

Titre: Développement d'une méthodologie statistique pour la détection de
Title: l'usure d'un outil avec des cartes de contrôle

Auteur: Moraima Béatriz Portillo
Author:

Date: 2008

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Portillo, M. B. (2008). Développement d'une méthodologie statistique pour la
Citation: détection de l'usure d'un outil avec des cartes de contrôle [Master's thesis, École
Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/8396/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/8396/>
PolyPublie URL:

**Directeurs de
recherche:** Soumaya Yacout, & Marc Bourdeau
Advisors:

Programme: Unspecified
Program:

UNIVERSITÉ DE MONTRÉAL

DÉVELOPPEMENT D'UNE MÉTHODOLOGIE STATISTIQUE
POUR LA DÉTECTION DE L'USURE D'UN OUTIL
AVEC DES CARTES DE CONTRÔLE

MORAIMA BÉATRIZ PORTILLO
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INDUSTRIEL)
NOVEMBRE 2008



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-47682-6
Our file Notre référence
ISBN: 978-0-494-47682-6

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■ ■ ■
Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

DÉVELOPPEMENT D'UNE MÉTHODOLOGIE STATISTIQUE
POUR LA DÉTECTION DE L'USURE D'UN OUTIL
AVEC DES CARTES DE CONTRÔLE

présenté par : PORTILLO Moraima Béatriz

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. CLÉMENT Bernard, Ph.D., président

Mme. YACOUT Soumaya, D. Sc., membre et directrice de recherche

M. BOURDEAU Marc, Ph.D., membre et codirecteur de recherche

M. QUALI Mohamed-Salah, Doctorat., membre

*À ma famille
qui me soutient toujours.*

REMERCIEMENTS

À ma directrice de recherche, Madame la professeure Soumaya Yacout, qui m'a offert des très bons conseils tout au long de mes études et qui m'a soutenue quand j'en avais le plus besoin.

À mon codirecteur, Monsieur le professeur Marc Bourdeau, qui m'a inspiré à poursuivre ma recherche et à toujours aller plus loin.

Aux membres du jury, qui ont contribué à l'amélioration de ce travail.

À l'École Polytechnique et son personnel, pour m'avoir fourni les outils et la collaboration nécessaire pour finir ma maîtrise.

RÉSUMÉ

Le développement croissant de la maintenance prédictive ou conditionnelle incite à la recherche concernant le diagnostic et le pronostic de défaillances, et dans ce contexte, les cartes de contrôle prennent une large place au sein du contrôle statistique de procédés pour la surveillance de l'usure, sujet dont nous allons traiter ici.

Cette recherche présente une méthodologie statistique détaillée pour l'utilisation des cartes de contrôle unidimensionnelles et multidimensionnelles dans la détection de l'usure d'un outil de coupe. Le travail ajoute à la recherche déjà existante le fait que les données observées sont non stationnaires et corrélées. On utilise ainsi un modèle autorégressif afin de convertir les données corrélées en résidus non corrélés. Comme première étape, nous utilisons une analyse statistique des composantes principales pour classer l'état de l'outil dans différentes classes selon le degré de détérioration. On utilise des données réelles pour illustrer la méthodologie. L'analyse et les résultats sont décrits en détails.

MOTS-CLÉS : Maintenance prédictive, analyse de l'usure, cartes de contrôle.

ABSTRACT

The increasing development of predictive maintenance or conditional based maintenance encourages research about fault diagnosis and prognosis. Control charts are used in wear monitoring within the statistical process control.

This research presents a detailed statistically based technology using univariate and multivariate control charts for tool wear detection. This work adds to the already existing research the fact that observed data are non stationary and they are correlated. An autoregressive model is used in order to convert correlated data into uncorrelated residuals. It starts by using a Principal Components Analysis to classify the degrees of deterioration of the tool. A real application is used to illustrate the method. Analysis and results are described.

KEY WORDS: Predictive maintenance, tool wear analysis, control charts.

TABLE DES MATIÈRES

DÉDICACE.....	iv
REMERCIEMENTS.....	v
RÉSUMÉ.....	vi
ABSTRACT.....	vii
TABLE DES MATIÈRES.....	viii
LISTE DES TABLEAUX.....	x
LISTE DES FIGURES.....	xii
LISTE DES ANNEXES.....	xv
INTRODUCTION.....	1
CHAPITRE 1 : LE PROBLÈME	
1.1.- DESCRIPTION DU PROBLÈME.....	4
1.2.- REVUE DE LA LITTÉRATURE.....	8
CHAPITRE 2 : CAS UNIDIMENSIONNEL	
2.1.- DÉTERMINATION DES OBSERVATIONS QUI SERVIRONT POUR CONSTRUIRE LES CARTES DE CONTRÔLE	22
2.2.- TRAITEMENT DE LA CORRÉLATION.....	29
2.2.1.- IDENTIFICATION DU MODÈLE DE LA SÉRIE TEMPORELLE.....	30
2.2.2.- ESTIMATION DES MODÈLES AUTORÉGRESSIFS.....	38
2.2.3.- CONFIRMATION DES MODÈLES ET OBTENTION DES RÉSIDUS.....	39
2.2.4.- OBTENTION DES RÉSIDUS DES PRÉVISIONS.....	43
2.3.- CONSTRUCTION DES CARTES DE CONTRÔLE.....	45
2.3.1.- CARTE X BAR.....	45
2.3.2.- CARTE ÉTENDUE R	47

2.4.- DÉTECTION DE L'USURE	49
----------------------------------	----

CHAPITRE 3 : CAS MULTIDIMENSIONNEL

3.1.- CARTES DE CONTRÔLE MULTIDIMENSIONNELLES.....	53
3.2.- PRÉPARATION DES DONNÉES.....	54
3.3.- CARTE DE CONTRÔLE DE HOTELLING.....	62
3.3.1- RÈGLES POUR LA RECONNAISSANCE DE PATRONS DANS UNE CARTE DE HOTELLING.....	63
3.3.2.- CARTE D'HOTELLING POUR $n=1$	64
3.4.- CARTE DE CONTRÔLE POUR LA DISPERSION, $n=1$	69
3.5.- RÉSULTATS SUR LA DÉTECTION DE L'USURE.....	73
3.5.1.- CARTE DE HOTELLING.....	74
3.5.2.- CARTE DE LA DISPERSION.....	77
CONCLUSION.....	80
RÉFÉRENCES.....	84
ANNEXES.....	92

LISTE DES TABLEAUX

Tableau 2.1-	Données de la variable de rugosité $O_i(t), i = 1, \dots, 6; t = 1, \dots, 50$. Six mesures de la surface usinée pour chaque valeur de t . T est le temps cumulatif.....	23
Tableau 2.2-	Valeurs propres de l'ACP sur les données du tableau 2.1.....	25
Tableau 2.3-	Moyenne des valeurs de rugosité selon intervalle de t tirés de la classification.....	27
Tableau 2.4-	Choix du Modèle selon le type de corrélation.....	37
Tableau 2.5-	Résultats des modèles autorégressifs données par Statistica.....	39
Tableau 2.6-	Calcul de résidus des modèles sur les observations : $\varepsilon_i(t) = Y_i(t) - \hat{Y}_i(t), t = 3, \dots, 25; i = 1, 2$	42
Tableau 2.7-	Calcul des résidus pour les prévisions $\hat{Y}_i(t), t = 26, \dots, 50; i = 1, 2$...	44
Tableau 2.8-	Résumé des règles de Western Électric concernant la carte $X\text{-bar}$ de la figure 2-17.....	46
Tableau 3.1-	Données de la variable $W_i(t)$	56
Tableau 3.2-	Résidus du modèle AR(2) pour la variable $W(t), t = 1, \dots, 25$	61
Tableau 3.3-	Résidus ε_W obtenus pour les prévisions de la série $\hat{W}(t), t=26, \dots, 50$	62
Tableau 3.4-	Situations associées aux patrons dans les cartes T^2	64
Tableau 3.5-	Résidus employés pour la construction des cartes multidimensionnelles.....	65
Tableau 3.6-	Valeurs de $Q(t), t=3, \dots, 25$ pour la construction de la carte de Hotelling.....	67
Tableau 3.7-	Valeurs de $M(t), t=4, \dots, 25$ obtenues dans la phase I de la carte	

	de dispersion.....	71
Tableau 3.8-	Valeurs de $M(t)$, $t=4,\dots,25$ employées pour la carte de dispersion après le tamisage de données.....	72
Tableau 3.9-	Résidus employés pour la détection de l'usure.....	74
Tableau 3.10-	Valeurs de $Q(t)$, $t=26,\dots,50$ employées pour la détection de l'usure avec la carte de Hotelling.....	75
Tableau 3.11-	Valeurs de $M(t)$, $t=26,\dots,50$ employées pour la détection de l'usure avec la carte de dispersion.....	78
Tableau A.1-	Revenus selon les années de scolarité des 6 individus.....	95
Tableau A.2-	Méthode de Ward : solution de 5 groupes.....	95
Tableau A.3-	Méthode de Ward : solution de 4 groupes.....	96
Tableau B.1-	Matrice des corrélations R	103
Tableau B.2-	Matrice des valeurs test.....	103
Tableau B.3-	Intervalle de confiance pour les λ_i	105
Tableau B.4A-	Coordonnées des échantillons.....	106
Tableau B.4B-	Corrélations des échantillons avec les facteurs.....	106
Tableau B.4C-	Anciens axes unitaires.....	107
Tableau B.5-	Données triées aléatoirement.....	109
Tableau B.6-	Résultats des permutations effectuées.....	112
Tableau C.1-	Résumé des règles de Western Electric concernant la carte \bar{X} -bar de la figure C.2.....	116
Tableau C.2-	Comparaison des résultats des cartes de contrôle construites pour la série des données $O_i(t)$, $i=1,\dots,6$	120

LISTE DES FIGURES

Figure 2-1.	Représentation des observations selon l'ACP.....	26
Figure 2-2.	Dendrogramme de la classification de Ward sur les deux premières coordonnées factorielles.....	26
Figure 2-3.	Corrélation positive des données : $\bar{O}(t)$ versus $\bar{O}(t-1)$. $t = 1, \dots, 25$..	30
Figure 2-4.	Série temporelle de l'état de fonctionnement normal de l'outil. Observations $\bar{O}(t)$, $t = 1, \dots, 25$	31
Figure 2-5.	Série temporelle non stationnaire. $Y_1(t)$, $t = 1, \dots, 25$	33
Figure 2-6.	Série temporelle non stationnaire. $Y_2(t)$, $t = 1, \dots, 25$	33
Figure 2-7.	Série temporelle stationnaire $Z_1(t)$, $t = 2, \dots, 25$	34
Figure 2-8.	Série temporelle stationnaire $Z_2(t)$, $t = 2, \dots, 25$	34
Figure 2-9.	Autocorrélations de la série $Z_1(t)$, $t = 2, \dots, 25$	36
Figure 2-10.	Autocorrélations de la série $Z_2(t)$, $t = 2, \dots, 25$	37
Figure 2-11.	Autocorrélation du modèle AR (4) pour la série $\hat{Z}_1(t)$	40
Figure 2-12.	Autocorrélation du modèle AR (4) pour la série $\hat{Z}_2(t)$	40
Figure 2-13.	Distribution des résultats du modèle AR (4) pour la série $\hat{Z}_1(t)$	41
Figure 2-14.	Distribution des résultats du modèle AR (4) pour la série $\hat{Z}_2(t)$	41
Figure 2-15.	Prévisions des observations $\hat{Y}_1(t)$	43
Figure 2-16.	Prévisions des observations pour $\hat{Y}_2(t)$	43
Figure 2-17.	Carte <i>X-bar</i>	46

Figure 2-18.	Carte $X\text{-bar}$ après le tamisage des données.....	47
Figure 2-19.	Carte R après le tamisage des données.....	48
Figure 2-20.	Détection de l'usure avec la carte $X\text{-bar}$	49
Figure 2-21.	Détection de l'usure avec la carte R.....	50
Figure 3-1.	Série temporelle non stationnaire $W(t)$, $t = 1, \dots, 25$	57
Figure 3-2.	Série temporelle stationnaire $Z_W(t)$	58
Figure 3-3.	Fonctions d'autocorrélation de la série des données $Z_W(t)$	59
Figure 3-4.	Fonctions d'autocorrélation du modèle AR(2). Série $\hat{W}(t)$	60
Figure 3-5.	Distribution des résultats du modèle AR(2). Série $\hat{W}(t)$	60
Figure 3-6.	Prévisions du modèle AR(2). Série $\hat{W}(t)$	61
Figure 3-7.	Carte de Hotelling. Phase I.....	69
Figure 3-8.	Carte de dispersion. Phase I.....	72
Figure 3-9.	Carte de dispersion révisée. Phase I.....	73
Figure 3-10.	Détection de l'usure avec la carte de Hotelling.....	76
Figure 3-11.	Détection de l'usure avec la carte de dispersion.....	78
Figure A.1.	Dendrogramme obtenu avec le logiciel SPAD.....	97
Figure B.1.	Plan de projection.....	99
Figure B.2.	Analyse des observations. Randomisation.....	108
Figure B.3.	Classification hiérarchique directe. Randomisation.....	111
Figure C.1.	Les rugosités moyennes \bar{O} versus t	114
Figure C.2.	Cartes de contrôle X- bar et R.....	115

Figure C.3.	Carte X-bar pour une distribution autre que normale.....	117
Figure C.4.	Carte MA X-bar.....	118
Figure C.5.	Carte EWMA X-bar.....	119
Figure C.6.	Carte CuSum X.....	120

LISTE DES ANNEXES

ANNEXE A :	CLASSIFICATION DES DONNÉES.....	89
ANNEXE B :	L'ANALYSE EN COMPOSANTES PRINCIPALES.....	95
ANNEXE C :	ANALYSE PRÉLIMINAIRE DES DONNÉES.....	111

INTRODUCTION

Les machines souffrent d'une dégradation continue, c'est pourquoi les industries sont toujours à la recherche de nouvelles techniques de maintenance qui permettent de déterminer la période de vie utile d'un équipement et la stratégie optimale pour l'entretien.

Selon Jardine et *Al.* (2006), la première technique employée en maintenance fut la *maintenance corrective ou non planifiée*, appliquée quand l'équipement tombait en panne pour le remettre en marche. Ensuite, on a connu la *maintenance préventive ou planifiée*, où une routine de maintenance générale est appliquée à certains intervalles de temps du fonctionnement de la machine ; bien qu'innovatrice, cette technique peut être coûteuse, c'est pourquoi la *maintenance prédictive ou conditionnelle* l'a remplacée. La maintenance conditionnelle tient compte des avantages recherchés par l'industrie dont on a fait mention auparavant. La décision d'effectuer la maintenance est prise selon les conditions actuelles et futures de l'équipement par le biais d'indicateurs qui reflètent l'état de dégradation de la machine ; la maintenance étant effectuée quand le besoin se présente.

Parmi les méthodes de prise de décision employées dans le cadre de la maintenance conditionnelle, on trouve : la méthode de la Maîtrise statistique des procédés (MSP) (en anglais : Statistical Process Control ou SPC), la méthode du Modèle Caché de Markov (Hidden Markov Model - HMM) et la méthode du Modèle des Risques Proportionnels (Proportional Hazards Model - PHM). Les données dont on dispose pour notre recherche nous orientent vers la MSP. Comme on ne dispose uniquement que de mesures de rugosité des pièces coupées par un outil, les coûts de non qualité, et les données sur la fiabilité de cet outil ne sont pas disponibles. Nous allons utiliser cette caractéristique uniquement pour surveiller la performance de l'outil d'usinage en question et prédire la

défaillance ou l'usure excessive ainsi que le besoin de remplacement de l'outil, à l'aide des cartes de contrôle.

Pour utiliser les cartes de contrôle de façon efficace, le processus d'usinage sous contrôle statistique doit représenter la caractéristique de la qualité, qui est dans notre cas, la mesure de rugosité de la surface, selon le modèle de Shewhart : $Y_t \sim N(\mu, \sigma)$ pour le cas unidimensionnel et $Y_t \sim N_p(\mu, \Sigma)$ dans le cas multidimensionnel, à des temps plus ou moins régulièrement espacés. Les données du processus doivent être indépendantes dans le temps et normalement distribuées pour chaque temps, ce qui n'est pas le cas pour l'indépendance dans l'application présentée dans ce travail, puisque l'usure est un phénomène qui augmente continuellement avec le temps. Cela nous mène à l'objectif principal de cette recherche : *développer une méthodologie statistique pour la détection de l'usure d'un outil avec des cartes de contrôle.*

La démarche à suivre consiste d'abord en l'utilisation d'une analyse en composantes principales (ACP) suivie d'une classification afin de grouper les observations de la rugosité selon l'état de dégradation de l'outil. Cela permettra de déterminer les observations qui seront utilisées pour construire les cartes de contrôle. Ensuite, comme les observations sont corrélées dans le temps, on estimera un modèle autorégressif sur ces données. Les résidus non corrélés qui proviennent de la soustraction des données estimées par le modèle des données originelles nous permettront de construire les cartes de contrôle, unidimensionnelles et multidimensionnelles, qui serviront à surveiller l'usure et à identifier le temps de remplacement de l'outil de coupe en termes de minutes.

Nous supposons ainsi que pour un outil neuf, les résidus suivront un certain modèle de probabilité avec une moyenne μ et une variance σ^2 et qu'à un certain degré d'usure, le changement de la moyenne, de la variance ou des deux se manifestera par des points à l'extérieur des limites de contrôle ou par l'observation d'une tendance, signalant par là les temps où l'usure est excessive.

Notre travail comporte trois chapitres : le chapitre 1 présente une description détaillée du problème, ainsi qu'une compilation des travaux sur le sujet. Les chapitres 2 et 3 expliquent le développement de la méthodologie implantée et les résultats obtenus pour les cas unidimensionnel et multidimensionnel respectivement.

CHAPITRE 1

LE PROBLÈME

On présente ce préambule au chapitre afin de mieux décrire le sujet de la recherche.

1.1.- DESCRIPTION DU PROBLÈME

On rencontre le problème décrit dans ce paragraphe dans les compagnies qui pratiquent des opérations d'usinage.

Une compagnie de production des produits en aluminium a acquis de nouvelles machines à commandes numériques de marque Makino tournant à 18000 tours par minute (RPM). Ces nouvelles machines possèdent un magasin interne contenant 146 outils. Certains d'entre eux sont nouveaux, ils sont achetés spécifiquement pour les nouvelles machines, tandis que les autres fonctionnaient déjà sur les anciennes machines à commandes numériques qui tournent à 13000 RPM. (Caron, 2002.)

À cause de la différence entre les RPM des nouveaux outils et des anciens, il n'est pas possible de déterminer a priori la durée de vie qu'aurait un ancien outil dans une nouvelle machine, seulement en se basant sur l'historique des données. Cet historique fixait la durée de vie de l'outil à 200 minutes d'utilisation, période au-delà de laquelle les outils sont systématiquement changés.

Ne connaissant pas la durée de vie des anciens outils dans les nouvelles machines, deux situations ont lieu : d'une part, certains outils qui auraient la capacité de continuer au-delà de 200 minutes seraient retirés, causant ainsi un coût supplémentaire au niveau des achats d'outils et un gaspillage des ressources qu'il faut éviter. D'autre part, certains outils ne seraient pas en mesure de produire des pièces conformes à cause du niveau d'usure qu'ils

auraient subi avant la fin de la durée de 200 minutes, produisant ainsi des coûts de non qualité à cause de défaillances internes, que cela soit issu des coûts de rebut ou des coûts de reprise des produits.

Le problème est donc de déterminer la durée de vie utile d'un outil en termes de minutes, ou de trouver le temps auquel l'outil doit être remplacé.

Pour résoudre le problème, il a été décidé de surveiller la performance de l'outil sur la ligne de production à l'aide des cartes de contrôle afin de pouvoir prédire sa défaillance ou son usure.

Une carte de contrôle est un graphe statistique d'une caractéristique moyenne de la qualité qui est mesurée à partir d'échantillons cueillis à des intervalles de temps réguliers. Elles contiennent une ligne centrale (CL) qui représente la valeur moyenne globale de la caractéristique de la qualité. Deux lignes horizontales, appelées limites de contrôle supérieure (UCL) et inférieure (LCL) sont aussi présentées. Ces limites représentent la dispersion $\kappa\sigma$ autour de la moyenne, où traditionnellement κ est égale à 3. Les observations moyennes qui tombent hors des limites à un temps donné sont considérées comme une évidence d'un processus hors de contrôle, évidence donc du phénomène d'usure dans notre cas d'étude.

Même si toutes les observations moyennes se situent à l'intérieur des limites de contrôle, le processus est considéré hors de contrôle si les observations exhibent certains comportements. Il existe une série de critères, appelés règles de Western Electric pour identifier ce genre de dérèglement. Quand la cause de ce patron est trouvée et éliminée, le processus peut être amélioré ou redevenir sous contrôle.

Puisqu'il y a deux paramètres dans la distribution normale, la moyenne (μ) et la variance (σ), une observation à l'extérieur des limites de contrôle peut être causée par un

changement de l'un ou de l'autre, ou des deux. C'est pourquoi on utilisera conjointement deux cartes de contrôle : la carte qui mesure la tendance centrale et la carte qui mesure la dispersion.

L'expérience provenant de l'application des cartes de contrôle de la qualité en industrie a permis de spécifier des critères pour leur construction afin d'obtenir des bons résultats : *Les données du processus sous contrôle statistique doivent appartenir à une distribution normale, varier autour d'une moyenne fixe d'une manière stable (variance), dit l'état stationnaire, et ne pas être corrélées* (Wheeler, 1995 ; Montgomery, 2005 ; Mason et Young, 2007). Un tel processus, sous contrôle, représente une caractéristique de la qualité selon le modèle statistique de Shewhart : $Y_t \sim N(\mu, \sigma)$ pour le cas unidimensionnel et $Y_t \sim N_p(\mu, \Sigma)$ dans le cas multidimensionnel.

Quand ces critères sont satisfaits, nous pouvons employer les cartes de contrôle conventionnelles et obtenir des conclusions sur l'état du processus. Même dans des situations où le critère de normalité est partiellement violé, les cartes fonctionnent raisonnablement bien, mais si les observations sont corrélées, les cartes de contrôle présenteront un niveau trop élevé de fausses alarmes (Montgomery, 2005).

Notre cas d'étude comporte des données corrélées puisque l'usure augmente graduellement avec le temps. Ceci complique la problématique à laquelle on fait face, puisque dorénavant il ne s'agit pas seulement d'utiliser les cartes de contrôle pour déterminer le temps auquel un outil de coupage doit être remplacé, mais plutôt de *développer une méthodologie statistique, une technique de surveillance de l'usure d'un outil avec des cartes de contrôle lorsque les données sont corrélées.*

La méthodologie statistique que nous proposons est une démarche en trois étapes :

Premièrement, on emploie une technique d'analyse statistique multidimensionnelle, spécifiquement, l'analyse de composantes principales (ACP) suivie d'une classification ascendante hiérarchique selon le critère d'agrégation de Ward obtenue sur les axes principaux significatifs de l'ACP, pour *choisir les observations qui serviront pour la construction des cartes de contrôle.*

Deuxièmement, nous allons *développer un modèle mathématique autorégressif selon la méthodologie Box-Jenkins construit* sur l'intervalle des données qui serviront pour la construction des cartes, afin de modéliser leur structure corrélative.

Troisièmement, en utilisant les résidus, non corrélés et normalement distribués, résultant de la soustraction des observations théoriques obtenues par le modèle autorégressif aux données originelles, nous allons *construire les cartes de contrôle des résidus (unidimensionnelles et multidimensionnelles) et détecter le phénomène d'usure* lorsque les résidus montrent une tendance ou qu'ils se trouvent à l'extérieur des limites de contrôle des cartes.

Finalement, on pourra interpréter les résultats et prendre une décision à propos du temps auquel l'outil de coupage doit être remplacé.

Remarque importante. Il est important de remarquer que cette recherche pose les bases d'une méthodologie statistique, d'une technologie de surveillance de l'usure. Les données ne sont pas parfaites pour l'illustration, mais elles ont l'avantage d'être réelles. C'est ainsi que l'on admettra tout au long du texte que certaines hypothèses statistiques sont valides. Et nous continuerons de développer notre approche comme si elles étaient parfaites pour l'illustration. Notre objectif est *avant tout méthodologique*. Chaque application est un cas d'espèce et nos données demandent des adaptations *ad hoc*, que l'on appelle tamisage des données.

1.2.- REVUE DE LA LITTÉRATURE

Maintenant que l'on connaît les dimensions du problème issu de cette recherche ainsi que la méthodologie envisagée pour trouver une solution, on présente un bref aperçu des connaissances actuelles reliées à notre travail.

La maintenance prédictive ou conditionnelle

Ce travail se situe dans le cadre de la maintenance conditionnelle qui consiste à mesurer de manière périodique ou en continu des paramètres décrivant au mieux l'état de l'équipement afin de déterminer le moment où il faudra effectuer la maintenance. Au lieu de se fier à des valeurs statistiques moyennes, comme dans le cas de l'analyse de fiabilité (Goode et *Al.*, 2000), la maintenance prédictive utilise différents indicateurs pour déterminer le temps pour effectuer un entretien. Les valeurs de ces indicateurs sont analysées périodiquement afin de détecter des signes de dégradation de l'équipement. L'objectif est de prendre la bonne décision pour entretenir l'équipement au bon moment et avant la défaillance.

Récemment, plusieurs techniques de détection et de diagnostic de l'état d'un équipement ont été développées (Jardine et *Al.*, 2006). Ces techniques sont principalement basées sur deux approches : la modélisation de la dégradation et l'extraction des attributs spécifiques à la dégradation à partir des observations collectées. Chacune de ces approches présente des avantages et des inconvénients. Le premier ne s'applique qu'à certains phénomènes de dégradation mais ne requiert pas la surveillance d'un grand nombre d'indicateurs. Le deuxième, en se basant principalement sur le pouvoir d'extraction de l'information, demande une base de données plus volumineuse.

Dans ce travail, nous présentons une technique d'extraction des attributs spécifiques au phénomène de dégradation. Nous appliquons cette technique au problème de la

surveillance de l'usure des outils d'usinage basée sur l'utilisation des cartes de contrôle pour la maîtrise statistique de procédés (MSP). Notre travail s'intéresse à la surveillance d'un procédé de coupage afin qu'il donne des produits conformes aux spécifications à travers l'analyse séquentielle des échantillons du produit. Les cartes de contrôle permettront de vérifier si le processus exhibe une certaine variabilité naturelle ou s'il est hors de contrôle et que l'usure de l'outil de coupage est manifeste.

Les cartes de contrôle ont été amplement utilisées à cause des bénéfices en termes d'amélioration de la productivité et de détection des défauts. Triantafyllopoulos (2006) offre un résumé des travaux dans le domaine de la MSP pour le lecteur intéressé. Montgomery (2005) discute amplement des cartes de contrôle. On utilise cet ouvrage comme une référence tout au long de notre travail.

Les cartes de contrôle et la surveillance de l'usure

L'utilisation des cartes de contrôle pour la surveillance de l'usure fut présentée pour la première fois par Manuele (1945). Cet article mentionne que l'usure est manifeste dans les cartes de contrôle à travers un changement de moyenne sous la forme d'une tendance. Afin de produire conformément aux spécifications du processus, une modification de la carte de contrôle *X-bar* est proposée pour remettre à zéro le processus lorsque les observations, suite à une tendance, se retrouvent à l'extérieur des limites de contrôle, celles-ci étant établies à 3σ à partir des limites de spécification du produit. Mandel (1969) emploie le même modèle de tendance pour l'implantation d'une carte de contrôle qu'incorpore une analyse de régression dans une situation dans laquelle la charge de travail est variable mais fortement corrélée au nombre d'heures de travail.

Duncan (1986) présente un exemple sur la méthode développée par Manuele (connue sous le nom de cartes de contrôle d'acceptation). Il explique que, sous l'effet de l'usure, la dispersion normale d'un processus diminue et que la durée de vie de l'outil augmentera

ou diminuera selon le pourcentage d'articles défectueux tolérés dans le processus. C'est dans ce contexte que les limites de contrôle spéciales qui ont été développées par Manuele s'appliquent.

Quesenberry (1988) emploie les conclusions des recherches précédentes pour développer une stratégie pour l'ajustement de la moyenne d'un procédé sujet à l'usure d'un outil afin de préserver la qualité du produit fini. Il expose que dans un processus d'usinage, le but est de produire les pièces le mieux possible avec l'équipement disponible. Les causes de variation présentes dans le processus sont attribuées à l'environnement, à la qualité de la matière première, à la qualité de l'outil employé et à l'usure de cet outil.

Selon Quesenberry, l'usure se manifeste dans la caractéristique de la qualité du produit fini sous la forme d'une tendance linéaire qui augmente selon l'ordre de production. Dans l'article, ce phénomène est lié à l'autocorrélation. L'auteur se sert d'un modèle de régression pour décrire le processus, à la place des méthodes d'essais et erreurs qui sont employées communément en processus d'usinage.

L'intérêt croissant pour la maintenance prédictive a donné lieu à plusieurs articles de recherches récentes qui visent le diagnostic et le pronostic des défaillances. Ces recherches étudient des méthodes de détection de la dégradation. Les cartes de contrôle sont utilisées pour la détection de l'usure. Fugate *et al.* (2001), utilisent les cartes de contrôle pour détecter le changement de niveaux de vibrations dans une structure en béton, signalant ainsi la détérioration de son état. Liao *et al.* (2006) comparent les cartes de contrôle pour l'exploration des données extraites d'une série chronologique avec une méthode de classification basée sur un algorithme génétique. Makis (1999) présente un modèle mathématique pour le phénomène de l'usure d'un outil et en tire une politique de remplacement optimal.

Notre travail repose sur la démarche suivante. D'abord, on considère que l'usure se manifeste dans la rugosité des pièces coupées par un outil sous la forme d'une tendance qui augmente selon l'ordre de production. Ce phénomène peut être modélisé grâce à un modèle autorégressif. Ensuite, on utilise les résidus indépendants et normalement distribués obtenus par le biais du modèle pour la construction des cartes de contrôle. Le phénomène d'usure se visualise alors lorsque survient un changement de moyenne sous la forme de tendance.

Les cartes de contrôle et l'autocorrélation

Le contrôle statistique de processus repose sur le postulat selon lequel la fabrication des produits est faite avec des caractéristiques indépendantes et identiquement distribuées (*iid*) dans le temps tant que le processus de production est affecté seulement par des causes communes de variation. Avec l'arrivée des systèmes de cueillette de données à grande vitesse, le postulat de l'indépendance est habituellement violé. Dans la plupart des cas, les mesures de processus ne sont plus *iid* même s'il n'y a aucune cause spéciale affectant le processus (Atienza et Al. 1997.) Pour faire face à ce problème, les praticiens sont divisés.

Par exemple, Wheeler (1992) signale que l'autocorrélation extrême (supérieure à +0.80) serrera les limites de contrôle et que celle-ci crée simultanément des données d'interprétation sans risque, parce que le processus communique un manque de contrôle statistique sous la forme de déplacement. La présence d'autocorrélation élevée ne serait donc plus qu'un mythe dans la construction des cartes de contrôle. Il faut remarquer qu'un traitement complet des effets de l'autocorrélation est au-delà de la portée du texte publié en 1992, et que cette robustesse de la méthode sous autocorrélation n'est pas signalée à notre connaissance par d'autres auteurs.

Dans l'édition de 1995, l'auteur précise que les effets de l'autocorrélation sur les cartes de contrôle seront les plus graves quand la taille de l'échantillon (utilisé pour la construction des cartes de contrôle) est $n=2$. Il ajoute que les limites de contrôle obtenues par les formules habituelles seront essentiellement correctes ou souillées par l'auto-corrélation. Il déclare aussi que la décision concernant le point de vue à adopter dépendra de la connaissance du processus et de l'interprétation soigneuse des cartes de contrôle. Finalement, la multiplication de la valeur des limites de contrôle par un facteur de correction qui tient compte de la quantité d'auto-corrélation est recommandée. Cette procédure de correction ne deviendra efficace que si l'autocorrélation dépasse la valeur de 0.7.

D'autres praticiens concluent que le problème de l'indépendance est dû au prélèvement inexact des observations. Par exemple, Hoerl et Palm (1992) ont proposé de prendre les échantillons moins fréquemment afin de pallier la structure corrélative des données, ce qui peut se vérifier par des tests. Avec l'automatisation de la cueillette des données, cette suggestion est équivalente à jeter une certaine partie des données recueillies. (Dooley et Kapoor, 1990.)

À l'autre extrémité du spectre, Montgomery (2005) signale que la plus importante des hypothèses faites au sujet des cartes de contrôle est celle de l'indépendance des observations, parce que les cartes de contrôle conventionnelles donnent des résultats fallacieux sous la forme d'un nombre élevé de fausses alertes lorsque les données sont corrélées, même si cette corrélation est basse, souligne-t-il. Montgomery cite de nombreux auteurs reconnus qui ont démontré ce fait et recommande ensuite divers traitements pour les données corrélées dont l'utilisation des cartes de contrôle unidimensionnelles construites à partir des résidus.

D'après Atienza et *Al.* (1997), Montgomery préconise la modélisation de la structure inhérente d'autocorrélation utilisant des modèles de Box-Jenkins pour détecter les

changements d'un processus en surveillant les résidus issus des différents modèles. L'article souligne que la plupart des recherches actuelles en MSP sont centrées sur le traitement de processus corrélés et l'application des modèles de Box-Jenkins. Ce qui n'est pas surprenant si l'on considère que l'utilisation des modèles statistiques dans la MSP ne détourne pas des idées originales de Shewhart qui déclare que la MSP vise à réaliser des prévisions et que, à cet effet on a besoin de certaines lois ou modèles, soit empiriques, soit statistiques (Wheeler, 1992.)

Young et Winistorfer (2001) signalent que tous les types des cartes de contrôle pour surveiller la moyenne de processus sont affectés par l'autocorrélation positive, y compris la carte de contrôle de Shewhart, la somme cumulative (CUSUM), et la carte de la moyenne mobile pondérée exponentiellement (EWMA).

Apley et Tsung (2002) donnent un résumé des travaux à propos de l'emploi des résidus d'un modèle mathématique pour la construction des cartes de contrôle. Notre intérêt se centre sur ceux qui emploient un modèle autorégressif parce qu'on l'emploie dans le traitement de l'autocorrélation, tel que celui de Fugate *et al.* (2001) -rapporté plus haut- et celui de Barone *et al.* (2006), qui utilisent les cartes de contrôle pour la détection des données aberrantes ainsi que pour la détermination des problèmes associés aux bases de données de véhicules de promenade.

Mason et Young (2007) ont publié une révision des considérations nécessaires pour implanter le contrôle statistique de processus unidimensionnels. Ils signalent aussi l'importance d'avoir des données indépendantes et issues d'une distribution normale. Pour confirmer la dépendance, ils conseillent d'observer le comportement des données par le biais d'une série temporelle et du coefficient d'autocorrélation, et c'est cette méthode que nous allons employer pour démontrer que nos données sont corrélées.

La consultation des références que l'on a discutées dans cette section nous permet d'établir les conclusions suivantes :

- La présence significative d'autocorrélation parmi les données est nuisible pour la construction des cartes de contrôle ;
- Si une carte est employée sous de telles conditions, on obtiendra trop souvent des fausses alarmes quand en réalité le processus se trouve sous contrôle statistique (Charnes, 1995; Montgomery, 2005.) ;
- Une méthodologie pour le traitement de l'autocorrélation est nécessaire en vue de la construction appropriée des cartes de contrôle.

On remarque que même si la quantité d'auto-corrélation présente dans les données du cas d'étude employé n'est pas trop élevée d'après certains auteurs, nous allons nous en servir ici afin d'illustrer la méthodologie de surveillance de l'usure proposée.

L'utilisation non traditionnelle de l'ACP

L'analyse en composantes principales (ACP) est utilisée dans plusieurs domaines de recherche (Yoon et McGregor, 2000) dont le contrôle statistique multidimensionnel de procédés (CSMP) (en anglais : MSPC) (Kano *et al.* 2003). La plupart des travaux dans ce domaine emploient l'ACP de façon *traditionnelle* pour réduire la dimension des données et construire des cartes de contrôle à partir des résidus, où le tableau rectangulaire soumis à l'analyse est un tableau « sujets x variables ». Un exemple est montré par Baydar et *Al.* (2001). Dans cet article, les signaux de vibration obtenus à partir des senseurs différents, sont résumés grâce à l'ACP. Ensuite, les statistiques Q et T^2 sont calculées sur les résidus des composantes principales qui permettent de formuler un modèle de référence sur l'état normal du processus. C'est ainsi que les conditions de panne, qui sont interprétées comme des déviations du modèle, sont détectées en surveillant les statistiques. On peut

citer d'autres travaux sur le sujet, comme celui de He *et Al.* (2006), Sohn *et Al.* (2000), ainsi que celui de Prins et Mader (1997).

Une autre application consiste en l'utilisation de l'ACP pour déterminer les variables clés d'un processus. Kosanovich *et Al.* (1994) emploient cette technique parce qu'elle tient compte des relations parmi les variables et de leur influence sur la qualité du produit. Ils emploient les résultats pour implanter une stratégie améliorée de contrôle automatisé.

En maintenance, les observations qui sont employées pour la construction des cartes de contrôle sont choisies de manière subjective. On trace la série temporelle des observations et leur nombre (utilisé pour la construction des cartes) est choisi d'après l'allure de la courbe. Dans ce travail, nous allons nous servir de l'ACP pour effectuer une analyse exploratoire qui permettra de visualiser et de découvrir des phénomènes tels qu'ils sont portés par les données (Spad 5.0, 2002.) Il s'agit de regrouper les observations (sous un critère de classification) afin de trouver celles qui font partie de l'état de fonctionnement normal de l'outil de coupage sous étude. C'est ainsi que les observations qui serviront pour construire les cartes de contrôle seront sélectionnées par le biais d'un algorithme formalisé et non de façon arbitraire.

Une telle approche a été largement utilisée dans le contexte reconnaissance des formes et de l'exploration des données (Siripatrawan, 2008). Par exemple, Carvalho *et Al.* (2003) emploient l'ACP afin de classifier différents types de solvants selon leur degré de toxicité. Tong *et Al.* (2006) démontrent qu'une classification faite sur les deux premières composantes principales de l'ACP est efficace dans l'inspection des carreaux murals. L'ACP a permis de classifier les carreaux par rapport à leur degré d'adhérence en deux groupes bien séparés.

Siripatrawan (2008) emploie l'ACP pour différencier deux sources de bactéries. Il remarque que l'ACP fournit de l'information complémentaire qui n'est pas accessible

dans l'analyse conventionnelle des données. Lozano et Al. (2008) utilisent l'ACP pour la reconnaissance des formes. Ils ont obtenu un taux de succès de 97% quant à la classification des vins par rapport aux différents processus de vieillissement mis en oeuvre.

La détection de la dégradation à l'aide de l'ACP suivi d'une classification de Ward semble donc un élément nouveau en maintenance. Autre nouvel élément que l'on présente dans ce travail est la nature de la matrice des données employée pour effectuer l'analyse. L'ACP est une technique employée pour la réduction des données généralement lorsqu'il y a *plusieurs variables* (Buroche et Saporta, 1980). Notre étude met en jeu six échantillons d'une même variable, la rugosité d'une surface usinée, cueillis à 50 temps différents. Les six mesures de cette variable seront traitées comme s'il s'agissait des variables différentes.

On trouve cet esprit dans l'origine de l'analyse. Dans le titre de son ouvrage, Karl Pearson (1901) laisse ouverte la question sur la forme de la matrice. Les données sont considérées comme un système de points dans l'espace. On peut donc soumettre à une ACP n'importe quel tableau rectangulaire $n \times p$. L'ACP, c'est de la géométrie. La question qui se pose est de trouver un sens aux axes d'inertie, donner un sens aux projections. Le problème ne se pose pas dans notre cas, où l'interprétation et l'utilité sont bien établies.

On peut citer quelques exemples qui se rapprochent de la façon dont on applique l'analyse. Kirby (2001) emploie des matrices où les éléments $p(i, j)$, sont les valeurs des pixels (intensité) dans une image $n \times p$ dont on cherche des caractéristiques. Ce serait solliciter les termes que de voir dans ce tableau une matrice (sujets x variables). Morineau et Aluja-Banet (1999) cherchent à réguler automatiquement la température dans la grotte de Lascaux. La matrice comporte 418 dates (lignes) et 30 stations (colonnes) de mesurage de la température. Lorsque pour un jour donné, on a un état

stable, plusieurs ou toutes les stations mesurent la même chose de façon qu'on peut supposer indépendante, on aurait alors des réalisations *iid*, soit des échantillons. C'est ainsi que plusieurs colonnes en réalité, ou toutes c'est selon, sont des échantillons de la même variable pour ces lignes-là. Les auteurs ne font évidemment pas mention du fait que des parties de leurs lignes sont en effet des échantillons. Ils ne pensent pas en ces termes. On a des points-lignes et des points-colonnes. On examine un nuage de 418 points dans \mathbb{R}^{30} .

Garcia-Allende et *al.* (2008) ajustent les caractéristiques d'un spectromètre par réflectivité qui produit des images spectrographiques. Les images qu'on soumet à des ACP pour compresser l'information sont des matrices 640×480 . À la verticale, 640 éléments du spectre, l'axe spectral, pour les 480 lignes de l'axe spatial horizontal. Évidemment, aucune référence à des tableaux sujets \times variables. Les projections sur des espaces de dimensions réduites sont des images compressées. Et bien sûr, pour une bonne partie des fréquences (lignes), les colonnes sont des échantillons de la même mesure avec ou non l'indépendance pour une grande part d'entre eux. Les concepts d'échantillons n'ont d'ailleurs aucune espèce d'intérêt : on veut compresser l'image, avant de la soumettre à des analyses en aval. Dans d'autres applications, on prend parfois la peine de déconnecter (rendre les échantillons indépendants) les points sur l'axe horizontal (les colonnes) en en prenant seulement un certain nombre.

Quoi qu'il en soit de ce qu'on peut trouver dans la littérature où beaucoup de tableaux soumis aux ACP ne sont pas du type (sujets \times variables), il est probable que ce soit la première fois qu'on trouve une application ACP à des tableaux où les colonnes sont systématiquement et uniquement des échantillons d'une même mesure. Mais comme partout dans les techniques de réduction de la dimension, c'est l'utilité des résultats et interprétations qu'on en tire qui la justifie.

Dans notre cas, on a en effet des valeurs échantillonales en colonnes qui constituent un nuage de points dans R^6 . On compresse l'information en deux dimensions, R^2 . Les projections des lignes dans R^2 sont étiquetées par le temps. Le premier axe devient un axe d'évolution dans le temps, c'est l'évolution de la rugosité, l'usure. Le second ne nous intéresse pas vraiment. Ce sont les résultats de la classification de Ward sur les deux premières coordonnées issues de l'ACP, les seules significatives, qui sont intéressantes pour nous : les projections sont groupées en trois (sous la classification non supervisée de Ward).

Les validations, par randomisation des colonnes de chaque ligne, pour simplettes qu'elles apparaissent, et peut-être trop peu nombreuses (quoique la cohérence des résultats plaident que non), nous indiquent que ces trois groupes ne sont probablement pas des artefacts des ordres pour chaque temps dans lequel les échantillons sont mesurés, lesquels ordres sont sans importance (voir l'annexe B), mais auraient pu perturber les résultats de l'ACP+Ward, ce qui n'est vraisemblablement pas le cas.

Cette cohérence issue des validations décrites à l'annexe B nous semble intéressante et apporte du poids à la conclusion concernant le nombre des observations, dites de l'état de fonctionnement normal de l'outil, qui serviront pour construire les cartes de contrôle. Quoi qu'il en soit, on peut raisonnablement conclure que les résultats du mémoire ne sont pas des artefacts de la procédure de cueillette des données.

Partant donc de cet état neuf de l'outil, on est en mesure maintenant de mettre en action des cartes de contrôle qu'il faudra établir avec soin.

Les cartes de contrôle multidimensionnelles

Dans les dernières années, le CSMP a repris de l'importance à cause de la quantité de processus qui sont observés par le biais de vecteurs (Quesenberry, 2001; Montgomery,

2005). Les cartes de contrôle multidimensionnelles sont discutées dans plusieurs articles : par exemple, Tracy *et al.* (1992), Vargas (2003), Prins et Mader (1997), Sullivan et Woodall (2000), Bersimis *et Al.* (2007), Mason *et Al.* (1997). Yoon et Macgregor (2000) signalent les avantages et les inconvénients du MSPC. La méthode est facile à implanter à cause de la disponibilité de l'historique de données et de la possibilité de détecter presque tous les types de défaillances. Cependant, l'interprétation d'un signal hors de contrôle n'est pas toujours simple puisqu'il faut découvrir ce qui a causé un tel état.

La plupart des recherches pour le contrôle de la moyenne en CSMP sont centrées sur la carte de Hotelling (Hotelling, 1947). Nous avons consulté différents articles pour tirer des renseignements importants à propos de cette carte. Prins et Mader (1997) présentent sa construction pour des groupes d'observations et pour des observations individuelles. Mason *et al.* (2003) fournissent des instructions pour l'identification de patrons dans cette carte. Mason *et al.* (1997) expliquent une méthode de décomposition de signaux multidimensionnels qui permet d'identifier la variable qui cause une situation hors de contrôle dans une carte. Gonzalez et Rodriguez (2003) présentent aussi un exemple sur la méthode de décomposition de Mason *et al.* On remarque dans ces articles l'importance d'effectuer une évaluation des observations aberrantes, de même qu'une analyse de la grandeur de l'autocorrélation présente parmi les données afin de bien construire la carte de contrôle.

Chen *et al.* (2005) utilisent des simulations pour mener une étude sur la performance des cartes de contrôle par rapport aux différents degrés de changement. Notre recherche exclut la simulation.

L'usure se manifeste dans la carte de contrôle lorsque survient un changement graduel de la moyenne. Malgré le pouvoir que possède la carte de Hotelling pour détecter ces changements (Chan et Li, 1994), sa performance peut diminuer puisque la matrice de covariance employée dans le calcul du statistique T^2 est constante dans le temps (Vargas

et Lagos, 2007). À ce propos, on utilise les différences successives des observations (Sullivan et Woodall, 2000). Vargas (2003) conseille l'utilisation d'un estimateur de la matrice de variances et de covariances qui soit basé sur ces différences pour construire une deuxième carte de Hotelling et comparer les résultats avec la carte construite avec l'estimateur classique de la matrice de variances et covariances. D'autre part, Surtihadi et *Al.* (2004) donnent des exemples où l'on constate l'influence de la matrice de variances et covariances et ils montrent que les cartes de dispersion sont en mesure de détecter ces changements.

Nous avons décidé d'utiliser l'estimateur classique de la matrice de variances et covariances pour construire la carte de Hotelling, et conjointement, on se servira de la recommandation proposée par Koo et Quah (2003) pour la construction d'une carte de dispersion qui est basée sur la différence successive des observations afin de constater l'impact que génère l'utilisation de cet estimateur.

CHAPITRE 2

CAS UNIDIMENSIONNEL

Notre cas d'étude comporte des données corrélées. Ceci exige l'implantation d'une méthodologie statistique pour éliminer autant que possible l'autocorrélation présente dans les données, et examiner la performance de l'outil sur la ligne de production avec des cartes de contrôle construites sur les résidus du modèle afin de prédire sa défaillance ou son usure excessive. (Voir l'annexe C pour obtenir des renseignements sur le processus lorsque la méthodologie proposée dans ce mémoire n'est pas appliquée.)

Ce chapitre présente les détails concernant le développement des étapes de la méthodologie lorsqu'on dispose d'*une seule variable* pour effectuer l'analyse :

- étape 1 : choix des observations de l'état de fonctionnement normal de l'outil qui serviront pour la détermination des cartes de contrôle, en utilisant l'analyse de composantes principales suivi d'une classification ;
- étape 2 : développement d'un modèle autorégressif selon la méthodologie Box-Jenkins sur ces données de l'outil à l'état de fonctionnement normal pour réduire sinon éliminer la corrélation présente dans les données ;
- étape 3 : construction des cartes de contrôle unidimensionnelles à partir des résidus du modèle.

Les résultats sur la détection du phénomène d'usure seront interprétés afin de trouver le temps auquel l'outil de coupage doit être remplacé selon les cartes unidimensionnelles *X-bar* et *R*.

2.1.- IDENTIFICATION DES OBSERVATIONS QUI SERVIRONT POUR CONSTRUIRE LES CARTES DE CONTRÔLE

Dans cette section, on illustre la première étape de la méthodologie. L'objectif est d'utiliser une analyse en composantes principales, suivie d'une classification ascendante hiérarchique selon le critère d'agrégation de Ward, pour choisir le nombre d'observations qui seront utilisées pour déterminer les cartes de contrôle.

Pour le cas d'étude unidimensionnel, on dispose de 50 observations réelles qui sont composées de six mesures d'une variable de rugosité de la surface usinée des pièces coupées par l'outil à des intervalles réguliers. Le tableau 2.1 présente les données des six mesures de rugosité prises toutes les 240 secondes. On appellera cette variable $O_i(t), i = 1, \dots, 6; t = 1, \dots, 50$, t étant le moment de prélèvement de l'observation. T est le temps cumulatif en secondes. Actuellement, au temps $T=12000$ secondes, l'outil est remplacé.

Si on voulait représenter les 50 observations dans l'espace, celles-ci formeraient un nuage peu visible dans un espace à six dimensions qu'on est donc incapable de classifier visuellement. Pourtant, ces observations comportent les renseignements sur l'état de dégradation de l'outil de coupe puisque l'usure augmente avec le temps. Nous avons donc décidé d'employer une analyse en composantes principales (ACP) afin de trouver l'information essentielle à propos de l'état de fonctionnement normal de l'outil contenu dans le tableau des données d'une façon cachée. L'ACP sert uniquement pour trouver deux axes qui serviront pour représenter les observations dans un espace de deux dimensions ; une classification de Ward sur les observations permettra de les regrouper en classes homogènes et de déterminer le nombre d'observations qui serviront pour construire les cartes de contrôle. On utilisera des permutations des échantillons de chaque observation ou temps t pour valider cette conclusion.

Tableau 2.1- Données de la variable de rugosité $O_i(t), i = 1, \dots, 6; t = 1, \dots, 50$.Six mesures de la surface usinée pour chaque valeur de t . T est le temps cumulatif en secondes. (Caron, 2002.)

t	T	O_1	O_2	O_3	O_4	O_5	O_6
1	240	0,1080	0,1142	0,1085	0,1074	0,1094	0,1103
2	480	0,1132	0,1118	0,1106	0,1105	0,1108	0,1119
3	720	0,1143	0,1122	0,1071	0,1086	0,1105	0,1105
4	960	0,1111	0,1136	0,1076	0,1080	0,1072	0,1126
5	1200	0,1103	0,1119	0,1071	0,1077	0,1079	0,1123
6	1440	0,1092	0,1140	0,1071	0,1081	0,1084	0,1100
7	1680	0,1107	0,1081	0,1060	0,1078	0,1086	0,1105
8	1920	0,1149	0,1118	0,1054	0,1069	0,1062	0,1106
9	2160	0,1144	0,1145	0,1053	0,1097	0,1058	0,1080
10	2400	0,1094	0,1142	0,1056	0,1065	0,1054	0,1083
11	2640	0,1040	0,1111	0,1096	0,1053	0,1074	0,1072
12	2880	0,1099	0,1083	0,1095	0,1062	0,1081	0,1079
13	3120	0,1112	0,1088	0,1075	0,1045	0,1081	0,1064
14	3360	0,1076	0,1105	0,1071	0,1064	0,1040	0,1082
15	3600	0,1066	0,1085	0,1075	0,1062	0,1041	0,1083
16	3840	0,1054	0,1109	0,1047	0,1048	0,1067	0,1087
17	4080	0,1071	0,1041	0,1082	0,1079	0,1056	0,1090
18	4320	0,1119	0,1084	0,1065	0,1076	0,1086	0,1079
19	4560	0,1113	0,1114	0,1088	0,1085	0,1051	0,1062
20	4800	0,1056	0,1111	0,1063	0,1053	0,1087	0,1084
21	5040	0,1060	0,1158	0,1073	0,1077	0,1070	0,1079
22	5280	0,1141	0,1120	0,1064	0,1075	0,1074	0,1062
23	5520	0,1159	0,1126	0,1082	0,1065	0,1088	0,1084
24	5760	0,1109	0,1149	0,1077	0,1089	0,1068	0,1083
25	6000	0,1096	0,1121	0,1083	0,1070	0,1082	0,1079

Tableau 2.1- Données de la variable de rugosité

$$O_i(t), i = 1, \dots, 6; t = 1, \dots, 50.$$

Six mesures de la surface usinée pour chaque valeur de t .

T est le temps cumulatif en secondes. (Caron, 2002.) (Suite et fin.)

26	6240	0,1079	0,1155	0,1096	0,1111	0,1112	0,1086
27	6480	0,1102	0,1062	0,1094	0,1100	0,1091	0,1071
28	6720	0,1168	0,1120	0,1106	0,1077	0,1111	0,1097
29	6960	0,1160	0,1162	0,1100	0,1097	0,1078	0,1098
30	7200	0,1082	0,1157	0,1086	0,1099	0,1118	0,1063
31	7440	0,1097	0,1161	0,1101	0,1093	0,1108	0,1084
32	7680	0,1124	0,1084	0,1117	0,1117	0,1119	0,1073
33	7920	0,1110	0,1139	0,1090	0,1102	0,1116	0,1074
34	8160	0,1080	0,1068	0,1094	0,1097	0,1091	0,1078
35	8400	0,1160	0,1085	0,1091	0,1095	0,1097	0,1061
36	8640	0,1088	0,1094	0,1091	0,1099	0,1100	0,1067
37	8880	0,1156	0,1075	0,1098	0,1113	0,1119	0,1098
38	9120	0,1126	0,1122	0,1093	0,1105	0,1100	0,1100
39	9360	0,1153	0,1084	0,1092	0,1127	0,1096	0,1066
40	9600	0,1133	0,1124	0,1095	0,1102	0,1093	0,1099
41	9840	0,1130	0,1175	0,1127	0,1092	0,1109	0,1102
42	10080	0,1167	0,1157	0,1126	0,1100	0,1114	0,1094
43	10320	0,1175	0,1160	0,1142	0,1106	0,1149	0,1119
44	10560	0,1152	0,1170	0,1137	0,1128	0,1110	0,1104
45	10800	0,1146	0,1158	0,1143	0,1126	0,1101	0,1123
46	11040	0,1138	0,1173	0,1108	0,1110	0,1132	0,1112
47	11280	0,1149	0,1131	0,1151	0,1147	0,1119	0,1129
48	11520	0,1179	0,1157	0,1130	0,1143	0,1156	0,1097
49	11760	0,1175	0,1176	0,1138	0,1135	0,1101	0,1102
50	12000	0,1140	0,1174	0,1113	0,1103	0,1137	0,1129
Moyenne		0,1118	0,1124	0,1092	0,1091	0,1093	0,1091

Le logiciel SPAD (2002) a été employé sous la procédure d'ACP et de classification hiérarchique pour cette analyse.

L'analyse effectuée par le logiciel garantit que la plupart de l'information contenue dans les données est conservée. D'après les résultats obtenus (voir tableau 2.2), il y a autant de composantes (6) que de variables employées pour effectuer l'analyse (6 mesures de rugosité). Nous pouvons résumer les données par les deux premières composantes principales, vu qu'elles expliquent 71,86 % de la variance totale contenue dans les données et que les valeurs propres associées sont les deux seules supérieures à l'unité (tableau 2.2). Deux axes nous permettent ainsi de reconstituer près des $\frac{3}{4}$ de l'inertie contenue dans les données, ce qui est suffisant pour l'analyse et permet de représenter les données.

Table 2.2- Valeurs propres de l'ACP sur les données du tableau 2.1.

Axe	Valeur propre	Pourcentage	Pourcentage cumulé
1	3,3866	56,44	56,44
2	0,9250	15,42	71,86
3	0,5982	9,97	81,83
4	0,5225	8,71	90,54
5	0,3365	5,61	96,15
6	0,2311	3,85	100,00

Grâce à l'ACP, nous avons trouvé deux composantes pour représenter les observations dans un espace à deux dimensions telles que montrées à la figure 2 - 1.

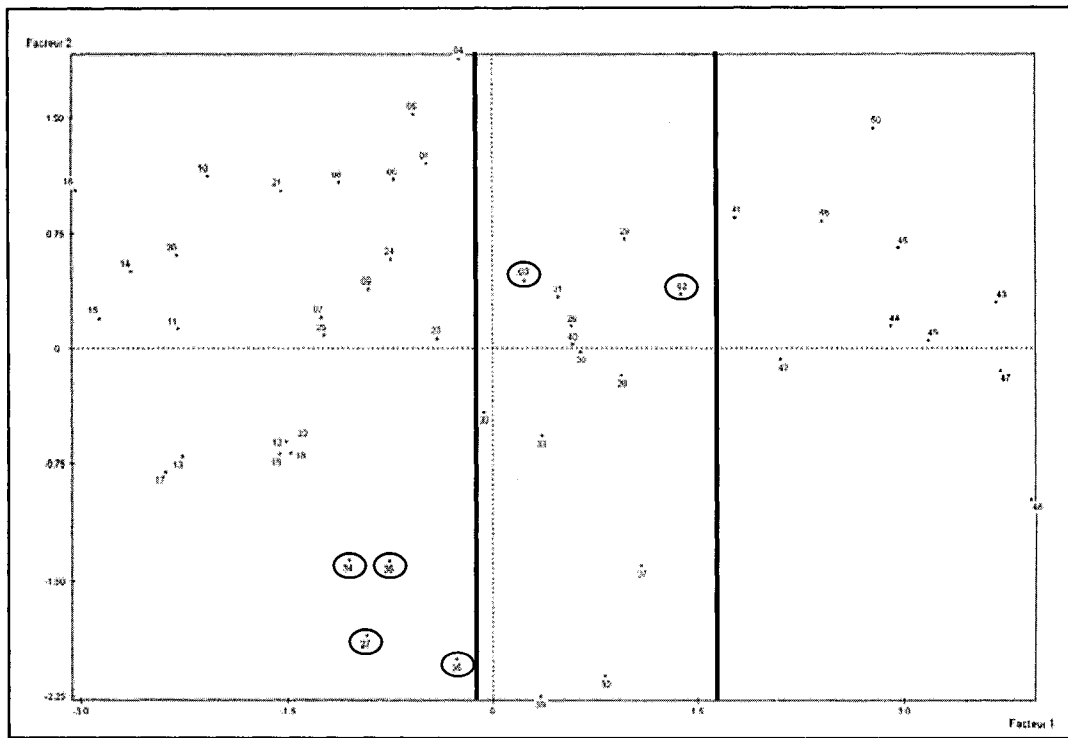


Figure 2-1. Représentation des observations selon l'ACP.

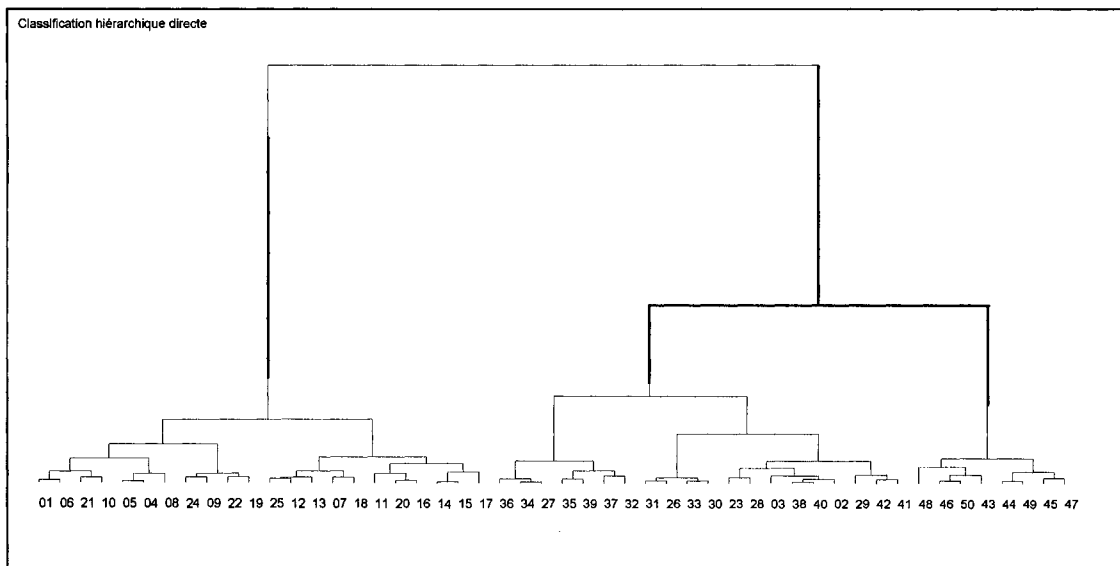


Figure 2-2. Dendrogramme de la classification de Ward sur les deux premières coordonnées factorielles.

La classification hiérarchique ascendante sur les coordonnées résultantes de l'ACP nous a permis d'obtenir un arbre de classification ou dendrogramme qui regroupe les observations par catégories selon le critère d'agrégation de Ward (voir figure 2-2). On peut donc étudier simultanément les groupes formés par la classification de la figure 2-2 et interpréter les axes de la figure 2-1 pour savoir si l'un des groupes correspond à l'état de fonctionnement normal de l'outil.

Procédons à l'interprétation des résultats de la classification : si on fait la coupure de l'arbre de classification de la figure 2-2 en trois groupes, on trouve trois intervalles de temps t , le premier groupe comptant les observations 1-25, le deuxième, celles de 26-40 et le troisième, celles de 41-50.

Si on calcule les moyennes pour les trois intervalles de temps obtenus en faisant la somme des valeurs des observations (tableau 2.1) qui appartiennent à chacun des intervalles, et en divisant par leur nombre (voir les résultats dans le tableau 2.3), on observe que la magnitude de la rugosité augmente par rapport au temps.

Tableau 2.3- Moyenne des valeurs de rugosité selon les intervalles de t
tirés de la classification

Intervalle de t	\bar{O}_1	\bar{O}_2	\bar{O}_3	\bar{O}_4	\bar{O}_5	\bar{O}_6
1 – 25	0,1100	0,1105	0,1073	0,1071	0,1070	0,1082
26 – 40	0,1126	0,1110	0,1098	0,1101	0,1103	0,1083
41 – 50	0,1137	0,1134	0,1109	0,1109	0,1107	0,1098

On peut donc définir trois catégories de pièces : les pièces avec peu de rugosité qui se retrouvent dans l'intervalle initial d'observations, soit les observations 1-25, les pièces

qui exhibent beaucoup de rugosité et qui se trouvent dans l'intervalle de temps final, de 41-50, et le reste des pièces qui sont au milieu. Ce comportement est un indice de l'état de dégradation de l'outil.

Si on regroupe les observations des intervalles dans la figure 2-1 et qu'on fait l'inspection visuelle, on peut interpréter que le premier axe du nuage est un axe d'augmentation générale de l'usure : de gauche à droite, on aura des pièces dont la rugosité est faible, moyenne puis élevée. Alors, on peut penser identifier le premier groupe, comprenant les observations de 1-25, comme l'état de fonctionnement normal de l'outil, suivi du groupe d'observations de 26-40 qui constituent une étape de transition au milieu des données telle que décrite par d'autres auteurs (Huo et Al 1997) et d'un intervalle d'observations, de 41-50, qui correspondent à l'étape usé de l'outil.

On remarque aussi dans la représentation visuelle de la figure 2-1, que les observations 2, 3, 27, 34, 35, et 36 sont mal classées. On attribue ces résultats à l'information manquante dans l'analyse (seulement 71,86 % de la variance des données sont expliqués par les 2 axes ou composantes).

Les résultats indiquent qu'on peut admettre que l'état de fonctionnement normal de l'outil correspond aux observations 1-25. L'annexe B explique que des résultats similaires peuvent être obtenus sous permutation aléatoires des $O_i(t)$ pour chaque T , de sorte que l'on a ce qui peut nous sembler une validation de ces conclusions. Il est noté que cette classification est subjective mais suffisante pour notre objectif de déterminer les observations qui seront utilisées pour construire les cartes de contrôle.

Ceci conclut la première étape de la méthodologie : on a identifié les observations qui serviront à construire les cartes de contrôle (observations 1-25) grâce à l'ACP suivi d'une classification. Maintenant, nous allons nous servir des résultats sur la détermination de l'état sous contrôle statistique pour entreprendre l'étape suivante.

Pour plus de renseignements sur la classification hiérarchique ascendante et les calculs pour l'ACP ainsi que la validation, le lecteur peut consulter les annexes A et B respectivement de ce mémoire.

2.2.- TRAITEMENT DE L'AUTOCORRÉLATION

Cette section explique l'étape 2 de la méthodologie développée dans ce travail. On développe un modèle autorégressif sur les données de l'état de fonctionnement normal de l'outil selon la méthodologie Box-Jenkins. Les résultats serviront pour le calcul des résidus indépendants et normalement distribués, de sorte que les cartes de contrôle seront employées correctement pour la détection de l'usure de l'outil.

La méthodologie Box-Jenkins considère un modèle mathématique général pour représenter un phénomène, l'usure de l'outil dans notre cas. Une observation est décomposée en termes d'une valeur moyenne et des paramètres qui sont liés aux composantes des erreurs présentes et précédentes, ce qui permet d'enlever l'auto-corrélation présente dans une série temporelle des données quelconques.

D'abord, il faut identifier le modèle de la série temporelle. Ensuite, les paramètres sont estimés et le modèle est confirmé. Selon le genre de série temporelle déterminé, le modèle peut être autorégressif, de moyenne mobile ou combiné. On décrira chacune des étapes de la méthodologie Box-Jenkins telle qu'explicitée dans l'ouvrage de Bowerman et O'Connell (1979).

2.2.1.- IDENTIFICATION DU MODÈLE DE LA SÉRIE TEMPORELLE

On a calculé la moyenne $\bar{O}(t)$ à partir des six mesures de la variable $O_i(t)$, pour $t=1, \dots, 25$ (les 25 premières observations du tableau 2.1), afin d'effectuer l'analyse des données de l'état de fonctionnement normal de l'outil de coupage.

$$\bar{O}(t) = \frac{1}{6} \sum_{i=1}^6 O_i(t) \quad t=1, \dots, 25 \quad (2-1)$$

Premièrement, un graphe des observations de la rugosité moyenne à t , $\bar{O}(t)$, versus les observations à $t-1$, $\bar{O}(t-1)$, est employé pour détecter la présence d'une autocorrélation de premier ordre. Une corrélation positive peut être confirmée dans la figure 2-3 qui montre un nuage d'observations autour d'une ligne ascendante.

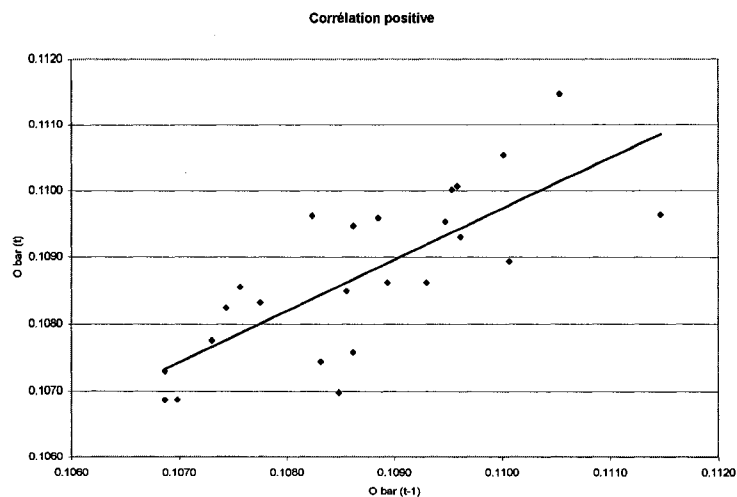


Figure 2-3. Corrélation positive des données : $\bar{O}(t)$ versus $\bar{O}(t-1)$, $t = 1, \dots, 25$.

Ensuite, la série temporelle de l'état de fonctionnement normal de l'outil (figure 2-4) constituée par les observations de rugosité moyenne $\bar{O}(t)$, $t=1, \dots, 25$ est analysée. La série temporelle montrée dans la figure 2-4 exhibe un comportement non stationnaire et corrélé. On observe aussi une période de rodage au début.

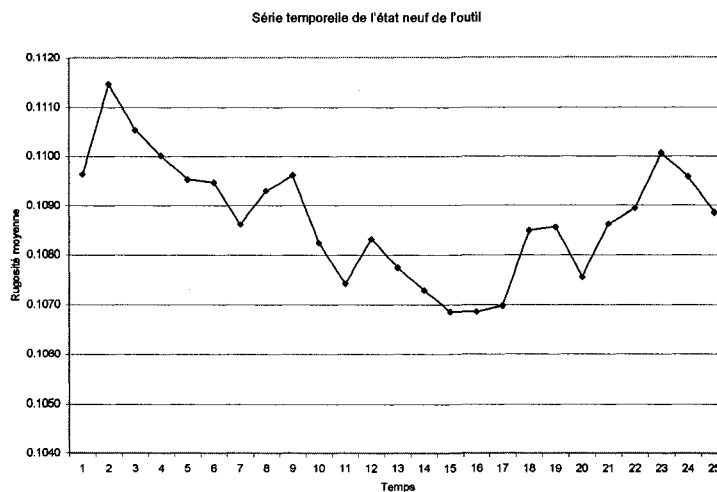


Figure 2-4. Série temporelle de l'état de fonctionnement normal de l'outil.

Observations $\bar{O}(t)$, $t = 1, \dots, 25$.

La série temporelle de la figure 2-4 ne permet pas d'obtenir un modèle selon la méthodologie Box-Jenkins, car elle ne remplit pas les conditions d'inversion requises dans les calculs matriciels effectués pour trouver les coefficients du modèle mathématique. Alors, un processus de *tamissage des données* a été employé afin d'identifier les séries temporelles dont les résultats des modèles obtenus remplissent les conditions : les différentes combinaisons parmi les six mesures de rugosité ont été testées en groupes de deux et trois. Par exemple, la moyenne de O_1 et O_2 ; la moyenne de O_3 , O_4 et O_5 , entre autres. Cette procédure nous a permis d'identifier deux séries : la moyenne des observations appartenant aux mesures O_3 et O_5 , qui constitue la série que l'on appellera $Y_1(t)$, et la moyenne des O_2 et O_4 , qui constitue la série appelée $Y_2(t)$.

Le calcul des observations moyennes est le suivant :

$$Y_1(t) = \frac{O_3(t) + O_5(t)}{2}, t = 1, \dots, 25 \quad (2-2)$$

$$Y_2(t) = \frac{O_2(t) + O_4(t)}{2}, t = 1, \dots, 25 \quad (2-3)$$

Les figures 2-5 et 2-6 montrent les deux séries temporelles convenables pour notre étude : $Y_1(t)$ et $Y_2(t), t = 1, \dots, 25$. L'inspection visuelle indique qu'elles ne sont pas stationnaires car les observations ne se situent pas autour d'une moyenne constante. Comme ceci est un prérequis de la méthodologie Box-Jenkins, il est nécessaire d'appliquer une transformation, et généralement, quand une série ne possède pas des patrons qui se répètent en cycles ou variations saisonnières, la transformation est simplement la différenciation de premier ordre, ce qui revient à utiliser:

$$Z_i(t) = Y_i(t) - Y_i(t-1), i = 1, 2. \quad (2-4)$$

Les figures 2-7 et 2-8 exhibent les nouvelles séries qui sont devenues stationnaires. Le logiciel Statistica (2001) a été employé à cet effet.

Une fois les données stationarisées, on peut estimer le genre de modèle qui génère la série et qui nous permettra d'éliminer la corrélation présente, selon le comportement exhibé par l'auto-corrélation et l'auto-corrélation partielle.

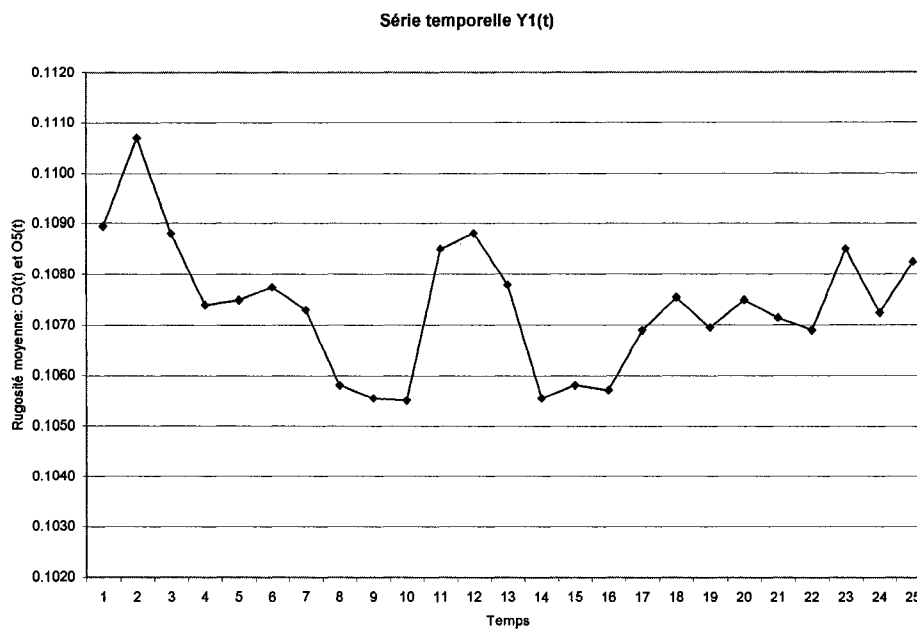


Figure 2-5. Série temporelle non stationnaire $Y_1(t), t = 1, \dots, 25$.

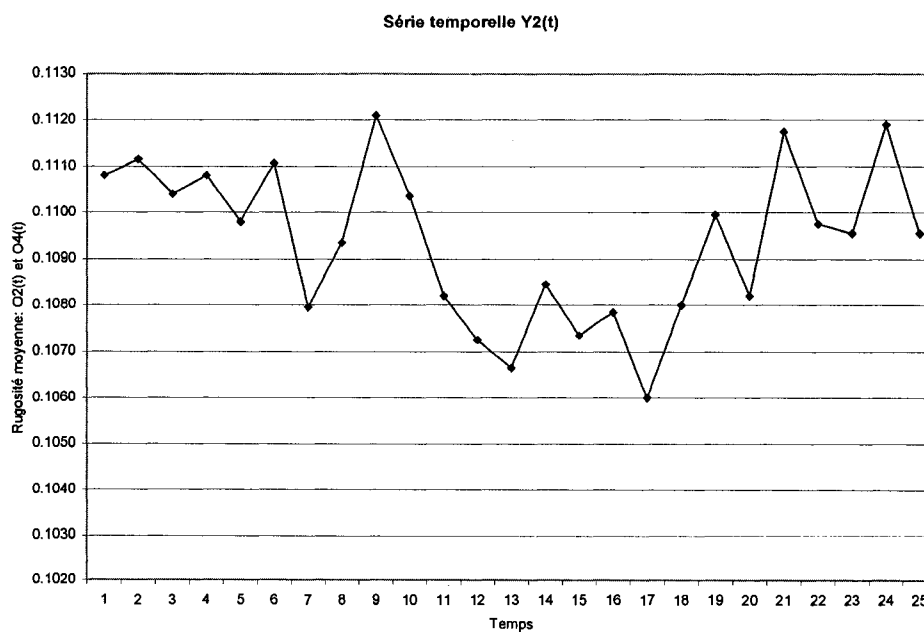


Figure 2-6. Série temporelle non stationnaire $Y_2(t), t = 1, \dots, 25$.

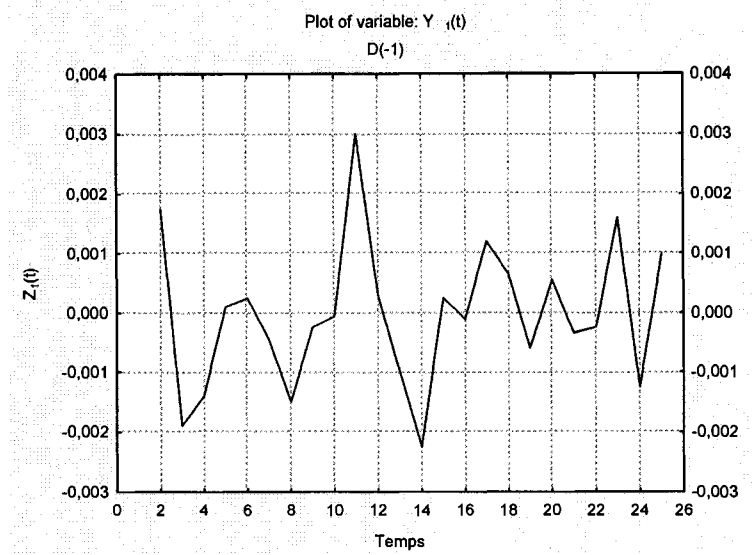


Figure 2-7. Série temporelle stationnaire $Z_1(t), t=2, \dots, 25$.

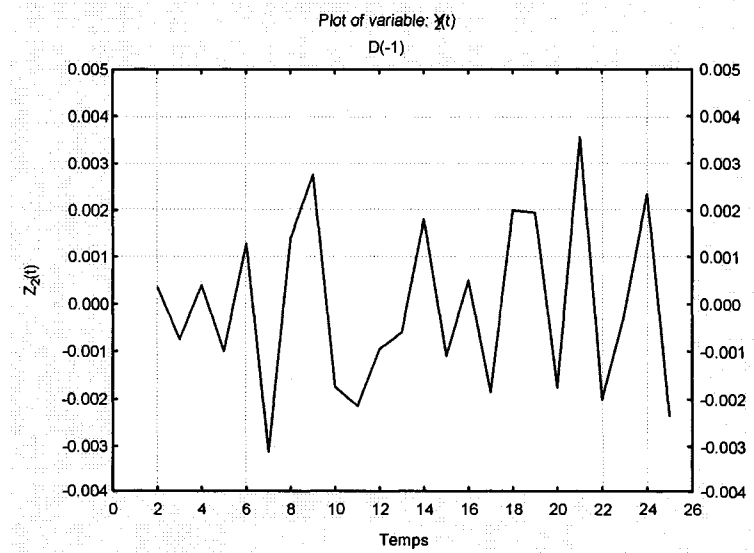


Figure 2-8. Série temporelle stationnaire $Z_2(t), t=2, \dots, 25$.

Les autocorrélations sont représentées par un graphe, le corrélogramme, qui exprime la corrélation ρ_k d'une série avec elle-même en fonction des retards (délais) de $k = 1, 2, \dots$. Normalement le logiciel Statistica calcule de 12 à 15 délais. Elles sont calculées sur les données par la formule suivante :

$$r_k = \frac{\sum_{t=2}^{n-k} [Z(t) - \bar{Z}][Z(t-k) - \bar{Z}]}{\sum_{t=2}^n [Z(t) - \bar{Z}]^2} \quad (2-5)$$

où r_k est l'estimation de ρ_k et \bar{Z} est la moyenne des observations $Z(t), Z(t+1), \dots, t = 2, \dots, 25$. Après un délai spécifique $k = q$, la fonction d'autocorrélation théorique d'une série stationnaire peut être infime ou s'atténuer lentement. Nous pouvons déterminer ce délai où la série devient infime, si elle le devient, par inspection visuelle de l'autocorrélation.

Les autocorrélations partielles sont considérées comme une autre méthode pour examiner les dépendances périodiques. Comme c'est une extension de l'autocorrélation, elle tient compte de l'ajustement du modèle utilisant les délais d'ordre inférieur. Si un retard de 1 est indiqué, alors l'auto-corrélation partielle est équivalent à l'auto-corrélation ($r_{kk} = r_1$, Si $k=1$). Pour un retard de 2 ou plus,

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \quad \text{pour } j=1, 2, \dots, k-1 \quad (2-6)$$

$$\text{Et, } r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j},$$

La fonction d'autocorrélation partielle d'une série stationnaire peut être infime ou s'atténuer lentement après un délai spécifique $k = q$ dépendamment de la nature des données. Nous pouvons déterminer ce délai où la série devient non significative par simple inspection visuelle du diagramme des autocorrélations partielles. Les intervalles de confiance illustrés servent de référence à ce propos.

Par la suite, on illustre les fonctions d'autocorrélation pour les différentes séries qui furent calculées à l'aide du logiciel Statistica (2001). On remarque que Statistica a calculé jusqu'à un délai 15.

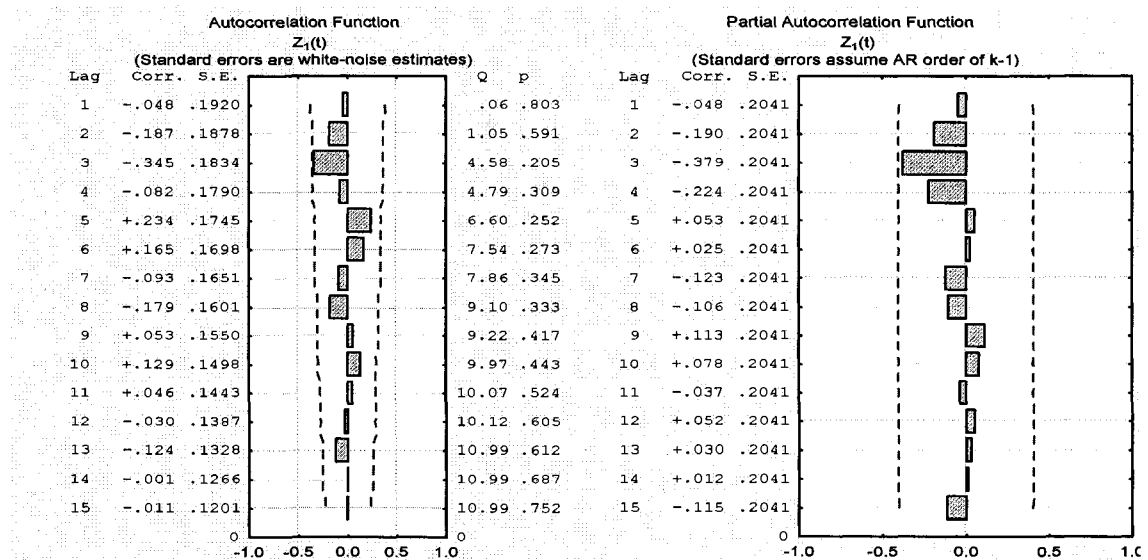


Figure 2-9. Auto-corrélations de la série $Z_1(t), t=2, \dots, 25$.

La figure 2-9 indique que la fonction d'auto-corrélation diminue lentement et que l'auto-corrélation partielle devient non significative après le délai d'ordre 4. Selon le tableau 2.4, on peut conclure que le modèle qui génère la série $Z_1(t)$ est un modèle autorégressif d'ordre 4, AR(4) (voir le tableau 2.4) qui sera estimé par Statistica. De la même manière,

si on analyse les fonctions de la figure 2-10, on trouve le même comportement alors le modèle qui génère la série $Z_2(t)$ est aussi un AR(4), qui sera estimé lui aussi par Statistica.

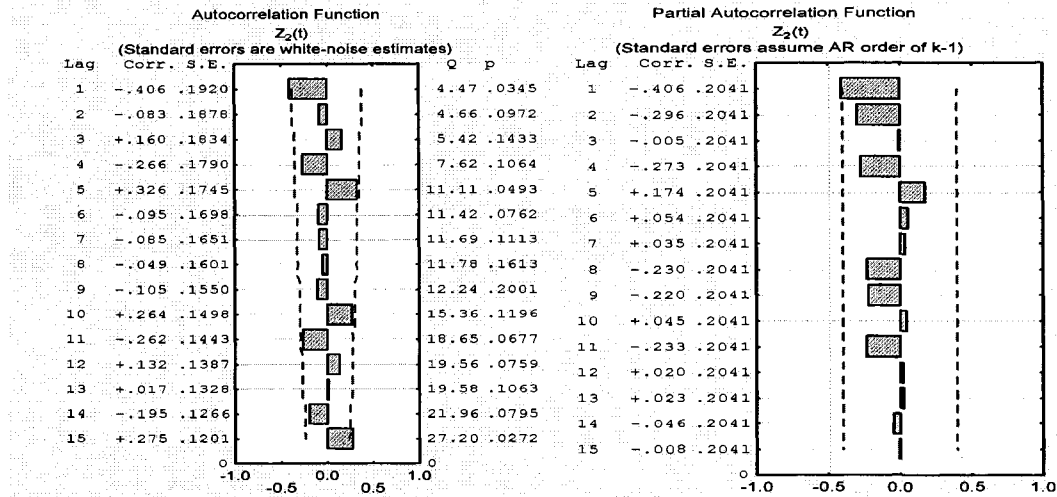


Figure 2-10. Auto-corrélations de la série $Z_2(t), t=2, \dots, 25$.

Tableau 2.4- Choix du modèle selon le type de corrélation.

Modèle	Auto-corrélation théorique	Auto-corrélation partielle
Moyenne mobile d'ordre q	Arrête après un délai q	Diminue lentement
Autorégressive d'ordre q	Diminue lentement	Arrête après un délai q
Combiné	Diminue lentement	Diminue lentement

2.2.2.- ESTIMATION DES MODÈLES AUTORÉGRESSIFS

Un modèle autorégressif AR d'ordre p exprime la valeur d'une observation en fonction des p observations précédentes affectées par des valeurs φ , une constante δ et une composante d'erreur ε_t .

$$\hat{Z}(t) = \delta + \varphi_1 Z(t-1) + \varphi_2 Z(t-2) + \dots + \varphi_p Z(t-p) + \varepsilon_t \quad (2-7)$$

D'après les fonctions d'autocorrélation, les modèles qui génèrent les séries $Z_1(t)$ et $Z_2(t)$ sont autorégressifs d'ordre 4. Alors, une observation est décrite en fonction des quatre observations précédentes.

En utilisant les équations de Yule-Walker, à l'aide du logiciel Statistica (2001), nous avons obtenu les paramètres des modèles suivants :

- Pour la série $Z_1(t), t=2, \dots, 25$:

$$\hat{Z}_1(t) = -0.2308 Z_1(t-1) - 0.2386 Z_1(t-2) - 0.4372 Z_1(t-3) - 0.2244 Z_1(t-4) \quad (2-8)$$

- Pour la série $Z_2(t), t=2, \dots, 25$:

$$\hat{Z}_2(t) = -0.5503 Z_2(t-1) - 0.3769 Z_2(t-2) - 0.1445 Z_2(t-3) - 0.3346 Z_2(t-4) \quad (2-9)$$

Il faut remarquer que dans les modèles obtenus, la constante δ et la composante d'erreur ε_t sont égales à zéro. Afin d'assurer que les modèles trouvés sont conformes aux exigences de la méthodologie Box-Jenkins, nous allons vérifier la diminution de l'autocorrélation et la distribution des résultats dans la section suivante.

2.2.3.- CONFIRMATION DES MODÈLES ET OBTENTION DES RÉSIDUS

Les résultats des modèles sont obtenus par Statistica à partir des équations d'estimation et des méthodes d'estimation non linéaire, et ils sont montrés dans le tableau 2.5. Le logiciel donne les valeurs $\hat{Y}_i(t), t = 3, \dots, 25; i = 1, 2$; on passe des valeurs $\hat{Z}_i(t)$ à $\hat{Y}_i(t)$ par intégration, car les $\hat{Z}_i(t)$ sont la différentiation des $\hat{Y}_i(t)$ (Voir l'équation 2-4). On obtient des résultats à partir de l'observation 3 parce que le modèle AR est d'ordre 4.

Tableau 2.5- Résultats des modèles autorégressifs données par Statistica.

t	\hat{Y}_1	t	\hat{Y}_1	t	\hat{Y}_2	t	\hat{Y}_2
3	0.109	15	0.105	3	0.111	15	0.103
4	0.108	16	0.104	4	0.111	16	0.103
5	0.108	17	0.104	5	0.110	17	0.101
6	0.107	18	0.104	6	0.111	18	0.103
7	0.106	19	0.104	7	0.108	19	0.105
8	0.104	20	0.105	8	0.108	20	0.105
9	0.104	21	0.105	9	0.110	21	0.108
10	0.103	22	0.105	10	0.110	22	0.108
11	0.105	23	0.107	11	0.107	23	0.109
12	0.106	24	0.106	12	0.105	24	0.110
13	0.105	25	0.106	13	0.104	25	0.110
14	0.104			14	0.103		

Statistica fournit aussi les moyens pour confirmer si les modèles obtenus sont convenables d'après la méthodologie Box-Jenkins : les figures 2-11 et 2-12 montrent la

diminution des fonctions d'auto-corrélation pour les résultats des deux modèles. D'autre part, les figures 2-13 et 2-14 présentent des diagrammes quantile-quantile sur la loi normale ou droite de Henry, qui permettent de vérifier la normalité de la distribution des résultats.

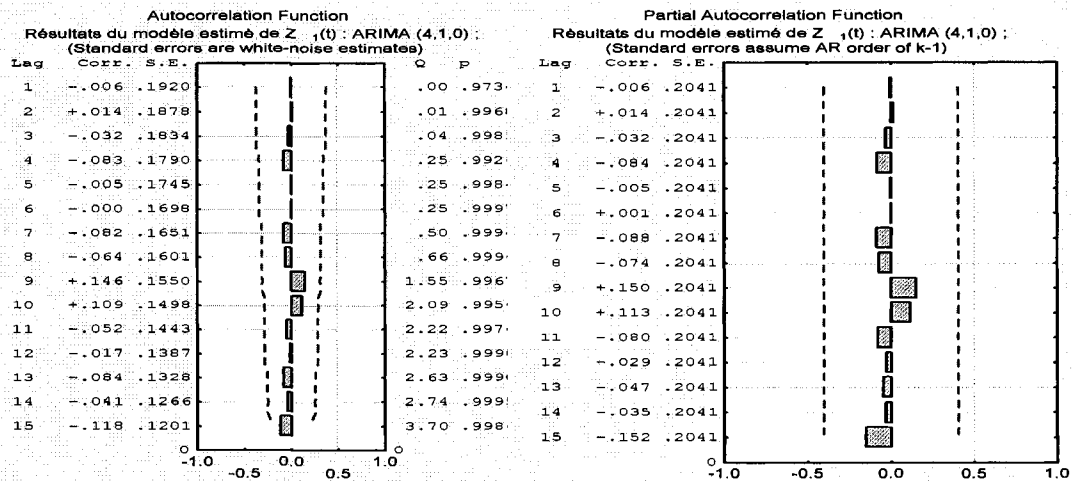


Figure 2-11. Auto-corrélation du modèle AR (4) pour la série $Z_1(t)$.

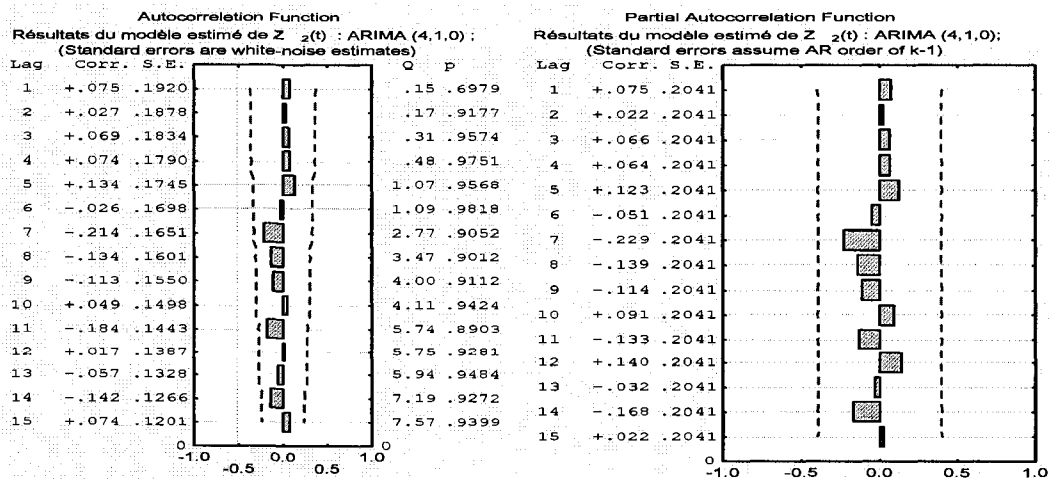


Figure 2-12. Auto-corrélation du modèle AR (4) pour la série $Z_2(t)$.

Il faut remarquer que les observations dans la figure 2-14 ne sont pas tout à fait groupées autour de la ligne droite du diagramme ; un test de Shapiro-Wilk a donné comme résultat que la statistique w , égale à 0.96813, donc se trouve entre 0.7 et 1. On admettra donc que les résultats sont *assez* gaussiens pour poursuivre le développement de la méthodologie de cette recherche.

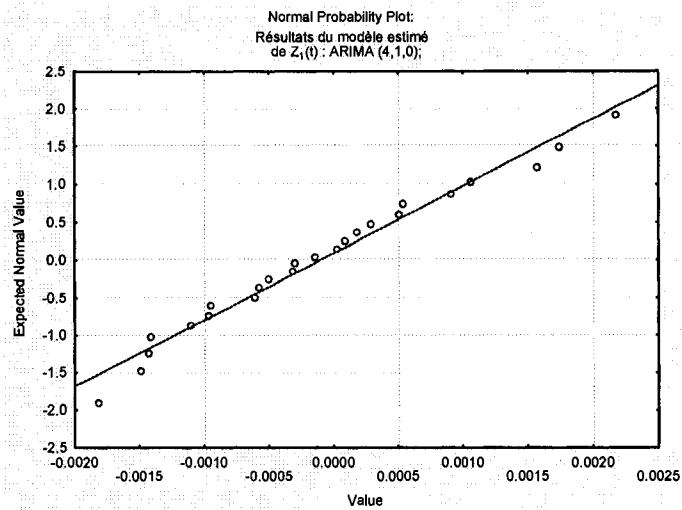


Figure 2-13. Distribution des résultats du modèle AR (4) pour la série $Z_1(t)$.

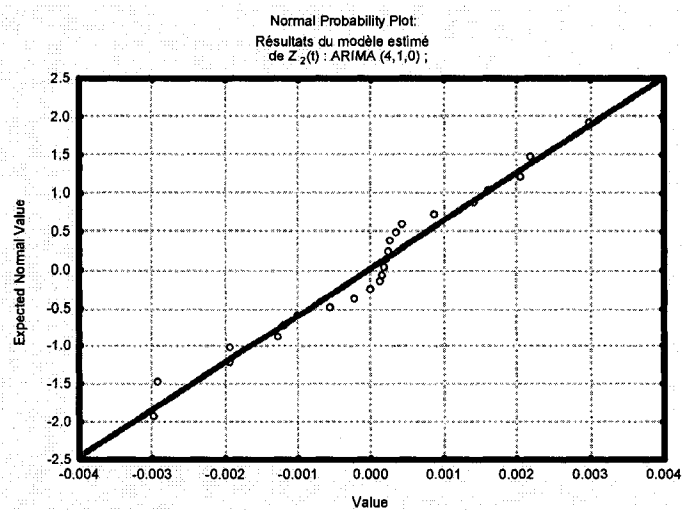


Figure 2-14. Distribution des résultats du modèle AR (4) pour la série $Z_2(t)$.

Maintenant que l'on a confirmé les modèles, nous allons utiliser les valeurs des observations originelles $Y_i(t), t = 3, \dots, 25; i = 1, 2$ qui ont été calculées avec les équations 2-2 et 2-3 pour obtenir les résidus $\varepsilon_i(t) = Y_i(t) - \hat{Y}_i(t), t = 3, \dots, 25; i = 1, 2$. Le tableau 2.6 montre les résultats. Ce sont ces résidus qui serviront pour la construction des cartes de contrôle.

Tableau 2.6- Calcul de résidus des modèles sur les observations :

$$\varepsilon_i(t) = Y_i(t) - \hat{Y}_i(t), t = 3, \dots, 25; i = 1, 2$$

t	Y_1	Y_2	$\varepsilon_1 = Y_1 - \hat{Y}_1$	$\varepsilon_2 = Y_2 - \hat{Y}_2$
3	0.1088	0.1104	-0.0004	-0.0002
4	0.1074	0.1108	-0.0004	0.0001
5	0.1075	0.1098	-0.0004	0.0001
6	0.1078	0.1111	0.0004	0.0005
7	0.1073	0.1080	0.0013	0.0004
8	0.1058	0.1094	0.0016	0.0016
9	0.1056	0.1121	0.0020	0.0022
10	0.1055	0.1104	0.0025	0.0002
11	0.1085	0.1082	0.0034	0.0009
12	0.1088	0.1073	0.0031	0.0019
13	0.1078	0.1067	0.0024	0.0026
14	0.1056	0.1085	0.0013	0.0041
15	0.1058	0.1074	0.0012	0.0042
16	0.1057	0.1079	0.0021	0.0046
17	0.1069	0.1060	0.0032	0.0047
18	0.1076	0.1080	0.0034	0.0050
19	0.1070	0.1100	0.0029	0.0049
20	0.1075	0.1082	0.0024	0.0032
21	0.1072	0.1118	0.0019	0.0038
22	0.1069	0.1098	0.0019	0.0015
23	0.1085	0.1096	0.0020	0.0009
24	0.1073	0.1119	0.0017	0.0018
25	0.1083	0.1096	0.0018	-0.0003

2.2.4.- OBTENTION DES RÉSIDUS DES PRÉVISIONS

Les prévisions pour de nouvelles observations $\hat{Y}_i(t), t = 26, \dots, 50, i = 1, 2$ sont calculées par Statistica en utilisant les paramètres du modèle AR(4) trouvé. Les figures 2-15 et 2-16 montrent les prévisions avec leurs intervalles de confiance.

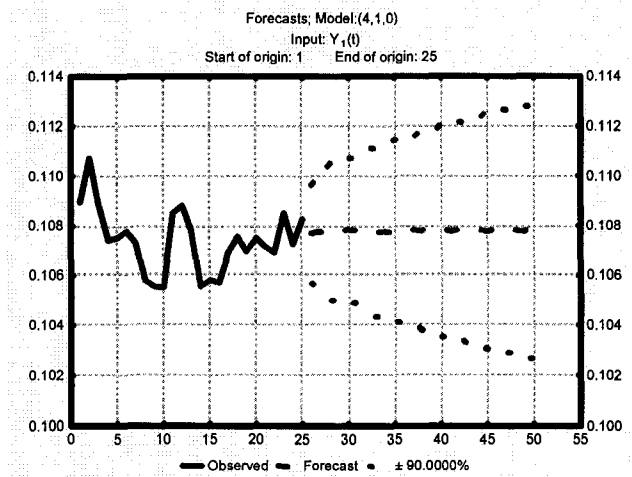


Figure 2-15. Prévisions des observations $\hat{Y}_1(t)$.

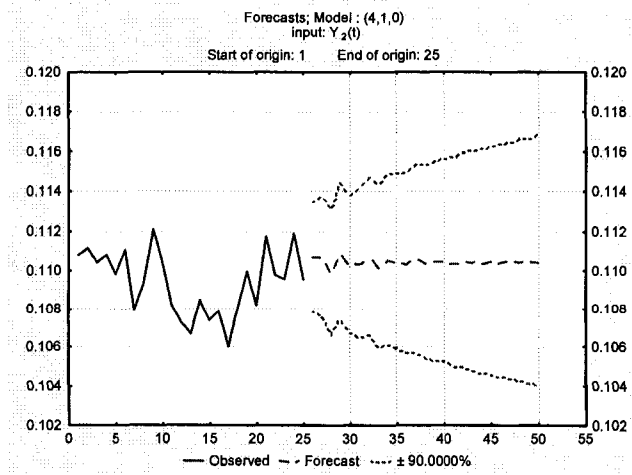


Figure 2-16. Prévisions des observations pour $\hat{Y}_2(t)$.

Pour calculer les résidus des prévisions $\varepsilon_i(t), t = 26, \dots, 50; i = 1, 2$ nous devons soustraire les valeurs estimées $\hat{Y}_i(t)$ aux observations originelles $Y_i(t)$. Le tableau 2.7 présente les résidus obtenus pour les prévisions données par le logiciel Statistica. Ces résidus seront tracés sur les cartes de contrôle pour la détermination de l'usure.

Tableau 2.7- Calcul des résidus pour les prévisions

$$\hat{Y}_i(t), t = 26, \dots, 50; i = 1, 2$$

t	Y_1	Y_2	\hat{Y}_1	\hat{Y}_2	$\varepsilon_1 = Y_1 - \hat{Y}_1$	$\varepsilon_2 = Y_2 - \hat{Y}_2$
26	0.1104	0.1133	0.1077	0.1107	0.0027	0.0026
27	0.1093	0.1081	0.1078	0.1107	0.0015	-0.0026
28	0.1109	0.1099	0.1077	0.1098	0.0031	0.0001
29	0.1089	0.1130	0.1077	0.1109	0.0012	0.0021
30	0.1102	0.1128	0.1078	0.1102	0.0024	0.0026
31	0.1105	0.1127	0.1078	0.1103	0.0026	0.0024
32	0.1118	0.1101	0.1078	0.1107	0.0040	-0.0006
33	0.1103	0.1121	0.1078	0.1102	0.0025	0.0019
34	0.1093	0.1083	0.1078	0.1105	0.0015	-0.0023
35	0.1094	0.1090	0.1078	0.1104	0.0016	-0.0014
36	0.1096	0.1097	0.1078	0.1103	0.0018	-0.0007
37	0.1109	0.1094	0.1078	0.1105	0.0031	-0.0011
38	0.1097	0.1114	0.1078	0.1103	0.0019	0.0010
39	0.1094	0.1106	0.1078	0.1104	0.0016	0.0001
40	0.1094	0.1113	0.1078	0.1104	0.0016	0.0009
41	0.1118	0.1134	0.1078	0.1104	0.0040	0.0030
42	0.1120	0.1129	0.1078	0.1104	0.0042	0.0024
43	0.1146	0.1133	0.1078	0.1104	0.0068	0.0029
44	0.1124	0.1149	0.1078	0.1104	0.0046	0.0045
45	0.1122	0.1142	0.1078	0.1104	0.0044	0.0038
46	0.1120	0.1142	0.1078	0.1104	0.0042	0.0038
47	0.1135	0.1139	0.1078	0.1104	0.0057	0.0035
48	0.1143	0.1150	0.1078	0.1104	0.0065	0.0046
49	0.1120	0.1156	0.1078	0.1104	0.0042	0.0052
50	0.1125	0.1139	0.1078	0.1104	0.0047	0.0034

2.3.- CONSTRUCTION DES CARTES DE CONTRÔLE

La construction des cartes $X\text{-bar}$ et R est faite conjointement en utilisant les résidus indépendants et normalement distribués, tels que calculés plus haut lorsque l'outil est neuf (données du tableau 2.6). Le test de Shapiro-Wilk indique une distribution normale pour les résidus du tableau 2.6, puisque la valeur de la statistique w , égale à 0.96245, se trouve entre 0.7 et 1.

2.3.1.- CARTE X-BAR

Afin de construire la carte $X\text{-bar}$, la moyenne de la population des résidus $\overline{\varepsilon(t)}$, $t=3, \dots, 25$, et son écart type seront estimés par $\bar{\varepsilon} = \frac{\sum_{t=3}^{m=25} \varepsilon(t)}{m-3+1}$ et $S = \bar{R} / d_2 \sqrt{n}$ respectivement. Où $\bar{R} = \frac{\sum_{t=3}^{m=25} |\varepsilon_1(t) - \varepsilon_2(t)|}{m-3+1}$ est la moyenne des étendues et $d_2 = 1.128$ est une constante pour la taille de l'échantillon $n=2$.

Dans la carte $X\text{-bar}$ exposée dans la figure 2-17, la ligne centrale est égale à la moyenne $\bar{\varepsilon} = 0.00196$, et les limites de contrôle sont $\bar{\varepsilon} \pm 3S$, à savoir : $UCL = 0.00428$ et $LCL = -0.36E-3$. Il n'y a pas d'observations en dehors des limites de contrôle. Toutefois, on observe qu'il y a certaines règles de 'Western Electric' qui ne sont pas respectées (Voir tableau 2.8), un comportement que l'on attribue à d'autres causes que l'usure. On se permet d'éliminer les observations numéro 3-5 et 15-19 pour enlever ce comportement. (Montgomery, 2005).

Les limites de contrôle seront recalculées avec les 15 observations restantes. Une nouvelle carte est obtenue après ce tamisage de données, la figure 2-18 montre la

nouvelle carte où les observations sont sous contrôle statistique, les nouvelles limites sont : $UCL=0.0031$, $CL=0.00183$ et $LCL=-0.25E-3$.

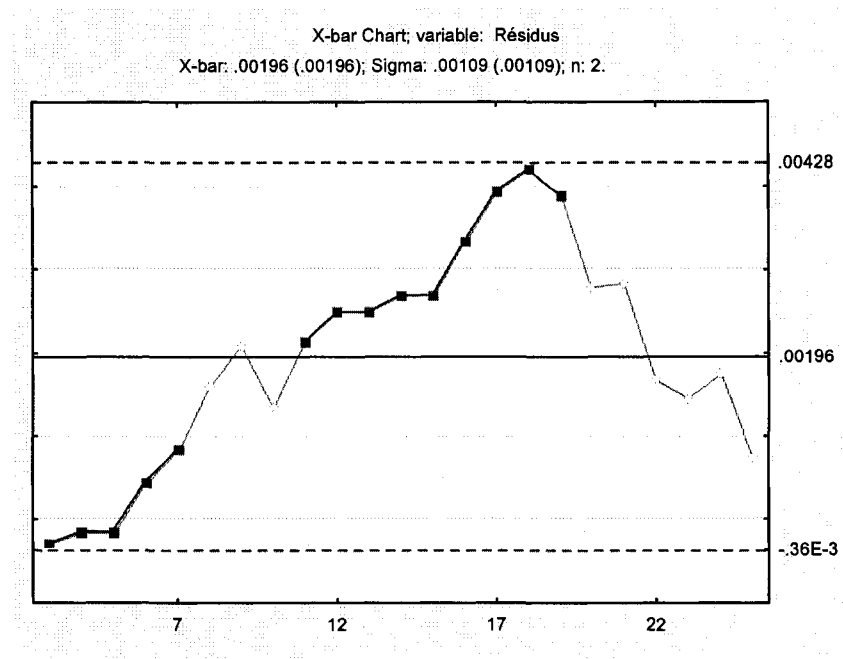


Figure 2-17. Carte X -bar.

Tableau 2.8- Résumé des règles de Western Electric concernant la carte X -bar de la figure 2-17.

Règles de 'Western Electric'	Intervalle de résidus affectée
9 observations consécutives dans la zone C ou au-delà (positionnées du même côté de la ligne centrale).	11-19
2 de 3 observations dans la zone A ou au-delà	3-5 ; 16-18
4 de 5 observations dans la zone B ou au-delà	3-7 ; 15-19

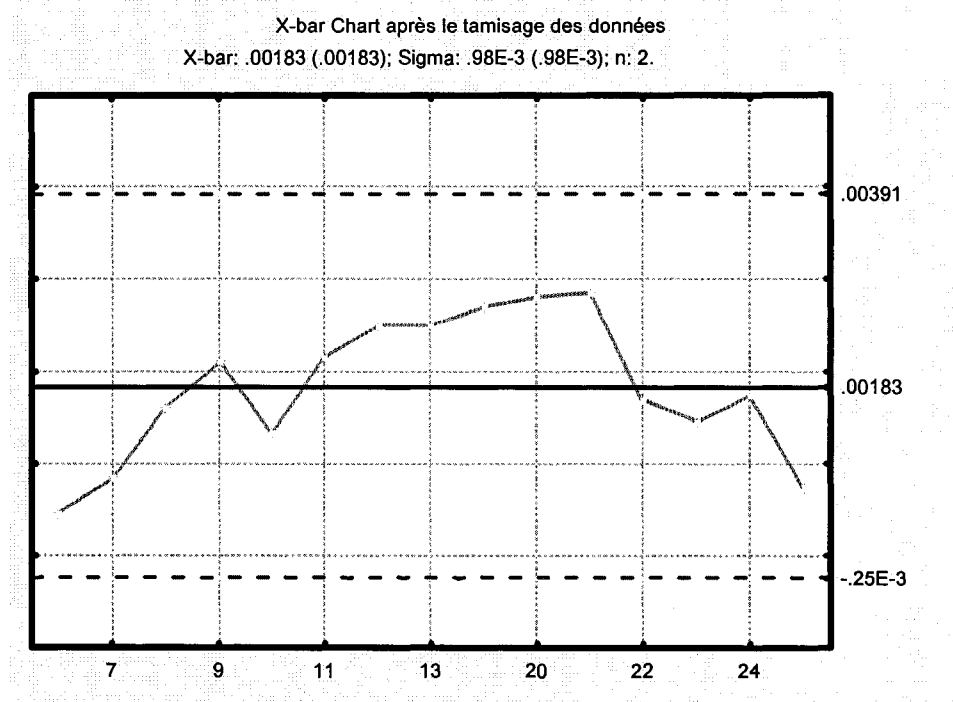


Figure 2-18. Carte $X\text{-bar}$ après le tamisage des données.

Nous allons employer la carte montrée de la figure 2-18 pour effectuer la détection de l'usure.

2.3.2.- CARTE ÉTENDUE R

On emploie les données du tableau 2.6 pour la construction de cette carte en tenant compte du tamisage effectué dans la carte $X\text{-bar}$, alors que les observations numéro 3-5 et 15-19 sont éliminées avant de construire cette carte. La moyenne de la population des résidus $\overline{\varepsilon(t)}$, $t=3, \dots, 25$, est estimée par $\bar{R} = \sum_{t=3}^{m=25} |\varepsilon_1(t) - \varepsilon_2(t)| / (m - 3 + 1) = 0.00111$. Les

constantes $D_3=3.267$ et $D_4=0$ sont employées car la taille de l'échantillon $n= 2$. Les limites de contrôle obtenues sont $UCL = D_4\bar{R} = 0.00361$, $CL = \bar{R} = 0.00111$ et $LCL = D_3\bar{R} = 0$.

La carte obtenue est présentée à la figure 2-19, il n'y a pas de points en dehors des limites de contrôle ou d'autres patrons. Alors, on peut utiliser cette carte pour la détection de l'usure.

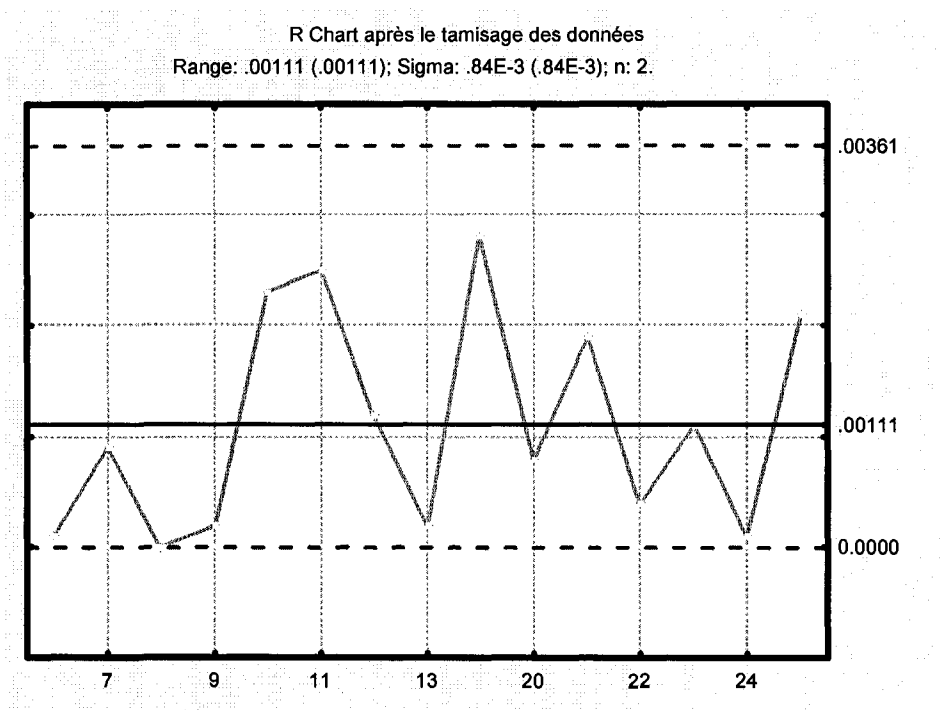


Figure 2-19. Carte R après le tamisage des données.

2.4.- DÉTECTION DE L'USURE

Nous allons tracer sur les cartes de contrôle déjà établies (Figures 2-18 et 2-19) les résidus des valeurs des prévisions trouvées à partir du modèle autorégressif (Tableau 2.7) pour les données suivantes. Cela permettra d'identifier le moment où l'usure est manifeste lorsque les observations dépassent les limites de contrôle. Les figures 2-20 et 2-21 montrent les résultats.

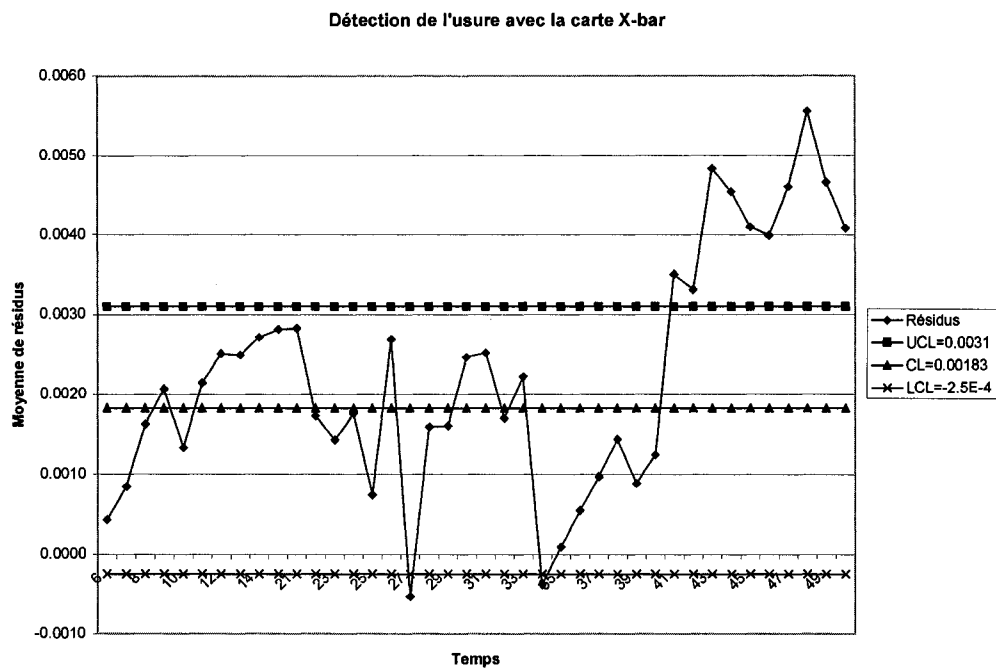


Figure 2-20. Détection de l'usure avec la carte *X-bar* construite sur les résidus du modèle de Box-Jenkins AR(4).

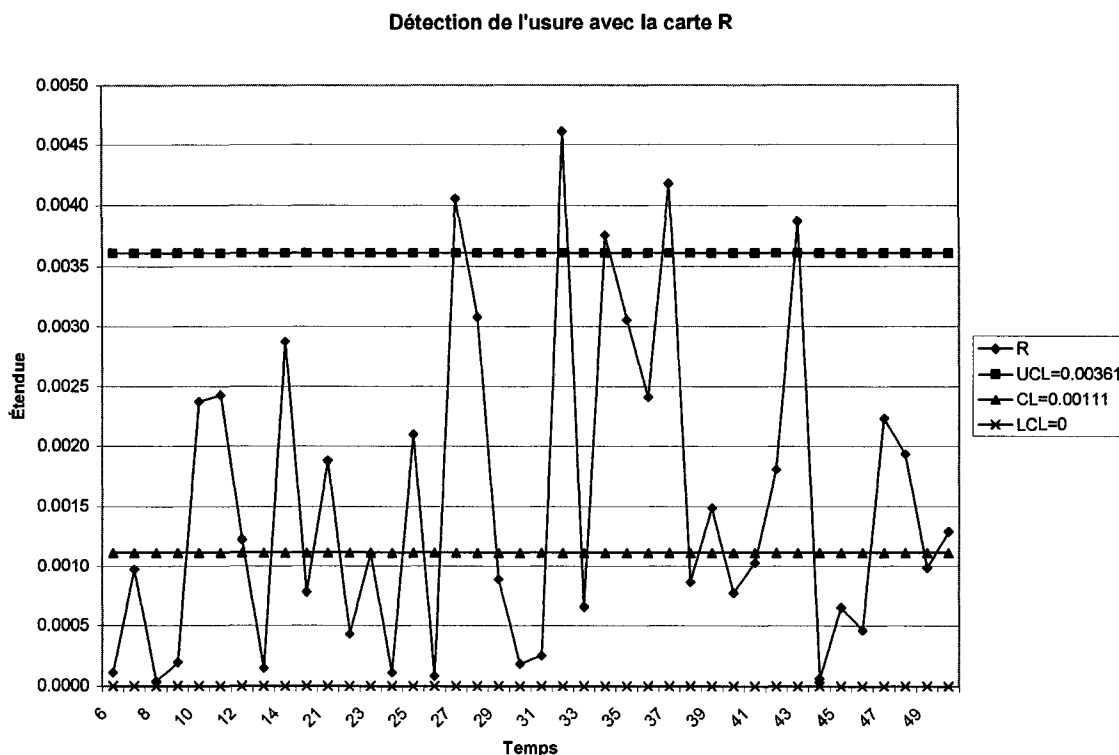


Figure 2-21. Détection de l'usure avec la carte R bar construite sur les résidus du modèle de Box-Jenkins $AR(4)$.

On peut distinguer trois zones dans les cartes :

- Zone de jeunesse : contient les premières observations dans les deux cartes. Celles-ci sont sous contrôle statistique lorsque le tamisage des données a été effectué (observations 3 à 5 et 15 à 19), ce sont elles qui ont été employées pour établir les limites de contrôle.
- Zone de transition : ce sont les observations 26-34. Des problèmes commencent à faire surface dans les cartes car l'observation numéros 35 est hors de contrôle dans la carte $X\text{-bar}$; et les numéros 28 et 33, dans la carte R . On considère ces

observations comme des signaux d'alerte puisque l'usure est manifeste plus clairement dans la troisième zone. La carte R permet donc l'identification de la zone de transition lorsqu'il y a une variation remarquable des étendues.

- Zone de vieillesse ou d'usure : ce sont les observations 35-50. L'usure est manifeste quasi entièrement par la carte $X\text{-bar}$ car à partir de l'observation 35 une tendance apparaît, les résidus augmentent sa valeur jusqu'à ce que les observations deviennent hors de contrôle plus continuellement : observations 42-50. La carte R montre les observations 35, 38 et 44 hors de contrôle ainsi que des variations significatives des étendues.

Le début de l'usure est donc détecté par la carte $X\text{-bar}$ à partir de l'observation 35. Le temps qui correspond selon le tableau 2.1 est $T=8400$ secondes, alors 140 minutes est le temps prescrit de remplacement de l'outil selon la carte unidimensionnelle construite ici.

On pourrait aussi considérer remplacer l'outil lorsqu'il est complètement usé, bien que cela entraînerait des coûts de non qualité du produit. La carte $X\text{-bar}$ permet d'identifier ce moment lorsque l'observation 44 se trouve au-dessus de la limite de contrôle supérieure ; le temps T qui correspond est de 10560 seconds selon le tableau 2.1, c'est-à-dire, qu'on remplacerait l'outil de coupage toutes les 176 minutes d'utilisation avec un outil nouveau.

Vu que l'historique des données fixait la durée de vie d'un outil tournant à 13000 RPM à 200 minutes, il est raisonnable de remplacer plus tôt cet outil lorsqu'il tourne plus rapidement (18000 RPM). Ces résultats confirment l'utilité de la méthodologie que l'on a mise en place pour la détection de l'usure de l'outil avec des cartes de contrôle unidimensionnelles.

La remarque suivante est importante: bien que certaines décisions puissent sembler arbitraires, il faut tenir compte que notre méthodologie est une démarche qu'on vise à expliciter avec des données réelles et que toutes les données réelles ont leur particularité. Pourtant, tels qu'exposés, avec les choix qui s'imposaient, les résultats sont cohérents avec les attentes issues de l'instrument précédent.

Maintenant, il reste à confirmer l'approche dans le cas multidimensionnel, le chapitre III présente la démarche.

CHAPITRE 3

CAS MULTIDIMENSIONNEL

Dans le chapitre précédent, nous avons expliqué les détails concernant le développement d'une méthodologie pour prédire la défaillance d'un outil de coupage qui tourne à 18000 RPM en utilisant les cartes de contrôle unidimensionnelles *X-bar* et *R* construites sur les résidus d'un modèle de *Box-Jenkins*. On a alors déterminé que le temps de remplacement le plus approprié pour l'outil est au début de l'usure, soit à 140 minutes d'utilisation.

Ce chapitre explique le développement des étapes de la méthodologie lorsqu'on dispose de *plusieurs variables*, ou indicateurs de l'usure, pour effectuer l'analyse. On utilisera la carte de contrôle multidimensionnelle de Hotelling à cet effet. Nous présentons d'abord une introduction aux cartes de contrôle multidimensionnelles. Ensuite, on exposera les étapes concernant la méthodologie. Finalement, les résultats sur la détection du phénomène d'usure seront interprétés afin de trouver le temps auquel l'outil de coupage devrait être remplacé.

3.1.- CARTES DE CONTRÔLE MULTIDIMENSIONNELLES

Il existe des situations où la surveillance statistique d'un procédé nécessite l'étude simultanée de plusieurs caractéristiques, cela est connu comme contrôle statistique multidimensionnel des procédés.

Les cartes de contrôle multidimensionnelles sont utilisées pour réduire la variabilité d'un procédé en surveillant plusieurs caractéristiques. Le pouvoir de leur utilisation vient de la différenciation qu'on peut observer grâce à elles entre la variation due à des causes spéciales et la variation aléatoire ou naturelle du processus, ce qui nous permet de distinguer des situations où le processus devient hors du contrôle statistique.

L'importance des cartes de contrôle multidimensionnelles réside dans le fait que lorsqu'on fait face à plusieurs variables, la seule utilisation des cartes unidimensionnelles construites sur chaque série de mesures ne fournit pas une base solide pour affirmer que *le procédé est sous contrôle statistique par rapport à une cible collective à atteindre* parce que l'erreur de type I et la probabilité de trouver une observation sous contrôle statistique ne correspondent pas aux niveaux d'alarme des cartes de contrôle individuelles, et que les limites de contrôle ne peuvent pas être ajustées à cause des corrélations présentes parmi les variables. Cette distorsion augmente conformément à l'augmentation du nombre des variables à surveiller. De plus, quand le nombre de variables est grand, la tâche de maintenir des cartes individuelles devient difficile à cause de la grande quantité de cartes à analyser.

Parmi les cartes multidimensionnelles les plus utilisées, on trouve la carte de Hotelling, la carte MCUSUM (Multivariate Cumulative Sum Control Chart en anglais) et la carte MEWMA (Multivariate Exponentially Weighted Moving Average Control Chart en anglais). Nous avons choisi pour nos besoins la carte de Hotelling pour effectuer la détection de l'usure de l'outil dans le cas multidimensionnel parce que la plupart des recherches en CSMP sont centrées sur cette carte, et parce qu'elle détecte très bien les changements de moyenne (Chan et Li, 1994), forme sous laquelle se manifeste le phénomène d'usure.

Brecka (1995) présente une liste de logiciels disponibles pour la construction des cartes multidimensionnelles. Dans ce travail, nous allons construire la carte de Hotelling en utilisant le logiciel Matlab ainsi que des feuilles de calcul de Microsoft Excel.

3.2.- PRÉPARATION DES DONNÉES

Afin de compléter notre exposé méthodologique, nous allons employer deux variables, indicateurs de l'usure, deux types de rugosité de la surface usinée qui constituent la clé du

processus de coupage et qui nous permettent de mener une étude multidimensionnelle en utilisant la carte de contrôle de Hotelling : la première variable, $O_i(t)$, fut utilisée dans le chapitre II pour le cas unidimensionnel où on a obtenu des résidus indépendants de modèles de Box-Jenkins pour la construction des cartes de contrôle (tableau 2.6) et pour la détection de l'usure (tableau 2.7); la deuxième variable, $W_i(t), i = 1, \dots, 3; t = 1, \dots, 50$, est un autre type de rugosité. Dans cette section du travail, on présente l'application des deux premières étapes de la méthodologie pour cette variable.

Le tableau 3.1 montre les 50 observations ou échantillons de la variable $W_i(t)$, $i = 1, \dots, 3; t = 1, \dots, 50$ qui ont été recueillis à toutes les 240 secondes, sur les mêmes échantillons que la variable précédente ; une observation est composée par trois mesures de rugosité de la surface usinée. (Caron, 2002.)

L'étape 1 de la méthodologie consiste en l'identification du nombre d'observations qui servent pour construire les cartes de contrôle : afin de comparer les résultats du cas unidimensionnel et multidimensionnel, nous allons employer le même intervalle d'observations dans les deux cas, ce sont les observations comprises entre 1 et 25, d'après les résultats obtenus dans le chapitre II.

L'étape 2 de la méthodologie comprend l'estimation d'un modèle mathématique de Box-Jenkins pour l'obtention des résidus non corrélés. Nous allons travailler avec la nouvelle variable $W_i(t)$ de la même façon que dans le cas de la variable $O_i(t)$. Premièrement, on analyse la série temporelle pour constater si les données sont en état stationnaire.

La série temporelle qui expose la moyenne des données de $W_i(t)$ (figure 3-1) et que l'on appellera $W(t)$, où $W(t) = \frac{1}{3} \sum_{i=1}^3 W_i(t), t = 1, \dots, 25$, exhibe un comportement non stationnaire parce qu'on observe que les données ne sont pas situées autour d'une valeur

moyenne constante. Deux lignes ont été tracées pour faciliter la compréhension du changement de moyenne dans la figure.

Tableau 3.1- Données de la variable $W_i(t)$. (Caron, 2002.)

t	T	W_1	W_2	W_3
1	240	0.1050	0.1174	0.1160
2	480	0.1121	0.1138	0.1130
3	720	0.1109	0.1117	0.1137
4	960	0.1093	0.1175	0.1131
5	1200	0.1122	0.1149	0.1089
6	1440	0.1098	0.1147	0.1123
7	1680	0.1086	0.1174	0.1127
8	1920	0.1115	0.1174	0.1137
9	2160	0.1140	0.1118	0.1068
10	2400	0.1158	0.1126	0.1136
11	2640	0.1064	0.1046	0.1082
12	2880	0.1119	0.1166	0.1086
13	3120	0.1081	0.1163	0.1137
14	3360	0.1051	0.1098	0.1108
15	3600	0.1140	0.1180	0.1111
16	3840	0.1149	0.1118	0.1056
17	4080	0.1158	0.1107	0.1095
18	4320	0.1141	0.1044	0.1084
19	4560	0.1120	0.1147	0.1104
20	4800	0.1159	0.1148	0.1116
21	5040	0.1126	0.1055	0.1115
22	5280	0.1109	0.1137	0.1118
23	5520	0.1149	0.1137	0.1081
24	5760	0.1096	0.1097	0.1096
25	6000	0.1100	0.1180	0.1134
26	6240	0.1162	0.1075	0.1112
27	6480	0.1152	0.1161	0.1087
28	6720	0.1138	0.1171	0.1135
29	6960	0.1163	0.1175	0.1140
30	7200	0.1142	0.1042	0.1117
31	7440	0.1131	0.1175	0.1105
32	7680	0.1156	0.1122	0.1120
33	7920	0.1100	0.1059	0.1098
34	8160	0.1189	0.1082	0.1099
35	8400	0.1174	0.1153	0.1102
36	8640	0.1192	0.1070	0.1097
37	8880	0.1181	0.1078	0.1101

Tableau 3.1- Données de la variable $W_i(t)$. (Caron, 2002.) (Suite et fin.)

38	9120	0.1206	0.1114	0.1125
39	9360	0.1180	0.1118	0.1107
40	9600	0.1216	0.1139	0.1137
41	9840	0.1210	0.1149	0.1129
42	10080	0.1139	0.1041	0.1149
43	10320	0.1166	0.1101	0.1142
44	10560	0.1156	0.1048	0.1142
45	10800	0.1147	0.1078	0.1121
46	11040	0.1153	0.1080	0.1129
47	11280	0.1162	0.1067	0.1127
48	11520	0.1136	0.1048	0.1123
49	11760	0.1153	0.1155	0.1150
50	12000	0.1157	0.1064	0.1121

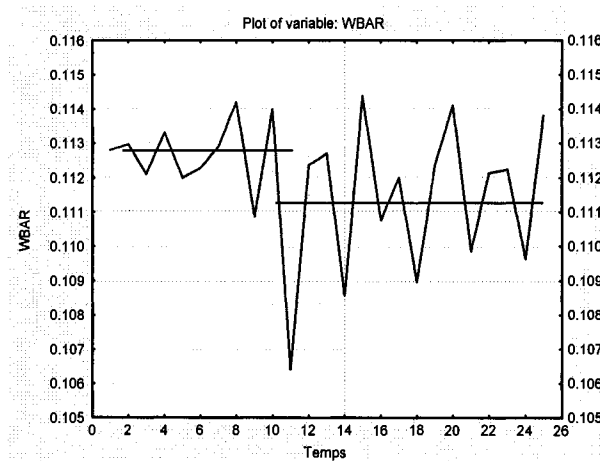


Figure 3-1. Série temporelle non stationnaire : $W(t)$, $t = 1, \dots, 25$.

Comme la méthodologie Box-Jenkins requiert que la série temporelle soit stationnaire, nous devons appliquer une différentiation de premier ordre qui se définit par la formule :

$$Z_w(t) = W(t) - W(t-1), \quad t = 2, \dots, 25. \quad (3-1)$$

La figure 3-2 montre le résultat de la transformation. La nouvelle série temporelle $Z_w(t)$ est devenue stationnaire. Maintenant, nous pouvons examiner la corrélation présente parmi les données et identifier le genre de modèle qui génère la série et qui enlèvera la corrélation.

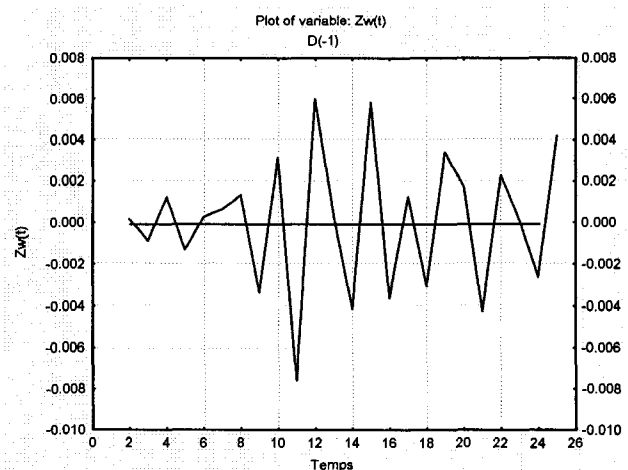


Figure 3-2. Série temporelle stationnaire : $Z_w(t)$

La figure 3-3 indique que la fonction d'autocorrélation de la série temporelle transformée, $Z_w(t)$, diminue lentement et que l'autocorrélation partielle n'est pas significative après le délai 2, on peut donc conclure que le modèle qui génère la série est autorégressif d'ordre 2, AR(2) (voir le tableau 2.4).

Voici l'équation du modèle AR(2) de la série $Z_w(t)$ obtenu avec Statistica (2001) :

$$\hat{Z}_w(t) = -1.12035 Z_w(t-1) - 0.60926 Z_w(t-2) \quad (3-2)$$

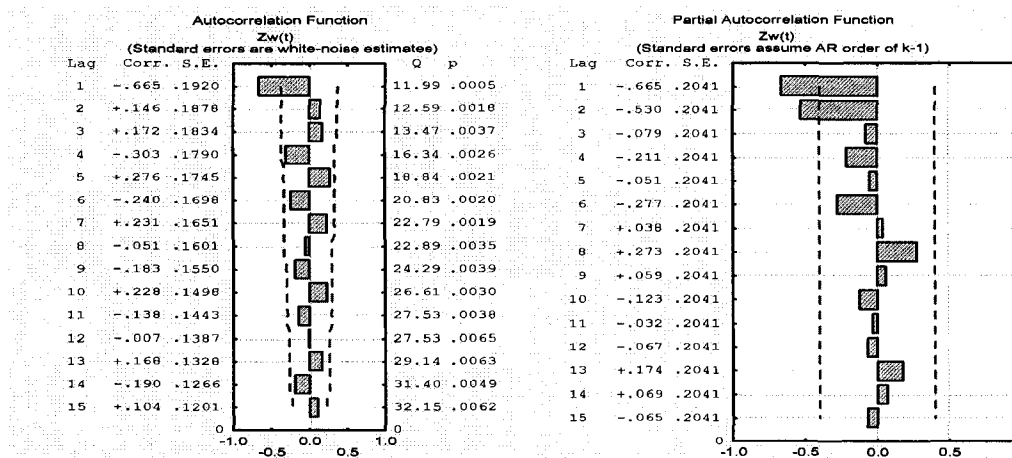


Figure 3-3. Fonctions d'auto-corrélation de la série des données $Z_w(t)$.

La figure 3-4 présente des graphes où on constate la diminution de l'autocorrélation une fois que le modèle autorégressif est appliqué. La figure 3-5 ne nous permet pas d'assurer la présence d'une distribution normale pour les résultats du modèle. Deux observations en particulier sont suspectes. Par contre, le test de Shapiro-Wilk confirme la normalité de la distribution car la valeur de la statistique w , égale à 0.93016, est située entre 0.7 et 1. Nous allons admettre que l'hypothèse est de validité plausible.

Comme le logiciel donne les valeurs intégrées du modèle $\hat{W}(t), t = 1, \dots, 25$, nous pouvons calculer les résidus $\varepsilon_w(t) = W(t) - \hat{W}(t), t = 1, \dots, 25$ en utilisant les valeurs des observations originelles $W(t)$. Le tableau 3.2 présente les résidus qui ont été calculés à partir du modèle et que nous allons employer pour la construction de la carte de contrôle de Hotelling à partir de $t=3$ car les résidus $\varepsilon_o(t)$ ont été obtenus à partir de cette valeur de t .

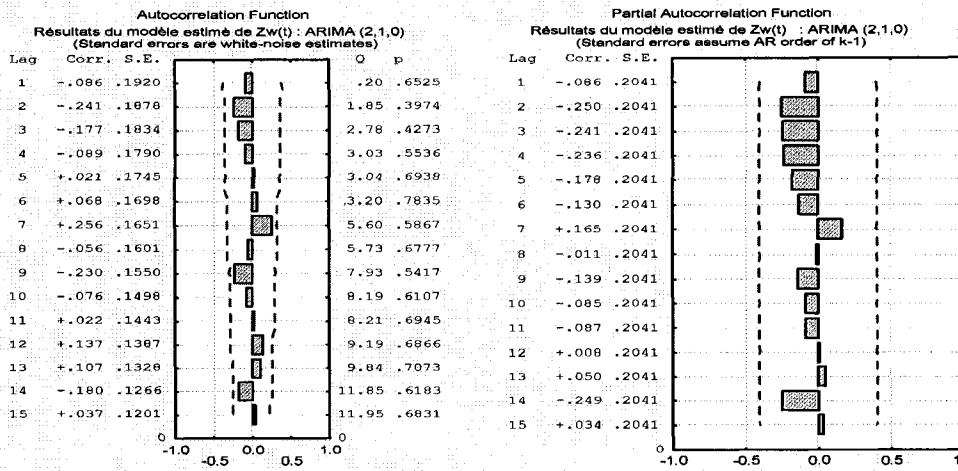


Figure 3-4. Graphe des autocorrélations du modèle AR(2). Série $\hat{W}(t)$.

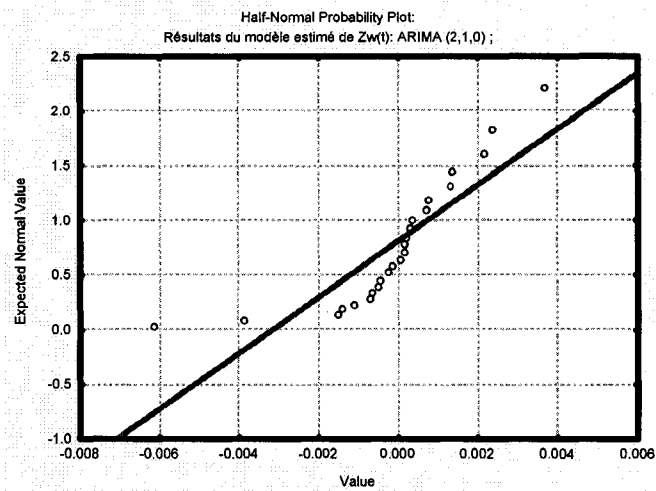


Figure 3-5. Distribution des résultats du modèle AR(2). Série $\hat{W}(t)$.

Tableau 3.2- Résidus du modèle AR(2) pour la variable $W(t), t = 1, \dots, 25$.

t	ε_W	t	ε_W	t	ε_W
1	0.0000	10	0.0013	19	0.0048
2	0.0000	11	-0.0002	20	0.0028
3	-0.0002	12	0.0064	21	-0.0012
4	0.0007	13	0.0044	22	0.0025
5	-0.0001	14	0.0004	23	0.0025
6	0.0006	15	0.0048	24	0.0010
7	0.0011	16	0.0008	25	0.0039
8	0.0002	17	0.0013		
9	-0.0016	18	0.0021		

Statistica donne des prévisions pour les nouvelles observations $t = 26, \dots, 50$; cela permet d'employer les valeurs originelles pour obtenir le reste des résidus qui seront employés pour la détection de l'usure avec la carte de contrôle de Hotelling. La figure 3-6 montre les prévisions pour la série des données intégrées $\hat{W}(t), t = 26, \dots, 50$ et le tableau 3.3 exhibe les valeurs ε_W obtenues. Ceci conclut l'étape 2 de la méthodologie.

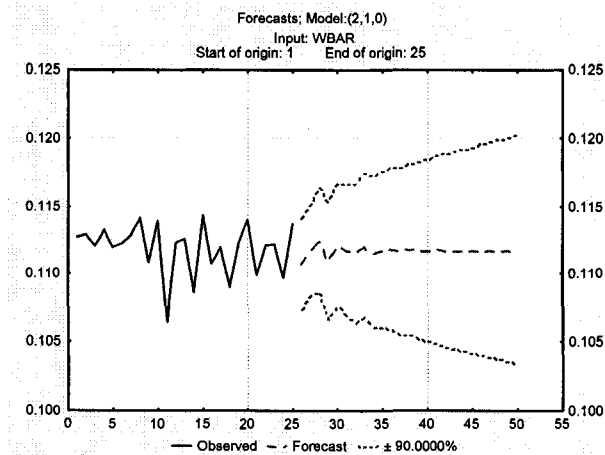
Figure 3-6. Prévisions du modèle AR(2). Série $\hat{W}(t)$.

Tableau 3.3- Résidus ε_W obtenus pour les prévisions de la série $\hat{W}(t)$, $t=26,\dots,50$.

t	ε_W	t	ε_W	t	ε_W
26	0.0009	35	0.0026	44	-0.0002
27	0.0017	36	0.0001	45	-0.0002
28	0.0023	37	0.0004	46	0.0003
29	0.0050	38	0.0030	47	0.0001
30	-0.0021	39	0.0018	48	-0.0015
31	0.0019	40	0.0047	49	0.0035
32	0.0018	41	0.0045	50	-0.0003
33	-0.0035	42	-0.0007		
34	0.0008	43	0.0019		

Les prochaines sections présentent les détails concernant l'étape 3 de la méthodologie : la construction des cartes de contrôle multidimensionnelles à partir des résidus indépendants et normalement distribués pour la détection de l'usure de l'outil.

3.3.- CARTE DE CONTRÔLE DE HOTELLING

En 1947, Hotelling a proposé une statistique, nommée T^2 , pour réaliser une carte de contrôle dans le cas multidimensionnel.

T^2 est une statistique unidimensionnelle qui représente le carré de la distance entre un vecteur d'observations normal p -dimensionnel « Y » et le centre du processus, représenté par le vecteur « \bar{Y} », moyenne de la population. Montgomery (2005) en donne l'expression :

$$T_i^2 = n (Y_i - \bar{Y})' \Sigma^{-1} (Y_i - \bar{Y}), i = 1, \dots, m \quad (3-3)$$

où Σ^{-1} est l'inverse de la matrice des variances et covariances ; n , est la taille de l'échantillon et m est le nombre total d'observations ou d'échantillons.

Plus la valeur de T_i^2 est élevée, plus éloignée se trouve le vecteur d'observations de la moyenne ou cible. Quand une valeur T_i^2 excède la limite de contrôle supérieure, UCL , le processus est qualifié de hors contrôle. Les limites, LCL et UCL , sont calculées sur un quantile d'une loi F pour l'erreur de type I, α , le nombre des variables, p , le nombre total d'échantillons (ou observations), m , et la taille des échantillons, n . La limite inférieure, LCL , est zéro parce que la statistique T_i^2 est une distance.

$$LCL=0 \quad \text{et}$$

$$UCL = F_{\alpha, p, mn-m-p+1} \left[\frac{p(m-1)(n-1)}{mn-m-p+1} \right]. \quad (3-4)$$

3.3.1.- RÈGLES POUR LA RECONNAISSANCE DE PATRONS DANS UNE CARTE DE HOTELLING

L'analyse de patrons dans les cartes T^2 demande la connaissance du processus et de l'historique des données. Comme les règles diagnostiques des cartes de contrôle spécifiées par «The Western Electric Handbook» ne sont pas appropriées pour détecter les observations qui causent le patron, à cause de la distribution non normale de T^2 , Mason *et al.* (2003) conseillent l'utilisation de la distribution T^2 en tant qu'outil de diagnostic dans ce cas.

Le tableau 3.4 explique les patrons qu'ils ont identifiés et les problèmes qui y sont associés.

Tableau 3.4- Situations associées aux patrons dans les cartes T^2 .

Patrons	Problème
Forme U répétée (période courte)	Effet de cycle: Données corrélées
Valeurs de T^2 regroupées au-dessus de la ligne zéro	Mélange de populations
Valeurs de T^2 avec tendance à la hausse/baisse	Tendance dans les observations
Groupements de valeurs de T^2	Changement dans le processus

3.3.2.- CARTE DE HOTELLING POUR $n=1$.

Dans la pratique, quand *la taille de l'échantillon n est égale à 1*, comme dans notre cas d'étude, un traitement spécial est requis. Tracy *et al.* (1992) ont développé une procédure basée sur la mesure de T^2 pour ce cas.

Yang et Trewn (2004) expliquent que cette procédure se déroule en deux phases. La première sert à estimer le vecteur de moyennes et la matrice de variances et covariances pour établir une référence du procédé sous contrôle statistique. Les données multidimensionnelles de l'état stable du processus sont examinées et les observations aberrantes sont éliminées. Cette référence est employée dans la deuxième phase, où le reste des observations recueillies sont utilisées.

Pour *la phase I*, on dispose de 23 observations multidimensionnelles qui sont les résidus non corrélés qui appartiennent à l'état de fonctionnement normal de l'outil de coupage et qui ont été obtenus avec la méthodologie Box-Jenkins pour les variables $O_i(t)$ et $W_i(t)$, les

deux types de rugosité. On a déjà précisé que l'emploi de résidus est nécessaire pour éliminer la présence d'autocorrélation parmi les données, autrement la puissance de la carte de contrôle serait diminuée (Montgomery et Mastrangelo, 1991).

Comme dans le cas du premier type de rugosité, on a retenu deux modèles, on a deux séries de résidus : $\varepsilon_1(t)$ et $\varepsilon_2(t)$, $t=3,\dots,25$ (tableau 2.6 et 2.7 dans le cas des prévisions).

Nous allons employer ici la moyenne de ces données, calculée selon la formule

suivante : $\varepsilon_o(t) = \frac{1}{2} \sum_{t=3}^{25} [\varepsilon_1(t) + \varepsilon_2(t)]$; $t = 3,\dots,25$. Ensuite, conjointement avec les résidus

$\varepsilon_w(t)$, $t = 3,\dots,25$ de la deuxième variable, on peut former 23 groupes de résidus

$\varepsilon(t) = [\varepsilon_w(t), \varepsilon_o(t)]^T$; $t = 3,\dots,25$. Les observations sont regroupées dans le tableau 3.5.

Tableau 3.5- Résidus employés pour la construction des cartes multidimensionnelles.

t	ε_w	ε_o	t	ε_w	ε_o
3	-0.0002	-0.0003	15	0.0048	0.0027
4	0.0007	-0.0001	16	0.0008	0.0033
5	-0.0001	-0.0001	17	0.0013	0.0039
6	0.0006	0.0004	18	0.0021	0.0042
7	0.0011	0.0008	19	0.0048	0.0039
8	0.0002	0.0016	20	0.0028	0.0028
9	-0.0016	0.0021	21	-0.0012	0.0028
10	0.0013	0.0013	22	0.0025	0.0017
11	-0.0002	0.0021	23	0.0025	0.0014
12	0.0064	0.0025	24	0.0010	0.0018
13	0.0044	0.0025	25	0.0039	0.0007
14	0.0004	0.0027			

Le vecteur de moyennes pour le groupe des 23 résidus est $\bar{\varepsilon} = [\bar{\varepsilon}_W, \bar{\varepsilon}_O] = [0.0017; 0.0019]$

où $\bar{\varepsilon}_W = \frac{1}{m} \sum_{t=3}^{25} \varepsilon_W(t)$ et $\bar{\varepsilon}_O = \frac{1}{m} \sum_{t=3}^{25} \varepsilon_O(t)$, avec $m=23$.

Pour le cas d'observations individuelles, l'estimation de la matrice de variances et covariances est très importante (Vargas, 2003.) Nous allons utiliser un estimateur classique, S , qui donne une matrice $p \times p$, et une carte de contrôle pour la dispersion pour valider ce choix.

$$S = \frac{1}{m-1} \sum_{t=3}^{25} [\varepsilon(t) - \bar{\varepsilon}] * [\varepsilon(t) - \bar{\varepsilon}]^T \quad (3-5)$$

On calcule :

$$S^{-1} = \left\{ \frac{1}{23-1} \sum_{t=3}^{25} [\varepsilon(t) - \bar{\varepsilon}] * [\varepsilon(t) - \bar{\varepsilon}]^T \right\}^{-1} = 1.0e-005 * \begin{bmatrix} 2.5785 & -1.2387 \\ -1.2387 & 6.4779 \end{bmatrix} \quad (3-6)$$

La distance T^2 pour les observations individuelles est alors,

$$Q(t) = [\varepsilon(t) - \bar{\varepsilon}]^T S^{-1} [\varepsilon(t) - \bar{\varepsilon}] \quad (3-7)$$

Alors, pour le calcul de la valeur de $Q(t)$, en utilisant l'équation 3-6 dans la formule 3-7, nous avons la première valeur du tableau 3.5,

$$Q(3) = [(-0.0002 - 0.0017); (-0.0003 - 0.0019)]^T * 1.0e + 005 * \begin{bmatrix} 2.5785 & -1.2387 \\ -1.2387 & 6.4779 \end{bmatrix} \\ * [-0.0019, -0.0022] = 3.0306$$

Le calcul pour les 22 observations restantes est similaire, les résultats obtenus sont montrés au tableau 3.6.

Tableau 3.6- Valeurs de $Q(t)$, $t=3, \dots, 25$ pour la construction de la carte de Hotelling.

t	$Q(t)$	t	$Q(t)$	t	$Q(t)$
3	3.0306	11	1.0509	19	3.5331
4	2.3536	12	5.2304	20	0.5915
5	2.5348	13	1.7116	21	3.3398
6	1.3608	14	1.1080	22	0.2306
7	0.7132	15	2.2781	23	0.4261
8	0.5270	16	1.7907	24	0.1155
9	2.9974	17	2.8306	25	2.8348
10	0.2150	18	3.2402		

Tracy, Young et Mason (1992) ont démontré que $Q(t)$ suit une loi Beta, ce qui permet de trouver une expression pour l'estimation des limites de contrôle de la phase I. Nous allons utiliser l'expression équivalente qui emploie la distribution F. Dans notre cas $m=23$, $p=2$ et l'erreur type I, $\alpha=0.05$.

$$LCL_Q \cong 0 \quad (3-8)$$

$$UCL_Q = \frac{(m-1)^2}{m} \frac{[p/(m-p-1)]F_{\alpha,p,m-p-1}}{1 + [p/(m-p-1)]F_{\alpha,p,m-p-1}} \quad (3-9)$$

$$UCL_Q = \frac{(23-1)^2}{23} \frac{[2/(23-2-1)] * F_{0.05,2,23-2-1}}{1 + [2/(23-2-1)] * 3.42} = 5.3628$$

Ici, l'attention est dirigée vers la limite supérieure. Si une observation se trouve au-dessous de la limite inférieure, ceci indiquerait un problème dans la cueillette des données. Mais si une observation se trouve au-dessus de la limite supérieure, celle-ci est qualifiée d'aberrante, elle est éliminée et les paramètres du processus sont recalculés : le vecteur de moyennes, la matrice de variances et covariances et les limites de contrôle. La procédure est répétée jusqu'à ce que l'on obtienne l'ensemble des données qui vont servir pour la carte de contrôle dans la phase II.

Voici la carte obtenue (figure 3-7) dans la phase I. Les observations montrent un comportement aléatoire et il n'y a pas de points aberrants. On peut donc considérer que l'on a établi l'ensemble des données nécessaires pour la détection de l'usure dans la phase II.

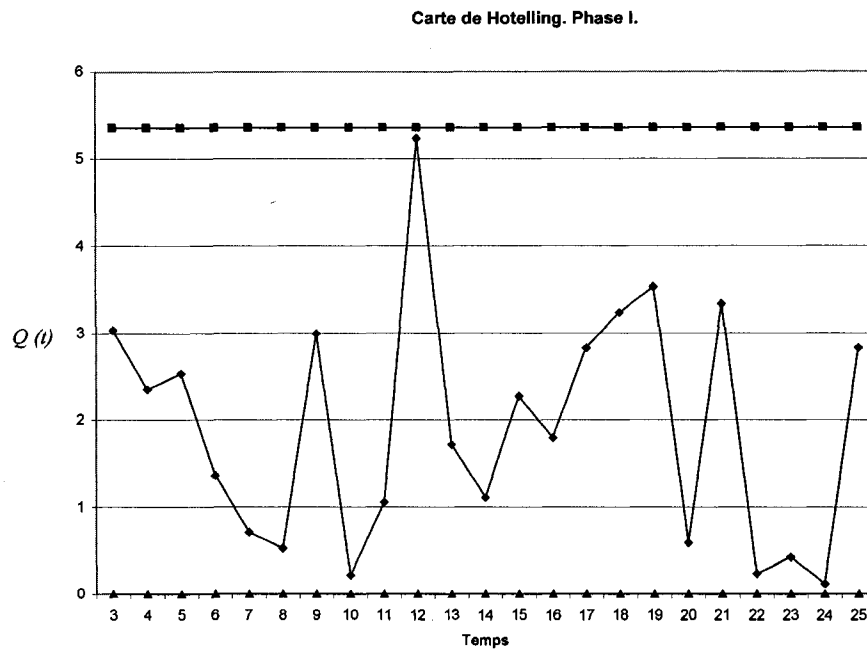


Figure 3-7. Carte de Hotelling. Phase I.

3.4.- CARTE DE CONTRÔLE POUR LA DISPERSION, $n=1$.

Cette carte multidimensionnelle est nécessaire afin de confirmer si la matrice de variances et covariances, S^{-1} , employée pour l'estimation de $Q(t)$, ne change pas avec le temps, puisque telle est la condition qu'il faut vérifier pour utiliser la carte de Hotelling.

Khoo et Quah (2003) proposent une carte de contrôle axée sur un paramètre qui tient compte des différences successives en observations multidimensionnelles pour l'étude de plusieurs caractéristiques de la qualité quand $n=1$. Ce paramètre est dans notre cas,

$$M(t) = \frac{1}{2} [\varepsilon(t) - \varepsilon(t-1)]^T S^{-1} [\varepsilon(t) - \varepsilon(t-1)], t=4, \dots, 25. \quad (3-10)$$

Pour cette carte, on dispose de 23 vecteurs d'observations multidimensionnelles $\varepsilon(t) = [\varepsilon_W(t), \varepsilon_O(t)]^T, t = 3, \dots, 25$. où $n=1$, qui sont les résidus obtenus. Selon la méthode de Koo et Quah, celles-ci doivent provenir d'un historique de données fiables avec une distribution multidimensionnelle $N_p(\mu_\varepsilon, \Sigma_\varepsilon)$ où μ_ε est le vecteur nominal de moyennes et Σ_ε la matrice de variances et covariances.

L'inverse de l'estimation de la matrice de variances et covariances qu'on emploie ici est la même qu'à l'équation 3-6. Alors, pour le calcul de la valeur de $M(t)$, nous avons,

$$M(4) = \frac{1}{2} [\varepsilon(4) - \varepsilon(3)]^T S^{-1} [\varepsilon(4) - \varepsilon(3)]$$

$$M(4) = \frac{1}{2} \left\{ [0.0009; 0.0002]^T * 1.0e + 005 * \begin{bmatrix} 2.5785 & -1.2387 \\ -1.2387 & 6.4779 \end{bmatrix} * [0.0009; 0.0002] \right\} = 0.1902$$

Le tableau 3.7 rapporte le reste des résultats. Comme le paramètre $M(t)$ suit la distribution chi-deux à 2 degrés de liberté, les limites de contrôle pour $n=1$ sont,

$$UCL_M = \chi_{p,\alpha}^2 = \chi_{2,0.05}^2 = 5.991 \quad (3-11)$$

$$LCL_M = \chi_{p,1-\alpha}^2 = \chi_{2,1-0.05}^2 = 0.103 \quad (3-12)$$

où $\chi_{p,\alpha}^2$ est le quantile $1-\alpha$ de la distribution avec $p=2$ degrés de liberté, $\alpha=0.05$ étant l'erreur de type I. Alors, si une observation se retrouve au-dessus de la limite supérieure, le processus est qualifié de hors contrôle. Mais si l'observation se retrouve au-dessous de la limite inférieure, une diminution générale de la dispersion du processus serait attribuée, ce qui est une indication qui montre que le processus s'améliore.

Tableau 3.7- Valeurs de $M(t)$, $t=4, \dots, 25$ obtenues dans la phase I de la carte de dispersion.

t	$M(t)$	t	$M(t)$	T	$M(t)$
4	0.1902	12	10.6814	20	1.2702
5	0.1650	13	1.0314	21	4.1255
6	0.2016	14	4.3496	22	5.3220
7	0.1186	15	4.9919	23	0.0583
8	0.8018	16	4.9533	24	0.8324
9	1.2203	17	0.2233	25	3.7426
10	3.1578	18	0.1639		
11	1.2920	19	2.1387		

La figure 3-8 montre la carte qui a été construite à partir des données du tableau 3.7. Les observations suivent un comportement aléatoire. Cependant les observations numéro 12 et 23 se trouvent en dehors des limites, ce pourquoi nous allons supprimer de la liste des données les observations qui servent à les obtenir, c'est-à-dire, les observations 11-12, et les observations 22-23.

Avec la nouvelle liste de $m=19$ observations, on obtient une nouvelle estimation de la matrice de variances et covariances, on utilisera l'inverse de cette matrice par la suite dans les calculs.

$$S^{-1} = \left\{ \frac{1}{19-1} \sum_{t=3}^{25} [\varepsilon(t) - \bar{\varepsilon}] * [\varepsilon(t) - \bar{\varepsilon}]^T \right\}^{-1} = 1.0e+005 * \begin{bmatrix} 3.0853 & -1.3697 \\ -1.3697 & 5.5117 \end{bmatrix} \quad (3-13)$$

On présente maintenant les nouvelles valeurs obtenues pour le paramètre $M(t)$ dans le tableau 3.8.

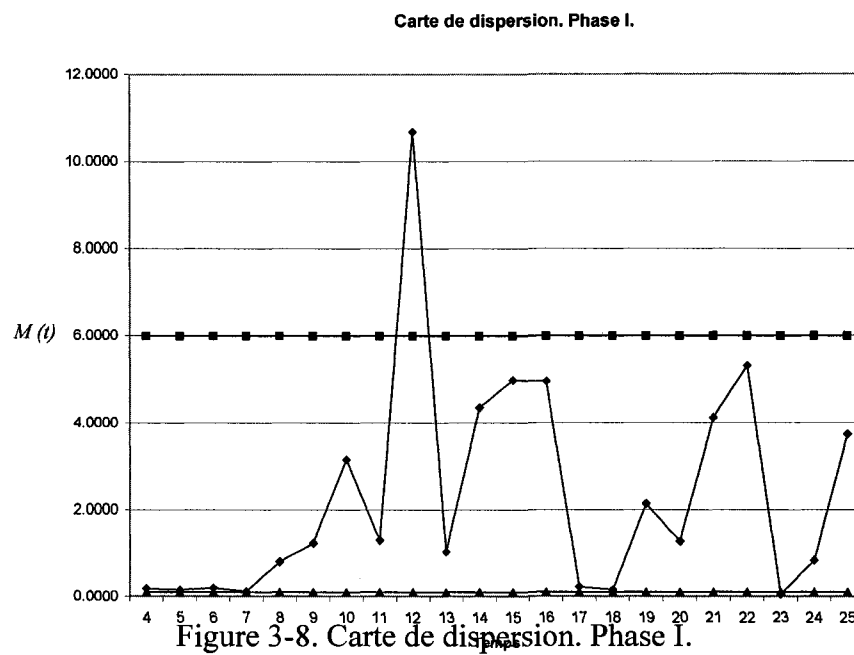


Tableau 3.8- Valeurs de $M(t)$, $t=4, \dots, 25$ employées pour la carte de dispersion après le tamisage des données.

t	$M(t)$	t	$M(t)$	t	$M(t)$
4	0.2226	14	5.1776	24	2.6471
5	0.1975	15	5.9730	25	4.1355
6	0.1931	16	5.7023		
7	0.1105	17	0.1934		
8	0.7999	18	0.1813		
9	1.3840	19	2.5206		
10	3.5830	20	1.2984		
13	2.7396	21	4.9364		

Les limites de contrôle gardent la même valeur (équations 3-11 et 3-12), alors la carte révisée est montrée dans la figure 3-9. Nous allons employer l'inverse de la matrice de variances et covariances recalculée (Équation 3-13) après tamisage des données, et les limites de contrôle obtenues pour la détection de l'usure.

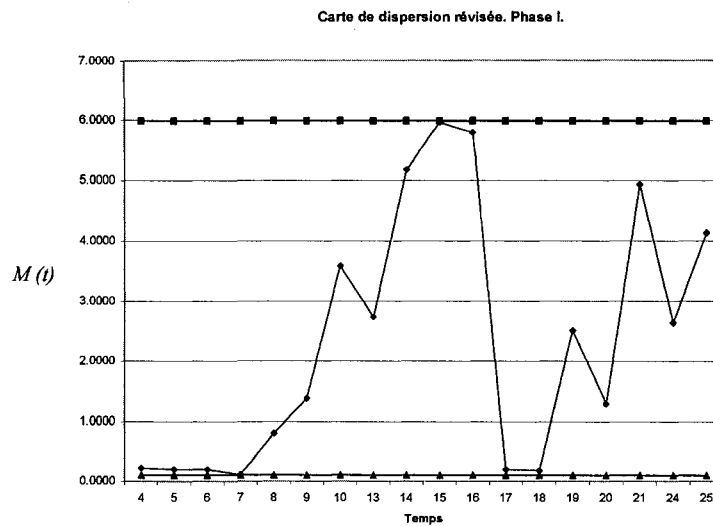


Figure 3-9. Carte de dispersion révisée. Phase I.

3.5.- RÉSULTATS SUR LA DÉTECTION DE L'USURE

Ici, nous allons utiliser le reste des observations des deux types de rugosité pour la détection de l'usure, c'est-à-dire, les observations 26-50 du tableau 3.3 qui correspondent aux résidus de la variable $W_i(t)$, que nous avons dénommés $\varepsilon_w(t)$, ainsi que la moyenne des observations 26-50 montrées dans le tableau 2.7 qui constituent les résidus de la variable $O_i(t)$ et que nous avons dénommés $\varepsilon_o(t)$. Le tableau 3.9 regroupe les données.

Tableau 3.9- Résidus employés pour la détection de l'usure.

t	ε_W	ε_O	t	ε_W	ε_O
26	0.0009	0.0027	39	0.0018	0.0009
27	0.0017	-0.0005	40	0.0047	0.0012
28	0.0023	0.0016	41	0.0045	0.0035
29	0.0050	0.0016	42	-0.0007	0.0033
30	-0.0021	0.0025	43	0.0019	0.0048
31	0.0019	0.0025	44	-0.0002	0.0045
32	0.0018	0.0017	45	-0.0002	0.0041
33	-0.0035	0.0022	46	0.0003	0.0040
34	0.0008	-0.0004	47	0.0001	0.0046
35	0.0026	0.0001	48	-0.0015	0.0056
36	0.0001	0.0006	49	0.0035	0.0047
37	0.0004	0.0010	50	-0.0003	0.0041
38	0.0030	0.0014			

Maintenant, on présente les détails pour la détection de l'usure dans chacune des cartes.

3.5.1.- CARTE DE HOTELLING

La *phase II* traite de l'établissement des nouvelles limites pour des nouvelles observations. À cette phase, chacune des nouvelles observations est un vecteur nommé $\varepsilon^{(new)} = (\varepsilon_W, \varepsilon_O)$. La distance T^2 est alors,

$$Q^{(new)} = (\varepsilon^{(new)} - \bar{\varepsilon})^T S^{-1} (\varepsilon^{(new)} - \bar{\varepsilon}) \quad (3-14)$$

En utilisant la matrice de variances et covariances de l'équation 3-6 dans l'équation 3-14, et le vecteur de moyennes $\bar{\varepsilon} = [\bar{\varepsilon}_W, \bar{\varepsilon}_O] = [0.0017; 0.0019]$, nous avons la valeur de $Q(26)$,

$$Q(26) = [-0.0008; 0.0008]^T * 1.0e + 005 * \begin{bmatrix} 2.5785 & -1.2387 \\ -1.2387 & 6.4779 \end{bmatrix} \\ * [-0.0008; 0.0008] = 0.7382$$

Le calcul pour le reste des observations est similaire, le tableau 3.10 montre les résultats obtenus.

Tableau 3.10- Valeurs de $Q(t), t=26, \dots, 50$ employées pour la détection de l'usure avec la carte de Hotelling.

t	$Q(t)$	T	$Q(t)$
26	0.7382	39	0.6751
27	3.7313	40	3.1583
28	0.1957	41	2.5700
29	3.1115	42	3.5872
30	4.5213	43	5.3146
31	0.2138	44	6.5337
32	0.0334	45	5.1017
33	7.4169	46	4.0905
34	3.1229	47	6.4527
35	2.7090	48	14.4418
36	1.2396	49	4.6656
37	0.6706	50	5.2567
38	0.7587		

On calcule des nouvelles limites de contrôle pour la phase II avec $m=25, p=2$ et l'erreur type I, $\alpha = 0.05$. Elles sont données par :

$$LCL = 0 \quad (3-15)$$

$$UCL = \frac{p(m+1)(m-1)}{m(m-p)} F_{\alpha, p, m-p} \quad (3-16)$$

$$UCL = \frac{2(25+1)(25-1)}{25(25-2)} F_{0.05;2,23} = \frac{1248}{575} * 3.42 = 7.4229$$

Si une observation se retrouvait au-dessus de la limite supérieure, le processus serait qualifié de hors contrôle et la cause devrait être trouvée. Dans notre cas, cela serait la manifestation de l'usure de l'outil.

Voici la carte obtenue (figure 3-10). On identifie une zone de transition et une zone de vieillesse comme dans le cas unidimensionnel. Dans la zone de transition (observations 26-38), il n'y a pas d'observations hors contrôle, seulement l'observation 33 est proche de la limite supérieure. La zone de vieillesse, à partir de l'observation 39, exhibe une tendance qui se développe jusqu'à ce que l'observation 48 se trouve hors des limites. Ce que l'on considère comme la manifestation du phénomène d'usure.

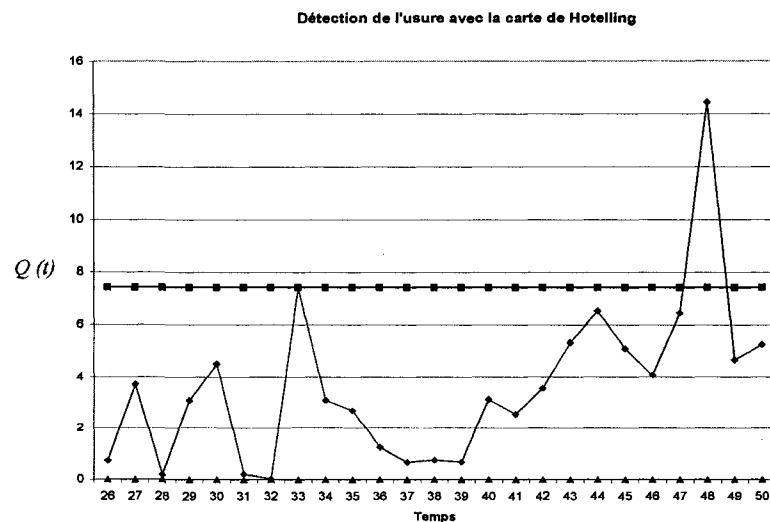


Figure 3-10. Détection de l'usure avec la Carte de Hotelling.

Il est raisonnable de localiser le début du phénomène d'usure à partir de l'observation 39 : c'est le début d'une tendance à l'augmentation. Le temps de remplacement correspond à 9360 secondes (156 minutes) d'utilisation, et ce chiffre est inférieur aux 200 minutes fixées par l'historique des données du même outil lorsqu'il tourne moins rapidement (13000 RPM).

On pourrait considérer aussi de faire le remplacement lorsque l'outil est complètement usé bien que cela entraînerait des coûts de non qualité du produit. Le temps de remplacement serait choisi à l'observation 48 aux *11520 secondes* (192 minutes) d'utilisation selon le tableau 2.1.

3.5.2.- CARTE DE LA DISPERSION

Dans cette étape, on emploie la matrice de variances et covariances révisée pour le calcul des paramètres $M(t)$, $t=26, \dots, 50$. (Équation 3-13). Alors pour le calcul de la valeur de $M(t)$, nous avons,

$$M(26) = \frac{1}{2} [\varepsilon(26) - \varepsilon(25)]^T S^{-1} [\varepsilon(26) - \varepsilon(25)]$$

$$M(26) = \frac{1}{2} \left\{ [-0.003, 0.002]^T * 1.0e + 005 * \begin{bmatrix} 3.0853 & -1.3697 \\ -1.3697 & 5.5117 \end{bmatrix} * [-0.003, 0.002] \right\} = 6.6251$$

Le calcul pour les observations restantes est similaire, les résultats obtenus sont présentés au tableau 3.11. Les limites de contrôle pour cette carte sont celles que l'on a obtenues dans la première étape : $UCL_M = \chi_{2,0.05}^2 = 5.991$ et $LCL_M = \chi_{2,1-0.05}^2 = 0.103$.

La figure 3-11 montre la carte obtenue. Afin de poursuivre l'analyse selon les mêmes principes que l'on a appliqués dans les cartes précédentes, on sépare les observations de la zone de transition (observations 26-38) et de la zone de vieillesse (observations 39-50).

Tableau 3.11- Valeurs de $M(t)$, $t=26,\dots,50$ employées pour la détection de l'usure avec la carte de dispersion.

T	$M(t)$	t	$M(t)$
26	6.6251	39	0.4177
27	6.5428	40	2.4060
28	2.1966	41	3.0541
29	2.2491	42	8.0797
30	17.7497	43	2.2574
31	4.9364	44	1.2376
32	0.3339	45	0.0882
33	9.5302	46	0.0963
34	12.4932	47	0.2436
35	0.8909	48	1.7793
36	2.4085	49	9.3923
37	0.0831	50	4.0289
38	1.8889		

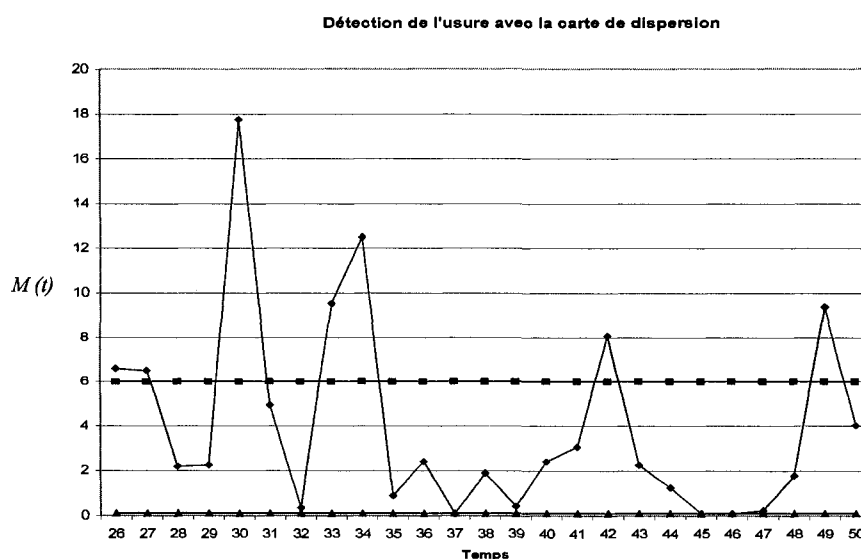


Figure 3-11. Détection de l'usure avec la carte de dispersion.

Dans la carte de la figure 3-11, on remarque que les observations numéro 26, 27, 30, 33 et 34 se trouvent hors de contrôle dans la zone de transition. Ce comportement ressemble à celui de la carte R dont on retrouve des entrées et sorties des observations par rapport à la limite supérieure, ce qui pourrait s'interpréter comme étant des signaux qui indiquent l'entrée dans la zone de vieillesse. La dispersion change donc avec le temps.

Dans la zone de vieillesse, on remarque uniquement les observations 42 et 49 hors contrôle. Les résultats de cette carte sont donc plus utiles pour identifier des signaux dans la zone de transition que dans la zone de vieillesse : comme dans le cas unidimensionnel, on obtient plus d'observations hors de contrôle dans la zone de transition ce qui permet de bien l'identifier.

Ceci conclut la dernière étape de la méthodologie, le phénomène d'usure a été détecté efficacement avec la carte de Hotelling, et la carte de dispersion a permis de confirmer que la dispersion change avec le temps et que l'utilisation d'un estimateur classique pour la matrice de variances et covariances, S^{-1} , dans la carte de Hotelling est convenable.

C'est ainsi que le temps de remplacement de l'outil peut être localisé au début de l'usure à 156 minutes, ou à 192 minutes d'utilisation, si on peut tolérer les coûts de non qualité du produit. Le choix dépendra des conditions de la politique de maintenance à implanter au sein de l'entreprise.

CONCLUSION

La méthodologie statistique développée ici a permis de détecter le phénomène d'usure efficacement avec les cartes de contrôle unidimensionnelles et multidimensionnelles.

Plusieurs constatations ont été réalisées pendant le développement de la méthodologie de trois étapes :

- L'utilisation des cartes de contrôle pour l'identification de l'usure implique une analyse statistique rigoureuse lorsque les données ne sont ni indépendantes ni stationnaires.
- Les données de cette étude ne sont pas parfaites pour l'illustration de la méthodologie, probablement aucune donnée réelle ne le serait. Cette problématique est commune aux données dans le cadre de la maintenance. Malgré ce fait, les données ont l'avantage d'être réelles, ce pourquoi on a admis tout au long du texte la validité des certaines hypothèses statistiques, alors qu'elles tenaient plus ou moins. Ce qui se produirait probablement dans toute application réelle.
- Dans *l'étape 1 de la méthodologie*, les premières 25 observations ont été identifiées comme la période pendant laquelle le processus est sous contrôle statistique, là où l'usure ne se manifeste pas.

On remarque dans cette étape, une utilisation non traditionnelle d'une analyse de composantes principales (ACP) suivie de la classification ascendante hiérarchique selon le critère d'agrégation de Ward pour effectuer le regroupement des observations selon l'état de dégradation. Cela nous a permis de déterminer l'information à propos de l'état de fonctionnement normal de l'outil contenu dans le tableau des données d'une façon cachée. Généralement, on utilise l'ACP pour la réduction des dimensions des données

lorsqu'il y a plusieurs variables. Notre étude a employé six mesures d'une même variable comme s'il s'agissait des variables différentes. Nous avons cependant validé sommairement cette procédure par des permutations aléatoires des échantillons à chaque temps d'observation. C'est la cohérence des résultats qui a décidé des observations se rapportant au fonctionnement normal de l'outil.

- Dans *l'étape 2 de la méthodologie*, nous avons éliminé la structure corrélative des données de chacune des variables grâce à l'emploi d'un modèle autorégressif de Box-Jenkins qui nous a fourni des résidus indépendants. Dans cette étape, on a démontré la procédure pour le tamisage d'une série temporelle lorsqu'elle ne remplit pas les conditions d'inversion requises par les calculs matriciels effectués dans le cadre de la méthodologie Box-Jenkins pour trouver les coefficients du modèle mathématique.
- Dans *l'étape 3 de la méthodologie*, on a effectué la construction des cartes de contrôle unidimensionnelles et multidimensionnelles avec des résidus indépendants et normalement distribués lorsque l'outil est neuf. On a pris en considération la taille de l'échantillon n dont on disposait : les cartes $X\text{-bar}$ et R sont utilisées quand $n < 10$ ($n=2$) dans le cas unidimensionnel, la carte de Hotelling et la carte multidimensionnelle de dispersion lorsque $n=1$. Les cartes ont été construites avec une erreur de type I égale à 0.05.

On a identifié trois zones selon le cycle de vie pour l'analyse des cartes de contrôle: la *zone de jeunesse* qu'on a déjà expliqué dans l'étape 1; la *zone de transition*, définie par un patron d'entrées et sorties d'observations par rapport aux limites dans les cartes de dispersion; et la *zone de vieillesse*, où les résidus montrent une tendance dans la carte $X\text{-bar}$ et de Hotelling.

L'usure a été détectée principalement par les cartes qui mesurent la tendance centrale : la carte $X\text{-bar}$ dans le cas unidimensionnel, et la carte de Hotelling dans le cas

multidimensionnel. Les cartes de dispersion ont plutôt permis de confirmer la présence d'une zone de transition au milieu des données (Huo et Al. 1997) qui fut identifiée aussi avec l'ACP suivi de la classification. La carte de dispersion multidimensionnelle a permis de confirmer que la dispersion change avec le temps et que l'utilisation d'un estimateur classique pour la matrice de variances et covariances, S^{-1} , dans la carte de Hotelling est convenable.

- **Les résultats** de cette recherche ont permis de déterminer la durée de vie utile d'un outil en termes de minutes pour un cas spécifique. Toutefois, le protocole statistique établi ici pour détecter l'usure avec des cartes de contrôle lorsque les données sont corrélées est applicable à tous les procédés, quand le phénomène d'usure, manifeste sous la forme d'une tendance linéaire qui augmente selon l'ordre de production, peut être modélisé grâce à des modèles de séries chronologiques, ici l'utilisation d'un modèle autorégressif.
- En utilisant la méthodologie pour la surveillance de l'usure proposée dans ce travail, on a obtenu des renseignements pour développer une **politique de remplacement**. L'entreprise peut choisir de *remplacer l'outil au début de l'usure* (le temps de remplacement est de 140 minutes selon la carte *X-bar*, et de 156 minutes d'après la carte de Hotelling) où de le *remplacer lorsqu'il est complètement usé*, si on peut se permettre d'inclure des coûts de non qualité du produit (le remplacement doit être effectué à toutes les 176 minutes d'utilisation selon la carte *X-bar* ou aux 192 minutes d'après la carte de Hotelling), mais cela ne tient pas compte des coûts de maintenance et de qualité. Le choix demanderait une analyse de coûts de non remplacement et de remplacement, puis une analyse où le temps de remplacement minimiserait une combinaison de ces coûts. Ce point peut être le sujet d'une prochaine recherche.

On peut comparer les résultats des cartes unidimensionnelles et multidimensionnelles parce qu'on s'est servi du même intervalle d'observations (État de fonctionnement

normal de l'outil : observations 1-25) pour établir les cartes. Lors de la comparaison, on a remarqué une différence de 16 minutes entre les temps de remplacement obtenus par la carte *X-bar* dans le cas unidimensionnel et par la carte de Hotelling dans le cas multidimensionnel. On recommande l'adoption du temps de remplacement proposé par la carte de Hotelling puisque l'approche multidimensionnelle permet de prendre une décision en tenant compte de plus d'information, deux types de rugosité furent employés par cette approche. En tout cas, une étude des coûts doit être effectuée, ce sont les résultats d'une telle analyse qui détermineront le choix le plus approprié.

- Vu que l'historique des données fixait la durée de vie de l'outil tournant à 13000 RPM à 200 minutes, il est raisonnable de le remplacer plus tôt lorsqu'il tourne plus rapidement, à 18000 RPM. Ces résultats confirment *l'utilité de la méthodologie* que l'on a mise en place. Alors, on recommande l'utilisation de cette approche pour la détection de l'usure lorsque les observations ne sont pas stationnaires et corrélées.
- Sujets pour des *recherches futures* : l'emploi d'autres cartes de contrôle multidimensionnelles pourrait être exploré. Des analyses de coûts de qualité et de la capacité du processus sont conseillées dès que les données deviennent disponibles.

RÉFÉRENCES

APPLEY, D. et TSUNG, F. (2002). 'The Autoregressive T² Chart for Monitoring Univariate Autocorrelated Processes'. Journal of Quality Technology. **34** (1): 80-96.

ATIENZA, O. ANG, B. et TANG, L. (1997). 'Statistical process control and forecasting'. International Journal of Quality Science. **2** (1): 37-51

BARONE, S. D'AMBROSIO, P. et ERTO, P. (2007). 'A Statistical Monitoring Approach for Automotive On-Board Diagnostic Systems'. Quality and Reliability Engineering International. **23**(5): 565-575.

BAYDAR, N. CHEN, Q. BALL, A. et KRUGER, U. (2001). 'Detection of incipient tooth defect in helical gears using multivariate statistics'. Mechanical Systems and Signal Processing. **15**(2): 303-321.

BERSIMIS, S. PSARAKIS, S. et PANARETOS, J. (2007). 'Multivariate Statistical Process Control Charts: An Overview'. Quality and Reliability Engineering International. **23**: 517-543.

BOUROCHE, J. et SAPORTA, G. (1980), 'L'analyse de données'. Presses universitaires de France. 1854: 126.

BOWERMAN, B. et O'CONNELL, R. (1979). 'Time Series & Forecasting: an applied approach'. O'connell. Wadsworth, Inc. California.

CARON, ANTOINE (2002). 'Implantation d'un système six sigma pour la détermination du temps de remplacement d'un outil de coupage'. Rapport présenté dans le cadre d'un travail effectué en vue d'obtention d'un diplôme de maîtrise à l'École Polytechnique de Montréal.

CARVALHO, C. PONS, M. et DA FONSECA, M. (2003). 'Principal Components Analysis as a Tool to Summarise Biotransformation Data : Influence on Cells of Solvent Type and Phase Ratio'. Biocatalysis and Biotransformation. **21**(6): 305-314.

CHAMP, C. JONES, L. et RIGDON, S. (2005). 'Properties of the t^2 Control Chart When Parameters are estimated'. Technometrics. **47**(4): 437.

CHAN, L. et LI, G-Y. (1994). 'A Multivariate Control Chart for Detecting Linear Trends'. Communications in Statistics and Simulation. **23**(4): 997-1012.

CHARNES, J. (1995). 'Tests for Special Causes With Multivariate Autocorrelated Data'. Computers Operational Research. **22**(4): 443-453.

CHEN, G. CHENG, S. et XIE, H. (2005). 'A New Multivariate Control Chart for Monitoring Both Location and Dispersion'. Communications in Statistics-Simulation and Computation. **34**: 203-217.

DOOLEY, K. et KAPOOR, S. (1990). 'An enhanced quality evaluation systems for continuous manufacturing processes, part 1: theory'. Transactions of the ASME, Journal of Engineering for Industry. **112**: 57-62.

DUNCAN, A. J. (1986). 'Quality control and industrial statistics', Irwin, Homewood, Sième édition.

FUGATE, M. SOHN, H. et FARRAR, C. (2001). 'Vibration-based damage detection using statistical process control', Mechanical Systems and Signal processing. **15**(4): 707-721.

GARCIA-ALLENDE, P. CONDE, O. CUBILLAS, A. et LOPEZ, J. (2008). 'Data processing method applying PCA and Spectral Angle Mapper (SAM) for imaging spectroscopic sensors'. IEEE Sensors, **8**(7): 1310 – 1316.

GONZALEZ, M. et RODRIGUEZ, P. (2003). 'Application of the Multivariate T^2 Control Chart and the Mason-Tracy-Young Decomposition Procedure to the Study of the Consistency of mpurity Profiles of Drug Substances'. Quality Engineering. **16**(1): 127-142.

GOODE, K. MOORE, J. et ROYLANCE, B. (2000). 'Plant machinery working life prediction method utilizing reliability and condition-monitoring data'. IMechE. **214**: 109-122.

HE, T. XIE, W-R. WU, Q-H. et SHI, T-L. (2006). 'Process Fault Detection and Diagnosis Based on Principal Component Analysis'. IEEE, 3551-3556.

HOERL, R. et PALM, A. (1992). 'Discussion: integrating SPC and APC'. Technometrics. **34**: 268-72.

HOTELLING, H. (1947). 'Multivariate Quality Control - Illustrated by the Air Testing of Sample Bombsights'. Techniques of Statistical Analysis. McGraw Hill, 111-184.

HUO, Y. CHEN, D. et WEN, S. (1997). 'Monitoring of the Wear Condition and Research on the Wear Process for Running Equipment'. Tribology Transactions. 40(1): 87-90.

JARDINE, A. LIN, D. et BANJEVIC, D. (2006). 'A review on machinery diagnostics and prognostics implementing condition-based maintenance'. Mechanical Systems and Signal Processing 20: 1483-1510.

KANO, M. TANAKA, S. HASEBE, S. HASHIMOTO, I. et OHNO, H. (2003). 'Monitoring Independent Components for Fault Detection'. AIChE Journal. 49(4): 969-976.

KIRBY, M. (2001). 'Geometric data analysis'. New York, NY : Wiley-Interscience Publication.

KOOH, M. et QUAH, S. (2003). 'Multivariate Control Chart for Process Dispersion Based on Individual Observations'. Quality Engineering. 15(4): 639-642.

KOOH, M. et QUAH, S. (2004). 'Alternatives to the Multivariate Control Chart for Process Dispersion'. Quality Engineering. 16(3): 423-435.

KOSANOVICH, K. PIOVOSO, M. DAHL, K. MACGREGOR, J. et NOMIKOS, P. (1994). 'Multi-Way PCA Applied to an Industrial Batch Process'. American Control Conference, 1294-1296.

LIAO, W. TING, C-F. et CHANG, P-C. (2006). 'An adaptive genetic clustering method for exploratory mining of feature vector and time series data', International Journal of production Research, 44(14): 2731-2748.

LOWRY, C. WOODALL, W. CHAMP, C. et RIGDON, S. (1992). 'A Multivariate Exponentially Weighted Moving Average Control Chart'. Technometrics. **34**(1): 46-53.

LOZANO, J. ARROYO, T. SANTOS, J. CABELLOS, J. et HORILLO, M. (2008). 'Electronic nose for wine aging detection'. Sensors and Actuators. B133: 180-186.

MAKIS, V. (1999). 'Optimal control policy for a tool-wear process subject to shocks and random failures', International Journal of Production Economics, 60-61, 613-621.

MANDEL, J. (1969). 'The regression control chart', Journal of Quality Technology, **1**(1): 1-9.

MANUELE, J. (1945). 'Control chart for determining tool wear', Industrial Quality Control. **1**(6): 7-10.

MASON, R. et YOUNG, J. (2007). 'Dependent Univariate Observations And Statistical Control'. Quality Progress. April, 62-64.

MASON, R. CHAMP, C. TRACY, N. WIERDA, S. et YOUNG, J. (2007). 'Assessment of Multivariate Process Control Techniques'. Journal of Quality Technology. **29**(2): 140-143.

MASON, R. CHOU, Y-M. SULLIVAN, J. STOUMBOS, Z. et YOUNG, J. (2003) 'Systematic Patterns in T^2 Charts'. Journal of Quality Technology, **35**(1): 47-58.

MASON, R. TRACY, N. et YOUNG, J. (1997). 'A practical Approach for Interpreting Multivariate T^2 Control Chart Signals'. Journal of Quality Technology. **29**(4): 396-406.

Matlab 7.0 The Mathworks, Inc 1984-2004. www.mathworks.com

Microsoft Excel 2003. Microsoft Corporation. 1985-2003. www.microsoft.com

MONTGOMERY D., (2005). 'Introduction to statistical quality control', John Wiley & Sons Inc. 5^o Edition.

MORINEAU, A. et ALUJA-BANET, T. (1999). 'Analyse en composantes principales'. Paris F : Éditions Lavoisier.

PEARSON, K. (1901). 'On lines and planes of closest fit to systems of points in space'. Philosophical Magazine. 2(11): 559 – 5792.

PRINS, J. et MADER, D. (1997). 'Multivariate Control Charts for Grouped and Individual Observations'. Quality Engineering. 10(1): 49-57.

QUESENBERY, C. P. (1988). 'An SPC approach to compensating a tool-wear process', Journal of Quality Technology, 20(4): 220-229.

QUESENBERY, C. P. (2001). 'The Multivariate Short-Run Snapshot Q Chart', Quality Engineering, 13(4): 679-683.

SIRIPATRAWAN, U. (2008). 'Rapid differentiation between E. coli and Salmonella Typhimurium using metal oxide sensors integrated with pattern recognition'. Elsevier. Sensors and Actuators. B133: 414-419.

RAHEJA D., LLINAS J., (2006). 'Data fusion/Data mining-based architecture for condition-based maintenance', International Journal of Production Research, 44(14).

SOHN, H. CZARNEEKI, J. et FARRAR, C. (2000). 'Structural Health Monitoring Using Statistical Process Control'. Journal of Structural Engineering, 126(11): 1356-1363.

Spad 5.5 Decisia. 1996-2002. www.decisia.fr

Statistica 6.0 Stat Soft, Inc. 1984-2001. www.statsoft.com

SUBHASH SHARMA (1996). 'Applied Multivariate techniques'. John Wiley & Sons Inc. New York, 493p

SULLIVAN, J. et WOODALL, W. (2000). 'Change-point detection of mean vector or covariance matrix shifts using multivariate individual observations'. IIE Transactions. **32**:537-549.

SURTIHADI, J. RAGHAVACHARI, M. et RUNGER, G. (2004). 'Multivariate Control Charts for Process Dispersion'. International Journal of Production Research, **42**(15): 2993-3009.

TONG, F. TSO, S. et HUNG, M. (2006). 'Impact-acoustics-based health monitoring of tile-wall bonding integrity using principal components analysis'. Journal of Sound and Vibration. 294: 329-340.

TRANTAFYLLOPOULOS, K. (2006). 'Multivariate Control Charts Based on Bayesian State Space Models'. Quality and Reliability Engineering International. **22**: 693-707.

TRACY, N. YOUNG, J. et MASON, R. (1992). 'Multivariate Control Charts for individual observations'. Journal of Quality Technology. **24**(2): 88-95.

VARGAS, J. (2003). 'Robust Estimation in Multivariate Control Charts for Individual Observations'. Journal of Quality Technology. **35**(4): 367.

VARGAS, J. et LAGOS, J. (2007). 'Comparison of Multivariate Control Charts for Process Dispersion'. Quality Engineering, **19**:191-196.

WHEELER, D. (1995). 'Advanced Topics in Statistical Process Control'. SPC Press, Inc. Knoxville, Tennessee.

WHEELER, D. et CHAMBERS, D. (1992). 'Understanding Statistical Process Control'. SPC Press, Inc. 2^o Edition. Knoxville, Tennessee.

YANG, K. et TREWN, J. (2004). 'Multivariate Statistical Methods in Quality Management'. McGraw Hill, Inc, 296.

YOON, S. et MACGREGOR, J. (2000). 'Statistical and Causal Model-Based Approaches to Fault Detection and Isolation'. AIChE Journal. **46**(9): 1813-1824.

YOUNG, T. et WINISTORFER, P. (2001). 'The effects of autocorrelation on real-time statistical process control with solutions for forest products manufacturers'. Forest products Journal. **51**(11-12): 70-77.

ANNEXE A

CLASSIFICATION DES DONNÉES

La classification est une technique qui sert à regrouper des observations en groupes les plus homogènes possible, groupant des observations similaires par rapport aux variables et représentant des différences entre les groupes.

Géométriquement, chaque observation peut être représentée comme un point dans l'espace à p dimensions, où p est le nombre des variables. Le regroupement d'observations par simple inspection visuelle devient impossible avec l'augmentation du nombre d'observations et des variables considérées. Les groupes doivent être obtenus par le biais d'algorithmes formalisés et non par des méthodes subjectives ou visuelles.

La classification des données peut être supervisée ou non supervisée. La classification non supervisée des données comprend deux approches : les méthodes hiérarchiques qui produisent une série de partitions ou groupes qui sont emboîtées, et les méthodes de partitionnement qui produisent un nombre fixé a priori de groupes. Nous présentons une brève explication de la classification hiérarchique ascendante qui est utilisée par SPAD, le logiciel employé pour effectuer la classification des données de notre étude.

A.1. LA CLASSIFICATION HIÉRARCHIQUE ASCENDANTE

Cette technique a pour objectif le groupement d'observations ayant un comportement similaire sur un ensemble de variables. Elle est nommée hiérarchique, puisqu'elle crée une suite de partitions emboîtées agrégeant à chaque étape et 2 à 2, les sous-ensembles les plus proches, de sorte que chaque groupe d'une partition est inclus dans la suivante.

Tous les algorithmes de classification hiérarchique ascendante se déroulent de la même façon. Leur principe consiste à construire une suite de partitions en partant de n groupes ou classes, soit les observations elles-mêmes, puis $n-1$ classes, $n-2$ classes..., emboîtées les unes dans les autres ; la partition en $n-k$ classes est obtenue en regroupant deux des classes de la partition en $n-k-1$ classes. Nous recherchons à chaque étape les deux classes les plus proches selon un certain critère, on les fusionne, et on continue d'étape en étape jusqu'à ce qu'il n'y ait plus qu'une seule classe ou groupe.

L'algorithme fournit une hiérarchie de partitions, que l'on représente par un arbre ou dendrogramme contenant l'historique de la classification, analogue à l'organigramme d'une entreprise, qui permet de retrouver $n-1$ partitions. Chaque coupure de l'arbre fournit une partition d'autant plus fine que l'on coupe plus bas, et à chaque partition correspond une valeur numérique représentant le niveau auquel ont lieu les regroupements ; plus l'indice est élevé, plus les parties regroupées sont hétérogènes. Cet indice est nommé le niveau d'agrégation. La proximité entre deux éléments est déterminée par une mesure de divergence (distance) ou par un critère d'agrégation.

A.2. AGRÉGATION SELON LE CRITÈRE DE WARD GÉNÉRALISÉ

La méthode de Ward ne calcule pas les distances entre des éléments d'une partition à une étape donnée. Les groupes sont formés par la maximisation de l'homogénéité interne aux groupes. Vu qu'une partition est d'autant meilleure que les classes créées sont homogènes, et que leur qualité est définie par une forte inertie interclasse (et donc une inertie intra classe faible, puisque la somme des deux inerties est fixe et vaut l'inertie totale de l'ensemble des données), le critère de regroupement de Ward cherche à fusionner les deux classes pour lesquelles la perte de l'inertie est la plus faible.

L'inertie interclasse est la moyenne des carrés des distances des centres de gravité de chaque classe au centre de gravité total, elle est connue aussi comme la somme carrée des

erreurs (ESS), et c'est cette mesure que la méthode cherche à minimiser. Dans le cas d'un nuage de points Y_i de masses respectives m_i , répartis en q classes de centres de gravité respectifs G_q , l'inertie intra classes est définie par :

$$I_{INTRA} = \sum_q \sum_i m_i \|Y_i - G_q\|^2$$

L'inertie intra classes, nulle à l'étape initiale de l'algorithme, s'accroît inévitablement à chaque agrégation. Le principe de l'agrégation selon la variance consiste à rechercher, à chaque étape de l'algorithme, les deux éléments dont l'agrégation produit l'augmentation d'inertie intra classes la plus faible. À chaque étape, deux éléments Y_i et Y_j de masses respectives m_i et m_j sont agrégés en un nouvel élément Y de masse m tel que :

$$Y = (m_i Y_i + m_j Y_j) / (m_i + m_j)$$

$$m = m_i + m_j$$

C'est-à-dire en leur barycentre. On montre (Buroche & Saporta, 1980) que l'accroissement de variance intra classe résultant de cette agrégation vaut :

$$\Delta_{ij} = (m_i m_j / m_i + m_j) \|Y_i - Y_j\|^2$$

Cet accroissement définit la distance entre deux éléments évoquée dans l'algorithme. À chaque étape, on agrège comme indiqué ci-dessus les deux éléments Y_i et Y_j de masses respectives m_i et m_j tels que Δ_{ij} est minimal.

Considérons l'exemple donné par Sharma (1996) dans le tableau A-1 pour illustrer la méthode. Initialement, chaque observation est considérée comme un groupe et la somme carrée des erreurs (ESS) est zéro. Pour la première étape de la classification, cinq groupes seront formés. On procède au calcul de la valeur d'ESS pour les différentes combinaisons d'individus. Le tableau A-2 expose les résultats.

Tableau A-1. Revenus selon les années de scolarité des 6 individus.

Individu	X=Revenu (milliers de \$)	Y=Education (années)
S1	5	5
S2	6	6
S3	15	14
S4	16	15
S5	25	20
S6	30	19

Tableau A-2. Méthode de Ward : solution de 5 groupes

Solution/Groupe	1	2	3	4	5	ESS
1	S1, S2	S3	S4	S5	S6	1.0
2	S1, S3	S2	S4	S5	S6	90.5
3	S1, S4	S2	S3	S5	S6	110.5
4	S1, S5	S2	S3	S4	S6	312.5
5	S1, S6	S2	S3	S4	S5	410.5
6	S2, S3	S1	S4	S5	S6	72.5
7	S2, S4	S1	S3	S5	S6	90.5
8	S2, S5	S1	S3	S4	S6	278.5
9	S2, S6	S1	S3	S4	S5	372.5
10	S3, S4	S1	S4	S5	S6	1.0
11	S3, S5	S1	S4	S4	S5	68.0
12	S3, S6	S1	S4	S4	S6	125.0
13	S4, S5	S1	S4	S3	S5	53.0
14	S4, S6	S1	S4	S3	S6	106.0
15	S5, S6	S1	S2	S3	S4	13.0

On montre ici le calcul de *ESS* pour le groupe *S1* et *S2* : *n* est le nombre d'individus qui interviennent dans la formation du groupe, \bar{X} est la moyenne des X_i et \bar{Y} est la moyenne des Y_i .

$$ESS = \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$ESS = (5 - 5.5)^2 + (6 - 5.5)^2 + (5 - 5.5)^2 + (6 - 5.5)^2 = 1.$$

Comme le montre le tableau A-2, la combinaison des individus '*S3, S4*' possède aussi le même *ESS* minimum que '*S1, S2*', la sélection est réalisée aléatoirement. La prochaine étape consiste en la formation de quatre groupes (Tableau A-3), des 10 combinaisons, la solution numéro 5 présente l'*ESS* plus faible, alors '*S3, S4*' est le groupe résultant.

Tableau A-3. Méthode de Ward : solution de 4 groupes.

Solution/Groupe	1	2	3	4	ESS
1	S1, S2, S3	S4	S5	S6	109.3
2	S1, S2, S4	S3	S5	S6	134.6
3	S1, S2, S5	S3	S4	S6	394.6
4	S1, S2, S6	S3	S4	S5	522.6
5	S1, S2	S3, S4	S5	S6	2.0
6	S1, S2	S3, S5	S4	S6	69.0
7	S1, S2	S3, S6	S4	S5	126.0
8	S1, S2	S4, S5	S3	S6	54.0
9	S1, S2	S4, S6	S3	S5	152.0
10	S1, S2	S5, S6	S3	S4	15.0

Le même processus est répété pour former une solution à trois groupes, puis deux groupes et à un seul en dernière étape. Le dendrogramme obtenu est montré par la figure A.1.

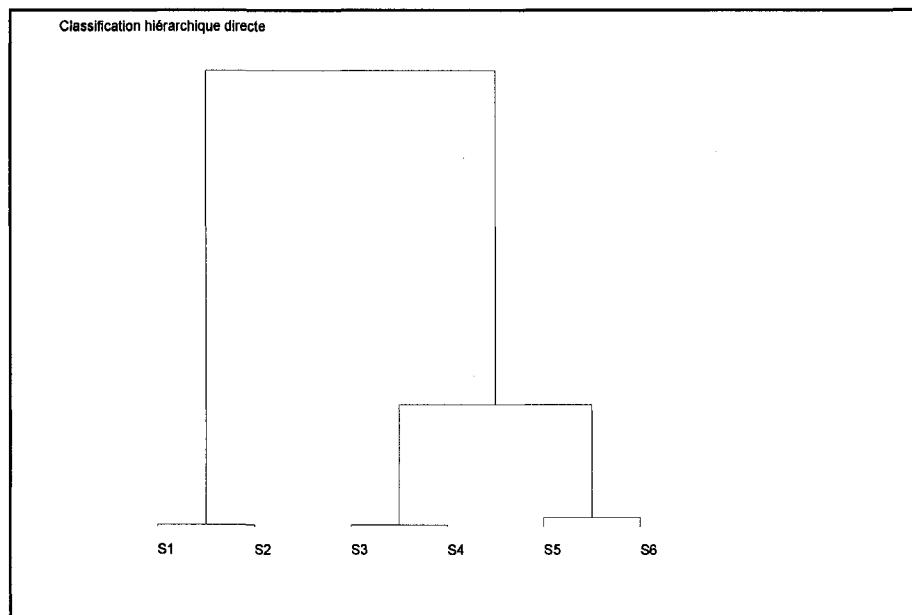


Figure A.1. Dendrogramme obtenu avec le logiciel SPAD.

ANNEXE B

L'ANALYSE EN COMPOSANTES PRINCIPALES

L'analyse en composantes principales (Buroche et Saporta, 1980) est une technique de réduction des données qui vise la formation de nouvelles variables qui sont des combinaisons linéaires des variables originelles. Le nombre maximum des nouvelles variables pouvant être formées est égal au nombre des variables originelles, et les nouvelles variables ne sont pas corrélées entre elles.

Cette méthode d'ordination classique a pour but la recherche d'axes orthogonaux (indépendants) dans un espace réduit à k dimensions ($k \ll p$) qui auront la propriété d'extraire le maximum de la variance (ou inertie projetée) des observations, donnant le meilleur résumé d'un ensemble d'observations situées dans un espace de p variables.

B.1. APPROCHE GÉOMÉTRIQUE

Lorsqu'il n'y a que deux variables, la représentation des données sur un plan est facile, de plus l'examen visuel de l'allure du nuage permet d'étudier l'intensité de la liaison ou relation entre les variables, ainsi que de repérer les observations ou groupes d'observations présentant des caractéristiques voisines.

S'il y a trois variables, l'étude visuelle est encore possible, mais dès que le nombre de p variables devient supérieur, cela devient impossible. Dans notre cas d'étude, une observation est composée par 6 mesures de rugosité de la surface usinée, prises toutes les 240 secondes. Les 50 observations forment un nuage peu visible dans un espace à six dimensions.

Si on veut représenter nos 50 observations sur un plan, il y aura des déformations de la configuration réelle qu'il faudra rendre minimales. Le dessin s'obtient géométriquement en projetant les observations e_1, e_2, \dots, e_{50} sur un plan de projection pour lequel les distances seront en moyenne le mieux conservées, tel que montré par la figure B.1.

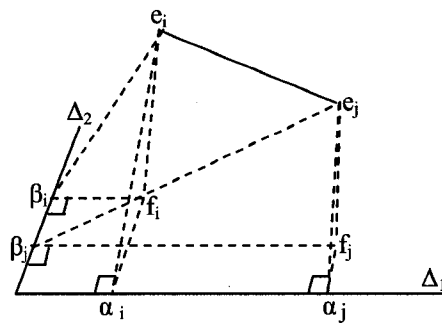


Figure B.1. Plan de projection.

Comme l'opération de projection raccourcit toujours les distances $d(f_i; f_j) \leq d(e_i; e_j)$, il faudra rendre maximale la moyenne des carrés des distances entre les projections f_1, f_2, \dots, f_{50} .

Pour retrouver ce plan que l'on appelle le plan principal, il suffit de trouver deux axes Δ_1 et Δ_2 de façon à ce que les observations projetées f_i puissent être définies par les composantes α_i et β_i . Si les axes sont perpendiculaires, on a :

$$d^2(f_i; f_j) = d^2(\alpha_i; \alpha_j) + d^2(\beta_i; \beta_j)$$

Et alors, la moyenne des carrés des distances entre les points projetés f_i est égale à la moyenne des carrés des distances entre les α_i plus la moyenne des carrés des distances entre les β_i .

L'analyse de composantes principales procède séquentiellement : il s'agit de trouver le Δ_1 qui rend maximale la distance $d^2(\alpha_i; \alpha_j)$ puis Δ_2 , perpendiculaire à Δ_1 , rendant maximale la distance résiduelle $d^2(\beta_i; \beta_j)$. On peut trouver alors p axes perpendiculaires entre eux qui seront les axes principaux du nuage.

En projetant e_i qui avait pour coordonnées initiales $(O_i^1, O_i^2, \dots, O_i^p)$ sur les axes principaux, on obtient de nouvelles coordonnées $(c_i^1, c_i^2, \dots, c_i^p)$. On construit ainsi de nouveaux caractères (variables) que l'on appelle les composantes principales, où chaque composante est une combinaison linéaire des variables initiales :

$$e^k = \varphi_1^k O^1 + \varphi_2^k O^2 + \dots + \varphi_p^k O^p$$

Les coefficients $(\varphi_1^k, \varphi_2^k, \dots, \varphi_p^k)$ forment le k -ième facteur principal φ^k . C'est ainsi que la meilleure représentation des données au moyen de q variables seulement ($q < p$) s'obtient en prenant les q premières composantes principales.

Vu que la distance entre les observations à projeter est la base de la méthode, comment faut-il la calculer ? L'approche analytique présente la procédure à suivre pour faire les calculs requis par l'analyse.

B.2. APPROCHE ANALYTIQUE

La procédure générale comporte l'utilisation des notions de l'algèbre linéaire et du calcul matriciel comme on le décrit maintenant :

- Calcul de la matrice des corrélations variables/variables ;
- Calcul de la matrice de vecteurs propres et du vecteur de valeurs propres ;
- Calcul des coordonnées des observations sur les nouveaux axes ;
- Calcul des corrélations entre les variables et les nouveaux axes.

Le tableau 2-1 du chapitre II montre les données de l'étude, le premier pas de la méthode consiste à *définir une matrice des données* $m \times p$: $m = 50$ observations et $p = 6$ variables.

Remarque : Dans notre cas d'étude, on emploie les six mesures de rugosité comme s'il s'agissait de variables différentes.

Les six mesures peuvent être résumées par leur moyenne et leur écart type. L'observation fictive dont les variables ont par valeur leurs moyennes s'appelle centre de gravité du nuage, g . Dans notre cas, g serait un temps fictif ayant six mesures de rugosité moyennes : 0.1118, 0.1124, 0.1092, 0.1091, 0.1093 et 0.1091. Les écarts types des six mesures sont : 0.0036, 0.0034, 0.0025, 0.0024, 0.0026 et 0.0019.

Les liaisons ou relations entre les variables prises deux à deux sont résumées par leurs covariances et par leur coefficients de corrélation, au nombre de $[p*(p-1)] / 2$, 15 coefficients à calculer. L'ensemble des variances et de covariances est regroupé dans un tableau S appelé matrice de variance des p mesures où le terme situé à l'intersection de la j -ième ligne et de la k -ième colonne est la covariance s_{jk} . Les termes diagonaux sont alors les variances s_j^2 des p mesures.

$$S = \begin{pmatrix} s_1^2 & & & & & \\ & s_{12} & \dots & & & s_{16} \\ & & & & & \vdots \\ & & s_2^2 & \dots & & \\ & & & & & \vdots \\ & & & & & s_6^2 \end{pmatrix}$$

De même, l'ensemble des coefficients de corrélation est regroupé dans la matrice de corrélation R dont les termes diagonaux valent 1 puisque $r(x^j; x^j) = 1$.

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{16} \\ & 1 & & \vdots \\ & & & 1 \end{pmatrix}$$

R et S sont des matrices carrées d'ordre $p=6$, symétriques. On pourra écrire seulement la moitié des termes. Si on note $D_{1/s}$ la matrice diagonale suivante :

$$D_{1/s} = \begin{pmatrix} 1/s_1 & & & 0 \\ & 1/s_2 & & \\ & & & \vdots \\ 0 & & & 1/s_p \end{pmatrix}$$

on a la relation matricielle :

$$R = D_{1/s} S D_{1/s}$$

Cette métrique diagonale, $D_{1/s}$, revient à diviser chaque caractère par son écart type, donc la distance entre deux observations ne dépend plus des unités de mesure et chaque variable a la même importance quelle que soit sa dispersion (variance 1.) Si X est le tableau à 50 lignes et 6 colonnes des données centrées, on a les relations matricielles suivantes :

$$S = X^t D_{1/s} X$$

Ainsi, on obtient la *matrice de corrélation* R de notre cas d'étude, montrée au tableau A1. Les 15 coefficients de corrélation linéaire indiquent la force de la liaison entre deux variables, ils prennent des valeurs entre -1 et 1. Pour ce cas, les liaisons sont positives,

plus la valeur est proche de l'unité, plus la relation est forte et croissante. Sur le tableau A1 on voit que les corrélations sont toujours supérieures au 30%.

Tableau B-1. Matrice des corrélations R .

	O ₁	O ₂	O ₃	O ₄	O ₅	O ₆
O ₁	1,00					
O ₂	0,31	1,00				
O ₃	0,52	0,39	1,00			
O ₄	0,60	0,33	0,75	1,00		
O ₅	0,52	0,38	0,70	0,66	1,00	
O ₆	0,34	0,42	0,39	0,32	0,34	1,00

La matrice des valeurs tests (tableau B-1) du logiciel sert à confirmer que les liaisons sont significatives, puisqu'aucune valeur n'est inférieure à 2 nous sommes en présence de variables linéairement dépendantes. On pourrait même établir une hiérarchie dans la force des liaisons.

Tableau B-2. Matrice des valeurs tests.

	O ₁	O ₂	O ₃	O ₄	O ₅	O ₆
O ₁	99,99					
O ₂	2,25	99,99				
O ₃	4,11	2,93	99,99			
O ₄	4,86	2,44	6,88	99,99		
O ₅	4,03	2,81	6,17	5,67	99,99	
O ₆	2,54	3,16	2,90	2,35	2,54	99,99

Puisque l'inertie totale du nuage de points est la moyenne des carrés des distances des m observations au centre de gravité (origine), elle mesure d'une certaine manière l'éloignement des points par rapport à leur centre de gravité. Autrement dit, la dispersion globale du nuage. Alors, on montre que la moyenne minimale des distances entre les points originaux e_i et les points projetés f_i de la figure B.1 peut être obtenue quand l'inertie du nuage projeté est maximale.

En conséquence, on peut interpréter que le plan principal du nuage d'observations est celui qui rend maximale l'inertie de l'ensemble d'observations projetées sur lui et que les axes principaux passent par le centre de gravité, dont l'inertie est la trace de R , égale au nombre de mesures p .

Cela équivaut à rendre maximale l'inertie des projections qui vaut $\sum p_i c_i^2$, où les c_i sont les mesures algébriques des projections des e_i sur Δ_i , les c_i seront les plus dispersés possible, fait qui garantit une perte d'information minimale. En d'autres termes, c est une combinaison linéaire des x_i de variance maximale.

Pour la recherche de composantes, d'axes et de facteurs principaux, il nous faut calculer la *matrice de vecteurs propres*, les *coordonnées des observations sur les nouveaux axes* et les *corrélations des échantillons avec les nouveaux axes*. Bouroche & Saporta (1980) expliquent les calculs matriciels.

L'ACP remplace les valeurs initiales des données par des variables non corrélées de variance maximale et d'importance décroissante. La somme des valeurs propres est une constante égale à l'inertie totale.

Le logiciel employé (SPAD version 5.5) utilise la métrique $D_{1/s}^2$, ce qui revient à centrer et réduire les six mesures de rugosité. Les facteurs principaux s'obtiennent en diagonalisant la matrice de corrélation R .

La somme de la variance associée à la projection des observations sur les axes, mieux connue comme la somme des valeurs propres, est égale au nombre de mesures $p = 6$ à cause de la métrique choisie. Comme les deux premières valeurs propres représentent environ 72% de l'inertie (Voir tableau 2-7, Chapitre 2), alors nous pouvons résumer les données par les deux premières composantes principales. Comme les variables avaient des corrélations bien significatives (plus de 30%, voir tableau B-1 et B-2) la réduction peut être considérée efficace.

Ayant pour objectif la représentation graphique des données, deux axes semblent le plus convenables car ils nous permettent de reconstituer près du $\frac{3}{4}$ de l'information contenue dans les données. Une classification hiérarchique appliquant la méthode d'agrégation de Ward sera effectuée à partir des deux premières coordonnées factorielles sauvegardées pour la typologie afin, si possible, d'identifier l'état de fonctionnement normal de l'outil.

Les intervalles laplaciens d'Anderson nous confèrent la certitude de trouver un axe qui interceptera une inertie comprise entre les bornes avec 95% de confiance. Selon le tableau B-3, les deux premiers axes intercepteront la plupart de l'inertie.

Tableau B-3. Intervalles de confiance pour les λ_i .

Numéro	Borne inférieure	Valeur propre	Borne supérieure
1	2,0721	3,3866	4,7011
2	0,5660	0,9250	1,2841
3	0,3660	0,5982	0,8304
4	0,3197	0,5225	0,7254
5	0,2059	0,3365	0,4671

Le tableau B-4A présente les coordonnées des six variables sur chacun des axes factoriels. Pour le premier axe, ils ont tous le même signe et sont compris entre 0.58 et 0.86, ce sont des valeurs élevées donc le nuage des points ne sera guère centré. Le fait le plus marquant de cette analyse est porté par la première composante, ce facteur reflète pratiquement de manière exclusive le phénomène d'usure. Le nuage d'observations montre le fait plus visiblement.

Tableau B-4A. Coordonnées des échantillons.

Libellé de la variable	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
O ₁	0,74	-0,17	0,32	-0,55	-0,08
O ₂	0,59	0,60	-0,49	-0,24	0,04
O ₃	0,86	-0,19	-0,10	0,23	0,19
O ₄	0,85	-0,31	-0,02	0,05	0,29
O ₅	0,83	-0,20	-0,15	0,20	-0,46
O ₆	0,58	0,61	0,47	0,24	0,00

Tableau B-4B. Corrélations des échantillons avec les facteurs.

Libellé de la variable	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
O ₁	0,74	-0,17	0,32	-0,55	-0,08
O ₂	0,59	0,60	-0,49	-0,24	0,04
O ₃	0,86	-0,19	-0,10	0,23	0,19
O ₄	0,85	-0,31	-0,02	0,05	0,29
O ₅	0,83	-0,20	-0,15	0,20	-0,46
O ₆	0,58	0,61	0,47	0,24	0,00

Tableau B-4C. Anciens axes unitaires.

Libellé de la variable	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
O ₁	0,40	-0,18	0,42	-0,77	-0,13
O ₂	0,32	0,62	-0,63	-0,33	0,06
O ₃	0,47	-0,19	-0,13	0,32	0,32
O ₄	0,46	-0,32	-0,03	0,07	0,51
O ₅	0,45	-0,21	-0,19	0,28	-0,79
O ₆	0,32	0,64	0,61	0,34	0,00

Comme il s'agit d'une analyse normée, les coordonnées des six mesures (tableau B-4A) coïncident avec la projection des observations sur l'axe factoriel de même rang (tableau B-4B).

Les anciens axes unitaires montrés par le tableau B-4C sont les coefficients de la liaison linéaire entre les six mesures de rugosité et les axes. On pourrait obtenir la contribution de chacune des variables à la formation des axes en élevant au carré chaque composante du vecteur φ . L'addition de toutes les contributions sur un axe est égale à 1.

Les composantes de φ donnent aussi les combinaisons linéaires des six mesures de rugosité d'origine qui définissent les nouvelles variables de variance maximale. Par exemple, pour la première composante :

$$\Psi_1 = 0.40Z_{O1} + 0.32Z_{O2} + 0.47Z_{O3} + 0.46Z_{O4} + 0.45Z_{O5} + 0.32Z_{O6}$$

où Z_{O_i} indique la valeur de la mesure centrée et réduite. Ces composantes définissent aussi la projection des anciens axes unitaires sur les nouveaux axes obtenus.

L'analyse visuelle du nuage est décrite dans le chapitre 2 de cet ouvrage.

B.3. RANDOMISATION DES ÉCHANTILLONS

Cette analyse sert à vérifier la pertinence de l'ACP. Vu que les pièces coupées sont sélectionnées au hasard pendant l'inspection, que l'échantillonnage les choisit de façon interchangeable et que la méthode exige la formation de vecteurs d'observations, nous allons introduire ce facteur aléatoire dans l'analyse en randomisant l'ordre des pièces. L'expérience comporte plusieurs répétitions, ici on en présente une à titre d'exemple.

On a assigné un nombre aléatoire à chacune des mesures de rugosité, ensuite on les a triés en ordre décroissant utilisant ce nombre pour garantir l'ordonnancement au hasard (Voir tableau B-5) et subséquemment, on a effectué l'ACP et la classification de Ward (Figure B.2).

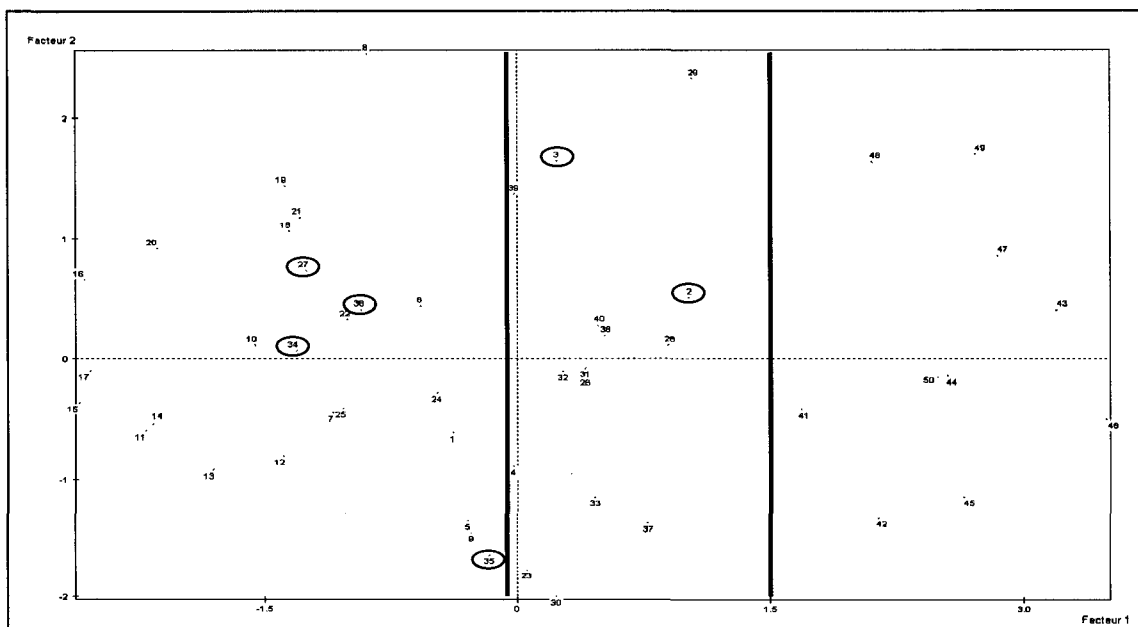


Figure B.2. Analyse des observations. Randomisation.

Les résultats sont les suivants : les deux premiers axes comportent la plupart de l'inertie (61.49%) et le nuage des individus (Figure B.2) exhibe le même comportement que l'analyse sur les données d'origine. On identifie trois intervalles pour les observations (1-25 ; 26-40 ; 41-50), dont les données numéro 2, 3, 27, 34, 35 et 36 sont mal classées.

Tableau B-5. Données triées aléatoirement.

<i>t</i>	O ₁	O ₂	O ₃	O ₄	O ₅	O ₆
1	0,1103	0,1094	0,1085	0,1074	0,1142	0,1080
2	0,1108	0,1106	0,1105	0,1118	0,1132	0,1119
3	0,1105	0,1071	0,1086	0,1122	0,1105	0,1143
4	0,1111	0,1080	0,1136	0,1076	0,1126	0,1072
5	0,1123	0,1119	0,1103	0,1079	0,1071	0,1077
6	0,1084	0,1081	0,1100	0,1071	0,1092	0,1140
7	0,1105	0,1081	0,1107	0,1078	0,1060	0,1086
8	0,1062	0,1069	0,1054	0,1149	0,1118	0,1106
9	0,1144	0,1097	0,1058	0,1053	0,1145	0,1080
10	0,1056	0,1065	0,1142	0,1094	0,1083	0,1054
11	0,1053	0,1072	0,1096	0,1040	0,1111	0,1074
12	0,1083	0,1095	0,1099	0,1062	0,1081	0,1079
13	0,1081	0,1088	0,1112	0,1075	0,1064	0,1045
14	0,1082	0,1076	0,1105	0,1071	0,1040	0,1064
15	0,1066	0,1083	0,1062	0,1075	0,1085	0,1041
16	0,1087	0,1054	0,1067	0,1109	0,1048	0,1047
17	0,1071	0,1079	0,1041	0,1056	0,1090	0,1082
18	0,1084	0,1079	0,1065	0,1119	0,1076	0,1086
19	0,1051	0,1085	0,1088	0,1113	0,1062	0,1114
20	0,1056	0,1063	0,1084	0,1111	0,1087	0,1053
21	0,1070	0,1079	0,1073	0,1077	0,1060	0,1158

Tableau B-5. Données triées aléatoirement. (Suite et fin).

22	0,1062	0,1120	0,1074	0,1075	0,1064	0,1141
23	0,1084	0,1159	0,1126	0,1082	0,1088	0,1065
24	0,1089	0,1077	0,1109	0,1083	0,1149	0,1068
25	0,1082	0,1096	0,1083	0,1079	0,1121	0,1070
26	0,1079	0,1155	0,1096	0,1112	0,1086	0,1111
27	0,1071	0,1091	0,1062	0,1100	0,1102	0,1094
28	0,1111	0,1097	0,1106	0,1120	0,1168	0,1077
29	0,1097	0,1098	0,1100	0,1162	0,1078	0,1160
30	0,1157	0,1118	0,1099	0,1063	0,1086	0,1082
31	0,1093	0,1101	0,1097	0,1084	0,1161	0,1108
32	0,1073	0,1117	0,1117	0,1084	0,1124	0,1119
33	0,1139	0,1110	0,1102	0,1090	0,1116	0,1074
34	0,1078	0,1091	0,1097	0,1094	0,1068	0,1080
35	0,1085	0,1160	0,1095	0,1061	0,1091	0,1097
36	0,1094	0,1088	0,1067	0,1099	0,1100	0,1091
37	0,1113	0,1156	0,1098	0,1098	0,1119	0,1075
38	0,1105	0,1100	0,1122	0,1100	0,1093	0,1126
39	0,1084	0,1066	0,1096	0,1127	0,1153	0,1092
40	0,1093	0,1102	0,1124	0,1095	0,1099	0,1133
41	0,1102	0,1109	0,1175	0,1092	0,1130	0,1127
42	0,1126	0,1157	0,1167	0,1114	0,1094	0,1100
43	0,1142	0,1106	0,1149	0,1119	0,1175	0,1160
44	0,1128	0,1152	0,1104	0,1137	0,1170	0,1110
45	0,1158	0,1146	0,1123	0,1101	0,1143	0,1126
46	0,1108	0,1110	0,1138	0,1173	0,1112	0,1132
47	0,1131	0,1129	0,1119	0,1149	0,1151	0,1147
48	0,1179	0,1156	0,1097	0,1130	0,1157	0,1143
49	0,1101	0,1102	0,1135	0,1138	0,1175	0,1176
50	0,1137	0,1113	0,1129	0,1103	0,1174	0,1140

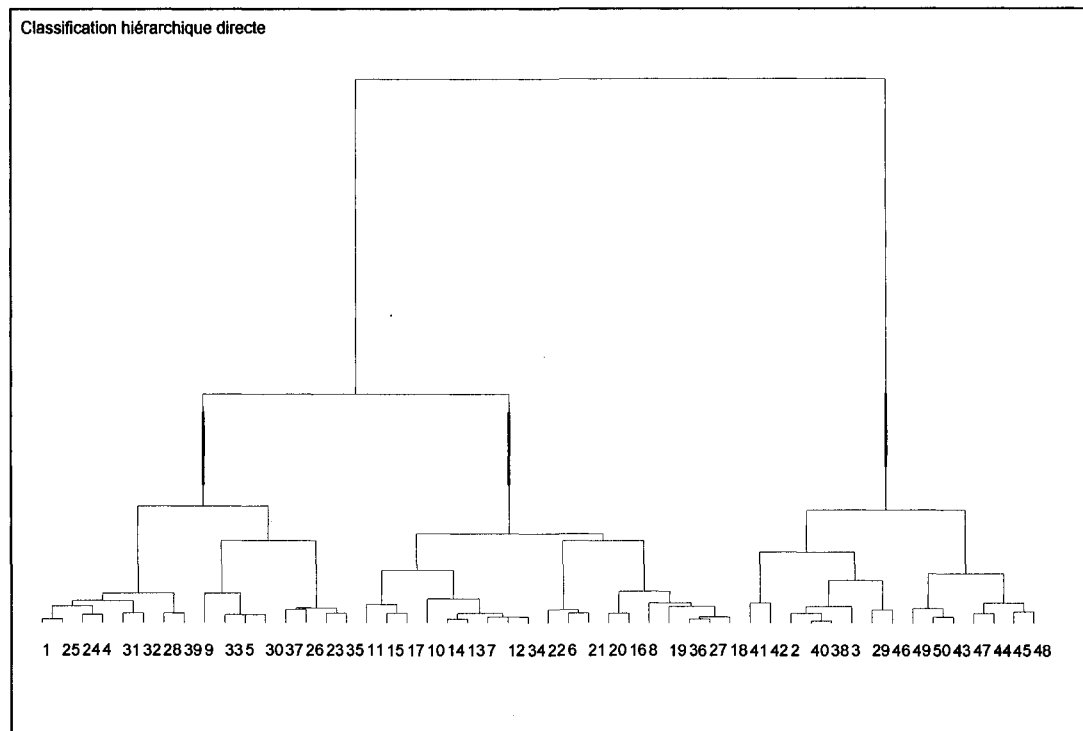


Figure B.3. Classification hiérarchique directe. Randomisation.

Il est donc évident que sous permutation des $O_i(t)$, l'état de fonctionnement normal de l'outil correspond aux observations 1-25 et qu'une zone intermédiaire se trouve au milieu démarquant l'état d'usure avancé de l'outil, de sorte que l'on a une validation des conclusions.

Quelques autres permutations aléatoires des colonnes, une différente pour chaque ligne, ont donné des résultats voisins à la suite ACP+Ward, toujours sur les deux premiers axes factoriels. On trouve au tableau B-6 les principaux résultats d'autres permutations effectuées. Dans la plupart des cas, on identifie la suite des premières 25 observations. De cette façon simple, on peut croire que la partition en trois groupes dont le premier comprend les observations 1 à 25 n'est pas un artéfact dû à l'échantillonnage. Le premier groupe s'interprète alors dans le sens qu'il constitue l'état de fonctionnement normal de

l'outil; le groupe intermédiaire serait un état de passage entre le premier et le troisième; qui serait, lui, l'état « usé » de l'outil.

Tableau B-6. Résultats des permutations effectuées.

Expérience	% Variance (expliquée par 2 axes)	Intervalle des observations
P	68,26	1-27
Q	60,74	1-25
O	59,87	1-25
R	58,02	1-27

Il va sans dire qu'une telle ACP diffère de celles qu'on peut trouver habituellement en analyse de données, puisque le plus souvent (mais pas toujours) on y trouve des tableaux 'sujets x variables', alors qu'ici il s'agit d'une matrice dont les colonnes sont des échantillons de la même mesure de rugosité. Mais comme toujours en ACP, ce qui justifie l'utilisation d'un tableau rectangulaire dans l'extraction de facteurs propres est que la décomposition en composantes principales dans un espace de basse dimension donne des résultats aisément interprétables.

Ici, le premier axe est clairement identifié comme un axe temporel. La classification de Ward sur les deux coordonnées principales présente invariablement trois groupes à peu

près composés des mêmes temps d'usure. Au fond, c'est cette cohérence qui est intéressante.

On pourrait voir que les classifications de Ward, validées de la même façon que celle décrite plus haut par permutations aléatoires des colonnes, ne donnent pas des groupes cohérents de validation en validation, et qu'une classification bien identifiée en trois groupes ne se présente jamais.

ANNEXE C

ANALYSE PRÉLIMINAIRE DES DONNÉES

Cet annexe présente une analyse de la série de données originelles $O_i(t)$, $i=1,\dots,6$, présentée au tableau 2.1. Aucun traitement n'a pas été appliqué sur ces données, nous allons tout simplement présenter les résultats de différentes cartes de contrôle unidimensionnelles afin d'obtenir des renseignements préliminaires à propos du processus.

D'abord, on trace les données de rugosité moyennes versus le temps. On remarque une tendance positive dans la figure C.1 puisque les valeurs de rugosité moyenne augmentent dans le temps, tel que mesuré par le modèle linéaire associé (qui au fait n'est valide qu'en première approximation). La performance de l'outil se dégrade et ceci est une indication du phénomène d'usure qui affecte l'outil employé dans le processus de coupe des pièces.

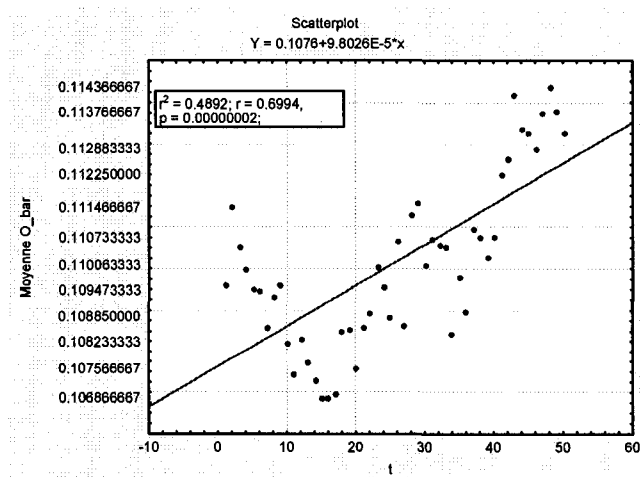


Figure C.1. Les rugosités moyennes \bar{O} versus t .

Ensuite, on emploie les cartes de contrôle unidimensionnelles afin de surveiller la performance de l'outil sur la ligne de production. Les cartes, obtenues avec le logiciel Statistica (2001), sont analysées pour l'obtention des renseignements sur le phénomène d'usure présente dans le processus. Le lecteur peut consulter l'ouvrage de Montgomery (2005) pour plus de détails sur leur construction.

La combinaison des cartes \bar{X} et R est utilisée parce que la taille de l'échantillon $n = 6$ est inférieure à 10 (on dispose de six mesures de rugosité prises toutes les 240 secondes.) Dans la carte de contrôle \bar{X} qui s'expose dans la figure C.2, on retrouve deux observations au-dessous de la limite inférieure au début du processus (observations numéro 15-16) et cinq observations au-dessus de la limite supérieure à la fin (observations numéro 43-44 et 47-49).

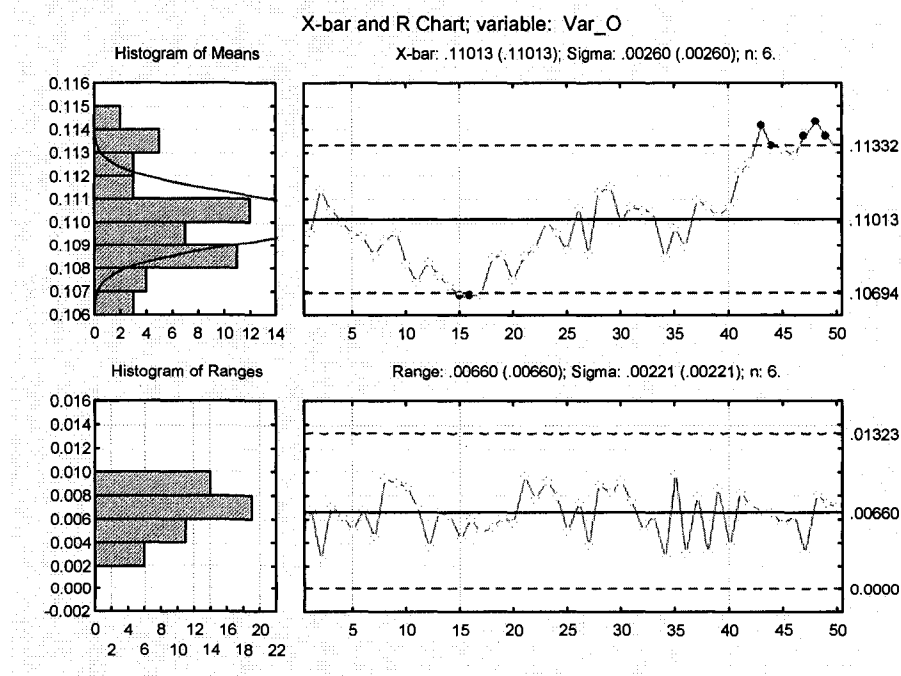


Figure C.2. Cartes de contrôle \bar{X} -bar et R .

Les observations présentent principalement deux tendances: la moyenne du processus expérimente une diminution et ensuite une croissance. De plus, les observations qui se trouvent à l'intérieur des limites présentent un comportement autre qu'aléatoire (Voir le tableau C-1), et presque toutes les observations sont suspectes, excepté pour l'intervalle d'observations 26-36, tel qu'indiqué au tableau C-1 qui présente les règles de Western Electric qui sont invoquées pour justifier notre jugement ici.

Tableau C-1. Résumé des règles de Western Electric concernant la carte *X-bar* de la figure C.2

Règles de Western Electric	Intervalle d'observations
9 observations consécutives positionnées du même côté de la ligne centrale	4-12
	13-21
	37-45
2 de 3 observations consécutives dans la zone A ou au-delà	11-13
	14-16
	41-43
	44-46
	47-49
4 de 5 observations consécutives dans la zone B ou au-delà	9-13
	14-18
	19-23
	40-44
	45-49
8 observations consécutives au-delà de la zone C	10-17
	41-48

Selon la carte R (Figure C.2), la dispersion du processus est sous contrôle statistique. On peut considérer la tendance positive que l'on retrouve à partir de l'observation 37 dans la carte $X\text{-bar}$ comme une manifestation du phénomène d'usure. Cependant, il est évident qu'une étude approfondie est nécessaire car il y a beaucoup d'observations hors de contrôle.

Le type de distribution des données employées dans la construction des cartes de contrôle peut affecter les limites calculées (Wheeler, 1995 ; Montgomery, 2005). La carte $X\text{-bar}$ de la figure C.3 prend compte des changements lorsque la distribution n'est pas normale. On observe un déplacement des limites vers le haut : la limite inférieure passe de 0.1069 à 0.1073 et la limite supérieure passe de 0.1133 à 0.1135. Or, le comportement des observations est similaire à celui qu'on retrouve dans la carte de contrôle précédente (Figure C.2). Ceci suggère que la normalité de la distribution n'est pas un facteur important pour la construction de cette carte.

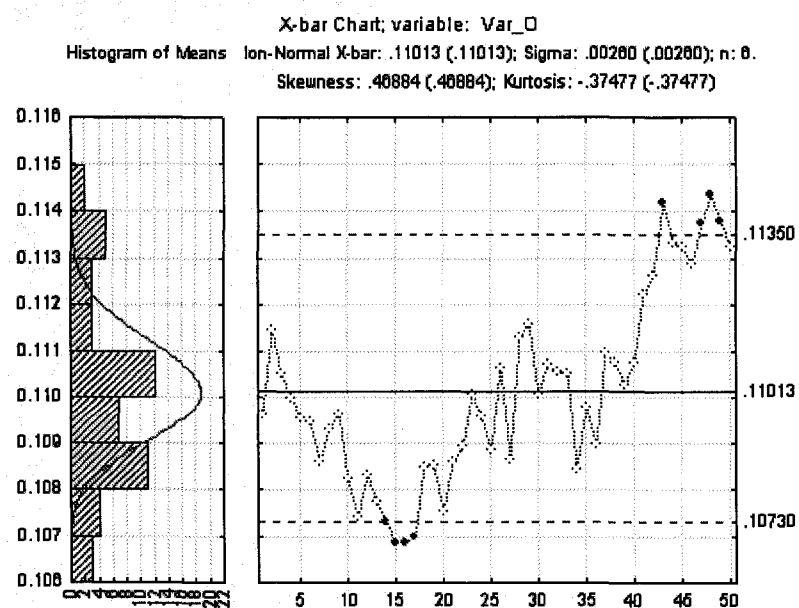


Figure C.3. Carte $X\text{-bar}$ pour une distribution autre que normale.

Par la suite, nous allons étudier les données du processus en utilisant des cartes de contrôle qui ont été conçues spécialement pour la détection de petits changements permanents dans la moyenne du processus et des tendances : la carte de moyenne mobile '*MA X-bar*', la carte de moyenne mobile pondérée exponentiellement '*EWMA X-bar*' et la carte des sommes cumulatives '*CuSum X*' sont présentées dans les figures C.4, C.5 et C.6 respectivement.

Dans la carte de moyenne mobile '*MA X-bar*' de la figure C.4, les observations 11 à 23 se trouvent au-dessous de la limite inférieure et une tendance positive apparaît jusqu'à ce que les observations 43 à 50 se retrouvent au-dessus de la limite supérieure.

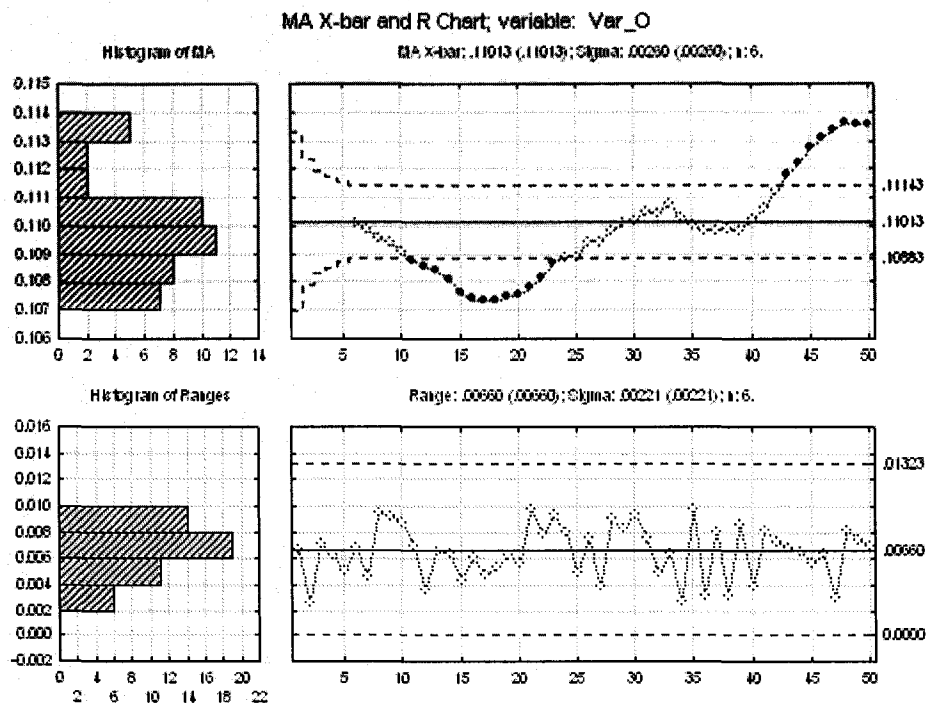
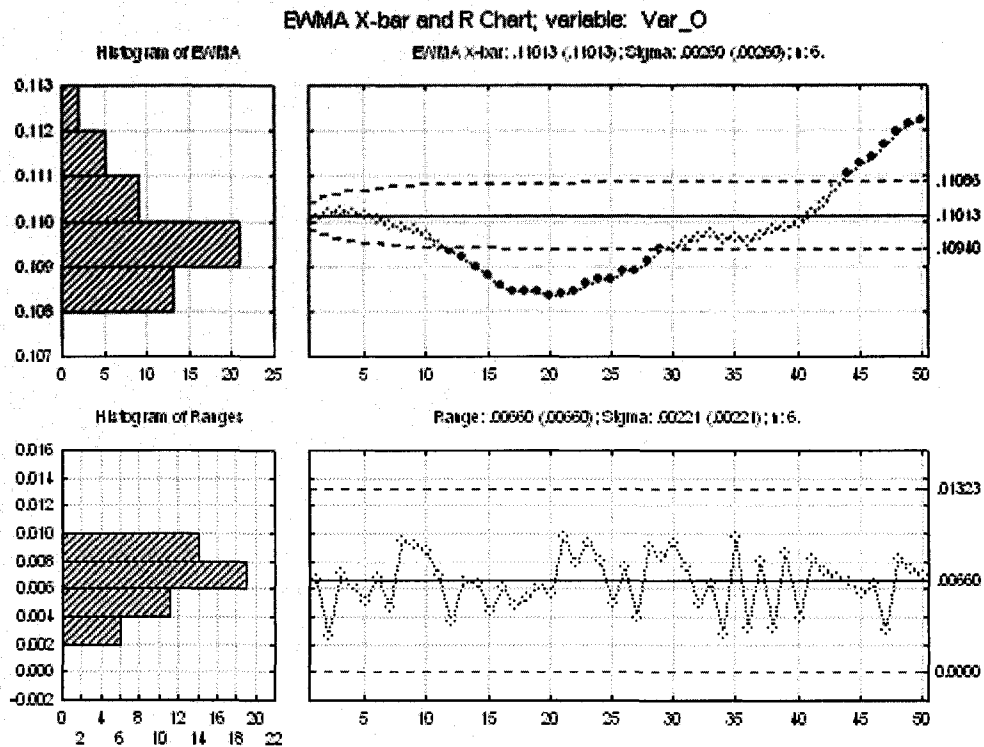


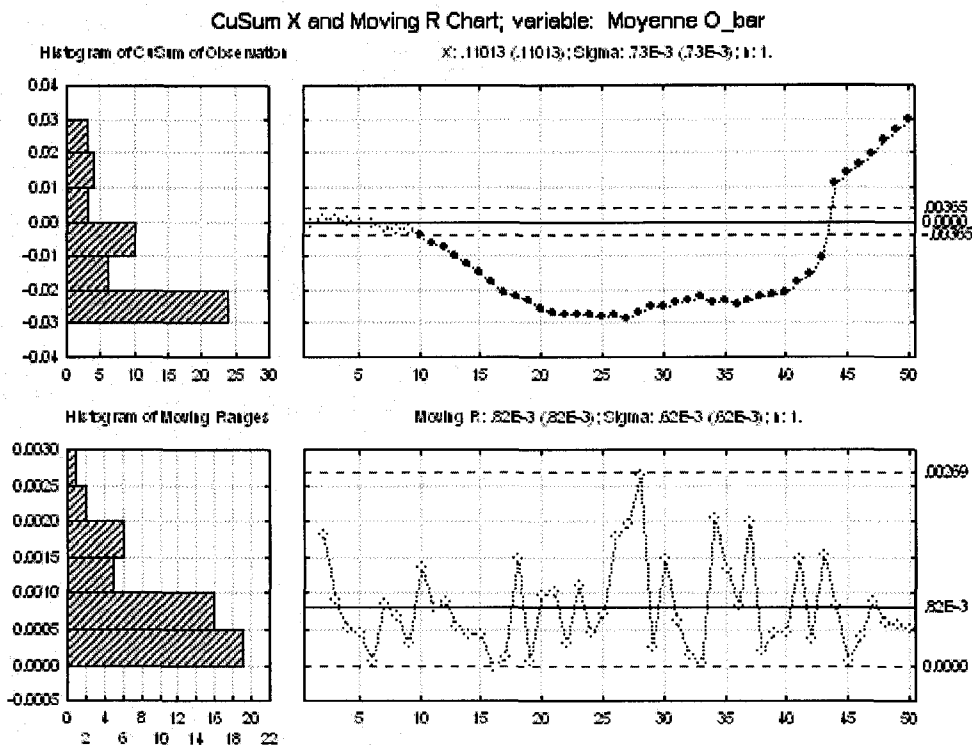
Figure C.4. Carte *MA X-bar*.

Figure C.5. Carte *EWMA X-bar*.

La carte des moyennes mobiles pondérées exponentiellement '*EWMA X-bar*' de la figure C.5 présente un comportement similaire à celui de la carte de moyenne mobile '*MA X-bar*', car les observations numéro 11-29 se trouvent au-dessous de la limite inférieure et une tendance positive apparaît jusqu'à ce que les observations numéro 44-50 se retrouvent au-dessus de la limite supérieure.

Enfin, la carte des sommes cumulatives '*CuSum X*' montre les observations 11 à 43 au-dessous de la limite inférieure et tout de suite, les observations numéro 44-50 qui se trouvent au-dessus de la limite supérieure (voir la figure C.6).

Le tableau C-2 présente un résumé des résultats de l'ensemble des cartes de contrôle analysées.

Figure C.6. Carte *CuSum X*.Tableau C-2. Comparaison des résultats des cartes de contrôle construites pour la série des données $O_i(t)$, $i=1, \dots, 6$.

Carte de contrôle	Intervalle d'observations hors de contrôle
<i>X-bar</i>	15-16, 43-44 et 47-49
<i>X-bar non normal</i>	14-17 et 43, 47-49
<i>MA X-bar</i>	11-23 et 43-50
<i>EWMA X-bar</i>	11-29 et 44-50
<i>CuSum X</i>	11-43 et 44-50

Suite à l'analyse des différentes cartes, on peut tirer les renseignements suivants :

- Les observations hors de contrôle au début du processus indiquent une diminution de la moyenne du processus qu'il faut investiguer ;
- La normalité de la distribution des données n'est pas un facteur déterminant pour la performance des cartes de contrôle ;
- Presque toutes les observations sont suspectes d'après la carte \bar{X} , excepté pour l'intervalle des observations 26-36. Il est clair qu'il existe un nombre élevé d'alarmes dont il faut chercher la cause. Le grand nombre d'observations hors de contrôle que l'on trouve dans le reste des cartes appuie cette conclusion ;
- Le développement d'une tendance positive, à partir de $t=37$ d'après la carte \bar{X} , signale l'apparition du phénomène d'usure. Cependant, une étude plus approfondie est nécessaire afin de définir les observations qui marquent le début de cet état, puisque le reste des cartes n'est pas très utile à ce propos ;
- Les observations qui se trouvent hors de contrôle à la fin du processus révèlent que l'état usagé de l'outil se trouverait à partir de l'observation $t=43$ selon la plupart des cartes.

Le grand nombre d'alarmes que l'on trouve dans les cartes de contrôle analysées nous donne l'occasion d'effectuer une recherche plus approfondie. La littérature sur le sujet nous informe que ce comportement peut avoir lieu à cause de la présence du phénomène d'autocorrélation parmi les données (Charnes ,1995 ; Wheeler, 1995 ; Montgomery, 2005 ; Mason et Young, 2007).

Notre travail propose une technologie pour la surveillance de l'usure spécialement conçue pour le traitement de données corrélées. Nous allons employer les données que l'on vient d'analyser pour « *illustrer notre nouvelle méthode* », même si celles-ci ne sont

pas parfaites pour l'application, parce qu'elles ont l'avantage d'être réelles. La décision de mener une étude de ce genre dépendra toujours du « jugement de l'analyste » et des objectifs de la recherche. C'est ainsi que les résultats que l'on vient d'obtenir concernant la détection de l'usure serviront pour valider la performance de la méthodologie proposée et comparer les résultats.