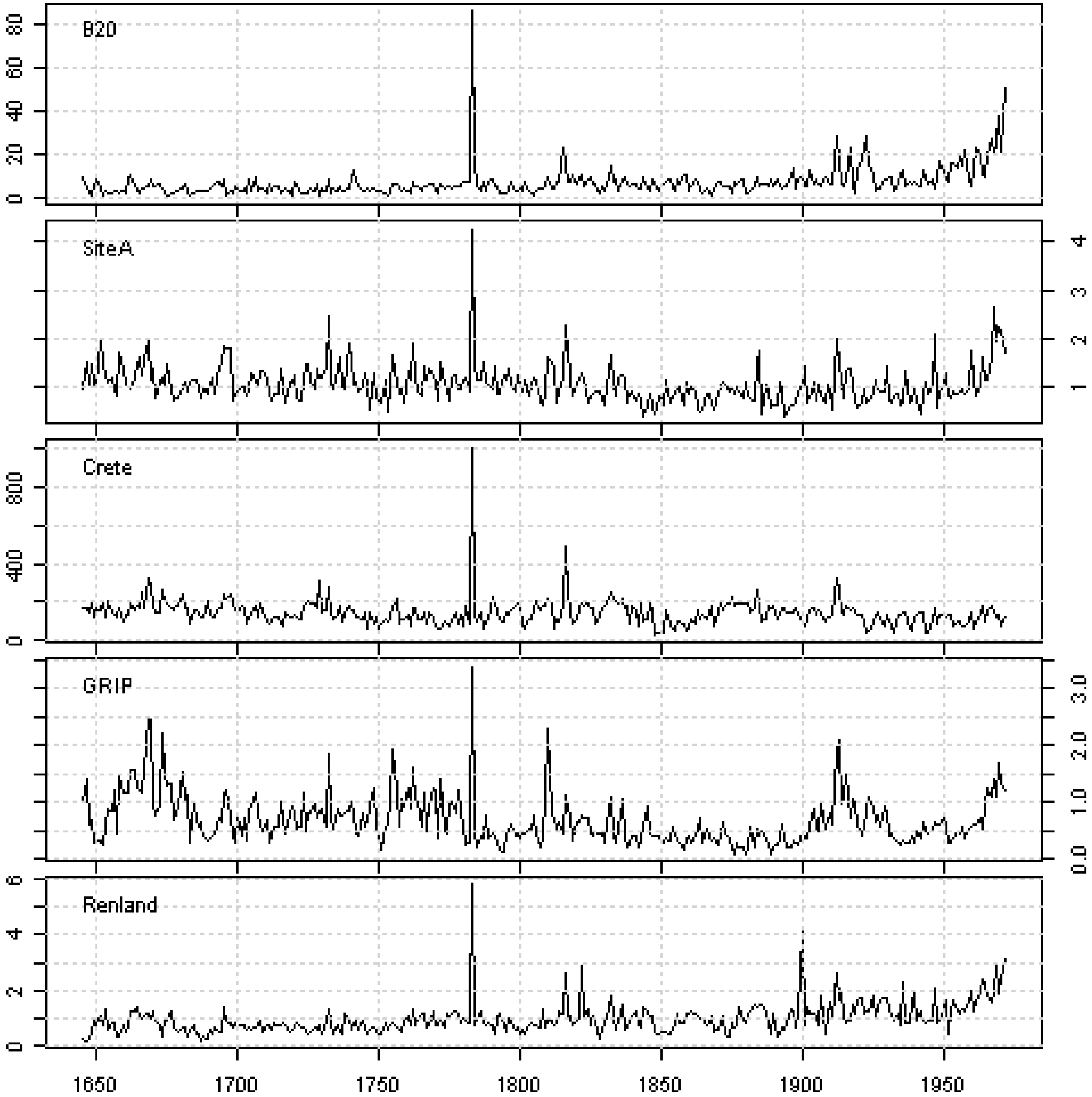
Extracting a Common Pulse Like Signal from Time series

J.Gazeaux – SA/IPSL-CNRS
D.Batista – University of Colorado
CM.Ammann – NCAR Boulder
P.Naveau – LSCE/IPSL- CNRS
C.Jégat - ENSM
C.Gao – Rutgers University

Colloque Nationale d'Assimilation, December, 2nd 2008

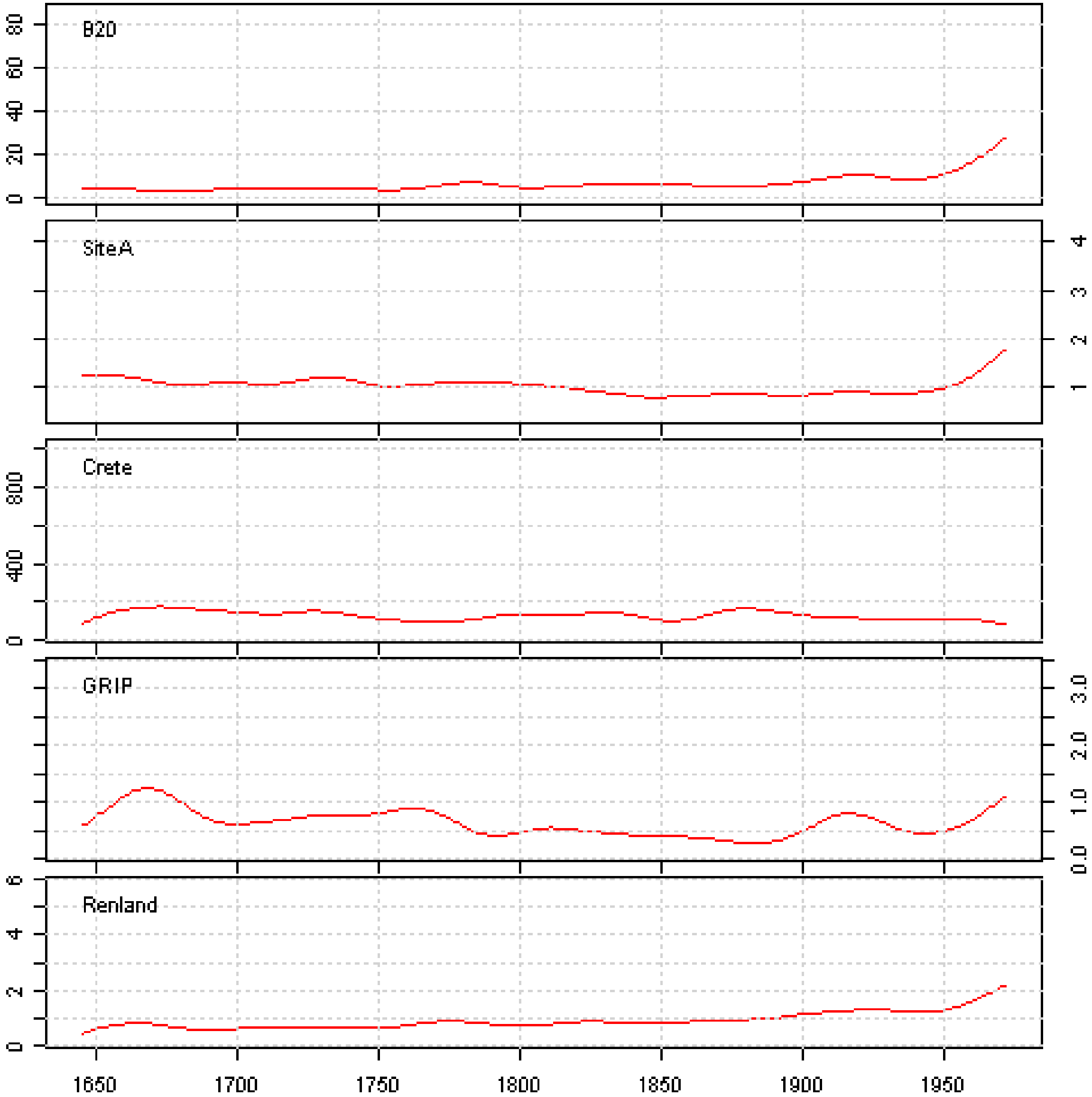
Motivations

Time Series over 1645 to 1990 of Sufate from Ice Core drilled in Greenland in 5 sites



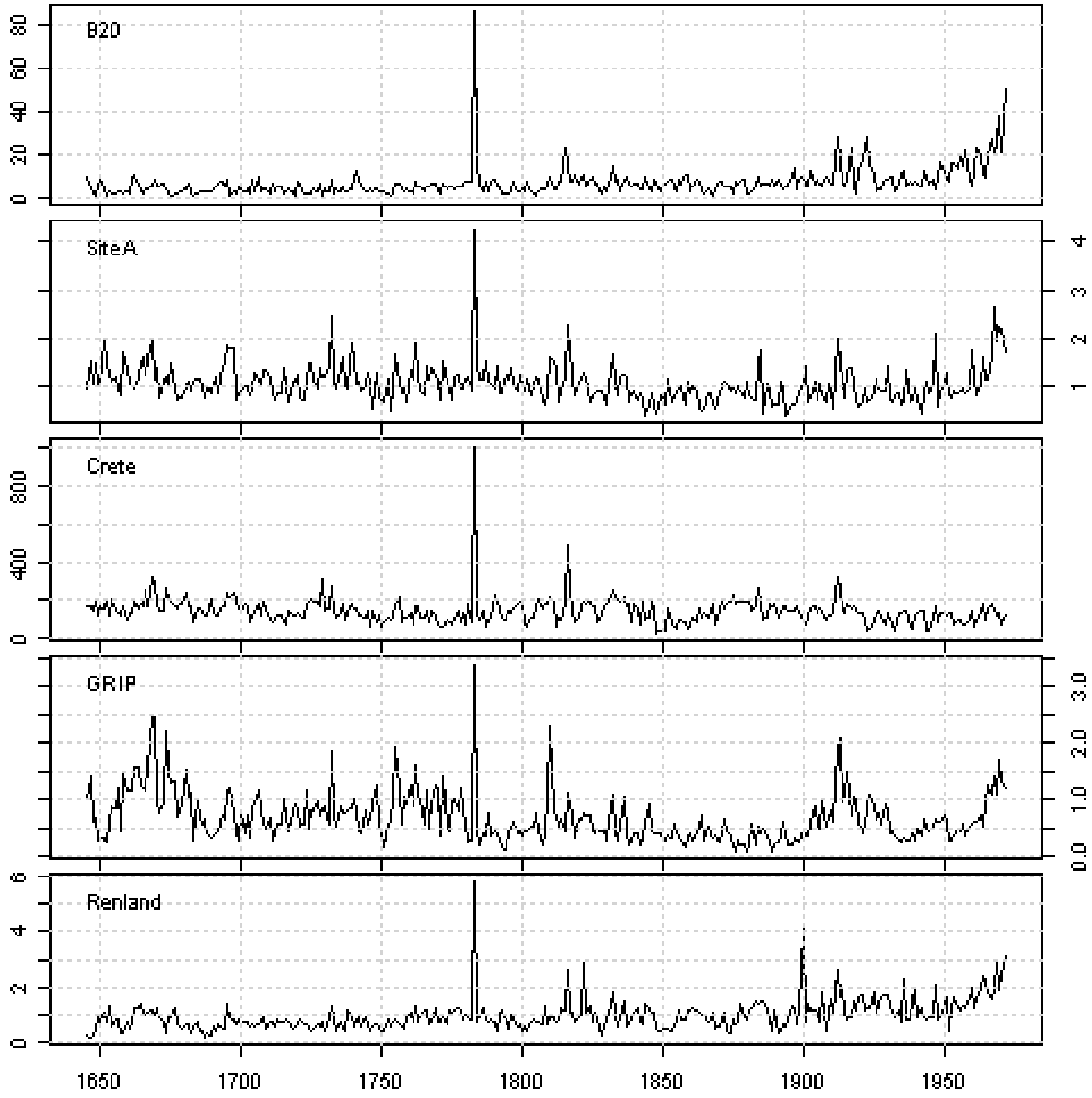
Motivations

Time Series over 1645 to 1990 of Sufate from Ice Core drilled in Greenland in 5 sites



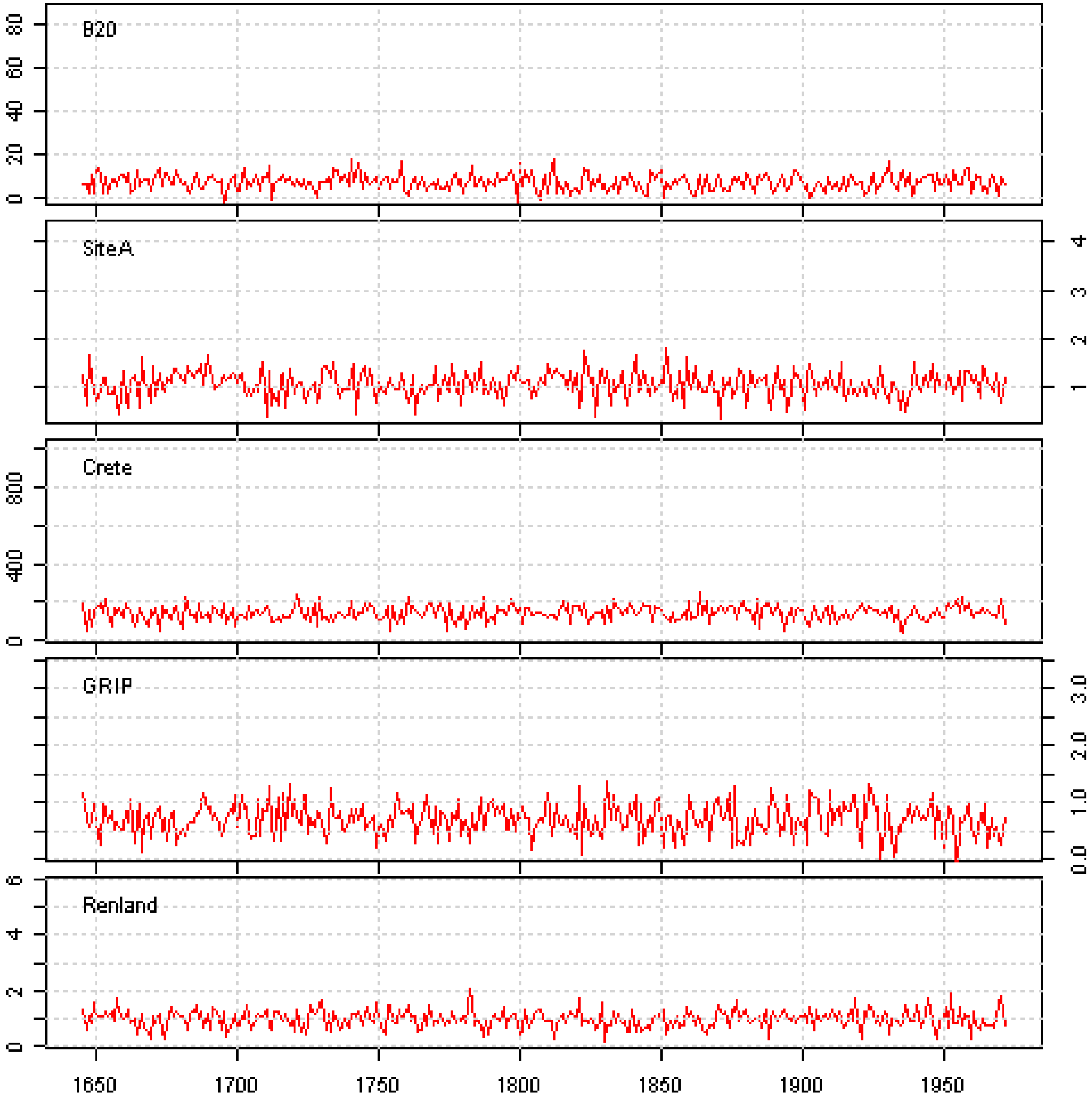
Motivations

Time Series over 1645 to 1990 of Sufate from Ice Core drilled in Greenland in 5 sites



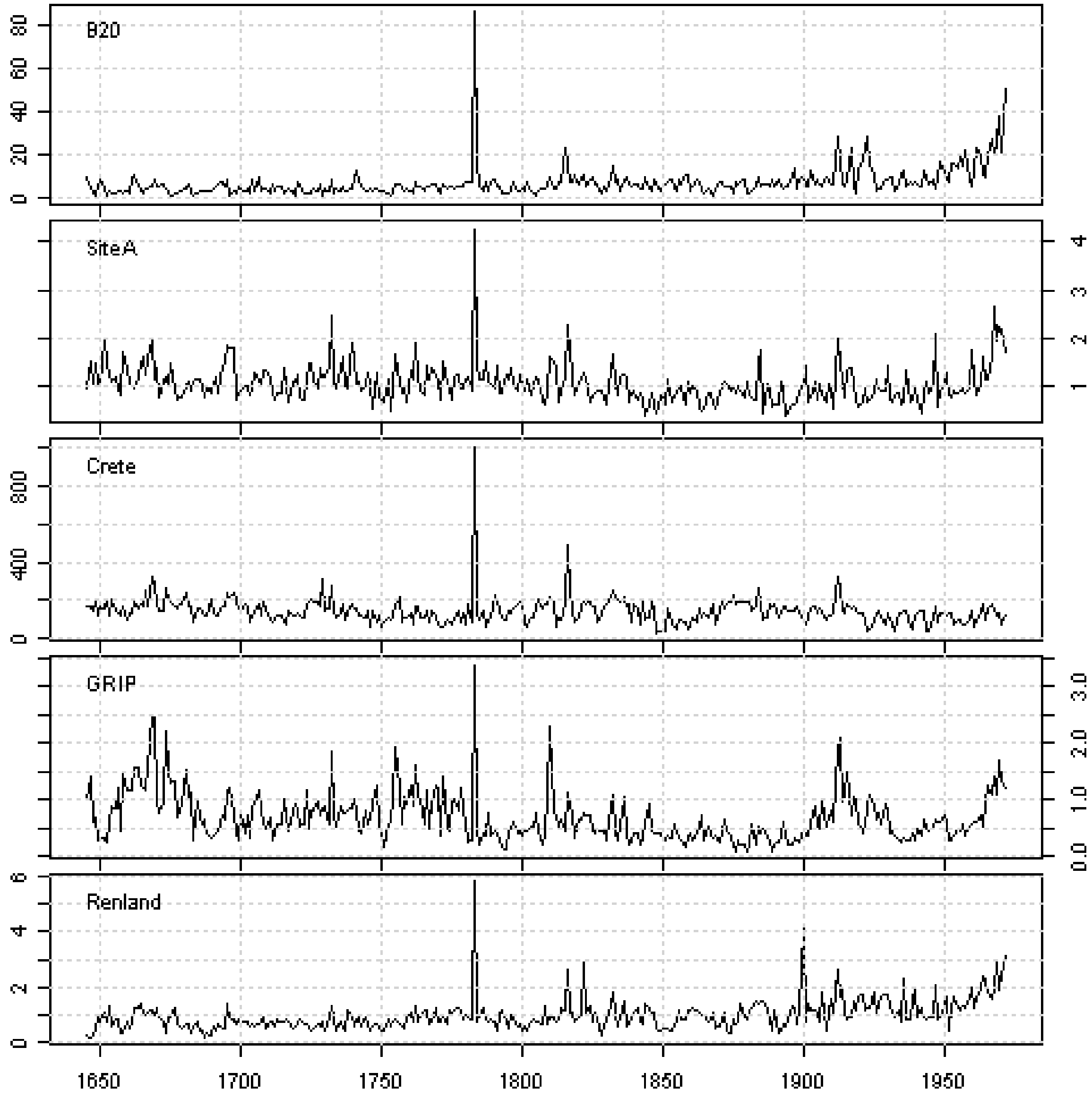
Motivations

Time Series over 1645 to 1990 of Sufate from Ice Core drilled in Greenland in 5 sites



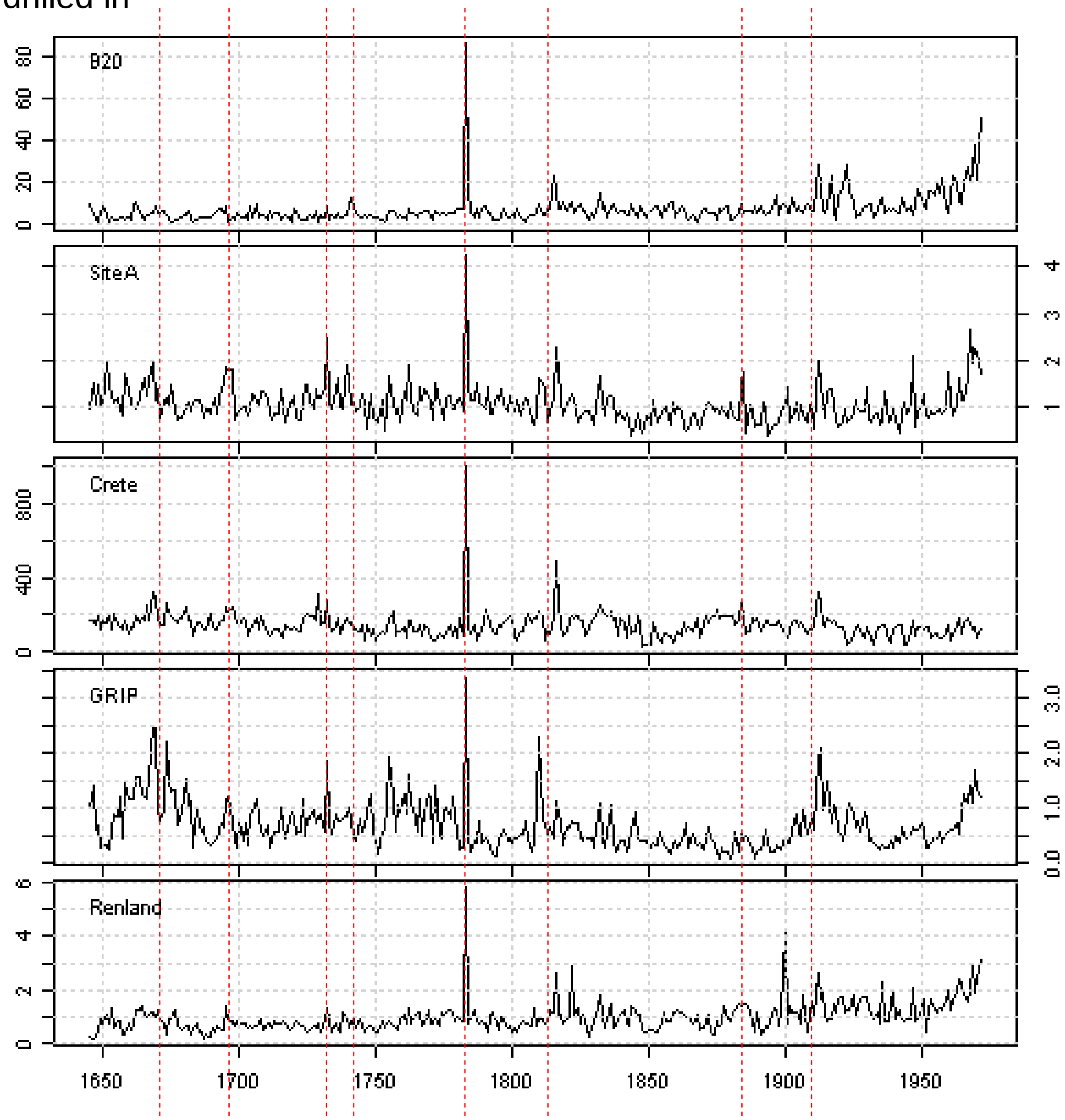
Motivations

Time Series over 1645 to 1990 of Sufate from Ice Core drilled in Greenland in 5 sites



Motivations

Time Series over 1645 to 1990 of Sufate from Ice Core drilled in Greenland in 5 sites



References of this work

- Decomposition methods of Time Series *([Wecker and Ansley, 1998])*

- Previous work : iterative method *([Robock and Free, 1995; Gao et al.,2008*), stationnary case, trend issue

- Guo : A signal extraction approach to modeling hormone time series with pulses and a changing baseline (1998) *(univariate)*

State-Space model and Non-Linear Kalman Filter

$$y_j(t) = \beta_j\, x(t) + f_j(t) + \epsilon_j(t) \quad \text{for } j = 1, \ldots, J \text{ and } t = 1, \ldots, T. \qquad (1)$$

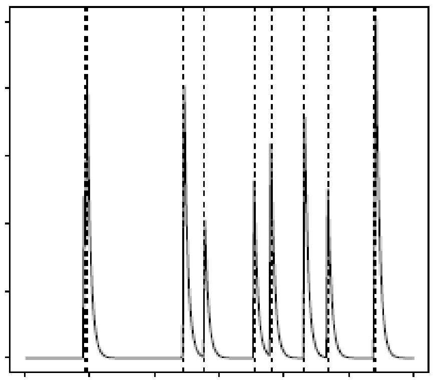
Parameters to estimate : β_j, …

State-Space model and Non-Linear Kalman Filter

$$y_j(t) = \beta_j\, x(t) + f_j(t) + \epsilon_j(t) \quad \text{for } j = 1, \ldots, J \tag{1}$$

$$x(t) = \alpha\, x(t-1) + v(t), \ \text{for } t = 1, \ldots, T \tag{2}$$

v : hidden signal : $$v(t) = \begin{cases} N(\mu_v, \sigma_v^2) & \text{, if } I_t = 1, \\ 0 & \text{, if } I_t = 0. \end{cases} , \tag{3}$$



Parameters to estimate : β_j, μ_v, σ_v, α, π,…

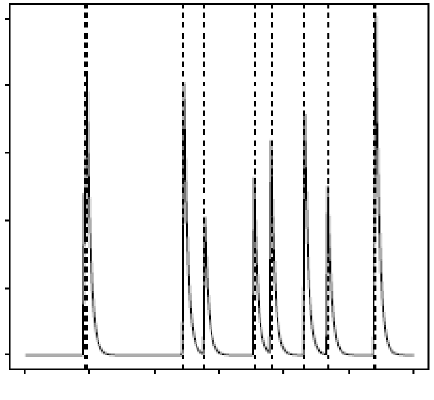
State-Space model and Non-Linear Kalman Filter

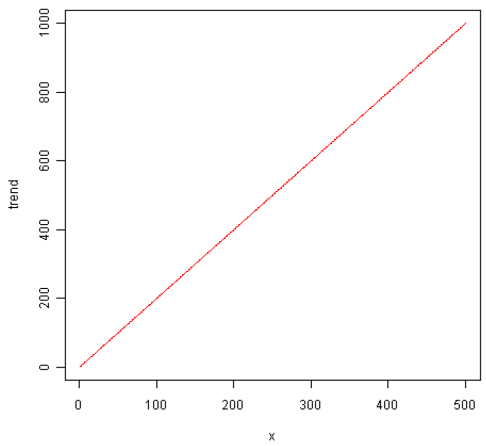
$$y_j(t) = \beta_j\, x(t) + f_j(t) + \epsilon_j(t) \quad \text{for } j = 1, \ldots, J \tag{1}$$

$$x(t) = \alpha\, x(t-1) + v(t), \text{ for } t = 1, \ldots, T \tag{2}$$

v : hidden signal : $v(t) = \begin{cases} N(\mu_v, \sigma_v^2) & \text{, if } I_t = 1, \\ 0 & \text{, if } I_t = 0. \end{cases}$ (3)

f : smoothing spline and its state space model $\mathbf{F_j}(t) = B\mathbf{F_j}(t-1) + \mathbf{E}$ (4)





Parameters to estimate : β_j, μ_v, σ_v, α, π,…

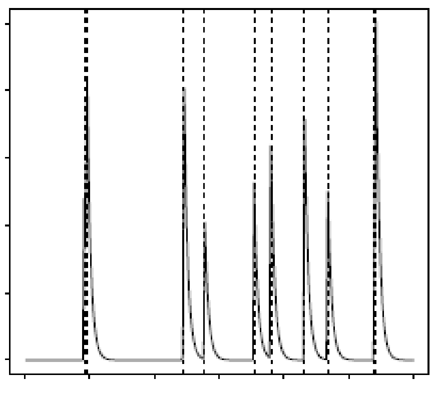
State-Space model and Non-Linear Kalman Filter

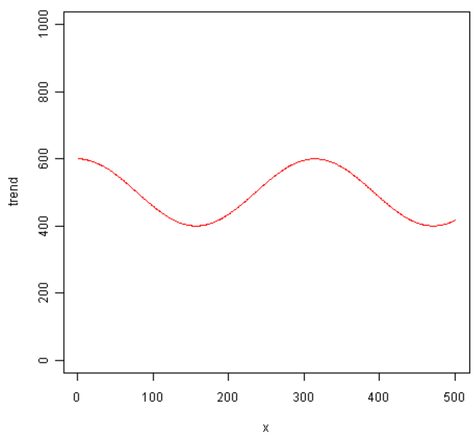
$$y_j(t) = \beta_j\, x(t) + f_j(t) + \epsilon_j(t) \quad \text{for } j = 1, \ldots, J \tag{1}$$

$$x(t) = \alpha\, x(t-1) + v(t), \ \text{for } t = 1, \ldots, T \tag{2}$$

v : hidden signal : $$v(t) = \begin{cases} N(\mu_v, \sigma_v^2) & , \text{if } I_t = 1, \\ 0 & , \text{if } I_t = 0. \end{cases} , \tag{3}$$

f : smoothing spline and its state space model $$\mathbf{F_j}(t) = B\mathbf{F_j}(t-1) + \mathbf{E}. \tag{4}$$





Parameters to estimate : β_j, μ_v, б$_v$, α, π,…

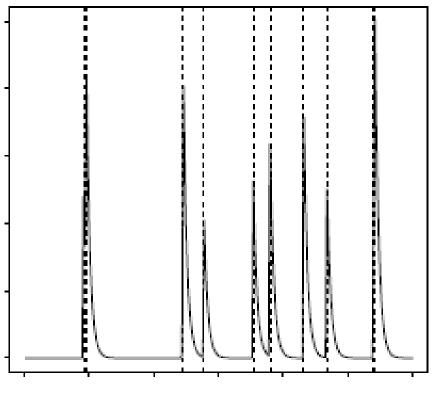
State-Space model and Non-Linear Kalman Filter

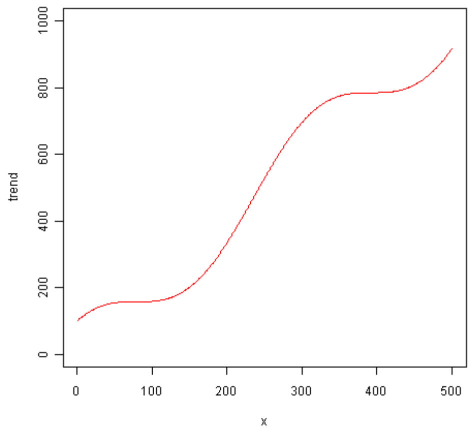
$$y_j(t) = \beta_j\, x(t) + f_j(t) + \epsilon_j(t) \quad \text{for } j = 1, \ldots, J \tag{1}$$

$$x(t) = \alpha\, x(t-1) + v(t), \ \text{for } t = 1, \ldots, T \tag{2}$$

v : hidden signal : $v(t) = \begin{cases} N(\mu_v, \sigma_v^2) & \text{, if } I_t = 1, \\ 0 & \text{, if } I_t = 0. \end{cases}$, (3)

f : smoothing spline and its state space model $\mathbf{F_j}(t) = B\mathbf{F_j}(t-1) + \mathbf{E}$. (4)





Parameters to estimate : β_j, μ_v, б$_v$, α, π,…

State-Space model and Non-Linear Kalman Filter

$$y_j(t) = \beta_j\, x(t) + f_j(t) + \epsilon_j(t) \quad \text{for } j = 1, \ldots, J \tag{1}$$

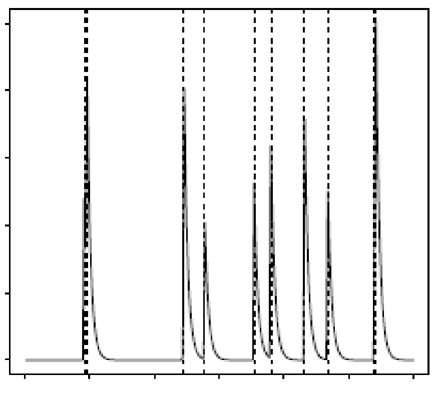
$$x(t) = \alpha\, x(t-1) + v(t), \quad \text{for } t = 1, \ldots, T \tag{2}$$

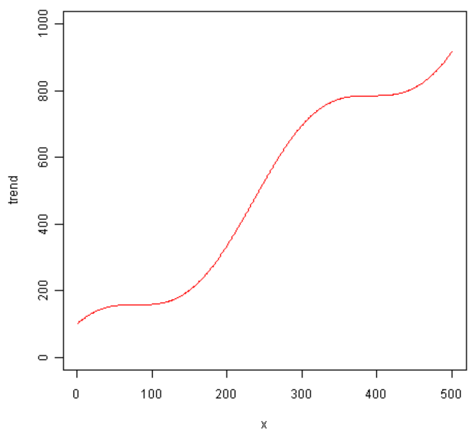
v : hidden signal :

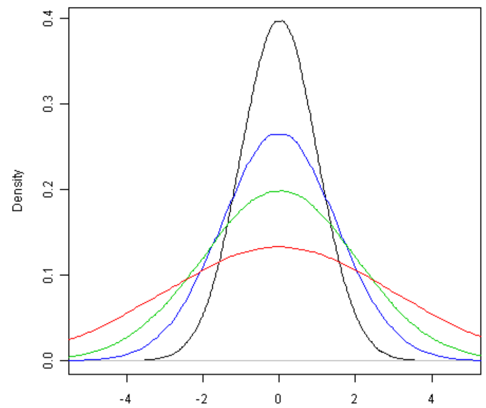
$$v(t) = \begin{cases} N(\mu_v, \sigma_v^2) & , \text{if } I_t = 1, \\ 0 & , \text{if } I_t = 0. \end{cases} , \tag{3}$$

f : smoothing spline and its state space model

$$\mathbf{F_j}(t) = B\mathbf{F_j}(t-1) + \mathbf{E} \tag{4}$$







Parameters to estimate : β_j, μ_v, б$_v$, α, π, б$_e$

State-Space model and Non-Linear Kalman Filter

$$y_j(t) = \beta_j\, x(t) + f_j(t) + \epsilon_j(t) \quad \text{for } j = 1, \ldots, J \tag{1}$$

$$x(t) = \alpha\, x(t-1) + v(t), \text{ for } t = 1, \ldots, T \tag{2}$$

v : hidden signal : $v(t) = \begin{cases} N(\mu_v, \sigma_v^2) & \text{, if } I_t = 1, \\ 0 & \text{, if } I_t = 0. \end{cases}$, (3)

f : smoothing spline / state space model : $\mathbf{F_j}(t) = B\mathbf{F_j}(t-1) + \mathbf{E}_{f_j}(t),$ (4)

$$Y_t = H X_t + E_t, \tag{5}$$

$$X_t = \Phi X_{t-1} + E_t^* \tag{6}$$

β_j, μ_v, $б_v$, α, π, $б_e$

$$X_t = (v(t), x(t), \mathbf{F_1}(t), \ldots, \mathbf{F_J}(t))^T$$

$$Y_t = (y_1(t), \ldots, y_J(t))$$

$$E_t = (\epsilon_1, \epsilon_2, \ldots, \epsilon_J)^T \text{ and } E_t^* = (0, 0, E_{f_1}^T(t), \ldots, E_{f_J}^T(t))^T$$

Parameter Estimation

- Estimation method of μ_v, $б_v$, α, π , $б_e$, β_j :

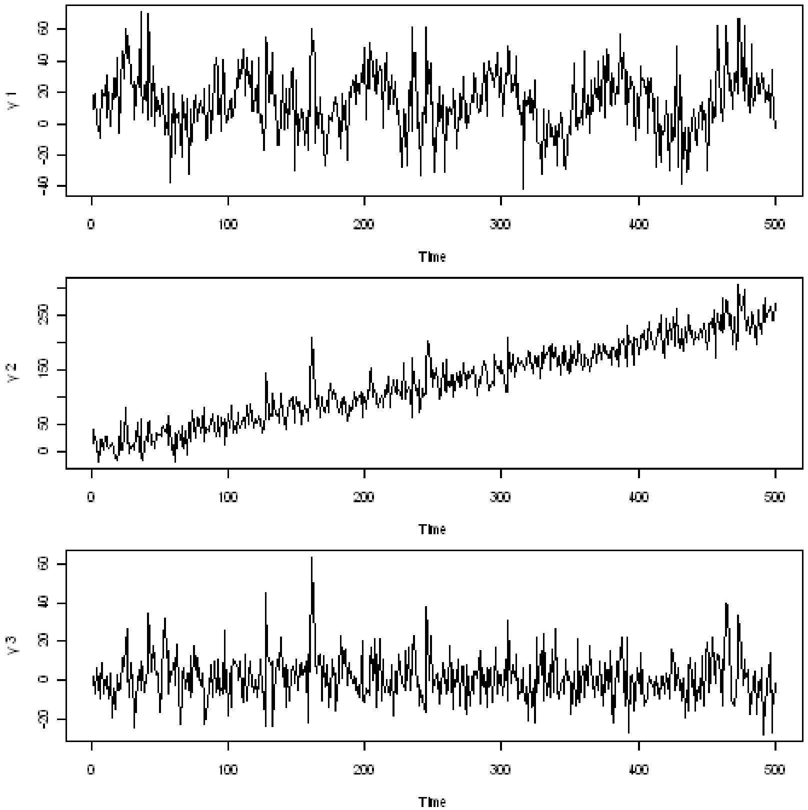
A priori : Rough Estimates

A posteriori : Maximum Likelihood Estimation of the parameters

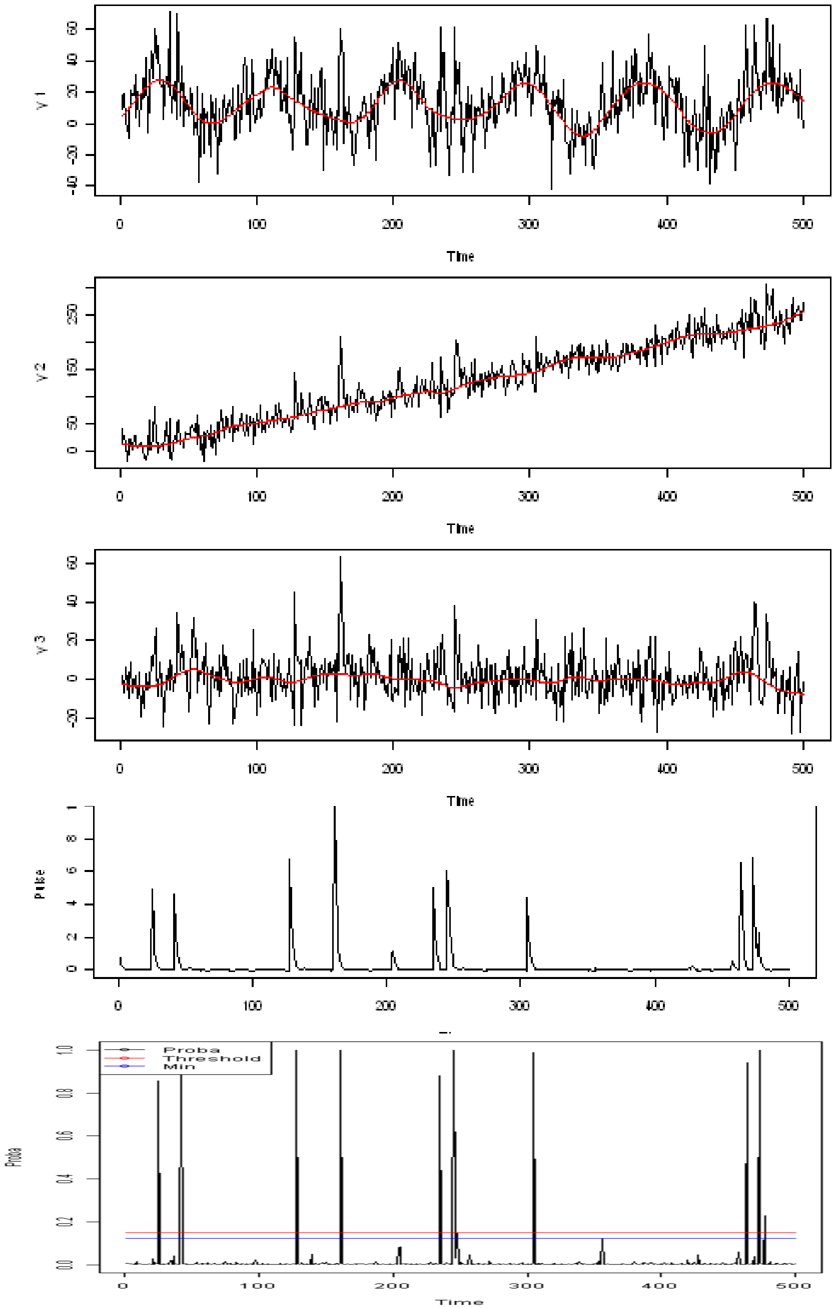
- FIS : Fixed Intervall Smoother

β_j, μ_v, $б_v$, α, π, $б_e$

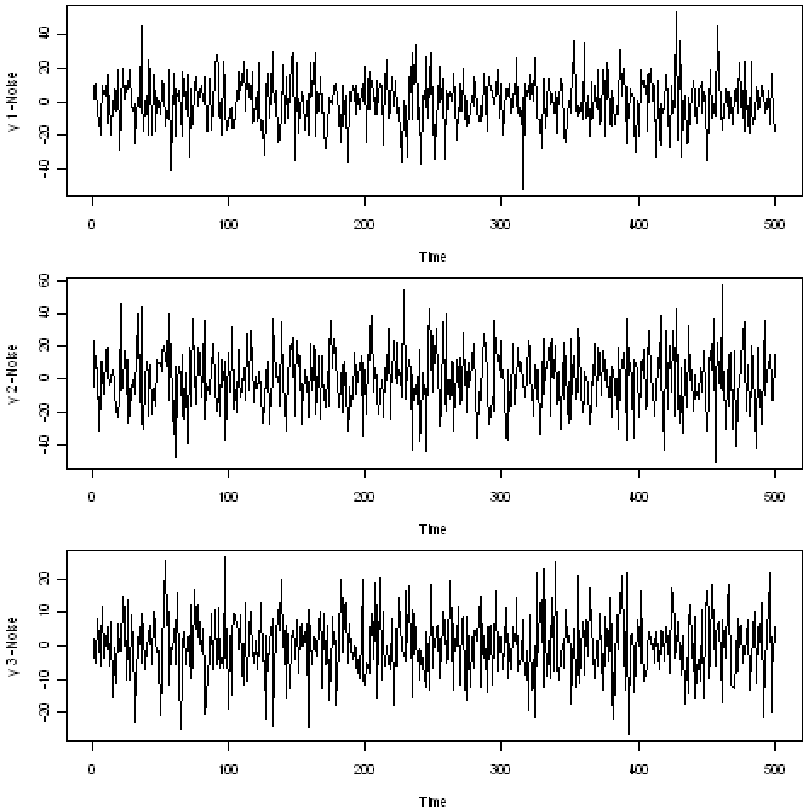
Input /

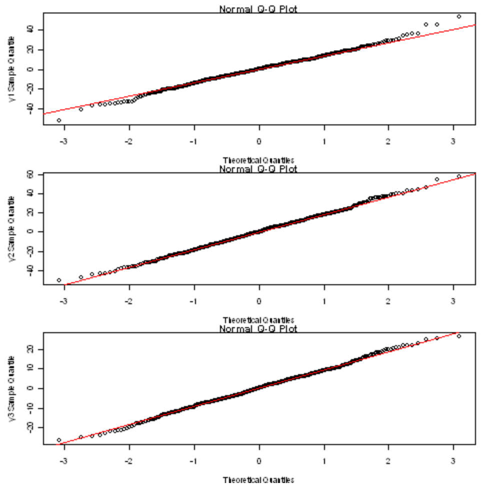


Input / Output



Input / Output



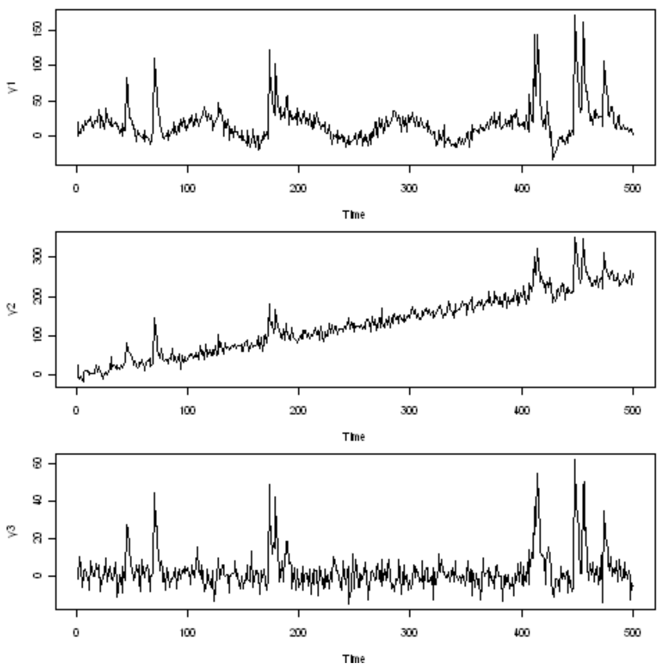


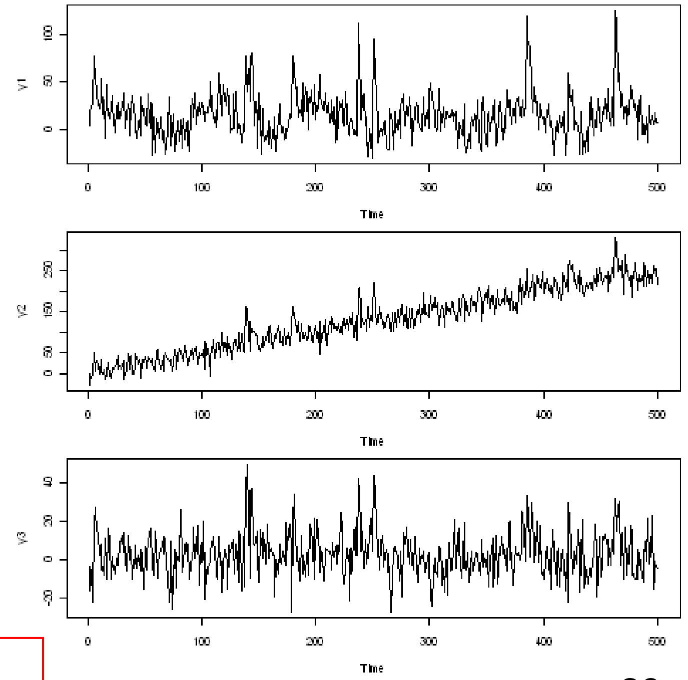
Monte Carlo Method to test the robustness of the model

6 different simulations

means 6 parameter sets tested

(400 time series over 500 time steps each)



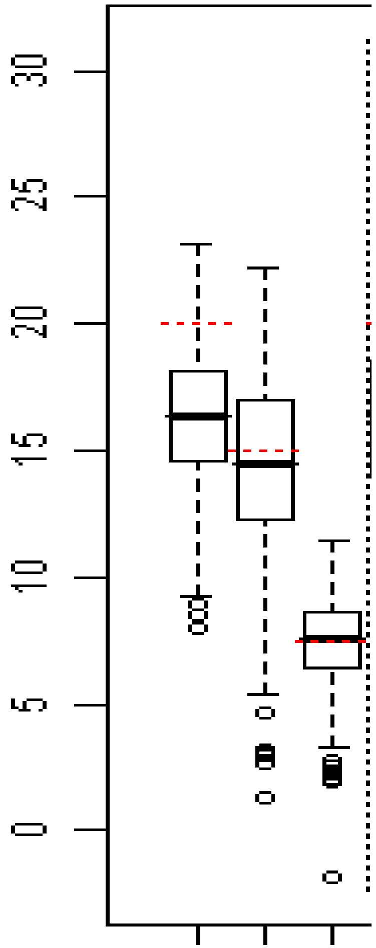


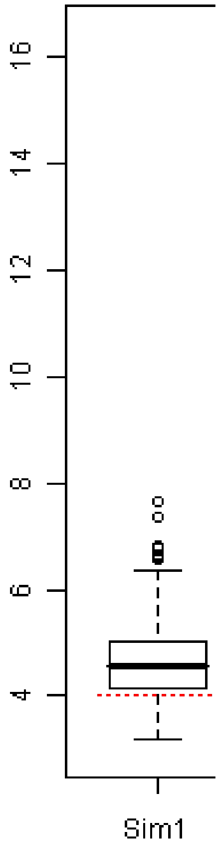
β_j, μ_v, б$_v$, α, π, б$_e$

Boxplots of the parameters

$$y_j(t) = \beta_j\, x(t) + f_j(t) + \epsilon_j(t) \quad \text{for } j = 1, \ldots, J \text{ and } t = 1, \ldots, T. \tag{1}$$

β_j

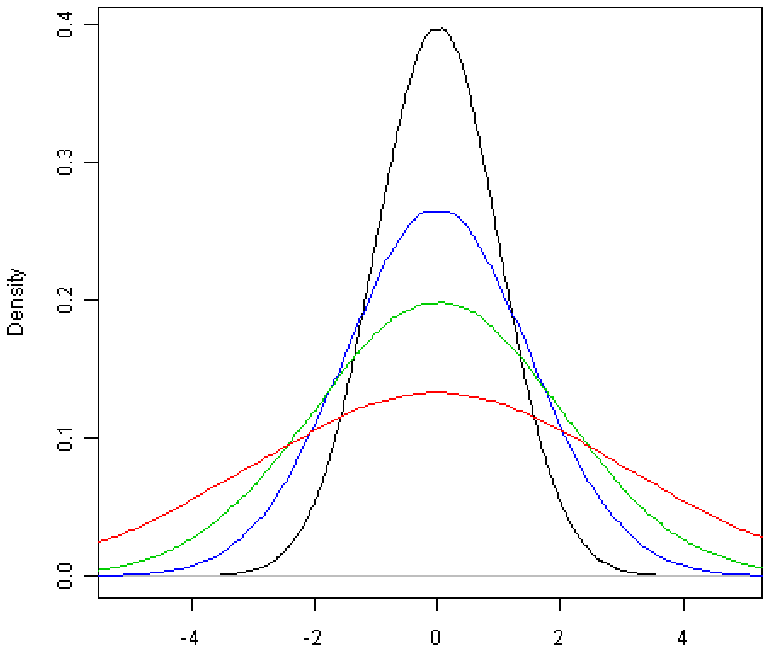


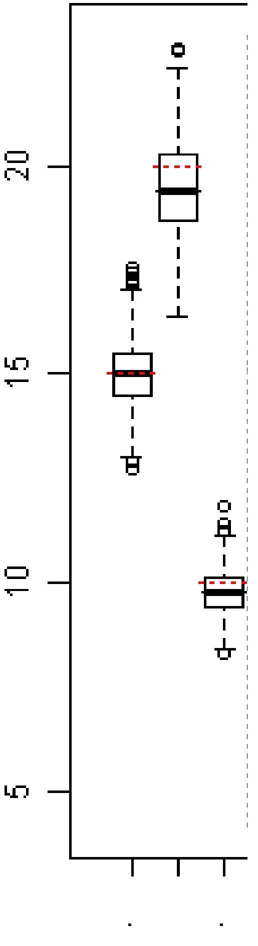


μ_v

Boxplots of the parameters

$$y_j(t) = \beta_j \, x(t) + f_j(t) + \epsilon_j(t) \text{ for } j = 1, \ldots, J \qquad (1)$$





σ_e

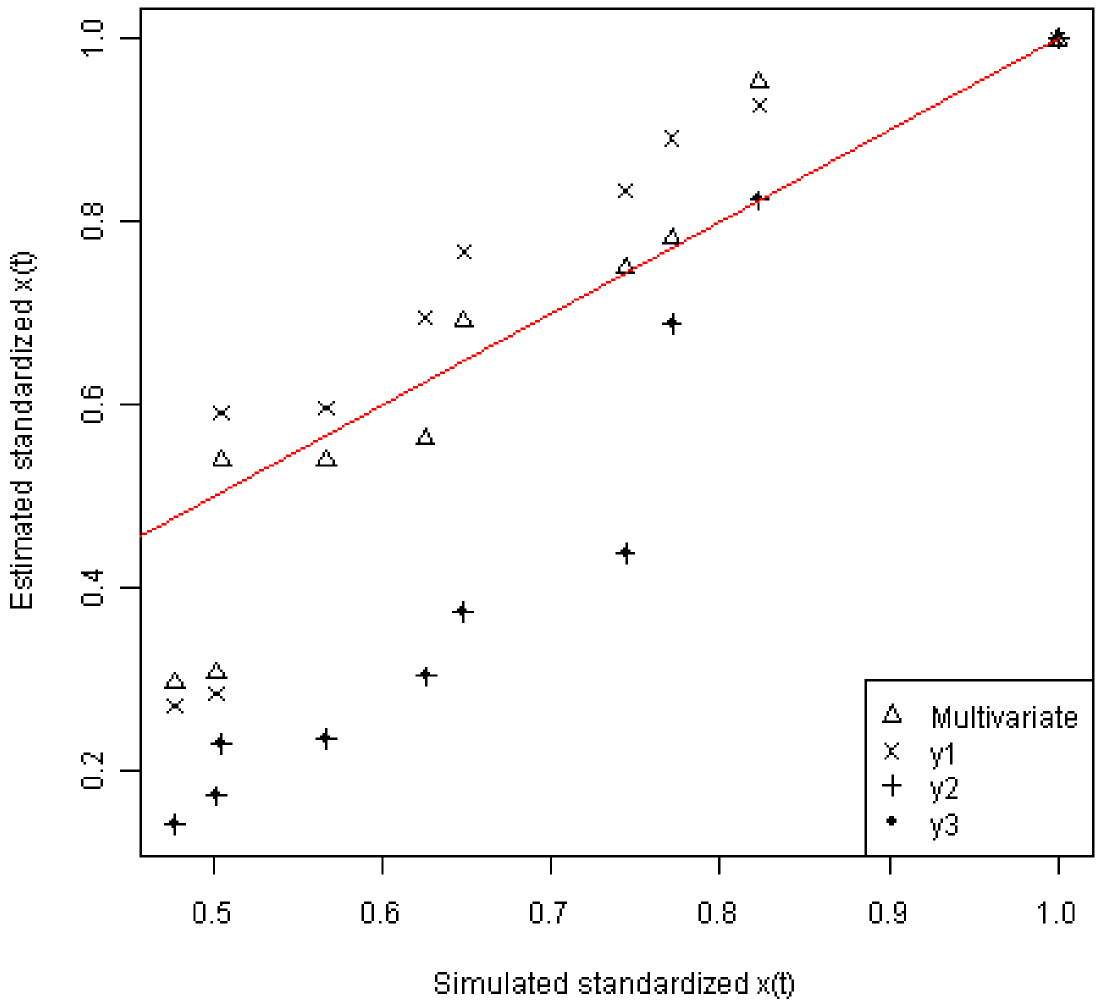
Improvement compare to univariate detection

- Mean of First and Second Type Errors on Simulation Time Series :

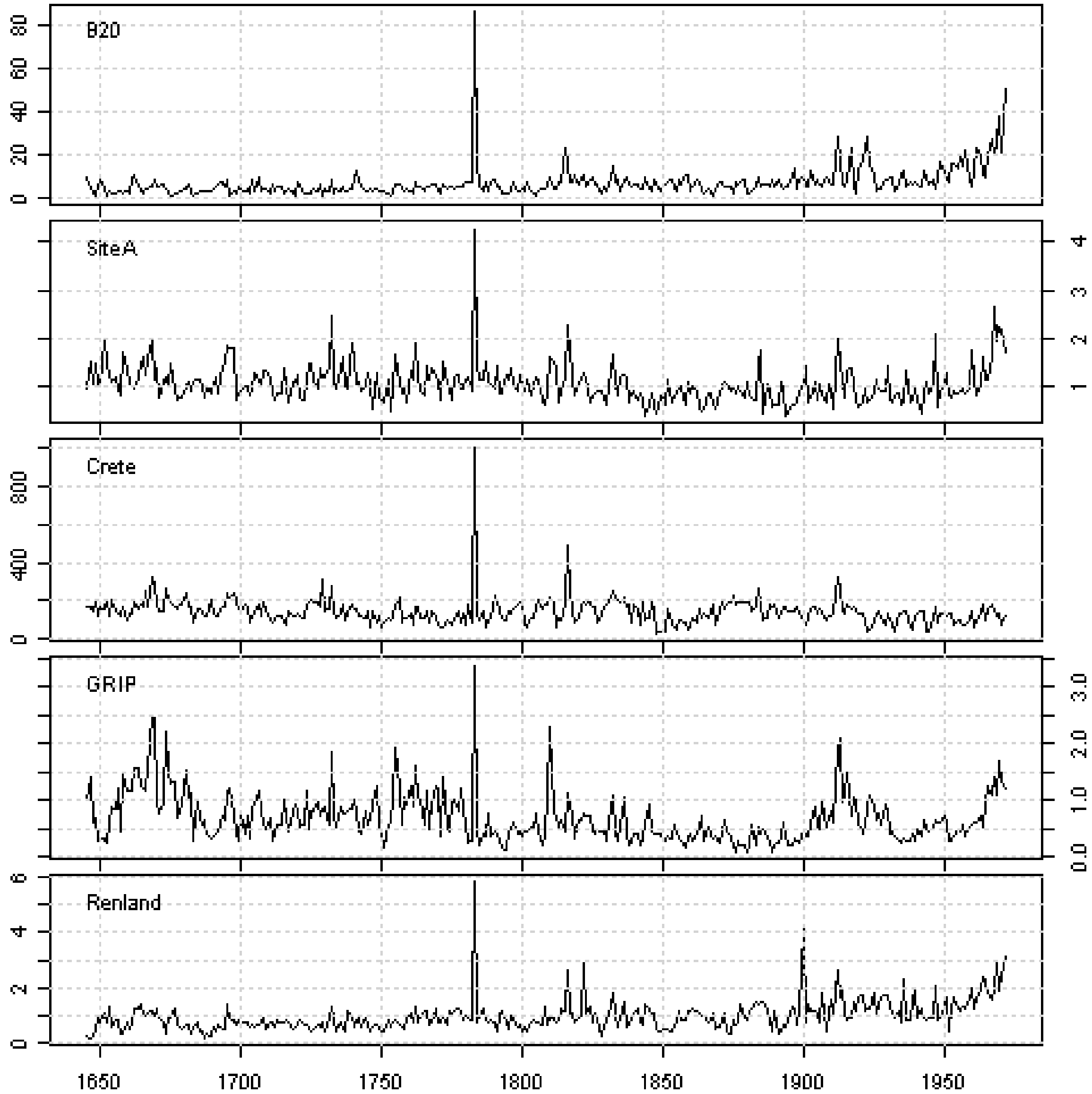
Sim1	Univariate	Multivariate
First Type Error	1.6 →	1.3
Second Type Error	6.3	1.4

Sim4	Univariate	Multivariate
First Type Error	0.14	0.10
Second Type Error	7.6	1.5

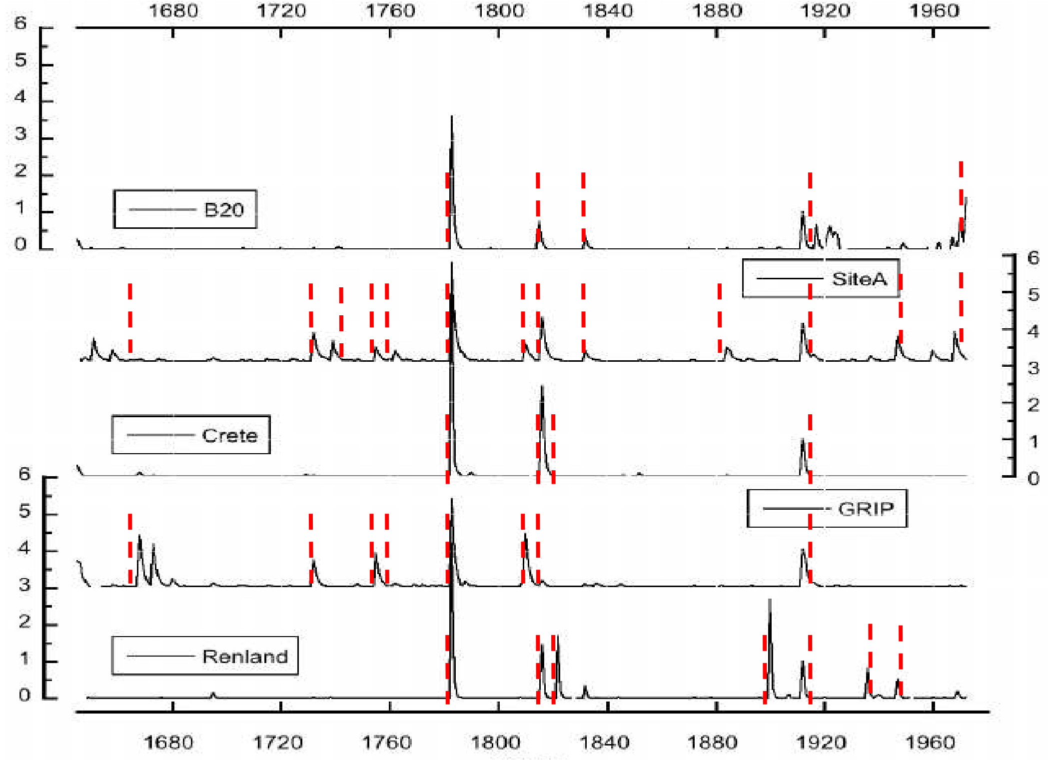
Multivariate vs Univariate



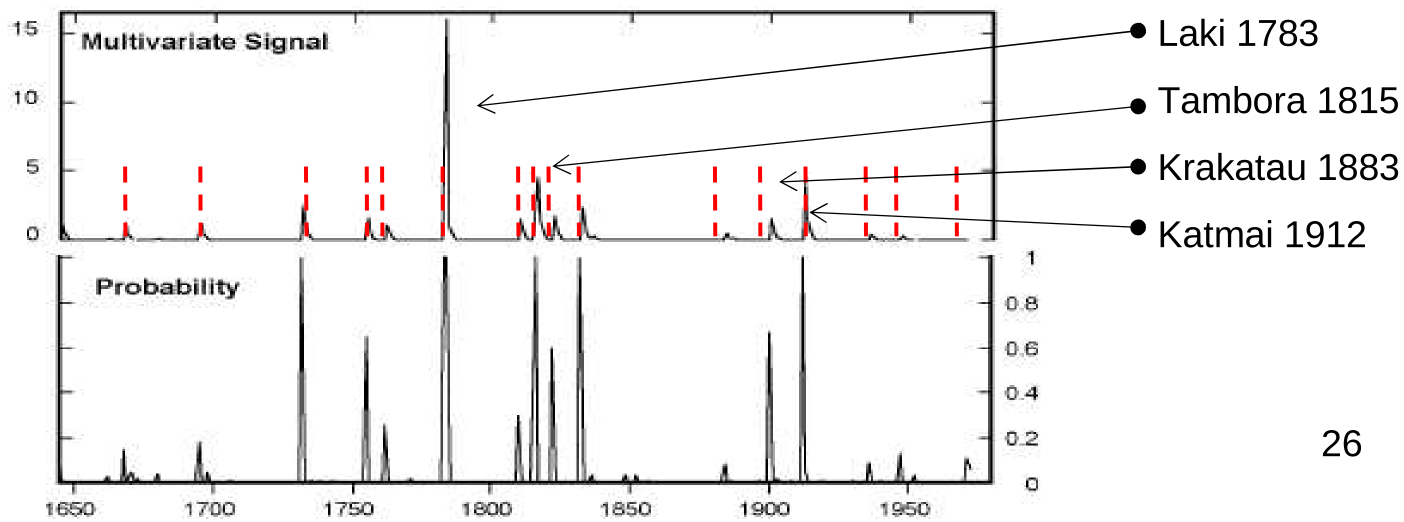
Result On Ice Cores Time Series



Univariate



Multivariate



Discussion on the model

I. <u>interesting points :</u>

- Global method vs iterative method
- Trend calculation : cycle, linear, whatever
- Multivariate : Improvement compare to univariate case
- Non linearities
- Different variances of noise

II. <u>Limitations :</u>

- Relative amplitudes
- Approximation of the distribution

Conclusion

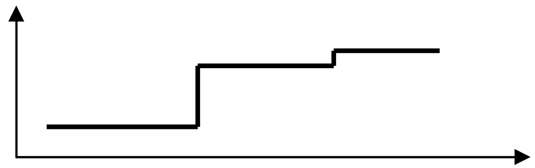
- R-package available (mep.R : julien.gazeaux@aero.jussieu.fr)

  http://www.r-project.org/

- Article : *Extracting Common Pulse-Like Signals from Multiple Ice Core Time Series, Batista, Gazeaux & al 2008* (Submitted)

- Applicatios :NO2 lightnings, aerosols, Dendrochronology, Volcanic eruptions, Ice cores …

- Similar model can be used to detect change point in time series ?



Ref :

1) Wecker, H.L., and Ansley, C.F., *The signal extraction approach to nonlinear regression and spline smoothing*, J. of the Amer Stat. Ass., 1983.

2) Guo, W., Y. Wang, and M. Brown, *A signal extraction approach to modeling hormone time series with pulses and a changing baseline*, J. of Am. Stat. Ass.,1998.