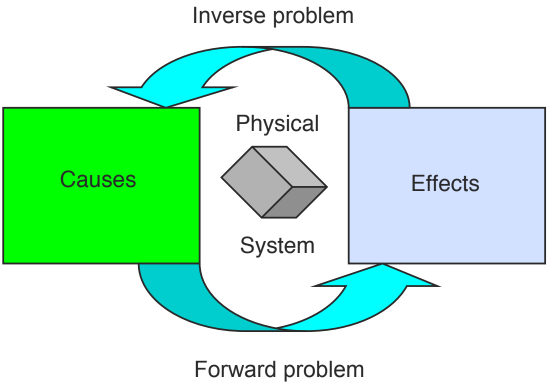
Sequential estimation Kalman filtering and a few ways to go beyond the Gaussian linear framework

Philippe Naveau

Laboratoire des Sciences du Climat et l'Environnement (LSCE)
Gif-sur-Yvette, France

17 octobre 2012

Alexis' view on D&A



Pierre Simon Laplace's view on D&A

"If an event can be produced by a number of n different causes, then the probabilities of the causes given the event ... are equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of the causes."

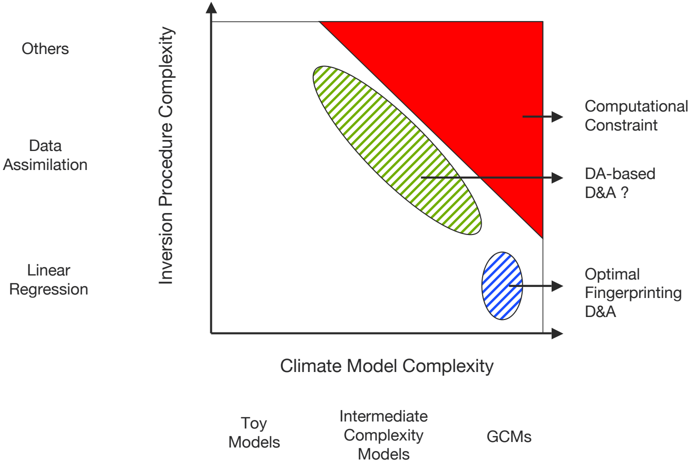
(1749-1827)

Pierre Simon Laplace (1749-1827)

"If an event can be produced by a number of n different causes, then the probabilities of the causes given the event ... are equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of the causes."

$$\mathbb{P}(\text{cause}_i|\text{event}) = \frac{\mathbb{P}(\text{event}|\text{cause}_i) \times \mathbb{P}(\text{cause}_i)}{\sum_{j=1}^{n} \mathbb{P}(\text{event}|\text{cause}_j) \times \mathbb{P}(\text{cause}_j)}$$

Where am I ?


Others
Data Assimilation
Linear Regression
Inversion Procedure Complexity
Computational Constraint
DA-based D&A ?
Optimal Fingerprinting D&A
Climate Model Complexity
Toy Models
Intermediate Complexity Models
GCMs

Leaving the Deutschmark for the Euro

Why was it hard to leave the Deutschmark ?

- The assumption of normality is very prevalent in the theoretical and applied statistical research
- Asymptotic justification : Central Limit Theorem
- Nice properties of Gaussian vectors

- Completely characterized by its first two moments
- Stability under linearity
- Stability under summation
- Stability under conditioning

Bayesian Kalman filter (Meinhold and Singpurwalla, 1983)

Observation equation

$$Y_t = F X_t + V_t \text{ with } V_t \sim N[\mathbf{0}, V]$$

State equation

$$X_t = G X_{t-1} + W_t \text{ with } W_t \sim N[\mathbf{0}, W]$$

Conditional distribution of X_t given $\mathbf{Y}_{1:t}$

If we assume

$$\left[X_{t-1} | Y_{1:(t-1)}\right] \sim N\left[\hat{X}_{t-1}, \Sigma_{t-1}\right]$$

then

$$[X_t | Y_{1:t}] \sim N\left[\hat{X}_t, \Sigma_t\right]$$

with

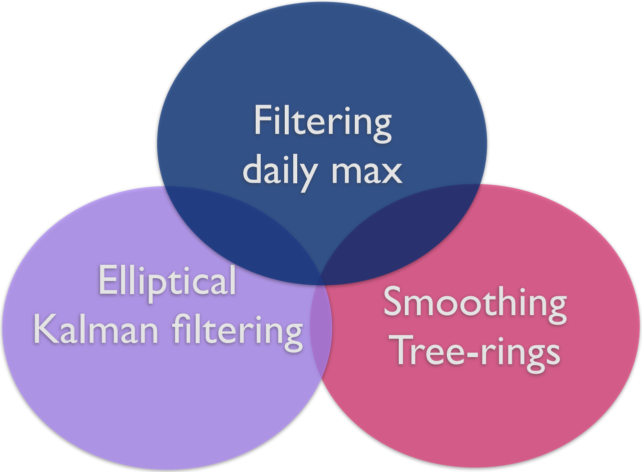
$$\hat{X}_t = G\hat{X}_{t-1} + R_t F^T (V + F R_t F)^{-1} e_t \text{ and } \Sigma_t = R_t - R_t F^T (V + F R_t F)^{-1} F R_t$$

where $R_t = G\Sigma_{t-1} G^T + W$ and $e_t = Y_t - F G \hat{X}_{t-1}$.

[1]

1. Brockwell and Davis, 2002 (chap 8) and 1991 (chap 12), Meinhold and Singpurwalla, (1983)

Three examples



Three examples

Daily maxima of methane and nitrous oxide at LSCE

- Joint work with Gwladys Toulemonde and Armelle Guillou

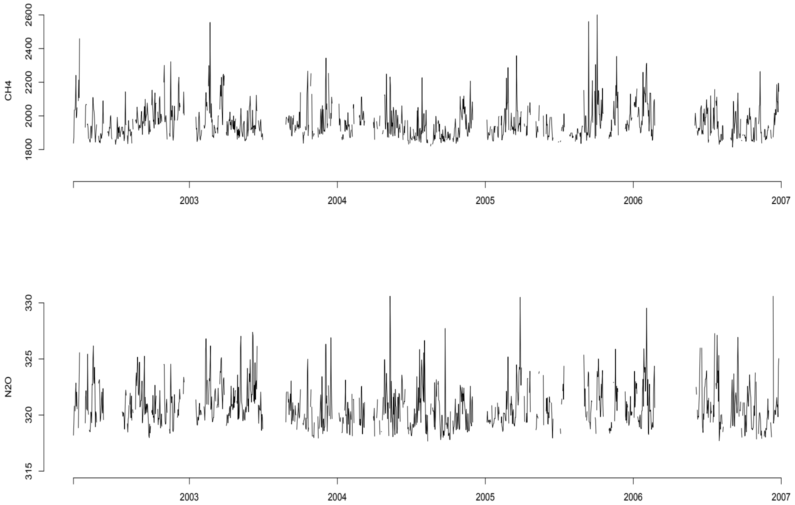


Figure 1: Daily maxima of CH_4 and N_2O during the period 2002-2007. Measurements in parts per billion by volume (ppbv) were made at LSCE, a laboratory located at Gif-sur-Yvette, a city south west of Paris, France. Data are missing during a few time lags and

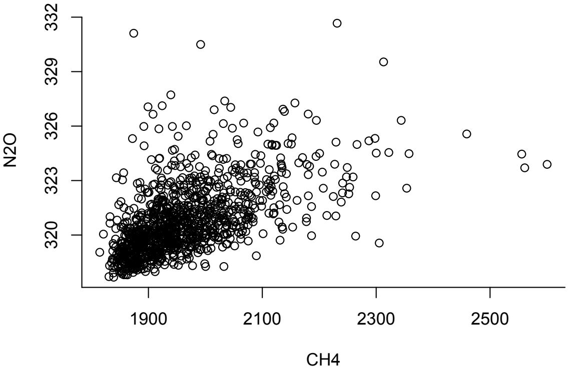
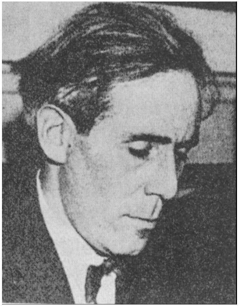


Figure 2: Scatterplot between daily maxima concentrations of CH_4 (x-axis) and N_2O (y-axis), see Figure 1.

Gumbel



CDF

$F(x) = \exp(-\exp(-(x-\mu)/\sigma))$ for all real x.

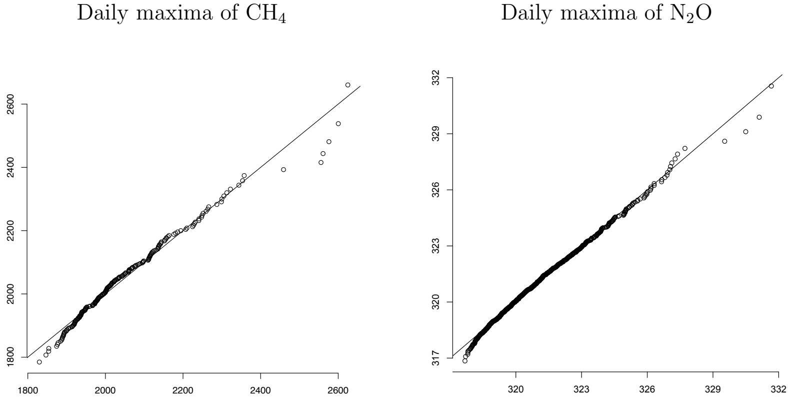


Figure 3: QQ-plots of daily maxima of CH_4 and N_2O obtained after fitting a Gumbel distribution via a method-of-moment technique proposed in Toulemonde *et al.* (2010). In (1), daily maxima of methane have estimates with 95% confidence intervals: $\hat{\sigma} = 79.8 \in [73.3; 86.4]$, $\hat{\mu} = 1915.9 \in [1904.4; 1927.4]$, and, for nitrous oxide, $\hat{\sigma} = 1.52 \in [1.39; 1.64]$, $\hat{\mu} = 320.0 \in [319.7; 320.2]$. The x-axis and y-axis represent the observed and expected ranked values.

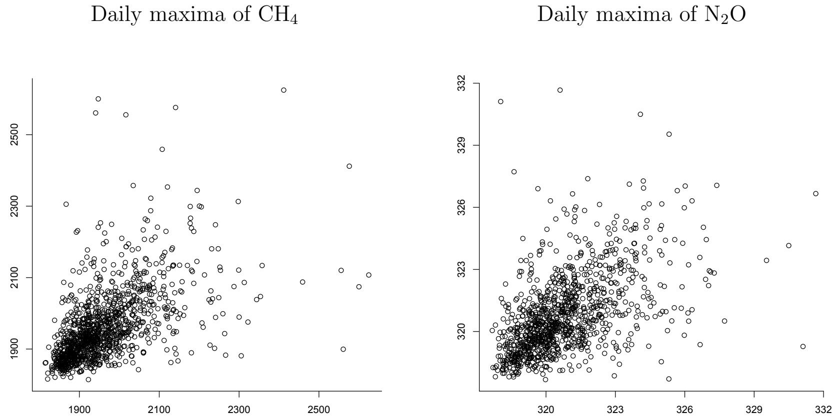


Figure 4: Scatter plots of consecutive maxima of CH_4 and N_2O. The x-axis corresponds to day t and the y-axis to day $t+1$. The empirical estimate of the lag 1 autocorrelation is equal to 0.55 for the CH_4 and to 0.52 for the N_2O.

The problems at hand

The Scientific Problem Under Study

- How to reconstruct missing maxima from one of each time series ?

The statistical Problem Under Study

- How to make on-line forecasts with Gumbel distributed random variables ?

A key linear relationship

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

where Gumbel(μ_1, σ) denotes a Gumbel r.v. which is independent of S that is a positive α-stable r.v. ($\alpha \in (0, 1]$) with Laplace transform

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

- A random variable S is said to be (α)-stable if and only if for all $k > 1$ there exist $c_k > 0$ and d_k such that $S_1 + \ldots + S_k \stackrel{d}{=} c_k S + d_k$ where S_1, S_2... are iid copies of S.
- Examples and special cases where one can write down explicit expressions for the density : Gaussian, Cauchy, Levy distributions.

Another state-space Gumbel maxima model

PROPOSED MODEL: *Let* $\{Z_t, t \in \mathbb{Z}\}$ *and* $\{Y_t, t \in \mathbb{Z}\}$ *be two stochastic processes defined as follows*

$$\begin{cases} Y_t = \nu_t + H_t Z_t + \eta_{t,\alpha_2} & \textit{(observational equation)} \\ Z_t = \alpha_1 Z_{t-1} + \varepsilon_{t,\alpha_1} & \textit{(state equation)} \end{cases} \tag{3}$$

where $H_t > 0$, $\alpha_1 \in (0,1)$, $\alpha_2 \in (0,1)$ *and the sequences* $\{\varepsilon_{t,\alpha_1}\}_t$ *and* $\{\eta_{t,\alpha_2}\}_t$ *correspond to two independent samples of Exponential-Stable variable,* $ExpS\,(\alpha_1, -\sigma\gamma(1-\alpha_1), \alpha_1\sigma)$ *and* $ExpS\,(\alpha_2, -H_t\sigma\gamma(1/\alpha_2 - 1), H_t\sigma)$*, respectively. The variable* ε_{t,α_1} *is independent of* $\{Z_{t'}\}_{t' \leq t-1}$ *and the variable* η_{t,α_2} *is independent of* $\{Z_{t'}\}_{t' \leq t}$*. The scalar* γ *is the Euler's constant.*

Properties

Margins

the variables Z_t and Y_t are Gumbel distributed with parameters $(-\gamma\sigma, \sigma)$ and $(\nu_t - \frac{H_t \gamma\sigma}{\alpha_2}, H_t \frac{\sigma}{\alpha_2})$

Covariances

$$\mathbb{C}ov(Z_t, Z_{t-h}) = \alpha_1^{|h|} \mathbb{V}ar(Z_t),$$
$$\mathbb{C}ov(Y_t, Z_t) = H_t \mathbb{V}ar(Z_t),$$
$$\mathbb{C}or(Y_t, Z_t) = \alpha_2$$

Filtering

$$p(Z_{k-1}|Y_{1:k-1}) \xrightarrow[p(Z_k|Z_{k-1})]{prediction} p(Z_k|Y_{1:k-1}) \underset{p(Y_k|Z_k)}{\overset{\overset{Y_k}{\downarrow}}{\xrightarrow{correction}}} p(Z_k|Y_{1:k})$$

Prediction and filtering densities

$$p(Z_k|Y_{1:k-1}) = \int p(Z_k|Z_{k-1})p(Z_{k-1}|Y_{1:k-1})dZ_{k-1} \qquad \text{(Prediction step)}$$

$$p(Z_k|Y_{1:k}) = \frac{p(Y_k|Z_k)p(Z_k|Y_{1:k-1})}{\int p(Y_k|Z_k)p(Z_k|Y_{1:k-1})dZ_k} \qquad \text{(Correction step)}$$

Particle filtering

Auxiliary particle filter (APF)

At time $t = t_0$

$$\xi_{t_0}^{1:N} \overset{iid}{\sim} p(X_{t_0})$$
$$w_{t_0}^{1:N} \leftarrow \frac{1}{N}$$

At time $t_0 < k \leq T$,

1) Selection step

$$\beta_k^i \leftarrow w_{k-1}^i \widehat{p}(Y_k|\xi_{k-1}^i)$$
$$j^{1:N} \leftarrow \text{resample}(\beta_k^{1:N}, 1:N)$$

2) Propagation

$$\xi_k^i \sim p(X_k|\xi_{k-1}^{j^i}) \text{ for } i = 1, ..., N$$

3) Computation of the weights for $i = 1, ..., N$

$$w_k^i \leftarrow \frac{p(Y_k|\xi_k^i)}{\widehat{p}(Y_k|\xi_{k-1}^{j^i})}$$
$$w_k^i \leftarrow \frac{w_k^i}{\sum_{i=1}^N w_k^i}$$

Particle filtering, a few references (source Olivier Cappé)

- Doucet, A., De Freitas, N. and Gordon, N. (eds.) (2001) *Sequential Monte Carlo Methods in Practice*. Springer.
- Ristic, B., Arulampalam, M. and Gordon, A. (2004) *Beyond Kalman Filters: Particle Filters for Target Tracking*. Artech House.
- Cappé, 0., Moulines, E. and Rydén, T. (2005) *Inference in Hidden Markov Models*. Springer.
- Doucet, A., Godsill, S. and Andrieu, C. (2000) On sequential Monte-Carlo sampling methods for Bayesian filtering. Stat. Comput., **10**, 197-208.
- Arulampalam, M., Maskell, S., Gordon, N. and Clapp, T. (2002) A tutorial on particle filters for on line non-linear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process., **50**, 241–254.
- Cappé, O., Godsill, S. J. and Moulines, E. (2007) An overview of existing methods and recent advances in sequential Monte Carlo, IEEE Proc., **95**, 899–924.

Particle Filtering

Weight particles for our Gumbel model (reducing the computational cost)

$$p(y_t|\xi^i_{t-1}) = \frac{1}{H_t\sigma} f_{U_{t,\alpha_1,\alpha_2}}\left(\frac{y_t - C}{H_t\sigma}\right)$$

where

$$U_{t,\alpha_1,\alpha_2} = \alpha_1 \log S_{t,\alpha_1} + \log S_{t,\alpha_2}$$

and

$$C = \nu_t - \frac{H_t\gamma\sigma}{\alpha_2} + H_t\alpha_1\gamma\sigma + H_t\alpha_1\xi^i_{t-1}$$

Comparing MSE for different methods

	KF	BF_{500}	APF-PS_{500}	APF-Opt_{500}
$\alpha_1 = 0.1$ and $\alpha_2 = 0.4$	1.354	1.317	1.317	**1.314**
$\alpha_1 = 0.1$ and $\alpha_2 = 0.6$	1.036	1.017	1.096	**1.013**
$\alpha_1 = 0.5$ and $\alpha_2 = 0.4$	1.336	1.296	1.233	**1.222**
$\alpha_1 = 0.5$ and $\alpha_2 = 0.6$	0.994	0.959	0.905	**0.841**
$\alpha_1 = 0.9$ and $\alpha_2 = 0.4$	0.984	0.873	**0.764**	**0.764**
$\alpha_1 = 0.9$ and $\alpha_2 = 0.6$	0.665	0.569	**0.434**	**0.434**

Table 1: Mean of the MSEs based on 100 replica.

Missing two weeks out of three months

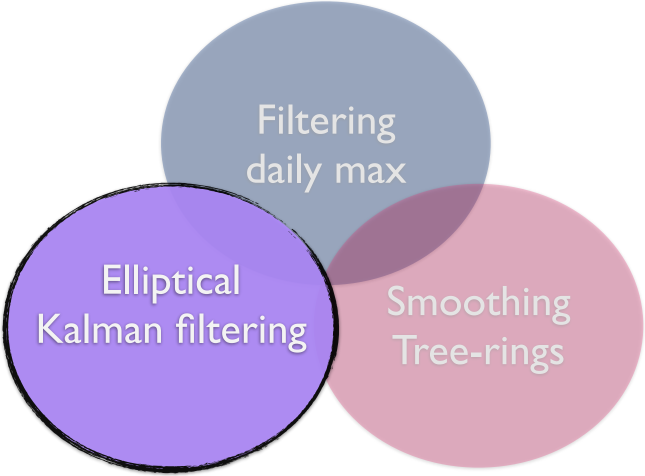
MSEs for return levels based on 100 replica for $\alpha_1 = 0.5$ and $\alpha_2 = 0.6$ with 500 particles.

Return period	Incomplete data	APF-PS$_N$	APF-Opt$_N$	Whole data
1 year (5.9)	0.77	0.65	0.62	0.61
5 year (7.5)	1.16	0.99	0.95	0.92
10 year (8.2)	1.35	1.16	1.11	1.08
50 year (9.8)	1.84	1.61	1.54	1.48

Conclusions about Gumbel state-space model

- Estimating hidden Gumbel distributed maxima is possible by using particle filtering techniques
- Optimizing the weights improves the MSE
- Very much tailored to Gumbel distributed maxima

Three examples



Elliptical distributions, GP tailed and Kalman filtering

- Joint work with Anne Sabourin

Elliptical distributions

A wide class, allowing for bounded or heavy tailed laws.

Definition

A random vector : $X \in \mathbb{R}^n$ with density f is elliptical with

- *Parameters* : $\mu \in \mathbb{R}^n$, $\Sigma \in \mathcal{M}_{n\times n}(\mathbb{R})$ a positive definite symmetric matrix
- *Density generator* g such that $\int_0^{+\infty} t^{n/2-1} g(t)dt < \infty$,

iff

$$f(x) = c_n|\Sigma|^{-1/2} g((x-\mu)'\Sigma^{-1}(x-\mu)),$$

$$c_n = \frac{\Gamma(n/2)}{\pi^{n/2}\int_0^{+\infty} t^{n/2-1} g(t)dt}$$

Gaussian vectors : a specific case of elliptical vectors with generator $g(s) = \exp(-\frac{s}{2})$ (see *e.g*[5] or [7])

Elliptical distributions

Any elliptical vector can be written as :

$$X = \mu + RA'U$$

where

- $U \in \mathbb{R}^n$ is uniformly distributed on the unit sphere
- $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is such that $A'A = \Sigma$
- R (called the radial variable) is a positive real random variable, independent from U and with density

$$h(r) = \frac{2}{\int t^{n/2-1} g(t)dt} r^{n-1} g(r^2) I_{[0,\infty[}(r)$$

An easy way to simulate elliptical distributions. see e.g [5] or [7]

Notations for conditioning

Crucial for filtering data !

Let $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $X_1 \in \mathbb{R}^p$, $X_2 \in \mathbb{R}^{n-p}$

Corresponding blocks for μ and Σ

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

see [5]

Margins, still elliptical

$$X_1 \sim \mathcal{E}_p(\mu_1, \Sigma_{11}, g_{(1)})$$

with

$$g_{(1)}(s) = \int_0^{+\infty} w^{\frac{n-p}{2}-1} g(s+w)dw$$

Conditioning, still elliptical

$$X_2 | (X_1 = x_1) \sim \mathcal{E}_{n-p}(\mu_{2|1}, \Sigma_{2|1}, g_{2|1})$$

with :

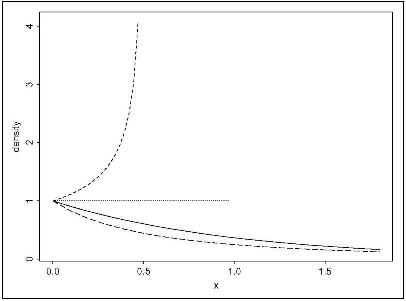
$$\begin{aligned}
\mu_{2|1} &= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1) \\
\Sigma_{2|1} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\
g_{2|1}(s) &= g(q_1 + s), q_1 = (X_1 - \mu_1)'\Sigma_{11}^{-1}(X_1 - \mu_1)
\end{aligned}$$

Same equations as for conditionals from Gaussian laws !

Thresholding : the Generalized Pareto Distribution (GPD)

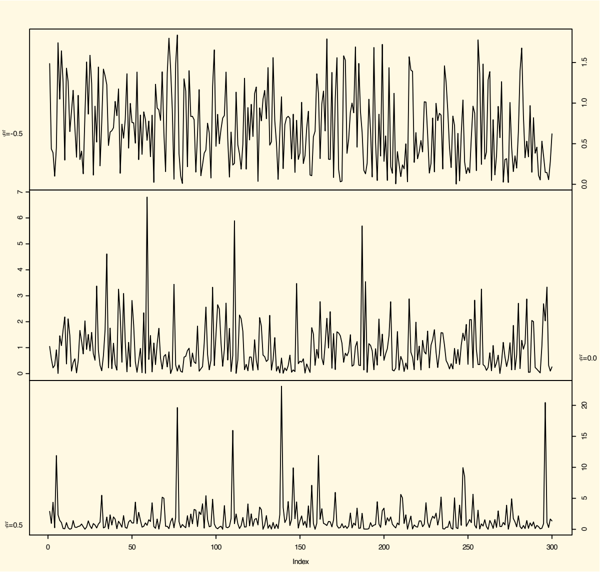
$$\mathbb{P}\{Y-u > y | Y > u\} = \left(1 + \frac{\xi\, y}{\sigma_u}\right)_+^{-1/\xi}$$

Vilfredo Pareto : 1848-1923



Born in France and trained as an engineer in Italy, he turned to the social sciences and ended his career in Switzerland. He formulated the power-law distribution (or "Pareto's Law"), as a model for how income or wealth is distributed across society.

GPD : “From Bounded to Heavy tails”



Elliptical Generator = the Generalized Pareto Tail

$$g_{\sigma,\xi}(s) = \mathbb{P}\{Y > s\} = \left(1 + \frac{\xi\, s}{\sigma}\right)_{+}^{-1/\xi}$$

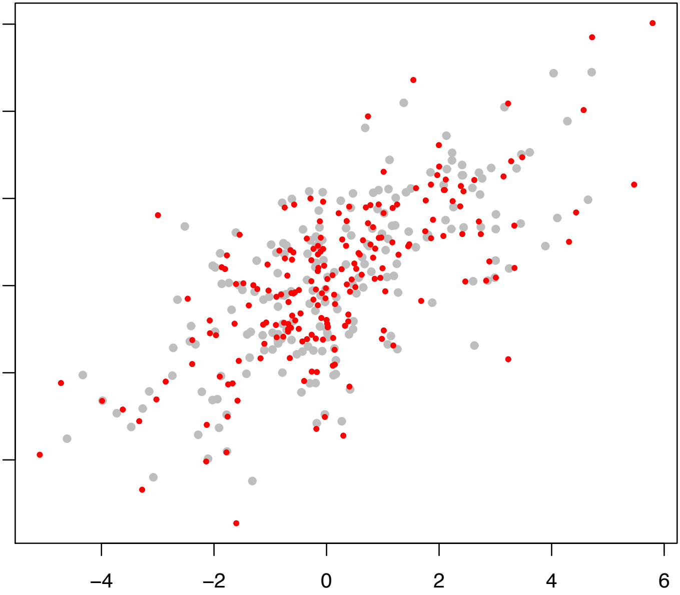
Elliptical distributions and Pareto generator

Fundamental property

$$g_{\sigma,\xi}(s+u) = g_{\sigma+\xi u}(s) g_{\sigma}(u)$$

A key to obtain explicit expressions for conditional and margins

Pareto versus exponential generators



AR(1) $X_t = FX_{t-1} + \epsilon_t$

Lemma

Elliptical innovations $\epsilon_t = (X_t - FX_{t-1} | x_{t-1})$ *have GPD generator with parameters :*

$$\tilde{\sigma} = \frac{\sigma + \xi q_{t-1}(x_{t-1})}{1 - \alpha\xi}, \qquad \tilde{\xi} = \frac{\xi}{1 - \alpha\xi}$$

Note : $q_{t-1}(x_{t-1})$ is as in (5)
Upper bound for $\tilde{\xi}$: $\tilde{\xi}_{sup} = \frac{1}{n}$

AR(1) $X_t = FX_{t-1} + \epsilon_t$

for $\xi > 0$

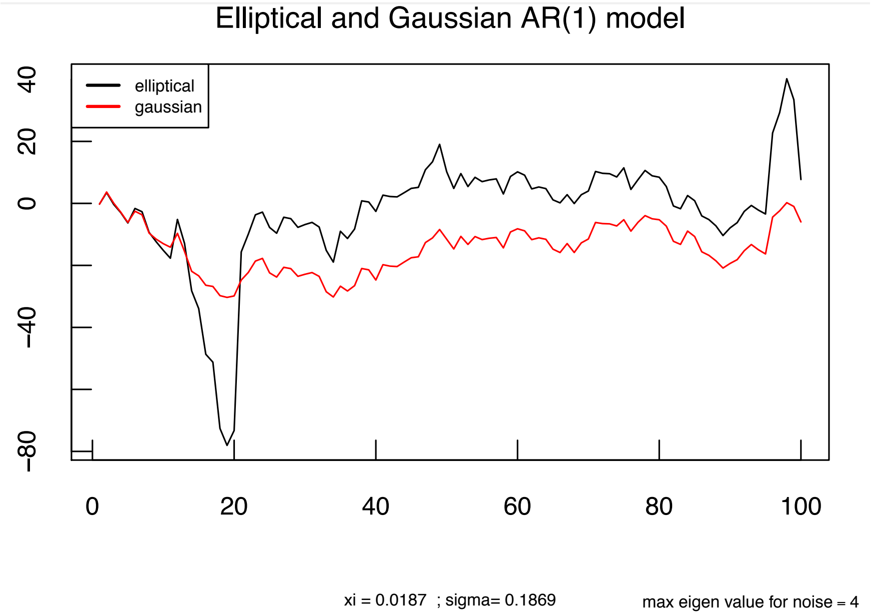
$$H_t(R) = \text{pbeta}_{(\frac{n}{2},\frac{1}{\xi}-\frac{n(T-1)}{2}-\frac{n}{2})}\left(\frac{\xi R^2}{\sigma + \xi(q_{t-1}(x_{t-1}) + R^2)}\right)$$

$$H_t^{-1}(u) = \sqrt{(\frac{\sigma}{\xi} + q_{t-1}(x_{t-1}))\frac{w_t^X(u)}{1 - w_t^X(u)}}$$

where

$$w_t^X(u) = pbeta^{-1}_{(\frac{n}{2},\frac{1}{\xi}-\frac{n(T-1)}{2}-\frac{n}{2})}(u)$$

AR(1) $X_t = FX_{t-1} + \epsilon_t$



Kalman filters and elliptical distributions

block matrices in Σ^W

$$
\begin{aligned}
\Sigma_{X_t} &= \Sigma_\epsilon + F\Sigma_{X_{t-1}}F' \\
\Sigma_{X_t,X_{t-k}} &= F^k\Sigma_{X_{t-k}} \\
\Sigma_{Y_t,X_{t-k}} &= GF^k\Sigma_{X_{t-k}} \\
\Sigma_{Y_t} &= G\Sigma_{X_t}G' + \Sigma_\nu \\
\Sigma_{Y_t,Y_{t-k}} &= GF^k\Sigma'_{X_{t-k}}
\end{aligned}
$$

Kalman filters and elliptical distributions

Generators

$$
\begin{aligned}
g_t^{\nu}(s) &= \int_0^{+\infty} w^{\frac{nT+p-n}{2}-1} g^W(s+w+q_t(x_t)) \\
g_t^{\epsilon}(s) &= \int_0^{+\infty} w^{\frac{nT-p}{2}-1} g^W(s+w+q_{t-1}(x_{t-1}))
\end{aligned}
$$

with $q_t(x_t) = x_t'(\Sigma_{X_t})^{-1}x_t$

Kalman filters : bringing the GPD

Choose $g^W(s) = g_{\sigma,\xi}(s)$ as a global generator for W.
Upper bound for ξ : $\xi_{\text{sup}} = \frac{2}{nT+p}$

Lemma

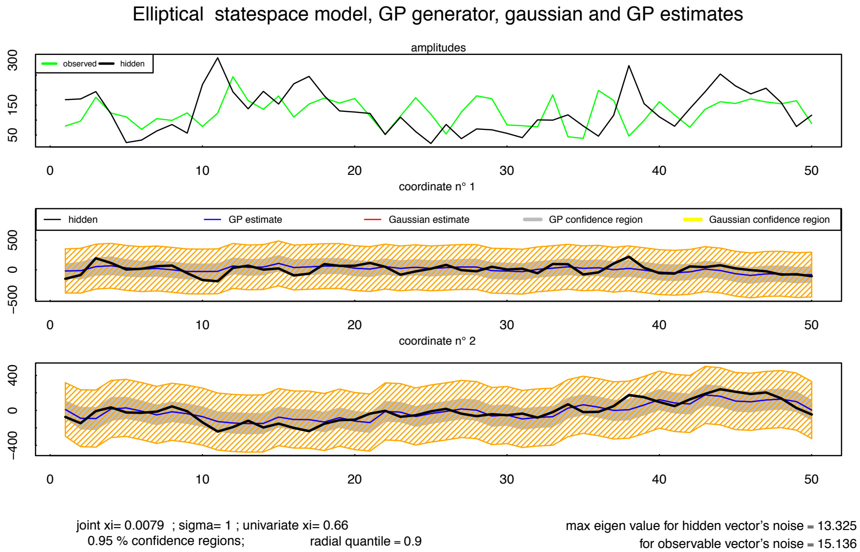
Elliptical innovations $\epsilon_t = (X_t - FX_{t-1}|x_{t-1})$ *and* $\nu_t = (Y_t - GX_t|x_t)$ *have GPD generator with parameters :*

$$\sigma^\epsilon = \frac{\sigma+\xi q_{t-1}(x_{t-1})}{1-\alpha^\epsilon\xi} \qquad \xi^\epsilon = \frac{\xi}{1-\alpha^\epsilon\xi}$$
$$\sigma^\nu = \frac{\sigma+\xi q_t(x_t)}{1-\alpha^\nu\xi} \qquad \xi^\nu = \frac{\xi}{1-\alpha^\nu\xi}$$

with $\alpha^\epsilon = \frac{nT-p}{2}, \quad \alpha^\nu = \frac{nT+p-n}{2}$

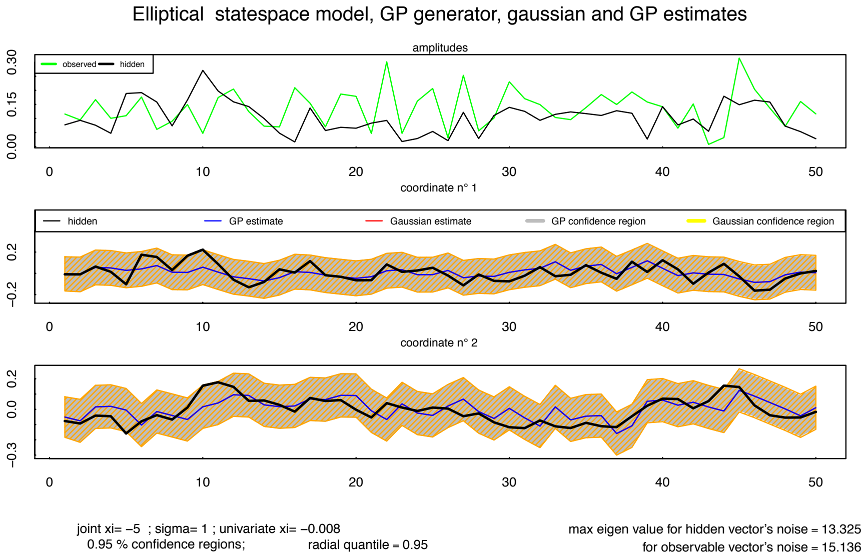
Simulations

FIGURE 4.1. $\xi > 0$, extreme radial quantile



Simulations

FIGURE 4.3. $\xi < 0$



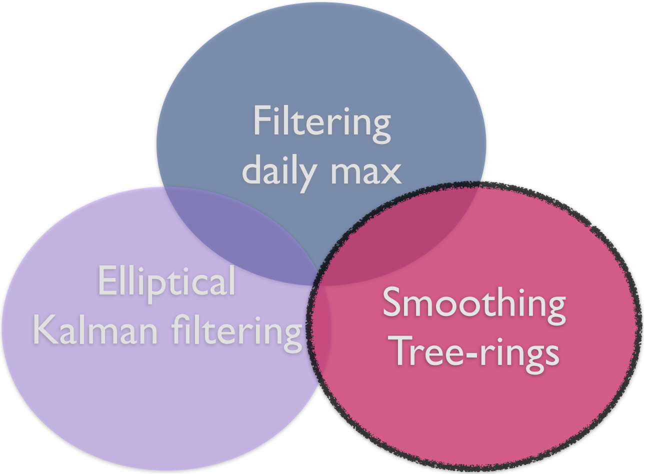
Conclusions about Elliptical KF

- Elliptical distributions with GPD generators provide explicit KF equations
- It can handle bounded, Gaussian and heavy tails
- Restricted to finite times series
- Looking for applications with symmetrical distributions

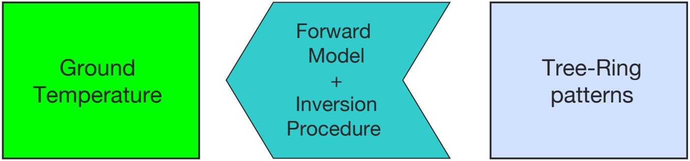
[7] E. Gómez, M.A. Gómez-Villegas, and J.M. Marín. Continuous elliptical and exponential power linear dynamic models. *Journal of Multivariate Analysis*, 83(1):22 – 36, 2002.

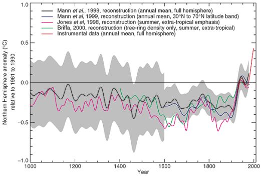
[8] E. Gómez, M.A. Gómez-Villegas, and J.M. Marín. A survey on continuous elliptical vector distributions. *Rev. Mat. Complut*, 16:345–361, 2003.

Three examples

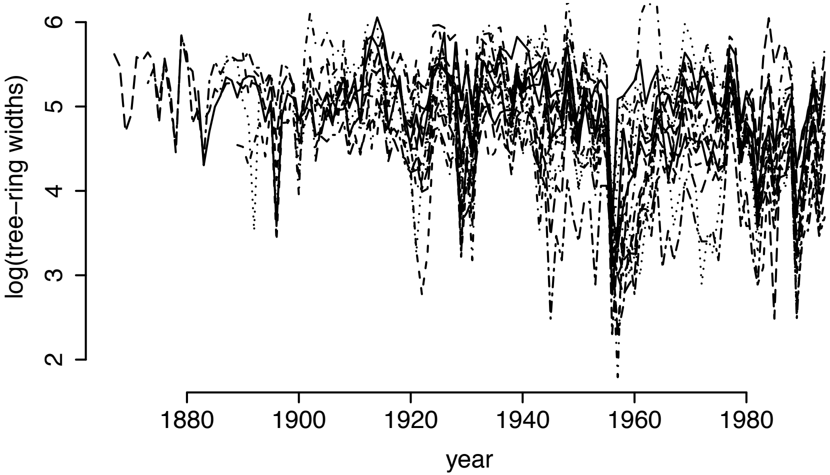


Dendro : an attempt to leave the linear world ?





Seventeen *Pinus halepensis Mill* tree ring width logarithms from the "Rognac" site (1867 – 1993)



The problems at hand

The Scientific Problem Under Study

- How to extract a common signal among 17 tree ring widths ?

The statistical Problem Under Study

- How to calculate the posterior distribution of a common signal ?

Similar BHM approaches

Hooten and Wikle, 2007

a BHM for the spatio-temporal growth dynamics of shortleaf pine but with chronology indices. They linked these chronologies with drought information like the Palmer Drought Severity Index.

Haslett, 2005

investigated the problem of reconstructing prehistoric climates from lake sediment cores.

The “linear aggregate model” (Cook 1990, Buckley 2009)

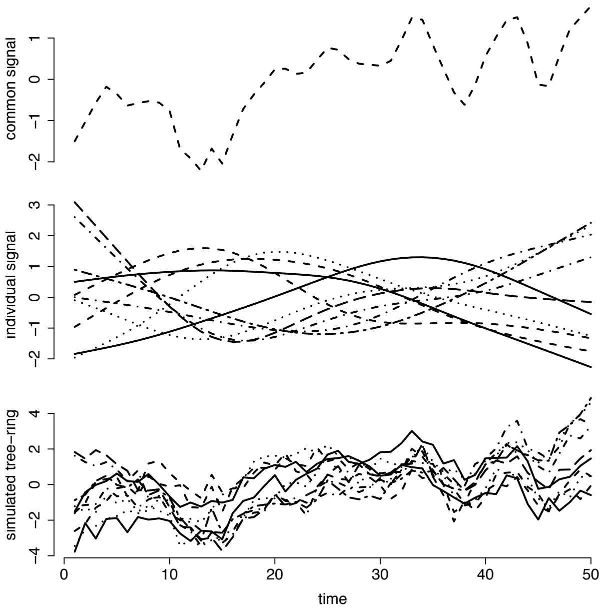
A log-additive model

$$\log(\text{ring width}) = F_t + G_t + D_t + \text{unexplained variability}$$

where

- t =year
- G_t the age-related trend due to normal physiological aging processes
- F_t to the climatically-related environmental signal
- $D_t (= 0)$ to disturbance factors, either within the forest stand or outside of it (e.g., insect outbreaks or fires).

Simulations : finding f and g_j from the y_j's (low panel)



Our main assumptions

$$\text{log(ring width)} = F_t + G_t + 0 + \text{unexplained variability}$$

Notations

- $\mathbf{y}_j = (y_j(t_1), ..., y_j(t_n))^T$ = log(ring width) produced by tree j
- $\mathbf{f} = (f(t_1), \ldots, f(t_n))^T$ = the hidden common signal,
- $\mathbf{g}_j = (g_j(t_1), \ldots, g_j(t_n))^T$ = individual age effect for tree j
- unexplained variability = a zero mean Gaussian vector with covariance $\sigma^2 \mathbf{I}_n$

Hierarchical Bayesian Model with three layers

$$
\begin{aligned}
[\text{process}, \text{parameters}|\text{data}] \quad \propto \quad & [\text{data}|\text{process}, \text{parameters}] \\
& \times [\text{process}|\text{parameters}] \\
& \times [\text{parameters}]
\end{aligned}
$$

Hierarchical Bayesian layers

Important statistical modeling questions

A) **Data layer** := [data|process, parameters]=

$$\mathbf{y}_j|\mathbf{g}_j, \mathbf{f}, \sigma^2 \sim \mathbf{g}_j + \mathbf{f} + \sigma^2 \mathcal{N}_n(\mathbf{0}_n, \mathbf{I}_n)$$

B) **Process layer :=** [process|parameters] =??

C) **Parameters layer (priors) :=** [parameters] =??

[process|parameters] = smoothing splines

Splines and BHM, Kimeldorf and Wahba (1970) and Wahba (1978)

$\mathbf{y} = \mathbf{f} + \sigma^2\mathcal{N}(\mathbf{0}, \mathbf{I})$ with improper Gaussian prior for the trend $\mathbf{f}$

$$\mathbf{f}|\tau^2 \sim \mathcal{N}_n(0, \tau^2\mathbf{K}^-)$$

where $\tau^2 = \sigma^2/\lambda$ and $\lambda \geq 0$ the classical smooth parameter that minimizes $\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int (f''(\mathbf{x}))^2 d\mathbf{x}$

Priors on variance components

Hastie (1990,2000) suggested to use proper inverse gamma priors $\sigma^2 \sim \mathcal{IG}(a_\sigma, b_\sigma)$ and $\tau^2 \sim \mathcal{IG}(a, b)$.

[process|parameters] = smoothing splines

Splines and BHM, Kimeldorf and Wahba (1970) and Wahba (1978)

$$\mathbf{f}|\tau_0^2 \sim \mathcal{N}_n(0, \tau_0^2\mathbf{K}^-) \text{ and } \mathbf{g}_j|\tau_j^2 \sim \mathcal{N}_n(0, \tau_j^2\mathbf{K}^-), \text{ for all } j = 1, \ldots, p.$$

Parameters layer (priors) := [**parameters**] =??

Variables changes

$$\phi_j = \frac{\sigma^2}{\tau_j^2 + \sigma^2}, \text{ for all } j = 0, \ldots, p.$$

If ϕ_j takes a value near one, then it means that the curve is very smooth.

Parameters layer (priors) := [**parameters**] $=??$

Variables changes

$$\phi_j = \frac{\sigma^2}{\tau_j^2 + \sigma^2}, \text{ for all } j = 0, \ldots, p.$$

If ϕ_j takes a value near one, then it means that the curve is very smooth.

Identifiability issues

- if all $\mathbf{g}_j$ proportional to **f**, it is impossible to distinguish **f** from $\mathbf{g}_j$
- the function **f** constrained to have a zero mean and unit variance (dimensionless)

Parameters layer (priors) := [**parameters**] =??

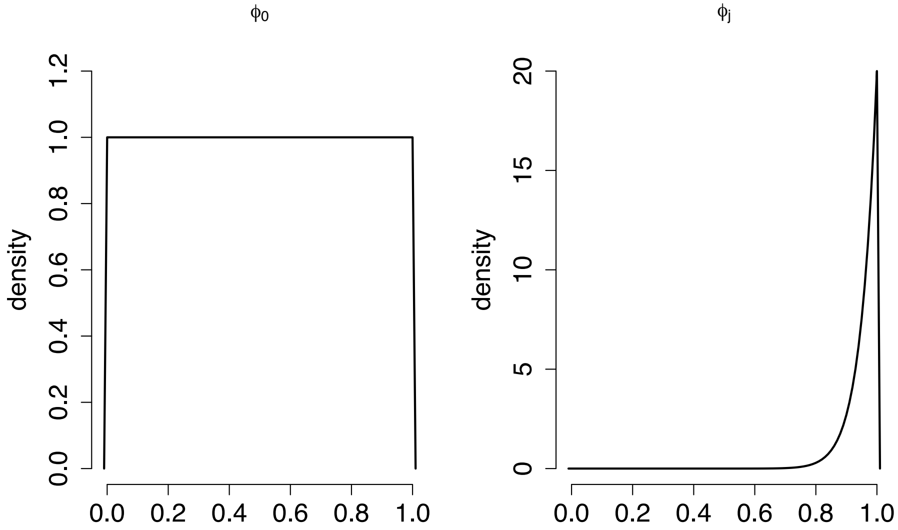
"Environmental information"

- the individual age effect function $\mathbf{g}_j$ should be very smooth because individual tree growth is a rather slow and cumulative process (Fang, 2010).
- the hidden signal shared by all trees $\mathbf{f}$ should capture environmental variabilities that correspond to rapid (yearly or decadal) or slow (centennial) changes.

Prior constraints

the frequency range of $\mathbf{g}_j$ is assumed to be much narrower than the one of $\mathbf{f}$.

Parameters layer (priors) := [parameters] $= \phi_j \sim$ Beta$(1, 1)$



Posteriors computations

Explicitly posterior distribution (Hastie, 2000)

$$\mathbf{f}|\mathbf{g}, \lambda_0 \mathbf{Y}, \sigma^2 \sim \mathcal{N}_n(\mathbf{B}(\mathbf{B}^T\mathbf{R}\mathbf{B} + \lambda_0 \boldsymbol{\Omega})^{-1}\mathbf{B}^T\mathbf{s}, \sigma^2\mathbf{B}(\mathbf{B}^T\mathbf{R}\mathbf{B} + \lambda_0\boldsymbol{\Omega})^{-1}\mathbf{B})$$

with

$$\mathbf{s} = \sum_{j=1}^{p}(\mathbf{y}_j - \mathbf{g}_j), \ \lambda_0 = \phi_0/(1-\phi_0), \ \mathbf{R} = \sum_{j=1}^{p}\mathbf{I}$$

and

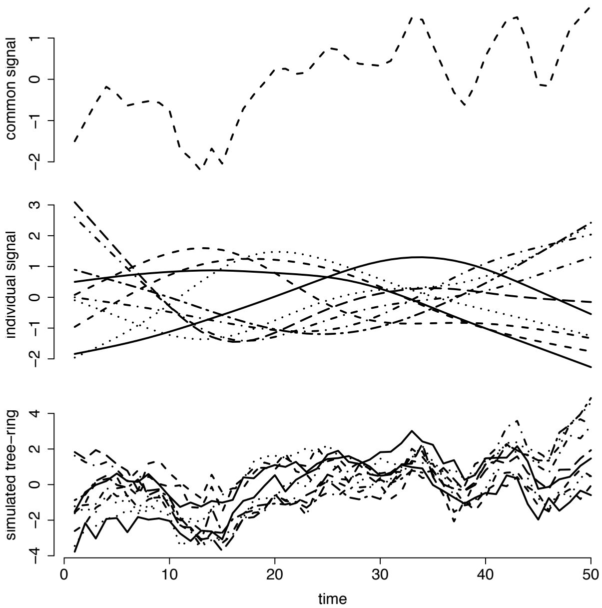
$$\mathbf{g}_j|, \mathbf{f}, \lambda_j \mathbf{y}_j, \sigma^2 \sim \mathcal{N}_n(\mathbf{B}(\mathbf{B}^T\mathbf{B} + \lambda_j \boldsymbol{\Omega})^{-1}\mathbf{B}^T\mathbf{d}, \sigma^2\mathbf{B}(\mathbf{B}^T\mathbf{B} + \lambda_j\boldsymbol{\Omega})^{-1}\mathbf{B})$$

with $\mathbf{d} = \mathbf{y}_j - \mathbf{f}$ and $\lambda_j = \phi_j/(1-\phi_j)$. It is also possible to show that σ^2 have an inverse gamma posterior distribution.

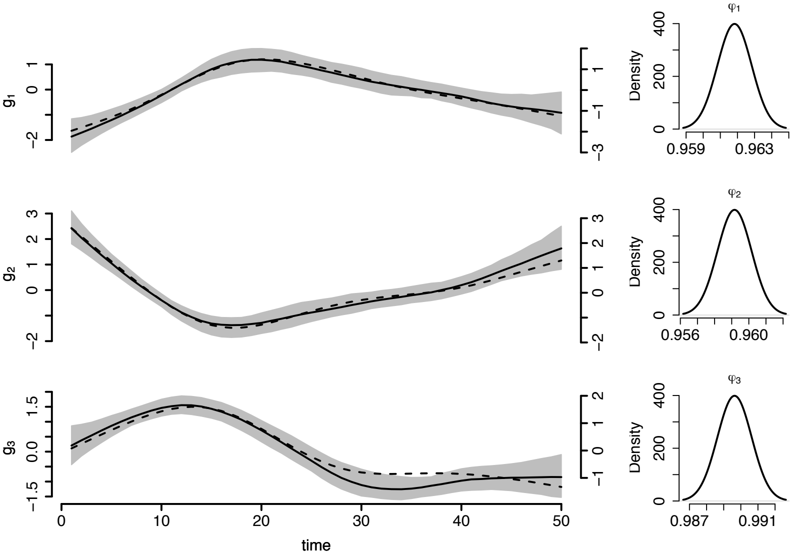
Gibbs and MH sampler

The parameters ϕ_0 and ϕ_j don't have standard posterior distributions so we use Metropolis-Hasting algorithm to estimate them.

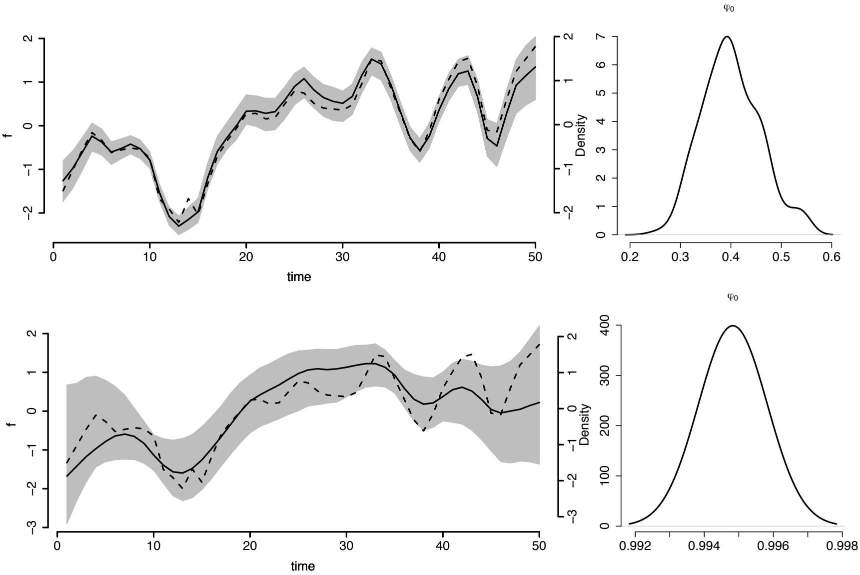
Simulations : finding f and g_j from the y_j's (low panel)



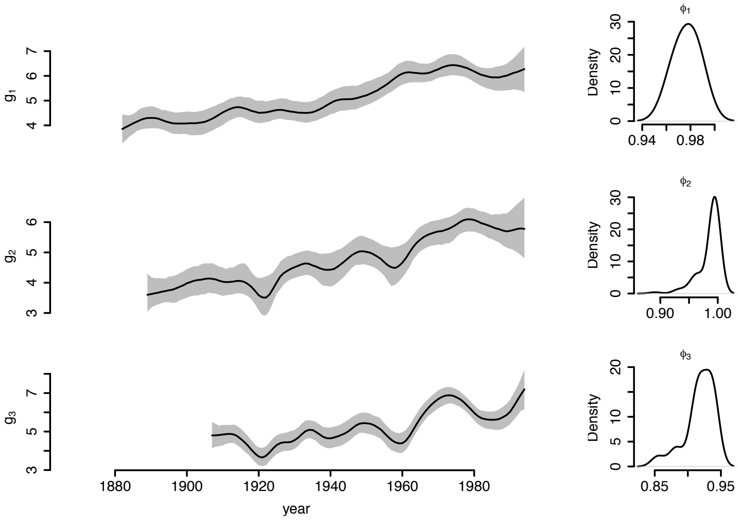
Simulations posteriors



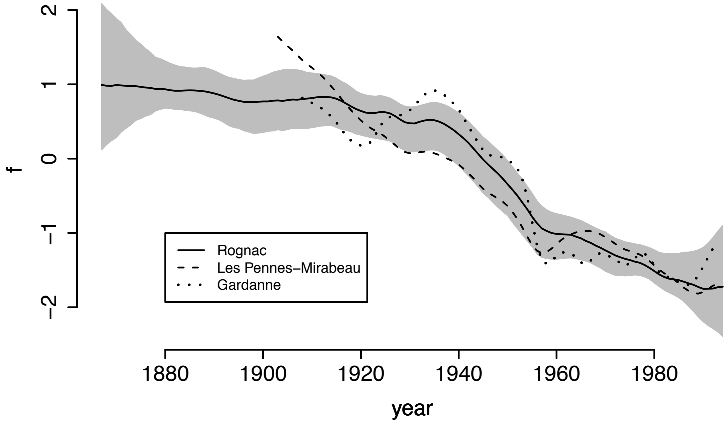
Simulations posteriors (noise variance 0.1 (top) and 0.5 (bottom))



The seventeen tree ring width logarithms



The seventeen tree ring width logarithms



Take-home messages from this dendro example

Positive points

- Outputs are probability distribution (i.e., easy to compute CI)
- Extract signal distribution is independently found from covariates like precip or temperatures
- Possibility to include more dynamical equations
- Package in R (upon request)

Drawbacks

- Only one site but a bigger set is under study
- Choice of the priors important (but is this a minus ?)

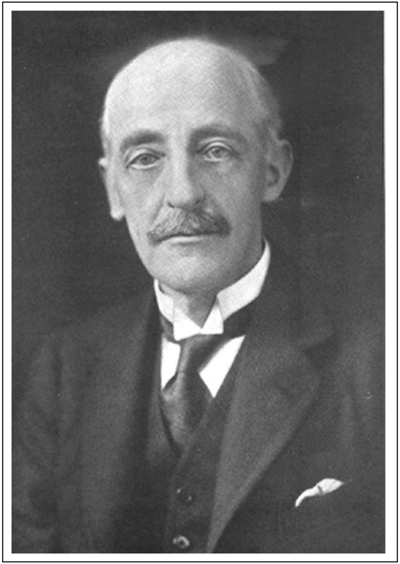
A few references

- Guin O., Naveau P. and Boreux J.J. (2012). Extracting a common signal in tree ring widths with a semi-parametric Bayesian hierarchical model (in revision).
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). *Bayesian Data Analysis, 2nd ed.* Chapman & Hall.
- Hooten M.B. and Wikle C.K. (2007). Shifts in the spatio-temporal growth dynamics of shortleaf pine. *Environmental and Ecological Statistics.* 14 :3.

Statistics and Earth sciences

"There is, today, always a risk that specialists in two subjects, using languages full of words that are unintelligible without study, will grow up not only, without knowledge of each other's work, but also will ignore the problems which require mutual assistance".

Sir Gilbert T. Walker
(Walker, 1927b, page 321)



A very short biblio

- J. Anderson, An ensemble adjustment Kalman filter for data assimilation, *Mont. Weath. R.*, **129**, 2001.
- T. Bengtsson, D. Nychka, C. Snyder A frame work for data assimilation and forecasting in high-dimensional non-linear dynamical systems. (2002).
- Doucet A., N. Freitas, N. Gordon, (2001). Sequential Monte Carlo Methods in Practice.
- Wikle, C.K. and N. Cressie, (1999) A dimension reduced approach to space-time Kalman filtering. *Biometrika* , 86 , 815-829.

Kalman filters : bringing the GPD

Proposition 4.13. *Inverse conditional uni variate cdf's for centered conditional variables* $\left((X_{t,i}|y_{1:t}) - \mu_{t+1}^{\theta_i}\right)$ *are:*

For $\xi > 0$

$$F^{-1}_{X_{t,i}|y_{1:t}}(u) = \begin{cases} \sqrt{\frac{\tilde{\sigma}_t \Sigma_{X_t,[ii]}}{\tilde{\xi}} \frac{qbeta_{\frac{1}{2},\frac{1}{\tilde{\xi}_t}-\frac{1}{2}}(2u-1)}{1-qbeta_{\frac{1}{2},\frac{1}{\tilde{\xi}}-\frac{1}{2}}(2u-1)}} & if\ u \geq \frac{1}{2} \\ -F^{-1}_{X_{t,i}|y_{1:t}}(1-u) & if\ u < \frac{1}{2} \end{cases}$$

For $\xi < 0$*:*

$$F^{-1}_{X_{ti}|y_{1:t}}(u) = \begin{cases} \sqrt{\frac{\tilde{\sigma}_t \Sigma_{X_t,[ii]}}{-\tilde{\xi}_t} qbeta_{\frac{1}{2},-\frac{1}{\tilde{\xi}_t}+1}(2u-1)} & if\ u \geq \frac{1}{2} \\ -F^{-1}_{X_{t;i}|y_{1:t}}(1-u) & if\ u < \frac{1}{2} \end{cases}$$

with $\alpha = \frac{nT+p-t(n-p)-1}{2}$, $\tilde{\sigma}_t = \frac{\sigma+\xi Q_t(y_{1:t})}{1-\alpha\xi}$ $\tilde{\xi}_t = \frac{\xi}{1-\alpha\xi}$

- Elliptical innovations $\epsilon_t = (X_t - FX_{t-1} | x_{t-1})$ and $\nu_t = (Y_t - GX_t | x_t)$ with

$$\begin{aligned} \epsilon_t &\sim \mathcal{E}\left(0, \Sigma^{\epsilon}, g^{\epsilon}_{t,x_{t-1}}\right) \\ \nu_t &\sim \mathcal{E}\left(0, \Sigma^{\nu}, g^{\nu}_{t,x_t}\right) \end{aligned}$$

- Finite time process : $t \in \{0 : T\}$
- Elliptical global vector

$$\begin{aligned} W &= \left(X'_0, X'_1 \ldots X'_T, Y'_1 \ldots Y'_T\right)' \\ &\sim \mathcal{E}_{nT+p}\left(0, \Sigma^W, g^W\right) \end{aligned}$$

- Result : Equations for estimates $\hat{x}_t$ and $\hat{\Sigma}_{X_t}$ are similar to those of the gaussian filter. Additional equations for conditional generators.

Comparing “one-fits-all” with the 17 g_j