Gaussian Model Selection with Unknown Variance

Y. Baraud, C. Giraud and S. Huet

Université de Nice - Sophia Antipolis,
INRA Jouy en Josas

Luminy, 13-17 novembre 2006

The statistical setting

The statistical model

Observations: $Y_i = \mu_i + \sigma \varepsilon_i, \ \ i = 1, \ldots, n$

- $\mu = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$ and $\sigma > 0$ are unknown
- $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d standard Gaussian

Collection of models / estimators

- $\mathcal{S} = \{S_m, \ m \in \mathcal{M}\}$ a countable collection of linear subspaces of $\mathbb{R}^n$ (models)
- $\hat{\mu}_m =$ least-squares estimator of μ on S_m

Example: change-points detection

- $\mu_i = f(x_i)$ with $f : [0,1] \mapsto \mathbb{R}$, piecewise constant.

- $\mathcal{M}$ is the set of increasing sequences $m = (t_0, \ldots, t_q)$ with $q \in \{1, \ldots, p\}$, $t_0 = 0$, $t_q = 1$, and $\{t_1, \ldots, t_{q-1}\} \subset \{x_1, \ldots, x_n\}$.

- models:
$$S_m = \left\{ (g(x_1), \ldots, g(x_n))', \ g \in \mathcal{F}_m \right\},$$
where
$$\mathcal{F}_{(t_0, \ldots, t_q)} = \left\{ g = \sum_{j=1}^{q} a_j \mathbf{1}_{[t_{j-1}, t_j[} \ \text{with} \ \ (a_1, \ldots, a_q) \in \mathbb{R}^q \right\}.$$

- No residual squares to estimate the variance.

Risk on a single model

Euclidean risk on S_m:

$$\mathbb{E}\left[\|\mu - \hat{\mu}_m\|^2\right] = \underbrace{\|\mu - \mu_m\|^2}_{\text{bias}} + \underbrace{D_m \sigma^2}_{\text{variance}}$$

Ideal: estimate μ with $\hat{\mu}_{m^*}$, where m^* minimizes $m \mapsto \mathbb{E}\left[\|\mu - \hat{\mu}_m\|^2\right] \ldots$

Model selection

Selection rule: we set $D_m = \dim(S_m)$ and select $\hat{m}$ minimizing

$$\mathrm{Crit}_L(m) = \|Y - \hat{\mu}_m\|^2 \left(1 + \frac{\mathrm{pen}(m)}{n - D_m}\right) \tag{1}$$

or

$$\mathrm{Crit}_K(m) = \frac{n}{2} \log \left(\frac{\|Y - \hat{\mu}_m\|^2}{n} \right) + \frac{1}{2} \mathrm{pen}'(m). \tag{2}$$

Some classical penalties:

FPE	AIC	BIC	AMDL
$\mathrm{pen}(m) = 2D_m$	$\mathrm{pen}'(m) = 2D_m$	$\mathrm{pen}'(m) = D_m \log n$	$\mathrm{pen}'(m) = 3D_m \log n$

Model selection

Selection rule: we select $\hat{m}$ minimizing

$$\mathrm{Crit}_L(m) = \|Y - \hat{\mu}_m\|^2 \left(1 + \frac{\mathrm{pen}(m)}{n - D_m}\right)$$

or

$$\mathrm{Crit}_K(m) = \frac{n}{2} \log\left(\frac{\|Y - \hat{\mu}_m\|^2}{n}\right) + \frac{1}{2}\mathrm{pen}'(m).$$

Criteria (1) and (2) are equivalent with

$$\mathrm{pen}'(m) = n \log\left(1 + \frac{\mathrm{pen}(m)}{n - D_m}\right).$$

Objectives

- for classical criteria: to analyze the Euclidean risk of $\hat{\mu}_{\hat{m}}$ with regard to the complexity of the family of model $\mathcal{S}$, and compare this risk to

$$\inf_{m \in \mathcal{M}} \mathbb{E}\left[\|\mu - \hat{\mu}_m\|\right]^2 .$$

- to propose penalties versatile enough to take into account the complexity of $\mathcal{S}$ and the sample size.

Complexity:

We say that $\mathcal{S}$ has an index of complexity (M, a) if for all $D \geq 1$

$$\mathrm{card}\left\{m \in \mathcal{M},\ D_m = D\right\} \leq M e^{aD}.$$

Theorem 1: Performances of classical penalties

Let $K > 1$ and $\mathcal{S}$ with complexity $(M, a) \in \mathbb{R}_+^2$. If for all $m \in \mathcal{M}$,

$$D_m \leq D_{\max}(K, M, a) \qquad \text{(explicit)}$$

and

$$\text{pen}(m) \geq K^2 \phi^{-1}(a) D_m,$$

with $\phi(x) = (x - 1 - \log x)/2$ for $x \geq 1$, then

$$\mathbb{E}\left[\|\mu - \hat{\mu}_{\hat{m}}\|^2\right] \leq \frac{K}{K-1} \inf_{m \in \mathcal{M}} \left[\|\mu - \mu_m\|^2 \left(1 + \frac{\text{pen}(m)}{n - D_m}\right) + \text{pen}(m)\sigma^2\right] + R$$

where

$$R = \frac{K\sigma^2}{K-1}\left[K^2\phi^{-1}(a) + 2K + \frac{8KMe^{-a}}{\left(e^{\phi(K)/2} - 1\right)^2}\right].$$

Performances of $\hat{\mu}_{\hat{m}}$

- under the above hypotheses if $\mathrm{pen}(m) = K\phi^{-1}(a)D_m$ with $K > 1$

$$\mathbb{E}\left[\|\mu - \hat{\mu}_{\hat{m}}\|^2\right] \leq c(K,M)\,\phi^{-1}(a)\,\left[\inf_{m\in\mathcal{M}} \mathbb{E}\left[\|\mu - \hat{\mu}_m\|^2\right] + \sigma^2\right]$$

- The condition "$\mathrm{pen}(m) \geq K^2\phi^{-1}(a)D_m$ with $K > 1$" is sharp (at least when $a = 0$ and $a = \log n$).

Roughly, for large values of n this imposes the restrictions:

Criteria	FPE	AIC	BIC	AMDL
Complexity	$a < 0.15$	$a < 0.15$	$a < \frac{1}{2}\log(n)$	$a < \frac{3}{2}\log(n)$

Dkhi function

For $x \geq 0$, we define

$$\mathrm{Dkhi}[D,N,x] = \frac{1}{\mathbb{E}(X_D)} \times \mathbb{E}\left[\left(X_D - x\frac{X_N}{N}\right)_+\right] \quad \in]0,1].$$

where X_D and X_N are two independent $\chi^2(D)$ and $\chi^2(N)$.

Computation: $x \mapsto \mathrm{Dkhi}[D,N,x]$ is decreasing and

$$\mathrm{Dkhi}[D,N,x] = \mathbb{P}\left(F_{D+2,N} \geq \frac{x}{D+2}\right) - \frac{x}{D}\mathbb{P}\left(F_{D,N+2} \geq \frac{(N+2)x}{DN}\right),$$

where $F_{D,N}$ is a Fischer random variable with D and N degrees of freedom.

Theorem 2: a general risk bound

Let pen be an arbitrary non-negative penalty function and assume that $N_m = n - D_m \geq 2$ for all $m \in \mathcal{M}$. If $\hat{m}$ exists a.s., then for any $K > 1$

$$\mathbb{E}\left[\|\mu - \hat{\mu}_{\hat{m}}\|^2\right] \leq \frac{K}{K-1} \inf_{m \in \mathcal{M}} \left[\|\mu - \mu_m\|^2 \left(1 + \frac{\mathrm{pen}(m)}{N_m}\right) + \mathrm{pen}(m)\sigma^2\right] + \Sigma \qquad (3)$$

where

$$\Sigma = \frac{K^2\sigma^2}{K-1} \sum_{m \in \mathcal{M}} (D_m + 1)\mathrm{Dkhi}\left[D_m + 1, N_m - 1, \frac{N_m - 1}{K N_m}\mathrm{pen}(m)\right].$$

Minimal penalties

- Choose $K > 1$ and $\mathcal{L} = \{L_m,\ m \in \mathcal{M}\}$ non-negative numbers (weights) such that

$$\Sigma' = \sum_{m \in \mathcal{M}} (D_m + 1) e^{-L_m} < +\infty.$$

- For any $m \in \mathcal{M}$ set

$$\mathrm{pen}^{L}_{K,\mathcal{L}}(m) \quad = \quad K \frac{N_m}{N_m - 1} \mathrm{Dkhi}^{-1} \left[D_m + 1, N_m - 1, e^{-L_m} \right]$$

- When $L_m \vee D_m \leq \kappa n$ with $\kappa < 1$:

$$\mathrm{pen}^{L}_{K,\mathcal{L}}(m) \leq C(K, \kappa) \left(L_m \vee D_m \right).$$

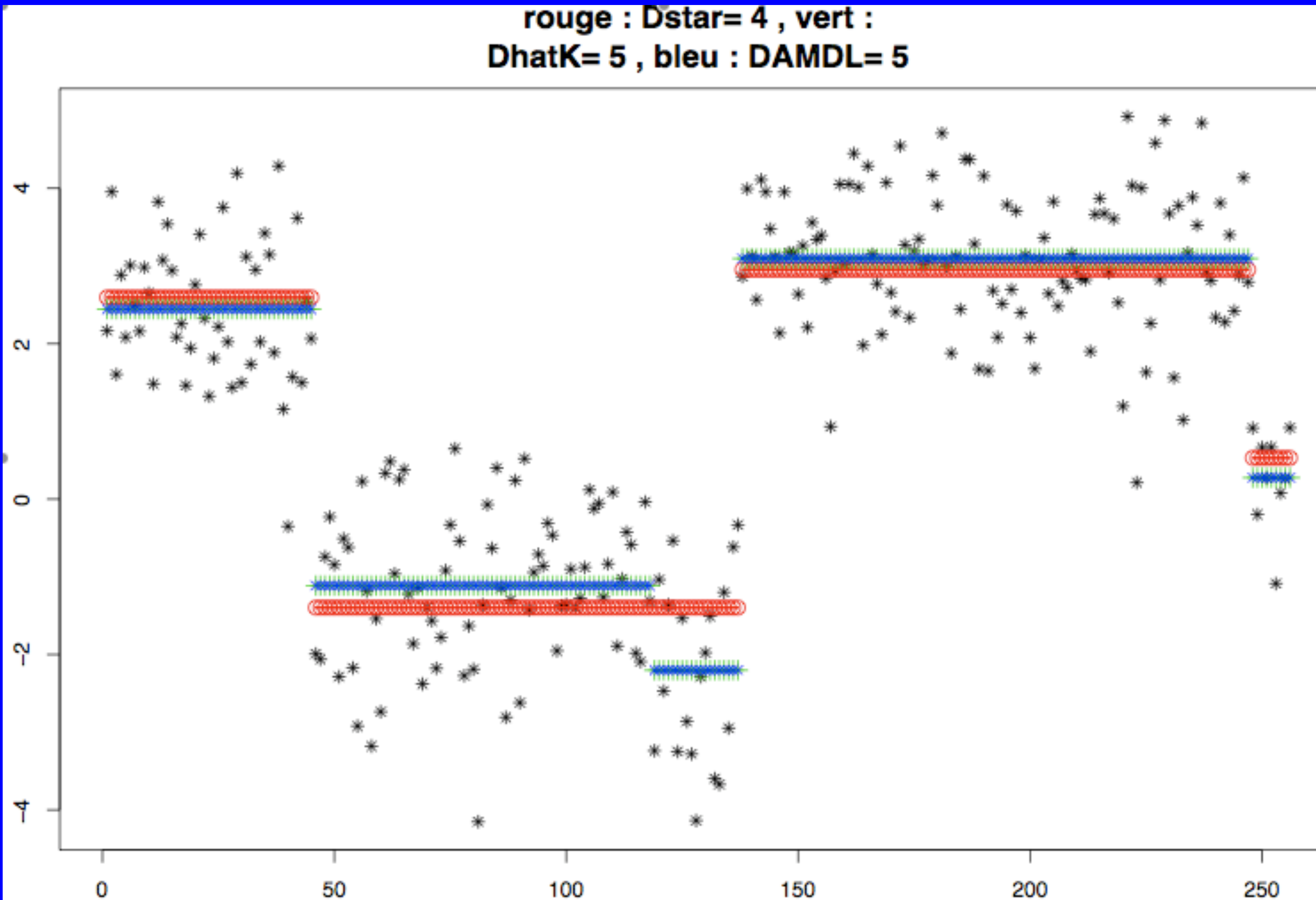
How to choose the L_m?

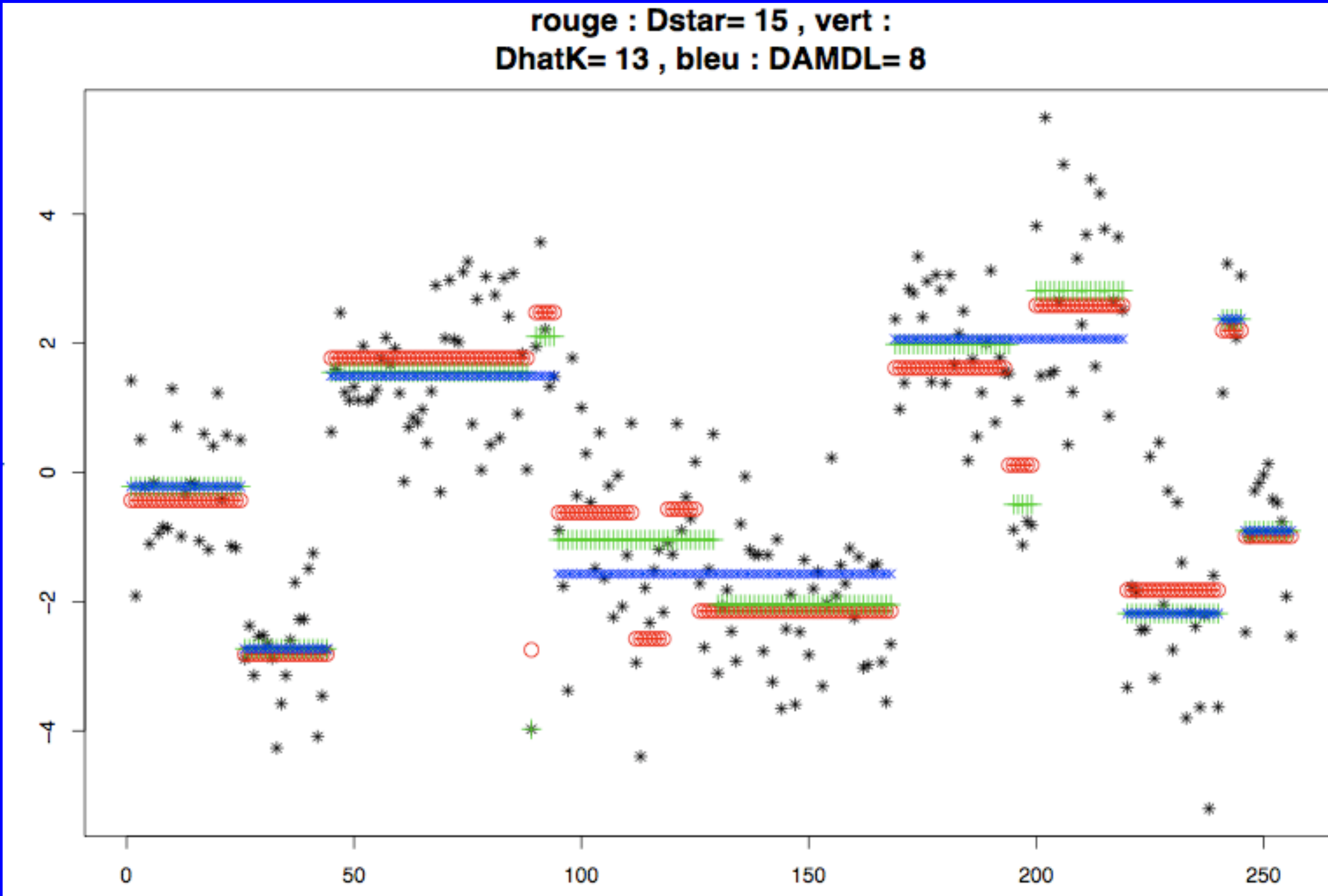
- When $\mathcal{S}$ has a complexity (M,a): a possible choice is $L_m = aD_m + 3\log(D_{m+1})$. Then

$$\Sigma' = \sum_{m\in\mathcal{M}} (D_m+1)e^{-L_m} \leq M \sum_{D\geq 1} D^{-2}$$

- For change-point detection: We choose $L_m = L(|m|) = \log\left[\binom{n}{|m|-2}\right] + 2\log(|m|)$, for which

$$\Sigma' = \sum_{D=2}^{p+1} \binom{n}{D-2} De^{-L(D)} = \sum_{D=2}^{p+1} \frac{1}{D} \leq \log(p+1).$$


rouge : Dstar= 4 , vert :
DhatK= 5 , bleu : DAMDL= 5
4
2
0
−2
−4
0
50
100
150
200
250


rouge : Dstar= 15 , vert :
DhatK= 13 , bleu : DAMDL= 8
4
2
0
-2
-4
0
50
100
150
200
250