

# Controle multivariado de processos em batelada com duração variável

Flávio S. Fogliatto UFRGS

Ndèye Niang CNAM/Paris

## RESUMO

Processos em batelada são utilizados em diversos setores industriais (por exemplo, na manufatura de alimentos e fármacos). Nesses processos, matérias-primas são carregadas em uma unidade de processamento e submetidas a uma série de transformações até a obtenção do produto final. O desempenho do processo é descrito por variáveis, monitoradas ao longo da batelada. Dados resultantes desses processos tendem a apresentar uma estrutura de correlação e autocorrelação significativa, sendo usualmente monitorados usando cartas de controle baseadas na análise de componentes principais (CCPs). Neste artigo, investiga-se o caso especial, bastante freqüente na prática, de bateladas com duração variável, as quais não podem ser diretamente monitoradas através das CCPs. Para tanto, propõe-se uma nova estratégia de controle multivariado da qualidade. No procedimento proposto, bateladas não são alinhadas ou *time warped* relativamente a suas trajetórias, mas completadas utilizando um esquema de fácil implementação prática. Desta forma, preserva-se toda a informação sobre a variabilidade ao longo do eixo do tempo nos perfis das variáveis de processo. O conjunto de dados completados é analisado utilizando o método Statis e o monitoramento do desempenho da batelada é realizado diretamente nos gráficos de planos fatoriais, a partir dos quais cartas de controle não-paramétricas são derivadas. Um exemplo utilizando dados simulados ilustra a proposta metodológica.

## PALAVRAS-CHAVE

Bateladas não-sincronizadas, CEP multivariado, método Statis.

## *Multivariate statistical control of unsynchronized batch processes*

### ABSTRACT

*Batch processes are widely used in several industrial sectors, such as food and pharmaceutical manufacturing. In a typical batch, raw materials are loaded in the processing unit and submitted to a series of transformations, yielding the final product. Process performance is described by variables which are monitored as the batch progresses. Data arising from such processes are likely to display a strong correlation-autocorrelation structure, and are usually monitored using control charts based on multiway principal components analysis (MPCA charts). In this paper we investigate the special (and rather frequent) case of batches with varying duration, which cannot be directly monitored using MPCA charts. We propose a new quality control strategy for monitoring such batches. In our proposition, batches are not aligned or time warped with respect to their trajectories, but are rather completed using a straightforward scheme. Thus all information on the variability in batch profiles along the time axis is preserved. The data set completed is reduced using the Statis method and monitoring of batch performance is accomplished directly on principal plane graphs, from which non-parametric control charts are derived. A simulated example illustrates the proposed method.*

### KEY WORDS

*Unsynchronized batches, Multivariate quality control, Statis method.*

## 1. INTRODUÇÃO

Cartas de controle (CCs) univariadas de Shewhart são utilizadas comumente no controle estatístico de processos industriais (MONTGOMERY, 2001). Tal estratégia de controle univariado, 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 apresentados por CCs multivariadas e univariadas são particularmente diferentes quando as variáveis monitoradas apresentam-se 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  $T^2$ . Outros tipos de CCs multivariadas são apresentados por Jackson (1991) e revisados por Wierda (1994), Lowry e Montgomery (1995), Harris, Seppala e Desborough (1999).

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, dificilmente verificada em processos monitorados com coleta automatizada de dados. Outra situação especial ocorre quando a variação natural de uma variável monitorada ao longo do tempo é descrita por um perfil (e não por uma variação aleatória em torno de um valor médio). CCs univariadas não podem ser aplicadas nessa situação, já que pressupõem variáveis com média constante. A alternativa mais utilizada, nesses casos, para o monitoramento de processos são as cartas de controle multivariadas baseadas em componentes principais multidirecionais (CCPs).

CCPs são operacionalizadas utilizando análise de componentes principais multidirecionais e CCs multivariadas de Hotelling. Essas cartas admitem pontos amostrais sucessivos dependentes, variáveis com estrutura de correlação linear significativa, além de permitir o monitoramento de variáveis que não apresentam valores fixos de média ao longo do tempo. CCPs são particularmente eficientes no monitoramento de processos discretos (tais como os que ocorrem em bateladas) e contínuos, quando um grande número de variáveis de processo estiver envolvido. O desenvolvimento das CCPs deve-se a Jackson e Mudholkar (1979), Kourti e MacGregor (1996) e MacGregor (1997). Aplicações das CCPs no monitoramento de processos podem ser encontradas na literatura nos trabalhos de Flores-Cerrillo e MacGregor (2002) e Kourti (2003), entre outros.

Duas das suposições que restringem a utilização das CCPs no monitoramento de alguns processos em bateladas são de

interesse neste artigo. A primeira diz respeito à duração da batelada. Para aplicar CCPs no monitoramento de processos em bateladas, todas as bateladas consideradas na análise devem apresentar uma mesma duração, além de estarem alinhadas em relação aos estágios do processo. Quando tal suposição não se verifica, as CCPs devem ser adaptadas de forma a considerar diferenças na duração das bateladas. Entretanto, as propostas de solução para o problema disponíveis na literatura nem sempre são eficientes, conforme discutido na Seção 2. A segunda suposição está relacionada à distribuição que caracteriza os dados de processo analisados na CCP. Para um correto uso da carta  $T^2$  no monitoramento do conjunto de dados reduzidos, estes devem ser caracterizados por uma distribuição normal multivariada. Quando tal suposição não se verificar, o uso de cartas de controle não-paramétricas deveria ser considerado.

## Processos em batelada são utilizados em diversos setores industriais; por exemplo, na manufatura de alimentos e produtos farmacêuticos

Neste artigo, propõe-se uma nova estratégia para o monitoramento de processos em bateladas com duração variável. Nessa proposta, bateladas não são alinhadas ou *time warped* em relação a suas trajetórias; como alternativa, completam-se bateladas de curta duração através de um procedimento simplificado e de fácil implementação prática. Assim, toda a informação acerca da variabilidade das bateladas ao longo do eixo do tempo é preservada. O conjunto de dados completos é então reduzido utilizando o método Statis (LAVIT *et al.*, 1994). Duas representações sumárias das bateladas tornam-se disponíveis: uma, que permite verificar a adequação da progressão da batelada à trajetória de referência de suas variáveis de controle, e outra, que resume as trajetórias das variáveis em cada instante da batelada, permitindo detectar desvios significativos em seu comportamento esperado ao longo do eixo do tempo. O monitoramento do desempenho da batelada com relação às duas representações acima é realizado diretamente em gráficos de planos fatoriais, a partir dos quais CCs não-paramétricas são derivadas. O método proposto é ilustrado utilizando dados simulados disponíveis em Rosa (2005).

A proposta aqui apresentada é uma extensão do trabalho de Scepi (2002), onde o uso do método Statis no controle multivariado da qualidade foi inicialmente proposto. Entretanto, ao menos duas contribuições distinguem o presente trabalho da proposta em Scepi (2002): (i) o método aqui proposto permite o monitoramento de bateladas com du-

ração variada, não discutido pela autora; e (ii) este trabalho apresenta o desenvolvimento de cartas de controle e estratégias para complementação de dados adequados para o monitoramento *on line* de processos em batelada. Os métodos apresentados em Scepi (2002) aplicam-se diretamente no controle *off line* de processos.

O restante do artigo está organizado em cinco seções. Na Seção 2, apresenta-se uma revisão de literatura sobre monitoramento multivariado de processos em bateladas com duração variável. Na Seção 3, apresenta-se a estratégia de controle multivariado da qualidade proposta. A seção é subdividida em três partes: inicialmente apresenta-se o esquema de complementação de bateladas; na seqüência, detalha-se a utilização do método Statis na redução da dimensionalidade dos dados de processo; por fim, cartas de controle não-paramétricas para o monitoramento *off-line* e *on-line* de processos em batelada são apresentadas. Na Seção 4 apresenta-se um exemplo numérico onde a aplicação das cartas de controle propostas é ilustrada. A Seção 5 apresenta a conclusão do artigo.

## 2. REFERENCIAL TEÓRICO

A aplicação de CCPs no monitoramento de processos em batelada de duração constante foi inicialmente proposta por Jackson e Mudholkar (1979), sendo posteriormente investigada por Nomikos e MacGregor (1995), Kourti e MacGregor (1996) e MacGregor (1997). Aplicações de CCPs no monitoramento de processos em batelada podem ser encontradas em Kourti (2003), entre outros.

De forma resumida, o monitoramento de processos em batelada através de CCPs é conduzido a partir da análise de duas CCs. A primeira é uma CC do tipo  $T^2$  onde se plotam os escores obtidos projetando bateladas futuras sobre os  $q$  componentes principais retidos na distribuição de referência. Tal distribuição é obtida rodando-se uma ACPM (Análise de Componentes Principais Multidirecionais). Ao contrário da análise tradicional de componentes principais, a ACPM permite a análise de matrizes tridimensionais de dados (para um comparativo entre técnicas, ver Jolliffe, 2004, p. 397) sobre dados obtidos a partir de bateladas conformes emergentes do processo. A segunda é uma CC do tipo  $Q$  para os resíduos do modelo de referência. A primeira CC permite monitorar o comportamento de fontes conhecidas de variabilidade no processo; a segunda permite detectar eventos atípicos que possam perturbar a estrutura de correlação/autocorrelação das variáveis em monitoramento.

No esquema de monitoramento baseado em CCPs,

pressupõem-se bateladas sincronizadas e com a mesma duração; isto é, todos os vetores de dados da distribuição de referência, bem como aqueles obtidos em bateladas futuras, têm a mesma dimensão. De outra forma, têm-se bateladas não alinhadas no eixo do tempo, não permitindo a utilização direta do procedimento descrito anteriormente. Tal situação especial foi inicialmente identificada por Nomikos (1995), que sugeriu o uso de uma variável indexadora alternativa (isto é, outra que não o tempo) para monitorar a progressão temporal da batelada. Tal variável deve (i) ser diretamente mensurável no processo, (ii) crescer monotonicamente com o tempo e (iii) apresentar os mesmos valores iniciais e finais em cada batelada. Apesar de serem bastante restritivos, tais requisitos foram satisfeitos nas aplicações reportadas por Kourti et al. (1996) e Neogi e Schlags (1998).

Kassidas et al. (1998) propuseram uma abordagem alternativa para o problema de bateladas com duração distinta, onde bateladas são alinhadas usando algoritmos de deflexão temporal dinâmica (*dynamic time warping* – DTW). Dados oriundos das bateladas alinhadas são então analisados usando ACPM, sendo o processo monitorado através de CCPs. Os algoritmos DTW, entretanto, apresentam algumas limitações intrínsecas. Por exemplo, trajetórias são alinhadas aos pares, o que pode resultar: (i) em uma trajetória comum diferente para cada par de bateladas (no caso de algoritmos simétricos), ou (ii) em trajetórias alinhadas para coincidir com uma trajetória de referência comum a todos os pares de trajetórias (no caso de algoritmos assimétricos). Na situação (i) o banco de dados resultante será ainda constituído de trajetórias com durações distintas, não resolvendo o problema inicial; na situação (ii) permite-se que dados sejam desconsiderados durante o alinhamento de algumas bateladas possibilitando, assim, que a duração dos pares de bateladas coincida com a trajetória de referência. Neste caso em particular, os autores propõem uma solução de compromisso, em que conjuntos de pontos são reduzidos a um valor de média e nenhum dado é ignorado durante o procedimento de alinhamento. Entretanto, a maior limitação no procedimento de Kassidas et al. (1998) diz respeito à representação da variação das bateladas ao longo do eixo do tempo, a qual é alterada quando estágios do processo em batelada são sincronizados.

Kaistha et al. (2004) utilizam um algoritmo DTW *ad hoc* para alinhar bateladas de diferentes durações. Em sua abordagem, supõe-se que a progressão das bateladas possa ser dividida em etapas distintas. Pontos de referência em cada etapa (tais como o seu início e término) são utilizados como pontos de âncora em um procedimento simplificado de interpolação, resultando em bateladas que são sincronizadas etapa a etapa. Uma vez alinhados, os dados das bate-

ladas são analisados usando ACPM e CCPs, com pequenas modificações.

Finalmente, Castagliola & Rosa (2006) investigaram o problema de processos em batelada com duração distinta a partir de uma nova abordagem analítica, sem a utilização de técnicas para o alinhamento de bateladas e de redução dimensional de dados. Os autores propõem o uso da distância de Hausdorff como medida de similaridade entre bateladas e uma batelada de referência (contendo a trajetória nominal das variáveis de processo). Tal distância corresponde à mediana do quadrado das distâncias euclidianas entre pontos em uma dada trajetória e todos os pontos na trajetória de referência. A distribuição de referência usada no monitoramento da batelada é baseada nas distâncias de Hausdorff. Apesar de sua simplicidade e dos bons resultados obtidos na aplicação do método a um conjunto de dados simulados, não se tem evidência de que a referida distância capture a estrutura de correlação/autocorrelação presente nas variáveis originais.

### 3. MÉTODO PROPOSTO

Seja  $N$  o número total de bateladas ( $b = 1, \dots, N$ ) usadas na amostra de referência, a partir da qual será determinada uma distribuição de referência a ser utilizada no monitoramento de bateladas futuras. A amostra de referência deve incluir somente bateladas que resultaram em produtos conformes com as especificações. No caso de bateladas fortemente desalinhadas e com durações muito distintas,  $N$  deve ser suficientemente grande para que todos os perfis e durações já observados no processo sob controle estejam bem representados na amostra de referência.

Associada a cada batelada existe uma tabela de dados descrevendo o desempenho de variáveis de processo ao longo da batelada. Um total de  $P$  variáveis de processo são observadas em  $T$  momentos no tempo. Dados coletados a partir da batelada  $b$  são organizados em uma matriz  $X^b$  de dimensão ( $T \times P$ ); cada elemento  $X_{t,p}^b$  ( $t = 1, \dots, T; p = 1, \dots, P$ ) da matriz traz uma observação da  $p$ -ésima variável de processo no instante  $t$ . Seja  $T_{\min}$  a duração da batelada mais curta e  $T_{\max}$  a duração da batelada mais longa.

O método aqui proposto é operacionalizado em três etapas: (i) tratamento inicial das bateladas de referência, onde matrizes de dados provenientes de bateladas com duração total  $T < T_{\max}$  (ou seja, bateladas mais curtas do que a batelada mais longa na amostra de referência) são completadas de forma a apresentar a mesma duração [isto é, todas as matrizes de dados  $X^b$ ,  $b = 1, \dots, N$ , passam a ter dimensão ( $T_{\max} \times P$ )]; (ii) redução da dimensionalidade dos dados usando o método Statis; e (iii) desenvolvimento de cartas

de controle para o monitoramento *off-line* e *on-line* das bateladas futuras. Nas próximas subseções, cada etapa é detalhada.

#### 3.1 Tratamento inicial de dados nas bateladas de referência

O objetivo nesta etapa é completar bateladas na amostra de referência de modo que todas passem a apresentar a duração da batelada mais longa ( $T_{\max}$ ). Este procedimento se justifica, já que matrizes de dados de mesma dimensão são necessárias para a operacionalização da etapa seguinte do método, baseada no método Statis. Optando-se por matrizes de dimensão ( $T_{\max} \times P$ ), isto é, com duração igual à da batelada mais longa, nenhuma informação sobre o comportamento das variáveis nas bateladas da amostra de referência é descartada.

Uma das abordagens mais utilizadas para completar dados faltantes em bateladas, oriunda da proposta de monitoramento *on-line* de processos de Nomikos & MacGregor (1995), consiste em substituir valores faltantes pelo último valor observado de cada variável de processo. Por exemplo, se  $T_{\max} = 25$ , a duração total observada em uma determinada batelada  $b$  for  $T = 20$  e  $X_{20,1} = 3,5$ ,  $X_{20,1}$  a  $X_{25,1}$  apresentariam valor igual a 3,5. Tal procedimento, utilizado devido à sua simplicidade, traz como desvantagem um potencial amortecimento e descaracterização da variabilidade nos tempos para os quais as variáveis de processo forem completadas na matriz de dados. A proposta apresentada na seqüência contorna este problema.

Inicialmente, calculam-se os desvios-padrão amostrais  $S_{t,p}$  de cada variável  $p$  ( $p = 1, \dots, P$ ) em todos os momentos no tempo  $t$  para os quais  $t \leq T_{\min}$ , através das expressões:

$$S_{t,p} = \sqrt{\frac{\sum_{b=1}^N (X_{t,p}^b - \bar{X}_{t,p})^2}{N-1}} \quad e \quad (1)$$

$$\bar{X}_{t,p} = \frac{\sum_{b=1}^N X_{t,p}^b}{N} \quad (2)$$

Para cada variável  $p$  existirão  $T_{\min}$  desvios-padrão, calculados a partir das observações da variável em questão nas bateladas da amostra de referência em cada instante do tempo. Seja  $s_p$  um vetor de dimensão ( $T_{\min} \times 1$ ) contendo os desvios-padrão amostrais associados à  $p$ -ésima variável, ordenados de  $S_{1,p}$  até  $S_{T_{\min},p}$ . Propõe-se ajustar um modelo de previsão à série temporal de observações de  $S_p$  e utilizá-lo para prever os valores  $S_{t,p}$  de  $t = T_{\min+1}$  a  $t = T_{\max}$ . Caso os desvios-padrão não sejam modeláveis como função do tempo, a média dos desvios em  $s_p$  passa a ser usada como estimativa de  $S_{t,p}$  de  $t = T_{\min+1}$  a  $t = T_{\max}$ .

Seja  $n_t$  o número de bateladas na amostra de referência com tempo de término  $T \geq t$ . Quando  $t = T_{\min}$ ,  $n_t = N$ , isto é, o número total de bateladas na amostra de referência; para valores de  $t > T_{\min}$ ,  $n_t < N$ . Utilizando a eq. (2), com somatório adaptado ao número de bateladas disponíveis, é possível calcular as médias amostrais da variável  $p$  a partir das observações da variável somente em bateladas que finalizaram após o tempo  $T_{\min}$ ; isto é  $\bar{X}_{t,p}$ ,  $t > T_{\min}$ . Analogamente, utilizando uma adaptação da eq. (1) é possível calcular os desvios-padrão amostrais da variável  $p$  a partir de suas observações somente em bateladas que finalizam após o tempo  $T_{\min}$ ; isto é  $S_{t,p}$ ,  $t > T_{\min}$ . Cabe salientar que o número de observações incluídas no cálculo de  $\bar{X}_{t,p}$  e  $S_{t,p}$  será variável, dependendo do número de bateladas cujo término ocorre após  $T_{\min}$ .

Na seqüência, propõe-se calcular um desvio-padrão combinado utilizando a seguinte expressão:

$$S_{t,p}^C = \left[ \frac{n_t}{N} \times S_{t,p} \right] + \left[ \left( 1 - \frac{n_t}{N} \right) \times \hat{S}_{t,p} \right], \quad t > T_{\min} \quad (3)$$

onde  $\hat{S}_{t,p}$  designa os valores previstos de  $S_{t,p}$  a partir do modelo de previsão determinado anteriormente. Trata-se de uma média ponderada de duas estimativas de desvio-padrão para a variável  $p$  no tempo  $t$ , onde o elemento ponderador considera o número de bateladas finalizadas em  $t$ .

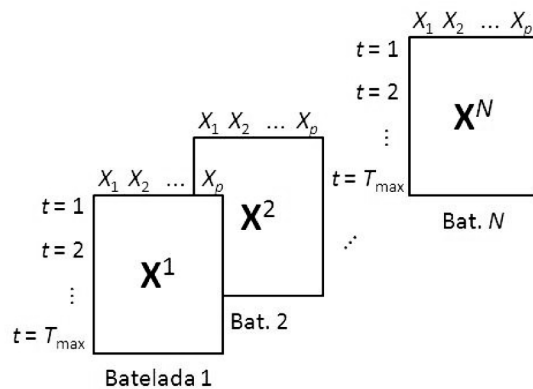
A partir dos desenvolvimentos acima, para cada tempo encontram-se disponíveis pares  $(\bar{X}_{t,p}, S_{t,p}^C)$  de parâmetros, os quais são então utilizados para simular observações de  $(T_{\max} - T_{\min})$  variáveis aleatórias normalmente distribuídas. Assim, para um dado tempo  $t^* > T_{\min}$ , serão geradas  $(N - n_{t^*})$  realizações de uma variável aleatória normalmente distribuída, com média  $\bar{X}_{t^*,p}$  e desvio-padrão  $S_{t^*,p}^C$ . Tais valores

simulados são utilizados no preenchimento das  $(N - n_{t^*})$  matrizes de dados para as quais não se dispõem de observações da  $p$ -ésima variável no tempo  $t^*$ . O mesmo procedimento é então repetido para as demais variáveis monitoradas nas bateladas.

### 3.2 Redução dimensional dos dados na amostra de referência usando o método Statis

O método Statis (ESCOUFIER, 1987; LAVIT *et al.*, 1994) é uma técnica de análise de dados multivariados que permite reduzir a dimensionalidade de arranjos tridimensionais de dados, a partir da utilização de uma medida de similaridade baseada na distância Euclidiana entre matrizes bidimensionais de dados que compõem o arranjo, duas a duas. A Figura 1 ilustra o arranjo tridimensional (e suas matrizes bidimensionais componentes) no contexto da aplicação aqui proposta.

Ao contrário da ACPM, a utilização do Statis em dados oriundos de bateladas preserva a informação sobre a variabilidade do processo no eixo do tempo. Variáveis de processo monitoradas ao longo de uma batelada são variáveis de natureza funcional, cuja evolução é função do tempo. Na ACPM, uma variável de processo observada no tempo  $t$  é tratada como uma variável aleatória cujas realizações encontram-se disponíveis nas bateladas da amostra de referência, no tempo  $t$  em questão. Sendo assim, uma mesma variável observada em  $T$  instantes no tempo é tratada como  $T$  variáveis aleatórias distintas. No método Statis, a natureza funcional das variáveis é considerada na medida em que suas realizações são indexadas no tempo e a informação sobre o comportamento de sua evolução temporal é um dos principais resultados da análise de dados realizada.



**Figura 1: Matrizes de dados das bateladas na amostra de referência.**



O método Statist foi originalmente proposto por Escoufier (1973). Sua utilização no controle da qualidade encontra-se restrita à proposição em Scepi (2002). O método Statist é delineado na seqüência, contextualizado à presente aplicação.

A redução de dimensionalidade promovida pelo Statist cumpre dois objetivos: (i) permitir uma representação sumária de matrizes de dados obtidas a partir de situações de interesse e (ii) oferecer uma representação da trajetória (ou comportamento) de variáveis de interesse no decorrer do tempo. O objetivo (i) é realizado através da análise da interestrutura, ao passo que o objetivo (ii), através da análise da intra-estrutura.

A lógica da análise da interestrutura, de forma resumida, é a seguinte: calculam-se as distâncias entre matrizes de dados, duas a duas, oriundas das diferentes bateladas (isto é, as matrizes  $\mathbf{X}^b$ ); tais resultados são organizados em uma matriz de distâncias a qual é diagonalizada via análise de componentes principais (ACP), sendo retidos os componentes relevantes na sua representação. Finalmente, as matrizes  $\mathbf{X}^b$  são projetadas nos planos fatoriais associados aos componentes retidos na ACP. O estudo da inter-estrutura não permite visualizar o comportamento das variáveis em um dado instante no tempo, mas o seu comportamento geral, considerado sobre todos os instantes de tempo.

A análise da intra-estrutura utiliza elementos da análise da interestrutura, apresentando a seguinte lógica: matrizes de dados  $\mathbf{X}^b$  são combinadas em uma matriz única, denominada matriz de compromisso. Essa matriz é obtida através do cálculo da média ponderada das matrizes  $\mathbf{X}^b$ , utilizando como elementos ponderadores os valores do autovetor principal obtido na ACP da análise da inter-estrutura. A matriz de compromisso é então diagonalizada via ACP, sendo retidos os componentes relevantes na sua representação. Finalmente, as observações de  $\mathbf{X}^b$  em cada instante do tempo são projetadas nos planos fatoriais formados pelos componentes retidos na ACP. O estudo da intraestrutura permite comparar a trajetória das variáveis em  $\mathbf{X}^b$  em cada instante do tempo com a trajetória de compromisso (que corresponde à sua trajetória esperada).

Nas próximas subseções são detalhados os passos operacionais dos estudos de interestrutura e intra-estrutura. Pressupõe-se a existência de  $N$  matrizes de dados  $\mathbf{X}^b$  ( $b=1, \dots, N$ ), já submetidas ao procedimento na seção 3.1, sendo seus dados centralizados e reduzidos para a análise que se segue.

#### Análise da interestrutura

A análise da interestrutura está baseada na definição de

uma medida de similaridade entre matrizes de dados  $\mathbf{X}^b$ , tendo seus passos operacionais esquematizados na Figura 2. Inicialmente, reescrevem-se as matrizes  $\mathbf{X}^b$  como matrizes quadradas  $\mathbf{W}^b$ , de dimensão  $(T_{\max} \times T_{\max})$ , obtidas através da seguinte operação:

$$\mathbf{W}^b = \mathbf{X}^b (\mathbf{X}^b)', \quad (4)$$

onde  $(\mathbf{X}^b)'$  designa a transposta de  $\mathbf{X}^b$ . Cabe ressaltar que na ACP tradicional, as matrizes padronizadas de dados  $\mathbf{X}^b$  seriam submetidas a uma operação matricial do tipo  $(\mathbf{X}^b)' \mathbf{X}^b$ , resultando na matriz de correlações entre variáveis, de dimensão  $(P \times P)$ . A operação na eq. (4), por outro lado, é apropriada para variáveis funcionais cujas realizações são indexadas no tempo, já que gera como resultado uma matriz onde a correlação entre as variáveis é explicitada em cada instante  $t$ .

## Investiga-se o caso especial de bateladas com duração variável, as quais não podem ser diretamente monitoradas através das cartas tradicionais

Calcula-se, na seqüência, uma medida de similaridade entre matrizes  $\mathbf{W}^b$ , duas a duas. A medida de similaridade utilizada é função do produto escalar de Hilbert-Schmidt, dado por:

$$\langle \mathbf{W}^i | \mathbf{W}^j \rangle_{HS} = Tr(\mathbf{D} \mathbf{W}^i \mathbf{D} \mathbf{W}^j), \quad (5)$$

onde  $i$  e  $j$  referem-se a duas bateladas quaisquer da amostra de referência,  $Tr(\ast)$  é o operador de traço matricial, e  $\mathbf{D}$  é uma matriz de pesos de importância para os instantes do tempo, com a seguinte configuração:

$$\mathbf{D} = \begin{pmatrix} D_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_{T_{\max}} \end{pmatrix}, \quad (6)$$

e elemento dado por:

$$D_t = \frac{nt}{N}, \quad (7)$$

tal que  $\sum_{t=1}^{T_{\max}} D_t = 1$ . A lógica da utilização da matriz  $\mathbf{D}$  é atribuir um maior peso de importância a instantes no tempo presentes em um maior número de bateladas na amostra de referência. Caso se deseje atribuir o mesmo peso de importância para todos os instantes do tempo, independentemente da quantidade de informação neles disponíveis sobre as variáveis de processo, então  $\mathbf{D} = \mathbf{I}$ , onde  $\mathbf{I}$  é uma matriz identidade de mesma dimensão de  $\mathbf{D}$ .

Se  $\langle \mathbf{W}^i | \mathbf{W}^j \rangle_{HS} = 0$ , variáveis em  $\mathbf{X}^i$  e  $\mathbf{X}^j$  não apresentam correlação significativa em nenhum instante nas bateladas. Usualmente utilizam-se matrizes  $\mathbf{W}^b$  normalizadas na eq. (5); tais matrizes são obtidas através da operação  $\mathbf{W}^b / \sqrt{\text{Tr}(\mathbf{W}^b)}$ . Nesse caso, a eq. (5) fornece coeficientes de correlação vetorial (coeficientes RV) entre matrizes de dados como resultado. No desenvolvimento que se segue, pressupõe-se a utilização de matrizes  $\mathbf{W}^b$  normalizadas.

Os coeficientes RV devem ser organizados em uma matriz  $\mathbf{S}$ , a qual, após multiplicada por uma matriz  $\Delta$  de pesos de importância para as bateladas, é diagonalizada através de uma ACP. A matriz  $\Delta$ , de dimensão  $(N \times N)$ , é determinada analisando a conformidade do produto resultante de cada batelada com as suas especificações técnicas, tal que bateladas gerando produtos próximos dos valores-alvo das especificações recebam pesos mais altos de importância. Caso tais informações não estejam disponíveis, recomenda-se atribuir o mesmo peso de importância a todas as bateladas, ou seja,  $\Delta = \mathbf{I}/N$ .

$N$  componentes principais (CPs), caracterizados por pares de autovalores ( $\lambda_b$ ) e autovetores ( $\mathbf{u}_b$ ), resultam da ACP realizada na matriz  $\mathbf{S}\Delta$ . A percentagem da variância em  $\mathbf{S}\Delta$  explicada pelo  $j$ -ésimo CP é dada por  $(\lambda_j / \sum_{b=1}^N \lambda_b) \times 100$ . Analisando a capacidade explicativa de cada CP, é possível determinar o número ideal de CPs a serem retidos na

representação da interestrutura das bateladas, seja através de opinião de especialistas conhecedores do processo ou utilizando o conjunto de critérios apresentado em Mingoti (2005; p. 89).

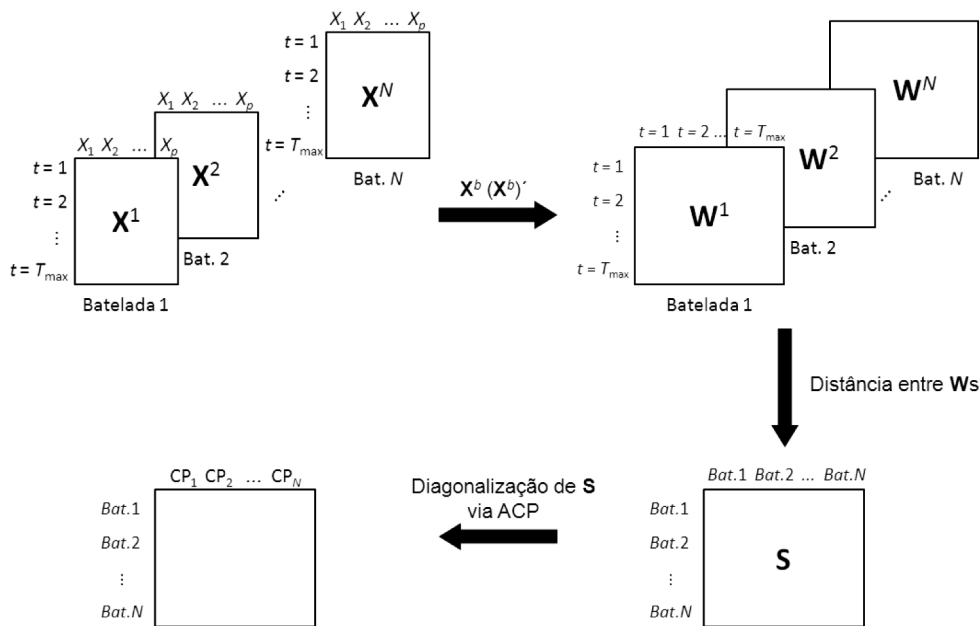
As distâncias entre matrizes  $\mathbf{W}^b$ , calculadas através da eq. (5), podem ser visualizadas de forma agregada ao projetar-se as matrizes  $\mathbf{W}^b$  nos eixos fatoriais retidos na representação da inter-estrutura, através da expressão:

$$a_{b,i} = \sqrt{\lambda_i} u_{b,i} \tag{8}$$

onde  $a_{b,i}$  é a coordenada da batelada  $b$  no  $i$ -ésimo eixo fatorial,  $\lambda_i$  é o autovalor associado ao  $i$ -ésimo CP de  $\mathbf{S}\Delta$  e  $\mathbf{u}_{b,i}$  é o  $b$ -ésimo elemento do autovetor  $\mathbf{u}_i$  associado ao  $i$ -ésimo CP de  $\mathbf{S}\Delta$ . Tipicamente, é possível obter uma boa representação gráfica das bateladas projetando suas respectivas matrizes  $\mathbf{W}^b$  nos eixos fatoriais derivados dos dois primeiros CPs (isto é, aqueles com maiores valores de  $\lambda_b$ ).

**Análise da intra-estrutura**

A análise da intra-estrutura está baseada na determinação de uma matriz de compromisso formada pela soma ponderada das matrizes. Ao contrário da matriz de distâncias  $\mathbf{S}$ , a matriz  $\mathbf{W}^{CO}$  preserva o indexador temporal. Assim, a partir de sua diagonalização, é possível obter projeções gráficas do comportamento das bateladas em cada instante do tempo.



**Figura 2: Passos operacionais da análise da interestrutura.**

Um resumo dos passos operacionais da análise da intra-estrutura vem apresentado na Figura 3.

A matriz  $W^{CO}$  é obtida através da seguinte expressão:

$$W^{CO} = \sum_{b=1}^N \alpha_b W^b. \quad (9)$$

Os pesos  $\alpha_b$  são definidos utilizando elementos do auto-par (*i.e.*, autovalor e autovetor) principal de  $S\Delta$  e da matriz de pesos  $\Delta$ ; isto é:

$$\alpha_b = (1/\sqrt{\lambda_1})\pi_b u_{b,1} \quad (10)$$

onde  $\pi_b$  é o  $b$ -ésimo elemento da diagonal principal da matriz  $\Delta$ ,  $\lambda_1$  é o autovalor principal de  $S\Delta$  e  $u_{b,1}$  é o  $b$ -ésimo elemento do autovetor principal de  $S\Delta$ .

A representação da intra-estrutura das variáveis analisadas pode ser obtida de duas formas. A primeira consiste em uma representação espacial das trajetórias de compromisso das variáveis analisadas, conforme exemplificado na Figura 10. Tal representação é dada pela projeção da matriz  $W^{CO}$  nos planos fatoriais associados aos CPs retidos na ACP aplicada à matriz  $W^{CO}D$ . Para tanto, utiliza-se a seguinte expressão:

$$z_{t,i} = \sqrt{\delta_i} v_{t,i} \quad (11)$$

onde  $z_{t,i}$  é a coordenada de representação do instante  $t$  no  $i$ -ésimo eixo fatorial,  $\delta_i$  é o autovalor associado ao  $i$ -ésimo

CP de  $W^{CO}D$  e  $u_{t,i}$  é o  $t$ -ésimo elemento do autovetor  $v_t$  associado ao  $i$ -ésimo CP de  $W^{CO}D$ . O exemplo na Figura 10 traz uma representação de trajetórias de compromisso no primeiro plano fatorial, obtida operacionalizando a eq. (11) para  $i = 1$  e 2.

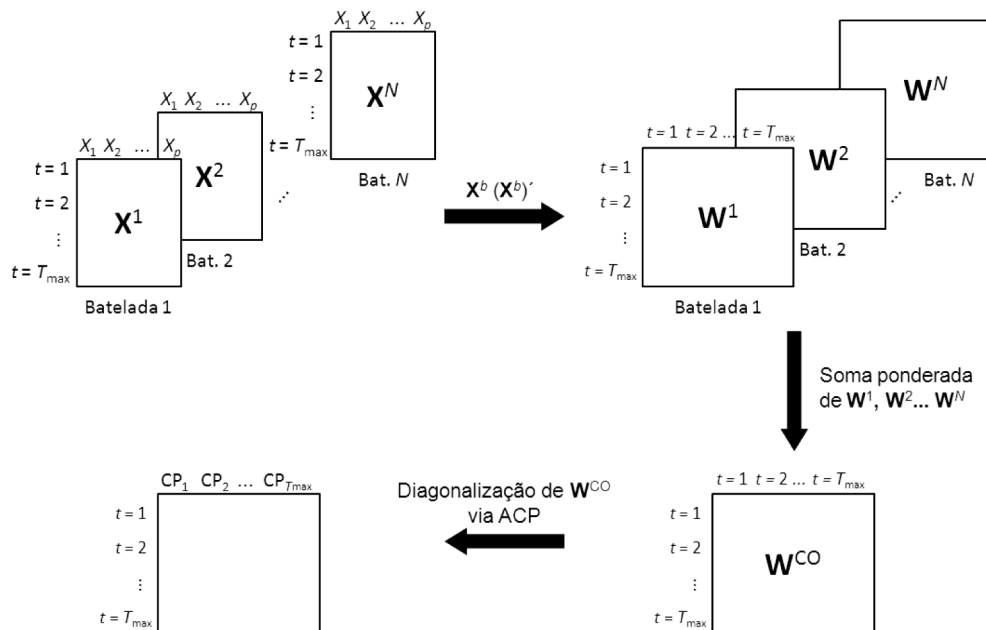
Além disso, é possível obter-se representações detalhadas do comportamento das  $P$  variáveis em cada batelada e em cada momento do tempo, conforme exemplificado na Figura 12. Para tanto, projeta-se cada matriz de dados  $W^b$  nos planos fatoriais associados aos CPs retidos na ACP realizada na matriz  $W^{CO}D$ . Para o  $i$ -ésimo plano fatorial, utiliza-se a seguinte expressão:

$$z_{t,i}^b = (1/\sqrt{\delta_i})w_t^b v_i, \quad (12)$$

onde  $z_{t,i}^b$  é a coordenada da representação da  $b$ -ésima batelada no instante  $t$ , no  $i$ -ésimo eixo fatorial;  $w_t^b$  é a  $t$ -ésima linha da matriz  $W^b$ ;  $\delta_i$  e  $v_i$  são como definidos na eq. (11). O exemplo na Figura 12 traz uma projeção das bateladas nos instantes  $t=1, 3$  e 5, no primeiro plano fatorial; tal representação foi obtida operacionalizando a eq. (12) para  $i = 1, 2$  e  $t = 1, 3$  e 5.

### 3.3 Cartas de controle IS e CO<sub>t</sub>

Nesta etapa, propõe-se a construção de duas cartas de controle (CCs) usando projeções de dados nos planos fa-



**Figura 3: Passos operacionais da análise da intra-estrutura.**



toriais resultantes da aplicação do método Statis; são elas a carta IS, derivada da análise da interestrutura, e a carta CO<sub>p</sub>, derivada da análise da intra-estrutura das bateladas na amostra de referência.

## Neste artigo, propõe-se uma nova estratégia para o monitoramento de processos em bateladas com duração variável.

A carta IS permite o controle *off-line* de bateladas futuras, sendo utilizada após o término da batelada, a partir da matriz de dados  $X^b$  completada seguindo o procedimento na seção 3.1. A carta CO<sub>i</sub> permite o controle *on-line* das bateladas futuras, podendo ser utilizada a partir de matrizes incompletas de dados  $X^b$ ; para concretizar o controle *on-line* das bateladas, um conjunto de  $T$  cartas CO<sub>i</sub> deverá ser utilizado. No caso da carta IS sinalizar uma batelada fora de controle, a identificação do momento no tempo onde o descontrole passou a ocorrer só será possível analisando retroativamente as cartas CO<sub>i</sub> associadas à batelada em questão.

A construção das cartas IS e CO<sub>i</sub> envolve a utilização do formulário apresentado na seção 3.2. O ponto de partida é a obtenção de representações gráficas da interestrutura e intra-estrutura das bateladas na amostra de controle, as quais servirão de base para a construção das CCs. Uma vez disponíveis tais representações, as CCs são formalizadas a partir da determinação de suas regiões de controle. Na exposição que se segue, será considerada a obtenção de CCs a partir de projeções das bateladas da amostra de referência no primeiro plano fatorial (isto é, associadas aos dois primeiros CPs) obtido das aplicações da ACP no escopo do método Statis. A obtenção de CCs para os demais planos fatoriais utilizaria a mesma seqüência de passos.

A base da carta IS são as projeções das bateladas da amostra de referência no primeiro plano fatorial obtido na análise da interestrutura; tais projeções resultam da aplicação da eq. (8) para  $b = 1, \dots, N$  e  $i = 1, 2$ . A base das cartas CO<sub>i</sub> são as projeções do comportamento das  $P$  variáveis monitoradas no processo em cada batelada da amostra de referência, em cada instante no tempo; tais projeções resultam da aplicação da eq. (12) para  $t = 1, \dots, T$ ,  $i = 1, 2$  e  $b = 1, \dots, N$ . Para a obtenção das CCs, em ambos os casos devem-se ajustar regiões de controle nas nuvens de pontos de referência resultantes, seguindo o procedimento descrito a seguir. [Em toda a exposição que se segue, parte-se do pressuposto que a representação da variabilidade total dos dados proporcionada pelos dois primeiros CPs obtidos nas ACPs realizadas nas análises da

inter e intra-estrutura é suficiente para realizar o controle do processo. Caso contrário, as CCs deveriam ser criadas contemplando um número maior de planos fatoriais.]

Considere os gráficos básicos das CCs IS e CO<sub>i</sub> mencionados no parágrafo anterior. Para estabelecer uma região de controle com significância  $(1 - \alpha)$  nesses gráficos, propõe-se uma adaptação do procedimento em Zani et al. (1998), originalmente concebido para a detecção gráfica de valores espúrios em conjuntos de dados bidimensionais. Na apresentação a seguir, consideram-se as projeções no primeiro plano fatorial utilizadas como base para a carta IS; o procedimento para determinação das regiões de controle nas cartas CO<sub>i</sub> é análogo.

Inicialmente, calcula-se a distância de Mahalanobis entre cada ponto no plano fatorial e o ponto médio. Dado que os pontos no plano fatorial são representações gráficas de pares de elementos oriundos de autovetores independentes por definição, a distância de Mahalanobis é obtida por (JOLLIFFE, 2004; p. 33):

$$d(\mathbf{a}_b, \bar{\mathbf{u}}_{b,i}) = \sqrt{(a_{b,1} - \bar{u}_{b,1})^2 + (a_{b,2} - \bar{u}_{b,2})^2}, b = 1, \dots, N \tag{13}$$

onde  $d(\mathbf{a}_b, \bar{\mathbf{u}}_{b,i})$  é a distância de Mahalanobis entre o ponto  $\mathbf{a}_b$ , com coordenadas dadas pela eq. (8) para  $i = 1$  e  $2$ , e o ponto  $\bar{\mathbf{u}}_{b,i}$ , com coordenadas  $(\bar{u}_{b,i}, \bar{u}_{b,2})$  dadas pela média aritmética dos elementos nos autovetores  $\bar{\mathbf{u}}_{b,1}$  e  $\bar{\mathbf{u}}_{b,2}$ , respectivamente.

Ordenam-se, na seqüência, as  $N$  distâncias de Mahalanobis obtidas na eq. (13) em ordem decrescente e retêm-se as 50% menores distâncias; os pontos correspondentes a essas distâncias formarão um *convex hull* (polígono) de abrangência 50% no plano fatorial (isto é, 50% dos pontos no gráfico estarão situados dentro de seus limites).

Ajusta-se então uma curva do tipo *B-spline* aos pontos limítrofes (isto é, que definem o contorno) do *convex hull* de abrangência 50% (para uma breve introdução sobre *B-splines*, recomenda-se Hastie et al., 2001; p. 160). O objetivo do ajuste de uma *B-spline* é unir esses pontos criando um contorno suavizado para o *hull*; uma alternativa mais simples seria ligar os pontos limítrofes do *hull* através de retas, mas o contorno resultante dessa operação tende a ser irregular. No exemplo da Figura 4 tem-se a projeção de 50 pontos em um plano fatorial, 25 dos quais encontram-se situados dentro dos limites do *convex hull* de abrangência 50%; esses pontos correspondem aos 25 menores valores de distância de Mahalanobis do grupo de dados. Do conjunto de 25 pontos, 9 são pontos limítrofes, apresentando-se

ligados por retas. A superfície formada pela união das retas é irregular. Assim, ajustou-se uma curva *B-spline* aos 9 pontos para suavizar o contorno do *convex hull* de abrangência 50%; a curva corresponde ao contorno mais interno na figura.

Finalmente, a região de controle é estabelecida na carta definindo um múltiplo  $l$  da distância entre o centróide (dado pela média aritmética das observações bivariadas no interior do *hull*) e os limites do *convex hull* de abrangência 50%, correspondendo à probabilidade de falso alarme ( $\alpha$  ou erro tipo I) desejada para a CC. Para um valor de  $\alpha = 0,10$ , por exemplo,  $l = 0,86$ . Demais valores de  $l$  para diferentes valores de  $\alpha$  são dados na Tabela 1 [baseados na suposição de dados seguindo uma distribuição normal bivariada e derivados no Apêndice de Zani et al. (1998; p. 267)]. Como a região interna não apresenta uma distância regular do ponto centróide, a região de controle também resultará irregular no formato. A Figura 4 exemplifica a região de controle, com  $\alpha = 0,10$ , obtida para os 50 pontos no plano fatorial (contorno mais externo). Sete pontos são externos a essa região, ao invés dos 5 esperados, considerando a sua abrangência de 90%. Esse comportamento decorre da expansão da região de controle a partir do *hull* mais interno, sendo mais freqüente em amostras contendo um grande número de potenciais valores espúrios.

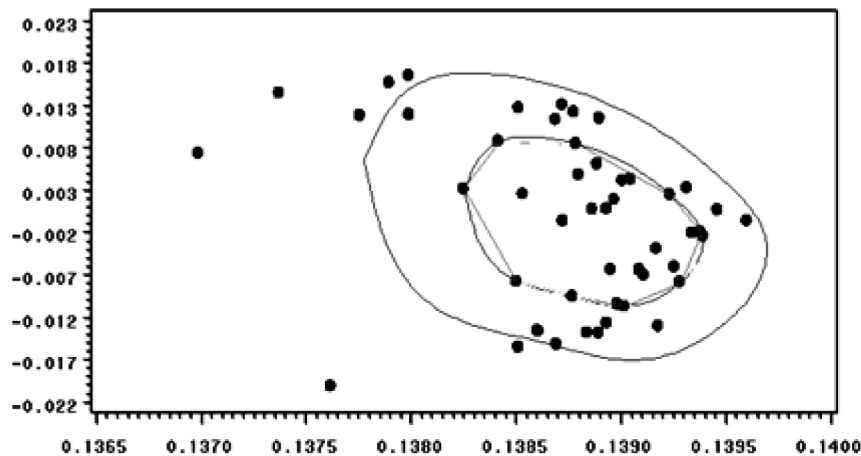
A região de controle da CC poderia ser determinada, alternativamente, definindo o *convex hull* de abrangência  $[(1 - \alpha) \times 100]\%$  correspondente à probabilidade de falso alarme  $\alpha$  desejada para as cartas, identificando seus pontos limítrofes e ajustando uma curva *B-spline* a esses pontos. A abordagem apresentada anteriormente, na qual a

região de controle é derivada expandindo o *convex hull* de abrangência 50%, entretanto, gera regiões de controle com contornos mais estáveis, menos susceptíveis à influência de pontos extremos na nuvem de dados (ZANI et al., 1998). É importante salientar que as CCs resultantes da utilização do procedimento aqui apresentado independem da distribuição de probabilidade que caracteriza os dados na amostra de referência, o que confere um caráter não-paramétrico às cartas.

O controle *off-line* de uma batelada futura ( $b = N + 1$ ) tem início com a projeção da sua matriz de dados correspondente,  $X^{N+1}$ , na carta IS. É importante ressaltar que, nesse caso, a matriz  $X^{N+1}$  estará plenamente disponível. Se a duração da batelada for inferior a  $T_{max}$ , será necessário completar a matriz de dados usando o procedimento na Seção 3.1. Se sua duração for superior a  $T_{max}$ , tem-se a situação de uma batelada mais longa do que as disponíveis na amostra de referência. Nesse caso, deve-se analisar se a nova batelada gerou produtos conformes e, nesse caso, inseri-la na amostra de referência, atualizando o valor de  $T_{max}$  para que corresponda ao da nova batelada.

**Tabela 1: Valores de  $\alpha$  e  $l$  usados na definição da região de controle nas CCs.**

Probabilidade $\alpha$	$l$
0,01	1,68
0,05	1,13
0,10	0,86
0,25	0,43



**Figura 4: Exemplo de *convex hull* e região de controle.**

O controle *on-line* do processo ocorre à medida que a nova batelada progride no tempo. Considere uma nova batelada disponível até um tempo  $t = t^* \leq T_{\max}$ , onde  $t^*$  designa o mais recente instante de tempo no qual as variáveis foram amostradas do processo. Assim, somente uma fração da matriz de dados  $\mathbf{X}^{N+1}$  estará disponível. Como o objetivo é verificar o comportamento das variáveis até o instante  $t^*$ , pode-se supor que seu comportamento em instantes futuros siga o padrão de compromisso de comportamento das bateladas na amostra de referência. Para tanto, devem-se combinar as informações disponíveis na matriz  $\mathbf{X}^{N+1}$  até o tempo  $t = t^*$  com as informações da matriz de compromisso  $\mathbf{W}^{CO}$ .

## Preserva-se toda a informação sobre a variabilidade ao longo do eixo do tempo nos perfis das variáveis de processo

Considere a matriz incompleta  $\tilde{\mathbf{X}}^{N+1}$  contendo observações das  $P$  variáveis de processo até o tempo  $t = t^*$ . Dispõe-se, assim, de dados apenas nas linhas iniciais da matriz, de  $t = 1$  a  $t = t^*$ . Seja  $(\mathbf{x}_t^{N+1})'$  o vetor de linha, de dimensão  $(1 \times P)$ , contendo as observações das  $P$  variáveis de processo na batelada  $b = N + 1$ , no tempo  $t$ . A matriz incompleta  $\tilde{\mathbf{X}}^{N+1}$  é constituída de  $t^*$  vetores  $(\mathbf{x}_t^{N+1})'$ , isto é:

$$\tilde{\mathbf{X}}^{N+1} = \begin{pmatrix} (\mathbf{x}_1^{N+1})' \\ \vdots \\ (\mathbf{x}_{t^*}^{N+1})' \end{pmatrix}. \quad (14)$$

Para avaliar o desempenho da batelada  $b = N + 1$  em um tempo  $t (\leq t^*)$ , utiliza-se a eq. (12), cuja operacionalização demanda o conhecimento do vetor  $\mathbf{w}_t^{NB}$  (e, conseqüentemente, da matriz  $\tilde{\mathbf{W}}^{N+1}$  associada a  $\tilde{\mathbf{X}}^{N+1}$ ). Suponha uma matriz  $\tilde{\mathbf{X}}^{N+1}$  com  $t^*$  linhas de dados disponíveis. A matriz  $\tilde{\mathbf{W}}^{N+1}$  associada a essa matriz terá os seguintes elementos:

$$\tilde{\mathbf{W}}^{N+1} = \begin{matrix} & \begin{matrix} t = 1 & t = 2 & \dots & t = t^* & \dots & t = T_{\max} \end{matrix} \\ \begin{matrix} t = 1 \\ t = 2 \\ \vdots \\ t = t^* \\ \vdots \\ t = T_{\max} \end{matrix} & \left[ \begin{array}{cccc|cc} \frac{1}{Tr(\mathbf{W}^{CO})} \tilde{\mathbf{X}}^{N+1} (\tilde{\mathbf{X}}^{N+1})' & & & & & \mathbf{W}^{CO} \\ \hline & & & & & \\ \mathbf{W}^{CO} & & & & & \mathbf{W}^{CO} \end{array} \right] \end{matrix} \quad (15)$$

ou seja, uma submatriz bloco diagonal de dimensão  $(t^* \times t^*)$  traz as informações disponíveis na matriz  $\tilde{\mathbf{X}}^{N+1}$ ; as demais porções da matriz  $\tilde{\mathbf{W}}^{N+1}$  utilizam porções da matriz de compromisso  $\mathbf{W}^{CO}$ .

A nova batelada, realizada até o instante  $t^*$ , é monitorada usando as cartas  $CO_t$ . Observe que haverá uma carta  $CO_t$  para cada instante de tempo até  $t^*$ , permitindo uma identificação precisa de estados fora-de-controle no processo. A nova batelada também pode ser monitorada usando a carta IS. Entretanto, devido aos dados faltantes em  $\tilde{\mathbf{X}}^{N+1}$ , quanto menor o valor de  $t^*$ , maior a probabilidade de ocorrência de alarmes falsos na carta.

Cabe ressaltar que tanto a carta IS como as  $CO_t$  são de natureza não-paramétrica, ampliando a potencial aplicabilidade do método aqui proposto. Todas as alternativas metodológicas abordadas na Seção 2 estão baseadas em suposições acerca da distribuição de probabilidade dos dados monitorados nas cartas de controle.

## 4. EXEMPLO NUMÉRICO

O método proposto na Seção 3 é exemplificado utilizando um conjunto de dados simulados, disponíveis em Rosa (2005). O conjunto de dados na amostra de referência é composto por 50 bateladas de duração variável ( $T_{\min} = 19$ ;  $T_{\max} = 25$ ); três variáveis de processo ( $X_1, X_2, X_3$ ) são consideradas. Uma visão parcial dos dados é apresentada na Tabela 2; as matrizes  $\mathbf{X}^b$ ,  $b = 1, \dots, 50$ , estão dispostas lado a lado na tabela; as células escurecidas correspondem a valores não disponíveis nas bateladas originais e completados utilizando o procedimento na Seção 3.1.

As trajetórias médias das variáveis  $X_1, X_2$  e  $X_3$  nas bateladas da amostra de referência são apresentadas na Figura 5 (gráficos à esquerda). Os desvios-padrão amostrais calculados a partir das 50 bateladas da amostra de referência, para os tempos  $t = 1$  a 19, encontram-se também grafados na Figura 5 (gráficos à direita).

As séries temporais de desvios-padrão amostrais apresentadas na Figura 5 foram modeladas utilizando o aplicativo NCSS, resultando nos modelos apresentados na Tabela 3. Nos três casos, o melhor ajuste foi obtido utilizando um modelo ARMA (8,7), apesar dos ajustes terem sido distintos, como indicam os coeficientes de determinação ( $R^2$ ) dos modelos na tabela. O modelo acima foi selecionado automaticamente pelo NCSS, utilizando como critério de desempenho na comparação o  $R^2$ , sendo utilizado neste exemplo a título de ilustração. Em aplicações reais, deve-se investigar modelos mais simples e parcimoniosos (por exemplo, modelos de suavização exponencial e de decomposição), em particular no caso de séries de desvios-padrão com número reduzido de observações.

Como mencionado anteriormente, as bateladas foram completadas utilizando o procedimento na seção 3.1. Ilustre-se, na Figura 6, as informações utilizadas para o preenchimento das realizações faltantes da variável  $X_2$  no tempo  $t = 21$ . Os valores de  $\bar{X}_{21,2}$  e  $S_{21,2}$  foram calculados a partir das ( $n_t =$ ) 46 observações de  $X_{21,2}$  disponíveis nas bateladas com duração  $T \geq 21$ . O valor de  $\hat{S}_{21,2}$  foi obtido utilizando o modelo de previsão com parâmetros apresentados na Tabela 3. No cálculo do desvio combinado  $S_{21,2}^C$ , considerouse  $n_t = 46$ . Finalmente, foram simuladas 4 observações de uma variável normalmente distribuída, com parâmetros ( $\mu = 10,3491$ ;  $\sigma = 0,5323$ ), os quais foram utilizados para completar as bateladas  $b = 8, 44, 46$  e  $49$ .

Uma batelada adicional ( $b = 51$ ) também é considerada na análise. A batelada, atipicamente longa, é considerada como fora-de-controle por Rosa (2005); as trajetórias de  $X_1, X_2$  e

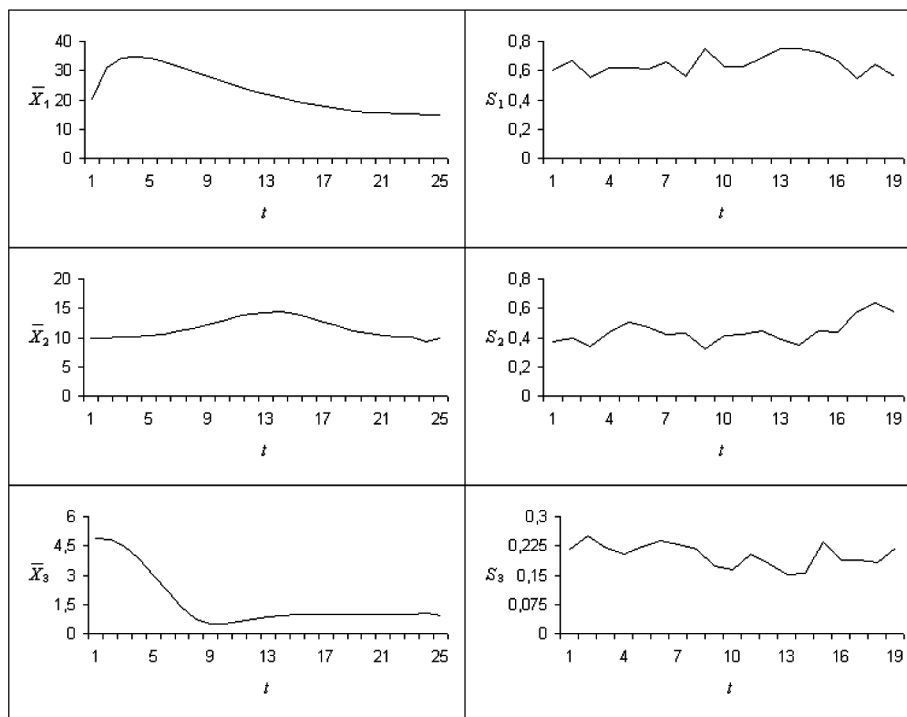
$X_3$  em  $b = 51$ , truncadas no tempo  $t = 25$ , são apresentadas na Figura 7, sobrepostas às trajetórias médias das variáveis (em azul). Como o método aqui proposto foi concebido para capturar comportamentos anormais em bateladas no eixo do tempo, espera-se que o mesmo sinalize a batelada 51 como fora-de-controle.

**O**s resultados ilustram o bom desempenho das novas cartas de controle, em situações de monitoramento on-line e off-line

Analisando-se o conjunto de dados tridimensionais na Tabela 2 a partir do método Statis, obtiveram-se representações gráficas da sua interestrutura e intra-estrutura, as quais permitiram construir cartas IS e  $CO_t$  para os dados do exemplo. O formulário na Seção 3.2 foi programado em

**Tabela 2: Visão parcial dos dados utilizados no exemplo.**

	b = 1			b = 2			...	b = 49			b = 50		
	$X_1$	$X_2$	$X_3$	$X_1$	$X_2$	$X_3$		$X_1$	$X_2$	$X_3$	$X_1$	$X_2$	$X_3$
1	20,1680	9,9725	5,0828	19,8859	9,8154	5,3395		20,0039	9,6434	4,8173	20,3341	10,1021	4,7173
2	30,6463	9,6774	4,7621	31,1813	9,7535	4,7621		30,6021	10,2216	4,7734	31,0140	9,7588	4,6462
3	33,4412	9,7785	4,3783	34,9439	10,0953	4,6369		34,2805	10,4348	4,3873	34,1975	10,0666	4,5224
4	34,4015	9,7956	3,6935	35,9946	10,2895	3,6385		34,5567	10,1914	3,6624	34,7616	9,7200	3,9165
5	33,7507	9,8031	2,5759	34,0381	9,7826	2,9741		33,7057	10,9325	2,7583	33,3380	10,9878	2,6400
6	33,6699	10,3921	2,1503	32,8543	9,9720	1,8127		32,6997	10,5301	2,0756	32,9003	11,2153	2,0004
7	31,9083	11,0228	1,4721	32,3810	10,9252	1,3642		31,0992	10,9244	1,2192	30,5132	10,4505	1,3675
8	29,4106	12,1054	0,5944	29,5731	11,8666	0,6309		29,5213	11,0232	0,4001	30,1130	11,9729	0,4377
9	29,0430	11,6161	0,4388	27,3642	12,0429	0,3107		27,5166	12,7353	0,5452	27,0736	12,1164	0,5380
10	25,8588	12,6815	0,5984	26,0493	13,5929	0,4583		25,9795	13,5765	0,6634	25,3395	12,6901	0,5486
11	24,7040	14,3199	0,5083	24,4405	13,7277	0,7580		24,4111	13,7361	0,7864	23,6610	13,8311	0,5066
12	22,7125	12,8929	0,5827	23,2986	15,2080	0,8062		21,3431	14,0059	0,6073	23,0994	14,6470	0,5876
13	21,5320	14,2421	0,8698	21,3154	14,2328	0,7666	...	21,9317	14,2480	1,1295	21,5096	14,3719	1,0810
14	21,2630	14,8028	1,2564	20,6867	14,2856	0,9800		20,4264	14,0258	0,9438	19,2471	14,1698	0,9284
15	20,2487	14,0446	1,2904	19,2732	13,9680	1,0305		20,0159	13,4707	1,0140	17,4432	13,7792	1,0155
16	18,7849	13,8531	0,9835	18,7198	12,9827	1,1180		18,1941	12,4187	1,0435	18,3357	12,5850	0,7778
17	17,9426	12,7168	1,0768	17,5892	12,3398	0,7996		16,7633	11,8401	0,9212	16,5321	12,1831	0,9070
18	17,1035	12,1050	0,8981	17,4377	11,7935	1,4066		16,2110	10,6074	1,0398	16,0609	11,2122	1,3134
19	16,0876	11,4472	1,1515	15,8024	10,4622	0,9277		16,6555	10,3804	1,1563	16,9305	10,1192	0,9578
20	15,6245	10,8485	0,8353	15,6219	10,4907	1,3175		15,5187	10,2095	0,8696	15,0049	10,3730	1,2501
21	15,9283	10,3298	0,5902	15,1361	10,4360	0,9370		15,1081	10,7820	0,7923	15,7893	10,2611	0,8969
22	15,8000	9,8619	1,0896	15,8107	10,0279	0,8297		14,2212	10,4467	0,4705	13,6878	9,9076	0,6572
23	15,9128	10,0203	0,9572	14,9383	9,8974	0,7786		15,7773	10,4098	0,8105	15,0436	10,1442	0,8548
24	14,5678	10,0122	0,9636	15,4760	9,5468	0,8707		15,7528	10,2906	0,9451	15,7919	9,2679	0,7641
25	14,8610	9,8011	0,7941	13,9039	9,7271	1,3226		15,4616	9,3539	0,9027	15,2958	10,3697	0,7251



**Figura 5: Trajetórias médias das variáveis de processo (gráficos da esquerda) e desvios-padrão das variáveis em cada instante no tempo (gráficos da direita).**

**Tabela 3: Modelos de previsão para as séries temporais de desvios-padrão de  $X_1$ ,  $X_2$  e  $X_3$ .**

		$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$
MODELO		ARMA (8,7)	ARMA (8,7)	ARMA (8,7)
Constante do modelo	AR (1)	0,1268	0,2919	0,0434
	AR (2)	0,2142	0,4132	-0,3588
	AR (3)	-0,1724	-0,2691	-0,0468
	AR (4)	-0,2432	-0,7709	0,4322
	AR (5)	-0,0165	0,0197	-0,4351
	AR (6)	-0,1720	0,3475	-0,1533
	AR (7)	0,0549	-0,0360	-0,3906
	AR (8)	-0,3152	-0,6466	-0,4009
	MA (1)	-0,3720	-0,3239	-0,2380
	MA (2)	0,2530	0,2424	-0,5912
	MA (3)	0,0448	-0,2367	0,1152
	MA (4)	0,0686	0,1290	-0,6153
	MA (5)	-0,1302	-0,1284	0,2862
	MA (6)	-0,0534	-0,3707	0,0464
MA (7)	-0,0460	0,1779	0,1482	
ajuste	$R^2$	37,5065	63,7963	83,9349

um aplicativo de análise estatística, a partir do qual foram obtidos tanto os resultados numéricos quanto os gráficos apresentados na seqüência (as programações estão disponíveis mediante solicitação aos autores).

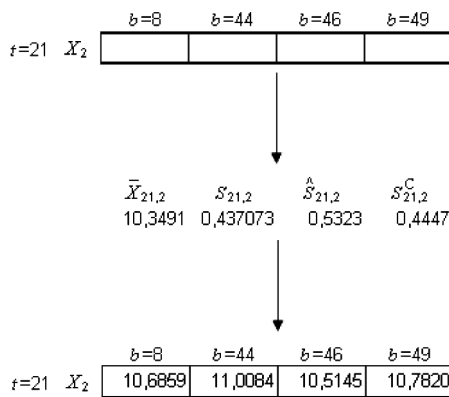
A análise da interestrutura utilizou 50 matrizes  $W^b (b = 1, \dots, 50)$  normalizadas. As distâncias entre matrizes (duas a duas) foram calculadas usando reiteradamente a eq. (5), e considerando uma matriz  $D$  com elementos dados pela eq. (7). As distâncias foram organizadas em uma matriz  $S$ . Utilizou-se uma matriz de pesos  $\Delta = I/50$ . A matriz  $S\Delta$  foi diagonalizada via ACP, tendo sido retidos os dois primeiros auto-pares que, juntos, capturam 96,8% da variabilidade total presente nos dados; os autopares estão apresentados na Tabela 4.

As 50 bateladas da amostra de referência foram projetadas no primeiro plano fatorial, correspondente aos

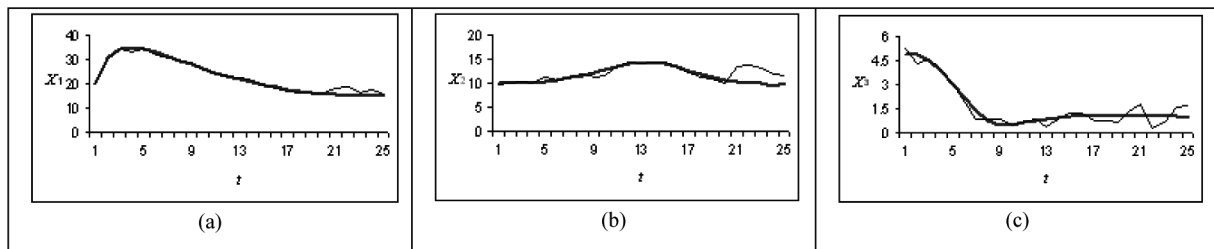
autovetores na Tabela 4, usando a eq. (8). O gráfico resultante vem apresentado na Figura 8. A figura também apresenta a região de controle da carta IS para um valor de  $\alpha = 0,01$ . A determinação da carta combinou programações em aplicativo de análise estatística (determinação do *convex hull* de abrangência 50%, bem como o ajuste das curvas *B-spline*) e cálculos em planilha eletrônica [envolvendo a transformação de coordenadas cartesianas dos pontos limítrofes do *hull* interno em coordenadas polares, as quais foram expandidas na proporção do valor de  $l$  (Tabela 1) e transformadas novamente em coordenadas cartesianas; uma curva *b-spline* foi, então, ajustada aos pontos resultantes].

Na Figura 9, formaliza-se a carta IS (somente os eixos fatoriais e a região de controle para  $\alpha = 0,01$  são apresentados no gráfico). Projetando-se a batelada 51 na carta utilizando a eq. (8), obtêm-se as coordenadas  $(0,1247; -0,0031)$ , correspondentes ao ponto 51 indicado na figura. A carta IS claramente sinaliza  $b=51$  como uma batelada fora-de-controle, conforme esperado.

A Figura 10 apresenta a representação espacial das trajetórias de compromisso das variáveis na amostra de referência, no plano fatorial associado aos primeiros dois autovetores resultantes da ACP aplicada à matriz  $W^{CO}D$ . Os dois primeiros autopares de  $W^{CO}D$  capturam 96,8% da variabilidade total presente na matriz de compromisso, não sendo aqui apresentados para manter a brevidade da exposição. Cada ponto no gráfico da Figura 10 está associado a um tempo  $t$ , conforme indicado na figura, e concentra a informação do comportamento de compromisso das três variáveis de processo nas 50 bateladas da amostra de referência; tais pontos são obtidos operacionalizando a eq. (11) para  $i = 1, 2$  e  $t = 1, \dots, 25$ . Observa-se que os primeiros e últimos instantes das bateladas têm sua variabilidade bem representada pelo segundo eixo fatorial. Já os instantes de tempo intermediários



**Figura 6: Preenchimento dos valores de  $X_{21,2}$  nas bateladas com duração inferior a  $t = 21$ .**

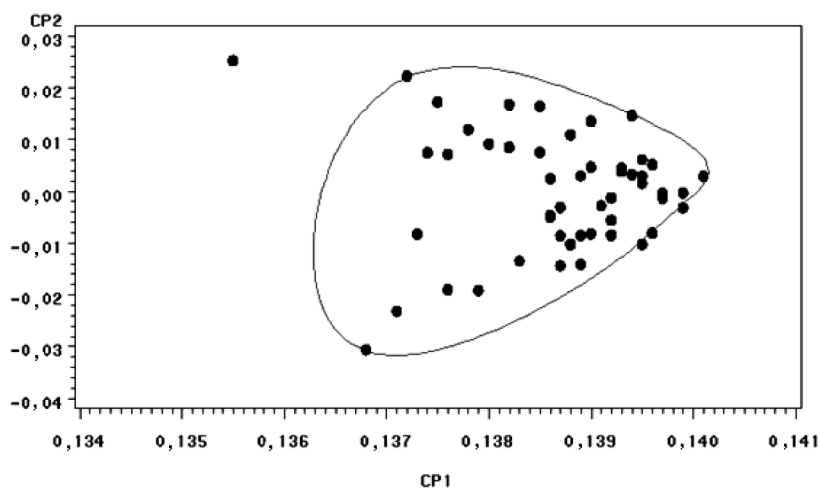


**Figura 7: Trajetórias médias de (a)  $X_1$ , (b)  $X_2$ , (c)  $X_3$  (linha cinza), e comportamento das variáveis em  $b=51$  (linha preta).**



**Tabela 4: Primeiros dois auto-pares obtidos da análise da inter-estrutura.**

1º autovetor ( $\lambda_1 = 0,9614$ )				2º autovetor ( $\lambda_2 = 0,0066$ )			
1	0,1419	26	0,1415	1	-0,0681	26	0,1345
2	0,1422	27	0,1413	2	0,1798	27	0,0309
3	0,1419	28	0,1409	3	-0,0335	28	0,1061
4	0,1425	29	0,1398	4	-0,0021	29	-0,2830
5	0,1415	30	0,1414	5	-0,1760	30	-0,0606
6	0,1425	31	0,1418	6	-0,0162	31	-0,1000
7	0,1415	32	0,1404	7	-0,0366	32	-0,2315
8	0,1382	33	0,1395	8	0,3102	33	-0,3741
9	0,1426	34	0,1400	9	-0,0379	34	-0,1007
10	0,1421	35	0,1414	10	0,0484	35	-0,0559
11	0,1423	36	0,1406	11	-0,0983	36	0,1476
12	0,1421	37	0,1420	12	0,0556	37	-0,0153
13	0,1429	38	0,1406	13	0,0365	38	-0,2342
14	0,1413	39	0,1426	14	0,0931	39	-0,0030
15	0,1414	40	0,1417	15	-0,1039	40	0,0377
16	0,1419	41	0,1422	16	-0,1034	41	-0,1254
17	0,1423	42	0,1404	17	0,0757	42	0,0879
18	0,1418	43	0,1410	18	0,0582	43	-0,1650
19	0,1422	44	0,1401	19	0,0400	44	0,0917
20	0,1417	45	0,1423	20	-0,1714	45	0,0195
21	0,1416	46	0,1413	21	-0,1257	46	0,2030
22	0,1424	47	0,1417	22	0,0638	47	0,1668
23	0,1422	48	0,1403	23	0,0365	48	0,2128
24	0,1408	49	0,1400	24	0,1129	49	0,2735
25	0,1417	50	0,1409	25	-0,1040	50	0,2061



**Figura 8: Projeção das bateladas da amostra de referência no 1º plano fatorial e região de controle da carta IS.**

rios, de  $t = 3$  a  $t = 14$ , têm sua variabilidade bem representada pelo primeiro eixo fatorial.

As cartas  $CO_t$  permitem efetuar o controle *on-line* do processo. Para ilustrar a utilização dessas cartas, simula-se uma nova batelada ( $b = NB$ ) para a qual se dispõe de dados até o quinto instante do tempo. A Figura 11 ilustra a trajetória parcial das variáveis  $X_1$ ,  $X_2$  e  $X_3$  em  $b = NB$  (a linha mais espessa corresponde à trajetória média das variáveis na amostra de referência). Em  $t=1$  e 2, a trajetória da nova batelada cor-

responde à trajetória média das variáveis. A partir de  $t=3$ , as trajetórias das três variáveis passam a apresentar desvios em relação às trajetórias médias.

Para obter as projeções de  $b = NB$  nas cartas  $CO_t$  é preciso completar a matriz de dados  $X^{NB}$  para valores de  $t > 5$ . Optou-se por utilizar os elementos de  $W^{CO}$  para  $t = 6, \dots, 25$  para completar a matriz  $W^{NB}$ , a qual, através da eq. (12), gerou as informações necessárias para construir as cartas  $CO_t$ , para  $t = 1, \dots, 5$ . Tais cartas são apresentadas na

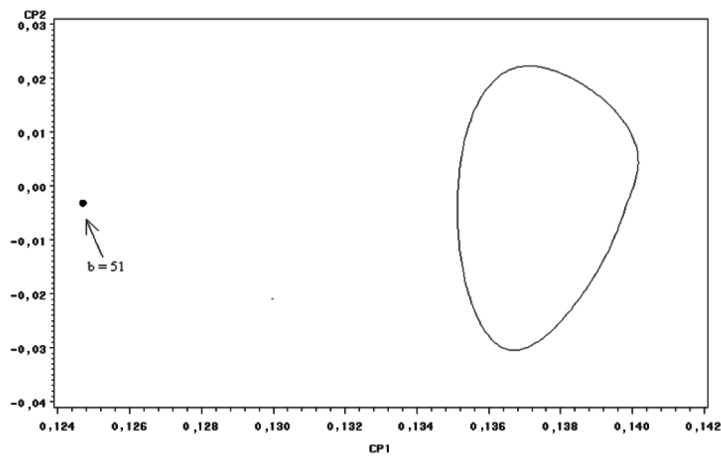


Figura 9: Projeção da batelada  $b=51$  na carta IS.

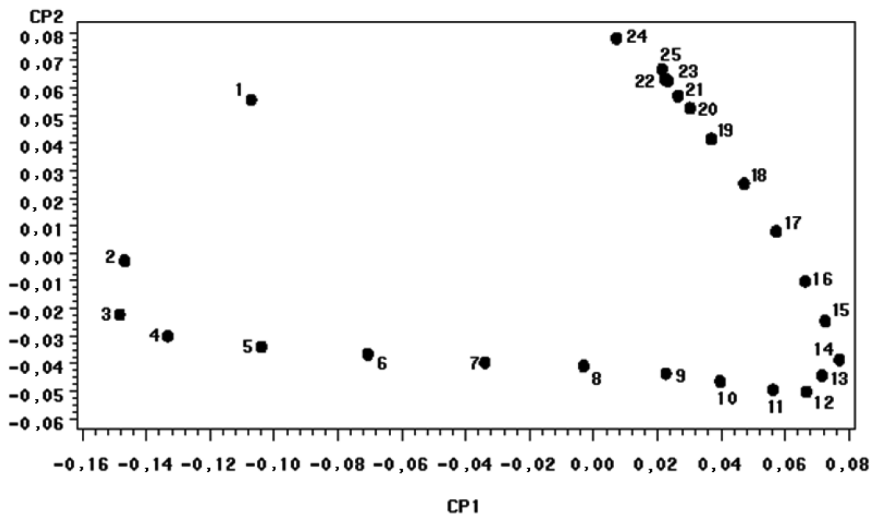
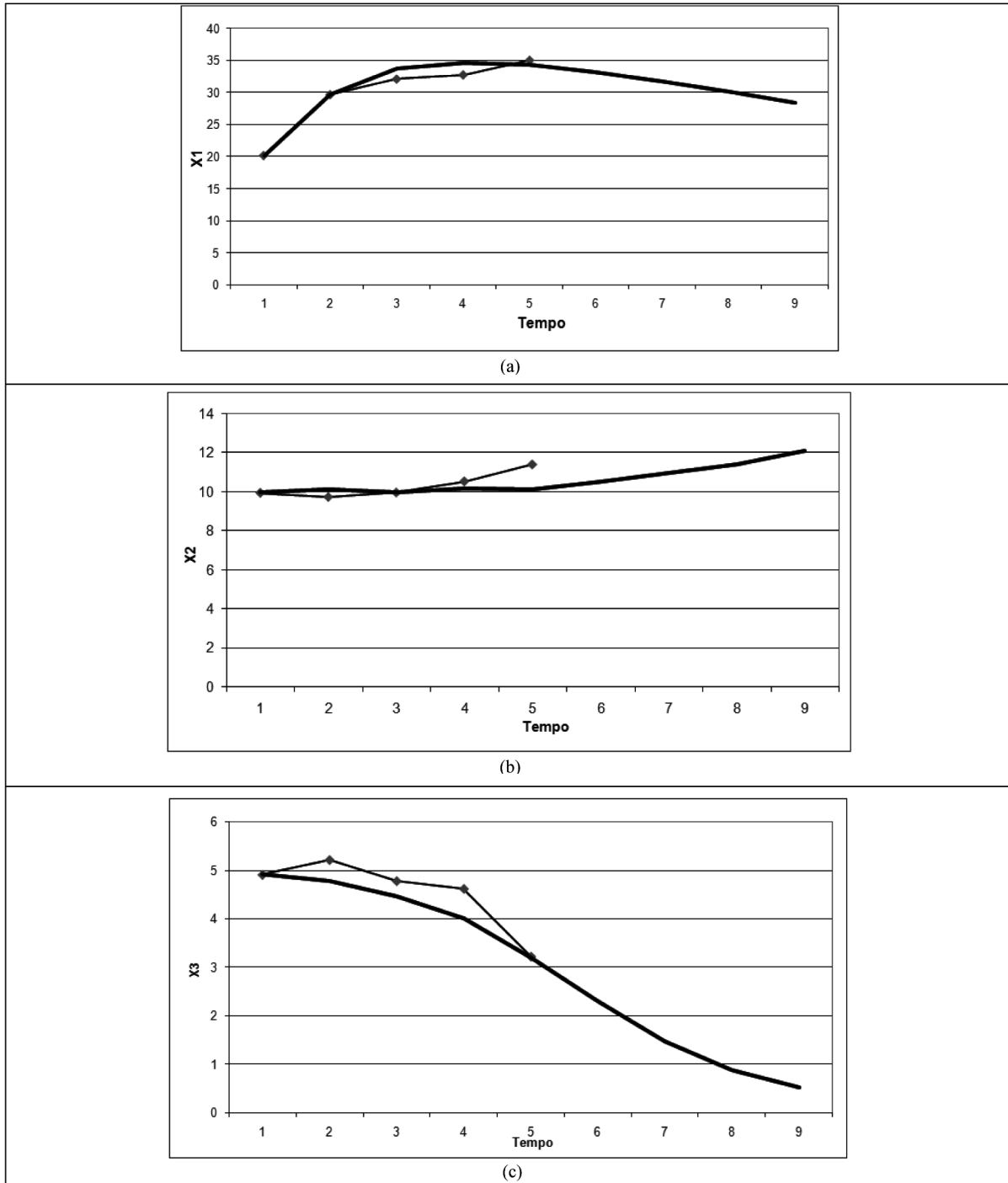


Figura 10: Representação das trajetórias de compromisso das variáveis na amostra de referência.

Figura 12, para  $t = 1, 3$  e 5. A nova batelada corresponde ao ponto NB.

Observa-se que o primeiro sinal ocorre na carta  $CO_2$ ,

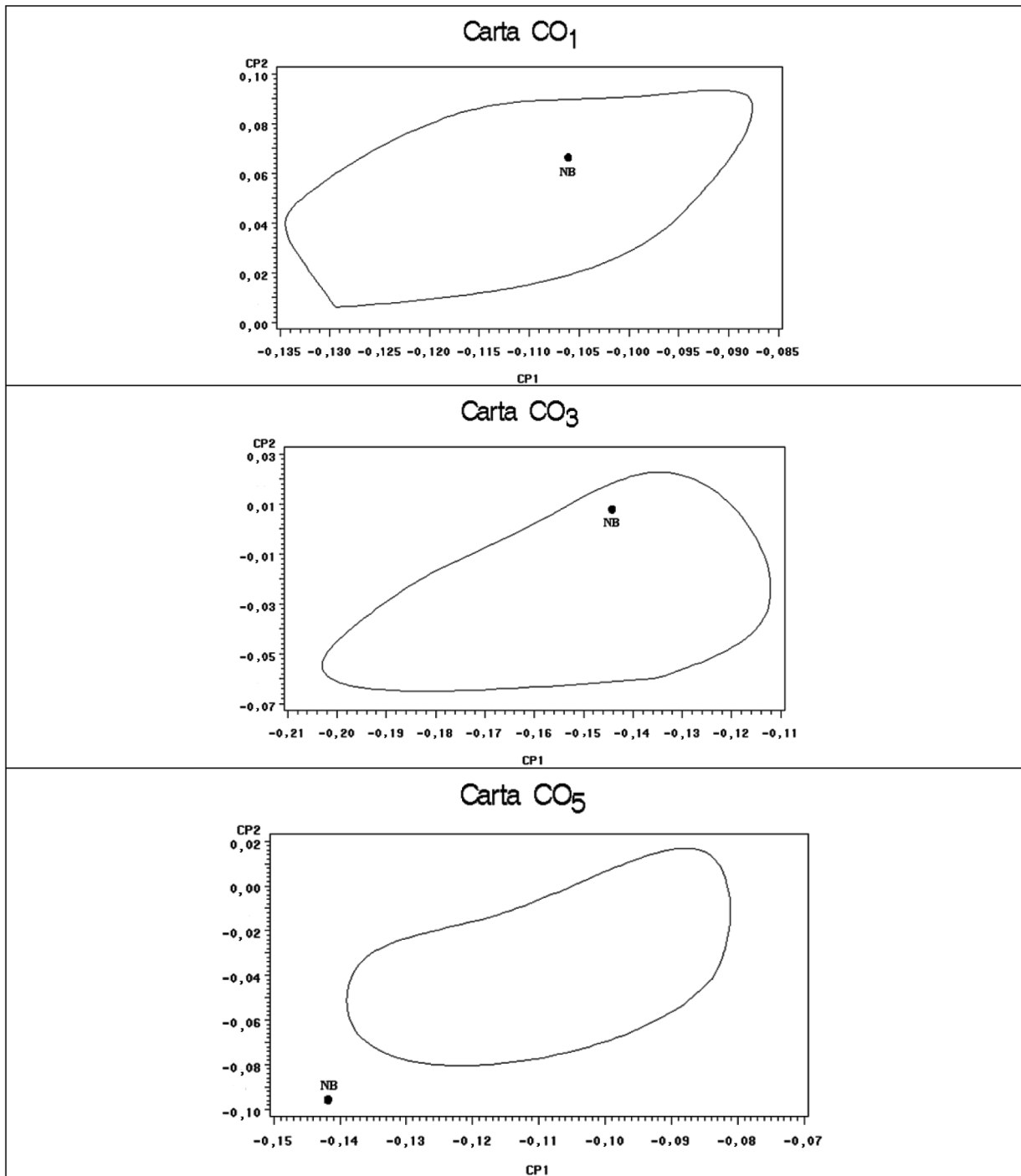
apesar da nova batelada desviar, com maior intensidade, do comportamento esperado a partir de  $t = 3$  (ver Figura 11). Essa defasagem na ocorrência do primeiro sinal justi-



**Figura 11: Trajetórias médias das variáveis de processo e suas 5 medições em  $b=NB$ .**

fica-se pela natureza acumulativa das projeções nas cartas  $CO_t$ . Em  $t = 5$ , por exemplo, o histórico de desempenho das variáveis de processo nos instantes 1 a 5 é considerado.

Conseqüentemente, é preciso reunir evidência acerca do comportamento atípico da nova batelada em vários instantes para que a carta sinalize o descontrolo.



**Figura 12: Cartas  $CO_t$  para  $t = 1, 3$  e  $5$ .**

## 5. CONCLUSÃO

Neste artigo, apresenta-se um novo método para o controle da qualidade de processos em batelada com duração variável. No método proposto, matrizes de dados são completadas de forma a coincidir com a dimensão das matrizes correspondentes às bateladas de mais longa duração, sendo então analisadas utilizando o método Statist (ESCOUFIER, 1987).

O monitoramento do processo é realizado através de duas cartas de controle. Na carta IS, o comportamento global da batelada, com relação às suas variáveis de controle, agregadas em todos os instantes no tempo, é verificado em relação à distribuição de referência; tal carta é apropriada para o controle *off-line* das bateladas. Nas cartas CO<sub>p</sub>, as trajetórias do conjunto de variáveis em cada instante do tempo são

analisadas e desvios significativos com relação às trajetórias de referência são detectados; tal carta é apropriada para o controle *on-line* das bateladas.

O método proposto é aplicado a um conjunto de dados simulados, onde três variáveis de processo são monitoradas em 25 instantes do tempo. Os resultados obtidos ilustram o bom desempenho apresentado pelas novas cartas de controle, tanto em situações de monitoramento *on-line* como *off-line*.

Extensões naturais do trabalho aqui apresentado incluem (i) um estudo comparativo dos resultados obtidos utilizando o método proposto e outros métodos disponíveis na literatura, com especial destaque para a estratégia baseada na deflexão temporal dinâmica proposta por Kassidas *et al.* (1998); e (ii) o desenvolvimento de métodos de diagnóstico dos pontos fora-de-controle, sinalizados nas cartas de controle propostas.

**Artigo recebido em 25/06/2007**

**Aprovado para publicação em 26/03/2008**

## REFERÊNCIAS

- CASTAGLIOLA, P.; ROSA, A. F. P. Monitoring of Batch Processes with Varying Durations Based on the Hausdorff Distance. *International Journal of Reliability, Quality and Safety Engineering*, Vol. 13, p. 213-236, 2006.
- ESCOUFIER, Y. Le traitement des variables vectorielles. *Biometrics*, v. 29, p. 751-760, 1973.
- \_\_\_\_\_. Three-mode data analysis: the STATIS method. In *Methods for multidimensional data analysis*, ECAS, FICHET B.; LAURO N. C. (eds.), p. 259-272, 1987.
- FLORES-CERRILLO, J.; MACGREGOR, J.F. Control of particle size distribution in emulsion semibatch polymerization using mid-course correction policies. *Industrial & Engineering Chemistry Research*, v. 41, p. 1805-1814, 2002.
- HARRIS, T. C.; SEPPALA, C. T.; DESBOROUGH, L. D. A review of performance monitoring and assessment techniques for univariate and multivariate control systems. *Journal of Process Control*, v. 9, p. 1-17, 1999.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning - Data mining, inference, and prediction*. Springer Science: New York, 2001.
- JACKSON, J. E. *A User's Guide to Principal Components*. Wiley: New York, 1991.
- JACKSON, J. E.; MUDHOLKAR, G. S. Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics*, v. 21, p. 341-34, 1979.
- JOLLIFFE, I. T. *Principal component analysis*. 2. ed. Springer Science: New York, 2004.
- KAISTHA, N.; MOORE, C. F.; LEITNAKER, M. G. A statistical process control framework for the characterization of variation in batch profiles. *Technometrics*, v. 46, p. 53-68, 2004.
- KASSIDAS, A.; MACGREGOR, J. F.; TAYLOR, P. A. Synchronization of batch trajectories using dynamic time warping. *AIChE Journal*, v. 44, p. 864-875, 1998.
- KOURTI, T. Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. *Journal of Chemometrics*, v. 17, p. 93-109, 2003.
- KOURTI, T.; MACGREGOR, J. F. Multivariate SPC Methods for Process and Product Monitoring. *Journal of Quality Technology*, v. 28, p. 409-428, 1996.
- KOURTI, T.; LEE, J.; MACGREGOR, J. F. Experiences with industrial applications of projection methods for multivariate statistical process control. *Computers and Chemical Engineering*, v. 20, S745-S750, 1996.
- LAVIT, C.; ESCOUFIER, Y.; SABATIER, R.; TRAISSAC, P. The ACT (STATIS method). *Computational Statistics & Data Analysis*, v. 18, p. 97-119, 1994.
- LOWRY, C. A.; MONTGOMERY, D. C. A review of multivariate control charts. *IIE Transactions*, v. 27, p. 800-810, 1995.
- MACGREGOR, J. F. Using on-line process data to improve quality: challenges for statisticians. *International Statistical Review*, v. 65, p. 309-323, 1997.
- MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada - uma abordagem aplicada*. Editora UFMG: Belo Horizonte, 2005.
- MONTGOMERY, D. C. *Introduction to Statistical Quality Control*. 4. ed. Wiley: New York, 2001.
- NEOGI, D.; SCHLAGS, C. E. Multivariate statistical analysis of an emulsion batch process.

*Industrial & Engineering Chemistry Research*, v. 37, p. 3971-3979, 1998.

NOMIKOS, P. *Statistical process control of batch processes*. Ph.D. Thesis, McMaster University, Hamilton (Canada), 1995.

NOMIKOS, P.; MACGREGOR, J. F. Multivariate SPC charts for monitoring batch processes. *Technometrics*, v. 37, p. 41-59, 1995.

ROSA, A. F. P. *Maîtrise statistique de procédés par lots à temps variable*. Thèse de Doctorat de l'Université de Nantes, Nantes (França), 2005.

SCEPI, G. Parametric and non parametric multivariate quality control charts. In *Multivariate Total Quality Control*, Physica-Verlag, LAURO C. et al. (eds.), p. 163-189, 2002.

WIERDA, S. J. Multivariate statistical process control - Recent results and directions for future research. *Statistica Neerlandica*, v. 48, p. 147-168, 1994.

ZANI, S.; RIANI, M.; CORBELLINI, A. Robust bivariate boxplots and multiple outlier detection. *Computational Statistics & Data Analysis*, v. 28, p. 257-270, 1998.

## AGRADECIMENTOS

A pesquisa do Prof. Fogliatto é financiada pelo CNPq, através do processo PQ 301433/2005-4. Os autores gostariam de agradecer aos *referees* anônimos pela atenção dispensada ao artigo e pelas valiosas sugestões feitas visando a sua melhoria. Os autores também agradecem a Rafael Alcoba pelo apoio no desenvolvimento do exemplo numérico

## SOBRE OS AUTORES

### Flávio S. Fogliatto

Departamento de Engenharia de Produção e Transportes – UFRGS  
End.: Av. Osvaldo Aranha, 99 – 5º andar – Porto Alegre – RS – 90040-020  
Fone: (51) 3308-4294 Fax: (51) 3308-4007  
E-mail: ffogliatto@producao.ufrgs.br

### Ndèye Niang

Conservatoire National des Arts et Metiers – CNAM  
End.: 292 Rue Saint Martin, F-75141 – Paris – Cédex 03 – França  
Fone: +33 1 40 27 20 00 Fax: +33 1 40 27 27 46  
E-mail: niang@cnam.fr