LI Simpósio Brasileiro de Pesquisa Operacional
Limeira–SP, 02 a 06 de setembro de 2019

SELEÇÃO DE VARIÁVEIS APLICADA AO CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS EM BATELADAS

Autor: Fernanda Araujo Pimentel Peres
fernanda.peres@ufrgs.br

Orientador: Flávio Sanson Fogliatto
ffogliatto@producao.ufrgs.br

Programa de Pós-Graduação em Engenharia de Produção
Universidade Federal do Rio Grande do Sul – UFRGS
Nível: Tese
Defesa em: Setembro/2018

RESUMO

O desenvolvimento de métodos de monitoramento adequados para o controle estatístico multivariado de processos em bateladas é um tópico emergente. Nesse contexto, a presente tese propôs o método PVS-MPCA, o qual integra a seleção de variáveis Pareto à Análise de Componentes Principais Multidirecionais. O objetivo é aprimorar a detecção de falhas em bancos de dados de elevada dimensionalidade, oriundos de processos industriais em bateladas de duração variável. O novo método foi aplicado em um banco de dados obtido de um processo industrial alimentício. Os resultados demonstraram que o método proposto reduziu drasticamente a taxa de alarmes falsos, melhorando o desempenho de monitoramento e superando as limitações encontradas pelo método MPCA tradicional na detecção de falhas em bancos em bateladas de elevada dimensionalidade. Conclui-se que o método PVS-MPCA agrega inovação à área de monitoramento de processos em bateladas e contribui para a geração de produtos de elevado padrão de qualidade.

PALAVRAS CHAVE. Seleção de variáveis, Detecção de falhas, Processo em bateladas.

Eixo temático: Pesquisa Operacional na Indústria

ABSTRACT

The development of monitoring methods to handle with multivariate statistical process control in batch processes is an emerging topic. To tackle the problem the PVS-MPCA method was proposed, which integrates Pareto Variable Selection and Multiway Principal Component Analysis. The aim is fault detection improvement in high-dimensional datasets from industrial batch processes with variable duration. The new method was applied to a real dataset from an industrial food process. Results showed that the proposed method dramatically reduced the false alarm rate, improving the monitoring performance and overcoming limitations found when classic MPCA method is applied to fault detection in batch processes with high dimensional datasets. We conclude that PVS-MPCA method adds innovation to the area of batch process monitoring and contributes to the generation of high-quality products.

KEYWORDS. Variable selection, Fault detection, Batch processes.

Thematic axis: Industry Operational Research

1. Apresentação

A presente tese se propôs a analisar as lacunas relacionadas a 3 áreas relevantes para o controle de processos industriais: (*i*) os processos em bateladas, (*ii*) o controle estatístico de processos multivariados (MSPC) de elevada dimensionalidade e (*iii*) a seleção de variáveis integrada ao MSPC.

O foco dado aos processos industriais em bateladas se justifica pelo fato de serem amplamente utilizados por indústrias químicas, farmacêuticas e de alimentos, mas pouco explorados na literatura devido a suas características, como: comportamento não linear, estado não estacionário e duração de processo variável [Impe & Gins, 2015].

Na área do MSPC, o foco foi na melhoria da detecção de falhas pelas cartas de controle multivariadas (CCM) T^2 de Hotelling e Q baseadas no método de projeção da Análise de Componentes Principais Multidirecionais (MPCA) [Woodall & Montgomery, 2014]. O tipo de monitoramento alvo deste estudo foi o *off-line* [Kourti, 2003] para processos industriais em bateladas de elevada dimensionalidade.

No que tange a seleção de variáveis, somente métodos com fins de classificação foram avaliados [Anzanello et al., 2012]. Sua implementação ocorreu mediante o uso da abordagem *wrapper*, na qual o subconjunto de variáveis relevantes é determinado através de um procedimento iterativo [Ghosh et al., 2014; Mehmod et al., 2012], que envolveu um *ranking* de importância de variáveis e uma técnica de classificação de bateladas.

O objetivo geral da tese foi a proposição de um método que integra a seleção de variáveis ao MSPC a fim de aprimorar a detecção de falhas em bancos de dados de elevada dimensionalidade, oriundos de processos industriais em bateladas de duração variável.

Para o desenvolvimento da pesquisa, foram explorados tanto os métodos qualitativos quanto os quantitativos. Primeiramente, uma revisão sistemática de literatura foi conduzida visando o melhor entendimento sobre o tema e as lacunas existentes na área de estudo. Esse conhecimento foi base para realização das etapas quantitativas da pesquisa, que se desdobraram em: (*i*) definição do melhor método para pré-processamento de bancos de dados em bateladas de duração variável e (*ii*) proposição de um novo método para monitoramento de processos em bateladas de elevada dimensionalidade que possa mitigar as lacunas dos métodos anteriormente proposto na literatura.

Nas próximas seções serão apresentadas as publicações e premiações desdobradas desta tese.

2. Artigos Publicados em Periódicos Internacionais

Uma revisão sistemática de literatura para a identificação, classificação e descrição dos métodos publicados, os quais integram a seleção de variáveis ao MSPC, foi resultado desta tese. A revisão, intitulada '*Variable selection methods in multivariate statistical process control: a systematic literature review*' (Métodos de seleção de variáveis no controle estatístico de processo multivariado: uma revisão sistemática de literatura), foi publicada no periódico internacional Qualis A2 em Engenharias III *Computers & Industrial Engineering* (IF 2018: 3,518) [Peres & Fogliatto, 2018] e proporcionou o entendimento do problema fomentando e justificando a escolha do tema da tese.

O objetivo do artigo foi responder às seguintes questões de pesquisa: (*i*) 'Como ocorre a evolução dos métodos que utilizam a seleção de variáveis para aprimorar o uso de MSPC em bancos de dados de elevada dimensionalidade?', (*ii*) 'Quantos são e qual a abrangência dos métodos integrados disponíveis na literatura?' e (*iii*) 'Quais são as oportunidades de pesquisa desdobradas dos achados da revisão?'. Esse estudo foi embasado em uma revisão teórica sobre as limitações dos métodos de MSPC na análise de bancos de dados industriais de elevada

dimensionalidade, os métodos de Seleção de Variáveis existentes, bem como as etapas de monitoramento no Controle Estatístico de Processo.

A evolução do estado da arte nesse tópico foi demonstrada, sendo cada um dos 30 métodos identificados na literatura brevemente descritos e discutidos. O artigo inovou ao propor uma classificação (clusterização) dos métodos propostos de acordo com a abordagem de seleção de variáveis implementada, e ao categorizar os 10 clusters obtidos de acordo com seus objetivos e a etapa de monitoramento de processo para a qual os métodos foram desenvolvidos. Assim, clusters de trabalhos foram propostos, auxiliando na identificação de lacunas e desdobramento de oportunidades de pesquisa sobre o tema. Dentre as contribuições do artigo podem ser citadas o auxílio à pesquisadores no desenvolvimento desse tópico mediante a definição das lacunas existentes sinalizando oportunidades de pesquisas futuras, bem como o apoio a profissionais responsáveis por departamentos de qualidade na identificação de métodos que possam auxiliar na solução de problemas industriais reais.

No que tange o objetivo geral desta Tese, conforme descrito na seção 1, o mesmo foi alcançado mediante a proposição de um novo método de detecção de falhas, baseado nas CCM T^2 de Hotteling e Q, capaz de lidar com bancos de dados em bateladas de elevada dimensionalidade. O método PVS-MPCA propôs que, prévio a construção das CCM baseadas em PCA, fosse conduzida a seleção das variáveis capazes de maximizar a acurácia de classificação das bateladas em conformes e não conformes. O artigo intitulado '*Fault detection in batch processes through variable selection integrated to multiway principal component analysis*' (Detecção de falhas em processos em bateladas através da integração da seleção de variáveis à análise de componentes principais multidirecionais) apresenta o método PVS-MPCA e foi publicado no periódico internacional Qualis A2 em Engenharias III *Journal of Process Control* (IF 2018: 3,316) [Peres et al., 2019].

Essa pesquisa quantitativa foi desenvolvida com o objetivo de responder às seguintes questões: (i) 'Quais variáveis melhor classificam as bateladas nas categorias conforme e não conforme?' e (ii) 'Qual o desempenho das CCM quando o método proposto é comparado às cartas de controle tradicionais T^2 de Hotteling e Q baseadas em PCA?'. O método proposto foi aplicado ao banco de dados do processo industrial de conchagem do chocolate ao leite, o qual foi pré-tratado, alinhado e sincronizado de acordo com o método de Alinhamento Temporal Dinâmico (DTW ou *Dynamic Time Warping*) escolhido conforme descrito na seção 3 [Peres et al., 2018a]. A seleção de variáveis seguiu uma abordagem *wrapper* envolvendo um índice de importância de variáveis, baseado nos parâmetros de saída da Análise Discriminante em Mínimos Quadrados Parciais (PLS-DA ou *Partial Least Squares – Discriminant Analysis*), e a técnica de classificação por k-vizinhos mais próximos (kNN ou *k-nearest neighbors*). Dessa forma, a cada iteração a variável menos importante para a classificação das bateladas foi descartada e a acurácia de classificação do subgrupo remanescente avaliada. Após a implementação da análise de Pareto Ótimo, o subgrupo com 0,17% de variáveis retidas foi considerado o que maximizava a acurácia em 100% com o menor número de variáveis (5 variáveis desdobradas retidas). A escolha das variáveis selecionadas pelo método PVS foi corroborada pela análise técnica do processo de conchagem do chocolate ao leite. A melhora do desempenho da detecção de falhas obtida pela aplicação do método proposto, quando comparado ao método de CCM baseadas em MPCA, demonstrou que as limitações do método tradicional foram superadas quando um banco de dados com um número de variáveis muito superior ao de observações foi analisado. Tendo em vista que os métodos publicados para detecção e isolamento de falhas em bancos de dados em bateladas de elevada dimensionalidade têm sido propostos para substituir o uso das cartas de controle baseadas em PCA [Chu et al., 2004; Yan et al., 2017; Zarzo & Ferrer, 2004], o método PVS-MPCA surge como uma proposta inovadora para estender o uso das tradicionais CCM T^2 de Hotteling e Q na detecção de falhas, mitigando suas limitações.

3. Artigo Completo Apresentado e Publicado em Simpósio Nacional

Os métodos tradicionais de MSPC em bateladas, como a MPCA, se baseiam na suposição de que todas as bateladas têm a mesma duração e estão sincronizadas (ou seja, eventos similares ocorrem nos mesmos instantes). Porém, na maioria das situações industriais reais, o tempo total de duração das bateladas e/ou a duração de vários estágios dentro das bateladas não é o mesmo [González-Martínez et al., 2011; Kassidas et al., 1998; Ramaker et al., 2003]. Um exemplo de processo com duração variável é o processo de conchagem do chocolate, o qual foi analisado como estudo de caso desta tese.

O artigo intitulado "*Estratégias para sincronização e alinhamento de bancos de dados de processos multivariados em bateladas utilizando algoritmos de alinhamento temporal dinâmico*" foi apresentado e publicado nos anais do *L Simpósio Brasileiro de Pesquisa Operacional - SOBRAPO* [Peres et al., 2018a]. O evento ocorreu entre os dias 06 a 09 de agosto de 2018 na cidade do Rio de Janeiro (Brasil) e os custos de deslocamento e hospedagem da autora foram financiados pela Taxa de Bancada para bolsista de Doutorado fornecida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Essa pesquisa quantitativa foi desenvolvida com o objetivo de responder às seguintes questões: (*i*) 'Qual dos 3 métodos de alinhamento e sincronização selecionados, baseados no algoritmo de DTW, é mais adequado para tratar um banco de dados do processo de conchagem do chocolate ao leite, o qual apresenta bateladas com duração variável?', e (*ii*) 'O uso de métodos de sincronização afeta o poder de classificação das bateladas em conformes e não-conformes?'. Os resultados foram discutidos sob três pontos de vista: (*i*) da tecnologia de fabricação do chocolate, (*ii*) do poder de classificação das bateladas em conformes e não conformes mediante aplicação da técnica de classificação kNN, e (*iii*) do método mais adequado para tratamento de dados a serem analisados por métodos de MSPC, visando aprimorar a construção do modelo de monitoramento na Fase I. Os três métodos analisados se mostraram hábeis para promover o alinhamento e a sincronização, sendo que o que apresentou os melhores valores de métricas de desempenho de classificação (acurácia, sensibilidade e especificidade) foi indicado como o mais adequado ao banco de dados analisado. O método escolhido manteve o poder de classificação das bateladas após as mesmas serem ajustadas para um mesmo tempo de duração, além de ter requerido o menor número de vizinhos (k=3) para obtenção desses resultados. A maioria dos métodos de monitoramento de processos tem seu foco no desempenho da Fase II do monitoramento [Peres & Fogliatto, 2018; Woodall & Montgomery, 2014]. Nesse contexto, até onde se tem conhecimento, a análise do impacto dos diferentes métodos de alinhamento e sincronização na determinação do conjunto de referência de bateladas conformes não foi previamente explorada, fazendo com que esse artigo contribua de maneira inovadora para o desenvolvimento de modelos sob-controle mais adequados para a Fase I do CEP.

4. Resumo de Artigo Apresentado e Publicado em Conferência Internacional

O resumo da pesquisa quantitativa descrita na seção 3 foi apresentado e publicado nos anais do *20th International Conference on Food Processing and Technology* [Peres et al., 2018b]. O evento ocorreu entre os dias 28 e 29 de junho de 2018 na cidade de Londres (Grã-Bretanha) e os custos de deslocamento e hospedagem da autora foram financiados pela Taxa de Bancada para bolsista de Doutorado fornecida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

5. Artigo em Revisão em Periódico Internacional

Uma versão ampliada e reescrita em inglês do artigo publicado nos anais do SBPO (descrito na seção 3) se encontra atualmente em revisão (*second round*) no periódico

internacional Qualis B1 em Engenharias III *Journal of Food Science and Technology* (IF 2018: 1,849).

6. Premiações Obtidas

A autora foi agraciada com o prêmio "*Best Paper Presentation*", conforme Anexo 1, no *20th International Conference on Food Processing and Technology* [Peres et al., 2018b], ocorrido entre os dias 28 e 29 de junho de 2018 na cidade de Londres (Grã-Bretanha).

7. Desdobramentos da Pesquisa

Atualmente a autora é bolsista de Pós Doutorado Júnior pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Seu projeto de pesquisa intitulado "*Seleção de variáveis integradas ao isolamento e diagnóstico de falhas em processos em bateladas de elevado dimensional*" tem antecedentes nos trabalhos já desenvolvidos pela proponente em sua tese de Doutorado e visa a consolidação de seus conhecimentos nessa área. A vigência da bolsa é de 12 meses (01/11/2018 até 31/10/2019).

No que tange as entregas desse projeto, estão a publicação de artigos em periódicos científicos internacionais e a apresentação de outros artigos, com resultados parciais, em congressos nacionais e/ou internacionais. O principal produto desse projeto de pesquisa é o desenvolvimento de um método inovador para o controle de processos em bateladas informatizados que gere real impacto na garantia da qualidade fabril.

Um segundo produto é o conhecimento adquirido e disseminado através desse projeto contribuindo para a formação de recursos humanos e evolução desta área de conhecimento. Nesse contexto, a autora ministrou parcialmente (no 1° semestre de 2019) a disciplina de Engenharia de Qualidade do curso de Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul, totalizando uma carga horária de 20 horas. Nessa disciplina, a autora tem possibilidade de disseminar os conhecimentos adquiridos ao longo de seu Doutorado e de reforçar sua maturidade profissional e independência em direção a uma carreira de longo prazo como pesquisadora-docente de alto nível. Somado a isso, a candidatura para concorrer ao Prêmio de Tese do 51° SOBRAPO é também uma importante forma de disseminação dos conhecimentos adquiridos para a comunidade científica nacional e internacional, buscando atingimento de parcerias científicas de alto nível, que tragam benefícios concretos à sociedade.

Pesquisas futuras se baseiam no aprofundamento dos métodos de monitoramento de processos em bateladas de elevada dimensionalidade e na aplicação desses no ambiente industrial. Além disso, o uso do MSPC como ferramenta para auxiliar as organizações na tomada de decisões baseadas em dados, explorando o grande volume de informações disponíveis em indústrias automatizadas e evitando empirismos ou necessidade de *experts* com amplo conhecimento sobre os processos, é um tema emergente e que contribuirá para a construção de organizações mais analíticas e atingimento de resultados de excelência.

Referências

Anzanello, M. J., Albin, S. L. e Chaovalitwongse, W. A. (2012). Multicriteria variable selection for classification of production batches. *European Journal of Operational Research*, 218 (1):97–105.

Chu, Y-H, Qin, S.J., e Han, C. (2004). Fault Detection and Operation Mode Identification Based on Pattern Classification with Variable Selection. *Industrial & Engineering Chemistry Research*, 43:1701-1710.

Ghosh, K., Ramteke, M. e Srinivasan, R. (2014). Optimal variable selection for effective statistical process monitoring. *Computers and Chemical Engineering*, 60:260–276.

González-Martínez, J. M., Ferrer, A. e Westerhuis, J. A. (2011). Real-time synchronization of batch trajectories for on-line multivariate statistical process control using Dynamic Time Warping. *Chemometrics and Intelligent Laboratory Systems*, 105:195–206.

Impe, J. V. e Gins, G. (2015) An extensive reference dataset for fault detection and identification in batch processes. *Chemometrics and Intelligent Laboratory Systems*, 148:20-31.

Kassidas, A., MacGregor, J. F. e Taylor, P. A. (1998). Synchronization of batch trajectories using dynamic time warping. *AIChE Journal*, 44(4):864–875.

Kourti, T. (2003). Abnormal situation detection, three-way data and projection methods; robust data archiving and modeling for industrial applications. *Annual Reviews in Control*, 27: 131–139.

Mehmood, T., Liland, K. H., Snipen, L., e Sæbø, S. (2012). A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems*, 118:62–69.

Peres, F. A. P. e Fogliatto, F. S. (2018). Variable selection methods in multivariate statistical process control: A systematic literature review. *Computers and Industrial Engineering*, 115:603–619.

Peres, F. A. P., Peres, T. N. P., Fogliatto, F. S. e Anzanello, M. J. (2018a). Estratégias para Sincronização e Alinhamento de Bancos de Dados de Processos Multivariados em Bateladas Utilizando Algoritmos de Alinhamento Temporal Dinâmico. In *Anais do L SBPO*, n° 84634, Rio de Janeiro. SOBRAPO. Disponível em: https://proceedings.science/sbpo/papers

Peres, F. A. P., Peres, T. N. P., Fogliatto, F. S. e Anzanello, M. J. (2018b). Strategies for Synchronizing Chocolate Conching Data Using Dynamic Time Warping. *International Scholarly and Scientific Research & Innovation* 12(6):117. Disponível em: https://waset.org/pdf/books/?id=84178&pageNumber=117

Peres, F. A. P., Peres, T. N. P., Fogliatto, F. S. e Anzanello, M. J. (2019). Fault detection in batch processes through variable selection integrated to multiway principal component analysis. *Journal of Process Control*, 80:223-234.

Ramaker, H. J., Van Sprang, E. N. M., Westerhuis, J. A. e Smilde, A. K. (2003). Dynamic time warping of spectroscopic batch data. *Analytica Chimica Acta*, 498:133–153.

Woodall, W. e Montgomery, D. (2014). Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology*, 46(1):78–94.

Yan, Z. e Kuang, T. H. e Yao, Y. (2017). Multivariate fault isolation of batch processes via variable selection in partial least squares discriminant analysis. *ISA Transactions*, 70: 389–399.

Zarzo, M., e Ferrer, A. (2004). Batch process diagnosis: PLS with variable selection versus block-wise PCR. *Chemometrics and Intelligent Laboratory Systems*, 73(1):15–27.

ANEXO 1



INTERNATIONAL SCIENTIFIC RESEARCH AND EXPERIMENTAL DEVELOPMENT

CERTIFICATE OF BEST PRESENTATION AWARD

ICFPT 2018: 20th International Conference on
Food Processing and Technology

hereby certifies that the below mentioned paper
has been selected as the best presentation

Strategies for Synchronizing Chocolate Conching Data Using Dynamic Time Warping

Fernanda A. P. Peres, Thiago N. Peres, Flavio S. Fogliatto, Michel J. Anzanello

London, United Kingdom

Authorized Signature

June 28-29, 2018