

Impact of noise on the detection of chromosomal abnormalities in array CGH measurements

Bénédicte Naudet-Briand, Delphine Maucort-Boulch et Pascal Roy

27 mai 2010

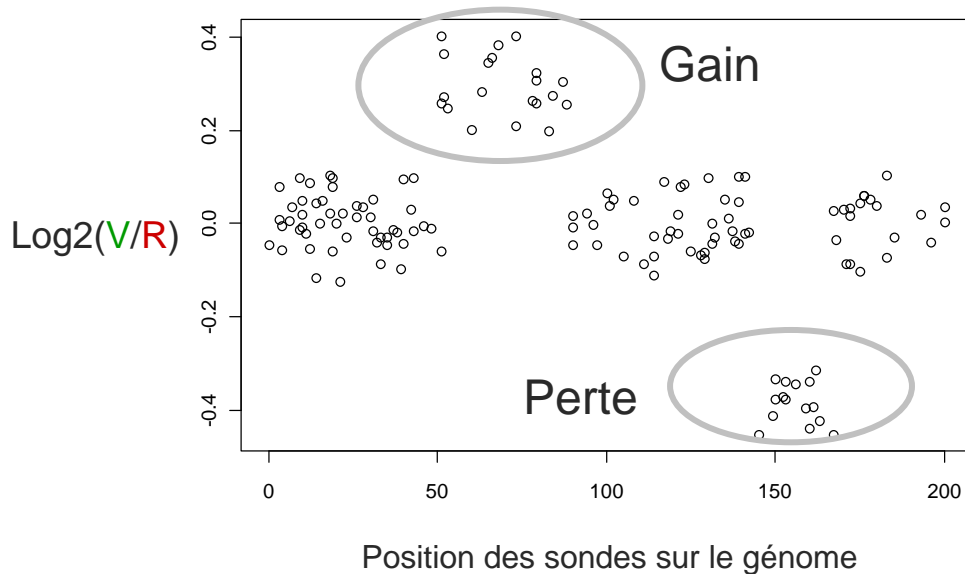
4ème Conférence d'Epidémiologie Clinique 2010

XVIème Réunion des Statisticiens des Centres de Luttés contre le Cancer

Contexte

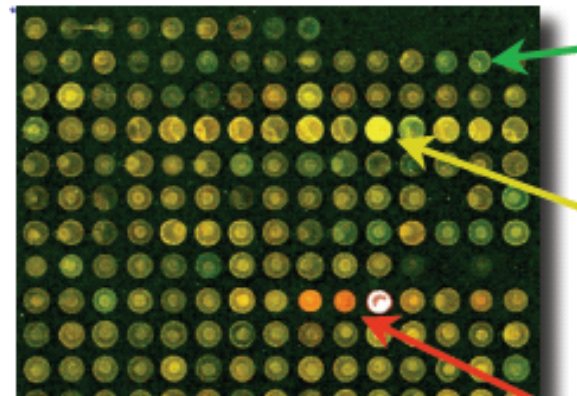
☐ Puces CGH

- ☐ Marquage des ADN
- ☐ Hybridation sur la puce
- ☐ Lecture de la lame par un scanner
- ☐ Analyse des rapports de fluorescence



ADN test

ADN témoin



Vert : ADN test > ADN témoin

Jaune : ADN test = ADN témoin

Rouge : ADN test < ADN témoin

Contexte

- Puces CGH

- Sources de variabilité
 - Biologique (signal)
 - Technique (traitement de l'ADN + erreur de mesure) → Bruit

- Selon le niveau de bruit, le signal est difficilement identifiable
 - Un niveau de bruit élevé peut fausser les résultats d'une analyse statistique d'un profil CGH

Objectif

- ❑ Étudier quel niveau de bruit empêche la détection d'anomalies

- ❑ Deux approches sont proposées
 - Analyse sur données simulées
 - Analyse sur données réelles

- ❑ Deux méthodes de segmentation sont utilisées
 - CGHseg (Picard et al., 2005)
 - CBS (Olshen et al., 2004)

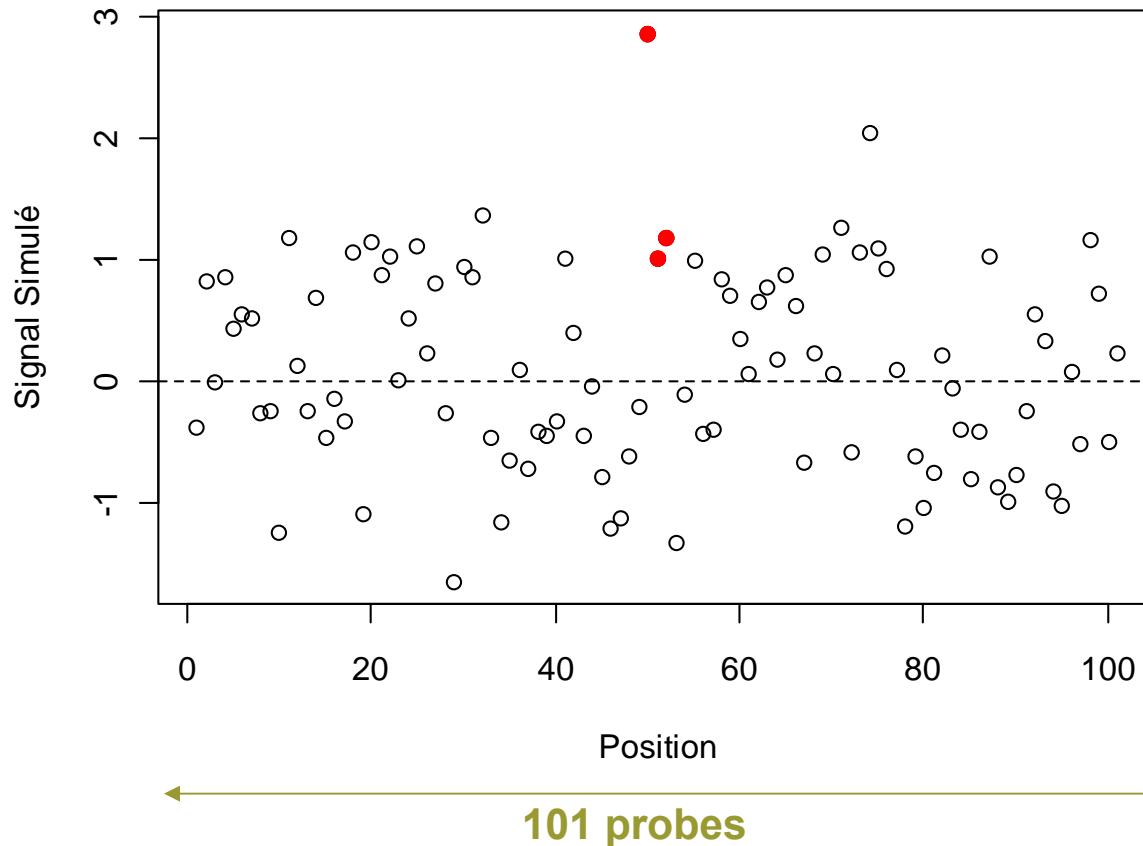
Matériels et méthodes

□ Données simulées

- $48 \times 6 \times 9 = 2592$ profils CGH (101 probes)
 - 48 anomalies centrées (3 à 97 probes, par 2 probes)
 - 6 différents nombre de copies (1, 3, 4, 5, 6 et 7)
 - 9 niveaux de bruit (0.1 à 0.9, par 0.1 unité)

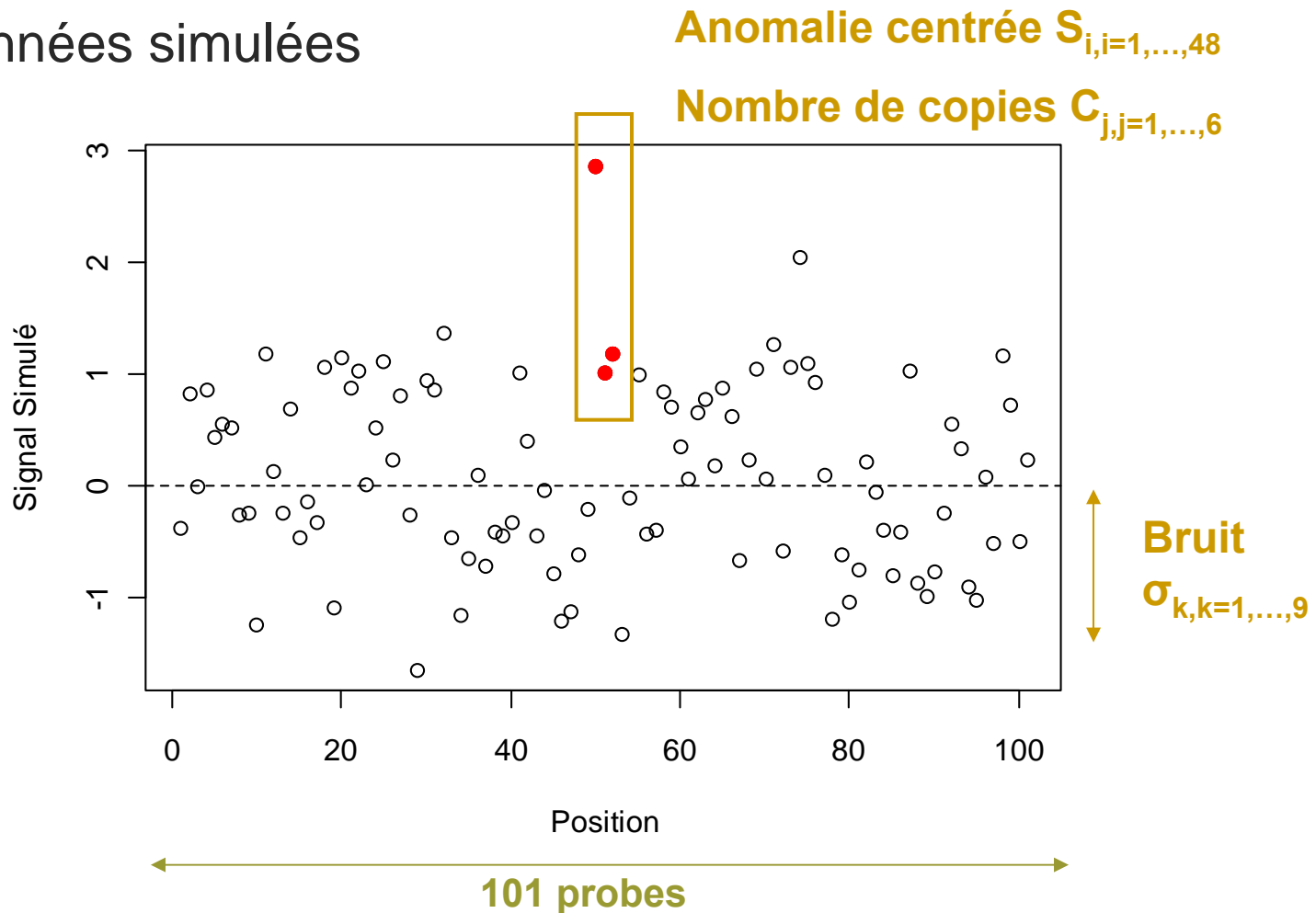
Matériels et méthodes

□ Données simulées



Matériels et méthodes

□ Données simulées



Matériels et méthodes

□ Données réelles

Abnormality Number	1	2	3	4	5	6	7
Patient	A	B	C	D	E	F	F
Chromosomal position	2q12.1	22q11.21	9q21.1	Xq11	16p11.2	2q35	2q35
Type of abnormality	Duplication	Duplication	Deletion	Deletion	Deletion	Triplication	Duplication
Length of the abnormality	2.2Mb	1.3 Mb	219.65 Kb	1.29 Mb	577.9 Kb	974.2 Kb	471.6 Kb

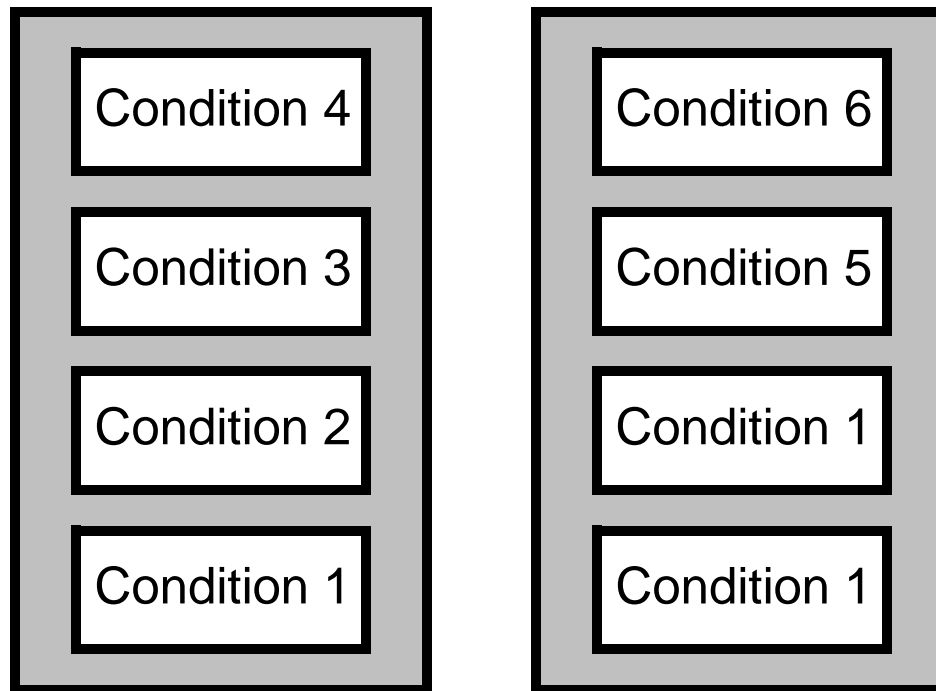
Matériels et méthodes

□ Données réelles

- Condition 1 : Normale
 - Condition 2 : Ajout d'1 μ L de chloroforme
 - Condition 3 : ADN nettoyé avec DNA Clean and Concentrator
 - Condition 4 : Fragmentation de l'ADN 30 min à 95°C
 - Condition 5 : Fragmentation de l'ADN 60 min à 95°C
 - Condition 6 : Ajout d'1 μ L d'acétate de sodium
-
- Plateforme Agilent (puces CGH 4x44K)
 - Hybridation des patients les uns contre les autres : B|A, C|D et E|F
 - Deux patients dans toutes les conditions → 6 puces soit 2 lames

Matériels et méthodes

□ Données réelles



Matériels et méthodes

□ Données réelles : Une mesure pour caractériser le bruit

■ Y_1, \dots, Y_n valeurs des log ratios successifs aux positions X_1, \dots, X_n

■ $Y_i = \mu_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma)$ pour tout $i = 1, \dots, n$

■ $Z = (Z_1, \dots, Z_{n-1})$ où $Z_i = Y_{i+1} - Y_i$

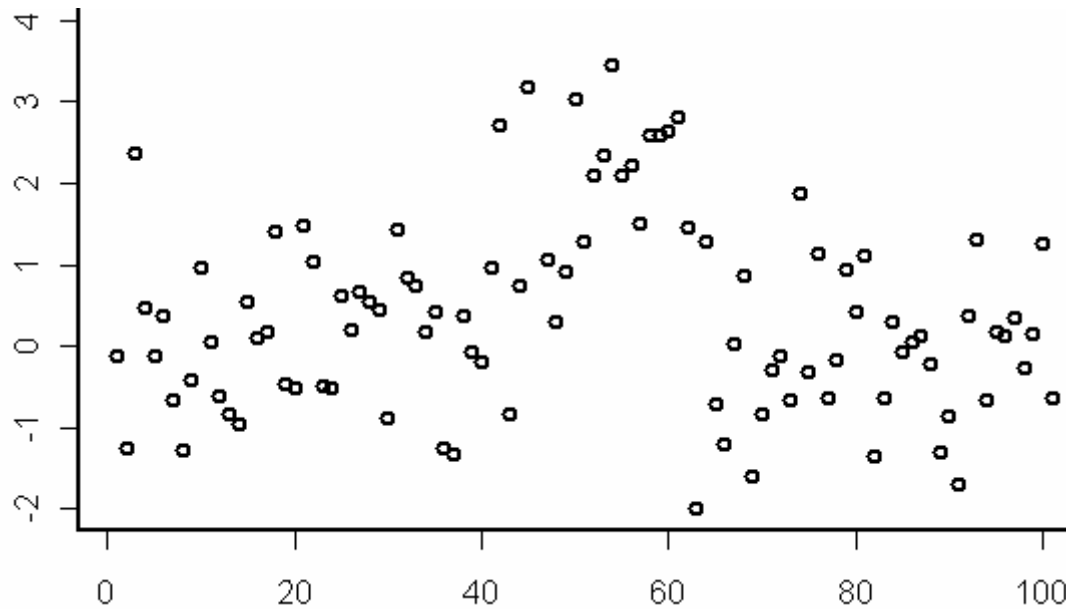
■ Une estimation de σ est donnée par :
$$\hat{\sigma} = \frac{\hat{IQR}(Z)}{\sqrt{2} \cdot IQR(N(0,1))}$$

→ DLRS (Derivative Log-Ratio Spread), Agilent Genomic Workbench®

Matériels et méthodes

☐ Méthodes de segmentation

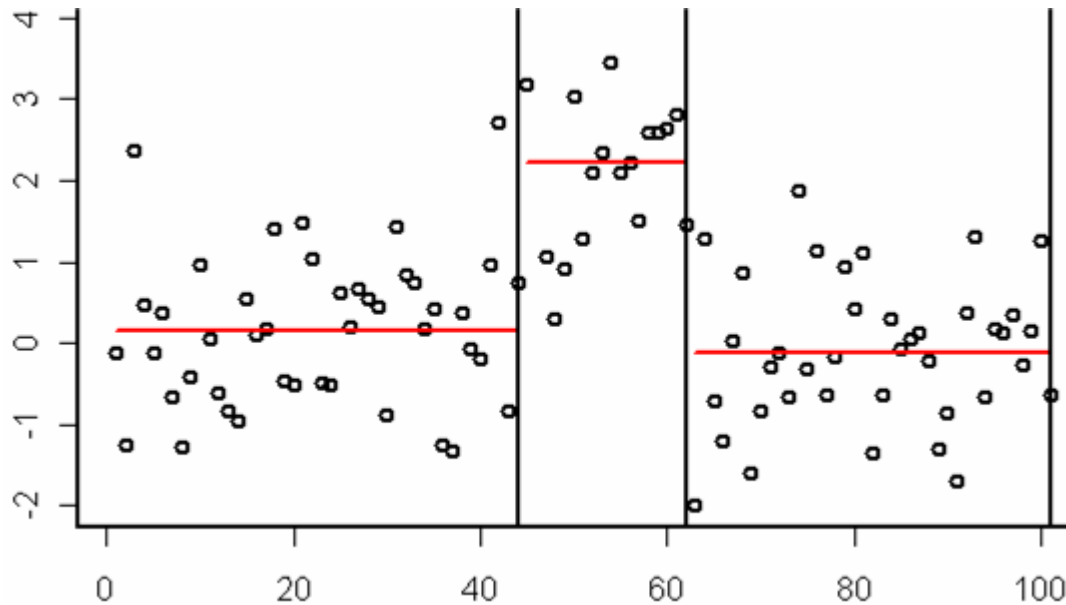
- CGHseg (Picard et al., 2005)
- CBS (Olshen et al., 2004)



Matériels et méthodes

☐ Méthodes de segmentation

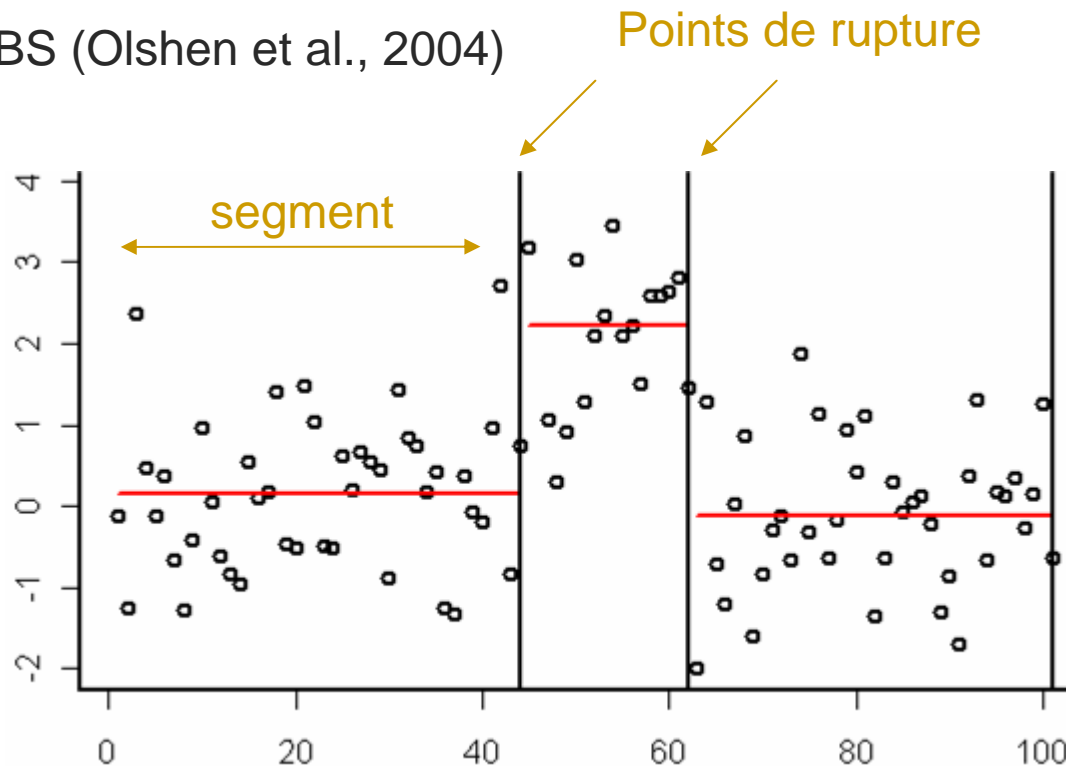
- CGHseg (Picard et al., 2005)
- CBS (Olshen et al., 2004)



Matériels et méthodes

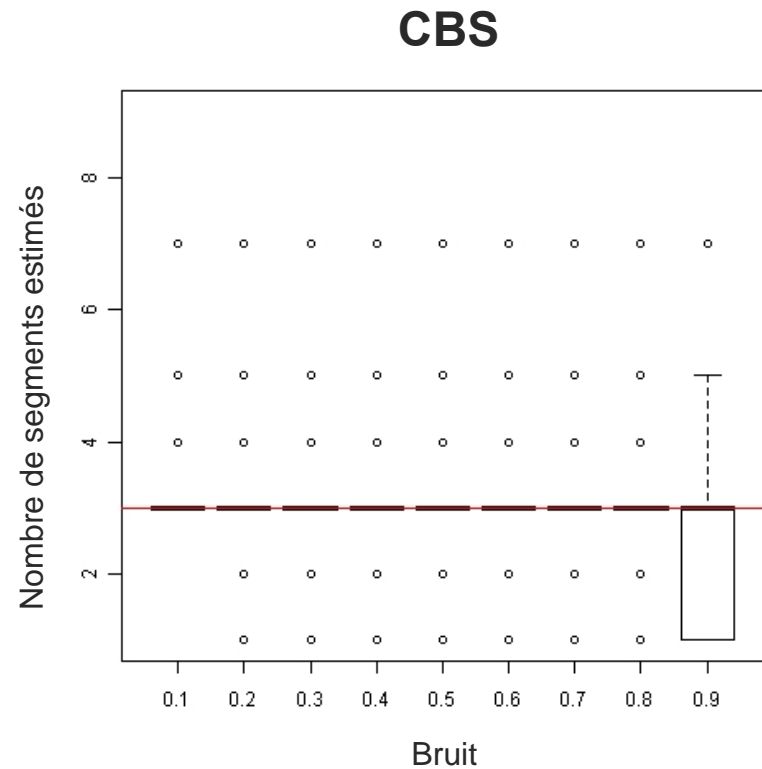
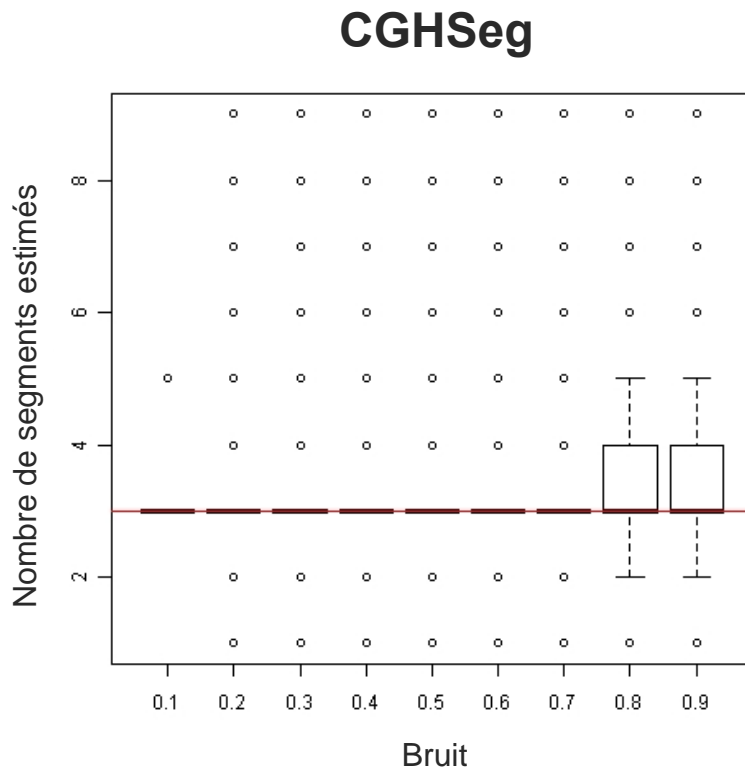
☐ Méthodes de segmentation

- CGHseg (Picard et al., 2005)
- CBS (Olshen et al., 2004)



Résultats – Données simulées

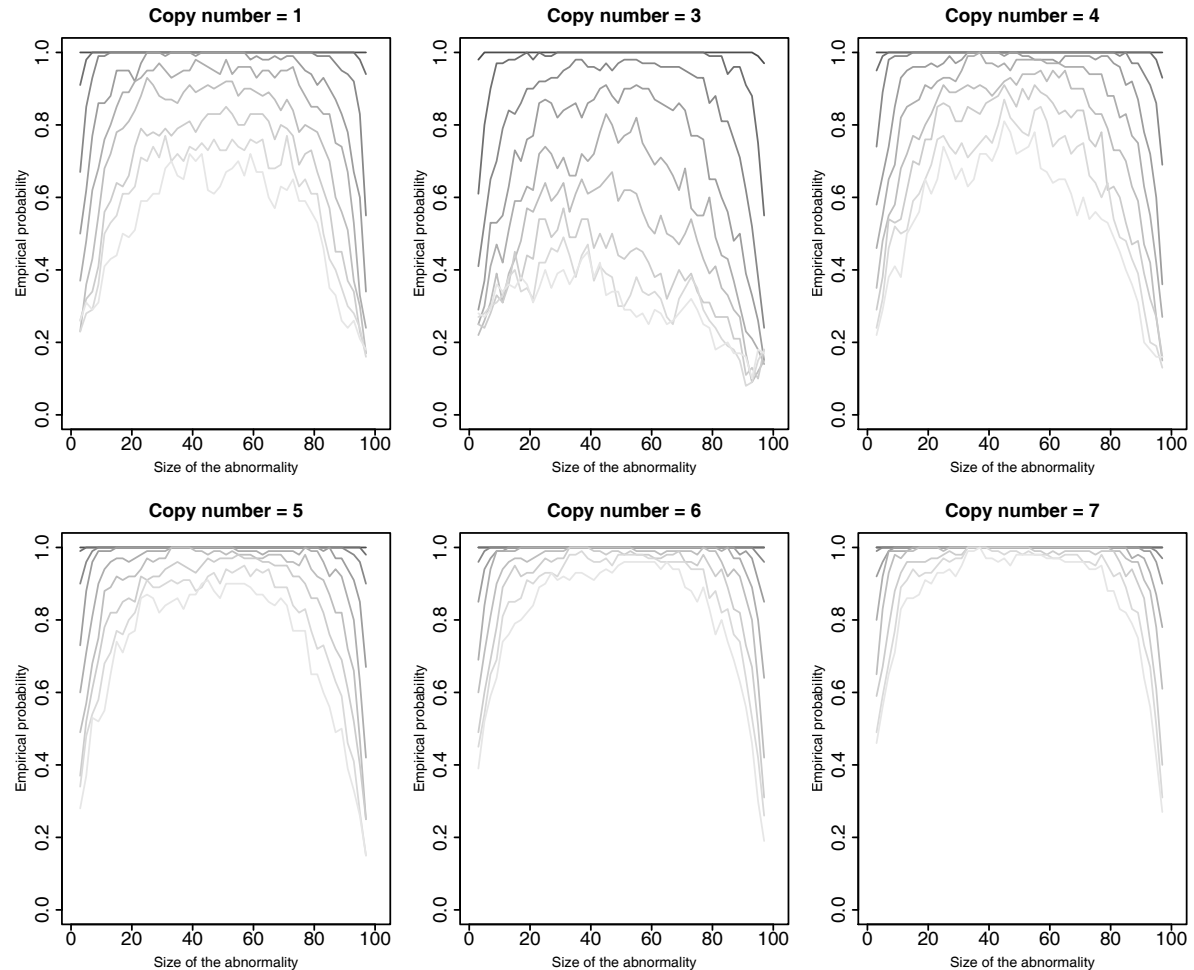
- Nombre de segments estimés par les deux méthodes



Résultats – Données simulées

□ Probabilité de détecter le nombre correct de segments

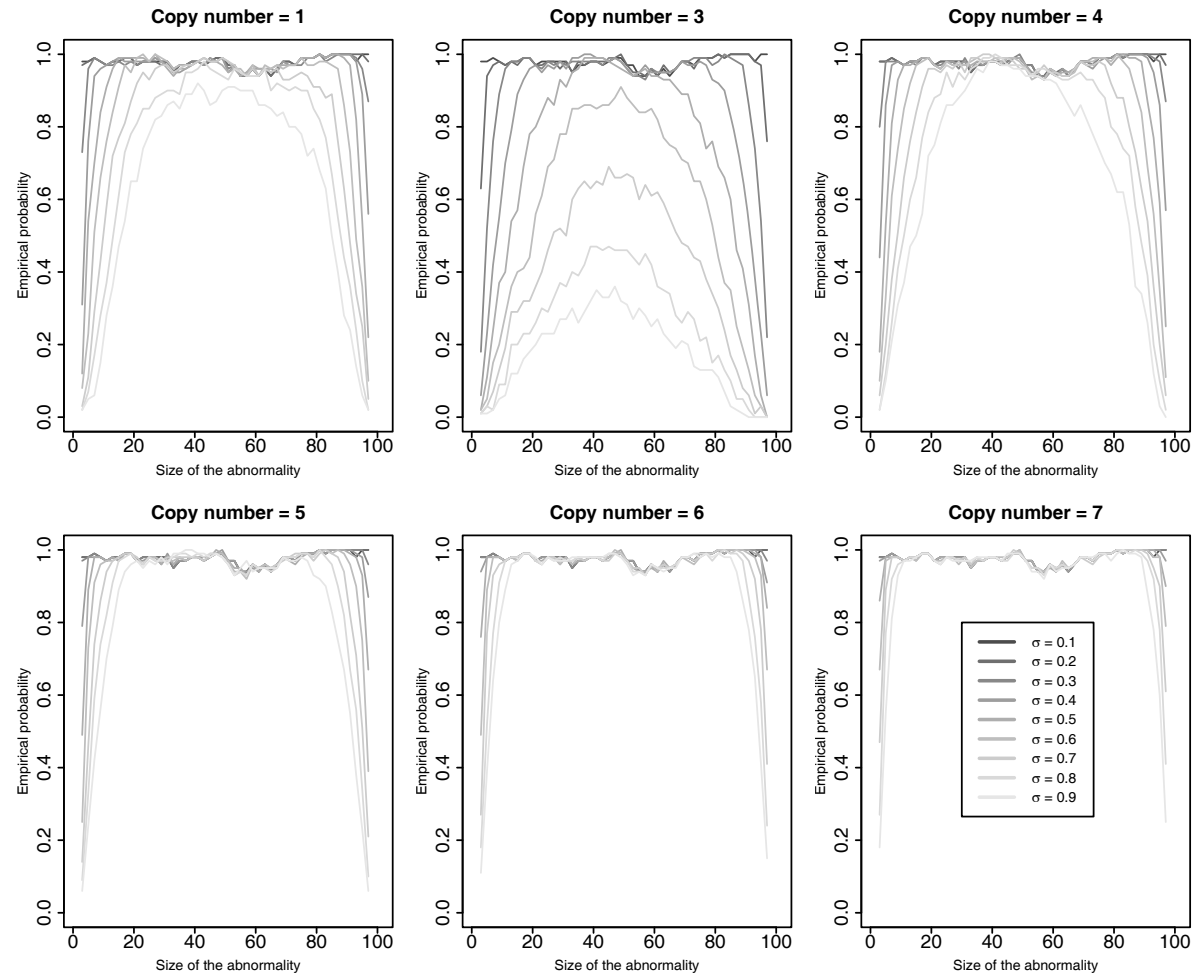
CGHSeg



Résultats – Données simulées

□ Probabilité de détecter le nombre correct de segments

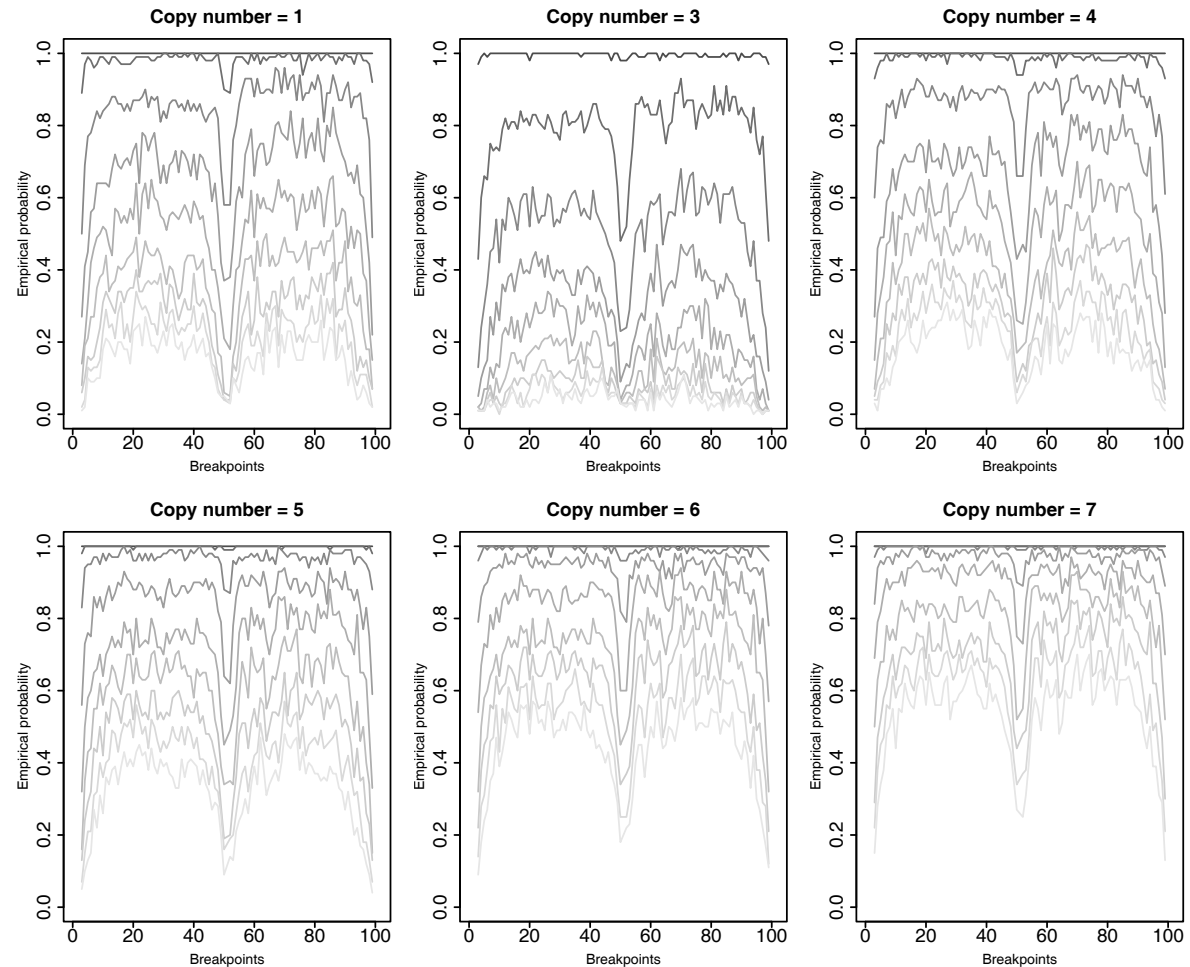
CBS



Résultats – Données simulées

□ Probabilité de détecter le nombre correct de segments et de prédire correctement les points de rupture

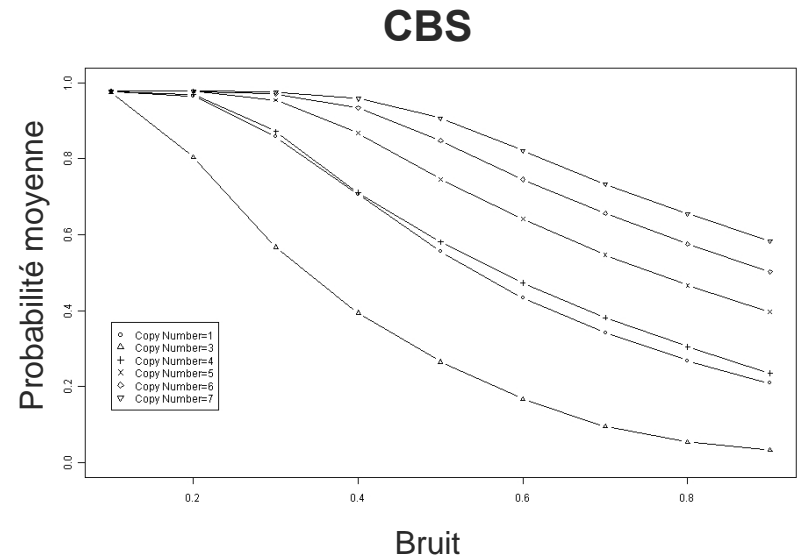
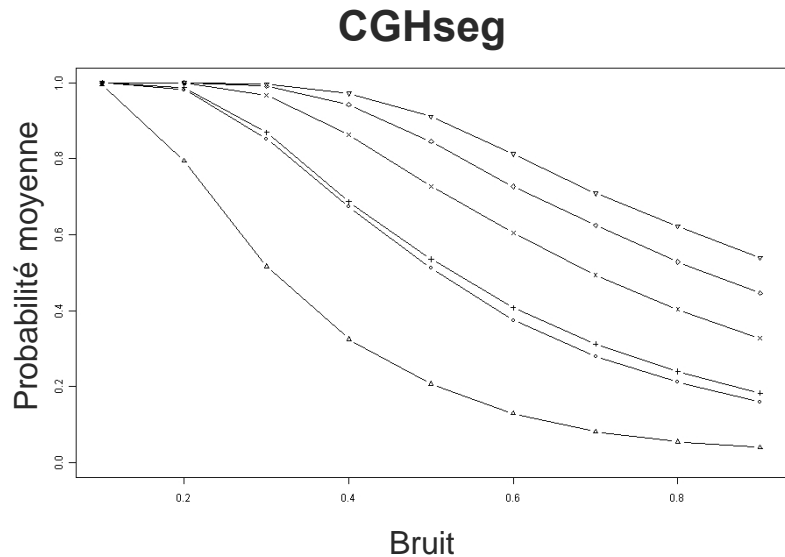
CGHseg



Résultats – Données simulées

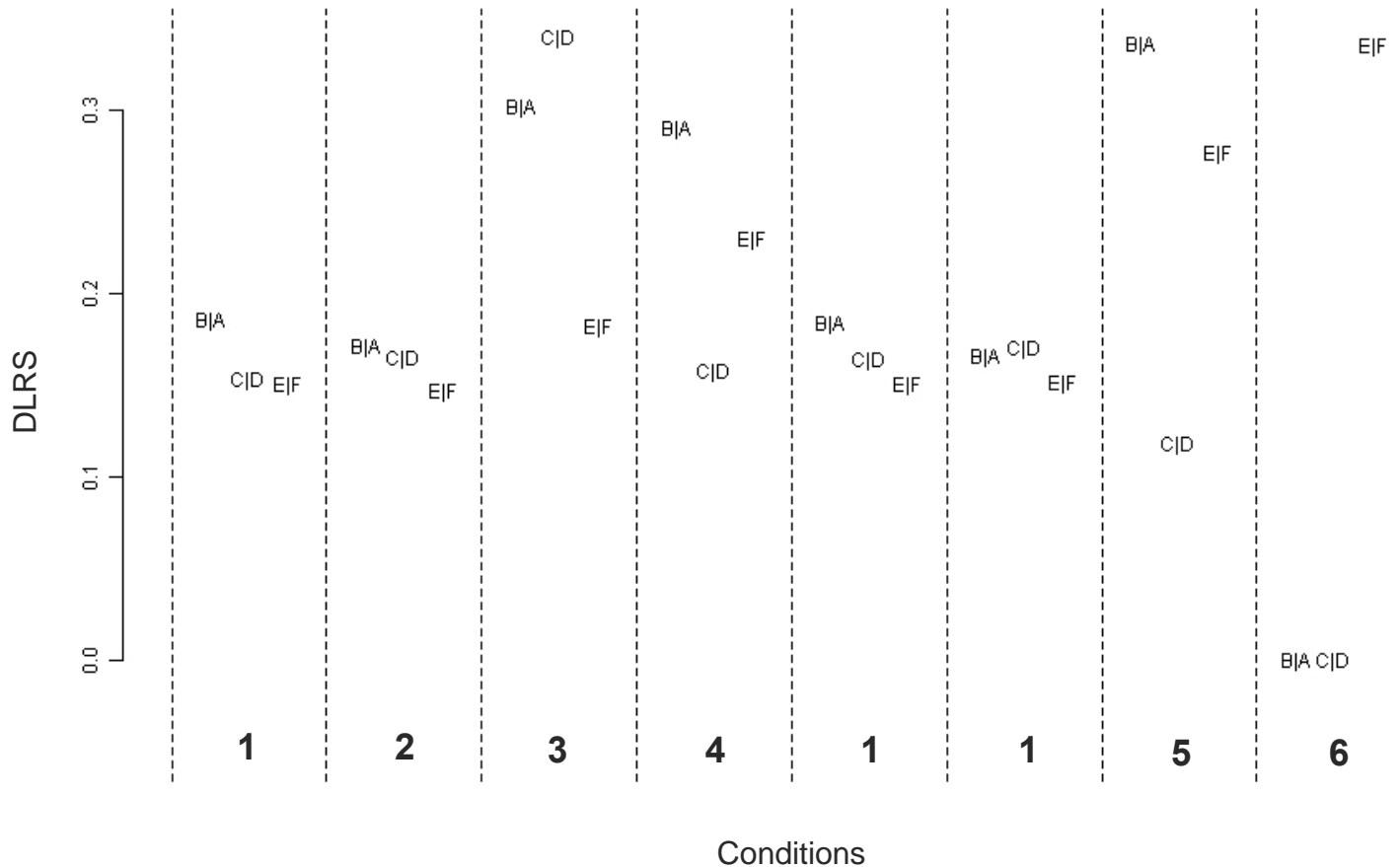
□ Synthèse des résultats

Probabilité moyenne de détecter le nombre correct de segments et de prédire correctement les points de ruptures



Résultats – Données réelles

Estimation du bruit : DLRS



Résultats – Données réelles

Conditions	1	2	3	4	1	1	5	6
------------	---	---	---	---	---	---	---	---

CGHseg

Anomaly Number	Hybridization	1	2	3	4	1	1	5	6
7	(E F)	21 : 1	21 : 1	21 : -2	22 : 1	22 : 1	22 : 1	22 : 1	
6	(E F)	-1 : -3	-1 : -3	-1 : -6	0 : -3	0 : -3	0 : -3	0 : -3	0 : 0
5	(E F)	-1 : 1	-1 : 0	-1 : 0	-1 : 2	-1 : 0	-1 : 1	-1 : 1	-1 : 0
4	(C D)	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1		
3	(C D)	0 : 0	0 : 0	0 : -29	0 : 0	0 : 0	0 : 0		
2	(B A)	-1 : 2	-1 : 2	-1 : 2	-1 : -1	-1 : 2	-1 : 2	-1 : -1	
1	(B A)	0 : 0	0 : 0	3 : 2	0 : 0	0 : 0	0 : 0	0 : 0	

CBS

Anomaly Number	Hybridization	1	2	3	4	1	1	5	6
7	(E F)	21 : 1	1 : 1	21 : -2	22 : 1	1 : 1	1 : 1	22 : 1	
6	(E F)	-1 : -3	-1 : -3	-1 : -6	0 : -3	0 : -3	0 : -3	0 : -3	0 : 0
5	(E F)	-1 : 1	-1 : 0	-1 : 0	-1 : 2	-1 : 0	-1 : 1	-1 : 1	-1 : 0
4	(C D)	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1		
3	(C D)				0 : 0	0 : 0			
2	(B A)	-1 : 2	-1 : 2	-4 : 2	-1 : -1	-1 : 2	-1 : -1	-1 : -1	
1	(B A)	3 : 0	1 : 0	3 : 2	0 : 0	2 : 0	0 : 0	0 : 0	

Résultats – Données réelles

Conditions		1	2	3	4	1	1	5	6
CGHseg									
Anomaly Number	Hybridization								
7	(E F)	21 : 1	21 : 1	21 : -2	22 : 1	22 : 1	22 : 1	22 : 1	
6	(E F)	-1 : -3	-1 : -3	-1 : -6	0 : -3	0 : -3	0 : -3	0 : -3	0 : 0
5	(E F)	-1 : 1	-1 : 0	-1 : 0	-1 : 2	-1 : 0	-1 : 1	-1 : 1	-1 : 0
4	(C D)	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1		
3	(C D)	0 : 0	0 : 0	0 : -29	0 : 0	0 : 0	0 : 0		
2	(B A)	-1 : 2	-1 : 2	-1 : 2	-1 : -1	-1 : 2	-1 : 2	-1 : -1	
1	(B A)	0 : 0	0 : 0	3 : 2	0 : 0	0 : 0	0 : 0	0 : 0	
CBS									
Anomaly Number	Hybridization								
7	(E F)	21 : 1	1 : 1	21 : -2	22 : 1	1 : 1	1 : 1	22 : 1	
6	(E F)	-1 : -3	-1 : -3	-1 : -6	0 : -3	0 : -3	0 : -3	0 : -3	0 : 0
5	(E F)	-1 : 1	-1 : 0	-1 : 0	-1 : 2	-1 : 0	-1 : 1	-1 : 1	-1 : 0
4	(C D)	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1		
3	(C D)				0 : 0	0 : 0			
2	(B A)	-1 : 2	-1 : 2	-4 : 2	-1 : -1	-1 : 2	-1 : -1	-1 : -1	
1	(B A)	3 : 0	1 : 0	3 : 2	0 : 0	2 : 0	0 : 0	0 : 0	

Résultats – Données réelles

Conditions		1	2	3	4	1	1	5	6
CGHseg									
Anomaly Number	Hybridization								
7	(E F)	21 : 1	21 : 1	21 : -2	22 : 1	22 : 1	22 : 1	22 : 1	
6	(E F)	-1 : -3	-1 : -3	-1 : -6	0 : -3	0 : -3	0 : -3	0 : -3	0 : 0
5	(E F)	-1 : 1	-1 : 0	-1 : 0	-1 : 2	-1 : 0	-1 : 1	-1 : 1	-1 : 0
4	(C D)	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1		
3	(C D)	0 : 0	0 : 0	0 : -29	0 : 0	0 : 0	0 : 0		
2	(B A)	-1 : 2	-1 : 2	-1 : 2	-1 : -1	-1 : 2	-1 : 2	-1 : -1	
1	(B A)	0 : 0	0 : 0	3 : 2	0 : 0	0 : 0	0 : 0	0 : 0	

CBS									
Anomaly Number	Hybridization								
7	(E F)	21 : 1	1 : 1	21 : -2	22 : 1	1 : 1	1 : 1	22 : 1	
6	(E F)	-1 : -3	-1 : -3	-1 : -6	0 : -3	0 : -3	0 : -3	0 : -3	0 : 0
5	(E F)	-1 : 1	-1 : 0	-1 : 0	-1 : 2	-1 : 0	-1 : 1	-1 : 1	-1 : 0
4	(C D)	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1	0 : 1		
3	(C D)				0 : 0	0 : 0			
2	(B A)	-1 : 2	-1 : 2	-4 : 2	-1 : -1	-1 : 2	-1 : -1	-1 : -1	
1	(B A)	3 : 0	1 : 0	3 : 2	0 : 0	2 : 0	0 : 0	0 : 0	

Conclusion

□ Données Simulées

- Quelque soit le niveau de bruit, le nombre de segments prédit est correct
- Les segments anormaux petits ou grand et proche des bornes du profil sont plus difficilement détectés par la méthode CBS en présence d'un niveau de bruit élevé
- Les points de rupture sont prédits de façon équivalente par les deux méthodes mais plus de difficulté à les estimer en présence d'un niveau de bruit élevé

□ Données Réelles

- Quelque soit le niveau de bruit, CGHseg est plus performante que CBS
 - Détection des anomalies
 - Détection de petit segments

Conclusion

- La détection de segments anormaux dépend :
 - Du niveau de bruit
 - De l'ampleur de l'anomalie (nombre de copies)



Merci de votre attention