

## MONITORAMENTO DA DISPERSÃO DE PROCESSOS MULTIVARIADOS POR VARIÁVEIS NÃO OBSERVÁVEIS

Sergio Ferreira Bastos<sup>a\*</sup>, Eugenio Kahn Epprecht<sup>a</sup>

*<sup>a</sup>Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, Brasil*

### Resumo

Em processos multivariados, existem diversas variáveis observáveis para serem controladas. Muitas vezes a variabilidade do processo, tanto a natural como a devida a causas especiais de variação, advém de fontes de variação independentes, que podem ser representadas por variáveis aleatórias não observáveis (ou “latentes”), independentes entre si. As realizações das variáveis observáveis são nesse caso em grande parte função das realizações das variáveis latentes, mais ruídos aleatórios. Uma alteração na média de uma das variáveis latentes não observáveis ou um aumento na sua dispersão resultam em, respectivamente, um deslocamento da média ou um aumento da variabilidade do vetor de variáveis observáveis ao longo de uma direção específica (ou “atribuível”). Um artigo recente propôs o controle da média de processos dessa natureza por gráficos de controle de estimativas dos valores das variáveis latentes, obtidas por projeções do vetor de variáveis observadas nas direções atribuíveis. No presente artigo propõe-se controlar a dispersão de tais processos por gráficos de controle do desvio-padrão (gráficos de S) dos valores das projeções do vetor de variáveis observadas nas direções atribuíveis. O desempenho dos gráficos propostos foi comparado, por simulação, com o de outros esquemas de monitoramento da matriz de covariâncias de processos multivariados; a saber, os gráficos da variância generalizada  $|\bar{S}|$  e da máxima variância amostral VMAX. O esquema proposto tem a vantagem da simplicidade e mostrou-se o mais eficiente dos esquemas analisados em sinalizar aumentos na dispersão de tal tipo de processos, com a vantagem adicional de indicar diretamente a fonte de variação sobre a qual está atuando a causa especial.

Palavras-chave: Variáveis latentes; Projeções; Controle estatístico de processos; Gráficos de controle; Dispersão; Monitoramento; Variáveis não observáveis; Processos multivariados; Direções atribuíveis

### Abstract

In multivariate processes there are several variables to be monitored. Many times both the natural process variability and the variability due to special causes as well arise from independent sources of variation which can be represented by non-observable (or “latent”) random variables, independent between themselves. The realizations of the observable variables are in this case largely due to the realizations of the latent variables (of

\* **Autor para correspondência:** e-mail: sbastos1@gmail.com

## PESQUISA OPERACIONAL PARA O DESENVOLVIMENTO

which they are a function), plus random noise. A shift in the mean of one of the latent variables or an increase in its spread result, respectively, in a shift in the mean or an increase in the variability of the vector of observable variables along a specific (or “assignable”) direction. A recent paper proposed controlling the mean of this kind of processes with control charts on the estimates of the values of the latent variables, obtained by projections of the vector of observed variables on the assignable directions. In the present article we propose controlling the spread of such processes with standard deviation control charts (S charts) of the values of the projections of the vector of observed variables on the assignable directions. The performance of the proposed charts was compared, via simulation, with other schemes for monitoring the variance-covariance matrix of multivariate processes; namely, the generalized variance  $|\bar{S}|$  chart and the VMAX chart, on the maximum sample variance. The proposed scheme has the advantage of simplicity and shown to be the most efficient of the three control schemes analyzed at signalling increases in the spread of this kind of process, with the additional advantage of directly indicating the variation source over which the special cause is acting.

**Keywords:** Latent variables; Projections; Statistical process control; Control charts; Spread; Monitoring; Non-observable variables; Multivariate processes; Assignable directions

## 1. Introdução

A variabilidade de um processo industrial está ligada à existência de dois tipos de causas de variação: as chamadas aleatórias, internas ou naturais e as chamadas especiais, externas ou atribuíveis. Devido à natureza das causas aleatórias, que são inerentes ao processo, não há como prevê-las e muito menos exercer controle sobre elas (aliás, uma das definições de causa aleatória é uma causa sobre a qual não se consegue agir); por outro lado, é possível identificar e controlar a variabilidade do processo devido às causas especiais.

Um processo está em controle estatístico quando está operando livre de causas especiais, ou seja, quando apenas causas aleatórias estão atuando no processo, e está fora de controle estatístico quando na presença das causas especiais. Assim, para que ocorra uma melhoria no processo deve-se promover uma redução na variabilidade provocada pelas causas especiais.

A causa especial (ou causas especiais) pode(m) não estar agindo diretamente em uma ou mais variáveis observáveis, mas sim em alguma variável latente, não diretamente observável, e seu efeito sobre essa variável não observável se refletir sobre as variáveis observáveis em conjunto, alterando a média e/ou variâncias dos elementos do vetor de variáveis observáveis ao longo de direções específicas, denominadas aqui de "direções atribuíveis", já que cada um dos elementos do vetor está associado a uma causa atribuível específica. Reconhecendo que este pode ser o caso, Barton e González-Barreto (1996) propuseram uma forma alternativa de representar um processo multivariado. Partindo do princípio de que é possível identificar padrões de erro relacionados com as prováveis causas especiais de variação, propuseram representar o vetor de variáveis observáveis sobre uma nova base vetorial, cujos elementos são definidos por esses padrões de erro. Assim, propuseram controlar a média dos componentes do vetor de variáveis observáveis nesta base por gráficos de controle separados. No entanto, não realizaram análises de desempenho abrangentes, e não focaram no caso (embora tenham mencionado a sua possibilidade) de haver menos variáveis latentes que observáveis, reduzindo a dimensão do problema. Numa linha relacionada com essa, Runger (1996) propôs o gráfico  $U^2$  (onde  $U$  é o vetor das  $k$  primeiras componentes principais) para detectar mudanças na média das variáveis observáveis ao longo de um vetor arbitrário de dimensão  $k$  no subespaço de  $U$ . Assim, o autor definiu  $U^2$  como o quadrado do comprimento da projeção ortogonal das variáveis observáveis no subespaço definido pelas componentes principais, ambas normalizadas pela matriz de

## PESQUISA OPERACIONAL PARA O DESENVOLVIMENTO

covariâncias. O gráfico  $U^2$  é, assim, um análogo do gráfico de  $T^2$ , porém num subespaço de dimensão reduzida.

Com motivação nesse tipo de processos, em que as causas especiais atuam em variáveis não observáveis (latentes), resultando em variações do vetor de variáveis observáveis em algumas direções específicas (que aqui chamaremos de “direções atribuíveis”, pois cada uma está associada univocamente a uma variável latente), Maranhão e Epprecht (2011) propuseram controlar a média do processo por meio de um gráfico de  $\bar{x}$  em cada projeção do vetor  $\mathbf{x}$  em uma direção atribuível (tantos gráficos quantas as variáveis latentes). Sua proposta difere da de Runger (1996), que, ao agregar a informação de todas as projeções em uma única estatística, tem a vantagem de usar um único gráfico, mas perde a capacidade de discriminar a variável latente afetada pela causa especial. O trabalho de Maranhão e Epprecht (2011) está mais na linha da proposta de Barton e González-Barreto (1996); eles porém detalham todos os aspectos da operação do esquema proposto e realizam uma análise de desempenho do mesmo. Pressupuseram ainda que as direções naturais de variação do processo em controle são as mesmas direções atribuíveis (com vetor médio nulo), já que as fontes naturais de variação do processo atuam nas variáveis latentes, de modo que, quando não houver conhecimento suficiente do processo para determinar essas direções, uma das maneiras de identificá-las é aplicar uma análise de componentes principais (ACP) a dados históricos do processo. Por isso, também analisaram a eficiência da ACP em identificar essas direções.

Os esquemas de monitoramento mencionados destinam-se a monitorar a média de processos multivariados. No entanto, a qualidade dos processos é tão (ou mais) afetada por aumentos na dispersão dos mesmos quanto por alterações na sua média. Assim, é importante controlar também a sua dispersão. No caso univariado, é usual controlar o processo por um par de gráficos: um para monitorar a sua média e outro para monitorar a sua dispersão. Com essa motivação, e dado que Maranhão e Epprecht (2011) se concentraram num esquema destinado a sinalizar alterações na média das variáveis latentes de processos multivariados, aqui propomos um esquema complementar, para monitorar aumentos na dispersão de tais variáveis. Especificamente, baseados no pressuposto de que aumentos na dispersão do processo se dão por aumentos na dispersão das variáveis latentes, propomos monitorar a *variância* desses processos por gráficos do desvio-padrão (gráficos de  $S$ ) dos valores das projeções do vetor observável nas  $k$  direções atribuíveis. O procedimento consiste em: retirar periodicamente uma amostra de tamanho  $n > 1$  do vetor de variáveis observáveis; projetar

## PESQUISA OPERACIONAL PARA O DESENVOLVIMENTO

esses  $n$  vetores em cada uma das  $k$  direções atribuíveis, obtendo assim, para cada direção atribuível,  $n$  valores (de comprimento e sinal) da projeção; calcular, para cada direção atribuível, o desvio-padrão amostral ( $S$ ) desses  $n$  valores, e lançá-lo num gráfico de  $S$ ; ter-se-á, assim,  $k$  gráficos de  $S$  simultâneos: um para cada direção atribuível.

Desde que se tenha o conhecimento das direções atribuíveis, o esquema proposto é uma alternativa prática e eficaz em sinalizar possíveis desvios no processo com a vantagem não desprezível de já indicar diretamente a variável latente cuja dispersão aumentou.

O nosso trabalho fundamenta-se nos seguintes pressupostos:

- i. As direções atribuíveis de variação do vetor de variáveis observáveis estão relacionadas às fontes primárias de variabilidade do processo. A variação devida a cada uma dessas fontes primárias pode ser representada por uma variável latente. Assim, quando o processo está em controle, as observações do vetor de variáveis observáveis são a resultante das realizações das variáveis latentes, mais ruídos aleatórios e erros de medição nas próprias variáveis observáveis.
- ii. As ditas direções atribuíveis de variação do vetor de variáveis observáveis são ortogonais entre si.
- iii. Uma causa especial que atue em uma variável latente, provocando uma mudança no seu valor médio e/ou um aumento da sua variância, resulta numa alteração da média do vetor de variáveis observáveis, ou no aumento de sua variância, ao longo da direção atribuível de variação correspondente àquela variável latente. No caso de diversas causas especiais atuando sobre as variáveis latentes, seus efeitos sobre as variáveis observáveis se combinam.

Uma das vantagens do método proposto é o benefício de redução da dimensionalidade do problema. Em alguns processos, o número de variáveis observáveis pode ser muito grande; o número de variáveis latentes pode ser bem menor. Tal vantagem, contudo, não é exclusiva do método proposto: outros esquemas de controle multivariado da dispersão, como o gráfico da variância generalizada e o gráfico  $V_{max}$ , também possuem essa vantagem — e nesse caso, ainda mais acentuada, pois utilizam uma única estatística e um único gráfico de controle. Porém, é de se esperar que o esquema proposto, por monitorar a dispersão das estimativas das próprias variáveis latentes (não observáveis) que correspondem às fontes de variabilidade do

processo, seja mais eficiente que esquemas que confundam (ou misturem) os efeitos de diferentes fontes de variação. Finalmente, ao se ter um gráfico separado associado a cada fonte de variação, a questão da interpretação do sinal dos gráficos fica trivialmente resolvida.

### 2. Projeções

Em diversos processos multivariados, a relação entre as características de qualidade, as variações aleatórias e as direções das perturbações conhecidas, assim como os ruídos próprios de medição, pode ser representada pelo modelo matricial abaixo:

$$\mathbf{x} = \mathbf{C}\mathbf{d} + \mathbf{e} \quad (1)$$

onde:

$\mathbf{x}$  é o vetor  $p \times 1$  das variáveis observáveis;

$\mathbf{C}$  é a matriz  $p \times q$  das direções atribuíveis de variação ( $q \leq p$ );

$\mathbf{d}$  é o vetor  $q \times 1$  de variáveis não observáveis que representa as fontes de variação que afetam diretamente o vetor das variáveis observáveis;

$\mathbf{e}$  é o vetor  $p \times 1$  que representa os ruídos de medição e possíveis pequenas variações aleatórias nas variáveis observáveis.

As colunas de  $\mathbf{C}$  são linearmente independentes e representam vetores unitários na direção de cada uma das fontes de variação atribuíveis conhecidas. Que esses vetores sejam unitários pode ser sempre conseguido por uma definição apropriada das unidades de medida dos elementos dos vetores  $\mathbf{d}$  e  $\mathbf{x}$ . No caso típico em que as unidades de  $\mathbf{x}$  são pré-estabelecidas, isso determina a escolha apropriada das unidades de  $\mathbf{d}$  tornam as colunas de  $\mathbf{C}$  vetores unitários.

Para o modelo apresentado na Equação (1) assume-se que  $\mathbf{d} \sim N(0, \Sigma_{\mathbf{d}})$ ,  $\Sigma_{\mathbf{d}}$  = diagonal  $(\sigma_1^2, \dots, \sigma_q^2)$ ,  $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I}_{p \times p})$ , independente de  $\mathbf{d}$ . Causas especiais de variação podem afetar  $E(\mathbf{d})$  ou aumentar  $\sigma_j^2$ , porém  $\Sigma_{\mathbf{d}}$  permanece diagonal. Tais pressupostos são válidos em muitas situações reais (ver por exemplo Apley e Shi (2001)).

Baseados no modelo de processo descrito pelas hipóteses apresentadas anteriormente, faz sentido monitorar o processo pelas variáveis latentes em vez de pelas variáveis observáveis. No entanto, como o vetor de variáveis não observáveis  $\mathbf{d}$  não é diretamente observável, ele precisa ser estimado. Para tanto, pode-se utilizar como estimador a projeção do vetor de variáveis observáveis  $\mathbf{x}$  no subespaço das direções atribuíveis representadas pelas colunas de  $\mathbf{C}$  (na base vetorial formada pelas colunas de  $\mathbf{C}$ ), ou seja:

$$\hat{\mathbf{d}} = Proj_{\mathbf{C}}^{\mathbf{x}} = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{x}$$

Aplicando a condição de ortogonalidade  $\mathbf{C}^T \mathbf{C} = \mathbf{I}_{q \times q}$ , vem:

$$\hat{\mathbf{d}} = \mathbf{C}^T \mathbf{x} \quad (2)$$

Para verificar a relação dessas projeções com as variáveis latentes e os ruídos, ao substituir a Equação (1) na Equação (2), vem:

$$\hat{\mathbf{d}} = \mathbf{d} + \mathbf{C}^T \mathbf{e} \quad (3)$$

Aqui nos preocupamos com o monitoramento da dispersão do processo. Da Equação (3) é trivial obter que

$$\Sigma_{\hat{\mathbf{d}}} = \Sigma_{\mathbf{d}} + \mathbf{C}^T \Sigma_{\mathbf{e}} \mathbf{C} = \Sigma_{\mathbf{d}} + \sigma^2 \mathbf{I}_{q \times q}. \quad (4)$$

A Equação (4) mostra que é válido o monitoramento da variabilidade do processo por gráficos separados nas variâncias (ou nos desvios-padrão) amostrais dos elementos de  $\hat{\mathbf{d}}$ . O monitoramento por gráficos separados traz a vantagem adicional de que um sinal em um dos gráficos já indicará automaticamente qual a variável latente que teve sua dispersão aumentada (ou seja, qual a fonte de variação atribuível afetada pela causa especial). Assim, aqui propomos e analisamos o desempenho de um conjunto de gráficos de  $S$  nos elementos de  $\hat{\mathbf{d}}$ .

### 3. O esquemas de controle proposto, $|S|$ e VMAX

Periodicamente, retira-se uma amostra de tamanho  $n > 1$  do vetor de variáveis observáveis  $\mathbf{x}$ , e projetam-se esses  $n$  vetores em cada uma das direções atribuíveis, pela Equação (2). Obtêm-se, assim,  $n$  vetores  $\hat{\mathbf{d}}$ . Calculam-se então os  $q$  desvios-padrão amostrais de cada elemento de  $\hat{\mathbf{d}}$ :

$$S_j = \sqrt{\frac{(\hat{d}_{ij} - \bar{\hat{d}}_j)^2}{n-1}}; \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, q$$

onde

$$\bar{\hat{d}}_j = \frac{1}{n} \sum_{i=1}^n \hat{d}_{ij}$$

é a média amostral das  $n$  estimativas de  $d_j$  (projeções de  $\mathbf{x}$  na  $j$ -ésima direção atribuível) na amostra.

Para cada direção atribuível  $j$ , os valores de  $S_j$  nas amostras sucessivas são lançados em um gráfico de  $S$  unilateral (já que o propósito é detectar rapidamente aumentos na variância de  $d_j$ ) com limite superior para uma probabilidade de alarme falso  $\alpha$  dado por

$$LSC_j = S_{j \text{ pooled}} \sqrt{\frac{\chi_{n-1, \alpha}^2}{n-1}} \quad (5)$$

onde  $S_{j \text{ pooled}}$  é o desvio-padrão combinado dos  $m$  valores de  $S_j$  ( $S_{j1}, S_{j2}, \dots, S_{jm}$ ) em  $m$  amostras iniciais do processo suposto em controle, ou seja,

$$S_{j \text{ pooled}} = \sqrt{\frac{\sum_{k=1}^m S_{jk}^2}{m}}; \quad k = 1, 2, \dots, m$$

e constitui uma estimativa do desvio-padrão de  $\hat{d}_j$ . De fato, como o desvio-padrão de  $\hat{d}_j$  é desconhecido, precisa ser estimado. A Equação (4) mostra que, mesmo no caso em que a matriz de variância-covariância de  $\hat{\mathbf{d}}$  é diagonal, ela é a soma da matriz de variância-covariância dos ruídos  $\mathbf{e}$  (diagonal, e que pode, no caso restrito de esses ruídos deverem-se apenas ao sistema de medição, ser estimada por experimentos específicos) com a matriz de variância-covariância de  $\mathbf{d}$ , que, essa, não pode ser estimada diretamente. A solução é estimar diretamente as variâncias (e os desvios-padrão) dos  $q$  elementos de  $\mathbf{d}$ . Outros estimadores poderiam ser usados (como  $\bar{S}$ ), mas propõe-se usar o desvio-padrão combinado por ser mais eficiente (Mahmoud et al., 2010).

Note-se, ainda, que, para ajustar a probabilidade de alarme falso conjunta dos  $q$  gráficos em  $S_j$  a um valor  $\alpha_{conj}$  especificado, a probabilidade de alarme falso em cada gráfico individual (a ser utilizada na determinação dos limites de controle para os gráficos, pela Equação (5)), deve ser calculada por

$$\alpha_{indiv} = 1 - (1 - \alpha_{conj})^{1/q}. \quad (6)$$

Duas outras estatísticas utilizadas para monitorar a dispersão de um processo multivariado são: a Variância Generalizada  $|\mathbf{S}|$  (Alt, 1985) e a Variância Máxima VMAX (Costa e Machado, 2009). Essas duas estatísticas concorrentes são sensíveis a alterações nos parâmetros da matriz de covariância do processo e conseguem sinalizar, com um único gráfico de controle, desvios no processo de produção.

A variância generalizada  $|\mathbf{S}|$  é definida como o determinante da matriz de covariâncias teórica do processo. Periodicamente, extrai-se uma amostra de tamanho  $n > 1$  do vetor de variáveis observáveis, calcula-se a sua matriz de covariâncias amostral  $\mathbf{S}$ , e o seu determinante  $|\mathbf{S}|$  é calculado e comparado com os limites de controle estabelecidos.

Por sua vez, Costa e Machado (2009) propuseram monitorar a variabilidade do processo pelo maior valor das variâncias amostrais padronizadas. Periodicamente, extrai-se uma amostra de tamanho  $n > 1$  do vetor de variáveis e calculam-se as variâncias amostrais de



## PESQUISA OPERACIONAL PARA O DESENVOLVIMENTO

cada uma das  $p$  variáveis separadamente, ou seja,  $\{S_1^2, S_2^2, \dots, S_p^2\}$ . A estatística VMAX é simplesmente:

$$VMAX = \text{máx}\{S_1^2, S_2^2, \dots, S_p^2\}.$$

Se a estatística VMAX apresentar valor maior que o limite superior de controle, isso é considerado sinal de que a dispersão de uma ou mais das variáveis aumentou.

### 4. Definindo o Experimento

Para ilustrar a aplicação do procedimento proposto, bem como realizar uma análise de desempenho do mesmo na sinalização de aumentos na dispersão das variáveis latentes, será utilizado o exemplo de montagem de portas traseiras em automóveis apresentado em Maranhão e Epprecht (2011). Nesse exemplo os autores mostraram que durante o processo de ajuste entre a porta traseira de um automóvel, representada na Figura 1 pelo retângulo menor – interno, e a carroceria (retângulo maior – externo), diferenças da posição da porta em relação à posição especificada podem ser representadas por variáveis latentes associadas a certos padrões de desvio (como rotação ou deslocamento horizontal) que não são as variáveis diretamente observadas (estas, no caso, são folgas entre a porta e a carroceria, medidas em pontos específicos).

Entre a carroceria e a porta são realizadas medições de folgas em quatro pontos específicos, aqui representados pelas letras A, B, C e D (carroceria - parte externa) e 1, 2, 3 e 4 (porta - parte interna). As folgas  $\overline{A1}$ ,  $\overline{B2}$ ,  $\overline{C3}$  e  $\overline{D4}$  são medidas pelas distâncias horizontais do pontos externos e internos, assim as folgas obtidas medem as variações sofridas pela porta na direção horizontal (esquerda e direita).

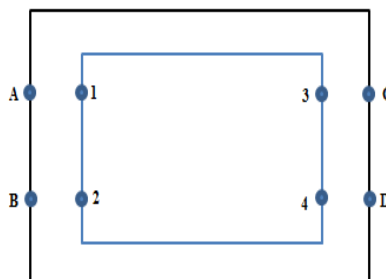


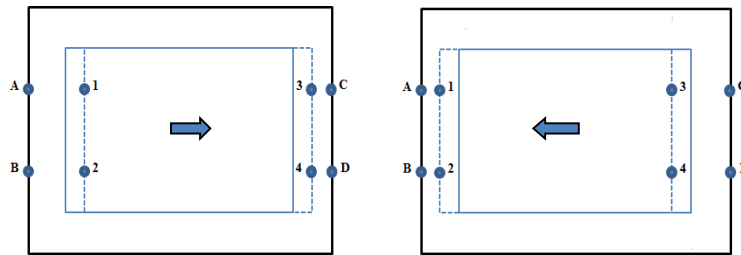
Figura 1- Exemplo proposto por Maranhão e Epprecht

As variações nas medidas das folgas são provocadas por variações nos padrões de montagem, como por exemplo: um deslocamento, uma rotação, uma translação lateral, uma

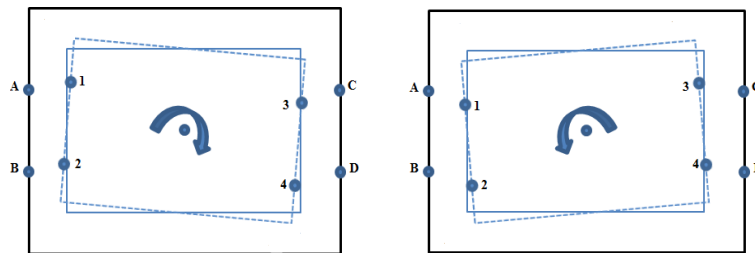
## PESQUISA OPERACIONAL PARA O DESENVOLVIMENTO

torção ou até mesmo uma contração/dilatação na porta traseira. No exemplo, Maranhão e Epprecht (2011) por simplificação, utilizaram apenas os padrões de variação de montagem associados à rotação, ao deslocamento e à contração.

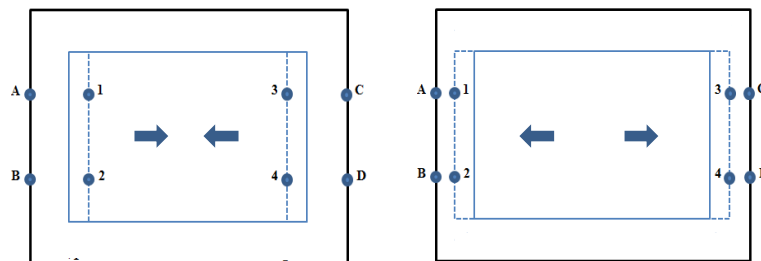
A Figura 3 apresenta o padrão de variação associado à rotação, a Figura 2 ao deslocamento e a Figura 4 à contração/dilatação.



**Figura 2 - Deslocamento lateral positivo e negativo**



**Figura 3 – Rotação positiva e negativa**



**Figura 4 - Contração/dilatação**

Cada uma dessas variações corresponde a uma direção específica do vetor de variáveis observáveis, ou seja, uma direção atribuível de variação, pois cada uma corresponde a uma causa especial de variação.

Na rotação as variações nas variáveis observáveis não são lineares, pois dependem das relações trigonométricas associadas ao ângulo de rotação. Com uma pequena perda de

## PESQUISA OPERACIONAL PARA O DESENVOLVIMENTO

informação, pode-se generalizar o modelo para uma variação linear, levando-se em conta apenas pequenos ângulos de rotação.

Na rotação anti-horária os valores das folgas medidas nos pontos  $\overline{A1}$  e  $\overline{D4}$  diminuem, enquanto que os valores das folgas medidas nos pontos  $\overline{B2}$  e  $\overline{C3}$  aumentam. Da mesma forma teríamos para a rotação horária uma diminuição nos valores das folgas medidas nos pontos  $\overline{B2}$  e  $\overline{C3}$  e um aumento no valor das folgas medidas nos pontos  $\overline{A1}$  e  $\overline{D4}$ .

No deslocamento para a direita os valores das folgas medidas nos pontos  $\overline{A1}$  e  $\overline{B2}$  aumentam, enquanto que os valores das folgas medidas nos pontos  $\overline{C3}$  e  $\overline{D4}$  diminuem e para a contração os valores das folgas medidas nos pontos  $\overline{A1}$ ,  $\overline{B2}$ ,  $\overline{C3}$  e  $\overline{D4}$  aumentam. As variáveis observáveis  $X_1$ ,  $X_2$ ,  $X_3$  e  $X_4$  são definidas aqui como os desvios (diferenças) em relação ao valor nominal das folgas medidas nos pontos  $\overline{A1}$ ,  $\overline{B2}$ ,  $\overline{C3}$  e  $\overline{D4}$ .

A rotação e o deslocamento alteram as variáveis observadas de maneira correlacionada. Por exemplo, no caso de uma rotação no sentido anti-horário (veja a Figura 3), uma diferença de  $d$  unidades para menos nas folgas nos pontos  $\overline{A1}$  e  $\overline{D4}$  em relação ao valor nominal coincide com uma diferença de  $d$  unidades para mais das folgas nos pontos  $\overline{B2}$  e  $\overline{C3}$  em relação ao valor nominal. Isso faz com que a média do vetor das variáveis observáveis se desloque  $2d$  unidades na direção do vetor unitário  $[-0,5 \ 0,5 \ 0,5 \ -0,5]^T$ . O mesmo raciocínio deve ser levado em conta para as outras direções atribuíveis. Assim, a matriz dos padrões de variação é:

$$\mathbf{C} = 0,5 \begin{bmatrix} -1 & +1 & +1 \\ +1 & +1 & +1 \\ +1 & -1 & +1 \\ -1 & -1 & +1 \end{bmatrix} \quad (7)$$

Nesse exemplo, indexaremos os elementos do vetor  $\mathbf{d}$  com letras, indicando a rotação central da porta em relação à carroceria pela letra T, o seu deslocamento horizontal pela letra D, e a contração ou dilatação da porta pela letra C.

Nesse caso específico, as Equações (1) e (3) tomam a forma:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = 0,5 \begin{bmatrix} -1 & +1 & +1 \\ +1 & +1 & +1 \\ +1 & -1 & +1 \\ -1 & -1 & +1 \end{bmatrix} \begin{bmatrix} d_T \\ d_D \\ d_C \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \quad (8)$$

$$\begin{bmatrix} \hat{d}_T \\ \hat{d}_D \\ \hat{d}_C \end{bmatrix} = 0,5 \begin{bmatrix} -1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & +1 & +1 & +1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}. \quad (9)$$

## 5. Simulação

A análise de desempenho foi realizada por simulação do usual procedimento em duas fases de Controle Estatístico do Processo — a *Fase I* de estimação de parâmetros e determinação de limites de controle, e a *Fase II* de monitoramento do processo usando os limites determinados na Fase I. Para simular a Fase I, geraram-se dados simulando o processo em controle para estimar os limites de controle de todos os esquemas de controle aqui analisados. Para a Fase II, geraram-se dados simulando aumentos na dispersão das variáveis latentes.

Foram comparados 3 esquemas de controle: o conjunto de gráficos de  $S$  nos desvios-padrão amostrais das projeções  $\hat{d}_j$ , o gráfico da variância amostral generalizada (denotado aqui por  $\text{Det}(S)$  ou  $|\bar{S}|$ ) e o gráfico da máxima variância amostral (VMAX). Para que os três esquemas de controle fossem comparados em igualdade de condições estabeleceu-se uma probabilidade de alarme falso igual a  $1/370,4$  para cada um dos esquemas. Assim, para os gráficos de  $S$  das projeções  $\hat{d}_j$ , os limites de controle foram calculados pelas Equações (5) e (6), de maneira a fornecer uma probabilidade de alarme falso em cada gráfico individual (para cada projeção  $\hat{d}_j$ ) que resulte numa probabilidade de alarmes falsos conjunta igual a  $1/370,4$ . Já os limites de controle dos gráficos de  $|\bar{S}|$  e de VMAX foram obtidos pelo quantil empírico das amostras da Fase I, de forma a obter a probabilidade de alarme falso igual ao estabelecido. Os limites de controle para  $|\bar{S}|$  e VMAX foram aqui definidos pelo quantil devido à não-trivialidade dos cálculos e das respostas aproximadas para o NMA.

Aqui, na simulação da Fase I, foram geradas 3704 amostras pseudo-aleatórias  $\sim N_3(\mathbf{0}, \Sigma_d)$  do vetor  $\mathbf{d}$ , de tamanho 5, assim como 3704 amostras pseudo-aleatórias do vetor  $\mathbf{e}$  de erros aleatórios,  $N_4(\mathbf{0}, \Sigma_e)$ , com  $\Sigma_d = \mathbf{I}_{2 \times 2}$  e  $\Sigma_e = \sigma^2 \mathbf{I}_{4 \times 4}$ . Esse número de amostras não é realista em muitos casos práticos (embora possa sê-lo em ambientes com coleta de dados altamente automatizada e em grande frequência, o que vem se tornando cada vez mais comum), mas o propósito aqui é de comparação do desempenho dos gráficos; com este propósito, decidimos utilizar um número de amostras iniciais (da Fase I) que garantisse boa precisão na determinação dos limites (levando em conta que, no caso de menor número de amostras iniciais, todos os esquemas propostos sofreriam, da mesma forma).

## PESQUISA OPERACIONAL PARA O DESENVOLVIMENTO

Para cada ponto amostral (uma realização do vetor  $\mathbf{d}$  e uma realização do vetor  $\mathbf{e}$ ) obtém-se uma realização do vetor  $\mathbf{x}$  pela Equação (8). Com essa realização, simula-se a aplicação de cada um dos procedimentos de controle, conforme o caso: para os gráficos de S das projeções nas direções atribuíveis, o cálculo dos elementos de  $\hat{\mathbf{d}}$  pela Equação (9) seguido pelo cálculo dos desvios-padrão amostrais desses elementos; o cálculo de VMAX, e o cálculo da variância generalizada.

Os limites dos esquemas de controle foram determinados conforme o procedimento a seguir:

### *Procedimento para o cálculo dos limites de controle*

- *Dados: o número de variáveis observáveis  $p$ , o valor da taxa de probabilidade de alarmes falsos  $\alpha$ , o tamanho das amostras  $n$ , o número de amostras  $m$  e o números de replicações  $R$ ;*
- *Gerar os  $q*n*m$  números pseudo-aleatórios  $N(0,1)$  para representar o vetor das variáveis não observáveis;*
- *Gerar os  $p*n*m$  números pseudo-aleatórios com média zero e desvio-padrão 0,1 para representar o vetor de ruídos aleatórios;*
- *Calcular os valores dos elementos do vetor de variáveis observáveis, pela equação (7);*
- *Agrupar os valores gerados em  $m$  grupos de tamanho  $n$ ;*
- *Calcular VMAX e Det\_S para esses grupos;*
- *Ordenar os valores obtidos nos grupos;*
- *Calcular o quantil  $(1-\alpha)$  dos valores calculados VMAX e Det\_S;*
- *Repetir todos os passos anteriores  $R$  vezes;*
- *Calcular a média dos  $R$  valores dos quantis calculados;*
- *Tomar a média dos quantis calculados igual ao limite de controle.*

Nesse trabalho com o processo em controle, espera-se para os gráficos de Shewhart nos 3 esquemas analisados que a probabilidade de alarme falso seja igual a  $1/370,4$ , ou seja 0,0027. A Tabela 1 mostra que os valores médios obtidos para a probabilidade de alarme falso é praticamente igual para todos os três casos analisados; conjunto dos gráficos nas projeções T e D, denotado por  $\hat{d}_T \cup \hat{d}_D$ , o gráfico de VMAX e gráfico Det(S). Além disso, mostra também os valores médios obtidos para a probabilidade de alarme falso para cada projeção  $\hat{d}_T$  e  $\hat{d}_D$  de forma individual. As probabilidades de alarme falso individuais foram calculadas aplicando a correção de Dunn-Sidak apresentada na Equação (6).

## PESQUISA OPERACIONAL PARA O DESENVOLVIMENTO

Tabela 1 – Probabilidade de alarme falso nos esquemas avaliados

$\sigma_e$ em controle	$\sigma_{\hat{d}_T}$	$\sigma_{\hat{d}_D}$	$\hat{d}_T$	$\hat{d}_D$	$\hat{d}_T \cup \hat{d}_D$	Vmax	Det(S)
0,1	1	1	0,0014	0,0015	0,0029	0,0027	0,0028
0,5	1	1	0,0013	0,0014	0,0027	0,0027	0,0027
1	1	1	0,0013	0,0013	0,0027	0,0027	0,0027

Para a Fase II, foram realizadas 3704 amostras de tamanho 5, com os parâmetros apresentados na Tabela 2.

Para os gráficos de Shewhart é usual medir o desempenho do processo pela probabilidade de ocorrência de alarmes verdadeiros em situações em que o processo apresenta-se fora de controle estatístico, e é esta a medida de desempenho adotada aqui. A probabilidade de sinal foi estimada pela razão entre o número total de alarmes e o número total de amostras simuladas em cada caso. No caso dos gráficos nas projeções, consideramos a probabilidade de sinal em *pelo menos um* dos gráficos bem como em cada gráfico individualmente (o que fornece informação da capacidade do esquema de identificar automaticamente a causa especial, ou pelo menos a variável latente que foi afetada).

A Tabela 2 apresenta os resultados dos dados simulados na Fase II.

Tabela 2 – Probabilidade de sinal nos esquemas avaliados

$\sigma_e$ em controle	$\sigma_T$	$\sigma_D$	$\hat{d}_T$	$\hat{d}_D$	$\hat{d}_T \cup \hat{d}_D$	Det(S)	VMAX
0,1	1	1,5	0,0013	0,0863	<b>0,0875</b>	0,0117	0,0505
0,1	1	2	0,0014	0,3308	<b>0,3318</b>	0,0272	0,2132
0,1	1,5	1,5	0,0925	0,0860	0,1707	0,0364	<b>0,1761</b>
0,1	1,5	2	0,0932	0,3308	<b>0,3931</b>	0,0681	0,3807
0,1	2	2	0,3458	0,3309	0,5623	0,1166	<b>0,5677</b>
0,5	1	1,5	0,0012	0,0610	<b>0,0621</b>	0,0097	0,0306
0,5	1	2	0,0013	0,2596	<b>0,2606</b>	0,0222	0,1334
0,5	1,5	1,5	0,0612	0,0611	<b>0,1185</b>	0,0272	0,1039
0,5	1,5	2	0,0611	0,2584	<b>0,3038</b>	0,0523	0,2497
0,5	2	2	0,2585	0,2582	<b>0,4499</b>	0,0894	0,4099
1	1	1,5	0,0015	0,0258	<b>0,0273</b>	0,0068	0,0121
1	1	2	0,0014	0,1266	<b>0,1278</b>	0,0142	0,0437
1	1,5	1,5	0,0285	0,0258	<b>0,0536</b>	0,0148	0,0321
1	1,5	2	0,0285	0,1269	<b>0,1517</b>	0,0272	0,0825
1	2	2	0,1343	0,1278	<b>0,2450</b>	0,0469	0,1561

### 6. Resultados

Observa-se na Tabela 2 que o esquema proposto (conjunto de gráficos nas projeções) apresenta vantagem em relação aos demais esquemas concorrentes. A probabilidade conjunta de sinal nesse esquema é praticamente sempre superior aos outros esquemas, além da vantagem adicional de identificar a variável latente que se alterou.

Para maiores valores de  $\sigma_e$ , o desempenho de todos os esquemas se deteriora, mas o dos gráficos nas projeções se deteriora menos que os dos gráficos de VMAX e Det(S). Comparando estes últimos, verifica-se que, quando só uma das variáveis latentes tem sua dispersão aumentada, o desempenho do gráfico de Det(S) se deteriora mais (com o aumento de  $\sigma_e$ ) que o do gráfico de VMAX; quando ambas variáveis latentes têm a dispersão aumentada, a situação se inverte: é o desempenho do gráfico de VMAX que se deteriora mais rapidamente.

O esquema proposto só perde para algum outro (a saber, para o gráfico de Det(S)) no caso de grandes aumentos nas variâncias de ambas as variáveis latentes, mas mesmo assim só no caso de  $\sigma_e$  pequeno, e a diferença (entre o número de sinais do esquema proposto e do gráfico de Det(S)) é bem pequena. Quando  $\sigma_e$  é maior a deterioração do desempenho dos esquemas analisados faz com que o esquema proposto seja mais eficiente que todos os outros.

Os resultados obtidos na Tabela 2 também comprovam a independência entre os sinais nos gráficos nas projeções propostas de T e D. Para os mesmos valores de  $\sigma_e$ , o número de sinais nas projeções de T e D depende apenas do aumento da variância da variável latente correspondente.

### 7. Conclusão

Foi proposto neste trabalho um esquema de controle estatístico da variabilidade de processos multivariados, idealizado para o caso em que as variações do processo devem-se a fontes que atuam sobre variáveis não diretamente observáveis, e que consiste em gráficos de S de projeções do vetor das variáveis observáveis.

Na comparação de desempenho com esquemas concorrentes (gráficos de controle da variância generalizada e de VMAX), o esquema proposto mostrou considerável vantagem: sua probabilidade de sinal é praticamente sempre superior à destes últimos; ele ainda apresenta a vantagem de identificar a variável latente cuja dispersão aumentou.

## PESQUISA OPERACIONAL PARA O DESENVOLVIMENTO

Desta forma, conclui-se que o esquema proposto é uma alternativa muito interessante para o controle da variabilidade de processos multivariados.

### Agradecimentos

O segundo autor agradece o apoio do CNPq.

### Referências

- Alt, F. B. (1985). *Multivariate Quality Control*. Encyclopedia of the Statistical Sciences. New York, John Wiley & Sons.
- Apley, W. D. & Shi, J. (2001). A Factor-Analysis Method for Diagnosing Variability in Multivariate Manufacturing Processes. American Statistical Association and the American Society for Quality. *Technometrics*, 43, 84-95.
- Barton, R. R. & González-Barreto, D. R. (1996). Process-oriented basis representations for multivariate process diagnostics. *Quality Engineering*, 9, 107-118.
- Costa, A. F. B. & Machado, M. A. G. (2009). A new chart based on the sample variances for monitoring the covariance matrix of multivariate processes. *International Journal of Advanced Manufacturing Technology*, 41, 770-779.
- Djauhari, M. A. (2005). Improved monitoring of multivariate process variability. *Journal of Quality Technology*, 37, 32-39.
- Mahmoud, M. A.; Henderson, G. R.; Epprecht, E. K.; & Woodall, W. H. (2010) Estimating the Standard Deviation in Quality Control Applications. *Journal of Quality Technology*, 42(4), pp. 348-357.
- Maranhão, P. H. & Epprecht, E. K. (2011). O Uso de Variáveis Transformadas no Controle Estatístico de Processos Multivariados. *PODes*, 3, 201-217.
- Runger, G. C (1996). Projections and  $U^2$  Multivariate Control Chart. *Journal of Quality Technology*, 28(3), pp. 313-319.
- Yeh, A. B., Lin, D. K. J & McGrath, R. N. (2006). Multivariate Control Charts for Monitoring Covariance Matrix: A Review. *Quality Technology & Quantitative Management*, 4, 415-436.