



# DETECÇÃO DE FALHAS EM PROCESSOS QUÍMICOS CONTÍNUOS: UMA PROPOSTA VIA REDUÇÃO DE DIMENSIONALIDADE BASEADA EM ANÁLISE POR COMPONENTES PRINCIPAIS.

G. F. J. BARCELOS<sup>1</sup>, L. S. ALMADA<sup>2</sup>, D. S. REIS<sup>1</sup>, G. S. COSTA<sup>1</sup> e G. M. ALMEIDA<sup>1</sup>

<sup>1</sup>Universidade Federal de Minas Gerais, Departamento de Engenharia Química

<sup>2</sup>Imperial College London, Department of Chemical Engineering

E-mail para contato: gustavoalmeida0001@gmail.com

RESUMO – Propõe-se, neste trabalho, uma metodologia para a construção de sistemas de detecção de falhas, para processos químicos contínuos, baseado em Análise por Componentes Principais (PCA). Essa é a técnica estatística multivariada mais usual em aplicações de monitoramento em processos químicos. A contribuição deste trabalho é o uso individual de limites de controle para cada componente principal em particular Utilizou-se o *benchmark* Tennessee como estudo de caso. Investigou-se todo o seu conjunto de vinte e uma falhas. É possível classificá-las em quatro grupos, conforme o grau de dificuldade de detecção. Os resultados mais significativos desse trabalho foram para duas das três falhas do grupo de detecção intermediária, com ganhos entre 3,8% e 64,8%, e para todas as seis falhas do grupo de detecção difícil, com ganhos entre 41,9% e 99,8%, em comparação com a Literatura. No geral, o resultado é superior para 16 de 21 falhas, em comparação com a estatística  $T^2$ ; e equivalente para 3 de 21 falhas, e superior para 13 de 21 falhas, em comparação com a estatística  $Q$ .

## 1. INTRODUÇÃO

O monitoramento de processos químicos é importante para garantir operações mais seguras, eficientes e limpas. Com a crescente complexidade dos processos, a demanda por sistemas automáticos e confiáveis de monitoramento é ainda maior. Mesmo para operadores e engenheiros experientes, a percepção de sinais de baixa magnitude, que são característicos de falhas em estágio inicial, é difícil. Atualmente, com o *big data*, e a possibilidade de uso de gigantescas massas de dados à disposição nas indústrias químicas, o uso do computador é ainda mais crucial para se ter êxito (Severson *et al.*, 2016; Yin *et al.*, 2014).

Entre as duas tarefas de monitoramento, a saber, a detecção, relativa ao reconhecimento de uma falha, e o diagnóstico, à descoberta de sua causa, a primeira ainda é um desafio do ponto de vista prático. O objetivo final é realização de ambas as etapas, de modo satisfatório, com uma taxa relativamente baixa de alarmes falsos (Chiang *et al.*, 2001).

Em indústrias químicas em geral, em todo o mundo, é usual realizar o monitoramento de variáveis de modo independente, como por exemplo, através de cartas de controle de Shewhart. Porém, dada a significativa correlação espacial entre variáveis de processo, o ideal



é considerá-las em conjunto. Ter um número significativo de variáveis, altamente correlacionadas, é algo inerente aos processos químicos. Nesse ponto, é interessante empregar técnicas de redução de dimensionalidade, por meio de um processo de extração de características, sendo essas características combinações das variáveis originais de processo, como por exemplo, vazão, pressão e temperatura. Essa redução facilita o processamento computacional e a visualização e entendimento do problema (Montgomery, 2016).

A técnica estatística multivariada, de redução de dimensionalidade, com o maior número de aplicações em monitoramento de processos químicos, é aquela denominada Análise por Componentes Principais (PCA; *Principal Component Analysis*). Ela busca explicar a variância total de um sistema multivariado, com  $p$  dimensões (variáveis originais), a partir de  $k$  dimensões (componentes principais), com  $k \ll p$ . As componentes, ortogonais entre si, são os eixos de um novo sistema de coordenadas, dado pela rotação daquele original. Elas são combinações (características) lineares das variáveis originais (Sharma, 1996).

## 1.1. OBJETIVOS

Propõe-se uma metodologia para a construção de sistemas de detecção de falhas, para processos químicos contínuos, baseado em Análise por Componentes Principais (PCA), uma técnica estatística multivariada. PCA é a técnica mais usual em aplicações de monitoramento em processos químicos contínuos (Severson *et al.*, 2016). A contribuição deste trabalho é o uso individual de limites de controle para cada componente principal em particular. Investigam-se falhas abruptas de baixa magnitude e falhas incipientes, que são de maior interesse prático e de detecção mais difícil num estágio inicial. Nessa direção, os objetivos específicos são, o conhecimento do processo e de seu bancos de dados, a extração de características, e a construção e análise de desempenho do sistema de detecção de falhas. Comparam-se os resultados com a referência clássica de Russell *et al.* (2000). Geraram-se os resultados com o *Toolbox Dimensionality Reduction* para Matlab (van der Maaten, 2018).

## 2. ESTUDO DE CASO

O estudo de caso é referente ao *benchmark* Tennessee (Downs e Vogel, 1993), baseado em um processo químico real. Ele é o mais utilizado pela comunidade de Engenharia Química em todo o mundo; por exemplo, para a construção e comparação de estratégias de sistemas de monitoramento. O processo contém cinco equipamentos principais: um reator, um condensador, um compressor, uma coluna *stripper*, e um separador líquido/vapor. O banco de dados é composto por 52 variáveis, entre medições de campo e análises de laboratório. O intervalo de amostragem é de 3 minutos, com um total de 960 observações, ou 48h de operação. Investigou-se o conjunto completo de 21 falhas, entre abruptas e incipientes, sempre com início em  $t$  (instante de tempo ou observação) igual a 160, ou 8h de operação.

## 3. METODOLOGIA

A metodologia é composta por seis etapas, conforme a descrição a seguir. (1) *Análise Exploratória dos Dados*: A partir dos conjuntos de dados, construíram-se gráficos temporais, histogramas e gráficos de dispersão, com o objetivo de se familiarizar com as variáveis e com os comportamentos do processo durante a operação normal e após a ocorrência das falhas.

Calcularam-se também estatísticas descritivas para as variáveis, como média e desvio-padrão, antes e após as falhas. (2) *Extração de características*: Após alguns testes, variou-se a redução de dimensionalidade ( $d$ ) de 5 até 45, com um incremento de 10. Com isso, obtiveram-se cinco modelos PCA, com 5, 15, 25, 35 e 45 componentes principais. (3) *Obtenção dos modelos PCA*: O conjunto de dados de entrada, característico de operações normais, é composto por 52 variáveis e 500 observações. Para evitar a influência indesejada de variáveis com maiores variâncias, em termos absolutos, sobre a definição das direções das componentes principais, padronizaram-se as variáveis (média zero e variância 1). (4) *Determinação dos limites de controle*: Propõe-se o cálculo do limite de controle superior, de modo individual, para cada componente principal, ao invés de um valor único, conforme realizado usualmente para as estatísticas  $T^2$  e  $Q$ . Para se comparar os resultados com Russell *et al.* (2000), adotou-se uma taxa de alarmes falsos de 1%. Para isso, o Limite Inferior de Controle ( $LIC$ ) é dado pelo percentil 0,5, e o Limite Superior de Controle ( $LSC$ ), pelo percentil 99,5. De modo a manter a independência necessária para o cálculo dos limites de controle, empregou-se um segundo conjunto de dados, também característico de operações normais, composto por 52 variáveis e 960 observações. (5) *Testes de detecção de falhas*: Testaram-se cada um dos cinco modelos PCA com as 21 falhas do *benchmark* Tennessee (Downs e Vogel, 1993). (6) *Validação dos sistemas de detecção de falhas*: Compararam-se os resultados obtidos com aqueles em Russell *et al.* (2000). Empregou-se a métrica, Taxa de Detecções Perdidas ( $MDR$ ; *Missed Detection Rate*; Equação 1), em que  $X$  é o número de detecções perdidas entre  $t = 160$  e  $t = 960$ , que é referente ao período sob falha.

$$MDR(\%) = X/800 \quad (1)$$

#### 4. RESULTADOS E DISCUSSÃO

A Tabela 1 contém os resultados obtidos com os cinco sistemas de detecção de falhas baseados em PCA, cada um com um número diferente de componentes principais ( $d = 5:10:45$ ). Apresentam-se as falhas em quatro grupos, segundo o grau de dificuldade de detecção. Baseou-se essa classificação nos valores de  $MDR$  (Taxa de Detecções Perdidas) em Russell *et al.* (2000). Grupo 1 (detecção fácil): Falhas 1, 2, 4, 6, 7, 8, 12, 13 e 14; Grupo 2 (detecção intermediária): Falhas 11, 17 e 18; Grupo 3 (detecção difícil): Falhas 5, 10, 16, 19, 20 e 21; e Grupo 4 (detecção muito difícil): Falhas 3, 9 e 15. O foco desse trabalho são os Grupos 2 a 4, cuja detecção ainda é um desafio. Inicialmente, observa-se, em geral, o decaimento do  $MDR$  com o aumento do número de componentes principais ( $d$ ).

Compararam-se os resultados com aqueles em Russell *et al.* (2000), para  $T^2$  e  $Q$ . Essas estatísticas são usuais como métricas em aplicações de detecção de falhas em processos químicos. Medem, respectivamente, a variação dentro e fora do modelo PCA. Em relação à  $T^2$ , calculou-se a diferença entre o valor de  $MDR$  segundo  $T^2$ , em Russell *et al.* (2000), e o menor valor de  $MDR$  entre os cinco modelos PCA deste trabalho (coluna Diferença Absoluta (%)). Calculou-se também o ganho relativo da proposta do presente trabalho em relação a Russell *et al.* (2000), ao se dividir essa diferença pelo valor de  $MDR$  segundo  $T^2$  (coluna Ganho Relativo (%)). Quanto mais negativo o ganho, melhor o resultado da proposta deste trabalho em relação à Russell *et al.* (2000). Procedeu-se, de modo análogo, para a estatística  $Q$ . Os melhores resultados encontram-se em negrito (colunas, Diferença Absoluta (%) e Ganho Relativo (%)). Descrevem-se os resultados por grupo de falhas.



Tabela 1 – Resultados dos sistemas de detecção de falhas baseados em PCA desse trabalho, em comparação com Russell *et al.* (2000).

[MDR (*Mixed Detection Rate*) = Taxa de Detecções Perdidas, em %;  $d$  (número de componentes principais) = Dimensão do sistema de detecção de falhas baseado em PCA]

Classe	Falha	MDR do presente trabalho ( $d = 5:10:45$ )					Russell <i>et al.</i> (2000)		Em comparação com $T^2$		Em comparação com $Q$	
		5	15	25	35	45	$T^2$	$Q$	Diferença Absoluta (%)	Ganho Relativo (%)	Diferença Absoluta (%)	Ganho Relativo (%)
1	1	0,319	0,010	0,010	0,010	0,010	0,008	0,003	0,2	25,0	0,7	233,3
	2	0,063	0,013	0,013	0,013	0,013	0,020	0,014	-0,8	-37,5	-0,2	-10,7
	4	0,935	0,335	0,136	0,136	0,003	0,956	0,038	-95,4	-99,7	-3,6	-93,4
	6	0,020	0,000	0,000	0,000	0,000	0,011	0,000	-1,1	-100,0	0,0	- <sup>†</sup>
	7	0,609	0,000	0,000	0,000	0,000	0,085	0,000	-8,5	-100,0	0,0	- <sup>†</sup>
	8	0,289	0,204	0,204	0,204	0,204	0,034	0,024	17,0	499,3	18,0	749,0
	12	0,211	0,211	0,211	0,211	0,193	0,029	0,025	16,4	563,8	16,8	670,0
	13	0,126	0,126	0,126	0,126	0,126	0,060	0,045	6,6	110,4	8,1	180,6
	14	0,970	0,196	0,149	0,149	0,088	0,158	0,000	-7,1	-44,6	8,8	- <sup>†</sup>
2	11	0,920	0,569	0,453	0,453	0,343	0,794	0,356	-45,2	-56,9	-1,4	-3,8
	17	0,864	0,595	0,545	0,418	0,091	0,259	0,108	-16,8	-64,8	-1,7	-15,5
	18	0,119	0,119	0,119	0,119	0,116	0,113	0,101	0,3	2,9	1,5	15,1
3	5	0,719	0,719	0,719	0,719	0,001	0,775	0,746	-77,4	-99,8	-74,5	-99,8
	10	0,611	0,611	0,611	0,611	0,179	0,666	0,659	-48,7	-73,2	-48,0	-72,9
	16	0,613	0,534	0,219	0,219	0,198	0,834	0,755	-63,7	-76,3	-55,8	-73,8
	19	0,930	0,930	0,849	0,181	0,181	0,996	0,873	-81,5	-81,8	-69,2	-79,2
	20	0,763	0,488	0,239	0,239	0,239	0,701	0,550	-46,2	-65,9	-31,1	-56,6
	21	0,331	0,331	0,331	0,331	0,331	0,736	0,570	-40,5	-55,0	-23,9	-41,9
4	3	0,841	0,841	0,841	0,841	0,841	0,998	0,991	-15,7	-15,7	-15,0	-15,1
	9	0,928	0,928	0,928	0,928	0,928	0,994	0,981	-6,7	-6,7	-5,4	-5,5
	15	0,930	0,930	0,930	0,930	0,930	0,988	0,973	-5,8	-5,9	-4,3	-4,4

<sup>†</sup> Divisão por zero.

- *Grupo 1 (Falhas fáceis: 1, 2, 4, 6, 7, 8, 12, 13 e 14):* Em relação à  $T^2$ , o resultado é inferior para as falhas 1, 8, 12 e 13, sendo a maior diferença para a falha 8, igual a 17,0%; e superior para as falhas 2, 4, 6, 7 e 14, sendo a maior diferença para a falha 4, igual a -95,4%. Em relação à  $Q$ , o resultado é inferior para as falhas 1, 8, 12, 13 e 14, sendo a maior diferença para a falha 8, igual a 18%; é equivalente para as falhas 6 e 7, ou seja, não há diferenças; e superior para as falhas 2 e 4, sendo a maior diferença para a falha 4, igual a -3,6%. Mas observa-se que o maior MDR (resultado menos satisfatório) entre essas falhas, nesse trabalho, é um valor razoável, igual a 20,4% (falha 8).

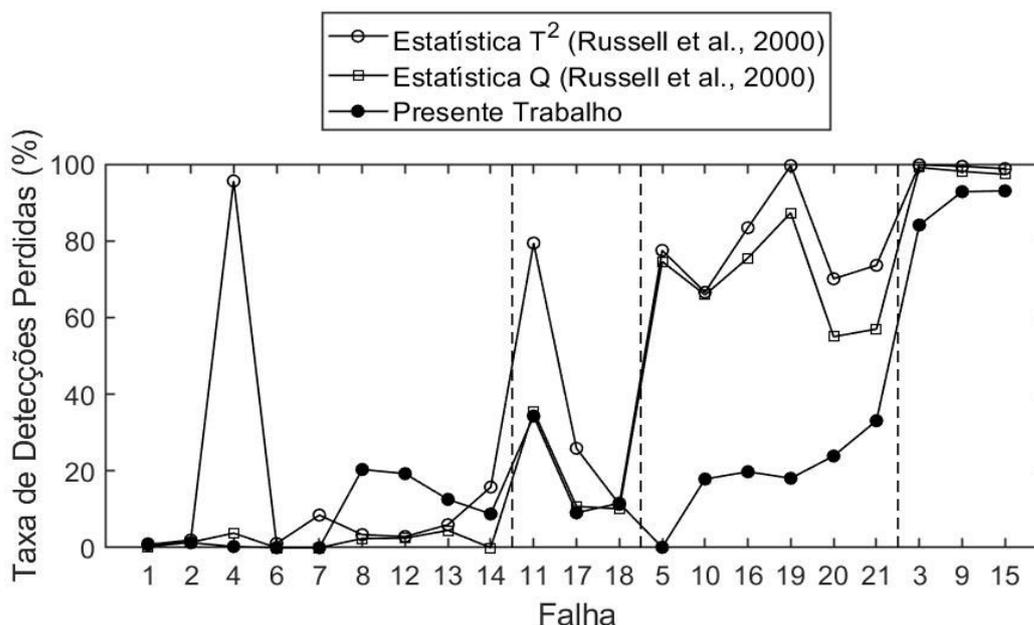
- *Grupo 2 (Falhas Intermediárias: 11, 17 e 18):* Em relação à  $T^2$ , o resultado para a falha 18 é equivalente, com *MDRs* significativamente próximos, iguais a 11,3% (Russell *et al.*, 2000) e 11,6% (este trabalho); e há um ganho significativo quanto à detecção das falhas 11 e 17, igual a -56,9% e -64,8%, respectivamente. Em relação à  $Q$ , o resultado é inferior para a falha 18, com um ganho de 15,1%; equivalente para a falha 11, com *MDRs* significativamente próximos, iguais a 34,3% (Russell *et al.*, 2000) e 35,6% (este trabalho); e superior para a falha 17, com um ganho significativo de -15,5%.
- *Grupo 3 (Falhas difíceis: 5, 10, 16, 19, 20 e 21):* O resultado obtido com esse trabalho é superior para todo o conjunto de seis falhas, em comparação com os *MDRs* de  $T^2$  e de  $Q$  em Russell *et al.* (2000). Em relação à  $T^2$ , os ganhos são iguais a -99,8%, -73,2%, -76,3%, -81,8%, -65,9% e -55,0%, respectivamente; e, em relação à  $Q$ , -99,8%, -72,9%, -73,8%, -79,2%, -56,6% e -41,9%, respectivamente. Os menores ganhos são para a falha 21, iguais a -55,0% (em relação à  $T^2$ ) e -41,9% (em relação à  $Q$ ), que são consideravelmente altos.
- *Grupo 4 (Falhas muito difíceis: 3, 9 e 15):* Os valores de *MDR* com esse trabalho ainda são consideravelmente altos, iguais a 84,1%, 92,8% e 93,0%, respectivamente; porém, houve uma redução geral em relação àqueles em Russell *et al.* (2000), principalmente para a falha 3 (*MDR* igual a 84,1%), com ganhos iguais a -15,7% (em relação à  $T^2$ , cujo *MDR* é igual a 99,8%) e -15,1% (em relação à  $Q$ , cujo *MDR* é igual a 99,1%).

A Figura 1 mostra os *MDRs* (Taxas de Detecções Perdidas, em %) relativos às estatísticas  $T^2$  e  $Q$ , em Russell *et al.* (2000), e aqueles obtidos neste trabalho (Tabela 1). As linhas verticais tracejadas delimitam os grupos de falhas, quanto ao grau de dificuldade de detecção, conforme descrição anterior. Pode-se observar que os resultados mais significativos desse trabalho, em comparação com Russell *et al.* (2000), são para as falhas 11 e 17 do Grupo 2 (detecção intermediária), e para todas as seis falhas (5, 10, 16, 19, 20 e 21) do Grupo 3 (detecção difícil). No geral, o resultado é superior para 16 de 21 falhas, em comparação com a estatística  $T^2$ ; e equivalente para 3 das 21 falhas, e superior para 13 de 21 falhas, em comparação com a estatística  $Q$ .

## 5. CONCLUSÃO

Propõe-se, neste trabalho, uma metodologia para a construção de sistemas de detecção de falhas, para processos químicos contínuos, baseado em Análise por Componentes Principais (PCA). Essa é a técnica estatística multivariada mais usual em aplicações de monitoramento em processos químicos. A contribuição deste trabalho é o uso individual de limites de controle para cada componente principal em particular. Utilizou-se como estudo de caso, por completo, o *benchmark* Tennessee. Investigou-se todo o conjunto de vinte e uma falhas, entre abruptas e incipientes. É possível classificá-las em quatro grupos, conforme o grau de dificuldade de detecção. Os resultados mais significativos desse trabalho, em comparação com Russell *et al.* (2000), foram para duas das três falhas do grupo de detecção intermediária, com ganhos entre 3,8% e 64,8%, e para todas as seis falhas do grupo de detecção difícil, com ganhos entre 41,9% e 99,8%. Como trabalhos futuros, sugerem-se considerar, a correlação (dependência) temporal entre as observações, e a não linearidade do processo, por exemplo, com o uso de KPCA (*Kernel PCA*).

Figura 1 – Resultado gráfico relativo à Tabela 1.



## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- CHIANG, L. H.; RUSSELL, E. L.; BRAATZ, R. D. *Fault detection and diagnosis in industrial systems*, Editora Springer, 2001.
- DOWNS, J. J.; VOGEL, E. F. A plant-wide industrial process problem control. *Comput. Chem. Eng.*, v. 17(3), p. 245-255, 1993.
- MONTGOMERY, D. C. *Introdução ao controle estatístico da qualidade*, 7a ed., Editora LTC, 2016.
- RUSSELL, E. L.; CHIANG, L. H.; BRAATZ, R. D. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometr. Intell. Lab. Syst.*, v. 51, p. 81-93, 2000.
- SEVERSON, K.; CHAIWATANODOM, P.; BRAATZ, R. D. Perspectives on process monitoring of industrial systems. *Annu. Rev. Control*, v. 42, p. 190-200, 2016.
- SHARMA, S. *Applied multivariate techniques*, Editora Wiley, 1996.
- VAN DER MAATEN, L. J. P. *Matlab Toolbox for Dimensionality Reduction*, Acesso em 05-03-2018, Disponível em <http://lvdmaaten.github.io/drtoolbox/>.
- YIN, S., DING, S. X., XIE, X., LUO, H. A review on basic data-driven approaches for industrial process monitoring. *IEEE T. Ind. Electron.*, v. 61(11), p. 6418-6428, 2014.