



# **Change detection for uncertain autoregressive dynamic models through nonparametric estimation**

Nadine Hilgert, Ghislain Verdier, Jean-Pierre Vila

## **► To cite this version:**

Nadine Hilgert, Ghislain Verdier, Jean-Pierre Vila. Change detection for uncertain autoregressive dynamic models through nonparametric estimation. *Statistical Methodology*, 2016, 33, pp.96-113. <10.1016/j.stamet.2016.08.003>. <hal-01581265>

**HAL Id: hal-01581265**

**<https://hal.science/hal-01581265v1>**

Submitted on 28 May 2020

**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.



HAL Authorization

# Change detection for uncertain autoregressive dynamic models through nonparametric estimation

Nadine Hilgert<sup>a</sup>, Ghislain Verdier<sup>b,\*</sup>, Jean-Pierre Vila<sup>a</sup>

<sup>a</sup>UMR 729 MISTEA, INRA-SupAgro, 2 place Pierre Viala, 34060 Montpellier, France

<sup>b</sup>LMAP, UMR CNRS 5142, Université de Pau et des Pays de l'Adour, Avenue de l'Université, 64000 Pau, France

---

## Abstract

A new statistical approach for on-line change detection in uncertain dynamic system is proposed. In change detection problem, the distribution of a sequence of observations can change at some unknown instant. The goal is to detect this change, for example a parameter change, as quickly as possible with a minimal risk of false detection. In this paper, the observations come from an uncertain system modeled by an autoregressive model containing an unknown functional component. The popular Page's CUSUM rule is not applicable anymore since it requires the full knowledge of the model. A new detection CUSUM-like scheme is proposed, which is based on the nonparametric estimation of the unknown component from a learning sample. Moreover, the estimation procedure can be updated on line which ensures a better detection, especially at the beginning of the monitoring procedure. Simulation trials were performed on a model describing a water treatment process and show the interest of this new procedure with respect to the classic CUSUM rule.

## Keywords:

CUSUM rule, Model change detection, Nonparametric estimation, Learning sample updating

---

---

\*Corresponding author

Email addresses: [nadine.hilgert@supagro.inra.fr](mailto:nadine.hilgert@supagro.inra.fr) (Nadine Hilgert),  
[ghislain.verdier@univ-pau.fr](mailto:ghislain.verdier@univ-pau.fr) (Ghislain Verdier),  
[jean-pierre.vila@supagro.inra.fr](mailto:jean-pierre.vila@supagro.inra.fr) (Jean-Pierre Vila)

## 1. INTRODUCTION

For many years, statistical “on-line” change detection methods have given rise to an abundant literature. This is due to the large number of applications, particularly in process supervision: quality control in agro-food industries, seismology or biotechnological processes as water depollution, etc. Among the various statistical approaches, one of the most efficient is to construct sequential rules deriving from test theory (see [3] or [17] for an overview). However despite a significant demand for methods capable of dealing with uncertain dynamic systems, a few work has been carried in this direction. Fouladirad and Nikiforov [8], Harrou et al. [13] and Gombay [12], for example, treat systems with nuisance parameters in linear or AR(1) models, but not unknown functional component in a general nonlinear model as it is the case in this paper.

One of the most suited rule to parameter change detection when the system can be described by a model is the CUSUM rule. It was introduced by Page [28], initially in the case of independent and identically distributed observations before and after the change. The study of the CUSUM rule convergence properties has been the subject of an extensive literature for fifty years. Among the most striking results, the works of Lorden [19] or Moustakides [24] can be mentioned in an independent and identically distributed (*iid*) context and Bansal and Papantoni-Kazakos [2], Lai [16], Fuh [9] or Mei [22] treat the case of dependent data.

The major drawback of this rule is that it requires the knowledge of the exact probability distributions of the observations before and after the change. Unfortunately, in practice, this assumption can be restrictive since a system can never be perfectly modelled. The simulation trials performed in this paper show that the quality of the CUSUM test is deteriorating rapidly with small modelling errors.

The aim of this paper is to propose a robust adaptation of the CUSUM rule when a component of the model describing the system is badly known. For example, in biotechnological processes such as water depollution, this component may represent the growth rate of biomass (see below).

The models considered in the sequel have the following form:

$$\forall n \geq 0, X_{n+1} = f(X_n) + F(X_n, \theta) + \epsilon_{n+1}, \quad (1)$$

with  $X_n \in \mathbb{R}^d$  the observation vector and  $X_0 = x_0$  the initial condition,  $\epsilon_n$  a Gaussian white noise with covariance matrix  $R$  (supposed invertible),

$F : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$  a known function and  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  the badly known component.  $\theta \in \Theta \subset \mathbb{R}^s$  is the parameter characterizing the change. It is a component of  $F$ , the known part of model (1). At the change time  $t_0$  (unknown but not random), the parameter  $\theta$  moves from a nominal value  $\theta_0$  to  $\theta_1$ . The two values  $\theta_0$  and  $\theta_1$  are supposed to be known.

The robust CUSUM approximation proposed in this paper consists in replacing, in the CUSUM procedure, the traditional increments (not computable since  $f$  is unknown) by convergent approximations of them. In Verdier et al. [38], the authors study such a type of rule when the system follows restrictive assumptions (satisfied, for example, by uniformly recurrent Markov process). But model (1) does not satisfy such assumptions and the results of [38] can not be applied here.

In this paper, the approximations of the increments are based on the use of a nonparametric kernel estimation of  $f$ . Kernel-based methods were first used for density function estimation ([29], [33]), then for regression function estimation [27]. Most of the regression estimators met in the literature deal with non-controlled models and treat the case of stationary processes [6]. Duflo [7] and Senoussi [34] were the first to give convergence results for regression estimation in a controlled framework and Portier and Oulidi [31] obtained the convergence of a kernel estimator over dilating sets (see below). In a change detection context, nonparametric kernel approaches were notably used for *iid* observations. For example, Baillo and Cuevas [1] propose a control chart based on a kernel estimate of the density level set of the observations. More recently, Su et al. [36] construct a CUSUM like algorithm from kernel estimates of the densities (in-control and out-of-control).

Moreover, in this paper, an updating of the nonparametric estimator during the monitoring procedure is proposed. Indeed, a part of the new observations collected during the monitoring phase are added to the learning sample, which increases the quality of the estimation and so the quality of the detection scheme. The resulting test statistic is then not stationary and the use of a fixed threshold (like it is traditionally the case for the CUSUM) is no more adapted. The computational procedure proposed in Verdier et al. [37] in order to obtain a constant probability of false alarm is then used.

The paper is organized as follows. In section 2, the CUSUM detection rule is presented and applied to model (1) when function  $f$  is supposed to be known. In section 3, the function  $f$  is supposed to be unknown. A nonparametric kernel method is used for the construction of a new detection rule: the nonparametric CUSUM. The convergence of the increments of the

rule is investigated and the updating procedure of the estimator is presented. In section 4, approximation methods for the ARL function of the detection rules are given. Finally, in section 5, simulation trials are performed in order to compare the nonparametric approach with the classical CUSUM rule used with a small modeling error.

In the following, the matrix norm is the Frobenius norm, which reduces to the Euclidean norm when dealing with vectors.

## 2. THE CUSUM ALGORITHM

### 2.1. General case

Let  $X_1, \dots, X_n$  be a sequence of random variables describing a dynamic system submitted to an abrupt change. This change is characterized by a move from  $\theta_0$  to  $\theta_1$  of the value of a parameter  $\theta$  of the distribution function. This change occurs at an unknown time  $t_0$ . The conditional density function of  $X_n$  given the past values  $X_{1:n-1}$  for  $n \geq t_0$  (out-of-control mode  $H_1$ ) is denoted  $p_{\theta_1}(\cdot|X_{1:n-1})$ , with the usual notation  $X_{1:n}$  for  $X_1, \dots, X_n$ . For  $n < t_0$  (nominal mode  $H_0$ ), the conditional density is  $p_{\theta_0}(\cdot|X_{1:n-1})$ .

We use  $P^{(t_0)}$  to denote the probability measure of such a distribution. The notation  $P_{\theta_0}$  is used in the case  $t_0 = \infty$  (no change). The stopping time of the CUSUM procedure [28] is defined by

$$t_C = \inf \left\{ n : \max_{1 \leq k \leq n} \sum_{i=k}^n Z_i \geq h \right\}, \quad (2)$$

with,

$$Z_i = \log \frac{p_{\theta_1}(X_i|X_{1:i-1})}{p_{\theta_0}(X_i|X_{1:i-1})}, \quad (3)$$

where  $h$  is a given threshold and  $g_n := \max_{1 \leq k \leq n} \sum_{i=k}^n Z_i$  is the CUSUM test statistic.

**Remark 1.** There exists (see [28]) an equivalent test statistic  $\tilde{g}_n$ , written in a recursive manner:

$$\forall n \geq 1, \quad \tilde{g}_n = \max(\tilde{g}_{n-1} + Z_n, 0) \quad \text{and} \quad \tilde{g}_0 = 0, \quad (4)$$

which is useful in practice. Note however that these two statistics are equivalent only when the threshold  $h$  is positive, which is frequently the case in

practice to ensure a low rate of false alarms, and in particular for the model (1) considered in this paper, due to the Gaussian nature of the conditional densities (see below).

Note also that Moustakides et al. [25] propose a slightly different writing of the CUSUM rule with a recursive version available whatever the value of the threshold.

Several authors (for example Bansal and Papantoni-Kazakos [2], Lai [16] or Fuh [9]) showed, under various hypothesis on the distribution  $P^{(t_0)}$ , that the rule  $t_C$  minimizes asymptotically (as  $\gamma \rightarrow \infty$ ) the worst mean delay for detection:

$$\bar{E}_{\theta_1}(T) = \sup_{t_0 \geq 1} \text{ess sup } E^{(t_0)}[(T - t_0 + 1)^+ | X_{1:t_0-1}], \quad (5)$$

over all detection rules  $T$  with a mean time between false alarms such that, for a given  $\gamma > 0$ ,

$$E_{\theta_0}(T) \geq \gamma \quad (\text{ARL constraint}). \quad (6)$$

In the criterion  $\bar{E}_{\theta_1}(T)$  (firstly introduced by Lorden [19]), the essential supremum is taken with respect to  $X_{1:t_0-1}$ .

## 2.2. CUSUM rule applied to the nonlinear autoregressive model (1)

In this part, the function  $f$  of model (1) is assumed to be known. Then, the model can be simplified as follows:

$$\forall n, \quad X_{n+1} = u(X_n, \theta) + \epsilon_{n+1}, \quad (\text{with } u = f + F) \quad (7)$$

where  $u$  is a known function. Given the Gaussian nature of the noise  $\epsilon_n$ , the conditional law of observations is also Gaussian. The Markovian property implies:

$$p_{\theta}(x_i | X_{1:i-1}) = p_{\theta}(x_i | X_{i-1}) = \frac{\exp \left[ -\frac{(x_i - u(X_{i-1}, \theta))^t R^{-1} (x_i - u(X_{i-1}, \theta))}{2} \right]}{(2\pi)^{d/2} |R|^{1/2}},$$

where  $|R|^{1/2}$  is the square root of the determinant of matrix  $R$ .

Then the increments of the CUSUM rule take the following form:

$$\begin{aligned} Z_i &= \log \left( \frac{p_{\theta_1}(X_i | X_{i-1})}{p_{\theta_0}(X_i | X_{i-1})} \right) \\ &= \frac{1}{2} [(\eta_i^0)^t R^{-1} \eta_i^0 - (\eta_i^1)^t R^{-1} \eta_i^1], \end{aligned} \quad (8)$$

with  $\eta_i^r := X_i - u(X_{i-1}, \theta_r) = X_i - (f(X_{i-1}) + F(X_{i-1}, \theta_r))$ , for  $r = 0, 1$ .

The following proposition states the optimality of the CUSUM rule for model (1), when  $d$ , the dimension of  $X$ , is equal to 1.

**Proposition 1.** *Consider the following model:*

$$\forall n \geq 0, X_{n+1} = u(X_n, \theta) + \epsilon_{n+1}, \quad (9)$$

with observation  $X_n \in \mathbb{R}$  and  $\epsilon_n$  a Gaussian white noise ( $\epsilon_n \sim N(0, \sigma^2)$ ).

1. Suppose that, for all  $\theta \in \Theta$ ,  $u(\cdot, \theta)$  is bounded. Then the Markov process  $(X_n)_{n \geq 1}$  is uniformly ergodic and therefore admits a stationary distribution  $\Pi$ .
2. If, moreover, the initial law of the Markov process  $(X_n)_{n \geq 1}$  is the stationary distribution ( $X_0 \sim \Pi$ ), then the CUSUM rule is asymptotically optimal in the sense of Lorden (it minimizes the criterion (5) subject to the constraint (6)).

PROOF. 1. The first part of the proposition is the direct application of Theorem 2, page 467 of Herkenrath [14] which ensures the uniform ergodicity of model (9) as soon as  $u(\cdot, \theta)$  is bounded, and obviously the existence of a stationary distribution.

2. The second point is a particular case of Fuh's result [9], who states the asymptotic optimality of CUSUM rule for hidden Markov models. Indeed model (9) can be seen as a hidden Markov model in which the state space of the hidden part of the model reduce to one element. The result of Fuh relies on two conditions (conditions  $C1$  and  $C2$ , page 953 of [9]). Condition  $C1$  concerns the  $w$ -uniform ergodicity of the Markov process (and uniform ergodicity implies  $w$ -uniform ergodicity). Condition  $C2$  is a moment condition for the likelihood function and reduces, for model (9), to:

$$\forall \theta \in \Theta, \sup_{x_0 \in \mathbb{R}} E[|p_{X_0}(x_0)X_1 p_\theta(X_1|X_0 = x_0)|] < \infty,$$

where  $p_{X_0}(\cdot)$  is the density of the stationary distribution.

It is easy to show that  $p_{X_0}(\cdot)$  is bounded by a constant  $0 < C < +\infty$ .

We then have, for all  $\theta \in \Theta$ :

$$\begin{aligned} E[p_{X_0}(x_0)X_1p_\theta(X_1|X_0 = x_0)] &= p_{X_0}(x_0)E[X_1p_\theta(X_1|X_0 = x_0)] \\ &= p_{X_0}(x_0) \int_{-\infty}^{+\infty} \left| x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - u(x_0, \theta))^2}{2\sigma^2}\right) \cdot p_{X_1}(x) \right| dx \\ &\leq C^2 \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - u(x_0, \theta))^2}{2\sigma^2}\right) dx \\ &\leq C^2 E[|Z|], \end{aligned}$$

with  $Z \sim N(u(x_0, \theta), \sigma^2)$ . Moreover,

$$E[|Z|] \leq E[Z^2 + 1] \leq u^2(x_0, \theta) + \sigma^2 + 1 \leq M,$$

since  $u(\cdot, \theta)$  is bounded, which completes the proof.

**Remark 2.** Following Fuh [9], it is supposed, in the previous Proposition, that the initial law of the Markov process  $(X_n)_{n \geq 1}$  is the stationary distribution. But this assumption can be removed by considering another optimality criterion than that of Lorden. Indeed, in a recent submitted paper, Pergamenchtchikov and Tartakovsky [30] show the asymptotic optimality of the Shiryaev-Roberts rule (SR rule). More precisely the SR rule minimizes the average delay:

$$E^{t_0}(\tau - t_0 + 1)^+$$

among all the rules  $\tau$  satisfying the following constraint:

$$\sup_{1 \leq k \leq k^* - m^*} P_{\theta_0}(k \leq \tau < k + m^*) \leq \beta$$

when  $\beta$  tends to 0 and  $k^*$ ,  $m^*$  and  $m^* - k^*$  tend to 0.

According to Pergamenchtchikov and Tartakovsky [30], the CUSUM and SR rules "are first-order asymptotically optimal under the same general conditions". The condition imposed by Pergamenchtchikov and Tartakovsky [30] in order to obtain the optimality can be easily verified for model (9) under the assumption of point 1 in Proposition 1, by following the example 6.2 (p. 24) of [30] and without imposing that the initial law of  $(X_n)_{n \geq 1}$  is the stationary distribution.

In conclusion, the CUSUM rule applied to model (9) is asymptotically optimal for the criterion of the average delay whatever may be the initial law of  $x_0$ .



### 3. NONPARAMETRIC CUSUM RULE

In this section, the function  $f$  of the model (1) is now supposed to be unknown. The increments  $(Z_n)_{n \geq 1}$  of the CUSUM rule (8) can no longer be calculated. The detection approach then relies on an estimation scheme of the  $(Z_n)_{n \geq 1}$ , which induces a CUSUM approximation.

#### 3.1. Definition of the stopping time

The estimation of the  $(Z_n)_{n \geq 1}$  is based on the use of a nonparametric kernel estimator  $\hat{f}_N$  of the unknown function  $f$  defined as follows. Let  $K$  be a kernel function, a function from  $\mathbb{R}^d$  to  $\mathbb{R}$ , continuous, bounded, symmetric and integrating to one. Let  $\delta_i$  be a sequence of real numbers converging to 0, called the bandwidth parameter. From a learning sample of  $N$  observations  $(X_i^s)_{1 \leq i \leq N}$  of model (1), for  $x \in \mathbb{R}^d$ ,  $f(x)$  is estimated by

$$\hat{f}_N(x) = \frac{\sum_{i=1}^{N-1} \delta_i^{-d} K\left(\frac{x-X_i^s}{\delta_i}\right) (X_{i+1}^s - F(X_i^s, \theta))}{\sum_{i=1}^{N-1} \delta_i^{-d} K\left(\frac{x-X_i^s}{\delta_i}\right)}, \quad (10)$$

if the denominator of (10) is not equal to 0, and by 0, otherwise. Here the value of  $\theta$  is known.

It can be noticed that the bandwidth parameter  $\delta_i$  in (10) depends on the index  $i$ , which allows to write separately the numerator and the denominator of  $\hat{f}_N$  in a recursive manner.

The CUSUM approximation, named nonparametric CUSUM rule, is defined in the same way as the CUSUM one by replacing, in (8),  $f(X_i)$  by its estimation  $\hat{f}_N(X_i)$ . The test statistic then writes:

$$\hat{g}_n^N = \max_{1 \leq k \leq n} \sum_{i=k}^n \hat{Z}_i^N,$$

with  $\hat{Z}_i^N = [(\hat{\eta}_i^0)^t R^{-1} \hat{\eta}_i^0 - (\hat{\eta}_i^1)^t R^{-1} \hat{\eta}_i^1] / 2$ , and  $\hat{\eta}_i^r = X_i - \hat{f}_N(X_{i-1}) - F(X_{i-1}, \theta_r)$ ,  $r = 0, 1$ .

The stopping time is defined as:

$$\hat{t}_{NP} = \inf\{n : \hat{g}_n^N \geq h\},$$

with  $h$  a chosen threshold.

**Remark 3.** As for the CUSUM rule (see section 2), there is an equivalent test statistic  $\tilde{g}_n^N$ , written in a recursive manner:

$$\forall n \geq 1, \quad \tilde{g}_n^N = \max\left(\tilde{g}_{n-1}^N + \hat{Z}_n^N, 0\right) \quad \text{and} \quad \tilde{g}_0^N = 0,$$

as soon as the threshold  $h$  is positive. We will use this recursive form in the following.

### 3.2. Remarks on the learning sample $(X_i^s)_{1 \leq i \leq N}$

Let us remark that  $\hat{f}_N(x)$  can be seen as a weighted sum of  $(X_{i+1}^s - F(X_i^s, \theta))_{i=1, \dots, N}$ . The closer  $X_i^s$  is to  $x$ , the greater is the contribution of  $X_{i+1}^s - F(X_i^s, \theta)$  for  $\hat{f}_N$  in the neighborhood of  $x$  thanks to the kernel  $K$  and the bandwidth parameter. The quality of the estimation of  $f(x)$  will therefore depend on the number of observations in the near neighborhood of the point  $x$ . In order to achieve a good level of detection, the estimator  $\hat{f}_N$  needs to be efficient throughout its application domain, especially in areas where  $X_i$  takes its values when the system is under  $H_0$ , but also in areas where  $X_i$  takes its values when the system is in the fault mode  $H_1$ . In other words, the learning sample must contain data from the system in the  $H_0$  mode but also in the  $H_1$  mode, especially if the application domains of the two modes are different (for example if an out-of-control situation leads to a decrease of the variable of interest  $X_i$ , for  $i \geq t_0$ ).

In a controlled framework, observations under fault mode  $H_1$  are not necessarily needed. Indeed, the control variable could be used to explore the domain of function  $f$  including those regions of the state space reached without control (under the  $H_1$  mode), and then to ensure a sufficiently rich sample in the regions of interest.

### 3.3. Convergence of the nonparametric CUSUM increments

The following theorem provides the convergence of the increments  $\hat{Z}_i^N$  of the nonparametric CUSUM rule:

**Theorem 2.** Consider the autoregressive model defined in (1). Suppose that the unknown function  $f$  is a contraction:

$$\exists r_f < 1, \quad \forall x \in \mathbb{R}^d, \quad \forall y \in \mathbb{R}^d, \quad \|f(x) - f(y)\| \leq r_f \|x - y\|,$$

and  $F$  is continuous and bounded:  $\forall x, \quad \forall \theta, \quad \|F(x, \theta)\| \leq M$ . Then,

$$\forall n \geq 1, \quad \lim_{N \rightarrow \infty} |\hat{Z}_n^N - Z_n| = 0 \quad \text{a.s.} \quad (11)$$

PROOF. See Appendix A.

Theorem 2 establishes the convergence of the nonparametric increments  $\hat{Z}_n^N$  to those of the CUSUM, which is asymptotically optimal as seen previously (Proposition 1). Note that this result is nevertheless not sufficient to obtain the asymptotic optimality of the nonparametric CUSUM scheme  $\hat{t}_{NP}$ . Theorem 2 must be regarded as an intuitive argument that the proposed approach has good properties and it then justifies the approach proposed in this paper. The simulations performed in section 5 confirm the good behaviour of the rule  $\hat{t}_N P$ .

The optimality of the nonparametric CUSUM rule is an interesting open problem, which still awaits solution.

#### 3.4. Learning sample updating

The nonparametric CUSUM rule can be used in two different ways: a classic approach with the two separated phases of implementation and monitoring, and another one with an on-line update of the learning sample.

##### 3.4.1. Classic approach

We follow here the standard Statistical Process Control (SPC) approach, see for example [23], in which two phases are separated: Phase I consists in the construction of the learning sample and the model estimation whereas Phase II is concerned with the on-line detection. Therefore, the nonparametric estimator  $\hat{f}_N$  is first computed, before the beginning of the monitoring, from a learning sample of size  $N$  for which the  $H_0$  and  $H_1$  modes have been fully identified.

##### 3.4.2. Updating approach

Updating the kernel estimation is easy thanks to the recursive writing of the numerator and the denominator of the estimator  $\hat{f}_N(x)$ .

For adding data to the learning sample, it is necessary to know the status of the process ( $H_0$  or  $H_1$ ): the process is in-control before the change time  $t_0$  and out-of-control after  $t_0$ . By assumption,  $t_0$  is unknown but can be estimated as soon as an alarm arises (see for example [3]) by:

$$\hat{t}_0 = t - T_t + 1,$$

with

$$T_k = T_{k-1} \cdot \mathbf{1}_{\{\hat{g}_{k-1}^N > 0\}} + 1 \quad \text{and} \quad T_1 = 1,$$

and  $t$  is the stopping time of the change detection rule considered (the CUSUM or its nonparametric approximation). More precisely,  $T_k$  counts the observations since the last zero of the CUSUM statistic.

In other words, with the recursive writing of  $\tilde{g}_k^N$ ,  $\hat{t}_0$  is the first time step after the last return to 0 of the test statistic. Therefore if we consider that, during the monitoring, all the observations before a return to 0 of the test statistics are in-control observations, we can expect that only few observations will be misclassified (observations out-of-control declared in control).

The update of the learning sample is then done as follows:

1. As long as the statistics  $\tilde{g}_n$  is equal to 0, the learning sample is updated with the current data  $X_n$  considering it was under  $H_0$ .
2. As soon as  $\tilde{g}_n$  is positive and less than the threshold  $h$ , the update is suspended.
3. If the test statistic returns to 0, we add to the learning sample all the previous observations, considering that all the past was under  $H_0$ .

The updating approach (choice of the threshold and detection performances) is investigated through numerical simulations in section 5.2.

#### 4. Approximation of the ARL functions

The aim of this section is to propose two methods to evaluate the Average Run Length of the CUSUM scheme (or its nonparametric version without the learning sample updating), which are defined by:

$$ARL_{H_0} = E_{\theta_0}(t_C) \quad \text{and} \quad ARL_{H_1} = E^{(1)}(t_C),$$

and correspond, respectively, to the mean number of observations before a false alarm and the mean delay for detection when  $t_0 = 1$ . The evaluation of the two  $ARL$  allows in particular to compare the performances of different detection rules. Moreover the  $ARL_{H_0}$  can be used to determine the value of the threshold  $h$  (see Section 4.3).

##### 4.1. Analytical approximation

Let us follow the original paper of Page [28], in which the observations  $X_i$  are *iid*, and consider the following extension of the CUSUM rule (2),

$$t_C(s) = \inf\{n : \max_{0 \leq k \leq n} \sum_{i=k}^n Z_i \geq h\}, \quad (12)$$

with  $Z_0 = s$ , with  $s \in [0; h[$ . Let  $J(s)$  denote the ARL of the rule (12), where  $s$  is the starting point of the rule. Note that the ARL of rule (2) is  $J(0)$ .

Page [28], in a context of *iid* observations, establishes the following result:

$$J(s) = 1 + J(0)Q(-s) + \int_0^h J(z)q(z-s) dz \quad (13)$$

where  $h$  is the threshold of the rule (12) and  $Q$  and  $q$  are, respectively, the cumulative distribution function and the density function of  $Z_i$ .

We propose to treat the case of the autoregressive model (1) by considering that (13) is approximatively true for this model. Note that (13) is true as soon as the increments  $Z_i$  are *iid*. For model (1), the  $Z_i$  are identically distributed if the initial distribution  $X_0$  is the stationary distribution, or at least asymptotically identically distributed because of the uniform ergodicity of the process. But they are not independent. The quality of the approximation (13) will then rely on the degree of the serial dependence of the  $Z_i$ . If the serial dependence is negligible, we can expect that the approximation (13) will give relatively good results. If it is not the case, equation (13) is not justified.

An approximation of the ARL of the CUSUM procedure can then be obtained by the resolution of the integral Fredholm equation (13). The numerical resolution of Fredholm equations is a very large topic in applied mathematics and is beyond the scope of this paper. For the ARL evaluation of a detection scheme, the paper of Goel and Wu [11] can be mentioned.

To illustrate the main idea of the resolution of this type of equations, we follow Moustakides et al. [26] and use a very simple numerical technique. Keeping in mind that, as mentioned in Moustakides et al. [26], "if one adopts more powerful numerical integration methods, the results will be of higher accuracy".

Let us consider the following sequence, in order to approach the integral (13):

$$0 = a_1 < a_2 < \dots < a_m = h.$$

For  $z \in [a_i; a_{i+1}]$  and inspired by Moustakides et al. [26]:

$$q(z-s) \approx \frac{Q(a_{i+1}-s) - Q(a_i-s)}{a_{i+1} - a_i}$$

and

$$\int_0^h J(z)q(z-s) dz \approx \sum_{i=1}^m (Q(a_{i+1}-s) - Q(a_i-s)) \frac{J(a_{i+1}) + J(a_i)}{2} \quad (14)$$

Then, by combining (13) and (14), we consider for all  $k$  in  $1, \dots, m$ :

$$j(a_k) = 1 + j(a_1) \cdot Q(-a_k) + \frac{1}{2} \sum_{i=1}^m (Q(a_{i+1}-a_k) - Q(a_i-a_k))(j(a_{i+1}) + j(a_i)),$$

with  $j$  denoting an approximation for  $J$ . It then leads to a system of  $m$  linear equations:

$$j_{m,1} = 1_{m,1} + R_{m,m} j_{m,1}, \quad (15)$$

with  $j_{m,1} = [j(a_1), \dots, j(a_m)]^t$  and  $R_{m,m}$  a  $m \times m$ -matrix, whose solution is given by:

$$j_{m,1} = (I_{m,m} - R_{m,m})^{-1} \cdot 1_{m,1}.$$

To finish, the cumulative distribution function  $Q$  of  $Z_i$  under  $H_0$  (respectively under  $H_1$ ) is unknown but can be replaced, in the linear system (15), by the empirical cumulative distribution obtained by simulating a large number of realizations of the trajectory  $(X_i)_{i \geq 1}$ , and by computing the resulting series of  $(Z_i)_{i \geq 1}$ .

This algorithm leads to an estimation of the  $ARL_{H_0}$  (respectively the  $ARL_{H_1}$ ) for the CUSUM algorithm (see the example developed in Section 5.1). As already mentioned, the quality of the previous procedure depends on the degree of the serial dependence of the  $Z_i$ , but also highly relies on a very good approximation of  $Q$ .

For the nonparametric scheme, the system cannot be exactly simulated since function  $f$  is unknown and is replaced by the nonparametric estimator. If the learning sample is not large enough, the approximation of  $Q$  is not sufficiently accurate and leads to poor ARL estimations (see Section 5.1). For the nonparametric rule, we then recommend to use the numerical approximation evoked in the following subsection in order to evaluate the ARL.

#### 4.2. Numerical approximation

A simple way, but often effective is to approach the ARL ( $ARL_{H_0}$  and  $ARL_{H_1}$ ) by Monte Carlo method. Indeed, since the model is known (or can be

estimated using  $\hat{f}_N$ ), it is possible to generate a large number of trajectories of the model and apply the detection rule to these trajectories for a given threshold  $h$ . When the system is simulated under  $H_0$ , the empirical mean time before the first false alarm is an approximation of the  $ARL_{H_0}$ . An estimation of  $ARL_{H_1}$  is obtained by simulating the model under  $H_1$ .

#### 4.3. Threshold choice

The two previous methods (sections 4.1 and 4.2) allow to determine the threshold value which corresponds to a chosen (by the experimenter)  $ARL_{H_0}$ , i.e. the mean time before the first false alarm, by varying the value  $h$  until the desired value for  $ARL_{H_0}$  is reached (see for example Qiu [32] for an effective algorithm).

The previous methods are applied in section 5.1 for comparing the CUSUM and nonparametric CUSUM rules. For the nonparametric CUSUM rule applied with the learning sample updating, the situation is more complex since a fixed threshold value is not the best solution. In subsection 5.2, an adaptive threshold is used.

## 5. SIMULATIONS

The aim of this section is twofold: firstly the performances of the nonparametric CUSUM rule are compared to that of the classical CUSUM used with or without modeling error. In particular we illustrate, through numerical simulations, the convergence properties (Theorem 2) by showing the good behaviour of the nonparametric CUSUM rule.

Secondly, the learning sample updating approach is studied.

Let us consider the following simplified model of a generic wastewater treatment, obtained by discretization of an original model in continuous time ([4], [35]):

$$\begin{cases} X_{n+1} = (TX_n)\mu(S_n) + X_n - (1 - \theta)U_nTX_n + \epsilon_{n+1}^{(1)} \\ S_{n+1} = (-\frac{TX_n}{\tau})\mu(S_n) + S_n + (1 - \theta)U_n(S_0 - S_n)T + \epsilon_{n+1}^{(2)}. \end{cases} \quad (16)$$

It corresponds to the consumption of a substrate with concentration  $S$  by a bacterial population with concentration  $X$ .  $\mu$  is the biomass growth rate, depending on  $S$ .  $T$  is the time discretization step.  $U$  is the control variable (the feed rate) of the process. The parameter  $\theta$  represents the clog rate of

the feed pump. The objective is to detect a slight feed pump clogging, that is, a change of the value  $\theta$  from  $\theta_0 = 0$  to a predefined value  $\theta_1$ .

In practice, the function  $\mu$  is often badly known. Many models are available in the literature for the growth rate (see [4]), and the choice is not always easy. Moreover, these models involve parameters that are difficult to estimate. These modeling uncertainties must be taken into account by the detection rules. In the sequel, the nonparametric rule will be used assuming that  $\mu$  is unknown.

We are interested in the second equation, dealing with the evolution of the substrate concentration. This equation can be rewritten in the form:

$$S_{n+1} = H_n(S_n)\mu(S_n) + F(S_n, U_n, \theta) + \epsilon_{n+1} \quad (17)$$

with,

$$H_n(S_n) = \frac{-TX_n}{\tau},$$

and

$$F(S_n, U_n, \theta) = S_n + (1 - \theta)U_n(S_0 - S_n)T.$$

According to the approach described in section 3, the function  $\mu$  can be estimated by the following nonparametric estimator:

$$\hat{\mu}(x) = \begin{cases} \frac{\sum_{i=1}^{N-1} \delta_i^{-1} K(\delta_i^{-1}(x - S_i)) H_i^{-1}(S_i) (S_{i+1} - F(S_i, U_i, \theta))}{\sum_{i=1}^{N-1} \delta_i^{-1} K(\delta_i^{-1}(x - S_i))} \\ 0 \text{ if the denominator is 0,} \end{cases}$$

which is a direct adaptation of (10) to the model (17).

The model was simulated with:  $U_n = 0.04$  (the control variable was constant),  $T = 0.17$ ,  $X_0 = 46$  and  $S_0 = 4$ .

The choice of  $\mu$  is discussed in the next subsection. The interested reader can find a full description of the process in [5].

### 5.1. Comparisons CUSUM vs nonparametric CUSUM

In this subsection, the classical CUSUM rule was applied to data generated from model (16), assuming that  $\mu$  follows a Monod model whereas the true biomass growth model was in fact simulated with:

$$\mu(S_n) = (1 - a)\mu^{Monod}(S_n) + a\mu^{Tessier}(S_n)$$



where  $a \in [0; 1]$  is a mixing rate of Monod and Tessier laws, corresponding respectively to:

$$\mu^{Monod}(S_n) = \mu_{max} \frac{S_n}{k_s + S_n}$$

and

$$\mu^{Tessier}(S_n) = \mu_{max} \left( 1 - \exp \left( \frac{-S_n}{k_s} \right) \right),$$

displayed in figure 1.

$k_s$  is a saturation constant and  $\mu_{max}$  is the maximum growth rate of biomass. The simulations were done with  $\mu_{max} = 0.05$  and  $k_s = 1$ .

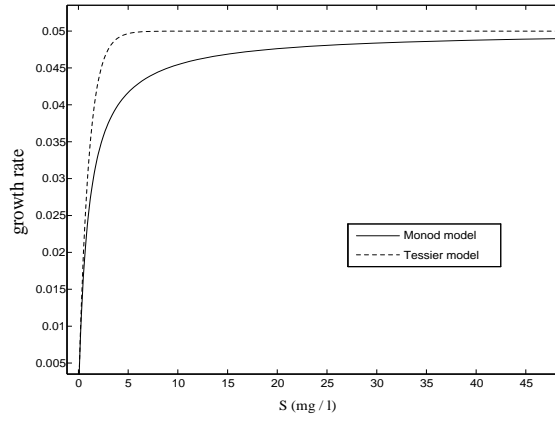


Figure 1: Monod and Tessier laws ( $k_s = 1$ ,  $\mu_{max} = 0.05$ )

Simultaneously, the nonparametric CUSUM was applied with the estimator  $\hat{\mu}$  instead of  $\mu$ , supposed unknown. The usual Gaussian kernel was used ( $K(x) = (1/\sqrt{2\pi}) \cdot \exp(-x^2/2)$ ) with the bandwidth parameter  $\delta_i = 0.5i^{-0.4}$ . The in-control and out-of-control modes were defined by  $H_0 : \theta = \theta_0 = 0$  and  $H_1 : \theta = \theta_1 = 0.01$  respectively. The  $H_1$  hypothesis corresponds to a slight feed pump clogging, inducing a decrease of the effect of the feed rate  $U_n$ . The first five hundred data  $(X_i, S_i)_{i=1, \dots, 500}$  constituted the learning sample for the nonparametric estimator.

Between  $t = 50$  and  $t = 250$ , the system was simulated under  $H_1$  for the reasons mentioned in the remark of part 3.2. The supervision started from  $t = 501$ .

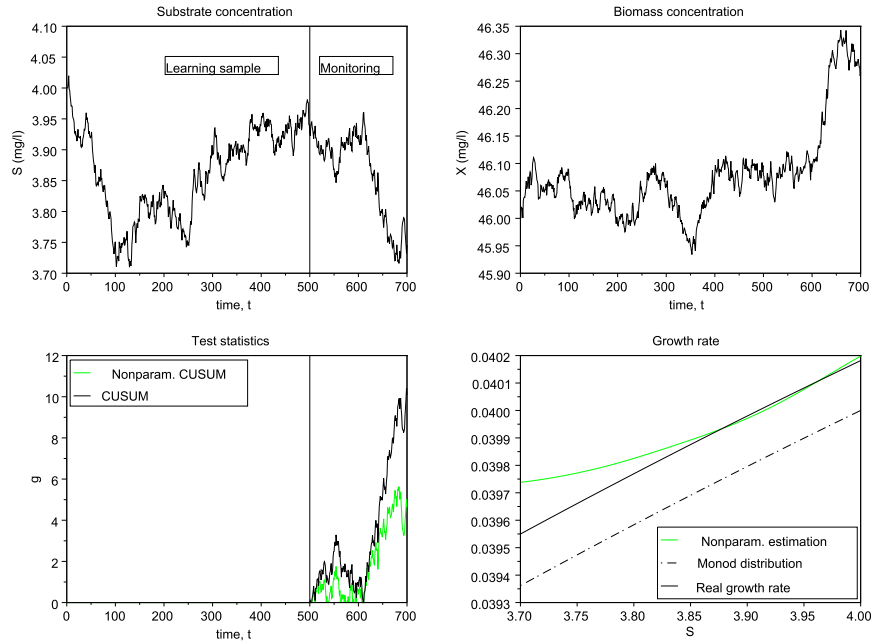


Figure 2: CUSUM and Nonparametric CUSUM rules applied to model (16)

For this first simulation, the mixing rate of the growth model was  $a = 0.02$  and the parameter change to be detected occurred at  $t_0 = 601$ . Graph 4 of Figure 2 represents the different biomass growth rates. The nonparametric estimation is very close to the real growth rate used for the simulation of the observations. This explains why the nonparametric rule is performing better than the original CUSUM rule applied with a false growth rate (Monod law). Indeed, the two rules detect the change without problem (Graph 3) but although the CUSUM test statistic is growing faster than the nonparametric one, the second is in fact more efficient,  $t_0 < \hat{t}_{NP} < t_C$ . Indeed, for a mean time before the first false alarm fixed to 200, the estimated threshold was  $h = 3.83$  for the CUSUM rule and  $h = 2.38$  for the nonparametric CUSUM rule (the method proposed in section 4.3 was used to determine the threshold). Consequently, the respective computed alarm times of the two rules were:  $t_C = 636$  and  $\hat{t}_{NP} = 629$ .

To check the interest of using preferably the nonparametric estimator as suggested by the first simulation, a series of simulations (of the same type as the previous one) were performed to assess the respective detection delays of

the two algorithms, for different values of  $\theta$  and mixing rate  $a$ . The results are presented in table 1. Each case represents an estimation of the mean delay for detection (i.e. the  $ARL_{H_1}$ ) with thresholds chosen such that the mean time before the first false alarm (i.e. the  $ARL_{H_0}$ ) is fixed to 200. Note that the results of table 1 are obtained with a change time fixed to  $t_0 = 501$ , which gives actually the zero-state ARL.

$a$	Rules	$\theta_1 = 0.02$	$\theta_1 = 0.01$	$\theta_1 = 0.005$
0%	<i>CUSUM</i>	14.35	33.80	67.21
	<i>NP – CUSUM</i>	14.40	34.99	78.00
1%	<i>CUSUM</i>	14.69	34.91	68.55
	<i>NP – CUSUM</i>	14.52	35.24	78.43
2%	<i>CUSUM</i>	15.23	38.90	80.14
	<i>NP – CUSUM</i>	14.58	35.70	80.07
5%	<i>CUSUM</i>	26.49	82.85	129.07
	<i>NP – CUSUM</i>	14.12	35.22	79.97

Table 1: Mean time delay for detection ( $ARL_{H_1}$ ).

When the CUSUM is applied with the true model ( $a = 0\%$ ), it gives the better results even if the nonparametric CUSUM is close for  $\theta_1 = 0.02$  and  $\theta_1 = 0.01$ . But when the CUSUM is applied with a modeling error ( $a > 0$ ), its behaviour is rapidly deteriorating as the mixing rate  $a$  increases and the nonparametric CUSUM becomes more efficient (for  $\theta_1 = 0.02$  when  $a = 1\%$  and for  $\theta_1 = 0.01$  and  $\theta_1 = 0.005$  when  $a = 2\%$ ).

Remark that for the very small change ( $\theta_1 = 0.005$ ) and without modeling error ( $a = 0\%$ ), the nonparametric CUSUM is far from the CUSUM (78.00 vs 67.21). We can improve the performance of the nonparametric approach by increasing the size of the learning sample. For example, with a learning sample of 3000 observations, the performance of the two rules are equivalent. This confirms the intuitive argument induced by Theorem 2.

**Remark 4.** Note that all the estimated ARLs of the table 1 were obtained using Monte Carlo simulations (subsection 4.2), with  $M = 30000$  repetitions.

The results obtained by the analytical approximation (subsection 4.1) are presented in Table 2. Note first that as mentioned in subsection 4.1, the

quality of the analytical approximation depends on the serial dependence degree of the series of increments. We then first generate 500 observations of the model (16) (with  $\theta_1 = 0.01$  and  $a = 0\%$ ) and compute the resulting increments  $Z_i, i = 1, \dots, 500$ . The Ljung-Box test [18] but also the nonparametric test for serial independence of Genest and Rémillard [10] are applied on the series  $Z_i, i = 1, \dots, 500$ . The previous procedure is repeated 1000 times. At the significance level 5%, the independence hypothesis is rejected only 42 times with the Ljung-Box test and 53 times with the nonparametric test. This shows that, for model (16), the serial dependence of the increments is negligible, thus justifying the use of the analytical approximation (subsection 4.1).

Using the same thresholds as in Table 1 (only the case  $a = 0\%$  is considered here), the  $ARL_{H_0}$  and  $ARL_{H_1}$  are estimated with  $m = 200$ . For the CUSUM rule,  $10^5$  points of the trajectory  $X$  are generated which leads to a series  $(Z_i)_{i=1, \dots, 10^5}$ . The cumulative distribution function (cdf)  $Q$  is estimated by the empirical cdf of the  $(Z_i)$ . We first remark that the estimated  $ARL_{H_0}$  is close to the expected value ( $ARL_{H_0} = 200$ ) and the estimated mean delays for detection are very close to those obtained in Table 1. It can therefore be seen that the technique employed to solve the Fredholm equation provides, despite its simplicity, valuable results.

For the nonparametric CUSUM rule, the cdf of  $\hat{Z}_i^N$  (for example under  $H_0$ ) is estimated as follows. A kernel density estimation is obtained from the  $\hat{Z}_i^N$  of the learning sample (under  $H_0$ , i.e. the observations 1 to 50 and 251 to 500). The empirical cdf of the  $10^5$  realizations of the distribution induced by this kernel estimator is then used as an estimator of  $Q$ . This estimator obviously depends on the learning sample (300 observations for estimating  $Q$  under  $H_0$  and 200 observations under  $H_1$  for the example considered in this section) and the estimated  $ARLs$  vary from a learning sample to another. The  $ARLs$  presented in Table 2 are, for the non parametric CUSUM rule, the mean of 1000 repetitions of the analytical approximation. We can see that the results are not so good as the previous ones. For example, the estimated  $ARL_{H_0}$  is not very close to 200 for  $\theta_1 = 0.005$ . The reason could be the choice of the bandwidth parameter of the kernel density estimator. For the nonparametric CUSUM rule, we then recommend to use Monte Carlo method to estimate the  $ARLs$ .

$a$	Rules	Mode	$\theta_1 = 0.02$	$\theta_1 = 0.01$	$\theta_1 = 0.005$
0%	<i>CUSUM</i>	$H_0$	198.70	199.55	201.61
		$H_1$	14.15	34.02	67.93
	<i>NP – CUSUM</i>	$H_0$	203.19	198.63	187.65
		$H_1$	15.75	41.54	79.77

Table 2: Estimated  $ARL_{H_0}$  and  $ARL_{H_1}$  obtained by analytical approximation.

### 5.2. Study of the learning sample updating approach

The aim of this subsection is to analyse the potential gain brought by the on-line update of the learning sample presented in section 3.4. Obviously, when the learning sample is large, the kernel estimator can be very accurate and the nonparametric CUSUM shows performances very close to those of the CUSUM. The update of the learning sample is of limited use in this case.

As in part 5.1, the size of the reference learning sample is equal to 500. The monitoring begins at  $t=501$  and the learning sample update is therefore applied until an alarm is triggered. The updated nonparametric CUSUM rule (noted *UNP – CUSUM*) is now compared to the nonparametric CUSUM rule.

The procedure used in part 5.1 for obtaining the thresholds according to the criterion of the mean time before the first false alarm can no longer be employed for the updated nonparametric rule. Indeed, from a theoretical point of view, the estimator of the unknown functional component will be better and better with time. In other words, the distribution of the increments of the rule may change in time. Then, there is no reason to keep a fixed threshold. Margavio et al. [21] proposed, in a control chart context, to consider an adaptive threshold satisfying the following constraint:

$$P_{\theta_0}[g_1 \geq h_1] = \alpha,$$

and  $\forall n \geq 2$ ,

$$P_{\theta_0}[g_n \geq h_n | g_1 < h_1, \dots, g_{n-1} < h_{n-1}] = \alpha, \quad (18)$$

where  $\alpha$  is fixed by the supervisor and  $g$  is the test statistic. At each time step  $n$ ,  $h_n$  is chosen such that the probability to have a false alarm at time  $n$ ,

while there were none before, is equal to  $\alpha$ . The value  $h_n$  which satisfies the constraint (18) is nothing else than the  $(1 - \alpha)$ -quantile of the distribution of  $g_n$  conditional on  $\{g_1 < h_1, \dots, g_{n-1} < h_{n-1}\}$ . An estimate  $\hat{h}_n$  of this quantile can be obtained by simulation, see [37].

In order to properly compare the two rules, the adaptive threshold constraint was applied to each of them with  $\alpha = 0.005$ . Constraint (18), which is more stringent than the ARL constraint (6) led to a mean time before first false alarm equal to  $ARL_{H_0} = 1/\alpha = 200$  (see for example [37]). The mean time delay for detection is presented in table 3, for several instants of change ( $t_0 = 501$ ,  $t_0 = 601$ ,  $t_0 = 701$  and  $t_0 = 801$ ), and for  $\theta_1 = 0.005$ . For each case, 30000 trajectories have been simulated and the mean delay was computed from the valid trajectories, that is to say from the trajectories for which the detection occurred after the change  $t_0$ .

Rules	$t_0 = 501$	$t_0 = 601$	$t_0 = 701$	$t_0 = 801$
$NP - CUSUM$	76.60	76.22	77.81	78.49
$UNP - CUSUM$	80.46	76.04	75.48	75.69

Table 3: Mean time delay for detection ( $ARL_{H_1}$ ) for  $\theta_1 = 0.005$  and  $a = 0\%$ .

The results obtained in table 3 show that the updated nonparametric rule needs, in average, four additional time steps to detect the change for  $t_0 = 501$ , compared to the conventional nonparametric rule. In this case, the update of the learning sample is not correctly done since the observations, classified in  $H_0$  mode, comes actually from a fault mode. This may explain the difference between the two rules. On the contrary, when the change happens later, the update of the learning sample allows to obtain better results than those obtained with the  $NP - CUSUM$ .

Figure 3 presents the final size of the learning samples for the  $UNP - CUSUM$  rule, for the case  $t_0 = 701$ , and obviously for valid trajectories, that is to say for the trajectories for which the detection came after the change time (the false alarms have been removed). For 54% of the (valid) trajectories, the final learning sample size is smaller than 701, and therefore without wrong classified observation. Nearly 75% of the trajectories have less than 30 misclassified observations.

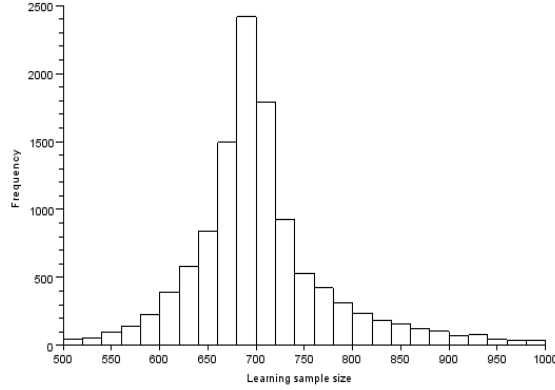


Figure 3: Final size of the learning sample of the  $\hat{t}_{UNP}$  rule for valid trajectoires ( $t_0 = 701$ )

## 6. Conclusion

The CUSUM rule is a popular detection rule widely used in industry. To be applied, it necessitates the exact knowledge of the probability distribution of the observations, in other words, the model describing the evolution of the system. We proposed in this paper a new change detection procedure inspired by the CUSUM rule, for detecting a sudden parameter change in a model with unknown functional component, as it is often the case in real situation. The procedure consists in replacing, in the CUSUM increments calculation, the unknown function by its kernel-based nonparametric estimator built from a learning sample. We show the convergence of the nonparametric CUSUM increments which ensures good performances for our rule, especially as we also showed the asymptotic optimality of the CUSUM rule itself for the model considered. Note however that the convergence of the nonparametric increments is not sufficient to prove the asymptotic optimality of the nonparametric CUSUM rule, which remains an interesting open problem.

The simulations carried out in section 5 show the relevance of this approach on a generic biotechnological model, typical of processes for which there are often modelling uncertainties, as in the bacteria growth rates. The comparison between the respective detection delays of the nonparametric CUSUM rule and the classic CUSUM rule applied on slightly erroneous mod-

els, shows that it is preferable to use the nonparametric approach in case of uncertainty on the system model. Indeed, the standard CUSUM rule is very sensitive to modelling errors. Moreover, the updating procedure of the rule can be a good solution at the beginning of the monitoring procedure, when the learning sample is not rich enough, and needs to be strengthened.

Throughout this paper, we have discussed the problem of parameter change detection when the values  $\theta_0$  and  $\theta_1$  are known. In practice, this assumption can be restrictive. Even if we can imagine that the value  $\theta_0$  characterizing the in-control mode can be estimated in a preliminary phase of observation, the value  $\theta_1$  characterizing the out-of-control mode is most often unknown. Our nonparametric rule can be adapted to this more general case, following a GLR-like approach [20].

### Acknowledgements

The authors would like to thank the associate editor and the reviewers for their valuable comments and suggestions to improve the quality of this paper.

### Appendix A. Proof of Theorem 2

This proof requires the use of the two following lemmas:

**Lemma 3 (Duflo [7]).** *If  $\epsilon = (\epsilon_n)_{n \geq 1}$  is a Gaussian white noise, then*

$$\sup_{k \leq n} \|\epsilon_k\| = o((\log n)^{1/2}) \quad \text{a.s.}$$

**Lemma 4.** *Let  $\hat{f}_N(x)$  be the nonparametric estimate of  $f$  defined in (10). Assume that the two following assumptions hold:*

*Assumption 1 : Function  $f$  is contracting.*

*Assumption 2 :  $\epsilon = (\epsilon_n)_{n \geq 1}$  is a sequence of Gaussian independent and identically distributed random vectors with mean 0 and invertible covariance matrix  $R$ .*

*Then, for  $\delta_i = i^{-1/2(d+1)}$ , any initial law and any  $A < \infty$ ,*

$$\sup_{\|x\| \leq v_N} \|\hat{f}_N(x) - f(x)\| = o(N^{-\lambda}), \quad (\text{A.1})$$



with  $\lambda \in ]0; 1/2(d+1)[$ , and  $v_N$  a sequence converging to infinity defined by:

$$v_N := A(\log \log N)^{1/2}.$$

PROOF OF LEMMA 4. This result states the convergence of the estimator  $\hat{f}_N$  over dilating sets. It is a particular case of a more general result of [31] who proved the convergence of the nonparametric estimator  $\hat{f}_N$  in a controlled model. Complete proof with general noises is also provided in [15].

The proof of Theorem 2 can now be established.

PROOF OF THEOREM 2. For  $r = 0, 1$  and for all  $n \geq 1$ , we have:

$$\begin{aligned} \hat{\eta}_n^r &:= X_n - \hat{f}_N(X_{n-1}) - F(X_{n-1}, \theta_r) \\ &= X_n - \hat{f}_N(X_{n-1}) - F(X_{n-1}, \theta_r) - f(X_{n-1}) + f(X_{n-1}) \\ &= \eta_n^r + (f(X_{n-1}) - \hat{f}_N(X_{n-1})). \end{aligned}$$

Let  $b_{n-1}^N := f(X_{n-1}) - \hat{f}_N(X_{n-1})$ , then  $\hat{\eta}_n^r = \eta_n^r + b_{n-1}^N$ . The increments of the nonparametric rule satisfy, for all  $n$ ,

$$\begin{aligned} \hat{Z}_n^N &= \frac{1}{2} [(\eta_n^0 + b_{n-1}^N)^t R^{-1}(\eta_n^0 + b_{n-1}^N) - (\eta_n^1 + b_{n-1}^N)^t R^{-1}(\eta_n^1 + b_{n-1}^N)] \\ &= \frac{1}{2} [(\eta_n^0)^t R^{-1} \eta_n^0 - (\eta_n^1)^t R^{-1} \eta_n^1 + 2(\eta_n^0)^t R^{-1} b_{n-1}^N - 2(\eta_n^1)^t R^{-1} b_{n-1}^N], \end{aligned}$$

because  $R^{-1}$  is symmetric. Therefore, for all  $n$ ,

$$\begin{aligned} \hat{Z}_n^N &= Z_n + \left[ \left( (\eta_n^0)^t - (\eta_n^1)^t \right) R^{-1} b_{n-1}^N \right] \\ &= Z_n + \left[ (F(X_{n-1}, \theta_1) - F(X_{n-1}, \theta_0))^t R^{-1} b_{n-1}^N \right], \end{aligned}$$

and,

$$\begin{aligned} \|\hat{Z}_n^N - Z_n\| &= \| (F(X_{n-1}, \theta_1) - F(X_{n-1}, \theta_0))^t R^{-1} b_{n-1}^N \| \\ &\leq \| (F(X_{n-1}, \theta_1) - F(X_{n-1}, \theta_0)) \| \cdot \| R^{-1} \| \cdot \| b_{n-1}^N \| \\ &\leq 2M \| R^{-1} \| \cdot \| b_{n-1}^N \|. \end{aligned} \tag{A.2}$$

Since the noise in model (1) is Gaussian, the observations  $(X_n)_{n \geq 1}$  are not bounded. Therefore, a classic convergence result of  $\hat{f}_N$  on compact sets

is not sufficient to control the quantity  $\|b_{n-1}^N\|$ . In order to use the convergence result over dilated sets with the sequence  $v_N$  defined in Lemma 4, the following decomposition of (A.2) is considered:

$$\begin{aligned}\|\hat{Z}_n^N - Z_n\| &\leq 2M\|R^{-1}\| \cdot \|b_{n-1}^N\| \\ &\leq 2M\|R^{-1}\| \cdot \|f(X_{n-1}) - \hat{f}_N(X_{n-1})\| \mathbf{1}_{\{\|X_n\| \leq v_N\}} \\ &\quad + 2M\|R^{-1}\| \cdot \|f(X_{n-1}) - \hat{f}_N(X_{n-1})\| \mathbf{1}_{\{\|X_n\| > v_N\}} \quad (\text{A.3})\end{aligned}$$

- For the first part of the decomposition (A.3), it comes:  $\forall n \geq 1$ ,

$$\begin{aligned}2M\|R^{-1}\| \cdot \|f(X_{n-1}) - \hat{f}_N(X_{n-1})\| \mathbf{1}_{\{\|X_n\| \leq v_N\}} \\ \leq 2M\|R^{-1}\| \sup_{\|x\| \leq v_N} \|\hat{f}_N(x) - f(x)\| \\ \rightarrow 0\end{aligned}$$

when  $N \rightarrow \infty$ , according to Lemma 4.

- The study of the second part is as follows: since  $f$  is Lipschitz and  $F$  is continuous and bounded, it holds:

$$\|f(x) + F(x, \theta)\| \leq r_f \|x\| + M'.$$

By definition of model (1), it comes for all  $m \geq 1$ :

$$\|X_m\| \leq r_f \|X_{m-1}\| + M' + \|\epsilon_m\|.$$

Therefore

$$\begin{aligned}\|X_m\| &\leq r_f^m \|X_0\| + B + \sum_{j=1}^m r_f^{m-j} \|\epsilon_j\| \\ &\leq r_f^m \|X_0\| + B + \sup_{k \leq m} \|\epsilon_k\| \cdot \left( \sum_{j=1}^m r_f^{m-j} \right) \\ &\leq r_f^m \|X_0\| + B + C \cdot \sup_{k \leq m} \|\epsilon_k\|\end{aligned}$$

with  $B$  and  $C$  two constants and  $X_0 = x_0$  the initial condition. Finally,

$$\|X_m\| = O(\sup_{k \leq m} \|\epsilon_k\|) \quad \text{a.s.,}$$

Lemma 3 yields,

$$\|X_m\| = O((\log m)^{1/2}) \quad \text{a.s.}$$

By a similar argument to that used in [7] (Lemma 1.3.20, p.23), it results:

$$\sup_{k \leq m} \|X_k\| = O((\log m)^{1/2}) \quad \text{a.s.}$$

Taking  $N = \exp(m)$ ,

$$\sup_{k \leq m} \|X_k\| = O((\log \log N)^{1/2}) \quad \text{a.s.},$$

which yields for  $n \leq \log(N)$ ,

$$\limsup_{N \rightarrow \infty} \frac{\|X_n\|}{(\log \log N)^{1/2}} < A \quad \text{a.s.}$$

for a constant  $A$ . With  $v_N = A(\log \log N)^{1/2}$ , it follows

$$\limsup_{N \rightarrow \infty} \frac{\|X_n\|}{v_N} < 1 \quad \text{a.s.},$$

and  $\mathbf{1}_{\{\|X_n\| > v_N\}} = 0$  for all  $n \leq \log(N)$  and for sufficiently large  $N$ . Therefore the second part of the decomposition (A.3) is also zero for sufficiently large  $N$ . The proof of the Theorem 2 is completed.

## References

- [1] A. Baillo, A. Cuevas, Parametric versus nonparametric tolerance regions in detection problems, Computational Statistics 21 (3-4) (2006) 523–536.
- [2] R. K. Bansal, P. Papantoni-Kazakos, An algorithm for detecting a change in a stochastic process, IEEE Transactions on Information Theory IT-32 (1986) 227–235.
- [3] M. Basseville, I. Nikiforov, Detection of Abrupt Changes. Theory and Application, Prentice-Hall, 1993.
- [4] G. Bastin, D. Dochain, On-line Estimation and Adaptive Control of Bioreactors, Elsevier, 1990.

- [5] O. Bernard, Z. Hadj-Sadok, D. Dochain, A. Genovesi, J.-P. Steyer, Dynamical model development and parameter identification for an anaerobic wastewater treatment process, *Biotechnology and Bioengineering* 75, 4 (2001) 424–438.
- [6] D. Bosq, *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics 110, Springer, 1996.
- [7] M. Duflo, *Random Iterative Models*, Springer-Verlag, 1997.
- [8] M. Fouladirad, I. Nikiforov, Optimal statistical fault detection with nuisance parameters, *Automatica* 41 (2005) 1157–1171.
- [9] C.-D. Fuh, SPRT and CUSUM in hidden markov models, *Annals of Statistics* 31 (2003) 942–977.
- [10] C. Genest, B. Rémillard, Tests of independence and randomness based on the empirical copula process, *Test* 13 (2) (2004) 335–369.
- [11] A. L. Goel, S. M. Wu, Determination of ARL and a contour nomogram for CUSUM charts to control normal mean, *Technometrics* 13 (2) (1971) 221–230.
- [12] E. Gombay, D. Serban, Monitoring parameter change in  $AR(p)$  time series models, *Journal of multivariate analysis* 100 (2009) 715–725.
- [13] F. Harrou, L. Fillatre, I. Nikiforov, Anomaly detection/detectability for a linear model with a bounded nuisance parameter, *Annual Reviews in Control* 38 (2014) 32–44.
- [14] U. Herkenrath, On the uniform ergodicity of Markov processes of order 2, *Journal of Applied Probability* 40 (2003) 455–472.
- [15] N. Hilgert, B. Portier, Strong uniform consistency and asymptotic normality of a kernel based error density estimator in functional autoregressive models, *Statistical Inference for Stochastic Processes* 15 (2) (2011) 105–125.
- [16] T. L. Lai, Information bounds and quick detection of parameter changes in stochastic systems, *IEEE Transactions on Information Theory* 44 (1998) 2917–2929.

- [17] T. L. Lai, Sequential analysis : Some classical problems and new challenges, *Statistica Sinica* 11 (2001) 303–408.
- [18] G. M. Ljung, G. E. Box, On a measure of lack of fit in time series models, *Biometrika* 65 (2) (1978) 297–303
- [19] G. Lorden, Procedures for reacting to a change in distribution, *Annals of Mathematical Statistics* 42 (1971) 1897–1908.
- [20] G. Lorden, Open-ended tests for Koopman-Darmois families, *Annals of Statistics* 1 (1973) 633–643.
- [21] T. M. Margavio, M. D. Conerly, W. H. Woodall, L. G. Drake, Alarm rates for quality control charts, *Statistics & Probability Letters* 24 (3) (1995) 219–224.
- [22] Y. Mei, Sequential change-point detection when unknown parameters are present in the pre-change distribution, *Annals of Statistics* 34 (2006) 92–122.
- [23] D. Montgomery, *Introduction to Statistical Quality Control*, Wiley, New York, 1996.
- [24] G. Moustakides, Optimal procedures for detecting change in distribution, *Annals of Statistics* 14 (1986) 1379–1387.
- [25] G. Moustakides, A. S. Polunchenko, A. G. Tartakovsky, Numerical comparison of cusum and shiryaev–roberts procedures for detecting changes in distributions, *Communications in Statistics - Theory and Methods* 38 (16-17) (2009) 3225–3239.
- [26] G. Moustakides, A. S. Polunchenko, A. G. Tartakovsky, A numerical approach to performance analysis of quickest change-point detection procedures, *Statistica Sinica* 21 (2011) 571–596.
- [27] E. A. Nadaraya, On estimating regression, *Theory of Probability and its Applications* 9 (1964) 141–142.
- [28] E. S. Page, Continuous inspection schemes, *Biometrika* 41 (1954) 100–115.

- [29] E. Parzen, On estimation of a probability density function and mode, *Annals of Mathematical Statistics* 33 (1962) 1065–1076.
- [30] S. M. Pergamenchchikov, A. G. Tartakovsky, Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data, *arXiv preprint arXiv:1510.02903v2*, 2016.
- [31] B. Portier, A. Oulidi, Nonparametric estimation and adaptive control of functional autoregressive models, *SIAM Journal on Control and Optimization* 39 (2000) 411–432.
- [32] P. Qiu, Distribution-free multivariate process control based on log-linear modeling, *IIE Transactions* 40 (7) (2008) 664–677.
- [33] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* 27 (1956) 832–837.
- [34] R. Senoussi, Uniform iterated logarithm laws for martingales and their application to functional estimation in controlled Markov chains, *Stochastic Processes and their Applications* 89 (2) (2000) 193–211.
- [35] J.-P. Steyer, O. Bernard, An example of the benefits obtained from the long term use of mathematical models in wastewater biological treatment, in: *Proc. of the 4th MATHMOD International Symposium on Mathematical Modelling*, Vienna, Austria, 2003, pp. 245–251.
- [36] J. Y. Su, F. F. Gan, X. Tang, Optimal cumulative sum charting procedures based on kernel densities, in: *Frontiers in Statistical Quality Control* 11, Springer, 2015, pp. 119–134.
- [37] G. Verdier, N. Hilgert, J.-P. Vila, Adaptive threshold computation for CUSUM-type procedures in change detection and isolation problems, *Computational Statistics and Data Analysis* 52 (9) (2008) 4161–4174.
- [38] G. Verdier, N. Hilgert, J.-P. Vila, Optimality of CUSUM Rule Approximations in Change-Point Detection Problems: Application to Nonlinear State-Space Systems, *IEEE transactions on Information Theory* 54 (11) (2008) 5102–5112.