Package CAPUSHE pour le logiciel R

Vincent Brault [1,2] & Jean-Patrick Baudry [3] & Cathy Maugis-Rabusseau [4] & Bertrand Michel [3]

[1] *Université Paris-Sud* 11
[2] *INRIA Saclay Île de France Projet* SELECT
[3] *LSTA Paris 6*
[4] *Institut de Mathématiques de Toulouse*
http://www.math.univ-toulouse.fr/∼maugis/CAPUSHE.html

1er juillet 2012

Introduction

Cadre : Sélection de modèle par minimisation d'un critère pénalisé.

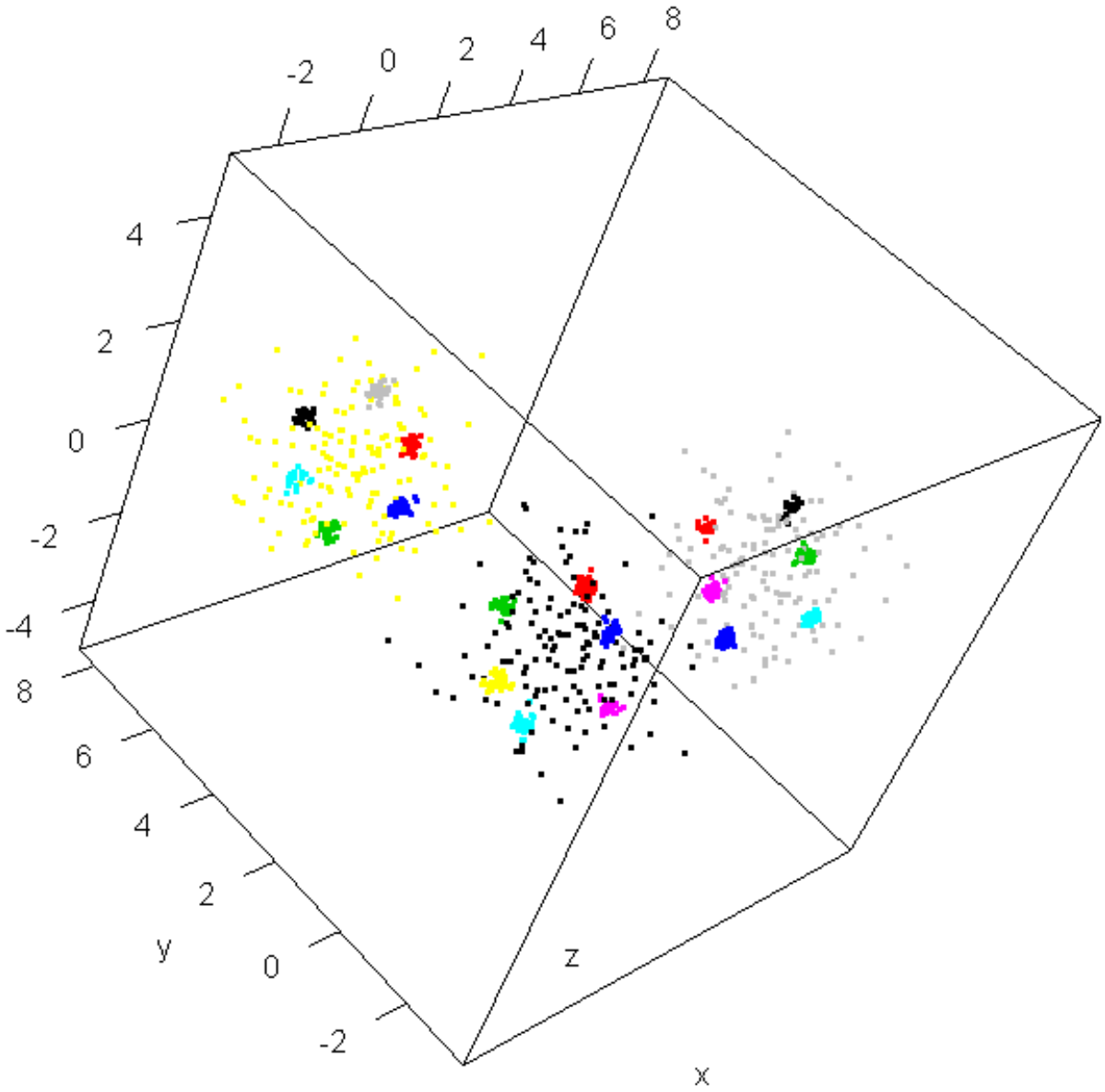
Hypothèse : La pénalité est connue à une constante près.

Objectif : Capushe permet de calibrer cette dernière.

Plan

1. Théorie
2. Heuristique de pente
3. DDSE
4. Djump
5. Capushe

Introduction



Plan

1. Théorie
2. Heuristique de pente
3. DDSE
4. Djump
5. Capushe

n-échantillon $X_1, \ldots, X_n$ de $\mathcal{X}$

Paramètre d'intérêt $s \in \mathscr{S}$

Fonction de contraste $\gamma : \mathscr{S} \times \mathcal{X} \to \mathbb{R}$ telle que

$$s \in \operatorname*{argmin}_{t \in \mathscr{S}} \mathbb{E}_X \left[\gamma(t, X)\right]$$

<u>Exemple</u> :

$n = 1000$ et $\mathcal{X} = \mathbb{R}^3$

s la densité de l'échantillon

$\gamma(s, X) = -\log s(X)$

Contraste empirique $\forall t \in \mathcal{S}, \gamma_n(t) = \frac{1}{n} \sum_{i=1}^{n} \gamma(t, X_i)$

Exemple :

$$\begin{aligned} \gamma_n(s_m) &= -\log\left(\prod_{i=1}^{n} s(X_i;\theta)\right) \\ &= -\sum_{i=1}^{n} \log s(X_i;\theta) \end{aligned}$$

Famille de modèles $(\mathcal{S}_m)_{m\in\mathcal{M}}$ de complexité respective C_m.
Estimateur $\widehat{s}_m$ de s dans $\mathcal{S}_m$:

$$\widehat{s}_m \in \underset{t\in\mathcal{S}_m}{\operatorname{argmin}}\ \gamma_n(t).$$

Exemple :

$\mathcal{S}_m$ ensemble des mélanges sphériques avec m composantes.

$$X \sim \sum_{i=1}^{m} p_i \mathcal{N}\left(\mu_i, \sigma_i^2 I_3\right)$$

$\Rightarrow C_m = 4m + (m-1)$ nombre de paramètres libres de tout mélange de $\mathcal{S}_m$

$$\gamma_n(s_m) = -\frac{1}{n}\sum_{i=1}^{n} \log\left[\sum_{i=1}^{m} p_i (2\pi)^{-3/2}\sigma_i^{-3} \exp\left(-\frac{1}{2\sigma_i^2}||X_i-\mu_i||_2^2\right)\right]$$

Modèle oracle :

$$m^\star \in \underset{m \in \mathcal{M}}{\operatorname{argmin}}\, \mathbb{E}_X \left[\ell \left(s, \widehat{s}_m \right) \right]$$

où $\ell\left(s,t\right) = \mathbb{E}_X\left[\gamma(t,X)\right] - \mathbb{E}_X\left[\gamma(s,X)\right]$

ℓ est la divergence de Kullback-Leibler

Modèle oracle :

$$m^\star \in \underset{m \in \mathcal{M}}{\operatorname{argmin}}\, \mathbb{E}_X \left[\ell\left(s, \widehat{s}_m\right)\right]$$

où $\ell\left(s,t\right) = \mathbb{E}_X\left[\gamma(t,X)\right] - \mathbb{E}_X\left[\gamma(s,X)\right]$

ℓ est la divergence de Kullback-Leibler

$$\mathrm{crit}(m) = \gamma_n(\widehat{s}_m)$$

Modèle oracle :

$$m^\star \in \operatorname*{argmin}_{m\in\mathcal{M}} \mathbb{E}_X \left[\ell\left(s, \widehat{s}_m\right)\right]$$

où $\ell\left(s,t\right) = \mathbb{E}_X\left[\gamma(t,X)\right] - \mathbb{E}_X\left[\gamma(s,X)\right]$

ℓ est la divergence de Kullback-Leibler

Fonction de pénalité pen : $\mathcal{M} \to \mathbb{R}^+$ telle que le modèle $\widehat{m}$ minimisant le critère pénalisé associé

$$\text{crit}(m) = \gamma_n(\widehat{s}_m) + \text{pen}(m)$$

ait un risque $\mathbb{E}_X\left[\ell(s, \widehat{s}_{\widehat{m}})\right]$ proche de celui de l'oracle.

Plan

1. Théorie
2. Heuristique de pente
3. DDSE
4. Djump
5. Capushe

Pénalité optimale

$$\underset{m\in\mathcal{M}}{\operatorname{argmin}}\ \mathrm{crit}_{\mathrm{opt}}(m) = \underset{m\in\mathcal{M}}{\operatorname{argmin}}\ \ell(s,\hat{s}_m)$$

La pénalité idéale serait :

$$\begin{aligned}\mathrm{crit}_{\mathrm{opt}}(m) = \ell(s,\hat{s}_m) \quad &\Leftrightarrow \quad \gamma_n(\hat{s}_m) + \mathrm{pen}_{\mathrm{opt}}(m) = \ell(s,\hat{s}_m)\\ &\Leftrightarrow \quad \mathrm{pen}_{\mathrm{opt}}(m) = \ell(s,\hat{s}_m) - \gamma_n(\hat{s}_m)\end{aligned}$$

Pénalité optimale

$$
\begin{aligned}
\mathrm{pen}_{\mathrm{opt}}(m) &= \ell(s, \hat{s}_m) - \gamma_n(\hat{s}_m) \\
&= \underbrace{\left(\mathbb{E}_X\left[\gamma(\hat{s}_m, X)\right] - \mathbb{E}_X\left[\gamma(s_m, X)\right]\right)}_{=v_m} \\
&\quad + \left(\mathbb{E}_X\left[\gamma(s_m, X)\right] - \mathbb{E}_X\left[\gamma(s, X)\right]\right) \\
&\quad + \underbrace{\left(\gamma_n(s_m) - \gamma\left(\hat{s}_m\right)\right)}_{=\widehat{v}_m} \\
&\quad - \left(\gamma_n(s_m) - \gamma_n(s)\right) \\
&\quad - \gamma_n(s) \\
&= v_m + \widehat{v}_m + \Delta_n(s_m) - \gamma_n(s)
\end{aligned}
$$

Heuristique de pente [SH1]

$$\text{pen}_\kappa = \kappa \widehat{v}_m$$

$$\begin{aligned}
\text{crit}_\kappa(m) &= \gamma_n(\widehat{s}_m) + \kappa \widehat{v}_m \\
&= \gamma_n(\widehat{s}_m) + \kappa \left(\gamma_n(s_m) - \gamma\left(\widehat{s}_m\right)\right) \\
&= (1-\kappa)\gamma_n(\widehat{s}_m) + \kappa \gamma_n(s_m)
\end{aligned}$$

Birgé et Massart (2006) :

$$\text{pen}_{\min}(m) \approx \widehat{v}_m$$

Heuristique de pente [SH2]

$$v_m \approx \widehat{v}_m$$

$$\Delta_n(s_m) \approx 0$$

Nous obtenons :

$$\begin{aligned} \mathrm{pen}_{\mathrm{opt}}(m) &= v_m + \widehat{v}_m + \Delta_n(s_m) \\ &\approx \widehat{v}_m + \widehat{v}_m \\ &\approx 2 \times \mathrm{pen}_{\mathrm{min}}(m) \end{aligned}$$

Forme de la pénalité $\text{pen}_{\text{shape}}$

La pénalité à calibrer est de la forme

$$\begin{aligned} \text{pen}_{\text{shape}}(m) &= \kappa \hat{v}_m \\ \text{pen}_{\text{min}}(m) &= \kappa_{\text{min}}\, \text{pen}_{\text{shape}}(m) \end{aligned}$$

Pour les mélanges gaussiens :

$$\text{pen}_{\text{shape}}(m) = C_m$$

Données

```
> datacapushe
   model   pen complexity contrast
1    K=1 0.004          4  7.02975
2    K=2 0.009          9  6.16194
3    K=3 0.014         14  5.22071
4    K=4 0.019         19  5.17469
5    K=5 0.024         24  5.13399
6    K=6 0.029         29  5.08707
7    K=7 0.034         34  5.05738
8    K=8 0.039         39  5.02772
9    K=9 0.044         44  4.99665
10  K=10 0.049         49  4.95389
```

Plan

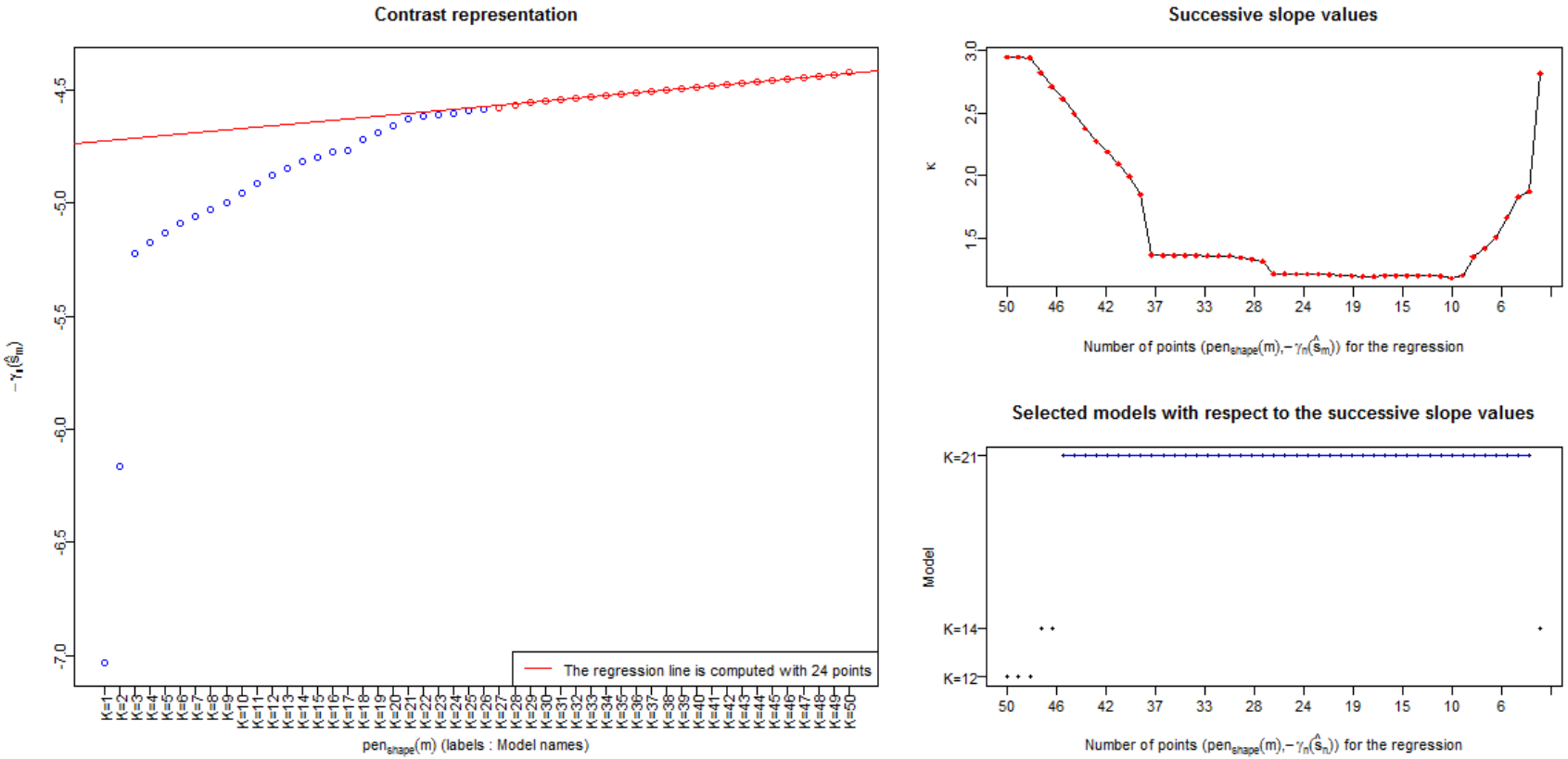
1. Théorie
2. Heuristique de pente
3. DDSE
4. Djump
5. Capushe

DDSE

$$
\begin{aligned}
\mathrm{pen}_{\min}(m) &\approx \hat{v}_m \\
&\approx \gamma_n(s_m) - \gamma_n(\hat{s}_m) \\
&\approx \gamma_n(s_m) - \gamma_n(s) + \gamma_n(s) - \gamma_n(\hat{s}_m) \\
&\approx \mathbb{E}_X\left[\ell(s, s_m)\right] + \gamma_n(s) - \gamma_n(\hat{s}_m)
\end{aligned}
$$

Plot

$$\gamma_n(\widehat{s}_m) - \gamma_n(s) \approx \mathbb{E}_X\left[\ell(s, s_m)\right] - \kappa_{\min}\mathrm{pen}_{\mathrm{shape}}(m)$$



Commande

```
> ResDDSE=DDSE(datacapushe,pct=0.15,point=0,psi.rlm=psi.bisquare,scoef=2)
```

Commande

```
> ResDDSE=DDSE(datacapushe,pct=0.15,point=0,psi.rlm=psi.bisquare,scoef=2)
> ResDDSE
[1] "K=21"
```

Commande

```
> ResDDSE=DDSE(datacapushe,pct=0.15,point=0,psi.rlm=psi.bisquare,scoef=2)
> ResDDSE
[1] "K=21"
> summary(ResDDSE)

Selected model: K=21

Corresponding Slope interval: [1.18;2.62](88% of points)

Selected model complexity: 104
```

Validation

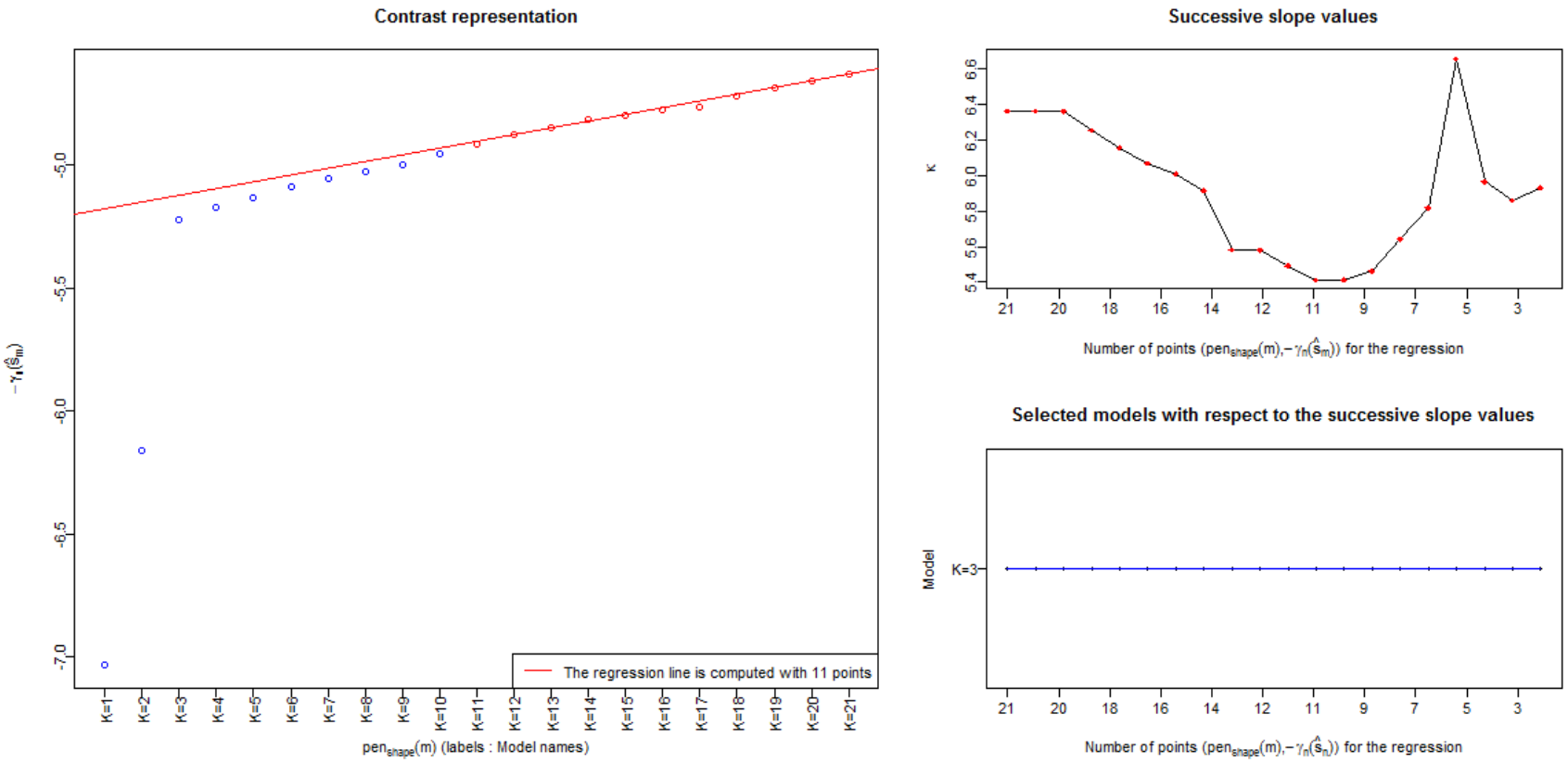
```
> ResDDSEpartial=DDSE(datapartialcapushe)
```

Validation

```
> ResDDSEpartial=DDSE(datapartialcapushe)
> plot(ResDDSEpartial)
```

Validation

```
> ResDDSEpartial=DDSE(datapartialcapushe)
> plot(ResDDSEpartial)
```

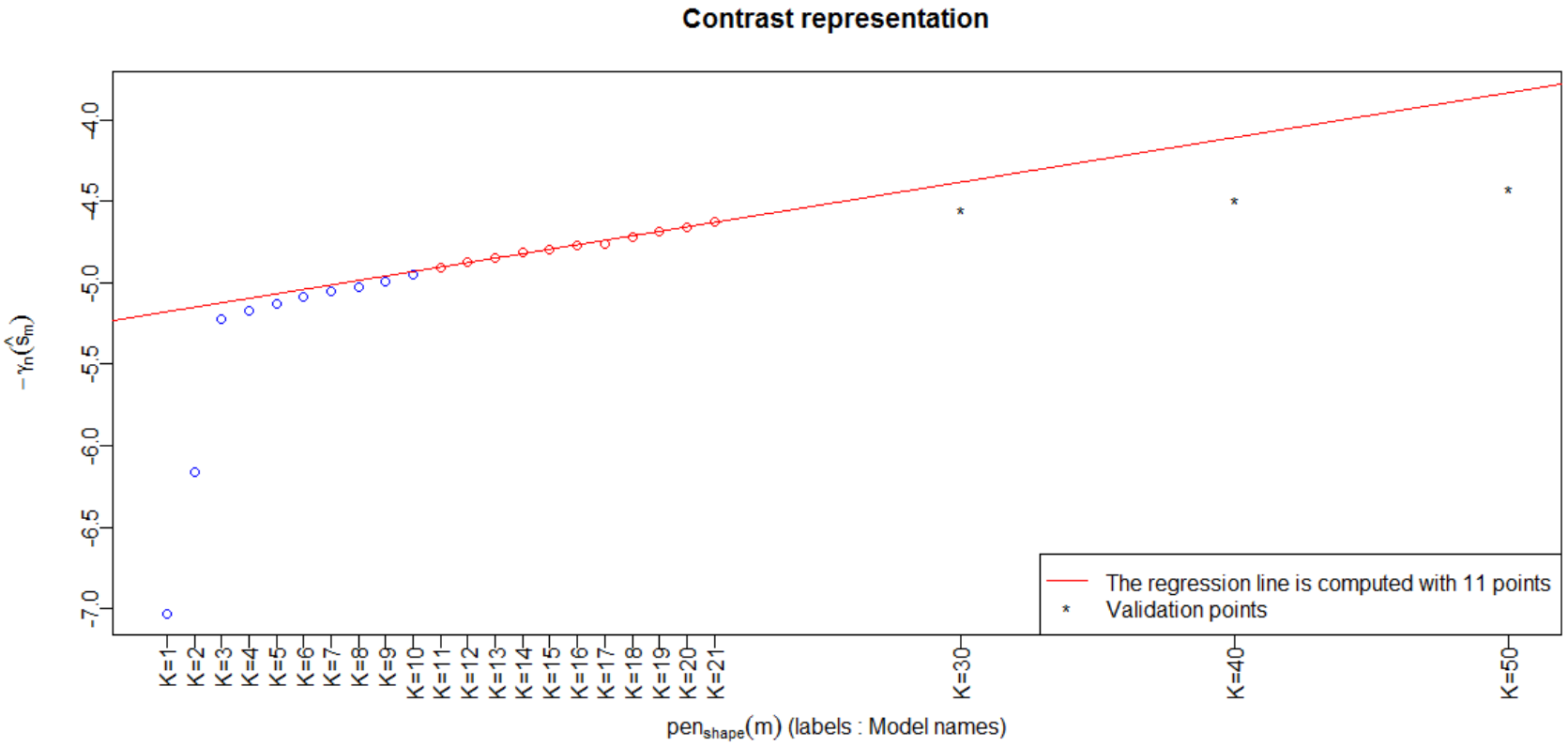


Validation

```
> ResDDSEpartial=DDSE(datapartialcapushe)
> plot(ResDDSEpartial)
> validation(ResDDSEpartial,datavalidcapushe)
```

Validation

```
> ResDDSEpartial=DDSE(datapartialcapushe)
> plot(ResDDSEpartial)
> validation(ResDDSEpartial,datavalidcapushe)
```

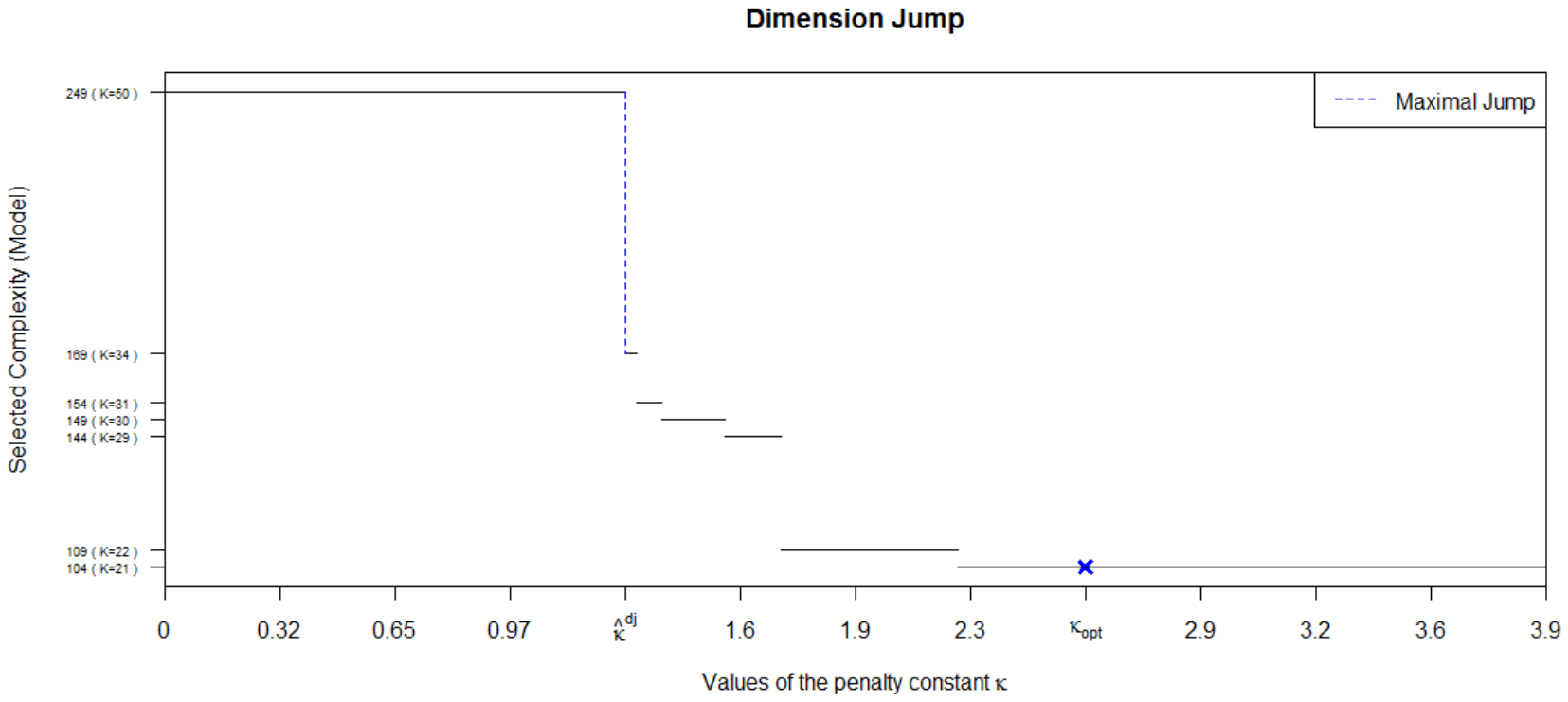


Plan

1. Théorie
2. Heuristique de pente
3. DDSE
4. Djump
5. Capushe

Djump

$C_{m(\kappa)}$ la complexité du modèle minimisant le critère $\text{crit}_\kappa(\cdot)$
$\kappa \mapsto C_{m(\kappa)}$ doit faire un saut au voisinage du κ_{min}

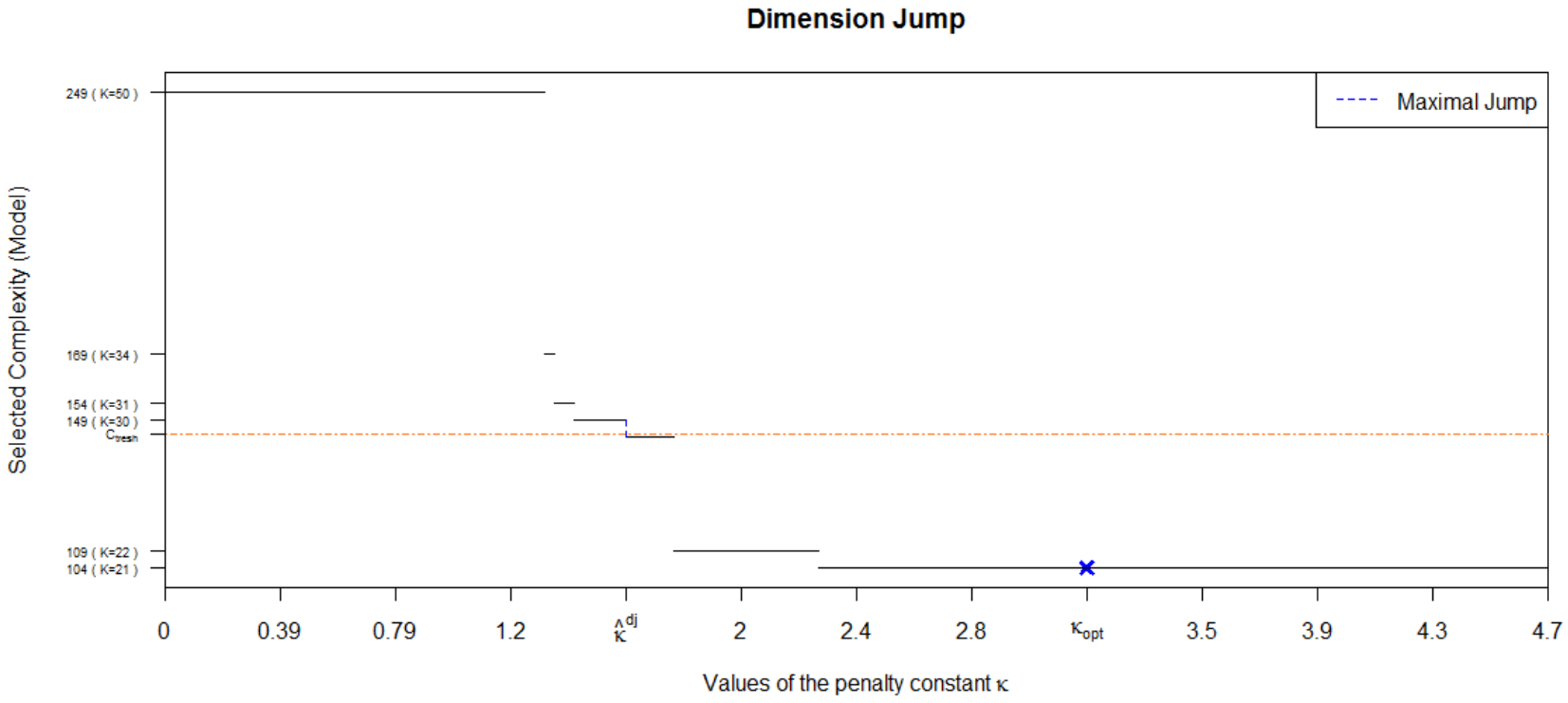


C_{tresh}

```
> plot(Djump(datacapushe,Ctresh=1000/log(1000)))
```

C_{tresh}

```
> plot(Djump(datacapushe,Ctresh=1000/log(1000)))
```

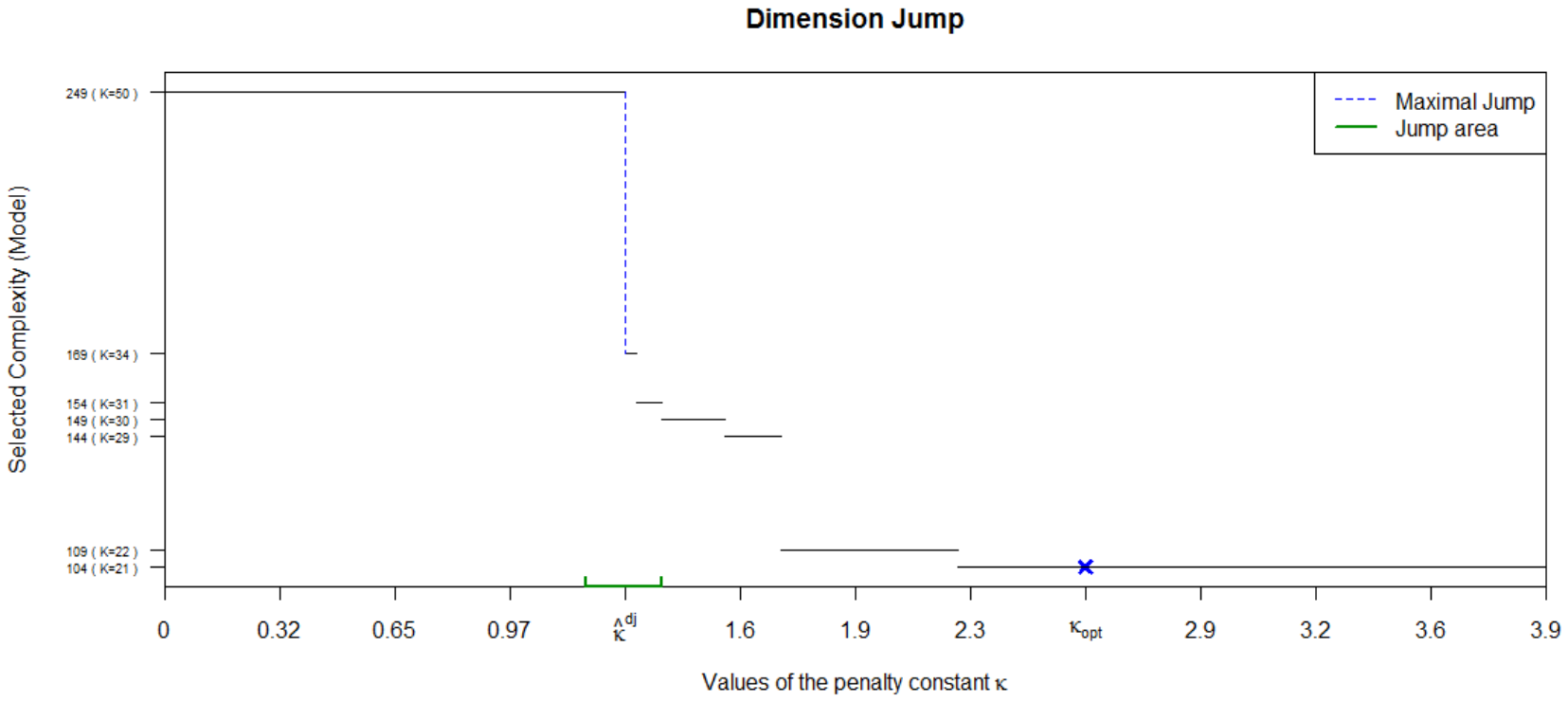


$C_{AreaJump}$

```
> plot(Djump(datacapushe,Careajump=sqrt(log(1000)/1000)))
```

$C_{AreaJump}$

```
> plot(Djump(datacapushe,Careajump=sqrt(log(1000)/1000)))
```



Commande

```
> ResDjump=Djump(datacapushe,scoef=2,Careajump=sqrt(log(1000)/1000),Ctresh=0)
```

Commande

```
> ResDjump=Djump(datacapushe,scoef=2,Careajump=sqrt(log(1000)/1000),Ctresh=0)
> ResDjump
[1] "K=21"
```

Commande

```
> ResDjump=Djump(datacapushe,scoef=2,Careajump=sqrt(log(1000)/1000),Ctresh=0)
> ResDjump
[1] "K=21"
> summary(ResDjump)

Selected model: K=21

Estimated penalty constant: 1.29
Selected model complexity: 104

Jump area: [1.18;1.4]
Model selected by jump area: K=21
```

Plan

1. Théorie
2. Heuristique de pente
3. DDSE
4. Djump
5. Capushe

Commande

```
> ResCapushe=capushe(datacapushe,n=1000)
```

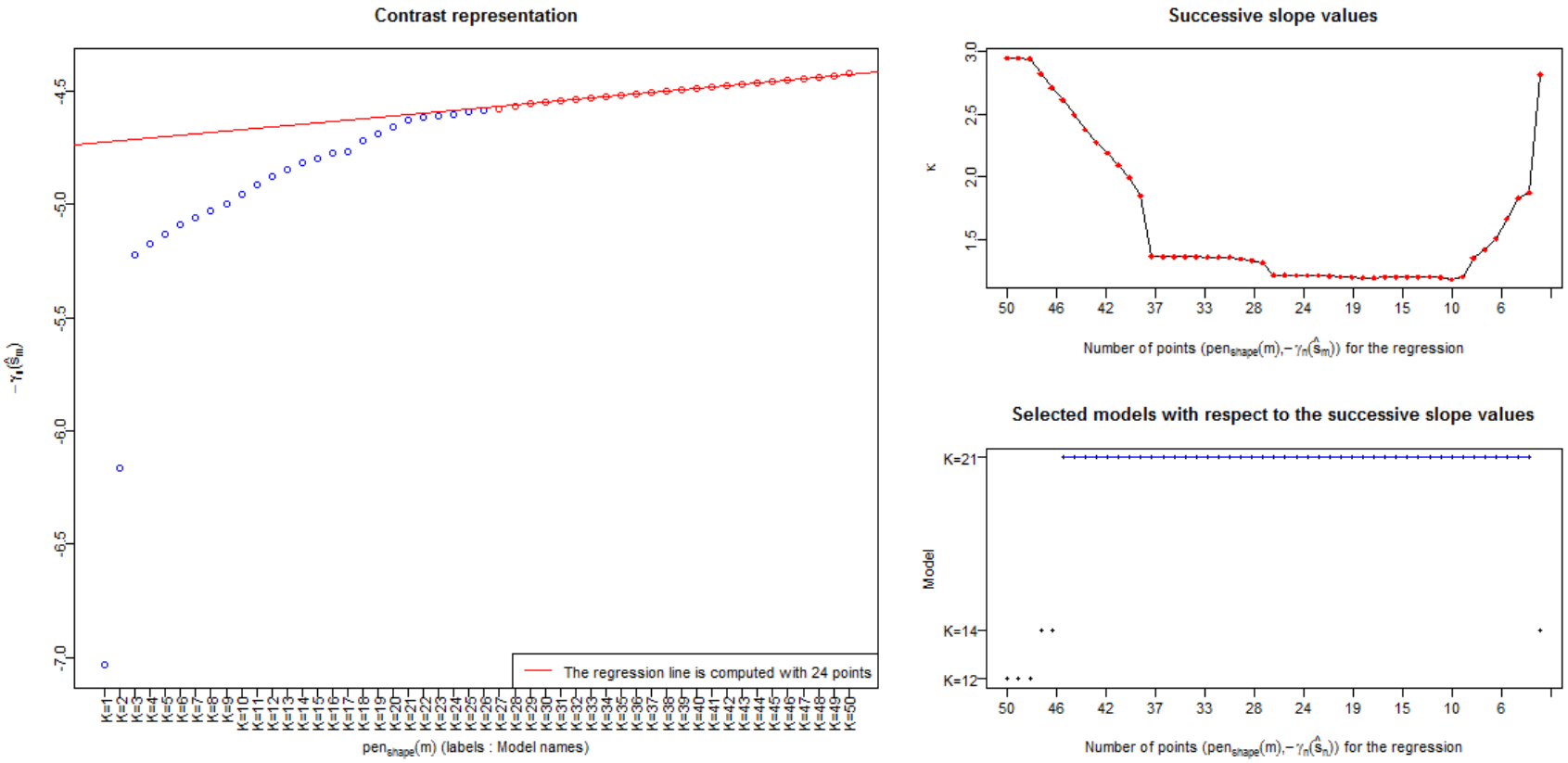
Commande

```
> ResCapushe=capushe(datacapushe,n=1000)
> summary(ResCapushe)

Model selected by DDSE: K=21

Corresponding Slope interval: [1.18;2.62](88% of points)
Selected model complexity: 104


Model selected by Djump: K=21

Estimated penalty constant: 1.29
Selected model complexity: 104

Jump area: [1.18;1.4]
Model selected by jump area: K=21


Model selected by AIC: K=50

Model selected by BIC: K=21
```

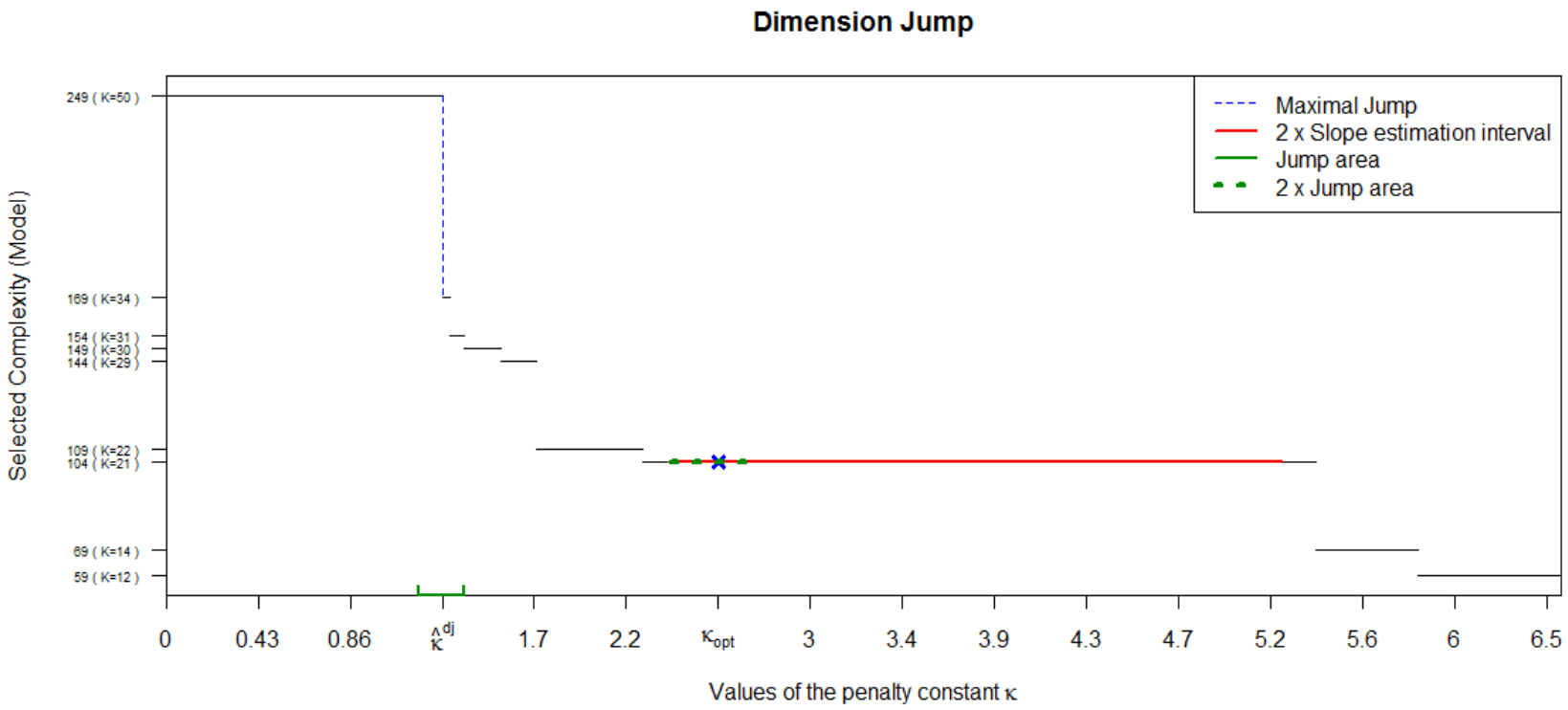
Commande

```
> ResCapushe=capushe(datacapushe,n=1000)
> summary(ResCapushe)

Model selected by DDSE: K=21

Corresponding Slope interval: [1.18;2.62](88% of points)
Selected model complexity: 104


Model selected by Djump: K=21

Estimated penalty constant: 1.29
Selected model complexity: 104

Jump area: [1.18;1.4]
Model selected by jump area: K=21


Model selected by AIC: K=50

Model selected by BIC: K=21
> plot(ResCapushe)
```

Plot


Contrast representation
The regression line is computed with 24 points
pen_shape(m) (labels : Model names)
Successive slope values
κ
Number of points (pen_shape(m), −γ_n(ŝ_m)) for the regression
Selected models with respect to the successive slope values
Model
K=21
K=14
K=12
Number of points (pen_shape(m), −γ_n(ŝ_m)) for the regression

Plot


Dimension Jump
Maximal Jump
2 x Slope estimation interval
Jump area
2 x Jump area
Selected Complexity (Model)
249 (K=50)
169 (K=34)
154 (K=31)
149 (K=30)
144 (K=29)
109 (K=22)
104 (K=21)
69 (K=14)
59 (K=12)
0
0.43
0.86
$\hat{\kappa}^{dj}$
1.7
2.2
κ_{opt}
3
3.4
3.9
4.3
4.7
5.2
5.6
6
6.5
Values of the penalty constant κ

Question ?

Birgé, L. et Massart, P. (2001). Gaussian model selection. *Journal of the European Mathematical Society*, 3(3) :203-268.

Birgé, L. et Massart, P. (2006). Minimal penalties for Gaussian model selection. *Probability Theory and Related Fields*, 138(1-2) :33-73.

Lebarbier, E. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing*, 85(4) :717-736.

Massart, P. (2007). *Concentration Inequalities and Model Selection. École d'été de Probabilités de Saint-Flour 2003. Lecture Notes in Mathematics*. Springer.

Maugis, C. et Michel, B. (2011). A non asymptotic penalized criterion for Gaussian mixture model selection. *ESAIM* : P & S, 15, p 41-68.

Baudry, J.-P., Maugis, C. and Michel, B. (2011) Slope Heuristics : Overview and Implementation. *Statistics and Computing*, 22(2), 455-470.