Yvette

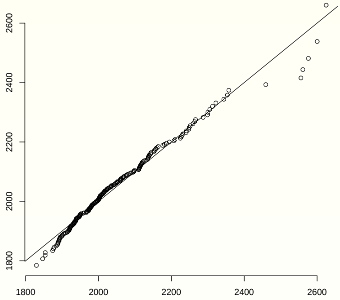


# The GEV distribution

$$\lim_{n\to\infty} \mathbb{P}\{a_n^{-1}(\max_{1\leq i\leq n} X_i - b_n) \leq x\} = \lim_{n\to\infty} F^n(a_n x + b_n) = H_\gamma(x)$$

where

$$H_\gamma(x) = \begin{cases} \exp\Big(-(1+\gamma x)^{-\frac{1}{\gamma}}\Big) & \text{with } x \text{ such that } 1+\gamma x > 0, \text{ if } \gamma \neq 0 \\ \exp\Big(-\exp(-x)\Big) & \text{for all } x \in \mathbb{R}, \text{ if } \gamma = 0 \end{cases}$$



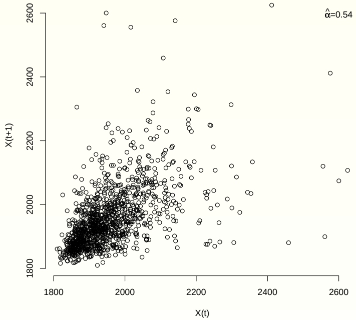




FIG. 1: Gumbel QQplot (on the left) and scatter plot of successive values, i.e. $(X_t, X_{t+1})$ (on the right) corresponding to the daily maxima of CH4.

# State-space models

Our objective : Proposing state-space models that preserve the nature of maxima distributions.

Classical state-space models used in geophysics :

$$Y_t = F_t(X_t, \varepsilon_t) \qquad \text{observation equation}$$
$$X_t = G_t(X_{t-1}, \eta_t) \qquad \text{state equation}$$

where we suppose independence between and within noises $\varepsilon_t$ and $\eta_t$.

Classical extra assumptions :

- Gaussian noises
- Linearity for the observation and state equations

If a record occurs in $Y_t$, then one expects $Y_t$ to be a GEV. The problem is that a GEV distribution cannot be retrieved through a Gaussian additive state-space model.

# A max-stable state-space model

Max-stable processes : Davis and Resnick (1989)

Naveau and Poncet (2007)

$$Y_t = F_t X_t \vee \varepsilon_t$$
$$X_t = G_t X_{t-1} \vee \eta_t$$

where $\varepsilon_t$ and $\eta_t$ correspond to Frechet noises

They proposed lower and upper bounds for $X_t$.

# A key linear relationship (Tawn, 1990)

$$\text{Gumbel}(\mu_1 + \mu_2, \sigma/\alpha) = \mu_2 + \sigma \log S + \text{Gumbel}(\mu_1, \sigma)$$

where Gumbel$(\mu_1, \sigma)$ is a Gumbel r.v. and independent of $S$ that is a positive $\alpha$-stable r.v. ($\alpha \in (0, 1]$) with Laplace transform

$$\mathbb{E}(\exp(-uS)) = \exp(-u^\alpha), \text{ for all } u > 0$$

Fougères et al. (2007) : If

$$Y_t = F_t \log\left(\sum_{a\in A} c_{t,a} S_a\right) + \varepsilon_t, \text{ with } t = 1,\dots,T,$$

where $\{c_{t,a} \geq 0\}$, $\{S_a, a \in A\}$ are independent positive $\alpha$-stable variables, $\varepsilon_t$ follows an iid Gumbel$(\mu_t, F_t)$, and all variables are mutually independent, then

$$\mathbb{P}(Y_1 \leq x_1, \dots, Y_T \leq x_T) = \prod_{a\in A} \exp\left(-\left(\sum_{t\in T} c_{t,a} e^{-\frac{x_t-\mu_t}{F_t}}\right)^{\alpha}\right)$$

# Gumbel state-space model

Naveau and Poncet (2007)

$$Y_t = F_t \log U_t + \varepsilon_t$$
$$U_t = G_t U_{t-1} + S_t$$

where $\varepsilon_t$ corresponds to an iid Gumbel noise and $S_t$ represents an iid positive $\alpha$-stable noise.

This implies

$$Y_t = F_t \log \left( \sum_{i=0}^{\infty} c_{t,i} S_{t-i} \right) + \varepsilon_t$$

i.e. $Y_t$ are Gumbel distributed. Moreover, the state-space model is linear !

# State-space models

Our objective : To propose state-space models that preserve the nature of maxima distributions.

Classical linear state-space models :

$$Y_t = F_t X_t + \varepsilon_t$$
$$X_t = G_t X_{t-1} + \eta_t$$

where we suppose independence between and within noises $\varepsilon_t$ and $\eta_t$.

A key linear relationship (Tawn, 1990)

$$\text{Gumbel}(\mu_1 + \mu_2, \sigma/\alpha) = \mu_2 + \sigma \log S + \text{Gumbel}(\mu_1, \sigma)$$

where Gumbel$(\mu_1, \sigma)$ is a Gumbel r.v. and independent of $S$ that is a positive $\alpha$-stable ($\alpha \in (0, 1]$) with Laplace transform

$$\mathbb{E}(\exp(-uS)) = \exp(-u^{\alpha}), \text{ for all } u > 0$$

# Gumbel linear model

## Result (1)

*Let* $\{X_t, t \in \mathbb{Z}\}$ *be a stochastic process defined by the recurrence relation*

$$X_t = \alpha X_{t-1} + \alpha\sigma \log S_t \tag{1}$$

*where* $\sigma \in \mathbb{R}_*^+$.
*Equation (1) has a unique strictly stationary solution,*

$$X_t = \sigma \sum_{j=0}^{\infty} \alpha^{j+1} \log S_{t-j} \tag{2}$$

*and* $X_t$ *follows a Gumbel* $(0, \sigma)$ *distribution,* $\forall\, t \in \mathbb{Z}$.

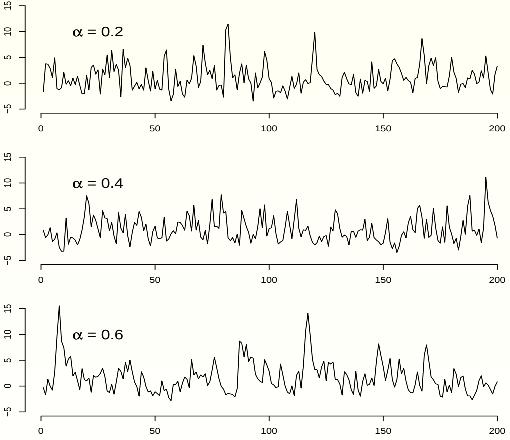



FIG. 2: Simulated time series $X_t$ from proposed model, with $t = 1, \ldots, 200$. We set $\sigma = 2$.

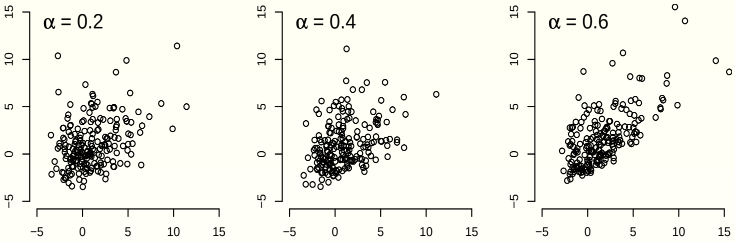



FIG. 3: Scatter plots of successive values, i.e. $X_t$ versus $X_{t+1}$ from the three simulated time series of the previous figure

Covariance $\mathbb{C}ov(X_t, X_{t-h}) = \mathbb{V}ar(X_0)\alpha^{|h|}$

# Estimators of the three unknown parameters

Natural estimators of $\mu$ and $\sigma$ based on the method of moments are

$$\widehat{\mu} = \overline{X} - \frac{\delta\sqrt{6}s}{\pi}$$

$$\widehat{\sigma} = \frac{\sqrt{6}s}{\pi}$$

where $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$, $s^2 = \frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{X})^2$ and where $\delta$ is the Euler's constant.

An estimator for $\alpha$ could be the following

$$\widehat{\alpha} = \frac{\frac{1}{T}\sum_{t=1}^{T-1}(X_t - \overline{X})(X_{t+1} - \overline{X})}{s^2}.$$

# Asymptotic behavior of the three estimators

## Result (2)

*The estimators of the three parameters $\mu$, $\sigma$ and $\alpha$ are consistent (a.s.).*

## Result (3)

$$\sqrt{T}\begin{pmatrix}\widehat{\mu}-\mu\\ \widehat{\sigma}-\sigma\\ \widehat{\alpha}-\alpha\end{pmatrix}$$

*is asymptotically normal with null expectation and covariance matrix defined as follows*

$$\begin{pmatrix} \frac{\pi^2\sigma^2}{6}\frac{1+\alpha}{1-\alpha}-\frac{12\delta\sigma^2\zeta(3)(1+\alpha+\alpha^2)}{\pi^2(1-\alpha^2)}+\frac{11\delta^2\sigma^2(1+\alpha^2)}{10(1-\alpha^2)} & \frac{6\sigma^2\zeta(3)(1+\alpha+\alpha^2)}{\pi^2(1-\alpha^2)}-\frac{11\delta\sigma^2(1+\alpha^2)}{10(1-\alpha^2)} & -\alpha\sigma\delta \\ \frac{6\sigma^2\zeta(3)(1+\alpha+\alpha^2)}{\pi^2(1-\alpha^2)}-\frac{11\delta\sigma^2(1+\alpha^2)}{10(1-\alpha^2)} & \frac{11\sigma^2(1+\alpha^2)}{10(1-\alpha^2)} & \alpha\sigma \\ -\alpha\sigma\delta & \alpha\sigma & 1-\alpha^2 \end{pmatrix}$$

# Asymptotic dependence parameters

## Result (5)

- *The upper tail dependence parameter $\chi$ defined as follows*

$$\chi = \lim_{x\to\infty} \frac{\mathbb{P}(X_{t-1} > x \quad \textit{and} \quad X_t > x)}{\mathbb{P}(X_{t-1} > x)}$$

  *is equal to zero (asymptotic independence).*

- *The dependence parameter $\overline{\chi}$ defined as follows*

$$\overline{\chi} = \lim_{x\to\infty} \frac{2\log \mathbb{P}(X_{t-1} > x)}{\log \mathbb{P}(X_{t-1} > x, X_t > x)} - 1$$

  *provides a measure which increases with dependence strength and is equal to $\alpha/(2-\alpha) \in (0,1)$.*

# Simulations

- $1000$ samples
- Sample sizes $n = 50, 100, \ldots, 1000$
- $\alpha \in \{0.2, 0.5, 0.8\}$
- $\mu = 0$ and $\sigma = 2$
- for each parameter : the mean of the estimations and also the first and the third quartiles

# Finite sample behavior of the estimators of $\mu$ and $\sigma$

$\alpha = 0.2$ $\alpha = 0.5$ $\alpha = 0.8$

FIG. 4: Mean (dashed line), first and third quartiles (dotted-dashed lines) for different sample sizes in the abscissa $n \in \{50, 100, \ldots, 1000\}$.

# Daily maxima of CH4

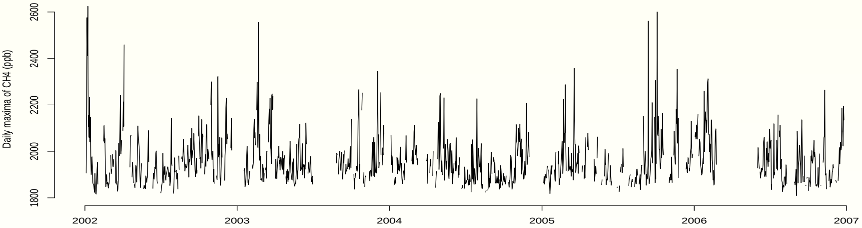



FIG. 5: Series of daily maxima of CH4 in Gif-sur-Yvette.

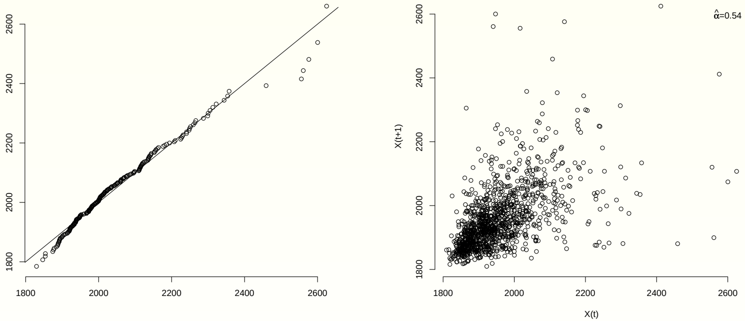



FIG. 6: Gumbel QQplot (on the left) and scatter plot of successive values, i.e. $(X_t, X_{t+1})$ (on the right) corresponding to the daily maxima of CH4.

# One-step prediction(1/2)

$1000$ estimations of $X_t|X_{t-1} = x$, $t = 1, \ldots, T-1$

- According to the proposed Gumbel model

$$X_t = \alpha X_{t-1} + \alpha\sigma \log S_t$$

- According to the classical Gaussian model

$$X_t = \alpha X_{t-1} + \varepsilon_t$$

$\hookrightarrow$ White histograms
We compare these white histograms with shaded areas which are the histograms of our daily maxima of CH4.

# Daily maxima of CH4

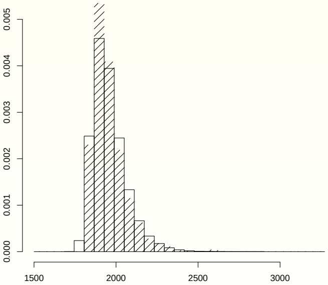



FIG. 7: Histograms of CH4 daily maxima previsions according to Gumbel model (on the left) and Gaussian model (on the right). The shaded histograms represent CH4 daily maxima.

# One-step prediction(2/2)

- We compute the estimators of the three parameters on a first period, here from 2002 to the middle of 2006
- For the second part of 2006, we compute 1000 $\widehat{X}_{t+1} = \widehat{\alpha} x_t + \widehat{\alpha}\widehat{\sigma} \log S_{t+1}$ with $x_t$ the observed value at time $t$ and $S_{t+1}$ a positive $\widehat{\alpha}$-stable random variable
- We deduce the empirical quartiles of the distribution of $\widehat{X}_{t+1} | X_t = x_t$.

# Daily maxima of CH4

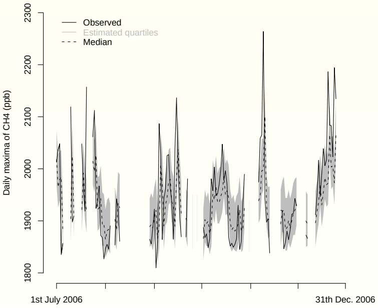



FIG. 8: Validation of the one-step previsions of methane daily maxima on the second part of the year 2006.

# Generalization to the GEV distribution

Let E be a random variable from a GEV$(\mu, \sigma, \gamma)$.

- If $\gamma$ is strictly negative then $-\log(\mu - \sigma/\gamma - E)$ follows a Gumbel$(\log(-\gamma/\sigma), -\gamma)$.
- If $\gamma$ is strictly positive then $\log(E - \mu + \sigma/\gamma)$ follows a Gumbel$(\log(\sigma/\gamma), \gamma)$.

**Result (6)**

$$X_t = \mu - \frac{\sigma}{\gamma} + \left(X_{t-1} - \mu + \frac{\sigma}{\gamma}\right)^{\alpha} \times S_t^{\alpha\gamma} \times \left(\frac{\sigma}{\gamma}\right)^{1-\alpha} \qquad (3)$$

*where* $(\mu, \sigma, \gamma) \in \mathbb{R} \times \mathbb{R}_*^+ \times \mathbb{R}_*$.
*Equation (3) has a unique strictly stationary solution given by*

$$X_t = \mu - \frac{\sigma}{\gamma} + \frac{\sigma}{\gamma} \prod_{j=0}^{\infty} (S_{t-j})^{\gamma\alpha^{j+1}}$$

*and* $X_t$ *follows a* $GEV(\mu, \sigma, \gamma)$ *distribution,* $\forall\, t \in \mathbb{Z}$.

# Future works

Gumbel state-space model

$$Y_t = F_t(\alpha X_t + \alpha\sigma \log \varepsilon_t)$$
$$X_t = \alpha X_{t-1} + \alpha\sigma \log S_t$$

where $\varepsilon_t$ and $S_t$ two iid positive $\alpha$-stable noises.
This implies that $Y_t$ are Gumbel$(0, F_t\sigma)$.

- Additive model, $X_t$ and $Y_t$ are both Gumbel
- $G_t$ has to be equal to $\alpha \in (0, 1)$ and $Y_t$ has a specific form.

Thank you for your attention !

# References


- Davis, R.A., Resnick, S.I., 1989. Basic properties and prediction of Max-Arma Processes. *Adv. Appl. Prob.*, **21**, 781-803.
- Fougères, A.L., Nolan, J.P., Rootzén, H, 2005. Mixture models for extremes, *submitted*.
- Joe, H., 1993. Parametric families of multivariate distributions with given margins. *J. Multivariate Anal.*, **46**, 262-282.
- Leadbetter, M. R., 1974. On extreme values in stationary sequences. *Z. Wahrsch. Verw. Gebiete.*, **28**, 289-303.
- Leadbetter, M. R., Lindgren, G., Rootzén, H., 1983. *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York.
- Morales, F.,C., 2005. Estimation of Max-Stable Processes Using Monte Carlo Methods with Applications to Financial Risk Assessment. Ph.D. Dissertation.
- Naveau, P., Poncet, P., 2007. State-space models for maxima precipitation. *Journal de la Société française de statistique*, **148**, 107-120.
- Samorodnitsky, G., Taqqu, M.S., 1994. *Stable non-Gaussian random processes*, Chapman & Hall, New York.
- Tawn, J.A., 1990. Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245-253.
- Zolotarev, V.M., 1986. *One-Dimensional Stable Distributions*. American Mathematical Society Translations of Mathematical Monographs, Vol. 65. American Mathematical Society, Providence, RI. Translation of the original 1983 (in Russian).

# Morales' state-space model (2005)

$$Y_t = \mu + \sigma \frac{X_t^\gamma - 1}{\gamma} + \varepsilon_t$$
$$X_t = [\beta \eta_t] \vee [(1-\beta)\eta_{t-1}]$$

where $a \vee b = max(a,b)$, $\eta_t$ unit Fréchet iid r.v. and $\varepsilon_t$ Gaussian noise.

Note : $\mu + \sigma \frac{X_t^\gamma - 1}{\gamma}$ is GEV$(\mu, \sigma, \xi)$.

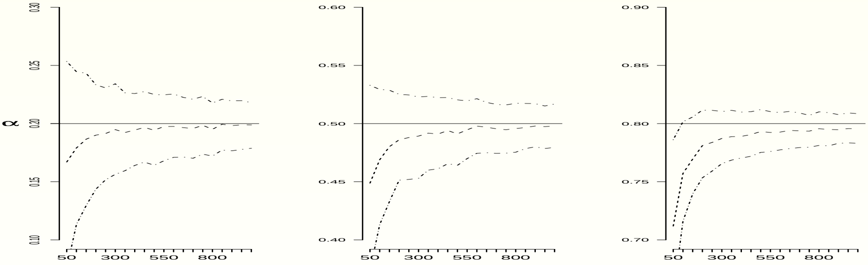
# Joint distribution of the vector $\mathbf{X_h} = (X_t, \ldots, X_{t-h})^t$, $h > 0$

## Result (4)

*The characteristic function of* $\mathbf{X_h} = (X_t, \ldots, X_{t-h})^t$, $h > 0$, *noted* $\mathbb{E}\left(e^{i<u,\mathbf{X_h}>}\right)$ *is*

$$\Gamma\left(1 - i\sigma \sum_{j=0}^{h} u_j \alpha^{h-j}\right) \prod_{j=0}^{h-1} \frac{\Gamma\left(1 - i\sigma \sum_{k=0}^{j} u_k \alpha^{j-k}\right)}{\Gamma\left(1 - i\sigma \sum_{k=0}^{j} u_k \alpha^{j-k+1}\right)}.$$

# Finite sample behavior of the estimator of $\alpha$




$\alpha = 0.2$ $\alpha = 0.5$ $\alpha = 0.8$

FIG. 9: Mean (dashed line), first and third quartiles (dotted-dashed lines) for different sample sizes in the abscissa $n \in \{50, 100, \ldots, 1000\}$.