



INTEGRAÇÃO HUMANO-MÁQUINA PARA O MONITORAMENTO DE PROCESSOS INDUSTRIAIS BASEADO EM DADOS

Afrânio José de Melo Junior

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Química, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Química.

Orientadores: José Carlos Costa da Silva Pinto
Maurício Melo Câmara

Rio de Janeiro

Maio de 2023

INTEGRAÇÃO HUMANO-MÁQUINA PARA O MONITORAMENTO DE
PROCESSOS INDUSTRIAIS BASEADO EM DADOS

Afrânio José de Melo Junior

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA QUÍMICA.

Orientadores: José Carlos Costa da Silva Pinto
Maurício Melo Câmara

Aprovada por: Prof. José Carlos Costa da Silva Pinto
Dr. Maurício Melo Câmara
Prof. Maurício Bezerra de Souza Júnior
Prof. Cristiano Hora de Oliveira Fontes
Dr. Fábio César Diehl

RIO DE JANEIRO, RJ – BRASIL

MAIO DE 2023

Melo Junior, Afrânio José de
Integração humano-máquina para o monitoramento de
processos industriais baseado em dados/Afrânio José de
Melo Junior. – Rio de Janeiro: UFRJ/COPPE, 2023.
XIX, 347 p.: il.; 29, 7cm.
Orientadores: José Carlos Costa da Silva Pinto
Maurício Melo Câmara
Tese (doutorado) – UFRJ/COPPE/Programa de
Engenharia Química, 2023.
Referências Bibliográficas: p. 190 – 334.
1. integração humano-máquina. 2. monitoramento
de processos. 3. ciência de dados. 4. aprendizado de
máquina. 5. indústria 4.0. 6. transformação digital. 7.
detecção e diagnóstico de falhas. 8. análise exploratória
de dados. 9. orientação a objetos. 10. matrizes de
distâncias. I. Pinto, José Carlos Costa da Silva *et al.* II.
Universidade Federal do Rio de Janeiro, COPPE, Programa
de Engenharia Química. III. Título.

“Muriel Rukeyser says that ‘the universe is made of stories, not of atoms.’ I describe myself as a ‘patternist,’ someone who views patterns of information as the fundamental reality. For example, the particles composing my brain and body change within weeks, but there is a continuity to the patterns that these particles make. A story can be regarded as a meaningful pattern of information, so we can interpret Muriel Rukeyser’s aphorism from this perspective. (...) I believe that it’s the evolution of patterns that constitutes the ultimate story of our world.”

Ray Kurzweil

“For politicians, business people and ordinary consumers, Dataism offers groundbreaking technologies and immense new powers. For scholars and intellectuals it also promises to provide the scientific holy grail that has eluded us for centuries: a single overarching theory that unifies all the scientific disciplines from musicology through economics to biology. According to Dataism, Beethoven’s Fifth Symphony, a stock-exchange bubble and the flu virus are just three patterns of data flow that can be analysed using the same basic concepts and tools. This idea is extremely attractive. It gives all scientists a common language, builds bridges over academic rifts and easily exports insights across disciplinary borders. Musicologists, economists and cell biologists can finally understand each other.”

Yuval Noah Harari

Agradecimentos

A minha família, em especial minha mãe, meu pai e minhas avós, por me fazerem aprender o quanto aprender é precioso.

A minha amada Tayná pelo carinho, paciência e apoio essenciais no final deste ciclo.

Aos orientadores Zé e Maurício pela inspiração, amizade e confiança.

Aos colegas do Grupo de Monitoramento Digital de Processos do LMSCP/Engepol: Carol Sá, Carol Dias, Deris, Diego, Gabriel, Gildeir, Giovani, Jesus, Juliana, Luiz, Marcelo, Nayher, Rafael, Tahyná, Thainá, Tiago Lemos, Thiago Feital e Wallthynay. Sem essa galera, esta tese não existiria.

Aos interlocutores técnicos do CENPES/Petrobras com quem interagi em projetos de pesquisa: Thiago Anzai, Fábio Diehl e Pedro Thompson.

Aos amigos Letícia, Bruna, Pedro, Carol e Fellipe, que me acompanham desde o mestrado.

Aos alunos do curso de Data Science e Machine Learning da Escola Piloto Prof. Giulio Massarani pelo período incrível de trocas e aprendizado.

Aos membros da banca examinadora pela disponibilidade e interesse.

Ao PEQ pela estrutura e suporte, em especial a Vera Cruz pelos atendimentos solícitos e eficientes.

A CAPES pelo apoio financeiro.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

INTEGRAÇÃO HUMANO-MÁQUINA PARA O MONITORAMENTO DE PROCESSOS INDUSTRIAIS BASEADO EM DADOS

Afrânio José de Melo Junior

Maio/2023

Orientadores: José Carlos Costa da Silva Pinto

Maurício Melo Câmara

Programa: Engenharia Química

Este trabalho propõe metodologias que compartilham o objetivo geral de aumentar a integração humano-máquina durante a aplicação de técnicas de monitoramento de processos baseadas em dados. Em particular, três etapas do fluxo de monitoramento foram consideradas: (i) desenvolvimento de *software*; (ii) análise exploratória de dados; e (iii) detecção e diagnóstico de falhas. Para a etapa (i), desenvolve-se uma arquitetura baseada no paradigma de orientação a objetos de modo a permitir implementações facilmente extensíveis e aplicáveis a sistemas industriais. Essa arquitetura foi empregada para a elaboração de um software que tem sido utilizado com sucesso em várias aplicações de monitoramento em tempo real na indústria. Para a etapa (ii), propõe-se uma metodologia de análise exploratória de dados voltada para séries temporais provenientes de processos industriais, aplicando-se à análise crítica de conjuntos de dados de *benchmarks* abertos da literatura. Essa análise resultou em informações valiosas para apoiar a escolha adequada de um *benchmark* para uma aplicação específica, ou vice-versa. Para a etapa (iii), desenvolve-se uma técnica baseada em matrizes de distâncias para representação de padrões intrínsecos e interpretáveis. Essa abordagem resultou em uma metodologia de análise e monitoramento de processos fácil de ser compreendida e interpretada por operadores. Todos os métodos propostos têm o potencial de promover uma sinergia mais eficiente entre as habilidades humanas e as capacidades das máquinas, visando aprimorar os sistemas de monitoramento de processos e aumentar a aceitação e disseminação das técnicas baseadas em dados na indústria.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

HUMAN-MACHINE INTEGRATION FOR INDUSTRIAL DATA-DRIVEN PROCESS MONITORING

Afrânio José de Melo Junior

May/2023

Advisors: José Carlos Costa da Silva Pinto
Maurício Melo Câmara

Department: Chemical Engineering

The present work proposes methodologies to enhance human-machine integration during the application of data-driven process monitoring techniques. In particular, three stages of the monitoring pipeline are considered: (i) software development; (ii) exploratory data analysis; and (iii) fault detection and diagnosis. For step (i), an architecture based on the object-oriented paradigm is developed in order to allow implementations that are easily extensible and applicable to industrial systems. This architecture was applied in the development of a software that has been used successfully in various real-time monitoring applications in the industry. For step (ii), an exploratory data analysis methodology is proposed with focus on time series generated by industrial processes and is applied to the critical analysis of open benchmark datasets that are available in the literature. This analysis resulted in valuable information to support the proper choice of a benchmark for a specific application, or vice versa. For step (iii), a technique based on distance matrices is developed to represent intrinsic and interpretable patterns. This approach has resulted in a methodology for process analysis and monitoring that is easy to understand and interpret by operators. All proposed methods have the potential to promote a more efficient synergy between human skills and machine capabilities, aiming to enhance process monitoring systems and increase acceptance and dissemination of data-driven techniques in the industry.

Sumário

Agradecimentos	v
Lista de Figuras	xiii
Lista de Tabelas	xix
1 Introdução	1
1.1 Objetivo do trabalho	3
1.2 Estrutura da tese	4
2 Referencial teórico e revisão da literatura	5
2.1 Breve resumo histórico	6
2.2 Monitoramento univariado	8
2.3 Monitoramento multivariado	10
2.4 Modelagem em variáveis latentes	12
2.4.1 Análise em componentes principais (PCA)	14
2.4.2 Mínimos quadrados parciais (PLS) e análise de correlações canônicas (CCA)	27
2.4.3 Outras técnicas	30
2.5 Aprendizado de máquina	32
2.5.1 Definições e histórico na área de monitoramento	33
2.5.2 Tipos de aprendizado	35
2.5.3 Redes neurais artificiais (ANN)	36

2.5.4	Máquinas de vetores suporte (SVM)	48
2.5.5	Outras técnicas	55
2.6	Aplicações na indústria	59
2.7	Ferramentas computacionais	70
2.7.1	Ferramentas comerciais	70
2.7.2	Ferramentas abertas	72
2.8	<i>Benchmarks</i>	73
2.8.1	Tennessee Eastman Process (1993)	74
2.8.1.1	Desenvolvimento do <i>benchmark</i> TEP	75
2.8.2	PenSim (2002)	77
2.8.3	DAMADICS (2006)	78
2.8.4	RAYMOND (2014)	79
2.8.5	IndPenSim (2015)	79
2.8.6	Cranfield Multiphase Flow Facility (2015)	80
2.8.7	PRONTO (2019)	80
2.8.8	Conjunto de dados 3W (2019)	81
2.8.9	Fracionador FCC (2022)	82
2.8.10	Outros conjuntos de dados	82
2.8.11	Modelos	83
2.9	Exploração, caracterização e tratamento de dados de processo	83
2.10	Oportunidades em outros domínios	87
2.11	Integração humano-máquina	91
2.12	Considerações finais	93
3	Implementação de técnicas de monitoramento	94
3.1	Programação orientada a objetos	94
3.1.1	Tipos abstratos de dados e classes	95
3.1.2	Herança	96

3.1.3	Composição	96
3.1.4	Polimorfismo	96
3.2	Metodologia	97
3.2.1	Modelos de monitoramento	97
3.2.1.1	Extensibilidade por herança	97
3.2.2	Tratamento de dados	98
3.2.3	Funções de conveniência	98
3.3	Estudo de caso: BibMon	99
3.4	Considerações finais	101
4	Análise exploratória de dados de processo	105
4.1	Metodologia	105
4.1.1	Visualização de dados	106
4.1.2	Teste de gaussianidade	106
4.1.3	Análise de autocorrelação	107
4.1.4	Estimação da razão sinal/ruído	108
4.1.5	Análise de correlações lineares e não lineares	109
4.1.6	Análise de falhas	110
4.2	Aplicações e resultados	110
4.2.1	Tennessee Eastman Process (TEP)	110
4.2.1.1	Dados normais	110
4.2.1.2	Dados de falha	119
4.2.2	DAMADICS	123
4.2.2.1	Dados normais	124
4.2.2.2	Dados de falha	131
4.2.3	IndPenSim	132
4.2.3.1	Dados normais	135
4.2.3.2	Dados de falha	138

4.2.4	Cranfield Multiphase Flow Facility	141
4.2.4.1	Dados normais	142
4.2.4.2	Dados de falha	146
4.2.5	3W	149
4.3	Discussão	157
4.3.1	Modos de operação	157
4.3.2	Disponibilidade de um modelo fenomenológico	157
4.3.3	Disponibilidade e fonte de dados	159
4.3.4	Variabilidade de dados	159
4.3.5	Não linearidade	160
4.3.6	Dinâmica	160
4.3.7	Gaussianidade	161
4.3.8	Qualidade de dados	161
4.4	Considerações finais	161
5	Análise de matrizes de distâncias para detecção e diagnóstico de falhas	163
5.1	<i>Framework teórico</i>	164
5.1.1	Matrizes de distâncias	164
5.1.2	Algoritmo de detecção de falhas	168
5.2	Resultados e discussão	170
5.2.1	Exemplo 1 - Caso numérico	170
5.2.2	Exemplo 2 - CSTR	173
5.2.3	Exemplo 3 - <i>Benchmark</i> TEP	177
5.3	Considerações finais	183
6	Conclusões e perspectivas	188
	Referências Bibliográficas	190

A Trabalhos-chave utilizados na revisão da literatura	335
Apêndices	335

Lista de Figuras

2.1	Ilustração da carta de Shewhart para a média, com violação da regra de Nelson nº 8 (GMCGLINN, 2008).	10
2.2	Região de confiança para duas variáveis correlacionadas, ilustrando a necessidade do monitoramento multivariado (FEITAL, 2011).	11
2.3	Modelo de um neurônio biológico (VU-QUOC, 2018).	37
2.4	Modelo de um neurônio artificial (CHRISLB, 2005).	37
2.5	Rede de percéptrons com camadas ocultas (adaptado de TSEKICHUN (2021)).	39
2.6	Comparação entre as funções de ativação degrau binária e sigmóide. .	40
2.7	Ilustração da operação de convolução. Adaptado de NG e FENG (2020).	44
2.8	Geração de um novo mapa de características a partir da aplicação de um filtro de convolução. Adaptado de GONG <i>et al.</i> (2019).	45
2.9	Ilustração da operação de <i>pooling</i> (YANI <i>et al.</i> , 2019).	45
2.10	Arquitetura típica de uma rede convolucional. Adaptado de PHUNG e RHEE (2019).	46
2.11	Ilustração em espaço bidimensional do problema de classificação binária linearmente separável.	49
2.12	Ilustração em espaço bidimensional da resolução do problema de classificação binária com contornos não lineares utilizando SVM e <i>kernel</i> RBF gaussiano.	53
2.13	Ilustração da criação de fronteira multimodal. O gradiente de cores tem a ver com a confiança da classificação, relacionada diretamente à distância para a fronteira de decisão.	54

3.1	Estrutura do pacote BibMon.	100
3.2	Diagrama UML de classes do pacote BibMon.	102
3.3	Resultados de detecção de falhas para a Falha 1 do <i>benchmark</i> TEP. .	103
3.4	Resultados de diagnóstico de falhas para a Falha 1 do <i>benchmark</i> TEP.103	
3.5	Resultados de detecção de falhas para um conjunto de dados reais de planta de processo.	104
4.1	Séries temporais dos dados normais do <i>benchmark</i> TEP.	113
4.2	Gráficos de dispersão de variáveis contínuas medidas e manipuladas de uma simulação do <i>benchmark</i> TEP.	114
4.3	Gráficos de dispersão de variáveis contínuas medidas e manipuladas de quinhentas simulação do <i>benchmark</i> TEP.	115
4.4	Efeito da adição de mais simulações do <i>benchmark</i> TEP no gráfico de dispersão “Válvula de purga” X “Válvula de vapor do esgotador”. . . .	116
4.5	Testes de gaussianidade do <i>benchmark</i> TEP usando 1 simulação (500 amostras).	116
4.6	Testes de gaussianidade do <i>benchmark</i> TEP usando 3 simulações (1500 amostras).	116
4.7	Gráficos de função de autocorrelação para algumas variáveis do <i>benchmark</i> TEP.	117
4.8	Tempo necessário para atingir 0,5 de autocorrelação no <i>benchmark</i> TEP.117	
4.9	Razões sinal/ruído no <i>benchmark</i> TEP.	117
4.10	Coeficientes de correlação para 1 simulação normal do <i>benchmark</i> TEP.118	
4.11	Evolução dos coeficientes de correlação do <i>benchmark</i> TEP à medida que o tamanho do filtro aumenta.	119
4.12	Evolução da relação entre duas variáveis no <i>benchmark</i> TEP à medida que o tamanho do filtro aumenta.	119
4.13	Comparação de séries temporais de conjuntos de dados normais (linhas pretas) e em falha (linhas vermelhas) do <i>benchmark</i> TEP (falha número 5). A reta tracejada azul indica o início da falha.	122
4.14	Carta de controle PCA para a falha 5 do <i>benchmark</i> TEP.	123

4.15	Séries temporais dos dados normais do <i>benchmark</i> DAMADICS.	125
4.16	Gráficos de dispersão de todos os dados do <i>benchmark</i> DAMADICS. .	126
4.17	Testes de gaussianidade do <i>benchmark</i> DAMADICS para períodos de duas horas.	127
4.18	Testes de gaussianidade do <i>benchmark</i> DAMADICS para períodos de seis horas.	127
4.19	Tempo necessário para atingir 0,5 de autocorrelação no <i>benchmark</i> DAMADICS.	127
4.20	Tempo necessário para atingir 0,5 de autocorrelação no <i>benchmark</i> DAMADICS.	128
4.21	Razões sinal/ruído no <i>benchmark</i> DAMADICS.	128
4.22	Razões sinal/ruído contra taxa de amostragem no <i>benchmark</i> DAMADICS.	128
4.23	Coeficientes de correlação para um período de quatro horas do <i>benchmark</i> DAMADICS.	129
4.24	Coeficientes de correlação não linear diários para o <i>benchmark</i> DAMADICS.	130
4.25	Coeficientes de correlação linear diários para o <i>benchmark</i> DAMADICS.	130
4.26	Efeito do filtro de média móvel nos coeficientes de correlação no <i>benchmark</i> DAMADICS.	131
4.27	Efeito do filtro de média móvel em um gráfico de dispersão do <i>benchmark</i> DAMADICS.	131
4.28	Cartas de controle PCA de treinamento e validação para o dia 09/11 do DAMADICS.	132
4.29	Cartas de controle PCA de teste para o dia 09/11 do DAMADICS. .	133
4.30	Cartas de controle PCA de treinamento e validação para o dia 17/11 do DAMADICS.	133
4.31	Cartas de controle PCA de teste para o dia 17/11 do DAMADICS. .	133
4.32	Séries temporais dos dados normais do <i>benchmark</i> IndPenSim.	136
4.33	Gráficos de dispersão dos lotes 1–30 do <i>benchmark</i> IndPenSim.	137

4.34	Tempo necessário para se atingir 0,5 de autocorrelação no <i>benchmark</i> IndPenSim.	138
4.35	Razões sinal/ruído no <i>benchmark</i> IndPenSim.	138
4.36	Testes de gaussianidade do <i>benchmark</i> IndPenSim.	138
4.37	Coeficientes de correlação para 1 lote do <i>benchmark</i> IndPenSim.	139
4.38	Coeficientes de correlação global para todos os lotes do <i>benchmark</i> IndPenSim.	140
4.39	Cartas de controle PCA de treino e validação do <i>benchmark</i> IndPenSim.	141
4.40	Cartas de controle PCA de teste do <i>benchmark</i> IndPenSim.	141
4.41	Séries temporais dos dados normais do <i>benchmark</i> Cranfield.	143
4.42	Gráficos de dispersão do conjunto de dados normais T1 do <i>benchmark</i> Cranfield.	144
4.43	Gráficos de dispersão selecionados do conjunto de dados normais T1 do <i>benchmark</i> Cranfield.	145
4.44	Testes de gaussianidade dos conjuntos de dados normais do <i>benchmark</i> Cranfield.	145
4.45	Tempo necessário para atingir 0,5 de autocorrelação no <i>benchmark</i> Cranfield.	145
4.46	Razões sinal/ruído no <i>benchmark</i> Cranfield.	145
4.47	Razão sinal/ruído contra taxa de amostragem no <i>benchmark</i> Cranfield.	146
4.48	Coeficientes de correlação para o conjunto de dados normais T1 do <i>benchmark</i> Cranfield.	147
4.49	Coeficientes de correlação contra taxa de amostragem no <i>benchmark</i> Cranfield.	149
4.50	Coeficientes de correlação contra tamanho de filtro de média móvel no <i>benchmark</i> Cranfield.	149
4.51	Cartas de controle para os três conjuntos de dados da falha 2 do <i>benchmark</i> Cranfield.	149
4.52	Gráficos de séries temporais da variável 4 para todos os 8 poços do <i>benchmark</i> 3W.	151
4.53	Gráficos de dispersão de todos os dados normais do <i>benchmark</i> 3W. .	151

4.54 Concatenação de períodos de tempo identificados a partir de dados normais do Poço 1. As instâncias são separadas por linhas tracejadas azuis e os períodos de tempo são separados por linhas vermelhas.	153
4.55 Séries temporais e gráficos de dispersão selecionados de um par de variáveis de dados W_1P_2	154
4.56 Resultados da aplicação de um algoritmo detector de pico aos dados descritos na Figura 4.55.	155
4.57 Faixa de tempo selecionada da variável 6, período W_1P_2	155
4.58 Resultados de uma tentativa de se realizar análise exploratória de dados no período W_1P_2	156
5.1 Ilustração esquemática de uma série temporal e sua matriz de distâncias correspondente.	165
5.2 Matrizes de distâncias e recorrências para sinais de processo típicos. As distâncias medianas foram usadas como limiares nos gráficos de recorrência. H refere-se à função de Heaviside e N refere-se à distribuição normal.	166
5.3 Matrizes de distâncias para combinações típicas de sinais de processo. .	167
5.4 Etapa de treinamento do algoritmo de detecção de falhas (FADEL, 2018).	169
5.5 Etapa de teste do algoritmo de detecção de falhas (FADEL, 2018). . .	170
5.6 Séries temporais do Exemplo 1.	171
5.7 Carta de controle para o cenário normal do Exemplo 1.	171
5.8 Cenários em falha analisados no Exemplo 1: degrau, aumento de variabilidade, oscilação, tendência e tendência seguido de estabilização.	172
5.9 Matrizes de distâncias para os cenários analisados no Exemplo 1. . . .	173
5.10 Função de desvio aplicada a κ_0 no cenário de desativação do catalisador do Exemplo 2 (CSTR).	176
5.11 Matrizes de distâncias para os casos analisados no Exemplo 2 (CSTR).	176
5.12 Carta de controle para o cenário normal do Exemplo 2 (CSTR). . . .	177
5.13 Cartas de controle para os cenários de falha do Exemplo 2 (CSTR). .	177
5.14 Matrizes de distâncias para os casos analisados no Exemplo 3 (TEP). .	179

5.15 Matrizes de distâncias para diferentes simulações do evento IDV(12) do <i>benchmark</i> TEP.	180
5.16 Matrizes de distâncias para diferentes simulações do evento IDV(16) do <i>benchmark</i> TEP.	180
5.17 Carta de controle para o cenário normal do Exemplo 3 (TEP).	180
5.18 Carta de controle para os cenários de falha do Exemplo 3 (TEP).	181
5.19 Taxas de detecção de falhas em função dos tamanhos de janela para os cenários de falha do Exemplo 3 (TEP).	183
5.20 Arquitetura da rede neural convolucional usada para diagnosticar falhas de TEP. <i>None</i> corresponde à dimensão referente ao tamanho do lote, que é independente da arquitetura da rede e escolhida no momento de ajuste do modelo.	185
5.21 Curvas de aprendizado apresentando taxas de verdadeiros positivos em função do tamanho da janela de análise.	186
5.22 Curvas de aprendizado apresentando acurácia em função do tamanho da janela de análise.	186
5.23 Matriz de confusão para análise de diagnóstico com tamanho de janela de 250.	187

Listas de Tabelas

2.1	Monitoramento de processos: algumas aplicações industriais reais ou estudos envolvendo dados reais de planta.	59
4.1	Variáveis do <i>benchmark</i> TEP.	111
4.2	Falhas do <i>benchmark</i> TEP.	121
4.3	Variáveis do <i>benchmark</i> DAMADICS.	123
4.4	Resultados diários de detecção utilizando o modelo PCA para o <i>benchmark</i> DAMADICS.	132
4.5	Variáveis do <i>benchmark</i> IndPenSim.	134
4.6	Variáveis do <i>benchmark</i> Cranfield.	141
4.7	Análise das falhas do <i>benchmark</i> Cranfield.	148
4.8	Variáveis do <i>benchmark</i> 3W.	150
4.9	Períodos de tempo identificados nos dados normais do <i>benchmark</i> 3W.	152
4.10	Resumo das características dos <i>benchmarks</i>	158
4.11	Resumo da análise exploratória de dados efetuada.	159
5.1	Parâmetros da simulação do CSTR.	175
5.2	Medições da simulação do CSTR.	175
5.3	Taxas de detecção de falhas do Exemplo 3 (TEP).	182
A.1	Artigos de revisão e que apresentam visão global sobre técnicas de monitoramento de processos baseado em dados.	335
A.2	Artigos de comparação de métodos de monitoramento baseado em dados.	344

A.3 Livros sobre monitoramento de processos baseado em dados. 345

Capítulo 1

Introdução

Tecnologias relacionadas a computação e sistemas de informação vêm revolucionando o mundo nas últimas décadas e as perspectivas apontam que tudo o que ainda está por vir tende a ser cada vez mais intenso e disruptivo (GIRASA, 2020). Podem ser mencionados, como exemplos, os evidentes avanços nas áreas de comunicações (GRAHAM e DUTTON, 2019), tecnologias *blockchain* (ASTE *et al.*, 2017), computação quântica (BERNAL *et al.*, 2022) e inteligência artificial (ACEMOGLU e RESTREPO, 2018; CHIANG *et al.*, 2022; KAPLAN e HAENLEIN, 2019; KUMAR, 2017; LEE *et al.*, 2018a; LI *et al.*, 2017a; OLIVEIRA *et al.*, 2021; PAN, 2016; RUSSELL *et al.*, 2015; VENKATASUBRAMANIAN, 2019). No meio corporativo, em particular, o termo *transformação digital* adquiriu recentemente importância considerável, tendo sido aplicado com bastante frequência para simbolizar esses avanços.

Nesse contexto, ganha destaque o conceito de Indústria 4.0, proposto na Alemanha no início da década de 2010 para simbolizar uma série de mudanças de paradigma interpretadas como a concretização da Quarta Revolução Industrial (CAÑAS *et al.*, 2021; DIEZ-OLIVAN *et al.*, 2019; ESCOBAR *et al.*, 2021; LU, 2017; LÓPEZ-GUAJARDO *et al.*, 2022; OZTEMEL e GURSEV, 2020; REIS e KENETT, 2018; REIS e SARAIVA, 2021; ROJKO, 2017; STEINWANDTER *et al.*, 2019; TADEU *et al.*, 2019; UDUGAMA *et al.*, 2022). Tendências da Indústria 4.0 incluem sistemas ciberfísicos (*cyber-physical systems*, *CPS*), internet das coisas (*internet of things*, *IoT*), interação humano-computador (*human-computer interaction*, *HCI*), automação de decisões, realidade aumentada, robôs autônomos, integração de sistemas, cibersegurança e computação de alto desempenho integrada na planta ou em nuvem.

Um tipo de bem que se torna cada vez mais relevante e estratégico nesse ambiente são os *dados*. Pode-se definir dados como “um conjunto de símbolos

sem qualquer significado além de sua existência” (MARQUARDT *et al.*, 2010). A extração de valor por meio da atribuição de significado aos dados, transformando-os em *informação*, é uma das operações mais rentáveis do mundo contemporâneo. No campo das mídias sociais, companhias bilionárias como *Facebook* constroem modelos de negócio com base nesse procedimento.

Apesar de a análise e a interpretação de dados serem atividades antigas, estando na base do processo cognitivo humano e do desenvolvimento da ciência, o ambiente altamente informatizado e inovador descrito nos parágrafos anteriores requer uma mudança de perspectiva. É nesse cenário que emerge uma nova disciplina, conhecida como *ciência de dados*¹ (ou, mais comumente pelo termo em inglês, *data science*), com metodologias e técnicas de análise próprias para integrar (i) ferramentas computacionais, (ii) conhecimento matemático/estatístico e (iii) conhecimento do domínio de origem e/ou aplicação dos dados (BECK *et al.*, 2016; CONWAY, 2010; GRUS, 2015; MOWBRAY *et al.*, 2022; SKIENA, 2017; VANDERPLAS, 2017). Em alguns casos, como durante a geração de quantidades massivas e variadas de dados a altas taxas (constituindo o cenário usualmente conhecido como *big data*), a aplicação de técnicas desenvolvidas no âmbito da ciência de dados torna-se imperativa (CHIANG *et al.*, 2017; GANDOMI e HAIDER, 2015; GAO *et al.*, 2020a; GROVER *et al.*, 2018; GUNTHER *et al.*, 2017; JONES, 2019; LAVASANI *et al.*, 2023; OUSSOUS *et al.*, 2018; QIN, 2014; SIVARAJAH *et al.*, 2017; UDUGAMA *et al.*, 2020; YANG e GE, 2022).

No mundo corporativo e no ambiente de negócios, a análise de dados para fins estratégicos costuma ser referenciada pelo termo *analytics* (ACITO e KHATRI, 2014; CHIANG *et al.*, 2018; GROVER *et al.*, 2018; KUNC e O'BRIEN, 2019; REIS e SARAIWA, 2022), que infelizmente não possui tradução adequada para o português (o uso da palavra “analítica” como substantivo não é disseminado na língua). A indústria de processos vem adotando esse jargão e enfatizando a importância de seu significado e aplicação, não sendo difícil encontrar trabalhos recentes da área cujos títulos contenham o termo (CHIANG *et al.*, 2017; GE *et al.*, 2017; GOEL *et al.*, 2017, 2020; HE e WANG, 2018; MOYNE e ISKANDAR, 2017; QIN, 2014; QIN e CHIANG, 2019; REIS, 2019; STIEF *et al.*, 2019b; SUN e BRAATZ, 2021; TSAI *et al.*, 2018).

Dentre as possibilidades de aplicação do *analytics* na indústria de processos, destaca-se a área de *monitoramento de processos* (FISHER *et al.*, 2020; HE e WANG, 2018; QIN e CHIANG, 2019; REIS e GINS, 2017; SHU *et al.*, 2016). No

¹Talvez o nome ciência de dados não seja o mais adequado do ponto de vista semântico e histórico, já que os termos “ciência” e “dados” existem há séculos e são intimamente relacionados. É o nome, porém, que se consagrou, sendo improvável que caia em desuso (SAUNDERS, 2013).

monitoramento, dados de processo são analisados de modo a permitir a obtenção de conclusões sobre a condição de operação, classificada como saudável² ou não. Em caso negativo, deve-se identificar a causa do comportamento anormal e procurar maneiras para corrigi-lo (REIS e KENETT, 2018). Além dos problemas tradicionais, como por exemplo os relacionados a falhas de processo e de medição, novos desafios vêm surgindo, como a possibilidade de ataques cibernéticos a sistemas de controle (CHEN *et al.*, 2020a; KRAVCHIK *et al.*, 2022). No Capítulo 2 são apresentados detalhes da teoria e do histórico que embasam os procedimentos de monitoramento.

Técnicas baseadas em dados (ou, como é comum dizer, técnicas *data-driven*) são técnicas matemáticas que identificam padrões nos dados e os convertem em informação, segundo o esquema a seguir:

$$\text{Dados} \rightarrow \textbf{Padrões} \rightarrow \text{Informação}$$

O objetivo da aplicação das técnicas é, em última análise, gerar valor a partir dos dados. No contexto do aprendizado humano, o fluxo seguiria a partir da *informação* por meio da sequência *conhecimento, entendimento e sabedoria* (JONES, 2019; MARQUARDT *et al.*, 2010). Todos os três conceitos podem ser encarados como formas de valor. Na indústria de processos, a informação pode ser usada para geração de valor sem a necessidade da participação humana direta. Um exemplo é a aplicação associada a sistemas automáticos de detecção de falhas e/ou prevenção de perdas. No entanto, dado que a inteligência artificial ainda está longe de conseguir o grau de cognição da inteligência natural, a participação humana se mostra ainda imprescindível em vários aspectos (EMMANOUILIDIS *et al.*, 2019; RIETH *et al.*, 2018; SACHA *et al.*, 2017; VILLALBA-DIEZ e ORDIERES-MERÉ, 2021). A hipótese que norteia o presente trabalho é de que, além do desenvolvimento de algoritmos mais eficientes, ainda há muito a se contribuir no que tange à integração entre humanos e sistemas de inteligência artificial para análise e monitoramento de processos.

1.1 Objetivo do trabalho

O objetivo geral do trabalho foi fornecer uma ampla gama de contribuições com o propósito comum de propiciar a participação humana eficiente e eficaz nos *pipelines* de análise e monitoramento de processos baseado em dados.

Os objetivos específicos foram:

²Algumas referências utilizam o termo *estável* ao invés de *saudável*, mas pode haver confusão com o conceito de estabilidade dinâmica. Uma boa alternativa é o uso do termo *estável estatisticamente*.

- proposição e utilização de uma arquitetura de *software* para monitoramento de processos que permita uma implementação facilmente extensível e aplicável a sistemas industriais em tempo real;
- desenvolvimento, implementação e aplicação de uma metodologia para análise exploratória de dados especialmente voltada para dados de processos industriais;
- desenvolvimento, implementação e aplicação de uma técnica baseada em análise de padrões intrínsecos dos dados para detecção e diagnóstico de falhas, eficaz do ponto de vista numérico e facilmente interpretável por humanos.

Cada um dos objetivos resultou em um dos capítulos do trabalho, como detalhado na subseção a seguir.

1.2 Estrutura da tese

No Capítulo 2 apresentam-se o referencial teórico e a revisão da literatura sobre técnicas baseadas em dados, utilizadas para monitoramento e análise de processos. No Capítulo 3 propõem-se uma metodologia para implementação de técnicas de monitoramento de processos e um estudo de caso utilizando o *software* BibMon (**Biblioteca de Monitoramento de Processos**) (MELO *et al.*, 2023). No Capítulo 4 apresenta-se uma metodologia para análise exploratória de dados, implementada no *software* KydLIB (**Know Your Data Library**) (MELO, 2023) e aplicada a vários conjuntos de dados industriais abertos disponíveis na literatura. No Capítulo 5 desenvolve-se uma técnica baseada em matrizes de distâncias para monitoramento de processos baseado em padrões intrínsecos e interpretáveis. No Capítulo 6, apresentam-se as conclusões e perspectivas futuras.

A presente tese foi desenvolvida no Laboratório de Modelagem, Simulação e Controle de Processos (LMSCP) do Programa de Engenharia Química (PEQ) do Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia (COPPE). O autor agradece à instituição pela estrutura e suporte e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Petrobras pelas bolsas de estudos fornecidas durante a realização do trabalho.

Capítulo 2

Referencial teórico e revisão da literatura

Neste capítulo são apresentados a revisão da literatura e o referencial teórico do trabalho proposto. Trata-se em específico do monitoramento de processos industriais baseado em dados, com ênfase no desenvolvimento histórico das várias metodologias, na base teórica das principais técnicas, nas aplicações em tempo real reportadas na literatura, nas ferramentas e dados disponíveis abertamente na literatura e nas oportunidades de sinergia com outros domínios do conhecimento.

O capítulo está estruturado como segue. Inicia-se na Seção 2.1 com um breve resumo histórico, em que se discute de forma ampla a evolução da área de monitoramento de processos industriais ao longo dos séculos XX e XXI. Na Seção 2.2 é introduzido o procedimento de monitoramento univariado. Na Seção 2.3 são discutidas a necessidade do monitoramento multivariado e o problema da colinearidade da matriz de dados que surge nesse contexto. A principal solução para o problema, a modelagem em variáveis latentes, é apresentada na Seção 2.4. As cada vez mais relevantes técnicas baseadas em aprendizado de máquina são apresentadas na Seção 2.5. Na Seção 2.6 é apresentado um levantamento de aplicações reais de técnicas de monitoramento baseado em dados reportadas na literatura. Nas Seções 2.7 e 2.8 apresentam-se as principais ferramentas computacionais e *benchmarks*, respectivamente, utilizadas pela comunidade de monitoramento de processos. Na Seção 2.9 apresenta-se uma revisão sobre caracterização e tratamento de dados de processo, uma etapa que se mostra fundamental durante a implementação de um sistema de monitoramento. Na Seção 2.10 discutem-se as oportunidades mapeadas de sinergia de outros domínios do conhecimento com a área de monitoramento de processos. Na Seção 2.11 apresenta-se um levantamento de estudos envolvendo integração humano-máquina em monitoramento. Na Seção 2.12 são apresentadas as

considerações finais.

No Apêndice A encontram-se listados em ordem cronológica e brevemente descritos os principais livros e artigos de revisão, que apresentam visões globais e comparações, utilizados como referência para a escrita deste capítulo.

2.1 Breve resumo histórico

O termo monitoramento de processos, em sua aplicação cotidiana e na prática de engenharia, pode ter um sentido bastante amplo. O dicionário *on-line* <https://www.dictionary.com/> apresenta uma definição bem adequada para o presente contexto:

monitorar: “*observar, gravar ou detectar (uma operação ou condição) com instrumentos que não têm efeito na operação ou na condição*”.

Nesse sentido, o monitoramento dos processos industriais acontece desde que eles foram inventados, mesmo que na forma do simples acompanhamento da condição do processo pelos operadores e engenheiros responsáveis.

SHEWHART (1926, 1931), durante trabalho no *Bell Telephone Laboratories*, propôs o conceito de *cartas de controle*, ferramentas gráficas que, com embasamento em conceitos estatísticos, permitem o monitoramento de variáveis importantes do processo com o objetivo de detectar qualquer evento associado a variações de causa especial (em contraste às variações de causa comum, aquelas inevitáveis e que sempre ocorrem durante a operação de um processo). Esse é considerado o marco fundador da área de *monitoramento estatístico de processos* (MONTGOMERY, 2012).

Outro marco importante foi a proposição da estatística T^2 por HOTELLING (1947), ferramenta que possibilitou pela primeira vez a análise de múltiplas variáveis (e suas interações) de forma simultânea. Esse trabalho constitui o início da área de *monitoramento estatístico multivariado de processos* (DAS *et al.*, 2012).

Apesar de Hotelling ter introduzido a base teórica para o monitoramento multivariado em 1947, apenas no fim da década de 1980 essa aplicação se disseminou, por conta principalmente de dois fatores: (i) a popularização do uso de computadores e (ii) a proposição da aplicação da técnica PCA (análise em componentes principais) para fins de monitoramento (WISE *et al.*, 1988), o que deu início à aplicação da *modelagem em variáveis latentes* no contexto do monitoramento, como descrito na Seção 2.4.

Da década de 1990 em diante, a aplicação de variáveis latentes para modelagem dos dados tornou-se a metodologia padrão para a abordagem do problema de monitoramento multivariado baseado em dados, como evidenciado em muitos artigos de revisão e de visão global da época (MACGREGOR e KOURTI, 1995; QIN, 2003; WISE e GALLAGHER, 1996).

A partir da década de 2000, o avanço da capacidade computacional vem fazendo com que abordagens baseadas em *aprendizado de máquina*³ ganhem cada vez mais relevância, especialmente por conta da capacidade em lidar com grandes conjuntos de dados e com a identificação de padrões complexos (MD NOR *et al.*, 2019; QIN e CHIANG, 2019).

Monitoramento baseado em modelos fenomenológicos

O monitoramento baseado em modelos fenomenológicos, desenvolvido a partir da década de 1970 (DAS *et al.*, 2012), é utilizado quando se dispõe de um modelo fenomenológico que descreve adequadamente o processo. Nesses casos, uma quantidade relativamente baixa de dados é necessária para a construção do sistema de monitoramento, em contraste com os métodos baseados em dados, que requerem quantidades massivas de dados e consequente esforço computacional para o processamento e tratamento (TIDRIRI *et al.*, 2016). No entanto, quanto maior é a complexidade do processo a ser monitorado, maior também é a dificuldade para desenvolver ou aplicar um modelo fenomenológico para fins de monitoramento (CHIANG *et al.*, 2001).

Os métodos baseados em modelos fenomenológicos podem ser classificados como (i) baseados em estimativa de parâmetros, (ii) em observadores e (iii) em relações de paridade (DAS *et al.*, 2012). De modo geral, esses métodos estabelecem uma comparação entre o comportamento observado do processo e o descrito pelo modelo, sendo a falha detectada quando os comportamentos não coincidem. O presente trabalho, por tratar exclusivamente do monitoramento baseado em dados, não abordará esses métodos. Caso o leitor ou leitora tenha interesse na literatura dessa área específica, pode consultar as excelentes revisões realizadas por DAS *et al.* (2012); DING *et al.* (2011); ISERMANN (2005); LI *et al.* (2020c); TIDRIRI *et al.* (2016); VENKATASUBRAMANIAN *et al.* (2003b); ZHONG *et al.* (2018). Para trabalhos sobre abordagens híbridas, que mesclam métodos baseados em modelos fenomenológicos e em dados, pode-se consultar BIKMUKHAMETOV e JÄSCHKE (2020); BRADLEY *et al.* (2022); DESTRO *et al.* (2020); HASSANPOUR *et al.* (2020);

³Apesar da evidente diferença histórica, é possível classificar os métodos de modelagem em variáveis latentes também como algoritmos de aprendizado de máquina (QIN e CHIANG, 2019).

RAJULAPATI *et al.* (2022); SANSANA *et al.* (2021); SOARES (2022); TIDRIRI *et al.* (2016); YANG *et al.* (2020b).

É importante ressaltar que a grande maioria dos métodos de monitoramento baseado em dados também utilizam modelos, embora estes apresentem natureza empírica e sejam construídos a partir dos dados.

Nota sobre nomenclatura

O termo “monitoramento de processos” muitas vezes é usado como sinônimo de “controle estatístico de processos”. Esse uso parece ser derivado do termo “controle estatístico de qualidade”, que surgiu historicamente com os trabalhos de Shewart (SHEWHART, 1926, 1931). KOURTI (2002) interpreta a diferença entre os termos como resultante da diferença entre as variáveis incluídas na análise: o controle estatístico de processos analisa as variáveis de processo, além das de qualidade, o que aumenta significativamente a capacidade de diagnóstico dos comportamentos anormais das próprias variáveis de qualidade.

Na área de engenharia, o termo “controle” é mais empregado no sentido de “controle automático”, em que o objetivo é automatizar a manutenção de condições de operação frente à perturbações e mudanças que ocorrem rotineiramente no processo (CHIANG *et al.*, 2001). Por esse motivo, no presente texto o termo utilizado preferencialmente é “monitoramento de processos”.

Para uma proposta consistente de definições e terminologias em língua portuguesa na área de monitoramento de processos, consultar SARTORI *et al.* (2012).

2.2 Monitoramento univariado

O monitoramento univariado é efetuado por meio de *cartas de controle*, dentre as quais as pioneiras e mais usadas são as *cartas de Shewhart* (SHEWHART, 1926, 1931). Essas cartas se destinam a monitorar valores de referência e variabilidades das variáveis de qualidade. A hipótese básica é de que o processo está estável (ou, de acordo com o jargão, “sob controle estatístico”) se cada uma dessas variáveis segue uma distribuição normal com média e variância constantes. Desse modo, duas cartas são construídas para cada variável: uma para a média e outra para o desvio-padrão.

Várias regras heurísticas podem ser propostas para determinar quando uma variável indica instabilidade estatística no processo, como as *regras de Nelson* (NELSON,

1984), aplicadas por exemplo à carta de média:

- **1:** um ponto se encontra a mais que três desvios-padrão da média;
- **2:** nove ou mais pontos em sequência se encontram do mesmo lado em relação à média;
- **3:** seis ou mais pontos em sequência evoluem em um mesmo sentido (continuamente aumentando ou diminuindo);
- **4:** catorze ou mais pontos em sequência alternam em sentido;
- **5:** dois (ou três) de uma sequência de três pontos estão a mais que dois desvios-padrão da média, no mesmo sentido;
- **6:** quatro (ou cinco) de uma sequência de cinco pontos estão a mais que um desvio-padrão da média, no mesmo sentido;
- **7:** quinze pontos em sequência estão a um desvio padrão da média;
- **8:** oito pontos em sequência se encontram a mais que um desvio padrão da média, em ambos os lados em relação à média.

Todas essas regras podem ser justificadas com base na Teoria Estatística. Por exemplo, sob a hipótese da distribuição normal, a situação descrita na primeira regra corresponde a uma probabilidade de ocorrência de apenas 0,3% (SCHWAAB e PINTO, 2007)⁴. As três primeiras regras são análogas aos conhecidos alarmes de *outlier*, *bias* e *drift*, respectivamente (CÂMARA, 2018). A Figura 2.1 mostra o exemplo de uma carta de Shewhart para a média, com violação da regra de Nelson nº 8. Na Figura 2.1, UCL e LCL referem-se a *upper control limit* e *lower control limit*, respectivamente, correspondentes ao valor típico de três vezes o desvio-padrão além da média.

As cartas de Shewhart tradicionais são úteis para a detecção de mudanças grandes e abruptas de processo (DAS *et al.*, 2012). Para detecção de mudanças mais sutis, foram propostas várias cartas alternativas, dentre as quais se destacam a CUSUM (PAGE, 1954) e a EWMA (ROBERTS, 1959), que utilizam informações correntes e passadas recentes (FEITAL, 2011). A carta CUSUM envolve o cálculo de somas cumulativas, enquanto a EWMA faz uso de uma média móvel ponderada exponencialmente para diminuição do ruído.

⁴Justificativas para as demais regras podem ser encontradas utilizando as distribuições binomial e de Poisson (NELSON, 1984).

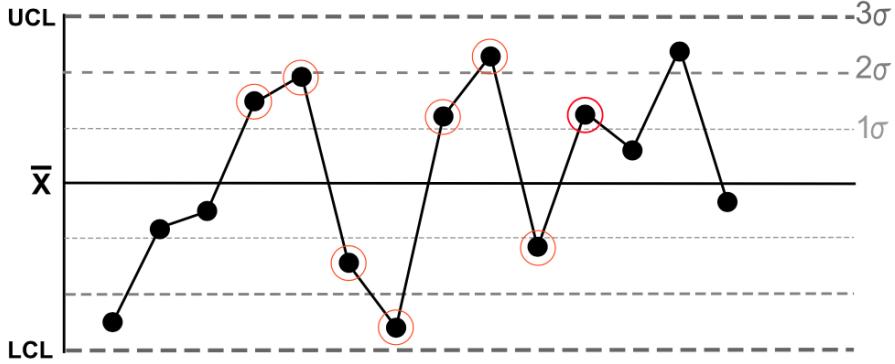


Figura 2.1: Ilustração da carta de Shewhart para a média, com violação da regra de Nelson nº 8 (GMCGLINN, 2008).

Apesar de o foco do presente trabalho ser o monitoramento multivariado, o monitoramento univariado continua a constituir tema de pesquisas e aplicações em várias áreas, inclusive na indústria de processos (BERLEMANN *et al.*, 2019; EISSA e HAMED, 2019; LIM e ANTONY, 2019; RAMESH e VASU, 2019; YAO e CHAKRABORTI, 2020). Para detalhes sobre os fundamentos das técnicas de monitoramento univariado, vale a consulta ao excelente livro de MONTGOMERY (2012).

2.3 Monitoramento multivariado

O monitoramento univariado é suficiente apenas para a análise de variáveis que não apresentam correlações entre si. A Figura 2.2 ilustra o porquê de a correlação entre variáveis introduzir a necessidade da abordagem multivariada para o monitoramento. O contorno ao redor da curva de pontos no espaço bidimensional consiste da região de confiança bivariada induzida pela correlação. Se novos dados de processo estão fora desse contorno, pode-se concluir que a correlação entre as variáveis foi quebrada, o que corresponde a um comportamento anormal. Na Figura 2.2 percebem-se regiões que se encontrariam dentro dos limites de controle univariados, mas fora do contorno bivariado, evidenciando a limitação do monitoramento univariado.

Considera-se a partir daqui o caso genérico (multivariado). Se as observações $\mathbf{x} \in \mathbb{R}^m$ podem ser descritas por uma distribuição de probabilidades, regiões de confiança são equivalentes a contornos da distribuição para valores de probabilidade fixos. Em particular, se a distribuição de probabilidades é a distribuição normal:

$$p(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^m |\boldsymbol{\Sigma}|}}, \quad (2.1)$$

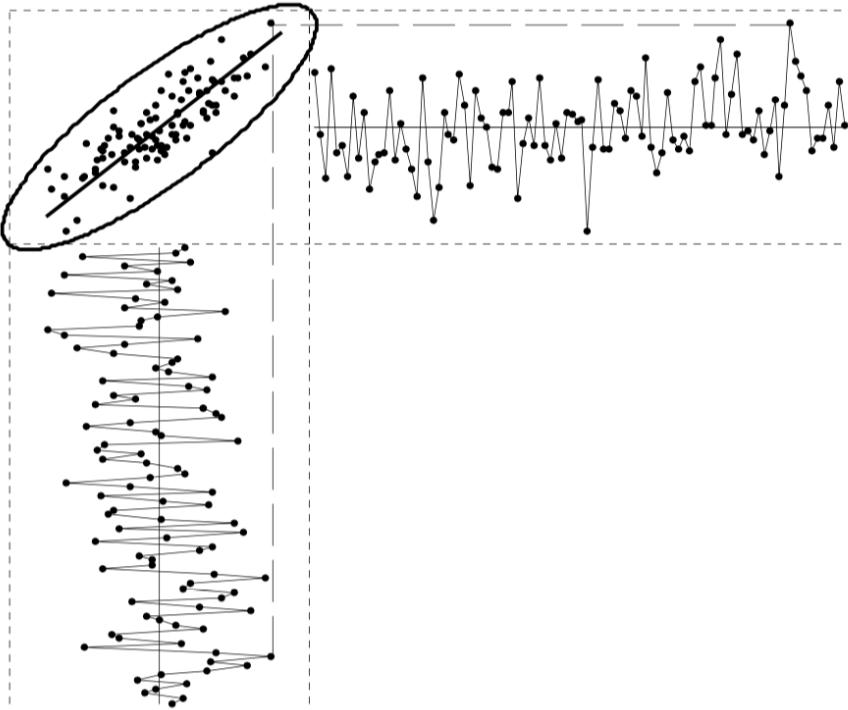


Figura 2.2: Região de confiança para duas variáveis correlacionadas, ilustrando a necessidade do monitoramento multivariado (FEITAL, 2011).

as regiões de confiança são hiperelipses, cujas equações têm a forma:

$$T^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}). \quad (2.2)$$

Nas Equações 2.1 e 2.2, $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são o vetor de médias e a matriz de covariâncias da distribuição, respectivamente.

A elipse representada por T^2 é comumente chamada e interpretada de duas maneiras (MASON e YOUNG, 2002):

- distância de Mahalanobis, que define os lugares geométricos dos pontos *estatisticamente equidistantes* da média $\boldsymbol{\mu}$ (MAHALANOBIS, 1936);
- estatística T^2 de Hotelling, que é uma espécie de generalização da estatística χ^2 (HOTELLING, 1947).

A estatística T^2 pode ser utilizada para a determinação de regiões de confiança multivariadas para fins de monitoramento. No entanto, para o cálculo é necessária a inversão da matriz de covariâncias $\boldsymbol{\Sigma}$, o que é difícil quando há alta correlação entre as variáveis, tornando-se impossível no limite em que a correlação é igual (ou próxima) a 1, por conta da consequente singularidade de $\boldsymbol{\Sigma}$. A solução mais utilizada para lidar com esse problema na indústria de processos é aplicar transformações nos

dados que eliminam a correlação, por meio de técnicas de modelagem em variáveis latentes.

2.4 Modelagem em variáveis latentes

A modelagem em variáveis latentes consiste em efetuar uma transformação nos dados, de modo a expressá-los como um novo conjunto de variáveis que explicitam características que se encontram ocultas (latentes) nos dados originais. A depender de que características se deseja explicitar, várias técnicas podem ser propostas, dentre as quais vale destacar (BLEI, 2014; BORGA *et al.*, 1992; CHIANG *et al.*, 2001; RENCHER e CHRISTENSEN, 2012):

- **Análise em componentes principais** (*principal component analysis, PCA*): gera-se um conjunto ortogonal de combinações lineares das variáveis originais, com o objetivo de selecionar um subconjunto dessas combinações que summariza apropriadamente a variabilidade dos dados.
- **Mínimos quadrados parciais** (*partial least squares, PLS*)/**Análise de correlações canônicas** (*canonical correlation analysis, CCA*): quando as variáveis estão divididas *a priori* em duas categorias (a porção preditora e a porção predita), geram-se dois conjuntos ortogonais de combinações lineares das variáveis originais (um para cada categoria) com o objetivo de selecionar as combinações lineares que melhor summarizam a relação linear entre as duas categorias.

PCA, PLS e CCA são as técnicas baseadas em dados mais utilizadas em aplicações de monitoramento de processos industriais, pelo fato de essas técnicas gerarem variáveis latentes *ortogonais*, o que elimina a correlação entre os dados, característica recorrente em dados industriais (KOURTI, 2002).

A modelagem em variáveis latentes é tão predominante do ponto de vista histórico que vários trabalhos que se propõem a apresentar ou revisar a literatura de monitoramento baseada em dados tratam apenas desse tipo de metodologia (CHIANG *et al.*, 2001; KRUGER e XIE, 2012; MACGREGOR e CINAR, 2012; QIN, 2009, 2003, 2012; YIN *et al.*, 2014a). Pode-se então inferir duas possibilidades de visão tácita por parte desses autores: de que a área de monitoramento de processos inteira é constituída apenas por esses métodos ou de que esses são os únicos métodos dignos de atenção.

MACGREGOR e CINAR (2012) são os únicos que forneceram uma justificativa para essa preferência:

In this paper, we focus on the proper use of data-driven (empirical) models for the monitoring, control and optimization of processes. In particular, we focus on latent variable models because, as we will show, they provide the proper structure to allow them to be built from plant data and be used for monitoring, control and optimization.

Ao longo do artigo, os autores justificam a escolha argumentando que métodos de modelagem em variáveis latentes são os únicos que modelam tanto a porção preditora (**X**) quanto a porção predita (**Y**) de um conjunto de dados; todos os outros métodos de regressão, segundo os autores, modelariam apenas o espaço **Y**, o que equivaleria a assumir que o espaço **X** apresenta posto completo, hipótese raramente concretizada em um cenário real. Por conta disso, os métodos de variáveis latentes seriam os mais adequados para lidar com problemas de dados faltantes e seriam os únicos métodos de monitoramento baseados em dados a proporcionarem modelos únicos, interpretáveis e causais.

Várias objeções podem ser feitas a esse ponto de vista. A mudança para o espaço de variáveis latentes não é a única maneira de lidar com a colinearidade na matriz **X**. A própria não linearidade de alguns métodos de regressão pode diminuir ou eliminar esse problema e, mesmo no contexto de métodos lineares, existem outras estratégias para lidar com a colinearidade, como a seleção de subconjuntos e a aplicação de métodos de penalização como *ridge regression* e *Lasso* (HASTIE *et al.*, 2009). Técnicas de reconciliação de dados também podem ser usadas, já que permitem tratar de flutuações e incertezas tanto de **X** quanto de **Y** (ANDRADE *et al.*, 2022; CÂMARA *et al.*, 2017; MENEZES *et al.*, 2020; PRATA *et al.*, 2009). Além do mais, não parece haver na literatura qualquer trabalho publicado pelos autores em que essa forte argumentação seja comprovada com resultados concretos. Pode-se perguntar se a argumentação não reflete uma afeição pessoal dos autores pelas técnicas, resultante da experiência de décadas de trabalho bem-sucedido com sua aplicação.

Como a modelagem em variáveis latentes é uma metodologia originada do domínio da estatística multivariada, muitos autores da área de monitoramento de processos se referem a elas com termos como “*multivariate statistical analysis*”, “*statistical based approaches*”, dentre outras (DAS *et al.*, 2012; MD NOR *et al.*, 2019). Apesar de fazerem sentido do ponto de vista histórico, atualmente esses termos podem causar confusão, já que o outro grande conjunto de técnicas de monitoramento

(baseadas no aprendizado de máquina) também podem ser multivariadas e ter base estatística⁵. Por esse motivo, este trabalho dará preferência ao termo “modelagem em variáveis latentes”, com exceção de contextos em que se apresentem referências a trabalhos da literatura.

A seguir, as principais técnicas de modelagem em variáveis latentes são descritas do ponto de vista teórico e histórico, focando nos desenvolvimentos relacionados ao monitoramento de processos. Uma novidade na presente descrição, em relação à maioria das referências da literatura, diz respeito à apresentação focada nas hipóteses que cada técnica traz em relação à natureza dos dados. Essas hipóteses são fundamentais para caracterizar o uso das técnicas, embora sejam frequentemente negligenciadas pela maior parte dos autores e dos usuários.

2.4.1 Análise em componentes principais (PCA)

PCA – Apresentação da técnica

O PCA foi proposto originalmente por PEARSON (1901), sendo que a primeira aplicação na área de monitoramento de processos foi proposta apenas muito mais tarde por WISE *et al.* (1988). A técnica efetua uma transformação linear no conjunto de dados a ser analisado, de modo a gerar um conjunto ortogonal de combinações lineares das variáveis originais, cujas direções são chamadas de componentes principais (SHLENS, 2014). A depender da natureza dos dados, a expressão em termos de componentes principais pode introduzir uma série de vantagens em relação à forma original, que são descritas adiante.

O fato da técnica de PCA propor e aplicar uma transformação linear nos dados introduz a primeira hipótese da técnica:

PCA, Hipótese 1: *As variáveis latentes podem ser obtidas por meio da aplicação de transformações lineares no conjunto de dados.*

Seja um conjunto de dados $\mathbf{X} \in \mathbb{R}^{n \times m}$, em que cada linha corresponde a uma observação e cada coluna corresponde a uma variável medida. A transformação linear

⁵Outra ambiguidade semântica pode advir do fato de que certos modelos de aprendizado de máquina, como os autocodificadores, podem ser também interpretados como modelos de variáveis latentes (KONG *et al.*, 2022).

efetuada pelo PCA pode ser representada por meio de uma mudança de base⁶:

$$\mathbf{T} = \mathbf{XP}, \quad (2.3)$$

em que as colunas de $\mathbf{P} \in \mathbb{R}^{m \times m}$, $\{\mathbf{p}_1, \dots, \mathbf{p}_m\}$, denominadas de componentes principais, representam uma nova base para expressar cada observação (linha) da matriz \mathbf{X} . As colunas de $\mathbf{T} \in \mathbb{R}^{n \times m}$ são as variáveis latentes e constituem projeções das colunas de \mathbf{X} na base $\{\mathbf{p}_1, \dots, \mathbf{p}_m\}$.

Para escolher a nova base \mathbf{P} , é preciso avaliar quais características da matriz transformada \mathbf{T} levam à melhor representação dos dados. Essa avaliação só é possível se são enunciadas mais hipóteses acerca da natureza dos dados:

PCA, Hipótese 2: *Há um grau de dependência linear entre as variáveis, o que atrapalha a descrição da essência dos dados por conta da presença de informação redundante e desnecessária.*

PCA, Hipótese 3: *Os dados possuem alta razão sinal/ruído⁷, o que significa que as variâncias maiores apresentam maior importância relativa da informação.*

A hipótese PCA-H2 indica que é desejável que as colunas de \mathbf{T} sejam linearmente independentes. Em outras palavras, a matriz de covariâncias empírica $\mathbf{C}_{\mathbf{T}}$:

$$\mathbf{C}_{\mathbf{T}} = \frac{(\mathbf{T} - \bar{\mathbf{T}})'(\mathbf{T} - \bar{\mathbf{T}})}{n - 1}, \quad (2.4)$$

deve conter apenas termos diagonais. Na Equação 2.4, $\bar{\mathbf{T}}$ é o vetor de médias das colunas de \mathbf{T} . Antes de prosseguir, é necessário para o desenvolvimento matemático enunciar mais uma hipótese⁸:

PCA, Hipótese 4: *Cada variável da matriz de dados apresenta média nula: $\bar{\mathbf{X}} = \bar{\mathbf{T}} = 0$.*

Nesse caso, a matriz de covariâncias de \mathbf{T} toma a forma:

$$\mathbf{C}_{\mathbf{T}} = \frac{\mathbf{T}'\mathbf{T}}{n - 1}. \quad (2.5)$$

⁶Se a matriz de dados \mathbf{X} estivesse arranjada no formato $m \times n$, a transformação linear poderia ser representada na intuitiva forma de um operador matricial \mathbf{P} atuando sobre os dados: $\mathbf{T} = \mathbf{PX}$, em que $T \in \mathbb{R}^{m \times n}$, no caso. Para mais detalhes, conferir SHLENS (2014).

⁷Na área de engenharia de sistemas em processos, o termo “sinal” é comumente utilizado para descrever informações de natureza dinâmica, caracterizadas por variações mais lentas e maior variância ao longo de períodos grandes de tempo.

⁸Na prática, força-se a validade dessa hipótese por meio de um processo conhecido como *centralização*, em que se subtrai a média de cada variável.

Fazendo uso da Equação 2.3, a Equação 2.5 é rescrita em termos da matriz de componentes principais \mathbf{P} :

$$\mathbf{C}_T = \mathbf{P}' \mathbf{C}_X \mathbf{P}. \quad (2.6)$$

De acordo com o Teorema Espectral (IZENMAN, 2008), a matriz \mathbf{C}_X , por ser simétrica, pode ser decomposta em valores e vetores característicos, na forma:

$$\mathbf{C}_X = \mathbf{E} \Lambda \mathbf{E}', \quad (2.7)$$

em que \mathbf{E} é a matriz que contém nas colunas os vetores característicos de \mathbf{C}_X e Λ é uma matriz diagonal que contém os valores característicos de \mathbf{C}_X em ordem decrescente de magnitude.

Para tornar a matriz \mathbf{C}_T diagonal, uma escolha conveniente para a nova base é fazer $\mathbf{P} = \mathbf{E}$, já que, nesse caso, substituindo a Equação 2.7 na Equação 2.6, chega-se a:

$$\mathbf{C}_T = \Lambda. \quad (2.8)$$

Se válidas as hipóteses PCA-H2 e H3, essa escolha faz com que a matriz transformada \mathbf{T} esteja em uma forma muito conveniente do ponto de vista da análise de dados, já que (i) a redundância causada pela dependência linear entre as variáveis é eliminada e (ii) as variáveis estão organizadas em ordem decrescente de importância, quantificada pelos valores de variância presentes na matriz diagonal \mathbf{C}_T .

É conveniente interpretar o PCA também do ponto de vista de um problema de otimização. O primeiro componente principal \mathbf{p}_1 deve fazer com que a variável latente $\mathbf{t}_1 = \mathbf{X}\mathbf{p}_1$ apresente máxima variância:

$$\begin{aligned} \mathbf{p}_1 &= \arg \max_{\|\mathbf{p}\|_2=1} \left\{ (\mathbf{X}\mathbf{p})' \mathbf{X}\mathbf{p} \right\} \\ &= \arg \max_{\|\mathbf{p}\|_2=1} \left\{ \mathbf{p}' \mathbf{X}' \mathbf{X}\mathbf{p} \right\} \\ &= \arg \max_{\|\mathbf{p}\|_2=1} \left\{ \mathbf{p}' \mathbf{C}_X \mathbf{p} \right\} \\ &= \arg \max \left\{ \frac{\mathbf{p}' \mathbf{C}_X \mathbf{p}}{\mathbf{p}' \mathbf{p}} \right\}, \end{aligned} \quad (2.9)$$

em que na última igualdade, a restrição de norma unitária é relaxada quando se divide a função objetivo por $\|\mathbf{p}\|_2^2$. Os demais componentes principais podem ser procurados nos sucessivos espaços ortogonais, deflacionando a matriz \mathbf{C}_X com respeito

a \mathbf{p} (SHawe-Taylor e CRISTIANINI, 2004).

A função objetivo na Equação 2.9 assume uma forma conhecida na matemática, o *quociente de Rayleigh*:

$$r(\mathbf{A}, \mathbf{B}, \mathbf{w}) = \frac{\mathbf{w}' \mathbf{A} \mathbf{w}}{\mathbf{w}' \mathbf{B} \mathbf{w}}. \quad (2.10)$$

Pode-se mostrar que os pontos críticos de $r(\mathbf{A}, \mathbf{B}, \mathbf{w})$ correspondem à solução do *problema característico generalizado* (BORGÀ *et al.*, 1992; DE BIE *et al.*, 2005):

$$\mathbf{A}\mathbf{e} = \lambda \mathbf{B}\mathbf{e}. \quad (2.11)$$

Como na função objetivo em questão $\mathbf{B} = \mathbf{I}$, a Equação 2.11 se reduz a um problema característico⁹ ordinário, o que é consistente com a dedução da Equação 2.8. As demais técnicas de modelagem em variáveis latentes são descritas adiante nesta seção com base no problema característico generalizado definido na Equação 2.11.

PCA – Redução de dimensionalidade

Para fins de redução de dimensionalidade, é selecionada uma quantidade a de componentes principais¹⁰ que acumulam certa porcentagem da variância explicada (tipicamente, 90%), dando origem à matriz reduzida $\mathbf{P}_a \in \mathbb{R}^{m \times a}$. A matriz de variáveis latentes reduzida $\mathbf{T}_a \in \mathbb{R}^{n \times a}$ é obtida por:

$$\mathbf{T}_a = \mathbf{X}\mathbf{P}_a. \quad (2.12)$$

De acordo com a hipótese PCA-H3, as variáveis mais relevantes são aquelas que variam mais no espaço multivariado estudado. Portanto, a matriz \mathbf{T}_a seria uma representação mais eficiente e adequada dos dados, já que nela não estão presentes as variáveis associadas às menores variâncias (correspondentes, segundo a hipótese, aos ruídos de medição). Essa matriz pode ser projetada nas variáveis originais:

$$\hat{\mathbf{X}} = \mathbf{T}_a \mathbf{P}'_a = \mathbf{X}(\mathbf{P}_a \mathbf{P}'_a). \quad (2.13)$$

⁹Pode-se mostrar que a decomposição característica aplicada à matriz de covariâncias \mathbf{C}_x equivale à decomposição singular aplicada à matriz de dados \mathbf{X} (SHLENS, 2014). Na verdade, este último procedimento é mais eficiente e estável do ponto de vista computacional, por evitar o cálculo de \mathbf{C}_x .

¹⁰Para discussões sobre métodos de seleção do parâmetro a , consultar FADEL (2018); HIMES *et al.* (1994); QIN e DUNIA (2000); SACCENTI e CAMACHO (2015); TAMURA e TSUJITA (2007); VALLE *et al.* (1999); WU *et al.* (2022).

Dessa maneira, recuperam-se na matriz $\hat{\mathbf{X}}$ as variáveis originais, mas contendo apenas a variação dita principal do conjunto de dados, correspondente aos a primeiros componentes principais. Nesse sentido, pode-se encarar o PCA como um filtro, já que a transformação de \mathbf{X} para $\hat{\mathbf{X}}$, em tese, reduz o nível de ruído do conjunto de dados.

A diferença entre \mathbf{X} e $\hat{\mathbf{X}}$ é a *matriz residual* \mathbf{E} :

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{X}(\mathbf{I} - \mathbf{P}_a \mathbf{P}_a'). \quad (2.14)$$

Os subespaços gerados por $\hat{\mathbf{X}}$ e \mathbf{E} são conhecidos por *espaço principal* e *espaço residual*, respectivamente (CHIANG *et al.*, 2001).

PCA – Monitoramento

No problema de monitoramento, a matriz de dados \mathbf{X} a ser modelada é proveniente de uma *série temporal*. O objetivo da aplicação do PCA, no caso, é identificar a estrutura do espaço principal durante um período normal de operação e usá-la como referência para avaliação de novos pontos, de modo a verificar se a estrutura se mantém.

Quando o processo que gera a série temporal é não estacionária, ou seja, apresenta características dinâmicas e dependência com o tempo, o PCA pode não ser capaz de identificar a estrutura do espaço principal. As variáveis latentes, no caso, podem resultar autocorrelacionadas e possivelmente apresentar correlação cruzada (KU *et al.*, 1995). Sendo assim, para aplicar a forma básica do PCA em problemas de monitoramento, é preciso enunciar a seguinte hipótese:

PCA, Hipótese 5: *A matriz de dados é obtida a partir de um processo em estado estacionário.*

O procedimento de monitoramento utilizando PCA, nesse caso, pode ser efetuado em duas etapas. A primeira etapa, chamada comumente de *treino*, consiste em aplicar o PCA a um conjunto de dados de operação normal, o *conjunto de treino*. Os resultados do treino são as matrizes de projeção \mathbf{P}_a e de variâncias explicadas Λ_a . Na segunda etapa, chamada comumente de *teste*, novas observações \mathbf{x} são testadas por meio da análise de *estatísticas de monitoramento*. As duas estatísticas mais usadas são os índices T^2 e Q , descritos a seguir.

A estatística T^2 , no contexto do PCA, consiste da aplicação da Equação 2.2 à variável latente $\mathbf{t}_a = \mathbf{P}_a' \mathbf{x}$ (a projeção da observação \mathbf{x} nos primeiros a componentes

principais) (KETELAERE *et al.*, 2015; KRUGER e XIE, 2012):

$$T^2 = \mathbf{t}'_{\mathbf{a}} \boldsymbol{\Lambda}_{\mathbf{a}}^{-1} \mathbf{t}_{\mathbf{a}} = \mathbf{x}' \mathbf{P}_{\mathbf{a}} \boldsymbol{\Lambda}_{\mathbf{a}}^{-1} \mathbf{P}'_{\mathbf{a}} \cdot \mathbf{x} \quad (2.15)$$

A estatística T^2 indica uma medida global da variabilidade capturada pelo modelo PCA nos primeiros a componentes principais, associada às variações sistemáticas do processo (CHIANG *et al.*, 2001). Portanto, no monitoramento, o aumento no valor de T^2 em novas observações sugere a ocorrência de distúrbios de mesma natureza que as variações identificadas pelo modelo PCA, ou seja, distúrbios nas direções do plano descrito pelo modelo (KUNDU *et al.*, 2017).

A estatística Q é definida como a distância euclidiana da observação \mathbf{x} à respectiva projeção no espaço principal, $\hat{\mathbf{x}} = (\mathbf{P}_{\mathbf{a}} \mathbf{P}'_{\mathbf{a}}) \mathbf{x}$:

$$Q = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \|(\mathbf{I} - \mathbf{P}_{\mathbf{a}} \mathbf{P}'_{\mathbf{a}}) \mathbf{x}\|^2 = (\mathbf{x} - \hat{\mathbf{x}})' (\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{x}' (\mathbf{I} - \mathbf{P}_{\mathbf{a}} \mathbf{P}'_{\mathbf{a}}) \mathbf{x}. \quad (2.16)$$

A estatística Q consiste do quadrado dos módulos dos erros de predição (resíduos) do modelo PCA, sendo por isso também conhecida como SPE (*Square Prediction Error*). Essa estatística indica uma medida global da variabilidade *não* capturada pelo modelo PCA (variabilidade presente no espaço residual), associada às variações não usuais ou não explicadas do processo (CHIANG *et al.*, 2001). Portanto, no monitoramento, o aumento no valor de Q em novas observações sugere uma mudança nas relações entre as variáveis (por exemplo, violação nas correlações), o que resulta na quebra da estrutura capturada pelo modelo durante o treino. Em outras palavras, o plano descrito pelo modelo PCA não inclui a direção do distúrbio (KUNDU *et al.*, 2017). Por conta da forma da Equação 2.16, estabelecer um limite de detecção utilizando a estatística Q equivale a delimitar uma hiper-esfera como região de confiança no espaço residual (LI e QIN, 2016).

O uso da estatística T^2 , derivada da distribuição normal (Equação 2.1), impõe a enunciação de mais uma hipótese:

PCA, Hipótese 6: *O conjunto de dados \mathbf{X} segue uma distribuição normal.*

Em alguns casos, a estatística T^2 pode ser usada mesmo em situações que violem a hipótese PCA-H6, por conta de as variáveis latentes serem combinações lineares das variáveis originais; a hipótese, no caso, é de que sigam aproximadamente uma distribuição normal devido ao Teorema do Limite Central (FUENTES-GARCÍA *et al.*, 2018), que estabelece que a soma de variáveis aleatórias se aproxima assintoticamente da representação normal, à medida que o número de variáveis aumenta.

Algumas referências consideram a normalidade como requisito intrínseco para a aplicação da técnica PCA e não como referente à utilização da estatística T^2 para o monitoramento, como considerado no presente texto. Isso ocorre porque essas referências aplicam e interpretam o PCA no contexto da (in)dependência estatística¹¹, enquanto no presente texto a discussão envolveu o conceito menos rigoroso de (in)dependência linear (vide hipótese PCA-H2). Para garantir que a independência linear entre as variáveis latentes do PCA equivalha à independência estatística, é preciso enunciar alguma hipótese sobre a distribuição de probabilidades dos dados. Pode-se mostrar que admitir que os dados seguem uma distribuição normal é condição suficiente para garantir essa equivalência (MILMAN, 2018).

Essa questão é sutil e frequentemente fonte de confusões conceituais. Jonathon Shlens, por exemplo, incluía a hipótese da normalidade nas primeiras versões de seu excelente tutorial sobre PCA (SHLENS, 2005), mas a retirou nas versões mais recentes (SHLENS, 2014). Como suporte à compreensão do assunto, vale a pena a consulta a uma comparação sucinta e elucidativa entre os termos *independência linear*, *ortogonalidade* e *não correlação* apresentada por RODGERS *et al.* (1984). Para as diferenças entre os conceitos de *dependência estatística* e *linear*, pode-se consultar SCHWAAB e PINTO (2007).

As estatísticas T^2 e Q do PCA podem ser usadas para as duas principais atividades tradicionalmente associadas ao monitoramento de processos: detecção e diagnóstico de falhas.

PCA – Detecção de falhas

De modo a efetuar a detecção de falhas, limites de controle devem ser determinados para as estatísticas T^2 e Q . Para um nível de confiança α , os limites mais usados são (KOURTI, 2019; KUNDU *et al.*, 2017):

$$T_{\alpha}^2 = \frac{a(n^2 - 1)}{n(n - a)} F_{\alpha(a, n-a)}, \quad (2.17)$$

$$Q_{\alpha} = \theta_1 \left[\frac{h_0 c_{\alpha} \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0}, \quad (2.18)$$

sendo :

¹¹Duas variáveis \mathbf{x}_1 e \mathbf{x}_2 são ditas estatisticamente independentes se a densidade de probabilidade conjunta $p(\mathbf{x}_1, \mathbf{x}_2) = p(\mathbf{x}_1)p(\mathbf{x}_2)$.

- $F_{\alpha(a,n-a)}$ o limite superior do percentil α da distribuição F de Fisher-Snedecor (SCHWAAB e PINTO, 2007), com graus de liberdade a e $n - a$;
- $\theta_i = \sum_{j=a+1}^m \lambda_j^i$;
- $h_0 = 1 - \left(\frac{2\theta_1\theta_3}{3\theta_2^2} \right)$;
- $c_\alpha = 1,645$ é o desvio normal correspondente ao percentil $1 - \alpha$.

A Equação 2.17 decorre diretamente da Teoria Estatística clássica. A Equação 2.18 é baseada na aproximação da distribuição de formas quadráticas, tendo sido primeiramente proposta por JACKSON e MUDHOLKAR (1979). Quando as hipóteses de normalidade assumidas não são válidas, as Equações 2.17 e 2.18 podem não ser bem sucedidas, e limites de controle baseados nas características estatísticas particulares de cada conjunto de dados devem ser propostos.

PCA – Identificação e Diagnóstico de falhas

Há certa confusão na literatura no que se refere às definições dos termos “identificação” e “diagnóstico” de falhas. CHIANG *et al.* (2001), por exemplo, definiram identificação como a indicação das variáveis mais relevantes para o diagnóstico, definido como a determinação da causa do estado anormal. Já QIN (2012) propôs uma definição em que os significados dos termos estão invertidos: a identificação, no caso, se refere à identificação de uma falha dentre um conjunto de possíveis falhas conhecidas, enquanto o diagnóstico refere-se à identificação das variáveis que contribuíram para a ocorrência da falha. No presente trabalho, a definição de QIN (2012) será adotada.

Há vários métodos de diagnóstico no contexto do PCA, dentre os quais os mais populares são os métodos de contribuições (ALCALA e QIN, 2011, 2009; CHIANG *et al.*, 2001; FUENTES-GARCÍA *et al.*, 2018; VIDAL-PUIG *et al.*, 2019). Esses métodos têm como base a análise da contribuição de cada variável para o índice de detecção. Para que sejam aplicados, é preciso enunciar mais uma hipótese:

PCA, Hipótese 7: *No momento da falha, a magnitude da contribuição de uma variável para o aumento do índice de detecção é proporcional à influência da variável na ocorrência do evento anormal.*

Os métodos de contribuições podem ser divididos em três grandes classes (ALCALA e QIN, 2011):

- **contribuições decompositivas:** decompõem o índice de detecção como a soma das contribuições das variáveis;
- **contribuições baseadas na reconstrução:** consideram a contribuição da variável como a magnitude da reconstrução do índice de detecção ao longo da direção dessa variável;
- **contribuições diagonais:** em problemas multibloco, variáveis semelhantes são agrupadas em blocos utilizando conhecimento de domínio para simplificar cenários com muitas variáveis. Contribuições diagonais mantêm apenas os termos de bloco no índice de detecção.

A título de ilustração, será apresentado o equacionamento da técnica mais disseminada, o método de contribuições decompositivas. Os índices T^2 e Q podem ser escritos na forma de um índice generalizado, idx :

$$idx = \mathbf{x}' \mathbf{M} \mathbf{x}, \quad (2.19)$$

sendo:

- $\mathbf{M} = \mathbf{P}_a \Lambda_a^{-1} \mathbf{P}'_a$ para $idx = T^2$;
- $\mathbf{M} = \mathbf{I} - \mathbf{P}_a \mathbf{P}'_a$ para $idx = Q$.

A contribuição decompositiva da variável i , cd_i , para o índice generalizado é escrita como (ALCALA e QIN, 2011):

$$cd_i = \mathbf{x}' \mathbf{M}^{1-\beta} \mathbf{I}_i \mathbf{I}'_i \mathbf{M}^\beta \mathbf{x}, \quad (2.20)$$

sendo \mathbf{I}_i a i -ésima coluna da matriz identidade e $0 \leq \beta \leq 1$. De acordo com a escolha do parâmetro β , diversos graus de decomposição são definidos. Na literatura, tradicionalmente utiliza-se $\beta = 0$ ou $\beta = 1$ sob o nome de “contribuição parcial” e $\beta = 1/2$ sob o nome de “contribuição completa”.

Os métodos de contribuição mencionados, apesar do disseminado uso, podem sofrer de um problema denominado na literatura como *smearing-out effect* (efeito de espalhamento, em tradução livre). Nessa situação, primeiramente demonstrada por WESTERHUIS *et al.* (2000), contribuições de variáveis afetadas pela falha aumentam as contribuições de variáveis não afetadas. VAN DEN KERKHOF *et al.* (2013) reportaram uma análise detalhada desse efeito, concluindo que ele é mais pronunciado ao longo dos componentes principais. Os autores recomendaram a aplicação de métodos de contribuição apenas no caso de haver uma interpretação

física clara dos componentes principais; caso não haja essa interpretação, sugeriram como alternativa para diagnóstico a utilização do desvio de cada variável em relação ao valor esperado (o equivalente a um diagnóstico univariado). ALAKENT (2023) propôs uma metodologia para reduzir o efeito de espalhamento por meio da aplicação de um método de redução de dimensionalidade não linear.

PCA – Histórico e tendências

Das técnicas de monitoramento multivariado, o PCA é a mais estudada e aplicada em sistemas simulados ou reais. Porém, em sua forma básica, como descrita nas páginas anteriores, ela não se adequa a uma série de situações relevantes do ponto de vista industrial. Isso levou a propostas de aprimoramento, sumarizadas nos pontos a seguir.

- **Processos em batelada:** desde a popularização da aplicação do PCA, no início da década de 1990, a adaptação para processos em batelada existe, por meio da aplicação do *multi-way* PCA (KOSANOVICH *et al.*, 1994; NOMIKOS e MACGREGOR, 1994). Nessa técnica, transforma-se a matriz de dados de dimensão $n \times m \times b$ (em que a terceira dimensão b diz respeito às diversas bateladas) em uma matriz bidimensional. Essa redução de dimensões pode ocorrer tanto por tempo quanto por batelada (QIN, 2012). Para comparação com outras estratégias usadas para monitoramento de processos em batelada, como a análise em fatores paralelos e decomposição trilinear, recomenda-se a consulta aos artigos de comparação publicados por CAMACHO *et al.* (2009); WESTERHUIS *et al.* (1999); WISE *et al.* (1999) e aos artigos de revisão publicados por RENDALL *et al.* (2019); YAO e GAO (2009). Uma técnica que parece interessante, mas não foi citada nessas revisões, é a 2D-PCA (LU *et al.*, 2005; ZHU *et al.*, 2020a), em que se monitoram separadamente as dinâmicas relacionadas às variáveis e às bateladas.
- **Dinâmica:** quando o processo analisado viola a Hipótese PCA-H5, não se encontrando em estado estacionário e apresentando características dinâmicas (como dependência temporal e autocorrelação), as regiões de confiança provenientes das estatísticas de monitoramento do PCA resultam muito largas (QIN, 2012). Isso pode levar a altas taxas de falhas não detectadas. A estratégia mais adotada para lidar com esse problema é a expansão da matriz de dados com os valores das variáveis originais defasados no tempo, como proposto por KU *et al.* (1995), o que equivale a ajustar um modelo auto-regressivo aos dados. A principal vantagem dessa abordagem é a simplicidade. As desvantagens são

(DONG e QIN, 2018b): (i) aumento no número de parâmetros e componentes principais, eliminando umas das principais vantagens do PCA, que é a redução de dimensionalidade; (ii) falta de uma representação explícita da relação entre as variáveis latentes e as originais; (iii) falta de relação entre as variáveis latentes e o conteúdo dinâmico dos dados. Uma tendência recente para reduzir essas desvantagens são os métodos que criam componentes principais explicitamente dinâmicos (DONG e QIN, 2018b; LI *et al.*, 2011, 2014). Em especial, o método DiPCA (*Dynamic-inner PCA*), proposto por DONG e QIN (2018a,b), é particularmente interessante por conta da separação das variáveis latentes em dois grupos: um conjunto de séries temporais principais e um resíduo com pouca ou nenhuma autocovariância, que pode ser tratado como um conjunto de dados estáticos por métodos tradicionais (como o próprio PCA). Ainda em relação ao DiPCA, vale ressaltar os trabalhos de GUO *et al.* (2020), que o utilizaram para montar uma estrutura sequencial para extração e modelagem de características dinâmicas e não lineares dos dados, e LU e HUANG (2020), que o aplicaram no problema de segmentação de séries temporais, com o objetivo de gerar subsequências de acordo com mudanças na dinâmica. Outra estratégia para lidar com a dinâmica, especialmente adequada para aplicação em processos com *drifts* ou mudanças de pontos de operação, é a utilização de algoritmos adaptativos, que se ajustam automaticamente a mudanças nas estruturas de média, variância e/ou correlação nos dados (QIN, 2012). Nesse contexto, RPCA (PCA recursivo), MW-PCA (*moving window PCA*) e EWMA-PCA (*exponentially weighted moving average PCA*) são três exemplos bastante aplicados (GALLAGHER *et al.*, 1997; JENG, 2010; KETELAERE *et al.*, 2015; LI *et al.*, 2000; WOLD, 1994). Uma revisão sobre avanços recentes de métodos para monitoramento de processos dinâmicos foi publicada por ALDRICH (2019).

- **Não linearidade:** há duas grandes classes de estratégias tradicionalmente utilizadas para incorporar não linearidade ao PCA. A primeira é o uso de redes neurais, como proposto por (i) KRAMER (1991), que se baseia em redes associativas; (ii) DONG e MCAVOY (1996), que combina redes ao uso do algoritmo de curvas principais (HASTIE e STUETZLE, 1989); (iii) JIA *et al.* (2000), que aplica ao monitoramento uma rede chamada IT-net (TAN e MAYROVOUNIOTIS, 1995); (iv) LIU *et al.* (2011), que combina redes à técnica SVDD (*support vector data description*), um método de aprendizado de máquina descrito na Seção 2.5.4. A segunda estratégia é o uso da técnica *kernel*-PCA (KPCA)¹², desenvolvida por SCHOLKOPF *et al.* (1998), aplicada na área de monitoramento por LEE *et al.* (2004a) e CHOI *et al.* (2005) e

¹²Uma explicação do conceito de *kernel* é apresentada na Seção 2.5.4, no contexto da aplicação a métodos do tipo máquina de vetores suporte.

posteriormente desenvolvida, por exemplo, por BERNAL-DE LÁZARO *et al.* (2022); DENG *et al.* (2013, 2018a,b); JIANG e YAN (2014, 2018); WANG *et al.* (2022a); YAO e WANG (2015). Além dessas duas estratégias principais, há outras estratégias para adaptação do PCA para não linearidade, como por exemplo o uso do algoritmo genético (HIDEN, 1997). Revisões sobre aplicações da técnica *kernel* em monitoramento de processos foram publicadas recentemente por PILARIO *et al.* (2019), APSEMIDIS *et al.* (2020) e PANI (2022).

- **Multimodalidade:** para lidar com a multimodalidade, podem ser utilizados algoritmos de agrupamento (ou *clustering*, como é mais comumente referenciado) para discriminar os modos de operação (QUIÑONES-GRUEIRO *et al.*, 2019). A estratégia mais simples é a identificação dos grupos *a priori* por meio de conhecimento prévio sobre o processo ou com algoritmos especializados, seguida da construção de modelos PCA individuais para cada grupo (QIN, 2012). Outra estratégia popular é o uso de modelos de mistura gaussianos, como proposto por CHEN e ZHANG (2010); CHEN *et al.* (2006); CHOI *et al.* (2004); FEITAL *et al.* (2013); LYU *et al.* (2023), que levam à identificação automática dos grupos na etapa de treinamento do PCA. Uma ampla revisão sobre monitoramento de processos multimodais foi publicada por QUIÑONES-GRUEIRO *et al.* (2019).
- **Múltiplas escalas de tempo:** BAKSHI *et al.* (1994) foi o primeiro a propor a análise de dados de processo em diferentes escalas de tempo, utilizando uma metodologia de aplicação de *wavelets* para o processamento de dados, descrita por RIOUL e VETTERLI (1991). KOSANOVICH e PIOVOSO (1997); KOSANOVICH (1997) utilizaram essa abordagem para pré-processar dados antes da aplicação do monitoramento com PCA. BAKSHI (1998) integrou PCA e *wavelets* em um só método, chamado de MSPCA (*multiscale PCA*), que funciona gerando um modelo PCA para cada uma das escalas de tempo de interesse e depois juntando os modelos para criação de um único modelo multiescala. Desde esse trabalho, vários estudos e aplicações foram reportados, utilizando ou adaptando essa abordagem. Por exemplo, YOON e MACGREGOR (2001) e MISRA *et al.* (2002) estenderam a metodologia para o problema de diagnóstico; ARADHYE *et al.* (2003) apresentaram uma análise teórica da metodologia; YOON e MACGREGOR (2004) utilizaram a metodologia para geração de gráficos de contribuição; REIS e SARAIVA (2006) analisaram dados com características de multi-resolução; e ZHANG e MA (2011) juntaram a abordagem multiescala com a técnica de *kernel* para modelagem da não linearidade. Trabalhos recentes incluem a aplicação em processos de múltiplas unidades e em batelada (ALAWI *et al.*, 2015), a integração com o teste de razão

de verossimilhança generalizada (SHERIFF *et al.*, 2017) e o uso em conjunto com técnicas de *kernel* para se lidar com não linearidades (NAWAZ *et al.*, 2020). Amplas revisões sobre a aplicação de métodos multiescala em análise de dados de processo foram publicadas por REIS (2019) e NAWAZ *et al.* (2022).

- **Múltiplas unidades e/ou estágios:** para lidar com essa questão de maneira descentralizada, as estratégias mais usadas são os métodos multibloco, como CPCCA (*consensus PCA*) e HPCA (*hierarchical PCA*) (QIN, 2012). Nesses métodos, a matriz de dados é dividida em vários blocos, com o objetivo de obter informação local e global simultaneamente. Um importante marco foi o trabalho de QIN *et al.* (2001), em que se demonstrou a equivalência algorítmica entre os métodos regulares e multibloco. Uma tendência recente é o desenvolvimento de esquemas totalmente baseados em dados para particionamento das unidades/estágios (HUANG e YAN, 2018). Trabalhos a destacar incluem os de GE e SONG (2013a), em que se propôs a construção de sub-blocos ao longo dos componentes principais; JIANG e YAN (2015), em que se combinou agrupamento espectral baseado em informação mútua, inferência bayesiana e KPCA para lidar com a não linearidade; HUANG e YAN (2018), em que se combinou KPCA com SVDD, resultando em um método capaz de determinar as variáveis independentes do processo; e CAO *et al.* (2021), em que se propôs um método híbrido para categorização de variáveis. Duas amplas revisões sobre monitoramento distribuído/de toda a planta foram publicadas por GE (2017) e JIANG *et al.* (2019).
- **Tratamento dos erros:** no caso dos erros de processo não serem independentes e identicamente distribuídos, é útil integrar a técnica com informações de densidade de probabilidade das variáveis para uma modelagem mais efetiva do espaço residual (FEITAL, 2011; MONTEIRO *et al.*, 2017). Abordagens para esse problema incluem o PCA probabilístico (TIPPING e BISHOP, 1999), PCA bayesiano (GE e SONG, 2010) e MLPCA (*maximum likelihood PCA*) (CHOI *et al.*, 2005; FEITAL *et al.*, 2010; NARASIMHAN e SHAH, 2008; WENTZELL e LOHNES, 1999; WENTZELL *et al.*, 1997). Vale destacar que o tratamento de dados heterocedásticos (com variância dependente do tempo ou das condições de operação) pode ser efetuado com MLPCA (FEITAL, 2011) ou regressão em componentes principais (MONTEIRO *et al.*, 2021). Uma proposta interessante nesse sentido é a unificação da metodologia PCA com a de reconciliação de dados, como reportado por NARASIMHAN e BHATT (2015).
- **Baixa qualidade dos dados:** para lidar com problemas na qualidade dos dados, tipicamente relacionados a observações faltantes ou espúrias, uma classe de técnicas denominadas de PCA robusto foi desenvolvida. Trabalhos

pioneiros incluem WALCZAK e MASSART (1995); XIE *et al.* (1993) e algumas aplicações/extensões representativas podem ser encontradas nas aplicações de JIN *et al.* (2006); LEE e LEE (2015); THARRAULT *et al.* (2008); WANG e ROMAGNOLI (2005); XIAO *et al.* (2019); ZHENG *et al.* (2020); ZHONG *et al.* (2022). Um fato curioso é a disseminação da aplicação do PCA robusto nas áreas de visão computacional, processamento de imagens e monitoramento por vídeo (BOUWMANS e ZAHZAH, 2014). Uma revisão sobre a aplicação do PCA a dados de processo com observações faltantes foi publicada por SEVERSON *et al.* (2017). Uma discussão interessante, em que se compara o PCA a outros métodos robustos, sob a perspectiva *big data*, foi publicada por ZHU *et al.* (2018). Para uma revisão em forma de tutorial sobre a aplicação de modelagem probabilística em variáveis latentes para a análise de dados de processo, consultar GE (2018).

2.4.2 Mínimos quadrados parciais (PLS) e análise de correlações canônicas (CCA)

PLS e CCA são duas técnicas que, em linhas gerais, compartilham o mesmo objetivo: gerar dois conjuntos ortogonais de combinações lineares das variáveis originais (um para cada categoria de variáveis, preditoras e preditas, definidas *a priori*), com o objetivo de selecionar as combinações lineares que melhor sumarizam a relação linear entre as duas categorias (BORGÀ *et al.*, 1992; DE BIE *et al.*, 2005; SHAWE-TAYLOR e CRISTIANINI, 2004). O CCA foi primeiramente proposto por HOTELLING (1936) e o PLS foi originalmente proposto por WOLD (1966). O PLS é amplamente utilizado na quimiometria, sendo essa a provável razão da predominância nos trabalhos da área de monitoramento de processos¹³. A preferência pelo PLS em relação ao CCA na quimiometria, no entanto, vem sendo contestada, como exemplificado por GATIUS *et al.* (2017).

As seguintes hipóteses acerca da natureza dos dados são compartilhadas por ambas as técnicas:

PLS/CCA, Hipótese 1: *As variáveis latentes podem ser obtidas por meio da aplicação de transformação linear no conjunto de dados.*

PLS/CCA, Hipótese 2: *O conjunto de dados pode ser dividido a priori em duas classes, o conjunto preditor **X** e o conjunto predito **Y**.*

¹³O PLS também é popular em outras áreas, sendo por exemplo o modelo baseado em dados mais utilizado no setor biofarmacêutico (BANNER *et al.*, 2021).

PLS/CCA, Hipótese 3: *Em cada um dos conjuntos \mathbf{X} e \mathbf{Y} , há um grau de dependência linear entre as variáveis, o que atrapalha a descrição da essência dos dados por conta da presença de informação redundante e desnecessária.*

PLS/CCA, Hipótese 4: *Há uma relação linear latente entre o conjunto \mathbf{X} e o conjunto \mathbf{Y} , que é de interesse que seja explicitada.*

A diferença entre o PLS e o CCA está na forma de obter a relação linear latente citada na hipótese PLS/CCA-H4: enquanto o PLS busca as direções que maximizam a covariância entre \mathbf{X} e \mathbf{Y} , o CCA busca as direções que maximizam a correlação entre \mathbf{X} e \mathbf{Y} (BORGA *et al.*, 1992):

PLS, Hipótese 5: *A relação linear latente da hipótese 4 é melhor representada nas direções que maximizam a covariância entre \mathbf{X} e \mathbf{Y} .*

CCA, Hipótese 5: *A relação linear latente da hipótese 4 é melhor representada nas direções que maximizam a correlação entre \mathbf{X} e \mathbf{Y} .*

Por último, mais uma hipótese compartilhada, necessária para o desenvolvimento matemático das técnicas:

PLS/CCA, Hipótese 6: *Cada variável das matrizes de dados apresenta média nula: $\bar{\mathbf{X}} = \bar{\mathbf{Y}} = 0$.*

Assim como o PCA, os problemas PLS e CCA podem ser postos na forma da maximização do quociente de Rayleigh.

No caso PLS:

$$\mathbf{p}_{1,x}, \mathbf{p}_{1,y} = \arg \max_{\|\mathbf{p}\|_2=1} \left\{ \frac{\mathbf{p}_x^T \mathbf{C}_{xy} \mathbf{p}_y}{\sqrt{\mathbf{p}_x^T \mathbf{p}_x} \sqrt{\mathbf{p}_y^T \mathbf{p}_y}} \right\}. \quad (2.21)$$

No caso CCA:

$$\mathbf{p}_{1,x}, \mathbf{p}_{1,y} = \arg \max_{\|\mathbf{p}\|_2=1} \left\{ \frac{\mathbf{p}_x^T \mathbf{C}_{xy} \mathbf{p}_y}{\sqrt{\mathbf{p}_x^T \mathbf{C}_x \mathbf{p}_x} \sqrt{\mathbf{p}_y^T \mathbf{C}_y \mathbf{p}_y}} \right\}. \quad (2.22)$$

Isso permite que os problemas sejam resolvidos utilizando o contexto do problema característico generalizado, Equação 2.11 (DE BIE *et al.*, 2005). As estatísticas de monitoramento T^2 e Q são elaboradas de forma análoga às do PCA. Para uma

descrição detalhada, recomenda-se a consulta a CHIANG *et al.* (2001) ou JIANG *et al.* (2019).

Não parece haver um consenso sobre em que situações deve-se usar o PLS ou o CCA. Por exemplo, CHEN (2017), ZHANG *et al.* (2018) e JIANG *et al.* (2019) indicam o CCA para modelar relações de entrada e saída e o PLS para monitoramento de variáveis de qualidade (medidas com mais dificuldade ou com atraso em relação às variáveis de processo). ZHANG *et al.* (2018) acrescentam ainda que o CCA é geralmente aplicado em um nível local, específico do processo, enquanto PCA e PLS são aplicados em um nível global, abrangendo toda a planta. Já SHAWE-TAYLOR e CRISTIANINI (2004), do domínio da ciência da computação, afirmam que o PLS é adequado para obter relações entre representações de entrada e saída e que o CCA é adequado para obter relações entre diferentes representações de um mesmo conjunto de dados. Uma abordagem possível para comparação envolve o estudo da relação matemática entre o PLS e o CCA; BARKER e RAYENS (2003), por exemplo, encaram o PLS como um CCA penalizado, sendo as penalizações representadas por um modelo PCA no espaço **X** e um modelo PCA no espaço **Y**.

Também há confusão na literatura no que se refere à nomenclatura dessas técnicas. O acrônimo PLS, por exemplo, pode ter dois significados: *partial least squares*, o nome original, ou *projection to latent structures*, uma reinterpretação (ABDI, 2010). Além do mais, há muitas variantes do PLS (ROSIPAL e KRÄMER, 2006), e muitas vezes não fica claro a qual delas está sendo feita a referência quando se usa a sigla. O termo *canonical correlation analysis* (CCA) é usado como sinônimo de *canonical variate analysis* (CVA) em algumas referências (CHIANG *et al.*, 2001; DING, 2014). CVA, porém, é melhor entendido e aceito na literatura como a aplicação da metodologia do CCA para a identificação de subespaços, de modo a encontrar uma combinação linear de entradas e saídas passadas que seja mais preditivas das saídas futuras (LI *et al.*, 2022d; SEVERSON *et al.*, 2016). Sendo assim, o CVA possui um sentido bem mais restrito, já que, do ponto de vista matemático, nada na técnica CCA impõe a aplicação nesse contexto específico. A confusão aumenta quando se analisa a literatura de outros domínios: por exemplo, em um livro sobre ecologia, GITTINS (1985) define o CVA como uma aplicação do CCA em que são adicionadas ao conjunto de dados variáveis com valores arbitrários, a quais são atribuídos significados pelo analista, de modo a incluir artificialmente alguma característica dos dados na análise (informações sobre classes, dentre outras). Essa é uma definição do termo completamente alheia ao que se discute nas áreas de quimiometria e monitoramento de processos.

O histórico e as tendências das técnicas PLS/CCA, em especial do PLS, seguem as mesmas linhas do desenvolvimento do PCA. Trabalhos vêm sendo desenvolvidos

para aplicações em processos:

- em batelada (AGARWAL *et al.*, 2022a; CHEN e LIU, 2002; FERRER *et al.*, 2008; GREGERSEN *et al.*, 1997; KOURTI *et al.*, 1995; VAN KOLLENBURG *et al.*, 2020; ZHANG e BAO, 2023);
- com dinâmica (CHEN e LIU, 2002; DONG e QIN, 2015, 2020; LAN *et al.*, 2020; LAVOIE *et al.*, 2019; LIU *et al.*, 2014b; LOU *et al.*, 2020, 2023; QIN, 1998; QIN *et al.*, 2021; QIN e MCAVOY, 1996; WANG *et al.*, 2003; XU e ZHU, 2020; YANG *et al.*, 2020a; ZHANG e LUO, 2020; ZHANG *et al.*, 2023b; ZHU, 2020; ZHU *et al.*, 2020b);
- com não linearidade (CHEN e WANG, 2021; CHU *et al.*, 2022; KONG *et al.*, 2023; PENG *et al.*, 2020b; WU *et al.*, 2020b; ZHANG e LUO, 2021; ZHANG e MA, 2011; ZHANG *et al.*, 2010);
- com multimodalidade (CHEN *et al.*, 2021; SUN e WANG, 2021; ZHANG *et al.*, 2023a; ZHAO *et al.*, 2006);
- com múltiplas escalas de tempo (BOTRE *et al.*, 2017; LEE *et al.*, 2009; TEP-POLA e MINKKINEN, 2000; ZHANG e MA, 2011);
- com múltiplas unidades e/ou estágios (CHEN *et al.*, 2020b; CHOI e LEE, 2005; FUENTE *et al.*, 2023; JIANG e YAN, 2020; KOURTI *et al.*, 1995; MACGREGOR *et al.*, 1994; QIN *et al.*, 2001; SONG *et al.*, 2020; SUN, 2020; ZHANG *et al.*, 2010; ZHU *et al.*, 2022);
- com tratamento dos erros (LI *et al.*, 2017b, 2018a; QIN *et al.*, 2022; ZHENG *et al.*, 2016);
- com dados de baixa qualidade (CHENG *et al.*, 2021; LEE *et al.*, 2007; LI *et al.*, 2018a; YI *et al.*, 2013; YIN *et al.*, 2014c).

Além do mais, tópicos novos vêm surgindo, como o estudo do efeito do tamanho do conjunto de dados na estrutura do modelo, proposto por LI *et al.* (2020f). Para uma revisão ampla da técnica CCA, recomenda-se a leitura de YANG *et al.* (2021b).

2.4.3 Outras técnicas

Como afirmado anteriormente, PCA e PLS/CCA são as mais aplicadas técnicas de modelagem em variáveis latentes no contexto do monitoramento de processos. No entanto, outras técnicas existem, dentre as quais vale destacar:

- **Análise discriminante de Fisher** (*Fisher discriminant analysis, FDA*): proposta por FISHER (1936), essa técnica admite que as variáveis estão divididas *a priori* em duas categorias, devendo-se gerar um conjunto ortogonal de combinações lineares das variáveis originais que melhor separa essas categorias. O FDA pode ser mais eficaz quando há dados sobre a operação normal e múltiplos casos de falha (QIN, 2012). Exemplos de aplicação do FDA em problemas de monitoramento podem ser consultados em ADIL *et al.* (2016); CHO (2007); DENG *et al.* (2016); FENG *et al.* (2016); HE *et al.* (2005, 2009); JIA *et al.* (2020); JIANG *et al.* (2015); LOU *et al.* (2022); MD NOR *et al.* (2015, 2017); SHANG *et al.* (2021); WANG *et al.* (2009); ZHONG *et al.* (2014). Para mais detalhes sobre a técnica, vale a consulta a CHIANG *et al.* (2001).
- **Análise em componentes independentes** (*Independent component analysis, ICA*): proposto por JUTTEN e HERAULT (1991), o ICA gera um conjunto de combinações lineares das variáveis originais de modo que a independência estatística entre as combinações seja máxima. O ICA é considerado uma alternativa para o PCA na hipótese de que a distribuição dos dados não é normal (ZHANG *et al.*, 2014). Aplicações no monitoramento de processos são recentes, tendo os primeiros trabalhos sido propostos no começo da década de 2000 (ALBAZZAZ e WANG, 2004; KANO *et al.*, 2003, 2004; LEE *et al.*, 2003, 2004b,c). Exemplos de avanços subsequentes podem ser consultados em CHEN e YU (2014); HSU *et al.* (2010a,b,c); LI *et al.* (2017c); LUGHOFER *et al.* (2020); XU *et al.* (2021a, 2023a); XU e DENG (2016); ZHU *et al.* (2017). Em particular, ZHANG *et al.* (2016) propuseram um interessante método para escolha automática da aplicação exclusiva ou mista das técnicas PCA e ICA, com ou sem uso de funções *kernel* para incorporação de não linearidade. Para mais detalhes sobre a técnica, consultar THARWAT (2018) ou COMON (1994). Para uma revisão da aplicação da técnica ICA ao monitoramento de processos, recomenda-se a leitura de PALLA e PANI (2023).
- **Análise de características lentas** (*Slow feature analysis, SFA*): proposta por WISKOTT e SEJNOWSKI (2002), essa técnica é baseada em uma formulação de otimização não linear que busca extrair um conjunto de variáveis independentes que variam lentamente ao longo do tempo, sob a hipótese de que variações lentas são mais relevantes e informativas do que as rápidas. O SFA vem sendo uma das mais técnicas mais utilizadas na literatura recente de monitoramento de processos, como evidenciam os trabalhos de GAO e SHARDT (2021, 2022); HUANG *et al.* (2020a); LI *et al.* (2022a, 2023b, 2022c); LIU *et al.* (2023a); PULI *et al.* (2021); SAAFAN e ZHU (2022); WANG *et al.* (2020a); XU e DING (2021); XU *et al.* (2022a,b); ZHANG *et al.* (2022a, 2021c, 2022e); ZHONG

et al. (2020). Para mais detalhes sobre a técnica, vale a consulta a SONG e ZHAO (2022).

Uma descrição de aplicações e avanços das técnicas FDA e ICA na área de monitoramento pode ser encontrada na revisão publicada por MD NOR *et al.* (2019).

2.5 Aprendizado de máquina

A inteligência artificial (IA) é um dos mais recentes campos da ciência e engenharia e vem adquirindo importância crescente nos últimos anos. De acordo com numerosas projeções, a IA é, junto com a biotecnologia, um dos domínios que mais apresentam potencial para avanço, inovação e impacto na sociedade (PAN, 2016; RUSSELL e NORVIG, 2009; VENKATASUBRAMANIAN, 2019). O termo surgiu em 1956, proposto pelos professores J. McCarthy, M. L. Minsky, H. Simon e A. Newell como (PAN, 2016):

(...) the ability of machines to understand, think, and learn in a similar way to human beings.

KAPLAN e HAENLEIN (2019) propõem uma definição mais específica:

We define AI as a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation.

Dessas definições, pode-se perceber a importância que tem o processo de aprendizado para a IA. O subcampo da IA relacionado ao aprendizado é conhecido como aprendizado de máquina (mais comumente referido pelo termo em inglês, *machine learning*). Apesar de em muitos contextos os termos se confundirem, é importante ressaltar que IA é mais amplo do que aprendizado de máquina, pois também envolve percepção dos dados, representação do conhecimento e operacionalização de dispositivos inteligentes (KAPLAN e HAENLEIN, 2019; RUSSELL e NORVIG, 2009). Para uma discussão ampla a respeito do impacto da IA na engenharia química, vale a leitura de VENKATASUBRAMANIAN (2019). Para discussões mais específicas a respeito do impacto do aprendizado de máquina, vale consultar DOBBELAERE *et al.* (2021); FUENTES-CORTÉS *et al.* (2022); GAO *et al.* (2022); KHAN e AMMAR TAQVI (2023); LEE *et al.* (2018b); MOWBRAY *et al.* (2022); QIN e CHIANG (2019); SCHWEIDTMANN *et al.* (2021).

As seções seguintes são dedicadas a aplicações de metodologias de aprendizado de máquina na área de monitoramento de processos.

2.5.1 Definições e histórico na área de monitoramento

Para entender com mais profundidade a natureza dos métodos de aprendizado de máquina, é útil analisar algumas definições propostas para o termo (MAYO, 2018):

A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E (MITCHELL, 1997).

Machine learning is essentially a form of applied statistics with increased emphasis on the use of computers to statistically estimate complicated functions and a decreased emphasis on proving confidence intervals around these functions (GOODFELLOW et al., 2016).

A primeira definição, mais pragmática, tem como foco o significado da palavra *aprendizado* no contexto do termo “aprendizado de máquina”, descrevendo as características operacionais da execução de um programa de computador para que se possa dizer que este esteja “aprendendo”. A segunda definição, mais teórica, é útil para visualizar a relação entre o aprendizado de máquina e a estatística. O aprendizado de máquina, no caso, é interpretado como uma *aplicação* da estatística, sendo a principal diferença em relação à estatística tradicional¹⁴ uma diferença de *abordagem*. O uso do termo *crescente* é particularmente interessante, já que com ele se inclui na definição uma tendência histórica: conforme a capacidade de processamento computacional evolui, mais os algoritmos de aprendizado se apoiam nessa capacidade para estimar funções desconhecidas, em detrimento da interpretabilidade dos mecanismos de cálculo e dos resultados por parte dos seres humanos. Essa tendência passou a se acentuar de forma significativa a partir da popularização das chamadas redes neurais profundas, principal técnica da família de aprendizado profundo (*deep learning*) (GOODFELLOW et al., 2016; VENKATASUBRAMANIAN, 2019).

A aplicação de abordagens de aprendizado de máquina em problemas de monitoramento de processos datam do final da década de 1980, com o uso de redes neurais (HOSKINS et al., 1991; KRAMER, 1992; VENKATASUBRAMANIAN e

¹⁴Nesse contexto, o termo “tradicional” diz respeito tanto à estatística frequentista quanto à bayesiana.

CHAN, 1989). Tais aplicações, no entanto, se restringiram a alguns trabalhos, sendo o enfoque da área muito mais voltado para a modelagem em variáveis latentes interpretada do ponto de vista da teoria estatística multivariada. Com a tendência recente de desenvolvimento da ciência de dados e com a aplicação bem-sucedida de métodos de aprendizado de máquina (em especial o *deep learning*) em diversos domínios, a comunidade de monitoramento de processos vem, especialmente a partir da década de 2010, mudando o enfoque e claramente colocando as abordagens de ciência de dados e aprendizado de máquina em posição de destaque em relação à estatística multivariada tradicional.

A análise dos principais trabalhos de Joe Qin, um dos pesquisadores mais importantes da área, ilustra bem o desenrolar desse processo histórico. Suas revisões sobre técnicas de monitoramento (QIN, 2009, 2003, 2012) apresentam apenas métodos de modelagem em variáveis latentes na perspectiva da estatística multivariada. No artigo de 2003 não é feita qualquer menção ao termo *machine learning*. No de 2009, há apenas uma menção, na segunda linha do artigo, com sentido vago e meramente contextualizador. No de 2012, o termo, apesar de já figurar nas palavras-chave, é utilizado no sentido de uma abordagem alternativa à abordagem principal:

“A related category of nonlinear data-driven methods suitable for process monitoring is from the machine learning literature including neural networks (HOSKINS et al., 1991) and support vector machines (SVM) (JACK e NANDI, 2001). Work in this area is rich by itself and a significant amount of work related to process monitoring is reviewed in VENKATASUBRAMANIAN et al. (2003a)”.

É significativo o fato de o autor, em um artigo cujo título é “*Survey on data-driven industrial process monitoring and diagnosis*”, dedicar treze páginas a metodologias relacionadas ao PCA e PLS e apenas um parágrafo para técnicas de aprendizado de máquina, encaminhando o leitor que deseje mais detalhes a uma revisão de quase uma década atrás.

Dois anos depois, em um artigo de perspectiva (QIN, 2014), mais atenção foi dada aos métodos de aprendizado, ainda com ênfase nas diferenças para a abordagem tradicional e já com a percepção de que era necessário incorporar, na engenharia de sistemas em processos, avanços alcançados em outros domínios:

“Although multivariate statistical approaches have been the favorite choice in the most recent 2 decades, it should be noted that data have been an integral part of process engineering solutions since time models have

been used for process optimization and control. (...) However, these aforementioned data analytics and practice in process systems engineering have apparently not connected to the recent development in machine learning, data mining, and big-data analytics. They differ not only in terms of sizes but also in how and what data should be used in solving real operation problems. In some ways, process systems engineering solutions are confined to one set of principles that are believed to be sound, whereas the machine learning and data-mining communities take the other way and achieve unexpected results and solutions that defy conventional wisdom”.

Já no mais recente trabalho (QIN e CHIANG, 2019), o discurso é claramente outro. O aprendizado de máquina passou para o foco da discussão, como fica claro no título, “*Advances and opportunities in machine learning for process data analytics*”. As técnicas de modelagem em variáveis latentes, antes a ideia central, agora foram classificadas como parte do conjunto das técnicas de aprendizado de máquina:

Although we consider multivariate latent variable methods as a subset of the statistical learning methods, the field of machine learning has grown tremendously in the last two decades (...) Depending on the characteristics of data and models, such as linear, collinear, or nonlinear data, there are respective supervised and unsupervised learning methods to choose from. To deal with collinearity in the data, for example, PCA is unsupervised learning and PLS is supervised learning.

Considerar técnicas de modelagem em variáveis latentes como subconjunto das técnicas de aprendizado de máquina é uma abordagem elegante e faz sentido do ponto de vista teórico, mas no presente trabalho optou-se pelo rigor na descrição do desenvolvimento histórico, sendo portanto mais adequada a apresentação das técnicas em separado.

2.5.2 Tipos de aprendizado

A principal forma de classificar as técnicas de aprendizado de máquina diz respeito ao tipo de aprendizado:

- **supervisionado:** o conjunto de treinamento possui a solução desejada (tipicamente na forma de rótulos);
- **não supervisionado:** o conjunto de treinamento não possui a solução desejada;

- **semi-supervisionado**: alguns dados de treinamento possuem a solução desejada e outros não;
- **por reforço**¹⁵: o algoritmo observa o ambiente, seleciona e executa ações e ganha recompensas (ou sofre penalidades) em retorno; com base nessas recompensas ou penalidades, o algoritmo decide o melhor caminho a seguir.

Outra classificação comum diz respeito à natureza das variáveis de predição. Se elas são discretas, a técnica é dita de *classificação*; se são contínuas, a técnica é de *regressão*.

Nas Seções 2.5.3 e 2.5.4, a título ilustrativo, são apresentadas com detalhes a base teórica e a evolução das aplicações de duas das técnicas de aprendizado de máquina mais usadas na área de monitoramento: redes neurais artificiais (*artificial neural networks*, ANN) e máquinas de vetores suporte (*support vector machines*, SVM). Na Seção 2.5.5, outras técnicas importantes são descritas sucintamente.

2.5.3 Redes neurais artificiais (ANN)

Redes neurais artificiais constituem uma classe de técnicas de aprendizado inspiradas vagamente no funcionamento do cérebro humano (AGGARWAL, 2018; ALDRICH e AURET, 2013). A Figura 2.3 ilustra a estrutura de um neurônio biológico. Os neurônios se ligam por meio de conexões entre dendritos (entradas) e axônios (saídas), conhecidas como sinapses. As forças entre as conexões são moduladas de acordo com estímulos externos. Essas modulações constituem o mecanismo por meio do qual ocorre o processo de aprendizado em seres vivos (AGGARWAL, 2018).

A Figura 2.4 apresenta a estrutura de um neurônio artificial. Cada neurônio j na rede é uma unidade de processamento local, que aceita n conexões por meio das entradas x_1, x_2, \dots, x_n e fornece uma saída o_j . Cada conexão i tem um peso associado, denotado por w_{ij} .

A saída o_j do neurônio é expressa por:

$$o_j = \phi \left(\sum_{i=1}^n w_{ij} x_i + \theta_j \right), \quad (2.23)$$

sendo ϕ uma função de ativação não linear. A *rede neural* é formada pela conexão

¹⁵Para uma interessante demonstração da capacidade de aprendizado dos algoritmos por reforço, vale assistir ao curto vídeo youtu.be/V1eYniJ0Rnk, em que se demonstra a aplicação do sistema *Google DeepMind* a um tradicional jogo de vídeo.

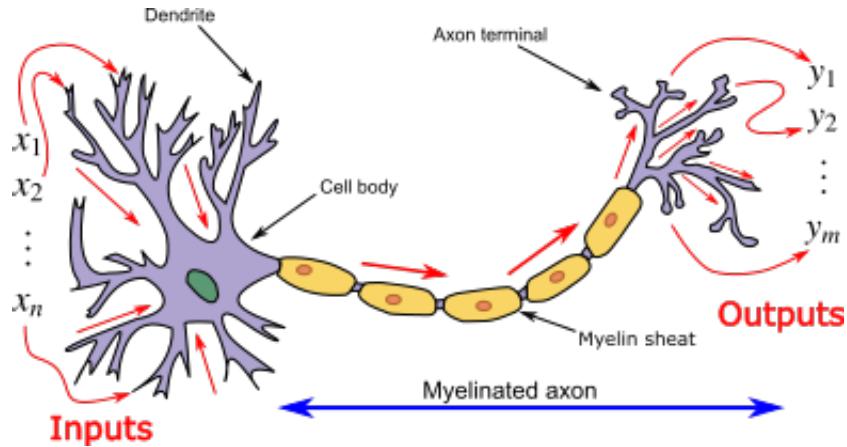


Figura 2.3: Modelo de um neurônio biológico (VU-QUOC, 2018).

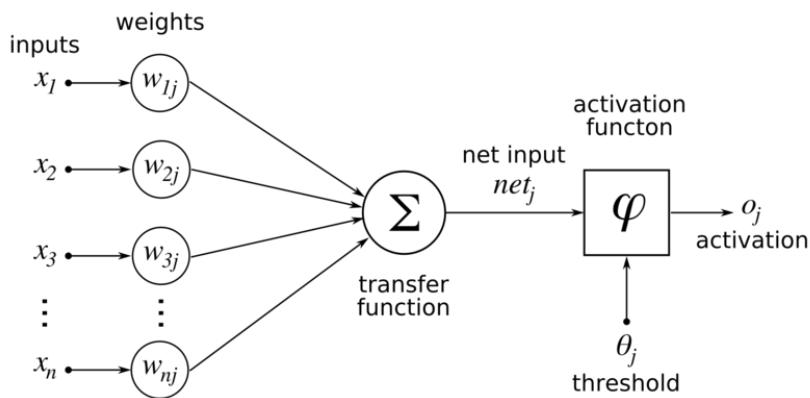


Figura 2.4: Modelo de um neurônio artificial (CHRISLB, 2005).

entre os vários neurônios. O aprendizado da rede, em semelhança ao que ocorre nos organismos vivos, se dá por meio da modulação dos pesos para adequar dados de entrada a dados de saída, especificados durante a etapa de treinamento. O aprendizado é bem-sucedido quando a rede tem a capacidade de *generalizar*: fornecer a saída correta para dados de entrada de teste, que não foram usados na etapa de treino. Existem vários algoritmos usados para o treinamento de redes neurais, que podem ser categorizados de acordo com as quatro regras básicas de aprendizado, descritas a seguir.

ANN - Regras de aprendizado

As quatro regras básicas de aprendizado utilizadas no treinamento de redes neurais artificiais são (CHOWDHARY, 2020):

- **Aprendizado por correção de erros (Regra Delta):** neste tipo de algoritmo supervisionado, os pesos são ajustados para minimizar o erro de predição da rede, que consiste na diferença entre a saída da rede e o resultado esperado.

- **Aprendizado competitivo:** neste tipo de algoritmo não supervisionado, os pesos são ajustados por meio da interação competitiva entre os neurônios, sendo que o objetivo de cada neurônio é alcançar o maior poder de representação em relação aos padrões dos dados de entrada.
- **Aprendizado de Boltzmann:** neste tipo de algoritmo não supervisionado, os pesos são ajustados utilizando o princípio da minimização da energia, de modo a capturar dependências estatísticas nos dados e possibilitar a criação de um modelo para gerar amostras semelhantes ao conjunto de treinamento.
- **Aprendizado Hebbiano:** nesse tipo de aprendizado não supervisionado, os pesos são ajustados com base no princípio de Hebb, que postula que se neurônios a jusante e a montante de uma conexão são ativados simultaneamente, essa conexão deve ser fortalecida, enquanto que se a ativação é assíncrona, a conexão deve ser enfraquecida.

Dos quatro tipos de aprendizado mencionados, o aprendizado por correções de erros é o mais amplamente aplicado e será o foco principal desta seção. Nesse contexto, a seguir serão descritos os *percéptrons*, que são os neurônios mais simples capazes de tomar decisões.

ANN - Percéptrons

O *percéptron* é o neurônio cuja função de ativação é a *função degrau binária*:

$$f(u) = \begin{cases} 0, & \text{se } u \leq 0 \\ 1, & \text{se } u > 0. \end{cases} \quad (2.24)$$

Com essa função, é possível treinar os pesos w do neurônio, de modo a criar um classificador binário¹⁶.

Para introduzir mais complexidade e flexibilidade na classificação, pode-se associar os percéptrons em rede, como na Figura 2.5, em que é apresentada uma rede com duas camadas ocultas.

Pode-se mostrar teoricamente que é possível calcular *qualquer* função com redes de percéptrons, por mais complicadas que sejam (HORNIK *et al.*, 1989; NIELSEN, 2015). Nesse sentido, redes de percéptrons são computacionalmente universais; ou

¹⁶Há com esse intuito um algoritmo simples, chamado de *regra de aprendizado do percéptron*, com convergência garantida para o caso linearmente separável e que pode ser conferido em RUSSELL e NORVIG (2009).

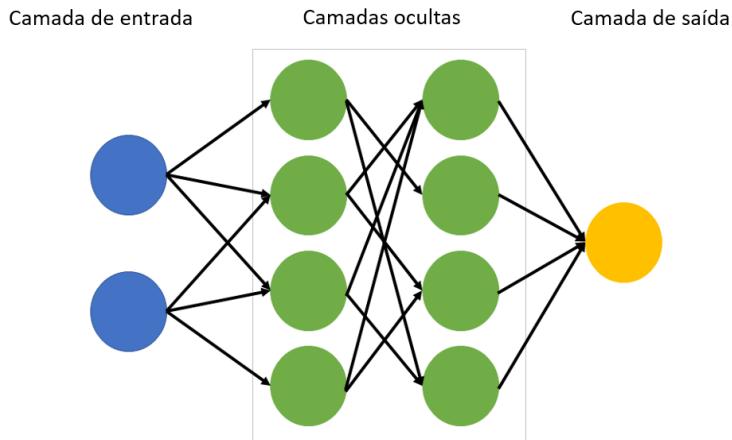


Figura 2.5: Rede de percéptrons com camadas ocultas (adaptado de TSEKICHUN (2021)).

seja, toda operação computacional pode ser representada por meio de uma rede de percéptrons. Apesar dessa forte garantia teórica, por anos as redes neurais foram pouco utilizadas na comunidade científica, por conta da falta de um algoritmo eficiente para o treinamento. O cenário mudou com o trabalho pioneiro de RUMELHART *et al.* (1986), que propuseram o *neurônio sigmóide* e o *algoritmo de retropropagação do erro*¹⁷.

ANN - Neurônio sigmóide e algoritmo de retropropagação do erro

Treinar uma rede por meio da Regra Delta significa ajustar seus pesos, de modo a minimizar o erro de predição; ou seja, adequar os pesos para que a rede resulte na saída desejada dentro de uma margem mínima de erro. Portanto, para que o treinamento seja possível, a relação entre pesos e saída precisa ser suave (pequenas mudanças nos pesos devem levar a pequenas mudanças na saída) (NIELSEN, 2015). Utilizando percéptrons como neurônios, a descontinuidade da função de ativação degrau binária (Equação 2.24) faz com que a relação entre pesos e saídas seja brusca e descontínua. RUMELHART *et al.* (1986) propuseram o neurônio sigmóide, em que a função de ativação é a função sigmóide¹⁸:

$$f(u) = \frac{1}{1 + \exp(-u)}. \quad (2.25)$$

Como mostrado na Figura 2.6, a função sigmóide pode ser interpretada como uma

¹⁷Segundo GERON (2017), esse algoritmo foi inventado várias vezes ao longo do tempo em vários domínios, começando por WERBOS (1974) nas ciências comportamentais.

¹⁸Há evidências de que neurônios biológicos utilizam funções de ativação sigmóide (GERON, 2017).

suavização da função degrau binária, o que facilita o processo de treinamento.

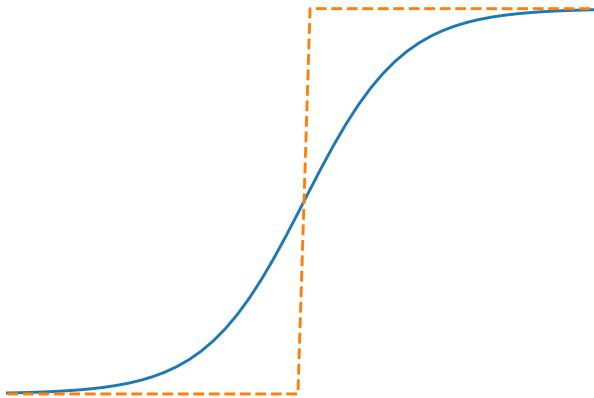


Figura 2.6: Comparação entre as funções de ativação degrau binária e sigmóide.

A outra grande contribuição do trabalho de RUMELHART *et al.* (1986) foi o algoritmo de retropropagação do erro. Cada iteração desse algoritmo é dividida em duas etapas: na primeira, chamada de passe direto, calcula-se a saída da rede; na segunda, o passe reverso, calcula-se o quanto cada camada contribui para o erro da camada anterior (isso é efetuado de trás para frente, ou seja, da saída para a entrada). O resultado dessa segunda etapa é uma medida eficiente do gradiente dos erros ao longo da rede, o que torna possível a aplicação do algoritmo do gradiente descendente para a minimização dos erros.

Diversos avanços nos algoritmos de treinamento tradicionais vêm sendo propostos. Um exemplo são as *máquinas de aprendizado extremo* (*extreme learning machines*, EML), uma metodologia proposta por HUANG *et al.* (2006) para aceleração do aprendizado das redes por meio da escolha aleatória dos nós em que serão treinados os pesos. Outra abordagem ainda mais bem-sucedida são os métodos de aprendizado profundo, descritos a seguir.

ANN - Aprendizado profundo

O aprendizado profundo é uma metodologia de aprendizado em que múltiplas camadas de processamento da informação possibilitam que conceitos complicados sejam representados por meio da associação de conceitos simples (GOODFELLOW *et al.*, 2016). Há evidências de que redes neurais biológicas processam informações dessa maneira, como por exemplo a organização das células no córtex cerebral (GERON, 2017). No contexto de redes neurais artificiais, redes profundas são aquelas que possuem mais de uma camada oculta.

Apesar de as propostas de RUMELHART *et al.* (1986) terem provocado grande interesse da comunidade científica pelo uso de redes neurais artificiais, dificuldades referentes ao treinamento de redes com múltiplas camadas persistiram por muito tempo. Apenas recentemente essas dificuldades vêm sendo contornadas. Trabalhos importantes incluem os de HINTON *et al.* (2006), que propuseram uma estratégia de treinamento por camada utilizando unidades de processamento gráfico (GPU's), e GLOROT e BENGIO (2010), que propuseram alternativas de função de ativação (além da popular sigmóide) e novas estratégias de inicialização dos pesos, de modo a atenuar o problema da dissipação ou explosão de gradientes ao longo das camadas. A grande maioria dos avanços de ponta do aprendizado de máquina em vários domínios, como a vitória do sistema AlphaGo no notoriamente complicado jogo chinês de tabuleiro Go (CHEN, 2016; WANG *et al.*, 2016), é resultante da aplicação de redes neurais profundas (VENKATASUBRAMANIAN, 2019).

Uma crítica comum a métodos dessa classe é a falta de interpretabilidade dos resultados por parte dos seres humanos, devido à falta de entendimento sobre os mecanismos internos de modelagem (VENKATASUBRAMANIAN, 2019). Recentemente, SIVARAM *et al.* (2020) e DAS *et al.* (2020) tentaram clarificar essa questão por meio de um estudo sistemático, utilizando modelos cuidadosamente escolhidos. CARTER *et al.* (2023) apresentaram uma revisão do aprendizado de máquina interpretável aplicado à indústria de processos.

Para uma revisão ampla da área de aprendizado profundo, recomenda-se a consulta a SENGUPTA *et al.* (2020). Para uma revisão da aplicação do aprendizado profundo à área de monitoramento de processos, consultar (YU e ZHANG, 2023). Para uma análise crítica sobre a aplicação de redes neurais profundas para a modelagem de sistemas dinâmicos, consultar ZHOU e ZHU (2023).

ANN - Arquiteturas

Diversos tipos de arquitetura podem ser propostos para as redes neurais (AGGARWAL, 2018; PIRDASHTI *et al.*, 2013). A divisão mais comum é entre as redes de alimentação direta¹⁹ (*feed-forward*) ou acíclicas, e as redes recorrentes ou cíclicas. Nas redes de alimentação direta, a informação se propaga em apenas um sentido, da entrada para a saída, como na rede da Figura 2.5. Nas redes recorrentes, as entradas são alimentadas sequencialmente e convertidas em estados ocultos da rede, que são retroalimentados à rede. Isso permite a interação de uma entrada fornecida no tempo

¹⁹Para uma visualização interativa de uma rede com essa arquitetura, recomenda-se a visitação da página playground.tensorflow.org.

t com entradas fornecidas nos tempos anteriores, como $t - 1, t - 2$ e assim por diante. Por esse motivo, redes recorrentes são especialmente úteis para a modelagem de séries temporais, sentenças textuais ou sequências biológicas (AGGARWAL, 2018).

Mecanismos de atenção, propostos por BAHDANAU *et al.* (2016), têm o objetivo de aprimorar o desempenho de redes neurais profundas por meio da ponderação e combinação dos dados de entrada com base na relevância para a tarefa em questão. Em particular, a autoatenção, proposta por CHENG *et al.* (2016), permite que cada posição nos dados de entrada se relacione a todas as outras posições de maneira independente. Essa ideia revolucionou o campo do aprendizado profundo, em especial o tratamento de dados sequenciais, permitindo a criação de arquiteturas de alimentação direta capazes de modelar dependências de longo alcance com algoritmos de treinamento paralelizáveis (NIU *et al.*, 2021). A arquitetura *Transformer* (VASWANI *et al.*, 2017), amplamente usada em modelos de linguagem como o *Generative pre-trained Transformer* (GPT) (RADFORD *et al.*, 2018), é baseada em mecanismos de autoatenção.

Autocodificadores são redes neurais designadas para reproduzir na saída, sob certas restrições, os dados alimentados na entrada (CHARTE *et al.*, 2020). A utilidade da metodologia vem da natureza das restrições. Por exemplo, se a camada oculta possui menos neurônios que a entrada, a rede será obrigada a representar os dados em um espaço latente de menor dimensão (chamado de código), de maneira semelhante ao procedimento de redução de dimensionalidade efetuado pelo modelo PCA (YANG e GE, 2020). Por isso, autocodificadores podem ser usados como modelos para redução de dimensionalidade não linear. Para uma revisão sobre a aplicação de autocodificadores para o monitoramento de processos, consultar QIAN *et al.* (2022).

Redes gerativas adversariais, propostas primeiramente por GOODFELLOW *et al.* (2014), são modelos compostos por duas redes neurais – a geradora e a discriminadora – treinadas em conjunto para gerar dados sintéticos com crescente realismo; a rede geradora tenta gerar dados cada vez mais parecidos com os dados reais e recebe *feedback* da rede discriminadora, que por sua vez é treinada para maximizar a capacidade de diferenciação entre dados reais e sintéticos. Para revisões sobre aplicações de redes gerativas adversariais a problemas com séries temporais e detecção de anomalias, respectivamente, consultar BROPHY *et al.* (2023); SABUHI *et al.* (2021).

Máquinas de Boltzmann restritas utilizam a regra de aprendizado de Boltzmann em conjunto com restrições de topologia de modo a tornar o processo de treinamento mais eficiente (FISCHER e IGEL, 2012). Uma conhecida classe de

redes de aprendizado profundo que aprendem de forma não supervisionada são as redes *deep-belief*, formadas por composição de máquinas de Boltzmann restritas (YU e YAN, 2019). Além disso, é comum usar máquinas de Boltzmann restritas na chamada etapa de pré-treino das redes profundas tradicionais (supervisionadas) (AGGARWAL, 2018).

Mapas auto-organizáveis combinam os princípios do aprendizado competitivo e Hebbiano para criar representações (tipicamente bidimensionais) de padrões complexos, preservando a estrutura topológica dos dados de entrada (KOHONEN, 1990; YIN, 2008). As representações permitem a visualização de padrões, em especial no contexto de tarefas como agrupamento ou modelagem em variáveis latentes.

Arquiteturas difusas (*fuzzy*), conhecidas como *neuro-fuzzy*, utilizam redes neurais como método de aprendizado para ajuste de parâmetros em um sistema de lógica difusa²⁰ (MD NOR *et al.*, 2019). Vale ressaltar que a lógica difusa por si só, sem o uso das redes, pode ser usada para elaboração de sistemas de aprendizado, como ilustram aplicações descritas na Seção 2.5.5.

A arquitetura de redes convolucionais, por ser utilizada no presente trabalho, é descrita em detalhes na próxima subseção.

ANN - Redes convolucionais

Redes convolucionais são redes de arquitetura esparsa, em que filtros são aplicados para ativar apenas os neurônios que levam ao aprendizado ótimo (VENKATASUBRAMANIAN, 2019). São especialmente bem-sucedidas em aplicações de visão computacional, tendo excedido recentemente a capacidade humana nessa área (HE *et al.*, 2015). Por serem esparsas, as redes convolucionais são capazes de aprender padrões locais e reconhecer esses padrões em qualquer posição de outra amostra. No caso de imagens, esses padrões podem corresponder a bordas, curvas e outros formatos geométricos²¹.

Para ser alimentada na entrada da rede, uma imagem deve ser representada como um tensor numérico no formato $l \times h \times c$, em que l diz respeito à largura, h diz respeito à altura e c diz respeito à profundidade, onde se distribuem os canais de cores. A primeira camada de processamento é geralmente uma camada convolucional, onde se aplica a operação de convolução, esquematizada na Figura 2.7.

²⁰A lógica difusa é uma generalização da lógica booleana, em que os valores de verdade das variáveis podem admitir valores no intervalo contínuo entre 0 (falso) e 1 (verdadeiro).

²¹Para visualizações interativas de redes convolucionais, recomenda-se a visitação das páginas adamharley.com/nm_vis/cnn/3d.html e poloclub.github.io/cnn-explainer/.

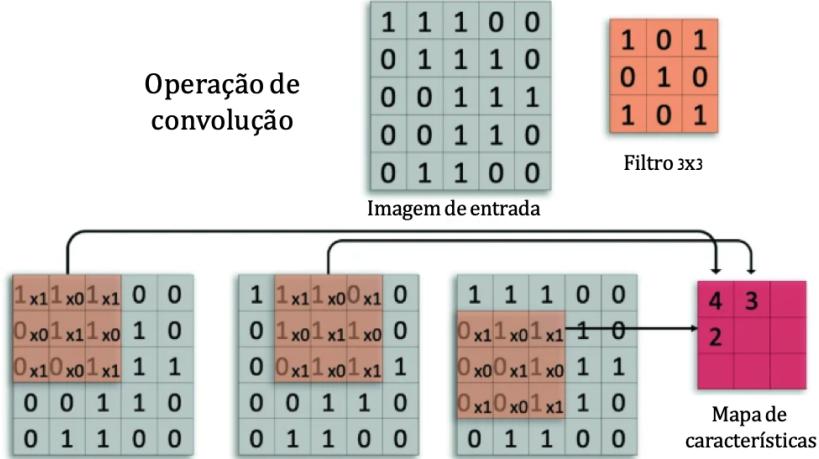


Figura 2.7: Ilustração da operação de convolução. Adaptado de NG e FENG (2020).

Na operação de convolução, um filtro (também chamado de *kernel*) percorre os pixels da imagem computando a cada passo o produto escalar e armazenando os resultados em uma matriz chamada de mapa de características. O mapa de características é a estrutura que reconhece os vários padrões (características) relativos à imagem. A natureza do padrão reconhecido depende da estrutura do filtro. Quando a imagem possui vários canais, o mapa de características resultante do filtro é a soma dos mapas de características correspondentes a cada canal. Em geral, os filtros têm dimensões 3×3 (a mais comum), 5×5 e 7×7 .

A camada convolucional pode ter vários filtros, cada um dando origem a um mapa de características. Cada elemento de um mapa de características corresponde a um neurônio. Os neurônios em um mesmo mapa de características compartilham os mesmos parâmetros (pesos). Em geral, a matriz de números que constitui cada filtro de dada camada é um parâmetro da rede, ou seja, é ajustada junto com os pesos de modo a resultar no aprendizado ótimo. O número de filtros de dada camada é um hiperparâmetro (deve ser determinado a priori).

Os múltiplos mapas de características podem ser alimentados a uma nova camada convolucional, resultando em novos mapas, como ilustrado na Figura 2.8. Cada neurônio de um mapa de características da saída se conecta a somente uma porção dos neurônios dos mapas da entrada. A porção da camada anterior que se conecta a um neurônio é chamada de seu campo receptivo.

A operação de convolução faz com que o mapa de saída tenha formato menor do que o mapa de entrada. De modo a fazer com que o formato seja o mesmo, pode-se utilizar estratégias de preenchimento (*padding*). Na estratégia conhecida como *zero padding* ou *same padding*, adicionam-se zeros às bordas do mapa de entrada de modo que o mapa de saída possua o mesmo formato da entrada. É uma maneira de

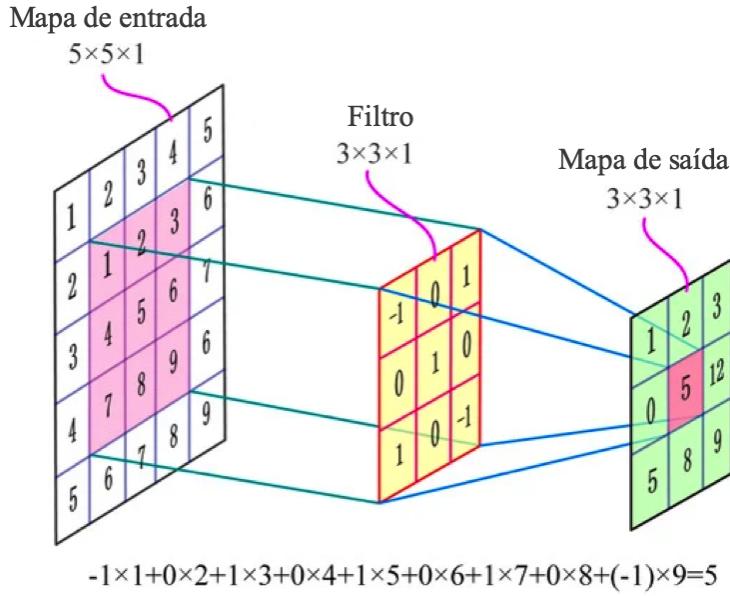


Figura 2.8: Geração de um novo mapa de características a partir da aplicação de um filtro de convolução. Adaptado de GONG *et al.* (2019).

preservar as informações das bordas ao longo da rede. Caso deseje-se gerar mapas de saída menores do que os de entrada, podem ser utilizados passos largos (*strides*).

Além da camada convolucional, é comum adicionar às redes camadas de *pooling*. Assim como as camadas convolucionais, os neurônios das camadas de *pooling* possuem campos receptivos, mas não há pesos e a operação efetuada não é a convolução, mas alguma operação de agregação, como a média ou a escolha do valor máximo (sendo essa última a mais usada). A operação de *pooling* é ilustrada na Figura 2.9.

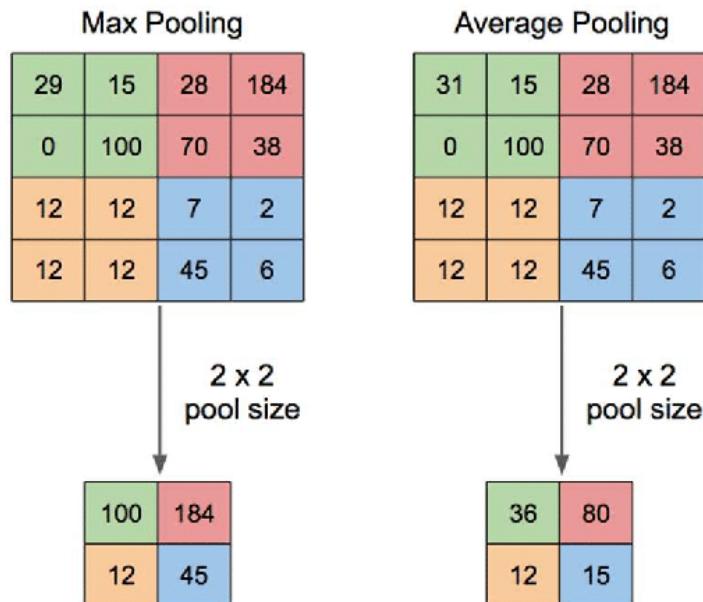


Figura 2.9: Ilustração da operação de *pooling* (YANI *et al.*, 2019).

Em geral, os campos receptivos de *pooling* têm tamanho 2×2 e *stride* 2, como mostrado acima. As camadas de *pooling* diminuem a quantidade de parâmetros da rede, atenuando o sobreajuste e o custo computacional. Além do mais, como há agregação da informação de janelas maiores em janelas menores, os campos receptivos das camadas mais profundas da rede, mesmo sendo pequenos, possuem informações relativas à toda a imagem original. Isso possibilita a criação de padrões globais e mais complexos/abstratos nas camadas profundas.

A Figura 2.10 ilustra a arquitetura típica de uma rede convolucional. A primeira parte da rede efetua a extração de características, ou seja, identifica os diversos padrões presentes na imagem. Tipicamente, utilizam-se vários pares de camadas convolução/*pooling*, o que define a profundidade da rede. A segunda parte consiste de uma camada densa (totalmente conectada) e é responsável por efetuar a classificação.

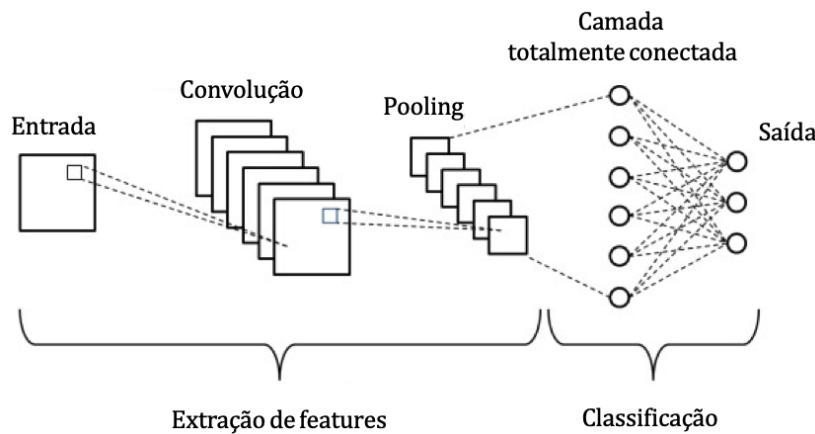


Figura 2.10: Arquitetura típica de uma rede convolucional. Adaptado de PHUNG e RHEE (2019).

No Capítulo 5 do presente trabalho, uma rede convolucional é utilizada para efetuar diagnóstico de falhas com base em análise de matrizes de distâncias em dados de processos.

Para uma discussão sobre a aplicação de redes neurais convolucionais a problemas da engenharia química, consultar JIANG e ZAVALA (2021).

ANN - Aplicações em monitoramento de processos

Aplicações de redes neurais para fins de monitoramento podem ser classificadas historicamente em duas abordagens. Na primeira abordagem, do tipo supervisionado, a cada nó de saída da rede é atribuído um cenário (normal ou de falha) e a rede deve classificar um dado conjunto de dados como correspondendo a cada cenário (AHMAD e HAMID, 2001; HOSKINS *et al.*, 1991; WU e ZHAO, 2018). Na segunda

abordagem, do tipo não supervisionado, as redes são incorporadas a métodos clássicos, como PCA, para incorporar não linearidade (DONG e MCAVOY, 1996; KRAMER, 1991). Evoluções dessa segunda abordagem incluem o uso das redes para extração de *features*, de maneira não linear, sem a dependência do formalismo de técnicas tradicionais, como proposto por ZHAO e LAI (2019), e o uso de redes *deep-belief* para aprendizado não supervisionado (ZHANG e ZHAO, 2017).

Diversos tipos de arquitetura das redes vêm sendo utilizadas em aplicações de monitoramento de processos, como:

- redes recorrentes (AGARWAL *et al.*, 2020, 2022b; DENG *et al.*, 2023; HAN *et al.*, 2020; JI *et al.*, 2023; KANG, 2020; KHALIFA *et al.*, 2023; LEMOS *et al.*, 2021; LI e QIN, 2023; OUYANG *et al.*, 2020; PENG *et al.*, 2020a; SHAHNAZARI, 2020; XAVIER e DE SEIXAS, 2018; XIONG *et al.*, 2022; YUAN e TIAN, 2019; ZHAO *et al.*, 2018);
- redes convolucionais (AL-WAHAIBI *et al.*, 2023; BAO *et al.*, 2022; GAO *et al.*, 2020b; GUO *et al.*, 2022; LI *et al.*, 2020b, 2022b; MA *et al.*, 2023; OLIVIER e ALDRICH, 2020; WU e ZHAO, 2018; XING e XU, 2022; YU *et al.*, 2020, 2021; ZHANG *et al.*, 2022d; ZHENG e ZHAO, 2020);
- autocodificadores (CACCIARELLI e KULAHCI, 2022; DENG *et al.*, 2023; GAO *et al.*, 2023; GUO *et al.*, 2021b; HALLGRIMSSON *et al.*, 2021; JANG *et al.*, 2023; LI *et al.*, 2023a; LIU *et al.*, 2022a,b; MA *et al.*, 2023; PENG *et al.*, 2022a; RUAN *et al.*, 2022; YANG e WANG, 2023; YIN e YAN, 2021; YU *et al.*, 2022, 2019; ZHANG *et al.*, 2021a; ZHANG e ZHAO, 2022; ZHANG *et al.*, 2023c; ZHU *et al.*, 2021a,b);
- redes generativas adversariais (CHAI e ZHAO, 2020; DU *et al.*, 2023; DZAFERAGIC *et al.*, 2022; JANG *et al.*, 2021; LI *et al.*, 2020a; LOMOV *et al.*, 2021; MA *et al.*, 2021; TIAN *et al.*, 2020; XU *et al.*, 2023b; ZHUO e GE, 2020);
- *Transformers* (BAI e ZHAO, 2023; BI *et al.*, 2023; WEI *et al.*, 2022; YI *et al.*, 2023; ZHANG *et al.*, 2022c; ZHOU *et al.*, 2023);
- máquinas de Boltzmann restritas/*deep-belief* (MOODY, 2014; PENG *et al.*, 2019; TANG *et al.*, 2018; TIAN *et al.*, 2020; WANG *et al.*, 2020b; WEI e WENG, 2020; YANG *et al.*, 2021a; YU e YAN, 2019; ZHANG e ZHAO, 2017);
- mapas auto-organizáveis (CHEN e YAN, 2012; LU e YAN, 2020, 2021; NG e SRINIVASAN, 2008; SILVA e WILCOX, 2019; VON BIRGELEN *et al.*, 2018; YAN *et al.*, 2019; YING *et al.*, 2020; YU *et al.*, 2014);

- redes *neuro-fuzzy* (AYOUBI e ISERMANN, 1997; EVSUKOFF e GENTIL, 2005; GHOLIZADEH *et al.*, 2019; KHALID *et al.*, 2011; KORBICZ e KOWAL, 2007; LAU *et al.*, 2013; PALADE *et al.*, 2002; SALAHSHOOR *et al.*, 2010; SHABANIAN e MONTAZERI, 2011; SUBBARAJ e KANNAPIRAN, 2014);
- máquinas de aprendizado extremo (BOLDT *et al.*, 2017; ZOU *et al.*, 2018), inclusive recentemente na forma não supervisionada (PENG *et al.*, 2022b; YONGYONG e XIAOQIANG, 2020; ZHANG *et al.*, 2019).

Uma abordagem usada em outros domínios (CZAJKOWSKI e PATAN, 2016; WOOTTON *et al.*, 2015), mas que não parece estar muito presente na área de processos, é o uso de redes neurais para criação de modelos de regressão e predição de variáveis de qualidade ou de saída (analogamente ao mecanismo do PLS e CCA), com subsequente monitoramento do erro de predição das observações de teste. Tal abordagem pode ser promissora por conta da grande capacidade das redes neurais capturarem relações não lineares. Um exemplo recente foi proposto por LEMOS *et al.* (2021), que aplicaram uma rede neural recorrente do tipo *echo state* para monitoramente de um sistema de medição fiscal em plataforma de petróleo.

Para revisões amplas sobre aplicações de redes neurais na engenharia química, recomenda-se a leitura de PIRDASHTI *et al.* (2013), CAVALCANTI *et al.* (2021), JIANG e ZAVALA (2021) e SUN *et al.* (2021).

2.5.4 Máquinas de vetores suporte (SVM)

Máquinas de vetores suporte, primeiramente propostas por BOSER *et al.* (1992), se propõem a encontrar hiperplanos que separam o espaço de variáveis em regiões de interesse (GERON, 2017; MD NOR *et al.*, 2019). Os hiperplanos são calculados com base em um pequeno subconjunto das observações, chamadas de *vetores suporte*. Para tal, é definida uma grandeza denominada *margem*, que corresponde à distância entre o hiperplano e o vetor suporte. O treinamento se dá pela resolução de um problema de otimização: na classificação, as margens são maximizadas, enquanto na regressão elas são minimizadas. Para introduzir não linearidade na técnica, o formalismo com hiperplanos permanece, porém com a utilização de mapeamentos para outros espaços, como explicado adiante.

A seguir, a técnica é descrita nos contextos de classificação supervisionada linear e não linear, classificação não supervisionada e regressão.

SVM – Classificação supervisionada linear

A Figura 2.11 ilustra o problema mais simples possível de classificação supervisionada, em que os dados se encontram no espaço bidimensional e há duas classes linearmente separáveis. As três retas da figura são igualmente válidas como modelos de discriminação, ao menos no que se refere aos pontos apresentados. A reta contínua, porém, intuitivamente é percebida como a melhor escolha, já que ela se encontra à maior distância de ambos os conjuntos, o que em tese diminuiria as chances de erro de classificação no caso de os pontos da figura serem representativos da população de que foram amostrados. Esse critério intuitivo de maximização da distância entre dados e modelo é a base da técnica SVM.

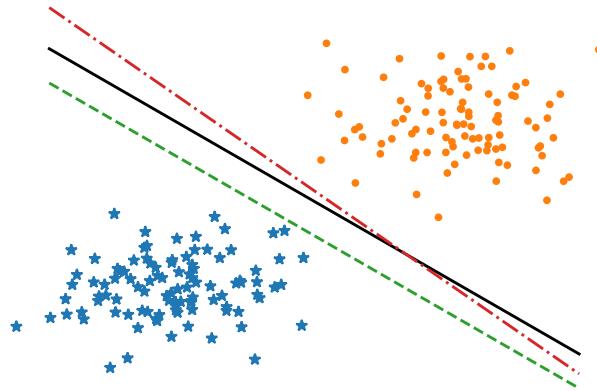


Figura 2.11: Ilustração em espaço bidimensional do problema de classificação binária linearmente separável.

Os pontos mais próximos da reta de classificação, em cada categoria, são chamados de *vetores suporte*. As distâncias entre esses pontos e a reta são chamadas de *margens*.

Ao generalizar o problema para um espaço m -dimensional, a reta de separação torna-se um hiperplano, descrito pela equação $\mathbf{w}^T \cdot \mathbf{x} + b = 0$. O conjunto de dados de treinamento S é escrito como (GERON, 2017):

$$S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}, \quad (2.26)$$

sendo $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$ uma das n observações. O booleano $y_i \in (0, 1)$ é usado

como critério para classificação:

$$y = \begin{cases} 0 & \text{se } \mathbf{w}^T \cdot \mathbf{x} + b < 0 \\ 1 & \text{se } \mathbf{w}^T \cdot \mathbf{x} + b \geq 0, \end{cases} \quad (2.27)$$

em que é atribuído a cada uma das classes o valor 0 ou 1. Geometricamente, pode-se mostrar que, quanto menor a inclinação $\|\mathbf{w}\|$ do hiperplano, maior é o tamanho da margem (GERON, 2017; LORENA e CARVALHO, 2003). Essa é a base do algoritmo de treinamento do SVM, constituído por um método de otimização. Há dois tipos de abordagem para montar o algoritmo de otimização SVM linear: com margens *rígidas* ou *flexíveis*.

No caso de margens rígidas, o objetivo é simplesmente maximizar as margens. Sendo assim, o problema de classificação SVM linear com margens rígidas pode ser posto como (GERON, 2017):

$$\begin{aligned} \text{Minimizar}_{\mathbf{w}, b} \quad & \frac{1}{2} \mathbf{w}^T \cdot \mathbf{w} \\ \text{sujeito a:} \quad & t^{(i)} (\mathbf{w}^T \cdot \mathbf{x}^{(i)} + b) \geq 1 \quad \forall i = 1, \dots, m, \end{aligned} \quad (2.28)$$

em que $t^{(i)} = -1$ para $y^{(i)} = 0$ e $t^{(i)} = +1$ para $y^{(i)} = 1$. A substituição de $\|\mathbf{w}\|$ por $\frac{1}{2} \mathbf{w}^T \cdot \mathbf{w}$ na função objetivo é conveniente para facilitar a resolução numérica. A restrição de desigualdade é a tradução matemática do conceito de margem rígida: nenhum ponto pode se encontrar além da margem.

No caso de margens flexíveis, violações são aceitas: é permitido que alguns pontos se encontrem além da margem, e em certos casos, do lado errado da classificação. Isso pode ser necessário para tornar a técnica mais robusta (menos sensível a *outliers*) e possibilitar a aplicação do formalismo linear mesmo em dados que não sejam linearmente separáveis.

O problema de classificação SVM com margens flexíveis pode ser posto como (GERON, 2017):

$$\begin{aligned} \text{Minimizar}_{\mathbf{w}, b, \zeta} \quad & \frac{1}{2} \mathbf{w}^T \cdot \mathbf{w} + C \sum_{i=1}^m \zeta^{(i)} \\ \text{sujeito a:} \quad & t^{(i)} (\mathbf{w}^T \cdot \mathbf{x}^{(i)} + b) \geq 1 - \zeta^{(i)} \quad \forall i = 1, \dots, m, \end{aligned} \quad (2.29)$$

em que $\zeta^{(i)} \geq 0$ são variáveis de relaxamento, que representam o quanto à instância i é permitido violar a margem.

Nesse caso, há um balanço entre dois objetivos conflitantes: diminuir $\frac{1}{2} \mathbf{w}^T \cdot \mathbf{w}$, de modo a aumentar a margem, e diminuir $\zeta^{(i)}$, para que haja poucas violações. O

hiperparâmetro C possibilita que se estabeleça um compromisso entre esses dois objetivos.

SVM – Classificação supervisionada não linear

A essência da classificação não linear está no mapeamento do conjunto de treinamento S , do espaço de entrada para um novo espaço, denominado *espaço de características*, resultando em um novo conjunto de treinamento $\Phi(S)$:

$$\Phi(S) = \{(\Phi(\mathbf{x}_1), y_1), (\Phi(\mathbf{x}_2), y_2), \dots, (\Phi(\mathbf{x}_n), y_n)\}. \quad (2.30)$$

Esse mapeamento é feito com a função Φ :

$$\Phi(\mathbf{x}_i) = (\Phi_1(\mathbf{x}_i), \Phi_2(\mathbf{x}_i), \dots, \Phi_M(\mathbf{x}_i)). \quad (2.31)$$

A dimensão M do espaço de características pode ser muito maior do que a dimensão m do espaço de entrada. No novo espaço eventualmente ampliado, é bem possível que os dados sejam linearmente separáveis. Na verdade, o teorema de Cover (COVER, 1965) afirma que a probabilidade de que isso aconteça é bastante alta. Tudo depende da escolha apropriada da função Φ .

Para que se possa lidar com espaços de características de grande dimensão (potencialmente infinita), deve-se usar o conceito de funções *kernel*, em um procedimento conhecido na literatura como truque *kernel* (*kernel trick*). Uma função *kernel* K é uma função que aceita dois pontos no espaço de entradas, \mathbf{x}_i e \mathbf{x}_j , e fornece o produto escalar $\Phi_i(\mathbf{x}_i)\Phi_j(\mathbf{x}_j)$ no espaço de características:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi_i(\mathbf{x}_i)\Phi_j(\mathbf{x}_j). \quad (2.32)$$

As funções *kernel* são úteis, pois:

- são em geral mais simples que as funções Φ originais;
- na resolução do problema dual da minimização SVM no espaço de características, pode-se usar apenas o produto interno e abrir mão da avaliação individual das funções Φ , o que torna o *kernel* útil na utilização de espaços complicados e mesmo de dimensão infinita (LORENA e CARVALHO, 2003).

Esse último item, aliado ao fato de que os cálculos durante o treinamento são realizados apenas com os vetores suporte e não com o conjunto inteiro de dados,

torna possível a implementação da classificação não linear com razoável eficiência computacional.

Existem vários tipos de *kernel*, dentre os quais destacam-se (GERON, 2017):

- Linear: $K(\mathbf{a}, \mathbf{b}) = \mathbf{a}^T \mathbf{b}$
- Polinomial: $K(\mathbf{a}, \mathbf{b}) = (\gamma \mathbf{a}^T \mathbf{b} + r)^d$
- RBF gaussiano: $K(\mathbf{a}, \mathbf{b}) = \exp(-\gamma \|\mathbf{a} - \mathbf{b}\|^2)$
- Sigmóide: $K(\mathbf{a}, \mathbf{b}) = \tanh(\gamma \mathbf{a}^T \mathbf{b} + r)$

A Figura 2.12 ilustra a aplicação da classificação com uso do *kernel* RBF em um conjunto não linearmente separável. O algoritmo é capaz de efetuar qualquer separação arbitrariamente, como é possível perceber com a criação de regiões de classificação exclusivas para adequar pontos na região não linearmente separável, no gráfico do canto inferior direito. Apesar de essa capacidade ser um recurso poderoso, é preciso tomar cuidado, pois é comum ocorrer um problema conhecido como *sobreajuste*, em que o modelo é flexível demais e consegue ajustar variações aleatórias e ruídos, interpretando-os como padrões a serem aprendidos²².

SVM – Classificação não supervisionada

Dois métodos são utilizados para extensão do SVM para o caso da classificação não supervisionada: SVDD (*support vector domain description*), proposto por TAX e DUIN (1999), e *one-class* SVM, proposto por SCHOLKOPF *et al.* (2001). A diferença entre os dois está na forma geométrica utilizada como modelo de separação: enquanto este utiliza um hiper-plano, aquele utiliza uma esfera.

A Figura 2.13, adaptada do manual do pacote computacional *scikit-learn* (PEDREGOSA *et al.*, 2011), ilustra a aplicação do método *one-class* SVM para a adequação de um modelo multimodal a um conjunto de dados. Os dados de treinamento e de teste regulares são provenientes de uma distribuição multimodal (obtida por amostragem de uma mistura de gaussianas) e os dados irregulares, provenientes da distribuição uniforme. Pode-se perceber que os dados de teste irregulares caem em sua maioria fora da região definida pelo modelo SVM.

²²O *subajuste* é o problema contrário: o modelo é simples demais para aprender padrões demasiado complexos.

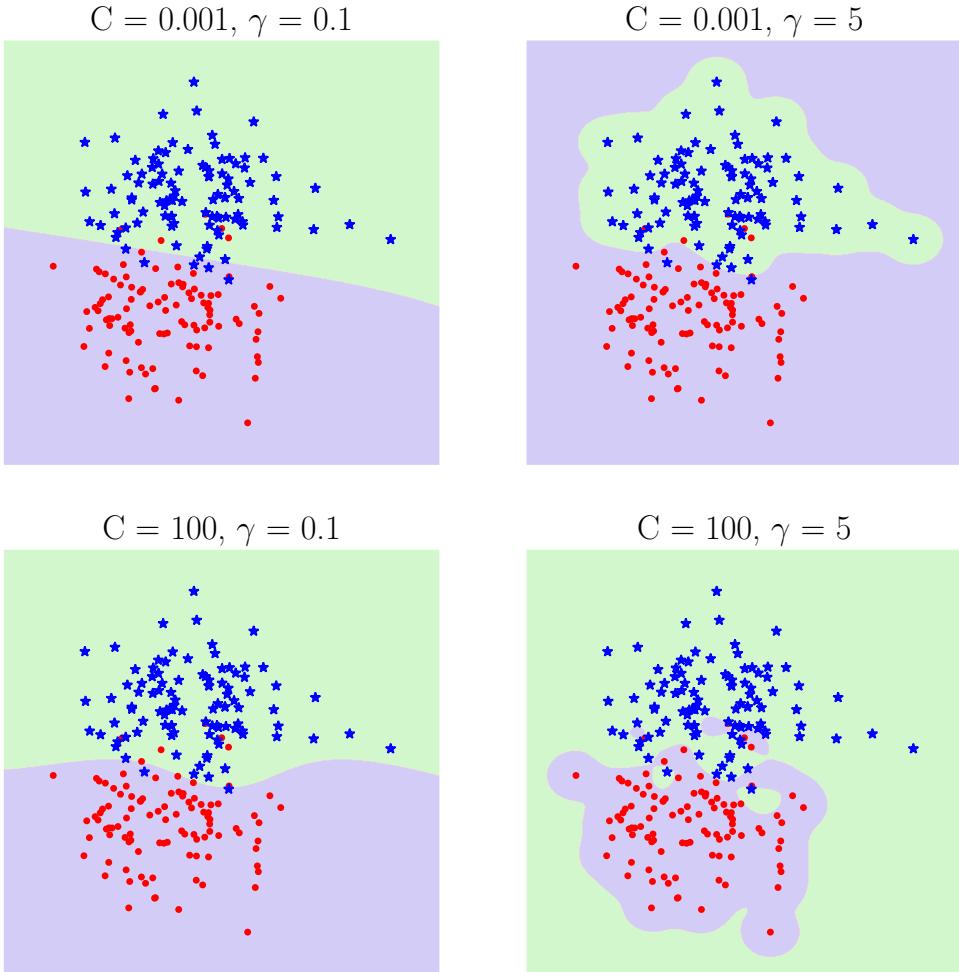


Figura 2.12: Ilustração em espaço bidimensional da resolução do problema de classificação binária com contornos não lineares utilizando SVM e *kernel* RBF gaussiano.

SVM – Regressão

O método SVM também pode ser usado para regressão, como proposto por DRUCKER *et al.* (1997). Nesse caso, o objetivo do algoritmo passa a ser o oposto ao da classificação: ao invés de maximizar a margem entre o hiperplano e os pontos, deve-se diminui-la. Para mais detalhes, checar SMOLA e SCHOLKOPF (2004).

SVM – Aplicações em monitoramento de processos

A aplicação do SVM em problemas de monitoramento ocorre tipicamente de forma supervisionada, por meio do uso da classificação para o diagnóstico de falhas (o que requer bancos de dados com informações abundantes sobre as falhas) (CHIANG *et al.*, 2004; YIN *et al.*, 2014b), ou não supervisionada, por meio da aplicação do

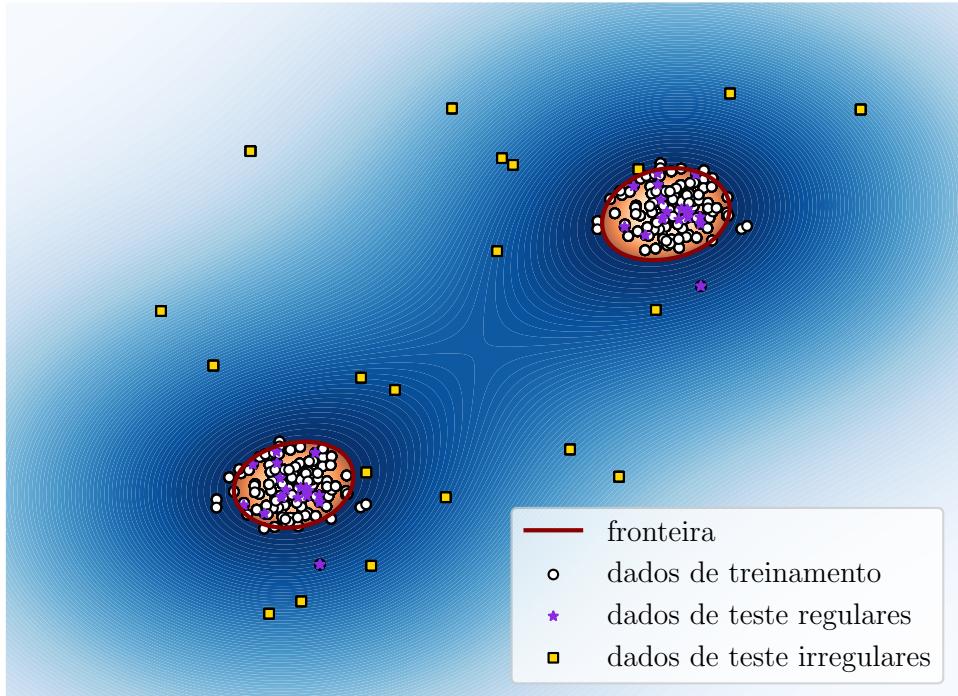


Figura 2.13: Ilustração da criação de fronteira multimodal. O gradiente de cores tem a ver com a confiança da classificação, relacionada diretamente à distância para a fronteira de decisão.

one-class SVM (GE e SONG, 2011; MAHADEVAN e SHAH, 2009; XIAO *et al.*, 2016) ou SVDD (GE *et al.*, 2011; LI e QIN, 2016).

Diversas modificações e extensões no algoritmo tradicional foram propostas para lidar com problemas específicos da área de monitoramento. Por exemplo, XIAO *et al.* (2014) propuseram metodologias para selecionar parâmetros do *kernel* gaussiano para aplicações em detecção de falhas. MD NOR *et al.* (2017) integraram o SVM com a análise discriminante de Fisher e *wavelets* para aplicação em problemas multiescala. MAHADEVAN e SHAH (2009) e YANG e HOU (2016) integraram o SVM com a técnica de seleção de variáveis RFE (*recursive feature elimination*). LEE e KIM (2018) propuseram uma metodologia adaptativa para aplicação em processos não estacionários. YUAN *et al.* (2020) apresentaram um algoritmo de SVDD com poda para lidar com *outliers* e aumentar a robustez. ZHANG e LI (2020) propuseram um método SVDD em duas etapas com otimização por algoritmo genético para lidar com características dinâmicas, não lineares e não gaussianas. ZHANG *et al.* (2022b) desenvolveram uma metodologia baseada em SVDD para lidar com dados incompletos advindos de unidades industriais com múltiplas unidades locais.

É interessante destacar que o SVM tem aplicação bem mais disseminada na área de engenharia mecânica do que na engenharia química (WIDODO e YANG, 2007; YIN e HOU, 2016), possivelmente por conta das características dos dados

provenientes de máquinas mecânicas, que contêm menos redundância e colinearidade do que dados de processos químicos.

2.5.5 Outras técnicas

A literatura da área de aprendizado de máquina é rica e disponibiliza várias outras técnicas, dentre as quais vale destacar:

- **k-Vizinhos mais próximos (*k*-Nearest Neighbors, **kNN**):** nesse método, a classificação de um ponto no espaço de variáveis é efetuada com base na classe da maioria dos k pontos mais próximos. Trata-se de um dos métodos mais simples e eficazes de aprendizado de máquina, adequado para conjuntos de dados não tão grandes em que não se conhece a distribuição de probabilidades (MD NOR *et al.*, 2019). Exemplos de aplicações incuem os de HE e WANG (2007), em que se aplica a classificação k -NN a um processo de manufatura de semicondutores; STOCKMANN *et al.* (2012), em que se propõe a identificação de falhas que afetam a planta inteira por meio da aplicação do algoritmo k -NN para a estimativa de atrasos; WANG *et al.* (2015a), ZHOU *et al.* (2016b), que propõem métodos de diagnóstico inspirados nas ideias de reconstrução e contribuição, respectivamente; SI *et al.* (2020), que utilizam um algoritmo de *Dynamic time warping* (DTW)²³ para definir a distância de Mahalanobis entre as séries temporais a ser usada na técnica; e FENG e LI (2020), que propõem um método para lidar com taxas de amostragem variáveis.
- **Métodos baseados em árvore:** a árvore de decisão é uma representação, baseada em estrutura de árvore, de um número de possíveis caminhos de decisão e da saída resultante de cada caminho (GRUS, 2015). O uso direto da árvore de decisão como algoritmo de aprendizado tem a grande vantagem da completa transparência e interpretabilidade da decisão tomada. Talvez por conta de sua simplicidade, não há muitas aplicações desse tipo na área de monitoramento de processos, sendo mais comuns na área de controle de qualidade (GUH e SHIUE, 2008; HE *et al.*, 2013). Uma estratégia comum é o uso de um comitê (*ensemble*) de árvores, como no algoritmo de *florestas aleatórias* (*random forests*), em que várias árvores são treinadas com porções aleatórias do conjunto de dados de treino e o resultado final resulta de um valor representativo (por exemplo, a média dos resultados de cada árvore). Tal abordagem é mais comumente

²³*Dynamic time warping* é uma técnica para encontrar o alinhamento ótimo de séries temporais por meio da minimização dos efeitos de deslocamento e distorção no tempo. Para mais informações, consultar BERNDT e CLIFFORD (1994); MULLER (2007); SENIN (2008).

encontrada em aplicações de monitoramento de processos, como apresentado por ALDRICH e AURET (2010); AURET e ALDRICH (2010); FAN *et al.* (2020); GAJJAR *et al.* (2020); LIU e GE (2019, 2018); NOGUEIRA *et al.* (2022); PUGGINI *et al.* (2015); SHRIVASTAVA (2021); XU *et al.* (2021b); ZHENG *et al.* (2022b).

- **Aprendizado de variedades:** nessa classe de técnicas, mais conhecida pelo termo em inglês *manifold learning*, o objetivo é efetuar redução de dimensionalidade dos dados preservando as relações não lineares entre eles (AYESHA *et al.*, 2020; GISBRECHT e HAMMER, 2015). As técnicas funcionam projetando os dados em variedades, subespaços topológicos que podem ser considerados localmente como euclídeanos. O mecanismo é útil para modelar a estrutura geométrica local dos dados, indo além das limitações de modelos como o PCA, que efetuam redução de dimensionalidade apenas usando a estrutura euclidiana global. Exemplos de aplicações na área de monitoramento incluem os de HE *et al.* (2018), que desenvolveram um método baseado em funções *kernel* especialmente designado para processos em batelada; LÓPEZ *et al.* (2020), que utilizaram aprendizado de variedades em conjunto com métodos de agrupamento para alinhamento temporal e identificação de fases de processos em batelada; e REIS *et al.* (2021), que propuseram um novo índice de monitoramento para cartas de controle baseadas em aprendizado de variedades. Uma tendência nos últimos anos é a proposição de metodologias que representam simultaneamente estruturas locais e globais nos dados, como evidenciado pelos estudos de CUI *et al.* (2022); FU e LUO (2019); LI *et al.* (2022e); LUO *et al.* (2016); YU (2016); ZHAN *et al.* (2019); ZHANG *et al.* (2011).
- **Dicionários esparsos:** primeiramente proposto por OLSHAUSEN e FIELD (1997) para aplicações em problemas de reconhecimento de imagens, o aprendizado por dicionários esparsos utiliza bases de dimensão maior que o espaço original para criação de uma representação esparsa dos dados (MAIRAL *et al.*, 2008). A base é chamada de *dicionário* e cada elemento da base é chamado de *átomo*. MAIRAL *et al.* (2008) afirmam que a representação esparsa leva a uma melhor captura de correlações de ordem superior. A aplicação desse tipo de metodologia em problemas de monitoramento é recente. WU *et al.* (2012) utilizaram o método de classificação por representação esparsa para detecção e diagnóstico de falhas; REN e LV (2014) ilustraram a aplicação de dicionários a um processo de fabricação de semicondutores; NING *et al.* (2015) desenvolveram gráficos de contribuição esparsos; PENG *et al.* (2017) propuseram a aplicação para detecção de falhas em processos multimodais; HUANG *et al.* (2019) propuseram o uso de funções *kernel* para captura da não

linearidade; YANG *et al.* (2019) desenvolveram uma metodologia robusta para monitoramento de processos com multimodalidade, ruído e *outliers*; HUANG *et al.* (2020b) apresentaram uma metodologia para decomposição de sistemas de elevada dimensão em módulos, de modo a facilitar a detecção de pequenas falhas; HUANG *et al.* (2021) propuseram uma metodologia de computação distribuída para lidar com dados no cenário *big data*; e HUANG *et al.* (2022) desenvolveram uma técnica adaptativa para lidar com processos multimodais.

- **Métodos baseados em lógica difusa:** a lógica difusa é uma generalização da lógica booleana, em que os valores de verdade das variáveis podem se encontrar em um intervalo contínuo entre 0 (falso) e 1 (verdadeiro). Exemplos de trabalhos que desenvolvem sistemas de monitoramento baseados em lógica difusa incluem: LEMOS *et al.* (2013), que desenvolveram um sistema de classificação adaptativo capaz de incorporar informações sobre modos de operação; RAMOS *et al.* (2018), em que se desenvolveu um algoritmo de agrupamento baseado em lógica difusa aplicado no *benchmark Tennessee Eastman*; RIBEIRO *et al.* (2020), que aplicaram o algoritmo genético e técnicas de agrupamento baseado em lógica difusa para identificação de padrões em séries temporais multivariadas; e ELSHENAWY *et al.* (2022), que combinaram as técnicas k-NN e C-means difuso em um novo algoritmo de agrupamento. Sistemas de lógica difusa podem ser integrados à capacidade de aprendizado das redes neurais, como descrito na Seção 2.5.3.
- **Redes bayesianas:** uma rede bayesiana é um modelo gráfico acíclico direto (*directed acyclic graph*, DAG) que representa as relações de dependência probabilística entre um grupo de variáveis (JUN e KIM, 2017; MD NOR *et al.*, 2019). Por conta disso, é capaz de lidar bem com incertezas e proporcionar inferências causais, descrevendo relações entre causas e sintomas (CAI *et al.*, 2009; LAMPIS e ANDREWS, 2009). Exemplos de trabalhos da área de monitoramento relacionados a redes bayesianas incluem: LIU *et al.* (2010), que utilizaram conhecimentos de especialistas para montar uma rede bayesiana para o monitoramento de um reator tubular; VERRON *et al.* (2010), que aplicaram redes bayesianas para construção e decomposição causal da carta de controle T^2 ; YU e RASHID (2013), que desenvolveram um método de monitoramento para processos dinâmicos; LIU *et al.* (2014a), que utilizaram redes bayesianas para filtrar o efeito de espalhamento que ocorre no diagnóstico de falhas pelo método da contribuição; CHEN e GE (2020), que desenvolveram uma rede bayesiana robusta para lidar com dados de baixa qualidade; LUCKE *et al.* (2020), que propuseram uma rede de classificação bayesiana baseada em alarmes estatísticos para lidar com o efeito do ruído; NHAT *et al.* (2020), que aplicaram

uma rede bayesiana para detecção de *kicks* (fluxos indesejados da formação do poço para o interior) de hidrocarbonetos em operações de perfuração; LIU *et al.* (2022c), que propuseram um método combinando análise mecanística e de transição de estados para reconhecer caminhos de propagação de falhas de processo; e YANG *et al.* (2022), que propuseram uma forma gráfica de representar as interações causais de um processo utilizando uma rede bayesiana interpretável. Uma tendência recente é a integração de redes bayesianas com métodos tradicionais, como PCA e ICA (GE e SONG, 2009; JIANG *et al.*, 2016a; WANG *et al.*, 2017c). Para uma revisão em formato de tutorial sobre a aplicação de modelos bayesianos no problema de monitoramento de processos, consulte RAVEENDRAN e HUANG (2020).

Para além das metodologias tradicionais, uma tendência nos últimos anos vem sendo a aplicação de técnicas menos restritas por estruturas de modelos e mais focadas na exploração de padrões intrínsecos dos próprios dados, dentre as quais destacam-se:

- **Agrupamento:** nessa abordagem, que constitui uma das metodologias de aprendizado não-supervisionado mais populares, dados são divididos em conjuntos com base em alguma medida de similaridade (EZUGWU *et al.*, 2022). Aplicações na área de monitoramento de processos incluem: BARRAGAN *et al.* (2016), que aplicaram uma metodologia *fuzzy* para efetuar agrupamento com base em *wavelets* e a aplicação da técnica PCA; FONTES e BUDMAN (2017), que propuseram um método híbrido que combina medidas de similaridade euclidiana e baseada em PCA; e ELSHENAWY *et al.* (2022), que combinaram as técnicas **kNN** e C-means difuso em um novo algoritmo de agrupamento.
- **Análise lógica de dados (*Logical analysis of data, LAD*):** nessa técnica, proposta por HAMMER (1986), os dados são binarizados para posterior descoberta de padrões interpretáveis por meio da aplicação de modelagem do tipo booleana (LEJEUNE *et al.*, 2019). A vantagem da técnica é a interpretabilidade dos padrões descobertos, o que torna mais viável a transformação da informação em *entendimento* e *conhecimento* humanos. Exemplos de aplicações da técnica na área de monitoramento podem ser encontradas em MORTADA *et al.* (2012), BAI *et al.* (2018), RAGAB *et al.* (2017, 2018, 2019) e KHALIFA *et al.* (2023).
- **Empirical machine learning:** essa abordagem, desenvolvida em trabalhos recentes (ANGELOV, 2014; ANGELOV *et al.*, 2016, 2017; ANGELOV, 2013; ANGELOV e GU, 2019), tem com base a densidade dos dados no espaço de variáveis e resulta em algoritmos totalmente baseados em dados, livres

de modelos probabilísticos e hiperparâmetros a serem definidos *a priori* pelo usuário. Novos conceitos são definidos na teoria proposta, como *proximidade cumulativa* (medida de centralidade), *excentricidade* (medida de anomalia) e *tipicidade* (medida de densidade). Algumas aplicações em monitoramento de processos foram desenvolvidas (BEZERRA *et al.*, 2016; COSTA *et al.*, 2014, 2015a,b; GERMANO *et al.*, 2016).

- **Matrizes de recorrência e matrizes de distância:** Matrizes de recorrência e matrizes de distância são representações bidimensionais de padrões de distância de dados. Aplicações dessa metodologia em monitoramento de processos vêm sendo propostas em combinação com várias técnicas, como redes neurais convolucionais (OLIVIER e ALDRICH, 2020), quantificação de recorrência (ZIAEI-HALIMEJANI *et al.*, 2021a,b,c,d) e análise de textura (BARDINAS, 2018; BARDINAS *et al.*, 2018; KOK *et al.*, 2022). No Capítulo 5 do presente trabalho é desenvolvida uma abordagem para uso de matrizes de distância em problemas de monitoramento.

2.6 Aplicações na indústria

Na presente seção, discute-se o monitoramento de processos baseado em dados do ponto de vista de aplicações industriais reais ou estudos envolvendo dados reais de planta. Algumas aplicações em áreas selecionadas da indústria encontram-se na Tabela 2.1. No corpo do texto são apresentados o desenvolvimento histórico das aplicações industriais e descritos os principais trabalhos.

Tabela 2.1: Monitoramento de processos: algumas aplicações industriais reais ou estudos envolvendo dados reais de planta.

Área	Referências
Nuclear	WISE <i>et al.</i> (1988, 1991), WISE e GALLAGHER (1996), GALLAGHER <i>et al.</i> (1996)
Petroquímica	SLAMA (1991), THOMAS <i>et al.</i> (1996), ALGHAZ-ZAWI e LENNOX (2008), LIU <i>et al.</i> (2009), KIRAN <i>et al.</i> (2012), SUN <i>et al.</i> (2013), KUMAR <i>et al.</i> (2020)

Continua na próxima página

Tabela 2.1 – *Continuada da página anterior*

Área	Referências
Polimerização	PIOVOSO <i>et al.</i> (1992), PIOVOSO e KOSANOVICH (1994), KOSANOVICH <i>et al.</i> (1994), KOURTI <i>et al.</i> (1995), MACGREGOR e KOURTI (1995), NOMIKOS e MACGREGOR (1995), KOURTI <i>et al.</i> (1996), MARTIN <i>et al.</i> (1996), MARTIN e MORRIS (1996), ALBERT <i>et al.</i> (1996), NEOGI e SCHLAGS (1998), RANNAR <i>et al.</i> (1998), TATES <i>et al.</i> (1999), QIN <i>et al.</i> (2001), KUMAR <i>et al.</i> (2003), SHARMIN <i>et al.</i> (2008), SIVALINGAM <i>et al.</i> (2015), PARK <i>et al.</i> (2020a)
Papel e Celulose	MILLER <i>et al.</i> (1993), TEPPOLA <i>et al.</i> (1998b), BISSESSUR <i>et al.</i> (1999), KESAVAN <i>et al.</i> (2000), MILETIC <i>et al.</i> (2004), CHIOUA <i>et al.</i> (2016), PARENTE <i>et al.</i> (2019)
Mineração	HODOUIN <i>et al.</i> (1993), BERGH (2010), LE <i>et al.</i> (2020)
Bioprocessos	ALBERT <i>et al.</i> (1996), IGNOVA <i>et al.</i> (1997), GREGERSEN <i>et al.</i> (1997), MONTAGUE <i>et al.</i> (1998), UNDEY <i>et al.</i> (2000), ALBERT e KINLEY (2001), LENNOX <i>et al.</i> (2001), KARIM <i>et al.</i> (2003), NUCCI <i>et al.</i> (2009, 2010)
Química	SIMOGLOU <i>et al.</i> (1997), GOULDING <i>et al.</i> (2000), RALSTON <i>et al.</i> (2001), CHIANG e COLE-GROVE (2007), DARKOW <i>et al.</i> (2014), DUNIA <i>et al.</i> (2014), LI <i>et al.</i> (2020d),
Semicondutores	GALLAGHER <i>et al.</i> (1997), WISE <i>et al.</i> (1997), WISE <i>et al.</i> (1999), QIN <i>et al.</i> (2006)
Águas e Resíduos	ROBERTSON <i>et al.</i> (1998), TEPPOLA <i>et al.</i> (1998a), TEPPOLA e MINKKINEN (1999), ROSEN e LENNOX (2001), GARCIA-ALVAREZ <i>et al.</i> (2009), LEE <i>et al.</i> (2009), PADHEE <i>et al.</i> (2012), YAMANAKA <i>et al.</i> (2012), HAIMI <i>et al.</i> (2016), HARROU <i>et al.</i> (2018), KLANDERMAN <i>et al.</i> (2020), LIU <i>et al.</i> (2020a) ZHOU <i>et al.</i> (2021), LIU <i>et al.</i> (2021a), HAN <i>et al.</i> (2021)

Continua na próxima página

Tabela 2.1 – *Continuada da página anterior*

Área	Referências
Metalúrgica	TAYLOR (1998), DUDZIC <i>et al.</i> (2000), KOURTI (2002), MILETIC <i>et al.</i> (2004), ZHANG e DUDZIC (2006), KANO e NAKAGAWA (2008), MILETIC <i>et al.</i> (2008), VANHATALO (2010), JI <i>et al.</i> (2010), MAJID <i>et al.</i> (2011, 2012), TESSIER <i>et al.</i> (2012), QIANG LIU <i>et al.</i> (2013), MAJID <i>et al.</i> (2015), ZHANG <i>et al.</i> (2016), ZHOU <i>et al.</i> (2016a), MA <i>et al.</i> (2017), YAO <i>et al.</i> (2018), ZHOU <i>et al.</i> (2020), ZHANG <i>et al.</i> (2021b), RIPPON <i>et al.</i> (2021)
Farmacêutica	UNDEY e ÇINAR (2002), MARKL <i>et al.</i> (2013), TÔRRES <i>et al.</i> (2015), KHARBACH <i>et al.</i> (2017), SILVA <i>et al.</i> (2017, 2018), TÔRRES <i>et al.</i> (2018), DUMAREY <i>et al.</i> (2018), ZOMER <i>et al.</i> (2018), COLUCCI <i>et al.</i> (2019), SILVA <i>et al.</i> (2019)
Madeireira	AMAZOUZ e PANTEA (2006)
Automobilística	FERRER (2007), SCHMID <i>et al.</i> (2020), SANCHEZ-MARQUEZ e JABALOYES VIVAS (2020)
Plásticos (3 ^a geração)	KAZMER <i>et al.</i> (2008)
Termoelétrica	BERRIOS <i>et al.</i> (2009), FERREIRA <i>et al.</i> (2015), TANG e ZHANG (2020), FONTES <i>et al.</i> (2021)
Hidrelétrica	LAZZAROTTO <i>et al.</i> (2016)
Cimento	BAKDI <i>et al.</i> (2017)
Petróleo	CLAVIJO <i>et al.</i> (2019), SAD <i>et al.</i> (2019)
Biorrefinarias	NACHTERGAELE <i>et al.</i> (2020)

Aplicações da técnica PCA existem desde o fim da década de 1980. O trabalho que introduziu o uso do PCA para monitoramento ilustrou a técnica com a aplicação em dados de um equipamento fundidor de cerâmica para reprocessamento de resíduos da indústria nuclear (WISE *et al.*, 1988), em que se mostrou pela primeira vez que a aplicação do PCA em dados históricos pode levar a um aumento do conhecimento sobre um processo industrial. No caso em questão, isso se deu por meio da análise do peso de cada variável na geração dos componentes principais. Resultados da aplicação industrial em tempo real nesse equipamento foram reportados por WISE

et al. (1991).

Vários relatos de aplicação do PCA e PLS em sistemas industriais surgiram nos anos seguintes. Em particular, SLAMA (1991) estudou a operação de um leito fluidizado de craqueamento catalítico e a seção de fracionamento de uma refinaria, relatando vários aspectos práticos relacionados à análise de dados reais, como validade das hipóteses acerca dos dados (em especial sobre a ausência de autocorrelação), tempo de amostragem e relação com as constantes de tempo do processo, entre outros. Os resultados do trabalho contribuíram para o aumento do conhecimento sobre o processo, em questões como diferenças entre os períodos de operação e detecção de um *drift* na dinâmica global da planta, a despeito de os produtos estarem sendo produzidos na janela de especificação. No entanto, a existência de não linearidades prejudicaram o desempenho do sistema de monitoramento em tempo real.

Um tipo de aplicação comum dessa época está relacionado aos métodos *multiblock* e *multi-way* em sistemas de polimerização em batelada e semi-batelada (KOURTI *et al.*, 1995; MACGREGOR e KOURTI, 1995; NOMIKOS e MACGREGOR, 1995). Esses trabalhos, no entanto, a despeito de usarem dados reais, focam mais na metodologia e apresentação matemática das técnicas do que na aplicação prática. O mesmo não se pode dizer dos trabalhos da companhia *DuPont* (KOSANOVICH *et al.*, 1994; PIOVOSO *et al.*, 1992; PIOVOSO e KOSANOVICH, 1994), que datam da mesma época e se destacam pelo foco nos estudos de caso. PIOVOSO e KOSANOVICH (1994), por exemplo, discutiram o problema da falta de dados representativos e a resistência do pessoal da operação à realização de experimentos programados para mapeamento de toda a faixa de operação, principalmente por conta do medo da produção de quantidade excessiva de produtos fora de especificação.

KOURTI *et al.* (1996) apresentaram um levantamento das aplicações dos métodos PCA e PLS até então e duas novas aplicações em processos de polimerização contínuo e semi-batelada. Os autores enfatizaram a rápida aceitação e utilização do monitoramento com variáveis latentes pela indústria. Em um dos estudos de caso, relataram que o uso dessa metodologia indicou rapidamente variáveis que levavam a uma baixa pureza e/ou recuperação, tarefa que havia custado considerável tempo e esforço por parte dos engenheiros responsáveis. No outro estudo, realizou-se a detecção de uma falha causada por um erro de digitação do operador. Não se apresentaram, no entanto, estudos de caso com implementações em tempo real.

WISE e GALLAGHER (1996), em um artigo muito influente, consolidaram o trabalho desenvolvido pelo grupo em anos anteriores e apresentaram de forma didática para a comunidade de controle de processos a metodologia de monitoramento por meio da abordagem quimiométrica. Exemplos de aplicação presentes nesse

trabalho envolvem: (i) o equipamento fundidor de cerâmica, como em trabalhos anteriores dos mesmos autores (WISE *et al.*, 1988, 1991); (ii) o monitoramento da mitigação de resíduos nucleares em tanques de estocagem, em que se fez a identificação de tendências sutis por meio da análise dos componentes principais e (iii) o monitoramento de um processo batelada por meio de espectrometria no infravermelho próximo (*near-infrared*, NIR), utilizando as técnicas *evolving factor analysis* (EFA) e *multivariate curve resolution* (MCR).

Talvez por conta da influência do trabalho de WISE e GALLAGHER (1996), nos anos seguintes a variedade de trabalhos com exemplos de aplicação industrial das técnicas de modelagem em variáveis latentes aumentou de forma significativa. IGNOVA *et al.* (1997), por exemplo, aplicaram *multi-way* PCA e PLS a dados de um bioprocesso de fermentação, chegando à conclusão de que uma grande redução de custos seria possível por conta da capacidade dos modelos desenvolvidos de determinar a qualidade da semente alimentada ao processo e predizer a concentração do produto final. A aplicação em tempo real da metodologia, que foi integrada a um sistema especialista denominado *G2 real-time expert system*, foi reportada em trabalhos posteriores (LENNOX *et al.*, 2001; MONTAGUE *et al.*, 1998).

GURDEN *et al.* (1998) relataram a introdução das técnicas quimiométricas em uma planta piloto industrial. Dentre os resultados obtidos, vale destacar: (i) detecção e diagnóstico de falhas; (ii) o uso de componentes principais para produzir um gráfico da trajetória do processo, possibilitando o monitoramento do progresso da operação da planta, e (iii) uso de variáveis latentes em análise de correlação cruzada, de modo a aumentar a capacidade interpretativa da análise de efeitos relacionados a atrasos no tempo. Os autores enfatizaram o fato de que a experiência e o aumento do entendimento sobre o processo capacitaram o pessoal da planta a propor a instalação de novos sensores e analisadores, tendo em vista a vantagem competitiva que trariam.

TAYLOR (1998) reportou a aplicação do PCA para detecção de instabilidades no alto-forno da companhia *British Steel*, tendo aumentado consideravelmente o tempo de antecipação em relação aos métodos de detecção tradicionais. Duas características interessantes desse trabalho são: (i) a descrição da implementação em tempo real na planta, bem mais detalhada do que a presente em trabalhos anteriores; e (ii) ser um trabalho exclusivamente industrial, sem a participação de especialistas universitários como autores, o que não ocorria desde os trabalhos da *DuPont* na primeira metade da década de 90.

NEOGI e SCHLAGS (1998), ambos provenientes da indústria *Air Products and Chemicals*, aplicaram MPCCA e MPLS a um processo em batelada de polimerização em emulsão. Pela primeira vez o critério usado para comparação das trajetórias de

cada batelada foi a extensão de reação, e não o tempo de batelada. Apesar de não ter havido implementação em tempo real, a aplicação dos métodos de monitoramento nos dados históricos resultou na detecção e diagnóstico de falhas bem-sucedidos. Além do mais, obteve-se um modelo PLS capaz de prever as viscosidades do produto, o que resultou na melhora do entendimento sobre o processo e no aumento da eficiência da operação.

MARTIN *et al.* (1998) reportaram a aplicação bem-sucedida do PCA a dados de um reator de leito fluidizado e discutiram questões em aberto, que consideravam essenciais para que o uso dessas técnicas se tornasse rotineiro na indústria, a saber: (i) o desenvolvimento de esquemas de monitoramento com quantidade mínima de dados de processo, (ii) o uso de técnicas multibloco para monitoramento de toda a planta, (iii) o desenvolvimento de modelos genéricos para o monitoramento de múltiplos produtos ou *grades*, (iv) o estudo do efeito da dinâmica e dos laços de controle no desempenho do sistema de monitoramento e no potencial para mascarar falhas, (vi) o estudo da sensibilidade do PCA e do PLS e da adequabilidade de seus limites de confiança e (vii) a extensão dos limites de confiança definidos pela densidade para o monitoramento de processos dinâmicos.

BISSESSUR *et al.* (1999) aplicaram duas abordagens para detecção de falhas em uma planta de fabricação de papel. A primeira se referiu ao monitoramento de um único equipamento, utilizando redes neurais aplicadas a dados de vibração. A segunda se referiu à utilização do PCA para o monitoramento do processo como um todo. Vale destacar a importante observação de que as falhas foram melhor detectadas analisando-se os componentes principais de menor ordem, já que os primeiros componentes capturavam variações nos balanços de massa e energia e movimentos de *setpoints* da planta. Trata-se de um resultado que vai de encontro ao que é normalmente admitido na utilização da técnica PCA em relação à importância relativa dos componentes principais.

DUDZIC *et al.* (2000) reportaram com detalhes sem precedentes a implementação das técnicas PCA e PLS a processos industriais da *Dofasco* (da área siderúrgica), que, segundo os autores (todos empregados da companhia), ocorria desde 1993. Duas aplicações em específico foram descritas: (i) o monitoramento de um processo de fundição na área de moldagem da principal máquina de lingotamento contínuo da companhia e (ii) um sistema de controle para determinação da quantidade ótima de reagentes necessária para remoção de enxofre do ferro-gusa. Os autores destacaram a progressão, ao longo da década de 1990, de análises *off-line* para implementações *on-line*, suportadas pelo contato próximo com o Consórcio de Controle Avançado da Universidade McMaster. As técnicas empregadas ofereceram um balanço adequado entre facilidade de desenvolvimento e *insights* adquiridos sobre o processo (uma

vantage sobre a aplicação de modelos totalmente fenomenológicos ou puramente empíricos) e apresentaram robustez perante ruído ou dados faltantes. Destacou-se a importância do pré-tratamento dos dados (seleção e aplicação de filtros, agrupamento de medições selecionadas para remoção de tendências, tratamento de dados faltantes, dentre outros). No monitoramento do processo de fundição, aplicou-se o PCA para reduzir a dimensionalidade da matriz de treino (já com aplicação do atraso para tratamento da dinâmica) de 240 para 10 variáveis. Duas cartas de controle baseadas no índice T^2 e uma baseada no índice Q foram disponibilizadas para o operador, com um esquema de alarmes progressivos (cores amarela e vermelha, respectivamente). Os autores reportaram uma redução de 50% nas paradas após dois anos de funcionamento do sistema de monitoramento e o alcance da maior produtividade em 12 anos de operação da máquina.

ALBERT e KINLEY (2001) aplicaram PCA e PLS para o monitoramento da biossíntese por fermentação em batelada do antibiótico tilosina. Aspectos interessantes do relato envolvem a baixa qualidade dos dados, que impuseram desafios significativos para o pré-tratamento, e a variação dessa qualidade em diferentes pontos da planta, o que, segundo os autores, constitui um impedimento para a integração da tecnologia de monitoramento à planta.

KARIM *et al.* (2003) relataram experiências na aplicação de métodos baseados em dados na indústria de bioprocessos. Desafios particulares dessa indústria foram destacados, como a extração de informação biologicamente significativa a partir de poucos estados de processo medidos em tempo real. Essa extração é complicada pelo fato de que os dados medidos geralmente representam condições do biorreator, enquanto a informação de interesse diz respeito à condição fisiológica da célula, o que tradicionalmente requer dispendiosas análises *offline*. Os autores mostraram como a aplicação do MPCA ajudou a diminuir essas limitações por meio de um aumento na capacidade de interpretação dos dados.

KUMAR *et al.* (2003) reportaram a aplicação do PCA a um processo de polimerização em alta pressão. Mostrou-se com detalhes que a técnica é capaz de capturar várias falhas de natureza multivariada (relacionadas a correlações), em que as variáveis se mantinham em seus níveis normais de operação nas cartas univariadas. Transições de *grade* e mudanças no processo foram melhor visualizadas no espaço de variáveis latentes. Apesar de ter sido informada a instalação de um sistema de monitoramento na planta, a operação *online* não foi descrita.

Em 2003, ocorreu em Ontario, Canadá, um simpósio denominado “Detecção de situações anormais e métodos de projeção: aplicações industriais”. KOURTI (2005a) reportou um resumo do evento, com descrições das apresentações realizadas. Estive-

ram presentes profissionais das companhias *Dofasco*, *DuPont*, *Perceptive Engineering* e do consórcio COREM (*Mineral Research Consortium*), além de pesquisadores de várias universidades. Há interessantes resumos das quatro palestras proferidas, sumarizadas a seguir:

- Paul Nomikos, da *DuPont*, relatou a experiência bem-sucedida no uso do PCA e PLS para análises *off-line* (como identificação e eliminação de causas de variabilidade, melhorias em processos fenomenológicos e novas estratégias de controle), mas também destacou a dificuldade na implementação *on-line*, devido ao desafio da manutenção de longo prazo dos modelos.
- Theodora Kourtzi fez várias recomendações práticas para aplicação de sistemas de monitoramento baseados em dados. Destacou a importância da aceitação dos sistemas pelos operadores, citando o fato de o mesmo sistema em unidades semelhantes da mesma companhia poderem ter telas potencialmente diferentes, por conta dos requerimentos dos operadores. Recomendou o começo das implementações com problemas pequenos, evitando o monitoramento inicial já de toda a planta. Enfatizou a importância do conhecimento sobre o processo, que ajuda a decidir sobre transformações e pesos aplicados nas variáveis e fornece informações sobre a qualidade dos dados obtidos. Relatou que, para monitorar desvios, um modelo linear costuma ser suficiente. Finalizou citando os fatores que potencialmente motivam uma indústria a aplicar um sistema de monitoramento multivariável: retorno financeiro, questões de segurança e regulamentações ambientais.
- Barry Lenox destacou que muitos dos projetos de monitoramento são os primeiros do tipo na indústria em que são desenvolvidos e que, após a primeira aplicação, é relativamente simples transferi-la para outros processos similares. Relatou também que sua experiência mostrou que a confiança dos operadores, engenheiros e gerentes nos sistemas pode evoluir a tal ponto, que paradas automáticas podem ser programadas apenas com informações fornecidas pelos sistemas de monitoramento.
- Vit Vaculisk, da *Dofasco*, enfatizou os perigos de uma aplicação mal desenvolvida, as barreiras para o desenvolvimento (disponibilidade e conteúdo dos dados, conhecimento técnico do pessoal e aceitação da tecnologia) e a decisão às vezes difícil entre a compra ou desenvolvimento dos sistemas dentro da própria companhia.

MILETIC *et al.* (2004), trabalhando para as companhias canadenses *Dofasco* e *Tembec* (da área de papel e celulose), apresentaram um artigo relatando a perspectiva

industrial da implementação em linha das técnicas PCA e PLS. Os autores propuseram um método genérico de desenvolvimento de sistemas de monitoramento, composto de etapas *off-line* e *on-line*. A etapa *off-line* consiste da seleção de dados e preparação, desenvolvimento e avaliação do modelo. A etapa de seleção de dados, apesar de talvez ser vista como trivial por pesquisadores do mundo acadêmico, é descrita pelos autores como dispendiosa e vital para o sucesso da aplicação. A etapa *on-line* consiste do desenvolvimento do sistema, integração e manutenção. Os autores enfatizaram a importância da integração com os operadores nessa fase, de modo a entender suas necessidades e como ocorrerá a interação com o sistema. A metodologia proposta é ilustrada com três exemplos de aplicações: (i) detecção de falhas em uma máquina de lingotamento contínuo usando PCA, (ii) desenvolvimento de modelos preditivos para um processo de dessulfurização usando PLS e (iii) detecção de falhas para um digestor contínuo usando PCA.

KOURTI (2005b) realizou uma revisão das aplicações industriais de monitoramento utilizando métodos de variáveis latentes. Tópicos interessantes tratados incluem como lidar com observações futuras de dados faltantes, modelos adaptativos, monitoramento de processos em batelada, monitoramento de transições, controle utilizando métodos de variáveis latentes, análise de imagens e arquivamento, compressão e reconstrução de dados.

ZHANG e DUDZIC (2006) descreveram mais um trabalho de implementação na *Dofasco*, dessa vez com a novidade de um esquema de sincronização de trajetórias para monitoramento de operações de transição. Como em trabalhos anteriores publicados pela companhia, vários detalhes interessantes de implementação foram descritos, como funcionamento dos alarmes e telas do sistema de controle.

QIN *et al.* (2006) propuseram um *framework* hierárquico e abrangente para controle e monitoramento de fábricas de semicondutores de nova geração à época. Conclusões interessantes sobre esse tipo de processo em específico foram obtidas, como a maior importância relativa de *drifts* em relação a falhas de sensores.

CHIANG e COLEGROVE (2007) relataram a implementação de PCA robusto na companhia *Dow Chemical* para monitoramento da manufatura do produto *Solid Epoxy Resins*, discutindo aspectos práticos como desempenho, manutenção e transferência de modelos. A implementação *on-line* foi descrita, com destaque para uma detecção de problemas na qualidade da matéria-prima efetuada pouco após o início do funcionamento do sistema.

KANO e NAKAGAWA (2008) relataram desenvolvimentos e aplicações do monitoramento baseado em dados na indústria siderúrgica, ilustrando com exemplos da companhia *Sumitomo Metals Kokura*. Os autores propuseram um novo método

capaz de lidar com informações qualitativas sobre qualidade e de relacionar condições operacionais ao rendimento da produção, além da análise simultânea de múltiplas unidades de processamento.

MILETIC *et al.* (2008) reportaram mais experiências no âmbito de aplicações da *Dofasco*. O foco deste trabalho, em particular, foi a estrutura computacional flexível que facilitou a manipulação dos dados e implementação de novos modelos. Foram descritas também extensões do sistema com outras técnicas, além de PCA e PLS, como redes bayesianas e técnicas de análise de imagens.

ALGHAZZAWI e LENNOX (2008) relataram o desenvolvimento de um sistema de monitoramento para um processo de refino de petróleo da *Saudi Aramco*, com o objetivo de acompanhar a dinâmica não estacionária do processo. Um método PCA recursivo foi utilizado e, para uma maior capacidade de isolamento de falhas, aplicou-se uma abordagem multibloco. Destacou-se a importância, para o bom desenvolvimento do sistema, do uso de conhecimento do pessoal que trabalha na planta. Não foi reportada implementação em tempo real.

VANHATALO (2010) descreveram a aplicação da técnica PCA para o monitoramento de um alto-forno em escala piloto. Os engenheiros responsáveis reportaram como ponto positivo da aplicação o fornecimento de uma visão global do estado térmico do equipamento, com consequente padronização da avaliação de sua condição e a possibilidade de desenvolvimento de critérios formais para quando se executar ações de controle. Como pontos negativos, reportou-se o alto custo de homem-hora relativo ao desenvolvimento, manutenção e interpretação dos modelos e a abstração desnecessária das representações multivariadas.

DARKOW *et al.* (2014) relataram a aplicação do método PCA para a detecção precoce de entupimentos ou bloqueios em um esgotador industrial. Os pontos-chave para o sucesso da aplicação foram: (i) implementação de etapas prévias à geração do modelo como a seleção de variáveis com base em conhecimento de especialistas e a remoção de *outliers* univariados por inspeção visual; (ii) envolvimento do pessoal da planta e de TI (tecnologia da informação) desde as etapas iniciais do projeto; (iii) desenvolvimento de uma interface gráfica de simples entendimento para diferentes usuários.

DUMAREY *et al.* (2018) descreveram a aplicação do método PCA para o monitoramento em tempo real da síntese contínua de um ingrediente ativo na indústria farmacêutica. Os autores identificaram causas para a disseminação lenta de métodos de monitoramento em tempo real na indústria farmacêutica, relacionadas principalmente às dificuldades de integração com os rígidos sistemas de controle de qualidade já existentes. A aplicação proposta resultou em um aumento do

entendimento do processo, incluindo a compreensão da variabilidade natural e a identificação de tipos e causas de falhas.

PATWARDHAN *et al.* (2019) reportaram aplicações envolvendo ciência de dados na empresa *Saudi Aramco*. Dentre elas, aplicações envolvendo monitoramento de processos foram reportadas, como monitoramento do desempenho de controladores e manutenção preditiva de equipamentos. Os autores ressaltaram a importância do gerenciamento de alarmes utilizando técnicas de análise de dados para maior eficácia dos sistemas de monitoramento junto aos operadores.

KLANDERMAN *et al.* (2020) reportaram a aplicação da técnica PCA para desenvolver uma metodologia adaptativa de monitoramento especialmente voltada para estações de tratamento de efluentes. Aspectos práticos descritos incluem: (i) aumento do entendimento da natureza das mudanças de processo por parte dos operadores; (ii) possibilidade de resultados inconsistentes com as avaliações dos operadores, mesmo quando os métodos funcionam de acordo com o esperado; (iii) recomendação de revisão prévia dos dados de treinamento por parte dos operadores e especialistas; (iv) recomendação de reinicializar o sistema de monitoramento após a detecção de uma falha para evitar contaminação com dados não normais.

KUMAR *et al.* (2020) relataram o desenvolvimento e aplicação de um sistema replicável baseado em modelagem em variáveis latentes para monitoramento de reformadores de plantas de produção de hidrogênio. Os autores ressaltaram a importância do retreinamento frequente dos modelos, propondo para esse fim uma metodologia baseada em *feedbacks* do usuário sobre falsos positivos.

Das descrições dos parágrafos anteriores, percebe-se que os principais trabalhos descrevendo implementações práticas de técnicas de monitoramento foram publicados na década de 2000, em particular os estudos de DUDZIC *et al.* (2000), MILETIC *et al.* (2004), KOURTI (2005a), ALGHAZZAWI e LENNOX (2008), KANO e NAKAGAWA (2008) e MILETIC *et al.* (2008). A partir da década de 2010, ocorre uma explosão de estudos na área de monitoramento de processos, com uma tendência de publicação cada vez maior de trabalhos com dados reais. No entanto, a maior parte desses trabalhos enfatiza análises *off-line* de técnicas novas ou variantes das técnicas tradicionais aplicadas nesses dados, com pouca ênfase nos aspectos práticos das implementações. Além do mais, apesar da crescente utilização de dados reais, a maior parte dos trabalhos ainda utiliza dados simulados, em especial do *benchmark Tennessee Eastman Process*, descrito na Seção 2.8. Também é importante perceber que os principais trabalhos que descrevem implementações práticas utilizam metodologias baseadas em variáveis latentes, em particular PCA e PLS, indicando que as demais técnicas ainda não permearam o setor de maneira intensa.

Em relação a estudos recentes envolvendo análises *off-line* de dados históricos reais, vale destacar alguns trabalhos, como de: SIVALINGAM *et al.* (2015), que desenvolveram uma metodologia para detecção de decomposição de reagentes em uma autoclave de copolimerização utilizando o método PCA; HARROU *et al.* (2018), que aplicaram métodos de aprendizado de máquina para o monitoramento de uma planta descentralizada de tratamento de efluentes; SAD *et al.* (2019), que aplicaram o método PCA para monitorar as propriedades físico-químicas de *blends* de óleo cru; CLAVIJO *et al.* (2019), que reportaram a aplicação bem sucedida do método CCA em dados históricos de processo para detecção e diagnóstico de uma falha de sensor em uma estação de medição fiscal de uma planta *onshore*, destacando o potencial de ganhos econômicos que pode resultar em uma aplicação em tempo real dessa natureza; NACHTERGAELE *et al.* (2020), que aplicaram diferentes técnicas de modelagem em variáveis latentes para desenvolver uma estratégia visando o aumento do entendimento sobre processos em biorrefinarias; LI *et al.* (2020d), que propuseram a extensão do PCA recursivo com metodologias *big data* para monitoramento de uma planta fluoroquímica; RIPPON *et al.* (2021), que desenvolveram um sensor virtual baseado em aprendizado de máquina para prognosticar falhas em um processo metalúrgico; e TIAN *et al.* (2022), que propuseram um sistema que combina técnicas variadas como redes neurais recorrentes e PCA para a predição de condições anormais em processos de craqueamento catalítico.

2.7 Ferramentas computacionais

Existem algumas ferramentas computacionais disponíveis para a aplicação de técnicas de monitoramento de processos baseado em dados. Nas duas próximas subseções, é feita uma apresentação das ferramentas, divididas entre *softwares* comerciais e abertos.

2.7.1 Ferramentas comerciais

A seguir tem-se uma lista não exaustiva de *softwares* comerciais que oferecem soluções para o monitoramento de processos baseado em dados.

- **PLS_Toolbox:** Criado em 1991 por Barry Wise, o propositor da aplicação da técnica PCA para monitoramento de processos (WISE *et al.*, 1988), este é um dos *softwares* pioneiros e talvez o mais disseminado da área da Quimiometria, como evidenciam as aplicações propostas por BARRA *et al.* (2019); NESPECA *et al.* (2019); SUN *et al.* (2019); WANG *et al.* (2015c), por exemplo. O pacote

contém uma vasta gama de técnicas aplicáveis a problemas da Quimiometria, como análise exploratória, classificação, regressão e monitoramento de processos. Funciona como *toolbox* do ambiente de computação Matlab e é comercializado pela empresa Eigenvector Research.

- **AspenTech Suite:** dentre as soluções oferecidas pela empresa AspenTech, várias envolvem aplicações de monitoramento de processos:
 - aspenOne Process Explorer: permite a visualização, análise, monitoramento e acesso a dados de operações de planta em tempo real;
 - Aspen ProMV: fornece a criação de alerta sobre a saúde do processo;
 - Aspen OnLine: oferece a possibilidade de calibrar modelos preditivos para monitorar processos em tempo real em conjunto com simulações de processo;
 - Aspen Process Pulse: possibilita a geração de avisos sobre problemas relacionados à qualidade de produtos com vistas à otimizar e melhoria de qualidade;
 - Aspen Mtell: foca em manutenção preditiva de equipamentos.
- **Aveva Insight:** comercializado pela empresa Aveva, esse *software* possibilita o monitoramento remoto e em tempo real de ativos de produção.
- **Precognize:** a empresa Precognize disponibiliza um *software* homônimo com soluções específicas para manutenção e monitoramento preditivos e diagnóstico de válvulas de controle.
- **Honeywell Forge:** oferecida pela empresa Honeywell, essa plataforma reúne diversas soluções no âmbito da transformação digital que incluem monitoramento, diagnóstico e previsão de desempenho de plantas de processo.
- **IBM Maximo Suite:** comercializada pela empresa IBM, essa plataforma oferece soluções para manutenção preditiva baseada na análise da condição de processos.
- **Braincube Suite:** a empresa Braincube oferece diversos aplicativos para visualização de dados, monitoramento da condição de processos e controle de qualidade.
- **Seeq Suite:** a empresa Seeq disponibiliza uma plataforma para análise de séries temporais que contempla várias etapas do fluxo de monitoramento de processos, como análise exploratória de dados, modelagem preditiva e geração de alarmes.

- **Proficy CSense**: desenvolvido pela empresa GE, possibilita a identificação de problemas, a busca por causas raiz, predição de desempenho futuro e automatização de ações.
- **JMP Software**: a empresa JMP oferece um conjunto de soluções para todo o fluxo de análise de dados industriais, incluindo o monitoramento estatístico.
- **Visplore**: comercializado pela empresa Visplore GmbH, esse software fornece soluções para análise exploratória de dados, incluindo visualização e análises de correlação, causa raiz, regressão e comparação.
- **Statistica**: desenvolvido pela empresa TIBCO, esse software possui um módulo para controle estatístico multivariado de processos que disponibiliza as técnicas PCA e PLS, bem como algumas de suas variantes.

2.7.2 Ferramentas abertas

A criação de ferramentas abertas para o monitoramento de processos em geral se dá por meio do desenvolvimento de extensões do ambiente Matlab. CAMACHO *et al.* (2015) apresentaram a MEDA, dedicada à análise multivariada no contexto da quimiometria utilizando métodos PCA e PLS e com algumas aplicações de monitoramento. JIANG *et al.* (2016b) e JIANG e YIN (2019) propuseram a DB-KIT, focada em indicadores-chave de desempenho (KPI, *key performance indicators*), com algoritmos de regressão e modelagem em variáveis latentes. GONZÁLEZ-MARTÍNEZ *et al.* (2018) propuseram a MVBatch, focada nos processos em batelada e que possui características bem interessantes, como pré-processamento dos dados, validação cruzada e desenvolvimento na plataforma GitHub. YI *et al.* (2019) apresentaram a Pre-Screen, dedicada ao pré-tratamento de dados de processo. VILLALBA *et al.* (2019) desenvolveram um tutorial, em forma de interface gráfica, para entendimento do monitoramento baseado em PCA. SCHAEFFER e BRAATZ (2022) disponibilizaram um aplicativo didático em Matlab para auxiliar no entendimento de modelos de variáveis latentes e de regressão. SUN *et al.* (2017) e ALIZADEH *et al.* (2018) reportaram *toolboxes* do Matlab utilizando métodos de variáveis latentes e análise de causalidade, respectivamente; porém, não foram reportados os nomes das ferramentas nem *links* para *download* e o presente autor não os encontrou buscando na Internet.

Como é possível perceber, as ferramentas disponíveis atualmente para o monitoramento de processos são predominantemente comerciais ou, quando desenvolvidas em código aberto, são extensões do ambiente Matlab. Isso representa uma dificuldade para o desenvolvimento colaborativo, uma vez que o ambiente computacional fechado limita o desenvolvimento e utilização em diversas indústrias que não possuem a

licença do software. Nesse contexto, a presente tese desenvolveu dois softwares - BibMon (MELO *et al.*, 2023) e KydLIB (MELO, 2023) - para preencher essa lacuna na literatura, fornecendo soluções eficientes e de fácil aplicação e extensão para o monitoramento de processos baseado em dados.

2.8 *Benchmarks*

Benchmarks são testes padronizados usados para medir e comparar desempenhos de sistemas computacionais do tipo *hardware* ou *software* (FLEMING e WALLACE, 1986). Apesar da longa tradição de uso de *benchmarks* na literatura de aprendizado de máquina (ABUBAKAR *et al.*, 2015; DENG *et al.*, 2009; DENG, 2012; HU *et al.*, 2021; XIAO *et al.*, 2017), na área específica de análise de dados de processo o número de *benchmarks* é significativamente menor. Uma possível razão é o uso relativamente mais recente de dados para gerar valor por meio de atividades de análise de processos. Outra possível razão é a tradicional mentalidade de confidencialidade no mundo corporativo, que tende a considerar todos os dados disponíveis como estratégicos e, portanto, sujeitos a restrições de confidencialidade. No entanto, o sucesso das abordagens de código aberto em outras áreas, especialmente na área de desenvolvimento de software (KALLIAMVAKOU *et al.*, 2015; STEINMACHER *et al.*, 2017), tem levado as empresas a compartilharem seus dados ao menos parcialmente, pois fica claro que esta abordagem pode trazer muitos benefícios para o campo industrial como um todo e para cada ator em particular. Vários dos *benchmarks* listados a seguir, especialmente os mais recentes, ilustram esta tendência.

Contradicitoriamente, com a crescente disponibilidade de *benchmarks*, escolher o estudo de caso mais adequado para cada aplicação individual pode se tornar bastante desafiador. Por esse motivo, as razões que explicam por que alguns *benchmarks* foram selecionados para um estudo específico às vezes não são muito bem definidas. Por exemplo, pode-se argumentar que o *benchmark* *Tennessee Eastman Process* (DOWNS e VOGEL, 1993) comumente usado é muito popular não apenas por causa de suas qualidades, mas por causa da familiaridade da comunidade de pesquisa com ele. Essa linha de pensamento sugere que pode haver alguma resistência para a adoção de novos *benchmarks*, levando ao desperdício de oportunidades. Na verdade, à medida que aumenta o número de *benchmarks* distintos disponíveis, o número de possíveis avaliações técnicas significativas também aumenta e mais pode ser aprendido sobre as metodologias de monitoramento de processos. Mesmo no âmbito de um único trabalho, a utilização de mais de um *benchmark* pode ser útil para evidenciar com mais clareza os pontos fortes e fracos da metodologia analisada.

Nesta seção, são descritos os *benchmarks* disponíveis na literatura para monitoramento de processos, provenientes de simulações ou conjuntos de dados. Para os conjuntos de dados, análises exploratórias são apresentadas no Capítulo 4.

2.8.1 Tennessee Eastman Process (1993)

O *benchmark* *Tennessee Eastman Process* (TEP) é uma simulação realista de um processo da *Eastman Chemical Company*, proposto como referência para aplicações de monitoramento e controle de processos (DOWNS e VOGEL, 1993). O modelo TEP tem sido amplamente aceito pela comunidade acadêmica para estudos de modelagem e atualmente pode ser considerado o *benchmark* mais utilizado para testes de técnicas de monitoramento de processos. O modelo TEP foi disponibilizado inicialmente como um simulador de processos Fortran, mas agora está implementado em várias plataformas, incluindo MATLAB/Simulink (BATHELT *et al.*, 2015) e Modelica (MARTIN-VILLALBA *et al.*, 2018). Conjuntos de dados de simulação independentes também estão disponíveis para o desenvolvimento de estudos numéricos (CHIANG *et al.*, 2001; REINARTZ *et al.*, 2021a; RIETH *et al.*, 2017), conforme descrito na próxima subseção.

O principal objetivo do processo TEP é gerar dois produtos (G e H) a partir de quatro reagentes (A, C, D e E), embora um inerte (B) e um subproduto (F) também estejam presentes nas correntes de processo. As reações são assumidas como irreversíveis e exotérmicas. Os reagentes gasosos A, C, D e E e o inerte B são alimentados ao reator, onde se formam os produtos G e H, no estado líquido, e o subproduto F. A corrente de produto passa por um condensador e um separador líquido-vapor, respectivamente. O vapor resultante é reciclado, enquanto o condensado passa por um esgotador, da base do qual saem os produtos. O vapor é purgado do separador líquido-vapor para remover o inerte e o subproduto.

A simulação original assume a existência de 41 variáveis medidas e 12 variáveis manipuladas, com intervalos de amostragem de 3, 6 e 15 minutos. Os estudos de referência originais compreendem 6 modos de operação e 21 perturbações pré-programadas, que incluem a introdução de degraus, modificação da variabilidade de medição, ocorrência de travamento da válvula, desvio da cinética de reação e perturbações desconhecidas. Simuladores mais recentes estenderam esses recursos para incluir medições e distúrbios de processo adicionais (GINS *et al.*, 2014; REINARTZ *et al.*, 2021a), conforme detalhado na próxima subseção.

2.8.1.1 Desenvolvimento do *benchmark* TEP

Nos anos que se seguiram à introdução do *benchmark* TEP, a comunidade de pesquisa explorou as muitas características do modelo em trabalhos relacionados ao controle de processos (BANERJEE e ARKUN, 1995; LYMAN e GEORGAKIS, 1995; MCAVOY *et al.*, 1995; RICKER, 1996), otimização de processos (RICKER, 1995; YAN e RICKER, 1997) e estimativa de estado (RICKER e LEE, 1995), por exemplo. Após o desenvolvimento e disseminação das estratégias de controle, foi possível utilizar o TEP como *benchmark* para monitoramento de processos, conforme proposto por HOWELL *et al.* (1997) utilizando a estratégia apresentada por MCAVOY *et al.* (1995). Um simulador de TEP frequentemente usado foi disponibilizado por Braatz e colaboradores, usando a estratégia de controle apresentada por LYMAN e GEORGAKIS (1995). Esse simulador está disponível no endereço <https://web.mit.edu/braatzgroup/links.html>.

Simuladores

Adaptações do código em Fortran original para as linguagens C e MATLAB vem sendo implementadas e disponibilizadas, como pode se encontrar no repositório denominado *Tennessee Eastman Challenge Archive*, mantido por Ricker e colaboradores e disponível em <https://depts.washington.edu/control/LARRY/TE/download.html>. Notavelmente, uma revisão de código (BATHELT *et al.*, 2015) foi proposta para possibilitar o uso de métodos de integração de passo variável em ambientes de simulação como o MATLAB. Antes dessa revisão, esses métodos podiam levar a resultados inconsistentes quando o fluxo de código originalmente proposto era utilizado. O novo simulador resultante incluiu mais distúrbios e medições de processo, documentação sobre variáveis internas, opções para definir magnitudes de falhas e a possibilidade de selecionar estruturas de controle distintas, como as propostas por RICKER (1996) e LARSSON *et al.* (2001). Em particular, CAPACI *et al.* (2019) propuseram recentemente um conjunto de diretrizes para o uso adequado do simulador de Ricker como referência para aplicações de controle estatístico de processos e projeto de experimentos. A principal desvantagem do simulador de Ricker é a necessidade de se obter uma licença MATLAB, que não é gratuita. Esse problema foi um pouco minimizado com a publicações de conjuntos de dados baseados no simulador, como descrito a seguir.

Dados

Conforme descrito, o simulador disponibilizado por Braatz e colaboradores é uma das ferramentas baseadas no *benchmark* TEP mais populares até hoje. Em complemento ao simulador, um conjunto de dados também foi disponibilizado para uso como *benchmark* de monitoramento de processos. O conjunto de dados foi gerado para o modo de operação 1 proposto no artigo original, compreendendo 25 h de operação para arquivos de dados de treinamento e 48 h de operação para arquivos de dados de teste. Nos arquivos de teste, as falhas foram introduzidas após 8 h de operação. CHIANG *et al.* (2001) discutiram o uso do simulador, dos conjuntos de dados e algumas aplicações orientadas a dados. A incorporação de uma estratégia de controle ao simulador e a disponibilização de um conjunto de dados tornaram o modelo TEP ainda mais fácil de usar do que antes, ajudando a explicar o crescimento significativo do número de publicações baseadas no modelo.

Apesar do uso generalizado, o conjunto de dados do Braatz apresenta algumas desvantagens. Talvez o principal problema esteja relacionado à escassez de informações, que está ficando mais clara agora na era de *big data*. Particularmente, foi argumentado (RIETH *et al.*, 2018) que os resultados obtidos com o conjunto de dados do Braatz podem ser tendenciosos por conterem apenas uma simulação por evento, o que pode não ser representativo da multiplicidade de possíveis simulações. Com base nisso, RIETH *et al.* (2017) forneceram um novo conjunto de dados contendo 500 simulações por evento, usando a mesma configuração de simulação originalmente descrita no conjunto de dados do Braatz. A disponibilidade de conjuntos de dados maiores, como o fornecido por RIETH *et al.* (2017), é mandatório no contexto de técnicas modernas que exigem grandes quantidades de dados, incluindo o desenvolvimento de redes neurais profundas, como ilustrado por HEO e LEE (2019) com o uso de redes neurais autoassociativas paralelas. O conjunto de dados disponibilizado por RIETH *et al.* (2017) é analisado em detalhes no Capítulo 4.

Novos conjuntos de dados estão surgindo, baseados também na simulação proposta por BATHELT *et al.* (2015). Por exemplo, MANCA (2020) forneceu um novo conjunto de dados para uso em estudos de gerenciamento de alarmes relacionados ao controle estatístico de processo univariado. Além disso, REINARTZ *et al.* (2021a,b) também forneceram um conjunto de dados estendido padronizado e abrangente, baseado no simulador de Ricker, para estudos numéricos genéricos. Esse novo conjunto de dados amplia os anteriores em escopo e escala, reduzindo os problemas relacionados à escassez de informações e melhorando o potencial de avanços científicos. Novas possibilidades também foram exploradas, relacionadas principalmente a transições de modo de operação e magnitudes de falhas. No entanto,

isso também torna o novo conjunto de dados mais difícil de usar; por exemplo, ele é significativamente maior que os demais conjuntos de dados, com tamanho total de 132,96 GB (aproximadamente 22 GB por modo de operação), o que pode inviabilizar sua manipulação para alguns usuários.

2.8.2 PenSim (2002)

O *benchmark* PenSim é um ambiente de simulação relacionado a um processo de produção de penicilina. O *benchmark* consiste de um modelo não estruturado de fermentação da penicilina, originalmente proposto por BAJPAI e REUB (1980) e posteriormente estendido por BIROL *et al.* (2002). É importante enfatizar que este se tornou o *benchmark* mais popular para testagem de técnicas projetadas e destinadas à aplicação em processos em batelada.

O processo opera em duas etapas: uma etapa de pré-cultura e uma etapa de alimentação descontínua. Durante a fase de pré-cultura, a biomassa é gerada; durante a fase de alimentação descontínua, a penicilina é produzida. O modelo representa essencialmente efeitos dinâmicos relacionados ao crescimento da biomassa, variação e controle de pH, variação e controle de temperatura, produção de penicilina, consumo de substrato, mudanças de volume, liberação de calor de reação e evolução da produção de CO₂. Embora os estudos originais descrevam valores-padrão razoáveis para as diversas variáveis de processo, o simulador disponível oferece muitas possibilidades para o usuário modificar as condições iniciais, os parâmetros cinéticos, o tipo de controlador e respectivos parâmetros de controle, as políticas de operação e os efeitos das variáveis ambientais. Flutuações e ruídos também são adicionados a algumas variáveis (mas não todas), para simular a variabilidade da medição. Particularmente, três exemplos de falhas foram apresentados por BIROL *et al.* (2002), incluindo uma falha do controlador de pH e distúrbios do tipo degrau na potência de agitação e taxa de fluxo de alimentação do substrato.

O *benchmark* PenSim também está disponível no pacote de simulação RAYMOND, conforme descrito na Subseção 2.8.4. Uma vantagem significativa de se usar o RAYMOND para simular o modelo PenSim é que o ruído pode ser adicionado a todas as medições de variáveis, tornando os dados mais realistas. Particularmente, VAN IMPE e GINS (2015) usaram o RAYMOND para gerar um conjunto de dados de referência contendo 1.600 lotes normais e 90.400 lotes defeituosos, que foram divididos em quatro subconjuntos de complexidade diferente. O primeiro subconjunto compreendeu o caso base, com condições iniciais amostradas de distribuições gaussianas; o segundo subconjunto introduziu variações não gaussianas nas condições iniciais; o terceiro subconjunto introduziu variabilidade de lote para lote durante as ações de

controle de pH; e o quarto subconjunto introduziu distúrbios multimodais assumindo a existência de duas populações de microrganismos. 23 medições estão disponíveis e 15 tipos diferentes de falhas de processo foram simuladas com diferentes magnitudes, incluindo mudanças abruptas e desvios lentos. Este conjunto de dados constitui uma importante contribuição para a área, pois aumentou a riqueza de informações, a facilidade de uso, os desafios considerados e o potencial de avanço científico baseado no modelo PenSim. Porém, infelizmente, o *link* para *download* apresentado no artigo de VAN IMPE e GINS (2015) está quebrado e não foi encontrado nenhum outro.

2.8.3 DAMADICS (2006)

DAMADICS (*Development and Application of Methods for Actuator Diagnosis in Industrial Control Systems*) é um *benchmark* composto por um simulador de processo e conjuntos de dados reais que representam atuadores eletropneumáticos de um processo de produção de açúcar (BARTYS *et al.*, 2006). O simulador original foi implementado em MATLAB/Simulink e os dados reais são provenientes da Lublin Sugar Factory, Polônia. Três atuadores distintos colocados ao longo do processo foram selecionados para representar o sistema: dois atuadores são usados para controlar a entrada e saída de caldo fino do evaporador e um atuador controla a entrada de água em um gerador de vapor da caldeira. Os modelos, validados com dados históricos e baseados em descrições termodinâmicas e mecatrônicas, foram divididos em quatro blocos funcionais principais que se referem respectivamente ao posicionador, servomotor, válvula e *bypass*. Há 19 falhas no *benchmark*, incluindo falhas de posicionador, de servomotor, de válvula de controle e falhas gerais/externas. Os dados históricos, que são descritos em detalhes no Capítulo 4, consistem de 32 medições amostradas a uma taxa de 1 Hz. O *benchmark* DAMADICS foi proposto explicitamente como referência para a avaliação de técnicas de monitoramento de processos. Por essa razão, uma sequência de etapas para detecção e diagnóstico de falhas foi proposta no artigo original, compreendendo a avaliação preliminar e a posterior análise da detectabilidade e isolabilidade das falhas nos ambientes de dados simulados e reais. Comparado aos *benchmarks* TEP e PenSim, a riqueza de informações é muito maior por um lado, pois o *benchmark* fornece dados reais e uma metodologia de testes com o simulador; no entanto, por outro lado, o escopo do *benchmark* é um pouco limitado, pois está relacionado especificamente a um tipo de equipamento dentro do processo.

2.8.4 RAYMOND (2014)

RAYMOND (GINS *et al.*, 2014) é um pacote MATLAB para geração de dados representativos de monitoramento por meio de simulações. Vários *benchmarks*, notadamente os modelos TEP e PenSim, estão prontamente disponíveis para uso no pacote. Além disso, os usuários também podem adicionar seus próprios modelos e especificar características de sensores, variabilidade do processo, flutuações de entrada, estratégias de controle e vários tipos de falhas. A principal contribuição do RAYMOND é a melhoria significativa da riqueza da informação disponível e o aumento do potencial de avanços científicos baseados em *benchmarks* existentes.

2.8.5 IndPenSim (2015)

IndPenSim é um modelo fenomenológico usado para descrever o processo de fermentação do *Penicillium chrysogenum* (GOLDRICK *et al.*, 2015, 2019). As principais diferenças em relação ao *benchmark* PenSim tradicional são a escala industrial (volumes de 100.000 L, em oposição aos vasos típicos de 100 L no *benchmark* PenSim), a disponibilidade de dados históricos reais e o uso de um modelo estruturado (PAUL e THOMAS, 1996), que consideram crescimento, morfologia, produção metabólica e degeneração da biomassa. Há no processo sensores para oxigênio dissolvido, pH, temperatura, formação de espuma e pressão. Medições *offline* incluem concentrações de penicilina, biomassa, amônia e ácido fenilacético.

A simulação foi implementada em MATLAB R2013B e posteriormente atualizada para MATLAB R2018B. Um conjunto de dados com 100 lotes também foi gerado e disponibilizado em um trabalho posterior (GOLDRICK *et al.*, 2019), que também estendeu a simulação com medições de espectroscopia Raman. Além disso, o modelo oferece muitas opções para o usuário, como estratégias de controle distintas, distúrbios nas concentrações e efeitos de inibição. Possíveis falhas de processo envolvem agitação, aeração e falhas de sensor. O trabalho original utilizou dados históricos para validar as simulações, apresentando as condições de operação (e respectivas precisões) em que os resultados da simulação poderiam ser considerados confiáveis.

O conjunto de dados disponibilizado por GOLDRICK *et al.* (2019) é analisado em detalhes no Capítulo 4.

2.8.6 Cranfield Multiphase Flow Facility (2015)

Cranfield Multiphase Flow Facility (daqui em diante denominado *benchmark Cranfield*) foi proposto como uma referência de dados reais para o desenvolvimento e validação de procedimentos estatísticos de monitoramento de processos (RUIZ-CÁRCEL *et al.*, 2015). Trata-se de um processo de separação de fluxo trifásico em escala piloto, composto por dutos, um separador bifásico gás-líquido e um separador trifásico água-óleo-ar. O *benchmark* foi projetado para fornecer uma vazão medida e controlada de água, óleo e ar para um sistema pressurizado. O conjunto de dados contém 24 variáveis de processo, 2 delas sendo entradas de processo (*set points* para vazões de ar e água) a uma taxa de amostragem de 1 Hz. Distintas condições de operação foram disponibilizadas por meio da manipulação das entradas, garantindo a representação de uma ampla gama de características dinâmicas nos dados coletados. 6 falhas padrão podem ser usadas para fins de validação, incluindo condições de operação anômalas, operação incorreta e bloqueios de linha. O artigo original também apresentou uma aplicação do método CVA ao *benchmark*, o que ajudou a estabelecer uma referência de resultado e apontar desafios específicos e pontos de melhoria. O *benchmark Cranfield* é um conjunto de dados fácil de usar contendo dados reais e riqueza de informações, o que pode explicar sua adoção relativamente rápida pela comunidade de pesquisa. O conjunto de dados é analisado em detalhes no Capítulo 4.

2.8.7 PRONTO (2019)

O conjunto de dados heterogêneo PRONTO (*PROcess NeTwork Optimization*) (STIEF *et al.*, 2019a,b) foi proposto para apresentar desafios relacionados às diferentes formas em que os dados podem ser adquiridos e armazenados. Especificamente, o *benchmark* contém medições de processo (incluindo algumas obtidas com sensores ultrassônicos de vazão e pressão de alta frequência), registros de alarmes, eventos e registros de alterações, um registro de operação e gravações de vídeo. Ele é baseado na mesma instalação descrita na Subseção 2.8.6, localizada na Cranfield University, e pode ser vista como uma extensão do *benchmark* anterior. Uma variedade de condições operacionais também está presente, garantindo a representação de uma ampla gama de características dinâmicas nos dados. Os dados de processo contêm 27 variáveis de processo e 2 entradas de processo (*set points* para vazões de ar e água) a uma taxa de amostragem de 1 Hz. Os registros e eventos de alarme e *logs* de alterações são amostrados por meio de uma política de amostragem orientada a eventos. A vazão ultrassônica e a pressão de alta frequência foram amostradas a 10 e 5 kHz, respectivamente. 5 tipos de falhas podem ser usados para validações

de monitoramento de processo: condições de *slugging*, falhas induzidas por válvulas manuais, vazamento de ar, bloqueio de ar e desvio de fluxo. Algumas características dos dados foram descritas no artigo original, notadamente os fatos de não haver dados ausentes ou medições corrompidas e de a principal fonte de incerteza ser o ruído de medição e as perturbações de processo. Foram apresentados estudos de caso efetuados com técnicas do tipo CVA e bayesianas. Os autores sugeriram abordagens interessantes para explorar todo o potencial possibilitado pela natureza heterogênea do *benchmark*, como o uso dos vídeos e do *log* de operação para rotular os dados de modo a alimentar algoritmos de aprendizado de máquina supervisionados. Comparado com o *benchmark* Cranfield, o *benchmark* PRONTO aumentou a riqueza de informações, os desafios propostos e o potencial de avanço científico, mas é consideravelmente mais difícil de usar.

Como o *benchmark* PRONTO é muito semelhante ao *benchmark* Cranfield, por se basear na mesma planta, os resultados da análise exploratória de dados não serão apresentados no Capítulo 4. Leitores interessados podem consultar MELO *et al.* (2022).

2.8.8 Conjunto de dados 3W (2019)

O conjunto de dados 3W (VARGAS *et al.*, 2019) é um repositório contendo dados reais, simulados e desenhados à mão que representam períodos de operação normais e anômalos de poços de petróleo *offshore*. 8 tipos de falhas estão documentados no repositório: aumento abrupto da proporção de sedimentos básicos e água (BSW), fechamento espúrio da válvula de segurança de fundo de poço (DHSV), condições severas de *slugging*, instabilidade de fluxo, perda rápida de produtividade, restrição rápida na válvula de controle de estrangulamento da produção (PCK), *scaling* na válvula PCK e hidrato na linha de produção. Oito medições estão disponíveis ao longo da operação, amostradas a uma taxa de 1 Hz. Os dados são divididos em 1984 blocos de tempo denominados pelos autores como instâncias, tipicamente com durações de algumas horas. As instâncias reais somam 1025, abrangendo um período entre os anos de 2012 e 2018 e um conjunto de 21 poços. Simulações e instâncias desenhadas à mão não possuem perturbações de processo e ruído de medição, sendo facilmente distinguíveis de instâncias reais. Os autores propõem dois *benchmarks* específicos para orientar a aplicação: i) investigação do impacto do uso de instâncias simuladas e desenhadas à mão e ii) detecção de anomalias. O conjunto de dados 3W traz desafios relacionados a dados do mundo real, como um número significativo de observações ausentes e variáveis congeladas ou ausentes. Isso é uma vantagem para quem deseja estudar como lidar com questões de qualidade de dados, mas pode

desencorajar pesquisadores que desejam avaliar o desempenho de uma determinada técnica apenas diante de incertezas inerentes aos processos. O conjunto de dados é analisado em detalhes no Capítulo 4.

2.8.9 Fracionador FCC (2022)

O Fracionador FCC é um modelo recentemente proposto para simular uma operação de refino composta por um craqueador catalítico de leito fluidizado e um fracionador (SANTANDER *et al.*, 2022). O processo é composto de pré-aquecedor, reator, fracionador, regenerador, compressor de gás úmido, sopradores de ar de elevação, soprador de ar de combustão, linhas de circulação de catalisador e controlador de processo. As inovações do modelo em relação aos modelos existentes na literatura incluem a distribuição realista do produto do craqueamento do reator e a integração entre o reator e o fracionamento. Um simulador MATLAB foi disponibilizado. Por se tratar de um *benchmark* muito recente, ainda há grande possibilidade de avanços em relação ao seu uso e divulgação. Em particular, a comunidade ainda precisa discutir como serão simuladas e avaliadas as situações inerentes ao monitoramento de processos, incluindo a natureza das falhas, pois esses tópicos específicos não foram abordados no artigo original.

2.8.10 Outros conjuntos de dados

A presente revisão focou no monitoramento de processos com base em séries temporais, mas vale a pena mencionar alguns *benchmarks* industriais interessantes para aplicações de diagnóstico de falhas em outros contextos. Na indústria manufatureira, alguns conjuntos de dados são gerados com cada observação correspondente a uma peça ou produto fabricado. Exemplos de *benchmarks* com essa configuração são o *benchmark* SECOM (MCCANN *et al.*, 2010), o desafio *Bosch Production Line Performance* do site Kaggle e o conjunto de dados *Three-Dimensional Scan Strategies* do *Additive Manufacturing Metrology Testbed* (LANE e YEUNG, 2019). Alguns conjuntos de dados, especialmente os da indústria siderúrgica, são constituídos por fotografias de peças; um exemplo é o desafio *Severstal: Steel Defect Detection* do site Kaggle. Outros conjuntos de dados que podem ser úteis para detecção de falhas, mas não relacionados ao monitoramento de processos, podem ser encontrados em indústrias como a engenharia aeroespacial (GAO e LIU, 2021) e geração de energia (LIU *et al.*, 2020b); bons exemplos são os conjuntos de dados PCoE da NASA e o desafio *VSB Power Line Fault Detection* do site Kaggle. O diagnóstico de falhas em máquinas rotativas também é um campo ativo de pesquisa, como ilustrado por um

estudo comparativo de técnicas de aprendizado profundo aplicadas a nove *benchmarks* abertos (ZHAO *et al.*, 2020). O projeto *Open Industrial Data* (LABORIE *et al.*, 2019), desenvolvido pelas empresas Aker BP e Cognite, fornece a primeira transmissão ao vivo de dados industriais reais e tem grande potencial para ser utilizado em estudos relacionados ao monitoramento de processos. Outros conjuntos de dados industriais relacionados a uma variedade de tópicos e setores da indústria podem ser encontrados no repositório fornecido pela companhia MakinaRocks no *site* GitHub.

2.8.11 Modelos

Além dos conjuntos de dados e simuladores já disponíveis, existem vários modelos conhecidos que os usuários podem implementar para uso como *benchmarks*. Essa prática é especialmente comum na comunidade de controle de processos. O *benchmark* BSM1 – *Benchmark Simulation Model No. 1* (ALEX *et al.*, 2008), – uma especificação detalhada de um ambiente de simulação para uma estação de tratamento de águas residuais projetada para avaliar estratégias de controle, também vem sendo usado para avaliar estratégias de monitoramento (CHEN *et al.*, 2016; HONG *et al.*, 2020). O reator tanque agitado contínuo (*continuous stirred-tank reactors*, CSTR) constitui outra referência comum, embora em contraste com o modelo BSM1, as especificações geralmente variem de trabalho para trabalho. Vale ressaltar que modelos CSTR específicos são populares na comunidade de controle de processos, mas não são comumente usados em aplicações de monitoramento de processos, como o Reator van der Vusse (AGUILAR-LOPEZ *et al.*, 2021; VAN DE VUSSE, 1964) e o Reator Williams-Otto (MUKKULA e ENGELL, 2021; WILLIAMS e OTTO, 1960). Em particular, o trabalho de WILLIAMS e OTTO (1960) oferece muitas possibilidades para estudos de monitoramento de processos, pois foi proposto como um modelo de planta de processo completo que também poderia ser usado como referência para monitoramento de processos. Outros *benchmarks* de modelos comuns incluem controle de nível de tanque, colunas de destilação e sistemas de aquecimento (ADEDIGBA *et al.*, 2017; AMIN *et al.*, 2020; GHOSH *et al.*, 2019).

2.9 Exploração, caracterização e tratamento de dados de processo

Na estatística clássica, uma das premissas básicas para o desenvolvimento da teoria é de que os dados a serem analisados são provenientes de experimentos planejados para testar hipóteses estatísticas específicas (SCHWAAB e PINTO, 2007,

2011). Dados classificados como históricos, por sua vez, resultam da observação não planejada da realidade, sendo a análise desse tipo de dado fortemente desencorajada na literatura estatística clássica (PEARSON, 2001). Portanto, a simples aplicação de técnicas clássicas à análise de dados históricos pode não levar a resultados satisfatórios.

A análise exploratória de dados (EDA, do inglês *exploratory data analysis*) fornece meios para se descobrir quaisquer fatos interessantes sobre os dados sem recorrer a uma linha específica de modelagem. De acordo com MORGENTHALER (2009), na EDA o analista “é livre para escolher qualquer procedimento para analisar os dados, e os principais objetivos são olhar para os dados e pensar sobre eles de vários pontos de vista”. O autor também cita exemplos de procedimentos estatísticos clássicos como gráficos de caixa e *median polishes*. Em tempos mais recentes, os procedimentos emergentes de ciência de dados oferecem soluções que vão além da estatística clássica, adequando-se melhor ao cenário de *big data*. Muitos desses procedimentos estão sendo disponibilizados em pacotes de código aberto, como DataPrep.EDA (PENG *et al.*, 2021), que oferece recursos de EDA centrados em tarefas, como análise bivariada e análise de valores faltantes. No entanto, apesar da abundância de opções de ferramentas no ecossistema de ciência de dados, nenhuma delas correspondia às necessidades específicas para se explorar dados históricos de processos.

A aplicação de EDA a dados de processo vem sendo discutida na literatura há vários anos. PEARSON (2001) pesquisou procedimentos simples como gráficos de quantil-quantil, gráficos de comparação de dados, gráficos de caixa e caracterizações de momentos. O autor argumentou que tal exploração pode ser essencial porque certas características inesperadas dos dados, como *outliers* provenientes de diversas fontes, podem invalidar completamente os resultados obtidos por técnicas de monitoramento, em especial as baseadas em critérios de erro quadrático. Suas principais recomendações foram: (i) a aplicação de testes de adequação dos dados a distribuições específicas (em especial a normal) e de comparação entre distribuições de diferentes conjuntos; (ii) o uso de métodos robustos para caracterização estatística, como mediana no lugar da média, MAD (*median absolute deviation from the median*) no lugar do desvio-padrão e coeficiente de correlação de Spearman no lugar do de Pearson²⁴; (iii) uso de métodos de visualização como *box-plots* para comparação entre variáveis. ABONYI (2007) gerou gráficos de caixa e de quantil-quantil para analisar dados históricos reais de uma planta de produção de polietileno. ZHANG *et al.* (2016) propuseram um teste de características dos dados para selecionar métodos de monitoramento de processos automaticamente; as características avaliadas foram

²⁴O coeficiente de correlação de Pearson é uma medida da associação linear entre variáveis, enquanto o coeficiente de correlação de Spearman é uma medida da associação monótona entre elas.

desvios de gaussianidade e linearidade, e os métodos selecionados foram variações e combinações das técnicas PCA e ICA.

Na área de ciência de dados, uma tendência é a integração de métodos de aprendizado de máquina com técnicas de visualização, dando origem a uma área conhecida como *visual analytics* (SACHA *et al.*, 2017; YOUSEF *et al.*, 2022). No que se refere à indústria de processos, alguns métodos de visualização de dados históricos foram propostos. WANG *et al.* (2004) e ALBAZZAZ *et al.* (2005) propuseram a visualização multidimensional das variáveis de processo por meio de gráficos de coordenadas paralelas (INSELBERG, 2009), mostrando que a metodologia é capaz de separar visualmente dados normais e não normais. THORNHILL *et al.* (2006) utilizaram uma árvore de classificação hierárquica para propor uma abordagem visual baseada em grupos para detecção automática de *outliers*. GAJJAR e PALAZOGLU (2016) aplicaram a visualização por coordenadas paralelas para propor uma metodologia de detecção e diagnóstico de falhas, integrando as técnicas PCA e *Random Forest*. WANG *et al.* (2017a,b, 2018a, 2015b) propuseram o uso de diagramas de Kiviat (mais conhecidos como gráficos de radar) para visualização e detecção de falhas baseada em visualização.

XU *et al.* (2015) publicaram ampla revisão a respeito do tratamento de dados industriais (o termo empregado pelos autores foi *data cleaning*). O problema foi dividido em quatro passos: preenchimento de valores faltantes, detecção de *outliers*, remoção de ruído e alinhamento temporal/estimação de atraso. Uma vasta gama de métodos disponíveis na literatura para cada um desses problemas foi apresentada. Os autores argumentaram que, na prática, a escolha dos métodos de tratamento de dados deve ser guiada pelas propriedades do próprio conjunto (como proporção de dados faltantes) e do conhecimento que já se tem sobre o processo (por exemplo, se for conhecido um modelo para o processo, o método deve ser capaz de incorporá-lo). Os métodos também devem ser capazes de preservar e revelar as características dos dados regulares e identificar os mecanismos que levam à contaminação (como a natureza dos *outliers*).

Na mesma linha de metodologias robustas, com foco em *outliers* e dados faltantes, ZHU *et al.* (2018) publicaram uma interessante revisão, sob a perspectiva *big data*, de técnicas comumente aplicadas para pré-processamento dos dados, e de variantes robustas do PCA e outras técnicas de monitoramento.

Uma importante etapa anterior à aplicação de uma técnica de monitoramento é a seleção das variáveis de entrada. A área do aprendizado de máquina que lida com essa questão é conhecida por *feature engineering*. Uma abordagem proposta por WANG e HE (2010) e HE e WANG (2011), chamada de *análise de padrões estatísticos*

(*statistical pattern analysis*, SPA), vem chamando atenção nos últimos anos. No SPA, selecionam-se como variáveis de entrada das técnicas (tipicamente, PCA) medidas estatísticas das variáveis de processo (média, variância, autocorrelação, correlação cruzada, dentre outras), o que melhora o desempenho de detecção e diagnóstico em muitos casos, segundo os autores. Trabalhos recentes sobre seleção de variáveis incluem os de LUCKE *et al.* (2019), que aplicaram critérios baseados em informação mútua para selecionar as variáveis; RAUBER *et al.* (2020), que propuseram uma técnica eficiente baseada na análise sequencial de contribuição para as falhas; KWAK *et al.* (2020), que propuseram uma técnica para extrair variáveis não estacionárias e a aplicaram ao problema de detecção de incrustação; WU *et al.* (2020a), que integraram métodos de seleção de variáveis com técnicas de janela móvel para monitoramento multimodal, atingindo bons resultados em uma falha particularmente desafiadora do *benchmark Tennessee Eastman*; SHAH *et al.* (2020), que compararam métodos de aprendizado raso e profundo para seleção de variáveis em uma plataforma de testes de natureza industrial focada em conceitos de *big data* e internet das coisas; e CLAVIJO *et al.* (2021), dentre os quais o presente autor se inclui, que propuseram a aplicação de métodos baseados em causalidade para seleção de features no *benchmark TEP* e em uma planta de medição fiscal de plataforma de petróleo. PERES e FOGLIATTO (2018) publicaram revisão sobre o assunto, destacando que os métodos mais usados são a regressão LASSO, o algoritmo genético e a regressão por seleção direta (*forward selection*).

SIANG *et al.* (2023) descreveram boas práticas para aquisição e preparação de dados de processo. Os autores enfatizaram a importância de (i) compreender a natureza dos dados por meio de análise exploratória; (ii) selecionar apropriadamente o subconjunto dos dados a ser utilizado na modelagem; (iii) atentar às condições de malha aberta ou fechada; (iv) alinhar sinais amostrados a diferentes taxas; e (v) adequar a escolha das métricas às necessidades práticas dos problemas de negócio.

Sobre o tema amplo de exploração, caracterização e tratamento de dados de processos, vale ainda destacar os trabalhos recentes de LI *et al.* (2018b), que propuseram a divisão do conjunto de dados com base na natureza de correlação (linear ou não linear) detectada entre as variáveis e a aplicação de uma estratégia hierárquica nesses conjuntos para monitoramento de características lineares e não lineares em diferentes níveis; THOMAS *et al.* (2018), que apresentaram uma abordagem integrada de agrupamento de dados e extração de *features*, como auxílio ao aumento do conhecimento sobre os dados de processo; PARENTE *et al.* (2019), que propuseram a aplicação da técnica de Monte Carlo para aumentar a quantidade de dados usados no treinamento de um modelo de monitoramento de um processo de produção de papel; e OFFERMANS *et al.* (2020), que compararam diversos métodos para sincronização

dinâmica das variáveis de processo.

No Capítulo 4, uma metodologia para análise exploratória de dados de processo é proposta e aplicada a vários dos *benchmarks* descritos na Seção 2.8.

2.10 Oportunidades em outros domínios

Análise de séries temporais

Dados provenientes de sistemas de monitoramento são resultantes de medidas efetuadas ao longo do tempo, sendo portanto classificados como *séries temporais*. Há toda uma área da estatística dedicada ao estudo de séries temporais (BOX *et al.*, 2015; WEI, 2005), com aplicações em diversas áreas, notavelmente na área de econometria (ENDERS, 2014). Surpreendentemente, há pouca interação entre as literaturas de análise de séries temporais e de monitoramento de processos. Artigos sobre monitoramento em geral são publicados em revistas das áreas de processos e quimiometria, em particular *Journal of Process Control*, *Control Engineering Practice*, *Journal of Chemometrics*, *Chemometrics and Intelligent Laboratory Systems*, *Computers and Chemical Engineering*, *Industrial & Engineering Chemistry Research* e *AICHE Journal*. Ao que parece, essas revistas não costumam ser acessadas por um público de interesse mais genérico em séries temporais, e nem parece haver muitas citações a esse público por parte dos autores dessas revistas. Uma sugestão para trabalhos futuros é a análise e confirmação dessa percepção utilizando técnicas de bibliometria.

QIN e CHIANG (2019) recentemente reconheceram a pouca atenção dada, por parte da comunidade de engenharia de sistemas em processos, aos demais domínios que estudam séries temporais:

Another notable contribution from process system engineering is the time series trend analysis by BAKSHI e STEPHANOPOULOS (1994a). Time series pattern matching for disturbance mining was studied by SINGHAL e SEBORG (2002). It is unfortunate that these efforts have not brought enough attention in process systems engineering. In the recent years time series data mining, segmentation, and pattern matching have been very active in creating simpler representations of time series that perform indexing, clustering, and classification (FU, 2011; KEOGH e KASETTY, 2003).

FU (2011), mencionado na transcrição acima, realizou ampla revisão sobre o trabalho de mineração de dados em séries temporais, categorizando as linhas de estudo em representação e indexação, medida de similaridade, segmentação, visualização e mineração. Nesse importante trabalho, publicado na revista “*Engineering Applications of Artificial Intelligence*”, são citados apenas alguns poucos artigos da área de engenharia de processos.

A análise de tendências qualitativas em séries temporais, citada na transcrição acima, foi proposta em uma sequência de quatro artigos (BAKSHI e STEPHANOPOULOS, 1994a,b; CHEUNG e STEPHANOPOULOS, 1990a,b) e posteriormente desenvolvida, por exemplo, por CHARBONNIER e PORTET (2012); MAESTRI *et al.* (2014); MAURYA *et al.* (2007, 2010, 2005); THURLIMANN e VILLEZ (2017); THURLIMANN *et al.* (2018); VILLEZ *et al.* (2008); VILLEZ (2015); VILLEZ *et al.* (2009). Nessa classe de técnicas, as séries temporais são analisadas de modo a obter, de maneira automática, características qualitativas que sejam indicativas do estado em que se encontra o processo. Muitas vezes essas características podem ser representadas de forma gráfica, o que pode ser relevante para aplicações em chão de fábrica. Enquanto QIN e CHIANG (2019) relataram uma falta de atenção à técnica por parte da comunidade de engenharia de sistemas em processos, pode-se observar também que o mesmo vale para a comunidade de análise de séries temporais.

No restante da seção, são descritas quatro vertentes da análise de séries temporais que têm especial potencial de sinergia com a área de monitoramento de processos: detecção de novidades/anomalias/pontos de mudança, *streaming data* e análise de causalidade e análise não linear.

Detecção de anomalias/novidades/pontos de mudança

No domínio dos processos industriais, a área de monitoramento de processos é praticamente sinônimo de detecção e diagnóstico de falhas (CHIANG *et al.*, 2001). Em outros domínios, porém, como processamento de sinais, análise de dados médicos, dentre outras, o mesmo problema é conhecido por outros nomes, como:

- detecção de anomalias (*anomaly detection*) (AHMAD e PURDY, 2016; AHMAD *et al.*, 2017; AKCAY *et al.*, 2022; ANDRADE *et al.*, 2019; ARIYALURAN HABEEB *et al.*, 2022; BARBARIOL *et al.*, 2022; BLÁZQUEZ-GARCÍA *et al.*, 2022; CARLETTI *et al.*, 2023; CHALAPATHY e CHAWLA, 2019; FORTHUIS, 2021; GARG *et al.*, 2022; GOLDSTEIN e UCHIDA, 2016; HAN *et al.*, 2022; HILAL *et al.*, 2022; LAKSHMINARAYAN *et al.*, 2015; LAVIN e AHMAD, 2015; RUFF *et al.*, 2021; SCHMIDL *et al.*, 2022; SGUEGLIA *et al.*,

2022; SPIES *et al.*, 2023; TALAGALA *et al.*, 2019; UL HASSAN *et al.*, 2023; XIA *et al.*, 2022);

- detecção de novidades (*novelty detection*) (BOUGUELIA *et al.*, 2018; CEJNEK e BUKOVSKY, 2018; CLIFTON *et al.*, 2011; DECKE *et al.*, 2022; DING *et al.*, 2014; DOMINGUES *et al.*, 2020; FARIA *et al.*, 2015; GHOSAL *et al.*, 2022; GRUHL, 2022; LA GRASSA *et al.*, 2022; MARKOU e SINGH, 2003a,b; PIMENTEL *et al.*, 2014; SKVARA *et al.*, 2018; TARASSENKO *et al.*, 2009);
- detecção de ponto de mudança (*change-point detection*) (ADAMS e MACKAY, 2007; ALIPPI *et al.*, 2015; AMINIKHANGHAHI e COOK, 2017; CABRIETO *et al.*, 2017, 2018a,b; CHEN *et al.*, 2022; GUPTA *et al.*, 2022; KUNCHEVA, 2013; NIU *et al.*, 2016; PREUSS *et al.*, 2015; STAUDACHER *et al.*, 2005; TRAN, 2019; TRUONG *et al.*, 2020; XIE *et al.*, 2021, 2013; XUAN e MURPHY, 2007).

Apesar de tratarem do mesmo problema (ao menos dos pontos de vista computacional e matemático), a comunicação entre esses domínios e o de monitoramento de processos é praticamente nula.

Talvez a principal razão para esse isolamento esteja na natureza dos dados analisados e nas metodologias empregadas na resolução dos problemas. Por exemplo, na área de diagnóstico de doenças a partir de dados médicos, os métodos são especialmente desenvolvidos para análises em batelada (CABRIETO *et al.*, 2018b), já que a detecção de uma doença (anomalia) não requer a análise da sequência ao mesmo tempo em que ela é gerada. Os dados também frequentemente são univariados, como por exemplo as batidas do coração (STAUDACHER *et al.*, 2005). Essas duas condições não são comuns em problemas de monitoramento de processos. Além do mais, dados de processo apresentam algumas características especiais, em particular a alta redundância e colinearidade entre as variáveis (KOURTI, 2002).

Pode-se afirmar que essa premissa continua em parte se justificando, pelo fato de que a própria natureza dos problemas nos diferentes domínios não mudou. No entanto, o momento intensamente disruptivo pelo qual todos os domínios estão passando (em especial o industrial, com a emergência da Indústria 4.0) faz com que a quase completa separação entre suas literaturas, no que se refere a um mesmo tema, passe a ser cada vez mais questionável. É improvável que não haja pelo menos alguma possibilidade de intercâmbio produtivo de ideias, mesmo sob a hipótese de que os métodos de resolução sejam diferentes. Essa hipótese, aliás, pode estar cada vez menos suportada pela realidade: trabalhos como o de ALIPPI *et al.* (2015), que estuda limitações teóricas da aplicação de métodos de detecção de mudanças

em dados multidimensionais gerados de forma contínua, têm escopos evidentemente sobrepostos com problemas de monitoramento.

Vale ressaltar que há outros domínios mais próximos que utilizam a mesma nomenclatura para o problema de monitoramento (detecção e diagnóstico de falhas) em que a comunicação é um pouco melhor; mas, de acordo com a percepção do presente autor ao analisar as citações dos artigos, essa comunicação ainda é insatisfatória. Exemplos típicos são as engenharias mecânica (FERNANDES *et al.*, 2022a; LIU *et al.*, 2018; SAINI *et al.*, 2022), nuclear (KATSER *et al.*, 2021; MA e JIANG, 2011), elétrica (ABBASI, 2022; KHOUKHI e KHALID, 2015) e mecatrônica (ABID *et al.*, 2020; ALOBAIDY *et al.*, 2020).

Streaming data

Na ciência de dados, a área que lida com a análise de dados gerados de forma contínua é conhecida como *streaming data* (dados em corrente, em tradução livre) (GOMES *et al.*, 2019; PHRIDVIRAJ e GURURAO, 2014; PSALTIS, 2017). Uma interessante ideia dessa área que pode ter bastante valia na área de monitoramento é o *concept drift*, nome pelo qual os cientistas de dados se referem à evolução, com o passar do tempo, de padrões e relações capturados pelas técnicas de aprendizado (AGRAHARI e SINGH, 2022; JANARDAN e MEHTA, 2017; KHAMASSI *et al.*, 2018; SOUZA *et al.*, 2020; TSYMBAL, 2004; ZLIOBAITE *et al.*, 2016). Essa ideia pode ser de valor significativo para a engenharia de sistemas em processos, pois o *drift* de processo constitui um dos estados de processo mais importantes e geralmente pode ser associado a problemas operacionais (KADLEC *et al.*, 2009). O jargão vem recentemente sendo adotado por trabalhos na área de processos, como evidenciado por JEONG e LEE (2020); SUN *et al.* (2020). Uma tendência recente é o uso de comitês (*ensembles*) de métodos para maior eficiência na adaptação à evolução dos dados (BARROS e SANTOS, 2019; CANO e KRAWCZYK, 2020; KRAWCZYK *et al.*, 2017; PIETRUCZUK *et al.*, 2017). Há trabalhos em *streaming data* potencialmente úteis no que se refere a agrupamento de dados (MOSHTAGHI *et al.*, 2016), pré-processamento (RAMIREZ-GALLEGO *et al.*, 2017), tratamento de *outliers* (GUPTA *et al.*, 2014; SADIK e GRUENWALD, 2014) e predição de eventos (LI *et al.*, 2020e).

Análise de causalidade

A análise de causalidade é um exemplo de aplicação baseada em análise de séries temporais que vem se destacando na área de monitoramento de processos, como

evidenciado por numerosos estudos (CHIANG e BRAATZ, 2003; CLAVIJO *et al.*, 2021; HU *et al.*, 2015; KATHARI e TANGIRALA, 2022; LI *et al.*, 2016; NAGHOOSI *et al.*, 2013; WANG *et al.*, 2023; YANG e BEQUETTE, 2023; YUAN e QIN, 2012; ZOPE *et al.*, 2023). Essa técnica usa testes estatísticos, como o teste de causalidade de Granger, para determinar se uma série temporal é útil para prever outra. Apesar dos avanços significativos feitos na literatura de engenharia de processos, parece haver uma falta de atenção de outros campos para esses trabalhos. Isso é particularmente evidente em estudos de revisão recentes sobre causalidade conduzidos por SHOJAIE e FOX (2022) e GUO *et al.* (2021a), em que não é feita referência a nenhum estudo recente da área de engenharia de processos.

Análise dinâmica não linear

A teoria da análise dinâmica não linear possui uma rica história na engenharia de processos (CLAVIJO *et al.*, 2022a,b; CLAVIJO, 2017; MELO *et al.*, 2003; OURIQUE *et al.*, 2002; PINTO e RAY, 1995; UPPAL *et al.*, 1974) e na análise de séries temporais (DIKS, 1999; GOOIJER, 2017; HUFFAKER *et al.*, 2017; KANTZ e SCHREIBER, 2004), apresentando potencial para mais aplicações nessa nova fase de Indústria 4.0. No entanto, na área de monitoramento de processos, esse potencial é pouco explorado. ALDRICH (2019) cita alguns métodos baseados em espaços de fase, como dimensão da correlação, expoentes de Lyapunov e entropia de informação, ressaltando que a adoção pela indústria ainda é baixa, por conta da influência do ruído na qualidade das descrições. Portanto, o potencial uso de ferramentas não lineares para representação e avaliação de estados de processo permanece amplamente inexplorado. As características não lineares de uma série temporal podem fornecer informações valiosas sobre a saúde do processo, indicar possíveis mudanças no seu comportamento e sugerir o uso de diferentes ferramentas numéricas para descrever qualitativa e quantitativamente o seu comportamento (ALDRICH, 2019).

2.11 Integração humano-máquina

A integração humano-máquina é um tópico que vem ganhando cada vez mais relevância em discussões relacionadas à Indústria 4.0 (VILLALBA-DIEZ e ORDIERES-MERÉ, 2021). Muitas vezes o ideal a ser atingido é a criação de sistemas totalmente automatizados, em que a participação humana em situações normais seja mínima ou mesmo inexistente. No entanto, isso é infactível em futuro próximo na maioria dos casos, o que leva à necessidade de desenvolver, implementar e usar estratégias eficientes

para se integrar a ação humana aos processos industriais cada vez mais digitalizados. Um termo muito utilizado na literatura para essa integração é *human-in-the-loop* ou HITL (GIL *et al.*, 2020).

Frameworks para a viabilização da integração humano-máquina vêm sendo propostos em vários contextos. LONGO *et al.* (2017) propuseram uma solução prática para integrar conteúdos de realidade aumentada de modo a potencializar as habilidades de operadores. Essa solução envolveu o fornecimento de *feedbacks* em tempo real, exposição a conteúdos de realidade aumentada e disponibilização de um assistente pessoal inteligente para interação com os operadores. COHEN *et al.* (2018) propuseram um *framework* para a interface entre o operador e a estação de trabalho. O sistema proposto coleta informações sobre o estado fisiológico, cognitivo e emocional do operador, bem como informações relacionadas ao ambiente de trabalho. Com base nessas informações, o objetivo é melhorar tanto as métricas objetivas quanto subjetivas de produção. GIL *et al.* (2020) elaboraram técnicas para auxiliar na análise e projeto de soluções HITL, identificando os desafios tecnológicos associados a essas soluções e propondo metodologias com foco em estratégias de controle e aspectos interacionais. BOUSDEKIS *et al.* (2020) desenvolveram uma arquitetura conceitual para uma simbiose ótima entre operadores e inteligência artificial. Nesse contexto, TAYLOR *et al.* (2020) ressaltaram a necessidade de se estabelecer como objetivo da integração humano-máquina a catálise do potencial criativo humano.

Um dos principais desafios relacionados à integração humano-máquina é a aceitação das metodologias por parte dos operadores. QIN e CHIANG (2019) destacaram a importância da interpretabilidade para que se estabeleça confiança nos modelos e algoritmos levando a que sejam adotados na prática industrial. BI *et al.* (2022) argumentaram que a interpretabilidade é o que constrói uma ponte de confiança entre os humanos e a inteligência artificial, devendo portanto ser uma componente fundamental de metodologias inteligentes de detecção e diagnóstico de falhas.

Outras abordagens propostas na literatura para integração humano-máquina incluem: aprendizado de máquina interativo, em que se integra o comportamento do usuário ao fluxo de aprendizado da inteligência artificial (AMERSHI *et al.*, 2014); técnicas de visualização interativa, em que o feedback humano aos sistemas de inteligência artificial se dá por meio de visualização (SACHA *et al.*, 2017); e controle automático não-linear utilizando aprendizado de comportamento humano (PERRUSQUIA *et al.*, 2021).

Do ponto de vista específico de monitoramento de processos, não foram muitos os trabalhos que discutiram o tema da integração humano-máquina. RANJIT *et al.*

(2015) propuseram uma metodologia para incorporar conhecimento de especialistas a um modelo misto de componentes principais e vizinhos mais próximos aplicado a um processo de manufatura de semicondutores. RIETH *et al.* (2018) estudaram métodos de avaliação de desempenho de técnicas de detecção de anomalias. Modelos baseados em análise lógica de dados, descritos na Seção 2.4.3, resultam em padrões interpretáveis que podem ser úteis para a geração de entendimento humano (RAGAB *et al.*, 2019). Uma das principais contribuições do presente trabalho é propor metodologias que propiciam integração humano-máquina em sistemas de monitoramento de processos, por meio do desenvolvimento de metodologias que incorporam i) alto nível de abstração durante a atividade de implementação computacional, ii) entendimento dos dados, e iii) interpretabilidade de resultados.

2.12 Considerações finais

O objetivo principal da revisão bibliográfica apresentada neste capítulo foi mapear, da maneira mais ampla possível, o estado da arte da área de monitoramento de processos baseado em dados, citando referências representativas de todas as vertentes e abordagens encontradas na literatura. Trata-se de uma área vasta e com muitos artigos de revisão recentemente publicados²⁵; esses artigos, porém, tratam de temas e/ou técnicas específicos e não se propõem a apresentar uma visão global de toda a área. Alguns tópicos estudados no presente trabalho não são objeto de revisão em nenhum desses artigos, como aplicações industriais, ferramentas computacionais e *benchmarks*, por exemplo.

A área de monitoramento de processos, com o advento da Indústria 4.0 e da transformação digital, vem adquirindo importância crescente nos ambientes acadêmico e corporativo. Na presente revisão, ficou clara a tendência histórica de ênfase em metodologias que extraem informações diretamente dos dados por meio de técnicas baseadas em aprendizado de máquina. Além do mais, as aplicações em sistemas industriais reais são na maioria baseadas na técnica PCA, o que indica simultaneamente conservadorismo por parte da comunidade e oportunidade para trabalhos inovadores. O presente trabalho pretende preencher algumas lacunas identificadas no que diz respeito à integração humano-máquina no *pipeline* de monitoramento, relacionadas em particular à arquitetura de *software*, exploração de dados e interpretabilidade de resultados de detecção e diagnóstico de falhas.

²⁵A lista completa de artigos de revisão analisados encontra-se no Apêndice A.

Capítulo 3

Implementação de técnicas de monitoramento

No presente capítulo é apresentada uma metodologia para implementação de técnicas de monitoramento de processos baseada no paradigma da orientação a objetos (OOP, *object-oriented programming*). Durante a atividade de programação, a integração humano-máquina se efetua por meio de níveis de abstração: quanto maior esse nível, mais as representações utilizadas se aproximam de conceitos humanos e estão distantes de conceitos específicos à computação. O paradigma OOP possibilita a implementação de modelos utilizando altos níveis de abstração de processos e dados, o que se traduz em softwares com maior legibilidade, reusabilidade, manutenibilidade e extensibilidade. A estratégia proposta no presente capítulo resulta em arquiteturas de monitoramento de processos com todas essas vantagens.

Nas próximas seções, serão apresentados o conceito de orientação a objetos, a metodologia proposta e um exemplo de aplicação utilizando o *software* BibMon (Biblioteca de Monitoramento de Processos) (MELO *et al.*, 2023).

3.1 Programação orientada a objetos

Um programa é um conjunto de instruções a serem fornecidas a um computador para que produza uma saída e/ou mude o próprio estado.

No paradigma da programação procedural, as instruções são organizadas em procedimentos conhecidos como *funções* ou *rotinas*, designados para receber dados de entrada, atuar sobre eles e produzir dados de saída. As funções em geral têm existência própria e se relacionam por meio dos dados da entrada e saída (a entrada

de uma função f pode ser a saída de outra função g , por exemplo) ou por meio de chamadas internas (um dos comandos contidos na função f pode ser a chamada da g). Encapsular funcionalidades em funções é uma maneira de efetuar *abstração de processos*: ao empregar funções, a complexidade dos procedimentos é abstraída, tornando visíveis no código de uso apenas suas entradas e saídas (SEBESTA, 2018). Apesar da grande vantagem advinda da organização da complexidade, o paradigma procedural encoraja o desenvolvimento com foco nos procedimentos computacionais, resultando em pouca ênfase nos dados processados ou na memória (estado) do computador. Para programas de grande porte, isso pode ser um limitador devido a problemas de comunicação entre as funções e vulnerabilidade dos dados, compartilhados livremente entre as várias funções.

No paradigma da orientação a objetos, dados e funções que operam nesses dados são encapsulados em entidades computacionais conhecidas como *objetos*. Encapsular dados e as respectivas funcionalidades em objetos é uma maneira de efetuar *abstração de dados*, permitindo um nível de gerenciamento de complexidade que vai um passo além da pura abstração de processos (SEBESTA, 2018). Ao contrário do paradigma procedural, o paradigma OOP permite o desenvolvimento com foco nos dados: cada objeto contém seu próprio conjunto de dados e funções designadas especialmente para operar sobre eles. No jargão do paradigma OOP, os dados de um objeto são chamados de *atributos* e as funções são chamadas de *métodos*. O paradigma está hoje entre os mais utilizados na indústria de *software* (LIAN *et al.*, 2022; PAZEL, 2022). No contexto da engenharia química, aplicações foram propostas por exemplo por SHI e LIANG (2011), que montaram uma estrutura para cálculo de propriedades termodinâmicas via equações de estado, e FONTOURA *et al.* (2022), que modelaram unidades *offshore* de separação de gás por membranas em diferentes níveis de configuração.

Nas próximas subseções, são apresentados os principais conceitos do paradigma OOP.

3.1.1 Tipos abstratos de dados e classes

Tipos abstratos de dados são modelos para a criação de objetos. Na maioria das linguagens de programação com suporte ao paradigma OOP, um tipo abstrato de dado é definido por meio de uma *classe*, que é um código descrevendo a interface e a implementação de um objeto (GAMMA *et al.*, 1994). Do ponto de vista da modelagem, classes podem ser encaradas como responsáveis por representar alguma ideia ou entidade do mundo real. Por exemplo, SHI e LIANG (2011) propõem a utilização de classes para a representação de modelos termodinâmicos do tipo

equações de estado.

3.1.2 Herança

A *herança* ou *derivação* é o mecanismo por meio do qual subtipos particulares são definidos a partir de tipos genéricos. A classe que representa o subtipo é denominada *classe derivada*, enquanto a classe correspondente ao tipo genérico é denominada *classe base*. Uma classe derivada herda todos os atributos e métodos de sua classe base, abrindo possibilidade para escrita de novos atributos e métodos ou reescrita dos antigos. Na aplicação proposta por SHI e LIANG (2011), por exemplo, a classe genérica que representa modelos de equação de estado dá origem às classes particulares que representam equações de estado cúbicas, que por sua vez originam classes ainda mais específicas representando equações como Soave-Redlich-Kwong (SOAVE, 1972) ou Peng-Robinson (PENG e ROBINSON, 1976). A herança também é conhecida como relação *is-a*, já que a classe derivada é uma representação particularizada da classe base.

3.1.3 Composição

A *composição* é um mecanismo que permite que uma classe defina um de seus atributos como um objeto instanciado de outra classe. Na aplicação proposta por FONTOURA *et al.* (2022), por exemplo, as relações entre os vários componentes de uma unidade de separação por membranas são modeladas por meio de relações de composição, uma escolha natural pois fisicamente os componentes mais internos estão contidos dentro dos mais externos. A composição também é conhecida como relação *has-a*, já que uma classe *contém* uma instância de outra classe.

3.1.4 Polimorfismo

O *polimorfismo* se refere ao uso de uma interface comum entre as classes base e derivadas, mas com diferentes implementações em cada classe. Na aplicação proposta por SHI e LIANG (2011), por exemplo, a classe base que representa as equações de estado possui um método para calcular o volume utilizando um algoritmo numérico genérico de busca de raízes de equações; na classe derivada que representa as equações de estado cúbicas, o método é substituído por um algoritmo analítico específico mantendo a mesma interface (ou seja, o acesso ao método é feito da mesma maneira em ambas as classes).

3.2 Metodologia

A metodologia é apresentada nesta seção na forma de recomendações para implementações relacionadas a modelagem, tratamento de dados e funções de conveniência. Na próxima seção, utiliza-se como exemplo específico de implementação o *software* BibMon (Biblioteca de Monitoramento de Processos) (MELO *et al.*, 2023).

3.2.1 Modelos de monitoramento

Modelos de monitoramento de processos com base em controle estatístico multivariado possuem algumas características em comum que podem ser utilizadas para implementação de classes abstratas:

- cartas de controle com índices de detecção de falhas;
- contribuições para o índice de detecção de falhas, de modo a possibilitar o diagnóstico;
- lógicas de alarmes;
- *pipelines* de tratamento de dados;
- métodos para treino e aplicação (teste) de modelos;
- ajuste de hiperparâmetros.

A proposta é que se implemente uma classe genérica com todas essas funcionalidades, abstraindo-se ao máximo as especificidades, de modo que sejam implantadas em classes específicas. Em particular, as características particulares de cada modelo devem ser implementáveis por meio do mecanismo de herança e os *pipelines* de tratamento de dados devem fazer parte da classe genérica por meio do mecanismo de composição, conforme descrito a seguir.

3.2.1.1 Extensibilidade por herança

Especificidades de modelos de monitoramento de processos baseados em controle estatístico multivariado podem ser divididas para implementação em duas funções:

- **Função de preparação:** após aplicação do pré-processamento de dados, esta função prepara o modelo para efetuar a predição ou reconstrução dos dados. No

caso de um modelo PCA, por exemplo, essa função consistiria da decomposição em valores e vetores característicos para cálculo dos componentes principais, conforme descrito na Seção 2.4.1.

- **Função de predição ou reconstrução:** recebe um conjunto de dados de entrada e retorna o conjunto de dados predito ou reconstruído. No caso de um modelo PCA, essa etapa consiste na projeção dos dados na direção dos componentes principais e reprojeção no espaço original, o que resulta no conjunto de dados reconstruído. É importante ressaltar que a aplicação do polimorfismo possibilita que a mesma interface na classe abstrata seja utilizada tanto para modelos de predição quanto de reconstrução.

A vantagem dessa estratégia é a resultante facilidade de extensão do *software*: para adicionar um novo modelo, basta que se implemente duas funções contendo apenas suas correspondentes especificidades algorítmicas.

3.2.2 Tratamento de dados

A ideia é que se implemente uma classe com métodos de tratamento de dados cujos objetos sejam atributos da classe de monitoramento. Dessa maneira, diferentes atributos podem ser criados, de acordo com as diferentes aplicações ao longo do *pipeline* de monitoramento. Além do mais, por meio da estratégia do polimorfismo, uma mesma interface pode ser utilizada para utilização de diferentes técnicas (tratamento de dados faltantes, filtro de ruído, etc.), com dados de diferentes naturezas (preditores, preditos ou reconstruídos) e em diferentes etapas (treino ou teste). Também é fácil estender o *software* com novas funções de tratamento de dados, bastando-se programar um novo método para a referida classe.

3.2.3 Funções de conveniência

Além das classes de modelos e tratamento de dados, o software pode conter outras funcionalidades que fazem uso das classes ou são usadas por elas, como carregamento de dados e geração de tabelas comparativas de resultados, por exemplo. Essas funcionalidades podem ser implementadas à parte e eventualmente disponibilizadas no espaço de nomes do *software* para acesso do usuário final.

Na próxima seção, é apresentado um exemplo de aplicação da metodologia proposta.

3.3 Estudo de caso: BibMon

A BibMon (Biblioteca de Monitoramento de Processos) (MELO *et al.*, 2023) é um pacote Python desenvolvido em projetos de parceria entre a COPPE/UFRJ e o CENPES/Petrobras. Seu propósito é fornecer modelos preditivos baseados em desvios para detecção e diagnóstico de falhas e acompanhamento da condição de processos. Sua arquitetura segue o padrão de implementação proposto neste capítulo.

Em essência, a utilização da biblioteca se dá em duas etapas:

1. na etapa de treinamento, gera-se um modelo que captura as relações entre as variáveis na condição normal do processo;
2. na etapa de predição, dados de processo são comparados com as previsões do modelo, resultando em desvios; caso estes estejam acima de um limite pré-definido, alarmes são acionados.

Especificamente, as cartas de controle implementadas baseiam-se no desvio quadrático (*Squared Prediction Error*, SPE).

Os recursos oferecidos pela BibMon são:

- Modelos PCA (Principal Component Analysis), SBM (Similarity-Based Models), ESN (Echo-State Network) e qualquer regressor que utilize a interface do pacote *scikit-learn*.
- Aplicação em sistemas *online*: um modelo na BibMon, após treinado, pode ser utilizado para análises *online* tanto com amostras individuais quanto com janelas de dados. A cada amostra ou janela, uma previsão é efetuada, o estado do modelo é atualizado e alarmes são calculados.
- Compatibilização, em uma mesma arquitetura, de modelos de regressão (ou seja, sensores virtuais, contendo dados separados em X e Y , como a regressão linear) e reconstrução (contendo apenas dados X , como o modelo PCA).
- *Pipelines* de pré-processamento que levam em conta as diferenças entre dados X e Y e entre as etapas de treino e teste.
- Disponibilidade de diversas lógicas de alarmes.
- Funções de conveniência para efetuar análises *offline* automáticas e traçar cartas de controle, diagnóstico, entre outras.

- Conjuntos de dados de processos reais e simulados disponíveis na própria biblioteca.
- Tabelas comparativas para automatizar a análise de desempenho de diferentes modelos.
- Sintonização automática de hiperparâmetros.

A Figura 3.1 apresenta a estrutura da BibMon. Os módulos disponíveis para serem utilizados pelo usuário são aqueles relacionados aos modelos (`esn`, `pca`, `sbm` e `sklearn_regressor`), dados (`carregar_dados` e `pre_process`) e funções de conveniência (`bibmon_tools`). A classe genérica mencionada na Seção 3.2.1, responsável por todas as funcionalidades comuns aos diferentes modelos de regressão ou reconstrução, é chamada de `ModeloGenerico` e está presente no módulo `modelo_generico`. É importante notar que esse módulo não está disponível no espaço de nomes da biblioteca, não podendo ser importado pelo usuário, mas pelos módulos que implementam os modelos específicos.

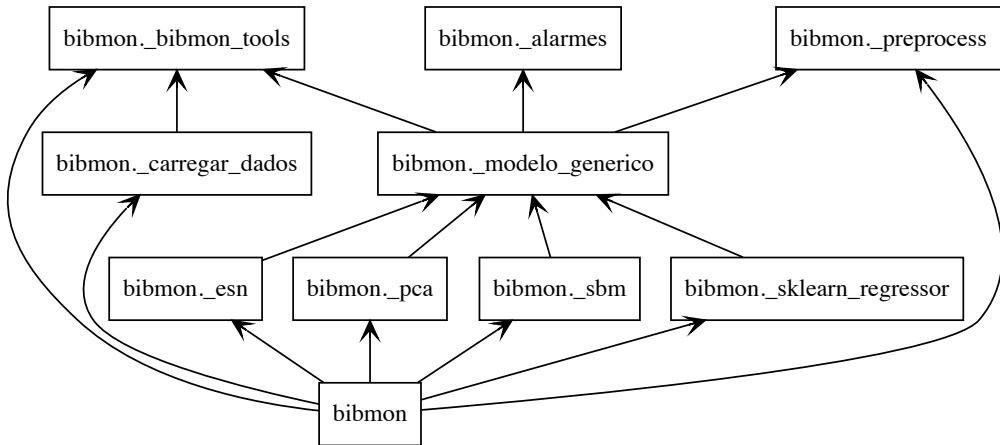


Figura 3.1: Estrutura do pacote BibMon.

The Unified Modeling Language (UML) é uma linguagem de modelagem útil para representação de sistemas orientados a objetos, sendo o padrão de representação adotado pela indústria (BOOCH *et al.*, 2005). A Figura 3.2 apresenta o diagrama UML da BibMon. Cada retângulo correspondente a uma classe é dividido em três porções, contendo respectivamente nome, atributos e métodos. Relações de herança são representadas por setas com triângulos nas pontas, enquanto relações de composição são representadas por setas com losangos nas pontas. Os métodos descritos na Seção 3.2.1.1 estão presentes nas classes dos modelos: a função de

preparação recebe o nome de `core_treino`, enquanto a função de predição ou reconstrução recebe o nome de `predicao`. A estratégia de implementação de técnicas de pré-processamento por meio da relação de composição, descrita na Seção 3.2.2, é representada pelo par de variáveis `pre_proc_X` e `pre_proc_Y`, atributos da classe `ModeloGenerico`.

As Figuras 3.3 a 3.5 apresentam exemplos de resultados gerados pela biblioteca. Na Figura 3.3 representa-se uma carta de controle de detecção de falhas gerada com o método PCA aplicado à falha 1 do *benchmark Tennessee Eastman Process*. Na Figura 3.4, observa-se o mapa de diagnóstico por contribuições parciais para o mesmo caso da figura anterior. Na Figura 3.5, apresentam-se cartas de controle e de sensor virtual de treino, validação e teste geradas com o modelo *Random Forest* aplicado a dados reais de processo. Nota-se nos dados de treino e validação o baixo erro da regressão e nos dados de teste o descolamento do modelo quando o processo entra em falha, de acordo com o esperado. Esses exemplos encontram-se disponibilizados em um tutorial que faz parte do manual da biblioteca.

É importante ressaltar que a BibMon vem sendo utilizada com sucesso há mais de um ano no sistema *Smart Monitor* da Petrobras, fornecendo suporte a dezenas de aplicações de monitoramento em tempo real de ativos sensíveis no setor *upstream*, como medidores fiscais.

3.4 Considerações finais

Neste capítulo apresentou-se uma estratégia para arquitetura de *software* utilizando o paradigma da orientação a objetos. A estratégia é voltada especialmente para a implementação de técnicas de monitoramento de processos e tem como objetivos principais: i) geração de códigos mais organizados, modulares e extensíveis, aumentando a integração entre as pessoas desenvolvedoras e a base de implementação, e ii) facilitar a implementação em sistemas em tempo real. A BibMon (MELO *et al.*, 2023), uma biblioteca para monitoramento de processos contendo modelos preditivos baseados em desvios, foi apresentada como exemplo bem-sucedido de aplicação da estratégia na indústria.

No capítulo a seguir, é proposta uma metodologia para análise exploratória de dados de processos que deu origem a outro software, a KydLIB (Know Your Data Library) (MELO, 2023).

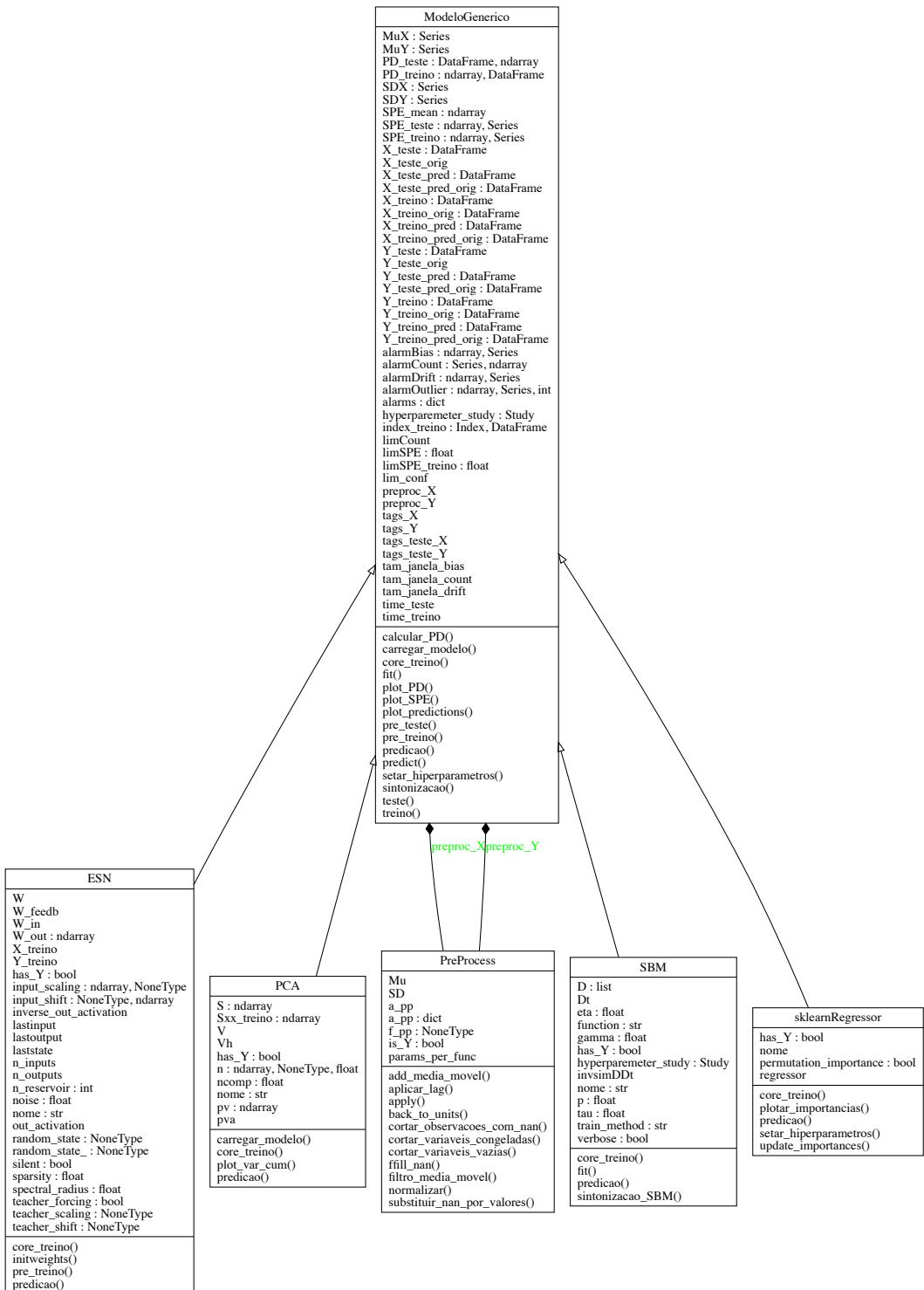


Figura 3.2: Diagrama UML de classes do pacote BibMon.

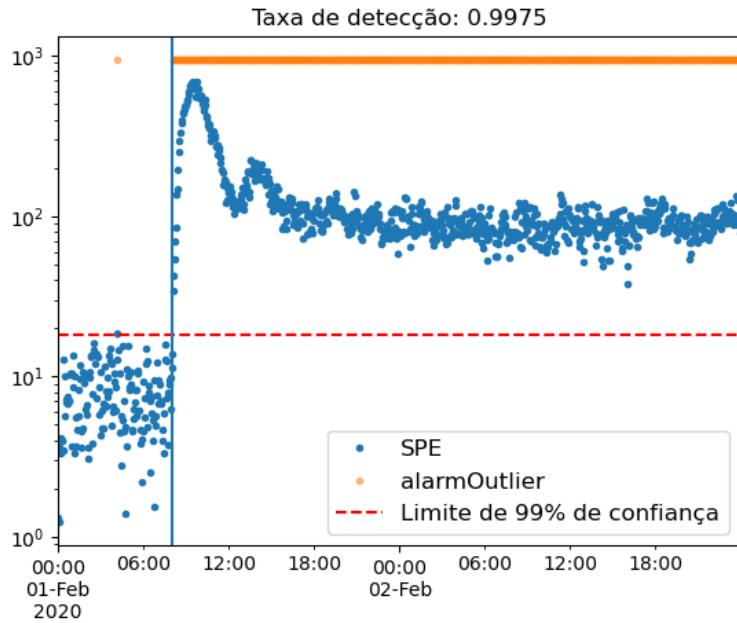


Figura 3.3: Resultados de detecção de falhas para a Falha 1 do *benchmark* TEP.

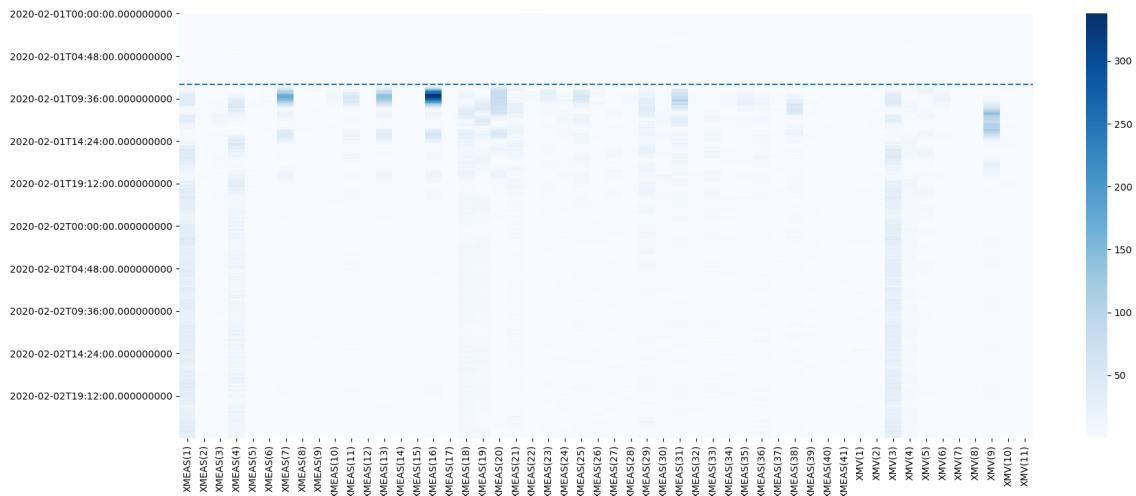


Figura 3.4: Resultados de diagnóstico de falhas para a Falha 1 do *benchmark* TEP.

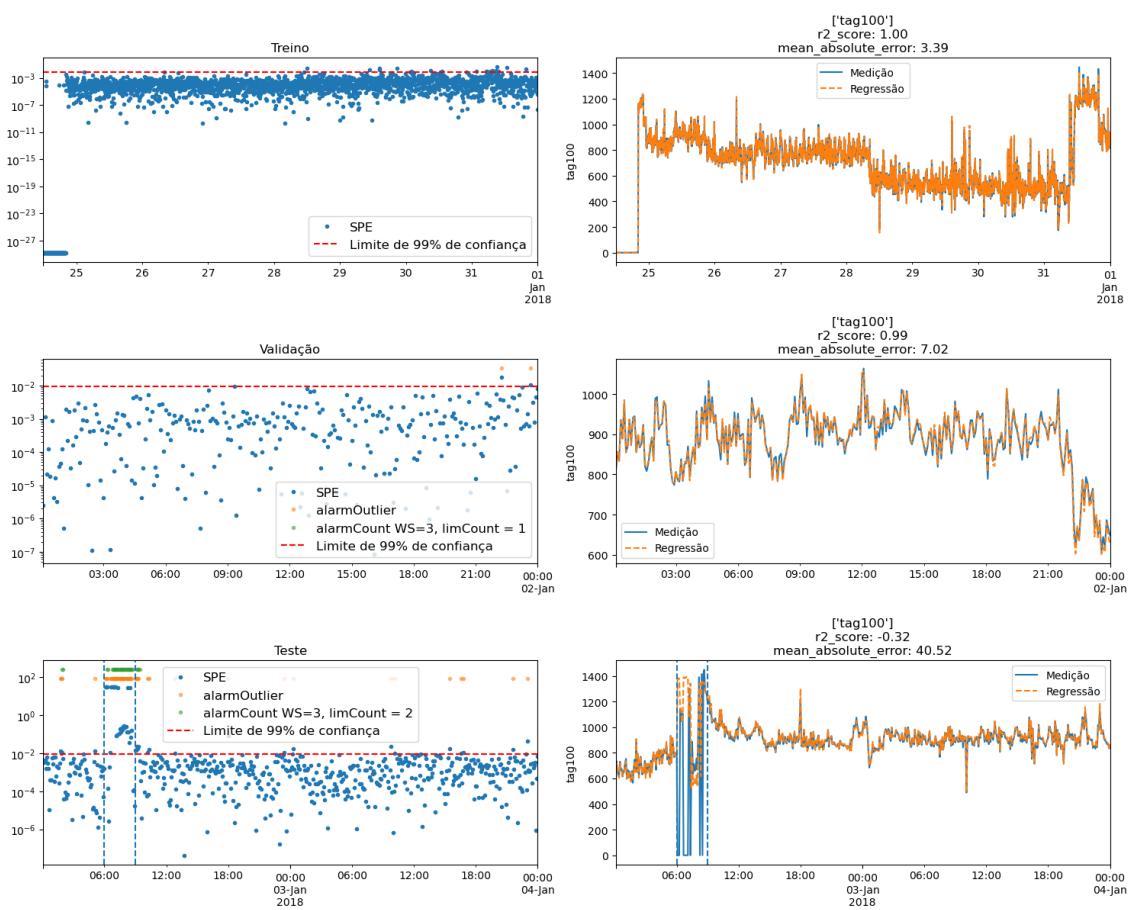


Figura 3.5: Resultados de detecção de falhas para um conjunto de dados reais de planta de processo.

Capítulo 4

Análise exploratória de dados de processo

Neste capítulo uma metodologia para análise exploratória de dados de processo é proposta. A metodologia é baseada na combinação de várias técnicas de análise de dados disponíveis na literatura, em particular relacionadas a correlação, gaussianidade, dinâmica e ruído. A proposta foi ilustrada com auxílio de aplicações em *benchmarks* abertos da literatura de monitoramento de processos. As aplicações levaram a várias conclusões interessantes sobre os *benchmarks* que não estavam claras na literatura até então. Uma biblioteca Python chamada KydLIB (Know Your Data Library) (MELO, 2023) encapsulando a metodologia foi disponibilizada no Python Package Index (PyPI, acessível em <https://pypi.org/>) e no seguinte repositório GitHub: <https://github.com/afraniomelo/kydlib>.

Nas seções a seguir, são apresentadas a metodologia proposta e os resultados obtidos, respectivamente.

4.1 Metodologia

Nesta seção é descrita a metodologia utilizada para caracterizar os conjuntos de dados. O teste de gaussianidade e o método de quantificação de não linearidade foram adaptados do teste de características dos dados proposto por ZHANG *et al.* (2016). A estimativa da razão sinal/ruído foi inspirada na técnica de espectros de variância, proposta por FEITAL e PINTO (2015). A quantificação da dinâmica é baseada na função de autocorrelação amostral de cada variável. Conjuntos de dados com altos níveis de não gaussianidade, dinâmica, ruído e não linearidade são particularmente desafiadores para o desempenho de algoritmos de monitoramento

de processos. Então, em teoria, eles seriam sensíveis a diferenças entre algoritmos, servindo como bons *benchmarks*.

Em todas as subseções a seguir, considera-se que os dados estão organizados em uma matriz \mathbf{X} com n linhas representando observações, m colunas representando variáveis e posto $p \leq m$.

4.1.1 Visualização de dados

Duas estratégias foram usadas para realizar a visualização de dados:

- Gráficos de séries temporais para avaliar visualmente a dinâmica das variáveis individuais.
- Gráficos de dispersão para avaliar visualmente as relações bidimensionais entre as variáveis.

Gráficos de séries temporais foram obtidos com o pacote Matplotlib. Os gráficos de dispersão foram obtidos com o pacote datashader (HOLOVIZ, 2022), que utiliza uma técnica conhecida como rasterização para visualizar grandes conjuntos de dados, plotando um pixel por ponto de dados e indicando a superposição de pontos com intensidade de cor.

4.1.2 Teste de gaussianidade

O teste multivariado de gaussianidade adotado neste trabalho foi proposto por ZHANG *et al.* (2016) e será brevemente descrito aqui. O teste funciona assumindo que, se os dados seguem uma distribuição gaussiana multivariada, o quadrado da distância de Mahalanobis da média segue a distribuição F , de acordo com a equação:

$$D(\mathbf{x}, \mu) = (\mathbf{x} - \bar{\mu})^T \mathbf{S}^{-1} (\mathbf{x} - \bar{\mu}) \sim \frac{m(n^2 - 1)}{n(n - m)} F(m, n - m), \quad (4.1)$$

onde $\bar{\mu}$ e \mathbf{S} são estimativas para a média populacional e matriz de covariância, respectivamente. Quando \mathbf{S} é singular, m pode ser substituído pelo posto de \mathbf{S} , e \mathbf{S}^{-1} pode ser substituído pela inversa generalizada de Moore-Penrose (BAKSALARY e TRENKLER, 2021). Computacionalmente, o teste verifica se as estatísticas de ordem de D são equivalentes aos respectivos quantis da distribuição F , ou seja, se a

relação $D \times F$ se ajusta à reta $F_t = a + bD_t$, com:

$$b = \frac{\sum_{t=1}^n (D_t - \bar{D})(F_t - \bar{F})}{\sum_{t=1}^n (D_t - \bar{D})^2}, \quad (4.2)$$

$$a = \bar{F} - b\bar{D}, \quad (4.3)$$

$$\bar{D} = \frac{1}{n} \sum_{t=1}^n D_t, \quad (4.4)$$

$$\bar{F} = \frac{1}{n} \sum_{t=1}^n F_t. \quad (4.5)$$

O desvio padrão de regressão pode ser calculado como:

$$s = \sqrt{\frac{\sum_{t=1}^n (F_t - (a + bD_t))^2}{n - 2}}. \quad (4.6)$$

Seguindo ZHANG *et al.* (2016), o modelo de regressão linear é considerado estatisticamente significativo se $s/\bar{F} < 0,15$. Se for o caso, a e b são comparados com 0 e 1, respectivamente; isso é feito avaliando-se as desigualdades:

$$|a| \leq \sigma, \quad (4.7)$$

$$|b - 1| \leq \beta, \quad (4.8)$$

onde σ e β são limites definidos como $0,05 \times \bar{F}$ e $0,2$, respectivamente. Se as desigualdades acima forem válidas, pode-se admitir que $F_t \sim D_t$ e os dados seguem uma distribuição gaussiana multivariada.

4.1.3 Análise de autocorrelação

A autocorrelação mede a dependência entre observações atrasadas de uma série temporal (BOX *et al.*, 2015). A função de autocorrelação amostral (FAC) é definida

como:

$$\rho_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (4.9)$$

onde k é a defasagem entre as observações. A FAC pode ser usada para avaliar o nível de comportamento dinâmico de um conjunto de dados. WEI e SONG (2020), por exemplo, calcularam a média das autocorrelações de cada variável usando defasagens de 1 a 5 para obter valores representativos para cada defasagem. No presente trabalho, a métrica adotada foi o tempo, identificado a partir das defasagens, necessário para se atingir²⁶ um valor de autocorrelação de 0,5. Isso permite a análise em uma ampla gama de defasagens, pois há grande variabilidade no comportamento das autocorrelações entre diferentes variáveis e diferentes conjuntos de dados.

4.1.4 Estimação da razão sinal/ruído

A relação sinal-ruído (RSR) é uma medida do nível de um sinal em relação ao nível de ruído de fundo. A técnica aqui escolhida para estimar a RSR foi o espectro de variâncias, proposto por FEITAL e PINTO (2015). Um espectro de variâncias é um conjunto de variâncias calculado à medida que alguma variável relacionada evolui. Ao se analisar uma série temporal, pode-se correr uma janela móvel ao longo dos dados e escolher o comprimento da janela (WL) como a variável relacionada, computando um valor representativo para cada comprimento de janela. A agregação pode ser feita com o uso de médias ou medianas, por exemplo. No caso das médias, o espectro de variâncias é descrito pela equação:

$$\sigma_{WL}^2 = \frac{\sum_{k=1}^{NW} \sum_{i=k}^{k+WL-1} (x_i - \bar{x})^2}{NW}, \quad (4.10)$$

onde NW é o número de janelas usadas em cada ponto do espectro. FEITAL e PINTO (2015) analisaram todo o espectro, para todos os tamanhos, mas aqui o interesse reside nos dois casos extremos:

- $WL = 2$: Nesse caso, na hipótese de duas observações consecutivas serem réplicas de uma mesma medição, a variância calculada dentro da janela será equivalente à variância do ruído de medição.
- $WL = n$: Esta é a variância total da série temporal analisada, ou seja, a variância do sinal.

²⁶É interessante ressaltar que tipicamente os valores de autocorrelação diminuem com o aumento das defasagens, já que amostras mais distantes são em geral menos dependentes entre si.

A razão $\sigma_{WL=n}^2/\sigma_{WL=2}^2$ constitui assim uma estimativa da razão sinal/ruído.

4.1.5 Análise de correlações lineares e não lineares

A análise de correlação linear foi realizada usando o coeficiente de correlação amostral de Pearson (SCHWAAB e PINTO, 2007):

$$\rho_k = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (4.11)$$

A análise de correlação não linear foi realizada usando o pacote ennemi (LAARNE *et al.*, 2021). Esse pacote usa informações mútuas (IM) para estimar a dependência geral entre duas variáveis. A informação mútua é definida em relação à entropia da informação como:

$$I(X, Y) = H(X) + H(Y) - H(X, Y), \quad (4.12)$$

onde $H(X)$ é a entropia de informação da variável X e $H(X, Y)$ é a entropia de informação conjunta entre X e Y . $I(X, Y)$ varia entre 0 e ∞ , mas LAARNE *et al.* (2021) propõem um método de normalização para gerar um coeficiente de correlação IM entre 0 e 1:

$$\rho_{I(X,Y)} = \sqrt{1 - \exp -2I(X, Y)}. \quad (4.13)$$

Quando X e Y seguem conjuntamente uma distribuição normal, o coeficiente de correlação IM é igual ao coeficiente de correlação de Pearson.

ZHANG *et al.* (2016) definiram um coeficiente de correlação não linear como a diferença entre informações mútuas e coeficientes de Pearson:

$$r_{X,Y} = \rho_{I(X,Y)} (1 - |\rho_{X,Y}|). \quad (4.14)$$

Além disso, os autores também propuseram um coeficiente de correlação não linear geral para fornecer uma medida de não linearidade para todo o conjunto de dados, definido como:

$$r = \sqrt{\frac{\sum_{i,j=1}^m r_{i,j}^2}{m^2 - m}}. \quad (4.15)$$

No presente trabalho, um coeficiente de correlação linear geral análogo foi

definido para permitir a comparação de comportamentos lineares e não lineares:

$$\rho = \sqrt{\frac{\sum_{i,j=1}^m \rho_{i,j}^2 - m}{m^2 - m}}. \quad (4.16)$$

4.1.6 Análise de falhas

As falhas são sumarizadas por:

- como as falhas se desenvolvem ao longo do tempo (abruptas, incipientes ou intermitentes);
- uma descrição qualitativa do evento relacionado ao comportamento anormal, quando disponível;
- resultados de detecção de falhas para um modelo PCA de referência. Todos os modelos foram treinados definindo-se o número de componentes principais como o que faz o modelo capturar 90% da variância dos dados.

4.2 Aplicações e resultados

Serão apresentados resultados da aplicação da metodologia para cinco *benchmarks* abertos da literatura: Tennessee Eastman Process, DAMADICS, IndPenSim, Cranfield Multiphase Flow Facility, PRONTO e 3W. Os *benchmarks* foram descritos com detalhes no Capítulo 2.

4.2.1 Tennessee Eastman Process (TEP)

O conjunto de dados TEP analisado nesta seção foi o conjunto proposto por RIETH *et al.* (2017).

4.2.1.1 Dados normais

Existem 52 variáveis medidas no conjunto de dados, listadas na Tabela 4.1.

Tabela 4.1: Variáveis do *benchmark* TEP.

No.	Tag	Nome	Unidades
1	xmeas_1	A Feed (stream 1)	kscmh
2	xmeas_2	D Feed (stream 2)	kg/hr
3	xmeas_3	E Feed (stream 3)	kg/hr
4	xmeas_4	A and C Feed (stream 4)	kscmh
5	xmeas_5	Recycle Flow (stream 8)	kscmh
6	xmeas_6	Reactor Feed Rate (stream 6)	kscmh
7	xmeas_7	Reactor Pressure	kPa gage
8	xmeas_8	Reactor Level	%
9	xmeas_9	Reactor Temperature	Deg C
10	xmeas_10	Purge Rate (stream 9)	kscmh
11	xmeas_11	Product Sep Temp	Deg C
12	xmeas_12	Product Sep Level	%
13	xmeas_13	Prod Sep Pressure	kPa gage
14	xmeas_14	Prod Sep Underflow (stream 10)	m3/hr
15	xmeas_15	Stripper Level	%
16	xmeas_16	Stripper Pressure	kPa gage
17	xmeas_17	Stripper Underflow (stream 11)	m3/hr
18	xmeas_18	Stripper Temperature	Deg C
19	xmeas_19	Stripper Steam Flow	kg/hr
20	xmeas_20	Compressor Work	kW
21	xmeas_21	Reactor Cooling Water Outlet Temp	Deg C
22	xmeas_22	Separator Cooling Water Outlet Temp	Deg C
23	xmeas_23	6A	% mol
24	xmeas_24	6B	% mol
25	xmeas_25	6C	% mol
26	xmeas_26	6D	% mol
27	xmeas_27	6E	% mol
28	xmeas_28	6F	% mol
29	xmeas_29	9A	% mol
30	xmeas_30	9B	% mol
31	xmeas_31	9C	% mol
32	xmeas_32	9D	% mol
33	xmeas_33	9E	% mol
34	xmeas_34	9F	% mol
35	xmeas_35	9G	% mol

Continua na próxima página

Tabela 4.1 – *Continuada da página anterior*

No.	Tag	Nome	Unidades
36	xmeas_36	9H	% mol
37	xmeas_37	11D	% mol
38	xmeas_38	11E	% mol
39	xmeas_39	11F	% mol
40	xmeas_40	11G	% mol
41	xmeas_41	11H	% mol
42	xmv_1	D Feed Flow (stream 2)	kg/h
43	xmv_2	E Feed Flow (stream 3)	kg/h
44	xmv_3	A Feed Flow (stream 1)	kscmh
45	xmv_4	A and C Feed Flow (stream 4)	kscmh
46	xmv_5	Compressor Recycle Valve	%
47	xmv_6	Purge Valve (stream 9)	%
48	xmv_7	Separator Pot Liquid Flow (stream 10)	m3/hr
49	xmv_8	Stripper Liquid Product Flow (stream 11)	m3/hr
50	xmv_9	Stripper Steam Valve	%
51	xmv_10	Reactor Cooling Water Flow	m3/hr
52	xmv_11	Condenser Cooling Water Flow	m3/hr

Abaixo são apresentadas as classificações das variáveis e os respectivos intervalos de amostragem:

- 1 a 22: medições de processo contínuas (3 min);
- 23 a 28: análise de alimentação do reator (6 min);
- 29 a 36: análise de gás de purga (6 min);
- 37 a 41: análise do produto (15 min);
- 42 a 52: variáveis manipuladas (3 min).

A Figura 4.1 mostra as séries temporais para uma simulação típica do *benchmark* TEP em condições normais. A maioria das variáveis consiste de flutuações de ruído em torno dos respectivos *set-points*. Oscilações de longo prazo em algumas variáveis são identificáveis por inspeção visual.

A Figura 4.2 mostra os gráficos de dispersão para medições de processos contínuos e variáveis manipuladas de uma simulação, que normalmente é usada como

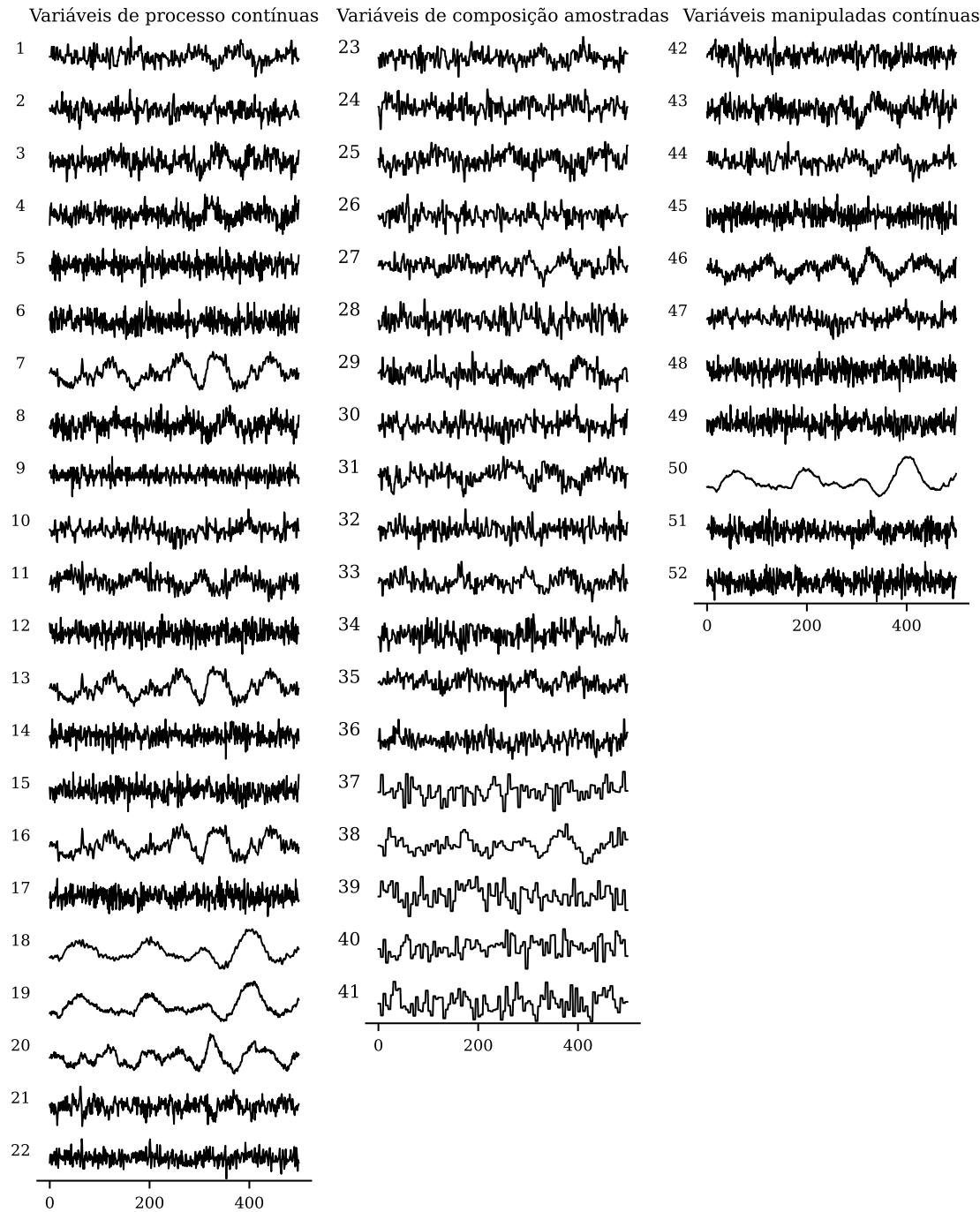


Figura 4.1: Séries temporais dos dados normais do *benchmark* TEP.

dados de treinamento em aplicações de monitoramento de processos. A Figura 4.3 mostra gráficos de dispersão equivalentes obtidos ao utilizar todas as 250 simulações disponíveis no conjunto de dados. Ambas as figuras mostram que as distribuições parecem ser de natureza gaussiana, algumas com certo grau de correlação linear; no entanto, no gráfico de dispersão com 250 simulações, isso fica evidente. A Figura 4.4 mostra como o aumento do número de simulações destaca os comportamentos gaussianos simples e relativamente uniformes das variáveis.



Figura 4.2: Gráficos de dispersão de variáveis contínuas medidas e manipuladas de uma simulação do *benchmark* TEP.

As Figuras 4.5 e 4.6 mostram testes de gaussianidade para conjuntos de dados com uma e três simulações, respectivamente. Na Figura 4.5, é evidente que as distribuições multivariadas são aproximadamente gaussianas, já que os pontos se distribuem de maneira linear ao longo da reta $D_t = F_t$. Os testes, no entanto, não são inequívocos: para as medições contínuas e amostradas, os resultados dos testes são falsos, de acordo com as Equações 4.7 e 4.8. Já na Figura 4.6, pode-se observar que três simulações são suficientes para que a gaussianidade dos dados seja atestada

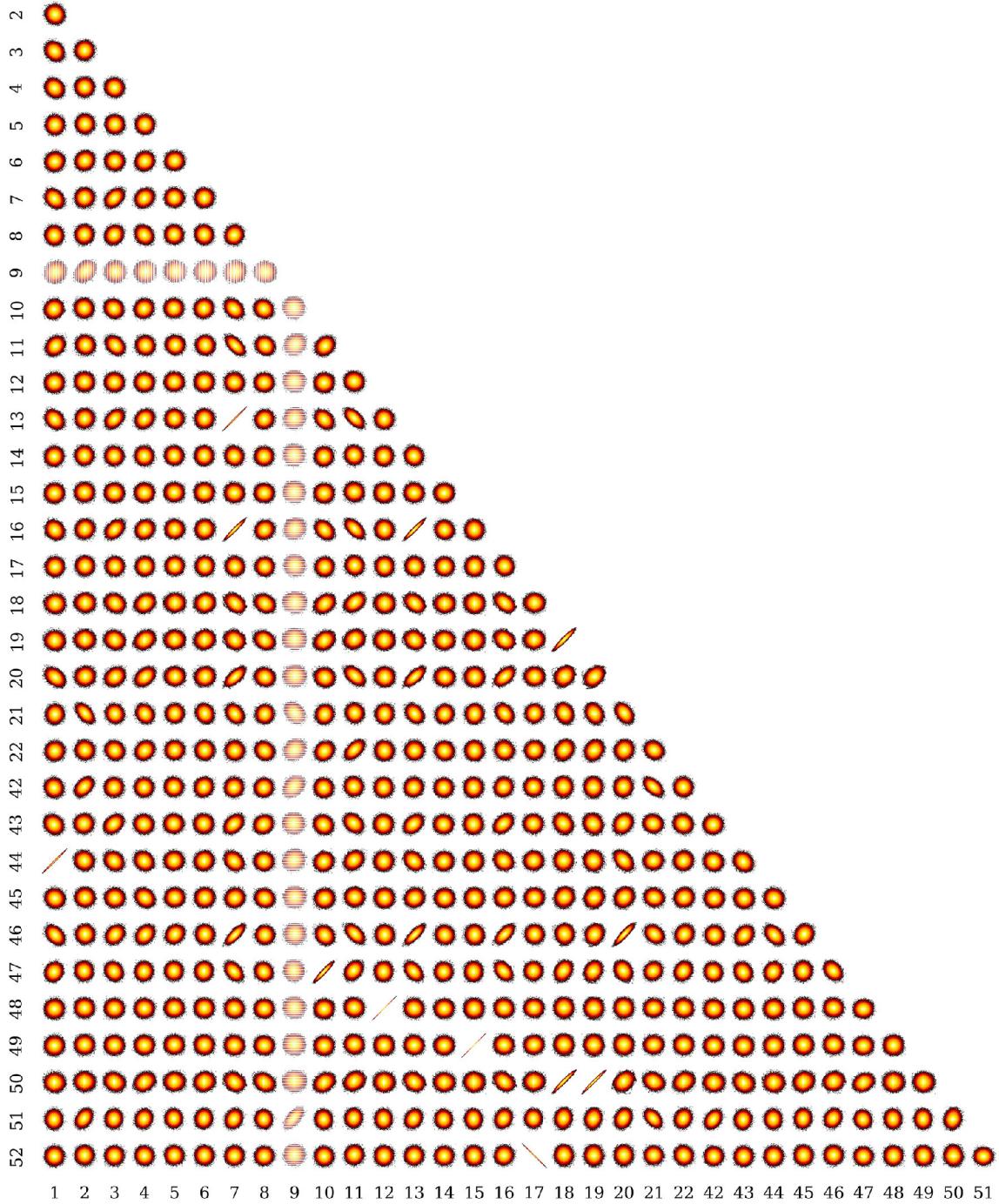


Figura 4.3: Gráficos de dispersão de variáveis contínuas medidas e manipuladas de quinhentas simulação do *benchmark* TEP.

definitivamente.

A maioria das variáveis tem um baixo grau de autocorrelação. Na Figura 4.7, os gráficos de autocorrelação são plotados para três variáveis. Nos dois primeiros gráficos, cujos perfis são semelhantes aos da maioria dos outros, a autocorrelação é baixa mesmo em valores de defasagem baixos. Em algumas variáveis que apresentam flutuações de longo prazo, a autocorrelação começa alta e diminui progressivamente, como na

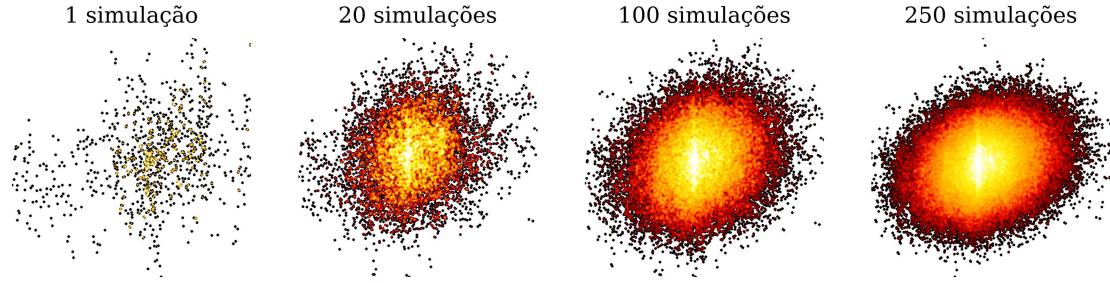


Figura 4.4: Efeito da adição de mais simulações do *benchmark* TEP no gráfico de dispersão “Válvula de purga” X “Válvula de vapor do esgotador”.

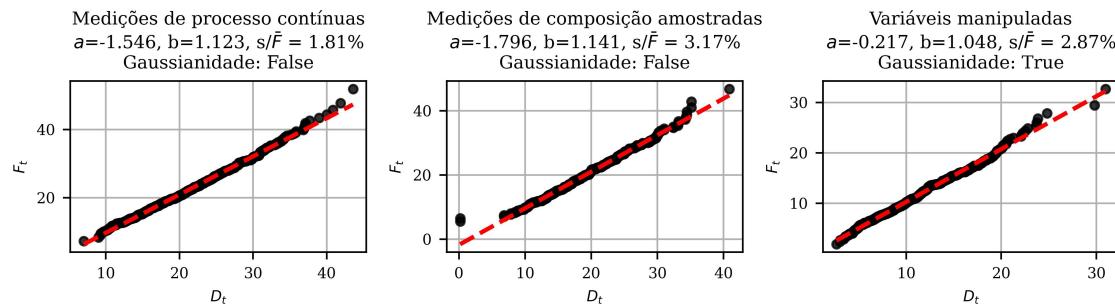


Figura 4.5: Testes de gaussianidade do *benchmark* TEP usando 1 simulação (500 amostras).

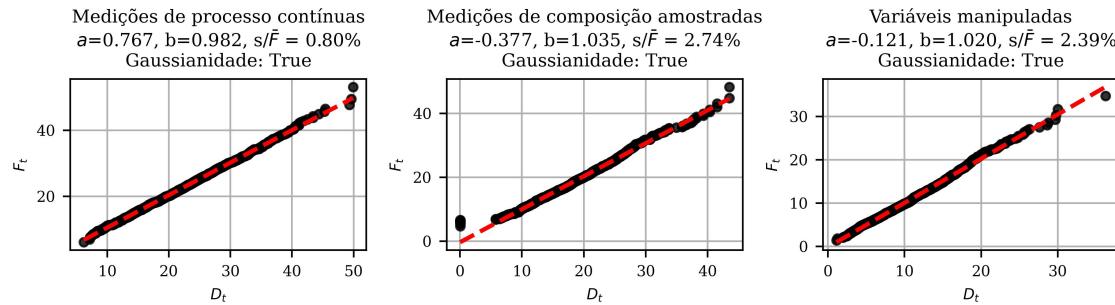


Figura 4.6: Testes de gaussianidade do *benchmark* TEP usando 3 simulações (1500 amostras).

variável 19. A Figura 4.8 mostra o tempo necessário para que a autocorrelação se reduza ao valor de 0,5. A mediana é de 3 min, exatamente o intervalo de amostragem para a maioria das variáveis.

As razões sinal/ruído também são baixas em geral, como pode ser visto na Figura 4.9. Isso indica que a variância total é comparável na maioria dos casos à variância do ruído de medição.

Os coeficientes de correlação são mostrados na Figura 4.10. Os níveis de correlações lineares e não lineares são baixos, com um nível de linearidade ligeiramente maior. Isso está de acordo com as intuições que se tem ao visualizar as séries temporais

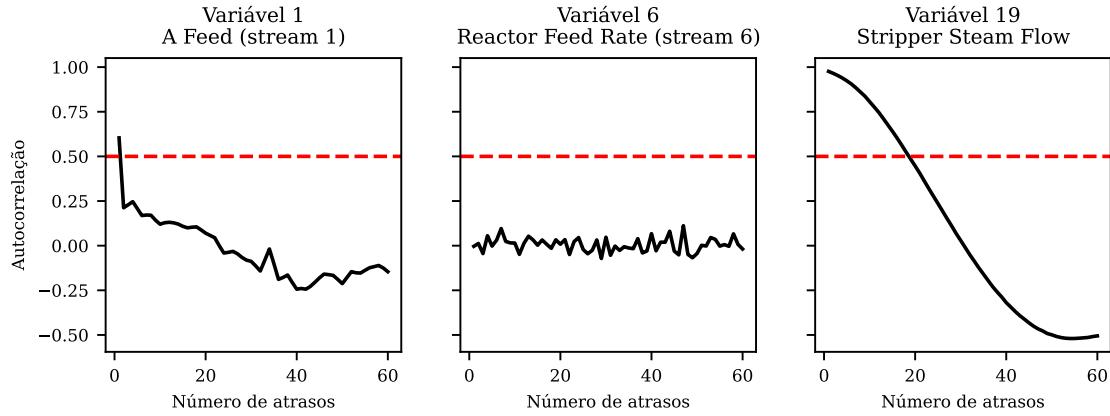


Figura 4.7: Gráficos de função de autocorrelação para algumas variáveis do *benchmark* TEP.

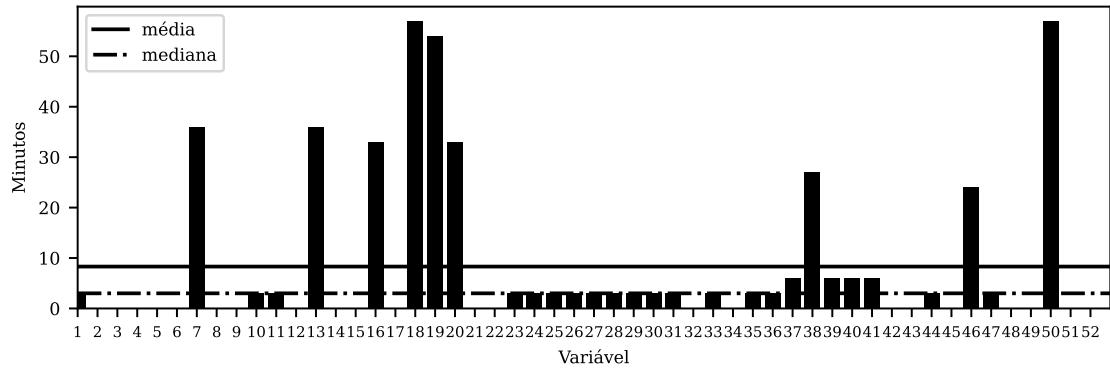


Figura 4.8: Tempo necessário para atingir 0,5 de autocorrelação no *benchmark* TEP.

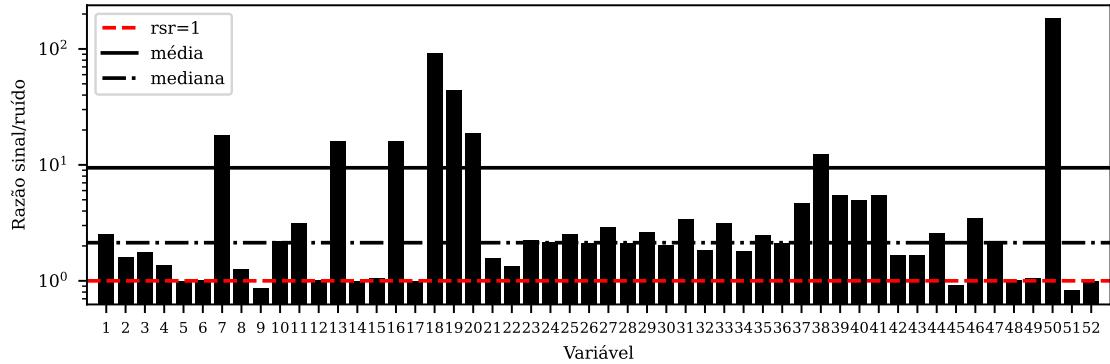


Figura 4.9: Razões sinal/ruído no *benchmark* TEP.

e os gráficos de dispersão do conjunto de dados.

As correlações tendem a ser sensíveis aos níveis de ruído, que são grandes neste conjunto de dados. Uma análise de redução de ruído foi realizada usando um filtro de média móvel. A Figura 4.11 mostra como os coeficientes gerais de correlação aumentam com o tamanho da janela do filtro.

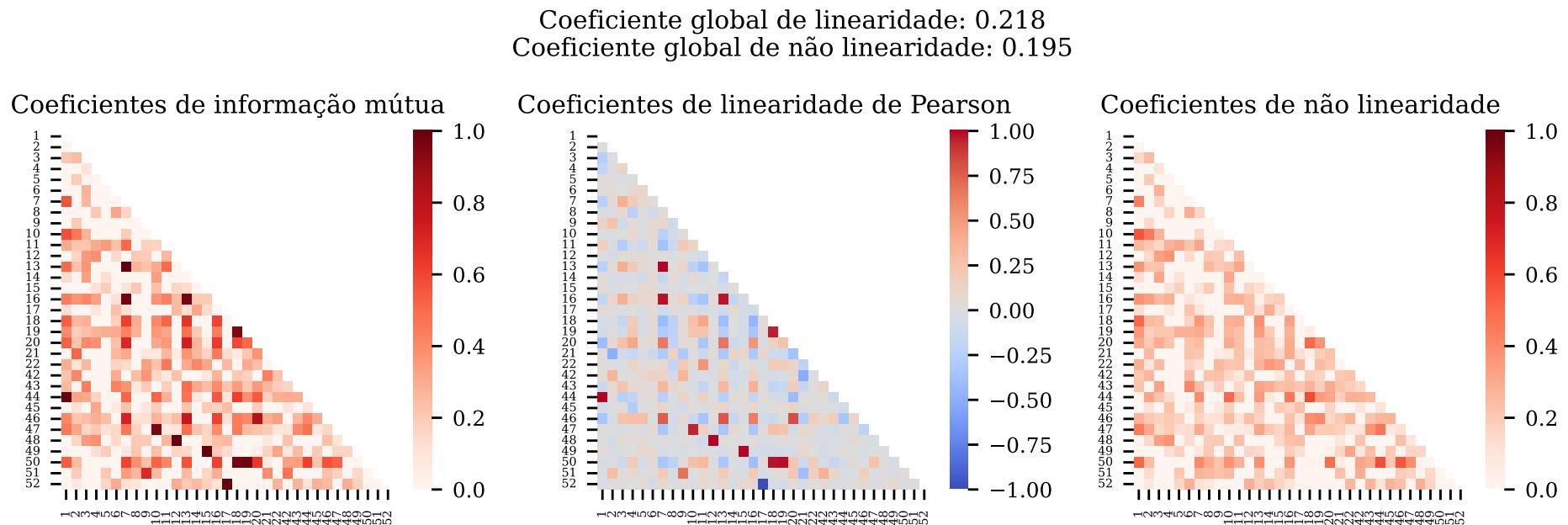


Figura 4.10: Coeficientes de correlação para 1 simulação normal do *benchmark* TEP.

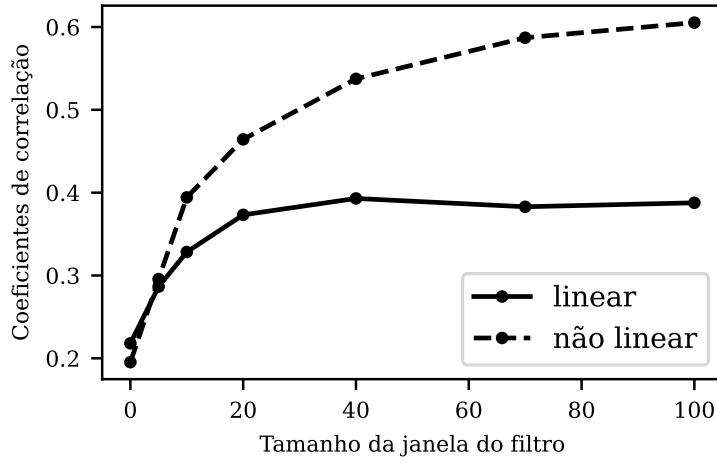


Figura 4.11: Evolução dos coeficientes de correlação do *benchmark* TEP à medida que o tamanho do filtro aumenta.

Há um aumento significativo nos coeficientes de correlação globais linear e não linear, o que de fato indica que o ruído artificial da medição oculta as relações entre as variáveis nos dados. No entanto, deve-se ter cuidado com a interpretação do gráfico da Figura 4.11, pois a informação mútua aumenta quando as dependências entre as variáveis são destacadas, mas não apenas quando isso acontece. Na Figura 4.12 a evolução da relação entre duas variáveis no conjunto de dados é mostrada, à medida que o tamanho do filtro aumenta. A partir da figura, pode-se concluir que a filtragem é realmente capaz de melhorar a trajetória dinâmica, mas apenas até certo ponto, pois nos últimos gráficos a trajetória fica distorcida. No entanto, apesar da distorção, a informação mútua continua aumentando. Assim, os platôs na Figura 4.11 não correspondem a tamanhos de janela ideais para aumento de ruído, embora em outras circunstâncias isso possa ser o caso.

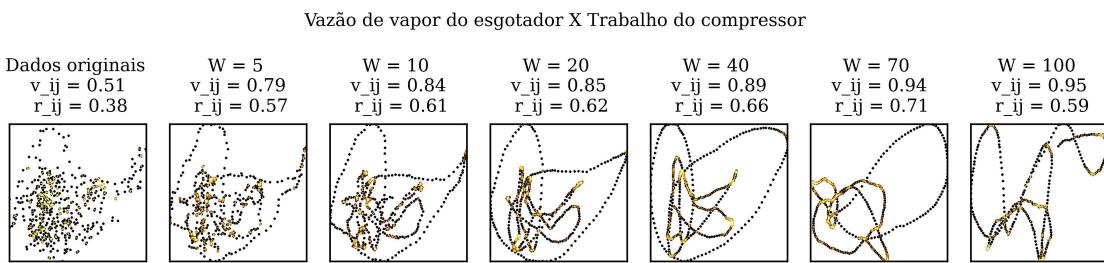


Figura 4.12: Evolução da relação entre duas variáveis no *benchmark* TEP à medida que o tamanho do filtro aumenta.

4.2.1.2 Dados de falha

Existem 5 tipos de falhas no conjunto de dados:

- degrau
- variação aleatória;
- tendência lenta;
- agarramento de válvula;
- desconhecida.

As falhas estão listadas na Tabela 4.2. As colunas “Distribuição”, “Início” e “Fim” foram construídas com inspeção visual usando gráficos como a Figura 4.13, em que há um gráfico para a Falha 5. Um gráfico de controle PCA é mostrado na Figura 4.14.

Ao contrário dos dados normais, onde há pouca variabilidade, as falhas do *benchmark* TEP são muito mais variadas. Na Tabela 4.2, as falhas são classificadas de acordo com alguns critérios:

- Tipo: a descrição fornecida no artigo original.
- Distribuição: se o efeito da falha é local (restringido a poucas variáveis) ou distribuído (espalhado ao longo do processo).
- Início: como a falha se desenvolve ao longo do tempo em seu início. Com exceção da Falha 13, todas as falhas para as quais há uma descrição podem ser classificadas como de partida abrupta.
- Fim: como a falha se comporta nos estágios finais. A maioria das falhas termina com flutuações variáveis, embora algumas se estabeleçam no mesmo estado estacionário (SS) ou em um novo. Esses são comportamentos típicos de malhas de controle ao responder a distúrbios.

Além disso, na última coluna, são apresentados os resultados da detecção de falhas com um modelo PCA simples de referência. Existem falhas fáceis de detectar, como as falhas 6 e 7, e falhas particularmente difíceis, como as falhas 3, 9 e 15. Em alguns casos, como mostrado na Figura 4.13 para a Falha 5, há detecção satisfatória no estágio inicial, mas à medida que os controladores trazem as variáveis de volta aos *set points* originais, suavizando as flutuações, as taxas de detecção caem novamente abaixo dos limites. Observar apenas os resultados numéricos, como é feito em muitos trabalhos (YIN *et al.*, 2012, 2014b; ZOU *et al.*, 2018), não esclarece essas nuances. É sempre aconselhável, portanto, analisar os resultados de detecção e diagnóstico de

Tabela 4.2: Falhas do *benchmark* TEP.

Falha	Descrição	Tipo	Distribuição	Início	Fim	PCA FDR (%)
1	A/C Feed Ratio, B Composition Constant (Stream 4) Step	Degrau	Distribuída	Abrupto	Novo SS	99,88
2	B Composition, A/C Ratio Constant (Stream 4) Step	Degrau	Distribuída	Abrupto	Novo SS	99,00
3	D Feed Temperature (Stream 2) Step	Degrau	—	Abrupto	—	2,00
4	Reactor Cooling Water Inlet Temperature Step	Degrau	Local	Abrupto	Novo SS	99,88
5	Condenser Cooling Water Inlet Temperature Step	Degrau	Distribuída	Abrupto	Mesmo SS	29,25
6	A Feed Loss (Stream 1) Step	Degrau	Distribuída	Abrupto	Shutdown	100,00
7	C Header Pressure Loss - Reduced Availability (Stream 4) Step	Degrau	Distribuída	Abrupto	Mesmo SS	100,00
8	A, B, C Feed Composition (Stream 4) Random Variation	Variação aleatória	Distribuída	Abrupto	Flutuações	98,25
9	D Feed Temperature (Stream 2) Random Variation	Variação aleatória	—	Abrupto	—	2,13
10	C Feed Temperature (Stream 4) Random Variation	Variação aleatória	Local	Abrupto	Flutuações	18,63
11	Reactor Cooling Water Inlet Temperature Random Variation	Variação aleatória	Local	Abrupto	Flutuações	72,25
12	Condenser Cooling Water Inlet Temperature Random Variation	Variação aleatória	Distribuída	Abrupto	Flutuações	98,63
13	Reaction Kinetics Slow Drift	Tendência lenta	Distribuída	Incipiente	Flutuações	93,63
14	Reactor Cooling Water Valve Sticking	Agarramento	Local	Abrupto	Flutuações	100,00
15	Condenser Cooling Water Valve Sticking	Agarramento	—	Abrupto	—	2,00
16	Desconhecido	Desconhecido	Local	Indefinido	Flutuações	23,25
17	Desconhecido	Desconhecido	Local	Indefinido	Flutuações	87,13
18	Desconhecido	Desconhecido	Distribuída	Indefinido	Shutdown	93,38
19	Desconhecido	Desconhecido	Local	Indefinido	Flutuações	16,00
20	Desconhecido	Desconhecido	Local	Indefinido	Flutuações	52,13

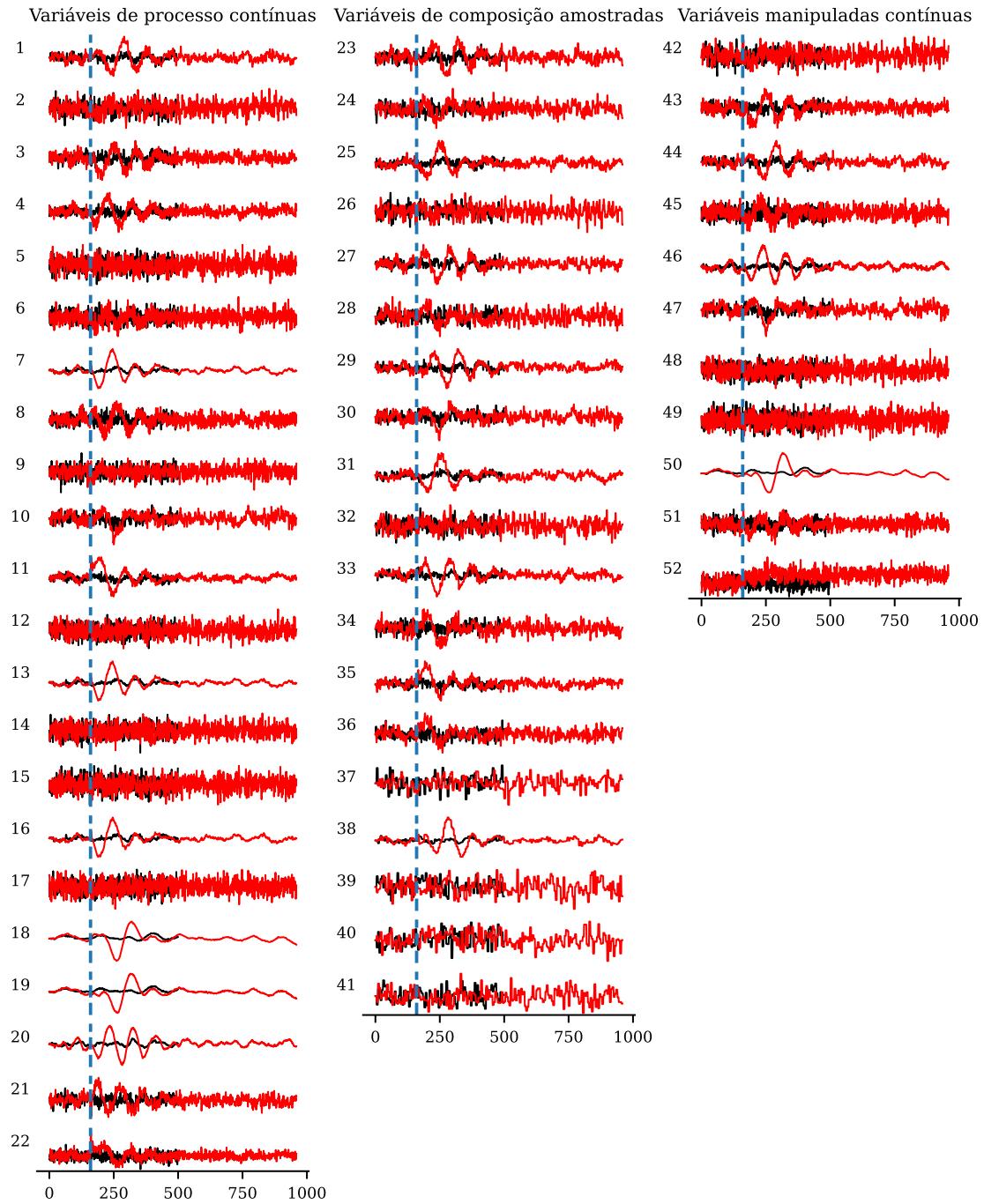


Figura 4.13: Comparação de séries temporais de conjuntos de dados normais (linhas pretas) e em falha (linhas vermelhas) do *benchmark* TEP (falha número 5). A reta tracejada azul indica o início da falha.

falhas em contexto para compreendê-los e interpretá-los adequadamente. Para isso, as técnicas de visualização costumam ser úteis.

O leitor que desejar um conjunto completo de descrições, visualizações e discussões de falhas do *benchmark* TEP pode consultar os notebooks Jupyter disponibilizados por MELO *et al.* (2022).

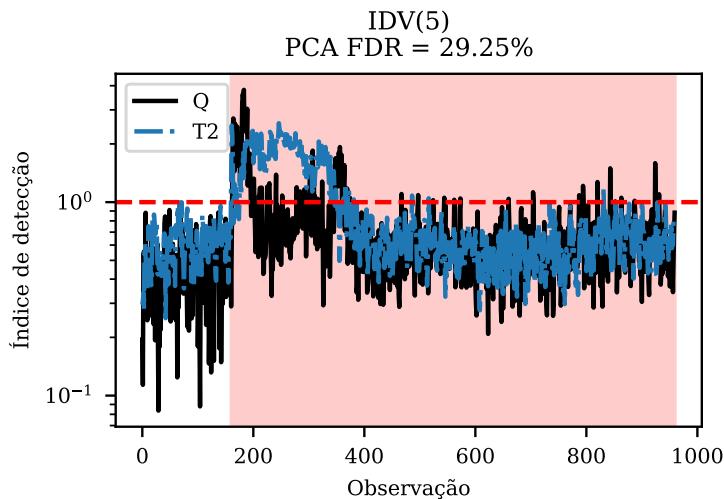


Figura 4.14: Carta de controle PCA para a falha 5 do *benchmark* TEP.

4.2.2 DAMADICS

Além do simulador descrito no Capítulo 2, o *benchmark* DAMADICS disponibiliza dados reais obtidos na Lublin Sugar Factory, Polônia, de 29 de outubro a 22 de novembro de 2001. A taxa de amostragem é de 1 Hz. A Tabela 4.3 lista as 32 medições presentes neste conjunto de dados.

Tabela 4.3: Variáveis do *benchmark* DAMADICS.

No.	Tag	Nome	Unidades
1	P51_05	P1 – juice pressure (valve inlet)	kPa
2	P51_06	P2 - juice pressure (valve outlet)	kPa
3	T51_01	T - juice temperature (valve outlet)	degC
4	F51_01	F - juice flow (1st evaporator inlet)	m ³ /h
5	LC51_03CV	CV – control value (controller output)	%
6	LC51_03X	X – servomotor rod displacement	%
7	LC51_03PV	PV - process value (juice level in 1st evaporator)	%
8	TC51_05	Juice temperature (1st evaporator inlet)	degC
9	T51_08	Juice temperature (1st evaporator outlet)	degC
10	D51_01	Juice density (1st evaporator inlet)	Bx
11	D51_02	Juice density (1st evaporator outlet)	Bx
12	F51_02	Steam flow	t/h
13	PC51_01	Steam pressure	kPa
14	T51_06	Steam temperature	degC
15	p51_03	vapor pressure	kPa

Continua na próxima página

Tabela 4.3 – *Continuada da página anterior*

No.	Tag	Nome	Unidades
16	t51_07	vapor temperature	degC
17	P57_03	P1 – juice pressure (valve inlet)	kPa
18	P57_04	P2 - juice pressure (valve outlet)	kPa
19	T57_03	T - juice temperature (valve inlet)	degC
20	FC57_03PV	PV - process value (juice flow, 5th evaporator outlet)	m ³ /h
21	FC57_03CV	CV – control value (controller output)	%
22	FC57_03X	X – servomotor rod displacement	%
23	P74_00	P1 – water pressure (valve inlet)	kPa
24	P74_01	P2 – water pressure (valve outlet)	kPa
25	T74_00	T - water temperature (valve outlet)	degC
26	F74_00	F – water flow (steam boiler inlet)	t/h
27	LC74_20CV	CV – control value (controller output)	%
28	LC74_20X	X – servomotor rod displacement	%
29	LC74_20PV	PV - process value (water level in steam boiler)	%
30	F74_30	Steam flow (steam boiler outlet)	t/h
31	P74_30	Steam pressure (steam boiler outlet)	kPa
32	T74_30	Steam temperature (steam boiler outlet)	degC

4.2.2.1 Dados normais

A Figura 4.15 mostra a série temporal para todos os dados normais plotados com um intervalo de amostragem de 5 minutos.

Neste conjunto de dados, há períodos de comportamento anormal não identificados; por exemplo, no intervalo entre os dias 31/10 a 03/11 na Figura 4.15, quando muitas variáveis congelam ou oscilam anormalmente. A visualização de séries temporais permite uma rápida identificação deste problema, típico de dados de processo reais.

A Figura 4.16 mostra os gráficos de dispersão para todas as variáveis do conjunto de dados. Há uma gama de comportamentos muito mais ampla do que no caso do *benchmark* TEP, o que certamente é reflexo de este ser um conjunto de dados reais e não uma simulação.

A Figura 4.17 mostra os testes de gaussianidade para quatro períodos de duas horas de dias diferentes. O comportamento dos pontos é claramente linear, embora os testes não sejam inequívocos, resultando verdadeiro para apenas dois deles, de

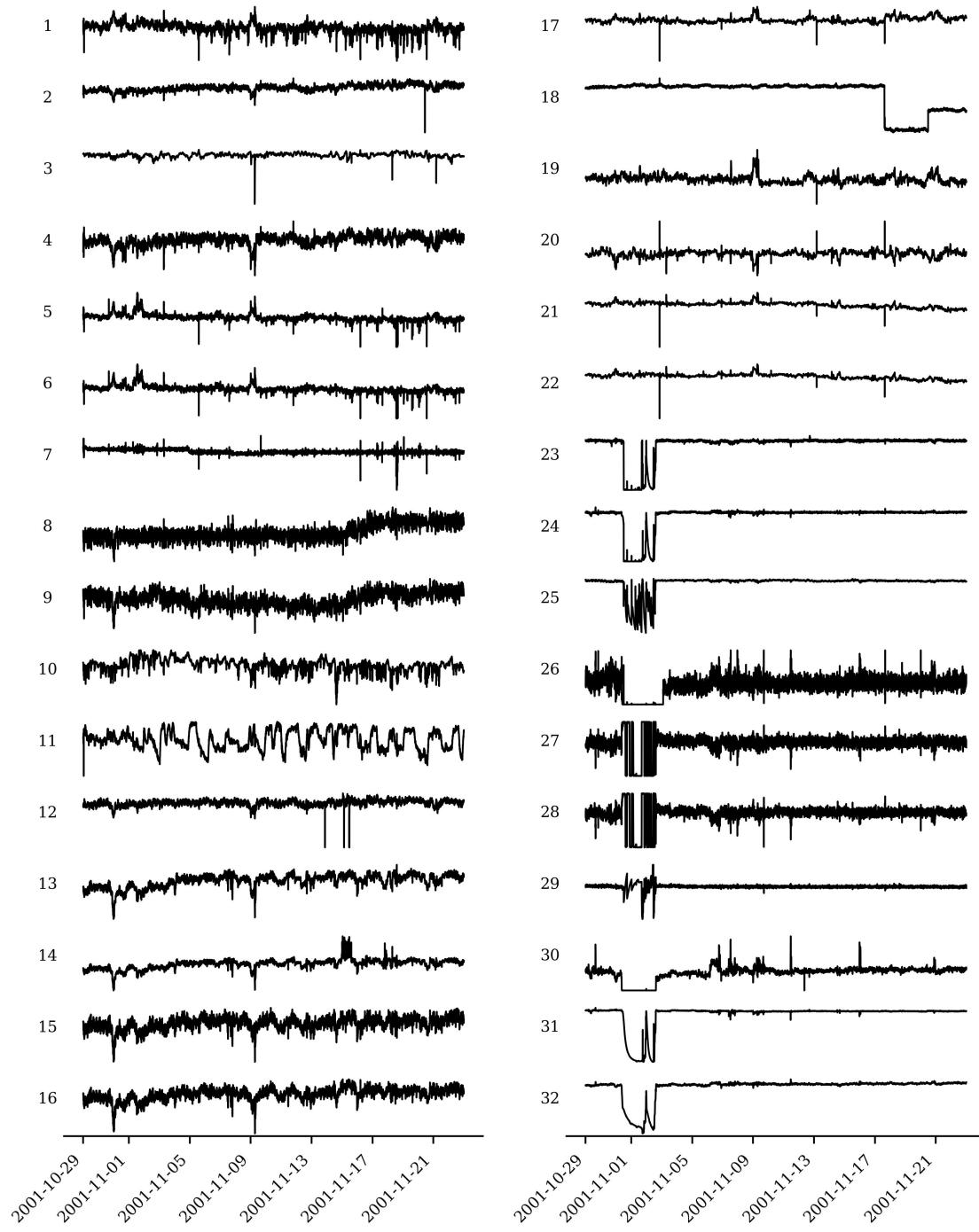


Figura 4.15: Séries temporais dos dados normais do *benchmark* DAMADICS.

acordo com as Equações 4.7 e 4.8. Na Figura 4.18, os mesmos dados são estendidos por períodos de seis horas, o que faz os dados se desviarem claramente da distribuição gaussiana. *Outliers*, comportamento anormal e *concept drift* podem estar entre as causas de as distribuições de dados por longos períodos desviarem-se da distribuição gaussiana.

Como o conjunto de dados é grande e possui ampla variedade de comportamen-



Figura 4.16: Gráficos de dispersão de todos os dados do *benchmark* DAMADICS.

tos, os padrões de autocorrelação variam consideravelmente. Em particular, sabe-se que a função de autocorrelação é sensível a *outliers* (DURRE *et al.*, 2014). Dois períodos de oito horas (00:00 às 08:00 dos dias 08/11 e 12/11) aparentemente livres de *outliers* foram encontrados e analisados. A duração de oito horas foi escolhida por representar um compromisso entre um intervalo não muito curto, que não levaria a análises representativas, e um intervalo não muito longo, que resultaria em uma grande probabilidade de ocorrência de observações não representativas. Os resultados encontram-se nas Figuras 4.19 e 4.20. Os padrões de autocorrelação são

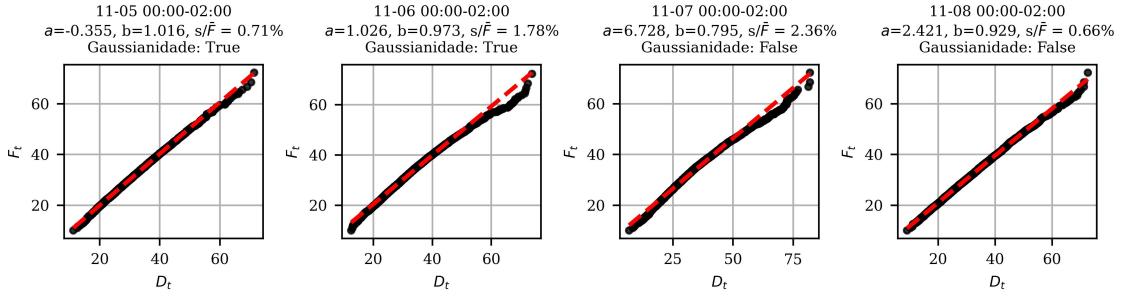


Figura 4.17: Testes de gaussianidade do *benchmark* DAMADICS para períodos de duas horas.

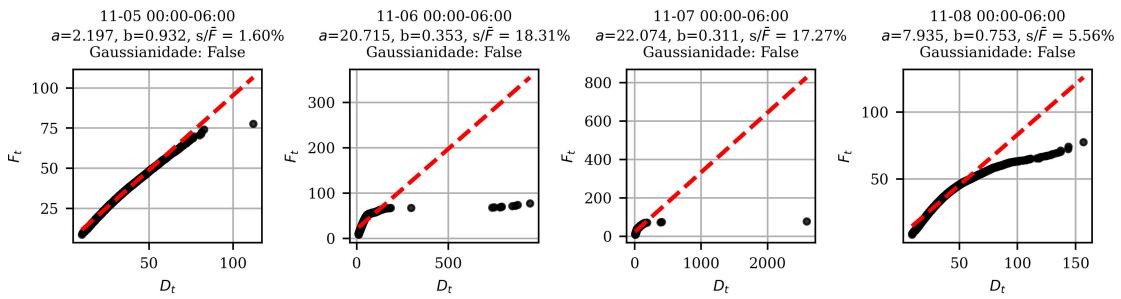


Figura 4.18: Testes de gaussianidade do *benchmark* DAMADICS para períodos de seis horas.

semelhantes, embora não exatamente iguais, evidenciando uma variação considerável nos comportamentos das variáveis.

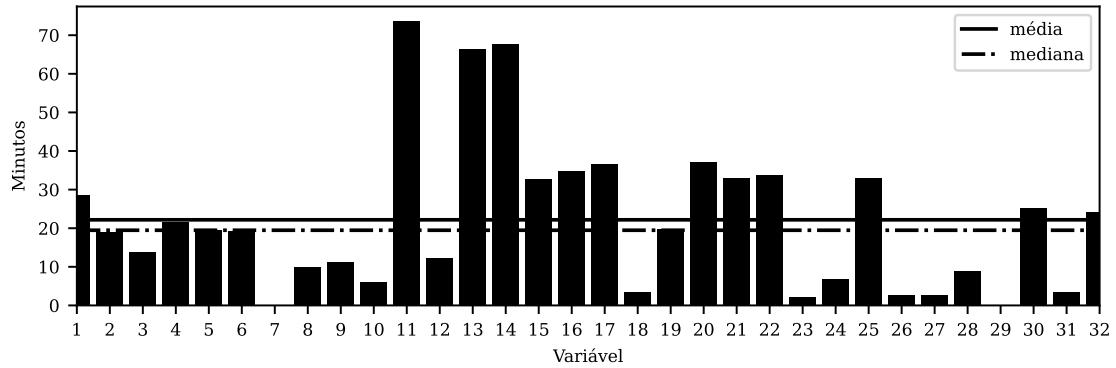


Figura 4.19: Tempo necessário para atingir 0,5 de autocorrelação no *benchmark* DAMADICS.

As razões sinal/ruído são apresentadas na Figura 4.21. A maioria das variáveis tem uma alta razão sinal-ruído. A taxa de amostragem original do conjunto de dados é de 1 Hz. Como esta é uma taxa de amostragem muito alta, é interessante ver como a diminuição da taxa faz a razão sinal/ruído diminuir, como representado na Figura 4.22. Este resultado, combinado com os gráficos de autocorrelação, mostra o quão significativas são as dinâmicas de curto prazo nesse conjunto de dados.

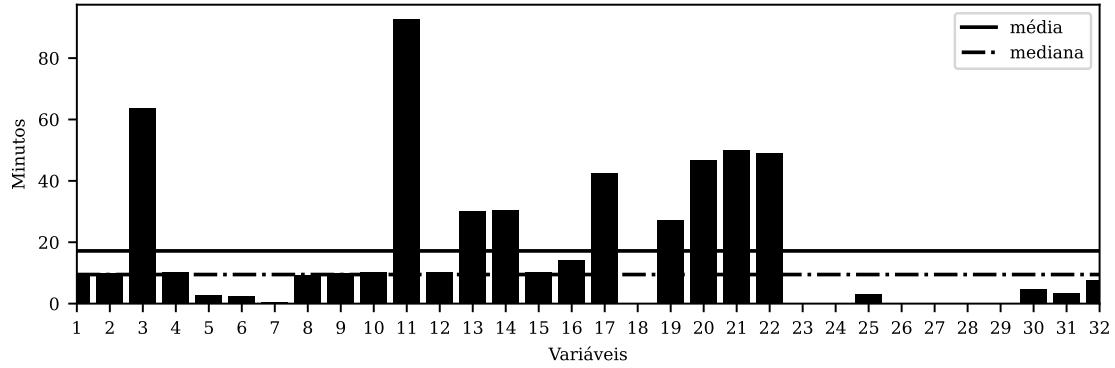


Figura 4.20: Tempo necessário para atingir 0,5 de autocorrelação no *benchmark* DAMADICS.

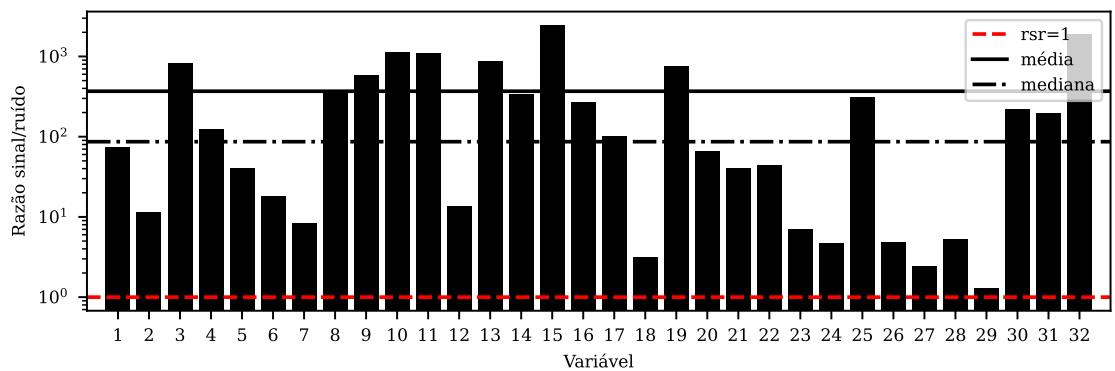


Figura 4.21: Razões sinal/ruído no *benchmark* DAMADICS.

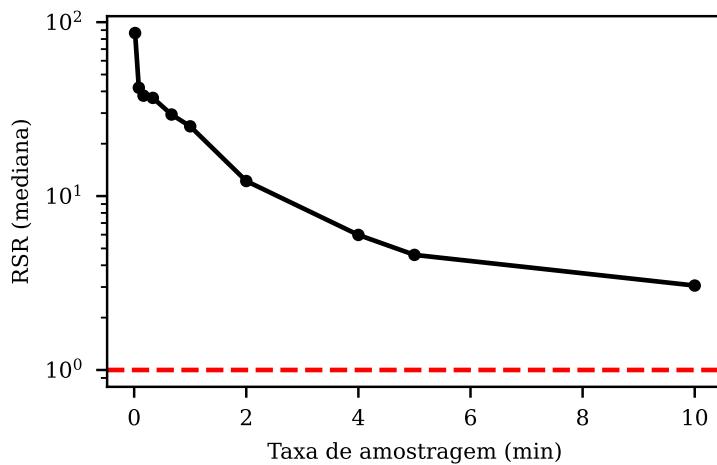


Figura 4.22: Razões sinal/ruído contra taxa de amostragem no *benchmark* DAMADICS.

A Figura 4.23 apresenta os coeficientes de correlação para um período de quatro horas (00:00 às 04:00 de 04/11), duração que se mostrou eficaz em representar as relações entre as variáveis após alguns testes. O grau de correlação é maior do que no *benchmark* TEP, tanto para correlações lineares quanto não lineares.

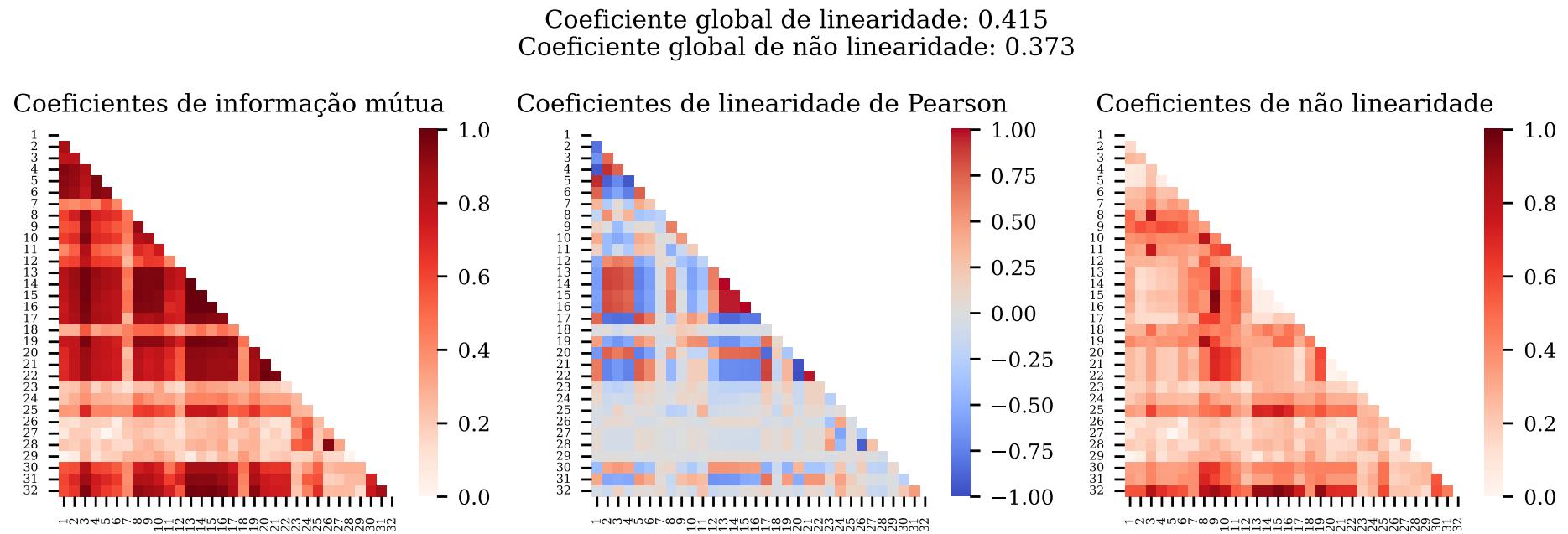


Figura 4.23: Coeficientes de correlação para um período de quatro horas do *benchmark* DAMADICS.

É interessante observar como os coeficientes variam ao longo do tempo no conjunto de dados. As Figuras 4.24 e 4.25 mostram valores diários de coeficientes de correlação não linear e linear. Os dados foram reduzidos por um reamostragem por um fator de 10 para atenuar o esforço computacional. Em certos dias, não foi possível calcular coeficientes não lineares devido a problemas desconhecidos, aparentemente relacionados à qualidade dos dados. Os coeficientes têm magnitudes semelhantes nos dois casos, oscilando entre 0,3 e 0,5.

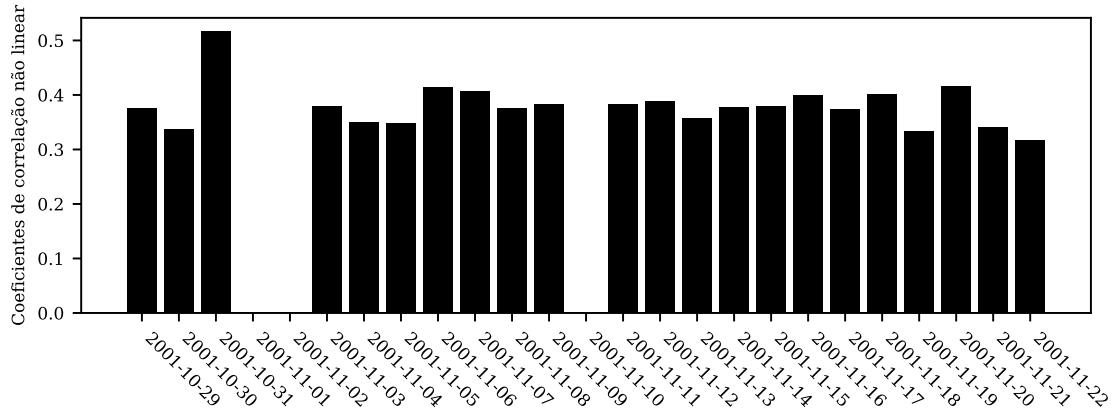


Figura 4.24: Coeficientes de correlação não linear diários para o *benchmark* DAMADICS.

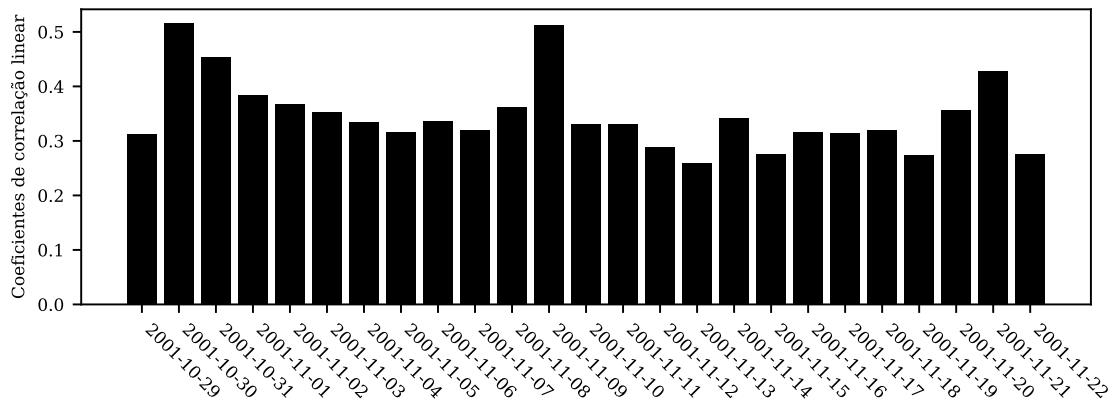


Figura 4.25: Coeficientes de correlação linear diários para o *benchmark* DAMADICS.

A Figura 4.26 mostra como os coeficientes de correlação aumentam com o tamanho de janela de filtro de média móvel. Nos três casos há aumentos consideráveis no coeficiente não linear e aumentos de menor magnitude no coeficiente linear.

Na Figura 4.27, a evolução da relação entre duas variáveis no conjunto de dados é mostrada, à medida que o tamanho do filtro aumenta. O filtro é capaz de melhorar uma trajetória dinâmica e, diferentemente do resultado no *benchmark* TEP, não há distorção para grandes tamanhos de janela, possivelmente como resultado da alta taxa de amostragem.

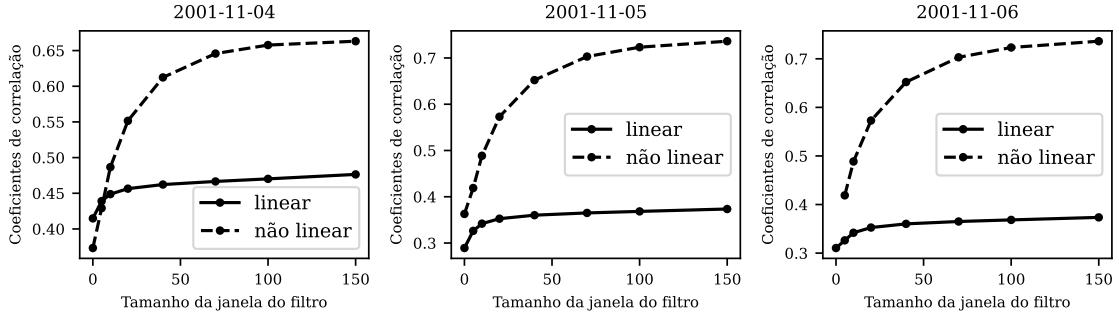


Figura 4.26: Efeito do filtro de média móvel nos coeficientes de correlação no *benchmark* DAMADICS.

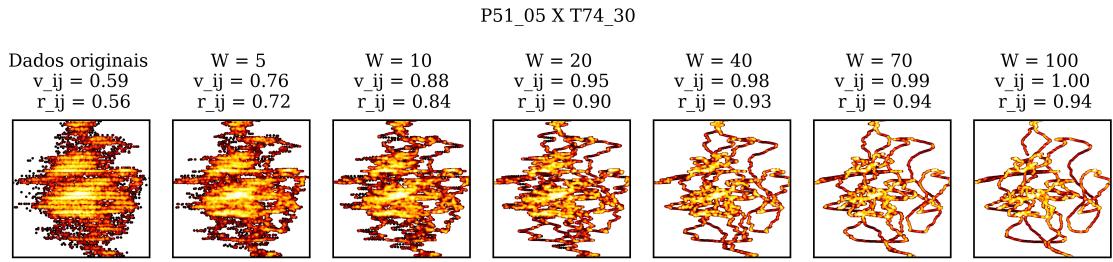


Figura 4.27: Efeito do filtro de média móvel em um gráfico de dispersão do *benchmark* DAMADICS.

4.2.2.2 Dados de falha

Todo o *benchmark* DAMADICS (que diz respeito ao simulador e conjunto de dados) contém 19 tipos de falhas do atuador (divididas em falhas na válvula de controle, falhas no servomotor pneumático, falhas no posicionador e falhas gerais/externas). No conjunto de dados, apenas quatro tipos de falhas gerais/externas foram introduzidas artificialmente nos dias 30/10, 9/11, 17/11, 20/11.

A Tabela 4.4 mostra os resultados do monitoramento de PCA por dia. Diferentemente do caso no *benchmark* TEP, escolher um período de treinamento não constitui uma tarefa trivial devido à alta variabilidade dos dados ao longo do tempo e à duração relativamente curta dos períodos de falha (da ordem de minutos) quando comparados ao conjunto de dados completo (da ordem de dias). Os resultados desta subseção foram obtidos por tentativa e erro de modo a chegar a bons conjuntos de treinamento em períodos anteriores às falhas. As Figuras de 4.28 até 4.31 mostram exemplos de resultados de monitoramento por dois dias em que alguns desafios comuns estão presentes. Na Figura 4.28, o gráfico de validação para o dia 09/11 mostra que o modelo se degrada devido ao *concept drift*, exigindo que o limite de detecção seja aumentado. Esse foi um problema comum encontrado ao treinar modelos ao longo de todo o conjunto de dados. A Figura 4.29 mostra o gráfico de teste para

o dia 09/11, onde fica claro que o modelo PCA é sensível a falhas; no entanto, ao final, um alarme falso é acionado, indicando possível comportamento anormal não identificado, uma ocorrência muito comum no conjunto de dados. Na Figura 4.30, para o dia 17/11, o gráfico de validação não mostra um *concept drift* pronunciado como o anterior, mas na Figura 4.31 surge outro problema: quando ocorre a primeira falha, o desempenho do modelo se degrada completamente e o índice de detecção não retorna para baixo do limite. Esses problemas indicam que o conjunto de dados DAMADICS, diferentemente da maioria dos outros *benchmarks*, apresenta desafios consideráveis relacionados à seleção de períodos de treinamento.

Tabela 4.4: Resultados diários de detecção utilizando o modelo PCA para o *benchmark* DAMADICS.

	PCA FDR (%)	PCA FAR (%)
10–30	94.19	12.57
11–09	83.71	5.50
11–17	100.00	71.60
11–20	100.00	68.44

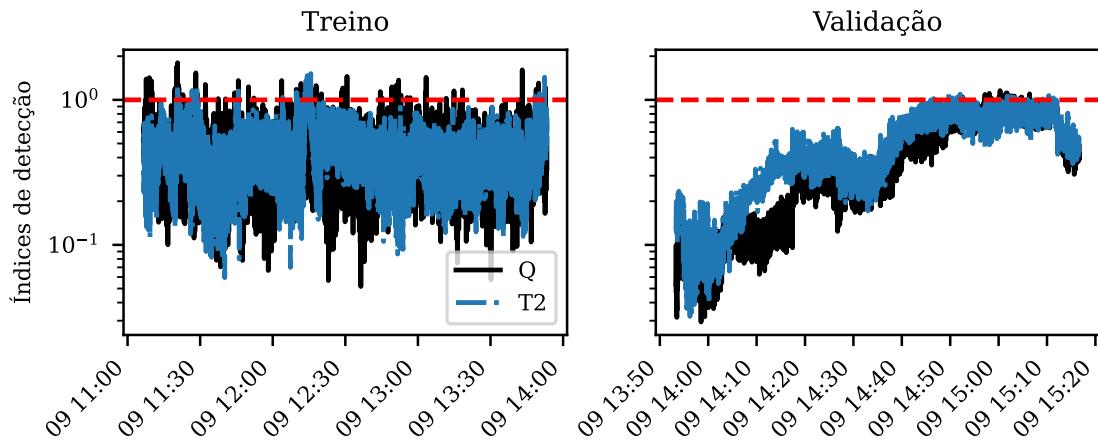


Figura 4.28: Cartas de controle PCA de treinamento e validação para o dia 09/11 do DAMADICS.

O leitor que desejar um conjunto completo de visualizações e discussões de falhas do *benchmark* DAMADICS pode consultar os notebooks Jupyter disponibilizados por MELO *et al.* (2022).

4.2.3 IndPenSim

O *benchmark* IndPenSim fornece dados obtidos por meio do simulador descrito no Capítulo 2. A Tabela 4.5 mostra as 38 variáveis disponíveis no conjunto de dados.

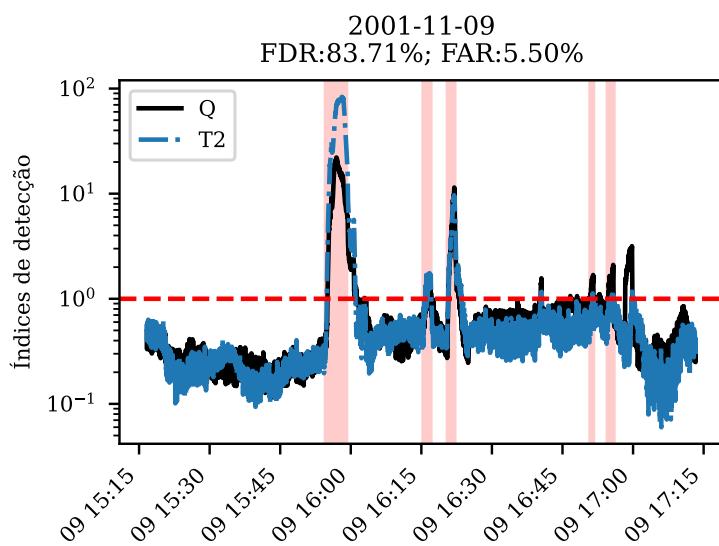


Figura 4.29: Cartas de controle PCA de teste para o dia 09/11 do DAMADICS.

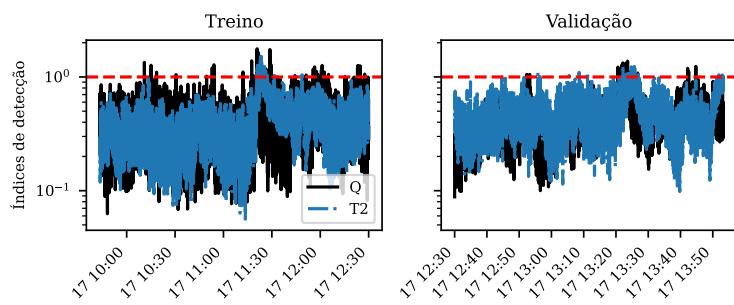


Figura 4.30: Cartas de controle PCA de treinamento e validação para o dia 17/11 do DAMADICS.

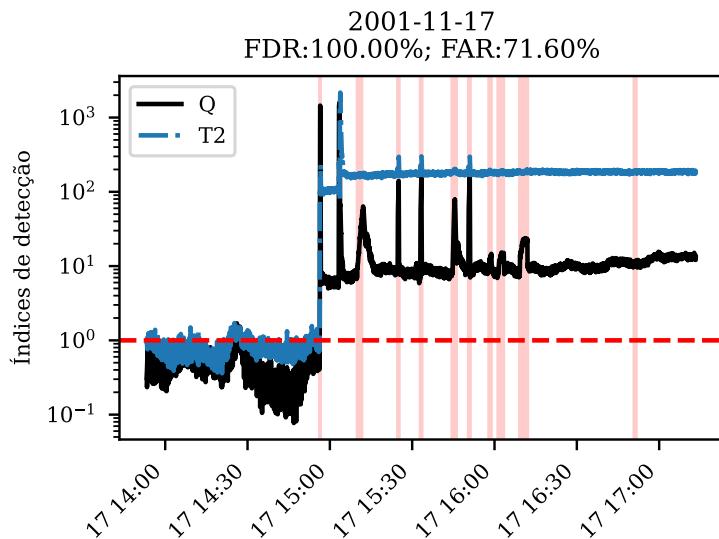


Figura 4.31: Cartas de controle PCA de teste para o dia 17/11 do DAMADICS.

Tabela 4.5: Variáveis do *benchmark* IndPenSim.

No.	Nome	Tag	Unidades
1	Aeration rate	Fg	L/h
2	Agitator RPM	RPM	–
3	Sugar feed rate	Fs	L/h
4	Acid flow rate	Fa	L/h
5	Base flow rate	Fb	L/h
6	Heating/cooling water flow rate	Fc	L/h
7	Heating water flow rate	Fh	L/h
8	Water for injection/dilution	Fw	L/h
9	Air head pressure	pressure	bar
10	Dumped broth flow	Fremoved	L/h
11	Substrate concentration	S	g/L
12	Dissolved oxygen concentration	DO2	mg/L
13	Penicillin concentration	P	g/L
14	Vessel Volume	V	L
15	Vessel Weight	Wt	kg
16	pH	pH	–
17	Temperature	T	K
18	Generated heat	Q	kJ
19	Carbon dioxide percent in off-gas	CO2outgas	%
20	PAA flow	Fpaa	L/h
21	PAA concentration offline	PAA_offline	g/L
22	Oil flow	Foil	L/hr
23	NH ₃ concentration off-line	NH3_offline	g/L
24	Oxygen Uptake Rate	OUR	g/min
25	Oxygen in percent in off-gas	O2	%
26	Offline Penicillin concentration	P_offline	g/L
27	Offline Biomass concentration	X_offline	g/L
28	Carbon evolution rate	CER	g/h
29	Ammonia shots	NH3_shots	kg
30	Viscosity	Viscosity_offline	centPoise
31	Fault reference	Fault_ref	–
32	0 - Recipe driven 1 - Operator controlled	Control_ref	–
33	1- No Raman spec	–	–
34	1-Raman spec recorded	–	–
35	Batch reference	Batch_ref	–

Continua na próxima página

Tabela 4.5 – *Continuada da página anterior*

No.	Nome	Tag	Unidades
36	2-PAT control	PAT_ref	–
37	Batch ID	–	–
38	Fault flag	–	–

As variáveis podem ser classificadas de acordo com sua natureza:

- Medições de processo contínuas: 1 a 10, 12, 14, 15 a 20, 22, 24, 25 e 28;
- Cálculos de sensores virtuais: 11 e 13;
- Medições offline: 21, 23, 26 e 27;
- Identificadores e especificadores de condição: 31 a 38.

Há também outras 2200 variáveis correspondentes aos dados de espectroscopia, disponíveis para uso em casos especiais. Essas variáveis não foram analisadas neste trabalho.

O processo ocorre em batelada. Existem 100 lotes (rodadas de batelada) disponíveis no conjunto de dados:

- Lotes 1–30: Controlados por uma abordagem orientada por receita (*recipe driven approach*);
- Lotes 31–60: Controlados pelos operadores;
- Lotes 61–90: Controlado por uma solução de controle avançado de processo utilizando espectroscopia Raman;
- Lotes 91–100: Contêm falhas que resultam em desvios do processo.

4.2.3.1 Dados normais

A Figura 4.32 mostra a série temporal para os dados dos lotes 1 a 30. Uma escala de cores cinza foi usada para diferenciar os lotes. Muitas das variáveis assumem uma faixa de valores discretos. Parece que há um problema com as variáveis 34, 37 e 38: a variável 34 (*1-Raman spec recorded*) é uma cópia da variável 35 (*Batch reference*); as variáveis 37 e 38 são iguais e seguem de perto a forma da variável 13 (concentração de penicilina). Esses comportamentos, que não são compatíveis com

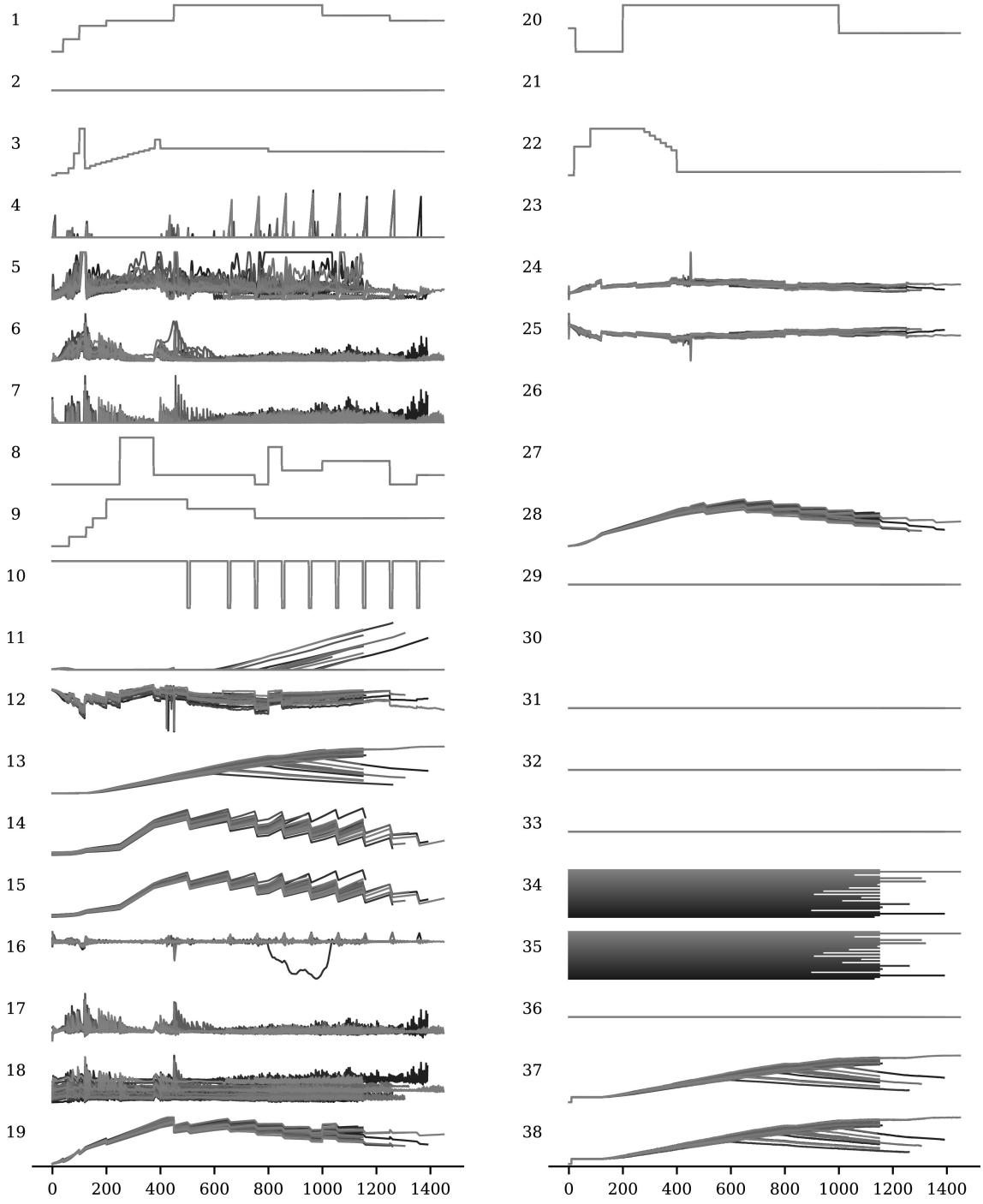


Figura 4.32: Séries temporais dos dados normais do *benchmark* IndPenSim.

os tipos indicados nos nomes das variáveis (ID e *flag*), também estão presentes nos próximos dois tipos de lotes.

A Figura 4.33 mostra os gráficos de dispersão para as medições contínuas do processo. A natureza discreta de muitas das variáveis também fica evidente. Os testes de gaussianidade apresentados na Figura 4.36 resultaram falsos para todos os três conjuntos de lotes, com grandes desvios em relação ao comportamento gaussiano. Os resultados da autocorrelação e da razão sinal/ruído foram calculados

para os três conjuntos de lotes, com os resultados mostrados nas Figuras 4.34 e 4.35, respectivamente. A autocorrelação é significativa para a maioria das variáveis, com valores semelhantes para cada variável nos três conjuntos de lotes. As razões sinal/ruído apresentam a mesma consistência. As correlações lineares e não lineares são geralmente altas, como mostrado nas Figuras 4.37 e 4.38. A Figura 4.37 apresenta matrizes de correlação para o Lote 1; a Figura 4.38 apresenta correlações globais lineares e não lineares para todos os lotes, onde a consistência entre os lotes também é observada.

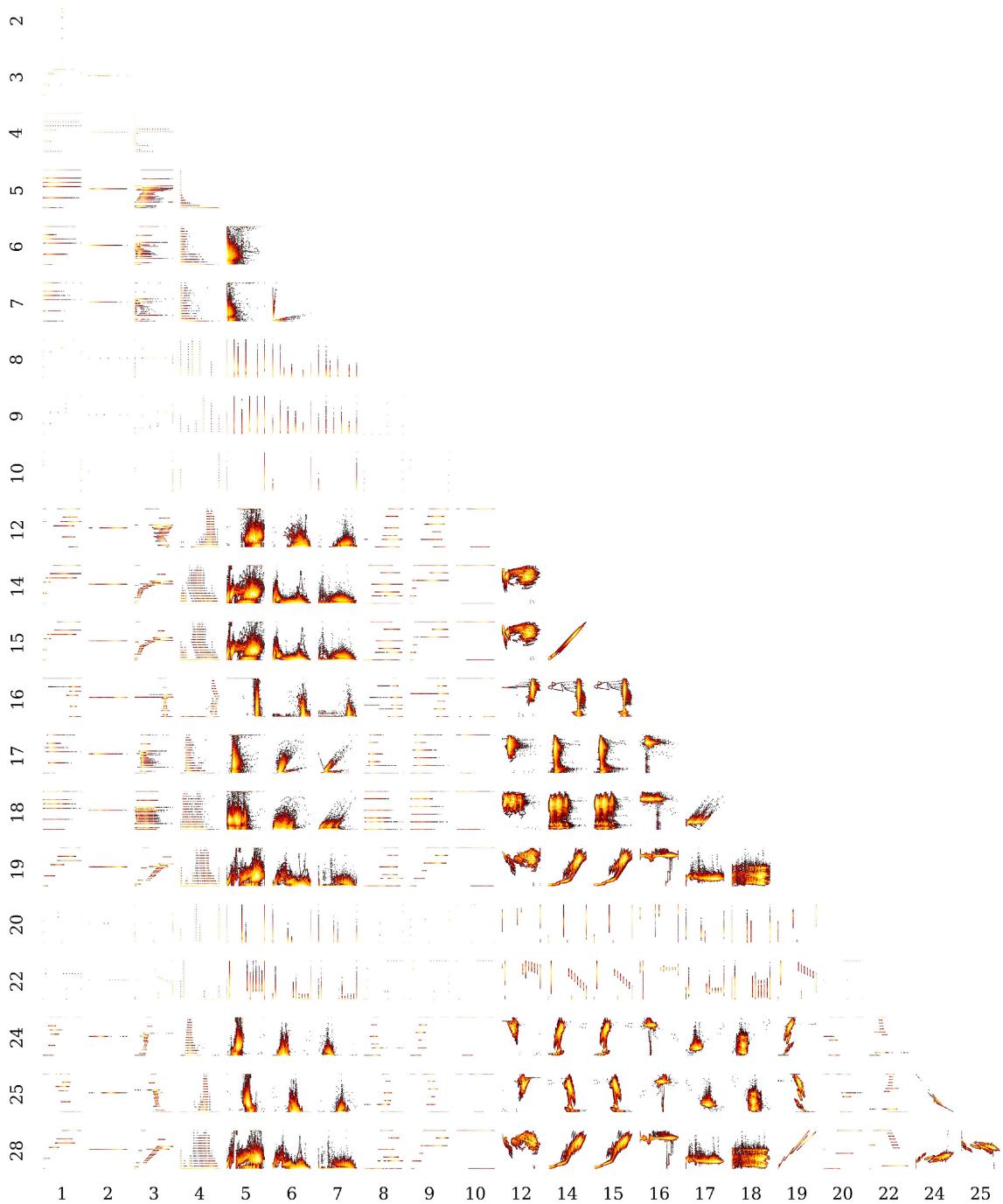


Figura 4.33: Gráficos de dispersão dos lotes 1–30 do *benchmark* IndPenSim.

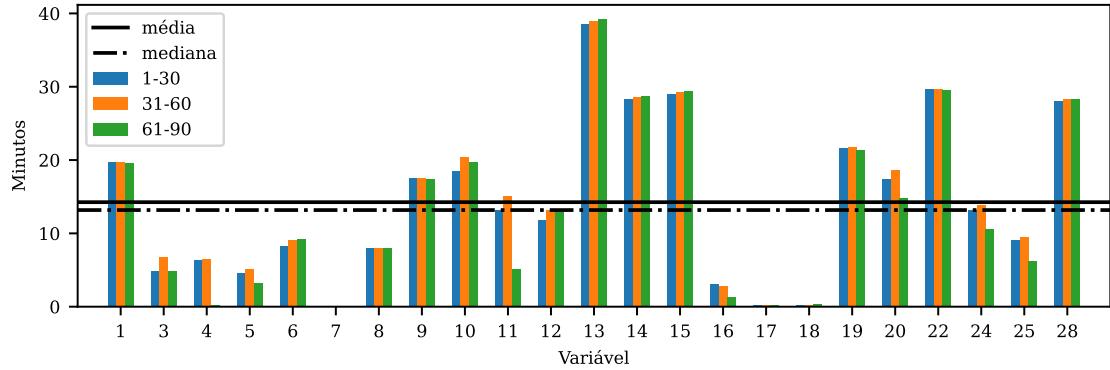


Figura 4.34: Tempo necessário para se atingir 0,5 de autocorrelação no *benchmark* IndPenSim.

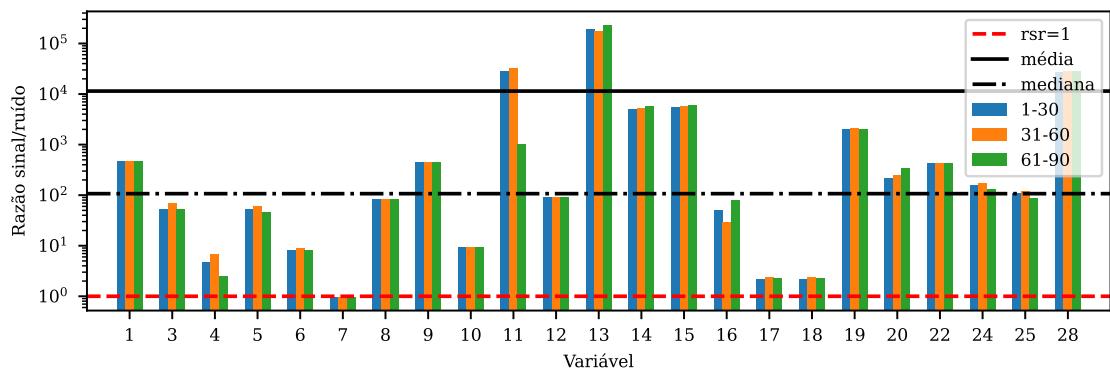


Figura 4.35: Razões sinal/ruído no *benchmark* IndPenSim.

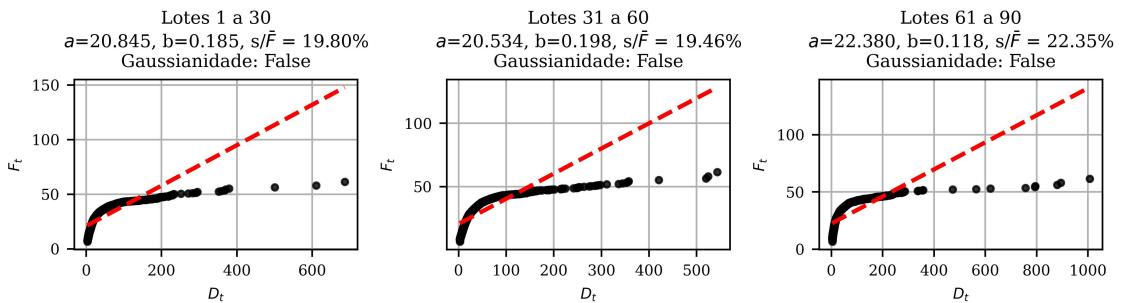


Figura 4.36: Testes de gaussianidade do *benchmark* IndPenSim.

4.2.3.2 Dados de falha

Infelizmente, não foi possível encontrar identificações de falhas neste conjunto de dados. Talvez essa informação devesse estar na variável 38, dado o seu nome (*fault flag*), mas não é o caso. O artigo original (GOLDRICK *et al.*, 2019) lista um conjunto de falhas, mas os dados utilizados não são exatamente os mesmos que o conjunto de dados disponibilizado no *benchmark*.

O treinamento e a validação foram realizados usando dados dos lotes 1–45 e

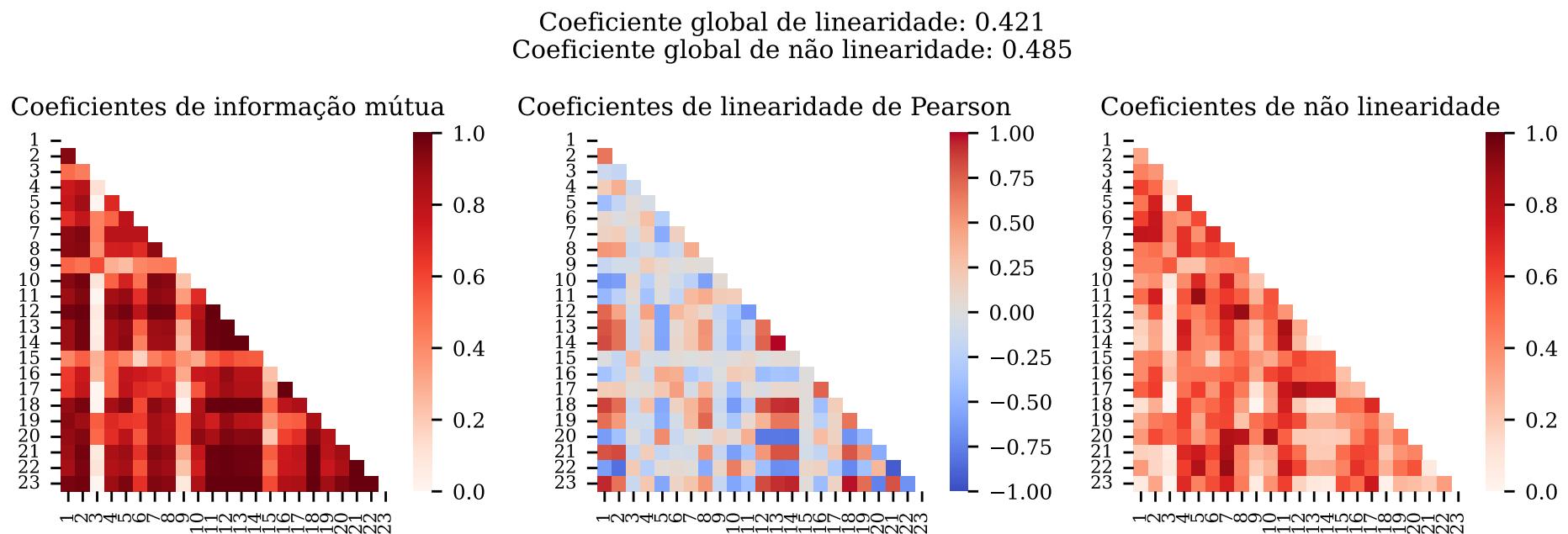


Figura 4.37: Coeficientes de correlação para 1 lote do *benchmark* IndPenSim.



Figura 4.38: Coeficientes de correlação global para todos os lotes do *benchmark* IndPenSim.

46–90, respectivamente. Os gráficos de treinamento e validação são apresentados na Figura 4.39. Os resultados para as dez falhas são apresentados na Figura 4.40. As taxas de detecção de PCA geralmente mostram boa sensibilidade a eventos anormais. No entanto, em alguns casos, apesar do aumento das taxas de detecção, elas não ultrapassam o limite de detecção. Há também muitos alarmes falsos em alguns lotes. É provável que a engenharia de características e uma seleção adequada de dados de

treinamento possam aliviar esses problemas.

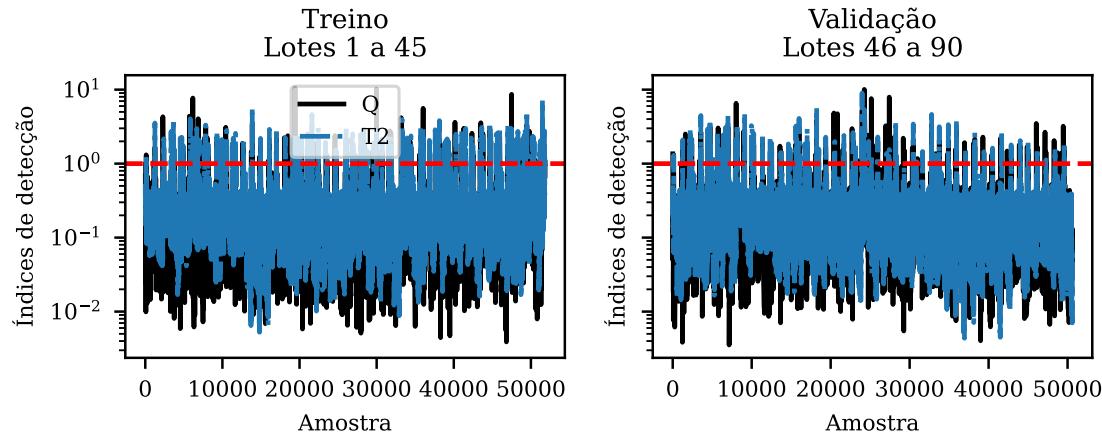


Figura 4.39: Cartas de controle PCA de treino e validação do *benchmark* IndPenSim.

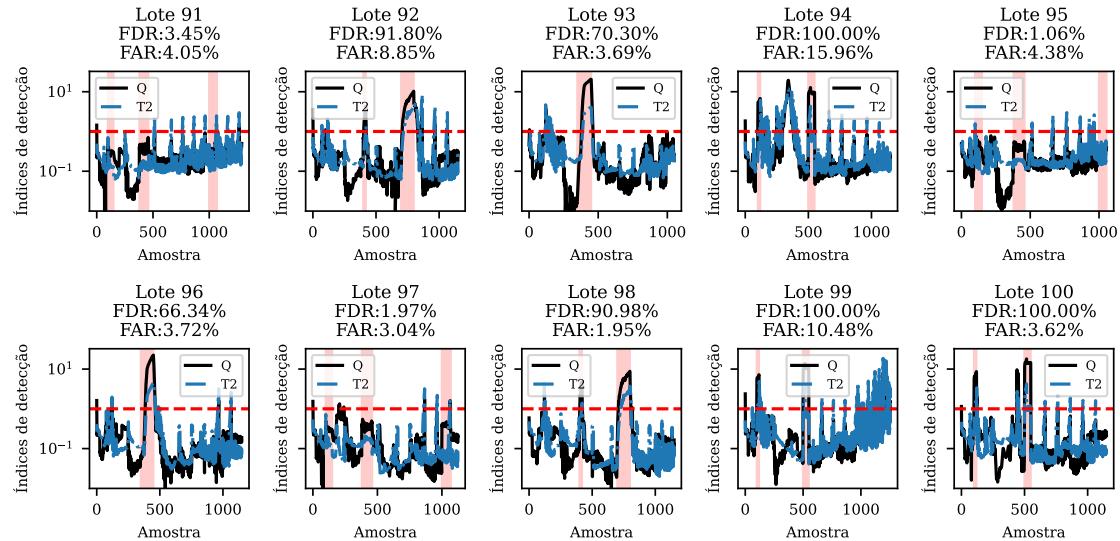


Figura 4.40: Cartas de controle PCA de teste do *benchmark* IndPenSim.

4.2.4 Cranfield Multiphase Flow Facility

A Tabela 4.6 lista as 24 medições presentes neste conjunto de dados.

Tabela 4.6: Variáveis do *benchmark* Cranfield.

No.	Local	Nome	Unidades
1	PT312	Air delivery pressure	MPa
2	PT401	Pressure in the bottom of the riser	MPa

Continua na próxima página

Tabela 4.6 – *Continuada da página anterior*

No.	Local	Nome	Unidades
3	PT408	Pressure in top of the riser	MPa
4	PT403	Pressure in top separator	MPa
5	PT501	Pressure in 3 phase separator	MPa
6	PT408	Diff. pressure (PT401-PT408)	MPa
7	PT403	Differential pressure over VC404	MPa
8	FT305	Flow rate input air	Sm ³ /s
9	FT104	Flow rate input water	kg/s
10	FT407	Flow rate top riser	kg/s
11	LI405	Level top separator	m
12	FT406	Flow rate top separator output	kg/s
13	FT407	Density top riser	kg/m ³
14	FT406	Density top separator output	kg/m ³
15	FT104	Density water input	kg/m ³
16	FT407	Temperature top riser	degC
17	FT406	Temperature top separator output	degC
18	FT104	Temperature water input	degC
19	LI504	Level gas-liquid 3 phase separator	%
20	VC501	Position of valve VC501	%
21	VC302	Position of valve VC302	%
22	VC101	Position of valve VC101	%
23	PO1	Water pump current	A
24	PT417	Pressure in mixture zone 2" line	MPa

4.2.4.1 Dados normais

Existem três conjuntos de dados normais a serem usados como dados de treinamento, chamados T1, T2 e T3. Esses conjuntos de dados são representados na Figura 4.41, na qual as mudanças de *set points* são perceptíveis. Na Figura 4.42, os gráficos de dispersão para todas as 24 variáveis no primeiro conjunto de dados de treinamento (T1) são apresentados. Observa-se uma ampla gama de comportamentos nas relações entre variáveis, como trajetórias dinâmicas e transições entre modos de operação. A Figura 4.43 apresenta alguns desses gráficos com mais detalhes.

Os testes de gaussianidade resultaram falsos para todos os três conjuntos de dados, com grandes desvios em relação ao comportamento gaussiano, conforme mostrado na Figura 4.44.

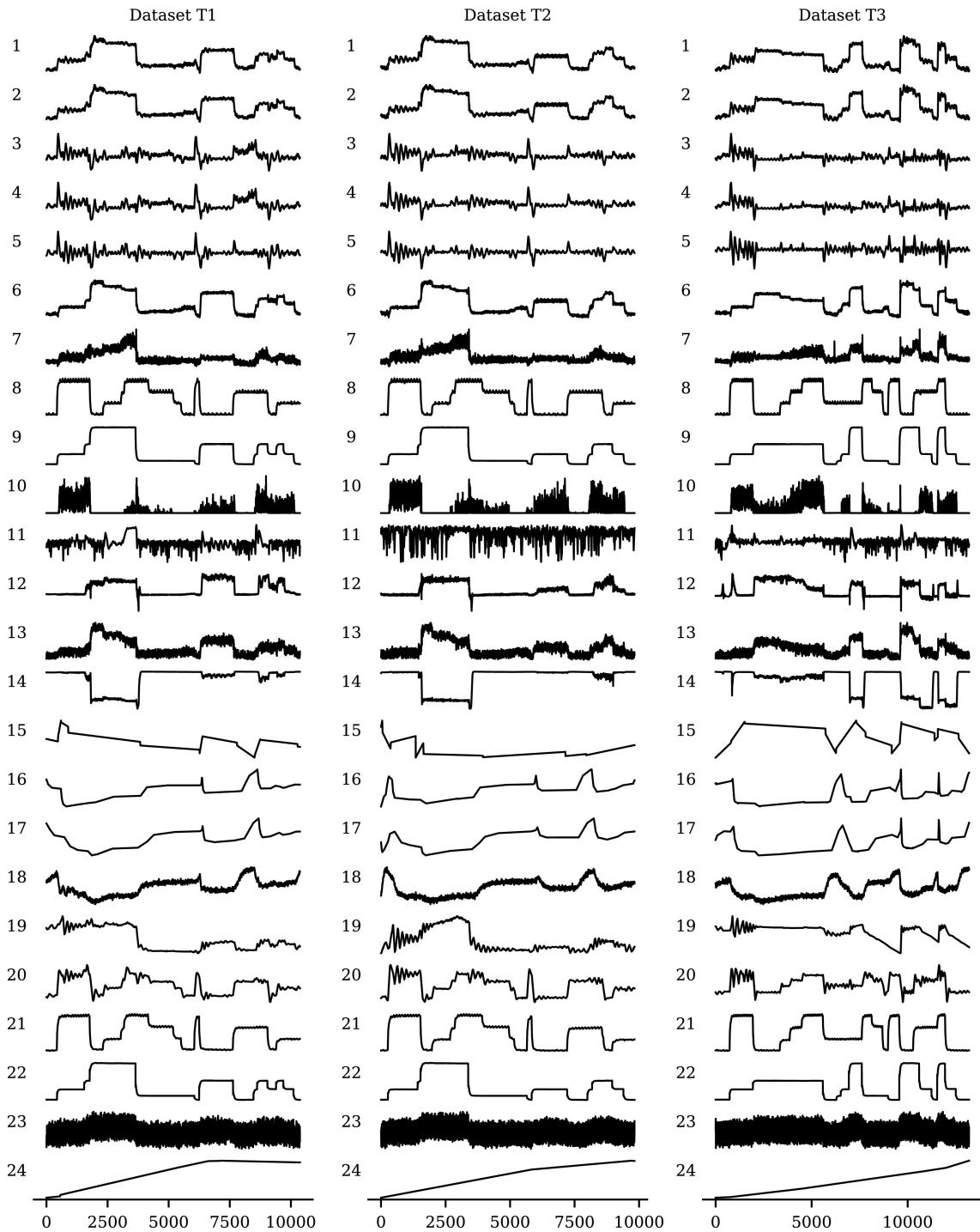


Figura 4.41: Séries temporais dos dados normais do *benchmark* Cranfield.

A autocorrelação é alta para a maioria das variáveis, como mostrado na Figura 4.45. Há alguma variação entre os conjuntos de dados, especialmente os valores menores no conjunto T3.

As razões sinal/ruído mostradas na Figura 4.46 também são altas, com exceção da variável 23, cujo comportamento ruidoso pode ser notado observando sua série temporal. A Figura 4.47 mostra como a diminuição da taxa de amostragem faz com



Figura 4.42: Gráficos de dispersão do conjunto de dados normais T1 do *benchmark* Cranfield.

que a razão sinal/ruído diminua.

A Figura 4.48 mostra os coeficientes de correlação para o conjunto de dados T1. Tanto os coeficientes lineares quanto os não lineares são altos, o que confirma a intuição de grande riqueza de relações entre as variáveis que se apreende pelos gráficos de dispersão. As Figuras 4.49 e 4.50 mostram como os coeficientes de correlação globais variam com a taxa de amostragem e o tamanho da janela de filtro

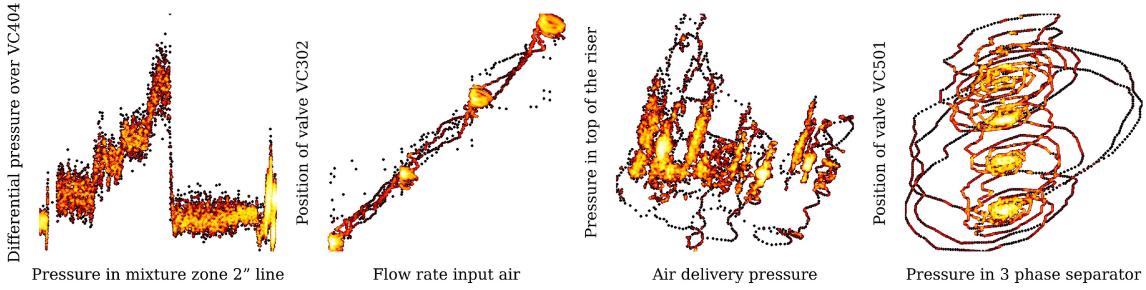


Figura 4.43: Gráficos de dispersão selecionados do conjunto de dados normais T1 do *benchmark* Cranfield.

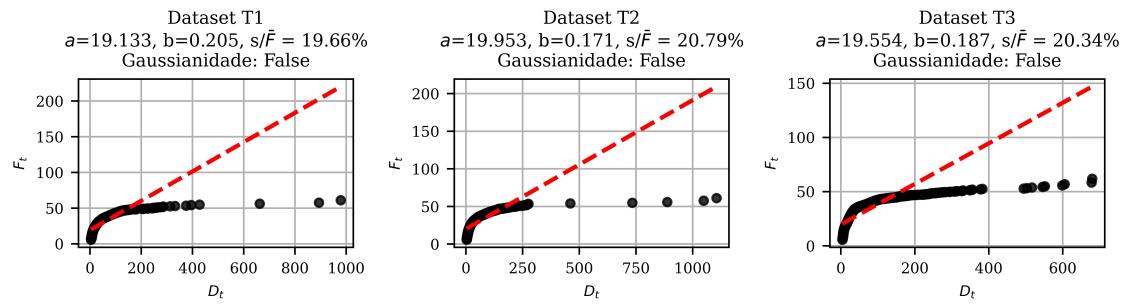


Figura 4.44: Testes de gaussianidade dos conjuntos de dados normais do *benchmark* Cranfield.

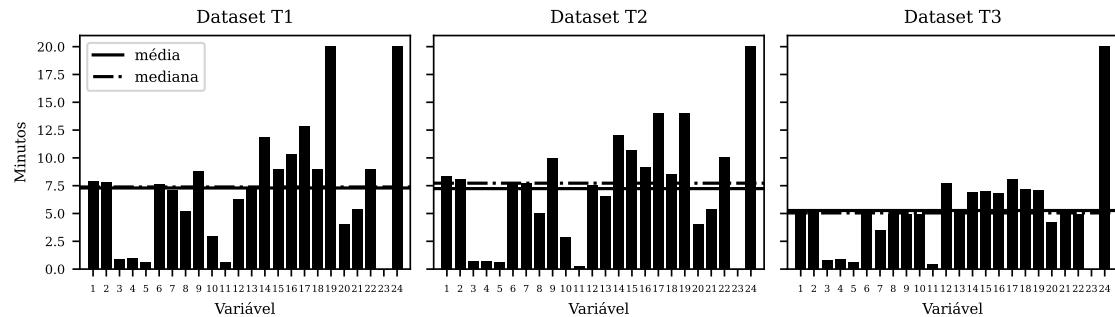


Figura 4.45: Tempo necessário para atingir 0,5 de autocorrelação no *benchmark* Cranfield.

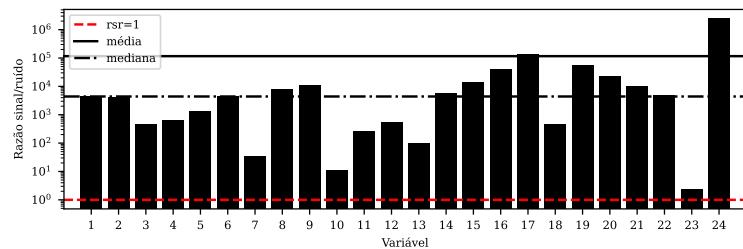


Figura 4.46: Razões sinal/ruído no *benchmark* Cranfield.

de média móvel, respectivamente. Os coeficientes lineares não são sensíveis à taxa de amostragem e os coeficientes não lineares diminuem quando a taxa de amostragem aumenta. Essa é uma indicação do quanto as relações não lineares são de curto

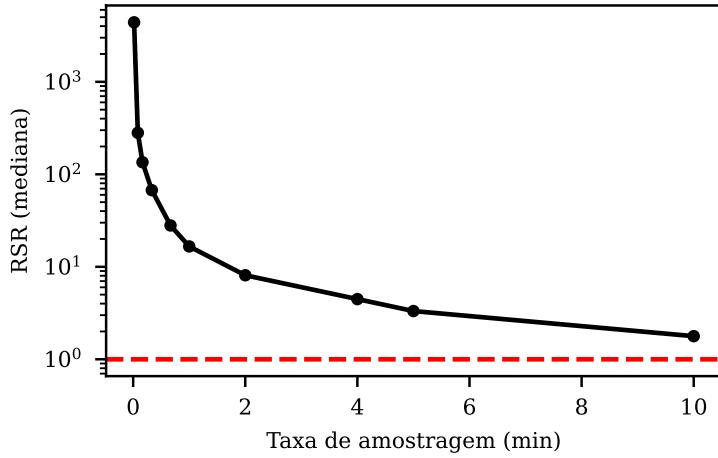


Figura 4.47: Razão sinal/ruído contra taxa de amostragem no *benchmark* Cranfield.

prazo, em oposição às relações lineares. No que se refere ao tamanho da janela do filtro, há pouca variação nos coeficientes de correlação, o que também é esperado, pois as trajetórias dinâmicas já são muito claras mesmo com o conjunto de dados não filtrado.

4.2.4.2 Dados de falha

Existem seis falhas no conjunto de dados, divididas em quatro tipos: bloqueio de linha (1 a 3), vazamento de linha (4), condições de *slugging* (5) e pressurização de linha (6). As falhas 1 a 4 têm cada uma três conjuntos de dados associados e as falhas 5 e 6 cada uma têm dois. A Tabela 4.7 lista as falhas juntamente com a descrição, distribuição, desenvolvimento ao longo do tempo e resultados de detecção com o modelo PCA.

A Figura 4.51 apresenta gráficos de controle para os três conjuntos de dados de Falha 2. No primeiro conjunto de dados, há um aumento gradual no índice de detecção Q , seguido por um aumento acentuado; entretanto, após o término da falha, os valores não retornam abaixo do limite, levando a uma alta taxa de falsos alarmes. Nos outros dois conjuntos de dados, o comportamento gradual não é observado, com aumento acentuado no último quarto do período em falha; além disso, não há problema de altas taxas de falsos alarmes. Essa variabilidade, mesmo entre conjuntos de dados que representam os mesmos fenômenos, é um desafio interessante para os profissionais que utilizam este benchmark.

O leitor que desejar um conjunto completo de descrições, visualizações e discussões de falhas do *benchmark* Cranfield pode consultar os notebooks Jupyter disponibilizados por MELO *et al.* (2022).

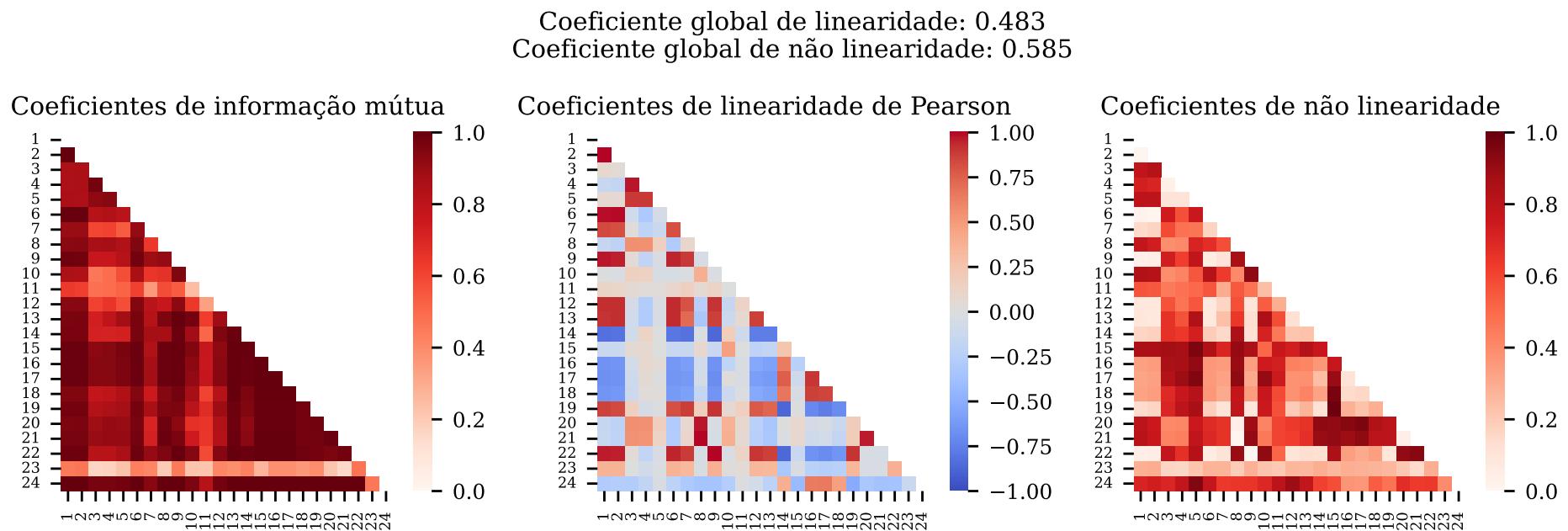


Figura 4.48: Coeficientes de correlação para o conjunto de dados normais T1 do *benchmark* Cranfield.

Tabela 4.7: Análise das falhas do *benchmark* Cranfield.

No.	Descrição	Distribuição	Desenvolvimento	PCA FDR (%)	PCA FAR (%)
1	Bloqueio na linha de ar	Distribuída	Incipiente	49.43, 42.21, 29.97	3.78, 13.36, 6.89
2	Bloqueio na linha de água	Distribuída	Incipiente	51.94, 15.96, 24.53	54.59, 3.5, 6.07
3	Bloqueio na entrada do sep. topo	Distribuída	Incipiente	100.0, 100.0, 99.52	35.43, 100.0, 19.96
4	Abertura de <i>bypass</i>	Distribuída	Incipiente	91.82, 24.7, 84.0	28.28, 8.96, 28.14
5	Condições de slugging	Distribuída	Intermitente	100.0, 87.14	100.0, 17.04
6	Pressurização na linha de 2"	Local	Abrupto	99.81, 99.95	0.81, 0.1

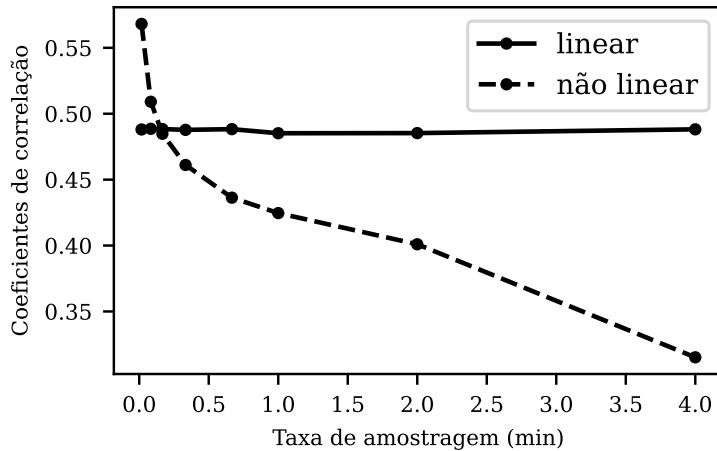


Figura 4.49: Coeficientes de correlação contra taxa de amostragem no *benchmark* Cranfield.

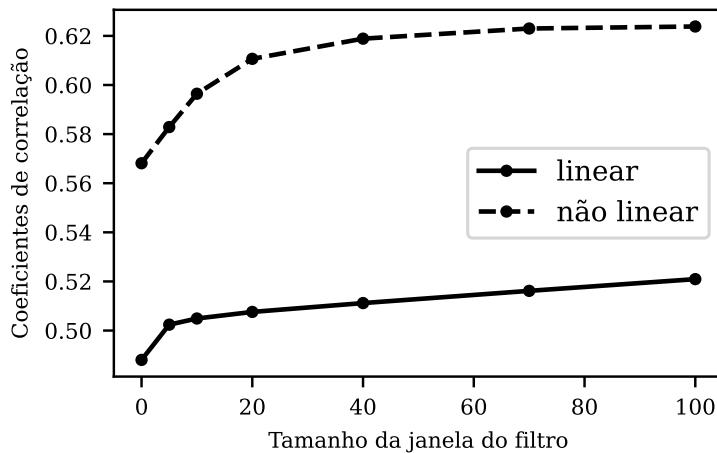


Figura 4.50: Coeficientes de correlação contra tamanho de filtro de média móvel no *benchmark* Cranfield.

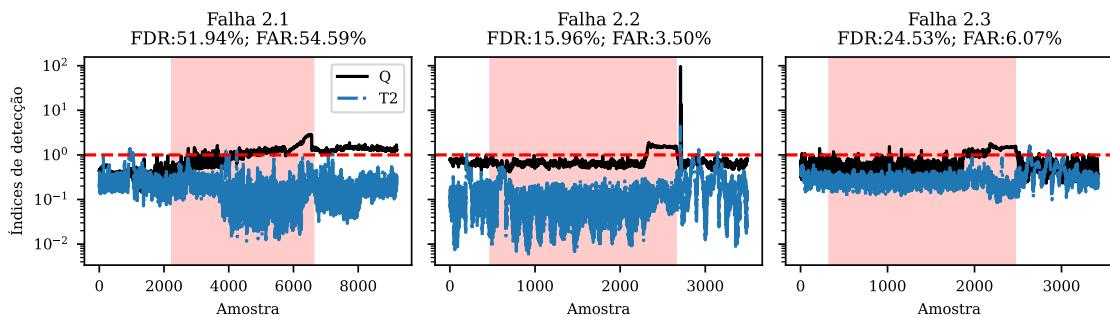


Figura 4.51: Cartas de controle para os três conjuntos de dados da falha 2 do *benchmark* Cranfield.

4.2.5 3W

As oito variáveis disponíveis no *benchmark* 3W estão listadas na Tabela 4.8.

Tabela 4.8: Variáveis do *benchmark* 3W.

No.	Tag	Nome	Unidades
1	P-PDG	Pressão na PDG	Pa
2	P-TPT	Pressão na TPT	Pa
3	T-TPT	Temperatura na TPT	degC
4	P-MON-CK	Pressão à montante da PCK	Pa
5	T-JUS-CKP	Temperatura à jusante da PCK	degC
6	P-JUS-CKGL	Pressão à jusante da GLCK	Pa
7	T-JUS-CKGL	Temperatura à jusante da GLCK	degC
8	QGL	Vazão de <i>gas lift</i>	sm ³ /s

As abreviaturas utilizadas nos nomes referem-se a:

- PDG: *permanent downhole gage*;
- TPT: *temperature and pressure transducer*;
- PCK: *production choke valve*;
- GLCK: *gas lift choke valve*.

As seguintes abreviaturas também são utilizadas nas tags: MONtante; JUSante; CKP = *choke* de produção; CKGL = *choke* de *gas lift*.

Os dados são divididos em blocos correspondentes a intervalos de tempo específicos, denominados pelos autores como “instâncias”. Algumas instâncias são quase consecutivas entre si, com um intervalo tipicamente de um a três minutos do final de uma até o início da outra.

Dados de vários poços estão disponíveis no conjunto de dados. É útil visualizar quão diferentes são os poços. A Figura 4.52 apresenta um gráfico da variável P-MON-CKP utilizando cores diferentes para identificação dos poços. A média de cada instância é plotada como uma linha tracejada. Cada poço tem uma ou mais faixas de regiões características dentro das quais a variável flutua. Há também diferenças entre as variações dessas flutuações.

A Figura 4.53 apresenta um gráfico de dispersão para todos os dados normais. Como esperado, os dados estão concentrados em porções esparsas do espaço de dados. Isso se deve às diferenças entre os poços e entre os períodos de tempo. A variável 7 (temperatura à jusante do GLCK) não contém dados e a variável 8 (vazão de *gas lift*) contém muito pouco.

Um intervalo de tempo em que existem várias instâncias quase consecutivas entre si é chamado no presente trabalho de *período de tempo*. A Tabela 4.9 apresenta

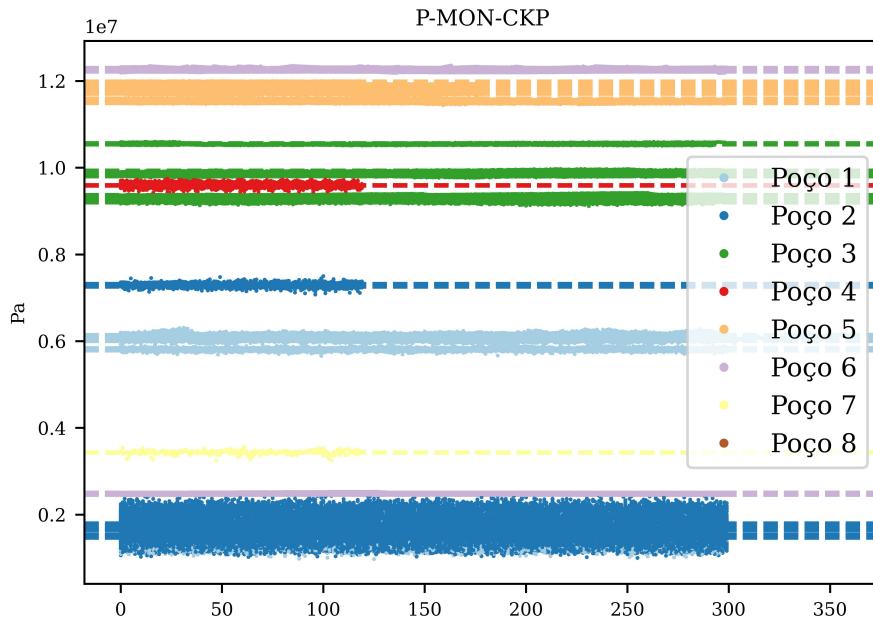


Figura 4.52: Gráficos de séries temporais da variável 4 para todos os 8 poços do benchmark 3W.

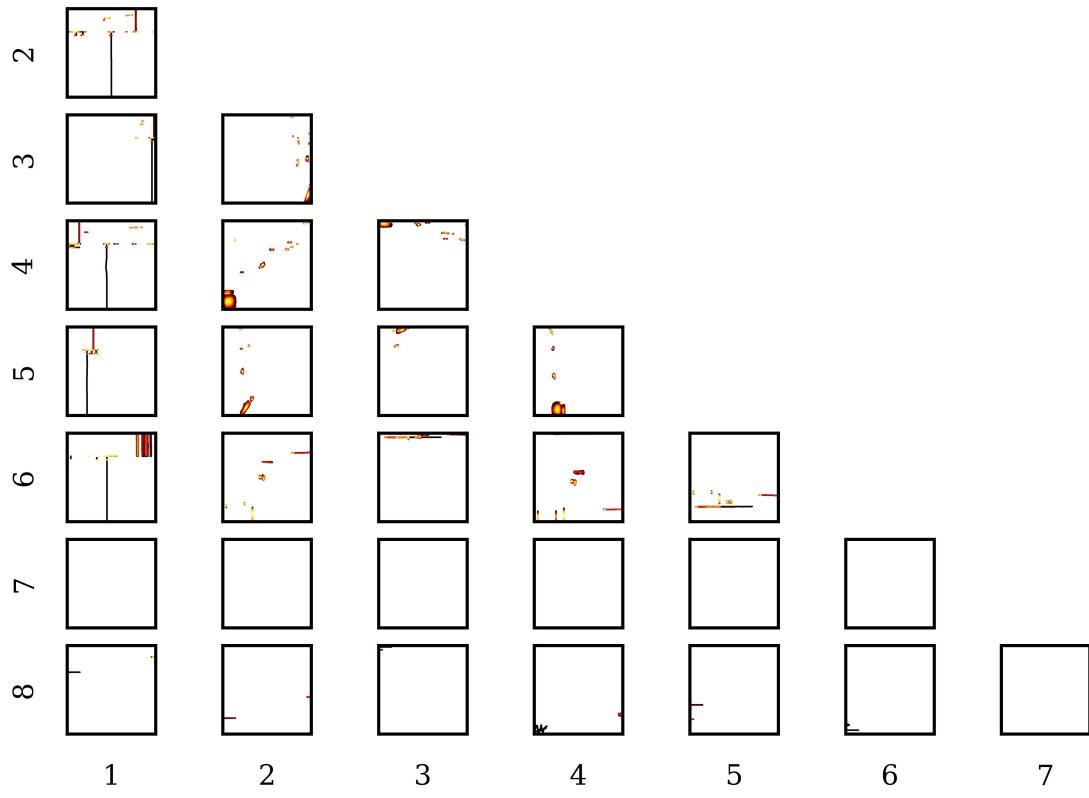


Figura 4.53: Gráficos de dispersão de todos os dados normais do benchmark 3W.

os períodos de tempo identificados nos dados normais por inspeção manual. O ID se encontra no formato $W_i P_j$, onde i se refere ao poço e j ao período. As outras colunas referem-se respectivamente aos *timestamps* de início e término, quais variáveis estão

disponíveis e o número de instâncias que o período contém.

Tabela 4.9: Períodos de tempo identificados nos dados normais do *benchmark* 3W.

ID	Início	Fim	Variáveis	No. inst.
W1P1	2017–02–01 02:02:07	2017–02–02 13:00:00	2, 3, 4, 5, 6	7
W1P2	2017–02–18 01:01:46	2017–02–20 03:00:00	2, 3, 4, 5, 6	10
W1P3	2017–04–21 10:02:51	2017–04–22 09:57:37	2, 3, 4, 5, 6	2
W1P4	2017–04–24 18:00:42	2017–04–25 19:00:00	2, 3, 4, 5, 6	5
W1P5	2017–05–20 19:00:00	2017–05–29 02:57:00	2, 3, 4, 5, 6	40
W1P6	2017–06–26 10:02:46	2017–06–27 05:56:26	2, 3, 4, 5, 6	4
W1P7	2017–08–23 05:00:00	2017–08–28 14:56:17	2, 3, 4, 5, 6	26
W2P1	2013–12–11 23:00:15	2013–12–12 08:59:18	2, 3, 4, 6	5
W2P2	2013–12–14 16:01:27	2013–12–14 22:00:00	2, 3, 4, 6	3
W2P3	2017–02–09 12:00:22	2017–02–20 03:00:00	2, 3, 4, 5, 6	51
W2P4	2017–03–07 20:00:31	2017–03–11 04:00:00	2, 3, 4, 5, 6	16
W2P5	2017–04–04 03:00:00	2017–04–07 00:59:37	2, 3, 4, 5, 6	14
W2P6	2017–06–11 00:01:08	2017–06–27 04:00:00	2, 3, 4, 5, 6	78
W2P7	2017–08–04 21:00:00	2017–08–13 19:58:47	2, 3, 4, 5, 6	43
W3P1	2017–05–14 01:00:00	2017–05–14 20:53:46	1, 2, 3, 4	4
W3P2	2017–08–07 10:00:00	2017–08–07 14:59:34	1, 2, 3, 4	1
W3P3	2017–08–12 02:01:16	2017–08–13 22:58:41	1, 2, 3, 4	9
W3P4	2017–08–20 23:00:53	2017–08–21 19:00:00	1, 2, 3, 4	4
W3P5	2017–09–14 11:00:00	2017–09–16 02:59:08	1, 2, 3, 4	8
W4P1	2014–08–06 22:00:33	2014–08–07 22:00:00	2, 3, 4, 5	12
W5P1	2017–03–31 06:00:14	2017–04–04 05:00:00	2, 3, 4, 5	19
W5P2	2017–05–16 05:00:46	2017–05–17 01:59:13	2, 3, 4, 5	7
W5P3	2017–08–12 11:00:00	2017–08–17 22:57:28	2, 3, 4, 5	43
W5P4	2017–08–30 06:00:00	2017–08–30 23:58:25	2, 3, 4, 5	6
W5P5	2017–09–10 03:00:04	2017–09–10 14:57:47	2, 3, 4, 5	6
W6P1	2017–02–06 22:00:31	2017–02–10 06:00:00	2, 3, 4, 5, 6	16
W6P2	2017–02–22 00:00:00	2017–02–24 21:59:21	2, 3, 4, 5, 6	14
W6P3	2017–05–05 21:00:21	2017–05–10 01:00:00	1, 2, 3, 4, 5, 6	20
W6P4	2017–08–15 06:00:00	2017–08–28 19:00:00	2, 3, 4, 5, 6	65
W7P1	2017–08–01 18:00:00	2017–08–01 21:59:44	1, 2, 3, 4, 5, 8	2
W8P1	2017–06–10 22:02:46	2017–06–15 23:56:26	1, 2, 3	21
W8P2	2017–06–29 17:01:23	2017–07–03 10:56:49	1, 2, 3	18
W8P3	2017–08–16 14:02:22	2017–08–19 16:57:01	1, 2, 3	15

Continua na próxima página

Tabela 4.9 – *Continuada da página anterior*

ID	Início	Fim	Variáveis	No. inst.
W8P4	2017-09-14 02:03:21	2017-09-14 11:57:01	1, 2, 3	2

A Figura 4.54 apresenta uma concatenação dos sete períodos de tempo que compõem os dados do Poço 1 (dos ID's W1P1 a W1P7). As instâncias são separadas por linhas tracejadas azuis e os períodos de tempo são separados por linhas vermelhas. Os períodos de tempo variam em duração. Os comportamentos das variáveis podem diferir muito de um período de tempo para outro.

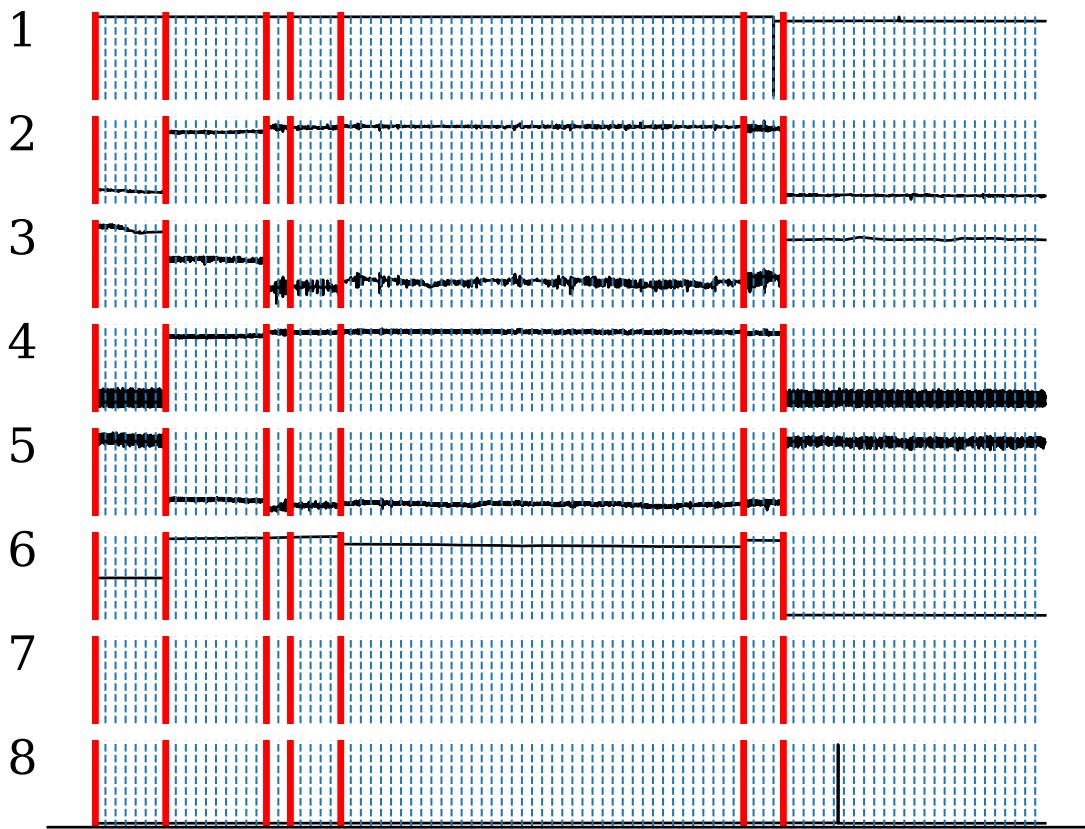


Figura 4.54: Concatenação de períodos de tempo identificados a partir de dados normais do Poço 1. As instâncias são separadas por linhas tracejadas azuis e os períodos de tempo são separados por linhas vermelhas.

A taxa de amostragem do conjunto de dados é de 1 Hz, mas isso não significa que as medições foram efetuadas nessa taxa. Isso fica claro ao observar a Figura 4.55, obtida a partir do período de tempo W_1P_2 . O gráfico à esquerda indica que os sensores efetuaram medições em uma frequência menor que a taxa de amostragem do conjunto de dados. Cada medição provavelmente corresponde a um dos picos e vales observados na série temporal. Quando não há medição, pode-se admitir que o

sistema Osisoft PI, utilizado na geração do conjunto de dados, interpolou os valores com linhas retas. É um comportamento semelhante à plotagem de dados de baixa resolução com linhas em *softwares* de plotagem como o Matplotlib. Mas na Figura 4.55, apenas pontos são usados, sem linhas, portanto o efeito não é devido ao software de plotagem.

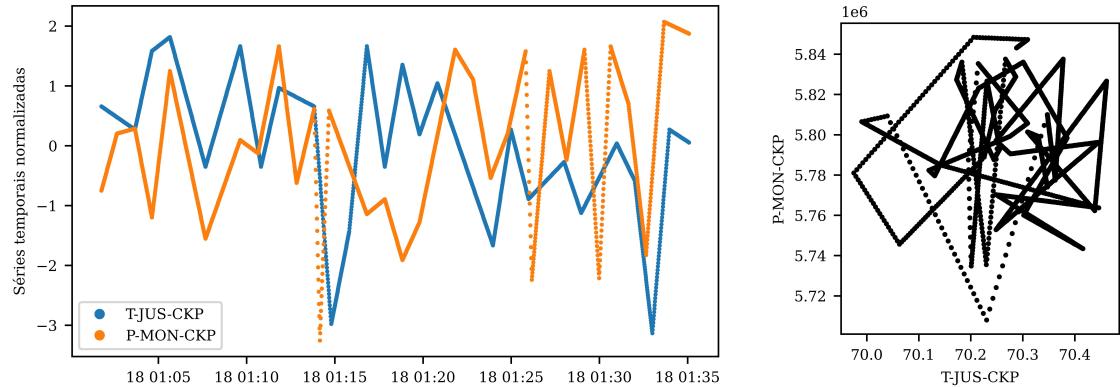


Figura 4.55: Séries temporais e gráficos de dispersão selecionados de um par de variáveis de dados $W_1 P_2$.

O gráfico de dispersão à direita mostra como essas interpolações introduzem correlações espúrias nos dados. Parece que existem trajetórias dinâmicas no espaço de fase entre as duas variáveis, mas isso é apenas o resultado dos altos e baixos das interpolações.

Seria possível a princípio proceder à identificação da taxa de medição real, mas essa não parece ser uma tarefa simples. A Figura 4.56 mostra o resultado da aplicação do pacote findpeaks (TASKESEN, 2022) para identificar os instantes reais de medição. As estrelas correspondem aos extremos detectados na mesma coordenada x para ambas as séries. Os pontos correspondem aos extremos detectados em apenas uma das séries para dada coordenada x . O resultado levanta algumas questões:

- A taxa de erro do algoritmo de identificação de picos é baixa, mas como lidar com os poucos erros?
- A taxa de amostragem não é regular. Em alguns pontos existem pequenas lacunas entre as medições dos dois sensores. Em outros lugares, apenas um dos sensores parece medir. Como alinhar observações de todas as variáveis?

Em alguns casos, o comportamento da interpolação se estende por longos períodos de tempo, como no caso da série temporal representada na Figura 4.57, extraída do período $W_1 P_2$.

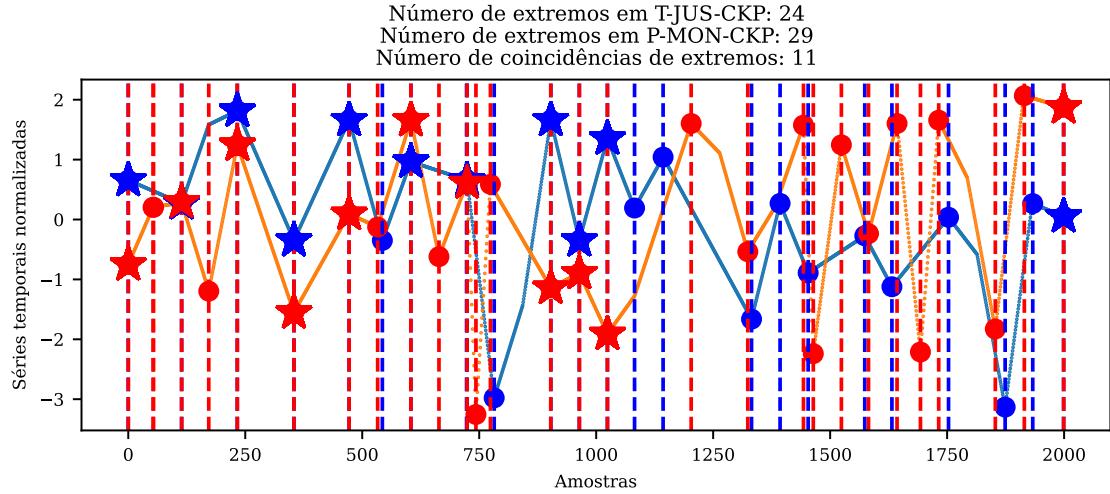


Figura 4.56: Resultados da aplicação de um algoritmo detector de pico aos dados descritos na Figura 4.55.

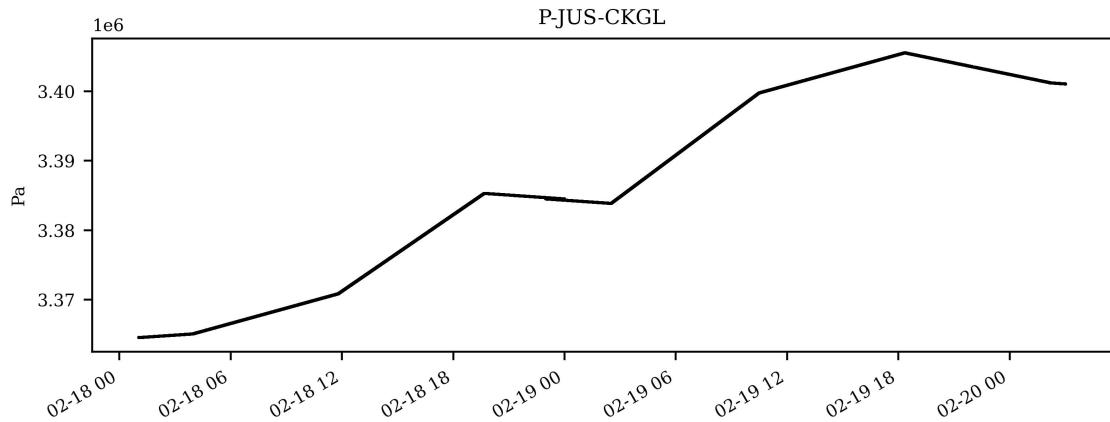


Figura 4.57: Faixa de tempo selecionada da variável 6, período $W_1 P_2$.

A Figura 4.57 faz suspeitar que as medições estão disponíveis apenas em determinados horários do dia. O resto do tempo, o sistema PI também parece interpolar os resultados.

As questões relativas a interpolações e taxas de amostragem não serão necessariamente problemáticos em todos as aplicações, mas qualquer pessoa que lide com o conjunto de dados precisa estar ciente dessas características. Em vários casos, o *downsampling* pode ser importante, embora não seja óbvio como executá-lo de maneira otimizada, pois a dinâmica e as taxas de amostragem diferem de variável para variável.

Para a metodologia de análise exploratória de dados adotada no presente trabalho, a questão da interpolação constitui um impedimento. A título de ilustração, os resultados serão mostrados para a análise do período $W_1 P_2$, representado com séries

temporais e gráficos de dispersão na Figura 4.58. A partir das características das séries temporais observadas, podem ser esperadas baixas autocorrelações, correlações não lineares e razões sinal/ruído. Mas estes não são os resultados das análises, apresentadas na Figura 4.58. Os dados parecem ter uma estrutura muito mais complexa do que realmente têm. Uma solução simples poderia ser a de reduzir a amostragem do conjunto de dados por um fator homogêneo, mas não seria eficaz, pois a dinâmica e o alcance das flutuações são diferentes nas variáveis 2 e 3 em comparação com as variáveis 4 a 5.

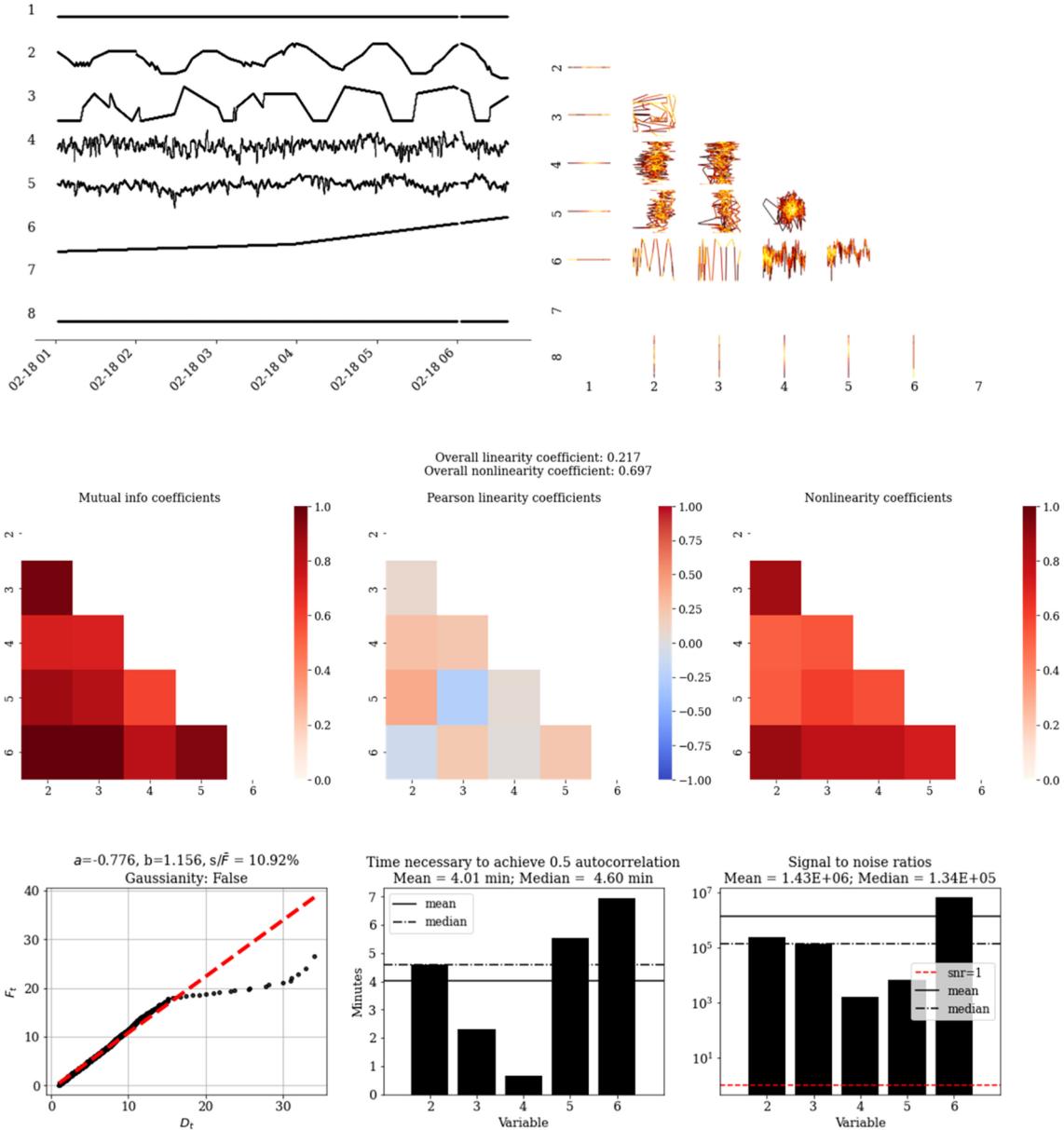


Figura 4.58: Resultados de uma tentativa de se realizar análise exploratória de dados no período $W_1 P_2$.

Como a validade dos resultados não poderia ser garantida sem abordar a questão da taxa de amostragem, decidiu-se não prosseguir com a análise do con-

junto de dados completo. Para uma análise mais detalhada, incluindo os dados em falha, o leitor interessado pode consultar os notebooks Jupyter disponibilizados por MELO *et al.* (2022), além dos notebooks disponíveis no endereço <https://www.kaggle.com/afrniomelo/code>.

4.3 Discussão

Nesta seção, os *benchmarks* analisados neste capítulo, além daqueles revisados no Capítulo 2, são discutidos e comparados de acordo com vários critérios que podem ser úteis para orientar a escolha para uma determinada aplicação. A Tabela 4.10 lista todos os *benchmarks* juntamente com algumas de suas características. A Tabela 4.11 resume os resultados relacionados à análise exploratória de dados.

4.3.1 Modos de operação

Os processos analisados neste trabalho podem ser divididos de acordo com os dois possíveis modos de operação:

- Processos contínuos: TEP, DAMADICS, Cranfield, PRONTO, 3W e FCC.
- Processos em batelada: PenSim e IndPenSim.

Os *benchmarks* para processos contínuos são mais numerosos e variados, refletindo um interesse histórico mais acentuado neste tipo de processo. No entanto, o interesse em processos descontínuos está crescendo, o que pode ser uma motivação para a comunidade gerar mais *benchmarks* e ir além da já conhecida simulação da fermentação da penicilina.

4.3.2 Disponibilidade de um modelo fenomenológico

Quando um modelo fenomenológico está disponível, os estudos podem ser mais flexíveis, permitindo a geração de dados de acordo com a necessidade do usuário. É possível, por exemplo, implementar metodologias de engenharia de sistemas em processos como projeto de processos, otimização de processos e projeto de controladores, integrando essas metodologias com o estudo de monitoramento de processos e diagnóstico de falhas. Os *benchmarks* TEP, PenSim, IndPenSim e FCC fornecem esse recurso.

Tabela 4.10: Resumo das características dos *benchmarks*.

Benchmark	Operação	Referências	Natureza	Plataforma	Tamanho (GB)	No. de vars.	No. de falhas
TEP	Contínua	CHIANG <i>et al.</i> (2001)	S/D	Fortran	0.0265	52	20
		RIETH <i>et al.</i> (2017)	D	—	1.338	52	20
		BATHELT <i>et al.</i> (2015)	S	MATLAB	—	—	—
		MARTIN-VILLALBA <i>et al.</i> (2018)	S	Modelica	—	—	—
		MANCA (2020)	D	—	15.81	85	20
		REINARTZ <i>et al.</i> (2021a)	D	—	132.96	107	28
Pensim	Batelada	BIROL <i>et al.</i> (2002)	S	MATLAB	—	—	—
		VAN IMPE e GINS (2015)	D	—	4	22	15
DAMADICS	Contínua	BARTYS <i>et al.</i> (2006)	S/D {R}	MATLAB	0.03	32	19
RAYMOND	Contínua	GINS <i>et al.</i> (2014)	S	MATLAB	—	—	—
IndPenSim	Batelada	GOLDRICK <i>et al.</i> (2015, 2019)	S/D	MATLAB	2.57	2238	4
Cranfield	Contínua	RUIZ-CÁRCEL <i>et al.</i> (2015)	D {R}	—	0.02	26	6
PRONTO	Contínua	STIEF <i>et al.</i> (2019a,b)	D {R}	—	4.37	29	5
3W	Contínua	VARGAS <i>et al.</i> (2019)	—	—	8	8	—
FCC	Contínua	SANTANDER <i>et al.</i> (2022)	S	MATLAB	—	—	—

Tabela 4.11: Resumo da análise exploratória de dados efetuada.

Benchmark	Dados analisados	Lin.	Não lin.	Gauss.
TEP	RIETH <i>et al.</i> (2017)	0.218	0.195	Verd.
DAMADICS	BARTYS <i>et al.</i> (2006)	0.415	0.373	Verd. em pequenos períodos
IndPenSim	GOLDRICK <i>et al.</i> (2019)	0.421	0.485	Falso
Cranfield	RUIZ-CÁRCEL <i>et al.</i> (2015)	0.488	0.568	Falso
3W	VARGAS <i>et al.</i> (2019)	—	—	Falso

4.3.3 Disponibilidade e fonte de dados

A disponibilidade de dados refere-se à existência de conjuntos de dados consolidados para *download*. Isso é importante já que os usuários não precisam ser proficientes no uso de softwares de simulação, permitindo que mais pessoas usem o *benchmark*. Além disso, pode-se efetuar comparações padronizadas pois há a garantia de toda a comunidade estar utilizando os mesmos dados.

Usualmente, quando os dados estão disponíveis, eles podem vir de duas fontes diferentes: simulações ou medições reais. Ao utilizar dados simulados, uma vantagem é a maior clareza: como o processo de geração de dados é totalmente conhecido, os usuários podem ter mais conhecimento sobre o comportamento dos dados e entender melhor o que ocorre em determinada aplicação. Os dados reais, por outro lado, apresentam desafios de aplicação do mundo real que os dados simulados não possuem, aproximando mais a aplicação aos comportamentos dos algoritmos no uso industrial real.

O conjunto de dados 3W é o único que possui outro tipo de dados: desenhados à mão por especialistas em processos. No conjunto de dados 3W, eles são úteis para refletir tendências gerais identificáveis em gráficos a olho nu, mas não possuem as nuances e a variabilidade refinadas de dados reais ou alguns dados simulados.

4.3.4 Variabilidade de dados

A variabilidade dos dados pode se referir às variabilidades dos períodos normais e dos períodos em falha. Uma alta variabilidade nos dados normais introduzirá mais desafios ao treinamento dos algoritmos. Os dados normais do *benchmark* TEP, utilizados na maioria dos trabalhos da literatura, apresentam pouca variabilidade. O trabalho de REINARTZ *et al.* (2021a) tenta remediar esse problema introduzindo um conjunto de dados significativamente mais variado. Os dados do *benchmark* DAMADICS têm uma grande abrangência temporal quando comparado à taxa de

amostragem e dinâmica natural do processo, levando a variabilidades decorrentes de comportamento anômalo não identificado e *concept drift*. O *benchmark* Cranfield tem variabilidade causada principalmente por mudanças de *set point*. O *benchmark* IndPenSim apresenta a variabilidade comum aos processos em batelada, tanto a dinâmica natural de uma rodada de batelada quanto a variabilidade entre os lotes resultantes. O *benchmark* 3W, apesar de possuir dados de muitos poços de petróleo em um longo intervalo de tempo, apresenta um conjunto reduzido de variáveis disponíveis, além de significativos problemas de qualidade, como dados perdidos e variáveis congeladas, o que reduz consideravelmente a riqueza potencial de comportamento dos dados.

Dados em falha ricos são importantes para se testar como um algoritmo treinado responderá a uma variedade de cenários. Todos os *benchmarks* analisados têm um bom conjunto de possibilidades nesse sentido. Vale ressaltar que o simulador DAMADICS possibilita a geração de uma variedade bem maior de comportamentos do que aqueles encontrados em seus dados reais.

4.3.5 Não linearidade

A não linearidade é importante porque muitas técnicas de monitoramento são propostas especificamente para resolver esse problema. A maioria dos trabalhos sobre essas técnicas utiliza o *benchmark* TEP, embora esse *benchmark* tenha um baixo nível de não linearidade quando comparado aos demais, conforme mostrado na Tabela 4.11. Esse resultado pode ser um incentivo para a comunidade utilizar os outros conjuntos de dados com mais frequência.

4.3.6 Dinâmica

O comportamento dinâmico também é um tópico central de pesquisa na literatura de monitoramento de processos. Todos os *benchmarks* têm algum nível de comportamento dinâmico, embora difiram em natureza e intensidade. A maioria das variáveis no *benchmark* TEP oscila em torno de um *set point* de estado estacionário, apresentando pouca ou nenhuma dependência temporal; entretanto, algumas variáveis apresentam flutuações de longo prazo que são bastante destacadas pela aplicação de filtros para redução do nível de ruído. Com exceção do *benchmark* 3W, os outros benchmarks, especialmente o Cranfield, apresentam comportamento dinâmico mais claro, incluindo dinâmica de curto prazo, o que não está presente no TEP devido à taxa de amostragem comparativamente baixa.

4.3.7 Gaussianidade

Desvios em relação ao comportamento gaussiano introduzem desafios significativos aos algoritmos de monitoramento. O comportamento do *benchmark* TEP é claramente gaussiano. Os dados do *benchmark* DAMADICS testam positivo para o teste gaussiano em períodos curtos, mas observam-se desvios a longo prazo. Todos os outros conjuntos de dados analisados desviam-se amplamente do comportamento gaussiano, o que os torna mais adequados para testar algoritmos que não levam em consideração a hipótese gaussiana, como o algoritmo ICA.

4.3.8 Qualidade de dados

Problemas de qualidade de dados podem ser encarados como vantagem ou desvantagem dependendo da aplicação. Se o objetivo é testar a robustez de procedimentos e algoritmos frente a cenários do mundo real, é bom que os desafios relacionados a qualidade estejam presentes no *benchmark*. O *benchmark* TEP apresenta alto nível de ruído e baixa taxa de amostragem quando comparado aos demais *benchmarks* contínuos disponíveis, apresentando desafios para a identificação de dinâmicas ou não linearidades, por exemplo. O *benchmark* 3W possui dados reais adquiridos em ambientes hostis, o que causa problemas como dados ausentes e variáveis congeladas. Testar algoritmos com este *benchmark* produzirá, portanto, resultados que podem representar com mais precisão aplicações reais que possuem características semelhantes.

4.4 Considerações finais

Neste capítulo propôs-se uma metodologia de análise exploratória de dados voltada especificamente para dados de processos industriais. A aplicação dessa metodologia gera informações valiosas que podem orientar a escolha de técnicas específicas de análise e monitoramento de processos para um determinado conjunto de dados. Além disso, a metodologia foi aplicada em dados de *benchmarks* disponíveis na literatura, possibilitando uma escolha consciente do melhor *benchmark* a ser usado em cada contexto durante a análise de desempenho de novas técnicas.

O *benchmark* TEP é o mais antigo e bem estabelecido, o que se reflete no grande número de opções de simuladores e conjuntos de dados disponíveis, dos mais simples aos mais avançados e complexos. PenSim é o *benchmark* mais popular para processos em batelada. Apesar de suas popularidades, a principal desvantagem do TEP e do

PenSim é a falta de dados reais. DAMADICS é a opção disponível quando se deseja estudar fenômenos no contexto de um único equipamento. IndPenSim propõe um modelo de produção de penicilina em larga escala com dados reais, fornecendo uma alternativa promissora ao PenSim no contexto de monitoramento de processos em batelada. Cranfield e PRONTO tratam de um processo de escoamento multifásico, sendo o primeiro mais simples e o segundo mais complexo, com características únicas como dados heterogêneos. 3W fornece uma ampla gama de representações para os fenômenos envolvidos na produção de petróleo em plataformas *offshore*, incluindo dados reais; também apresenta desafios únicos do mundo real relacionados a dificuldades em lidar com dados adquiridos em ambientes hostis.

Espera-se que o trabalho deste capítulo contribua para a comunidade tanto no que se refere à aplicação de técnicas de análise exploratória de dados de processo quanto para um maior conhecimento dos diferentes *benchmarks* disponíveis na literatura.

Após a obtenção do conhecimento sobre os dados utilizando a análise exploratória, o próximo passo natural é a aplicação de um método de monitoramento. Nesse contexto, propõe-se no próximo capítulo um novo método para o monitoramento com alto nível de interpretabilidade e baseado em padrões intrínsecos de dados.

Capítulo 5

Análise de matrizes de distâncias para detecção e diagnóstico de falhas

Neste capítulo estende-se o trabalho de FADEL (2018) com o objetivo de desenvolver e implementar uma metodologia para detecção e diagnóstico de falhas baseada em padrões de matrizes de distâncias. Uma matriz de distância é uma representação bidimensional que reflete padrões de dados intrínsecos e não é restrinida por uma estrutura de modelo específica. Mostra-se que padrões básicos de dados de processo como degraus, tendências e oscilações podem ser identificados distintamente em matrizes de distância, permitindo seu uso em algoritmos de monitoramento de processos e como ferramentas para auxiliar na compreensão e interpretação de processos. Redes neurais convolucionais são aplicadas para realizar o diagnóstico de falhas. A eficácia da metodologia é ilustrada por meio de exemplos numéricos, que envolvem um reator CSTR e o benchmark Tennessee Eastman Process.

O trabalho deste capítulo aborda alguns dos problemas identificados por BI *et al.* (2022) como desafios para alcançar o monitoramento inteligente de processos, como a adaptabilidade em vários modos de operação, a capacidade de lidar com dados não rotulados e a interpretabilidade do modelo.

As seções do capítulo estão organizadas como descrito a seguir. A Seção 5.1 define o conceito de matriz de distâncias, discute padrões comumente encontrados em dados de sensores de processos e apresenta o algoritmo de detecção de falhas proposto por FADEL (2018). A Seção 5.2 apresenta os resultados das aplicações em um exemplo numérico e dois exemplos de simulação de processos. A Seção 5.3 conclui com as considerações finais.

5.1 Framework teórico

5.1.1 Matrizes de distâncias

Matrizes de distâncias são matrizes compostas de distâncias entre pontos de dados. Embora qualquer norma possa ser usada como distância, geralmente a distância euclidiana é apropriada devido a sua intuitividade e eficiência. Seja uma série temporal multivariada $\mathbf{X} \in \mathbb{R}^{n \times m} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$, com $\mathbf{x}_i \in \mathbb{R}^m$. As linhas de \mathbf{X} correspondem às observações, enquanto as colunas correspondem às variáveis. A matriz de distâncias euclidianas \mathbf{D} é definida como:

$$D_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|, \quad (5.1)$$

para $i, j = 1, \dots, n$. Cada linha ou coluna i de \mathbf{D} armazena as distâncias da amostra i em relação a todas as outras amostras. A Figura 5.1 apresenta uma série temporal junto com sua correspondente matriz de distâncias. Nesse caso, o valor de n é igual a 10, indicando o número de amostras, enquanto m é igual a 1, pois a série temporal é unidimensional. A matriz de distâncias representa visualmente as distâncias por meio de um mapa de cores em tons de cinza, onde tons mais escuros indicam distâncias maiores entre as amostras. A cor vermelha destaca a amostra 8 no gráfico da série temporal, enquanto a oitava linha da matriz de distâncias exibe suas distâncias em relação a todas as outras amostras. Alternativamente, a oitava coluna poderia ser usada, dado que a matriz de distâncias é simétrica ($D_{ij} = D_{ji}$). Vale notar que as coordenadas (8, 3) e (3, 8) aparecem com cores mais escuras, indicando que essas amostras são as mais distantesumas das outras. Além disso, a diagonal principal da matriz tem a cor branca, pois a distância de uma amostra a ela mesma é zero. Apesar da simplicidade conceitual e computacional, as matrizes de distâncias são utilizadas em diversas aplicações importantes, incluindo algoritmos de aprendizado de máquina, localização em redes de sensores e conformação molecular (DOKMANIC *et al.*, 2015).

A Figura 5.2 mostra exemplos de matrizes de distâncias, na forma de mapas de cores, calculados usando a Equação 5.1 para alguns sinais de séries temporais comumente encontrados em dados de processo. As matrizes de distâncias são simétricas; portanto, para interpretar os gráficos, pode-se focar na metade triangular inferior. Como pode-se perceber, cada sinal representado gera seu próprio padrão visual de distâncias. Por exemplo, um sinal degrau aparece como blocos retangulares; padrões de tendência são percebidos como gradientes de cores perpendiculares à diagonal e que se tornam mais escuros perto da borda; tendências após uma constante são compostos por uma combinação dos padrões presentes nos sinais de degrau e

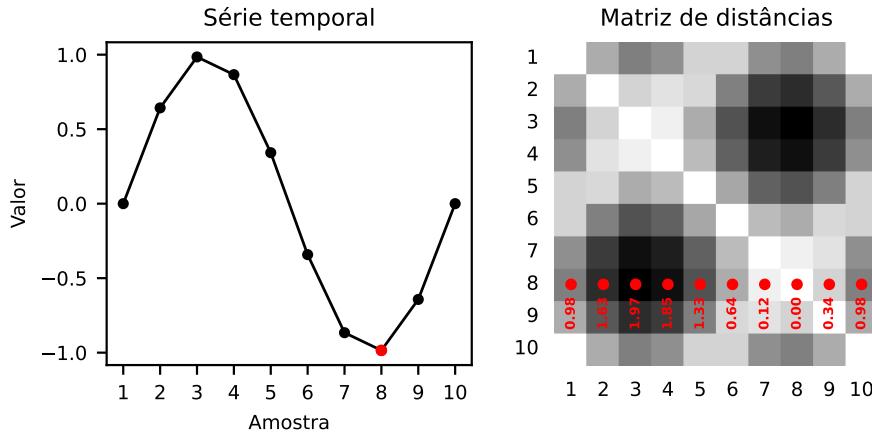


Figura 5.1: Ilustração esquemática de uma série temporal e sua matriz de distâncias correspondente.

tendência; oscilações aparecem como pontos redondos, sendo o número de pontos relacionado à frequência de oscilação; e o ruído distorce os padrões anteriores, conforme ilustrado pelas últimas quatro linhas da Figura 5.2, que mostram uma sequência que muda de um sinal oscilatório puro para ruído puro com magnitude relativa crescente do ruído.

Matrizes de distâncias podem ser interpretadas como gráficos de recorrências sem limite (KOK *et al.*, 2022). O gráfico de recorrências é uma ferramenta usada para avaliar o comportamento dinâmico de um sistema, representando a natureza periódica de uma trajetória no espaço de fases em um gráfico bidimensional (ECKMANN *et al.*, 1987; GOSWAMI, 2019). O gráfico de recorrências é construído como uma representação visual da matriz de recorrências, que é definida como:

$$R_{ij} = \Theta(\epsilon - ||\mathbf{x}_i - \mathbf{x}_j||), \quad (5.2)$$

onde Θ é a função Heaviside e ϵ é o raio de uma hiperesfera que define um limiar de proximidade entre duas observações. A função Heaviside torna \mathbf{R} uma matriz booleana, com elementos positivos nas posições (i, j) onde o sistema se repete, ou seja, onde \mathbf{x}_i é suficientemente próximo a \mathbf{x}_j . A Figura 5.2 apresenta exemplos de gráficos de recorrências junto com as matrizes de distâncias correspondentes. Ao desconsiderar o limite, eliminando a trabalhosa tarefa de sintonizá-lo (SCHINKEL *et al.*, 2008), a Equação 5.2 se torna a Equação 5.1 e \mathbf{R} se torna \mathbf{D} .

Pode-se argumentar que a matriz de distâncias \mathbf{D} , apesar de mais simples, possui ainda mais informações dinâmicas do que \mathbf{R} , que é binária e influenciada pelo escolha arbitrária do limite. Nesta perspectiva, a matriz de distâncias é convenientemente interpretada como uma rica representação do comportamento dinâmico de um sistema.

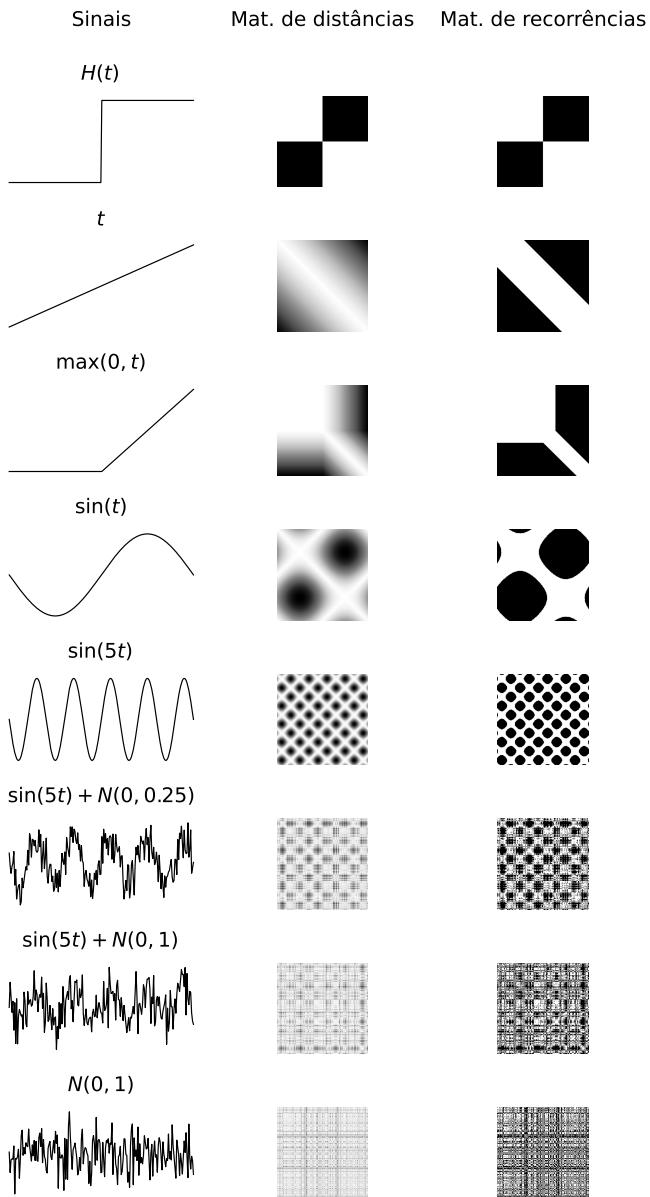


Figura 5.2: Matrizes de distâncias e recorrências para sinais de processo típicos. As distâncias medianas foram usadas como limiares nos gráficos de recorrência. H refere-se à função de Heaviside e N refere-se à distribuição normal.

Os gráficos da Figura 5.2 reforçam essa noção intuitiva, pois as matrizes de distâncias são versões mais sutis das informações presentes nas matrizes de recorrências.

Em dados de processos reais, os sinais típicos representados na Figura 5.2 aparecem combinados em diferentes proporções. Essas combinações se refletem nas matrizes de distâncias, o que as torna uma ferramenta poderosa para avaliar visualmente o comportamento de uma série temporal multivariada. A Figura 5.3 apresenta três matrizes de distâncias obtidas de sinais que combinam ruído, degraus, tendências e oscilações; pode-se identificar nas matrizes de distâncias as combinações dos padrões correspondentes.

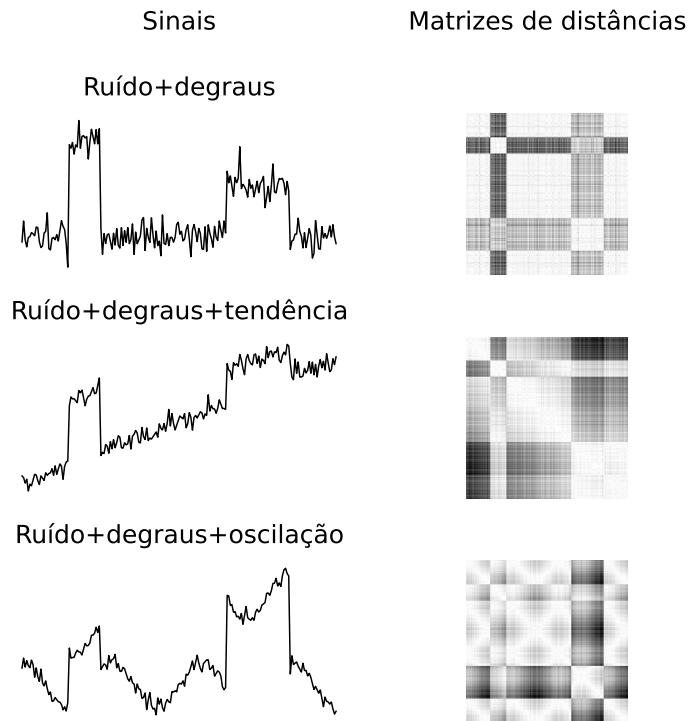


Figura 5.3: Matrizes de distâncias para combinações típicas de sinais de processo.

As matrizes de distâncias também podem ser calculadas para uma série temporal multidimensional. Isso traz a principal vantagem de lidar com a matriz de distâncias no lugar da série temporal original: a redução dos dados para 2 dimensões em uma representação que pode ser mais facilmente interpretada por humanos e também analisada por algoritmos de processamento de imagens. Análises de dados multidimensionais a partir de um exemplo numérico e simulações de processo são apresentadas na Seção 5.2.

Outros autores usaram anteriormente a análise de matrizes de distâncias para fins de monitoramento de processos. Por exemplo, BARDINAS (2018); BARDINAS *et al.* (2018); KOK e ALDRICH (2019); KOK *et al.* (2022) empregaram métodos de análise de textura para extrair variáveis a partir de padrões de matrizes de distâncias. Especificamente, esses estudos envolveram descritores de textura, como níveis de cinza ou textons²⁷, que foram extraídos de matrizes de distâncias e usados como entrada para algoritmos de aprendizado de máquina, como redes neurais convolucionais. Embora eficaz, essa abordagem é limitada a representações que podem ser expressas em termos de descritores de textura. As representações resultantes podem ser difíceis de interpretar e geralmente requerem etapas de processamento adicionais, como a decomposição em PCA. Em contraste, a abordagem do presente

²⁷Textons são as microestruturas gráficas que definem o menor nível de atomicidade para a percepção visual humana (ZHU *et al.*, 2005).

trabalho foca em padrões de processo interpretáveis que podem ser usados na análise exploratória visual de dados. Ao analisar matrizes de distâncias diretamente, a metodologia proposta fornece uma representação mais abrangente da estrutura de dados subjacente, capturando relações sutis presentes nos dados. Portanto, essa abordagem alternativa resulta em um fluxo de análise simplificado e tem potencial para oferecer novas ideias para aplicações em que os descritores de textura podem não ser ideais.

No presente trabalho, utiliza-se o algoritmo de detecção de falhas proposto por FADEL (2018) como ponto de partida para: (i) o desenvolvimento de uma nova metodologia de visualização de dados, e (ii) a inclusão de diagnóstico de falhas. O algoritmo original se baseia em medidas de dissimilaridade entre matrizes de distâncias de janelas móveis. As métricas de monitoramento resultam da comparação direta de matrizes de distância, permitindo controlar a adaptabilidade escolhendo a posição relativa entre as janelas móveis. Na subseção a seguir, o algoritmo é detalhado.

5.1.2 Algoritmo de detecção de falhas

O algoritmo descrito nesta subseção foi proposto por FADEL (2018). Ele é baseado na definição de duas janelas de tempo de tamanho WS , chamadas W_1 e W_2 , separadas por uma distância D . A cada iteração k da etapa de treinamento, uma janela estendida é gerada pela concatenação de W_1 e W_2 e é usada para calcular uma matriz de distâncias D_k . Em seguida, as duas janelas avançam d passos e o processo é repetido por K iterações até que todos os dados de treinamento sejam contemplados. O resultado da etapa de treinamento é um banco de dados de K matrizes de distâncias $\{D_1, \dots, D_K\}$ que representa os padrões de distâncias do sistema. A Figura 5.4 ilustra a etapa de treinamento do algoritmo.

O índice de detecção de falhas é baseado na dissimilaridade média de uma determinada matriz de distâncias em relação a todas as outras. A métrica de dissimilaridade selecionada para este trabalho é a norma de Frobenius, que tem a forma:

$$FN(\mathbf{X}, \mathbf{Y}) = \sqrt{\text{Tr}[(\mathbf{X} - \mathbf{Y})(\mathbf{X} - \mathbf{Y})^T]}. \quad (5.3)$$

A norma de Frobenius é uma medida da distância euclidiana entre matrizes quando o espaço matricial é considerado euclidiano (COHEN, 2021).

O índice de detecção de falhas para uma determinada janela de dados é assim definido como a norma de Frobenius média da matriz de distâncias que representa a

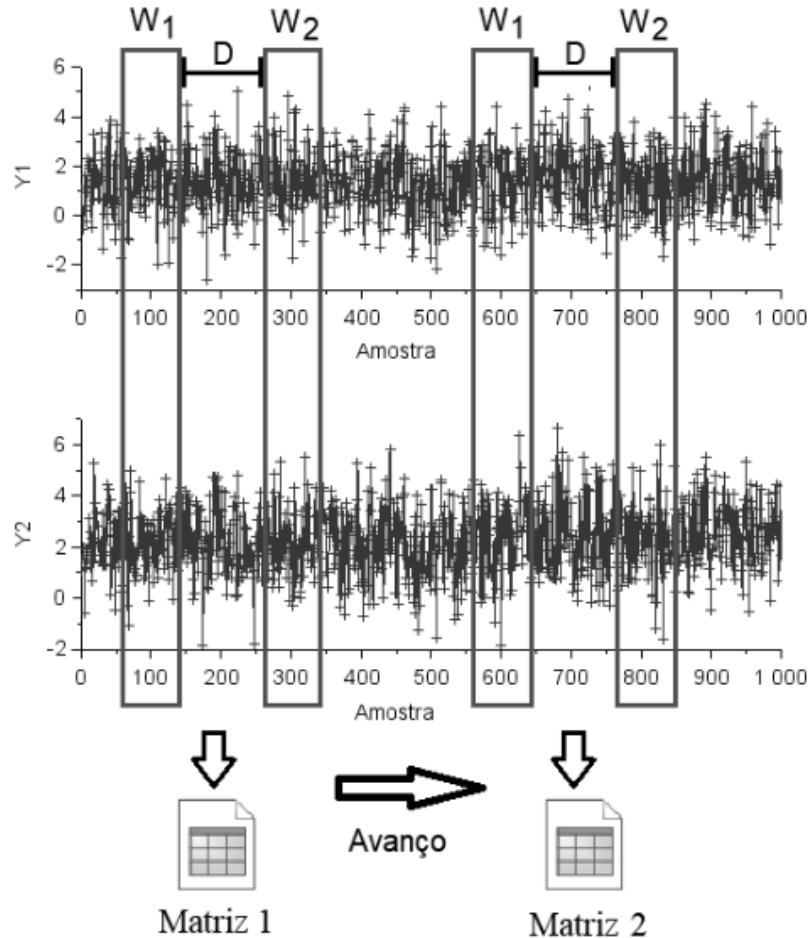


Figura 5.4: Etapa de treinamento do algoritmo de detecção de falhas (FADEL, 2018).

janela presente em relação a todas as outras matrizes que compõem a base de dados de treinamento $\{D_1, \dots, D_K\}$. O limite de detecção é definido como um percentil de índices de detecção gerados com outros dados normais durante a etapa de validação. Normalmente, esse percentil é selecionado como 95 ou 99%.

Durante a fase de monitoramento *online*, as matrizes de distâncias obtidas em tempo real são comparadas com as do banco de dados de treinamento para detectar anormalidades, conforme mostrado na Figura 5.5. A comparação pode ser feita de duas maneiras:

1. *Modo de referência móvel*: as duas janelas que geram as matrizes de distâncias *online* se movem juntas. Dessa maneira, apenas a distribuição relativa das distâncias entre as observações é monitorada: quando as observações se movem juntas de uma região de operação para outra, mas as distâncias relativas permanecem as mesmas, o algoritmo não indica comportamento anormal. Este modo, portanto, resulta em um esquema de monitoramento que é adaptável às mudanças.

2. *Modo de referência fixa*: a primeira janela W_1 permanece fixa enquanto a segunda janela W_2 se move. O índice de detecção resultante torna-se sensível a alterações de regiões operacionais, indicando-as como anormalidades.

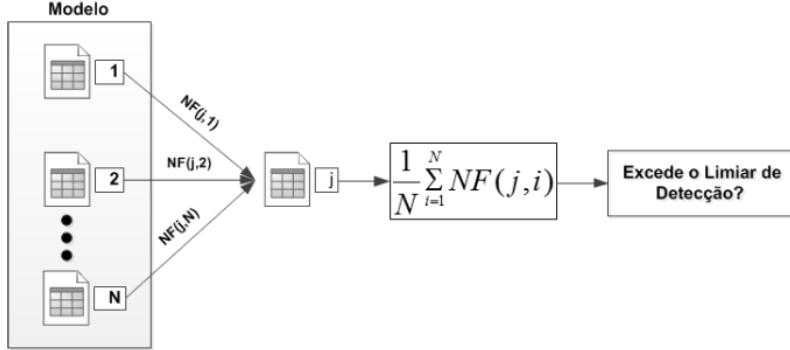


Figura 5.5: Etapa de teste do algoritmo de detecção de falhas (FADEL, 2018).

Ao se usar os dois modos simultaneamente, obtém-se um método poderoso e prático para o monitoramento de ambos os aspectos de um processo: a distribuição relativa das distâncias, que é independente do referencial, e a posição absoluta das observações, que é dependente do referencial. Essa é a característica que dá ao algoritmo suas capacidades adaptativas.

Os hiperparâmetros da estrutura proposta podem ser ajustados, dependendo dos tempos característicos da dinâmica do sistema. Quanto maiores forem os valores de WS e D , maior será o intervalo de dados considerado para o cálculo dos padrões de distância, permitindo assim a caracterização de dinâmicas lentas.

As principais contribuições do presente trabalho para aumentar a aplicabilidade do algoritmo descrito são: i) a interpretação dos padrões de distâncias de acordo com o descrito na Subseção 5.1.1; e ii) a proposta de utilização de redes neurais convolucionais para diagnóstico de falhas. Na próxima seção, casos ilustrativos serão apresentados para ilustrar a aplicação desses dois pontos.

5.2 Resultados e discussão

5.2.1 Exemplo 1 - Caso numérico

O primeiro exemplo, também usado por FADEL (2018), é um caso numérico bidimensional simples (AFZAL *et al.*, 2017) descrito pela Equação 5.4:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0,2 \\ 0,3 & 1 \end{bmatrix} \cdot N \left(\begin{bmatrix} 1 & 2 \end{bmatrix}, \begin{bmatrix} 1,3 & 0,3 \\ 0,3 & 1,5 \end{bmatrix} \right) + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}, \quad (5.4)$$

onde N denota a distribuição normal e e_1 e e_2 são ruídos independentes com distribuição $N(0, 10^{-6})$. Quatro mil amostras foram geradas e são apresentadas na Figura 5.6. Uma carta de controle para o cenário normal é apresentada na Figura 5.7. A linha tracejada horizontal indica o limite de detecção.

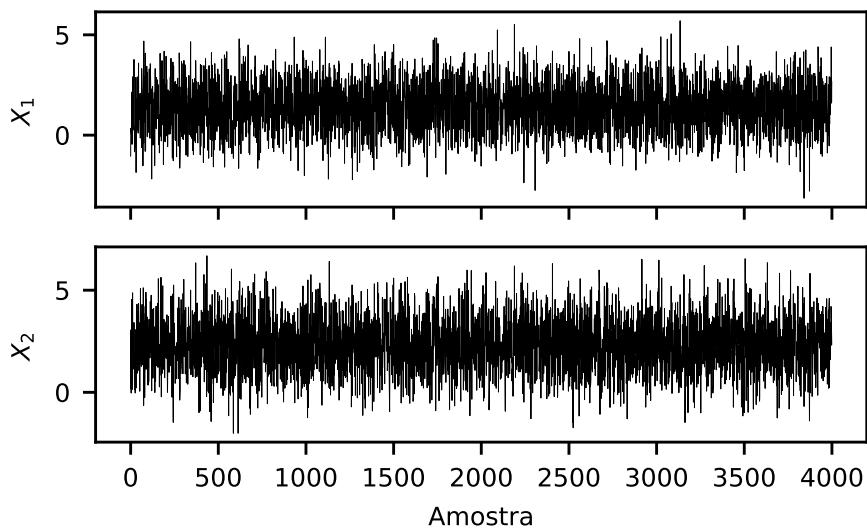


Figura 5.6: Séries temporais do Exemplo 1.

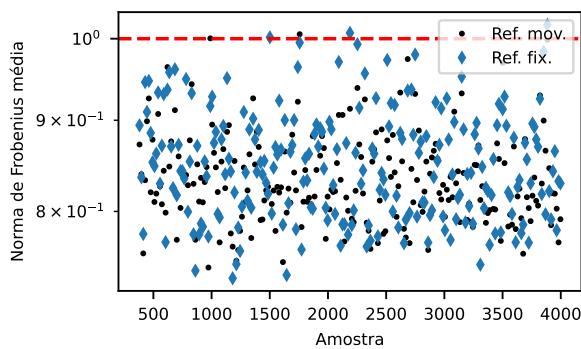


Figura 5.7: Carta de controle para o cenário normal do Exemplo 1.

Foram introduzidos cinco comportamentos anormais tipicamente encontrados em dados de processo: degrau, aumento de variabilidade, oscilação, tendência e tendência seguida de estabilização. Gráficos de linhas e de dispersão para os cinco cenários, juntamente com cartas de controle com resultados de detecção de falhas, são apresentados na Figura 5.8. Planos de fundo e pontos vermelhos indicam comportamento anormal. Os hiperparâmetros usados no algoritmo de detecção de

falhas foram $WS = 16$, $D = 350$ e $d = 16$.

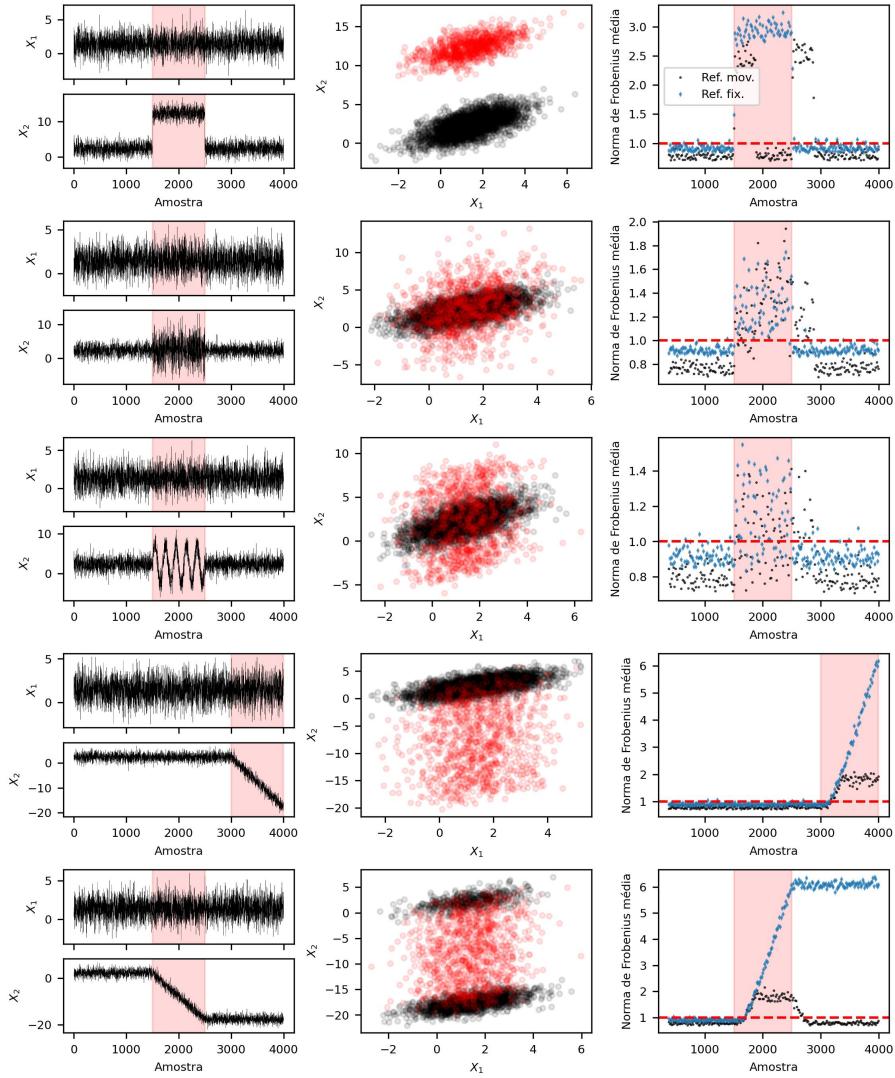


Figura 5.8: Cenários em falha analisados no Exemplo 1: degrau, aumento de variabilidade, oscilação, tendência e tendência seguido de estabilização.

O índice obtido pelo modo de referência fixa apresenta sensibilidade adequada a eventos em todos os cenários. Utilizando o modo de referência móvel, diferentes tipos de comportamentos adaptativos são observados nos casos de degrau e tendência. No caso de degraus, o índice acusa a falha, depois se adapta, depois acusa a nova mudança quando o sistema retorna ao estado estacionário original e depois se adapta novamente. No caso da tendência, o índice do modo de referência móvel permanece sustentado acima do limite de detecção, mas sem aumentar mais, ao contrário do modo de referência fixa, em que os desvios se refletem diretamente no comportamento do índice. Quando a estabilização ocorre após a tendência, o índice do modo de referência fixo se estabiliza bem acima do limite, enquanto o índice do modo de referência móvel retorna para baixo do limite. O algoritmo, portanto, é capaz de se adaptar a diferentes regiões de operação, mas ainda indica as transições e as distâncias

entre elas em uma carta de controle simples de entender. Para fins ilustrativos, as matrizes de distâncias são mostradas na Figura 5.9 para os cenários analisados no Exemplo 1. Os padrões identificados na Seção 5.1 podem ser claramente observados de novo.

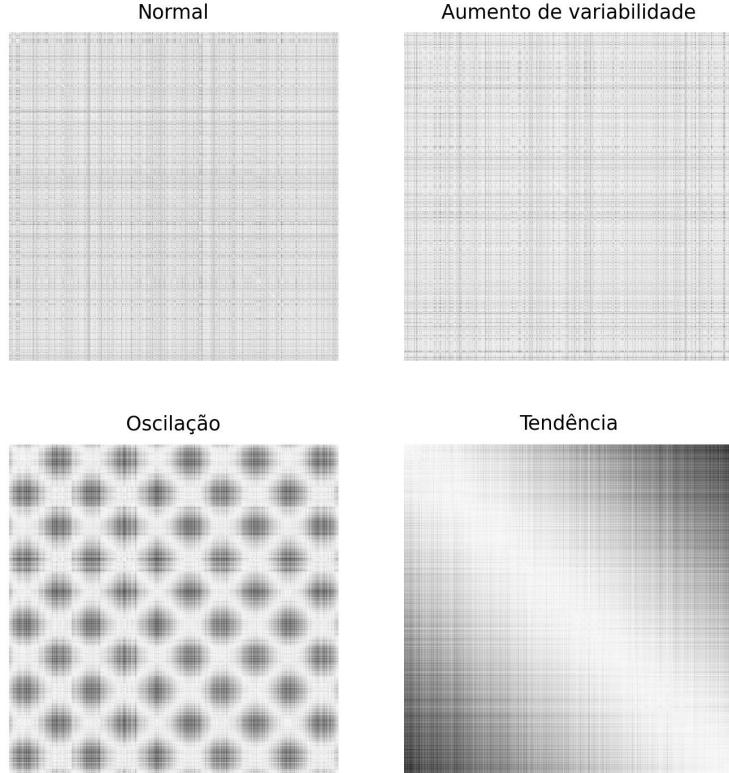


Figura 5.9: Matrizes de distâncias para os cenários analisados no Exemplo 1.

5.2.2 Exemplo 2 - CSTR

Este exemplo consiste em um processo CSTR simulado para a reação $A \rightarrow B$. O modelo de processo é descrito pelas Equações 5.5–5.11:

- Balanço material global:

$$\frac{dV}{dt} = F_i - F \quad (5.5)$$

- Balanço material por componente:

$$\frac{dVC_A}{dt} = F_iC_{Ai} - FC_A - VrC_A \quad (5.6)$$

- Balanço de energia:

$$\frac{dVT}{dt} = F_i T_i - FT - \frac{\Delta H}{\rho c_p} V r C_A - \frac{UA}{\rho c_p} (T - T_j) \quad (5.7)$$

- Balanço de energia na jaqueta de resfriamento:

$$\frac{dT_j}{dt} = \frac{F_j(T_{ji} - T_j)}{V_j} + \frac{UA}{\rho_j c_{pj} V_j} (T - T_j) \quad (5.8)$$

- Equação de Arrhenius para a taxa de reação:

$$r = \kappa_0 e^{-E/RT} \quad (5.9)$$

- Equações de controle:

$$F = F_{set} + P_V(V_{set} - V) \quad (5.10)$$

$$F_j = F_{j,set} + P_T(T_{set} - T) \quad (5.11)$$

As Tabelas 5.1 e 5.2 listam parâmetros e medições do processo (FEITAL *et al.*, 2013). O volume V e a temperatura T do reator são controlados pela manipulação das vazões de saída do reator e da jaqueta de resfriamento, F e F_j , respectivamente. Há três estados estacionários no processo, representados na Tabela 5.2 pelos três conjuntos de valores possíveis para as variáveis C_A , T e T_j . O estado estacionário operacional é selecionado por meio do *set point* para T . Incerteza de medição (em variáveis medidas) e distúrbios de processo (em variáveis de entrada) foram simulados por meio de sequências gaussianas independentemente distribuídas. O desvio-padrão do ruído de medição foi definido como 10% da variabilidade de cada variável no período de treinamento. Os desvios-padrão dos distúrbios de processo foram definidos como 1,5 ft³/h, 0,02 mol/ft³, 1 R e 1 R para as variáveis de entrada F_i , C_{ai} , T_i , T_{ji} , respectivamente. É importante ressaltar que, ao incluir ruídos de medição e distúrbios de processo nas séries temporais, a robustez do método proposto pode ser demonstrada ao lidar com diferentes incertezas inerentes ou fatores externos.

Tabela 5.1: Parâmetros da simulação do CSTR.

Parâmetros	Descrição	Valores	Unidades
V_j	Volume da jaqueta	3,85	ft^3
ρ	Densidade do líquido	50	lb_m/ft^3
ρ_j	Densidade do líquido de resfr.	62,5	lb_m/ft^3
k_0	Fator pré-exponencial	$7,08 \times 10^{10}$	1/h
E	Energia de ativação	3×10^4	BTU/mol
R	Constante de gás universal	1,99	BTU/mol R
ΔH	Entalpia de reação	-3×10^4	BTU/mol
C_p	Cap. calorífica do líquido	0,75	BTU/ lb_m R
C_{pj}	Cap. calorífica do líquido de resfr.	1	BTU/ lb_m R
U	Coef. de transf. de calor na jaqueta	150	BTU/h ft^2 R
A	Area de transf. de calor	250	ft^2
n_k	Número de pontos de amostragem	2880	-
t_s	Intervalo de amostragem	2	min
P_V	Ganho do controlador de volume	-10	1/h
P_T	Ganho do controlador de temperatura	-4	ft^3/h R

Tabela 5.2: Medições da simulação do CSTR.

Variáveis	Descrição	Valores	Unidades	Classificação
F_i	Vazão de entrada	40	ft^3/h	Entrada
C_{ai}	Conc. do reagente na entrada	0,5	mol/ ft^3	Entrada
T_i	Temp. da entrada	530	R	Entrada
T_{ji}	Temp. da entrada da jaqueta	530	R	Entrada
F_j	Vazão na jaqueta	49,9	ft^3/h	Manipulada
F	Vazão de saída do reator	40	ft^3/h	Manipulada
V	Volume do reator	48	ft^3	Estado
C_A	Conc. do reagente	0,4739 0,2451 0,0591	mol/ ft^3	Estado
T	Temp. do reator	537,16 599,99 651,06	R	Estado
T_j	Temp. da jaqueta	536,62 594,63 641,79	R	Estado

Três cenários anormais foram simulados: 1) mudança de *set point*, 2) desativação do catalisador e 3) perda de controle. O cenário 1 foi simulado alterando-se o *set point* da temperatura de 651,06 R para 537,16 R. O cenário 2 foi simulado aplicando-se a função de desvio representada na Figura 5.10 ao fator pré-exponencial κ_0 da Lei de Arrhenius (Equação 5.9). O cenário 3 foi simulado desligando o controle após metade do tempo de simulação. Matrizes de distâncias representativas calculadas com

observações coletadas após o início de cada evento²⁸ são mostradas na Figura 5.11. Nota-se que todos os três cenários têm padrões de distância associados característicos. O padrão de tendência é claramente reconhecível no caso da desativação do catalisador. O padrão também está presente no caso de perda de controle, mas junto com os característicos pontos redondos das oscilações.

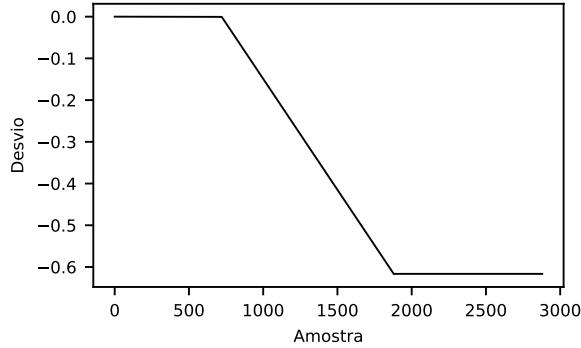


Figura 5.10: Função de desvio aplicada a κ_0 no cenário de desativação do catalisador do Exemplo 2 (CSTR).



Figura 5.11: Matrizes de distâncias para os casos analisados no Exemplo 2 (CSTR).

As cartas de controle resultantes são apresentadas nas Figuras 5.12 e 5.13. Os hiperparâmetros utilizados foram os mesmos do primeiro exemplo: $WS = 16$, $D = 350$ e $d = 16$. Os resultados são consistentes com os encontrados no primeiro exemplo. Usando o modo de referência móvel, o método se adapta a um novo estado operacional quando os padrões de distâncias se tornam semelhantes aos do estado anterior. Isso acontece quando os controladores são capazes de rejeitar a perturbação e apresentar um desempenho compatível com o desempenho que foi aprendido. Por outro lado, usando o modo de referência fixa, a estabilização do índice de detecção ocorre acima do limiar, indicando que as observações residem em uma região de operação diferente da anterior. Isso ocorre nos cenários de mudança de *set point* e desativação de catalisador. No caso da perda de controle, as distâncias não se

²⁸A mudança de set point não foi apresentada pois, como estão representados apenas os cenários após os eventos, a matriz de distâncias resultaria igual ao caso normal.

estabilizam, portanto ambos os índices permanecem acima do limite de detecção após o evento.

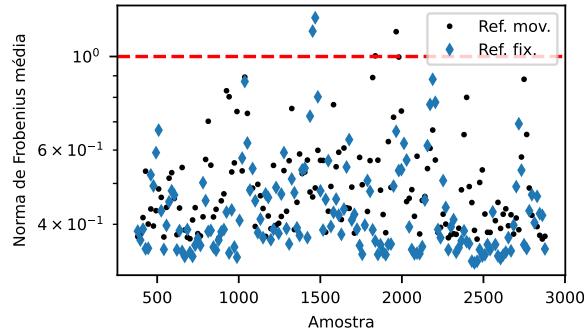


Figura 5.12: Carta de controle para o cenário normal do Exemplo 2 (CSTR).

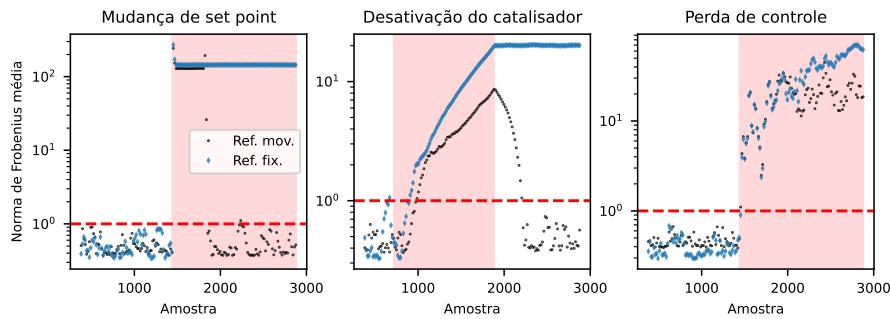


Figura 5.13: Cartas de controle para os cenários de falha do Exemplo 2 (CSTR).

5.2.3 Exemplo 3 - *Benchmark TEP*

O último exemplo diz respeito ao benchmark *Tennessee Eastman Process*, descrito em detalhes nos Capítulos 2 e 4. O conjunto de dados utilizado é o mesmo do Capítulo 4 (RIETH *et al.*, 2017).

Matrizes de distâncias representativas para o cenário normal e os cenários de falhas são apresentadas na Figura 5.14. Cada gráfico foi gerado usando 500 pontos de simulação. A figura permite apreender algumas das características mais importantes do comportamento dos dados: (i) dados normais – IDV(0) – contém ruído, oscilações e degraus; (ii) as falhas 3, 4, 9, 15 e 19 apresentam padrões de distância semelhantes aos dados normais; (iii) as falhas 1, 5 e 7 se manifestam como oscilações regulares que amortecem ao longo do tempo; (iv) as falhas 8 e 10 levam a oscilações que não amortecem com o tempo; (v) as falhas 2, 6, 12, 13 e 18 levam a afastamentos do ponto de operação original, com oscilações prévias no caso das falhas 12, 13 e 18; (vi) as falhas 11 e 14 estão relacionadas a mudanças na variabilidade; (vii) as falhas 17 e 20

são compostas por ciclos de tendências e degraus²⁹. Portanto, a Figura 5.14 constitui uma forma muito eficiente de se visualizar a essência dos dados do *benchmark* TEP, considerando que se trata de uma única representação pictórica de 21 simulações, cada uma composta por 500 observações e 33 características, totalizando 346.500 observações.

Além da aplicação para comparação de diferentes conjuntos de dados, a visualização das matrizes de distâncias pode ser útil para comparar diferentes realizações de um mesmo tipo de evento. No conjunto de dados TEP, quatro falhas específicas - 12, 13, 16 e 18 - podem levar a resultados consideravelmente diferentes, dependendo da simulação. A Figura 5.15 apresenta as matrizes de distâncias para diferentes simulações da falha 12. Há oscilações em todos os casos, mas o bloco retangular escuro que indica um afastamento considerável do ponto de operação inicial não está necessariamente presente. Além disso, os blocos retangulares podem aparecer em diferentes tamanhos e intensidades de cor, dependendo de quando acontece a partida e do quanto longe ela vai, respectivamente. Um comportamento semelhante é observado nos casos das falhas 13 e 18. A Figura 5.16 apresenta as matrizes de distâncias para diferentes simulações da falha 16. Esta é a falha cujas simulações levam aos resultados mais diversos, como se pode perceber por meio dos diferentes padrões oscilatórios das matrizes na figura.

As cartas de controle são apresentadas nas Figuras 5.17 e 5.18. Os hiperparâmetros usados no algoritmo de detecção de falhas foram $WS = 40$, $D = 10$ e $d = 10$. Nas falhas 1, 2, 5 e 7, o modo de referência móvel do algoritmo indica que novos estados estacionários foram alcançados com distribuições de distâncias compatíveis com os iniciais. Na falha 5, o modo de referência fixa também retorna à região abaixo do limite de detecção, indicando que o sistema retornou ao seu ponto operacional original. A falha 4 é detectada apenas pelo modo de referência fixa. As falhas 6, 8, 11, 12, 13, 17, 18 e 20 são detectadas por ambos os modos. O método detecta as falhas 10, 16 e 19 com dificuldade. As falhas 3, 9 e 15 não são observáveis por este método, pois os padrões de distância são os mesmos que os padrões normais devido à ação do sistema de controle. É inclusive questionável do ponto de vista semântico denominar esses cenários como falha, como praticado na literatura, já que o controle mitiga os efeitos anormais nas variáveis de processo. A Tabela 5.3 lista as taxas de detecção de falhas alcançadas pelo modo de referência fixa do algoritmo.

O tamanho da janela é o hiperparâmetro mais importante na estrutura do modelo, pois determina o intervalo de dados considerado para calcular os padrões de distâncias. A Figura 5.19 mostra a variação nas taxas de detecção de falhas com

²⁹A falha 16 apresenta um comportamento peculiar que será detalhado adiante.

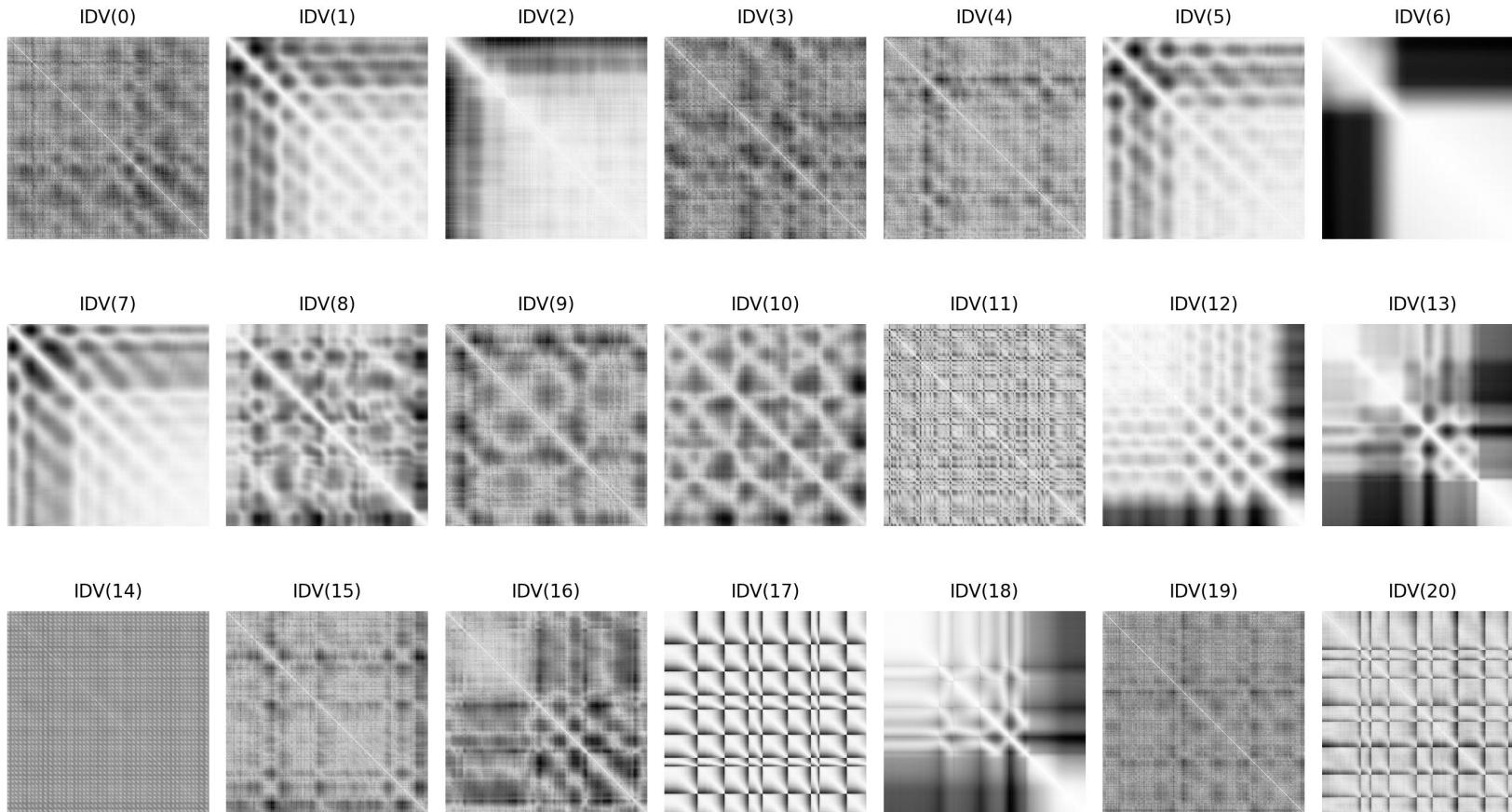


Figura 5.14: Matrizes de distâncias para os casos analisados no Exemplo 3 (TEP).

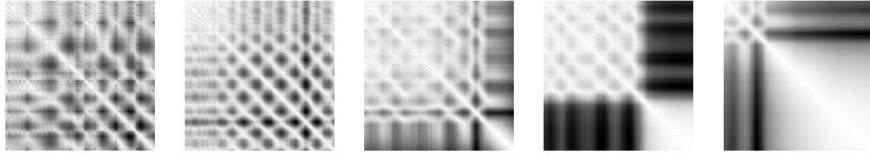


Figura 5.15: Matrizes de distâncias para diferentes simulações do evento IDV(12) do *benchmark* TEP.

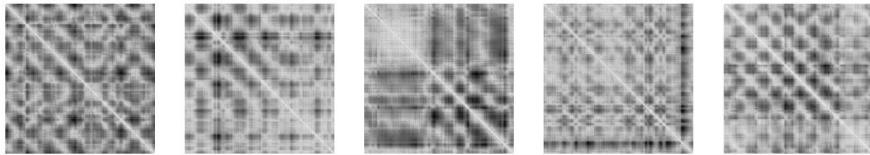


Figura 5.16: Matrizes de distâncias para diferentes simulações do evento IDV(16) do *benchmark* TEP.

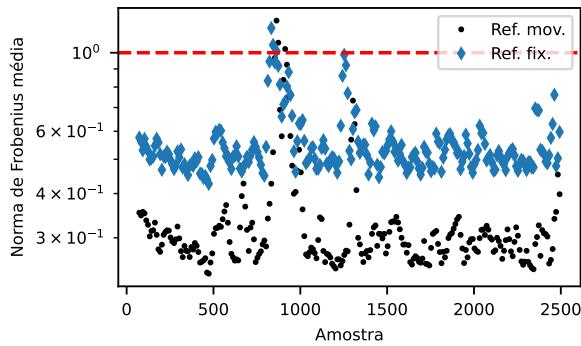


Figura 5.17: Carta de controle para o cenário normal do Exemplo 3 (TEP).

base em diferentes tamanhos de janela para os cenários de falha do TEP. Algumas falhas, como as falhas 1, 2 e 6, exibem dinâmica rápida e podem ser capturadas efetivamente com tamanhos de janela relativamente curtos. Por outro lado, as falhas 4, 11 e 20 se desenvolvem com dinâmica mais lenta e se beneficiam de tamanhos de janela maiores. Neste estudo de caso específico, o tamanho ideal da janela pode ser determinado selecionando o tamanho mínimo que atinge o melhor desempenho. Para outras aplicações, é importante conduzir a análise de hiperparâmetros para identificar os valores ideais específicos.

O diagnóstico de falhas usando matrizes de distâncias pode ser efetuado com a aplicação de redes neurais convolucionais, como mostrado por OLIVIER e ALDRICH (2020). As redes neurais convolucionais são modelos de aprendizado profundo inspirados na arquitetura do córtex visual cerebral e designados para processamento de imagens. Elas têm sido amplamente aplicadas no contexto de monitoramento de processos (LIU *et al.*, 2021b; XING e XU, 2022; ZHANG *et al.*, 2021a). A arquitetura da rede empregada no presente trabalho foi construída utilizando-se a biblioteca

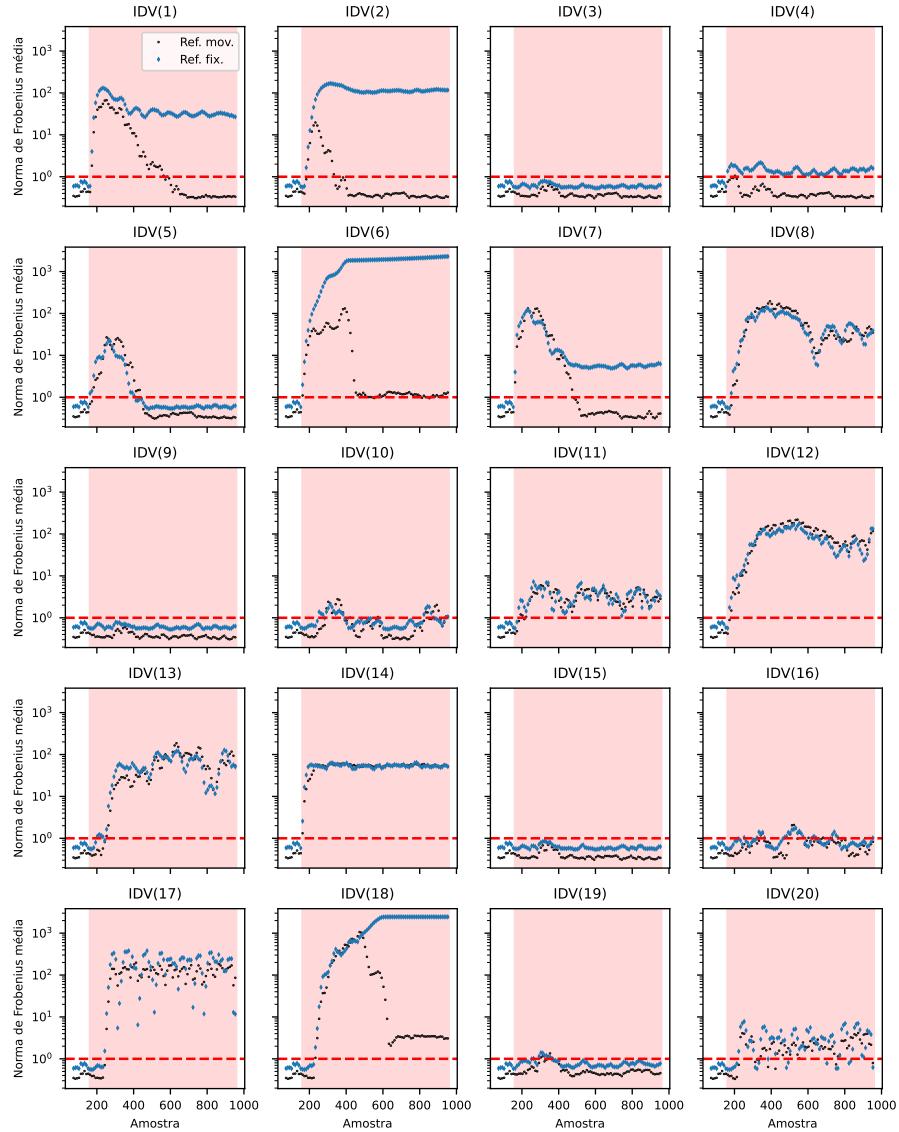


Figura 5.18: Carta de controle para os cenário de falha do Exemplo 3 (TEP).

Keras (CHOLLET, 2021) e é apresentada na Figura 5.20. As matrizes de entrada foram redimensionadas para 64×64 pixels usando interpolação bilinear, seguindo para identificação de características por meio de uma sequência de camadas de convolução e *pooling* e finalmente passando por três camadas densas para classificação (CHOLLET, 2021). O modelo é projetado para classificar uma dada matriz de distâncias em uma das 17 falhas observáveis. As falhas 3, 9 e 15 foram excluídas da análise, pois suas matrizes de distâncias são indistinguíveis para o algoritmo.

A Figura 5.21 apresenta as curvas de aprendizado obtidas treinando-se a rede com matrizes de tamanhos de janela *WS* crescentes. A métrica escolhida foi a revocação ou taxa de verdadeiro-positivo. As falhas 1, 2, 4, 5, 6, 7 e 14 podem ser identificadas com baixo sobreajuste usando-se poucas amostras. A maioria das outras falhas alcança boa revocação com a adição de mais amostras. A Figura 5.22 apresenta

Tabela 5.3: Taxas de detecção de falhas do Exemplo 3 (TEP).

Falha	FDR
IDV(1)	0.988
IDV(2)	0.975
IDV(3)	0.00
IDV(4)	1.00
IDV(5)	0.316
IDV(6)	1.00
IDV(7)	1.00
IDV(8)	0.975
IDV(9)	0.00
IDV(10)	0.241
IDV(11)	0.975
IDV(12)	0.987
IDV(13)	0.937
IDV(14)	1.00
IDV(15)	0.00
IDV(16)	0.304
IDV(17)	0.899
IDV(18)	0.911
IDV(19)	0.101
IDV(20)	0.785

as curvas de aprendizado para a acurácia, que pode ser usada sem problemas neste caso, pois as classes são balanceadas³⁰. Existe um ótimo global em torno de 250 amostras, metade do tamanho total do conjunto de dados. A partir desse ponto, as taxas de verdadeiros positivos das classes mais problemáticas começam a diminuir. Isso acontece porque, em geral, o período mais distinguível para a identificação de falhas é o início, quando os efeitos das falhas estão mais intensos. Após certo tempo, a ação da malha de controle mitiga os efeitos das falhas. A Figura 5.23 apresenta a matriz de confusão dos resultados para as etapas de treinamento e teste. Todos os erros estão relacionados a pelo menos uma das quatro falhas previamente identificadas por inspeção visual como aquelas que geram os padrões mais diversos

³⁰A acurácia refere-se à proporção total de instâncias classificadas corretamente, o que dá uma estimativa não enviesada do desempenho do classificador apenas quando o número de instâncias em cada classe é aproximadamente igual.

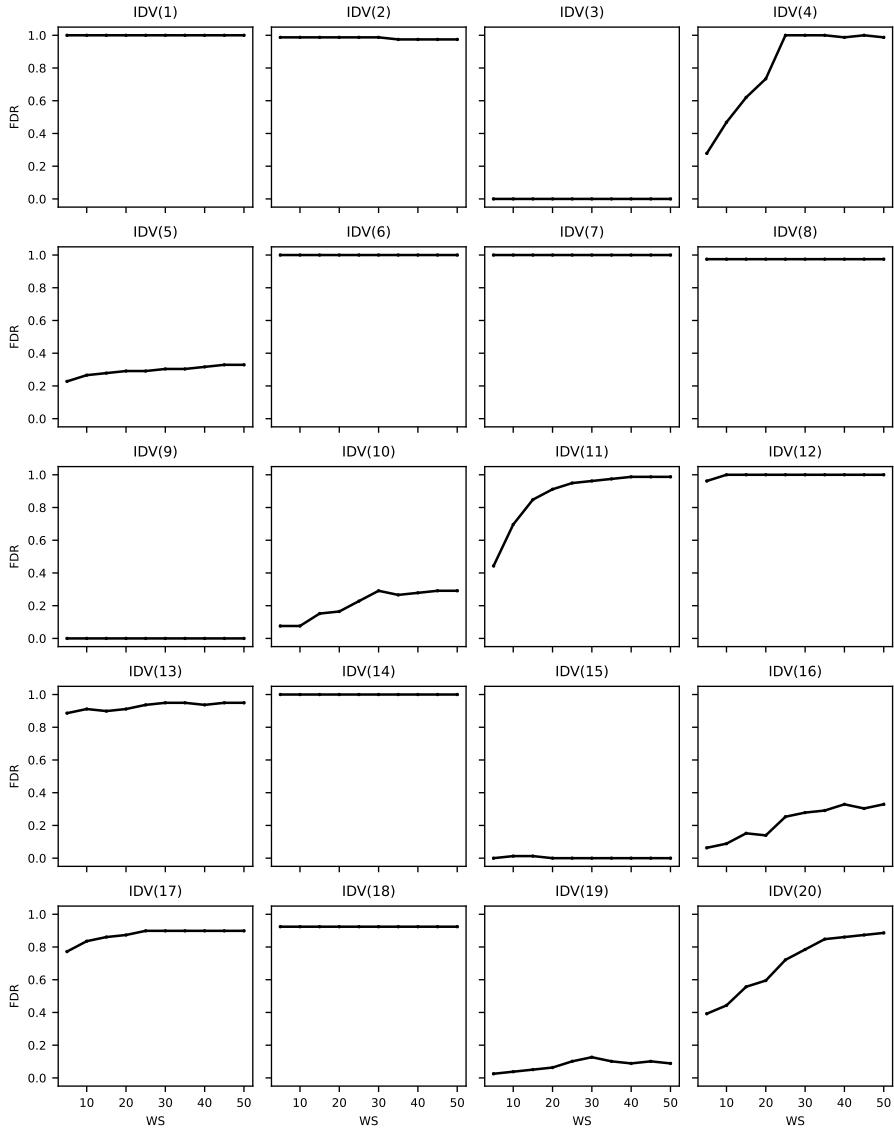


Figura 5.19: Taxas de detecção de falhas em função dos tamanhos de janela para os cenários de falha do Exemplo 3 (TEP).

de distâncias: 12, 13, 16 e 18. A falha 16, em particular, gera o maior problema para o classificador, o que faz sentido já que essa é a falha com os padrões mais variados. Além desses erros esperados, o modelo apresenta excelente desempenho de diagnóstico. É importante ressaltar que nenhum dos hiperparâmetros nos algoritmos de detecção e de diagnóstico foi ajustado. Isso possivelmente tornaria os resultados ainda melhores por uma pequena margem.

5.3 Considerações finais

Neste capítulo aplicou-se a análise de padrões de matrizes de distâncias para fins de monitoramento de processos. Mostrou-se que padrões comuns de dados de

processo podem ser representados como padrões distintos em matrizes de distâncias, permitindo tanto uma fácil interpretação visual humana quanto uma fácil manipulação por algoritmos de processamento de imagens. Aplicou-se um algoritmo de detecção de falhas adaptativo desenvolvido em um trabalho anterior (FADEL, 2018) a novos estudos de caso e, por fim, propôs-se o diagnóstico de falhas baseado em uma combinação de padrões de distâncias e redes neurais convolucionais. Além da eficácia numérica refletida nas altas acuráncias, a abordagem leva a representações simples que podem ser facilmente interpretadas pelos operadores do processo, ajudando a melhorar a compreensão do processo e a integração humano-máquina.

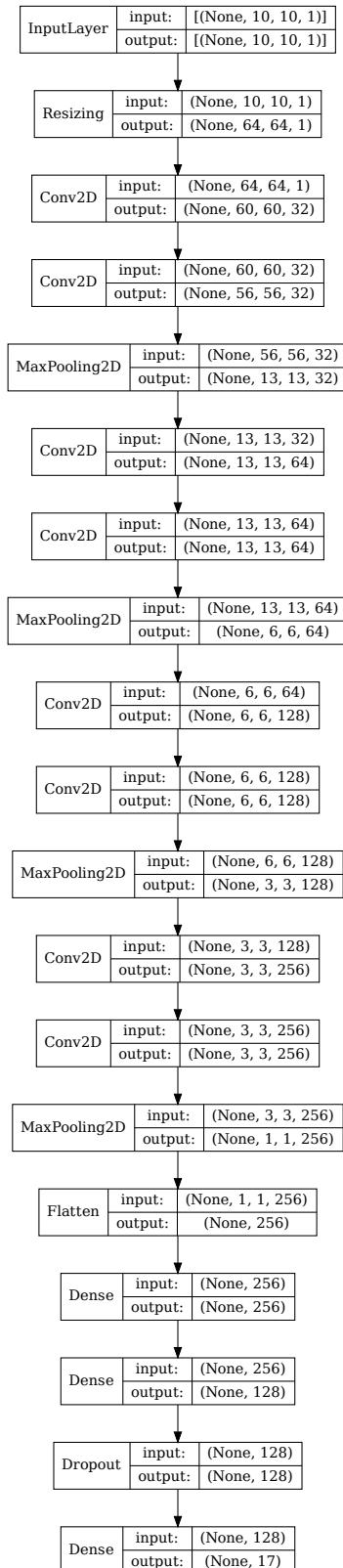


Figura 5.20: Arquitetura da rede neural convolucional usada para diagnosticar falhas de TEP. *None* corresponde à dimensão referente ao tamanho do lote, que é independente da arquitetura da rede e escolhida no momento de ajuste do modelo.

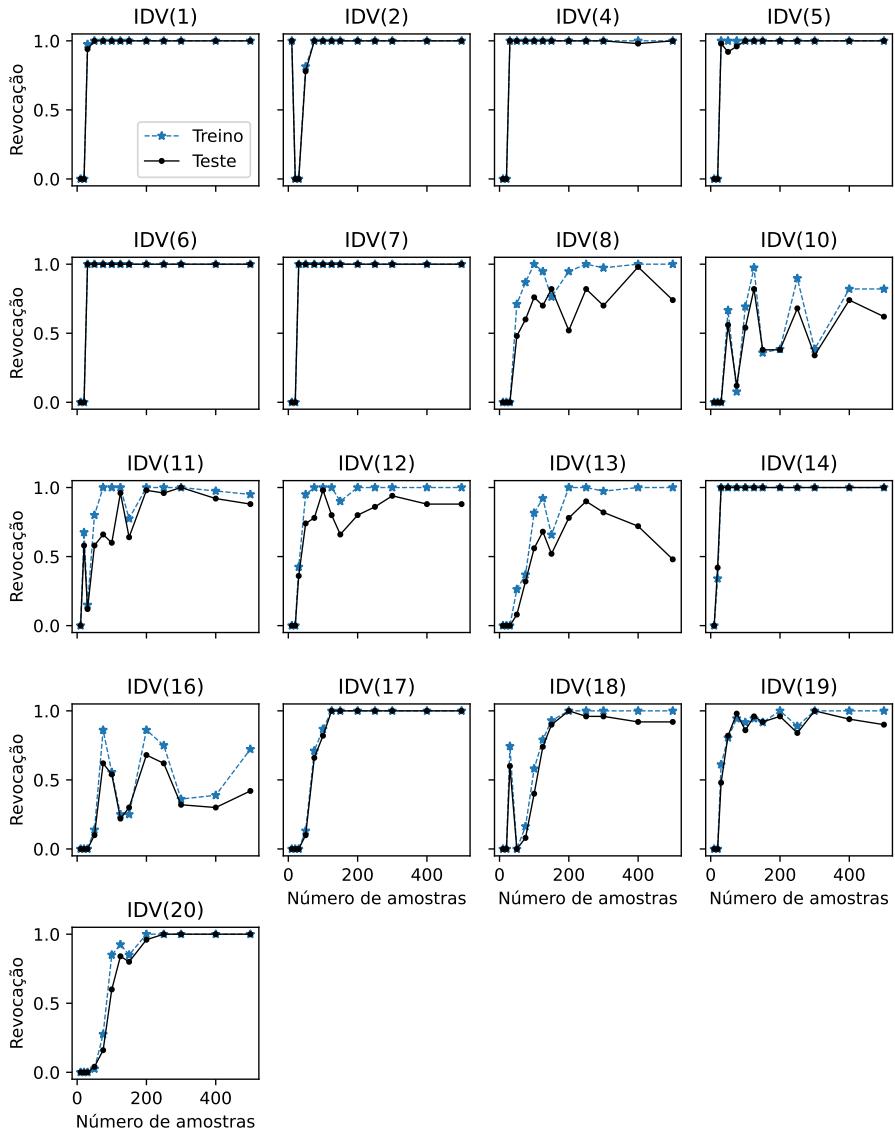


Figura 5.21: Curvas de aprendizado apresentando taxas de verdadeiros positivos em função do tamanho da janela de análise.

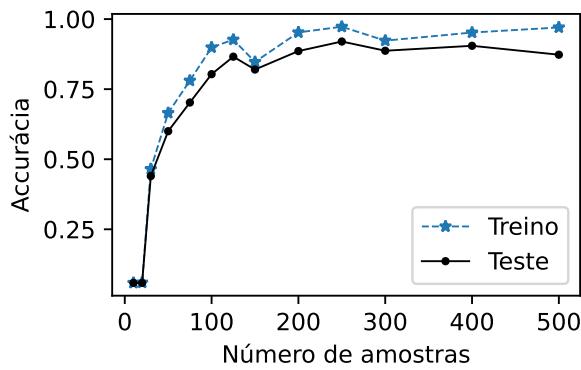


Figura 5.22: Curvas de aprendizado apresentando acuráncias em função do tamanho da janela de análise.

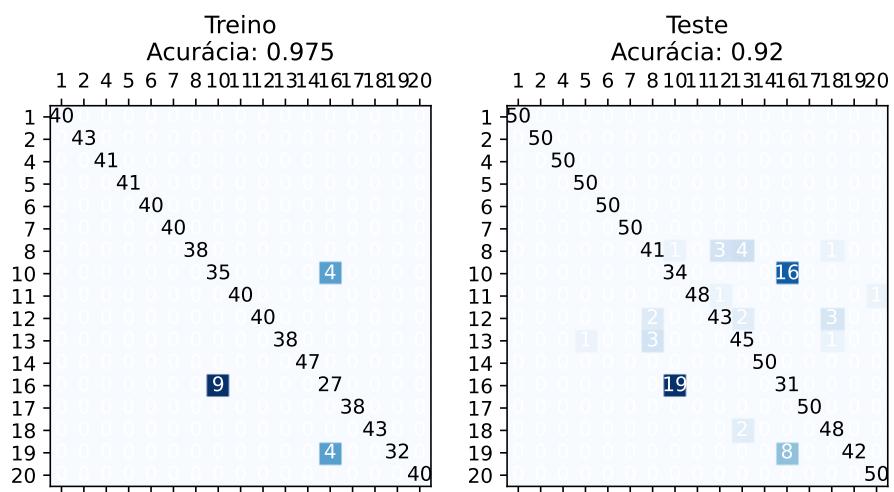


Figura 5.23: Matriz de confusão para análise de diagnóstico com tamanho de janela de 250.

Capítulo 6

Conclusões e perspectivas

O presente trabalho teve como objetivo geral contribuir com metodologias para propiciar uma maior integração humano-máquina durante a aplicação de técnicas de monitoramento de processos industriais baseado em dados. Dado o avanço da utilização da inteligência artificial (IA) no contexto da adoção do paradigma da Indústria 4.0, estudos nesse sentido são importantes para aumentar a compreensão e aceitação das técnicas baseadas em dados na indústria.

É possível que se tenha a ilusão de que aplicar técnicas de IA resulta na completa automatização e consequente exclusão da inteligência humana dos processos produtivos. O que se verifica no presente estágio de desenvolvimento, no entanto, é o contrário: quanto maior a digitalização, maior a necessidade de integrar humanos e máquinas para que o potencial atual da IA seja alcançado em sua plenitude. Ao longo do fluxo de monitoramento de processos industriais, a integração humano-máquina se dá em várias etapas, dentre as quais no presente trabalho foram exploradas: (i) desenvolvimento de *software*, (ii) análise exploratória de dados e (iii) detecção e diagnóstico de falhas. No caso (i), a arquitetura de *software* proposta, baseada no paradigma da orientação a objetos, resulta em códigos mais organizados, modulares e extensíveis, propiciando uma maior integração da pessoa desenvolvedora com a base de implementação. No caso (ii), a metodologia de análise exploratória de dados proposta possibilita um maior conhecimento a respeito dos dados, o que traz como consequência escolhas e interpretações mais racionais e eficientes nas etapas de modelagem e análise em tempo real. No caso (iii), a técnica desenvolvida para detecção e diagnóstico de falhas permite que o operador identifique a natureza de um evento anormal visualizando padrões intrínsecos dos dados representados em um diagrama bidimensional de fácil compreensão e reconhecimento. Todas as metodologias propostas têm o potencial de aumentar o entendimento e aceitação de técnicas baseadas em dados por parte de especialistas e operadores de plantas

industriais.

A seguir lista-se as perspectivas para trabalhos futuros.

- **Desenvolvimento da BibMon:** a biblioteca vem sendo estendida no âmbito de projetos de pesquisa de parceria entre a COPPE (UFRJ) e o CENPES (Petrobras). Em particular, além da adição contínua de modelos e técnicas de pré-processamento, funcionalidades de detecção e diagnóstico também podem ser adicionadas, como alarmes, índices de detecção e cálculos de contribuição. Outra estratégia é ir além de modelos baseados em desvios, implementando-se por exemplo modelos de classificação.
- **Desenvolvimento da KydLIB:** a biblioteca será desenvolvida continuamente como um projeto de código aberto na comunidade GitHub. Há planos para adicionar mais avaliações de análise exploratória, como correlações em diferentes níveis de atraso, estacionariedade, perfis de matrizes, espectros de variância e correlação e adequação à distribuições estatísticas além da gaussiana.
- **Aplicação da análise de padrões de distâncias a outros conjuntos de dados e *benchmarks*.**
- **Comparação e possível integração da análise de padrões de distância com outras técnicas:** outras técnicas descritas na revisão bibliográfica, como os mapas auto-organizáveis e os autocodificadores, são frequentemente utilizadas para criar representações bidimensionais de padrões presentes em conjuntos de dados de alta dimensionalidade. Essas representações podem ser comparadas com os padrões de distância e, potencialmente, integradas e utilizadas de forma conjunta em um determinado contexto.
- **Identificação direta de padrões:** pode-se estender a técnica de análise de padrões de distâncias de modo a desenvolver uma arquitetura de aprendizado profundo especializada na identificação de padrões de distância comumente encontrados em dados de processo, como degraus, tendências e oscilações. Em tese, os modelos resultantes serão capazes de diagnosticar eventos anormais não por seus rótulos individuais, mas pelos padrões genéricos que compõem seu comportamento.
- **Estudos com desenvolvedores e operadores na indústria,** de modo a avaliar os impactos das metodologias propostas durante a atividade de monitoramento.

Referências Bibliográficas

- ABBASI, A. R., 2022, “Fault detection and diagnosis in power transformers: a comprehensive review and classification of publications and methods”, *Electric Power Systems Research*, v. 209 (ago.), pp. 107990. ISSN: 0378-7796. doi: 10.1016/j.epsr.2022.107990. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378779622002176>>.
- ABDI, H., 2010, “Partial least squares regression and projection on latent structure regression (PLS Regression)”, *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 2, n. 1 (jan.), pp. 97–106. ISSN: 19395108. doi: 10.1002/wics.51. Disponível em: <<http://doi.wiley.com/10.1002/wics.51>>.
- ABID, A., KHAN, M. T., IQBAL, J., 2020, “A review on fault detection and diagnosis techniques: basics and beyond”, *Artificial Intelligence Review*, (nov.). ISSN: 0269-2821, 1573-7462. doi: 10.1007/s10462-020-09934-2. Disponível em: <<http://link.springer.com/10.1007/s10462-020-09934-2>>.
- ABONYI, J., 2007, “Application of Exploratory Data Analysis to Historical Process Data of Polyethylene Production”, *Bulgarian Journal of Industrial Chemistry*, v. 35, pp. 85–93.
- ABUBAKAR, A. I., CHIROMA, H., MUAZ, S. A., et al., 2015, “A Review of the Advances in Cyber Security Benchmark Datasets for Evaluating Data-Driven Based Intrusion Detection Systems”, *Procedia Computer Science*, v. 62 (jan.), pp. 221–227. ISSN: 1877-0509. doi: 10.1016/j.procs.2015.08.443. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050915025788>>.
- ACEMOGLU, D., RESTREPO, P., 2018, *Artificial Intelligence, Automation and Work*. Relatório Técnico 24196, National Bureau of Economic Research, jan. Disponível em: <<http://www.nber.org/papers/w24196>>.
- ACITO, F., KHATRI, V., 2014, “Business analytics: Why now and what next?” *Business Horizons*, v. 57, n. 5 (set.), pp. 565–570. ISSN: 00076813. doi:

10.1016/j.bushor.2014.06.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0007681314000871>>.

ADAMS, R. P., MACKAY, D. J. C., 2007, “Bayesian Online Changepoint Detection”, *arXiv:0710.3742 [stat]*, (out.). Disponível em: <<http://arxiv.org/abs/0710.3742>>.

ADEDIGBA, S. A., KHAN, F., YANG, M., 2017, “Dynamic Failure Analysis of Process Systems Using Principal Component Analysis and Bayesian Network”, *Industrial & Engineering Chemistry Research*, v. 56, n. 8 (mar.), pp. 2094–2106. ISSN: 0888-5885. doi: 10.1021/acs.iecr.6b03356. Disponível em: <<https://doi.org/10.1021/acs.iecr.6b03356>>.

ADIL, M., ABID, M., KHAN, A., et al., 2016, “Exponential discriminant analysis for fault diagnosis”, *Neurocomputing*, v. 171 (jan.), pp. 1344–1353. ISSN: 09252312. doi: 10.1016/j.neucom.2015.07.099. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231215011078>>.

AFZAL, M. S., TAN, W., CHEN, T., 2017, “Process Monitoring for Multimodal Processes With Mode-Reachability Constraints”, *IEEE Transactions on Industrial Electronics*, v. 64, n. 5 (maio), pp. 4325–4335. ISSN: 1557-9948. doi: 10.1109/TIE.2017.2677351. Disponível em: <<https://ieeexplore.ieee.org/document/7869281>>.

AGARWAL, P., GONZALEZ, J. I. M., ELKAMEL, A., et al., 2020, “Hierarchical Deep Recurrent Neural Network based Method for Fault Detection and Diagnosis”, *arXiv:2012.03861 [cs]*, (dez.). Disponível em: <<http://arxiv.org/abs/2012.03861>>.

AGARWAL, P., AGHAEE, M., TAMER, M., et al., 2022a, “A novel unsupervised approach for batch process monitoring using deep learning”, *Computers & Chemical Engineering*, v. 159 (mar.), pp. 107694. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.107694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422000382>>.

AGARWAL, P., GONZALEZ, J. I. M., ELKAMEL, A., et al., 2022b, “Hierarchical Deep LSTM for Fault Detection and Diagnosis for a Chemical Process”, *Processes*, v. 10, n. 12 (dez.), pp. 2557. ISSN: 2227-9717. doi: 10.3390/pr10122557. Disponível em: <<https://www.mdpi.com/2227-9717/10/12/2557>>.

AGGARWAL, C. C., 2018, *Neural Networks and Deep Learning: A Textbook*. Cham, Springer International Publishing. ISBN: 9783319944623

9783319944630. doi: 10.1007/978-3-319-94463-0. Disponível em: <<http://link.springer.com/10.1007/978-3-319-94463-0>>.

AGRAHARI, S., SINGH, A. K., 2022, “Concept Drift Detection in Data Stream Mining : A literature review”, *Journal of King Saud University - Computer and Information Sciences*, v. 34, n. 10, Part B (nov.), pp. 9523–9540. ISSN: 1319-1578. doi: 10.1016/j.jksuci.2021.11.006. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1319157821003062>>.

AGUILAR-LOPEZ, R., MATA-MACHUCA, J. L., GODINEZ-CANTILLO, V., 2021, “A TITO Control Strategy to Increase Productivity in Uncertain Exothermic Continuous Chemical Reactors”, *Processes*, v. 9, n. 5 (maio), pp. 873. ISSN: 2227-9717. doi: 10.3390/pr9050873. Disponível em: <<https://www.mdpi.com/2227-9717/9/5/873>>.

AHMAD, A., HAMID, M. K. A., 2001, “Neural Networks for Process Monitoring, Control and Fault Detection: Application to Tennessee Eastman Plant”. In: *Malaysian Science and Technology Congress*, jan. Disponível em: <<http://www.researchgate.net/publication/255635179>>.

AHMAD, S., PURDY, S., 2016, “Real-Time Anomaly Detection for Streaming Analytics”, *arXiv:1607.02480 [cs]*, (jul.). Disponível em: <<http://arxiv.org/abs/1607.02480>>.

AHMAD, S., LAVIN, A., PURDY, S., et al., 2017, “Unsupervised real-time anomaly detection for streaming data”, *Neurocomputing*, v. 262 (nov.), pp. 134–147. ISSN: 09252312. doi: 10.1016/j.neucom.2017.04.070. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231217309864>>.

AHMED, Q., RAZA, S. A., ALANAZI, D. M., 2020, “Reliability based fault analysis models with industrial applications: A systematic literature review”, *Quality and Reliability Engineering International*, (nov.). ISSN: 0748-8017, 1099-1638. doi: 10.1002/qre.2797. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/qre.2797>>.

AKCAY, S., AMELN, D., VAIDYA, A., et al., 2022, “Anomalib: A Deep Learning Library for Anomaly Detection”. In: *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 1706–1710, out. doi: 10.1109/ICIP46576.2022.9897283.

- AL-WAHABI, S. S. S., ABIOLA, S., CHOWDHURY, M. A., et al., 2023, "Improving convolutional neural networks for fault diagnosis in chemical processes by incorporating global correlations", *Computers & Chemical Engineering*, (maio), pp. 108289. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2023.108289. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S009813542300159X>>.
- ALAKENT, B., 2023, "Reducing smearing effect in contribution plots and improving fault detection via polynomial approximated isomap embeddings", *The Canadian Journal of Chemical Engineering*, v. 101, n. 5 (maio), pp. 2768–2789. ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.24689. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.24689>>.
- ALAUDIN, M., KHAN, F., IMTIAZ, S., et al., 2018, "A Bibliometric Review and Analysis of Data-Driven Fault Detection and Diagnosis Methods for Process Systems", *Industrial & Engineering Chemistry Research*, v. 57, n. 32 (ago.), pp. 10719–10735. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.8b00936. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.8b00936>>.
- ALAWI, A., ZHANG, J., MORRIS, J., 2015, "Multiscale Multiblock Batch Monitoring: Sensor and Process Drift and Degradation", *Organic Process Research & Development*, v. 19, n. 1 (jan.), pp. 145–157. ISSN: 1083-6160, 1520-586X. doi: 10.1021/op400337x. Disponível em: <<http://pubs.acs.org/doi/10.1021/op400337x>>.
- ALBAZZAZ, H., WANG, X. Z., 2004, "Statistical Process Control Charts for Batch Operations Based on Independent Component Analysis", *Industrial & Engineering Chemistry Research*, v. 43, n. 21 (out.), pp. 6731–6741. ISSN: 0888-5885. doi: 10.1021/ie049582+. Disponível em: <<https://pubs.acs.org/doi/abs/10.1021/ie049582%2B>>.
- ALBAZZAZ, H., WANG, X. Z., MARHOON, F., 2005, "Multidimensional visualisation for process historical data analysis: a comparative study with multivariate statistical process control", *Journal of Process Control*, v. 15, n. 3 (abr.), pp. 285–294. ISSN: 09591524. doi: 10.1016/j.jprocont.2004.06.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152404000678>>.
- ALBERT, S., MONTAGUE, G., MORRIS, A., et al., 1996, "Multivariate Statistical Process Control in Batch Process Monitoring", *IFAC Proceedings Volumes*, v. 29, n. 1 (jun.), pp. 6708–6713. ISSN: 14746670. doi:

10.1016/S1474-6670(17)58760-1. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017587601>>.

ALBERT, S., KINLEY, R. D., 2001, “Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision”, *Trends in Biotechnology*, v. 19, n. 2 (fev.), pp. 53–62. ISSN: 01677799. doi: 10.1016/S0167-7799(00)01528-6. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0167779900015286>>.

ALCALA, C. F., QIN, S. J., 2011, “Analysis and generalization of fault diagnosis methods for process monitoring”, *Journal of Process Control*, v. 21, n. 3 (mar.), pp. 322–330. ISSN: 09591524. doi: 10.1016/j.jprocont.2010.10.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152410001915>>.

ALCALA, C. F., QIN, S. J., 2009, “Reconstruction-based contribution for process monitoring”, *Automatica*, v. 45, n. 7 (jul.), pp. 1593–1600. ISSN: 00051098. doi: 10.1016/j.automatica.2009.02.027. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0005109809001277>>.

ALDRICH, C., AURET, L., 2010, “Fault detection and diagnosis with random forest feature extraction and variable importance methods”, *IFAC Proceedings Volumes*, v. 43, n. 9, pp. 79–86. ISSN: 14746670. doi: 10.3182/20100802-3-ZA-2014.00020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S147466701632818X>>.

ALDRICH, C., 2019, “Process Fault Diagnosis for Continuous Dynamic Systems Over Multivariate Time Series”. In: *Time Series Analysis [Working Title]*, IntechOpen, abr. doi: 10.5772/intechopen.85456. Disponível em: <<http://doi.org/10.5772/intechopen.85456>>.

ALDRICH, C., AURET, L., 2013, *Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods*. London, Springer London. ISBN: 978-1-4471-5184-5 978-1-4471-5185-2. doi: 10.1007/978-1-4471-5185-2. Disponível em: <<http://link.springer.com/10.1007/978-1-4471-5185-2>>.

ALEX, J., BENEDETTI, L., COPP, J., et al., 2008, *Benchmark Simulation Model no. 1 (BSM1)*. Relatório técnico, IWA Taskgroup on benchmarking of control strategies for WWTPs - Lund University. Disponível em: <http://iwa-mia.org/wp-content/uploads/2019/04/BSM_TG_Tech_Report_no_1_BSM1_General_Description.pdf>.

- ALGHAZZAWI, A., LENNOX, B., 2008, “Monitoring a complex refining process using multivariate statistics”, *Control Engineering Practice*, v. 16, n. 3 (mar.), pp. 294–307. ISSN: 09670661. doi: 10.1016/j.conengprac.2007.04.014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066107001049>>.
- ALIPPI, C., BORACCHI, G., CARRERA, D., et al., 2015, “Change Detection in Multivariate Datastreams: Likelihood and Detectability Loss”, *arXiv:1510.04850 [stat]*, (out.). Disponível em: <<http://arxiv.org/abs/1510.04850>>.
- ALIZADEH, E., KOUJOK, M. E., RAGAB, A., et al., 2018, “A Data-Driven Causality Analysis Tool for Fault Diagnosis in Industrial Processes”, *IFAC-PapersOnLine*, v. 51, n. 24, pp. 147–152. ISSN: 24058963. doi: 10.1016/j.ifacol.2018.09.548. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405896318322390>>.
- ALOBAIDY, M., ABDUL JABBAR, J., AL KHAYYT, S., 2020, “Faults Diagnosis in Robot Systems: A Review”, *Al-Rafidain Engineering Journal (AREJ)*, v. 25, n. 2 (dez.), pp. 166–177. ISSN: 2220-1270. doi: 10.33899/renj.2020.127782.1051. Disponível em: <https://mosuljournals.com/article_167295.html>.
- AMAZOUZ, M., PANTEA, R., 2006, “Use of multivariate data analysis for lumber drying process monitoring and fault detection”. In: *Proceedings of the 2006 International Conference on Data Mining*, Las Vegas, jun. Disponível em: <<https://www.researchgate.net/publication/220705046>>.
- AMERSHI, S., CAKMAK, M., KNOX, W. B., et al., 2014, “Power to the People: The Role of Humans in Interactive Machine Learning”, *AI Magazine*, v. 35, n. 4 (dez.), pp. 105–120. ISSN: 2371-9621, 0738-4602. doi: 10.1609/aimag.v35i4.2513. Disponível em: <<https://ojs.aaai.org/index.php/aimagazine/article/view/2513>>.
- AMIN, M. T., KHAN, F., AHMED, S., et al., 2020, “A novel data-driven methodology for fault detection and dynamic risk assessment”, *The Canadian Journal of Chemical Engineering*, v. 98, n. 11, pp. 2397–2416. ISSN: 1939-019X. doi: 10.1002/cjce.23760. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.23760>>.
- AMINIKHANGHAHI, S., COOK, D. J., 2017, “A survey of methods for time series change point detection”, *Knowledge and Information Systems*, v. 51,

n. 2 (maio), pp. 339–367. ISSN: 0219-1377, 0219-3116. doi: 10.1007/s10115-016-0987-z. Disponível em: <<http://link.springer.com/10.1007/s10115-016-0987-z>>.

ANDRADE, G. M., DE MENEZES, D. Q., SOARES, R. M., et al., 2022, “Virtual flow metering of production flow rates of individual wells in oil and gas platforms through data reconciliation”, *Journal of Petroleum Science and Engineering*, v. 208 (jan.), pp. 109772. ISSN: 09204105. doi: 10.1016/j.petrol.2021.109772. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0920410521013930>>.

ANDRADE, T., GAMA, J., RIBEIRO, R. P., et al., 2019, “Anomaly Detection in Sequential Data: Principles and Case Studies”. In: *Wiley Encyclopedia of Electrical and Electronics Engineering*, John Wiley & Sons, Inc., pp. 1–14, Hoboken, NJ, USA, fev. ISBN: 978-0-471-34608-1. doi: 10.1002/047134608X.W8382. Disponível em: <<http://doi.wiley.com/10.1002/047134608X.W8382>>.

ANGELOV, P., 2014, “Anomaly detection based on eccentricity analysis”. In: *2014 IEEE Symposium on Evolving and Autonomous Learning Systems (EALS)*, pp. 1–8, Orlando, FL, USA, dez. IEEE. ISBN: 978-1-4799-4494-1. doi: 10.1109/EALS.2014.7009497. Disponível em: <<http://ieeexplore.ieee.org/document/7009497/>>.

ANGELOV, P., GU, X., KANGIN, D., et al., 2016, “Empirical data analysis: A new tool for data analytics”. In: *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 000052–000059, Budapest, Hungary, out. IEEE. ISBN: 978-1-5090-1897-0. doi: 10.1109/SMC.2016.7844219. Disponível em: <<http://ieeexplore.ieee.org/document/7844219/>>.

ANGELOV, P., GU, X., KANGIN, D., 2017, “Empirical Data Analytics”, *International Journal of Intelligent Systems*, v. 32, n. 12 (dez.), pp. 1261–1284. ISSN: 08848173. doi: 10.1002/int.21899. Disponível em: <<http://doi.wiley.com/10.1002/int.21899>>.

ANGELOV, P. P., 2013, *Autonomous learning systems: from data streams to knowledge in real-time*. Chichester, West Sussex, United Kingdom, Wiley, a John Wiley & Sons, Ltd., Publication. ISBN: 978-1-119-95152-0.

ANGELOV, P. P., GU, X., 2019, *Empirical Approach to Machine Learning*, v. 800, *Studies in Computational Intelligence*. Cham, Springer International

Publishing. ISBN: 978-3-030-02383-6 978-3-030-02384-3. doi: 10.1007/978-3-030-02384-3. Disponível em: <<http://link.springer.com/10.1007/978-3-030-02384-3>>.

APSEMIDIS, A., PSARAKIS, S., MOGUERZA, J. M., 2020, “A review of machine learning kernel methods in statistical process monitoring”, *Computers & Industrial Engineering*, v. 142 (abr.), pp. 106376. ISSN: 03608352. doi: 10.1016/j.cie.2020.106376. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0360835220301108>>.

ARADHYE, H. B., BAKSHI, B. R., STRAUSS, R. A., et al., 2003, “Multiscale SPC using wavelets: Theoretical analysis and properties”, *AIChE Journal*, v. 49, n. 4 (abr.), pp. 939–958. ISSN: 00011541, 15475905. doi: 10.1002/aic.690490412. Disponível em: <<http://doi.wiley.com/10.1002/aic.690490412>>.

ARIYALURAN HABEEB, R. A., NASARUDDIN, F., GANI, A., et al., 2022, “Clustering-based real-time anomaly detection—A breakthrough in big data technologies”, *Transactions on Emerging Telecommunications Technologies*, v. 33, n. 8 (ago.). ISSN: 2161-3915, 2161-3915. doi: 10.1002/ett.3647. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/ett.3647>>.

ARUNTHAVANATHAN, R., KHAN, F., AHMED, S., et al., 2021, “An analysis of process fault diagnosis methods from safety perspectives”, *Computers & Chemical Engineering*, v. 145 (fev.), pp. 107197. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2020.107197. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135420312400>>.

ASKARIAN, M., ESCUDERO, G., GRAELLS, M., et al., 2016, “Fault diagnosis of chemical processes with incomplete observations: A comparative study”, *Computers & Chemical Engineering*, v. 84 (jan.), pp. 104–116. ISSN: 00981354. doi: 10.1016/j.compchemeng.2015.08.018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135415002793>>.

ASTE, T., TASCA, P., MATTEO, T. D., 2017, “Blockchain Technologies: foreseeable impact on industry and society”, *Computer*, v. 50, n. 9. doi: 10.1109/MC.2017.3571064. Disponível em: <<https://ieeexplore.ieee.org/document/8048633>>.

AURET, L., ALDRICH, C., 2010, “Unsupervised Process Fault Detection with Random Forests”, *Industrial & Engineering Chemistry Research*, v. 49, n. 19

(out.), pp. 9184–9194. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie901975c.
Disponível em: <<https://pubs.acs.org/doi/10.1021/ie901975c>>.

AYESHA, S., HANIF, M. K., TALIB, R., 2020, “Overview and comparative study of dimensionality reduction techniques for high dimensional data”, *Information Fusion*, v. 59 (jul.), pp. 44–58. ISSN: 1566-2535. doi: 10.1016/j.inffus.2020.01.005. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S156625351930377X>>.

AYOUBI, M., ISERMANN, R., 1997, “Neuro-fuzzy systems for diagnosis”, *Fuzzy Sets and Systems*, v. 89, n. 3 (ago.), pp. 289–307. ISSN: 01650114. doi: 10.1016/S0165-0114(97)00011-0. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0165011497000110>>.

BAHDANAU, D., CHO, K., BENGIO, Y., 2016. “Neural machine translation by jointly learning to align and translate”. maio. Disponível em: <<http://arxiv.org/abs/1409.0473>>. arXiv:1409.0473 [cs, stat].

BAI, X., TAN, J., WANG, X., 2018, “Fault Diagnosis and Knowledge Extraction Using Fast Logical Analysis of Data with Multiple Rules Discovery Ability”. In: Shi, Z., Pennartz, C., Huang, T. (Eds.), *2nd International Conference on Intelligence Science (ICIS)*, IFIP Advances in Information and Communication Technology, pp. 412–421, Beijing, China. Springer International Publishing. ISBN: 978-3-030-01313-4. doi: 10.1007/978-3-030-01313-4_44. Disponível em: <<https://hal.inria.fr/hal-02118838/>>.

BAI, Y., ZHAO, J., 2023, “A novel transformer-based multi-variable multi-step prediction method for chemical process fault prognosis”, *Process Safety and Environmental Protection*, v. 169 (jan.), pp. 937–947. ISSN: 0957-5820. doi: 10.1016/j.psep.2022.11.062. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957582022010369>>.

BAJPAI, R. K., REUB, M., 1980, “A mechanistic model for penicillin production”, *Journal of Chemical Technology and Biotechnology*, v. 30, n. 1, pp. 332–344. ISSN: 1935-181X. doi: 10.1002/jctb.503300140. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/jctb.503300140>>.

BAKDI, A., KOUADRI, A., BENSMAIL, A., 2017, “Fault detection and diagnosis in a cement rotary kiln using PCA with EWMA-based adaptive threshold monitoring scheme”, *Control Engineering Practice*, v. 66 (set.), pp. 64–75. ISSN: 09670661. doi: 10.1016/j.conengprac.2017.06.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066117301296>>.

- BAKSALARY, O. M., TRENKLER, G., 2021, “The Moore–Penrose inverse: a hundred years on a frontline of physics research”, *The European Physical Journal H*, v. 46, n. 1 (abr.), pp. 9. ISSN: 2102-6467. doi: 10.1140/epjh/s13129-021-00011-y. Disponível em: <<https://doi.org/10.1140/epjh/s13129-021-00011-y>>.
- BAKSHI, B. R., 1998, “Multiscale PCA with application to multivariate statistical process monitoring”, *AIChe Journal*, v. 44, n. 7 (jul.), pp. 1596–1610. ISSN: 00011541, 15475905. doi: 10.1002/aic.690440712. Disponível em: <<http://doi.wiley.com/10.1002/aic.690440712>>.
- BAKSHI, B. R., LOCHER, G., STEPHANOPOULOS, G., et al., 1994, “Analysis of operating data for evaluation, diagnosis and control of batch operations”, *Journal of Process Control*, v. 4, n. 4 (jan.), pp. 179–194. ISSN: 09591524. doi: 10.1016/0959-1524(94)80040-5. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0959152494800405>>.
- BAKSHI, B., STEPHANOPOULOS, G., 1994a, “Representation of process trends—III. Multiscale extraction of trends from process data”, *Computers & Chemical Engineering*, v. 18, n. 4 (abr.), pp. 267–302. ISSN: 00981354. doi: 10.1016/0098-1354(94)85028-3. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0098135494850283>>.
- BAKSHI, B., STEPHANOPOULOS, G., 1994b, “Representation of process trends—IV. Induction of real-time patterns from operating data for diagnosis and supervisory control”, *Computers & Chemical Engineering*, v. 18, n. 4 (abr.), pp. 303–332. ISSN: 00981354. doi: 10.1016/0098-1354(94)85029-1. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0098135494850291>>.
- BANERJEE, A., ARKUN, Y., 1995, “Control configuration design applied to the Tennessee Eastman plant-wide control problem”, *Computers & Chemical Engineering*, v. 19, n. 4 (abr.), pp. 453–480. ISSN: 0098-1354. doi: 10.1016/0098-1354(94)00058-V. Disponível em: <<https://www.sciencedirect.com/science/article/pii/009813549400058V>>.
- BANNER, M., ALOSSERT, H., SPENCER, C., et al., 2021, “A decade in review: use of data analytics within the biopharmaceutical sector”, *Current Opinion in Chemical Engineering*, v. 34 (dez.), pp. 100758. ISSN: 2211-3398. doi: 10.1016/j.coche.2021.100758. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2211339821000903>>.

- BAO, Y., WANG, B., GUO, P., et al., 2022, “Chemical process fault diagnosis based on a combined deep learning method”, *The Canadian Journal of Chemical Engineering*, v. 100, n. 1 (jan.), pp. 54–66. ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.24153. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.24153>>.
- BARBARIOL, T., CHIARA, F. D., MARCATO, D., et al., 2022, “A Review of Tree-Based Approaches for Anomaly Detection”. In: Tran, K. P. (Ed.), *Control Charts and Machine Learning for Anomaly Detection in Manufacturing*, Springer Series in Reliability Engineering, Springer International Publishing, pp. 149–185, Cham. ISBN: 9783030838195. doi: 10.1007/978-3-030-83819-5_7. Disponível em: <https://doi.org/10.1007/978-3-030-83819-5_7>.
- BARDINAS, J., 2018, *Characterisation of Dynamic Process Systems by Use of Recurrence Texture Analysis*. Tese de Doutorado, Curtin University, Australia, maio. Disponível em: <<https://espace.curtin.edu.au/handle/20.500.11937/73577>>.
- BARDINAS, J., ALDRICH, C., NAPIER, L., 2018, “Predicting the Operating States of Grinding Circuits by Use of Recurrence Texture Analysis of Time Series Data”, *Processes*, v. 6, n. 2 (fev.), pp. 17. ISSN: 2227-9717. doi: 10.3390/pr6020017. Disponível em: <<http://www.mdpi.com/2227-9717/6/2/17>>.
- BARKER, M., RAYENS, W., 2003, “Partial least squares for discrimination”, *Journal of Chemometrics*, v. 17, n. 3 (mar.), pp. 166–173. ISSN: 0886-9383, 1099-128X. doi: 10.1002/cem.785. Disponível em: <<http://doi.wiley.com/10.1002/cem.785>>.
- BARRA, I., MANSOURI, M. A., CHERRAH, Y., et al., 2019, “FTIR fingerprints associated to a PLS-DA model for rapid detection of smuggled non-compliant diesel marketed in Morocco”, *Vibrational Spectroscopy*, v. 101 (mar.), pp. 40–45. ISSN: 09242031. doi: 10.1016/j.vibspec.2019.02.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0924203118302601>>.
- BARRAGAN, J. F., FONTES, C. H., EMBIRUÇU, M., 2016, “A wavelet-based clustering of multivariate time series using a Multiscale SPCA approach”, *Computers & Industrial Engineering*, v. 95 (maio), pp. 144–155. ISSN: 03608352. doi: 10.1016/j.cie.2016.03.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0360835216300560>>.

- BARROS, R. S. M. D., SANTOS, S. G. T. D. C., 2019, “An overview and comprehensive comparison of ensembles for concept drift”, *Information Fusion*, v. 52 (dez.), pp. 213–244. ISSN: 1566-2535. doi: 10.1016/j.inffus.2019.03.006. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1566253518308066>>.
- BARTYS, M., PATTON, R., SYFERT, M., et al., 2006, “Introduction to the DAMADICS actuator FDI benchmark study”, *Control Engineering Practice*, v. 14, n. 6 (jun.), pp. 577–596. ISSN: 09670661. doi: 10.1016/j.conengprac.2005.06.015. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066105001796>>.
- BATHELT, A., RICKER, N. L., JELALI, M., 2015, “Revision of the Tennessee Eastman Process Model”, *IFAC-PapersOnLine*, v. 48, n. 8, pp. 309–314. ISSN: 24058963. doi: 10.1016/j.ifacol.2015.08.199. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405896315010666>>.
- BECK, D. A. C., CAROTHERS, J. M., SUBRAMANIAN, V. R., et al., 2016, “Data science: Accelerating innovation and discovery in chemical engineering”, *AIChe Journal*, v. 62, n. 5 (maio), pp. 1402–1416. ISSN: 00011541. doi: 10.1002/aic.15192. Disponível em: <<http://doi.wiley.com/10.1002/aic.15192>>.
- BERGH, L. G., 2010, “Restrictions on MMM Industrial Data to Build PCA Models”, *IFAC Proceedings Volumes*, v. 43, n. 9 (jan.), pp. 98–103. ISSN: 1474-6670. doi: 10.3182/20100802-3-ZA-2014.00024. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S147466701632821X>>.
- BERLEMANN, M., FREESE, J., KNOTH, S., 2019, “Dating the start of the US house price bubble: an application of statistical process control”, *Empirical Economics*, (mar.). ISSN: 0377-7332, 1435-8921. doi: 10.1007/s00181-019-01648-x. Disponível em: <<http://link.springer.com/10.1007/s00181-019-01648-x>>.
- BERNAL, D. E., AJAGEKAR, A., HARWOOD, S. M., et al., 2022, “Perspectives of quantum computing for chemical engineering”, *AIChe Journal*, v. 68, n. 6 (jun.). ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.17651. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/aic.17651>>.
- BERNAL-DE LÁZARO, J. M., CRUZ-CORONA, C., SILVA-NETO, A. J., et al., 2022, “Criteria for optimizing kernel methods in fault monitoring process:

- A survey”, *ISA Transactions*, v. 127 (ago.), pp. 259–272. ISSN: 0019-0578. doi: 10.1016/j.isatra.2021.08.040. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0019057821004687>>.
- BERNDT, D. J., CLIFFORD, J., 1994, “Using Dynamic Time Warping to Find Patterns in Time Series”. In: *AAAIWS’94: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, pp. 359–370. Disponível em: <<https://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-031.pdf>>.
- BERRIOS, R., PAREDES, R., CIPRIANO, A., 2009, “Sensor fault detection and diagnosis in the gas turbine of San Isidro combined cycle power plant”, *IFAC Proceedings Volumes*, v. 42, n. 8, pp. 1354–1359. ISSN: 14746670. doi: 10.3182/20090630-4-ES-2003.00220. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667016359638>>.
- BEZERRA, C. G., COSTA, B. S. J., GUEDES, L. A., et al., 2016, “An evolving approach to unsupervised and Real-Time fault detection in industrial processes”, *Expert Systems with Applications*, v. 63 (nov.), pp. 134–144. ISSN: 09574174. doi: 10.1016/j.eswa.2016.06.035. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417416303153>>.
- BI, X., QIN, R., WU, D., et al., 2022, “One step forward for smart chemical process fault detection and diagnosis”, *Computers & Chemical Engineering*, v. 164 (ago.), pp. 107884. ISSN: 00981354. doi: 10.1016/j.compchemeng.2022.107884. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135422002228>>.
- BI, X., WU, D., XIE, D., et al., 2023, “Large-scale chemical process causal discovery from big data with transformer-based deep learning”, *Process Safety and Environmental Protection*, v. 173 (maio), pp. 163–177. ISSN: 0957-5820. doi: 10.1016/j.psep.2023.03.017. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957582023002100>>.
- BIKMUKHAMETOV, T., JÄSCHKE, J., 2020, “Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models”, *Computers & Chemical Engineering*, v. 138 (jul.), pp. 106834. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106834. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419313675>>.
- BIROL, G., ÜNDEY, C., ÇINAR, A., 2002, “A modular simulation package for fed-batch fermentation: penicillin production”, *Computers & Chemical*

Engineering, v. 26, n. 11 (nov.), pp. 1553–1565. ISSN: 00981354. doi: 10.1016/S0098-1354(02)00127-8. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135402001278>>.

BISSESSUR, Y., MARTIN, E., MORRIS, A., 1999, “Monitoring the performance of the paper making process”, *Control Engineering Practice*, v. 7, n. 11 (nov.), pp. 1357–1368. ISSN: 09670661. doi: 10.1016/S0967-0661(99)00105-7. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066199001057>>.

BLEI, D. M., 2014, “Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models”, *Annual Review of Statistics and Its Application*, v. 1, n. 1 (jan.), pp. 203–232. ISSN: 2326-8298, 2326-831X. doi: 10.1146/annurev-statistics-022513-115657. Disponível em: <<https://www.annualreviews.org/doi/10.1146/annurev-statistics-022513-115657>>.

BLÁZQUEZ-GARCÍA, A., CONDE, A., MORI, U., et al., 2022, “A Review on Outlier/Anomaly Detection in Time Series Data”, *ACM Computing Surveys*, v. 54, n. 3 (abr.), pp. 1–33. ISSN: 0360-0300, 1557-7341. doi: 10.1145/3444690. Disponível em: <<https://dl.acm.org/doi/10.1145/3444690>>.

BOLDT, F. D. A., RAUBER, T. W., VAREJÃO, F. M., 2017, “Cascade Feature Selection and ELM for automatic fault diagnosis of the Tennessee Eastman process”, *Neurocomputing*, v. 239 (maio), pp. 238–248. ISSN: 09252312. doi: 10.1016/j.neucom.2017.02.025. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231217302953>>.

BOOCH, G., RUMBAUGH, J., JACOBSON, I., 2005, *The Unified Modeling Language User Guide*. 2nd edition ed. Upper Saddle River, NJ, Addison-Wesley Professional. ISBN: 978-0-321-26797-9.

BORGA, M., LANDELIUS, T., KNUTSSON, H., 1992, “A Unified Approach to PCA, PLS, MLR and CCA”, Disponível em: <<https://www.researchgate.net/publication/2305978>>.

BOSER, B. E., GUYON, I. M., VAPNIK, V. N., 1992, “A training algorithm for optimal margin classifiers”. In: *COLT '92: Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152. ACM, jul. ISBN: 978-0-89791-497-0. doi: 10.1145/130385.130401. Disponível em: <<https://dl.acm.org/citation.cfm?id=130401>>.

- BOTRE, C., MANSOURI, M., KARIM, M. N., et al., 2017, “Multiscale PLS-based GLRT for fault detection of chemical processes”, *Journal of Loss Prevention in the Process Industries*, v. 46 (mar.), pp. 143–153. ISSN: 09504230. doi: 10.1016/j.jlp.2017.01.008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0950423017300268>>.
- BOUGUELIA, M.-R., NOWACZYK, S., PAYBERAH, A. H., 2018, “An adaptive algorithm for anomaly and novelty detection in evolving data streams”, *Data Mining and Knowledge Discovery*, v. 32, n. 6 (nov.), pp. 1597–1633. ISSN: 1384-5810, 1573-756X. doi: 10.1007/s10618-018-0571-0. Disponível em: <<http://link.springer.com/10.1007/s10618-018-0571-0>>.
- BOUSDEKIS, A., APOSTOLOU, D., MENTZAS, G., 2020, “A human cyber physical system framework for operator 4.0 – artificial intelligence symbiosis”, *Manufacturing Letters*, v. 25 (ago.), pp. 10–15. ISSN: 2213-8463. doi: 10.1016/j.mfglet.2020.06.001. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2213846320301279>>.
- BOUWMANS, T., ZAHZAH, E. H., 2014, “Robust PCA via Principal Component Pursuit: A review for a comparative evaluation in video surveillance”, *Computer Vision and Image Understanding*, v. 122 (maio), pp. 22–34. ISSN: 10773142. doi: 10.1016/j.cviu.2013.11.009. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1077314213002294>>.
- BOX, G. E. P., JENKINS, G. M., REINSEL, G. C., et al., 2015, *Time Series Analysis: Forecasting and Control*. 5 edition ed. Hoboken, New Jersey, Wiley. ISBN: 978-1-118-67502-1.
- BRADLEY, W., KIM, J., KILWEIN, Z., et al., 2022, “Perspectives on the integration between first-principles and data-driven modeling”, *Computers & Chemical Engineering*, v. 166 (out.), pp. 107898. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.107898. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422002368>>.
- BROPHY, E., WANG, Z., SHE, Q., et al., 2023, “Generative Adversarial Networks in Time Series: A Systematic Literature Review”, *ACM Computing Surveys*, v. 55, n. 10 (fev.), pp. 199:1–199:31. ISSN: 0360-0300. doi: 10.1145/3559540. Disponível em: <<https://dl.acm.org/doi/10.1145/3559540>>.
- CABRIETO, J., TUERLINCKX, F., KUPPENS, P., et al., 2017, “Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods”, *Behavior Research*

Methods, v. 49, n. 3 (jun.), pp. 988–1005. ISSN: 1554-3528. doi: 10.3758/s13428-016-0754-9. Disponível em: <<http://link.springer.com/10.3758/s13428-016-0754-9>>.

CABRIETO, J., TUERLINCKX, F., KUPPENS, P., et al., 2018a, “Testing for the Presence of Correlation Changes in a Multivariate Time Series: A Permutation Based Approach”, *Scientific Reports*, v. 8, n. 1 (dez.). ISSN: 2045-2322. doi: 10.1038/s41598-017-19067-2. Disponível em: <<http://www.nature.com/articles/s41598-017-19067-2>>.

CABRIETO, J., TUERLINCKX, F., KUPPENS, P., et al., 2018b, “Capturing correlation changes by applying kernel change point detection on the running correlations”, *Information Sciences*, v. 447 (jun.), pp. 117–139. ISSN: 00200255. doi: 10.1016/j.ins.2018.03.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0020025516316929>>.

CACCIARELLI, D., KULAHCI, M., 2022, “A novel fault detection and diagnosis approach based on orthogonal autoencoders”, *Computers & Chemical Engineering*, v. 163 (jul.), pp. 107853. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.107853. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422001910>>.

CAI, Z., SUN, S., SI, S., et al., 2009, “Research of failure prediction Bayesian network model”. In: *2009 16th International Conference on Industrial Engineering and Engineering Management*, pp. 2021–2025, out. doi: 10.1109/ICIEEM.2009.5344265. Disponível em: <<http://ieeexplore.ieee.org/document/5344265>>.

CAMACHO, J., PICÓ, J., FERRER, A., 2009, “The best approaches in the on-line monitoring of batch processes based on PCA: Does the modelling structure matter?” *Analytica Chimica Acta*, v. 642, n. 1 (maio), pp. 59–68. ISSN: 0003-2670. doi: 10.1016/j.aca.2009.02.001. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0003267009002189>>.

CAMACHO, J., PÉREZ-VILLEGAS, A., RODRÍGUEZ-GÓMEZ, R. A., et al., 2015, “Multivariate Exploratory Data Analysis (MEDA) Toolbox for Matlab”, *Chemometrics and Intelligent Laboratory Systems*, v. 143 (abr.), pp. 49–57. ISSN: 01697439. doi: 10.1016/j.chemolab.2015.02.016. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743915000465>>.

CANO, A., KRAWCZYK, B., 2020, “Kappa Updated Ensemble for drifting data stream mining”, *Machine Learning*, v. 109, n. 1 (jan.), pp. 175–218. ISSN:

1573-0565. doi: 10.1007/s10994-019-05840-z. Disponível em: <<https://doi.org/10.1007/s10994-019-05840-z>>.

CAO, Y., YUAN, X., WANG, Y., et al., 2021, “Hierarchical hybrid distributed PCA for plant-wide monitoring of chemical processes”, *Control Engineering Practice*, v. 111 (jun.), pp. 104784. ISSN: 0967-0661. doi: 10.1016/j.conengprac.2021.104784. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0967066121000617>>.

CAPACI, F., VANHATALO, E., KULAHCI, M., et al., 2019, “The revised Tennessee Eastman process simulator as testbed for SPC and DoE methods”, *Quality Engineering*, v. 31, n. 2 (abr.), pp. 212–229. ISSN: 0898-2112, 1532-4222. doi: 10.1080/08982112.2018.1461905. Disponível em: <<https://tandfonline.com/doi/full/10.1080/08982112.2018.1461905>>.

CARLETTI, M., TERZI, M., SUSTO, G. A., 2023, “Interpretable Anomaly Detection with DIFFI: Depth-based feature importance of Isolation Forest”, *Engineering Applications of Artificial Intelligence*, v. 119 (mar.), pp. 105730. ISSN: 0952-1976. doi: 10.1016/j.engappai.2022.105730. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197622007205>>.

CARTER, A., IMTIAZ, S., NATERER, G. F., 2023, “Review of interpretable machine learning for process industries”, *Process Safety and Environmental Protection*, v. 170 (fev.), pp. 647–659. ISSN: 0957-5820. doi: 10.1016/j.psep.2022.12.018. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957582022010837>>.

CAVALCANTI, F. M., KOZONOE, C. E., PACHECO, K. A., et al., 2021, “Application of Artificial Neural Networks to Chemical and Process Engineering”, *Artificial Neural Networks and Deep Learning - Applications and Perspective*, (mar.). doi: 10.5772/intechopen.96641. Disponível em: <<https://www.intechopen.com/online-first/application-of-artificial-neural-networks-to-chemical-and-process-engineering>>.

CAÑAS, H., MULA, J., DÍAZ-MADROÑERO, M., et al., 2021, “Implementing Industry 4.0 principles”, *Computers & Industrial Engineering*, v. 158 (ago.), pp. 107379. ISSN: 0360-8352. doi: 10.1016/j.cie.2021.107379. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360835221002837>>.

CEJNEK, M., BUKOVSKY, I., 2018, “Concept drift robust adaptive novelty detection for data streams”, *Neurocomputing*, v. 309 (out.), pp. 46–53. ISSN:

09252312. doi: 10.1016/j.neucom.2018.04.069. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231218305253>>.

CHAI, Z., ZHAO, C., 2020, “A Fine-Grained Adversarial Network Method for Cross-Domain Industrial Fault Diagnosis”, *IEEE Transactions on Automation Science and Engineering*, v. 17, n. 3 (jul.), pp. 1432–1442. ISSN: 1545-5955, 1558-3783. doi: 10.1109/TASE.2019.2957232. Disponível em: <<https://ieeexplore.ieee.org/document/8950281/>>.

CHALAPATHY, R., CHAWLA, S., 2019, “Deep Learning for Anomaly Detection: A Survey”, *arXiv:1901.03407 [cs, stat]*, (jan.). Disponível em: <<http://arxiv.org/abs/1901.03407>>.

CHARBONNIER, S., PORTET, F., 2012, “A self-tuning adaptive trend extraction method for process monitoring and diagnosis”, *Journal of Process Control*, v. 22, n. 6 (jul.), pp. 1127–1138. ISSN: 09591524. doi: 10.1016/j.jprocont.2012.03.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152412000753>>.

CHARTE, D., CHARTE, F., DEL JESUS, M. J., et al., 2020, “An analysis on the use of autoencoders for representation learning: Fundamentals, learning task case studies, explainability and challenges”, *Neurocomputing*, v. 404 (set.), pp. 93–107. ISSN: 09252312. doi: 10.1016/j.neucom.2020.04.057. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S092523122030624X>>.

CHEN, A., ZHOU, H., AN, Y., et al., 2016, “PCA and PLS monitoring approaches for fault detection of wastewater treatment process”. In: *2016 IEEE 25th International Symposium on Industrial Electronics (ISIE)*, pp. 1022–1027, jun. doi: 10.1109/ISIE.2016.7745032.

CHEN, G., GE, Z., 2020, “Robust Bayesian networks for low-quality data modeling and process monitoring applications”, *Control Engineering Practice*, v. 97 (abr.), pp. 104344. ISSN: 09670661. doi: 10.1016/j.conengprac.2020.104344. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066120300289>>.

CHEN, J. X., 2016, “The Evolution of Computing: AlphaGo”, *Computing in Science & Engineering*, v. 18, n. 4 (jul.), pp. 4–7. ISSN: 1521-9615. doi: 10.1109/MCSE.2016.74. Disponível em: <<http://ieeexplore.ieee.org/document/7499782/>>.

- CHEN, J., YU, J., 2014, “Independent Component Analysis Mixture Model Based Dissimilarity Method for Performance Monitoring of Non-Gaussian Dynamic Processes with Shifting Operating Conditions”, *Industrial & Engineering Chemistry Research*, v. 53, n. 13 (abr.), pp. 5055–5066. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie401027b. Disponível em: <<http://pubs.acs.org/doi/10.1021/ie401027b>>.
- CHEN, J., LIU, K.-C., 2002, “On-line batch process monitoring using dynamic PCA and dynamic PLS models”, *Chemical Engineering Science*, v. 57, n. 1 (jan.), pp. 63–75. ISSN: 00092509. doi: 10.1016/S0009-2509(01)00366-9. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0009250901003669>>.
- CHEN, Q., WANG, Y., 2021, “Key performance indicator related state monitoring based on kernel canonical correlation analysis”, *Control Engineering Practice*, v. 107 (fev.), pp. 104692. ISSN: 09670661. doi: 10.1016/j.conengprac.2020.104692. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066120302628>>.
- CHEN, S., WU, Z., CHRISTOFIDES, P. D., 2020a, “Cyber-attack detection and resilient operation of nonlinear processes under economic model predictive control”, *Computers & Chemical Engineering*, v. 136 (maio), pp. 106806. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106806. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135420300855>>.
- CHEN, T., ZHANG, J., 2010, “On-line multivariate statistical monitoring of batch processes using Gaussian mixture model”, *Computers & Chemical Engineering*, v. 34, n. 4 (abr.), pp. 500–507. ISSN: 00981354. doi: 10.1016/j.compchemeng.2009.08.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S009813540900218X>>.
- CHEN, T., MORRIS, J., MARTIN, E., 2006, “Probability density estimation via an infinite Gaussian mixture model: application to statistical process monitoring”, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, v. 55, n. 5 (nov.), pp. 699–715. ISSN: 0035-9254, 1467-9876. doi: 10.1111/j.1467-9876.2006.00560.x. Disponível em: <<http://doi.wiley.com/10.1111/j.1467-9876.2006.00560.x>>.
- CHEN, X., WANG, J., DING, S. X., 2020b, “Complex System Monitoring Based on Distributed Least Squares Method”, *IEEE Transactions on Automation Science and Engineering*, pp. 1–9. ISSN: 1545-5955, 1558-3783. doi: 10.

1109/TASE.2020.3022924. Disponível em: <<https://ieeexplore.ieee.org/document/9205200/>>.

CHEN, X., YAN, X., 2012, “Using improved self-organizing map for fault diagnosis in chemical industry process”, *Chemical Engineering Research and Design*, v. 90, n. 12 (dez.), pp. 2262–2277. ISSN: 0263-8762. doi: 10.1016/j.cherd.2012.06.004. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263876212002493>>.

CHEN, Y., WANG, T., SAMWORTH, R. J., 2022, “High-Dimensional, Multiscale Online Changepoint Detection”, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, v. 84, n. 1 (fev.), pp. 234–266. ISSN: 1369-7412, 1467-9868. doi: 10.1111/rssb.12447. Disponível em: <<https://academic.oup.com/jrsssb/article/84/1/234/7056123>>.

CHEN, Z., 2017, *Data-Driven Fault Detection for Industrial Processes*. Wiesbaden, Springer Fachmedien Wiesbaden. ISBN: 978-3-658-16755-4 978-3-658-16756-1. doi: 10.1007/978-3-658-16756-1. Disponível em: <<http://link.springer.com/10.1007/978-3-658-16756-1>>.

CHEN, Z., LIU, C., DING, S. X., et al., 2021, “A Just In Time Learning Aided Canonical Correlation Analysis Method for Multimode Process Monitoring and Fault Detection”, *IEEE Transactions on Industrial Electronics*, v. 68, n. 6, pp. 12. doi: 10.1109/TIE.2020.2989708.

CHENG, H., WU, J., HUANG, D., et al., 2021, “Robust adaptive boosted canonical correlation analysis for quality-relevant process monitoring of wastewater treatment”, *ISA Transactions*, v. 117 (nov.), pp. 210–220. ISSN: 0019-0578. doi: 10.1016/j.isatra.2021.01.039. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0019057821000495>>.

CHENG, J., DONG, L., LAPATA, M., 2016. “Long Short-Term Memory-Networks for Machine Reading”. set. Disponível em: <<http://arxiv.org/abs/1601.06733>>. arXiv:1601.06733 [cs].

CHEUNG, J. T. Y., STEPHANOPOULOS, G., 1990a, “Representation of process trends—Part I. A formal representation framework”, *Computers & Chemical Engineering*, v. 14, n. 4 (maio), pp. 495–510. ISSN: 0098-1354. doi: 10.1016/0098-1354(90)87023-I. Disponível em: <<http://www.sciencedirect.com/science/article/pii/009813549087023I>>.

CHEUNG, J. T. Y., STEPHANOPOULOS, G., 1990b, “Representation of process trends—Part II. The problem of scale and qualitative scaling”, *Computers*

& Chemical Engineering, v. 14, n. 4 (maio), pp. 511–539. ISSN: 0098-1354. doi: 10.1016/0098-1354(90)87024-J. Disponível em: <<http://www.sciencedirect.com/science/article/pii/009813549087024J>>.

CHIANG, L. H., RUSSELL, E. L., BRAATZ, R. D., 2001, *Fault Detection and Diagnosis in Industrial Systems*. Softcover reprint of the original 1st ed. 2001 edition ed. London ; New York, Springer. ISBN: 978-1-85233-327-0.

CHIANG, L., LU, B., CASTILLO, I., 2017, “Big Data Analytics in Chemical Engineering”, *Annual Review of Chemical and Biomolecular Engineering*, v. 8, n. 1 (jun.), pp. 63–85. ISSN: 1947-5438, 1947-5446. doi: 10.1146/annurev-chembioeng-060816-101555. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev-chembioeng-060816-101555>>.

CHIANG, L. H., BRAATZ, R. D., 2003, “Process monitoring using causal map and multivariate statistics: fault detection and identification”, *Chemometrics and Intelligent Laboratory Systems*, v. 65, n. 2 (fev.), pp. 159–178. ISSN: 01697439. doi: 10.1016/S0169-7439(02)00140-5. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743902001405>>.

CHIANG, L. H., COLEGROVE, L. F., 2007, “Industrial implementation of on-line multivariate quality control”, *Chemometrics and Intelligent Laboratory Systems*, v. 88, n. 2 (set.), pp. 143–153. ISSN: 01697439. doi: 10.1016/j.chemolab.2007.02.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743907000445>>.

CHIANG, L. H., KOTANCHEK, M. E., KORDON, A. K., 2004, “Fault diagnosis based on Fisher discriminant analysis and support vector machines”, *Computers & Chemical Engineering*, v. 28, n. 8 (jul.), pp. 1389–1401. ISSN: 00981354. doi: 10.1016/j.compchemeng.2003.10.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135403002618>>.

CHIANG, L. H., BRAUN, B., WANG, Z., et al., 2022, “Towards artificial intelligence at scale in the chemical industry”, *AICHE Journal*, v. 68, n. 6 (jun.). ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.17644. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/aic.17644>>.

CHIANG, R. H., GROVER, V., LIANG, T.-P., et al., 2018, “Special Issue: Strategic Value of Big Data and Business Analytics”, *Journal of Management*

Information Systems, v. 35, n. 2 (abr.), pp. 383–387. ISSN: 0742-1222, 1557-928X. doi: 10.1080/07421222.2018.1451950. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/07421222.2018.1451950>>.

CHIOUA, M., BAUER, M., CHEN, S.-L., et al., 2016, “Plant-wide root cause identification using plant key performance indicators (KPIs) with application to a paper machine”, *Control Engineering Practice*, v. 49 (abr.), pp. 149–158. ISSN: 09670661. doi: 10.1016/j.conengprac.2015.10.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066115300381>>.

CHO, H.-W., 2007, “Identification of contributing variables using kernel-based discriminant modeling and reconstruction”, *Expert Systems with Applications*, v. 33, n. 2 (ago.), pp. 274–285. ISSN: 09574174. doi: 10.1016/j.eswa.2006.05.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417406001412>>.

CHOI, S. W., LEE, I.-B., 2005, “Multiblock PLS-based localized process diagnosis”, *Journal of Process Control*, v. 15, n. 3 (abr.), pp. 295–306. ISSN: 09591524. doi: 10.1016/j.jprocont.2004.06.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095915240400085X>>.

CHOI, S. W., PARK, J. H., LEE, I.-B., 2004, “Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis”, *Computers & Chemical Engineering*, v. 28, n. 8 (jul.), pp. 1377–1387. ISSN: 00981354. doi: 10.1016/j.compchemeng.2003.09.031. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135403002606>>.

CHOI, S. W., LEE, C., LEE, J.-M., et al., 2005, “Fault detection and identification of nonlinear processes based on kernel PCA”, *Chemometrics and Intelligent Laboratory Systems*, v. 75, n. 1 (jan.), pp. 55–67. ISSN: 01697439. doi: 10.1016/j.chemolab.2004.05.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743904001224>>.

CHOLLET, F., 2021, *Deep Learning with Python*. 2 ed. Shelter Island, Manning Publications. ISBN: 978-1-61729-686-4.

CHOWDHARY, K. R., 2020, “Statistical Learning Theory”. In: Chowdhary, K. (Ed.), *Fundamentals of Artificial Intelligence*, Springer India, pp. 415–443, New Delhi. ISBN: 9788132239727. doi: 10.1007/978-81-322-3972-7_14. Disponível em: <https://doi.org/10.1007/978-81-322-3972-7_14>.

- CHRISLB, 2005. “Diagram of an artificial neuron.” jul.
- CHU, F., MO, S., HAO, L., et al., 2022, “Operating Performance Assessment of Complex Nonlinear Industrial Process Based on Kernel Locally Linear Embedding PLS”, *IEEE Transactions on Automation Science and Engineering*, pp. 1–13. ISSN: 1558-3783. doi: 10.1109/TASE.2022.3230687.
- CLAVIJO, N., MELO, A., CÂMARA, M. M., et al., 2019, “Development and Application of a Data-Driven System for Sensor Fault Diagnosis in an Oil Processing Plant”, *Processes*, v. 7, n. 436 (jul.). doi: 10.3390/pr7070436. Disponível em: <<https://www.mdpi.com/2227-9717/7/7/436>>.
- CLAVIJO, N., MELO, A., SOARES, R. M., et al., 2021, “Variable Selection for Fault Detection Based on Causal Discovery Methods: Analysis of an Actual Industrial Case”, *Processes*, v. 9, n. 3 (mar.), pp. 544. doi: 10.3390/pr9030544. Disponível em: <<https://www.mdpi.com/2227-9717/9/3/544>>.
- CLAVIJO, N. A., DIEHL, F. C., GEREVINI, G. G., et al., 2022a, “Chaotic dynamics induced by anti-slug control in offshore oil production plants”, *Journal of Petroleum Science and Engineering*, v. 215 (ago.), pp. 110716. ISSN: 0920-4105. doi: 10.1016/j.petrol.2022.110716. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0920410522005812>>.
- CLAVIJO, N. A., PINTO, J. C., MELO, P. A., 2022b, “Continuous diabatic free-radical solution polymerization reactors: Search engines for nonlinear dynamical solutions”, *Chemical Engineering Science*, v. 248 (fev.), pp. 117221. ISSN: 0009-2509. doi: 10.1016/j.ces.2021.117221. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0009250921007867>>.
- CLAVIJO, N. A., 2017, *Análise e Caracterização de Dinâmica Caótica em Reatores Contínuos de Polimerização via Radicais Livres em Solução*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- CLIFTON, D. A., HUGUENY, S., TARASSENKO, L., 2011, “Novelty Detection with Multivariate Extreme Value Statistics”, *Journal of Signal Processing Systems*, v. 65, n. 3 (dez.), pp. 371–389. ISSN: 1939-8018, 1939-8115. doi: 10.1007/s11265-010-0513-6. Disponível em: <<http://link.springer.com/10.1007/s11265-010-0513-6>>.
- COHEN, A., ATOUI, M. A., 2020, “On wavelet based statistical process monitoring”, *Transactions of the Institute of Measurement and Control*,

(jul.), pp. 014233122093570. ISSN: 0142-3312, 1477-0369. doi: 10.1177/0142331220935708. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0142331220935708>>.

COHEN, M. X., 2021, *Linear Algebra: Theory, Intuition, Code*. Amsterdam, SincXpress BV. ISBN: 978-90-831366-0-8.

COHEN, Y., GOLAN, M., SINGER, G., et al., 2018, “Workstation-Operator Interaction in 4.0 Era: WOI 4.0”, *IFAC-PapersOnLine*, v. 51, n. 11 (jan.), pp. 399–404. ISSN: 2405-8963. doi: 10.1016/j.ifacol.2018.08.327. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2405896318314514>>.

COLUCCI, D., PRATS-MONTALBÁN, J. M., FISSORE, D., et al., 2019, “Application of multivariate image analysis for on-line monitoring of a freeze-drying process for pharmaceutical products in vials”, *Chemometrics and Intelligent Laboratory Systems*, v. 187 (abr.), pp. 19–27. ISSN: 01697439. doi: 10.1016/j.chemolab.2019.02.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S016974391830501X>>.

COMON, P., 1994, “Independent component analysis, A new concept?” *Signal Processing*, v. 36, n. 3 (abr.), pp. 287–314. ISSN: 01651684. doi: 10.1016/0165-1684(94)90029-9. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0165168494900299>>.

CONWAY, D., 2010. “The Data Science Venn Diagram”. Disponível em: <<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>>.

COSTA, B. S. J., ANGELOV, P. P., GUEDES, L. A., 2014, “Real-Time Fault Detection Using Recursive Density Estimation”, *Journal of Control, Automation and Electrical Systems*, v. 25, n. 4 (ago.), pp. 428–437. ISSN: 2195-3880, 2195-3899. doi: 10.1007/s40313-014-0128-4. Disponível em: <<http://link.springer.com/10.1007/s40313-014-0128-4>>.

COSTA, B. S. J., ANGELOV, P. P., GUEDES, L. A., 2015a, “Fully unsupervised fault detection and identification based on recursive density estimation and self-evolving cloud-based classifier”, *Neurocomputing*, v. 150 (fev.), pp. 289–303. ISSN: 09252312. doi: 10.1016/j.neucom.2014.05.086. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231214013174>>.

COSTA, B. S. J., BEZERRA, C. G., GUEDES, L. A., et al., 2015b, “Online fault detection based on Typicality and Eccentricity Data Analytics”.

In: *2015 International Joint Conference on Neural Networks (IJCNN)*, Killarney, Ireland, jul.b. IEEE. ISBN: 978-1-4799-1960-4. doi: 10.1109/IJCNN.2015.7280712. Disponível em: <<http://ieeexplore.ieee.org/document/7280712/>>.

COVER, T. M., 1965, “Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition”, *IEEE Transactions on Electronic Computers*, v. EC-14, n. 3 (jun.), pp. 326–334. ISSN: 0367-7508. doi: 10.1109/PGEC.1965.264137. Disponível em: <<http://ieeexplore.ieee.org/document/4038449/>>.

CUI, P., WANG, X., YANG, Y., 2022, “Nonparametric manifold learning approach for improved process monitoring”, *The Canadian Journal of Chemical Engineering*, v. 100, n. 1 (jan.), pp. 67–89. ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.24066. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.24066>>.

CURRERI, F., FIUMARA, G., XIBILIA, M. G., 2020, “Input Selection Methods for Soft Sensor Design: A Survey”, *Future Internet*, v. 12, n. 6 (jun.), pp. 97. ISSN: 1999-5903. doi: 10.3390/fi12060097. Disponível em: <<https://www.mdpi.com/1999-5903/12/6/97>>.

CZAJKOWSKI, A., PATAN, K., 2016, “Robust Fault Detection by Means of Echo State Neural Network”. In: Kowalcuk, Z. (Ed.), *Advanced and Intelligent Computations in Diagnosis and Control*, v. 386, Springer International Publishing, pp. 341–352, Cham. ISBN: 978-3-319-23179-2 978-3-319-23180-8. doi: 10.1007/978-3-319-23180-8_25. Disponível em: <http://link.springer.com/10.1007/978-3-319-23180-8_25>.

CÂMARA, M. M., SOARES, R. M., FEITAL, T., et al., 2017, “Numerical Aspects of Data Reconciliation in Industrial Applications”, *Processes*, v. 5, n. 4 (dez.), pp. 56. ISSN: 2227-9717. doi: 10.3390/pr5040056. Disponível em: <<https://www.mdpi.com/2227-9717/5/4/56>>.

CÂMARA, M. M., 2018, *Técnicas de Identificação Voltadas para a Otimização de Processos em Tempo Real*. Tese de Doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

DAI, X., GAO, Z., 2013, “From Model, Signal to Knowledge: A Data-Driven Perspective of Fault Detection and Diagnosis”, *IEEE Transactions on Industrial Informatics*, v. 9, n. 4 (nov.), pp. 2226–2238. ISSN: 1551-3203, 1941-0050. doi: 10.1109/TII.2013.2243743. Disponível em: <<http://ieeexplore.ieee.org/document/6423903/>>.

- DARKOW, T., DITTMAR, R., TIMM, H., 2014, “Real-time application of multivariate statistical methods for early event detection in an industrial slurry stripper”, *IFAC Proceedings Volumes*, v. 47, n. 3, pp. 8879–8884. ISSN: 14746670. doi: 10.3182/20140824-6-ZA-1003.00399. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667016430144>>.
- DAS, A., MAITI, J., BANERJEE, R., 2012, “Process monitoring and fault detection strategies: a review”, *International Journal of Quality & Reliability Management*, v. 29, n. 7 (jul.), pp. 720–752. ISSN: 0265-671X. doi: 10.1108/02656711211258508. Disponível em: <<https://www.emeraldinsight.com/doi/10.1108/02656711211258508>>.
- DAS, L., SIVARAM, A., VENKATASUBRAMANIAN, V., 2020, “Hidden representations in deep neural networks: Part 2. Regression problems”, *Computers & Chemical Engineering*, v. 139 (ago.), pp. 106895. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106895. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419311871>>.
- DE BIE, T., CRISTIANINI, N., ROSIPAL, R., 2005, “Eigenproblems in Pattern Recognition”. In: *Handbook of Geometric Computing*, Springer Berlin Heidelberg, pp. 129–167, Berlin, Heidelberg. ISBN: 978-3-540-20595-1 978-3-540-28247-1. doi: 10.1007/3-540-28247-5_5. Disponível em: <http://link.springer.com/10.1007/3-540-28247-5_5>.
- DECKE, J., SCHMEISSING, J., BOTACHE, D., et al., 2022, “NDNET: A Unified Framework for Anomaly and Novelty Detection”. In: Schulz, M., Trinitis, C., Papadopoulou, N., et al. (Eds.), *Architecture of Computing Systems*, Lecture Notes in Computer Science, pp. 197–210, Cham. Springer International Publishing. ISBN: 9783031218675. doi: 10.1007/978-3-031-21867-5_13.
- DENG, J., DONG, W., SOCHER, R., et al., 2009, “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, jun. doi: 10.1109/CVPR.2009.5206848.
- DENG, L., 2012, “The MNIST Database of Handwritten Digit Images for Machine Learning Research”, *IEEE Signal Processing Magazine*, v. 29, n. 6 (nov.), pp. 141–142. ISSN: 1558-0792. doi: 10.1109/MSP.2012.2211477.
- DENG, W., LI, Y., HUANG, K., et al., 2023, “LSTMED: An uneven dynamic process monitoring method based on LSTM and Autoencoder neural network”,

Neural Networks, v. 158 (jan.), pp. 30–41. ISSN: 0893-6080. doi: 10.1016/j.neunet.2022.11.001. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0893608022004415>>.

DENG, X., TIAN, X., CHEN, S., 2013, “Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault diagnosis”, *Chemometrics and Intelligent Laboratory Systems*, v. 127 (ago.), pp. 195–209. ISSN: 01697439. doi: 10.1016/j.chemolab.2013.07.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743913001330>>.

DENG, X., TIAN, X., CHEN, S., et al., 2016, “Statistics local fisher discriminant analysis for industrial process fault classification”. In: *2016 UKACC 11th International Conference on Control (CONTROL)*, Belfast, United Kingdom, ago. IEEE. ISBN: 978-1-4673-9891-6. doi: 10.1109/CONTROL.2016.7737588. Disponível em: <<http://ieeexplore.ieee.org/document/7737588/>>.

DENG, X., TIAN, X., CHEN, S., et al., 2018a, “Deep Principal Component Analysis Based on Layerwise Feature Extraction and Its Application to Nonlinear Process Monitoring”, *IEEE Transactions on Control Systems Technology*. ISSN: 1063-6536, 1558-0865, 2374-0159. doi: 10.1109/TCST.2018.2865413. Disponível em: <<https://ieeexplore.ieee.org/document/8456838/>>.

DENG, X., TIAN, X., CHEN, S., et al., 2018b, “Nonlinear Process Fault Diagnosis Based on Serial Principal Component Analysis”, *IEEE Transactions on Neural Networks and Learning Systems*, v. 29, n. 3 (mar.), pp. 560–572. ISSN: 2162-237X, 2162-2388. doi: 10.1109/TNNLS.2016.2635111. Disponível em: <<http://ieeexplore.ieee.org/document/7795255/>>.

DESTRO, F., FACCO, P., GARCIA MUÑOZ, S., et al., 2020, “A hybrid framework for process monitoring: Enhancing data-driven methodologies with state and parameter estimation”, *Journal of Process Control*, v. 92 (ago.), pp. 333–351. ISSN: 09591524. doi: 10.1016/j.jprocont.2020.06.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152420302365>>.

DIEZ-OLIVAN, A., DEL SER, J., GALAR, D., et al., 2019, “Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0”, *Information Fusion*, v. 50 (out.), pp. 92–111. ISSN: 15662535.

doi: 10.1016/j.inffus.2018.10.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1566253518304706>>.

DIKS, C., 1999, *Nonlinear Time Series Analysis: Methods And Applications*. Singapore ; River Edge, NJ, Wspc. ISBN: 978-981-02-3505-5.

DING, S., 2021, *Advanced methods for fault diagnosis and fault tolerant control*. Berlin, Heidelberg, Springer Berlin Heidelberg. ISBN: 978-3-662-62003-8 978-3-662-62004-5. doi: 10.1007/978-3-662-62004-5. Disponível em: <<http://link.springer.com/10.1007/978-3-662-62004-5>>.

DING, S. X., 2014, *Data-driven Design of Fault Diagnosis and Fault-tolerant Control Systems*. Advances in Industrial Control. London, Springer London. ISBN: 978-1-4471-6409-8 978-1-4471-6410-4. doi: 10.1007/978-1-4471-6410-4. Disponível em: <<http://link.springer.com/10.1007/978-1-4471-6410-4>>.

DING, S., ZHANG, P., JEINSCH, T., et al., 2011, “A survey of the application of basic data-driven and model-based methods in process monitoring and fault diagnosis”, *IFAC Proceedings Volumes*, v. 44, n. 1 (jan.), pp. 12380–12388. ISSN: 14746670. doi: 10.3182/20110828-6-IT-1002.02842. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667016456077>>.

DING, X., LI, Y., BELATRECHE, A., et al., 2014, “An experimental evaluation of novelty detection methods”, *Neurocomputing*, v. 135 (jul.), pp. 313–327. ISSN: 09252312. doi: 10.1016/j.neucom.2013.12.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231213011314>>.

DOBBELAERE, M. R., PLEHIERS, P. P., VAN DE VIJVER, R., et al., 2021, “Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats”, *Engineering*, v. 7, n. 9 (set.), pp. 1201–1211. ISSN: 2095-8099. doi: 10.1016/j.eng.2021.03.019. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2095809921002010>>.

DOKMANIC, I., PARHIZKAR, R., RANIERI, J., et al., 2015, “Euclidean Distance Matrices: Essential Theory, Algorithms and Applications”, *IEEE Signal Processing Magazine*, v. 32, n. 6 (nov.), pp. 12–30. ISSN: 1053-5888. doi: 10.1109/MSP.2015.2398954. Disponível em: <<http://arxiv.org/abs/1502.07541>>.

- DOMINGUES, R., MICHIARDI, P., BARLET, J., et al., 2020, “A comparative evaluation of novelty detection algorithms for discrete sequences”, *Artificial Intelligence Review*, v. 53, n. 5 (jun.), pp. 3787–3812. ISSN: 1573-7462. doi: 10.1007/s10462-019-09779-4. Disponível em: <<https://doi.org/10.1007/s10462-019-09779-4>>.
- DONG, D., MCAVOY, T., 1996, “Nonlinear principal component analysis—Based on principal curves and neural networks”, *Computers & Chemical Engineering*, v. 20, n. 1 (jan.), pp. 65–78. ISSN: 0098-1354. doi: 10.1016/0098-1354(95)0003-K. Disponível em: <<https://www.sciencedirect.com/science/article/pii/009813549500003K>>.
- DONG, Y., QIN, S. J., 2015, “Dynamic-Inner Partial Least Squares for Dynamic Data Modeling”, *IFAC-PapersOnLine*, v. 48, n. 8, pp. 117–122. ISSN: 24058963. doi: 10.1016/j.ifacol.2015.08.167. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405896315010344>>.
- DONG, Y., QIN, S. J., 2018a, “Dynamic latent variable analytics for process operations and control”, *Computers & Chemical Engineering*, v. 114 (jun.), pp. 69–80. ISSN: 00981354. doi: 10.1016/j.compchemeng.2017.10.029. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135417303848>>.
- DONG, Y., QIN, S. J., 2020, “Dynamic Inner Canonical Correlation Analysis based Process Monitoring”. In: *2020 American Control Conference (ACC)*, pp. 3553–3558, Denver, CO, USA, jul. IEEE. ISBN: 978-1-5386-8266-1. doi: 10.23919/ACC45564.2020.9147890. Disponível em: <<https://ieeexplore.ieee.org/document/9147890/>>.
- DONG, Y., QIN, S. J., 2018b, “A novel dynamic PCA algorithm for dynamic data modeling and process monitoring”, *Journal of Process Control*, v. 67 (jul.), pp. 1–11. ISSN: 09591524. doi: 10.1016/j.jprocont.2017.05.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095915241730094X>>.
- DOWNS, J., VOGEL, E., 1993, “A plant-wide industrial process control problem”, *Computers & Chemical Engineering*, v. 17, n. 3 (mar.), pp. 245–255. ISSN: 0098-1354. doi: 10.1016/0098-1354(93)80018-I. Disponível em: <[https://doi.org/10.1016/0098-1354\(93\)80018-I](https://doi.org/10.1016/0098-1354(93)80018-I)>.
- DRUCKER, H., BURGES, C. J. C., KAUFMAN, L., et al., 1997, “Support Vector Regression Machines”, *Advances in neural information pro-*

cessing systems. Disponível em: <<https://papers.nips.cc/paper/1238-support-vector-regression-machines>>.

DU, W., YANG, J., MENG, G., 2023, ‘Fault diagnosis for dynamic system based on the independent latent space reconstruction of generative adversarial network”, *Journal of Process Control*, v. 125 (maio), pp. 28–40. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2023.04.001. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152423000598>>.

DUDZIC, M., VACULIK, V., MILETIC, I., 2000, “On-line applications of multivariate statistics at Dofasco”, *IFAC Proceedings Volumes*, v. 33, n. 22 (ago.), pp. 425–430. ISSN: 14746670. doi: 10.1016/S1474-6670(17)37032-5. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017370325>>.

DUMAREY, M., HERMANTO, M., AIRIAU, C., et al., 2018, “Advances in Continuous Active Pharmaceutical Ingredient (API) Manufacturing: Real-time Monitoring Using Multivariate Tools”, *Journal of Pharmaceutical Innovation*, (out.). ISSN: 1872-5120, 1939-8042. doi: 10.1007/s12247-018-9348-7. Disponível em: <<http://link.springer.com/10.1007/s12247-018-9348-7>>.

DUNIA, R., ROCHELLE, G., EDGAR, T. F., et al., 2014, “Multivariate monitoring of a carbon dioxide removal process”, *Computers & Chemical Engineering*, v. 60 (jan.), pp. 381–395. ISSN: 00981354. doi: 10.1016/j.compchemeng.2013.09.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135413002792>>.

DURRE, A., FRIED, R., LIBOSCHIK, T., 2014, “Robust estimation of (partial) autocorrelation”, *Discussion papers SFB 823*, (abr.). doi: 10.17877/DE290R-13701. Disponível em: <<https://eldorado.tu-dortmund.de/handle/2003/33011>>.

DZAFERAGIC, M., MARCHETTI, N., MACALUSO, I., 2022, “Fault Detection and Classification in Industrial IoT in Case of Missing Sensor Data”, *IEEE Internet of Things Journal*, v. 9, n. 11 (jun.), pp. 8892–8900. ISSN: 2327-4662. doi: 10.1109/JIOT.2021.3116785.

ECKMANN, J.-P., KAMPHORST, S. O., RUELLE, D., 1987, “Recurrence Plots of Dynamical Systems”, *Europhysics Letters (EPL)*, v. 4, n. 9 (nov.), pp. 973–977. ISSN: 0295-5075. doi: 10.1209/0295-5075/4/9/004. Disponível em: <<https://doi.org/10.1209/0295-5075/4/9/004>>.

- EISSA, M. E., HAMED, H. S., 2019, “Application of Statistical Quality Control Tools for Monitoring of Pharmaceutical Product Properties”, *Biological Sciences-PJSIR*, v. 62, n. 1, pp. 39–48. Disponível em: <<https://v2.pjsir.org/index.php/biological-sciences/article/download/350/213>>.
- ELSHENAWY, L. M., CHAKOUR, C., MAHMOUD, T. A., 2022, “Fault detection and diagnosis strategy based on k-nearest neighbors and fuzzy C-means clustering algorithm for industrial processes”, *Journal of the Franklin Institute*, (jul.). ISSN: 0016-0032. doi: 10.1016/j.jfranklin.2022.06.022. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0016003222004252>>.
- EMMANOUILIDIS, C., PISTOFIDIS, P., BERTONCELJ, L., et al., 2019, “Enabling the human in the loop: Linked data and knowledge in industrial cyber-physical systems”, *Annual Reviews in Control*, v. 47, pp. 249–265. ISSN: 13675788. doi: 10.1016/j.arcontrol.2019.03.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1367578818301366>>.
- ENDERS, W., 2014, *Applied Econometric Time Series*. 4 edition ed. Hoboken, NJ, Wiley. ISBN: 978-1-118-80856-6.
- ESCOBAR, C. A., MCGOVERN, M. E., MORALES-MENENDEZ, R., 2021, “Quality 4.0: a review of big data challenges in manufacturing”, *Journal of Intelligent Manufacturing*, v. 32, n. 8 (dez.), pp. 2319–2334. ISSN: 1572-8145. doi: 10.1007/s10845-021-01765-4. Disponível em: <<https://doi.org/10.1007/s10845-021-01765-4>>.
- EVSUKOFF, A., GENTIL, S., 2005, “Recurrent neuro-fuzzy system for fault detection and isolation in nuclear reactors”, *Advanced Engineering Informatics*, v. 19, n. 1 (jan.), pp. 55–66. ISSN: 14740346. doi: 10.1016/j.aei.2005.01.009. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S147403460500008X>>.
- EZUGWU, A. E., IKOTUN, A. M., OYELADE, O. O., et al., 2022, “A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects”, *Engineering Applications of Artificial Intelligence*, v. 110 (abr.), pp. 104743. ISSN: 09521976. doi: 10.1016/j.engappai.2022.104743. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095219762200046X>>.
- FADEL, F. E. D. F., 2018, *Uma Avaliação Crítica Sobre Técnicas Baseadas em PCA para Detecção de Falhas em Processos da Indústria Química*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, jun.

- FAN, S.-K. S., HSU, C.-Y., TSAI, D.-M., et al., 2020, “Data-Driven Approach for Fault Detection and Diagnostic in Semiconductor Manufacturing”, *IEEE Transactions on Automation Science and Engineering*, v. 17, n. 4 (out.), pp. 1925–1936. ISSN: 1545-5955, 1558-3783. doi: 10.1109/TASE.2020.2983061. Disponível em: <<https://ieeexplore.ieee.org/document/9066890/>>.
- FARIA, E. R., GONCALVES, I. R., GAMA, J. A., et al., 2015, “Evaluation of Multiclass Novelty Detection Algorithms for Data Streams”, *IEEE Transactions on Knowledge and Data Engineering*, v. 27, n. 11 (nov.), pp. 2961–2973. ISSN: 1041-4347. doi: 10.1109/TKDE.2015.2441713. Disponível em: <<http://ieeexplore.ieee.org/document/7118190/>>.
- FEITAL, T., PINTO, J. C., 2015, “Use of variance spectra for in-line validation of process measurements in continuous processes”, *The Canadian Journal of Chemical Engineering*, v. 93, n. 8, pp. 1426–1437. ISSN: 1939-019X. doi: 10.1002/cjce.22219. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cjce.22219>>.
- FEITAL, T., KRUGER, U., LEI XIE, et al., 2010, “A unified statistical framework for monitoring multivariate systems with unknown source and error signals”, *Chemometrics and Intelligent Laboratory Systems*, v. 104, n. 2 (dez.), pp. 223–232. ISSN: 0169-7439. doi: 10.1016/j.chemolab.2010.08.012. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169743910001607>>.
- FEITAL, T., KRUGER, U., DUTRA, J., et al., 2013, “Modeling and performance monitoring of multivariate multimodal processes”, *AICHE Journal*, v. 59, n. 5, pp. 1557–1569. ISSN: 1547-5905. doi: 10.1002/aic.13953. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/aic.13953>>.
- FEITAL, T. D. S., 2011, *Monitoramento da Condição de Processos Químicos Industriais*. Tese de Doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, nov.
- FENG, J., LI, K., 2020, “MRS kNN fault detection method for multirate sampling process based variable grouping threshold”, *Journal of Process Control*, v. 85 (jan.), pp. 149–158. ISSN: 09591524. doi: 10.1016/j.jprocont.2019.11.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152419305104>>.
- FENG, J., WANG, J., ZHANG, H., et al., 2016, “Fault Diagnosis Method of Joint Fisher Discriminant Analysis Based on the Local and Global Manifold

Learning and Its Kernel Version”, *IEEE Transactions on Automation Science and Engineering*, v. 13, n. 1 (jan.), pp. 122–133. ISSN: 1545-5955, 1558-3783. doi: 10.1109/TASE.2015.2417882. Disponível em: <<http://ieeexplore.ieee.org/document/7091038/>>.

FERNANDES, M., CORCHADO, J. M., MARREIROS, G., 2022a, “Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review”, *Applied Intelligence*, v. 52, n. 12 (set.), pp. 14246–14280. ISSN: 1573-7497. doi: 10.1007/s10489-022-03344-3. Disponível em: <<https://doi.org/10.1007/s10489-022-03344-3>>.

FERNANDES, N. C. P., RATO, T. J., REIS, M. S., 2022b, “Modeling in the observable or latent space? A comparison of dynamic latent variable based monitoring methods for sensor fault detection”, *Chemometrics and Intelligent Laboratory Systems*, v. 231 (dez.), pp. 104684. ISSN: 0169-7439. doi: 10.1016/j.chemolab.2022.104684. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169743922001952>>.

FERREIRA, A. M. S., DE OLIVEIRA FONTES, C. H., CAVALCANTE, C. A. M. T., et al., 2015, “Pattern recognition as a tool to support decision making in the management of the electric sector. Part II: A new method based on clustering of multivariate time series”, *International Journal of Electrical Power & Energy Systems*, v. 67 (maio), pp. 613–626. ISSN: 01420615. doi: 10.1016/j.ijepes.2014.12.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0142061514007285>>.

FERRER, A., 2007, “Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process”, *Quality Engineering*, v. 19, n. 4 (out.), pp. 311–325. ISSN: 0898-2112, 1532-4222. doi: 10.1080/08982110701621304. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/08982110701621304>>.

FERRER, A., AGUADO, D., VIDAL-PUIG, S., et al., 2008, “PLS: A versatile tool for industrial process improvement and optimization”, *Applied Stochastic Models in Business and Industry*, v. 24, n. 6 (nov.), pp. 551–567. ISSN: 15241904, 15264025. doi: 10.1002/asmb.716. Disponível em: <<http://doi.wiley.com/10.1002/asmb.716>>.

FISCHER, A., IGEL, C., 2012, “An Introduction to Restricted Boltzmann Machines”. In: Alvarez, L., Mejail, M., Gomez, L., et al. (Eds.), *Progress in Pattern*

Recognition, Image Analysis, Computer Vision, and Applications, Lecture Notes in Computer Science, pp. 14–36, Berlin, Heidelberg. Springer. ISBN: 9783642332753. doi: 10.1007/978-3-642-33275-3_2.

FISHER, O. J., WATSON, N. J., ESCRIG, J. E., et al., 2020, “Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems”, *Computers & Chemical Engineering*, v. 140 (set.), pp. 106881. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106881. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419308373>>.

FISHER, R. A., 1936, “The Use of Multiple Measurements in Taxonomic Problems”, *Annals of Eugenics*, v. 7, n. 2 (set.), pp. 179–188. ISSN: 20501420. doi: 10.1111/j.1469-1809.1936.tb02137.x. Disponível em: <<http://doi.wiley.com/10.1111/j.1469-1809.1936.tb02137.x>>.

FLEMING, P. J., WALLACE, J. J., 1986, “How not to lie with statistics: the correct way to summarize benchmark results”, *Communications of the ACM*, v. 29, n. 3 (mar.), pp. 218–221. ISSN: 00010782. doi: 10.1145/5666.5673. Disponível em: <<http://portal.acm.org/citation.cfm?doid=5666.5673>>.

FONTES, C. H., BUDMAN, H., 2017, “A hybrid clustering approach for multivariate time series – A case study applied to failure analysis in a gas turbine”, *ISA Transactions*, v. 71 (nov.), pp. 513–529. ISSN: 00190578. doi: 10.1016/j.isatra.2017.09.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0019057817305530>>.

FONTES, C. H., SANTOS, I. C., EMBIRUÇU, M., et al., 2021, “Pattern reconciliation: A new approach involving constrained clustering of time series”, *Computers & Chemical Engineering*, v. 145 (fev.), pp. 107169. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.107169. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135420307614>>.

FONTOURA, T. B., DE SÁ, M. C. C., DE MENEZES, D. Q. F., et al., 2022, “Modeling of spiral wound membranes for gas separations. Part III: A nonisothermal 2D permeation model”, *Chemical Engineering Research and Design*, v. 177 (jan.), pp. 376–393. ISSN: 0263-8762. doi: 10.1016/j.cherd.2021.10.036. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263876221004469>>.

FOORTHUIS, R., 2021, “On the nature and types of anomalies: a review of deviations in data”, *International Journal of Data Science and Analytics*, v. 12,

n. 4 (out.), pp. 297–331. ISSN: 2364-4168. doi: 10.1007/s41060-021-00265-1.
Disponível em: <<https://doi.org/10.1007/s41060-021-00265-1>>.

FU, T.-C., 2011, “A review on time series data mining”, *Engineering Applications of Artificial Intelligence*, v. 24, n. 1 (fev.), pp. 164–181. ISSN: 09521976. doi: 10.1016/j.engappai.2010.09.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0952197610001727>>.

FU, Y., LUO, C., 2019, “Joint Structure Preserving Embedding Model and Its Application for Process Monitoring”, *Industrial & Engineering Chemistry Research*, v. 58, n. 45 (nov.), pp. 20667–20679. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.9b03077. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.9b03077>>.

FUENTE, M. J. D. L., SAINZ-PALMERO, G. I., GALENDE-HERNÁNDEZ, M., 2023, “Dynamic Decentralized Monitoring for Large-Scale Industrial Processes Using Multiblock Canonical Variate Analysis Based Regression”, *IEEE Access*, v. 11, pp. 26611–26623. ISSN: 2169-3536. doi: 10.1109/ACCESS.2023.3256719.

FUENTES-CORTÉS, L. F., FLORES-TLACUAHUAC, A., NIGAM, K. D. P., 2022, “Machine Learning Algorithms Used in PSE Environments: A Didactic Approach and Critical Perspective”, *Industrial & Engineering Chemistry Research*, v. 61, n. 25 (jun.), pp. 8932–8962. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.2c00335. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.2c00335>>.

FUENTES-GARCÍA, M., MACIÁ-FERNÁNDEZ, G., CAMACHO, J., 2018, “Evaluation of diagnosis methods in PCA-based Multivariate Statistical Process Control”, *Chemometrics and Intelligent Laboratory Systems*, v. 172 (jan.), pp. 194–210. ISSN: 01697439. doi: 10.1016/j.chemolab.2017.12.008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743917302046>>.

GAJJAR, S., PALAZOGLU, A., 2016, “A data-driven multidimensional visualization technique for process fault detection and diagnosis”, *Chemometrics and Intelligent Laboratory Systems*, v. 154 (maio), pp. 122–136. ISSN: 01697439. doi: 10.1016/j.chemolab.2016.03.027. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743916300703>>.

GAJJAR, S., KULAHCI, M., PALAZOGLU, A., 2020, “Least Squares Sparse Principal Component Analysis and Parallel Coordinates for Real-Time

Process Monitoring”, *Industrial & Engineering Chemistry Research*, v. 59, n. 35 (set.), pp. 15656–15670. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c01749. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.0c01749>>.

GALLAGHER, N. B., WISE, B. M., STEWART, C. W., 1996, “Application of multi-way principal components analysis to nuclear waste storage tank monitoring”, *Computers & Chemical Engineering*, v. 20 (jan.), pp. S739–S744. ISSN: 00981354. doi: 10.1016/0098-1354(96)00131-7. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0098135496001317>>.

GALLAGHER, N. B., WISE, B. M., BUTLER, S. W., et al., 1997, “Development and Benchmarking of Multivariate Statistical Process Control Tools for a Semiconductor Etch Process: Improving Robustness through Model Updating”, *IFAC Proceedings Volumes*, v. 30, n. 9 (jun.), pp. 79–84. ISSN: 14746670. doi: 10.1016/S1474-6670(17)43143-0. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017431430>>.

GAMMA, E., HELM, R., DR, R. J., et al., 1994, *Design Patterns: Elements of Reusable Object-Oriented Software*. 1^a edição ed. Reading, Mass, Addison-Wesley Professional. ISBN: 978-0-201-63361-0.

GANDOMI, A., HAIDER, M., 2015, “Beyond the hype: Big data concepts, methods, and analytics”, *International Journal of Information Management*, v. 35, n. 2 (abr.), pp. 137–144. ISSN: 02684012. doi: 10.1016/j.ijinfomgt.2014.10.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0268401214001066>>.

GANESAN, R., DAS, T., VENKATARAMAN, V., 2004, “Wavelet-based multiscale statistical process monitoring: A literature review”, *IIE Transactions*, v. 36, n. 9 (set.), pp. 787–806. ISSN: 0740-817X. doi: 10.1080/07408170490473060. Disponível em: <<https://doi.org/10.1080/07408170490473060>>.

GAO, H., ZHU, L.-T., LUO, Z.-H., et al., 2022, “Machine Learning and Data Science in Chemical Engineering”, *Industrial & Engineering Chemistry Research*, v. 61, n. 24 (jun.), pp. 8357–8358. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.2c01788. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.2c01788>>.

GAO, H., HUANG, W., GAO, X., et al., 2023, “Decentralized adaptively weighted stacked autoencoder-based incipient fault detection for nonlinear industrial

- processes”, *ISA Transactions*, (maio). ISSN: 0019-0578. doi: 10.1016/j.isatra.2023.04.035. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0019057823002045>>.
- GAO, R. X., WANG, L., HELU, M., et al., 2020a, “Big data analytics for smart factories of the future”, *CIRP Annals*, (jun.). ISSN: 00078506. doi: 10.1016/j.cirp.2020.05.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0007850620301359>>.
- GAO, X., SHARDT, Y. A. W., 2021, “Dynamic system modelling and process monitoring based on long-term dependency slow feature analysis”, *Journal of Process Control*, v. 105 (set.), pp. 27–47. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2021.07.007. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152421001086>>.
- GAO, X., SHARDT, Y. A. W., 2022, “EVOLVE-INFOMAX: An Unsupervised Learning Principle of Invariances for Nonlinear Dynamic Systems”, *Industrial & Engineering Chemistry Research*, v. 61, n. 49 (dez.), pp. 18004–18016. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.2c03330. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.2c03330>>.
- GAO, X., YANG, F., FENG, E., 2020b, “A process fault diagnosis method using multi-time scale dynamic feature extraction based on convolutional neural network”, *The Canadian Journal of Chemical Engineering*, v. n/a, n. n/a (mar.). ISSN: 1939-019X. doi: 10.1002/cjce.23740. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cjce.23740>>.
- GAO, Z., LIU, X., 2021, “An Overview on Fault Diagnosis, Prognosis and Resilient Control for Wind Turbine Systems”, *Processes*, v. 9, n. 2 (fev.), pp. 300. doi: 10.3390/pr9020300. Disponível em: <<https://www.mdpi.com/2227-9717/9/2/300>>.
- GARCIA-ALVAREZ, D., FUENTE, M., VEGA, P., et al., 2009, “Fault Detection and Diagnosis using Multivariate Statistical Techniques in a Wastewater Treatment Plant”, *IFAC Proceedings Volumes*, v. 42, n. 11, pp. 952–957. ISSN: 14746670. doi: 10.3182/20090712-4-TR-2008.00156. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667015303992>>.
- GARG, A., ZHANG, W., SAMARAN, J., et al., 2022, “An Evaluation of Anomaly Detection and Diagnosis in Multivariate Time Series”, *IEEE Transactions on Neural Networks and Learning Systems*, v. 33, n. 6 (jun.), pp. 2508–2517. ISSN: 2162-2388. doi: 10.1109/TNNLS.2021.3105827.

- GATIUS, F., MIRALBÉS, C., DAVID, C., et al., 2017, “Comparison of CCA and PLS to explore and model NIR data”, *Chemometrics and Intelligent Laboratory Systems*, v. 164 (maio), pp. 76–82. ISSN: 01697439. doi: 10.1016/j.chemolab.2017.03.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743916305226>>.
- GE, Z., 2018, “Process Data Analytics via Probabilistic Latent Variable Models: A Tutorial Review”, *Industrial & Engineering Chemistry Research*, v. 57, n. 38 (set.), pp. 12646–12661. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.8b02913. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.8b02913>>.
- GE, Z., 2017, “Review on data-driven modeling and monitoring for plant-wide industrial processes”, *Chemometrics and Intelligent Laboratory Systems*, v. 171 (dez.), pp. 16–25. ISSN: 01697439. doi: 10.1016/j.chemolab.2017.09.021. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743916303811>>.
- GE, Z., SONG, Z., 2013a, “Distributed PCA Model for Plant-Wide Process Monitoring”, *Industrial & Engineering Chemistry Research*, v. 52, n. 5 (fev.), pp. 1947–1957. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie301945s. Disponível em: <<http://pubs.acs.org/doi/10.1021/ie301945s>>.
- GE, Z., SONG, Z., 2011, “A distribution-free method for process monitoring”, *Expert Systems with Applications*, v. 38, n. 8 (ago.), pp. 9821–9829. ISSN: 09574174. doi: 10.1016/j.eswa.2011.02.048. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417411002545>>.
- GE, Z., SONG, Z., 2010, “Mixture Bayesian regularization method of PPCA for multimode process monitoring”, *AICHE Journal*, v. 56, n. 11 (nov.), pp. 2838–2849. ISSN: 00011541. doi: 10.1002/aic.12200. Disponível em: <<http://doi.wiley.com/10.1002/aic.12200>>.
- GE, Z., SONG, Z., 2009, “Multimode process monitoring based on Bayesian method”, *Journal of Chemometrics*, v. 23 (nov.), pp. 636–650. ISSN: 08869383, 1099128X. doi: 10.1002/cem.1262. Disponível em: <<http://doi.wiley.com/10.1002/cem.1262>>.
- GE, Z., SONG, Z., 2013b, *Multivariate Statistical Process Control. Advances in Industrial Control*. London, Springer London. ISBN: 978-1-4471-4512-7 978-1-4471-4513-4. doi: 10.1007/978-1-4471-4513-4. Disponível em: <<http://link.springer.com/10.1007/978-1-4471-4513-4>>.

- GE, Z., GAO, F., SONG, Z., 2011, “Batch process monitoring based on support vector data description method”, *Journal of Process Control*, v. 21, n. 6 (jul.), pp. 949–959. ISSN: 09591524. doi: 10.1016/j.jprocont.2011.02.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152411000321>>.
- GE, Z., SONG, Z., GAO, F., 2013, “Review of Recent Research on Data-Based Process Monitoring”, *Industrial & Engineering Chemistry Research*, v. 52, n. 10 (mar.), pp. 3543–3562. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie302069q. Disponível em: <<http://pubs.acs.org/doi/10.1021/ie302069q>>.
- GE, Z., SONG, Z., DING, S. X., et al., 2017, “Data Mining and Analytics in the Process Industry: The Role of Machine Learning”, *IEEE Access*, v. 5, pp. 20590–20616. ISSN: 2169-3536. doi: 10.1109/ACCESS.2017.2756872. Disponível em: <<http://ieeexplore.ieee.org/document/8051033/>>.
- GERMANO, A. L., GUEDES, L. A., COSTA, B. S. J., et al., 2016, “Detecção de falhas no processo Tennessee Eastman Utilizando Métricas de Tipicidade e Excentricidade”. In: *XXI Congresso Brasileiro de Automática - CBA2016*. Disponível em: <<https://www.researchgate.net/publication/311104211>>.
- GERON, A., 2017, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 1 edition ed. Beijing ; Boston, O'Reilly Media. ISBN: 978-1-4919-6229-9.
- GHOLIZADEH, M., YAZDIZADEH, A., MOHAMMAD-BAGHERPOUR, H., 2019, “Fault detection and identification using combination of EKF and neuro-fuzzy network applied to a chemical process (CSTR)”, *Pattern Analysis and Applications*, v. 22, n. 2 (maio), pp. 359–373. ISSN: 1433-755X. doi: 10.1007/s10044-017-0634-7. Disponível em: <<https://doi.org/10.1007/s10044-017-0634-7>>.
- GHOSAL, T., SAIKH, T., BISWAS, T., et al., 2022, “Novelty Detection: A Perspective from Natural Language Processing”, *Computational Linguistics*, v. 48, n. 1 (abr.), pp. 77–117. ISSN: 0891-2017, 1530-9312. doi: 10.1162/coli_a_00429. Disponível em: <<https://direct.mit.edu/coli/article/48/1/77/108847/Novelty-Detection-A-Perspective-from-Natural>>.
- GHOSH, A., AHMED, S., KHAN, F., 2019, “Modeling and Testing of Temporal Dependency in the Failure of a Process System”, *Industrial & Engineering Chemistry Research*, v. 58, n. 19 (maio), pp. 8162–8171. ISSN: 0888-5885.

doi: 10.1021/acs.iecr.8b06300. Disponível em: <<https://doi.org/10.1021/acs.iecr.8b06300>>.

GIL, M., ALBERT, M., FONS, J., et al., 2020, “Engineering human-in-the-loop interactions in cyber-physical systems”, *Information and Software Technology*, v. 126 (out.), pp. 106349. ISSN: 0950-5849. doi: 10.1016/j.infsof.2020.106349. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950584920301166>>.

GINS, G., VANLAER, J., VAN DEN KERKHOF, P., et al., 2014, “The RAYMOND simulation package — Generating RAYpresentative MONitoring Data to design advanced process monitoring and control algorithms”, *Computers & Chemical Engineering*, v. 69 (out.), pp. 108–118. ISSN: 00981354. doi: 10.1016/j.compchemeng.2014.07.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135414002142>>.

GIRASA, R. J., 2020, *Artificial intelligence as a disruptive technology: economic transformation and government regulation*. Cham, Palgrave Macmillan. ISBN: 9783030359751 9783030359744 9783030359775.

GISBRECHT, A., HAMMER, B., 2015, “Data visualization by nonlinear dimensionality reduction”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 5, n. 2 (mar.), pp. 51–73. ISSN: 19424787. doi: 10.1002/widm.1147. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/widm.1147>>.

GITTINS, R., 1985, *Canonical Analysis: a Review with Applications in Ecology*. Berlin, Heidelberg, Springer Berlin Heidelberg. ISBN: 978-3-642-69878-1 978-3-642-69880-4.

GLOROT, X., BENGIO, Y., 2010, “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256, mar. Disponível em: <<http://proceedings.mlr.press/v9/glorot10a.html>>.

GMCGLINN, 2008. “Rule 8 - Control Charts for Nelson Rules.svg”..

GOEL, P., DATTA, A., MANNAN, M. S., 2017, “Application of big data analytics in process safety and risk management”. In: *2017 IEEE International Conference on Big Data (Big Data)*, pp. 1143–1152, Boston, MA, dez. IEEE. ISBN: 978-1-5386-2715-0. doi: 10.1109/BigData.2017.8258040. Disponível em: <<http://ieeexplore.ieee.org/document/8258040/>>.

- GOEL, P., JAIN, P., PASMAN, H. J., et al., 2020, “Integration of data analytics with cloud services for safer process systems, application examples and implementation challenges”, *Journal of Loss Prevention in the Process Industries*, v. 68 (nov.), pp. 104316. ISSN: 09504230. doi: 10.1016/j.jlp.2020.104316. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0950423020306033>>.
- GOLDRICK, S., STEFAN, A., LOVETT, D., et al., 2015, “The development of an industrial-scale fed-batch fermentation simulation”, *Journal of Biotechnology*, v. 193 (jan.), pp. 70–82. ISSN: 01681656. doi: 10.1016/j.jbiotec.2014.10.029. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0168165614009377>>.
- GOLDRICK, S., DURAN-VILLALOBOS, C. A., JANKAUSKAS, K., et al., 2019, “Modern day monitoring and control challenges outlined on an industrial-scale benchmark fermentation process”, *Computers & Chemical Engineering*, v. 130 (nov.), pp. 106471. ISSN: 00981354. doi: 10.1016/j.compchemeng.2019.05.037. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135418305106>>.
- GOLDSTEIN, M., UCHIDA, S., 2016, “A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data”, *PLOS ONE*, v. 11, n. 4 (abr.). ISSN: 1932-6203. doi: 10.1371/journal.pone.0152173. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0152173>>.
- GOMES, H. M., READ, J., BIFET, A., et al., 2019, “Machine learning for streaming data: state of the art, challenges, and opportunities”, *ACM SIGKDD Explorations Newsletter*, v. 21, n. 2 (nov.), pp. 6–22. ISSN: 1931-0145, 1931-0153. doi: 10.1145/3373464.3373470. Disponível em: <<https://dl.acm.org/doi/10.1145/3373464.3373470>>.
- GONG, W., CHEN, H., ZHANG, Z., et al., 2019, “A Novel Deep Learning Method for Intelligent Fault Diagnosis of Rotating Machinery Based on Improved CNN-SVM and Multichannel Data Fusion”, *Sensors*, v. 19, n. 7 (jan.), pp. 1693. ISSN: 1424-8220. doi: 10.3390/s19071693. Disponível em: <<https://www.mdpi.com/1424-8220/19/7/1693>>.
- GONZÁLEZ-MARTÍNEZ, J., CAMACHO, J., FERRER, A., 2018, “MVBatch: A matlab toolbox for batch process modeling and monitoring”, *Chemometrics and Intelligent Laboratory Systems*, v. 183 (dez.), pp. 122–133. ISSN: 01697439. doi: 10.1016/j.chemolab.2018.11.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743918304003>>.

- GOODFELLOW, I., BENGIO, Y., COURVILLE, A., 2016, *Deep Learning*. Cambridge, MIT Press. ISBN: 978-0-262-03561-3. Disponível em: <<http://www.deeplearningbook.org/>>.
- GOODFELLOW, I. J., POUGET-ABADIE, J., MIRZA, M., et al., 2014, “Generative Adversarial Networks”, *arXiv:1406.2661 [cs, stat]*, (jun.). Disponível em: <<http://arxiv.org/abs/1406.2661>>.
- GOOIJER, J. G. D., 2017, *Elements of Nonlinear Time Series Analysis and Forecasting*. Cham, Springer. ISBN: 978-3-319-43252-6.
- GOPALUNI, R. B., TULSYAN, A., CHACHUAT, B., et al., 2020, “Modern Machine Learning Tools for Monitoring and Control of Industrial Processes: A Survey”, *IFAC-PapersOnLine*, v. 53, n. 2 (jan.), pp. 218–229. ISSN: 2405-8963. doi: 10.1016/j.ifacol.2020.12.126. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2405896320303827>>.
- GOSWAMI, B., 2019, “A Brief Introduction to Nonlinear Time Series Analysis and Recurrence Plots”, *Vibration*, v. 2, n. 4 (dez.), pp. 332–368. ISSN: 2571-631X. doi: 10.3390/vibration2040021. Disponível em: <<https://www.mdpi.com/2571-631X/2/4/21>>.
- GOULDING, P. R., LENNOX, B., SANDOZ, D. J., et al., 2000, “Fault detection in continuous processes using multivariate statistical methods”, *International Journal of Systems Science*, v. 31, n. 11 (jan.), pp. 1459–1471. ISSN: 0020-7721, 1464-5319. doi: 10.1080/00207720050197839. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/00207720050197839>>.
- Graham, M., Dutton, W. H. (Eds.), 2019, *Society and the Internet: How Networks of Information and Communication are Changing Our Lives*. Oxford, Oxford University Press. ISBN: 978-0-19-258156-3.
- GREGERSEN, L., JØRGENSEN, S. B., ANDERSEN, M. Y., 1997, “Industrial Fed-Batch Fermentation Monitoring”, *IFAC Proceedings Volumes*, v. 30, n. 9 (jun.), pp. 49–54. ISSN: 14746670. doi: 10.1016/S1474-6670(17)43138-7. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017431387>>.
- GROVER, V., CHIANG, R. H., LIANG, T.-P., et al., 2018, “Creating Strategic Business Value from Big Data Analytics: A Research Framework”, *Journal of Management Information Systems*, v. 35, n. 2 (abr.), pp. 388–423. ISSN: 0742-1222, 1557-928X. doi: 10.1080/07421222.2018.1451951.

Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/07421222.2018.1451951>>.

GRUHL, C. M., 2022, *Novelty Detection for Multivariate Data Streams with Probabilistic Models*. Kassel. doi: 10.17170/KOBRA-202205106160. Disponível em: <<https://kobra.uni-kassel.de/handle/123456789/13902>>.

GRUS, J., 2015, *Data science from scratch: first principles with Python*. First edition ed. Sebastopol, CA, O'Reilly. ISBN: 978-1-4919-0142-7.

GUH, R.-S., SHIUE, Y.-R., 2008, “An effective application of decision tree learning for on-line detection of mean shifts in multivariate control charts”, *Computers & Industrial Engineering*, v. 55, n. 2 (set.), pp. 475–493. ISSN: 03608352. doi: 10.1016/j.cie.2008.01.013. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0360835208000107>>.

GUNTHER, W. A., REZAZADE MEHRIZI, M. H., HUYSMAN, M., et al., 2017, “Debating big data: A literature review on realizing value from big data”, *The Journal of Strategic Information Systems*, v. 26, n. 3 (set.), pp. 191–209. ISSN: 09638687. doi: 10.1016/j.jsis.2017.07.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0963868717302615>>.

GUO, L., WU, P., LOU, S., et al., 2020, “A multi-feature extraction technique based on principal component analysis for nonlinear dynamic process monitoring”, *Journal of Process Control*, v. 85 (jan.), pp. 159–172. ISSN: 09591524. doi: 10.1016/j.jprocont.2019.11.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152419305062>>.

GUO, P., RAO, S., HAO, L., et al., 2022, “Fault diagnosis of a semi-batch crystallization process through deep learning method”, *Computers & Chemical Engineering*, v. 164 (ago.), pp. 107807. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.107807. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422001454>>.

GUO, R., CHENG, L., LI, J., et al., 2021a, “A Survey of Learning Causality with Data: Problems and Methods”, *ACM Computing Surveys*, v. 53, n. 4 (jul.), pp. 1–37. ISSN: 0360-0300, 1557-7341. doi: 10.1145/3397269. Disponível em: <<https://dl.acm.org/doi/10.1145/3397269>>.

GUO, Z., SHANG, C., YE, H., 2021b, “A novel similarity metric with application to big process data analytics”, *Control Engineering Practice*, v. 113 (ago.), pp. 104843. ISSN: 0967-0661. doi: 10.1016/j.conengprac.2021.104843.

Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0967066121001209>>.

GUPTA, M., GAO, J., AGGARWAL, C. C., et al., 2014, “Outlier Detection for Temporal Data: A Survey”, *IEEE Transactions on Knowledge and Data Engineering*, v. 26, n. 9 (set.), pp. 2250–2267. ISSN: 1041-4347. doi: 10.1109/TKDE.2013.184. Disponível em: <<http://ieeexplore.ieee.org/document/6684530/>>.

GUPTA, M., WADHVANI, R., RASOOL, A., 2022, “Real-time Change-Point Detection: A deep neural network-based adaptive approach for detecting changes in multivariate time series data”, *Expert Systems with Applications*, v. 209 (dez.), pp. 118260. ISSN: 0957-4174. doi: 10.1016/j.eswa.2022.118260. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417422014026>>.

GURDEN, S., MARTIN, E., MORRIS, A., 1998, “The introduction of process chemometrics into an industrial pilot plant laboratory”, *Chemometrics and Intelligent Laboratory Systems*, v. 44, n. 1-2 (dez.), pp. 319–330. ISSN: 01697439. doi: 10.1016/S0169-7439(98)00119-1. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743998001191>>.

HAIMI, H., MULAS, M., CORONA, F., et al., 2016, “Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant”, *Engineering Applications of Artificial Intelligence*, v. 52 (jun.), pp. 65–80. ISSN: 09521976. doi: 10.1016/j.engappai.2016.02.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0952197616300124>>.

HALLGRIMSSON, A. D., NIEMANN, H. H., LIND, M., 2021, “Unsupervised isolation of abnormal process variables using sparse autoencoders”, *Journal of Process Control*, v. 99 (mar.), pp. 107–119. ISSN: 09591524. doi: 10.1016/j.jprocont.2021.01.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152421000056>>.

HAMMER, P. L., 1986, “Partially defined Boolean functions and cause-effect relationships”. In: *Proceedings of the International Conference on Multi-Attribute Decision Making via OR-Based Expert Systems*, University of Passau.

HAN, H.-G., DONG, L.-X., QIAO, J.-F., 2021, “Data-knowledge-driven diagnosis method for sludge bulking of wastewater treatment process”, *Journal of*

- Process Control*, v. 98 (fev.), pp. 106–115. ISSN: 09591524. doi: 10.1016/j.jprocont.2021.01.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152421000019>>.
- HAN, S., HU, X., HUANG, H., et al., 2022, “ADBench: Anomaly Detection Benchmark”. In: Koyejo, S., Mohamed, S., Agarwal, A., et al. (Eds.), *Advances in Neural Information Processing Systems*, v. 35, pp. 32142–32159. Curran Associates, Inc.
- HAN, Y., DING, N., GENG, Z., et al., 2020, “An optimized long short-term memory network based fault diagnosis model for chemical processes”, *Journal of Process Control*, v. 92 (ago.), pp. 161–168. ISSN: 09591524. doi: 10.1016/j.jprocont.2020.06.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152420302390>>.
- HARROU, F., DAIRI, A., SUN, Y., et al., 2018, “Statistical monitoring of a wastewater treatment plant: A case study”, *Journal of Environmental Management*, v. 223 (out.), pp. 807–814. ISSN: 03014797. doi: 10.1016/j.jenvman.2018.06.087. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0301479718307394>>.
- HASSANPOUR, H., MHASKAR, P., HOUSE, J. M., et al., 2020, “A hybrid modeling approach integrating first-principles knowledge with statistical methods for fault detection in HVAC systems”, *Computers & Chemical Engineering*, v. 142 (nov.), pp. 107022. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.107022. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135420303975>>.
- HASTIE, T., STUETZLE, W., 1989, “Principal Curves”, *Journal of the American Statistical Association*, v. 84, n. 406 (jun.), pp. 502–516. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/01621459.1989.10478797>>.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. 2 ed. New York, Springer-Verlag. ISBN: 978-0-387-84857-0. Disponível em: <<https://www.springer.com/gp/book/9780387848570>>.
- HE, F., WANG, C., FAN, S.-K. S., 2018, “Nonlinear fault detection of batch processes based on functional kernel locality preserving projections”, *Chemometrics and Intelligent Laboratory Systems*, v. 183 (dez.), pp. 79–89. ISSN:

01697439. doi: 10.1016/j.chemolab.2018.10.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S016974391830296X>>.

HE, K., ZHANG, X., REN, S., et al., 2015, “Deep Residual Learning for Image Recognition”, *arXiv:1512.03385 [cs]*, (dez.). Disponível em: <<http://arxiv.org/abs/1512.03385>>.

HE, Q. P., WANG, J., 2007, “Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes”, *IEEE Transactions on Semiconductor Manufacturing*, v. 20, n. 4 (nov.), pp. 345–354. ISSN: 0894-6507. doi: 10.1109/TSM.2007.907607. Disponível em: <<http://ieeexplore.ieee.org/document/4369338/>>.

HE, Q. P., WANG, J., 2018, “Statistical process monitoring as a big data analytics tool for smart manufacturing”, *Journal of Process Control*, v. 67 (jul.), pp. 35–43. ISSN: 09591524. doi: 10.1016/j.jprocont.2017.06.012. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152417301257>>.

HE, Q. P., WANG, J., 2011, “Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes”, *AIChe Journal*, v. 57, n. 1 (jan.), pp. 107–121. ISSN: 00011541. doi: 10.1002/aic.12247. Disponível em: <<http://doi.wiley.com/10.1002/aic.12247>>.

HE, Q. P., QIN, S. J., WANG, J., 2005, “A new fault diagnosis method using fault directions in Fisher discriminant analysis”, *AIChe Journal*, v. 51, n. 2 (fev.), pp. 555–571. ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.10325. Disponível em: <<http://doi.wiley.com/10.1002/aic.10325>>.

HE, S.-G., HE, Z., WANG, G. A., 2013, “Online monitoring and fault identification of mean shifts in bivariate processes using decision tree learning techniques”, *Journal of Intelligent Manufacturing*, v. 24, n. 1 (fev.), pp. 25–34. ISSN: 0956-5515, 1572-8145. doi: 10.1007/s10845-011-0533-5. Disponível em: <<http://link.springer.com/10.1007/s10845-011-0533-5>>.

HE, X. B., WANG, W., YANG, Y. P., et al., 2009, “Variable-weighted Fisher discriminant analysis for process fault diagnosis”, *Journal of Process Control*, v. 19, n. 6 (jun.), pp. 923–931. ISSN: 09591524. doi: 10.1016/j.jprocont.2008.12.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095915240800190X>>.

HEO, S., LEE, J. H., 2019, “Statistical Process Monitoring of the Tennessee Eastman Process Using Parallel Autoassociative Neural Networks and a Large

- Dataset”, *Processes*, v. 7, n. 7 (jul.), pp. 411. ISSN: 2227-9717. doi: 10.3390/pr7070411. Disponível em: <<https://www.mdpi.com/2227-9717/7/7/411>>.
- HIDEN, H., 1997, “Non-linear principal components analysis using genetic programming”. In: *Second International Conference on Genetic Algorithms in Engineering Systems*, v. 1997, pp. 302–307, Glasgow, UK. IEE. ISBN: 978-0-85296-693-8. doi: 10.1049/cp:19971197. Disponível em: <https://digital-library.theiet.org/content/conferences/10.1049/cp_19971197>.
- HILAL, W., GADSDEN, S. A., YAWNEY, J., 2022, “Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances”, *Expert Systems with Applications: An International Journal*, v. 193, n. C (maio). ISSN: 0957-4174. doi: 10.1016/j.eswa.2021.116429. Disponível em: <<https://doi.org/10.1016/j.eswa.2021.116429>>.
- HIMES, D. M., STORER, R. H., GEORGAKIS, C., 1994, “Determination of the number of principal components for disturbance detection and isolation”. In: *Proceedings of 1994 American Control Conference - ACC '94*, v. 2, pp. 1279–1283, jun. doi: 10.1109/ACC.1994.752265. Disponível em: <<https://ieeexplore.ieee.org/document/752265>>.
- HIMMELBLAU, D., 1978, *Fault Detection and Diagnosis in Chemical and Petrochemical Processes*. Chemical engineering monographs. Amsterdam, Elsevier. ISBN: 9780444412959. Disponível em: <<https://books.google.com.br/books?id=mgG8oAEACAAJ>>.
- HINTON, G. E., OSINDERO, S., TEH, Y.-W., 2006, “A Fast Learning Algorithm for Deep Belief Nets”, *Neural Computation*, v. 18, n. 7 (jul.), pp. 1527–1554. ISSN: 0899-7667, 1530-888X. doi: 10.1162/neco.2006.18.7.1527. Disponível em: <<http://www.mitpressjournals.org/doi/10.1162/neco.2006.18.7.1527>>.
- HODOUIN, D., MACGREGOR, J., HOU, M., et al., 1993, “Multivariate Statistical-Analysis of Mineral processing Data”, *CIM Bulletin*, v. 86, n. 975, pp. 23–34.
- HOLOVIZ, 2022. “Datashader”. Disponível em: <<https://github.com/holoviz/datashader>>.
- HONG, H., JIANG, C., PENG, X., et al., 2020, “Concurrent Monitoring Strategy for Static and Dynamic Deviations Based on Selective Ensemble

Learning Using Slow Feature Analysis”, *Industrial & Engineering Chemistry Research*, v. 59, n. 10 (mar.), pp. 4620–4635. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.9b05547. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.9b05547>>.

HORNIK, K., STINCHCOMBE, M., WHITE, H., 1989, “Multilayer feedforward networks are universal approximators”, *Neural Networks*, v. 2, n. 5 (jan.), pp. 359–366. ISSN: 08936080. doi: 10.1016/0893-6080(89)90020-8. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0893608089900208>>.

HOSKINS, J. C., KALIYUR, K. M., HIMMELBLAU, D. M., 1991, “Fault diagnosis in complex chemical plants using artificial neural networks”, *AIChE Journal*, v. 37, n. 1 (jan.), pp. 137–141. ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.690370112. Disponível em: <<http://doi.wiley.com/10.1002/aic.690370112>>.

HOTELLING, H., 1947, “Multivariate quality control illustrated by the air testing of sample bombsites”. In: Eisenhart, C., Hastay, M., Wallis, W. (Eds.), *Techniques of Statistical Analysis*, McGraw-Hill, New York.

HOTELLING, H., 1936, “Relations Between Two Sets of Variates”, *Biometrika*, v. 28, n. 3-4, pp. 321–377. Disponível em: <<https://doi.org/10.1093/biomet/28.3-4.321>>.

HOWELL, J., CHEN, J., ZHANG, J., 1997, “The Tennessee Eastman Problem as a Process Monitoring Benchmark”, *IFAC Proceedings Volumes*, v. 30, n. 18 (ago.), pp. 223–228. ISSN: 1474-6670. doi: 10.1016/S1474-6670(17)42405-0. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1474667017424050>>.

HSU, C.-C., CHEN, M.-C., CHEN, L.-S., 2010a, “Integrating independent component analysis and support vector machine for multivariate process monitoring”, *Computers & Industrial Engineering*, v. 59, n. 1 (ago.), pp. 145–156. ISSN: 03608352. doi: 10.1016/j.cie.2010.03.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0360835210000823>>.

HSU, C.-C., CHEN, M.-C., CHEN, L.-S., 2010b, “Intelligent ICA-SVM fault detector for non-Gaussian multivariate process monitoring”, *Expert Systems with Applications*, v. 37, n. 4 (abr.), pp. 3264–3273. ISSN: 09574174. doi: 10.1016/j.eswa.2009.09.053. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417409008343>>.

- HSU, C.-C., CHEN, M.-C., CHEN, L.-S., 2010c, “A novel process monitoring approach with dynamic independent component analysis”, *Control Engineering Practice*, v. 18, n. 3 (mar.), pp. 242–253. ISSN: 09670661. doi: 10.1016/j.conengprac.2009.11.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066109002093>>.
- HU, J., ZHANG, L., CAI, Z., et al., 2015, “Fault propagation behavior study and root cause reasoning with dynamic Bayesian network based framework”, *Process Safety and Environmental Protection*, v. 97 (set.), pp. 25–36. ISSN: 09575820. doi: 10.1016/j.psep.2015.02.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957582015000166>>.
- HU, W., FEY, M., ZITNIK, M., et al., 2021, “Open Graph Benchmark: Datasets for Machine Learning on Graphs”, *arXiv:2005.00687 [cs, stat]*, (fev.). Disponível em: <<http://arxiv.org/abs/2005.00687>>.
- HUANG, G.-B., ZHU, Q.-Y., SIEW, C.-K., 2006, “Extreme learning machine: Theory and applications”, *Neurocomputing*, v. 70, n. 1-3 (dez.), pp. 489–501. ISSN: 09252312. doi: 10.1016/j.neucom.2005.12.126. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231206000385>>.
- HUANG, J., YANG, X., YAN, X., 2020a, “Slow feature analysis-independent component analysis based integrated monitoring approach for industrial processes incorporating dynamic and static characteristics”, *Control Engineering Practice*, v. 102 (set.), pp. 104558. ISSN: 09670661. doi: 10.1016/j.conengprac.2020.104558. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066120301532>>.
- HUANG, J., YAN, X., 2018, “Relevant and independent multi-block approach for plant-wide process and quality-related monitoring based on KPCA and SVDD”, *ISA Transactions*, v. 73 (fev.), pp. 257–267. ISSN: 00190578. doi: 10.1016/j.isatra.2018.01.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S001905781830003X>>.
- HUANG, K., WEN, H., JI, H., et al., 2019, “Nonlinear process monitoring using kernel dictionary learning with application to aluminum electrolysis process”, *Control Engineering Practice*, v. 89 (ago.), pp. 94–102. ISSN: 09670661. doi: 10.1016/j.conengprac.2019.05.021. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066119300760>>.
- HUANG, K., WU, Y., WEN, H., et al., 2020b, “Distributed dictionary learning for high-dimensional process monitoring”, *Control Engineering Practice*,

v. 98 (maio), pp. 104386. ISSN: 09670661. doi: 10.1016/j.conengprac. 2020.104386. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066120300526>>.

HUANG, K., WEI, K., LI, Y., et al., 2021, “Distributed dictionary learning for industrial process monitoring with big data”, *Applied Intelligence*, (mar.). ISSN: 1573-7497. doi: 10.1007/s10489-020-02128-x. Disponível em: <<https://doi.org/10.1007/s10489-020-02128-x>>.

HUANG, K., TAO, Z., LIU, Y., et al., 2022, “Adaptive Multimode Process Monitoring Based on Mode-Matching and Similarity-Preserving Dictionary Learning”, *IEEE Transactions on Cybernetics*, pp. 1–14. ISSN: 2168-2275. doi: 10.1109/TCYB.2022.3178878.

HUFFAKER, R., BITTELLI, M., ROSA, R., 2017, *Nonlinear Time Series Analysis with R*. 1 edition ed. Oxford ; New York, Oxford University Press. ISBN: 978-0-19-880825-1.

IGNOVA, M., GLASSEY, J., WARD, A., et al., 1997, “Multivariate statistical methods in bioprocess fault detection and performance forecasting”, *Transactions of the Institute of Measurement and Control*, v. 19, n. 5 (dez.), pp. 271–279. ISSN: 0142-3312, 1477-0369. doi: 10.1177/014233129701900507. Disponível em: <<http://journals.sagepub.com/doi/10.1177/014233129701900507>>.

INSELBERG, A., 2009, *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. New York, Springer-Verlag. ISBN: 978-0-387-21507-5. Disponível em: <<https://www.springer.com/gp/book/9780387215075>>.

ISERMANN, R., 2005, “Model-based fault-detection and diagnosis – status and applications”, *Annual Reviews in Control*, v. 29, n. 1 (jan.), pp. 71–85. ISSN: 13675788. doi: 10.1016/j.arcontrol.2004.12.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1367578805000052>>.

IZENMAN, A. J., 2008, *Modern multivariate statistical techniques: regression, classification, and manifold learning*. Springer texts in statistics. New York ; London, Springer. ISBN: 978-0-387-78188-4.

JACK, L. B., NANDI, A. K., 2001, “Support vector machines for detection and characterization of rolling element bearing faults”, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, v. 215, n. 9 (set.), pp. 1065–1074. ISSN: 0954-4062,

2041-2983. doi: 10.1177/095440620121500907. Disponível em: <<http://journals.sagepub.com/doi/10.1177/095440620121500907>>.

JACKSON, J. E., MUDHOLKAR, G. S., 1979, “Control Procedures for Residuals Associated with Principal Component Analysis”, *Technometrics*, v. 21, n. 3 (ago.), pp. 341–349. ISSN: 00401706. doi: 10.2307/1267757. Disponível em: <<https://www.jstor.org/stable/1267757?origin=crossref>>.

JANARDAN, MEHTA, S., 2017, “Concept drift in Streaming Data Classification: Algorithms, Platforms and Issues”, *Procedia Computer Science*, v. 122, pp. 804–811. ISSN: 18770509. doi: 10.1016/j.procs.2017.11.440. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1877050917326881>>.

JANG, K., KIM, M., YOON, H., et al., 2021, “Adversarial Autoencoder Based Nonlinear Process Monitoring”. In: Türkay, M., Gani, R. (Eds.), *Computer Aided Chemical Engineering*, v. 50, *31 European Symposium on Computer Aided Process Engineering*, Elsevier, pp. 1195–1201, jan. doi: 10.1016/B978-0-323-88506-5.50184-4. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780323885065501844>>.

JANG, K., PILARIO, K. E. S., LEE, N., et al., 2023, “Explainable Artificial Intelligence for Fault Diagnosis of Industrial Processes”, *IEEE Transactions on Industrial Informatics*, pp. 1–8. ISSN: 1941-0050. doi: 10.1109/TII.2023.3240601.

JENG, J.-C., 2010, “Adaptive process monitoring using efficient recursive PCA and moving window PCA algorithms”, *Journal of the Taiwan Institute of Chemical Engineers*, v. 41, n. 4 (jul.), pp. 475–481. ISSN: 18761070. doi: 10.1016/j.jtice.2010.03.015. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1876107010000532>>.

JEONG, D. H., LEE, J. M., 2020, “Ensemble learning based latent variable model predictive control for batch trajectory tracking under concept drift”, *Computers & Chemical Engineering*, v. 139 (ago.), pp. 106875. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106875. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419313602>>.

JI, C., MA, F., WANG, J., et al., 2023, “Profitability related industrial-scale batch processes monitoring via deep learning based soft sensor development”, *Computers & Chemical Engineering*, v. 170 (fev.), pp. 108125.

ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.108125. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422004586>>.

JI, Z., ZHANG, X., WANG, C., 2010, “Monitoring of continuous steel casting process based on independent component analysis”. In: *2010 Chinese Control and Decision Conference*, pp. 3920–3923, Xuzhou, China, maio. IEEE. ISBN: 978-1-4244-5181-4. doi: 10.1109/CCDC.2010.5498446. Disponível em: <<http://ieeexplore.ieee.org/document/5498446/>>.

JIA, F., MARTIN, E. B., MORRIS, A. J., 2000, “Non-linear principal components analysis with application to process fault detection”, *International Journal of Systems Science*, v. 31, n. 11 (jan.), pp. 1473–1487. ISSN: 0020-7721, 1464-5319. doi: 10.1080/00207720050197848. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/00207720050197848>>.

JIA, Z., ZENG, Y., ZHANG, Y., et al., 2020, “Local class-specific discriminant analysis with variable weighting and its application in fault diagnosis”, *Computers & Chemical Engineering*, v. 141 (out.), pp. 107023. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.107023. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135420302763>>.

JIANG, B., ZHU, X., HUANG, D., et al., 2015, “A combined canonical variate analysis and Fisher discriminant analysis (CVA–FDA) approach for fault diagnosis”, *Computers & Chemical Engineering*, v. 77 (jun.), pp. 1–9. ISSN: 00981354. doi: 10.1016/j.compchemeng.2015.03.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135415000721>>.

JIANG, Q., YAN, X., 2014, “Improved fault detection in nonlinear chemical processes using WKPCA-SVDD”, *Korean Journal of Chemical Engineering*, v. 31, n. 11 (nov.), pp. 1935–1942. ISSN: 0256-1115, 1975-7220. doi: 10.1007/s11814-014-0151-y. Disponível em: <<http://link.springer.com/10.1007/s11814-014-0151-y>>.

JIANG, Q., YAN, X., 2020, “Neighborhood Stable Correlation Analysis for Robust Monitoring of Multiunit Chemical Processes”, *Industrial & Engineering Chemistry Research*, v. 59, n. 38 (set.), pp. 16695–16707. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c02552. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.0c02552>>.

JIANG, Q., YAN, X., 2015, “Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA”, *Jour-*

nal of Process Control, v. 32 (ago.), pp. 38–50. ISSN: 09591524. doi: 10.1016/j.jprocont.2015.04.014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095915241500089X>>.

JIANG, Q., YAN, X., 2018, “Parallel PCA–KPCA for nonlinear process monitoring”, *Control Engineering Practice*, v. 80 (nov.), pp. 17–25. ISSN: 09670661. doi: 10.1016/j.conengprac.2018.07.012. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066118303289>>.

JIANG, Q., YAN, X., LI, J., 2016a, “PCA-ICA Integrated with Bayesian Method for Non-Gaussian Fault Diagnosis”, *Industrial & Engineering Chemistry Research*, v. 55, n. 17 (maio), pp. 4979–4986. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.5b04023. Disponível em: <<http://pubs.acs.org/doi/10.1021/acs.iecr.5b04023>>.

JIANG, Q., YAN, X., HUANG, B., 2019, “Review and Perspectives of Data-Driven Distributed Monitoring for Industrial Plant-Wide Processes”, *Industrial & Engineering Chemistry Research*, v. 58, n. 29 (jul.), pp. 12899–12912. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.9b02391. Disponível em: <<http://pubs.acs.org/doi/10.1021/acs.iecr.9b02391>>.

JIANG, S., ZAVALA, V. M., 2021, “Convolutional neural nets in chemical engineering: Foundations, computations, and applications”, *AICHE Journal*, v. 67, n. 9 (set.). ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.17282. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/aic.17282>>.

JIANG, Y., YIN, S., DONG, J., et al., 2020, “A Review on Soft Sensors for Monitoring, Control and Optimization of Industrial Processes”, *IEEE Sensors Journal*, pp. 1–1. ISSN: 1558-1748. doi: 10.1109/JSEN.2020.3033153. Disponível em: <<https://ieeexplore.ieee.org/document/9235582>>.

JIANG, Y., YIN, S., 2019, “Recent Advances in Key-Performance-Indicator Oriented Prognosis and Diagnosis With a MATLAB Toolbox: DB-KIT”, *IEEE Transactions on Industrial Informatics*, v. 15, n. 5 (maio), pp. 2849–2858. ISSN: 1551-3203, 1941-0050. doi: 10.1109/TII.2018.2875067. Disponível em: <<https://ieeexplore.ieee.org/document/8486739/>>.

JIANG, Y., YIN, S., YANG, Y., 2016b, “Comparison of KPI related fault detection algorithms using a newly developed MATLAB toolbox: DB-KIT”. In: *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics*

Society, pp. 7149–7154, Florence, Italy, out.b. IEEE. ISBN: 978-1-5090-3474-1. doi: 10.1109/IECON.2016.7792957. Disponível em: <<http://ieeexplore.ieee.org/document/7792957/>>.

JIAO, Z., HU, P., XU, H., et al., 2020, “Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications”, *ACS Chemical Health & Safety*, v. 27, n. 6 (nov.), pp. 316–334. ISSN: 1871-5532, 1878-0504. doi: 10.1021/acs.chas.0c00075. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.chas.0c00075>>.

JIN, H. D., LEE, Y.-H., LEE, G., et al., 2006, “Robust Recursive Principal Component Analysis Modeling for Adaptive Monitoring”, *Industrial & Engineering Chemistry Research*, v. 45, n. 2 (jan.), pp. 696–703. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie050850t. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie050850t>>.

JING, C., HOU, J., 2015, “SVM and PCA based fault classification approaches for complicated industrial process”, *Neurocomputing*, v. 167 (nov.), pp. 636–642. ISSN: 0925-2312. doi: 10.1016/j.neucom.2015.03.082. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231215004610>>.

JONES, M., 2019, “What we talk about when we talk about (big) data”, *The Journal of Strategic Information Systems*, v. 28, n. 1 (mar.), pp. 3–16. ISSN: 09638687. doi: 10.1016/j.jsis.2018.10.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0963868718302622>>.

JUN, H.-B., KIM, D., 2017, “A Bayesian network-based approach for fault analysis”, *Expert Systems with Applications*, v. 81 (set.), pp. 332–348. ISSN: 09574174. doi: 10.1016/j.eswa.2017.03.056. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417417302191>>.

JUTTEN, C., HERAULT, J., 1991, “Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture”, *Signal Processing*, v. 24, n. 1 (jul.), pp. 1–10. ISSN: 01651684. doi: 10.1016/0165-1684(91)90079-X. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/016516849190079X>>.

KADLEC, P., GABRYS, B., STRANDT, S., 2009, “Data-driven Soft Sensors in the process industry”, *Computers & Chemical Engineering*, v. 33, n. 4 (abr.), pp. 795–814. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2008.12.012. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135409000076>>.

- KADLEC, P., GRBIĆ, R., GABRYS, B., 2011, “Review of adaptation mechanisms for data-driven soft sensors”, *Computers & Chemical Engineering*, v. 35, n. 1 (jan.), pp. 1–24. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2010.07.034. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135410002838>>.
- KALLIAMVAKOU, E., DAMIAN, D., BLINCOE, K., et al., 2015, “Open Source-style Collaborative Development Practices in Commercial Projects Using GitHub”. In: *Proceedings of the 37th International Conference on Software Engineering - Volume 1*, ICSE ’15, pp. 574–585, Piscataway, NJ, USA. IEEE Press. ISBN: 978-1-4799-1934-5. Disponível em: <<http://dl.acm.org/citation.cfm?id=2818754.2818825>>.
- KANG, J.-L., 2020, “Visualization analysis for fault diagnosis in chemical processes using recurrent neural networks”, *Journal of the Taiwan Institute of Chemical Engineers*, v. 112 (jul.), pp. 137–151. ISSN: 18761070. doi: 10.1016/j.jtice.2020.06.016. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1876107020301668>>.
- KANO, M., NAKAGAWA, Y., 2008, “Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry”, *Computers & Chemical Engineering*, v. 32, n. 1-2 (jan.), pp. 12–24. ISSN: 00981354. doi: 10.1016/j.compchemeng.2007.07.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135407001986>>.
- KANO, M., TANAKA, S., HASEBE, S., et al., 2003, “Monitoring independent components for fault detection”, *AICHE Journal*, v. 49, n. 4 (abr.), pp. 969–976. ISSN: 00011541, 15475905. doi: 10.1002/aic.690490414. Disponível em: <<http://doi.wiley.com/10.1002/aic.690490414>>.
- KANO, M., HASEBE, S., HASHIMOTO, I., et al., 2004, “Evolution of multivariate statistical process control: application of independent component analysis and external analysis”, *Computers & Chemical Engineering*, v. 28, n. 6-7 (jun.), pp. 1157–1166. ISSN: 00981354. doi: 10.1016/j.compchemeng.2003.09.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135403002497>>.
- KANTZ, H., SCHREIBER, T., 2004, *Nonlinear Time Series Analysis*. 2 edition ed. Cambridge, UK ; New York, Cambridge University Press. ISBN: 978-0-521-52902-0.

- KAPLAN, A., HAENLEIN, M., 2019, “Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence”, *Business Horizons*, v. 62, n. 1 (jan.), pp. 15–25. ISSN: 00076813. doi: 10.1016/j.bushor.2018.08.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0007681318301393>>.
- KARIM, M., HODGE, D., SIMON, L., 2003, “Data-Based Modeling and Analysis of Bioprocesses: Some Real Experiences”, *Biotechnology Progress*, v. 19, n. 5 (out.), pp. 1591–1605. ISSN: 8756-7938. doi: 10.1021/bp015514w. Disponível em: <<http://doi.wiley.com/10.1021/bp015514w>>.
- KATHARI, S., TANGIRALA, A. K., 2022, “A Novel Framework for Causality Analysis of Deterministic Dynamical Processes”, *Industrial & Engineering Chemistry Research*, v. 61, n. 50 (dez.), pp. 18426–18444. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.2c02072. Disponível em: <<https://pubs.acs.org/doi/abs/10.1021/acs.iecr.2c02072>>.
- KATSER, I. D., KOZITSIN, V. O., MAKSIMOV, I. V., et al., 2021, “Data pre-processing methods for NPP equipment diagnostics algorithms: an overview”, *Nuclear Energy and Technology*, v. 7, n. 2 (jun.), pp. 111–125. ISSN: 2452-3038. doi: 10.3897/nucet.7.63675. Disponível em: <<https://nucet.pensoft.net/article/63675/>>.
- KAZMER, D. O., WESTERDALE, S., HAZEN, D., 2008, “A Comparison of Statistical Process Control (SPC) and On-Line Multivariate Analyses (MVA) for Injection Molding”, *International Polymer Processing*, v. 23, n. 5 (nov.), pp. 447–458. ISSN: 0930-777X, 2195-8602. doi: 10.3139/217.2192. Disponível em: <<http://www.hanser-elibrary.com/doi/abs/10.3139/217.2192>>.
- KEOGH, E., KASETTY, S., 2003, “On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration”, *Data Mining and Knowledge Discovery*, v. 7, pp. 349–371. Disponível em: <<https://link.springer.com/article/10.1023/A:1024988512476>>.
- KESAVAN, P., LEE, J. H., SAUCEDO, V., et al., 2000, “Partial least squares (PLS) based monitoring and control of batch digesters”, *Journal of Process Control*, v. 10, n. 2-3 (abr.), pp. 229–236. ISSN: 09591524. doi: 10.1016/S0959-1524(99)00028-1. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152499000281>>.
- KETELAERE, B., HUBERT, M., SCHMITT, E., 2015, “Overview of PCA-Based Statistical Process-Monitoring Methods for Time-Dependent, High-

Dimensional Data”, *Journal of Quality Technology*, v. 47, n. 4 (out.), pp. 318–335. ISSN: 0022-4065, 2575-6230. doi: 10.1080/00224065.2015.11918137. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/00224065.2015.11918137>>.

KHALID, H. M., KHOUKHI, A., AL-SUNNI, F. M., 2011, “Fault detection and classification using Kalman filter and genetic neuro-fuzzy systems”. In: *2011 Annual Meeting of the North American Fuzzy Information Processing Society*, El Paso, TX, USA, mar. IEEE. ISBN: 978-1-61284-968-3. doi: 10.1109/NAFIPS.2011.5751925. Disponível em: <<http://ieeexplore.ieee.org/document/5751925/>>.

KHALIFA, R. M., YACOUT, S., BASSETTO, S., 2023, “Root cause analysis of an out-of-control process using a logical analysis of data regression model and exponential weighted moving average”, *Journal of Intelligent Manufacturing*, (abr.). ISSN: 1572-8145. doi: 10.1007/s10845-023-02118-z. Disponível em: <<https://doi.org/10.1007/s10845-023-02118-z>>.

KHAMASSI, I., SAYED-MOUCHAWEH, M., HAMMAMI, M., et al., 2018, “Discussion and review on evolving data streams and concept drift adapting”, *Evolving Systems*, v. 9, n. 1 (mar.), pp. 1–23. ISSN: 1868-6478, 1868-6486. doi: 10.1007/s12530-016-9168-2. Disponível em: <<http://link.springer.com/10.1007/s12530-016-9168-2>>.

KHAN, N., AMMAR TAQVI, S. A., 2023, “Machine Learning an Intelligent Approach in Process Industries: A Perspective and Overview”, *Chem-BioEng Reviews*, v. 10, n. 2 (abr.), pp. 195–221. ISSN: 2196-9744, 2196-9744. doi: 10.1002/cben.202200030. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cben.202200030>>.

KHARBACH, M., CHERRAH, Y., VANDER HEYDEN, Y., et al., 2017, “Multivariate statistical process control in product quality review assessment – A case study”, *Annales Pharmaceutiques Françaises*, v. 75, n. 6 (nov.), pp. 446–454. ISSN: 00034509. doi: 10.1016/j.pharma.2017.07.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S000345091630133X>>.

KHATIBISEPEHR, S., HUANG, B., KHARE, S., 2013, “Design of inferential sensors in the process industry: A review of Bayesian methods”, *Journal of Process Control*, v. 23, n. 10 (nov.), pp. 1575–1596. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2013.05.007. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S095915241300156X>>.

- KHOUKHI, A., KHALID, M. H., 2015, “Hybrid computing techniques for fault detection and isolation, a review”, *Computers & Electrical Engineering*, v. 43 (abr.), pp. 17–32. ISSN: 00457906. doi: 10.1016/j.compeleceng.2014.12.015. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S004579061500004X>>.
- KIRAN, K. L., SELVARAJ, S., LEE, J., et al., 2012, “Application of fault monitoring and diagnostic techniques and their challenges in petrochemical industries”, *IFAC Proceedings Volumes*, v. 45, n. 15 (jan.), pp. 702–707. ISSN: 1474-6670. doi: 10.3182/20120710-4-SG-2026.00182. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1474667016305262>>.
- KLANDERMAN, M. C., NEWHART, K. B., CATH, T. Y., et al., 2020, “Case studies in real-time fault isolation in a decentralized wastewater treatment facility”, *Journal of Water Process Engineering*, v. 38 (dez.), pp. 101556. ISSN: 22147144. doi: 10.1016/j.jwpe.2020.101556. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2214714420304347>>.
- KOHONEN, T., 1990, “The self-organizing map”, *Proceedings of the IEEE*, v. 78, n. 9 (set.), pp. 1464–1480. ISSN: 1558-2256. doi: 10.1109/5.58325.
- KOK, T. L., ALDRICH, C., 2019, “Analysis of Electrochemical Noise for Corrosion Type Identification by Use of Global Recurrence Plots and Texture Analysis”, *IFAC-PapersOnLine*, v. 52, n. 14 (jan.), pp. 36–41. ISSN: 2405-8963. doi: 10.1016/j.ifacol.2019.09.160. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S240589631930792X>>.
- KOK, T. L., ALDRICH, C., ZABIRI, H., et al., 2022, “Application of unthresholded recurrence plots and texture analysis for industrial loops with faulty valves”, *Soft Computing*, v. 26, n. 19 (out.), pp. 10477–10492. ISSN: 1433-7479. doi: 10.1007/s00500-022-06894-3. Disponível em: <<https://doi.org/10.1007/s00500-022-06894-3>>.
- KONG, X., JIANG, X., ZHANG, B., et al., 2022, “Latent variable models in the era of industrial big data: Extension and beyond”, *Annual Reviews in Control*, v. 54 (jan.), pp. 167–199. ISSN: 1367-5788. doi: 10.1016/j.arcontrol.2022.09.005. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1367578822000992>>.
- KONG, X., LUO, J., FENG, X., et al., 2023, “A General Quality-Related Nonlinear Process Monitoring Approach Based on Input–Output Kernel PLS”, *IEEE*

Transactions on Instrumentation and Measurement, v. 72, pp. 1–12. ISSN: 1557-9662. doi: 10.1109/TIM.2023.3238692.

KORBICZ, J., KOWAL, M., 2007, “Neuro-fuzzy networks and their application to fault detection of dynamical systems”, *Engineering Applications of Artificial Intelligence*, v. 20, n. 5 (ago.), pp. 609–617. ISSN: 0952-1976. doi: 10.1016/j.engappai.2006.11.009. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197606002156>>.

KOSANOVICH, K. A., PIOVOSO, M. J., 1997, “PCA of wavelet transformed process data for monitoring”, *Intelligent Data Analysis*, v. 1, n. 1-4 (jan.), pp. 85–99. ISSN: 1088-467X. doi: 10.1016/S1088-467X(97)00004-8. Disponível em: <<https://www.sciencedirect.ez29.capes.proxy.ufrj.br/science/article/pii/S1088467X97000048>>.

KOSANOVICH, K., PIOVOSO, M., DAHL, K., et al., 1994, “Multi-way PCA applied to an industrial batch process”. In: *Proceedings of 1994 American Control Conference - ACC '94*, v. 2, pp. 1294–1298, Baltimore, MD, USA. IEEE. ISBN: 978-0-7803-1783-3. doi: 10.1109/ACC.1994.752268. Disponível em: <<http://ieeexplore.ieee.org/document/752268/>>.

KOSANOVICH, K. A., 1997, “Improvements in the Development of Models for Monitoring & Detection”, *IFAC Proceedings Volumes*, v. 30, n. 18 (ago.), pp. 29–34. ISSN: 14746670. doi: 10.1016/S1474-6670(17)42376-7. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017423767>>.

KOURTI, T., 2005a, “Abnormal situation detection and projection methods—industrial applications. October 28–29, 2003. Hamilton, Ontario, Canada”, *Chemometrics and Intelligent Laboratory Systems*, v. 76, n. 2 (abr.), pp. 215–220. ISSN: 01697439. doi: 10.1016/j.chemolab.2005.01.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743905000171>>.

KOURTI, T., 2002, “Process analysis and abnormal situation detection: from theory to practice”, *IEEE Control Systems*, v. 22, n. 5 (out.), pp. 10–25. ISSN: 1066-033X, 1941-000X. doi: 10.1109/MCS.2002.1035214. Disponível em: <[https://ieeexplore.ieee.org/document/1035214/](http://ieeexplore.ieee.org/document/1035214/)>.

KOURTI, T., 2019, “Pharmaceutical Manufacturing: The Role of Multivariate Analysis in Design Space, Control Strategy, Process Understanding, Troubleshooting, and Optimization”. In: am Ende, D. J., am Ende, M. T.

(Eds.), *Chemical Engineering in the Pharmaceutical Industry*, John Wiley & Sons, Inc., pp. 601–629, Hoboken, NJ, USA, abr. ISBN: 978-1-119-60080-0 978-1-119-28586-1. doi: 10.1002/9781119600800.ch75. Disponível em: <<http://doi.wiley.com/10.1002/9781119600800.ch75>>.

KOURTI, T., 2005b, “Application of latent variable methods to process control and multivariate statistical process control in industry”, *International Journal of Adaptive Control and Signal Processing*, v. 19, n. 4 (maio), pp. 213–246. ISSN: 0890-6327, 1099-1115. doi: 10.1002/acs.859. Disponível em: <<http://doi.wiley.com/10.1002/acs.859>>.

KOURTI, T., NOMIKOS, P., MACGREGOR, J. F., 1995, “Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS”, *Journal of Process Control*, v. 5, n. 4 (ago.), pp. 277–284. ISSN: 09591524. doi: 10.1016/0959-1524(95)00019-M. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/095915249500019M>>.

KOURTI, T., LEE, J., MACGREGOR, J. F., 1996, “Experiences with industrial applications of projection methods for multivariate statistical process control”, *Computers & Chemical Engineering*, v. 20 (jan.), pp. S745–S750. ISSN: 00981354. doi: 10.1016/0098-1354(96)00132-9. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0098135496001329>>.

KRAMER, M., 1992, “Autoassociative neural networks”, *Computers & Chemical Engineering*, v. 16, n. 4 (abr.), pp. 313–328. ISSN: 00981354. doi: 10.1016/0098-1354(92)80051-A. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/009813549280051A>>.

KRAMER, M. A., 1991, “Nonlinear principal component analysis using autoassociative neural networks”, *AICHE Journal*, v. 37, n. 2 (fev.), pp. 233–243. ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.690370209. Disponível em: <<http://doi.wiley.com/10.1002/aic.690370209>>.

KRAVCHIK, M., DEMETRIO, L., BIGGIO, B., et al., 2022, “Practical Evaluation of Poisoning Attacks on Online Anomaly Detectors in Industrial Control Systems”, *Computers & Security*, v. 122 (nov.), pp. 102901. ISSN: 0167-4048. doi: 10.1016/j.cose.2022.102901. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167404822002942>>.

KRAWCZYK, B., MINKU, L. L., GAMA, J., et al., 2017, “Ensemble learning for data stream analysis: A survey”, *Information Fusion*, v. 37 (set.), pp. 132–156. ISSN: 15662535. doi: 10.1016/j.inffus.2017.02.004.

Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1566253516302329>>.

KRUGER, U., XIE, L., 2012, *Statistical Monitoring of Complex Multivariate Processes: With Applications in Industrial Process Control*. Chichester, UK, John Wiley & Sons, Ltd. ISBN: 978-0-470-51725-3 978-0-470-02819-3. doi: 10.1002/9780470517253. Disponível em: <<http://doi.wiley.com/10.1002/9780470517253>>.

KU, W., STORER, R. H., GEORGAKIS, C., 1995, “Disturbance detection and isolation by dynamic principal component analysis”, *Chemometrics and Intelligent Laboratory Systems*, v. 30, n. 1 (nov.), pp. 179–196. ISSN: 0169-7439. doi: 10.1016/0169-7439(95)00076-3. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0169743995000763>>.

KUMAR, A., BHATTACHARYA, A., FLORES-CERRILLO, J., 2020, “Data-driven process monitoring and fault analysis of reformer units in hydrogen plants: Industrial application and perspectives”, *Computers & Chemical Engineering*, v. 136 (maio), pp. 106756. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106756. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419309482>>.

KUMAR, S. L., 2017, “State of The Art-Intense Review on Artificial Intelligence Systems Application in Process Planning and Manufacturing”, *Engineering Applications of Artificial Intelligence*, v. 65 (out.), pp. 294–329. ISSN: 09521976. doi: 10.1016/j.engappai.2017.08.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0952197617301896>>.

KUMAR, V., SUNDARARAJ, U., SHAH, S. L., et al., 2003, “Multivariate Statistical Monitoring of a High-Pressure Polymerization Process”, *Polymer Reaction Engineering*, v. 11, n. 4 (dez.), pp. 1017–1052. ISSN: 1054-3414, 1532-2408. doi: 10.1081/PRE-120026883. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1081/PRE-120026883>>.

KUNC, M., O'BRIEN, F. A., 2019, “The role of business analytics in supporting strategy processes: Opportunities and limitations”, *Journal of the Operational Research Society*, v. 70, n. 6 (jun.), pp. 974–985. ISSN: 0160-5682, 1476-9360. doi: 10.1080/01605682.2018.1475104. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/01605682.2018.1475104>>.

KUNCHEVA, L. I., 2013, “Change Detection in Streaming Multivariate Data Using Likelihood Detectors”, *IEEE Transactions on Knowledge and Data*

Engineering, v. 25, n. 5 (maio), pp. 1175–1180. ISSN: 1041-4347. Disponível em: <<http://doi.org/10.1109/TKDE.2011.226>>.

KUNDU, M., KUNDU, P. K., DAMARLA, S. K., 2017, *Chemometric Monitoring: Product Quality Assessment, Process Fault Detection, and Applications*. Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742, CRC Press. ISBN: 978-1-4987-8007-0 978-1-4987-8008-7. doi: 10.1201/9781315155135. Disponível em: <<https://www.taylorfrancis.com/books/9781315155135>>.

KWAK, S., MA, Y., HUANG, B., 2020, “Extracting nonstationary features for process data analytics and application in fouling detection”, *Computers & Chemical Engineering*, v. 135 (abr.), pp. 106762. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106762. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419305927>>.

LA GRASSA, R., GALLO, I., LANDRO, N., 2022, “OCmst: One-class novelty detection using convolutional neural network and minimum spanning trees”, *Pattern Recognition Letters*, v. 155 (mar.), pp. 114–120. ISSN: 0167-8655. doi: 10.1016/j.patrec.2021.11.013. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865521004049>>.

LAARNE, P., ZAIDAN, M. A., NIEMINEN, T., 2021, “ennemi: Non-linear correlation detection with mutual information”, *SoftwareX*, v. 14 (jun.), pp. 100686. ISSN: 2352-7110. doi: 10.1016/j.softx.2021.100686. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352711021000315>>.

LABORIE, F., RØED, O. C., ENGDAHL, G., et al., 2019, “Extracting Value from Data Using an Industrial Data Platform to Provide a Foundational Digital Twin”. In: *Day 1 Mon, May 06, 2019*, p. D011S010R005, Houston, Texas, abr. OTC. doi: 10.4043/29576-MS. Disponível em: <<https://onepetro.org/OTCONF/proceedings/19OTC/1-19OTC/Houston,%20Texas/181452>>.

LAKSHMINARAYAN, C., ALVARADO, A. S., PRINCIPE, J. C., et al., 2015. “Anomaly detection in streaming data”. dez. Disponível em: <<http://patents.google.com/patent/US9218527B2/en>>.

LAMPIIS, M., ANDREWS, J. D., 2009, “Bayesian belief networks for system fault diagnostics”, *Quality and Reliability Engineering International*, v. 25, n. 4 (jun.), pp. 409–426. ISSN: 07488017, 10991638. doi: 10.1002/qre.978. Disponível em: <<http://doi.wiley.com/10.1002/qre.978>>.

- LAN, T., TONG, C., SHI, X., et al., 2020, “Dynamic statistical process monitoring based on generalized canonical variate analysis”, *Journal of the Taiwan Institute of Chemical Engineers*, v. 112 (jul.), pp. 78–86. ISSN: 18761070. doi: 10.1016/j.jtice.2020.07.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1876107020301723>>.
- LANE, B., YEUNG, H., 2019, “Process Monitoring Dataset from the Additive Manufacturing Metrology Testbed (AMMT): “Three-Dimensional Scan Strategies””, *Journal of Research of the National Institute of Standards and Technology*, v. 124 (nov.), pp. 124033. ISSN: 2165-7254. doi: 10.6028/jres.124.033. Disponível em: <<https://nvlpubs.nist.gov/nistpubs/jres/124/jres.124.033.pdf>>.
- LARSSON, T., HESTETUN, K., HOVLAND, E., et al., 2001, “Self-Optimizing Control of a Large-Scale Plant: The Tennessee Eastman Process”, *Industrial & Engineering Chemistry Research*, v. 40, n. 22 (out.), pp. 4889–4901. ISSN: 0888-5885. doi: 10.1021/ie000586y. Disponível em: <<https://doi.org/10.1021/ie000586y>>.
- LAU, C. K., GHOSH, K., HUSSAIN, M. A., et al., 2013, “Fault diagnosis of Tennessee Eastman process with multi-scale PCA and ANFIS”, *Chemometrics and Intelligent Laboratory Systems*, v. 120 (jan.), pp. 1–14. ISSN: 0169-7439. doi: 10.1016/j.chemolab.2012.10.005. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169743912002080>>.
- LAVASANI, M. S., ARDALI, N. R., SOTUDEH-GHAREBAGH, R., et al., 2023, “Big data analytics opportunities for applications in process engineering”, *Reviews in Chemical Engineering*, v. 39, n. 3 (abr.), pp. 479–511. ISSN: 2191-0235. doi: 10.1515/revce-2020-0054. Disponível em: <<https://www.degruyter.com/document/doi/10.1515/revce-2020-0054/html>>.
- LAVIN, A., AHMAD, S., 2015, “Evaluating Real-Time Anomaly Detection Algorithms – The Numenta Anomaly Benchmark”. In: *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pp. 38–44, Miami, FL, USA, dez. IEEE. ISBN: 978-1-5090-0287-0. doi: 10.1109/ICMLA.2015.141. Disponível em: <<http://ieeexplore.ieee.org/document/7424283/>>.
- LAVOIE, F. B., MUTEKI, K., GOSELIN, R., 2019, “Computing meaningful models of continuous data”, *Chemical Engineering Research and Design*, v. 148 (ago.), pp. 321–335. ISSN: 02638762. doi: 10.1016/j.cherd.2019.06.

016. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0263876219302989>>.
- LAZZAROTTO, E., GRAMANI, L. M., NETO, A. C., et al., 2016, “Principal Components in Multivariate Control Charts Applied to Dara Instrumentation of Dams”, *Independent Journal of Management & Production*, v. 7, n. 1 (mar.), pp. 17–37. ISSN: 2236-269X. doi: 10.14807/ijmp.v7i1.369. Disponível em: <<http://www.ijmp.jor.br/index.php/ijmp/article/view/369>>.
- LE, T. M. K., MÄKELÄ, M., SCHREITHOFER, N., et al., 2020, “A multivariate approach for evaluation and monitoring of water quality in mining and minerals processing industry”, *Minerals Engineering*, v. 157 (out.), pp. 106582. ISSN: 08926875. doi: 10.1016/j.mineng.2020.106582. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0892687520304027>>.
- LEE, H. W., LEE, M. W., PARK, J. M., 2007, “Robust Adaptive Partial Least Squares Modeling of a Full-Scale Industrial Wastewater Treatment Process”, *Industrial & Engineering Chemistry Research*, v. 46, n. 3 (jan.), pp. 955–964. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie061094+. Disponível em: <<https://doi.org/10.1021/ie061094+>>.
- LEE, H. W., LEE, M. W., PARK, J. M., 2009, “Multi-scale extension of PLS algorithm for advanced on-line process monitoring”, *Chemometrics and Intelligent Laboratory Systems*, v. 98, n. 2 (out.), pp. 201–212. ISSN: 01697439. doi: 10.1016/j.chemolab.2009.07.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743909001415>>.
- LEE, H., LEE, J., 2015, “Online update techniques for projection based Robust Principal Component Analysis”, *ICT Express*, v. 1, n. 2 (set.), pp. 59–62. ISSN: 24059595. doi: 10.1016/j.icte.2015.09.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405959515000041>>.
- LEE, J.-M., YOO, C.-K., LEE, I., 2003, “New monitoring technique with an ICA algorithm in the wastewater treatment process”, *Water Science and Technology*, v. 47, n. 12 (jun.), pp. 49–56. ISSN: 0273-1223, 1996-9732. doi: 10.2166/wst.2003.0627. Disponível em: <<https://iwaponline.com/wst/article/47/12/49/30937/New-monitoring-technique-with-an-ICA-algorithm-in>>.
- LEE, J., DAVARI, H., SINGH, J., et al., 2018a, “Industrial Artificial Intelligence for industry 4.0-based manufacturing systems”, *Manufacturing Letters*,

- v. 18 (out.), pp. 20–23. ISSN: 22138463. doi: 10.1016/j.mfglet.2018.09.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2213846318301081>>.
- LEE, J. H., SHIN, J., REALFF, M. J., 2018b, “Machine learning: Overview of the recent progresses and implications for the process systems engineering field”, *Computers & Chemical Engineering*, v. 114 (jun.), pp. 111–121. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2017.10.008. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135417303538>>.
- LEE, J.-M., YOO, C., CHOI, S. W., et al., 2004a, “Nonlinear process monitoring using kernel principal component analysis”, *Chemical Engineering Science*, v. 59, n. 1 (jan.), pp. 223–234. ISSN: 00092509. doi: 10.1016/j.ces.2003.09.012. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0009250903004652>>.
- LEE, J.-M., YOO, C., LEE, I.-B., 2004b, “Statistical process monitoring with independent component analysis”, *Journal of Process Control*, v. 14, n. 5 (ago.), pp. 467–485. ISSN: 09591524. doi: 10.1016/j.jprocont.2003.09.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152403000994>>.
- LEE, J.-M., YOO, C., LEE, I.-B., 2004c, “Statistical monitoring of dynamic processes based on dynamic independent component analysis”, *Chemical Engineering Science*, v. 59, n. 14 (jul.), pp. 2995–3006. ISSN: 00092509. doi: 10.1016/j.ces.2004.04.031. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0009250904002660>>.
- LEE, S., KIM, S. B., 2018, “Time-adaptive support vector data description for nonstationary process monitoring”, *Engineering Applications of Artificial Intelligence*, v. 68 (fev.), pp. 18–31. ISSN: 09521976. doi: 10.1016/j.engappai.2017.10.016. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0952197617302683>>.
- LEJEUNE, M., LOZIN, V., LOZINA, I., et al., 2019, “Recent advances in the theory and practice of Logical Analysis of Data”, *European Journal of Operational Research*, v. 275, n. 1 (maio), pp. 1–15. ISSN: 03772217. doi: 10.1016/j.ejor.2018.06.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0377221718305344>>.
- LEMOS, A., CAMINHAS, W., GOMIDE, F., 2013, “Adaptive fault detection and diagnosis using an evolving fuzzy classifier”, *Information Sciences*,

- v. 220 (jan.), pp. 64–85. ISSN: 00200255. doi: 10.1016/j.ins.2011.08.030. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S002002551100449X>>.
- LEMOS, T., CAMPOS, L. F., MELO, A., et al., 2021, “Echo State network based soft sensor for Monitoring and Fault Detection of Industrial Processes”, *Computers & Chemical Engineering*, v. 155 (dez.), pp. 107512. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107512. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135421002908>>.
- LENNOX, B., MONTAGUE, G., HIDEN, H., et al., 2001, “Process monitoring of an industrial fed-batch fermentation”, *Biotechnology and Bioengineering*, v. 74, n. 2 (jul.), pp. 125–135. ISSN: 0006-3592, 1097-0290. doi: 10.1002/bit.1102. Disponível em: <<http://doi.wiley.com/10.1002/bit.1102>>.
- LI, B.-H., HOU, B.-C., YU, W.-T., et al., 2017a, “Applications of artificial intelligence in intelligent manufacturing: a review”, *Frontiers of Information Technology & Electronic Engineering*, v. 18, n. 1 (jan.), pp. 86–96. ISSN: 2095-9184, 2095-9230. doi: 10.1631/FITEE.1601885. Disponível em: <<http://link.springer.com/10.1631/FITEE.1601885>>.
- LI, C., ZHOU, Z., WEN, C., et al., 2022a, “Fault Detection of Non-Gaussian and Nonlinear Processes Based on Independent Slow Feature Analysis”, *ACS Omega*, v. 7, n. 8 (mar.), pp. 6978–6990. ISSN: 2470-1343, 2470-1343. doi: 10.1021/acsomega.1c06649. Disponível em: <<https://pubs.acs.org/doi/10.1021/acsomega.1c06649>>.
- LI, D., WANG, H., ZHOU, J., 2020a, “Novelty Detection for Multimode Process Using GANs with Learning Disentangled Representation”. In: *2020 Chinese Control And Decision Conference (CCDC)*, pp. 2536–2541, Hefei, China, ago.a. IEEE. ISBN: 978-1-72815-855-6. doi: 10.1109/CCDC49329.2020.9164262. Disponível em: <<https://ieeexplore.ieee.org/document/9164262/>>.
- LI, D., LIU, J., MA, X., et al., 2023a, “Stacked supervised auto-encoder with graph regularization for feature extraction and fault classification in chemical processes”, *Journal of Process Control*, v. 127 (jul.), pp. 102999. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2023.102999. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152423000781>>.
- LI, G., QIN, S. J., 2016, “Comparative study on monitoring schemes for non-Gaussian distributed processes”, *Journal of Process Control*, v. 67, pp. 69–82. ISSN:

09591524. doi: 10.1016/j.jprocont.2016.08.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152416301081>>.

LI, G., LIU, B., QIN, S. J., et al., 2011, “Dynamic latent variable modeling for statistical process monitoring”, *IFAC Proceedings Volumes*, v. 44, n. 1 (jan.), pp. 12886–12891. ISSN: 14746670. doi: 10.3182/20110828-6-IT-1002.00934. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667016456909>>.

LI, G., QIN, S. J., ZHOU, D., 2014, “A New Method of Dynamic Latent-Variable Modeling for Process Monitoring”, *IEEE Transactions on Industrial Electronics*, v. 61, n. 11 (nov.), pp. 6438–6445. ISSN: 0278-0046, 1557-9948. doi: 10.1109/TIE.2014.2301761. Disponível em: <<http://ieeexplore.ieee.org/document/6718065/>>.

LI, G., QIN, S. J., YUAN, T., 2016, “Data-driven root cause diagnosis of faults in process industries”, *Chemometrics and Intelligent Laboratory Systems*, v. 159 (dez.), pp. 1–11. ISSN: 01697439. doi: 10.1016/j.chemolab.2016.09.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743916303203>>.

LI, J., QIN, S. J., 2023, “Applying and dissecting LSTM neural networks and regularized learning for dynamic inferential modeling”, *Computers & Chemical Engineering*, (abr.), pp. 108264. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2023.108264. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135423001345>>.

LI, Q., PAN, F., ZHAO, Z., 2017b, “Concurrent probabilistic PLS regression model and its applications in process monitoring”, *Chemometrics and Intelligent Laboratory Systems*, v. 171 (dez.), pp. 40–54. ISSN: 0169-7439. doi: 10.1016/j.chemolab.2017.09.015. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169743916304531>>.

LI, Q., PAN, F., ZHAO, Z., et al., 2018a, “Process Modeling and Monitoring With Incomplete Data Based on Robust Probabilistic Partial Least Square Method”, *IEEE Access*, v. 6, pp. 10160–10168. ISSN: 2169-3536. doi: 10.1109/ACCESS.2018.2810079.

LI, Q., ZHAO, Z., WANG, L., 2023b, “Slowness or Autocorrelation? A serial correlation feature analysis method and its application in process monitoring”, *Journal of Process Control*, v. 121 (jan.), pp. 1–12. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2022.11.010. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152422002128>>.

- LI, S., LUO, J., HU, Y., 2020b, “Semi-supervised process fault classification based on convolutional ladder network with local and global feature fusion”, *Computers & Chemical Engineering*, v. 140 (set.), pp. 106843. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106843. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419309901>>.
- LI, S., LUO, J., HU, Y., 2022b, “Nonlinear process modeling via unidimensional convolutional neural networks with self-attention on global and local inter-variable structures and its application to process monitoring”, *ISA Transactions*, v. 121 (fev.), pp. 105–118. ISSN: 0019-0578. doi: 10.1016/j.isatra.2021.04.014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0019057821002111>>.
- LI, S., LUO, J., HU, Y., 2022c, “Toward Interpretable Process Monitoring: Slow Feature Analysis-Aided Autoencoder for Spatiotemporal Process Feature Learning”, *IEEE Transactions on Instrumentation and Measurement*, v. 71, pp. 1–11. ISSN: 1557-9662. doi: 10.1109/TIM.2021.3127284.
- LI, S., ZHOU, X., PAN, F., et al., 2017c, “Correlated and weakly correlated fault detection based on variable division and ICA”, *Computers & Industrial Engineering*, v. 112 (out.), pp. 320–335. ISSN: 03608352. doi: 10.1016/j.cie.2017.08.035. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0360835217303947>>.
- LI, W., YUE, H. H., VALLE-CERVANTES, S., et al., 2000, “Recursive PCA for adaptive process monitoring”, *Journal of Process Control*, v. 10, pp. 471–486. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152400000226>>.
- LI, W., LI, H., GU, S., et al., 2020c, “Process fault diagnosis with model- and knowledge-based approaches: Advances and opportunities”, *Control Engineering Practice*, v. 105 (dez.), pp. 104637. ISSN: 09670661. doi: 10.1016/j.conengprac.2020.104637. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066120302070>>.
- LI, W., ZHAO, C., GAO, F., 2018b, “Linearity Evaluation and Variable Subset Partition Based Hierarchical Process Modeling and Monitoring”, *IEEE Transactions on Industrial Electronics*, v. 65, n. 3 (mar.), pp. 2683–2692. ISSN: 0278-0046, 1557-9948. doi: 10.1109/TIE.2017.2745452. Disponível em: <<http://ieeexplore.ieee.org/document/8017442/>>.
- LI, X., LIN, T., MBA, D., 2022d, “A Tutorial on Canonical Variate Analysis for Diagnosis and Prognosis”. In: Chaari, F., Chiementin, X., Zimroz, R., et al.

(Eds.), *Smart Monitoring of Rotating Machinery for Industry 4.0*, Applied Condition Monitoring, Springer International Publishing, pp. 11–31, Cham, d. ISBN: 9783030795191. doi: 10.1007/978-3-030-79519-1_2. Disponível em: <https://doi.org/10.1007/978-3-030-79519-1_2>.

LI, X., XUE, F., QIN, L., et al., 2020d, “A recursively updated Map-Reduce based PCA for monitoring the time-varying fluorochemical engineering processes with big data”, *Chemometrics and Intelligent Laboratory Systems*, v. 206 (nov.), pp. 104167. ISSN: 01697439. doi: 10.1016/j.chemolab.2020.104167. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743920304007>>.

LI, Y., GE, T., CHEN, C., 2020e, “Data Stream Event Prediction Based on Timing Knowledge and State Transitions”. In: *Proceedings of the VLDB Endowment*, v. 13, e. Disponível em: <<http://www.vldb.org/pvldb/vol13/p1779-li.pdf>>.

LI, Y., MA, F., JI, C., et al., 2022e, “Fault Detection Method Based on Global-Local Marginal Discriminant Preserving Projection for Chemical Process”, *Processes*, v. 10, n. 1 (jan.), pp. 122. ISSN: 2227-9717. doi: 10.3390/pr10010122. Disponível em: <<https://www.mdpi.com/2227-9717/10/1/122>>.

LI, Z., YU, Y., PAN, X., et al., 2020f, “Effect of dataset size on modeling and monitoring of chemical processes”, *Chemical Engineering Science*, v. 227 (dez.), pp. 115928. ISSN: 00092509. doi: 10.1016/j.ces.2020.115928. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0009250920304607>>.

LIAN, V., VAROY, E., GIACAMAN, N., 2022, “Learning Object-Oriented Programming Concepts Through Visual Analogies”, *IEEE Transactions on Learning Technologies*, v. 15, n. 1 (fev.), pp. 78–92. ISSN: 1939-1382. doi: 10.1109/TLT.2022.3154805. Disponível em: <<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9722963>>.

LIM, S. A., ANTONY, J., 2019, *Statistical Process Control for the Food Industry: A Guide for Practitioners and Managers*. 1 edition ed. Hoboken, NJ, Wiley. ISBN: 978-1-119-15198-2.

LIU, H., BUVAT, J.-C., ESTEL, L., et al., 2010, “Bayesian Network Method for Fault Diagnosis in a Continuous Tubular Reactor”, *Chemical Product and Process Modeling*, v. 5, n. 1 (jan.). ISSN: 1934-2659. doi: 10.2202/

- 1934-2659.1491. Disponível em: <<https://www.degruyter.com/view/j/cppm.2010.5.1/cppm.2010.5.1.1491/cppm.2010.5.1.1491.xml>>.
- LIU, H., ZHANG, H., ZHANG, Y., et al., 2020a, “Modeling of Wastewater Treatment Processes Using Dynamic Bayesian Networks Based on Fuzzy PLS”, *IEEE Access*, pp. 1–1. ISSN: 2169-3536. doi: 10.1109/ACCESS.2020.2995068. Disponível em: <<https://ieeexplore.ieee.org/document/9094234/>>.
- LIU, H., YANG, J., ZHANG, Y., et al., 2021a, “Monitoring of wastewater treatment processes using dynamic concurrent kernel partial least squares”, *Process Safety and Environmental Protection*, v. 147 (mar.), pp. 274–282. ISSN: 09575820. doi: 10.1016/j.psep.2020.09.034. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957582020317511>>.
- LIU, J., WONG, D. S. H., CHEN, D.-S., 2014a, “Bayesian filtering of the smearing effect: Fault isolation in chemical process monitoring”, *Journal of Process Control*, v. 24, n. 3 (mar.), pp. 1–21. ISSN: 09591524. doi: 10.1016/j.jprocont.2013.12.018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152414000043>>.
- LIU, J., HOU, L., WANG, X., et al., 2021b, “Explainable fault diagnosis of gas-liquid separator based on fully convolutional neural network”, *Computers & Chemical Engineering*, v. 155 (dez.), pp. 107535. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107535. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135421003136>>.
- LIU, J., MU, G., CHEN, J., 2023a, “Tensor slow feature analysis and its applications for batch process monitoring”, *Computers & Chemical Engineering*, v. 173 (maio), pp. 108207. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2023.108207. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135423000765>>.
- LIU, J., WU, J., XIE, Y., et al., 2022a, “Toward robust process monitoring of complex process industries based on denoising sparse auto-encoder”, *Journal of Industrial Information Integration*, v. 30 (nov.), pp. 100410. ISSN: 2452-414X. doi: 10.1016/j.jii.2022.100410. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2452414X22000772>>.
- LIU, L., ZHAO, H., HU, Z., 2022b, “Graph dynamic autoencoder for fault detection”, *Chemical Engineering Science*, v. 254 (jun.), pp. 117637. ISSN: 0009-2509. doi: 10.1016/j.ces.2022.117637. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0009250922002214>>.

- LIU, N., HU, M., WANG, J., et al., 2022c, “Fault detection and diagnosis using Bayesian network model combining mechanism correlation analysis and process data: Application to unmonitored root cause variables type faults”, *Process Safety and Environmental Protection*, v. 164 (ago.), pp. 15–29. ISSN: 0957-5820. doi: 10.1016/j.psep.2022.05.073. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957582022004761>>.
- LIU, Q., QIN, S. J., CHAI, T., 2014b, “Quality-Relevant Monitoring and Diagnosis with Dynamic Concurrent Projection to Latent Structures”, *IFAC Proceedings Volumes*, v. 47, n. 3, pp. 2740–2745. ISSN: 14746670. doi: 10.3182/20140824-6-ZA-1003.02329. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667016420240>>.
- LIU, R., YANG, B., ZIO, E., et al., 2018, “Artificial intelligence for fault diagnosis of rotating machinery: A review”, *Mechanical Systems and Signal Processing*, v. 108 (ago.), pp. 33–47. ISSN: 08883270. doi: 10.1016/j.ymssp.2018.02.016. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0888327018300748>>.
- LIU, S., LEI, F., ZHAO, D., et al., 2023b, “Abnormal Situation Management in Chemical Processes: Recent Research Progress and Future Prospects”, *Processes*, v. 11, n. 6 (jun.), pp. 1608. ISSN: 2227-9717. doi: 10.3390/pr11061608. Disponível em: <<https://www.mdpi.com/2227-9717/11/6/1608>>.
- LIU, X., MIAO, X., JIANG, H., et al., 2020b, “Data analysis in visual power line inspection: An in-depth review of deep learning for component detection and fault diagnosis”, *Annual Reviews in Control*, v. 50, pp. 253–277. ISSN: 13675788. doi: 10.1016/j.arcontrol.2020.09.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1367578820300596>>.
- LIU, X., KRUGER, U., LITTLER, T., et al., 2009, “Moving window kernel PCA for adaptive monitoring of nonlinear processes”, *Chemometrics and Intelligent Laboratory Systems*, v. 96, n. 2 (abr.), pp. 132–143. ISSN: 01697439. doi: 10.1016/j.chemolab.2009.01.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743909000033>>.
- LIU, X., LI, K., MCAFEE, M., et al., 2011, “Improved nonlinear PCA for process monitoring using support vector data description”, *Journal of Process Control*, v. 21, n. 9 (out.), pp. 1306–1317. ISSN: 09591524. doi: 10.1016/j.jprocont.2011.07.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152411001442>>.

- LIU, Y., XIE, M., 2020, “Rebooting data-driven soft-sensors in process industries: A review of kernel methods”, *Journal of Process Control*, v. 89 (maio), pp. 58–73. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2020.03.012. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152420300330>>.
- LIU, Y., GE, Z., 2019, “Deep ensemble forests for industrial fault classification”, *IFAC Journal of Systems and Control*, v. 10 (dez.), pp. 100071. ISSN: 24686018. doi: 10.1016/j.ifacsc.2019.100071. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S246860181930080X>>.
- LIU, Y., GE, Z., 2018, “Weighted random forests for fault classification in industrial processes with hierarchical clustering model selection”, *Journal of Process Control*, v. 64 (abr.), pp. 62–70. ISSN: 09591524. doi: 10.1016/j.jprocont.2018.02.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152418300271>>.
- LOMOV, I., LYUBIMOV, M., MAKAROV, I., et al., 2021, “Fault detection in Tennessee Eastman process with temporal deep learning models”, *Journal of Industrial Information Integration*, v. 23 (set.), pp. 100216. ISSN: 2452-414X. doi: 10.1016/j.jii.2021.100216. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2452414X21000145>>.
- LONGO, F., NICOLETTI, L., PADOVANO, A., 2017, “Smart operators in industry 4.0: A human-centered approach to enhance operators’ capabilities and competencies within the new smart factory context”, *Computers & Industrial Engineering*, v. 113 (nov.), pp. 144–159. ISSN: 0360-8352. doi: 10.1016/j.cie.2017.09.016. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360835217304291>>.
- LORENA, A. C., CARVALHO, A. C., 2003, *Introdução às Máquinas de Votores Suporte*. Relatório Técnico 192, Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo, São Carlos, abr. Disponível em: <http://conteudo.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_192.pdf>.
- LOU, C., ATOUI, M. A., LI, X., 2022, “Novel online discriminant analysis based schemes to deal with observations from known and new classes: Application to industrial systems”, *Engineering Applications of Artificial Intelligence*, v. 111 (maio), pp. 104811. ISSN: 0952-1976. doi: 10.1016/j.engappai.2022.104811. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S095219762200080X>>.

- LOU, S., WU, P., GUO, L., et al., 2020, “Dynamic process monitoring using dynamic latent variable and canonical correlation analysis model”, *The Canadian Journal of Chemical Engineering*, (nov.). ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.23923. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.23923>>.
- LOU, S., WU, P., YANG, C., et al., 2023, “Structured fault information-aided canonical variate analysis model for dynamic process monitoring”, *Journal of Process Control*, v. 124 (abr.), pp. 54–69. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2023.01.011. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152423000112>>.
- LU, N., YAO, Y., GAO, F., et al., 2005, “Two-dimensional dynamic PCA for batch process monitoring”, *AIChE Journal*, v. 51, n. 12 (dez.), pp. 3300–3304. ISSN: 00011541. doi: 10.1002/aic.10568. Disponível em: <<http://doi.wiley.com/10.1002/aic.10568>>.
- LU, S., HUANG, S., 2020, “Segmentation of Multivariate Industrial Time Series Data Based on Dynamic Latent Variable Predictability”, *IEEE Access*, v. 8, pp. 112092–112103. ISSN: 2169-3536. doi: 10.1109/ACCESS.2020.3002257. Disponível em: <<https://ieeexplore.ieee.org/document/9116988>>.
- LU, W., YAN, X., 2020, “Deep fisher autoencoder combined with self-organizing map for visual industrial process monitoring”, *Journal of Manufacturing Systems*, v. 56 (jul.), pp. 241–251. ISSN: 0278-6125. doi: 10.1016/j.jmsy.2020.05.005. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0278612520300698>>.
- LU, W., YAN, X., 2021, “Deep model based on mode elimination and Fisher criterion combined with self-organizing map for visual multimodal chemical process monitoring”, *Information Sciences*, (fev.), pp. S0020025521000748. ISSN: 00200255. doi: 10.1016/j.ins.2021.01.036. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0020025521000748>>.
- LU, Y., 2017, “Industry 4.0: A survey on technologies, applications and open research issues”, *Journal of Industrial Information Integration*, v. 6 (jun.), pp. 1–10. ISSN: 2452414X. doi: 10.1016/j.jii.2017.04.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2452414X17300043>>.
- LUCKE, M., MEI, X., STIEF, A., et al., 2019, “Variable Selection for Fault Detection and Identification based on Mutual Information of Alarm Series”, *IFAC-PapersOnLine*, v. 52, n. 1, pp. 673–678. ISSN: 24058963. doi: 10.1016/

j.ifacol.2019.06.140. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405896319302277>>.

LUCKE, M., STIEF, A., CHIOUA, M., et al., 2020, “Fault detection and identification combining process measurements and statistical alarms”, *Control Engineering Practice*, v. 94 (jan.), pp. 104195. ISSN: 09670661. doi: 10.1016/j.conengprac.2019.104195. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066119301753>>.

LUGHOFER, E., ZAVOIANU, A.-C., POLLAK, R., et al., 2020, “On-line anomaly detection with advanced independent component analysis of multi-variate residual signals from causal relation networks”, *Information Sciences*, v. 537 (out.), pp. 425–451. ISSN: 00200255. doi: 10.1016/j.ins.2020.06.034. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0020025520306204>>.

LUO, L., BAO, S., MAO, J., et al., 2016, “Nonlocal and local structure preserving projection and its application to fault detection”, *Chemometrics and Intelligent Laboratory Systems*, v. 157 (out.), pp. 177–188. ISSN: 01697439. doi: 10.1016/j.chemolab.2016.07.014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S016974391630185X>>.

LYMAN, P., GEORGAKIS, C., 1995, “Plant-wide control of the Tennessee Eastman problem”, *Computers & Chemical Engineering*, v. 19, n. 3 (mar.), pp. 321–331. ISSN: 00981354. doi: 10.1016/0098-1354(94)00057-U. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/009813549400057U>>.

LYU, Y., ZHOU, L., CONG, Y., et al., 2023, “Multirate Mixture Probability Principal Component Analysis for Process Monitoring in Multimode Processes”, *IEEE Transactions on Automation Science and Engineering*, pp. 1–12. ISSN: 1558-3783. doi: 10.1109/TASE.2023.3253285.

LÓPEZ, C. A. M., BHONSALE, S., PEETERS, K., et al., 2020, “Manifold Learning and Clustering for Automated Phase Identification and Alignment in Data Driven Modeling of Batch Processes”, *Frontiers in Chemical Engineering*, v. 2 (nov.). ISSN: 2673-2718. doi: 10.3389/fceng.2020.582126. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fceng.2020.582126/full>>.

LÓPEZ-GUAJARDO, E. A., DELGADO-LICONA, F., ÁLVAREZ, A. J., et al., 2022, “Process intensification 4.0: A new approach for attaining new,

sustainable and circular processes enabled by machine learning”, *Chemical Engineering and Processing - Process Intensification*, v. 180 (out.), pp. 108671. ISSN: 0255-2701. doi: 10.1016/j.cep.2021.108671. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0255270121003597>>.

MA, F., JI, C., XU, M., et al., 2023, “Spatial Correlation Extraction for Chemical Process Fault Detection Using Image Enhancement Technique aided Convolutional Autoencoder”, *Chemical Engineering Science*, (maio), pp. 118900. ISSN: 0009-2509. doi: 10.1016/j.ces.2023.118900. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0009250923004566>>.

MA, J., JIANG, J., 2011, “Applications of fault detection and diagnosis methods in nuclear power plants: A review”, *Progress in Nuclear Energy*, v. 53, n. 3 (abr.), pp. 255–266. ISSN: 01491970. doi: 10.1016/j.pnucene.2010.12.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0149197010001769>>.

MA, L., DONG, J., PENG, K., et al., 2017, “A novel data-based quality-related fault diagnosis scheme for fault detection and root cause diagnosis with application to hot strip mill process”, *Control Engineering Practice*, v. 67 (out.), pp. 43–51. ISSN: 09670661. doi: 10.1016/j.conengprac.2017.07.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066117301569>>.

MA, L., DING, Y., WANG, Z., et al., 2021, “An interpretable data augmentation scheme for machine fault diagnosis based on a sparsity-constrained generative adversarial network”, *Expert Systems with Applications*, v. 182 (nov.), pp. 115234. ISSN: 0957-4174. doi: 10.1016/j.eswa.2021.115234. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417421006667>>.

MACGREGOR, J. F., KOURTI, T., 1995, “Statistical process control of multivariate processes”, *Control Engineering Practice*, v. 3, n. 3 (mar.), pp. 403–414. ISSN: 0967-0661. doi: 10.1016/0967-0661(95)00014-L. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/096706619500014L>>.

MACGREGOR, J., CINAR, A., 2012, “Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods”, *Computers & Chemical Engineering*, v. 47 (dez.), pp. 111–120. ISSN: 00981354. doi: 10.

1016/j.compchemeng.2012.06.017. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135412001937>>.

MACGREGOR, J. F., JAECKLE, C., KIPARISSIDES, C., et al., 1994, “Process monitoring and diagnosis by multiblock PLS methods”, *AICHE Journal*, v. 40, n. 5 (maio), pp. 826–838. ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.690400509. Disponível em: <<http://doi.wiley.com/10.1002/aic.690400509>>.

MAESTRI, M., ZIELLA, D., CASSANELLO, M., et al., 2014, “Automatic qualitative trend simulation method for diagnosing faults in industrial processes”, *Computers & Chemical Engineering*, v. 64 (maio), pp. 55–62. ISSN: 00981354. doi: 10.1016/j.compchemeng.2014.01.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S009813541400009X>>.

MAHADEVAN, S., SHAH, S. L., 2009, “Fault detection and diagnosis in process data using one-class support vector machines”, *Journal of Process Control*, v. 19, n. 10 (dez.), pp. 1627–1639. ISSN: 09591524. doi: 10.1016/j.jprocont.2009.07.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152409001334>>.

MAHALANOBIS, P. C., 1936, “On the generalized distance in Statistics”, *Proceedings of the National Institute of Science of India*, (abr.). Disponível em: <http://library.isical.ac.in:8080/jspui/bitstream/123456789/6765/1/Vol02_1936_1_Art05-pcm.pdf>.

MAIRAL, J., BACH, F., PONCE, J., et al., 2008, “Supervised Dictionary Learning”, *arXiv:0809.3083 [cs]*, (set.). Disponível em: <<http://arxiv.org/abs/0809.3083>>.

MAJID, N. A. A., TAYLOR, M. P., CHEN, J. J., et al., 2011, “Multivariate statistical monitoring of the aluminium smelting process”, *Computers & Chemical Engineering*, v. 35, n. 11 (nov.), pp. 2457–2468. ISSN: 00981354. doi: 10.1016/j.compchemeng.2011.03.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135411000913>>.

MAJID, N. A. A., TAYLOR, M. P., CHEN, J. J. J. J., et al., 2012, “Diagnosing faults in aluminium processing by using multivariate statistical approaches”, *Journal of Materials Science*, v. 47, n. 3 (fev.), pp. 1268–1279. ISSN: 0022-2461, 1573-4803. doi: 10.1007/s10853-011-5884-y. Disponível em: <<http://link.springer.com/10.1007/s10853-011-5884-y>>.

- MAJID, N. A. A., TAYLOR, M. P., CHEN, J. J. J., et al., 2015, “Aluminium Process Fault Detection and Diagnosis”, *Advances in Materials Science and Engineering*, v. 2015, pp. 1–11. ISSN: 1687-8434, 1687-8442. doi: 10.1155/2015/682786. Disponível em: <<http://www.hindawi.com/journals/amse/2015/682786/>>.
- MANCA, G., 2020, “Tennessee-Eastman-Process Alarm Management Dataset”, (nov.). doi: 10.21227/326k-qr90. Disponível em: <<https://ieee-dataport.org/open-access/tennessee-eastman-process-alarm-management-dataset>>.
- MARKL, D., WAHL, P. R., MENEZES, J. C., et al., 2013, “Supervisory Control System for Monitoring a Pharmaceutical Hot Melt Extrusion Process”, *AAPS PharmSciTech*, v. 14, n. 3 (set.), pp. 1034–1044. ISSN: 1530-9932. doi: 10.1208/s12249-013-9992-7. Disponível em: <<http://link.springer.com/10.1208/s12249-013-9992-7>>.
- MARKOU, M., SINGH, S., 2003a, “Novelty detection: a review—part 1: statistical approaches”, *Signal Processing*, v. 83, n. 12 (dez.), pp. 2481–2497. ISSN: 01651684. doi: 10.1016/j.sigpro.2003.07.018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0165168403002020>>.
- MARKOU, M., SINGH, S., 2003b, “Novelty detection: a review—part 2:”, *Signal Processing*, v. 83, n. 12 (dez.), pp. 2499–2521. ISSN: 01651684. doi: 10.1016/j.sigpro.2003.07.019. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0165168403002032>>.
- MARQUARDT, W., MORBACH, J., WIESNER, A., et al., 2010, *OntoCAPE: a re-usable ontology for chemical process engineering*. RWTH edition. Heidelberg ; New York, Springer. ISBN: 978-3-642-04654-4 978-3-642-04655-1.
- MARTIN, E., MORRIS, A., 1996, “Non-parametric confidence bounds for process performance monitoring charts”, *Journal of Process Control*, v. 6, n. 6 (dez.), pp. 349–358. ISSN: 09591524. doi: 10.1016/0959-1524(96)00010-8. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0959152496000108>>.
- MARTIN, E., MORRIS, A., PAPAZOGLOU, M., et al., 1996, “Batch process monitoring for consistent production”, *Computers & Chemical Engineering*, v. 20 (jan.), pp. S599–S604. ISSN: 00981354. doi: 10.1016/0098-1354(96)00109-3. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0098135496001093>>.

- MARTIN, E., MORRIS, A., KIPARISSIDES, C., 1998, "Multivariate Statistical Process Control and Process Performance Monitoring", *IFAC Proceedings Volumes*, v. 31, n. 11 (jun.), pp. 347–356. ISSN: 14746670. doi: 10.1016/S1474-6670(17)44952-4. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017449524>>.
- MARTIN-VILLALBA, C., URQUIA, A., SHAO, G., 2018, "Implementations of the Tennessee Eastman Process in Modelica", *IFAC-PapersOnLine*, v. 51, n. 2, pp. 619–624. ISSN: 24058963. doi: 10.1016/j.ifacol.2018.03.105. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405896318301095>>.
- MASON, R. L., YOUNG, J. C., 2002, *Multivariate Statistical Process Control with Industrial Applications*. Philadelphia, SIAM. ISBN: 978-0-89871-846-1.
- MAURYA, M. R., RENGASWAMY, R., VENKATASUBRAMANIAN, V., 2007, "Fault diagnosis using dynamic trend analysis: A review and recent developments", *Engineering Applications of Artificial Intelligence*, v. 20, n. 2 (mar.), pp. 133–146. ISSN: 09521976. doi: 10.1016/j.engappai.2006.06.020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0952197606001126>>.
- MAURYA, M. R., PARITOSH, P. K., RENGASWAMY, R., et al., 2010, "A framework for on-line trend extraction and fault diagnosis", *Engineering Applications of Artificial Intelligence*, v. 23, n. 6 (set.), pp. 950–960. ISSN: 09521976. doi: 10.1016/j.engappai.2010.01.027. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0952197610000680>>.
- MAURYA, M., RENGASWAMY, R., VENKATASUBRAMANIAN, V., 2005, "Fault Diagnosis by Qualitative Trend Analysis of the Principal Components", *Chemical Engineering Research and Design*, v. 83, n. 9 (set.), pp. 1122–1132. ISSN: 02638762. doi: 10.1205/cherd.04280. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0263876205728094>>.
- MAYO, M., 2018. "The Essence of Machine Learning". dez. Disponível em: <<https://www.kdnuggets.com/2018/12/essence-machine-learning.html>>.
- MCAVOY, T., YE, N., GANG, C., 1995, "An improved base control for the Tennessee Eastman problem". In: *Proceedings of 1995 American Control Conference - ACC'95*, v. 1, pp. 240–244 vol.1, jun. doi: 10.1109/ACC.1995.529245.
- MCCANN, M., LI, Y., MAGUIRE, L., et al., 2010, "Causality Challenge: Benchmarking relevant signal components for effective monitoring and pro-

cess control". In: *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, pp. 277–288. PMLR, fev. Disponível em: <<https://proceedings.mlr.press/v6/mccann10a.html>>.

MD NOR, N., HUSSAIN, M. A., HASSAN, C. R. C., 2015, "Process Monitoring and Fault Detection in Non-Linear Chemical Process Based On Multi-Scale Kernel Fisher Discriminant Analysis". In: *Computer Aided Chemical Engineering*, v. 37, Elsevier, pp. 1823–1828. ISBN: 978-0-444-63429-0. doi: 10.1016/B978-0-444-63577-8.50149-2. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/B9780444635778501492>>.

MD NOR, N., HUSSAIN, M. A., CHE HASSAN, C. R., 2017, "Fault diagnosis and classification framework using multi-scale classification based on kernel Fisher discriminant analysis for chemical process system", *Applied Soft Computing*, v. 61 (dez.), pp. 959–972. ISSN: 15684946. doi: 10.1016/j.asoc.2017.09.019. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1568494617305586>>.

MD NOR, N., CHE HASSAN, C. R., HUSSAIN, M. A., 2019, "A review of data-driven fault detection and diagnosis methods: applications in chemical process systems", *Reviews in Chemical Engineering*, (jan.). ISSN: 2191-0235, 0167-8299. doi: 10.1515/revce-2017-0069. Disponível em: <<http://www.degruyter.com/view/j/revce.ahead-of-print/revce-2017-0069/revce-2017-0069.xml>>.

MELO, A., 2023. "KydLIB: Know Your Data Library". Disponível em: <<https://github.com/afraniomelo/KydLIB>>.

MELO, A., CÂMARA, M. M., CLAVIJO, N., et al., 2022, "Open benchmarks for assessment of process monitoring and fault diagnosis techniques: A review and critical analysis", *Computers & Chemical Engineering*, v. 165 (set.), pp. 107964. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.107964. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422003003>>.

MELO, A., CLAVIJO, N., SOARES, R., et al., 2023. "BibMon: Biblioteca de Monitoramento de Processos". Em processo de registro.

MELO, P. A., BISCAIA, E. C., PINTO, J. C., 2003, "The bifurcation behavior of continuous free-radical solution loop polymerization reactors", *Chemical Engineering Science*, v. 58, n. 13 (jul.), pp. 2805–2821. ISSN: 00092509. doi: 10.1016/S0009-2509(03)00132-5. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0009250903001325>>.

- MENEZES, D. Q. F. D., DE SÁ, M. C. C., FONTOURA, T. B., et al., 2020, “Modeling of Spiral Wound Membranes for Gas Separations—Part II: Data Reconciliation for Online Monitoring”, *Processes*, v. 8, n. 9 (set.), pp. 1035. ISSN: 2227-9717. doi: 10.3390/pr8091035. Disponível em: <<https://www.mdpi.com/2227-9717/8/9/1035>>.
- MLETIC, I., QUINN, S., DUDZIC, M., et al., 2004, “An industrial perspective on implementing on-line applications of multivariate statistics”, *Journal of Process Control*, v. 14, n. 8 (dez.), pp. 821–836. ISSN: 09591524. doi: 10.1016/j.jprocont.2004.02.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152404000174>>.
- MLETIC, I., BOUDREAU, F., DUDZIC, M., et al., 2008, “Experiences in applying data-driven modelling technology to steelmaking processes”, *The Canadian Journal of Chemical Engineering*, v. 86, n. 5 (out.), pp. 937–946. ISSN: 00084034. doi: 10.1002/cjce.20090. Disponível em: <<http://doi.wiley.com/10.1002/cjce.20090>>.
- MILLER, P., SWANSON, R. E., HECKLER, C. E., 1993, “Contribution plots: a missing link in multivariate quality control”. Rochester, NY, USA.
- MILMAN, O., 2018. “Why does PCA assume Gaussian Distribution?” ago. Disponível em: <<https://datascience.stackexchange.com/a/37178>>.
- MISRA, M., YUE, H., QIN, S., et al., 2002, “Multivariate process monitoring and fault diagnosis by multi-scale PCA”, *Computers & Chemical Engineering*, v. 26, n. 9 (set.), pp. 1281–1293. ISSN: 00981354. doi: 10.1016/S0098-1354(02)00093-5. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135402000935>>.
- MITCHELL, T. M., 1997, *Machine Learning*. 1 ed. New York, McGraw-Hill Science/Engineering/Math. ISBN: 978-0-07-042807-2.
- MONTAGUE, G., HIDEN, H., KORNFIELD, G., 1998, “Multivariate Statistical Monitoring Procedures for Fermentation Supervision: An Industrial Case Study”, *IFAC Proceedings Volumes*, v. 31, n. 8 (maio), pp. 399–404. ISSN: 14746670. doi: 10.1016/S1474-6670(17)40218-7. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017402187>>.
- MONTEIRO, A. D. R. D., FEITAL, T. D. S., PINTO, J. C., 2017, “Statistical Aspects of Near-Infrared Spectroscopy for the Characterization of Errors and Model Building”, *Applied Spectroscopy*, v. 71, n. 7 (jul.), pp. 1665–1676. ISSN: 0003-7028, 1943-3530. doi: 10.1177/

0003702817704587. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0003702817704587>>.

MONTEIRO, A. D. R. D., FEITAL, T. D. S., PINTO, J. C., 2021, “A Numerical Procedure for Multivariate Calibration Using Heteroscedastic Principal Components Regression”, *Processes*, v. 9, n. 9 (set.), pp. 1686. ISSN: 2227-9717. doi: 10.3390/pr9091686. Disponível em: <<https://www.mdpi.com/2227-9717/9/9/1686>>.

MONTGOMERY, D. C., 2012, *Introduction to Statistical Quality Control*. 7 edition ed. Hoboken, NJ, Wiley. ISBN: 978-1-118-14681-1.

MOODY, J. M., 2014, *Process Monitoring with Restricted Boltzmann Machines*. Dissertação de Mestrado, Stellenbosch University, Stellenbosch, abr. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1029.9120&rep=rep1&type=pdf>>.

MORGENTHALER, S., 2009, “Exploratory data analysis”, *WIREs Computational Statistics*, v. 1, n. 1, pp. 33–44. ISSN: 1939-0068. doi: 10.1002/wics.2. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.2>>.

MORTADA, M.-A., CARROLL, T., YACOUT, S., et al., 2012, “Rogue components: their effect and control using logical analysis of data”, *Journal of Intelligent Manufacturing*, v. 23, n. 2 (abr.), pp. 289–302. ISSN: 0956-5515, 1572-8145. doi: 10.1007/s10845-009-0351-1. Disponível em: <<http://link.springer.com/10.1007/s10845-009-0351-1>>.

MOSHTAGHI, M., LECKIE, C., BEZDEK, J. C., 2016, “Online Clustering of Multivariate Time-series”. In: *Proceedings of the 2016 SIAM International Conference on Data Mining*, pp. 360–368. Society for Industrial and Applied Mathematics, jun. ISBN: 978-1-61197-434-8. doi: 10.1137/1.9781611974348.41. Disponível em: <<https://pubs.siam.org/doi/10.1137/1.9781611974348.41>>.

MOWBRAY, M., VALLERIO, M., PEREZ-GALVAN, C., et al., 2022, “Industrial data science – a review of machine learning applications for chemical and process industries”, *Reaction Chemistry & Engineering*, v. 7, n. 7, pp. 1471–1509. ISSN: 2058-9883. doi: 10.1039/D1RE00541C. Disponível em: <<http://xlink.rsc.org/?DOI=D1RE00541C>>.

MOYNE, J., ISKANDAR, J., 2017, “Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing”, *Processes*, v. 5, n. 39

(jul.). ISSN: 2227-9717. doi: 10.3390/pr5030039. Disponível em: <<http://www.mdpi.com/2227-9717/5/3/39>>.

MUKKULA, A. R. G., ENGELL, S., 2021, “Handling Measurement Delay in Iterative Real-Time Optimization Methods”, *Processes*, v. 9, n. 10 (out.), pp. 1800. doi: 10.3390/pr9101800. Disponível em: <<https://www.mdpi.com/2227-9717/9/10/1800>>.

MULLER, M., 2007, “Dynamic Time Warping”. In: *Information Retrieval for Music and Motion*, Springer, pp. 69–84, Berlin, Heidelberg. ISBN: 978-3-540-74048-3. doi: 10.1007/978-3-540-74048-3_4. Disponível em: <https://doi.org/10.1007/978-3-540-74048-3_4>.

NACHTERGAELE, P., THYBAUT, J. W., DE MEESTER, S., et al., 2020, “Multivariate Analysis of Industrial Biorefinery Processes: Strategy for Improved Process Understanding with Case Studies in Fatty Acid Production”, *Industrial & Engineering Chemistry Research*, (mar.). ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c00515. Disponível em: <<https://pubs.acs.org/doi/abs/10.1021/acs.iecr.0c00515>>.

NAGHOOSI, E., HUANG, B., DOMLAN, E., et al., 2013, “Information transfer methods in causality analysis of process variables with an industrial application”, *Journal of Process Control*, v. 23, n. 9 (out.), pp. 1296–1305. ISSN: 09591524. doi: 10.1016/j.jprocont.2013.02.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152413000310>>.

NARASIMHAN, S., BHATT, N., 2015, “Deconstructing principal component analysis using a data reconciliation perspective”, *Computers & Chemical Engineering*, v. 77 (jun.), pp. 74–84. ISSN: 00981354. doi: 10.1016/j.compchemeng.2015.03.016. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135415000873>>.

NARASIMHAN, S., SHAH, S. L., 2008, “Model identification and error covariance matrix estimation from noisy data using PCA”, *Control Engineering Practice*, v. 16, n. 1 (jan.), pp. 146–155. ISSN: 09670661. doi: 10.1016/j.conengprac.2007.04.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066107000925>>.

NAWAZ, M., MAULUD, A. S., ZABIRI, H., et al., 2020, “Multiscale Framework for Real-Time Process Monitoring of Nonlinear Chemical Process Systems”, *Industrial & Engineering Chemistry Research*, v. 59, n. 41 (out.), pp. 18595–18606. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c02288. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.0c02288>>.

- NAWAZ, M., MAULUD, A. S., ZABIRI, H., 2022, “Analysis of multiscale process monitoring in industrial processes from a bibliometric perspective”, *Computers & Chemical Engineering*, v. 168 (dez.), pp. 108055. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.108055. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422003878>>.
- NELSON, L. S., 1984, “The Shewhart Control Chart—Tests for Special Causes”, *Journal of Quality Technology*, v. 16, n. 4 (out.), pp. 237–239. ISSN: 0022-4065. doi: 10.1080/00224065.1984.11978921. Disponível em: <<https://doi.org/10.1080/00224065.1984.11978921>>.
- NEOGI, D., SCHLAGS, C. E., 1998, “Multivariate Statistical Analysis of an Emulsion Batch Process”, *Industrial & Engineering Chemistry Research*, v. 37, n. 10 (out.), pp. 3971–3979. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie980243o. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie980243o>>.
- NESPECA, M. G., PAVINI, W. D., DE OLIVEIRA, J. E., 2019, “Multivariate filters combined with interval partial least square method: A strategy for optimizing PLS models developed with near infrared data of multicomponent solutions”, *Vibrational Spectroscopy*, v. 102 (maio), pp. 97–102. ISSN: 09242031. doi: 10.1016/j.vibspec.2019.05.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0924203119300037>>.
- NG, D., FENG, M., 2020, “Medical Image Recognition: An Explanation and Hands-On Example of Convolutional Networks”. In: Celi, L. A., Majumder, M. S., Ordóñez, P., et al. (Eds.), *Leveraging Data Science for Global Health*, Springer International Publishing, pp. 263–284, Cham. ISBN: 978-3-030-47994-7. doi: 10.1007/978-3-030-47994-7_16. Disponível em: <https://doi.org/10.1007/978-3-030-47994-7_16>.
- NG, Y. S., SRINIVASAN, R., 2008, “Multivariate Temporal Data Analysis Using Self-Organizing Maps. 2. Monitoring and Diagnosis of Multistate Operations”, *Industrial & Engineering Chemistry Research*, v. 47, n. 20 (out.), pp. 7758–7771. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie071022y. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie071022y>>.
- NHAT, D. M., VENKATESAN, R., KHAN, F., 2020, “Data-driven Bayesian network model for early kick detection in industrial drilling process”, *Process Safety and Environmental Protection*, v. 138 (jun.), pp. 130–138. ISSN: 09575820. doi: 10.1016/j.psep.2020.03.017. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957582020301002>>.

- NIELSEN, M., 2015, *Neural Networks and Deep Learning*. Determination Press.
Disponível em: <<http://neuralnetworksanddeeplearning.com/>>.
- NING, C., CHEN, M., ZHOU, D., 2015, “Sparse Contribution Plot for Fault Diagnosis of Multimodal Chemical Processes”, *IFAC-PapersOnLine*, v. 48, n. 21, pp. 619–626. ISSN: 24058963. doi: 10.1016/j.ifacol.2015.09.595.
Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405896315017243>>.
- NIU, Y. S., HAO, N., ZHANG, H., 2016, “Multiple Change-Point Detection: A Selective Overview”, *Statistical Science*, v. 31, n. 4 (nov.). ISSN: 0883-4237. doi: 10.1214/16-STS587. Disponível em: <<https://projecteuclid.org/journals/statistical-science/volume-31/issue-4/Multiple-Change-Point-Detection-A-Selective-Overview/10.1214/16-STS587.full>>.
- NIU, Z., ZHONG, G., YU, H., 2021, “A review on the attention mechanism of deep learning”, *Neurocomputing*, v. 452 (set.), pp. 48–62. ISSN: 0925-2312. doi: 10.1016/j.neucom.2021.03.091. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S092523122100477X>>.
- NOGUEIRA, J. D. N. P., MELO, P. A., DE SOUZA JR., M. B., 2022, “Faulty Scenarios in Sour Water Treatment Units: Simulation and AI-Based Diagnosis”, *Process Safety and Environmental Protection*, (jul.). ISSN: 0957-5820. doi: 10.1016/j.psep.2022.07.043. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957582022006553>>.
- NOMIKOS, P., MACGREGOR, J. F., 1995, “Multivariate SPC Charts for Monitoring Batch Processes”, *Technometrics*, v. 37, n. 1, pp. 41–59. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/00401706.1995.10485888>>.
- NOMIKOS, P., MACGREGOR, J. F., 1994, “Monitoring batch processes using multiway principal component analysis”, *AICHE Journal*, v. 40, n. 8 (ago.), pp. 1361–1375. ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.690400809. Disponível em: <<http://doi.wiley.com/10.1002/aic.690400809>>.
- NUCCI, E. R., CRUZ, A. J., GIORDANO, R. L., et al., 2009, “Monitoring Penicillin G Acylase (PGA) Production Using Principal Component Analysis (PCA)”. In: *Computer Aided Chemical Engineering*, v. 27, Elsevier, pp. 1629–1634. ISBN: 978-0-444-53472-9. doi: 10.1016/S1570-7946(09)70662-5. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1570794609706625>>.

NUCCI, E. R., CRUZ, A. J. G., GIORDANO, R. C., 2010, “Monitoring bio-reactors using principal component analysis: production of penicillin G acylase as a case study”, *Bioprocess and Biosystems Engineering*, v. 33, n. 5 (jun.), pp. 557–564. ISSN: 1615-7591, 1615-7605. doi: 10.1007/s00449-009-0377-y. Disponível em: <<http://link.springer.com/10.1007/s00449-009-0377-y>>.

OFFERMANS, T., SZYMANSKA, E., BUYDENS, L. M., et al., 2020, “Synchronizing process variables in time for industrial process monitoring and control”, *Computers & Chemical Engineering*, v. 140 (set.), pp. 106938. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106938. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135420301010>>.

OKADA, K. F. A., SILVA DE MORAIS, A., OLIVEIRA-LOPES, L. C., et al., 2021, “A Survey on Fault Detection and Diagnosis Methods”. In: *2021 14th IEEE International Conference on Industry Applications (INDUSCON)*, pp. 1422–1429, ago. doi: 10.1109/INDUSCON51756.2021.9529495. Disponível em: <<https://ieeexplore.ieee.org/document/9529495>>.

OLIVEIRA, L. M. C., DIAS, R., REBELLO, C. M., et al., 2021, “Artificial Intelligence and Cyber-Physical Systems: A Review and Perspectives for the Future in the Chemical Industry”, *AI*, v. 2, n. 3 (set.), pp. 429–443. ISSN: 2673-2688. doi: 10.3390/ai2030027. Disponível em: <<https://www.mdpi.com/2673-2688/2/3/27>>.

OLIVIER, J., ALDRICH, C., 2020, “Dynamic Monitoring of Grinding Circuits by Use of Global Recurrence Plots and Convolutional Neural Networks”, *Minerals*, v. 10, n. 11 (out.), pp. 958. ISSN: 2075-163X. doi: 10.3390/min10110958. Disponível em: <<https://www.mdpi.com/2075-163X/10/11/958>>.

OLSHAUSEN, B. A., FIELD, D. J., 1997, “Sparse coding with an overcomplete basis set: A strategy employed by V1?” *Vision Research*, v. 37, n. 23 (dez.), pp. 3311–3325. ISSN: 00426989. doi: 10.1016/S0042-6989(97)00169-7. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0042698997001697>>.

OURIQUE, C. O., BISCAIA, E. C., PINTO, J. C., 2002, “The use of particle swarm optimization for dynamical analysis in chemical processes”, *Computers & Chemical Engineering*, v. 26, n. 12 (dez.), pp. 1783–1793. ISSN:

00981354. doi: 10.1016/S0098-1354(02)00153-9. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135402001539>>.

OUSSOUS, A., BENJELLOUN, F.-Z., AIT LAHCEN, A., et al., 2018, “Big Data technologies: A survey”, *Journal of King Saud University - Computer and Information Sciences*, v. 30, n. 4 (out.), pp. 431–448. ISSN: 13191578. doi: 10.1016/j.jksuci.2017.06.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1319157817300034>>.

OUYANG, H., ZENG, J., LI, Y., et al., 2020, “Fault Detection and Identification of Blast Furnace Ironmaking Process Using the Gated Recurrent Unit Network”, *Processes*, v. 8, n. 4 (abr.), pp. 391. ISSN: 2227-9717. doi: 10.3390/pr8040391. Disponível em: <<https://www.mdpi.com/2227-9717/8/4/391>>.

OZTEMEL, E., GURSEV, S., 2020, “Literature review of Industry 4.0 and related technologies”, *Journal of Intelligent Manufacturing*, v. 31, n. 1 (jan.), pp. 127–182. ISSN: 0956-5515, 1572-8145. doi: 10.1007/s10845-018-1433-8. Disponível em: <<http://link.springer.com/10.1007/s10845-018-1433-8>>.

PADHEE, S., GUPTA, N., KAUR, G., 2012, “Data Driven Multivariate Technique for Fault Detection of Waste Water Treatment Plant”, *International Journal of Engineering and Advanced Technology*, v. 1, n. 4 (abr.), pp. 45–50. Disponível em: <<https://www.researchgate.net/publication/266080220>>.

PAGE, E. S., 1954, “Continuous Inspection Schemes”, *Biometrika*, v. 41, n. 1/2, pp. 100–115. ISSN: 0006-3444. doi: 10.2307/2333009. Disponível em: <<https://www-jstor-org.ez29.periodicos.capes.gov.br/stable/2333009>>.

PALADE, V., PATTON, R. J., UPPAL, F. J., et al., 2002, “Fault diagnosis of an industrial gas turbine using neuro-fuzzy methods”, *IFAC Proceedings Volumes*, v. 35, n. 1 (jan.), pp. 471–476. ISSN: 1474-6670. doi: 10.3182/20020721-6-ES-1901.01632. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1474667015400539>>.

PALLA, G. P. L., PANI, A. K., 2023, “Independent component analysis application for fault detection in process industries: Literature review and an application case study for fault detection in multiphase flow systems”, *Measurement*, v. 209 (mar.), pp. 112504. ISSN: 0263-2241. doi: 10.1016/j.measurement.2023.112504. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263224123000684>>.

- PAN, Y., 2016, "Heading toward Artificial Intelligence 2.0", *Engineering*, v. 2, n. 4 (dez.), pp. 409–413. ISSN: 20958099. doi: 10.1016/J.ENG.2016.04.018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2095809917300772>>.
- PANI, A. K., 2022, "Non-linear process monitoring using kernel principal component analysis: A review of the basic and modified techniques with industrial applications", *Brazilian Journal of Chemical Engineering*, v. 39, n. 2 (jun.), pp. 327–344. ISSN: 1678-4383. doi: 10.1007/s43153-021-00125-2. Disponível em: <<https://doi.org/10.1007/s43153-021-00125-2>>.
- PARENTE, A. P., DE SOUZA JR., M. B., VALDMAN, A., et al., 2019, "Data Augmentation Applied to Machine Learning-Based Monitoring of a Pulp and Paper Process", *Processes*, v. 7, n. 12 (dez.), pp. 958. ISSN: 2227-9717. doi: 10.3390/pr7120958. Disponível em: <<https://www.mdpi.com/2227-9717/7/12/958>>.
- PARK, B. E., KIM, J. S., LEE, J.-K., et al., 2020a, "Anomaly detection in a hyper-compressor in low-density polyethylene manufacturing processes using WPCA-based principal component control limit", *Korean Journal of Chemical Engineering*, v. 37, n. 1 (jan.), pp. 11–18. ISSN: 1975-7220. doi: 10.1007/s11814-019-0403-y. Disponível em: <<https://doi.org/10.1007/s11814-019-0403-y>>.
- PARK, Y.-J., FAN, S.-K. S., HSU, C.-Y., 2020b, "A Review on Fault Detection and Process Diagnostics in Industrial Processes", *Processes*, v. 8, n. 9 (set.), pp. 1123. ISSN: 2227-9717. doi: 10.3390/pr8091123. Disponível em: <<https://www.mdpi.com/2227-9717/8/9/1123>>.
- PATWARDHAN, R. S., HAMADAH, H. A., PATEL, K. M., et al., 2019, "Applications of Advanced Analytics at Saudi Aramco: A Practitioners' Perspective", *Industrial & Engineering Chemistry Research*, v. 58, n. 26 (jul.), pp. 11338–11351. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.8b06205. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.8b06205>>.
- PAUL, G. C., THOMAS, C. R., 1996, "A structured model for hyphal differentiation and penicillin production using *Penicillium chrysogenum*", *Biotechnology and Bioengineering*, v. 51, n. 5, pp. 558–572. ISSN: 1097-0290. doi: 10.1002/(SICI)1097-0290(19960905)51:5\%3C558::AID-BIT8\%3E3.0.CO;2-B. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0290%2819960905%2951%3A5%3C558%3A%3AAID-BIT8%3E3.0.CO%3B2-B>>.

- PAZEL, D. P., 2022, “Software Modeling and Object-Oriented Design”. In: Pazel, D. P. (Ed.), *Music Representation and Transformation in Software: Structure and Algorithms in Python*, Springer International Publishing, pp. 3–10, Cham. ISBN: 978-3-030-97472-5. doi: 10.1007/978-3-030-97472-5_1. Disponível em: <https://doi.org/10.1007/978-3-030-97472-5_1>.
- PEARSON, K., 1901, “LIII - On lines and planes of closest fit to systems of points in space”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, v. 2, n. 11 (nov.), pp. 559–572. ISSN: 1941-5982. doi: 10.1080/14786440109462720. Disponível em: <<https://doi.org/10.1080/14786440109462720>>.
- PEARSON, R., 2001, “Exploring process data”, *Journal of Process Control*, v. 11, n. 2 (abr.), pp. 179–194. ISSN: 09591524. doi: 10.1016/S0959-1524(00)00046-9. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152400000469>>.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., et al., 2011, “Scikit-learn: Machine Learning in Python”, *Journal of Machine Learning Research*, v. 12 (out.), pp. 2825–2830. Disponível em: <<http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>>.
- PENG, C., KAI, W., KUN, Z., et al., 2022a, “Monitoring of wastewater treatment process based on multi-stage variational autoencoder”, *Expert Systems with Applications*, v. 207 (nov.), pp. 117919. ISSN: 0957-4174. doi: 10.1016/j.eswa.2022.117919. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417422011642>>.
- PENG, D.-Y., ROBINSON, D. B., 1976, “A New Two-Constant Equation of State”, *Industrial & Engineering Chemistry Fundamentals*, v. 15, n. 1 (fev.), pp. 59–64. ISSN: 0196-4313, 1541-4833. doi: 10.1021/i160057a011. Disponível em: <<https://pubs.acs.org/doi/abs/10.1021/i160057a011>>.
- PENG, J., WU, W., LOCKHART, B., et al., 2021, “DataPrep.EDA: Task-Centric Exploratory Data Analysis for Statistical Modeling in Python”. In: *Proceedings of the 2021 International Conference on Management of Data*, pp. 2271–2280, jun. doi: 10.1145/3448016.3457330. Disponível em: <<http://arxiv.org/abs/2104.00841>>.
- PENG, P., WU, Y., ZHANG, Y., et al., 2019, “A Novel Fault Detection and Diagnosis Method Based on Gaussian-Bernoulli Restricted Boltzmann Machine”. In: *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 1349–1354, out. doi: 10.1109/SMC.2019.8914596.

- PENG, P., ZHANG, W., ZHANG, Y., et al., 2020a, “Cost sensitive active learning using bidirectional gated recurrent neural networks for imbalanced fault diagnosis”, *Neurocomputing*, v. 407 (set.), pp. 232–245. ISSN: 09252312. doi: 10.1016/j.neucom.2020.04.075. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231220306603>>.
- PENG, X., TANG, Y., DU, W., et al., 2017, “Multimode Process Monitoring and Fault Detection: A Sparse Modeling and Dictionary Learning Method”, *IEEE Transactions on Industrial Electronics*, v. 64, n. 6 (jun.), pp. 4866–4875. ISSN: 0278-0046, 1557-9948. doi: 10.1109/TIE.2017.2668987. Disponível em: <<http://ieeexplore.ieee.org/document/7855804/>>.
- PENG, X., LI, Z., ZHONG, W., et al., 2020b, “Concurrent Quality-Relevant Canonical Correlation Analysis for Nonlinear Continuous Process Decomposition and Monitoring”, *Industrial & Engineering Chemistry Research*, v. 59, n. 18 (maio), pp. 8757–8768. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c00895. Disponível em: <<https://pubs.acs.org/doi/abs/10.1021/acs.iecr.0c00895>>.
- PENG, X., LI, H., YUAN, F., et al., 2022b, “An extreme learning machine for unsupervised online anomaly detection in multivariate time series”, *Neurocomputing*, v. 501 (ago.), pp. 596–608. ISSN: 0925-2312. doi: 10.1016/j.neucom.2022.06.042. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231222007615>>.
- PERERA, Y. S., RATNAWEERA, D. A. A. C., DASANAYAKA, C. H., et al., 2023, “The role of artificial intelligence-driven soft sensors in advanced sustainable process industries: A critical review”, *Engineering Applications of Artificial Intelligence*, v. 121 (maio), pp. 105988. ISSN: 0952-1976. doi: 10.1016/j.engappai.2023.105988. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197623001720>>.
- PERES, F. A. P., FOGLIATTO, F. S., 2018, “Variable selection methods in multivariate statistical process control: A systematic literature review”, *Computers & Industrial Engineering*, v. 115 (jan.), pp. 603–619. ISSN: 03608352. doi: 10.1016/j.cie.2017.12.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0360835217305727>>.
- PERRUSQUIA, A., YU, W., LI, X., 2021, “Nonlinear Control Using Human Behavior Learning”, *Information Sciences*, (mar.). ISSN: 0020-0255. doi: 10.1016/j.ins.2021.03.043. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0020025521002887>>.

- PHRIDVIRAJ, M., GURURAO, C., 2014, “Data Mining – Past, Present and Future – A Typical Survey on Data Streams”, *Procedia Technology*, v. 12, pp. 255–263. ISSN: 22120173. doi: 10.1016/j.protcy.2013.12.483. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2212017313006683>>.
- PHUNG, V. H., RHEE, E. J., 2019, “A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets”, *Applied Sciences*, v. 9, n. 21 (jan.), pp. 4500. ISSN: 2076-3417. doi: 10.3390/app9214500. Disponível em: <<https://www.mdpi.com/2076-3417/9/21/4500>>.
- PIETRUCZUK, L., RUTKOWSKI, L., JAWORSKI, M., et al., 2017, “How to adjust an ensemble size in stream data mining?” *Information Sciences*, v. 381 (mar.), pp. 46–54. ISSN: 00200255. doi: 10.1016/j.ins.2016.10.028. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0020025516313445>>.
- PILARIO, K. E., SHAFIEE, M., CAO, Y., et al., 2019, “A Review of Kernel Methods for Feature Extraction in Nonlinear Process Monitoring”, *Processes*, v. 8, n. 1 (dez.), pp. 24. ISSN: 2227-9717. doi: 10.3390/pr8010024. Disponível em: <<https://www.mdpi.com/2227-9717/8/1/24>>.
- PIMENTEL, M. A., CLIFTON, D. A., CLIFTON, L., et al., 2014, “A review of novelty detection”, *Signal Processing*, v. 99 (jun.), pp. 215–249. ISSN: 01651684. doi: 10.1016/j.sigpro.2013.12.026. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S016516841300515X>>.
- PINTO, J., RAY, W., 1995, “The dynamic behavior of continuous solution polymerization reactors—VII. Experimental study of a copolymerization reactor”, *Chemical Engineering Science*, v. 50, n. 4 (fev.), pp. 715–736. ISSN: 00092509. doi: 10.1016/0009-2509(94)00456-2. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0009250994004562>>.
- PIOVOSO, M. J., KOSANOVICH, K. A., PEARSON, R. K., 1992, “Monitoring Process Performance in Real-Time”. In: *1992 American Control Conference*, pp. 2359–2363, Chicago, IL, USA, jun. IEEE. ISBN: 978-0-7803-0210-5. doi: 10.23919/ACC.1992.4792557. Disponível em: <<https://ieeexplore.ieee.org/document/4792557/>>.
- PIOVOSO, M. J., KOSANOVICH, K. A., 1994, “Applications of multivariate statistical methods to process monitoring and controller design”, *International*

Journal of Control, v. 59, n. 3 (mar.), pp. 743–765. ISSN: 0020-7179, 1366-5820. doi: 10.1080/00207179408923103. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/00207179408923103>>.

PIRDASHTI, M., CURTEANU, S., KAMANGAR, M. H., et al., 2013, “Artificial neural networks: applications in chemical engineering”, *Reviews in Chemical Engineering*, v. 29, n. 4 (jan.). ISSN: 2191-0235, 0167-8299. doi: 10.1515/revce-2013-0013. Disponível em: <<https://www.degruyter.com/view/j/revce.2013.29.issue-4/revce-2013-0013/revce-2013-0013.xml>>.

PRATA, D. M., LIMA, E. L., PINTO, J. C., 2009, “Nonlinear Dynamic Data Reconciliation in Real Time in Actual Processes”. In: *Computer Aided Chemical Engineering*, v. 27, Elsevier, pp. 47–54. ISBN: 9780444534729. doi: 10.1016/S1570-7946(09)70228-7. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1570794609702287>>.

PREUSS, P., PUCHSTEIN, R., DETTE, H., 2015, “Detection of Multiple Structural Breaks in Multivariate Time Series”, *Journal of the American Statistical Association*, v. 110, n. 510 (abr.), pp. 654–668. ISSN: 0162-1459, 1537-274X. doi: 10.1080/01621459.2014.920613. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/01621459.2014.920613>>.

PSALTIS, A., 2017, *Streaming Data: Understanding the real-time pipeline*. 1 edition ed. Shelter Island, NY, Manning Publications. ISBN: 978-1-61729-228-6.

PUGGINI, L., DOYLE, J., MCLOONE, S., 2015, “Fault Detection using Random Forest Similarity Distance”, *IFAC-PapersOnLine*, v. 48, n. 21, pp. 583–588. ISSN: 24058963. doi: 10.1016/j.ifacol.2015.09.589. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405896315017188>>.

PULI, V. K., RAVEENDRAN, R., HUANG, B., 2021, “Complex probabilistic slow feature extraction with applications in process data analytics”, *Computers & Chemical Engineering*, v. 154 (nov.), pp. 107456. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107456. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135421002349>>.

QIAN, J., SONG, Z., YAO, Y., et al., 2022, “A review on autoencoder based representation learning for fault detection and diagnosis in industrial processes”, *Chemometrics and Intelligent Laboratory Systems*, v. 231 (dez.), pp. 104711. ISSN: 0169-7439. doi: 10.1016/j.chemolab.2022.104711. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169743922002222>>.

- QIANG LIU, QIN, S. J., TIANYOU CHAI, 2013, “Decentralized Fault Diagnosis of Continuous Annealing Processes Based on Multilevel PCA”, *IEEE Transactions on Automation Science and Engineering*, v. 10, n. 3 (jul.), pp. 687–698. ISSN: 1545-5955, 1558-3783. doi: 10.1109/TASE.2012.2230628. Disponível em: <<http://ieeexplore.ieee.org/document/6419855/>>.
- QIN, S. J., 2009, “Data-driven Fault Detection and Diagnosis for Complex Industrial Processes”, *IFAC Proceedings Volumes*, v. 42, n. 8, pp. 1115–1125. ISSN: 14746670. doi: 10.3182/20090630-4-ES-2003.00184. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667016359274>>.
- QIN, S. J., 2014, “Process data analytics in the era of big data”, *AIChe Journal*, v. 60, n. 9 (set.), pp. 3092–3100. ISSN: 1547-5905. doi: 10.1002/aic.14523. Disponível em: <<https://aiche-onlinelibrary-wiley.ez29.capes.proxy.ufrj.br/doi/abs/10.1002/aic.14523>>.
- QIN, S. J., 1998, “Recursive PLS algorithms for adaptive data modeling”, *Computers & Chemical Engineering*, v. 22, n. 4-5 (jan.), pp. 503–514. ISSN: 00981354. doi: 10.1016/S0098-1354(97)00262-7. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135497002627>>.
- QIN, S. J., 2003, “Statistical process monitoring: basics and beyond”, *Journal of Chemometrics*, v. 17, n. 8-9, pp. 480–502. ISSN: 1099-128X. doi: 10.1002/cem.800. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.800>>.
- QIN, S. J., 2012, “Survey on data-driven industrial process monitoring and diagnosis”, *Annual Reviews in Control*, v. 36, n. 2 (dez.), pp. 220–234. ISSN: 13675788. doi: 10.1016/j.arcontrol.2012.09.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1367578812000399>>.
- QIN, S. J., CHIANG, L. H., 2019, “Advances and opportunities in machine learning for process data analytics”, *Computers & Chemical Engineering*, v. 126 (jul.), pp. 465–473. ISSN: 00981354. doi: 10.1016/j.compchemeng.2019.04.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419302248>>.
- QIN, S. J., DUNIA, R., 2000, “Determining the number of principal components for best reconstruction”, *Journal of Process Control*, p. 6.

- QIN, S. J., VALLE, S., PIOVOSO, M. J., 2001, “On unifying multiblock analysis with application to decentralized process monitoring”, *Journal of Chemometrics*, v. 15, n. 9 (out.), pp. 715–742. ISSN: 0886-9383, 1099-128X. doi: 10.1002/cem.667. Disponível em: <<http://doi.wiley.com/10.1002/cem.667>>.
- QIN, S. J., CHERRY, G., GOOD, R., et al., 2006, “Semiconductor manufacturing process control and monitoring: A fab-wide framework”, *Journal of Process Control*, v. 16, n. 3 (mar.), pp. 179–191. ISSN: 09591524. doi: 10.1016/j.jprocont.2005.06.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152405000600>>.
- QIN, S. J., DONG, Y., ZHU, Q., et al., 2020, “Bridging systems theory and data science: A unifying review of dynamic latent variable analytics and process monitoring”, *Annual Reviews in Control*, v. 50, pp. 29–48. ISSN: 13675788. doi: 10.1016/j.arcontrol.2020.09.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1367578820300602>>.
- QIN, S. J., LIU, Y., DONG, Y., 2021, “Plant-wide troubleshooting and diagnosis using dynamic embedded latent feature analysis”, *Computers & Chemical Engineering*, v. 152 (set.), pp. 107392. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107392. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135421001708>>.
- QIN, S., MCAVOY, T., 1996, “Nonlinear FIR modeling via a neural net PLS approach”, *Computers & Chemical Engineering*, v. 20, n. 2 (fev.), pp. 147–159. ISSN: 00981354. doi: 10.1016/0098-1354(95)00011-P. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/009813549500011P>>.
- QIN, Y., LOU, Z., WANG, Y., et al., 2022, “An analytical partial least squares method for process monitoring”, *Control Engineering Practice*, v. 124 (jul.), pp. 105182. ISSN: 0967-0661. doi: 10.1016/j.conengprac.2022.105182. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0967066122000740>>.
- QUIÑONES-GRUEIRO, M., PRIETO-MORENO, A., VERDE, C., et al., 2019, “Data-driven monitoring of multimode continuous processes: A review”, *Chemometrics and Intelligent Laboratory Systems*, v. 189 (jun.), pp. 56–71. ISSN: 01697439. doi: 10.1016/j.chemolab.2019.03.012. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743918306774>>.

- RADFORD, A., NARASIMHAN, K., SALIMANS, T., et al., 2018, “Improving language understanding by generative pre-training”, Disponível em: <<https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf>>.
- RAGAB, A., EL-KOUJOK, M., AMAZOUZ, M., et al., 2017, “Fault detection and diagnosis in the Tennessee Eastman Process using interpretable knowledge discovery”. In: *2017 Annual Reliability and Maintainability Symposium (RAMS)*, Orlando, FL, USA. IEEE. ISBN: 978-1-5090-5284-4. doi: 10.1109/RAM.2017.7889650. Disponível em: <<http://ieeexplore.ieee.org/document/7889650/>>.
- RAGAB, A., EL-KOUJOK, M., POULIN, B., et al., 2018, “Fault diagnosis in industrial chemical processes using interpretable patterns based on Logical Analysis of Data”, *Expert Systems with Applications*, v. 95 (abr.), pp. 368–383. ISSN: 09574174. doi: 10.1016/j.eswa.2017.11.045. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417417307984>>.
- RAGAB, A., EL-KOUJOK, M., GHEZZAZ, H., et al., 2019, “Deep understanding in industrial processes by complementing human expertise with interpretable patterns of machine learning”, *Expert Systems with Applications*, v. 122 (maio), pp. 388–405. ISSN: 09574174. doi: 10.1016/j.eswa.2019.01.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417419300107>>.
- RAJULAPATI, L., CHINTA, S., SHYAMALA, B., et al., 2022, “Integration of machine learning and first principles models”, *AIChe Journal*, v. 68, n. 6 (jun.). ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.17715. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/aic.17715>>.
- RALSTON, P., DEPUY, G., GRAHAM, J. H., 2001, “Computer-based monitoring and fault diagnosis: a chemical process case study”, *ISA Transactions*, v. 40, n. 1 (jan.), pp. 85–98. ISSN: 00190578. doi: 10.1016/S0019-0578(00)00033-1. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0019057800000331>>.
- RAMESH, S., VASU, B., 2019, “Application of EWMA chart for monitoring process mean in paper industry”, *Management Science Letters*, pp. 571–576. ISSN: 19239335, 19239343. doi: 10.5267/j.msl.2019.1.006. Disponível em: <http://www.growingscience.com/msl/Vol9/msl_2019_47.pdf>.
- RAMIREZ-GALLEGOS, S., KRAWCZYK, B., GARCÍA, S., et al., 2017, “A survey on data preprocessing for data stream mining: Current status and future

- directions”, *Neurocomputing*, v. 239 (maio), pp. 39–57. ISSN: 09252312. doi: 10.1016/j.neucom.2017.01.078. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231217302631>>.
- RAMOS, A. R., CORONA, C. C., VERDEGAY, J. L., et al., 2018, “An approach for fault diagnosis using a novel hybrid fuzzy clustering algorithm”. In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Rio de Janeiro, jul. IEEE. ISBN: 978-1-5090-6020-7. doi: 10.1109/FUZZ-IEEE.2018.8491547. Disponível em: <<https://ieeexplore.ieee.org/document/8491547/>>.
- RANJIT, M., GAZULA, H., HSIANG, S. M., et al., 2015, “Fault Detection Using Human–Machine Co-Construct Intelligence in Semiconductor Manufacturing Processes”, *IEEE Transactions on Semiconductor Manufacturing*, v. 28, n. 3 (ago.), pp. 297–305. ISSN: 1558-2345. doi: 10.1109/TSM.2015.2432770.
- RANNAR, S., MACGREGOR, J. F., WOLD, S., 1998, “Adaptive batch monitoring using hierarchical PCA”, *Chemometrics and Intelligent Laboratory Systems*, v. 41, n. 1 (jul.), pp. 73–81. ISSN: 01697439. doi: 10.1016/S0169-7439(98)00024-0. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743998000240>>.
- RATO, T., REIS, M., SCHMITT, E., et al., 2016, “A systematic comparison of PCA-based Statistical Process Monitoring methods for high-dimensional, time-dependent Processes”, *AIChE Journal*, v. 62, n. 5 (maio), pp. 1478–1493. ISSN: 00011541. doi: 10.1002/aic.15062. Disponível em: <<http://doi.wiley.com/10.1002/aic.15062>>.
- RAUBER, T., BOLDT, F., MUNARO, C., 2020, “Feature Selection for Multivariate Contribution Analysis in Fault Detection and Isolation”, *Journal of the Franklin Institute*, (mar.). ISSN: 00160032. doi: 10.1016/j.jfranklin.2020.03.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0016003220301617>>.
- RAVEENDRAN, R., HUANG, B., 2020, “Conjugate exponential family graphical models in process monitoring: A tutorial review”, *Chemometrics and Intelligent Laboratory Systems*, v. 204 (set.), pp. 104095. ISSN: 01697439. doi: 10.1016/j.chemolab.2020.104095. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S016974392030126X>>.
- REINARTZ, C., KULAHCI, M., RAVN, O., 2021a, “An Extended Tennessee Eastman Simulation Dataset for Fault Detection and Decision Support Systems”,

Computers & Chemical Engineering, (mar.), pp. 107281. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107281. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135421000594>>.

REINARTZ, C., KULAHCI, M., RAVN, O., 2021b, “Tennessee Eastman Reference Data for Fault-Detection and Decision Support Systems”, doi: 10.11583/DTU.13385936.v1.

REIS, M., 2019, “Multiscale and Multi-Granularity Process Analytics: A Review”, *Processes*, v. 7, n. 61 (jan.). ISSN: 2227-9717. doi: 10.3390/pr7020061. Disponível em: <<http://www.mdpi.com/2227-9717/7/2/61>>.

REIS, M., GINS, G., 2017, “Industrial Process Monitoring in the Big Data/Industry 4.0 Era: from Detection, to Diagnosis, to Prognosis”, *Processes*, v. 5, n. 35 (jun.). ISSN: 2227-9717. doi: 10.3390/pr5030035. Disponível em: <<http://www.mdpi.com/2227-9717/5/3/35>>.

REIS, M. S., KENETT, R., 2018, “Assessing the value of information of data-centric activities in the chemical processing industry 4.0”, *AICHE Journal*, v. 64, n. 11 (nov.), pp. 3868–3881. ISSN: 00011541. doi: 10.1002/aic.16203. Disponível em: <<http://doi.wiley.com/10.1002/aic.16203>>.

REIS, M. S., SARAIVA, P. M., 2021, “Data-centric process systems engineering: A push towards PSE 4.0”, *Computers & Chemical Engineering*, v. 155 (dez.), pp. 107529. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107529. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135421003070>>.

REIS, M. S., SARAIVA, P. M., 2022, “Data-Driven Process System Engineering—Contributions to its consolidation following the path laid down by George Stephanopoulos”, *Computers & Chemical Engineering*, v. 159 (mar.), pp. 107675. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.107675. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422000199>>.

REIS, M. S., SARAIVA, P. M., 2006, “Multiscale statistical process control with multiresolution data”, *AICHE Journal*, v. 52, n. 6 (jun.), pp. 2107–2119. ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.10805. Disponível em: <<http://doi.wiley.com/10.1002/aic.10805>>.

REIS, M. S., RENDALL, R., RATO, T. J., et al., 2021, “Improving the sensitivity of statistical process monitoring of manifolds embedded in high-dimensional spaces: The truncated-Q statistic”, *Chemometrics and In-*

- telligent Laboratory Systems*, v. 215 (ago.), pp. 104369. ISSN: 0169-7439. doi: 10.1016/j.chemolab.2021.104369. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169743921001374>>.
- REN, L., LV, W., 2014, “Fault Detection via Sparse Representation for Semiconductor Manufacturing Processes”, *IEEE Transactions on Semiconductor Manufacturing*, v. 27, n. 2 (maio), pp. 252–259. ISSN: 0894-6507, 1558-2345. doi: 10.1109/TSM.2014.2302011. Disponível em: <<http://ieeexplore.ieee.org/document/6719488/>>.
- RENCHER, A. C., CHRISTENSEN, W. F., 2012, *Methods of multivariate analysis*. Wiley series in probability and statistics. Third edition ed. Hoboken, New Jersey, Wiley. ISBN: 978-0-470-17896-6.
- RENDALL, R., CHIANG, L. H., REIS, M. S., 2019, “Data-driven methods for batch data analysis – A critical overview and mapping on the complexity scale”, *Computers & Chemical Engineering*, v. 124 (maio), pp. 1–13. ISSN: 00981354. doi: 10.1016/j.compchemeng.2019.01.014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135418311104>>.
- RIBEIRO, K., FONTES, C. H., DE MELO, G. J. A., 2020, “Genetic algorithm-based fuzzy clustering applied to multivariate time series”, *Evolutionary Intelligence*, (maio). ISSN: 1864-5909, 1864-5917. doi: 10.1007/s12065-020-00422-8. Disponível em: <<http://link.springer.com/10.1007/s12065-020-00422-8>>.
- RICKER, N. L., 1996, “Decentralized control of the Tennessee Eastman Challenge Process”, *Journal of Process Control*, v. 6, n. 4 (ago.), pp. 205–221. ISSN: 0959-1524. doi: 10.1016/0959-1524(96)00031-5. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0959152496000315>>.
- RICKER, N. L., LEE, J. H., 1995, “Nonlinear modeling and state estimation for the Tennessee Eastman challenge process”, *Computers & Chemical Engineering*, v. 19, n. 9 (set.), pp. 983–1005. ISSN: 0098-1354. doi: 10.1016/0098-1354(94)00113-3. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0098135494001133>>.
- RICKER, N., 1995, “Optimal steady-state operation of the Tennessee Eastman challenge process”, *Computers & Chemical Engineering*, v. 19, n. 9 (set.), pp. 949–959. ISSN: 00981354. doi: 10.1016/0098-1354(94)00043-N. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/009813549400043N>>.

- RIETH, C. A., AMSEL, B. D., TRAN, R., et al., 2017, “Additional Tennessee Eastman Process Simulation Data for Anomaly Detection Evaluation”, (jul.). doi: 10.7910/DVN/6C3JR1. Disponível em: <<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6C3JR1>>.
- RIETH, C. A., AMSEL, B. D., TRAN, R., et al., 2018, “Issues and Advances in Anomaly Detection Evaluation for Joint Human-Automated Systems”. In: Chen, J. (Ed.), *Advances in Human Factors in Robots and Unmanned Systems*, v. 595, Springer International Publishing, pp. 52–63, Cham. ISBN: 978-3-319-60383-4 978-3-319-60384-1. doi: 10.1007/978-3-319-60384-1_6. Disponível em: <http://link.springer.com/10.1007/978-3-319-60384-1_6>.
- RIOUL, O., VETTERLI, M., 1991, “Wavelets and signal processing”, *Signal Processing Magazine, IEEE*, v. 8 (out.), pp. 14–38. doi: 10.1109/79.91217. Disponível em: <<https://ieeexplore.ieee.org/document/91217>>.
- RIPPON, L. D., YOUSEF, I., HOSSEINI, B., et al., 2021, “Representation Learning and Predictive Classification: Application with an Electric Arc Furnace”, *Computers & Chemical Engineering*, (mar.), pp. 107304. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107304. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S009813542100082X>>.
- ROBERTS, S. W., 1959, “Control Chart Tests Based on Geometric Moving Averages”, *Technometrics*, v. 1, n. 3 (ago.), pp. 239–250. ISSN: 0040-1706. doi: 10.1080/00401706.1959.10489860. Disponível em: <<https://amstat.tandfonline.com/doi/abs/10.1080/00401706.1959.10489860>>.
- ROBERTSON, T., CHEN, J., ROMAGNOLI, J., et al., 1998, “Intelligent Monitoring for Quality Control in a Biological Nutrient Removal Wastewater Treatment Pilot Plant”, *IFAC Proceedings Volumes*, v. 31, n. 11 (jun.), pp. 559–564. ISSN: 14746670. doi: 10.1016/S1474-6670(17)44985-8. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017449858>>.
- RODGERS, J. L., NICEWANDER, W. A., TOOTHAKER, L., 1984, “Linearly Independent, Orthogonal, and Uncorrelated Variables”, *The American Statistician*, v. 38, n. 2 (maio), pp. 133–134. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/00031305.1984.10483183>>.
- ROJKO, A., 2017, “Industry 4.0 Concept: Background and Overview”, *International Journal of Interactive Mobile Technologies (iJIM)*, v. 11, n. 5 (jul.), pp. 77–

90. ISSN: 1865-7923. doi: 10.3991/ijim.v11i5.7072. Disponível em: <<http://online-journals.org/index.php/i-jim/article/view/7072>>.
- ROSEN, C., LENNOX, J., 2001, “Multivariate and multiscale monitoring of wastewater treatment operation”, *Water Research*, v. 35, n. 14 (out.), pp. 3402–3410. ISSN: 00431354. doi: 10.1016/S0043-1354(01)00069-0. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0043135401000690>>.
- ROSIPAL, R., KRÄMER, N., 2006, “Overview and Recent Advances in Partial Least Squares”. In: Saunders, C., Grobelnik, M., Gunn, S., et al. (Eds.), *Subspace, Latent Structure and Feature Selection*, v. 3940, Springer Berlin Heidelberg, pp. 34–51, Berlin, Heidelberg. ISBN: 978-3-540-34137-6 978-3-540-34138-3. doi: 10.1007/11752790_2. Disponível em: <http://link.springer.com/10.1007/11752790_2>.
- RUAN, H., YU, J., SHU, F., et al., 2022, “A Joint Stacked Autoencoder Approach with Silhouette Information for Industrial Fault Detection”, *Processes*, v. 10, n. 11 (nov.), pp. 2408. ISSN: 2227-9717. doi: 10.3390/pr10112408. Disponível em: <<https://www.mdpi.com/2227-9717/10/11/2408>>.
- RUFF, L., KAUFFMANN, J. R., VANDERMEULEN, R. A., et al., 2021, “A Unifying Review of Deep and Shallow Anomaly Detection”, *Proceedings of the IEEE*, v. 109, n. 5 (maio), pp. 756–795. ISSN: 0018-9219, 1558-2256. doi: 10.1109/JPROC.2021.3052449. Disponível em: <<https://ieeexplore.ieee.org/document/9347460/>>.
- RUIZ-CÁRCEL, C., CAO, Y., MBA, D., et al., 2015, “Statistical process monitoring of a multiphase flow facility”, *Control Engineering Practice*, v. 42 (set.), pp. 74–88. ISSN: 09670661. doi: 10.1016/j.conengprac.2015.04.012. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066115000866>>.
- RUMELHART, D. E., HINTON, G. E., WILLIAMS, R. J., 1986, “Learning representations by back-propagating errors”, *Nature*, v. 323, n. 6088 (out.), pp. 533–536. ISSN: 1476-4687. doi: 10.1038/323533a0. Disponível em: <<https://www.nature.com/articles/323533a0>>.
- RUSSELL, S., DEWEY, D., TEGMARK, M., 2015, “Research Priorities for Robust and Beneficial Artificial Intelligence”, *AI Magazine*, v. 36, n. 4 (dez.), pp. 105–114. ISSN: 0738-4602, 0738-4602. doi: 10.1609/aimag.v36i4.2577. Disponível em: <<https://aaai.org/ojs/index.php/aimagazine/article/view/2577>>.

- RUSSELL, S. J., NORVIG, P., 2009, *Artificial Intelligence: A Modern Approach*. Edição: 3 ed. Upper Saddle River, Prentice Hall. ISBN: 978-0-13-604259-4.
- SAAFAN, H., ZHU, Q., 2022, “Improved manifold sparse slow feature analysis for process monitoring”, *Computers & Chemical Engineering*, v. 164 (ago.), pp. 107905. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.107905. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422002435>>.
- SABUHI, M., ZHOU, M., BEZEMER, C.-P., et al., 2021, “Applications of Generative Adversarial Networks in Anomaly Detection: A Systematic Literature Review”, *IEEE Access*, v. 9, pp. 161003–161029. ISSN: 2169-3536. doi: 10.1109/ACCESS.2021.3131949.
- SACCENTI, E., CAMACHO, J., 2015, “Determining the number of components in principal components analysis: A comparison of statistical, crossvalidation and approximated methods”, *Chemometrics and Intelligent Laboratory Systems*, v. 149 (dez.), pp. 99–116. ISSN: 01697439. doi: 10.1016/j.chemolab.2015.10.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743915002579>>.
- SACHA, D., SEDLMAIR, M., ZHANG, L., et al., 2017, “What you see is what you can change: Human-centered machine learning by interactive visualization”, *Neurocomputing*, v. 268 (dez.), pp. 164–175. ISSN: 09252312. doi: 10.1016/j.neucom.2017.01.105. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231217307609>>.
- SAD, C. M., DA SILVA, M., DOS SANTOS, F. D., et al., 2019, “Multivariate data analysis applied in the evaluation of crude oil blends”, *Fuel*, v. 239 (mar.), pp. 421–428. ISSN: 00162361. doi: 10.1016/j.fuel.2018.11.045. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0016236118319264>>.
- SADIK, S., GRUENWALD, L., 2014, “Research issues in outlier detection for data streams”, *ACM SIGKDD Explorations Newsletter*, v. 15, n. 1 (mar.), pp. 33–40. ISSN: 19310145. doi: 10.1145/2594473.2594479. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2594473.2594479>>.
- SAINI, K., DHAMI, S. S., VANRAJ, 2022, “Predictive Monitoring of Incipient Faults in Rotating Machinery: A Systematic Review from Data Acquisition to Artificial Intelligence”, *Archives of Computational Methods in Engineering*, v. 29, n. 6 (out.), pp. 4005–4026. ISSN: 1886-1784. doi:

10.1007/s11831-022-09727-6. Disponível em: <<https://doi.org/10.1007/s11831-022-09727-6>>.

SALAHSHOOR, K., KORDESTANI, M., KHOSHRO, M. S., 2010, “Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers”, *Energy*, v. 35, n. 12 (dez.), pp. 5472–5482. ISSN: 0360-5442. doi: 10.1016/j.energy.2010.06.001. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360544210003154>>.

SANCHEZ-MARQUEZ, R., JABALOYES VIVAS, J. M., 2020, “Multivariate SPC methods for controlling manufacturing processes using predictive models – A case study in the automotive sector”, *Computers in Industry*, v. 123 (dez.), pp. 103307. ISSN: 01663615. doi: 10.1016/j.compind.2020.103307. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0166361520305418>>.

SANSANA, J., JOSWIAK, M. N., CASTILLO, I., et al., 2021, “Recent trends on hybrid modeling for Industry 4.0”, *Computers & Chemical Engineering*, v. 151 (ago.), pp. 107365. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107365. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135421001435>>.

SANTANDER, O., KUPPURAJ, V., HARRISON, C. A., et al., 2022, “An open source fluid catalytic cracker - fractionator model to support the development and benchmarking of process control, machine learning and operation strategies”, *Computers & Chemical Engineering*, v. 164 (ago.), pp. 107900. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2022.107900. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422002381>>.

SARTORI, I., AMARO, C. A., SOUZA JÚNIOR, M. B., et al., 2012, “Detecção, Diagnóstico e Correção de Falhas: Uma Proposição Consistente de Definições e Terminologias”, *Ciência & Engenharia*, v. 21 (out.), pp. p. 41–53. ISSN: 1983-4071. Disponível em: <<http://www.seer.ufu.br/index.php/cieng/article/view/13183>>.

SAUNDERS, T., 2013, “Data Science and Data Scientists - What’s in a Name?” *Information Management*, v. 11, pp. 1–3.

SCHAEFFER, J., BRAATZ, R. D., 2022, “Latent Variable Method Demonstrator — software for understanding multivariate data analytics algorithms”, *Computers & Chemical Engineering*, v. 167 (nov.), pp. 108014. ISSN: 0098-1354.

doi: 10.1016/j.compchemeng.2022.108014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135422003489>>.

SCHINKEL, S., DIMIGEN, O., MARWAN, N., 2008, “Selection of recurrence threshold for signal detection”, *The European Physical Journal Special Topics*, v. 164, n. 1 (out.), pp. 45–53. ISSN: 1951-6355, 1951-6401. doi: 10.1140/epjst/e2008-00833-5. Disponível em: <<http://link.springer.com/10.1140/epjst/e2008-00833-5>>.

SCHMID, M., KNEIDINGER, H.-G., ENDISCH, C., 2020, “Data-Driven Fault Diagnosis in Battery Systems through Cross-Cell Monitoring”, *IEEE Sensors Journal*, pp. 1–1. ISSN: 1530-437X, 1558-1748, 2379-9153. doi: 10.1109/JSEN.2020.3017812. Disponível em: <<https://ieeexplore.ieee.org/document/9171345/>>.

SCHMIDL, S., WENIG, P., PAPENBROCK, T., 2022, “Anomaly detection in time series: a comprehensive evaluation”, *Proceedings of the VLDB Endowment*, v. 15, n. 9 (maio), pp. 1779–1797. ISSN: 2150-8097. doi: 10.14778/3538598.3538602. Disponível em: <<https://dl.acm.org/doi/10.14778/3538598.3538602>>.

SCHOLKOPF, B., SMOLA, A., MÜLLER, K.-R., 1998, “Nonlinear Component Analysis as a Kernel Eigenvalue Problem”, *Neural Computation*, v. 10, n. 5 (jul.), pp. 1299–1319. ISSN: 0899-7667. doi: 10.1162/089976698300017467. Disponível em: <<https://www.mitpressjournals.org/doi/10.1162/089976698300017467>>.

SCHOLKOPF, B., PLATT, J. C., SHawe-Taylor, J., et al., 2001, “Estimating the Support of a High-Dimensional Distribution”, *Neural Computation*, v. 13, n. 7 (jul.), pp. 1443–1471. ISSN: 0899-7667, 1530-888X. doi: 10.1162/089976601750264965. Disponível em: <<http://www.mitpressjournals.org/doi/10.1162/089976601750264965>>.

SCHWAAB, M., PINTO, J. C., 2007, *Análise de Dados Experimentais I. Fundamentos de Estatística e Estimação de Parâmetros*, v. 1. Rio de Janeiro, E-papers. Disponível em: <http://www.e-papers.com.br/produtos.asp?codigo_produto=2792>.

SCHWAAB, M., PINTO, J. C., 2011, *Análise de Dados Experimentais II. Planejamento de Experimentos*, v. 2. Rio de Janeiro, E-papers. Disponível em: <www.e-papers.com.br/produtos.asp?codigo_produto=2794&promo=0>.

- SCHWEIDTMANN, A. M., ESCHE, E., FISCHER, A., et al., 2021, “Machine Learning in Chemical Engineering: A Perspective”, *Chemie Ingenieur Technik*, v. 93, n. 12 (dez.), pp. 2029–2039. ISSN: 0009-286X, 1522-2640. doi: 10.1002/cite.202100083. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cite.202100083>>.
- SEBESTA, R. W., 2018, *Concepts of Programming Languages*. 12th edition ed. NY, NY, Pearson. ISBN: 978-0-13-499718-6.
- SENGUPTA, S., BASAK, S., SAIKIA, P., et al., 2020, “A review of deep learning with special emphasis on architectures, applications and recent trends”, *Knowledge-Based Systems*, v. 194 (abr.), pp. 105596. ISSN: 09507051. doi: 10.1016/j.knosys.2020.105596. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095070512030071X>>.
- SENIN, P., 2008, “Dynamic Time Warping Algorithm Review”, Disponível em: <<http://citesearx.ist.psu.edu/viewdoc/download?doi=10.1.465.4905&rep=rep1&type=pdf>>.
- SEVERSON, K., CHAIWATANODOM, P., BRAATZ, R. D., 2016, “Perspectives on process monitoring of industrial systems”, *Annual Reviews in Control*, v. 42, pp. 190–200. ISSN: 13675788. doi: 10.1016/j.arcontrol.2016.09.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S136757881630027X>>.
- SEVERSON, K., MOLARO, M., BRAATZ, R., 2017, “Principal Component Analysis of Process Datasets with Missing Values”, *Processes*, v. 5, n. 38 (jul.). ISSN: 2227-9717. doi: 10.3390/pr5030038. Disponível em: <<http://www.mdpi.com/2227-9717/5/3/38>>.
- SGUEGLIA, A., DI SORBO, A., VISAGGIO, C. A., et al., 2022, “A systematic literature review of IoT time series anomaly detection solutions”, *Future Generation Computer Systems*, v. 134 (set.), pp. 170–186. ISSN: 0167-739X. doi: 10.1016/j.future.2022.04.005. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167739X22001285>>.
- SHABANIAN, M., MONTAZERI, M., 2011, “A neuro-fuzzy online fault detection and diagnosis algorithm for nonlinear and dynamic systems”, *International Journal of Control, Automation and Systems*, v. 9, n. 4 (ago.), pp. 665–670. ISSN: 2005-4092. doi: 10.1007/s12555-011-0407-9. Disponível em: <<https://doi.org/10.1007/s12555-011-0407-9>>.

- SHAH, D., WANG, J., HE, Q. P., 2020, “Feature engineering in big data analytics for IoT-enabled smart manufacturing – Comparison between deep learning and statistical learning”, *Computers & Chemical Engineering*, v. 141 (out.), pp. 106970. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106970. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135420300363>>.
- SHAHNAZARI, H., 2020, “Fault diagnosis of nonlinear systems using recurrent neural networks”, *Chemical Engineering Research and Design*, v. 153 (jan.), pp. 233–245. ISSN: 02638762. doi: 10.1016/j.cherd.2019.09.026. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0263876219304435>>.
- SHANG, C., ZHAO, L., HUANG, X., et al., 2021, “Group-sparsity-enforcing fault discrimination and estimation with dynamic process data”, *Journal of Process Control*, v. 105 (set.), pp. 236–249. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2021.08.003. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152421001281>>.
- SHARMIN, R., SHAH, S. L., SUNDARARAJ, U., 2008, “A PCA Based Fault Detection Scheme for an Industrial High Pressure Polyethylene Reactor”, *Macromolecular Reaction Engineering*, v. 2, n. 1 (fev.), pp. 12–30. ISSN: 1862832X. doi: 10.1002/mren.200700023. Disponível em: <<http://doi.wiley.com/10.1002/mren.200700023>>.
- SHAWE-TAYLOR, J., CRISTIANINI, N., 2004, “Pattern analysis using eigen-decompositions”. In: *Kernel Methods for Pattern Analysis*, Cambridge University Press, pp. 140–194. Disponível em: <<https://doi.org/10.1017/CBO9780511809682.007>>.
- SHERIFF, M. Z., MANSOURI, M., KARIM, M. N., et al., 2017, “Fault detection using multiscale PCA-based moving window GLRT”, *Journal of Process Control*, v. 54 (jun.), pp. 47–64. ISSN: 09591524. doi: 10.1016/j.jprocont.2017.03.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152417300689>>.
- SHEWHART, W. A., 1926, “Quality control charts”, *The Bell System Technical Journal*, v. 5, n. 4 (out.), pp. 593–603. ISSN: 0005-8580. doi: 10.1002/j.1538-7305.1926.tb00125.x. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1926.tb00125.x>>.
- SHEWHART, W. A., 1931, *Economic control of quality of manufactured product*. Lancaster, D. Van Nostrand Company, Inc.

- SHI, L., LIANG, X., 2011, “An Object-Oriented Approach for Structure Design of Property Calculation Programs Using Equations of State”, *Industrial & Engineering Chemistry Research*, v. 50, n. 10 (maio), pp. 6404–6412. ISSN: 0888-5885. doi: 10.1021/ie101547t. Disponível em: <<https://doi.org/10.1021/ie101547t>>.
- SHLENS, J., 2005, “A Tutorial on Principal Component Analysis - v2.0”, Disponível em: <<https://www.cs.cmu.edu/~elaw/papers/pca.pdf>>.
- SHLENS, J., 2014, “A Tutorial on Principal Component Analysis - v3.0.2”, *arXiv:1404.1100 [cs, stat]*, (abr.). doi: 10.48550/arXiv.1404.1100. Disponível em: <<http://arxiv.org/abs/1404.1100>>.
- SHOJAIE, A., FOX, E. B., 2022, “Granger Causality: A Review and Recent Advances”, *Annual Review of Statistics and Its Application*, v. 9, n. 1 (mar.), pp. 289–319. ISSN: 2326-8298, 2326-831X. doi: 10.1146/annurev-statistics-040120-010930. Disponível em: <<https://www.annualreviews.org/doi/10.1146/annurev-statistics-040120-010930>>.
- SHRIVASTAVA, R., 2021, “Comparative study of boosting and bagging based methods for fault detection in a chemical process”. In: *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 674–679, mar. doi: 10.1109/ICAIS50930.2021.9395905.
- SHU, Y., MING, L., CHENG, F., et al., 2016, “Abnormal situation management: Challenges and opportunities in the big data era”, *Computers & Chemical Engineering*, v. 91 (ago.), pp. 104–113. ISSN: 00981354. doi: 10.1016/j.compchemeng.2016.04.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135416301077>>.
- SI, Y., CHEN, Z., SUN, J., et al., 2020, “A Data-Driven Fault Detection Framework Using Mahalanobis Distance Based Dynamic Time Warping”, *IEEE Access*, pp. 1–1. ISSN: 2169-3536. doi: 10.1109/ACCESS.2020.3001379. Disponível em: <<https://ieeexplore.ieee.org/document/9113317/>>.
- SIANG, L. C., ELNAAWI, S., RIPPON, L. D., et al., 2023. “Data Quality Over Quantity: Pitfalls and Guidelines for Process Analytics”. abr. Disponível em: <<http://arxiv.org/abs/2211.06440>>. arXiv:2211.06440 [cs, eess].
- SILVA, A., SARRAGUÇA, M., FONTEYNE, M., et al., 2017, “Multivariate statistical process control of a continuous pharmaceutical twin-screw granulation and fluid bed drying process”, *International Journal of Phar-*

maceutics, v. 528, n. 1-2 (ago.), pp. 242–252. ISSN: 03785173. doi: 10.1016/j.ijpharm.2017.05.075. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0378517317304994>>.

SILVA, A., VERCROYSSE, J., VERVAET, C., et al., 2018, “Process monitoring and evaluation of a continuous pharmaceutical twin-screw granulation and drying process using multivariate data analysis”, *European Journal of Pharmaceutics and Biopharmaceutics*, v. 128 (jul.), pp. 36–47. ISSN: 09396411. doi: 10.1016/j.ejpb.2018.04.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0939641117312869>>.

SILVA, A. F., VERCROYSSE, J., VERVAET, C., et al., 2019, “In-Depth Evaluation of Data Collected During a Continuous Pharmaceutical Manufacturing Process: A Multivariate Statistical Process Monitoring Approach”, *Journal of Pharmaceutical Sciences*, v. 108, n. 1 (jan.), pp. 439–450. ISSN: 00223549. doi: 10.1016/j.xphs.2018.07.033. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0022354918305112>>.

SILVA, R. G., WILCOX, S. J., 2019, “Feature evaluation and selection for condition monitoring using a self-organizing map and spatial statistics”, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, v. 33, n. 1 (fev.), pp. 1–10. ISSN: 0890-0604, 1469-1760. doi: 10.1017/S0890060417000518. Disponível em: <https://www.cambridge.org/core/product/identifier/S0890060417000518/type/journal_article>.

SIMOGLOU, A., MARTIN, E., MORRIS, A., et al., 1997, “Multivariate Statistical Process Control in Chemicals Manufacturing”, *IFAC Proceedings Volumes*, v. 30, n. 18 (ago.), pp. 21–28. ISSN: 14746670. doi: 10.1016/S1474-6670(17)42375-5. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017423755>>.

SINGHAL, A., SEBORG, D. E., 2002, “Pattern Matching in Multivariate Time Series Databases Using a Moving-Window Approach”, *Industrial & Engineering Chemistry Research*, v. 41, n. 16 (ago.), pp. 3822–3838. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie010517z. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie010517z>>.

SIVALINGAM, G., SONI, N. J., VAKIL, S. M., 2015, “Detection of decomposition for high pressure ethylene/vinyl acetate copolymerization in autoclave reactor using principal component analysis on heat balance model”, *The*

Canadian Journal of Chemical Engineering, v. 93, n. 6 (jun.), pp. 1063–1075. ISSN: 00084034. doi: 10.1002/cjce.22200. Disponível em: <<http://doi.wiley.com/10.1002/cjce.22200>>.

SIVARAJAH, U., KAMAL, M. M., IRANI, Z., et al., 2017, “Critical analysis of Big Data challenges and analytical methods”, *Journal of Business Research*, v. 70 (jan.), pp. 263–286. ISSN: 01482963. doi: 10.1016/j.jbusres.2016.08.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S014829631630488X>>.

SIVARAM, A., DAS, L., VENKATASUBRAMANIAN, V., 2020, “Hidden representations in deep neural networks: Part 1. Classification problems”, *Computers & Chemical Engineering*, v. 134 (mar.), pp. 106669. ISSN: 00981354. doi: 10.1016/j.compchemeng.2019.106669. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419306878>>.

SKIENA, S. S., 2017, *The data science design manual*. Texts in Computer Science. New York, NY, Springer Berlin Heidelberg. ISBN: 978-3-319-55443-3.

SKVARA, V., PEVNY, T., SMIDL, V., 2018, “Are generative deep models for novelty detection truly better?” *arXiv:1807.05027 [cs, stat]*, (jul.). Disponível em: <<http://arxiv.org/abs/1807.05027>>.

SLAMA, C. F., 1991, *Multivariate Statistical Analysis of Data from an Industrial Fluidized Catalytic Cracking Process Using PCA and PLS*. Dissertação de Mestrado, McMaster University. Disponível em: <<https://macsphere.mcmaster.ca/handle/11375/8066>>.

SLISKOVIC, D., GRBIC, R., HOCENSKI, Z., 2011, “Methods for Plant Data-Based Process Modeling in Soft-Sensor Development”, *Automatika*, v. 52, n. 4 (jan.), pp. 306–318. ISSN: 0005-1144, 1848-3380. doi: 10.1080/00051144.2011.11828430. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/00051144.2011.11828430>>.

SMOLA, A. J., SCHOLKOPF, B., 2004, “A tutorial on support vector regression”, *Statistics and Computing*, v. 14, n. 3 (ago.), pp. 199–222. ISSN: 0960-3174. doi: 10.1023/B:STCO.0000035301.49549.88. Disponível em: <<http://link.springer.com/10.1023/B:STCO.0000035301.49549.88>>.

SOARES, R. M., 2022, *Data-Driven Monitoring Enhanced by Process Knowledge*. Tese de Doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

- SOAVE, G., 1972, "Equilibrium constants from a modified Redlich-Kwong equation of state", *Chemical Engineering Science*, v. 27, n. 6 (jun.), pp. 1197–1203. ISSN: 0009-2509. doi: 10.1016/0009-2509(72)80096-4. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0009250972800964>>.
- SONG, B., SHI, H., TAN, S., et al., 2020, "Multi-Subspace Orthogonal Canonical Correlation Analysis for Quality Related Plant Wide Process Monitoring", *IEEE Transactions on Industrial Informatics*, pp. 1–1. ISSN: 1551-3203, 1941-0050. doi: 10.1109/TII.2020.3015034. Disponível em: <<https://ieeexplore.ieee.org/document/9162498/>>.
- SONG, P., ZHAO, C., 2022, "Slow Down to Go Better: A Survey on Slow Feature Analysis", *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–21. ISSN: 2162-2388. doi: 10.1109/TNNLS.2022.3201621.
- SOUZA, F. A. A., ARAÚJO, R., MENDES, J., 2016, "Review of soft sensor methods for regression applications", *Chemometrics and Intelligent Laboratory Systems*, v. 152 (mar.), pp. 69–79. ISSN: 0169-7439. doi: 10.1016/j.chemolab.2015.12.011. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169743915003263>>.
- SOUZA, V. M. A., DOS REIS, D. M., MALETZKE, A. G., et al., 2020, "Challenges in benchmarking stream learning algorithms with real-world data", *Data Mining and Knowledge Discovery*, v. 34, n. 6 (nov.), pp. 1805–1858. ISSN: 1573-756X. doi: 10.1007/s10618-020-00698-5. Disponível em: <<https://doi.org/10.1007/s10618-020-00698-5>>.
- SPIES, N. C., FARNSWORTH, C. W., JACKUPS, R., 2023, "Data-Driven Anomaly Detection in Laboratory Medicine: Past, Present, and Future", *The Journal of Applied Laboratory Medicine*, v. 8, n. 1 (jan.), pp. 162–179. ISSN: 2475-7241. doi: 10.1093/jalm/jfac114. Disponível em: <<https://academic.oup.com/jalm/article/8/1/162/6887080>>.
- STAUDACHER, M., TELSER, S., AMANN, A., et al., 2005, "A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep", *Physica A: Statistical Mechanics and its Applications*, v. 349, n. 3-4 (abr.), pp. 582–596. ISSN: 03784371. doi: 10.1016/j.physa.2004.10.026. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0378437104013640>>.
- STEINMACHER, I., ROBLES, G., FITZGERALD, B., et al., 2017, "Free and open source software development: the end of the teenage ye-

ars”, *Journal of Internet Services and Applications*, v. 8, n. 1 (dez.). ISSN: 1867-4828, 1869-0238. doi: 10.1186/s13174-017-0069-9. Disponível em: <<https://jisajournal.springeropen.com/articles/10.1186/s13174-017-0069-9>>.

STEINWANDTER, V., BORCHERT, D., HERWIG, C., 2019, “Data science tools and applications on the way to Pharma 4.0”, *Drug Discovery Today*, (jun.). ISSN: 13596446. doi: 10.1016/j.drudis.2019.06.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1359644618305324>>.

STIEF, A., TAN, R., CAO, Y., et al., 2019a, “PRONTO heterogeneous benchmark dataset”, (maio). doi: 10.5281/zenodo.1341583. Disponível em: <<https://zenodo.org/record/1341583>>.

STIEF, A., TAN, R., CAO, Y., et al., 2019b, “A heterogeneous benchmark dataset for data analytics: Multiphase flow facility case study”, *Journal of Process Control*, v. 79 (jul.), pp. 41–55. ISSN: 09591524. doi: 10.1016/j.jprocont.2019.04.009. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152418303603>>.

STOCKMANN, M., HABER, R., SCHMITZ, U., 2012, “Source identification of plant-wide faults based on k nearest neighbor time delay estimation”, *Journal of Process Control*, v. 22, n. 3 (mar.), pp. 583–598. ISSN: 09591524. doi: 10.1016/j.jprocont.2012.01.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095915241200008X>>.

SUBBARAJ, P., KANNAPIRAN, B., 2014, “Fault detection and diagnosis of pneumatic valve using Adaptive Neuro-Fuzzy Inference System approach”, *Applied Soft Computing*, v. 19 (jun.), pp. 362–371. ISSN: 1568-4946. doi: 10.1016/j.asoc.2014.02.008. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494614000775>>.

SUN, B., WANG, J., HOU, B., et al., 2017, “A data-driven fault detection toolbox based on MATLAB GUIDE”. In: *2017 Chinese Automation Congress (CAC)*, pp. 4326–4331, Jinan, out. IEEE. ISBN: 978-1-5386-3524-7. doi: 10.1109/CAC.2017.8243540. Disponível em: <<http://ieeexplore.ieee.org/document/8243540/>>.

SUN, D., LU, G., ZHOU, H., et al., 2013, “Condition Monitoring of Combustion Processes Through Flame Imaging and Kernel Principal Component Analysis”, *Combustion Science and Technology*, v. 185, n. 9 (set.), pp. 1400–1413.

ISSN: 0010-2202, 1563-521X. doi: 10.1080/00102202.2013.798316. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/00102202.2013.798316>>.

SUN, L., LIANG, F., CUI, W., 2021, “Artificial Neural Network and Its Application Research Progress in Chemical Process”, *arXiv:2110.09021 [cs, eess]*, (out.). Disponível em: <<http://arxiv.org/abs/2110.09021>>.

SUN, L., YOU, G., CAO, X., et al., 2019, “Comparative investigation for raw and processed Aconiti Lateralis Radix using chemical UPLC-MS profiling and multivariate classification techniques”, *Journal of Food and Drug Analysis*, v. 27, n. 1 (jan.), pp. 365–372. ISSN: 10219498. doi: 10.1016/j.jfda.2018.10.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1021949818301649>>.

SUN, Q., GE, Z., 2021, “A Survey on Deep Learning for Data-Driven Soft Sensors”, *IEEE Transactions on Industrial Informatics*, v. 17, n. 9 (set.), pp. 5853–5866. ISSN: 1941-0050. doi: 10.1109/TII.2021.3053128. Disponível em: <<https://ieeexplore.ieee.org/document/9329169>>.

SUN, R., 2020, “Multiblock global orthogonal projections to latent structures for fault diagnosis”, *Chemometrics and Intelligent Laboratory Systems*, v. 204 (set.), pp. 104092. ISSN: 01697439. doi: 10.1016/j.chemolab.2020.104092. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743920300617>>.

SUN, R., WANG, Y., 2021, “C IPLS IKPLS for Modeling and Detecting Nonlinear Multimode Processes”, *Industrial & Engineering Chemistry Research*, v. 60, n. 4 (fev.), pp. 1684–1698. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c04038. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.0c04038>>.

SUN, W., BRAATZ, R. D., 2021, “Smart process analytics for predictive modeling”, *Computers & Chemical Engineering*, v. 144 (jan.), pp. 107134. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.107134. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135420304555>>.

SUN, Z., TANG, J., QIAO, J., et al., 2020, “Review of Concept Drift Detection Method for Industrial Process Modeling”. In: *2020 39th Chinese Control Conference (CCC)*, pp. 5754–5759, jul. doi: 10.23919/CCC50068.2020.9189106.

- TADEU, H. F. B., DE CASTRO MOURA DUARTE, A. L., TAURION, C., et al., 2019, “Digital Transformation: Digital Maturity Applied to Study Brazilian Perspective for Industry 4.0”. In: Alcaraz, J. L. G., Rivera Cadavid, L., González-Ramírez, R. G., et al. (Eds.), *Best Practices in Manufacturing Processes*, Springer International Publishing, pp. 3–27, Cham. ISBN: 978-3-319-99189-4 978-3-319-99190-0. doi: 10.1007/978-3-319-99190-0_1. Disponível em: <http://link.springer.com/10.1007/978-3-319-99190-0_1>.
- TALAGALA, P. D., HYNDMAN, R. J., SMITH-MILES, K., et al., 2019, “Anomaly Detection in Streaming Nonstationary Temporal Data”, *Journal of Computational and Graphical Statistics*, (maio). ISSN: 1061-8600, 1537-2715. doi: 10.1080/10618600.2019.1617160. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/10618600.2019.1617160>>.
- TAMURA, M., TSUJITA, S., 2007, “A study on the number of principal components and sensitivity of fault detection using PCA”, *Computers & Chemical Engineering*, v. 31, n. 9 (set.), pp. 1035–1046. ISSN: 00981354. doi: 10.1016/j.compchemeng.2006.09.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135406002419>>.
- TAN, S., MAYROVOUNIOTIS, M. L., 1995, “Reducing data dimensionality through optimizing neural network inputs”, *AICHE Journal*, v. 41, n. 6 (jun.), pp. 1471–1480. ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.690410612. Disponível em: <<http://doi.wiley.com/10.1002/aic.690410612>>.
- TANG, P., PENG, K., ZHANG, K., et al., 2018, “A Deep Belief Network-based Fault Detection Method for Nonlinear Processes”, *IFAC-PapersOnLine*, v. 51, n. 24 (jan.), pp. 9–14. ISSN: 2405-8963. doi: 10.1016/j.ifacol.2018.09.522. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2405896318322134>>.
- TANG, Y., ZHANG, S., 2020, “Fault detection of FWTPs in coal-fired power plants using K-WD-KPCA in consideration of multiple operation conditions”, *Asia-Pacific Journal of Chemical Engineering*, (nov.). ISSN: 1932-2135, 1932-2143. doi: 10.1002/apj.2599. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/apj.2599>>.
- TAQVI, S. A. A., ZABIRI, H., TUFA, L. D., et al., 2021, “A Review on Data Driven Learning Approaches for Fault Detection and Diagnosis in Chemical Processes”, *ChemBioEng Reviews*, (fev.), pp. cben.202000027. ISSN: 2196-

9744, 2196-9744. doi: 10.1002/cben.202000027. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cben.202000027>>.

TARASSENKO, L., CLIFTON, D. A., BANNISTER, P. R., et al., 2009, “Novelty Detection”. In: Boller, C., Chang, F.-K., Fujino, Y. (Eds.), *Encyclopedia of structural health monitoring*, John Wiley, Chichester, West Sussex, U.K. ISBN: 978-0-470-05822-0 978-0-470-06162-6. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470061626.shm183>>.

TASKESEN, E., 2022. “findpeaks”. Disponível em: <<https://github.com/erdogant/findpeaks>>.

TATES, A. A., LOUWERSE, D. J., SMILDE, A. K., et al., 1999, “Monitoring a PVC Batch Process with Multivariate Statistical Process Control Charts”, *Industrial & Engineering Chemistry Research*, v. 38, n. 12 (dez.), pp. 4769–4776. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie9901067. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie9901067>>.

TAX, D. M., DUIN, R. P., 1999, “Support vector domain description”, *Pattern Recognition Letters*, v. 20, n. 11-13 (nov.), pp. 1191–1199. ISSN: 01678655. doi: 10.1016/S0167-8655(99)00087-2. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0167865599000872>>.

TAYLOR, A., 1998, “The Application of Principal Component Analysis for Predicting Blast Furnace Stability”, *IFAC Proceedings Volumes*, v. 31, n. 23 (set.), pp. 223–226. ISSN: 14746670. doi: 10.1016/S1474-6670(17)35884-6. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017358846>>.

TAYLOR, M. P., BOXALL, P., CHEN, J. J. J., et al., 2020, “Operator 4.0 or Maker 1.0? Exploring the implications of Industrie 4.0 for innovation, safety and quality of work in small economies and enterprises”, *Computers & Industrial Engineering*, v. 139 (jan.), pp. 105486. ISSN: 0360-8352. doi: 10.1016/j.cie.2018.10.047. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360835218305278>>.

TEPPOLA, P., MINKKINEN, P., 1999, “Possibilistic and fuzzy C-means clustering for process monitoring in an activated sludge waste-water treatment plant”, *Journal of Chemometrics*, v. 13, n. 3-4 (maio), pp. 445–459. ISSN: 0886-9383, 1099-128X. doi: 10.1002/(SICI)1099-128X(199905/08)13:3/4<445::AID-CEM557>3.0.CO;2-W. Disponível em: <[http:](http://)

//doi.wiley.com/10.1002/%28SICI%291099-128X%28199905/08%
2913%3A3/4%3C445%3A%3AAID-CEM557%3E3.0.CO%3B2-W>.

TEPPOLA, P., MINKKINEN, P., 2000, "Wavelet-PLS regression models for both exploratory data analysis and process monitoring", *Journal of Chemometrics*, v. 14, n. 5-6 (set.), pp. 383–399. ISSN: 0886-9383, 1099-128X. doi: 10.1002/1099-128X(200009/12)14:5/6<383::AID-CEM616>3.0.CO;2-5. Disponível em: <<http://doi.wiley.com/10.1002/1099-128X%28200009/12%2914%3A5/6%3C383%3A%3AAID-CEM616%3E3.0.CO%3B2-5>>.

TEPPOLA, P., MUJUNEN, S.-P., MINKKINEN, P., 1998a, "A combined approach of partial least squares and fuzzy c-means clustering for the monitoring of an activated-sludge waste-water treatment plant", *Chemometrics and Intelligent Laboratory Systems*, v. 41, n. 1 (jul.), pp. 95–103. ISSN: 01697439. doi: 10.1016/S0169-7439(98)00026-4. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743998000264>>.

TEPPOLA, P., MUJUNEN, S.-P., MINKKINEN, P., et al., 1998b, "Principal component analysis, contribution plots and feature weights in the monitoring of sequential process data from a paper machine's wet end", *Chemometrics and Intelligent Laboratory Systems*, v. 44, n. 1-2 (dez.), pp. 307–317. ISSN: 01697439. doi: 10.1016/S0169-7439(98)00188-9. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743998001889>>.

TESSIER, J., DUCHESNE, C., TARCY, G. P., et al., 2012, "Multivariate Analysis and Monitoring of the Performance of Aluminum Reduction Cells", *Industrial & Engineering Chemistry Research*, v. 51, n. 3 (jan.), pp. 1311–1323. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie201258b. Disponível em: <<http://pubs.acs.org/doi/10.1021/ie201258b>>.

THARRAULT, Y., MOUROT, G., RAGOT, J., et al., 2008, "Fault Detection and Isolation with Robust Principal Component Analysis", *International Journal of Applied Mathematics and Computer Science*, v. 18, n. 4 (dez.), pp. 429–442. ISSN: 1641-876X. doi: 10.2478/v10006-008-0038-3. Disponível em: <<http://content.sciendo.com/view/journals/amcs/18/4/article-p429.xml>>.

THARWAT, A., 2018, "Independent component analysis: An introduction", *Applied Computing and Informatics*, (ago.). ISSN: 22108327. doi: 10.1016/j.aci.2018.08.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2210832718301819>>.

- THOMAS, C., WADA, T., SEBORG, D. E., 1996, “Principal Component Analysis Applied to Process Monitoring of an Industrial Distillation Column”, *IFAC Proceedings Volumes*, v. 29, n. 1 (jun.), pp. 5859–5864. ISSN: 14746670. doi: 10.1016/S1474-6670(17)58618-8. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017586188>>.
- THOMAS, M. C., ZHU, W., ROMAGNOLI, J. A., 2018, “Data mining and clustering in chemical process databases for monitoring and knowledge discovery”, *Journal of Process Control*, v. 67 (jul.), pp. 160–175. ISSN: 09591524. doi: 10.1016/j.jprocont.2017.02.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095915241730032X>>.
- THORNHILL, N. F., MELBØ, H., WIIK, J., 2006, “Multidimensional Visualization and Clustering of Historical Process Data”, *Industrial & Engineering Chemistry Research*, v. 45, n. 17 (ago.), pp. 5971–5985. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie051054q. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie051054q>>.
- THURLIMANN, C. M., VILLEZ, K., 2017, “Input estimation as a qualitative trend analysis problem”, *Computers & Chemical Engineering*, v. 107 (dez.), pp. 333–342. ISSN: 00981354. doi: 10.1016/j.compchemeng.2017.04.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135417301680>>.
- THURLIMANN, C. M., DÜRRENMATT, D. J., VILLEZ, K., 2018, “Soft-sensing with qualitative trend analysis for wastewater treatment plant control”, *Control Engineering Practice*, v. 70 (jan.), pp. 121–133. ISSN: 09670661. doi: 10.1016/j.conengprac.2017.09.015. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066117302186>>.
- TIAN, W., LIU, Z., LI, L., et al., 2020, “Identification of abnormal conditions in high-dimensional chemical process based on feature selection and deep learning”, *Chinese Journal of Chemical Engineering*, v. 28, n. 7 (jul.), pp. 1875–1883. ISSN: 1004-9541. doi: 10.1016/j.cjche.2020.05.003. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1004954120302135>>.
- TIAN, W., WANG, S., SUN, S., et al., 2022, “Intelligent prediction and early warning of abnormal conditions for fluid catalytic cracking process”, *Chemical Engineering Research and Design*, v. 181 (maio), pp. 304–320. ISSN: 0263-8762. doi: 10.1016/j.cherd.2022.03.031. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S026387622200137X>>.

- TIDRIRI, K., CHATTI, N., VERRON, S., et al., 2016, “Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges”, *Annual Reviews in Control*, v. 42, pp. 63–81. ISSN: 13675788. doi: 10.1016/j.arcontrol.2016.09.008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1367578816300669>>.
- TIEN, D., KHIANG-WEE LIM, LIU JUN, 2004, “Comparative study of PCA approaches in process monitoring and fault detection”. In: *30th Annual Conference of IEEE Industrial Electronics Society, 2004. IECON 2004*, v. 3, pp. 2594–2599, Busan, South Korea. IEEE. ISBN: 978-0-7803-8730-0. doi: 10.1109/IECON.2004.1432212. Disponível em: <<http://ieeexplore.ieee.org/document/1432212/>>.
- TIPPING, M. E., BISHOP, C. M., 1999, “Probabilistic principal component analysis”, *Journal of Royal Statistical Society*, v. 61, n. 3, pp. 611–622. Disponível em: <<https://www.jstor.org/stable/2680726>>.
- TRAN, D., 2019, “Automated Change Detection and Reactive Clustering in Multivariate Streaming Data”. In: *2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF)*, mar. doi: 10.1109/RIVF.2019.8713738. Disponível em: <<https://arxiv.org/abs/1311.0505>>.
- TRUONG, C., OUDRE, L., VAYATIS, N., 2020, “Selective review of offline change point detection methods”, *Signal Processing*, v. 167 (fev.), pp. 107299. ISSN: 0165-1684. doi: 10.1016/j.sigpro.2019.107299. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0165168419303494>>.
- TSAI, Y., LU, Q., RIPPON, L., et al., 2018, “Pattern and Knowledge Extraction using Process Data Analytics: A Tutorial”, *IFAC-PapersOnLine*, v. 51, n. 18, pp. 13–18. ISSN: 24058963. doi: 10.1016/j.ifacol.2018.09.237. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S240589631831913X>>.
- TSEKICHUN, 2021. “Neural network explain”.
- TSYMBAL, A., 2004, “The problem of concept drift: definitions and related work”, maio. Disponível em: <<https://www.researchgate.net/publication/228723141>>.

- TÔRRES, A. R., GRANGEIRO, S., FRAGOSO, W. D., 2015, “Multivariate control charts for monitoring captopril stability”, *Microchemical Journal*, v. 118 (jan.), pp. 259–265. ISSN: 0026265X. doi: 10.1016/j.microc.2014.07.017. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0026265X14001507>>.
- TÔRRES, A. R., DE OLIVEIRA, A. D. P., GRANGEIRO, S., et al., 2018, “Multivariate statistical process control in annual pharmaceutical product review”, *Journal of Process Control*, v. 69 (set.), pp. 97–102. ISSN: 09591524. doi: 10.1016/j.jprocont.2018.06.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095915241830115X>>.
- UDUGAMA, I. A., GARGALO, C. L., YAMASHITA, Y., et al., 2020, “The Role of Big Data in Industrial (Bio)chemical Process Operations”, *Industrial & Engineering Chemistry Research*, v. 59, n. 34 (ago.), pp. 15283–15297. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c01872. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.0c01872>>.
- UDUGAMA, I. A., BAYER, C., BAROUTIAN, S., et al., 2022, “Digitalisation in chemical engineering: Industrial needs, academic best practice, and curriculum limitations”, *Education for Chemical Engineers*, v. 39 (abr.), pp. 94–107. ISSN: 1749-7728. doi: 10.1016/j.ece.2022.03.003. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1749772822000112>>.
- UL HASSAN, M., REHMANI, M. H., CHEN, J., 2023, “Anomaly Detection in Blockchain Networks: A Comprehensive Survey”, *IEEE Communications Surveys & Tutorials*, v. 25, n. 1, pp. 289–318. ISSN: 1553-877X. doi: 10.1109/COMST.2022.3205643.
- UNDEY, C., TATARA, E., WILLIAMS, B., et al., 2000, “A hybrid supervisory knowledge-based system for monitoring penicillin fermentation”. In: *Proceedings of the 2000 American Control Conference. ACC (IEEE Cat. No.00CH36334)*, pp. 3944–3948 vol.6, Chicago, IL, USA. IEEE. ISBN: 978-0-7803-5519-4. doi: 10.1109/ACC.2000.876962. Disponível em: <<http://ieeexplore.ieee.org/document/876962/>>.
- UNDEY, C., ÇINAR, A., 2002, “Statistical monitoring of multistage, multiphase batch processes”, *IEEE Control Systems*, v. 22, n. 5 (out.), pp. 40–52. ISSN: 1066-033X, 1941-000X. doi: 10.1109/MCS.2002.1035216. Disponível em: <<https://ieeexplore.ieee.org/document/1035216/>>.

- UPPAL, A., RAY, W., POORE, A., 1974, “On the dynamic behavior of continuous stirred tank reactors”, *Chemical Engineering Science*, v. 29, n. 4 (abr.), pp. 967–985. ISSN: 00092509. doi: 10.1016/0009-2509(74)80089-8. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0009250974800898>>.
- VALLE, S., LI, W., QIN, S. J., 1999, “Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods”, *Industrial & Engineering Chemistry Research*, v. 38, n. 11 (nov.), pp. 4389–4401. ISSN: 0888-5885. doi: 10.1021/ie990110i. Disponível em: <<https://doi.org/10.1021/ie990110i>>.
- VAN DE VUSSE, J. G., 1964, “Plug-flow type reactor versus tank reactor”, *Chemical Engineering Science*, v. 19, n. 12 (dez.), pp. 994–996. ISSN: 0009-2509. doi: 10.1016/0009-2509(64)85109-5. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0009250964851095>>.
- VAN DEN KERKHOF, P., VANLAER, J., GINS, G., et al., 2013, “Analysis of smearing-out in contribution plot based fault isolation for Statistical Process Control”, *Chemical Engineering Science*, v. 104 (dez.), pp. 285–293. ISSN: 00092509. doi: 10.1016/j.ces.2013.08.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0009250913005502>>.
- VAN IMPE, J., GINS, G., 2015, “An extensive reference dataset for fault detection and identification in batch processes”, *Chemometrics and Intelligent Laboratory Systems*, v. 148 (nov.), pp. 20–31. ISSN: 01697439. doi: 10.1016/j.chemolab.2015.08.019. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743915002105>>.
- VAN KOLLENBURG, G. H., VAN ES, J., GERRETZEN, J., et al., 2020, “Understanding chemical production processes by using PLS path model parameters as soft sensors”, *Computers & Chemical Engineering*, v. 139 (ago.), pp. 106841. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106841. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419312669>>.
- VANDERPLAS, J., 2017, *Python Data Science Handbook*. Sebastopol, O'Reilly Media, Incorporated. ISBN: 978-1-4919-1205-8.
- VANHATALO, E., 2010, “Multivariate process monitoring of an experimental blast furnace”, *Quality and Reliability Engineering International*, v. 26, n. 5 (jul.), pp. 495–508. ISSN: 07488017. doi: 10.1002/qre.1070. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/qre.1070>>.

- VARGAS, R. E. V., MUNARO, C. J., CIARELLI, P. M., et al., 2019, “A realistic and public dataset with rare undesirable real events in oil wells”, *Journal of Petroleum Science and Engineering*, v. 181 (out.). ISSN: 09204105. doi: 10.1016/j.petrol.2019.106223. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0920410519306357>>.
- VASWANI, A., SHAZEER, N., PARMAR, N., et al., 2017. “Attention Is All You Need”. dez. Disponível em: <<http://arxiv.org/abs/1706.03762>>. arXiv:1706.03762 [cs].
- VENKATASUBRAMANIAN, V., 2019, “The promise of artificial intelligence in chemical engineering: Is it here, finally?” *AIChe Journal*, v. 65, n. 2 (fev.), pp. 466–478. ISSN: 00011541. doi: 10.1002/aic.16489. Disponível em: <<http://doi.wiley.com/10.1002/aic.16489>>.
- VENKATASUBRAMANIAN, V., CHAN, K., 1989, “A neural network methodology for process fault diagnosis”, *AIChe Journal*, v. 35, n. 12 (dez.), pp. 1993–2002. ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.690351210. Disponível em: <<http://doi.wiley.com/10.1002/aic.690351210>>.
- VENKATASUBRAMANIAN, V., RENGASWAMY, R., KA, S. N., et al., 2003a, “A review of process fault detection and diagnosis Part III: Process history based methods”, *Computers and Chemical Engineering*, pp. 327–346. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S009813540200162X>>.
- VENKATASUBRAMANIAN, V., RENGASWAMY, R., YIN, K., et al., 2003b, “A review of process fault detection and diagnosis Part I: Quantitative model-based methods”, *Computers and Chemical Engineering*, pp. 293–311. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135402001606>>.
- VERRON, S., LI, J., TIPLICA, T., 2010, “Fault detection and isolation of faults in a multivariate process with Bayesian network”, *Journal of Process Control*, v. 20, n. 8 (set.), pp. 902–911. ISSN: 09591524. doi: 10.1016/j.jprocont.2010.06.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152410001125>>.
- VIDAL-PUIG, S., VITALE, R., FERRER, A., 2019, “Data-driven supervised fault diagnosis methods based on latent variable models: a comparative study”, *Chemometrics and Intelligent Laboratory Systems*, v. 187 (abr.), pp. 41–52. ISSN: 01697439. doi: 10.1016/j.chemolab.2019.02.006.

Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743918304076>>.

VILLALBA, P., SANCHIS, J., FERRER, A., 2019, “A graphical user interface for PCA-based MSPC: A benchmark software for multivariate statistical process control in MATLAB”, *Chemometrics and Intelligent Laboratory Systems*, v. 185 (fev.), pp. 135–152. ISSN: 01697439. doi: 10.1016/j.chemolab.2018.12.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743917308110>>.

VILLALBA-DIEZ, J., ORDIERES-MERÉ, J., 2021, “Human–Machine Integration in Processes within Industry 4.0 Management”, *Sensors*, v. 21, n. 17 (jan.), pp. 5928. ISSN: 1424-8220. doi: 10.3390/s21175928. Disponível em: <<https://www.mdpi.com/1424-8220/21/17/5928>>.

VILLEZ, K., ROSÉN, C., ANCTIL, F., et al., 2008, “Qualitative representation of trends: an alternative approach to process diagnosis and control”, *Water Science and Technology*, v. 57, n. 10 (maio), pp. 1525–1532. ISSN: 0273-1223, 1996-9732. doi: 10.2166/wst.2008.141. Disponível em: <<https://iwaponline.com/wst/article/57/10/1525/13662/Qualitative-representation-of-trends-an>>.

VILLEZ, K., 2015, “Qualitative path estimation: A fast and reliable algorithm for qualitative trend analysis”, *AIChE Journal*, v. 61, n. 5 (maio), pp. 1535–1546. ISSN: 00011541. doi: 10.1002/aic.14736. Disponível em: <<http://doi.wiley.com/10.1002/aic.14736>>.

VILLEZ, K., KESER, B., RIEGER, L., 2009, “Qualitative representation of trends (QRT) as a tool for automated data-driven monitoring of on-line sensors”, *IFAC Proceedings Volumes*, v. 42, n. 8, pp. 1276–1281. ISSN: 14746670. doi: 10.3182/20090630-4-ES-2003.00209. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667016359523>>.

VON BIRGELEN, A., BURATTI, D., MAGER, J., et al., 2018, “Self-Organizing Maps for Anomaly Localization and Predictive Maintenance in Cyber-Physical Production Systems”, *Procedia CIRP*, v. 72 (jan.), pp. 480–485. ISSN: 2212-8271. doi: 10.1016/j.procir.2018.03.150. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S221282711830307X>>.

VU-QUOC, L., 2018. “Biological neuron model”. Disponível em: <https://en.wikipedia.org/w/index.php?title=Biological_neuron_model&oldid=906448410>.

- WALCZAK, B., MASSART, D. L., 1995, “Robust principal components regression as a detection tool for outliers”, *Chemometrics and Intelligent Laboratory Systems*, v. 27, n. 1 (jan.), pp. 41–54. ISSN: 0169-7439. doi: 10.1016/0169-7439(95)80006-U. Disponível em: <<http://www.sciencedirect.com/science/article/pii/016974399580006U>>.
- WANG, D., ROMAGNOLI, J., 2005, “Robust multi-scale principal components analysis with applications to process monitoring”, *Journal of Process Control*, v. 15, n. 8 (dez.), pp. 869–882. ISSN: 09591524. doi: 10.1016/j.jprocont.2005.04.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152405000363>>.
- WANG, F., ZHANG, J. J., ZHENG, X., et al., 2016, “Where does AlphaGo go: from church-turing thesis to AlphaGo thesis and beyond”, *IEEE/CAA Journal of Automatica Sinica*, v. 3, n. 2 (abr.), pp. 113–120. ISSN: 2329-9266. Disponível em: <<https://ieeexplore.ieee.org/document/7471613>>.
- WANG, G., LIU, J., LI, Y., 2015a, “Fault diagnosis using kNN reconstruction on MRI variables: Fault diagnosis”, *Journal of Chemometrics*, v. 29, n. 7 (jul.), pp. 399–410. ISSN: 08869383. doi: 10.1002/cem.2719. Disponível em: <<http://doi.wiley.com/10.1002/cem.2719>>.
- WANG, J., ZHAO, Z., LIU, F., 2020a, “Robust Slow Feature Analysis for Statistical Process Monitoring”, *Industrial & Engineering Chemistry Research*, v. 59, n. 27 (jul.), pp. 12504–12513. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c01512. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.0c01512>>.
- WANG, J., HE, Q. P., 2010, “Multivariate Statistical Process Monitoring Based on Statistics Pattern Analysis”, *Industrial & Engineering Chemistry Research*, v. 49, n. 17 (set.), pp. 7858–7869. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie901911p. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie901911p>>.
- WANG, K., YAN, C., YUAN, X., et al., 2022a, “A reduced nonstationary discrete convolution kernel for multimode process monitoring”, *International Journal of Machine Learning and Cybernetics*, (ago.). ISSN: 1868-808X. doi: 10.1007/s13042-022-01621-8. Disponível em: <<https://doi.org/10.1007/s13042-022-01621-8>>.
- WANG, N., YUAN, Z., WANG, D., 2009, “Improving Process Fault Detection and Diagnosis Using Robust PCA and Robust FDA”. In: *2009 WRI*

World Congress on Computer Science and Information Engineering, pp. 54–59, Los Angeles, California USA. IEEE. ISBN: 978-0-7695-3507-4. doi: 10.1109/CSIE.2009.348. Disponível em: <<http://ieeexplore.ieee.org/document/5171300/>>.

WANG, R., BALDEA, M., EDGAR, T., 2017a, “Data Visualization and Visualization-Based Fault Detection for Chemical Processes”, *Processes*, v. 5, n. 45 (ago.). ISSN: 2227-9717. doi: 10.3390/pr5030045. Disponível em: <<http://www.mdpi.com/2227-9717/5/3/45>>.

WANG, R., EDGAR, T. F., BALDEA, M., 2017b, “A geometric framework for monitoring and fault detection for periodic processes”, *AIChe Journal*, v. 63, n. 7 (jul.), pp. 2719–2730. ISSN: 00011541. doi: 10.1002/aic.15638. Disponível em: <<http://doi.wiley.com/10.1002/aic.15638>>.

WANG, R., EDGAR, T. F., BALDEA, M., et al., 2018a, “A geometric method for batch data visualization, process monitoring and fault detection”, *Journal of Process Control*, v. 67 (jul.), pp. 197–205. ISSN: 09591524. doi: 10.1016/j.jprocont.2017.05.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152417301105>>.

WANG, R. C., EDGAR, T. F., BALDEA, M., et al., 2015b, “Process fault detection using time-explicit Kiviat diagrams”, *AIChe Journal*, v. 61, n. 12 (dez.), pp. 4277–4293. ISSN: 00011541. doi: 10.1002/aic.15054. Disponível em: <<http://doi.wiley.com/10.1002/aic.15054>>.

WANG, S., ZHAO, Q., HAN, Y., et al., 2023, “Root cause diagnosis for complex industrial process faults via spatiotemporal coalescent based time series prediction and optimized Granger causality”, *Chemometrics and Intelligent Laboratory Systems*, v. 233 (fev.), pp. 104728. ISSN: 0169-7439. doi: 10.1016/j.chemolab.2022.104728. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169743922002398>>.

WANG, X. Z., 1999, *Data Mining and Knowledge Discovery for Process Monitoring and Control*. Advances in Industrial Control. London, Springer. ISBN: 9781447111375 9781447104216. doi: 10.1007/978-1-4471-0421-6. Disponível em: <<http://link.springer.com/10.1007/978-1-4471-0421-6>>.

WANG, X. Z., MEDASANI, S., MARHOON, F., et al., 2004, “Multidimensional Visualization of Principal Component Scores for Process Historical Data Analysis”, *Industrial & Engineering Chemistry Research*, v. 43, n. 22 (out.), pp. 7036–7048. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie030816j. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie030816j>>.

- WANG, X., KRUGER, U., LENNOX, B., 2003, “Recursive partial least squares algorithms for monitoring complex industrial processes”, *Control Engineering Practice*, v. 11, n. 6 (jun.), pp. 613–632. ISSN: 09670661. doi: 10.1016/S0967-0661(02)00096-5. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066102000965>>.
- WANG, Y., PAN, Z., YUAN, X., et al., 2020b, “A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network”, *ISA Transactions*, v. 96 (jan.), pp. 457–467. ISSN: 0019-0578. doi: 10.1016/j.isatra.2019.07.001. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0019057819302903>>.
- WANG, Y., JIANG, Q., FU, J., 2017c, “Efficient Monitoring of Nonlinear Chemical Processes based on Fault-Relevant Kernel Principal Component Subspace Construction and Bayesian Inference”, *Journal of Chemical Engineering of Japan*, v. 50, n. 8, pp. 648–656. ISSN: 0021-9592, 1881-1299. doi: 10.1252/jcej.16we280. Disponível em: <https://www.jstage.jst.go.jp/article/jcej/50/8/50_16we280/_article>.
- WANG, Y., SI, Y., HUANG, B., et al., 2018b, “Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008–2017”, *The Canadian Journal of Chemical Engineering*, v. 96, n. 10, pp. 2073–2085. ISSN: 1939-019X. doi: 10.1002/cjce.23249. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cjce.23249>>.
- WANG, Y., QIN, Y., LOU, Z., et al., 2022b, “Survey on recursive statistical process monitoring methods”, *The Canadian Journal of Chemical Engineering*, v. 100, n. 9, pp. 2093–2103. ISSN: 1939-019X. doi: 10.1002/cjce.24494. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cjce.24494>>.
- WANG, Z. X., HE, Q. P., WANG, J., 2015c, “Comparison of variable selection methods for PLS-based soft sensor modeling”, *Journal of Process Control*, v. 26 (fev.), pp. 56–72. ISSN: 09591524. doi: 10.1016/j.jprocont.2015.01.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152415000050>>.
- WEBERT, H., DÖSS, T., KAUPP, L., et al., 2022, “Fault Handling in Industry 4.0: Definition, Process and Applications”, *Sensors*, v. 22, n. 6 (jan.), pp. 2205. ISSN: 1424-8220. doi: 10.3390/s22062205. Disponível em: <<https://www.mdpi.com/1424-8220/22/6/2205>>.

- WEESE, M., MARTINEZ, W., MEGAHED, F. M., et al., 2016, “Statistical Learning Methods Applied to Process Monitoring: An Overview and Perspective”, *Journal of Quality Technology*, v. 48, n. 1 (jan.), pp. 4–24. ISSN: 0022-4065, 2575-6230. doi: 10.1080/00224065.2016.11918148. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/00224065.2016.11918148>>.
- WEI, C., SONG, Z., 2020, “Generalized Semisupervised Self-Optimizing Kernel Model for Quality-Related Industrial Process Monitoring”, *IEEE Transactions on Industrial Electronics*, v. 67, n. 12 (dez.), pp. 10876–10886. ISSN: 1557-9948. doi: 10.1109/TIE.2019.2962468.
- WEI, W. W. S., 2005, *Time Series Analysis : Univariate and Multivariate Methods*. 2 edition ed. Boston, Pearson. ISBN: 978-0-321-32216-6.
- WEI, Y., WENG, Z., 2020, “Research on TE process fault diagnosis method based on DBN and Dropout”, *The Canadian Journal of Chemical Engineering*, (mar.). ISSN: 00084034. doi: 10.1002/cjce.23750. Disponível em: <<http://doi.wiley.com/10.1002/cjce.23750>>.
- WEI, Z., JI, X., ZHOU, L., et al., 2022, “A novel deep learning model based on target transformer for fault diagnosis of chemical process”, *Process Safety and Environmental Protection*, v. 167 (nov.), pp. 480–492. ISSN: 0957-5820. doi: 10.1016/j.psep.2022.09.039. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957582022008138>>.
- WENTZELL, P. D., LOHNES, M. T., 1999, “Maximum likelihood principal component analysis with correlated measurement errors: theoretical and practical considerations”, *Chemometrics and Intelligent Laboratory Systems*, v. 45, n. 1-2 (jan.), pp. 65–85. ISSN: 01697439. doi: 10.1016/S0169-7439(98)00090-2. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743998000902>>.
- WENTZELL, P. D., ANDREWS, D. T., HAMILTON, D. C., et al., 1997, “Maximum likelihood principal component analysis”, *Journal of Chemometrics*, v. 11, n. 4 (jul.), pp. 339–366. ISSN: 0886-9383, 1099-128X. doi: 10.1002/(SICI)1099-128X(199707)11:4<339::AID-CEM476>3.0.CO;2-L. Disponível em: <<http://doi.wiley.com/10.1002/%28SICI%291099-128X%28199707%2911%3A4%3C339%3A%3AAID-CEM476%3E3.0.CO%3B2-L>>.
- WERBOS, P. J., 1974, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Tese de Doutorado, Harvard University. Disponível em: <<https://www.researchgate.net/publication/35657389>>.

- WESTERHUIS, J. A., KOURTI, T., MACGREGOR, J. F., 1999, "Comparing alternative approaches for multivariate statistical analysis of batch process data", *Journal of Chemometrics*, v. 13, n. 3-4 (maio), pp. 397–413. ISSN: 0886-9383, 1099-128X. doi: 10.1002/(SICI)1099-128X(199905/08)13:3/4<397::AID-CEM559>3.0.CO;2-I. Disponível em: <<http://doi.wiley.com/10.1002/%28SICI%291099-128X%28199905/08%2913%3A3/4%3C397%3A%3AAID-CEM559%3E3.0.CO%3B2-I>>.
- WESTERHUIS, J. A., GURDEN, S. P., SMILDE, A. K., 2000, "Generalized contribution plots in multivariate statistical process monitoring", *Chemometrics and Intelligent Laboratory Systems*, v. 51, n. 1 (maio), pp. 95–114. ISSN: 01697439. doi: 10.1016/S0169-7439(00)00062-9. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743900000629>>.
- WIDODO, A., YANG, B.-S., 2007, "Support vector machine in machine condition monitoring and fault diagnosis", *Mechanical Systems and Signal Processing*, v. 21, n. 6 (ago.), pp. 2560–2574. ISSN: 08883270. doi: 10.1016/j.ymssp.2006.12.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0888327007000027>>.
- WILLIAMS, T. J., OTTO, R. E., 1960, "A generalized chemical processing model for the investigation of computer control", *Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics*, v. 79, n. 5 (nov.), pp. 458–473. ISSN: 2379-674X. doi: 10.1109/TCE.1960.6367296.
- WISE, B. M., GALLAGHER, N. B., 1996, "The process chemometrics approach to process monitoring and fault detection", *Journal of Process Control*, v. 6, n. 6 (dez.), pp. 329–348. ISSN: 0959-1524. doi: 10.1016/0959-1524(96)00009-1. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0959152496000091>>.
- WISE, B. M., GALLAGHER, N. B., WATTS BUTLER, S., et al., 1997, "Development and Benchmarking of Multivariate Statistical Process Control Tools for a Semiconductor ETCH Process: Impact of Measurement Selection and Data Treatment on Sensitivity", *IFAC Proceedings Volumes*, v. 30, n. 18 (ago.), pp. 35–42. ISSN: 14746670. doi: 10.1016/S1474-6670(17)42377-9. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017423779>>.
- WISE, B. M., GALLAGHER, N. B., BUTLER, S. W., et al., 1999, "A comparison of principal component analysis, multiway principal component analysis,

- trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process”, *Journal of Chemometrics*, v. 13, n. 3-4 (maio), pp. 379–396. ISSN: 0886-9383, 1099-128X. Disponível em: <<http://doi.wiley.com/10.1002/%28SICI%291099-128X%28199905/08%2913%3A3/4%3C379%3A%3AAID-CEM556%3E3.0.CO%3B2-N>>.
- WISE, B., VELTKAMP, D., DAVIS, B., et al., 1988, “Principal components analysis for monitoring the West Valley Liquid Fed Ceramic Melter”. In: *Waste Management '88 Proceedings*, pp. 811–818. Disponível em: <<https://www.wmsym.org/archives/1988/V2/117.pdf>>.
- WISE, B., VELTKAMP, D., RICKER, N., et al., 1991, “Application of Multivariate Statistical Process Control (MSPC) to the West Valley Slurry-Fed Ceramic Melter Process”. In: *Waste Management '91 Proceedings*, Tucson. Disponível em: <<http://www.wmsym.org/archives/1991/V2/26.pdf>>.
- WISKOTT, L., SEJNOWSKI, T. J., 2002, “Slow Feature Analysis: Unsupervised Learning of Invariances”, *Neural Computation*, v. 14, n. 4 (abr.), pp. 715–770. ISSN: 0899-7667. doi: 10.1162/089976602317318938.
- WOLD, H., 1966, “Estimation of principal components and related models by iterative least squares”, *Multivariate analysis*, pp. 391–420.
- WOLD, S., 1994, “Exponentially weighted moving principal components analysis and projections to latent structures”, *Chemometrics and Intelligent Laboratory Systems*, v. 23, n. 1 (abr.), pp. 149–161. ISSN: 0169-7439. doi: 10.1016/0169-7439(93)E0075-F. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0169743993E0075F>>.
- WOOTTON, A. J., DAY, C. R., HAYCOCK, P. W., 2015, “An Echo State Network approach to structural health monitoring”. In: *2015 International Joint Conference on Neural Networks (IJCNN)*, Killarney, Ireland, jul. IEEE. ISBN: 978-1-4799-1960-4. doi: 10.1109/IJCNN.2015.7280627. Disponível em: <<http://ieeexplore.ieee.org/document/7280627/>>.
- WU, D., ZHOU, D., ZHANG, J., et al., 2020a, “Multimode process monitoring based on fault dependent variable selection and moving window-negative log likelihood probability”, *Computers & Chemical Engineering*, v. 136 (maio), pp. 106787. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106787. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419312414>>.

- WU, D., ZHOU, D., CHEN, M., 2022, “Performance-Driven Component Selection in the Framework of PCA for Process Monitoring: A Dynamic Selection Approach”, *IEEE Transactions on Control Systems Technology*, v. 30, n. 3 (maio), pp. 1171–1185. ISSN: 1558-0865. doi: 10.1109/TCST.2021.3094512.
- WU, H., ZHAO, J., 2018, “Deep convolutional neural network model based chemical process fault diagnosis”, *Computers & Chemical Engineering*, v. 115 (jul.), pp. 185–197. ISSN: 00981354. doi: 10.1016/j.compchemeng.2018.04.009. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135418302990>>.
- WU, L., CHEN, X., PENG, Y., et al., 2012, “Fault detection and diagnosis based on sparse representation classification (SRC)”. In: *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 926–931, Guangzhou, China, dez. IEEE. ISBN: 978-1-4673-2127-3 978-1-4673-2125-9 978-1-4673-2126-6. doi: 10.1109/ROBIO.2012.6491087. Disponível em: <<http://ieeexplore.ieee.org/document/6491087/>>.
- WU, P., LOU, S., ZHANG, X., et al., 2020b, “Novel Quality Relevant Process Monitoring based on Dynamic Locally Linear Embedding Concurrent Canonical Correlation Analysis”, *Industrial & Engineering Chemistry Research*, v. 59, n. 49 (dez.), pp. 21439–21457. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c03492. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.0c03492>>.
- XAVIER, G. M., DE SEIXAS, J. M., 2018, “Fault Detection and Diagnosis in a Chemical Process using Long Short-Term Memory Recurrent Neural Network”. In: *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, jul. IEEE. ISBN: 978-1-5090-6014-6. doi: 10.1109/IJCNN.2018.8489385. Disponível em: <<https://ieeexplore.ieee.org/document/8489385/>>.
- XIA, X., PAN, X., LI, N., et al., 2022, “GAN-based anomaly detection: A review”, *Neurocomputing*, v. 493 (jul.), pp. 497–535. ISSN: 0925-2312. doi: 10.1016/j.neucom.2021.12.093. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231221019482>>.
- XIAO, H., RASUL, K., VOLLMGRAF, R., 2017, “Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms”, *arXiv:1708.07747 [cs, stat]*, (set.). Disponível em: <<http://arxiv.org/abs/1708.07747>>.
- XIAO, W., HUANG, X., HE, F., et al., 2019, “Online Robust Principal Component Analysis with Change Point Detection”, *IEEE Transactions on Multimedia*.

ISSN: 1520-9210, 1941-0077. doi: 10.1109/TMM.2019.2923097. Disponível em: <<https://ieeexplore.ieee.org/document/8736886/>>.

XIAO, Y., WANG, H., ZHANG, L., et al., 2014, “Two methods of selecting Gaussian kernel parameters for one-class SVM and their application to fault detection”, *Knowledge-Based Systems*, v. 59 (mar.), pp. 75–84. ISSN: 09507051. doi: 10.1016/j.knosys.2014.01.020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0950705114000379>>.

XIAO, Y., WANG, H., XU, W., et al., 2016, “Robust one-class SVM for fault detection”, *Chemometrics and Intelligent Laboratory Systems*, v. 151 (fev.), pp. 15–25. ISSN: 01697439. doi: 10.1016/j.chemolab.2015.11.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743915003056>>.

XIE, L., ZOU, S., XIE, Y., et al., 2021, “Sequential (Quickest) Change Detection: Classical Results and New Directions”, *IEEE Journal on Selected Areas in Information Theory*, v. 2, n. 2 (jun.), pp. 494–514. ISSN: 2641-8770. doi: 10.1109/JSAIT.2021.3072962.

XIE, Y., HUANG, J., WILLETT, R., 2013, “Change-Point Detection for High-Dimensional Time Series With Missing Data”, *IEEE Journal of Selected Topics in Signal Processing*, v. 7, n. 1 (fev.), pp. 12–27. ISSN: 1932-4553. doi: 10.1109/JSTSP.2012.2234082. Disponível em: <<https://ieeexplore.ieee.org/document/6381435>>.

XIE, Y.-L., WANG, J.-H., LIANG, Y.-Z., et al., 1993, “Robust principal component analysis by projection pursuit”, *Journal of Chemometrics*, v. 7, n. 6 (nov.), pp. 527–541. ISSN: 0886-9383, 1099-128X. doi: 10.1002/cem.1180070606. Disponível em: <<http://doi.wiley.com/10.1002/cem.1180070606>>.

XING, J., XU, J., 2022, “An Improved Convolutional Neural Network for Recognition of Incipient Faults”, *IEEE Sensors Journal*, v. 22, n. 16 (ago.), pp. 16314–16322. ISSN: 1558-1748. doi: 10.1109/JSEN.2022.3189484. Disponível em: <<https://ieeexplore.ieee.org/document/9830529/>>.

XIONG, S., ZHOU, L., DAI, Y., et al., 2022, “Attention-based LSTM fully convolutional network for chemical process fault diagnosis”, *Chinese Journal of Chemical Engineering*, (jul.). ISSN: 1004-9541. doi: 10.1016/j.cjche.2022.06.029. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1004954122002890>>.

- XU, B., ZHU, Q., 2020, “Online Quality Relevant Monitoring with Dynamic Weighted Partial Least Squares”, *Industrial & Engineering Chemistry Research*, v. 59, n. 48 (dez.), pp. 21124–21132. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.0c03901. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.0c03901>>.
- XU, C., HUANG, D., LI, D., et al., 2021a, “Novel Process Monitoring Approach Enhanced by a Complex Independent Component Analysis Algorithm with Applications for Wastewater Treatment”, *Industrial & Engineering Chemistry Research*, v. 60, n. 38 (set.), pp. 13914–13926. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.1c01990. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.1c01990>>.
- XU, C., HUANG, D., CAI, B., et al., 2023a, “A complex-valued slow independent component analysis based incipient fault detection and diagnosis method with applications to wastewater treatment processes”, *ISA Transactions*, v. 135 (abr.), pp. 213–232. ISSN: 0019-0578. doi: 10.1016/j.isatra.2022.09.019. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0019057822004773>>.
- XU, S., LU, B., BALDEA, M., et al., 2015, “Data cleaning in the process industries”, *Reviews in Chemical Engineering*, v. 31, n. 5 (jan.). ISSN: 2191-0235, 0167-8299. doi: 10.1515/revce-2015-0022. Disponível em: <<https://www.degruyter.com/view/j/revce.2015.31.issue-5/revce-2015-0022/revce-2015-0022.xml>>.
- XU, X., DING, J., 2021, “Decentralized dynamic process monitoring based on manifold regularized slow feature analysis”, *Journal of Process Control*, v. 98 (fev.), pp. 79–91. ISSN: 09591524. doi: 10.1016/j.jprocont.2020.12.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152420303437>>.
- XU, Y., DENG, X., 2016, “Fault detection of multimode non-Gaussian dynamic process using dynamic Bayesian independent component analysis”, *Neurocomputing*, v. 200 (ago.), pp. 70–79. ISSN: 09252312. doi: 10.1016/j.neucom.2016.03.015. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231216300352>>.
- XU, Y., CONG, K., ZHU, Q., et al., 2021b, “A novel AdaBoost ensemble model based on the reconstruction of local tangent space alignment and its application to multiple faults recognition”, *Journal of Process Control*, v. 104 (ago.), pp. 158–167. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2021.

07.004. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152421001050>>.

XU, Y., ZOU, X., KE, W., et al., 2023b, “A novel imbalanced fault diagnosis method based on area identification conditional generative adversarial networks”, *The Canadian Journal of Chemical Engineering*, (maio), pp. cjce.24949. ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.24949. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.24949>>.

XU, Y., JIA, M., MAO, Z., 2022a, “A novel auto-regressive dynamic slow feature analysis method for dynamic chemical process monitoring”, *Chemical Engineering Science*, v. 248 (fev.), pp. 117236. ISSN: 0009-2509. doi: 10.1016/j.ces.2021.117236. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0009250921008010>>.

XU, Y., JIA, M., MAO, Z., et al., 2022b, “A novel kernel dynamic inner slow feature analysis method for dynamic nonlinear process concurrent monitoring of operating point deviations and process dynamics anomalies”, *Journal of Process Control*, v. 110 (fev.), pp. 59–75. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2021.12.005. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152421002213>>.

XUAN, X., MURPHY, K., 2007, “Modeling changing dependency structure in multivariate time series”. In: *Proceedings of the 24th international conference on Machine learning - ICML '07*, pp. 1055–1062, Corvalis, Oregon. ACM Press. ISBN: 978-1-59593-793-3. doi: 10.1145/1273496.1273629. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1273496.1273629>>.

YAMANAKA, O., YOSHIZAWA, N., HIRAOKA, Y., et al., 2012, “A Monitoring Technique Using Multivariate Statistical Process Control Method for Performance Improvement with Application to Wastewater Treatment Plant Operation”, *IFAC Proceedings Volumes*, v. 45, n. 16 (jul.), pp. 619–624. ISSN: 14746670. doi: 10.3182/20120711-3-BE-2027.00223. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667015380204>>.

YAN, M., RICKER, N., 1997, “On-line optimization of the Tennessee Eastman challenge process”. In: *Proceedings of the 1997 American Control Conference (Cat. No.97CH36041)*, v. 5, pp. 2960–2965 vol.5, jun. doi: 10.1109/ACC.1997.612000.

YAN, S., HUANG, J., YAN, X., 2019, “Monitoring of quality-relevant and quality-irrelevant blocks with characteristic-similar variables based on self-

organizing map and kernel approaches”, *Journal of Process Control*, v. 73 (jan.), pp. 103–112. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2018.12.009. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152418305602>>.

YAN, W., WANG, J., LU, S., et al., 2023, “A Review of Real-Time Fault Diagnosis Methods for Industrial Smart Manufacturing”, *Processes*, v. 11, n. 2 (fev.), pp. 369. ISSN: 2227-9717. doi: 10.3390/pr11020369. Disponível em: <<https://www.mdpi.com/2227-9717/11/2/369>>.

YANG, C., HOU, J., 2016, “Fed-batch fermentation penicillin process fault diagnosis and detection based on support vector machine”, *Neurocomputing*, v. 190 (maio), pp. 117–123. ISSN: 09252312. doi: 10.1016/j.neucom.2016.01.027. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231216000746>>.

YANG, C., ZHOU, L., HUANG, K., et al., 2019, “Multimode process monitoring based on robust dictionary learning with application to aluminium electrolysis process”, *Neurocomputing*, v. 332 (mar.), pp. 305–319. ISSN: 09252312. doi: 10.1016/j.neucom.2018.12.024. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231218314784>>.

YANG, J., DONG, J., SHI, H., et al., 2020a, “Quality monitoring method based on enhanced canonical component analysis”, *ISA Transactions*, (jun.). ISSN: 00190578. doi: 10.1016/j.isatra.2020.06.008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0019057820302548>>.

YANG, J., BAO, W., LIU, Y., et al., 2021a, “A pairwise graph regularized constraint based on deep belief network for fault diagnosis”, *Digital Signal Processing*, v. 108 (jan.), pp. 102868. ISSN: 1051-2004. doi: 10.1016/j.dsp.2020.102868. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S105120042030213X>>.

YANG, J., WANG, L., 2023, “Nonlocal, local and global preserving stacked autoencoder based fault detection method for nonlinear process monitoring”, *Chemometrics and Intelligent Laboratory Systems*, v. 235 (abr.), pp. 104758. ISSN: 0169-7439. doi: 10.1016/j.chemolab.2023.104758. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169743923000084>>.

YANG, S., BEQUETTE, B. W., 2023, “Observational process data analytics using causal inference”, *AICHE Journal*, v. 69, n. 4 (abr.). ISSN: 0001-1541, 1547-

5905. doi: 10.1002/aic.17986. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/aic.17986>>.

YANG, S., NAVARATHNA, P., GHOSH, S., et al., 2020b, “Hybrid Modeling in the Era of Smart Manufacturing”, *Computers & Chemical Engineering*, v. 140 (set.), pp. 106874. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106874. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419313122>>.

YANG, W.-T., REIS, M. S., BORODIN, V., et al., 2022, “An interpretable unsupervised Bayesian network model for fault detection and diagnosis”, *Control Engineering Practice*, v. 127 (out.), pp. 105304. ISSN: 0967-0661. doi: 10.1016/j.conengprac.2022.105304. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0967066122001502>>.

YANG, X., LIU, W., LIU, W., et al., 2021b, “A Survey on Canonical Correlation Analysis”, *IEEE Transactions on Knowledge and Data Engineering*, v. 33, n. 6 (jun.), pp. 2349–2368. ISSN: 1041-4347, 1558-2191, 2326-3865. doi: 10.1109/TKDE.2019.2958342. Disponível em: <<https://ieeexplore.ieee.org/document/8928538/>>.

YANG, Z., GE, Z., 2020, “Monitoring and prediction of big process data with deep latent variable models and parallel computing”, *Journal of Process Control*, v. 92 (ago.), pp. 19–34. ISSN: 09591524. doi: 10.1016/j.jprocont.2020.05.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152420302171>>.

YANG, Z., GE, Z., 2022, “On Paradigm of Industrial Big Data Analytics: From Evolution to Revolution”, *IEEE Transactions on Industrial Informatics*, pp. 1–15. ISSN: 1941-0050. doi: 10.1109/TII.2022.3190394.

YANI, M., S, SI., M. B. I., S.T., M. C. S., 2019, “Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry’s Nail”, *Journal of Physics: Conference Series*, v. 1201, n. 1 (maio), pp. 012052. ISSN: 1742-6596. doi: 10.1088/1742-6596/1201/1/012052. Disponível em: <<https://doi.org/10.1088/1742-6596/1201/1/012052>>.

YAO, M., WANG, H., 2015, “On-line monitoring of batch processes using generalized additive kernel principal component analysis”, *Journal of Process Control*, v. 28 (abr.), pp. 56–72. ISSN: 09591524. doi: 10.1016/j.jprocont.2015.02.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152415000347>>.

- YAO, Y., GAO, F., 2009, “A survey on multistage/multiphase statistical modeling methods for batch processes”, *Annual Reviews in Control*, v. 33, n. 2 (dez.), pp. 172–183. ISSN: 13675788. doi: 10.1016/j.arcontrol.2009.08.001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1367578809000455>>.
- YAO, Y., BAO, J., SKYLLAS-KAZACOS, M., et al., 2018, “Fault Detection and Diagnosis In Hall–Héroult Cells Based on Individual Anode Current Measurements Using Dynamic Kernel PCA”, *Metallurgical and Materials Transactions B*, v. 49, n. 4 (ago.), pp. 2077–2088. ISSN: 1073-5615, 1543-1916. doi: 10.1007/s11663-018-1254-3. Disponível em: <<http://link.springer.com/10.1007/s11663-018-1254-3>>.
- YAO, Y., CHAKRABORTI, S., 2020, “Phase I process monitoring: The case of the balanced one way random effects model”, *Quality and Reliability Engineering International*, (nov.). ISSN: 0748-8017, 1099-1638. doi: 10.1002/qre.2793. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/qre.2793>>.
- YI, G., HERDSMAN, C., MORRIS, J., 2019, “A MATLAB toolbox for data pre-processing and multivariate statistical process control”, *Chemometrics and Intelligent Laboratory Systems*, v. 194 (nov.), pp. 103863. ISSN: 01697439. doi: 10.1016/j.chemolab.2019.103863. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743919304551>>.
- YI, H., HEHE, M., HONGBO, S., 2013, “Fault detection for chemical process based on robust PLS”. In: *2013 25th Chinese Control and Decision Conference (CCDC)*, pp. 4947–4952, maio. doi: 10.1109/CCDC.2013.6561830.
- YI, Y., ZHAO, H., HU, Z., et al., 2023, “A local–global transformer for distributed monitoring of multi-unit nonlinear processes”, *Journal of Process Control*, v. 122 (fev.), pp. 13–26. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2022.12.007. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S095915242200230X>>.
- YIN, H., 2008, “The Self-Organizing Maps: Background, Theories, Extensions and Applications”. In: Fulcher, J., Jain, L. C. (Eds.), *Computational Intelligence: A Compendium*, Studies in Computational Intelligence, Springer, pp. 715–762, Berlin, Heidelberg. ISBN: 9783540782933. doi: 10.1007/978-3-540-78293-3_17. Disponível em: <https://doi.org/10.1007/978-3-540-78293-3_17>.

- YIN, J., YAN, X., 2021, “Stacked sparse autoencoders monitoring model based on fault-related variable selection”, *Soft Computing*, (jan.). ISSN: 1432-7643, 1433-7479. doi: 10.1007/s00500-020-05384-8. Disponível em: <<http://link.springer.com/10.1007/s00500-020-05384-8>>.
- YIN, S., DING, S. X., HAGHANI, A., et al., 2012, “A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process”, *Journal of Process Control*, v. 22, n. 9 (out.), pp. 1567–1581. ISSN: 09591524. doi: 10.1016/j.jprocont.2012.06.009. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152412001503>>.
- YIN, S., DING, S. X., XIE, X., et al., 2014a, “A Review on Basic Data-Driven Approaches for Industrial Process Monitoring”, *IEEE Transactions on Industrial Electronics*, v. 61, n. 11 (nov.), pp. 6418–6428. ISSN: 0278-0046, 1557-9948. doi: 10.1109/TIE.2014.2301773. Disponível em: <<http://ieeexplore.ieee.org/document/6717991/>>.
- YIN, S., GAO, X., KARIMI, H. R., et al., 2014b, “Study on Support Vector Machine-Based Fault Detection in Tennessee Eastman Process”, v. 2014. ISSN: 1085-3375, 1687-0409. doi: 10.1155/2014/836895. Disponível em: <<http://www.hindawi.com/journals/aaa/2014/836895/>>.
- YIN, S., WANG, G., YANG, X., 2014c, “Robust PLS approach for KPI-related prediction and diagnosis against outliers and missing data”, *International Journal of Systems Science*, v. 45, n. 7 (jul.), pp. 1375–1382. ISSN: 0020-7721, 1464-5319. doi: 10.1080/00207721.2014.886136. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/00207721.2014.886136>>.
- YIN, S., LI, X., GAO, H., et al., 2015, “Data-Based Techniques Focused on Modern Industry: An Overview”, *IEEE Transactions on Industrial Electronics*, v. 62, n. 1 (jan.), pp. 657–667. ISSN: 1557-9948. doi: 10.1109/TIE.2014.2308133.
- YIN, Z., HOU, J., 2016, “Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes”, *Neurocomputing*, v. 174 (jan.), pp. 643–650. ISSN: 09252312. doi: 10.1016/j.neucom.2015.09.081. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231215014149>>.
- YING, Y., LI, Z., YANG, M., et al., 2020, “Multimode Operating Performance Visualization and Nonoptimal Cause Identification”, *Processes*, v. 8, n. 1

(jan.), pp. 123. ISSN: 2227-9717. doi: 10.3390/pr8010123. Disponível em: <<https://www.mdpi.com/2227-9717/8/1/123>>.

YONGYONG, H., XIAOQIANG, Z., 2020, “Sparse representation preserving embedding based on extreme learning machine for process monitoring”, *Transactions of the Institute of Measurement and Control*, (mar.), pp. 014233121989893. ISSN: 0142-3312, 1477-0369. doi: 10.1177/0142331219898937. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0142331219898937>>.

YOON, S., MACGREGOR, J. F., 2004, “Principal-component analysis of multiscale data for process monitoring and fault diagnosis”, *AICHE Journal*, v. 50, n. 11 (nov.), pp. 2891–2903. ISSN: 0001-1541, 1547-5905. doi: 10.1002/aic.10260. Disponível em: <<http://doi.wiley.com/10.1002/aic.10260>>.

YOON, S., MACGREGOR, J. F., 2001, “Unifying PCA and multiscale approaches to fault detection and isolation”, *IFAC Proceedings Volumes*, v. 34, n. 25 (jun.), pp. 433–438. ISSN: 14746670. doi: 10.1016/S1474-6670(17)33862-4. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667017338624>>.

YOUSEF, I., SHAH, S. L., GOPALUNI, R. B., 2022, “Visual Analytics: A New Paradigm for Process Monitoring”, *IFAC-PapersOnLine*, v. 55, n. 7 (jan.), pp. 376–383. ISSN: 2405-8963. doi: 10.1016/j.ifacol.2022.07.473. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2405896322008795>>.

YU, F., LIU, J., LIU, D., et al., 2022, “Feature learning based on entropy estimation density peak clustering and stacked autoencoder for industrial process monitoring”, *The Canadian Journal of Chemical Engineering*, (dez.), pp. cjce.24750. ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.24750. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.24750>>.

YU, H., KHAN, F., GARANIYA, V., et al., 2014, “Self-Organizing Map Based Fault Diagnosis Technique for Non-Gaussian Processes”, *Industrial & Engineering Chemistry Research*, v. 53, n. 21 (maio), pp. 8831–8843. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie500815a. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie500815a>>.

YU, J., 2016, “Process monitoring through manifold regularization-based GMM with global/local information”, *Journal of Process Control*, v. 45 (set.),

- pp. 84–99. ISSN: 09591524. doi: 10.1016/j.jprocont.2016.07.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152416300890>>.
- YU, J., YAN, X., 2019, “Active features extracted by deep belief network for process monitoring”, *ISA Transactions*, v. 84 (jan.), pp. 247–261. ISSN: 00190578. doi: 10.1016/j.isatra.2018.10.011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0019057818303902>>.
- YU, J., ZHANG, Y., 2023, “Challenges and opportunities of deep learning-based process fault detection and diagnosis: a review”, *Neural Computing and Applications*, v. 35, n. 1 (jan.), pp. 211–252. ISSN: 1433-3058. doi: 10.1007/s00521-022-08017-3. Disponível em: <<https://doi.org/10.1007/s00521-022-08017-3>>.
- YU, J., ZHENG, X., WANG, S., 2019, “A deep autoencoder feature learning method for process pattern recognition”, *Journal of Process Control*, v. 79 (jul.), pp. 1–15. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2019.05.002. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152419303002>>.
- YU, J., ZHANG, C., WANG, S., 2020, “Multichannel one-dimensional convolutional neural network-based feature learning for fault diagnosis of industrial processes”, *Neural Computing and Applications*, (jul.). ISSN: 0941-0643, 1433-3058. doi: 10.1007/s00521-020-05171-4. Disponível em: <<http://link.springer.com/10.1007/s00521-020-05171-4>>.
- YU, J., RASHID, M. M., 2013, “A novel dynamic bayesian network-based networked process monitoring approach for fault detection, propagation identification, and root cause diagnosis”, *AICHE Journal*, v. 59, n. 7 (jul.), pp. 2348–2365. ISSN: 00011541. doi: 10.1002/aic.14013. Disponível em: <<http://doi.wiley.com/10.1002/aic.14013>>.
- YU, W., WU, M., LU, C., 2021, “Meticulous process monitoring with multiscale convolutional feature extraction”, *Journal of Process Control*, v. 106 (out.), pp. 20–28. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2021.08.014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152421001396>>.
- YUAN, J., TIAN, Y., 2019, “An Intelligent Fault Diagnosis Method Using GRU Neural Network towards Sequential Data in Dynamic Processes”, *Processes*, v. 7, n. 152 (mar.). ISSN: 2227-9717. doi: 10.3390/pr7030152. Disponível em: <<https://www.mdpi.com/2227-9717/7/3/152>>.

- YUAN, P., MAO, Z., WANG, B., 2020, “A pruned support vector data description -based outlier detection method: Applied to robust process monitoring”, *Transactions of the Institute of Measurement and Control*, (mar.), pp. 014233122090595. ISSN: 0142-3312, 1477-0369. doi: 10.1177/0142331220905951. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0142331220905951>>.
- YUAN, T., QIN, S. J., 2012, “Root cause diagnosis of plant-wide oscillations using Granger causality”, *IFAC Proceedings Volumes*, v. 45, n. 15, pp. 160–165. ISSN: 14746670. doi: 10.3182/20120710-4-SG-2026.00172. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1474667016304360>>.
- ZHAN, C., LI, S., YANG, Y., 2019, “Improved process monitoring based on global-local manifold analysis and statistical local approach for industrial process”, *Journal of Process Control*, v. 75 (mar.), pp. 107–119. ISSN: 09591524. doi: 10.1016/j.jprocont.2018.12.016. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152418305717>>.
- ZHANG, C., YU, J., YE, L., 2021a, “Sparsity and manifold regularized convolutional auto-encoders-based feature learning for fault detection of multivariate processes”, *Control Engineering Practice*, v. 111 (jun.), pp. 104811. ISSN: 0967-0661. doi: 10.1016/j.conengprac.2021.104811. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0967066121000885>>.
- ZHANG, C., DONG, J., PENG, K., et al., 2022a, “Dynamic industrial process monitoring based on concurrent fast and slow-time-varying feature analytics”, *The Canadian Journal of Chemical Engineering*, v. 100, n. S1 (abr.). ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.24207. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.24207>>.
- ZHANG, C., PENG, K., DONG, J., 2021b, “A lifecycle operating performance assessment framework for hot strip mill process based on robust kernel canonical variable analysis”, *Control Engineering Practice*, v. 107 (fev.), pp. 104698. ISSN: 09670661. doi: 10.1016/j.conengprac.2020.104698. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066120302689>>.
- ZHANG, C., DONG, J., PENG, K., et al., 2022b, “A novel quality-related process monitoring method for multi-unit industrial processes under incomplete data conditions”, *The Canadian Journal of Chemical Engineering*, v. n/a, n. n/a

(jun.). ISSN: 1939-019X. doi: 10.1002/cjce.24511. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cjce.24511>>.

ZHANG, H., DENG, X., ZHANG, Y., et al., 2019, “Nonlinear Process Monitoring Based on Global Preserving Unsupervised Kernel Extreme Learning Machine”, *IEEE Access*. ISSN: 2169-3536. doi: 10.1109/ACCESS.2019.2932017. Disponível em: <<https://ieeexplore.ieee.org/document/8781877/>>.

ZHANG, H., DENG, X., ZHANG, Y., et al., 2021c, “Dynamic nonlinear batch process fault detection and identification based on two-directional dynamic kernel slow feature analysis”, *The Canadian Journal of Chemical Engineering*, v. 99, n. 1 (jan.), pp. 306–333. ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.23832. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.23832>>.

ZHANG, J., ZHAO, C., 2022, “Condition-driven probabilistic adversarial autoencoder with nonlinear Gaussian feature learning for nonstationary process monitoring”, *Journal of Process Control*, v. 117 (set.), pp. 140–156. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2022.07.012. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152422001408>>.

ZHANG, K., HAO, H., CHEN, Z., et al., 2015, “A comparison and evaluation of key performance indicator-based multivariate statistics process monitoring approaches”, *Journal of Process Control*, v. 33 (set.), pp. 112–126. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2015.06.007. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152415001341>>.

ZHANG, K., DING, S. X., SHARDT, Y. A., et al., 2017, “Assessment of T2- and Q-statistics for detecting additive and multiplicative faults in multivariate statistical process monitoring”, *Journal of the Franklin Institute*, v. 354, n. 2 (jan.), pp. 668–688. ISSN: 00160032. doi: 10.1016/j.jfranklin.2016.10.033. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S001600321630401X>>.

ZHANG, K., PENG, K., CHU, R., et al., 2018, “Implementing multivariate statistics-based process monitoring: A comparison of basic data modeling approaches”, *Neurocomputing*, v. 290 (maio), pp. 172–184. ISSN: 09252312. doi: 10.1016/j.neucom.2018.02.051. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231218301917>>.

ZHANG, L., SONG, Z., ZHANG, Q., et al., 2022c, “Generalized transformer in fault diagnosis of Tennessee Eastman process”, *Neural Computing and*

Applications, v. 34, n. 11 (jun.), pp. 8575–8585. ISSN: 1433-3058. doi: 10.1007/s00521-021-06711-2. Disponível em: <<https://doi.org/10.1007/s00521-021-06711-2>>.

ZHANG, M.-Q., LUO, X.-L., 2020, “Modified canonical variate analysis based on dynamic kernel decomposition for dynamic nonlinear process quality monitoring”, *ISA Transactions*, (ago.). ISSN: 00190578. doi: 10.1016/j.isatra.2020.08.017. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0019057820303402>>.

ZHANG, M.-Q., LUO, X.-L., 2021, “Modified canonical variate analysis based on dynamic kernel decomposition for dynamic nonlinear process quality monitoring”, *ISA Transactions*, v. 108 (fev.), pp. 106–120. ISSN: 00190578. doi: 10.1016/j.isatra.2020.08.017. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0019057820303402>>.

ZHANG, M., GE, Z., SONG, Z., et al., 2011, “Global–Local Structure Analysis Model and Its Application for Fault Detection and Identification”, *Industrial & Engineering Chemistry Research*, v. 50, n. 11 (jun.), pp. 6837–6848. ISSN: 0888-5885, 1520-5045. doi: 10.1021/ie102564d. Disponível em: <<https://pubs.acs.org/doi/10.1021/ie102564d>>.

ZHANG, S., BAO, X., 2023, “Slow-varying batch process monitoring based on canonical variate analysis”, *The Canadian Journal of Chemical Engineering*, v. 101, n. 1 (jan.), pp. 400–419. ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.24401. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.24401>>.

ZHANG, S., WANG, F., WANG, S., et al., 2014, “Multivariate process monitoring based on the distribution test of the data”. In: *Proceeding of the 11th World Congress on Intelligent Control and Automation*, pp. 2723–2728, Shenyang, China, jun. IEEE. ISBN: 978-1-4799-5825-2. doi: 10.1109/WCICA.2014.7053156. Disponível em: <<http://ieeexplore.ieee.org/document/7053156/>>.

ZHANG, S., WANG, F., ZHAO, L., et al., 2016, “A Novel Strategy of the Data Characteristics Test for Selecting a Process Monitoring Method Automatically”, *Industrial & Engineering Chemistry Research*, v. 55, n. 6 (fev.), pp. 1642–1654. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.5b03525. Disponível em: <<http://pubs.acs.org/doi/10.1021/acs.iecr.5b03525>>.

ZHANG, S., BAO, X., WANG, S., 2023a, “Common canonical variate analysis (CCVA) based modeling and monitoring for multimode processes”,

Chemical Engineering Science, v. 271 (maio), pp. 118581. ISSN: 0009-2509. doi: 10.1016/j.ces.2023.118581. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0009250923001379>>.

ZHANG, S., WEI, H.-L., DING, J., 2022d, “An effective zero-shot learning approach for intelligent fault detection using 1D CNN”, *Applied Intelligence*, (dez.). ISSN: 1573-7497. doi: 10.1007/s10489-022-04342-1. Disponível em: <<https://doi.org/10.1007/s10489-022-04342-1>>.

ZHANG, T., YAN, G., REN, M., et al., 2023b, “Dynamic transfer soft sensor for concept drift adaptation”, *Journal of Process Control*, v. 123 (mar.), pp. 50–63. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2023.01.012. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152423000203>>.

ZHANG, X., MA, L., PENG, K., et al., 2022e, “A quality-related distributed fault detection method for large-scale sequential processes”, *Control Engineering Practice*, v. 127 (out.), pp. 105308. ISSN: 0967-0661. doi: 10.1016/j.conengprac.2022.105308. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0967066122001526>>.

ZHANG, Y., DUDZIC, M. S., 2006, “Online monitoring of steel casting processes using multivariate statistical technologies: From continuous to transitional operations”, *Journal of Process Control*, v. 16, n. 8 (set.), pp. 819–829. ISSN: 09591524. doi: 10.1016/j.jprocont.2006.03.005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152406000412>>.

ZHANG, Y., MA, C., 2011, “Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS”, *Chemical Engineering Science*, v. 66, n. 1 (jan.), pp. 64–72. ISSN: 00092509. doi: 10.1016/j.ces.2010.10.008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0009250910005944>>.

ZHANG, Y., ZHOU, H., QIN, S. J., et al., 2010, “Decentralized Fault Diagnosis of Large-Scale Processes Using Multiblock Kernel Partial Least Squares”, *IEEE Transactions on Industrial Informatics*, v. 6, n. 1 (fev.), pp. 3–10. ISSN: 1551-3203, 1941-0050. doi: 10.1109/TII.2009.2033181. Disponível em: <<http://ieeexplore.ieee.org/document/5340619/>>.

ZHANG, Y., LI, X., 2020, “Two step support vector data description for dynamic, non linear, and non Gaussian processes monitoring”, *The Canadian Journal*

of Chemical Engineering, v. 98, n. 10 (out.), pp. 2109–2124. ISSN: 0008-4034, 1939-019X. doi: 10.1002/cjce.23762. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cjce.23762>>.

ZHANG, Z., ZHAO, J., 2017, “A deep belief network based fault diagnosis model for complex chemical processes”, *Computers & Chemical Engineering*, v. 107 (dez.), pp. 395–407. ISSN: 00981354. doi: 10.1016/j.compchemeng.2017.02.041. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135417301059>>.

ZHANG, Z., ZHU, J., ZHANG, S., et al., 2023c, “Process monitoring using recurrent Kalman variational auto-encoder for general complex dynamic processes”, *Engineering Applications of Artificial Intelligence*, v. 123 (ago.), pp. 106424. ISSN: 0952-1976. doi: 10.1016/j.engappai.2023.106424. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197623006085>>.

ZHAO, C., 2022, “Perspectives on nonstationary process monitoring in the era of industrial artificial intelligence”, *Journal of Process Control*, v. 116 (ago.), pp. 255–272. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2022.06.011. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152422001184>>.

ZHAO, H., LAI, Z., 2019, “Neighborhood preserving neural network for fault detection”, *Neural Networks*, v. 109 (jan.), pp. 6–18. ISSN: 08936080. doi: 10.1016/j.neunet.2018.09.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0893608018302727>>.

ZHAO, H., SUN, S., JIN, B., 2018, “Sequential Fault Diagnosis Based on LSTM Neural Network”, *IEEE Access*, v. 6, pp. 12929–12939. ISSN: 2169-3536. doi: 10.1109/ACCESS.2018.2794765. Disponível em: <<http://ieeexplore.ieee.org/document/8272354/>>.

ZHAO, S. J., ZHANG, J., XU, Y. M., 2006, “Performance monitoring of processes with multiple operating modes through multiple PLS models”, *Journal of Process Control*, v. 16, n. 7 (ago.), pp. 763–772. ISSN: 09591524. doi: 10.1016/j.jprocont.2005.12.002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152406000187>>.

ZHAO, Z., LI, T., WU, J., et al., 2020, “Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study”, *ISA Transactions*, v. 107 (dez.), pp. 224–255. ISSN: 00190578. doi: 10.1016/

j.isatra.2020.08.010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0019057820303335>>.

ZHENG, J., ZHAO, C., GAO, F., 2022a, “Retrospective comparison of several typical linear dynamic latent variable models for industrial process monitoring”, *Computers & Chemical Engineering*, v. 157 (jan.), pp. 107587. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107587. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135421003653>>.

ZHENG, J., SONG, Z., GE, Z., 2016, “Probabilistic learning of partial least squares regression model: Theory and industrial applications”, *Chemometrics and Intelligent Laboratory Systems*, v. 158 (nov.), pp. 80–90. ISSN: 01697439. doi: 10.1016/j.chemolab.2016.08.014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743916302672>>.

ZHENG, J., ZHU, J., CHEN, G., et al., 2020, “Dynamic Bayesian network for robust latent variable modeling and fault classification”, *Engineering Applications of Artificial Intelligence*, v. 89 (mar.), pp. 103475. ISSN: 09521976. doi: 10.1016/j.engappai.2020.103475. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0952197620300026>>.

ZHENG, J., LIU, Y., GE, Z., 2022b, “Dynamic ensemble selection based improved random forests for fault classification in industrial processes”, *IFAC Journal of Systems and Control*, v. 20 (jun.), pp. 100189. ISSN: 2468-6018. doi: 10.1016/j.ifacsc.2022.100189. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2468601822000049>>.

ZHENG, S., ZHAO, J., 2020, “A new unsupervised data mining method based on the stacked autoencoder for chemical process fault diagnosis”, *Computers & Chemical Engineering*, v. 135 (abr.), pp. 106755. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106755. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S009813541930986X>>.

ZHONG, K., MA, D., HAN, M., 2020, “Distributed dynamic process monitoring based on dynamic slow feature analysis with minimal redundancy maximal relevance”, *Control Engineering Practice*, v. 104 (nov.), pp. 104627. ISSN: 09670661. doi: 10.1016/j.conengprac.2020.104627. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066120301970>>.

ZHONG, L., CHANG, Y., WANG, F., et al., 2022, “Distributed Missing Values Imputation Schemes for Plant-Wide Industrial Process Using Va-

- riational Bayesian Principal Component Analysis”, *Industrial & Engineering Chemistry Research*, v. 61, n. 1 (jan.), pp. 580–593. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.1c03860. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.1c03860>>.
- ZHONG, M., XUE, T., DING, S. X., 2018, “A survey on model-based fault diagnosis for linear discrete time-varying systems”, *Neurocomputing*, v. 306 (set.), pp. 51–60. ISSN: 09252312. doi: 10.1016/j.neucom.2018.04.037. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231218304715>>.
- ZHONG, S., WEN, Q., GE, Z., 2014, “Semi-supervised Fisher discriminant analysis model for fault classification in industrial processes”, *Chemometrics and Intelligent Laboratory Systems*, v. 138 (nov.), pp. 203–211. ISSN: 01697439. doi: 10.1016/j.chemolab.2014.08.008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169743914001816>>.
- ZHOU, B., YE, H., ZHANG, H., et al., 2016a, “Process monitoring of iron-making process in a blast furnace with PCA-based methods”, *Control Engineering Practice*, v. 47 (fev.), pp. 1–14. ISSN: 09670661. doi: 10.1016/j.conengprac.2015.11.006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066115300472>>.
- ZHOU, J., HUANG, F., SHEN, W., et al., 2021, “Sub-period division strategies combined with multiway principle component analysis for fault diagnosis on sequence batch reactor of wastewater treatment process in paper mill”, *Process Safety and Environmental Protection*, v. 146 (fev.), pp. 9–19. ISSN: 09575820. doi: 10.1016/j.psep.2020.08.032. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957582020317018>>.
- ZHOU, J., ZHU, Y., 2023. “A critical look at deep neural network for dynamic system modeling”. jan. Disponível em: <<http://arxiv.org/abs/2301.11604>>.
- ZHOU, K., TONG, Y., LI, X., et al., 2023, “Exploring global attention mechanism on fault detection and diagnosis for complex engineering processes”, *Process Safety and Environmental Protection*, v. 170 (fev.), pp. 660–669. ISSN: 0957-5820. doi: 10.1016/j.psep.2022.12.055. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957582022011223>>.
- ZHOU, P., ZHANG, R., LIANG, M., et al., 2020, “Fault identification for quality monitoring of molten iron in blast furnace ironmaking based on KPLS with improved contribution rate”, *Control Engineering Practice*, v. 97 (abr.),

pp. 104354. ISSN: 09670661. doi: 10.1016/j.conengprac.2020.104354. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0967066120300332>>.

ZHOU, Z., WEN, C., YANG, C., 2016b, “Fault Isolation Based On k-Nearest Neighbor Rule For Industrial Processes”, *IEEE Transactions on Industrial Electronics*. ISSN: 0278-0046, 1557-9948. doi: 10.1109/TIE.2016.2520898. Disponível em: <<http://ieeexplore.ieee.org/document/7390039/>>.

ZHU, J., SHI, H., SONG, B., et al., 2021a, “Nonlinear process monitoring based on load weighted denoising autoencoder”, *Measurement*, v. 171 (fev.), pp. 108782. ISSN: 02632241. doi: 10.1016/j.measurement.2020.108782. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S026322412031280X>>.

ZHU, J., GE, Z., SONG, Z., et al., 2018, “Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data”, *Annual Reviews in Control*, v. 46, pp. 107–133. ISSN: 13675788. doi: 10.1016/j.arcontrol.2018.09.003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1367578818301056>>.

ZHU, J., YAO, Y., GAO, F., 2020a, “Multiphase two-dimensional time-slice dynamic system for batch process monitoring”, *Journal of Process Control*, v. 85 (jan.), pp. 184–198. ISSN: 09591524. doi: 10.1016/j.jprocont.2019.12.004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152419302197>>.

ZHU, J., JIANG, M., LIU, Z., 2021b, “Fault Detection and Diagnosis in Industrial Processes with Variational Autoencoder: A Comprehensive Study”, *Sensors*, v. 22, n. 1 (dez.), pp. 227. ISSN: 1424-8220. doi: 10.3390/s22010227. Disponível em: <<https://www.mdpi.com/1424-8220/22/1/227>>.

ZHU, Q., 2020, “Auto-regressive modeling with dynamic weighted canonical correlation analysis”, *Journal of Process Control*, v. 95 (nov.), pp. 32–44. ISSN: 09591524. doi: 10.1016/j.jprocont.2020.09.007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959152420302924>>.

ZHU, Q., JOE QIN, S., DONG, Y., 2020b, “Dynamic Latent Variable Regression for Inferential Sensor Modeling and Monitoring”, *Computers & Chemical Engineering*, (mar.), pp. 106809. ISSN: 00981354. doi: 10.1016/j.compchemeng.2020.106809. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0098135419312323>>.

- ZHU, Q.-X., MENG, Q.-Q., HE, Y.-L., 2017, “Novel Multidimensional Feature Pattern Classification Method and Its Application to Fault Diagnosis”, *Industrial & Engineering Chemistry Research*, v. 56, n. 31 (ago.), pp. 8906–8916. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.7b00027. Disponível em: <<http://pubs.acs.org/doi/10.1021/acs.iecr.7b00027>>.
- ZHU, Q.-X., WANG, X.-W., LI, K., et al., 2022, “Enhanced multicorrelation block process monitoring and abnormality root cause analysis for distributed industrial process: A visual data-driven approach”, *Journal of Process Control*, v. 118 (out.), pp. 1–15. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2022.08.008. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152422001500>>.
- ZHU, S.-C., GUO, C.-E., WANG, Y., et al., 2005, “What are Textons?” *International Journal of Computer Vision*, v. 62, n. 1 (abr.), pp. 121–143. ISSN: 1573-1405. doi: 10.1023/B:VISI.0000046592.70770.61. Disponível em: <<https://doi.org/10.1023/B:VISI.0000046592.70770.61>>.
- ZHUO, Y., GE, Z., 2020, “Gaussian Discriminative Analysis aided GAN for imbalanced big data augmentation and fault classification”, *Journal of Process Control*, v. 92 (ago.), pp. 271–287. ISSN: 09591524. doi: 10.1016/j.jprocont.2020.06.014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S095915242030250X>>.
- ZIAEI-HALIMEJANI, H., KOUZBOUR, S., ZARGHAMI, R., et al., 2021a, “Monitoring of the bubble columns hydrodynamics by recurrence quantification data analysis”, *Chemical Engineering Research and Design*, v. 171 (jul.), pp. 100–110. ISSN: 0263-8762. doi: 10.1016/j.cherd.2021.05.002. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263876221001957>>.
- ZIAEI-HALIMEJANI, H., NAZEMZADEH, N., ZARGHAMI, R., et al., 2021b, “Fault Diagnosis of Chemical Processes based on Joint Recurrence Quantification Analysis”, *Computers & Chemical Engineering*, (set.), pp. 107549. ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2021.107549. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135421003276>>.
- ZIAEI-HALIMEJANI, H., ZARGHAMI, R., MANSOURI, S. S., et al., 2021c, “Data-Driven Fault Diagnosis of Chemical Processes Based on Recurrence Plots”, *Industrial & Engineering Chemistry Research*, v. 60, n. 7 (fev.),

pp. 3038–3055. ISSN: 0888-5885. doi: 10.1021/acs.iecr.0c06307. Disponível em: <<https://doi.org/10.1021/acs.iecr.0c06307>>.

ZIAEI-HALIMEJANI, H., ZARGHAMI, R., MOSTOUFI, N., 2021d, “Joint recurrence based root cause analysis of nonlinear multivariate chemical processes”, *Journal of Process Control*, v. 103 (jul.), pp. 19–33. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2021.05.008. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152421000779>>.

ZLIOBAITE, I., PECHENIZKIY, M., GAMA, J., 2016, “An Overview of Concept Drift Applications”. In: Japkowicz, N., Stefanowski, J. (Eds.), *Big Data Analysis: New Algorithms for a New Society*, v. 16, Springer International Publishing, pp. 91–114, Cham. ISBN: 978-3-319-26987-0 978-3-319-26989-4. doi: 10.1007/978-3-319-26989-4_4. Disponível em: <http://link.springer.com/10.1007/978-3-319-26989-4_4>.

ZOMER, S., ZHANG, J., TALWAR, S., et al., 2018, “Multivariate monitoring for the industrialisation of a continuous wet granulation tabletting process”, *International Journal of Pharmaceutics*, v. 547, n. 1 (ago.), pp. 506–519. ISSN: 0378-5173. doi: 10.1016/j.ijpharm.2018.06.034. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378517318304320>>.

ZOPE, K., SINGHAL, T., NISTALA, S. H., et al., 2023, “Transfer Entropy-Based Automated Fault Traversal and Root Cause Identification in Complex Nonlinear Industrial Processes”, *Industrial & Engineering Chemistry Research*, v. 62, n. 9 (mar.), pp. 4002–4018. ISSN: 0888-5885, 1520-5045. doi: 10.1021/acs.iecr.2c03570. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.iecr.2c03570>>.

ZOU, W., XIA, Y., LI, H., 2018, “Fault Diagnosis of Tennessee-Eastman Process Using Orthogonal Incremental Extreme Learning Machine Based on Driving Amount”, *IEEE Transactions on Cybernetics*, v. 48, n. 12 (dez.), pp. 3403–3410. ISSN: 2168-2267, 2168-2275. doi: 10.1109/TCYB.2018.2830338. Disponível em: <<https://ieeexplore.ieee.org/document/8361786/>>.

Apêndice A

Trabalhos-chave utilizados na revisão da literatura

Tabela A.1: Artigos de revisão e que apresentam visão global sobre técnicas de monitoramento de processos baseado em dados.

Referência	Descrição
MACGREGOR e KOURTI (1995)	Ressalta a diferença entre controle de qualidade (extensões multivariadas das cartas de Shewhart) e de processo (métodos de variáveis latentes, PCA e PLS).
WISE e GALLAGHER (1996)	Trata de variáveis latentes (PCA e PLS), incluindo uma discussão mais aprofundada sobre regressão, com alguns métodos não comumente citados nos demais trabalhos analisados, como regressão rígida.
KOURTI (2002)	Trata de variáveis latentes (PCA e PLS), incluindo uma discussão sobre aspectos práticos, como características das bases de dados, frequência de amostragem, entre outros.

Continua na próxima página

Tabela A.1 – *Continuada da página anterior*

Referência	Descrição
QIN (2003)	Trata de variáveis latentes (em particular, o PCA); foco na comparação entre contribuição e reconstrução e em estudos de detecabilidade, reconstrutibilidade e identificabilidade.
VENKATASUBRAMANIAN <i>et al.</i> (2003a)	Trata de variáveis latentes (PCA e PLS), análise qualitativa de tendências, classificadores estatísticos e redes neurais. Compara essas técnicas com métodos baseados em modelos (observadores, filtros de Kalman, paridade, além de métodos baseados em modelos qualitativos), apresentados nas duas partes anteriores da série de artigos.
MILETIC <i>et al.</i> (2004)	Trata de variáveis latentes (PCA e PLS), tendo como foco aplicações industriais. Apresenta também uma metodologia para o desenvolvimento de sistemas industriais de monitoramento úteis e de longo prazo.
GANESAN <i>et al.</i> (2004)	Revisa estratégias para monitoramento de processos usando <i>wavelets</i> .
KOURTI (2005b)	Trata de variáveis latentes (PCA e PLS), destacando algumas questões características de aplicações reais, como cálculos em linha, aplicações em controle e arquivamento, compressão e reconstrução dos dados.

Continua na próxima página

Tabela A.1 – *Continuada da página anterior*

Referência	Descrição
QIN (2009)	Tem o escopo parecido com a revisão anterior (QIN, 2003) do mesmo autor.
YAO e GAO (2009)	Revisa estratégias para monitoramento de processos em batelada e multiestágio.
KADLEC <i>et al.</i> (2009)	Revisa técnicas e aplicações de sensores virtuais baseados em dados.
KADLEC <i>et al.</i> (2011)	Revisa mecanismos adaptativos para a construção de sensores virtuais baseados em dados.
DING <i>et al.</i> (2011)	Trata de variáveis latentes (PCA e PLS) e compara essas técnicas com métodos baseados em modelos (filtro de Kalman, observadores e paridade).
SLISKOVIC <i>et al.</i> (2011)	No contexto de modelos para aplicação em sensores virtuais, divide os métodos em “métodos estatísticos multivariados”, “redes neurais artificiais”, “máquinas de vetores suporte”, “métodos híbridos” e “métodos adaptativos”.
DAS <i>et al.</i> (2012)	Divide os métodos “baseados em dados” em duas classes: “baseados na estatística” (que incluem os métodos univariados e os de variável latente, quando multivariados) e “baseados na inteligência artificial” (que incluem as redes neurais e a lógica <i>fuzzy</i>). Compara essas técnicas com métodos “baseados em conhecimentos <i>a priori</i> ” (estimação de parâmetros, observadores e paridade).

Continua na próxima página

Tabela A.1 – *Continuada da página anterior*

Referência	Descrição
QIN (2012)	Tem o escopo parecido com as revisões anteriores do mesmo autor (QIN, 2009, 2003), com a adição do modelo PLS.
MACGREGOR e CINAR (2012)	Fornece uma visão global sobre a aplicação de métodos de variáveis latentes em problemas de monitoramento, controle tolerante a falhas e otimização.
GE <i>et al.</i> (2013)	Faz uma classificação dos métodos por processo, categorizando em processos contínuos, em bate-lada, não gaussianos, não lineares, variantes no tempo/multimodais e dinâmicos. Inclui uma vasta gama de metodologias não citadas em revisões anteriores.
DAI e GAO (2013)	Faz uma revisão de métodos de detecção e diagnóstico de falhas sob a perspectiva do processamento de dados.
KHATIBISEPEHR <i>et al.</i> (2013)	Revisa o projeto de sensores virtuais utilizando métodos bayesianos.
YIN <i>et al.</i> (2014a)	Trata de variáveis latentes (PCA e PLS, apesar de citar alguns outros modelos), com foco em modificações induzidas por condições de operação industrial, em particular relacionadas à incerteza e ao comportamento dinâmico.
YIN <i>et al.</i> (2015)	Revisa técnicas baseadas em dados aplicadas ao controle e monitoramento industriais.
SEVERSON <i>et al.</i> (2016)	Foca em variáveis latentes. Cita brevemente algumas outras abordagens, em especial a SVM.

Continua na próxima página

Tabela A.1 – *Continuada da página anterior*

Referência	Descrição
TIDRIRI <i>et al.</i> (2016)	Compara métodos baseados em dados e em modelos, com foco nas tentativas de se criar metodologias híbridas. Separa os métodos baseados em dados em redes bayesianas, redes neurais, cartas de controle univariadas e modelagem em variáveis latentes (citando apenas PCA e PLS).
WEESE <i>et al.</i> (2016)	Apresenta uma revisão sobre a aplicação de métodos de aprendizado de máquina no monitoramento de processos.
SHU <i>et al.</i> (2016)	Apresenta desafios e perspectivas para a área de gerenciamento de situações anormais sob a ótica <i>big data</i> .
SOUZA <i>et al.</i> (2016)	Revisa métodos de sensores virtuais para aplicações em problemas de regressão.
GE (2017)	Revisa estratégias adotadas para monitoramento de toda a planta industrial.
REIS e GINS (2017)	Discorre sobre perspectivas e tendência histórica das aplicações de monitoramento, da detecção para o diagnóstico e para o prognóstico.
ALAUDDIN <i>et al.</i> (2018)	Análise bibliométrica da área de monitoramento baseado em dados.
WANG <i>et al.</i> (2018b)	Apresenta uma revisão sistemática da área de estatística multivariada aplicada a monitoramento de processos entre 2008 e 2017, com foco em trabalhos chineses.

Continua na próxima página

Tabela A.1 – *Continuada da página anterior*

Referência	Descrição
PERES e FOGLIATTO (2018)	Revisa estratégias adotadas para seleção de variáveis.
HE e WANG (2018)	Apresenta perspectivas para o monitoramento no cenário <i>big data</i> .
QUIÑONES-GRUEIRO <i>et al.</i> (2019)	Revisa estratégias adotadas para processos multimodais.
ALDRICH (2019)	Revisa estratégias adotadas para processos dinâmicos.
JIANG <i>et al.</i> (2019)	Revisa estratégias adotadas para monitoramento de toda a planta industrial.
REIS (2019)	Revisa estratégias adotadas para monitoramento multiescala.
RENDALL <i>et al.</i> (2019)	Revisa estratégias para monitoramento de processos em batelada, mapeando-as em uma escala de complexidade de modelagem e implementação.
MD NOR <i>et al.</i> (2019)	Divide as abordagens em “análise estatística multivariada” e “aprendizado de máquina”, focando nas diferenças entre as aplicações dos métodos a depender do tipo de fábrica analisada.
PILARIO <i>et al.</i> (2019)	Revisa aplicações da técnica <i>kernel</i> para extração de <i>features</i> em monitoramento de processos.
DIEZ-OLIVAN <i>et al.</i> (2019)	Revisa aplicações baseadas em fusão de dados e aprendizado de máquina no âmbito de atividades de prognóstico.
GOPALUNI <i>et al.</i> (2020)	Revisa as aplicações de técnicas de aprendizado de máquina para o monitoramento e controle de processos.

Continua na próxima página

Tabela A.1 – *Continuada da página anterior*

Referência	Descrição
APSEMIDIS <i>et al.</i> (2020)	Revisa aplicações da técnica <i>kernel</i> em monitoramento de processos.
PARK <i>et al.</i> (2020b)	Categoriza os métodos de acordo com o tipo de sistema em que são aplicados: dinâmicos, não lineares, não gaussianos, multimodais/variantes no tempo e não estacionários.
QIN <i>et al.</i> (2020)	Revisa estratégias de modelagem em variáveis latentes no contexto de processos dinâmicos, propondo uma nova estrutura matemática para unificar as abordagens já existentes.
COHEN e ATOUI (2020)	Revisa estratégias para monitoramento de processos usando <i>wavelets</i> .
AHMED <i>et al.</i> (2020)	Análise bibliométrica no âmbito de técnicas classificadas como “baseadas em Inteligência Artificial”, o que segundo o critério dos autores exclui modelos de variáveis latentes como PCA e PLS/CCA.
JIANG <i>et al.</i> (2020)	Apresenta uma revisão sobre sensores virtuais aplicados, dentre outros problemas, ao monitoramento de processos.
LIU e XIE (2020)	Revisa a aplicação de técnicas <i>kernel</i> para a construção de sensores virtuais baseados em dados.
CURRERI <i>et al.</i> (2020)	Revisa métodos de seleção de variáveis para construção de sensores virtuais.

Continua na próxima página

Tabela A.1 – *Continuada da página anterior*

Referência	Descrição
JIAO <i>et al.</i> (2020)	Revisa a aplicação de métodos de aprendizado de máquina para garantia da saúde e segurança de processos.
ARUNTHAVANATHAN <i>et al.</i> (2021)	Apresenta uma análise global dos três aspectos considerados pelos autores essenciais para a segurança de processos: detecção e diagnóstico de falhas, avaliação de risco e gerenciamento de situações anormais.
TAQVI <i>et al.</i> (2021)	Revisa vários métodos focando nas categorizações supervisionados/-não supervisionados.
OKADA <i>et al.</i> (2021)	Divide as abordagens em análise de sinais, baseadas em modelos, baseadas em dados e híbridas.
SUN e GE (2021)	Revisa métodos de aprendizado profundo para aplicações em sensores virtuais.
PANI (2022)	Revisa extensões do modelo PCA utilizando a técnica <i>kernel</i> para monitoramento não linear.
ZHAO (2022)	Revisa métodos para monitoramento não estacionário e apresenta perspectivas para a área.
WANG <i>et al.</i> (2022b)	Revisa técnicas de natureza recursiva.
QIAN <i>et al.</i> (2022)	Revisa a aplicação de autocodificadores para o aprendizado de representação.
BI <i>et al.</i> (2022)	Apresenta uma revisão abrangente, além de perspectivas e desafios para o desenvolvimento futuro de técnicas de detecção e diagnóstico de falhas.

Continua na próxima página

Tabela A.1 – *Continuada da página anterior*

Referência	Descrição
NAWAZ <i>et al.</i> (2022)	Aplica técnicas bibliométricas para revisar especificamente estratégias adotadas para monitoramento multiescala.
KONG <i>et al.</i> (2022)	Revisa a aplicação de métodos baseados em variáveis latentes sob a perspectiva <i>big data</i> .
WEBERT <i>et al.</i> (2022)	Revisa estratégias para tratamento de falhas no contexto das atividades de detecção, classificação e priorização.
MELO <i>et al.</i> (2022)	Revisa <i>benchmarks</i> abertos para testes de técnicas de monitoramento de processos.
PERERA <i>et al.</i> (2023)	Apresenta uma revisão crítica do uso industrial de sensores virtuais baseados em inteligência artificial.
YAN <i>et al.</i> (2023)	Revisa os métodos para diagnóstico de falhas em tempo real aplicados ao monitoramento de equipamentos e processos.
PALLA e PANI (2023)	Revisa a aplicação da técnica ICA a problemas de monitoramento.
YU e ZHANG (2023)	Revisa a aplicação de técnicas de aprendizado profundo a problemas de monitoramento.
LIU <i>et al.</i> (2023b)	Apresenta uma revisão e perspectivas acerca da área de gerenciamento de situações anormais em processos químicos.

Tabela A.2: Artigos de comparação de métodos de monitoramento baseado em dados.

Referência	Descrição
WESTERHUIS <i>et al.</i> (1999)	Compara diversas abordagens para monitoramento de processos em batelada.
TIEN <i>et al.</i> (2004)	Compara diversas abordagens do PCA (convencional, MPCPA, APCA e EWPCA).
ALCALA e QIN (2011)	Analisa cinco métodos comumente usados para diagnóstico e propõe sua unificação em três métodos gerais.
YIN <i>et al.</i> (2012)	Compara as aplicações de diversos métodos (PCA, PLS, ICA, FDA, SAP).
ZHANG <i>et al.</i> (2015)	Compara os desempenhos de métodos de monitoramento baseados em KPI (<i>key performance indicators</i>).
JING e HOU (2015)	Compara as aplicações dos métodos PCA e SVM.
LI e QIN (2016)	Compara diversos esquemas de monitoramento para processos com distribuições não gaussianas, incluindo ICA, SPA, KDE e SVDD.
ASKARIAN <i>et al.</i> (2016)	Compara diferentes abordagens para lidar com dados faltantes e aos desempenhos de diferentes esquemas de diagnóstico.
RATO <i>et al.</i> (2016)	Compara diferentes abordagens do PCA para lidar com processos de grande dimensão e dependentes do tempo.
ZHANG <i>et al.</i> (2017)	Compara o desempenho das estatísticas T^2 e Q para detecção de falhas aditivas e multiplicativas.

Continua na próxima página

Tabela A.2 – *Continuada da página anterior*

Referência	Descrição
FUENTES-GARCÍA <i>et al.</i> (2018)	Propõe uma metodologia para comparação experimental entre diversos esquemas de diagnóstico, no contexto da aplicação do PCA.
ZHANG <i>et al.</i> (2018)	Compara a aplicação do PCA, CCA e PLS, bem como suas extensões dinâmicas.
VIDAL-PUIG <i>et al.</i> (2019)	Compara esquemas de diagnóstico aplicados a modelos de variáveis latentes (PCA e PLS).
SHRIVASTAVA (2021)	Compara diferentes estratégias para agregação de modelos baseados em árvores de decisão.
LOMOV <i>et al.</i> (2021)	Compara as aplicações de diferentes arquiteturas de aprendizado profundo.
FERNANDES <i>et al.</i> (2022b)	Compara modelos dinâmicos para monitoramento baseado em variáveis latentes.
ZHENG <i>et al.</i> (2022a)	Compara modelos dinâmicos lineares para monitoramento baseado em variáveis latentes.

Tabela A.3: Livros sobre monitoramento de processos baseado em dados.

Referência	Descrição
Fault Detection and Diagnosis in Chemical and Petrochemical Processes (HIMMELBLAU, 1978)	Descreve aplicações baseadas em dados e em modelos, incluindo cartas de controle, reconhecimento de padrões e estimativa de parâmetros.

Continua na próxima página

Tabela A.3 – *Continuada da página anterior*

Referência	Descrição
Data Mining and Knowledge Discovery for Process Monitoring and Control (WANG, 1999)	Aborda tópicos como processamento de sinais, extração de variáveis, identificação de estados operacionais, dentre outros.
Fault Detection and Diagnosis in Industrial Systems (CHIANG <i>et al.</i> , 2001)	Descreve métodos de variáveis latentes (PCA, FDA, PLS e CVA) e aplica no modelo Tennessee Eastman. Nos últimos capítulos, dedica-se a métodos “analíticos” e “baseados no conhecimento”.
Multivariate Statistical Process Control with Industrial Applications (MASON e YOUNG, 2002)	Descreve várias formas de aplicação da estatística T^2 na área de monitoramento.
Statistical Monitoring of Complex Multivariate Processes: With Applications in Industrial Process Control (KRUGER e XIE, 2012)	Descreve os métodos PCA e PLS em detalhes, desde suas versões mais básicas até avanços recentes.
Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods (ALDRICH e AURET, 2013)	É o primeiro livro a apresentar o problema de monitoramento do ponto de vista da área de aprendizado de máquina, tratando de redes neurais, teoria do aprendizado estatístico, métodos de kernel, métodos de árvore, dentre outros.
Multivariate Statistical Process Control: Process Monitoring Methods and Applications (GE e SONG, 2013b)	De maneira semelhante à revisão publicada no mesmo ano por dois dos mesmos autores (GE <i>et al.</i> , 2013), o livro faz uma classificação e apresentação dos métodos por processo, categorizando em processos contínuos, em batelada, não gaussianos, não lineares, variantes no tempo/multimodais e dinâmicos.

Continua na próxima página

Tabela A.3 – *Continuada da página anterior*

Referência	Descrição
Data-driven Design of Fault Diagnosis and Fault-tolerant Control Systems (DING, 2014)	Trata de três temas: (i) métodos básicos de análise multivariada aplicados ao problema de monitoramento; (ii) projeto orientado por dados de modelos de dados de entrada/saída e esquemas de detecção e isolamento de falhas baseados em observadores; (iii) projeto orientado por dados de sistemas de controle tolerantes a falhas no contexto do gerenciamento do tempo de vida de sistemas de controle.
Chemometric Monitoring: Product Quality Assessment, Process Fault Detection, and Applications (KUNDU <i>et al.</i> , 2017)	Focado no contexto da quimiometria, apresenta variadas técnicas (PCA, PLS, SVM, <i>clustering</i> , etc.) e aborda alguns itens não comentados nos demais livros (geração e pré-processamento dos dados, etc.), ilustrando com vários exemplos de aplicações reais.
Advanced methods for fault diagnosis and fault-tolerant control (DING, 2021)	Trata de técnicas baseadas em dados e em modelos, com ênfase na aplicação de métodos de aprendizado de máquina e no problema do controle tolerante a falhas.