CARTAS DE CONTROLE MULTIVARIADAS BASEADAS EM COMPONENTES PRINCIPAIS

Ariane Ferreira Porto Rosa, *Mestranda*
Flávio S. Fogliatto, *Ph.D.*
Programa de Pós -Graduação em Engenharia de Produção - PPGEP
UFRGS - Praça Argentina, 9 – Sala LOPP, Porto Alegre - RS - 90040-020
e-mail: aprosa@ppgep.ufrgs.br; ffogliatto@ppgep.ufrgs.br

ABSTRACT:

Statistical control charts have been widely used for monitoring the performance of industrial processes. Several modifications on the classic Shewhart charts have been proposed over the years in the literature; they aim at providing better tools for monitoring increasingly complex processes. This paper presents one such modification: a multivariate control chart based on principal components. The key idea is to provide efficient control over several variables by monitoring only a few independent linear combinations of them, called principal components. The principal components control charts are particularly efficient for on-line monitoring of batch processes, as discussed in the paper.

KEYWORDS: multivariate control charts, principal components, batch processes.

RESUMO:

As cartas de controle estatístico têm sido amplamente utilizadas no monitoramento do desempenho de processos. Com a crescente informatização dos processos industriais, tem-se verificado um aumento sensível na quantidade de informações disponíveis sobre variáveis de processo. Via de regra, essas variáveis apresentam-se fortemente correlacionadas. Em casos especiais, como nos processos em batelada, tais variáveis descrevem um perfil de variação ao longo do tempo, caracterizando o comportamento normal do processo. Nessas condições especiais, as cartas de controle tradicionais não proporcionam um monitoramento eficaz sobre o processo. Este artigo apresenta uma alternativa para o monitoramento on-line de processos: as cartas de controle multivariadas baseadas em componentes principais. A idéia central dessas cartas é monitorar simultaneamente diversas variáveis controlando somente algumas poucas combinações lineares

independentes delas; tais combinações são denominadas componentes principais. O artigo enfatiza a aplicação das cartas de componentes principais no monitoramento de processos em batelada.

PALAVRAS-CHAVE: Cartas de controle multivariadas, componentes principais, Processos em batelada.

1. INTRODUÇÃO

As cartas de controle (CCs) univariadas de Shewhart são comumente utilizadas no controle estatístico de processos industriais [ver Montgomery (1996) e Duncan (1986), entre outros]. Essas cartas utilizam como dados de entrada medições de variáveis que influenciam o desempenho do processo em questão. As medições são realizadas em pontos espaçados no tempo e registradas nas cartas; este registro resulta em gráficos temporais apresentando os valores de medição da variável no eixo vertical e os pontos no tempo nos quais as medições são efetuadas no eixo horizontal. A cada medição, compara-se o resultado obtido com limites de controle. Medições fora dos limites indicam a presença de causas especiais de variabilidade, anômalas ao processo, que prejudicam seu desempenho. Uma vez identificadas as causas especiais, pode-se atuar sobre elas, melhorando continuamente o desempenho do processo. As CCs permitem a redução sistemática da variabilidade nas características críticas de processo (representadas pelas variáveis monitoradas nas cartas).

A estratégia de controle univariado de Shewhart, acrescida de algumas suposições, pode ser estendida para o caso multivariado. CCs multivariadas são indicadas em situações onde variáveis de processo ou produto são monitoradas simultaneamente. Os resultados das CCs multivariadas e univariadas são particularmente diferentes quando do monitoramento simultâneo de variáveis correlacionadas; nestes casos o uso de CCs multivariadas é fortemente recomendado. O tipo de CC multivariada de utilização mais comum é a carta de Hotelling, ou carta de T^2; ver Hotelling (1947). Outros tipos de CCs multivariadas são apresentadas em Jackson (1991).

Algumas situações especiais dificultam ou impossibilitam a utilização das CCs tradicionais. Por exemplo, as CCs multivariadas (assim como as univariadas) pressupõem independência entre

pontos amostrais. Com a crescente automatização de processos, variáveis podem ser medidas em tempo real e em pontos temporais próximos; nessas condições, a suposição de independência dificilmente é verificada. Uma outra situação especial ocorre quando a variação normal de uma variável monitorada ao longo do tempo é descrita por um perfil; ou seja, a média da variável, em condições normais de processo, muda ao longo do tempo. As CCs univariadas não podem ser aplicadas nessa situação, já que pressupõem variáveis com média constante.

Neste artigo, apresenta-se uma ferramenta para o monitoramento, simultâneo e em tempo real, de variáveis determinantes da qualidade de produtos e processos: as cartas de controle multivariadas baseadas em componentes principais (CCPs). Essas cartas são operacionalizadas utilizando a análise de componentes principais e as CCs multivariadas de Hotelling. As CCPs admitem pontos amostrais sucessivos dependentes, além de permitir o monitoramento de variáveis que não apresentam valores fixos de média ao longo do tempo. As CCPs são particularmente eficientes no monitoramento *on line* de processos em batelada, onde um grande número de variáveis de processo esteja envolvido.

A análise de componentes principais tem por objetivo reduzir a dimensionalidade de conjuntos de dados multivariados. Em outras palavras, busca-se representar de maneira reduzida um conjunto de variáveis aleatórias, sem que isso acarrete perda significativa de informações. Tal representação reduzida é obtida utilizando componentes principais. Matematicamente, os componentes principais são combinações lineares das variáveis aleatórias, com a propriedade especial de independência.

A idéia-chave das CCPs consiste em utilizar uma estratégia de controle similar àquela das CCMs utilizando, ao invés das variáveis de processo, componentes principais delas derivados. Desta forma, simplifica-se a tarefa de monitoramento de processos, já que este passa a ser feito sobre um número menor de variáveis (ou seja, sobre os componentes principais). Como os componentes principais são independentes entre si, as CCPs podem ser utilizadas no monitoramento de processos com as características especiais descritas anteriormente.

Este artigo encontra-se organizado em cinco seções, incluindo a presente introdução. A seção 2 traz uma introdução à análise de componentes principais. A seção 3 apresenta uma introdução às cartas de controle multivariadas de Hotelling. As seções 2 e 3 são necessárias para a compreensão das cartas de controle baseadas em componentes principais, apresentadas na seção 4. Uma conclusão encerra o artigo na seção 5.

2. ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais (CPs) é uma técnica estatística utilizada para resumir informações em conjuntos de dados multivariados. Considere um conjunto de dados composto de realizações de p variáveis aleatórias. Tal conjunto de dados possui p CPs, sendo cada CP formado por uma combinação linear distinta das p variáveis aleatórias originais. Essas combinações são determinadas através de manipulações algébricas na matriz de covariâncias das p variáveis, conforme apresentado mais adiante. Cada CP captura uma *direção de variabilidade* do conjunto de dados originais. As direções capturadas por cada CP são ortogonais entre si. Isso equivale a dizer que os CPs são variáveis aleatórias independentes.

O objetivo da determinação de CPs pode ser melhor entendido ao considerar-se que, via de regra, grande parte da variabilidade do conjunto de dados pode ser descrito por menos que a totalidade dos CPs. Assim, pode-se reduzir a dimensionalidade do conjunto de dados (isto é, o número de variáveis aleatórias que o descrevem), sem perda significativa de informações. Além disso, substitui-se um grande número de variáveis aleatórias *dependentes* por um número menor de variáveis aleatórias *independentes*. A Análise de CPs tem sido utilizada com sucesso em diversas áreas do conhecimento; alguns exemplos de aplicação são apresentados em Hair, Jr. *et al.* (1995).

Conforme mencionado acima, CPs são combinações lineares de variáveis aleatórias, identificadas por X_i, $i = 1,\ldots, p$, com realizações designadas por $(x_{i1}, x_{i2}, \ldots, x_{in})$. Essas combinações são obtidas a partir da matriz de covariâncias, associadas às p variáveis aleatórias. A determinação algébrica dos CPs é apresentada na sequência. Na exposição que se segue, matrizes são identificadas por letras maiúsculas em negrito, p.ex. $\mathbf{A}$, e seus transpostos por $\mathbf{A}^t$; vetores são identificados por letras minúsculas em negrito, p.ex. $\mathbf{a}$, e seus transpostos por $\mathbf{a}^t$.

Seja $\mathbf{\Sigma}$ a matriz de covariâncias, de dimensão $(p \times p)$, associada a matriz de variáveis aleatórias $\mathbf{X} = [X_1, X_2, ..., X_p]$. A dimensão de $\mathbf{X}$ é $(n \times p)$, ou seja, dispõe-se de n observações de cada variável aleatória. O vetor $\mathbf{x}^t$ denota uma linha qualquer de $\mathbf{X}$. Os p autovalores de $\mathbf{\Sigma}$ são designados por λ_i, $i = 1,\ldots, p$, e os p autovetores designados por $\mathbf{e}_i$, $i = 1,\ldots, p$, com elementos dados por $(e_{i1},\ldots, e_{ip})$. Os autovalores e autovetores de uma matriz quadrada não-singular qualquer são determinados através de sua equação característica; ver Strang (1988), p. 246. Associado a cada autovalor λ_i existe um autovetor $\mathbf{e}_i$. Assim, os pares $(\lambda_1, \mathbf{e}_1)$, $(\lambda_2, \mathbf{e}_2)$, ..., $(\lambda_p, \mathbf{e}_p)$ correspondem aos autovalores e autovetores de $\mathbf{\Sigma}$, com autovalores arranjados tal que $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p$.

O $i^{\text{ésimo}}$ CP pode ser obtido pela expressão [Seber (1984)]:

$$Y_i = \mathbf{x}^t \mathbf{e}_i = e_{i1} X_1 + e_{i2} X_2 + \mathrm{K} + e_{ip} X_p \,. \qquad (1)$$

Na eq. (1) os elementos do autovetor $\mathbf{e}_i$ funcionam como pesos de importância de $X_1,\ldots, X_p$ na composição do $i^{ésimo}$ CP, e são denominados *cargas* do CP. Em CPs obtidos a partir de variáveis aleatórias padronizadas, a magnitude da carga associada a uma variável descreve sua importância relativa na composição dos CPs. É bastante comum, como apresentado no exemplo mais adiante nesta seção, que alguns CPs sejam majoritariamente descritos por uma única variável aleatória. O valor observado de Y_i para um determinado vetor $(x_1,\ldots, x_p)$ de realizações de $(X_1,\ldots, X_p)$ é denominado *escore* do CP. Cada CP apresenta n escores, correspondendo ao número total de realizações disponíveis de cada variável aleatória.

Cada CP descreve uma das direções de variabilidade do conjunto de dados. Em outras palavras, cada CP descreve uma porção da variância total apresentada pelo conjunto das p variáveis aleatórias. Sabe-se que a diagonal principal de $\mathbf{\Sigma}$ apresenta os valores de variância associados a cada variável aleatória. Logo, a variância total no conjunto de dados é dada pela soma dos elementos da diagonal principal de $\mathbf{\Sigma}$; este valor é igual à soma dos autovalores de $\mathbf{\Sigma}$ [já que o traço de uma matriz quadrada é igual à soma de seus autovalores; ver Rencher (1995), p. 39]. Assim, pode-se utilizar os autovalores de $\mathbf{\Sigma}$ para determinar quanto da variabilidade total presente no conjunto de dados está sendo descrita por cada CP. Mais especificamente, a proporção da variância total descrita pelo $i^{ésimo}$ CP é dada por:

$$\text{Proporção da variância descrita pelo } i^{ésimo} \text{ CP} = \frac{\lambda_i}{\lambda_1 + \mathrm{K} + \lambda_p} \,. \qquad (2)$$

Através da eq. (2), pode-se definir o número de CPs a serem utilizados na representação do conjunto de dados originais. Essa definição é essencialmente qualitativa. Por exemplo, um analista pode considerar satisfatória uma representação dada por um grupo de CPs que descreva 90% da variabilidade original. Algumas regras auxiliares na definição do número de CPs que melhor representam o conjunto de dados em estudo são apresentadas em Rencher (1995), p. 434.

Geometricamente, os CPs podem ser representados por um sistema de eixos ortogonais com origem posicionada na média do conjunto de dados. O primeiro CP segue a direção de maior variabilidade dos dados. O segundo CP segue a segunda direção de maior variabilidade, ortogonal ao primeiro

CP. Os demais CPs são posicionados no espaço *p*-dimensional seguindo a mesma lógica. à primeira e assim por diante.

Por exemplo, suponha duas variáveis aleatórias X_1, X_2, seguindo uma distribuição Normal bivariada $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com vetor de médias dado por $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. A função de densidade de $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ é dada pela equação da elipse centrada em $\boldsymbol{\mu}$:

$$(\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c^2, \qquad (3)$$

onde $\mathbf{x}$ denota o vetor que contém as variáveis aleatórias X_1 e X_2. Os eixos da elipse em (3) são dados por $\pm c(\lambda_i)^{1/2}\mathbf{e}_i$, com i = 1,2. Os CPs $Y_1 = \mathbf{x}^t\mathbf{e}_1$ e $Y_2 = \mathbf{x}^t\mathbf{e}_2$ estão posicionados nas direções dos eixos da elipse de densidade constante, conforme ilustrado na Figura 1.

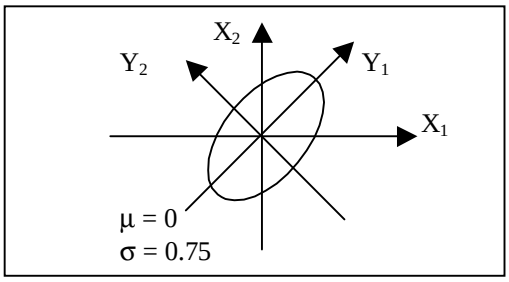

Figura 1. Elipse de densidade constante e CPs Y_1 e Y_2.

Os conceitos apresentados acima são agora ilustrados através de um exemplo [Johnson & Wichern (1992)]. Considere três variáveis aleatórias X_1, X_2, X_3, com matriz de covariâncias $\boldsymbol{\Sigma}$ dada por:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Os autovalores e autovetores de $\boldsymbol{\Sigma}$ são:

$\lambda_1 = 5.83$ $\quad \mathbf{e}_1' = [0.383, -.924, 0]$,

$\lambda_2 = 2.00$ $\quad \mathbf{e}_2' = [0, 0, 1]$,

$\lambda_3 = 0.17$ $\quad \mathbf{e}_3' = [0.924, 0.383, 0]$.

Os CPs são obtidos diretamente dos autovetores, sendo dados abaixo:

$$Y_1 = \mathbf{x}\mathbf{e}_1^t = 0{,}383X_1 - 0{,}924X_2$$

$$Y_2 = \mathbf{x}\mathbf{e}_2^t = 1{,}000X_3$$

$$Y_3 = \mathbf{x}\mathbf{e}_3^t = 0{,}924X_1 + 0{,}383X_2$$

Observe que o 2º CP captura uma direção de variância representada exclusivamente pela variável aleatória X_3.

A proporção da variância total explicada pelos CPs, obtida utilizando a eq. (2) é:

CP_1 = 72,87% CP_2 = 25,00% CP_3 = 2,12%.

É fácil verificar que aproximadamente 98% da variabilidade total presente nos dados está representada pelos dois primeiros CPs. Assim, pode-se utilizar Y_1 e Y_2 no lugar das variáveis originais, sem grandes perdas de informação. Utilizando os CPs, reduz-se a dimensionalidade do problema de três variáveis (variáveis aleatórias originais) para duas variáveis (CPs).

3. CARTAS DE CONTROLE MULTIVARIADAS PARA MÉDIAS

As Cartas de Controle Multivariadas (CCMs) são utilizadas no monitoramento conjunto de duas ou mais variáveis de processo ou produto. Nessas cartas, pontos *p*-dimensionais (ou seja, relativos a *p* variáveis aleatórias ou estatísticas de interesse delas derivadas) são representados uni-dimensionalmente e plotados em gráficos similares às cartas de Shewhart. Desta forma, simplifica-se a tarefa de controle simultâneo de variáveis.

As CCMs são particularmente recomendadas em situações onde exista correlação significativa entre as variáveis a serem monitoradas. A significância das correlações pode ser determinada utilizando o teste de Fisher e as correlações amostrais; ver Freund & Simon (1997), p. 538. No caso de correlações significativas, o monitoramento das *p* variáveis de interesse utilizando *p* cartas univariadas resulta em grande número de alarmes falsos, o que é indesejável em qualquer esquema de monitoramento. Na verdade, mesmo variáveis independentes monitoradas individualmente apresentam alarmes falsos, o que torna seu monitoramento simultâneo, usando CCMs, recomendado [Jackson (1991)].

As CCMs foram originalmente introduzidas por Hotelling (1947) para o monitoramento de médias amostrais; estas cartas são descritas a seguir.

Suponha variáveis de interesse seguindo uma distribuição Normal *p*-variada, com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. Tomam-se amostras de tamanho *n* para cada uma das *p* variáveis de interesse (a serem monitoradas). Calcula-se a média amostral de cada variável e escreve-se o resultado num vetor de médias $\bar{\mathbf{x}}^t = \left[\bar{X}_1, \mathrm{K}\ , \bar{X}_p\right]$. A estatística a ser monitorada na CCM é:

$$(i) \qquad \chi_0^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \quad \text{ou}$$

$$(ii) \qquad T^2 = n(\bar{\mathbf{x}} - \bar{\bar{\mathbf{x}}})^t \mathbf{S}^{-1} (\bar{\mathbf{x}} - \bar{\bar{\mathbf{x}}}). \tag{4}$$

A estatística (i) será utilizada sempre que os parâmetros populacionais $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ forem conhecidos. A estatística (ii) será utilizada quando os parâmetros populacionais $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ não forem conhecidos, sendo estimados por $\bar{\bar{\mathbf{x}}}$ e $\mathbf{S}$. O estimador $\bar{\bar{\mathbf{x}}}$ é dado por [Montgomery (1996)]:

$$\bar{\bar{\mathbf{x}}}^t = \left[\bar{\bar{X}}_1, \ldots, \bar{\bar{X}}_p\right] \quad \text{e} \quad \bar{\bar{X}}_i = \frac{1}{mn} \sum_{k=1}^{m} \sum_{j=1}^{n} x_{jki}\,, \quad i = 1, \ldots, p, \tag{5}$$

onde m denota o número total de amostras de tamanho n utilizadas na estimação. O estimador da matriz $\mathbf{S} = \left[\bar{S}_{qh}\right]$ é:

$$\bar{S}_{qh} = \frac{1}{m} \frac{1}{(n-1)} \sum_{k=1}^{m} \sum_{j=1}^{n} \left(x_{jqk} - \bar{x}_{qk}\right)\left(x_{jhk} - \bar{x}_{hk}\right) \tag{6}$$

Via de regra, os parâmetros populacionais não são conhecidos, sendo estimados utilizando as eqs. (5) e (6) e dados coletados do processo em estudo. O limite de controle superior (LCS) da CCM para esta situação mais usual é dado por:

$$LCS = \frac{p(m+1)(n-1)}{mn - m - p + 1} F_{\alpha, p, mn-m-p+1}\,, \tag{7}$$

onde (1-α) denota o nível de significância do teste de hipóteses representado pela CCM. O limite de controle inferior é zero, por definição. As estimativas dos parâmetros populacionais caracterizarão a distribuição de referência, contra a qual observações futuras do processo serão comparadas. Desta forma, os dados utilizados na determinação das estimativas devem ser coletados quando o processo apresentar operação estável e desejável.

Para ilustrar a operacionalização das CCMs, considere duas variáveis de interesse, medidas de um mesmo processo [Montgomery (1996)]. Coletam-se amostras de tamanho n = 10. Os limites de controle foram calculados a partir de 20 amostras obtidas do processo em condições desejáveis de operação. A estatística amostral utilizada para monitorar o desempenho das variáveis é:

$$T^2 = \left(\frac{n}{\bar{S}_1^2 \bar{S}_2^2 - \bar{S}_{12}^2}\right)\left[\bar{S}_2\left(\bar{x}_1 - \bar{X}_1\right) + \bar{S}_1\left(\bar{x}_2 - \bar{X}_2\right) - 2\bar{S}_{12}\left(\left(\bar{x}_1 - \bar{X}_1\right)\left(\bar{x}_2 - \bar{X}_2\right)\right)\right], \tag{8}$$

ou seja, a expressão em (4) adpatada para o caso especial onde p = 2. Os dados utilizados, as estimativas dos parâmetros populacionais $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ e o valor de T^2 para cada amostra estão apresentadas na Tabela 1. A carta de controle resultante, com LCS calculado pela eq. (7), está presentada na Figura 2 (α = 0.001). Como esperado, os pontos na carta caracterizam um processo estável.

Número da Amostra	*Médias Amostrais*		*Variâncias e Covariâncias*			
k	$\bar{x}_{1k}$	$\bar{x}_{2k}$	S_{1k}^2	S_{2k}^2	S_{12k}	T^2
1	115.25	1.04	1.25	0.87	0.80	2.16
2	115.91	1.06	1.26	.085	0.81	2.14
3	115.05	1.09	1.30	0.90	0.82	6.77
4	116.21	1.05	1.02	0.85	0.81	8.29
5	115.90	1.07	1.16	0.73	0.80	1.89
6	115.55	1.06	1.01	0.80	0.76	0.03
7	114.98	1.05	1.25	0.78	0.75	7.54
8	115.25	1.10	1.40	0.83	0.80	3.01
8	116.15	1.09	1.19	0.87	0.83	5.92
10	115.92	1.05	1.17	0.86	0.95	2.41
11	115.75	0.99	1.45	0.79	0.78	1.13
12	114.90	1.06	1.24	0.82	0.81	9.96
13	116.01	1.05	1.26	0.55	0.72	3.86
14	115.83	1.07	1.17	0.76	0.75	1.11
15	115.29	1.11	1.23	0.89	0.82	2.56
16	115.63	1.04	1.24	0.91	0.83	0.70
17	115.47	1.03	1.20	0.95	0.70	0.19
18	115.58	1.05	1.18	0.83	0.79	0.00
19	115.72	1.06	1.31	0.89	0.76	0.35
20	115.40	1.04	1.29	0.85	0.68	0.62

Tabela 1. Dados do exemplo de CCM.

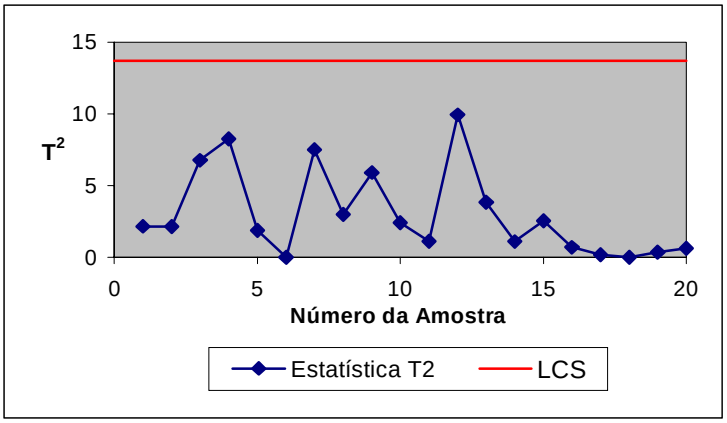

Figura 2. Exemplo de CCM de Hotelling.

4. CARTAS DE CONTROLE MULTIVARIADAS BASEADAS EM COMPONENTES PRINCIPAIS

As CCM baseadas em componentes principais (doravante designadas por CCP) utilizam a estratégia de redução de dimensionalidade apresentada na seção 3, em conjunto com a CCM de Hotelling, apresentada na seção 4. As CCPs foram inicialmente desenvolvidas para monitoramento de processos contínuos, onde diversas variáveis de processo/produto são monitoradas *on line*; ver Kresta *et al.* (1991), Skagerberg *et al.* (1992), Miller *et al.* (1993) e McGregor *et al*. (1994). Mais recentemente, a aplicação das CCPs no monitoramento de processos em batelada foi sugerida por Nomikos & McGregor (1994, 1995). O desenvolvimento aqui apresentado aplica-se a processos em batelada, estando baseado nos trabalhos de McGregor (1995), Nomikos & McGregor, (1994, 1995) e Kourti & McGregor (1996).

As CCPs para processos em bateladas são ilustradas através do seguinte exemplo. Deseja-se monitorar um processo de mistura de borracha para manufatura de bandas de rodagem. Vários componentes são misturados mecânicamente em um equipamento Bambury. As variáveis de interesse, monitoradas *on line*, são a temperatura da mistura (medida em três pontos distintos do equipamento) e a energia integrada à massa. As realizações dessas variáveis de interesse são registradas a cada 50 segundos no decorrer da batelada (que dura 240 segundos) e armazenadas em um equipamento supervisório. Diversas variáveis de produto, com especificações definidas, são monitoradas em laboratório. O resultado da análise laboratorial das variáveis caracteriza a qualidade da mistura, identificando se uma batelada transcorreu em condições normais ou sob efeito de causas especiais.

Dispõe-se de informações sobre 50 bateladas, transcorridas em condições normais. Essas informações encontram-se organizadas na Tabela 2. As realizações de cada variável de processo são representadas pela matriz **X**. Os 48 pontos de medição ao longo do tempo, designados por *período*, correspondem às medições das variáveis de processo, do tempo 1 ao tempo 240 segundos, tomadas em intervalos de 5 segundos.

						Período						
	1				…				48			
Batelada/Variáveis	T^1	T^2	T^3	EI					T^1	T^2	T^3	EI
1	**X**											
N												
50												

Tabela 2. Informações sobre as variáveis de processo obtidas em 50 bateladas.

Observe que as mesmas variáveis são medidas repetidamente em todos os períodos. Uma estratégia de controle seria formar amostras a cada 10 segundos, com medições das 4 variáveis, e plotar os resultados em uma CCM de Hotelling. Todavia, dois problemas surgem dessa estratégia: (*i*) a proximidade temporal das medições compromete a suposição de independência entre pontos amostrais na CCM, e (*ii*) as variáveis não apresentam valores fixos de média ao longo do tempo total da batelada, conforme ilustrado para T^1 na 0. Observe também que estratégias univariadas de monitoramento não podem ser consideradas, tendo em vista o problema (*i*). No contexto acima, as CCPs parecem ser a melhor estratégia de monitoramento, conforme apresentado a seguir.

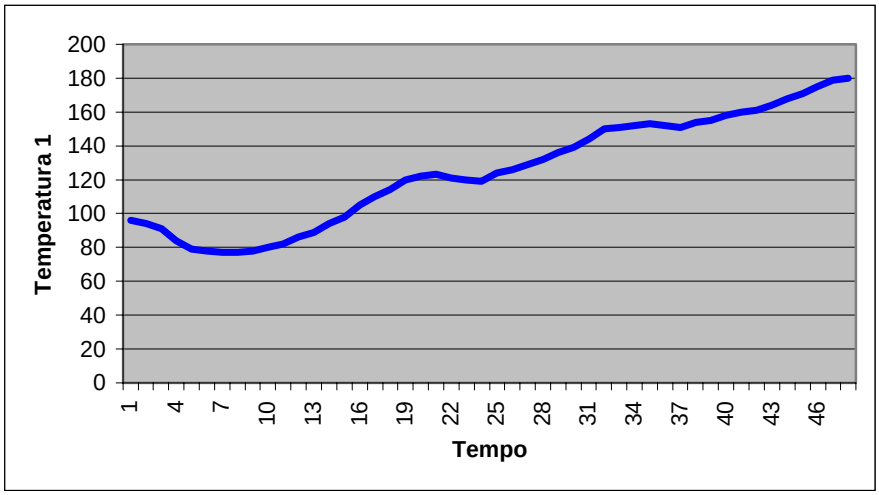

Figura 3 Variação de T^1 ao longo da batelada.

Na estratégia de monitoramento utilizando CCPs, 192 variáveis aleatórias são consideradas. Essas variáveis correspondem às medições de T^1, T^2, T^3 e EI em cada período. A CCP é inicialmente implementada reduzindo o número total de variáveis a serem monitoradas, através da determinação de seus CPs. Os CPs são obtidos das colunas de **X** (a matriz de dados na Tabela 2), correspondentes às 192 variáveis de processo. Suponha, para fins de ilustração, que as 192 variáveis sejam bem representadas por 5 CPs. Suponha também que os CPs foram obtidos de

variáveis previamente padronizadas (eliminando, assim, problemas decorrentes da diferença de escala entre variáveis).

Os CPs foram obtidos de uma matriz **X** que caracteriza a operação normal (ou desejável) do processo. Assim, parece natural utilizar a distribuição de probabilidade dos CPs como referência, contra a qual futuras observações dos CPs serão comparadas, visando determinar o *status* do processo.

Considere uma batelada recém iniciada. Deseja-se detectar a ocorrência de causas especiais no decorrer da batelada e atuar corretivamente sobre elas, evitando a manufatura de produtos não-conformes. O monitoramento da batelada utilizará como estatística amostral um vetor contendo os escores obtidos para os 5 CPs. Pontos amostrais serão comparados contra uma distribuição de referência (obtida a partir das observações das 50 bateladas na Tabela 2), utilizando uma CCM de Hotelling, com LSC obtido pela expressão em (7). Todavia, como durante o monitoramento a batelada ainda está em curso, só será possível obter os escores de cada CP após sua conclusão. Isso ocorre porque os CPs são combinações lineares de *todas* as 192 variáveis de processos; assim, num período $k < 48$, as observações das variáveis de processo durante os períodos (k, 48] ainda não se encontram disponíveis.

Para contornar o problema acima, Nomikos & McGregor (1995) sugerem antecipar os valores ainda não observados das variáveis, supondo processo operando em condições normais. Dessa forma, calculam-se os escores dos CPs utilizando as observações das variáveis disponíveis até o período atual e supõe-se que o valor das observações futuras (ainda não disponíveis) de todas as variáveis será zero. Isso equivale a supor que o processo transcorrerá normalmente no futuro, já que as variáveis foram previamente padronizadas (i.e., apresentam média zero) antes da determinação dos CPs.

Suponha que, num dado período k, um ponto amostral é plotado além dos limites de controle na CCM de Hotelling. Neste caso, um ou mais dos CPs apresentou desvio anormal de seus valores de média, devendo ser identificados. O objetivo do monitoramento é determinar qual variável de processo desviou-se de sua trajetória esperada e atuar corretivamente sobre ela. Uma vez identificado o CP responsável pelo sinal na CCM de Hotelling, é necessário identificar qual (ou quais) das variáveis de processo contribuiu para o escore anormal apresentado pelo CP.

A abordagem mais usual para determinação da contribuição de variáveis de processo para o comportamento anormal de escores em CPs utiliza gráficos de contribuição [MacGregor *et al.*

(1994), Kourti & McGregor (1996)]. Os gráficos de contribuição são gráficos de barra onde são plotados os valores observados das variáveis de processo no período onde a CCM de Hotelling sinalizou uma causa especial. Como as variáveis foram padronizadas, os gráficos apresentam uma linha central em zero. Variáveis com valores positivos ou negativos muito diferentes de zero (por exemplo, valores maiores que $|2|$) são candidatos à investigação detalhada.

Algumas vantagens da estratégia de monitoramento descrita acima são apresentadas na sequência:

(i) As CCPs permitem o controle simultâneo de um grande número de variáveis de processo, oferecendo-se como ferramenta eficiente no monitoramento de processos automatizados.

(ii) As CCPs permitem o controle de variáveis que descrevem um perfil de variação ao longo do tempo, conforme ilustrado na 0; essas variáveis são de ocorrência bastante comum na industria.

(iii) A utilização das CCPs não pressupõe variáveis Normalmente distribuídas; na prática, as variáveis podem apresentar qualquer distribuição de probabilidade.

(iv) As CCPs podem ser aplicadas quando sucessivos pontos amostrais das variáveis de processo apresentam dependência.

5. CONCLUSÃO

As cartas de controle estatístico são ferramentas de monitoramento do desempenho de processos industriais. Cartas de controle são muito utilizadas na prática, tendo em vista a sua simplicidade e eficácia. Neste artigo, apresentam-se as cartas de controle multivariadas baseadas em componentes principais (CCPs). As CCPs são uma ferramenta de monitoramento, simultâneo e em tempo real, de variáveis determinantes da qualidade de produtos e processos; sua aplicação no monitoramento de processos em batelada é enfatizada neste trabalho.

A aplicação das CCPs apresentam diversas vantagens. Elas permitem, por exemplo, o controle simultâneo de um grande número de variáveis de processo, oferecendo-se como ferramenta eficiente no monitoramento de processos automatizados. As CCPs permitem, também, o controle de variáveis que descrevem um perfil de variação ao longo do tempo, o que é muito frequente na indústria. Finalmente, sua utilização não pressupõe variáveis Normalmente distribuídas. Tal suposição é violada com bastante frequência em aplicações práticas das cartas de controle multivariadas tradicionais.

BIBLIOGRAFIA

DUNCAN, A.J., (1986). *Quality Control And Industrial Statistics.* 5ª Ed., Richard D. Irwin,Inc.

FREUND, J.E. & SIMON, G. A.(1997). *Modern Elementary Statistics.* 9ª Ed., Englewood Cliffs, NJ: Prentice Hall.

HAIR, JR., J.F., ANDERSON, R.E., TATHAM, R.L. & BLACK, W.C. (1995). *Multivariate Data Analysis with Readings.* 4ª Ed., Englewood Cliffs, NJ: Prentice Hall.

HOTELLING, H. (1947). Multivariate Quality Control. Em *Techniques of Statistical Analysis*, editado por Eisenhart, Hastay & Wallis. New York: McGraw-Hill.

JACKSON, J. E. (1991). Multivariate Quality Control: 40 Years Later. Em *Statistical Process Control in Manufacturing*, editado por J.B.Keats & D.C. Montgomery. New York: Marcel Dekker, p. 123-138.

JOHNSON, R.A. & WICHERN, D.W. (1992). *Applied Multivariate Statistical Analysis.* 3ª Ed, Englewood Cliffs, NJ: Prentice Hall.

KOURTI, T. & McGREGOR, J.F. (1996). Multivariate SPC Methods for Process and Product Monitoring, **28** (4), p. 409-428.

KRESTA, J., MACGREGOR, J.F. & MARLIN, T.E. (1991). *Multivariate Statistical Monitoring of Process Operating Performance, Canadian Journal of Chemical Engineering*, 69, 35-47.

MACGREGOR, J.F. (1995). *Using on-line Process Data to Improve Quality*, 39ª Annual Fall Conference.

MACGREGOR, J.F., JAECKLE, C., KIPARISSIDES, C. & KOUTOUDI, M. (1994). Process Monitoring and Diagnosis by Multi-Block PLS Methods. *Journal of the American Institute of Chemical Engineers*, **40**, p. 826-838.

MILLER, P., SWANSON, R.E., and HECKLER, C.E. (1993). *Contribution Plots: The Missing Link in Multivariate Quality Control, unpublished manuscript to be submitted to Journal of Quality Technology*.

MONTGOMERY, D.C. (1996). *Introduction to Statistical Quality Control*, 3ª ed. New York: John Wiley.

NOMIKOS, P. and MACGREGOR, J.F. (1994). *Monitoring of Batch Using Multiway Principal Componentes Analysis . AIChE Journal*, 40, 1361-1375.

RENCHER, A.C. (1995). Methods of Multivariate Analysis. New York: John Wiley.

SEBER, G.A.F. (1986). *Multivariate Observations*. New York: John Wiley.

SKAGERBER, B., MACGREGOR, J.F. and KIPARISSIDES, C. (1992). *Multivariate Data Analysis Applied to Low Density Polyethylene Reactors. Journal of Chemometrics*, 5, 345-360.

STRANG, G. (1988). *Linear Algebra and Its Applications.* 3ª Ed., Forth Worth: Saunders College Publishing.