Detecting Performance Degradation of Software-Intensive Systems in the Presence of Trends and Long-Range Dependence

Alexey Artemov[2], *Evgeny Burnaev*[1]

[1]Institute for Information Transmission Problems RAS, Russia
[2]Lomonosov Moscow State University, Complex Systems Modelling Laboratory, Moscow, Russia

Outline

1. Motivation and objectives
2. Optimal Estimation of a Signal Perturbed by a Fractional Brownian Noise
3. Trend Estimation
4. Change-point detection
5. Ensembles of Weak Detectors
6. Applications to Real Data
7. Conclusions

- Software-intensive systems are widespread now:
  — broadband communications systems,
  — Internet systems (including devices and networks for data transmission, internet-services).
  — information systems and call-centers, etc.
- Smooth and efficient operation of these systems is a top priority (big number of users!)
- Hardware and human failures are now the norm for such systems, not the exception, due to their large-scale
- It is proposed (e.g. Casas 2010, Tartakovsky 2013) to detect/predict failures and malicious activity based on analysis of data (number of requests processed, average response time, volume of transmitted network traffic, etc.), collected during the operation of a system

- Mathematically considered applied problem is reduced to change-point detection problem
- Measured characteristics usually exhibits quasi-periodic stochastic cycles on a number of time-scales (day, week, year)
- Stochastic components exhibits long-range dependence, being a main cause of spikes in measurements
- Classical change-point detection statistics are optimal under restrictive assumptions, in particular, it is assumed that pre-/post-change distributions and change-point model are known

The aim of this work is to:

- develop methods for trend estimation observed in a fractional noise, modeling long-range dependence
- develop approaches for constructing ensembles of weak detectors for on-line detection of transient changes
- develop a methodology for modeling and estimation of signals with trends and propose a change-point detection framework on its basis

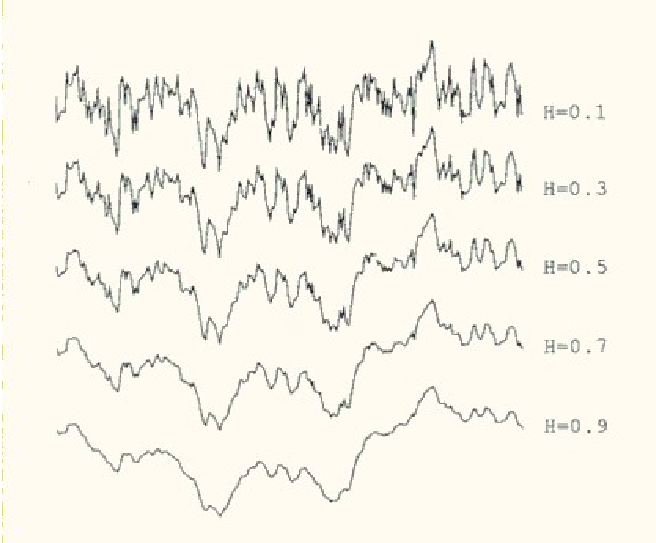
Fractional Brownian Motion

- Fractional Brownian Motion $B^H = \left(B^H_t\right)_{0 \leqslant t \leqslant T}$ with a Hurst parameter $H \in (0,1)$ is a Gaussian process with continuous trajectories, such that

$$\begin{aligned} B^H_0 &= 0, \qquad \mathbf{E} B^H_t = 0, \\ \mathbf{E} B^H_s B^H_t &= \frac{1}{2}\left(s^{2H} - t^{2H} + |t-s|^{2H}\right). \end{aligned}$$

- In case $H = 1/2$, Fractional Brownian Motion is an ordinary Brownian Motion
- Corresponding fractional noise processes can efficiently model many financial, economic, natural and technical systems (e.g. Keshner, 1982; Dubovikov et al., 2004)

Simulated trajectories of a Fractional Brownian Motion



Description of a Model

- $(\Omega, \mathcal{F}, (\mathcal{F}_t^{\xi})_{t\geq 0}, \mathrm{P})$ is a filtered probability space,
- Observed process $\xi = (\xi_t)_{0\leqslant t\leqslant T}$ is represented as

$$\xi_t = a(t) + \sigma(t) B_t^H,$$

where

— $B^H = \left(B_t^H\right)_{0\leqslant t\leqslant T}$ is a fractional Brownian motion with a Hurst parameter $H \in (0,1)$,

— diffusion coefficient $\sigma(t)$ is considered to be known,

— drift coefficient $a(t)$ is represented as

$$a(t) = \sum_{i=0}^{n} \theta_i \varphi_i(t)$$

w.r.t. to a given dictionary of functions $\{\varphi_i(t)\}_{i=0,\,\ldots,n}$,

— coefficients $\theta_i, i = 0,\ \ldots, n$ are not known.

We denote $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_n)^{\mathsf{T}}$, $\boldsymbol{\varphi}(t) = (\varphi_0(t), \ldots, \varphi_n(t))^{\mathsf{T}}$, then

$$a(t) = \boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{\varphi}(t)$$

Problem Statement

- Observations $\{\xi_s, 0 \leqslant s \leqslant t\}$ up to t are given
- **Problem 1.** Assuming $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$ find MLE
- **Problem 2.** Assuming prior distribution of $\boldsymbol{\theta}$ to be
  - — $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$, or
  - — uniform on $\mathbf{r} = [a_0, b_0] \times \cdots \times [a_n, b_n]$,

  find strategy $\delta^* = (\tau^*, \widehat{\boldsymbol{\theta}}^*)$, such that

$$\inf_{\delta \in D} \mathbf{E}\left[c\tau + \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|^2\right] = \mathbf{E}\left[c\tau^* + \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^*\|^2\right],$$

  where
  - — $c \geq 0$ is a given cost of observations,
  - — D is a set of strategies $\delta = (\tau, \widehat{\boldsymbol{\theta}})$ containing finite stopping times w.r.t. to the filtration $\mathcal{F}_t^\xi$

Notations

We denote by

$$\begin{aligned}
\kappa_H &= 2H\Gamma\left(\frac{3}{2}-H\right)\Gamma\left(\frac{1}{2}+H\right),\\
k_H(t,s) &= \kappa_H^{-1} s^{1/2-H}(t-s)^{1/2-H},\\
\lambda_H &= \frac{2H\Gamma(3-2H)\Gamma\left(\frac{1}{2}+H\right)}{\Gamma\left(\frac{3}{2}-H\right)},\\
w_H(t) &= \lambda_H^{-1} t^{2-2H},\\
dw_t^H &= d(w_H(t)) = \lambda_H^{-1}(2-2H)t^{1-2H}dt
\end{aligned}$$

Let us define processes $M^H = \left(M_t^H\right)_{0\leqslant t\leqslant T}, m^H = \left(m_t^H\right)_{0\leqslant t\leqslant T}$ according to

$$M_t^H \equiv \int_0^t k_H(t,s)\,d\xi_s, \qquad m_t^H = M_t^H / w_H(t).$$

Theorem 1: MLE

MLE $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ for $\boldsymbol{\theta}$ is defined by

$$\widehat{\boldsymbol{\theta}}_{\mathrm{ML}} = \boldsymbol{R}_H^{-1}(t)\boldsymbol{\psi}_t^H,$$

where $\boldsymbol{R}_H(t)$ is a nonrandom matrix with elements

$$(\boldsymbol{R}_H(t))_{ij} = \int_0^t \psi_i(s)\psi_j(s)\,dw_s^H, \qquad i, j = 0, \ldots, n,$$

and $\boldsymbol{\psi}^H = \left(\boldsymbol{\psi}_t^H\right)_{0 \leqslant t \leqslant T}$ is a stochastic process taking values in $\mathbb{R}^{n+1}$ with coordinates defined by

$$(\boldsymbol{\psi}_t^H)_i = \int_0^t \psi_i(s)\,dM_s^H, \qquad i = 0, \ldots, n,$$

the functions $\psi_i(t)$, $i = 0, \ldots, n$, are given by

$$\psi_i(t) = \frac{d}{dw_t^H}\left(\int_0^t k_H(t,s)\sigma^{-1}(s)\varphi_i'(s)\,ds\right), \qquad i = 0, \ldots, n.$$

Corollary: polynomial drift

Let $\varphi_i(t) = t^i$, $i = 0, \dots, n$, $\sigma(t) \equiv \sigma$, then

- $\psi_i(t) = \beta_H(i)/\sigma t^{i-1}$, $i = 0, \dots, n$,
- elements of $\boldsymbol{\psi}^H$ and $\boldsymbol{R}_H(t)$ are defined by

$$(\boldsymbol{\psi}_t^H)_i = \frac{\beta_H(i)}{\sigma} \int_0^t s^{i-1}\, dM_s^H, \quad (\boldsymbol{R}_H(t))_{ij} = \frac{\alpha_H(i,j)}{\sigma^2}\, t^{i+j-2H},$$

  where

$$\begin{aligned} \alpha_H(i,j) &= \lambda_H^{-1} \beta_H(i) \beta_H(j)\, \frac{2-2H}{i+j-2H}, \\ \beta_H(i) &= i\, \frac{2-2H+i-1}{2-2H}\, \frac{\Gamma(3-2H)}{\Gamma(3-2H+i-1)}\, \frac{\Gamma(3/2-H+i-1)}{\Gamma(3/2-H)}, \end{aligned}$$

- MLE $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ is obtained as a solution of $\boldsymbol{\psi}_t^H - \boldsymbol{R}_H(t)\boldsymbol{\theta} = 0$.

Corollary: linear drift

- In case of a linear trend $d\xi_s = \theta_1 ds + \sigma\, dB_s^H$, $s \in [0, t]$
- MLE has the form

$$(\widehat{\theta}_1)_{\mathrm{ML}} = \frac{\sigma M_t^H}{w_H(t)}.$$

- This particular result (for $\sigma = 1$) was obtained in (Norros I. et al. 1999).

Theorem 2: Normal prior distribution

Let $\boldsymbol{\theta}$ has a prior $\mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$. Then the optimal Bayesian estimate $\widehat{\boldsymbol{\theta}}_{\mathrm{BAYES}}$ is the posterior mean

$$\widehat{\boldsymbol{\theta}}_{\mathrm{BAYES}} = \mathbf{E}\,[\boldsymbol{\theta} \,|\, \mathcal{F}_t^{\xi}] = (\boldsymbol{R}_H(t) + \boldsymbol{\Sigma}^{-1})^{-1}(\boldsymbol{\psi}_t^H + \boldsymbol{\Sigma}^{-1}\mathbf{m}).$$

The conditional estimation error $\mathbf{E}\,(\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\mathrm{BAYES}}\|^2 \,|\, \mathcal{F}_t^{\xi})$ is defined by the trace of the posterior covariance matrix

$$\operatorname{cov}[\boldsymbol{\theta} \,|\, \mathcal{F}_t^{\xi}] = (\boldsymbol{R}_H(t) + \boldsymbol{\Sigma}^{-1})^{-1}.$$

Corollary: optimal stopping time

- The optimal stopping time

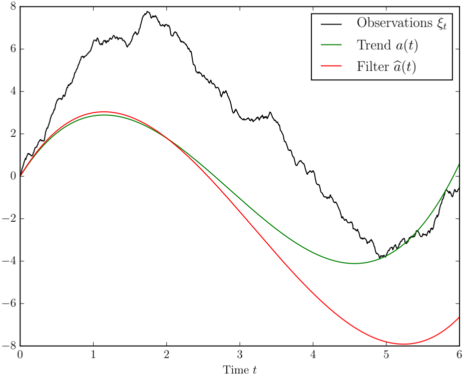
$$\begin{aligned}\widehat{\tau}_{\mathrm{BAYES}} &= \arg\inf_{\tau\in D} \mathbf{E}\left[c\tau + \mathbf{E}\left(\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\mathrm{BAYES}}\|^2 \,|\, \mathcal{F}_\tau^\xi\right)\right] \\ &= \arg\inf_{t\in[0,T]} F_H(t),\end{aligned}$$

  where the function

$$\begin{aligned}F_H(t) &= ct + \mathbf{E}\left(\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\mathrm{BAYES}}\|^2 \,|\, \mathcal{F}_t^\xi\right) \\ &= ct + \mathrm{tr}\left((\boldsymbol{R}_H(t) + \boldsymbol{\Sigma}^{-1})^{-1}\right), \qquad t \in [0,T],\end{aligned}$$

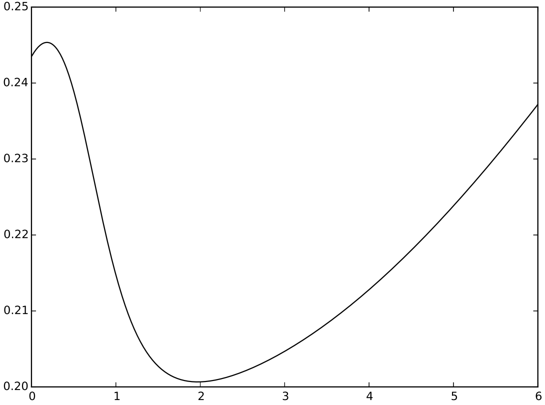
- Let $\varphi_i(t) = t^i$, $i = 0, \ldots, n$, $\sigma(t) \equiv \sigma$, $\boldsymbol{\Sigma} = \mathrm{diag}\,(\gamma_0^2, \ldots, \gamma_n^2)$, then the function $F_H(t)$ has a single minimum for $t \in [0,T]$,
- These results generalize those from (Cetin et al. 2013), obtained for the case of a linear trend

Numerical modeling



- $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$, $\boldsymbol{m} = [4, -4, 0.4]$, $\boldsymbol{\Sigma} = \mathrm{diag}(0.1, 0.1, 0.1)$,
- $\boldsymbol{\theta} = [5.5, -3.0, 0.35]$, $\sigma = 3$, $H = 0.8$,
- $\widehat{\boldsymbol{\theta}}_{\mathrm{BAYES}} = [5.73, -3.05, 0.32]$.

Numerical modeling



Cost function $F_H(t)$ for $n = 2, H = 0.2, c = 0.02$.

Theorem 3: Uniform prior distribution

Let $d\xi_t = \theta_1\,dt + \sigma\,dB_t^H$ with $\theta_1 \sim U(a,b)$. Then the optimal Bayesian estimate has the form

$$(\widehat{\theta}_1)_{\mathrm{BAYES}} = m_t^H + [Z_t^H w_H(t)]^{-1}[\Lambda_t^H(a) - \Lambda_t^H(b)],$$

the conditional mean square estimation error is given by

$$\begin{aligned}\gamma_t^H &= \mathbf{E}\left((\theta_1 - (\widehat{\theta}_1)_{\mathrm{BAYES}})^2 \,|\, \mathcal{F}_t^\xi\right) = [w_H(t)]^{-1} + \\ &+ [Z_H(t)w_H(t)]^{-1}[\Lambda_t^H(a)(a - m_t^H) - \Lambda_t^H(b)(b - m_t^H)] \\ &\quad - [Z_H(t)w_H(t)]^{-2}[\Lambda_t^H(a) - \Lambda_t^H(b)]^2,\end{aligned}$$

where

$$\begin{aligned}Z_t^H &= \sqrt{\frac{2\pi}{w_H(t)}}\,\exp\left\{\frac{1}{2}\,(m_t^H)^2 w_H(t)\right\}C_t^H, \\ C_t^H &= \Phi\left((b - m_t^H)\sqrt{w_H(t)}\right) - \Phi\left((a - m_t^H)\sqrt{w_H(t)}\right).\end{aligned}$$

Optimal stopping time

- To determine the optimal stopping time, it is necessary to solve

$$
\begin{aligned}
\widehat{\tau}_{\mathrm{BAYES}} &= \arg\inf_{\tau} \mathbf{E}\left[c\tau + \mathbf{E}\left((\theta_1 - (\widehat{\theta}_1)_{\mathrm{BAYES}})^2 \,|\, \mathcal{F}_\tau^\xi\right)\right] = \\
&= \arg\inf_{\tau} \mathbf{E}\,[c\tau + \gamma_\tau^H].
\end{aligned}
$$

- Explicit expression for an optimal stopping time is not possible even in case of a linear trend

Modelling trends and quasi-periodicity

Problems involving quasi-periodic systems, exhibiting significant trends, are extremely widespread:

- analysis and prediction of faults in distributed information systems
- prediction of electricity load
- detection of changes in natural systems
- analysis of fluctuations of seasonal temperature
- prediction of the amount of transmitted traffic
- analysis and prediction of macroeconomic data

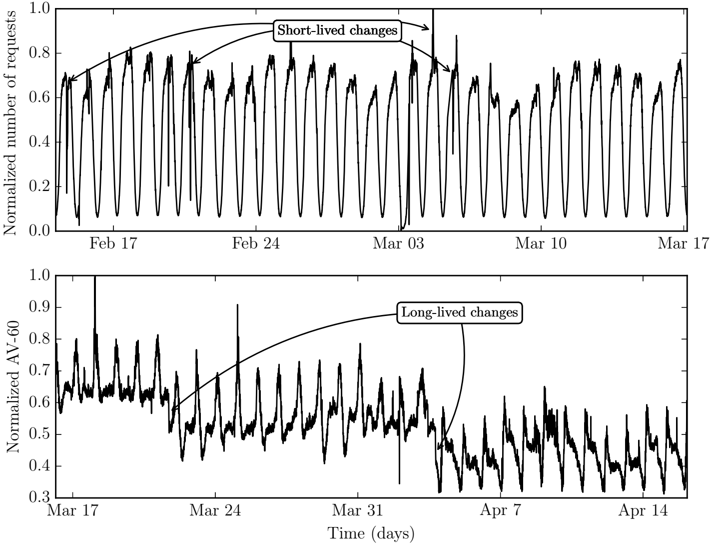


Figure : Web service load

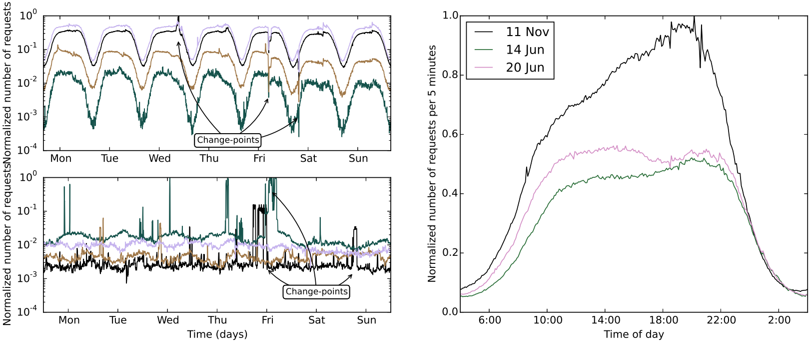


Figure : Top-left: weekly load profile of a geoinformation system. Right: daily load of a system aggregated over consecutive 5-minute intervals for three different days. Bottom-left: weekly load shape of the traffic in the Abilene network

Trend estimation based on MLE

- We assume that observations are generated by

$$X_t = f(t) + \eta_t^H, \qquad t \geqslant 0,$$

  where

  — the trend $f(t)$ is a some smooth function

  — $\eta_t^H = \sigma Z_t^H$, Z_t^H is a fBM noise

- We can approximate $f(t)$ using some finite-order polynomial $\sum_{i=0}^{n} \theta_i (t - t_0)^i$ in the neighbourhood of any $t_0 > 0$,
- Given noisy observations $\{(X_k, t_k)\}_{k=1}^{l}$, the goal is to estimate the expected value $f(t) = \mathbf{E} X_t$ for any $t \geqslant 0$
- We assume σ and H to be known, since in practice

  — it is sufficient to initialize σ roughly,

  — H can be easily estimated e.g. by an approach from (Dubovikov M.M. et al. 2004),

1. Set $W(a,b) = \{(X_{t_k}, t_k) : a \leqslant t_k \leqslant b\}$
2. Assuming

$$X_{t_k} \approx \sum_{i=0}^{3} \theta_i (t_k - t_0)^i + \sigma Z_{t_k}^H, \qquad (X_{t_k}, t_k) \in W(a,b),$$

   where $t_0 = (a+b)/2$, estimate $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_3)$ using MLE
3. Using $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$, compute the expected sample path by $\widehat{f}(t) = \sum_{i=0}^{3} (\widehat{\boldsymbol{\theta}}_{\mathrm{ML}})_i (t - t_0)^i$ for each $t \in [a,b]$
4. To obtain the estimate $\widehat{f}(t)$ for all $t \geqslant 0$, we use the sliding window $[a, a+\Delta]$ with sufficiently large Δ.

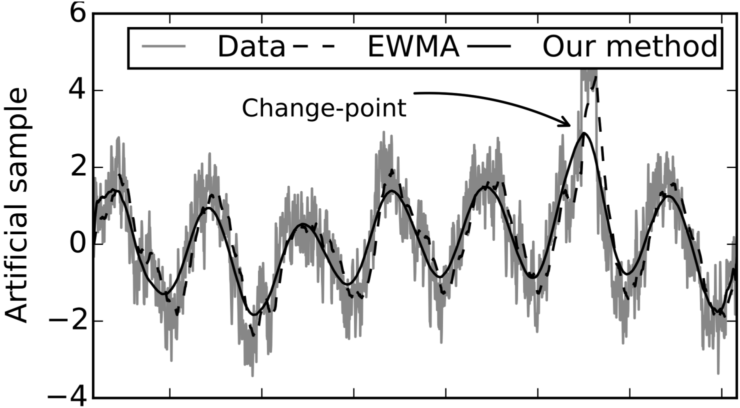


Figure : Artificial data: $\{(X_k, t_k)\}_{k=1}^{l}, l = 2016$, measured at consecutive 5-minute intervals according to the model $X_k = A\sin(2\pi t_k/T) + \eta^H(t)$, where $A = 1.5, T = 288$, and $\eta^H(t)$ is the LRD noise process with $H = 0.8$

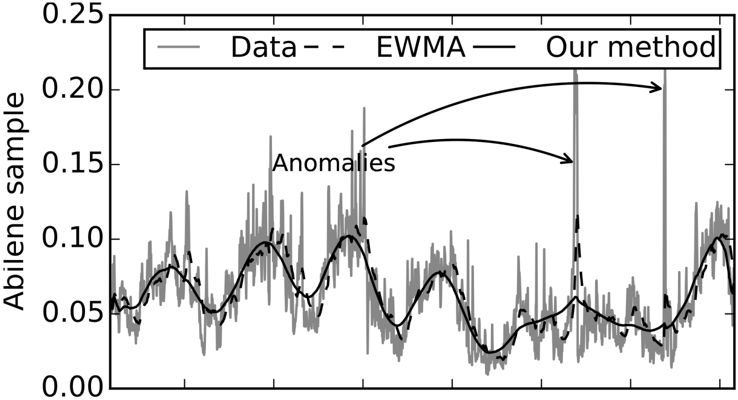


Figure : Abilene network service load

Change-point model and residual process

- Change-point (CP) model for the noise process

$$\eta^H(t) = \mu 1_{[\theta, \theta+\Delta t]}(t) + \sigma Z_t^H, \qquad t \geqslant 0,$$

  where
  — θ is an unknown time of a change,
  — μ is an unknown change magnitude,
  — σ is a unknown (non-random) variance,
  — Z_t^H is a fGn

- To detect the change, we introduce a residual process $R = (R_t)_{t \geqslant 0}$

$$R_t = \sigma^{-1}(X_t - \widehat{X}_t), \qquad t \geqslant 0,$$

  where X_t is an observed signal and $\widehat{X}_t$ is an estimate of its trend

- $\mathbf{E}R_t \neq 0$ for $t \in [\theta, \theta + \Delta t]$, and $\mathbf{E}R_t \approx 0$ otherwise

Challenges in CP detection

Many existing CP detection procedures *(control charts, CUSUM, EWMA, Shiryaev-Roberts procedure, etc.)* assume about the residual process that

- pre- and post-change distributions are known
- no repeated changes occur (change lasts forever)
- observations are stationary
- observations are independent

Violation of these assumptions $\rightarrow$ poor performance!

Why ensembles?

- Different detection algorithms have different properties in different conditions
- In machine learning, ensembles improve performance of weak predictors

Goals of the work:

- Define an ensemble of CP detection procedures
- Propose an ensemble learning algorithm
- Evaluate ensemble performance via experiments

Ensembles of CP detection procedures

- Let $\Pi_1, \ldots, \Pi_n$ denote n CP detection procedures, such as the CUSUM procedure based on the process $s = (s_t)_{t \geqslant 0}$:

$$s_t = \max\left\{0, \max_{1 \le \theta \le t} \sum_{k=\theta}^{t} \zeta_k \right\} = \max(0, s_{t-1} + \zeta_t), \quad T_0 = 0,$$

  where $\zeta_t = \log(f_0(X_t)/f_\infty(X_t))$ is the log-likelihood ratio, and $f_\infty(\cdot)$ and $f_0(\cdot)$ are 1d pre/post-change distributions
- Π_k : alarm at time $\tau_k = \inf\{t \ge 0 : s_t^k \ge h_k\}$
- **Ensemble:** alarm at $\tau_{\mathrm{A}} = \inf\{t \geqslant 0 : a_t \geqslant h_{\mathrm{A}}\}$

$$a_t = \psi(\boldsymbol{\lambda}; \mathbf{S}_t^1, \ldots, \mathbf{S}_t^n),$$

  where
  - $\mathbf{S}_t^k = \left\{s_u^k, 0 \le u \le t\right\}$: sample path of s^k
  - $\boldsymbol{\lambda} \in \mathbb{R}^n$: parameters

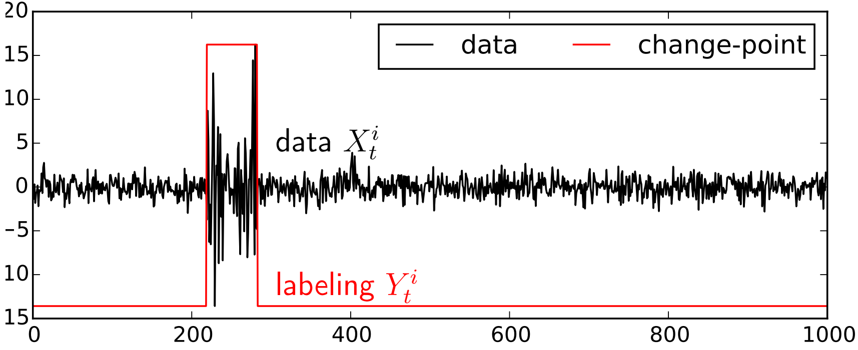
Examples of specific ensembles

- Majority voting: $\psi(\boldsymbol{\lambda}; \mathbf{S}) = \frac{2}{n} \sum_{k=1}^{n} 1_{\{s_t^k \geq 1\}}(t)$
- Weighted voting of "probabilities"
  $\psi(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{S}) = \sum_{k=1}^{n} \alpha_k \sigma\big(\beta_k (s_t^k - 1/2)\big)$,
  $\sigma(x) = 1/(1 + e^{-x})$
- Logistic regression based classifier

$$\psi(\boldsymbol{\lambda}; \mathbf{S}) = \sigma\Big(\sum_{j=0}^{p} \sum_{k=1}^{n} \lambda_{kj} s_{t-j}^{k} - \lambda_0\Big)$$

Learning the ensemble parameters

- $\mathcal{X} = \{(X_t^i, Y_t^i), i = 1, \ldots, N\}$: the training set
- We have: $X_t^i = \begin{cases} (X_t^\infty)^i, & \text{if } t \in \mathcal{T}_\infty^i, \\ (X_t^0)^i, & \text{if } t \in \mathcal{T}_0^i \end{cases}$



Learning problem for an ensemble

- Consider the performance measure

$$\mathbf{F}(h, c_\infty, c_0) = c_\infty \mathbf{F}_\infty(h) + c_0 \mathbf{F}_0(h)$$

- The quantities $\mathbf{F}_\infty(h)$ и $\mathbf{F}_0(h)$ are defined by

$$\mathbf{F}_\infty(h) = \mathbf{E}_\infty \left[\frac{\overbrace{\int_{\mathcal{T}_\infty} 1_{\{a_t \geqslant h\}}(t) dt}^{\text{duration of the false alarm}}}{\underbrace{\int_0^T 1_{\mathcal{T}_\infty}(t) dt}_{\text{in-control duration}}} \right], \qquad \mathbf{F}_0(h) = \mathbf{E}_0 \left[\frac{\overbrace{\int_{\mathcal{T}_0} 1_{\{a_t < h\}}(t) dt}^{\text{duration of the false silence}}}{\underbrace{\int_0^T 1_{\mathcal{T}_0}(t) dt}_{\text{out-of-control duration}}} \right]$$

- Learning: optimize $\mathbf{F}(h, c_\infty, c_0) \to \inf\limits_{\boldsymbol{\lambda} \in \mathbb{R}^n}$

Learning the ensemble parameters

- We consider the measure

$$\mathbf{F}(h,1,1) = \mathbf{E}_\infty\left[\frac{\int_{\mathcal{T}_\infty} 1_{\{a_t \geqslant h\}}(t)dt}{\int_0^T 1_{\mathcal{T}_\infty}(t)dt}\right] + \mathbf{E}_0\left[\frac{\int_{\mathcal{T}_0} 1_{\{a_t < h\}}(t)dt}{\int_0^T 1_{\mathcal{T}_0}(t)dt}\right]$$

- Its empirical approximation (nondifferentiable)

$$\widehat{\mathbf{F}}(h,1,1) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T_\infty^i}\sum_{t\in\mathcal{T}_\infty^i} 1_{\{a_t \geqslant h\}}(t) + \frac{1}{T_0^i}\sum_{t\in\mathcal{T}_0^i} 1_{\{a_t < h\}}(t)\right\}$$

- Differentiable approximation

$$\widehat{\mathbf{F}}_{\mathrm{D}}(h,1,1) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T_\infty^i}\sum_{t\in\mathcal{T}_\infty^i} \sigma(a_t - h) + \frac{1}{T_0^i}\sum_{t\in\mathcal{T}_0^i} \sigma(h - a_t)\right\}$$

can be optimized using gradient descent

Performance of ensemble for change-point detection

Example: mean shift, long range dependent process

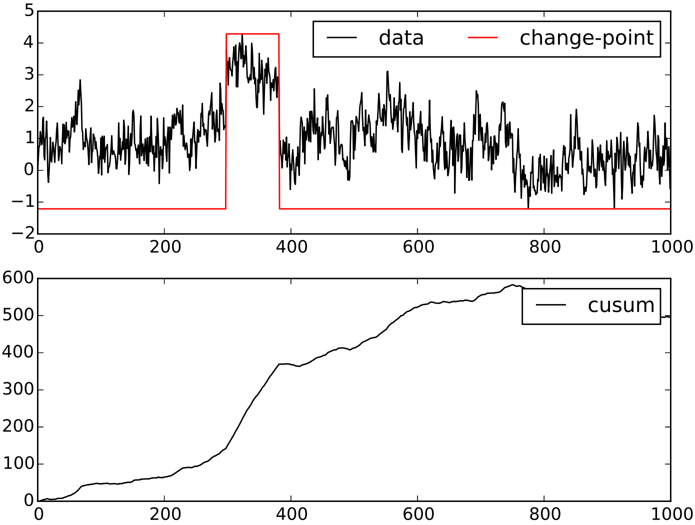
$$X_t = \begin{cases} B_t^H, & \text{if } t \notin [\theta, \theta + \Delta t], \\ \mu + B_t^H, & \text{if } t \in [\theta, \theta + \Delta t] \end{cases}$$

- $B^H = \left(B_t^H\right)_{0 \leqslant t \leqslant T}$ is a fractional gaussian noise (fGn) process with Hurst index $H = 0.95$
- change magnitude $\mu > 0$ is unknown
- change time $\theta \in [0, T]$ is unknown

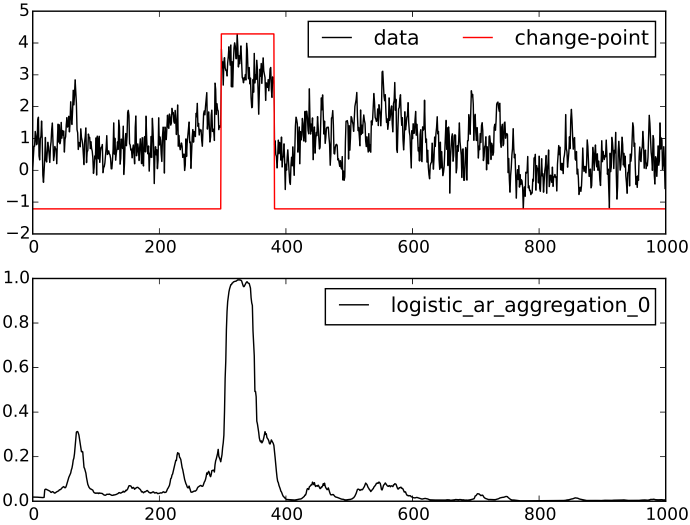
Learn the ensemble assuming

$$X_t = \mu 1_{\{t \geqslant \theta\}}(t) + W_t, \qquad \mu = 1$$

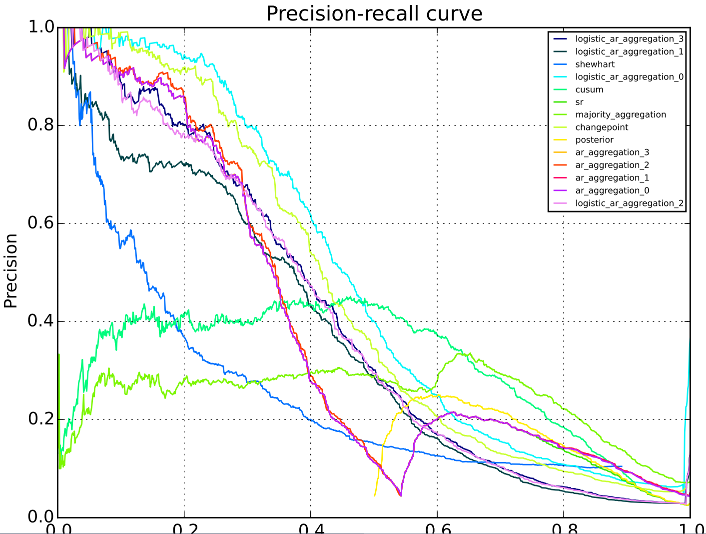
Example. fGN process



Example. fGn process



Performance of ensemble for change-point detection



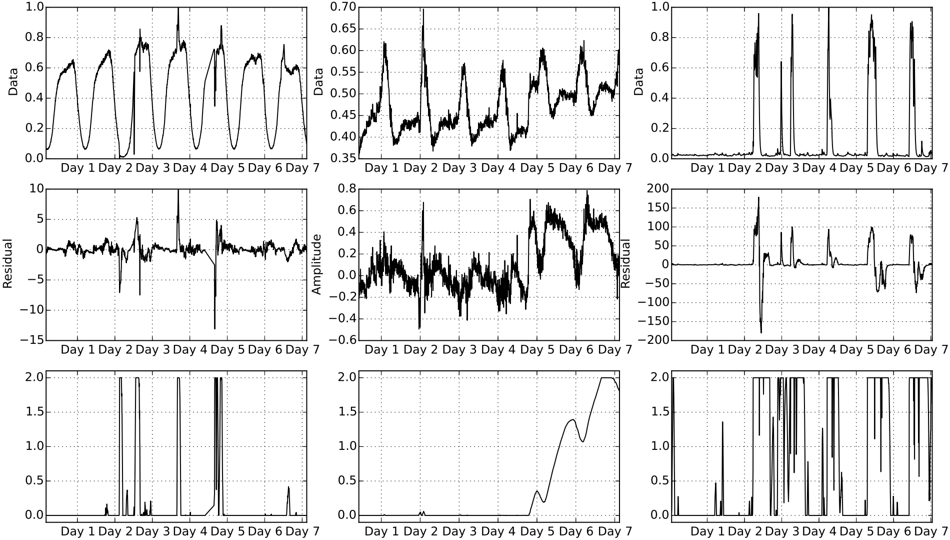


FIGURE : From top to bottom: real signal, residuals, detector

We developed and analysed

- methods for trend estimation observed in a fractional noise, modeling long-range dependence
- approaches for constructing ensembles of weak detectors for on-line detection of transient changes
- methodology for modeling and estimation of signals with trends and propose a change-point detection framework on its basis

Quality of ensembles for change-point detection

Example 1: mean change for heavy-tailed process

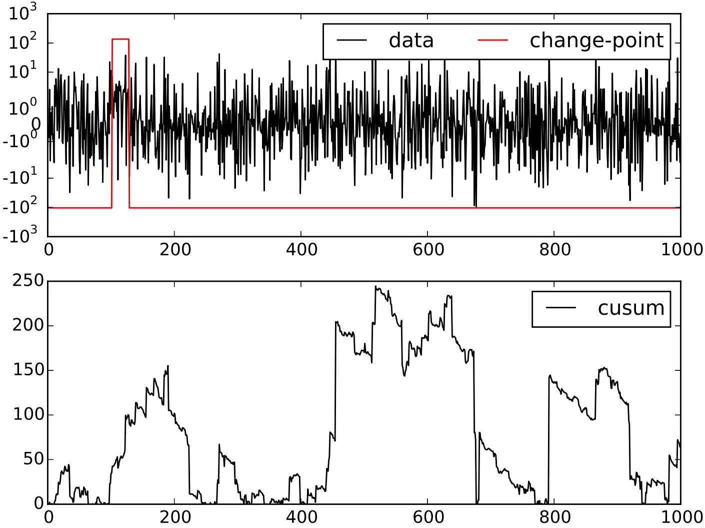
$$X_t = \begin{cases} C_t, & \text{if } t \notin [\theta, \theta + \Delta t], \\ \mu + C_t, & \text{if } t \in [\theta, \theta + \Delta t] \end{cases}$$

- $C = (C_t)_{0 \leqslant t \leqslant T}$ are Cauchy-distributed i.i.d.r.v. with shift $x = 0$ and scale $\gamma = 1$
- change magnitude $\mu > 0$ is unknown
- change time $\theta \in [0, T]$ is unknown

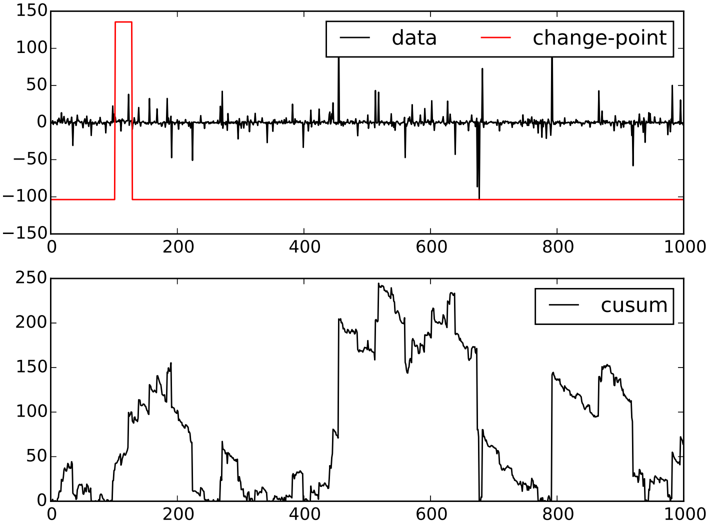
Learn the ensemble assuming

$$X_t = \mu 1_{\{t \geqslant \theta\}}(t) + W_t, \qquad \mu = 1$$

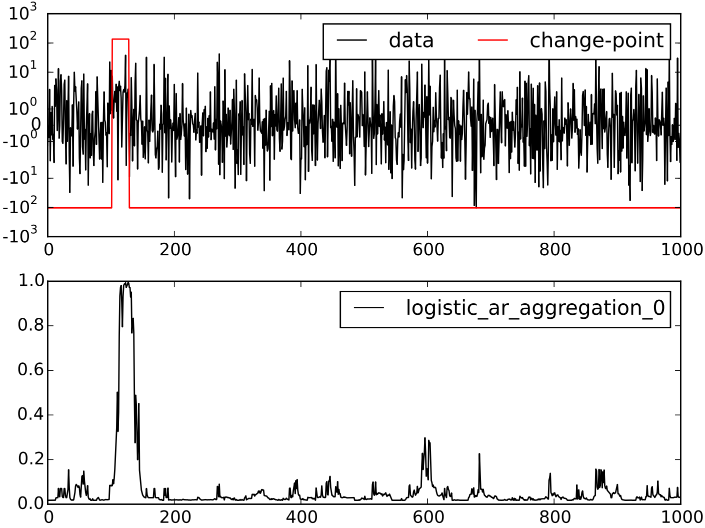
Example 1. Cauchy process



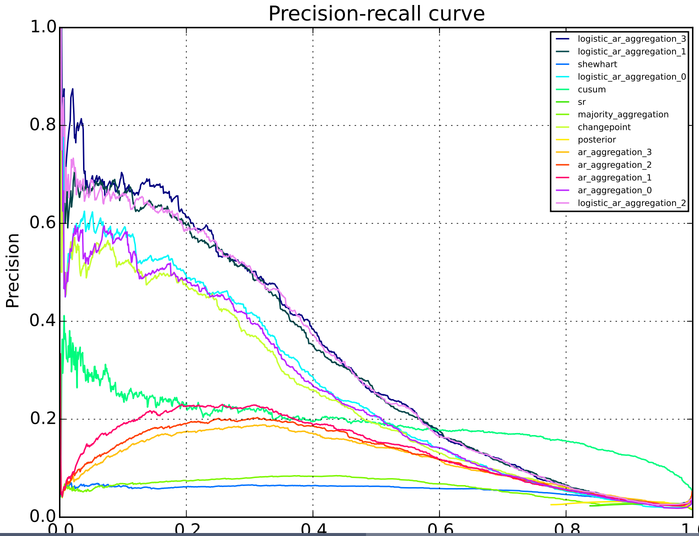
Example 1. Cauchy process



Example 1. Cauchy process



Performance of ensemble for change-point detection



Artifical datasets used in the evaluation

Dataset	Type of data	Parameters subject to change	Change time θ and change duration Δ	Change magnitude
WhiteNoise	Uncorrelated Gaussian Noise	mean	random, $\theta \sim U(200, 800)$, $\Delta \sim U(5, 100)$	random, $\mu \sim U(0.1, 2.0)$
Fractal	Fractional Gaussian Noise			
Cauchy	Uncorrelated Cauchy Noise			
GARCH1	GARCH$(1, 1)$ process	α_1, β_1		random, $\alpha_1 \sim U(.4, .8)$, $\beta_1 \sim U(.1, .2)$,
ARMA-AR	ARMA(10, 3) process	AR terms φ_i		random
ARMA-MA	ARMA(10, 3) process	MR terms θ_j		random
GARCH1 + ARMA	GARCH$(1, 1)$ + ARMA(10, 3) process	$\alpha_1, \beta_1, \varphi_i, \theta_j$		random

Performance of ensemble for change-point detection

	WhiteNoise	**Fractal**	**Cauchy**	**GARCH1**	**ARMA-AR**	**ARMA-MA**	**GARCH1 + ARMA**
Shewhart	77.52	24.44	05.45	32.08	19.80	76.37	40.00
CUSUM	61.11	30.70	19.24	59.88	28.74	89.90	**75.44**
SR	22.22	6.11	.40	50.06	24.17	7.15	72.72
Changepoint	60.62	45.42	24.18	21.94	13.03	57.15	22.98
Posterior π_t	27.38	7.76	.66	53.60	29.00	35.69	74.58
Maj	62.13	24.62	6.11	47.80	28.74	**92.71**	67.01
Weight − 0	71.73	38.94	25.08	55.62	24.94	79.48	67.23
Weight − 1	71.58	38.97	13.46	57.65	29.60	91.92	71.28
Weight − 2	73.89	39.63	12.29	56.57	**30.45**	91.13	69.28
Weight − 3	73.25	38.98	11.61	**57.83**	26.24	90.70	72.11
Log − 0	77.25	**48.64**	25.90	51.35	23.29	87.72	68.35
Log − 1	76.27	36.20	31.03	50.03	23.49	88.47	65.97
Log − 2	78.01	39.74	31.30	49.35	27.99	88.93	66.43
Log − 3	**78.85**	40.31	**32.24**	49.08	27.99	88.88	66.77