

Université de Montréal

**CONTRÔLE STATISTIQUE DES PROCÉDÉS  
MULTIVARIÉS**

par

**François Martel**

Département de mathématiques et de statistique  
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de Maître ès sciences (M.Sc.)  
en Statistique

Juillet 1998

© MARTEL, MCMXCVIII



QA

3

U54

1998

V.023



Université de Montréal

Bibliothèque



Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**CONTRÔLE STATISTIQUE DES PROCÉDÉS  
MULTIVARIÉS**

présenté par

**François Martel**

a été évalué par un jury composé des personnes suivantes :

*Robert Cléroux*

---

(président-rapporteur)

*Christian Léger*

---

(directeur de recherche)

*Gina Bezeau*

---

(co-directeur)

*François Perron*

---

(membre du jury)

Mémoire accepté le :

98.10.13

## SOMMAIRE

---

Dans un procédé industriel, il est fort probable que des données sur plusieurs caractéristiques de qualité soient prises simultanément. La relation entre les différentes variables est mesurée, entre autre, par la matrice de variance-covariance du procédé. Ainsi, des cartes de contrôle univariées pour chacune des variables ne sont pas appropriées puisqu'elles ne tiennent pas compte de la corrélation. Le but de ce mémoire est de présenter les principales procédures multivariées pour le contrôle statistique des procédés pour détecter un changement dans le vecteur des moyennes. Aussi, pour être en mesure d'interpréter un signal, on doit être capable d'identifier les variables responsables. On présentera donc également les principales procédures pour la détection des variables hors-contrôle.

## REMERCIEMENTS

---

Je tiens d'abord à remercier mon directeur de recherche Christian Léger pour sa très grande disponibilité, son sens critique aiguisé, ses conseils judicieux et pour avoir partagé avec moi ses connaissances statistiques.

Je tiens également à remercier de façon particulière la compagnie Stone-Consolidated pour leur appui financier mais je suis surtout reconnaissant envers Gina Bezeau et Martin Carignan pour leurs conseils, leur appui et pour avoir partagé avec moi leur expérience pertinente quant à la pratique du CSP. Sans eux, ce projet n'aurait pas vu le jour. Merci à vous de l'audace dont vous avez fait preuve en vous impliquant dans un tel projet.

Merci au Conseil de Recherche en Sciences Naturelles et en Génie du Canada pour le soutien financier et pour m'avoir donné la possibilité de réaliser ce projet dans le cadre d'une bourse à incidence industrielle.

Je ne saurais oublier ma mère et mon père pour leur soutien financier mais surtout pour les encouragements et la confiance qu'ils m'ont témoignés tout au long de ces enrichissantes années d'études. Merci à vous deux.

Finalement, je veux remercier de façon spéciale ma conjointe Mijanou pour la confiance et la patience dont elle a fait preuve et pour m'avoir soutenu de façon remarquable dans ce retour aux études.

Plus personnellement, je veux lui dire un MERCI du fond du coeur pour le cadeau merveilleux et inestimable qu'elle m'a fait au cours de cette dernière année d'étude, notre petit amour Félix.

# TABLE DES MATIÈRES

---

SOMMAIRE	ii
REMERCIEMENTS	iii
LISTE DES FIGURES	x
LISTE DES TABLEAUX	xiii
INTRODUCTION	1
CHAPITRE 1. INTRODUCTION AU CONTRÔLE STATISTI- QUE DES PROCÉDÉS	4
1.1. ORIGINE ET PHILOSOPHIES DU CSP	5
1.1.1. Historique du CSP	5
1.1.2. Le concept de variation	6
1.1.3. Les quatre états d'un procédé	8
1.2. LES OUTILS DU CSP	9
1.2.1. Les sept outils principaux	9

	vi
1.2.2. La carte de contrôle	9
1.3. LA CARTE DE SHEWHART	12
1.3.1. La carte $\bar{X}R$	12
1.3.2. La carte $\bar{X}Rm$	19
1.3.3. <i>LMS</i> de la carte de Shewhart	22
1.4. LA CARTE CUSUM	24
1.4.1. Principes de base	25
1.4.2. <i>LMS</i> de la procédure cusum	29
1.4.3. Choix des paramètres	30
1.4.4. Autres caractéristiques	31
1.5. LA CARTE EWMA	31
1.5.1. Principes de base	32
1.5.2. <i>LMS</i> de la procédure ewma	35
1.5.3. Choix des paramètres	35
1.5.4. Autres caractéristiques	36
1.6. COMPARAISON DES TROIS CARTES	37



	vii
1.7. AUTRES CARACTÉRISTIQUES	44
1.7.1. Données non normales	44
1.7.2. Données auto-corrélées	44
1.7.3. Données multivariées	45
<b>CHAPITRE 2.  CONTRÔLE STATISTIQUE DES PROCÉDÉS MULTIVARIÉS</b>	<b>46</b>
2.1. INTRODUCTION	46
2.2. LA CARTE $T^2$	47
2.2.1. Modèle théorique	48
2.2.2. Application à la carte $T^2$	49
2.2.3. Données individuelles	59
2.2.4. Estimation de $\Sigma$ à partir des étendues mobiles	61
2.2.5. Résumé des statistiques $T^2$ et des limites de contrôle	66
2.2.6. <i>LMS</i> de la carte $T^2$	68
2.3. LA CARTE CUSUM MULTIVARIÉE	69
2.3.1. La carte cusum multivariée de Crosier	69
2.3.2. Paramètres et <i>LMS</i> de la carte cusum multivariée	75
2.3.3. Autres procédures cusum multivariés	75

	viii
2.4. LA CARTE EWMA MULTIVARIÉE	77
2.4.1. Procédure pour la carte ewma multivariée	77
2.4.2. Paramètres et <i>LMS</i> de la carte ewma multivariée	80
2.4.3. Autres procédures ewma multivariées	80
2.5. COMPARAISONS DES TROIS CARTES	81
2.6. AUTRES CARTES ET TECHNIQUES EXISTANTES	85
2.7. DÉTECTION DES VARIABLES HORS-CONTRÔLE	87
2.7.1. Utilisation de $p$ procédures univariées	88
2.7.2. L'analyse en composantes principales	90
2.7.3. Décomposition de la statistique $T^2$	96
2.7.4. Choix d'une procédure pour détecter les variables hors-contrôle	99
2.8. DONNÉES AUTO-CORRÉLÉES	106
<b>CHAPITRE 3. APPLICATION À UN JEU DE DONNÉES</b>	<b>109</b>
3.1. PRÉSENTATION DES VARIABLES ET DU JEU DE DONNÉES	109
3.2. ANALYSE DU JEU DE DONNÉES	110
3.3. APPLICATION DE LA CARTE MULTIVARIÉE $T^2$	122
3.3.1. Application de la carte aux quatre variables	123

3.3.2. Application de la carte aux deux premières composantes principales	125
3.4. DÉTECTION DES VARIABLES HORS-CONTRÔLE	127
3.4.1. Construction des cartes univariées	127
3.4.2. Application de la méthode d'acp de Kourti et MacGregor	131
3.5. CONCLUSION	138
CONCLUSION	142
RÉFÉRENCES	145

## LISTE DES FIGURES

---

1.1	Carte de contrôle typique.	11
1.2	Carte de contrôle $\bar{X}R$ pour 20 observations de la phase un ( $n=5$ et $L=3$ ).	18
1.3	Carte de contrôle de Shewhart pour 10 observations en contrôle ( $\delta=0$ ) avec $L=3,09$ (phase deux et $n=1$ ).	21
1.4	Carte de contrôle cusum pour 10 observations en contrôle ( $\delta=0$ ) avec $h_{cusum}=1,72$ et $k=1,5$ (phase deux et $n=1$ ).	28
1.5	Carte de contrôle ewma pour 10 observations en contrôle ( $\delta=0$ ) avec $r=0,75$ et $h_{ewma}=3,087$ (phase deux et $n=1$ ).	34
1.6	Cartes de contrôle de Shewhart, cusum et ewma pour 10 observations ayant subi un déplacement de $\delta=2$ (phase deux et $n=1$ ).	42
1.7	Cartes de contrôle de Shewhart, cusum et ewma pour 10 observations ayant subi un déplacement de $\delta=3$ (phase deux et $n=1$ ).	43
2.8	Carte de contrôle $T^2$ des 100 échantillons de la phase un pour $n > 1$ .	54
2.9	Cartes de contrôle multivariée et univariées des 10 observations de la phase deux pour $n > 1$ .	56

2.10 Régions de contrôle multivariée et univariée pour $n > 1$ .	57
2.11 Carte de contrôle $T^2$ pour 10 observations de la phase deux ( $n=1$ ) en utilisant les étendues mobiles pour l'estimation de $\Sigma$ .	65
2.12 Régions de contrôle multivariée et univariée pour les 10 observations de la phase deux pour des données individuelles.	66
2.13 Carte de contrôle cusum multivariée pour 10 observations de la phase deux avec $k=1/2$ et $h_{cusum}=5,50$ ( $n=1$ ).	73
2.14 Région de contrôle de la carte cusum multivariée avec $k=1/2$ et $h_{cusum}=5,50$ ( $n=1$ ).	74
2.15 Carte de contrôle ewma multivariée pour 10 observations de la phase deux avec $r=0,2$ et $h_{ewma}=9,65$ ( $n=1$ ).	79
2.16 Cartes de contrôle pour (A) $T^2$ , (B) cusum et (C) ewma pour 10 observations de la phase deux pour des données individuelles.	82
2.17 Cartes de contrôle individuelles de Shewhart pour les deux variables de l'exemple 2.1 ( $L=3,02$ ).	102
2.18 Représentation de l'analyse en composantes principales pour $p=2$ (6 représente le point 6 de l'exemple 2.1).	103
2.19 Cartes de contrôle individuelles de Shewhart pour les deux variables de l'exemple 2.2 ( $L=3,02$ ).	105
3.20 Série d'observations du jeu1 pour les 4 variables du procédé.	111
3.21 Fonction d'auto-corrélation pour les 4 variables du jeu1.	113

3.22	Fonction d'auto-corrélation pour les résidus du modèle $AR(1)$ multivarié pour les 4 variables du jeu1.	114
3.23	Série d'observations du jeu2 pour les 4 variables du procédé.	116
3.24	Comparaisons des diagrammes en boîte du jeu1 (gauche) et du jeu2 (droit) pour les 4 variables.	118
3.25	Carte de contrôle $T^2$ pour les 201 observations du jeu1 (pour les 4 variables).	124
3.26	Carte de contrôle $T^2$ pour les 122 observations du jeu2 (pour les 4 variables).	126
3.27	Cartes de contrôle de Shewhart univariées pour les 4 variables du jeu1 (avec $L=3,40$ ).	129
3.28	Cartes de contrôle de Shewhart univariées pour les 4 variables du jeu2 (avec $L=3,40$ ).	130
3.29	Représentation fixe de la rotation des variables $BR$ , $A$ et $B$ pour les observations du jeu1.	137
3.30	Représentation fixe de la rotation des variables $BR$ , $L$ et $B$ pour les observations du jeu1.	138
3.31	Représentation fixe de la rotation des variables $BR$ , $L$ et $B$ pour les observations du jeu2.	139

## LISTE DES TABLEAUX

---

1.1	Efficacité relative des étendues par rapport à l'écart type.	15
1.2	Données de l'exemple 1.1 (déplacement de $\delta=0$ ) pour 10 observations de la phase deux pour $n=1$ .	20
1.3	Données de l'exemple 1.2 (déplacement de $\delta=2\sigma$ ) pour 10 observations de la phase deux pour $n=1$ .	22
1.4	Données de l'exemple 1.3 (déplacement de $\delta=3\sigma$ ) pour 10 observations de la phase deux pour $n=1$ .	22
1.5	$LMS_1$ de la carte de Shewhart pour une $LMS_0$ fixée et pour un déplacement de $\delta\sigma$ de la moyenne (données individuelles).	24
1.6	Paramètres $k$ et $h_{cusum}$ de la carte cusum pour un déplacement $\delta$ de la moyenne et pour une $LMS_0$ égale à 370.	31
1.7	Paramètres $r$ et $h_{ewma}$ de la carte ewma et la $LMS_1$ correspondante pour un déplacement $\delta$ de la moyenne pour une $LMS_0$ fixée à 300 et 500.	36
1.8	Comparaisons qualitatives des cartes de Shewhart, cusum et ewma.	37

- 1.9  $LMS_1$  des cartes de Shewhart, cusum et ewma pour une  $LMS_0=500$  et un déplacement  $\delta$  de la moyenne (la valeur avec un astérique provient de simulations). 40
- 2.1 Moyennes échantillonnales et valeur de la statistique  $T_{n,f}^2$  pour les 10 échantillons de la phase deux ( $n > 1$ ). 54
- 2.2 Probabilité d'avoir un point hors-contrôle si on utilise simultanément  $p$  procédures univariées de Shewhart pour différentes valeurs de  $p$  et de  $\rho$  et pour  $\alpha=0,0027$  et  $\alpha=0,0027/p$ . 58
- 2.3 Les 10 observations de la phase deux pour  $n=1$  ayant subi un déplacement de  $\tau^2=3,22$ . 64
- 2.4 Les 10 observations de la phase deux pour  $n=1$  ayant subi un déplacement de  $\tau^2=4,07$ . 64
- 2.5 Résumé des statistiques  $T^2$  et des limites de contrôle;  $b = (2(m - 1)^2)/(3m - 4)$ . 67
- 2.6  $LMS_1$  de la procédure  $T^2$  pour des données individuelles de la phase deux, pour  $p=2$  et  $m=100$ , lorsque l'on utilise les étendues mobiles pour l'estimation de  $\Sigma$ . 69
- 2.7 Valeurs du paramètre  $h_{cusum}$  à utiliser pour  $k=0,5$ ,  $p$  et  $LMS_0$  donnés pour la procédure cusum multivariée. 75
- 2.8 Valeurs des paramètres  $r$  et  $h_{ewma}$  à utiliser pour différents déplacements  $\tau$  et différentes valeurs de  $p$  pour une  $LMS_0=200$  pour la procédure ewma multivariée. 80



- 2.9  $LMS_1$  des cartes multivariées  $T^2$ , cusum et ewma pour une  $LMS_0=200$   
et un déplacement  $\tau$  du vecteur des moyennes pour  $p=2$  et  $m=100$ . 84
- 3.1 Valeurs de la statistique de Student pour les observations hors-  
contrôle des jeu1 (14, 29, 97, 100, 105 et 183) et jeu2 (7 et 30). 131
- 3.2 Scores normalisés des 4 composantes principales pour les 6 points  
hors-contrôle du jeu1 (14, 29, 97, 100, 105 et 183) et les 2 points  
hors-contrôle du jeu2 (7 et 30). 132
- 3.3 Contributions des variables pour les 6 points hors-contrôle du jeu1  
(14, 29, 97, 100, 105 et 183) et les 2 points du jeu2 (7 et 30) pour  
les scores normalisés supérieurs à 2,5. 133

## INTRODUCTION

---

Dans ce mémoire, nous aborderons le sujet du contrôle statistique des procédés, communément appelé le CSP.

Avec les nouvelles technologies, il est maintenant possible d'obtenir des données sur plusieurs variables à la fois et ce à des intervalles de temps très courts. Pour cette raison, l'utilisation d'une procédure multivariée est essentielle. Le but de ce mémoire est de présenter les principales procédures multivariées pour le CSP, procédures qui tiennent compte de la structure de corrélation entre les variables.

Les procédures multivariées les plus répandues sont des adaptations multivariées de procédures univariées. Il est donc important de bien comprendre ces procédures univariées avant d'analyser celles multivariées. On débutera donc par une analyse des trois principales cartes au niveau univarié, qui sont la carte de Shewhart, introduite par Shewhart dans les années 1920, la carte à somme cumulée, introduite par Page en 1954, et la carte à moyenne mobile et à pondération exponentielle, introduite par Roberts en 1959.

La carte de Shewhart a longtemps été considérée comme la seule carte pour faire du CSP, en raison principalement de sa robustesse et de sa facilité d'utilisation. Cependant, ces dernières années avec la prolifération des ordinateurs, la facilité avec laquelle une carte peut se calculer perd de son importance puisque

ces calculs peuvent tous être programmés. Il est donc important de considérer de nouvelles cartes de contrôle dans notre analyse, comme les cartes cusum et ewma.

Les cartes de contrôle multivariées sont beaucoup plus récentes. Il y a cependant un engouement pour ces cartes ces dernières années. La version multivariée de la carte de Shewhart, la carte  $T^2$ , a été introduite en 1947 par Hotelling. La carte cusum multivariée a pour sa part été introduite en 1988 par Crosier. Finalement, la carte ewma multivariée fut introduite par Lowry, Woodall, Champ et Rigdon en 1992.

Pour comparer les cartes de contrôle entre elles, on utilisera principalement une mesure quantitative à savoir la longueur moyenne d'une suite, dénotée par  $LMS$ . La  $LMS$  est le nombre espéré d'observations avant que la carte n'émette un signal. Évidemment, on voudra cette  $LMS$  la plus longue possible s'il n'y a pas de changement dans la moyenne (pour ne pas que la carte émette de faux signaux trop souvent) et on la voudra la plus petite possible s'il y a véritablement un changement dans la moyenne du procédé (pour être informé le plus rapidement possible). Ces analyses des longueurs moyennes d'une suite seront effectuées autant au niveau univarié que multivarié.

Les cartes multivariées nous permettent de détecter un point hors-contrôle en tenant compte des  $p$  caractéristiques simultanément. Elles ne nous permettent cependant pas de déterminer quelle variable ou groupe de variables est responsable de ce signal. Il est pourtant primordial d'être en mesure d'interpréter un signal obtenu sur une carte multivariée. Ainsi, des méthodes de détection des variables hors-contrôle ont été introduites par différents auteurs. Ces méthodes sont un complément aux cartes de contrôle multivariées et ne les remplacent pas. On doit suivre la moyenne d'un procédé sur une carte multivariée et quand il y a un signal, alors on applique les techniques de détection des variables hors-contrôle.

Dans ce mémoire, on présentera donc les méthodes les plus efficaces pour détecter quelle variable ou groupe de variables est responsable du signal hors-contrôle.

Aussi, très souvent les données obtenues d'un procédé industriel ne respectent pas les hypothèses d'indépendance que l'on a supposées pour la construction des cartes de contrôle. On vérifiera donc l'impact de l'auto-corrélation sur les données dans le cas d'un modèle auto-régressif d'ordre 1.

Le chapitre 1 sera consacré aux cartes de contrôle univariées; on y verra entre autres l'origine et la philosophie du CSP, les outils du CSP, la carte de Shewhart, la carte cusum et la carte ewma. Le chapitre 2 sera pour sa part consacré aux procédures multivariées. On y verra d'abord la carte  $T^2$ , puis la carte cusum multivariée ainsi que la carte ewma multivariée. On y présentera aussi les méthodes les plus efficaces pour détecter les variables hors-contrôle. On terminera ce chapitre en parlant des données auto-corrélées. Le chapitre 3 sera réservé à l'application de ces procédures multivariées à des jeux de données tirés du domaine des pâtes et papier, en particulier aux 4 variables qui ont trait à la blancheur du papier, c'est-à-dire la brillance et les variables  $L$ ,  $A$  et  $B$ .

## CHAPITRE 1

---

### INTRODUCTION AU CONTRÔLE STATISTIQUE DES PROCÉDÉS

Dans ce premier chapitre, on fera un survol du contrôle statistique des procédés (CSP) univarié, c'est-à-dire lorsqu'une seule caractéristique de qualité est étudiée. Nous débuterons en parlant de l'origine et de la philosophie du CSP. Nous verrons également les sept outils principaux du CSP et plus particulièrement la carte de contrôle. Les trois sections suivantes nous permettront de voir en détail le fonctionnement des trois cartes les plus répandues, c'est-à-dire la carte de Shewhart, la carte à somme cumulée et la carte à moyenne mobile et à pondération exponentielle. La section suivante nous permettra de faire une comparaison de ces trois cartes. Nous terminerons ce chapitre en parlant de l'effet des données non normales, auto-corrélées et multivariées.

Le présent chapitre est principalement inspiré de l'ouvrage de Montgomery (1996) qui s'intitule *Introduction to Statistical Quality Control*.

## 1.1. ORIGINE ET PHILOSOPHIES DU CSP

### 1.1.1. Historique du CSP

Le contrôle statistique des procédés (aussi appelé contrôle statistique de qualité) a trouvé son origine dans les années 1920 avec l'apparition des cartes de contrôle. C'est un Américain du nom de Walter A. Shewhart, qui travaillait pour la compagnie Bell Telephone Laboratories, qui développa le concept de ces cartes statistiques. Malgré l'amélioration de la qualité des produits et l'augmentation du volume de production, il fut incapable d'implanter le concept dans l'industrie car le CSP était considéré comme trop technique et la demande était tellement grande à cette époque que la qualité importait peu pour les directeurs d'usine.

Ce fut plutôt W. Edwards Deming, un collègue de travail de Shewhart, qui pendant la deuxième guerre mondiale eut la chance d'appliquer les concepts de Shewhart à la fabrication du matériel de guerre. Cette tentative ne connut cependant pas un grand succès.

Ce n'est donc qu'à la fin des années 1940 au Japon que Deming réussit à faire la preuve de l'efficacité de ces nouvelles méthodes. C'est en convainquant les hauts dirigeants japonais de l'impact que pouvait avoir le CSP sur la qualité et la productivité qu'il réussit à implanter les cartes de contrôle et que le Japon prit le virage de l'amélioration continue. Dans les années 1970, les Américains prirent eux aussi ce virage devant le succès remporté par ces nouvelles méthodes au Japon. Depuis lors, le CSP a pris une place de plus en plus importante dans la façon de contrôler les procédés de fabrication.

### 1.1.2. Le concept de variation

Lors d'un procédé de fabrication, on sait qu'il est pratiquement impossible de fabriquer deux objets parfaitement identiques. On s'attend plutôt à produire des produits semblables mais non identiques. Il est donc tout à fait normal qu'une certaine variation soit présente dans le procédé de fabrication. Cette variation est directement liée à la qualité puisque Taguchi définit la qualité comme un minimum de variation sur l'objectif (voir Taguchi et Wu, 1980 et Taguchi, 1986).

Le CSP est un ensemble d'outils qui nous permettra d'atteindre les deux objectifs primordiaux du contrôle de la qualité, c'est-à-dire continuellement chercher à améliorer la performance du procédé et réduire la variation sur les paramètres clés.

Pour bien comprendre la situation, il serait utile de faire la distinction entre deux concepts de variation, celui en ingénierie et celui de Shewhart.

Le concept de variation en ingénierie repose sur des spécifications établies par le client pour s'assurer que les produits sont conformes. Le principe est simple, tout ce qui rencontre les spécifications est bon, tout ce qui ne rencontre pas les spécifications est mauvais. Le but est de se conformer aux spécifications. Le principal problème est que les spécifications ne sont généralement pas reliées au procédé de fabrication, celles-ci n'empêchent donc pas la fabrication de mauvais produits et ne réduit en rien la variation.

C'est pourquoi Shewhart a établi son propre concept de variation. Selon lui, un procédé possède deux types de variation. Une première est due à des causes naturelles (variation dans les matières premières, variation normale de la machinerie, ...) et ne peut être évitée car elle est inhérente au processus de fabrication. Des changements au processus habituel comme un bris d'équipement

ou un mauvais approvisionnement sont une deuxième source de variation qu'on appelle variation due à une cause assignable. Cette variation peut produire un changement dans la moyenne du procédé ou dans sa variance, changement que l'on voudra noter. Une fois le changement noté, on voudra trouver l'origine du problème, la cause assignable, afin de pouvoir corriger le problème.

Il a donc créé un outil, la carte de contrôle, qui permet de distinguer la variation naturelle (due au hasard) de la variation assignable (due à une cause assignable). Si la variation est due au seul facteur de hasard, on dira que le procédé est en contrôle. Par contre, si elle est due à des causes assignables, on dira que le procédé est hors-contrôle. Le but d'une carte de contrôle est de détecter la variation due aux causes assignables. Par la suite, lorsqu'une variation assignable a été identifiée, on se doit de chercher la cause de cette variation. Pour améliorer notre procédé, on doit réduire ce genre de variation en trouvant et en corrigeant ces causes assignables.

Il faut également mentionner que le fait qu'un procédé soit hors-contrôle n'est pas nécessairement négatif. En effet, si la variance du procédé diminue, la carte se doit d'émettre un signal pour nous aviser de ce changement, et ce même si celui-ci augmente la performance du procédé.

La force de la carte tient dans sa façon de calculer les limites, c'est-à-dire à partir de mesures qui proviennent du procédé et non à partir des spécifications du client. En cherchant et en éliminant les causes assignables, on sera en mesure de rendre la variation du procédé cohérente et prévisible et une fois la maîtrise de la variation atteinte, on peut essayer de rencontrer les spécifications.



### 1.1.3. Les quatre états d'un procédé

Avec la distinction préalablement établie entre les spécifications et les limites de contrôle, on peut s'apercevoir qu'un procédé de fabrication peut se trouver dans quatre états possibles. Selon une nomenclature particulière à Wheeler et Chambers (1992), ces quatre états sont, du plus favorable au moins favorable, l'état idéal (100 % de produits à l'intérieur des spécifications et 100 % à l'intérieur des limites de contrôle); l'état de seuil (produits hors-spécifications mais à l'intérieur des limites); l'état bord du chaos (produits dans les spécifications mais hors-limites) et finalement l'état de chaos (produits hors-spécifications et hors-limites).

Dépendant de l'état dans lequel le procédé se trouve, il existe des actions possibles. Dans l'état idéal, puisque tout va bien, on doit continuer à utiliser les cartes de contrôle pour maintenir le procédé en contrôle et nous avertir si des problèmes surviennent. L'état suivant est celui dit de seuil. Cet état sera préféré à l'état bord du chaos puisqu'ici, on est en contrôle statistique, le procédé est donc prévisible. Cependant, malgré cette consistance, des produits se retrouvent à l'extérieur des spécifications, la solution pour remédier à ce problème est de changer le procédé (ajuster la cible ou réduire la variation) ou élargir les spécifications. Une solution temporaire serait de trier les produits.

L'état bord du chaos semble à prime abord correct puisqu'il produit 100 % de produits conformes, mais cette façon de voir les choses mène à une négligence du procédé puisque celui-ci n'est pas en contrôle et est donc imprévisible. Ainsi, le taux de 100 % de produits conformes peut changer à tout moment. Puisque l'imprévisibilité de ce procédé est due à un manque de contrôle statistique, la façon de sortir de cet état est d'éliminer les causes assignables par l'utilisation des cartes de contrôle. Le dernier état est celui du chaos. Ici, en plus de ne pas être en contrôle statistique, il y a production de produits non conformes. La seule

façon d'amener un procédé hors de cet état est d'éliminer les causes assignables. Ce qui nécessite l'utilisation des cartes de contrôle.

Maintenant que l'origine et les philosophies du CSP ont été présentées, dans la prochaine section on verra plus en détail les outils qui nous permettent d'atteindre un procédé prévisible.

## 1.2. LES OUTILS DU CSP

### 1.2.1. Les sept outils principaux

Comme on l'a déjà mentionné, le CSP est un ensemble puissant d'outils permettant d'atteindre la stabilité d'un procédé par la réduction de la variation. Il existe sept outils principaux, que l'on peut regrouper en trois groupes. Le premier groupe contient le formulaire de collecte de données et le diagramme de concentration des défauts qui sont une énumération descriptive. Le deuxième comprend deux schémas, soit le diagramme de Pareto et le diagramme de cause à effet, alors que le dernier groupe contient trois outils statistiques, l'histogramme, le diagramme en nuage de points et la carte de contrôle.

Pour un bon résumé des six premiers outils voir Montgomery (1996, pp 150 à 158). Par contre, l'outil le plus utilisé et le plus sophistiqué est la carte de contrôle puisque c'est le seul outil qui donne la possibilité d'obtenir un procédé prévisible. Nous allons donc à partir de maintenant traiter plus en détail cet outil essentiel au bon fonctionnement du CSP.

### 1.2.2. La carte de contrôle

Une carte de contrôle nous permet de suivre la moyenne d'un procédé ou sa variation et de décider si ce procédé est en contrôle ou non. Le but est de détecter

la variation assignable par l'émission d'un signal sur la carte. Lorsque la carte a détecté ce problème, on doit l'étudier pour trouver la cause de cette variation. Lorsque la cause assignable est bien identifiée, on doit l'éliminer pour améliorer notre procédé.

La figure 1.1 montre une carte de contrôle standard où une mesure d'une caractéristique de qualité  $X$  est placée sur un graphique en fonction du numéro de l'observation ou du temps. Chacune des observations est reliée à l'aide de segments pour faciliter la visualisation de la séquence de points.

La carte contient une ligne centrale ( $LC$ ) qui correspond généralement à la moyenne échantillonnale de la caractéristique de qualité lorsque le procédé est en contrôle. Il y a également deux autres lignes qui correspondent à la limite supérieure de contrôle ( $LSC$ ) et la limite inférieure de contrôle ( $LIC$ ).

Ces limites sont choisies de manière à ce que si le procédé est en contrôle la grande majorité des points soient à l'intérieur de ces deux limites. Tant que le point est entre les deux limites et qu'il n'y a aucune tendance non-aléatoire, on dira que le procédé est en contrôle et aucun correctif n'est nécessaire. Cependant, lorsqu'un point se trouve à l'extérieur de ces limites ou qu'une tendance non aléatoire est identifiable, le procédé sera jugé hors-contrôle et une action devra être entreprise pour trouver et éliminer la cause assignable responsable de ce signal.

Il existe principalement trois types de cartes de contrôle univariées. Il s'agit de la carte de Shewhart, de la carte à somme cumulée (cusum) et de la carte à moyenne mobile et à pondération exponentielle (ewma). Nous verrons respectivement chacune de ces cartes dans les trois prochaines sections. Afin de comparer la performance de chacune des cartes, on a besoin d'une mesure commune. La plus utilisée en pratique est la longueur moyenne d'une suite (de l'anglais *average run length*).

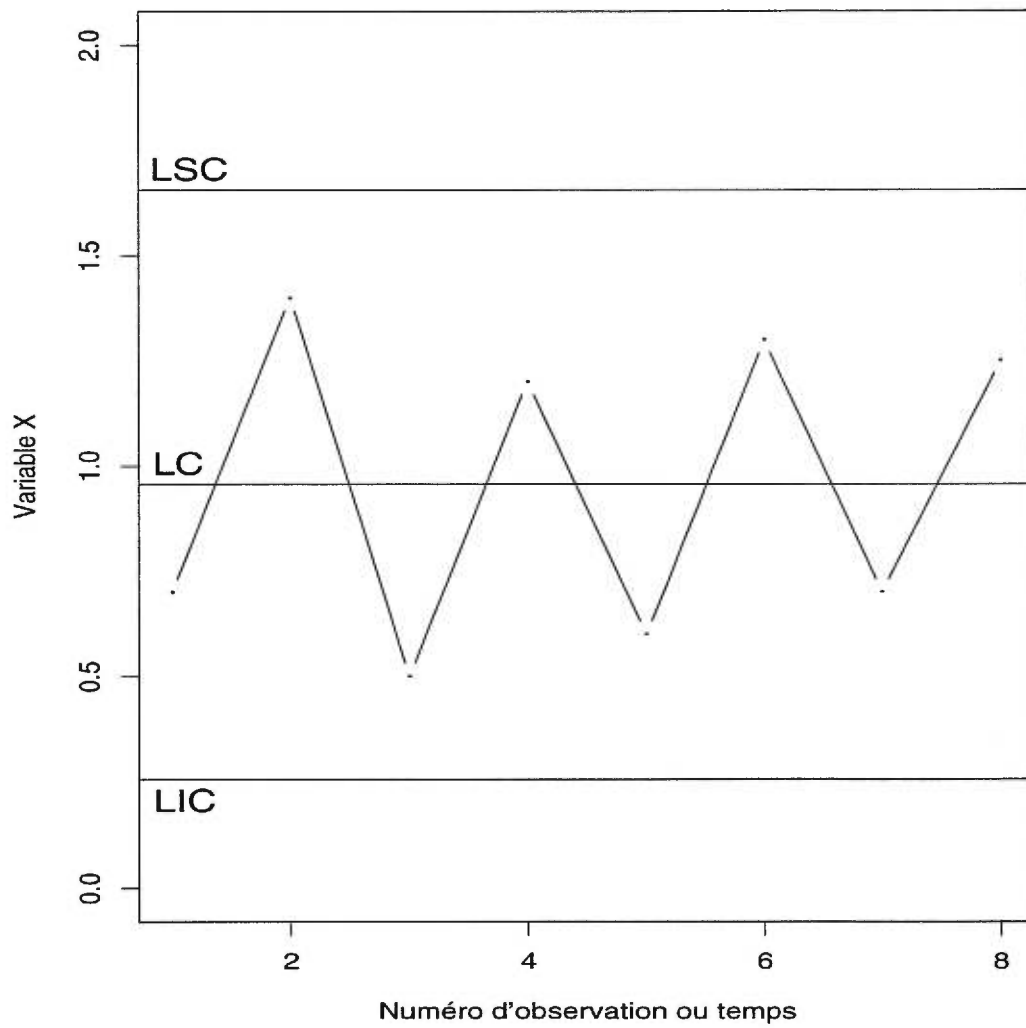


FIGURE 1.1. Carte de contrôle typique.

Par définition, la  $LMS$  est le nombre espéré d'observations requis avant que la carte n'émette un signal. Il existe deux types de  $LMS$ , celle lorsque le procédé est en contrôle, que l'on dénotera par  $LMS_0$  et celle pour un procédé qui a subi un déplacement, que l'on dénotera par  $LMS_1$ . La carte affichant la meilleure performance sera donc celle pour laquelle (pour une  $LMS_0$  fixée) la  $LMS_1$  sera la plus petite s'il y a véritablement un changement dans la moyenne du procédé. Par contre, s'il n'y a pas de changement dans le procédé, on voudra que la  $LMS_0$  soit la plus élevée possible (puisque l'on veut minimiser  $\alpha$  le risque d'erreur de type 1, c'est-à-dire de faussement déclarer le procédé hors-contrôle).

### 1.3. LA CARTE DE SHEWHART

Cette carte de contrôle est la plus utilisée dans l'industrie et la plus facilement applicable. Il existe principalement deux types de carte de Shewhart, la carte  $\bar{X}R$  et la carte  $XRm$ . La première est utilisée avec des données groupées et la deuxième avec des données individuelles. Nous les verrons plus en détail dans les deux prochaines sections.

#### 1.3.1. La carte $\bar{X}R$

Cette série de cartes est composée de deux cartes distinctes, la première pour suivre la moyenne du procédé, la carte  $\bar{X}$  et la deuxième pour suivre la variation, la carte  $R$ .

Pour appliquer la théorie de Shewhart et ainsi être en mesure de développer ces cartes on doit tout d'abord avoir le modèle suivant.

Modèle 1.3.1 *Soit une caractéristique de qualité  $X$ . On recueille  $m$  échantillons indépendants chacun contenant  $n$  observations indépendantes et identiquement*

distribuées  $x_{j1}, \dots, x_{jn}$  où  $x_{ji} \sim N(\mu, \sigma^2)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ ,  $\mu$  et  $\sigma^2$  inconnus.

La moyenne échantillonnale pour le  $j$ -ième échantillon est,

$$\bar{x}_j = \frac{x_{j1} + \dots + x_{jn}}{n}, \quad j = 1, \dots, m.$$

On sait alors que  $\bar{x}_j \sim N(\mu, \sigma^2/n)$ ,  $j = 1, \dots, m$ .

Puisque  $\mu$  et  $\sigma^2$  sont inconnus, ils devront être estimés à partir des échantillons que l'on a recueillis. De plus, la taille de chaque échantillon  $n$  sera habituellement petite, c'est-à-dire,  $n \leq 6$ .

Le meilleur estimateur de  $\mu$ , la moyenne du procédé, est la moyenne globale, soit

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \dots + \bar{x}_m}{m}.$$

Cette moyenne sera utilisée comme ligne centrale ( $LC_{\bar{X}}$ ) sur la carte  $\bar{X}$ . Pour construire les limites de contrôle, on a besoin d'un estimé de  $\sigma$ . On doit noter que le Modèle 1.3.1 est valable seulement si le procédé est en contrôle. Mais en réalité le procédé n'est que très rarement en contrôle. On se doit donc de définir un modèle alternatif qui répondra à cette réalité. Ce modèle sera utilisé dans le temps pour l'amélioration continue. Alors soit le modèle d'analyse de variance avec effets aléatoires suivant,

$$X_{ji} = \mu_j + \epsilon_{ji}, \quad j = 1, \dots, m, \quad i = 1, \dots, n,$$

où les  $\mu_j$  sont indépendants de loi  $N(\mu, \sigma_\mu^2)$ , les  $\epsilon_{ji}$  sont indépendants de loi  $N(0, \sigma_\epsilon^2)$  et les  $\mu_j, \epsilon_{ji}$  sont également indépendants entre eux.

On note donc qu'en réalité, les observations sont soumises à deux types de variation. Une première entre les échantillons puisque la moyenne de chaque échantillon n'est pas nécessairement la même. On l'appellera variation inter-groupes ou globale et on la dénote par  $\sigma_\mu^2$ . La deuxième est la variation à

l'intérieur de chaque échantillon que l'on appellera variation intra-groupe ou locale et on la dénote par  $\sigma_\epsilon^2$ .

Le but d'une carte de contrôle est de détecter une variation de la moyenne, c'est-à-dire un changement de la variation inter-groupes. Pour ce faire, on a besoin de connaître la variation intra-groupe qui elle est toujours présente dans notre modèle alternatif. Notre estimation de  $\sigma$  doit donc représenter une estimation de la variation intra-groupe. Si on est en contrôle (Modèle 1.3.1), il est naturel d'utiliser  $s$ , l'écart type échantillonnal, puisque c'est le meilleur estimateur de  $\sigma$ . Cependant, en réalité, on ne sera que très rarement en contrôle, alors en utilisant  $s$  de l'échantillon complet, on va surestimer notre variation intra-groupe.

Alors pour avoir une estimation de la variation locale  $\sigma_\epsilon^2$ , on va utiliser une moyenne des écart types de chaque échantillon ou une moyenne des étendues des  $m$  échantillons. La méthode la plus efficace est d'utiliser les écart types. Cependant, pour des raisons historiques (simplicité du calcul de  $R$  par rapport à  $s$ ), la méthode la plus répandue est d'utiliser les étendues. L'étendue de l'échantillon  $j$  est la différence entre la plus grande et la plus petite observation, soit

$$R_j = x_{(jn)} - x_{(j1)}, \quad j = 1, \dots, m,$$

où  $x_{(j1)} \leq x_{(j2)} \leq \dots \leq x_{(jn)}$  sont les statistiques d'ordre des observations de l'échantillon  $j$ .

L'estimation de la variation à court terme est l'étendue moyenne des  $m$  échantillons, soit

$$\bar{R} = \frac{R_1 + \dots + R_m}{m}.$$

Pour des raisons historiques, on utilise donc  $\bar{R}$ . Si  $n < 10$ , la perte d'efficacité de l'utilisation des étendues par rapport à l'écart type est faible. Le tableau 1.1 (Montgomery, 1996) nous montre l'efficacité relative de l'utilisation de l'étendue  $R$  par rapport à l'écart type échantillonnal  $s$  pour l'estimation de l'écart type  $\sigma$  en fonction de la taille de l'échantillon et ce pour des données normales.

$n$	Efficacité relative
2	1,000
3	0,992
4	0,975
5	0,955
6	0,930
10	0,850

TABLEAU 1.1. Efficacité relative des étendues par rapport à l'écart type.

Si  $n$  est large,  $R$  perd rapidement de l'efficacité par rapport à  $s$ . Par contre, plus  $n$  augmente et plus l'estimation de  $\sigma$  par  $R$  ou  $s$  est précise de sorte que l'utilisation de l'un ou l'autre changera peu de chose en pratique. Donc, en conformité avec la pratique, on utilisera  $\bar{R}$  comme estimateur de  $\sigma$ .

Les limites de contrôle supérieure et inférieure ainsi que la ligne centrale de la carte  $\bar{X}$  sont données par

$$\begin{aligned}
 LSC_{\bar{X}} &= \bar{\bar{x}} + \frac{L}{d_{2;n}\sqrt{n}}\bar{R} \\
 LC_{\bar{X}} &= \bar{\bar{x}} \\
 LIC_{\bar{X}} &= \bar{\bar{x}} - \frac{L}{d_{2;n}\sqrt{n}}\bar{R},
 \end{aligned}
 \tag{1.3.1}$$

où le terme  $(\frac{\bar{R}}{d_{2;n}\sqrt{n}})$  est l'estimation de  $\sigma_{\bar{x}}$  et le paramètre  $L$  est choisi en fonction de la probabilité de faussement détecter un signal. Cette valeur est choisie de façon à ce que si un point se trouve à l'extérieur de ces limites, on est pratiquement certain qu'une cause assignable est reliée à ce signal et on aura que très peu de faux signaux. C'est pour cette raison que pour les cartes de Shewhart, on prendra  $L=3$ , qui est la valeur usuelle et qui correspond "aux limites trois-sigma". De plus ce choix est robuste à l'hypothèse de normalité, c'est-à-dire que même si la distribution des observations n'est pas normale, la probabilité de faussement détecter un signal est faible. Pour des données normales, si  $L=3$ , cette probabilité est de 0,0027 lorsque  $\sigma$  est connu.

La constante  $d_{2;n}$  dépend de la taille de l'échantillon  $n$  et est basée sur l'hypothèse de normalité des données. Pour les valeurs de ces constantes en fonction



de  $n$ , voir Montgomery (1996, Appendice VI). Cette constante provient de la relation qui existe entre l'étendue d'un échantillon d'une distribution normale et l'écart type de cette distribution. Elle est calculée de façon à ce que l'estimé de  $\sigma$  soit sans biais. Une estimation de  $\sigma$  est donc  $R/d_{2;n}$ . Si  $\bar{R}$  est la moyenne des étendues des  $m$  échantillons, on utilisera  $\hat{\sigma} = \bar{R}/d_{2;n}$  et  $\hat{\sigma}_{\bar{x}} = \bar{R}/d_{2;n}\sqrt{n}$ .

La carte de contrôle pour suivre la moyenne d'un procédé est ainsi construite avec  $LSC_{\bar{X}}$ ,  $LC_{\bar{X}}$  et  $LIC_{\bar{X}}$ . On peut maintenant calculer la moyenne d'un nouvel échantillon et vérifier si le procédé est en contrôle ou non, c'est-à-dire si la moyenne est à l'intérieur des limites ou non et par la suite procéder de la même façon pour tous les échantillons suivants.

La variation du procédé peut être suivie à l'aide des étendues sur la carte  $R$ . Les limites de contrôle supérieure et inférieure ainsi que la ligne centrale de la carte  $R$  sont données par

$$\begin{aligned} LSC_R &= \bar{R} + L d_{3;n} \frac{\bar{R}}{d_{2;n}} \\ LC_R &= \bar{R} \\ LIC_R &= \max[0, \bar{R} - L d_{3;n} \frac{\bar{R}}{d_{2;n}}], \end{aligned} \tag{1.3.2}$$

où le terme  $(d_{3;n} \frac{\bar{R}}{d_{2;n}})$  est l'estimation de  $\sigma_R$  et  $L$  sera encore une fois choisi en fonction de la probabilité de faussement détecter un signal et habituellement pour les cartes de Shewhart, on posera  $L=3$ . La constante  $d_{3;n}$  dépend de  $n$  et est basée sur l'hypothèse de normalité des données. Puisque les limites de contrôle de la carte  $R$  sont de la forme  $\bar{R} \pm L\sqrt{\text{var}(R)}$ , on veut un estimé de la variance de  $R$ . On a  $R = \sigma W$  où  $W$  est l'étendue de  $n$   $N(0, 1)$ . On a donc  $\sqrt{\text{var}(R)} = \sigma\sqrt{\text{var}(W)}$ . On a trouvé que le terme  $\bar{R}/d_{2;n}$  estime  $\sigma$  alors la constante  $d_{3;n} = \sqrt{\text{var}(W)}$ . Pour les valeurs de ces constantes en fonction de  $n$ , voir Montgomery (1996, Appendice VI).

On doit souligner ici la présence d'une borne inférieure pour la carte  $R$ . Celle-ci est présente car, comme on l'a déjà mentionné, si la variation du procédé diminue de façon importante, on doit en être informé puisque le procédé a subi un changement (positif) et on doit trouver la cause assignable responsable de ce changement pour pouvoir être en mesure d'améliorer notre procédé.

Pour détecter un point hors-contrôle sur la carte  $R$ , le principe est le même que sur la carte  $\bar{X}$ . Ces deux cartes sont jumelées pour former la carte  $\bar{X}R$ , qui nous permet de suivre à la fois la moyenne d'un procédé et sa variation pour des échantillons de taille  $n$  pour une caractéristique de qualité donnée. Les deux cartes sont construites simultanément et sont placées l'une sous l'autre pour mieux suivre le procédé.

La figure 1.2 montre un exemple de carte  $\bar{X}R$ . Les données proviennent d'une loi  $N(10, 1)$ . On a simulé  $m=20$  échantillons, chacun de taille  $n=5$  de cette loi normale et on a calculé nos estimés de  $\mu$  et de  $\sigma$  à partir de ceux-ci. Par la suite, on a calculé les limites de contrôle à l'aide des équations (1.3.1) et (1.3.2). On a utilisé  $L=3$  pour le calcul de ces limites. À partir des moyennes des échantillons ( $\bar{x}_j$ ), des étendues ( $R_j$ ) et des limites de contrôle, il nous a été possible de construire une carte  $\bar{X}R$  pour suivre la moyenne et la variation du procédé.

On s'aperçoit donc qu'il n'y a aucun point hors-contrôle et que l'on peut ainsi déclarer le procédé en contrôle statistique. Par la suite, on va obtenir de nouveaux échantillons, et on va vérifier si le procédé est toujours en contrôle en plaçant sur la carte de contrôle la valeur de la moyenne et de l'étendue de ce nouvel échantillon. On doit noter que l'on ne recalcule pas de nouvelles limites de contrôle, car pour tous les échantillons suivants, on va se servir des limites de contrôle calculées précédemment et donc en se servant des estimés de  $\mu$  et de  $\sigma$  calculés précédemment.

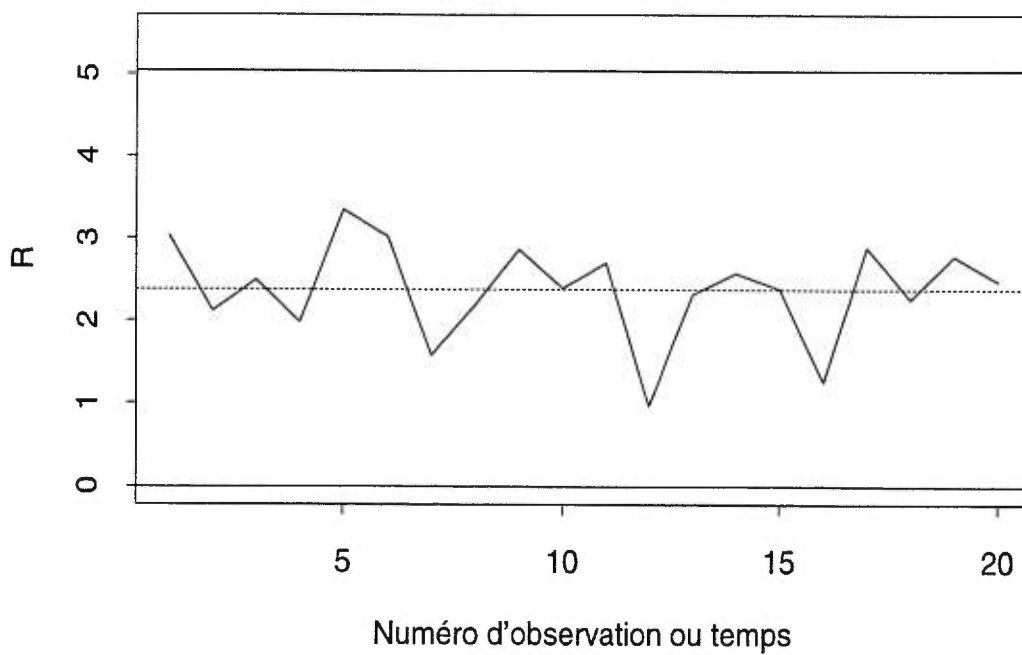
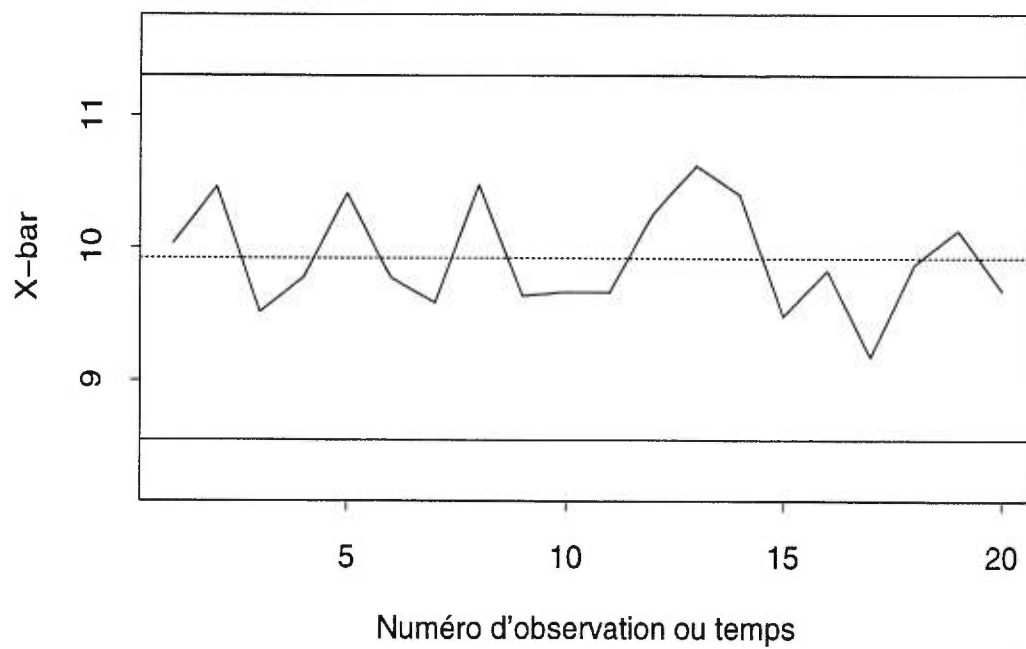


FIGURE 1.2. Carte de contrôle  $\bar{X}R$  pour 20 observations de la phase un ( $n=5$  et  $L=3$ ).

### 1.3.2. La carte $XRm$

Lorsque la taille de chaque sous-groupe est de  $n=1$ , la méthode utilisant la carte  $\bar{X}R$  n'est pas appropriée puisqu'il devient impossible de calculer des étendues à l'intérieur des sous-groupes. On a alors recours à la carte  $XRm$  lorsque l'on doit analyser un procédé dont le regroupement rationnel est de taille 1.

Ici, on estimera la variation à court terme par la moyenne de toutes les étendues mobiles obtenues par les différences successives entre les valeurs individuelles. C'est cette approximation qui sera utilisée afin de construire les limites de contrôle pour les valeurs individuelles (carte  $X$ ) et pour les étendues mobiles (carte  $Rm$ ).

Sous les hypothèses du Modèle 1.3.1 avec cependant  $n=1$ , on peut construire la carte  $XRm$ . On simplifiera la notation en utilisant  $x_j$  à la place de  $x_{ji}$  puisque  $i=1$  pour tous les échantillons.

On calcule tout d'abord les étendues mobiles en prenant la différence entre chacune des valeurs individuelles

$$Rm_j = |x_j - x_{j-1}|, \quad j = 2, \dots, m. \quad (1.3.3)$$

On calcule ensuite la moyenne des valeurs individuelles et l'étendue mobile moyenne du procédé,

$$\bar{x} = \frac{x_1 + \dots + x_m}{m}, \quad \overline{Rm} = \frac{Rm_2 + \dots + Rm_m}{m-1}. \quad (1.3.4)$$

On calcule finalement les limites de contrôle et la ligne centrale pour la carte  $X$ ,

$$\begin{aligned} LSC_X &= \bar{x} + L \frac{\overline{Rm}}{d_{2;2}} \\ LC_X &= \bar{x} \\ LIC_X &= \bar{x} - L \frac{\overline{Rm}}{d_{2;2}}. \end{aligned} \quad (1.3.5)$$

Et pour la carte  $Rm$ ,

$$\begin{aligned}
 LSC_{Rm} &= \overline{Rm} + L d_{3;2} \frac{\overline{Rm}}{d_{2;2}} \\
 LC_{Rm} &= \overline{Rm} \\
 LIC_{Rm} &= 0.
 \end{aligned}
 \tag{1.3.6}$$

Encore une fois, on prendra très souvent  $L=3$ . Aussi pour des observations individuelles, les étendues mobiles sont calculées à partir de  $n=2$  et ainsi on a que  $d_{2;2}=1,128$  et  $d_{3;2}=0,853$ .

Nous allons maintenant construire un exemple de carte  $XRm$ . Les données ont été simulées à partir d'une loi normale avec  $n=1$  observation par échantillon. On a tout d'abord simulé  $m=30$  observations provenant de la loi  $N(10,1)$ . À partir de celles-ci, on a calculé les estimés de  $\mu$  et de  $\sigma$ . On a obtenu  $\hat{\mu}=10,01$  et  $\hat{\sigma}=1,03$ . Par la suite, on a calculé les limites de contrôle pour suivre la moyenne de ce procédé à partir de l'équation (1.3.5). On est maintenant en mesure d'obtenir de nouveaux échantillons et de vérifier s'ils sont en contrôle en utilisant les limites de contrôle que l'on vient de calculer.

Nous allons faire trois exemples différents. Le premier pour un déplacement de la moyenne de  $\delta=0$ . Le deuxième pour un déplacement de  $\delta=2\sigma$  et le troisième pour  $\delta=3\sigma$ .

Pour le premier exemple, supposons donc que l'on obtient 10 nouvelles observations avec la même moyenne (10) et la même variance (1). Le tableau 1.2 contient les 10 observations de l'exemple 1.1.

$j$	1	2	3	4	5	6	7	8	9	10
$x_j$	11,9	9,0	8,4	11,3	9,7	8,4	9,6	10,4	9,4	10,5

TABLEAU 1.2. Données de l'exemple 1.1 (déplacement de  $\delta=0$ ) pour 10 observations de la phase deux pour  $n=1$ .

La figure 1.3 représente la carte de contrôle de Shewhart pour les données de l'exemple 1.1 et pour suivre la moyenne du procédé seulement et avec les limites de contrôle calculées à partir des estimés des paramètres obtenus des 30 échantillons précédents. Les limites ont été calculées à partir de  $L=3,09$  pour être en mesure de comparer avec les cartes cusum et ewma. Les lignes pleines représentent les limites de contrôle alors que la ligne pointillée représente la ligne centrale.

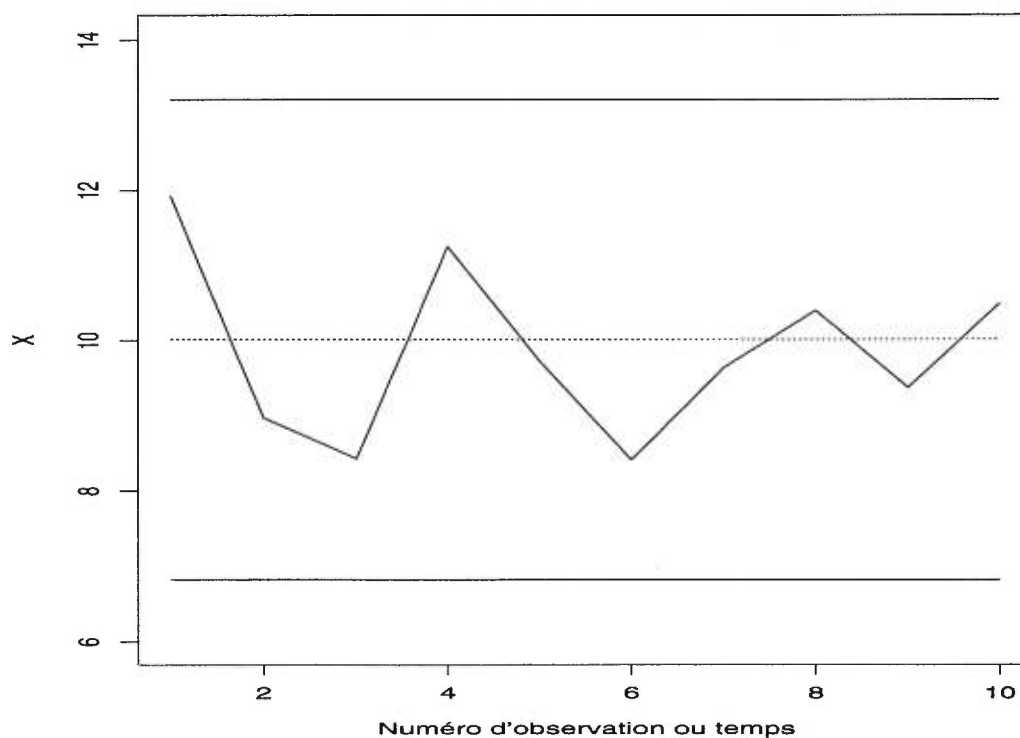


FIGURE 1.3. Carte de contrôle de Shewhart pour 10 observations en contrôle ( $\delta=0$ ) avec  $L=3,09$  (phase deux et  $n=1$ ).

On s'aperçoit qu'il n'y a aucun point hors-contrôle. Ce qui est prévisible puisque les 10 nouvelles observations étaient sur la cible.

Supposons maintenant que l'on obtient 10 nouvelles observations qui ont subi un déplacement de la moyenne de  $\delta=2\sigma$  et sont donc de loi  $N(12,1)$ . Le tableau 1.3 contient les 10 observations de l'exemple 1.2.

$j$	1	2	3	4	5	6	7	8	9	10
$x_j$	10,8	10,5	12,9	12,8	12,1	11,7	14,1	10,4	11,4	11,3

TABLEAU 1.3. Données de l'exemple 1.2 (déplacement de  $\delta=2\sigma$ ) pour 10 observations de la phase deux pour  $n=1$ .

Alors à partir de ces données et des limites calculées précédemment, on a construit la carte  $XRm$  (pour la moyenne seulement) qui est présentée à la figure 1.6 (A) à la page 42. Pour être en mesure d'établir une comparaison avec les cartes cusum et ewma, on a utilisé  $L=3,09$ .

On voit donc que la carte a détecté un point hors-contrôle à l'observation 7 sur la carte  $X$ . Une action devrait donc être entreprise pour trouver et éliminer la cause assignable responsable de ce signal.

On a finalement obtenu 10 autres observations qui cette fois ont subi un déplacement de la moyenne de  $\delta=3\sigma$  et sont donc de loi  $N(13,1)$ . Le tableau 1.4 contient les 10 observations de l'exemple 1.3.

$j$	1	2	3	4	5	6	7	8	9	10
$x_j$	12,7	13,6	12,9	13,2	13,5	14,3	12,7	13,1	14,1	13,0

TABLEAU 1.4. Données de l'exemple 1.3 (déplacement de  $\delta=3\sigma$ ) pour 10 observations de la phase deux pour  $n=1$ .

La carte de contrôle pour la moyenne est présentée à la figure 1.7 (A) à la page 43. Ici, la carte a détecté le changement plus rapidement, c'est-à-dire à l'observation 2, mais également aux observations 5, 6 et 9.

### 1.3.3. *LMS* de la carte de Shewhart

La longueur moyenne d'une suite est une caractéristique très importante d'une carte de contrôle puisqu'elle permet de comparer entre elles la performance des cartes. Comme on l'a déjà mentionné la *LMS* est le nombre espéré de points requis avant qu'un de ceux-ci ne se situe à l'extérieur des limites de contrôle. Le

fait qu'un point soit à l'extérieur des limites peut donc être vu comme une Bernoulli avec une probabilité de succès  $p = P(\text{un point soit à l'extérieur des limites})$ . Cette probabilité est fixe si la moyenne utilisée pour le calcul des limites est la vraie moyenne (fixe) alors que c'est une probabilité conditionnelle qui dépend de la moyenne (aléatoire) de l'échantillon historique, sinon. C'est ainsi que le nombre d'essais avant d'obtenir un point hors-contrôle suit une distribution géométrique de paramètre  $p$  tel que décrit précédemment avec une espérance égale à  $1/p$ .

On obtient donc pour les cartes de Shewhart,

$$LMS = \frac{1}{P(\text{un point soit à l'extérieur des limites})}.$$

Il existe deux types de longueur moyenne d'une suite, qui répondent à deux questions distinctes. On veut tout d'abord savoir combien il faudra d'observations à la carte avant qu'elle ne donne une fausse alarme (point hors-contrôle mais procédé en contrôle). Cette  $LMS$  correspond à la  $LMS_0$  pour identifier celle calculée à partir d'un procédé qui est sur la cible et elle correspond à la robustesse de la carte. Par la suite, on veut savoir combien cela prend-il d'observations avant que la carte ne détecte un signal lorsque le procédé est hors-contrôle. Cette  $LMS$  correspond à la  $LMS_1$  pour identifier celle calculée à partir d'un procédé hors-cible et elle correspond à la sensibilité de la carte.

Alors si on est sur la cible, la probabilité qu'un point soit à l'extérieur des limites est  $\alpha$ , qui représente l'erreur de type 1. Si on est hors-cible, la probabilité qu'un point soit à l'extérieur des limites est  $1 - \beta$ , où  $\beta$  représente l'erreur de type 2.

On obtient ainsi pour Shewhart,

$$LMS_0 = \frac{1}{\alpha} \quad \text{et} \quad LMS_1 = \frac{1}{1 - \beta}.$$

Pour obtenir des  $LMS$  exactes, on va supposer que  $\mu$  et  $\sigma$  sont connus ce qui n'est pas le cas du Modèle 1.3.1. Ainsi les limites de contrôle de l'équation (1.3.1) seront



donc calculées à partir de  $\mu$  et  $\sigma$  et non de leurs estimations. Puisque les  $LMS$  sont exactes pour  $\mu$  et  $\sigma$  connus, il faut noter qu'elles deviennent approximatives lorsque les estimations sont utilisées. On obtient donc

$$\begin{aligned}\alpha &= 2(1 - \Phi(L)) \\ \beta &= \Phi(L - \delta\sqrt{n}) - \Phi(-L - \delta\sqrt{n}),\end{aligned}$$

où  $\Phi$  est la fonction de répartition d'une  $N(0, 1)$ ,  $L$  est le paramètre choisi pour contrôler la largeur des limites de contrôle,  $\delta$  est le déplacement que l'on veut détecter en unités d'écart type et  $n$  est la taille de chaque sous-groupe.

Par exemple, avec le modèle des données de l'exemple 1.2, avec  $L=3$ ,  $n=1$  et  $\delta=2$ , on obtient  $LMS_0=370$  et  $LMS_1=6,30$ .

Le tableau 1.5 donne pour un déplacement de la moyenne de  $\delta\sigma$  et une  $LMS_0$  fixée, la valeur de la  $LMS_1$  pour des données individuelles. Ces valeurs des  $LMS$  sont évidemment obtenues en supposant une distribution normale des données.

$LMS_0$	$\delta$				
	0,5	1,0	2,0	3,0	4,0
300 L=2,935	129	37,7	5,72	1,90	1,17
370 L=3,000	155	43,9	6,30	2,00	1,19
500 L=3,090	202	54,6	7,26	2,15	1,22

TABLEAU 1.5.  $LMS_1$  de la carte de Shewhart pour une  $LMS_0$  fixée et pour un déplacement de  $\delta\sigma$  de la moyenne (données individuelles).

#### 1.4. LA CARTE CUSUM

Le désavantage majeur de l'utilisation des cartes de Shewhart c'est que pour décider si un point est hors-contrôle on ne se sert que de ce dernier et non de la séquence entière de points. Cela a pour effet que les cartes  $\bar{X}R$  et  $XRm$  sont relativement insensibles aux petits déplacements dans la moyenne ou dans

la variation du procédé, c'est-à-dire un changement de l'ordre de  $1,5\sigma$  ou moins dans la moyenne.

Si on est intéressé à détecter de petits déplacements, il existe deux alternatives aux cartes de Shewhart, il s'agit des cartes cusum et ewma. Dans cette section, on introduit donc un deuxième type de carte, la carte à somme cumulée, de l'anglais *cumulative sum*.

### 1.4.1. Principes de base

Cette carte a été introduite pour la première fois par Page (1954). Le principe général derrière cette carte, c'est que celle-ci tient directement compte de l'information contenue dans la séquence de points en calculant la somme cumulative des écarts entre la valeur obtenue  $x$  et une valeur cible pour la moyenne  $\mu$ . C'est pour cette raison qu'elle est très utile pour détecter de petits changements dans la moyenne. En effet, la procédure est dite séquentielle, c'est-à-dire qu'elle accumule à chaque point un écart (si celui-ci est plus grand qu'une certaine valeur) qui nous permet de détecter un changement qui se produirait plus en douceur dans la moyenne du procédé. La carte de Shewhart ne permet pas cela puisque pour décider si un point est hors-contrôle, on ne se sert que de ce dernier point. Un petit changement est donc plus difficile à détecter.

Cette carte est particulièrement utile avec des données individuelles. Si on a des données groupées, on doit prendre la moyenne de chaque échantillon pour pouvoir appliquer la procédure cusum. Pour cette raison, on va employer  $x_j$ ,  $j = 1, \dots, m$ , qui peut aussi bien représenter une seule observation que la moyenne d'un sous-groupe de taille  $n$ . De plus, il faut mentionner que la carte cusum est principalement construite pour suivre la moyenne d'un procédé. On se concentrera donc sur ce cas.

On doit également mentionner qu'il existe deux types de cartes cusum, la forme "tabular" et la forme "v-mask". On n'expliquera cependant pas ce dernier type puisque Montgomery (1996) décourage fortement son utilisation pour les raisons suivantes: il y a ambiguïté dans le choix des paramètres et l'interprétation de la carte est difficile.

Alors pour la carte cusum de type "tabular" et sous les hypothèses du Modèle 1.3.1, soit

$$\begin{aligned} C_j^+ &= \max[0, x_j - (\mu + K) + C_{j-1}^+] \\ C_j^- &= \max[0, (\mu - K) - x_j + C_{j-1}^-], \end{aligned} \quad (1.4.1)$$

où  $j = 1, \dots, m$ , avec les valeurs initiales  $C_0^+ = C_0^- = 0$ .

En fait,  $C_j^+$  ( $C_j^-$ ) représente l'accumulation des écarts entre  $\mu$  et les valeurs qui sont supérieures (inférieures) à  $\mu$ , pour autant que cet écart soit plus grand qu'un certain  $K$ .

Le paramètre  $K$  est une valeur de référence et il est fonction de l'écart type, c'est-à-dire  $K = k\sigma$ . Souvent on prendra  $K = \delta\sigma/2 = |\mu_1 - \mu|/2$  où  $\delta$  est le changement en unités d'écart type et  $\mu_1$  est la valeur de la moyenne correspondante à ce changement.

Si  $C_j^+$  ou  $C_j^-$  est plus grand qu'un certain  $H$ , on considère le procédé hors-contrôle. Encore ici, le paramètre  $H$  est fonction de  $\sigma$  et on a  $H = h_{cusum}\sigma$ . Le choix des paramètres  $H$  et  $K$  est important puisqu'il a un impact direct sur la performance de la carte cusum. Les actions qui doivent être entreprises suivant un signal sont les mêmes pour toutes les cartes de contrôle. On doit trouver la cause assignable responsable et l'éliminer. De plus, après un tel signal, on réinitialise la procédure cusum à 0.

Le graphique est construit en plaçant sur la carte les deux statistiques  $C_j^+$  et  $C_j^-$  en fonction de  $j$ , l'une au-dessus de la  $LC$  de zéro et l'autre au-dessous de celle-ci avec un intervalle de décision de  $H$ .

Puisque  $\mu$  et  $\sigma$  sont inconnus, on va les estimer à partir de nos échantillons. On recommande d'utiliser les mêmes estimateurs que pour la carte de Shewhart puisqu'ici encore on veut une idée de la variation à court terme. On utilisera donc

$$\hat{\mu} = \bar{x} = \frac{x_1 + \dots + x_m}{m} \quad \text{et} \quad \hat{\sigma} = \frac{\overline{Rm}}{d_{2;2}}, \quad (1.4.2)$$

où  $\overline{Rm} = (Rm_2 + \dots + Rm_m)/(m - 1)$ ,  $d_{2;2}=1,128$  et  $Rm_j$ ,  $j = 2, \dots, m$ , est défini à partir de l'équation (1.3.3).

On utilisera donc

$$\begin{aligned} \hat{C}_j^+ &= \max[0, x_j - (\hat{\mu} + K) + C_{j-1}^+] \\ \hat{C}_j^- &= \max[0, (\hat{\mu} - K) - x_j + C_{j-1}^-], \end{aligned} \quad (1.4.3)$$

où  $K = k\hat{\sigma}$  et il y a hors-contrôle si  $\hat{C}_j^+$  ou  $\hat{C}_j^-$  est plus grand que  $H = h_{cusum}\hat{\sigma}$ .

Avec les observations des exemples 1.1, 1.2 et 1.3 et les équations (1.4.3), on a pu construire des cartes cusum. Les estimés de  $\mu$  et de  $\sigma$  ont été calculés à partir des 30 échantillons préliminaires de la procédure de Shewhart. Les paramètres de la carte cusum sont  $h_{cusum}=1,72$  et  $k=1,5$ . Tout comme pour la procédure de Shewhart pour les données individuelles, on a construit une carte pour un déplacement de  $\delta=0$  (figure 1.4), une autre pour  $\delta=2\sigma$  (figure 1.6 (B), page 42) et une dernière pour un déplacement de  $\delta=3\sigma$  (figure 1.7 (B), page 43). Les observations sont celles de l'exemple 1.1 pour  $\delta=0$ , de l'exemple 1.2 pour  $\delta=2$  et finalement de l'exemple 1.3 pour  $\delta=3$ .

Pour  $\delta=0$ , on s'aperçoit qu'il n'y a aucun point hors-contrôle. Par contre, la procédure cusum a détecté un signal aux observations 4 et 7 pour le déplacement de  $2\sigma$  et aux observations 2, 4, 6, 8 et 9 pour le déplacement de  $3\sigma$ .

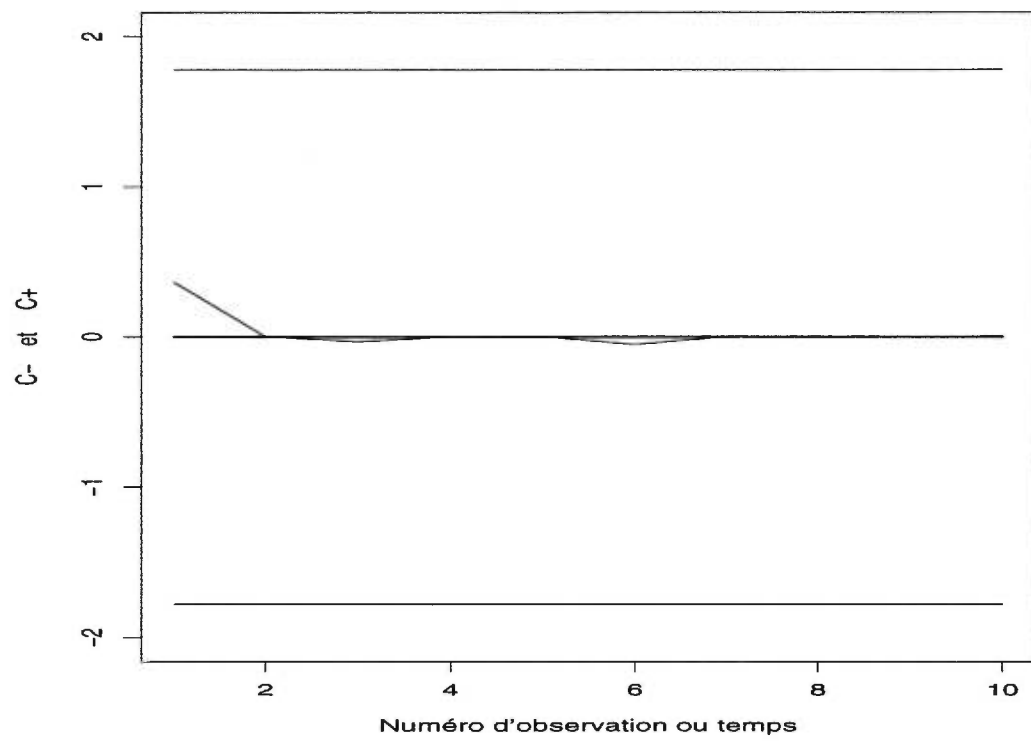


FIGURE 1.4. Carte de contrôle cusum pour 10 observations en contrôle ( $\delta=0$ ) avec  $h_{cusum}=1,72$  et  $k=1,5$  (phase deux et  $n=1$ ).

On peut noter ici une différence entre la carte de Shewhart et cusum. En effet, l'échelle utilisée sur les cartes n'est pas la même. La carte de Shewhart est centrée sur la moyenne du procédé tandis que la carte cusum est centrée sur la valeur 0. Cependant, il est facile de ramener la carte cusum avec une échelle centrée sur la moyenne en ajoutant à la statistique  $C_j$ , la valeur de la moyenne.

#### 1.4.2. *LMS* de la procédure cusum

Il existe plusieurs techniques pour calculer approximativement la *LMS* de la procédure cusum étant donné que le calcul exact est inabordable. Pour une formule simple et rapide à calculer on recommande l'approximation donnée par Hawkins (1992). Cette approximation est une application du modèle de Mandel (1969) et elle fait appel aux moindres carrés pondérés. Hawkins fournit deux tables, une pour les valeurs de  $h_{cusum}$  comprises entre 0 et 9 avec des intervalles de 0,5 et l'autre pour les valeurs de  $k$  comprises entre -2 et 3 avec des intervalles de 0,125. Pour un  $h_{cusum}$  fixé dans la table, il existe des paramètres  $\alpha_h$ ,  $\epsilon_h$  et  $\epsilon_h^*$  correspondants. De la même manière, pour un  $k$  fixé, il existe  $\beta_k$ ,  $\eta_k$  et  $\eta_k^*$  correspondants. À partir de ceux-ci, on calcule une statistique  $Y_{hk}$ , où

$$Y_{hk} = \alpha_h + \beta_k + \epsilon_h \eta_k + \epsilon_h^* \eta_k^*.$$

Enfin, la *LMS* est obtenue à l'aide de la fonction de répartition d'une loi  $N(0,1)$ . On a donc

$$LMS = \frac{1}{\Phi(-Y_{hk})}.$$

Les approximations des *LMS* sont généralement à l'intérieur de 3 % des vraies valeurs des *LMS* avec une moyenne pour les différences égale à -0,2 % avec un écart type de 0,7 %.

Alors avec le modèle des données de l'exemple 1.2, avec  $k=1$ ,  $h_{cusum}=2,52$  et  $\delta=2$ , on obtient avec l'approximation de Hawkins (1992),  $LMS_0=375$  et  $LMS_1=3,26$ .

### 1.4.3. Choix des paramètres

Le choix des paramètres  $K$  et  $H$  (et donc  $k$  et  $h_{cusum}$ ) de la carte cusum est très important car il a une incidence directe sur la performance de la carte. On recommande habituellement de les choisir en fonction de la  $LMS$  désirée. Il y a eu plusieurs études sur le choix des paramètres, entre autre par Hawkins (1993a).

D'après les études,  $h_{cusum}=4$  ou  $h_{cusum}=5$  et  $k=1/2$  va généralement donner de bons résultats pour la  $LMS$  pour un déplacement de  $1\sigma$  dans la moyenne du procédé.

Généralement, on choisira ces paramètres de façon à vouloir minimiser  $LMS_1$  pour détecter un déplacement d'amplitude  $\delta$  pour une  $LMS_0$  fixée. D'après les études, on s'est aperçu que l'on doit choisir  $k$  proportionnellement à l'amplitude du déplacement que l'on veut détecter. Le choix de  $k = \frac{1}{2}\delta$ , où  $\delta$  est l'amplitude du déplacement calculée à partir d'unités d'écart type, semble être optimal. Lorsque  $k$  est ainsi fixée, on choisit  $h_{cusum}$  en fonction de la  $LMS_0$  désirée.

Hawkins (1993a) a fourni une table pour les valeurs de  $k$  et  $h_{cusum}$  pour obtenir une  $LMS_0$  de 370 qui est égale à celle pour la carte de Shewhart lorsque  $L=3$ . Le tableau 1.6 présente donc les valeurs optimales de  $k$  et  $h_{cusum}$  à utiliser pour un déplacement  $\delta$  donné. Pour les valeurs de  $h_{cusum}$  choisies en fonction de d'autres  $LMS_0$ , voir l'article de Gan (1991).

Alors même si la carte cusum a d'abord été introduite pour détecter de petits changements dans la moyenne, elle s'ajuste facilement pour des changements plus larges. On doit également noter que contrairement à la carte de Shewhart

$\delta$	0,5	1,0	1,5	2,0	2,5	3,0
$k$	0,25	0,50	0,75	1,00	1,25	1,50
$h$	8,01	4,77	3,34	2,52	1,99	1,61

TABLEAU 1.6. Paramètres  $k$  et  $h_{cusum}$  de la carte cusum pour un déplacement  $\delta$  de la moyenne et pour une  $LM S_0$  égale à 370.

où un seul paramètre est présent, la carte cusum permet l'ajustement de deux paramètres.

#### 1.4.4. Autres caractéristiques

Il existe des ajustements qui ont été apportés à la procédure cusum pour augmenter sa performance. Parmi ceux-ci, il y a celui pour lutter contre l'inertie en période de démarrage. En effet, puisque la procédure est de type séquentiel, on peut manquer des changements dans la moyenne en période de démarrage puisque les deux statistiques sont égales à zéro. Il existe donc une procédure introduite par Lucas et Crosier (1982) pour augmenter la sensibilité en de telles circonstances. Essentiellement, les valeurs de départ  $C_0^+$  et  $C_0^-$  seront différentes de zéro, souvent d'une valeur de  $H/2$ .

Aussi, si la taille de chaque échantillon est de  $n > 1$ , on a qu'à remplacer  $x_j$  par  $\bar{x}_j$  et  $\sigma$  par  $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ .

### 1.5. LA CARTE EWMA

Nous allons maintenant introduire le deuxième type de carte nous permettant de détecter les petits déplacements plus facilement que la carte de Shewhart, il s'agit de la carte à moyenne mobile et à pondération exponentielle (de l'anglais *exponentially weighted moving average*). Cette carte est très semblable à la carte cusum pour la détection de petits déplacements ( $1,5 \sigma$  ou moins).



### 1.5.1. Principes de base

Cette carte a été introduite par Roberts (1959). Elle a été principalement construite pour des données individuelles et pour suivre la moyenne d'un procédé et non sa variation. On va donc se concentrer sur le cas de la moyenne d'un procédé avec  $n=1$ . Elle s'ajuste cependant facilement si  $n > 1$ .

Alors sous les hypothèses du modèle 1.3.1 avec  $n=1$ , la moyenne mobile à pondération exponentielle est définie comme suit,

$$z_j = rx_j + (1 - r)z_{j-1}, \quad j = 1, \dots, m, \quad (1.5.1)$$

où  $0 < r \leq 1$  est une constante et la valeur de départ est égale à la cible du procédé, c'est-à-dire  $z_0 = \mu$ . Encore une fois, on va estimer  $\mu$  par la moyenne échantillonnale  $\bar{x}$ . On peut également noter que si  $r=1$ , on se ramène à la carte de Shewhart.

L'équation 1.5.1 peut être également écrite sous la forme

$$z_j = r \sum_{t=0}^{j-1} (1 - r)^t x_{j-t} + (1 - r)^j z_0, \quad j = 1, \dots, m.$$

Puisque la procédure ewma peut être vue comme une moyenne pondérée de toutes les observations précédentes et présente, elle est robuste à la non normalité. Elle est donc une procédure à recommander si on a des données individuelles.

Si les observations  $x_j$  sont des variables aléatoires indépendantes de variance  $\sigma^2$ , la variance de  $z_j$  est

$$\sigma_{z_j}^2 = \sigma^2 \left( \frac{r}{2 - r} \right) (1 - (1 - r)^{2j}), \quad j = 1, \dots, m.$$

La carte ewma est construite en plaçant la valeur de la statistique  $z_j$  en fonction de  $j$ . La  $LC$  et les limites de contrôle sont construites ainsi,

$$\begin{aligned} LSC &= \mu + h_{ewma}\sigma\sqrt{\left(\frac{r}{2-r}\right)(1-(1-r)^{2j})} \\ LC &= \mu \\ LIC &= \mu - h_{ewma}\sigma\sqrt{\left(\frac{r}{2-r}\right)(1-(1-r)^{2j})}. \end{aligned} \tag{1.5.2}$$

Les limites sont donc variables, mais elles se stabilisent rapidement à

$$\mu \pm h_{ewma}\sigma\sqrt{\frac{r}{2-r}}.$$

On recommande cependant d'utiliser les limites de contrôle de l'équation (1.5.2). Les limites sont variables et plus petites au début du procédé car la variance est plus petite.

Si pour un point on a  $z_j > LSC$  ou  $z_j < LIC$ , il y a hors-contrôle et on doit chercher la cause assignable responsable de ce signal. Le choix des paramètres  $h_{ewma}$  et  $r$  dépend encore une fois de la  $LMS$  désirée.

Encore une fois, on devra estimer  $\sigma$  à partir des échantillons. On recommande de l'estimer à l'aide des étendues mobiles, puisque l'on veut une estimation de la variation à court terme. On utilisera donc  $\hat{\sigma} = \overline{Rm}/d_{2,2}$ . Les limites de contrôle à utiliser seront donc

$$\begin{aligned} LSC &= \hat{\mu} + h_{ewma}\hat{\sigma}\sqrt{\left(\frac{r}{2-r}\right)(1-(1-r)^{2j})} \\ LC &= \hat{\mu} \\ LIC &= \hat{\mu} - h_{ewma}\hat{\sigma}\sqrt{\left(\frac{r}{2-r}\right)(1-(1-r)^{2j})}. \end{aligned} \tag{1.5.3}$$

On est donc prêt à construire un exemple de carte ewma. On prend les mêmes observations que pour les cartes de Shewhart et cusum. Nos estimés de

$\mu$  et  $\sigma$  sont donc 10,01 et 1,03 respectivement. Les paramètres de la procédure ewma sont  $r=0,75$  et  $h_{ewma}=3,087$ . Encore une fois, on a testé trois types de déplacement. Un déplacement de  $\delta=0$  (figure 1.5), un de  $\delta=2$  (figure 1.6 (C), page 42) et finalement un dernier pour  $\delta=3$  (figure 1.7 (C), page 43). Pour  $\delta=0$ , il n'y a aucun point hors-contrôle (tout comme pour Shewhart et cusum). Pour  $\delta=2$ , il y a hors-contrôle aux observations 4 et 7 et pour  $\delta=3$ , il y a hors-contrôle aux observations 2, 4, 5, 6, 8 et 9.

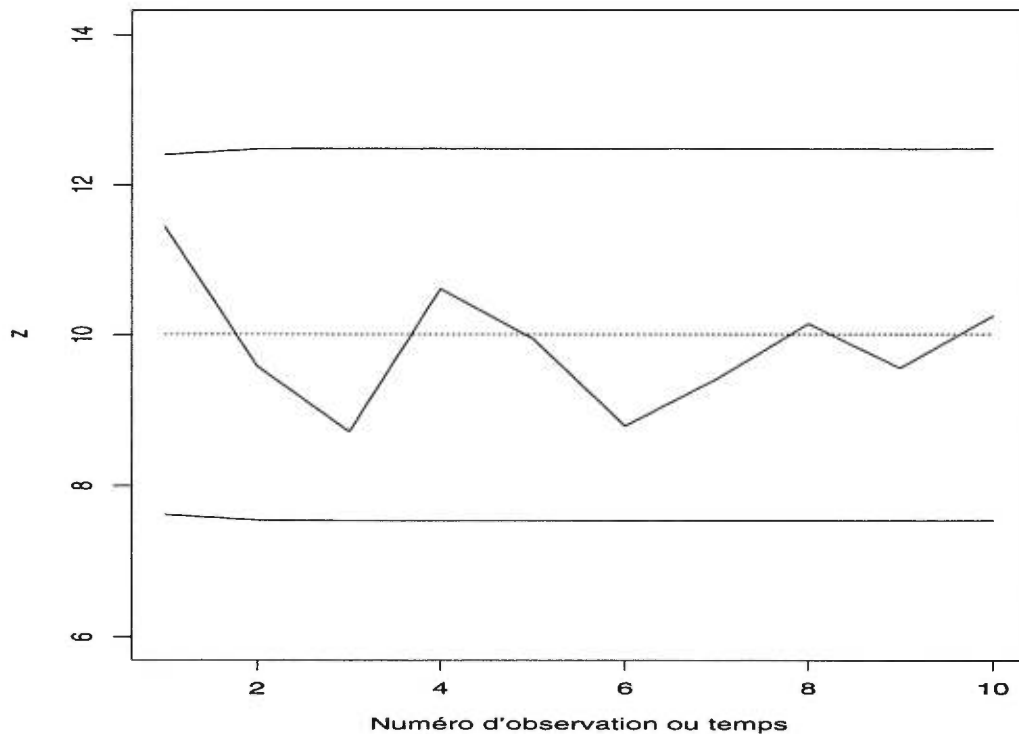


FIGURE 1.5. Carte de contrôle ewma pour 10 observations en contrôle ( $\delta=0$ ) avec  $r=0,75$  et  $h_{ewma}=3,087$  (phase deux et  $n=1$ ).

Sur la carte ewma, on peut s'apercevoir que contrairement à la carte cusum, l'échelle utilisée est centrée non pas sur la valeur 0, mais sur la moyenne du procédé tout comme la carte de Shewhart.

### 1.5.2. *LMS* de la procédure ewma

Il existe plusieurs façons de calculer la *LMS* de la carte ewma. Une de celles-ci consiste à représenter la statistique ewma comme une chaîne de Markov. On peut ainsi faire l'approximation des propriétés de la carte ewma par une chaîne de Markov discrète suivant une procédure similaire à celle employée par Brook et Evans (1972) dans le cas de la carte cusum. Pour une discussion détaillée de l'approche de la chaîne de Markov, voir Lucas et Saccucci (1990).

### 1.5.3. Choix des paramètres

Tout comme pour la procédure cusum, le choix des paramètres  $h_{ewma}$  et  $r$  de la procédure ewma est très important puisqu'il a une influence directe sur la performance de la carte. Il est possible de les choisir pour obtenir une performance de la *LMS* qui sera semblable à celle de cusum pour la détection de petits déplacements.

Il y a eu plusieurs études à propos du choix de ces paramètres, entre autre Lucas et Saccucci (1990). La procédure optimale consiste à fixer la  $LMS_0$  et l'amplitude du déplacement que l'on veut détecter et par la suite on sélectionne les paramètres  $h_{ewma}$  et  $r$  pour obtenir une  $LMS_1$  minimale. Une bonne règle est de choisir des plus petites valeurs de  $r$  pour détecter de plus petits déplacements.

Le tableau 1.7 donne la procédure optimale lorsque l'on est intéressé à détecter un déplacement de  $\delta$ -sigma dans la moyenne d'un procédé. On donne les valeurs des paramètres à choisir pour deux  $LMS_0$  différentes, soit de 300 et 500 observations lorsque le procédé est en contrôle. De plus, puisque la valeur optimale de  $LMS_1$  est à peu près équivalente sur un intervalle de valeur de  $r$  et de  $h_{ewma}$ , on a fourni l'intervalle de ces valeurs. Une procédure ewma construite à partir de la plus petite valeur de ces intervalles va donner plus de protection contre les petits

déplacements alors que la plus grande valeur de ces intervalles va donner plus de protection contre les gros déplacements. Les  $LMS$  de cette table ont été calculées à partir de la méthode approximative de la chaîne de Markov et proviennent de Lucas et Saccucci (1990).

$LMS_0$		$\delta$				
		0,5	1,0	2,0	3,0	4,0
300	$r$	0,05-0,06	0,14-0,15	0,38-0,42	0,71-0,74	0,84-0,97
	$h$	2,462-2,399	2,723-2,707	2,895-2,885	2,931-2,930	2,935-2,934
	$LMS_1$	24,9	9,14	3,23	1,72	1,16
500	$r$	0,05	0,12-0,15	0,36-0,37	0,66-0,70	0,82-0,95
	$h$	2,616	2,858-2,907	3,044-3,047	3,084-3,086	3,089-3,090
	$LMS_1$	28,7	10,2	3,51	1,86	1,21

TABLEAU 1.7. Paramètres  $r$  et  $h_{ewma}$  de la carte ewma et la  $LMS_1$  correspondante pour un déplacement  $\delta$  de la moyenne pour une  $LMS_0$  fixée à 300 et 500.

Alors tout comme cusum, on voit que même si la carte a d'abord été introduite pour détecter de petits déplacements, elle s'ajuste facilement si on est intéressé à des déplacements plus larges. On doit noter que cette carte, tout comme cusum, permet l'ajustement de deux paramètres contrairement à la carte de Shewhart qui n'a qu'un seul paramètre.

#### 1.5.4. Autres caractéristiques

Tout comme pour la carte cusum, il existe une procédure pour réagir plus rapidement en période de démarrage. Cette procédure sera principalement appliquée si  $r < 0,25$ . On l'obtient en implantant simultanément deux statistiques unilatérales ewma, chacune avec une valeur de départ différente. C'est-à-dire que l'on remplacera la valeur de départ  $z_0$  par une valeur de départ plus basse que la cible pour la statistique  $z_j^-$  et par une valeur plus haute que la cible pour la statistique  $z_j^+$ . S'il y a hors-contrôle au départ, une des deux statistiques va réagir plus rapidement. Si le procédé est en contrôle, les deux statistiques vont

converger et lorsqu'elles seront suffisamment proches, on pourra en laisser tomber une. Pour plus de détails, voir Lucas et Saccucci (1990).

Un dernier point à souligner pour ewma c'est que celle-ci peut être utilisée pour prédire où la moyenne du procédé sera à la prochaine période. En fait,  $z_j$  est actuellement une prévision de la valeur de la moyenne du procédé au temps  $j + 1$ . Pour des explications plus détaillées, voir Montgomery (1996).

Aussi, si la taille de chaque échantillon est de  $n > 1$ , on a qu'à remplacer  $x_j$  par  $\bar{x}_j$  et  $\sigma$  par  $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ .

## 1.6. COMPARAISON DES TROIS CARTES

Cette section nous permettra d'établir une comparaison entre les trois types de cartes que nous venons de voir. Nous établirons deux types de comparaisons, une qualitative à partir des caractéristiques des cartes et l'autre quantitative à partir des *LMS*.

Le tableau 1.8 nous permet de bien visualiser les comparaisons qualitatives que l'on peut établir entre les trois cartes. On a comparé les caractéristiques suivantes: le nombre de paramètres, si la statistique est facilement interprétable, si la statistique utilise la séquence de points entière, si la procédure détecte bien les petits ( $< 1,5\sigma$ ) et les gros ( $> 3\sigma$ ) déplacements.

Caractéristique	Shewhart	cusum	ewma
paramètres	1	2	2
facilement interprétable	oui	non	non
séquence entière	non	oui	oui
petits déplacements	oui*	oui	oui
gros déplacements	oui	oui	oui

TABLEAU 1.8. Comparaisons qualitatives des cartes de Shewhart, cusum et ewma.

Cette comparaison qualitative nous permet de bien cerner certains des avantages et désavantages des cartes de Shewhart, cusum et ewma. L'avantage majeur

de la carte de Shewhart réside au niveau de l'interprétation de la statistique. En effet, c'est la valeur de l'observation (pour des données individuelles) ou la moyenne d'un échantillon (pour des données groupées) qui est placée sur un graphique et non pas une statistique cumulée comme pour cusum et ewma. Pour cusum et ewma, la valeur obtenue est accumulée et c'est la statistique cumulée qui nous permet de juger si le procédé est en contrôle. Il est donc plus difficile à partir de cette statistique seule d'avoir une idée si le procédé est en contrôle ou non puisque son interprétation n'est pas directe.

L'observation 7 de la figure 1.6, qui est hors-contrôle pour les trois cartes, illustre bien ce fait. Les valeurs sur l'axe vertical correspondant à la 7-ième observation sont, pour les cartes de Shewhart, cusum et ewma 14,06, 3,13 et 13,46 respectivement. Alors même si on n'a pas les limites de contrôle pour la carte de Shewhart, l'utilisateur peut se douter que le procédé est hors-contrôle puisqu'il sait si une valeur de 14,06 est acceptable ou non pour cette variable. Par contre, sans les limites de contrôle de la carte cusum, on ne peut avoir une idée si le procédé est en contrôle ou non, car une valeur de 3,13 n'a aucune signification pour l'utilisateur. La procédure ewma se situe quelque part entre Shewhart et cusum en ce qui a trait à son interprétation. En effet, une valeur de 13,46 nous indique que la moyenne a certainement augmentée, mais cette valeur n'est pas celle de la variable elle-même comme pour Shewhart, elle représente plutôt une valeur accumulée à partir des observations précédentes. Il est donc plus difficile pour l'utilisateur de savoir si 13,46 est acceptable ou non pour une valeur accumulée.

Par contre, l'avantage majeur des cartes cusum et ewma c'est qu'elles permettent l'ajustement de deux paramètres. Cela permet donc à ces deux cartes de détecter assez facilement les gros et les petits déplacements. De plus, leurs statistiques utilisent la séquence de points entière ce qui permet de détecter les

petits déplacements beaucoup plus rapidement que la carte de Shewhart puisque les petits effets s'accumulent.

On doit cependant mentionner que l'on a ajouté un astérisque pour la carte de Shewhart en ce qui a trait à la détection des petits déplacements, car il est possible, en ajoutant des critères de décision supplémentaires, de détecter les petits déplacements avec la carte de Shewhart. Cependant, cet ajout fait en sorte que la  $LMS_0$  est modifiée.

La comparaison quantitative a été effectuée à l'aide des  $LMS$ . On a comparé les  $LMS_1$  des trois cartes pour une  $LMS_0$  fixée à 500. On a pris une  $LMS_0=500$  car pour la procédure ewma, où aucune approximation n'est fournie, on a tiré les résultats directement de l'article de Lucas et Saccucci (1990), où les  $LMS_1$  ont été calculées seulement pour des  $LMS_0$  égales à 300 et 500. Puisque les cartes cusum et ewma permettent l'ajustement de 2 paramètres, il nous est possible d'obtenir plus d'une procédure qui correspond à une  $LMS_0$  de 500. On a donc choisi 3 procédures pour chacune des cartes qui sont optimales pour des déplacements de la moyenne de 1, 2 et 3  $\sigma$ . On les note respectivement cusum1 et ewma1, cusum2 et ewma2, cusum3 et ewma3.

Pour cusum, les paramètres ont été choisis à partir de la procédure décrite à la section 1.4.3 et à l'aide de l'article de Gan (1991). Les  $LMS$  ont été calculées à partir de la méthode de Hawkins (1992) et sont donc approximatives. La valeur avec un astérisque provient de simulations puisque la table fournie par Hawkins (1992), ne contient pas toutes les valeurs possibles pour  $h_{cusum}$  et  $k$ .

Pour ewma, les paramètres ont été choisis à partir de l'article de Lucas et Saccucci (1990). Les  $LMS$  sont également tirées de cet article et sont calculées approximativement à partir de la méthode de la chaîne de Markov. Le tableau 1.9 donne les  $LMS_1$  correspondant à un déplacement  $\delta$  pour une  $LMS_0=500$ .



$\delta$	Shewhart	cusum1	cusum2	cusum3	ewma1	ewma2	ewma3
	$L=3,09$	$k=0,5$ $h=5,08$	$k=1$ $h=2,66$	$k=1,5$ $h=1,72$	$h=2,814$ $r=0,1$	$h=3,054$ $r=0,4$	$h=3,087$ $r=0,75$
0	500	510	494	513	500	500	500
0,5	201	38,5	80,0	141	31,3	71,2	140
1,0	54,6	10,5	14,6	26,9	10,3	14,3	30,6
2,0	7,26	4,06	3,40	3,89	4,36	3,52	4,54
3,0	2,15	2,61*	1,93	1,78	2,87	2,02	1,88

TABLEAU 1.9.  $LMS_1$  des cartes de Shewhart, cusum et ewma pour une  $LMS_0=500$  et un déplacement  $\delta$  de la moyenne (la valeur avec un astérique provient de simulations).

On peut tirer deux conclusions à partir de ce tableau. La première est que si on compare la procédure de Shewhart avec cusum3 et ewma3, on s'aperçoit que les  $LMS_1$  sont toujours plus petites pour cusum3 et ewma3 que pour Shewhart. Il n'y a donc aucune raison de préférer la procédure Shewhart à celles de cusum ou ewma si on se fie uniquement à la performance des cartes.

De plus, si on choisit une procédure qui est optimale pour les petits déplacements (cusum1 ou ewma1), on s'aperçoit que les  $LMS_1$  pour les petits déplacements ont considérablement diminuées (en comparaison avec les procédures de Shewhart, cusum3 et ewma3), alors que pour un déplacement de  $\delta=3$ , elle augmente mais de façon très minime (moins de une observation supplémentaire). On s'aperçoit donc qu'il est possible de considérablement diminuer la  $LMS_1$  pour les petits déplacements sans affecter de manière significative la  $LMS_1$  pour les gros déplacements. Il paraît donc tout à notre avantage d'utiliser une procédure qui est optimale pour les petits déplacements, comme cusum1 et ewma1.

On sait que pour la procédure de Shewhart, il existe des critères de décision supplémentaires pour diminuer la  $LMS_1$ . Cependant, ces critères, tout en diminuant la  $LMS_1$ , vont faire diminuer également la  $LMS_0$ . Il deviendra donc difficile avec ces critères de fixer  $LMS_0$  à une valeur souhaitée puisque la procédure de Shewhart utilise un seul paramètre. Pour une discussion plus détaillée de ces critères, voir Montgomery (1996).

Finalement, les figures 1.6 et 1.7 nous permettent de bien visualiser les différences graphiques entre les 3 cartes. Comme on l'a déjà mentionné, ces figures représentent un procédé provenant d'une loi  $N(10, 1)$ . On a d'abord calculé les estimés de  $\mu$  et de  $\sigma$  à partir de 30 observations en contrôle et par la suite, on a simulé 10 nouvelles observations ayant subi un déplacement. Les observations de la figure 1.6 ont subi un déplacement de  $\delta=2\sigma$  et celles de la figure 1.7, un déplacement de  $\delta=3\sigma$ . Il faut mentionner que chacune de ces cartes représentent un seul jeu de données et que l'on n'obtiendra pas toujours les mêmes résultats avec d'autres jeux de données. Cependant, après avoir simulé plusieurs jeux de données, on se rend compte que les conclusions que l'on va tirer à partir des graphiques restent les mêmes.

On note d'abord la grande similitude entre les cartes cusum et ewma. Car en plus de détecter les points hors-contrôle aux mêmes endroits, les graphiques des cartes sont sensiblement les mêmes.

Pour la carte de Shewhart, on note tout d'abord que pour  $\delta=2$ , le temps requis avant de détecter un point hors-contrôle (à  $t=7$ ) est plus long que pour cusum et ewma (à  $t=4$ ) alors que pour  $\delta=3$ , on obtient le même temps pour les trois cartes soit 2 observations. On remarque également que le graphique de la carte de Shewhart se rapproche très fortement des deux autres cartes.

Alors en conclusion, on peut dire que même si la carte de Shewhart est très utilisée en pratique, vu sa simplicité et son interprétation (et aussi pour des raisons historiques), il vaut la peine de porter une attention particulière aux deux autres types de cartes, cusum et ewma, puisque leurs performances dépassent celle de Shewhart. Le choix de la carte dépendra de l'utilisation que l'on veut en faire, des priorités de l'utilisateur et des ressources mises à notre disposition car chacune des trois cartes présentent des avantages et inconvénients. À l'utilisateur de bien cerner ce qu'il veut vraiment.

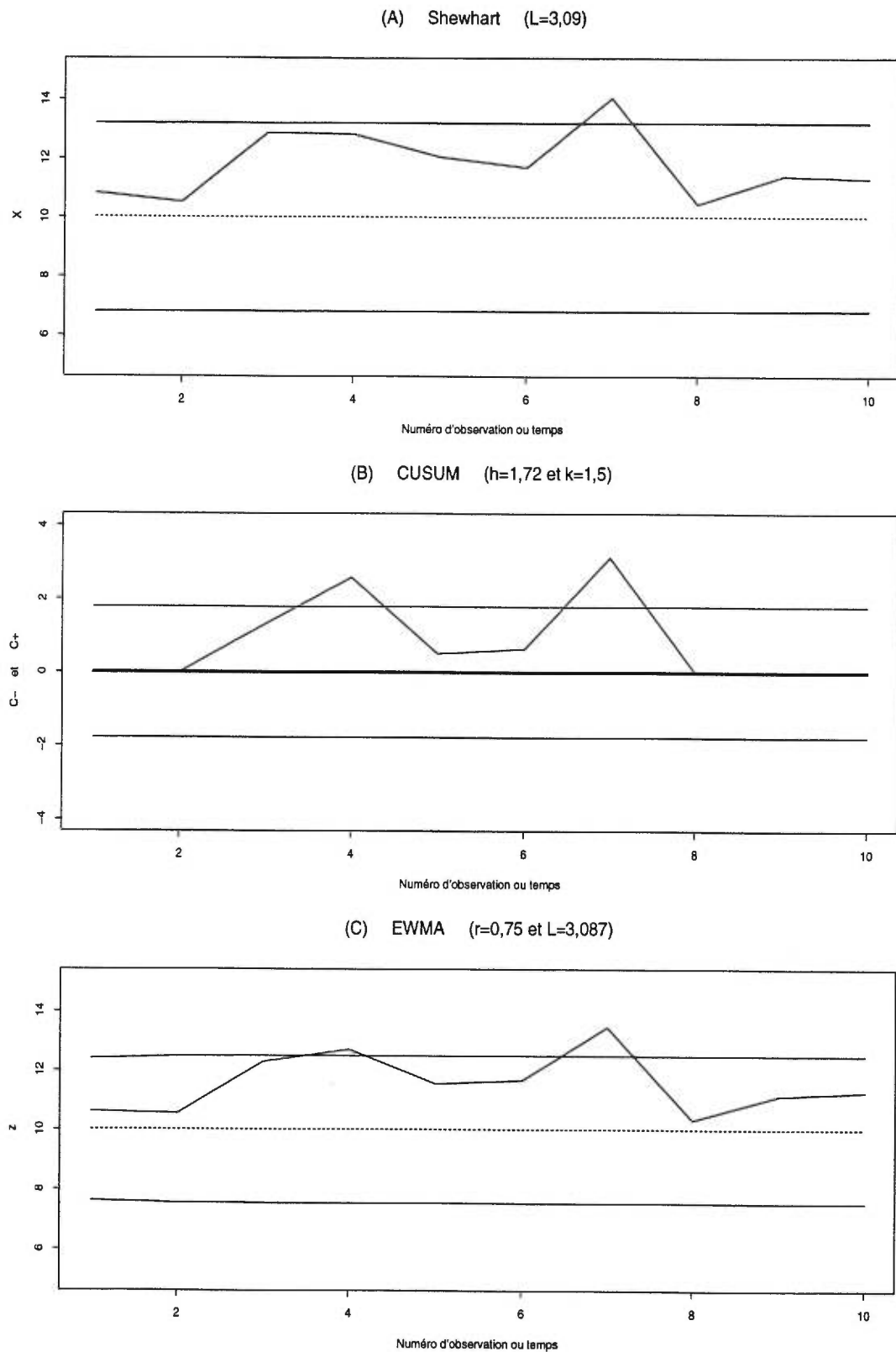


FIGURE 1.6. Cartes de contrôle de Shewhart, cusum et ewma pour 10 observations ayant subi un déplacement de  $\delta=2$  (phase deux et  $n=1$ ).

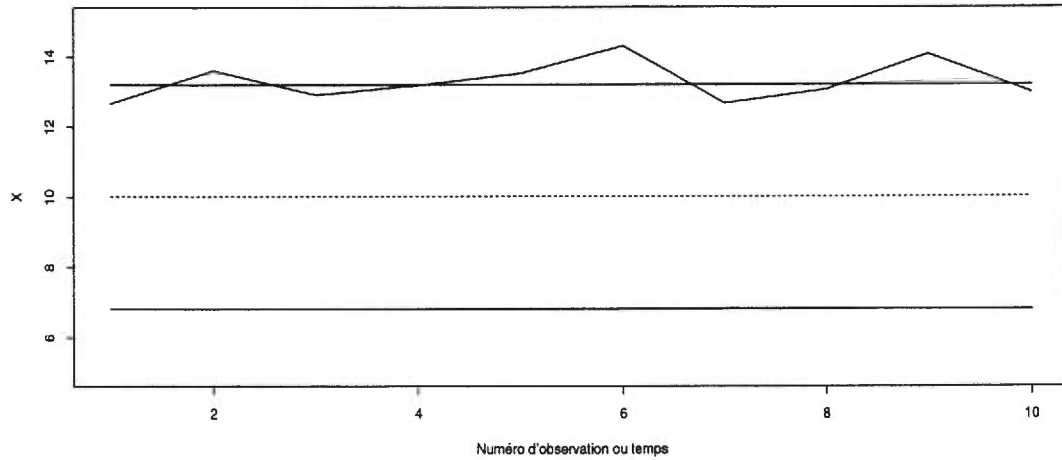
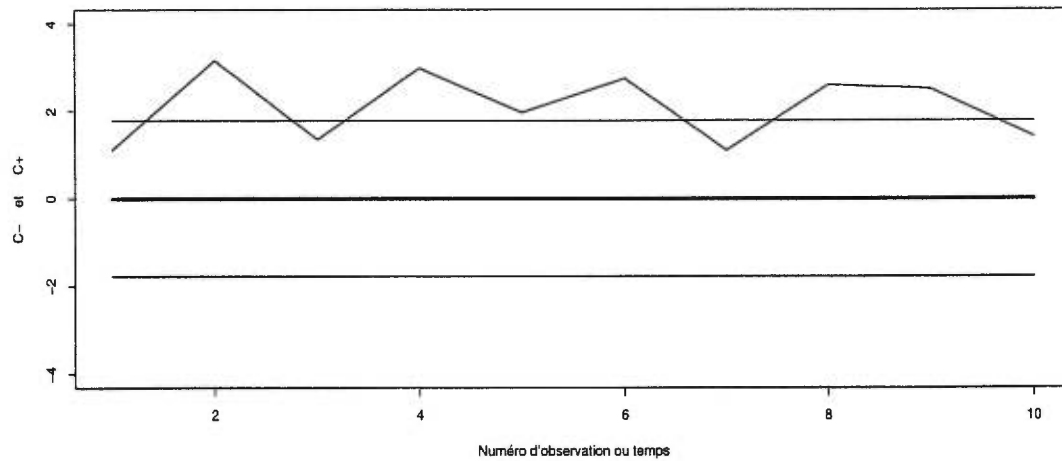
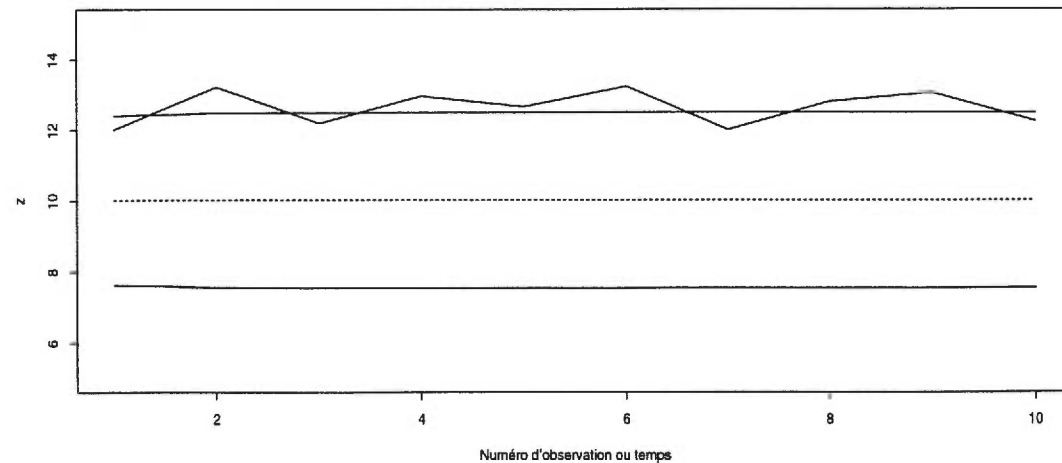
(A) Shewhart ( $L=3,09$ )(B) CUSUM ( $h=1,72$  et  $k=1,5$ )(C) EWMA ( $r=0,75$  et  $L=3,087$ )

FIGURE 1.7. Cartes de contrôle de Shewhart, cusum et ewma pour 10 observations ayant subi un déplacement de  $\delta=3$  (phase deux et  $n=1$ ).

## 1.7. AUTRES CARACTÉRISTIQUES

Dans cette dernière section, on abordera trois points que l'on n'a pas encore traités. On parlera de l'effet de la non normalité, de l'effet des données auto-corrélées et de l'utilisation des données multivariées.

### 1.7.1. Données non normales

Les trois types de cartes présentées précédemment ont été construites sous les hypothèses du Modèle 1.3.1, c'est-à-dire en supposant des données normales. Cependant, dans plusieurs situations, on aura raison de douter de la validité de cette hypothèse.

Pour les cartes de Shewhart, la non-normalité a peu d'effet sur sa performance. En effet, si le choix du  $L=3$  est un choix très utilisé en pratique c'est justement pour ne pas diminuer la performance de la carte même si les données ne sont pas normales. Car quelle que soit la distribution des données, le choix de  $L=3$  nous assurera que lorsqu'un point est hors-contrôle, on est pratiquement certain qu'une cause assignable est associée à ce signal. Les fausses alarmes seront donc très rares. Pour une bonne discussion à ce sujet, voir Wheeler et Chambers (1992).

### 1.7.2. Données auto-corrélées

Les hypothèses du Modèle 1.3.1 impliquaient également que les données étaient indépendantes. Alors si les données sont auto-corrélées, cela a un effet sur la performance des cartes.

Une approche utilisée pour travailler avec de telles données est de modéliser la procédure à l'aide d'un modèle de série chronologique. On utilisera ce modèle

pour éliminer l'auto-corrélation des données et ensuite on appliquera les cartes de contrôle aux résidus. Pour un résumé de cette technique, voir Montgomery (1996).

### **1.7.3. Données multivariées**

Le dernier point à soulever dans cette section, est le fait que très souvent en pratique, on aura à travailler avec des données multivariées. En effet, dans un procédé, il est fort probable que des données sur plusieurs variables soient prises simultanément. Ces variables sont reliées entre elles et les cartes de contrôle univariées sur chacune des variables sont donc inappropriées puisqu'elles ne tiennent pas compte de la corrélation entre les variables. On aura donc besoin du contrôle statistique des procédés multivariés qui sera présenté au chapitre deux.

## CHAPITRE 2

---

# CONTRÔLE STATISTIQUE DES PROCÉDÉS MULTIVARIÉS

Dans ce deuxième chapitre, on parlera du contrôle statistique des procédés multivariés (CSPM). La structure sera semblable à celle du chapitre un. On commencera par une introduction du CSPM. Par la suite, on présentera la carte  $T^2$ , puis la carte à somme cumulée multivariée et finalement la carte à moyenne mobile et à pondération exponentielle multivariée. On établira également une comparaison entre ces trois types de cartes. La section six nous permettra de faire un survol des autres cartes et techniques existantes au niveau multivarié. Finalement, on verra les méthodes pour détecter les variables hors-contrôle pour un procédé multivarié.

### 2.1. INTRODUCTION

Comme on l'a mentionné à la fin du chapitre un, le contrôle statistique des procédés multivariés (CSPM) a été introduit parce que très souvent la performance d'un produit est déterminée en fonction de plusieurs caractéristiques de qualité. De plus, en raison des nouvelles technologies, beaucoup de données sont disponibles sur plusieurs variables. On a donc recours au CSPM. Le but premier est de tenir compte de la corrélation qui existe entre les variables.

Il survient dès lors un problème sur l'utilisation simultanée de ces données. Soit un procédé décrit à l'aide de  $p$  caractéristiques de qualité. Il existe plusieurs façons univariées de tester l'hypothèse que le procédé est en contrôle. Une première est d'appliquer la procédure à chacune des  $p$  variables séparément et de rejeter si l'une des procédures mène à un rejet de l'hypothèse de contrôle. Alors, sous l'hypothèse d'indépendance entre les variables, l'erreur globale de type 1 devient  $1 - (1 - \alpha)^p$ . Une deuxième solution est de rejeter l'hypothèse lorsque toutes les procédures sont rejetées. L'erreur globale de type 1, sous l'hypothèse d'indépendance entre les variables, devient alors  $\alpha^p$ . Plusieurs autres possibilités existent.

Pour contrer ce problème, on pourrait modifier l'erreur globale en changeant la valeur de  $\alpha$ . Cependant cette façon de faire n'est pas évidente puisque dans la majorité des cas les variables ne sont pas indépendantes. On doit donc appliquer une procédure multivariée pour tenir compte de la corrélation entre les variables.

Le premier à avoir introduit le CSPM fut Hotelling (1947). Dans la grande majorité des cas une procédure multivariée sert uniquement à suivre la moyenne d'un procédé et non sa variation. Nous nous concentrerons donc uniquement sur cet aspect dans le présent chapitre. Dans les trois prochaines sections nous verrons les versions multivariées des trois cartes présentées au chapitre un, c'est-à-dire la carte  $T^2$ , la carte cusum multivariée et la carte ewma multivariée.

## 2.2. LA CARTE $T^2$

La carte  $T^2$  est la version multivariée de la carte de Shewhart au niveau univarié. Elle a été introduite par Hotelling (1947). Le principe est le même que pour la carte de Shewhart. On calcule une statistique pour chaque observation donnée, la statistique  $T^2$ , et on compare celle-ci avec une limite supérieure de contrôle pour vérifier si le procédé est en contrôle. Cette statistique tiendra



compte de la corrélation qui existe entre les  $p$  variables du procédé par la matrice de variance-covariance. La présente section, est principalement inspirée de l'article de Wierda (1994).

### 2.2.1. Modèle théorique

Avant d'introduire la carte de contrôle proprement dite, on va présenter la théorie sur laquelle la carte  $T^2$  de Hotelling s'appuie. De façon générale, un vecteur  $x$  sera dénoté par  $\underline{x}$  et une lettre majuscule sera utilisée pour dénoter une matrice.

Le théorème 2.2.1 est un résultat fondamental en analyse multivariée (voir Johnson et Wichern, 1992)

**THÉORÈME 2.2.1.** *Soit  $Q^2 = \underline{u}'S^{-1}\underline{u}$  où  $\underline{u}$  est un vecteur aléatoire et  $S$  une matrice aléatoire indépendante de  $\underline{u}$  avec  $\underline{u} \sim N_p(\underline{\mu}, \Sigma)$  et  $fS \sim W_p(f, \Sigma)$  où  $f \geq p$ ,  $N_p$  est une loi normale multivariée de dimension  $p$  et  $W_p$  est une loi de Wishart. Alors*

$$Q^2 \sim \frac{fp}{f-p+1} F(p, f-p+1; \tau^2),$$

où le paramètre de décentralité de cette loi de Fisher est  $\tau^2 = \underline{\mu}'\Sigma^{-1}\underline{\mu}$ .

Aussi, si on a un échantillon  $\underline{x}_1, \dots, \underline{x}_n$  de loi  $N_p(\underline{\mu}, \Sigma)$ , alors on sait que la moyenne échantillonnale  $\underline{\bar{x}}$  et la matrice de covariance échantillonnale  $S$ , définie par

$$S = \frac{1}{n-1} \sum_{i=1}^n (\underline{x}_i - \underline{\bar{x}})'(\underline{x}_i - \underline{\bar{x}}),$$

sont un vecteur et une matrice aléatoires indépendants de distribution

$$\underline{\bar{x}} \sim N_p(\underline{\mu}, (1/n)\Sigma) \quad \text{et} \quad (n-1)S \sim W_p(n-1, \Sigma).$$

Le résultat suivant est une application directe du théorème 2.2.1. Soit  $\underline{\mu}_0$ , un vecteur postulé comme étant la moyenne du procédé. Alors

$$T^2 = n(\underline{\bar{x}} - \underline{\mu}_0)' S^{-1} (\underline{\bar{x}} - \underline{\mu}_0)$$

suit une distribution

$$\frac{(n-1)p}{n-p} F(p, n-p; \tau^2),$$

où  $\tau^2 = n(\underline{\mu} - \underline{\mu}_0)' \Sigma^{-1} (\underline{\mu} - \underline{\mu}_0)$  et  $\underline{\mu}$  est le véritable vecteur moyen du procédé.

Le résultat permet de tester l'hypothèse  $H_0 : \underline{\mu} = \underline{\mu}_0$  contre  $H_1 : \underline{\mu} \neq \underline{\mu}_0$ . Le test  $T^2$  de niveau  $\alpha$  est de rejeter  $H_0$  si

$$T^2 > \frac{(n-1)p}{n-p} F_\alpha(p, n-p)$$

où  $F_\alpha(\nu_1, \nu_2)$  est la valeur critique au niveau  $\alpha$  d'une distribution Fisher de paramètres  $\nu_1$  et  $\nu_2$ . Ce test est le test du rapport de vraisemblance.

On va donc appliquer cette théorie à la carte de contrôle  $T^2$  dans la prochaine sous-section.

### 2.2.2. Application à la carte $T^2$

Pour la construction de la carte  $T^2$ , on se doit de distinguer deux situations possibles. La première concerne l'analyse des données passées (que l'on appellera phase un) et la deuxième concerne l'analyse des données futures (que l'on appellera phase deux). On estimera la moyenne et la variance de notre procédé à partir des observations de la phase un et on se servira de ces estimés pour les deux phases. La distribution de la statistique  $T^2$  sera dépendante de la phase dans laquelle on se trouve puisque la distribution de la différence entre une observation et une moyenne empirique change selon que l'observation est utilisée dans le calcul de la moyenne ou est indépendante de celle-ci.

On va tout d'abord introduire la carte  $T^2$  pour des données groupées et par la suite on va appliquer la procédure au cas où  $n=1$ , c'est-à-dire pour des données individuelles.

On définit le modèle suivant:

Modèle 2.2.1. *Soit un procédé de fabrication caractérisé par  $p$  mesures de qualité  $x_1, \dots, x_p$  et pour lequel on a recueilli  $m$  échantillons indépendants chacun contenant  $n$  observations indépendantes. On dénote alors  $\underline{x}_{ji}$  le  $i$ -ième vecteur de dimension  $p$  du  $j$ -ième échantillon,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ , où  $\underline{x}_{ji} \sim N_p(\underline{\mu}_j, \Sigma)$ ,  $\underline{\mu}_j$  et  $\Sigma$  inconnus. Ainsi  $x_{lji}$ , représente l'observation du vecteur  $\underline{x}_{ji}$  correspondant à la variable  $l$ ,  $l = 1, \dots, p$ .*

Alors pour les  $m$  échantillons aléatoires de taille  $n > p$  de loi  $N_p(\underline{\mu}_j, \Sigma)$  recueillis à la phase un, on veut tester  $H_{0j} : \underline{\mu}_j = \underline{\bar{\mu}}$ ,  $j = 1, \dots, m$  et où  $\underline{\bar{\mu}} = (\underline{\mu}_1 + \dots + \underline{\mu}_m)/m$ . Pour ce faire, on va calculer la moyenne échantillonnale et la matrice de covariance échantillonnale des  $m$  échantillons. Le vecteur moyen de dimension  $p$  pour le  $j$ -ième échantillon est donné par

$$\bar{\underline{x}}_j = \frac{\underline{x}_{j1} + \dots + \underline{x}_{jn}}{n}, \quad j = 1, \dots, m, \quad (2.2.1)$$

et sa matrice de covariance est obtenue à l'aide de la formule

$$S_j = \frac{1}{n-1} \sum_{i=1}^n (\underline{x}_{ji} - \bar{\underline{x}}_j)' (\underline{x}_{ji} - \bar{\underline{x}}_j), \quad j = 1, \dots, m. \quad (2.2.2)$$

La moyenne globale échantillonnale est donnée par

$$\bar{\bar{\underline{x}}} = \frac{\bar{\underline{x}}_1 + \dots + \bar{\underline{x}}_m}{m}, \quad (2.2.3)$$

De plus, un estimateur sans biais pour  $\Sigma$  qui décrit la variation interne du procédé est

$$\bar{S} = \frac{S_1 + \dots + S_m}{m}. \quad (2.2.4)$$

Avant d'aller plus loin, il serait bon d'expliquer la notation qui sera utilisée tout au long de cette section. La statistique  $T^2$  et la limite de contrôle pour des données groupées de la phase un seront dénotées respectivement par  $T_{n;j}^2$  et  $LSC_{n;j}$ . Tandis qu'on utilisera  $T_{1;j}^2$  et  $LSC_{1;j}$  pour le cas des données individuelles de la phase un. Aussi, on dénotera la statistique pour les données groupées de la phase deux par  $T_{n;f}^2$  et pour les données individuelles par  $T_{1;f}^2$ . De la même façon, la limite de contrôle pour les données groupées de la phase deux sera dénotée par  $LSC_{n;f}$ , alors que pour les données individuelles ce sera  $LSC_{1;f}$ .

Alors par le théorème 2.2.1, on a que

$$T_{n;j}^2 = n(\bar{x}_j - \bar{\bar{x}})' \bar{S}^{-1} (\bar{x}_j - \bar{\bar{x}}), \quad (2.2.5)$$

suit une distribution

$$\frac{p(m-1)(n-1)}{m(n-1) - p + 1} F(p, m(n-1) - p + 1; \tau_{n;j}^2), \quad (2.2.6)$$

où  $\tau_{n;j}^2 = \frac{mn}{m-1} (\underline{\mu}_j - \underline{\bar{\mu}})' \Sigma^{-1} (\underline{\mu}_j - \underline{\bar{\mu}})$ ,  $j = 1, \dots, m$ .

Par la suite, on place sur un graphique la valeur de  $T_{n;j}^2$  en fonction du numéro d'observation ou du temps avec une limite supérieure de contrôle de

$$LSC_{n;j} = \frac{p(m-1)(n-1)}{m(n-1) - p + 1} F_{\alpha}(p, m(n-1) - p + 1). \quad (2.2.7)$$

On peut maintenant passer à la phase deux en faisant l'hypothèse que l'on a un échantillon en contrôle obtenu de la phase un. On obtient un nouvel échantillon de taille  $n$  de loi  $N_p(\underline{\mu}_f, \Sigma)$  obtenu indépendamment des échantillons précédents. Soit  $\bar{x}_f$  sa moyenne échantillonnale,

$$\bar{x}_f = \frac{\bar{x}_{f1} + \dots + \bar{x}_{fn}}{n},$$

alors à l'aide du théorème 2.2.1, on a que

$$T_{n;f}^2 = n(\bar{x}_f - \bar{\bar{x}})' \bar{S}^{-1} (\bar{x}_f - \bar{\bar{x}}), \quad (2.2.8)$$

suit une distribution

$$\frac{p(m+1)(n-1)}{m(n-1)-p+1} F(p, m(n-1)-p+1, \tau_{n,f}^2),$$

où  $\bar{S}$  est calculée à partir de l'équation (2.2.4) et  $\tau_{n,f}^2 = \frac{mn}{m+1} (\underline{\mu}_f - \underline{\bar{\mu}})' \Sigma^{-1} (\underline{\mu}_f - \underline{\bar{\mu}})$ . Notez que la variabilité de ce nouvel échantillon n'entre donc pas dans le calcul de  $\bar{S}$ .

La carte de contrôle est construite en plaçant sur un graphique la valeur de  $T_{n,f}^2$  avec une limite supérieure de contrôle égale à

$$LSC_{n,f} = \frac{p(m+1)(n-1)}{m(n-1)-p+1} F_\alpha(p, m(n-1)-p+1). \quad (2.2.9)$$

Par la suite, on recueille de nouveaux échantillons et on répète la procédure que l'on vient d'appliquer à ces nouveaux échantillons.

Il serait bon de souligner la différence entre les limites de contrôle de la phase un et deux. On remarque que la distribution de Fisher est la même pour les deux limites. En fait, la seule différence se situe au niveau du paramètre  $m$ , les paramètres  $p$  et  $n$  demeurant inchangés. Puisque  $LSC_{n,j}/LSC_{n,f} = (m-1)/(m+1)$ , plus  $m$ , le nombre d'échantillons recueillis à la phase un, augmente et plus cette différence devient minime. Par exemple, pour  $m=200$ , on obtient un ratio de 0,99.

On va maintenant construire un exemple de carte à partir de données simulées pour  $p=2$ . On va également démontrer qu'une procédure multivariée est en mesure de détecter un signal que les cartes univariées des variables prises une à la fois ne sont pas en mesure de détecter.

Alors pour notre exemple, soit  $\underline{x}_j \sim N_2(\underline{\mu}_j, \Sigma)$  où

$$\underline{\mu}_j = \begin{pmatrix} 10 \\ 20 \end{pmatrix}, j = 1, \dots, m \quad \text{et} \quad \Sigma = \begin{pmatrix} 12 & 8 \\ 8 & 16 \end{pmatrix}.$$

On va donc simuler  $m=100$  échantillons aléatoires de taille  $n=5$  de cette loi normale multivariée. Pour chacun de ces échantillons, on a calculé le vecteur des moyennes et la matrice de covariance échantillonnale donnée par l'équation (2.2.2). Par la suite, on a pu calculer les estimateurs pour le vecteur des moyennes et pour la matrice de covariance à l'aide des équations (2.2.3) et (2.2.4). On a obtenu,

$$\hat{\underline{\mu}} = \bar{\underline{x}} = \begin{pmatrix} 9,90 \\ 19,92 \end{pmatrix} \quad \text{et} \quad \hat{\underline{\Sigma}} = \bar{\underline{S}} = \begin{pmatrix} 11,94 & 7,50 \\ 7,50 & 15,01 \end{pmatrix}.$$

Puisque l'on est à la phase un, on a calculé la valeur de la statistique  $T_{n;j}^2$  pour chacun des 100 échantillons, à l'aide de l'équation (2.2.5).

Pour décider si le procédé est en contrôle ou non, on doit comparer ces valeurs avec la limite supérieure de contrôle donnée par l'équation (2.2.7), c'est-à-dire

$$LSC_{n;j} = \frac{792}{399} F_{\alpha}(2, 399).$$

Alors si  $\alpha=0,0027$  tout comme pour le cas univarié, on obtient  $LSC_{n;j}=11,92$ . La figure 2.8 représente la carte de contrôle  $T^2$  pour ces 100 échantillons.

On s'aperçoit donc que  $T_{n;j}^2 < LSC_{n;j}, \forall j$ . Le procédé était donc en contrôle lorsque les échantillons ont été recueillis. La phase deux consiste à recueillir de nouveaux échantillons et de vérifier si ceux-ci sont en contrôle en utilisant les estimateurs de  $\underline{\mu}$  et  $\underline{\Sigma}$  calculés à la phase un.

Par exemple, supposons que pour les prochaines observations, il se produit un déplacement de  $0,5\sigma$  vers le haut pour la première variable, alors que la deuxième variable subit un déplacement de  $0,5\sigma$  vers le bas. On simule donc 10 nouveaux échantillons provenant de cette distribution  $N_2(\underline{\mu}, \underline{\Sigma})$  où  $\underline{\mu} = (10+0,5*\sqrt{12}, 20-0,5*\sqrt{16})'$  et  $\underline{\Sigma}$  est défini comme précédemment. Ces échantillons sont indépendants entre eux et indépendants des 100 échantillons précédents. La moyenne de chacun des 10 échantillons est présentée au tableau 2.1.

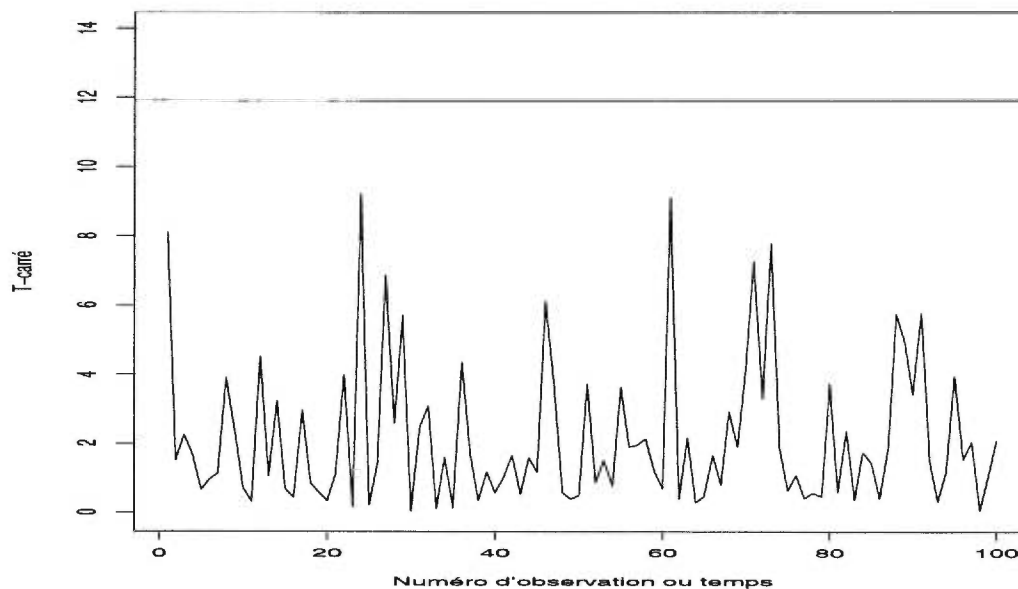


FIGURE 2.8. Carte de contrôle  $T^2$  des 100 échantillons de la phase un pour  $n > 1$ .

no. obs.	$\bar{x}_1$	$\bar{x}_2$	$T_{n,f}^2$
1	12,00	18,02	7,61
2	10,63	18,26	2,94
3	11,16	16,62	9,70
4	13,28	16,97	18,24
5	11,02	18,24	3,85
6	11,71	19,76	2,59
7	10,72	16,99	6,87
8	11,60	18,39	5,10
9	10,54	17,36	5,13
10	12,68	21,30	3,64

TABLEAU 2.1. Moyennes échantillonales et valeur de la statistique  $T_{n,f}^2$  pour les 10 échantillons de la phase deux ( $n > 1$ ).

Puisque l'on se trouve à la phase deux, la statistique  $T_{n,f}^2$  a été calculée à l'aide de l'équation (2.2.8). La valeur de  $T_{n,f}^2$  pour les 10 nouveaux échantillons est présentée au tableau 2.1. On compare ensuite ces valeurs avec la limite de contrôle de l'équation (2.2.9), c'est-à-dire,

$$LSC_{n,f} = \frac{808}{399} F_{\alpha}(2, 399).$$

Alors si  $\alpha=0,0027$ , on obtient  $LSC_{n,f}=12,16$ .

La figure 2.9 (A) montre la carte de contrôle multivariée  $T^2$  des 10 échantillons recueillis à la phase deux. On voit que la carte a détecté un signal à l'observation 4.

Maintenant que l'on a vu que la procédure  $T^2$  a été capable de détecter un signal, il serait intéressant de voir ce qui serait arrivé si on avait appliqué la procédure univariée de Shewhart à chacune des deux caractéristiques de qualité. La figure 2.9 (B) et 2.9 (C) montre les cartes univariées de Shewhart, avec  $L=3$ , présentées au chapitre 1, pour chacune des deux variables. On voit qu'il n'y a aucun point hors-contrôle. Pourquoi alors la procédure multivariée signale-t-elle un point hors-contrôle à l'observation 4, alors que celle univariée ne le fait pas ?

La figure 2.10 répond justement à cette question. On a représenté la variable  $X_1$  sur l'axe horizontal et la variable  $X_2$  sur l'axe vertical. Le rectangle sur cette figure représente les limites de contrôle de Shewhart ( $L=3$ ) lorsque les deux cartes univariées sont présentées simultanément. L'ellipse représente les limites de contrôle de la carte  $T^2$ . Comme prévu, le point 4 est hors-contrôle pour la procédure multivariée, mais en contrôle pour les procédures univariées. Évidemment, pour plus de trois variables, il devient impossible de faire un tel graphique.

La différence entre cette ellipse et le rectangle est due à la dépendance entre les deux variables. L'avantage majeur de la carte  $T^2$  est donc que celle-ci tient compte



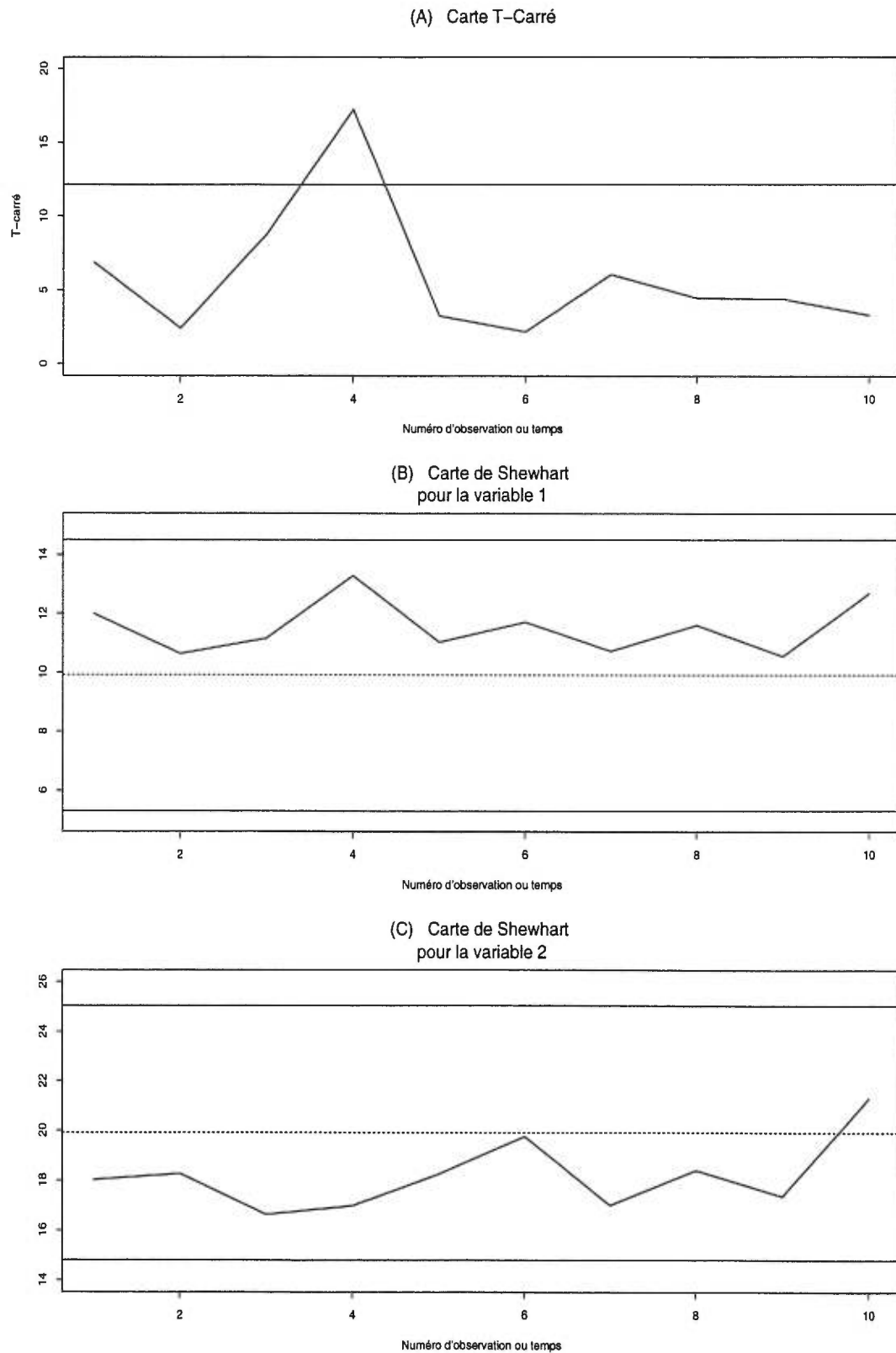


FIGURE 2.9. Cartes de contrôle multivariée et univariées des 10 observations de la phase deux pour  $n > 1$ .

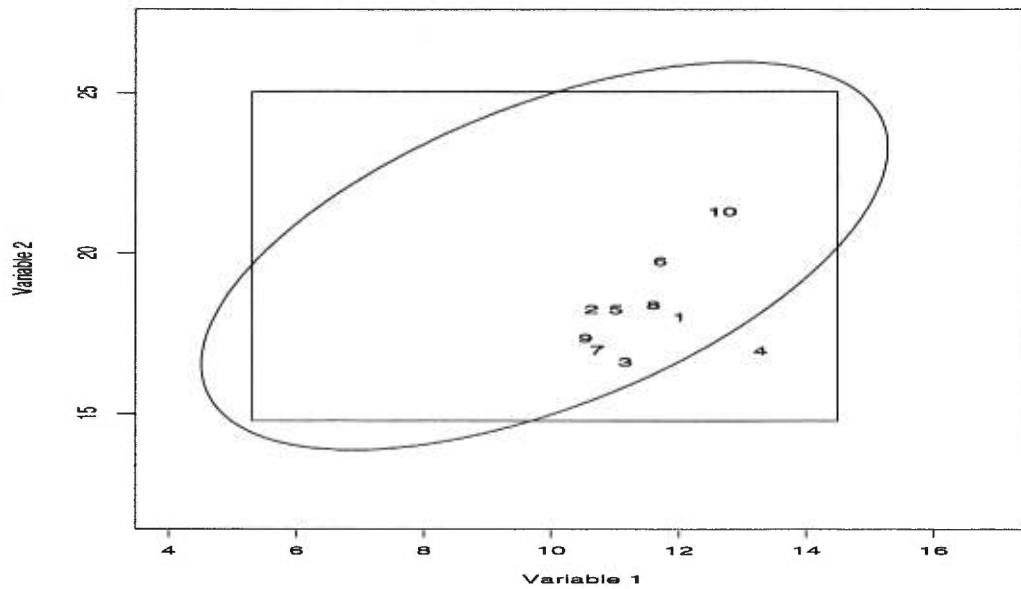


FIGURE 2.10. Régions de contrôle multivariée et univariée pour  $n > 1$ .

de la structure de corrélation des caractéristiques de qualité et c'est pour cette raison que la procédure multivariée permet de détecter des points hors-contrôle que les procédures univariées ignorent. D'ailleurs, pour  $n$ ,  $m$  et  $p$  constants, le ratio de l'aire du rectangle sur l'aire de l'ellipse est proportionnel à  $1/\sqrt{1-\rho^2}$ , où  $\rho$  est le coefficient de corrélation entre les deux variables. Ainsi, plus  $\rho$  augmente et plus l'aire du rectangle est grand par rapport à l'aire de l'ellipse. Par exemple, si  $\rho=0,5$ , l'aire du rectangle est 1,15 fois plus grand que l'aire de l'ellipse et si  $\rho=0,9$ , ce ratio devient 2,29. On s'aperçoit donc que plus la corrélation augmente et plus on risque de manquer des signaux si on utilise  $p$  procédures univariées pour décrire un procédé multivarié.

Finalement, on a calculé, pour différentes valeurs de  $p$  et de  $\rho$ , la probabilité d'avoir un point hors-contrôle en utilisant simultanément  $p$  procédures univariées. Cette probabilité correspond à celle d'avoir au moins une variable hors-contrôle. Pour le cas multivarié, cette probabilité  $\alpha$  est constante quel que soit les valeurs de  $p$  et de  $\rho$ . Ici, on suppose les variances et covariances connues et ainsi les

matrices de covariance sont respectivement pour  $p=2, 3$  et  $4$ ,

$$\Sigma_2 = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

$$\Sigma_3 = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix},$$

et

$$\Sigma_4 = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \rho^3\sigma_1\sigma_4 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho^2\sigma_2\sigma_4 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho^3\sigma_3\sigma_4 \\ \rho^3\sigma_1\sigma_4 & \rho^2\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}.$$

Pour les procédures univariées, on a utilisé les variances connues pour le calcul des limites de contrôle. On a calculé ces probabilités pour deux  $\alpha$  différents. Le premier en prenant  $\alpha=0,0027$ , qui correspond à une limite de contrôle de  $L=3$  sur les cartes univariées et l'autre qui tient compte du nombre de variables, c'est-à-dire en prenant  $\alpha=0,0027/p$ , qui correspond à  $L=3,21, 3,32$  et  $3,40$ , respectivement pour  $p=2, 3$  et  $4$ . Le tableau 2.2 donne, pour  $\rho=0, 0,5$  et  $0,9$ , la probabilité d'avoir un point hors-contrôle lorsque l'on utilise simultanément  $p$  procédures univariées de Shewhart.

	$\rho$	0,0027	0,0027/ $p$
$p = 2$	0	0,0054	0,0027
	0,5	0,0052	0,0026
	0,9	0,0042	0,0021
$p = 3$	0	0,0081	0,0027
	0,5	0,0078	0,0026
	0,9	0,0055	0,0019
$p = 4$	0	0,0108	0,0027
	0,5	0,0103	0,0026

TABLEAU 2.2. Probabilité d'avoir un point hors-contrôle si on utilise simultanément  $p$  procédures univariées de Shewhart pour différentes valeurs de  $p$  et de  $\rho$  et pour  $\alpha=0,0027$  et  $\alpha=0,0027/p$ .

On voit donc que plus  $\rho$  augmente et plus la probabilité d'avoir un point hors-contrôle diminue. Par exemple, pour  $p=3$ ,  $\alpha=0,0027/p$  et  $\rho=0,9$ , la probabilité d'avoir un point hors-contrôle est de 0,0019 et elle équivaut à utiliser des limites de contrôle de  $L=3,11$  sur les cartes univariées, alors que pour une corrélation nulle cette probabilité est 0,0027 et elle correspond à  $L=3$ . Alors plus la corrélation est élevée et plus l'avantage de l'utilisation des cartes multivariées est important.

### 2.2.3. Données individuelles

Lorsque l'on est à la phase un et que l'on a des données individuelles, il n'est pas possible d'appliquer le théorème 1.2.1 pour obtenir la distribution de  $T^2$  puisque  $\underline{x}_j$  et  $S$  ne sont pas indépendants. De plus, pour estimer la matrice de covariance  $\Sigma$ , il est normal de penser à prendre la matrice de covariance échantillonnale donnée par

$$S = \frac{1}{m-1} \sum_{j=1}^m (\underline{x}_j - \bar{\underline{x}})' (\underline{x}_j - \bar{\underline{x}}). \quad (2.2.10)$$

Cependant tout comme dans le cas univarié, il n'est pas toujours recommandé d'utiliser cette estimation puisqu'elle ne représente pas une bonne estimation de la variation à court terme. On va donc présenter dans la présente section la façon de construire la carte  $T^2$  à l'aide de  $S$  et dans la prochaine section en prenant les étendues mobiles pour l'estimation de  $\Sigma$ . Nous expliquerons dans une section ultérieure laquelle des méthodes utiliser dépendant de la nature de nos observations.

Voyons donc la méthode pour construire la carte  $T^2$  si on prend  $S$  comme estimateur de  $\Sigma$ . On considère  $m > p$  échantillons aléatoires obtenus sous les hypothèses du Modèle 2.2.1 avec cependant  $n=1$ . La moyenne échantillonnale est donnée par

$$\bar{\underline{x}} = \frac{\underline{x}_1 + \dots + \underline{x}_m}{m},$$

En prenant  $S$  comme estimation de  $\Sigma$ , on obtient selon une extension des résultats de Tracy, Young et Mason (1992), que

$$T_{1;j}^2 = (\underline{x}_j - \bar{x})' S^{-1} (\underline{x}_j - \bar{x}), \quad (2.2.11)$$

suit une distribution

$$\frac{(m-1)^2}{m} B\left(\frac{p}{2}, \frac{m-p-1}{2}; \tau_{1;j}^2\right), \quad (2.2.12)$$

où  $B(\nu_1, \nu_2)$  est une distribution Bêta de paramètre  $\nu_1$  et  $\nu_2$  et

$$\tau_{1;j}^2 = \frac{m}{m-1} (\underline{\mu}_j - \underline{\bar{\mu}})' \Sigma^{-1} (\underline{\mu}_j - \underline{\bar{\mu}}), \quad j = 1, \dots, m.$$

La carte de contrôle est par la suite construite en plaçant sur un graphique les valeurs de  $T_{1;j}^2$ ,  $j = 1, \dots, m$  en fonction du numéro d'observation ou du temps avec une limite supérieure de contrôle de

$$LSC_{1;f} = \frac{(m-1)^2}{m} B_\alpha\left(\frac{p}{2}, \frac{m-p-1}{2}\right), \quad (2.2.13)$$

où  $B_\alpha(\nu_1, \nu_2)$  est la valeur critique au niveau  $\alpha$  d'une distribution Bêta de paramètre  $\nu_1$  et  $\nu_2$ .

On est donc prêt à passer à la phase deux. On considère maintenant une nouvelle observation  $\underline{x}_f$  provenant de la loi  $N_p(\underline{\mu}_f, \Sigma)$  obtenue indépendamment des  $m$  observations précédentes. Alors à l'aide du théorème 1.2.1,

$$T_{1;f}^2 = (\underline{x}_f - \bar{x})' S^{-1} (\underline{x}_f - \bar{x}), \quad (2.2.14)$$

suit une distribution

$$\frac{p(m+1)(m-1)}{m(m-p)} F(p, m-p; \tau_{1;f}^2), \quad (2.2.15)$$

où  $\tau_{1;f}^2 = \frac{m}{m+1} (\underline{\mu}_f - \underline{\bar{\mu}})' \Sigma^{-1} (\underline{\mu}_f - \underline{\bar{\mu}})$ .

L'hypothèse  $H_0 : \underline{\mu}_f = \underline{\mu}$  est rejetée au niveau  $\alpha$  si  $T_{1;f}^2$  est plus grand que la limite supérieure de contrôle, c'est-à-dire plus grand que

$$LSC_{1;f} = \frac{p(m+1)(m-1)}{m(m-p)} F_\alpha(p, m-p).$$

#### 2.2.4. Estimation de $\Sigma$ à partir des étendues mobiles

Comme on l'a mentionné à la section précédente, l'estimation de  $\Sigma$  par  $S$  la matrice de covariance échantillonnale, n'est pas toujours une bonne estimation à utiliser puisqu'elle ne représente pas la variation à court terme du procédé. Holmes et Mergen (1993) ont proposé une procédure plus puissante pour l'estimation de  $\Sigma$  qui tient compte de la nature en série des observations et qui sera une estimation plus robuste de la variation à court terme.

L'estimateur qu'ils proposent est une extension de l'approche des étendues mobiles utilisée dans les applications univariées et on le dénote par  $\tilde{S}$ . Cet estimateur utilise les différences entre chaque paire d'observations successives de  $\underline{x}$ . Soit  $\underline{v}_j$  cette différence, un vecteur ligne de dimension  $p$ , on a donc

$$\underline{v}_j = \underline{x}_{j+1} - \underline{x}_j, \quad j = 1, \dots, m-1.$$

Un estimateur sans biais pour  $\Sigma$  est alors donné par,

$$\tilde{S} = \frac{1}{2} \frac{V'V}{(m-1)}, \quad (2.2.16)$$

où  $V$  est une matrice de dimension  $(m-1)$  par  $p$  formée à partir des vecteurs  $\underline{v}_j$ ,  $j = 1, \dots, m-1$ .

Selon Scholz et Tosch (1994), la distribution approximative de  $\tilde{S}$  est une distribution Wishart, c'est-à-dire

$$b\tilde{S} \approx W_p(b, \Sigma), \quad \text{où } b = \frac{2(m-1)^2}{3m-4}.$$

Pour distinguer, la statistique  $T^2$  et la limite de contrôle calculée à partir de  $\tilde{S}$ , on utilisera  $\tilde{T}^2$  et  $L\tilde{S}C$ . Une approximation de la distribution de  $\tilde{T}^2$  pour les données de la phase un est présentée dans Sullivan et Woodall (1996):

$$\tilde{T}_{1,j}^2 = (\underline{x}_j - \bar{\underline{x}})' \tilde{S}^{-1} (\underline{x}_j - \bar{\underline{x}}), \quad (2.2.17)$$

suit approximativement une distribution

$$\frac{(m-1)^2}{m} B\left(\frac{p}{2}, \frac{b-p-1}{2}\right). \quad (2.2.18)$$

La limite supérieure de contrôle pour la carte  $T^2$  basée sur  $\tilde{S}$  est alors

$$L\tilde{S}C_{1;j} = \frac{(m-1)^2}{m} B_\alpha\left(\frac{p}{2}, \frac{b-p-1}{2}\right).$$

Si  $\tilde{T}_{1;j}^2 > L\tilde{S}C_{1;j}$ ,  $j = 1, \dots, m$ , il y a hors-contrôle. On peut maintenant passer à la phase deux en recueillant une nouvelle observation. Selon Scholz et Tosch (1994), la statistique,

$$\tilde{T}_{1,f}^2 = (\underline{x}_f - \bar{x})' \tilde{S}^{-1} (\underline{x}_f - \bar{x}), \quad (2.2.19)$$

suit une distribution approximative,

$$\frac{bp}{b-p+1} \frac{m+1}{m} F(p, b-p+1). \quad (2.2.20)$$

La limite supérieure de contrôle pour la phase deux en utilisant toutes les observations pour l'estimation de  $\Sigma$  est donc

$$L\tilde{S}C_{1;f} = \frac{bp}{b-p+1} \frac{m+1}{m} F_\alpha(p, b-p+1), \quad (2.2.21)$$

où  $b = (2(m-1)^2)/(3m-4)$ .

On va maintenant construire un exemple de carte  $T^2$  pour des données individuelles. Les données vont être simulées à partir d'une loi normale multivariée pour  $p=2$ . Alors pour notre exemple, soit  $\underline{x}_j \sim N_2(\underline{\mu}_j, \Sigma)$ , où

$$\underline{\mu}_j = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad j = 1, \dots, m \quad \text{et} \quad \Sigma = \begin{pmatrix} 9 & 3 \\ 3 & 7 \end{pmatrix}.$$

On va donc simuler  $m=100$  observations aléatoires de cette loi normale multivariée. On va se servir des ces observations pour estimer le vecteur des moyennes et la matrice de covariance du procédé. Le vecteur moyen de ces 100 observations

est  $\bar{x} = (0,244, -0,346)'$ . On va appliquer la procédure  $T^2$  à partir de l'estimation de  $\Sigma$  calculée à partir des étendues mobiles. Alors à l'aide de la procédure décrite à la section 2.2.4, on a calculé les différences successives entre chaque observation. Par la suite, à l'aide de l'équation (2.2.16) on a obtenu l'estimation suivante pour  $\Sigma$ ,

$$\tilde{S} = \begin{pmatrix} 8,79 & 2,53 \\ 2,53 & 7,14 \end{pmatrix}. \quad (2.2.22)$$

On a ensuite calculé la valeur de  $\tilde{T}_{1;j}^2$ ,  $j = 1, \dots, 100$ , à partir de l'équation (2.2.17). La limite de contrôle supérieure correspondante est

$$L\tilde{S}C_{1;j} = \frac{(100 - 1)^2}{100} B_{\alpha}(1; 32, 61).$$

Pour calculer les limites de contrôle, on prendra  $\alpha=0,005$  pour pouvoir faire une comparaison avec les cartes cusum et ewma. Ce choix de  $\alpha$  correspond à une  $LMS_0=200$ . Donc si  $\alpha=0,005$ , on obtient  $L\tilde{S}C_{1;j}=14,70$ . Puisque toutes les valeurs de  $\tilde{T}_{1;j}^2 \leq L\tilde{S}C_{1;j}$ , on conclut que le procédé est en contrôle. On peut donc passer à la phase deux.

On va donc simuler 2 nouvelles séries de 10 observations provenant d'une loi normale ayant subi un déplacement de sa moyenne. La carte de contrôle pour les 10 premières observations sera présentée dans la présente section, alors que la carte pour les 10 dernières sera présentée dans la section comparaison des cartes (section 2.5). Supposons donc que les 10 premières observations proviennent d'une loi  $N_2(\underline{\mu}_f, \Sigma)$  avec

$$\underline{\mu}_f = \begin{pmatrix} 0 - \sqrt{9} \\ 0 + \sqrt{7} \end{pmatrix},$$

alors que les 10 autres proviennent d'une loi  $N_2(\underline{\mu}_f, \Sigma)$  avec

$$\underline{\mu}_f = \begin{pmatrix} 0 + \sqrt{9} \\ 0 + 2\sqrt{7} \end{pmatrix}.$$



Alors, pour les 10 premières observations, la première variable a subi un déplacement de  $1\sigma$  vers le bas alors que la deuxième a subi un déplacement de  $1\sigma$  vers le haut. Ce déplacement équivaut à un paramètre de décentralité de  $\tau^2=3,22$ . Le tableau 2.3 contient ces 10 nouvelles observations.

no. obs.	$x_1$	$x_2$
1	-3,37	0,39
2	-0,93	3,19
3	-2,64	2,66
4	-2,10	2,01
5	-2,27	4,35
6	-6,12	5,51
7	-2,77	3,11
8	4,41	3,62
9	0,88	5,27
10	2,51	2,34

TABLEAU 2.3. Les 10 observations de la phase deux pour  $n=1$  ayant subi un déplacement de  $\tau^2=3,22$ .

Pour les 10 dernières, la première variable a subi un déplacement de  $1\sigma$  vers le haut, alors que la deuxième a subi un déplacement de  $2\sigma$  vers le haut. Ce déplacement équivaut à un paramètre de décentralité de  $\tau^2=4,07$ . Le tableau 2.4 contient la valeur de ces 10 observations.

no. obs.	$x_1$	$x_2$
1	5,29	7,35
2	3,99	11,36
3	5,61	6,81
4	0,54	7,80
5	2,72	4,09
6	8,69	9,42
7	4,11	2,86
8	-1,31	2,79
9	5,84	8,28
10	-3,07	4,07

TABLEAU 2.4. Les 10 observations de la phase deux pour  $n=1$  ayant subi un déplacement de  $\tau^2=4,07$ .

Pour ces 20 nouvelles observations, on a calculé la valeur de la statistique  $\tilde{T}_{1,f}^2$  à l'aide de l'équation (2.2.19).

La limite supérieure de contrôle correspondante calculée à partir de l'équation (2.2.21) et pour  $\alpha=0,005$  est égale à  $L\tilde{S}C_{1,f}=11,80$ .

La figure 2.11 représente la carte  $T^2$  pour les 10 premières observations.

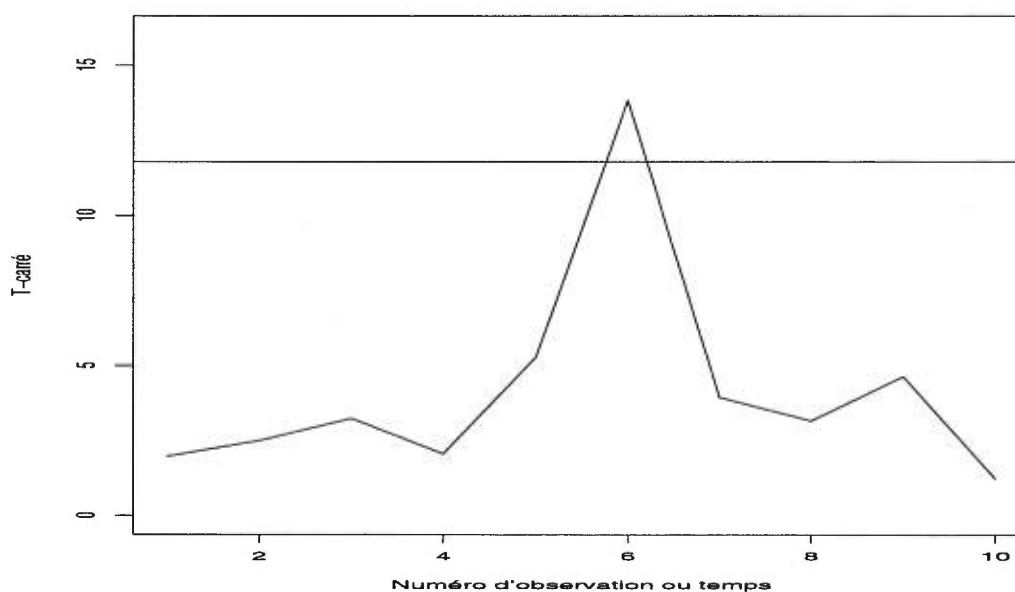


FIGURE 2.11. Carte de contrôle  $T^2$  pour 10 observations de la phase deux ( $n=1$ ) en utilisant les étendues mobiles pour l'estimation de  $\Sigma$ .

On s'aperçoit alors qu'il y a hors-contrôle pour l'observation 6. Puisque l'on est à deux dimensions, il serait intéressant de faire le graphique de la région de contrôle multivariée (ellipse) et de celle univariée (rectangle) obtenue en appliquant la procédure univariée de Shewhart ( $\alpha=0,005$  et donc  $L=2,81$ ) à chacune des deux caractéristiques de qualité. La figure 2.12 représente ces deux régions de contrôle ainsi que les 10 premières observations.

Encore ici, on s'aperçoit que le point 6 est hors-contrôle pour la procédure multivariée. Cependant pour les procédures univariées, il est à l'intérieur des limites et donc en contrôle.

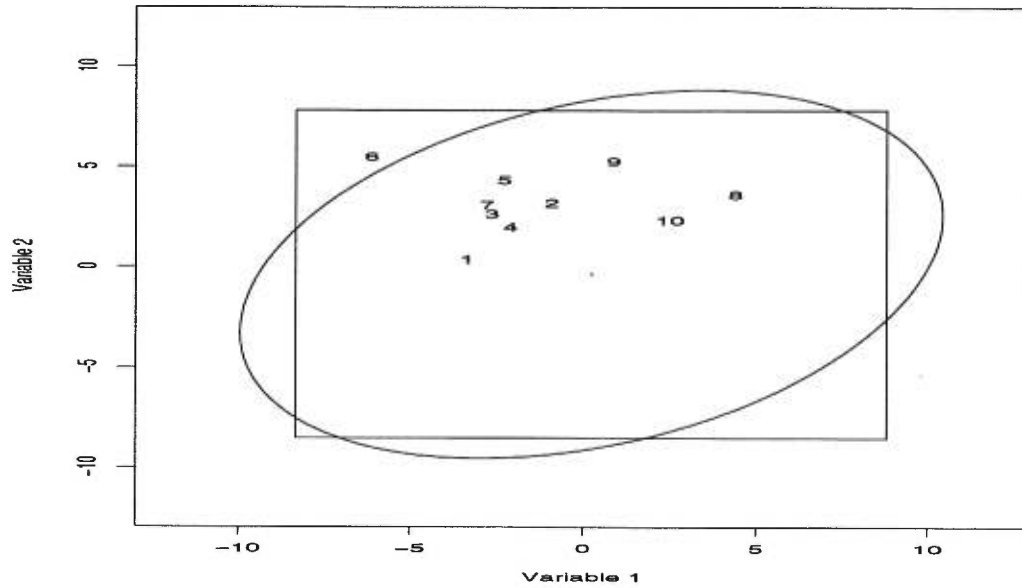


FIGURE 2.12. Régions de contrôle multivariée et univariée pour les 10 observations de la phase deux pour des données individuelles.

La carte  $T^2$  pour les 10 dernières observations est présentée à la figure 2.16 (section 2.5).

### 2.2.5. Résumé des statistiques $T^2$ et des limites de contrôle

Dans la présente section, on va résumer la procédure  $T^2$  en présentant la statistique  $T^2$  et la limite de contrôle qu'on doit utiliser dépendant si on a des données individuelles ou groupées et si on est à la phase un ou deux. Voir le tableau 2.5. De plus, pour les données individuelles, seule la procédure à partir des étendues mobiles ( $\tilde{S}$ ) est présentée puisque Sullivan et Woodall (1996), après avoir testé six estimateurs de  $\Sigma$ , recommandent l'utilisation de  $\tilde{S}$ .

Données	Phase	Notation	Formules
Groupées	un	$T_{n;j}^2$ $LSC_{n;j}$	$n(\bar{x}_j - \bar{x})' \bar{S}^{-1} (\bar{x}_j - \bar{x})$ $\frac{p(m-1)(n-1)}{m(n-1)-p+1} F_\alpha(p, m(n-1) - p + 1)$
	deux	$T_{n;j}^2$ $LSC_{n;j}$	$n(\bar{x}_f - \bar{x})' \bar{S}^{-1} (\bar{x}_f - \bar{x})$ $\frac{p(m+1)(n-1)}{m(n-1)-p+1} F_\alpha(p, m(n-1) - p + 1)$
Ind.	un	$T_{1;j}^2$ $LSC_{1;j}$	$(x_j - \bar{x})' \tilde{S}^{-1} (x_j - \bar{x})$ $\frac{(m-1)^2}{m} B_\alpha\left(\frac{p}{2}, \frac{b-p-1}{2}\right)$
	deux	$T_{1;j}^2$ $LSC_{1;j}$	$(x_f - \bar{x})' \tilde{S}^{-1} (x_f - \bar{x})$ $\frac{bp}{b-p+1} \frac{m+1}{m} F_\alpha(p, b - p + 1)$

TABLEAU 2.5. Résumé des statistiques  $T^2$  et des limites de contrôle;  $b = (2(m-1)^2)/(3m-4)$ .

### 2.2.6. LMS de la carte $T^2$

Tout comme pour le cas univarié, la *LMS* nous permettra de comparer la performance des cartes entre elles. Puisque la procédure  $T^2$  est l'extension multivariée de la carte de Shewhart au niveau univarié, le calcul des *LMS* est similaire.

Ainsi on a  $LMS_0 = 1/\alpha$  et  $LMS_1 = 1/(1 - \beta)$ , où  $\alpha$  est l'erreur de type 1, c'est-à-dire la probabilité de faussement déclarer le procédé hors-contrôle et  $\beta$  est l'erreur de type 2, c'est-à-dire la probabilité de faussement juger le procédé en contrôle.

Ainsi, pour les données individuelles de la phase deux lorsque l'on utilise les étendues mobiles pour l'estimation de  $\Sigma$ , on a que  $\alpha$  est approximativement égale à

$$1 - F_{(p, f-p+1)}^{-1} \left( \left( \frac{m}{m+1} \right) \left( \frac{f-p+1}{fp} \right) LSC_{1;f} \right) \quad (2.2.23)$$

et  $\beta$  est approximativement égale à

$$F_{(p, f-p+1; \tau^2)}^{-1} \left( \left( \frac{m}{m+1} \right) \left( \frac{f-p+1}{fp} \right) LSC_{1;f} \right), \quad (2.2.24)$$

où  $F_{(a,b)}^{-1}(x)$  est la fonction de répartition d'une loi de Fisher de paramètre  $a$  et  $b$  et  $\tau^2$  représente le déplacement que l'on veut détecter. On peut donc obtenir la longueur moyenne d'une suite pour différents déplacements du vecteur des moyennes. On doit noter que même si le vecteur des moyennes  $\underline{\mu}$  et la matrice de covariance  $\Sigma$  sont inconnus, les calculs sont possibles (bien qu'en utilisant  $\bar{S}$ , les résultats sont approximatifs).

Les calculs des  $LMS_1$  sont donnés en fonction de  $\tau$  qui est la racine carré du paramètre de décentralité. Le paramètre  $\tau$  est très souvent utilisé pour représenter une mesure de distance entre la moyenne du procédé et l'observation.

Le tableau 2.6 donne pour des déplacements  $\tau$  du vecteur des moyennes, la valeur des longueurs moyennes de la suite pour  $p=2$  et  $m=100$ .

$LMS_0$	$\tau$				
	0,5	1,0	2,0	3,0	4,0
200	120	45,5	7,76	2,40	1,31
370	211	74,2	10,9	2,94	1,43
500	279	94,5	12,9	3,26	1,51

TABLEAU 2.6.  $LMS_1$  de la procédure  $T^2$  pour des données individuelles de la phase deux, pour  $p=2$  et  $m=100$ , lorsque l'on utilise les étendues mobiles pour l'estimation de  $\Sigma$ .

### 2.3. LA CARTE CUSUM MULTIVARIÉE

Tout comme pour le cas univarié, la carte  $T^2$  est relativement insensible aux petits déplacements de la moyenne puisque la statistique pour le point  $j$  est calculée seulement à partir de ce point et non de la séquence de point entière. Alors si on est intéressé à détecter les petits déplacements de la moyenne pour le cas multivarié, il existe deux alternatives à la carte  $T^2$ , il s'agit de la carte cusum multivariée et de la carte ewma multivariée. Dans cette section, on présente donc la carte cusum multivariée qui a été introduite par Crosier (1988).

En fait, il existe plusieurs types de cartes cusum multivariées. On va présenter celle introduite par Crosier (1988) puisque c'est celle qui est la plus souvent représentée lorsque l'on aborde le sujet des cartes de contrôle cusum multivariées.

#### 2.3.1. La carte cusum multivariée de Crosier

Crosier (1988) a introduit deux types de cartes cusum multivariées. La première procédure consiste à réduire chaque observation multivariée en un scalaire (la statistique  $T^2$ ) et par la suite d'appliquer la procédure cusum univariée à ces scalaires. La deuxième procédure consiste plutôt à former un vecteur cusum directement des observations multivariées.

On va présenter la deuxième procédure puisque Crosier (1988) a démontré qu'elle obtient une meilleure performance pour la *LMS* et qu'en plus, elle donne une certaine indication de la direction qu'a pris la moyenne après avoir subi le déplacement.

Cette procédure est présentée pour le cas des données individuelles, mais elle s'ajuste bien pour le cas des données groupées. On n'aura qu'à remplacer le vecteur  $\underline{x}_j$  par le vecteur  $\bar{\underline{x}}_j$  et la matrice de covariance  $\Sigma$  par  $\Sigma/n$ .

La procédure cusum multivariée de Crosier consiste à définir un plan multivarié cusum comme suit. Soit le vecteur d'observations  $\underline{x}_j$  obtenu sous les hypothèses du Modèle 2.2.1, avec cependant  $n=1$ . On calcule tout d'abord la statistique  $C_j$ , où

$$C_j = [(\underline{d}_{j-1} + \underline{x}_j - \underline{\mu})' \Sigma^{-1} (\underline{d}_{j-1} + \underline{x}_j - \underline{\mu})]^{1/2}, \quad j = 1, \dots, m,$$

avec comme valeur de départ  $\underline{d}_0 = \underline{0}$ . Par la suite, on doit calculer la valeur du vecteur  $\underline{d}_j$ , où

$$\underline{d}_j = \begin{cases} (\underline{d}_{j-1} + \underline{x}_j - \underline{\mu})(1 - k/C_j) & \text{si } C_j > k \\ 0 & \text{si } C_j \leq k \end{cases}$$

$j = 1, \dots, m$ . Finalement, on calcule la statistique cusum  $Y_j$ , où

$$Y_j = [\underline{d}_j' \Sigma^{-1} \underline{d}_j]^{1/2}.$$

On doit noter qu'ici, on a utilisé le vecteur  $\underline{d}_j$  et non  $\underline{S}_j$  comme le fait Crosier (1988) pour éviter toute confusion avec  $S$  la matrice de covariance échantillonnale.

Il y a hors-contrôle à l'observation  $j$  si  $Y_j > h_{cusum}$ . La carte de contrôle est construite en plaçant sur un graphique la valeur de  $Y_j$ ,  $j = 1, \dots, m$ , en fonction du numéro d'observation ou du temps avec une limite supérieure de contrôle de  $h_{cusum}$ .

Pour mieux comprendre la procédure cusum, examinons-la donc plus en détail. On calcule tout d'abord une statistique ( $C_1$ ) qui est la longueur du vecteur  $(\underline{x}_1 - \underline{\mu})$  par rapport à la matrice de covariance  $\Sigma$ . Si ce déplacement ( $C_1$ ) est assez important, donc plus grand qu'un certain  $k$ , on doit tenir compte de ce déplacement en calculant le vecteur  $\underline{d}_1$ . Alors si  $C_1 > k$ , on calcule le vecteur  $\underline{d}_1$  en réduisant le vecteur  $(\underline{x}_1 - \underline{\mu})$  d'un facteur de  $(1 - k/C_1)$ . C'est la longueur de ce nouveau vecteur qui nous permettra de décider si le procédé est en contrôle. On dénote la longueur du vecteur  $\underline{d}_1$  par  $Y_1$ . Alors si  $Y_1$  n'est pas assez grand et donc qu'il n'y a pas hors-contrôle, on recommence le processus avec le vecteur  $\underline{d}_1$  qui est différent de  $\underline{0}$  puisqu'il tient compte que  $\underline{x}_1$  était différent de  $\underline{\mu}$ . On calcule maintenant la longueur du vecteur  $(\underline{d}_1 + \underline{x}_2 - \underline{\mu})$  par rapport à  $\Sigma$ . Encore une fois, si  $C_2 > k$ , on continue à accumuler les différences. On répète ce processus jusqu'à ce que  $Y_n > h_{cusum}$  (hors-contrôle) ou jusqu'à ce que la longueur du vecteur  $(\underline{d}_{n-1} + \underline{x}_n - \underline{\mu})$  soit redevenue plus petite que  $k$ . Si tel est le cas, on réinitialise le vecteur  $\underline{d}_n$  à  $\underline{0}$  et on recommence la procédure décrite précédemment.

On s'aperçoit que cette procédure cusum multivariée est semblable à celle univariée puisqu'elle permet l'ajustement de deux paramètres, soit  $k$  et  $h_{cusum}$ . Le paramètre  $k$  est choisi en fonction du déplacement de la moyenne que l'on veut détecter. Tout comme la procédure cusum univariée, on recommande de choisir  $k = \tau/2$ , où  $\tau$  est la racine carrée du paramètre de décentralité ou le déplacement que l'on veut détecter. Ce choix minimise la valeur de la  $LMS_1$  pour un déplacement de la moyenne équivalent à  $\tau$  et pour une  $LMS_0$  fixe. Le paramètre  $h_{cusum}$  est choisi en fonction de la  $LMS_0$  désirée, en utilisant une approximation de la chaîne de Markov. On doit noter que ces deux paramètres sont également fonction de  $p$ , le nombre de variables.

Tout comme pour la procédure cusum univariée, il faudra estimer  $\underline{\mu}$  et  $\Sigma$  à partir des observations de la phase un. Pour le cas des données individuelles, on estimera  $\Sigma$  en utilisant les étendues mobiles, c'est-à-dire à l'aide de l'équation



(2.2.16). Si  $n > 1$ , on se servira plutôt de la moyenne des matrices de covariances (équation 2.2.4). Pour  $\underline{\mu}$ , on prendra évidemment le vecteur de la moyenne échantillonnale  $\bar{x}$ . Alors pour le cas des données individuelles, on utilisera plutôt les équations suivantes pour la procédure cusum multivariée,

$$\hat{C}_j = [(\hat{d}_{j-1} + \underline{x}_j - \bar{x})' \tilde{S}^{-1} (\hat{d}_{j-1} + \underline{x}_j - \bar{x})]^{1/2}, \quad (2.3.1)$$

$$\hat{d}_j = \begin{cases} (\hat{d}_{j-1} + \underline{x}_j - \bar{x})(1 - k/\hat{C}_j) & \text{si } \hat{C}_j > k \\ 0 & \text{si } \hat{C}_j \leq k \end{cases} \quad (2.3.2)$$

et

$$\hat{Y}_j = [\hat{d}_j' \tilde{S}^{-1} \hat{d}_j]^{1/2}. \quad (2.3.3)$$

On va donc construire un exemple de carte cusum multivariée à partir des données individuelles de la phase deux. Les estimations de  $\underline{\mu}$  et de  $\Sigma$  ont été calculées à partir des données de la phase un pour l'exemple des données individuelles de la procédure  $T^2$ . On va calculer la valeur de la statistique  $\hat{C}_j$ , puis du vecteur  $\hat{d}_j$  et finalement de la statistique cusum  $\hat{Y}_j$  à l'aide respectivement des équations (2.3.1), (2.3.2) et (2.3.3).

On a utilisé comme paramètres  $k=1/2$  et la limite supérieure de contrôle a été fixée à  $h_{cusum}=5,50$ . Ces choix correspondent à une  $LMS_0=200$ , tout comme l'exemple de la procédure  $T^2$ . On a construit la carte cusum pour les deux mêmes exemples que  $T^2$  (pour les données individuelles). La carte pour les 10 premières observations est présentée dans la présente section alors que celle pour les 10 dernières est présentée dans la section comparaison (2.5).

La figure 2.13 représente la carte cusum multivariée pour les 10 premières observations qui correspondent à des observations provenant d'une loi normale multivariée avec un vecteur des moyennes ayant subi un déplacement de  $1\sigma$  vers le bas pour la première variable et de  $1\sigma$  vers le haut pour la deuxième variable.

On voit donc que la procédure cusum multivariée a détecté un point hors-contrôle à l'observation 5 et à l'observation 9, comparativement à la carte  $T^2$  qui avait détecté un signal seulement à l'observation 6. Il est important de souligner que lorsque la procédure cusum détecte un point hors-contrôle, on réinitialise la procédure à 0 en égalant le vecteur  $\underline{d}_j$  au vecteur nul  $\underline{0}$ . L'avantage de la procédure cusum c'est que celle-ci accumule les petits déplacements dans la même direction pour détecter des points hors-contrôle plus rapidement que la carte  $T^2$ .

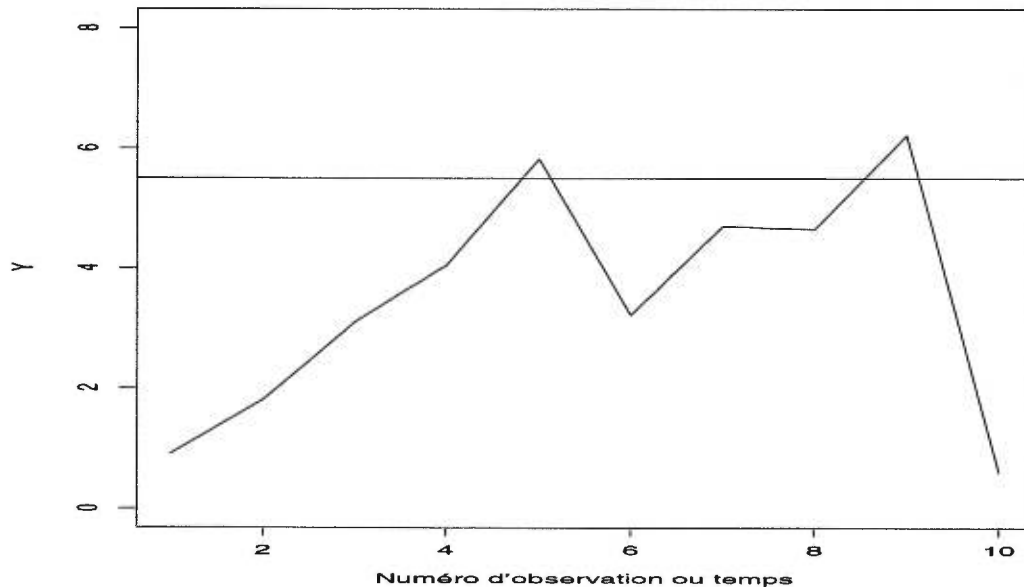


FIGURE 2.13. Carte de contrôle cusum multivariée pour 10 observations de la phase deux avec  $k=1/2$  et  $h_{cusum}=5,50$  ( $n=1$ ).

Puisque l'on est à deux dimensions, il est possible de visualiser la carte cusum à l'aide d'une région de contrôle ellipsoïdale. La figure 2.14 représente cette carte. La plus petite ellipse représente la région délimitée par le paramètre  $k=1/2$ . C'est cette ellipse qui nous permettra de juger si le déplacement d'une observation par rapport à sa moyenne est suffisamment important pour accumuler ce déplacement. On doit noter que si pour l'observation  $j$ , la longueur du vecteur  $(\hat{d}_{j-1} + \underline{x}_j - \hat{\mu})$  redevient plus petite que  $k=1/2$ , et donc que le point représenté par  $(\hat{d}_{j-1} + \underline{x}_j - \hat{\mu})$  est à l'intérieur de la plus petite ellipse, on réinitialise la procédure à 0. La

plus grande ellipse représente la région de contrôle délimitée par le paramètre  $h_{cusum}=5,50$ . Il y a hors-contrôle à l'observation  $j$  si la longueur du vecteur  $\hat{d}_j = (\hat{d}_{j-1} + \underline{x}_j - \hat{\mu})(1 - k/\hat{C}_j)$  est plus grande que  $h_{cusum}=5,50$  et donc si le point représentée par  $\hat{d}_j$  est à l'extérieur de la plus grande ellipse. Sur cette figure, les nombres 1 à 10 représentent les observations de la phase deux (les  $\underline{x}_j$ ), alors que les lettres A à J représentent les points du vecteur  $\hat{d}_j$ ,  $j = 1 \dots, 10$ .

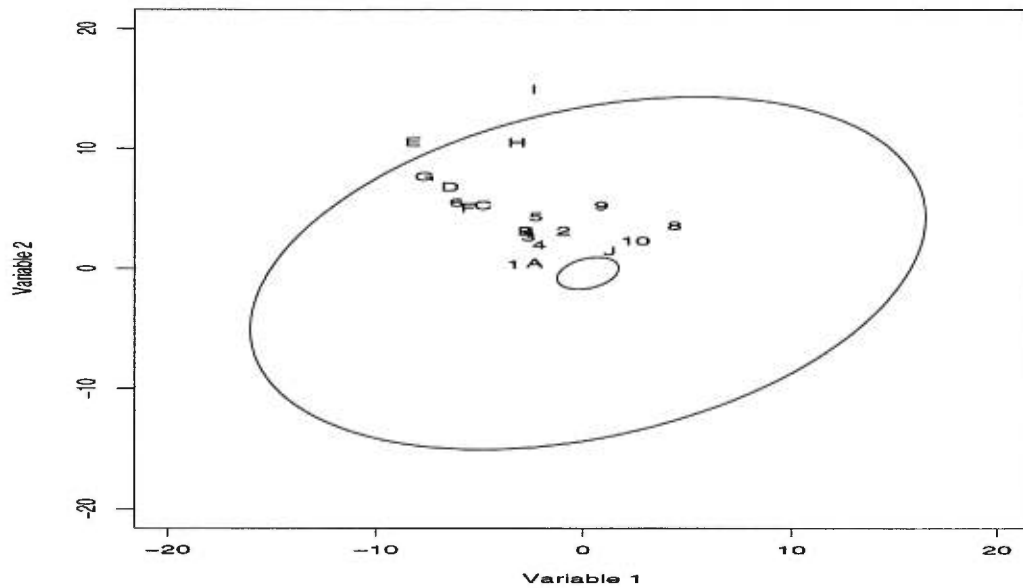


FIGURE 2.14. Région de contrôle de la carte cusum multivariée avec  $k=1/2$  et  $h_{cusum}=5,50$  ( $n=1$ ).

Comme prévu, les observations 5 (point E) et 9 (point I) sont hors-contrôle puisqu'ils sont à l'extérieur de la grande ellipse.

Il faut noter que cette représentation graphique de la carte cusum multivariée est possible pour  $p=2$ . Pour plus de 3 variables, il devient impossible de faire un tel graphique.

La carte de contrôle pour la deuxième série d'observations est présentée à la section comparaison (section 2.5).

### 2.3.2. Paramètres et $LMS$ de la carte cusum multivariée

On doit tout d'abord mentionner que les  $LMS$  de la procédure cusum multivariée ne dépendent de la moyenne  $\underline{\mu}$  et de la matrice de covariance  $\Sigma$  qu'à travers le paramètre de décentralité  $\tau^2$ .

Le choix des paramètres de la procédure cusum multivariée s'obtient de la même façon que pour le cas univarié. On choisira le paramètre  $k$  en fonction du déplacement du vecteur des moyennes que l'on veut détecter, qui nous est donné par la racine carré du paramètre de décentralité  $\tau^2$ . Le choix de  $k = \tau/2$  semble minimiser la  $LMS_1$  pour une  $LMS_0$  fixe. Le paramètre  $h_{cusum}$  est choisi pour obtenir une  $LMS_0$  acceptable.

Crosier (1988) a utilisé l'approximation de la chaîne de Markov pour déterminer les  $LMS_0$  et donc pour fixer la valeur du paramètre  $h_{cusum}$ . Les  $LMS_1$  ont été obtenues par simulations.

Le tableau 2.7 donne pour différentes valeurs de  $p$  et des  $LMS_0$  égales à 200 et 500, les valeurs de  $h_{cusum}$  à utiliser pour un  $k$  fixé à 0,5.

$p$	$LMS_0=200$	$LMS_0=500$
2	5,50	6,65
5	9,46	10,9
10	14,9	17,2
20	24,7	28,0

TABLEAU 2.7. Valeurs du paramètre  $h_{cusum}$  à utiliser pour  $k=0,5$ ,  $p$  et  $LMS_0$  donnés pour la procédure cusum multivariée.

### 2.3.3. Autres procédures cusum multivariés

Dans les sections précédentes, on a présenté la carte cusum multivariée de Crosier. Il existe cependant d'autres types de cartes cusum qui ont été introduites par d'autres auteurs. Dans cette section, on fera un survol de ces autres cartes.

Les premiers à avoir parlé de procédure cusum multivariée furent Woodall et Ncube (1985). Dans cet article, l'utilisation simultanée de  $p$  procédures cusum univariées bilatérales est considérée comme étant une seule procédure cusum multivariée. La procédure multivariée est hors-contrôle si au moins une des procédures univariées est hors-contrôle. Ils utilisent la matrice de covariance  $\Sigma$  pour étudier la performance de la *LMS*.

Healy (1987) a démontré qu'une procédure cusum univariée basée sur une combinaison linéaire des variables peut être utilisée pour une situation multivariée si on est intéressé à détecter un changement dans le vecteur de la moyenne dans une seule direction connue.

Comme on l'a mentionné précédemment, Crosier (1988) avait introduit deux types de cartes cusum multivariées. On a présenté la deuxième. La première consiste à tout d'abord réduire chaque observation en un scalaire (la statistique  $T = \sqrt{T^2}$ ) et par la suite d'appliquer une procédure cusum univariée à ces scalaires.

Mohebbi et Hayre (1989) a appliqué les fonctions de pertes à la procédure cusum multivariée.

Pignatiello et Runger (1990) ont quant à eux proposé deux autres cartes cusum multivariées qui sont très semblables en performance aux deux cartes de Crosier (1988). La première procédure accumule le vecteur d'observations  $X$  avant de produire la forme quadratique (comparable à la procédure de Crosier que l'on a présentée). La deuxième calcule une forme quadratique et par la suite accumule ces formes quadratiques (comparable à la première procédure de Crosier).

Hawkins (1991, 1993b) a introduit une carte cusum basée sur le vecteur des résidus de la régression de chaque variable sur toutes les autres.

Finalement, Sparks (1992) a appliqué l'analyse en composantes principales à la carte cusum.

## 2.4. LA CARTE EWMA MULTIVARIÉE

On va maintenant présenter la carte ewma multivariée, qui tout comme la procédure cusum, nous permet de détecter les petits changements dans le vecteur des moyennes plus rapidement que la carte  $T^2$ .

Elle a été introduite par Lowry, Woodall, Champ et Rigdon (1992). Elle est présentée pour des données individuelles, mais elle s'ajuste bien pour des données groupées. On aura qu'à remplacer le vecteur  $\underline{x}_j$  par le vecteur  $\bar{\underline{x}}_j$  et la matrice de covariance  $\Sigma$  par  $\Sigma/n$ .

### 2.4.1. Procédure pour la carte ewma multivariée

Soit le vecteur d'observations  $\underline{x}_j$  obtenu sous les hypothèses du Modèle 2.2.1, avec cependant  $n=1$ . On calcule tout d'abord la statistique  $\underline{z}_j$  comme suit,

$$\underline{z}_j = r\underline{x}_j + (1 - r)\underline{z}_{j-1}, \quad j = 1, \dots, m,$$

où  $\underline{z}_0 = \underline{\mu}$  et  $0 < r \leq 1$ . La carte émet un signal à l'observation  $j$  si

$$Y_j^2 = \underline{z}_j' \Sigma_{\underline{z}_j}^{-1} \underline{z}_j > h_{ewma},$$

où  $h_{ewma}$  est la limite de contrôle et  $\Sigma_{\underline{z}_j}$  est la matrice de covariance des  $\underline{z}_j$  et est égale à

$$\Sigma_{\underline{z}_j} = [r(1 - (1 - r)^{2j}) / (2 - r)] \Sigma, \quad j = 1, \dots, m. \quad (2.4.1)$$

La carte de contrôle est construite en plaçant sur un graphique la valeur de  $Y^2$  en fonction du numéro d'observation ou du temps avec une limite de contrôle de  $h_{ewma}$ .

Tout comme pour le cas univarié, on peut prendre la variance asymptotique, c'est-à-dire,

$$\Sigma_{\hat{z}_j} = \frac{r}{2-r} \Sigma, \quad (2.4.2)$$

mais on recommande de prendre celle calculée à partir de l'équation (2.4.1).

De façon analogue à la situation pour le cas univarié, la procédure ewma multivariée est équivalente à la procédure  $T^2$  si  $r=1$ .

Les paramètres  $r$  et  $h_{ewma}$  sont choisis en fonction respectivement du changement de la moyenne que l'on veut détecter et de la  $LMS_0$  désirée. Les plus petites valeurs de  $r$  sont plus efficaces pour détecter de plus petits déplacements dans le vecteur de la moyenne et donc pour de plus petites valeurs du paramètre de décentralité. Ces paramètres sont également fonction du nombre de caractéristiques de qualité  $p$ .

Comme pour toutes les procédures, il faudra estimer  $\underline{\mu}$  et  $\Sigma$  à partir des observations de la phase un. Pour le cas des données individuelles, on estimera  $\Sigma$  en utilisant les étendues mobiles, c'est-à-dire à l'aide de l'équation (2.2.16). Si  $n > 1$ , on se servira plutôt de l'équation (2.2.4). Pour  $\underline{\mu}$ , on prendra évidemment le vecteur de la moyenne échantillonnale  $\bar{x}$ . Alors pour le cas des données individuelles, on utilisera plutôt les équations suivantes pour la procédure ewma multivariée.

$$\hat{z}_j = r x_j + (1-r) \hat{z}_{j-1} \quad \text{et} \quad \hat{Y}_j^2 = \hat{z}_j' \tilde{S}_{\hat{z}_j}^{-1} \hat{z}_j, \quad (2.4.3)$$

où  $\hat{z}_0 = \hat{\underline{\mu}}$  et  $\tilde{S}_{\hat{z}_j} = [r(1 - (1-r)^{2j}) / (2-r)] \tilde{S}$ ,  $j = 1, \dots, m$ .

On va construire un exemple de carte ewma multivariée à partir des données individuelles de la phase deux. Les estimations de  $\underline{\mu}$  et de  $\Sigma$  ont été calculées à partir des données de la phase un pour l'exemple des données individuelles de la

procédure  $T^2$ . On va calculer la valeur du vecteur  $\hat{z}_j$ , puis de la statistique  $\hat{Y}_j^2$  à l'aide de l'équation (2.4.3).

Les paramètres  $r=0,2$  et  $h_{ewma}=9,65$  ont été choisis pour obtenir une  $LMS_0=200$ . Tout comme  $T^2$  et cusum on a construit deux cartes ewma. Une qui est présentée dans la présente section (celle pour la première série d'observations) et l'autre qui est présentée dans la section comparaison (section 2.5). La figure (2.15) représente la carte de contrôle multivariée ewma pour les 10 premières observations.

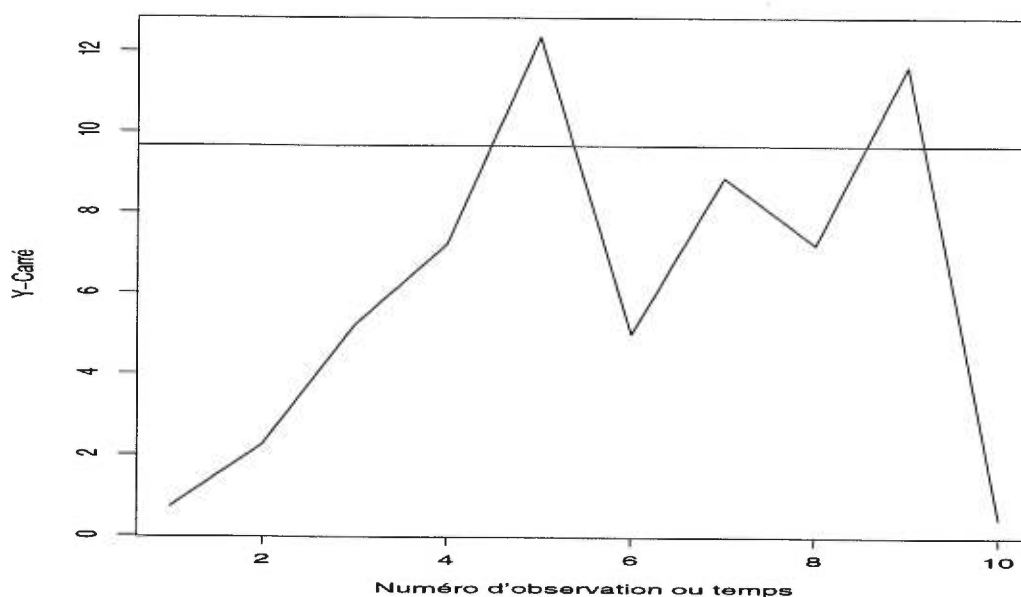


FIGURE 2.15. Carte de contrôle ewma multivariée pour 10 observations de la phase deux avec  $r=0,2$  et  $h_{ewma}=9,65$  ( $n=1$ ).

Tout comme la procédure cusum, il y a hors-contrôle aux observations 5 et 9. L'avantage de cette procédure par rapport à  $T^2$ , c'est que tout comme cusum, elle tient compte de la séquence de point entière et est donc en mesure d'accumuler les petits déplacements dans la même direction pour signaler en moyenne plus rapidement que la procédure  $T^2$ .



### 2.4.2. Paramètres et $LMS$ de la carte ewma multivariée

On doit tout d'abord mentionner que, tout comme la procédure  $T^2$  et cusum, les  $LMS$  de la procédure ewma multivariée ne dépendent que de la moyenne et de la matrice de covariance qu'à travers le paramètre de décentralité  $\tau^2$ . De plus, les  $LMS$  sont basées sur la variance asymptotique calculée à partir de l'équation (2.4.2).

Tout comme le cas univarié, de plus petites valeurs de  $r$  seront plus efficaces pour détecter de plus petits déplacements.

Le tableau 2.8 donne pour différentes valeurs de  $p$  et une  $LMS_0=200$ , les valeurs de  $r$  et de  $h_{ewma}$  à utiliser pour un déplacement  $\tau$ . Les valeurs de ces paramètres proviennent de l'article de Lowry, Woodall, Champ et Rigdon (1992).

$\tau$		$p=2$	$p=3$	$p=4$
0,5	$r$	0,06	0,06	0,06
	$h$	7,70	9,80	11,66
1	$r$	0,16	0,16	0,14
	$h$	9,35	11,52	13,34
2	$r$	0,34	0,30	0,28
	$h$	10,17	12,31	14,25

TABLEAU 2.8. Valeurs des paramètres  $r$  et  $h_{ewma}$  à utiliser pour différents déplacements  $\tau$  et différentes valeurs de  $p$  pour une  $LMS_0=200$  pour la procédure ewma multivariée.

### 2.4.3. Autres procédures ewma multivariées

D'autres auteurs ont abordé le sujet des cartes de contrôle ewma multivariées.

D'abord Sparks (1992) a appliqué l'analyse en composantes principales à la procédure ewma multivariée.

Tsui et Woodall (1993) ont par ailleurs proposé une nouvelle méthode pour la construction de la carte ewma multivariée qui est basée sur les fonctions de pertes.

Pour leur part, Prabhu et Runger (1996) et Runger et Prabhu (1996) ont travaillé sur l'utilisation de la chaîne de Markov pour l'analyse de la carte ewma multivariée.

Finalement, Prabhu et Runger (1997) fournissent des recommandations pour le choix des paramètres de la carte ewma multivariée.

## 2.5. COMPARAISONS DES TROIS CARTES

Cette section nous permettra d'établir des comparaisons entre les trois cartes multivariées.

La première constatation que l'on peut faire concerne l'interprétation des statistiques  $T^2$ , cusum et ewma. Contrairement au cas univarié, où la statistique de Shewhart était plus facilement interprétable que les statistiques cusum et ewma, pour le cas multivarié, il n'y a aucun avantage à utiliser une statistique par rapport à une autre si on se fie uniquement à l'interprétabilité. Les trois statistiques représentent une mesure de distance entre une observation multivariée et sa moyenne; l'échelle de la statistique n'est pas particulièrement facile à interpréter pour l'une ou l'autre des méthodes.

En regardant les cartes de contrôle pour ces trois procédures, on constate la même chose. La figure 2.16 représente les cartes de contrôle pour les procédures (A)  $T^2$ , (B) cusum et (C) ewma pour la deuxième série d'observations pour les données individuelles. Ces 10 observations ont subi un déplacement du vecteur des moyennes de  $1\sigma$  vers le haut pour la première variable et de  $2\sigma$  vers le haut pour la deuxième variable, pour un paramètre de décentralité de  $\tau^2=4,07$ .

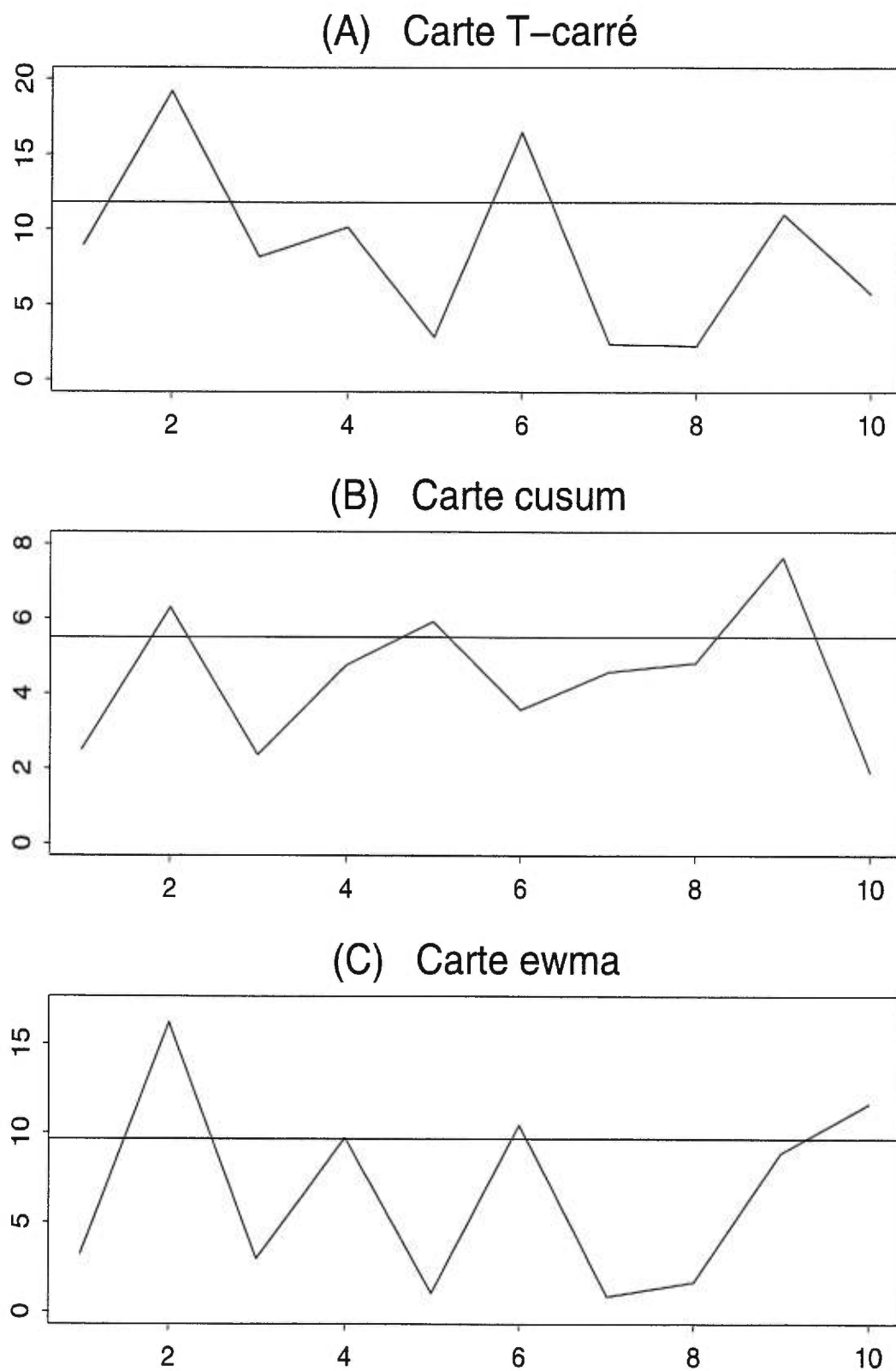


FIGURE 2.16. Cartes de contrôle pour (A)  $T^2$ , (B) cusum et (C) ewma pour 10 observations de la phase deux pour des données individuelles.

On constate donc que les trois cartes sont construites de la même façon. On place sur un graphique la valeur d'une statistique en fonction du numéro d'observation ou du temps avec une limite supérieure de contrôle. Pour les trois cartes, il y a hors-contrôle à l'observation  $j$  si la statistique pour ce point est plus grande que la limite.

Les données sont présentées au tableau 2.4. Les paramètres ont été choisis en fonction d'une  $LMS_0=200$ . Pour  $T^2$ , cela équivaut à  $\alpha=0,005$ , pour cusum on prendra alors  $k=1/2$  et  $h_{cusum}=5,50$ , tandis que pour ewma  $r=0,2$  et  $h_{ewma}=9,65$ .

On note également que la carte  $T^2$  a détecté des points hors-contrôle aux observations 2 et 6, alors que cusum a détecté des signaux aux points 2, 5 et 9 tandis que la carte ewma a signalé aux points 2, 4, 6 et 10.

Alors puisque l'interprétation des cartes est sensiblement la même, pour comparer les cartes de contrôle multivariées, on devra se fier uniquement à la performance des cartes et donc en analysant les  $LMS$  pour ces cartes.

Pour la comparaison des  $LMS_1$ , on a fixé  $p$  le nombre de variables à deux et la  $LMS_0=200$ . Ces choix sont attribuables au fait que ces calculs ont déjà été effectués par les auteurs qui ont introduit les cartes multivariées cusum et ewma.

Pour la carte  $T^2$ , les  $LMS$  ont été obtenues à l'aide des équations de la section 2.2.6, en supposant des données individuelles de la phase deux et l'estimation de  $\Sigma$  à partir des étendues mobiles à l'aide de  $m=100$  observations.

Pour cusum, on a présenté trois procédures différentes, la première qui est optimale pour un déplacement de  $\tau=1$  (cusum1), la deuxième optimale pour  $\tau=2$  (cusum2) et la dernière pour  $\tau=3$  (cusum3). Les  $LMS$  sont tirées de l'article de Crosier (1988) et sont obtenues à l'aide de simulations. Pour  $k$ , on a pris  $k=\tau/2$ . Le paramètre  $h_{cusum}$  est choisi à l'aide de l'approximation de la chaîne de Markov.

Pour ewma, on a également présenté trois procédures qui sont aussi optimales pour des déplacements de  $\tau=1, 2$  et  $3$ . On les nomme ewma1, ewma2 et ewma3. Les  $LMS$  sont tirées de l'article de Lowry, Woodall, Champ et Rigdon (1992) et sont également obtenues par simulations. Les calculs sont faits à partir de la matrice de covariance asymptotique. La valeur de la limite supérieure  $h_{ewma}$  est obtenue par simulations.

Le tableau 2.9 contient les valeurs des  $LMS_1$  correspondant aux 7 procédures décrites précédemment.

$\tau$	$T^2$	cusum1	cusum2	cusum3	ewma1	ewma2	ewma3
	$\alpha = 0,005$	$k=0,5$	$k=1$	$k=1,5$	$h=9,65$	$h=10,29$	$h=10,53$
	$LSC=11,80$	$h=5,50$	$h=2,99$	$h=1,87$	$r=0,2$	$r=0,4$	$r=0,6$
0	200	200	200	200	201	199	200
0,5	120	28,8	48,0	78,7	35,1	51,9	73,6
1,0	45,5	9,35	11,0	18,4	10,1	13,2	19,3
2,0	7,76	4,20	3,48	3,72	3,80	3,54	3,86
3,0	2,40	2,78	2,08	1,69	2,42	2,04	1,88

TABLEAU 2.9.  $LMS_1$  des cartes multivariées  $T^2$ , cusum et ewma pour une  $LMS_0=200$  et un déplacement  $\tau$  du vecteur des moyennes pour  $p=2$  et  $m=100$ .

La conclusion à tirer de ce tableau est la même qu'au niveau univarié. On s'aperçoit que les procédures cusum et ewma optimales pour des déplacements de  $\tau=3$  (cusum3 et ewma3) et  $\tau=2$  (cusum2 et ewma2) obtiennent toujours de meilleurs résultats que la procédure  $T^2$ . Il est donc tout à notre avantage d'utiliser ces procédures puisqu'en moyenne elles détectent un déplacement beaucoup plus rapidement que  $T^2$ .

Par contre, si on compare cusum1 et ewma1 à  $T^2$ , on s'aperçoit que celles-ci détectent un changement plus rapidement pour les petits déplacements, mais que pour les gros déplacements ( $\tau=3$ ), la procédure  $T^2$  réagit un peu plus rapidement (2,40 observations pour  $T^2$  comparativement à 2,78 et 2,42 pour cusum1 et ewma1 respectivement). On conclut donc qu'on a avantage à utiliser les procédures optimales pour des déplacements de  $\tau=1$ , puisque celles-ci réduisent considérablement

les  $LMS_1$  pour les petits déplacements alors qu'en moyenne cela va prendre moins d'une observation supplémentaire pour détecter les gros déplacements ( $\tau > 2$ ).

## 2.6. AUTRES CARTES ET TECHNIQUES EXISTANTES

Dans cette section on fera un bref survol des autres cartes ou techniques existantes pour le CSPM.

Tout d'abord, Patel (1973) a abordé le sujet du CSPM lorsque le vecteur d'observations provient d'une loi binomiale ou Poisson multivariée.

MacGregor et Wong (1980) nous démontrent qu'il est possible de traiter le problème du CSPM de façon plus efficace si on a une bonne connaissance du procédé. Ils utilisent des séries chronologiques multivariées et des méthodes d'identification du procédé pour développer un modèle stochastique dynamique pour un réacteur chimique.

Fuchs et Kenett (1987) présentent une nouvelle méthode basée sur une région de tolérance multivariée. Ils définissent une région couverte par les données en contrôle et on teste si une nouvelle observation est contenue dans cette région.

Hawkins (1991, 1993b) a présenté le CSPM basé sur l'ajustement des variables à partir de la régression. Il applique la méthode à la carte  $T^2$  ainsi qu'à la carte cusum. Cette méthode est recommandée si on est en mesure de connaître la structure du déplacement du vecteur des moyennes. En effet, la statistique  $T^2$  est le test optimal pour un déplacement général du vecteur des moyennes, mais si le déplacement a une certaine structure connue, le test n'est pas optimal. Hawkins (1991, 1993b) propose la carte  $T^2$  et la carte cusum basée sur le vecteur des résidus de la régression de chaque variable sur toutes les autres. Ainsi, chaque composante du vecteur des résidus représente la statistique optimale pour tester si cette variable a subi un déplacement de sa moyenne. Alors dans le cas bien précis

d'un déplacement quelconque connu d'avance, les cartes basées sur la régression obtiennent de meilleurs résultats que les cartes habituelles.

Chan et Li (1994) proposent une carte de contrôle multivariée, la carte M, pour surveiller si la moyenne d'un procédé a un comportement linéaire pendant la période où les sous-groupes sont formés en utilisant la poursuite de projection (de l'anglais *projection pursuit*).

Liu (1995) propose la construction de nouvelles cartes multivariées (r, Q et S) basées sur une approche non paramétrique en utilisant le concept de profondeur des données.

Tracy, Young et Mason (1995) utilisent l'analyse en composantes principales pour que le signal soit plus facilement interprétable pour le cas bivarié.

Nomikos et MacGregor (1995) considèrent le problème des données qui suivent une trajectoire qui varie dans le temps et qui sont obtenues tout au long d'un procédé en batch et non seulement à la fin de celui-ci sur le produit final. Ils utilisent l'analyse en composantes principales multiple pour décrire les batchs passées en contrôle ce qui par le fait même facilite l'analyse des nouvelles batchs. Leur méthode consiste à tester si on est en contrôle pendant une batch et non seulement à la fin de celle-ci en récoltant des données sur trois composantes: les variables, le temps et les batchs. En fait, on va mesurer  $J$  variables à  $K$  temps différents pendant une batch et ce sur  $I$  batchs.

Runger (1996) propose une nouvelle carte, la carte  $U^2$ , qui est sensible à des causes assignables spécifiques et qui affectent seulement un sous-ensemble de variables. Cette méthode est basée sur la projection.

Finalement, Timm (1996) propose lui aussi une méthode lorsque le déplacement affecte seulement un sous-ensemble de variables, une à la fois. Cette méthode est basée sur le test des intersections finies.

## 2.7. DÉTECTION DES VARIABLES HORS-CONTRÔLE

Depuis le début, le seul point que l'on a analysé pour juger la performance d'une carte est sa capacité à détecter une situation hors-contrôle par l'émission d'un signal sur la carte. Plus rapidement un changement sera détecté et meilleure sera la carte. Pour un procédé univarié, c'est d'ailleurs la seule caractéristique quantitative sur laquelle on se base pour juger de la performance d'une carte. Pour une procédure multivariée, on doit cependant analyser un autre aspect, il s'agit de la capacité d'une procédure à interpréter un signal et donc sa capacité à détecter quelle variable ou groupe de variables est hors-contrôle lorsqu'un signal se présente. En fait, les cartes elles-mêmes ne nous indiquent pas ces variables hors-contrôle, il faudra trouver une procédure en surplus de cette carte pour identifier les variables hors-contrôle.

Il est primordial d'être en mesure d'identifier ces variables pour pouvoir par la suite identifier les causes assignables, les éliminer et ainsi améliorer le procédé par la réduction de la variation.

Il existe d'ailleurs plusieurs méthodes pour détecter les variables hors-contrôle. Dans les prochaines sections, nous présenterons les principales en détail. Toutes ces méthodes peuvent être regroupées en trois catégories. La première qui traite des techniques qui analysent les  $p$  variables du procédé séparément, la deuxième qui discute des méthodes qui utilisent l'analyse en composantes principales et la dernière qui se base sur une décomposition de la statistique  $T^2$ . On note finalement que ces méthodes sont soit applicables aux trois cartes ( $T^2$ , cusum et ewma), soit à la carte  $T^2$  seulement.



### 2.7.1. Utilisation de $p$ procédures univariées

La première méthode à laquelle on pense et certainement la plus simple, est de suivre parallèlement à une carte multivariée, chacune des  $p$  cartes univariées. La procédure consiste alors à construire les  $p+1$  cartes simultanément et lorsqu'un signal survient sur la carte multivariée, de se référer aux cartes univariées pour savoir quelle(s) variable(s) est(sont) responsables(s) de ce signal.

Il existe alors différents choix pour le niveau critique  $\alpha_u$  des cartes univariées, qui dépendra du niveau critique  $\alpha_m$  de la carte multivariée. Le premier choix est de prendre  $\alpha_u = \alpha_m$ . Cependant, ce choix nous contraint à avoir une probabilité de fausse alarme supérieure à  $\alpha_m$ , car on utilise  $p$  procédures simultanément.

Une des méthodes proposées pour tenter de régler ce problème est d'utiliser les intervalles de Bonferroni pour contrôler le niveau critique  $\alpha_u$ . Une procédure simple proposée par Alt (1985), consiste à construire des intervalles de confiance pour chacune des  $p$  moyennes du procédé avec un niveau de confiance ajusté à  $\alpha_u = \alpha_m/p$ . Cependant, Wierda (1994) conclut que ces intervalles sont en général trop larges, on conclut ainsi que les variables sont en contrôle plus souvent qu'elles ne le sont en réalité car on s'assure d'avoir au plus

$$\alpha_m + \sum_{t=2}^p (-1)^{t-1} \binom{p}{t} \left(\frac{\alpha_m}{p}\right)^t$$

comme probabilité de fausse alarme.

Pour contrer ce problème Doganaksoy, Faltin et Tucker (1991) ont proposé l'utilisation des intervalles simultanés de Bonferroni où le volume de la région de confiance de Bonferroni est plus petite et donc avec un niveau de confiance différent de  $\alpha_m/p$ . La procédure consiste à tout d'abord effectuer le test  $T^2$  à un niveau  $\alpha_m$ . S'il y a hors-contrôle pour l'observation  $j$ , on doit calculer pour chacune des  $p$  variables, le niveau de confiance le plus petit, que l'on dénotera

$\alpha_{ind;l}$ ,  $l = 1, \dots, p$ , pour lequel l'intervalle  $\mu_l - \mu_{lj}$  contient 0 et où  $\mu_l$  et  $\mu_{lj}$  dénote respectivement la moyenne en contrôle pour la variable  $l$  et la moyenne de la variable  $l$  pour l'observation  $j$ . On calcule ensuite un niveau de confiance  $\alpha_{bonf}$  qui mène au niveau confiance désiré pour les intervalles simultanés de Bonferroni pour  $\mu_l - \mu_{lj}$ . Finalement, les variables pour lesquelles  $\alpha_{ind;l} > \alpha_{bonf}$ ,  $l = 1, \dots, p$ , sont celles qui sont susceptibles d'être responsable du signal hors-contrôle. Cette méthode n'est pas recommandée car, tout comme les précédentes, elle ignore la structure de corrélation et le choix de  $\alpha_{bonf}$  n'est pas clair car il dépend de plusieurs facteurs (incluant  $p$ , la structure de corrélation,  $\alpha$  et  $\beta$ ).

Sous les hypothèses de normalité des données, Hayter et Tsui (1994) proposent une procédure plus puissante en choisissant un  $\alpha_u$  plus grand que  $\alpha_m/p$ . Ce choix est  $\alpha_u = 1 - (1 - \alpha_m)^{1/p}$ . Cependant, même si ce choix est plus puissant, en pratique la différence est négligeable.

Il survient cependant deux problèmes reliés à l'utilisation de ces méthodes. Le premier étant que l'on ignore la corrélation qui existe entre les variables en les traitant comme si elles étaient indépendantes. Le niveau critique  $\alpha_u$ , n'est donc, pour aucune des méthodes énumérées ci-haut, égale à  $\alpha_m$ . Aussi, il se peut très bien qu'aucune variable ne soit hors-contrôle sur les cartes univariées puisque le signal n'est peut-être pas dû à une seule variable, mais bien à une combinaison de celles-ci.

On décourage donc leur utilisation comme méthode unique de détection des variables hors-contrôle, particulièrement si la corrélation est élevée. Néanmoins, elles peuvent être utilisées de façon efficace pour illustrer comment la moyenne de chaque variable se comporte.

Ce qui pourrait être utile également, serait d'avoir sur un même graphique, la carte multivariée et chacune des cartes univariées. Fuchs et Benjamini (1994)

optent pour cette méthode avec l'introduction de la carte MP qui présente simultanément sur le même graphique, les statistiques univariées et multivariée. La statistique multivariée est la statistique  $T^2$  alors que celles univariées sont les statistiques centrées réduites pour chacune des  $p$  variables.

La dernière méthode que l'on présente et qui utilise  $p$  procédures univariées est celle introduite par Hayter et Tsui (1994). Ils présentent une nouvelle procédure basée sur la construction d'intervalles de confiance simultanés et exacts pour chacune des variables. Le procédé est en contrôle tant que tous les intervalles contiennent  $\hat{\mu}_l$ , la moyenne estimée de la variable  $l$ . Cependant, si au moins un des intervalles ne contient pas  $\hat{\mu}_l$ , le procédé est hors-contrôle et on est en mesure de savoir immédiatement quelle(s) variable(s) est(sont) hors-contrôle.

L'avantage de cette méthode c'est que l'on s'assure d'obtenir un niveau critique égale à  $\alpha_m$ . La procédure est basée sur le choix d'un point critique  $C_{\Sigma, \alpha_m}$  choisi en fonction de  $\Sigma$  et  $\alpha_m$ . La constante  $C_{\Sigma, \alpha_m}$  est telle que

$$P \left( \frac{|x_l - \hat{\mu}_l|}{\Sigma_{ll}^{1/2}} \leq C_{\Sigma, \alpha_m}, l = 1, \dots, p \right) = 1 - \alpha_m$$

Le désavantage majeur de cette méthode réside dans le choix du point critique  $C_{\Sigma, \alpha}$ , qui est fonction de  $\Sigma$  connu. Pour  $p=2$ , on peut le trouver à partir de tables existantes. Pour  $p > 2$ , il devra être calculé par intégration numérique ou par approximation à partir de tables existantes. Sinon, ils peuvent être obtenus par simulation. Pour une discussion détaillée, voir Hayter et Tsui (1994).

### 2.7.2. L'analyse en composantes principales

L'analyse en composantes principales (acp) est utilisée en contrôle statistique des procédés multivariés pour répondre principalement à deux buts. Le premier

est pour réduire la dimensionalité du problème alors que le deuxième est pour détecter les variables hors-contrôle.

Tout d'abord, l'ACP a pour but de réduire la dimensionalité du problème en trouvant de nouvelles variables non corrélées (les composantes principales) qui sont des combinaisons linéaires des variables originales. Il peut donc être intéressant (si  $p$  est élevé, par exemple  $p=10$  et plus) de réduire le problème en ne conservant que les 4 ou 5 nouvelles variables qui expliquent le plus la variabilité du procédé.

Géométriquement, ces combinaisons linéaires représentent la sélection d'un nouveau système de coordonnées obtenu en appliquant une rotation au système de coordonnées originales. Ces nouveaux axes représentent les directions avec le plus de variabilité.

Comme on va le voir, l'ACP dépend entièrement et exclusivement de la structure de covariance ou de corrélation entre les variables et ne requiert pas l'hypothèse de normalité des données.

La procédure de l'analyse en composantes principales consiste à décomposer la matrice de covariance du procédé en vecteurs propres et valeurs propres à l'aide de l'équation matricielle suivante,

$$\Sigma = A \Lambda A', \quad (2.7.1)$$

où  $AA' = I$  est la matrice identité  $p \times p$ ,  $\Lambda$  est une matrice diagonale et les éléments de la diagonale sont en ordre décroissant.

Les colonnes de  $A$  sont les vecteurs propres dénotés par  $q_c$ ,  $c = 1, \dots, p$ , alors que les éléments sur la diagonale de  $\Lambda$ ,  $\lambda_c$ ,  $c = 1, \dots, p$ , sont appelés les valeurs propres de  $\Sigma$ . Les éléments qui ne sont pas sur la diagonale sont égaux à 0.

Ainsi, les nouvelles variables, c'est-à-dire les  $p$  composantes principales (cp) sont données par,

$$w_1 = \underline{a}'_1 \underline{x}, \dots, w_p = \underline{a}'_p \underline{x} \quad \text{où} \quad \underline{w} = A\underline{x}. \quad (2.7.2)$$

En utilisant (2.7.1), on voit que la matrice de covariance des cp est  $\Lambda$ . La variance de  $w_c$  est donc  $\lambda_c$  et puisque les valeurs propres sont en ordre décroissant, la variance des différentes cp est de plus en plus petite. De plus, les éléments hors de la diagonale de  $\Lambda$  étant nuls, les cp sont non corrélées. Lorsque  $\Sigma$  est inconnu, on doit appliquer la méthodologie sur un estimateur comme  $\tilde{S}$ .

Pour réduire la dimensionalité du problème, la méthode consiste à ramener les  $p$  variables du procédé à un nombre inférieur, disons  $p'$  à l'aide des composantes principales en ne gardant que celles qui expliquent le plus la variabilité du procédé. Pour ce faire, on peut tester si les  $p'$  premières cp comptent pour au moins  $100(1-\delta)$  % de la variation totale. On va donc tester

$$H_0 : \frac{\lambda_{p'+1} + \dots + \lambda_p}{\lambda_1 + \dots + \lambda_p} \geq \delta \quad \text{contre} \quad H_1 : \frac{\lambda_{p'+1} + \dots + \lambda_p}{\lambda_1 + \dots + \lambda_p} < \delta, \quad (2.7.3)$$

où  $\lambda_c, c = 1, \dots, p$  sont les valeurs propres de la matrice  $\tilde{S}$ . Anderson (1984) suggère que  $\delta$  soit fonction de  $p$  et donc que pour  $p = 3, \dots, 20$ ,

$$\delta = -0,37 + 0,33\sqrt{p} - 0,037p.$$

Si on rejette  $H_0$ , la procédure qui consiste à ne prendre que les  $p'$  premières cp est adéquate.

À partir des  $p'$  composantes principales que l'on a conservées, on a le choix d'appliquer la procédure  $T^2$  habituelle ou d'utiliser les intervalles de Bonferroni sur les nouvelles variables non corrélées.

On doit cependant noter que contrairement à la procédure  $T^2$ , l'acp n'est pas indépendante de l'échelle utilisée pour les variables. En effet, les résultats de l'acp seront différents en effectuant l'acp sur les variables originales ou sur les

variables ayant subi un changement d'échelle. En particulier, l'acp appliquée à la matrice de covariance ou à la matrice de corrélation ne donne pas les mêmes résultats.

Si l'échelle des variables n'est pas la même, on recommande habituellement de normaliser les données en posant

$$y_{lj} = \frac{(x_{lj} - \bar{x}_l)}{\tilde{s}_l}, \quad l = 1, \dots, p \text{ et } j = 1, \dots, m, \quad (2.7.4)$$

où  $x_{lj}$  est l'observation  $j$  pour la variable  $l$ ,  $\bar{x}_l$  représente l'estimation de la moyenne pour la variable  $l$  et  $\tilde{s}_l$  représente l'estimation de l'écart type pour la variable  $l$  et correspond à la racine carrée de l'élément  $(l, l)$  de la matrice  $\tilde{S}$ .

Par la suite, on applique la procédure de l'acp sur les nouvelles variables normalisées en décomposant la matrice de covariance des variables normalisées  $y_{lj}$ .

Cependant, il peut arriver que l'échelle utilisée doive être conservée en raison de sa signification. Il faut donc analyser les variables pour savoir si l'échelle utilisée a une importance quelconque et appliquer l'acp en fonction de ce critère.

Aussi si  $p=2$  ou  $3$ , il est possible de faire le graphique de la région de contrôle identifiée avec le nouveau système d'axes et de vérifier ce qui se passe point par point (ce que nous verrons à partir de notre exemple pour  $p=2$ , à la fin du chapitre 2).

Pour une discussion détaillée de l'acp appliquée à la carte  $T^2$ , voir Jackson (1985). Pour l'acp appliquée aux procédures cusum et ewma, on doit se référer à l'article de Sparks (1992).

On vient donc de présenter la procédure pour appliquer l'analyse en composantes principales au CSP. Ainsi, on va créer de nouvelles variables et les signaux sur les cartes de contrôle nous indiqueront lesquelles de ces nouvelles variables

sont hors-contrôle. Mais si en appliquant cette procédure, on sait que la 1<sup>ère</sup> composante principale qui est une combinaison linéaire des variables  $x_1, x_2$  et  $x_3$  est hors-contrôle, cela ne nous indique pas quelle(s) variable(s) est(sont) hors-contrôle. Ainsi, l'acp comme on vient de le voir est surtout utile pour réduire, si possible, le nombre de variables. Pour la détection des variables hors-contrôle, on doit pousser l'analyse un peu plus en profondeur.

Ainsi, Kourti et MacGregor (1996) proposent une méthode pour répondre au deuxième but de l'acp et donc pour véritablement détecter les variables hors-contrôle à l'aide de l'acp.

On doit tout d'abord savoir que la statistique  $T^2$  calculée sur les variables originales est équivalente à la statistique  $T^2$  calculée sur les scores lorsque toutes les composantes principales sont utilisées, où le score de la composante principale  $c$  est donné par

$$t_c = \underline{a}'_c(\underline{x} - \bar{\underline{x}}) \quad c = 1, \dots, p.$$

La statistique  $T^2$  peut donc être écrite en fonction des composantes principales de la façon suivante,

$$T^2 = \sum_{c=1}^p \frac{t_c^2}{\lambda_c}.$$

La procédure consiste à effectuer une carte multivariée comme à l'habitude et lorsqu'il y a un point hors-contrôle, disons le point  $j$ , on calcule les scores normalisés des  $p$  composantes principales pour ce point, qui sont

$$\frac{t_c}{\lambda_c^{1/2}}, \quad c = 1, \dots, p, \quad (2.7.5)$$

où  $\underline{a}_c$  est le vecteur propre et  $\lambda_c$  est la valeur propre pour la composante principale  $c$ .

Pour ce point hors-contrôle, on détecte alors les scores normalisés élevés à l'aide d'une limite supérieure. On pourrait prendre la limite de Bonferroni puisque ces scores sont non corrélés. Cependant, lorsqu'une carte multivariée a détecté

un point hors-contrôle, il n'y a plus lieu de se préoccuper de l'erreur de type 1 pour obtenir des limites de contrôle précises à  $\alpha_m$  sur les cartes univariées, un changement de moyenne au niveau désiré  $\alpha_m$  ayant déjà été détecté sur la carte multivariée. Les cartes univariées doivent servir uniquement à nous aider à décider quelles variables ont causé ce changement. Ainsi, Kourti et MacGregor (1996) proposent de prendre une limite égale à 2,5 pour détecter les scores normalisés élevés.

Maintenant pour les scores normalisés qui sont supérieurs à la limite, on utilise des graphiques de contribution pour trouver quelle variable ou groupe de variables est responsable du signal. Ce genre de graphique indique comment chaque variable impliquée dans un score contribue à celui-ci. Les variables qui ont les contributions les plus élevées sont susceptibles d'être responsables du signal. Ainsi la contribution de la variable  $l$  au score de la composante principale  $c$  est

$$a_{cl} (x_l - \bar{x}_l), \quad c = 1, \dots, p, \quad l = 1, \dots, p, \quad (2.7.6)$$

où  $a_{cl}$  est l'élément  $l$  du vecteur  $q_c$ .

On prend en considération les variables avec les contributions les plus élevées mais du même signe que le score. C'est-à-dire que si le score est positif on ne garde que les contributions positives et vice-versa. Cela est dû au fait qu'une contribution négative n'aide en rien un score positif à être hors-contrôle.

Souvent plus d'un score est élevé. On suggère alors de calculer une contribution moyenne globale par variable pour tous les scores avec des valeurs supérieures à la limite de 2,5. Ainsi pour ces scores élevés, on calcule la contribution en ne gardant que ceux du même signe que le score, puis on somme les valeurs sur les  $k$  scores élevés. On procède ainsi. On calcule la contribution de chaque variable,

$$cont_{cl} = \frac{t_c}{\lambda_c} a_{cl} (x_l - \bar{x}_l), \quad c = 1, \dots, p, \quad l = 1, \dots, p, \quad (2.7.7)$$



pour les  $k$  scores élevés. Si  $cont_{cl} < 0$ , on l'égalé à 0 pour ne garder que ceux du même signe que le score. Puis on somme pour la variable  $l$ , c'est-à-dire

$$cont_l = \sum_{c=1}^k (cont_{cl}).$$

Les variables avec les contributions moyennes globales les plus élevées doivent être prises en considération.

Évidemment encore ici, l'échelle influence sensiblement les résultats obtenus, on recommande encore une fois de normaliser les données si l'échelle utilisée n'a pas vraiment de signification particulière.

### 2.7.3. Décomposition de la statistique $T^2$

La dernière méthode que l'on présente pour détecter les variables hors-contrôle, est celle qui utilise une décomposition de la statistique  $T^2$ . On présentera tout d'abord la méthode plus générale présentée par Mason, Tracy et Young (1995) et on montrera que la méthode de Murphy (1987) est un cas particulier de cette méthode.

Mason, Tracy et Young (1995) ont décomposé la statistique  $T^2$  en  $p$  composantes indépendantes où chacune d'elles donnent de l'information sur les variables qui contribuent de façon significative au signal hors-contrôle. Ils suggèrent de décomposer la statistique  $T^2$  de la manière suivante,

$$\begin{aligned} T^2 &= T_1^2 + T_{2.1}^2 + T_{3.1,2}^2 + \dots + T_{p.1,2,\dots,p-1}^2 \\ &= T_1^2 + \sum_{l=1}^{p-1} T_{l+1.1,2,\dots,l}^2 \end{aligned} \quad (2.7.8)$$

où  $T_1^2$  est le carré de la statistique  $t$  univariée pour la première variable et  $T_{l+1.1,2,\dots,l}^2$  est le carré de la statistique  $t$  de la variable  $l + 1$  ajustée par les estimés de la moyenne et de l'écart-type de la distribution conditionnelle de la variable  $l + 1$  étant donné les  $l$  premières variables. La question est maintenant

de savoir comment utiliser l'information donnée par les différents termes de la décomposition.

Il y a  $p$  termes dans le calcul de la décomposition donnée par (2.7.8) et ceux-ci sont tous de signe positif. Ils contribuent donc tous à faire augmenter la valeur de la statistique  $T^2$ . Ainsi, si la statistique est hors-contrôle pour une observation, il est possible de déterminer quel(s) terme(s) de la décomposition est(sont) le plus responsable(s) de cette augmentation de la statistique  $T^2$ .

On doit noter que la décomposition donnée par (2.7.8) n'est pas unique. En fait, il existe  $p!$  différentes façons de partitionner la statistique  $T^2$ , dépendant de l'ordre avec lequel on conditionne les variables. Ainsi, cette procédure peut être intéressante si on a une idée à priori de l'ordre des variables car on n'aura que  $p$  termes conditionnels à calculer. Sinon, il y a  $p!$  ordres arbitraires qui peuvent mener à des conclusions différentes selon l'ordre.

Alors si on n'a pas d'indication particulière sur l'ordre, les auteurs proposent de calculer les  $p$  termes non conditionnels pour chacune des variables et les  $p$  termes conditionnels d'une variable étant donné toutes les autres. Ces  $2p$  termes seront comparés à une limite de contrôle à l'aide de la distribution de Fisher, qui est

$$\left(\frac{m+1}{m}\right) F_{\alpha,1,n-1}.$$

Si le terme non conditionnel de la variable  $l$  est plus grand que cette limite, cela indique que cette variable est à l'extérieur de ces limites univariées. Si le terme conditionnel qui représente la variable  $l$  étant donné toutes les autres est supérieur à cette même limite, on conclut qu'il y a un changement dans la relation qu'a la variable  $l$  avec les  $(l-1)$  autres variables.

Cette méthode quoique bien développée par les auteurs et utilisée par plusieurs autres de façon indirecte, n'est pas recommandée en raison du trop grand nombre de décompositions possibles pour la statistique  $T^2$  (il y en a  $p!$ ).

Une approche qui utilise indirectement la décomposition de la statistique  $T^2$  est celle proposée par Murphy (1987). Cette approche, recommandée par Wierda (1994), suppose tout d'abord que pour l'observation  $j$ , on ait un signal hors-contrôle à partir de la carte  $T^2$ . Par la suite, on doit partitionner le vecteur d'observation  $\underline{x}_j$  en deux parties. Un premier sous-ensemble de  $q$  variables, le vecteur  $\underline{x}_j^{(1)}$  que l'on suspecte être responsable du signal et un deuxième sous-ensemble de  $(p - q)$  variables, le vecteur  $\underline{x}_j^{(2)}$ .

On est alors en mesure de calculer deux statistiques  $T^2$ , une calculée à partir des  $p$  variables, que l'on dénotera  $T_p^2$  et qui nous a permis de juger l'observation hors-contrôle et une autre qui dénote la distance réduite correspondant au sous-ensemble de  $q$  variables, que l'on dénotera par  $T_q^2$ .

Pour tester si le sous-ensemble de  $q$  variables est responsable du signal, on doit calculer la statistique  $D_q$  où

$$D_q = \frac{T_p^2 - T_q^2}{(m + 1)(n - 1) + T_q^2},$$

que l'on compare à la limite supérieure de contrôle suivante,

$$LSC_q = \frac{p - q}{m(n - 1) - p + 1} F_\alpha(p - q, m(n - 1) - p + 1).$$

Si  $D_q > LSC_q$ , on rejette  $H_0$  et on conclut donc que le sous-ensemble de  $q$  variables n'est pas responsable du signal.

Puisqu'il est pratiquement impossible de suspecter un sous-ensemble de variables d'être hors-contrôle, le problème majeur avec cette méthode est le nombre de sous-ensemble possibles, qui dépend de  $p$  et qui est égale à  $2^p - 2$ . Si  $p=4$ , on

obtient 14 possibilités. On doit donc appliquer la méthode que l'on vient de voir aux  $2^p - 2$  possibilités.

Tout comme pour la méthode de décomposition, cette procédure n'est pas recommandée en raison du trop grand nombre de possibilités.

Comme on l'a dit précédemment, la méthode de Murphy (1987) est un cas particulier de la méthode de décomposition de la statistique  $T^2$ . En effet, Mason, Tracy et Young (1995) ont démontré que la statistique  $D_q$  se décompose de la manière suivante,

$$D_q = \sum_{l=q+1}^p T_{l,1,\dots,l-1}^2,$$

qui représente alors une somme de  $(p - q)$   $T^2$  conditionnels.

#### 2.7.4. Choix d'une procédure pour détecter les variables hors-contrôle

Dans cette section, on va présenter brièvement la procédure que l'on recommande pour la détection des variables hors-contrôle et également présenter un exemple à partir de données. À la lumière de ce qui a été présenté dans les trois sections précédentes, voici la démarche que l'on propose.

La première étape consiste à effectuer une carte multivariée avec un niveau de confiance égal à  $\alpha_m$ . Pour les points hors-contrôle, et seulement pour ceux-ci, on regardera sur les cartes univariées pour savoir ce qui se passe.

On se doit donc de construire simultanément les  $p + 1$  cartes de contrôle, mais de regarder les cartes univariées seulement pour les points hors-contrôle, cela nous assure ainsi de conserver un niveau critique global de  $\alpha_m$ . Pour le niveau critique des cartes univariées, on propose de prendre  $\alpha_u = \alpha_m / p$ . Ces cartes vont nous fournir une certaine indication de ce qui se passe avec les variables. Pour les

points hors-contrôle, les variables qui sont supérieures à la limite de Bonferroni sont considérées comme étant responsables du signal. Cependant, comme on l'a déjà mentionné cette méthode présente certaines lacunes car il se peut très bien qu'aucune variable ne soit hors-contrôle car le signal est peut-être dû à la combinaison de plusieurs des variables.

Ainsi, on propose de calculer les statistiques de Student pour chacune des variables pour être en mesure de savoir laquelle ou lesquelles se sont le plus éloignées de leurs moyennes respectives. Cela revient à normaliser, ce qui est fait à partir de l'équation (3.2.6).

Finalement, on propose en plus d'appliquer l'analyse en composantes principales, principalement la méthode de Kourti et MacGregor (1996).

On est donc prêt pour un exemple. Pour ce faire, on prendra les exemples sur lesquels on a déjà travaillé pour  $n=1$ . Ainsi pour le premier exemple, les 10 observations avaient subi un déplacement de  $1\sigma$  vers le bas pour la première variable et de  $1\sigma$  vers le haut pour la deuxième variable. Ces observations sont présentées au tableau 2.3. La carte  $T^2$  pour ces 10 observations, avec  $\alpha_m=0,005$ , avait donné un signal à l'observation 6.

Puisque l'on a un signal sur la carte multivariée, on va aller voir les cartes univariées de Shewhart des deux variables pour savoir ce qui se passe au point 6. On prendra  $\alpha_u=\alpha_m/p=0,005/2$ , ce qui donne un  $L=3,02$ .

La figure 2.17 représente les cartes individuelles de Shewhart pour les données de l'exemple 2.1. On voit que le point 6 n'est pas hors-contrôle pour aucune des deux variables. Cependant, on doit mentionner que même si celui-ci n'est pas hors-contrôle, un expert n'aura pas de difficulté à identifier un changement de moyenne pour la variable 2 en analysant la carte à l'aide de règles de décision supplémentaires. Mais si on regarde uniquement le point 6, celui-ci n'étant pas

hors-contrôle, il est donc impossible de déclarer une variable responsable du signal à cette étape-ci.

Calculons donc la valeur des variables normalisées pour ce point. Pour la variable 1, on obtient -2,04 et pour la variable 2, on a 2,08. On peut donc supposer que les deux variables sont également responsable du signal à l'observation 6.

On va maintenant pousser notre analyse un peu plus loin, en appliquant l'analyse en composantes principales et plus précisément la méthode de Kourti et MacGregor (1996).

Il serait bon tout d'abord de faire le graphique de l'ellipse que l'on a déjà effectué, mais en rajoutant les nouveaux axes des composantes principales. La figure 2.18 représente justement ce nouveau système d'axe composé à partir des deux composantes principales.

On doit mentionner que puisque les deux variables sont sensiblement sur la même échelle, on a effectué l'acp sur la matrice de covariance des observations de la phase un. On a obtenu les deux cp suivantes:

$$w_1 = -0,809x_1 - 0,588x_2 \quad \text{et} \quad w_2 = 0,588x_1 - 0,809x_2,$$

qui ont comme valeurs propres correspondantes,  $\lambda_1=10,63$  et  $\lambda_2=5,30$ .

La première cp, celle qui explique le plus de la variabilité, c'est-à-dire 66,7 %, est celle représentée par l'axe du sud-ouest au nord-est. La deuxième cp, celle qui explique 33,3 % de la variabilité est représentée par l'axe du sud-est au nord-ouest.

On s'aperçoit donc que le point 6 semble être hors-contrôle sur la 2e composante principale. Ce que nous vérifierons à l'aide de la méthode de Kourti et MacGregor (1996).

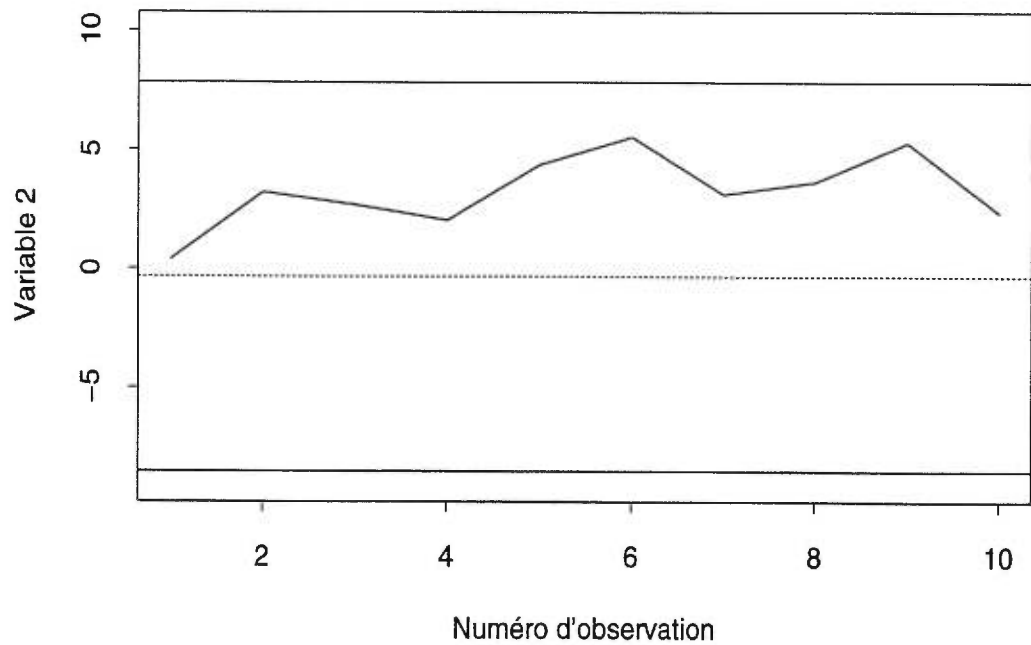
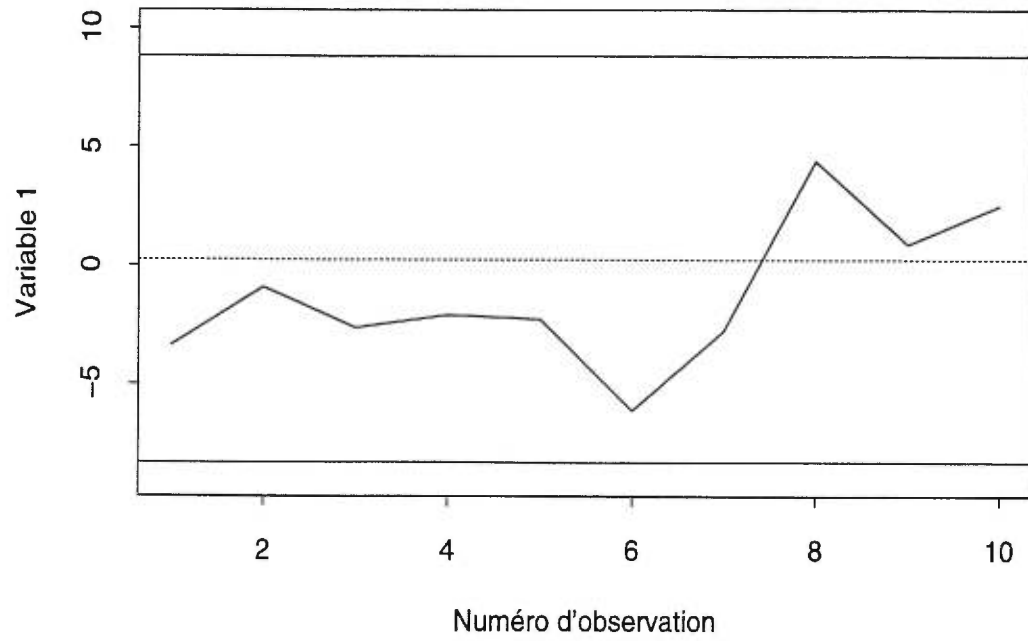


FIGURE 2.17. Cartes de contrôle individuelles de Shewhart pour les deux variables de l'exemple 2.1 ( $L=3,02$ ).

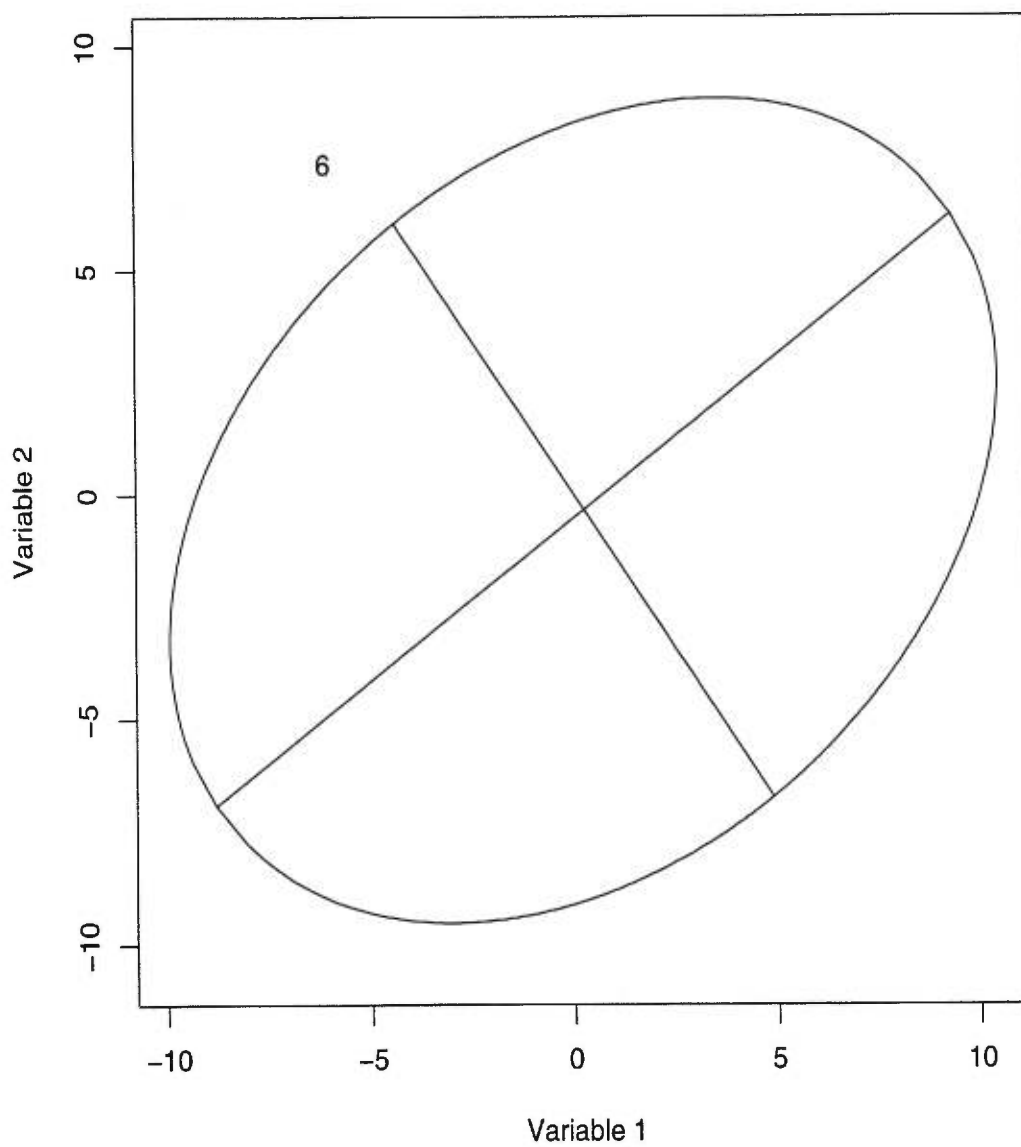


FIGURE 2.18. Représentation de l'analyse en composantes principales pour  $p=2$  (6 représente le point 6 de l'exemple 2.1).



Pour le point hors-contrôle 6, on doit tout d'abord calculer les scores normalisés des deux composantes principales à l'aide de l'équation (2.7.5). On obtient 0,52 pour la première cp et -3,68 pour la deuxième. On doit comparer ces valeurs avec la limite de 2,5. La deuxième cp est donc hors-contrôle et on doit calculer la contribution de chaque variable à ce score à l'aide de l'équation (2.7.7). On obtient 5,99 pour la contribution de la première variable et 7,58 pour la deuxième variable.

On conclut donc que les deux variables sont responsables du signal à l'observation 6 car les deux contributions sont élevées pour le score hors-contrôle de la deuxième composante principale. Cette conclusion est en accord avec le déplacement réel que les variables avaient subi (c'est-à-dire  $+1\sigma$  pour la première et  $-1\sigma$  pour la deuxième).

Reprenons maintenant la même procédure avec les données de l'exemple 2.2. Ces données avaient subi un déplacement de  $1\sigma$  vers le haut pour la première variable et de  $2\sigma$  vers le haut pour la deuxième. La carte  $T^2$  pour ces 10 observations, avec  $\alpha_m=0,005$ , avait donné des signaux aux observations 2 et 6.

On va donc construire les cartes univariées avec  $\alpha_u=0,005/2$  pour les deux variables. La figure 2.19 représente ces cartes. On voit ainsi que le point 2 et 6 sont en contrôle pour la première variable mais hors-contrôle pour la deuxième. Ceci s'explique par le fait que la première variable a subi un déplacement de seulement  $1\sigma$  alors que la deuxième en a subi un de  $2\sigma$ .

Voyons ce qui arrive si on applique la procédure de Kourti et MacGregor. Pour le point hors-contrôle 2, les scores normalisés des deux composantes principales sont respectivement -3,04 et -3,16. Puisque ces deux scores, en valeur absolue, sont supérieurs à 2,5, on va calculer une contribution moyenne globale pour les deux variables. La contribution globale de la variable 1 aux scores hors-contrôle 1 et 2 est de 2,83 et celle de la variable 2 est de 19,41.

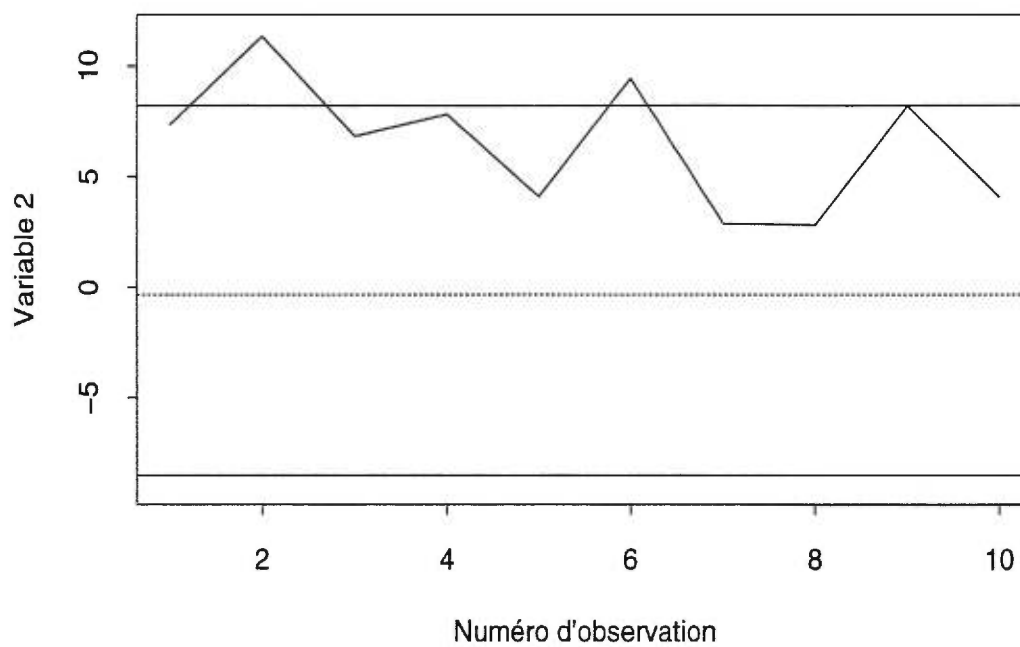
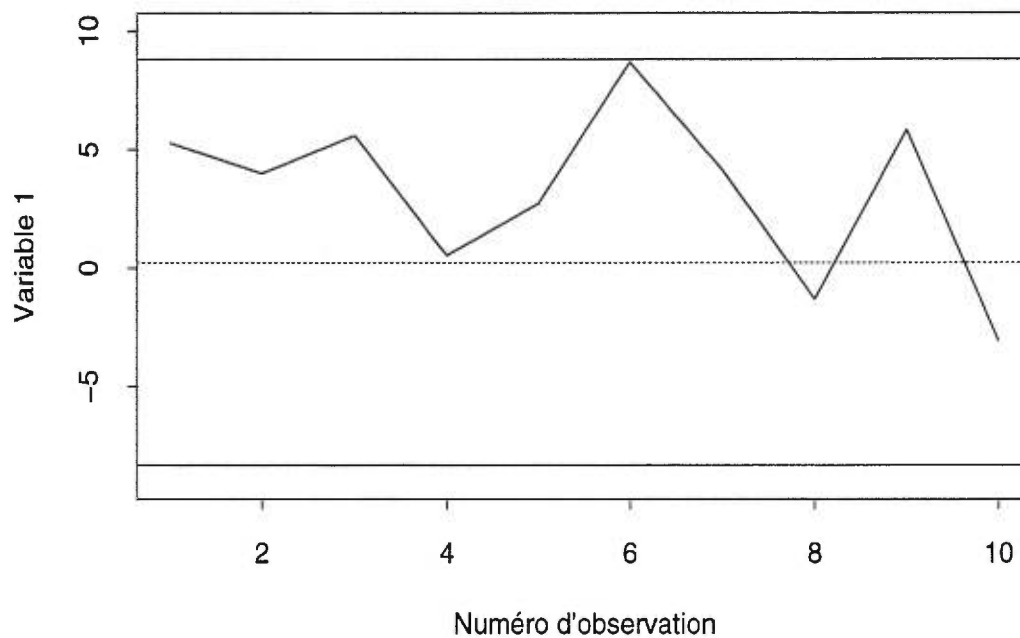


FIGURE 2.19. Cartes de contrôle individuelles de Shewhart pour les deux variables de l'exemple 2.2 ( $L=3,02$ ).

Il est donc clair que la variable deux est hors-contrôle car sa contribution est très élevée. Cette variable avait d'ailleurs subi un déplacement de deux sigmas de sa moyenne. On ne peut cependant déclarer la variable 1 responsable du signal puisque sa contribution n'est pas élevée.

Maintenant, pour le point hors-contrôle 6, les scores normalisés des deux composantes principales sont respectivement -3,86 et -1,28. La première cp est donc hors-contrôle et on va calculer la contribution de chaque variable. On obtient respectivement 8,09 et 6,80. On considère donc les deux variables responsables du signal à l'observation 6.

## 2.8. DONNÉES AUTO-CORRÉLÉES

Tous les résultats obtenus précédemment l'on été sous l'hypothèse d'indépendance entre les observations. Qu'arrive-t-il si cette hypothèse n'est pas respectée? Supposons donc que les observations sont auto-corrélées dans le temps et qu'elles suivent un certain modèle de série chronologique multivariée. De façon plus spécifique, nous supposons que les observations suivent un modèle  $AR(1)$  multivarié. On a donc

$$(\underline{x}_j - \underline{\mu}) = \Phi (\underline{x}_{j-1} - \underline{\mu}) + \underline{\epsilon}_j, \quad j = 2, \dots, m, \quad (2.8.1)$$

où  $\underline{\epsilon}_j$  est de loi normale multivariée de dimension  $p$  avec une moyenne de  $\underline{0}$  et une matrice de covariance égale à  $\Sigma_\epsilon$  et où

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1p} \\ \phi_{21} & \phi_{22} & \dots & \phi_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{p1} & \phi_{p2} & \dots & \phi_{pp} \end{pmatrix}. \quad (2.8.2)$$

Pour le cas univarié, le modèle  $AR(1)$  est le suivant,

$$(x_j - \mu) = \phi (x_{j-1} - \mu) + \epsilon_j, \quad j = 2, \dots, m, \quad (2.8.3)$$

où  $\epsilon_j \sim N(0, \sigma_\epsilon^2)$ .

En faisant l'hypothèse que les observations proviennent d'un modèle  $AR(1)$ , voyons ce qu'il advient avec les estimateurs de  $\Sigma$ . Les calculs seront effectués pour le cas univarié, mais ils se généralisent au cas multivarié. En utilisant (2.8.3), on a

$$Var(x) = \sigma_x^2 = \frac{\sigma_\epsilon^2}{1 - \phi^2} \quad \text{et} \quad Var(\bar{x}) = \frac{\sigma_x^2 (1 + \phi^2)}{n (1 - \phi^2)} + o(1/n),$$

où  $o(1/n)$  est un reste d'ordre  $1/n$ . Ainsi, les estimateurs de  $\sigma^2$ , qui pour le cas univarié sont de la forme

$$\tilde{S} = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 \quad \text{et} \quad S = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

ont comme espérance sous l'hypothèse du Modèle  $AR(1)$ ,

$$E(\tilde{S}) = \frac{\sigma_\epsilon^2}{1 + \phi} = (1 - \phi) Var(x)$$

et

$$E(S) \approx \frac{\sigma_\epsilon^2}{1 - \phi^2} \left( 1 - \frac{2\phi^2}{(n-1)(1 - \phi^2)} \right) = Var(x) \left( 1 - \frac{2\phi^2}{(n-1)(1 - \phi^2)} \right).$$

On s'aperçoit alors que l'espérance de l'estimateur  $\tilde{S}$ , qui sous l'hypothèse d'indépendance est égale à la variance des observations  $x$ , est grandement affectée par l'auto-corrélation. En fait, plus l'auto-corrélation est élevée et plus  $E(\tilde{S})$  sera petite. Pour  $S$ , l'espérance sera approximativement toujours égale à  $Var(x)$ .

En fait, le calcul a été démontré sous les hypothèses du modèle de série chronologique  $AR(1)$ , mais il se généralise à tout modèle de série chronologique  $ARMA(p, q)$ . Pour une référence à ce sujet, voir Wei (1990). L'espérance de  $\tilde{S}$  sera affectée par l'auto-corrélation et sera fonction du modèle spécifié. Par contre, l'espérance de  $S$  sera approximativement toujours égale à la variance des observations et ce quelque soit le modèle spécifié.

Pour ces raisons, pour des données auto-corrélées, on privilégiera  $S$  à  $\tilde{S}$  pour estimer la matrice de covariance du procédé.

La carte de contrôle que l'on utilisera sera elle aussi influencée par l'auto-corrélation. En effet, comme on vient de le voir l'espérance de  $S$  est approximativement égale à la variance des observations  $x$ . Or, pour la carte  $T^2$ , c'est justement cette variance que l'on doit estimer et que l'on a besoin dans le calcul de la statistique  $T^2$ . Cependant, pour ewma, on a besoin de la variance de la statistique  $z_j$ . Dans le cas univarié, et sous l'hypothèse d'indépendance, cette variance est égale à

$$Var(z_j) = \frac{\sigma_x^2}{2-r}(1 - (1-r)^{2j}),$$

et elle est donc facile à calculer. Sous l'hypothèse d'auto-corrélation, cette variance devient plus complexe à calculer et dépend du modèle que l'on va ajuster à nos données. Sous l'hypothèse du modèle  $AR(1)$ , on peut démontrer que

$$Var(z_j) = \frac{r^2\sigma_\epsilon^2}{1-\phi^2} \left( \sum_{t=0}^{j-1} (1-r)^{2t} + 2 \sum_{x=1}^{j-1} \sum_{t=1}^{j-x} \phi^t (1-r)^{t+2(x-1)} \right).$$

On voit donc que l'expression de  $Var(z_j)$  est plus complexe et qu'elle dépend du modèle de série chronologique. Pour la carte cusum, on obtient la même conclusion.

Alors pour ces raisons, on privilégiera la carte  $T^2$  à la carte ewma ou cusum si on est en présence de données auto-corrélées.

## CHAPITRE 3

---

### APPLICATION À UN JEU DE DONNÉES

Dans ce dernier chapitre, nous appliquerons les techniques présentées aux chapitres précédents à des données réelles. On expliquera tout d'abord quelles sont les variables impliquées et comment les jeux de données ont été obtenus. On présentera quelques statistiques descriptives de ces jeux. Par la suite, nous appliquerons la carte de contrôle multivariée à ces jeux de données. Finalement, on verra quelles variables sont responsables des signaux obtenus sur la carte multivariée.

#### 3.1. PRÉSENTATION DES VARIABLES ET DU JEU DE DONNÉES

Les données obtenues proviennent d'un procédé industriel continu du domaine des pâtes et papiers. Les données ont été obtenues sur 4 variables et elles se rapportent à la blancheur du papier. Ces 4 variables sont la brillance (identifiée par BR) et les trois composantes L, A et B qui servent à décrire la couleur du papier. La variable BR est une mesure de la réflexion de la lumière bleue, alors que L représente une mesure de la blancheur, 100 indiquant un blanc parfait et 0 un noir absolu. La variable A est quant à elle une mesure du degré de vert par rapport au rouge, allant du vert pour les valeurs négatives au rouge pour les valeurs positives. De la même façon, B représente le compromis entre le bleu (valeurs positives) et le jaune (valeurs négatives).

Les données sont les valeurs obtenues sur le procédé à toutes les trois heures et elles représentent des données individuelles et donc avec  $n=1$ . Nous avons reçu trois jeux de données différents. Nous allons travailler avec seulement deux de ceux-ci car les constatations et les conclusions tirées à partir des deux premiers englobent celles pour le troisième jeu. Le premier jeu contient  $m=201$  observations tandis que pour le deuxième,  $m=271$ . Nous appliquerons les procédures multivariées vues au chapitre 2 à ces deux jeux de données.

### 3.2. ANALYSE DU JEU DE DONNÉES

Dans cette section, on fera une analyse des données. On se concentrera principalement sur le premier jeu de données, que l'on nommera jeu1. Ce jeu contient 201 observations. Les constatations sur ce jeu sont également valables pour le jeu2 et le jeu3.

La figure 3.20 représente, pour chacune des 4 variables du procédé, les 201 observations du jeu1. La première constatation que l'on doit faire concerne l'auto-corrélation des données. En effet, celle-ci est très élevée pour chacune des variables.

On constate ainsi que pour une variable donnée, les observations ne sont pas indépendantes dans le temps. La ligne représente une courbe ajustée à l'aide d'une méthode de lissage (fonction *lowess* du logiciel S-PLUS). Pour une explication plus détaillée de la méthode, voir Cleveland (1979). Cette courbe peut nous être utile pour savoir si la série est stationnaire ou non. Ici, la série semble stationnaire, c'est-à-dire que la moyenne semble être à peu près la même pour toutes les observations et ce pour au moins 3 des 4 variables (mis à part la variable A).

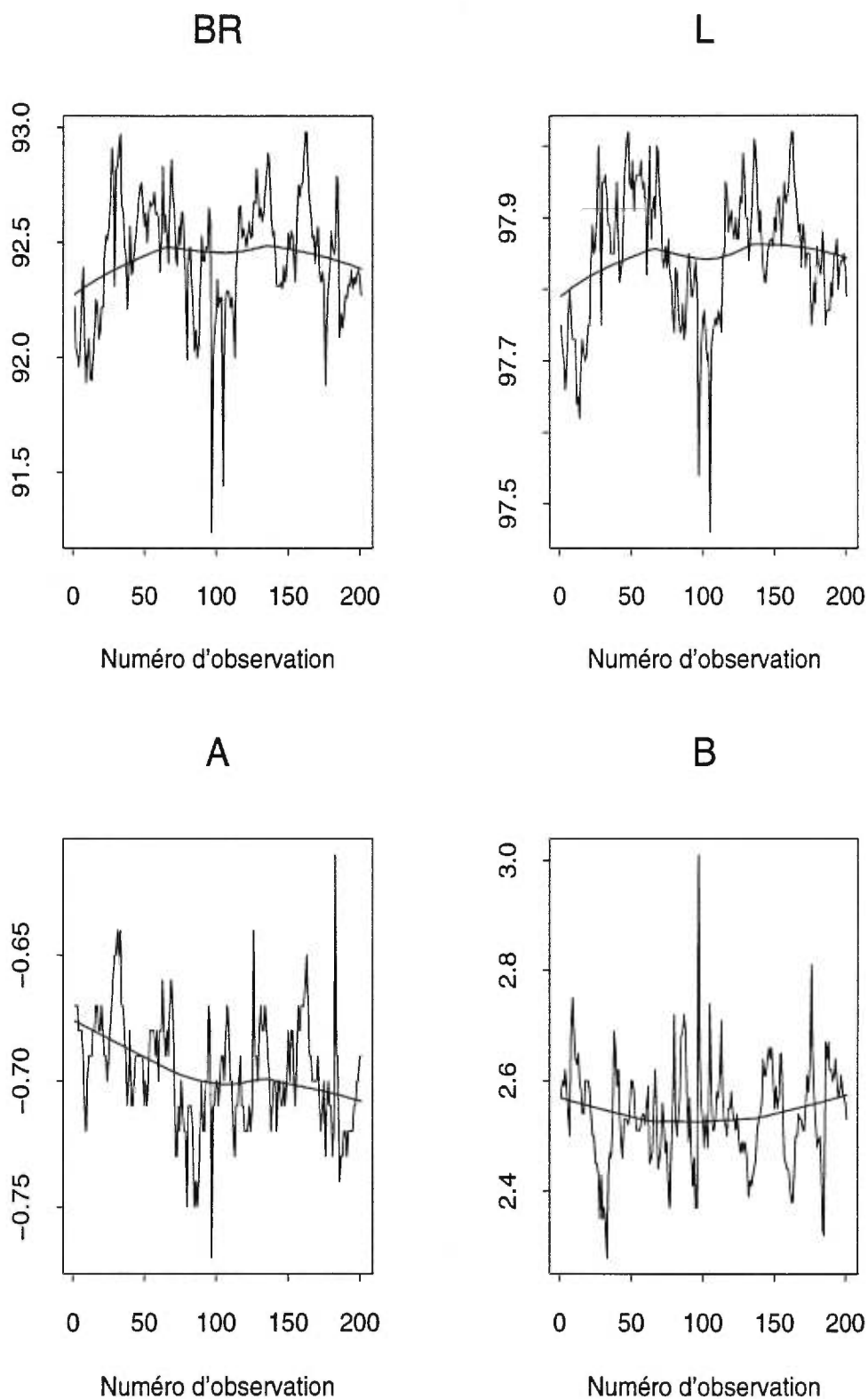


FIGURE 3.20. Série d'observations du jeu1 pour les 4 variables du procédé.



Par la suite, on a regardé l'auto-corrélation entre les observations. On a analysé l'auto-corrélation entre les observations d'une même variable mais également celle entre les différentes variables. Pour une question de présentation, seuls les graphiques des fonctions d'auto-corrélation (FAC) entre les observations d'une même variable seront présentés. Aussi, les constatations que l'on fait pour l'auto-corrélation entre les observations pour une même variable, sont également valables pour les auto-corrélations inter-variables.

La figure 3.21 nous montre la fonction d'auto-corrélation  $\rho_k$  en fonction du délai  $k$  entre les observations pour les 4 variables du jeu1.

On voit donc que même l'auto-corrélation pour les délais jusqu'à 8 est élevée, et ce pour au moins 3 variables. De plus, cette fonction d'auto-corrélation se rapproche fortement de celle d'un modèle de série chronologique  $AR(1)$  multivariée. Nous allons donc tenter d'appliquer un tel modèle à ces variables et voir ce qui advient des résidus. On posera donc,

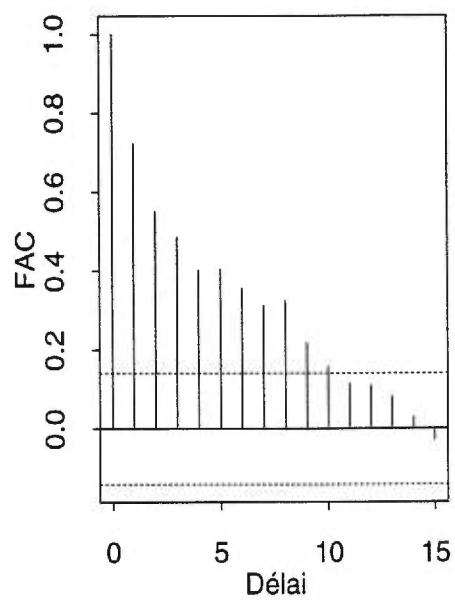
$$(\underline{x}_{j+1} - \underline{\mu}) = \Phi(\underline{x}_j - \underline{\mu}) + \underline{\epsilon}_{j+1}, \quad (3.2.1)$$

où  $\underline{\epsilon}_{j+1}$  est de loi  $N_p(\underline{0}, \Sigma_\epsilon)$  et  $\Phi$  est défini à l'équation (2.8.2).

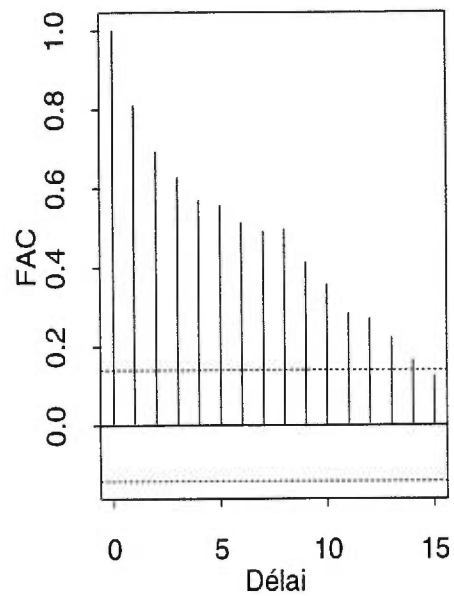
La figure 3.22 représente la fonction d'auto-corrélation (FAC) des résidus pour les 201 observations du jeu1 après avoir estimé les paramètres d'un modèle  $AR(1)$  multivarié. Encore une fois, l'analyse a été faite sur toutes les corrélations, mais on ne présente que celles entre les observations d'une même variable.

On voit donc clairement que l'auto-corrélation n'est pas significative (à une ou deux exceptions près). Un modèle auto-régressif d'ordre un semble donc être une bonne approximation. On obtient les mêmes résultats sur les auto-corrélations entre les variables. On supposera donc que les données suivent effectivement un

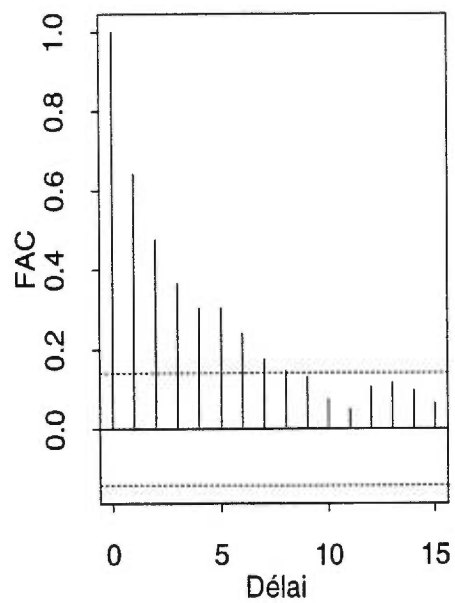
Série : BR



Série : L



Série : A



Série : B

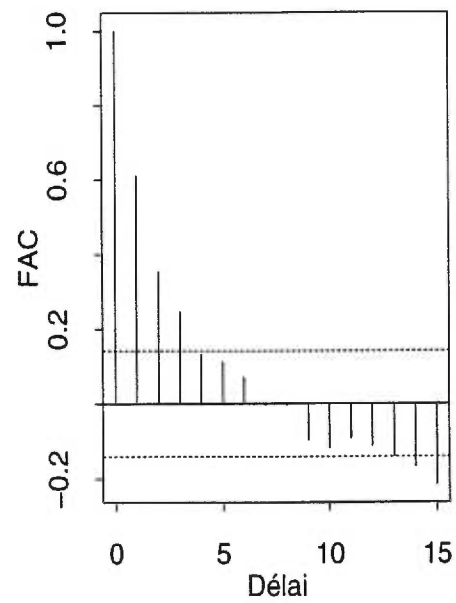
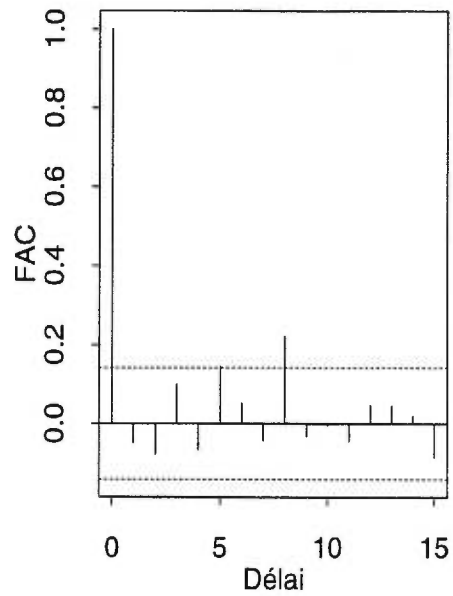
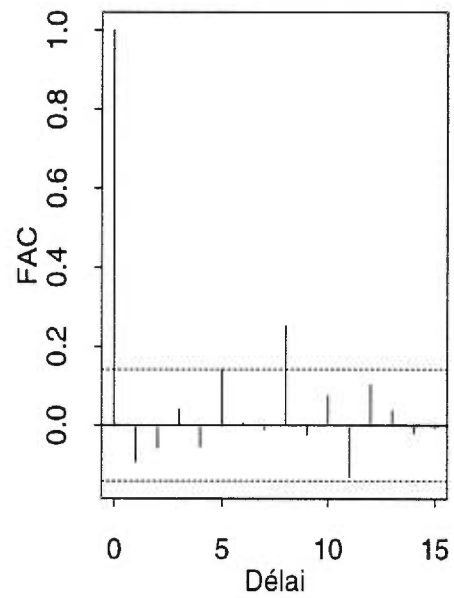


FIGURE 3.21. Fonction d'auto-corrélation pour les 4 variables du jeu1.

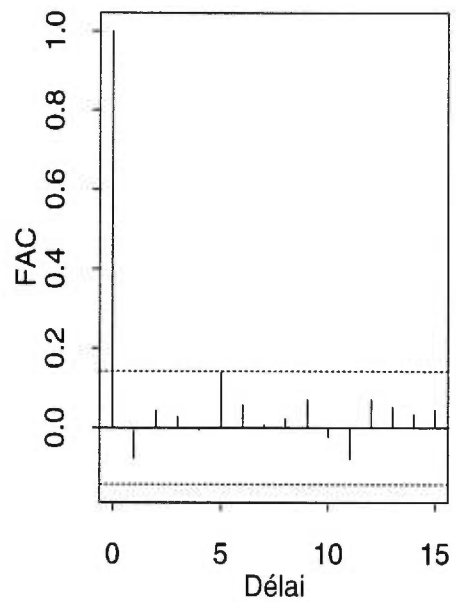
Série : BR



Série : L



Série : A



Série : B

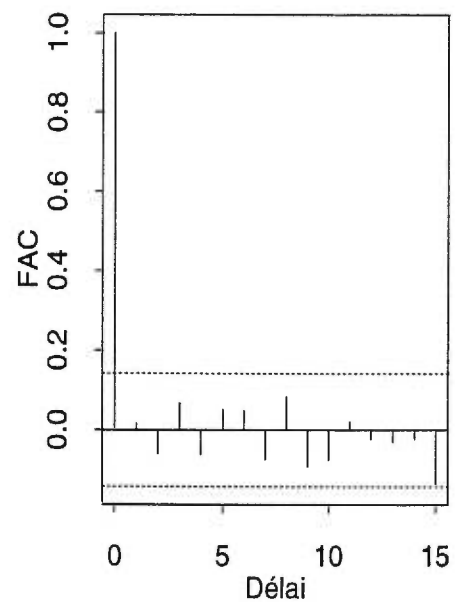


FIGURE 3.22. Fonction d'auto-corrélation pour les résidus du modèle  $AR(1)$  multivarié pour les 4 variables du jeu1.

modèle  $AR(1)$  multivariée avec l'estimé suivant pour  $\Phi$ ,

$$\hat{\Phi} = \begin{pmatrix} 0,077 & -0,049 & -0,086 & 0,17 \\ 1,67 & 0,97 & 0,18 & -0,43 \\ 0,85 & 0,24 & 0,66 & -0,30 \\ -0,26 & 0,089 & -0,091 & 0,73 \end{pmatrix}. \quad (3.2.2)$$

Ces estimés ont été calculés à partir de la méthode de Yule-Walker.

Voyons maintenant ce qui se passe avec le jeu2 qui contient 271 observations. La figure 3.23 représente ces 271 observations pour les 4 variables en plus de la courbe de lissage.

On s'aperçoit qu'il semble y avoir un changement de moyenne près de l'observation 150. Cette série ne semble donc pas stationnaire. Il ne serait donc pas approprié de faire une carte de contrôle à partir de ces 271 observations. Nous ne conserverons ainsi que les 122 dernières observations seulement (celles à partir de l'observation 150). De plus, tout comme le jeu1, le modèle de série chronologique  $AR(1)$  multivariée s'ajuste très bien à ces observations. Pour toutes les analyses suivantes, le jeu2 représentera donc les observations 150 à 271 du jeu2.

Puisque les données sont fortement auto-corrélées, notre choix de l'estimateur pour la matrice de covariance  $\Sigma$  doit en tenir compte. Comme on l'a démontré au chapitre 2, la matrice de covariance échantillonnale  $S$ , est un meilleur estimateur de  $\Sigma$  que  $\tilde{S}$  puisque l'espérance de  $S$  est approximativement toujours égale à la variance des observations  $x$  quelque soit le choix du modèle de série chronologique, contrairement à l'espérance de  $\tilde{S}$  qui est influencée par le modèle de série chronologique.

Voici donc les estimés du vecteur des moyennes, de la matrice de covariance et de la matrice de corrélation pour les 201 observations du jeu1. Les variables sont présentées dans l'ordre suivant: BR, L, A et B.

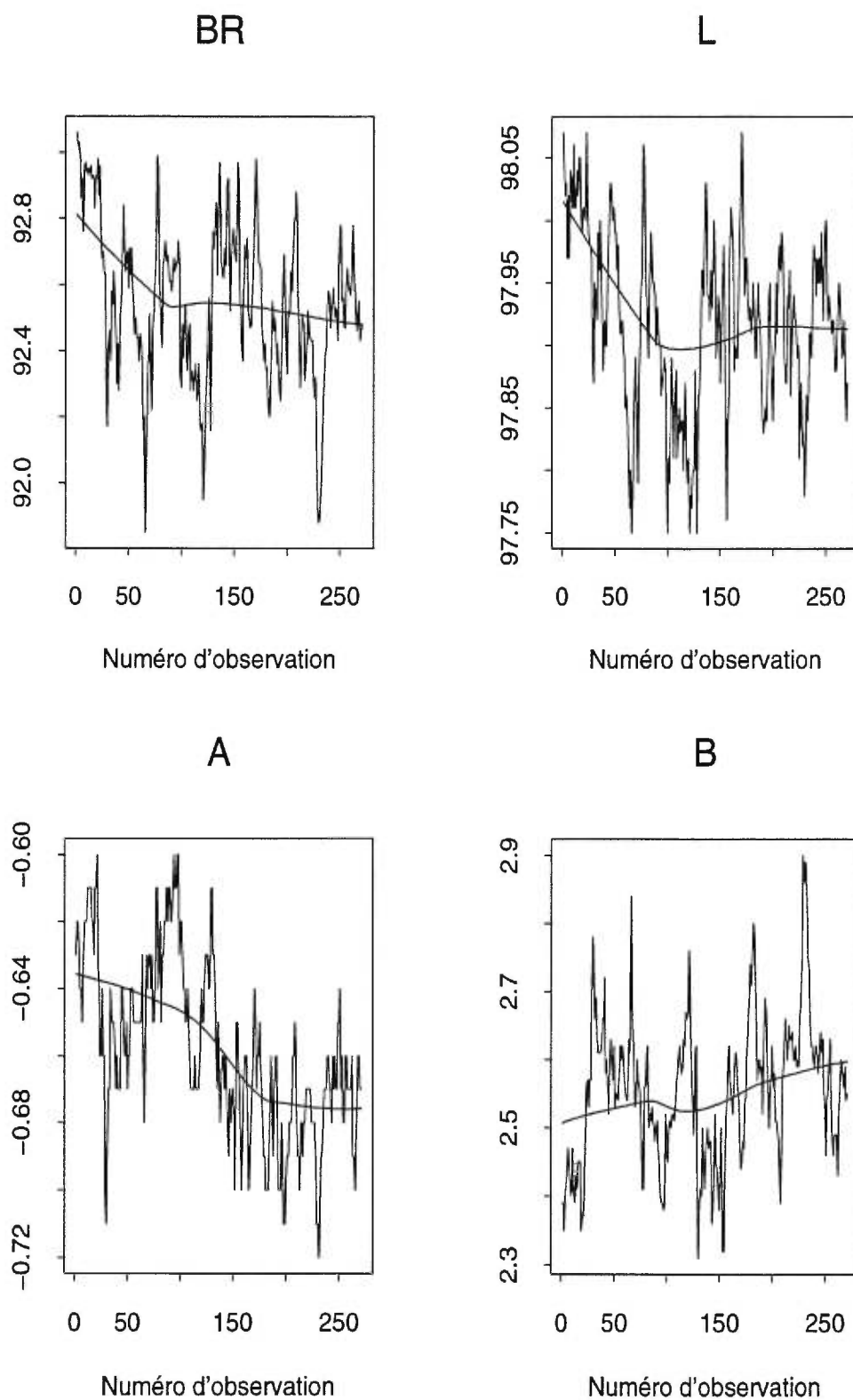


FIGURE 3.23. Série d'observations du jeu2 pour les 4 variables du procédé.

$$\hat{\underline{\mu}} = \underline{\bar{x}} = \begin{pmatrix} 92,44 \\ 97,85 \\ -0,70 \\ 2,54 \end{pmatrix}, \quad (3.2.3)$$

$$\hat{\Sigma} = S = \begin{pmatrix} 0,071 & 0,023 & 0,0030 & -0,021 \\ 0,023 & 0,0088 & 0,00070 & -0,0047 \\ 0,0030 & 0,00070 & 0,00051 & -0,0013 \\ -0,021 & -0,0047 & -0,0013 & 0,0090 \end{pmatrix}, \quad (3.2.4)$$

et

$$\hat{R} = \begin{pmatrix} 1,00 & 0,91 & 0,49 & -0,82 \\ 0,91 & 1,00 & 0,33 & -0,53 \\ 0,49 & 0,33 & 1,00 & -0,62 \\ -0,82 & -0,53 & -0,62 & 1,00 \end{pmatrix},$$

où  $\hat{R}$  est la matrice de corrélation échantillonnale entre les variables. On note que les variables sont corrélées positivement entre elles à l'exception de la variable  $B$  qui est corrélée négativement avec chacune des 3 autres variables.

Un fait maintenant qui est intéressant de noter est la différence entre les moyennes des deux échantillons, c'est-à-dire entre le jeu1 et les 122 observations que l'on a conservées pour le jeu2. En effet, si on regarde la figure 3.24 qui représente la comparaison entre les diagrammes en boîte du jeu1 (gauche) et du jeu2 (droit) pour les 4 variables, on s'aperçoit que les distributions ne sont pas les mêmes.

Il serait donc intéressant d'effectuer un test de moyennes pour être certain que l'on pourra comparer les jeu1 et jeu2. Alors pour chaque variable, le test des moyennes consiste à tester  $H_0 : \mu_1 = \mu_2$ , contre  $H_1 : \mu_1 \neq \mu_2$ . La statistique est alors

$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\hat{\sigma}_{\bar{x}_1}^2 + \hat{\sigma}_{\bar{x}_2}^2}}, \quad (3.2.5)$$

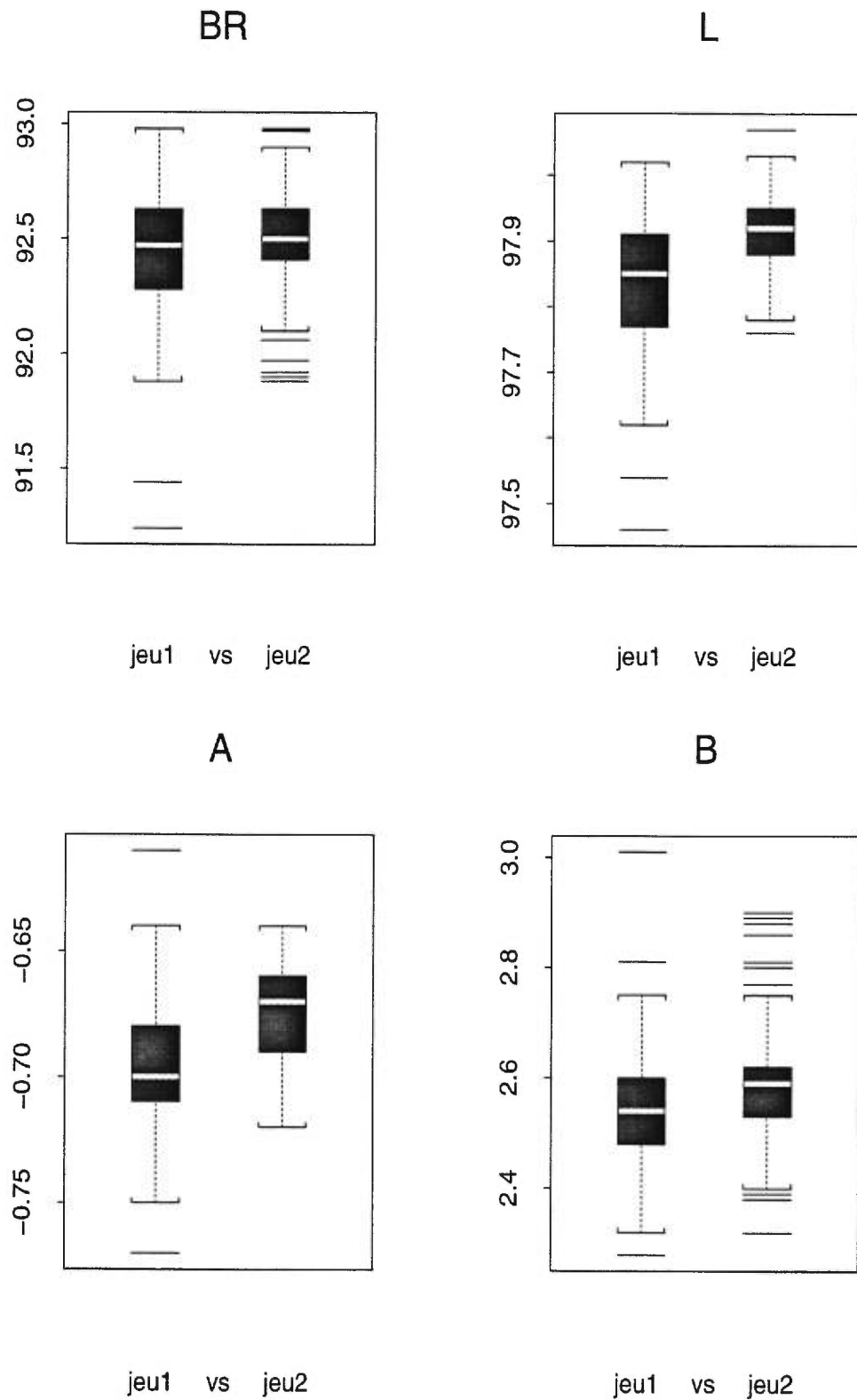


FIGURE 3.24. Comparaisons des diagrammes en boîte du jeu1 (gauche) et du jeu2 (droit) pour les 4 variables.

où pour un modèle  $AR(1)$ , on a vu au chapitre 2 que  $\hat{\sigma}_x^2 = [\hat{\sigma}^2(1 + \hat{\phi}^2)]/[n(1 - \hat{\phi}^2)]$ . Cette statistique, pour  $n$  grand, suit une loi  $N(0, 1)$ .

On obtient ainsi des statistiques  $Z$  égales à  $Z_{BR} = -1,14$ ,  $Z_L = -4,11$ ,  $Z_A = -5,48$  et  $Z_B = -1,73$ , qui correspondent à des valeurs- $p$  égales à 0,25,  $4,0 \times 10^{-5}$ ,  $4,3 \times 10^{-8}$  et 0,084. On constate donc que pour 2 des 4 variables, on rejette facilement l'hypothèse d'égalité des moyennes.

Pour cette raison, les cartes de contrôle ne seront appliquées que pour les données de la phase un. C'est-à-dire que l'on testera si la moyenne des observations du jeu1 (ou jeu2) est en contrôle à partir des estimés des paramètres calculés à partir du même jeu de données. Ainsi puisque les moyennes sont différentes pour le jeu1 versus le jeu2, il est impossible de surveiller par exemple si la moyenne des observations du jeu2 est en contrôle à partir des estimations du jeu1.

On doit noter que les mêmes analyses ont été effectuées à partir du troisième jeu de données et les mêmes conclusions s'appliquaient. C'est-à-dire que l'auto-corrélation est très élevée, que la série est non stationnaire et que les moyennes des variables entre les différents jeux de données sont différentes.

Ce qui peut être important de savoir maintenant concerne l'analyse en composantes principales. Il serait intéressant de savoir s'il est possible de conserver moins de variables dans notre analyse. Cependant, on doit noter que l'acp n'est pas indépendante de l'échelle utilisée. Il y a donc deux façons de traiter le problème. La première consiste à effectuer l'acp sur les variables originales et donc en décomposant la matrice de covariance  $\Sigma$  du procédé. La deuxième consiste à normaliser les données pour les rendre sur une échelle commune et de décomposer la matrice de covariance des variables normalisées ce qui correspond à la matrice de corrélation des données originales. La première sera privilégiée si l'échelle originale des variables semble avoir une signification particulière. Sinon, on recommande d'effectuer l'acp sur les variables normalisées.



Dans le cas qui nous concerne, l'échelle utilisée ne semble pas avoir de signification particulière, on appliquera donc la méthode de l'acp aux variables normalisées. Cependant, les résultats avec les variables originales seront également présentés. Nous allons donc mettre les variables sur la même échelle en posant,

$$z_{lj} = \frac{(x_{lj} - \bar{x}_l)}{s_{ll}}, \quad l = 1, 2, 3, 4 \text{ et } j = 1, \dots, m, \quad (3.2.6)$$

où  $x_{lj}$  est l'observation  $j$  pour la variable  $l$ ,  $\bar{x}_l$  représente l'estimation de la moyenne pour la variable  $l$  et  $s_{ll}$  représente l'estimation de l'écart-type pour la variable  $l$  et correspond à la racine carrée de l'élément  $(l, l)$  de la matrice  $S$ .

Si on applique l'analyse en composantes principales au jeu1 standardisé à l'aide de (3.2.6), on obtient  $\lambda$  le vecteur des valeurs propres qui correspondent aux variances des nouvelles variables et la matrice  $A$  des vecteurs propres (colonnes de  $A$ ),

$$\lambda = \begin{pmatrix} 2,88 \\ 0,77 \\ 0,34 \\ 0,0059 \end{pmatrix} \quad \text{et} \quad A = \begin{pmatrix} -0,570 & 0,275 & 0,0747 & -0,771 \\ -0,493 & 0,551 & -0,428 & 0,520 \\ -0,405 & -0,750 & -0,522 & -0,0195 \\ 0,518 & 0,241 & -0,734 & -0,368 \end{pmatrix}$$

On obtient donc les nouvelles variables  $y_l$ ,  $l = 1, 2, 3, 4$ , suivantes:

$$\begin{aligned} y_1 &= -0,570 BR - 0,493 L - 0,405 A + 0,518 B \\ y_2 &= 0,275 BR + 0,551 L - 0,750 A + 0,241 B \\ y_3 &= 0,0747 BR - 0,428 L - 0,522 A - 0,734 B \\ y_4 &= -0,771 BR + 0,520 L - 0,0195 A - 0,368 B, \end{aligned}$$

où la somme des carrés des coefficients d'une cp est égale à un. On remarque que pour la première cp, tous les coefficients sont du même ordre de grandeur, ainsi

à prime abord, on ne peut avoir d'idée de quelles variables seront hors-contrôle si cette cp est hors-contrôle. Pour cette cp, on note également que la variable  $B$  est opposée aux trois autres puisque son signe est positif alors que pour les trois autres variables les coefficients sont négatifs. Ce fait n'est pas surprenant, car  $B$  est corrélée négativement avec  $BR$ ,  $L$  et  $A$ , alors que ces dernières sont corrélées positivement entre elles.

Pour la 2e cp, c'est maintenant la variable  $A$  qui est de signe opposé aux trois autres. Pour la 3e, on remarque que le coefficient de la variable  $BR$  est très petit par rapport aux autres, ainsi si un signal survient sur cette cp, il n'y a que très peu de chance que cette variable soit responsable. Pour la 4e cp, c'est la variable  $A$  qui a un coefficient presque nul.

De plus le vecteur  $\underline{\lambda}$  nous indique le pourcentage de la variabilité du procédé expliquée par chacune des composantes principales à l'aide de

$$\frac{\lambda_c}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4}, \quad c = 1, 2, 3, 4.$$

Ainsi, la première cp explique 72,1 % de la variabilité totale, la 2e explique 19,4 %, la 3e 8,4 % et la 4e 0,1 %. Il est donc intéressant de noter que les deux premières composantes expliquent à elles seules plus de 91 % de la variabilité totale du procédé. On doit noter que les mêmes résultats sont obtenus pour les deux autres jeux de données. À la lumière de ces résultats, il serait donc important de se demander s'il vaut la peine de conserver les 4 composantes principales. La question est de savoir si on veut être informé d'un changement de moyenne dans la 3e ou surtout dans la 4e cp puisque celles-ci expliquent ensemble moins de 9 % de la variabilité totale du procédé.

Trois choix se présentent donc à nous pour appliquer une procédure multivariée. Le premier est d'appliquer la procédure habituelle en construisant une carte de contrôle sur les 4 variables et d'appliquer la procédure de détection des

variables hors-contrôle que l'on a présentée au chapitre 2. Cette façon de faire sera appliquée si évidemment on veut que la carte émette un signal même pour la 3e ou 4e composante principale et que l'on veut savoir quelles variables sont responsables de ces signaux.

Le deuxième choix est de faire une carte de contrôle sur les deux premières composantes principales seulement et d'appliquer la procédure de détection des variables hors-contrôle sur ces deux composantes seulement. Cette façon de faire sera appliquée si on décide que les deux dernières cp n'expliquent pas une partie assez importante de la variabilité pour que la carte signale s'il y a hors-contrôle sur ces axes.

La dernière façon de faire est de tout d'abord effectuer une carte de contrôle sur les 4 variables du procédé. Il y aura donc un signal si n'importe quelle cp est hors-contrôle. Cependant en appliquant la procédure de Kourti et MacGregor (1996), si on s'aperçoit que c'est la 3e ou la 4e cp qui est hors-contrôle, on arrête notre analyse. Si c'est la 1e ou la 2e, on continue la procédure pour savoir quelles variables sont responsables. Cette procédure peut être appliquée si on est intéressé à savoir quels points sont hors-contrôle (même pour les 2 dernières cp) mais que les efforts seront portés principalement sur les observations hors-contrôle sur les 2 premières cp.

### 3.3. APPLICATION DE LA CARTE MULTIVARIÉE $T^2$

Dans cette section, on appliquera la carte multivariée  $T^2$  aux jeu1 et aux 122 dernières observations du jeu2.

Puisque les données sont fortement auto-corrélées, on se devra de suivre les conclusions de la section 2.8 sur l'auto-corrélation. La première conclusion est que l'on devra prendre  $S$ , la matrice de covariance échantillonnale, pour estimer

la matrice de covariance du procédé. La deuxième est que l'on appliquera la carte  $T^2$  à ces observations et non les cartes cusum ou ewma.

Avant d'appliquer la carte  $T^2$ , il serait bon de redonner les équations pour le calcul de la statistique  $T^2$  et la limite de contrôle. Pour nos jeux de données, on a des données individuelles ( $n=1$ ), on est à la phase un et on prendra  $S$  comme estimateur, les équations à utiliser sont donc,

$$T_{1;j}^2 = (\underline{x}_j - \bar{\underline{x}})' S^{-1} (\underline{x}_j - \bar{\underline{x}}), \quad (3.3.1)$$

pour la statistique  $T^2$  et

$$LSC_{1;j} = \frac{(m-1)^2}{m} B_{\alpha} \left( \frac{p}{2}, \frac{m-p-1}{2} \right), \quad (3.3.2)$$

pour la limite de contrôle, où  $B_{\alpha}(\nu_1, \nu_2)$  est la valeur critique au niveau  $\alpha$  d'une distribution Beta de paramètre  $\nu_1$  et  $\nu_2$ ,  $m$  est le nombre d'observations et  $p=4$  est le nombre de variables impliquées.

### 3.3.1. Application de la carte aux quatre variables

Ici nous appliquerons la procédure multivariée  $T^2$  aux 4 variables.

Alors pour les  $m=201$  observations du jeu1, les estimateurs de  $\underline{\mu}$  et de  $\Sigma$  sont donnés par les équations (3.2.3) et (3.2.4) respectivement.

La limite de contrôle calculée à partir de l'équation (3.3.2), avec  $m=201$ ,  $p=4$  et  $\alpha=0,0027$ , est égale à 15,61.

La carte de contrôle  $T^2$  calculée pour les 201 observations du jeu1 est alors présentée à la figure 3.25.

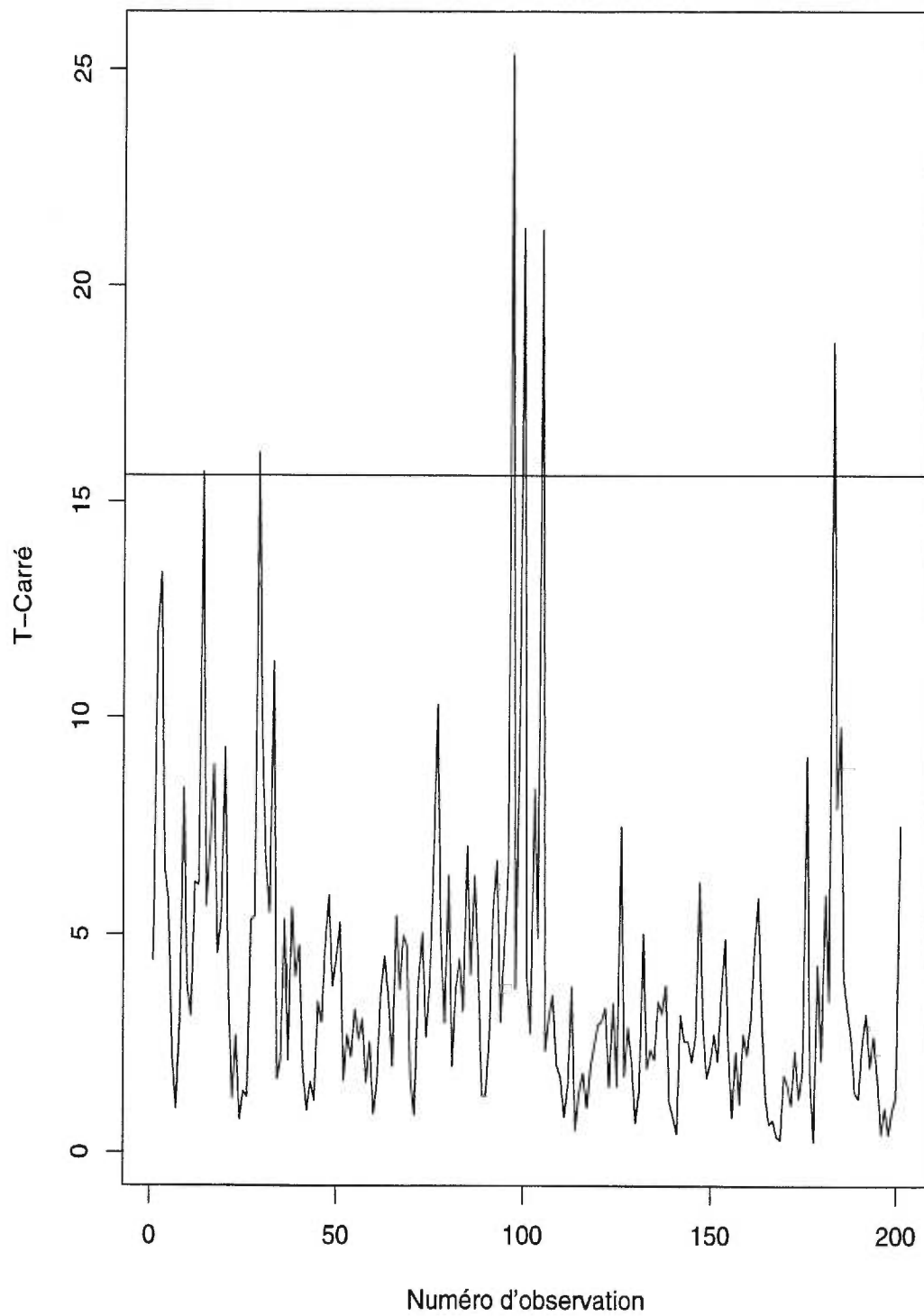


FIGURE 3.25. Carte de contrôle  $T^2$  pour les 201 observations du jeu1 (pour les 4 variables).

Ainsi, les observations numérotées 14, 29, 97, 100, 105 et 183 ont une valeur de la statistique  $T^2$  supérieure à la limite de contrôle de 15,61 et sont donc hors-contrôle. Ces valeurs sont respectivement égales à 15,7, 16,1, 25,3, 21,3, 21,3 et 18,7.

Appliquons maintenant la même procédure, mais pour le jeu2. Pour ces 122 observations, on obtient

$$\bar{\bar{x}} = \begin{pmatrix} 92,50 \\ 97,92 \\ -0,68 \\ 2,59 \end{pmatrix} \quad \text{et} \quad S = \begin{pmatrix} 0,041 & 0,0080 & 0,0023 & -0,019 \\ 0,0080 & 0,0029 & 0,00047 & -0,0023 \\ 0,0023 & 0,00047 & 0,00027 & -0,0011 \\ -0,019 & -0,0023 & -0,0011 & 0,011 \end{pmatrix}$$

La limite de contrôle avec  $m=122$ ,  $p=4$  et  $\alpha=0,0027$  est égale à 15,20. La figure 3.26 représente la carte de contrôle  $T^2$  pour les 122 dernières observations du jeu2.

Pour ce jeu, les observations 7 et 30 sont hors-contrôle puisque leur statistique  $T^2$  est respectivement égale à 37,6 et 27,8.

Maintenant que la carte  $T^2$  a détecté ces points hors-contrôle, il faut être en mesure de savoir quelles variables sont responsables de ces points hors-contrôle. C'est la prochaine étape à effectuer. Mais pour l'instant, nous allons construire la carte multivariée  $T^2$  pour les jeu1 et jeu2, dans le cas où nous nous intéressons aux deux premières composantes principales seulement.

### 3.3.2. Application de la carte aux deux premières composantes principales

Ici, nous appliquerons la carte  $T^2$ , mais en utilisant les deux premières composantes principales seulement puisque celles-ci expliquent 91,5 % de la variabilité totale dans le jeu1 et 89,8 % pour le jeu2.

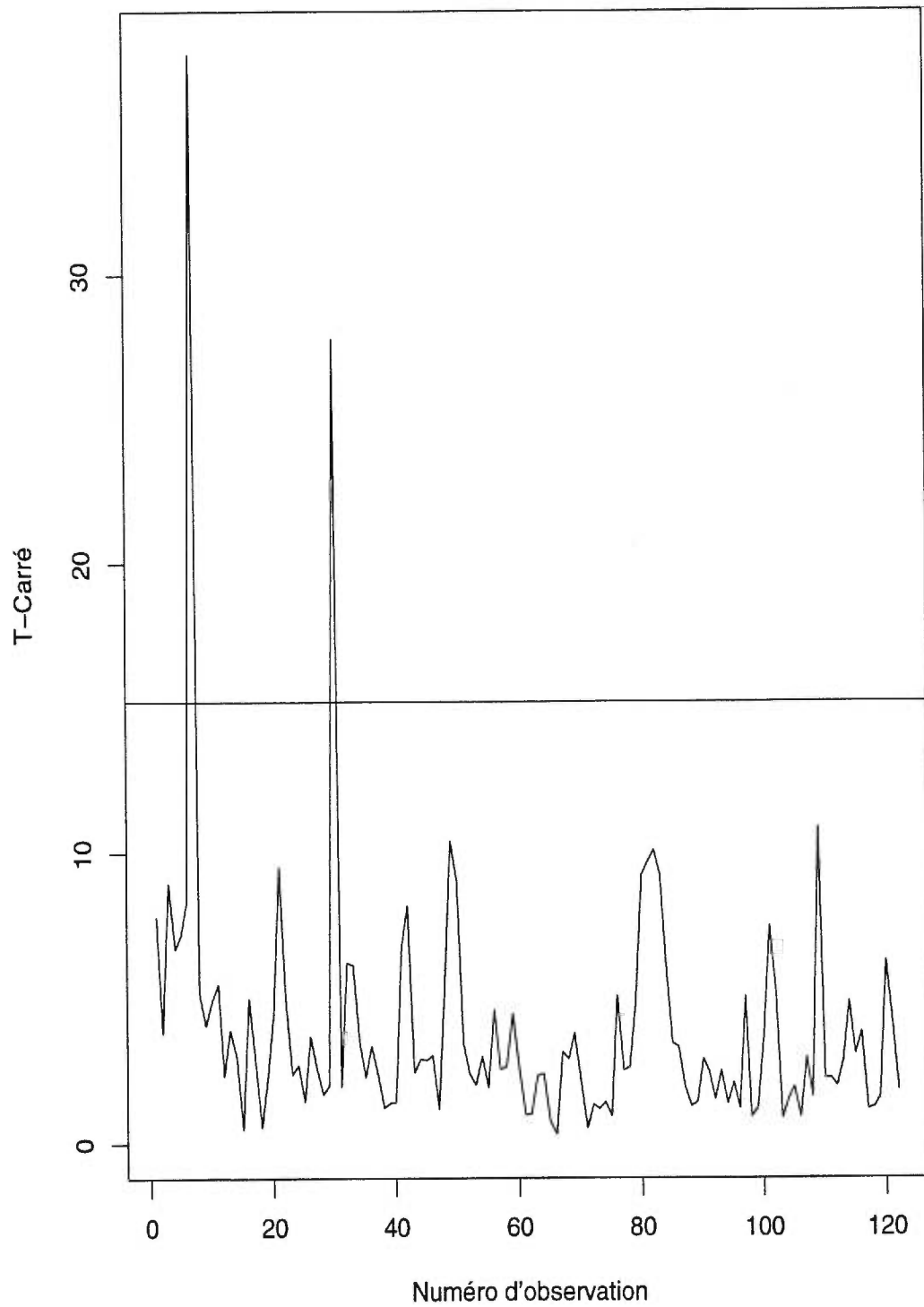


FIGURE 3.26. Carte de contrôle  $T^2$  pour les 122 observations du jeu2 (pour les 4 variables).

Alors en appliquant la procédure décrite au chapitre 2 avec les deux premières composantes principales on obtient les résultats suivants. La limite de contrôle pour le jeu1 est alors 11,37 et les points hors-contrôle pour ce jeu sont les observations numéros 97, 105 et 183, comparativement aux points 14, 29, 97, 100, 105 et 183 lorsque l'on utilise les 4 variables. On peut donc supposer que les observations 14, 29 et 100 sont hors-contrôle sur les deux dernières composantes principales. C'est ce que nous analyserons plus en profondeur dans la prochaine section. Pour le jeu2, la limite de contrôle est de 11,09 et seul le point 7 est hors-contrôle, comparativement aux points 7 et 30 lorsque les 4 variables étaient utilisées. Il semble donc que le point 7 est hors-contrôle sur à tout le moins la première ou deuxième cp alors que le point 30 est hors-contrôle sur la troisième ou quatrième.

Il reste donc à savoir quelles variables sont responsables de ces points hors-contrôle sur les deux premières composantes principales. C'est ce qui sera abordé dans la prochaine section.

### 3.4. DÉTECTION DES VARIABLES HORS-CONTRÔLE

Dans cette section, on appliquera les techniques de détection des variables hors-contrôle aux observations hors-contrôle du jeu1 et du jeu2 pour être en mesure d'interpréter ces signaux.

#### 3.4.1. Construction des cartes univariées

La première étape consiste à construire les cartes univariées de Shewhart pour savoir si les points hors-contrôle de la procédure multivariée sont également hors-contrôle sur les cartes univariées.



Puisque les données sont auto-corrélées, on a pris comme estimateur de la variance de la variable  $l$ , l'élément  $(l, l)$  de la matrice  $S$ . On a utilisé les limites de Bonferroni pour le calcul des limites de contrôle. On a donc pris  $\alpha_u = \alpha_m/4$ , qui avec  $\alpha_m=0,0027$ , équivaut à  $L=3,40$ . La figure 3.27 représente les cartes de contrôle univariées des 4 variables pour le jeu1.

Ainsi, seuls les points 97 (variables  $BR$  et  $B$ ), 105 (variables  $BR$  et  $L$ ) et 183 (variable  $A$ ) sont hors-contrôle sur les cartes univariées. Pour les points 14, 29 et 100, aucune variable ne semble donc être hors-contrôle.

Pour le jeu2, les cartes de contrôle univariées sont présentées à la figure 3.28. Pour ces points, aucune variable n'est hors-contrôle.

On voit que pour plusieurs points hors-contrôle sur la procédure multivariée, aucune variable n'est hors-contrôle sur les procédures univariées. On se doit donc de pousser plus loin cette analyse pour être en mesure de véritablement identifier des variables hors-contrôle pour tous les points.

La deuxième analyse à faire est de calculer les statistiques de Student pour chacun des points hors-contrôle, cela revient à normaliser les observations. Cela nous indiquera ainsi quelles variables se sont le plus éloignées de leurs moyennes. Le tableau 3.1 contient les statistiques de Student pour chacun des points hors-contrôle du jeu1 et du jeu2 et ce pour les 4 variables.

Si on fixe une limite supérieure à 3,40, qui est la limite de Bonferroni utilisée pour le calcul des limites pour les cartes de Shewhart univariées, on arrive évidemment au même résultat. C'est-à-dire que les points qui sont supérieurs à 3,40 sont ceux qui étaient hors-contrôle sur les procédures univariées de Shewhart. Avec un tel tableau, on est en mesure de savoir quelles variables sont le plus responsables des signaux même s'il n'y a pas hors-contrôle pour ces variables. Par exemple, pour le point 14 du jeu1, aucune variable n'est hors-contrôle,

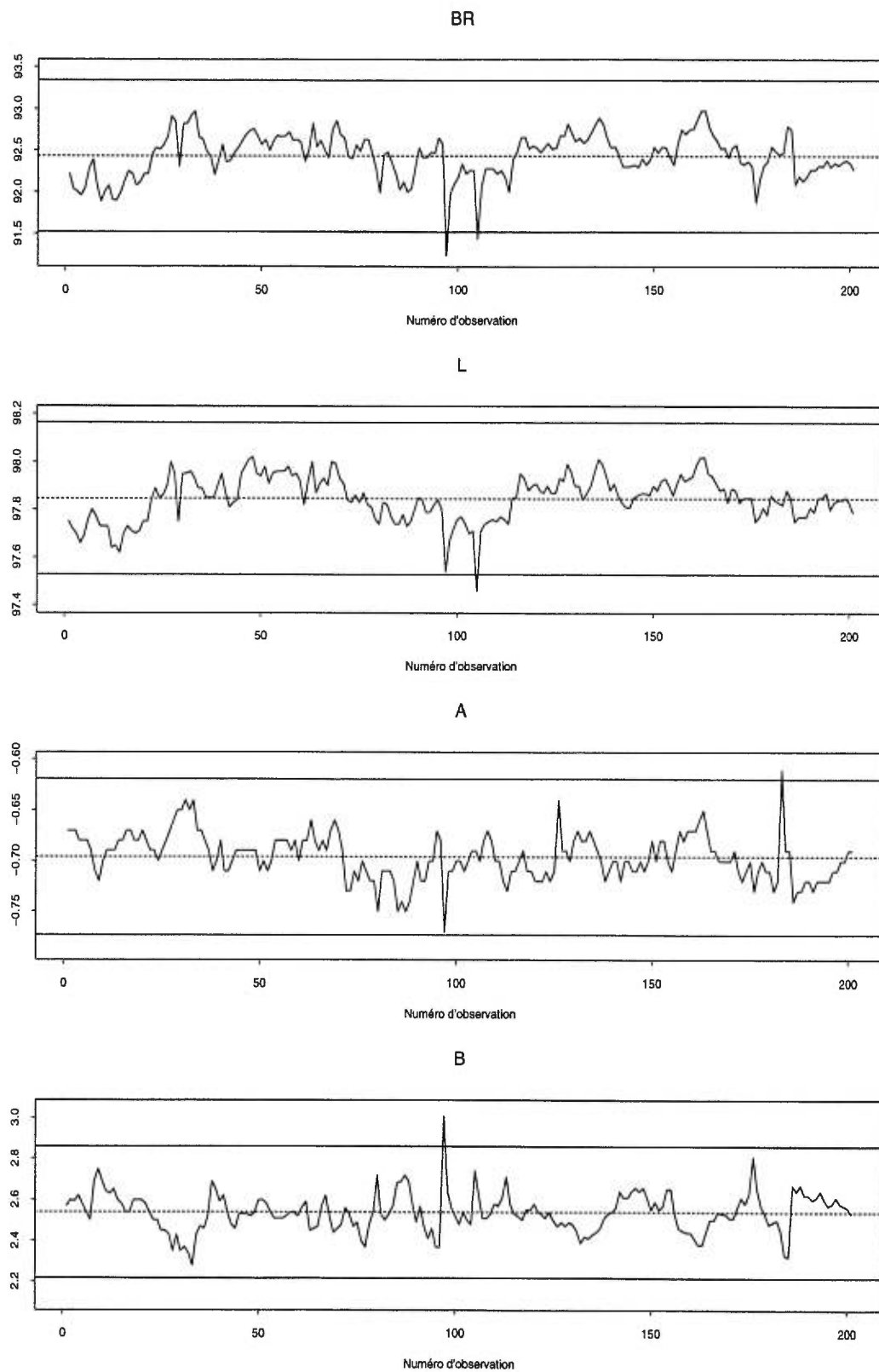


FIGURE 3.27. Cartes de contrôle de Shewhart univariées pour les 4 variables du jeu1 (avec  $L=3,40$ ).

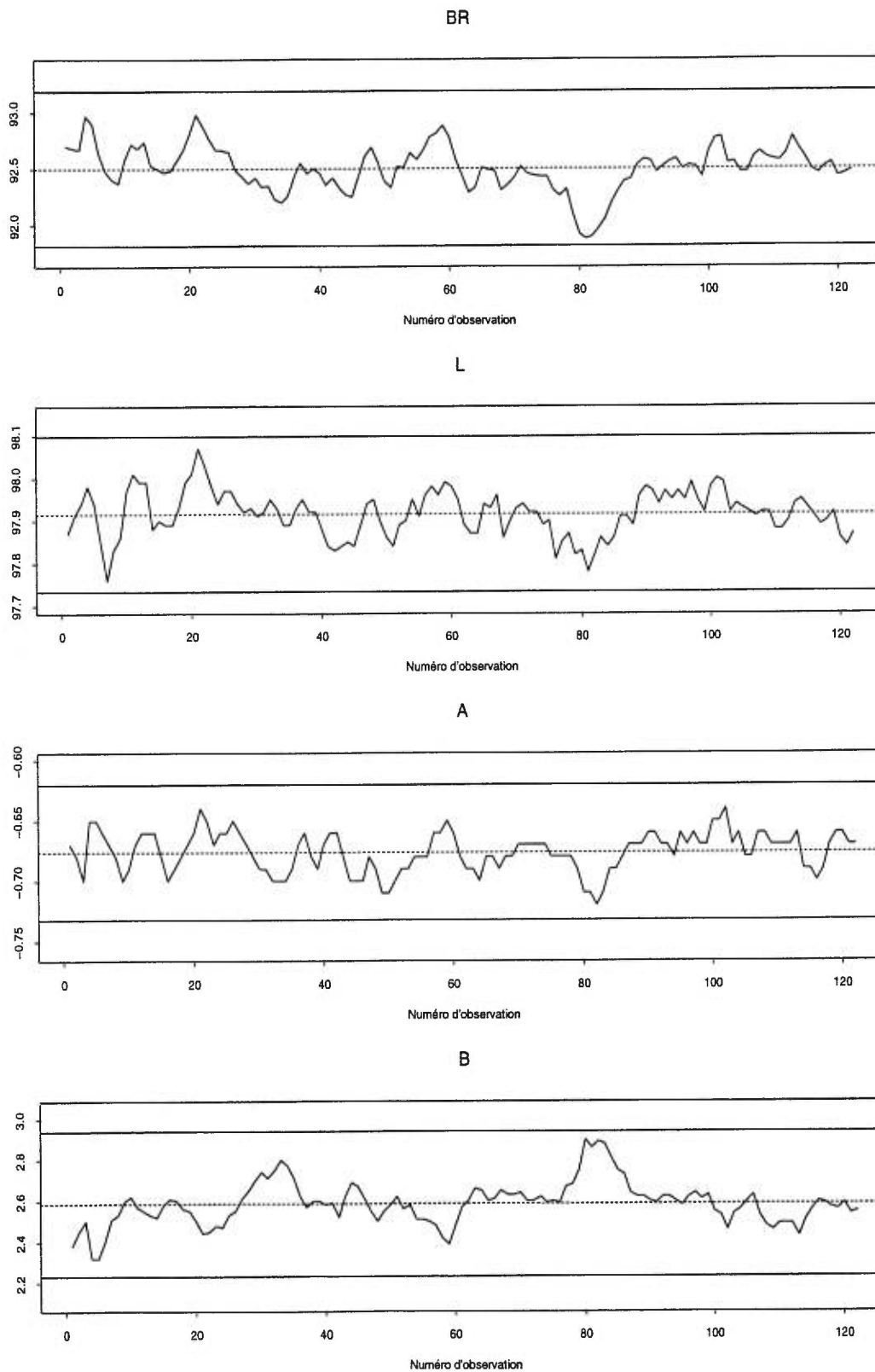


FIGURE 3.28. Cartes de contrôle de Shewhart univariées pour les 4 variables du jeu2 (avec  $L=3,40$ ).

no. obs.	<i>BR</i>	<i>L</i>	<i>A</i>	<i>B</i>
14	-1,67	-2,41	0,71	0,64
29	-0,47	-1,02	2,03	-1,16
97	-4,47	-3,27	-3,27	4,97
100	-0,96	-0,92	-0,18	-0,21
105	-3,73	-4,12	0,26	2,12
183	0,13	-0,28	3,80	-0,95
7	-0,10	-2,90	0,38	-0,73
30	-0,40	-0,10	-0,83	1,47

TABLEAU 3.1. Valeurs de la statistique de Student pour les observations hors-contrôle des jeu1 (14, 29, 97, 100, 105 et 183) et jeu2 (7 et 30).

cependant pour la variable *L*, la statistique de Student est égale à -2,41. On peut donc supposer que c'est cette variable qui est le plus responsable. Aussi, si on regarde le point 97, les statistiques de Student pour les 4 variables sont élevées et pourtant on avait déclaré que seules les variables *BR* et *B* étaient responsables.

Il est donc important de comprendre que les limites de Bonferroni sont des limites que l'on doit utiliser comme guide et non pas comme des limites absolues étant donné qu'un signal a déjà été détecté. Ainsi, toutes les valeurs qui sont élevées doivent être prises en considération.

### 3.4.2. Application de la méthode d'acp de Kourti et MacGregor

Dans cette section, nous appliquerons cette méthode en utilisant les 4 variables et également en utilisant seulement les deux premières composantes principales. Aussi, on verra les différences que l'on obtient si on applique l'acp sur les variables originales ou sur les variables normalisées.

Voyons tout d'abord ce qui arrive si on applique la méthode de Kourti et MacGregor (1996) aux 4 variables normalisées du procédé.

On commence par calculer les scores normalisés des composantes principales pour savoir quelles composantes principales sont hors-contrôle. Le tableau 3.2

contient les scores normalisés pour les 6 points hors-contrôle du jeu1 et pour les 2 points du jeu2 lorsque les 4 variables normalisées sont utilisées.

no. obs.	1ère cp	2e cp	3e cp	4e cp
14	1,29	-2,46	0,12	-2,83*
29	-0,38	-2,84*	0,33	2,80*
97	4,75*	0,71	-1,51	-0,17
100	0,57	-0,78	0,98	4,41*
105	3,03*	-3,39*	-0,36	-0,71
183	-1,16	-3,63*	-2,01	0,39
7	0,46	3,65*	-0,68	-4,86*
30	0,82	-1,10	0,086	-5,09*

TABLEAU 3.2. Scores normalisés des 4 composantes principales pour les 6 points hors-contrôle du jeu1 (14, 29, 97, 100, 105 et 183) et les 2 points hors-contrôle du jeu2 (7 et 30).

Comme limite, nous choisirons 2,5 qui est la limite recommandée par Kourti et MacGregor (1996). Comme on le constate, cette limite est différente de toutes les limites utilisées précédemment, qui tenait compte de  $\alpha$ , l'erreur de type 1, et qui tournait autour de la valeur 3. La raison est qu'ici ce n'est pas tant que l'on veut savoir quelles composantes principales est hors-contrôle que la contribution des variables aux scores élevés. Pour savoir quelles variables sont hors-contrôle, il est donc important de considérer tous les scores élevés. On n'a pas besoin ici de se préoccuper de l'erreur de type 1.

Ainsi les composantes principales avec un astérisque (\*) sont hors-contrôle. Pour le jeu1, les points 14 et 100 sont hors-contrôle sur la 4e cp seulement, alors que les points 97, 105 et 183 sont hors-contrôle sur une des deux premières cp (ou les deux). Finalement, le point 29, est hors-contrôle sur la 2e et 4e cp. Pour le jeu2, on en conclut que pour le point 7, les 2e et 4e cp sont hors-contrôle, alors que pour le point 30 seule la 4e est hors-contrôle.

Par la suite, pour ces composantes hors-contrôle, on va calculer une contribution moyenne globale des variables aux scores de ces cp. On se rappelle que l'on ne garde que les variables qui sont du même signe que le score. Le tableau

3.3 contient les contributions des variables pour les 6 points hors-contrôle du jeu1 et pour les 2 points du jeu2 pour les scores normalisés supérieurs à 2,5.

no. obs.	<i>BR</i>	<i>L</i>	<i>A</i>	<i>B</i>
14	0	46,04*	0,51	8,62
29	13,55*	1,82	4,91	16,39*
97	7,12*	4,50*	3,70*	7,19*
100	42,19*	0	0,20	4,36
105	7,74*	12,36*	0,76	1,96
183	0	0,63	11,76*	0,94
7	0	56,70*	0,70	1,90
30	0	1,71	0	39,66*

TABLEAU 3.3. Contributions des variables pour les 6 points hors-contrôle du jeu1 (14, 29, 97, 100, 105 et 183) et les 2 points du jeu2 (7 et 30) pour les scores normalisés supérieurs à 2,5.

On s'aperçoit donc clairement que pour tous les points hors-contrôle, il y a maintenant des variables responsables. Les astérisques (\*) représentent les variables que l'on suspecte être responsables du signal hors-contrôle. Ainsi, pour le point 14, on note que trois contributions sont différentes de 0, mais que leur ordre de grandeur est très différent. Il est clair qu'avec une contribution de 46,04, la variable *L* est responsable du signal. Cependant, on doit également donner une attention particulière à la variable *B* qui a une contribution de 8,62, mais on doit surtout concentrer nos efforts sur la variable *L*. Pour le point 29, on suspectera les variables *BR* et *B*, tandis que les 4 variables semblent être responsables pour le point 97. Seule la variable *BR* mérite une attention particulière pour le point 100. Finalement, pour le point 105, on suspecte *BR* et *L* et pour le point 183, c'est *A* qui semble être responsable. Pour le jeu2, on voit que c'est la variable *L* qui est responsable du signal à l'observation 7 tandis que c'est la variable *B* qui est responsable pour le point 30.

Aussi, ces résultats viennent confirmer ce que l'on a écrit auparavant au sujet des cp, c'est-à-dire que si la 4e cp est hors-contrôle, il est très peu probable que ce soit en raison de la variable *A*. Ce fait est confirmé par les points 14, 100 et 30 qui sortent sur la 4e cp et pour lesquels la variable *A* n'est pas responsable.

Si on établit une comparaison avec le tableau 3.1 qui présente les statistiques de Student, on s'aperçoit que ce n'est pas nécessairement les variables avec les statistiques de Student les plus élevées (en valeur absolue) qui sont responsables des signaux. Par exemple, pour le point 29, la variable  $A$  a la statistique de Student la plus élevée (2,03), mais n'est pas responsable du signal. Aussi, pour le point 100, les 4 statistiques de Student sont inférieures à 1 et on identifie quand même la variable  $BR$  comme étant responsable.

Comme on l'a déjà mentionné, on obtient des conclusions quelque peu différentes si on applique l'acp sur les données originales plutôt que sur les données normalisées. Voici les 4 cp lorsque l'acp est appliquée sur les données originales et donc en décomposant la matrice  $S$  du jeu1,

$$y_1 = -0,917 BR - 0,293 L - 0,0390 A + 0,268 B$$

$$y_2 = -0,0587 BR - 0,573 L + 0,128 A - 0,808 B$$

$$y_3 = -0,103 BR + 0,207 L + 0,973 A + 0,0145 B$$

$$y_4 = -0,381 BR + 0,737 L - 0,190 A - 0,525 B.$$

La première constatation que l'on fait, est que pour la première cp, le coefficient de  $BR$  est très élevé par rapport aux autres, ce qui n'était pas le cas pour la première cp des variables normalisées où les 4 coefficients avaient à peu près le même poids. Ainsi, si la 1<sup>ère</sup> cp est hors-contrôle dans ce cas-ci, il y a de fortes possibilités que la variable  $BR$  soit en cause.

Pour le jeu1, on obtient que pour les points 14, 29 et 100 les conclusions restent exactement les mêmes. C'est-à-dire que les mêmes composantes principales sont supérieures à 2,5 et les mêmes variables sont susceptibles d'être responsables des signaux. Pour le point 183, c'est maintenant la cp 3 qui est supérieure à 2,5 au lieu de la cp 2 lorsque l'acp est appliquée aux variables normalisées. Cependant, c'est la variable  $L$  qui semble être responsable dans les deux cas. Ces faits ne sont

pas surprenants puisque pour la cp 2 appliquée aux variables normalisées, c'est justement la variable  $A$  qui avait le coefficient le plus élevé soit 0,750 et pour la cp 3 appliquée aux variables originales, le coefficient de  $A$  est également très élevé à 0,973. Il n'est donc pas surprenant que si la variable  $A$  est hors-contrôle, ce soit la cp 2 et la cp 3 qui soient hors-contrôle pour respectivement les procédures appliquées aux variables normalisées et aux variables originales.

Pour le point 97, c'est la même cp qui est élevée, mais ce sont les variables hors-contrôle qui ne sont pas les mêmes. En effet, la première cp est hors-contrôle dans les deux cas, mais pour l'acp pour les variables normalisées les 4 variables étaient responsables, alors que pour les variables originales seule la variable  $BR$  est responsable du signal. Ceci peut s'expliquer en regardant les coefficients de la première cp. En effet, pour les variables normalisées, on avait remarqué que les coefficients des variables pour la première cp étaient à peu près tous égaux (en valeur absolue), ainsi a priori, on ne peut suspecter une variable plus qu'une autre d'être hors-contrôle lorsqu'il y a un signal sur la 1<sup>ière</sup> cp. Pour les variables originales, on a remarqué que le coefficient de  $BR$  est très élevé par rapport aux autres, il n'est donc pas surprenant qu'un signal sur cette cp implique que seule la variable  $BR$  soit responsable.

Pour le point 105, c'est un peu le même scénario. Pour l'acp appliquée aux variables originales, c'est la première cp qui est hors-contrôle et donc encore ici seule la variable  $BR$  semble être responsable. On avait cependant trouvé que les deux premières cp et les variables  $BR$  et  $L$  étaient hors-contrôle lorsque l'on a appliquée l'acp aux variables normalisées.

Pour le jeu2, en appliquant la procédure de l'acp aux variables originales, on obtient cette fois-ci, les mêmes conclusions. La variable  $L$  est responsable pour le point 7 tandis que c'est la variable  $B$  qui semble hors-contrôle pour le point 30.



Si on décide d'appliquer la carte de contrôle aux deux premières composantes principales seulement, on obtient sensiblement les mêmes résultats. En effet, pour le jeu1, seuls les points 97, 105 et 183 étaient hors-contrôle pour cette procédure et si on regarde au tableau 3.2, on s'aperçoit que pour ces points ce sont soit la 1ère, la 2e ou les deux premières cp qui sont hors-contrôle. Il est donc évident que pour ces points les mêmes variables vont être responsables des signaux. Pour les points 14 et 100, on a trouvé que c'est la 4e cp qui est hors-contrôle, c'est donc pour cette raison que ces points ne furent pas hors-contrôle lorsque l'on a fait la carte de contrôle sur les 2 premières cp seulement. Cependant, le point 29 qui n'avait pas été détecté hors-contrôle sur la carte  $T^2$  sur les deux premières cp, semble être hors-contrôle sur la 2e et 4e cp lorsque l'on applique la méthode de Kourti et MacGregor. C'est donc l'effet combiné sur ces deux cp qui explique le signal.

Maintenant, pour le jeu2, si on applique la procédure mais pour les deux premières cp seulement, on arrive à la même conclusion que pour le point 7, c'est la variable  $L$  qui semble être responsable du signal.

Essayons maintenant de visualiser ces conclusions à l'aide de graphiques en trois dimensions provenant du logiciel statistique S-PLUS et plus particulièrement de la fonction `xgobi`. À l'aide de cette fonction, il est possible d'effectuer une rotation tridimensionnelle de trois des variables du procédé. La figure 3.29 représente les variables  $BR$ ,  $A$  et  $B$  pour les observations du jeu1. Sur le logiciel, on a analysé cette rotation tridimensionnelle et la figure 3.29 est une représentation fixe de cette rotation.

On doit spécifier que les 3 composantes principales sur ce graphique sont les cp pour les variables  $BR$ ,  $A$  et  $B$  et ne sont donc pas les mêmes cp que l'on a présentées précédemment, qui elles représentaient les cp des 4 variables du procédé.

Sur ce graphique, on peut y voir que la 1<sup>ère</sup> composante principale est l'axe (non représentée) qui part du sud-ouest et qui va au nord-est. C'est évidemment elle qui explique le plus de variabilité. La deuxième cp est représentée par l'axe (non représentée) qui part du sud-est et qui se dirige vers le nord-ouest. Finalement, la 3<sup>e</sup> cp est celle représentée par l'épaisseur du papier, c'est donc l'axe qui sort (ou entre) de la feuille. C'est la cp qui explique le moins de variabilité. On peut ainsi voir clairement que le point 97 est hors-contrôle sur la première cp et que le point 183 semble hors-contrôle sur les deux premières cp.

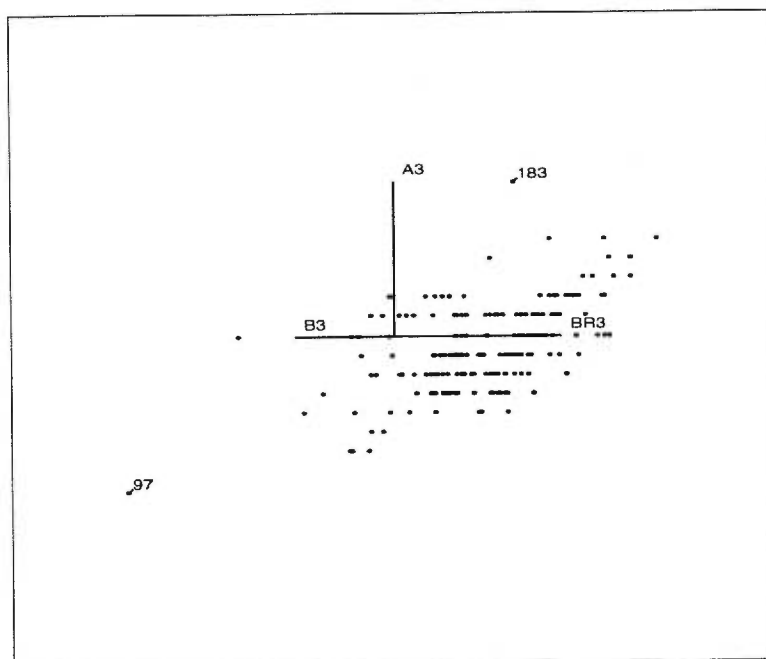


FIGURE 3.29. Représentation fixe de la rotation des variables  $BR$ ,  $A$  et  $B$  pour les observations du jeu1.

Si on prend la variable  $L$  à la place de la variable  $A$ , on sera en mesure de voir ce qui se passe sur les cp qui expliquent le moins de variabilité. La figure 3.30 représente le même graphique que précédemment mais pour les variables  $BR$ ,  $L$  et  $B$ .

On voit que les points 14 et 100 sont hors-contrôle, mais pour la dernière cp, celle qui explique le moins de variabilité, c'est-à-dire 0,2 % dans ce cas-ci. On s'aperçoit par ailleurs que même si ces points sont hors-contrôle, ils ne sortent

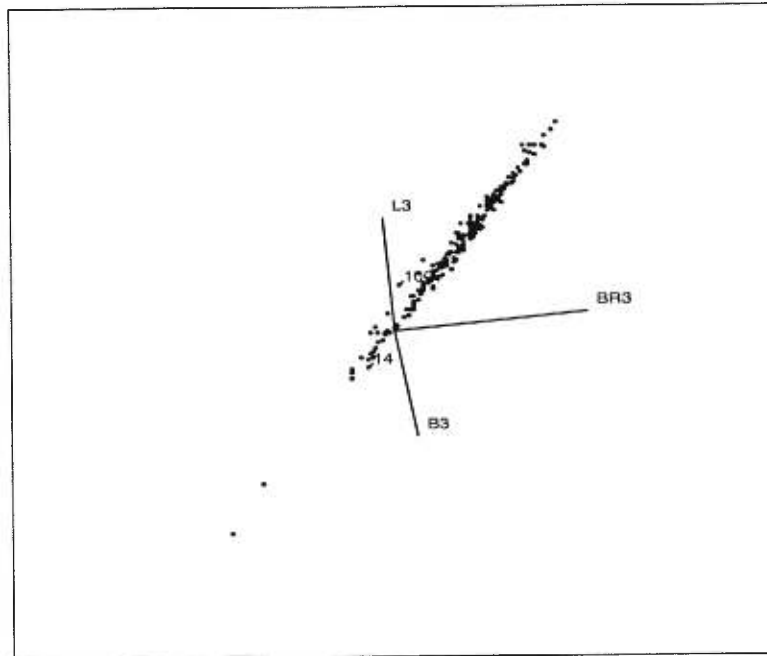


FIGURE 3.30. Représentation fixe de la rotation des variables  $BR$ ,  $L$  et  $B$  pour les observations du jeu1.

que très peu du nuage de points, car en fait ils sont situés près de l'origine des 3 variables. La question est donc de savoir s'il vaut la peine de signaler pour des points qui sont hors-contrôle sur les cp qui expliquent que très peu de variabilité, dans ce cas-ci 0,2 %, car en fait ces points ne sortent que très peu pour chacune des variables.

Pour le jeu1, voyons graphiquement ce que l'on obtient. La figure 3.31 est la représentation fixe de la rotation des variables  $BR$ ,  $L$  et  $B$  pour les 122 observations du jeu2. On voit clairement que les observations 7 et 30 sont hors-contrôle sur 3e cp, qui explique 0,4 % de la variabilité totale dans ce cas-ci.

### 3.5. CONCLUSION

À la suite de l'analyse des jeux de données, plusieurs conclusions peuvent être tirées.

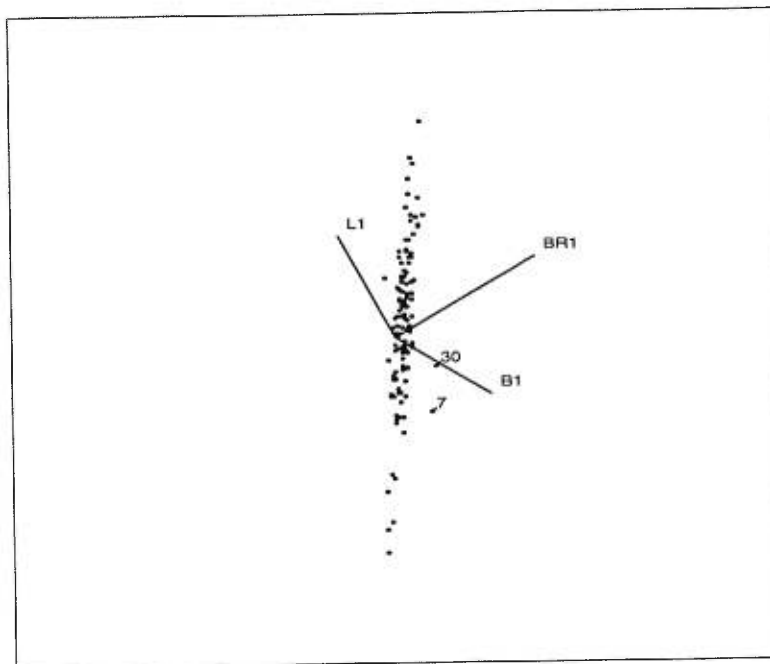


FIGURE 3.31. Représentation fixe de la rotation des variables  $BR$ ,  $L$  et  $B$  pour les observations du jeu2.

La première est que les données sont fortement auto-corrélées et que le modèle de série chronologique multivarié  $AR(1)$  s'ajuste bien à ces données. Dans un premier temps, cela fait en sorte que l'on devra privilégier comme estimateur de la matrice de covariance du procédé  $\Sigma$ , la matrice de covariance échantillonnale  $S$  et non  $\tilde{S}$  calculé à partir des étendues mobiles. Dans un deuxième temps, cela fait en sorte que l'on devra privilégier la carte  $T^2$  aux cartes cusum et ewma.

La deuxième conclusion est qu'il y avait des différences significatives entre les moyennes des variables des 3 jeux de données. Ainsi, il était impossible d'appliquer la carte aux données de la phase 2 (par exemple le jeu 2) à partir des estimés de la phase 1 (par exemple le jeu 1). On s'est donc contenté d'appliquer la carte aux données de la phase un, c'est-à-dire en analysant chacun des jeux séparément.

La troisième conclusion concerne l'analyse en composantes principales. On s'est aperçu que les 2 premières composantes principales comptent pour environ

90 % de la variabilité totale du procédé. On doit donc se questionner pour savoir s'il vaut la peine de considérer les 2 dernières cp puisqu'un tout petit déplacement sur l'un de ces axes qui varient peu sera rapidement détecté.

On doit également noter que pour l'acp, on doit se questionner pour savoir si on applique l'acp aux variables originales (et donc en décomposant la matrice de covariance  $\Sigma$ ) ou aux variables normalisées (et donc en décomposant la matrice de corrélation). Les résultats obtenus seront différents dépendant de la méthode choisie.

Pour ce qui est des observations hors-contrôle, on peut dire que les jeux de données étudiés étaient relativement en contrôle. Pour le jeu1, on a détecté 6 points hors-contrôle sur 201, alors que l'on en a détecté 2 sur 122 pour le jeu2 (lorsque l'acp est appliquée sur les 4 variables normalisées).

Pour la détection des variables hors-contrôle, on en conclut que la méthode d'acp de Kourti et MacGregor (1996) est la plus efficace. En effet, lorsque l'on applique les cartes de contrôle univariées aux 4 variables, seulement 3 points sur les 8 hors-contrôle identifient des variables responsables. Ainsi, pour 5 points sur 8, aucune variable n'est identifiée comme étant responsable du signal. Quant à elle, la statistique de Student nous donne une idée un peu plus précise des variables qui se sont le plus éloignées de leur moyenne respective, cependant cette statistique ne tient pas compte de la corrélation entre les variables. Ainsi, les conclusions tirées à partir des statistiques de Student ne sont pas les mêmes que celles tirées à partir de l'acp. Et on a pu démontrer graphiquement que les conclusions tirées à partir de l'acp sont valables.

Pour la méthode de Kourti et MacGregor appliquée aux variables normalisées, on s'est rendu compte que pour les 8 points hors-contrôle, aucune fois la 3e cp n'est hors-contrôle. Cependant, les 4 variables ont été identifiées comme pouvant

être responsable du signal dans une proportion à peu près équivalente (4 fois sur 8 pour les variables  $BR$  et  $L$ , 2 fois sur 8 pour  $A$  et 3 sur 8 pour  $B$ ).

On conclut donc à la suite de toute ces analyses qu'aucune variable ne semble être hors-contrôle plus souvent qu'une autre, qu'il n'y a pas de relation entre le fait qu'une variable soit hors-contrôle et le fait qu'une autre le soit ou pas, que si une cp est hors-contrôle, il est impossible de savoir si une variable ou une autre a plus de chance d'être hors-contrôle. La seule chose que l'on peut affirmer, c'est que si la 3e cp est hors-contrôle, il y a très peu de chance que ce soit attribuable à la variable  $BR$  et si la 4e cp est hors-contrôle, il y a peu de chance que la variable  $A$  soit en cause. Et finalement, que les conclusions sur les variables responsables des signaux tirées à partir des statistiques de Student ne sont pas les mêmes que celles tirées à partir de l'acp de Kourti et MacGregor et que l'on obtient également des conclusions différentes si on utilise les variables originales.

## CONCLUSION

---

Le but de ce mémoire était de présenter une procédure multivariée pour le contrôle statistique des procédés. Nous avons tout d'abord analysé les cartes de contrôle univariées au chapitre 1. Nous avons analysé les cartes Shewhart, cusum et ewma. Par la suite, on s'est concentré sur ces mêmes cartes mais au niveau multivarié.

On s'est tout d'abord aperçu que les cartes cusum et ewma obtenaient sensiblement les mêmes performances pour la longueur moyenne d'une suite pour un déplacement donné (pour une  $LMS_0$  fixe) et que ces performances étaient meilleures que celle de la carte  $T^2$ , spécialement pour les petits déplacements. Point de vue performance, on a donc tout avantage à privilégier les cartes cusum et ewma à la carte  $T^2$ .

Cependant, en considérant d'autres critères, on s'est aperçu que le meilleur choix n'est pas de choisir les cartes cusum et ewma pour les raisons suivantes. Premièrement, le choix des paramètres est plus facile pour la carte  $T^2$  qui ne dépend que de  $\alpha$ , l'erreur de type 1. On fixe donc  $\alpha$  et on calcule la limite de contrôle supérieure correspondante à l'aide d'une distribution connue (Bêta ou Fisher). Pour les cartes cusum et ewma, il est moins facile de modifier les limites de contrôle. Les auteurs qui ont introduit ces cartes ont certes fourni des tables de paramètres en fonction du déplacement  $\tau$  que l'on veut détecter, de la  $LMS$  désirée (et donc de  $\alpha$ ) et de  $p$  le nombre de variables, mais pour quelques valeurs

différentes seulement. Il n'est pas évident que pour un  $p$ , un  $\alpha$  et  $\tau$  quelconque, qu'il existe une table donnant la valeur des paramètres.

On a également conclu que l'auto-corrélation des données avait deux effets majeurs sur les cartes de contrôle. Comme premier effet, on a conclu que l'on devait prendre comme estimateur de la matrice de covariance du procédé, la matrice de covariance échantillonnale  $S$  et non  $\tilde{S}$  calculée à partir des étendues mobiles. La raison est que  $VAR(S)$  est toujours asymptotiquement égale à la variance des observations alors que  $VAR(\tilde{S})$  est fonction du modèle de série chronologique que l'on ajuste aux données.

Comme deuxième effet, on a conclu qu'il était préférable d'utiliser la carte  $T^2$  aux cartes cusum et ewma puisque cette carte se sert de la variance des observations pour le calcul des limites, tandis que pour les cartes cusum et ewma, on se doit de calculer la variance d'une autre statistique, qui est cumulée celle-là. Par exemple, pour la procédure ewma, l'expression de la variance de la statistique  $Z$  est fonction du modèle de série chronologique et est une expression assez complexe à calculer.

Ainsi, pour ces deux raisons, on préférera la carte  $T^2$  aux cartes cusum et ewma.

Pour la détection des variables hors-contrôle, on a conclu que la méthode de Kourti et MacGregor (1996) qui utilise l'analyse en composantes principales, est la plus efficace pour détecter les variables responsables d'un signal hors-contrôle. On suggère cependant de tout d'abord effectuer des cartes univariées des variables à l'aide des limites de Bonferroni, c'est-à-dire avec un niveau  $\alpha_u = \alpha_m/p$  pour avoir une idée de ce qui s'est passé avec les variables. Pour un point hors-contrôle, on peut également ordonner les variables en calculant la statistique de Student pour chacune des variables. Cela nous indiquera quelles variables se sont le plus éloignées de leur moyennes.



On doit également noter que l'acp n'est pas indépendante de l'échelle utilisée pour les variables et que l'on doit se questionner à savoir si on applique l'acp sur les variables originales ou sur les variables normalisées. On obtient des conclusions différentes si on utilise l'une ou l'autre.

## RÉFÉRENCES

---

- Alt, F.B. (1985). Multivariate Quality Control. Dans *Encyclopedia of Statistical Sciences*, Vol 6, Wiley and Sons, New York, 110-122.
- Anderson T.W. (1984). *An Introduction to Multivariate Statistical Analysis, 2e édition*, Wiley and Sons, New York.
- Brook, D. et Evans, D.A. (1972). An Approach to the Probability Distribution of CUSUM Run Length, *Biometrika*, **59**, 539-549.
- Chan, L.K. et Li, G.Y. (1994). A Multivariate Control Chart for Detecting Linear Trends, *Communications in Statistics: Simulation and Computation*, **23**, 997-1012.
- Cleveland W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association*, **74**, 829-836.
- Crosier, R.B. (1988). Multivariate Generalizations of Cumulative Sum Quality-Control Schemes, *Technometrics*, **30**, 291-303.
- Doganaksoy, N., Faltin, F.W. et Tucker, W.T. (1991). Identification of Out of Control Quality Characteristics in a Multivariate Manufacturing Environment, *Communications in Statistics: Theory and Methods*, **20**, 2775-2790.
- Fuchs, C. et Benjamini, Y. (1994). Multivariate Profile Charts for Statistical Process Control, *Technometrics*, **36**, 182-195.
- Fuchs, C. et Kenett, R.S. (1987). Multivariate Tolerance Regions and F-Tests, *Journal of Quality Technology*, **19**, 122-131.
- Gan, F.F. (1991). An Optimal Design of CUSUM Quality Control Charts, *Journal of Quality Technology*, **23**, 279-286.
- Hawkins, D.M. (1991). Multivariate Quality Control Based on Regression-Adjusted Variables, *Technometrics*, **33**, 61-75.
- Hawkins, D.M. (1992). A Fast Accurate Approximation for Average Run Lengths of CUSUM Control Charts, *Journal of Quality Technology*, **24**, 37-43.
- Hawkins, D.M. (1993a). Cumulative Sum Control Charting: An Underutilized SPC Tool, *Quality Engineering*, **5**, 463-477.
- Hawkins, D.M. (1993b). Regression Adjustment for Variables in Multivariate Quality Control, *Journal of Quality Technology*, **25**, 170-182.
- Hayter, A.J. et Tsui, K-L. (1994). Identification and Quantification in Multivariate Quality Control Problems, *Journal of Quality Technology*, **26**, 197-208.
- Healy, J.D. (1987). A Note on Multivariate CUSUM Procedures, *Technometrics*, **29**, 409-412.
- Holmes, D.S. et Mergen, A.E. (1993). Improving the Performance of the  $T^2$  Control Chart, *Quality Engineering*, **5**, 619-625.

- Hotelling, H. (1947). Multivariate Quality Control, Illustrated by the Air Testing of Sample Bombsights. Dans *Techniques of Statistical Analysis* (Eisenhart, Hasty et Wallis, éditeurs), McGraw-Hill, New York.
- Jackson, J.E. (1985). Multivariate Quality Control, *Communications in Statistics: Theory and Methods*, **14**, 2657-2688.
- Johnson, R.A. et Wichern, D.W. (1992) *Applied Multivariate Statistical Analysis, 3e édition* Prentice Hall, New Jersey.
- Kourti, T. et MacGregor, J.F. (1996). Multivariate SPC Methods for Process and Product Monitoring, *Journal of Quality Technology*, **28**, 409-428.
- Liu, R.Y. (1995). Control Charts for Multivariate Processes, *Journal of the American Statistical Association*, **90**, 1380-1387.
- Lowry, C.A., Woodall, W.H., Champ, C.W. et Rigdon, S.E. (1992). A Multivariate Exponentially Weighted Moving Average Control Chart, *Technometrics*, **34**, 46-53.
- Lucas, J.M. et Crosier, R.B. (1982). Fast Initial Response for CUSUM Quality Control Schemes, *Technometrics*, **24**, 199-205.
- Lucas, J.M. et Saccucci, M.S. (1990). Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements, *Technometrics*, **32**, 1-12.
- MacGregor, J.F. et Wong, A.K.L. (1980). Multivariate Model Identification and Stochastic Control of a Chemical Reactor, *Technometrics*, **22**, 453-464.
- Mandel, J. (1969). The Partitionning of Interaction in Analysis of Variance, *Journal of Research, National Bureau of Standards*, **73**, 309-327.
- Mason, R.L., Tracy, N.D. et Young, J.C. (1995). Decomposition of  $T^2$  for Multivariate Control Chart Interpretation, *Journal of Quality Technology*, **27**, 99-108.
- Mohebbi, C. et Hayre, L. (1989). Multivariate Control Charts: A Loss Function Approach, *Sequential Analysis*, **8**, 253-268.
- Montgomery, D.C. (1996). *Introduction to Statistical Quality Control, 3e édition*, Wiley, New York.
- Murphy, B.J. (1987). Selecting Out of Control Variables with the  $T^2$  Multivariate Quality Control Procedures, *The Statistician*, **36**, 571-583.
- Nomikos, P. et MacGregor, J.F. (1995). Multivariate SPC Charts for Monitoring Batch Processes, *Technometrics*, **37**, 41-59.
- Page, E.S. (1954). Continuous Inspection Schemes, *Biometrika*, **41**, 100-115.
- Patel, H.I. (1973). Quality Control Methods for Multivariate Binomial and Poisson Distributions, *Technometrics*, **15**, 103-112.
- Pignatiello, J.J. et Runger, G.C. (1990). Comparisons of Multivariate CUSUM Charts, *Journal of Quality Technology*, **22**, 173-186.
- Prabhu, S.S. et Runger, G.C. (1997). Designing a Multivariate EWMA Control Chart, *Journal of Quality Technology*, **29**, 8-15.
- Prabhu, S.S. et Runger, G.C. (1996). Analysis of a Two-Dimensional Markov-Chain, *Communications in Statistics: Simulation and Computation*, **25**, 75-79.
- Roberts, S.W. (1959). Control Charts Tests Based on Geometric Moving Average, *Technometrics*, **1**, 239-250.
- Runger, G.C. (1996). Projections and the  $U^2$  Multivariate Control Chart, *Technometrics*, **28**, 313-319.
- Runger, G.C. et Prabhu, S.S. (1996). A Markov Chain Model for the Multivariate Exponentially Weighted Moving Averages Control Chart, *Journal of the American Statistical Association*, **91**, 1701-1706.

- Scholz, F.W. et Tosch, T.J. (1994). Small Sample Uni and Multivariate Control Charts for Means, *Proceedings of the American Statistical Association, Quality and Productivity Section*, 17-22.
- Sparks, R.S. (1992). Quality Control with Multivariate Data, *Australian Journal of Statistics*, **34**, 375-390.
- Sullivan, J.H. et Woodall, W.H. (1996). A Comparison of Multivariate Control Charts for Individual Observations, *Journal of Quality Technology*, **28**, 398-408.
- Taguchi, G. (1986). *Introduction to Quality Engineering*, Asian Productivity Organization, White Plains, New York.
- Taguchi, G. et Wu, Y. (1980). *Introduction to Off-Line Quality Control*, Control Japan Quality Control Organization, Nagoya, Japan.
- Timm, N.H. (1996). Multivariate Quality Control Using Finite Intersections Tests, *Journal of Quality Technology*, **28**, 233-243.
- Tracy, N.D., Young, J.C. et Mason, R.L. (1995). A Bivariate Control Chart for Paired Measurements, *Journal of Quality Technology*, **27**, 370-376.
- Tracy, N.D., Young, J.C. et Mason, R.L. (1992). Multivariate Control Charts for Individual Observations, *Journal of Quality Technology*, **24**, 88-95.
- Tsui, K-L. et Woodall, W.H. (1993). Multivariate Control Charts Based on Loss Functions, *Sequential Analysis*, **12**, 79-92.
- Wei W.W. (1990). *Time Series Analysis, Univariate and Multivariate Methods.*, Addison-Wesley, Redwood City.
- Wheeler, D.J. et Chambers, D.S. (1992). *Understanding Statistical Process Control, 3e édition*, SPC Press, Knoxville.
- Wierda, S.J. (1994). Multivariate Statistical Process Control, Recent Results and Directions for Future Research, *Statistica Neerlandica*, **48**, 147-168.
- Woodall, W.H. et Ncube, M.M. (1985). Multivariate CUSUM Quality-Control Procedures, *Technometrics*, **27**, 285-292.