

CONTRÔLE MULTIVARIÉ DE PROCÉDÉS PAR LOTS À L'AIDE DE STATIS

Ndèye Niang¹, Flavio S. Fogliatto², Gilbert Saporta¹

¹Chaire de Statistique Appliquée & CEDRIC CNAM
292, rue Saint Martin, 75141 Paris Cedex 03, France,

ndeye.niang_keita@cnam.fr

gilbert.saporta@cnam.fr

²Departamento de Engenharia de Produção,
Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil,

ffogliatto@producao.ufrgs.br

RÉSUMÉ

Nous proposons une nouvelle approche du contrôle de qualité des procédés par lots basée sur la méthode STATIS et des cartes de contrôles non paramétriques à partir d'enveloppes convexes. Cette approche générale peut être utilisée pour le contrôle en fin et en cours de fabrication des procédés par lots. La méthode proposée est illustrée sur des données réelles.

MOTS-CLÉS: procédés par lots, contrôle de qualité multivarié, méthode STATIS

SUMMARY

We propose a new non parametric quality control strategy for monitoring batch processes based on the three way method STATIS and convex hull peeling. This general approach allows off line and on line monitoring of batch process. A real example illustrates the proposed method.

KEYWORDS: Batch process, Multivariate quality control, STATIS method.

1-Introduction

Les procédés par lots sont largement utilisés dans le secteur industriel notamment dans l'industrie agroalimentaire ou pharmaceutique. Dans ces procédés, les matières premières sont introduites dans un ordre spécifique et subissent une série de transformations pendant un certain temps qui peut être fixe ou variable donnant alors lieu à des procédés à temps fixe ou à temps variable. Le produit final obtenu est ensuite analysé pour vérifier s'il correspond à des standards de qualité désirés. Le suivi du procédé s'effectue à travers un ensemble de variables caractéristiques du procédé prélevées par un échantillonnage en ligne au fur et à mesure de son déroulement. Les données se présentent sous la forme d'un tableau à trois entrées ou « cube » de données. Due à la nature multidimensionnelle des données issues de tels procédés, les cartes de contrôle multivariées sont alors les seules adéquates pour le contrôle de leur qualité.

La carte multivariée la plus fréquemment utilisée est la carte T^2 de Hotelling, Lowry & Montgomery (1995). En général, ces cartes de contrôle sont basées sur l'hypothèse d'indépendance des observations et de multinormalité des caractéristiques du procédé. Mais dans la pratique ces hypothèses ne sont pas toujours vérifiées. De plus les cartes de contrôle classiques ne permettent pas un contrôle efficace lorsque les standards de qualité sont décrits par des profils ou courbes. Dans le cas de tels procédés, le contrôle s'effectue à travers des cartes multivariées basées sur une analyse en composantes principales particulière (multiway principal component analysis) Nomikos & MacGregor (1995). Ces cartes seront notées MPCA-CCs dans la suite.

L'application des MPCA-CCs pour le monitoring des procédés par lots a été initialement proposé by Jackson & Mudhokar (1979), et beaucoup étudiée par la suite par Nomikos & MacGregor (1995), Kourti & MacGregor (1996) et MacGregor (1997). Elle suppose, en plus de la normalité des variables, que tous les lots aient la même durée et ne peut donc pas être directement utilisée pour le contrôle des procédés à temps variable, ni pour le contrôle de procédés par lots en cours de fabrication. De nombreuses méthodes ont été proposées pour adapter les MPCA-CCs aux cas cités ci-dessus: Nomikos (1995), Kassidas *et al.* (1998), Doan & Srinivasan (2008). Elles peuvent être globalement considérées comme des méthodes de prétraitement dont le but est de

donner la même longueur à tous les lots afin d'appliquer ensuite les MPCA-CCs classiques. Cependant ces méthodes présentent toutes quelques limitations comme cela sera brièvement discuté plus loin.

Dans cette communication, nous proposons une nouvelle approche dans laquelle nous utilisons la méthode STATIS pour réduire la dimension des données sans aucun traitement préalable. Ensuite, nous construisons des cartes de contrôles non paramétriques à partir des enveloppes convexes construites sur les plans factoriels issus de l'application de la méthode STATIS. Cette approche générale peut être utilisée pour le contrôle en fin et en cours de fabrication des procédés par lots à temps fixe et à temps variable sans aucune hypothèse sur la distribution des variables. La méthode proposée est illustrée sur des données réelles d'un procédé à temps fixe.

2- Méthodes usuelles de contrôle de qualité des procédés par lots

Elle supposent que l'on dispose d'un historique de N lots de référence c'est à dire des lots ayant donné un produit de bonne qualité définissant ainsi une distribution de référence représentant le bon fonctionnement du procédé. Dans le cas des procédés à temps fixe, les données se présentent sous la forme de N tableaux à p variables prélevées à T instants. Les méthodes usuelles (multiway principal component analysis) consistent à organiser les données sous la forme d'une matrice ayant les lots en ligne et les variables en colonne pour les instants de temps successifs (figure 1).

	<i>instant 1</i>	<i>instant 2</i>	<i>instant T</i>	
$\mathbf{X} =$	VARIABLES	VARIABLES	VARIABLES	LOT 1
	$1,2,\dots,p$	$1,2,\dots,p$	LOT 2
				LOT N

Figure 1- Matrice des données

On est alors ramené dans le cadre du contrôle multivarié classique. On effectue une analyse en composantes principales pour réduire la dimension des données. Les q premières composantes retenues définissent la distribution de référence. Le contrôle en fin de fabrication d'un lot \mathbf{X}_f consiste à calculer la statistique T^2 de Hotelling à partir de ces composantes standardisées selon la formule suivante :

$$T_{q,f}^2 = \sum_{i=1}^q \frac{c_{i,f}^2}{\sqrt{\lambda_i}} = \sum_{i=1}^q y_{i,f}^2 \quad (1)$$

Le lot sera déclaré hors contrôle si cette valeur est supérieure à la limite de contrôle définie par:

$$LSC = (pT(N+1)(N-1)) / N(N-pT) F_{\alpha;pT,N-pT} \quad (2)$$

où $F_{\alpha;pT,N-pT}$ est le quantile de niveau α d'une loi de Fisher à pT et $N-pT$ degrés de liberté.

Il est aussi habituel d'utiliser des cartes Q des résidus basées sur les dernières composantes principales pour détecter les dérèglages du procédé qui affectent la structure de corrélation des variables Nomikos et MacGregor (1995).

La carte de Hotelling et les cartes Q sont des cartes paramétriques qui reposent sur l'hypothèse de normalité, ce qui constitue une première limitation de ces méthodes.

Par ailleurs, les cartes présentées ci dessus supposent que les lots ont tous la même longueur. Mais des variations dans les conditions initiales ou les matières premières peuvent conduire à des variations dans la longueur des lots, on a alors des procédés par lots à temps variable. D'autre part, lorsqu'on s'intéresse au contrôle en cours de fabrication, on ne dispose que d'une partie des données, celles jusqu'à l'instant $t < T$. Les MPCA-CCS ne permettent donc pas le contrôle des procédés à temps variable ni le contrôle en cours de fabrication et doivent alors être adaptées.

Pour le cas des procédés à temps variable, Nomikos suggère l'utilisation d'une variable

d'indexation autre que le temps pour le monitoring du procédé. Une telle variable doit être mesurable, monotone croissante dans le temps et avoir les mêmes valeurs de début et de fin pour tous les lots, ce qui est assez restrictif mais dans Kourti *et al.* (1996) on trouve des exemples d'application. D'autres auteurs tels que Kassidas *et al.* (1998) proposent des approches différentes dans lesquelles préalablement à l'ACP, les lots sont prétraités en utilisant des techniques de dynamic time warping (DTW) pour donner aux lots la même longueur. Ces méthodes complexes et coûteuses en calculs ont par ailleurs quelques inconvénients dont en particulier la perte de l'information temporelle et la non utilisation de toutes les observations des lots. On peut également citer les travaux de Rosa (2005) basés sur la distance de Hausdorff dans lesquels aucune réduction de dimension n'est nécessaire. Mais là aussi cela conduit à une perte de l'information temporelle et il ne nous paraît évident que la structure de corrélation soit bien prise en compte.

Concernant le contrôle en cours de fabrication, les méthodes proposées consistent à compléter le tableau des données du lot. Nomikos et MacGregor (1995) proposent trois stratégies:

i) la première consiste à supposer que pour les $T-t$ instants suivants l'instant t , les variables du procédé suivent leur trajectoire moyenne. On complète donc les valeurs indisponibles, inconnues par des zéros.

ii) Une deuxième méthode repose sur l'hypothèse que les variables aux $T-t$ instants suivants se comportent comme à l'instant t , donc que les résidus restent constant entre les instants t et T . Une solution intermédiaire entre ces deux stratégies consiste à considérer que les résidus suivants décroissent linéairement ou exponentiellement de leur valeur à l'instant t à zéro à la fin du lot.

iii) La troisième approche considère les $T-t$ observations comme des valeurs manquantes et utilise les composantes principales de l'ACP pour les prédire. Cette dernière méthode apparaît plus performante que les autres si on dispose d'au moins 10% des données du lot. Mais le choix de la stratégie à utiliser dépend des spécificités du procédé à étudier, et il peut parfois être utile de les combiner.

3- Méthode proposée

Nous proposons une méthodologie de contrôle des procédés par lots dans laquelle les données de chaque lot sont représentées sous la forme d'une matrice ayant en ligne les instants de mesure. On est donc en présence de plusieurs tableaux décrivant un ensemble d'individus sur p variables. Il est alors possible de les analyser directement sans aucun traitement préalable en utilisant la méthode STATIS. Nous expliquons plus en détail notre proposition après avoir rappelé brièvement la méthode STATIS.

3.1. STATIS

STATIS est une méthode d'analyse exploratoire simultanée de plusieurs tableaux de données recueillies à différentes occasions Escoufier(2006). A notre connaissance son utilisation en contrôle de qualité se limite aux travaux de Scepi (2002).

L'idée essentielle est la recherche d'une structure commune aux tableaux pour voir si les distances entre les individus sont stables d'un tableau à l'autre. Elle fonctionne en trois étapes. D'abord on effectue une analyse globale dans laquelle on cherche à comparer la structure des tableaux sans pouvoir donner une explication fine des éventuelles différences entre tableaux. Cette étape est appelée *interstructure*. L'étude fine s'effectue dans la deuxième analyse appelée *intrastructure*. Elle repose sur la détermination d'un résumé global des tableaux appelé compromis qui permet de trouver un espace commun de représentation. L'étude de l'évolution de chacun des individus des tableaux sur cet espace de représentation permet d'expliquer au niveau individuel les écarts mis en évidence par l'interstructure.

Plus précisément, on dispose de X_i ($i = 1, \dots, N$) matrices contenant T observations de p variables. Préalablement à l'analyse, les données sont centrées réduites. STATIS associe à chaque X_i la matrice ($T \times T$) des produits scalaires entre les individus $W_i = X_i X_i'$, où X_i' est la matrice

transposée de X_i . C'est un objet représentatif de X_i ; il contient tous les liens inter-individus. Pour comparer deux tableaux X_i et $X_{i'}$, on utilise le coefficient RV de corrélation vectorielle, Escoufier (2006) défini par:

$$RV_{ii'} = \text{trace}(W_i W_{i'}) / \sqrt{\text{trace}(W_i)^2 \text{trace}(W_{i'})^2} \quad (3)$$

RV varie entre 0 and 1; plus il est proche de 1, plus les deux matrices W_i et $W_{i'}$ sont similaires.

L'interstructure consiste à étudier graphiquement les ressemblances globales entre tableaux. Cela passe par la diagonalisation, comme en ACP, de la matrice S contenant les coefficients RV entre W_i et $W_{i'}$ ($i, i' = 1, \dots, N$). Les proximités entre tableaux sont visualisées en projetant les objets W_i sur le premier plan principal: la coordonnée du tableau W_i sur le k ème axe factoriel est donnée par $c_i^k = \sqrt{\lambda_k} u_i^k$ où λ_k est la valeur propre associée au k ème vecteur propre u^k . Cette représentation sera appelée carte-IS dans la suite. La matrice S ayant tous ses termes positifs, le premier vecteur propre u^1 définit un facteur taille.

L'étude de l'intrastructure consiste d'abord à rechercher le compromis qui résume au mieux l'ensemble de tableaux. La solution est une moyenne pondérée des objets W_i les coefficients α_i^1 étant les composantes du premier vecteur propre u^1 normalisé:

$$W = \sum_{i=1}^N \alpha_i^1 W_i \quad (4)$$

Les poids α_i^1 représentent alors le niveau d'accord entre les tableaux et le compromis. Cette définition du compromis confère à STATIS une robustesse aux valeurs aberrantes: plus un tableau est différent des autres, moins il a d'influence sur le compromis. Cette propriété est particulièrement intéressante en contrôle de qualité dont le but est la détection de valeurs anormales.

La matrice compromis W est ensuite diagonalisée permettant ainsi d'obtenir l'espace de représentation commun à l'ensemble des tableaux. Il est alors possible de visualiser sur le premier plan principal des points artificiels B_t ($t=1, \dots, T$) appelés points compromis. Les coordonnées sur le k -ième axe factoriel axis sont les éléments du vecteur suivant:

$$z_k = \sqrt{\delta_k} v^k = (1 / \sqrt{\delta_k}) W v^k \quad (5)$$

où δ_k est la valeur propre associée au k -ième vecteur propre v^k .

De plus, il est possible de représenter les individus de tous les tableaux W_i en les projetant sur le plan compromis par la technique des points supplémentaires. Les différentes positions d'un individu selon les tableaux définissent sa trajectoire qui permet de mettre en évidence des écarts entre tableaux au niveau individuel. Il est donc possible d'avoir une représentation détaillée du comportement commun des lots à un instant donné. Cette représentation sera notée carte- CO_t par la suite.

La représentation du compromis dans l'intrastructure est équivalente à celle qu'on aurait eue en faisant ACP de la matrice \underline{X} obtenue en juxtaposant les tableaux X_i pondérés par les α_i^1 :

$$\underline{X} = \left[\sqrt{\alpha_1^1} X_1 \mid \sqrt{\alpha_2^1} X_2 \mid \dots \mid \sqrt{\alpha_N^1} X_N \right] \quad (6)$$

Dans tout ce qui précède on a supposé que les tableaux avaient la même dimension. Cependant les analyses effectuées ne dépendent pas du nombre de variables. STATIS peut alors être utilisé pour étudier des matrices ayant le même nombre de lignes mais des nombres de colonnes différents.

3.3 Cartes de contrôle non paramétriques

Les cartes de contrôle non paramétriques (IS-CC and CO_t -CCs) que nous proposons sont basées sur des enveloppes convexes directement construites sur les plans principaux de STATIS:

interstructure (carte-IS) et intrastructure (carte-CO_i) plots. Pour établir une région de contrôle de confiance $(1-\alpha)$ on utilise une proposition de Zani *et al.* (1998), comprenant les trois étapes suivantes. On détermine sur le plan (carte-IS) une région intérieure qui contient une proportion π^* des points. Elle est obtenue par lissage par une B-spline des contours de l'enveloppe convexe contenant les points. Ensuite on détermine une estimation du centre de la région en prenant par exemple la moyenne arithmétique des observations dans la région. Finalement, la carte de contrôle est obtenue par dilatation de l'enveloppe lissée en multipliant la distance entre le centre et frontière de la région par un nombre l correspondant à la probabilité α de fausse alarme désirée. Pour $\alpha=0.01$, la valeur correspondante de l est 1.68.

4- Applications

On dispose de N lots de référence associés à N tableaux de dimension $T \times p$ que l'on utilise pour avoir la carte IS-CC. Le but du contrôle en fin de fabrication est de vérifier la conformité des données d'un nouveau lot représenté par la matrice X_{N+1} avec des standards de qualité résumés, représentés par la carte de contrôle issue des lots de référence. Le contrôle du nouveau lot s'effectue alors en projetant la matrice X_{N+1} sur la carte IS-CC. Le lot sera déclaré sous contrôle si la matrice se projette à l'intérieur de la région de contrôle. Dans le cas contraire, le lot sera hors contrôle et les cartes CO_{*t*}-CC peuvent être utiles pour identifier l'instant auquel a eu lieu cette sortie des limites de contrôle.

Le contrôle en cours de fabrication consiste suivre le procédé au fur et mesure de son déroulement pour détecter le plus tôt possible une sortie des limites de contrôle plutôt que d'attendre la fin du lot. Nous disposons donc $t = t' \leq T$, mesures où t' représente l'instant de mesure le plus récent. Ainsi seule une partie des données de la matrice X_{N+1} est disponible. Il est donc impossible de projeter la matrice sur la carte CO_{*t'*}-CC.

Mais puisqu'on souhaite vérifier le comportement du procédé jusqu'à l'instant t' nous proposons d'établir une distribution de référence du comportement du procédé jusqu'à l'instant t' ou de manière équivalente une carte de contrôle notée IS(t')-CC. Cela consiste donc à déterminer une carte IS-CC sur des tableaux partiels obtenus en sélectionnant les t' premières lignes des tableaux X_i . En principe il faudrait autant de cartes que d'instant de mesure mais il n'est pas nécessaire d'établir toutes les cartes, des cartes tous les 5 ou 10 instants de mesure suffisent au contrôle en cours de fabrication.

Nous illustrons notre proposition sur des données réelles utilisées dans la littérature des procédés par lots par Nomikos, Mc Gregor ou Eriksson et al par exemple. Les données sont issues d'un procédé de polymérisation et sont composées de 18 lots de référence sélectionnés comme représentant le comportement normal souhaité du procédé. Pour chaque lot, 10 variables ont été prélevées à 100 instants. Les variables X_1 X_2 X_3 X_6 et X_7 sont des mesures de température, X_4 X_8 et X_9 sont des mesures de pression et les variables X_5 et X_{10} représentent des vitesses d'écoulement de matières ajoutées au réacteur. On dispose de plus d'un ensemble de 11 lots supplémentaires pour tester les performances des méthodes. Il contient 4 lots de bonne qualité et 7 mauvais lots. Nous avons appliqué STATIS et les régions de contrôle avec un niveau de confiance de 99% sont construites sur les plans factoriels de l'interstructure.

La figure 2 montre les résultats du contrôle en fin de fabrication. Tous les 7 mauvais lots ont été signalés hors contrôle avec cependant un plus fort signal pour 6 d'entre eux (fig.2.a). Le mauvais lot proche de la limite a été diagnostiqué comme ayant un comportement différent des 6 autres et n'est en général pas détecté comme étant hors contrôle (Eriksson *et al*). La carte de contrôle (fig.2.b) montre les résultats pour les bons lots. 3 bons lots parmi les 4 sont signalés sous contrôle. On constate cependant une fausse alerte comme dans Eriksson et al.

* π est égale à 50% des points dans Zani *et al.* (1998), mais on peut utiliser une plus grande proportion.

L'application de la méthode séquentielle proposée (les figures ne sont pas présentées ici) permet de détecter les écarts aux modèles de référence à l'instant 20 pour tous les 6 lots fortement hors contrôle.

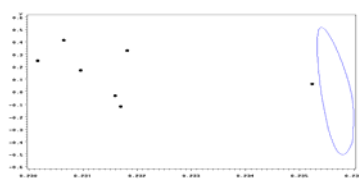


Fig. 2.a. Mauvais lots

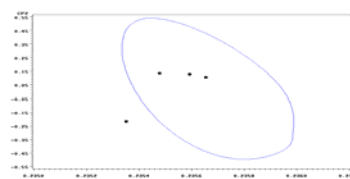


Fig. 2.b. Bons lots

Conclusion

Nous avons proposé une méthode pour le contrôle de qualité des procédés par lots en fin et en cours de fabrication basée sur la méthode STATIS. Le suivi du procédé est effectué à travers des cartes de contrôles non paramétriques utilisant toutes les observations disponibles pour le contrôle en fin de fabrication, et une partie des observations séquentiellement pour le contrôle en cours de fabrication. La méthode est illustrée sur des données réelles.

Des évaluations plus formelles des performances de la méthode sont nécessaires ainsi que des études comparatives avec d'autres méthodes proposées dans la littérature.

Bibliographie

- Doan, X-T., Srinivasan, R. (2008) Online monitoring of multi-phase batch processes using phase-based multivariate statistical control. *Computers and Chemical Engineering*, 32: 230-243.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., and Wold, S.(2001).*Multi- and Megavariate Data Analysis*. Umetrics
- Escoufier, Y. 2006. Operator related to a data matrix: a survey. *Proceedings in Computational Statistics* Rizzi A. et al. (eds), 285-297 Physica-Verlag.
- Jackson, J.E. and Mudholkar, G.S. (1979) Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics*, 21 (3), 341–34.
- Kassidas, A., MacGregor, J.F. and Taylor, P.A. (1998) Synchronization of batch trajectories using dynamic time warping. *AIChE Journal*, 44, 864–875.
- Kourti, T. and MacGregor, J.F. (1996) Multivariate SPC Methods for Process and Product Monitoring. *Journal of Quality Technology*, 28 (4), 409–428.
- Lowry, C.A. and Montgomery, D.C. (1995) A review of multivariate control charts. *IIE Transactions*, 27 (6), 800–810.
- MacGregor, J.F. (1997) Using on-line process data to improve quality: challenges for statisticians. *International Statistical Review*. 65 (3), 309–323.
- Nomikos, P. and MacGregor, J.F. (1995) Multivariate SPC charts for monitoring batch processes. *Technometrics*, 37 (1), p.41–59.
- Rosa, A.F.P. (2005) *Maîtrise statistique de procédés par lots à temps variable*. Thèse de Doctorat de l'Université de Nantes, Nantes (France).
- Scepi, G. (2002) Parametric and non parametric multivariate quality control charts. In *Multivariate Total Quality Control*, Physica-Verlag, Lauro C. et al. (eds), 163–189.
- Zani, S., Riani, M. and Corbellini, A. (1998) Robust bivariate boxplots and multiple outlier detection. *Computational Statistics & Data Analysis*, 28, 257-270.