

NOTES DE STATISTIQUE ET D'INFORMATIQUE

96/3

CARTES DE CONTRÔLE :
combinaison de résultats,
données corrélées ou multivariées

R. PALM

Faculté universitaire des
Sciences agronomiques

Centre de Recherches
agronomiques

GEMBLoux

(Belgique)

CARTES DE CONTRÔLE : **combinaison de résultats,** **données corrélées ou multivariées**

R. PALM⁽¹⁾

RÉSUMÉ

Dans cette note, on décrit d'abord les cartes de contrôle sur les moyennes mobiles, sur les résultats d'un lissage exponentiel et sur les sommes cumulées. Ensuite, on examine les cartes de contrôle pour des données non indépendantes et pour des données multivariées.

Ces cartes sont illustrées par des exemples numériques traités par le logiciel Minitab.

SUMMARY

This note first describes moving average, exponentially weighted moving average and cumulative sum control charts. Then, control charts for autocorrelated and for multivariate data are examined.

Minitab outputs are given for these control charts.

1. INTRODUCTION

Dans une première note [PALM, 1996], nous avons décrit quelques cartes de contrôle classiques, valables pour les contrôles par mesures et pour les contrôles par comptages. Ces cartes, qui correspondent aux cartes de SHEWHART proprement dites, sont basées sur l'examen des caractéristiques de chacun des échantillons pris isolément. De plus, pour l'établissement de ces cartes, nous avons supposé que les observations successives étaient indépendantes. Enfin, nous n'avons considéré que des cartes basées sur l'observation d'une seule caractéristique.

⁽¹⁾ Chargé de cours associé à la Faculté universitaire des Sciences agronomiques de Gembloux.

Nous examinons, dans la présente note, un ensemble de cartes de contrôle moins classiques. Nous consacrons d'abord un paragraphe aux cartes de contrôle combinant les observations réalisées sur des échantillons successifs (paragraphe 2). Ensuite, nous considérons le cas de données non indépendantes (paragraphe 3) et des données multivariées (paragraphe 4). Nous terminons par quelques informations complémentaires (paragraphe 5).

Les principes généraux relatifs à ces cartes sont décrits, notamment, par MONTGOMERY [1991] et des références spécifiques seront données au fur et à mesure de l'étude des différentes cartes.

Les exemples numériques, présentés à titre d'illustration, sont traités par le logiciel Minitab en utilisant principalement les commandes MACHART et EWMACHART et la macroprocédure %CUSUM. Les cartes de contrôle ont été réalisées à partir de l'interface à menu déroulant et nous avons gardé les options proposées par défaut pour les titres et les dénominations des axes qui apparaissent, par conséquent, en anglais dans les figures. Des informations concernant ces commandes sont données dans le manuel d'utilisation [X, 1994]. Des résultats pratiquement identiques auraient pu être obtenus par d'autres logiciels et notamment par SAS, par les procédures MACONTROL et CUSUM [SAS, 1995].

2. COMBINAISON DES RÉSULTATS DE PLUSIEURS ÉCHANTILLONS

2.1. Principe

Lors de l'étude des cartes de contrôle de la moyenne [PALM, 1996], nous avons insisté sur la difficulté de mettre en évidence des dérèglements du procédé de fabrication, du moins lorsque la modification de la moyenne n'est pas très importante. Les courbes d'efficacité ont montré, par exemple, que, pour une taille de l'échantillon égale à 5 et un risque de première espèce égal à 0,002, la probabilité de ne pas détecter une modification de la moyenne égale à la moitié de l'écart-type est égale à 0,98 et qu'il faudra, en moyenne, prélever 41 échantillons successifs avant de se rendre compte qu'il y a effectivement eu modification de cette moyenne.

Nous avons vu également que, pour diminuer le risque de ne pas mettre en évidence une modification de la moyenne sans augmenter la taille des échantillons et sans augmenter exagérément le risque d'une fausse alerte, une solution consiste à introduire dans les cartes de contrôle une série de règles complémentaires. Ces règles tiennent compte de la position des points successifs sur la carte. Il s'agit en fait d'une forme de combinaison des résultats de plusieurs échantillons successifs.

D'autres formes de combinaison peuvent être envisagées et nous allons examiner successivement les cartes de contrôle basées sur les moyennes mobiles, sur les lissages exponentiels et sur les sommes cumulées. Ces cartes de contrôle peuvent être utilisées à la fois pour des échantillons d'effectifs supérieurs à l'unité ou pour des données individuelles et constituent des alternatives intéressantes

aux cartes de SHEWHART, surtout pour la mise en évidence de faibles modifications de la moyenne. Elles sont particulièrement appropriées pour le contrôle de la matière première ou du produit fini résultant de procédés chimiques continus. L'effet de lissage caractérisant ces cartes traduit alors, dans une certaine mesure, l'effet lié au mélange et à l'homogénéisation qui a lieu au cours des procédés industriels ultérieurs [GRANT et LEAVENWORTH, 1980].

L'utilisation de ces cartes suppose, comme pour les cartes de SHEWHART, que les observations successives sont indépendantes et distribuées selon une loi normale d'écart-type σ_0 .

2.2. Cartes de contrôle sur les moyennes mobiles

Plutôt que de faire le contrôle sur les valeurs individuelles x_i ou sur les valeurs moyennes \bar{x}_i , on peut effectuer le contrôle sur des *moyennes mobiles*⁽²⁾, \bar{x}'_i , calculées sur les résultats des q échantillons les plus récents.

Dans le cas de données individuelles, la valeur \bar{x}'_i est la moyenne arithmétique de l'observation x_i et des $q - 1$ observations précédentes :

$$\bar{x}'_i = (x_{i-q+1} + \dots + x_{i-1} + x_i)/q,$$

tandis que dans le cas d'échantillons d'effectifs n_i , on remplace les valeurs individuelles par les valeurs moyennes correspondantes :

$$\bar{x}'_i = (\bar{x}_{i-q+1} + \dots + \bar{x}_{i-1} + \bar{x}_i)/q.$$

On notera que pour $i < q$, c'est-à-dire pour les premiers échantillons, les moyennes mobiles sont calculées à partir de moins de q valeurs. Ainsi, par exemple, si q est égal à 3, la valeur \bar{x}'_1 sera égale à x_1 ou à \bar{x}_1 et la valeur \bar{x}'_2 sera égale à la moyenne de x_1 et x_2 ou de \bar{x}_1 et \bar{x}_2 .

Si l'effectif de l'échantillon j est égal à n_j et en considérant que n_j vaut 1 s'il s'agit d'une valeur individuelle, la variance de x_j ou de \bar{x}_j est égale à σ_0^2/n_j et, pour $i \geq q$, la variance de \bar{x}'_i , est égale à :

$$v(\bar{x}'_i) = \frac{\sigma_0^2}{q^2} \left(\frac{1}{n_{i-q+1}} + \dots + \frac{1}{n_{i-1}} + \frac{1}{n_i} \right),$$

car \bar{x}'_i est, à la constante q près, la somme de q variables considérées comme indépendantes. Lorsque $i < q$, le nombre de termes intervenant dans l'expression de la variance est égal à i et non à q et le dénominateur, q^2 , est remplacé par i^2 .

Si les effectifs n_i sont égaux à n , la formule se simplifie et on a :

$$v(\bar{x}'_i) = \sigma_0^2/qn \quad \text{si } i \geq q$$

et

$$v(\bar{x}'_i) = \sigma_0^2/in \quad \text{si } i < q.$$

Les limites de contrôle des moyennes mobiles sont alors égales à :

$$m_0 \pm u_{1-\alpha/2} \sqrt{v(\bar{x}'_i)} \quad \text{ou} \quad m_0 \pm k \sqrt{v(\bar{x}'_i)},$$

⁽²⁾ En anglais : *moving average control chart*.

selon qu'elles sont exprimées en probabilité ou en fonction d'un multiple de l'erreur-standard.

Si m_0 et σ_0 sont inconnus, on les estime, par exemple à partir des h premiers échantillons, comme nous l'avons signalé à propos des cartes de contrôle de SHEWHART.

D'autre part, le choix de la constante q conditionne l'importance du lissage : plus cette constante est élevée, plus l'effet du lissage est important et plus la carte de contrôle permet d'apprécier la tendance générale au détriment des fluctuations de point à point. L'option par défaut proposée par Minitab est $q = 3$.

Ce choix de q peut également être réalisé sur la base de la *longueur moyenne de la suite homogène*⁽³⁾, appelée aussi période opérationnelle moyenne, qui représente le nombre moyen d'échantillons qu'il faut prélever avant qu'un point ne sorte des limites de contrôle. Cette longueur moyenne dépend du risque de première espèce, de l'importance de la modification de la moyenne, exprimée en proportion de l'erreur-standard de \bar{x}_i et de la constante q . NIANG et SAPORTA [1995] ont établi la formule de calcul et présentent une table donnant les longueurs moyennes pour diverses conditions. L'examen de ces tables montre que des faibles modifications de la moyenne du procédé sont plus rapidement mises en évidence avec des valeurs élevées de q . Par contre, les modifications importantes de la moyenne sont plus rapidement détectées avec des valeurs relativement faibles de q .

Il faut noter aussi que les moyennes mobiles ne sont pas indépendantes même si les observations individuelles ou les échantillons le sont. Par conséquent, l'interprétation de la succession des points n'est pas la même que pour des moyennes successives indépendantes. En particulier, on n'effectue généralement pas les différents tests de détection de situations particulières décrits antérieurement [PALM, 1996].

A titre d'illustration, nous avons repris les données traitées précédemment et relatives aux poids égouttés de boîtes de purée de tomates, qui ont servi à illustrer la détection de situations particulières à l'aide des règles complémentaires [PALM, 1996]. Nous considérons que la moyenne et l'écart-type du procédé valent :

$$m_0 = 612,17 \quad \text{et} \quad \sigma_0 = 40,185,$$

ces estimations ayant été déduites des observations réalisées au cours d'une phase préliminaire. Les données utilisées pour la phase de contrôle proprement dite ont été simulées, en considérant une augmentation de la moyenne de 10 unités.

Les moyennes des premiers échantillons sont les suivantes :

$$615,4, \quad 609,0, \quad 652,2, \quad 608,4, \quad 640,8, \quad 611,0, \quad 624,0 \dots$$

En prenant arbitrairement $q = 5$, on obtient les moyennes mobiles suivantes :

$$615,4, \quad 612,2, \quad 625,5, \quad 621,3, \quad 625,2, \quad 624,3, \quad 627,3 \dots$$

⁽³⁾ En anglais : *average run length, ARL*.

Tableau 1. Erreurs-standards et limites de contrôle des moyennes mobiles d'ordre 5.

i	Erreurs-standards	Limites inférieures	Limites supérieures
1	17,97	558,3	666,1
2	12,71	574,0	650,3
3	10,38	581,0	643,3
4	8,99	585,2	639,1
≥ 5	8,04	588,1	636,3

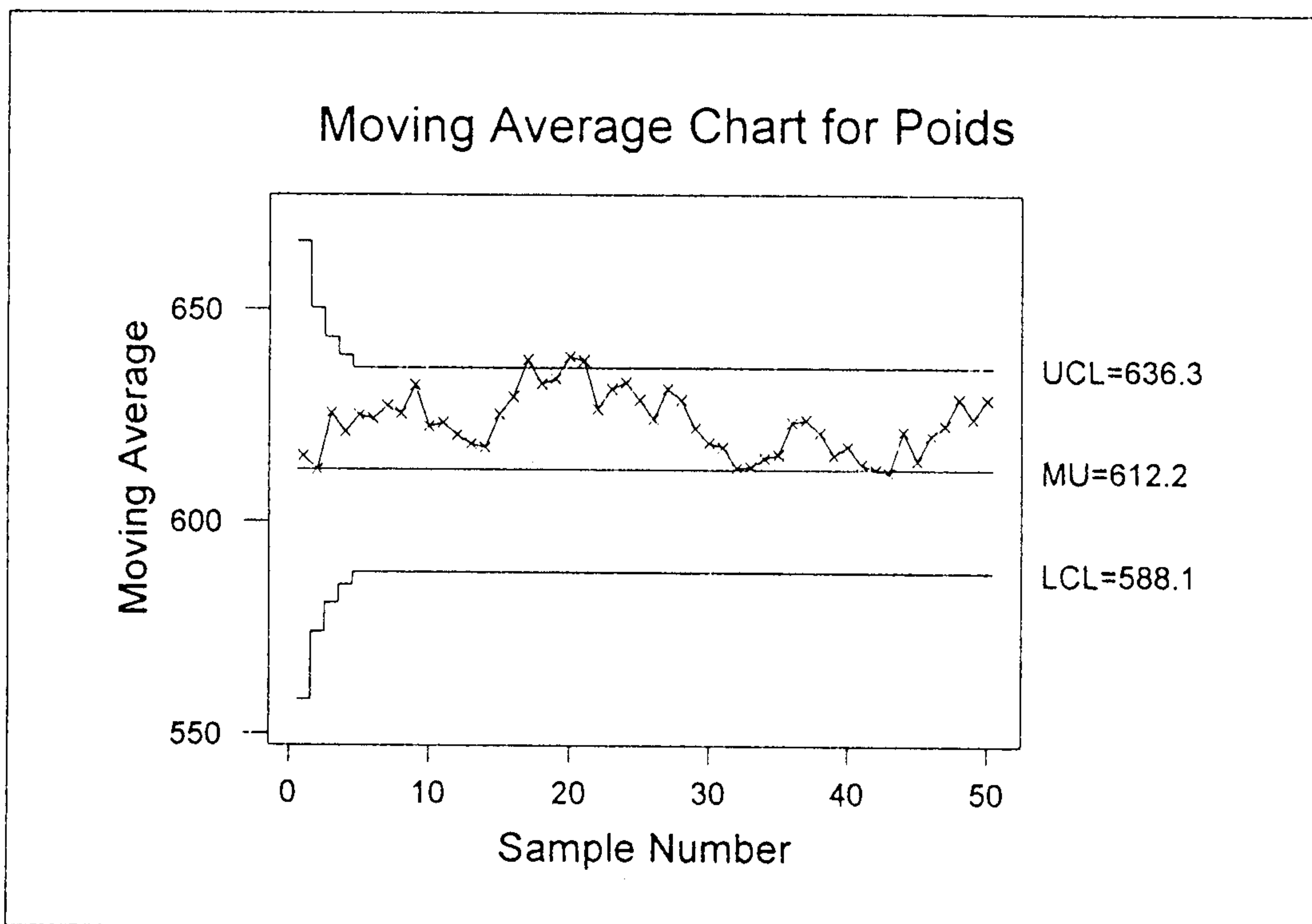


Figure 1. Carte de contrôle sur les moyennes mobiles ($q = 5$ et $m = m_0 + 10$).

Le tableau 1 reprend, en fonction du numéro d'ordre, i , de l'échantillon, l'erreur-standard de la moyenne mobile et les limites de confiance, calculées pour $k = 3$.

La figure 1, obtenue par la commande MACHART du logiciel Minitab, donne la carte de contrôle des moyennes mobiles d'ordre 5. On retrouve bien pour les premières valeurs, les résultats calculés ci-dessus. On constate que trois points sont hors des limites de contrôle, indiquant que le procédé n'est pas sous contrôle, du fait de l'augmentation de 10 unités de la moyenne du procédé. Une intervention aurait pu avoir lieu sur le procédé dès le 17^{ième} échantillon.

2.3. Cartes de contrôle après lissage exponentiel

Le lissage exponentiel simple est une technique développée dans le cadre de l'étude des séries chronologiques. La méthode s'apparente au calcul des moyennes mobiles dont il a été question au paragraphe précédent.

Rappelons que, dans le cas de la moyenne mobile et sauf pour les premières valeurs, la valeur \bar{x}'_i est la moyenne d'un nombre constant, q , de valeurs individuelles ou moyennes, chacune de ces valeurs ayant un poids identique.

Dans le cas de la carte de contrôle après *lissage exponentiel*⁽⁴⁾, par contre, on calcule une moyenne pondérée des observations passées, \bar{x}^*_i , en accordant à celles-ci des poids de plus en plus petits, au fur et à mesure que l'information vieillit. De façon plus précise, on a :

$$\bar{x}^*_i = \lambda \bar{x}_i + \lambda(1 - \lambda) \bar{x}_{i-1} + \lambda(1 - \lambda)^2 \bar{x}_{i-2} + \lambda(1 - \lambda)^3 \bar{x}_{i-3} + \dots,$$

λ étant une constante, comprise entre 0 et 1. Dans cette relation, ainsi que dans les relations qui vont suivre, \bar{x}_j désigne la moyenne arithmétique du $j^{\text{ième}}$ échantillon d'effectif n_j ou la $j^{\text{ième}}$ observation individuelle, n_j étant alors égal à l'unité.

La méthode porte le nom de lissage exponentiel car les poids :

$$w_j = \lambda(1 - \lambda)^j,$$

qui constituent une progression géométrique et qui sont fonction de la constante de lissage λ , se situent sur une courbe exponentielle.

De façon stricte, pour que \bar{x}^*_i soit réellement une moyenne pondérée des observations du passé, il faut que la somme des w_j soit égale à l'unité, ce qui n'est vrai que si le nombre d'observations du passé est égal à l'infini. Aussi, l'équation est-elle généralement présentée par la formule de récurrence :

$$\bar{x}^*_i = \lambda \bar{x}_i + (1 - \lambda) [\lambda \bar{x}_{i-1} + \lambda(1 - \lambda) \bar{x}_{i-2} + \lambda(1 - \lambda)^2 \bar{x}_{i-3} + \dots],$$

c'est-à-dire également :

$$\bar{x}^*_i = \lambda \bar{x}_i + (1 - \lambda) \bar{x}^*_{i-1}.$$

Pour débiter le processus de lissage par cette formule de récurrence, il faut disposer d'une valeur initiale \bar{x}^*_0 , qu'on posera par exemple égale à la moyenne m_0 . On aura alors :

$$\bar{x}^*_1 = \lambda \bar{x}_1 + (1 - \lambda) m_0$$

$$\bar{x}^*_2 = \lambda \bar{x}_2 + (1 - \lambda) \bar{x}^*_1,$$

et ainsi de suite.

Le choix du paramètre λ se pose de façon analogue au choix de q dans le cas de la moyenne mobile. En effet, il permet de donner un poids plus ou moins

⁽⁴⁾ En anglais : *exponentially weighted moving average control chart, EWMA control chart.*

grand aux données récentes. A la limite, si λ est égal à 1, la valeur lissée \bar{x}_i^* se confond avec la valeur \bar{x}_i .

Il existe d'ailleurs une relation entre λ et q par l'intermédiaire de la notion d'âge moyen de l'information. Ainsi, si les effectifs sont constants, pour la moyenne mobile d'ordre q , on peut considérer que l'âge moyen des observations intervenant dans le calcul de \bar{x}_i est égal à :

$$i - (q - 1)/2 \quad (i \geq q).$$

Ainsi, par exemple, la moyenne mobile d'ordre cinq correspondant au 25^{ième} échantillon fait intervenir, en nombre égal, les observations des échantillons de rangs 21, 22, 23, 24 et 25 et l'âge moyen correspond donc à 23.

Pour le lissage exponentiel, on peut montrer que l'âge moyen est égal à :

$$i - (1 - \lambda)/\lambda,$$

et on peut constater que l'âge moyen de la moyenne mobile correspond à l'âge moyen du lissage si :

$$(q - 1)/2 = (1 - \lambda)/\lambda,$$

c'est-à-dire si :

$$\lambda = 2/(q + 1).$$

Par exemple, à $q = 5$ correspond une valeur de λ de 0,33 et à $q = 3$ correspond une valeur de λ de 0,5.

La notion d'âge moyen de l'information permet, dans une certaine mesure, d'orienter le choix de λ , mais d'autres approches peuvent être utilisées.

La première de ces approches repose sur l'hypothèse que les moyennes observées successives, $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_i$, ne sont pas indépendantes mais sont générées par un processus stochastique particulier, le processus ARIMA(0,1,1), dont il sera question au paragraphe 3, lors de l'étude des données corrélées.

L'autre approche repose sur l'hypothèse d'indépendance des données successives et fait appel à la notion de longueur moyenne de la suite homogène. Le logiciel SAS propose la fonction EWMAARL permettant le calcul de cette longueur moyenne et un tableau de valeurs est donné dans la documentation relative à ce logiciel [SAS, 1995].

A titre d'exemple, reprenons le cas du contrôle du poids des boîtes de purée de tomates traité au paragraphe 2.2.

Lors du contrôle des moyennes successives et pour des limites calculées sur la base de $k = 3$, le risque α est égal à 0,0027 et la longueur moyenne de la suite homogène est égale à :

$$1/\alpha = 1/0,0027 = 370,4,$$

comme nous l'avons vu précédemment [PALM, 1996]. Si le procédé reste sous contrôle, il faudra donc observer 370 échantillons, en moyenne, avant qu'une fausse alerte ne soit déclenchée.

Pour la carte de contrôle basée sur le lissage exponentiel avec $\lambda = 1$, on obtient évidemment le même résultat, mais si λ diminue, la longueur moyenne augmente. Pour $\lambda = 0,25$ on a, par exemple, une longueur moyenne de 503 [SAS, 1995].

Si par contre la moyenne du procédé se modifie de 9 unités, soit environ de la moitié de l'erreur-standard d'une moyenne individuelle qui, pour l'exemple traité, vaut 17,9, la table des longueurs moyennes nous donne les résultats suivants : 155 si $\lambda = 1$ et 48,45 si $\lambda = 0,25$. Cela signifie donc que la modification de la moyenne du procédé sera détectée trois fois plus rapidement dans le cas du lissage avec $\lambda = 0,25$ que dans le cas de la carte de contrôle de la moyenne.

De façon générale, l'examen de la table donnée par SAS montre que les longueurs moyennes les plus faibles s'observent pour des valeurs faibles de λ si la modification de la moyenne du procédé est faible et pour des valeurs élevées de λ si la modification est importante. Un compromis acceptable correspond à $\lambda = 0,2$, qui est la valeur par défaut prise par le logiciel Minitab.

En considérant que les échantillons successifs sont indépendants, la variance de la valeur lissée est donnée par la relation suivante :

$$v(\bar{x}_i^*) = \sigma_0^2 \lambda^2 \left[\frac{1}{n_i} + \frac{(1-\lambda)^2}{n_{i-1}} + \frac{(1-\lambda)^4}{n_{i-2}} + \frac{(1-\lambda)^6}{n_{i-3}} + \dots + \frac{(1-\lambda)^{2(i-1)}}{n_1} \right],$$

et les limites de contrôle s'écrivent alors :

$$m_0 \pm u_{1-\alpha/2} \sqrt{v(\bar{x}_i^*)} \quad \text{ou} \quad m_0 \pm k \sqrt{v(\bar{x}_i^*)}.$$

Si les paramètres m_0 et σ_0 sont inconnus, ils sont estimés par les méthodes décrites dans le cadre des cartes de contrôle de SHEWHART.

La relation donnant la variance de la valeur lissée provient du fait que celle-ci est la somme de i termes considérés comme indépendants :

$$\lambda(1-\lambda)^j \bar{x}_{i-j} \quad (j = 0, \dots, i-1)$$

dont la variance vaut :

$$\lambda^2(1-\lambda)^{2j} \sigma_0^2 / n_{i-j}.$$

Dans les cas où les effectifs sont constants, l'expression de la variance se simplifie et on a :

$$v(\bar{x}_i^*) = \sigma_0^2 \lambda [1 - (1-\lambda)^{2i}] / n(2-\lambda).$$

Cette relation montre que la variance tend vers la constante :

$$\sigma_0^2 \lambda / n(2-\lambda)$$

lorsque i augmente, car $(1-\lambda)^{2i}$ tend vers zéro.

Si on compare cette constante à la variance de la moyenne mobile d'ordre q , on constate qu'il y a égalité des deux variances :

$$\frac{\sigma_0^2 \lambda}{n(2 - \lambda)} = \frac{\sigma_0^2}{nq}$$

lorsque :

$$\lambda = 2/(q + 1),$$

c'est-à-dire lorsque l'âge moyen de la valeur lissée est identique à l'âge moyen de la moyenne mobile, comme nous venons de le voir ci-dessus ; dans ce cas, les deux cartes de contrôle ont des propriétés fort proches.

La figure 2, obtenue par la commande EWMA CHART de Minitab, donne la carte de contrôle pour les données considérées au paragraphe précédent, en adoptant pour le paramètre de lissage la valeur proposée par défaut ($\lambda = 0,2$).

Les valeurs lissées successives sont les suivantes :

$$\begin{aligned}\bar{x}_1^* &= (0,2)(615,4) + (0,8)(612,17) = 612,82, \\ \bar{x}_2^* &= (0,2)(609,0) + (0,8)(612,82) = 612,06, \\ \bar{x}_3^* &= (0,2)(652,2) + (0,8)(612,06) = 620,09, \\ &\text{etc.}\end{aligned}$$

Quant aux variances de ces valeurs, elles sont égales à :

$$\begin{aligned}v(\bar{x}_1^*) &= (40,185)^2(0,2)^2/5 = 12,92, \\ v(\bar{x}_2^*) &= (40,185)^2(0,2)^2(1 + 0,8^2)/5 = 21,19, \\ v(\bar{x}_3^*) &= (40,185)^2(0,2)^2(1 + 0,8^2 + 0,8^4)/5 = 26,49, \\ &\text{etc.}\end{aligned}$$

ou, plus simplement encore :

$$\begin{aligned}v(\bar{x}_1^*) &= (40,185)^2(0,2)(1 - 0,8^2)/(5)(1,8) = 12,92, \\ v(\bar{x}_2^*) &= (40,185)^2(0,2)(1 - 0,8^4)/(5)(1,8) = 21,19, \\ v(\bar{x}_3^*) &= (40,185)^2(0,2)(1 - 0,8^6)/(5)(1,8) = 26,49, \\ &\text{etc.}\end{aligned}$$

Les limites de contrôle sont donc, respectivement :

$$\begin{aligned}612,17 \pm 3\sqrt{12,92}, &\text{ soit } 601,4 \text{ et } 623,0, \\ 612,17 \pm 3\sqrt{21,19}, &\text{ soit } 598,4 \text{ et } 626,0, \\ 612,17 \pm 3\sqrt{26,49}, &\text{ soit } 596,7 \text{ et } 627,6, \\ &\text{etc.}\end{aligned}$$

Elles tendent vers :

$$612,17 \pm 3\sqrt{35,89} \text{ soit } 594,2 \text{ et } 630,1,$$

puisque la variance de la valeur lissée tend vers :

$$v(\bar{x}^*) = (40,185)^2(0,2)/(5)(1,8) = 35,89.$$

La figure 2 montre, plus clairement encore que le graphique des moyennes mobiles, que le procédé n'est pas sous contrôle, puisque plusieurs points sont hors des limites.

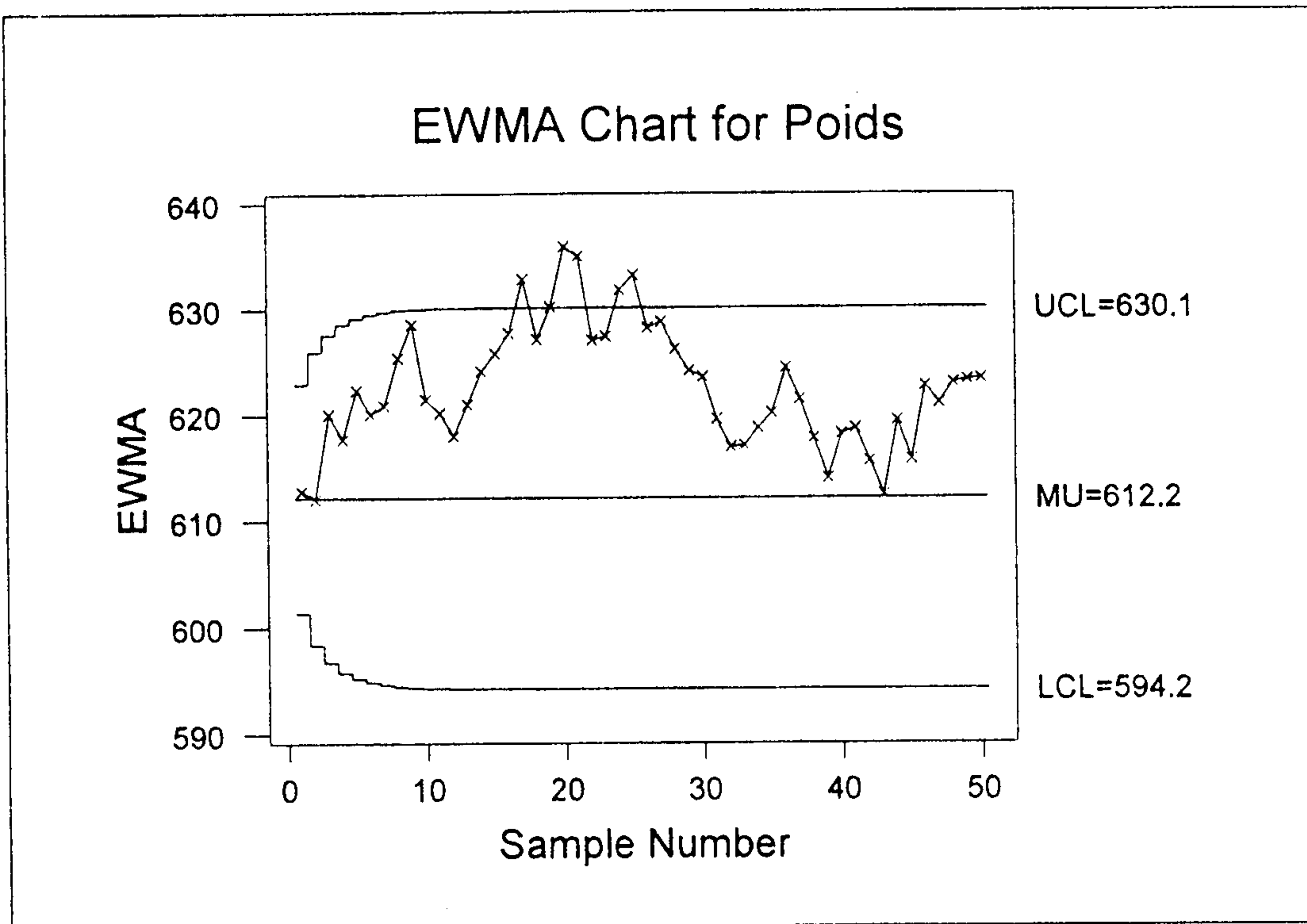


Figure 2. Carte de contrôle après lissage exponentiel ($\lambda = 0,2$ et $m = m_0 + 10$).

2.4. Cartes de contrôle basées sur les sommes cumulées

Une autre forme de combinaison des résultats d'échantillons successifs consiste à utiliser des procédures de contrôle basées sur les *sommes cumulées*⁽⁵⁾.

Pour chaque échantillon, on calcule l'écart réduit :

$$z_i = (\bar{x}_i - m_0) / (\sigma_0 / \sqrt{n_i}) .$$

Ces écarts réduits sont alors cumulés de proche en proche :

$$\begin{aligned} S_1 &= z_1 , \\ S_2 &= z_1 + z_2 , \\ &\dots \\ S_i &= z_1 + z_2 + \dots + z_i , \end{aligned}$$

et portés sur un graphique.

Si le procédé est sous contrôle, les points de ce graphique ne présentent pas de tendance particulière, car les écarts positifs et négatifs se compensent. Par contre, si la moyenne m du procédé est supérieure à m_0 alors, le graphique présente une tendance croissante. De même si la moyenne est inférieure à m_0 , la tendance est décroissante.

⁽⁵⁾ En anglais : *cumulative sum chart*, *CUSUM chart*.

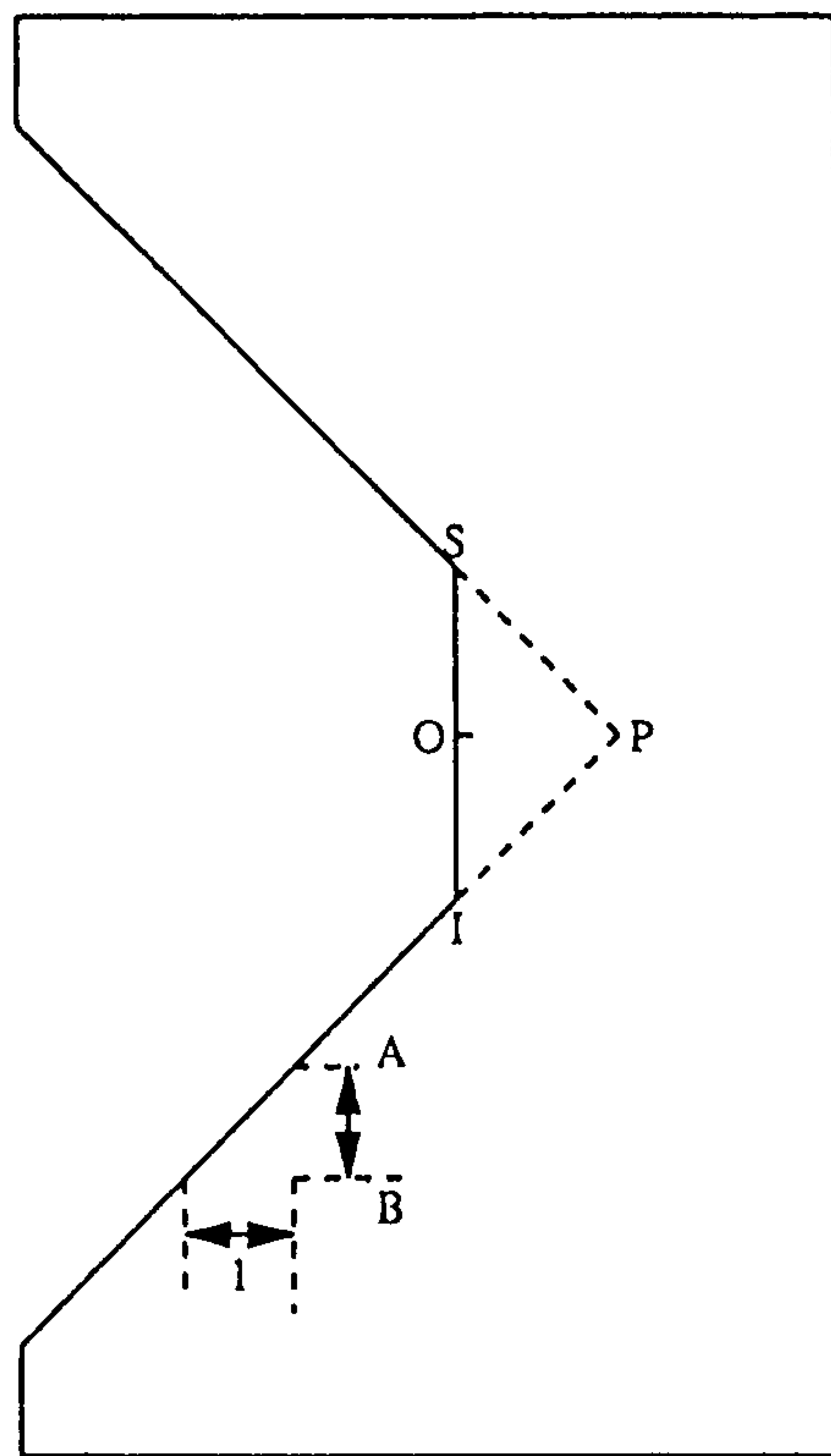


Figure 3. Représentation schématique du masque utilisé dans la carte de contrôle des sommes cumulées.

Pour suivre l'évolution du procédé et vérifier s'il est sous contrôle, on peut utiliser un masque en forme de V, découpé dans une feuille de carton (figure 3), que l'on place sur le graphique, de manière à faire coïncider le point O du masque avec le dernier point du graphique et en maintenant la bissectrice de l'angle du V en position horizontale. Tant que le graphique des sommes cumulées est entièrement visible, on considère que le procédé est sous contrôle. Par contre, dès que le masque cache un ou plusieurs points, on déclare que le procédé n'est plus sous contrôle.

Le contrôle basé sur les sommes cumulées revient à réaliser un test séquentiel de conformité de la moyenne, ce test étant appliqué à la série des moyennes prises dans l'ordre inverse des observations [JONHSON, 1961].

La figure 4 donne la carte de contrôle des sommes cumulées pour l'exemple déjà traité dans les paragraphes précédents et le masque a été positionné sur la somme cumulée obtenue à la suite des 50 échantillons. Cette figure a été réalisée par la commande %CUSUM du logiciel Minitab. Ce logiciel ne standardise pas les écarts par rapport à m_0 qui sont cumulés, ce qui aura une incidence sur la définition du masque, comme nous le verrons ci-dessous.

On constate que le bras inférieur du masque coupe la ligne des sommes cumulées, indiquant clairement qu'il y a augmentation de la moyenne par rapport à m_0 .

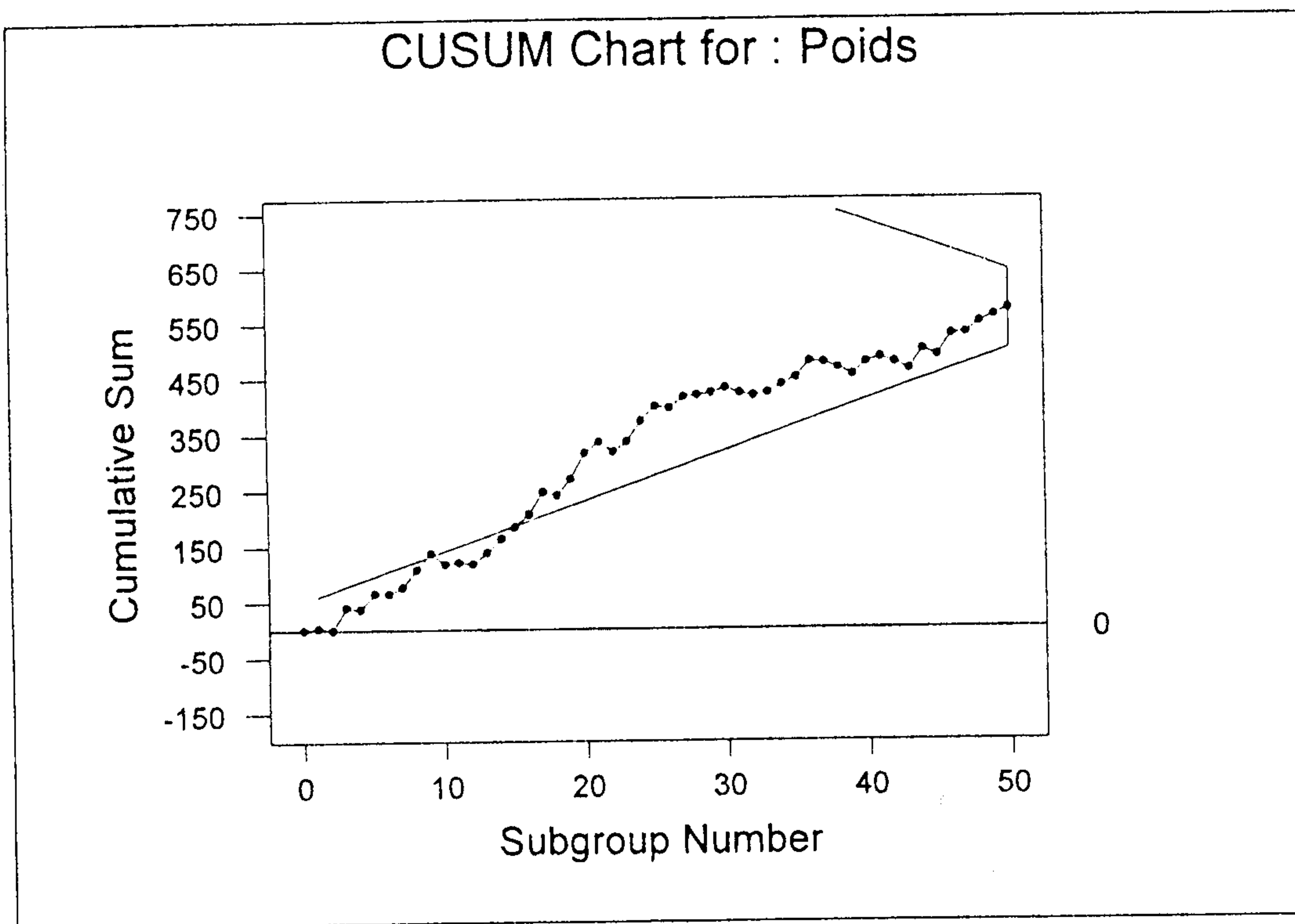


Figure 4. Carte de contrôle des sommes cumulées bilatérales ($h = 4$, $k = 0,5$ et $m = m_0 + 10$).

La procédure de contrôle des sommes cumulées est évidemment fonction de la forme du masque utilisé. Cette forme peut être décrite de différentes manières. Une solution consiste à définir les paramètres h et k .

Le premier paramètre, h , mesure la distance verticale entre l'origine et la branche inférieure ou supérieure du V. Sur la figure 3, cette distance est égale à OI ou OS. Le deuxième paramètre, k , est la pente de la branche inférieure du V. Il s'agit de la distance AB sur la figure 3. La figure 3 a été construite en utilisant des unités arbitraires, mais en pratique, il faut évidemment respecter les unités utilisées dans le graphique des sommes cumulées.

Les valeurs de h et de k qui viennent d'être définies supposent que les sommes cumulées ont été standardisées. Si, comme dans la figure 4, les écarts par rapport à la moyenne ne sont pas divisés par l'erreur-standard de la moyenne et pour autant que les effectifs soient constants, alors la demi-largeur du masque à l'origine et la pente de la branche inférieure du V valent respectivement :

$$H = h \sigma_0 / \sqrt{n} \quad \text{et} \quad K = k \sigma_0 / \sqrt{n}.$$

Ainsi pour la figure 4, on a retenu les valeurs suivantes :

$$h = 4 \quad \text{et} \quad k = 0,5,$$

qui sont les valeurs utilisées par défaut par la commande %CUSUM de Minitab.

On a donc :

$$H = (4)(40,185)/\sqrt{5} = 71,89$$

et

$$K = (0,5)(40,185)/\sqrt{5} = 8,99.$$

Le choix de h et de k peut être réalisé de manière relativement empirique ou de façon plus raisonnée en fonction des risques α et β . Rappelons que α est le risque de première espèce, c'est-à-dire la probabilité de détecter un changement de la moyenne alors qu'en réalité le procédé est sous contrôle, et que β est le risque de deuxième espèce, c'est-à-dire la probabilité de ne pas détecter une modification donnée de la moyenne.

Les valeurs h et k sont, en effet, liées aux valeurs α et β par les relations suivantes :

$$h = \log_e[(1 - \beta)/(\alpha/2)]/|\delta|$$

et

$$k = |\delta|/2.$$

Dans ces relations, δ est l'écart par rapport à la moyenne théorique qu'on souhaite mettre en évidence avec la probabilité $1 - \beta$. Cet écart est exprimé sous la forme d'un multiple de l'erreur-standard des moyennes \bar{x}_i :

$$\delta = \frac{m - m_0}{\sigma_0/\sqrt{n}}.$$

Si β est suffisamment petit, la formule donnant h peut être simplifiée de la manière suivante :

$$h = -\log_e(\alpha/2)/|\delta|.$$

On remarque que l'effectif de l'échantillon n n'intervient pas directement dans le calcul de h et k . Son incidence est indirecte, car les valeurs de h et k se mesurent sur l'axe des ordonnées et l'effectif intervient dans la standardisation de la somme cumulée.

Pour l'exemple considéré, au lieu d'utiliser les valeurs par défaut, on aurait pu déterminer les valeurs de h et k de manière à mettre en évidence une différence de 20 unités, par exemple, avec une probabilité égale à 0,9, en considérant que α vaut 0,002. On trouverait alors :

$$\delta = 20/(40,185/\sqrt{5}) = 1,112,$$

$$h = \log(0,9/0,001)/1,112 = 6,11,$$

$$k = 1,112/2 = 0,56.$$

Par rapport à la solution par défaut, il y aurait un déplacement vers le bas de la branche inférieure du V, presque parallèlement à elle-même puisqu'on aurait :

$$H = (6,1)(40,185)/\sqrt{5} = 109,6,$$

$$K = (0,56)(40,185)/\sqrt{5} = 10,06.$$

On mettrait donc moins vite en évidence la modification de moyenne que dans le cas précédent.

Les paramètres h et k peuvent également être déterminés en fonction de la longueur moyenne de la suite homogène, qui représente le nombre moyen d'échantillons prélevés avant de déclarer que le procédé n'est pas sous contrôle. Idéalement cette longueur devrait être grande si $m = m_0$ et petite si m est différent de m_0 . Les valeurs de cette longueur moyenne sont données dans des tables ou abaques. Des références bibliographiques sont données dans SAS [1995]. Ces longueurs moyennes peuvent également être obtenues à partir de la fonction CUSUMARL du logiciel SAS.

Pour $h = 4$ et $k = 0,5$, les tables en question montrent, par exemple, que la longueur moyenne de la suite homogène est de 168 si $\delta = 0$ et de 8,38 si $\delta = 1$. Par contre pour $h = 6$ et $k = 0,5$ les deux valeurs sont respectivement égales à 1.277 et 12,37.

La carte de contrôle que nous venons de présenter suppose qu'on souhaite détecter aussi bien une augmentation qu'une diminution de la moyenne du procédé et la technique de contrôle correspond en fait à des tests bilatéraux. La carte de contrôle sur les sommes cumulées a cependant initialement été mise au point pour détecter des modifications dans une seule direction, par des tests unilatéraux.

Si on désire détecter une éventuelle augmentation de la moyenne, on calcule les écarts suivants :

$$z_i^+ = [\bar{x}_i - (m_0 + k\sigma_0/\sqrt{n_i})] / (\sigma_0/\sqrt{n_i}),$$

k étant une constante positive fixée, appelée *valeur de référence*⁽⁶⁾.

Les z_i^+ sont alors cumulés pour conduire aux quantités S_i^+ , appelées *sommes cumulées supérieures*⁽⁷⁾, qui sont portées sur le graphique. Lorsque le cumul des écarts conduit à une valeur négative, on pose $S_i^+ = 0$. En posant $S_0^+ = 0$ on a donc, pour les différents échantillons :

$$S_i^+ = \max(0, S_{i-1}^+ + z_i^+).$$

On considère enfin qu'il y a augmentation de la moyenne du procédé lorsque S_i^+ dépasse une constante positive h , appelée *intervalle de décision*⁽⁸⁾. Lorsqu'un réglage du procédé a lieu suite à la mise en évidence d'une modification de la moyenne, la somme cumulée est remise à zéro.

Pour détecter une éventuelle diminution de la moyenne, on calcule les écarts :

$$z_i^- = [\bar{x}_i - (m_0 - k\sigma_0/\sqrt{n_i})] / (\sigma_0/\sqrt{n_i}),$$

qu'on cumule pour obtenir les quantités S_i^- , appelées *sommes cumulées inférieures*⁽⁹⁾ :

$$S_i^- = \max(0, S_{i-1}^- - z_i^-),$$

⁽⁶⁾ En anglais : *reference value*.

⁽⁷⁾ En anglais : *upper cumulative sums*.

⁽⁸⁾ En anglais : *decision interval*.

⁽⁹⁾ En anglais : *lower cumulative sums*.

et si S_i^- dépasse la valeur positive h , on considère qu'il y a eu diminution de la moyenne.

Certains auteurs utilisent une formulation différente pour les sommes cumulées inférieures, qui revient simplement à prendre l'opposé des sommes définies ci-dessus :

$$S_i^- = \min(0, S_{i-1}^- + z_i^-).$$

Ces sommes, qui sont alors négatives ou nulles, sont comparées à la valeur $-h$.

Enfin, comme dans le cas des tests bilatéraux, les écarts peuvent ne pas être divisés par l'erreur-standard de la moyenne, du moins si les effectifs des échantillons sont constants.

Comme pour les cartes des sommes bilatérales, le choix de h et k peut être réalisé de façon empirique ou en fonction de la longueur moyenne des suites homogènes. Des tables donnant cette longueur en fonction de h , k et δ sont données dans la documentation SAS et la fonction CUSUMARL de ce logiciel permet le calcul de ces valeurs [SAS, 1995]. Ainsi, pour $h = 4$ et $k = 0,5$, qui sont les valeurs prises par défaut par le logiciel Minitab, cette longueur moyenne vaut 335 si $\delta = 0$ et 8,38 si $\delta = 1$. Par rapport au cas bilatéral, la longueur est doublée pour $\delta = 0$ alors qu'elle est identique pour $\delta = 1$.

Pour l'exemple considéré, la figure 5 donne la carte de contrôle des sommes cumulées, à la fois pour la mise en évidence d'une augmentation ou d'une réduction de la moyenne. Ce graphique a été obtenu par le logiciel Minitab en utilisant les valeurs proposées par défaut et il n'y a pas eu remise à zéro de la somme cumulée lors de la mise en évidence d'une modification de la moyenne. Les écarts sommés n'ayant pas été divisés par l'écart-type de la moyenne les valeurs utilisées sont, comme pour les tests bilatéraux :

$$H = 71,89 \quad \text{et} \quad K = 8,99.$$

Le tableau 2 donne les détails relatifs au calcul de sommes cumulées S_i , S_i^+ et S_i^- pour les premiers échantillons. La première colonne correspond aux numéros des échantillons ; la deuxième colonne reprend les moyennes des échantillons, \bar{x}_i ; la troisième colonne donne les écarts par rapport à m_0 :

$$z_i = \bar{x}_i - 612,17$$

et la quatrième colonne donne les S_i . On retrouve bien les valeurs représentées sur la carte de contrôle des sommes cumulées bilatérales (figure 4).

La cinquième colonne reprend les écarts :

$$z_i^+ = \bar{x}_i - (612,17 + 8,99),$$

et les S_i^+ sont repris en sixième colonne. Ces sommes ont été représentées dans la figure 5 par des cercles. On constate que S_1^+ et S_2^+ sont nuls, car les deux premiers écarts, donc aussi les deux premiers cumuls, sont négatifs. On constate aussi, qu'au 9^{ième} échantillon, la somme cumulée dépasse la valeur H , qui, rappelons-le, est égale à 71,89. La carte de contrôle met donc en évidence une augmentation de la moyenne.

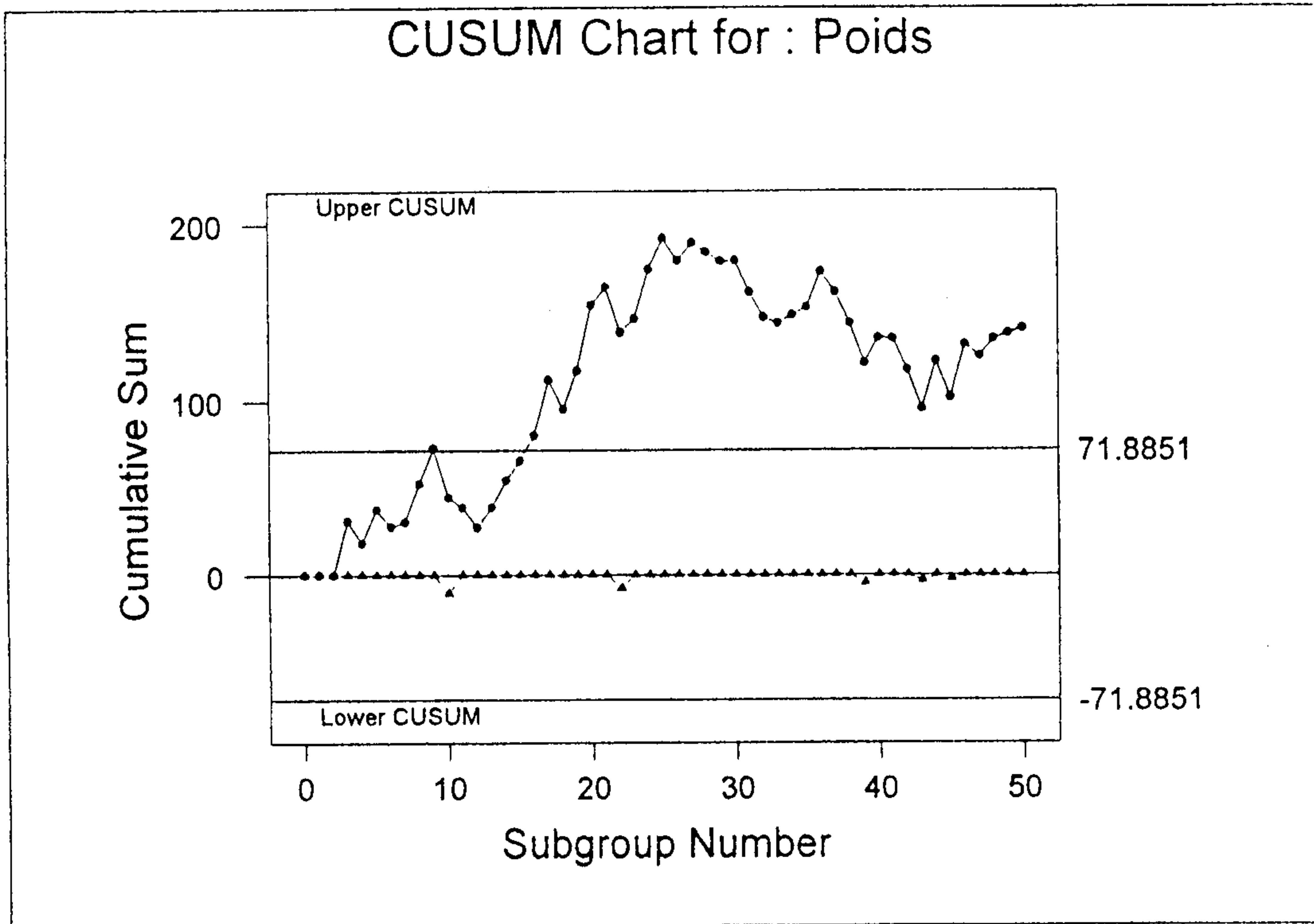


Figure 5. Carte de contrôle des sommes cumulées unilatérales ($h = 4$, $k = 0,5$ et $m = m_0 + 10$).

Tableau 2. Calcul des sommes cumulées bilatérales et unilatérales ($m = m_0 + 10$).

i	\bar{x}_i	z_i	S_i	z_i^+	S_i^+	z_i^-	S_i^-
1	615,4	3,23	3,23	-5,76	0	12,22	0
2	609,0	-3,17	0,06	-12,16	0	5,82	0
3	652,2	40,03	40,09	31,04	31,04	49,02	0
4	608,4	-3,77	36,32	-12,76	18,28	5,22	0
5	640,8	28,63	64,95	19,64	37,92	37,62	0
6	611,0	-1,17	63,78	-10,16	27,76	7,82	0
7	624,0	11,83	75,61	2,84	30,60	20,82	0
8	643,6	31,43	107,04	22,44	53,04	40,42	0
9	641,4	29,23	136,27	20,24	73,28	38,22	0
10	592,8	-19,37	116,90	-28,36	44,92	-10,38	-10,38

La septième colonne du tableau 2 donne les écarts :

$$z_i^- = \bar{x}_i - (612,17 - 8,99),$$

et, enfin, la huitième colonne reprend les valeurs de S_i^- . La plupart des S_i^- sont nuls car les z_i^- sont pratiquement tous positifs. Les S_i^- sont représentés, dans

la figure 5, par des triangles. Aucune valeur n'est inférieure à -71,89, et donc aucune diminution de la moyenne n'est mise en évidence.

Pour des valeurs fixées de H et de K (ou de h et de k), il existe une relation entre la carte de contrôle sur les sommes cumulées dans le cas bilatéral et dans les deux cas unilatéraux.

En effet, les écarts entre les sommes cumulées, S_i , et la branche inférieure du masque dans le cas du contrôle bilatéral sont tout à fait comparables aux écarts entre les sommes cumulées S_i^+ et H .

Pour s'en rendre compte, il suffit de se rappeler que la branche inférieure du masque a une pente égale à K . Pour une abscisse donnée, i , l'ordonnée du point situé sur la droite vaut :

$$S_{(i)} = a + iK,$$

et l'écart entre S_i et $S_{(i)}$ est égal à :

$$S_i - S_{(i)} = (z_1 + z_2 + \dots + z_i) - a - iK.$$

Pour les écarts entre les sommes cumulées supérieures et H on a :

$$\begin{aligned} S_i^+ - H &= (z_1^+ + z_2^+ + \dots + z_i^+) - H \\ &= (z_1 + z_2 + \dots + z_i) - iK - H, \end{aligned}$$

pour autant que le cumul des z_i^+ n'ait pas conduit à une valeur négative.

De même, à une constante près, il y a une analogie entre les écarts des S_i par rapport à la branche supérieure du masque et les écarts des S_i^- par rapport à H .

Ainsi donc, la carte de contrôle sur les sommes bilatérales est, en pratique, équivalente à l'examen simultané des sommes cumulées supérieures et des sommes cumulées inférieures.

Par rapport aux cartes classiques des moyennes, les cartes de contrôle sur les sommes cumulées ont l'avantage de mettre plus rapidement en évidence des modifications relativement faibles de la moyenne, de l'ordre de l'écart-type de la moyenne ($\delta = 1$), par exemple. Ainsi pour $h = 4$ et $k = 0,5$ on a vu que la longueur moyenne de la suite homogène est 8,4 si $\delta = 1$. Une carte de contrôle de la moyenne a, dans les mêmes conditions, une longueur moyenne de l'ordre de 55. Par contre, pour $\delta = 2$, la longueur moyenne est de 2,2 pour les sommes cumulées et de 1,0 pour la carte de SHEWHART, ce qui montre que la carte des sommes cumulées est moins efficace pour la détection de modifications importantes de la moyenne.

Des solutions ont été proposées pour améliorer la rapidité de détection dans une telle situation. Une première solution est d'utiliser conjointement la carte de contrôle de SHEWHART et la carte de contrôle des sommes cumulées. Une autre solution, proposée par LUCAS et CROSIER [1982], consiste à doter la carte de contrôle des sommes cumulées d'une procédure de *réponse initiale rapide*⁽¹⁰⁾, en

⁽¹⁰⁾ En anglais : *fast initial response*, FIR.

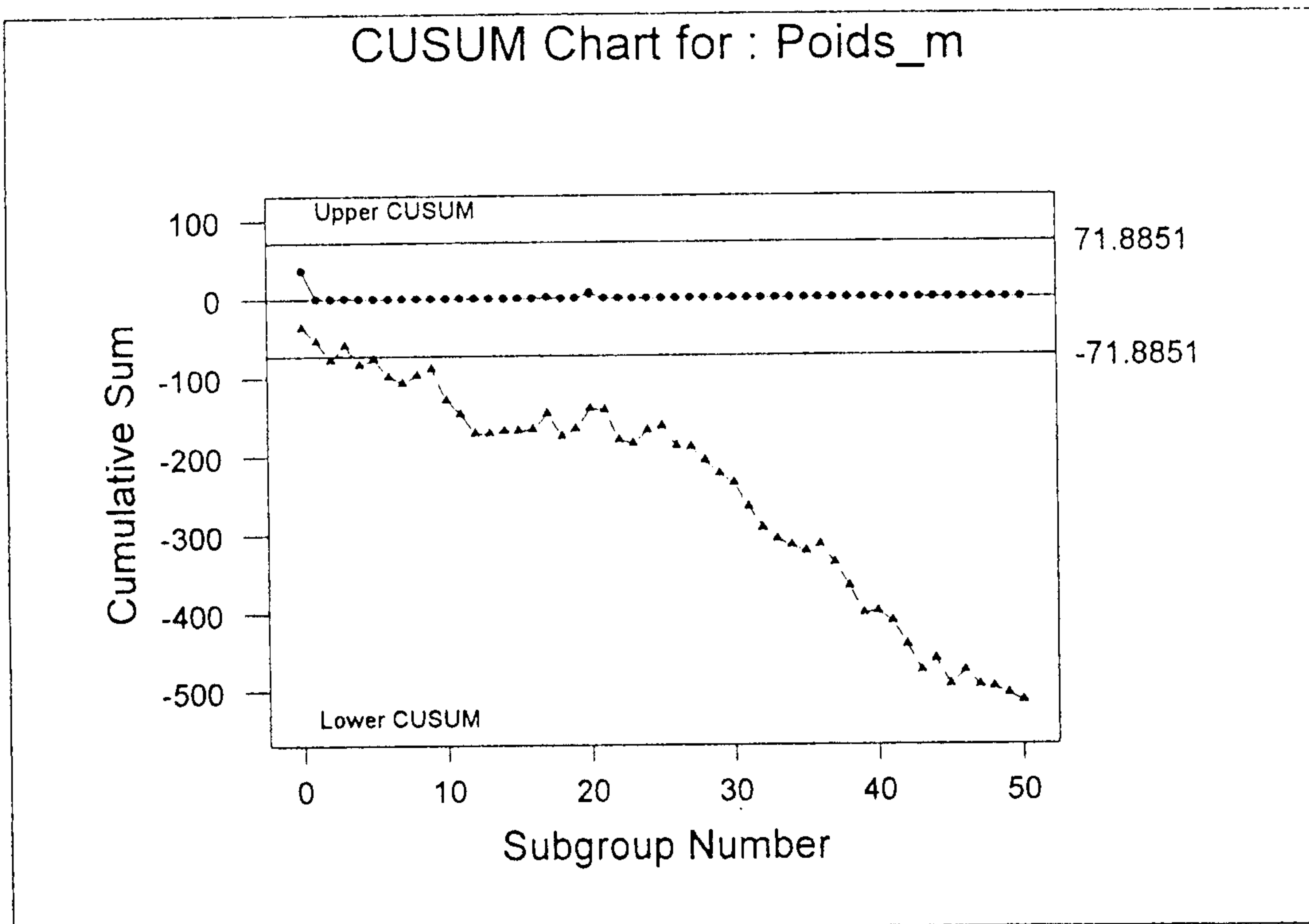


Figure 6. Carte de contrôle des sommes cumulées unilatérales ($h = 4$, $k = 0,5$, valeur de départ = 36,0 et $m = m_0 - 20$).

donnant à la somme cumulée une valeur fixée dès le départ⁽¹¹⁾. Cela revient à donner à S_0^+ et S_0^- une valeur différente de zéro, par exemple $h/2$. De même, lors de la détection d'une modification de la moyenne, la somme cumulée n'est pas remise à zéro mais à la valeur initiale. Si le procédé est sous contrôle cette valeur initiale a très peu d'effet, par contre si le procédé est hors de contrôle, le signal est donné plus rapidement avec une valeur non nulle de S_0 qu'avec une valeur nulle.

Pour illustrer l'effet de cette valeur initiale, nous avons repris les données traitées précédemment en considérant toutefois une modification de la moyenne plus importante. Pour cela nous avons soustrait 30 unités à toutes les observations. Cela revient à dire qu'on considère un procédé dont la moyenne est inférieure de 20 unités à la valeur théorique. Le fait de considérer une sous-estimation permet d'illustrer la notion de somme cumulée inférieure plus clairement que ne le permettait la figure 5. En outre, on a considéré que la valeur de départ est égale à $H/2$, soit 36,0.

Les résultats donnés par Minitab sont repris dans la figure 6. On peut constater que la modification de la moyenne est mise en évidence dès le deuxième échantillon, alors qu'en donnant à S_0 une valeur nulle, il aurait fallu attendre le dixième échantillon.

⁽¹¹⁾ En anglais : *headstart*.

3. CARTES DE CONTRÔLE POUR DONNÉES NON INDÉPENDANTES

3.1. Données avec tendance générale

Les cartes de contrôle de SHEWHART et les cartes présentées au paragraphe précédent ont été développées en supposant que les observations successives sont indépendantes. Cette hypothèse n'est cependant pas toujours vérifiée en pratique et la présence d'une corrélation entre données successives provoque une augmentation de la fréquence des fausses alarmes.

La non-indépendance des observations peut être due à diverses causes parmi lesquelles on peut citer l'usure de l'outil de production, les variations lentes et continues dans la matière première, les modifications liées aux conditions environnementales, comme la température par exemple. Le phénomène d'autocorrélation est encore accentué lorsqu'on réduit l'intervalle de temps entre les échantillons successifs, ce qui est notamment le cas lorsqu'on réalise des observations individuelles de façon relativement continue ou à intervalles de temps assez courts, grâce à des senseurs, par exemple.

En présence d'autocorrélation, deux attitudes différentes peuvent être envisagées. La première réaction consiste à considérer que cette autocorrélation indique la présence d'une source de variation identifiable, qu'on peut donc essayer de mettre en évidence afin de la supprimer et de réduire ainsi la variabilité du procédé. On peut aussi, d'autre part, considérer que cette autocorrélation est due à l'ensemble des causes communes et, par conséquent, en tenir compte pour définir les limites de contrôle du procédé. La distinction entre ces deux situations dépend du contexte. Ce problème est discuté par FALTIN et WOODALL [1991].

Lorsqu'on se trouve dans la seconde situation, on s'efforce de modéliser les données, de manière à éliminer la partie déterministe, responsable de l'autocorrélation. On construit alors la carte de contrôle sur les résidus de cette modélisation. Négliger la présence de l'autocorrélation et traiter les données comme si elles étaient indépendantes conduirait en effet à une augmentation de la fréquence des fausses alarmes.

Nous allons d'abord examiner le cas particulier de procédés de production présentant une tendance générale en moyenne, par exemple à cause de l'usure de l'outil de production. Dans une telle situation, les observations successives présentent un phénomène de corrélation et la variabilité des résultats relatifs à une courte période de production est nettement plus faible que la variabilité des résultats pour l'ensemble de la production. L'origine de l'autocorrélation et de la variabilité supplémentaire est connue et ne peut en pratique être supprimée. On va donc s'efforcer de tenir compte de cette source de variation lors du contrôle du procédé.

La figure 7, relative à des données simulées, schématise une telle situation. Les moyennes des 24 échantillons de 5 observations présentent une tendance linéaire. La pente de la droite est égale à 4,933 et l'écart-type résiduel

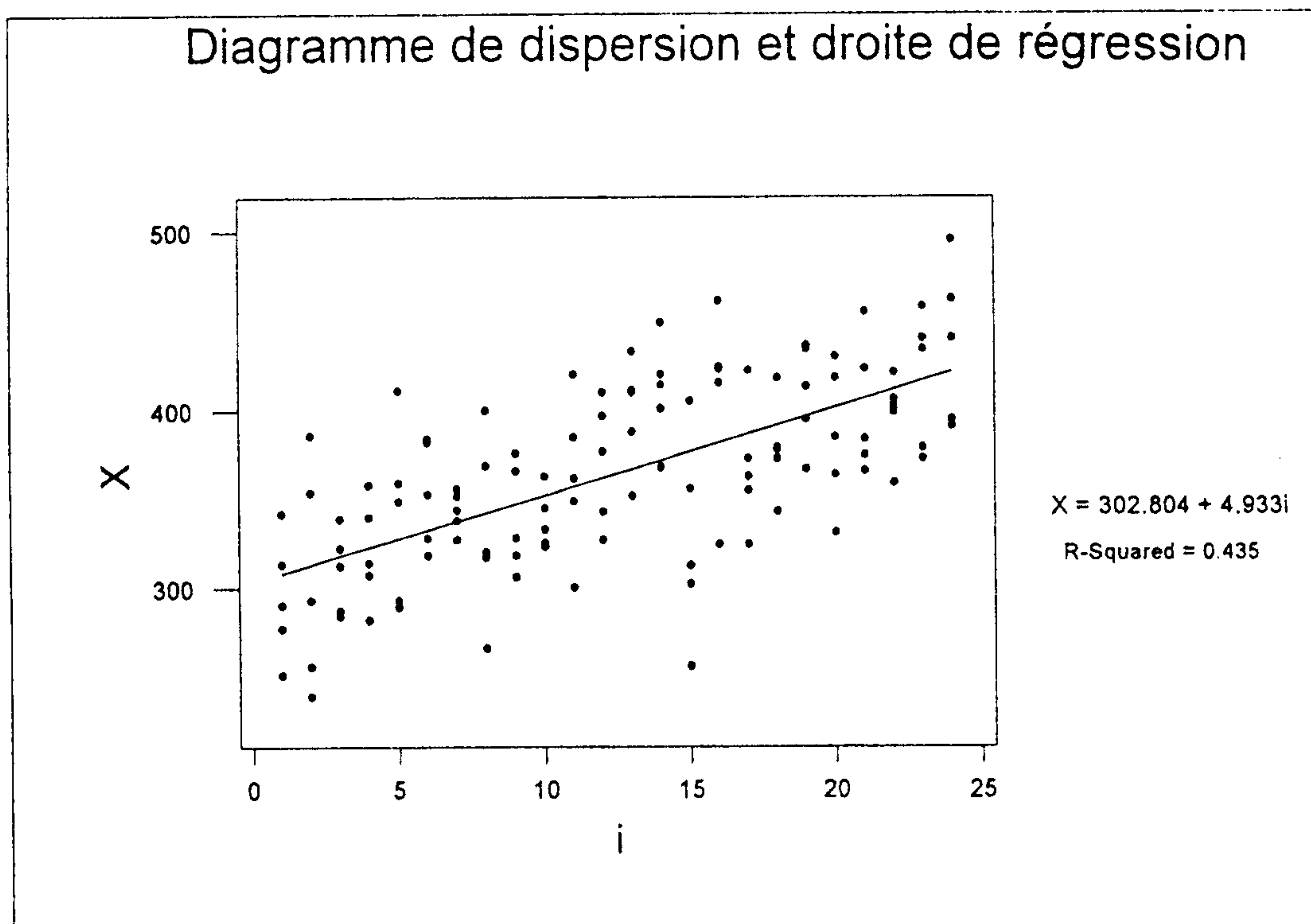


Figure 7. Diagramme de dispersion et droite de régression : données simulées présentant une tendance générale.

estimé est égal à 39,27, alors que l'écart-type de l'ensemble des observations vaut 52,0. On pourrait vérifier que durant la phase préliminaire correspondant aux 24 échantillons, le procédé est bien sous contrôle, tant en ce qui concerne la moyenne que l'écart-type. Il suffit pour cela d'établir les cartes de contrôle sur les écarts par rapport à la tendance générale :

$$x'_{ij} = x_{ij} - (302,80 + 4,93 i) \quad (i = 1, \dots, 24 ; j = 1, \dots, 5),$$

pour autant que ceux-ci soient indépendants.

Les informations recueillies au cours de la phase de précontrôle peuvent être utilisées, d'une part, pour le réglage initial de l'outil et, d'autre part, pour la détermination du délai entre deux interventions successives sur l'outil.

Ainsi, considérons par exemple que les limites de tolérance sont égales à 200 et à 600 et qu'on souhaite que la moyenne se situe au moins à trois écarts-types de ces limites, c'est-à-dire entre les deux valeurs extrêmes suivantes :

$$m_{inf} = 200 + (3)(39,27) = 317,81$$

et

$$m_{sup} = 600 - (3)(39,27) = 482,19,$$

En début de cycle, la moyenne sera donc réglée sur 317,81 et un remplacement, éventuellement partiel, ou un nouveau réglage de l'outil de production

devra avoir lieu avant que la moyenne n'atteigne 482,19. Etant donné qu'entre deux échantillons successifs l'augmentation de la moyenne est de 4,933, il faudra intervenir après le 33^{ième} échantillon, puisque, après 33 intervalles de temps, la moyenne sera de :

$$317,81 + (33)(4,933) = 480,60.$$

L'usure de l'outil étant fonction du nombre d'articles produits, le délai avant l'intervention qui vient d'être calculé n'est valable que si la production entre deux prélèvements successifs reste constante et identique à la production réalisée durant la phase de précontrôle. Si les volumes de production entre prélèvements devaient ne pas être constants, il y aurait lieu d'établir un modèle de régression en considérant comme variable explicative, non pas le numéro d'ordre de l'échantillon, mais bien le volume de la production depuis la dernière intervention sur l'outil.

La procédure qui vient d'être illustrée peut évidemment être étendue au cas où l'évolution de la moyenne n'est pas linéaire, en faisant appel, soit à des transformations de variables en vue de linéariser la relation, soit en utilisant des modèles de régression non linéaires.

Il est bien entendu aussi qu'il faudra intervenir plus tôt si, indépendamment de l'augmentation de la moyenne liée à l'usure de l'outil, d'autres causes de variation interviennent. Il sera donc utile de suivre l'évolution du procédé par une carte de contrôle des moyennes et des écarts-types par exemple.

Pour les moyennes, la ligne centrale se situera en :

$$m_i = 317,81 + 4,933 i$$

et les limites de contrôle correspondront à deux droites parallèles d'équation :

$$317,81 + 4,933 i \pm u_{1-\alpha/2} (39,27/\sqrt{5}).$$

Dans ces relations, i représente le numéro d'ordre de l'échantillon depuis le dernier réglage de l'outil.

Une solution tout à fait équivalente consisterait à effectuer le contrôle, non pas sur les moyennes observées, \bar{x}_i , mais sur une transformation linéaire de ces moyennes :

$$\bar{x}'_i = \bar{x}_i - (317,81 + 4,933 i).$$

La ligne centrale serait alors une horizontale située en 0 et les limites de contrôle seraient situées en :

$$\pm u_{1-\alpha/2} (39,27/\sqrt{5}),$$

soit en $\pm 54,27$, si $\alpha = 0,002$.

Pour les écarts-types, la carte de contrôle serait une carte classique, l'écart-type de référence, σ_0 , étant égal à l'écart-type résiduel, soit 39,27.

3.2. Modélisation de l'autocorrélation

Au paragraphe 3.1. nous avons implicitement considéré que les valeurs successives, \bar{x}_i , répondent au modèle suivant :

$$\bar{x}_i = a + b i + e_i,$$

e_i étant le résidu de la régression. On a supposé également que les e_i sont indépendants.

L'idée de soustraire des données initiales la partie modélisable et d'effectuer le contrôle sur les résidus de cette modélisation peut être étendue à d'autres situations que la régression.

En particulier, on peut faire appel aux techniques de modélisation développées dans le cadre de l'étude des séries chronologiques et notamment aux modèles ARIMA de BOX et JENKINS.

Notre objectif n'est pas de donner ici une description détaillée de ces techniques. Nous renvoyons le lecteur aux ouvrages spécialisés. La référence de base dans ce domaine est certainement le livre de BOX et JENKINS [1976], mais on peut mentionner aussi l'ouvrage de NAZEM [1988], qui s'adresse spécialement à des lecteurs ayant une formation limitée en mathématique et en statistique. Un résumé est également donné par PALM [1987]. Nous ne présenterons que quelques notions de base.

Les premiers outils à définir dans ce contexte sont l'autocorrélogramme et l'autocorrélogramme partiel. Ces outils permettent en effet de vérifier s'il y a indépendance entre données successives et, dans la négative, d'orienter la modélisation de ces dépendances.

Afin de rendre les notations plus conformes à celles généralement utilisées dans l'étude des séries chronologiques, nous allons considérer, dans ce paragraphe, que la suite des données s'écrit :

$$x_1, x_2, \dots, x_t, \dots, x_n,$$

x_t représentant la valeur de la caractéristique au temps t , qu'il s'agisse d'une observation individuelle ou de la moyenne de plusieurs observations.

L'autocorrélation de rang k se définit alors par la relation suivante :

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (k = 1, \dots, K),$$

\bar{x} étant la moyenne de la série.

Ce coefficient mesure l'intensité de la liaison qui existe entre des observations séparées par k intervalles de temps. En pratique, on se limite le plus souvent à l'examen des trois ou quatre premiers coefficients, du moins en l'absence de

phénomènes périodiques. A partir de ces autocorrélations, on peut définir les autocorrélations partielles de rang k , qui s'interprètent de la même façon que les coefficients de corrélation partielle en régression multiple : l'autocorrélation partielle de rang k est la corrélation entre des observations séparées par k intervalles de temps, lorsqu'on élimine de ces observations, l'influence des $k-1$ observations intermédiaires. Les autocorrélogrammes sont des représentations graphiques des autocorrélations en fonction de k .

Si la variable à contrôler suit une distribution normale, on peut tester la signification des coefficients d'autocorrélation. En effet, s'il y a indépendance entre les observations successives, les distributions d'échantillonnage de ces coefficients sont approximativement normales, de moyenne nulle et d'écart-type égal à $1/\sqrt{n}$. On rejette par conséquent l'hypothèse d'indépendance entre les observations si :

$$|r_k| > u_{1-\alpha/2}/\sqrt{n},$$

pour une ou plusieurs valeurs de k .

Lorsqu'il y a autocorrélation et qu'on souhaite en tenir compte pour définir les limites de contrôle, on peut utiliser les modèles ARIMA, car ils sont susceptibles de modéliser des données caractérisées par des autocorrélogrammes de nature très diverses.

Considérons d'abord les données pouvant être décrites par des processus stochastiques gaussiens stationnaires. Dans ce cas, on considère que les observations successives, x_1, x_2, \dots, x_n , ont été générées par une distribution de probabilité multinormale dont les moyennes, les variances et les covariances sont constantes dans le temps. Pour de telles séries, il n'y a donc pas de tendance en moyenne ni en variance et les autocorrélations et autocorrélations partielles ne sont fonction que du nombre d'intervalles de temps, k , qui séparent les observations. De telles données peuvent être modélisées par les modèles ARMA (p, q) dont l'expression générale est la suivante :

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}.$$

Cette expression contient en fait trois parties. La première partie est la partie dite autorégressive d'ordre p et correspond aux p termes du type :

$$\phi_i x_{t-i} \quad (i = 1, \dots, p).$$

Pour cette partie autorégressive, on considère que l'observation au temps t , x_t , est une fonction des p observations antérieures, les ϕ_i étant des paramètres.

La seconde partie du modèle est le résidu au temps t , qui est noté e_t . Les résidus relatifs aux observations successives sont considérés comme étant indépendants et de moyennes nulles. Ils sont comparables aux résidus définis dans les modèles de régression. Si ces résidus ne sont pas de moyennes nulles, il suffit d'ajouter une constante au modèle donné ci-dessus.

La troisième partie du modèle est la partie moyenne mobile d'ordre q et correspond aux q termes du type :

$$\theta_j e_{t-j} \quad (j = 1, \dots, q).$$

Dans les modèles ARMA, les valeurs de p et q sont fonction de la nature des dépendances entre les observations successives, mais en pratique, elles dépassent rarement 2. Le choix de ces valeurs se fait sur la base de l'examen des autocorrélations et des autocorrélations partielles. Il fait partie de la phase d'identification du modèle et peut éventuellement être remis en cause lors de la phase de validation, dont nous parlerons ci-dessous.

A cette phase d'identification succède la phase d'estimation, au cours de laquelle on détermine les valeurs numériques à donner aux paramètres ϕ_i et θ_j . Cette estimation se fait par la mise en oeuvre d'algorithmes des moindres carrés non linéaires.

Enfin, lorsqu'on dispose des estimations de ces paramètres, il reste à vérifier si le modèle est adéquat. Cette phase de validation est essentiellement basée sur l'analyse des erreurs de prévision à l'horizon 1 :

$$\hat{e}_t = x_t - \hat{x}_{t-1}(1),$$

$\hat{x}_{t-1}(1)$ représentant la prévision faite au temps $t-1$ pour l'horizon 1, c'est-à-dire pour le temps t , cette prévision étant réalisée à partir du modèle.

Si on trouve que ces erreurs de prévision présentent encore des corrélations importantes, on revient à la phase d'identification, on modifie les valeurs de p et/ou de q et on recommence la séquence.

Les modèles ARMA que nous venons de présenter très brièvement concernent, nous l'avons signalé plus haut, des séries stationnaires. La méthodologie peut être étendue sans difficulté aux séries avec une tendance en moyenne et conduit aux modèles ARIMA. Ces modèles sont en fait des modèles ARMA appliqués, non pas aux séries brutes, mais aux séries différenciées. Si la tendance est linéaire, on peut montrer que la différenciation première :

$$w_t = x_t - x_{t-1},$$

élimine la tendance et la série w_t peut être décrite par les modèles ARMA. Si la tendance est non linéaire, deux ou plusieurs différenciations successives permettent d'obtenir une série sans tendance. Lorsqu'on effectue des prévisions dans le cas de séries avec tendance, on fait d'abord une prévision de la série différenciée ; ensuite on repasse à la série initiale par le calcul d'une somme, c'est-à-dire aussi par une intégration, ce qui justifie la dénomination de modèles intégrés qu'on leur donne. On les représente par le sigle ARIMA (p, d, q), d étant l'ordre de la différenciation ($d = 1$ pour une différenciation première, $d = 2$ pour une différenciation seconde, etc.). En pratique, la valeur d dépasse rarement 2 et la détermination de d fait partie de la phase d'identification.

Disposant de la série des erreurs de prévision à l'horizon 1, \hat{e}_t , considérées comme des estimations des résidus, on peut construire les cartes de contrôle classiques pour données individuelles qui ont été décrites précédemment [PALM, 1996]. En particulier, pour le contrôle de la moyenne du procédé, la ligne centrale correspond à la moyenne des résidus :

$$\bar{e} = \frac{1}{n} \sum_{t=1}^n \hat{e}_t,$$

qui est généralement proche de zéro et les limites de contrôle sont situées en :

$$\bar{e} \pm u_{1-\alpha/2} \hat{\sigma}_e,$$

la valeur $u_{1-\alpha/2}$ pouvant être remplacée par k , si on souhaite exprimer les limites de contrôle en fonction d'un multiple de l'erreur-standard.

Dans cette relation, $\hat{\sigma}_e$ est une estimation de l'écart-type résiduel du modèle, cette caractéristique étant obtenue lors de la phase d'estimation des paramètres du modèle ARIMA.

Une telle carte de contrôle peut utilement être accompagnée de la carte de contrôle construite sur les données originales, cette dernière étant plus parlante pour la visualisation de la dynamique du procédé contrôlé, mais ne servant pas à déclencher une intervention sur le procédé lorsqu'un point se situe hors des limites de contrôle.

MONTGOMERY et MASTRANGELO [1991] ont d'ailleurs suggéré de combiner en une seule carte de contrôle l'information fournie par la carte de contrôle sur les résidus et par la carte de contrôle sur les données initiales. Ils proposent, en effet, de prendre comme ligne centrale les valeurs successives de $\hat{x}_{t-1}(1)$, c'est-à-dire les prédictions faites par le modèle au temps $t-1$ pour le temps t , et comme limites de contrôle les valeurs :

$$\hat{x}_{t-1}(1) \pm u_{1-\alpha} \hat{\sigma}_e.$$

La construction des cartes de contrôle après modélisation de la série par les modèles ARIMA est une technique relativement lourde à mettre en oeuvre à cause des difficultés liées à l'identification des modèles. Il faut remarquer cependant qu'il existe des logiciels d'analyse des séries chronologiques opérant le choix d'un modèle ARIMA de façon entièrement automatique, ce qui simplifie évidemment l'étape de modélisation.

D'autre part, diverses études relatives aux séries chronologiques ont montré que, pour de nombreuses séries présentant des autocorrélations positives et une moyenne n'évoluant pas de façon trop rapide, de bonnes prévisions à l'horizon 1 sont obtenues par le lissage exponentiel. Dans le cas particulier du modèle ARIMA (0,1,1) :

$$x_t = x_{t-1} + e_t - \theta e_{t-1},$$

les prévisions à l'horizon 1 sont d'ailleurs égales aux prévisions à l'horizon 1 obtenues à partir du lissage exponentiel défini au paragraphe 2.3, en prenant comme paramètre de lissage la valeur :

$$\lambda = 1 - \theta.$$

Sous les conditions qui viennent d'être énoncées, le lissage exponentiel simple peut dès lors remplacer la modélisation par les modèles ARIMA et la carte de contrôle est réalisée sur les erreurs de prévision à l'horizon 1 [MONTGOMERY et MASTRANGELO, 1991].

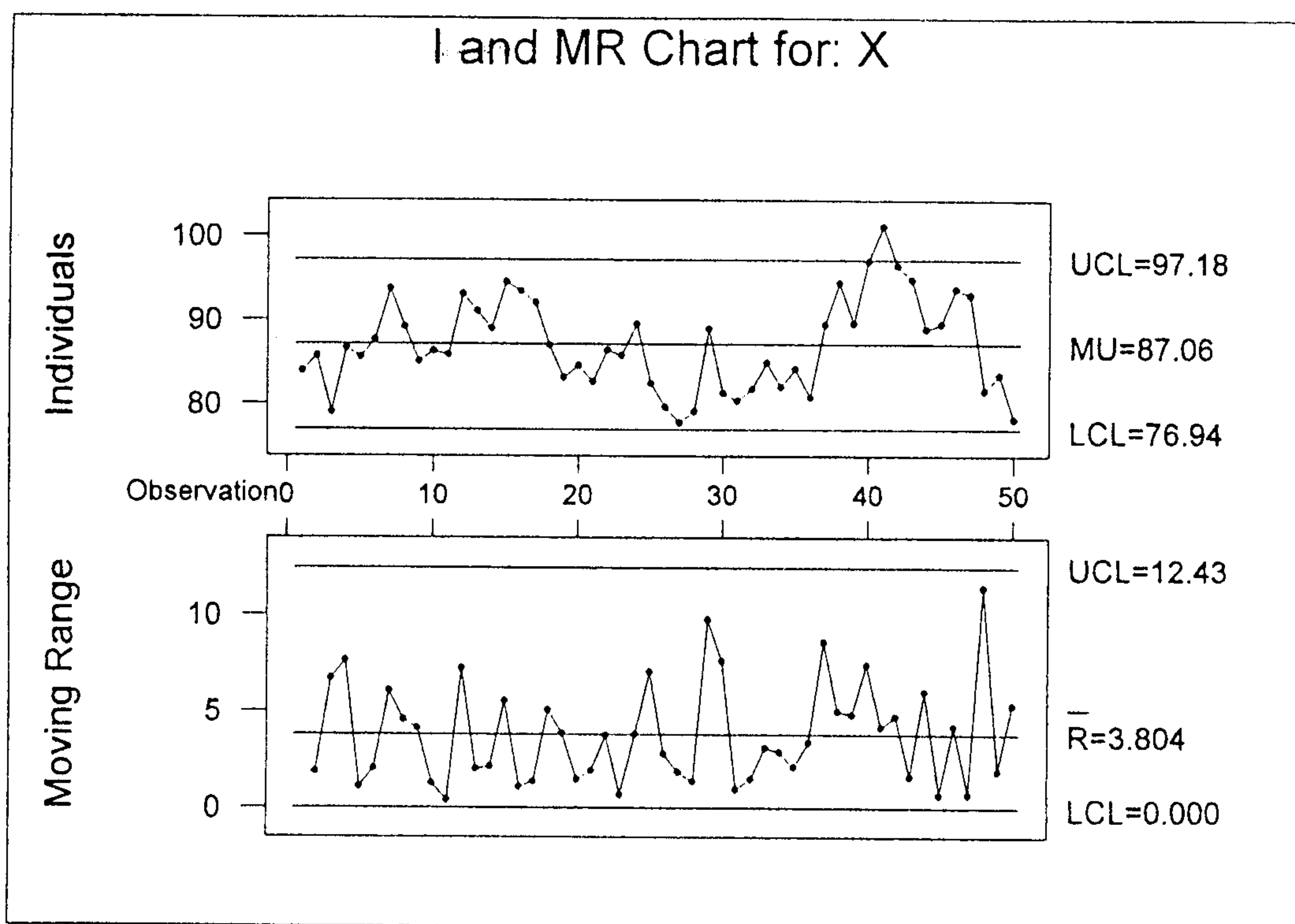


Figure 8. Cartes de contrôle des moyennes et des amplitudes mobiles : données brutes simulées à partir du modèle autorégressif d'ordre 1.

Une telle carte s'apparente donc aux cartes de contrôle décrites au paragraphe 2.3 pour des données non corrélées. Il ne faut cependant pas perdre de vue que, dans le cas des données non corrélées, on porte sur le graphique les valeurs lissées, \bar{x}_t^* , alors que, dans le cas des données corrélées, on porte les résidus de prévision c'est-à-dire aussi, en reprenant les notations du paragraphe 2.3, les valeurs :

$$\hat{e}_t = x_t - \bar{x}_{t-1}^*.$$

Pour illustrer le contrôle dans le cas de données autocorrélées, nous avons simulé 50 observations provenant d'un modèle autorégressif d'ordre 1 :

$$x_t = 13,04 + 0,847 x_{t-1} + e_t,$$

les e_t successifs étant des valeurs observées d'une variable normale de moyenne nulle et d'écart-type égal à 3,9. Ce modèle a été proposé par MONTGOMERY et MASTRANGELO [1991] pour l'analyse d'observations relatives à un procédé chimique particulier.

La figure 8 donne les cartes de contrôle pour données individuelles obtenues par Minitab, à l'aide de la commande %IMRCHART. Des informations concernant la construction de ces cartes sont données dans PALM [1996].

La moyenne du procédé a été estimée à partir de la moyenne des 50 observations et l'écart-type a été obtenu en calculant la moyenne des amplitudes mobiles

déterminées sur deux observations successives et en divisant cette moyenne par 1,12838, qui est l'espérance mathématique de la distribution d'échantillonnage de l'amplitude d'un échantillon aléatoire et simple de deux observations prélevé dans une population normale réduite. Les estimations suivantes ont été obtenues :

$$\hat{m}_0 = 87,06 \quad \text{et} \quad \hat{\sigma}_0 = 3,3712.$$

La carte des valeurs individuelles montre qu'une observation est en dehors des limites de contrôle et l'utilisation des règles complémentaires fait apparaître à plusieurs reprises l'absence de stabilité du procédé. En réalité, le procédé semble ne pas être sous contrôle car on n'a pas tenu compte des corrélations entre observations successives, qui sont importantes, puisque les autocorrélations de rang 1, 2 et 3 sont, respectivement, égales à 0,62, 0,42 et 0,26.

L'ajustement du modèle AR(1) a conduit aux résultats suivants :

$$x_t = 29,46 + 0,660 x_{t-1} + \hat{e}_t,$$

l'écart-type des résidus étant égal à 4,28. Le modèle ajusté est quelque peu différent du modèle théorique utilisé pour la simulation. Une telle discordance peut s'expliquer par le nombre relativement faible de données simulées.

La figure 9 donne les cartes de contrôle des résidus, établies comme dans le cas de la figure 8, et on constate qu'aucun point ne se trouve hors des limites de contrôle et donc que le procédé est bien sous contrôle.

La série d'observations a également été soumise au lissage exponentiel, en donnant au paramètre de lissage, λ , la valeur 0,707. Cette valeur a été obtenue en ajustant le modèle ARIMA(0,1,1) et en utilisant la relation existant entre λ et θ qui a été présentée ci-dessus. Les cartes de contrôle sont données à la figure 10. On peut constater que les résidus du lissage et les résidus du modèle AR(1) ont un comportement fort semblable. La corrélation entre les deux séries de résidus est d'ailleurs égale à 0,93 ; il en résulte que les cartes de contrôle des figures 9 et 10 sont tout à fait comparables.

4. CARTES DE CONTRÔLE POUR DONNÉES MULTIVARIÉES

4.1. Intérêt du contrôle multivarié

Jusqu'à présent, nous avons considéré que le contrôle du procédé se base sur une seule caractéristique quantitative. Il peut arriver cependant que plusieurs caractéristiques soient observées de manière à en suivre l'évolution simultanée.

A ce sujet, on notera tout d'abord que l'examen de cartes de contrôle séparées pour les diverses caractéristiques n'est pas une solution adéquate.

En effet, considérons d'abord le cas d'un procédé caractérisé par deux variables non corrélées. Supposons, en outre, que les distributions d'échantillonnage des moyennes d'échantillons successifs correspondent à deux variables normales

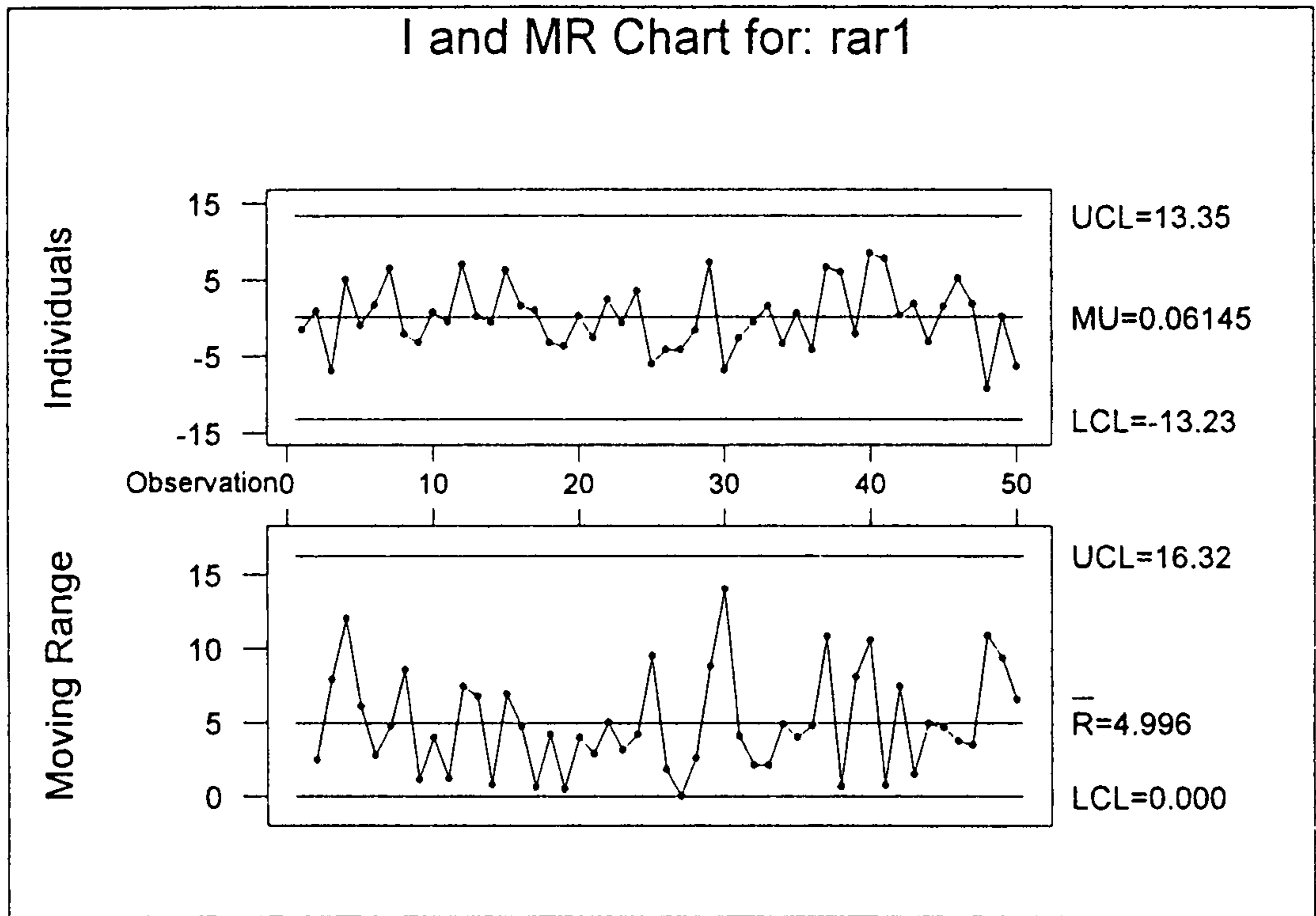


Figure 9. Cartes de contrôle des moyennes et des amplitudes mobiles : résidus obtenus après ajustement du modèle AR(1).

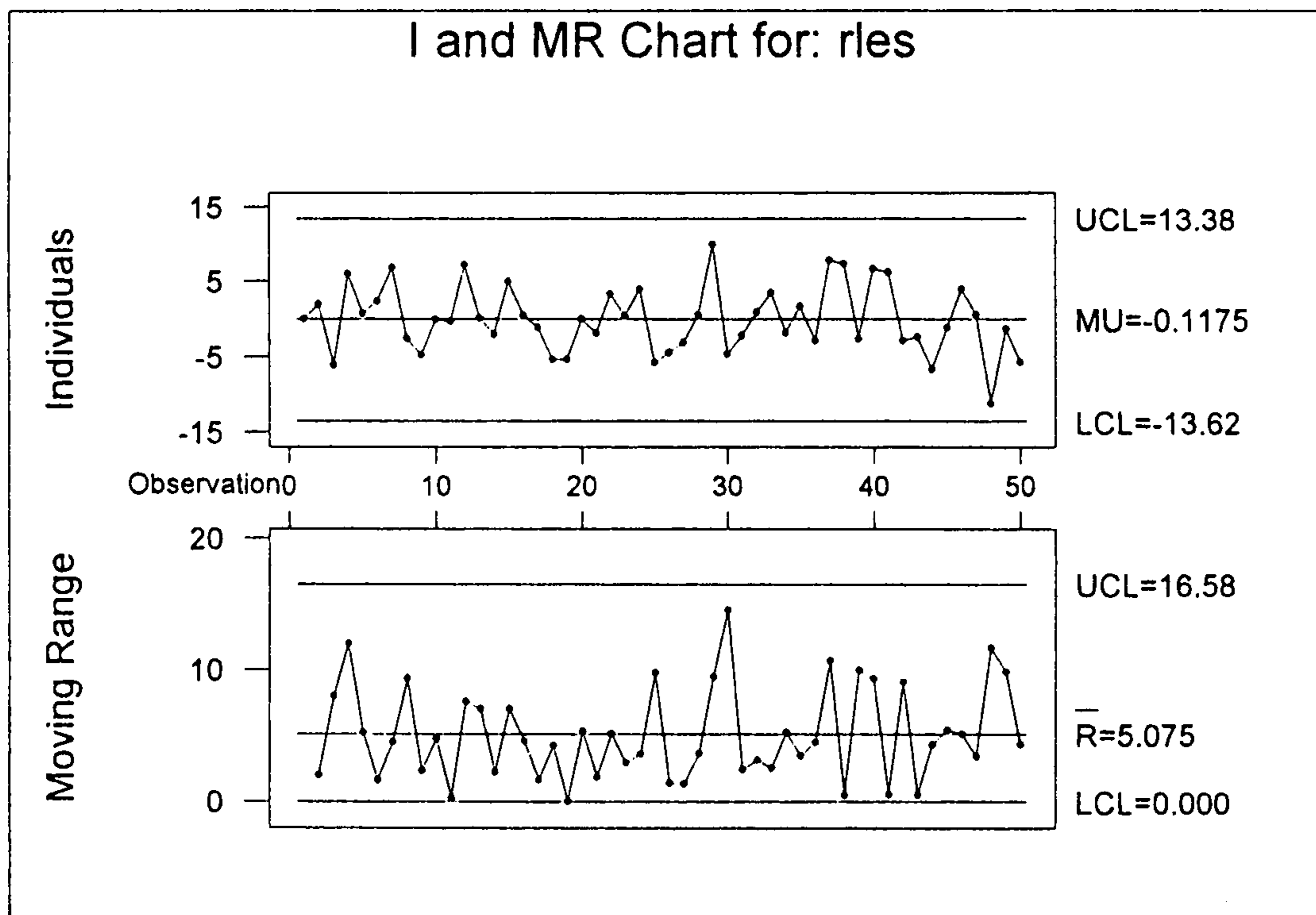


Figure 10. Cartes de contrôle des moyennes et des amplitudes mobiles : résidus obtenus après lissage exponentiel.

réduites, U_1 et U_2 . Si on examine indépendamment chacune des variables et pour un risque α égal à 0,002, les limites de contrôle se situent en $-3,09$ et en $3,09$. Dans ces conditions :

$$P(-3,09 < U_1 < 3,09) = P(-3,09 < U_2 < 3,09) = (1 - \alpha) = 0,998,$$

et la probabilité qu'aucune des deux variables ne soit hors des limites de contrôle s'écrit :

$$P(-3,09 < U_1 < 3,09 \quad \text{et} \quad -3,09 < U_2 < 3,09) = (1 - \alpha)^2 \cong 0,996,$$

puisque les deux caractéristiques sont indépendantes. La probabilité que le procédé soit déclaré hors de contrôle alors qu'il est sous contrôle est donc à peu près égale à 0,004.

On pourrait bien sûr ramener cette probabilité à une valeur proche de la valeur nominale α . Il suffirait de considérer, pour les limites de contrôle de chacune des variables, un risque égal à $\alpha/2$. Cela reviendrait donc à remplacer $u_{1-\alpha/2}$ par $u_{1-\alpha/4}$. Pour $\alpha = 0,002$, les limites de contrôle par variable se situeraient alors en $\pm 3,29$. Les intervalles de contrôle ainsi modifiés s'appellent intervalles de BONFERRONI⁽¹²⁾ [ALT, 1982]. La généralisation au cas de p variables indépendantes est immédiate ; il suffit, en effet, de remplacer $u_{1-\alpha/2}$ par $u_{1-\alpha/2p}$.

La solution n'est cependant pas encore satisfaisante, car des couples d'observations correspondant à des valeurs identiques de la densité de probabilité à deux dimensions peuvent conduire tantôt à accepter, tantôt à rejeter l'hypothèse que le procédé est sous contrôle. Ainsi, considérons, par exemple, un échantillon dont les deux moyennes seraient $u_{11} = 3,0$ et $u_{21} = 3,0$ et un second échantillon dont les moyennes seraient $u_{12} = 4,24$ et $u_{22} = 0$. Le premier couple de moyennes est dans les limites modifiées de contrôle pour chacune des variables alors que le deuxième couple est hors des limites de contrôle pour la variable u_1 . Et pourtant, la densité de probabilité à deux dimensions, $\phi(u_1, u_2)$, qui est égale au produit des densités de probabilité marginales lorsque les variables sont indépendantes, est identique pour les deux couples :

$$\phi(3) \phi(3) = \phi(4,24) \phi(0) \cong 0,00002.$$

La solution correcte pour définir les limites de contrôle est de déterminer, dans l'espace à deux dimensions, le cercle de rayon R tel que :

$$P(U_1^2 + U_2^2 \leq R^2) = 1 - \alpha.$$

Ce cercle, appelé cercle de concentration, est le lieu géométrique des points de même densité de probabilité à deux dimensions.

On sait que la somme des carrés de deux variables normales réduites indépendantes est une variable χ^2 à 2 degrés de liberté ; on trouve donc que :

$$R^2 = \chi_{1-\alpha}^2,$$

et, pour une valeur de α égale à 0,002, R vaut 3,53.

⁽¹²⁾ En anglais : *BONFERRONI intervals*.

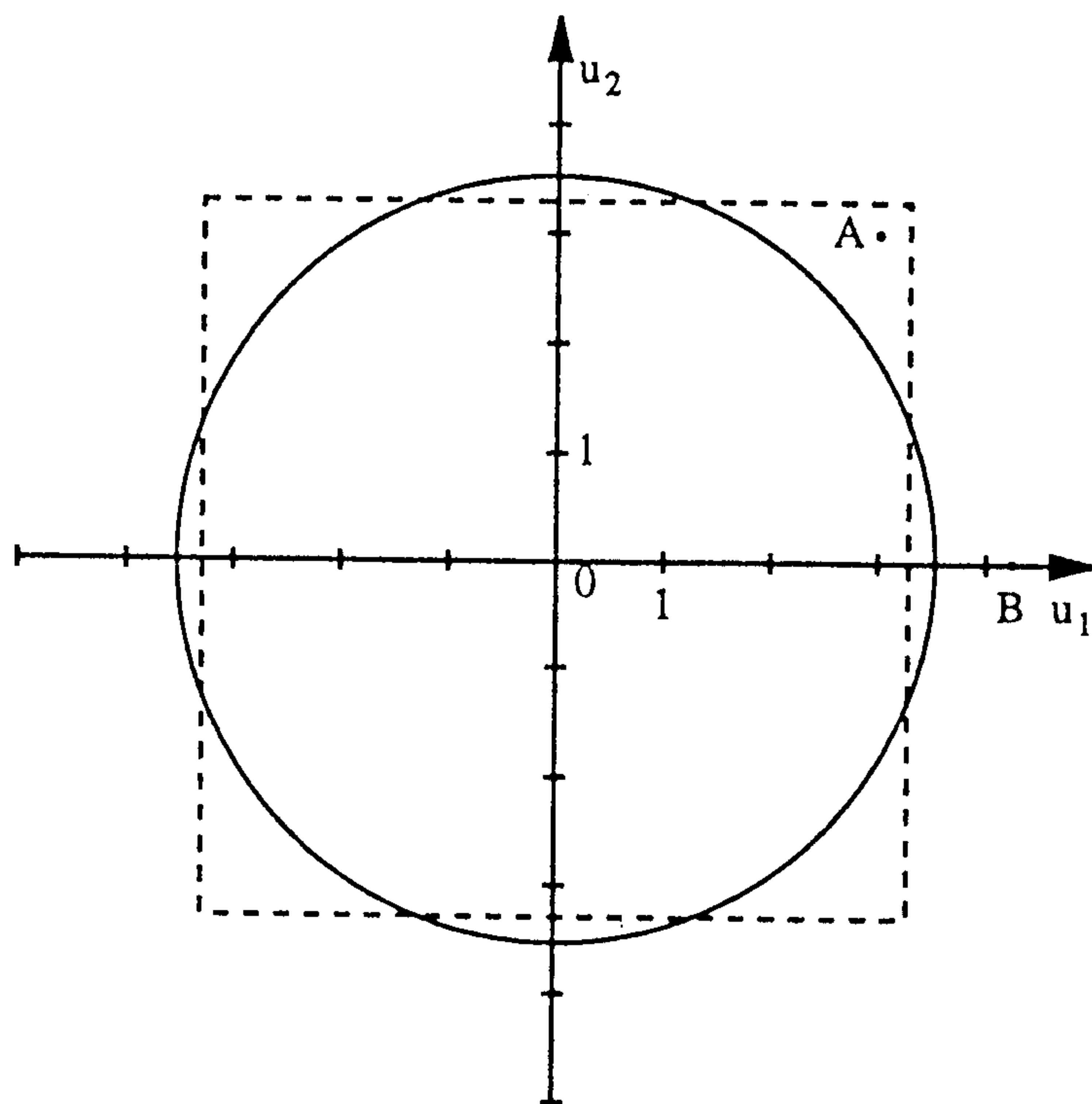


Figure 11. Intervalles de BONFERRONI et cercle de concentration, dans le cas de deux variables normales réduites indépendantes ($\alpha = 0,002$).

La figure 11 donne une représentation graphique de la zone exacte d'acceptation de l'hypothèse que le procédé est sous contrôle, représentée par le cercle, ainsi que la zone carrée définie par les limites modifiées de contrôle des deux variables considérées séparément. La figure 11 matérialise également la position des deux échantillons considérés ci-dessus (points A et B). On constate que pour chacun des échantillons, le procédé doit être considéré comme n'étant pas sous contrôle.

L'exemple ci-dessus correspond à une situation particulière, du fait de la non corrélation des variables U_1 et U_2 . De façon plus générale, les limites de contrôle exactes sont données, dans l'espace (u_1, u_2) , par des ellipses de concentration dont les axes ne sont pas parallèles aux axes des coordonnées. La discordance entre la probabilité qu'un couple d'observations se trouve dans cette ellipse et la probabilité que le couple se trouve dans le rectangle construit à partir des limites modifiées de contrôle, calculées variable par variable, est encore accentuée du fait de cette corrélation, ce qui justifie l'intérêt de l'utilisation de méthodes multivariées exactes.

4.2. Contrôle du vecteur des moyennes

L'approche multivariée du contrôle repose sur l'hypothèse d'une distribution normale à p variables, de vecteur de moyennes m_0 et de matrice de variances et covariances, appelée aussi matrice de dispersion, égale à Σ_0 .

Si les paramètres de cette distribution sont connus, l'hypothèse de la conformité du vecteur de moyenne d'un échantillon de n observations :

$$H_0 : m_i = m_0 ,$$

peut être testée en calculant la quantité :

$$\chi_{obs}^2 = n(\bar{x}_i - m_0)' \Sigma_0^{-1} (\bar{x}_i - m_0) ,$$

qui possède, si le procédé est sous contrôle, une distribution χ^2 à p degrés de liberté [DAGNELIE, 1975].

Par conséquent, une carte de contrôle du vecteur des moyennes peut être obtenue en portant sur un graphique les valeurs χ_{obs}^2 relatives aux échantillons successifs. La ligne centrale se trouve en p , la limite de contrôle inférieure se situe en zéro et la limite de contrôle supérieure se situe en $\chi_{1-\alpha}^2$.

Si les paramètres de la distribution multinormale ne sont pas connus, ils peuvent être estimés à partir des données du passé, comme dans le cas des cartes de contrôle univariées [PALM, 1996]. Différentes solutions peuvent être envisagées. Ainsi par exemple, si on dispose de h échantillons de n observations prélevés à intervalles réguliers, des estimations non biaisées du vecteur des moyennes et de la matrice des variances et covariances sont données par la moyenne des vecteurs des moyennes, \bar{x}_i , et par la moyenne des matrices de variances et covariances estimées, S_i :

$$\hat{m}_0 = \sum_{i=1}^h \bar{x}_i / h \quad \text{et} \quad \hat{\Sigma}_0 = \sum_{i=1}^h S_i / h .$$

Dans ces conditions, et si le procédé est sous contrôle, l'expression :

$$T^2 = n(\bar{x}_i - \hat{m}_0)' \hat{\Sigma}_0^{-1} (\bar{x}_i - \hat{m}_0)$$

est liée aux distributions de HOTELLING, qui sont, elles-mêmes, liées aux distributions F de FISHER-SNEDECOR [DAGNELIE, 1975]. De façon plus précise, la quantité :

$$F_{obs} = n(\bar{x} - \hat{m}_0)' \hat{\Sigma}_0^{-1} (\bar{x} - \hat{m}_0) \frac{h(n-1) - p + 1}{p(h+1)(n-1)} ,$$

possède une distribution F à $k_1 = p$ et $k_2 = h(n-1) - p + 1$ degrés de liberté [ALT, 1985].

La limite inférieure de contrôle de la carte des valeurs F est, par conséquent, égale à zéro et la limite supérieure de contrôle est égale à $F_{1-\alpha}$, la ligne centrale se situant en $k_2/(k_2 - 2)$, c'est-à-dire pratiquement en 1.

Si le nombre de degrés de liberté du dénominateur, k_2 , est suffisamment grand, la distribution de la variable T^2 de HOTELLING peut d'ailleurs être assimilée directement à une distribution χ^2 à p degrés de liberté.

4.3. Contrôle de la dispersion

En désignant par A_i les matrices des sommes de carrés et de produits des écarts du $i^{\text{ième}}$ échantillon :

$$A_i = (n - 1) S_i,$$

le test de conformité de la matrice de dispersion :

$$H_0 : \Sigma_i = \Sigma_0,$$

peut être réalisé, de façon approximative, en calculant l'expression :

$$\chi_{obs}^2 = -pn + pn \ln(n) - n \ln(|A_i| / |\Sigma_0|) + \text{tr}(\Sigma_0^{-1} A_i),$$

qui, si l'hypothèse nulle est vraie, possède une distribution χ^2 à $p(p+1)/2$ degrés de liberté [ALT, 1985]. Comme pour la carte de contrôle de la moyenne, on porte sur un graphique les valeurs successives de χ_{obs}^2 , la ligne centrale étant en $p(p+1)/2$, la limite de contrôle inférieure en 0 et la limite de contrôle supérieure en $\chi_{1-\alpha}^2$.

Une autre solution se base sur la distribution d'échantillonnage de la variance généralisée, c'est-à-dire du déterminant de la matrice de variances et covariances.

L'espérance mathématique et l'erreur-standard de la distribution d'échantillonnage de la variance généralisée, $|S_i|$, d'un échantillon de n observations sont données par les relations suivantes [ALT, 1985] :

$$E(|S_i|) = b_1 |\Sigma_0| \quad \text{et} \quad V(|S_i|) = b_2 |\Sigma_0|^2,$$

avec :

$$b_1 = (n - 1)^{-p} \prod_{i=1}^p (n - i)$$

$$\text{et} \quad b_2 = (n - 1)^{-2p} \prod_{i=1}^p (n - i) \left[\prod_{j=1}^p (n - j + 2) - \prod_{j=1}^p (n - j) \right].$$

La ligne centrale se trouve alors en $b_1 |\Sigma_0|$ et les limites de contrôle se situent en :

$$b_1 |\Sigma_0| \pm k \sqrt{b_2} |\Sigma_0|,$$

en remplaçant la limite inférieure par zéro si la formule ci-dessus conduit à une valeur négative. Cette solution ne permet cependant pas de chiffrer exactement le risque α , car la distribution d'échantillonnage de la variance généralisée n'est pas exactement normale.

Si la matrice Σ_0 est inconnue, on remplace, dans les formules ci-dessus, les valeurs théoriques par les estimations correspondantes. Ainsi, la matrice de variances et covariances, Σ_0 , est remplacée par la valeur définie au paragraphe précédent :

$$\hat{\Sigma}_0 = \sum_{i=1}^h S_i / h$$

et la variance généralisée $|\Sigma_0|$ est remplacée par l'estimation non biaisée suivante :

$$|\hat{\Sigma}_0^*| = \frac{1}{h} \sum_{i=1}^h (|S_i|/b_1) = |\hat{\Sigma}_0|/b_1.$$

La ligne centrale de la carte de contrôle sera donc située en $|\hat{\Sigma}_0|$, c'est-à-dire aussi en $b_1 |\hat{\Sigma}_0^*|$, et les limites de contrôle se placeront en :

$$b_1 |\hat{\Sigma}_0^*| \pm k \sqrt{b_2} |\hat{\Sigma}_0^*|.$$

4.4. Exemple

Pour illustrer les principes décrits ci-dessus, nous avons repris l'exemple proposé par MONTGOMERY [1991] concernant la force de tension (en livres par pouce carré) et le diamètre de fibres textiles (en centièmes de pouce). Au cours d'une phase préliminaire, 20 échantillons de 10 fibres ont été prélevés et les estimations suivantes ont été calculées :

$$\hat{m}_0 = \begin{pmatrix} 115,59 \\ 1,06 \end{pmatrix} \quad \text{et} \quad \hat{\Sigma}_0 = \begin{pmatrix} 1,23 & 0,79 \\ 0,79 & 0,83 \end{pmatrix}.$$

Le vecteur des moyennes contient les moyennes pour les deux caractéristiques des 200 fibres, et la matrice de variances et covariances est la moyenne des matrices de variances et covariances des 20 échantillons.

Supposons que, en phase de contrôle, le $i^{\text{ème}}$ échantillon de 10 observations soit caractérisé par les valeurs suivantes :

$$\bar{x}_i = \begin{pmatrix} 114,98 \\ 1,05 \end{pmatrix} \quad \text{et} \quad S_i = \begin{pmatrix} 1,25 & 0,75 \\ 0,75 & 0,78 \end{pmatrix},$$

et qu'on désire vérifier si les caractéristiques des fibres sont bien sous contrôle.

Si on considère que m_0 et Σ_0 sont suffisamment proches des paramètres théoriques du procédé, le contrôle du vecteur des moyennes et de la matrice de dispersion peut se faire par l'intermédiaire des distributions χ^2 .

Pour le vecteur des moyennes, il suffit de comparer la valeur :

$$\chi_{obs}^2 = 10 \begin{bmatrix} 114,98 - 115,59 \\ 1,05 - 1,06 \end{bmatrix}' \begin{bmatrix} 1,23 & 0,79 \\ 0,79 & 0,83 \end{bmatrix}^{-1} \begin{bmatrix} 114,98 - 115,59 \\ 1,05 - 1,06 \end{bmatrix}$$

à la valeur $\chi_{0,998}^2$, avec 2 degrés de liberté.

On a, pour α égal à 0,002 :

$$\chi_{obs}^2 = 7,54 < \chi_{0,998}^2 = 12,43.$$

On accepte donc l'hypothèse que le procédé est sous contrôle du point de vue de la moyenne.

La procédure utilisée n'est qu'approximative dans la mesure où le vecteur des moyennes et la matrice des variances et covariances théoriques ne sont pas connus mais estimés. De façon plus rigoureuse, il conviendrait de calculer la valeur :

$$F_{obs} = 7,54 \frac{20(10-1) - 2 + 1}{2(20+1)(10-1)} = \frac{179}{378} 7,54 = 3,57,$$

et de la comparer à $F_{0,998}$ avec 2 et 179 degrés de liberté, soit 6,44, ce qui permettrait également de conclure que le procédé est bien sous contrôle du point de vue de la moyenne.

On peut constater aussi que le contrôle sur la base de la valeur F est pratiquement identique au contrôle sur la base de la valeur χ^2 . En effet :

$$F_{obs}/F_{0,998} \cong 0,56 \quad \text{et} \quad \chi_{obs}^2/\chi_{0,998}^2 = 0,61.$$

Dans le premier cas, le point est, sur la carte de contrôle, à 56 % de la limite de contrôle, alors qu'il est à 61 % de la limite de contrôle dans le second cas. Cette faible différence est liée au nombre élevé d'observations réalisées en vue de l'estimation des paramètres théoriques. Supposons, par exemple, que les paramètres \hat{m}_0 et $\hat{\Sigma}_0$ aient été calculés à partir de 10 échantillons de 5 observations et que, en phase de contrôle, l'échantillon ayant donné \bar{x}_i et S_i soit également de 5 observations, on aurait obtenu des rapports de 0,23 et 0,30, respectivement pour F et pour χ^2 , ce qui correspond à une différence relative plus grande que dans le cas ci-dessus.

En ce qui concerne le test de conformité de la matrice de dispersion, on peut calculer la valeur χ_{obs}^2 , définie au paragraphe 4.3, en remplaçant cependant la variance généralisée $|\Sigma_0|$ par son estimation non biaisée :

$$\begin{aligned} |\hat{\Sigma}_0^*| &= [(1,23)(0,83) - (0,79)(0,79)] / [(9)(8)/(9^2)] \\ &= 0,3968/0,8889 = 0,4464, \end{aligned}$$

On a, en outre,

$$A_i = 9 \begin{bmatrix} 1,25 & 0,75 \\ 0,75 & 0,78 \end{bmatrix} = \begin{bmatrix} 11,25 & 6,75 \\ 6,75 & 7,02 \end{bmatrix},$$

$$|A_i| = 33,4125,$$

$$\hat{\Sigma}_0^{-1} A_i = \begin{pmatrix} 10,0932 & 0,1429 \\ -1,4743 & 8,3219 \end{pmatrix},$$

et $\text{tr}(\hat{\Sigma}_0^{-1} A_i) = 18,4151.$

On obtient alors :

$$\begin{aligned} \chi_{obs}^2 &= -(2)(10) + (2)(10) \ln(10) - 10 \ln(33,4125/0,4464) + 18,4151 \\ &= -20 + 46,0517 - 43,1547 + 18,4151 = 1,3121. \end{aligned}$$

Cette valeur doit être comparée à la valeur $\chi_{0,998}^2$ à 3 degrés de liberté, soit :

$$\chi_{0,998}^2 = 14,80,$$

et on conclut que le procédé est sous contrôle du point de vue de la dispersion.

Pour la variance généralisée, la ligne centrale se situe en 0,3968 et les limites de contrôle, pour $k = 3$, sont en :

$$0,3968 \pm (3)(0,4464)\sqrt{0,4170},$$

puisque :

$$\begin{aligned} b_2 &= (10 - 1)^{-4} [(10 - 1)(10 - 2)] [(10 - 1 + 2)(10 - 2 + 2) - (10 - 1)(10 - 2)] \\ &= 0,4170. \end{aligned}$$

La limite de contrôle inférieure est remplacée par zéro car la formule ci-dessus donne une valeur négative et la limite de contrôle supérieure vaut 1,26. On conclut donc que le procédé est sous contrôle du point de vue de la dispersion, puisque la variance généralisée de l'échantillon $|S_i|$, qui est égale à 0,4125, est bien située entre les limites de contrôle.

Si on disposait d'une succession d'échantillons prélevés à intervalles réguliers pendant la phase de contrôle, on pourrait, bien entendu, établir les cartes de contrôle : il suffirait de répéter, pour chacun des échantillons, les calculs qui viennent d'être réalisés et de porter les résultats sur un graphique.

5. QUELQUES INFORMATIONS COMPLÉMENTAIRES

Pour des données indépendantes, nous avons décrit, dans la présente note, trois procédures basées sur la combinaison des résultats de plusieurs échantillons : le contrôle des moyennes mobiles, le contrôle après lissage exponentiel et le contrôle sur les sommes cumulées. Nous avons également signalé que, par rapport aux cartes de contrôle de SHEWHART, ces cartes avaient l'avantage de permettre une détection plus rapide des modifications de moyennes inférieures à 1,5 ou 2 fois l'erreur-standard, du moins si elles sont conçues de façon adéquates.

MONTGOMERY [1991] signale que le contrôle sur les moyennes mobiles est en général moins performant que le contrôle après lissage exponentiel ou le contrôle sur les sommes cumulées, dans le cas de faibles modifications de moyennes. Cet auteur a également constaté que le contrôle après lissage est souvent supérieur au contrôle sur les sommes cumulées pour détecter des modifications plus importantes de la moyenne, surtout si λ est plus grand que 0,1.

Afin d'assurer une bonne sensibilité du contrôle, à la fois pour des modifications faibles et pour des modifications importantes des moyennes, MONTGOMERY [1991] suggère de combiner les cartes des moyennes proposées par SHEWHART avec le contrôle après lissage exponentiel. Dans ces cas, l'auteur propose d'écarter

légèrement les limites de contrôle, en donnant à la constante k , qui intervient dans la méthode de l'erreur-standard, une valeur de 3,25 ou même 3,50, au lieu de la valeur 3, traditionnellement retenue.

Au paragraphe 2, les cartes de contrôle ont été présentées dans le cas du contrôle par mesures. Des adaptations peuvent cependant être envisagées pour les contrôles par comptages. Considérons, par exemple, la situation où on dénombre les défauts sur des lots comportant un nombre constant d'articles. Pour une telle situation, nous avons vu que, en utilisant l'approximation normale, les limites de contrôle s'écrivent :

$$y_0 \pm u_{1-\alpha/2} \sqrt{y_0},$$

y_0 étant le nombre moyen de défauts par lot. Si on soumet les valeurs observées successives, y_i , au lissage exponentiel :

$$y_i^* = \lambda y_i + (1 - \lambda) y_{i-1}^*,$$

les limites de contrôle tendront vers les valeurs :

$$y_0 \pm u_{1-\alpha/2} \sqrt{\frac{\lambda y_0}{2 - \lambda}}.$$

Des références bibliographiques pour l'utilisation des cartes de sommes cumulées pour des distributions autres que la normale sont données par GOEL [1982].

D'autre part, pour le contrôle de données multivariées, nous nous sommes limités à la présentation d'une ou deux méthodes (paragraphe 4) mais d'autres solutions peuvent être envisagées. Il s'agit notamment de l'établissement de cartes basées sur la généralisation à p dimensions du lissage exponentiel⁽¹³⁾ et des sommes cumulées⁽¹⁴⁾. Une revue bibliographique de ces méthodes multivariées de contrôle a été réalisée par WIERDA [1994].

De façon générale, lorsqu'un procédé est considéré comme n'étant pas sous contrôle, se pose le problème de rechercher les raisons de cette situation et, plus particulièrement, la ou les variables responsables du signal. Une solution est évidemment d'analyser les cartes de contrôle univariées. Mais, comme nous l'avons vu au paragraphe 4.1, il peut arriver que ce soit la combinaison de deux ou plusieurs variables qui provoque le signal, chacune des variables prise indépendamment étant sous contrôle. Diverses techniques d'analyse multivariée peuvent être envisagées et notamment l'analyse en composantes principales [JACKSON, 1991].

Enfin, on notera que, ni le logiciel SAS, ni le logiciel Minitab n'offrent la possibilité de construire, de façon automatique, les cartes de contrôle multivariées. Pour les données bivariées, un programme SAS est cependant proposé dans la documentation SAS [SAS, 1995].

⁽¹³⁾ En anglais : *multivariate exponentially weighted moving average chart, MEWMA chart.*

⁽¹⁴⁾ En anglais : *multivariate sum chart, MCUSUM chart.*

BIBLIOGRAPHIE

- ALT F.B. [1982]. BONFERRONI inequalities and intervals. In : KOTZ S., JOHNSON N.L. (edit.). *Encyclopedia of statistical sciences* (vol. 1). New York, Wiley, 294-300.
- ALT F.B. [1985]. Multivariate quality control. In : KOTZ S., JOHNSON N.L. (edit.). *Encyclopedia of statistical sciences* (vol. 6). New York, Wiley, 110-112.
- BOX G.E.P., JENKINS G.M. [1976]. *Time series analysis: forecasting and control*. San Francisco, Holden-Day, 575 p.
- DAGNELIE P. [1975]. *Analyse statistique à plusieurs variables*. Gembloux, Presses Agronomiques, 362 p.
- FALTIN F.W., WOODALL W.H. [1991]. Some statistical process control methods for autocorrelated data : discussion. *Technometrics* 23, 194-197.
- GOEL A.L. [1982]. Cumulative sum control charts. In : KOTZ S., JOHNSON N.L. (edit.). *Encyclopedia of statistical sciences* (vol. 2). New York, Wiley, 233-241.
- GRANT E.L., LEAVENWORTH R.S. [1980]. *Statistical quality control*. New York, McGraw-Hill, 684 p.
- JACKSON J.E. [1991]. *A user's guide to principal components*. New York, Wiley, 569 p.
- JOHNSON N.L. [1961]. A simple theoretical approach to cumulative sum control charts. *J. Amer. Stat. Assoc.* 56, 835-840.
- LUCAS J.M., CROSIER R.B. [1982]. Fast initial response for CUSUM quality schemes : give your CUSUM a head start. *Technometrics* 24, 199-205.
- MONTGOMERY D.C. [1991]. *Introduction to statistical quality control*. New York, Wiley, 674 p.
- MONTGOMERY D.C., MASTRANGELO C.M. [1991]. Some statistical process control methods for autocorrelated data. *J. Qual. Technol.* 23, 179-193.
- NAZEM S. [1988]. *Applied time series analysis for business and economic forecasting*. New York, Dekker, 431 p.
- NIANG N., SAPORTA G. [1995]. Période opérationnelle moyenne de la carte des moyennes mobiles équipondérées pour le contrôle du centrage d'un procédé. *Rev. Stat. Appl.* 43, 5-20.
- PALM R. [1987]. Etude des séries chronologiques par la méthode de BOX et JENKINS. *Notes Stat. Inform.* (Gembloux) 87/2, 41 p.
- PALM R. [1996]. Cartes de contrôle : les cartes de SHEWHART. *Notes Stat. Inform.* (Gembloux) 96/2, 41 p.
- SAS INSTITUTE INC. [1995]. *SAS/QC^R Software: usage and reference, Version 6* (2 vol.). Cary, NC, SAS Institute Inc., 1697 p.
- WIERDA S.J. [1994]. Multivariate statistical process control - recent results and directions for future research. *Stat. Neerl.* 48, 2, 147-168.
- X [1994]. *Minitab reference manual, release 10 for windows*. State College, PA, Minitab, 1047 p.