



**HAL**  
open science

## Estimation de la concentration d'un agent biologique par détection de rupture sur vidéos de fluorescences issues de PCR

Abou Keita, Romain Hérault, Stéphane Canu

► **To cite this version:**

Abou Keita, Romain Hérault, Stéphane Canu. Estimation de la concentration d'un agent biologique par détection de rupture sur vidéos de fluorescences issues de PCR. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00656568

**HAL Id: hal-00656568**

**<https://hal.science/hal-00656568>**

Submitted on 17 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation de la concentration d'un agent biologique par détection de rupture sur vidéos de fluorescences issues de PCR.

M. KEITA Abou<sup>1</sup>

M. HERAULT Romain<sup>1</sup>

M. CANU Stephane<sup>1</sup>

<sup>1</sup>LITIS - INSA ROUEN

LITIS EA 4108, INSA de Rouen, Avenue de l'université BP 8  
76801 Saint Etienne du Rouvray

abou.keita@insa-rouen.fr / romain.herault@insa-rouen.fr / stephane.canu@insa-rouen.fr

## Résumé

*L'estimation de la concentration d'un agent pathologique est une étape primordiale pour déclencher une alerte en cas de menace biologique. En appliquant une PCR associée à un marqueur fluorescent sur un échantillon à analyser, nous observons une émission lumineuse qui est fonction de la concentration de l'agent. Cette concentration est alors déduite du nombre de cycles de réaction à partir duquel la fluorescence change de comportement. Dans le cadre du projet GENEASE, nous proposons d'utiliser une méthode de détection de rupture, le CUSUM, qui estime ce nombre de cycles. Nous comparons ces résultats à ceux trouvés par la méthode du maximum de la dérivée seconde, méthode qui est utilisée dans les appareils industriels et des laboratoires. Nous obtenons des résultats comparables, mais contrairement à cette dernière, sans disposer du signal complet et en minimisant le retard à la détection.*

## Mots Clef

Détection de rupture, Retard à la détection, Signal de fluorescence, PCR, Reconnaissance d'agents biologiques.

## Abstract

*Estimating concentration of a pathological agent is an essential step for triggering an alert in case of biological threat. By applying a PCR associated with a fluorescent marker, we can observe that emitted light depends on the concentration of the agent. This quantity can be deduced from the number of reaction cycles at which the fluorescence changes in behavior. As part of the GENEASE project, we propose to use the CUSUM method in order to detect the rupture. This method allows to estimate when the rupture occurs in term of cycle number. We compare these results with those found by the method of the second derivative maximum, method that is used in industrial appliances and laboratories. We obtain similar results ; however, compare to the latter, we do not need a complete signal which brings lower detection delay.*

## Keywords

Rupture detection, Detection delay, Fluorescence signal, PCR, Recognition of biological agents.

## 1 Introduction

L'émergence de la menace biologique est devenue une réalité, avec notamment les attaques à l'Anthrax en 2001, qui a conduit à la mise en place du plan Biotox complétant le dispositif de l'Organisation de la Réponse de Sécurité Civile (ORSEC) de gestion des risques et de la Crise face à une menace biologique. Dans ce cadre, la détection au plus tôt d'une agression et sa caractérisation de manière fiable permet une réponse optimale en terme d'organisation de secours et de dispositions à adopter afin d'en limiter les conséquences. A ce jour, il n'existe pas de système de surveillance et d'analyse biologique en continu de l'environnement.

Le projet GENEASE<sup>1</sup>, auquel nous participons, porte sur l'étude d'un système compact et sensible de surveillance et d'analyse biologique en semi-continu de l'environnement permettant de détecter et d'identifier plusieurs agents biologiques simultanément par biologie moléculaire.

Une *Polymerase Chain Reaction* (PCR) va être utilisée à cette fin. La quantité de matériel génétique est doublée à chaque cycle de la réaction. Si un marqueur fluorescent est lié à ce matériel, nous observons une émission lumineuse croissant exponentiellement. Nous pouvons alors déduire du point de montée de fluorescence la concentration initiale du matériel génétique.

Diverses méthodes de calcul de la concentration sont utilisées en laboratoires et dans l'industrie. Elles sont principalement basées sur le calcul d'un point caractéristique du signal de fluorescence, appelé  $C_t$ , point exprimé en nombre de cycles de réaction. Ce  $C_t$  est une fonction log linéaire de la concentration notée  $C_0$ . Après étalonnage, nous pouvons donc déduire de  $C_t$  la concentration  $C_0$ . Cependant

1. Projet financé par l'ANR et composé de Bertin Technologies, du CEA LETI, du CEA SBTN et du LITIS, cf. Remerciements

ces méthodes nécessitent la connaissance complète du signal de fluorescence.

Pour notre application, la détection doit s'effectuer le plus rapidement possible après que le signal de fluorescence ait atteint le point caractéristique,  $C_t$ , pour éventuellement déclencher une alerte. Ainsi, on doit à la fois être confiant dans le nombre de cycles retenus mais aussi minimiser le retard à la détection. Deux objectifs *a priori* contradictoires. De plus, nous souhaitons mettre en place une méthode robuste sur un équipement de terrain car les signaux de fluorescence peuvent être très bruités.

De ce fait, nous proposons une technique statistique robuste de détection de rupture, basée sur le CUSUM [1]. Dédit d'un rapport de vraisemblance dynamique entre deux populations, le nombre de cycles  $C_t$  est obtenu avec une confiance satisfaisante.

Nous allons d'abord faire un état de l'art en présentant le signal de fluorescence avec deux techniques de détermination du  $C_t$ , puis l'approche CUSUM et enfin son application sur les signaux de fluorescences.

## 2 Présentation d'un signal de fluorescence

Les données dont nous disposons sont obtenues par le suivi en temps réel d'une *Polymerase Chain Reaction* (PCR) [5, 6, 3, 10]. Il s'agit d'une technique découverte par Kary Mullis en 1983 [7]. C'est un outil universel dans le domaine de la biologie pour la détection et les quantifications des acides nucléiques.

Le suivi en temps-réel d'une réaction de PCR peut être représenté sous la forme d'une courbe de fluorescence (fig. 2). Ce signal correspond à la moyenne sur une zone d'intérêt de la lumière fluorescente captée par caméra à la fin de chaque cycle de réaction, chaque cycle durant approximativement 1 minute.

Les signaux de fluorescences sur lesquels nous travaillons sont issus de l'immuno-analyse d'agents biologiques, à partir d'appareils de laboratoire (CEA-LETI et LHVP<sup>2</sup>). Le CEA-LETI fabrique une puce « Smart Drop » sur laquelle se fait cette analyse biologique [2].

La cinétique de cette réaction met en jeu trois phases : une phase d'initiation, une phase exponentielle et une phase plateau. La position de la montée du signal, lorsque l'on passe de la phase d'initiation à la phase exponentielle est fonction de la concentration de l'espèce à qualifier dans la goutte analysée. Cette position est repérée par un nombre de cycles appelé  $C_t$ .

Il existe différentes méthodes de détermination du  $C_t$ , par exemple la méthode du maximum de la dérivée seconde (fig. 3(b)) et la méthode dite du seuil direct (fig. 3(a)) [6, 12]. La concentration  $C_0$  est log linéaire par rapport au cycle  $C_t$ . Un étalonnage préalable pour calculer les paramètres de la fonction linéaire, paramètres différents pour chaque méthode et condition d'expérimentation, permet de

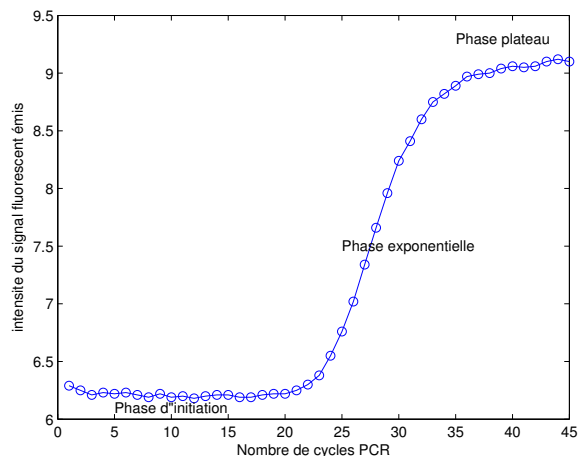


FIGURE 2 – Suivi en temps réel d'une réaction de fluorescence qui correspond à la moyenne de la fluorescence sur une région d'intérêt.

déduire  $C_0$  de  $C_t$  [9]. Il est ici à noter que ces deux méthodes sont consistantes si les deux séries de  $C_t$  trouvées montrent uniquement un biais linéaire.

Cependant ces méthodes requièrent la connaissance de l'intégralité du signal de fluorescence. L'un des buts de notre étude est de déterminer le point de montée avec un minimum de retard et sans la connaissance complète du signal. De ce fait nous allons utiliser la technique du CUSUM qui est une des méthodes de détection de changement sur un ensemble de données.

Sur l'ensemble des signaux de fluorescences (fig. 1(a) par exemple), nous observons deux types de comportements, des signaux avec un effet de saut, signifiant la présence de l'agent, et d'autres sans saut, signifiant son absence (fig. 1(b)). L'utilisation de cette technique nous permettra de distinguer ces deux cas et d'estimer l'instant de rupture.

**Remarque :** Nous comparerons nos résultats à ceux obtenus par la méthode de la dérivée seconde qui est actuellement la méthode utilisée par le CEA-LETI et le LHVP.

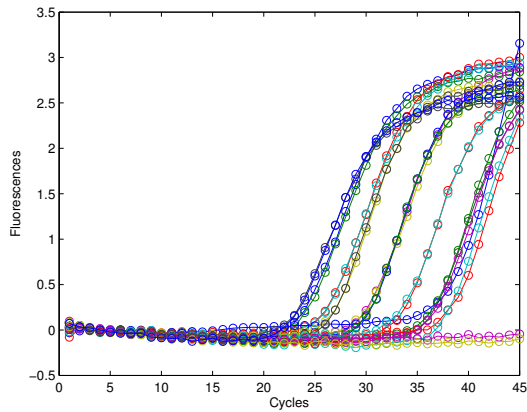
## 3 Présentation du CUSUM

### 3.1 Approche

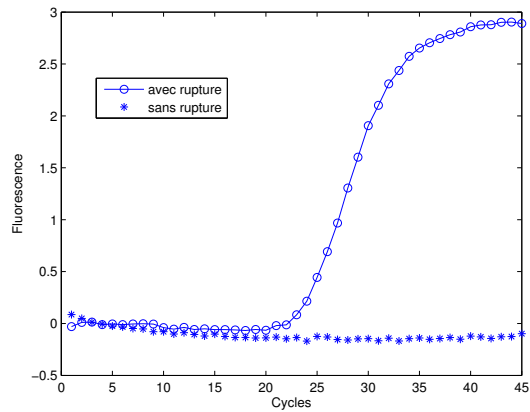
Le *Cumulative SUM* (CUSUM) [1, 8] est l'une des techniques statistiques utilisées pour la détection de rupture sur des données. Cette méthode est déduite d'après le rapport de vraisemblance dynamique entre deux populations. Il s'agit d'une technique d'analyse séquentielle. En considérant une distribution de probabilité  $P_\theta$ , paramétrée par  $\theta$ , (l'espérance par exemple), cette méthode consiste à détecter un changement sur ce paramètre.

La figure 4(a) présente un échantillon gaussien de variance constante et deux valeurs distinctes de la moyenne. Il existe un instant de rupture  $r$  tel que l'espérance de cet échan-

2. Cf. Remerciements

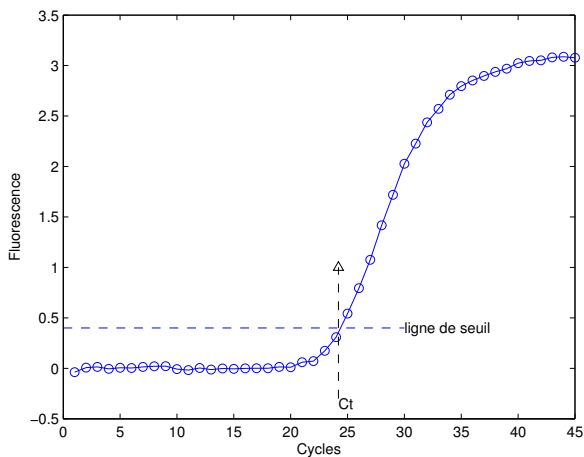


(a) Données.

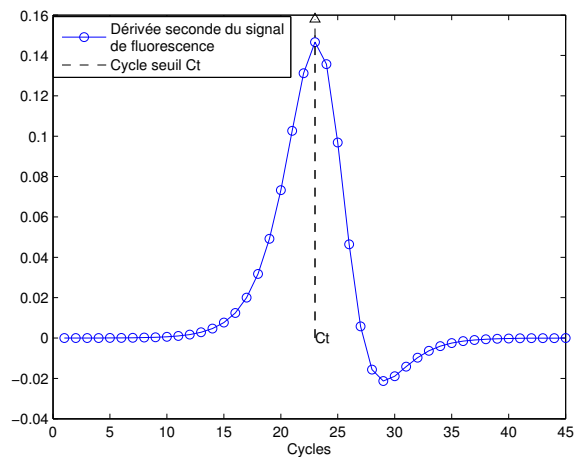


(b) Deux types de comportements.

FIGURE 1 – Représentation graphique des données de fluorescence (a) et ses deux types de comportements (b).



(a) Méthode du seuil. Le  $C_t$  est défini comme étant égal à l'intersection d'une ligne de seuil fixée avec la courbe de fluorescence.



(b) Maximum de la dérivée seconde. Le  $C_t$  est défini comme étant égal au maximum de la dérivée seconde de la courbe de fluorescence.

FIGURE 3 – Deux différentes méthodes de détermination d'un seuil.

tillon est égale à un réel  $\theta_0$  avant  $r$  et un réel  $\theta_1$  après  $r$ . Le CUSUM permet de détecter ce changement (fig. 4(b)) [1].

### 3.2 Cadre de travail

Le cœur de ce travail est basé sur un modèle statistique de test paramétrique. Il revient alors à bien poser le test d'hypothèse (éq. 1) et définir sa région critique (éq. 2). Dans la suite nous utilisons les notations suivantes :

#### Notations

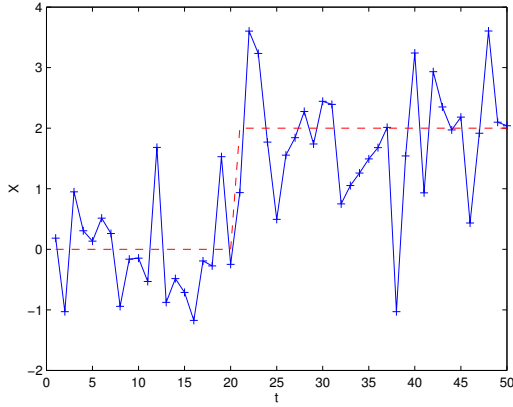
- $n$  : nombre de fluorescences,
  - $C_{max}$  : nombre de cycle maximal,
  - $i, j$  et  $k$  indices associés aux signaux,  $1 \leq i, j, k \leq n$ ,
  - $r, s$  et  $t$  indices associés aux cycles,  $1 \leq r, s, t \leq C_{max}$ .
- Notons  $\mathcal{X} = \{\mathbf{x}^i\}_{i=1, \dots, n}$  comme étant l'ensemble des fluorescences. Supposons que la fluorescence  $\mathbf{x}^i = \{x_t^i\}_{t=1, \dots, C_{max}}$ ,  $\forall i \in [1, n]$ , admet une densité

de probabilité  $\mathcal{P}_\theta$  paramétrée par  $\theta \in \Omega$  où  $\Omega$  est appelé espace des paramètres réels.

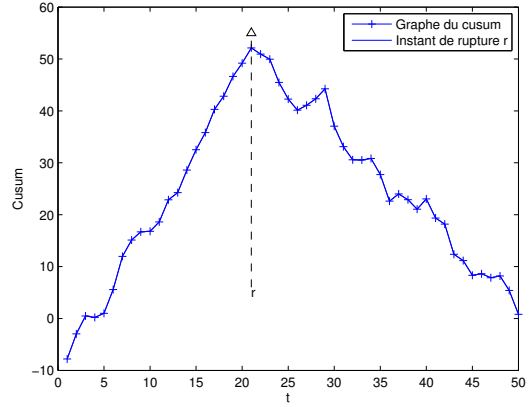
Deux situations se présentent :

- soit toutes les observations de l'échantillon ont la même densité de probabilité, caractérisée par  $\theta_0$ ,
- soit il existe  $r$  inconnu,  $1 \leq r \leq C_{max}$ , tels que  $\theta = \theta_0$  avant  $r$  et  $\theta = \theta_1 (\neq \theta_0)$  après  $r$ .

Nous allons représenter les deux situations (saut ou sans saut) sous la forme d'un test d'hypothèse.



(a) Échantillon gaussien



(b) Graphe du CUSUM associé l'échantillon gaussien (a)

FIGURE 4 – Un échantillon gaussien avec son graphe de CUSUM correspondant [1]. L'algorithme du CUSUM estime l'instant de rupture à  $t = 21$ .

**Test d'hypothèse de niveau  $\alpha$**  (voir ci après pour la définition de  $\alpha$ ) :  $\forall s$  tel que  $1 \leq s \leq C_{max}$

$$\left\{ \begin{array}{l} \mathcal{H}_0 : \quad \mathbf{x}^i = \{x_t^i\}_{t=1, \dots, s} \rightsquigarrow \mathcal{P}_{\theta_0} \\ \mathcal{H}_1 : \quad \exists r \text{ tel que} \\ \quad \quad \mathbf{x}^i = \{x_t^i\}_{t=1, \dots, r-1} \rightsquigarrow \mathcal{P}_{\theta_0} \\ \quad \quad \mathbf{x}^i = \{x_t^i\}_{t=r, \dots, s} \rightsquigarrow \mathcal{P}_{\theta_1} \end{array} \right. \quad (1)$$

Le principal problème est de comparer l'hypothèse nulle ( $\mathcal{H}_0$ ) à l'hypothèse alternative ( $\mathcal{H}_1$ ).

**Remarque :** Nous désirons détecter la rupture le plus tôt possible, le test se fait alors de manière récursive suivant  $s$ .

**Définition :**  $\alpha$  est appelé erreur de premier espèce, c'est à dire la probabilité de refuser l'hypothèse nulle  $\mathcal{H}_0$  à tort, c'est à dire :

$$\alpha = P(\mathcal{H}_1 | \mathcal{H}_0).$$

Nous allons maintenant définir la région critique ou région d'acceptation de l'hypothèse alternative [1, 8].

**Règle de décision ou région critique [11] :** Pour un signal  $i$ , ( $1 \leq i \leq n$ ),

$$W^i = \left\{ (x_1^i, \dots, x_s^i) / \log \frac{L(x_1^i, \dots, x_s^i | \mathcal{H}_1)}{L(x_1^i, \dots, x_s^i | \mathcal{H}_0)} \geq h_\alpha \right\}, \quad (2)$$

avec  $h_\alpha$  un seuil décision convenablement choisi [1], et  $L$  représente la vraisemblance [1, 4].

$W^i$  est une région d'acceptation de l'hypothèse  $\mathcal{H}_1$  pour le signal  $i$  ( $1 \leq i \leq n$ ).

La vraisemblance est mise en œuvre dans le but de quantifier l'adéquation entre une distribution de probabilité et un échantillon : plus grande est la vraisemblance de l'échantillon, meilleure est l'adéquation. Elle est utilisée par les tests d'hypothèses dont le but est de discriminer entre deux groupes de distributions qui affirment chacun contenir la distribution qui a engendré l'échantillon.

Le log du rapport de vraisemblance est égal à,

$$\log \frac{L(x_1^i, \dots, x_s^i | \mathcal{H}_1)}{L(x_1^i, \dots, x_s^i | \mathcal{H}_0)} = \log \frac{\prod_{t=1}^{r-1} P_{\theta_0}(x_t^i) \prod_{t=r}^s P_{\theta_1}(x_t^i)}{\prod_{t=1}^s P_{\theta_0}(x_t^i)}.$$

Après simplification, nous obtenons

$$\log \frac{L(x_1^i, \dots, x_s^i | \mathcal{H}_1)}{L(x_1^i, \dots, x_s^i | \mathcal{H}_0)} = \sum_{t=r}^s \log \frac{P_{\theta_1}(x_t^i)}{P_{\theta_0}(x_t^i)}. \quad (3)$$

Notons

$$S_s^i(r) = \sum_{t=r}^s \log \frac{P_{\theta_1}(x_t^i)}{P_{\theta_0}(x_t^i)}. \quad (4)$$

Pour résoudre un tel problème, nous utilisons le principe du maximum du log du rapport de vraisemblance généralisé car l'instant de rupture  $r$  est inconnu [1].

Posons alors,

$$S_s^i := \max_{1 \leq r \leq s} S_s^i(r), \quad (5)$$

et la règle de décision  $d^i$ , telle que  $d^i = 0$  ou 1 est accordée respectivement à  $\mathcal{H}_0$  ou  $\mathcal{H}_1$ , est donnée par,

$$d^i = \begin{cases} 0 & \text{si } S_s^i < h_\alpha \\ 1 & \text{si } S_s^i \geq h_\alpha. \end{cases}$$

Nous désirons détecter l'instant de rupture le plus tôt possible. Ainsi avant d'estimer cet instant  $r$ , nous allons d'abord déterminer le temps d'arrêt  $t_a$ , qui est l'instant minimum où le choix  $d^i = 1$  est pris. Cet instant  $t_a$  nous permet de minimiser le retard à la détection. Il vient que,

$$t_a = \min\{s : S_s^i \geq h_\alpha\}. \quad (6)$$

Une fois que l'instant d'arrêt  $t_a$  est déterminé, nous sommes alors sous l'hypothèse ( $\mathcal{H}_1$ ), il reste à déterminer l'instant de rupture.

**Estimation de  $r$  :** Il existe un instant  $r$  où il y a un saut, le problème revient alors à estimer cet instant. En utilisant le principe du maximum de vraisemblance, nous obtenons [1],

$$r^* = \arg \max_{1 \leq r \leq t_a} \log \left[ \prod_{t=1}^{r-1} P_{\theta_0}(x_t^i) \prod_{t=r}^{t_a} P_{\theta_1}(x_t^i) \right]. \quad (7)$$

où  $\log \left[ \prod_{t=1}^{r-1} P_{\theta_0}(x_t^i) \prod_{t=r}^{t_a} P_{\theta_1}(x_t^i) \right]$  est le log vraisemblance et  $r^*$  est l'estimateur de l'instant de rupture  $r$ .

En multipliant et en divisant par  $\prod_{t=r}^{t_a} P_{\theta_0}(x_t^i)$ , nous obtenons que,

$$r^* = \arg \max_{1 \leq r \leq t_a} \left[ \log \frac{\prod_{t=r}^{t_a} P_{\theta_1}(x_t^i)}{\prod_{t=r}^{t_a} P_{\theta_0}(x_t^i)} + \log \prod_{t=1}^{t_a} P_{\theta_0}(x_t^i) \right]. \quad (8)$$

Le dernier terme étant constant par rapport à  $r$ , nous avons,

$$r^* = \arg \max_{1 \leq r \leq t_a} \left[ \sum_{t=r}^{t_a} \log \frac{P_{\theta_1}(x_t^i)}{P_{\theta_0}(x_t^i)} \right]. \quad (9)$$

Finalement,

$$r^* = \arg \max_{1 \leq r \leq t_a} S_{t_a}^i(r), \quad (10)$$

$S_{t_a}^i(r)$  ainsi défini représente le CUSUM du signal  $i$ . Elle est la somme cumulée du log du rapport de vraisemblance pour les observations de  $r$  à  $t_a$ .

La figure 5 illustre en détails la procédure de l'estimation de  $r$  pour chaque signal  $i$ ,  $1 \leq i \leq n$  et  $\forall 1 \leq c \leq C_{max}$ .

**Remarque :** Les paramètres avant la rupture  $\theta_0$  et après la rupture  $\theta_1$  ne sont pas nécessairement connus. Nous utilisons alors le log du rapport de vraisemblance généralisé [1], qui est :

$$\log \frac{\hat{L}(x_1^i, \dots, x_s^i | \mathcal{H}_1)}{\hat{L}(x_1^i, \dots, x_s^i | \mathcal{H}_0)} = \sup_{\theta_1, \theta_0} \sum_{t=r}^s \log \frac{P_{\theta_1}(x_t^i)}{P_{\theta_0}(x_t^i)} =: \hat{S}_s^i(r). \quad (11)$$

Dans d'autres cas, il n'y a que le paramètre après la rupture qui est inconnu. De ce fait le log du rapport de vraisemblance généralisé [1] est :

$$\log \frac{\hat{L}(x_1^i, \dots, x_s^i | \mathcal{H}_1)}{\hat{L}(x_1^i, \dots, x_s^i | \mathcal{H}_0)} = \sup_{\theta_1} \sum_{t=r}^s \log \frac{P_{\theta_1}(x_t^i)}{P_{\theta_0}(x_t^i)} =: \hat{S}_s^i(r). \quad (12)$$

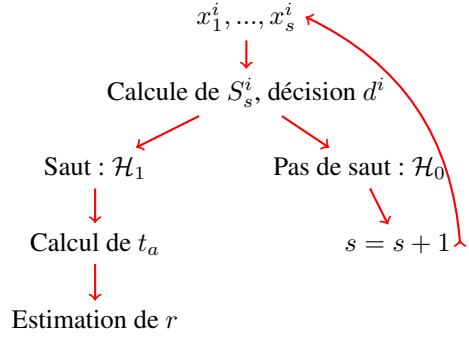


FIGURE 5 – Schéma illustrant l'algorithme de détection de rupture.

## 4 Résultats obtenus et étude comparative

Dans cette partie, nous appliquons l'algorithme du CUSUM sur les données fournies par le LHVP provenant de l'appareil d'enregistrement. Cet appareil mesure les points caractéristiques,  $C_t$ , par la méthode de la dérivée seconde. Ces derniers seront comparés à ceux que nous obtenons par la méthode du CUSUM.

Les signaux de fluorescence sont représentés sur la figure 1(a) et les points caractéristiques correspondants sur le tableau 1.

En appliquant la méthode du CUSUM, nous obtenons des instants de ruptures notés  $r$ . Les points caractéristiques réels sont déduits en faisant une interpolation (spline par exemple) ou une pondération sur les  $r$  entiers. Les résultats obtenus sont représentés sur le tableau 1 : la colonne  $\hat{C}_{\partial^2}$  représente les cycles livrés par l'appareil standard ; la colonne  $t_a$  représente le temps d'arrêt ; la colonne  $\hat{C}_{CUS}$  représente les points caractéristiques obtenus via une pondération autour des  $r$  et  $\hat{C}_{CUS}$  ceux obtenus via une interpolation spline.

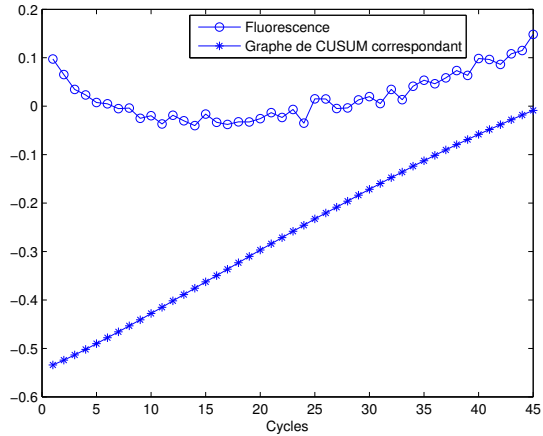
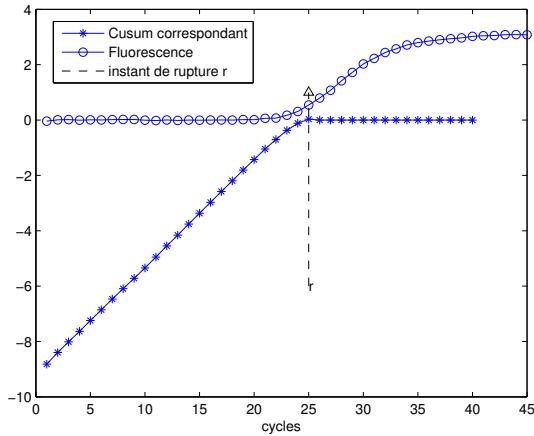
Nous constatons que les points caractéristiques obtenus via la méthode du CUSUM sont cohérents avec ceux livrés par l'appareil de mesure standard (tab. 1) car nous restons linéaire par rapport aux résultats fournis par ce dernier (fig. 9).

Deux exemples de comportement sur le graphe du CUSUM associé à sa fluorescence sont représentés sur les figures 6(a) et 6(b).

En représentant sur la même figure les points caractéristiques obtenus par la méthode de la dérivée seconde par rapport à ceux obtenus par la méthode du CUSUM, nous obtenons la figure 7.

Le graphe des cycles calculés via la pondération par rapport aux cycles obtenus par la méthode de la dérivée seconde, (fig. 9(b)), présente quelques fluctuations par rapport à ceux obtenus par l'interpolation spline (fig. 9(a)).

Nous pouvons alors déduire que l'interpolation est meilleure que la pondération. Dans les deux cas, nous res-



(a) Signal de fluorescence avec un effet de saut et son graphe de CUSUM correspondant

(b) Signal de fluorescence sans un effet de saut et son graphe de CUSUM correspondant

FIGURE 6 – Deux types de comportement des données et leur CUSUM correspondant. Un cycle seuil est déterminé à partir de l’instant de rupture pour des signaux avec saut et cet instant est l’instant auquel la courbe du CUSUM atteint son maximum.

tons linéaire par rapport aux points caractéristiques obtenus par l’appareil standard (fig. 9) donc le  $C_0$  peut en être déduit après étalonnage. Nous observons un biais constant sur les valeurs obtenues par rapport à celles de l’appareil de mesure standard. Nous allons faire de l’apprentissage par validation croisée afin d’estimer ce biais et d’en appliquer une correction.

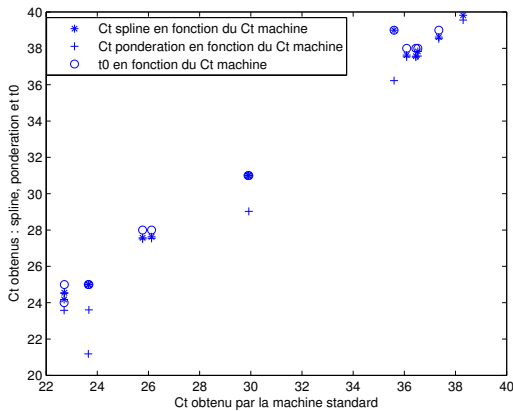


FIGURE 7 – Points caractéristiques obtenus par la méthode de la dérivée seconde par rapport aux notre et aux instants de rupture  $t_0$ .

### Validation croisée :

Le biais rencontré est constant. Nous allons alors calculer cette correction par apprentissage en utilisant la validation croisée par *leave one out*. Elle consiste à faire la moyenne de la différence entre le point caractéristique obtenu par la méthode de la dérivée seconde et celui que nous obtenons

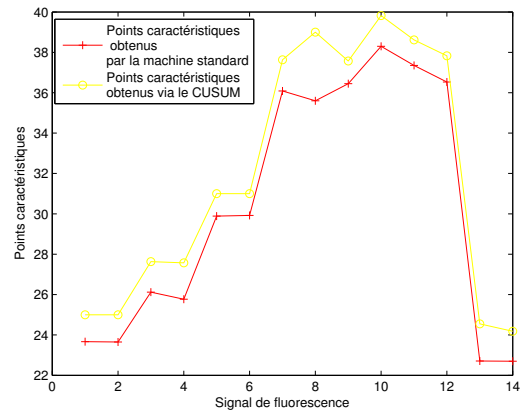


FIGURE 8 – Comparaison des points caractéristiques : ceux obtenus par la méthode de la dérivée seconde contre ceux que nous obtenons par la méthode du CUSUM.

sans la valeur courante (éq. 13).

Les résultats obtenus sont donnés sur le tableau 2.

$$correction(i) = \frac{1}{N-1} \sum_{j=1; j \neq i}^{N-1} difference(j), \forall 1 \leq i \leq N, \quad (13)$$

où  $N$  est le nombre de signaux avec rupture et *difference* est la différence entre nos cycles obtenus ( $\hat{C}_{CUS}$ ) et ceux trouvés par la méthode de la dérivée seconde ( $\hat{C}_{\partial^2}$ ).

$$difference(i) = \hat{C}_{CUS}(i) - \hat{C}_{\partial^2}(i), \forall 1 \leq i \leq N.$$

Ainsi, nous déterminons les points caractéristiques corrigés ( $\hat{C}_t$  corrigé) en retirant les valeurs de la correction à nos

Signaux	$\hat{C}_{\theta^2}$	$t_a$	$r$	$\hat{C}_{CUP}$	$\hat{C}_{CUS}$
1	23.22	25	25	24.79	25
2	23.57	25	25	24.63	24.79
3	26.96	28	28	26.63	28
4	26.86	28	28	26.75	28
5	28.72	30	30	30.99	30
6	28.67	30	30	29.63	29.85
7	34.64	36	36	35.67	36
8	34.80	36	36	35.83	36
9	36.06	38	38	37.58	37.75
10	35.89	37	37	37.09	36.91
11	35.53	37	37	36.62	36.80
12	35.68	37	36	35.58	36.17
13	36.10	37	37	33.48	37
14	36.23	38	38	37.62	37.79
15	36.01	37	37	38.36	37
16	35.85	37	37	36.53	36.74
17	22.05	24	24	23.72	24
18	22.10	24	24	23.54	23.78
19	0	45 (=n)	0	0	0
20	0	45 (=n)	0	0	0

TABLE 1 – Tableau des différents points caractéristiques obtenus en appliquant le CUSUM (2 dernières colonnes) et ceux de l'appareil d'enregistrement (2<sup>ieme</sup> colonne).  $r$  est l'instant de rupture déterminé par la méthode du CUSUM ;  $t_a$  représente le temps d'arrêt. Les deux derniers signaux n'ont pas de points caractéristiques. Ce sont des signaux sans rupture.

cycles obtenus ( $\hat{C}_{CUS}$ ) :

$$\hat{C}_t \text{ corrigé}(i) = \hat{C}_{CUS}(i) - \text{correction}(i), \forall 1 \leq i \leq N.$$

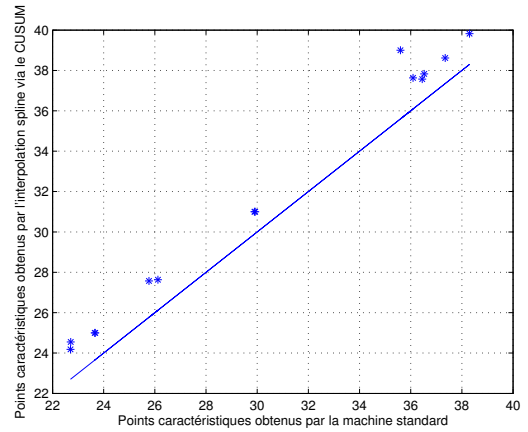
Un point caractéristique est obtenu avec un retard de 1.39 cycles en moyen. En se basant sur nos objectifs correspondant aux tolérances des appareils de mesure industriel (tab. 4). Nous pouvons donc conclure que nos points caractéristiques respectent les performances à atteindre en terme de retard à la détection et de précision du  $\hat{C}_t$ .

Avant la correction, les points caractéristiques que nous obtenons sont au dessus de ceux de l'appareil standard de 1.13 cycle en moyen et le retard à la détection est de 1.39 cycle en moyen (dernière ligne).

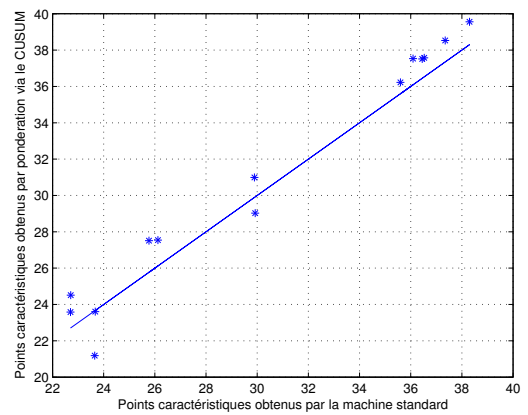
Après correction (tab. 2 et tab. 3) nous sommes de 0.26 cycle en moyen au dessus des points caractéristiques obtenus par la méthode de la dérivée seconde et de 1.65 cycles en moyen de retard à la détection.

	Minimum	Maximum	Moyenne
$ \hat{C}_{\theta^2} - \hat{C}_t \text{ corrigé} $ , précision du cycle seuil :	0.01 cycle	0.73 cycle	0.26 cycle

TABLE 3 – Résumé, tableau de précision des points caractéristiques obtenus par la méthode du CUSUM.



(a) Points caractéristiques obtenus par interpolation spline par rapport à ceux obtenus par la méthode de la dérivée seconde.



(b) Points caractéristiques obtenus par pondération par rapport à ceux obtenus par la méthode de la dérivée seconde.

FIGURE 9 – Comparaison des points caractéristiques obtenus par pondération et par interpolation.

	Minimum	Objectif	Obtenu
Entre 15 et 35 cycles, précision du cycle seuil	+ - 2 ct	+ - 0.5 ct	+ - 0.26 ct
Information de quantifications après le cycle seuil	10 cycles	4 cycles	1.65 ct

TABLE 4 – Performances à atteindre.



Retard détection	Différence	Correction	Cycle corrigé	Erreur
1.13	1.78	1.22	23.77	0.55
1.34	1.22	1.25	23.53	0.04
1.13	1.03	1.27	26.72	0.24
1.13	1.14	1.26	26.73	0.13
1.13	1.28	1.25	28.74	0.02
1.28	1.18	1.26	28.58	0.09
1.13	1.36	1.25	34.74	0.10
1.13	1.20	1.26	34.73	0.07
1.38	1.69	1.23	36.51	0.45
1.22	1.02	1.27	35.63	0.26
1.33	1.27	1.25	35.54	0.01
1.96	0.49	1.30	34.8	0.18
1.13	0.90	1.27	35.72	0.38
1.34	1.56	1.23	36.55	0.32
1.13	0.99	1.27	35.72	0.29
1.39	0.89	1.27	35.46	0.39
1.13	1.95	1.21	22.78	0.73
1.35	1.68	1.23	22.54	0.44
<b>1.39</b>	<b>1.13</b>	...	...	<b>0.26</b>

TABLE 2 – Tableau des différences entre les points caractéristiques obtenus par la méthode de la dérivée seconde ( $\hat{C}_{\partial^2}$ ) et les autres, ainsi que les points caractéristiques obtenus après correction ( $\hat{C}_t$  corrigé). Les valeurs en gras représentent les moyennes de la colonne; retard détection =  $t_a - \hat{C}_{CUS}$ ; différence =  $|\hat{C}_{\partial^2} - \hat{C}_{CUS}|$ ; cycle corrigé =  $\hat{C}_t$  corrigé; erreur =  $|\hat{C}_{\partial^2} - \hat{C}_t \text{ corrigé}|$ .

## 5 Conclusion et perspectives

La détermination d'un point de montée du signal de fluorescence peut être faite en temps continu et sans connaissance au préalable du comportement total du signal. Pour cela, nous avons utilisé la méthode du CUSUM qui est une technique de détection de rupture. Nous obtenons des points caractéristiques avec un biais constant par rapport à ceux fournis par l'appareil de mesure standard tout en minimisant le retard à la détection. Pour corriger ce de biais, nous avons alors utilisé une technique d'apprentissage à savoir la validation croisée. Avec cette méthode nous avons obtenu des résultats satisfaisants avec 1.65 cycles de retard à la détection.

Il reste à étendre l'algorithme en travaillant directement sur les images de fluorescences au lieu de la moyenne spatiale de ces dernières. Ce qui facilitera notamment l'estimation de la variance pour notre modèle statistique en supposant le signal ergodique.

## REMERCIEMENTS

Le projet Genetic Equipement for biothreat environmental Analysis and Surveillance (GENEASE) dont est issu ce travail est financé par l'Agence Nationale de la Recherche<sup>3</sup>. Il est composé

3. <http://www.agence-nationale-recherche.fr/>

de Bertin technologies<sup>4</sup>, du CEA LETI<sup>5</sup>, du CEA SBTN<sup>6</sup> et du LETIS; le LHVP<sup>7</sup> en est un prescripteur.

## Références

- [1] M. BASSEVILLE et I.V. NIKIFOROV : *Detection of Abrupt Changes : Theory and Application (Prentice Hall information and system sciences series)*. Prentice Hall, 1993.
- [2] J. BERTHIER et P. SILBERZAN : *Microfluidics for Biotechnology, Second Edition*. Artech House, 2009.
- [3] S.A. BUSTIN, V. BENES, J.A. GARSON, J. HELLEMANS, J. HUGGETT, M. KUBISTA, R. MUELLER, T. NOLAN, M.W. PFAFFL et G.L. SHIPLEY : The MIQE guidelines : minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry*, 55(4):611–622, 2009.
- [4] R.O. DUDA, P.E. HART et D.G. STORK : *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.
- [5] M. KUBISTA, J.M. ANDRADE, M. BENGTSSON, A. FOROOTAN, J. JONÁK, K. LIND, B. Sjögreen L. Strömbom A. Stahlberg SINDELKA, R. Sjöback et N. ZORIC : The real-time Polymerase Chain Reaction. *Molecular aspects of medicine*, 27(2-3):95–125, 2006.
- [6] K.J. LIVAK et T.D. SCHMITTGEN : Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method. *Methods*, 25(4):402–408, 2001.
- [7] K.B. MULLIS et F.A. FALOONA : Specific synthesis of dna in vitro via a polymerase-catalyzed chain reaction. *Methods in enzymology*, 155:335–350, 1987.
- [8] E.S. PAGE : Continuous inspection schemes. *Biometrika*, 41(1-2):100–115, 1954.
- [9] R.G. RUTLEDGE : Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. *Nucleic acids research*, 32(22):e178, 2004.
- [10] A. SCHAEFER, M. JUNG, H.J. MOLLENKOPF, I. WAGNER, C. STEPHAN, F. JENTZMIK, K. MILLER, M. LEIN, G. KRISTIANSEN et K. JUNG : Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. *International journal of cancer*, 126(5):1166–1176, 2010.
- [11] P. TASSI : *Méthodes statistiques*. Economica Paris, 1985.
- [12] C. TSE et J. CAPEAU : Quantification des acides nucléiques par PCR quantitative en temps réel. *Ann Biol Clin*, 61(3):279–293, 2003.

4. Bertin technologies <http://www.bertin.fr/>

5. Commissariat à l'Energie Atomique - Laboratoire d'Electronique et de Technologies de l'Information, <http://www.leti.fr>

6. Commissariat à l'Energie Atomique - Service de Biochimie et Toxicologie Nucléaire, <http://www-dsv.cea.fr/sbntn/>

7. Laboratoire d'Hygiène de la Ville de Paris