



HAL
open science

Détection Statistique de Rupture de Modèle dans les Systèmes Dynamiques - Application à la Supervision de Procédés de Dépollution Biologique

Ghislain Verdier

► **To cite this version:**

Ghislain Verdier. Détection Statistique de Rupture de Modèle dans les Systèmes Dynamiques - Application à la Supervision de Procédés de Dépollution Biologique. Mathématiques [math]. Université Montpellier II - Sciences et Techniques du Languedoc, 2007. Français. NNT: . tel-00221418

HAL Id: tel-00221418

<https://theses.hal.science/tel-00221418>

Submitted on 28 Jan 2008

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.

UNIVERSITÉ MONTPELLIER II

SCIENCES ET TECHNIQUES DU LANGUEDOC

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ MONTPELLIER II

Discipline : Mathématiques appliquées

Ecole Doctorale : Information, Structures et Systèmes

Formation Doctorale : Biostatistique

Détection Statistique de Rupture de Modèle dans les Systèmes Dynamiques - Application à la Supervision de Procédés de Dépollution Biologique

par

Ghislain VERDIER

soutenue publiquement le **30 novembre 2007** devant le jury composé de :

Mme	M. BASSEVILLE	Dir. de recherche, IRISA Rennes	Rapporteur
MM.	A. BERLINET	Professeur, Université Montpellier II	Président
	P. DEL MORAL	Dir. de recherche, INRIA Bordeaux	Rapporteur
Mme	N. HILGERT	Chargée de recherche, INRA Montpellier	Membre invitée
MM.	G. MOUSTAKIDES	Professeur, Université de Patras	Rapporteur
	I. NIKIFOROV	Professeur, Université de Troyes	Examineur
	B. PORTIER	Professeur, INSA Rouen	Examineur
	J-P. STEYER	Dir. de recherche, INRA Narbonne	Examineur
	J-P. VILA	Dir. de recherche, INRA Montpellier	Dir. de Thèse

Remerciements

Tout d'abord, je tiens à remercier très sincèrement mes directeurs de thèse Naline Hilgert et Jean-Pierre Vila. Ils ont fait preuve à mon égard d'une très grande disponibilité tout au long de ces trois années. Leurs compétences scientifiques, leur rigueur, leur ouverture d'esprit mais aussi leur bonne humeur m'ont permis de mener ce travail à terme. Travailler avec eux a été fort agréable et très enrichissant.

J'adresse un grand merci à Michèle Basseville, Pierre Del Moral et George Moustakides pour l'intérêt qu'ils ont porté à mon travail en acceptant d'en être les rapporteurs. Leurs remarques et leurs conseils m'ont été très utiles pour la rédaction de la version finale du manuscrit, mais aussi pour ouvrir des perspectives à ce travail.

Je remercie également Alain Berlinet d'avoir accepté spontanément d'être le président du jury, et Igor Nikiforov, Bruno Portier et Jean-Philippe Steyer pour avoir participé à ce jury.

Mon travail a été suivi tout au long de ces trois années par un comité de pilotage de thèse. Je tiens à remercier très chaleureusement l'ensemble des membres de ce comité. Fabien Campillo, Igor Nikiforov et Bruno Portier, de par leurs commentaires et leurs nombreux conseils judicieux, ont contribué à l'avancement de ma thèse. Jean-Philippe Steyer a été d'une grande disponibilité malgré un emploi du temps très chargé. Il m'a permis de tester les approches méthodologiques que j'ai développé sur le procédé pilote du LBE de Narbonne.

Je n'oublie pas de saluer tous les membres de l'UMR Analyse des Systèmes et Biométrie qui m'ont accueilli très chaleureusement et m'ont permis de travailler dans les meilleures conditions. Je remercie donc le directeur de l'UMR, Pierre Cartigny, mais aussi : Alain, Anne, Brigitte, Cécile, Christophe, Frédéric, Gilles, Luc, Martine, Nicolas, Pascal, Patrice, Philippe, Pierre-André, Véronique. J'ai aussi une pensée pour tous les thésards ou stagiaires que j'ai rencontré pendant cette thèse. Parmi eux : Ahmad, Hazael, Laurent, Thomas, Rémi, Véra, Virginie, Vivien.

Je remercie également les deux cofinanceurs de ma thèse : le département MIA de l'INRA et la région Languedoc-Roussillon.

Enfin, j'adresse une pensée à mes proches, membres de ma famille ou amis, qui m'ont toujours soutenu et encouragé tout au long de mes études. Je remercie particulièrement Marion, pour son infinie patience et son soutien affectif.

Table des matières

Introduction	9
I Etat de l'art	17
1 Méthodes de détection et de diagnostic	19
1.1 Approches sans modèle analytique	20
1.2 Approches basées sur un modèle analytique	21
1.3 Approches relevant de l'intelligence artificielle	25
1.4 Approches statistiques	26
2 Méthodes statistiques de détection	31
2.1 Introduction	31
2.2 De la règle de Shewart au CUSUM	33
2.2.1 Le rapport de vraisemblance	33
2.2.2 Les règles de décision élémentaires	33
2.3 La règle du CUSUM	34
2.3.1 Données indépendantes	35
2.3.2 Données dépendantes	37
2.3.3 Résultats d'optimalité	40
2.4 Dérivation du CUSUM	43
2.4.1 Paramètre caractérisant la panne inconnu	43
2.4.2 Problèmes plus complexes	46
2.5 Autres méthodes de détection	46
2.5.1 Les méthodes bayésiennes	46
2.5.2 L'approche asymptotique locale	47
2.6 Détection dans les modèles à espace d'état	48
2.6.1 Panne additive dans un modèle à espace d'état linéaire	48
2.6.2 Modèles plus généraux et filtrage particulière	49
2.7 Le diagnostic d'anomalie	50
2.7.1 L'algorithme récursif de Nikiforov [74]	51

2.7.2	Application	53
II	Nouvelles approches de la détection de rupture de modèle pour les modèles dynamiques incertains et les modèles dynamiques à espace d'état	57
3	Seuils adaptatifs	59
3.1	Introduction	59
3.2	Choix du seuil et intérêt de notre approche	60
3.3	Seuil adaptatif pour l'algorithme CUSUM	62
3.3.1	Description et propriétés	62
3.3.2	Propriétés de convergence	64
3.3.3	Simulations	67
3.4	Seuil adaptatif dans le cas du diagnostic	71
3.4.1	L'algorithme et les contraintes initiales	71
3.4.2	Estimation des seuils	74
3.4.3	Simulation	77
3.5	Conclusion	79
4	Une approximation de la règle du CUSUM	83
4.1	Le "CUSUM approché"	84
4.2	Optimalité de la règle du CUSUM	84
4.3	Optimalité de la règle du CUSUM approché au sens de Lorden	88
4.4	Autre critère d'optimalité	99
4.5	Conclusions	104
5	Le CUSUM non paramétrique avec apprentissage	105
5.1	Cas où la fonction f est connue	106
5.1.1	Ecriture de la règle du CUSUM	106
5.1.2	Optimalité de la règle du CUSUM pour le modèle (5.1) lorsque f est connue	107
5.2	Le CUSUM non paramétrique avec apprentissage	112
5.2.1	L'estimateur à noyau de convolution	113
5.2.2	Remarques sur l'échantillon d'apprentissage de l'estimateur	113
5.2.3	Test CUSUM non paramétrique	114
5.3	Propriétés théoriques du test	116
5.4	Simulations	120
5.4.1	Cas où f suit une loi de Monod	121
5.4.2	Cas où f est contaminée	124
5.5	Conclusions	128

6	Algorithmes de filtrage-détection	129
6.1	Introduction	129
6.2	Le cas général	131
6.2.1	La modélisation et ses conséquences	131
6.2.2	Estimation de la vraisemblance par intégration de Monte Carlo	132
6.2.3	Principe de l'algorithme et résultats	135
6.2.4	Notion d'optimalité dans les modèles à espace d'état	137
6.2.5	L'algorithme de filtrage-détection	137
6.2.6	Problème de l'accroissement du temps de calculs	139
6.2.7	Simulations	143
6.3	Paramètre caractérisant la panne inconnu	148
6.3.1	Principe de l'algorithme	148
6.3.2	Description de l'algorithme	151
6.3.3	Méthodes d'approximation	154
6.3.4	Simulations	155
6.4	L'algorithme de filtrage-détection comme concurrent du GLR	163
6.4.1	Introduction	163
6.4.2	Principe et détail de l'algorithme	164
6.4.3	Résultats d'optimalité	166
6.4.4	Une autre règle de décision	174
6.4.5	Simulations	175
6.5	L'algorithme de filtrage-détection pour le diagnostic de panne	178
6.5.1	La règle de décision à utiliser	178
6.5.2	Optimalité des approches proposées	180
6.5.3	Simulations	180
6.6	Conclusions	183
7	Applications	189
7.1	Introduction	189
7.1.1	Le modèle	190
7.1.2	Les données	191
7.1.3	Les types de panne étudiés	194
7.1.4	Le cahier des charges de la supervision du procédé	196
7.2	Panne de capteur	197
7.2.1	La modélisation stochastique	197
7.2.2	Détection sur des données simulées	198
7.2.3	Applications aux données réelles	202
7.2.4	Conclusions sur la panne de capteur	206
7.3	Le changement de modèle sur X_1	207
7.3.1	Les deux modèles	209
7.3.2	Détection du changement de régime en simulation	209

TABLE DES MATIÈRES

7.3.3	Détection du changement de régime sur les données réelles . . .	210
7.4	Conclusions	214
Conclusion		217
III	Annexes	219
A	Les estimateurs à noyaux de convolution	221
A.1	Principe de l'estimation non paramétrique	221
A.2	L'estimateur de la densité	222
A.3	L'estimateur de la régression	223
B	Méthodes de filtrage	225
B.1	Modèle à espace d'état et filtre optimal	225
B.2	Filtrage particulaire	227
B.2.1	Filtre de Monte Carlo classique	227
B.2.2	Filtre de Monte Carlo avec ré-échantillonnage	228
B.3	Filtres à noyaux de convolution (Rossi et Vila [91, 92])	229
B.3.1	Filtre à noyaux à mémoire complète	230
B.3.2	Filtre par noyaux avec ré-échantillonnage	231
B.3.3	Estimation de la densité conditionnelle du paramètre θ (Rossi et Vila [91])	231
C	Les procédés biotechnologiques	233
C.1	Généralités	233
C.2	Modélisation par bilan de matière	234
C.3	Les paramètres cinétiques	236
C.4	La méthanisation	237
Bibliographie		239

Introduction

Depuis le milieu du vingtième siècle, il existe un intérêt grandissant concernant la détection de rupture dans les modèles dynamiques, notamment en supervision des procédés. En effet, l'opérateur cherche constamment à maîtriser le plus d'aspects possibles du système auquel il est confronté dans le but de le rendre le plus efficace possible. On parle ici de système au sens général du terme, c'est à dire pouvant représenter une voiture, un procédé biologique, une rivière, ou autre...

Une étape importante de la supervision est le diagnostic de panne ou d'anomalie "en ligne", qui consiste d'une part, à détecter et localiser une panne ou un dysfonctionnement, puis à rechercher la cause de cette panne, et enfin à agir sur le système pour le remettre dans un état de fonctionnement normal. Dans cette thèse, on ne s'intéressera uniquement qu'à la première étape de détection et localisation.

Il y a donc une demande croissante pour que les systèmes soient plus sûrs. Evidemment, à l'origine, les méthodes de détection mises au point étaient destinées à des applications industrielles à haut niveau de risque, comme le nucléaire et l'aérospatiale. Mais rapidement, l'intérêt pour la sûreté s'est porté sur un grand nombre de procédés divers. En effet, compte tenu des coûts de fonctionnement et des objectifs fixés en terme de rentabilité d'une usine de fabrication quelconque, il apparaît important de pouvoir mettre en place une procédure de diagnostic, permettant de détecter une anomalie avant qu'elle dégénère et entraîne un arrêt total du système pendant un certain temps, ou plus simplement permettant d'effectuer une maintenance "en ligne" du procédé, ou encore et surtout d'assurer la bonne qualité du produit.

Les méthodes de détection, et en particulier les méthodes statistiques de détection, trouvent donc un grand nombre d'applications, comme le contrôle de qualité dans l'industrie agro-alimentaire, la détection de changement en économie ou encore la sismologie. Depuis le début du vingt-et-unième siècle, ce type de méthodes a trouvé un autre champ d'application concernant le contre-terrorisme et la détection d'attaque biologique (Stoto et al. [104]).

L'application principale de ce travail de thèse concerne la surveillance d'un procédé de dépollution biologique. Un tel procédé consiste en la dégradation du substrat (le polluant) par des bactéries (la biomasse). Les procédés de traitement

Introduction

des substances polluantes sont de nos jours utilisés à grande échelle, notamment dans les stations d'épuration. Les volumes à traiter sont importants (60 millions de m³ d'eaux usées sont produits par jour en Europe) et ils possèdent de plus une charge polluante à grande variabilité. A cette grande variabilité des intrants s'ajoute une connaissance insuffisante des dynamiques des populations bactériennes utilisées dans la dégradation des polluants.

La conduite des procédés de dépollution biologique est donc un problème complexe. La moindre défaillance sur le procédé (contamination bactérienne, défaut de capteur, colmatage de conduit,...) peut limiter les performances des installations et avoir de graves conséquences si elle n'est pas détectée rapidement. En effet, après un arrêt total du bioréacteur, il faut plusieurs semaines pour redémarrer le processus et retrouver un fonctionnement normal du système. Il s'avère donc fondamental de pouvoir surveiller ces procédés en temps réel de manière à garantir la sécurité matérielle et humaine, tout en atteignant les objectifs d'exploitation fixés. Revenons sur ce dernier point. Les eaux rejetées après traitement dans une station d'épuration ne sont pas destinées à produire directement de l'eau potable. Elles sont généralement renvoyées vers les rivières et réintègrent ainsi le cycle de l'eau où elles pourront éventuellement, plus tard, être pompées pour produire (après de nouveaux traitements) de l'eau potable. Il existe donc une réglementation européenne concernant la qualité des eaux rejetées par les stations d'épuration. Cette réglementation fixe entre autres les taux minimum de réduction des polluants (entre 70% et 90%). Une panne non détectée peut entraîner un rejet massif d'eaux usées non traitées, ce qui peut avoir de graves conséquences sur la vie d'un cours d'eau qui reçoit cette pollution, surtout en période d'étiage.

L'objectif de cette thèse est donc de mettre au point des méthodes statistiques de détection d'anomalie qui permettent de répondre au mieux aux spécificités des problèmes de dépollution. Ce travail s'inscrit dans une collaboration de plusieurs années déjà, entre le laboratoire d'Analyse des Systèmes et Biométrie de l'INRA de Montpellier et le Laboratoire de Biotechnologie de l'Environnement de Narbonne sur des problèmes de conduite de procédés biologiques de dépollution. La collaboration entre les deux unités a d'ailleurs permis d'étudier les problèmes du contrôle (thèses de Hilgert [40] et Wagner [111]) et du filtrage (thèse de Rossi [90]) pour ce type de procédés. Cette thèse s'inscrit donc dans la continuité de ces travaux.

*

* *

L'automatique et l'intelligence artificielle sont les grandes disciplines utilisées généralement pour traiter les problèmes de détection de panne. Mais la théorie des tests statistiques propose des approches compétitives qui animent une communauté

de statisticiens. On peut trouver dans Basseville et Nikiforov [6] et Lai [52] une large revue de tout ce qui a été fait depuis le milieu du vingtième siècle dans ce domaine. Pour une introduction plus générale à la détection et au diagnostic de panne, on peut citer Steyer, Génovési et Harmand [103].

Rentrons à présent un peu plus en détail sur les aspects méthodologiques des procédures de détection. Il convient tout d'abord de définir la terminologie que nous allons utiliser tout au long de ce document.

- Nous définirons une **anomalie** sur un procédé comme une déviation non permise d'une propriété ou d'un paramètre caractéristique du système par rapport aux conditions normales de fonctionnement. Dans la suite, nous utiliserons sans distinction les termes : panne, défaut, changement et anomalie pour signifier un comportement anormal du système bien que le vocabulaire usuel en diagnostic distingue ces notions ([103]).

- Nous définirons le problème de la **détection** de la façon suivante : on considère deux régimes de fonctionnement pour le système sous surveillance : un régime de fonctionnement normal et un régime de panne. Initialement, le système est sous le régime de fonctionnement normal et à un instant inconnu, il passe sous le régime de panne. L'objectif de la détection est de détecter le plus rapidement possible ce changement de régime.

- Le **diagnostic** d'anomalie sera la généralisation du problème de détection au cas où plusieurs pannes peuvent affecter le système. On considère ainsi un régime de fonctionnement normal et K régimes de fonctionnement anormal, chacun relié à une panne. L'objectif est alors de détecter un changement de régime et de localiser la panne correspondante.

Dans ce travail, la plupart des résultats obtenus traitent du cas de la détection. Cependant, nous avons cherché à appliquer les algorithmes de détection mis au point au diagnostic, puisque cela constitue le problème le plus fréquemment rencontré en pratique.

Il existe plusieurs types de pannes sur un procédé qui se différencient suivant le comportement au cours du temps de la variable d'intérêt du système. Nous avons représenté dans la figure 1 trois types de pannes sur un processus gaussien.

- **Changement abrupt** : Ce type de panne correspond à un saut brutal du signal. Il peut correspondre à un arrêt total d'un capteur, à une déconnection... Il est souvent modélisé par un changement brutal de paramètre.

- **Changement graduel** : Ce type de défaut correspond typiquement à un encrassement d'une pompe. Il y a une dérive lente et donc un éloignement progressif

Introduction

du signal de sa valeur nominale. Il est difficile à détecter rapidement notamment sur des processus non stationnaires.

- **Changement intermittent** : Il s'agit de défauts caractéristiques des faux contacts, par exemple sur un capteur. Le signal va et vient entre les valeurs normales et les valeurs sous panne.

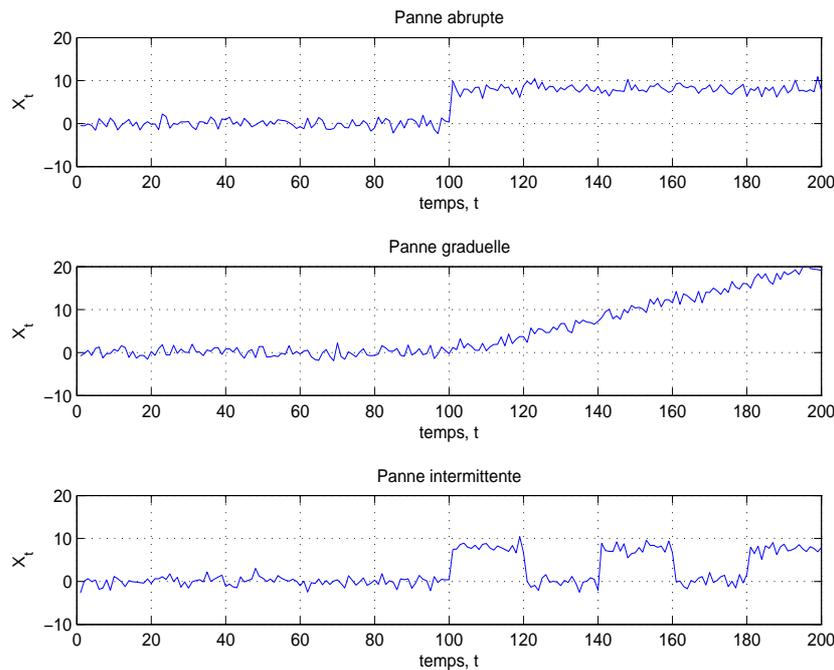


FIG. 1 – Evolution au cours du temps des trois types de pannes sur une variable X

Tout au long de ce document, les règles de décision que nous allons étudier vont concerner la détection de changements abrupts. A l'instant de panne inconnu, on va supposer que le système passe d'un modèle sous régime de fonctionnement normal à un modèle sous régime de panne. Il est donc nécessaire de pouvoir modéliser le système sous son comportement de panne. Une connaissance importante du procédé en fonctionnement anormal sera souvent primordiale. Ainsi, les méthodes que nous proposons ne pourront pas s'appliquer à la détection d'une panne jamais rencontrée auparavant.

Les premiers travaux statistiques concernant ce type de changement sont dus à Shewhart [96], mais une des premières règles de décision séquentielles, et qui reste encore aujourd'hui largement utilisée dans l'industrie, est la règle des sommes cumulées (règle du CUSUM). Initialement, cette règle était destinée à détecter un

changement de paramètre dans la densité d'une série d'observations indépendantes et identiquement distribuées avant et après l'instant de panne. Elle a été ensuite adaptée au cas de systèmes dynamiques, c'est-à-dire à des données dépendantes. La règle du CUSUM permet alors de détecter un changement de paramètre dans la densité conditionnelle des observations, supposée connue.

Aujourd'hui, il est nécessaire de mettre au point des méthodes de détection et de diagnostic s'appliquant à des systèmes de plus en plus complexes, notamment des systèmes où la densité conditionnelle des observations n'est pas connue. Les procédés de dépollution font partie de ces systèmes complexes puisqu'ils possèdent plusieurs particularités qui rendent l'utilisation directe de méthodes classiques de type CUSUM difficile, voire impossible.

La première particularité de ces procédés de dépollution concerne les incertitudes de modèle, notamment au niveau de la modélisation du taux de croissance de la biomasse. En effet, il est toujours difficile de choisir une loi caractérisant l'évolution de ce taux de croissance et ces incertitudes de modélisation doivent être prises en compte par les algorithmes de détection d'anomalie. Nous nous intéressons donc par la suite à des modèles de la forme :

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

où X_n est la variable d'état caractérisant le système, f est une fonction inconnue, F est connue, θ caractérise la panne sur le système (θ passe à l'instant de panne t_p d'une valeur nominale θ_0 à une valeur θ_1) et ϵ est un bruit blanc. La fonction f étant inconnue, il est alors impossible d'appliquer des règles classiques de type CUSUM. Notre approche consiste à utiliser un estimateur non paramétrique de la fonction f , de sorte à pouvoir construire une règle du CUSUM dite "non paramétrique", grâce à cet estimateur.

L'autre principale particularité des procédés de dépollution est que parfois, l'absence de capteur (pour des raisons financières ou autres) empêche d'obtenir directement les mesures des variables d'intérêt. Les modèles qui nous intéressent se décomposent alors en deux parties, une équation d'état décrivant l'évolution des variables d'intérêt non mesurées au cours du temps, et une équation d'observation qui établit le lien entre variables d'intérêt et variables d'observation. Un tel modèle s'écrit :

$$\begin{cases} X_n = f(X_{n-1}, \theta, v_n) \\ Y_n = g(X_n, \theta, w_n) \end{cases}, \quad (2)$$

avec X_n variable d'état (ou d'intérêt) non mesurée, Y_n variable d'observation, v et w des bruits blancs et f et g des fonctions connues. Pour des systèmes linéaires, une solution efficace au problème de détection de changement de paramètre est proposée par Willisky et Jones [112]. Lorsque le système est non linéaire, les choses se

Introduction

compliquent et la règle de Willsky et Jones n'est plus aussi efficace. Quelques auteurs se sont intéressés à ce problème et la majorité des solutions proposées reposent sur l'utilisation de filtres de Monte Carlo (ou filtres particuliers), très en vue depuis une dizaine d'années.

Les méthodes de détection que nous proposons par la suite sont basées sur l'estimation de la densité conditionnelle des observations à partir d'une méthode de filtrage particulière originale (Rossi [90], Rossi et Vila [91, 92]).

Nous traitons les deux situations que nous venons de présenter de manière identique. Nous allons utiliser des outils, estimateur non paramétrique et filtre à noyaux de convolution, pour tenter de construire des règles de décision sous la forme du CUSUM, et plus particulièrement qui convergent vers les statistiques du CUSUM.

Les deux aspects qui viennent d'être évoqués, à savoir, l'estimation non paramétrique d'une partie inconnue d'un système dynamique, et la problématique du filtrage traité par des approches particulières font partie des thématiques du laboratoire d'Analyse des Systèmes et Biométrie depuis déjà quelques années [40, 111, 90].

REMARQUES :

- Une autre particularité des procédés de dépollution est qu'ils sont relativement lents. Sur le jeu de données utilisé dans le chapitre 7, l'intervalle de temps entre deux mesures est de deux minutes. Cette particularité permet de mettre en place des procédures plus coûteuses en temps de calcul que celles utilisées sur des applications de type guidage d'avion où le temps entre deux mesures est beaucoup plus faible.

- Dans les modèles (1) et (2), il n'apparaît pas de variable de contrôle permettant de "guider" le procédé. Pourtant, cette variable existe dans les procédés de dépollution (débit d'alimentation du bioréacteur). Ce choix est délibéré de notre part. Pour étudier les propriétés d'optimalité des approches proposées, il est nécessaire d'effectuer des hypothèses de stabilité sur le modèle. La prise en compte dans le modèle d'une variable de contrôle complexifie alors la notion de stabilité. Un travail supplémentaire serait nécessaire pour étendre les résultats théoriques obtenus, aux modèles contrôlés.

En pratique, les méthodes que nous proposons s'adaptent tout de même directement au cas de modèles contrôlés, comme nous le verrons dans le chapitre 7.

- Nous traitons dans ce document deux particularités des procédés de dépollution. D'une part, le cas où une fonction f d'un modèle autorégressif est inconnue (modèle (1)), et d'autre part, le cas d'observation indirecte (modèle (2)). Il est évident qu'en pratique, les deux modèles peuvent se cumuler et on se retrouve alors avec une fonction inconnue dans un modèle à espace d'état. Nous n'avons pas traité ce problème plus complexe qui pourra faire l'objet d'un travail futur.

*

* *

Ce mémoire se décompose en trois parties, la dernière regroupant les annexes.

La première partie propose un état de l'art des procédures de détection et de diagnostic. Le premier chapitre présente rapidement les quatre grandes approches qui traitent de ce problème : le traitement du signal, l'automatique, l'IA et les statistiques.

Dans le deuxième chapitre, nous rentrons un peu plus en détail sur les méthodes statistiques de détection, notamment sur la construction de la règle du CUSUM, qui sera la règle de référence tout au long du document.

La deuxième partie rassemble l'ensemble des résultats obtenus au cours de cette thèse.

Le chapitre 3 concerne le choix du seuil pour des règles de décision construites sous la même forme que le CUSUM. En effet, les procédures de détection de rupture de modèle ne sont rien d'autre que des tests statistiques (séquentiels) et donc, comme tout test, elles nécessitent le choix d'un seuil qui va réaliser un compromis entre les erreurs de 1^{ière} et 2^{ième} espèces, qui seront respectivement associées en détection d'anomalie, à la fréquence des fausses alarmes et au retard à la détection. Nous avons proposé une méthode calculatoire qui permet d'obtenir un seuil adaptatif garantissant un taux constant de fausses alarmes tout au long du processus pour la règle du CUSUM. Nous avons étendu notre approche au cas du diagnostic.

Généralement, les procédures de détection statistiques utilisent un seuil fixe tout au long de la procédure. Nous verrons que cette façon de procéder n'est pas toujours adaptée, notamment pour les procédés de dépollution, procédés qui sont fortement non stationnaires. Le taux de fausses alarmes sur ce type de procédés n'est alors plus constant dans le temps en utilisant un seuil fixe, ce qui peut être préjudiciable.

Le chapitre 4 concerne des résultats d'optimalité asymptotique. A partir des travaux de Lai [50], nous avons montré qu'une règle de décision construite sous la forme du CUSUM, à partir d'estimations convergentes des vraisemblances conditionnelles, est asymptotiquement optimale. Nous avons obtenu l'optimalité pour deux critères, celui du pire retard moyen à la détection et celui du retard moyen.

Ce résultat nous paraît important puisqu'il assure que lorsque les vraisemblances conditionnelles des observations, sous hypothèse de fonctionnement normal et sous hypothèse de panne, sont inaccessibles, il suffit de pouvoir estimer ces quantités de façon convergente pour construire une règle de décision optimale (appelée CUSUM approché).

Nous appliquons ces résultats aux deux types de problèmes qui nous intéressent. D'une part, celui où une partie du modèle est inconnue (modèle (1)), on parle alors du "CUSUM non paramétrique". Cette règle de décision est présentée en détail dans le chapitre 5. L'estimation de la densité conditionnelle est possible grâce à

Introduction

l'utilisation d'un estimateur à noyaux construit à partir d'un échantillon d'apprentissage regroupant à la fois, des données sous régime de fonctionnement normal et des données sous régime de panne.

D'autre part, les résultats d'optimalité du CUSUM approché sont appliqués dans le cas où le système est indirectement observé (modèle (2)). Ce point est traité dans le chapitre 6. La densité conditionnelle des observations est estimée à partir du filtre à noyaux de convolution (Rossi [90]). Le cas plus complexe où le paramètre caractérisant la panne est inconnu, est traité et une règle de décision, concurrente de la règle du GLR (Lorden [62]) est proposée et son optimalité est obtenue. On parlera dans ce chapitre de procédures de filtrage-détection.

Le chapitre 7 concerne la mise en place des procédures de filtrage-détection et des seuils adaptatifs sur des données réelles provenant du LBE de Narbonne. Le système dynamique considéré modélise le fonctionnement d'un bioréacteur de traitement des eaux usées. Deux types d'anomalies ont été étudiés : d'une part, une panne de capteur sur la concentration en substrat et d'autre part, une anomalie plus artificielle, qui concerne un changement de modèle sur la concentration en biomasse.

La troisième partie regroupe les annexes. L'annexe A est une rapide présentation des estimateurs non paramétriques à noyaux de convolution. L'annexe B présente la problématique du filtrage et les solutions proposées par des méthodes de Monte Carlo. Le filtre à noyaux de convolution est détaillé. Enfin, le procédé de dépollution considéré est exposé en annexe C.

Première partie
Etat de l'art

Chapitre 1

Méthodes de détection et de diagnostic

La diversité des systèmes à surveiller, et parfois l'hétérogénéité des problèmes sur un même système ont entraîné la multiplication des approches de la détection et du diagnostic d'anomalie. Ainsi, on peut citer quatre grandes approches traitant ces sujets : l'automatique, le traitement du signal, l'intelligence artificielle et les statistiques.

Le choix d'une méthode de détection plutôt qu'une autre va tout d'abord dépendre de la forme des données et des informations à notre disposition sur le système. Il y a alors deux cas de figure : soit il n'existe pas de modélisation du système, soit un modèle, qu'il soit quantitatif ou qualitatif permet de décrire le fonctionnement de ce système. Les méthodes issues du traitement du signal traitent généralement du premier cas de figure tandis que les méthodes issues de l'Automatique nécessitent une modélisation du système. On trouve dans les approches relevant de l'IA et de celles relevant des statistiques des méthodes applicables avec modèle et d'autres sans.

Dans cette partie, nous allons plutôt insister sur les méthodes qui se rapprochent le plus des approches statistiques sur lesquelles nous avons travaillé, c'est à dire sur des méthodes qui se basent sur une modélisation du système considéré. En effet, pour des procédés de type retraitement des eaux usées, qui constituent l'application principale de ce travail de thèse, ou plus généralement pour tout ce qui concerne les procédés biotechnologiques, on peut espérer bénéficier d'une modélisation du système.

Généralement, dans la communauté du diagnostic de panne, les procédures de détection basées sur une modélisation du système se traitent en deux étapes, comme on peut le voir sur la figure 1.1 :

Etape 1 : Génération de résidus

Cette première étape consiste à calculer des quantités qui vont refléter la présence ou non d'une panne considérée sur le système. Ces quantités sont appelées résidus. Les résidus seront proches de 0 sous régime de fonctionnement normal, et significativement non nuls sous régime de panne.

Étape 2 : Evaluation des résidus

Les résidus sont examinés et une règle de décision est appliquée (comparaison à un seuil, test statistique, méthode floue,...) pour déterminer si une panne a eu lieu ou non.

Dans la communauté des statisticiens, la terminologie est légèrement différente puisqu'on ne va pas générer des résidus mais plutôt construire des statistiques de test et l'évaluation va consister en un test entre deux hypothèses (pour le problème de détection), ou plus (pour le diagnostic).

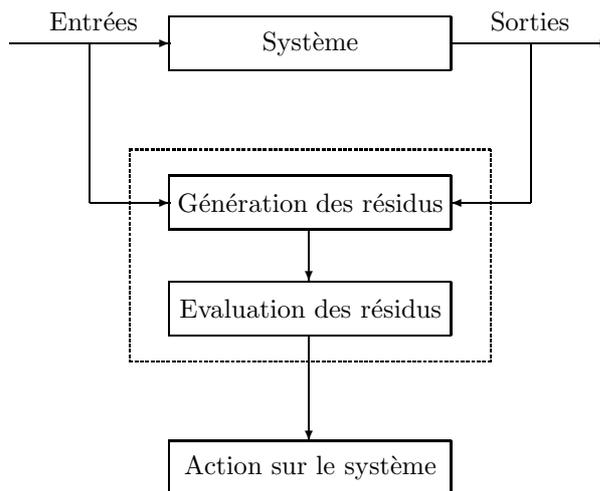


FIG. 1.1 – Structure générale d'une procédure de diagnostic de panne

1.1 Approches sans modèle analytique

L'utilisation de capteurs capables de mesurer des grandeurs en temps réel sur des procédés est le véritable point de départ de la surveillance en ligne. Jusque là, la surveillance se limitait à détecter une défaillance à partir des cinq sens de l'être humain : un bruit, une odeur anormale, un échauffement, etc... Il s'est ainsi rapidement posé la question de la fiabilité des capteurs, dont une réponse peut être apportée par la redondance matérielle.

Redondance matérielle

L'objectif est ici de déterminer si les capteurs présentent des anomalies. On va alors mettre en place plusieurs capteurs qui vont mesurer la même variable. On va ensuite comparer ces valeurs deux à deux. Avec deux capteurs (on parle alors de duplex), on pourra détecter une défaillance. Avec trois capteurs (triplex), on pourra détecter et localiser le capteur défaillant. Les inconvénients de cette méthode sont les coûts élevés et le fait qu'on ne peut détecter que des défaillances de capteurs et non des anomalies sur le système lui-même. De plus, l'augmentation du nombre de capteurs entraîne une augmentation du coût de maintenance. Pour revenir à l'exemple des procédés biotechnologiques, et plus particulièrement à la digestion anaérobie, les pannes de capteurs sont relativement fréquentes compte tenu de leur conditions d'utilisation, et donc plus il y a de capteurs, plus il y a de possibilités de pannes...

La redondance matérielle est donc surtout utilisée sur des systèmes critiques où les coûts de surveillance sont finalement secondaires par rapport à la sécurité matérielle et humaine, comme dans l'industrie nucléaire ou l'aérospatiale.

Comparaison à des seuils

Il s'agit d'une des méthodes les plus utilisées dans l'industrie. Elle consiste à comparer les mesures des variables ou leurs moyennes à des seuils. On peut alors considérer deux types de seuils : un premier niveau indiquant une légère variation et donc une présence probable d'une panne. L'opérateur peut alors intervenir sur le procédé pour réparer le système. Et un deuxième niveau indiquant une panne plus grave et entraînant l'arrêt du système. Le choix des seuils s'effectue généralement grâce à des connaissances expertes et des échantillons d'apprentissage.

Analyse spectrale

En analysant tout d'abord le spectre des signaux en état de fonctionnement normal, on peut détecter une panne sur le système dès qu'il y a un changement au niveau de la fréquence du signal. Ces méthodes sont bien adaptées dans le cas où les fréquences des régimes de panne sont connues.

1.2 Approches basées sur un modèle analytique

Il existe différents types de modèles : le modèle d'observation, qui reproduit le comportement externe du système (entrées-sorties) ; le modèle de connaissance qui s'appuie sur les lois de la physique régissant le système ; et enfin le modèle mixte s'appuyant à la fois sur des lois physiques et sur une représentation entrées-sorties.

Parmi les méthodes issues de l'automatique, on compte trois grandes approches pour générer les résidus : l'estimation paramétrique, l'espace de parité et les observateurs. On peut trouver une large revue de toutes ces méthodes dans les ouvrages

Chapitre 1. Méthodes de détection et de diagnostic

de Chen et Patton [15], Gertler [34] ou Maquin et Ragot [65], ou encore dans les publications de Frank [31] et Isermann [43].

La modélisation considérée ici sera sous un format entrées-sorties (par exemple de type ARMA) ou encore sous forme d'un modèle à espace d'état.

Considérons par exemple le modèle à espace d'état linéaire suivant :

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) + F_1 f(t) \mathbb{I}_{\{t \geq t_p\}} \\ y(t) = Cx(t) + F_2 f(t) \mathbb{I}_{\{t \geq t_p\}}, \end{cases} \quad (1.1)$$

où $x(t)$ est la variable d'état, $u(t)$ la variable de commande, $y(t)$ la variable d'observation et $f(t)$ représente ici une panne additive (t_p est l'instant d'apparition de la panne).

Espace de parité

La méthode d'espace de parité a été introduite par Mironovski [68]. On peut consulter Frank [30] pour un état de l'art.

A partir de relations algébriques issues de la modélisation du système (ces relations sont obtenues par exemple par fonctions de transfert comme la transformée de Laplace), on s'intéresse à la cohérence entre les entrées et les sorties du système. Le fait que des relations de parité ne soient pas vérifiées génère des résidus.

Il existe deux types de relations de parité :

- la redondance analytique directe, ne faisant intervenir que des mesures obtenues au même instant.
- la redondance analytique temporelle, faisant intervenir des mesures issues d'une fenêtre temporelle.

La redondance directe a tout d'abord été mise au point pour remplacer la redondance matérielle. Elle présente l'avantage d'avoir des coûts beaucoup plus faibles que cette dernière. Mais elle souffre du même désavantage, à savoir que des relations mettant en jeu des mesures d'un même instant ne permettent de surveiller que des éventuels défauts de capteurs. C'est pour cela que les relations temporelles ont été mises au point, puisqu'elle permettent, elles, de détecter des anomalies affectant le système lui même.

On peut citer ici les travaux de Domlan, Maquin et Ragot [24] qui utilisent la méthode de l'espace de parité pour traiter le problème des systèmes à commutation. L'objectif est de détecter les changements successifs de modèles lorsque la modélisation du système peut être simplifiée par des modèles locaux. Ce problème, qui n'est pas directement un problème de diagnostic de panne pourrait cependant s'appliquer à la surveillance si on s'intéressait à la fois à la détection d'une panne mais aussi à la détection d'un retour au régime normal lorsqu'on est sous régime de panne. Cela ne sera pas le cas dans ce document, où on suppose qu'après une détection, le système est réparé instantanément.

Observateurs

Les premiers travaux sur le diagnostic à partir d'observateurs sont dus à Clark, Fosth et Walton [16]. Le principe est le même que pour l'espace de parité. Il s'agit d'étudier la cohérence entre les valeurs mesurées et celles qu'on aura estimées. Grâce à un observateur, on va pouvoir estimer l'état x du système et donc obtenir une estimation $\hat{y}(t) = C\hat{x}(t)$ de la variable d'observation. Le résidu sera alors défini comme la différence entre la mesure obtenue par le capteur et l'estimation obtenue à partir de la modélisation du système :

$$e(t) := y(t) - \hat{y}(t) .$$

On peut différencier deux types d'approches : celle où on compare les observations au seul modèle représentant le fonctionnement normal du système. Et d'autre part, une approche où on compare les observations à tout un ensemble de modèles, chacun associé à une certaine situation de panne. On parle alors de "bank of observers".

Lorsque le modèle déterministe précédent est remplacé par un modèle stochastique, on ne parle plus d'observateurs mais de filtres : filtre de Kalman, filtre de Kalman étendu...

On peut noter la différence entre les deux premières approches de génération de résidus présentées : par espace de parité et observateurs. Dans un cas, on cherche à éliminer l'état x du modèle (1.1) grâce à des relations de parité tandis que pour une approche basée sur des observateurs, on estime cet état x par \hat{x} .

Estimation paramétrique

L'estimation paramétrique pour résoudre des problèmes de diagnostic a tout d'abord été étudié par Isermann [43]. On suppose ici qu'un défaut sur le système entraîne une variation d'un ou plusieurs paramètres. L'estimation paramétrique consiste alors à identifier en temps réel les paramètres du système et à les comparer à leurs valeurs de référence lorsque le système est sous régime de fonctionnement normal. Les résidus sont généralement définis comme la différence entre les valeurs connues des paramètres et leurs estimations.

Les trois approches qui viennent d'être présentées sont assez similaires dans leur fonctionnement. On peut même montrer que la méthode d'espace de parité peut être vue comme un cas particulier d'observateur et dans certains cas particuliers, l'estimation paramétrique s'apparente à des relations de parité (voir Frank [31]).

La robustesse

Lorsqu'on est confronté à un problème réel, les choses se compliquent. En effet, il n'est pas réaliste de considérer qu'un modèle est parfaitement ajusté au système à superviser. Une grande partie des travaux sur le diagnostic, ces dernières années,

ont consisté à prendre en compte, pour les méthodes de détection considérées, les erreurs de modélisation, les bruits sur le système et plus généralement toutes les incertitudes sur le système. On parle alors de diagnostic de panne robuste [15], dont l'objectif est de générer des résidus robustes.

On s'intéresse alors à des modèles s'écrivant sous la forme :

$$\begin{cases} \dot{x}(t) = (A + \Delta A)x(t) + (B + \Delta B)u(t) + (F_1 + \Delta F_1)f(t)\mathbb{I}_{\{t \geq t_p\}} \\ y(t) = (C + \Delta C)x(t) + (F_2 + \Delta F_2)f(t)\mathbb{I}_{\{t \geq t_p\}}, \end{cases} \quad (1.2)$$

où ΔA , ΔB , ΔC , ΔF_1 et ΔF_2 représentent l'incertitude sur le modèle, ou plutôt ce qui prend en compte le "mauvais" ajustement du modèle.

Pour traiter ce type de problème, on peut rapidement évoquer la mise au point d'observateurs robustes et de relations de parité robustes. Une troisième méthode consiste à utiliser des seuils adaptatifs lors de l'évaluation des résidus. Dans Ding, Zhang et Frank [23], le seuil est choisi à chaque instant en fonction du pire des cas possibles (au niveau des incertitudes sur le modèle). Ainsi, il n'y a pas de fausse alarme, mais le seuil étant très élevé, certaines anomalies deviennent indétectables. Les auteurs proposent alors d'utiliser l'approche probabiliste pour fixer un taux de fausses alarmes.

Nos travaux sur le seuil adaptatif pour des règles de type CUSUM (chapitre 3) partent donc de la même idée, à savoir contrôler le taux de fausses alarmes sur le système.

La non linéarité

Nous avons écrit les modèles (1.1) et (1.2) sous forme linéaire. En pratique, les problèmes à traiter sont souvent non linéaires. Il existe alors deux façons de procéder. La première se décompose en deux étapes :

- linéarisation autour du point de fonctionnement
- puis génération de résidus avec des méthodes robustes.

Cette méthode marche bien si la linéarisation ne cause pas un trop grand écart entre les modèles linéaires et non linéaires.

La deuxième façon de procéder consiste à appliquer directement des techniques non linéaires :

- observateurs non linéaires [93]
- relations de parité non linéaires [14, 100].

L'inconvénient de cette deuxième approche est qu'elle ne s'applique que pour certaines classes de modèles non linéaires.

D'une façon générale, la non linéarité pose problème pour ce type d'approches.

Conclusion et critique des approches "automatiques"

- Dans les modèles (1.1) et (1.2), la panne est additive. Ces méthodes sont capables de traiter des pannes multiplicatives, mais le problème est alors plus complexe.

- Les méthodes automatiques sont plus évoluées que les méthodes statistiques de détection pour ce qui concerne la robustesse et donc l'incertitude sur le modèle, puisque cet aspect intéresse la communauté automatique depuis déjà de nombreuses années. Malgré ça, l'approche stochastique, qui permet tout de même de prendre en compte certaines incertitudes (comme le bruit sur le système) possède des outils pouvant répondre à ce type de problème. L'utilisation de méthodes non paramétriques pour l'estimation d'une partie inconnue dans le modèle (chapitre 5) en est un exemple parmi d'autres.

- Comme le font remarquer Chen et Patton [15], la plupart des travaux dans le champ du diagnostic basés sur un modèle quantitatif concernent la génération des résidus. Et c'est peut être là que se trouve l'intérêt des méthodes stochastiques, qui s'intéressent réellement à l'évaluation des résidus en prenant en compte la probabilité de fausses alarmes, le retard à la détection, ou d'autres critères. Par exemple, dans Kinnaert et *all* [45], les auteurs génèrent des résidus par redondance puis ils les évaluent grâce à l'algorithme statistique de diagnostic de Nikiforov [74]. L'approche asymptotique locale (voir Basseville [4]) est utilisée par Li, Chan et Zhang [57] qui ont formé leur résidus grâce à un filtre prédictif ou encore dans Zhang, Basseville et Benveniste [116] après avoir utilisé une méthode de génération de résidus par espace de parité.

1.3 Approches relevant de l'intelligence artificielle

Les réseaux de neurones

Lorsqu'on n'a pas accès à une modélisation du système sous surveillance, il est possible d'utiliser les réseaux de neurones pour obtenir des relations entre les variables. A partir d'un échantillon d'apprentissage (supervisé ou non supervisé), il est possible d'effectuer des prédictions. On peut alors les utiliser à des fins de diagnostic en générant des résidus à partir de la comparaison entre les prédictions et les observations. On parle alors d'observateur neuronal. Les réseaux de neurones sont adaptés à des problèmes non linéaires et sont utilisables en ligne, d'où leur attractivité dans de nombreux problèmes.

Les systèmes d'inférences flous

La théorie des ensembles flous de Zadeh [115] est à la base des systèmes d'inférences flous qui trouvent de nombreuses applications dans la modélisation, la commande et la supervision de procédés. Ces méthodes sont basées sur des connaissances expertes. Un système d'inférence flou est constitué de :

- un univers de discours qui contient les fonctions d'appartenance des variables d'entrées et de sorties à des classes.

- une base de connaissances contenant les règles liant les variables d'entrées et de sorties sous la forme "SI...ALORS...".

- un mécanisme de raisonnement qui base son fonctionnement sur la logique du *modus ponens* généralisé.

Ces méthodes permettent de prendre en compte les notions d'incertitude et d'imprécision et se révèlent donc souvent plus efficaces que les méthodes robustes issues de l'automatique.

Dans Lardon, Punal et Steyer [55], les auteurs utilisent la théorie de l'évidence pour la surveillance d'un procédé de retraitement des eaux usés par digestion anaérobie. Dans Lardon [54], il s'agit de détecter les pannes du réseau de capteur du même digesteur anaérobie.

D'une manière générale, les méthodes floues sont beaucoup utilisées pour la deuxième étape de la procédure de détection : l'évaluation des résidus, et plus particulièrement pour la localisation de la panne. Comme nous venons de le voir, généralement, un résidu s'obtient en effectuant la différence entre une mesure et une estimation à partir du modèle. Lorsque plusieurs résidus, chacun reliés à un certain nombre de variables, sont significativement non nuls, il peut être intéressant d'utiliser des méthodes floues qui vont prendre en compte des connaissances expertes pour pondérer ces résidus (voir par exemple Ragot et Maquin [85]).

1.4 Approches statistiques

Parmi les approches statistiques de détection, on peut dégager deux grandes familles :

Approches non paramétriques

Pour cette première approche, on considère qu'il n'y a pas de modèle analytique qui décrit le système. Les méthodes de détection doivent donc uniquement se baser sur les données à disposition. Généralement, le but est de détecter un changement de moyenne, de médiane ou de variance dans la suite d'observations. Brodsky et Darkhovsky [12] proposent une large revue de ces méthodes pour des données indépendantes ou non, ou encore non stationnaires. Dans certains cas, ces méthodes sont des adaptations de méthodes paramétriques. Ainsi, si on considère une série aléatoire (X_n) de moyenne négative avant le temps de panne, et positive après, on parlera de CUSUM non paramétrique pour la procédure de détection dont le temps d'arrêt est donné par :

$$C_{NP} = \inf\{n : g_n > h\}, \quad \text{avec } g(n) = \max(g(n-1) + X_n, 0), \quad g(0) = 0, \quad (1.3)$$

et où h est le seuil du test. Ce test s'inspire de l'écriture récursive de la règle du CUSUM que nous allons présenter dans les parties suivantes.

Ce type de méthodes est particulièrement adapté à la finance et à l'économie, où il est généralement impossible de disposer d'une modélisation du système. C'est vers ces applications que se dirigent notamment les travaux de Steland [101], et Pawlak et al [78].

Approches paramétriques

On suppose maintenant que le système sous surveillance peut être modélisé sous forme analytique. Les méthodes de détection d'anomalie sont construites à partir de la théorie des tests d'hypothèses, la plupart grâce à une quantité bien connue en statistique mathématique, le rapport de vraisemblance. Les tests d'hypothèses mis en place seront des tests séquentiels pour prendre en compte la dynamique du système.

Généralement, une règle de décision statistique consiste à calculer une statistique de test, vérifiant de bonnes propriétés, à savoir par exemple, qu'elle est en moyenne, négative sous régime de fonctionnement normal, et positive et croissante sous régime de panne. Une alarme est alors déclenchée lorsque la statistique de test franchit un seuil fixé par l'expérimentateur.

La plus simple des modélisations pour un système est de supposer que les observations sont toutes indépendantes et qu'elles suivent toutes une même loi, par exemple une loi normale avec un changement de moyenne à l'instant de panne. C'est dans ce contexte qu'ont été mises au point les toutes premières règles de décision par Shewhart [96] et surtout par Page et sa règle du CUSUM [76].

Bien que cette modélisation du système soit simple (Basseville et Nikiforov [6] parlent de problème de détection basique), dans de nombreux cas incluant les pannes additives sur un modèle de type ARMAX avec bruit blanc gaussien, le problème de détection revient à détecter un changement de moyenne dans une suite d'observations suivant une loi gaussienne. Dans des cas plus généraux, on peut obtenir asymptotiquement le même résultat (approche asymptotique locale [4]).

Les règles de décision du type CUSUM sont adaptables à des systèmes dynamiques (et donc des observations non indépendantes) et le principe de base est de détecter un changement de paramètre dans le modèle, où plus exactement un changement de paramètre dans la densité conditionnelle des observations. Le paramètre θ caractérisant la panne vaut θ_0 avant la panne et θ_1 après. Dans le cas où le paramètre θ_1 est inconnu mais appartient à un intervalle Θ_1 qui lui, est connu, plusieurs procédures sont alors utilisables comme par exemple l'algorithme GLR [62] qui cherche à maximiser la vraisemblance sur Θ_1 , ou encore l'algorithme MLR (de l'anglais "mixture likelihood ratio"), où on utilise une distribution a priori du paramètre θ_1 sur Θ_1 . Ce cas de figure où le paramètre caractérisant la panne est inconnu est évidemment le plus réaliste en pratique puisqu'il est difficile de prévoir "l'amplitude" d'une anomalie.

Chapitre 1. Méthodes de détection et de diagnostic

De façon plus générale, une panne peut être caractérisée par un changement de modèle entre le régime normal et le régime de panne et non plus seulement un changement de paramètre. On parle ainsi de détection de rupture de modèle. Cette façon d'aborder le problème est plus générale que les cas de pannes additives ou multiplicatives puisqu'ici, on s'autorise un changement total de modélisation.

Ces dernières années, les recherches ont porté sur plusieurs domaines et le premier concerne le diagnostic d'anomalie. La première règle de décision traitant conjointement le problème de détection et de localisation est celle de Nikiforov [72]. Il s'agit d'un algorithme généralisant le test CUSUM au diagnostic. Par la suite, on peut citer les travaux de Lai [51] (données dépendantes), Nikiforov [73, 74] ou encore Oskiper et Poor [75] (algorithmes récursifs). Là encore, toutes ces règles de décision s'appliquent dans le cas où on possède une modélisation des différentes anomalies possibles.

Une autre orientation de recherche concerne la mise au point de règles de décision pour des systèmes de plus en plus complexes comme ceux modélisés par un modèle à espace d'état. Il s'agit pour nous d'un cas très important puisque les procédés biotechnologiques sont généralement représentés sous cette forme. Nous l'avons vu dans les parties précédentes, les méthodes relevant de l'Automatique traitent ce genre de problèmes depuis déjà de nombreuses années, les méthodes statistiques ont donc un retard à rattraper.

Pour les modèles à espace d'état linéaires avec bruit blanc gaussien additif, la règle de Willsky et Jones [112] est bien connue : l'utilisation du filtre de Kalman permet de transformer le problème de détection d'une panne additive sur le système en un changement de moyenne sur la suite des innovations, qui suivent une loi normale.

Pour des modèles à espace d'état non linéaires, de récentes règles de décision ont été proposées [1, 2, 58, 59, 117]. Ces méthodes utilisent la théorie du filtrage particulière qui est bien plus efficace qu'un filtre de Kalman étendu pour traiter les non linéarités.

Conclusion sur les approches paramétriques

- Les méthodes que nous allons étudier nécessitent la connaissance de la modélisation sous régime de panne et ne permettent donc pas de détecter des pannes jamais rencontrées.

- Comme nous l'avons noté dans une partie précédente, il n'existe que peu de travaux sur la robustesse des algorithmes comme c'est le cas pour les méthodes automatiques. On peut cependant citer les travaux de Basseville et Nikiforov [7] et Fouladirad et Nikiforov [29] concernant les méthodes statistiques de détection avec paramètres de nuisance dans le modèle.

- Dans ce document, les modélisations linéaires et non linéaires sont traitées indifféremment d'un point de vue mise en place de l'algorithme, que ce soit pour des modèles autorégressifs de type entrées-sorties ou pour des modèles à espace d'état. De même, les résultats obtenus montrent que les règles de décision sont aussi efficaces pour un modèle linéaire que pour un modèle fortement non linéaire.

- Le principal avantage des méthodes statistiques se situe dans le fait qu'il est possible de vérifier des résultats théoriques sur les règles proposées. En effet, il est possible d'obtenir l'optimalité d'une procédure. C'est ainsi que Moustakides [70] montre que parmi tous les algorithmes ayant un temps moyen entre deux fausses alarmes fixé, la règle du CUSUM minimise le pire retard moyen à la détection. Depuis une vingtaine d'années, on voit apparaître une "course à l'optimalité" au travers des résultats entre autres de Bansal et Papantoni-Kazakos [3], Lai [50], Fuh [32], Tartakovsky [107] qui établissent l'optimalité de règles de décision pour des systèmes de plus en plus complexes et des critères d'optimalité mieux adaptés.

Nous présentons un peu plus en détails ces méthodes statistiques de détection dans le chapitre suivant.

*
* *

Les différentes méthodes évoquées ici ne constituent évidemment pas une liste exhaustive des méthodes de détection et de diagnostic. Il s'agit plutôt d'une présentation des outils de base.

La figure qui suit (Basseville et Cordier [5]) permet de comprendre l'organisation générale d'une procédure de surveillance en ligne. On observe quatre niveaux de traitement :

- 1- Le premier niveau concerne le traitement des données capteurs et aussi la surveillance des capteurs.
- 2- Le second niveau concerne la génération d'alarme à partir des observations.
- 3- Le troisième niveau concerne l'interprétation des alarmes (localisation, ...).
- 4- Le dernier niveau concerne le système d'aide à la conduite à proprement parler.

Cette figure montre bien la complexité d'un système de surveillance, qui fait intervenir les quatre principales approches évoquées au début du chapitre, suivant le type de problème rencontré.

D'une façon générale, le choix d'une méthode plutôt qu'une autre sera intimement lié aux caractéristiques du système (type de données, budget, ...) ainsi qu'aux contraintes fixées par l'expérimentateur (retard à la détection inférieur à un certain laps de temps ou utilisation d'un critère plus souple comme le temps moyen entre deux fausses alarmes supérieur à une constante).

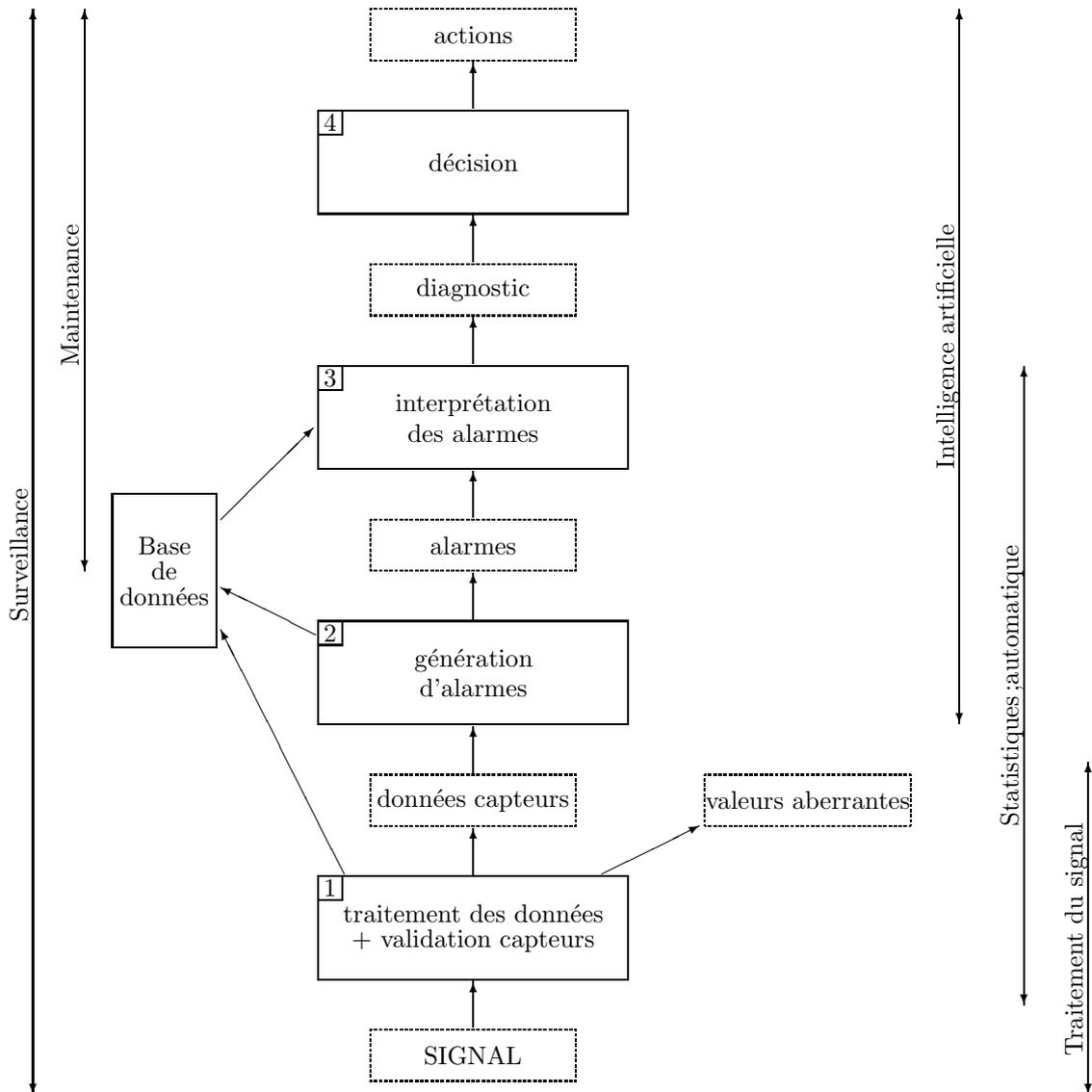


FIG. 1.2 – Organisation générale des traitements en-ligne. On distingue quatre niveaux de traitement permettant de passer des signaux issus des capteurs à des informations symboliques destinées à l'opérateur

Chapitre 2

Méthodes statistiques de détection

Dans cette partie, nous rentrons un peu plus en détails dans la présentation et la construction des méthodes statistiques de détection et de diagnostic, basées sur une modélisation analytique du système. L'objectif de cette partie est notamment de présenter une règle de décision très connue, la règle du CUSUM, qui a constitué le point de départ de ce travail de thèse.

2.1 Introduction

L'objectif de la détection d'anomalie est de repérer, à partir des mesures à notre disposition, le passage pour le système sous surveillance d'un état de fonctionnement normal, appelé régime H_0 , à un état de fonctionnement anormal (ou panne), appelé régime H_1 . La panne est caractérisée par un changement de paramètre dans le modèle. Avant l'instant de panne, le paramètre noté θ vaut θ_0 . Après l'instant de panne t_p inconnu, il prend la valeur θ_1 . Il existe alors deux façons d'aborder le problème :

- **La détection hors-ligne** : A partir d'un échantillon d'observations de taille N , le problème est alors de décider s'il y a eu ou non, un ou plusieurs changements de paramètre entre les instants 1 et N , et d'estimer ces temps de changement. Dans le cas d'un unique changement de valeur, on peut citer les travaux de Picard [79] pour détecter un changement dans la moyenne d'un processus autorégressif. Dans le cas de changements multiples, Lavielle [56] s'intéresse à la fois au cas où le nombre de changements est connu et aussi au cas où il est inconnu.

- **La détection en-ligne** : Les observations arrivent en continu et on doit détecter le plus rapidement possible un changement de paramètre, s'il a lieu. On va donc construire une règle de décision qui, à chaque nouvelle observation nous parvenant, tranchera entre les hypothèses H_0 et H_1 . Si l'hypothèse H_1 est retenue, une alarme sera déclenchée au pas de temps noté t_a . Si l'hypothèse H_0 est retenue,

Chapitre 2. Méthodes statistiques de détection

alors la surveillance continue. Si au temps d'alarme t_a , le processus est en fait encore sous le mode H_0 , on parlera de fausse alarme. Si l'alarme déclenchée au temps t_a n'est pas une fausse alarme, la quantité $t_a - t_p$ représente le retard à la détection.

La détection en-ligne est évidemment le problème le plus compliqué puisqu'on doit prendre une décision en ayant le moins d'observations possible sous H_1 , pour avoir un faible retard à la détection. D'autre part, il y aura une contrainte de plus par rapport au problème hors-ligne. Il s'agit de la fréquence d'acquisition des données, qui, si elle est trop élevée, demande à réaliser des calculs rapides. Ce n'est pas le cas des procédés biotechnologiques, où généralement, on peut compter plusieurs minutes entre deux mesures ce qui permet d'effectuer une grande quantité de calculs. On peut cependant rencontrer ce type de problème dans les systèmes de guidage de missile, ou d'avion (voir par exemple Fouladirad et Nikiforov [29]).

Si la détection hors-ligne est finalement assez proche des tests d'hypothèses, puisqu'on peut parler d'erreur de première espèce (détecter une panne alors qu'il n'y en a pas eu) et de puissance (probabilité de détecter une panne lorsqu'elle a bien eu lieu), c'est assez différent pour la détection en-ligne puisqu'il est difficile d'utiliser de tels concepts. On parlera plutôt de temps moyen entre deux fausses alarmes et de retard à la détection. La qualité d'une règle de décision sera donc généralement jugée sur sa capacité à détecter le plus rapidement possible une anomalie pour un temps moyen entre deux fausses alarmes fixé.

On trouve dans Basseville et Nikiforov [6] et Lai [49, 52] une large revue des méthodes de détection en-ligne que nous allons présenter maintenant. Dans ce travail de thèse, on ne s'intéresse pas à l'instant précis d'apparition de la panne. On peut tout de même citer les travaux de Wu [113] sur l'estimation du temps de panne t_p .

REMARQUE : Tout au long de ce document, nous avons pris le parti, pour des raisons de lisibilité, de caractériser une panne par un changement de paramètre (qui passe de la valeur θ_0 à θ_1) dans la loi des observations plutôt que par un changement total de la loi des observations caractérisée par deux densités : p sous H_0 et q sous H_1 . La deuxième modélisation peut être traitée comme un changement de paramètre en écrivant la densité des observations de la façon suivante :

$$(1 - \theta)p + \theta q,$$

avec sous H_0 , $\theta = \theta_0 = 0$, et sous H_1 , $\theta = \theta_1 = 1$.

Toutes les règles de décision proposées par la suite sont donc applicables à des problèmes de détection de changement de modèle (comme on le verra dans le chapitre 7) et non seulement à des changements de paramètres.

2.2 De la règle de Shewart au CUSUM

2.2.1 Le rapport de vraisemblance

Considérons à présent une suite (X_n) de variables aléatoires indépendantes ayant pour densité de probabilité $p_\theta(\cdot)$, supposée connue, dépendant du paramètre θ . Avant l'instant de panne t_p , θ vaut θ_0 et après l'instant de panne, le paramètre prend la valeur $\theta_1 \neq \theta_0$. Nous allons tout d'abord supposer que les valeurs θ_0 et θ_1 sont connues.

Une grande partie des règles de détection de changement sont basées sur un objet important en statistique mathématique, à savoir le logarithme du rapport de vraisemblance :

$$s(X) = \log \frac{p_{\theta_1}(X)}{p_{\theta_0}(X)}.$$

La propriété de ce rapport est la suivante : si on note respectivement E_{θ_0} et E_{θ_1} les espérances sous les distributions p_{θ_0} et p_{θ_1} , alors on a :

$$E_{\theta_0}(s) < 0 \quad \text{et} \quad E_{\theta_1}(s) > 0. \quad (2.1)$$

Ainsi, un changement sur le paramètre θ entraîne un changement de signe pour l'espérance du logarithme du rapport de vraisemblance. C'est à partir de cette propriété que bon nombres d'algorithmes sont construits.

2.2.2 Les règles de décision élémentaires

La règle de Shewart (1931)

La règle de Shewart [96] est une des toutes premières méthodes de détection de changement en ligne. Son principe est le suivant : on considère un échantillon de taille N fixée (les N premières valeurs observées). On effectue sur cet échantillon un test du rapport de vraisemblance entre les deux hypothèses : $H_0 : \theta = \theta_0$ et $H_1 : \theta = \theta_1$. Si l'hypothèse H_1 est retenue, une alarme est déclenchée. Sinon, on reconsidère un échantillon de taille N (donc les valeurs des instants $N + 1$ à $2N$ et ainsi de suite...) jusqu'à ce que la décision soit en faveur de H_1 .

Plus précisément, si on note :

$$S_j^k = \sum_{i=j}^k s_i \quad \text{où} \quad s_i = \log \frac{p_{\theta_1}(X_i)}{p_{\theta_0}(X_i)},$$

le logarithme du rapport de vraisemblance des observations X_j, \dots, X_k , alors pour un échantillon de taille N , la règle de décision du test du rapport de vraisemblance est la suivante :

$$d = \begin{cases} 0 & \text{si } S_1^N < h, \text{ on choisit } H_0 \\ 1 & \text{si } S_1^N \geq h, \text{ on choisit } H_1, \end{cases}$$

Chapitre 2. Méthodes statistiques de détection

où h est le seuil fixé par l'expérimentateur. La règle de décision de Shewart s'écrit donc :

$$t_a = N. \min\{K : S_{N(K-1)+1}^{NK} \geq h, K = 1, 2, \dots\}. \quad (2.2)$$

Les règles de la moyenne mobile

Les règles de la moyenne mobile utilisent la propriété fondamentale du rapport de vraisemblance (2.1) combinée avec un effet de mémoire des anciennes observations. Ainsi, un poids important va être attribué aux plus récentes valeurs contre des poids plus faibles pour les plus anciennes. Lorsque la mémoire sera finie (de taille N), on parlera de moyenne mobile finie (FMA), règle de décision introduite par Page [76], dont la statistique de test s'écrit :

$$g_n = \sum_{i=0}^{N-1} \gamma_i s_{n-i}$$

où γ_i est le poids associé à la $(n - i)$ -ième observations.

Roberts [87] utilise une mémoire infinie, la règle de la moyenne mobile géométrique (GMA). La statistique de test s'écrit alors :

$$g_n = \sum_{i=0}^{\infty} \gamma_i s_{n-i} \text{ avec } \gamma_i = \alpha(1 - \alpha)^i, 0 < \alpha \leq 1,$$

où $s_i = 0$ si $j < 0$.

Dans les deux précédents algorithmes, le temps d'arrêt est défini de la façon suivante :

$$t_a = \min\{n : g_n \geq h\}$$

où h est le seuil du test.

Ces algorithmes élémentaires sont beaucoup utilisés en pratique mais ne possèdent pas de propriété d'optimalité. Or, il est important de pouvoir montrer l'optimalité d'une approche, ne serait-ce que pour donner de la légitimité à cette dernière. Une des toutes premières règles possédant des propriétés d'optimalité est le CUSUM (ou règle des sommes cumulées).

2.3 La règle du CUSUM

La règle du CUSUM a été introduite par Page [76] en 1954, pour détecter un changement de paramètre dans une suite d'observations indépendantes. Des résultats d'optimalité pour cette règle ont été obtenus par Lorden [62], Moustakides [70] et

Ritov [86]. C'est encore aujourd'hui un algorithme très utilisé sous sa forme initiale, c'est à dire sous hypothèse d'indépendance des observations, même pour des systèmes dynamiques, en contrôle de qualité ou autres. Pourtant, depuis déjà de nombreuses années, la règle du CUSUM a été adaptée à des situations plus complexes : données dépendantes, chaînes de Markov ou encore chaînes de Markov cachées. Des résultats d'optimalité ont même parfois été obtenus. Comme le font remarquer Stoumbos et *al* [105], il y a un véritable fossé entre les outils utilisés par l'expérimentateur et les tous les travaux de recherche effectués depuis une cinquantaine d'années.

Dans cette partie, nous revenons sur la construction du CUSUM, évoquons quelques résultats d'optimalité importants, puis nous présentons des adaptations de ce test à des situations plus complexes.

2.3.1 Données indépendantes

Rappelons la modélisation du système : nous considérons une suite d'observations (X_n) indépendantes de densité p_θ telles que :

$$\begin{aligned} X_1, X_2, \dots, X_{t_p-1} & \text{ sont des variables aléatoires indépendantes de densité } p_{\theta_0} \\ X_{t_p}, X_{t_p+1}, \dots & \text{ sont des variables aléatoires indépendantes de densité } p_{\theta_1} \end{aligned}$$

et on note $P^{(t_p)}$ la mesure de probabilité pour une telle distribution et $E^{(t_p)}$ l'espérance associée à cette mesure. Lorsque t_p vaut l'infini, c'est à dire lorsqu'il n'y a pas de panne, on note P_{θ_0} et E_{θ_0} la mesure de probabilité et l'espérance correspondante. La densité p_θ est connue et ne dépend que du paramètre θ . Les valeurs θ_0 et θ_1 sont supposées connues.

Construction du test :

A l'instant n , on cherche à savoir s'il y a eu un changement de paramètre à un instant j compris entre 1 et n . L'idée de la règle du CUSUM est donc de construire, pour toutes les valeurs j comprises entre 1 et n , un rapport de vraisemblance entre les deux hypothèses :

$$\begin{aligned} & \theta = \theta_0 \text{ pour } i = 1, \dots, n \\ \text{contre } & \theta = \theta_0 \text{ pour } i < j \text{ et } \theta = \theta_1 \text{ pour } i \geq j. \end{aligned}$$

Le logarithme du rapport de vraisemblance entre ces deux hypothèses s'écrit

alors :

$$\begin{aligned} \log \frac{p_{\theta_0, \theta_1, j}(X_1, \dots, X_n)}{p_{\theta_0}(X_1, \dots, X_n)} &= \log \frac{\prod_{i=1}^{j-1} p_{\theta_0}(X_i) \cdot \prod_{i=j}^n p_{\theta_1}(X_i)}{\prod_{i=1}^n p_{\theta_0}(X_i)} \\ &= \log \frac{\prod_{i=j}^n p_{\theta_1}(X_i)}{\prod_{i=j}^n p_{\theta_0}(X_i)} \\ &= \sum_{i=j}^n \log \frac{p_{\theta_1}(X_i)}{p_{\theta_0}(X_i)}. \end{aligned}$$

On effectue ce rapport pour $j = 1, \dots, n$, et on obtient l'écriture de la statistique de test g_n de la règle du CUSUM, qui est la maximisation de ces rapports sur l'instant supposé de changement de régime j :

$$g_n = \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{p_{\theta_1}(X_i)}{p_{\theta_0}(X_i)}. \quad (2.3)$$

La statistique de test du CUSUM vérifie les propriétés suivantes : les incréments $\log(p_{\theta_1}(X_i)/p_{\theta_0}(X_i))$ sont en moyenne négatifs sous le régime H_0 et donc la statistique de test a tendance à être négative (ou du moins proche de 0) et sous le régime H_1 , g_n est en moyenne positive et croissante. Le temps d'arrêt est ainsi défini de la façon suivante :

$$t_{CUSUM} = \inf\{n : g_n \geq h\}. \quad (2.4)$$

C'est le premier instant où la statistique de test g_n franchit le seuil h fixé au préalable par l'expérimentateur. La figure 2.1 représente une application de la règle du CUSUM sur un modèle gaussien.

REMARQUES :

- On peut utiliser de façon équivalente le CUSUM sous forme récursive,

$$g_n = \left(g_{n-1} + \log \frac{p_{\theta_1}(X_n)}{p_{\theta_0}(X_n)} \right)^+$$

avec la notation $m^+ = m$ si $m > 0$ et 0 sinon. Cette écriture diminue fortement le nombre de calculs ce qui constitue un des grands avantages du test CUSUM. Il s'agit d'une écriture équivalente et non égale puisque la statistique de test sous sa forme récursive ne prend que des valeurs supérieures ou égales à 0. Les deux statistiques de test sont en fait égales dès lors qu'elles sont positives.

- On voit ici apparaître la ressemblance entre la règle du CUSUM écrite sous sa forme récursive et le CUSUM non paramétrique évoqué dans le chapitre 1 (règle (1.3)).

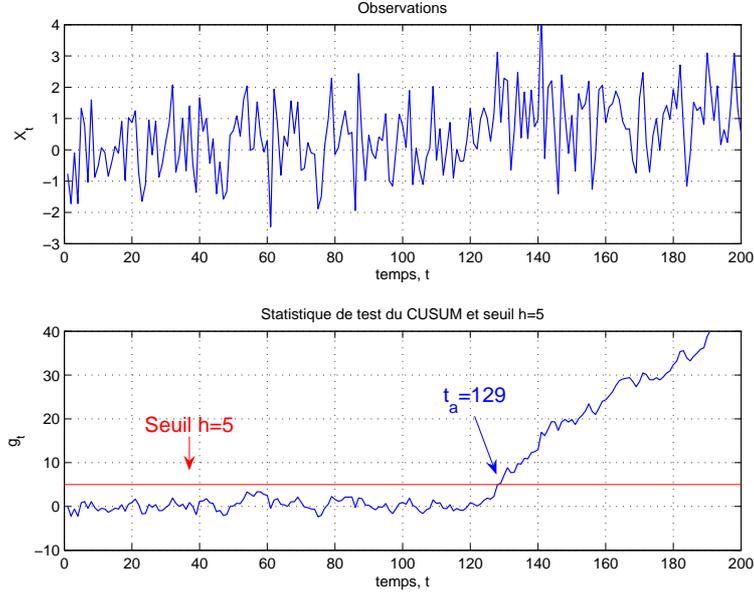


FIG. 2.1 – Mise en place du test CUSUM sur des observations simulées suivant une loi normale $N(0;1)$ avant l’instant de panne $t_p = 120$, et suivant une loi $N(1;1)$ après l’instant de panne. Si le seuil vaut $h = 5$, la détection a lieu à l’instant $t_{CUSUM} = 129$.

2.3.2 Données dépendantes

Le fait de considérer des observations indépendantes est assez restrictif en pratique. La plupart des systèmes à surveiller sont des systèmes dynamiques, c’est à dire que les variables qui interviennent sont liées par des relations temporelles. L’approche statistique classique consiste alors à modéliser une panne ou une anomalie sur un procédé par un changement de paramètre dans la densité conditionnelle des observations. Ainsi, sous P_{θ_0} , on a :

$$X_n | X_1, \dots, X_{n-1} \sim p_{\theta_0}(\cdot | X_1, \dots, X_{n-1}) \quad \forall n \geq 1$$

et sous $P^{(t_p)}$:

$$\begin{aligned} X_n | X_1, \dots, X_{n-1} &\sim p_{\theta_0}(\cdot | X_1, \dots, X_{n-1}) & \forall n < t_p \\ X_n | X_1, \dots, X_{n-1} &\sim p_{\theta_1}(\cdot | X_1, \dots, X_{n-1}) & \forall n \geq t_p . \end{aligned}$$

La notation précédente est assez floue notamment au niveau de la densité conditionnelle sous H_1 , $p_{\theta_1}(\cdot | X_1, \dots, X_{n-1})$. On peut en effet considérer plusieurs façons de générer le changement (voir Basseville et Nikiforov [6]). Suivant le type de modèle qu’on va traiter, il y aura des hypothèses supplémentaires à considérer. Nous évoquons ici trois des façons de générer un changement les plus utilisées : deux

Chapitre 2. Méthodes statistiques de détection

approches “dépendantes” et une troisième approche “indépendante”. Evidemment, on ne choisit pas la façon dont va être généré un changement. Ce choix est dicté par la nature du modèle considéré. Pour les trois façons présentées ci-dessous, des exemples de modélisation de système sont évoqués.

Première façon : Il s’agit de la génération de changement la plus générale que nous allons utiliser dans cette thèse. Les observations sont générées sous le régime H_0 suivant une loi conditionnelle qui admet une densité de probabilité $p_{\theta_0}(X_n|X_{1:n-1})$. Pour tout n supérieur à l’instant de panne inconnu t_p , la valeur θ_0 est remplacée par la valeur θ_1 et le passé, $X_{1:n-1}$, est utilisé comme condition initiale des observations. De plus, la densité conditionnelle $p_{\theta_1}(X_n|X_{1:n-1})$ dépend de la valeur t_p et ainsi, doit donc s’écrire de façon rigoureuse, $p_{\theta_0, \theta_1, t_p}(X_n|X_{1:n-1})$. L’indice “ θ_0, θ_1, t_p ” doit s’interpréter de la façon suivante : à l’instant t_p , θ passe de la valeur θ_0 à θ_1 . La densité de tout l’échantillon s’écrit alors :

$$p(X_{1:n}|n \geq t_p > 1) = p_{\theta_0}(X_1) \left[\prod_{i=2}^{t_p-1} p_{\theta_0}(X_i|X_{1:i-1}) \right] \left[\prod_{i=t_p}^n p_{\theta_0, \theta_1, t_p}(X_i|X_{1:i-1}) \right].$$

EXEMPLE : Nous rencontrons ce type de modélisation notamment sur les modèles à espace d’état. En effet, pour le modèle suivant, que nous utiliserons souvent en simulations,

$$\begin{cases} X_n = aX_{n-1} + b\frac{X_{n-1}}{1+X_{n-1}^2} + c \cdot \cos(1.2 \cdot n) + v_{n-1} \\ Y_n = dX_n^2 + w_n, \end{cases} \quad (2.5)$$

avec $\theta = (a, b, c, d)$, la densité conditionnelle des observations $p_{\theta_0, \theta_1, t_p}(Y_n|Y_{1:n-1})$ dépend de l’instant de changement de régime.

Deuxième façon : La deuxième façon est un cas particulier de la première. Le changement va être généré de la même façon que dans la première méthode mais cette fois-ci, la densité conditionnelle $p_{\theta_1}(X_n|X_{1:n-1})$ ne va pas dépendre de l’instant de changement de régime. La densité de tout l’échantillon $X_{1:n}$ sous l’hypothèse $n \geq t_p$ s’écrit alors :

$$p(X_{1:n}|n \geq t_p > 1) = p_{\theta_0}(X_1) \left[\prod_{i=2}^{t_p-1} p_{\theta_0}(X_i|X_{1:i-1}) \right] \left[\prod_{i=t_p}^n p_{\theta_1}(X_i|X_{1:i-1}) \right] \quad (2.6)$$

EXEMPLE : Il s’agit de la génération de changement la plus rencontrée dans la littérature. Un changement de paramètre dans un modèle autorégressif de la forme :

$$X_n = f(X_{n-1}, \theta) + \epsilon_n$$

est l'exemple basique de cette modélisation du changement. Plus précisément, considérons le modèle suivant :

$$X_n = \theta X_{n-1} + \epsilon_n, \quad \epsilon_n \sim N(0, 1),$$

alors on a :

- sous H_0 , $X_n|X_{1:n-1}$ a pour densité $p_{\theta_0}(x_n|X_{1:n-1}) = 1/\sqrt{2\pi} \exp(-(x_n - \theta_0 X_{n-1})^2/2)$
 - sous H_1 , $X_n|X_{1:n-1}$ a pour densité $p_{\theta_1}(x_n|X_{1:n-1}) = 1/\sqrt{2\pi} \exp(-(x_n - \theta_1 X_{n-1})^2/2)$,
- qui ne dépend pas de l'instant de changement.

Troisième façon : La troisième façon de générer un changement consiste à supposer qu'il y a indépendance entre les observations sous le régime H_0 et les observations sous le régime H_1 . La densité de $X_{1:n}$ sous l'hypothèse $n \geq t_p$ s'écrit :

$$p(X_{1:n}|n \geq t_p > 1) = p_{\theta_0}(X_1) \left[\prod_{i=2}^{t_p-1} p_{\theta_0}(X_i|X_{1:i-1}) \right] p_{\theta_1}(X_{t_p}) \left[\prod_{i=t_p+1}^n p_{\theta_1}(X_i|X_{t_p:i-1}) \right] \quad (2.7)$$

EXEMPLE : Il s'agit d'un cas particulier de la deuxième méthode où, dès qu'on passe sous régime de panne, on oublie tout le passé. Si on reprend le modèle autorégressif précédent, cela revient à considérer que :

$$\forall n < t_p \quad X_n = f(X_{n-1}, \theta_0) + \epsilon_n, \quad \forall n \geq t_p + 1, \quad X_n = f(X_{n-1}, \theta_1) + \epsilon_n,$$

et pour $n = t_p$, $X_n = f(X^0, \theta_1) + \epsilon_n$, où X^0 est la condition initiale du régime de panne.

REMARQUE : Dans ce document, on se placera généralement dans un des deux premiers cas suivant le chapitre concerné. Pour la partie sur le CUSUM non paramétrique (chapitre 5), on pourra se contenter de la deuxième méthode de génération puisque nous allons traiter des modèles autorégressifs simples. Pour tout ce qui concerne la partie filtrage (chapitre 6), on devra se placer dans le premier cas.

La généralisation naturelle de la règle du CUSUM à des données dépendantes consiste alors à remplacer $p_{\theta_d}(X_i)$, pour $d = 0, 1$, dans (2.3) par la densité conditionnelle des observations sous les hypothèses H_0 et H_1 . La statistique de test du CUSUM généralisé s'écrira alors :

$$g_n = \max_{1 \leq j \leq n} \sum_{i=j}^n Z_{i,j}, \quad \text{avec} \quad Z_{i,j} = \log \frac{p_{\theta_0, \theta_1, j}(X_i|X_{1:i-1})}{p_{\theta_0}(X_i|X_{1:i-1})}. \quad (2.8)$$

REMARQUE : Pour la deuxième méthode de génération, g_n s'écrit de la même façon en remplaçant dans (2.8), $p_{\theta_0, \theta_1, j}(X_i|X_{1:i-1})$ par $p_{\theta_1}(X_i|X_{1:i-1})$. De plus la statistique

Chapitre 2. Méthodes statistiques de détection

de test pourra encore s'écrire de façon équivalente sous forme récursive. Ce n'est pas le cas pour la première méthode de génération de changement.

Propriété 2.1 *La propriété (2.1) est toujours valable pour les incréments $Z_{i,j}$ calculés grâce aux densités conditionnelles.*

Démonstration : La fonction logarithme étant strictement concave, l'inégalité de Jensen entraîne : $\forall j, j \leq i$,

$$\begin{aligned}
 E_{\theta_0} \left[\log \frac{p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \right] &< \log \left(E_{\theta_0} \left[\frac{p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \right] \right) \\
 &< \log \left(E_{\theta_0} \left[\frac{p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1}) p_{\theta_0}(X_{1:i-1})}{p_{\theta_0}(X_{1:i})} \right] \right) \\
 &< \log \left(\int p_{\theta_0}(x_{1:i}) \frac{p_{\theta_0, \theta_{1,j}}(x_i | x_{1:i-1}) p_{\theta_0}(x_{1:i-1})}{p_{\theta_0}(x_{1:i})} dx_1 \dots dx_i \right) \\
 &< \log \left(\int \left(\int p_{\theta_0, \theta_{1,j}}(x_i | x_{1:i-1}) dx_i \right) p_{\theta_0}(x_{1:i-1}) dx_1 \dots dx_{i-1} \right) \\
 &< \log 1 \\
 &< 0.
 \end{aligned}$$

De même, on peut montrer que :

$$\begin{aligned}
 E^{(j)} \left[\log \frac{p_{\theta_0}(X_i | X_{1:i-1})}{p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1})} \right] &< 0 \\
 \Rightarrow E^{(j)} \left[\log \frac{p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \right] &> 0. \quad \square.
 \end{aligned}$$

Le temps d'arrêt du CUSUM généralisé est défini comme le premier instant où la statistique de test franchit un seuil h :

$$t_{CUSUM} = \inf\{n : g_n \geq h\}.$$

2.3.3 Résultats d'optimalité

Généralement, on définit l'optimalité d'une procédure de détection par le fait qu'elle minimise un critère choisi, parmi un ensemble de procédures vérifiant une contrainte fixée. C'est à partir de cette contrainte que l'on va déterminer le seuil h de la règle de décision. Evidemment, le choix d'une contrainte ou d'une autre va dépendre directement du type d'application. Pour des procédés issus des biotechnologies, on peut penser que la contrainte du temps moyen entre deux fausses alarmes

est un choix cohérent. En effet, sur ce type d'application, l'expérimentateur souhaite qu'il n'y ait pas trop de fausses alarmes mais d'un autre côté, il ne s'agit pas d'un procédé critique (système de navigation d'un avion) pour lequel on doit utiliser des contraintes plus sévères. Le temps moyen entre deux fausses alarmes est donc la contrainte que nous utiliserons généralement dans ce document. Ainsi, pour un temps moyen entre deux fausses alarmes égal à une constante γ :

$$E_{\theta_0}[t_a] = \gamma, \quad (2.9)$$

nous nous intéresserons au retard à la détection des différentes règles.

La fonction $\theta \mapsto E_{\theta}[t_a]$ est appelée *fonction ARL* (de l'anglais "average run length"). Elle a une grande importance pour tout ce qui concerne le choix du seuil. Nous y reviendrons dans le chapitre 3.

Evidemment, d'autres contraintes sont envisageables : dans le chapitre 3 sur le seuil adaptatif, nous considérons une contrainte plus restrictive qui fixe la probabilité de fausse alarme à chaque pas de temps. Lai [50] fixe la probabilité de fausse alarme sur une fenêtre de pas de temps. Enfin, on peut trouver des contraintes non pas liées aux fausses alarmes mais au retard à la détection : pour les systèmes de navigation (avion ou missile), on peut envisager que la contrainte soit que le retard à la détection ne dépasse pas un nombre de pas de temps fixé, le critère d'optimalité devenant alors le taux de fausses alarmes.

De la même façon, plusieurs critères sont envisageables pour évaluer le retard à la détection. Connaissant la trajectoire des observations avant l'instant de changement, le retard moyen à la détection est évalué par :

$$E_{\theta_1}(t_a - t_p + 1 | t_a \geq t_p, X_{1:t_p-1}) \quad (2.10)$$

REMARQUE : A partir de l'instant t_p , les observations X sont sous H_1 . Dans ce document, on va supposer que la panne a en fait eu lieu entre les instant $t_p - 1$ et t_p . Ainsi, si on détecte le changement en t_p , on considère que le retard à la détection vaut 1. Le retard à la détection est donc égal à $t_a - t_p + 1$. Dans la littérature, on peut voir la quantité $t_a - t_p$, qui indique donc que le changement a eu lieu exactement en t_p .

Le retard moyen défini en (2.10) dépend de la trajectoire avant l'instant de panne mais aussi du temps de panne lui-même qui est supposé inconnu, mais non aléatoire. Pour enlever ces dépendances, on trouve essentiellement deux critères :

- Le critère du pire retard moyen à la détection, qui a été introduit par Lorden [62]. Il consiste à prendre le pire des cas, tant au niveau de l'instant de panne que des observations du passé. Il est défini de la façon suivante :

$$\bar{E}_{\theta_1}(t_a) = \sup_{t_p \geq 1} \sup_{\text{ess}} E^{(t_p)}[(t_a - t_p + 1)^+ | X_{1:t_p-1}], \quad (2.11)$$

Chapitre 2. Méthodes statistiques de détection

où le supremum essentiel est pris par rapport à X_1, \dots, X_{t_p-1} .

- Le deuxième critère est un peu plus souple puisqu'on ne considère plus le supremum essentiel. Ce critère fut proposé par Pollak [80] et utilisé notamment par Pollak et Siegmund [83] et s'écrit :

$$\sup_{t_p \geq 1} E^{(t_p)}[t_a - t_p + 1 | t_a \geq t_p]. \quad (2.12)$$

*

* *

Le premier résultat d'optimalité pour la règle du CUSUM est obtenu par Lorden [62] pour des données indépendantes et identiquement distribuées, avant et après l'instant de changement. Il montre que, parmi toutes les règles de décision ayant un temps moyen entre deux fausses alarmes égal à une constante γ , la règle du CUSUM minimise, asymptotiquement sur γ , le critère du pire retard moyen à la détection $\bar{E}_{\theta_1}(t_a)$. Lorden obtient de plus, le résultat suivant :

$$\inf_{T: E_{\theta_0}(T) \geq \gamma} \bar{E}_{\theta_1}(T) \sim \frac{\log \gamma}{I(p_{\theta_1} | p_{\theta_0})} \quad \text{lorsque } \gamma \rightarrow \infty, \quad (2.13)$$

où $I(p_{\theta_1} | p_{\theta_0})$ représente la distance de Kullback-Leibler entre les deux distributions sous H_0 et H_1 . Ainsi, lorsque le temps moyen entre deux fausses alarmes tend vers l'infini, autrement dit, lorsque le seuil du test tend vers l'infini, la règle du CUSUM est optimale pour le critère du pire retard moyen à la détection. Plus tard, Moustakides [70] et Ritov [86] établissent ce résultat dans un cadre non asymptotique, c'est à dire à γ fixé.

Les premiers résultats d'optimalité pour des données dépendantes sont obtenues par Bansal et Papantoni-Kazakos [3]. Ils établissent l'optimalité asymptotique du CUSUM (généralisé à des données dépendantes) pour des données stationnaires ergodiques avant et après l'instant de changement. Ils supposent de plus qu'il y a indépendance entre $\{X_n, n < t_p\}$ et $\{X_n, n \geq t_p\}$. Autrement dit, ils se placent dans la troisième méthode de génération de changement évoquée dans la sous partie 2.3.2.

Lai [50] établit l'optimalité asymptotique du CUSUM au sens de Lorden pour des données dépendantes dans un contexte plus général que celui de Bansal et Papantoni-Kazakos. En effet, il considère la deuxième façon de générer le changement et il détermine une borne inférieure du pire retard moyen à la détection pour tous les algorithmes ayant un temps moyen entre deux fausses alarmes fixé, et il montre ensuite que la règle du CUSUM atteint cette borne, autrement dit, qu'elle est optimale. Pour établir cette optimalité, Lai effectue des hypothèses de stabilité du rapport de

vraisemblance du type :

$$\frac{1}{n} \sum_{i=t_p}^{n+t_p} \frac{p_{\theta_1}(X_i|X_{1:i-1})}{p_{\theta_0}(X_i|X_{1:i-1})} \rightarrow I \quad \text{en probabilité (sous } P^{(t_p)} \text{) lorsque } n \rightarrow \infty, \quad (2.14)$$

où I est une constante.

- Dans le cas *iid* (indépendant et identiquement distribués), la relation (2.14) est vérifiée lorsque $I = I(p_{\theta_1}|p_{\theta_0})$ est l'information de Kullback-Leibler entre les deux distributions sous H_0 et sous H_1 .

- Dans le cas de données dépendantes, on retrouve souvent des hypothèses du type (2.14) que ce soit dans le cas des tests d'hypothèses (Lai [48]) ou dans la détection de panne (Tartakovsky et Veeravalli [109]).

Nous reviendrons un peu plus en détails dans les chapitres suivants sur ces résultats d'optimalité et notamment sur les hypothèses de stabilité du rapport de vraisemblance.

2.4 Dérivation du CUSUM

2.4.1 Paramètre caractérisant la panne inconnu

Dans cette partie, nous donnons l'écriture des règles de décision dans le cas où le changement est généré suivant la deuxième méthode évoquée dans la partie 2.3.2 puisqu'il s'agit de l'écriture usuelle rencontrée dans la littérature et le cas pour lequel les résultats d'optimalité ont été obtenus. Cependant, l'écriture de toutes les règles qui suivent, pour la première méthode de génération, est directe, il suffit juste de remplacer p_{θ_1} par $p_{\theta_0, \theta_1, j}$.

La règle du CUSUM traite le cas où les valeurs prises par le paramètre θ : θ_0 et θ_1 , sont supposées connues. Supposer que la valeur θ_0 est connue n'est pas très restrictif en pratique puisque cette valeur correspond au fonctionnement normal du système. On peut donc imaginer que cette valeur a été bien estimée dans une phase préliminaire de la surveillance. Par contre, supposer que la valeur θ_1 est connue est beaucoup plus restrictif puisque c'est rarement le cas en pratique. En effet, on connaît plus ou moins "l'allure" d'une panne mais il y a toujours une incertitude sur son intensité. Si on revient au cas des procédés biotechnologiques, si l'anomalie considérée est une baisse de la concentration en entrée de la biomasse (S_{ini-H_0}), il est fréquent qu'il soit difficile de donner a priori la valeur exacte de la concentration sous régime de panne S_{ini-H_1} . Par contre, on pourra donner un intervalle pour cette concentration du type : $S_{ini-H_1} \in [a, b]$.

La mise au point de méthodes statistiques capables de répondre à ce type de problème est donc d'un grand intérêt. Supposons que sous l'hypothèse H_0 , les ob-

servations ont pour densité conditionnelles $p_{\theta_0}(\cdot|X_{1:i-1})$, θ_0 connue. Et, à partir de l'instant de changement t_p , la densité conditionnelle est $p_{\theta_1}(\cdot|X_{1:i-1})$ où θ_1 est inconnue mais appartient à un intervalle Θ_1 , qui lui, est connu. Une des règles de décision les plus connues est la règle du GLR (de l'anglais "generalized likelihood ratio"), introduite par Lorden [62] pour des données indépendantes. Cette règle est en fait une modification du CUSUM, c'est à dire qu'on va, dans l'expression du CUSUM (2.8), remplacer la vraisemblance inconnue p_{θ_1} par son estimateur de maximum de vraisemblance sur Θ_1 .

Le temps d'arrêt du GLR est ainsi défini de la façon suivante :

$$t_{GLR} = \inf \left\{ n : \max_{1 \leq j \leq n} \sup_{\theta \in \Theta_1} \sum_{i=j}^n \log \frac{p_{\theta}(X_i|X_{1:i-1})}{p_{\theta_0}(X_i|X_{1:i-1})} \geq h \right\}. \quad (2.15)$$

Dans le cas d'un changement de moyenne dans une suite de variables aléatoires gaussiennes (les observations sont indépendantes et donc le GLR s'écrit en remplaçant la densité conditionnelle par la densité des observations dans (2.15)), l'optimalité au sens de Lorden de la règle du GLR a été obtenue par Siegmund et Venkatraman [99].

Contrairement à la règle du CUSUM, le GLR ne peut pas être écrit sous forme récursive et d'autre part, la procédure d'optimisation de la vraisemblance sur Θ_1 est généralement coûteuse en temps de calcul, quel que soit le logiciel utilisé. Une solution pour diminuer les temps de calcul consiste alors à remplacer, dans l'expression (2.15), la maximisation $\max_{1 \leq j \leq n}$ par $\max_{n-M \leq j \leq n}$ où M sera appelée la fenêtre :

$$\tilde{t}_{GLR} = \inf \left\{ n : \max_{n-M \leq j \leq n} \sup_{\theta \in \Theta_1} \sum_{i=j}^n \log \frac{p_{\theta}(X_i|X_{1:i-1})}{p_{\theta_0}(X_i|X_{1:i-1})} \geq h \right\}. \quad (2.16)$$

On parle alors de GLR à fenêtre limitée, et cette règle présente l'avantage de garder la quantité de calcul constante au cours du temps. Cet algorithme a été introduit par Willsky et Jones [112] dans le cadre d'un changement additif dans un modèle à espace d'état linéaire (nous reviendrons sur cet algorithme dans la partie concernant les observations indirectes).

Lai [50] établit l'optimalité du GLR pour différents types de contraintes et Lai et Shan [53] s'intéressent aux choix de la fenêtre ainsi que du seuil du test en pratique.

Une alternative au GLR, et donc à la maximisation de la vraisemblance sur Θ_1 est proposée par Pollak et Siegmund [83] qui intègrent la vraisemblance p_{θ} par rapport à une distribution G , qui peut représenter la connaissance a priori qu'on a sur la vraie valeur de θ_1 . On parle alors de la règle du rapport de vraisemblance de mélange (algorithme MLR, "mixture likelihood ratio"). Dans le cas de données

indépendantes et identiquement distribuées avant et après l’instant de changement, Pollak et Siegmund [83] obtiennent des résultats d’optimalité pour le critère (2.12).

Dans le cas de données dépendantes, le temps d’arrêt de l’algorithme MLR s’écrit :

$$t_{MLR} = \inf \left\{ n : \max_{1 \leq k \leq n} \log \frac{\int_{\Theta_1} \prod_{i=k}^n p_{\theta}(X_i | X_{1:i-1}) dG(\theta)}{\prod_{i=k}^n p_{\theta_0}(X_i | X_{1:n})} \geq h \right\}. \quad (2.17)$$

Lai [50], sous des hypothèses de stabilité du logarithme du rapport de vraisemblance, montre l’optimalité de l’algorithme MLR au sens de Lorden, mais aussi pour une version à fenêtre limitée de cet algorithme.

Une autre méthode pour traiter le cas où la valeur θ_1 est inconnue, consiste à estimer cette valeur à partir des observations du passé pour calculer les incréments de l’algorithme de détection. Dragalin [26] ou encore Krieger, Pollak et Yakir [47] adapte le CUSUM de la façon suivante :

$$t_a = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \log(p_{\theta_{j,n}}(X_i)/p_{\theta_0}(X_i)) \geq h \right\},$$

où $\theta_{j,n}$ est un estimateur construit grâce à X_j, \dots, X_n .

Lorden et Pollak [64] et Krieger, Pollak et Yakir [47] adaptent de la même façon la règle de Shirayayev-Roberts. Cette règle de décision, introduite simultanément par Shirayayev [97] et Roberts [88] est similaire à l’algorithme CUSUM puisque le temps d’arrêt de ce test s’écrit :

$$t_{SR} = \inf \left\{ n : \sum_{k=1}^n \prod_{i=k}^n \frac{p_{\theta_1}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \geq h \right\}.$$

La version “mélangée” de cette règle a été introduite par Pollak [81] et on parle alors de la règle SRP.

La façon d’aborder le problème en estimant la valeur θ_1 est, selon Lorden et Pollak [64], assez similaire, du moins asymptotiquement, aux versions mélangées de ces mêmes tests. Elle trouve son intérêt dans les cas où il est difficile de choisir une mesure a priori pertinente pour l’intégration de la densité conditionnelle, comme dans le cas où la dimension de θ est élevée. On peut aussi remarquer que l’algorithme GLR est construit sous cette forme d’estimation de θ_1 , où dans ce cas, l’estimateur n’est rien d’autre que l’estimateur du maximum de vraisemblance.

On peut enfin citer les travaux de Mei [66, 67] qui s’intéresse au cas où la valeur θ_0 est inconnu et appartient à un intervalle Θ_0 . L’optimalité étudiée consiste alors à maximiser le temps moyen entre deux fausses alarmes pour un retard moyen à la détection fixé.

2.4.2 Problèmes plus complexes

La règle du CUSUM originelle est donc un algorithme de détection de changement de paramètre dans la densité, ou la densité conditionnelle des observations (ou plus généralement, un changement de structure de ces densités). Ce problème de changement de paramètre correspond donc à une panne abrupte sur le système. Plusieurs auteurs, parmi eux Krieger, Pollak et Yakir [47, 82], se sont intéressés à des types de panne graduelle. Le CUSUM est alors adapté à cette situation où la valeur θ_1 du paramètre après l'instant de panne peut être modélisée, par exemple, par une régression linéaire.

Un autre problème de détection qui a un intérêt grandissant est la détection multi-canaux, qui trouve des applications notamment en informatique avec tout ce qui concerne les attaques de virus sur internet. La problématique statistique est la suivante : il y a N populations $X_n^i, i = 1, \dots, N$ mutuellement indépendantes, qui ont chacune pour densité de probabilité $f_i(x)$ sous régime de fonctionnement normal. A un instant t_p , un changement se produit sur une et une seule population. Si le changement intervient sur la population j , alors $X_n^j, n \geq t_p$ a pour densité $g_j(x)$. Le CUSUM multi-canaux consiste alors à appliquer un test CUSUM pour chaque population puisque la panne n'intervenant que sur un seul canal, un unique test CUSUM sur l'ensemble des données pourrait entraîner un masquage de la panne par toutes les autres données, provenant d'un régime normal. On peut citer dans ce domaine les travaux de Tartakovsky [106, 108], qui établit notamment l'optimalité du CUSUM multicanaux au sens de Lorden et propose d'autres algorithmes pour aborder le problème, par exemple dans le cas où les distributions avant et après le changement sont inconnues.

2.5 Autres méthodes de détection

2.5.1 Les méthodes bayésiennes

Une autre façon d'aborder le problème est de se placer dans un contexte bayésien. Ce type de méthodes consiste à considérer une loi de probabilité a priori (généralement une loi géométrique) sur le temps de panne t_p . La règle de décision est alors construite de sorte qu'une alarme soit déclenchée lorsque la probabilité a posteriori qu'un changement a eu lieu dépasse un certain seuil. Ces méthodes ont été introduites par Girshick et Rubin [35] et des résultats d'optimalité ont été obtenus par Shiryaev [98] dans le cas de données *iid* et par Yakir [114] pour des modèles markoviens. Tartakovsky et Veeravalli [109] établissent l'optimalité de la règle de Shiryaev dans des cas plus généraux (données non *iid* et distribution a priori autre que géométrique).

Récemment, Tartakovsky [107] établit l'optimalité d'une procédure de détection bayésienne pour une contrainte différente de [109]. Les hypothèses de stabilité sur

le rapport de vraisemblance qu'il effectue sont alors moins restrictives que celles de Lai [50]. Nous reviendrons sur ces deux résultats d'optimalité et les hypothèses correspondantes dans le chapitre 4. La règle de décision que Tartakovsky considère est la suivante :

$$t_a = \inf \left\{ n : \sum_{k=1}^n \pi_k \prod_{i=k}^n \frac{p_{\theta_1}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \geq h \right\}, \quad h > 1 \quad (2.18)$$

où $\pi_k = P[t_p = k]$ est la loi a priori sur l'instant de panne.

2.5.2 L'approche asymptotique locale

Comme cela a été dit dans l'introduction, le problème basique de détection de changement (voir Basseville et Nikiforov [6]) revient à détecter une variation de la moyenne dans une série gaussienne. De nombreuses règles de décision, parmi lesquelles celles que nous venons de présenter, ont été mises au point pour ce type de problème.

Dans de nombreux cas, comme celui d'un changement additif dans un modèle de type ARMAX avec bruits blancs gaussiens additifs, le problème de détection se réduit à la surveillance de la moyenne d'un processus gaussien. On retombe donc sur le problème classique.

Pour un type de changement quelconque (et donc un changement multiplicatif), la réduction du problème de détection au problème basique s'effectue au travers de l'approche asymptotique locale. Rappelons brièvement le fonctionnement de cette méthode détaillée notamment dans Basseville [4] et dans Basseville et Nikiforov [6].

Sous fonctionnement normal H_0 , $\theta = \theta_0$, et sous régime de panne H_1 , $\theta = \theta_0 + \Delta/\sqrt{N}$ où N est la taille de l'échantillon (et correspond donc au pas de temps) et Δ est inconnu mais fixe.

On génère alors des résidus sous la forme :

$$\xi_N = \sqrt{N}U_N(\theta_0),$$

où la fonction d'estimation U_N , qui dépend des observations, vérifie entre autre : $E_{\theta}[U_N(\theta_0)] = 0$ si et seulement si $\theta = \theta_0$. Alors les résidus sont asymptotiquement (sur N) gaussiens :

$$\text{sous } H_0, \quad \xi \sim \mathcal{N}(0, \Sigma) \quad \text{et} \quad \text{sous } H_1, \quad \xi \sim \mathcal{N}(M\Delta, \Sigma),$$

où Σ et M peuvent être estimées.

Pour construire la fonction U_N , on peut utiliser notamment la fonction score (c'est-à-dire la dérivée de la fonction de log-vraisemblance par rapport au paramètre) ou d'autres fonctions.

2.6 Détection dans les modèles à espace d'état

Nous l'avons vu dans le premier chapitre, les méthodes de détection d'anomalie doivent s'adapter au cas où les variables d'intérêt ne sont pas observées directement. Depuis quelques années, des approches statistiques sont proposées pour répondre à ce problème. Considérons le modèle à espace d'état suivant :

$$\begin{cases} X_n = f(X_{n-1}, v_n) \\ Y_n = h(X_n, w_n) \end{cases}$$

où :

- X_n est la variable d'état,
- Y_n est la variable d'observation,
- v_n est le bruit sur le modèle d'état,
- w_n est le bruit sur le modèle d'observation.

Généralement, lorsqu'on se place en dehors du contexte de la détection d'anomalie, l'objectif est, pour ce type de modèle, de reconstruire l'état X en fonction des observations Y , autrement dit, dans un cadre stochastique, de trouver la loi conditionnelle de X sachant les observations du passé Y . Pour répondre à cette question, on utilise la théorie du filtrage. Pour des systèmes linéaires gaussiens, la solution est donnée par le filtre de Kalman. Pour des systèmes non linéaires, des outils d'approximation efficaces ont été mis au point récemment, il s'agit de la théorie du filtrage particulière, dont nous rappelons les principes en annexe B.

Revenons à la détection de changement. Comme dans les parties précédentes, un paramètre θ caractérisant la panne (θ_0 sous H_0 et θ_1 sous H_1) intervient dans f et/ou h .

2.6.1 Panne additive dans un modèle à espace d'état linéaire

Nous considérons le modèle à espace d'état linéaire suivant :

$$\begin{cases} X_n = aX_{n-1} + v_n \\ Y_n = bX_n + w_n \end{cases} \quad (2.19)$$

où a, b sont des réels (ou plus généralement des matrices) et v_n et w_n sont des bruits blancs gaussiens de variances respectives σ_v^2 et σ_w^2 et mutuellement indépendants. Pour des systèmes linéaires de ce type, l'approche statistique de détection la plus utilisée actuellement consiste à appliquer le filtre de Kalman afin de retomber sur un "problème basique" (Basseville et Nikiforov [6]) de changement de moyenne dans une suite de variables gaussiennes.

Le filtre de Kalman permet d'obtenir, à partir d'un algorithme récursif, la loi conditionnelle de X_n sachant Y_1, \dots, Y_n (cette loi est gaussienne). Le processus d'innovation, défini par :

$$e_n := Y_n - E[Y_n|Y_{1:n-1}] = Y_n - bE[X_n|Y_1, \dots, Y_{n-1}], \quad (2.20)$$

est alors un processus gaussien de moyenne nulle et de variance $\sigma_e^2(n)$.

Nous allons supposer qu'à un instant t_p inconnu, un changement additif affecte le système dans le sens où la quantité $\theta \mathbb{1}_{\{n \geq t_p\}}$ est ajoutée dans l'équation d'état ou dans l'équation d'observation du modèle (2.19). Le processus (2.20) avec X_n toujours défini par (2.19) n'est plus un processus d'innovation. Il s'agit toujours d'une suite de variables gaussiennes indépendantes de variance $\sigma_e^2(n)$ mais sa moyenne $m_n = E[e_n]$ n'est plus nulle pour $n \geq t_p$ et varie dans le temps. Lai et Shan [53] donnent un algorithme récursif pour évaluer la suite m_n qui dépend de l'instant de panne.

Le problème de détection revient alors à repérer un changement de moyenne dans une suite de variables aléatoires gaussiennes et la règle du CUSUM proposée par Willsky et Jones [112] et dont des résultats d'optimalité ont été obtenus par Lai et Shan [53] s'écrit alors :

$$t_W = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{f((e_i - m_i)/\sigma_e(i))}{f(e_i/\sigma_e(i))} \right\} \quad (2.21)$$

où $f(y) = \exp(-y^2/2)/(2\pi)^{1/2}$ est la densité d'une gaussienne.

L'algorithme de Willsky et Jones [112] est en fait plus général puisqu'il traite le cas où le paramètre θ_1 est inconnu, et donc, à la manière du GLR, on effectue une maximisation sur Θ_1 du paramètre θ_1 . De plus, pour contourner le problème de l'augmentation des calculs, les auteurs considèrent une fenêtre mobile $[n - M, n]$ sur laquelle va être effectuée la maximisation (au lieu de $[1, n]$). Cela permet de garder le nombre d'opérations constant tout au long du procédé.

Comme nous le verrons sur des simulations dans le chapitre 6, cette méthode marche très bien pour des systèmes linéaires avec bruits blancs gaussiens additifs et une panne additive. Lorsque le système est non linéaire ou que la panne n'est plus additive, plus rien ne garantit le bon fonctionnement de cette règle. En effet, pour traiter la non linéarité, on utilise alors un filtre de Kalman étendu en considérant toujours que le processus d'innovation est gaussien alors qu'il ne l'est plus. Pour ces cas là, il est donc nécessaire de mettre au point des algorithmes plus généraux.

2.6.2 Modèles plus généraux et filtrage particulaire

Dans cette partie, nous nous intéressons à des situations plus complexes que dans la partie précédente, c'est à dire à des systèmes :

- dont la modélisation est non linéaire,

- qui peuvent être affectés par des changements multiplicatifs,
- dont les bruits blancs peuvent être non gaussiens et/ou non additifs.

Dans ce genre de situations, les approches de détection basées sur l'utilisation de filtres particulières donnent de bons résultats. Cérou et Le Gland [18] génèrent des résidus. Plus précisément, ils évaluent la fonction score à partir d'un filtre à particules.

Toujours à partir de méthodes particulières, Vaswani [110] propose plusieurs statistiques de test capables de détecter des changements de paramètre. Les valeurs des paramètres sont inconnues et les changements peuvent être lents ou soudains.

Zhang, Campillo, Cérou et Le Gland [117], dans un contexte de diagnostic, utilisent des filtres particulières en parallèle qui ignorent des fautes possibles sur le système. Par exemple, un filtre est mis en place de sorte à rejeter la i -ème faute. Si ce filtre ne marche pas bien, c'est à dire si la vraisemblance du filtre est petite, on en conclut que la i -ème faute a eu lieu.

Enfin, plusieurs auteurs, dont Andrieu, Doucet, Singh et Tadic [1], Azimi-Sadjadi et Krishnaprasad [2], et Li et Kadiramanathan [58, 59], ont cherché à aborder le problème comme on le fait pour la règle du CUSUM. C'est l'approche à laquelle nous nous sommes intéressés. L'utilisation de filtres en parallèle décrivant le fonctionnement normal du système ainsi que le ou les régimes de panne permet de reconstruire (ou plutôt d'approcher) la vraisemblance conditionnelle $p_\theta(Y_k|Y_{1:k-1})$ sous H_0 et sous H_1 . La mise en place d'une règle de décision de type CUSUM est alors possible. La complexité de cette approche vient du fait que la vraisemblance conditionnelle sous régime de panne dépend de l'instant de changement de régime. On est donc dans le cas de figure où le changement est généré par la première méthode de la partie 2.3.2. Un seul filtre ne suffira pas à caractériser le fonctionnement anormal du système : il en faudra autant que d'instants possibles de panne. Pour répondre à ce problème, Azimi-Sadjadi et Krishnaprasad [2] considèrent une statistique de test légèrement différente pour obtenir un algorithme récursif. Les autres auteurs proposent d'utiliser, à la manière du GLR ou de la règle de Willsky et Jones [112], une fenêtre qui garantit un nombre de calculs constant au cours du temps. Nous reviendrons plus en détails sur ces algorithmes dans le chapitre 6.

Pour toutes ces méthodes, il n'existait pas, à notre connaissance, de résultat d'optimalité comme c'est le cas pour la règle du CUSUM. Nous avons cherché à en obtenir.

2.7 Le diagnostic d'anomalie

La surveillance d'un processus biologique ou autre ne se limite malheureusement pas à la détection d'une seule défaillance. De nombreuses pannes peuvent intervenir sur ces installations, certaines étant plus ou moins graves que d'autres. La mise en

place de méthodes capables de détecter et de localiser une panne parmi un ensemble de défaillances possibles est donc d'un grand intérêt. Le principe est le suivant : avant l'instant de panne, le système est sous le régime H_0 et à partir de l'instant t_p , il passe sous un régime $H_i, i = 1, \dots, K$, où K représente le nombre de pannes possibles. La réponse d'un algorithme de diagnostic consiste donc en un temps d'alarme ainsi qu'en un choix de type de panne.

Nikiforov [72] est un des tous premiers à traiter le problème du diagnostic, c'est à dire à traiter conjointement la détection et la localisation de panne. Il considère des observations *iid* et obtient des résultats d'optimalité asymptotique pour son algorithme, qui est une généralisation du CUSUM au problème multi-hypothèses. Plus tard, Lai [51] montre l'optimalité de l'algorithme de Nikiforov pour des données dépendantes en considérant la densité conditionnelle des observations. Il procède de la même manière que dans [50] en effectuant des hypothèses de stabilité des rapports de vraisemblances. Mais ces deux algorithmes ne possèdent pas d'écriture récursive ce qui peut rendre leur utilisation en pratique difficile, surtout s'il y a beaucoup d'hypothèses alternatives.

Deux algorithmes récursifs de diagnostic de panne ont été mis au point au cours des dernières années, il s'agit de ceux de Nikiforov [73, 74] et de Oskiper et Poor [75]. Dans les deux cas, les auteurs ont établi des propriétés d'optimalité dans le cas de données *iid*. Nous avons comparé ces deux méthodes en simulation sur des données dépendantes. Nous avons obtenu des comportements semblables en terme de retard à la détection, l'un ou l'autre algorithme prenant l'avantage suivant le type de panne considéré.

Comme l'est le CUSUM en détection, nous avons donc choisi un algorithme de référence pour le diagnostic. Notre choix s'est porté sur un algorithme récursif bien que les résultats d'optimalité ne soient valables que pour des données *iid*. Ainsi, pour la suite, nous considérerons plutôt l'algorithme de diagnostic de Nikiforov puisqu'il nous est apparu plus "lisible" que celui de Oskiper et Poor, où certaines statistiques de test ne sont pas forcément négatives en moyenne sous H_0 .

2.7.1 L'algorithme récursif de Nikiforov [74]

Nous considérons l'algorithme de Nikiforov pour des données non *iid* et pour un changement généré par la deuxième méthode, où la densité conditionnelle ne dépend pas de l'instant de changement de régime. Le but d'un algorithme de diagnostic est de fournir un temps d'arrêt (ou temps d'alarme, noté t_{rec}) et le type de panne correspondant, noté ν_{rec} . Pour K types de pannes possibles, on a donc

$$\begin{aligned} X_n | X_{1:n-1} &\sim p_{\theta_0}(\cdot | X_{1:n-1}) \text{ si } n < t_p \\ X_n | X_{1:n-1} &\sim p_{\theta_i}(\cdot | X_{1:n-1}) \text{ si } n \geq t_p \end{aligned}$$

Chapitre 2. Méthodes statistiques de détection

avec $l = 1, \dots, K$. Le temps de panne t_p et le type de panne l sont inconnus et on note :

$$Z_n(l, j) = \log(p_{\theta_l}(X_n|X_{1:n-1})/p_{\theta_j}(X_n|X_{1:n-1})) ,$$

le logarithme du rapport des vraisemblances conditionnelles entre les hypothèses $H_l : \theta = \theta_l$ et $H_j : \theta = \theta_j$.

L'Algorithme

Il s'agit d'un algorithme récursif qui n'est rien d'autre que l'association de K tests CUSUM parallèles. Il s'écrit :

$$t_a = \min \{N_1, \dots, N_K\} \quad (2.22)$$

$$\nu = \arg \min \{N_1, \dots, N_K\} \quad (2.23)$$

où N_l est défini par :

$$N_l = \inf \left\{ t \geq 1 : \min_{0 \leq j \neq l \leq K} [g_t(l, 0) - g_t(j, 0) - h_{l,j}] \geq 0 \right\} , \quad (2.24)$$

où les fonctions de décision $g_t(l, 0)$ sont définies comme pour le CUSUM écrit sous sa forme récursive :

$$g_t(l, 0) = (g_{t-1}(l, 0) + Z_t(l, 0))^+ , \quad g_0(l, 0) = 0 , \quad l = 1, \dots, K . \quad (2.25)$$

Les seuils $h_{l,j}$ sont choisis de la façon suivante :

$$h_{l,j} = \begin{cases} h_{det} & \text{si } l = 1, \dots, K \text{ et } j = 0 \\ h_{loc} & \text{si } j, l = 1, \dots, K \text{ et } j \neq l \end{cases} \quad (2.26)$$

où h_{det} est le *seuil de détection* et h_{loc} est le *seuil de localisation*, et avec $h_{det} \geq h_{loc}$.

Optimalité

Nikiforov [74] montre que l'algorithme précédent, dans le cas de données indépendantes, minimise le critère suivant :

$$\bar{E}(t_a) = \sup_{t_p \geq 1, 1 \leq l \leq K} E_{(t_p)}^l(t_a - t_p + 1 | t_a \geq t_p) \quad (2.27)$$

parmi toutes les règles de décision vérifiant les contraintes suivantes :

$$E_{\theta_0}(t_a) \geq \gamma \quad (2.28)$$

$$\text{et } \max_{1 \leq l \leq K} \max_{1 \leq j \neq l \leq K} \sup_{t_p \geq 1} P_{(t_p)}^l(\nu = j | t_a \geq t_p) \leq b. \quad (2.29)$$

2.7.2 Application

Nous allons appliquer l'algorithme de diagnostic à une simulation d'un modèle décrivant une réaction, issue des biotechnologies, mono-substrat mono-biomasse (cf. Annexe C). Il y a deux variables d'état considérées, les concentrations en biomasse $(X_n)_n$ et en substrat $(S_n)_n$ dont les évolutions au cours du temps sont décrites par le modèle suivant (Bastin et Dochain [8]) :

$$\begin{cases} X_{n+1} = X_n + (TX_n)\mu_n - (1 - \theta)(TU_nX_n) + \epsilon_{n+1}^{(1)} \\ S_{n+1} = S_n - \frac{TX_n}{\tau}\mu_n + (1 - \theta)U_n(S_0 - S_n)T + \epsilon_{n+1}^{(2)} \end{cases} \quad (2.30)$$

avec

$$\mu_n = \mu_{max} \frac{S_n}{K_s + S_n} \quad (\text{cinétique de Monod})$$

La quantité μ_n représente le taux de croissance de la biomasse. Les constantes ont pour valeurs : $T = 0.17h$, $S_0 = 50mg.l^{-1}$, $\tau = 1$, $K_s = 1g.l^{-1}$, $\mu_{max} = 0.05h^{-1}$. Les conditions initiales sont : $B_0 = 10.4mg.l^{-1}$, $S_0 = 39.5mg.l^{-1}$ et $U_n = 0.04h^{-1}$. Enfin $\epsilon_k^{(d)} \sim N(0, 10^{-4})$, $d = 1, 2$.

On va alors considérer trois pannes possibles. Les différentes valeurs du paramètre θ seront donc :

$$H_0 : \theta_0 = 0, \quad H_1 : \theta_1 = 0.01, \quad H_2 : \theta_2 = 0.02, \quad H_3 : \theta_3 = -0.02.$$

On va simuler notre modèle sur 5000 pas de temps en faisant intervenir une panne de type H_1 au temps $t = 4500$.

REMARQUE : Un même paramètre θ caractérise ici les trois pannes, qui peuvent être considérés comme des intensités de colmatage de la pompe différentes. Dans un tel cadre d'utilisation, l'algorithme de diagnostic peut être vu comme un algorithme de détection qui fournit, en plus d'un temps d'alarme, une intensité de panne.

Description de l'algorithme de Nikiforov

On va tout d'abord expliquer, de façon intuitive, le fonctionnement de l'algorithme (2.22)-(2.25).

Pour chaque type de panne possible l , on va effectuer toute une série de tests s'apparentant au CUSUM puisqu'on va comparer la différence entre les deux statistiques de test $g_t(l, 0)$ et $g_t(j, 0)$ à un seuil h pour toutes les hypothèses j différentes de H_l . Cette comparaison va s'apparenter à un test de l'hypothèse H_l contre l'hypothèse H_j .

Une panne sera alors détectée et localisée lorsque pour une hypothèse H_l , toutes les fonctions de décision $g_t(l, 0) - g_t(j, 0)$ de tous les tests H_l contre H_j , $0 \leq j \leq K$ seront supérieures aux seuils fixés.

Chapitre 2. Méthodes statistiques de détection

Dans notre cas et parce que nous avons trois hypothèses alternatives à H_0 , nous allons effectuer en tout, neuf tests :

Pour tester l'hypothèse H_1 , on aura :

$$H_1 \text{ vs } H_0, \quad H_1 \text{ vs } H_2, \quad \text{et} \quad H_1 \text{ vs } H_3.$$

Pour tester l'hypothèse H_2 , on aura :

$$H_2 \text{ vs } H_0, \quad H_2 \text{ vs } H_1, \quad \text{et} \quad H_2 \text{ vs } H_3,$$

et enfin, pour l'hypothèse H_3 .

$$H_3 \text{ vs } H_0, \quad H_3 \text{ vs } H_1, \quad \text{et} \quad H_3 \text{ vs } H_2,$$

Résultats de la simulation

La figure 2.2 représente l'allure des concentrations en biomasse et en substrat avec l'apparition d'une panne de type 1 au pas de temps 4500. La figure 2.3 représente

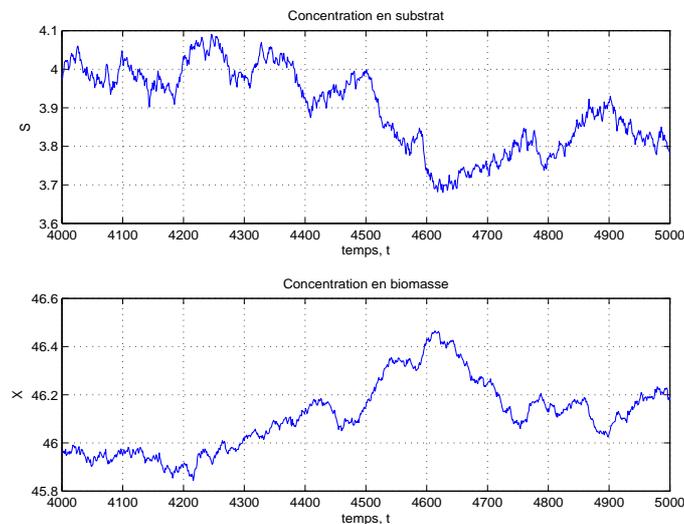


FIG. 2.2 – Graphes des données du processus biotechnologique

les graphes des différentes statistiques de test avec sur la première ligne, les trois tests concernant l'hypothèse H_1 , sur la deuxième, les tests de l'hypothèse H_2 et l'hypothèse H_3 sur la troisième.

Une anomalie sera détectée et identifiée comme une faute de type l dès que les trois tests de la l -ième ligne auront dépassé le seuil fixé par l'expérimentateur.

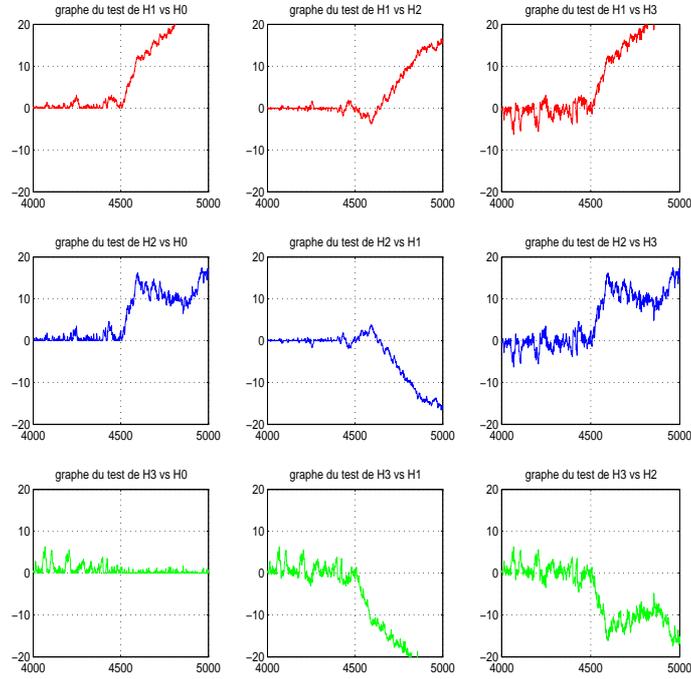


FIG. 2.3 – Graphes des statistiques des différents tests

Le tableau 2.1 donne les temps d'arrêt des neuf tests, c'est à dire les pas de temps (lorsqu'ils existent) où les statistiques de test dépassent un seuil fixé (empiriquement) à 10. On voit alors que l'alarme est déclenchée au pas de temps $t_a = 4812$ puisque c'est à ce moment là que les trois statistiques de tests de l'hypothèse H_1 ont franchi le seuil. Les temps d'arrêt noté NA correspondent à des non-détections : les statistiques de tests n'ont pas franchi les seuils.

H_1 vs H_0 $t_a = 4593$	H_1 vs H_2 $t_a = 4812$	H_1 vs H_3 $t_a = 4583$
H_2 vs H_0 $t_a = 4568$	H_2 vs H_1 $t_a = NA$	H_2 vs H_3 $t_a = 4568$
H_3 vs H_0 $t_a = NA$	H_3 vs H_2 $t_a = NA$	H_3 vs H_2 $t_a = NA$

TAB. 2.1 – Temps d'arrêt des neuf tests

On peut remarquer que les tests dits de localisation (H_1 vs H_2 , H_1 vs H_3 , H_2 vs H_1 ,...) jouent parfaitement leur rôle. En effet, si on ne s'intéresse qu'aux tests de détection H_1 vs H_0 , H_2 vs H_0 et H_3 vs H_0 , comme le font notamment Li

Chapitre 2. Méthodes statistiques de détection

et Kadiramanathan [59], nous aurions alors détecté la panne plus tôt, mais nous aurions mal localisé cette dernière (localisation de type 2 alors que la panne est de type 1).

Deuxième partie

Nouvelles approches de la
détection de rupture de modèle
pour les modèles dynamiques
incertains et les modèles
dynamiques à espace d'état

Chapitre 3

Seuils adaptatifs

3.1 Introduction

Comme nous l’avons évoqué dans le premier chapitre, les méthodes de détection et de diagnostic d’anomalie se décomposent généralement en deux étapes : une étape de génération des résidus, quantités proches de zéro sous régime de fonctionnement normal et significativement non nulles sous régime de panne, et une étape d’évaluation des résidus qui peut s’effectuer en comparant la valeur des résidus à un seuil choisi par l’expérimentateur.

Pour les méthodes statistiques étudiées dans cette thèse, l’approche va consister à comparer une statistique de test, construite de telle sorte qu’elle aura tendance à être négative sous H_0 , et positive et croissante sous H_1 , au seuil du test. La question de la valeur de ce seuil se pose alors.

Le choix du seuil est en fait un compromis entre le retard à la détection et le taux de fausses alarmes. Plus le seuil choisi est grand et moins il y aura de fausses alarmes mais plus le retard à la détection sera important. Au contraire, un seuil trop bas entraîne de nombreuses fausses alarmes. Généralement, le seuil est pris de sorte qu’une contrainte fixée par l’expérimentateur soit vérifiée : temps moyen entre deux fausses alarmes fixé, probabilité de fausse alarme sur une fenêtre de pas de temps fixée, retard à la détection inférieur à une certaine constante...

Dans la littérature statistique, on peut trouver un grand nombre de méthodes pour calculer le seuil dans le cas où les observations X_n sont indépendantes et identiquement distribuées avant et après l’instant de changement, en particulier grâce à la contrainte ARL (“average run length”). Dans le cas nous intéressant plus particulièrement, c’est à dire le cas des systèmes dynamiques, il y a très peu de travaux concernant le calcul du seuil.

Nous avons donc mis en place une méthode calculatoire qui nous permet d’obtenir un seuil pertinent pour des règles de décision construites de la même façon que la

règle du CUSUM. Cette méthode nous permet d'obtenir un seuil adaptatif, c'est à dire qui va varier au cours du temps en s'adaptant à la loi de la statistique de test pour garantir un test de niveau constant. Ce seuil présente donc l'avantage de maintenir le taux de fausses alarmes constant tout au long du processus, ce qui n'est pas forcément le cas pour des méthodes plus traditionnelles fournissant un seuil fixe. Nous reviendrons par la suite sur les problèmes qu'engendre l'utilisation d'un seuil fixe.

Nous avons étendu notre approche adaptative à un algorithme de diagnostic d'anomalie (Nikiforov [73]). Il est alors possible de contrôler à chaque instant le taux de fausses alarmes ainsi que la probabilité de faux diagnostic.

L'utilisation de seuils adaptatifs n'est pas originale puisque pour les méthodes issues de l'Automatique, certains auteurs (dont Frank [31]) l'utilisent depuis quelques années dans le but de prendre en compte l'incertitude sur le système, comme cela a été évoqué dans le chapitre 1.

REMARQUE : L'objectif de ce chapitre est de proposer un algorithme, facile à utiliser, et permettant d'obtenir un seuil qui garantit la vérification des contraintes imposées. Nous ne cherchons en aucun cas à obtenir des résultats théoriques et des approximations très précises sur la valeur du seuil, comme on peut en trouver dans la littérature dans le cas *iid*.

3.2 Choix du seuil et intérêt de notre approche

Le temps d'arrêt du CUSUM est défini de la façon suivante :

$$t_a = \inf\{n : g_n \geq h\} \quad \text{avec} \quad g_n = \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})}.$$

En pratique, on doit fixer une valeur pour le seuil h . Il est généralement choisi de sorte que le temps moyen entre deux fausses alarmes soit égal à :

$$E_{\theta_0}(t_a) = \gamma,$$

pour une valeur γ fixée. Cette contrainte, appelée contrainte ARL, est utilisée par Lorden [62] et Lai [50] notamment¹, pour établir l'optimalité asymptotique du CUSUM, et la fonction $\theta \mapsto E_{\theta}(t_a)$ est appelée fonction ARL. Des méthodes pour estimer la fonction ARL sont décrites dans Basseville et Nikiforov [6]. Mais la plupart de

¹Lai [50] utilise en fait la contrainte suivante : $E_{\theta_0}(t_a) \geq \gamma$. Pour obtenir l'optimalité de la règle du CUSUM dans un cadre dépendant, il est nécessaire de travailler avec cette contrainte, contrairement au cas indépendant, où l'égalité suffit.

ces méthodes traitent le cas gaussien indépendant. Ces approximations sont obtenues à partir de travaux sur la taille moyenne de l'échantillon ("average sample number") (Lorden [61, 63], Berk [9]) pour le test du rapport de vraisemblance séquentiel ("sequential probability ratio test"), dont le CUSUM peut être vu comme une extension pour répondre au problème de la détection de changement.

Pour le cas dépendant, Lai et Shan [53] proposent une nouvelle contrainte de la forme :

$$\sup_{k \geq 1} P_{\theta_0}(k \leq t_a \leq k + m_\alpha) \leq \alpha ,$$

et utilisent des simulations de Monte-Carlo pour obtenir le seuil correspondant. Cette contrainte permet de contrôler la probabilité de fausse alarme sur une fenêtre de taille m_α . Nous reviendrons sur cette contrainte dans le prochain chapitre.

Enfin, une méthode très simple pour obtenir le seuil du test consiste à estimer de façon empirique la fonction ARL à partir d'un grand nombre de simulations. Pour cela, on fixe une valeur h et on applique le test CUSUM à m trajectoires simulées sous H_0 . On retient alors les m instants d'alarmes t_a (qui sont en fait des fausses alarmes puisqu'on est sous H_0) et on a :

$$E_{\theta_0}(t_a) \simeq \sum_{i=1}^m \frac{t_a(m)}{m} . \quad (3.1)$$

Il suffit juste de répéter ces simulations en faisant varier la valeur de h jusqu'à obtenir le temps moyen souhaité avant la première fausse alarme. L'approximation consiste alors à supposer que ce temps moyen avant la première fausse alarme est le même que le temps moyen entre deux fausses alarmes.

Ces méthodes fournissent un seuil constant tout au long du procédé. Pour des processus indépendants ou stationnaires, un seuil fixe est parfaitement approprié comme nous le verrons en simulation, mais pour les processus non stationnaires, cela peut ne pas être le cas. En effet, pour ce type de processus, l'ordre de grandeur de la statistique de test g_n (de la règle du CUSUM ou d'une autre règle) peut varier fortement au cours du temps et il est donc inapproprié de comparer cette statistique de test au même seuil tout au long de la surveillance. De façon plus précise, si l'ordre de grandeur de la statistique de test change, la comparer à un seuil fixe va avoir pour conséquence le non respect de la contrainte fixée initialement, et ainsi, on peut s'attendre à une augmentation du taux de fausses alarmes ou au contraire, un retard à la détection plus grand.

L'approximation qui consiste à dire que le temps moyen avant la première fausse alarme est le même que le temps moyen entre deux fausses alarmes n'est clairement plus réaliste ici. Pour être certain de vérifier la contrainte ARL, il faudrait, en toute rigueur, recalculer la valeur du seuil après une fausse alarme.

Pour les systèmes avec une variable de commande, le fait de jouer sur cette variable pour conduire le procédé peut entraîner des variations de l'ordre de grandeur de la statistique de test, comme on le verra dans la dernière partie de ce chapitre. Il semble donc pertinent de pouvoir mettre au point une méthode de calcul qui fournisse un seuil s'adaptant à ce type de données et garantissant ainsi, un taux de fausses alarmes constant tout au long du procédé.

3.3 Seuil adaptatif pour l'algorithme CUSUM

3.3.1 Description et propriétés

Pour obtenir le seuil adaptatif, nous n'utilisons plus la contrainte ARL du temps moyen entre deux fausses alarmes mais nous considérons la probabilité conditionnelle de fausse alarme suivante :

$$\begin{aligned} \text{si } n &= 1, P_{\theta_0}[g_1 \geq h_1] = \alpha, \\ \text{et } \forall n &\geq 2, P_{\theta_0}[g_n \geq h_n | g_1 < h_1, \dots, g_{n-1} < h_{n-1}] = \alpha, \end{aligned} \quad (3.2)$$

où α est fixé par l'utilisateur. A chaque pas de temps n , nous allons choisir h_n de sorte que la probabilité d'avoir une fausse alarme en n , sachant qu'il n'y en a pas eu avant, soit égale à α . Le seuil h_n qui va vérifier cette contrainte n'est rien d'autre que le quantile d'ordre α de la loi conditionnelle de g_n sachant l'événement $\{g_1 < h_1, \dots, g_{n-1} < h_{n-1}\}$. Une estimation \hat{h}_n de ce quantile pourra être obtenue par simulation, à partir d'une estimation non paramétrique de la fonction de densité de la variable aléatoire g_n , ou tout simplement en utilisant le quantile empirique.

REMARQUE : La contrainte (3.2) est utilisée par Hawkins, Qiu et Kang [39] qui proposent une statistique de test non paramétrique pour détecter un changement de moyenne dans une suite de variables gaussiennes.

Lien avec la contrainte de Lorden

Les résultats d'optimalité de Lorden dépendent de la contrainte :

$$E_{\theta_0}(t_a) = \gamma. \quad (3.3)$$

Il est donc important de pouvoir faire le lien entre la contrainte ARL et celle que nous utilisons pour obtenir le seuil adaptatif. Supposons donc que la contrainte

(3.2) soit vérifiée pour un certain α . On a :

$$\begin{aligned}
 P_{\theta_0}[t_a = n] &= P_{\theta_0}[g_n \geq h_n, g_{n-1} < h_{n-1}, \dots, g_1 < h_1] \\
 &= P_{\theta_0}[g_n \geq h_n | g_1 < h_1, \dots, g_{n-1} < h_{n-1}] \cdot P_{\theta_0}[g_{n-1} < h_{n-1}, \dots, g_1 < h_1] \\
 &= \dots \\
 &= P_{\theta_0}[g_n \geq h_n | g_1 < h_1, \dots, g_{n-1} < h_{n-1}] \cdot \\
 &\quad \cdot P_{\theta_0}(g_1 < h_1) \cdot \prod_{i=2}^{n-1} P_{\theta_0}[g_i < h_i | g_1 < h_1, \dots, g_{i-1} < h_{i-1}]. \\
 &= \alpha(1 - \alpha)^{n-1}.
 \end{aligned}$$

Ainsi, t_a est un variable aléatoire géométrique de paramètre α . Son espérance vaut :

$$E_{\theta_0}[t_a] = \frac{1}{\alpha}.$$

Le résultat précédent établit un lien direct entre les contraintes (3.2) et (3.3) :

$$\gamma = 1/\alpha. \tag{3.4}$$

Estimation du seuil adaptatif

A partir de la modélisation connue du système, nous simulons M trajectoires $(X_i^m)_{i \geq 1, m=1, \dots, M}$ du processus sous l'hypothèse H_0 .

Pour $n = 1$: h_1 est le quantile d'ordre $(1 - \alpha)$ de la variable g_1 et il est donc estimé par \hat{h}_1^M à partir des statistiques de test g_1^m obtenues en appliquant la règle du CUSUM.

Pour $n \geq 2$: nous effectuons à chaque pas de temps les deux étapes suivantes :

Etape 1 : Nous appliquons le test CUSUM aux M trajectoires (sous H_0) (X_1^m, \dots, X_n^m) , $m = 1, \dots, M$, et nous utilisons les statistiques de test obtenues g_n^m , $m = 1, \dots, M$, qui sont *iid*, pour construire un estimateur \hat{h}_n^M du quantile d'ordre $1 - \alpha$ de la loi conditionnelle $g_n | g_1 < h_1, \dots, g_{n-1} < h_{n-1}$.

Etape 2 : Nous supprimons de l'échantillon les trajectoires dont la valeur de la statistique de test a dépassé le seuil \hat{h}_n^M à l'instant n puisqu'elles ne satisfont plus la condition de la contrainte (3.2). Nous re-simulons alors d'autres trajectoires ne dépassant pas les seuils $\hat{h}_1^M, \dots, \hat{h}_n^M$, en sorte d'avoir encore M trajectoires satisfaisant la condition de la contrainte, c'est à dire des trajectoires sans fausse alarme jusqu'à l'instant n .

- On retourne à l'étape 1 avec $n = n + 1$.

Chapitre 3. Seuils adaptatifs

REMARQUE : A chaque pas de temps, au moment du ré-échantillonnage, on supprime un pourcentage à peu près égal à α , de trajectoires ne vérifiant plus la contrainte fixée, que l'on remplace par des trajectoires n'ayant pas dépassées les seuils $\hat{h}_1^M, \dots, \hat{h}_n^M$. Mais plus on avance dans le temps, et plus il est difficile de simuler des trajectoires n'ayant jamais franchi le seuil adaptatif. Pour résoudre ce problème, on peut choisir de retenir des trajectoires n'ayant pas eu de fausse alarme depuis quelques pas de temps seulement.

Cette approximation revient en fait à remplacer la contrainte (3.2) par la suivante :

$$P_{\theta_0}[g_n \geq h_n | g_{n-k} < h_{n-k}, \dots, g_{n-1} < h_{n-1}] = \alpha,$$

pour k fixé.

Il s'agit d'une approximation qui, dans les simulations que nous avons pu mener, s'est avéré sans grande conséquence. Le CUSUM possède une propriété intéressante dans le cas de données indépendantes : il "oublie" tout le passé dès lors qu'il devient négatif (voir son écriture récursive), et il a tendance à être négatif sous H_0 (du moins, les incréments $Z_{i,j}$ sont négatifs en moyenne). Pour des données dépendantes, ce n'est plus vrai puisque à l'instant n , la vraisemblance conditionnelle dépend de tout le passé. Cependant, pour des procédés markoviens d'ordre fini, les observations ne dépendent par des plus anciennes valeurs et donc la remarque effectuée pour des données indépendantes est à nouveau valable. Pour des modèles plus généraux, il y a une dépendance et donc une erreur dans l'estimation du seuil. La fenêtre k doit alors être choisie suffisamment grande pour rendre l'erreur d'estimation très faible.

3.3.2 Propriétés de convergence

A l'étape 1, nous estimons en réalité le quantile de la loi conditionnelle de g_n sachant que $g_1 < \hat{h}_1^M, \dots, g_{n-1} < \hat{h}_{n-1}^M$. Il faut donc s'assurer qu'asymptotiquement, le quantile que nous estimons converge bien vers le quantile cherché, autrement dit que la contrainte (3.2) est bien vérifiée.

Choix de l'estimateur

Il existe de nombreuses méthodes pour estimer les quantiles d'une distribution. On peut par exemple utiliser le quantile empirique.

Définition 3.1 Soit Y_1, \dots, Y_M un échantillon de même loi μ sur \mathbb{R} de fonction de répartition F .

1- On appelle fonction de répartition empirique associée à Y_1, \dots, Y_M , la fonction aléatoire $F_M : \mathbb{R} \mapsto [0; 1]$ définie pour tout y par :

$$F_M(y) = \frac{1}{M} \sum_{i=1}^M \mathbb{I}_{\{Y_i \leq y\}}.$$

2- Pour tout $\beta \in]0; 1[$ le quantile empirique d'ordre $1 - \beta$ est par définition $Y_{([M\beta])}$ où $[M\beta]$ est la partie entière de $M\beta$ et $Y_{(1)}, \dots, Y_{(M)}$ sont les statistiques d'ordre de l'échantillon.

Théorème 3.1 (Convergence des quantiles empiriques) Soit Y_1, \dots, Y_M un échantillon de même loi μ sur \mathbb{R} de fonction de répartition F . Si F est continue, d'inverse généralisée F^{-1} , alors pour tout β dans $]0; 1[$, en notant $Q_M(\beta) := Y_{([M\beta])}$ et $Q(\beta) = F^{-1}(\beta)$, on a :

$$Q_M(\beta) \rightarrow Q(\beta) \text{ p.s. lorsque } M \rightarrow \infty$$

Le théorème précédent assure donc la convergence du quantile empirique vers le véritable quantile.

Le problème

Pour déterminer le seuil h_n , le quantile empirique \bar{h}_n^M de la distribution g_n sachant que $g_1 < h_1, \dots, g_{n-1} < h_{n-1}$ est donc consistant (Théorème précédent). Le problème est que notre façon d'utiliser l'échantillon d'apprentissage, notamment le ré-échantillonnage, ne nous permet pas de calculer le quantile empirique de cette distribution mais plutôt celui de g_n sachant que $g_1 < \hat{h}_1^M, \dots, g_{n-1} < \hat{h}_{n-1}^M$.

Théorème 3.2 On a :

$$\lim_{M \rightarrow \infty} P_{\theta_0}[g_n \geq h_n | g_1 < \hat{h}_1^M, \dots, g_{n-1} < \hat{h}_{n-1}^M] = P_{\theta_0}[g_n \geq h_n | g_1 < h_1, \dots, g_{n-1} < h_{n-1}].$$

Ainsi, lorsque M tend vers l'infini, le quantile d'ordre α de g_n sachant $\{g_1 < \hat{h}_1^M, \dots, g_{n-1} < \hat{h}_{n-1}^M\}$ est égal au quantile d'ordre α de g_n sachant $\{g_1 < h_1, \dots, g_{n-1} < h_{n-1}\}$. Donc, l'estimation \hat{h}_n^M vérifie bien la contrainte (3.2) asymptotiquement.

Démonstration :

Nous allons procéder par récurrence :

- n=2

On a :

$$P_{\theta_0}[g_2 \geq h_2 | g_1 < \hat{h}_1^M] = \frac{P_{\theta_0}[g_2 \geq h_2 \cap g_1 < \hat{h}_1^M]}{P_{\theta_0}[g_1 < \hat{h}_1^M]}$$

- Le dénominateur :

$$\begin{aligned}
 & \hat{h}_1^M \xrightarrow{\text{p.s.}} h_1 \\
 \Rightarrow & g_1 - \hat{h}_1^M \xrightarrow{\text{p.s.}} g_1 - h_1 \\
 \Rightarrow & g_1 - \hat{h}_1^M \xrightarrow{\mathcal{L}} g_1 - h_1 \\
 \Rightarrow & P_{\theta_0}[g_1 - \hat{h}_1^M < 0] \xrightarrow{M \rightarrow \infty} P_{\theta_0}[g_1 - h_1 < 0] \\
 \Rightarrow & P_{\theta_0}[g_1 < \hat{h}_1^M] \xrightarrow{M \rightarrow \infty} P_{\theta_0}[g_1 < h_1].
 \end{aligned}$$

• Le numérateur :

On a

$$g_1 - \hat{h}_1^M \xrightarrow{\text{p.s.}} g_1 - h_1,$$

donc, on obtient

$$\begin{aligned}
 & (g_2 - h_2; g_1 - \hat{h}_1^M) \xrightarrow{\mathcal{L}} (g_2 - h_2; g_1 - h_1) \\
 \Rightarrow & P_{\theta_0}[g_2 \geq h_2 \cap g_1 < \hat{h}_1^M] \xrightarrow{M \rightarrow \infty} P_{\theta_0}[g_2 \geq h_2 \cap g_1 < h_1].
 \end{aligned}$$

Autrement dit,

$$\lim_{M \rightarrow \infty} P_{\theta_0}[g_2 \geq h_2 | g_1 < \hat{h}_1^M] = P_{\theta_0}[g_2 \geq h_2 | g_1 < h_1],$$

et donc le quantile empirique \hat{h}_2^M obtenue grâce à la distribution de g_2 sachant $g_1 < \hat{h}_1^M$ tend vers le quantile de la distribution de g_2 sachant $g_1 < h_1$:

$$\hat{h}_2^M \xrightarrow{\text{p.s.}} h_2 .$$

- n quelconque

Hypothèse de récurrence :

$$\hat{h}_1^M \xrightarrow{\text{p.s.}} h_1, \hat{h}_2^M \xrightarrow{\text{p.s.}} h_2, \dots, \hat{h}_{n-1}^M \xrightarrow{\text{p.s.}} h_{n-1}$$

De part l'hypothèse de récurrence,

$$g_1 - \hat{h}_1^M \xrightarrow{\text{p.s.}} g_1 - h_1, g_2 - \hat{h}_2^M \xrightarrow{\text{p.s.}} g_2 - h_2, \dots, g_{n-1} - \hat{h}_{n-1}^M \xrightarrow{\text{p.s.}} g_{n-1} - h_{n-1}$$

ce qui entraîne que

$$(g_1 - \hat{h}_1^M, g_2 - \hat{h}_2^M, \dots, g_{n-1} - \hat{h}_{n-1}^M) \xrightarrow{\mathcal{L}} (g_1 - h_1, g_2 - h_2, \dots, g_{n-1} - h_{n-1})$$

et

$$(g_n - h_n, g_1 - \hat{h}_1^M, \dots, g_{n-1} - \hat{h}_{n-1}^M) \xrightarrow{\mathcal{L}} (g_n - h_n, g_1 - h_1, \dots, g_{n-1} - h_{n-1}),$$

et ainsi

$$\hat{h}_n^M \xrightarrow{\text{p.s.}} h_n . \quad \square$$

3.3.3 Simulations

Nous allons appliquer, en simulation, l'algorithme du seuil adaptatif, tout d'abord sur un modèle stationnaire et nous le comparerons avec une méthode fournissant un seuil fixe, et ensuite nous l'appliquerons sur un modèle fortement non stationnaire.

Modèle stationnaire

On considère le modèle AR suivant :

$$X_n = 0.9X_{n-1} + \theta + \epsilon_n,$$

où X représente les observations, et ϵ_n est un bruit blanc gaussien de variance $\sigma^2 = 0.008$. Le changement de paramètre considéré est le suivant : sous H_0 , $\theta = \theta_0 = 0.3$ et sous H_1 , $\theta = \theta_1 = 0.38$

Nous avons mis en place le seuil adaptatif (avec $M = 5000$ trajectoires simulées sous H_0) avec la contrainte suivante :

$$P_{\theta_0}[g_n \geq h_n | g_1 < h_1, \dots, g_{n-1} < h_{n-1}] = 1/100,$$

et un seuil fixe est calculé à partir de la méthode présentée dans la partie 3.2 où la valeur du seuil est estimée par (3.1). Pour ce seuil fixe, la contrainte à vérifier est la suivante :

$$E_{\theta_0}(t_a) = 100.$$

Les deux seuils obtenus sont du même niveau compte tenu du résultat (3.4).

Sur la figure 3.1, on remarque que le seuil adaptatif et le seuil fixe ont approximativement les mêmes valeurs. Cela confirme le résultat (3.4), et le fait qu'un seuil fixe est adapté pour un processus stationnaire, ou plus précisément pour un processus où la statistique de test de la règle de décision utilisée, a approximativement la même loi tout au long de l'expérience.

REMARQUE : Cette simulation indique donc que dans le cas d'un processus stationnaire, l'approximation qui consiste à dire que le temps moyen entre deux fausses alarmes est égal au temps moyen avant la première fausse alarme est raisonnable. Ce n'est pas le cas pour la simulation suivante.

Modèle non stationnaire

Nous nous sommes intéressés au modèle suivant :

$$X_n = \theta X_{n-1} \cdot \cos(0.02n) + 0.5 + \epsilon_n, \tag{3.5}$$

avec ϵ_n un bruit blanc gaussien de variance $\sigma^2 = 2.10^{-4}$. La valeur nominale du paramètre θ est $\theta_0 = 0.5$ et la valeur en régime de panne est : $\theta_1 = 0.4$.

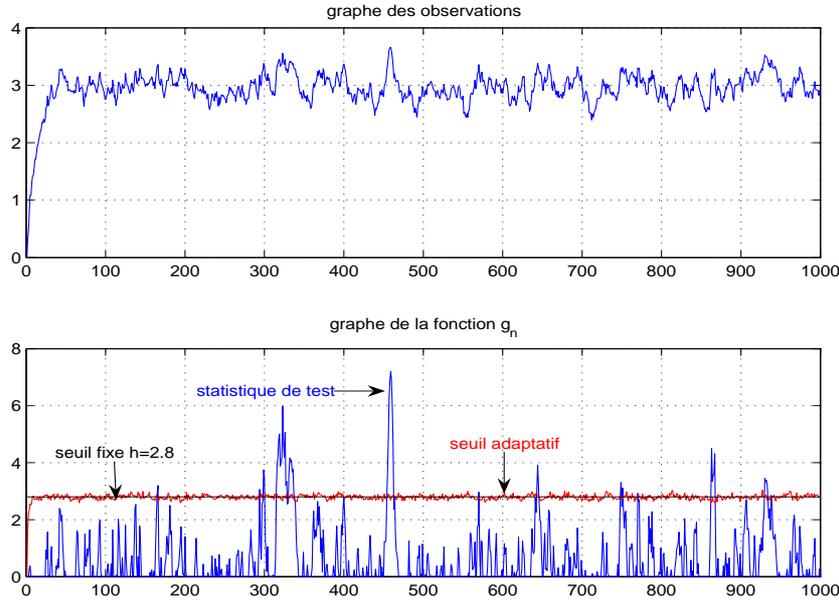


FIG. 3.1 – Test CUSUM appliqué aux données du modèle AR sous l’hypothèse H_0 , seuil fixe et seuil adaptatif, $M=5000$.

Le seuil adaptatif est calculé à partir de $M = 5000$ trajectoires sous H_0 du modèle (3.5) avec une probabilité de fausse alarme égale à $\alpha = 0.01$ (le seuil fixe basé sur le temps moyen avant la première fausse alarme ($\gamma = 100$) vaut ici 2.5).

Le premier graphique de la figure 3.2 représente l’évolution des observations sous régime H_0 . Le deuxième graphique donne la statistique de test correspondante et le seuil adaptatif. On voit que l’ordre de grandeur de la statistique de test varie fortement au cours du temps, de façon périodique à cause du cosinus. Dans une telle situation, l’utilisation d’un seuil fixe est clairement inapproprié. Au contraire, le seuil adaptatif est parfaitement ajusté aux données et permet de maintenir constant le taux moyen de fausses alarmes tout au long du processus.

Dans la figure 3.3, nous avons représenté une trajectoire du modèle (3.5) avec plusieurs pannes successives. La règle du CUSUM est appliquée en supposant à nouveau que $\theta_1 = 0.4$. A l’instant $t = 550$, un panne est simulé en prenant comme valeur de paramètre sous H_1 , $\theta'_1 = \theta_1 = 0.4$. A partir de $t = 590$, on revient à un régime de fonctionnement normal jusqu’à $t = 700$, où une panne est simulé avec $\theta''_1 = 0.35$ pendant 40 pas de temps. Enfin, un troisième panne est simulé ($\theta'''_1 = 0.44$) entre les instant 850 et 890. Pour chaque panne, la statistique de test franchit significativement le seuil (aux instants $t = 564$, $t = 715$ et $t = 883$). La détection a donc bien lieu même lorsque il y a une légère différence entre la valeur du paramètre utilisée dans le calcul du CUSUM et celle utilisée pour simuler la

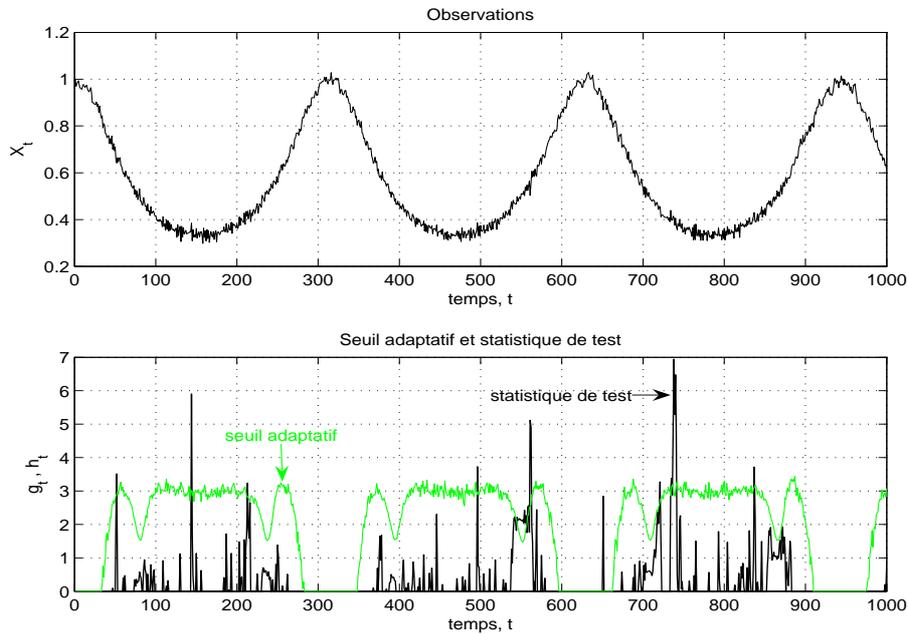


FIG. 3.2 – Test CUSUM appliqué au modèle (3.5) : le seuil adaptatif s'adapte à la distribution de la statistique de test

trajectoire (les deux pannes des instants $t = 700$ et $t = 850$).

Erreur d'estimation

Evidemment, si on relance l'algorithme de calcul du seuil plusieurs fois, nous obtiendrons à chaque fois des seuils différents pour un même instant t , à cause des erreurs d'estimation. Il est donc important de mesurer ces erreurs. La figure 3.4 représente un intervalle de confiance empirique de niveau 95%, obtenu en appliquant 2000 fois l'algorithme du seuil adaptatif. Cet intervalle de confiance est relativement étroit, signe que l'erreur d'estimation est faible pour un niveau de fausse alarme $\alpha = 0.01$ et $M = 5000$. Il est clair que plus le niveau α sera faible et plus il sera nécessaire de prendre un valeur M importante pour limiter l'erreur d'estimation.

Cet exemple basique montre la pertinence de la méthode adaptative qui permet de contrôler le taux de fausses alarmes comme souhaité.

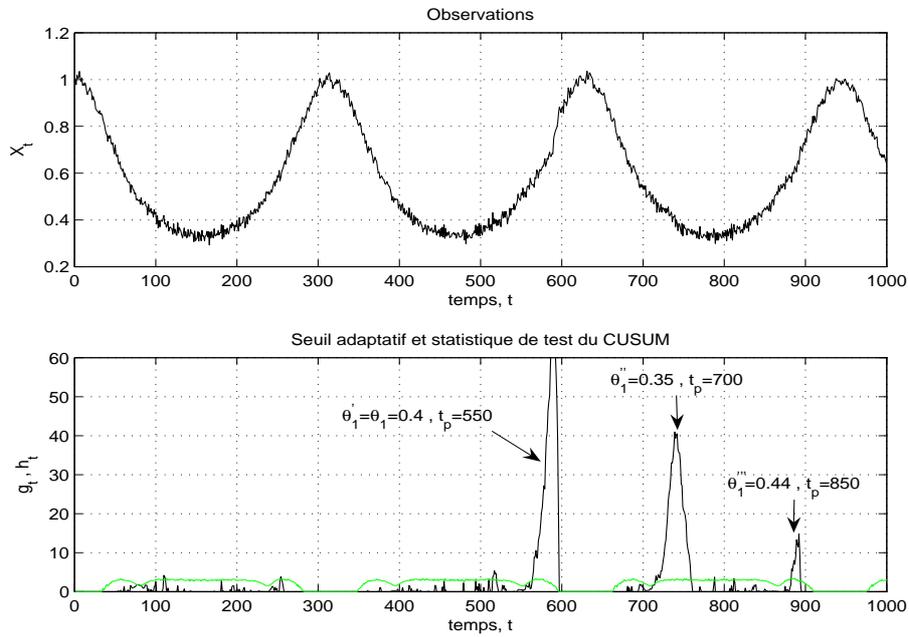


FIG. 3.3 – Test CUSUM appliqué au modèle (3.5) avec $\theta_1 = 0.4$. Trois pannes simulées avec $\theta_1' = 0.4$ ($t_p = 550$), $\theta_1'' = 0.35$ ($t_p = 700$) et $\theta_1''' = 0.44$ ($t_p = 850$)

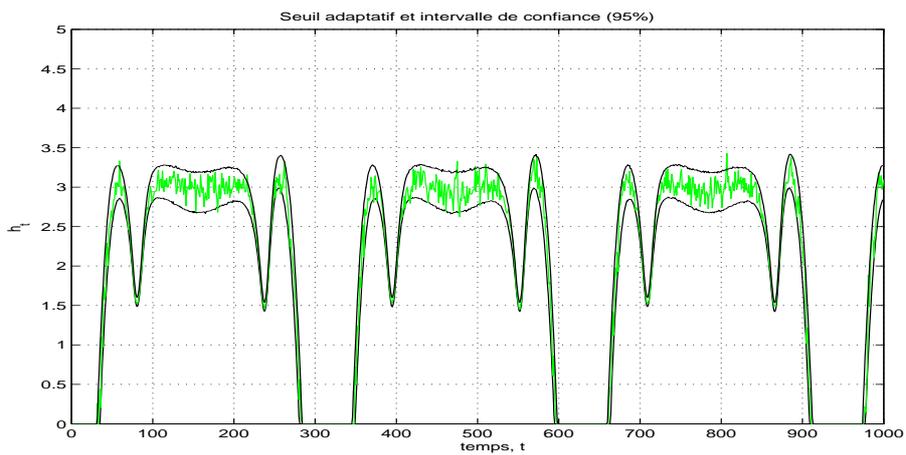


FIG. 3.4 – Seuil adaptatif pour le modèle (3.5) et intervalle de confiance de ce seuil à 95%

3.4 Seuil adaptatif dans le cas du diagnostic

Dans les parties précédentes, nous nous sommes intéressés au problème de la détection de changement et nous avons proposé une méthode efficace, au vue des premières simulations, pour calculer le seuil de règles de décision construites de la même façon que celle du CUSUM. Mais comme nous l'avons signalé dans les chapitres précédents, il est important de pouvoir adapter les méthodes de détection au cas du diagnostic de panne, où il est alors nécessaire de détecter un changement et de l'identifier parmi un ensemble de changements possibles. C'est la raison pour laquelle nous étendons dans cette partie l'approche du seuil adaptatif à un algorithme de diagnostic de changement, à savoir la règle de Nikiforov [73, 74], présentée dans le chapitre 2. Nous nous plaçons dans cette partie dans la deuxième façon de génération de changement. Nous reviendrons sur les algorithmes de diagnostic dans le cas de la première méthode de génération de changement dans le chapitre 6.

Comme dans le cas de la détection, le choix du seuil va s'effectuer en fonction de contraintes fixées par l'expérimentateur, concernant le taux moyen de fausses alarmes mais aussi la probabilité de mal localiser un changement.

Pour des raisons de présentation et de compréhension, nous nous limitons ici au cas où il n'y a seulement que deux types de changement possibles. Les trois hypothèses considérées sont alors les suivantes : H_0 pour le fonctionnement normal du système, H_1 pour le premier type de changement, et H_2 pour le second. Comme pour le problème de détection, la panne se traduit par un changement soudain du paramètre θ caractérisant la panne.

3.4.1 L'algorithme et les contraintes initiales

L'algorithme de Nikiforov [73, 74] a été initialement développé pour des données indépendantes. C'est en tout cas dans ce cas de figure que des propriétés d'optimalité ont été obtenues (voir chapitre 2). Cependant, comme nous l'avons vu dans les parties précédentes, l'intérêt du seuil adaptatif se situe dans sa capacité à traiter les processus non stationnaires. C'est pourquoi nous considérons l'algorithme de diagnostic dans un cadre de données dépendantes en utilisant les densités conditionnelles des observations.

Rappelons alors que cet algorithme met en jeu 4 statistiques de test dans le cas de deux hypothèses alternatives :

$$\forall n \geq 1 \quad g_n(1, 0), \quad g_n(1, 0) - g_n(2, 0), \quad g_n(2, 0) \quad \text{et} \quad g_n(2, 0) - g_n(1, 0)$$

$$\text{avec} \quad g_n(i, 0) = (g_{n-1}(i, 0) + Z_n(i, 0))^+ \quad \text{et} \quad Z_n(i, 0) = \log \frac{p_{\theta_i}(X_n | X_{1:n-1})}{p_{\theta_0}(X_n | X_{1:n-1})},$$

pour $i = 1, 2$.

Chapitre 3. Seuils adaptatifs

A partir des quantités suivantes :

$$N_1 = \inf\{n \geq 1 : g_n(1, 0) \geq h_d \text{ et } g_n(1, 0) - g_n(2, 0) \geq h_l\} \quad (3.6)$$

$$N_2 = \inf\{n \geq 1 : g_n(2, 0) \geq h_d \text{ et } g_n(2, 0) - g_n(1, 0) \geq h_l\} \quad (3.7)$$

avec $h_d \geq h_l$, on peut définir le temps d'arrêt et le type de panne :

Temps d'alarme

Le temps d'alarme vaut :

$$t_a = \min\{N_1, N_2\} \quad (3.8)$$

Type de panne

Soit ν le type de panne (1 ou 2) :

$$\nu = \arg \min\{N_1, N_2\} \quad (3.9)$$

Nikiforov [73] considère les contraintes suivantes :

$$- E_{\theta_0}(t_a) \geq \gamma, \quad (3.10)$$

$$- \max_{1 \leq j \neq l \leq 2} \sup_{t_p \geq 1} P_{(t_p)}^l(\nu = j | t_a \geq t_p) \leq b, \quad (3.11)$$

où $P_{(t_p)}^l$, $l = 1, 2$ représente la mesure de probabilité d'une distribution qui passe sous H_l au pas de temps t_p . La première contrainte indique que le temps moyen entre deux fausses alarmes est supérieur à γ . Il s'agit de la contrainte classique sur l'ARL. La deuxième impose que la pire probabilité de mal localiser une panne (suivant le type de panne et le moment où elle a lieu) est inférieure à b . L'expérimentateur fixe donc les valeurs γ et b et il faut déterminer les seuils h_d et h_l qui vont vérifier les contraintes précédentes.

Comme dans le cas du seuil adaptatif mono-panne, nous allons choisir de nouvelles contraintes permettant de calculer aisément les seuils à partir d'un grand nombre de simulations. Pour légitimer notre choix de contraintes, nous établirons un lien avec celles proposées par Nikiforov.

Dans l'algorithme présenté ci-dessus, le même seuil de détection h_d est comparé à deux statistiques de test : $g_n(1, 0)$ et $g_n(2, 0)$. Pourtant ces deux quantités n'ont pas de raison d'avoir le même ordre de grandeur et les comparer à un même seuil ne semble donc pas être la solution la mieux adaptée. C'est pourquoi il nous paraît intéressant de considérer, puisqu'il y a quatre statistiques de test, quatre seuils au lieu de deux. Chaque seuil sera donc évalué à partir de la distribution de la statistique de test correspondante.

Compte tenu de la forme du test et du choix de contraintes de Nikiforov [73], nous avons choisi les 4 contraintes suivantes :

$$\forall n \geq 1, \quad P_{\theta_0}[g_n(1, 0) \geq h_{10}^n \cap g_n(1, 0) - g_n(2, 0) \geq h_{12}^n | t_a \geq n] \leq \alpha_1 \quad (3.12)$$

$$P_{\theta_0}[g_n(2, 0) \geq h_{20}^n \cap g_n(2, 0) - g_n(1, 0) \geq h_{21}^n | t_a \geq n] \leq \alpha_2 \quad (3.13)$$

$$\sup_{1 \leq t_p \leq n} P_{(t_p)}^2[g_n(1, 0) \geq h_{10}^n \cap g_n(1, 0) - g_n(2, 0) \geq h_{12}^n | t_a \geq n] \leq \beta_1 \quad (3.14)$$

$$\sup_{1 \leq t_p \leq n} P_{(t_p)}^1[g_n(2, 0) \geq h_{20}^n \cap g_n(2, 0) - g_n(1, 0) \geq h_{21}^n | t_a \geq n] \leq \beta_2 \quad (3.15)$$

où $h_{10}^n \geq h_{12}^n > 0$, $h_{20}^n \geq h_{21}^n > 0$. Les 4 paramètres α_1 , α_2 , β_1 et β_2 sont choisis par l'expérimentateur. Ces contraintes permettent de contrôler le taux de fausses alarmes et la probabilité de faux diagnostic. L'équation (3.12) impose qu'à l'instant n , les seuils h_{10}^n et h_{12}^n soient choisis de sorte que la probabilité d'une fausse alarme, en ayant localisé un panne de type 1, soit inférieure à α_1 . L'équation (3.13) est l'analogie de (3.12) pour une panne de type 2. Enfin, (3.14) et (3.15) concernent la probabilité de faux diagnostic à l'instant n . Nous prenons le supremum sur l'instant de panne t_p pour éviter une dépendance de cette probabilité avec l'instant de changement.

Comme dans le problème de détection, il est aisé de faire le lien entre les contraintes (3.12) à (3.15) et (3.10), (3.11). Si on note α la probabilité d'avoir une fausse alarme à un instant donné, alors $\alpha \leq \alpha_1 + \alpha_2$, et :

$$\begin{aligned} P_{\theta_0}[t_a = n] &= P_{\theta_0}[t_a = n | t_a \geq n] \cdot P_{\theta_0}[t_a \geq n] \\ &= P_{\theta_0}[t_a = n | t_a \neq 1, \dots, t_a \neq n-1] \cdot P_{\theta_0}[t_a \neq 1, \dots, t_a \neq n-1] \\ &= \dots \\ &= P_{\theta_0}[t_a = n | t_a \geq n] \cdot P_{\theta_0}[t_a \neq n-1 | t_a \geq n-1] \\ &\quad \dots P_{\theta_0}[t_a \neq 2 | t_a \geq 2] \cdot P_{\theta_0}[t_a \neq 1] \\ &= \alpha(1 - \alpha)^{n-1} \end{aligned}$$

Nous obtenons :

$$E_{\theta_0}[t_a] = \frac{1}{\alpha}. \quad (3.16)$$

Ainsi, si les contraintes (3.12) et (3.13) sont vérifiées, nous avons $E_{\theta_0}[t_a] \geq 1/(\alpha_1 + \alpha_2)$.

A partir des contraintes (3.14) et (3.15), nous obtenons l'inégalité suivante :

$$\max_{1 \leq j \neq l \leq 2} \sup_{t_p \geq 1} P_{t_p}^l(\nu = j | t_a \geq t_p) \leq \max(\beta_1, \beta_2). \quad (3.17)$$

Chapitre 3. Seuils adaptatifs

Les contraintes (3.12) à (3.15) sont donc plus restrictives que (3.10) et (3.11). Mais leur principal avantage est qu'elles permettent une plus grande liberté dans le choix des probabilités de fausse détection et de mauvaise localisation suivant le type de panne. Ainsi, s'il est par exemple plus dangereux pour le système sous surveillance de mal diagnostiquer une panne de type 1 plutôt qu'une panne de type 2, on peut choisir une plus petite valeur pour β_2 que pour β_1 .

*
* *

Avant de passer à l'estimation des seuils, on peut évoquer le problème où le nombre d'hypothèses alternatives K est supérieur à deux. Dans ce cas général, l'algorithme de diagnostic de Nikiforov a été présenté dans la partie 2.7.1. Il met en jeu K^2 statistiques de test :

$$\forall l = 1, \dots, K \text{ et } \forall 0 \leq j \leq K, j \neq l, \text{ test de } H_l \text{ contre } H_j.$$

Il est alors nécessaire de considérer K^2 seuils h_{lj} et on peut utiliser les K^2 contraintes suivantes : $\forall n \geq 1$,

$$\forall l = 1, \dots, K$$

$$P_{\theta_0} \left(\min_{0 \leq j \neq l \leq K} [g_n(l) - g_n(j) - h_{lj}^n] \geq 0 | t_a \geq n \right) \leq \alpha_l, \quad (3.18)$$

et

$$\forall l = 1, \dots, K, \forall j = 1, \dots, K, j \neq l$$

$$\sup_{1 \leq t_p \leq n} P_{(t_p)}^l \left(\min_{0 \leq j \neq l \leq K} [g_n(l) - g_n(j) - h_{lj}^n] \geq 0 | t_a \geq n \right) \leq \beta_{lj} \quad (3.19)$$

avec $h_{l0}^n \geq h_{lj}^n > 0$. La méthode de calcul de seuils développée par la suite pour deux hypothèses alternatives peut théoriquement s'adapter à ce cas général. Cependant, on peut penser que le nombre élevé de contraintes et de seuils va rendre difficile son utilisation en-ligne.

3.4.2 Estimation des seuils

La méthode pour calculer les seuils est basée, comme dans la partie mono-panne, sur des simulations de trajectoires et des étapes de ré-échantillonnage. Cependant, dans le cas multi-panne, chaque contrainte met en jeu 2 seuils. Ainsi, à chaque pas de temps, nous allons déterminer des régions dans lesquelles les contraintes sont vérifiées. En prenant ensuite l'intersection des régions obtenues grâce aux équations (3.12) et (3.14), nous obtiendrons l'ensemble des couples (h_{10}^n, h_{12}^n) vérifiant les contraintes. Dans la suite, nous développons étape par étape, l'approche pour choisir un couple (h_{10}^n, h_{12}^n) pour $n \geq 2$. Les seuils (h_{20}^n, h_{21}^n) sont obtenus de façon identique.

Equation (3.12)

A l'instant n , nous considérons M trajectoires (X_1^m, \dots, X_n^m) , $m = 1, \dots, M$ sous H_0 n'ayant pas eu de fausse alarme. Pour une valeur fixé α_1 , nous devons déterminer l'ensemble :

$$\mathcal{A}_1 = \left\{ (h_{10}^n, h_{12}^n) : P_{\theta_0}[g_n(1, 0) \geq h_{10}^n \cap g_n(1, 0) - g_n(2, 0) \geq h_{12}^n | t_a \geq n] \leq \alpha_1 \right\}.$$

Tout d'abord, nous nous intéressons à la frontière de cet ensemble :

$$\partial \mathcal{A}_1 = \left\{ (h_{10}^n, h_{12}^n) : P_{\theta_0}[g_n(1, 0) \geq h_{10}^n \cap g_n(1, 0) - g_n(2, 0) \geq h_{12}^n | t_a \geq n] = \alpha_1 \right\}.$$

Nous avons

$$\begin{aligned} & P_{\theta_0}[g_n(1, 0) \geq h_{10}^n \cap g_n(1, 0) - g_n(2, 0) \geq h_{12}^n | t_a \geq n] = \alpha_1 \\ \Leftrightarrow & P_{\theta_0}[g_n(1, 0) \geq h_{10}^n | g_n(1, 0) - g_n(2, 0) \geq h_{12}^n, t_a \geq n] \\ & \quad \cdot P_{\theta_0}[g_n(1, 0) - g_n(2, 0) \geq h_{12}^n | t_a \geq n] = \alpha_1. \end{aligned}$$

Fixons une valeur pour h_{12}^n . La probabilité suivante :

$$P_{\theta_0}[g_n(1, 0) - g_n(2, 0) \geq h_{12}^n | t_a \geq n]$$

est estimée en appliquant l'algorithme de détection-localisation à chaque trajectoire. Notons $\hat{\alpha}$ cet estimateur. $\hat{\alpha}$ est la proportion de valeurs $g_n^m(1, 0) - g_n^m(2, 0)$ supérieures à h_{12}^n parmi les M trajectoires. Ainsi, nous obtenons :

$$\hat{P}_{\theta_0}[g_n(1, 0) \geq h_{10}^n | g_n(1, 0) - g_n(2, 0) \geq h_{12}^n, t_a \geq n] = \frac{\alpha_1}{\hat{\alpha}}, \quad (3.20)$$

ce qui implique que h_{10}^n peut être estimé par le quantile conditionnel d'ordre $(\alpha_1/\hat{\alpha})$ de la loi conditionnelle de $g_n(1, 0)$. Nous pouvons estimer ce quantile en sélectionnant, dans l'échantillon d'apprentissage, les trajectoires respectant la condition qui apparaît dans la probabilité (3.20) et en procédant comme dans la partie 3.3.1. Evidemment, le quantile conditionnel ci-dessus n'a de sens que si $\hat{\alpha}$ est plus grand que α_1 . Si ce n'est pas le cas, alors il est impossible de trouver une valeur \hat{h}_{10}^n telle que le couple $(\hat{h}_{10}^n, h_{12}^n)$ appartient à la frontière $\partial \mathcal{A}_1$.

Pour une valeur h_{12}^n , nous obtenons donc une valeur \hat{h}_{10}^n . En faisant varier petit à petit la première valeur, nous obtenons l'ensemble $\partial \mathcal{A}_1$ et par conséquent, la région \mathcal{A}_1 , qui est aussi délimitée par la droite d'équation $h_{10} = h_{12}$ (figure 3.5).

REMARQUE : Pour estimer le quantile d'ordre $(\alpha_1/\hat{\alpha})$ de la variable aléatoire $g_n(1, 0)$, on utilise les trajectoires vérifiant : $g_n(1, 0) - g_n(2, 0) \geq h_{12}^n$. Autrement dit, on utilise seulement un pourcentage $\hat{\alpha}$ des M trajectoires initiales, et $\hat{\alpha}$ peut être petit. L'échantillon de départ doit donc ici être beaucoup plus important que dans le cas mono-panne, où en simulation, un échantillon de 5000 trajectoires donne déjà de bons résultats (partie 3.3.3).

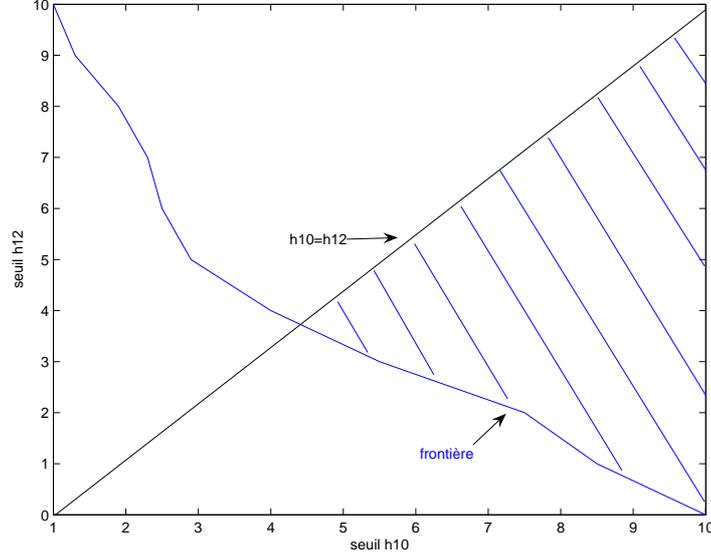


FIG. 3.5 – L'ensemble \mathcal{A}_1 des couples vérifiant (3.12) est représenté par la zone hachurée

Equation (3.14)

Nous allons procéder comme dans la partie précédente. A partir d'une valeur fixée pour h_{12}^n , nous obtenons une valeur h_{10}^n telle que la contrainte (3.14) est vérifiée. Mais la situation est à présent plus complexe parce que dans cette contrainte, le supremum est pris sur l'instant de changement. Ainsi, l'équation (3.14) n'est pas vraiment une contrainte mais plutôt n contraintes, chacune reliée à un instant de changement compris entre 1 et n . Il ne faut donc plus considérer seulement un échantillon d'apprentissage mais n échantillons, et il faut déterminer la frontière de chaque région associée à un temps de changement possible. La région associée à la contrainte (3.14) sera l'intersection de toutes les régions obtenues. Il est clair qu'en procédant de la sorte, la quantité de calculs va augmenter trop rapidement pour une utilisation de cette méthode en pratique (et surtout en ligne). Il faut donc simplifier le problème et une solution consiste à remplacer la contrainte (3.14) par :

$$\sup_{n-k \leq t_p \leq n} P_{(t_p)}^2 [g_n(1, 0) \geq h_{10}^n \cap g_n(1, 0) - g_n(2, 0) \geq h_{12}^n | t_a \geq n] \leq \beta_1. \quad (3.21)$$

L'utilisation d'une fenêtre nous permet de considérer, non plus n échantillons d'apprentissage, mais seulement $k + 1$.

Pour comprendre la raison de cette simplification, revenons à la description de l'algorithme et aux propriétés basiques de la règle du CUSUM. Sous l'hypothèse

H_0 , les incréments $Z_n(l, 0)$ sont négatifs en moyenne, ce qui implique que $g_n(1, 0)$ et $g_n(2, 0)$ sont proches de zéro, tout comme la quantité $g_n(1, 0) - g_n(2, 0)$ et son opposée. A partir de l'instant t_p , lorsque le système passe en régime de panne, par exemple sous H_1 , il est aisé de montrer que $E_{\theta_1}(s_n(1, 0)) > E_{\theta_1}(s_n(2, 0))$, $\forall n \geq t_p$. La quantité $g_n(1, 0) - g_n(2, 0)$ va donc avoir tendance à augmenter en moyenne sous H_1 . Ainsi, il semble que la probabilité de mauvaise localisation (dont une condition nécessaire est que $g_n(2, 0) - g_n(1, 0)$ soit grand) est plus importante juste après l'instant de panne. C'est pourquoi la contrainte (3.21) impose un taux de faux diagnostic très proche de celui imposé par (3.14).

Choix des seuils h_{10}^n et h_{12}^n

Lorsqu'on a déterminé les deux régions correspondant à (3.12) et (3.21), il faut choisir les seuils dans l'intersection des deux régions. Evidemment, il est plus judicieux, du point de vue du retard à la détection, de choisir ces valeurs près de la frontière.

Transposition aux seuils h_{20}^n et h_{21}^n

La même approche est adoptée pour l'estimation de h_{20}^n et h_{21}^n à partir des contraintes (3.13) et (3.15). L'inégalité (3.15) étant plus exactement remplacée par la contrainte suivante :

$$\sup_{n-k \leq t_p \leq n} P_{(t_p)}^1 [g_n(2, 0) \geq h_{20}^n \cap g_n(2, 0) - g_n(1, 0) \geq h_{21}^n | t_a \geq n] \leq \beta_2.$$

Une fois les quatre seuils estimés à l'instant n , on re-échantillonne les trajectoires et on répète la procédure pour obtenir les seuils à l'instant $n + 1$.

3.4.3 Simulation

Nous allons appliquer notre méthode sur le modèle de dépollution biologique présenté dans les chapitres précédents :

$$\begin{cases} X_{n+1} = X_n + (TX_n)\mu(S_n) - T(\phi U_n)X_n + \epsilon_{n+1}^{(1)} \\ S_{n+1} = S_n - \frac{TX_n}{\tau}\mu(S_n) + (\phi U_n)(\psi S_{ini} - S_n)T + \epsilon_{n+1}^{(2)}. \end{cases} \quad (3.22)$$

Dans cette simulation, nous considérons deux pannes caractérisées par le paramètre $\theta = (\phi, \psi)$. La première représente le colmatage de la pompe d'alimentation. La valeur réelle de U_n sera plus petite que la valeur nominale imposée et le paramètre ϕ représente le taux de colmatage de la pompe. Le deuxième type de panne

Chapitre 3. Seuils adaptatifs

est une variation imprévisible de la concentration en substrat dans l'alimentation du bioréacteur par rapport à sa valeur nominale S_{ini} . Ainsi, sous H_0 , $\theta = \theta_0 = (1, 1)$ et les deux types de fautes sont modélisés par : sous H_1 , $\theta = \theta_1 = (0.84, 1)$ et sous H_2 , $\theta = \theta_1 = (1, 1.12)$.

REMARQUE : Dans cette simulation, nous avons supposé que les valeurs caractérisant les pannes sont connues. Il s'agit des conditions normales d'application de la règle du CUSUM. Pour être plus réaliste, on aurait pu supposer que les valeurs θ_1 et θ_2 étaient inconnues et on aurait alors appliqué une règle de décision de type GLR ou encore MLR, puisque le calcul des seuils adaptatifs peut être étendu à ce type de règle. De plus, nous avons appliqué la règle de décision à la seule variable S_n .

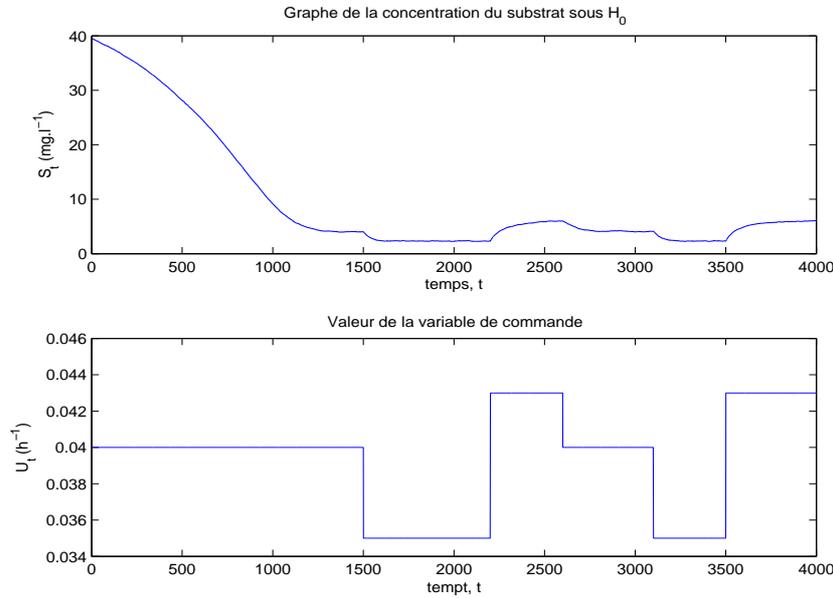


FIG. 3.6 – Graphe de la concentration du substrat et valeur de la variable de commande

La figure 3.6 représente la concentration en substrat sous H_0 et la valeur de la variable de commande U_n , que l'on fait varier au cours du temps.

On applique l'algorithme de détection-localisation à l'équation du substrat. Les valeurs des paramètres sont :

$$\alpha_1 = 0.005, \quad \alpha_2 = 0.005, \quad \beta_1 = 0.2 \quad \text{et} \quad \beta_2 = 0.2,$$

et on considère $M = 100000$ trajectoires simulées du modèle (3.22) pour le calcul des seuils, qui sont représentés dans la figure 3.7. Les valeurs de ces seuils varient au

cours du temps, en particulier en fonction de la valeur de la variable de commande. On voit donc, encore une fois, que l'utilisation de seuils adaptatifs est pertinente puisque ces derniers s'adaptent aux variations des distributions des statistiques de test dues à la variable de contrôle.

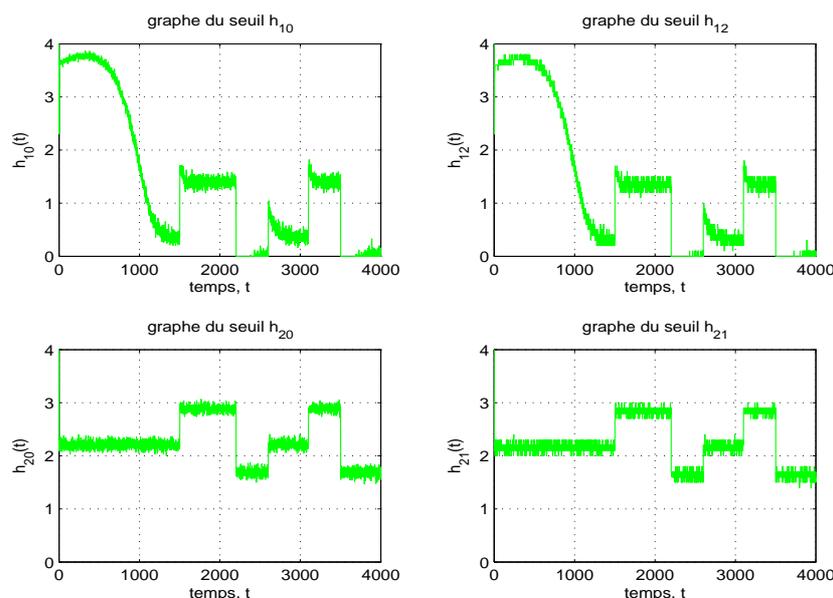


FIG. 3.7 – Graphes des 4 seuils adaptatifs de l'algorithme de détection-localisation

REMARQUE : Pour des raisons algorithmiques, nous avons choisi les couples de seuils près de la première bissectrice, ce qui explique pourquoi les valeurs h_{10} et h_{12} (idem pour h_{20} et h_{21}) sont similaires.

Dans la figure 3.8, nous avons représenté une trajectoire du modèle (3.22) et la règle de décision correspondante. Les deux régimes de fonctionnement anormal ont été simulés : entre les pas de temps 2000 et 2050 pour H_1 et entre 3400 et 3450 pour H_2 .

3.5 Conclusion

La méthode du seuil adaptatif paraît adaptée au type de modèle nous intéressant. Elle permet de calculer un seuil très facilement à partir de simulations et garantit un taux de fausses alarmes et une probabilité de mauvaise localisation constants tout au long du processus, ce qui n'est pas forcément le cas avec un seuil fixe. Pour des

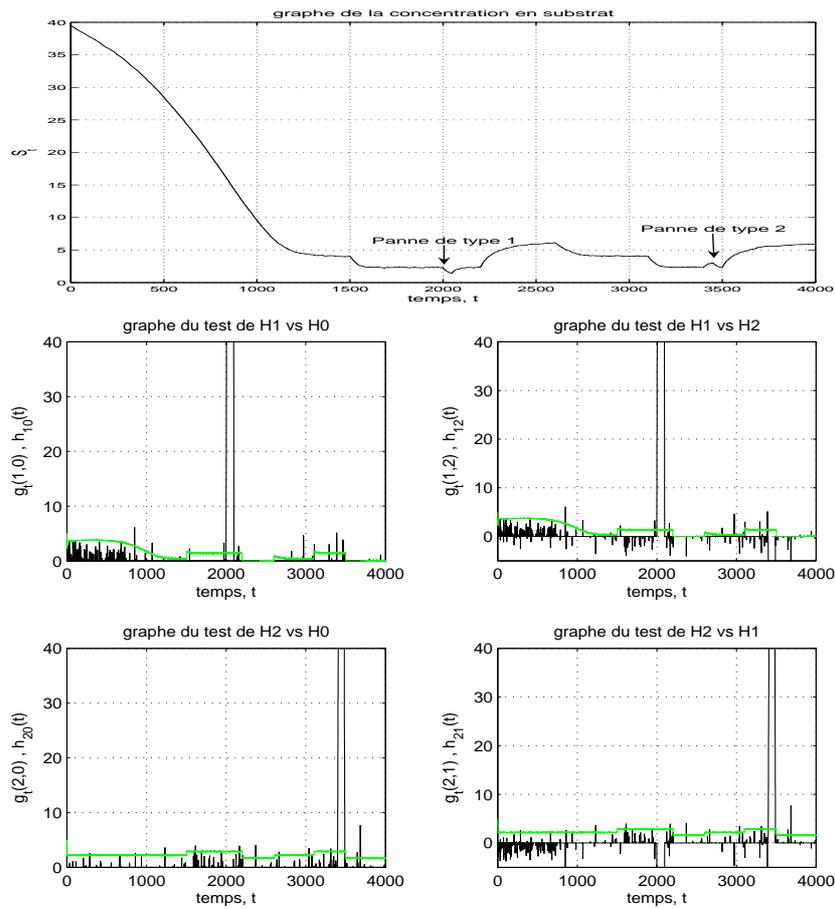


FIG. 3.8 – Règle de Nikiforov appliquée sur le substrat S_n avec deux types de panne : sous H_1 entre $t = 2000$ et $t = 2050$ et sous H_2 entre $t = 3400$ et $t = 3450$.

procédés de type dépollution biologique, qui sont généralement conduits grâce à une variable de contrôle, cela peut être une des solutions.

Nous sommes conscients des limites de notre approche. Estimer des quantiles associés à de faibles niveaux (0.01 ou 0.001) à partir de simulations n'est pas forcément évident. Notre objectif n'était pas d'obtenir des seuils très précis mais plutôt de bonnes approximations qui ne coûtent pas très cher en calculs et peuvent être obtenues en ligne. Malgré ça, les différentes simulations que nous avons pu faire ont donné de bons résultats. Dans le chapitre 7 sur le procédé de dépollution, le seuil adaptatif est calculé à chaque fois avec un niveau $\alpha = 0.0007$ pour répondre au cahier des charges du procédé de dépollution. Même pour ce faible niveau, le seuil obtenu est assez cohérent (pas trop "bruité"), signe que l'estimation du quantile est satisfaisante.

En ce qui concerne la quantité de calculs, la méthode dans le cas mono-panne est parfaitement adaptée et le calcul du seuil est quasiment instantané. Pour le diagnostic de panne, c'est un peu plus compliqué. Pour le cas de deux hypothèses alternatives, la méthode permet de traiter, avec un échantillon assez important de 100000 trajectoires, un pas de temps toute les deux secondes, avec une programmation sous Matlab sur un processeur à 2.6GHz. Dans de nombreux types de procédés, cette durée de calcul est acceptable, notamment pour les procédés de dépollution, où on peut compter plusieurs minutes entre chaque mesure. Théoriquement, on pourrait étendre la méthode à un nombre K d'hypothèses alternatives supérieur à 2. Mais la durée de calcul serait, à ce moment là, beaucoup plus importante puisqu'il y aurait alors, en utilisant la même méthode, K^2 contraintes à considérer. Un choix sur les contraintes les plus importantes aux yeux de l'utilisateur serait alors à envisager.

Chapitre 4

Une approximation de la règle du CUSUM

Dans les chapitres suivants, nous allons nous intéresser à deux cas régulièrement rencontrés dans les problèmes de dépollution biologique. Le premier va concerner les modèles dont une partie est inconnue ou plus simplement incertaine, comme cela peut être le cas pour le taux de croissance des bactéries. Le deuxième va concerner les systèmes indirectement observés, c'est à dire les modèles à espace d'état. Dans les deux cas, il n'est généralement pas possible d'écrire la densité conditionnelle des observations, donc il n'est pas possible de mettre en place les procédures classiques de détection d'anomalie, telles que les règles du CUSUM et du MLR.

Pour les deux situations que nous venons d'évoquer, notre approche va être similaire. Nous allons utiliser des outils, non paramétriques pour le premier cas et de type filtrage pour le deuxième, dans le but d'estimer les vraisemblances conditionnelles des observations de sorte à construire une règle sur le modèle du CUSUM mais en remplaçant les vraisemblances (sous H_0 et sous H_1) par leurs estimations. Nous parlerons alors de la règle du CUSUM approché.

Le plan de ce chapitre est le suivant : dans un premier temps, nous présentons la règle du CUSUM approché. La question des propriétés de cette règle se pose alors. Ainsi, nous revenons dans la deuxième partie sur des résultats d'optimalité pour des procédures classiques (dont le CUSUM), résultats que nous adaptons par la suite au CUSUM approché (partie 4.3 pour l'optimalité au sens de Lorden et partie 4.4 pour un critère d'optimalité plus souple).

4.1 Le “CUSUM approché”

Nous considérons ici un contexte général où le changement est généré suivant la première méthode présentée dans la partie 2.3.2. Ainsi,

$$\begin{aligned} X_n|X_{1:n-1} &\sim p_{\theta_0}(\cdot|X_{1:n-1}) && \text{si } n < t_p \\ \text{et } X_n|X_{1:n-1} &\sim p_{\theta_0, \theta_1, t_p}(\cdot|X_{1:n-1}) && \text{si } n \geq t_p. \end{aligned}$$

Nous nous plaçons dans le cas où les densités conditionnelles p_{θ_0} et $p_{\theta_0, \theta_1, j}$, $j = 1, \dots, n$ sont inaccessibles. Ces quantités sont inconnues mais on peut estimer les vraisemblances conditionnelles par :

$$\hat{l}_{0,n}^N \quad \text{et} \quad \hat{l}_{1,n,j}^N, \quad j = 1, \dots, n$$

où N est un entier qui conditionne l’approximation de la vraisemblance, et qui sera particularisé dans les chapitres suivants. Dans le chapitre 5 concernant l’estimation d’une fonction f inconnue, N correspond à la taille de l’échantillon d’apprentissage de l’estimateur à noyau. Dans le chapitre 6 où nous traitons les modèles à espace d’état, N correspond au nombre de particule du filtre.

A partir des estimations \hat{l} , nous allons construire une règle de décision sous la forme de la règle du CUSUM. On note :

$$\hat{Z}_{i,j}^N = \log \frac{\hat{l}_{1,i,j}^N}{\hat{l}_{0,i}^N},$$

et on définit le temps d’arrêt de la règle du “CUSUM approché” comme suit :

$$\hat{t} = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \hat{Z}_{i,j}^N \geq h \right\}. \quad (4.1)$$

Dans le chapitre 2, nous avons évoqué les nombreux résultats d’optimalité pour les règles classiques, et notamment pour celle du CUSUM. Une des premières questions essentielles est donc de savoir quelles sont les propriétés d’une règle de décision construite à partir des estimations des vraisemblances, et plus particulièrement, si l’optimalité est conservée, et sous quelles hypothèses. Pour cela, revenons un peu plus en détails sur ces résultats d’optimalité.

4.2 Optimalité de la règle du CUSUM

Généralement, l’optimalité d’une règle de décision s’obtient en deux étapes. D’une part, en considérant tout un ensemble de règles vérifiant une contrainte fixée,

on cherche à établir une borne inférieure pour le critère d'optimalité choisi. Et dans un second temps, on montre que la règle en question vérifie la contrainte, et atteint la borne inférieure.

Les premiers résultats d'optimalité pour la règle du CUSUM ont été obtenus par Lorden [62] dans un cadre asymptotique et pour des données indépendantes pour le critère du pire retard moyen à la détection $\bar{E}_{\theta_1}(\cdot)$ (défini en (2.11)), sous la contrainte ARL du temps moyen entre deux fausses alarmes. Toujours pour des données indépendantes, Moustakides [70] et Ritov [86] établissent l'optimalité dans un cadre non asymptotique.

Dans le cas de données dépendantes, l'étude des propriétés d'optimalité repose sur la stabilité du logarithme du rapport de vraisemblance. Tous les résultats d'optimalité proposés ces dernières années supposent donc plus ou moins fortement que la quantité $1/n \sum_{i=k}^{k+n-1} Z_i$ tend vers une constante I . Dans le cas de données indépendantes, cette hypothèse est vérifiée et I représente l'information de Kulbach-Leibler entre les distributions sous H_0 et H_1 .

Les premiers résultats d'optimalité pour la règle du CUSUM dans un cadre dépendant sont ceux de Bansal et Papantoni-Kazokos [3] qui considère tout de même des modèles assez restrictifs où il est nécessaire d'avoir indépendance entre les observations avant et après le temps de changement.

Nous allons maintenant nous intéresser aux résultats d'optimalité pour la règle CUSUM obtenus par Lai [50] puisque c'est à partir de ses travaux que nous allons étudier les propriétés de la règle du CUSUM approché. On considère un changement de paramètre dans la densité conditionnelle des observations qui est donc : $p_{\theta_0}(x_n|X_{1:n-1})$ sous H_0 et $p_{\theta_1}(x_n|X_{1:n-1})$ sous H_1 .

REMARQUE : Lai [50] se place dans le cas où la densité conditionnelle, sous régime de panne, ne dépend pas de l'instant d'apparition de la panne.

Sous l'hypothèse suivante :

$$\lim_{n \rightarrow \infty} \sup_{t_p \geq 1} \sup \text{ess } P^{(t_p)} \left\{ \max_{t \leq n} \sum_{i=t_p}^{t_p+t} Z_i \geq I(1 + \delta)n | X_{1:t_p-1} \right\} = 0 \quad \forall \delta > 0, \quad (4.2)$$

Lai [50] obtient la borne minimale du pire retard moyen à la détection :

$$\bar{E}_{\theta_1}(t_a) = \sup_{t_p \geq 1} \sup \text{ess } E^{(t_p)}[(t_a - t_p + 1)^+ | X_{1:t_p-1}], \quad (4.3)$$

pour toutes les règles t_a vérifiant la contrainte :

$$E_{\theta_0}(t_a) \geq \gamma, \quad (4.4)$$

c'est à dire parmi toutes les règles t_a ayant un temps moyen entre deux fausses alarmes supérieur à γ . Cette borne minimale est donnée par le théorème suivant :

Chapitre 4. Une approximation de la règle du CUSUM

Théorème 4.1 (Lai [50]) *Si l'hypothèse (4.2) est vérifiée, alors lorsque $\gamma \rightarrow \infty$, on a :*

$$\inf\{\bar{E}_{\theta_1}(t_a) : E_{\theta_0}(t_a) \geq \gamma\} \geq (I^{-1} + o(1)) \log \gamma. \quad (4.5)$$

Lai ([50] Théorème 4, ii)) montre ensuite que sous l'hypothèse suivante,

$$\lim_{n \rightarrow \infty} \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ n^{-1} \sum_{i=j}^{j+n} Z_i \leq I(1 - \delta) | X_{1:j-1} \right\} = 0, \quad \forall \delta > 0, \quad (4.6)$$

la règle du CUSUM atteint la borne inférieure (4.5) et est donc asymptotiquement optimale pour le critère $\bar{E}_{\theta_1}(\cdot)$ introduit par Lorden. On parle ainsi d'optimalité au sens de Lorden.

La règle du CUSUM à fenêtre limitée,

$$\tilde{t}_{CUSUM} = \inf \left\{ n : \max_{n-m_\alpha \leq j \leq n} \sum_{i=j}^n Z_i \geq h \right\}, \quad (4.7)$$

est, elle aussi, asymptotiquement optimale au sens de Lorden lorsque la taille de la fenêtre, m_α , tend vers l'infini.

Au vue de ces résultats, une question naturelle se pose. Quels sont les modèles qui vérifient les hypothèses (4.2) et (4.6) ? Pour des modèles autorégressifs de la forme $X_{n+1} = g(X_n, \theta) + \epsilon_n$, nous verrons dans le prochain chapitre que ces hypothèses sont vérifiées si le bruit est gaussien et si la fonction g vérifie de bonnes propriétés. Mais des auteurs comme Fuh [32], Mei [66] ou encore Tartakovsky [107] évoquent la difficulté de vérifier ces hypothèses en général.

Tartakovsky [107] établit l'optimalité asymptotique de l'approche bayésienne dont le temps d'arrêt est défini en (2.18). Les hypothèses de stabilité qu'il effectue sur le rapport de vraisemblance sont moins restrictives que les hypothèses (4.2) et (4.6) puisqu'il n'apparaît plus le supremum essentiel sur la trajectoire avant l'instant de panne. Et c'est justement, d'après Tartakovsky, le point gênant des hypothèses de Lai [50] puisque sans ce supremum essentiel, "les hypothèses sont alors vérifiables pour de nombreuses situations".

De la même façon, Lai [50] établit l'optimalité asymptotique du CUSUM à fenêtre limitée pour un critère plus souple que le pire retard moyen à la détection, le retard moyen :

$$E^{(t_p)}(t_a - t_p + 1)^+. \quad (4.8)$$

Lai montre que la règle du CUSUM à fenêtre limitée atteint la borne inférieure du critère (4.8) lorsqu'elle vérifie une contrainte plus sévère que la contrainte ARL (4.4), à savoir la contrainte sur la probabilité de fausse alarme suivante :

$$\sup_{t \geq 1} P_{\theta_0}(t \leq t_a < t + m_\alpha) \leq \alpha, \quad (4.9)$$

$$\text{où } \liminf \frac{m_\alpha}{|\log \alpha|} > I^{-1} \text{ mais } \log m_\alpha = o(\log \alpha) \text{ quand } \alpha \rightarrow \infty.$$

Il y a deux avantages à utiliser une contrainte plus sévère que la contrainte ARL : d'une part, le fait qu'il y ait un temps moyen fixé entre deux fausses alarmes n'entraîne pas forcément que la probabilité d'avoir une fausse alarme dans les tout premiers instants de la surveillance soit faible. La contrainte (4.9), tout comme la contrainte que nous avons utilisé lors de la construction du seuil adaptatif permet de contrôler la probabilité de fausse alarme (cette fois-ci sur un intervalle). D'autre part, la contrainte (4.9) associée au critère (4.8) permet d'assouplir l'hypothèse (4.2) sur la stabilité du rapport de vraisemblance puisque le supremum essentiel n'est plus nécessaire. L'hypothèse suivante est alors suffisante pour s'assurer de l'optimalité de la règle du CUSUM à fenêtre limitée :

$$\lim_{n \rightarrow \infty} \sup_{t_p \geq 1} P^{(t_p)} \left\{ \max_{t \leq n} \sum_{i=t_p}^{t_p+t} Z_i \geq I(1 + \delta)n \right\} = 0 \quad \forall \delta > 0. \quad (4.10)$$

La borne minimale pour le critère (4.8) est donnée par le théorème suivant :

Théorème 4.2 (Lai [50], Théorème 2) *Supposons que (4.9) et (4.10) soient vérifiées pour une constante positive I . Alors, lorsque $\alpha \rightarrow 0$,*

$$E^{(t_p)}(t_a - t_p + 1)^+ \geq \frac{P_{\theta_0}(t_a \geq t_p)}{I + o(1)} |\log \alpha| \quad (4.11)$$

uniformément sur $t_p \geq 1$.

Lai [50] montre que le CUSUM à fenêtre limitée atteint cette borne sous l'hypothèse (4.6).

REMARQUES SUR LA BORNE INFÉRIEURE (4.11) :

- L'avantage du critère précédent est que, contrairement au pire retard moyen à la détection (4.3), il est valable sur tous les instants de changement possibles et non pas sur l'instant de changement qui maximise le retard moyen.

- Le critère permet d'assouplir une des deux hypothèses de stabilité, (4.10), qui devient alors, d'après Tartakovsky, vérifiable dans de nombreuses situations.

*

* *

Dans les parties qui suivent, nous montrons l'optimalité du CUSUM approché construit à partir d'estimations consistantes des vraisemblances conditionnelles, d'une part au sens de Lorden, mais aussi en utilisant le critère (4.8) sous la contrainte (4.9) qui nécessite des hypothèses moins restrictives sur le logarithme du rapport de vraisemblance. Ces résultats d'optimalité très généraux seront appliqués dans les chapitres suivants.

4.3 Optimalité de la règle du CUSUM approché au sens de Lorden

Supposons que les hypothèses de stabilité du rapport de vraisemblance (4.2) et (4.6) soient vérifiées. Compte tenu du fait que la densité conditionnelle sous H_1 dépend de l'instant supposé de changement de régime, ces hypothèses de stabilité s'écrivent dans notre cas :

$$\limsup_{n \rightarrow \infty} \sup_{t_p \geq 1} \text{ess } P^{(t_p)} \left\{ \max_{t \leq n} \sum_{i=t_p}^{t_p+t} Z_{i,t_p} \geq I(1 + \delta)n | X_{1:t_p-1} \right\} = 0 \quad \forall \delta > 0, \quad (4.12)$$

$$\limsup_{n \rightarrow \infty} \sup_{t_p \geq 1} \sup_{j \geq t_p} \text{ess } P^{(t_p)} \left\{ n^{-1} \sum_{i=j}^{j+n} Z_{i,j} \leq I(1 - \delta) | X_{1:j-1} \right\} = 0, \quad \forall \delta > 0. \quad (4.13)$$

Le résultat du théorème 4.1 (Lai [50]) est toujours valable dans notre situation, bien qu'elle soit plus générale, et donc si l'hypothèse (4.12) est vérifiée, alors lorsque $\gamma \rightarrow \infty$, on a toujours la même borne inférieure :

$$\inf \{ \bar{E}_{\theta_1}(t_a) : E_{\theta_0}(t_a) \geq \gamma \} \geq (I^{-1} + o(1)) \log \gamma. \quad (4.14)$$

Supposons que les estimations des vraisemblances soient consistantes, à savoir :

$$\begin{cases} \hat{l}_{0,n}^N \rightarrow p_{\theta_0}(X_n | X_{1:n-1}) \text{ p.s.} \\ \hat{l}_{1,n,j}^N \rightarrow p_{\theta_0, \theta_{1,j}}(X_n | X_{1:n-1}) \text{ p.s. } \forall j \in [1; n], \end{cases} \quad (4.15)$$

Théorème 4.3 *Sous l'hypothèse (4.15), on a :*

$$\lim_{N \rightarrow \infty} E_{\theta_0}(\hat{t}) \geq e^h. \quad (4.16)$$

Théorème 4.4 *Lorsque les hypothèses (4.13) et (4.15) sont vérifiées et que N tend vers l'infini, le pire retard moyen à la détection de la règle \hat{t} vérifie :*

$$\bar{E}_{\theta_1}(\hat{t}) \leq (I^{-1} + o(1))h, \text{ lorsque } h \rightarrow \infty. \quad (4.17)$$

Corollaire 4.1 *Sous les hypothèses (4.12), (4.13) et (4.15), la règle \hat{t} est asymptotiquement (sur N et h) optimale au sens de Lorden.*

Démonstration : En prenant $h = \log \gamma$:

- Le théorème 4.3 assure que la règle \hat{t} vérifie asymptotiquement la contrainte ARL :

$$\lim_{N \rightarrow \infty} E_{\theta_0}(\hat{t}) \geq \gamma,$$

- Par le théorème 4.4, on montre que le pire retard moyen à la détection vérifie :

$$\lim_{N \rightarrow \infty} \bar{E}_{\theta_1}(\hat{t}) \leq (I^{-1} + o(1)) \log \gamma \quad \text{quand } \gamma \rightarrow \infty,$$

autrement dit que la règle \hat{t} atteint la borne inférieure de tous les algorithmes vérifiant la contrainte ARL définie en (4.14).

Donc la règle \hat{t} est asymptotiquement optimale sauf s'il existe une règle \hat{T}_N telle que :

$$\lim_{N \rightarrow \infty} E_{\theta_0}(\hat{T}_N) \geq \gamma, \quad (4.18)$$

et

$$\lim_{N \rightarrow \infty} \bar{E}_{\theta_1}(\hat{T}_N) < (I^{-1} + o(1)) \log \gamma.$$

Or, si (4.18) est vérifiée, alors,

$$\begin{aligned} & \forall \epsilon > 0, \exists N_0, \text{ tel que } \forall N \geq N_0, \quad E_{\theta_0}(\hat{T}_N + \epsilon) \geq \gamma \\ \Rightarrow & \forall \epsilon > 0, \exists N_0, \text{ tel que } \forall N \geq N_0, \quad \bar{E}_{\theta_1}(\hat{T}_N + \epsilon) \geq (I^{-1} + o(1)) \log \gamma \\ \Rightarrow & \bar{E}_{\theta_1}(\hat{T}_N) + \epsilon \geq (I^{-1} + o(1)) \log \gamma \\ \Rightarrow & \lim_{N \rightarrow \infty} \bar{E}_{\theta_1}(\hat{T}_N) \geq (I^{-1} + o(1)) \log \gamma. \quad \square \end{aligned}$$

Chapitre 4. Une approximation de la règle du CUSUM

Les théorèmes 4.3 et 4.4 établissent donc l'optimalité asymptotique (sur N et h) au sens de Lorden du CUSUM approché \hat{t} . La démonstration de ces théorèmes s'inspirent de Lai [50].

Démonstration du théorème 4.3 :

Nous allons noter \mathcal{F}_n la tribu engendrée par X_1, \dots, X_n . Lai [50], pour démontrer l'optimalité du CUSUM, utilise le fait que $\{\exp(\sum_{i=j}^n Z_{i,j}), \mathcal{F}_n, n \geq j\}$ est une martingale de moyenne 1 sous P_{θ_0} . Il peut ainsi utiliser l'inégalité des sous-martingales de Doob. Dans notre cas, $\{\exp(\sum_{i=j}^n \hat{Z}_{i,j}^N), \mathcal{F}_n, n \geq j\}$ n'est plus une martingale, et le lemme qui suit permet d'établir une inégalité analogue à celle de Doob.

Lemme 4.1 *Nous avons le résultat suivant : $\forall n, j \geq 1, j \leq n$,*

$$\forall \epsilon > 0, \forall \eta > 0, \exists N_0 \text{ tel que } \forall N \geq N_0,$$

$$\epsilon.P_{\theta_0} \left[\sup_{j \leq k \leq n} \exp \left(\sum_{i=j}^k \hat{Z}_{i,j}^N \right) \geq \epsilon \right] \leq 1 + \eta.$$

La démonstration du lemme précédent nécessite le résultat suivant, qui est une propriété bien connue des martingales :

Lemme 4.2 *Soit (M_n, \mathcal{F}_n) une martingale. Alors on a :*

$$E[M_n \mathbb{I}_A] = E[M_i \mathbb{I}_A] \quad \text{pour } A \in \mathcal{F}_i, \quad \forall i \leq n.$$

Démonstration du Lemme 4.1 :

Soit $\epsilon > 0$ et $j \leq n$. On considère les ensembles :

$$A = \left\{ \sup_{j \leq k \leq n} \exp \left(\sum_{i=j}^k Z_{i,j} \right) > \epsilon \right\}$$

et

$$\hat{A} = \left\{ \sup_{j \leq k \leq n} \exp \left(\sum_{i=j}^k \hat{Z}_{i,j}^N \right) > \epsilon \right\}.$$

Alors A et \hat{A} sont respectivement les unions disjointes des ensembles :

$$\begin{aligned} A_j &= \left\{ \exp \left(\sum_{i=j}^j Z_{i,j} \right) > \epsilon \right\} \\ &\vdots \\ A_k &= \left\{ \exp \left(\sum_{i=j}^k Z_{i,j} \right) > \epsilon \right\} \cap \left\{ \max_{j \leq t \leq k-1} \exp \left(\sum_{i=j}^t Z_{i,j} \right) \leq \epsilon \right\} \in \mathcal{F}_k \\ &\vdots \\ A_n &= \left\{ \exp \left(\sum_{i=j}^n Z_{i,j} \right) > \epsilon \right\} \cap \left\{ \max_{j \leq t \leq n-1} \exp \left(\sum_{i=j}^t Z_{i,j} \right) \leq \epsilon \right\} \in \mathcal{F}_n \end{aligned}$$

et

$$\begin{aligned} \hat{A}_j &= \left\{ \exp \left(\sum_{i=j}^j \hat{Z}_{i,j}^N \right) > \epsilon \right\} \\ &\vdots \\ \hat{A}_n &= \left\{ \exp \left(\sum_{i=j}^n \hat{Z}_{i,j}^N \right) > \epsilon \right\} \cap \left\{ \max_{j \leq t \leq n-1} \exp \left(\sum_{i=j}^t \hat{Z}_{i,j}^N \right) \leq \epsilon \right\} \in \mathcal{F}_n. \end{aligned}$$

Grâce au lemme 4.2 et puisque $\{\exp(\sum_{i=j}^n Z_{i,j}), n \geq j\}$ est une martingale (de moyenne 1) sous P_{θ_0} (voir Lai [50]), on a $\forall j \leq k \leq n$:

$$E_{\theta_0} \left[\exp \left(\sum_{i=j}^n Z_{i,j} \right) \mathbb{1}_{A_k} \right] = E_{\theta_0} \left[\exp \left(\sum_{i=j}^k Z_{i,j} \right) \mathbb{1}_{A_k} \right]. \quad (4.19)$$

Puisque par hypothèses les estimations $\hat{l}_{1,t,j}^N, 1 \leq j \leq t$ et $\hat{l}_{0,t}^N$ sont convergentes, le théorème des transformations continues assure que :

$$\lim_{N \rightarrow \infty} \exp \left(\sum_{i=j}^k \hat{Z}_{i,j}^N \right) = \exp \left(\sum_{i=j}^k Z_{i,j} \right) \quad \text{p.s.}$$

De plus, si on note $T_k = \exp(\sum_{i=j}^k Z_{i,j})$ et $\hat{T}_k = \exp(\sum_{i=j}^k \hat{Z}_{i,j}^N)$, les fonctions :

$$\begin{aligned} (T_j, \dots, T_k) &\mapsto T_k \mathbb{1}_{\{T_k > \epsilon\}} \mathbb{1}_{\{T_{k-1} \leq \epsilon\}} \dots \mathbb{1}_{\{T_j \leq \epsilon\}} \\ \text{et } (T_j, \dots, T_k, T_n) &\mapsto T_n \mathbb{1}_{\{T_k > \epsilon\}} \mathbb{1}_{\{T_{k-1} \leq \epsilon\}} \dots \mathbb{1}_{\{T_j \leq \epsilon\}} \end{aligned}$$

sont presque sûrement continues (les ensembles B_1 et B_2 de leurs points de discontinuité sont de mesure nulle) et donc par le théorème des transformations continues :

$$\begin{aligned} \hat{T}_k \mathbb{1}_{\{\hat{T}_k > \epsilon\}} \mathbb{1}_{\{\hat{T}_{k-1} \leq \epsilon\}} \dots \mathbb{1}_{\{\hat{T}_j \leq \epsilon\}} &\xrightarrow{N \rightarrow \infty} T_k \mathbb{1}_{\{T_k > \epsilon\}} \mathbb{1}_{\{T_{k-1} \leq \epsilon\}} \dots \mathbb{1}_{\{T_j \leq \epsilon\}} \quad \text{p.s.} \\ \hat{T}_n \mathbb{1}_{\{\hat{T}_k > \epsilon\}} \mathbb{1}_{\{\hat{T}_{k-1} \leq \epsilon\}} \dots \mathbb{1}_{\{\hat{T}_j \leq \epsilon\}} &\xrightarrow{N \rightarrow \infty} T_n \mathbb{1}_{\{T_k > \epsilon\}} \mathbb{1}_{\{T_{k-1} \leq \epsilon\}} \dots \mathbb{1}_{\{T_j \leq \epsilon\}} \quad \text{p.s.} \end{aligned}$$

Chapitre 4. Une approximation de la règle du CUSUM

Autrement dit,

$$\exp\left(\sum_{i=j}^k \hat{Z}_{i,j}^N\right) \mathbb{1}_{\hat{A}_k} \xrightarrow{N \rightarrow \infty} \exp\left(\sum_{i=j}^k Z_{i,j}\right) \mathbb{1}_{A_k} \quad \text{p.s.} \quad (4.20)$$

et

$$\exp\left(\sum_{i=j}^n \hat{Z}_{i,j}^N\right) \mathbb{1}_{\hat{A}_k} \xrightarrow{N \rightarrow \infty} \exp\left(\sum_{i=j}^n Z_{i,j}\right) \mathbb{1}_{A_k} \quad \text{p.s.} \quad (4.21)$$

Les deux quantités à gauche de (4.20) et (4.21) sont intégrables (voir ci-dessous), le théorème de convergence dominée de Lebesgue entraîne :

$$\begin{aligned} & \forall \epsilon, \forall \xi, \exists N_0 \text{ tel que } \forall N \geq N_0 : \\ & \quad \left| E_{\theta_0} \left[\hat{T}_k \mathbb{1}_{\hat{A}_k} \right] - E_{\theta_0} \left[T_k \mathbb{1}_{A_k} \right] \right| \leq \xi \\ & \quad \text{et} \quad \left| E_{\theta_0} \left[\hat{T}_n \mathbb{1}_{\hat{A}_k} \right] - E_{\theta_0} \left[T_n \mathbb{1}_{A_k} \right] \right| \leq \xi, \\ & \quad \text{ainsi,} \quad \left| E_{\theta_0} \left[\hat{T}_k \mathbb{1}_{\hat{A}_k} \right] - E_{\theta_0} \left[\hat{T}_n \mathbb{1}_{\hat{A}_k} \right] \right| < 2\xi \end{aligned} \quad (4.22)$$

grâce à l'égalité (4.19).

Or,

$$E_{\theta_0} \left[\hat{T}_k \mathbb{1}_{\hat{A}_k} \right] > \epsilon P_{\theta_0}(\hat{A}_k) \quad \forall k \leq n, \quad \forall j \leq k,$$

puisque sur \hat{A}_k , $\hat{T}_k > \epsilon$. Et donc pour N assez grand, nous obtenons :

$$\epsilon P_{\theta_0}(\hat{A}_k) - E_{\theta_0} \left[\hat{T}_n \mathbb{1}_{\hat{A}_k} \right] < 2\xi \quad \forall k \leq n, \quad (4.23)$$

en utilisant (4.22). Et en sommant de $k = j$ jusqu'à n l'inégalité (4.23) on obtient si N est suffisamment grand :

$$E_{\theta_0} \left[\hat{T}_n \mathbb{1}_{\hat{A}_n \cup \dots \cup \hat{A}_j} \right] + 2(n - j + 1)\xi > \epsilon P_{\theta_0}(\hat{A}_n \cup \dots \cup \hat{A}_j),$$

puisque les \hat{A}_k sont deux à deux disjoints, et donc,

$$E_{\theta_0} \left[\hat{T}_n \mathbb{1}_{\hat{A}} \right] + 2(n - j + 1)\xi > \epsilon P_{\theta_0} \left[\sup_{j \leq k \leq n} \hat{T}_k > \epsilon \right].$$

Finalement, si N est assez grand,

$$E_{\theta_0} \left[\hat{T}_n \right] + 2(n - j + 1)\xi > \epsilon P_{\theta_0} \left[\sup_{j \leq k \leq n} \hat{T}_k \geq \epsilon \right]. \quad (4.24)$$

Optimalité de la règle du CUSUM approché au sens de Lorden

Or, $\hat{T}_n \rightarrow T_n$ presque sûrement quand $N \rightarrow \infty$, donc le théorème de convergence dominée de Lebesgue entraîne :

$$\forall n \geq 1, \forall j \leq n, \forall \epsilon > 0, \exists N_0 \text{ tel que } \forall N \geq N_0 \\ E_{\theta_0} [\hat{T}_n] \leq E_{\theta_0} [T_n] + \epsilon = 1 + \epsilon.$$

On obtient donc, à partir de (4.24), le résultat cherché : $\forall j, n \geq 1$,

$$\forall \epsilon > 0, \forall \eta > 0, \exists N_0 \text{ tel que } \forall N \geq N_0, \\ \epsilon \cdot P_{\theta_0} \left[\sup_{j \leq k \leq n} \exp \left(\sum_{i=j}^k \hat{Z}_{i,j}^N \right) \geq \epsilon \right] \leq 1 + \eta.$$

Nous avons utilisé dans la démonstration du lemme, le théorème de convergence dominée. Il reste donc à vérifier que $\exp \left(\sum_{i=j}^n \hat{Z}_{i,j}^N \right)$ est intégrable. Or, puisque

$$\hat{l}_{1,i,j}^N \rightarrow p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1}) \quad \text{et} \quad \hat{l}_{0,i}^N \rightarrow p_{\theta_0}(X_i | X_{1:i-1}),$$

il vient :

$$\exists N_1 \text{ tel que } \forall N \geq N_1, \\ \hat{l}_{1,i,j}^N \leq 2p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1}) \quad \text{p.s.} \\ \text{et} \quad \hat{l}_{0,i}^N \geq \frac{1}{2}p_{\theta_0}(X_i | X_{1:i-1}) \quad \text{p.s.}$$

Ainsi, pour $N > N_1$,

$$\exp \left(\sum_{i=j}^n \hat{Z}_{i,j}^N \right) \leq \prod_{i=j}^n \frac{4 \cdot p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \\ \leq 4^{n-j} \prod_{i=j}^n \frac{p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \quad \text{p.s.}$$

et

$$\forall n \geq 1, \quad E_{\theta_0} \left[4^{n-j} \prod_{i=j}^n \frac{p_{\theta_0, \theta_{1,j}}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \right] = 4^{n-j} < \infty.$$

Les conditions d'applications du théorème de convergence dominée de Lebesgue sont donc vérifiées ce qui achève la démonstration du lemme 4.1. \square

Chapitre 4. Une approximation de la règle du CUSUM

On peut alors revenir à la démonstration du théorème 4.3. Nous définissons les temps d'arrêts suivants :

$$\tau_0 = 0, \quad \tau_1 = \inf \left\{ t : \sum_{i=1}^t \hat{Z}_{i,1}^N < 0 \right\} \quad (\inf \emptyset = \infty)$$

et,

$$\forall r \geq 1, \quad \tau_{r+1} = \inf \left\{ t > \tau_r : \sum_{i=\tau_r+1}^t \hat{Z}_{i,\tau_r+1}^N < 0 \right\}.$$

D'après le résultat du lemme 4.1,

$$\begin{aligned} \forall j, \forall n \geq j, \forall \eta > 0 \exists N_0 \text{ tel que } \forall N \geq N_0 \\ e^h \cdot P_{\theta_0} \left\{ \exp\left(\sum_{i=j}^n \hat{Z}_{i,j}^N\right) \geq e^h \right\} \leq 1 + \eta \\ \Rightarrow P_{\theta_0} \left\{ \sum_{i=j}^n \hat{Z}_{i,j}^N \geq h \right\} \leq (1 + \eta)e^{-h}. \end{aligned}$$

Ainsi, sur $\{\tau_r < \infty\}$, on obtient pour $n > \tau_r$ et $N \geq N_0$:

$$P_{\theta_0} \left\{ \sum_{i=\tau_r+1}^n \hat{Z}_{i,\tau_r+1}^N \geq h | \mathcal{F}_{1:\tau_r} \right\} \leq (1 + \eta)e^{-h}. \quad (4.25)$$

Soit

$$R = \inf \left\{ r \geq 0 : \tau_r < \infty \text{ et il existe } t > \tau_r \text{ tel que } \sum_{i=\tau_r+1}^t \hat{Z}_{i,\tau_r+1}^N \geq h \right\}.$$

Alors, soit r_0 fixé,

$$\begin{aligned} P_{\theta_0}(R > r_0) &= E_{\theta_0}[\mathbb{1}_{\{R > r_0\}}] \\ &= E_{\theta_0}[E_{\theta_0}[\mathbb{1}_{\{R > r_0\}} | \mathcal{F}_{1:\tau_{r_0}}]] \\ &= E_{\theta_0}[E_{\theta_0}[\mathbb{1}_{\{R > r_0\}} \cdot \mathbb{1}_{\{R \geq r_0\}} | \mathcal{F}_{1:\tau_{r_0}}]] \\ &= E_{\theta_0}[E_{\theta_0}[\mathbb{1}_{\{R > r_0\}} | \mathcal{F}_{1:\tau_{r_0}}] \mathbb{1}_{\{R \geq r_0\}}] \\ &= E_{\theta_0}[P_{\theta_0}(R \geq r_0 + 1 | \mathcal{F}_{1:\tau_{r_0}}) \mathbb{1}_{\{R \geq r_0\}}]. \end{aligned}$$

Or,

$$\begin{aligned} P_{\theta_0}(R \geq r_0 + 1 | \mathcal{F}_{1:\tau_{r_0}}) &= P_{\theta_0} \left(\sum_{i=\tau_{r_0}+1}^n \hat{Z}_{i,\tau_{r_0}+1}^N < h \text{ pour } \tau_{r_0} < \infty \text{ et } n > \tau_{r_0} | \mathcal{F}_{1:\tau_{r_0}} \right) \\ &\geq (1 - (1 + \eta)e^{-h}) \text{ par (4.25),} \end{aligned}$$

et donc,

$$\begin{aligned}
 P_{\theta_0}(R > r_0) &\geq E_{\theta_0} \left[(1 - (1 + \eta)e^{-h}) \mathbb{I}_{\{R \geq r_0\}} \right] \\
 &\geq (1 - (1 + \eta)e^{-h}) E_{\theta_0} \left[\mathbb{I}_{\{R \geq r_0\}} \right] \\
 &\geq (1 - (1 + \eta)e^{-h}) P_{\theta_0}(R \geq r_0) \\
 &\geq (1 - (1 + \eta)e^{-h}) P_{\theta_0}(R > r_0 - 1).
 \end{aligned}$$

Nous obtenons alors,

$$E_{\theta_0}(R) = \sum_{r_0=0}^{\infty} P_{\theta_0}(R > r_0) \geq \sum_{r_0=0}^{\infty} \left(1 - \frac{1 + \eta}{e^h} \right)^{r_0} = \frac{e^h}{1 + \eta}.$$

Autrement dit,

$$\forall \eta, \exists N_0 \text{ tel que } \forall N \geq N_0, \quad E_{\theta_0}(R) \geq \frac{e^h}{1 + \eta}.$$

Enfin, puisque

$$\hat{t} = \inf \left\{ n : \sum_{i=\tau_r+1}^n \hat{Z}_{i,\tau_r+1}^N \geq h \text{ pour } \tau_r < n \right\} \geq R,$$

on peut conclure que

$$\forall \eta, \exists N_0 \text{ tel que } \forall N \geq N_0, \quad E_{\theta_0}(\hat{t}) \geq \frac{e^h}{1 + \eta}.$$

Autrement dit,

$$\lim_{N \rightarrow \infty} E_{\theta_0}(\hat{t}) \geq e^h. \quad \square$$

Démonstration du théorème 4.4 :

Dans cette démonstration, nous allons utiliser le lemme suivant :

Lemme 4.3 *Supposons que l'hypothèse (4.13) soit vérifiée. Alors cette hypothèse est toujours vraie lorsqu'on remplace les incréments $Z_{i,j}$ par les estimations $\hat{Z}_{i,j}^N$ et qu'on fait tendre N vers l'infini.*

Démonstration du lemme :

On a, $\forall n \geq 1$,

$$\forall \epsilon' > 0, \exists N_0 \text{ tel que } \forall N \geq N_0, \quad \left| \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} - \frac{1}{n} \sum_{i=j}^{j+n} \hat{Z}_{i,j}^N \right| \leq \epsilon' \quad \text{p.s.} \quad (4.26)$$

Chapitre 4. Une approximation de la règle du CUSUM

De plus, l'hypothèse (4.13) assure que :

$$\forall \epsilon > 0, \forall \eta > 0, \exists n_0 \text{ tel que } \forall n \geq n_0$$

$$\sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ I - \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} \geq \epsilon | X_{1:j-1} \right\} \leq \eta. \quad (4.27)$$

Soient $\epsilon > 0$ et $n \geq n_0$. Alors,

$$\begin{aligned} & \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ I - \frac{1}{n} \sum_{i=j}^{j+n} \hat{Z}_{i,j}^N \geq \epsilon | X_{1:j-1} \right\} \\ &= \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ I - \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} + \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} - \frac{1}{n} \sum_{i=j}^{j+n} \hat{Z}_{i,j}^N \geq \epsilon | X_{1:j-1} \right\} \\ &\leq \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ I - \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} + \left| \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} - \frac{1}{n} \sum_{i=j}^{j+n} \hat{Z}_{i,j}^N \right| \geq \epsilon | X_{1:j-1} \right\} \\ &\leq \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ I - \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} \geq \epsilon - \left| \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} - \frac{1}{n} \sum_{i=j}^{j+n} \hat{Z}_{i,j}^N \right| | X_{1:j-1} \right\}, \end{aligned}$$

ce qui, combiné à (4.26), avec $\epsilon' \in]0; \epsilon[$ et $N \geq N_0$ donne :

$$\begin{aligned} & \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ I - \frac{1}{n} \sum_{i=j}^{j+n} \hat{Z}_{i,j}^N \geq \epsilon | X_{1:j-1} \right\} \\ &\leq \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ I - \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} \geq \epsilon - \epsilon' | X_{1:j-1} \right\}. \end{aligned}$$

Finalement, on a,

$$\forall 0 < \delta < \epsilon, \forall \eta > 0, \exists n_0 \text{ tel que } \forall n \geq n_0, \exists N_0 \text{ tel que } \forall N \geq N_0,$$

$$\begin{aligned} & \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ I - \frac{1}{n} \sum_{i=j}^{j+n} \hat{Z}_{i,j}^N \geq \epsilon | X_{1:j-1} \right\} \\ &\leq \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ I - \frac{1}{n} \sum_{i=j}^{j+n} Z_{i,j} \geq \delta | X_{1:j-1} \right\} \\ &\leq \eta \text{ par (4.27)}. \end{aligned}$$

Ainsi, le lemme est vérifié et on a donc :

$$\lim_{n \rightarrow \infty} \lim_{N \rightarrow \infty} \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ n^{-1} \sum_{i=j}^{j+n} \hat{Z}_{i,j}^N \leq I(1 - \delta) | X_{1:j-1} \right\} = 0, \quad \forall \delta > 0. \quad \square$$

Optimalité de la règle du CUSUM approché au sens de Lorden

On peut alors revenir à la démonstration du théorème. Pour montrer le résultat, il suffit de montrer que : $\forall 0 < \delta < 1$,

$$\sup_{t_p \geq 1} \sup_{\text{ess}} E^{(t_p)} \{(\hat{t} - t_p + 1)^+ | X_1, \dots, X_{t_p-1}\} \leq (1 + o(1))(1 - \delta)^{-2} I^{-1} h, \quad (4.28)$$

lorsque $h \rightarrow \infty$ et $N \rightarrow \infty$.

Soit δ compris strictement entre 0 et 1. Considérons à présent n_h le plus petit entier tel que $n_h - 1$ soit supérieur à $(1 - \delta)^{-1} I^{-1} h$. Alors on a grâce au lemme 4.3, si h (et donc n_h) et N sont suffisamment grands :

$$\begin{aligned} & \sup_{t \geq t_p \geq 1} \sup_{\text{ess}} P^{(t_p)} \left\{ (n_h - 1)^{-1} \sum_{i=t}^{t+n_h-1} \hat{Z}_{i,t}^N \leq I(1 - \delta) | X_{1:t-1} \right\} < \delta \\ \Rightarrow & \sup_{t \geq t_p \geq 1} \sup_{\text{ess}} P^{(t_p)} \left\{ \sum_{i=t}^{t+n_h-1} \hat{Z}_{i,t}^N \leq I(1 - \delta)(n_h - 1) | X_{1:t-1} \right\} < \delta. \end{aligned}$$

Or, $n_h - 1 > (1 - \delta)^{-1} I^{-1} h$, donc $h < I(1 - \delta)(n_h - 1)$. Ainsi,

$$\sup_{t \geq t_p \geq 1} \sup_{\text{ess}} P^{(t_p)} \left\{ \sum_{i=t}^{t+n_h-1} \hat{Z}_{i,t}^N \leq h | X_{1:t-1} \right\} < \delta. \quad (4.29)$$

Lorsque h et N sont suffisamment grands, nous obtenons quels que soient $k \geq 1$ et $t_p \geq 1$:

$$\begin{aligned} & \sup_{\text{ess}} P^{(t_p)} \{ \hat{t} - t_p + 1 > kn_h | X_{1:t_p-1} \} \\ &= \sup_{\text{ess}} P^{(t_p)} \{ \hat{t} - t_p + 1 > jn_h ; \forall 1 \leq j \leq k | X_{1:t_p-1} \} \\ &= \sup_{\text{ess}} P^{(t_p)} \{ \hat{t} > jn_h + t_p - 1 ; \forall 1 \leq j \leq k | X_{1:t_p-1} \} \\ &= \sup_{\text{ess}} P^{(t_p)} \left\{ \max_{1 \leq u \leq jn_h + t_p - 1} \sum_{i=u}^{jn_h + t_p - 1} \hat{Z}_{i,u}^N < h ; \forall 1 \leq j \leq k | X_{1:t_p-1} \right\} \\ &\leq \sup_{\text{ess}} P^{(t_p)} \left\{ \sum_{i=(j-1)n_h + t_p}^{jn_h + t_p - 1} \hat{Z}_{i,(j-1)n_h + t_p}^N < h ; \forall 1 \leq j \leq k | X_{1:t_p-1} \right\} \\ &\leq \sup_{\text{ess}} P^{(t_p)} \left\{ \sum_{i=t_p}^{n_h + t_p - 1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-1)n_h + t_p}^{kn_h + t_p - 1} \hat{Z}_{i,(k-1)n_h + t_p}^N < h | X_{1:t_p-1} \right\} \end{aligned}$$

Or,

$$\begin{aligned} & P^{(t_p)} \left\{ \sum_{i=t_p}^{n_h+t_p-1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-1)n_h+t_p}^{kn_h+t_p-1} \hat{Z}_{i,(k-1)n_h+t_p}^N < h \mid \mathcal{F}_{1:t_p-1} \right\} \\ &= E^{(t_p)} \left(P^{(t_p)} \left\{ \sum_{i=t_p}^{n_h+t_p-1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-1)n_h+t_p}^{kn_h+t_p-1} \hat{Z}_{i,(k-1)n_h+t_p}^N < h \mid \mathcal{F}_{1:(k-1)n_h+t_p-1} \right\} \mid \mathcal{F}_{1:t_p-1} \right) \end{aligned}$$

grâce à la propriété $P(A|\mathcal{G}) = E(P(A|\mathcal{F})|\mathcal{G})$ si $\mathcal{G} \subset \mathcal{F}$,

$$\begin{aligned} &= E^{(t_p)} \left(E^{(t_p)} \left\{ \mathbb{I}_{\{\sum_{i=t_p}^{n_h+t_p-1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-1)n_h+t_p}^{kn_h+t_p-1} \hat{Z}_{i,(k-1)n_h+t_p}^N < h\}} \mid \mathcal{F}_{1:(k-1)n_h+t_p-1} \right\} \mid \mathcal{F}_{1:t_p-1} \right) \\ &= E^{(t_p)} \left(E^{(t_p)} \left\{ \mathbb{I}_{\{\sum_{i=t_p}^{n_h+t_p-1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-2)n_h+t_p}^{(k-1)n_h+t_p-1} \hat{Z}_{i,(k-2)n_h+t_p}^N < h\}} \right. \right. \\ &\quad \left. \left. \mathbb{I}_{\{\sum_{i=(k-1)n_h+t_p}^{kn_h+t_p-1} \hat{Z}_{i,(k-1)n_h+t_p}^N < h\}} \mid \mathcal{F}_{1:(k-1)n_h+t_p-1} \right\} \mid \mathcal{F}_{1:t_p-1} \right) \\ &= E^{(t_p)} \left(\mathbb{I}_{\{\sum_{i=t_p}^{n_h+t_p-1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-2)n_h+t_p}^{(k-1)n_h+t_p-1} \hat{Z}_{i,(k-2)n_h+t_p}^N < h\}} \right. \\ &\quad \left. E^{(t_p)} \left\{ \mathbb{I}_{\{\sum_{i=(k-1)n_h+t_p}^{kn_h+t_p-1} \hat{Z}_{i,(k-1)n_h+t_p}^N < h\}} \mid \mathcal{F}_{1:(k-1)n_h+t_p-1} \right\} \mid \mathcal{F}_{1:t_p-1} \right) \\ &= E^{(t_p)} \left(\mathbb{I}_{\{\sum_{i=t_p}^{n_h+t_p-1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-2)n_h+t_p}^{(k-1)n_h+t_p-1} \hat{Z}_{i,(k-2)n_h+t_p}^N < h\}} \right. \\ &\quad \left. P^{(t_p)} \left\{ \sum_{i=(k-1)n_h+t_p}^{kn_h+t_p-1} \hat{Z}_{i,(k-1)n_h+t_p}^N < h \mid \mathcal{F}_{1:(k-1)n_h+t_p-1} \right\} \mid \mathcal{F}_{1:t_p-1} \right) \\ &\leq E^{(t_p)} \left(\mathbb{I}_{\{\sum_{i=t_p}^{n_h+t_p-1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-2)n_h+t_p}^{(k-1)n_h+t_p-1} \hat{Z}_{i,(k-2)n_h+t_p}^N < h\}} \cdot \delta \mid \mathcal{F}_{1:t_p-1} \right) \end{aligned}$$

en utilisant le résultat (4.29)

$$\leq \delta \cdot P^{(t_p)} \left\{ \sum_{i=t_p}^{n_h+t_p-1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-2)n_h+t_p}^{(k-1)n_h+t_p-1} \hat{Z}_{i,(k-2)n_h+t_p}^N < h \mid \mathcal{F}_{1:t_p-1} \right\}$$

En répétant ces étapes en conditionnant successivement par $X_1, \dots, X_{(j-1)n_h+t_p-1}$ pour $j = k-1, \dots, 1$, nous obtenons alors le résultat suivant :

$$P^{(t_p)} \left\{ \sum_{i=t_p}^{n_h+t_p-1} \hat{Z}_{i,t_p}^N < h, \dots, \sum_{i=(k-1)n_h+t_p}^{kn_h+t_p-1} \hat{Z}_{i,(k-1)n_h+t_p}^N < h \mid \mathcal{F}_{1:t_p-1} \right\} \leq \delta^k.$$

Ainsi

$$\sup \operatorname{ess} P^{(t_p)} \{ \hat{t} - t_p + 1 > kn_h | X_{1:t_p-1} \} \leq \delta^k,$$

et puisque $kn_h > 0$ on peut remplacer $\hat{t} - t_p + 1$ par $(\hat{t} - t_p + 1)^+$ ce qui permet d'obtenir :

$$\sup \operatorname{ess} P^{(t_p)} \{ (\hat{t} - t_p + 1)^+ > kn_h | X_{1:t_p-1} \} \leq \delta^k. \quad (4.30)$$

Ainsi, lorsque h et N sont suffisamment grands :

$$\begin{aligned} & \sup_{t_p \geq 1} \sup \operatorname{ess} E^{(t_p)} \{ n_h^{-1} (\hat{t} - t_p + 1)^+ | X_{1:t_p-1} \} \\ &= \sup_{t_p \geq 1} \sup \operatorname{ess} \left(\sum_{k=0}^{\infty} P^{(t_p)} (n_h^{-1} (\hat{t} - t_p + 1)^+ \geq k | X_{1:t_p-1}) \right) \\ &\leq \sum_{k=0}^{\infty} \sup_{t_p \geq 1} \sup \operatorname{ess} P^{(t_p)} (n_h^{-1} (\hat{t} - t_p + 1)^+ \geq k | X_{1:t_p-1}) \\ &\leq \sum_{k=0}^{\infty} \delta^k = (1 - \delta)^{-1}. \end{aligned}$$

De plus $n_h \leq (1 - \delta)^{-1} I^{-1} h + 2$, et alors :

$$\begin{aligned} \sup_{t_p \geq 1} \sup \operatorname{ess} E^{(t_p)} \{ (\hat{t} - t_p + 1)^+ | X_{1:t_p-1} \} &\leq (1 - \delta)^{-2} I^{-1} h + 2(1 - \delta)^{-1} \\ &\leq (1 + o(1))(1 - \delta)^{-2} I^{-1} h, \end{aligned}$$

ce qui nous permet de conclure à l'optimalité de la procédure de filtrage-détection, puisque nous obtenons le résultat (4.28). \square

4.4 Autre critère d'optimalité

Dans cette partie, nous allons nous intéresser au critère du retard moyen :

$$E^{(t_p)} (t_a - t_p + 1)^+, \quad (4.31)$$

lorsqu'on considère toutes les règles de décision qui vérifient la contrainte suivante :

$$\sup_{t \geq 1} P_{\theta_0} (t \leq t_a < t + m_\alpha) \leq \alpha, \quad (4.32)$$

où $\liminf \frac{m_\alpha}{|\log \alpha|} > I^{-1}$ mais $\log m_\alpha = o(\log \alpha)$ quand $\alpha \rightarrow 0$.

Chapitre 4. Une approximation de la règle du CUSUM

Sous l'hypothèse suivante, qui est l'adaptation de l'hypothèse (4.10), à la situation que nous considérons où la densité conditionnelle sous régime de panne dépend de l'instant de changement,

$$\limsup_{n \rightarrow \infty} \sup_{t_p \geq 1} P^{(t_p)} \left\{ \max_{t \leq n} \sum_{i=t_p}^{t_p+t} Z_{i,t_p} \geq I(1+\delta)n \right\} = 0 \quad \forall \delta > 0, \quad (4.33)$$

et si (4.32) est vérifiée, alors le théorème 4.2 est toujours valable et donne donc la borne inférieure de retard moyen à la détection :

$$E^{(t_p)}(t_a - t_p + 1)^+ \geq \frac{P_{\theta_0}(t_a \geq t_p)}{I + o(1)} |\log \alpha|. \quad (4.34)$$

L'objectif est alors de montrer que le CUSUM atteint cette borne inférieure. On considère m_α des entiers positifs tels que :

$$\liminf \frac{m_\alpha}{|\log \alpha|} > I^{-1} \quad \text{mais} \quad \log m_\alpha = o(\log \alpha) \quad \text{quand} \quad \alpha \rightarrow 0. \quad (4.35)$$

Lai [50] montre que dans le cas de données *iid*, le CUSUM vérifie la contrainte (4.32). En fait, pour une valeur α fixée, il est possible de choisir un seuil h de sorte que (4.32) soit vérifiée. Lorsqu'on se place dans le cas de données dépendantes, il semblerait que le CUSUM ne vérifie plus la contrainte fixée. Lai propose donc d'utiliser un CUSUM à fenêtre limitée (comme Willsky et Jones [112]) où la maximisation du logarithme du rapport de vraisemblance s'effectue sur $[n - m_\alpha; n]$. Nous procédons de la même façon en considérant un "CUSUM approché" à fenêtre limitée, que nous écrivons :

$$\hat{t}_{lim} = \inf \left\{ n : \max_{n-m_\alpha \leq j \leq n} \sum_{i=j}^n \hat{Z}_{i,j}^N \geq h \right\}$$

où $\hat{Z}_i^N = \log(\hat{l}_{1,i,j}^N / \hat{l}_{0,i}^N)$. Nous obtenons le théorème suivant :

Théorème 4.5 *Si l'hypothèse (4.15) assurant la consistance des estimations $\hat{l}_{1,i,j}^N$ et $\hat{l}_{0,i}^N$ est vérifiée, alors :*

i) L'estimation du CUSUM à fenêtre limitée \hat{t}_{lim} vérifie lorsque $N \rightarrow \infty$:

$$\sup_{t \geq 1} P_{\theta_0}(t \leq \hat{t}_{lim} \leq t + m_\alpha) \leq 2m_\alpha e^{-h}.$$

Autrement dit, la contrainte (4.32) est vérifiée si on prend $h = \log(2m_\alpha/\alpha)$ lorsque $N \rightarrow \infty$.

ii) De plus, si m_α vérifie (4.35), si $h = \log(2m_\alpha/\alpha)$, et si les hypothèses (4.33) et (4.13) sont vérifiées, alors lorsque $N \rightarrow \infty$ et $\alpha \rightarrow 0$:

$$E^{(t_p)}(\hat{t}_{lim} - t_p + 1)^+ \sim \frac{P_{\theta_0}(\hat{t}_{lim} \geq t_p)}{I + o(1)} |\log \alpha|,$$

uniformément sur $t_p \geq 1$. Autrement dit, le CUSUM approché à fenêtre limitée atteint la borne inférieure du retard moyen à la détection et est donc asymptotiquement optimal.

Démonstration du théorème 4.5, i)

Compte tenu de la définition de \hat{t}_{lim} , on a l'inclusion suivante : $\forall t \geq 1$,

$$\begin{aligned} \{t \leq \hat{t}_{lim} < t + m_\alpha\} &\subset \bigcup_{j=t-m_\alpha}^{t+m_\alpha-1} \left\{ \max_{j \leq k \leq k+m_\alpha} \sum_{i=j}^k \hat{Z}_{i,j}^N \geq h \right\} \\ &\subset \bigcup_{j=t-m_\alpha}^{t+m_\alpha-1} \left\{ \max_{j \leq k \leq n} \sum_{i=j}^k \hat{Z}_{i,j}^N \geq h \right\} \quad \text{où } n \geq t + m_\alpha, \end{aligned}$$

et donc,

$$P_{\theta_0}(t \leq \hat{t}_{lim} < t + m_\alpha) \leq \sum_{j=t-m_\alpha}^{t+m_\alpha-1} P_{\theta_0} \left[\max_{j \leq k \leq n} \sum_{i=j}^k \hat{Z}_{i,j}^N \geq h \right].$$

D'autre part, $\forall n$,

$$P_{\theta_0} \left[\max_{j \leq k \leq n} \sum_{i=j}^k \hat{Z}_{i,j}^N \geq h \right] = P_{\theta_0} \left[\max_{j \leq k \leq n} \exp \left(\sum_{i=j}^k \hat{Z}_{i,j}^N \right) \geq e^h \right],$$

et donc, on obtient grâce au lemme 4.1 :

$$\begin{aligned} \forall \eta > 0, \exists N_0 \text{ tel que } \forall N \geq N_0, \\ P_{\theta_0} \left[\max_{j \leq k \leq n} \sum_{i=j}^k \hat{Z}_{i,j}^N \geq h \right] \leq (1 + \eta)e^{-h}. \end{aligned}$$

Ainsi,

$$\forall \eta > 0, \exists N_0 \text{ tel que } \forall N \geq N_0, \\ P_{\theta_0}(t \leq \hat{t}_{lim} < t + m_\alpha) \leq \sum_{j=t-m_\alpha}^{t+m_\alpha-1} (1 + \eta)e^{-h}$$

$$\Rightarrow P_{\theta_0}(t \leq \hat{t}_{lim} < t + m_\alpha) \leq (1 + \eta)2m_\alpha e^{-h}.$$

Chapitre 4. Une approximation de la règle du CUSUM

Autrement dit,

$$\lim_{N \rightarrow \infty} P_{\theta_0}(t \leq \hat{t}_{lim} < t + m_\alpha) \leq 2m_\alpha e^{-h},$$

et le point i) est donc démontré.

Démonstration du théorème 4.5, ii)

- On a : $2m_\alpha e^{-h} = \alpha$ donc $h = \log(2m_\alpha/\alpha)$ et donc :

$$h = \log 2 + \log m_\alpha - \log \alpha,$$

mais comme, d'après (4.35), $\log m_\alpha = o(\log \alpha)$, on a :

$$h \sim |\log \alpha|.$$

- Pour obtenir le résultat cherché, puisque l'hypothèse (4.13) est vérifiée, il suffit de reprendre la démonstration du théorème 4.4 au niveau de (4.29) en définissant n_h de la même façon, et de remarquer la chose suivante :

$$n_h \sim (1 - \delta)^{-1} I^{-1} h \Rightarrow n_h \sim (1 - \delta)^{-1} I^{-1} |\log \alpha|.$$

Or,

$$\begin{aligned} \liminf \frac{m_\alpha}{|\log \alpha|} > I^{-1} &\Rightarrow \liminf \frac{m_\alpha}{I^{-1} |\log \alpha|} > 1 \\ &\Rightarrow \liminf \frac{m_\alpha}{n_h} > 1, \end{aligned}$$

si δ est suffisamment petit.

Alors, lorsque h et N sont suffisamment grands (h grand équivaut à α proche de 0), nous obtenons quels que soient $k \geq 1$ et $t_p \geq 1$:

$$\begin{aligned} &\sup \text{ess } P^{(t_p)} \{ \hat{t}_{lim} - t_p + 1 > kn_h | X_{1:t_p-1} \} \\ &= \sup \text{ess } P^{(t_p)} \{ \hat{t}_{lim} - t_p + 1 > jn_h ; \forall 1 \leq j \leq k | X_{1:t_p-1} \} \\ &= \sup \text{ess } P^{(t_p)} \{ \hat{t}_{lim} > jn_h + t_p - 1 ; \forall 1 \leq j \leq k | X_{1:t_p-1} \} \\ &= \sup \text{ess } P^{(t_p)} \left\{ \max_{jn_h + t_p - 1 - m_\alpha \leq u \leq jn_h + t_p - 1} \sum_{i=u}^{jn_h + t_p - 1} \hat{Z}_{i,u}^N < h ; \forall 1 \leq j \leq k | X_{1:t_p-1} \right\} \\ &\leq \sup \text{ess } P^{(t_p)} \left\{ \max_{(j-1)n_h + t_p - 1 \leq u \leq jn_h + t_p - 1} \sum_{i=u}^{jn_h + t_p - 1} \hat{Z}_{i,u}^N < h ; \forall 1 \leq j \leq k | X_{1:t_p-1} \right\}, \end{aligned}$$

où la dernière inégalité vient du fait que si α est suffisamment petit, alors $m_\alpha > n_h$ et donc :

$$\begin{aligned} & jn_h + t_p - 1 - m_\alpha < jn_h + t_p - 1 - n_h \\ \Rightarrow & P^{(t_p)} \left\{ \max_{jn_h + t_p - 1 - m_\alpha \leq u \leq jn_h + t_p - 1} \sum_{i=u}^{jn_h + t_p - 1} \hat{Z}_{i,u}^N < h \right\} \\ & \leq P^{(t_p)} \left\{ \max_{(j-1)n_h + t_p - 1 \leq u \leq jn_h + t_p - 1} \sum_{i=u}^{jn_h + t_p - 1} \hat{Z}_{i,u}^N < h \right\}, \end{aligned}$$

et on continue en suivant la démonstration du théorème 4.4. Ainsi, on obtient le résultat suivant grâce à (4.13) :

$$\sup_{t_p \geq 1} \sup \text{ess } E^{(t_p)} \{ (\hat{t}_{lim} - t_p + 1)^+ | X_1, \dots, X_{t_p-1} \} \leq (1 + o(1))(1 - \delta)^{-2} I^{-1} h, \quad (4.36)$$

si δ est suffisamment petit.

On a : $\forall t_p \geq 1$,

$$\begin{aligned} & E^{(t_p)} [(\hat{t}_{lim} - t_p + 1)^+] \\ & = E^{(t_p)} [(\hat{t}_{lim} - t_p + 1)^+ | \hat{t}_{lim} \geq t_p] \cdot P^{(t_p)} [\hat{t}_{lim} \geq t_p] \\ & \quad + E^{(t_p)} [(\hat{t}_{lim} - t_p + 1)^+ | \hat{t}_{lim} < t_p] \cdot P^{(t_p)} [\hat{t}_{lim} < t_p] \\ & = E^{(t_p)} [(\hat{t}_{lim} - t_p + 1)^+ | \hat{t}_{lim} \geq t_p] \cdot P^{(t_p)} [\hat{t}_{lim} \geq t_p] \end{aligned}$$

car $E[X] = E[X|A] \cdot P(A) + E[X|B] \cdot P(B)$ si A et B forment une partition de Ω . De plus $E^{(t_p)} [(\hat{t}_{lim} - t_p + 1)^+ | \hat{t}_{lim} < t_p] = E^{(t_p)} [0] = 0$.

Ainsi, $\forall t_p \geq 1$,

$$\begin{aligned} E^{(t_p)} [(\hat{t}_{lim} - t_p + 1)^+] & = E^{(t_p)} [\hat{t}_{lim} - t_p + 1 | \hat{t}_{lim} \geq t_p] \cdot P^{(t_p)} [\hat{t}_{lim} \geq t_p] \\ & \leq \sup \text{ess } E^{(t_p)} [(\hat{t}_{lim} - t_p + 1)^+ | X_1, \dots, X_{t_p-1}] \cdot P^{(t_p)} [\hat{t}_{lim} \geq t_p], \end{aligned}$$

puisque $\{\hat{t}_{lim} \geq t_p\} \in \mathcal{F}_{t_p-1}$.

Et donc, $\forall t_p \geq 1$, lorsque $\alpha \rightarrow 0$ et $N \rightarrow \infty$,

$$E^{(t_p)} [(\hat{t}_{lim} - t_p + 1)^+] \leq (1 + o(1))(1 - \delta)^{-2} I^{-1} h \cdot P^{(t_p)} [\hat{t}_{lim} \geq t_p].$$

On termine la démonstration en remarquant que :

$$P^{(t_p)} [\hat{t}_{lim} \geq t_p] = P_{\theta_0} [\hat{t}_{lim} \geq t_p]. \quad \square$$

Le résultat obtenu en (4.36) permet d'autre part d'obtenir le corollaire suivant :

Chapitre 4. Une approximation de la règle du CUSUM

Corollaire 4.2 *Le CUSUM approché à fenêtre limité est asymptotiquement optimal au sens de Lorden si les conditions de stabilité du logarithme du rapport de vraisemblance (4.12) et (4.13) sont vérifiées et si m_α vérifie (4.35), lorsque $h \sim |\log \alpha| \rightarrow \infty$. En effet, on a alors :*

$$\bar{E}_{\theta_1}(\hat{t}_{lim}) \leq (I^{-1} + o(1))h$$

lorsque $h \rightarrow \infty$.

4.5 Conclusions

Dans ce chapitre, nous avons établi l’optimalité asymptotique d’une règle de décision construite à partir des estimations des vraisemblances conditionnelles sous régime de fonctionnement normal et sous régime de panne.

Nous avons tout d’abord établi l’optimalité de la règle du “CUSUM approché” (corollaire 4.1) et de sa version à fenêtre limitée (corollaire 4.2) au sens de Lorden, c’est à dire pour le critère du pire retard moyen à la détection et la contrainte du temps moyen entre deux fausses alarmes. Les deux hypothèses de stabilité sur le logarithme du rapport de vraisemblance nécessaires pour obtenir ces résultats sont assez fortes mais dans le chapitre suivant, nous montrons qu’elles sont vérifiées pour des modèles auto régressifs relativement généraux. Pour ce qui concerne les modèles à espace d’état, la question de la vérification de ces hypothèses reste ouverte. Et c’est la raison pour laquelle nous avons cherché à obtenir un résultat d’optimalité pour des hypothèses plus faibles. Nous n’avons réussi qu’à moitié puisque nous avons obtenu l’optimalité de la procédure proposée pour un critère plus souple que le pire retard moyen à la détection et une contrainte sur la probabilité de fausse alarme sur une fenêtre glissante. L’utilisation de ce critère permet d’assouplir une des deux hypothèses de stabilité qui devient, d’après Tartakovsky [107], vérifiable dans de nombreuses situations. Cependant, nos résultats dépendent toujours de la contrainte (4.13).

Dans tout ce chapitre, les résultats d’optimalité ont été obtenus pour des seuils fixes, qu’on fait tendre vers l’infini. Ces résultats sont donc difficilement utilisables en pratique puisque un seuil infini entraîne un retard à la détection infini.

D’autre part, on a vu dans le chapitre précédent que le choix d’un seuil fixe n’est pas toujours adapté. Il serait intéressant d’établir l’optimalité des règles CUSUM ou CUSUM approché en considérant la contrainte sur la probabilité de fausse alarme utilisée pour le seuil adaptatif.

Chapitre 5

Le CUSUM non paramétrique avec apprentissage

Nous nous plaçons dans cette partie dans le cas présenté dans l'introduction où une partie du modèle décrivant le processus est inconnue. Dans le cas de procédés biotechnologiques comme les procédés de dépollution, cette composante inconnue est souvent la conséquence de l'incertitude qui affecte la modélisation des taux de croissance des biomasses. Nous considérons donc le modèle autorégressif suivant :

$$X_{n+1} = f(X_n) + F(X_n, \theta) + \epsilon_{n+1} \quad (5.1)$$

où

- $(X_n)_n \in \mathbb{R}^d$ est la suite des variables d'état,
- $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est une fonction connue,
- $(\epsilon_n)_n$ est un bruit blanc gaussien de matrice de covariance R inversible,
- $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est la fonction inconnue.

Nous supposons qu'il y a deux régimes de fonctionnement : nominal (régime $H_0 : \theta = \theta_0$) et panne (régime $H_1 : \theta = \theta_1$). Comme nous l'avons déjà évoqué, le test CUSUM nécessite pour son utilisation la connaissance complète du modèle sous H_0 et sous H_1 de sorte à pouvoir calculer les vraisemblances conditionnelles $p_{\theta_0}(X_n | X_{1:n-1})$ et $p_{\theta_1}(X_n | X_{1:n-1})$, ce qui n'est pas le cas ici puisque f est inconnue.

REMARQUES :

- Compte tenu de la structure du modèle considéré ici, le changement est généré par la deuxième méthode décrite dans la partie 2.3.2.

- Dans ce chapitre, nous utilisons comme norme vectorielle, la norme euclidienne définie pour un vecteur $X = (X^1, \dots, X^d)$ de \mathbb{R}^d par :

$$\|X\| = \sqrt{(X^1)^2 + \dots + (X^d)^2}.$$

Chapitre 5. Le CUSUM non paramétrique avec apprentissage

La norme matricielle utilisée est la norme de Frobenius définie pour une matrice $M \in \mathbb{R}^{r \times s}$ par :

$$\|M\| = \sqrt{\sum_{i=1}^r \sum_{j=1}^s M_{ij}^2}.$$

La norme de Frobenius est sous-multiplicative, c'est-à-dire que

$$\|MP\| \leq \|M\| \|P\|,$$

et cette relation s'applique que P soit une matrice ou un vecteur. Dans le cas d'un vecteur, la norme de Frobenius correspond à la norme euclidienne.

*
* *

Dans un premier temps, nous supposons que la fonction f est connue et nous montrons que la règle du CUSUM classique, appliquée au modèle (5.1), est optimale au sens de Lorden, c'est à dire que les conditions de stabilité (4.2) et (4.6) sont vérifiées par le modèle (5.1).

Nous revenons ensuite au cas nous intéressant où la fonction f est inconnue et nous introduisons une adaptation de la règle du CUSUM, que nous appellerons "CUSUM non paramétrique", qui va consister tout simplement à appliquer la procédure CUSUM en remplaçant, dans le calcul des incréments, la fonction f par un estimateur de cette dernière. Nous utiliserons pour cela une estimation non paramétrique à noyaux de convolution. Une rapide introduction sur les estimateurs à noyaux de convolution est proposée en annexe A.

5.1 Cas où la fonction f est connue

5.1.1 Ecriture de la règle du CUSUM

Rappelons brièvement l'écriture de la statistique de test du CUSUM :

$$\forall n, \quad g_n = \max_{1 \leq k \leq n} \sum_{i=k}^n Z_i \quad \text{avec} \quad Z_i = \log \frac{p_{\theta_1}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})}.$$

Pour écrire le test du CUSUM, il faut donc déterminer l'écriture analytique de la densité conditionnelle des observations $p_{\theta}(x_n | X_{1:n-1})$. Compte tenu de la nature gaussienne du bruit et du fait qu'il s'agit d'un bruit additif dans le modèle considéré, la loi conditionnelle des observations est une loi normale :

$$\mathcal{L}(X_i | X_{1:i-1}) = \mathcal{N}(f(X_{i-1}) + F(X_{i-1}, \theta), R).$$

La densité conditionnelle des observations s'écrit donc :

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

On peut alors calculer les incréments Z_i :

$$Z_i = \log \left(\frac{p_{\theta_1}(X_i|X_{1:i-1})}{p_{\theta_0}(X_i|X_{1:i-1})} \right) = \frac{1}{2} [(\eta_i^0)^t R^{-1} \eta_i^0 - (\eta_i^1)^t R^{-1} \eta_i^1], \quad (5.2)$$

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

5.1.2 Optimalité de la règle du CUSUM pour le modèle (5.1) lorsque f est connue

Nous avons vu dans le chapitre précédent, d'après les résultats de Lai [50], que si les deux hypothèses de stabilité du rapport de vraisemblances conditionnelles :

$$\lim_{n \rightarrow \infty} \sup_{t_p \geq 1} \sup \operatorname{ess} P^{(t_p)} \left\{ \max_{t \leq n} \sum_{i=t_p}^{t_p+t} Z_i \geq I(1 + \delta)n | X_{1:t_p-1} \right\} = 0 \quad \forall \delta > 0, \quad (5.3)$$

et

$$\lim_{n \rightarrow \infty} \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \operatorname{ess} P^{(t_p)} \left\{ n^{-1} \sum_{i=j}^{j+n} Z_i \leq I(1 - \delta) | X_{1:j-1} \right\} = 0, \quad \forall \delta > 0, \quad (5.4)$$

sont vérifiées, alors la règle du CUSUM est asymptotiquement optimale pour le critère du pire retard moyen à la détection. Montrons donc que ces hypothèses sont vérifiées pour le modèle (5.1).

Lai [50] formule la remarque suivante : supposons que le processus $\{X_n, n \geq 1\}$ soit une chaîne de Markov de transition $p_{\theta_0}(\cdot|\cdot)$ pour $t < t_p$ et $p_{\theta_1}(\cdot|\cdot)$ pour $t \geq t_p$. Alors les deux hypothèses (5.3) et (5.4) se simplifient par :

$$\begin{aligned} \exists I, \quad \forall \delta > 0, \quad \sup_x P^{(1)} \left\{ \max_{t \leq n} \sum_{i=1}^{t+1} Z_i \geq I(1 + \delta)n | X_0 = x \right\} &\rightarrow 0 \\ \text{et} \quad \sup_x P^{(1)} \left\{ \sum_{i=1}^{n+1} Z_i \leq (I - \delta)n | X_0 = x \right\} &\rightarrow 0 \quad \text{lorsque } n \rightarrow \infty. \end{aligned}$$

Théorème 5.1 (Duflo [28], p. 23) *Soit $Y = (Y_n)$ une suite de variables aléatoires. Supposons que (a_n) est une suite qui tend vers l'infini. Si Y_n/a_n converge vers 0 presque sûrement, alors, $\max_{k \leq n} Y_k/a_n$ converge vers 0 presque sûrement.*

Chapitre 5. Le CUSUM non paramétrique avec apprentissage

Corollaire 5.1 *Une condition suffisante pour que les deux hypothèses de stabilité précédentes soient vérifiées est que la quantité $1/n \sum_{i=1}^n Z_i$ converge presque sûrement vers une constante I .*

Démonstration : Il suffit alors d'appliquer le théorème 5.1 aux suites $Y_n = \sum_{i=1}^n Z_i$, $a_n = n$, et on conclut par le fait que la convergence presque sûre entraîne la convergence en probabilité. \square

Le théorème suivant (Duflo [27]) permet d'obtenir la convergence presque sûre de $1/n \sum_{i=1}^n Z_i$ pour des cas particuliers de modèles autorégressifs :

Théorème 5.2 (Duflo [27]) *Soit un modèle autorégressif fonctionnel sur \mathbb{R}^d :*

$$X_n = u(X_{n-1}) + \epsilon_n.$$

S1 -

$$(H) : \limsup_{\|x\| \rightarrow \infty} \frac{\|u(x)\|}{\|x\|} = \rho < 1$$

et la loi du bruit a une densité strictement positive par rapport à la mesure de Lebesgue.

S2 - Il existe une norme pour laquelle la fonction u est lipschitzienne de rapport $\rho < 1$.

Sous S1 ou S2, on a les propriétés suivantes :

a- Le modèle est stable. Sa loi stationnaire μ admet un moment d'ordre 2.

b- Pour toute fonction μ -presque sûrement continue v de \mathbb{R}^d dans \mathbb{R} (l'ensemble des points de discontinuité de v est de mesure nulle par rapport à μ), telle que $|v(x)| \leq A\|x\|^2 + B$, A et B étant des constantes, on a :

$$\frac{1}{n+1} \sum_{i=0}^n v(X_i) \rightarrow \mu(v) \quad \text{p.s.,}$$

pour toute loi initiale du processus X_n .

Revenons donc au modèle (5.1). Puisqu'on suppose, dans cette sous-partie, que la fonction f est connue, pour des raisons de lisibilité, on va ré-écrire ce modèle sous la forme :

$$\forall n, \quad X_n = u(X_{n-1}, \theta) + \epsilon_n, \quad (5.5)$$

avec u une fonction connue et en supposant, sans perte de généralité que $\epsilon_n \sim N(0, I_d)$, où I_d est la matrice identité de $\mathbb{R}^{d \times d}$.

Corollaire 5.2 *Si la fonction u est continue et vérifie l'hypothèse de stabilité :*

$$(H) : \limsup_{\|x\| \rightarrow \infty} \frac{\|u(x, \theta)\|}{\|x\|} = \rho < 1, \forall \theta,$$

alors,

$$\exists I \quad \text{tel que} \quad \frac{1}{n} \sum_{i=1}^n Z_i \rightarrow I \quad \text{p.s.}$$

Démonstration :

On a,

$$X_{n-1} = u(X_{n-2}, \theta) + \epsilon_{n-1}, \quad \epsilon_{n-1} \sim N(0, I_d),$$

et ϵ_n et ϵ_{n-1} sont indépendants. Soit T_n le couple : $T_n := (X_n, X_{n-1})$. On peut alors écrire que :

$$\begin{aligned} T_n &= U(T_{n-1}) + \epsilon'_n, & \text{avec} \quad U : \mathbb{R}^{2d} &\longrightarrow \mathbb{R}^{2d} \\ & & (x, y) &\longmapsto (u(x), u(y)) \\ & & \text{et} \quad \epsilon'_n &\sim N(0, I_{2d}). \end{aligned}$$

- Si U vérifie l'hypothèse de stabilité

$$\lim_{\|t\| \rightarrow \infty} \sup \frac{\|U(t)\|}{\|t\|} = \rho' < 1,$$

et si

$$\begin{aligned} v : \mathbb{R}^{2d} &\longrightarrow \mathbb{R} \\ t = (x, y) &\longmapsto \log \frac{p_{\theta_1}(x|y)}{p_{\theta_0}(x|y)} \end{aligned}$$

est presque sûrement continue et vérifie

$$|v(t)| \leq A\|t\|^2 + B, \quad A \text{ et } B \text{ constantes,}$$

alors on peut appliquer le théorème 5.2 et on obtient :

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n v(T_i) \rightarrow \mu(v) \quad \text{p.s.} \\ \Rightarrow &\frac{1}{n} \sum_{i=1}^n v(X_i, X_{i-1}) \rightarrow \mu(v) \quad \text{p.s.} \\ \Rightarrow &\frac{1}{n} \sum_{i=1}^n \log \frac{p_{\theta_1}(X_i|X_{i-1})}{p_{\theta_0}(X_i|X_{i-1})} \rightarrow \mu(v) = I \quad \text{p.s.} \end{aligned}$$

Chapitre 5. Le CUSUM non paramétrique avec apprentissage

- Montrons tout d'abord que si :

$$\limsup_{\|x\| \rightarrow \infty} \frac{\|u(x)\|}{\|x\|} = \rho < 1 ,$$

alors on a bien que :

$$\limsup_{\|t\| \rightarrow \infty} \frac{\|U(t)\|}{\|t\|} = \rho' < 1 .$$

Puisqu'on utilise la norme euclidienne :

$$\begin{aligned} \|U(t)\| &:= \|(u(x), u(y))\| \\ &= \sqrt{\|u(x)\|^2 + \|u(y)\|^2} \\ \text{et } \|t\| &:= \sqrt{\|x\|^2 + \|y\|^2}. \end{aligned}$$

Ainsi,

$$\frac{\|U(t)\|}{\|t\|} = \sqrt{\frac{\|u(x)\|^2 + \|u(y)\|^2}{\|x\|^2 + \|y\|^2}} . \quad (5.6)$$

Il faut à présent s'intéresser au comportement du rapport (5.6) lorsque $\|t\|$ tend vers l'infini. Pour cela, on va considérer les trois cas possibles :

Cas 1 : $\|x\|$ tend vers l'infini et $\|y\|$ ne tend pas vers l'infini. On considère donc qu'il existe une constante C bornant $\|y\|$.

$$\begin{aligned} \frac{\|U(t)\|}{\|t\|} &= \sqrt{\frac{\|u(x)\|^2}{\|x\|^2 + \|y\|^2} + \frac{\|u(y)\|^2}{\|x\|^2 + \|y\|^2}} \\ &\leq \sqrt{\frac{\|u(x)\|^2}{\|x\|^2 + \|y\|^2} + \frac{(\sup_{\|y\| \leq C} \|u(y)\|)^2}{\|x\|^2 + \|y\|^2}} \\ &\leq \sqrt{\frac{\|u(x)\|^2}{\|x\|^2} + \underbrace{\frac{Cte}{x^2}}_{\downarrow \text{ lorsque } \|x\| \rightarrow \infty}} \\ &\quad 0 \end{aligned}$$

et donc,

$$\limsup_{\|t\| \rightarrow \infty} \frac{\|U(t)\|}{\|t\|} \leq \limsup_{\|x\| \rightarrow \infty} \frac{\|u(x)\|}{\|x\|} = \rho < 1 .$$

(Nous avons utilisé le fait que u est continue, et donc ainsi, $\sup_{\|y\| \leq C} \|u(y)\|$ est fini.)

Cas 2 : $\|x\|$ ne tend pas vers l'infini et $\|y\|$ tend vers l'infini. Ce cas là se traite exactement comme le premier cas en inversant les rôles de x et y . On aboutit donc au même résultat.

Cas 3 : $\|x\|$ et $\|y\|$ tendent vers l'infini.

$$\begin{aligned} \frac{\|U(t)\|}{\|t\|} &= \sqrt{\frac{\|u(x)\|^2}{\|x\|^2 + \|y\|^2} + \frac{\|u(y)\|^2}{\|x\|^2 + \|y\|^2}} \\ &= \sqrt{\frac{\frac{\|x\|^2}{\|x\|^2 + \|y\|^2} \|u(x)\|^2}{\|x\|^2} + \frac{\frac{\|y\|^2}{\|x\|^2 + \|y\|^2} \|u(y)\|^2}{\|y\|^2}} \\ &= \sqrt{\frac{\|x\|^2}{\|x\|^2 + \|y\|^2} \left(\frac{\|u(x)\|^2}{\|x\|^2}\right) + \frac{\|y\|^2}{\|x\|^2 + \|y\|^2} \left(\frac{\|u(y)\|^2}{\|y\|^2}\right)}. \end{aligned}$$

Or, par hypothèse,

$$\limsup_{\|x\| \rightarrow \infty} \frac{\|u(x)\|}{\|x\|} = \rho < 1,$$

et donc,

$$\limsup_{\|t\| \rightarrow \infty} \frac{\|U(t)\|}{\|t\|} \leq \rho.$$

- Montrons finalement que v est μ -presque sûrement continue et qu'elle vérifie :

$$|v(t)| \leq A\|t\|^2 + B.$$

$$\begin{aligned} v(x, y) &= \log \frac{p_{\theta_1}(x|y)}{p_{\theta_0}(x|y)} \\ &= \frac{1}{2} [(x - u(y; \theta_0))^t R^{-1} (x - u(y; \theta_0)) - (x - u(y; \theta_1))^t R^{-1} (x - u(y; \theta_1))] \\ &= \frac{1}{2} (\|x - u(y, \theta_0)\|^2 - \|x - u(y, \theta_1)\|^2) \quad \text{car } R = I_d \\ &\leq \frac{(\|x\| + \|u(y, \theta_0)\|)^2 + (\|x\| + \|u(y, \theta_1)\|)^2}{2}. \end{aligned}$$

Cette fonction est continue sur \mathbb{R}^{2d} , donc μ -presque sûrement continue. De plus, soit $k > 1$:

Chapitre 5. Le CUSUM non paramétrique avec apprentissage

- Sur le pavé $[-k; k] \times \dots \times [-k; k]$, v est continue, donc bornée par une constante B .

$$\Rightarrow |v(x, y)| \leq A\|(x, y)\|^2 + B, \quad \forall A > 0.$$

- Sur l'espace $\mathbb{R}^{2d} \setminus [-k; k]^{2d}$, on a,

$$\begin{aligned} \frac{|v(x, y)|}{\|(x, y)\|^2} &\leq \frac{(\|x\| + \|u(y, \theta_0)\|)^2 + (\|x\| + \|u(y, \theta_1)\|)^2}{2\|(x, y)\|^2} \\ &\leq \frac{\|x\|^2}{\|(x, y)\|^2} + \frac{\|x\|(\|u(y, \theta_0)\| + \|u(y, \theta_1)\|)}{\|(x, y)\|^2} + \frac{\|u(y, \theta_0)\|^2}{2\|(x, y)\|^2} + \frac{\|u(y, \theta_1)\|^2}{2\|(x, y)\|^2} \\ &\leq \frac{\|x\|^2}{\|(x, y)\|^2} + \frac{\|x\|(\|u(y, \theta_0)\| + \|u(y, \theta_1)\|)}{\|x\|^2 + \|y\|^2} + \frac{\|u(y, \theta_0)\|^2}{2\|y\|^2} + \frac{\|u(y, \theta_1)\|^2}{2\|y\|^2} \\ &\leq \frac{\|x\|^2}{\|(x, y)\|^2} + \frac{\|x\|\|u(y, \theta_0)\|}{\|x\|^2 + \|y\|^2} + \frac{\|x\|\|u(y, \theta_1)\|}{\|x\|^2 + \|y\|^2} + \frac{\|u(y, \theta_0)\|^2}{2\|y\|^2} + \frac{\|u(y, \theta_1)\|^2}{2\|y\|^2} \\ &\leq \frac{\|x\|^2}{\|(x, y)\|^2} + \frac{\frac{\|x\|\|y\|}{\|x\|^2 + \|y\|^2} \|u(y; \theta_1)\|}{\|y\|} + \frac{\frac{\|x\|\|y\|}{\|x\|^2 + \|y\|^2} \|u(y; \theta_0)\|}{\|y\|} \\ &\quad + \frac{\|u(y, \theta_0)\|^2}{2\|y\|^2} + \frac{\|u(y, \theta_1)\|^2}{2\|y\|^2} \\ &\leq A, \end{aligned}$$

puisque

$$\frac{\|x\|^2}{\|(x, y)\|^2} \leq 1 \quad \text{et} \quad \frac{\|x\|\|y\|}{\|x\|^2 + \|y\|^2} \leq 1,$$

et si k est assez grand, par hypothèse, la quantité $u(y; \theta_j)/y$ est bornée. \square

Les résultats d'optimalité asymptotique de la règle du CUSUM sont alors valables pour le modèle (5.5), et donc pour le modèle (5.1) si f est connue.

REMARQUE : Nous avons considéré que $R = I_d$ sans perte de généralité puisque si R est une matrice de covariance quelconque, et si je note : $a = x - g(y, \theta_0)$, alors :

$$a^t R^{-1} a \leq |a^t R^{-1} a| \leq \|a^t\| \cdot \|R^{-1}\| \cdot \|a\| = \|a\|^2 \cdot \|R^{-1}\|.$$

La quantité $\|R^{-1}\|$ ne joue aucun rôle dans la continuité de v et peut être prise en compte dans la constante A .

5.2 Le CUSUM non paramétrique avec apprentissage

Dans cette partie on se place dans le cas où la fonction f du modèle (5.1) est inconnue. La densité conditionnelle des observations est alors, elle aussi, inconnue. Les

incréments de la règle du CUSUM (voir (5.2)) ne peuvent pas être écrits. L'approche de détection que nous allons développer se base sur une estimation non paramétrique de la fonction f inconnue.

5.2.1 L'estimateur à noyau de convolution

Les méthodes non paramétriques par noyaux de convolution ont connu leurs premières utilisations dans le cadre de l'estimation de fonction de densité de probabilité (Rosenblatt [89], Parzen [77]) puis dans celui de l'estimation de fonction de régression (Nadaraya [71]). L'estimation de la fonction f dans le modèle (5.1) a tout d'abord été traitée par Senoussi [95], qui a obtenu des résultats de convergence uniforme sur les compacts.

Supposons donc que nous disposions d'un échantillon d'apprentissage de N observations $(X_i)_{1 \leq i \leq N}$ du modèle (5.1), l'estimateur de la fonction f est le suivant :

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

K est une fonction bornée, positive, symétrique, intégrable et d'intégrale égale à 1, appelée noyau. δ_N est la fenêtre, quantité qui doit tendre vers zéro quand N tend vers l'infini. L'avantage des méthodes non paramétriques réside dans le fait qu'aucune hypothèse structurelle n'est à faire sur la fonction f mis à part une hypothèse de continuité. Un échantillon d'apprentissage est suffisant pour estimer cette fonction. Dans Senoussi [95], il est montré que sous des hypothèses générales sur le modèle, le noyau, le bruit blanc et avec une fenêtre de la forme $\delta_i = C.i^{-\alpha}$ (C constante), l'estimateur est uniformément convergent sur les compacts :

$$\forall v, \quad \sup_{\|x\| \leq v} \|\hat{f}_N(x) - f(x)\| = O(N^{-\alpha}(\log \log N)^{1/2}) \quad \text{p.s.} \quad \text{où } 0 < \alpha < 1/2(d+1),$$

où d est la dimension de la variable X .

Cet estimateur a déjà été étudié et introduit dans un algorithme de contrôle adaptatif pour la régulation d'un procédé de dépollution par digestion anaérobie (Hilgert et coll [41]). Il est de mise au point très simple et a montré de bonnes qualités d'estimation.

5.2.2 Remarques sur l'échantillon d'apprentissage de l'estimateur

L'estimateur à noyau \hat{f}_N est une fonction définie sur l'ensemble \mathbb{R}^d . L'estimateur sera d'autant meilleur en un point x qu'il y aura beaucoup de données d'apprentissage X_i dans le voisinage de ce point x . Il est nécessaire, pour atteindre un bon

niveau de détection, que notre estimateur soit performant dans la région où les variables X_i prennent leurs valeurs lorsque le système est sous le régime H_0 mais il faut qu'il le soit aussi dans la région où les variables X_i prennent leurs valeurs lorsque le système est sous le régime H_1 . Autrement dit, notre échantillon d'apprentissage devra contenir des données du système en fonctionnement normal mais aussi en fonctionnement anormal, ce qui peut être réalisé dans une phase préliminaire de simple observation du processus.

Dans le cas d'un modèle contrôlé, des données en régime de panne ne sont pas forcément nécessaires. On peut en effet se servir, quand c'est possible, de la variable de contrôle pour "envoyer" X_i dans les régions d'intérêt. Nous reviendrons dans la partie 5.4 sur cet aspect.

L'estimateur non paramétrique peut s'écrire sous forme récursive de la façon suivante :

$$\begin{aligned} H_n(x) &= H_{n-1}(x) + \delta_n^{-d} K \left(\frac{x - X_i}{\delta_i} \right), & H_{-1}(x) &= 0 \\ N_n(x) &= N_{n-1}(x) + \delta_n^{-d} K \left(\frac{x - X_i}{\delta_i} \right) (X_{n+1} - F(X_n, \theta)), & N_{-1}(x) &= 0, \end{aligned}$$

où $N_n(x)$ et $H_n(x)$ représentent respectivement le numérateur et le dénominateur de $\hat{f}_N(x)$. Cette écriture récursive permet d'augmenter très facilement, au cours du temps, la taille de l'échantillon d'apprentissage. Cependant, dans le cadre de la détection d'anomalie, il faut se montrer prudent puisque pour rajouter des observations à l'échantillon, il faut savoir dans quel régime se trouve le système, sous peine de risquer de contaminer l'échantillon. Or, lorsqu'on surveille le processus, on ne peut être sûr qu'il soit sous H_0 compte tenu du retard à la détection.

Dans ce document, on considère que l'estimateur non paramétrique \hat{f}_N de la fonction f a été calculé au préalable, c'est à dire avant la mise en place de la procédure de détection, à partir d'un échantillon de taille N pour lequel les régimes, normal et de panne, ont parfaitement été identifiés (voir figure 5.1). On suppose de plus que l'échantillon est suffisamment riche pour que l'estimateur soit performant à la fois sous régime normal et sous régime de panne.

Les données de l'échantillon d'apprentissage et les données recueillies pendant la procédure de surveillance sont donc totalement distinctes. On ne traite pas dans ce document la situation plus complexe où les observations obtenues au cours de la surveillance sont utilisées pour réactualiser l'estimateur (schéma de la figure 5.2).

5.2.3 Test CUSUM non paramétrique

Nous définissons la règle du CUSUM non paramétrique comme celle du CUSUM, en remplaçant dans le calcul des incréments (5.2), $f(X_i)$ par son estimation $\hat{f}_N(X_i)$.

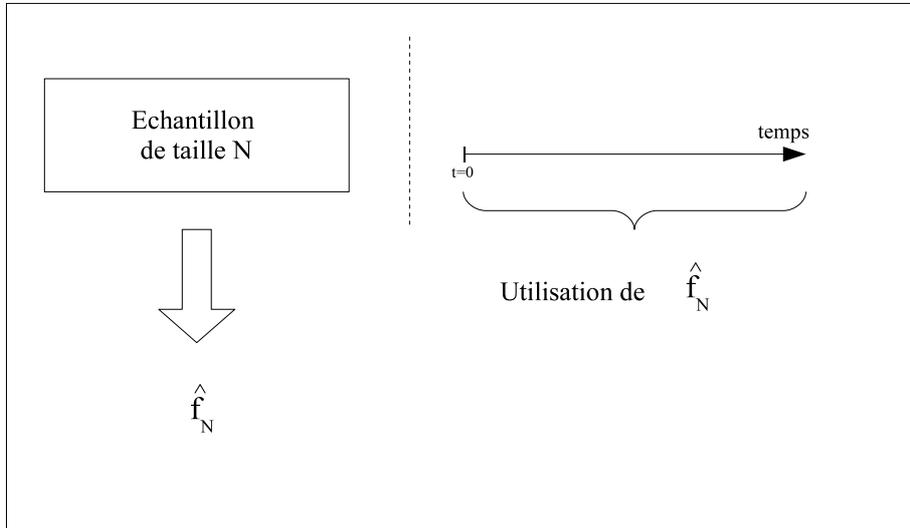


FIG. 5.1 – Situation où l'estimateur est construit à partir d'un échantillon d'apprentissage indépendant à la surveillance en cours

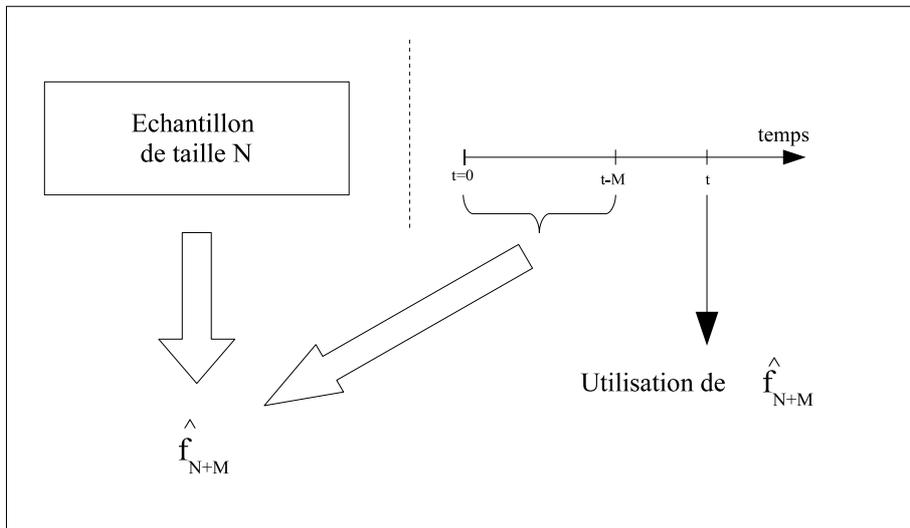


FIG. 5.2 – Situation où l'estimateur est construit à partir d'un échantillon d'apprentissage ainsi que de données provenant de la surveillance en cours

La statistique de test s'écrit :

$$\hat{g}_n^N = \max_{1 \leq k \leq n} \sum_{i=k}^n \hat{Z}_i^N \quad \text{avec} \quad \hat{Z}_i^N = \frac{1}{2} [(\hat{\eta}_i^0)^t R^{-1} \hat{\eta}_i^0 - (\hat{\eta}_i^1)^t R^{-1} \hat{\eta}_i^1]$$

et $\hat{\eta}_i^r = X_i - \hat{f}_N(X_{i-1}) - F(X_{i-1}, \theta_r), \quad r = 0, 1.$

Le temps d'arrêt est alors défini par :

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

5.3 Propriétés théoriques du test

Nous avons vu dans la partie 5.1 que la règle du CUSUM, pour des modèles autorégressifs vérifiant de bonnes hypothèses, est optimale pour détecter un changement de paramètre dans la densité conditionnelle des observations. Le résultat de Senoussi [95] assurant la convergence de l'estimateur à noyau laisse penser que la règle du CUSUM non paramétrique doit avoir un comportement similaire à celle du CUSUM lorsque la taille de l'échantillon d'apprentissage est suffisamment grande. Nous montrons par la suite que le CUSUM non paramétrique est lui aussi optimal au sens de Lorden en appliquant les résultats obtenus dans le chapitre 4. Pour cela, il faut tout d'abord établir la convergence des incréments de l'algorithme.

Théorème 5.3 *Considérons le modèle autorégressif :*

$$X_n = f(X_{n-1}) + F(X_{n-1}, \theta) + \epsilon_n \quad , \quad \epsilon_n \sim N(0, R). \quad (5.7)$$

Supposons que la fonction f inconnue soit lipschitzienne de rapport $r_f < 1$ et que la fonction F soit continue et bornée. Alors,

$$\forall n \geq 1 \quad \lim_{N \rightarrow \infty} |\hat{Z}_n^N - Z_n| = 0 \quad \text{p.s.} \quad (5.8)$$

La démonstration du Théorème précédent nécessite l'utilisation du Théorème et du Lemme qui suivent.

Théorème 5.4 (Portier et Oulidi ([84], corollaire 3.2)) *(Convergence sur les compacts dilatés)*

Si les deux hypothèses suivantes sont vérifiées :

Hyp 1 : *La fonction f est lipschitzienne de rapport $r_f < 1$, ie :*

$$\exists r_f < 1, \forall x, \forall y, \quad \|f(x) - f(y)\| \leq r_f \|x - y\|.$$

Hyp 2 : $\epsilon = (\epsilon_n)_{n \geq 1}$ est une suite de variables aléatoires gaussiennes indépendantes et identiquement distribuées de moyenne nulle.

alors, lorsque $\delta_i = i^{-1/2(d+1)}$, pour toute loi initiale et quel que soit $A \leq \infty$, on a :

$$\sup_{\|x\| \leq v_N} \|\hat{f}_N(x) - f(x)\| = O(N^{-\lambda}), \quad (5.9)$$

avec $\lambda \in]0; 1/2(d+1)[$, et v_N est une suite tendant vers l'infini définie par :

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

Lemme 5.1 (Duflo [28]) Si $\epsilon = (\epsilon_n)$ est un bruit blanc gaussien, alors :

$$\sup_{k \leq n} \|\epsilon_k\| = O(\sqrt{\log n}) \text{ p.s.}$$

Démonstration de Théorème 5.3 :

$\forall r = 0, 1$ et $\forall n \geq 1$:

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

On voit apparaître un biais entre les résidus du CUSUM et les résidus du CUSUM non paramétrique :

$$b_{n-1}^N := f(X_{n-1}) - \hat{f}_N(X_{n-1})$$

et $\hat{\eta}_n^r = \eta_n^r + b_{n-1}^N$.

Les incréments du CUSUM non paramétrique s'écrivent donc :

$$\begin{aligned} \forall n, \quad \hat{Z}_n^N &= \frac{1}{2} [(\eta_n^0 + b_{n-1}^N)^t R^{-1}(\eta_n^0 + b_{n-1}^N) - (\eta_n^1 + b_{n-1}^N)^t R^{-1}(\eta_n^1 + b_{n-1}^N)] \\ &= \frac{1}{2} \left[\left((\eta_n^0)^t + (b_{n-1}^N)^t \right) R^{-1}(\eta_n^0 + b_{n-1}^N) - \left((\eta_n^1)^t + (b_{n-1}^N)^t \right) R^{-1}(\eta_n^1 + b_{n-1}^N) \right] \\ &= \frac{1}{2} \left[(\eta_n^0)^t R^{-1} \eta_n^0 + (\eta_n^0)^t R^{-1} b_{n-1}^N + (b_{n-1}^N)^t R^{-1} \eta_n^0 + (b_{n-1}^N)^t R^{-1} b_{n-1}^N \right. \\ &\quad \left. - (\eta_n^1)^t R^{-1} \eta_n^1 - (\eta_n^1)^t R^{-1} b_{n-1}^N - (b_{n-1}^N)^t R^{-1} \eta_n^1 - (b_{n-1}^N)^t R^{-1} b_{n-1}^N \right] \\ &= \frac{1}{2} \left[(\eta_n^0)^t R^{-1} \eta_n^0 - (\eta_n^1)^t R^{-1} \eta_n^1 + 2(\eta_n^0)^t R^{-1} b_{n-1}^N - 2(\eta_n^1)^t R^{-1} b_{n-1}^N \right] \\ &\quad \text{car } R^{-1} \text{ est symétrique et donc } (R^{-1})^t = R^{-1} \\ &= Z_n + \left[\left((\eta_n^0)^t - (\eta_n^1)^t \right) R^{-1} b_{n-1}^N \right] \\ &= Z_n + \left[(F(X_{n-1}, \theta_1) - F(X_{n-1}, \theta_0))^t R^{-1} b_{n-1}^N \right]. \end{aligned}$$

Ainsi,

$$\begin{aligned}
 \|\hat{Z}_n^N - Z_n\| &= \|(F(X_{n-1}, \theta_1) - F(X_{n-1}, \theta_0))^t R^{-1} b_{n-1}^N\| \\
 &\leq \|(F(X_{n-1}, \theta_1) - F(X_{n-1}, \theta_0))\| \cdot \|R^{-1}\| \cdot \|b_{n-1}^N\| \\
 &\leq 2M_1 \|R^{-1}\| \cdot \|b_{n-1}^N\|,
 \end{aligned} \tag{5.10}$$

où M_1 est la constante qui, par hypothèse, borne la fonction F .

La nature gaussienne du bruit rend les observations non bornées. Nous ne pouvons donc pas appliquer ici le résultat de convergence de l'estimateur à noyau de Senoussi énoncé dans la partie 5.2.1 puisqu'il s'agit d'un résultat de convergence sur des compacts. Nous allons alors utiliser le Théorème 5.4 de convergence sur les compacts dilatés.

On peut ainsi décomposer l'expression (5.10) en deux parties grâce à la suite v_N définie dans le Théorème 5.4 :

$$\begin{aligned}
 \|\hat{Z}_n^N - Z_n\| &\leq 2M_1 \|R^{-1}\| \cdot \|b_{n-1}^N\| \\
 &\leq 2M_1 \|R^{-1}\| \cdot \|f(X_{n-1}) - \hat{f}_N(X_{n-1})\| \mathbf{1}_{\{\|X_n\| \leq v_N\}} \\
 &\quad + 2M_1 \|R^{-1}\| \cdot \|f(X_{n-1}) - \hat{f}_N(X_{n-1})\| \mathbf{1}_{\{\|X_n\| > v_N\}}
 \end{aligned} \tag{5.11}$$

- En ce qui concerne la première partie de la décomposition (5.11) : $\forall n \geq 1$,

$$\begin{aligned}
 2M_1 \|R^{-1}\| \cdot \|f(X_{n-1}) - \hat{f}_N(X_{n-1})\| \mathbf{1}_{\{\|X_n\| \leq v_N\}} \\
 \leq 2M_1 \|R^{-1}\| \sup_{\|x\| \leq v_N} \|\hat{f}_N(x) - f(x)\| \rightarrow 0 \text{ lorsque } N \rightarrow \infty,
 \end{aligned}$$

sous réserve que les 2 hypothèses du théorème 5.4 soient vérifiées. Et elles le sont par hypothèse sur notre modèle.

- Occupons nous maintenant de la deuxième partie de la décomposition.

Puisque f est lipschitzienne et que la fonction F est continue et bornée, on a :

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

Alors,

$$\begin{aligned}
 X_n &= f(X_{n-1}) + F(X_{n-1}, \theta) + \epsilon_n \\
 \Rightarrow \|X_n\| &\leq \|f(X_{n-1}) + F(X_{n-1}, \theta)\| + \|\epsilon_n\| \\
 &\leq r_f \|X_{n-1}\| + M + \|\epsilon_n\|
 \end{aligned}$$

Et donc,

$$\begin{aligned}
 \|X_n\| &\leq r_f^n \|X_0\| + B + \sum_{j=1}^n r_f^{n-j} \|\epsilon_j\| \\
 \Rightarrow \|X_n\| &\leq r_f^n \|X_0\| + B + \sup_{k \leq n} \|\epsilon_k\| \cdot \left(\sum_{j=1}^n r_f^{n-j} \right) \\
 \Rightarrow \|X_n\| &\leq r_f^n \|X_0\| + B + \sup_{k \leq n} \|\epsilon_k\| \cdot \left(\sum_{j=0}^{n-1} r_f^j \right) \\
 \Rightarrow \|X_n\| &\leq r_f^n \|X_0\| + B + C \cdot \sup_{k \leq n} \|\epsilon_k\|
 \end{aligned}$$

où B et C sont des constantes. Par suite, on obtient :

$$\begin{aligned}
 \Rightarrow \|X_n\| &= O\left(\sup_{k \leq n} \|\epsilon_k\|\right) \quad \text{p.s.} \\
 \Rightarrow \sup_{k \leq n} \|X_k\| &= O\left(\sup_{k \leq n} \|\epsilon_k\|\right) \quad \text{p.s.}
 \end{aligned}$$

Ainsi, par le Lemme 5.1, on obtient :

$$\sup_{k \leq n} \|X_k\| = O(\sqrt{\log n}) \quad \text{p.s.} \tag{5.12}$$

Le résultat (5.12) indique donc,

$$\exists c > 0, \exists N(c) \text{ tel que, } \forall n \geq N(c), \quad \sup_{k \leq n} \|X_k\| \leq c \cdot \sqrt{\log n} \quad \text{p.s.} \tag{5.13}$$

Soit n fixé.

- 1^{er} cas : si $n < N(c)$

On applique (5.13) en définissant une constante $M_n(c)$ qui est égale, pour ce premier cas, à $N(c)$, définie en (5.13) :

$\exists c > 0, \exists M_n(c)(:= N(c))$ tel que :

$$\begin{aligned}
 \sup_{k \leq M_n(c)} \|X_k\| &\leq c \cdot \sqrt{\log M_n(c)} \quad \text{p.s.} \\
 \Rightarrow \sup_{k \leq n} \|X_k\| &\leq c \cdot \sqrt{\log M_n(c)} \quad \text{p.s.}
 \end{aligned}$$

- 2^{ième} cas : si $n \geq N(c)$

Chapitre 5. Le CUSUM non paramétrique avec apprentissage

On considère $M_n(c) \geq n$. Puisque $M_n(c) \geq N(c)$, on a par (5.13) :

$$\begin{aligned} \sup_{k \leq M_n(c)} \|X_k\| &\leq c \sqrt{\log M_n(c)} \quad \text{p.s.} \\ \Rightarrow \sup_{k \leq n} \|X_k\| &\leq c \sqrt{\log M_n(c)} \quad \text{p.s.} \end{aligned}$$

En résumé, on obtient :

$$\forall n, \exists c > 0, \exists M_n(c), \text{ tels que } \sup_{k \leq n} \|X_k\| \leq c \sqrt{\log M_n(c)} \quad \text{p.s.} \quad (5.14)$$

Soit $n \geq 1$ fixé. Puisque $v_N = A(\log \log N)^{1/2}$, si N est suffisamment grand (de sorte que $A(\log \log N)^{1/2} \geq c \sqrt{\log M_n(c)}$), alors,

$$\sup_{k \leq n} \|X_k\| \leq v_N \quad \text{p.s.}$$

Ainsi la deuxième partie de la décomposition (5.11) est nulle.

On obtient bien le résultat cherché, à savoir :

$$\forall n \geq 1 \quad \lim_{N \rightarrow \infty} |\hat{Z}_n^N - Z_n| = 0 \quad \text{p.s.} \quad \square$$

Corollaire 5.3 *Sous les hypothèses du Corollaire 5.2 et du Théorème 5.3, le CUSUM non paramétrique est asymptotiquement (sur γ et sur N) optimal au sens de Lorden.*

Démonstration : Les incréments de l'algorithme étant convergents, l'application du corollaire 4.1 nous permet d'obtenir le résultat. \square

5.4 Simulations

Dans cette partie, nous allons comparer en simulations le comportement du CUSUM et du CUSUM non paramétrique. Pour cela nous reprenons le modèle de dépollution simplifié :

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

θ représente ici le taux de colmatage de la pompe d'alimentation du procédé. μ est le taux de croissance de la biomasse, fonction du substrat. En pratique, la fonction μ est

souvent mal connue. Il existe de nombreuses lois dans la littérature pour modéliser ce taux de croissance (loi de Monod, loi de Teissier, etc... cf. Bastin et Dochain [8]). Le choix d'une loi n'est pas toujours commode. Ces lois font de plus intervenir des paramètres qu'il n'est pas facile d'estimer. Ces incertitudes de modélisation doivent être prises en compte par les algorithmes de détection d'anomalies. Dans le cadre de l'application du CUSUM non paramétrique, nous allons supposer que cette fonction μ est inconnue.

Nous allons nous intéresser à la deuxième équation du modèle qui concerne la concentration en substrat. On peut ré-écrire cette deuxième équation sous la forme :

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

avec,

$$H_n(S_n) = \frac{-TX_n}{\tau}, \quad \text{et} \quad F(S_n, U_n, \theta) = S_n + (1 - \theta)U_n(S_0 - S_n)T.$$

Il est possible, puisque nous connaissons les fonctions H et F , d'estimer la fonction μ à l'aide d'un estimateur à noyau qui aura la forme suivante :

$$\hat{\mu}(x) = \begin{cases} \frac{\sum_{i=1}^{N-1} k_i^{-1} K(k_i^{-1}(x-S_i)) H_i^{-1}(S_i) (S_{i+1} - F(S_i, U_i, \theta))}{\sum_{i=1}^{N-1} k_i^{-1} K(k_i^{-1}(x-S_i))} \\ 0 \quad \text{si le dénominateur est égal à 0.} \end{cases}$$

Nous allons effectuer deux types de simulations : dans un premier temps, nous allons comparer le CUSUM non paramétrique avec le CUSUM calculé à partir de la vraie fonction f , à savoir ici une loi de Monod :

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

Dans un deuxième temps, nous allons appliquer le CUSUM en supposant que le taux de croissance suit toujours une loi de Monod alors que nous aurons simulé notre modèle en ayant contaminé cette loi par une loi de Tessier.

Dans toute cette partie, nous choisissons pour K le noyau gaussien ($K(x) = (1/\sqrt{2\pi}) \cdot e^{-x^2/2}$) et pour la fenêtre : $\delta_i = 10 \cdot i^{-0.4}$.

5.4.1 Cas où f suit une loi de Monod

Panne abrupte

Dans la première simulation nous simulons une panne à l'instant $t_p = 6500$ en faisant passer la valeur de θ de 0 à 1 pour voir comment se comporte le CUSUM ainsi que le CUSUM utilisé avec l'estimation non paramétrique de la fonction μ_n .

Pour que l'estimation non paramétrique de f soit satisfaisante, nous considérons que les pas de temps de 1 jusqu'à 6000 représentent l'échantillon d'apprentissage de l'estimateur à noyau. Nous simulons aussi une panne (panne contrôlée) entre les instants $t = 4000$ et $t = 5000$ pour les raisons évoquées concernant l'échantillon d'apprentissage. La surveillance du procédé est lancée à partir de $t = 6001$.

Les résultats de cette simulation sont regroupés dans la figure 5.3 où les deux premiers graphes décrivent le comportement des observations (biomasse et substrat). On voit que les pannes (d'apprentissage et la panne à détecter) sont très brutales puisqu'on passe quasiment instantanément d'une concentration de 4mg/L à 0 pour le substrat. Les deux modèles sous H_1 et H_0 sont donc très différents et c'est pourquoi les deux tests détectent facilement la panne.

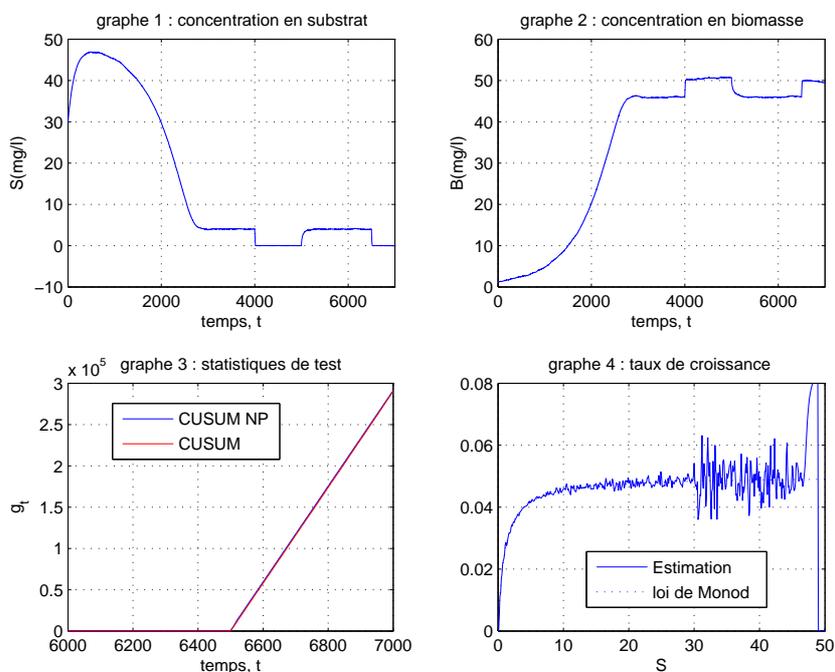


FIG. 5.3 – Comparaison des tests CUSUM et CUSUM non paramétrique

En effet, on peut voir sur le graphe 3 que les deux statistiques de tests sont pratiquement confondues et augmentent rapidement vers l'infini.

Panne "faible"

La panne de la figure 5.3 est très marquée si bien que les deux statistiques de test sont confondues. Pour comparer ces deux tests, il est donc nécessaire de simuler une panne moins abrupte.

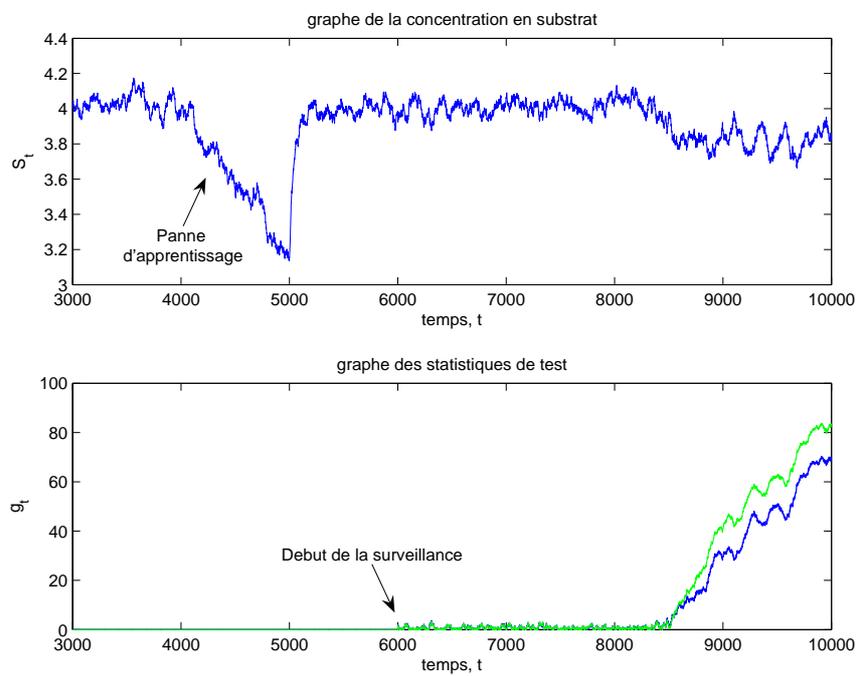


FIG. 5.4 – Comparaison des test CUSUM (en clair) et CUSUM non paramétrique (en foncé)

C'est l'objet de la simulation suivante (figure 5.4) où nous avons simulé le processus biotechnologique sur 10000 pas de temps. Nous utilisons à nouveau les 6000 premiers pas de temps pour calculer l'estimateur à noyau avec notamment une "panne d'apprentissage" entre $t = 4000$ et $t = 5000$ où nous avons fait croître de façon linéaire la valeur de θ pour la faire passer de 0 à 0.05. On a simulé une panne à l'instant $t_p = 8500$ avec : $\theta = \theta_1 = 0.01$. Il s'agit donc là, d'une panne très faible puisque cela revient à faire passer la valeur de la commande U_n de 0.04 à 0.0396.

A nouveau, les deux statistiques décollent simultanément et la détection s'effectue sans problème malgré un changement plus difficile à détecter. Il est ici intéressant de remarquer que l'échantillon d'apprentissage n'est pas seulement construit à partir des régimes H_0 et H_1 mais que nous avons, entre les pas de temps 4000 et 5000, fait varier la variable de commande de sorte à "balayer l'espace".

5.4.2 Cas où f est contaminée

Nous allons maintenant appliquer l'algorithme CUSUM classique à des données générées à partir du modèle de dépollution en supposant que μ suit une loi de Monod alors qu'en fait nous aurons simulé le modèle avec :

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

où $a \in [0; 1]$ sera le coefficient de mélange du taux de croissance. Les lois de Monod et de Tessier correspondent respectivement à

$$\mu^{Monod}(S_n) = \mu_{max} \frac{S_n}{K_s + S_n} \quad \text{et} \quad \mu^{Tessier}(S_n) = \mu_{max} \left(1 - \exp\left(\frac{-S_n}{K_s}\right) \right),$$

et sont représentées dans la figure 5.5.

Parallèlement, nous appliquerons notre algorithme CUSUM non paramétrique en utilisant l'estimation à noyau $\hat{\mu}$ en place de μ supposée inconnue. Nous simulons le système sur 10000 pas de temps en testant l'hypothèse $H_0 : \theta = \theta_0 = 0$ contre l'hypothèse $H_1 : \theta = \theta_1 = 0.01$ (léger colmatage de pompe caractérisé par une baisse de l'effet de la variable U_n). Pour que l'estimation non paramétrique soit satisfaisante, nous considérons que les pas de temps de $t = 1$ jusqu'à $t = 6000$ représentent l'échantillon d'apprentissage de l'estimateur à noyau avec à nouveau une "panne d'apprentissage" entre $t = 4000$ et $t = 5000$. Pour cette simulation, le taux de contamination est très faible et égal à :

$$a = 0.02,$$

et la panne à détecter intervient au pas de temps $t = 8500$. Les graphes 1 et 2 de la figure 5.6 décrivent le comportement du substrat et de la biomasse.

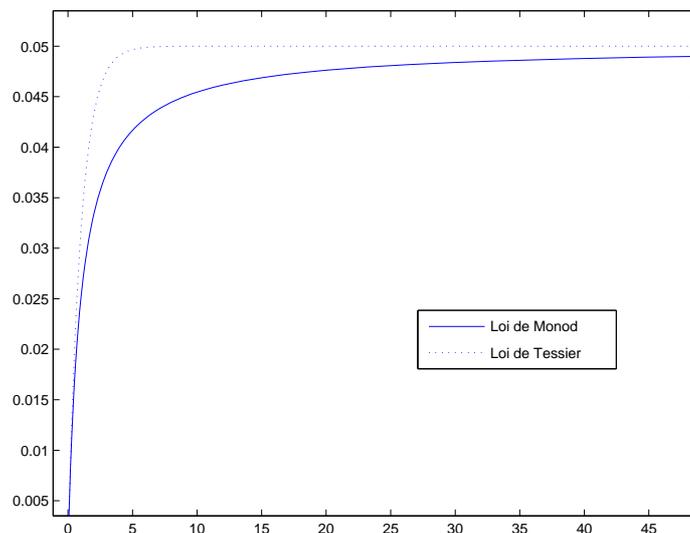


FIG. 5.5 – Représentation des lois de Monod et de Tessier

Pour comparer les deux tests, il est nécessaire de choisir un seuil pour chacun d’eux afin de comparer les retards à la détection. Le seuil de chaque test sera choisi de sorte que le temps moyen avant la première fausse alarme soit de 2000 pas de temps. Nous utilisons ici la méthode (3.1), présentée dans le chapitre 3, pour calculer ces seuils. Cette méthode donne les mêmes résultats que la méthode du seuil adaptatif (chapitre 3) compte tenu du régime pratiquement stationnaire atteint par le procédé après $t = 2500$. Pour le CUSUM, on obtient un seuil égal à 11 tandis que pour le CUSUM non paramétrique le seuil est égal à 2.9.

Nous voyons sur le graphe 3 de la figure 5.6 que le CUSUM, sous H_1 , augmente plus rapidement que le CUSUM non paramétrique. Cependant, son seuil élevé lui fait détecter la panne plus lentement que le CUSUM non paramétrique. En effet le CUSUM détecte la panne au temps $t = 8573$ alors que sa version non paramétrique la détecte au pas de temps $t = 8547$. Ces résultats trouvent leur explication dans le graphe 4 de la figure 5.6 qui représente la loi de Monod, l’estimation non paramétrique de μ_n utilisée pour le CUSUM non paramétrique et enfin la fonction μ (fonction de Monod contaminée). Nous avons choisi de présenter ces 3 fonctions dans l’intervalle $[3.6; 4.3]$ puisque il s’agit du voisinage dans lequel le substrat prend ses valeurs sous régime normal et sous régime de panne, et donc le voisinage dans lequel il est intéressant de comparer les trois fonctions.

Comme prévu, nous voyons alors que l’estimation non paramétrique de μ est plus proche du taux réel de croissance que le taux correspondant à la loi de Monod. De ce fait le test non paramétrique se comporte mieux que le CUSUM.

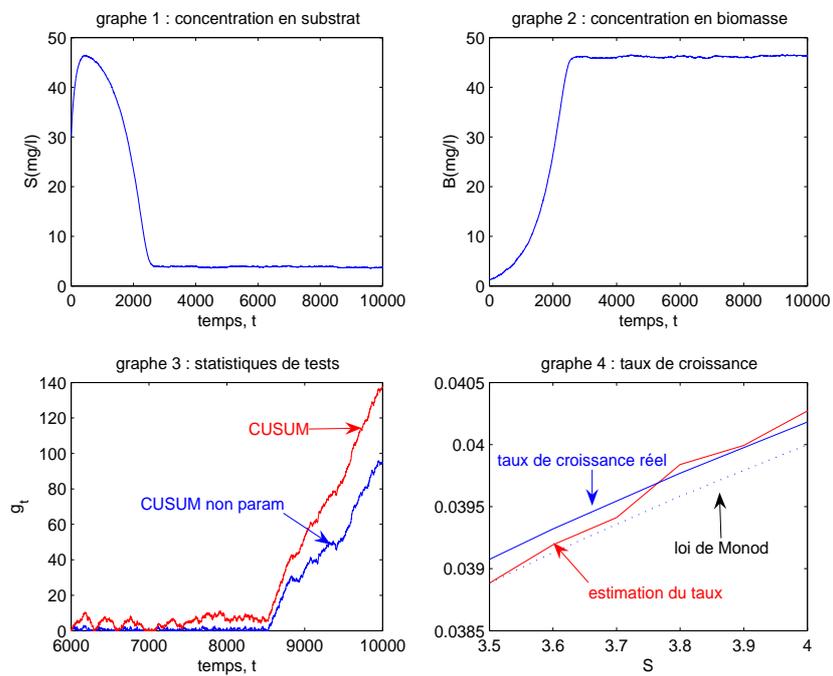


FIG. 5.6 – Comparaison entre les deux tests dans le cas d’une contamination de la loi de Monod. Panne simulée à l’instant $t_p = 8500$

Ces premiers résultats laissent penser que dans le cas d'une légère erreur de modélisation, la règle du CUSUM non paramétrique peut se montrer plus efficace que la règle du CUSUM. Nous nous proposons donc de calculer à partir d'un nombre important de simulations les retards moyens à la détection des deux tests pour des pannes plus ou moins prononcées et des taux de contamination plus ou moins élevés. Chaque valeur du tableau 5.1 est la moyenne des retards à la détection de 2000 tests identiques au précédent. Le seuil est à présent choisi de sorte que le temps moyen avant la première fausse alarme soit égal à 200 pas de temps (estimation réalisée à partir de 5000 répétitions). L'hypothèse de fonctionnement normal est toujours la même ($H_0 : \theta_0 = 0$), et on calcule le retard moyen pour différentes hypothèses alternatives $\theta_1 = 0.02, 0.01$ et 0.005 . On fait aussi varier le taux de contamination de $a = 0\%$ (pas de contamination, et donc situation idéale pour le CUSUM), jusqu'à un taux de contamination important de 5% .

Valeur de a	Tests	$\theta_1 = 0.02$	$\theta_1 = 0.01$	$\theta_1 = 0.005$
0%	t_{CUSUM}	15.28	34.00	66.08
	\hat{t}_{NP}	15.41	34.13	65.04
1%	t_{CUSUM}	15.38	35.64	68.05
	\hat{t}_{NP}	15.10	34.93	65.11
2%	t_{CUSUM}	15.55	40.80	81.09
	\hat{t}_{NP}	15.11	35.98	64.42
5%	t_{CUSUM}	27.01	82.45	129.42
	\hat{t}_{NP}	15.31	33.16	66.00

TAB. 5.1 – Retard moyen à la détection des algorithmes CUSUM et CUSUM non paramétrique pour différentes pannes et en fonction du taux de contamination de la loi de Monod.

Les résultats de cette série de simulations nous permettent de tirer un grand nombre d'enseignements. Lorsque le modèle de μ est la loi de Monod à paramètres connus exactement, c'est à dire lorsqu'il n'y a pas de contamination de la loi de Monod (1^{ière} ligne du tableau), le CUSUM se comporte bien. L'algorithme non paramétrique a un retard à la détection très proche de celui du CUSUM. Pour $\theta_1 = 0.005$, le CUSUM non paramétrique a un retard plus faible que le CUSUM mais cela doit être dû à un effet d'échantillonnage. Lorsque le taux de contamination est égal à 1% , on remarque que pour une panne assez marquée ($\theta_1 = 0.02$), les deux algorithmes se comportent à nouveau de façon identique. Mais plus la panne devient "fine", plus le CUSUM non paramétrique prend l'avantage sur sa version

paramétrique. Ce phénomène est encore plus marqué pour un taux de contamination de 2%. Pour une contamination de 5%, nous voyons clairement que le CUSUM non paramétrique est bien plus performant que le CUSUM standard.

La série de simulations qui vient d'être présentée montre bien que l'utilisation d'une méthode d'estimation non paramétrique est profitable dès lors qu'il y a une incertitude sur la modélisation du système considéré. Une modélisation exacte n'étant pas réalisable dans le cas de bioprocédés, il semble que cette méthode ait un grand intérêt dans la surveillance de systèmes biotechnologiques d'intérêt environnemental, systèmes pour lesquels il est possible d'avoir une grande base de données nécessaire à la construction de l'estimateur.

5.5 Conclusions

Dans ce chapitre, nous avons mis en place une procédure de détection, s'inspirant de la règle du CUSUM, sur des systèmes pour lesquels une partie de la modélisation est inconnue. La procédure consiste à remplacer, dans le calcul des incréments du CUSUM, la fonction f par une estimation à noyau de convolution, construite à partir d'un échantillon d'apprentissage. Nous avons obtenu, sous des hypothèses raisonnables en pratique, l'optimalité asymptotique au sens de Lorden de la procédure proposée. Le comportement en simulation de cette règle de décision laisse penser que son utilisation en pratique peut être bénéfique dans le cas d'incertitudes sur le modèle.

L'utilisation de l'estimateur à noyaux de convolution pour construire une règle de détection de changement, est également applicable à d'autres procédures telles que la règle du GLR, ou bien à un algorithme de diagnostic d'anomalie. On peut alors répondre à des situations plus fréquentes en pratique où les paramètres caractérisant les pannes sont inconnus.

Un travail supplémentaire est nécessaire pour le cas plus complexe où on souhaite utiliser les observations obtenues pendant la surveillance pour réactualiser l'estimateur à noyau. La question de l'optimalité asymptotique sur N (taille de l'échantillon d'apprentissage) se posant alors.

Chapitre 6

Algorithmes de filtrage-détection

6.1 Introduction

Dans ce chapitre, nous allons traiter d'un problème rencontré dans la plupart des applications : le cas où les variables d'intérêt du système, ou variables d'état, ne sont pas directement observées. Il faut alors se contenter de mesures subsidiaires, appelées variables d'observation. Dans le cas des bioréacteurs de dépollution, les concentrations en substrat et en biomasse jouent le rôle des variables d'état. Les coûts de mise en place et de maintenance des capteurs pour ce type de quantités étant élevés, les expérimentateurs se contentent souvent de mesurer des grandeurs indirectement liées à ces concentrations, comme les débits de gaz ou le pH à l'intérieur du bioréacteur.

Ce type de situation définit le problème du filtrage, dont l'objectif est de reconstruire l'état X grâce aux observations Y à partir de la modélisation en système dynamique suivante :

$$\begin{cases} X_n = f(X_{n-1}, \theta, v_n) \\ Y_n = g(X_n, \theta, w_n) \end{cases} \quad (6.1)$$

où les $X_n \in \mathbb{R}^d$ et $Y_n \in \mathbb{R}^q$ sont respectivement les variables d'état et d'observation, v_n et w_n des bruits sur le système, et θ un paramètre qui, dans notre cas, caractérise la panne.

Les méthodes statistiques de filtrage permettent d'obtenir ou d'estimer la densité conditionnelle de l'état $p_\theta(x_n|Y_{1:n})$, appelée : filtre optimal. Dans le cas de systèmes linéaires avec bruits blancs gaussiens additifs, le filtre de Kalman fournit le filtre optimal. Dès que le système est non linéaire ou que les bruits ne sont plus gaussiens, le filtre de Kalman (ou plutôt le filtre de Kalman étendu) n'est plus aussi efficace et il vaut mieux utiliser d'autres outils (voir par exemple Rossi [90], où une comparaison est effectué entre le filtre de Kalman étendu, les filtres particuliers usuels et le filtre à noyaux de convolution).

Chapitre 6. Algorithmes de filtrage-détection

Les méthodes de filtrage particulaire ont été introduites récemment [19, 36, 46]. L'approche générale est la suivante. A partir d'un grand nombre de simulations de l'état à chaque pas de temps appelées particules, le filtre permet de reconstituer le filtre optimal ou encore la densité prédictive $p_\theta(x_n|Y_{1:n-1})$, sous la forme d'une mesure empirique (à chaque particule est associé un poids correspondant à la vraisemblance de la particule au vue des observations).

Nous allons utiliser dans ce chapitre, une méthode de type filtrage particulaire originale, qui permet, non pas d'obtenir une mesure empirique de la distribution recherchée, mais une estimation fonctionnelle de la densité de probabilité, obtenue par noyaux de convolution. Ce filtre particulaire à noyaux de convolution a été mis au point par Rossi et Vila [91] et il présente l'avantage de posséder des propriétés théoriques intéressantes, notamment la convergence L_1 presque sûre de l'estimateur vers le filtre optimal. Il possède de plus des conditions d'utilisation plus souples que les filtres particuliers usuels.

Dans l'annexe B, nous rappelons brièvement le fonctionnement des filtres à particules usuels ainsi que celui du filtre à noyaux de convolution et ses résultats de convergence.

*

* *

Une anomalie sur un système décrit par un modèle à espace d'état sera caractérisée par un changement de paramètre dans l'équation d'état ou dans l'équation d'observation ou plus généralement, comme on le verra dans le chapitre 7 sur le traitement des données réelles, par un changement total de structure du modèle. Nous avons évoqué dans le chapitre 2 quelques méthodes proposées ces dernières années pour traiter le problème de la détection de changement dans les modèles à espace d'état.

Lorsque le modèle est linéaire avec bruits blancs gaussiens additifs et la panne additive, l'approche la plus pertinente consiste à utiliser la règle de Willsky et Jones (2.21) [112]. Dans des cas plus complexes, comme les modèles non linéaires, l'utilisation de filtres à particules donne de meilleurs résultats.

L'approche que nous avons développée s'inspire des travaux de Li et Kadiramanathan [58, 59]. Nous cherchons à détecter un changement dans la densité conditionnelle des observations. Pour obtenir cette densité, il est nécessaire d'utiliser une méthode de type filtrage. A partir de la méthode de filtrage de Rossi et Vila [91], nous obtenons l'optimalité asymptotique de la procédure proposée sous les hypothèses de stabilité classiques vues au chapitre 4, pour le cas où les valeurs caractérisant le changement, θ_0 et θ_1 , sont connues.

Nous étendons ensuite la procédure pour traiter le cas où le paramètre θ_1 est inconnu et nous obtenons l'optimalité asymptotique dans le cas où la variable d'état

est directement observée. Cela revient à avoir dans le modèle (6.1) pour équation d'observation : $Y_n = X_n$. Nous sommes alors dans le cadre d'utilisation de la règle du GLR, et nous montrons en simulation, que notre règle de décision, plus rapide en temps de calcul, est aussi efficace que le GLR.

Nous abordons ensuite une situation importante en pratique, le diagnostic dans les modèles à espace d'état.

6.2 Le cas général

6.2.1 La modélisation et ses conséquences

Nous considérons le modèle (6.1) et nous nous plaçons sous les hypothèses suivantes :

- π_0 , la distribution de probabilité de l'état initial X_0 est connue.
- Les fonctions f et g sont connues ainsi que les distributions \mathcal{L}_v et \mathcal{L}_w des bruits v et w de sorte à connaître analytiquement la densité conditionnelle des observations : $p_\theta(y_n|X_n)$ qui sera strictement positive. Cette hypothèse peut être assouplie (voir annexe B).
- Sous le fonctionnement normal du système, le paramètre θ vaut θ_0 , et en régime de panne, il vaut θ_1 . Ces deux valeurs sont connues.
- Nous supposons que les mesures de probabilité π_n et $\pi_{n|n-1}$, respectivement de $X_n|Y_{1:n}$ et $X_n|Y_{1:n-1}$, admettent des densités notées $p_\theta(x_n|Y_{1:n})$ et $p_\theta(x_n|Y_{1:n-1})$.

Pour détecter le changement de valeur du paramètre θ dans le modèle (6.1), nous allons chercher à détecter un changement dans la densité conditionnelle des observations $p_\theta(y_n|Y_{1:n-1})$. Généralement, il n'est pas possible d'écrire sous forme analytique cette densité. Nous allons donc l'estimer sous les deux hypothèses de fonctionnement : H_0 et H_1 . A partir de là, nous mettrons en place la procédure du CUSUM approché proposée dans le chapitre 4.

REMARQUE : Compte tenu de la modélisation du système, le changement de régime est généré suivant la première méthode décrite dans le chapitre 2, cas le plus général où la densité conditionnelle dépend de l'instant de changement de régime. Pour s'en convaincre, considérons le modèle linéaire suivant :

$$\begin{cases} X_n = A_n X_{n-1} + B_n w_n \\ Y_n = C_n X_n + D_n v_n \end{cases} \quad (6.2)$$

Le filtre de Kalman permet d'obtenir la vraie densité $p(x_n|Y_{1:n})$ et cette densité dépend de la suite des matrices A_n , B_n , C_n et D_n , qui dépendent évidemment de

l'indice n . Dans notre cas (modèle (6.1)), la densité $p_\theta(x_n|Y_{1:n})$ va donc dépendre, sous H_1 , de l'instant de changement de régime, le paramètre θ , passant de la valeur θ_0 à θ_1 , jouant le rôle d'une matrice qui varie à un instant t_p dans l'exemple ci-dessus.

6.2.2 Estimation de la vraisemblance par intégration de Monte Carlo

Revenons tout d'abord sur la méthode d'intégration de Monte Carlo. Supposons que la quantité suivante :

$$I(\psi) = \int \psi(y)p(y) dy,$$

où p est une densité de probabilité et ψ une fonction de y , est difficile à intégrer. On peut estimer cette intégrale, de façon consistante (loi des grands nombres), à partir d'un échantillon aléatoire $(y_k)_{1 \leq k \leq m}$ tiré de p en calculant la somme :

$$\frac{1}{m} \sum_{k=1}^m \psi(y_k) \xrightarrow{m \rightarrow \infty} I(\psi) \quad \text{p.s.}$$

C'est à partir de cette méthode que nous allons estimer les vraisemblances conditionnelles des observations sous H_0 et sous H_1 .

Sous H_0 , hypothèse de fonctionnement normal

A partir du modèle (6.1), et pour $\theta = \theta_0$, la vraisemblance conditionnelle des observations est donnée par l'intégrale suivante :

$$\begin{aligned} p_{\theta_0}(Y_n|Y_{1:n-1}) &= \int p_{\theta_0}(Y_n|Y_{1:n-1}, x_n) \cdot p_{\theta_0}(x_n|Y_{1:n-1}) dx_n \\ &= \int p_{\theta_0}(Y_n|x_n) \cdot p_{\theta_0}(x_n|Y_{1:n-1}) dx_n . \end{aligned} \quad (6.3)$$

REMARQUE : Une caractéristique des modèles à espace d'état de type (6.1) est que l'observation à l'instant n est indépendante des observations précédentes conditionnellement à la variable d'état X_n .

Par hypothèse sur le modèle, on peut écrire la densité conditionnelle $p_{\theta_0}(y_n|X_n)$. On n'a pas accès à la densité conditionnelle (qui est en fait une densité prédictive) $p_{\theta_0}(x_n|Y_{1:n-1})$ mais on peut l'estimer grâce à des méthodes de prédiction de type filtrage particulière (détaillées dans les parties suivantes). En tirant de façon aléatoire

un échantillon $(x_{n,0}(k))$, $k = 1, \dots, m$ de cette estimation $\hat{p}_{\theta_0}(x_n|Y_{1:n-1})$, on peut approcher la vraisemblance des observations par intégration de Monte Carlo :

$$\hat{l}_{0,n}^{N,m} = \frac{1}{m} \sum_{k=1}^m p_{\theta_0}(Y_n|x_{n,0}(k)) \simeq p_{\theta_0}(Y_n|Y_{1:n-1}). \quad (6.4)$$

L'estimateur $\hat{l}_{0,n}^{N,m}$, dépend de m , taille de l'échantillon pour l'intégration de Monte Carlo, et de N , le nombre de particules utilisées pour estimer la densité prédictive à partir de la méthode de filtrage considérée.

Sous H_1 , régime de panne

Comme nous l'avons remarqué dans la partie 6.2.1, la vraisemblance conditionnelle sous l'hypothèse H_1 dépend de l'instant de changement de régime, qui est inconnu. Il est alors nécessaire, pour construire le CUSUM approché, d'estimer cette vraisemblance pour tous les instants possibles de changement j , compris entre 1 et n .

L'estimation de la vraisemblance conditionnelle à l'instant n sous l'hypothèse d'un changement de régime à un instant j (compris entre 1 et n) va s'effectuer de la même manière que sous H_0 . En effet, on a :

$$p_{\theta_0, \theta_1, j}(Y_n|Y_{1:n-1}) = \int p_{\theta_1}(Y_n|x_n) \cdot p_{\theta_0, \theta_1, j}(x_n|Y_{1:n-1}) dx_n, \quad (6.5)$$

où $p_{\theta_1}(y_n|X_n)$ est connue d'après les hypothèses sur le modèle. La vraisemblance sera donc estimée par intégration de Monte Carlo :

$$\hat{l}_{1,n,j}^{N,m} = \frac{1}{m} \sum_{k=1}^m p_{\theta_1}(Y_n|x_{n,1,j}(k)) \simeq p_{\theta_0, \theta_1, j}(Y_n|Y_{1:n-1}), \quad (6.6)$$

où $x_{n,1,j}(k)$ sera un échantillon aléatoire tiré à partir de l'estimation de $p_{\theta_0, \theta_1, j}(x_n|Y_{1:n-1})$. Cette estimation est obtenue grâce à une méthode de type filtrage particulière appliquée à la modélisation suivante : pour $i = 1, \dots, j-1$, on utilisera le modèle (6.1) avec $\theta = \theta_0$ et à partir de l'instant j , avec $\theta = \theta_1$.

REMARQUE : Selon le modèle (6.1), la densité conditionnelle $p_{\theta_1}(y_n|X_n)$ ne dépend pas de l'instant de changement j . On pourrait, pour rester le plus général possible, envisager une équation d'observation évoluant au cours du temps sous régime de panne.

Le filtre à noyaux de convolution pour estimer la densité prédictive

Les estimations des densités $p_{\theta_0}(x_n|Y_{1:n-1})$ et $p_{\theta_0, \theta_{1,j}}(x_n|Y_{1:n-1})$ seront obtenues à partir de l'algorithme de filtrage à noyaux de convolution de Rossi et Vila [91] présenté en annexe B.

L'estimation du filtre optimal $p_{\theta}(x_n|Y_{1:n})$ est obtenue à partir de N couples de particules $(\tilde{X}_n, \tilde{Y}_n)$ ayant évolué suivant le modèle à espace d'état (6.1) et s'écrit :

$$\hat{p}_{\theta}(x_n|Y_{1:n}) = \frac{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_n(i) - Y_n) \cdot K_{\delta_N}^x(\tilde{X}_n(i) - x_n)}{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_n(i) - Y_n)}, \quad (6.7)$$

où K^x et K^y sont des noyaux de convolution et δ_N leurs fenêtres associées. Nous adoptons l'écriture usuelle en statistique non paramétrique : $K_{\delta_N}(x) = K(x/\delta_N)/\delta_N^d$ si $x \in \mathbb{R}^d$. L'estimateur de la densité prédictive s'écrit de façon analogue.

Le théorème suivant est dû à Rossi et Vila [91]. Il assure la convergence du filtre à noyaux de convolution vers le filtre optimal.

Théorème 6.1 (Rossi et Vila [91]) *Si K^x et K^y sont des noyaux de Parzen-Rosenblatt positifs et bornés, si $p_{\theta}(\cdot|Y_{1:n-1})$ est positive et continue en y_n pour tout n , s'il existe $M > 0$ tel que $0 < p_{\theta}(y_t|X_t) \leq M_2$ pour tout n et s'il existe $\alpha \in]0; 1[$ tel que $N(\delta_N)^{2q} = O(N^{\alpha})$, alors*

$$\begin{cases} \lim_{N \rightarrow \infty} \frac{N\delta_N^{d+q}}{\log N} = \infty \\ \lim_{N \rightarrow \infty} \delta_N = 0 \end{cases} \Rightarrow \lim_{N \rightarrow \infty} \int |\hat{p}_{\theta}(x_n|Y_{1:n}) - p_{\theta}(x_n|Y_{1:n})| dx_n = 0 \quad \text{p.s.}$$

Les propriétés théoriques du filtre à convolution étant les mêmes pour l'estimation de la densité prédictive que pour l'estimation du filtre optimal, nous avons aussi la convergence suivante sous les mêmes hypothèses :

$$\lim_{N \rightarrow \infty} \int |\hat{p}_{\theta}(x_n|Y_{1:n-1}) - p_{\theta_1}(x_n|Y_{1:n-1})| dx_n = 0 \quad \text{p.s.} \quad (6.8)$$

REMARQUE : Les conditions d'utilisation du filtre à noyaux de convolution sont plus souples que celles des filtres usuels. Il n'est notamment pas nécessaire de connaître analytiquement la densité conditionnelle $p_{\theta}(y_n|X_n)$. Mais pour estimer la vraisemblance conditionnelle des observations par intégration de Monte Carlo, nous avons besoin de cette écriture analytique. Cette connaissance est donc supposée ici. Cependant, Rossi [90] propose d'utiliser un estimateur à noyaux de la vraisemblance conditionnelle des observations qui ne nécessite pas l'écriture analytique de la densité $p_{\theta}(y_n|X_n)$. En simulation, les résultats obtenus se révèlent légèrement moins bons que l'estimation par intégration de Monte Carlo. C'est pourquoi nous avons choisi d'utiliser cette dernière tout en sachant qu'il existe une alternative pour des cas où, par exemple, la loi du bruit n'est pas connue mais juste simulable, et donc l'écriture de $p_{\theta}(y_n|X_n)$ inconnue.

6.2.3 Principe de l'algorithme et résultats

Nous allons considérer tout un ensemble de modèles qui vont caractériser tous les comportements possibles du système en régime normal ou non. Ainsi, un premier modèle décrira le fonctionnement normal du système :

$$\begin{cases} X_n = f(X_{n-1}, \theta_0, v_n) \\ Y_n = g(X_n, \theta_0, w_n). \end{cases} \quad (6.9)$$

Et à l'instant n , nous allons considérer n modèles, chacun relié à un instant possible de changement j compris entre 1 et n :

modele j : $1 \leq j \leq n$

$$\begin{cases} X_i = f(X_{i-1}, \theta_0, v_i) \\ Y_i = g(X_i, \theta_0, w_i) , \quad \text{pour } i < j \end{cases} \quad (6.10)$$

$$\begin{cases} X_i = f(X_{i-1}, \theta_1, v_i) \\ Y_i = g(X_i, \theta_1, w_i) , \quad \text{pour } i \geq j. \end{cases} \quad (6.11)$$

Grâce aux algorithmes de filtrage mis en parallèle et aux estimations de Monte Carlo, nous allons estimer les vraisemblances conditionnelles sous H_0 et sous H_1 (à l'instant n , une sous H_0 et n sous H_1), estimations obtenues à partir de (6.4) et (6.6) et notées $\hat{l}_{0,n}^{N,m}$ et $\hat{l}_{1,n,j}^{N,m}$, pour $1 \leq j \leq n$. Le théorème suivant assure la consistance de ces estimations :

Théorème 6.2 *Sous les hypothèses décrites dans la partie 6.2.1 sur le modèle et sous les hypothèses du Théorème 6.1 assurant la convergence du filtre à noyaux, on a lorsque m et N tendent vers l'infini :*

$$\hat{l}_{0,n}^{N,m} \rightarrow p_{\theta_0}(Y_n | Y_{1:n-1}) \text{ p.s.} \quad (6.12)$$

$$\hat{l}_{1,n,j}^{N,m} \rightarrow p_{\theta_0, \theta_1, j}(Y_n | Y_{1:n-1}, j) \text{ p.s.} \quad \forall 1 \leq j \leq n. \quad (6.13)$$

Démonstration : Nous montrons la convergence de $\hat{l}_{1,n,j}^{N,m}$ vers $p_{\theta_0, \theta_1, j}(Y_n | Y_{1:n-1})$. Le résultat pour $\hat{l}_{0,n}^{N,m}$ s'obtient de la même façon.

La vraisemblance conditionnelle des observations est donnée par l'intégrale suivante :

$$p_{\theta_0, \theta_1, j}(Y_n | Y_{1:n-1}) = \int p_{\theta_1}(Y_n | x_n) \cdot p_{\theta_0, \theta_1, j}(x_n | Y_{1:n-1}) dx_n .$$

On a défini

$$\hat{l}_{1,n,j}^{N,m} = \frac{1}{m} \sum_{i=1}^m p_{\theta_1}(Y_n | x_{n,1,j}(i)),$$

Chapitre 6. Algorithmes de filtrage-détection

où $(x_{n,1,j}(i), i = 1, \dots, m, j \leq i)$ est un échantillon aléatoire tiré de l'estimateur $\hat{p}_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1})$ de la densité prédictive obtenue par le filtre à noyaux de convolution.

Par la loi forte des grands nombres :

$$\hat{l}_{1,n,j}^{N,m} \xrightarrow{m \rightarrow \infty} \int p_{\theta_1}(Y_n | x_n) \cdot \hat{p}_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1}) dx_n \quad \text{p.s.} \quad (6.14)$$

De plus,

$$\begin{aligned} & \left| \int p_{\theta_1}(Y_n | x_n) \cdot \hat{p}_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1}) dx_n - \int p_{\theta_1}(Y_n | x_n) \cdot p_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1}) dx_n \right| \\ &= \left| \int p_{\theta_1}(Y_n | x_n) \cdot (\hat{p}_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1}) - p_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1})) dx_n \right| \\ &\leq M_2 \int |\hat{p}_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1}) - p_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1})| dx_n \\ &\rightarrow 0 \quad \text{p.s.} \quad \text{lorsque } N \rightarrow \infty, \end{aligned}$$

grâce au Théorème 6.1.

En résumé, on obtient donc :

$$\begin{aligned} \hat{l}_{1,n,j}^{N,m} &\xrightarrow[\text{p.s.}]{m \rightarrow \infty} \int p_{\theta_1}(Y_n | x_n) \cdot \hat{p}_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1}) dx_n \\ &\quad \text{p.s.} \downarrow N \rightarrow \infty \\ &\int p_{\theta_1}(Y_n | x_n) \cdot p_{\theta_0, \theta_{1,j}}(x_n | Y_{1:n-1}) dx_n \\ &= p_{\theta_0, \theta_{1,j}}(Y_n | Y_{1:n-1}) \end{aligned}$$

et donc,

$$\hat{l}_{1,n,j}^{N,m} \xrightarrow{N, m \rightarrow \infty} p_{\theta_0, \theta_{1,j}}(Y_n | Y_{1:n-1}) \quad \text{p.s.} \quad \square$$

Corollaire 6.1 *Sous les hypothèses sur le modèle effectuées dans la partie 6.2.1, sous les hypothèses du théorème 6.1, et si les conditions de stabilité (4.12) et (4.13) sont vérifiées, alors la règle de décision définie de la façon suivante :*

$$\hat{t} = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{\hat{l}_{1,i,j}^{N,m}}{\hat{l}_{0,i}^{N,m}} \geq h \right\}, \quad (6.15)$$

est asymptotiquement optimale (sur N , m et h) au sens de Lorden.

Démonstration :

La démonstration est immédiate. Les estimations des vraisemblances conditionnelles $\hat{l}_{0,n}^{N,m}$ et $\hat{l}_{1,n,j}^{N,m}$ étant convergentes, l'algorithme de filtrage-détection est donc, grâce au corollaire 4.1, asymptotiquement optimal au sens de Lorden. \square

6.2.4 Notion d'optimalité dans les modèles à espace d'état

Nous avons considéré le modèle (6.1) où les variables X_n ne sont pas mesurées et où seules les variables Y_n sont à notre disposition. Comment définir dans ce cas l'optimalité d'un algorithme de détection ? L'approche que nous utilisons, puisque nous effectuons un rapport de vraisemblance conditionnelle des observations, suppose que tout changement sur le modèle à espace d'état, que ce soit dans l'équation d'état ou dans l'équation d'observation, se traduit par un changement de paramètre θ dans la densité conditionnelle des observations $p_\theta(y_n|Y_{1:n-1})$ (ou plus généralement un changement de structure de la densité conditionnelle).

Cependant, est-il pertinent de considérer qu'une anomalie sur le procédé considéré, et donc un changement de paramètre θ sur le modèle précédent, va entraîner un changement dans la densité conditionnelle des observations ? A priori la réponse est oui, un changement dans la densité conditionnelle devrait avoir lieu.

6.2.5 L'algorithme de filtrage-détection

Nous détaillons dans cette partie le fonctionnement de notre algorithme de filtrage-détection. Cet algorithme se décompose en quatre étapes : les étapes 1- et 4- représentent le fonctionnement classique du filtre à noyaux de convolution de Rossi et Vila [91]. Nous avons rajouté l'étape 2- pour estimer la vraisemblance conditionnelle des observations. L'étape 3- correspond à la procédure de test. Nous détaillons dans les étapes 1-, 2- et 4- ci-dessous, les calculs correspondant au modèle (6.9) décrivant le fonctionnement normal du système. Les mêmes étapes sont à réaliser pour tous les autres modèles (6.10)-(6.11) caractérisant le passage sous régime de panne du système, dans lesquels les particules du filtre sont d'abord générées avec $\theta = \theta_0$ jusqu'à l'instant supposé de changement j , puis avec $\theta = \theta_1$.

Initialisation : $t = 1$

- Génération de N réalisations $\bar{X}_0(i)$, $i = 1, 2, \dots, N$ grâce à π_0 .
- Génération de N réalisations des bruits $v_1(i)$ et $w_1(i)$, $i = 1, 2, \dots, N$ grâce aux lois \mathcal{L}_v et \mathcal{L}_w supposées connues.

- Le modèle (6.9) conduit aux réalisations :

$$\begin{aligned}\tilde{X}_{1,H_0}(i) &= f(\bar{X}_0(i), \theta_0, v_1(i)) \\ \tilde{Y}_{1,H_0}(i) &= g(\tilde{X}_{1,H_0}(i), \theta_0, w_1(i)) , \quad i = 1, \dots, N.\end{aligned}$$

- A partir de l'observation Y_1 , les N couples de particules $(\tilde{X}_{1,H_0}(i), \tilde{Y}_{1,H_0}(i))$ permettent la construction d'une estimation par noyaux $\hat{p}_{\theta_0}(x_1|Y_1)$ de la densité conditionnelle $p_{\theta_0}(x_1|Y_1)$ selon (6.7).

Chapitre 6. Algorithmes de filtrage-détection

Temps t : $t > 1$

Etape 1 :

- A partir de l'estimation obtenue à l'instant $t - 1$, génération de :
 $\bar{X}_{t-1,H_0}(i) \sim \hat{p}_{\theta_0}(x_{t-1}|Y_{1:t-1})$, $v_t(i) \sim \mathcal{L}_v$, $w_t(i) \sim \mathcal{L}_w$ pour $i = 1, \dots, N$.

- Le modèle (6.9) conduit à :

$$\begin{aligned}\tilde{X}_{t,H_0}(i) &= f(\bar{X}_{t-1,H_0}(i), \theta_0, v_t(i)) \\ \tilde{Y}_{t,H_0}(i) &= g(\tilde{X}_{t,H_0}(i), \theta_0, w_t(i)), \quad i = 1, \dots, N.\end{aligned}$$

Etape 2 :

- A partir des particules $\tilde{X}_{t,H_0}(i)$, $\tilde{Y}_{t-1,H_0}(i)$, $i = 1, \dots, N$, et de l'observation Y_{t-1} , on peut construire un estimateur à noyaux convergent de la densité prédictive $p_{\theta_0}(x_t|Y_{1:t-1})$:

$$\hat{p}_{\theta_0}(x_t|Y_{1:t-1}) = \frac{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{t-1,H_0}(i) - Y_{t-1}) \cdot K_{\delta_N}^x(\tilde{X}_{t,H_0}(i) - x_t)}{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{t-1,H_0}(i) - Y_{t-1})}.$$

- On peut alors estimer la vraisemblance conditionnelle des observations sous l'hypothèse H_0 dès que la mesure Y_t est disponible : il suffit de générer un échantillon de m particules selon la densité prédictive estimée $\hat{p}_{\theta_0}(x_t|Y_{1:t-1})$, que l'on notera $(x_{t,0}(i), i = 1, \dots, m)$. La vraisemblance des observations est alors estimée par :

$$\hat{p}_{\theta_0}(Y_t|Y_{1:t-1}) = \hat{l}_{0,t}^{N,m} = \frac{1}{m} \sum_{i=1}^m p_{\theta_0}(Y_t|x_{t,0}(i)).$$

REMARQUE : Parallèlement, on estime les vraisemblances $p_{\theta_0,\theta_{1,j}}$ par $\hat{l}_{1,t,j}^{N,m}$ pour $j = 1, \dots, t$ à partir d'un échantillon $(x_{t,1,j}(i), i = 1, \dots, m)$ tiré de l'estimation $\hat{p}_{\theta_0,\theta_{1,j}}(x_t|Y_{1:t-1})$.

Etape 3 :

A partir des quantités $\hat{l}_{0,t}^{N,m}$ et $\hat{l}_{1,t,j}^{N,m}$, pour $j = 1, \dots, t$ obtenues à l'étape 2-, on peut calculer la statistique de test suivante :

$$\hat{g}_t = \max_{1 \leq j \leq t} \sum_{i=j}^t \hat{Z}_{i,j}^{N,m} \quad \text{avec} \quad \hat{Z}_{i,j}^{N,m} = \log \frac{\hat{l}_{1,i,j}^{N,m}}{\hat{l}_{0,i}^{N,m}},$$

et une alarme est déclenchée si la statistique de test \hat{g}_t dépasse un seuil h fixé par l'expérimentateur. Si l'alarme n'est pas déclenchée, on passe alors à l'étape 4.

Etape 4 :

- A partir des particules $\tilde{X}_{t,H_0}(i), \tilde{Y}_{t,H_0}(i), i = 1, \dots, N$, et de l'observation Y_t , nous estimons le filtre optimal par l'estimateur à noyau :

$$\hat{p}_{\theta_0}(x_t|Y_{1:t}) = \frac{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{t,H_0}(i) - Y_t) \cdot K_{\delta_N}^x(\tilde{X}_{t,H_0}(i) - x_t)}{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{t,H_0}(i) - Y_t)}.$$

REMARQUE : On estime de même, à partir de particules issues des modèles (6.10)-(6.11), le filtre optimal $p_{\theta_0, \theta_1, j}(x_t|Y_{1:t})$ pour $j = 1, \dots, t$ par un estimateur à noyaux.

- $t = t + 1$ et retour à l'étape 1.

6.2.6 Problème de l'accroissement du temps de calculs

Un problème se pose rapidement lorsqu'on met en place l'algorithme de filtrage-détection tel que nous venons de le décrire dans la partie 6.2.5. Les calculs croissent de façon régulière, puisqu'à chaque pas de temps, on considère un filtre supplémentaire. Nous avons envisagé deux méthodes pour réduire les calculs : l'utilisation d'une fenêtre glissante, méthode qui est la plus rigoureuse, et une approximation grossière qui donne tout de même des résultats satisfaisants en simulations.

Algorithme de filtrage-détection à fenêtre limitée

Plusieurs solutions sont utilisées dans la littérature pour diminuer la quantité de calculs. Dans [2], les auteurs, qui n'utilisent pas un filtre à noyaux de convolution mais un filtre particulaire classique, proposent d'utiliser une statistique de test légèrement différente afin de retrouver une écriture récursive de cette statistique de test. Li et Kadiramanathan [59] et Andrieu et al [1] proposent eux de travailler sur une fenêtre temporelle de taille M . Ainsi, à un instant t , on ne considère que les $M + 1$ filtres reliés aux hypothèses où $\theta = \theta_0$ pour $i < j$ et $\theta = \theta_1$ pour $i \geq j$ pour $j = t - M, \dots, t$, en plus du filtre sous H_0 . Cette méthode nous semble être la plus adaptée à condition de prendre, en pratique, une fenêtre suffisamment grande pour permettre de détecter les changements. En effet, comme le montre le petit exemple dans la figure 6.1, où on applique un CUSUM à fenêtre limitée sur un modèle très simple, la statistique de test de l'algorithme est croissante sous H_1 en moyenne pendant M pas de temps puis, a tendance à être constante. Il faut donc choisir une fenêtre suffisamment grande pour que la statistique de test puisse franchir le seuil fixé.

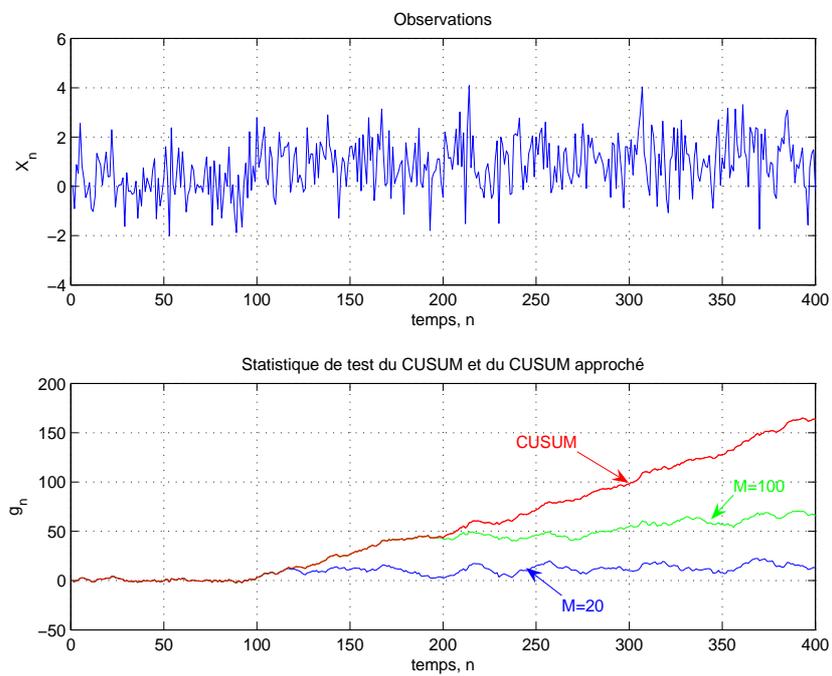


FIG. 6.1 – Mise en place du test CUSUM et du test CUSUM à fenêtre limitée (pour $M = 20$ et $M = 100$) sur des observations simulées suivant une loi normale $N(0;1)$ avant l'instant de panne $t_p = 100$, et suivant une loi $N(1;1)$ après l'instant de panne.

REMARQUE : Lorsqu'on s'intéresse à l'optimalité asymptotique du CUSUM approché à fenêtre limitée (chapitre 4), en raison du seuil h qui tend vers l'infini, on est obligé de considérer une fenêtre M , qui tend elle aussi, vers l'infini. C'est aussi le cas dans les deux corollaires qui suivent.

La statistique de test de l'algorithme de filtrage-détection à fenêtre limitée s'écrit alors sous la forme suivante :

$$\hat{t}_{lim} = \inf \left\{ n : \max_{n-M \leq j \leq n} \sum_{i=j}^n \log \frac{\hat{l}_{1,i,j}^{N,m}}{\hat{l}_{0,i}^{N,m}} \geq h \right\}, \quad (6.16)$$

au lieu de l'écriture (6.15), difficile à en mettre en place en pratique. On a alors les deux résultats d'optimalité suivants :

Corollaire 6.2 *Sous les hypothèses du Corollaire 6.1, la règle de décision \hat{t}_{lim} définie en (6.16) est asymptotiquement optimale au sens de Lorden lorsque N , m et h tendent vers l'infini et que la fenêtre M_α (dynamique) est choisie telle que :*

$$\liminf \frac{M_\alpha}{|\log \alpha|} > I^{-1} \quad \text{mais} \quad \log M_\alpha = o(\log \alpha) \quad \text{quand} \quad |\log \alpha| \sim h \rightarrow \infty$$

Démonstration : Le corollaire précédent est en fait l'application directe du corollaire 4.2. \square

Corollaire 6.3 *Sous les hypothèses sur le modèle effectuées dans la partie 6.2.1, sous les hypothèses du théorème 6.1, et si les conditions de stabilité (4.33) et (4.13) sont vérifiées, la règle de décision \hat{t}_{lim} est asymptotiquement optimale (pour N et m tendant vers l'infini et α vers 0) lorsque la fenêtre vérifie :*

$$\liminf \frac{M_\alpha}{|\log \alpha|} > I^{-1} \quad \text{mais} \quad \log M_\alpha = o(\log \alpha) \quad \text{quand} \quad \alpha \rightarrow 0$$

et que $2M_\alpha e^{-h} = \alpha$, pour le critère du retard moyen : $E^{(tp)}(t_a - t_p + 1)^+$, lorsqu'on considère l'ensemble des règles de décision telles que :

$$\sup_{t \geq 1} P_{\theta_0}(t \leq t_a < t + m_\alpha) \leq \alpha .$$

Démonstration : Il s'agit ici d'appliquer le théorème 4.5 du chapitre 4. \square

On obtient donc, pour la règle \hat{t}_{lim} , des résultats d'optimalité pour différents critères et contraintes.

Méthode plus approximative

La deuxième méthode que nous avons envisagée pour diminuer la quantité de calcul de l’algorithme de filtrage-détection consiste à ne considérer qu’un seul filtre sous l’hypothèse H_1 . La vraisemblance conditionnelle des observations sous H_1 sera alors estimée à partir de particules ayant évolué depuis l’instant initial $t = 0$ suivant le modèle suivant :

$$\begin{cases} X_n = f(X_{n-1}, \theta_1, v_n) \\ Y_n = h(X_n, \theta_1, w_n). \end{cases} \quad (6.17)$$

Cela revient à considérer que la densité conditionnelle sous H_1 ne dépend pas de l’instant de changement de régime, ce qui est incorrect. Le temps d’arrêt sera écrit :

$$\hat{t}_{app} = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{\hat{l}_{1,i,1}^{N,m}}{\hat{l}_{0,i}^{N,m}} \geq h \right\}. \quad (6.18)$$

Le fait de n’avoir que deux filtres en parallèle au lieu de $M + 2$ rend la procédure très rapide (avec, de plus, une possibilité d’écriture récursive). Mais la quantité $\hat{l}_{1,i,1}^{N,m}$ qui devrait estimer la vraisemblance conditionnelle $p_{\theta_0, \theta_1, j}(Y_i | Y_{1:i-1})$ ne converge pas vers cette dernière. Donc, a priori, rien ne garantit que cette approche fonctionne en pratique.

Nous avons réalisé toute une série de simulations pour comparer ces deux approches (\hat{t}_{lim} et \hat{t}_{app}) et sur les modèles utilisés, les résultats sont cependant assez semblables (voir partie 6.2.7).

Les hypothèses de stabilité

Pour obtenir les résultats d’optimalité précédents, il faut que les hypothèses de stabilité sur le logarithme du rapport de vraisemblance évoquées dans le chapitre 3 soient vérifiées. Nous avons vu dans le chapitre précédent que c’était le cas pour des modèles auto-régressifs de la forme :

$$Y_n = g(Y_{n-1}, \theta) + \epsilon_n$$

sous de bonnes conditions sur g et sur le bruit ϵ . En ce qui concerne les modèle à espace d’état, il est a priori plus difficile de vérifier ces hypothèses, notamment (4.12) et (4.13) à cause du supremum essentiel. Pour ce qui est de la contrainte (4.33), elle est, selon Tartakovsky [107], beaucoup plus facile à vérifier, comme c’est le cas dans certaines classes de modèles à espace d’état linéaires.

6.2.7 Simulations

Comportement de l'algorithme de filtrage-détection

Nous allons tout d'abord appliquer les deux versions de l'algorithme de filtrage-détection, \hat{t}_{lim} et \hat{t}_{app} , sur un modèle à espace d'état très utilisé comme modèle test par la communauté s'intéressant au problème du filtrage [46, 59, 90] :

$$\begin{cases} X_n = 0.5X_{n-1} + \theta \frac{X_{n-1}}{1+X_{n-1}^2} + 8 \cdot \cos(1.2 \cdot t) + v_{n-1} \\ Y_n = \frac{1}{20}X_n^2 + w_n \end{cases} \quad (6.19)$$

avec $v_n \sim \mathcal{N}(0; 1)$ et $w_n \sim \mathcal{N}(0; 10)$. La panne considérée est la suivante : sous H_0 , $\theta = \theta_0 = 25$ et sous H_1 , $\theta = \theta_1 = 16$. Nous avons simulé une panne à l'instant $t_p = 120$. La figure 6.2 représente l'évolution des variables d'état et d'observation au cours du temps. On peut remarquer que le changement considéré n'est pas grossier et donc difficilement visible à "l'œil nu".

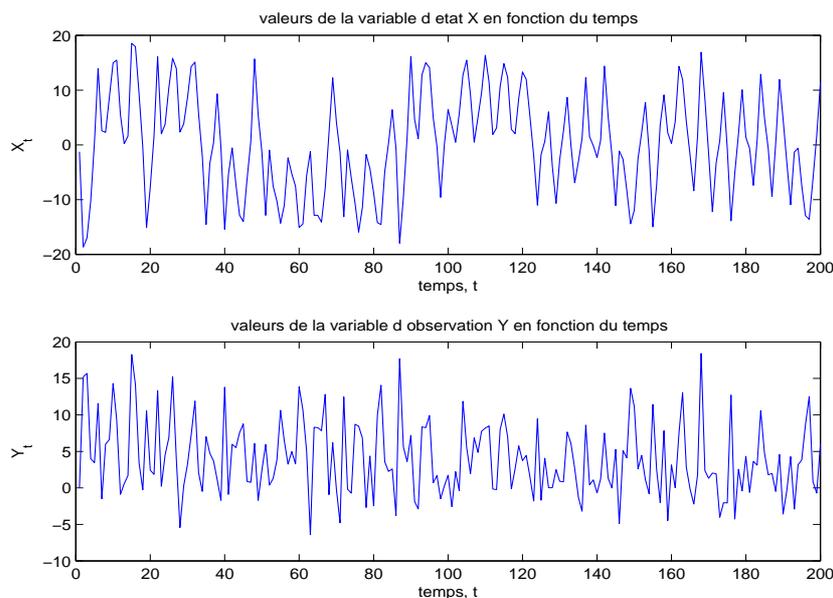


FIG. 6.2 – Evolution des variables d'état et d'observation au cours du temps avec un changement en $t_p = 120$

Sur la figure 6.3 sont représentées les statistiques de test de l'algorithme de filtrage-détection. \hat{t}_{app} a donc été obtenue à partir de deux filtres en parallèle tandis que \hat{t}_{lim} a été calculée en prenant une fenêtre M égale à 60. Nous avons choisi pour cette simulation $N = m = 500$. Les estimations à noyaux de convolution

Chapitre 6. Algorithmes de filtrage-détection

sont obtenues en utilisant des noyaux gaussiens avec pour paramètre de fenêtre : $\delta_N = \text{std}(\tilde{X})N^{(-1/5)}$.

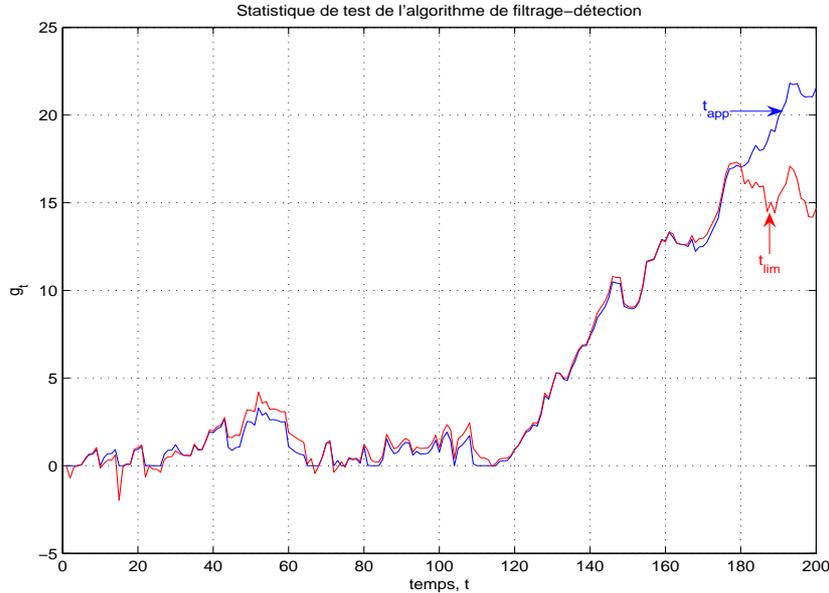


FIG. 6.3 – Statistiques de test de l'algorithme de filtrage-détection

Malgré le fait que le changement ne soit pas grossier, les deux statistiques de test détectent le changement de régime sans difficulté, avec un nombre de particules peu élevé. Comme nous l'avons évoqué dans la partie précédente, après 60 pas de temps sous H_1 , la statistique de test \hat{t}_{lim} se stabilise (à partir de $t = 180$). On peut remarquer que l'approximation grossière \hat{t}_{app} se comporte de façon quasi-similaire à \hat{t}_{lim} . C'est finalement un résultat assez surprenant, mais que les simulations suivantes vont confirmer.

Comparaison avec d'autres méthodes

Modèle linéaire :

Nous considérons le modèle à espace d'état linéaire suivant :

$$\begin{cases} X_n = aX_{n-1} + v_n \\ Y_n = bX_n + w_n, \end{cases} \quad (6.20)$$

où a, b sont des réels et w_n et v_n sont des bruits blancs gaussiens de variance σ_w^2 et σ_v^2 .

Nous allons supposer qu'à un instant t_p inconnu, un changement additif affecte le système dans le sens où la quantité $\theta \mathbb{1}_{\{n \geq t_p\}}$ est ajoutée dans l'équation d'état du modèle (6.20).

Pour des systèmes linéaires de ce type, l'approche statistique la plus utilisée actuellement consiste à appliquer l'algorithme de Willsky et Jones t_W (2.21) décrit dans le chapitre 2. Le principe est le suivant : en appliquant le filtre de Kalman, on retombe sur un "problème basique" de changement de moyenne dans une suite de variables gaussiennes.

Nous nous proposons de comparer les retards moyens à la détection entre notre algorithme de filtrage-détection et la méthode basée sur le filtre de Kalman. Pour cette simulation, les valeurs des paramètres sont les suivantes : $a = 1/2$, $b = 1$, $\sigma_w^2 = 1$, $\sigma_v^2 = 1$ et $\pi_0 \sim N(2, 1)$. Les estimations à noyaux de convolution sont obtenues en utilisant à nouveau des noyaux gaussiens K^x et K^y , avec pour paramètre de fenêtre : $\delta_N = std(\tilde{X})N^{(-1/5)}$ et $\delta_N = std(\tilde{Y})N^{(-1/5)}$. On choisit par simulation (5000 répétitions) le seuil de chaque algorithme de sorte que le temps moyen avant la première fausse alarme soit égal à 100, et on évalue le retard moyen à la détection $E_{\theta_1}^{(1)}(\cdot)$, là aussi, sur 5000 répétitions. Nous testons plusieurs valeurs pour θ et nous appliquons l'algorithme de filtrage-détection pour différentes valeurs des paramètres N et m (dans toutes ces simulations, on a par ailleurs choisi $N = m$). Pour la statistique de test \hat{t}_{lim} , on prend une fenêtre égale à $M = 100$.

Les résultats sont regroupés dans la table 6.1. On voit que l'algorithme t_W est plus performant que l'algorithme de filtrage-détection. Cela n'a rien d'étonnant compte tenu du fait que le filtre de Kalman donne la véritable densité conditionnelle contrairement au filtre à noyaux de convolution, qui fournit seulement une estimation, même convergente. Malgré cela, les résultats obtenus montrent le bon comportement de l'algorithme de filtrage détection, même avec un nombre de particules peu élevé, notamment lorsque la panne est très marquée ($\theta = 2$). Lorsque la panne est plus "fine", il est alors nécessaire de considérer un nombre de particules élevé pour atteindre un retard à la détection proche de la règle t_W .

Nous avons considéré dans cet exemple une panne additive sur notre modèle. Lorsque la panne n'est plus additive mais multiplicative, la règle t_W ne peut plus être appliquée, contrairement à l'algorithme de filtrage-détection.

Lorsque le modèle décrivant le fonctionnement du système n'est plus linéaire, une des approches les plus utilisées consiste à appliquer le test proposé par Willsky et Jones à l'aide d'un filtre de Kalman étendu. Mais deux problèmes peuvent alors se poser : d'une part le processus e_n n'est généralement plus gaussien et il est difficile de donner un sens au temps d'arrêt défini en (2.21) et d'autre part, le filtre de Kalman étendu n'est pas la solution idéale aux problèmes de non linéarité. Dans la plupart des cas, les filtres de Monte Carlo se montrent meilleurs.

Modèle non linéaire :

	Algorithme de filtrage détection								t_W défini
	N=50		N=100		N=500		N=5000		en (2.21)
	\hat{t}_{lim}	\hat{t}_{app}	\hat{t}_{lim}	\hat{t}_{app}	\hat{t}_{lim}	\hat{t}_{app}	\hat{t}_{lim}	\hat{t}_{app}	t_w
$\theta = 0.5$	27.27	26.10	25.77	24.46	23.05	22.26	21.65	20.90	13.47
$\theta = 1$	8.50	8.68	8.63	8.47	8.11	7.98	7.70	7.66	6.30
$\theta = 2$	3.31	3.30	3.23	3.23	3.18	3.13	3.06	3.06	2.96

TAB. 6.1 – Retard moyen $E_{\theta_1}^{(1)}$ à la détection pour différentes pannes

	$E_{\theta_1}^{(1)}(\hat{t}_{lim})$	$E_{\theta_1}^{(1)}(\hat{t}_{app})$	$E_{\theta_1}^{(1)}(t_W)$	$E_{\theta_1}^{(1)}(t_{LK})$
$\sigma_w^2 = \sigma_v^2 = 0.01, \theta_1 = 0.05$	15.51	15.45	14.23	16.85
$\sigma_w^2 = \sigma_v^2 = 0.1, \theta_1 = 0.2$	13.25	13.18	14.73	13.50
$\sigma_w^2 = \sigma_v^2 = 1, \theta_1 = 1$	11.83	11.69	20.91	10.94

TAB. 6.2 – Retard moyen à la détection

Considérons le modèle suivant :

$$\begin{cases} X_n = \frac{X_{n-1}}{2(1+X_{n-1}^2)} + \theta + v_n \\ Y_n = \exp(-X_n) + w_n, \end{cases} \quad (6.21)$$

avec v_n et w_n des bruits blancs gaussiens de variances respectives σ_v^2 et σ_w^2 . Le paramètre θ caractérise une panne additive sur l'équation d'état. Ainsi, $\theta = \theta_0 = 0$ avant l'instant de panne et $\theta = \theta_1$ après l'instant de panne. Nous allons comparer en simulation les retards moyens à la détection des deux algorithmes de filtrage-détection, \hat{t}_{lim} et \hat{t}_{app} (avec $N = m = 1000$ particules, et $M = 80$) avec la règle de Willsky et Jones appliquée grâce au filtre de Kalman étendu t_W , mais aussi avec l'algorithme de Li et Kadiramanathan [58], t_{LK} , dont s'inspire l'algorithme de filtrage-détection, et pour lequel on utilise un filtre de Monte Carlo classique (1000 particules également).

La table 6.2 représente les retards à la détection des quatre règles. Pour chaque algorithme et chaque valeur de paramètre (θ_1, σ_w^2 et σ_v^2), les seuils ont été obtenus par simulation de Monte Carlo de sorte que le temps moyen avant la première fausse alarme soit égal à 100. Les retards à la détection sont calculés à partir de 5000 simulations.

On peut voir que les retards à la détection de l'algorithme de filtrage-détection et de t_{LK} sont très proches. Nous l'avons déjà dit, la mise au point des deux règles est similaire. L'algorithme t_W se comporte aussi bien que les deux autres algorithmes

	$E_{\theta_1}^{(10)}(\hat{t}_{lim} - 10)$	$E_{\theta_1}^{(10)}(\hat{t}_{app} - 10)$
$\sigma_w^2 = \sigma_v^2 = 0.01, \theta_1 = 0.05$	13.61	13.80
$\sigma_w^2 = \sigma_v^2 = 0.1, \theta_1 = 0.2$	11.52	11.53
$\sigma_w^2 = \sigma_v^2 = 1, \theta_1 = 1$	8.79	8.77

TAB. 6.3 – Retard moyen à la détection sur 5000 répétitions

pour des petites valeurs de variance. Mais plus la variance augmente et plus il a du mal à détecter les changements.

*

* *

Ces simulations montrent que dans le cas d'un modèle linéaire avec bruit blanc gaussien additif, la règle t_W est très performante pour détecter une panne additive. Mais en considérant un nombre assez élevé de particules, l'algorithme de filtrage-détection donne de bons résultats. Dans un cadre non linéaire, les méthodes de filtrage particulière se montrent généralement plus efficaces que le filtre de Kalman étendu et les règles de décision basées sur ces méthodes (\hat{t}_{lim} , \hat{t}_{app} et t_{LK}) ont tendance à prendre l'avantage sur t_W . D'autre part, les règles de décision \hat{t}_{lim} , \hat{t}_{app} et t_{LK} peuvent être utilisées dans le cas de panne multiplicative, ou encore pour des bruits non gaussiens, ce qui n'est pas le cas de la règle t_W . Enfin, comme nous l'avons évoqué dans les parties précédentes, on peut envisager d'utiliser \hat{t}_{lim} (et \hat{t}_{app}), à l'aide d'une estimation à noyaux de la vraisemblance au lieu de l'intégration de Monte Carlo, ce qui rend ses conditions d'utilisation plus générales que celles de t_{LK} .

Retour sur \hat{t}_{lim} et \hat{t}_{app} :

Nous avons vu au cours de ces simulations que les deux règles de décision sont très proches l'une de l'autre. Si on examine de plus près les résultats des tables 6.1 et 6.2, on a même l'impression que la procédure approximative prend quasi-systématiquement l'avantage, même si l'écart est très faible. Une explication se trouve dans le fait que nous calculons le retard moyen lorsque $t_p = 1$. Ainsi les deux statistiques de test \hat{g}_{lim} et \hat{g}_{app} valent 0 à l'instant précédent le changement (donc à $t = 0$). Ce cas de figure avantage légèrement \hat{t}_{app} puisque la statistique de test \hat{g}_{app} a tendance à être légèrement inférieure à \hat{g}_{lim} sous H_0 . Nous avons repris le modèle précédent en calculant cette fois-ci les retards lorsque la panne intervient à l'instant $t_p = 10$. Les résultats obtenus sont regroupés dans la table 6.3.

	$t_p = 0$	$t_p = 10$	$t_p = 50$	$t_p = 100$	$t_p = 200$
$E_{\theta_1}^{(t_p)}(\hat{t}_{app} - t_p)$	21.26	19.61	18.92	19.54	18.91

TAB. 6.4 – Retard moyen à la détection pour \hat{t}_{app} suivant l’instant de panne

La tendance remarquée dans les premières simulations semble s’estomper mais les deux algorithmes restent très similaires.

Revenons à présent sur le principe de fonctionnement de la règle \hat{t}_{app} . Un seul filtre est considéré pour estimer la vraisemblance sous H_1 . Toutes les particules de ce filtre sont générées, depuis l’instant initial en supposant qu’on est sous le régime H_1 . Le retard $E_{\theta_1}^{(1)}$ utilisé dans les simulations précédentes rend ainsi l’approximation effectuée moins sévère puisque les trajectoires sont simulées sous H_1 depuis l’instant $t = 1$. On pourrait donc s’attendre à une dégradation de la règle de décision plus la panne intervient tard puisque les particules seraient alors propagées, sous le régime normal, à travers un faux modèle pendant un long moment. Nous avons repris le modèle linéaire avec le changement $\theta_1 = 0.5$ et pour $N = m = 500$, nous avons calculé le retard à la détection pour la règle \hat{t}_{app} suivant l’instant où apparaît la panne. La table 6.4 présente les résultats.

Nous voyons que la règle de décision ne se dégrade pas même lorsque le temps de panne apparaît tardivement.

6.3 Paramètre caractérisant la panne inconnu

Nous considérons dans cette partie, une adaptation de l’algorithme de filtrage-détection à un cas plus réaliste en pratique. Nous supposons maintenant que le paramètre θ_1 , caractérisant le régime de panne du système, est inconnu et appartient à un compact Θ_1 . Nous considérons toujours le même système (6.1) avec les mêmes hypothèses sur le modèle que dans la partie 6.2.1 si ce n’est que, sous H_0 le paramètre θ vaut θ_0 qui est connue, et sous H_1 , θ prend la valeur inconnue mais fixe θ_1 qui appartient à un compact Θ_1 . De plus, on suppose que θ_0 n’appartient pas au compact Θ_1 .

6.3.1 Principe de l’algorithme

Comme dans la première partie de ce chapitre, nous allons chercher à détecter un changement de paramètre dans la densité conditionnelle des observations. Nous allons donc estimer cette densité conditionnelle sous H_0 et sous H_1 en combinant le filtre à noyaux de convolution avec des intégrations de Monte Carlo pour construire

la règle de décision sous la même forme :

$$\hat{t} = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{\hat{l}_{1,i,j}^{N,m}}{\hat{l}_{0,i}^{N,m}} \geq h \right\}. \quad (6.22)$$

Estimation sous l'hypothèse H_0

Nous sommes là exactement dans les mêmes conditions que dans la partie 6.2 puisque le paramètre θ_0 est supposé connue. On a :

$$p_{\theta_0}(Y_n | Y_{1:n-1}) = \int p_{\theta_0}(Y_n | x_n) \cdot p_{\theta_0}(x_n | Y_{1:n-1}) dx_n. \quad (6.23)$$

On estime donc la vraisemblance conditionnelle des observations par

$$\hat{l}_{0,n}^{N,m} = \frac{1}{m} \sum_{k=1}^m p_{\theta_0}(Y_n | x_{n,0}(k)) \simeq p_{\theta_0}(Y_n | Y_{1:n-1}), \quad (6.24)$$

où $(x_{n,0}(k))$, $k = 1, \dots, m$ est un échantillon aléatoire de l'estimation $\hat{p}_{\theta_0}(x_n | Y_{1:n-1})$ de la densité prédictive obtenue en appliquant le filtre à noyaux de convolution à la modélisation suivante :

$$\begin{cases} X_n = f(X_{n-1}, \theta_0, v_n) \\ Y_n = h(X_n, \theta_0, w_n) \end{cases} \quad (6.25)$$

REMARQUE : Comme dans la partie précédente, l'estimation de cette vraisemblance conditionnelle est convergente.

Estimation sous l'hypothèse H_1

En théorie du filtrage, un moyen pour traiter les paramètres inconnus est de considérer qu'ils font partie de l'état du système. Nous allons utiliser cette approche. La modélisation du système sous l'hypothèse d'un changement de régime à un instant j compris entre 1 et n sera :

$$\begin{cases} X_i = f(X_{i-1}, \theta_0, v_i) \\ Y_i = h(X_i, \theta_0, w_i) \end{cases} \quad \forall i < j \quad (6.26)$$

et

$$\begin{cases} \begin{bmatrix} X_i \\ \theta_i \\ Y_i = h(X_i, \theta_i, w_i) \end{bmatrix} = \begin{bmatrix} f(X_{i-1}, \theta_{i-1}, v_i) \\ \theta_{i-1} \end{bmatrix} \end{cases} \quad \forall i \geq j. \quad (6.27)$$

Chapitre 6. Algorithmes de filtrage-détection

REMARQUES :

- Le paramètre θ_1 est fixe sous l'hypothèse H_1 , c'est pourquoi sa modélisation est la plus simple possible $\theta_i = \theta_{i-1}$. C'est l'algorithme de filtrage à noyaux de convolution qui nous permet d'écrire le modèle sous cette forme. En effet, pour les filtres à particules traditionnels, une telle écriture entraîne une dégénérescence des poids. On est alors obligé d'ajouter un bruit dans l'équation $\theta_i = \theta_{i-1}$ (voir Liu et West [60]) qui n'a aucune justification théorique.

- Nous avons pris le parti de considérer dans ce travail de thèse des changements abrupts. On passe à l'instant t_p , d'une valeur θ_0 à une valeur θ_1 , ces deux valeurs étant fixes. Mais rien n'empêche dans le modèle (6.27) de considérer une panne graduelle en posant un modèle de panne (linéaire ou autre) sur θ .

La vraisemblance conditionnelle des observations fournie par le modèle sous l'hypothèse d'un changement de régime à un instant j compris entre 1 et n est donnée par l'intégrale suivante :

$$p_{H_0, H_1, j}(Y_n | Y_{1:n-1}) = \int p_{H_1}(Y_n | x, \theta) \cdot p_{H_0, H_1, j}(x, \theta | Y_{1:n-1}) dx d\theta.$$

Le filtre à noyaux de convolution permet d'estimer la densité prédictive $p_{H_0, H_1, j}(x, \theta | Y_{1:n-1})$ à partir de particules ayant évolué grâce aux modèles (6.26) et (6.27). L'estimateur s'écrit :

$$\hat{p}_{H_0, H_1, j}(x, \theta | Y_{1:n-1}) = \frac{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{n-1, j}(i) - Y_{n-1}) \cdot K_{\delta_N}^x(\tilde{X}_{n, j}(i) - x) \cdot K_{\delta_N}^\theta(\tilde{\theta}_n(i) - \theta)}{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{n-1, H_1}(i) - Y_{n-1})},$$

où K^y , K^x et K^θ sont des noyaux de convolution et δ_N leurs fenêtres.

L'estimation du filtre optimal $p_{H_0, H_1, j}(x, \theta | Y_{1:n})$ s'écrit de façon analogue.

REMARQUES :

- Depuis le début de la partie 6.3, nous avons légèrement changé d'écriture : $p_{\theta_0, \theta_1, j}$ est devenue $p_{H_0, H_1, j}$ et nous ne parlons plus de vraisemblance conditionnelle sous hypothèse de régime de panne mais de vraisemblance fournie par les modèles (6.26) et (6.27). Dans les parties précédentes, $p_{\theta_0, \theta_1, j}$ représentée la densité pour $\theta = \theta_1$ à partir de $t = j$. Dans cette partie, on ne connaît pas la vraie valeur θ_1 et on considère θ comme une variable à part entière dans le modèle (6.27). C'est pourquoi nous avons changé d'écriture. Des résultats de convergence de la loi a posteriori montrent par ailleurs que lorsqu'on a une infinité d'observations, $p_{H_0, H_1, j}$ tend vers $p_{\theta_0, \theta_1, j}$. Nous reviendrons sur ce résultat dans la partie 6.4.

- La densité $p_{H_0, H_1, j}(x, \theta | Y_{1:n})$ a pour support $\mathbb{R} \times \Theta_1$ puisqu'on sait que la vraie valeur de θ_1 se trouve dans l'ensemble Θ_1 . Par contre, l'estimateur obtenu par le filtre à noyaux de convolution a pour support $\mathbb{R} \times \mathbb{R}$. Dans notre contexte, cela

pose un problème puisque cela veut dire qu'à l'étape de ré-échantillonnage du filtre à noyaux, on peut obtenir des particules $\bar{\theta}$ n'appartenant pas à Θ_1 . Cela revient en fait à supprimer l'hypothèse $\theta_1 \in \Theta_1$ et donc, sous régime H_0 , au bout de quelques pas de temps, toutes les particules ont tendance à se regrouper autour de la valeur de θ sous H_0 , c'est à dire θ_0 .

Pour remédier à ce problème, nous n'allons donc pas considérer comme estimateur de $p_{H_0, H_1, j}(x, \theta | Y_{1:n})$ la quantité $\hat{p}_{H_0, H_1, j}(x, \theta | Y_{1:n})$ mais sa troncature $\bar{p}_{H_0, H_1, j}(x, \theta | Y_{1:n})$ sur $\mathbb{R} \times \Theta_1$ (on tronque puis on normalise). Cet estimateur aura l'avantage d'avoir le même support que la densité ciblée.

De la même façon, nous pouvons considérer la troncature $\bar{p}_{H_0, H_1, j}(x, \theta | Y_{1:t-1})$ de l'estimateur $\hat{p}_{H_0, H_1, j}(x, \theta | Y_{1:t-1})$ de la densité prédictive.

De plus, le théorème suivant nous assure que ce nouvel estimateur est au moins aussi "bon" que le premier :

Théorème 6.3 (Devroye [22]) *Soit g une fonction telle que $\int g = 1$ et f une densité de support S . On définit :*

$$g^* = \frac{g \mathbb{I}_S}{\int_S g}.$$

Alors on a,

$$\int |g^* - f| \leq \int |g - f|.$$

Grâce au filtre à noyaux de convolution et à l'étape de troncature, nous sommes en mesure de calculer la quantité suivante :

$$\hat{l}_{1,n,j}^{N,m} = \frac{1}{m} \sum_{i=1}^m p_{H_0, H_1, j}(Y_n | x_{n,j}(i), \theta(i)) \quad (6.28)$$

où $(x_{n,j}(i), \theta(i))_{i=1, \dots, m}$ est un échantillon aléatoire de $\bar{p}_{H_0, H_1, j}(x_n, \theta | Y_{1:n-1})$ (estimateur de la densité prédictive) obtenue par le filtre prédictif. La quantité $\hat{l}_{1,n,j}^{N,m}$ est donc un estimateur de la vraisemblance fournie par le modèle.

6.3.2 Description de l'algorithme

Comme nous l'avons dit dans la partie précédente, pour estimer la vraisemblance sous l'hypothèse de fonctionnement normal, nous sommes dans les mêmes conditions que dans la partie 6.2, et donc le calcul de la quantité $\hat{l}_{0,n}^{N,m}$ va s'effectuer de la même manière en suivant l'algorithme décrit dans la partie 6.2.5.

Chapitre 6. Algorithmes de filtrage-détection

Nous allons ici détailler les différentes étapes pour le calcul des quantités $\hat{l}_{1,n,j}^{N,m}$ pour une valeur j fixée, vraisemblance obtenue par la modélisation (6.26) et (6.27) sous l'hypothèse d'un changement de régime à un instant $j \leq n$.

Initialisation : $t = 1$

- Génération de N réalisations $\bar{X}_0(i)$, $i = 1, 2, \dots, N$ grâce à π_0 .
- Génération de N réalisations des bruits $v_1(i)$ et $w_1(i)$, $i = 1, 2, \dots, N$ grâce aux lois \mathcal{L}_v et \mathcal{L}_w supposées connues.

- Le modèle (6.26) conduit aux réalisations :

$$\begin{aligned}\tilde{X}_1(i) &= f(\bar{X}_0(i), \theta_0, v_1(i)) \\ \tilde{Y}_1(i) &= g(\tilde{X}_1(i), \theta_0, w_1(i)) , \quad i = 1, \dots, N.\end{aligned}$$

- A partir de l'observation Y_1 , les N couples de particules $(\tilde{X}_1(i), \tilde{Y}_1(i))$ permettent la construction d'une estimation par noyau $\hat{p}_{H_0, H_1, j}(x_1 | Y_1)$ de la densité conditionnelle $p_{H_0, H_1, j}(x_1 | Y_1)$ (de la même façon que (6.7)).

Temps $t < j$:

On répète les étapes 1- et 4- de l'algorithme décrit dans la partie 6.2.5 pour faire évoluer les couples de particules suivant la modélisation (6.26).

Temps j :

Etape 1 :

- A partir de l'estimation obtenue à l'instant $j - 1$, génération de :
 $\bar{X}_{j-1}(i) \sim \hat{p}_{H_0, H_1, j}(x_{j-1} | Y_{1:j-1}) := \hat{p}_{\theta_0}(x_{j-1} | Y_{1:j-1})$, $w_j(i) \sim \mathcal{L}_w$, $v_j(i) \sim \mathcal{L}_v$
- Génération de N réalisations $\bar{\theta}_{j-1}(i)$, $i = 1, \dots, N$ (l'indice apparaissant à coté de θ indique le pas de temps) grâce à la loi à priori $p_0(\theta)$. Si on n'a pas d'information à priori, on se contente de tirer les particules suivant une loi uniforme sur Θ_1 .

- Le modèle (6.27) conduit à :

$$\begin{aligned}\tilde{X}_j(i) &= f(\bar{X}_{j-1}(i), \bar{\theta}_{j-1}(i), v_j(i)) \\ \tilde{\theta}_j(i) &= \bar{\theta}_{j-1}(i) \\ \tilde{Y}_j(i) &= g(\tilde{X}_j(i), \tilde{\theta}_j(i), w_j(i)) , \quad i = 1, \dots, N.\end{aligned}$$

Etape 2 :

- A partir des particules $\tilde{X}_j(i)$, $\tilde{\theta}_j(i)$, $\tilde{Y}_{j-1}(i)$, $i = 1, \dots, N$, et de l'observation Y_{j-1} , on peut construire un estimateur à noyaux consistant de la densité prédictive $p_{H_0, H_1, j}(x_j, \theta_j | Y_{1:j-1})$:

$$\hat{p}_{H_0, H_1, j}(x_j, \theta_j | Y_{1:j-1}) = \frac{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{j-1}(i) - Y_{j-1}) \cdot K_{\delta_N}^x(\tilde{X}_j(i) - x_j) \cdot K_{\delta_N}^\theta(\tilde{\theta}_j(i) - \theta_j)}{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{j-1}(i) - Y_{j-1})}.$$

- On détermine la troncature $\bar{p}_{H_0, H_1, j}(x_j, \theta_j | Y_{1:j-1})$ de $\hat{p}_{H_0, H_1, j}(x_j, \theta_j | Y_{1:j-1})$.

- On peut alors estimer la vraisemblance conditionnelle des observations fournies par le modèle dès que la mesure Y_j est disponible :

il suffit de générer un échantillon de m particules de la densité prédictive estimée $\bar{p}_{H_0, H_1, j}(x_j, \theta_j | Y_{1:j-1})$ que l'on notera $(x_j(k), \theta_j(k))$, $k = 1, \dots, m$. La vraisemblance des observations est alors estimée par :

$$\hat{l}_{1,j,j}^{N,m} = \frac{1}{m} \sum_{k=1}^m \hat{p}_{H_1}(Y_j | x_j(k), \theta_j(k)).$$

Etape 3 :

- A partir des particules $\tilde{X}_j(i)$, $\tilde{\theta}_j(i)$, $\tilde{Y}_j(i)$, $i = 1, \dots, N$, et de l'observation Y_j , nous estimons le filtre optimal par une approche à noyaux :

$$\hat{p}_{H_0, H_1, j}(x_j, \theta_j | Y_{1:j}) = \frac{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_j(i) - Y_j) \cdot K_{\delta_N}^x(\tilde{X}_j(i) - x_j) \cdot K_{\delta_N}^\theta(\tilde{\theta}_j(i) - \theta_j)}{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_j(i) - Y_j)}.$$

- On détermine la troncature du filtre optimal $\bar{p}_{H_0, H_1, j}(x_j, \theta_j | Y_{1:j})$.

Temps $j < t \leq n$:

Etape 1 :

- A partir de l'estimation obtenue à l'instant $t - 1$, génération de :
 $(\bar{X}_{t-1}(i), \bar{\theta}_{t-1}(i)) \sim \bar{p}_{H_0, H_1, j}(x_{t-1}, \theta_{t-1} | Y_{1:t-1})$, $w_t(i) \sim \mathcal{L}_w$, $v_t(i) \sim \mathcal{L}_v$

- On fait évoluer les particules grâce au modèle (6.27) pour obtenir $\tilde{X}_t(i)$, $\tilde{\theta}_t(i)$ et $\tilde{Y}_t(i)$, $i = 1, \dots, N$.

Etape 2 :

Chapitre 6. Algorithmes de filtrage-détection

- On estime la densité prédictive par :

$$\hat{p}_{H_0, H_1, j}(x_t, \theta_t | Y_{1:t-1}) = \frac{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{t-1}(i) - Y_{t-1}) \cdot K_{\delta_N}^x(\tilde{X}_t(i) - x_t) \cdot K_{\delta_N}^\theta(\tilde{\theta}_t(i) - \theta_t)}{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_{t-1}(i) - Y_{t-1})},$$

et on détermine la densité tronquée sur le domaine $\mathbb{R} \times \Theta_1$.

- On calcule la quantité $\hat{l}_{1,t,j}^{N,m}$ à partir d'un échantillon de m particules, $(x_t(k), \theta_t(k), i = k, \dots, m)$, de la densité prédictive estimée $\bar{p}_{H_0, H_1, j}(x_t, \theta_t | Y_{1:t-1})$.

Etape 3 :

- On estime le filtre optimal par :

$$\hat{p}_{H_0, H_1, j}(x_t, \theta_t | Y_{1:t}) = \frac{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_t(i) - Y_t) \cdot K_{\delta_N}^x(\tilde{X}_t(i) - x) \cdot K_{\delta_N}^\theta(\tilde{\theta}_t(i) - \theta)}{\sum_{i=1}^N K_{\delta_N}^y(\tilde{Y}_t(i) - Y_t)},$$

et on détermine la densité tronquée sur le domaine $\mathbb{R} \times \Theta_1$.

De plus, à l'instant $t = n$:

On dispose des quantités $\hat{l}_{1,j,j}^{N,m}, \hat{l}_{1,j+1,j}^{N,m}, \dots, \hat{l}_{1,n,j}^{N,m}$. On peut donc calculer $\sum_{i=j}^n \log \frac{\hat{l}_{1,i,j}^{N,m}}{\hat{l}_{0,i}^{N,m}}$ et obtenir la statistique de test \hat{g}_n définie par :

$$\hat{g}_n = \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{\hat{l}_{1,i,j}^{N,m}}{\hat{l}_{0,i}^{N,m}},$$

à partir de tous les autres filtres caractérisant un instant possible de changement compris entre 1 et n .

6.3.3 Méthodes d'approximation

Comme dans la partie 6.2, on ajoute à chaque pas de temps un filtre supplémentaire, ce qui rend difficile l'utilisation de l'algorithme précédent.

On peut à nouveau envisager les deux méthodes présentées précédemment pour diminuer la quantité de calcul. D'une part, la méthode qui consiste à effectuer la maximisation pour le calcul de la statistique de test sur une fenêtre de taille M . Le temps d'arrêt de la règle de décision s'écrit alors :

$$\hat{t}_{lim} = \inf \left\{ n : \max_{n-M \leq j \leq n} \sum_{i=j}^n \log \frac{\hat{l}_{1,i,j}^{N,m}}{\hat{l}_{0,i}^{N,m}} \geq h \right\}. \quad (6.29)$$

On peut aussi utiliser la méthode plus approximative qui consiste à ne considérer qu'un seul filtre pour le régime de panne. Les particules $\tilde{\theta}$ sont alors générées dès l'instant initial et non plus à partir d'un instant j , $j \geq 1$, et on applique le filtre à noyaux de convolution en ne considérant que le modèle (6.27) pour $j = 1$. Le temps d'arrêt de cette règle de décision s'écrit à nouveau :

$$\hat{t}'_{app} = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{\hat{l}_{1,i,1}^{N,m}}{\hat{l}_{0,i}^{N,m}} \geq h \right\}. \quad (6.30)$$

6.3.4 Simulations

Nous comparons en simulation les deux procédures de détection (6.29) et (6.30), et nous nous intéressons plus particulièrement au déplacement des particules $\tilde{\theta}$. Le système étudié est le modèle à espace d'état utilisé dans la partie précédente :

$$\begin{aligned} X_n &= 0.5X_{n-1} + \theta \frac{X_{n-1}}{1 + X_{n-1}^2} + 8 \cos(1.2n) + v_n \\ Y_n &= 0.05X_n^2 + w_n \end{aligned}$$

où v_n et w_n sont des bruits blancs gaussiens de moyenne 0, et cette fois-ci de variance 1 et 0.1 respectivement.

Première simulation

Dans un premier temps, nous allons tester l'algorithme sur un changement grossier pour voir comment se comportent les particules $\tilde{\theta}(i)$ que l'on obtient grâce au filtre à noyaux.

Nous allons donc simuler des données en posant que sous H_0 , le paramètre θ vaut :

$$\theta_0 = 25,$$

et sous H_1 , le paramètre θ est égal à :

$$\theta_1 = 12.$$

Nous supposons pour mettre en place les algorithmes que la valeur de θ_1 est inconnue et nous disposerons seulement de l'information suivante :

$$\theta_1 \in \Theta_1 = [10; 20].$$

La simulation est réalisée avec $N = 500$ particules et pour l'estimation des vraisemblances par intégration de Monte Carlo, nous utilisons un échantillon de taille $m = 500$.

Chapitre 6. Algorithmes de filtrage-détection

Pour la règle \hat{t}'_{app} , à l'instant $t = 1$, nous allons tirer les particules $\bar{\theta}$ suivant la loi uniforme sur Θ_1 . Pour la procédure \hat{t}'_{lim} , de la même façon, à l'instant j supposé de changement de régime, nous supposons que nous n'avons pas d'information a priori sur la valeur de θ_1 . La fenêtre utilisée vaut $M = 80$.

Nous simulons la panne à l'instant $t_p = 120$. La figure 6.4 représente l'évolution des variables X_n et Y_n au cours du temps. On remarque alors que la panne considérée est une panne grossière puisqu'on voit clairement le changement de régime sur la variable d'observation. La figure 6.5 donne l'allure de la statistique de test des deux règles de décision. La détection s'effectue sans problème.

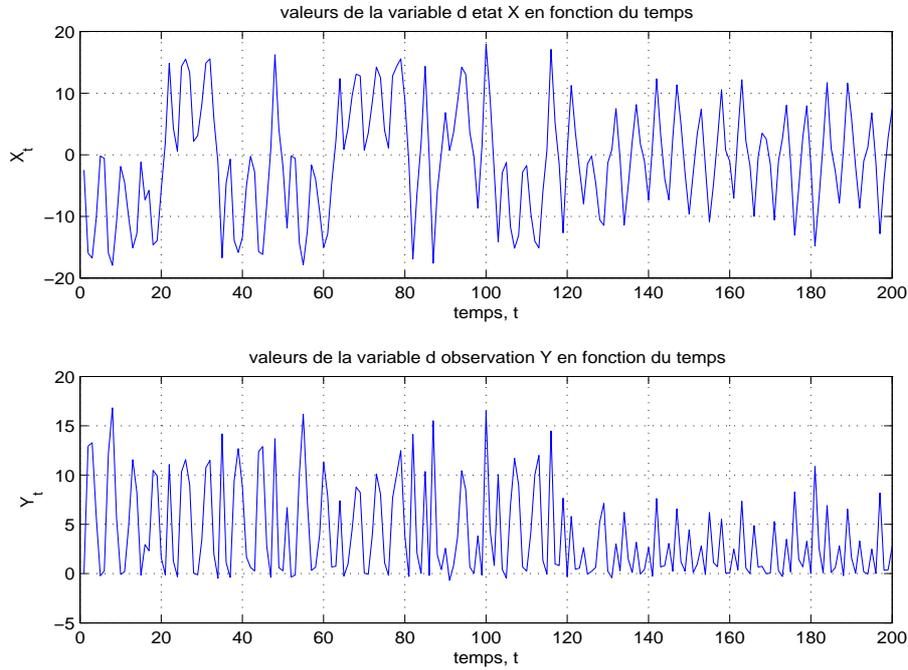


FIG. 6.4 – Graphiques des variables d'état et d'observation avec une panne à l'instant $T_p = 120$

La figure 6.6 est plus intéressante dans le sens où elle montre bien l'évolution des particules $\tilde{\theta}(i)$ au cours du temps pour la règle de décision \hat{t}'_{app} . Chaque courbe de la figure 6.6 représente l'estimation de la densité $p_{H_1}(\theta_t|Y_{1:t})$ pour différents pas de temps. L'estimateur utilisé est le suivant :

$$\hat{p}_{H_1}(\theta_t|Y_{1:t}) = \frac{\sum_{i=1}^N K_{\delta_n}^y(\tilde{Y}_{t,1}(i) - Y_t) \cdot K_{\delta_n}^\theta(\tilde{\theta}_t(i) - \theta_t)}{\sum_{i=1}^N K_{\delta_n}^y(\tilde{Y}_{t,1}(i) - Y_t)}. \quad (6.31)$$

D'après les résultats de Rossi [90], cet estimateur est consistant.

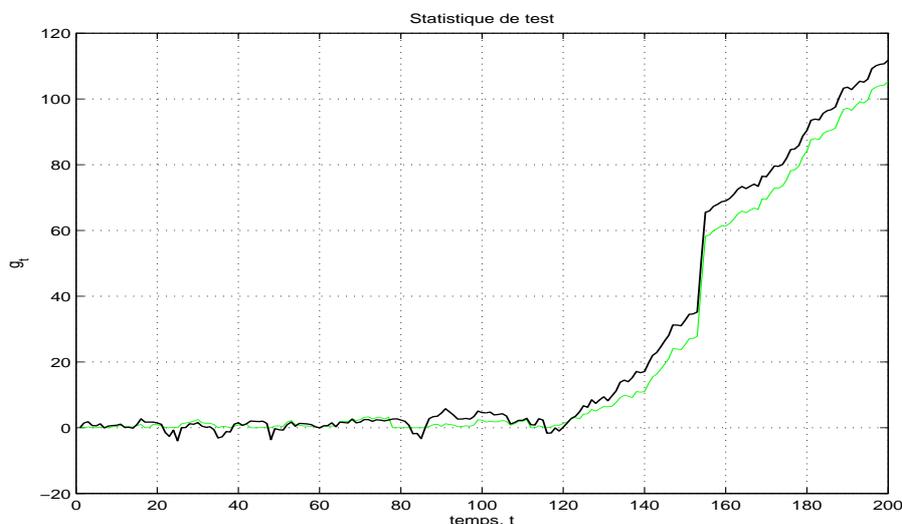


FIG. 6.5 – Graphe représentant les statistiques de test : \hat{t}'_{lim} en foncé et \hat{t}'_{app} en clair

On peut ainsi comprendre le fonctionnement de la règle \hat{t}'_{app} . Au tout début du processus et jusqu'à l'instant précédent la panne, la densité vient se coller sur la valeur 20 ($t = 116, 117, 118$ et 119), qui est la borne supérieure de l'intervalle Θ_1 . Ce n'est pas surprenant, puisqu'on est sous le régime H_0 , l'algorithme de filtrage cherche à estimer la vraie valeur de θ , c'est à dire θ_0 . C'est à ce moment là qu'il est important de réaliser l'étape de troncature, afin d'éviter que toutes les particules se retrouvent dans un voisinage de θ_0 . A partir de $t = 120$, on voit que le mode de la densité commence à se déplacer lentement vers la gauche et plus précisément vers la vraie valeur de θ sous régime de panne, c'est à dire $\theta_1 = 12$.

REMARQUE : Il est plus difficile de représenter le déplacement des particules pour la règle \hat{t}'_{lim} puisqu'à chaque pas de temps, l'instant j supposé de changement de régime et obtenu par maximum de vraisemblance, peut varier. Ainsi, d'un pas de temps à un autre, ce ne sont pas les mêmes jeux de particules utilisés pour le calcul de la statistique de test. Cependant, on peut imaginer le même comportement des particules : sous régime de panne, il y a un déplacement des particules vers la vraie valeur θ_1 .

Dans la figure 6.7, nous avons représenté l'estimateur

$$\hat{\theta}_t = \frac{1}{N} \sum_{i=1}^N \bar{\theta}_t(i),$$

de l'espérance conditionnelle $E[\theta|Y_{1:t}]$. Rossi [90] montre que cet estimateur est consistant lorsque le nombre de particules tend vers l'infini. Enfin, Schwartz [94]

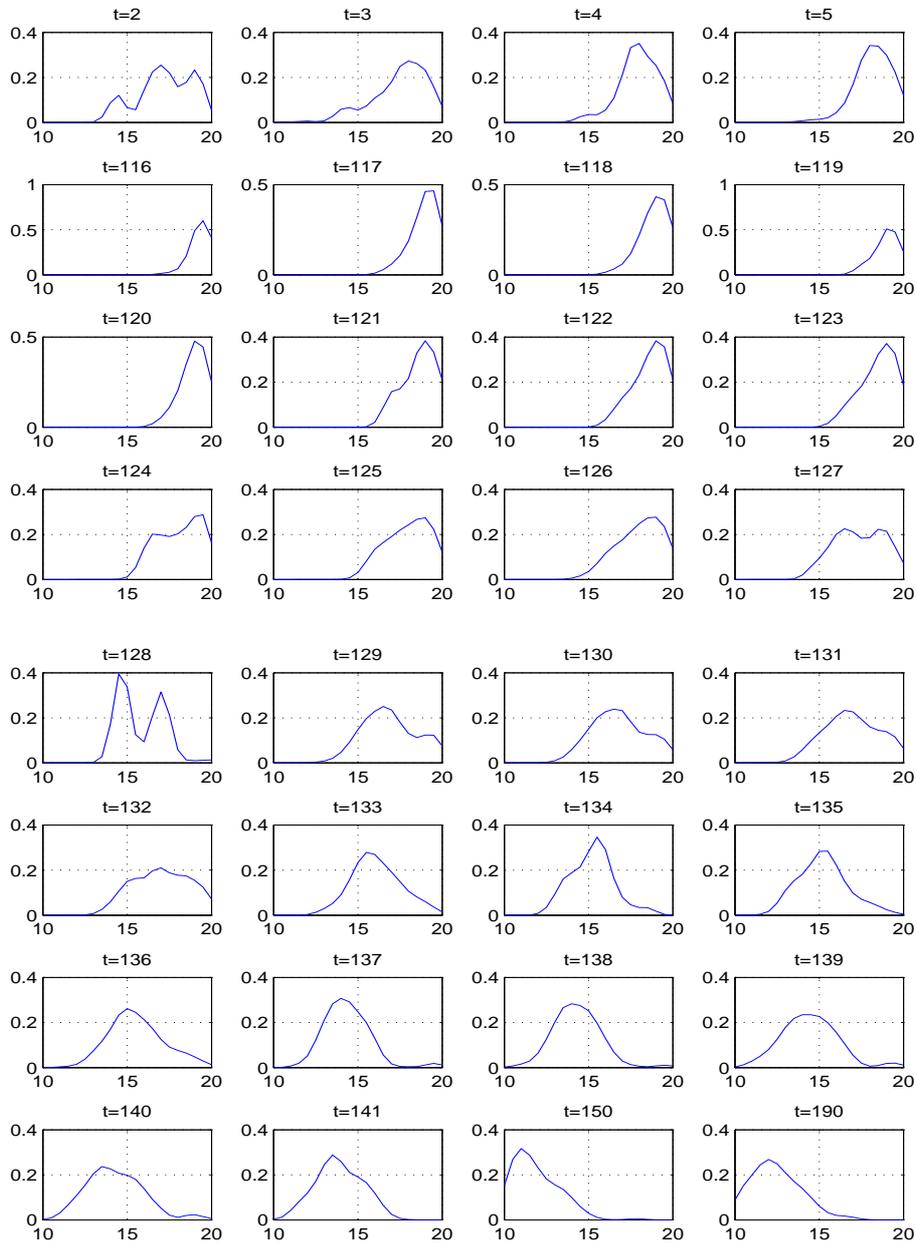


FIG. 6.6 – Estimation de la densité de θ sachant le passé

montre que :

$$\lim_{t \rightarrow \infty} E[\theta | Y_{1:t}] = \theta_1 \quad \text{p.s.} \quad (6.32)$$

Le résultat (6.32) ainsi que la convergence de $\hat{\theta}_t$ vers l'espérance conditionnelle ne sont valables que lorsque les observations Y_t proviennent de la bonne modélisation. Ce n'est pas le cas ici puisque les premières observations sont issues de la modélisation sous H_0 . Malgré cette contamination, on peut voir que l'estimateur se comporte bien.

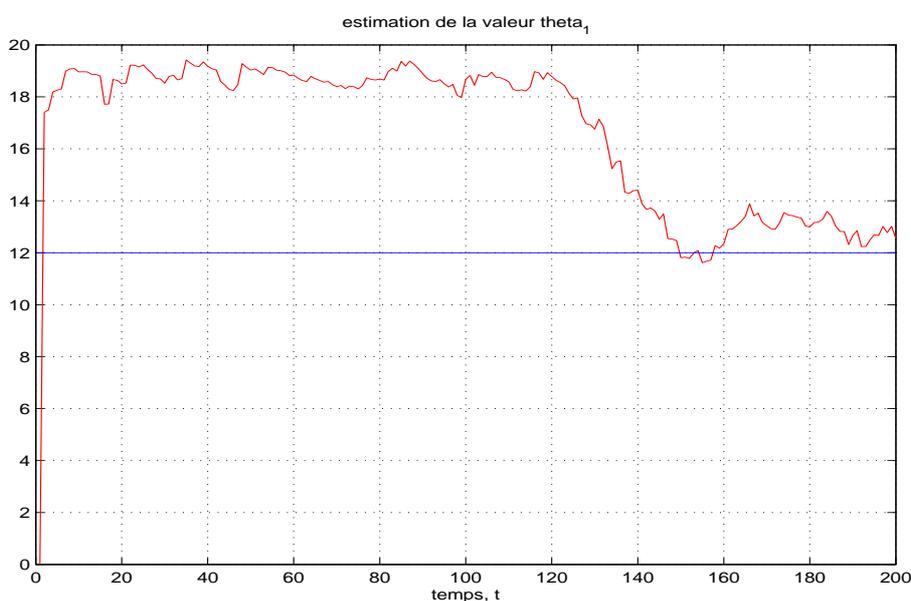


FIG. 6.7 – Estimation de la densité de θ sachant le passé

Revenons à la figure 6.6, à partir de l'instant $t = 140$, on peut dire que le mode de la densité estimée se situe sur la vraie valeur θ_1 . Le déplacement des particules a donc pris une vingtaine de pas de temps pour se retrouver dans un voisinage de la vraie valeur θ_1 . Ce déplacement de particules est fonction de plusieurs paramètres. Tout d'abord, le nombre de particules du filtre. Si ce nombre est élevé, on peut espérer avoir une meilleure estimation. Mais le ou les paramètres les plus importants sont les fenêtres des estimateurs à noyaux du filtre. La taille de la fenêtre va conditionner l'allure de l'estimateur et donc le ré-échantillonnage, autrement dit, le déplacement des particules. Toutes les simulations que nous avons menées ont abouti aux mêmes conclusions : pour les variables d'état et d'observation, le choix de la fenêtre n'est pas primordial (une fenêtre classique en estimation non paramétrique $\delta_N = \text{std}(\tilde{X}) \cdot N^{-1/5}$ donne la plupart du temps de bons résultats),

Chapitre 6. Algorithmes de filtrage-détection

mais au contraire, le choix de la fenêtre pour le paramètre θ est très important et doit être fixé au cas par cas. Ainsi, une fenêtre plutôt grande aura tendance à aplatir l'estimateur à noyau du filtre optimal et à permettre un déplacement plus rapide des particules, mais dans le même temps, lorsque les particules seront dans le voisinage de θ_1 , l'estimation sera moins précise ce qui peut rendre une panne très "fine", indétectable. Réciproquement, une fenêtre "petite" va permettre d'avoir une estimation plus précise du paramètre θ_1 mais le déplacement des particules sera plus lent.

REMARQUE : Pour la simulation précédente, nous avons choisi une fenêtre de taille $\delta_N = 0.5$, ce qui constitue, pour ce modèle, une fenêtre de taille moyenne. Lorsqu'on utilise une fenêtre plus grande (par exemple $\delta_N = 1.5$), il faut moins de 10 pas de temps pour que le mode de la distribution des particules soit sur la valeur θ_1 .

Deuxième simulation

Nous allons à présent tester l'efficacité des algorithmes sur une panne beaucoup moins grossière pour le même modèle à espace d'état. Pour cette simulation, nous avons :

$$\theta_0 = 21, \theta_1 = 18 \text{ et } \theta_1 \in \Theta = [10; 20]$$

Nous augmentons aussi le nombre de particules qui est à présent égal à $N = 1000$ et $m = 1000$; et la fenêtre pour la règle \hat{t}'_{lim} vaut encore $M = 80$. Les fenêtres des estimateurs à noyaux sont les mêmes que pour la simulation précédente.

Les résultats de cette simulation sont présentés dans les figures 6.8 et 6.9. Cette fois-ci, le changement considéré est difficilement visible "à l'oeil nu" mais les règles de décision parviennent à le détecter. On peut voir à nouveau que les deux algorithmes ont un comportement assez similaire.

Conclusions

La méthode proposée fonctionne bien sans avoir besoin d'utiliser un nombre trop élevé (qui ralentit fortement les calculs) de particules.

Concernant le choix de la fenêtre pour le noyaux K^θ , nous avons vu qu'il est nécessaire de faire un compromis entre la vitesse de déplacement des particules et la qualité de l'estimation obtenue. Suivant le type de changement auquel on est confronté (fin ou grossier), il apparaît un choix de seuil plus ou moins adapté. Malheureusement, en pratique, θ_1 appartient à Θ_1 et nous ne savons pas si le paramètre se situe plus dans une région de Θ_1 que dans une autre. On pourrait alors envisager d'utiliser deux fenêtres au lieu d'une. A chaque instant, on calculerait la vraisemblance conditionnelle avec les deux fenêtres (une petite pour la qualité de l'estima-

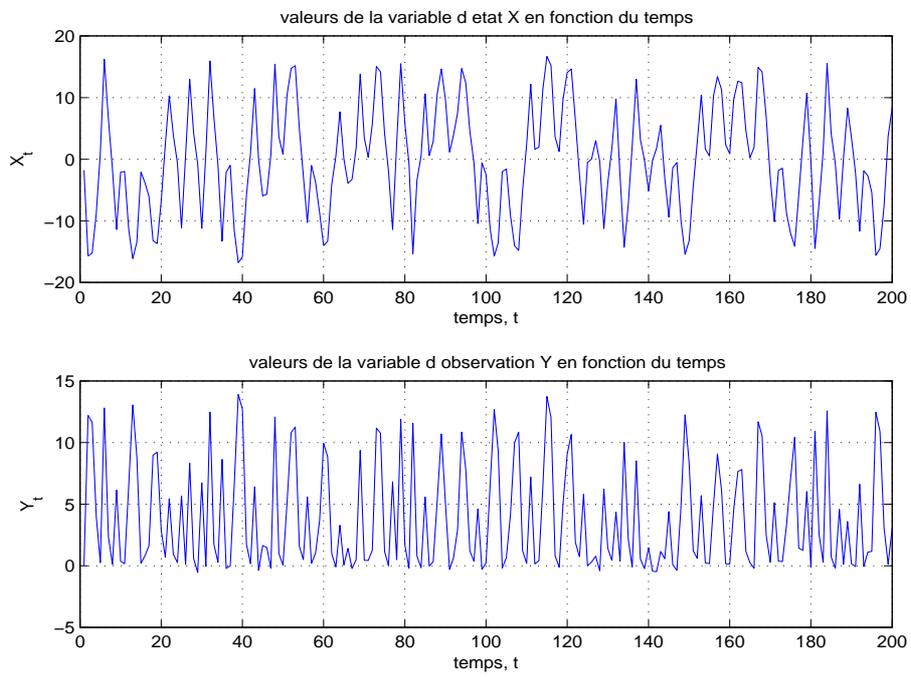


FIG. 6.8 – Graphiques des variables d'état et d'observation avec une panne à l'instant $t_p = 120$

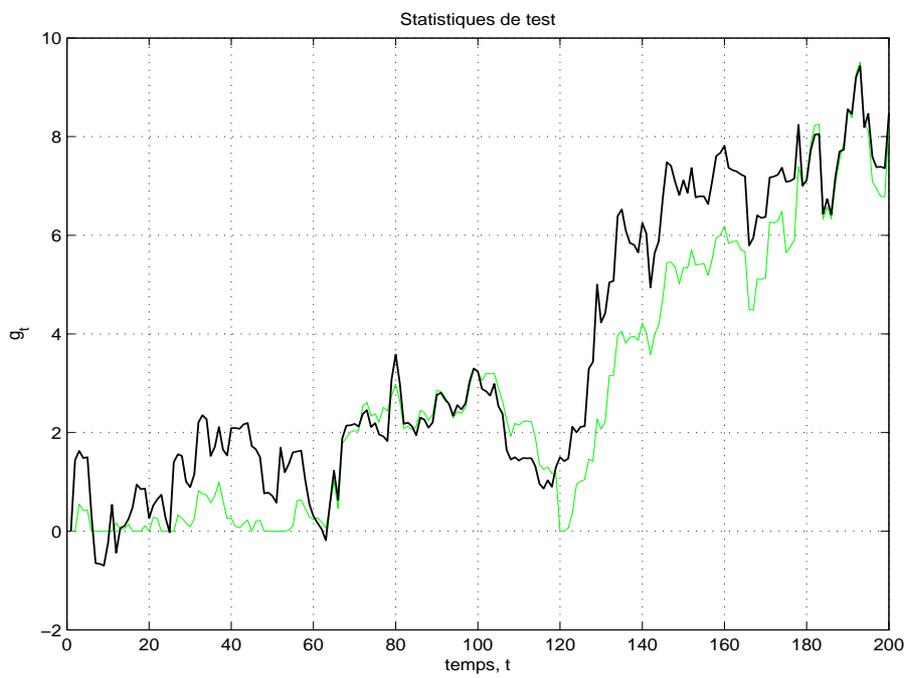


FIG. 6.9 – Graphe représentant les statistiques de test des règles de décision : \hat{t}_{lim} en foncé et \hat{t}_{app} en clair

tion et une grande pour la phase de déplacement des particules) et on choisirait de retenir l'estimation la plus élevée.

6.4 L'algorithme de filtrage-détection comme concurrent du GLR

6.4.1 Introduction

L'algorithme de filtrage-détection présenté dans les parties précédentes peut s'adapter à un cas très particulier, celui où les variables d'état sont observées. La modélisation du processus n'est plus dans ce cas un modèle à espace d'état, mais tout simplement un modèle auto-régressif de la forme :

$$X_n = f(X_{n-1}, \theta) + \epsilon_n, \quad \epsilon_n \sim \mathcal{N}(0, \sigma^2). \quad (6.33)$$

Nous allons supposer que la valeur θ_0 du paramètre caractérisant le changement est connue. On peut ainsi obtenir directement à partir du modèle (6.33) la vraisemblance conditionnelle des observations $p_{\theta_0}(X_n|X_{1:n-1})$. D'autre part la valeur du paramètre sous régime de panne, θ_1 , est inconnue. L'approche que nous proposons va consister à considérer un modèle à espace d'état "simplifié", uniquement dans le but d'estimer le paramètre inconnu, où l'équation d'observation est réduite à sa plus simple expression : $Y_n = X_n$. La modélisation du système sous l'hypothèse d'un changement de régime à un instant j sera la suivante :

$$X_n = f(X_{n-1}, \theta_0) + \epsilon_n, \quad \text{pour } n < j \quad (6.34)$$

et

$$\left\{ \begin{array}{l} \left[\begin{array}{c} X_n \\ \theta_n \end{array} \right] = \left[\begin{array}{c} f(X_{n-1}, \theta_{n-1}) + \epsilon_n \\ \theta_{n-1} \end{array} \right] \\ Y_n = X_n \end{array} \right. \quad \text{pour } n \geq j. \quad (6.35)$$

Nous allons chercher à estimer la quantité $p_{\theta_1}(X_t|X_{1:t-1})$ de sorte à construire, comme dans les parties précédentes, un test du rapport des vraisemblances.

L'algorithme ainsi défini est en fait un concurrent direct du test GLR puisqu'il s'utilise dans les mêmes conditions d'utilisation. Il possède un avantage certain sur le GLR, une écriture sous forme récursive.

REMARQUE : J'adopte dans cette partie deux écritures pour la même quantité, à savoir :

$$p_{H_1}(X_i|\theta, X_{1:i-1}) = p_{\theta}(X_i|X_{1:i-1})$$

On utilisera la première plutôt lorsqu'on considérera θ comme une variable aléatoire et la seconde lorsqu'on le traitera comme un paramètre.

6.4.2 Principe et détail de l'algorithme

Nous allons chercher à estimer la quantité suivante :

$$l_i = \int p_{H_1}(X_i|\theta_{i-1}, X_{1:i-1}) \cdot p_{H_1}(\theta_{i-1}|X_{1:i-1}) d\theta_{i-1}, \quad (6.36)$$

par estimation de Monte Carlo :

$$\hat{l}_i^{N,m} = \frac{1}{m} \sum_{j=1}^m p_{H_1}(X_i|\theta_{i-1}(j), X_{1:i-1}),$$

où $\theta_{i-1}(j)$, $j = 1, \dots, m$, est un échantillon de $\bar{p}_{H_1}(\theta_{i-1}|X_{1:i-1})$, troncature sur Θ_1 de l'estimation $\hat{p}_{H_1}(\theta_{i-1}|X_{1:i-1})$ de la densité $p_{H_1}(\theta_{i-1}|X_{1:i-1})$ obtenue grâce au filtre à noyaux de convolution.

Nous allons définir le temps d'arrêt de la règle de décision de la façon suivante :

$$\tilde{t}_{app}'' = \inf\{n : \max_{1 \leq k \leq n} \sum_{i=k}^n \hat{Z}_i \geq h\}, \quad (6.37)$$

où les incréments de l'algorithme s'écrivent :

$$\hat{Z}_i = \log \frac{\hat{l}_i^{N,m}}{p_{\theta_0}(X_i|X_{1:i-1})}. \quad (6.38)$$

REMARQUE : Nous considérons dans cette partie un modèle auto-régressif (modèle (6.33)). Nous nous situons donc dans le deuxième cas de génération de changement (voir partie 2.3.2) où la densité conditionnelle des observations sous régime de panne, ne dépend pas de l'instant de changement de régime.

Cependant, compte tenu de la modélisation que nous utilisons pour traduire le changement (modèle (6.34) et (6.35)), si nous souhaitons nous inscrire dans la lignée de la partie 6.3, nous pourrions nous intéresser à la vraisemblance :

$$p_{H_0, H_1, j}(X_n|X_{1:n-1}) = \int p_{H_1}(X_n|\theta_{n-1}, X_{1:n-1}) \cdot p_{H_0, H_1, j}(\theta_{n-1}|X_{1:n-1}) d\theta_{n-1}, \quad (6.39)$$

qui, elle, dépend de l'instant de changement de régime supposé j , puisque la densité $p_{H_0, H_1, j}(\theta_{n-1}|X_{1:n-1})$ dépend évidemment de cet instant j . En procédant de la sorte, il faudrait ainsi considérer à nouveau autant de filtres que d'instant de changement de régime possibles.

Dans cette partie, nous cherchons avant tout à construire un concurrent du GLR, sur le plan de l'efficacité pour la détection, mais aussi sur le plan de la rapidité de

calcul, c'est pourquoi nous n'allons considérer qu'un filtre sous l'hypothèse H_1 , celui où le changement à lieu à l'instant initial, pour estimer la quantité l_i . La règle de décision utilisée se rapproche donc de celle notée \hat{t}_{app} dans les parties précédentes. L'estimation sera a priori moins précise car contaminée par des observations issues du régime H_0 . La quantité l_i va cependant être un estimateur convergent (sur N et m mais aussi et surtout sur le temps n) de la vraisemblance conditionnelle $p_{\theta_1}(X_n|X_{1:n-1})$.

Dans les faits, quelles seront les différences entre les deux algorithmes? Si on considère le filtre supposant un changement de régime au pas de temps j , à cet instant, comme dans la partie 6.3, on va tirer, suivant une loi a priori p_0 sur le paramètre θ , des particules $\bar{\theta}$. Pour la procédure \hat{t}_{app}'' , à l'instant j , on va tirer des particules suivant l'estimation obtenue au pas de temps $j-1$, estimation éventuellement mauvaise si on était, à l'instant $j-1$, sous le régime H_0 . La différence entre les deux méthodes peut se voir comme une simple différence entre les lois a priori sur le paramètre θ_1 utilisées à l'instant j .

L'algorithme

Initialisation : $t = 1$

- Génération de N réalisations $\bar{\theta}_0(i), i = 1, \dots, N$ de la loi a priori $p_0(\theta)$ de θ_1 sur Θ_1 .
- Génération de N réalisations du bruit $\epsilon_1(i), i = 1, \dots, N$ selon la loi simulable \mathcal{L}_ϵ

Le modèle (6.35) conduit aux réalisations :

$$\tilde{\theta}_1(i) = \bar{\theta}_0(i) \quad \text{et} \quad \tilde{Y}_1(i) = \tilde{X}_1(i) = f(X_0, \bar{\theta}_0(i)) + \epsilon_1(i)$$

- A partir de l'observation Y_1 , les N particules $(\tilde{\theta}_1(i), \tilde{X}_1(i))$ permettent la construction d'une estimation par noyau $\hat{p}(\theta_1|X_1)$ de la densité conditionnelle $p(\theta_1|X_1)$.
- On détermine la troncature $\bar{p}(\theta_1|X_1)$ de $\hat{p}(\theta_1|X_1)$ sur Θ_1 .

Temps t : $t > 1$

Etape 1 :

- A partir des estimations obtenues à l'instant $t-1$, génération de :
 $\bar{\theta}_{t-1}(i) \sim \bar{p}_{H_1}(\theta_{t-1}|X_1, \dots, X_{t-1})$ et $\epsilon_t(i) \sim \mathcal{L}_{\epsilon_t}, i = 1, \dots, N$
- A partir de l'observation X_{t-1} obtenue à l'instant $t-1$, le modèle (6.35) conduit à :

$$\tilde{\theta}_t(i) = \bar{\theta}_{t-1}(i) \quad \text{et} \quad \tilde{Y}_t(i) = \tilde{X}_t(i) = f(X_{t-1}, \bar{\theta}_{t-1}(i)) + \epsilon_t(i), \quad i = 1, \dots, N$$

Etape 2 :

Chapitre 6. Algorithmes de filtrage-détection

- Génération de m réalisations : $\theta_{t-1}(j) \sim \bar{p}_{H_1}(\theta_{t-1}|X_1, \dots, X_{t-1})$, $j = 1, \dots, m$
- A partir du modèle (6.35), dès que la mesure X_t est disponible, calcul de :

$$\hat{l}_t^{N,m} = \frac{1}{m} \sum_{j=1}^m p_{H_1}(X_t|\theta_{t-1}(j), X_1, \dots, X_{t-1}) ,$$

Etape 3 :

- Calcul des incréments \hat{Z}_t :

$$\hat{Z}_t = \log \frac{\hat{l}_t^{N,m}}{p_{\theta_0}(X_t|X_1, \dots, X_{t-1})} ,$$

et si

$$\hat{g}_t = \max_{1 \leq k \leq t} \sum_{i=k}^t \hat{Z}_i \geq h$$

où h est le seuil fixé par l'expérimentateur, une alarme est déclenchée. Dans le cas contraire la surveillance se poursuit par l'étape 4.

Etape 4 :

- A partir de $\tilde{\theta}_t(i)$, $\tilde{X}_t(i)$ et de l'observation Y_t , nous estimons le filtre optimal par une approche à noyau :

$$\hat{p}_{H_1}(\theta_t|X_1, \dots, X_t) = \frac{\sum_{i=1}^N K_{\delta_N}^x(\tilde{X}_t(i) - X_t) \cdot K_{\delta_N}^\theta(\tilde{\theta}_t(i) - \theta_t)}{\sum_{i=1}^N K_{\delta_N}^x(\tilde{X}_t(i) - X_t)}$$

- On détermine la troncature $\bar{p}_{H_1}(\theta_t|X_1, \dots, X_t)$ de $\hat{p}_{H_1}(\theta_t|X_1, \dots, X_t)$ sur le compact Θ_1 .
- $t = t + 1$ et retour à l'étape 1.

REMARQUE : La statistique de test peut s'écrire de façon équivalente, sous forme récursive : $\hat{g}_t = (\hat{g}_{t-1} + \hat{Z}_t)^+$ avec la notation $m^+ = m$ si $m > 0$ et 0 sinon. Cette règle de décision est donc très avantageuse du point de vue de la durée des calculs.

6.4.3 Résultats d'optimalité

Nous n'allons pas pouvoir appliquer directement les résultats obtenus dans le chapitre 4 puisque la quantité $\hat{l}_n^{N,m}$ ne converge pas vers la vraisemblance conditionnelle des observations $p_{\theta_1}(X_n|X_{1:n-1})$ pour tout n mais seulement asymptotiquement sur n . Pour obtenir l'optimalité asymptotique au sens de Lorden de la procédure proposée, nous nous sommes inspirés des résultats obtenus par Lai [50],

sur l'optimalité d'une règle basée sur l'algorithme MLR ("mixture likelihood ratio") présentée dans le chapitre 2. L'algorithme MLR repose sur l'intégration de la densité $p_\theta(x_n|X_{1:n-1})$ par rapport à une loi a priori sur θ . La principale difficulté pour adapter cette démonstration réside dans le fait que, pour la règle \hat{t}_{app}'' , nous intégrons $p_\theta(x_n|X_{1:n-1})$ par rapport à une densité obtenue par filtrage mais impossible à minorer sur un voisinage de θ_1 lorsqu'on considère le supremum essentiel du critère de Lorden sur les trajectoires.

Pour résoudre ce problème, nous allons devoir faire des hypothèses un peu plus restrictives sur la densité conditionnelle $p_\theta(x_n|X_{1:n-1})$.

HYPOTHÈSE : Supposons qu'il existe c et C deux constantes strictement positives telles que :

$$\forall x_i, \forall \theta \in \Theta_1, \quad c \leq p_\theta(x_i|X_{1:i-1}) \leq C. \quad (6.40)$$

On note :

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

et on suppose que :

$$\frac{1}{n} \sum_{i=t_p}^{t_p+n-1} Z_i(\theta_1) \xrightarrow{P} I \quad \text{quand } n \rightarrow \infty.$$

Nous allons faire la même hypothèse que Lai ([50], équation (28)), à savoir :

$\forall \delta > 0, \exists \Theta_\delta \subset \Theta$ et $n(\delta) \geq 1$ tel que $\theta_1 \in \Theta_\delta$, et :

$$\sup_{n \geq n(\delta)} \sup_{k \geq t_p \geq 1} \sup_{\theta_1} \text{ess } P_{\theta_1}^{(t_p)} \left\{ \inf_{\theta \in \Theta_\delta} \sum_{i=k}^{k+n} Z_i(\theta) \leq (I - \delta)n | X_{1:k-1} \right\} \leq \delta. \quad (6.41)$$

Cette hypothèse nous impose que, pour tout δ (et donc pour δ petit), on peut trouver un domaine Θ_δ telle que la quantité $1/n \sum_{i=k}^{k+n-1} Z_i(\theta) - I$, soit inférieure à δ , quel que soit $\theta \in \Theta_\delta$. Cette hypothèse est la généralisation de l'hypothèse (4.13) utilisée dans le chapitre 2.

La règle de décision est définie de la façon suivante :

$$\hat{t}_{app}'' = \inf \left\{ n : \max_{1 \leq k \leq n} \sum_{i=k}^n \hat{Z}_i \geq h \right\}, \quad (6.42)$$

où les incréments s'écrivent :

$$\hat{Z}_i = \log \frac{\frac{1}{m} \sum_{j=1}^m p_{H_1}(X_i|X_{1:i-1}, \theta_{i-1}(j))}{p_{\theta_0}(X_i|X_{1:i-1})}, \quad (6.43)$$

Chapitre 6. Algorithmes de filtrage-détection

où $\theta_{i-1}(j)$ est un échantillon aléatoire tiré de $\bar{p}_{H_1}(\theta|X_{1:i-1})$, obtenue par filtrage.

Le théorème qui suit assure que la procédure de détection \hat{t}_{app}'' atteint la borne inférieure du pire retard moyen à la détection $\bar{E}_{\theta_1}(\cdot)$ défini dans le chapitre 4.

Théorème 6.4 *Si les hypothèses (6.40) et (6.41) sont vérifiées et sous les hypothèses de convergence du filtre à noyaux de convolution, on a :*

$$\bar{E}_{\theta_1}(\hat{t}_{app}'') \leq (I^{-1} + o(1))h,$$

lorsque h , N et m tendent vers l'infini.

Nous avons évoqué dans la partie précédente (partie 6.3) le résultat de Rossi, qui montre, en utilisant les travaux de Schwartz [94], la convergence de l'estimation $\hat{p}(\theta|X_{1:n})$ vers une mesure de Dirac en la vraie valeur du paramètre. Dans notre contexte, malgré la contamination due à X_1, \dots, X_{t_p-1} , asymptotiquement sur n , $\hat{p}_{H_1}(\theta|X_{1:n})$ converge vers un Dirac en θ_1 sous régime de panne. On a alors le lemme suivant, nécessaire pour la démonstration du Théorème 6.4 :

Lemme 6.1 *Soit t_p fini. Pour toute valeur δ , et donc quel que soit l'ensemble Θ_δ défini à partir de l'hypothèse (6.41), on a :*

$$\forall \epsilon > 0, \exists n', \exists m', \exists N' \text{ tel que pour } i \geq n', m > m', N > N' \\ \frac{1}{m} \sum_{j=1}^m p_{H_1}(X_i|X_{1:i-1}, \theta_{i-1}(j)) \geq \inf_{\theta \in \Theta_\delta} p_\theta(X_i|X_{1:i-1}) - \epsilon \quad \text{p.s.}$$

Démonstration du Théorème 6.4 :

On a :

$$\sum_{i=k}^n \hat{Z}_i = \sum_{i=k}^n \log \frac{\frac{1}{m} \sum_{j=1}^m p_{H_1}(X_i|X_{1:i-1}, \theta_{i-1}(j))}{p_{\theta_0}(X_i|X_{1:i-1})}.$$

Ainsi, on obtient : $\forall \delta, \forall \epsilon$, si n , m et N sont suffisamment grands :

$$\begin{aligned}
\sum_{i=k}^n \hat{Z}_i &= \sum_{i=k}^{n'-1} \log \frac{\frac{1}{m} \sum_{j=1}^m p_{H_1}(X_i | X_{1:i-1}, \theta_{i-1}(j))}{p_{\theta_0}(X_i | X_{1:i-1})} + \sum_{i=n'}^n \log \frac{\frac{1}{m} \sum_{j=1}^m p_{H_1}(X_i | X_{1:i-1}, \theta_{i-1}(j))}{p_{\theta_0}(X_i | X_{1:i-1})} \\
&\geq \sum_{i=k}^{n'-1} \log \frac{\frac{1}{m} \sum_{j=1}^m p_{H_1}(X_i | X_{1:i-1}, \theta_{i-1}(j))}{p_{\theta_0}(X_i | X_{1:i-1})} + \sum_{i=n'}^n \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1}) - \epsilon}{p_{\theta_0}(X_i | X_{1:i-1})} \\
&\geq \sum_{i=k}^n \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \\
&\quad + \sum_{i=k}^{n'-1} \log \frac{\frac{1}{m} \sum_{j=1}^m p_{H_1}(X_i | X_{1:i-1}, \theta_{i-1}(j))}{p_{\theta_0}(X_i | X_{1:i-1})} - \sum_{i=k}^{n'-1} \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \\
&\quad + \sum_{i=n'}^n \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1}) - \epsilon}{p_{\theta_0}(X_i | X_{1:i-1})} - \sum_{i=n'}^n \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \\
&\geq \sum_{i=k}^n \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \\
&\quad + \sum_{i=k}^{n'-1} \log \frac{\inf_{\theta \in \Theta_1} p_\theta(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} - \sum_{i=k}^{n'-1} \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \\
&\quad + \sum_{i=n'}^n \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1}) - \epsilon}{p_{\theta_0}(X_i | X_{1:i-1})} - \sum_{i=n'}^n \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} \\
&\geq \sum_{i=k}^n \log \frac{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})} + \sum_{i=k}^{n'-1} \log \frac{\inf_{\theta \in \Theta_1} p_\theta(X_i | X_{1:i-1})}{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})} \\
&\quad + \sum_{i=n'}^n \log \left(1 - \frac{\epsilon}{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})} \right)
\end{aligned}$$

Les deuxième et troisième termes à droite de l'inégalité dépendent ici des observations du passé X_1, \dots, X_{i-1} . C'est justement ce point là qui nous empêche de continuer puisque lorsqu'on va introduire le supremum essentiel sur le passé, on ne va pas pouvoir garantir que ces quantités sont finies. Il faut donc utiliser l'hypothèse (6.40).

Alors,

$$\sum_{i=k}^{n'-1} \log \frac{\inf_{\theta \in \Theta_1} p_\theta(X_i | X_{1:i-1})}{\inf_{\theta \in \Theta_\delta} p_\theta(X_i | X_{1:i-1})} \geq \sum_{i=k}^{n'-1} \log \left(\frac{c}{C} \right)$$

Chapitre 6. Algorithmes de filtrage-détection

Ainsi, $\forall \delta > 0, \forall \epsilon' > 0$, si n, m et N sont suffisamment grands :

$$\begin{aligned} \sum_{i=k}^n \hat{Z}_i &\geq \inf_{\theta \in \Theta_\delta} \left(\sum_{i=k}^n \log \frac{p_\theta(X_i|X_{1:i-1})}{p_{\theta_0}(X_i|X_{1:i-1})} \right) + \sum_{i=k}^{n'-1} \log \frac{\inf_{\theta \in \Theta_1} p_\theta(X_i|X_{1:i-1})}{\inf_{\theta \in \Theta_\delta} p_\theta(X_i|X_{1:i-1})} + \sum_{i=n'}^n \log(1 - \epsilon') \\ &\geq \inf_{\theta \in \Theta_\delta} \left(\sum_{i=k}^n Z_i(\theta) \right) + \sum_{i=k}^{n'-1} \log \frac{C}{C} + (n - k) \log(1 - \epsilon') \end{aligned}$$

Donc, $\forall k, \forall \delta$ et $\forall \epsilon'' > 0$, si n, m et N sont suffisamment grands,

$$\sum_{i=k}^n \hat{Z}_i \geq \inf_{\theta \in \Theta_\delta} \left(\sum_{i=k}^n Z_i(\theta) \right) + Cte - (n - k)\epsilon''. \quad (6.44)$$

En utilisant (6.41) et (6.44), il vient si m et N sont suffisamment grands :

$\forall \epsilon'' > 0, \forall \delta > 0, \exists \Theta_\delta \subset \Theta$ et $n(\delta, \epsilon) \geq 1$ tel que $\theta_1 \in \Theta_\delta$, et :

$$\begin{aligned} &\sup_{n \geq n(\delta)} \sup_{k \geq t_p \geq 1} \sup_{\theta_1} \text{ess} P_{\theta_1}^{(t_p)} \\ &\left\{ \inf_{\theta \in \Theta_\delta} \sum_{i=k}^{k+n} Z_i(\theta) + Cte - n\epsilon'' \leq (I - \delta)n + Cte - n\epsilon'' | X_{1:k-1} \right\} \leq \delta, \end{aligned}$$

c'est à dire,

$$\sup_{n \geq n(\delta)} \sup_{k \geq t_p \geq 1} \sup_{\theta_1} \text{ess} P_{\theta_1}^{(t_p)} \left\{ \sum_{i=k}^{k+n} \hat{Z}_i \leq (I - \delta - \epsilon'')n + Cte | X_{1:k-1} \right\} \leq \delta. \quad (6.45)$$

Soient δ compris strictement entre 0 et 1 et $\epsilon'' > 0$. Considérons n_h le plus petit entier tel que $n_h - 1$ soit supérieur à $I^{-1}(1 - \delta - \epsilon'')^{-1}(h - Cte)$. Alors, si h est assez grand,

$$\sup_{k \geq t_p \geq 1} \sup_{\theta_1} \text{ess} P_{\theta_1}^{(t_p)} \left\{ \sum_{i=k}^{k+n_h-1} \hat{Z}_i \leq I(1 - \delta - \epsilon'')(n_h - 1) + Cte | X_{1:k-1} \right\} \leq \delta$$

Or,

$$\begin{aligned} &I^{-1}(1 - \delta - \epsilon'')^{-1}(h - Cte) < n_h - 1 \\ \Rightarrow &h - Cte < I(1 - \delta - \epsilon'')(n_h - 1) \\ \Rightarrow &h < I(1 - \delta - \epsilon'')(n_h - 1) + Cte \end{aligned}$$

Et donc, si h est assez grand,

$$\sup_{k \geq t_p \geq 1} \sup_{\theta_1} \text{ess} P_{\theta_1}^{(t_p)} \left\{ \sum_{i=k}^{k+n_h-1} \hat{Z}_i \leq h | X_{1:k-1} \right\} \leq \delta. \quad (6.46)$$

On peut alors suivre la même démonstration que celle du Théorème 4.4 à partir de l'inégalité (4.29), qui est l'analogue de (6.46) et on obtient finalement si h est grand :

$$\bar{E}_{\theta_1}(\hat{t}_{app}'') \leq (1 - \delta)^{-1} \cdot n_h$$

Or,

$$n_h \sim I^{-1}(1 - \delta - \epsilon'')^{-1}(h - Cte),$$

et donc

$$\begin{aligned} \bar{E}_{\theta_1}(\hat{t}_{app}'') &\leq I^{-1}(1 - \delta)^{-1}(1 - \delta - \epsilon'')^{-1}(h - Cte) \\ &\leq I^{-1}(1 - \delta)^{-1}(1 - \delta - \epsilon'')^{-1}h \left(1 - \frac{Cte}{h}\right) \\ &\leq I^{-1}(1 - \delta)^{-1}(1 - \delta - \epsilon'')^{-1}h(1 + o(1)). \quad \square \end{aligned}$$

Pour établir l'optimalité de la procédure, il faut montrer qu'elle vérifie la contrainte ARL.

Théorème 6.5 *On a le résultat suivant :*

$$E_{\theta_0}(\hat{t}_{app}'') \geq e^h,$$

et donc, si on prend $h = \log \gamma$, la contrainte ARL : $E_{\theta_0}(\hat{t}_{app}'') \geq \gamma$ est vérifiée.

Démonstration : La démonstration de ce théorème est exactement la même que celle de Lai [50] (Théorème 4, ii) dès lors que l'on a le lemme suivant :

Lemme 6.2 *On note \mathcal{F}_n la tribu engendrée par X_1, \dots, X_n . Alors, pour tout k , $\{\exp(\sum_{i=k}^n \hat{Z}_i), \mathcal{F}_n, n \geq k\}$ est une martingale positive de moyenne 1 sous P_{θ_0} .*

Démonstration du lemme :

On a

$$\begin{aligned}
 E_{\theta_0} \left[\exp \sum_{i=k}^n \hat{Z}_i | \mathcal{F}_{n-1} \right] &= E_{\theta_0} \left[\exp \sum_{i=k}^n \log \frac{\hat{l}_i^{N,m}}{p_{\theta_0}(X_i | X_{1:i-1})} | \mathcal{F}_{n-1} \right] \\
 &= E_{\theta_0} \left[\prod_{i=k}^n \frac{\hat{l}_i^{N,m}}{p_{\theta_0}(X_i | X_{1:i-1})} | \mathcal{F}_{n-1} \right] \\
 &= E_{\theta_0} \left[\prod_{i=k}^{n-1} \frac{\hat{l}_i^{N,m}}{p_{\theta_0}(X_i | X_{1:i-1})} \cdot \frac{\hat{l}_n^{N,m}}{p_{\theta_0}(X_n | X_{1:n-1})} | \mathcal{F}_{n-1} \right] \\
 &= \prod_{i=k}^{n-1} \frac{\hat{l}_i^{N,m}}{p_{\theta_0}(X_i | X_{1:i-1})} \cdot E_{\theta_0} \left[\frac{\hat{l}_n^{N,m}}{p_{\theta_0}(X_n | X_{1:n-1})} | \mathcal{F}_{n-1} \right] \\
 &= \exp \sum_{i=k}^{n-1} \hat{Z}_i \cdot \int p_{\theta_0}(x_n | X_{1:n-1}) \frac{\hat{l}_n^{N,m}}{p_{\theta_0}(x_n | X_{1:n-1})} dx_n \\
 &= \exp \sum_{i=k}^{n-1} \hat{Z}_i \cdot \int \hat{l}_n^{N,m} dx_n \\
 &= \exp \sum_{i=k}^{n-1} \hat{Z}_i \cdot \int \frac{1}{m} \sum_{j=1}^m p_{\theta_1}(x_n | X_{1:n-1}, \theta_{i-1}(j)) dx_n \\
 &= \exp \sum_{i=k}^{n-1} \hat{Z}_i \cdot \frac{1}{m} \sum_{j=1}^m \int p_{\theta_1}(x_n | X_{1:n-1}, \theta_{i-1}(j)) dx_n \\
 &= \exp \sum_{i=k}^{n-1} \hat{Z}_i
 \end{aligned}$$

où la dernière égalité vient du fait que $p_{\theta_1}(\cdot | X_{n-1}, \theta_{n-1}(j))$ est une densité et donc d'intégrale égale à 1.

De plus, si nous calculons la moyenne du terme général, nous avons :

$$\begin{aligned}
 E_{\theta_0} \left[\prod_{i=k}^n \exp(\hat{Z}_i) \right] &= E_{\theta_0} \left[\prod_{i=k}^n \frac{\frac{1}{m} \sum_{t=1}^m p_{\theta_1}(X_i | X_{1:i-1}, \theta_{i-1}(t))}{p_{\theta_0}(X_i | X_{1:i-1})} \right] \\
 &= E_{\theta_0} \left[\frac{\prod_{i=k}^n \frac{1}{m} \sum_{t=1}^m p_{\theta_1}(X_i | X_{1:i-1}, \theta_{i-1}(t))}{p_{\theta_0}(X_k, \dots, X_n | X_{1:k-1})} \right] \\
 &= E_{\theta_0} \left[\prod_{i=k}^n \frac{1}{m} \sum_{t=1}^m p_{\theta_1}(X_i | X_{1:i-1}, \theta_{i-1}(t)) \cdot \frac{p_{\theta_0}(X_1, \dots, X_{k-1})}{p_{\theta_0}(X_1, \dots, X_n)} \right] \\
 &= \int p_{\theta_0}(x_1, \dots, x_n) \cdot \prod_{i=k}^n \frac{1}{m} \sum_{t=1}^m p_{\theta_1}(x_i | X_{1:i-1}, \theta_{i-1}(t)) \cdot \frac{p_{\theta_0}(x_1, \dots, x_{k-1})}{p_{\theta_0}(x_1, \dots, x_n)} dx_1 \dots dx_n \\
 &= \int \prod_{i=k}^n \frac{1}{m} \sum_{t=1}^m p_{\theta_1}(x_i | X_{1:i-1}, \theta_{i-1}(t)) \cdot p_{\theta_0}(x_1, \dots, x_{k-1}) dx_1 \dots dx_n \\
 &= \int \left(\int \frac{1}{m} \sum_{t=1}^m p_{\theta_1}(x_n | X_{1:n-1}, \theta_{n-1}(t)) dx_n \right) \cdot \\
 &\quad \prod_{i=k}^{n-1} \frac{1}{m} \sum_{t=1}^m p_{\theta_1}(x_i | X_{1:i-1}, \theta_{i-1}(t)) \cdot p_{\theta_0}(x_1, \dots, x_{k-1}) dx_1 \dots dx_{n-1} \\
 &= \int \prod_{i=k}^{n-1} \frac{1}{m} \sum_{t=1}^m p_{\theta_1}(x_i | X_{1:i-1}, \theta_{i-1}(t)) \cdot p_{\theta_0}(x_1, \dots, x_{k-1}) dx_1 \dots dx_{n-1} \\
 &= \dots \\
 &= \int p_{\theta_0}(x_1, \dots, x_{k-1}) dx_1 \dots dx_{k-1} \\
 &= 1
 \end{aligned}$$

où nous avons utilisé à plusieurs reprises le théorème de Fubini pour des fonctions positives. Nous avons donc montré que $\{\exp(\sum_{i=k}^n \hat{Z}_i), \mathcal{F}_n, n \geq k\}$ est une martingale positive de moyenne 1 sous P_0 . \square

Corollaire 6.4 *Sous l'hypothèse de stabilité (4.2) sur le rapport de vraisemblance du chapitre 4, sous les hypothèses (6.40) et (6.41), et sous les hypothèses de convergence du filtre à noyaux de convolution, la règle \hat{t}_{app}'' est asymptotiquement (sur γ , N et m) optimale pour le critère du pire retard moyen à la détection parmi toutes les règles vérifiant la contrainte ARL.*

Démonstration : Par le Théorème 6.4, on montre que lorsque $h = \log \gamma$, \hat{t}_{app}'' atteint la borne inférieure du pire retard moyen à la détection des règles vérifiant $E_{\theta_0}(t_a) \geq \gamma$. \square

REMARQUE : Concernant l'hypothèse supplémentaire (6.40) que nous venons d'effectuer, cela revient à prendre dans le modèle de départ (6.33), un bruit blanc borné, ce qui ne constitue pas en soit une hypothèse restrictive en pratique, au contraire, puisque par exemple, pour en revenir au modèle de dépollution biologique, les grandeurs comme les concentrations en biomasse et en substrat sont bornées.

6.4.4 Une autre règle de décision

Nous avons envisagé une légère modification de la règle de décision \hat{t}_{app}'' . On utilise, pour cette règle, à l'instant de changement t_p , une mauvaise estimation de la densité de θ puisque la densité a posteriori est construite à partir de données sous régime de fonctionnement normal, donc sous une mauvaise modélisation. Nous allons alors introduire une "roue de secours", par exemple la loi a priori $p_0(\theta)$, de sorte à construire une règle de décision, combinaison de l'algorithme de filtrage-détection et de la règle MLR, dont le temps d'arrêt s'écrit :

$$\hat{t}_{mix} = \inf\{n : \max_{1 \leq k \leq n} \sum_{i=k}^n \hat{Z}'_i \geq h\}, \quad (6.47)$$

où les incréments s'écrivent :

$$\hat{Z}'_i = \log \frac{l'_i}{p_{\theta_0}(X_i|X_{1:i-1})} \quad (6.48)$$

avec

$$l'_i = \max \left(\int p_{H_1}(X_i|\theta, X_{1:i-1}) \cdot p_{H_1}(\theta|X_{1:i-1}) d\theta; \int p_{H_1}(X_i|\theta, X_{1:i-1}) \cdot p_0(\theta) d\theta \right),$$

où $p_0(\theta)$ est la distribution a priori de θ_1 dans Θ_1 . A chaque pas de temps, nous allons prendre le maximum entre deux quantités :

- $p_{H_1}(X_i|X_{1:i-1}) := \int p_{H_1}(X_i|\theta, X_{1:i-1}) \cdot p_{H_1}(\theta|X_{1:i-1}) d\theta$ qui représente la vraisemblance conditionnelle fournie par le modèle. Et donc, lorsque n devient grand, la vraisemblance conditionnelle p_{H_1} tend vers la "vraie" vraisemblance conditionnelle des observations, p_{θ_1} .

- $p_{H_1}^{p_0}(X_i|X_{1:i-1}) := \int p_{H_1}(X_i|\theta, X_{1:i-1}) \cdot p_0(\theta) d\theta$ qui représente la vraisemblance conditionnelle des observations lorsqu'on suppose que le paramètre θ_1 , vu à nouveau comme une variable aléatoire, a une densité de probabilité égale à la distribution a priori $p_0(\theta)$. La quantité ainsi calculée ne sera généralement pas une estimation très fine de la vraisemblance conditionnelle des observations (à moins d'avoir une loi a priori sur θ_1 très précise).

Dans certain cas, l'estimation, à partir de la loi a priori, sera meilleure que celle basée sur la loi a posteriori. Par exemple, à l'instant de panne t_p , où $p_{H_1}(\theta|X_{1:t_p-1})$ n'a alors été construite qu'à partir d'observations sous régime H_0 . La distribution a posteriori peut alors être totalement aberrante, et dans ce cas, la vraisemblance calculée à partir de p_0 est préférable. Donc l'utilisation de cette "roue de secours" peut sembler-t-il améliorer l'efficacité de l'algorithme. On va voir que ce n'est pas vraiment le cas en simulations.

6.4.5 Simulations

Comparaison des deux règles de décision proposées

Dans cette partie, nous allons comparer, sur deux modèles non linéaires, les retards moyens à la détection des deux algorithmes qui viennent d'être présentés : d'une part l'algorithme de filtrage-détection \hat{t}_{app}'' et d'autre part l'algorithme de filtrage combiné avec l'algorithme MLR, \hat{t}_{mix} . Nous calculons aussi les retards à la détection pour deux algorithmes MLR appliqués directement au modèle (6.33) : un utilisant une loi a priori gaussienne ($t_{MLR-gau}$) sur Θ_1 et un autre une loi a priori uniforme ($t_{MLR-uni}$).

Considérons à présent le modèle suivant :

$$X_n = \frac{X_{n-1}}{2(1 + X_{n-1}^2)} + \theta + \epsilon_n \quad \text{avec} \quad \epsilon_n \sim N(0; \sigma^2).$$

La panne considérée est la suivante : sous H_0 , $\theta = \theta_0 = 0$ et sous H_1 , $\theta = \theta_1 = 0.5$. On considère que la valeur θ_1 est inconnue et appartient à l'intervalle : $\Theta_1 = [0.1; 0.7]$.

Pour le MLR avec loi a priori gaussienne, nous avons choisi une loi $N(0.5; 0.01)$ et nous utilisons la même loi a priori pour \hat{t}_{mix} . Pour les algorithmes de filtrage-détection, nous utilisons $N = m = 1000$ et la fenêtre pour l'estimateur à noyaux vaut $\delta_N = 0.2N^{-1/5}$.

Le retard moyen est calculé à partir de 5000 répétitions et le seuil de chaque test est calculé de sorte que le temps moyen avant la première fausse alarme soit égal à 100.

Nous avons répété cette simulation pour plusieurs valeurs de σ^2 . Les résultats sont regroupés dans la table 6.5.

Ces résultats montrent que finalement, l'algorithme \hat{t}_{mix} est le moins efficace des quatre. C'est en partie dû au fait que sa statistique de test prend de plus grandes valeurs que les trois autres sous H_0 et donc, le seuil obtenu pour vérifier la contrainte $E_{\theta_0}(\hat{t}_{mix}) = 100$, est assez élevé. L'introduction d'une roue de secours n'apporte rien. La règle \hat{t}_{app}'' apparaît comme la mieux adaptée.

Dans la deuxième simulation, nous allons comparer sur un autre modèle les quatre mêmes algorithmes lorsque la loi a priori choisie est moins pertinente. Ainsi

	\hat{t}_{app}''	\hat{t}_{mix}	$t_{MLR-gau}$	$t_{MLR-uni}$
$\sigma^2 = 0.1$	2.6	3.41	3.11	3.14
$\sigma^2 = 0.2$	4.35	5.73	5.23	5.25
$\sigma^2 = 0.5$	10.56	10.82	9.65	9.65

TAB. 6.5 – Comparaison des retard des différentes règles de décision

	\hat{t}_{app}''	\hat{t}_{mix}	$t_{MLR-gau}$	$t_{MLR-uni}$
$\theta_1 = 15, \sigma^2 = 0.2$	18.02	17.92	16.52	17.90
$\theta_1 = 15.8, \sigma^2 = 0.002$	8.27	11.80	14.99	16.51

TAB. 6.6 – Comparaison des retards moyens des différentes règles de décision

pour la loi normale, nous n'allons pas la centrer sur θ_1 comme dans la première simulation, et nous allons considérer un intervalle Θ_1 plus grand pour handicaper la loi a priori uniforme.

Considérons donc le modèle suivant :

$$X_n = \frac{X_{n-1}}{2} + \theta \frac{X_{n-1}}{1 + X_{n-1}^2} + \theta + \epsilon_n \quad \text{avec} \quad \epsilon_n \sim N(0; \sigma^2).$$

Sous H_0 , $\theta = \theta_0 = 16$ et nous allons considérer deux pannes. L'intervalle Θ_1 vaut $[14; 15.9]$ et la loi a priori gaussienne pour \hat{t}_{mix} et $t_{MLR-gau}$ est une $N(15.3; 0.05)$. Nous utilisons toujours le même nombre de particules ($N = m = 1000$).

Les résultats de cette simulation sont donnés dans la table 6.6. Ces résultats montrent la difficulté de choisir la loi a priori pour les algorithmes MLR. Si cette loi n'est pas pertinente, il devient difficile de détecter les pannes rapidement (deuxième ligne). D'autre part, la fenêtre pour les estimations à noyaux de convolution est ici égale à $\delta_N = 0.4N^{-1/5}$. Un meilleur choix de fenêtre améliorant sensiblement les résultats des méthodes à filtrage est certainement possible.

Comparaison avec le GLR

Les premières simulations laissent penser que la règle \hat{t}_{app}'' est la plus adaptée parmi les quatre étudiées. Nous allons maintenant la comparer à une des règles les plus connues et plus utilisées, la règle du GLR (notée t_{GLR} et définie dans le chapitre 2). Considérons le modèle suivant :

$$X_n = 0.5X_{n-1} + \theta + \epsilon_n$$

L'algorithme de filtrage-détection comme concurrent du GLR

où ϵ_n est un bruit blanc gaussien de variance $\sigma^2 = 0.1$.

Sous H_0 , $\theta = \theta_0 = 0$. Nous avons dans un premier temps comparé les deux algorithmes pour deux valeurs de θ_1 telles que :

$$\theta_1 \in [0.2; 0.7].$$

Ces deux valeurs représentent une panne “fine”, c’est à dire difficile à détecter et une panne plus grossière avec une grande variation du paramètre θ .

L'algorithme de filtrage détection est utilisé avec $N = m = 500$. Nous choisissons nos seuils de sorte que le temps moyen avant la première fausse alarme soit égal à 100. Pour l'algorithme de filtrage-détection, le seuil obtenu est $h = 1.3$ et pour le GLR, $h = 3.2$. Voici les résultats obtenus du retard moyen à la détection sur 2000 répétitions :

$$\underline{\theta_1 = 0.6} :$$

$$\text{retard moyen pour } t_{GLR} = 23.18$$

$$\text{retard moyen pour } \hat{t}_{app}'' = 18.97$$

$$\underline{\theta_1 = 0.25} :$$

$$\text{retard moyen pour } t_{GLR} = 3.57$$

$$\text{retard moyen pour } \hat{t}_{app}'' = 3.88$$

On peut voir que l'algorithme de filtrage détection se comporte très bien. Il est même supérieur au GLR pour $\theta_1 = 0.6$ et légèrement inférieur pour $\theta_1 = 0.25$. Ceci s'explique peut-être par le fait que nous avons ici choisi une fenêtre très petite de sorte à détecter les pannes “fines”. Ainsi, pour $\theta_1 = 0.25$, les particules $\tilde{\theta}$ se déplacent moins vite, ce qui entraîne un léger retard à la détection.

Suivant le type de panne (plus ou moins fine), l'algorithme de filtrage détection est donc plus ou moins efficace que le GLR. En pratique, on ne sait juste que $\theta_1 \in \Theta_1$, sans savoir si la panne est fine ou non. Nous avons donc effectué la simulation suivante. Nous avons gardé les mêmes choix de paramètres que dans la partie précédente, les valeurs des seuils de détection n'ont donc pas changé. Par contre, nous avons simulé 2000 trajectoires avec un temps de panne à l'instant $t_p = 50$. Et à cet instant t_p la valeur θ passe de $\theta_0 = 0.71$ à θ_1 où θ_1 est tiré de façon aléatoire pour chaque trajectoire suivant une loi uniforme sur $[0, 0.65]$ (nous ne dépassons pas la valeur 0.65 car les tests ont alors de grandes difficultés à détecter la panne). Lorsque la statistique de test du GLR ou de l'algorithme de filtrage-détection n'a pas franchi le seuil après 100 pas de temps (donc à $t_p + 100$), nous considérons que l'algorithme en question n'a pas réussi à détecter la panne.

Sur les 2000 trajectoires, l'algorithme de filtrage-détection n'a enregistré qu'une seule non-détection (la valeur de θ_1 pour cette trajectoire était de 0.6037) alors que l'algorithme GLR n'a pas détecté 4 pannes ($\theta_1 = 0.6236, 0.6085, 0.6173$ et 0.6418).

De plus, si on calcule les retards à la détection des deux algorithmes sur les trajectoires dont la panne a été détectée, on obtient pour l'algorithme de filtrage-détection un retard de 7.03 contre un retard de 7.41 pour l'algorithme GLR.

On voit donc que sur cet exemple simple, l'algorithme de filtrage-détection semble prendre l'avantage sur le GLR malgré un nombre peu élevé de particules ($N = m = 500$). De plus, au niveau de la quantité des calculs, l'algorithme de filtrage-détection est largement plus rapide que le GLR.

6.5 L'algorithme de filtrage-détection pour le diagnostic de panne

Dans cette partie, nous nous intéressons au problème du diagnostic de panne. Les trois situations qui viennent d'être présentées :

- modèle à espace d'état avec θ_1 connu,
- modèle à espace d'état avec θ_1 inconnu,
- modèle autorégressif avec θ_1 inconnu,

peuvent être le cadre du problème du diagnostic de panne. Nous traitons plus particulièrement le deuxième cas, qui est le plus général. Nous reprenons donc dans cette partie les notations utilisées dans la partie 6.3.

6.5.1 La règle de décision à utiliser

Jusqu'à présent, lorsque nous avons évoqué le problème du diagnostic, notamment pour le calcul du seuil adaptatif (chapitre 3), nous avons utilisé la règle de décision introduite par Nikiforov [73, 74] décrite dans le chapitre 2. Initialement, cette règle de décision a été mise au point pour des données indépendantes. Nous l'avons utilisée pour des données dépendantes en considérant la densité conditionnelle des observations. Mais chaque fois, nous avons considéré la deuxième façon de générer le changement, autrement dit lorsque la densité conditionnelle des observations sous l'hypothèse H_1 ne dépend pas de l'instant de changement de régime. Or ici, ce n'est plus le cas.

Les deux premières approches envisagées pour traiter le diagnostic de panne dans un modèle à espace d'état consistent à appliquer la règle de décision de Nikiforov à partir des statistiques de test construites pour les règles de détection \hat{t}_{lim} et \hat{t}_{app} . Ainsi, comme dans les autres parties, soit on considère une fenêtre glissante sur laquelle on estime par maximum de vraisemblance l'instant supposé de changement

de régime, soit on néglige cette dépendance de l'instant de changement de régime. Dans les deux cas, le temps d'arrêt de l'algorithme et le type de panne seront définis de la façon suivante :

$$\begin{aligned}\hat{t}_{diag} &= \min \{ \hat{t}_1, \dots, \hat{t}_K \} \\ \hat{\nu}_{diag} &= \arg \min \{ \hat{t}_1, \dots, \hat{t}_K \}\end{aligned}$$

où \hat{t}_l est défini par :

$$\hat{t}_l = \inf \left\{ t \geq 1 : \min_{0 \leq r \neq s \leq K} [g_t(r, 0) - g_t(s, 0) - h_{r,s}] \geq 0 \right\},$$

où $h_{r,s}$ représente le seuil du test et peut éventuellement être calculé à partir de la méthode du seuil adaptatif, et la quantité $g_t(r, 0) - g_t(s, 0)$ peut être vue comme la statistique du test de l'hypothèse r contre l'hypothèse s . K représente le nombre d'hypothèses alternatives, autrement dit le nombre de pannes possibles sur le système.

La différence entre les deux approches se situe donc dans le calcul des statistiques de test $g_t(r, 0)$. Dans le premier cas,

$$g_t(r, 0) = \max_{t-M \leq j \leq t} \sum_{i=j}^t \log \frac{\hat{l}_{r,i,j}^{N,m}}{\hat{l}_{0,i}^{N,m}},$$

et on notera \hat{t}_{diag}^{lim} et $\hat{\nu}_{diag}^{lim}$ le temps d'alarme et le type d'alarme correspondants.

Dans le deuxième cas,

$$g_t(r, 0) = \max \left(g_{t-1}(r, 0) + \log \frac{\hat{l}_{r,i,1}^{N,m}}{\hat{l}_{0,i}^{N,m}}, 0 \right),$$

avec les notations \hat{t}_{diag}^{app} et $\hat{\nu}_{diag}^{app}$ pour les temps d'arrêt et type de panne.

La mise au point de ces deux règles de décision nécessite donc d'obtenir, à l'instant t , les quantités $\hat{l}_{0,t}^{N,m}$ et $\hat{l}_{r,t,j}^{N,m}$ pour $r = 1, \dots, K$ et $j = t - M, \dots, t$ pour \hat{t}_{diag}^{lim} et $\hat{l}_{0,t}^{N,m}$ et $\hat{l}_{r,t,1}^{N,m}$ pour $r = 1, \dots, K$ pour \hat{t}_{diag}^{app} . L'ensemble des modèles décrivant le comportement du système sous l'hypothèse H_1 utilisés dans le cas de la détection devra donc être enrichi par les modèles décrivant toutes les autres hypothèses alternatives. Pour la règle \hat{t}_{diag}^{lim} , on aura donc $K(M + 1) + 1$ filtres à faire tourner en parallèle, suivant l'algorithme décrit dans la partie 6.3. Pour la règle \hat{t}_{diag}^{app} , il n'y aura que $K + 1$ filtres.

On peut encore envisager une troisième règle de décision, basée sur la procédure de diagnostic proposée par Lai [51]. Le temps d'arrêt de l'algorithme et le type de panne sont alors définis de la façon suivante :

$$\hat{t}_{diag}^{lai} = \min_{1 \leq r \leq K} \hat{t}_r^{lai}, \quad \hat{\nu} = \arg \min \{ \hat{t}_r^{lai} : 1 \leq r \leq K \}$$

avec

$$\hat{t}_r^{lai} = \inf \left\{ n : \max_{n-M \leq k \leq n} \min_{0 \leq s \neq r \leq K-1} \sum_{i=k}^n \frac{\hat{t}_{r,i,k}^{N,m}}{\hat{t}_{s,i,k}^{N,m}} \geq h \right\}.$$

6.5.2 Optimalité des approches proposées

Lai [51] a obtenu l'optimalité asymptotique de la règle \hat{t}_r^{lai} lorsqu'on a accès aux vraisemblances conditionnelles $p_{\theta_i}(Y_n|Y_{1:n-1})$ pour $i = 1, \dots, K$ sous des hypothèses de stabilité du rapport des vraisemblances entre les différentes hypothèses. Le résultat de Lai est en fait l'adaptation directe des résultats de Lai [50] au cas du diagnostic. On pourrait peut-être envisager d'étendre les résultats du chapitre 4 au cas du diagnostic, puisque ces résultats que nous avons obtenu sont la généralisation de ceux de Lai [50] au cas où la densité conditionnelle n'est pas accessible.

Pour les deux premières règles, $(\hat{t}_{diag}^{lim}, \hat{\nu}_{diag}^{lim})$ et $(\hat{t}_{diag}^{app}, \hat{\nu}_{diag}^{app})$, cela paraît plus compliqué puisque les seuls résultats d'optimalité obtenus lorsqu'on connaît la densité des observations, ne s'appliquent qu'au cas indépendant.

6.5.3 Simulations

Nous allons appliquer les méthodes de diagnostic proposées à un modèle à espace d'état non linéaire avec trois hypothèses alternatives et les paramètres caractérisant les pannes seront supposés inconnus.

Plus précisément, nous nous intéressons au modèle :

$$\begin{cases} X_n = 0.5X_{n-1} + b \frac{X_{n-1}}{1+X_{n-1}^2} + c \cdot \cos(1.2 \cdot n) + v_{n-1} \\ Y_n = \frac{1}{20}X_n^2 + w_n \end{cases} \quad (6.49)$$

avec $v_n \sim \mathcal{N}(0; 10)$ et $w_n \sim \mathcal{N}(0; 1)$. Nous considérons trois pannes sur les paramètres b et c :

- sous H_0 , $\theta_0 = [b_0; c_0] = [25, 8]$,
- sous H_1 , $\theta_1 = [b_1; c_1] = [25/2, 8]$,
- sous H_2 , $\theta_2 = [b_2; c_2] = [25, 4]$,
- sous H_3 , $\theta_3 = [b_3; c_3] = [25/2, 4]$.

Avant d'aller plus loin, il est intéressant de s'attarder sur la figure 6.10. Elle représente les graphes des variables d'état et d'observation pour les trois types de panne lorsque la panne intervient à l'instant $t_p = 120$. Les pannes sont visibles, on voit clairement le changement de régime et on pourrait même (à partir des variables d'état) localiser le type de panne puisque chaque panne semble entraîner un comportement particulier de la variable d'état. Par contre, il en va différemment pour la variable d'observation. En effet, les trois pannes entraînent un comportement

identique de la variable d'observation. En pratique, nous n'avons accès qu'à cette dernière, et si dans ce cas là, la panne est visible "à l'écran", il est impossible de différencier le type de panne. C'est un des exemples qui montrent tout l'intérêt de mettre au point une méthode de diagnostic efficace.

Les ensembles Θ_j , $j = 1, 2, 3$, supposés connus sont :

$$\Theta_1 = [10; 20] \times \{8\}, \quad \Theta_2 = \{25\} \times [0; 7], \quad \text{et} \quad \Theta_3 = [10; 20] \times [0; 7].$$

Panne de type 1 :

Tous les algorithmes de filtrage-détection pour obtenir les quantités $\hat{l}_{r,n,j}^{N,m}$ tournent avec $N = 2000$ et $m = 2000$. Nous simulons à l'instant $t_p = 120$, un panne de type 1 (passage sous H_1).

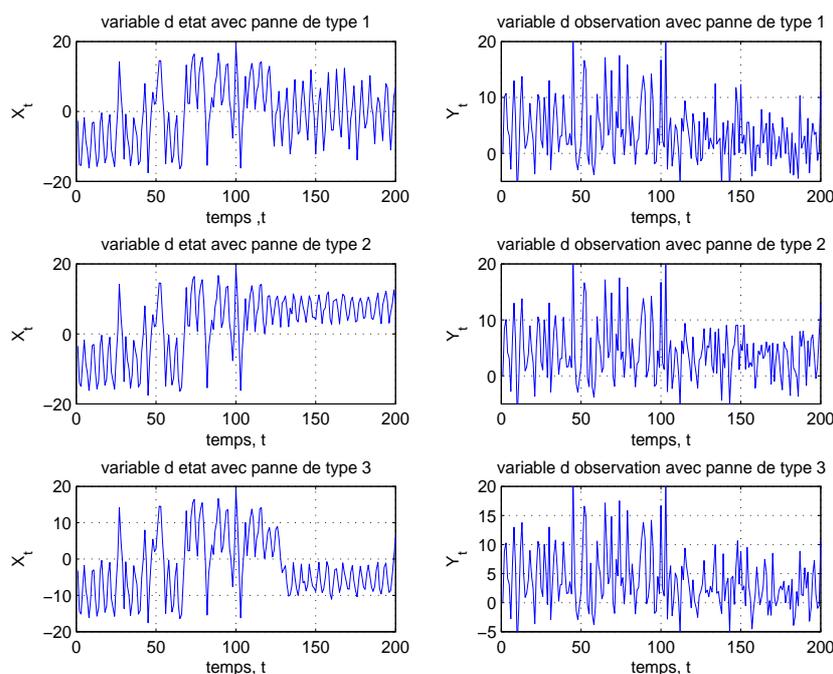


FIG. 6.10 – Graphes des variables d'état et d'observation pour les 3 types de panne

Nous avons représenté sur la figure 6.11 les neuf statistiques de test obtenues grâce à la règle \hat{l}_{diag}^{lim} (nous avons utilisé une fenêtre de taille $M = 80$). Ces neuf statistiques de test représentent les quantités $g_t(r, 0) - g_t(s, 0)$. Rappelons qu'il y a une alarme lorsque les trois statistiques de test d'une même ligne franchissent les seuils de détection. Nous n'avons pas ici calculé de seuil mais on peut voir clairement que

la panne de type 1 est localisée puisque les trois statistiques de test correspondantes “décollent” fortement à partir d’environ $t = 140$. De la même façon, sur la figure 6.12, on a représenté les tests de la règle \hat{t}_{diag}^{app} et là encore, la détection et la localisation ont lieu sans problème. Enfin, sur la figure 6.13, on a représenté les tests de la règle \hat{t}_{diag}^{lai} , mettant en jeu trois statistiques de test finales, et le diagnostic s’effectue bien.

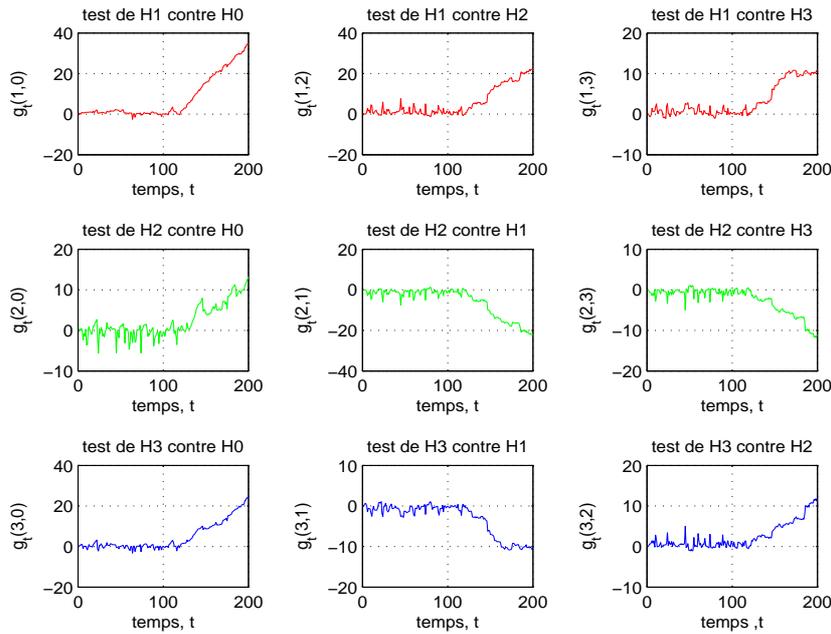


FIG. 6.11 – Statistiques de test de la règle \hat{t}_{diag}^{lim} pour une panne de type 1 simulée à $t_p = 120$

Nous avons représenté dans la figure 6.14 une estimation du paramètre b_1 obtenue grâce à la règle \hat{t}_{diag}^{app} . On peut voir que sous H_0 , la moyenne est élevée, les particules cherchent à atteindre la vraie valeur du paramètre, qui sous H_0 vaut 25. Dès que le système est en état de panne, les particules commencent à se déplacer vers la nouvelle vraie valeur de b ($= b_1 = 12.5$). On voit donc que l’estimation devient de plus en plus correcte au cours du temps.

Pannes de type 2 et 3 :

Les trois règles de décision semblent assez efficaces pour détecter et localiser le bon type de panne sur le premier exemple. Il reste à vérifier que les autres types de panne sont détectés et localisés. Les deux figures 6.15 et 6.16 représentent respectivement les statistiques de test de la règle \hat{t}_{diag}^{lim} pour une panne de type 2 et une

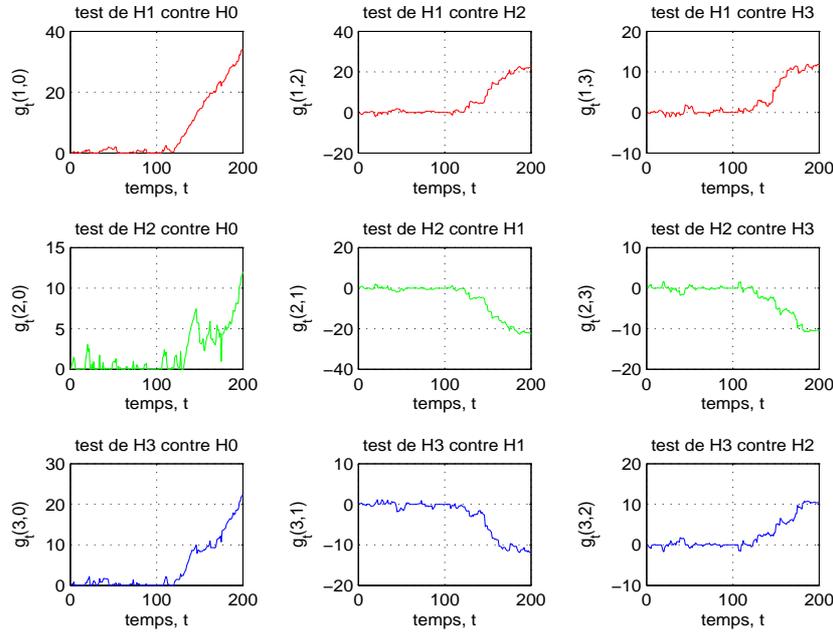


FIG. 6.12 – Statistiques de test de la règle \hat{t}_{diag}^{app} pour une panne de type 1 simulée à $t_p = 120$

panne de type 3. On voit dans les deux cas que la localisation s'effectue sans erreur dans les deux cas.

Nous n'avons pas représenté ici les résultats obtenus sur les pannes de type 2 et 3 pour les règles \hat{t}_{diag}^{app} et \hat{t}_{diag}^{lai} . En répétant les simulations plusieurs fois, nous avons remarqué que ces deux règles localisent bien la panne de type 3 mais ont tendance à commettre une erreur de localisation lorsque la panne est de type 2, en émettant une alarme de type 3.

Il serait peut-être nécessaire d'approfondir ces simulations, notamment en comparant les retards à la détection, pour voir si une d'elles prend le dessus sur les deux autres, la règle \hat{t}_{diag}^{app} ayant tout de même un avantage non négligeable, en terme de quantité de calcul puisque le nombre de filtres en parallèle est nettement inférieur à celui des deux autres approches. Mais d'autre part, sur cette simulation, la règle \hat{t}_{diag}^{lim} est la seule qui arrive à bien localiser une panne de type 2.

6.6 Conclusions

Nous avons proposé dans ce chapitre des règles de décision construites à partir d'un filtre à noyaux de convolution afin de répondre au problème de la détection et

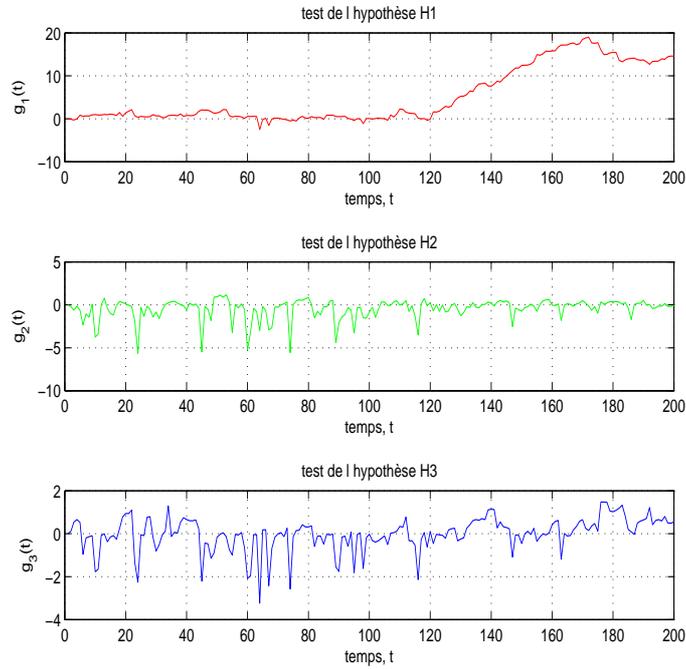


FIG. 6.13 – Statistiques de test de la règle \hat{t}_{diag}^{lai} pour une panne de type 1 simulée à $t_p = 120$

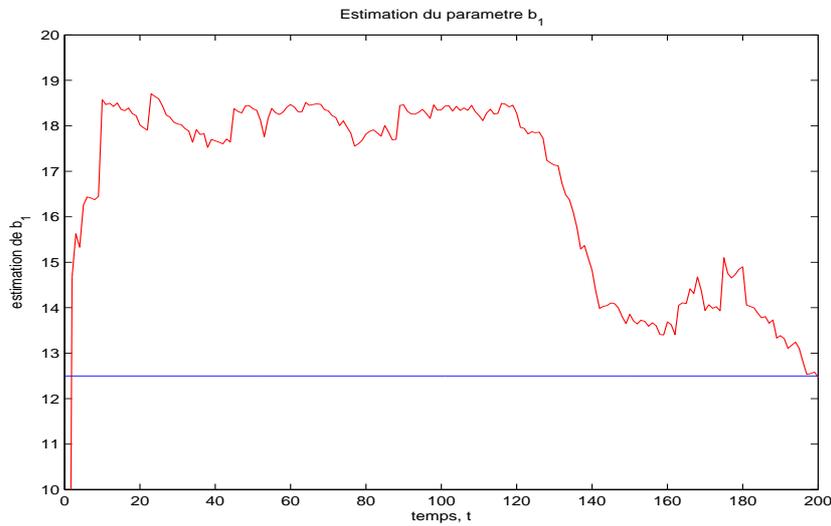


FIG. 6.14 – Estimation du paramètre b_1 au cours du temps obtenue grâce à la règle \hat{t}_{diag}^{app}

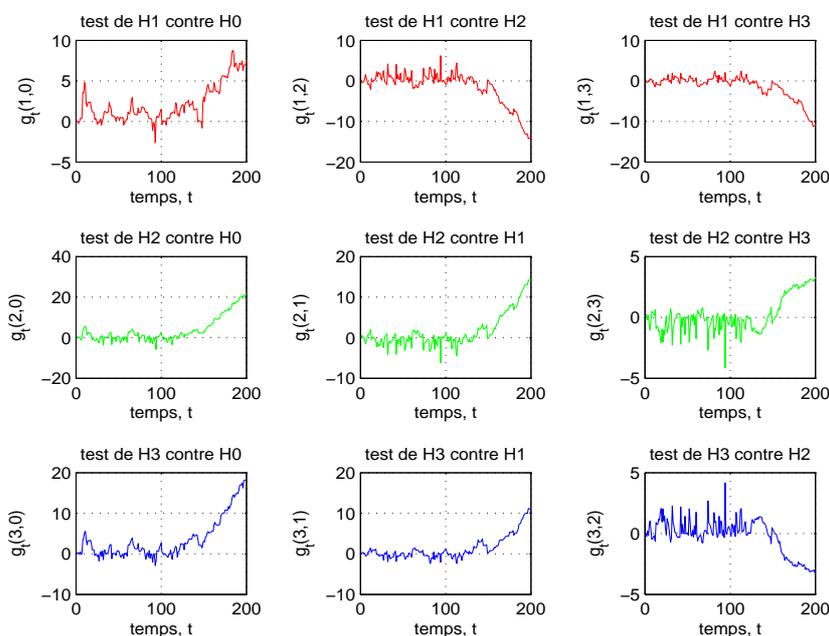


FIG. 6.15 – Statistiques de test de la règle \hat{t}_{diag}^{lim} pour une panne de type 2 simulée à $t_p = 120$

du diagnostic dans les modèles à espace d'état.

Nous avons tout d'abord considéré le cas le plus simple de la détection où les valeurs θ_0 et θ_1 du paramètre caractérisant la panne sont connues. A l'aide des résultats obtenus dans le chapitre 4, nous avons établi l'optimalité de la procédure proposée, sous des hypothèses de stabilité du rapport de vraisemblance qui reste malheureusement difficilement vérifiables en pratique.

L'algorithme initialement proposé nécessite d'ajouter, à chaque pas de temps, un filtre supplémentaire ce qui rend son utilisation en pratique difficile. Deux solutions ont été envisagées : d'une part l'utilisation d'une fenêtre glissante comme c'est généralement le cas pour le GLR, et qui permet de garder un nombre constant de filtres en parallèle, et d'autre part, une approximation qui consiste à négliger la dépendance de la densité conditionnelle des observations envers l'instant d'apparition de la panne. Nous avons obtenu l'optimalité asymptotique de la première solution pour deux critères, celui du pire retard moyen à la détection et celui, plus souple, du retard moyen. Pour la deuxième solution, il semblerait qu'il soit possible d'obtenir l'optimalité en supposant qu'asymptotiquement sur le temps, la densité conditionnelle ne dépende plus de l'instant de changement. Cette hypothèse n'étant finalement pas très incohérente. Les résultats obtenus en simulation montrent des comportements très proches des deux règles de décision.

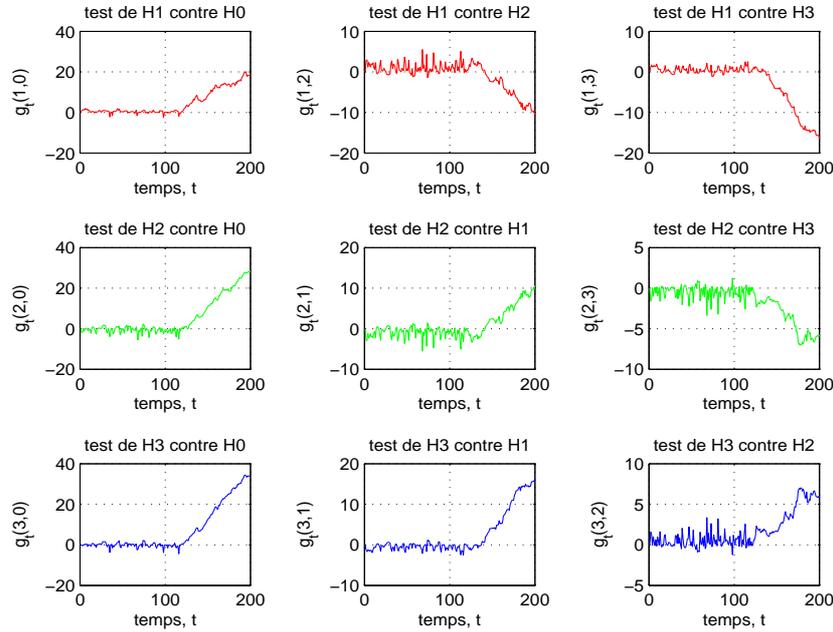


FIG. 6.16 – Statistiques de test de la règle \hat{t}_{diag}^{lim} pour une panne de type 3 simulée à $t_p = 120$

Nous avons ensuite étudié un cas plus réaliste en pratique, celui où la valeur θ_1 caractérisant la panne est inconnue et appartient à un compact Θ_1 connu. L'approche considérée consiste alors à traiter le paramètre θ comme une variable d'état à estimer.

Les premiers résultats obtenus en simulation sont relativement bons malgré l'augmentation de la dimension de l'état. Il semble cependant qu'il soit nécessaire de s'intéresser de plus près au choix de la fenêtre des estimateurs à noyaux.

L'optimalité de cette approche a été obtenue dans le cas d'une observation directe de la variable d'état X . Les conditions d'utilisation de l'algorithme de filtrage-détection sont alors les mêmes que le GLR, et il semblerait qu'il soit au moins aussi efficace que ce dernier, bénéficiant en outre d'une durée de calculs largement plus faible.

La dernière partie de ce chapitre est consacrée au problème du diagnostic. Nous avons adapté deux algorithmes, ceux de Nikiforov [73] et Lai [51], en calculant les statistiques de test à partir de la procédure de filtrage-détection. Il serait intéressant d'étudier les propriétés de ces algorithmes.

Un autre problème peut être traité à partir de l'algorithme de filtrage-détection. Il s'agit du cas où la valeur du paramètre θ_0 est inconnue et appartient à un compact Θ_0 . Il suffit alors de considérer dans la modélisation qui décrit le fonctionnement

normal du système, le paramètre θ comme une variable aléatoire, comme nous l'avons fait pour le paramètre θ_1 dans la partie 6.3.

Chapitre 7

Applications

7.1 Introduction

Dans ce chapitre, nous allons mettre en œuvre les outils mis au point dans les chapitres précédents sur un problème réel, à savoir un procédé biologique de dépollution, et plus particulièrement, un procédé de digestion anaérobie. Le système dynamique que nous considérons modélise le fonctionnement d'un bioréacteur de traitement des boues de vinification après les vendanges. Ce modèle a été mis au point au Laboratoire de Biotechnologie de l'Environnement (LBE) de l'INRA à Narbonne il y a une dizaine d'années. Il se montre robuste et efficace même dans des conditions "extrêmes", puisque comme nous allons le voir, il s'adapte bien aux "données-capturées" malgré de grosses variations appliquées par l'expérimentateur dans le débit d'alimentation. D'autre part, il est aussi robuste dans le temps puisque la plupart des paramètres du modèle pour les données que nous allons utiliser, qui datent de 2001, ont été identifiés en 1997 (voir Steyer et Bernard [102]).

La digestion anaérobie consiste en la transformation de la matière organique (les substrats) en biogaz, réalisée par des micro-organismes (les biomasses). Une présentation des procédés biotechnologiques est effectuée en annexe C.

De nombreuses pannes peuvent avoir lieu sur ce type de procédé : fuite d'une pompe, variation de la concentration initiale en substrat, problème de capteur... Il est donc nécessaire de pouvoir détecter ces défaillances qui peuvent parfois conduire à des conséquences plus graves comme l'arrêt du système.

D'autre part, certaines quantités ne sont pas toujours mesurables à l'intérieur du bioréacteur et il est donc nécessaire d'utiliser des outils de type filtrage. La mise en place du filtre à noyaux de convolution (utilisé dans l'algorithme de filtrage-détection) par Rossi [90] sur cette application a d'ailleurs donné de bons résultats.

7.1.1 Le modèle

Le modèle que nous considérons met en jeu deux biomasses (X_1 et X_2) et deux substrats (S_1 et S_2). Les équations du modèle sont les suivantes :

$$\begin{cases} \dot{X}_1 = (\mu_1(S_1) - \alpha D)X_1 \\ \dot{X}_2 = (\mu_2(S_2) - \alpha D)X_2 \\ \dot{S}_1 = D(S_1^{in} - S_1) - k_1\mu_1(S_1)X_1 \\ \dot{S}_2 = D(S_2^{in} - S_2) + k_2\mu_1(S_1)X_1 - k_3\mu_2(S_2)X_2 \\ \dot{Z} = D(Z_{in} - Z) \\ \dot{C}_{TI} = D(C_{TI}^{in} - C_{TI}) + k_4\mu_1(S_1)X_1 + k_5\mu_2(S_2)X_2 - q_{CO_2} \end{cases} \quad (7.1)$$

Il s'agit d'un système dynamique défini par des équations différentielles. Nous utiliserons une version discrétisée, selon le schéma d'Euler, de ce modèle tout en considérant des bruits sur chaque équation.

Description des variables

- X_1 est la concentration en bactéries acidogènes.
- X_2 est la concentration en bactéries méthanogènes.
- S_1 est la concentration en substrat organique caractérisée par sa demande chimique en oxygène.
- S_2 est la concentration en acide gras volatils.
- Z représente l'alcalinité totale.
- C_{TI} est la concentration en carbone inorganique.
- D est le taux de dilution et joue le rôle de la variable de contrôle
- q_{CO_2} représente le débit de CO_2 .

Les taux de croissance

Les taux de croissance $\mu_1(S_1)$ et $\mu_2(S_2)$ des biomasses X_1 et X_2 sont respectivement modélisés par des lois de Monod :

$$\mu_1(S_1) = \mu_{max1} \frac{S_1}{K_{S1} + S_1},$$

et de Haldane :

$$\mu_2(S_2) = \mu_{max2} \frac{S_2}{K_{S2} + S_2 + (S_2)^2/K_I},$$

avec pour valeurs de paramètres :

$$\mu_{max1} = 1.2 \quad \text{le taux de croissance maximal pour } X_1.$$

$$\mu_{max2} = 0.74 \quad \text{le taux de croissance maximal pour } X_2.$$

$K_{S_1} = 8.875$ la constante de saturation du substrat S_1 .

$K_{S_2} = 23.2$ la constante de saturation du substrat S_2 .

$K_I = 256$ la constante d'inhibition pour S_2 .

Les valeurs de paramètres utilisées sont celles de Steyer et Bernard [102]. Il y a eu un réajustement de K_{S_1} et K_{S_2} par rapport au modèle initial [10].

Valeurs des paramètres

Les valeurs des paramètres sont obtenues dans [10]. Le terme α représente la proportion de biomasse fixée et donc, non soumise à l'effet de dilution. Ce paramètre est fixé à $\alpha = 0.5$. Les constantes k_i représentent des rendements de conversion et ont pour valeurs :

$$k_1 = 42.14, \quad k_2 = 116.5, \quad k_3 = 268, \quad k_4 = 50.6, \quad k_5 = 343.6, \quad k_6 = 453.$$

Le coefficient k_6 apparaît dans le calcul du dégagement de méthane non détaillé ici. Les grandeurs S_1^{in} , S_2^{in} , Z_{in} et C_{TI}^{in} sont les concentrations initiales de l'alimentation et valent respectivement 9 g/L, 92 g/L, 88 meq/L et 9 mmol/L.

*

* *

Les principales caractéristiques du modèle ont été présentées. Une description plus détaillée du modèle est donnée dans Bernard et al. [10].

7.1.2 Les données

Les données utilisées correspondent à 17 jours de mesures (octobre 2001) du bioréacteur du LBE de Narbonne. Pour la plupart des capteurs, une mesure est effectuée toutes les deux minutes.

Les figures qui suivent représentent l'évolution de certaines variables.

La variable Q_{in} (figure 7.1) est le débit d'alimentation du bioréacteur et la figure 7.2 représente le volume de travail du bioréacteur, qui était initialement de 350 l et qui passe à 850 l après le redémarrage de la pompe de mélange de pied de cuve (environ au sixième jour). Ces deux quantités permettent d'obtenir le taux de dilution D (la variable de contrôle), puisque :

$$D = \frac{Q_{in}}{V}.$$

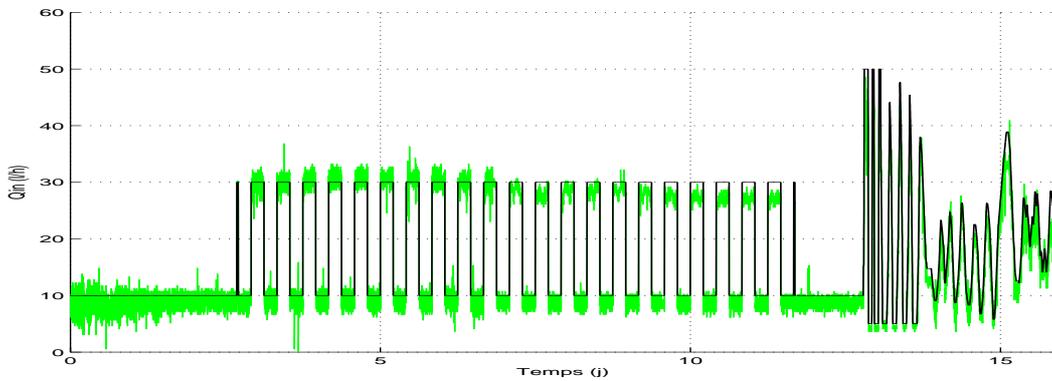


FIG. 7.1 – Tracé du débit théorique en foncé et du débit mesuré Q_{in} en clair.

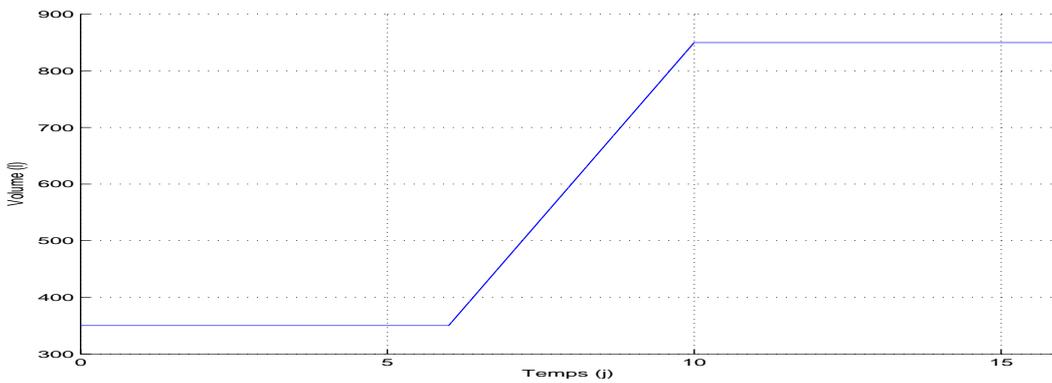


FIG. 7.2 – Volume de travail V du bioréacteur.

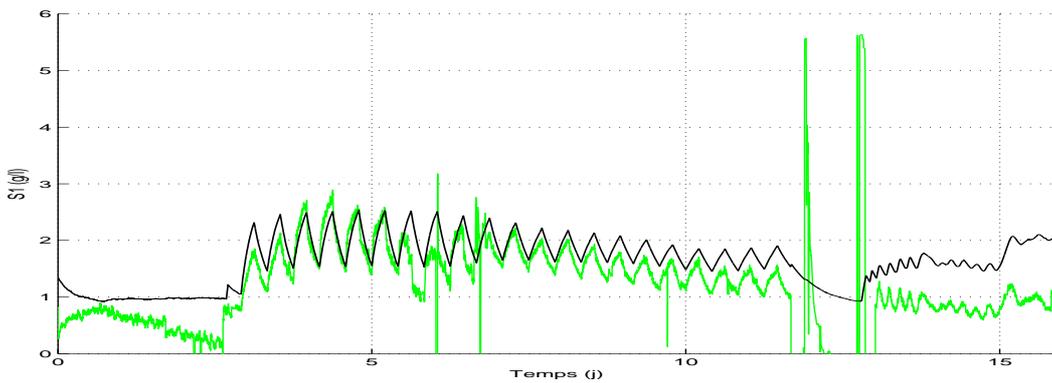


FIG. 7.3 – Evolution de la concentration S_1 : en foncé, le modèle et en clair, les données-capteur

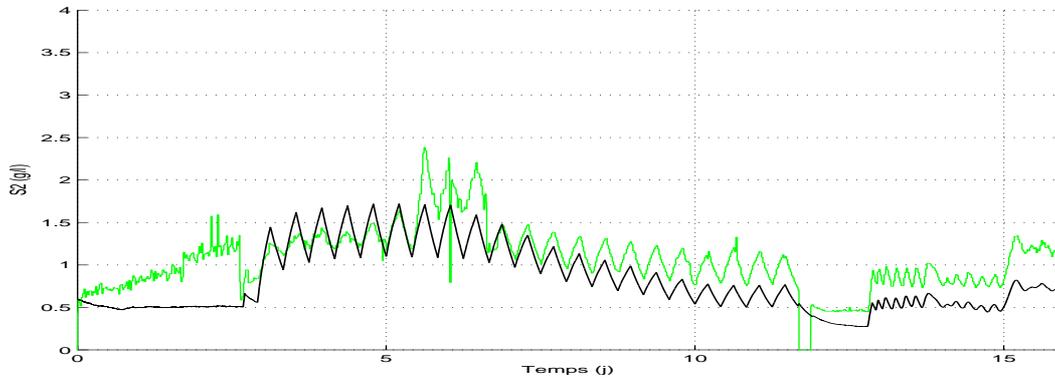


FIG. 7.4 – Evolution de la concentration S_2 : en foncé, le modèle et en clair, les données-capteur

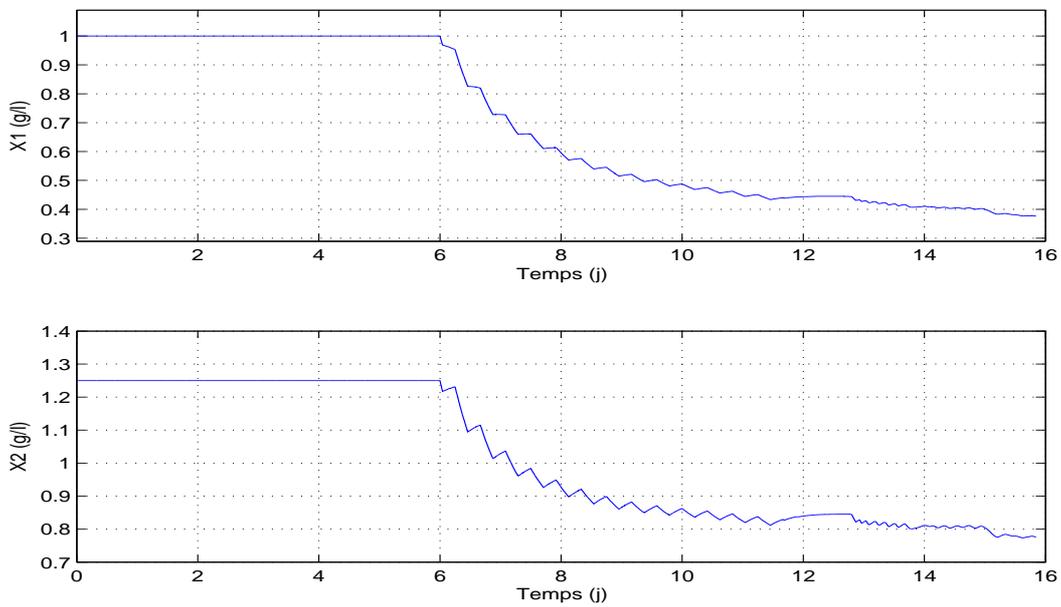


FIG. 7.5 – Evolution des concentrations en biomasse X_1 et X_2 .

La figure 7.5 représente l'évolution des concentrations en biomasses au cours du temps obtenues en simulant le modèle (7.1) du système. En effet, on ne mesure pas ces quantités, qui sont donc, ici, de véritables variables d'état. D'autre part, une hypothèse est effectuée sur l'évolution de ces concentrations : elles sont constantes jusqu'au redémarrage de la pompe. Cette hypothèse est irréaliste en pratique mais donne finalement, une meilleure modélisation du système que celle obtenue en utilisant les équations différentielles sur X_1 et X_2 du modèle (7.1). Après le redémarrage de la pompe, on utilise, pour modéliser le système, le modèle habituel :

$$\begin{aligned}\dot{X}_1 &= (\mu_1(S_1) - \alpha D)X_1 \\ \dot{X}_2 &= (\mu_2(S_2) - \alpha D)X_2.\end{aligned}$$

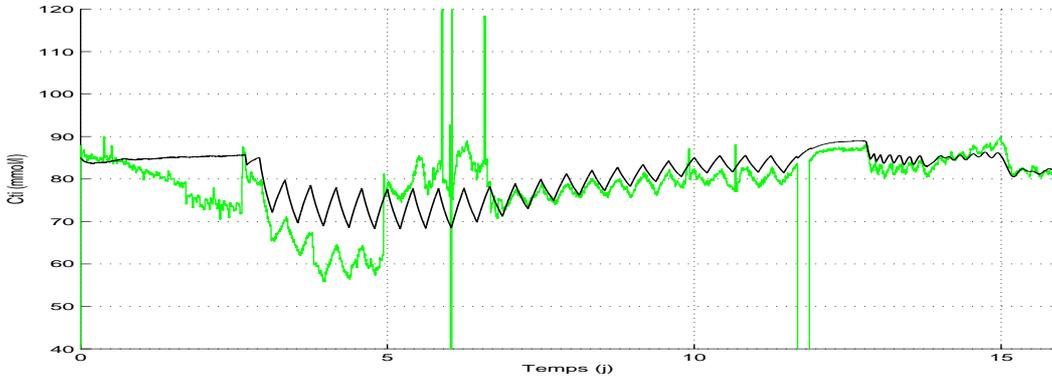


FIG. 7.6 – Evolution de la concentration C_{TI} : en foncé, le modèle et en clair, les données-capteur

L'ensemble de ces courbes fait apparaître des problèmes de capteurs avec des apparitions de biais entre le modèle et les données que nous allons chercher à mettre en évidence.

7.1.3 Les types de panne étudiés

Le système que nous venons de présenter consiste en la dégradation de deux substrats par deux biomasses. On est donc en présence de deux réactions. D'une part, les biomasses X_1 vont transformer le substrat S_1 (voir annexe C) en produit (du CO_2) mais aussi en substrat S_2 et d'autre part, les biomasses X_2 vont transformer le substrat S_2 en produit (CH_4 et CO_2). La deuxième réaction va donc dépendre de la première, comme on peut le voir dans le modèle (7.1). Au contraire, la première réaction, peut se traiter indépendamment de X_2 et S_2 .

Dans cette application nous n'allons considérer que la première réaction mono-substrat mono-biomasse, avec seulement deux variables : X_1 et S_1 . Nous nous sommes alors intéressés à deux types d'anomalies.

La panne de capteur

Dans un premier temps, nous allons chercher à détecter des pannes du capteur sur S_1 . On voit sur la figure 7.3 quelques décrochages (ou valeurs aberrantes) des données-capteur et on s'intéressera plus particulièrement au problème qui se manifeste entre les cinquième et sixième jours avec l'apparition d'un biais entre les valeurs données par le capteur et celles issues de la simple simulation du modèle. Nous utiliserons donc l'algorithme de filtrage détection pour mettre en évidence ces défauts.

Le changement de modèle

La deuxième anomalie considérée est artificielle. Comme nous l'avons dit précédemment, jusqu'au pas de temps $t = 4320$ (environ le sixième jour), on suppose que la concentration en biomasse X_1 est constante. A partir du changement de la pompe, on considère que le substrat X_1 évolue selon l'équation différentielle du modèle (7.1). Un changement de modèle a donc été introduit pour que le modèle suive au mieux les données. Si on suppose que la biomasse reste constante tout au long du processus, le modèle n'est plus aussi bon que pour la bonne modélisation (voir figures 7.3 et 7.7).

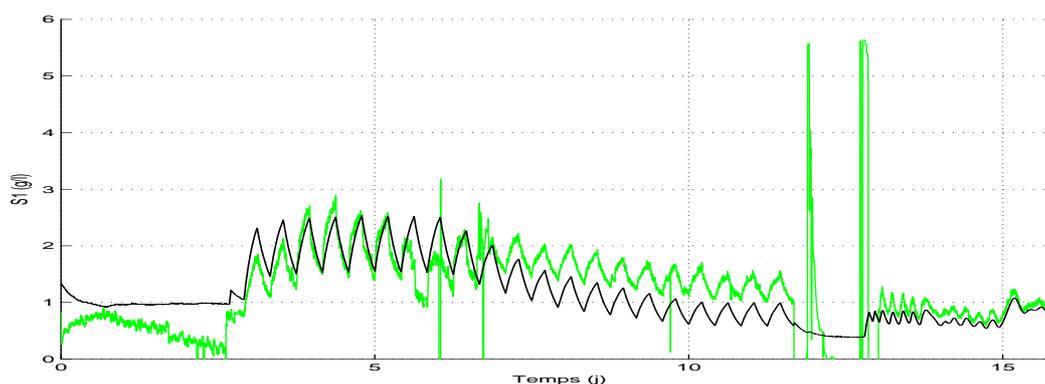


FIG. 7.7 – Evolution de la concentration S_1 : en foncé, le modèle, lorsqu'on suppose que la biomasse reste constante et égale à 1, et en clair, les données-capteur.

Nous allons donc chercher à détecter le changement de modèle sur X_1 à partir des données-capteur sur S_1 à l'aide de l'algorithme de filtrage-détection, en considérant

que le régime normal correspond au cas où la biomasse est constante et qu'à un instant inconnu, on change de modèle.

*
* *

Pour les deux anomalies considérées, nous allons dans un premier temps mettre en place la procédure de détection sur des données simulées pour voir comment se comporte la règle de décision, avant de l'appliquer sur les données réelles.

Jusqu'à présent, sur les graphes, l'unité de temps était les jours. Pour faciliter la lecture des différentes courbes, nous allons considérer comme unité, le pas de temps du modèle discrétisé (voir partie suivante), sachant qu'un pas de temps correspond à deux minutes.

7.1.4 Le cahier des charges de la supervision du procédé

Comme tout type de procédé sous surveillance, l'expérimentateur souhaite que le retard à la détection soit faible tout en ayant un faible taux de fausses alarmes. On peut ici tenter de donner des valeurs de référence pour ces quantités.

En ce qui concerne le retard à la détection, on considère le temps de séjour défini de la façon suivante :

$$\text{temps de séjour} = \frac{V}{Q_{in}}.$$

Le temps de séjour représente le temps théorique pendant lequel une molécule reste dans le réacteur. Pour pouvoir agir de manière efficace sur le système après l'apparition d'une panne, les biotechnologues considèrent que le retard à la détection doit être inférieur à un temps de séjour. Idéalement, il faut que ce retard à la détection soit compris entre 1/10 et 1/5 du temps de séjour.

Dans les données nous intéressant, le temps de séjour varie fortement entre 12h et 48h à cause des variations de Q_{in} . Et lorsque le volume du réacteur passe à 850L, ce temps de séjour varie entre 20h et 100h. Par commodité, nous allons supposer que le temps de séjour est constant et égal à 30h, autrement dit égal à 900 pas de temps. Ainsi, le retard à la détection doit être inférieur à 900 pas de temps, et idéalement inférieur à une centaine de pas de temps (voire deux cent pas de temps).

Pour ce qui concerne le taux de fausses alarmes, l'expérimentateur souhaite généralement avoir, au maximum, une fausse alarme par jour, et idéalement une fausse alarme tous les deux, trois jours.

Dans les simulations qui suivent, nous allons donc mettre en place les seuils adaptatifs proposés dans le chapitre 3, avec un niveau :

$$\alpha = 0.0007.$$

Ainsi, à chaque pas de temps, la probabilité d'avoir une fausse alarme est égale à 0.0007. Ce niveau correspond à un temps moyen entre deux fausses alarmes d'environ 1430 pas de temps, ce qui correspond approximativement à une moyenne de deux jours entre deux fausses alarmes.

Nous pourrions donc par la suite vérifier si les procédures que nous proposons vérifient sur les exemples traités, le cahier des charges de ce procédé de dépollution.

7.2 Panne de capteur

7.2.1 La modélisation stochastique

Le modèle que nous considérons est la version discrétisée des équations différentielles de X_1 et S_1 du modèle (7.1). La variable X_1 n'est pas mesurée. La variable S_1 est mesurée grâce à un capteur. Mais nous allons considérer que la concentration en substrat S_1 est inconnue, puisque nous n'avons accès qu'à une donnée-capteur de cette quantité. X_1 et S_1 sont les deux variables d'état du système. La seule variable d'observation est la donnée-capteur de S_1 , notée C_{S_1} . Le modèle à espace d'état est le suivant :

$$\begin{cases} \begin{bmatrix} X_1(t+1) \\ S_1(t+1) \end{bmatrix} = \begin{bmatrix} X_1(t) + T(\mu_1 - \alpha D(t))X_1(t) + v_{t+1} \\ S_1(t) + T(D(t)(S_1^{\text{in}} - S_1(t)) - k_1\mu_1 X_1(t)) + w_{t+1} \end{bmatrix} \\ C_{S_1}(t+1) = S_1(t+1) - \theta + \epsilon_{t+1}, \end{cases} \quad (7.2)$$

où v , w et ϵ sont des bruits blancs gaussiens indépendants de variance respective : $\sigma_v^2 = 10^{-6}$, $\sigma_w^2 = 10^{-4}$ et $\sigma_\epsilon^2 = 10^{-3}$. Ces valeurs ont été fixées de manière empirique de sorte que les trajectoires simulées par le modèle soient cohérentes. Les conditions initiales sont les suivantes : $X_1(1) = 1$, $S_1(1) = 1.35$.

REMARQUES :

- Nous distinguons ici la concentration réelle en substrat, de la mesure capteur correspondante, puisque nous cherchons justement à détecter une anomalie sur le capteur.

- Dans le modèle (7.2), nous avons utilisé pour équation d'état sur X_1 , la discrétisation de l'équation différentielle de (7.1). Cette équation d'état est valable pour $t > 4320$, c'est à dire après le remplacement de la pompe. Jusque là, la modélisation correcte est donc la suivante :

$$X_1(t+1) = X_1(t) + v_{t+1},$$

où la biomasse est supposée rester constante.

Le paramètre θ caractérise la panne de capteur, ou plus généralement, l'apparition d'un biais entre le modèle et les données.

7.2.2 Détection sur des données simulées

Nous allons étudier le comportement de l'algorithme de filtrage-détection sur la panne considérée. Dans un premier temps, nous nous placerons dans les conditions classiques d'utilisation de la règle de décision, à savoir un changement brutal de valeur du paramètre θ , de θ_0 à θ_1 . Dans un deuxième temps, nous ferons varier la valeur θ_1 de façon graduelle tout en supposant, pour la règle de décision, que le changement est brutal. Ce cas de figure s'apparente davantage au problème du capteur sur S_1 qui apparaît sur les données réelles aux alentours du sixième jour (voir figure 7.3).

Panne abrupte

Les données sont donc simulées à l'aide du modèle (7.2). Sous régime de fonctionnement normal, il n'y a pas de biais sur le capteur et donc le paramètre θ vaut $\theta_0 = 0$. A l'instant $t = 4000$ (pour être le plus proche possible de la vraie panne sur les données), nous simulons une panne de capteur avec $\theta_1 = 0.5$. Nous appliquons l'algorithme de filtrage-détection avec valeur du paramètre caractérisant la panne inconnue (partie 6.3). On suppose que :

$$\theta_1 \in \Theta_1 = [0.4, 1.5].$$

Nous utilisons les deux procédures \hat{t}'_{lim} et \hat{t}'_{app} définies dans le chapitre 6 avec $N = m = 200$ et une fenêtre de taille $M = 200$ pour la première règle. Les noyaux utilisés pour les estimations non paramétriques sont des noyaux gaussiens avec pour paramètre de fenêtre $\delta_N = std(\tilde{x})N^{-0.2}$ pour chaque variable et paramètre.

La surveillance est effectuée entre les pas de temps $t = 2600$ et $t = 4300$. La figure 7.8 représente les données simulées : en gras, les données avec la panne simulée et en clair les données si on ne simule pas de panne.

La figure 7.9 représente les statistiques de test de l'algorithme de filtrage-détection. On voit que les deux algorithmes n'ont pas de mal à détecter la panne de capteur. Chose plus surprenante, les deux statistiques de tests augmentent très fortement juste après l'apparition de la panne puis se mettent à décroître, alors que sous H_1 ces statistiques de test devraient être croissantes en moyenne. C'est la première fois que nous rencontrons, en simulation, ce problème. L'explication vient peut-être du phénomène déjà évoqué lors de l'étude de l'optimalité de l'algorithme de filtrage-détection. La règle de décision est construite de sorte à détecter un changement dans la densité conditionnelle des observations. Elle est d'ailleurs optimale parmi les règles de décision s'attaquant à un changement dans cette densité. Ainsi, pour qu'une panne puisse être détectable, il faut qu'elle entraîne un changement dans la densité conditionnelle des observations. C'est le cas à l'instant d'apparition du biais

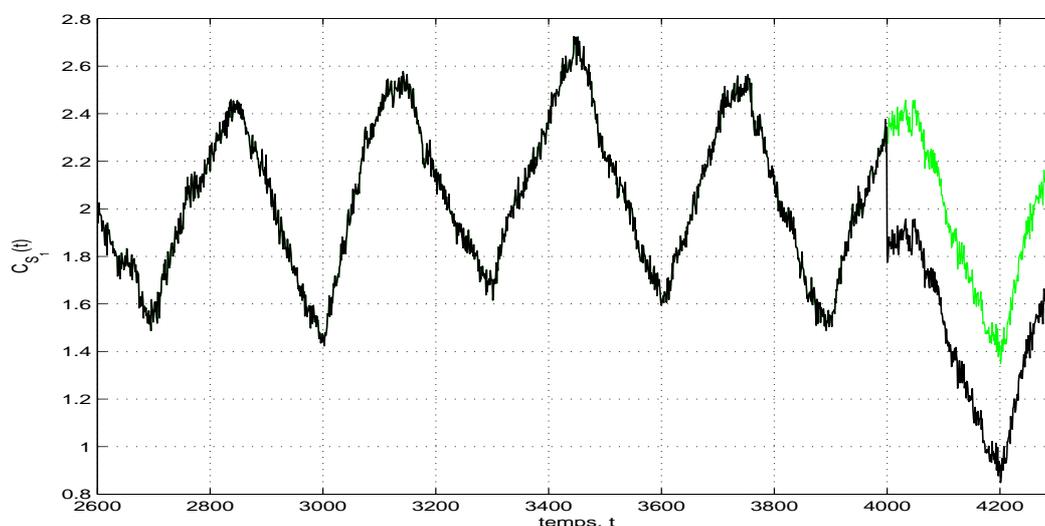


FIG. 7.8 – Evolution de S_1 : en foncé, données simulées avec panne de capteur, en clair, données simulées sans panne

sur le capteur. Mais au bout de quelques pas de temps, ce n'est peut-être plus le cas.

La figure 7.10 correspond à un “zoom” de la figure 7.9. Nous avons représenté la statistiques de test de la règle de décision \hat{t}'_{app} et seuil adaptatif correspondant, au niveau $\alpha = 0.0007$. On note une fausse alarme aux environs du pas de temps 3100. A l'instant 4000, la statistique de test augmente fortement et la panne est détectée au pas de temps 4001.

REMARQUES :

- Dans tout ce chapitre, le seuil adaptatif est calculé à partir de 20000 trajectoires sous H_0 . Si le calcul de ce seuil reste raisonnable en temps de calcul pour la règle \hat{t}'_{app} , ce n'est plus vraiment le cas pour la règle \hat{t}'_{lim} . C'est pourquoi nous ne donnons que le seuil adaptatif relié à la règle \hat{t}'_{app} .
- En toute rigueur, après avoir enregistré une fausse alarme comme c'est le cas dans l'exemple précédent, on doit re-initialiser la règle de décision. Cela n'a pas été fait dans ce chapitre.

Panne graduelle

Nous allons maintenant nous intéresser à l'apparition d'un biais graduel sur le capteur de S_1 . Ce type de panne se rapproche davantage de l'anomalie repérée sur les données réelles. Nous simulons une panne à l'instant $t = 4000$. Sous le régime

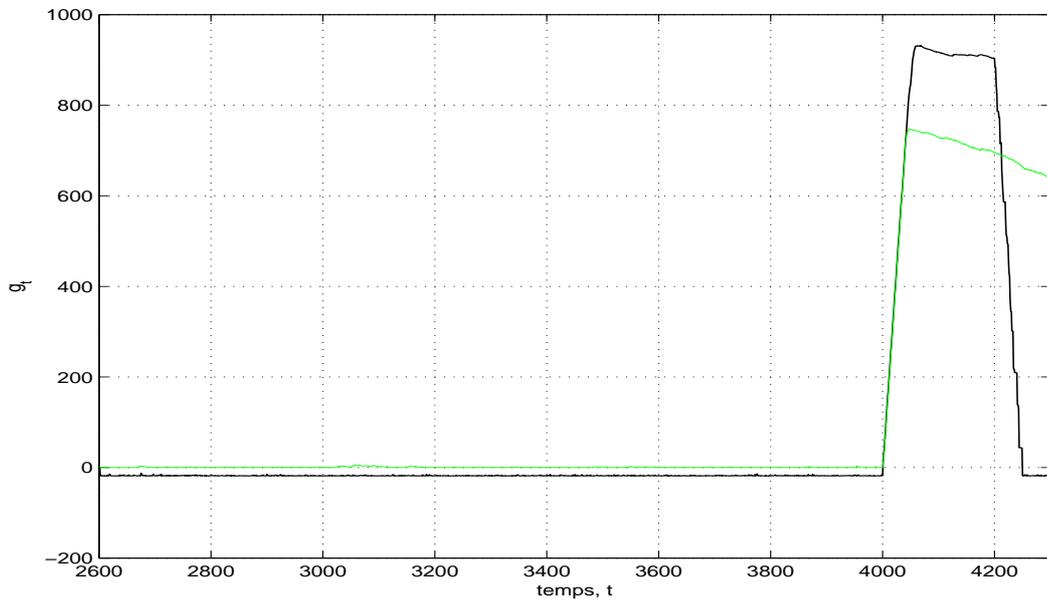


FIG. 7.9 – Statistiques de test des règles de décision \hat{t}'_{lim} (en foncé) et \hat{t}'_{app} en clair

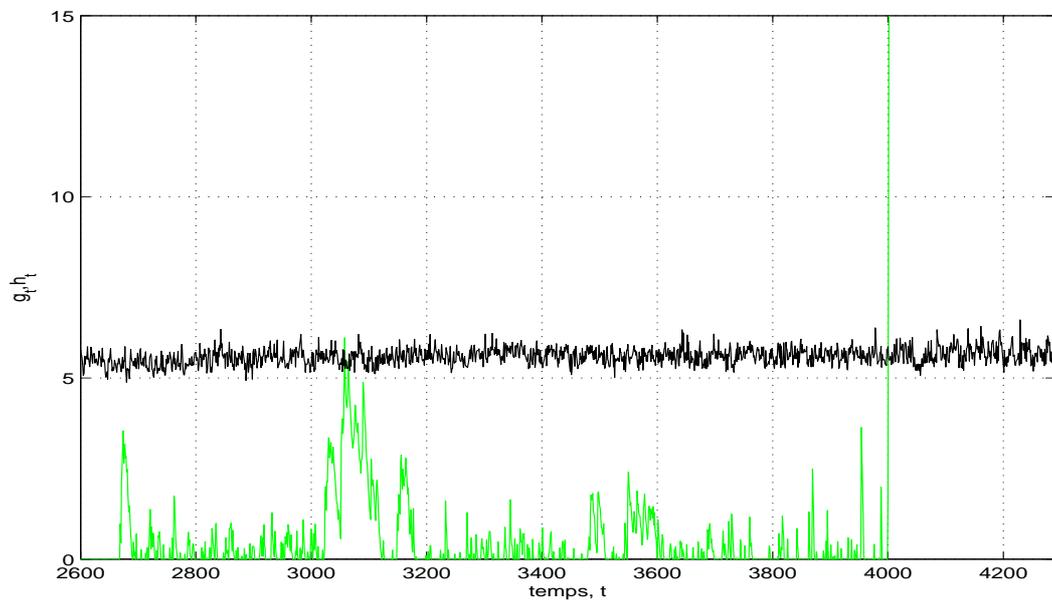


FIG. 7.10 – Statistique de test de la règle de décision \hat{t}'_{app} et seuil adaptatif au niveau $\alpha = 0.0007$.

normal, θ vaut $\theta_0 = 0$, et la panne est simulée de la façon suivante : entre les pas de temps 4000 et 4100, on introduit un biais qui va augmenter de façon linéaire de 0 à 1.2. A partir de $t = 4100$ le biais reste constant et égal à $\theta_1 = 1.2$.

La figure 7.11 représente les données simulées : en gras, les données simulées avec la panne et en clair les données si on ne simule pas de panne.

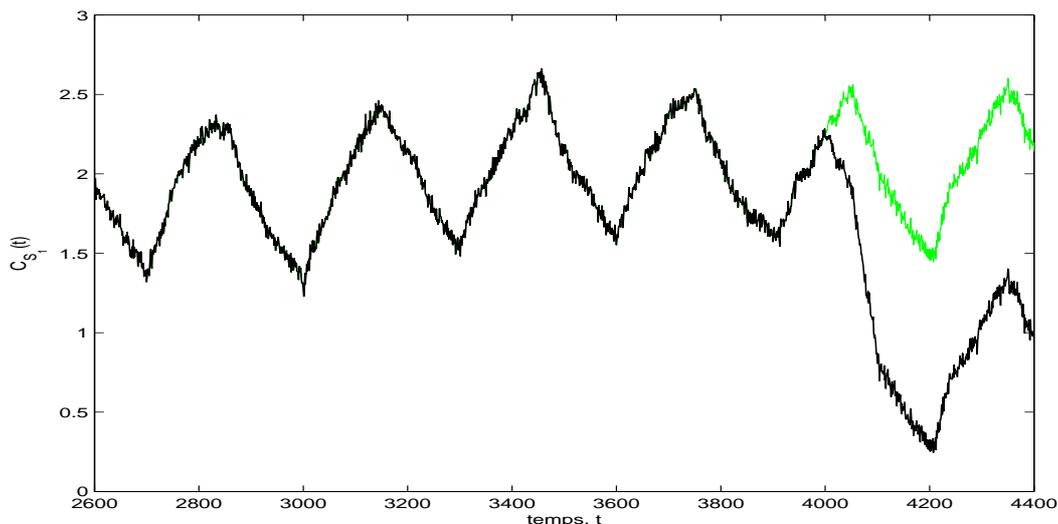


FIG. 7.11 – Evolution de S_1 : en foncé, données simulées avec panne de capteur graduelle, en clair, données simulées sans panne

Parallèlement, nous mettons en place les deux versions des statistiques de test de l’algorithme de filtrage-détection, \hat{t}'_{lim} et \hat{t}'_{app} , avec $N = m = 200$ pour les deux règles et une fenêtre de taille $M = 200$ pour la première règle. Le choix des noyaux et des paramètres est identique à la première simulation. Nous appliquons les algorithmes sous les hypothèses suivantes :

$$\text{sous } H_0, \theta = \theta_0 = 0, \quad \text{et sous } H_1, \theta = \theta_1 \in [0.4, 1.5],$$

mais θ_1 est supposé fixe. Nous allons pouvoir vérifier si les règles de décision sont capables de détecter une panne graduelle malgré l’hypothèse d’un changement abrupt.

La figure 7.12 représente les statistiques de test de l’algorithme de filtrage-détection. Les deux règles de décision détectent la panne, avec un léger retard par rapport à la première simulation. Les résultats obtenus sont donc encourageants puisque les règles de décision sont capables de détecter l’apparition graduel du biais alors que la modélisation de l’anomalie n’est pas bonne puisqu’on considère que la panne est abrupte.

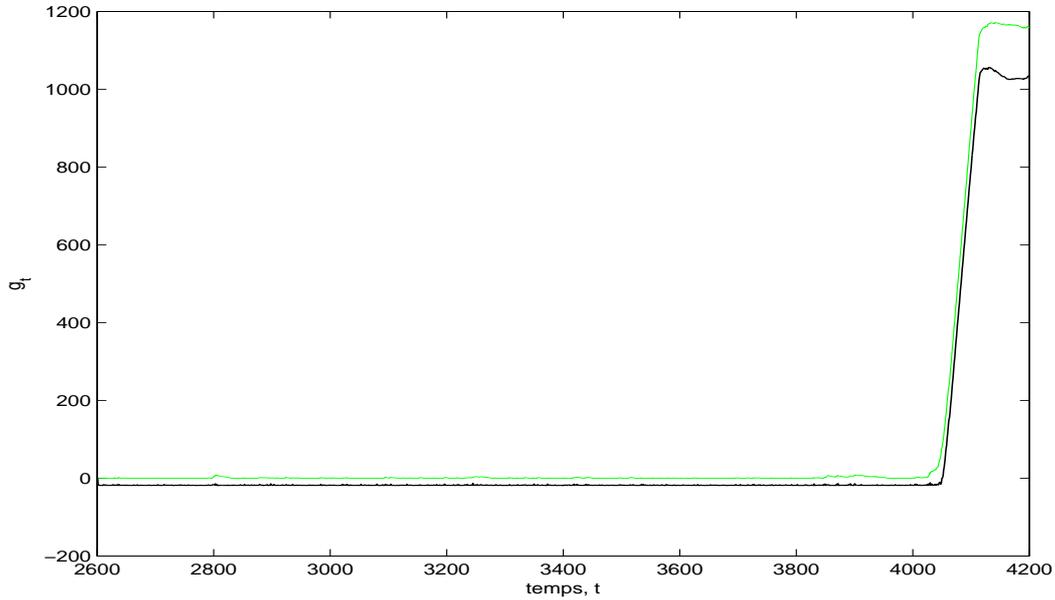


FIG. 7.12 – Statistiques de test des règles de décision \hat{t}'_{lim} (en foncé) et \hat{t}'_{app} en clair pour la panne de capteur graduelle

La figure 7.13 permet de juger le retard à la détection de la règle \hat{t}'_{app} qui est ici de 28 pas de temps (détection au pas de temps $t = 4028$). Ce retard, supérieur à la panne de capteur abrupte est certainement dû au fait que la panne est graduelle et donc, plus difficile à détecter. On peut aussi remarquer deux fausses alarmes avant l'instant de panne.

La figure 7.14 donne l'estimation du biais entre le capteur et S_1 . Cette estimation est assez cohérente.

En simulation, il n'y a donc aucun problème pour détecter ces pannes de capteur.

7.2.3 Applications aux données réelles

Nous allons maintenant utiliser les deux règles de décision sur les données réelles pour mettre en évidence un problème sur le capteur (voir [102]) qui apparaît aux alentours des pas de temps 3950 et 4000, comme on peut le voir sur la figure 7.15, où sont représentées les données-capteur réelles (en clair) et les données-capteur simulées grâce au modèle (7.2), lorsqu'il n'y a pas de panne.

Nous mettons en place les règles de décision \hat{t}'_{lim} et \hat{t}'_{app} avec $N = m = 1000$ et on considère à nouveau une fenêtre de taille $M = 200$ pour \hat{t}'_{lim} . On suppose que

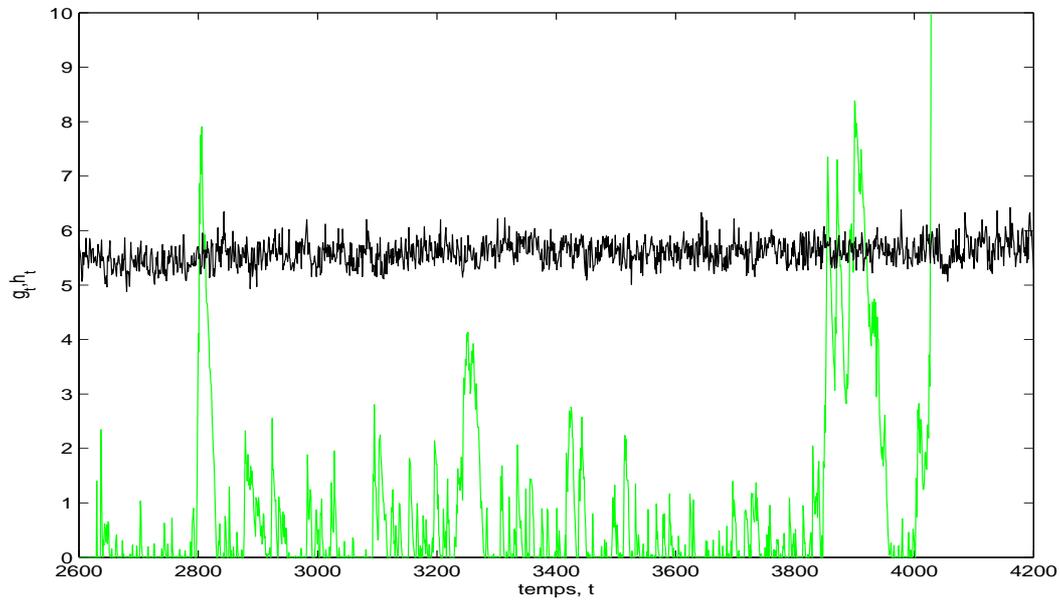


FIG. 7.13 – Statistique de test de la règle de décision \hat{t}'_{app} pour la panne de capteur graduelle et seuil adaptatif.

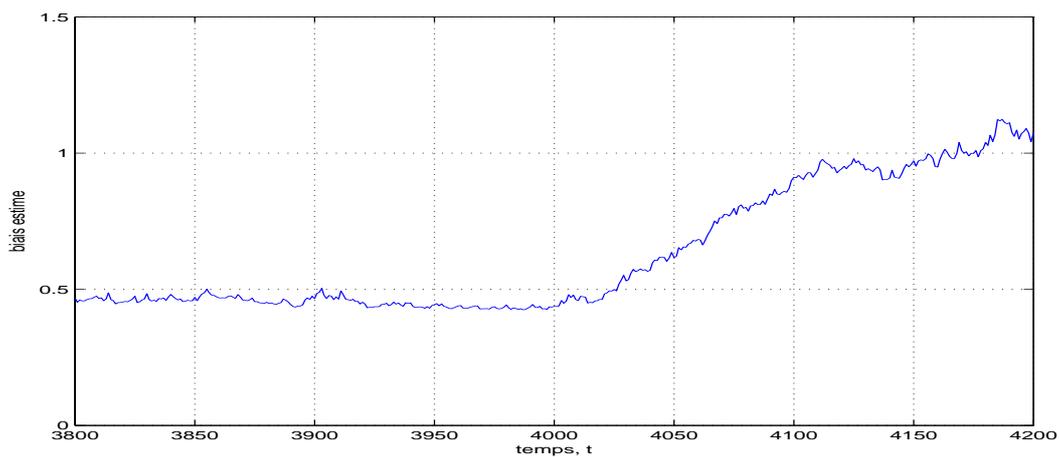


FIG. 7.14 – Estimation du paramètre θ_1 à partir de la règle de décision \hat{t}'_{app}

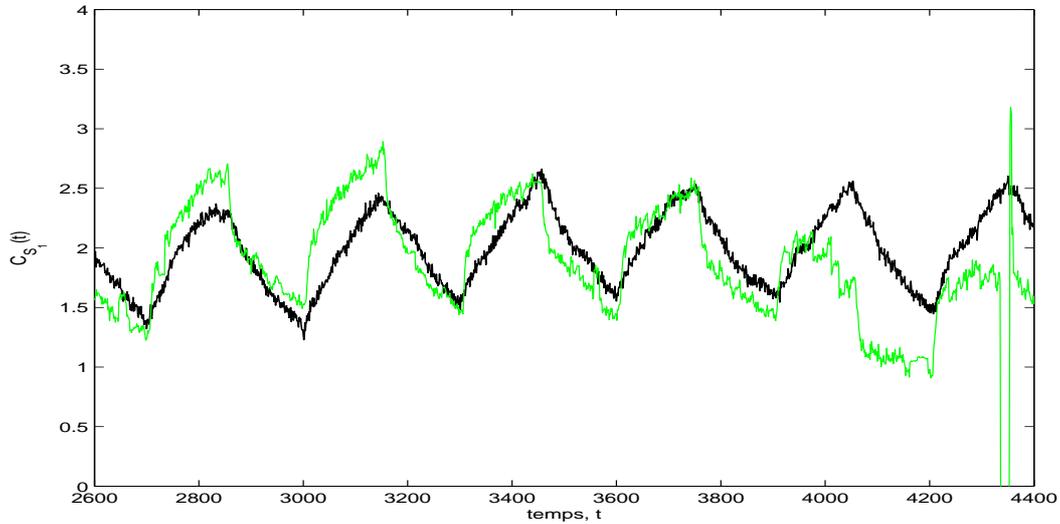


FIG. 7.15 – Evolution du capteur de S_1 : en foncé, données simulées sans panne, en clair, données réelles

sous régime normal : $\theta_0 = 0$ et sous régime de panne $\theta_1 \in [0.6, 1.5]$. Les noyaux et fenêtres utilisés sont identiques aux simulations précédentes.

Les deux statistiques de test sont représentées dans la figure 7.16. On voit tout de suite que les deux statistiques sont perturbées sous l’hypothèse H_0 , contrairement à ce qu’on a observé en simulation où elles étaient négatives ou nulles. Il semble que la règle \hat{t}_{lim} détecte la panne puisqu’elle décolle de façon significative peu après $t = 4000$. Mais son comportement sous H_0 n’est pas satisfaisant. La règle \hat{t}'_{app} , elle, ne fonctionne pas sur cet exemple puisqu’elle est croissante, même sous H_0 .

Comment expliquer le comportement médiocre des deux règles ? Les problèmes rencontrés sur cet exemple proviennent certainement du “mauvais” ajustement du modèle aux données sur certains intervalles alors qu’on est sous régime H_0 . Si on regarde de plus près la comparaison entre les données réelles et celles simulées par le modèle (7.2), on voit, sur la figure 7.17, qu’à partir du pas de temps $t = 3755$, la concentration mesurée diminue fortement pendant une dizaine de pas de temps avant de reprendre une allure plus “classique”. Le modèle (7.2) ne permet pas de suivre cette forte décroissance et la règle de décision détecte alors une panne sur le capteur. Ce type de phénomène se reproduit plusieurs fois sur les pas de temps considérés (entre 2600 et 3900), ce qui peut expliquer la médiocrité des résultats obtenus, notamment pour la règle \hat{t}'_{app} .

Nous avons obtenu l’explication de ces mauvais ajustements du modèle aux

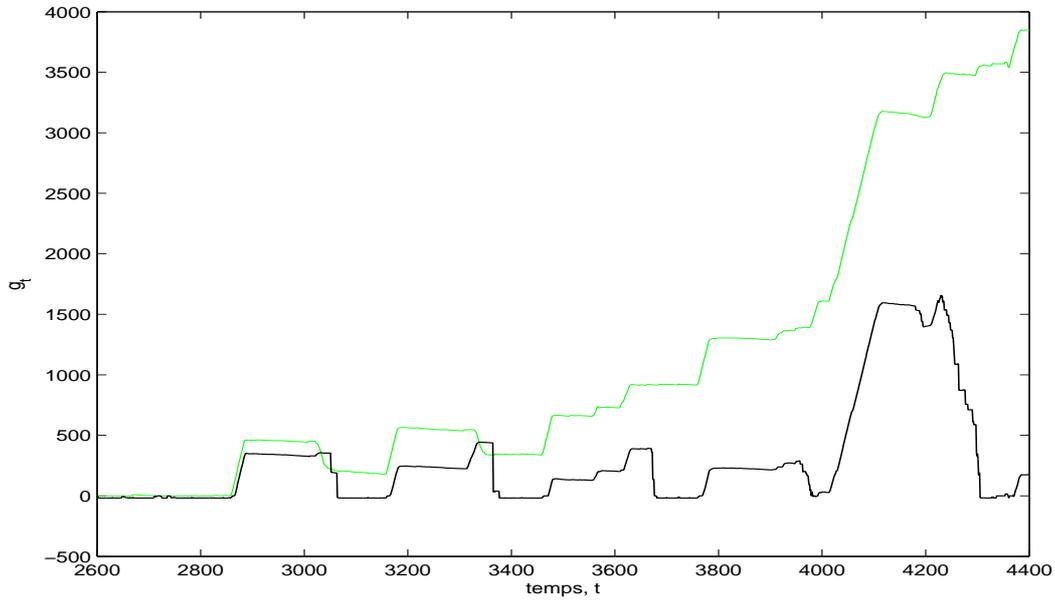


FIG. 7.16 – Statistiques de test des règles de décision \hat{t}_{lim} (en foncé) et \hat{t}_{app} en clair pour les données réelles

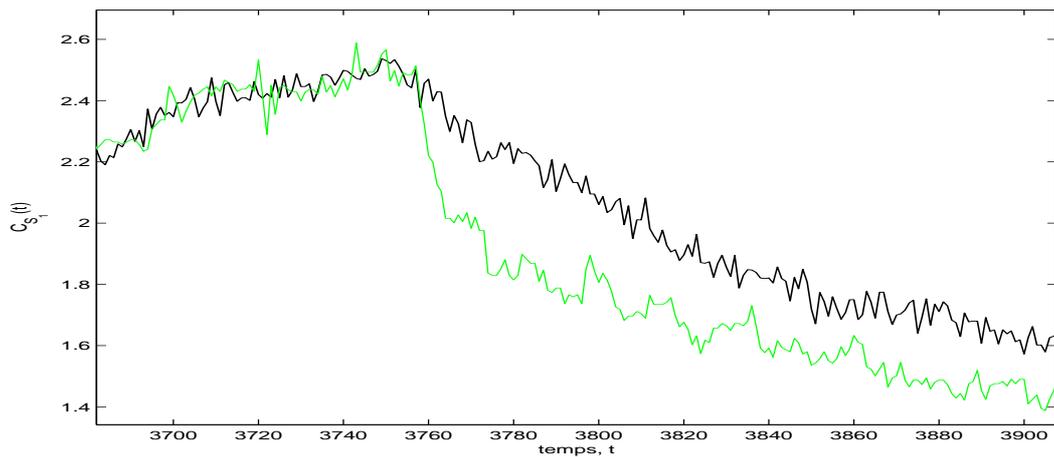


FIG. 7.17 – Evolution du capteur de S_1 sous H_0 : en foncé, données simulées sans panne, en clair, données réelles

données sur ces période d'à peu près dix pas de temps (par exemple entre $t = 3750$ et $t = 3760$ dans la figure 7.17) auprès de Jean-Philippe Steyer du LBE de Narbonne.

Jusqu'au pas de temps $t = 4320$, la pompe de mélange de pied de cuve ne fonctionne pas. Cette pompe permet d'assurer en temps normal un mélange efficace dans le bioréacteur. Ainsi, lorsqu'elle est en panne, chaque variation de Q_{in} (voir la figure 7.1) entraîne des variations de la mesure de la concentration en substrat S_1 car le mélange n'est pas assuré. Après l'instant $t = 4320$, la pompe est remise en marche et on n'observe plus ce phénomène.

Pour revenir à la figure 7.17, au pas de temps $t = 3750$, la variable Q_{in} passe d'une valeur de $30l.h^{-1}$ à $10l.h^{-1}$. Le mélange n'étant pas assuré, le capteur enregistre une grande variation de la concentration en S_1 .

Pour prendre en compte ce problème de pompe de mélange, nous choisissons d'utiliser une variance sur S_1 qui va dépendre du temps. Ainsi, σ_w^2 (voir modèle (7.2)) va être à nouveau égale à 10^{-4} sauf cinq pas de temps après une variation de Q_{in} , où, pendant dix pas de temps, nous utiliserons $\sigma_w^2 = 2.10^{-2}$ (par exemple entre $t = 3755$ et $t = 3765$).

En considérant cette variance, on donne la possibilité au modèle de "capter" les variations importantes de S_1 lors de la variation de Q_{in} .

Nous mettons à nouveau en place les procédures de détection, (figure 7.18) et nous voyons que cette fois-ci, la surveillance s'effectue de façon satisfaisante puisque sous H_0 , les statistiques de test sont largement moins perturbées.

Dans la figure 7.19, nous avons représenté la statistique de test de la règle \hat{t}_{app} et le seuil adaptatif (qui s'adapte par ailleurs aux variations de la variance) et on voit que la détection s'effectue au pas de temps $t = 3933$. La détection a donc lieu, semble t'il, très rapidement.

On voit par ailleurs qu'il y a un certain nombre de fausses alarmes (supérieur au taux fixé sur l'intervalle considéré). On peut l'expliquer par la présence de quelques données aberrantes comme c'est le cas vers les pas de temps $t = 2730$ et $t = 3550$, où le capteur enregistre, dans les deux cas, une dizaine de données "étranges". La statistique de test réagit alors (et même fortement pour $t = 3550$).

7.2.4 Conclusions sur la panne de capteur

Les règles de décision \hat{t}_{lim} et \hat{t}_{app} , semblent être efficaces pour détecter l'apparition d'un biais sur le capteur de la concentration en substrat S_1 . En simulation, les résultats sont bons, même lorsque la panne est graduelle et l'algorithme mis en place pour la détection d'une panne abrupte. Sur les données réelles, les résultats sont encourageants, comme on peut le voir sur la figure 7.18. Il semble, qu'en terme de retard à la détection, les règles de décision soient dans les conditions fixées par le cahier des charges.

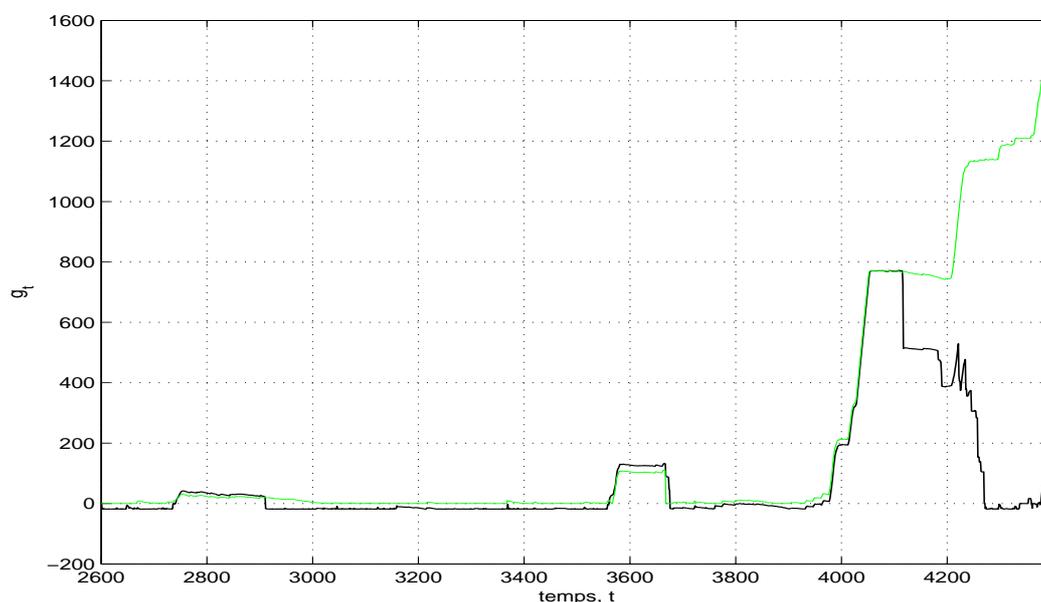


FIG. 7.18 – Statistiques de test des règles de décision \hat{t}'_{lim} (en foncé) et \hat{t}'_{app} en clair pour les données réelles

Généralement, pour traiter l'apparition d'un biais sur un capteur, une méthode simple mais efficace consiste à effectuer la différence entre le modèle et les données. C'est l'objet de la figure 7.20 où on a représenté la différence entre le modèle déterministe du procédé (le modèle (7.2) sans les bruits gaussiens) et les mesures réalisées par le capteur. On voit sur la courbe obtenue qu'il se passe quelque chose d'anormal après le pas de temps $t = 4000$. Il semble, que sur cette figure, le retard à la détection soit supérieur à ce que nous avons obtenu grâce aux algorithmes de filtrage-détection.

De plus, l'allure de la statistique de test \hat{t}'_{app} semble être plus nette dans le sens où il n'y a pas d'ambiguïté possible. A partir de l'instant $t = 3950$, la statistique de test augmente franchement.

7.3 Le changement de modèle sur X_1

Dans cette partie, nous allons nous intéresser au changement de modèle sur la biomasse X_1 . Avant le redémarrage de la pompe de pied de cuve et le changement de volume de travail du bioréacteur (voir figure 7.2), la concentration en biomasse est supposée constante et égale à 1. Ensuite elle évolue selon l'équation différentielle (7.1) (voir figure 7.5).

Nous allons donc chercher à détecter, grâce à l'algorithme de filtrage-détection, ce

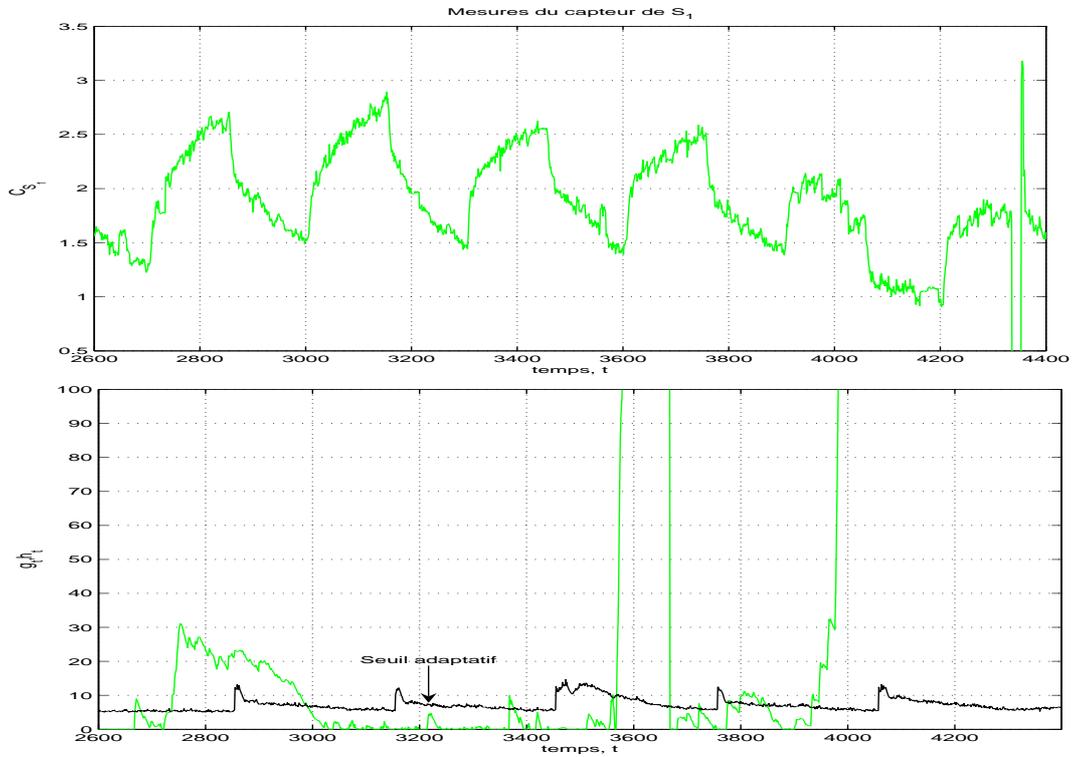


FIG. 7.19 – Statistiques de test des règles de décision \hat{t}_{lim} (en foncé) et \hat{t}_{app} en clair pour les données réelles

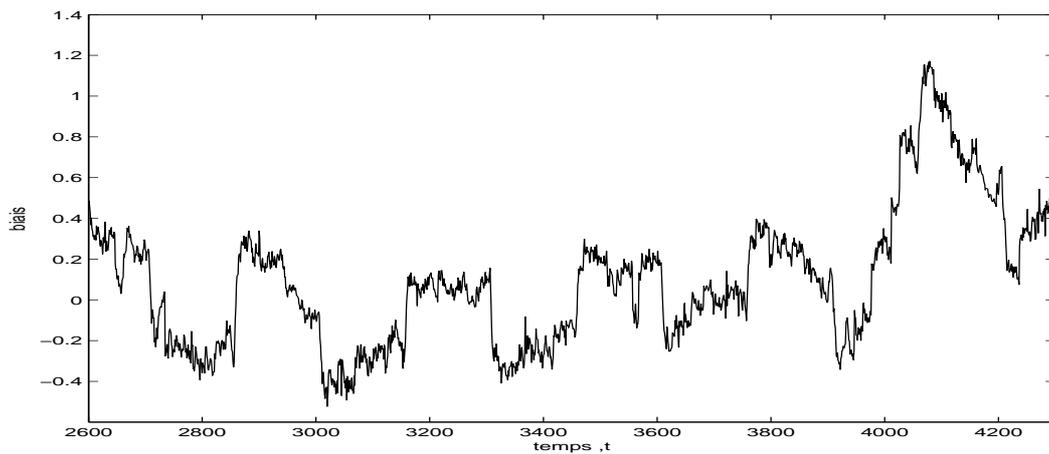


FIG. 7.20 – Valeur du biais entre le modèle et les données réelles

changement de modélisation, d'une part à partir de données simulées et ensuite sur les données réelles. Le régime normal va correspondre à la partie où la concentration en biomasse est constante.

7.3.1 Les deux modèles

Sous le régime de fonctionnement initial, le modèle du système est le suivant :

$$\left\{ \begin{array}{l} \left[\begin{array}{l} X_1(t+1) \\ S_1(t+1) \end{array} \right] = \left[\begin{array}{l} X_1(t) + v_{t+1} \\ S_1(t) + T(D(t)(S_1^{in} - S_1(t)) - k_1\mu_1 X_1(t)) + w_{t+1} \end{array} \right] \\ C_{S_1}(t+1) = \bar{S}_1(t+1) + \epsilon_{t+1}, \end{array} \right. \quad (7.3)$$

où v et ϵ sont des bruits blancs gaussiens indépendants de variances respectives : $\sigma_v^2 = 10^{-6}$ et $\sigma_\epsilon^2 = 10^{-3}$. A nouveau, w sera un bruit blanc gaussien de variance variant au cours du temps. Plus précisément, de $t = 1$ à $t = 4320$, cinq pas de temps après une variation de la variable Q_{in} , on pose $\sigma_w^2 = 0.02$ pendant dix pas de temps. Le reste du temps, $\sigma_w^2 = 10^{-4}$. Les conditions initiales sont les suivantes : $X_1(1) = 1$, $S_1(1) = 1.35$.

Sous régime après changement, le modèle est :

$$\left\{ \begin{array}{l} \left[\begin{array}{l} X_1(t+1) \\ S_1(t+1) \end{array} \right] = \left[\begin{array}{l} X_1(t) + T(\mu_1 - \alpha D(t))X_1(t) + v_{t+1} \\ S_1(t) + T(D(t)(S_1^{in} - S_1(t)) - k_1\mu_1 X_1(t)) + w_{t+1} \end{array} \right] \\ C_{S_1}(t+1) = \bar{S}_1(t+1) + \epsilon_{t+1}, \end{array} \right. \quad (7.4)$$

avec les mêmes valeurs de variance que dans le modèle précédent.

Les données réelles à notre disposition sont donc censées provenir du modèle (7.3) de $t = 1$ à $t = 4319$, et le modèle (7.4) à partir de $t = 4320$.

7.3.2 Détection du changement de régime en simulation

Nous simulons, à l'aide des modèles (7.3) et (7.4) un changement de régime à l'instant $t_p = 4320$. La figure 7.21 représente la trajectoire des données-captteur avec le changement de régime (en foncé) mais aussi, la même trajectoire où le changement de régime n'a pas eu lieu. Après l'instant de changement, on voit que les deux courbes sont très proches, et qu'il faut attendre quelques centaines de pas de temps pour voir une différence significative entre les deux. Le changement à détecter est donc très "fin".

Nous utilisons les deux procédures \hat{t}_{lim} et \hat{t}_{app} définies dans le chapitre 6 avec $N = m = 200$ pour les deux règles et une fenêtre de taille $M = 1000$ pour la première règle. Les noyaux utilisés pour les estimations non paramétriques sont des noyaux

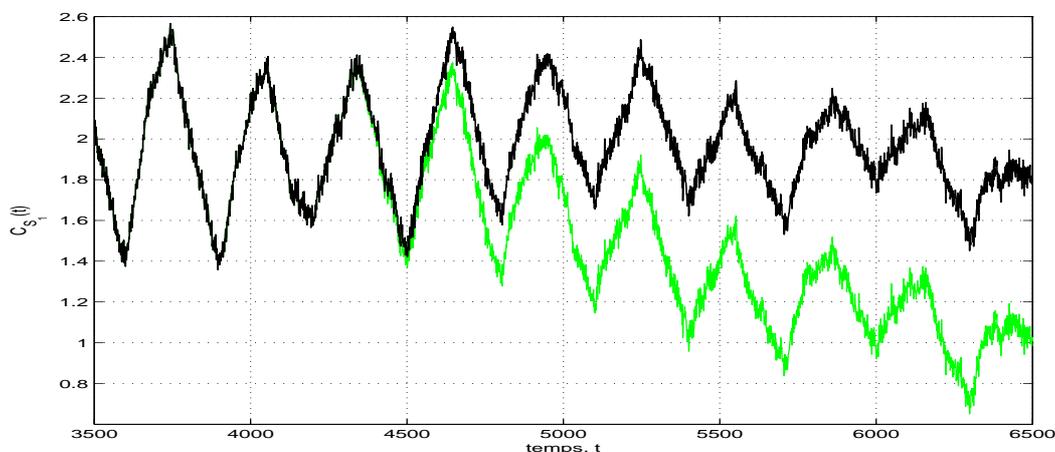


FIG. 7.21 – Evolution du capteur de S_1 : en foncé, données simulées avec changement de modèle, en clair, données simulées sans changement

gaussiens avec pour paramètre de fenêtre $\delta_N = std(\tilde{S})N^{-0.2}$ pour les variables liées au substrat S_1 et $\delta_N = 0.001$ pour les variables liées à la biomasse X_1 .

La figure 7.22 représente les statistiques de test des deux algorithmes. On voit que la détection a bien lieu. De façon plus précise concernant la règle \hat{t}_{app} (voir figure 7.23), il apparaît que la détection s'effectue au pas de temps $t = 4556$, donc à peu près deux cents pas de temps après l'apparition du changement de modèle. On est donc ici à la limite du retard à la détection imposée par le cahier des charges. Cependant, lorsqu'on revient aux données simulées (figure 7.21), on se rend compte qu'au pas de temps où a lieu la détection, la différence entre les deux modèles est très faible. La détection de changement, par l'expérimentateur, à la seule vue des observations, aurait été probablement impossible.

7.3.3 Détection du changement de régime sur les données réelles

Nous traitons maintenant les données réelles. L'anomalie considérée, c'est à dire le changement de modèle sur X_1 , est ici très difficile à détecter pour plusieurs raisons : d'une part, nous avons vu que même en simulation, il n'est pas évident de détecter cette anomalie rapidement. D'autre part, le changement de pompe a lieu au pas de temps $t_p = 4320$, et à cet instant là, on a l'impression, au vu des données (voir figure 7.24), qu'on est toujours sous le régime de panne du biais sur le capteur, étudié dans la première partie. Or les procédures de détection étudiées dans cette thèse nécessitent de ne traiter qu'une panne à la fois. Enfin, les données, entre les pas

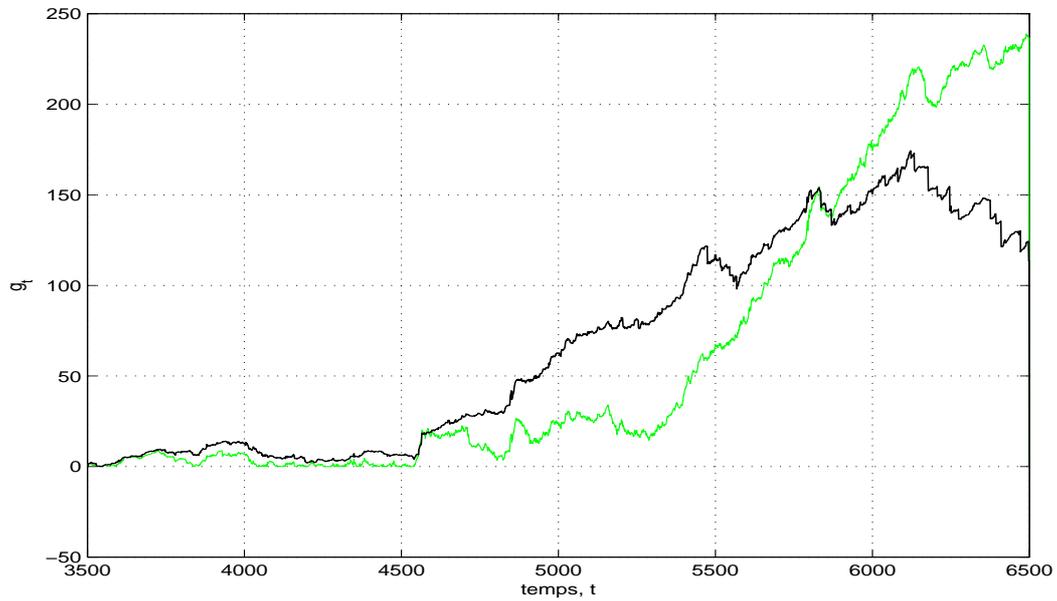


FIG. 7.22 – Statistiques de test des règles de décision \hat{t}_{lim} (en foncé) et \hat{t}_{app} en clair pour le changement de modèle simulé

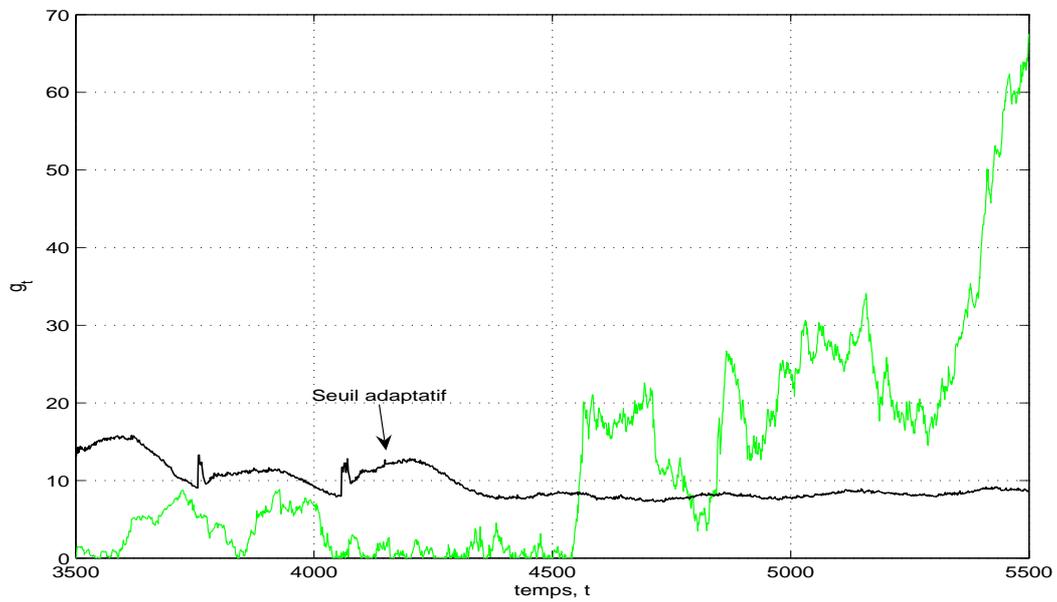


FIG. 7.23 – Statistique de test de la règle de décision \hat{t}_{app} et seuil adaptatif au niveau $\alpha = 0.0007$.

de temps $t = 4000$ et $t = 5000$ sont très perturbées (biais sur le capteur, données aberrantes), notamment entre les pas de temps 4700 et 4900. Lorsqu'on met en place la procédure de détection entre les pas de temps 4000 et 5000, la panne de capteur ainsi que les données aberrantes perturbent très fortement les algorithmes de filtrage-détection.

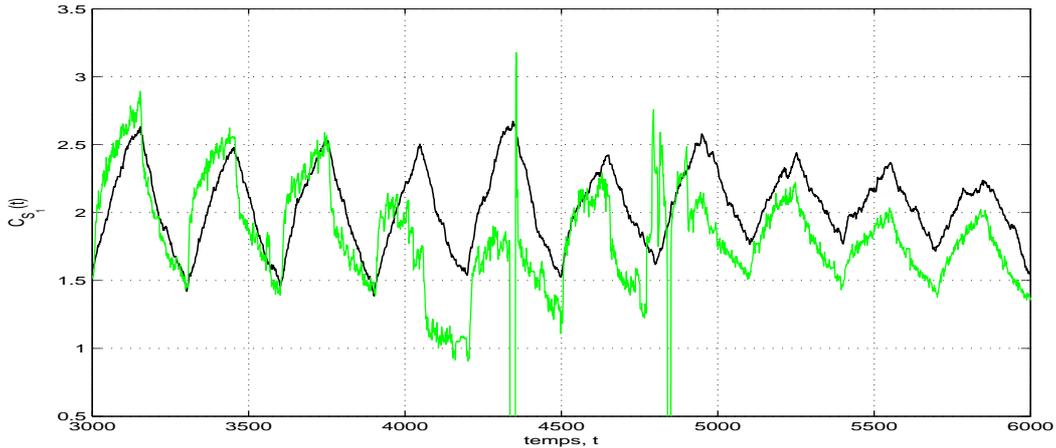


FIG. 7.24 – Evolution du capteur de S_1 : en foncé, données simulées avec changement de modèle sur S_1 , et en clair, données réelles

Nous allons donc aborder le problème de la façon suivante : nous allons tout d'abord mettre en place les procédures de détection entre les pas de temps $t = 2000$ et $t = 3950$, c'est à dire sous H_0 , pour voir comment se comporte les statistiques de tests. Ensuite, nous mettrons en place ces règles de décision entre les pas de temps $t = 5000$ et $t = 7000$ pour voir si la détection a bien lieu.

Nous mettons en place les règles de décision \hat{t}'_{lim} et \hat{t}'_{app} avec $N = m = 1000$ particules et une fenêtre de taille $M = 500$. Les noyaux utilisés sont toujours des noyaux gaussiens et les fenêtres sont les mêmes que dans la simulation précédente.

La figure 7.25 représente, d'une part, le modèle simulé et les données réelles (en clair) et d'autre part, les statistiques de test des deux algorithmes et le seuil adaptatif calculé au niveau $\alpha = 0.0007$ pour la règle \hat{t}'_{app} . Globalement, le résultat obtenu est conforme à ce qu'on attendait. On remarque essentiellement deux fausses alarmes, une au pas de temps $t = 2650$ et une autre au pas de temps $t = 3550$ comme dans le cas de la panne de capteur. Ces données correspondent à des données aberrantes comme on peut le voir sur le premier graphe de la figure 7.25.

La figure 7.26 représente les résultats obtenus sur la période $t = 5000$ à $t = 7000$. A nouveau, les résultats obtenus sont conformes à nos attentes. Les deux statistiques

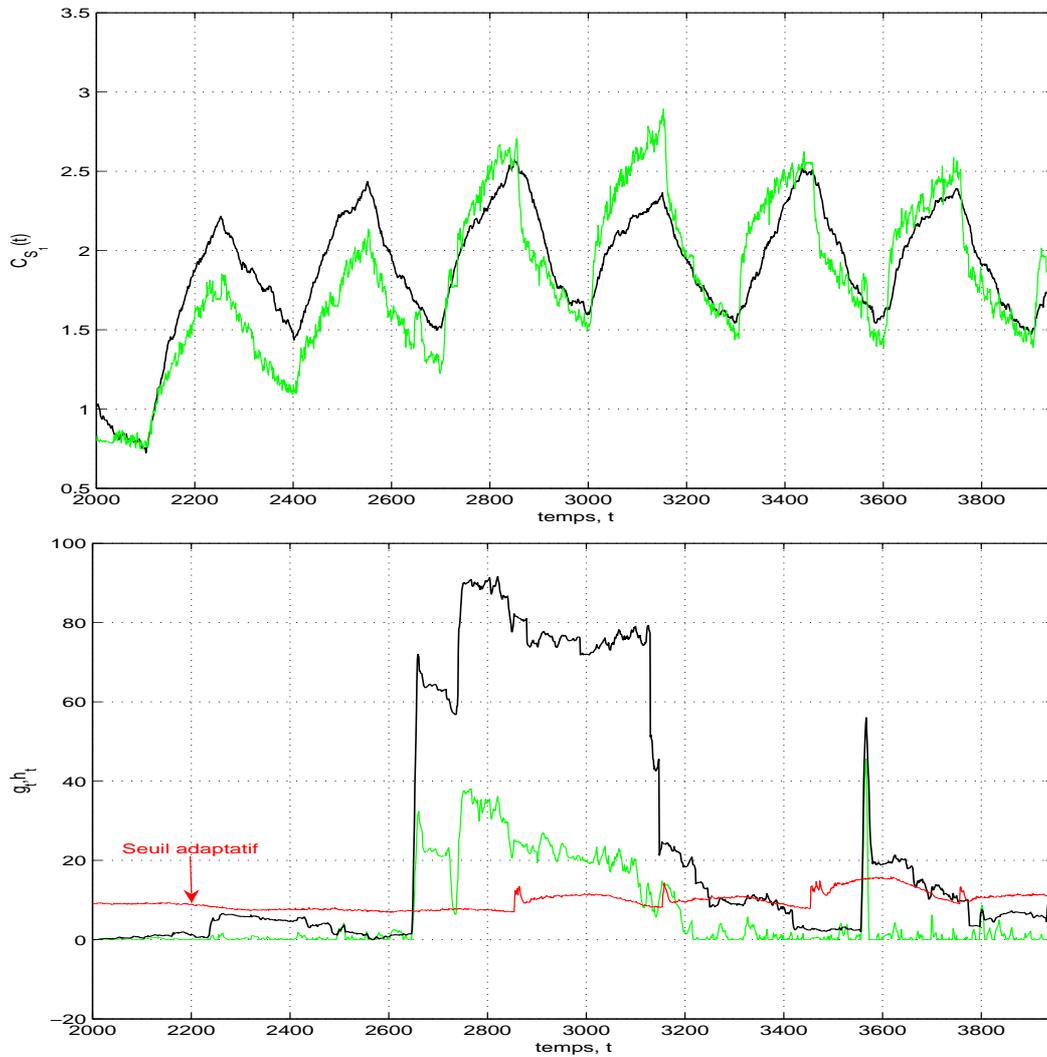


FIG. 7.25 – Evolution du capteur de S_1 : en foncé, données simulées, en clair, les données réelles. Statistiques de test des règles de décision \hat{t}'_{lim} (en foncé) et \hat{t}'_{app} (en clair), pour le changement de modèle sur les données réelles et seuil adaptatif au niveau $\alpha = 0.0007$.

de test augmentent en moyenne et sans ambiguïté. La statistique de test de la règle \hat{t}_{app} franchit le seuil adaptatif à l'instant $t = 5374$. Le retard à la détection est donc assez important ici, mais tout de même inférieur au temps de séjour (qui vaut 900 pas de temps).

REMARQUE : La statistique de test \hat{t}_{lim} n'augmente plus à partir du pas de temps $t = 5600$ hormis au pas de temps 6350 où, comme on peut le voir sur les données (graphe 1 de la figure 7.26), il y a probablement un petit problème sur le capteur. C'est la conséquence de l'utilisation de la fenêtre ($M = 500$). C'est le comportement normal que nous avons évoqué dans le chapitre 6. Sous H_1 , la statistique de test, en moyenne, augmente pendant M pas de temps, puis devient constante.

Sur cet exemple, on voit donc qu'il est donc nécessaire de considérer une fenêtre très large pour mettre en évidence le changement de régime. Une fenêtre de taille $M = 100$, ne permettrait pas de détecter la rupture de modèle.

7.4 Conclusions

Les premiers résultats du fonctionnement de l'algorithme de filtrage-détection sur les deux types d'anomalies considérés sont encourageants. Les seuils adaptatifs calculés à partir d'un grand nombre de simulations du modèle restent relativement cohérents lorsqu'ils sont appliqués sur les données.

Les deux types de panne considérés sont détectés sans trop de problème par les procédures de filtrage-détection. Et la condition fixée par le cahier des charges, retard à la détection inférieur à deux cents pas de temps pour une fausse alarme tous les 1500 pas de temps, semble respectée, pour le biais sur le capteur, et quasiment respectée pour le changement de modèle.

Il est apparu au cours du traitement de ces données, l'importance du choix des fenêtres des estimateurs à noyaux. On pourrait certainement améliorer le comportement des règles de décision en choisissant des noyaux et surtout des tailles de fenêtres plus adaptés. Une étude plus complète sur le choix de ces paramètres est donc nécessaire.

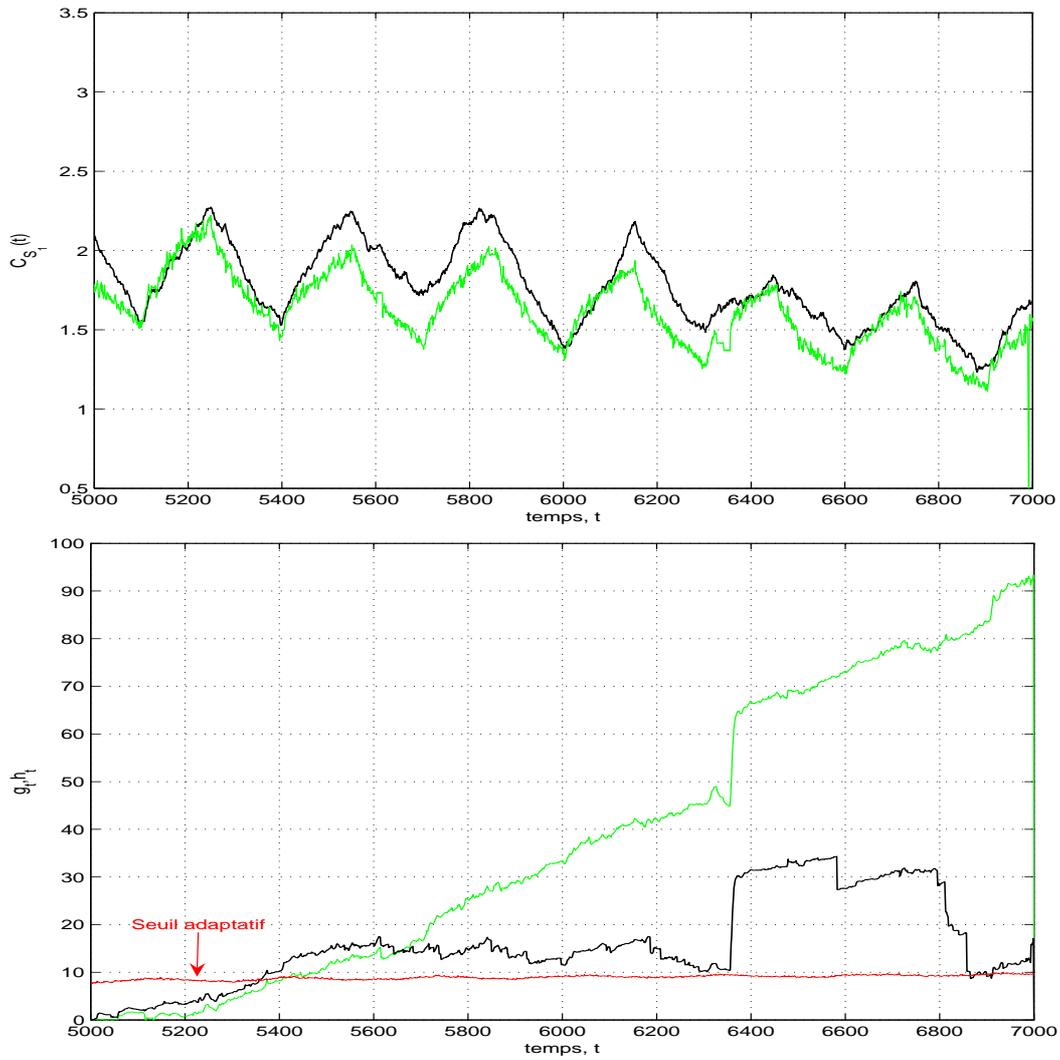


FIG. 7.26 – Evolution du capteur de S_1 : en foncé, données simulées avec panne, en clair, les données réelles. Statistiques de test des règles de décision \hat{t}_{lim} (en foncé) et \hat{t}_{app} (en clair), pour le changement de modèle sur les données réelles et seuil adaptatif au niveau $\alpha = 0.0007$.

Conclusion

Nous avons traité, dans ce travail, le problème de la détection et du diagnostic de panne, ou plutôt, pour être plus précis, le problème de la détection de rupture de modèle. L'objectif de départ était de mettre au point des méthodes statistiques de détection pouvant s'adapter aux particularités des systèmes complexes que sont les procédés de dépollution biologique. L'application de ces méthodes à un bioréacteur de traitement des eaux usées était une des motivations de cette thèse.

Pour répondre aux problèmes rencontrés, notre démarche a été la suivante : nous sommes partis de règles de décision classiques, règle du CUSUM, algorithme de diagnostic de Nikiforov, et nous avons cherché à adapter ces règles, à l'aide d'outils (non paramétriques et de type filtrage) déjà étudiés depuis quelques années dans le laboratoire, aux systèmes nous intéressant.

La première partie de ce travail nous a amené à nous interroger sur le choix du seuil d'une procédure de détection. Pour des procédés non stationnaires, l'utilisation classique d'un seuil fixe nous a paru inapproprié dans le sens où, l'ordre de grandeur de la statistique de test pouvant varier fortement sous régime normal, le taux de fausses alarmes a peu de chance de rester constant.

La méthode que nous avons proposé est basée sur l'estimation de la loi de la statistique de test à partir d'un grand nombre de simulations. Notre méthode entraîne, certes, une grande quantité de calcul, notamment dans le cas du diagnostic, mais elle permet de garder constant le taux de fausses alarmes.

Pour aborder les deux cas développés dans les chapitres 5 (partie inconnue dans le modèle) et 6 (observation indirecte), nous avons introduit la règle du "CUSUM approché". Si la densité conditionnelle des observations est inconnue, le CUSUM ne peut pas être calculé. L'approche considérée ici consiste à estimer les vraisemblances conditionnelles nécessaires à la construction du CUSUM. Nous avons établi l'optimalité de cette procédure sous différents critères classiques en détection de rupture, sous hypothèses de convergence des estimations vers les vraisemblances et sous stabilité du rapport de vraisemblance. Il s'agit là d'un résultat général qui peut être appliqué à de nombreux cas particuliers de modélisation.

Nous l'appliquons tout d'abord pour des modèles dont une partie est inconnue. Cette partie inconnue est estimée à l'aide d'un estimateur à noyau de convolution

Conclusion

construit à partir d'un échantillon d'apprentissage constitué de données en régime normal et en régime de panne. Nous montrons de plus que les hypothèses de stabilité sur le rapport de vraisemblance nécessaires à l'optimalité de l'approche sont vérifiées pour des modèles assez généraux.

Nous nous sommes intéressés par la suite au cas de systèmes indirectement observés. L'estimation des vraisemblances conditionnelles pour construire le CUSUM approché est alors réalisée en combinant une approche de filtrage particulière avec la méthode d'intégration de Monte Carlo. Plusieurs cas sont étudiés : paramètre caractérisant la panne connu, puis paramètre inconnu appartenant à un compact et enfin, nous proposons une règle de décision concurrente du GLR en écrivant un modèle à espace d'état simplifié. Des résultats d'optimalité sont obtenus en utilisant, pour certains d'entre eux, les résultats du chapitre 4. Finalement, le problème du diagnostic dans les modèles à espace d'état est évoqué.

L'application des procédures de "filtrage-détection" sur les données réelles fournies par le LBE de Narbonne a donné des résultats encourageants ce qui laisse à penser que ce type de méthodes pourrait constituer un complément aux approches issues de l'Automatique et de l'IA, généralement utilisées pour la surveillance des bioprocédés.

Les résultats obtenus dans cette thèse ouvrent des perspectives intéressantes pour les méthodes de détection de rupture dans les systèmes complexes.

En premier lieu, il est nécessaire d'effectuer un travail de recherche plus approfondie sur les méthodes de diagnostic de panne. Les résultats obtenus en simulations dans la dernière partie du chapitre 6 sur l'application de l'algorithme de filtrage-détection au cas multi-hypothèses doivent nous encourager d'une part, à étudier l'optimalité de cette procédure et d'autre part, à proposer d'autres règles peut être moins coûteuses en calculs.

De plus, comme nous l'avons mentionné dans l'introduction, la prise en compte d'une variable de contrôle dans le modèle lors de l'étude de l'optimalité du CUSUM approché est à considérer, compte tenu de l'omniprésence de ce type de variable sur tous les procédés industriels.

Au cours de ce travail de thèse, nous avons finalement traité indépendamment les problèmes d'estimation non paramétrique et de filtrage. On pourrait envisager de s'intéresser à la combinaison des deux en tentant de mettre en place une procédure de détection dans un modèle à espace d'état avec partie inconnue.

Troisième partie

Annexes

Annexe A

Les estimateurs à noyaux de convolution

Dans cette annexe, nous présentons rapidement le principe de l'estimation non paramétrique par noyaux de convolution ainsi que quelques résultats généraux. Pour rentrer plus en détails dans ces techniques, on peut consulter les ouvrages de Bosq [11] pour ce qui concerne l'estimation de la densité et de la régression, et Duflo [28] pour l'estimation fonctionnelle dans un modèle autorégressif.

A.1 Principe de l'estimation non paramétrique

Supposons qu'on cherche à estimer une densité de probabilité ψ d'une variable aléatoire X à partir d'un échantillon $(X_i)_{i=1,\dots,n}$ ou une fonction de régression r à partir de n couples de données $(X_i, Y_i)_{i=1,\dots,n}$ où la relation de régression de Y sur X est :

$$Y_i = r(X_i) + \epsilon_i.$$

Il existe alors deux grandes approches pour estimer les fonctions recherchées :

- Dans l'approche paramétrique, on suppose que la fonction cherchée a une forme fonctionnelle pré-spécifiée (par exemple une loi normale pour ψ ou une forme affine pour r) qui dépend d'un nombre fini de paramètres inconnus (moyenne de la normale, ordonnée à l'origine de la fonction r). L'approche consiste alors à estimer ces paramètres à partir des observations.

- Dans l'approche non paramétrique, on ne fait aucune hypothèse spécifique sur la forme de la fonction cherchée. Le principe d'estimation en un point x consiste à utiliser seulement l'information apportée par les observations X_i "proches" de x pour l'estimation de ψ et l'information apportée par les Y_i qui correspondent à des

Annexe A. Les estimateurs à noyaux de convolution

X_i “proches” de x pour l’estimation de r . Ainsi, plus une observation sera proche de x et plus sa contribution à la construction de l’estimateur sera importante.

Les estimateurs à noyaux de convolution font partie de la famille des estimateurs non paramétriques. En tout point d’observation le noyau de convolution va introduire une pondération fonction de la proximité entre l’observation et le point x .

Définition A.1 *Un noyau K est une application de \mathbb{R}^d dans \mathbb{R} , bornée, positive, symétrique, intégrable par rapport à la mesure de Lebesgue, et d’intégrale 1.*

La vitesse de convergence des estimateurs en non paramétrique est plus faible que celle en paramétrique, on doit donc disposer de plus d’information. De plus, cette vitesse est d’autant plus faible que la dimension de l’espace des observations est grande (*fléau de la dimension*).

A.2 L’estimateur de la densité

L’estimateur de la densité est proposé par Rosenblatt [89] en 1956 :

Définition A.2 *Soient X_1, \dots, X_n des variables aléatoires à valeurs dans \mathbb{R}^d , de densité commune ψ et K un noyau sur \mathbb{R}^d . L’estimateur à noyau de la densité ψ associé au noyau K est défini par :*

$$\psi_n(x) = \frac{1}{n\delta_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{\delta_n}\right).$$

La suite (δ_n) , appelée “fenêtre”, est une suite de réels strictement positifs telle que $\delta_n \rightarrow 0$ si $n \rightarrow \infty$. On rencontre souvent la notation suivante : $K_{\delta_n}(x) = K(x/\delta_n)/\delta_n^d$ par soucis de commodité.

Avant de passer à la suite, nous définissons une catégorie particulière de noyaux, les noyaux de Parzen Rosenblatt :

Définition A.3 *Soit K un noyau sur \mathbb{R}^d . On dit que K est un noyau de Parzen-Rosenblatt si :*

$$\lim_{\|x\| \rightarrow +\infty} \|x\|^d K(x) = 0.$$

Théorème A.1 *Soit K un noyau sur \mathbb{R}^d , lipschitzien et de Parzen-Rosenblatt et $\mathcal{C} \subset \mathbb{R}^d$ un compact. On suppose que f est continue sur \mathcal{C} . Alors,*

$$\lim_{n \rightarrow \infty} \delta_n = 0 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} \sup_{x \in \mathcal{C}} |\psi_n(x) - \psi(x)| = 0 \quad \text{p.s.}$$
$$\lim_{n \rightarrow \infty} \frac{n\delta_n^d}{\log n} = \infty$$

D’autres résultats de convergence ont été obtenus notamment par Parzen [77], comme les convergences en moyenne quadratique et en probabilité.

A.3 L'estimateur de la régression

Concernant le cadre de la régression, Nadaraya [71] propose en 1964 comme estimateur de la régression $r(x) = E[Y|X = x]$:

$$r_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{\delta_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{\delta_n}\right)},$$

où $(X_i, Y_i)_{i=1, \dots, n}$ sont des vecteurs aléatoires indépendants et de même loi que (X, Y) . Greblicki et Krzyzak [37] donnent des conditions suffisantes de convergence en probabilité et Collomb [17] énonce un théorème de convergence uniforme presque sûre pour cet estimateur de la régression.

Pour des processus dépendants, comme dans le modèle suivant :

$$X_{n+1} = f(X_n) + \epsilon_{n+1},$$

l'estimation de la fonction f a été l'objet d'une littérature abondante. Pour ces processus, il est alors nécessaire de faire des hypothèses de stationnarité ou de stabilité. On trouve notamment dans Bosq [11], différents types de résultats de convergence, notamment des résultats de convergence sur des compacts dilatés.

Ces dernières années, de nombreux auteurs se sont intéressés au problème de l'estimation non paramétrique dans les modèles contrôlés par une variable U . Ainsi, pour un modèle de la forme :

$$X_{n+1} = f(X_n) + F_n(X_n, U_n) + \epsilon_n,$$

un estimateur non paramétrique récursif s'écrit :

$$f_n(x) = \frac{\sum_{i=1}^{n-1} \delta_i^{-d} K\left(\frac{x-X_i}{\delta_i}\right) (X_{i+1} - F_i(X_i, U_i))}{\sum_{i=1}^{n-1} \delta_i^{-d} K\left(\frac{x-X_i}{\delta_i}\right)}.$$

REMARQUE : Dans le cadre des problèmes de contrôle, les estimateurs récursifs sont souvent privilégiés pour la rapidité de leurs calculs.

Pour les modèles contrôlés, la principale difficulté vient du fait qu'on ne peut plus parler de stabilité mais de stabilisabilité (voir Dufflo [27]). Les premiers résultats de convergence sont dus à Senoussi [95] qui montre la convergence presque sûre sur les compacts de l'estimateur f_n . Plus tard, Portier et Oulidi [84] établissent la convergence presque sûre sur des compacts dilatés, résultat que nous utilisons dans l'étude sur les propriétés du CUSUM non paramétrique (voir Théorème 5.4). On peut enfin citer les travaux de Hilgert, Senoussi et Vila [42] qui généralisent le résultat de Senoussi à une fonction f qui varie dans le temps de manière aléatoire ou déterministe.

Annexe A. Les estimateurs à noyaux de convolution

Annexe B

Méthodes de filtrage

Nous présentons dans cette partie les principes de bases des méthodes de filtrage particulière, et notamment ceux du filtre à noyaux de convolution, en nous inspirant, entre autres, de Rossi et Vila [92] et Campillo [13].

B.1 Modèle à espace d'état et filtre optimal

Définition B.1 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires. Le processus $(X_n)_{n \geq 1}$ est appelé chaîne de Markov si la densité conditionnelle de X vérifie la propriété suivante :

$$p(x_{n+1}|X_{1:n}) = p(x_{n+1}|X_n) \quad \forall n \geq 1. \quad (\text{B.1})$$

Un système à espace d'état est constitué de deux processus. Un processus d'état $(X_n)_{n \geq 0}$, qui vérifie la propriété de Markov et qui a pour loi initiale π_0 . Et d'autre part, un processus d'observation $(Y_n)_{n \geq 1}$. Les deux processus sont reliés par une hypothèse de canal sans mémoire, c'est à dire que les observations Y_n sont mutuellement indépendantes conditionnellement aux variables d'état (X_n) :

$$p(y_n|X_n, Y_{1:n-1}) = p(y_n|X_n).$$

Considérons alors le modèle suivant :

$$\begin{cases} X_n = f(X_{n-1}, v_n) \\ Y_n = g(X_n, w_n) \end{cases}, \quad (\text{B.2})$$

où v_n et w_n sont des bruits blancs, $X_n \in \mathbb{R}^d$, $Y_n \in \mathbb{R}^q$. L'objectif est donc d'estimer X_n en utilisant les observations Y_1, \dots, Y_n . Il s'agit de déterminer la densité conditionnelle $p(x_n|Y_{1:n})$, appelée filtre optimal. Cette quantité se calcule en deux étapes

Annexe B. Méthodes de filtrage

de façon récursive :

$$\begin{array}{ccccc}
 & & & & Y_n \\
 & & & & \downarrow \\
 p(x_{n-1}|Y_{1:n-1}) & \xrightarrow[p(x_n|X_{n-1})]{\text{prédiction}} & p(x_n|Y_{1:n-1}) & \xrightarrow[p(y_n|X_n)]{\text{correction}} & p(x_n|Y_{1:n}) \quad (\text{B.3})
 \end{array}$$

Une première étape de prédiction, où, à partir du filtre optimal de l'instant $n - 1$ et de l'équation d'état du modèle (B.2), on obtient la densité prédictive $p(x_n|Y_{1:n-1})$, et une deuxième étape de correction à partir de l'équation d'observation du modèle (B.2) et de l'observation Y_n .

Les équations fournissant la densité prédictive et le filtre optimal sont les suivantes :

Etape 1 : prédiction

$$p(x_n|Y_{1:n-1}) = \int p(x_n|x_{n-1})p(x_{n-1}|Y_{1:n-1}) dx_{n-1}. \quad (\text{B.4})$$

Etape 2 : correction

$$p(x_n|Y_{1:n}) = \frac{p(Y_n|x_n)p(x_n|Y_{1:n-1})}{\int p(Y_n|x_n)p(x_n|Y_{1:n-1}) dx_n}. \quad (\text{B.5})$$

Dans la pratique, on rencontre deux cas, selon que l'on peut résoudre ou pas les deux équations précédentes.

Dans le cas où le modèle (B.2) est linéaire avec bruits blancs gaussiens additifs, les deux équations (B.4) et (B.5) admettent une solution explicite qui est fournie par le filtre de Kalman [44].

Dans le cas non linéaire, en général ces équations n'admettent pas de solution explicite. On a alors recours à des approximations. Une des premières approches, largement étudiée et largement utilisée dans l'industrie, consiste à linéariser à chaque pas de temps et à appliquer ensuite le filtre de Kalman sur le système linéaire obtenu. La linéarisation s'effectue généralement autour du point correspondant à l'estimation obtenue au pas de temps précédent. On parle alors de filtre de Kalman étendu (voir [33] pour des détails). Dans de nombreux cas, cet algorithme donne de bons résultats. Cependant s'il y a de fortes non linéarités, il est mis en défaut.

Depuis une dizaine d'années, les méthodes de filtrage particulière ont fait leur apparition. Ces techniques, bien que coûteuses en calcul, sont simples à mettre en œuvre et donnent de bons résultats.

B.2 Filtrage particulière

Dans cette partie, nous présentons le principe des filtres de Monte Carlo (ou filtres particuliers). Ces méthodes s'appuient sur un grand nombre de simulations (les particules) qui sont des trajectoires possibles du système dynamique (B.2). Pour une étude détaillée de ces outils, on peut consulter Doucet, de Freitas et Gordon [25]. Dans Andrieu et al. [1], les auteurs appliquent le filtrage particulière à des problèmes de détection de changement, d'identification et de contrôle.

B.2.1 Filtre de Monte Carlo classique

Approximation particulière

On appelle approximation particulière d'une densité p , la mesure empirique suivante :

$$\hat{p}^N(x) := \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{X}(i)}(x),$$

où $\tilde{X}(i) \sim p(x)$ et $\delta_{\tilde{X}(i)}(\cdot)$ est la mesure de Dirac au point $\tilde{X}(i)$. De façon plus générale, on pourra attribuer un poids w_i à chaque particule $\tilde{X}(i)$:

$$\hat{p}^N(x) = \sum_{i=1}^N w_i \delta_{\tilde{X}(i)}(x),$$

avec $w_i \geq 0$ et $\sum_{i=1}^N w_i = 1$.

Le principe des filtres de Monte Carlo va être le suivant : un grand nombre de particules vont évoluer suivant le modèle (B.2). A chaque pas de temps, des poids vont être associés à ces particules en fonction de la vraisemblance de la trajectoire simulée par rapport aux observations Y . On va donc obtenir des mesures empiriques des quantités d'intérêt (filtre optimal, densité prédictive).

L'algorithme du filtre de Monte Carlo pondéré

Initialisation :

- A partir de la loi initiale π_0 de X_0 supposée connue, on génère N particules $(\tilde{X}_0(i))_{i=1, \dots, N}$ indépendantes et identiquement distribuées.

On peut calculer une estimation de π_0 :

$$\hat{\pi}_0 = \sum_{i=1}^N w_0(i) \delta_{\tilde{X}_0(i)} \quad \text{avec} \quad w_0(i) = \frac{1}{N}.$$

Annexe B. Méthodes de filtrage

A l'instant t quelconque :

(i) Evolution des particules suivant le modèle : pour chaque i , on génère $(\tilde{X}_t(i))$ à partir des particules $(\tilde{X}_{t-1}(i))$ et de l'équation d'état du modèle (B.2).

(ii) Correction des poids associés aux particules suivant leurs vraisemblances respectives :

$$w_t(i) = \frac{w_{t-1}(i)p(Y_t|\tilde{X}_t(i))}{\sum_{i=1}^N w_{t-1}(i)p(Y_t|\tilde{X}_t(i))}. \quad (\text{B.6})$$

On obtient ainsi l'approximation du filtre optimal :

$$\hat{p}(x_t|Y_{1:t}) = \sum_{i=1}^N w_t(i)\delta_{\tilde{X}_t(i)}(x_t).$$

REMARQUE : Le poids de chaque trajectoire est ici évalué à partir de la fonction de vraisemblance $p(y_t|X_t)$. D'une manière plus générale on peut évaluer ces poids à partir d'un échantillonnage d'importance en introduisant une loi d'importance. On parle alors de filtre SIS ("sequential importance sampling").

Il existe de nombreux résultats de convergence pour cette méthode de filtrage (voir par exemple Del Moral [20]). Cependant, en pratique, cet algorithme présente un défaut : tous les poids sauf un, deviennent plus ou moins rapidement, nuls, au profit d'une seule trajectoire. On parle de dégénérescence des poids. Pour remédier à ce problème, plusieurs solutions ont été envisagées : d'une part en limitant la mémoire du filtre, c'est à dire en ne prenant plus en compte les poids trop anciens d'une trajectoire. Ainsi les particules qui ont un poids très petit en début de filtrage ne sont plus désavantagées. On parle alors de filtre à mémoire limitée ou encore de filtre à oubli exponentiel.

L'autre solution envisagée et qui donne de très bons résultats consiste à effectuer, à chaque pas de temps, un ré-échantillonnage des particules : on parle alors de filtres avec interaction.

B.2.2 Filtre de Monte Carlo avec ré-échantillonnage

Les filtres particuliers avec interaction ont été introduit simultanément par Del Moral, Rigal et Salut [21], Gordon, Salmond et Smith [36] ("bootstrap filter") et Kitagawa [46] ("Monte Carlo filter"). Le principe est le suivant : à chaque pas de temps, on effectue un ré-échantillonnage des particules. Ainsi, les particules qui ont un poids élevé ont une grande chance d'évoluer tandis que celles qui ont un poids faible disparaissent. Ainsi, au cours du temps, les particules se positionnent dans la partie la plus intéressante de l'espace.

Description de l'algorithme

Initialisation :

- A partir de la loi initiale π_0 de X_0 supposée connue, on génère N particules $(\tilde{X}_0(i))_{i=1,\dots,N}$ indépendantes et identiquement distribuées.

On peut calculer une estimation de π_0 :

$$\hat{\pi}_0 = \sum_{i=1}^N w_0(i) \delta_{\tilde{X}_0(i)} \quad \text{avec} \quad w_0(i) = \frac{1}{N}.$$

A l'instant t quelconque :

(i) Echantillonnage : on génère $\tilde{X}_{t-1}(i) \sim \hat{p}(x_{t-1}|Y_{1:t-1})$, $i = 1, \dots, N$

(ii) Evolution des particules suivant le modèle : pour chaque i , on génère $(\tilde{X}_t(i))$ à partir des particules $(\tilde{X}_{t-1}(i))$ et de l'équation d'état du modèle (B.2).

(iii) Pondération des particules suivant leurs vraisemblances respectives :

$$w_t(i) = \frac{p(Y_t|\tilde{X}_t(i))}{\sum_{i=1}^N p(Y_t|\tilde{X}_t(i))}. \quad (\text{B.7})$$

On obtient ainsi l'approximation du filtre optimal :

$$\hat{p}(x_t|Y_{1:t}) = \sum_{i=1}^N w_t(i) \delta_{\tilde{X}_t(i)}(x_t).$$

Ce filtre se comporte mieux que les filtres de Monte Carlo classiques. Cependant, lorsque les bruits (d'observation et d'état) sur le système sont faibles, il diverge à nouveau à cause de la dégénérescence des particules.

B.3 Filtres à noyaux de convolution (Rossi et Vila [91, 92])

On considère à nouveau dans cette partie le modèle (B.2) et on effectue les hypothèses suivantes :

- π_0 , la distribution initiale de X_0 est connue.
- les fonctions f et g sont connues.
- on peut simuler les bruits v et w .

Annexe B. Méthodes de filtrage

Ces hypothèses sont semblables à celles des filtres de Monte Carlo, sauf la dernière, qui permet ici d'utiliser le filtre à noyaux de convolution dans des cas plus généraux. En effet, on n'a pas besoin, ici, de connaître l'écriture analytique de la densité $p(y_n|X_n)$.

B.3.1 Filtre à noyaux à mémoire complète

Le principe du filtre à noyaux de convolution est le suivant : le filtre optimal va être estimé par une approche non paramétrique (estimateurs à noyaux présentés dans l'annexe A) à partir de trajectoires simulées par le modèle à espace d'état. Ainsi,

$$p(X_t|Y_{1:t}) = \frac{p_{XY}(X_t, Y_{1:t})}{p_Y(Y_{1:t})},$$

où les densités p_{XY} et p_Y seront estimées par la méthode des noyaux de convolution.

Notons $Z_t = (X_t, Y_{1:t})$. L'algorithme de filtrage est le suivant :

- Génération de N états initiaux $\tilde{X}_0(i) \sim \pi_0$, $i = 1, \dots, N$.
- Evolution des N états initiaux $\tilde{X}_0(i)$ suivant le modèle (B.2), t fois successivement et génération des observations correspondantes $\tilde{Y}_{1:t}(i) = \tilde{Y}_1(i), \dots, \tilde{Y}_t(i)$. On obtient ainsi $\tilde{Z}_t(i) = (\tilde{X}_t(i), \tilde{Y}_{1:t}(i))$.
- Les estimations non paramétriques des densités p_{XY} et p_Y sont alors obtenues par :

$$\begin{aligned} \hat{p}_{XY}(Z_t) &= \frac{1}{N} \sum_{i=1}^N K_{\delta_n}(Z_t - \tilde{Z}_t(i)) \\ &= \frac{1}{N} \sum_{i=1}^N K_{\delta_n}^x(X_t - \tilde{X}_t(i)) K_{\delta_n}^y(Y_{1:t} - \tilde{Y}_{1:t}(i)), \end{aligned}$$

pour un noyau K de la forme $K^x.K^y$, et :

$$\hat{p}_Y(Y_{1:t}) = \frac{1}{N} \sum_{i=1}^N K_{\delta_n}(Y_{1:t} - \tilde{Y}_{1:t}(i)).$$

Finalement, l'estimateur du filtre optimal s'écrit :

$$\begin{aligned} \hat{p}(X_t|Y_{1:t}) &= \frac{\hat{p}_{XY}(X_t, Y_{1:t})}{\hat{p}_Y(Y_{1:t})} \\ &= \frac{\sum_{i=1}^N K_{\delta_n}^x(X_t - \tilde{X}_t(i)) K_{\delta_n}^y(Y_{1:t} - \tilde{Y}_{1:t}(i))}{\sum_{i=1}^N K_{\delta_n}(Y_{1:t} - \tilde{Y}_{1:t}(i))}. \end{aligned}$$

Rossi et Vila [92] établissent la convergence L_1 presque sûre de ce filtre :

Théorème B.1 (Rossi et Vila [92]) *Si les noyaux K^x et K^y sont de Parzen-Rosenblatt, positifs et bornés, si p_Y est positive et continue au point $Y_{1:t}$ et si $X_t \mapsto p(X_t, Y_{1:t})$ est continue presque partout, alors*

$$\lim_{n \rightarrow \infty} \delta_n = 0 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} \int |\hat{p}(x_t|Y_{1:t}) - p(x_t|Y_{1:t})| dx_t = 0 \quad \text{p.s.}$$

$$\lim_{n \rightarrow \infty} \frac{n\delta_n^{tq+d}}{\log n} = \infty$$

B.3.2 Filtre par noyaux avec ré-échantillonnage

De la même façon que pour les filtres de Monte Carlo, Rossi et Vila [92] ont étudié un filtre à noyaux avec ré-échantillonnage. Ce ré-échantillonnage est effectué à partir de l'estimation à noyaux du filtre optimal. Là encore, il n'est pas nécessaire de connaître l'écriture analytique de la densité des observations. L'algorithme s'écrit :

- (i) Génération de N états $\bar{X}_{t-1}(i) \sim \hat{p}(x_{t-1}|Y_{1:t-1})$.
- (ii) Evolution des particules : pour tout $i = 1, \dots, N$ on obtient les couples $(\tilde{X}_t(i), \tilde{Y}_t(i))$ en appliquant le système (B.2) à $\bar{X}_{t-1}(i)$.
- (iii) Estimation du filtre optimal :

$$\hat{p}(x_t|Y_{1:t}) = \frac{\sum_{i=1}^N K_{\delta_n}^x(x_t - \tilde{X}_t(i)) K_{\delta_n}^y(Y_t - \tilde{Y}_t(i))}{\sum_{i=1}^N K_{\delta_n}(Y_t - \tilde{Y}_t(i))}.$$

Rossi et Vila [92] établissent la convergence L_1 presque sûre de l'estimateur du filtre optimal $\hat{p}(x_t|Y_{1:t})$. Il s'agit du Théorème 6.1 utilisé dans le chapitre 6 de ce rapport pour montrer l'optimalité de l'algorithme de filtrage-détection.

B.3.3 Estimation de la densité conditionnelle du paramètre θ (Rossi et Vila [91])

Dans le chapitre 6, nous considérons le problème de la détection d'un changement de paramètre dans un modèle à espace à état. Une grande partie de ce chapitre traite le cas où la paramètre θ_1 caractérisant le régime de panne est inconnu. On se retrouve donc confronter à un problème d'estimation de paramètre dans un modèle à espace d'état.

L'approche habituellement utilisée consiste à traiter le paramètre θ_1 comme une variable d'état du système qui évolue selon le modèle :

$$\theta_{t+1} = \theta_t.$$

Annexe B. Méthodes de filtrage

Pour les filtres de Monte Carlo classique, un tel modèle entraîne une dégénérescence des poids. On est alors obligé d'introduire un bruit artificiel :

$$\theta_{t+1} = \theta_t + \epsilon_{t+1},$$

qui n'a aucune justification théorique.

Pour le filtre à noyaux de convolution, il ne sera pas nécessaire d'utiliser ce bruit puisque la diversité des particules va être respectée grâce à l'estimation fonctionnelle (estimation non paramétrique) de la densité $p(\theta_t|Y_{1:t})$.

L'algorithme s'écrit :

(i) Génération de N couple $(\bar{X}_{t-1}(i), \bar{\theta}_{t-1}(i)) \sim \hat{p}(x_{t-1}, \theta_{t-1}|Y_{1:t-1})$.

(ii) Evolution des particules : pour tout $i = 1, \dots, N$ on obtient les particules $(\tilde{X}_t(i), \tilde{\theta}_t(i), \tilde{Y}_t(i))$ en appliquant aux couples $(\bar{X}_{t-1}(i), \bar{\theta}_{t-1}(i))$ le modèle à espace d'état :

$$\left\{ \begin{array}{l} \left[\begin{array}{l} X_t = f(X_{t-1}, \theta_{t-1}, v_t) \\ \theta_t = \theta_{t-1} \end{array} \right] \\ Y_t = g(X_t, \theta_t, w_t) \end{array} \right.$$

(iii) Estimation des densités :

$$\hat{p}(x_t, \theta_t|Y_{1:t}) = \frac{\sum_{i=1}^N K_{\delta_n}^x(x_t - \tilde{X}_t(i)) K_{\delta_n}^\theta(\theta_t - \tilde{\theta}_t(i)) K_{\delta_n}^y(Y_t - \tilde{Y}_t(i))}{\sum_{i=1}^N K_{\delta_n}(Y_t - \tilde{Y}_t(i))}$$

$$\hat{p}(x_t|Y_{1:t}) = \frac{\sum_{i=1}^N K_{\delta_n}^x(x_t - \tilde{X}_t(i)) K_{\delta_n}^y(Y_t - \tilde{Y}_t(i))}{\sum_{i=1}^N K_{\delta_n}(Y_t - \tilde{Y}_t(i))}$$

$$\hat{p}(\theta_t|Y_{1:t}) = \frac{\sum_{i=1}^N K_{\delta_n}^\theta(\theta_t - \tilde{\theta}_t(i)) K_{\delta_n}^y(Y_t - \tilde{Y}_t(i))}{\sum_{i=1}^N K_{\delta_n}(Y_t - \tilde{Y}_t(i))}$$

Annexe C

Les procédés biotechnologiques

Cette annexe a pour but de présenter les grands principes des procédés biotechnologiques. Pour plus d'informations, le lecteur peut se rapporter au livre de Bastin et Dochain [8].

C.1 Généralités

Un procédé biotechnologique est constitué d'un ensemble de réactions biologiques qui ont lieu la plupart du temps dans un bioréacteur. Schématiquement, ce type de processus peut être décrit par la réaction suivante :



On introduit à l'intérieur du bioréacteur des nutriments, les substrats. Pour des procédés de dépollution biologique, ces substrats correspondent à la matière polluante à traiter. La biomasse présente à l'intérieur du bioréacteur est constituée de micro-organismes. Sous des conditions environnementales (pH, température...) satisfaisantes, les biomasses se développent en consommant les substrats, et génèrent des produits.

Ce type de réactions peut être utilisé à des fins assez variées :

- La consommation de substrat : on privilégie la dégradation du substrat. C'est le cadre de la dépollution biologique où les biomasses consomment les nutriments afin de rejeter des eaux propres.

- La croissance des biomasses : on cherche dans ce cas à produire des biomasses, comme c'est le cas dans les procédés qui permettent la production de levure.

Annexe C. Les procédés biotechnologiques

- La production d'un composé principal : c'est notamment le cas dans l'industrie agro-alimentaire (yaourt, alcool) ou encore dans l'industrie pharmaceutique (antibiotiques).

Dans la pratique, on trouve trois modes opératoires qui se différencient par la façon dont le bioréacteur est alimenté en substrat :

- Le mode discontinu ou batch : la totalité du substrat est introduit dès le démarrage de la réaction. Par la suite, aucun prélèvement ni apport n'est effectué. La réaction se termine quand tout le substrat est consommé par la biomasse. L'opérateur ne peut agir au cours de la réaction que sur les conditions environnementales.

- Le mode semi-continu ou fedbatch : contrairement au mode de fonctionnement précédent, les éléments nutritifs (le substrat) sont introduits au fur et à mesure des besoins des micro-organismes. La durée de ces procédés est limitée par la capacité volumique des réacteurs.

-Le mode continu : L'alimentation du réacteur en substrat est faite en continu et un volume égal au volume d'alimentation est soutiré de manière à garder constant le volume à l'intérieur du bioréacteur. Ce mode de fonctionnement est plus riche puisqu'il permet entre autre, d'étudier les caractéristiques d'un micro-organisme sur de longues périodes. Il permet d'autre part de traiter de gros volumes dans des réacteurs de petite taille, ce qui rend ce mode de fonctionnement très utilisé dans l'industrie agro-alimentaire ou pharmaceutique, et dans la dépollution notamment.

La figure C.1 représente un schéma de réaction type avec un seul substrat S et une seule biomasse X . S_{in} est la concentration en substrat dans l'alimentation. Q_{in} représente le débit d'alimentation en substrat et Q_{out} le débit de soutirage. Pour le mode discontinu, Q_{in} et Q_{out} sont tous les deux nuls. Pour le mode semi-continu, seul Q_{out} est nul.

C.2 Modélisation par bilan de matière

Considérons la réaction schématisée dans la figure C.1 où on ne considère qu'une biomasse et qu'un substrat. Le principe de la conservation de la matière étant vérifié, il n'y a ni création ni disparition de matière, les seules variations sont dues à l'alimentation et au soutirage du bioréacteur. Les évolutions de la biomasse et du substrat

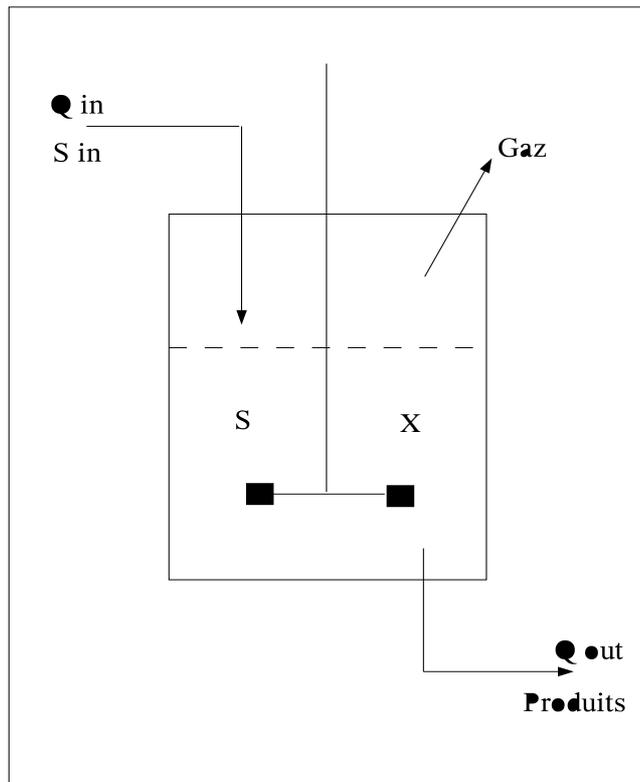


FIG. C.1 – Schéma d'un bioréacteur

Annexe C. Les procédés biotechnologiques

sont caractérisées par les équations différentielles suivantes :

$$\begin{cases} \frac{d(XV)}{dt} = r_X V - Q_{out} X \\ \frac{d(SV)}{dt} = -r_S V + Q_{in} S_{in} - Q_{out} S \\ \frac{d(V)}{dt} = Q_{in} - Q_{out} \end{cases} \quad (C.1)$$

où X et S sont les concentrations en substrat et en biomasse, V est le volume du milieu dans le réacteur et r_X et r_S sont respectivement les vitesses de croissance de la biomasse et de consommation du substrat.

C.3 Les paramètres cinétiques

Une hypothèse de modélisation classique consiste à supposer que la vitesse de consommation du substrat, r_S , est fonction de la vitesse de croissance de la biomasse, r_X . Ainsi, si on note τ le rendement de conversion substrat-biomasse, on a :

$$r_S = \frac{r_X}{\tau}.$$

D'autre part, on utilise généralement le taux de croissance μ de la biomasse plutôt que la vitesse de croissance. Ce taux est défini de la façon suivante :

$$\mu = \frac{r_X}{X}.$$

A partir de ces quantités, on peut, de façon équivalente au système C.1, donner les équations caractérisant l'évolution des concentrations en biomasse et en substrat :

$$\begin{cases} \frac{dX}{dt} = \left(\mu - \frac{Q_{out}}{V}\right) X \\ \frac{dS}{dt} = -\frac{\mu X}{\tau} + \frac{Q_{in}}{V} S_{in} - \frac{Q_{out}}{V} S \\ \frac{d(V)}{dt} = Q_{in} - Q_{out} \end{cases} \quad (C.2)$$

Dans un tel modèle, le choix du taux de croissance μ est très important. Mais sa caractérisation mathématique est très difficile compte tenu des nombreuses grandeurs qui interviennent dans le système. De très nombreux modèles ont été proposés dont la plupart sont recensés dans [8]. Nous évoquons rapidement les deux modélisations les plus utilisées :

- La première est la loi de Monod [69] :

$$\mu = \mu_{max} \frac{S}{S + K_S},$$

où μ_{max} est le taux de croissance maximal et K_S la constante de saturation. La particularité de cette loi est qu'elle est nulle lorsqu'il n'y a plus de substrat à consommer et elle modélise le fait que, plus il y a de substrat et plus la biomasse croît.

- Pour prendre en compte un phénomène d'inhibition de la croissance par la présence de trop fortes concentrations en substrat, on peut considérer la loi de Haldane [38] :

$$\mu = \mu_{max} \frac{S}{S + K_S + \frac{S^2}{K_i}},$$

où K_i est la constante d'inhibition.

C.4 La méthanisation

Le traitement biologique des eaux usées par méthanisation consiste à mettre en contact le substrat avec une population de micro-organismes qui vont donc consommer le polluant.

La méthanisation est une digestion anaérobie qui transforme le substrat essentiellement en boues et en biogaz. La découverte de la méthanisation est dû à A. Volta en 1776 qui observa un rejet de gaz dans un marais, provenant de la décomposition des déchets organiques végétaux.

Deux principaux groupes de bactéries interviennent au cours de la méthanisation. D'une part les acidogènes sont responsables de la transformation des molécules de la matière organique en acides gras volatils tandis que les méthanogènes sont responsables de la dégradation de ces acides gras volatils en biogaz essentiellement composé de méthane, CH_4 (entre 50 et 80%) et de dioxyde de carbone (entre 20 et 50%). On récupère aussi des boues dans une faible quantité (5 à 10 fois moindre qu'en voie aérobie).

Ce procédé possède de nombreux avantages :

- il produit du biogaz utilisable comme source d'énergie, par exemple en alimentant un moteur à gaz qui produit l'électricité.

- il produit de faibles quantités de boues qui restent utilisables comme amendement en agriculture.

- il est intéressant d'un point de vue économique puisqu'il nécessite que peu d'énergie pour son fonctionnement.

Annexe C. Les procédés biotechnologiques

La méthanisation se révèle donc être un procédé très efficace si elle est bien contrôlée. Il faut donc imposer un fonctionnement en temps réel proche d'un état de fonctionnement optimal. La question de la supervision et plus particulièrement de la surveillance est donc centrale puisqu'il est nécessaire de pouvoir gérer, en temps réel, le bioprocédé face à divers problèmes de fonctionnement (pannes de capteurs, d'actionneurs) ou encore de dysfonctionnement du bioprocédé lui-même dans le cas où il dérive de l'état nominal (contamination bactérienne par exemple).

Bibliographie

- [1] C. Andrieu, A. Doucet, S. S. Singh, and V. B. Tadic, “Particle methods for change detection, system identification, and control,” *Proc. IEEE*, vol. 92, pp. 423–438, Mar. 2004.
- [2] B. Azimi-Sadjadi and P. Krishnaprasad, “Change detection for nonlinear systems; a particle filtering approach,” in *Proc. of the American Control Conference*, Anchorage, Alaska, USA, 2002, pp. 4074–4079.
- [3] R. K. Bansal and P. Papantoni-Kazakos, “An algorithm for detecting a change in a stochastic process,” *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 227–235, Mar. 1986.
- [4] M. Basseville, “On-board component fault detection and isolation using the statistical local approach,” *Automatica*, vol. 34(11), pp. 1391–1415, 1998.
- [5] M. Basseville and M.-O. Cordier, “Surveillance et diagnostic de systèmes dynamiques : approches complémentaires du traitement du signal et de l’intelligence artificielle,” INRIA, Tech. Rep., Apr. 1996.
- [6] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes. Theory and Application*. Prentice-Hall, 1993.
- [7] —, “Fault isolation for diagnosis : nuisance rejection and multiple hypothesis testing,” *Annual reviews in control*, vol. 26(2), pp. 189–202, 2002.
- [8] G. Bastin and D. Dochain, *On-line Estimation and Adaptive Control of Bioreactors*. Elsevier, 1990.
- [9] R. H. Berk, “Some asymptotic aspects of sequential analysis,” *Annals of Statistics*, vol. 1, pp. 1126–1138, 1973.
- [10] O. Bernard, Z. Hadj-Sadok, D. Dochain, A. Genovesi, and J.-P. Steyer, “Dynamical model development and parameter identification for an anaerobic wastewater treatment process,” *Biotechnology and Bioengineering*, vol. 75, 4, pp. 424–438, 2001.
- [11] D. Bosq, Ed., *Non parametric Statistics for Stochastic Processes*, ser. Lecture Notes in Statistics. New York, USA : Springer, 1996, vol. 110.
- [12] B. E. Brodsky and B. S. Darkhovsky, *Non-Parametric Statistical Diagnosis, Problems and Methods*. Kluwer Academic Publishers, 2000.

Bibliographie

- [13] F. Campillo, “Filtrage particulaire et modèles de Markov cachés,” Cours Master recherche 2ème année, Université du Sud Toulon-Var, 2006.
- [14] J.-P. Cassar and C. Guernez, “Sensitivity considerations on structured polynomial residuals,” in *European Control Conference, ECC’99*, Karlsruhe, Germany, 1999.
- [15] J. Chen and R. J. Patton, *Robust model-based fault diagnosis for dynamic systems*. Kluwer Academic Publishers, Amsterdam, 1999.
- [16] R. N. Clark, D. C. Fosth, and V. M. Walton, “Detecting instrument malfunctions in control systems,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-11, pp. 465–473, 1975.
- [17] G. Collomb, “Conditions nécessaires et suffisantes de convergence uniforme d’un estimateur de la régression, estimation des dérivées de la régression,” *Comptes Rendus de l’Académie des Sciences de Paris*, vol. 288, pp. 161–164, 1979.
- [18] F. Cérou and F. Le Gland, “Efficient particle methods for residual generation in partially observed sde’s,” in *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia, 2000, pp. 1200–1205.
- [19] P. Del Moral, “Nonlinear filtering using random particles,” *Theory of Probability and its Application*, vol. 40, pp. 859–872, 1995.
- [20] —, “Nonlinear filtering :interacting particle solution,” *Markov Processes and Related Fields*, vol. 2(4), pp. 555–580, 1996.
- [21] P. Del Moral, G. Rigal, and G. Salut, “Estimation et commande optimale non linéaire : un cadre unifié pour la résolution particulaire,” Rapport technique LAAS/CNRS, contrat DRET-DIGILOG, Toulouse, Mar. 1992.
- [22] L. Devroye, *A course in density estimation*. Birkhauser, Boston, 1987.
- [23] S. X. Ding, P. Zhang, and P. M. Frank, “Application of probabilistic robustness technique to the fault detection system design,” in *Proceedings of the 42th IEEE Conference on Decision and Control*, Maui, Hawaii, USA, 2003, pp. 972–977.
- [24] E. A. Domlan, D. Maquin, and J. Ragot, “Diagnostic des systèmes à commutation, approche par la méthode de l’espace de parité,” in *Conférence Internationale Francophone d’Automatique, CIFA’2004*, Douz, Tunisia, 2004.
- [25] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science, Springer, 2001.
- [26] V. P. Dragalin, “The sequential change point problem,” *Economic Quality Control*, vol. 12, pp. 95–122, 1997.
- [27] M. Duflo, *Méthodes récursives aléatoires*. Masson, 1990.

-
- [28] ———, *Random Iterative Models*. Springer-Verlag, 1997.
- [29] M. Fouladirad and I. Nikiforov, “Optimal statistical fault detection with nuisance parameters,” *Automatica*, vol. 41, pp. 1157–1171, 2005.
- [30] P. M. Frank, “Fault diagnosis in dynamic systems using analytical and knowledge based redundancy : A survey and some new result,” *Automatica*, vol. 26 (3), pp. 459–474, 1990.
- [31] ———, “Analytical and qualitative model-based fault diagnosis - a survey and some new results,” *European Journal of Control*, vol. 2, pp. 6–28, 1996.
- [32] C.-D. Fuh, “Optimal change point detection in state space models,” Institute of statistical science, Academia sinica, Tech. Rep., Mar. 2006.
- [33] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA : MIT Press, 1974.
- [34] J. J. Gertler, *Fault detection and diagnosis in engineering systems*. Marcel Dekker Inc., 1998.
- [35] M. A. Girshick and H. Rubin, “A Bayes approach to a quality control model,” *Annals of Mathematical Statistics*, vol. 23, pp. 114–125, 1952.
- [36] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, “Novel approach to non-linear/non Gaussian Bayesian state estimation,” *IEE Proceedings-F*, vol. 140, pp. 107–113, 1993.
- [37] W. Greblicki and A. Krzyzak, “Asymptotic properties of kernel estimates of a regression fonction,” *Journal of Statistical Planning and Inference*, vol. 4, pp. 81–90, 1980.
- [38] J. B. S. Haldane, *Enzymes*. Longmans, 1930.
- [39] D. M. Hawkins, P. Qiu, and C. W. Kang, “The changepoint model for statistical process control,” *Journal of Quality Technology*, 35, vol. 35, pp. 355–365, 2003.
- [40] N. Hilgert, “Identification et contrôle de processus autorégressifs non linéaires incertains : application à des procédés biotechnologiques,” Ph.D. dissertation, Université Paris-Sud, Paris, France, 1997.
- [41] N. Hilgert, J. Harmand, J. P. Steyer, and J. P. Vila, “Nonparametric identification and adaptive control of an anaerobic fluidized bed digester,” *Control Engineering Practice*, vol. 8, pp. 367–376, 2000.
- [42] N. Hilgert, R. Senoussi, and J.-P. Vila, “Estimation non-paramétrique de suites de fonctions de processus autorégressifs non linéaires,” *Comptes Rendus de l’Academie des Sciences de Paris, Série I Math.*, vol. 323, pp. 1085–1090, 1996.
- [43] R. Isermann, “Supervision, fault-detection and fault-diagnosis methods - an introduction,” *Control Engineering Practice*, vol. 5, pp. 639–652, 1997.

Bibliographie

- [44] R. E. Kalman, "A new approach to linear filtering and prediction theory," *Transaction of the ASME, Journal of Basic Engineering*, vol. 82, pp. 34–45, 1960.
- [45] M. Kinnaert, D. Vrancic, E. Denolin, D. Juricic, and J. Petrovcic, "Model-based fault detection and isolation for a gas-liquid separation unit," *Control Engineering Practice*, vol. 8, pp. 1273–1283, 2000.
- [46] G. Kitagawa, "Monte carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, pp. 1–25, 1996.
- [47] A. M. Krieger, M. Pollak, and B. Yakir, "Surveillance of a simple linear regression," *Journal of the American Statistical Association*, vol. 98, pp. 456–469, June 2003.
- [48] T. L. Lai, "Asymptotic optimality of invariant sequential probability ratio tests," *Annals of Statistics*, vol. 9, pp. 318–333, 1981.
- [49] —, "Sequential changepoint detection in quality control and dynamical systems," *Journal of the Royal Statistical Society : Series B*, vol. 57, pp. 613–658, 1995.
- [50] —, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2917–2929, Nov. 1998.
- [51] —, "Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems," *IEEE Trans. Inform. Theory*, vol. 46, pp. 595–608, Mar. 2000.
- [52] —, "Sequential analysis : Some classical problems and new challenges," *Statistica Sinica*, vol. 11, pp. 303–408, 2001.
- [53] T. L. Lai and J. Z. Shan, "Efficient recursive algorithms for detection of abrupt changes in signals and systems," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 952–966, May 1999.
- [54] L. Lardon, "Représentation et gestion des incertitudes pour le diagnostic par la théorie de Dempster-Shafer : application aux procédés biologiques," Ph.D. dissertation, ENSAM, Montpellier, France, Dec. 2004.
- [55] L. Lardon, A. Punal, and J.-P. Steyer, "On-line diagnosis and uncertainty management using evidence theory - experimental illustration to anaerobic digestion processes," *Journal of Process Control*, vol. 14, pp. 747–763, 2004.
- [56] M. Lavielle, "Detection of multiple changes in a sequence of random variable," *Stochastic Processes and their Applications*, vol. 83, pp. 79–102, Sept. 1999.
- [57] J. Li, C. W. Chan, and H.-Y. Zhang, "Asymptotic local approach in fault detection based on predictive filters," *Journal of Guidance, Control, and Dynamics*, vol. 28, pp. 1112–1122, 2005.

-
- [58] P. Li and V. Kadiramanathan, “Particle filtering based likelihood ratio approach to fault diagnosis in nonlinear stochastic systems,” *IEEE Trans. Syst., Man, Cybern. C*, vol. 31, pp. 337–343, Aug. 2001.
- [59] —, “Fault detection and isolation in non-linear stochastic systems - a combined adaptive monte carlo filtering and likelihood approach,” *International Journal of Control*, vol. 77, pp. 1101–1114, 2004.
- [60] J. Liu and M. West, “Combined parameter and state estimation in simulation-based filtering,” in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. New York : Springer-Verlag, 2001, pp. 197–223.
- [61] G. Lorden, “On excess over the boundary,” *Annals of Mathematical Statistics*, vol. 41, pp. 520–527, 1970.
- [62] —, “Procedures for reacting to a change in distribution,” *Annals of Mathematical Statistics*, vol. 42, pp. 1897–1908, 1971.
- [63] —, “Open-ended tests for Koopman-Darmois families,” *Annals of Statistics*, vol. 1, pp. 633–643, 1973.
- [64] G. Lorden and M. Pollak, “Nonanticipating estimation applied to sequential analysis and changepoint detection,” *Annals of Statistics*, vol. 33, pp. 1422–1454, 2005.
- [65] D. Maquin and J. Ragot, *Diagnostic des systèmes linéaires*. Collection Pédagogique d’Automatique, Hermès Science Publications, 2000.
- [66] Y. Mei, “Asymptotically optimal methods for sequential change-point detection,” Ph.D. dissertation, California Institute of Technology, Pasadena, California, May 2003.
- [67] —, “Sequential change-point detection when unknown parameters are present in the pre-change distribution,” *Annals of Statistics*, vol. 34, pp. 92–122, 2006.
- [68] L. A. Mironovski, “Functional diagnosis of linear dynamic systems,” *Automation and Remote Control*, vol. 40, pp. 1198–1205, 1979.
- [69] J. Monod, *Recherches sur la Croissance des Cultures Bactériennes*. Hermann & Cie, 1942.
- [70] G. Moustakides, “Optimal procedures for detecting change in distribution,” *Annals of Statistics*, vol. 14, pp. 1379–1387, 1986.
- [71] E. A. Nadaraya, “On estimating regression,” *Theory of Probability and its Applications*, vol. 9, pp. 141–142, 1964.
- [72] I. Nikiforov, “A generalized change detection problem,” *IEEE Trans. Inform. Theory*, vol. 41, pp. 171–187, Jan. 1995.

Bibliographie

- [73] ———, “A simple recursive algorithm for diagnosis of abrupt changes in random signals,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 2740–2746, 2000.
- [74] ———, “A lower bound for the detection/isolation delay in a class of sequential tests,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 3037–3046, 2003.
- [75] T. Oskiper and H. Poor, “Online activity detection in a multiuser environment using the matrix CUSUM algorithm,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 477–493, 2002.
- [76] E. S. Page, “Continuous inspection schemes,” *Biometrika*, vol. 41, pp. 100–115, 1954.
- [77] E. Parzen, “On estimation of a probability density function and mode,” *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.
- [78] M. Pawlak, E. Rafajlowicz, and A. Steland, “On detecting jumps in time series : nonparametric setting,” *Journal of Nonparametric Statistics*, vol. 16, pp. 329–347, 2004.
- [79] D. Picard, “Testing and estimating change points in time series,” *Advances in Applied Probability*, vol. 17, pp. 841–867, Dec. 1985.
- [80] M. Pollak, “Optimal detection of a change in distribution,” *Annals of Statistics*, vol. 13, pp. 206–227, 1985.
- [81] ———, “Average run lengths of an optimal method of detecting a change in distribution,” *Annals of Statistics*, vol. 15, pp. 749–779, 1987.
- [82] M. Pollak, A. M. Krieger, and B. Yakir, “Detecting a change in regression : first-order optimality,” *Annals of Statistics*, vol. 27, pp. 1896–1913, 1999.
- [83] M. Pollak and D. Siegmund, “Approximations to the expected sample size of certain sequential tests,” *Annals of Statistics*, vol. 3, pp. 1267–1282, 1975.
- [84] B. Portier and A. Oulidi, “Nonparametric estimation and adaptive control of functional autoregressive models,” *SIAM Journal on Control and Optimization*, vol. 39, pp. 411–432, 2000.
- [85] J. Ragot and D. Maquin, “Fault measurement detection in an urban water supply network,” *Journal of Process Control*, vol. 16, pp. 887–902, 2006.
- [86] Y. Ritov, “Decision theoretic optimality of the CUSUM procedure,” *Annals of Statistics*, vol. 18, pp. 1464–1469, 1990.
- [87] S. W. Roberts, “Control charts based on geometric moving averages,” *Technometrics*, vol. 1, pp. 239–250, 1959.
- [88] ———, “A comparison of some control chart procedures,” *Technometrics*, vol. 8, pp. 411–430, 1966.
- [89] M. Rosenblatt, “Remarks on some nonparametric estimates of a density function,” *Annals of Mathematical Statistics*, vol. 27, pp. 832–837, 1956.

-
- [90] V. Rossi, “Filtrage non linéaire par noyaux de convolution : Application à un procédé de dépollution biologique,” Ph.D. dissertation, ENSAM, Montpellier, France, Dec. 2004.
- [91] V. Rossi and J.-P. Vila, “Approche non paramétrique du filtrage de système non linéaire à temps discret et à paramètres inconnus,” *C. R. Acad. Sci. Paris, Ser. I*, vol. 340, pp. 759–764, 2005.
- [92] —, “Nonlinear filtering in discrete time : a particle convolution approach,” *An. Inst. Stat. Univ. Paris*, vol. 3, pp. 71–102, 2006.
- [93] S. Schreier, J. Ragot, R. Patton, and P. M. Frank, “Observer design for a class of non-linear systems,” in *IFAC Symposium on Fault Detection, Supervision and Safety for Technical processes, SAFEPROCESS’97*, Hull, Great Britain, 1997.
- [94] L. Schwartz, “On Bayes procedures,” *Probability Theory and Related Fields*, vol. 4-1, pp. 10–26, Mar. 1965.
- [95] R. Senoussi, “Lois du logarithme itéré et identification,” Ph.D. dissertation, Université Paris-Sud, Paris, France, 1991.
- [96] W. A. Shewhart, *Economic Control of Quality Manufactured Product*. Macmillan, New York, 1931.
- [97] A. N. Shiryaev, “On optimum methods in quickest detection problems,” *Theory of Probability and its Applications*, vol. 8, pp. 22–46, 1963.
- [98] —, *Optimal Stopping Rules*. Springer-Verlag, New-York, 1978.
- [99] D. Siegmund and E. S. Venkatraman, “Using the generalized likelihood ratio statistic for sequential detection of a change-point,” *Annals of Statistics*, vol. 23, pp. 255–271, 1995.
- [100] M. Staroswiecki and G. Comtet-Varga, “Fault detectability and isolability in algebraic dynamic system,” in *European Control Conference, ECC’99*, Karlsruhe, Germany, 1999.
- [101] A. Steland, “Nonparametric monitoring of financial time series by jump-preserving estimators,” *Statistical Papers*, vol. 43, pp. 361–377, 2002.
- [102] J.-P. Steyer and O. Bernard, “An exemple 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.
- [103] J.-P. Steyer, A. Genovesi, and J. Harmand, “Outils d’aide au diagnostic et détection de pannes,” in *Automatique des bioprocédés*, D. Dochain, Ed. Paris : Hermès, 2001, ch. 8, pp. 215–244.

Bibliographie

- [104] M. Stoto, R. D. Fricker, A. Jain, A. Diamond, J. O. Davies-Cole, C. Glymph, G. Kidane, G. Lum, L. Jones, K. Dehan, and C. Yuan, “Evaluating statistical methods for syndromic surveillance,” in *Statistical Methods in Counter-Terrorism : Game Theory, Modeling, Syndromic Surveillance and Biometric Authentication*, D. Olwell, A. G. Wilson, and G. Wilson, Eds. New-York : Springer, 2006, ch. III, pp. 141–172.
- [105] Z. G. Stoumbos, M. R. Reynolds, T. P. Ryan, and W. H. Woodall, “The state of statistical process control as we proceed into the 21st century,” *Journal of the American Statistical Association*, vol. 95, pp. 992–998, Sept. 2000.
- [106] A. G. Tartakovsky, “Asymptotic performance of a multichart CUSUM test under false alarm probability constraint,” in *Proc. of the 44th IEEE Conference on Decision and Control and the European Control Conference*, Seville, Spain, 2005, pp. 320–325.
- [107] —, “Asymptotic optimality in bayesian change-point detection problems under global false alarm probability constraint,” *Theory of Probability and Its Applications (submitted)*, 2006.
- [108] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blazek, and H. Kim, “Detection of intrusions in information systems by sequential change-point methods,” *Statistical Methodology*, vol. 3, pp. 252–293, July 2006.
- [109] A. G. Tartakovsky and V. V. Veeravalli, “General asymptotic bayesian theory of quickest change detection,” *Theory of Probability and its Application*, vol. 49, pp. 538–582, 2005.
- [110] N. Vaswani, “Additive change detection in nonlinear systems with unknown change parameters,” *IEEE Trans. Signal Processing*, vol. 55, pp. 859–872, Mar. 2007.
- [111] V. Wagner, “Identification non paramétrique et contrôle prédictif neuronal de processus non linéaires incertains de type biotechnologique : application à un procédé de dépollution biologique,” Ph.D. dissertation, ENSAM, Montpellier, France, Sept. 2001.
- [112] A. S. Willsky and H. L. Jones, “A generalized likelihood ratio approach to detection and estimation of jumps in linear systems,” *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 108–112, Feb. 1976.
- [113] Y. Wu, Ed., *Inference for change point and post change means after a CUSUM test*, ser. Lecture Notes in Statistics. New York, USA : Springer, 2005, vol. 180.
- [114] B. Yakir, “Optimal detection of a change in distribution when the observations form a Markov chain with a finite state space,” in *Change-Point Problems*, E. Carlstein, H. Muller, and D. Siegmund, Eds. Edition Hayward, CA : Inst. Math. Statist., 1994, pp. 346–358.

- [115] L. Zadeh, “Fuzzy sets,” *Information and control*, vol. 8, pp. 338–353, 1965.
- [116] Q. Zhang, M. Basseville, and A. Benveniste, “Fault detection and isolation in nonlinear dynamic systems : a combined input-output and local approach,” *Automatica*, vol. 34, pp. 1359–1373, 1998.
- [117] Q. Zhang, F. Campillo, F. Cérou, and F. Le Gland, “Nonlinear system fault detection and isolation based on bootstrap particle filters,” in *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005*, Seville, Spain, 2005, pp. 3821–3826.