

Nayron Morais Almeida

**Proposta de *Framework* para  
Detecção de Eventos em Processos  
Industriais e Diagnóstico de Causa  
Raiz: Abordagens Baseadas em  
Aprendizado on-line**

Belo Horizonte - Minas Gerais

Novembro, 2020



Nayron Morais Almeida

**Proposta de *Framework* para Detecção de Eventos  
em Processos Industriais e Diagnóstico de Causa  
Raiz: Abordagens Baseadas em Aprendizado on-line**

Dissertação de mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Mestre em Engenharia Elétrica.

Universidade Federal de Minas Gerais - UFMG

Programa de Pós-Graduação em Engenharia Elétrica - PPGEE

Orientador: Walmir Matos Caminhas

Belo Horizonte - Minas Gerais

Novembro, 2020

A447p

Almeida, Nayron Morais.

Proposta de framework para detecção de eventos em processos industriais e diagnóstico de causa raiz [recurso eletrônico] : abordagens baseadas em aprendizado on-line / Nayron Morais Almeida. – 2020.  
1 recurso online (120 f. : il., color.) : pdf.

Orientador: Walmir Matos Caminhas.

Dissertação (mestrado) Universidade Federal de Minas Gerais, Escola de Engenharia.

Bibliografia: f. 115-120.

Exigências do sistema: Adobe Acrobat Reader.

1. Engenharia elétrica - Teses. 2. Framework (Programa de computador) - Teses. 3. Controle de processo - Teses. 4. Indústria - Teses. I. Caminhas, Walmir Matos. II. Universidade Federal de Minas Gerais. Escola de Engenharia.  
III. Título.

CDU: 621.3(043)

**"Proposta de Framework para Detecção de Eventos em  
Processos Industriais e Diagnóstico de Causa Raiz:  
Abordagens Baseadas em Aprendizado On-line"**

**NAYRON MORAIS ALMEIDA**

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Mestre em Engenharia Elétrica.

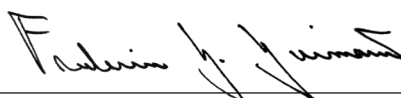
Aprovada em 12 de novembro de 2020.

Por:



---

**Prof. Dr. Walmir Matos Caminhas  
DELT (UFMG)**



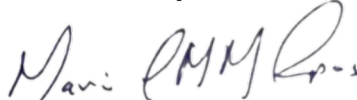
---

**Prof. Dr. Frederico Gadelha Guimarães  
DEE (UFMG)**



---

**Dr. Sergio Gregorio de Oliveira  
CENPES (Petrobras)**



---

**Dr. Mario Cesar Mello Massa de Campos  
CENPES (Petrobras)**

*Aos meus pais, Rita e José, que enfrentaram e superaram inúmeras dificuldades pelas  
quais me permitiram chegar a este ponto.*

# Agradecimentos

Agradeço à minha família pelo amor e apoio incondicional, são meu bem maior.

Agradeço ao meu orientador professor Walmir Matos Caminhas por ter sempre uma atitude compreensiva, pelas oportunidades e orientações.

As membros do grupo do projeto Petrobras pelas trocas de experiências e críticas construtivas que me permitiram evoluir consideravelmente, e que além de companheiros de trabalho tornaram-se grandes amigos: professor Bomfim, Murilo, Denis, Jefferson, professor Reinaldo, professor Benjamim, professor André Paim e novamente, professor Walmir.

Aos amigos da sala de mestrado com quem convivi a maior parte do tempo na Universidade e criei significativos laços de amizade: Lucas 1 e 2, Luíza, Hélio, Higor e Rafael.

E a todos que direta ou indiretamente me ajudaram a subir mais este degrau, sem a ajuda dada não teria conseguido dá mais este passo, obrigado.





*"Why is it that nobody understands me and everybody likes me?"*

Albert Einstein



# Resumo

Devido ao aumento da complexidade das plantas industriais e o alto índice de digitalização das mesmas, a técnica clássica de monitoramento baseada em lógica de limiares tem perdido espaço. Tal técnica aborda o problema de forma univariável, que em geral acarreta em um alto índice de geração de alarme. Em substituição as anteriores, estão sendo empregadas técnicas baseadas em aprendizado de máquina, utilizando ferramentas como redes neurais artificiais, sistemas *fuzzy* e métodos estatísticos. Com as mesmas é possível tratar a detecção de eventos de forma multivariável, sob a ótica de classificação de padrões, sendo o tema explorado no presente trabalho. Visando o desenvolvimento de uma ferramenta que possa ser usada em plantas industriais modernas, sem base de dados rotuladas com eventos ocorridos, ou não utilizar informação *a priori*, foi desenvolvida uma abordagem baseada em sistemas evolutivos (*Evolving Systems*). Em tais sistemas, a base de conhecimento sobre o comportamento do processo é construída a partir de aprendizado incremental e on-line. De forma que, além da adaptação de parâmetros, também é realizada a adaptação da estrutura dos modelos. Com base nisso, este trabalho propõe um *framework* portátil e escalável para uso no monitoramento de processos industriais. A estrutura proposta é totalmente baseada em métodos evolutivos, permitindo que qualquer método seja incorporado, independente de sua abordagem. Além disso, visando auxiliar o supervisor no entendimento do comportamento do processo por meio do método evolutivo, um método de causa raiz é utilizado para justificar comportamentos do processo que foram detectados. Para demonstrar seu uso, são implementados e incorporados quatro métodos propostos na literatura com fundamentações distintas. Experimentos são executados com um *benchmark* e com um sistema real de selagem de compressor de gases. Os resultados obtidos indicam que o *framework* é uma ferramenta promissora para uso no monitoramento de processos.

Palavras-chave: Evolutivo, Ferramenta, Processos, Monitoramento, Indústria.



# Abstract

Due to the increased complexity of industrial plants and their high digitization, the classic monitoring technique based on threshold logic has lost use. That technique addresses the problem in a univariable way, which in general leads to a high rate of alarm generation. In place of the previous ones, techniques based on machine learning are being used, using approaches such as artificial neural networks, fuzzy systems and statistical methods. With them, it is possible to handle the detection of events in a multivariable way, from the perspective of classification of patterns, being the theme explored in the present work. In order to develop a tool that can be used in modern industrial plants, without a database labeled with events that have occurred, or not to use information a priori, an approach based on evolving systems was developed. In such systems, the knowledge base about the process behavior is built from incremental and online learning. That way, in addition to adapting parameters, it carries out the adapt the models's structure. Based on this, this work proposes a portable and scalable framework for use in monitoring industrial processes. The proposed structure is based entirely on evolving methods, allowing any method to be incorporated, regardless of its approach. In addition, in order to assist the supervisor in understanding the behavior of the process through the evolving method, a method of root cause is used to justify behaviors of the process that were detected. To demonstrate its use, four methods proposed in the literature with different grounds are implemented and incorporated. Experiments are performed with a benchmark and with a real gas compressor's seal system. The results obtained indicate that the framework is a promising tool for use in monitoring processes.

*Keywords: Evolving, Tool, Processes, Monitoring, Industry.*



# Lista de Ilustrações

Figura 2.1 – Esquema geral do EMPF. . . . .	31
Figura 2.2 – Base de dados tridimensional sintética que simula um processo real. . .	33
Figura 2.3 – Resultado do monitoramento com EMPF para o período de criação do segundo grupo. . . . .	34
Figura 2.4 – Resultado do monitoramento com EMPF para o período de retorno do Grupo 2 para o Grupo 1. . . . .	35
Figura 2.5 – Resultado do monitoramento com EMPF para o período de criação do terceiro grupo. . . . .	35
Figura 2.6 – Resultado do monitoramento com EMPF para o período de retorno do Grupo 3 para o Grupo 1. . . . .	36
Figura 2.7 – Resultado do monitoramento com EMPF para o período de segunda ocorrência do primeiro evento. . . . .	37
Figura 2.8 – Resultado do monitoramento com EMPF para o período de segunda ocorrência do segundo evento. . . . .	37
Figura 3.1 – Processo genérico monitorado pelas leituras do sensor de temperatura e pressão com ocorrência de falha de alta temperatura e alta pressão (a) e estados contidos na matriz memória para representação da operação normal (b). . . . .	42
Figura 3.2 – Procedimento de inferência realizado pela SBM. Se o estado atual e estimativa não estiverem dentro da região delimitada por $\tau$ considera-se que esse estado representa um comportamento anômalo. . . . .	43
Figura 3.3 – Representação de cada modo operacional do processo na SBM-Multiclasse. . . . .	45
Figura 3.4 – Procedimento de inferência realizado pela SBM-Multiclasse. O estado atual é associado a classe que melhor aproximá-lo, nesse exemplo, seria associado a Classe 2 (Falha de Alta Pressão). . . . .	45
Figura 4.1 – Estrutura geral da eSBM. . . . .	48
Figura 4.2 – Ilustração do cálculo da matriz $G$ , o símbolo + em vermelho indica a média ou centro do grupo, $\bar{X}$ . A região de grupo (elipse em cinza) é deslocada fazendo cada ponto da matriz memória como centro (ponto circular em vermelho) para então ser efetuado o cálculo da similaridade entre os demais. . . . .	52

Figura 4.3 – Ilustração do procedimento de atualização de grupos. Na primeira linha é apresentada a situação em que as duas condições são satisfeitas, e tanto os atributos estatísticos quanto a matriz memória são atualizados. Na segunda linha a situação em que nenhuma das condições são satisfeitas, não ocorrendo nenhuma atualização no respectivo grupo. . . . .	54
Figura 4.4 – Ilustração do procedimento de união de grupos. Quando as condições de união não são satisfeitas (a), mantendo a estrutura atual, e quando são satisfeitas (b), nesse exemplo para o Grupo 1 e Grupo 3, efetuando a união e resultando no Grupo 1, 3. . . . .	56
Figura 4.5 – Ilustração do resultado da abordagem proposta para composição da matriz memória utilizando os extremos do intervalo recomendado para $\gamma_{point}$ . . . . .	57
Figura 4.6 – Base de dados sintética bidimensional (a), conjunto de amostras representativas selecionadas com $\gamma_{group} = 0.99$ e $\gamma_{point} = 0.15$ (b) e com $\gamma_{group} = 0.5$ e $\gamma_{point} = 0.15$ (c). As elipses nas últimas figuras representam os diferentes tamanhos da região de grupo em função da variação de $\gamma_{group}$ . . . . .	61
Figura 4.7 – Estado selecionado no conjunto inicial. As duas figuras apresentam o mesmo estado, optou-se por essa forma de apresentação para uma melhor perspectiva de distância do estado em relação as duas matriz memória. . . . .	62
Figura 4.8 – Qualidade da estimativa em função do tamanho da região de grupo e do operador de similaridade para as duas matriz memória. O eixo $y$ apresenta o erro relativo que trata-se da métrica utilizada na fase de Classificação. . . . .	62
Figura 4.9 – Influência do tamanho da região pontual no eSBM+. Impacto na quantidade de estados adicionados à matriz memória (a) e na qualidade da estimativa (b). . . . .	64
Figura 4.10–Tempo médio e mediana (em milissegundos) de 100 execuções no cálculo de uma estimativa pelo eSBM+ em função da quantidade de estados representativos para cada operador de similaridade. . . . .	65
Figura 4.11–Base de dados sintética com seis grupos. . . . .	65
Figura 4.12–Influência do tamanho da Janela de Criação na formação de grupos. . .	66
Figura 4.13–Comportamento do agrupamento em função do tamanho da Janela de Criação. As elipses representam a região de grupo de cada grupo e os pontos coloridos os estados representativos. . . . .	67
Figura 4.13–Comportamento do agrupamento em função do tamanho da Janela de Criação. As elipses representam a região de grupo de cada grupo e os pontos coloridos os estados representativos. . . . .	68



Figura 4.14–Influência do limiar mínimo de similaridade na qualidade de estimativa para cada operador de similaridade. . . . .	69
Figura 4.15–Influência do percentual de resíduo aceitável na formação de grupos. . .	69
Figura 4.16–Comportamento do agrupamento em função do percentual de resíduo aceitável. As elipses representam a região de grupo de cada grupo e os pontos coloridos os estados representativos. . . . .	70
Figura 4.16–Comportamento do agrupamento em função do percentual de resíduo aceitável. As elipses representam a região de grupo de cada grupo e os pontos coloridos os estados representativos. . . . .	71
Figura 4.17–Qualidade da estimativa em função da aleatoriedade do método de composição da matriz memória em 100 execuções. . . . .	72
Figura 5.1 – Conjunto com 10 realizações do vetor aleatório $X \sim \mathcal{N}_2(\mu, \Sigma)$ e seus respectivos eixos, $x_1$ e $x_2$ . . . . .	76
Figura 5.2 – Componentes obtidas para o conjunto de pontos ilustrativo. . . . .	77
Figura 5.3 – Interpretação dos valores dos coeficientes/autovetores relacionados à primeira componente, $PC_1$ . <i>rad</i> representa uma função que realiza a transformação de graus para radianos. . . . .	78
Figura 5.4 – Interpretação dos valores dos coeficientes/autovetores relacionados à segunda componente, $PC_2$ . <i>rad</i> representa uma função que realiza a transformação de graus para radianos. . . . .	78
Figura 5.5 – Ilustração do comportamento do PCA quando aplicado em variáveis descorrelacionadas (a) ou descorrelacionadas e com variâncias iguais entre si (b). . . . .	79
Figura 5.6 – CPV para o conjunto de dados ilustrativo. . . . .	80
Figura 5.7 – Ilustração dos valores críticos para cada índice considerando um problema bidimensional. O eixo horizontal representa o SP e o vertical o SR. As linhas pontilhadas representam o valor crítico do índice $T^2$ , as tracejadas do $SPE$ e a elipse da combinação dos dois anteriores. . . . .	86
Figura 6.1 – Estrutura do sistema do atuador utilizado no <i>benchmark</i> . . . . .	94
Figura 6.2 – Dados simulados do <i>benchmark</i> DAMADICS para a operação normal e para as falhas $f_1$ , $f_7$ e $f_{13}$ utilizando a biblioteca do Simulink®. . . . .	96
Figura 6.3 – Dados simulados duplicados e em função da média e desvio padrão móvel de cada variável. . . . .	97
Figura 6.4 – Resultado da aplicação do eSBM e eSBM+ para o conjunto simulado. .	98
Figura 6.5 – Tempo médio e mediana (em milissegundos) de 30 execuções do eSBM e eSBM+ em função do número de amostras processadas. . . . .	99
Figura 6.6 – Consumo de memória RAM (em megabytes) do eSBM e eSBM+ em função do número de amostras processadas. . . . .	100

Figura 6.7 – Falha 1: Desbalanceamento no nível de vazão do gás de selagem e elevação do diferencial de pressão do filtro de gás - LP. . . . .	105
Figura 6.8 – Falha 2: Vazamento no selo primário. . . . .	105
Figura 6.9 – Falha 3: Elevação das pressões de selagem. . . . .	106
Figura 6.10–Dados da operação normal intercalado com as falhas utilizado no experimento. . . . .	106
Figura 6.11–Resultado do monitoramento do sistema de selagem com os métodos considerados no experimento. . . . .	108
Figura 6.12–Contribuição de cada variável para a criação de grupos pelo eSBM+. . . . .	108
Figura 6.13–Contribuição de cada variável para a criação de grupos pelo OEC. . . . .	109
Figura 6.14–Contribuição de cada variável para a criação de grupos pelo OEC modificado. . . . .	109
Figura 6.15–Contribuição de cada variável para a criação de grupos pelo AutoCloud. . . . .	110
Figura 6.16–Contribuição de cada variável para a criação de grupos pelo MicroTEDA-Clus. . . . .	110
Figura 6.17–Contribuição de cada variável para a criação de grupos pelo MicroTEDA-Clus modificado. . . . .	111

# Lista de Tabelas

Tabela 1 – Valor da matriz $M$ para cada índice. . . . .	86
Tabela 2 – Matriz de decisão para avaliação de desempenho de métodos direcionados à detecção e classificação de falhas. . . . .	92
Tabela 3 – Tipos de falhas disponíveis no <i>benchmark</i> DAMADICS. . . . .	95
Tabela 4 – Tipo e intensidade possível para cada falha no <i>benchmark</i> DAMADICS. . . . .	95
Tabela 5 – Métricas de detecção e classificação de falhas para o resultado apresentado pelos métodos eSBM e eSBM+. . . . .	98
Tabela 6 – Métricas de tempo de detecção e classificação de falhas para o resultado apresentado pelos métodos eSBM e eSBM+. . . . .	99
Tabela 7 – Variáveis utilizadas para o monitoramento do sistema de selagem. . . . .	104
Tabela 8 – Três falhas que ocorreram no sistema de selagem. . . . .	104



# Lista de Algoritmos

1	Método proposto para composição da matriz memória. . . . .	55
2	<i>Evolving Similarity-Based Modeling Plus.</i> . . . . .	58
3	Procedimento para seleção de variáveis candidatas responsáveis pela falha. .	89



# Lista de Abreviaturas e Siglas

ACC	Accuracy
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CCK	Cauchy Kernel
CNG	Criando Novo Grupo
CPV	Cumulative Percentage Variance
DAMADICS	Development and Application of Methods for Actuator Diagnosis in Industrial Control Systems
DDT	Detection Delay Time
DDR	Double Data Rate
DPCA	Dynamic Principal Components Analysis
EMPF	Evolving Monitoring Processes Framework
eSBM	Evolving Similarity-Based Modeling
FIR	Fault Isolation Rate
IID	Independente e Identicamente Distribuída
INC	Inicializando
IDT	Isolation Delay Time
IHM	Interface Homem-Máquina
IMK	Inverse Multiquadric Function
LK	Laplacian Kernel
MOC	Modo de Operação Conhecido

MSPC	Multivariate Statistical Process Control
NOC	Normal Operation Condition
OEC	Online Elipsoidal Clustering
PCA	Principal Components Analysis
PC	Principal Components
POD	Probability of Detection
POFA	Probability of False Alarm
RAM	Random Access Memory
RBC	Reconstruction-Based Contribution
RBF	Radial Basis Function
SBM	Similarity-Based Modeling
SPE	Squared Prediction Error
SP	Subespaço Principal
SR	Subespaço Residual
STO	Saturated Triangular Operator
SVM	Support Vector Machine
TEDA	Typicality and Eccentricity Data Analytics
VA	Variável Aleatória
WSF	Wegerich Similarity Function



# Sumário

<b>1</b>	<b>Introdução</b>	<b>25</b>
1.1	Motivação	27
1.2	Objetivos	27
1.3	Organização do Trabalho	28
<b>2</b>	<b>Monitoramento de Processos Industriais</b>	<b>29</b>
2.1	<i>Framework</i> para monitoramento de processos industriais	30
2.1.1	Detalhamento de Níveis	30
2.1.1.1	Modelagem	30
2.1.1.2	Causa	32
2.1.1.3	Complementação	32
2.1.2	Exemplo ilustrativo de funcionamento	32
2.2	Resumo do capítulo	36
<b>3</b>	<b>Modelagem Baseada em Similaridade de Modelos</b>	<b>39</b>
3.1	Detector de anomalias	39
3.1.1	Operadores de similaridade	42
3.2	Classificador de modos operacionais	44
3.2.1	Composição da matriz memória	45
3.3	Resumo do capítulo	46
<b>4</b>	<b>Clusterização Evolutiva Baseada na SBM</b>	<b>47</b>
4.1	SBM evolutiva	47
4.2	Modificações propostas	49
4.2.1	Operação de Similaridade	49
4.2.2	Estimação	51
4.2.3	Classificação	51
4.2.4	Evolução	53
4.2.5	Composição da matriz memória	55
4.2.6	Sensibilidade da parametrização	60
4.2.6.1	Região de grupo	60
4.2.6.2	Região pontual	62
4.2.6.3	Janela de criação	63
4.2.6.4	Limiar mínimo de similaridade	63
4.2.6.5	Percentual de resíduo aceitável	64
4.2.6.6	Método de composição	65
4.3	Resumo do capítulo	66
<b>5</b>	<b>Diagnóstico de Causa Raiz de Falhas Baseado em Contribuição</b>	<b>73</b>
5.1	Controle de Processo Estatístico Multivariado	73

5.1.1	Análise de Componentes Principais . . . . .	73
5.1.2	Controle estatístico . . . . .	80
5.1.2.1	Estatística de <i>Hotteling</i> . . . . .	82
5.1.2.2	Erro Quadrático . . . . .	83
5.1.2.3	Combinação de estatísticas . . . . .	84
5.1.3	Análise Dinâmica de Componentes Principais . . . . .	85
5.2	Contribuição Baseada em Reconstrução . . . . .	87
5.3	RBC on-line . . . . .	89
5.4	Resumo do capítulo . . . . .	90
<b>6</b>	<b>Experimentos . . . . .</b>	<b>91</b>
6.1	Comparativo de desempenho entre eSBM e eSBM+ . . . . .	91
6.1.1	Métricas de desempenho . . . . .	92
6.1.2	Experimento com <i>benchmark</i> DAMADICS . . . . .	93
6.2	Demonstração de uso do EMPF . . . . .	98
6.2.1	Sugestões de modificações ao OEC e MicroTEDAClus . . . . .	100
6.2.2	Experimento com sistema de selagem . . . . .	102
6.3	Resumo do capítulo . . . . .	107
<b>7</b>	<b>Conclusão . . . . .</b>	<b>113</b>
7.1	Trabalhos Futuros . . . . .	114
	<b>Referências . . . . .</b>	<b>115</b>

# Capítulo 1

## Introdução

Com o nascimento da indústria moderna, por consenso, na segunda metade do século XVIII por meio da Revolução Industrial iniciada na Inglaterra no setor têxtil e metalúrgico, houve uma necessidade constante e crescente do desenvolvimento tecnológico e científico. Essa necessidade, adveio, em parte, para a melhoria dos processos de produção nas indústrias que na época constituía-se majoritariamente de manufatura, para a produção em grande escala e com menor custo que surgiu devido à alta demanda na época nos dois setores mencionados, que por sua vez foi ocasionada pelo êxodo rural, e para a expansão das fronteiras comerciais (Iglésias, 1981).

Atualmente essa necessidade ainda se mantêm forte, contudo, busca-se além de melhorias dos processos de produção e minimização de custos, atender a níveis satisfatórios de qualidade, segurança aos trabalhadores responsáveis e ao meio externo, evitando perdas de vidas, prejuízos econômicos e ambientais (Venkatasubramanian et al., 2003b). Para tanto, houve um aumento considerável no grau de complexidade dos processos industriais, tornando-se compostos por dezenas de subprocessos e/ou milhares de variáveis que representam medições de grandezas físicas utilizadas para indicar o estado de funcionamento do processo. Devido a isso, tornou-se impraticável seu gerenciamento apenas por esforço humano, fazendo surgir a necessidade de ferramentas computacionais capazes de auxiliar os profissionais responsáveis, ou simplesmente supervisores, nessa tarefa.

Nesse sentido, a comunidade científica da área tem direcionado constantemente seus esforços para a proposição de ferramentas voltadas a auxiliar o supervisor a manter o processo funcionando de forma estável, minimizando custos e maximizando a produtividade, a qualidade e a segurança nos ambientes de produção das indústrias modernas. Como exemplo dessas ferramentas, têm-se os métodos baseados em dados históricos, que tem ganhado maior atenção nas últimas décadas devido a facilidade de obtenção de dados/medições, que por sua vez possibilitam modelar o comportamento do processo sem a representação matemática de sua física, pois em geral, para plantas de larga de escala e com alta complexidade é extremamente custoso obter esse tipo de conhecimento (Chiang

et al., 2000).

A título de exemplo de métodos pertencentes a essa classe, têm-se as redes neurais clássicas (Venkatasubramanian et al., 1990; Samanta, 2004), redes neurais profundas (Heo and Lee, 2018; Zhao et al., 2019; Zhang et al., 2019), modelos neuro-fuzzy (Mendonça et al., 2009), Máquina de Vetores de Suporte (SVM) (Namdari and Jazayeri-Rad, 2014), Sistemas Especialistas (Cho and Park, 1997), Multivariate Statistical Process Control (MSPC) (Joe Qin, 2003; Montgomery, 2007; Jiang et al., 2019), Similarity-based Modeling (SBM) (Marins et al., 2018), dentre muitos outros. Uma visão geral pode ser obtida em (Venkatasubramanian et al., 2003a; Ge, 2017). De forma geral, para uso desses métodos, é necessário apenas que seja fornecida uma base de dados rotulada (*i.e.*, uma amostra/medição do processo acompanhada do identificador do estado operacional) e alguns valores para os parâmetros do método, após executa-se uma etapa de treinamento que busca ajustar os parâmetros internos do método maximizando a qualidade da modelagem.

Neste mesmo contexto, surgiu nas últimas décadas uma nova classe de métodos, os chamados métodos evolutivos (Angelov et al., 2010), ao contrário de alguns citados anteriormente, dispensa uma etapa de treinamento anterior ao uso. Seu treinamento, aqui chamado de aprendizagem, é executado de forma incremental e on-line, ponto-a-ponto, sem reapresentação. Sua aprendizagem envolve não apenas ajustar os parâmetros dos modelos, como também ajustar sua estrutura para acomodar informações que o estado atual não contém. Devido a essas características, tal classe de método tornou-se mais atrativa para uso no contexto industrial. Os baseados em aprendizagem não supervisionada são os mais utilizados, uma vez que a informação de rotulagem não está disponível *a priori*. Dessa forma, novos eventos são detectados e incorporados a base conhecimento dos métodos para rotulagem futura pelo supervisor (Lemos et al., 2013).

Tendo em mente as diferentes abordagens de métodos evolutivos existentes na literatura, e seu uso incipiente nas indústrias. Esse trabalho propõe um *framework* capaz de unificar o uso das diferentes abordagens, facilitando o uso dessa classe de métodos no contexto industrial, possuindo também a capacidade apontar a causa raiz de eventos detectados, permitindo uma maior transparência no comportamento dos métodos evolutivos para o supervisor.

Além disso, recentemente, Mariano and Caminhas (2019) propuseram o *Evolving Similarity-Based Modeling* (eSBM), uma versão evolutiva de uma técnica há bastante tempo utilizada na indústria, a SBM (Marins et al., 2018). O método proposto demonstrou desempenho satisfatório nos experimentos realizados pelos autores. Com isso, neste trabalho também é proposta uma nova versão do eSBM.

## 1.1 Motivação

A tarefa de monitoramento é vital para a supervisão, se não é realizada com qualidade, a supervisão estará dependente, em grande parte, de esforço humano, podendo não ser suficiente ou praticável. Nesse sentido, é importante que as ferramentas utilizadas no monitoramento forneçam ao supervisor informações precisas e com alta confiabilidade, idealmente em tempo real, sobre o comportamento do processo, para que diante de comportamentos anômalos o supervisor possa tomar conhecimento da situação o mais rápido possível, e tomar as medidas necessárias para a resolução do problema.

Na literatura existe a classe de métodos evolutivos, ideal para uso no monitoramento de processos industriais devido sua característica de autoadaptação. No entanto, o uso dessa classe de métodos ainda é incipiente nesse contexto, em parte, por ser relativamente recente se consideradas as demais classes já consolidadas nesses ambientes. Ademais, dentro da classe de métodos evolutivos são propostas diferentes abordagens que modelam a dinamicidade do processo por meio de estruturas distintas.

Embora na literatura já exista uma quantidade significativa de métodos evolutivos, a grande maioria é ideal apenas para aplicações e situações específicas, com isso, a proposição de novas abordagens capazes de lidar com toda e qualquer situação é altamente desejável.

Além da detecção e classificação, também é importante para o supervisor saber o que as gerou, do contrário, na ocorrência de uma operação anormal e indesejável, identificar a causa raiz para a resolução do problema pode vir a ser uma tarefa difícil e demorada, impactando negativamente na qualidade e segurança do processo.

## 1.2 Objetivos

Esse trabalho tem por objetivo geral propor um *framework* para unificar o uso das diferentes abordagens de métodos evolutivos com a capacidade de indicar a causa raiz de novidades capturadas pelos métodos. De forma secundária, busca-se investigar e implementar melhorias ao eSBM. Para tanto, os seguintes objetivos específicos são necessários:

- Investigar as diferentes abordagens de métodos evolutivos na literatura
- Desenvolver um *framework* para unificar o uso das diferentes abordagens de métodos evolutivos
- Agregar a capacidade de indicar a causa raiz de novidades capturadas no processo pelos métodos evolutivos ao *framework*

- Demonstrar o uso do *framework*
- Investigar o método eSBM
- Implementar melhorias ao método eSBM
- Realizar experimentos para a validação das mudanças implementadas no eSBM

### 1.3 Organização do Trabalho

O trabalho está dividido em oito capítulos, incluindo o atual. No capítulo atual foi apresentado o contexto histórico da necessidade do monitoramento de processos industriais, a motivação e os objetivos desse trabalho. No próximo capítulo, Capítulo 2, é apresentado o *framework* proposto. O Capítulo 3 contém toda a base teórica da técnica precursora do eSBM, a SBM, como detector de anormalidade e classificador. No Capítulo 4 é apresentado o eSBM e as melhorias que estão sendo propostas nesse trabalho. No Capítulo 5 é apresentado um método voltado ao diagnóstico de causa raiz de falhas e seus fundamentos teóricos, ao qual irá compor o *framework* proposto. No Capítulo 6 são apresentados os experimentos realizados para validar e demonstrar as proposições desse trabalho. Por fim, no Capítulo 7 são apresentadas as conclusões finais e perspectivas de trabalhos futuros.

## Capítulo 2

# Monitoramento de Processos Industriais

A tarefa de monitoramento de processos por técnicas computacionais consiste em detectar e diagnosticar eventos em tempo real, de forma a auxiliar o supervisor no controle de qualidade. Diagnosticar, de forma mais detalhada, consiste em *identificar* a qual situação já vivenciada o evento relaciona-se, *isolar* o instante que inicia e termina, e apontar a *causa raiz* do evento.

Na literatura são propostos diversos métodos para lidar separadamente com cada tarefa. Uma classe de métodos, os evolutivos, são capazes de executar quase todas as tarefas, apenas a causa raiz não é comumente realizada. Esses métodos, embora compartilhem da mesma filosofia de construção de conhecimento do processo, possuem abordagens distintas, representando os eventos por meio de diferentes estruturas. Tal característica dificulta o uso em ambiente industrial real, pois torna-se necessária a implementação de mecanismos para lidar com o gerenciamento de diferentes estruturas de modelagem do processo, apresentação da informação de saída no formato adequado à equipe de supervisão e demais requisitos atrelados as especificidades de cada tipo de processo.

Com base nisso, neste capítulo propõe-se um *framework* portátil e escalável que generaliza o uso das muitas abordagens na literatura. Nessa estrutura é permitida a adição de quaisquer tipos de métodos evolutivos, o que inclui métodos com aprendizagem supervisionada e não supervisionada. Além disso, também é proposto a adaptação de um método de causa raiz baseado em contribuição, o *Reconstruction-Based Contributions* (RBC) (Li et al., 2014), como uma ferramenta de explicabilidade de modelos, uma vez que é altamente desejável apresentar informações ao supervisor que o permita entender o que está ocorrendo no processo, e capturado pelo método evolutivo em uso. À estrutura proposta deu-se o nome de *Evolving Monitoring Processes Framework* (EMPF).

## 2.1 *Framework* para monitoramento de processos industriais

Considerando as diferentes e complementares informações de representação do processo, qualitativa (*e.g.*, nome do modo operacional, grau de severidade, procedimento executado em sua ocorrência, etc. ) e quantitativa (*i.e.*, representação matemática), o *framework* é dividido em três principais níveis, a saber:

- *Modelagem*: Contém os métodos evolutivos e sua modelagem do processo, isso inclui informações quantitativas e qualitativas;
- *Causa*: Contém uma adaptação do RBC para funcionamento on-line, ao qual é responsável por auxiliar o supervisor no entendimento do comportamento dos métodos evolutivos;
- *Complementação*: Contém interfaces, cujo principal objetivo é fornecer os meios necessários para que o supervisor complemente as informações quantitativas com qualitativas.

Na Figura 2.1 é apresentado um esquema geral do EMPF, uma visão mais detalhada de cada nível é dada na Seção 2.1.1. Na figura é possível observar que tanto o serviço de aquisição de dados quanto a Interface Homem-Máquina (IHM) é independente do *framework*, facilitando sua implantação em quaisquer ambientes, e com as devidas alterações, até mesmo em servidores dedicados para funcionamento como provedor de serviços. O serviço de aquisição de dados pode fornecer dados "brutos" do processo, pré-processados ou quaisquer outros. A IHM pode ser construída utilizando qualquer tecnologia, e se os resultados do *framework* forem fornecidos para consumo por meio de *Web Services* ou algo similar, pode ser desenvolvida de forma nativa em qualquer plataforma (*e.g.*, *web*, *mobile* e *desktop*).

### 2.1.1 Detalhamento de Níveis

#### 2.1.1.1 Modelagem

A Modelagem engloba os métodos evolutivos, o conhecimento quantitativo e qualitativo e um Gerenciador de Conhecimento (GC). O GC é responsável por gerenciar as informações quantitativas e qualitativas, tornando o gerenciamento independente do método. Para tanto, utiliza-se, dentro do paradigma de programação orientada a objetos, classes abstratas para representar a estrutura base de um/uma grupo/classe e a informação qualitativa, essa última de acordo com o interesse do supervisor. A classe abstrata referente ao/a grupo/classe, definirá as funções (*i.e.*, método dentro do paradigma) comuns a todo



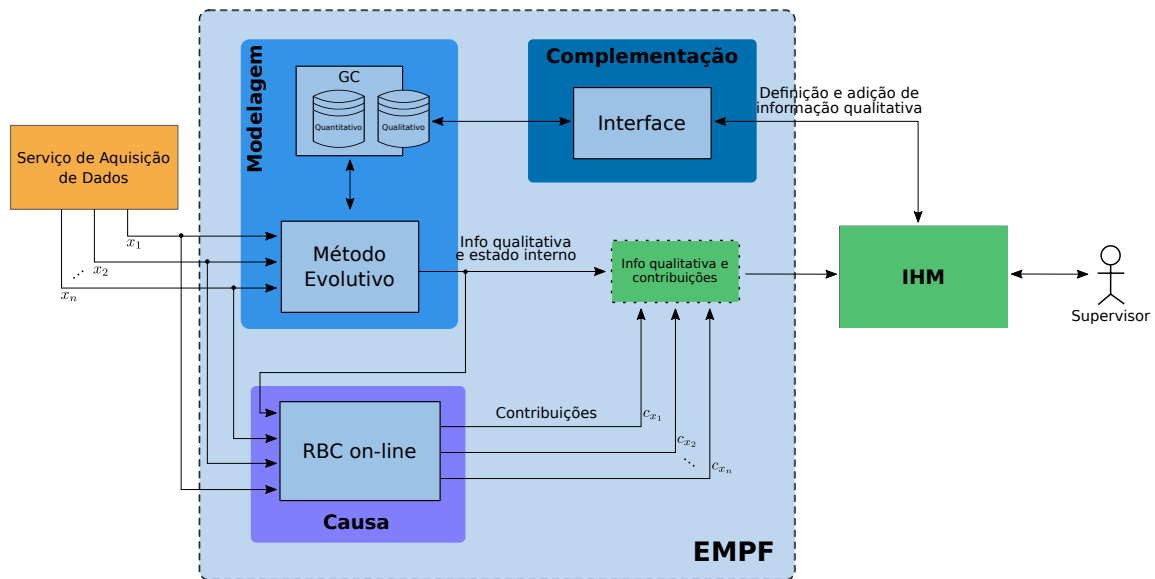


Figura 2.1 – Esquema geral do EMPF.

método evolutivo, sendo: atualização, classificação e união. Na classificação, além do grupo a qual a entrada melhor relaciona-se, também deverá ser retornado o estado interno do método, podendo ser um dos três descritos a seguir:

- *Inicializando (INC)*: O método ainda não possui grupos/classes criados/criadas e está criando o/a primeiro/primeira. Esse estado pode durar apenas uma amostra ou várias dependendo da abordagem utilizada pelo método.
- *Modo de Operação Conhecido (MOC)*: O método já possui grupos/classes efetivamente criados/criadas e a entrada foi associada a um/uma desses/dessas. É esperado que o método passe a maior parte de sua execução nesse estado.
- *Criando Novo Grupo (CNG)*: O método detectou uma mudança no processo e iniciou o procedimento de criação de um/uma novo/nova grupo/classe. De forma similar ao primeiro estado, esse pode durar uma amostra ou várias, dependendo apenas da abordagem utilizada.

O estado interno é necessário para o RBC distinguir entre as amostras que de fato pertencem ao/a grupo/classe dado/dada na classificação, para então criar o modelo de referência corretamente. A título de exemplo, considere um método qualquer com aprendizagem não supervisionada. O mesmo utiliza uma janela com  $k$  amostras consecutivas para criar um grupo. O grupo será criado apenas no fim da janela, contudo, durante esse período as amostras serão associadas a um dos grupos existentes, quando de fato pertencem ao novo grupo. Logo, sem o estado interno não há como identificar

que essas  $k$  amostras da janela não devem ser consideradas aos grupos em que foram previamente associadas.

### 2.1.1.2 Causa

Como comentado anteriormente, o nível de Causa engloba uma versão adaptada do método RBC para funcionamento on-line. Na versão proposta originalmente, o método indica, a partir de um modelo da operação normal, quais as variáveis mais afetadas pela falha. Na adaptação em questão, o modelo de referência é definido como o/a grupo/classe anterior a uma mudança, indicando quais as variáveis mais afetadas pelo novo grupo em relação ao anterior. Por exemplo, se o método evolutivo estiver associando o comportamento do processo ao Grupo 1, e a partir de um determinado instante de tempo houve a criação do Grupo 2, nesse momento o RBC é executado e fornecerá quais as variáveis que mais mudaram em termos de média e/ou variância em relação ao Grupo 1. Essa adaptação, chamada de RBC on-line, será apresentada em detalhes no Capítulo 5, mais especificamente na Seção 5.3.

### 2.1.1.3 Complementação

O nível de Complementação engloba uma interface que deve ser utilizada por uma IHM, permitindo ao supervisor interagir com o *framework*, ou mais especificamente com o GC, para complementar as informações quantitativas de cada grupo/classe, adicionando as qualitativas, tais como: Qual modo operacional o/a grupo/classe representa? Quando ocorreu? Qual a severidade desse modo operacional? Qual o procedimento a ser executado na ocorrência desse modo operacional? dentre outras informações que a equipe julgar relevante.

Esse nível é vital para maximizar a qualidade do monitoramento e por consequência da supervisão, uma vez que com o conhecimento técnico do especialista agregado ao conhecimento quantitativo dos métodos evolutivos, em futuras ocorrências das situações já vivenciadas o diagnóstico dar-se-á de forma mais eficiente e eficaz.

## 2.1.2 Exemplo ilustrativo de funcionamento

Considerando que o conjunto de dados tridimensional apresentado na Figura 2.2 representa medições de sensores de um determinado processo, e que nos instantes  $t = 500$  e  $t = 1500$  ocorreram pela primeira vez dois eventos distintos que duram 500 amostras. Após investigação do supervisor, o mesmo consegue encontrar e solucionar a causa dos eventos, retornando o processo ao modo de operação normal. Entretanto, após um período de normalidade, os eventos voltam a ocorrer de igual maneira.

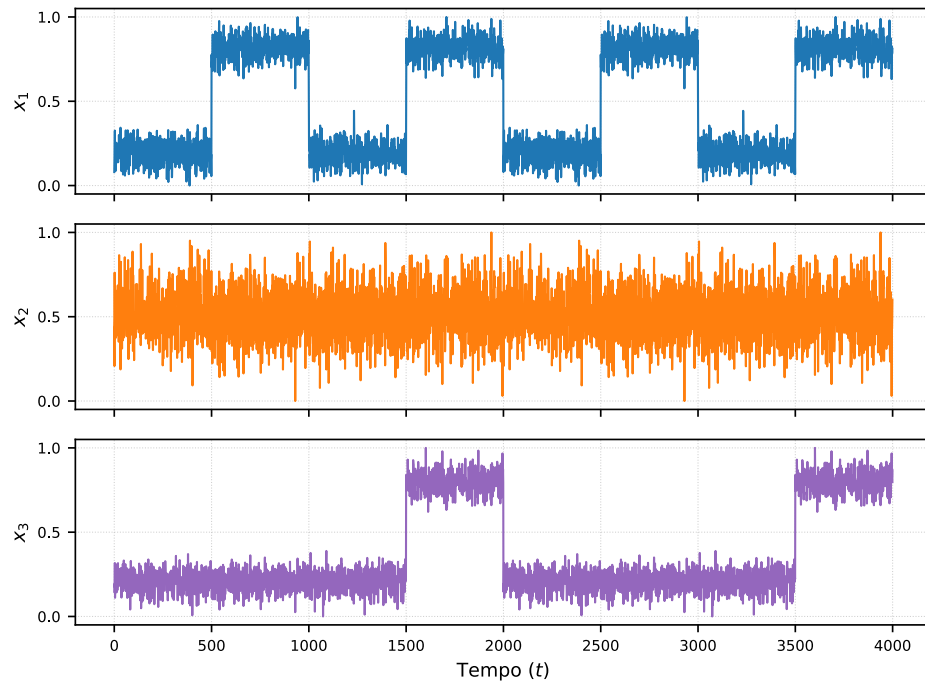


Figura 2.2 – Base de dados tridimensional sintética que simula um processo real.

Na ocorrência de cada evento, foi necessária uma determinada quantidade de tempo para que houvesse a percepção da situação, e também fosse avaliada as condições de cada grandeza física do processo (mensuradas pelos sensores) para encontrar a raiz do problema, de forma que as melhores e mais seguras medidas possam ser executadas para sua solução. No exemplo citado, devido a baixa dimensionalidade, é fácil perceber que o terceiro e quarto evento são os mesmos dois já ocorridos anteriormente. Contudo, em processos com alta dimensionalidade, não seria tão simples chegar a essa conclusão, em que provavelmente seria empenhado o mesmo esforço da primeira ocorrência para obtê-la.

Por meio do EMPF é possível monitorar o processo de forma mais eficiente e eficaz. Os métodos evolutivos podem detectar, identificar e incorporar novos eventos de forma simples, enquanto que o RBC on-line indica um conjunto significativamente menor de variáveis para o supervisor investigar. Além disso, uma vez encontrada a raiz real do problema e o procedimento para sua solução, tal informação pode ser adicionada a base de conhecimento do método evolutivo, para que em futuras ocorrências o supervisor saiba de forma imediata do que trata-se o evento e como resolvê-lo.

Incorporando o método evolutivo eSBM+ ao EMPF, tal método possui aprendizagem não supervisionada e modela eventos por meio de grupos, o mesmo é proposto como uma versão do método eSBM no Capítulo 4 desta dissertação. O monitoramento da base de dados apresentada com o EMPF ocorre da seguinte forma:

- 1º) Cada ponto é lido de um arquivo de dados e fornecido como entrada ao EMPF

2º) O primeiro evento, em  $t = 500$ , é detectado e incorporado a base de conhecimento com atraso de 18 amostras devido o uso janelas pelo eSBM+, conforme apresentado na Figura 2.3. O RBC indica corretamente a variável  $x_1$  como a mais influente (em percentual) na necessidade de criação do novo grupo. Em uso prático, essas informações serão entregues ao supervisor que terá a responsabilidade de complementá-las com as devidas informações qualitativas.

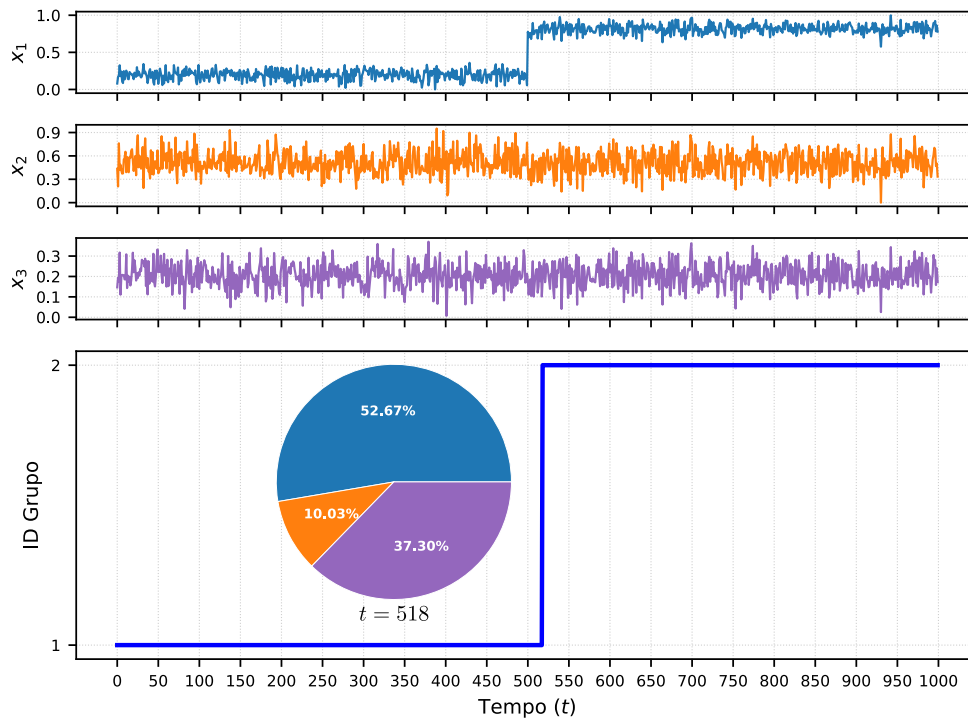


Figura 2.3 – Resultado do monitoramento com EMPF para o período de criação do segundo grupo.

3º) Em  $t = 1000$ , o EMPF detecta o encerramento do primeiro evento (Figura 2.4), e dada a mudança de grupo, é realizada o cálculo de contribuição. Novamente, o RBC indica corretamente a variável  $x_1$  como a mais influente na mudança. Perceba que em  $t = 1073$  é indicada uma mudança de grupo, entretanto, trata-se de um falso positivo. Sem considerar essa informação, o RBC realiza o cálculo das contribuições, obtendo um resultado similar ao da mudança anterior. Esse comportamento se dá porque o RBC ainda não obteve amostras suficientes referentes ao Grupo 1, com isso, toma como referência o Grupo 2.

4º) No instante  $t = 1500$  é detectado o segundo evento (Figura 2.5), o reflexo disso é a mudança do Grupo 1 para o Grupo 2, contudo, essa mudança ocorre apenas porque o Grupo 2 está mais próximo espacialmente do novo evento que o Grupo 1. Em  $t = 1514$  o Grupo 3 é efetivamente criado para representar o segundo evento. Nas duas mudanças, o RBC calcula e indica corretamente as variáveis  $x_1$  e  $x_3$  como as mais influentes no novo comportamento.

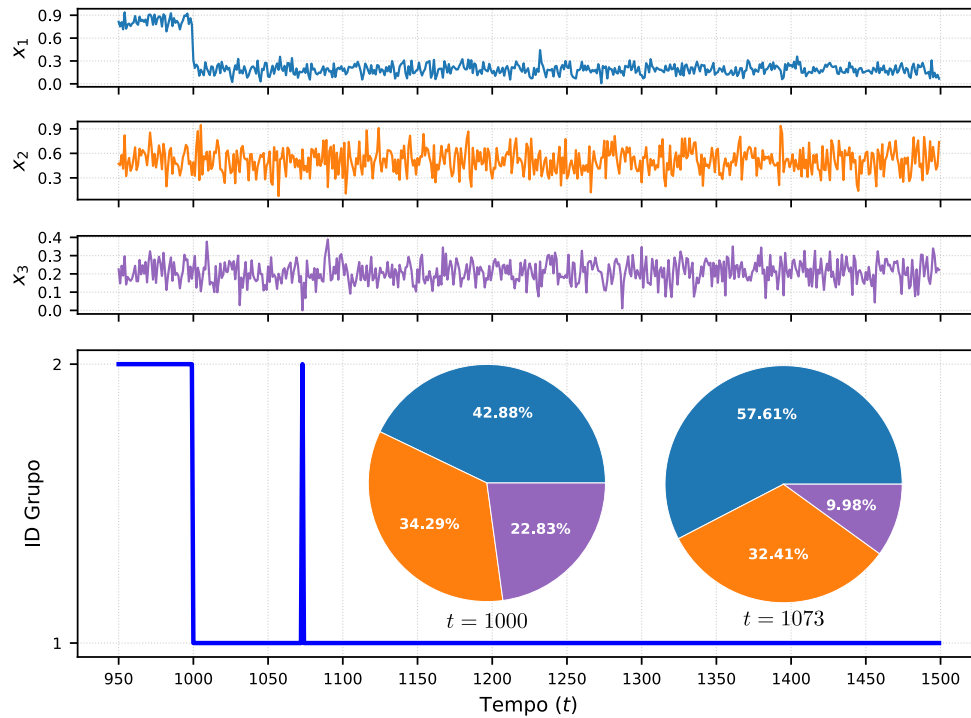


Figura 2.4 – Resultado do monitoramento com EMPF para o período de retorno do Grupo 2 para o Grupo 1.

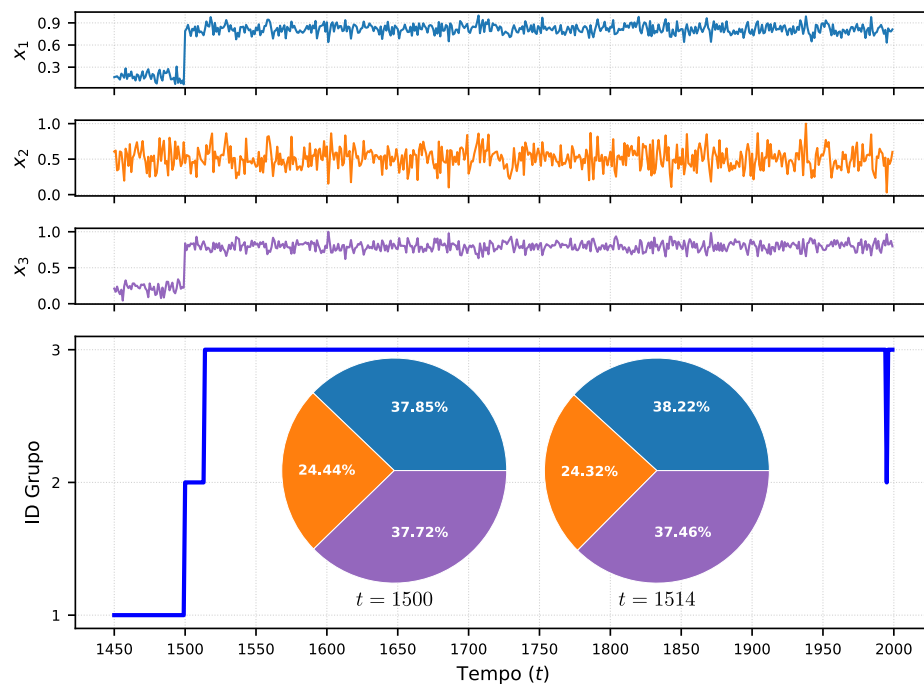


Figura 2.5 – Resultado do monitoramento com EMPF para o período de criação do terceiro grupo.

- 5º) Em  $t = 2000$  detecta-se o encerramento do novo evento, ocorrendo a transição do Grupo 3 para o Grupo 1 (Figura 2.6). Ocorre uma transição para o Grupo 2 antes de  $t = 2000$ , novamente um falso positivo devido as métricas de distância do eSBM+.

Nessa situação o RBC indica todas as variáveis como contribuintes, capturando apenas a variação natural das variáveis.

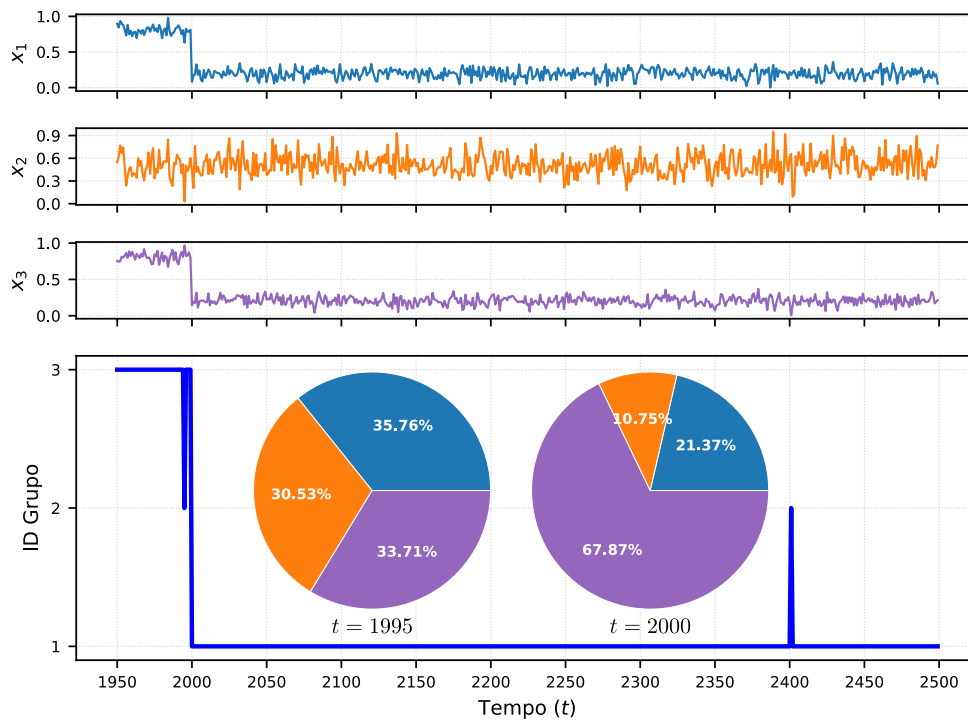


Figura 2.6 – Resultado do monitoramento com EMPF para o período de retorno do Grupo 3 para o Grupo 1.

6º) Na segunda vez em que ocorre cada evento, o EMPF indica de imediato que o comportamento já existe em sua base de conhecimento (Figura 2.7 e Figura 2.8), e considerando que o supervisor tenha adicionado as devidas informações qualitativas, as mesmas serão prontamente apresentadas.

Uma ilustração do EMPF mais próxima do uso real em um processo industrial está disponibilizada em <https://youtu.be/gX6D3mb1Uh0>.

## 2.2 Resumo do capítulo

O monitoramento de processos industriais tem demandado cada vez mais ferramentas capazes de modelar os diferentes modos operacionais possíveis com máxima qualidade e robustez, idealmente, à medida que surgem, pois devido a alta complexidade inerente aos processos industriais modernos é praticamente impossível conhecer todas as possibilidades de operação *a priori*.

Em vista disso, nesse capítulo foi proposto o *framework* EMPF, uma ferramenta capaz de efetuar o monitoramento por meio de quaisquer métodos evolutivos e ainda indicar a causa raiz de mudanças de comportamentos apontadas pelos mesmos. Para ilustrar o

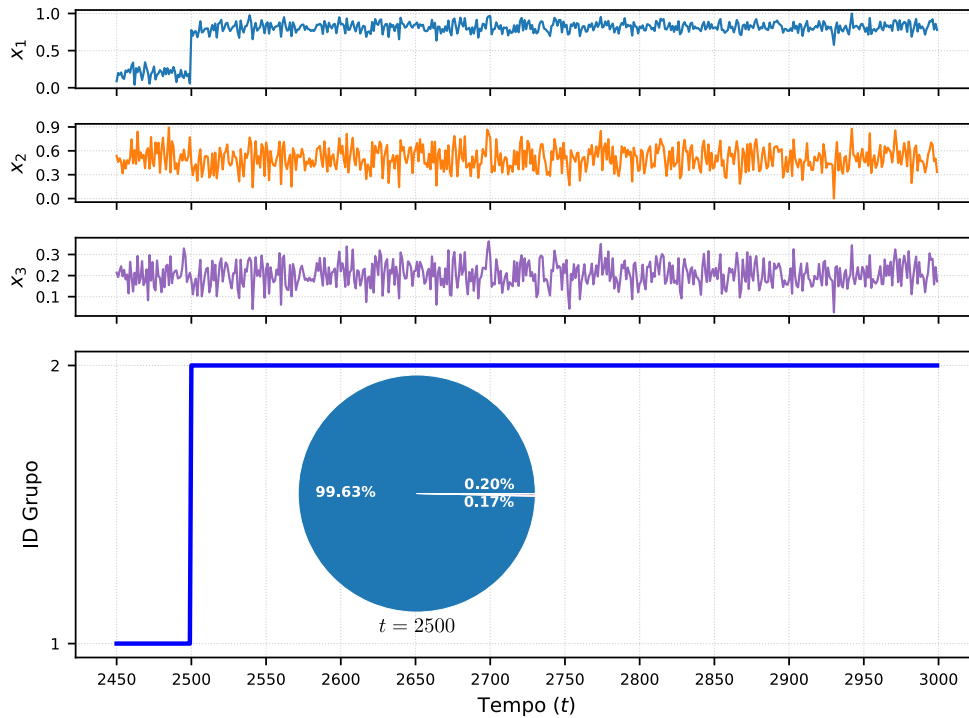


Figura 2.7 – Resultado do monitoramento com EMPF para o período de segunda ocorrência do primeiro evento.

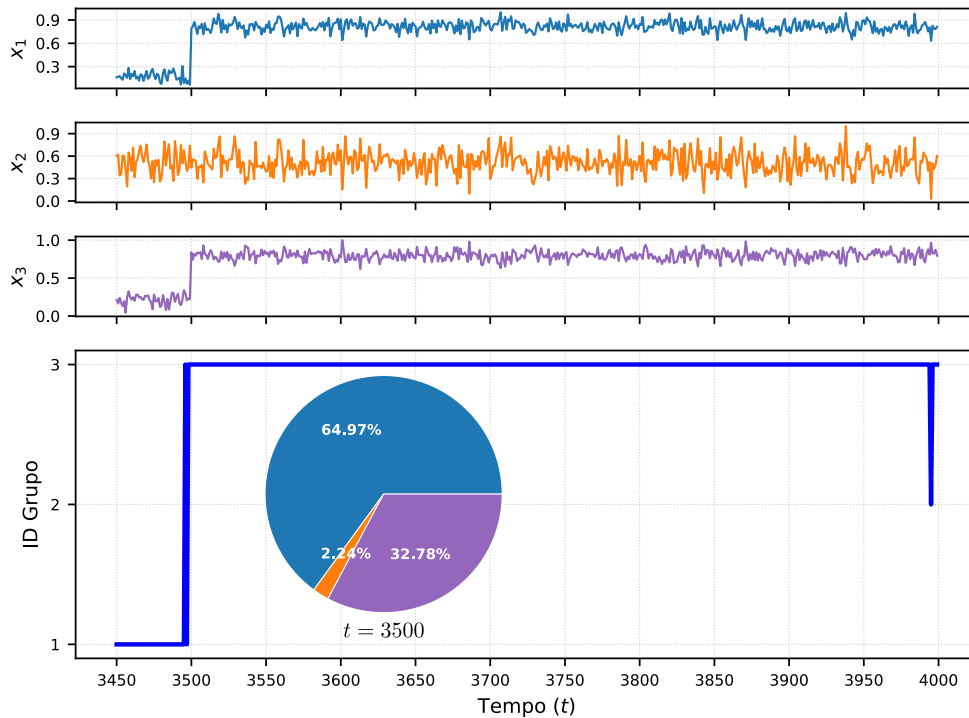


Figura 2.8 – Resultado do monitoramento com EMPF para o período de segunda ocorrência do segundo evento.

funcionamento interno da ferramenta foram apresentados os resultados do monitoramento de uma base de dados sintética tridimensional contendo três eventos distintos.





## Capítulo 3

# Modelagem Baseada em Similaridade de Modelos

A Modelagem Baseada em Similaridade (SBM) (do inglês, *Similarity-Based Modeling*) é uma técnica não paramétrica proposta por [Singer et al. \(1997\)](#) para detecção de comportamento anômalo em plantas de usinas nucleares. No entanto, com o passar dos anos seu uso estendeu-se às mais diversas áreas. A exemplo, [Tobar et al. \(2011\)](#) a utiliza para detecção de anomalias em plantas de geração de energia a gás. [Perez et al. \(2018, 2019\)](#) utilizam para caracterizar o processo de degradação de baterias de íons de lítio. Em ([Monkman and Wegerich, 2005](#); [Marins et al., 2018](#); [Wegerich et al., 2003](#)) é aplicada na detecção e classificação de falhas em máquinas elétricas. [Wegerich \(2004\)](#) a usa para detecção precoce de falhas incipientes em caixas de engrenagens intermediária, e em ([Pipke, 2014](#)) é criado e patenteado um método para monitorar o estado de saúde humana, capaz de detectar problemas de saúde incipientes.

Neste capítulo será apresentada a SBM, inicialmente em sua concepção original, como um detector de anomalias, e em seguida, em sua última versão, como um classificador multiclasse.

### 3.1 Detector de anomalias

Considerando um conjunto de  $k$  amostras ou medições ou estados referente ao modo de operação normal de um determinado processo, em que cada observação é uma agregação de medições de  $n$  variáveis ou sensores,  $x \in \mathbb{R}^n$ . Esse conjunto de observações

são acomodadas em uma matriz  $D$ , denominada matriz memória,

$$D^{k \times n} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & \dots & x_{kn} \end{bmatrix}, \quad (3.1)$$

e na SBM representará o modo de operação normal.

Considerando agora que já exista uma matriz memória referente a operação normal e que o estado do processo no instante  $t$  é representado por  $x_t \in \mathbb{R}^n$ , para avaliar se o processo está operando normalmente é necessário primeiramente gerar uma estimativa para o estado atual com base nas observações contidas na matriz memória,

$$\hat{x}_t = D^T w_t, \quad (3.2)$$

com  $T$  indicando a matriz transposta de  $D$  e  $w_t \in \mathbb{R}^k$  é o estimador linear de  $x_t$ .

Uma possibilidade para encontrar os coeficientes que compõem  $w_t$  consiste em minimizar a norma Euclidiana, fazendo isso são obtidas as seguintes expressões que fornecem o estimador linear ótimo, ou seja, que fornece a estimativa com menor norma possível (Singer et al., 1997):

$$w_t = (DD^T)^{-1} D x_t = G^{-1} a_t, \quad (3.3)$$

$$w_t^o = \frac{w_t}{\|w_t\|_1}, \quad (3.4)$$

$$\hat{x}_t = D^T w_t^o, \quad (3.5)$$

em que  $\|\cdot\|_1$  é a norma  $L1$ ,  $w_t^o \in \mathbb{R}^k$  é o estimador linear ótimo,  $G = DD^T \in \mathbb{R}^{k \times k}$  e  $a_t = D x_t \in \mathbb{R}^k$ .

Porém, conforme é apontado por Singer et al. (1997), essa abordagem possui algumas limitações, como: a matriz  $G$  deve ser não singular; incapacidade de acomodar ruídos; necessidade de uma grande quantidade de amostras na matriz memória. Limitações desse tipo impactam significativamente nas possibilidades de aplicações, em especial aplicações com necessidade de processamento em tempo real, em que a realização de tratamentos para os problemas citados podem não ser viáveis ou até mesmo ideais.

Para contornar o problema, Singer et al. (1997) propõe substituir algumas das operações de produto matricial em (3.5) por um operador não linear, ao qual denomina de

operador de similaridade. O objetivo é utilizá-lo para fornecer uma medida de similaridade  $0 \leq s \leq 1$  entre dois estados  $x_i$  e  $x_j$ , estados não similares terão  $s \approx 0$  e estados muito similares terão  $s \approx 1$ ,

$$s(x_i, x_j) = x_i \circ x_j, \quad 0 \leq s(x_i, x_j) \leq 1, \quad (3.6)$$

com  $\circ$  representando uma operação de similaridade.

Dessa forma, com as modificações propostas,  $G$  e  $a_t$  serão constituídas por medidas de similaridade calculadas conforme (3.6). O vetor  $a_t$  avaliará a similaridade entre  $x_t$  e os estados representativos na matriz  $D$ , e  $G$  conterà a similaridade entre os estados representativos contidos na matriz  $D$ , transformando o vetor de similaridades  $a_t$  em um conjunto de pesos para cada observação em  $D$  (Wegerich et al., 2003). Dessa forma, a estimativa linear ótima para o estado atual passa a ser calculada conforme expressão a seguir:

$$w_t = (D \circ D^T)^{-1}(D \circ x_t) = G^{-1}a_t, \quad (3.7)$$

$$w_t^o = \frac{w_t}{\|w_t\|_1}, \quad (3.8)$$

$$\hat{x}_t = D^T w_t^o, \quad (3.9)$$

em que  $G = D \circ D^T$ ,  $\{G \in \mathbb{R}^{k \times k} \mid 0 \leq a_{ij} \leq 1\}$ , e  $a_t = D \circ x_t$ ,  $\{a_t \in \mathbb{R}^k \mid 0 \leq a_i \leq 1\}$ .

Além disso, para preservar as mesmas propriedades lineares da estimativa calculada por meio de (3.5), é importante que o operador de similaridade escolhido tenha idealmente as seguintes propriedades (Singer et al., 1997):

- A matriz  $G$  deve ser não singular;
- Se algum(ns) elemento(s) do vetor do estado atual não estiver dentro dos limites dos estados na matriz memória, a estimativa ainda deve representar uma estimativa ótima;
- Se o estado atual é idêntico a algum estado em  $D$ , a estimativa ótima deve ser idêntica ao estado atual;
- O erro deve ser minimizado.

Tendo obtido  $\hat{x}_t$  por meio de (3.9), o segundo e último passo para avaliar se o processo está operando em situação anormal consiste em avaliar a similaridade entre  $x_t$  e  $\hat{x}_t$ . Se a similaridade resultante for maior ou igual a um limiar  $\tau$ ,  $\{\tau \in \mathbb{R} \mid 0 < \tau < 1\}$ , definido *a priori*,

$$s(x_t, \hat{x}_t) \geq \tau, \quad (3.10)$$

considera-se que o processo está operando normalmente, caso contrário, o processo está operando de forma anormal, cabendo ao supervisor investigar a(s) causa(s) do mau funcionamento.

Para um melhor entendimento, considere um processo qualquer monitorado pelo nível de pressão e temperatura, conforme apresentado na Figura 3.1a. Na figura, as linhas tracejadas em vermelho na horizontal determinam os níveis máximos de temperatura e pressão permitidos no processo, valores acima desse limiar configuram uma falha de operação. A região de operação normal do processo é representada pelos estados contidos na matriz memória, apresentados na Figura 3.1b. Ao ser apresentado um novo estado,  $x_t$ , a estimativa é calculada,  $\hat{x}_t$ , para então avaliar a similaridade entre o novo estado e essa estimativa, como ilustrado na Figura 3.2. Esse procedimento é repetido para toda entrada dada.

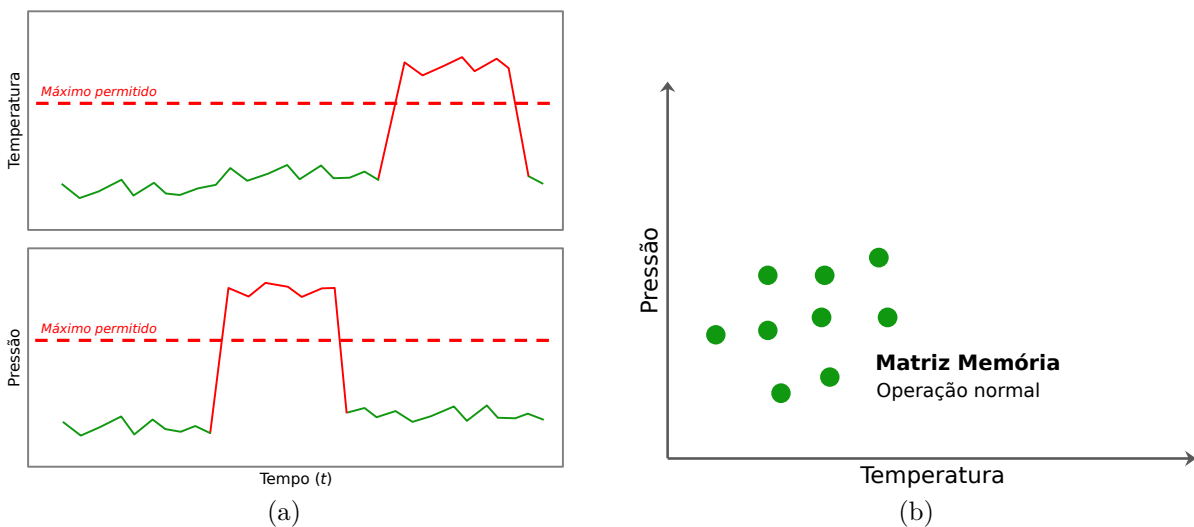


Figura 3.1 – Processo genérico monitorado pelas leituras do sensor de temperatura e pressão com ocorrência de falha de alta temperatura e alta pressão (a) e estados contidos na matriz memória para representação da operação normal (b).

### 3.1.1 Operadores de similaridade

Os operadores de similaridade são funções que usam como métrica a distância entre os vetores; as distâncias mais utilizadas são as pertencentes a família  $l_p$  norma (Golub

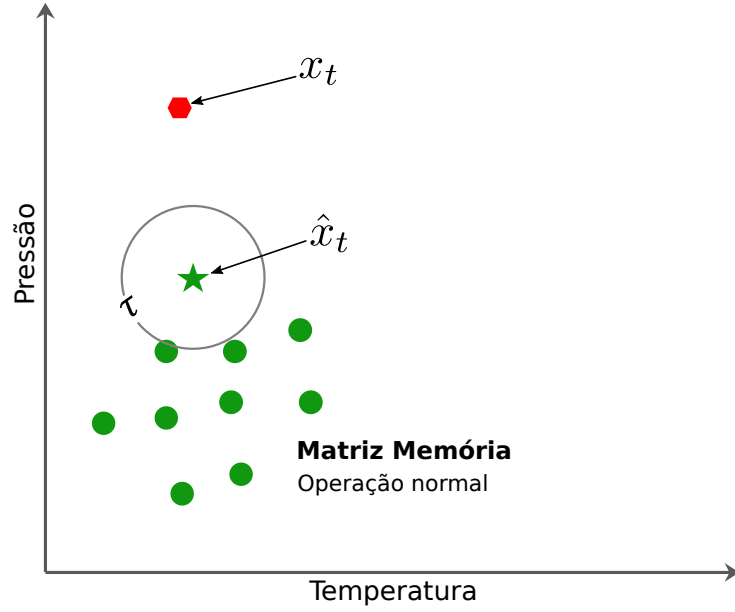


Figura 3.2 – Procedimento de inferência realizado pela SBM. Se o estado atual e estimativa não estiverem dentro da região delimitada por  $\tau$  considera-se que esse estado representa um comportamento anômalo.

and Van Loan, 1996), com maior destaque a Euclidiana e Manhattan, definidas quando  $p = 2$  e  $p = 1$ , respectivamente. A forma geral da norma  $l_p$  é definida conforme expressão a seguir:

$$d_{ij} = d(x_i, x_j) = \|x_i - x_j\|_p = \left( \sum_{l=1}^n |x_{il} - x_{jl}|^p \right)^{\frac{1}{p}}, \quad (3.11)$$

e os operadores mais comuns são (Marins et al., 2018):

- *Inverse Multiquadric Function (IMK)*

$$s_{imk}(d_{ij}) = \frac{1}{\sqrt{1 + \alpha^2 d_{ij}^2}}. \quad (3.12)$$

- *Cauchy Kernel (CCK)*

$$s_{cck}(d_{ij}) = s_{imk}^2(d_{ij}) = \frac{1}{1 + \alpha^2 d_{ij}^2}. \quad (3.13)$$

- *Wegerich Similarity Function (WSF)*

$$s_{wsf}(d_{ij}) = \frac{1}{1 + \alpha d_{ij}}. \quad (3.14)$$

- *Laplacian Kernel (LK)*

$$s_{lk}(d_{ij}) = e^{-\alpha d_{ij}}. \quad (3.15)$$

- *Radial Basis Function Kernel (RBF)*

$$s_{rbf}(d_{ij}) = e^{-\alpha d_{ij}^2}. \quad (3.16)$$

- *Saturated Triangular Operator (STO)*

$$s_{sto}(d_{ij}) = \begin{cases} \delta - d_{ij}, & d_{ij} \leq \delta - \epsilon. \\ \epsilon, & c.c. \end{cases} \quad (3.17)$$

O parâmetro  $\alpha \in \mathbb{R}$  nos cinco primeiro operadores controla o decaimento da função e deve ser definido de forma a maximizar a qualidade da detecção, o mesmo critério deve ser utilizado para definir os valores de  $\delta \in \mathbb{R}$  e  $\epsilon \in \mathbb{R}$  no operador STO. Nesse último, também deve-se considerar que o valor de  $\epsilon$  deve ser pequeno e positivo para garantir que  $s(x_i, x_j) > 0$  em situações que a amostra ultrapasse levemente a distância máxima.

## 3.2 Classificador de modos operacionais

Embora seja razoavelmente simples adaptar a SBM apresentada anteriormente para classificação envolvendo muitas classes, foi apenas [Marins et al. \(2018\)](#) que propuseram tal formalização. Nessa abordagem, denominada SBM-Multiclasse, cada modo operacional será visto como uma classe e cada uma terá sua própria matriz memória,

$$D_i^{k \times n} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & \dots & x_{kn} \end{bmatrix}_i, \quad i = 1, 2, \dots, c, \quad (3.18)$$

em que  $c$  indica a quantidade de classes ou modos operacionais. Utilizando o processo genérico apresentado anteriormente na Figura 3.1, ter-se-á uma representação como a apresentada na Figura 3.3, em que cada modo operacional, normal e falhas, terá sua região resumida por pontos representativos que por sua vez irão compor a respectiva matriz memória.

Com isso, para classificar uma nova amostra, será gerada uma estimativa por meio de (3.9) em função de cada classe,

$$\hat{x}_{ti} = D_i^T w_{ti}^o, \quad \forall i \in \{1, 2, \dots, c\}, \quad (3.19)$$

associa-se a entrada com a classe que apresentar melhor estimativa, identificada por  $c^*$ ,

$$c^* = \operatorname{argmax}_{i \in \{1, 2, \dots, c\}} s(x_t, \hat{x}_{ti}), \quad (3.20)$$

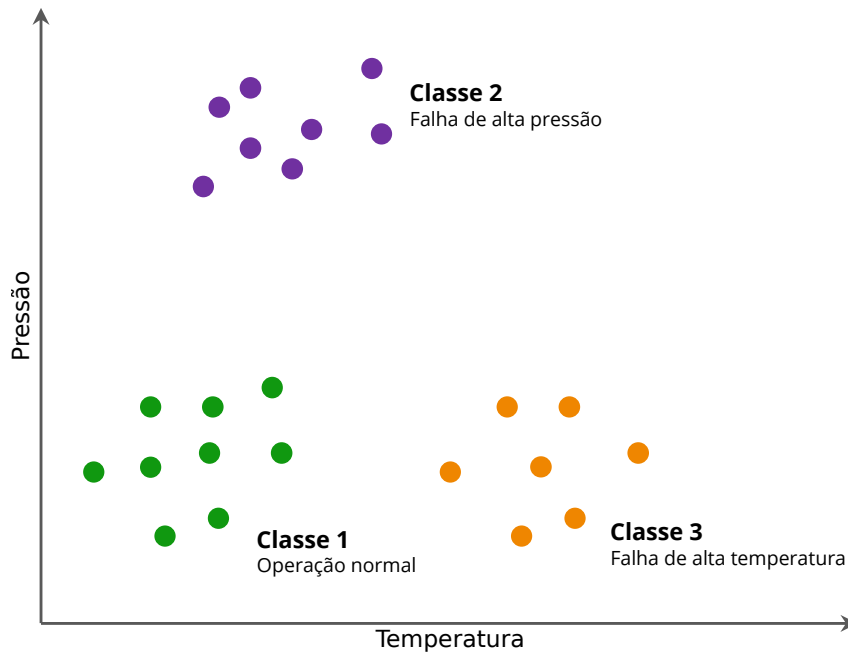


Figura 3.3 – Representação de cada modo operacional do processo na SBM-Multiclasse.

como ilustrado na Figura 3.4.

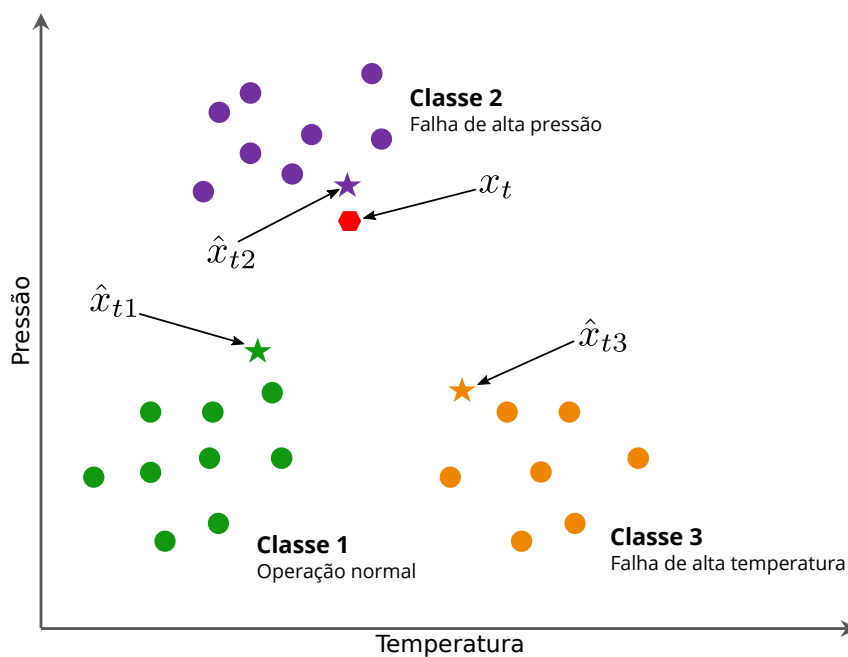


Figura 3.4 – Procedimento de inferência realizado pela SBM-Multiclasse. O estado atual é associado a classe que melhor aproximá-lo, nesse exemplo, seria associado a Classe 2 (Falha de Alta Pressão).

### 3.2.1 Composição da matriz memória

Considerando a necessidade de modelos parcimoniosos, isto é, modelos que equilibram o custo computacional com a qualidade da modelagem, [Marins et al. \(2018\)](#) também

propuseram um método iterativo para remover amostras redundantes da matriz memória. No método proposto considera-se um conjunto de vetores inicial,  $\Upsilon_i$ , pertencente a  $i$ -ésima classe, a partir desse conjunto seleciona-se a primeira amostra representativa como a mediana geométrica,  $v_i$ . Para obtê-la é necessário solucionar o seguinte problema de otimização:

$$v_i = \operatorname{argmin}_{z \in \Upsilon_i} \sum_{x \in \Upsilon_i} \|x - z\|_2, \quad (3.21)$$

ou utilizar a aproximação proposta em (Cardot et al., 2013), que embora também seja necessário solucionar um modelo de otimização, reduz significativamente o custo computacional.

As próximas amostras em  $\Upsilon_i$  serão selecionadas comparando-as entre si. Inicia-se removendo todas as amostras com similaridade maior que um limiar  $\zeta$  com a mediana,  $\{\zeta \in \mathbb{R} \mid 0 < \zeta < 1\}$ . Após, seleciona-se uma amostra aleatoriamente dentre as restantes e novamente são removidas as que tiverem similaridade maior que  $\zeta$ , a selecionada também é removida. O procedimento é repetido até não haver mais amostras em  $\Upsilon_i$ . Resumidamente, uma amostra  $x_i \in \Upsilon_i$  é incorporada à matriz memória da  $i$ -ésima classe,  $D_i$ , se

$$s(x_i, x_j) \leq \zeta, \quad \forall x_j \in D_i. \quad (3.22)$$

### 3.3 Resumo do capítulo

Nesse capítulo foi apresentada a técnica de Modelagem Baseada em Similaridade de Modelos sob as duas perspectivas mais comumente utilizadas na literatura, como um detector de anomalias e como um classificador multiclasse. A mesma tem sido empregada com sucesso nos mais diversos tipos de problemas. Tal amplitude de uso se dá, em grande parte, por sua simplicidade, pois sua dinâmica de inferência se resume apenas a cálculos de aproximações lineares a partir de conjuntos de pontos representativos e comparações de similaridades.



## Capítulo 4

# Clusterização Evolutiva Baseada na SBM

Como visto no capítulo anterior, a SBM tem sido utilizada com sucesso nas mais diversas áreas. No entanto, seu uso se restringe a aplicações em que todas as classes de interesse são conhecidas previamente, tornando-a inviável para uso na grande maioria dos processos industriais modernos, em que a quantidade de modos operacionais é desconhecida, geralmente pela complexidade intrínseca aos processos. Na tentativa de torná-la viável para uso nesse contexto, [Mariano and Caminhas \(2019\)](#) propuseram a SBM Evolutiva (eSBM) (do inglês, *Evolving Similarity-Based Modeling*), com a capacidade de iniciar sem qualquer informação dos modos de operação, construindo-a de forma autônoma e incremental.

Dito isso, este capítulo será focado na apresentação do eSBM, em seguida, também será apresentado algumas propostas de melhorias, que objetivam tornar o eSBM mais intuitivo no que diz respeito a definição de valores para os parâmetros e menos dependente de conhecimento técnico do processo.

### 4.1 SBM evolutiva

Conceitualmente, a SBM-Multiclasse é um método adaptativo, pois apenas seus parâmetros são ajustados, mantendo estrutura fixa. O eSBM, por outro lado, é um método evolutivo, ajustando tanto seus parâmetros quanto sua estrutura de forma on-line, ou seja, os ajustes, aqui chamado de aprendizagem, ocorrem processando cada entrada uma única vez. Essa abordagem também é comumente chamada de aprendizagem incremental ou recursiva.

Ao contrário da SBM-Multiclasse, em que a aprendizagem ocorre de forma supervisionada, no eSBM é realizado de maneira não supervisionada, em função disso, o termo "classe" será substituído por "grupo", uma vez que a informação de rotulagem não é utilizada previamente ou até mesma necessária.

O eSBM é caracterizado em três fases: Estimação, Classificação e Adaptação. As duas primeiras fases correspondem a SBM-Multiclasse. Na fase de Estimação é gerada as estimativas para a entrada considerando cada grupo existente. Na classificação é fornecido o grupo a qual a entrada melhor relaciona-se, e na Adaptação são realizadas atualizações na estrutura atual do método, quando necessário.

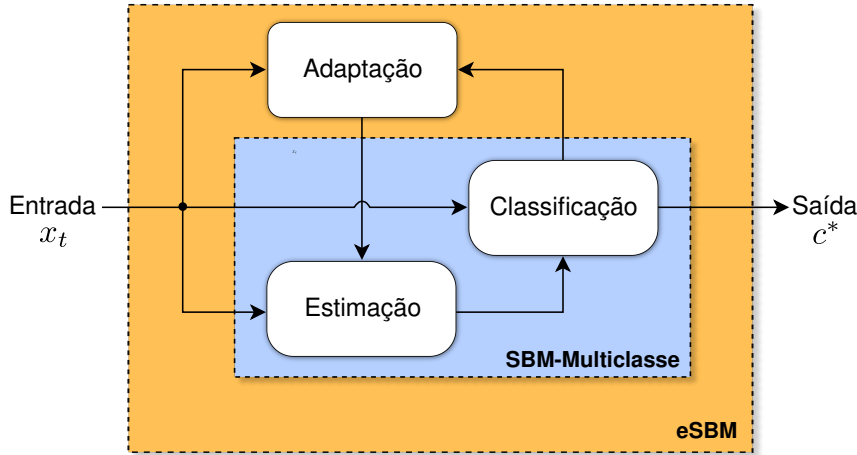


Figura 4.1 – Estrutura geral da eSBM.

Fonte: Adaptado de [Mariano and Caminhas \(2019\)](#).

Os autores realizaram uma pequena modificação na Classificação, o limiar mínimo de similaridade  $\tau$  antes utilizado na detecção pela SBM, passa a ser utilizado para indicar quando uma amostra deve ser considerada na atualização ou criação de grupos. Quando a similaridade entre a estimativa do grupo  $c^*$  e a entrada for maior ou igual a  $\tau$ , considera-se que o processo está em um modo operacional conhecido, caso contrário, considera-se desconhecido. Em ambas as situações é disparada uma *flag* para a etapa de Adaptação que decidirá se deve atualizar ou criar um grupo,

$$flag = \begin{cases} c^*, & s(x_t, \hat{x}_{tc^*}) \geq \tau. \\ desconhecido, & c.c. \end{cases} \quad (4.1)$$

Na Adaptação a *flag* será utilizada em um sistema de janelas com tamanho  $win_{adapt} \in \mathbb{N}^*$  para a atualização e  $win_{novo} \in \mathbb{N}^*$  para a criação. Quando o valor da *flag* for "desconhecido", a amostra que disparou a *flag* é adicionada em uma janela de dados  $W$ , quando atingir  $win_{novo}$  amostras é avaliado se pelo menos a quantidade de  $win_{novo}^{min} \in \mathbb{N}^*$  foram classificadas como desconhecidas, se sim, essas amostras são utilizadas para criar o novo grupo, se não, a janela é descartada e uma nova é iniciada. A composição da matriz memória é realizada utilizando o método descrito na Seção 3.2 com uma pequena

modificação. Além de avaliar a similaridade entre as amostras também é avaliada a similaridade com a mediana geométrica,

$$s(x_i, x_p) \leq \zeta_1 \text{ e } s(v_{c+1}, x_i) \leq \zeta_2, \quad \forall x_i \in W, \forall x_p \in D_{c+1}, \quad (4.2)$$

em que  $\zeta_1$  e  $\zeta_2$  são os limiares de similaridade,  $c + 1$  identifica o novo grupo. Os autores argumentam que essa avaliação é necessária para evitar sobreposições entre as regiões dos grupos e incorporação de *outliers* na matriz memória.

Quando o valor da *flag* for  $c^*$  a atualização é iniciada para esse grupo. A amostra que disparou a *flag* é incluída em uma janela de dados  $M$ , mas apenas se a similaridade entre a estimativa estiver dentro de um intervalo,  $[p_{min}, p_{max}]$ , tal que  $0 < p_{min}, p_{max} < 1$ ; também é necessário que a similaridade entre os demais grupos seja menor que  $\tau$ . Quando a quantidade de amostras em  $M$  atingir  $win_{adapt}^{min} \in \mathbb{N}^*$  é aplicado o método de seleção de amostras representativas nessa janela, o resultado é incorporado na matriz memória atual e reaplica-se o método.

É importante ressaltar que o primeiro grupo é criado com as  $win_{novo}$  amostras.

## 4.2 Modificações propostas

A parametrização do eSBM definirá a qualidade do agrupamento, demandando conhecimento prévio sobre a dinâmica do processo alvo para fazê-la. Por consequência, será única para cada aplicação, ou seja, se configurada para um determinado processo, tal configuração será útil em outro processo apenas se houver o mesmo número de variáveis e com a mesma dinâmica de funcionamento (*i.e.*, no que diz respeito a média e variância de cada variável), do contrário, o desempenho do método será completamente diferente da primeira aplicação.

Nesta seção serão apresentadas algumas propostas de mudanças ao eSBM, objetivando facilitar a parametrização e redução do custo computacional. A versão com esse conjunto de mudanças foi dado o nome de eSBM+. As modificações contemplaram todas as fases, mas manteve-se a essência original do método. Vale ressaltar que dada a natureza do método, a fase de Adaptação aqui é chamada de Evolução.

### 4.2.1 Operação de Similaridade

Na maioria das abordagens com a SBM encontradas na literatura são utilizadas como métrica no cálculo da similaridade a distância Euclidiana ou Manhattan, porém, ambas são dependentes da escala dos dados, gerando a incapacidade de manter o desempenho do método similar em diferentes aplicações com a mesma parametrização.

Como uma forma de contornar esse problema, propõe-se o uso da distância de Mahalanobis (De Maesschalck et al., 2000), uma distância estatística invariante de escala,

$$d_{ij} = d(x_i, x_j) = \sqrt{(x_i - x_j)^T S^{-1} (x_i - x_j)}, \quad (4.3)$$

em que  $S^{-1}$  é a inversa da matriz de covariância amostral, podendo ser calculada incrementalmente conforme as expressões a seguir (Lughofer, 2011):

$$[S_i^{-1}]_{l_{i+1}} = \frac{[S_i^{-1}]_{l_i}}{1 - \beta} - \frac{\beta}{(1 - \beta)} \times \left[ \frac{1}{1 + \beta} \times \frac{[[S_i^{-1}]_{l_i}(x_{l_{i+1}} - [\bar{X}_i]_{l_i})][[S_i^{-1}]_{l_i}(x_{l_{i+1}} - [\bar{X}_i]_{l_i})]^T}{(x_{l_{i+1}} - [\bar{X}_i]_{l_i})^T [S_i^{-1}]_{l_i} (x_{l_{i+1}} - [\bar{X}_i]_{l_i})} \right], \quad [S_i^{-1}]_0 = I_n, \quad (4.4)$$

$$[\bar{X}_i]_{l_{i+1}} = \frac{l_i [\bar{X}_i]_{l_i} + x_{l_{i+1}}}{l_i + 1}, \quad [\bar{X}_i]_0 = x_0, \quad (4.5)$$

com  $\bar{X}$  sendo a média aritmética amostral,  $i = 1, 2, \dots, c$ ,  $l_i$  é a quantidade de amostras associadas ao  $i$ -ésimo grupo no instante  $t$ ,  $\beta = 1/(l_i + 1)$  e  $I_n$  é uma matriz identidade de ordem  $n$ .

O uso dessa distância possibilita a definição de limiares estatísticos, que por sua vez definem uma região com formato elipsoidal centrada em  $\bar{X}$  que conterá  $100\gamma\%$  dos dados, baseando-se na premissa de que os dados seguem uma distribuição gaussiana multivariada, mesmo que não sigam, ainda será uma aproximação razoável para qualquer distribuição unimodal (Tax and Duin, 2000):

$$D_{max} = \sqrt{\chi_{n,\gamma}^2}, \quad (4.6)$$

em que  $D_{max}$  é o limiar máximo de distância,  $\chi_{n,\gamma}^2$  é a distribuição Qui-quadrado (Abramowitz and Stegun, 1948; Wasserman, 2013) com  $n$  graus de liberdade e  $\gamma$ ,  $\{\gamma \in \mathbb{R} \mid 0 < \gamma < 1\}$ , o percentil desejado, sendo comumente utilizado o valor de 0.99.

Dessa forma, os operadores de similaridade apresentados na Subseção 3.1.1 podem ser ajustados conforme o limiar de distância em (4.6) e o de similaridade  $\tau$ . Para os operadores não lineares com o parâmetro  $\alpha$ , faz-se:

$$s(D_{max}) = \tau, \quad (4.7)$$

e desenvolve-se a igualdade isolando  $\alpha$ . A expressão resultante fornece o valor de  $\alpha$  que ajusta a curva da função de modo a fazer a igualdade ser verdadeira, indicando que

quando a distância for maior que  $D_{max}$  a similaridade será menor que  $\tau$ , por consequência, subtende-se que o processo está operando fora da região conhecida.

A título de exemplo, considere o operador LK:

$$\begin{aligned} s_{lk}(D_{max}) &= \tau \\ e^{-\alpha D_{max}} &= \tau \\ -\alpha D_{max} &= \ln(\tau) \\ \alpha &= -\frac{\ln(\tau)}{D_{max}}, \end{aligned} \tag{4.8}$$

o parâmetro  $\alpha$  do operador de similaridade pode ser definido diretamente por (4.8) em função dos limiares desejados, necessitando apenas de um pouco de conhecimento sobre intervalos estatísticos para interpretar o resultado da ação tomada.

Para o operador linear STO, a igualdade estabelece-se definindo  $\delta = D_{max}$  e  $\epsilon = \tau$ .

## 4.2.2 Estimação

Na estimação é definida uma região chamada de Região de Grupo para ser utilizada como referência no cálculo das operações de similaridade em (3.9),

$$D_{group} = \sqrt{\chi_{n, \gamma_{group}}^2}. \tag{4.9}$$

Embora não haja validade estatística, pois o centro não será mais  $\bar{X}$ , ainda é fornecida uma métrica razoável de distância. Na Figura 4.2 é ilustrada a abordagem simulando o cálculo da matriz  $G$ , cada ponto da matriz memória será tomado como o centro da região para calcular a similaridade entre os demais utilizando (4.7). O cálculo do vetor  $a_t$  ocorre da mesma forma.

Em testes experimentais,  $0.8 \leq \gamma_{group} \leq 0.999$ , demonstrou-se ser um intervalo razoável para a definição do tamanho da região de grupo.

## 4.2.3 Classificação

Na Classificação propõe-se utilizar apenas a média do erro relativo entre a entrada e a respectiva estimativa em cada variável,

$$[E_t]_i = \frac{1}{n} \sum_{j=1}^n \left[ \frac{|[x_t]_j - [\hat{x}_t]_j|}{|[x_t]_j|} \right]_i, \quad i = 1, 2, \dots, c, \tag{4.10}$$

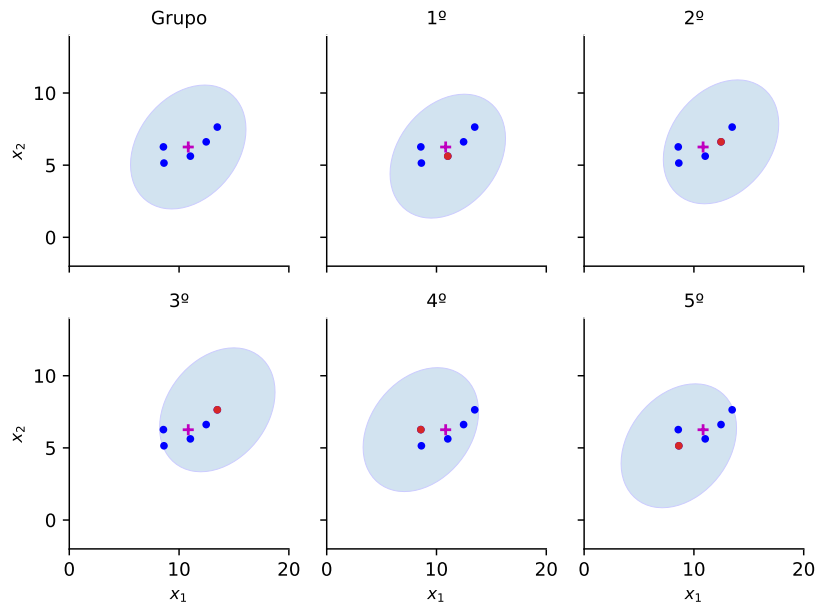


Figura 4.2 – Ilustração do cálculo da matriz  $G$ , o símbolo  $+$  em vermelho indica a média ou centro do grupo,  $\bar{X}$ . A região de grupo (elipse em cinza) é deslocada fazendo cada ponto da matriz memória como centro (ponto circular em vermelho) para então ser efetuado o cálculo da similaridade entre os demais.

em que  $x_t$  será associado ao grupo que apresentar menor erro relativo médio,

$$c^* = \underset{i \in \{1, 2, \dots, c\}}{\operatorname{argmin}} [E_t]_i, \quad (4.11)$$

e de forma similar ao eSBM, será enviada uma *flag* para a fase de Evolução indicando o que deve ser feito, atualizar ou criar grupos, para isso é definido o parâmetro  $\theta$ ,  $\{\theta \in \mathbb{R} \mid 0 < \theta < 1\}$ , responsável por indicar o percentual de resíduo tolerável na estimativa, se o erro for menor ou igual a  $\theta$  considera-se que o processo está operando em um modo operacional conhecido, do contrário, é indicada a Evolução a necessidade de criação de um novo grupo,

$$flag = \begin{cases} c^*, & [E_t]_{c^*} \leq \theta \\ desconhecido, & c.c. \end{cases} \quad (4.12)$$

O parâmetro  $\theta$  será um dos principais meios para o controle da sensibilidade do método. Em testes experimentais, o intervalo  $0.05 \leq \theta \leq 0.1$  mostrou-se razoável.

## 4.2.4 Evolução

### Criação de grupos

De forma similar a eSBM, também utiliza-se uma janela de dados  $W$ , com tamanho  $k$ , porém, essa janela define uma quantidade mínima de amostras que devem ser classificadas como *desconhecidas* de forma consecutiva. Se for indicado o grupo  $c^*$  durante a formação da janela, descarta-se a mesma e uma nova será iniciada na próxima classificação *desconhecida*. Além da definição de uma quantidade mínima de pontos necessários para uma estimativa razoável, seu uso também tem o intuito de evitar criação de grupos por *outliers*. O primeiro grupo é criado com as  $k$  primeiras amostras apresentadas ao método.

### Atualização de grupos

Nessa versão descartou-se o uso de janelas na atualização. A cada nova amostra associada ao grupo  $c^*$ , é avaliado se a mesma encontra-se dentro da região de grupo  $c^*$ , centrada em  $\bar{X}_{c^*}$ , e fora da região de grupo dos demais, também centrada em seus respectivos centros,

$$d_{x_t, \bar{X}_i} > D_{group} \leq d_{x_t, \bar{X}_{c^*}} \quad \forall i \neq c^*, \quad (4.13)$$

se a condição anterior for satisfeita, utiliza-se (4.5) e (4.4) para atualizar o centro e a matriz de covariância, respectivamente. No que diz respeito a atualização da matriz memória, além da condição anterior, também é necessário que a nova amostra esteja fora de uma nova região, chamada de Região Pontual,

$$D_{point} = \sqrt{\chi_{n, \gamma_{point}}^2}, \quad (4.14)$$

tal região define uma área mínima entorno de cada amostra da matriz memória que não pode haver outra,

$$d_{x_t, x_p} \geq D_{point} \quad \forall x_p \in D_{c^*}, \quad (4.15)$$

evitando a adição de amostras que estejam dentro da região de grupo de outros grupos e também de amostras redundantes. Em testes experimentais, o intervalo  $0.05 \leq \gamma_{point} \leq 0.2$  mostrou-se razoável. A Figura 4.3 ilustra as duas condições necessárias.

### União de grupos

Visando a remoção de grupos redundantes, também é proposto uma abordagem para união de grupos. Se a distância entre os centros de dois grupos,  $i$  e  $j$ , em pelo menos

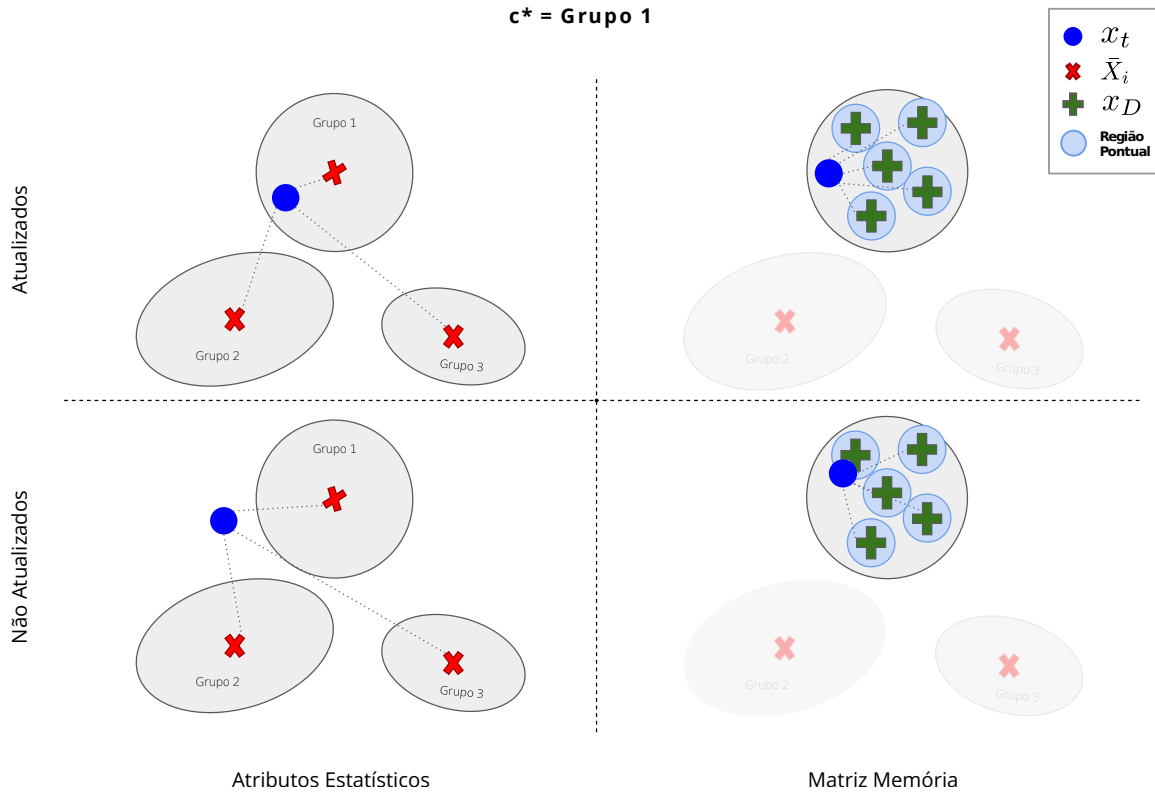


Figura 4.3 – Ilustração do procedimento de atualização de grupos. Na primeira linha é apresentada a situação em que as duas condições são satisfeitas, e tanto os atributos estatísticos quanto a matriz memória são atualizados. Na segunda linha a situação em que nenhuma das condições são satisfeitas, não ocorrendo nenhuma atualização no respectivo grupo.

uma das direções, for menor ou igual a metade de  $D_{group}$ ,

$$d_{\bar{X}_i \bar{X}_j} \leq \frac{D_{group}}{2} \vee d_{\bar{X}_j \bar{X}_i} \leq \frac{D_{group}}{2}, \quad (4.16)$$

considera-se que os dois grupos são redundantes e o procedimento de união é executado.

Como as restrições de atualização de grupo garantem a ausência de amostras representativas redundantes e dentro da região de outros grupos, a matriz memória do grupo resultante é apenas a união dos grupos alvo. Para as variáveis estatísticas, centro e matriz de covariância, utiliza-se o procedimento proposto em (Kelly, 1994), ao qual é resumido nas expressões a seguir:

$$l_i = l_i + l_j, \quad (4.17)$$

$$\bar{X}_i = \frac{l_i}{l_i + l_j} \bar{X}_i + \frac{l_j}{l_i + l_j} \bar{X}_j, \quad (4.18)$$



$$S_i = \frac{l_i - 1}{l_i + l_j + 1} S_i + \frac{l_j - 1}{l_i + l_j + 1} S_j + \bar{X}_j \frac{l_i l_j}{l_i + l_j (l_i + l_j - 1)} (\bar{X}_i - \bar{X}_j)^T (\bar{X}_i - \bar{X}_j). \quad (4.19)$$

Uma ilustração do procedimento em função dos atributos estatísticos e da matriz memória é apresentado na Figura 4.4.

### 4.2.5 Composição da matriz memória

Propõe-se também um novo procedimento para a seleção das amostras representativas utilizando a Região Pontual definida anteriormente.

Dado um conjunto inicial de amostras  $\Upsilon_i$  referente ao  $i$ -ésimo grupo, seleciona-se como primeira amostra o centro,  $\bar{X}$ . O próximo passo é remover todas as amostras que estejam dentro da Região Pontual de  $\bar{X}$ , das amostras restantes, seleciona-se uma aleatoriamente e novamente são removidas as amostras dentro de sua Região Pontual. O procedimento é repetido até não haver amostras em  $\Upsilon_i$ . Uma ilustração do resultado do procedimento é fornecida na Figura 4.5 e um resumo no Algoritmo 1.

---

**Algoritmo 1:** Método proposto para composição da matriz memória.

---

**Input:**  $\Upsilon_i, D_{point}, \bar{X}_i$  e  $S_i^{-1}$ .

**Output:** Matriz memória,  $D_i$ .

```

1  $D_i \leftarrow \bar{X}_i$ .
2  $\Upsilon_i \leftarrow$  amostras com distância entre  $\bar{X}$  maior ou igual a  $D_{point}$ .
3 while  $\Upsilon_i \neq \emptyset$  do
4    $z \leftarrow$  selecione aleatoriamente uma amostra em  $\Upsilon_i$ .
5    $D_i \leftarrow D_i \cup z$ .
6    $\Upsilon_i \leftarrow$  amostras com distância entre  $z$  maior ou igual a  $D_{point}$ .
7 end

```

---

Por fim, um resumo do eSBM+ é fornecido no Algoritmo 2.

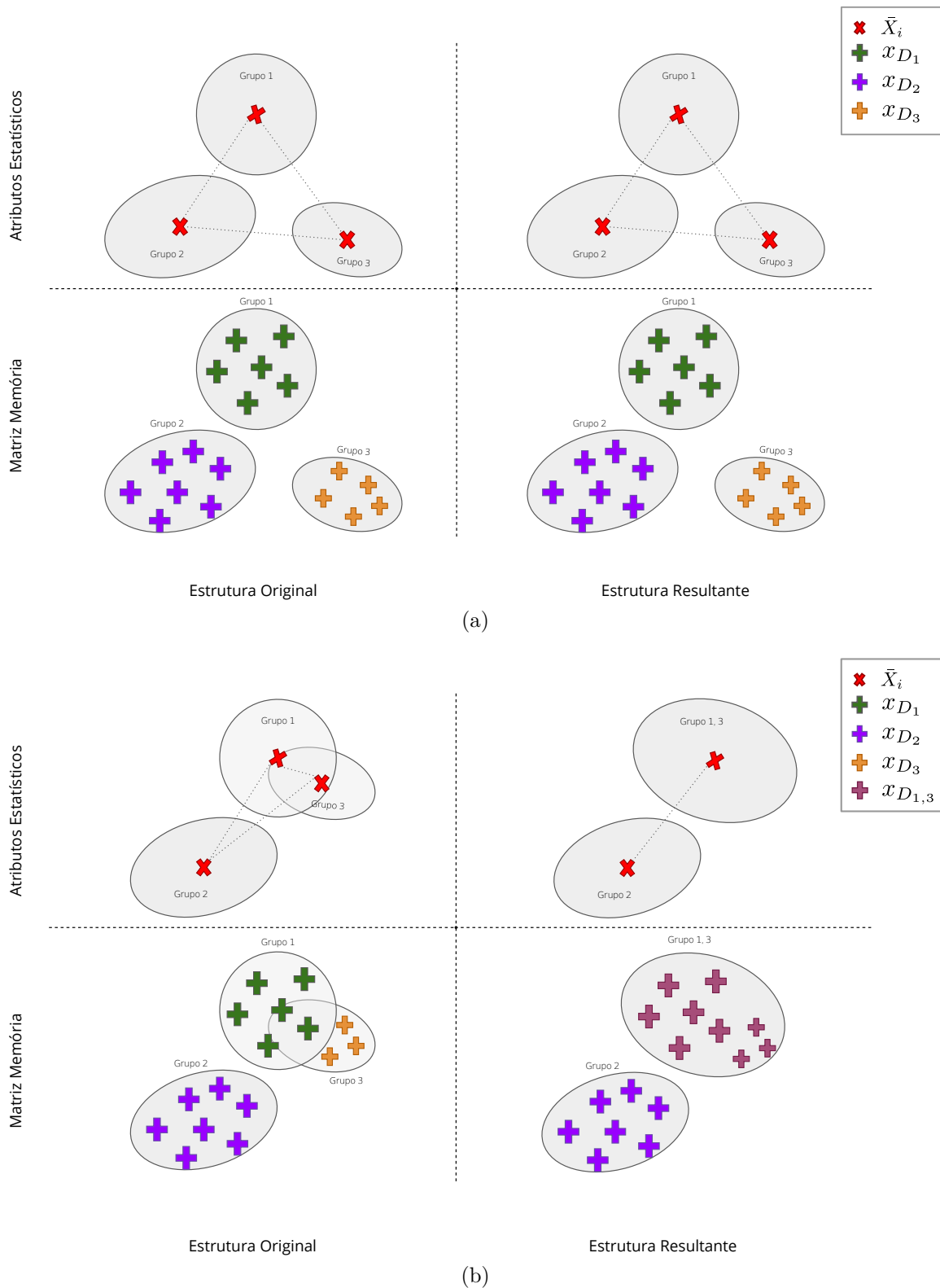


Figura 4.4 – Ilustração do procedimento de união de grupos. Quando as condições de união não são satisfeitas (a), mantendo a estrutura atual, e quando são satisfeitas (b), nesse exemplo para o Grupo 1 e Grupo 3, efetuando a união e resultando no Grupo 1, 3.

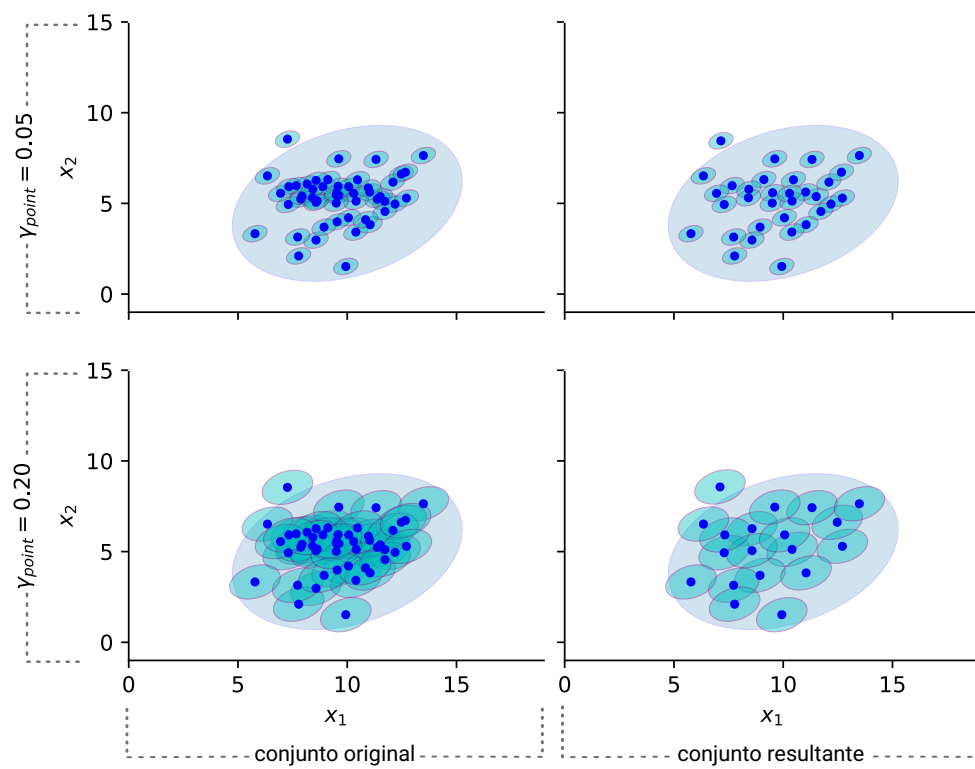


Figura 4.5 – Ilustração do resultado da abordagem proposta para composição da matriz memória utilizando os extremos do intervalo recomendado para  $\gamma_{point}$ .

---

**Algoritmo 2:** *Evolving Similarity-Based Modeling Plus.*


---

**Input:**  $x_t \mid t = 1, 2, \dots, +\infty$ .  
**Output:**  $c^*$ .

- 1 **Initialization**
- 2     defina  $\gamma_{group}, \gamma_{point}, \theta, k$  e  $\tau$ .
- 3     calcule  $D_{group}, D_{point}$  e  $\alpha$  para o operador escolhido.
- 4      $W \leftarrow \emptyset$ .
- 5      $c \leftarrow 0$ .
- 6 **end**
- 7 **if** *não existir grupo* **then**
- 8     Aguarde  $k$  amostras e crie o primeiro,  $criarGrupo(x_{1,\dots,k})$ .
- 9      $c \leftarrow c + 1$ .
- 10 **else**
- 11      $\hat{x}_{t,i} \leftarrow$  calcule a estimativa de cada grupo usando (3.9).
- 12      $c^* \leftarrow$  obtenha a classificação usando (4.11).
- 13     **if** *grupo é desconhecido* **then**
- 14          $W \leftarrow W \cup x_t$ .
- 15         **if** *W tem k amostras* **then**
- 16             crie um novo grupo,  $criarGrupo(W)$ .
- 17              $c \leftarrow c + 1$ .
- 18              $W \leftarrow \emptyset$ .
- 19         **end**
- 20     **else**
- 21          $W \leftarrow \emptyset$ .
- 22          $atualizarGrupo(x_t, c^*)$ .
- 23     **end**
- 24 **end**
- 25 **function**  $criarGrupo(\Upsilon)$
- 26      $\bar{X}_{c+1} \leftarrow$  Calcule a média pela forma comum.
- 27      $[S^{-1}]_{c+1} \leftarrow$  Calcule a matriz de covariância e dependendo do tamanho da janela calcule a inversa ou pseudoinversa.
- 28      $l_{c+1} \leftarrow k$ .
- 29      $D_{c+1} \leftarrow$  usando o Algoritmo 1.
- 30      $G_{c+1} \leftarrow$  calcule G.
- 31      $[G^{-1}]_{c+1} \leftarrow$  calcule a inversa de G.
- 32 **end**

---

---

```

33 function atualizarGrupo( $x_t, c^*$ )
34   if (4.13) for satisfeita then
35      $\bar{X}_{c^*} \leftarrow$  usando (4.5).
36      $[S^{-1}]_{c^*} \leftarrow$  usando (4.4).
37      $l_{c^*} \leftarrow l_{c^*} + 1.$ 
38     if (4.15) for satisfeita then
39        $s \leftarrow$  calc. a similaridade entre  $x_t$  e cada amostra em  $D_{c^*}.$ 
40        $G_{c^*} \leftarrow G_{c^*} \cup s.$ 
41        $D_{c^*} \leftarrow D_{c^*} \cup x_t.$ 
42        $[G^{-1}]_{c^*} \leftarrow$  recalcule a inversa de  $G_{c^*}.$ 
43     end
44     for todas as combinações de grupos i e j do
45       if (4.16) for satisfeita then
46          $\bar{X}_i \leftarrow$  usando (4.18).
47          $[S^{-1}]_i \leftarrow$  usando (4.19).
48          $l_i \leftarrow l_i + l_j.$ 
49          $D_i \leftarrow D_i \cup D_j.$ 
50          $G_i \leftarrow$  calcule G.
51          $[G^{-1}]_i \leftarrow$  calcule a inversa de  $G_i.$ 
52         end
53       end
54     end
55 end

```

---

## 4.2.6 Sensibilidade da parametrização

O eSBM+ possui um total de seis parâmetros:

- $\gamma_{group}$ : Define o tamanho da região de grupo, utilizada como referência no cálculo das operações de similaridades e também como delimitadora da região de ocupação das amostras da matriz memória. Essa região também influencia indiretamente na sensibilidade do método, um vez que com regiões de grupos menores, a tendência é que seja criada uma maior quantidade de grupos.
- $\gamma_{point}$ : Define o tamanho da Região de Pontual, que é a distância mínima que deve existir entre amostras da matriz memória no mesmo grupo.
- $k$ : Define a quantidade de amostras iniciais para a criação do primeiro grupo, bem como a quantidade de amostras consecutivas que devem ter valor de *flag* igual a *desconhecido*, utilizadas na criação de novos grupos.
- $\tau$ : Define o limiar mínimo de similaridade, amostras com distância entre o centro ou entre outras amostras maior que  $D_{group}$  terão similaridade menor que esse limiar.
- $\theta$ : Define o percentual,  $100\theta\%$ , de resíduo aceitável na estimativa, isto é, o percentual de resíduo entre a entrada e a melhor estimativa (*i.e.*, referente ao grupo  $c^*$ ) para considerar que a entrada pertence a um grupo existente.
- *Operador de Similaridade*: Define qual o operador será utilizado para computar a similaridade. Aqui está sendo considerado apenas os seis apresentados na Seção 3.1.1.

O restante dessa subseção busca discutir a influência de cada parâmetro citado na qualidade do agrupamento. Também é apresentada uma breve discussão sobre a característica de aleatoriedade do método de composição da matriz memória.

### 4.2.6.1 Região de grupo

Para o experimento envolvendo a região de grupo, considere o conjunto de dados sintéticos apresentado na Figura 4.6. Na figura é apresentada duas matriz memória obtidas com tamanhos diferentes da região de grupo,  $\gamma_{group} = 0.99$  e  $\gamma_{group} = 0.5$ , mas com o mesmo tamanho da região pontual,  $\gamma_{point} = 0.15$ . Para avaliar como o tamanho da região de grupo influencia no resultado final do eSBM+, optou-se por variar o tamanho da mesma (conforme Figura 4.6b e 4.6c) e verificar o peso dessa mudança em termos de qualidade da aproximação linear do estado atual. As duas diferentes matriz memória foram consideradas para avaliar se uma região de grupo que não esteja bem representada pela matriz memória também possui forte influência na aproximação, isto é, se aumentar ou diminuir seu tamanho sem modificar a matriz memória prejudica ou melhora a estimativa.

Para tanto, um estado foi selecionado aleatoriamente no conjunto inicial, o mesmo é apresentado na Figura 4.7. Na Figura 4.8 é apresentado o resultado obtido considerando os seis operadores de similaridades listados previamente, em que  $\tau = 0.001$ . Observa-se que os operadores RBF e STO sofrem forte influência do tamanho da região de grupo, de forma que quanto maior o tamanho dessa região, pior a aproximação. Os demais operadores demonstram que a qualidade da estimativa praticamente não sofre influência, sendo mais atrelada a matriz memória em si.

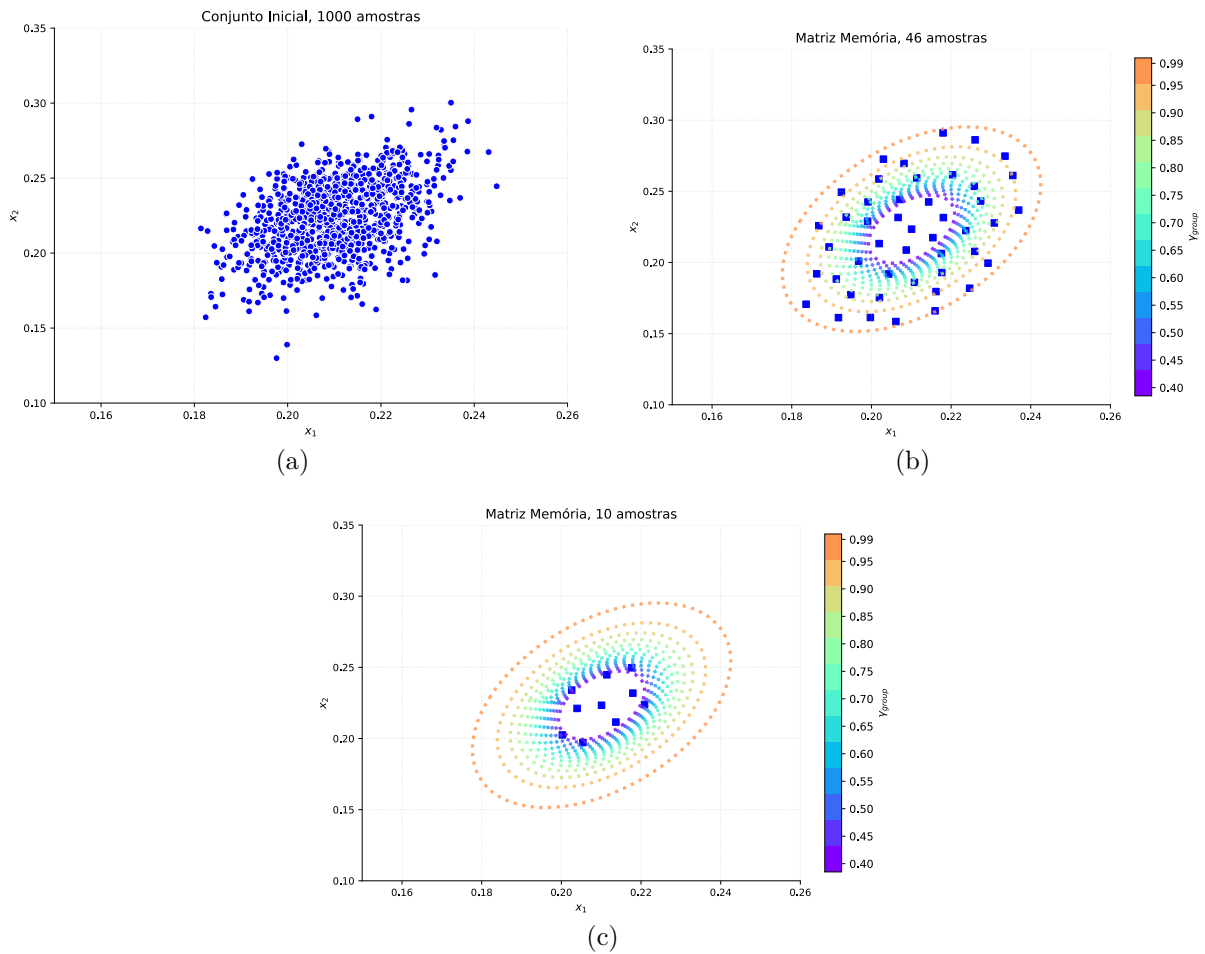


Figura 4.6 – Base de dados sintética bidimensional (a), conjunto de amostras representativas selecionadas com  $\gamma_{group} = 0.99$  e  $\gamma_{point} = 0.15$  (b) e com  $\gamma_{group} = 0.5$  e  $\gamma_{point} = 0.15$  (c). As elipses nas últimas figuras representam os diferentes tamanhos da região de grupo em função da variação de  $\gamma_{group}$ .

Dessa forma, é possível concluir que, desde que essa região esteja bem representada pela matriz memória, aumentá-la provocará uma melhora na qualidade da estimativa. Além disso, aumentá-la também diminuirá a sensibilidade do método, pois uma maior região será relacionada a um único grupo.

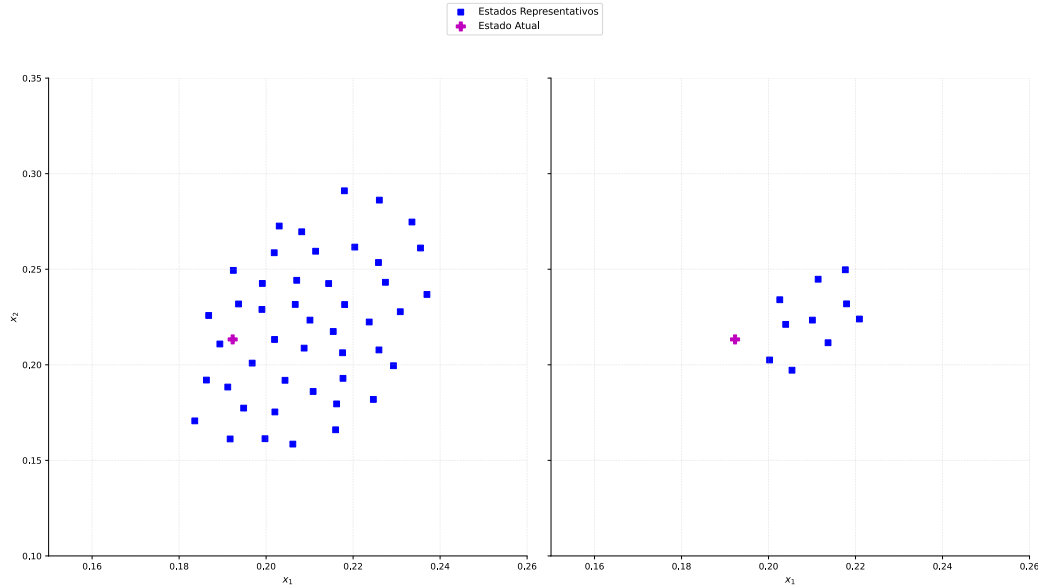


Figura 4.7 – Estado selecionado no conjunto inicial. As duas figuras apresentam o mesmo estado, optou-se por essa forma de apresentação para uma melhor perspectiva de distância do estado em relação as duas matriz memória.

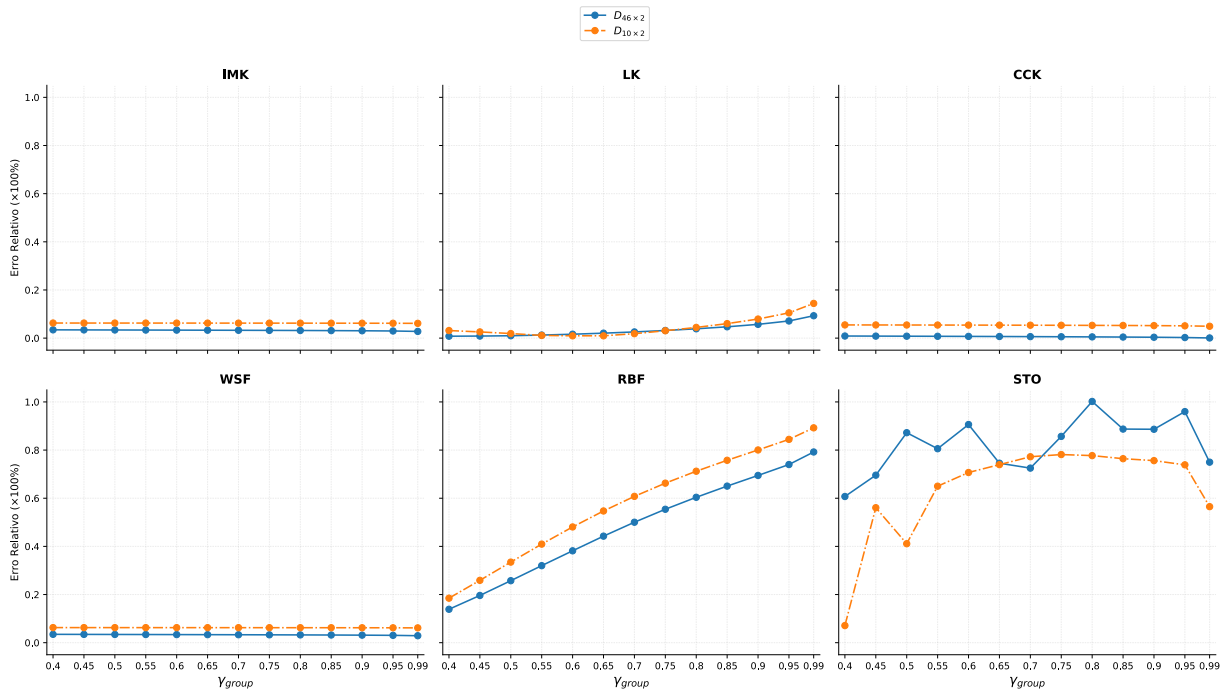


Figura 4.8 – Qualidade da estimativa em função do tamanho da região de grupo e do operador de similaridade para as duas matriz memória. O eixo  $y$  apresenta o erro relativo que trata-se da métrica utilizada na fase de Classificação.

#### 4.2.6.2 Região pontual

Para avaliar a influência da região pontual, optou-se por utilizar uma região de grupo maior e fixa,  $\gamma_{group} = 0.99$ , e variar o tamanho região pontual, ou seja, variar a quantidade de estados representativos na estimativa. O estado utilizado como referência na estimativa foi o mesmo do experimento anterior (Figura 4.7). O limiar mínimo de



similaridade foi definido em  $\tau = 0.001$ .

Na Figura 4.9 é apresentado o resultado obtido, sendo que na Figura 4.9a tem-se a quantidade de estados representativos selecionados pelo método de composição em função do tamanho da região pontual, e na Figura 4.9b o impacto das diferentes quantidades de estados representativos na estimativa do estado atual. A primeira figura indica, como esperado e desejado, que aumentar o tamanho dessa região reduz a quantidade de estados adicionados à matriz memória. A segunda, que desde que a região de grupo seja bem representada, para a maioria dos operadores uma quantidade menor de estados representativos traz mais ganhos à qualidade da estimativa, ou seja, valores maiores para  $\gamma_{point}$  são mais preferíveis. Na Figura 4.10 tem-se apresentado o tempo médio e a mediana de 100 execuções para o cálculo da estimativa em função da quantidade de estados na matriz memória. Como esperado, quanto menor o valor de  $\gamma_{point}$ , maior será o tempo de processamento no cálculo de uma estimativa, pois será realizada uma maior quantidade de operações lógicas e matemáticas.

#### 4.2.6.3 Janela de criação

O tamanho da janela de criação,  $k$ , influenciará tanto na qualidade da estimativa, uma vez que a mesma define a quantidade máxima de amostras que podem ser incluídas inicialmente na matriz memória de um grupo, como também na sensibilidade, pois essa quantidade de amostras refletirá no tamanho máximo da região de grupo.

Considere, por exemplo, a base dados sintética apresentada na Figura 4.11 com 6 grupos, levando em consideração que cada mudança gradual é um grupo. Para avaliar a influência de  $k$  no agrupamento, o mesmo foi variado de 5 até 205, e os demais parâmetros foram fixados em:  $\gamma_{group} = 0.9$ ,  $\gamma_{point} = 0.2$ ,  $\tau = 0.01$ ,  $\theta = 0.1$  e o operador de similaridade no WSF. É possível observar por meio da Figura 4.12 que esse parâmetro influencia diretamente no número final de grupos, de forma que quanto maior o tamanho dessa janela, menor a quantidade de grupos. Contudo, valores exagerados prejudicam consideravelmente a qualidade do agrupamento, como observado na Figura 4.13, em que é apresentado o resultado final do agrupamento para diferentes tamanhos da janela de criação.

Esse comportamento se dá em função do tamanho da região de grupo, uma vez que quanto maior a quantidade de amostras utilizadas na estimação da matriz de covariância, maior a chance da região que essa matriz representará englobar grupos distintos. Além disso, esse comportamento também é influenciado pelo valor de  $\theta$ .

#### 4.2.6.4 Limiar mínimo de similaridade

O limiar mínimo de similaridade,  $\tau$ , define o peso de referência de amostras que estejam exatamente na borda da região de grupo. Nesse sentido, valores menores, próximo de zero, são mais recomendáveis, do contrário, há grandes chances da estimativa sair da

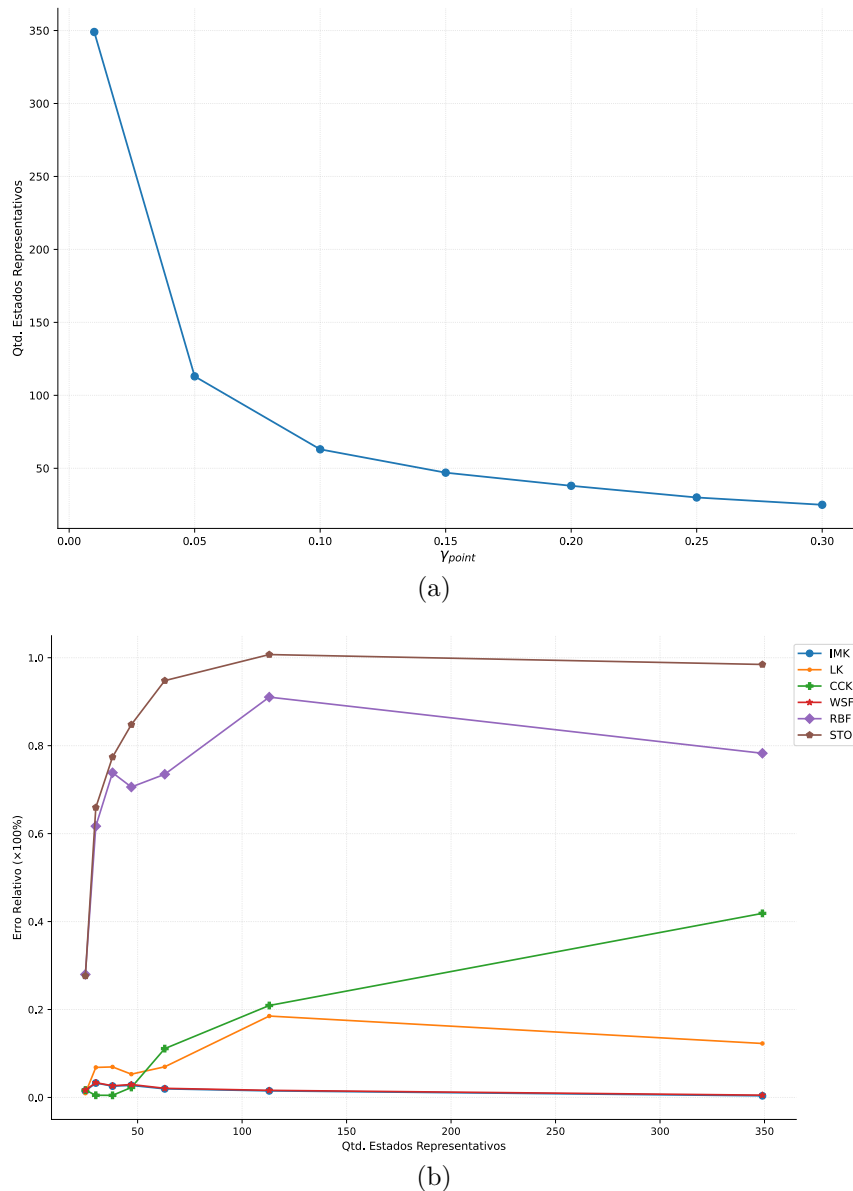


Figura 4.9 – Influência do tamanho da região pontual no eSBM+. Impacto na quantidade de estados adicionados à matriz memória (a) e na qualidade da estimativa (b).

região de grupo, pois o peso de referência não retrata a real proximidade. Essa afirmação é corroborada por meio dos resultados apresentados na Figura 4.14, em que percebe-se que uma melhor estimativa é gerada quando  $\tau$  assume valores menores. Esses resultados foram obtidos utilizando a matriz memória apresentada na Figura 4.6b, o estado atual da Figura 4.7 e  $\gamma_{group} = 0.99$ .

#### 4.2.6.5 Percentual de resíduo aceitável

O percentual de resíduo aceitável, como dito, define a quantidade (em percentual) de resíduo para considerar que uma estimativa é suficientemente boa para ser associada a um grupo. É o principal parâmetro para controlar a sensibilidade do método. Quanto maior

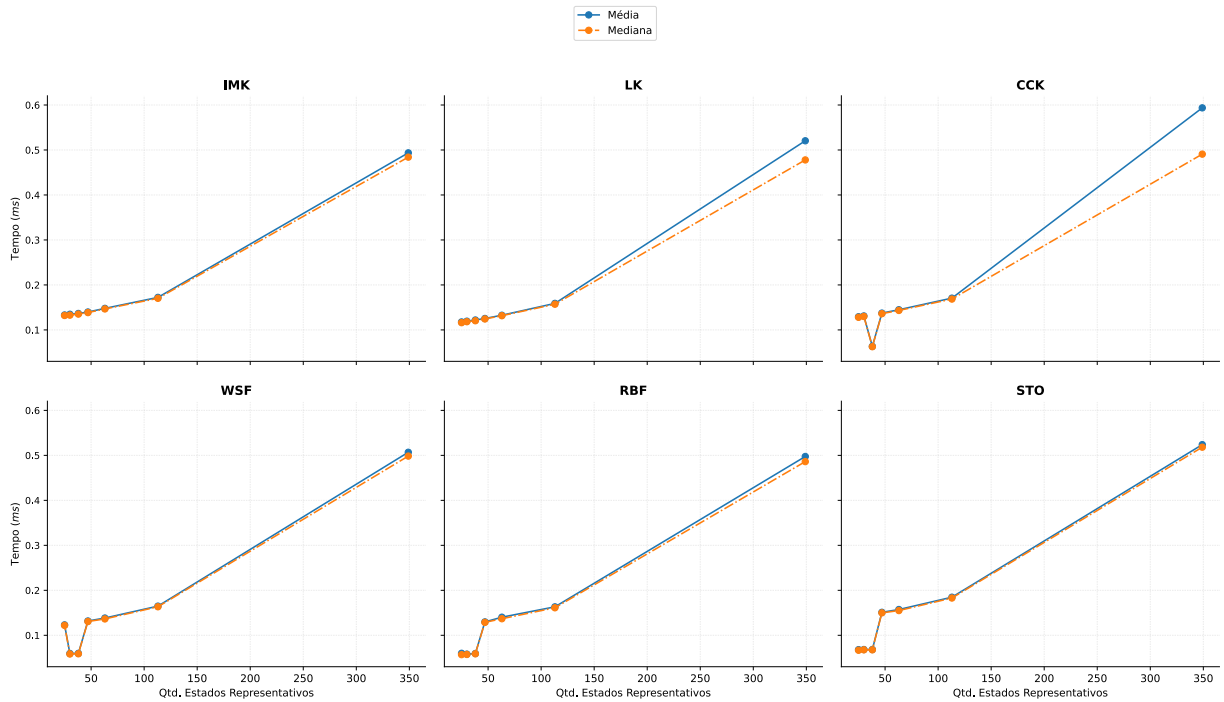


Figura 4.10 – Tempo médio e mediana (em milissegundos) de 100 execuções no cálculo de uma estimativa pelo eSBM+ em função da quantidade de estados representativos para cada operador de similaridade.

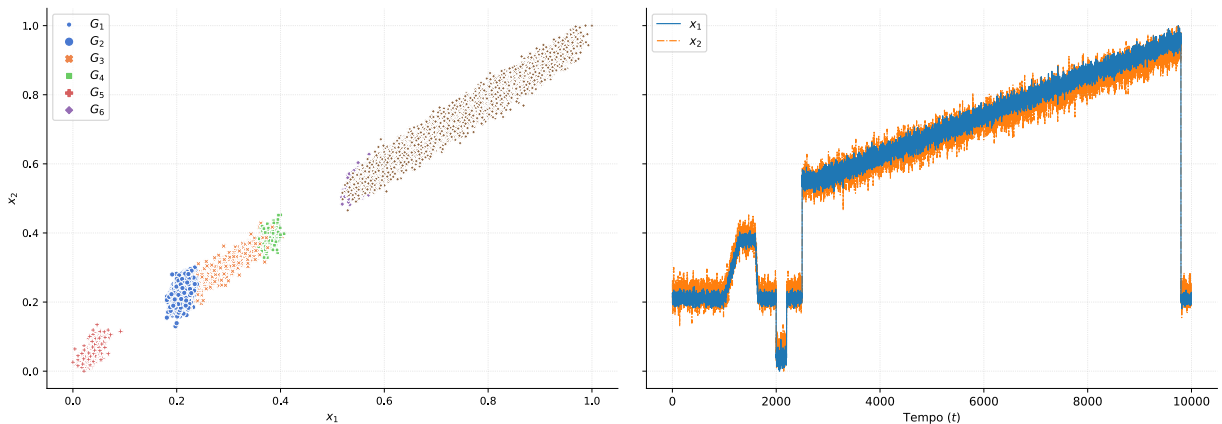


Figura 4.11 – Base de dados sintética com seis grupos.

seu valor, menor a quantidade de grupos, como apresentado na Figura 4.15. Porém, de igual maneira a janela de criação, valores extremos prejudicam a qualidade do agrupamento, como apresentado na Figura 4.16. A parametrização do eSBM+ nesse experimento foi a mesma do *janela de criação*, com exceção ao tamanho da janela que aqui foi definido em  $k = 20$ .

#### 4.2.6.6 Método de composição

Embora não exista uma dependência significativa do método de composição da matriz memória, uma vez que o mesmo é aplicado apenas na criação do grupo, nessa subseção é apresentado o resultado de um experimento que busca avaliar a influência dessa

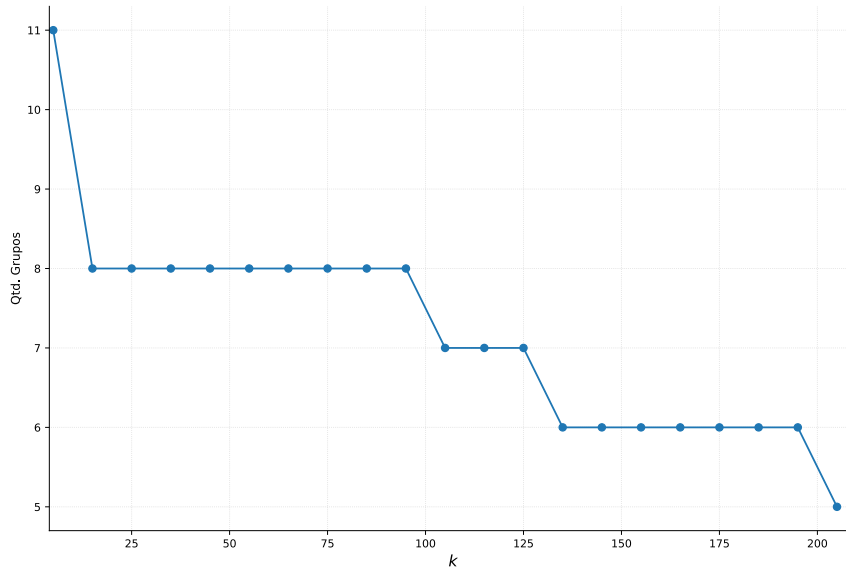


Figura 4.12 – Influência do tamanho da Janela de Criação na formação de grupos.

característica. Para tanto, utilizando o mesmo estado da Figura 4.7 foi computado o erro relativo de 100 execuções.

O resultado obtido com  $\gamma_{group} = 0.99$ ,  $\gamma_{point} = 0.2$  e  $\tau = 0.001$  é apresentado na Figura 4.17, percebe-se que houve uma influência considerável apenas nos dois operadores que apresentaram o pior desempenho em todos os experimentos realizados, RBF e STO. Nos demais, embora exista influência, não é tão significativa ao ponto de prejudicar a qualidade da modelagem. Percebe-se também que a quantidade de estados selecionados varia entre um mínimo e máximo, indicando o melhor e pior caso da representatividade da região de grupo.

### 4.3 Resumo do capítulo

Nesse capítulo foi apresentado o método eSBM, proposto como uma versão evolutiva da SBM para o monitoramento de processos industriais de forma mais independente de dados rotulados. Foram propostas um conjunto de mudanças ao eSBM. A versão com essas mudanças foi denominada de eSBM+, e para um melhor entendimento dos parâmetros dessa nova versão, foram realizados diversos experimentos, demonstrando que embora todos os parâmetros influenciem na qualidade do agrupamento, a sensibilidade do método pode ser controlada de forma mais direta pelo tamanho da janela de criação e pelo percentual de resíduo aceitável, mantendo os demais fixos. Pelos experimentos também foi possível perceber que o eSBM+ consegue lidar de forma satisfatória com mudanças de graduais/lentas no processo, sendo capaz de detectar os diferentes níveis das mesmas.

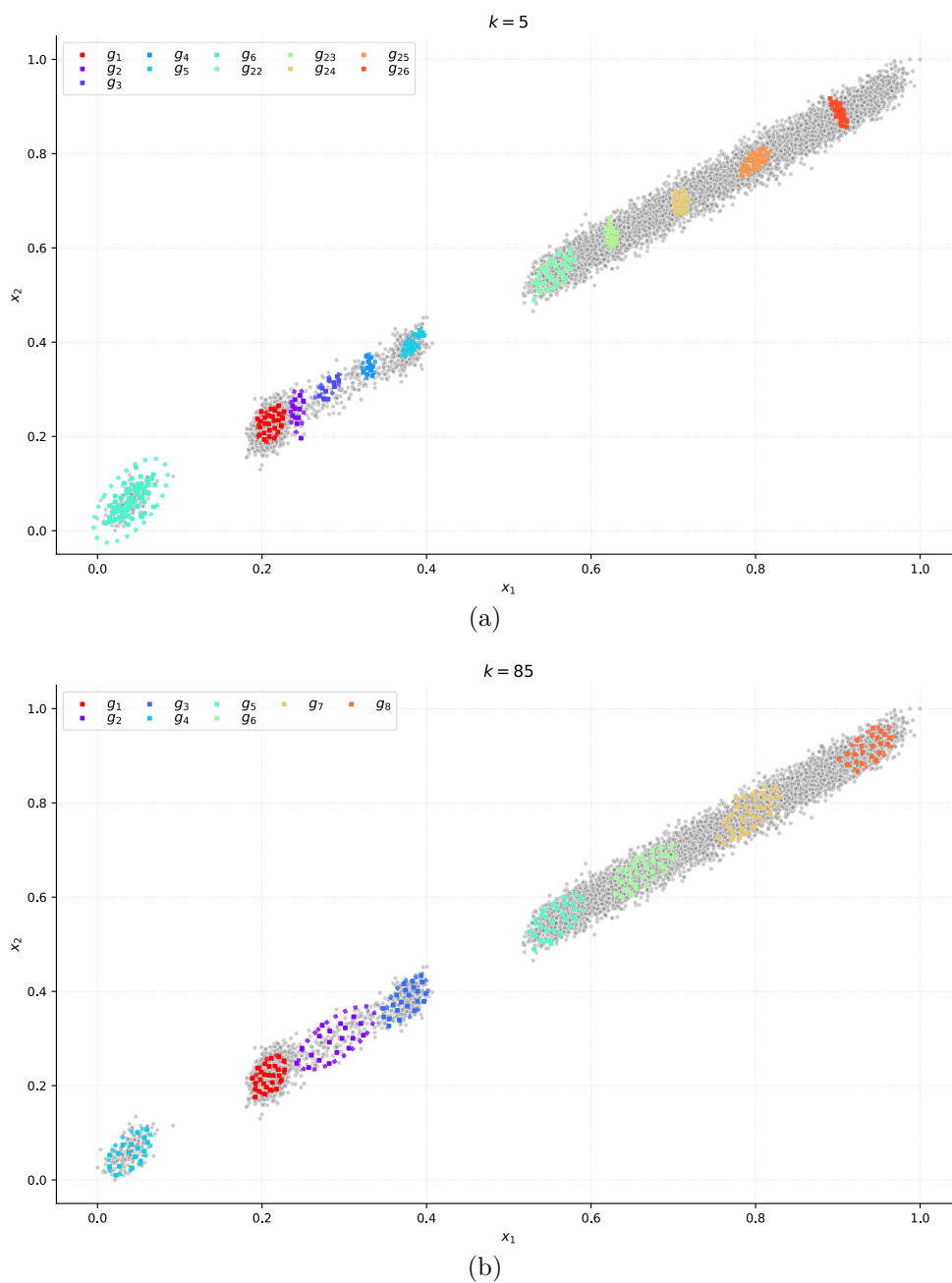
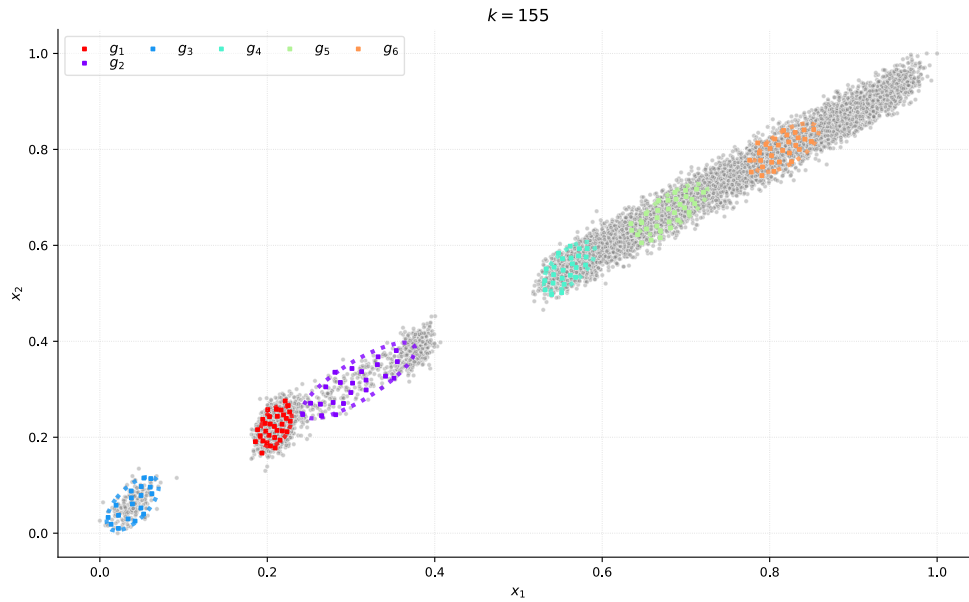
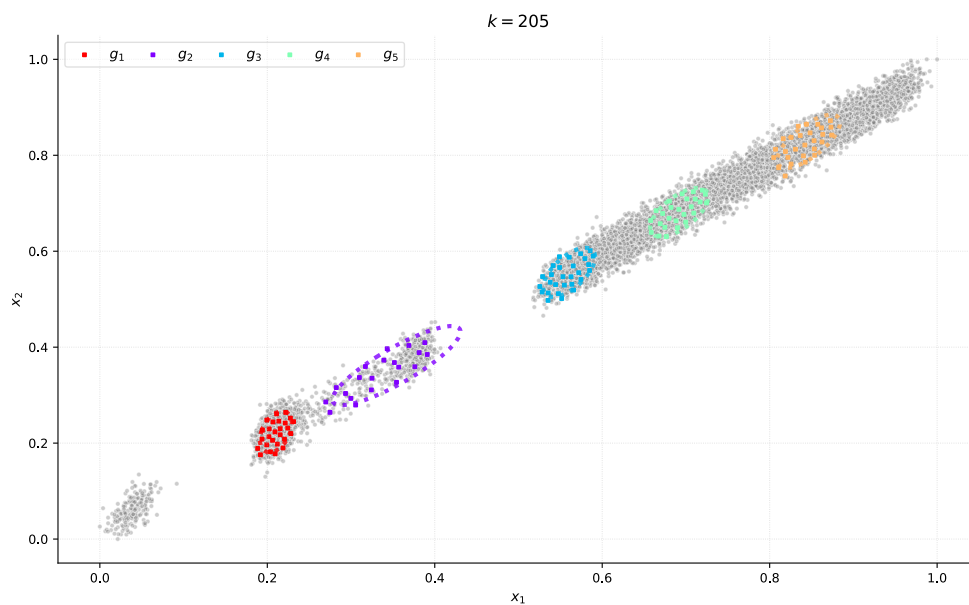


Figura 4.13 – Comportamento do agrupamento em função do tamanho da Janela de Criação. As elipses representam a região de grupo de cada grupo e os pontos coloridos os estados representativos.



(c)



(d)

Figura 4.13 – Comportamento do agrupamento em função do tamanho da Janela de Criação. As elipses representam a região de grupo de cada grupo e os pontos coloridos os estados representativos.

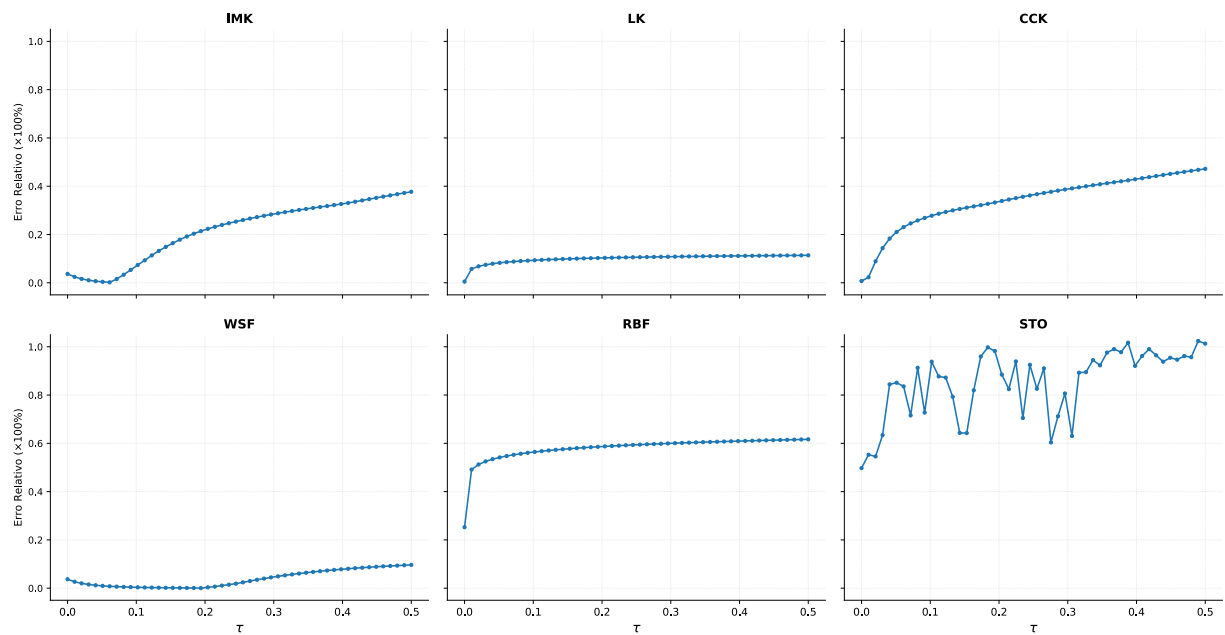


Figura 4.14 – Influência do limiar mínimo de similaridade na qualidade de estimativa para cada operador de similaridade.

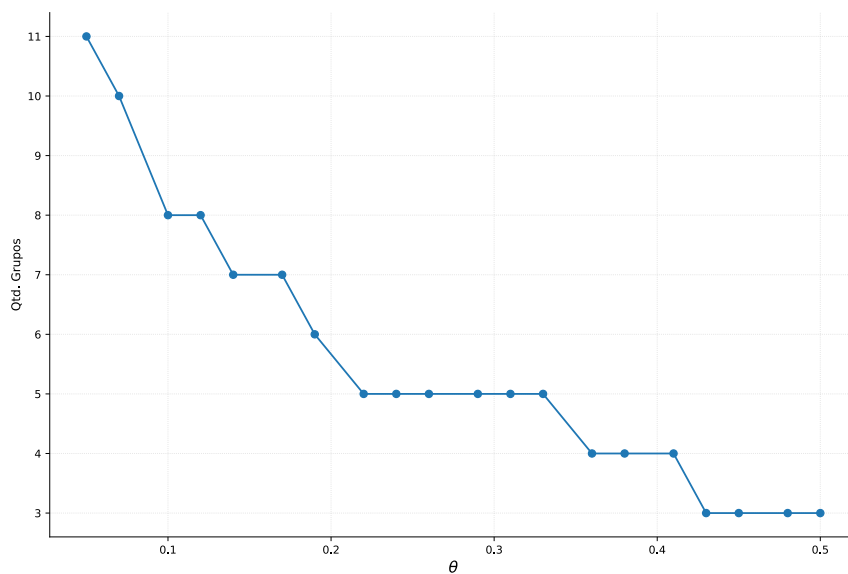


Figura 4.15 – Influência do percentual de resíduo aceitável na formação de grupos.

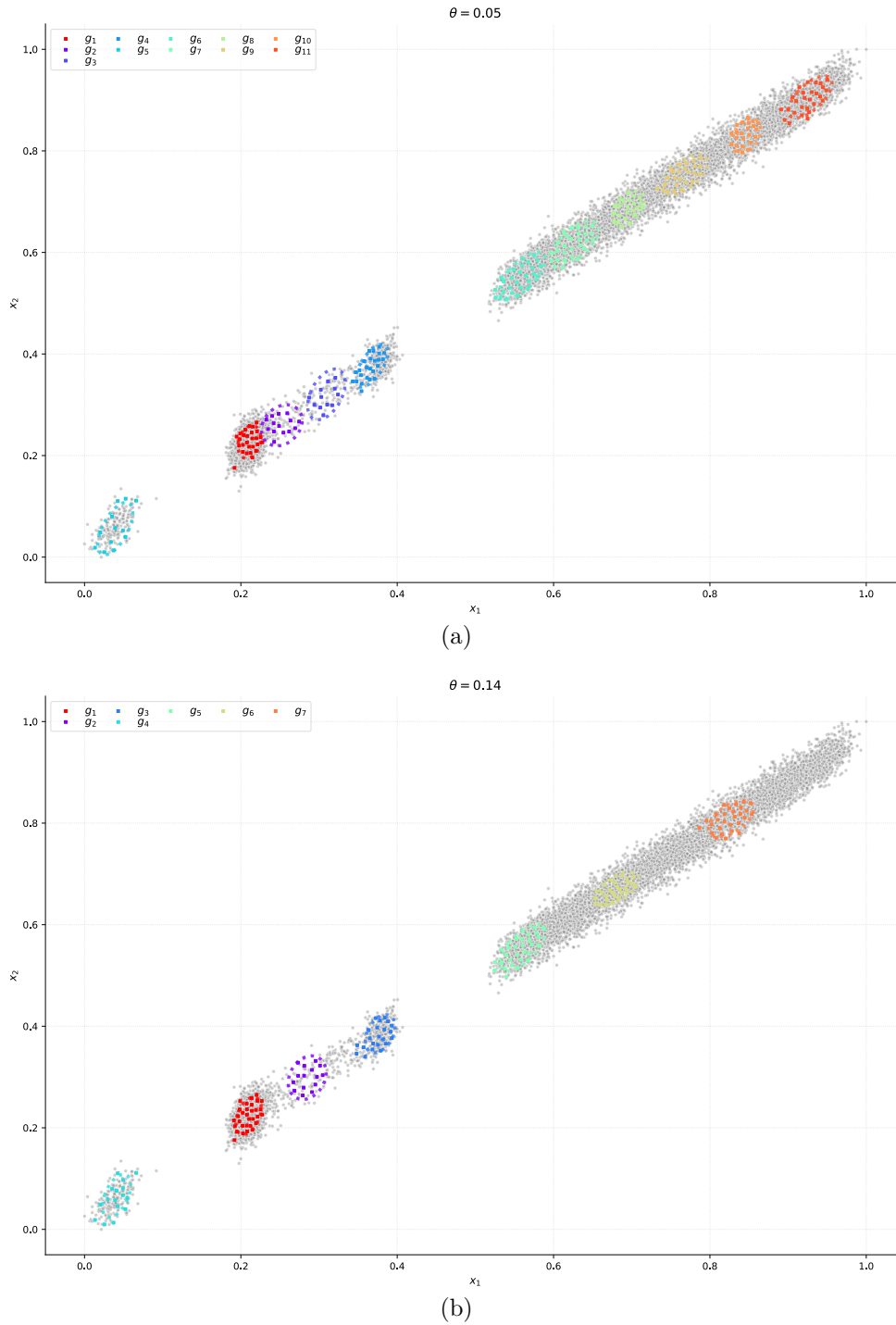


Figura 4.16 – Comportamento do agrupamento em função do percentual de resíduo aceitável. As elipses representam a região de grupo de cada grupo e os pontos coloridos os estados representativos.



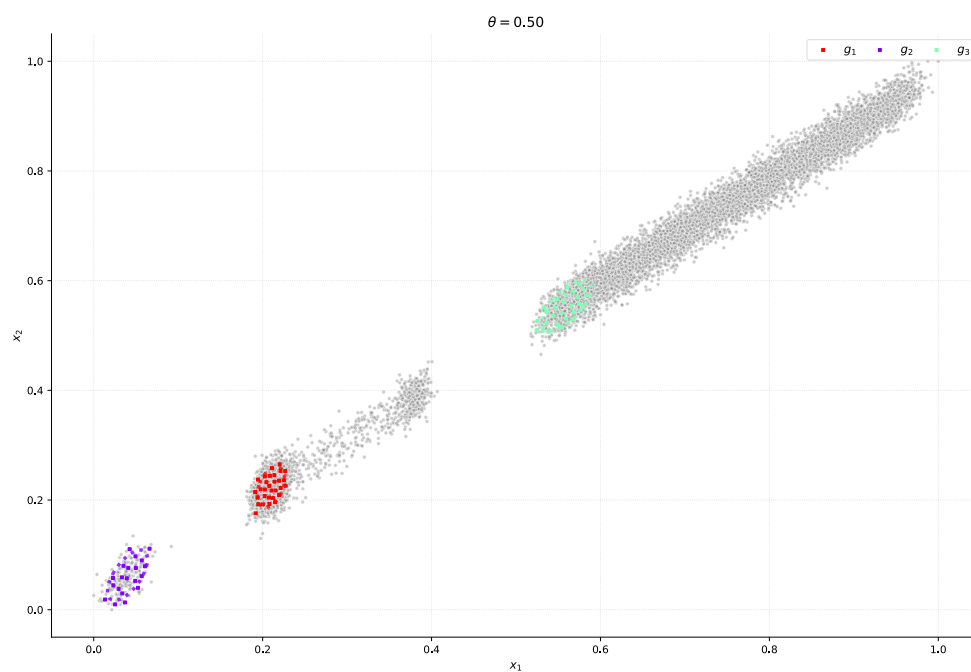
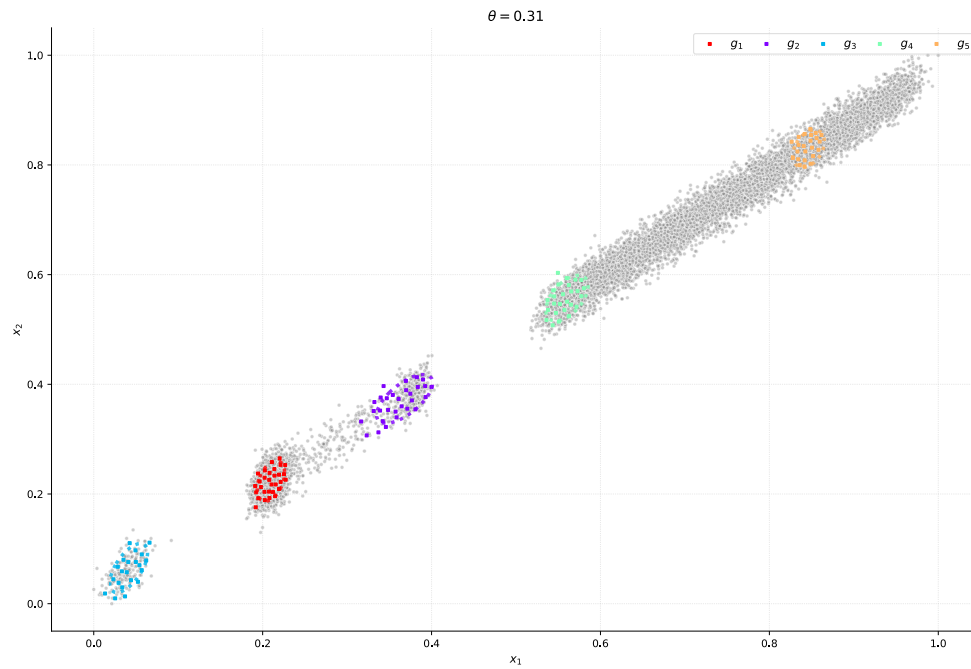


Figura 4.16 – Comportamento do agrupamento em função do percentual de resíduo aceitável. As elipses representam a região de grupo de cada grupo e os pontos coloridos os estados representativos.



Figura 4.17 – Qualidade da estimativa em função da aleatoriedade do método de composição da matriz memória em 100 execuções.

## Capítulo 5

# Diagnóstico de Causa Raiz de Falhas Baseado em Contribuição

Após a detecção e identificação de uma falha no processo, é ideal que informações sobre a causa da falha sejam fornecidas ao supervisor, para que a partir das informações dadas sejam executadas medidas eficazes e seguras para a resolução do problema. Na literatura, existe a área de pesquisa de Diagnóstico de Causa Raiz, direcionada à pesquisa e desenvolvimento de técnicas computacionais para auxiliar o supervisor a identificar a origem do problema de forma mais eficiente. Neste capítulo será apresentado o método Contribuição Baseada em Reconstrução (RBC) (do inglês, *Reconstruction-Based Contribution*), pertencente ao grupo dos métodos de diagnóstico de causa raiz baseados em contribuição. Nesse grupo é utilizada a filosofia de que, a partir de um modo de operação de referência, a causa raiz de uma determinada falha é dada as variáveis que mais contribuem em termos de magnitude de mudança de média e/ou variância.

Inicialmente é apresentado o método de Controle de Processo Estatístico Multivariado (MSPC) (do inglês, *Multivariate Statistical Process Control*) que é a base do RBC, e após, o método em questão.

## 5.1 Controle de Processo Estatístico Multivariado

A teoria probabilística pelo MSPC tem como ponto de partida a Análise de Componentes Principais (PCA) (do inglês, *Principal Components Analysis*), para tanto, a próxima seção é focada em sua apresentação. Após, os fundamentos estatísticos em que baseia-se o monitoramento envolvendo a MSPC são também apresentados.

### 5.1.1 Análise de Componentes Principais

A PCA é uma técnica multivariada que transforma um conjunto de Variáveis Aleatórias (VAs) correlacionadas em um conjunto descorrelacionado de igual ou menor

tamanho por meio de combinações lineares. Em geral, seu uso emprega-se com o objetivo de redução de dimensionalidade de dados e inferência estatística (Jackson, 1991; Johnson et al., 2002).

Algebricamente, as transformações aplicadas são uma forma particular da combinação linear de  $n$  variáveis aleatórias,  $X_1, X_2, \dots, X_n$ . Geometricamente, essas combinações lineares representam a seleção de um novo sistema de coordenadas obtidos a partir da rotação do sistema original. As Componentes Principais (PCs) são as  $p$  componentes do novo eixo,  $\{p \in \mathbb{N}^* | 1 \leq p \leq n\}$ , representando as direções com maior variância e fornecendo uma simples e mais parcimoniosa descrição da matriz de covariância das VAs que compõem o sistema original (Johnson et al., 2002).

A técnica é baseada unicamente na matriz de covariância  $\Sigma$  ou na matriz de correlação  $\rho$ . No entanto, seu uso não depende das variáveis em questão serem originárias de uma distribuição normal multivariada. Contudo, PCs derivadas dessa distribuição permitem interpretações úteis em termos de densidade de hiper-elipsoides, além disso, possibilita a realização de inferências estatísticas (Johnson et al., 2002).

Considerando o vetor aleatório  $X^T = [X_1, X_2, \dots, X_n]$  com valor esperado  $\mu$ , matriz de covariância  $\Sigma$  e também as seguintes combinações lineares:

$$\begin{aligned} Y_1 &= a_1^T X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n \\ Y_2 &= a_2^T X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2n}X_n \\ &\quad \vdots \\ Y_p &= a_p^T X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pn}X_n, \end{aligned} \tag{5.1}$$

em que, assim como demonstrado em (Johnson et al., 2002), a variância e a covariância dessa combinação linear, ou qualquer outra, podem ser definidas conforme (5.2) e (5.3), respectivamente,

$$Var(Y_i) = a_i^T \Sigma a_i, \quad i = 1, 2, \dots, p, \tag{5.2}$$

$$Cov(Y_i, Y_k) = a_i^T \Sigma a_k, \quad i, k = 1, 2, \dots, p, \tag{5.3}$$

as componentes principais serão exatamente essas combinações lineares descorrelacionadas,  $Y_1, Y_2, \dots, Y_p$ , cuja variância é tão grande quanto possível. A primeira componente é a combinação linear com a maior variância, isto é, a que maximiza

$$Var(Y_1) = a_1^T \Sigma a_1. \tag{5.4}$$

Percebe-se facilmente que (5.4) pode crescer indefinidamente apenas definindo um valor para  $a_1$  tão grande quanto se queira. Uma forma de eliminar esse comportamento é restringindo o comprimento dos vetores formados pelos coeficientes de  $a$  para um tamanho unitário. Cada componente será obtida sob a perspectiva de um problema de otimização matemática:

$$PC_1 = a_1^T X \text{ que maximiza } Var(a_1^T X), \text{ sujeito a } a_1^T a_1 = 1.$$

$$PC_2 = a_2^T X \text{ que maximiza } Var(a_2^T X), \text{ sujeito a } a_2^T a_2 = 1 \text{ e } Cov(a_1^T X, a_2^T X) = 0.$$

$$\vdots$$

$$PC_p = a_p^T X \text{ que maximiza } Var(a_p^T X), \text{ sujeito a } a_p^T a_p = 1 \text{ e } Cov(a_p^T X, a_k^T X) = 0$$

para todo  $k < p$ .

Uma segunda forma mais comum de obter as componentes principais ocorre por meio da decomposição espectral da matriz de covariância ou correlação (Jackson, 1991; Johnson et al., 2002), sendo essa utilizada no decorrer dessa dissertação. Considera-se que a matriz de covariância  $\Sigma$  é diagonalizável e como tal é decomposta nos seguintes pares de autovalores e autovetores:  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_n, e_n)$ , isto é,

$$\Sigma = P \Lambda P^T = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pn} \end{bmatrix} \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ 0 & \lambda_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{np} \end{bmatrix} \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{1p} \\ e_{12} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1n} & e_{2n} & \cdots & e_{np} \end{bmatrix}^T, \quad (5.5)$$

$$\lambda_{11} \geq \lambda_{22} \geq \dots \lambda_{np} \geq 0, \quad (5.6)$$

a  $p$ -ésima componente principal será dada por:

$$PC_p = Y_p = e_p^T X = e_{p1}^T X_1 + e_{p2}^T X_2 + \dots + e_{pn}^T X_n, \quad p = 1, 2, \dots, n, \quad (5.7)$$

de forma que,

$$Var(Y_i) = e_i^T \Sigma e_i = \lambda_i, \quad i = 1, 2, \dots, p, \quad (5.8)$$

$$Cov(Y_i, Y_k) = e_i^T \Sigma e_k = 0, \quad \forall i \neq k \mid i, k = 1, 2, \dots, p. \quad (5.9)$$

Para um resultado coerente é crucial que as variáveis tenham a mesma escala, pois as componentes são obtidas em função da variância, dessa maneira, variáveis com maior escala transmitiram a falsa sensação de alta variância, muito embora seja apenas uma diferença nas unidades de medidas. A padronização mais comum nesse contexto é a de média zero e desvio padrão unitário, porém, quaisquer outras podem ser utilizadas desde que as variáveis padronizadas fiquem com média zero, um requisito para a correta execução da rotação do eixo original e validade de inferências estatísticas.

Para ilustrar o PCA considere o conjunto bidimensional de 10 amostras apresentado na Figura 5.1, essas amostras são realizações do vetor aleatório  $X^T = [x_1, x_2] \sim \mathcal{N}_2(\mu, \Sigma)$ , em que  $\mu^T = [10, 20]$  e  $\Sigma = \begin{bmatrix} 26 & 19 \\ 19 & 40 \end{bmatrix}$ . O centro utilizado na figura foi computado apenas com as amostras, na figura também são evidenciados os dois eixos do conjunto,  $x_1$  e  $x_2$ .

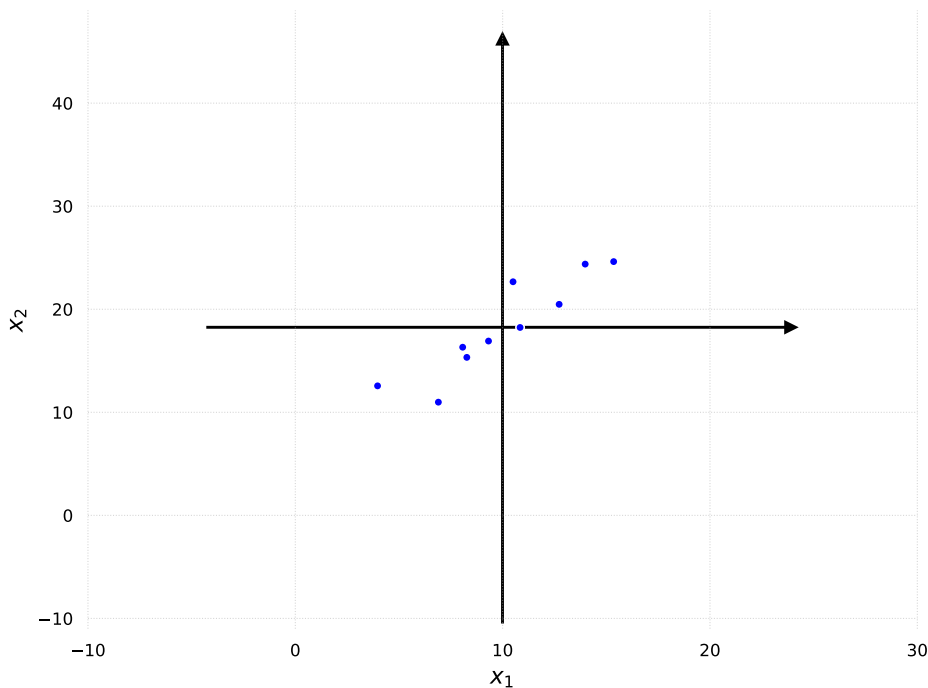


Figura 5.1 – Conjunto com 10 realizações do vetor aleatório  $X \sim \mathcal{N}_2(\mu, \Sigma)$  e seus respectivos eixos,  $x_1$  e  $x_2$ .

Utilizando a média e matriz de correlação amostrais obteve-se o resultado apresentado na Figura 5.2 com as duas abordagens descritas anteriormente. Optou-se por utilizar a matriz de correlação amostral porque, primeiro, assim como será discutido na Seção 5.1.2, dificilmente em problemas reais a matriz de covariância/correlação será conhecida, sendo necessária estimá-la a partir dos dados disponíveis. Segundo, se os dados estiverem padronizados para média zero e desvio padrão unitário, calcular a matriz de covariância irá gerar a de correlação.

Como foi dito no início da seção, geometricamente as componentes resultantes

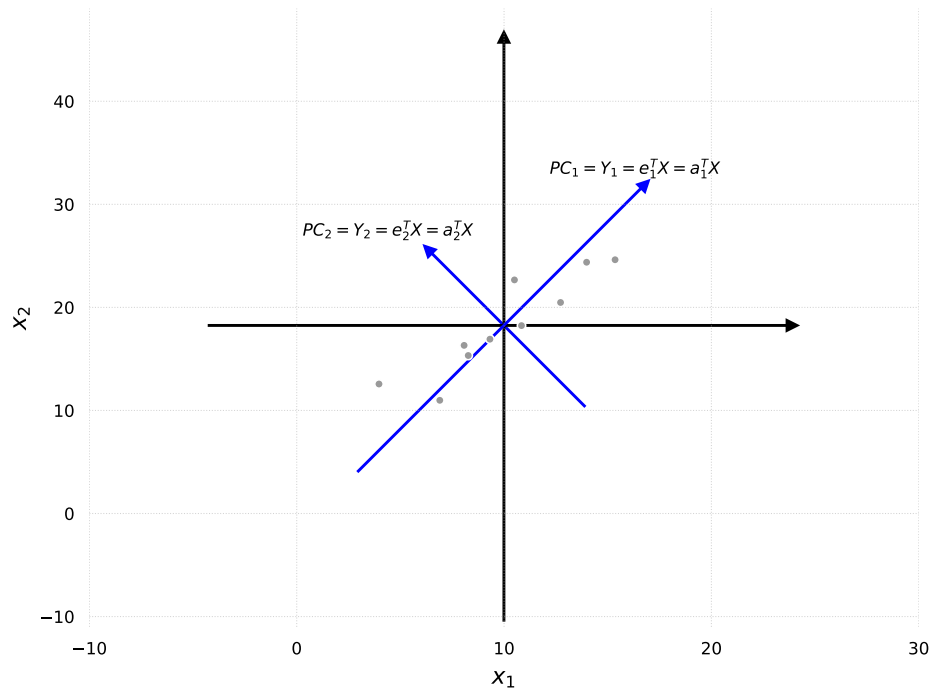


Figura 5.2 – Componentes obtidas para o conjunto de pontos ilustrativo.

são o resultado da rotação dos eixos originais de forma a maximizar a variância gerada da interação entre as variáveis. Tal rotação é caracterizada pelos coeficientes  $a_{11}, a_{12}, \dots, a_{pn}$  ou pelos autovetores  $e_{11}, e_{12}, \dots, e_{pn}$ , que por sua vez representam o cosseno do ângulo em radianos entre o novo eixo e os antigos, assim como ilustrado na Figura 5.3 e 5.4 para a primeira ( $PC_1$ ) e segunda ( $PC_2$ ) componente, respectivamente.

No que diz respeito as diferentes abordagens para a busca das componentes, desde que as variáveis sejam correlacionadas ou as variâncias sejam diferentes entre si, ambas as abordagens fornecem o mesmo resultado. Cabe ressaltar que quando não existir correlação entre as variáveis, a aplicação do PCA não irá agregar ganhos, pois as direções que maximizam a variância serão as mesmas do sistema de coordenadas original, além disso, se as variáveis também tiverem as mesmas variâncias, existirá infinitas combinações lineares que, em termos de qualidade de modelagem, será a mesma obtida pelo sistema de coordenadas original, e nessa situação, o resultado pode divergir entre as duas abordagens (Johnson et al., 2002). As duas situações são ilustradas na Figura 5.5.

A seleção da quantidade de componentes principais ocorre, em geral, por meio do critério da Variância Percentual Acumulada (CPV) (do inglês, *Cumulative Percentage Variance*), devido sua simplicidade (Johnson et al., 2002):

$$CPV = 100 \times \sum_{i=1}^n \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}, \quad (5.10)$$

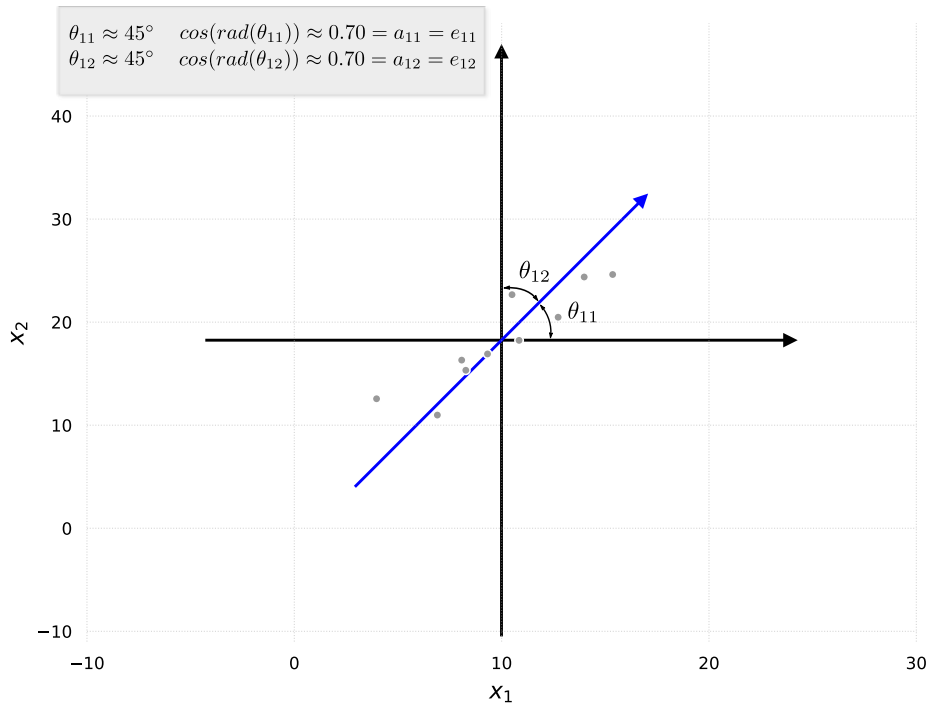


Figura 5.3 – Interpretação dos valores dos coeficientes/autovetores relacionados à primeira componente,  $PC_1$ .  $\text{rad}$  representa uma função que realiza a transformação de graus para radianos.

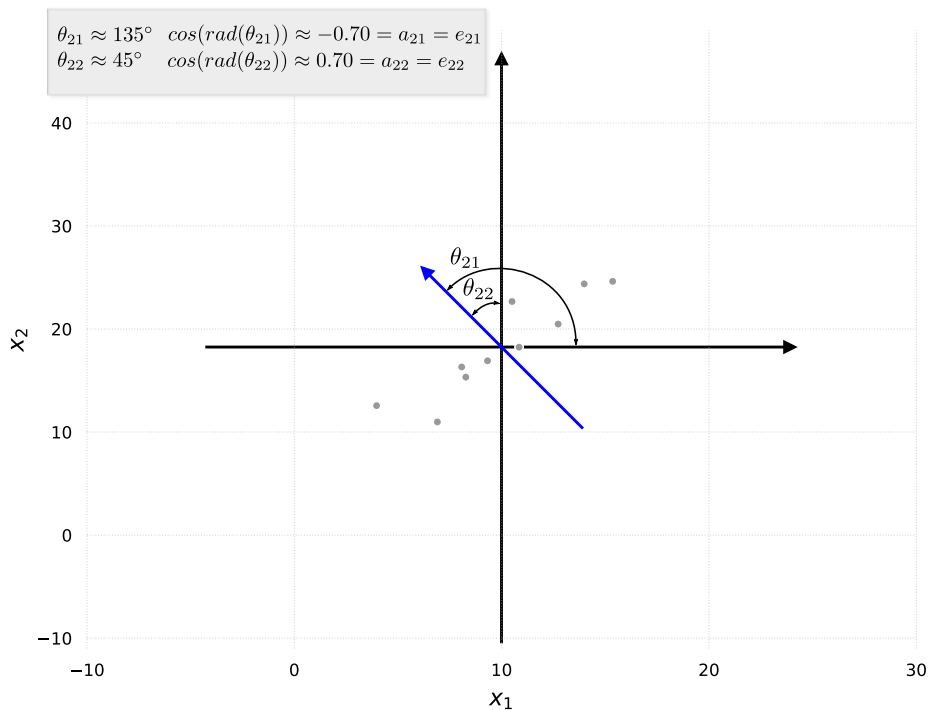
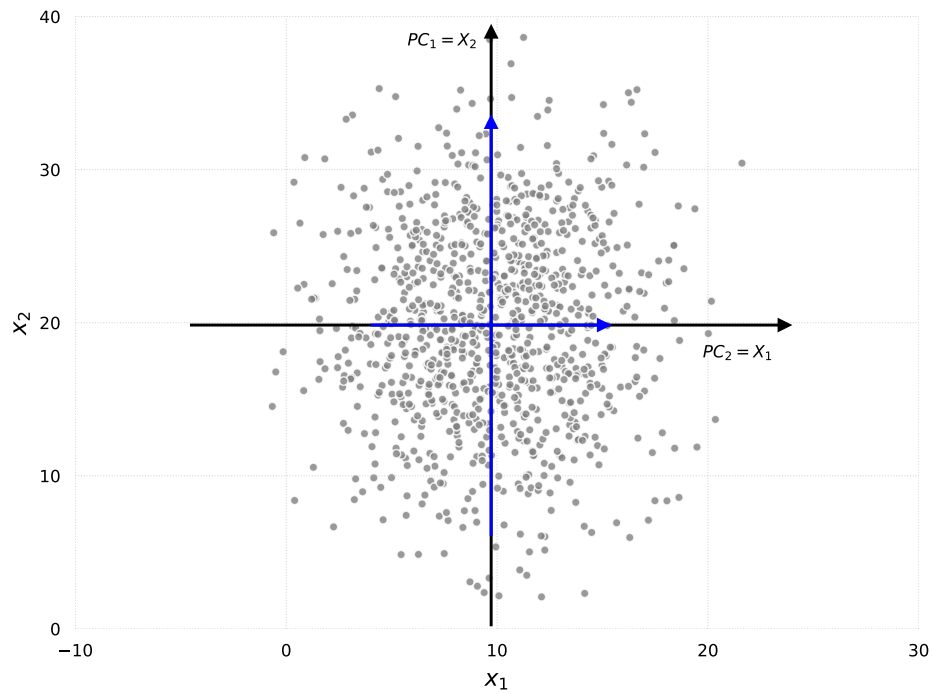
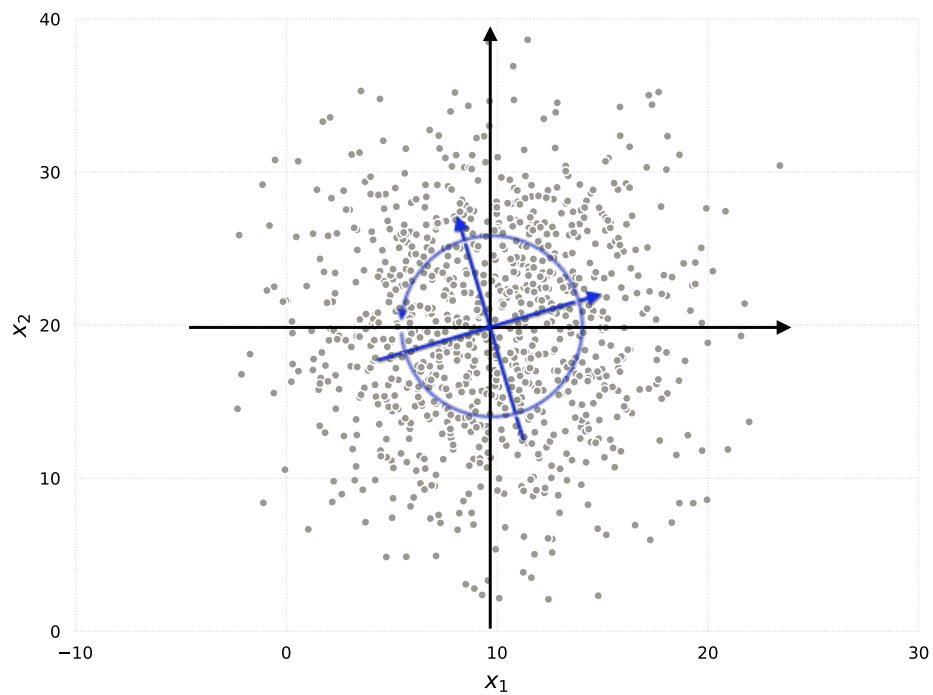


Figura 5.4 – Interpretação dos valores dos coeficientes/autovetores relacionados à segunda componente,  $PC_2$ .  $\text{rad}$  representa uma função que realiza a transformação de graus para radianos.





(a)



(b)

Figura 5.5 – Ilustração do comportamento do PCA quando aplicado em variáveis descorrelacionadas (a) ou descorrelacionadas e com variâncias iguais entre si (b).

$$CPV \geq 100 \times \alpha_{cpv}, \quad (5.11)$$

em que  $\alpha_{cpv}$ ,  $\{\alpha_{cpv} \in \mathbb{R} \mid 0 < \alpha_{cpv} < 1\}$ . Mas, em aplicações em que o objetivo é a qualidade da aproximação, a validação cruzada, critério de Akaike (AIC) (do inglês, *Akaike Information Criterion*) ou de informação bayesiana (BIC) (do inglês, *Bayesian information criterion*) (Burnham and Anderson, 2004) serão métodos mais apropriados para a avaliação da qualidade do modelo.

Recomenda-se na literatura  $\alpha_{cpv} = 0.95$ , de forma que o restante da variância seja considerada como ruído descorrelacionado. Por exemplo, utilizando esse critério para o conjunto de dados ilustrativo, apenas a primeira componente seria suficiente para explicar 95% da variância (Figura 5.6).

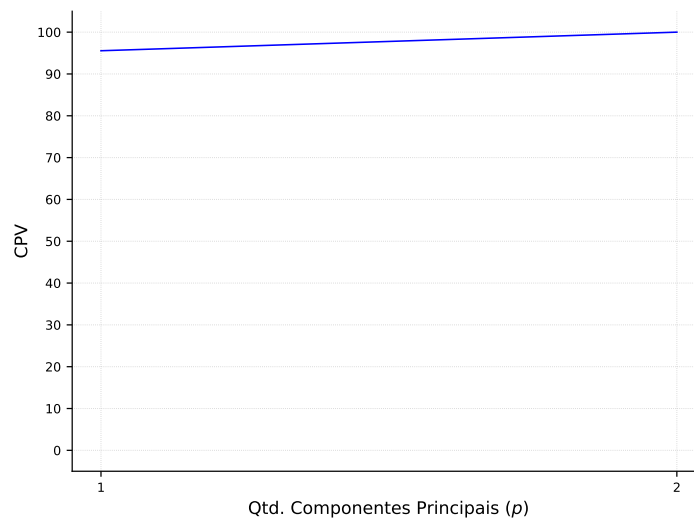


Figura 5.6 – CPV para o conjunto de dados ilustrativo.

Definida a quantidade de PCs, novas amostras são transformadas conforme (5.12),

$$z = Y_p^T X, \quad (5.12)$$

em que  $X$  pode ser reconstruído de forma aproximada para o espaço original utilizando o inverso da expressão anterior,

$$\hat{X} = Y_p z, \quad (5.13)$$

e o erro na reconstrução será proporcional a variância acumulada pelas  $p$  componentes principais.

### 5.1.2 Controle estatístico

Partindo da premissa de que as  $n$  variáveis do processo seguem uma distribuição normal multivariada,  $x = [x_1, x_2, \dots, x_n]^T$ ,  $\{x \sim \mathcal{N}_n(\mu, \Sigma)\}$ , estabelece-se medidas estatísticas que indicam a probabilidade do processo estar fora da região de operação normal.

Contudo, em geral,  $\mu$  e  $\Sigma$  não são conhecidas *a priori*, nessa situação utiliza-se amostras do processo para estimá-las,

$$\bar{X} = \frac{1}{k-1} \sum_{i=1}^k x_i, \quad (5.14)$$

$$S = \frac{1}{k-1} \sum_{i=1}^k (x_i - \bar{X})(x_i - \bar{X})^T, \quad (5.15)$$

em que  $k$  indica a quantidade de amostras,  $\bar{X} \in \mathbb{R}^n$  a média amostral, e  $S \in \mathbb{R}^{n \times n}$  a matriz de covariância amostral.

Com isso, utiliza-se o PCA para subdividir o espaço cartesiano original em um Subespaço Principal (SP) contendo as  $p$  componentes principais, e um Subespaço Residual (SR) contendo as  $(n-p)$  componentes restantes (Joe Qin, 2003):

$$S = P\Lambda P^T = \hat{P}\hat{\Lambda}\hat{P}^T + \tilde{P}\tilde{\Lambda}\tilde{P}^T = SP + SR, \quad (5.16)$$

em que  $\hat{P} \in \mathbb{R}^{n \times p}$  e  $\hat{\Lambda} \in \mathbb{R}^{p \times p}$  são respectivamente os autovetores e autovalores que compõem o SP,  $\tilde{P} \in \mathbb{R}^{n \times (n-p)}$  e  $\tilde{\Lambda} \in \mathbb{R}^{(n-p) \times (n-p)}$  os que compõem o SR.

Por meio dessa abordagem o monitoramento do processo pode ser executado apenas com o SP ou SR, pois ambos, dentro de suas especificidades, resumem a relação linear existente em um conjunto menor de variáveis "artificiais" descorrelacionadas, e que devido a falta de correlação, possibilitam uma forma mais fácil de monitoração individual, e ainda, estimar a probabilidade real do erro Tipo I e II (Jackson, 1991).

A transformação para cada subespaço ocorre por meio de (5.12). Para o SP tem-se,

$$\hat{z} = \hat{P}^T x, \quad (5.17)$$

e para o SR,

$$\tilde{z} = \tilde{P}^T x, \quad (5.18)$$

e  $x$  pode ser completamente reconstruído utilizando (5.13), porém, agora considerando não apenas as componentes principais,

$$x = \hat{P}\hat{z} + \tilde{P}\tilde{z}. \quad (5.19)$$

Por conseguinte define-se valores críticos para indicar quando o processo está operando em situação anormal. Esses valores são fornecidos por funções estatísticas que

na literatura de monitoramento estatístico são chamadas de índices de detecção. Em geral, são utilizadas três, a saber: Estatística de *Hottelling*, Erro Quadrático ou a combinação das duas anteriores.

### 5.1.2.1 Estatística de *Hotteling*

Em situações que  $\mu$  e  $\Sigma$  são conhecidos, ou pelo menos  $k$  é significativamente grande, é possível, sob uma perspectiva global, definir o valor crítico pela seguinte medida:

$$D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu), \quad (5.20)$$

pois

$$D^2 \sim \chi_n^2, \quad (5.21)$$

em que  $\chi_n^2$  é a distribuição Qui-quadrado com  $n$  graus de liberdade. O processo estará operando normalmente se as amostras processadas tiverem a distância expressa por (5.20) menor ou igual ao valor crítico,

$$\delta_{chi}^2 = \chi_{n,\alpha}^2, \quad (5.22)$$

$$D^2 \leq \delta_{chi}^2, \quad (5.23)$$

em que  $\alpha$  define o valor crítico ou nível de significância, e é a probabilidade assumida para o erro Tipo I, ou seja, a probabilidade da distância computada ultrapassar o valor crítico mesmo com o processo operando normalmente. Em geral, utiliza-se  $\alpha = 0.01$ .

No entanto, em situações práticas, recomenda-se utilizar a distribuição *T-squared* de *Hotelling*, que é uma generalização da distribuição *t* de *Student* (Abramowitz and Stegun, 1948; Wasserman, 2013) e fornecerá o mesmo resultado da Qui-quadrado quando  $k$  for grande. Nesse sentido, Hotelling (1931) demonstra que:

$$T^2 = (x - \bar{X})^T S^{-1} (x - \bar{X}), \quad (5.24)$$

$$T^2 \sim \frac{n(k-1)}{k-n} F_{n,k-n}, \quad (5.25)$$

em que  $F_{n,k-n}$  é a distribuição *F* de Snedecor (Abramowitz and Stegun, 1948) com  $k$  e  $(k-n)$  graus de liberdade. Dessa forma tem-se um valor crítico menos conservador para o monitoramento e em função da quantidade de amostras disponíveis. A avaliação de

anormalidade ocorre de forma similar a (5.23), considera-se que o processo está operando normalmente se (Jackson, 1991; Joe Qin, 2003)

$$T^2 \leq \delta_{T^2}^2, \quad (5.26)$$

com

$$\delta_{T^2}^2 = \frac{n(k-1)}{k-n} F_{n,k-n,\alpha}. \quad (5.27)$$

Para monitorar o processo apenas pelo SP são utilizadas as mesmas abordagens,

$$T^2 = \hat{z}^T \hat{\Lambda}^{-1} \hat{z}, \quad (5.28)$$

e o valor crítico definido por meio de (5.22) ou (5.27), dependendo da quantidade de amostras disponíveis.

### 5.1.2.2 Erro Quadrático

O Erro Quadrático de Previsão (SPE) (do inglês, *Squared Prediction Error*) ou ainda estatística Q, é uma segunda forma de monitorar o processo, apenas por meio do SR. Jackson and Mudholkar (1979) o define como a média da soma do quadrado da diferença entre  $x$  e sua aproximação utilizando apenas o SP,  $\hat{x}$ , esse último expresso por (5.13) com as devidas alterações:

$$SPE = (x - \hat{x})^T (x - \hat{x}) = x^T \tilde{P} \tilde{P}^T x = \tilde{z}^T \tilde{z} = \|\tilde{z}\|^2, \quad (5.29)$$

em que essa quantidade representa o quão bem o SP representa a variação do processo, ou em outras palavras, representa o quão bem uma observação ajusta-se ao modelo linear com coeficientes iguais a SP.

Para obter o valor crítico considere

$$\theta_1 = \sum_{i=p+1}^n \tilde{\Lambda}_i,$$

$$\theta_2 = \sum_{i=p+1}^n \tilde{\Lambda}_i^2,$$

$$\theta_3 = \sum_{i=p+1}^n \tilde{\Lambda}_i^3,$$

e por último

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}.$$

Com base nesses valores é possível definir a quantidade

$$c = \theta_1 \frac{\left[ \left( \frac{SPE}{\theta_1} \right)^{h_0} - \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} - 1 \right]}{\sqrt{2\theta_2 h_0^2}}, \quad (5.30)$$

que aproxima-se de uma distribuição normal com média zero e variância unitária (Jackson and Mudholkar, 1979). Dessa forma, o valor crítico para o SPE pode ser definido como:

$$\delta_{SPE}^2 = \theta_1 \left[ \frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1 \right]^{1/h_0}, \quad (5.31)$$

em que  $c_\alpha$  é o desvio de uma distribuição normal relativo ao percentil superior  $1 - \alpha$ , ou seguindo a mesma interpretação da estatística de *Hotteling*,  $c_\alpha$  é a quantidade de desvio positivo que define a probabilidade de erro Tipo I igual a  $\alpha$  para uma distribuição normal (Jackson, 1991).

O processo estará operando normalmente se

$$SPE \leq \delta_{SPE}^2. \quad (5.32)$$

### 5.1.2.3 Combinação de estatísticas

Assim como demonstrado por Dunia and Joe Qin (1998); Yue and Qin (2001), existem anormalidades que podem ser capturadas apenas por um dos índices anteriores,  $T^2$  ou  $SPE$ . Uma forma de contornar esse problema seria utilizar ambas de forma paralela. No entanto, em termos práticos, torna-se preferível o uso de apenas um. Yue and Qin (1998, 2001) propuseram uma versão em que os dois são combinados em um só, o chamaram de Índice Combinado, sendo definido como:

$$\varphi = \frac{T^2}{\delta_{T^2}^2} + \frac{SPE}{\delta_{SPE}^2} = x^T \Phi x, \quad (5.33)$$

em que

$$\Phi = \frac{\hat{P}\hat{\Lambda}^{-1}\hat{P}^T}{\delta_{T^2}^2} + \frac{\tilde{P}\tilde{P}^T}{\delta_{SPE}^2}. \quad (5.34)$$

A partir da forma quadrática definida em (5.33) e dos resultados de [Box \(1954\)](#), que discute sobre a distribuição de formas quadráticas originárias de uma distribuição normal multivariada, os autores definem o valor crítico,

$$\delta_{comb}^2 = g\chi_{h,\alpha}^2, \quad (5.35)$$

em que  $g$  é definido por

$$g = \frac{tr(S\Phi)^2}{tr(S\Phi)}, \quad (5.36)$$

e os graus de liberdade,  $h$ , por

$$h = \frac{[tr(S\Phi)]^2}{tr(S\Phi)^2}, \quad (5.37)$$

$tr(S\Phi)$  e  $tr(S\Phi)^2$  podem ser simplificadas em ([Joe Qin, 2003](#)):

$$tr(S\Phi) = \frac{p}{\delta_{T^2}^2} + \frac{\theta_1}{\delta_{SPE}^2}, \quad (5.38)$$

$$tr(S\Phi)^2 = \frac{p}{[\delta_{T^2}^2]^2} + \frac{\theta_2}{[\delta_{SPE}^2]^2}. \quad (5.39)$$

Além das justificativas dadas anteriormente, considerando o preceito de que o processo segue uma distribuição normal multivariada, o uso do Índice Combinado torna-se preferível ao  $T^2$  e  $SPE$ , pois o valor crítico é definido com o formato de um hiper-elipsoide, ao contrário do outros dois, que o definem como retas contínuas. Uma ilustração do comportamento dos índices apresentados é dado na Figura 5.7.

O processo estará operando normalmente se

$$\varphi \leq \delta_{comb}^2. \quad (5.40)$$

A expressão (5.33) permite ainda definir uma forma genérica para computar o índice considerando as três abordagens apresentadas ([Alcala and Qin, 2009](#)):

$$Index(x) = x^T M x, \quad (5.41)$$

em que  $M$  é definida conforme Tabela 1.

### 5.1.3 Análise Dinâmica de Componentes Principais

As inferências estatísticas realizadas utilizando a abordagem clássica do PCA (Seção 5.1.1) baseiam-se na premissa de que as amostras são independentes e identicamente

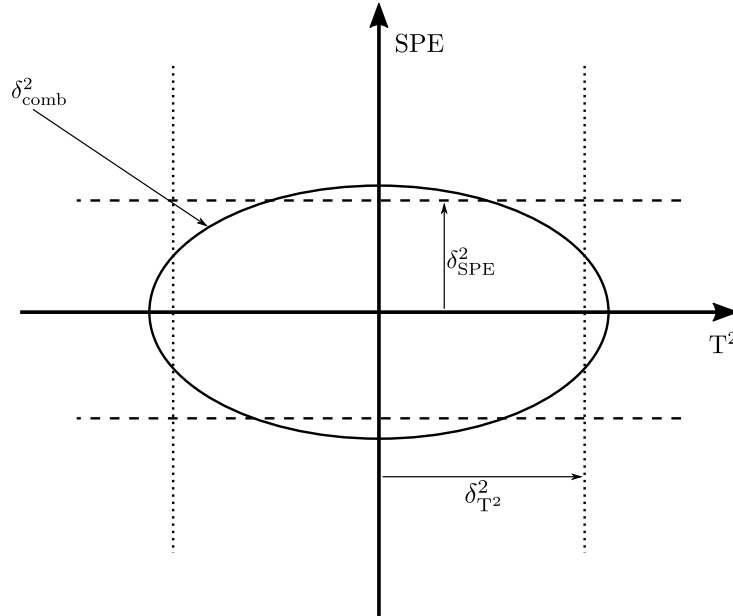


Figura 5.7 – Ilustração dos valores críticos para cada índice considerando um problema bidimensional. O eixo horizontal representa o SP e o vertical o SR. As linhas pontilhadas representam o valor crítico do índice  $T^2$ , as tracejadas do  $SPE$  e a elipse da combinação dos dois anteriores.

Fonte: Adaptado de [Yue and Qin \(2001\)](#).

	$T^2$	SPE	Combinação
$M$	$\hat{P}\hat{\Lambda}^{-1}\hat{P}^T$	$\tilde{P}\tilde{P}^T$	$\phi$

Tabela 1 – Valor da matriz  $M$  para cada índice.

distribuídas (i.i.d.), contudo, em processos industriais, raramente essa suposição é satisfeita, pelo menos no que diz respeito a independência.

Para tanto, considere a amostra  $x_t \in \mathbb{R}^n$  referente ao instante  $t$  e  $x_{t-1} \in \mathbb{R}^n$  ao instante anterior,  $t - 1$ . A inferência realizada utilizando o PCA é válida, ainda que de forma relaxada, se

$$\text{cov}(x_{ti}, x_{(t-1)i}) \approx 0, \quad \forall i \mid i = 1, 2, \dots, n, \quad (5.42)$$

caso essa condição não seja satisfeita, existirá autocorrelação e/ou correlação cruzada, que em termos simples, significa que o estado atual do processo, instante  $t$ , é resultado da situação em que encontrava-se anteriormente, instante  $t - 1$ , e isso, nesse contexto, é uma suposição totalmente válida e realista. Uma falha incipiente, por exemplo, descreve perfeitamente essa situação. Porém, o monitoramento não possui validade estatística, pois a premissa de independência não foi satisfeita.

Uma forma de lidar com essa situação, proposta por [Ku et al. \(1995\)](#), consiste em adicionar variáveis "extras" na construção do modelo PCA, essas variáveis extras serão



valores atrasados no tempo. Essa abordagem foi denominada pelo autor como Análise Dinâmica de Componentes Principais (DPCA) (do inglês, *Dynamic Principal Component Analysis*). Para o caso geral, ter-se-á

$$x_1 = [x_{11}, x_{12}, \dots, x_{1n}]^T, \quad (5.43)$$

$$S_{lag} = cov(M) = \begin{bmatrix} x_1 & | & x_{1-1} & | & \cdots & | & x_{1-l} \\ x_2 & | & x_{2-1} & | & \cdots & | & x_{2-l} \\ \vdots & | & \vdots & | & \ddots & | & \vdots \\ x_t & | & x_{t-1} & | & \cdots & | & x_{t-l} \end{bmatrix}^T, \quad x_0 = 0, \quad (5.44)$$

em que  $M \in \mathbb{R}^{k \times (n+nl)}$  é uma matriz de dados com as variáveis atrasadas no tempo,  $t = 1, 2, \dots, k$  e  $l \in \mathbb{N}^*$  representa a quantidade de amostras atrasadas no tempo e deve ser determinada por algum método de avaliação de qualidade de modelos, como validação cruzada, AIC ou BIC, por exemplo.

Dessa forma, ao utilizar o PCA baseado na  $S_{lag}$ , elimina-se a correlação, autocorrelação e correlação cruzada, garantindo a base estatística do monitoramento.

## 5.2 Contribuição Baseada em Reconstrução

O RBC (Yue and Qin, 1998; Dunia and Joe Qin, 1998; Yue and Qin, 2001; Alcalá and Qin, 2009, 2010; Li et al., 2010, 2014) baseia-se na MSPC e vem sendo utilizado com sucesso como uma ferramenta para auxiliar o supervisor do processo na identificação da causa raiz de falhas (Ge et al., 2010; He et al., 2014; Zhu et al., 2017; Elshenawy and Mahmoud, 2018). O método indica quais as variáveis mais influenciaram por meio da minimização dos efeitos de cada variável no valor do índice.

No espaço de variáveis esse procedimento representa reconstruir a amostra sem os efeitos da falha, isto é, sem a parcela que a faz ultrapassar o valor crítico do respectivo índice,

$$z_i = x - \xi_i f_i, \quad i = 1, 2, \dots, n, \quad (5.45)$$

em que  $z_i$  representa a estimativa para a  $i$ -ésima variável sem os efeitos da falha,  $\xi_i$  é  $i$ -ésima coluna de uma matriz identidade e representa a direção da falha,  $f_i$  é a estimativa da magnitude da falha.

O índice de  $z_i$  será definido como:

$$Index(z_i) = (x - \xi_i f_i)^T M (x - \xi_i f_i). \quad (5.46)$$

A tarefa de reconstruir consiste em buscar o valor de  $f_i$  que faz o índice,  $Index(z_i)$ , ser minimizado. Uma solução analítica pode ser obtida calculando a primeira derivada de  $Index(z_i)$  em relação a  $f_i$  e igualar a zero (Yue and Qin, 2001; Alcalá and Qin, 2009):

$$\frac{\partial Index(z_i)}{\partial f_i} = -2(x - \xi_i f_i)^T M \xi_i, \quad (5.47)$$

igualando a zero e isolando  $f_i$ ,

$$f_i = (\xi_i^T M \xi_i)^{-1} \xi_i^T M x. \quad (5.48)$$

O RBC é essa quantidade reconstruída na direção  $\xi_i$  e que a partir de (5.46) pode ser expressa como (Alcalá and Qin, 2009):

$$RBC_i = (\xi_i f_i)^T M (\xi_i f_i). \quad (5.49)$$

Substituindo (5.48) em (5.49) e realizando algumas manipulações tem-se:

$$RBC_i = x^T M \xi_i (\xi_i^T M \xi_i)^{-1} \xi_i^T M x, \quad (5.50)$$

e sob uma perspectiva geral,

$$\begin{aligned} Index(z_i) &= x^T M x - x^T M \xi_i (\xi_i^T M \xi_i)^{-1} \xi_i^T M x \\ &= Index(x) - RBC_i, \end{aligned} \quad (5.51)$$

ou seja,

$$Index(x) = Index(z_i) + RBC_i. \quad (5.52)$$

Cabe ressaltar que a direção para a reconstrução da falha,  $\xi_i$ , é fator determinante para a indicação correta da contribuição das variáveis. Dunia and Joe Qin (1998) sugerem que seja extraída de uma base de dados históricos da falha. Contudo, na maioria das situações não existirá uma base de dados histórica referente a falha. Para essas ocasiões, Li et al. (2010) propuseram estimar a direção considerando uma falha que afeta múltiplas variáveis, e utilizando-se das propriedades do RBC, apresentadas pelos próprios autores, avaliar se o valor do índice da parcela reconstruída está abaixo do valor crítico. A combinação mínima de variáveis que faz o valor ficar igual ou menor ao valor crítico será o conjunto de variáveis candidatas que causaram a falha, representado por  $f^* \subset \{1, 2, \dots, n\}$ . Para evitar indeterminações, os autores substituem a inversão de matriz em (5.50) pela pseudoinversa. O Algoritmo 3 resume a abordagem proposta.

---

**Algoritmo 3:** Procedimento para seleção de variáveis candidatas responsáveis pela falha.

---

**Input:**  $x_t, \delta_{Index}^2$ .  
**Output:** Conjunto de variáveis candidatas.

```

1 begin
2    $f^* \leftarrow \emptyset$ 
3    $indice \leftarrow$  calcule o valor do índice por meio de (5.41).
4   do
5     /* reconstrua o valor índice por meio de (5.50) substituindo
6     a inversa pela pseudoinversa e selecione a variável com maior
7     contribuição */
8      $j \leftarrow \operatorname{argmax}_{i \in \{1,2,\dots,n\} \mid i \notin f^*} (RBC_i)$ 
9      $f^* \leftarrow f^* \cup j$ 
10    while  $[RBC_j \geq (indice - \delta_{Index}^2)]$ 
11  return  $f^*$ 
12 end
```

---

O percentual de contribuição da  $i$ -ésima variável no conjunto candidato utilizando o PCA será definido como:

$$Cont_i^{RBC} = 100 \times \frac{RBC_i}{\sum_{j \in f^*} RBC_j}, \quad i \in f^*, \quad (5.53)$$

e utilizando DPCA,

$$Cont_i^{RBC} = 100 \times \frac{RBC_i + \sum_{j \in lag(i)} RBC_j}{\sum_{j \in f^*} RBC_j}, \quad i \in f^*, \quad (5.54)$$

em que  $lag(i)$  representa o conjunto variáveis atrasadas da  $i$ -ésima variável que esteja contido em  $f^*$ .

## 5.3 RBC on-line

Como apresentado, o RBC indica um conjunto de variáveis que divergem da operação normal, contudo, tornando o modelo de referência não estático obtêm-se uma ferramenta de explicabilidade universal, isto é, aplicável em todo e qualquer método independentemente de sua fundamentação e estrutura.

Para dinamicamente atualizar o modelo de referência é realizado o seguinte procedimento: em uma janela  $Q$  são armazenadas as últimas  $d \in \mathbb{N}^*$  amostras associadas ao grupo/classe em que o processo encontra-se atualmente e o método evolutivo não encontrava-se no estado CNG. Na ocorrência de uma mudança de grupo/classe, cria-se o modelo de referência (*i.e.*, o modelo PCA) com a janela armazenada do/da grupo/classe

anterior para então calcular a contribuição de  $w \in \mathbb{N}^*$  amostras a partir do ponto de mudança. Uma nova janela começa a ser armazenada com amostras do/da novo/nova grupo/classe e quando ocorrer uma nova mudança o procedimento é repetido.

Dessa forma, o RBC irá indicar as variáveis que mais contribuiram para a criação e transição de grupo/classe, funcionando como um método de explicabilidade de modelos no contexto da área de Aprendizagem de Máquina. Essa informação torna-se útil ao supervisor, pois indicará o que o método evolutivo está detectando de mudança no processo que justifica a criação de um novo grupo/classe.

## **5.4 Resumo do capítulo**

O RBC tem sido amplamente utilizado na literatura em aplicações envolvendo processos industriais. Nesse capítulo, a base teórica da técnica foi apresentada de forma detalhada. Além disso, foi proposta uma nova versão para o RBC, a RBC on-line, voltada ao diagnóstico de causa de raiz de mudanças entre diferentes modos operacionais, não apenas em relação a operação normal, como executado no RBC.

# Capítulo 6

## Experimentos

Neste capítulo serão executados dois experimentos para demonstrar e avaliar as proposições desse trabalho. Inicialmente, um experimento com o *benchmark* DAMADICS é realizado para comparação de desempenho entre o eSBM e eSBM+. Por fim, um experimento com um sistema real de selagem de compressor de gases é realizado para demonstrar o uso do EMPF em um ambiente real.

### 6.1 Comparativo de desempenho entre eSBM e eSBM+

Em [Mariano and Caminhas \(2019\)](#) o eSBM é avaliado no *benchmark* DAMADICS e em um processo real de selagem de compressor; dada a característica do método no que diz respeito a relação da parametrização com o processo alvo, optou-se, com o objetivo de eliminar eventuais vieses na comparação de desempenho entre o eSBM e eSBM+, reproduzir o experimento executado pelos autores, contudo, apenas para o *benchmark*, pois é o único de domínio público, possibilitando eventuais necessidades de reprodução do experimento aqui executado. A implementação utilizada do eSBM foi a mesma do trabalho original que o propunha, ao qual foi gentilmente disponibilizada pelos autores para os testes nesse trabalho.

A avaliação será com base em algumas métricas descritas na próxima seção e pelo consumo de recursos computacionais, esse último será resumido pelo tempo médio de processamento de 30 execuções de cada amostra e pelo consumo de memória RAM. Para fins de referência, ambas as implementações estavam na linguagem de programação Python em sua versão 3.6. A máquina utilizada foi um *laptop* com processador i5-5200U 2.20GHz (5<sup>a</sup> geração) e memória RAM de 8GB DDR3 1600MHz de canal único.

### 6.1.1 Métricas de desempenho

Para avaliar o desempenho do eSBM e eSBM+ serão utilizadas algumas métricas voltadas ao contexto de detecção e classificação de falhas descritas em Vachtsevanos (2006) e Bartyś et al. (2006). Algumas métricas são calculadas a partir de uma matriz de decisão (Tabela 2). Por meio da matriz de decisão serão calculadas três métricas: Probabilidade de Detecção (POD) (do inglês, *Probability of Detection*); Probabilidade de Alarme Falso (POFA) (do inglês, *Probability of False Alarm*; e, Acurácia (ACC) (do inglês, *Accuracy*).

Saída	Falha	Operação Normal	Total
<i>Positivo (detectado)</i>	$a$ Número de falhas detectadas	$b$ Número de alarmes falsos	$a + b$ Total de casos de alarmes
<i>Negativo (não detectado)</i>	$c$ Número total de falhas não detectadas	$d$ Número de rejeições corretas	$c + d$ Total de casos sem alarmes
<b>Total</b>	$a + c$ Número total de casos de falhas	$b + d$ Número total de casos de operação normal	$a + b + c + d$ Número total de casos

Tabela 2 – Matriz de decisão para avaliação de desempenho de métodos direcionados à detecção e classificação de falhas.

Fonte: Adaptado de Vachtsevanos (2006).

- POD - Indica a probabilidade de verdadeiros positivos, sendo definida como a proporção de falhas detectadas corretamente em relação a todos os casos de falhas potenciais,

$$POD = \frac{a}{a + c}. \quad (6.1)$$

- POFA - Indica a probabilidade de falsos positivos, sendo definida como a proporção de falhas detectadas erroneamente em relação a todos os casos de operação normal,

$$POFA = \frac{b}{b + d}. \quad (6.2)$$

- ACC - Indica a probabilidade do algoritmo distinguir corretamente entre uma falha e a operação normal, sendo definida como a proporção de situações corretamente detectadas em relação a todos os casos possíveis de ocorrência,

$$ACC = \frac{a + d}{a + b + c + d}. \quad (6.3)$$

Também é desejável avaliar o quão rápido os métodos detectam uma falha em sua primeira e segunda ocorrência, ou seja, em quanto tempo o método indica a existência de um comportamento ainda não observado no processo, e após já ter modelado o comportamento, em quanto tempo o mesmo consegue identificá-lo. Para essa avaliação serão utilizadas duas métricas: Tempo de Detecção de Falha (DDT) (do inglês, *Detection Delay Time*); e, Tempo de Isolação da Falha (IDT) (do inglês, *Isolation Delay Time*),

- DDT - Representa o intervalo de tempo entre o instante que a falha ocorre pela primeira vez e sua detecção pelo método.
- IDT - Representa o intervalo de tempo entre o instante que a falha ocorre pela segunda vez e sua classificação pelo método.

Uma última métrica que será utilizada na avaliação dos métodos será a Taxa de Isolação de Falha (FIR) (do inglês, *Fault Isolation Rate*), que indica o percentual de falhas que o método foi capaz de isolar inequivocadamente, sendo definida como:

$$FIR = \frac{A}{A + B}, \quad (6.4)$$

em que  $A$  representa o total de falhas detectadas e classificadas corretamente, e  $B$  o total de falhas detectadas, mas classificadas incorretamente. No experimento que será realizado na próxima seção, também foi considerado a operação normal no cálculo dessa métrica, visto que os métodos em questão são evolutivos com aprendizagem não supervisionada, de forma que a operação normal está passível das mesmas limitações de modelagem das operações que representam falhas.

### 6.1.2 Experimento com *benchmark* DAMADICS

O *benchmark* em foco simula condições reais de operação de uma fábrica de açúcar polonesa, sua autoria é de um grupo de pesquisa europeu chamado de DAMADICS (do inglês, *Development and Application of Methods for Actuator Diagnosis in Industrial Control Systems*). O *benchmark* simula um atuador que controla o fluxo de entrada de água em uma caldeira, que por sua vez faz parte de um processo de 5 estágios de evaporação na produção de açúcar. O atuador é composto por uma válvula de controle, por um servo motor pneumático e por um posicionador (Figura 6.1). A válvula controla o fluxo de água que passa por um cano que alimenta a caldeira. O servo motor pneumático realiza variações da válvula, agindo na taxa de fluxo de água. O servo motor é composto por um diafragma preenchido por um fluido, de maneira que, comprimindo o fluido, é possível realizar o deslocamento da haste do motor. A haste do motor está conectada ao plugue da válvula, controlando sua abertura. O posicionador é um dispositivo utilizado para controlar

a abertura da válvula em malha fechada por meio da haste do motor, evitando desvios causados por fontes externas ou internas, tais como: atrito, variações na pressão fornecida ao servo motor, etc. Uma descrição completa do processo pode ser encontrada em (Bartys et al., 2006), trabalho que o propõe.

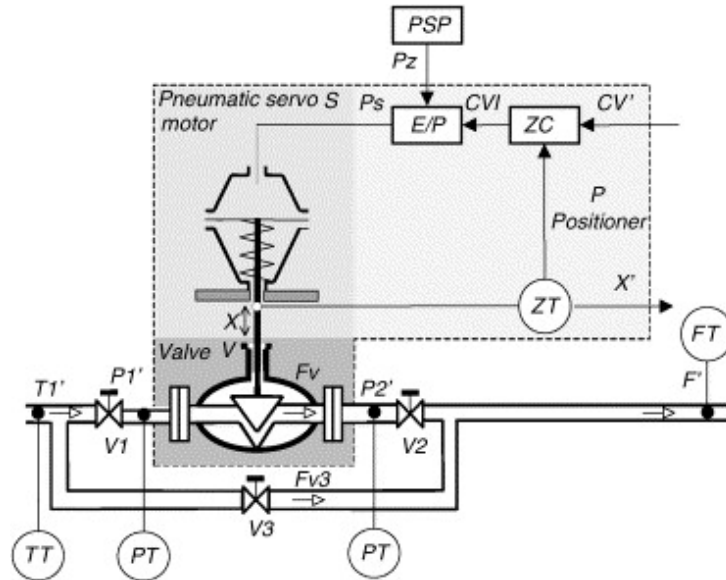


Figura 6.1 – Estrutura do sistema do atuador utilizado no *benchmark*.

A implementação do *benchmark* foi disponibilizada por seus autores em domínio público no formato de biblioteca do Simulink<sup>®</sup>. Em tentativas recentes não foi possível acessar o domínio apontado pelos autores, porém, a biblioteca já encontrava-se armazenada localmente nos computadores dos autores desse trabalho. E novamente, considerando uma eventual necessidade de reprodução dos experimentos, os arquivos do *benchmark* estão disponibilizados sob os mesmos termos do original em: <https://www.github.com/nayronmorais/DAMADICS-Benchmark>.

No *benchmark* são fornecidas 6 variáveis para o monitoramento: 1) Sinal de saída do controlador (CV); 2) Pressão de entrada da água (P1); 3) Pressão de saída da água (P2); 4) Temperatura da água (T1); 5) Fluxo médio de água (F); 6) Abertura do plug da válvula (X). Todas são sinais periódicos no intervalo  $[0, 1]$ . É possível simular um total de 20 modos operacionais, a normal (NOC) (do inglês, *Normal Operation Condition*) e 19 falhas (Tabela 3); permite também, em algumas falhas, escolher entre incipiente ou abrupta. Nessa última ainda é possível, também apenas em algumas, ter diferentes intensidades: pequena, média ou alta (Tabela 4). As intensidades são referente ao valor máximo utilizado na simulação da falha, sendo 25%, 50% e 75%, respectivamente.

No experimento original, os autores do eSBM utilizam apenas as falhas  $f_1$ ,  $f_7$  e  $f_{13}$ , de forma que as falhas abruptas,  $f_1$  e  $f_7$ , são avaliadas de forma independente da incipiente,  $f_{13}$ , pois para a detecção da falha incipiente foi necessário aumentar a sensibilidade método. Em ambos os casos, as falhas foram intercaladas com a operação normal. Para o primeiro



Falha	Localização no atuador	Descrição
$f_1$	Válvula de Controle	Entupimento
$f_2$		Sedimentação na válvula ou em sua base
$f_3$		Corrosão na válvula ou em sua base
$f_4$		Aumento do atrito na válvula
$f_5$		Vazamento externo
$f_6$		Vazamento interno
$f_7$		Fluxo crítico
$f_8$	Servo Motor	Servo motor com haste torcida
$f_9$		Agarramento da haste do servo motor
$f_{10}$		Perfuração do diafragma do servo motor
$f_{11}$		Falha na mola do servo motor
$f_{12}$	Posicionador	Falha no transdutor eletropneumático
$f_{13}$		Falha no sensor de posicionamento da haste
$f_{14}$		Falha no sensor de pressão
$f_{15}$		Falha na mola do posicionador
$f_{16}$	Geral/Externa	Queda no suprimento de pressão para o posicionador
$f_{17}$		Variação inesperada de pressão ao longo da válvula
$f_{18}$		Abertura parcial ou total da válvula <i>bypass</i>
$f_{19}$		Falha no sensor de vazão

Tabela 3 – Tipos de falhas disponíveis no *benchmark* DAMADICS.

Falha	Abrupta			Incipiente
	Pequena	Média	Alta	
$f_1$	x	x	x	-
$f_2$	-	-	x	x <sup>1)</sup>
$f_3$	-	-	-	x <sup>1)</sup>
$f_4$	-	-	-	x <sup>2)</sup>
$f_5$	-	-	-	x <sup>1)</sup>
$f_6$	-	-	-	x <sup>1)</sup>
$f_7$	x	x	x	-
$f_8$	x	x	x	-
$f_9$	-	-	-	x <sup>1)</sup>
$f_{10}$	x	x	x	-
$f_{11}$	-	-	x	x <sup>1)</sup>
$f_{12}$	x	x	x	-
$f_{13}$	x	x	x	x <sup>3)</sup>
$f_{14}$	x	x	x	-
$f_{15}$	-	-	x	-
$f_{16}$	x	x	x	-
$f_{17}$	-	-	x	x <sup>2)</sup>
$f_{18}$	x	x	x	x <sup>1)</sup>
$f_{19}$	x	x	x	-

1) : 1 dia      2) : 1 hora      3) : 15 minutos

Tabela 4 – Tipo e intensidade possível para cada falha no *benchmark* DAMADICS.

com as falhas abruptas foi utilizado a seguinte ordem: *operação normal*, falha  $f_1$ , *operação normal*, falha  $f_7$ , *operação normal*, falha  $f_1$ , *operação normal* e falha  $f_7$ . Para o segundo: *operação normal*, falha  $f_{13}$ , *operação normal* e falha  $f_{13}$ .

Ainda no mesmo contexto, o experimento com melhor resultado não utilizou as entradas originais do processo simulado, isso porque, segundo os autores, devido a natureza periódica dos sinais foram criados vários grupos ao longo do período, à vista disso, foi utilizada a média e desvio padrão de uma janela móvel de 100 amostras para cada variável. Desse modo, o conjunto de variáveis original para entrada no método foi duplicado.

Com base nisso, reproduziu-se, nesse trabalho, o experimento com melhor resultado, porém, considerando um período maior de amostras para cada modo operacional e de forma conjunta, isto é, considerando falhas abruptas e incipiente em sequência. O resultado da simulação é apresentado na Figura 6.2 mesclando a ordem dos experimentos originais. Os dados utilizados como entrada nos métodos, a média e desvio padrão móvel, são apresentados na Figura 6.3.

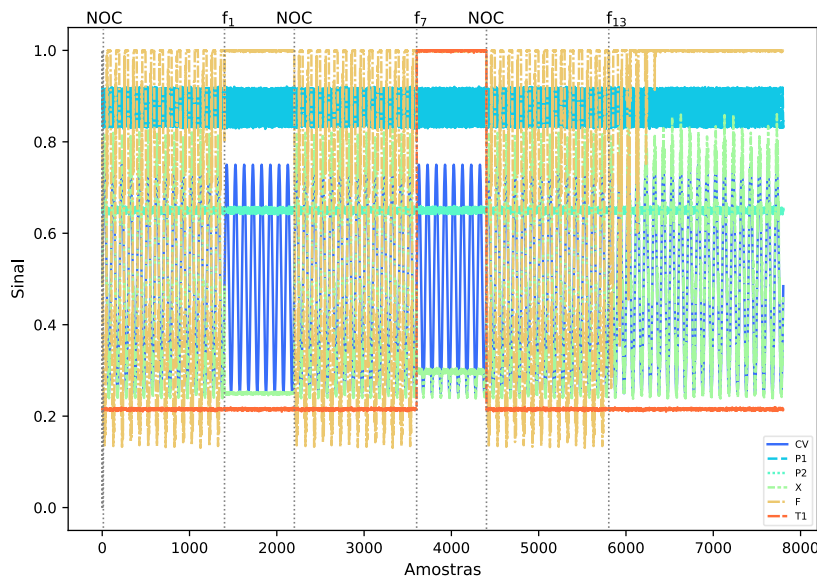


Figura 6.2 – Dados simulados do *benchmark* DAMADICS para a operação normal e para as falhas  $f_1$ ,  $f_7$  e  $f_{13}$  utilizando a biblioteca do Simulink®.

A configuração de parametrização utilizada para o eSBM (mesma do experimento original) foi:  $win_{novo} = 50$ ,  $win_{novo}^{min} = 25$ ,  $win_{adapt} = 50$ ,  $win_{adapt}^{min} = 25$ ,  $\alpha = 45$ ,  $\sigma = 0.05$ ,  $v = 0.01$ ,  $\tau = 0.5$ ,  $\mu_{min} = 0.05$ ,  $\mu_{max} = 0.5$  e distância Euclidiana. Cabe ressaltar que o valor do parâmetro  $\alpha$  apresentado não é o mesmo do experimento original, os autores indicaram o valor 25, porém, nos experimentos executados, o método não foi capaz de detectar a falha incipiente com esse nível de sensibilidade, assim, o valor foi aumentado empiricamente acrescentando 5 unidades até o mesmo ser capaz de realizar a detecção de forma semelhante a apresentada no trabalho original. Para o eSBM+ foi utilizada a mesma parametrização do exemplo ilustrativo da Seção ?? com exceção ao tamanho da

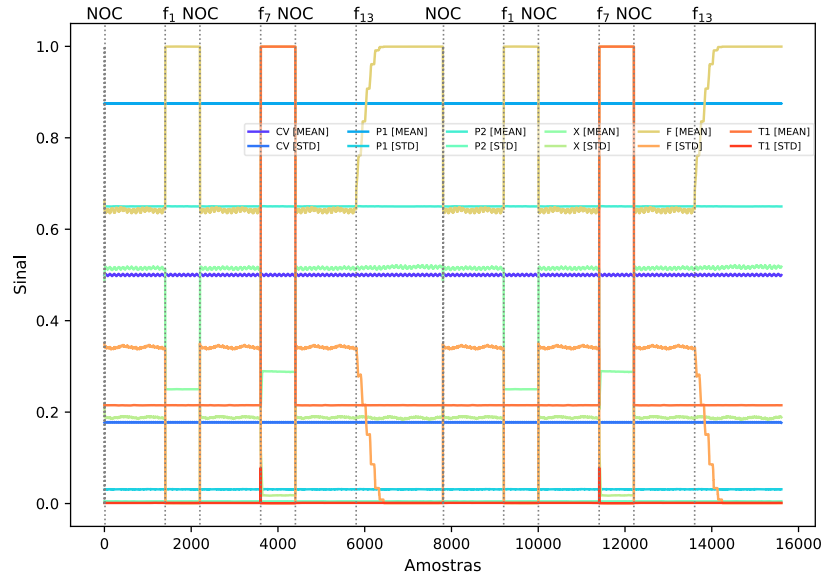


Figura 6.3 – Dados simulados duplicados e em função da média e desvio padrão móvel de cada variável.

janela de criação, que foi utilizado a dimensão dos dados:  $\gamma_{group} = 0.9$ ,  $\gamma_{point} = 0.2$ ,  $\theta = 0.1$ ,  $\tau = 1 \times 10^{-5}$ ,  $k = 12$  e operador WSF.

O resultado da aplicação de ambos os métodos é apresentado na Figura 6.4. A saída real/esperada é representada pela linha contínua na cor laranja, sendo 1 indicando a operação normal, 2 a falha  $f_1$ , 3 a falha  $f_7$  e 4 a falha  $f_{13}$ . A linha tracejada na cor verde representa a saída/classificação do eSBM. A linha pontilhada na cor azul representa a saída/classificação do eSBM+. Por meio da figura é possível perceber que ambos os métodos detectam e classificam os diferentes modos operacionais de forma satisfatória e também que o eSBM+ modela a falhas por meio de mais de um grupo, o mesmo ocorre com o eSBM para a falha  $f_{13}$ . Em uso prático, com o EMPF por exemplo, seria responsabilidade do supervisor indicar qual (is) o (s) grupo (s) representa/representam cada modo operacional, para que em ocorrências futuras de quaisquer dos grupos o supervisor saiba a qual o modo operacional e qual (is) a (s) medidas (s) a ser/serem tomada (s).

Na Tabela 5 e 6 são apresentados os valores das métricas de detecção e classificação de falhas para ambos os métodos. Por meio dos valores apresentados é possível notar que o desempenho dos métodos foi bem próximo, contudo, no eSBM+ não foi necessário quaisquer conhecimentos sobre a dinâmica das variáveis do processo, inclusive utiliza a mesma parametrização do exemplo ilustrativo que possui número de variáveis menor e comportamento diferente.

No que diz respeito ao consumo de recursos de computacionais, em geral, o eSBM+ também apresentou melhor desempenho, mantendo-o linear a quantidade de grupos. A abordagem de atualização de grupo no eSBM demonstrou ser um grande ponto negativo, pois necessita de bem mais processamento que a da eSBM+, uma vez que necessita

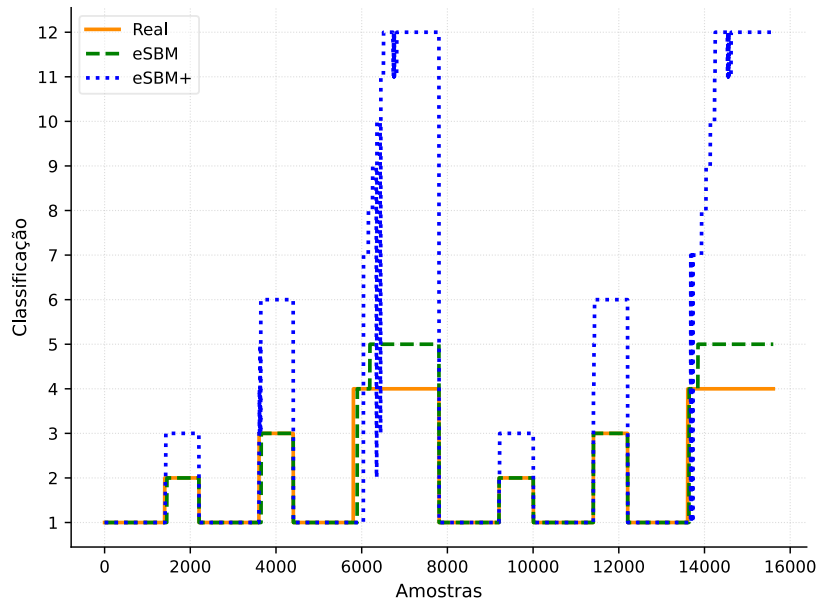


Figura 6.4 – Resultado da aplicação do eSBM e eSBM+ para o conjunto simulado.

calcular o mediana geométrica por meio de modelos de otimização, enquanto que o eSBM+ executa apenas uma avaliação de distâncias. Essa informação é corroborada por meio da Figura 6.5 que contém o tempo médio de 30 execuções de ambos os métodos, observa-se que no período referente a falha incipiente, em que há mais atualizações dos modelos, o eSBM necessitou em média de aproximadamente 40 vezes a mais de tempo que apenas no procedimento de classificação. A mediana das 30 execuções confirma o que foi dito, indicando que a média no período não foi resultado da presença de *outliers*. O consumo de memória (Figura 6.6) do eSBM+ também manteve-se inferior ao eSBM, no entanto, provavelmente o consumo mais elevado do eSBM esteja mais relacionado as estruturas de dados utilizadas na implementação, visto que o eSBM+ armazena informações a mais por grupo (*e.g.*, centro e a inversa da matriz de covariância).

	eSBM	eSBM+
<b>POD</b>	0.96	0.95
<b>POFA</b>	0.00	0.00
<b>ACC</b>	0.98	0.97
<b>FIR</b>	0.96	0.94

Tabela 5 – Métricas de detecção e classificação de falhas para o resultado apresentado pelos métodos eSBM e eSBM+.

## 6.2 Demonstração de uso do EMPF

Para demonstrar o uso do EMPF em um processo real, optou-se por utilizá-lo em um sistema de selagem de compressor de gases. Foi utilizado três métodos evolutivos além

Falha	eSBM		eSBM+	
	DDT	IDT	DDT	IDT
$f_1$	50	0	12	0
$f_7$	50	0	12	0
$f_{13}$	91	32	236	70

Tabela 6 – Métricas de tempo de detecção e classificação de falhas para o resultado apresentado pelos métodos eSBM e eSBM+.

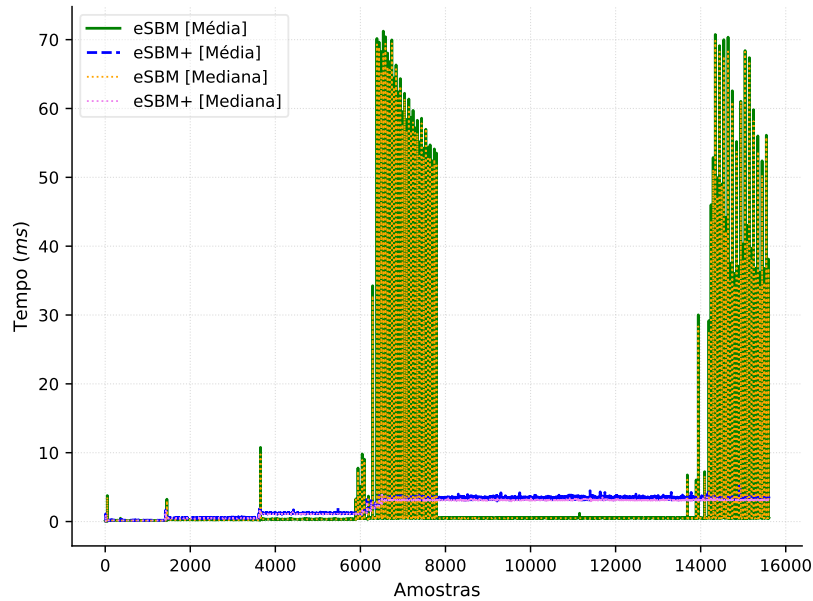


Figura 6.5 – Tempo médio e mediana (em milissegundos) de 30 execuções do eSBM e eSBM+ em função do número de amostras processadas.

do eSBM+ com abordagens distintas entre si, a saber: *Online Ellipsoidal Clustering* (OEC) (Moshtaghi et al., 2016), utiliza modelos gaussianos para modelar os grupos, um por grupo; *AutoCloud* (Bezerra et al., 2020), baseado no *framework* para análise de dados TEDA (Angelov, 2014), que por sua vez utiliza a distância média entre pontos; *MicroTEDAClus* (Maia et al., 2020), também baseado em TEDA, porém, utiliza a abordagem de macro e micro grupos, isto é, um grupo é uma composição de outros pequenos grupos. Com exceção ao MicroTEDAClus, que utilizou-se a implementação disponibilizada pelos autores com pequenas adaptações ao EMPF, todos os demais foram implementados a partir do trabalho que o propunha.

Durante o experimento foi observado que o método OEC e MicroTEDAClus, possuem limitações quanto ao uso no monitoramento contínuo de processos reais. O primeiro sofre de instabilidade numérica na inversa da matriz de covariância de um "grupo especial". No segundo há a criação excessiva de grupos, gerando muitos alarmes falsos. Algumas pequenas modificações serão sugeridas na próxima seção para tentar lidar com as questões levantadas. A seção seguinte apresenta o experimento em si, incluindo os

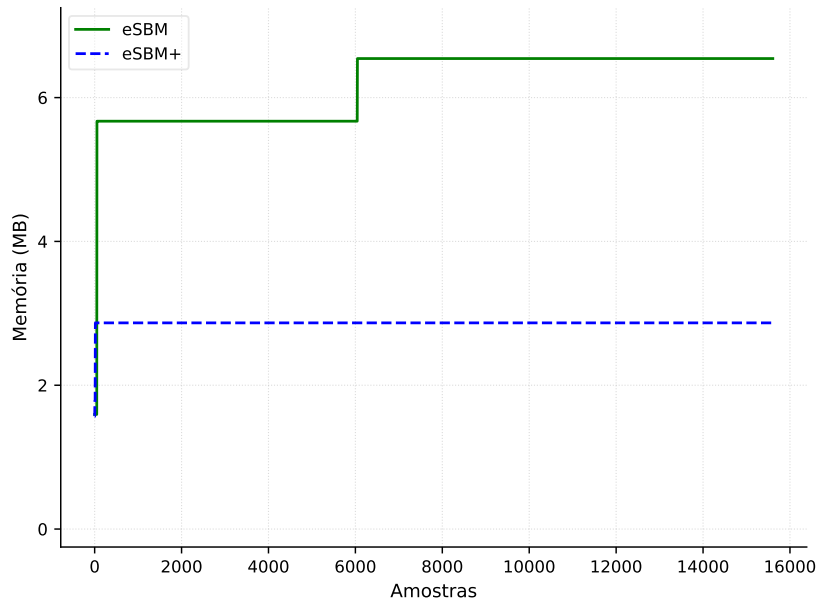


Figura 6.6 – Consumo de memória RAM (em megabytes) do eSBM e eSBM+ em função do número de amostras processadas.

resultados do OEC e MicroTEDAClus com as modificações sugeridas.

## 6.2.1 Sugestões de modificações ao OEC e MicroTEDAClus

### OEC

O OEC possui um grupo denominado *Tracker*, que desloca-se continuamente de modo a seguir a dinâmica do processo. Quando o *Tracker* está totalmente separado dos grupos existentes, considera-se que houve uma mudança da dinâmica do processo e um novo grupo é criado para representar o comportamento. A velocidade com que o *Tracker* segue a dinâmica do processo é definida por um parâmetro  $\lambda$  chamado de *Fator de Esquecimento*. O parâmetro define um *horizonte de memória* do passado que o *Tracker* deve "lembrar", de forma que se esse horizonte for muito grande, novas observações terão pouca influência no deslocamento do *Tracker*.

Em termos matemáticos, o Fator de Esquecimento define o peso de novas observações na atualização incremental do centro e da inversa da matriz de covariância do *Tracker* (Moshtaghi et al., 2016),

$$m_{t+1} = \lambda m_t + (1 - \lambda)x_{t+1}, \quad (6.5)$$

$$S_{t+1}^{-1} = \frac{tS_t^{-1}}{\lambda(t-1)} \left[ I - \frac{(x_{t+1} - m_t)(x_{t+1} - m_t)^T S_t^{-1}}{\frac{(t-1)}{\lambda} + (x_{t+1} - m_t)^T S_t^{-1}(x_{t+1} - m_t)} \right], \quad (6.6)$$

em que  $m_{t+1}$  é o novo centro,  $S_{t+1}^{-1}$  a nova inversa da matriz de covariância e  $I$  é uma matriz identidade de ordem  $n$ .

A instabilidade numérica surge quando o *horizonte de memória* tende a ser pequeno. Nos experimentos executados,  $\lambda < 0.99$  fez surgir o problema processando aproximadamente 5000 amostras. O maior problema é que os valores possíveis (*i.e.*, que evitam ou pelo menos atrasam a instabilidade) definem um *horizonte de memória* com mais de 300 amostras, que além de ser considerado muito grande para o experimento em questão com amostragem de 1 minuto, impedem que comportamentos com curto período de duração sejam capturados.

Como uma forma de contornar o problema, propõe-se substituir a inversa da matriz de covariância pela matriz de covariância, pois a instabilidade ocorre na inversão, conforme discutido em Moshtaghi et al. (2011). Em termos práticos, não há qualquer diferença no uso de uma ou outra, pois o método necessita apenas dos autovalores. Dessa forma, a matriz de covariância do *Tracker* pode ser estimada incrementalmente conforme a expressão a seguir (Moshtaghi et al., 2011):

$$S_{t+1} = \frac{\lambda(t-1)}{t} S_t + \frac{\lambda^2}{t} (x_{t+1} - m_t)(x_{t+1} - m_t)^T. \quad (6.7)$$

### MicroTEDAClus

No MicroTEDAClus há apenas um parâmetro definido pelo usuário,  $r_0$ , esse parâmetro define a diferença máxima entre o primeiro e segundo ponto utilizado para criar e atualizar um micro-grupo, respectivamente. Se o segundo ponto estiver a uma distância maior que  $r_0$ , o mesmo é utilizado para criar um novo micro-grupo e não atualizar o já existente. Segundo os autores, essa abordagem busca evitar micro-grupos com área muito grande, uma vez que o segundo ponto pode estar significativamente distante do primeiro.

Um micro-grupo é efetivamente criado quando possuir mais que 2 amostras, e sua densidade é maior que a média dos micro-grupos que compõem o mesmo macro-grupo. Se um micro-grupo não estiver sobreposto a nenhum outro, torna-se um macro-grupo. Dessa forma, podem existir grupos com apenas três pontos. Essa abordagem demonstra-se promissora, contudo, no uso prático em processos industriais, gera muitos alarmes falsos. Além disso, observou-se que o limiar ao qual define quando um ponto deve ser utilizado na atualização é único de cada micro-grupo e em função da quantidade de pontos associados ao mesmo. O limiar inicia pequeno e cresce a medida que novos pontos são associados. No entanto, esse procedimento faz gerar muitos micro-grupos, pois os micro-grupos terão sensibilidade muita alta assim que são criados, e pode ocasionar também micro-grupos com área significativamente grande.

Dessa forma, propõe-se a inserção de um parâmetro  $k$  para determinar a quantidade

mínima de pontos para a efetivação da criação de um grupo, como também a inversão da definição do limiar. A expressão a seguir é utilizada para definir o valor do limiar atualmente:

$$\frac{3}{1 + e^{-0.007(s_t^l - 100)}}, \quad (6.8)$$

em que  $s_t^l$  é a quantidade de amostras associadas ao micro-cluster  $l$  no instante  $t$ .

Para inverter o limiar, isto é, iniciar alto e diminuir a medida que novos pontos são associados, é utilizada a seguinte expressão:

$$\frac{3}{3 - e^{-0.007(s_t^l - 100)}}. \quad (6.9)$$

### 6.2.2 Experimento com sistema de selagem

Para demonstrar o uso e efetividade da estrutura proposta, simula-se o monitoramento em tempo real de um sistema real de selagem de compressor de gases por meio de dados históricos com tempo de amostragem de 1 minuto. O sistema em questão consiste em vedar o conjunto rotativo do compressor para evitar que o gás do processo interaja com o meio externo. Dentre os principais motivos de escolha e uso estão a economia no processo, segurança e proteção ambiental (Bloch and Geitner, 2012).

Existem diversos arranjos e configurações de selagem utilizados na indústria, os quatro principais são: selo de anel flutuante ou de buchas, selo de bucha guiada, selo mecânico de contato e selo seco a gás (Bloch and Geitner, 2012). Cada um tem suas vantagens e desvantagens. O compressor monitorado utiliza o selo seco a gás. Para este tipo de selo é feita a injeção de um gás limpo numa dada posição do selo. Este gás injetado irá atuar para impedir o vazamento do gás de processo através da selagem do compressor em direção aos mancais e/ou à atmosfera. Para isto, o gás de selagem deve ser fornecido em pressões acima da pressão de operação do compressor, vista pelo selo. Esta pressão deve ser alta o suficiente para que o gás de selagem flua na direção do gás de processo, porém não muito alta de forma a não haver uma vazão elevada, que pode tanto ser ruim para o processo como implica em um consumo elevado deste insumo. O gás de selagem é fornecido em condições de pressão adequada sendo controlada a diferença de pressão entre o suprimento do gás e a pressão da selagem. Parte do gás de selagem flui também através do selo em direção aos mancais sendo que a configuração do selo prevê um ponto de saída deste gás, chamado vent. Normalmente há um segundo gás de separação, muitas vezes nitrogênio, que faz uma barreira impedindo o gás de selagem de seguir para os mancais. As variações das vazões de gás de selagem e das condições do vent de gás são indicativos da saúde do selo. Uma necessidade neste processo é o de fornecimento de um gás de selagem limpo e seco para não causar danos ao sistema de selagem do



compressor. Para isto são utilizados sistemas de filtragem que preparam o gás para injeção. Os filtros do sistema devem ser monitorados visando manter sua eficiência e identificação de necessidade de limpeza ou de troca de filtro. As condições de operação do compressor também são importantes no desempenho do sistema de selagem e por isto variáveis que indicam condições operacionais e de esforço mecânico também devem ser monitoradas. Nesta aplicação foi feita, utilizando o conhecimento de especialistas em máquinas, a escolha de quais variáveis devem ser monitoradas neste processo não se fazendo a introdução de um grande número de variáveis para daí obter conhecimento.

As variáveis monitoradas estão dispostas na Tabela 7. Para simplificação, foi utilizado no experimento apenas quatro modos operacionais, em que um é a operação normal e os outros três são falhas (Tabela 8). As falhas são apresentadas nas Fig.6.7–6.9 por meio das variáveis mais afetadas, de acordo com a equipe de supervisão. A linha vermelha e pontilhada na vertical delimita o início da falha.

Variável	Descrição
$x_1$	Pressão a jusante do compressor D - CO2
$x_2$	Pressão diferencial do filtro de gás - LP
$x_3$	Pressão diferencial do vent primário - baixa pressão LP
$x_4$	Pressão diferencial do filtro 2 - alta pressão
$x_5$	Pressão diferencial do filtro de suprimento de gás 1 - alta pressão
$x_6$	Pressão diferencial do vent primário - HP
$x_7$	Pressão diferencial vent primário - NDE HP
$x_8$	Pressão primaria vazamento selo - DE HP
$x_9$	Pressão primaria vazamento selo - NDE HP
$x_{10}$	Pressão primaria vazamento selo - NDE LP
$x_{11}$	Pressão primaria vazamento selo - DE LP
$x_{12}$	Pressão de gás de suprimento - separação HP
$x_{13}$	Pressão diferencial do vent primário - LP
$x_{14}$	Pressão diferencial vent primário - NDE LP
$x_{15}$	Vibração eixo x - NDE HP
$x_{16}$	Vibração eixo x - DE HP
$x_{17}$	Vibração eixo x - DE LP
$x_{18}$	Vibração eixo x - NDE LP
$x_{19}$	Vibração eixo y - NDE HP
$x_{20}$	Vibração eixo y - DE HP
$x_{21}$	Vibração eixo y - NDE LP
$x_{22}$	Vibração eixo y - DE LP
$x_{23}$	Deslocamento axial A - HP
$x_{24}$	Deslocamento axial B - HP
$x_{25}$	Deslocamento axial A - LP

$x_{26}$	Deslocamento axial B - LP
$x_{27}$	Vazão de suprimento de gás de selagem - DE LP
$x_{28}$	Vazão de suprimento de gás de selagem - NDE LP
$x_{29}$	Pressão a montante do compressor A - CO2A
$x_{30}$	Pressão a jusante do compressor A - CO2A
$x_{31}$	Pressão a jusante do compressor B - CO2A
$x_{32}$	Pressão a montante do compressor B - CO2A
$x_{33}$	Pressão a jusante do compressor C - CO2A
$x_{34}$	Temperatura a jusante compressor A
$x_{35}$	Temperatura de entrada do terceiro estágio
$x_{36}$	Temperatura de entrada do primeiro estágio
$x_{37}$	Temperatura a jusante compressor C
$x_{38}$	Temperatura de entrada do segundo estágio

Tabela 7 – Variáveis utilizadas para o monitoramento do sistema de selagem.

ID	Principais Variáveis Relacionadas	Descrição
$f_1$	$x_2, x_{27}, x_{28}$	Desbalanceamento no nível de vazão do gás de selagem e elevação do diferencial de pressão do filtro de gás - LP
$f_2$	$x_{13}$	Vazamento no selo primário que gerou o aumento do diferencial de pressão
$f_3$	$x_8, x_9, x_{10}, x_{11}, x_{27}, x_{28}$	Elevação das pressões de selagem

Tabela 8 – Três falhas que ocorreram no sistema de selagem.

Para o experimento, optou-se por intercalar as falhas com a operação normal. O período de dados resultante é apresentado na Figura 6.10 com as devidas demarcações de início e fim das falhas. Vale ressaltar que no período de dados referente a operação normal existem distúrbios que visualmente indicam uma mudança significativa em termos de média, contudo, de acordo com a equipe de supervisão as mudanças não representam uma anormalidade no funcionamento do processo.

O monitoramento foi executado com as variáveis normalizadas pelo mínimo e máximo do equipamento, com exceção ao OEC que foi utilizado os valores originais, por especificidades do método. Para tanto, a parametrização utilizada para cada método e que forneceu melhor resultado foi:

- eSBM+:  $\gamma_{group} = 0.9, \gamma_{point} = 0.15, \theta = 0.1, \tau = 0.01 \times 10^{-5}, k = 20$  e operador WSF.
- OEC:  $\gamma_1 = 0.1, \gamma_2 = 0.2, n_s = 10$  e  $\lambda = 0.99$ .

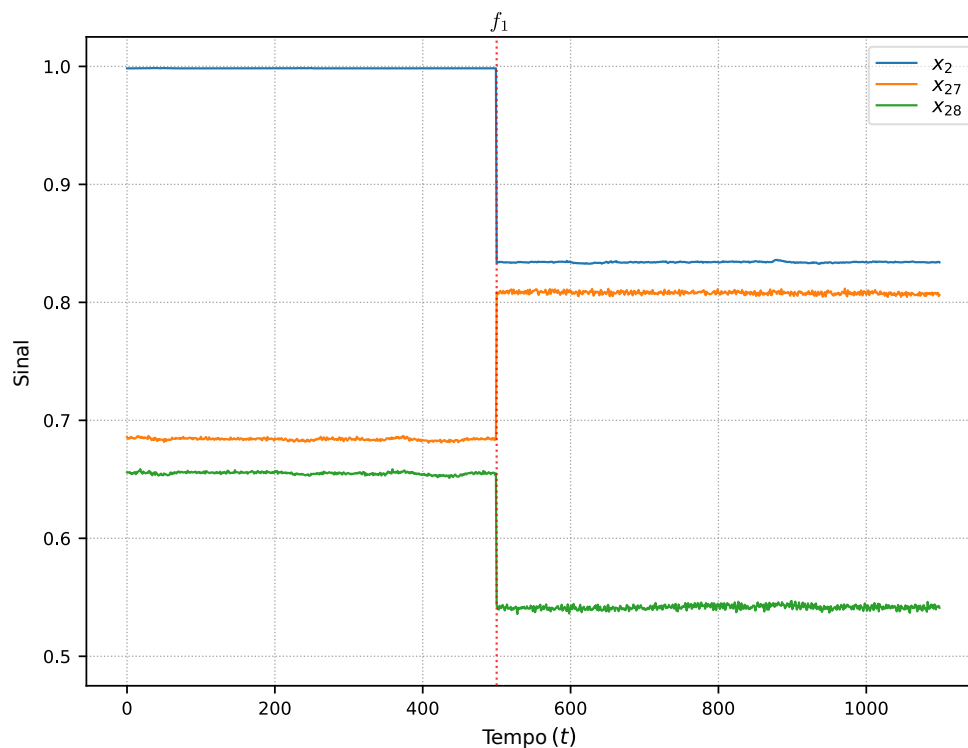


Figura 6.7 – Falha 1: Desbalanceamento no nível de vazão do gás de selagem e elevação do diferencial de pressão do filtro de gás - LP.

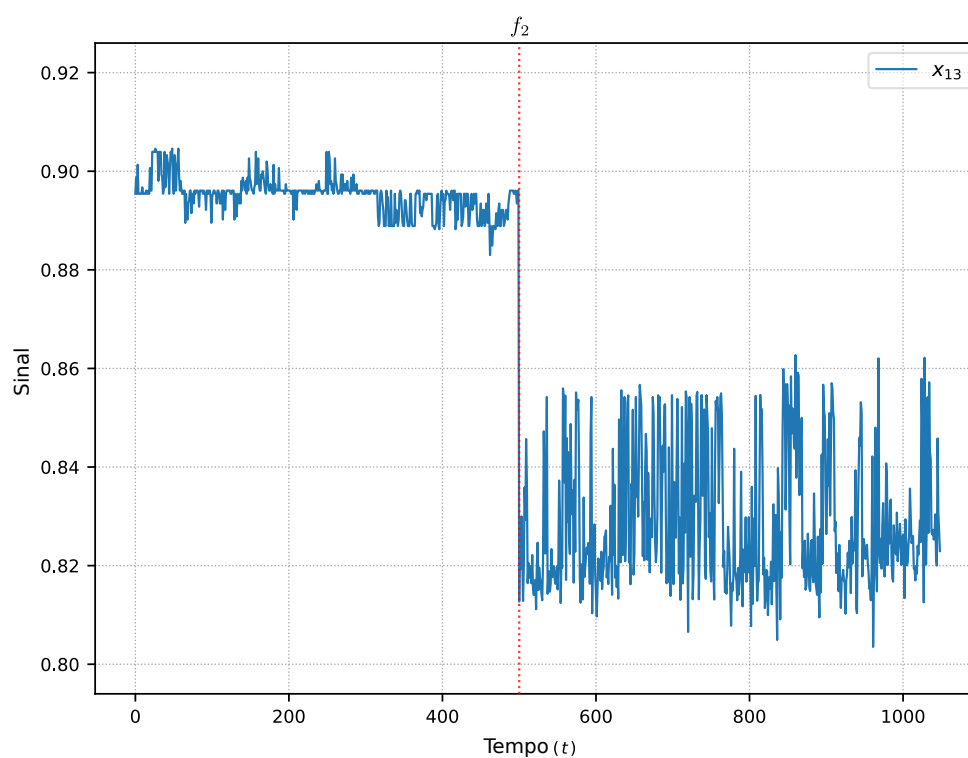


Figura 6.8 – Falha 2: Vazamento no selo primário.

- OEC modificado:  $\gamma_1 = 0.85$ ,  $\gamma_2 = 0.9$ ,  $n_s = 10$  e  $\lambda = 0.8$ .
- AutoCloud:  $m = 3$ .

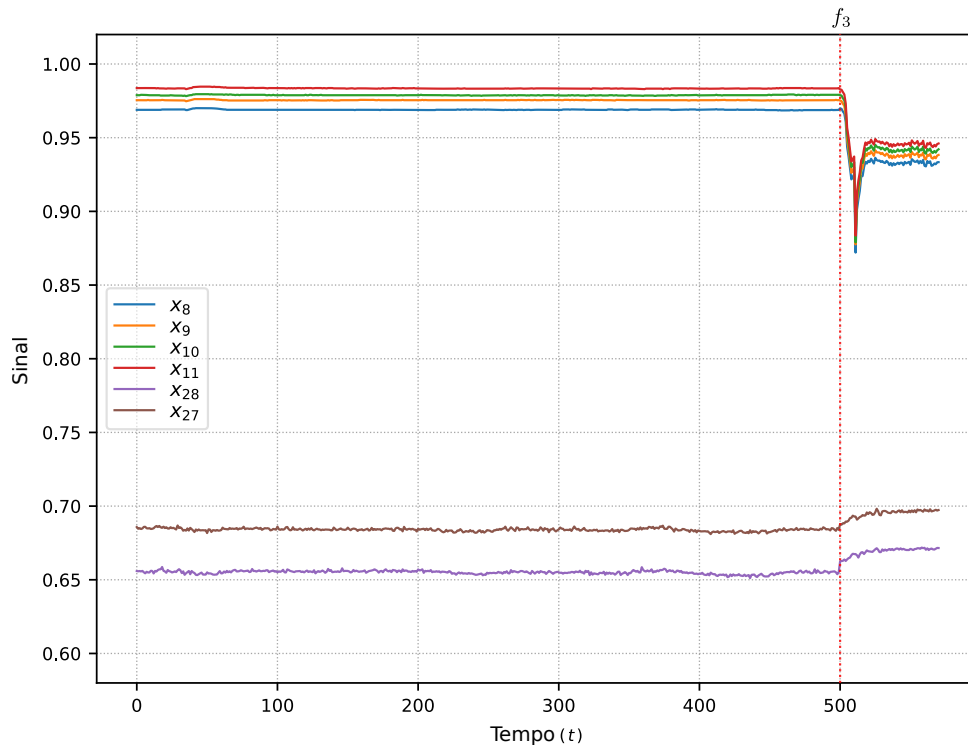


Figura 6.9 – Falha 3: Elevação das pressões de selagem.

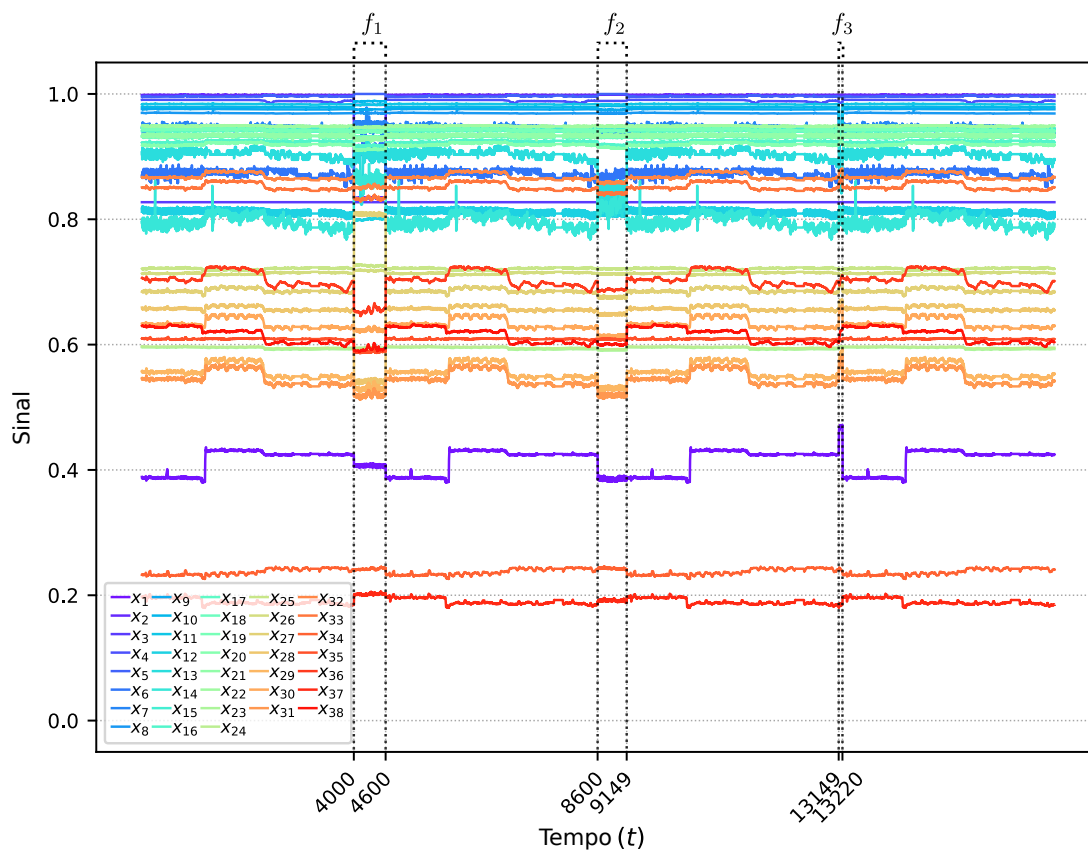


Figura 6.10 – Dados da operação normal intercalado com as falhas utilizado no experimento.

- MicroTEDAClus:  $r_0 = 0.001$ .
- MicroTEDAClus modificado:  $r_0 = 0.001$  e  $k = 10$ .
- RBC:  $d = 50$ ,  $w = 10$ ,  $\alpha = 0.99$ ,  $\alpha_{cpv} = 0.95$ ,  $l = 0$  e o índice *combined* (para todos os métodos).

O resultados obtidos são apresentados na Figura 6.11. Observa-se que os métodos capturaram razoavelmente a dinâmica do processo, contudo, apenas o eSBM+ e o Auto-Cloud conseguiram distinguir entre os quatro modos operacionais considerados. Apenas a Falha 3 não foi detectada pelo MicroTEDAClus. O OEC modificado apresentou melhores resultados que sua versão original, mas não distinguiu a Falha 2 e 3 dos demais modos operacionais. A versão modificada do MicroTEDAClus criou menos grupos, como desejado, mas não conseguiu distinguir os comportamentos relativos as Falha 2 e 3.

Nas Figuras 6.12 - 6.17 são apresentados, por simplicidade apenas no instante de criação de grupo, as contribuições de cada variável, percebe-se que para os métodos que capturaram os modos operacionais corretamente, as variáveis indicadas na Tabela 8 como principais responsáveis estão no conjunto apontado pelo RBC on-line.

## 6.3 Resumo do capítulo

Nesse capítulo foram realizados dois experimentos para validação das proposições desse trabalho, com um *benchmark* e com um sistema de selagem de compressor de gases. Inicialmente, para avaliar o conjunto de mudanças propostas ao eSBM, o eSBM e eSBM+ foram aplicados no *benchmark* DAMADICS para então calcular uma série de métricas de desempenho. Após, adicionando quatro métodos evolutivos distintos ao EMPF, o mesmo foi utilizado para monitorar um processo de selagem de compressor de gases. Em ambos os experimentos, os resultados demonstraram que as duas principais proposições desse trabalho são promissoras para o monitoramento de processos industriais.

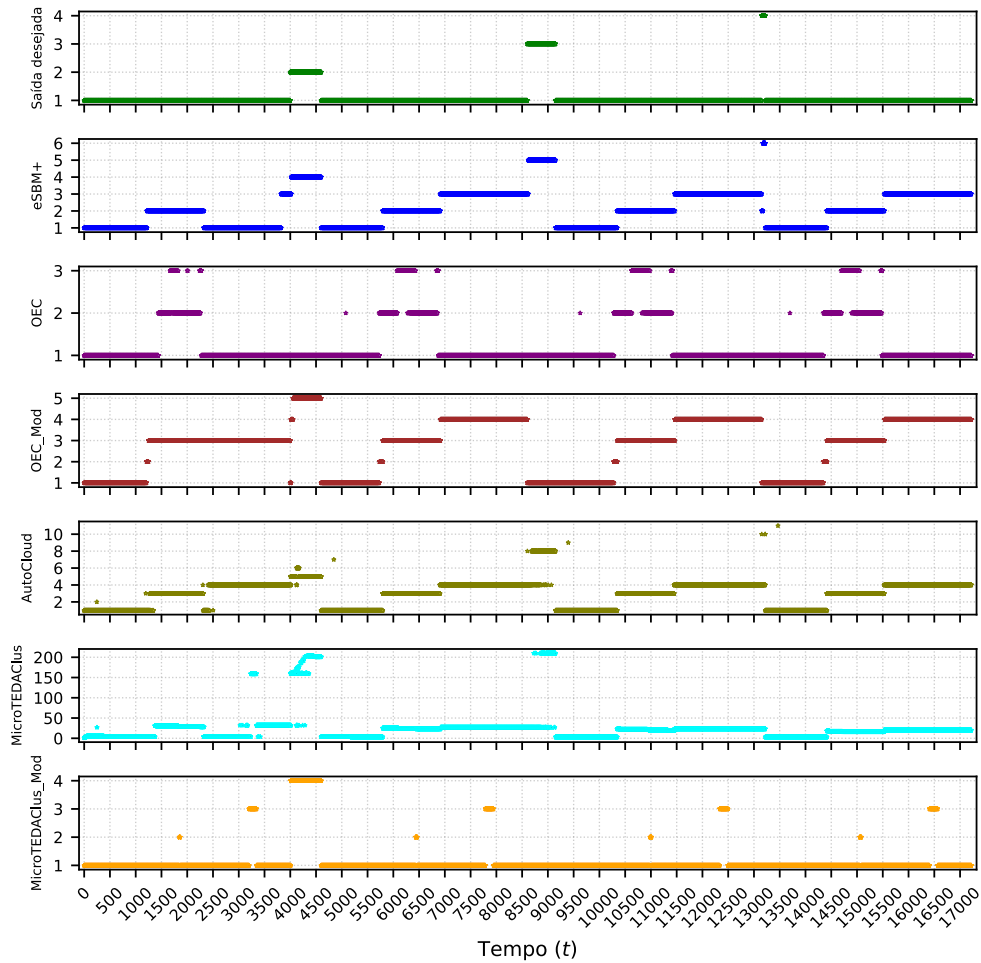


Figura 6.11 – Resultado do monitoramento do sistema de selagem com os métodos considerados no experimento.

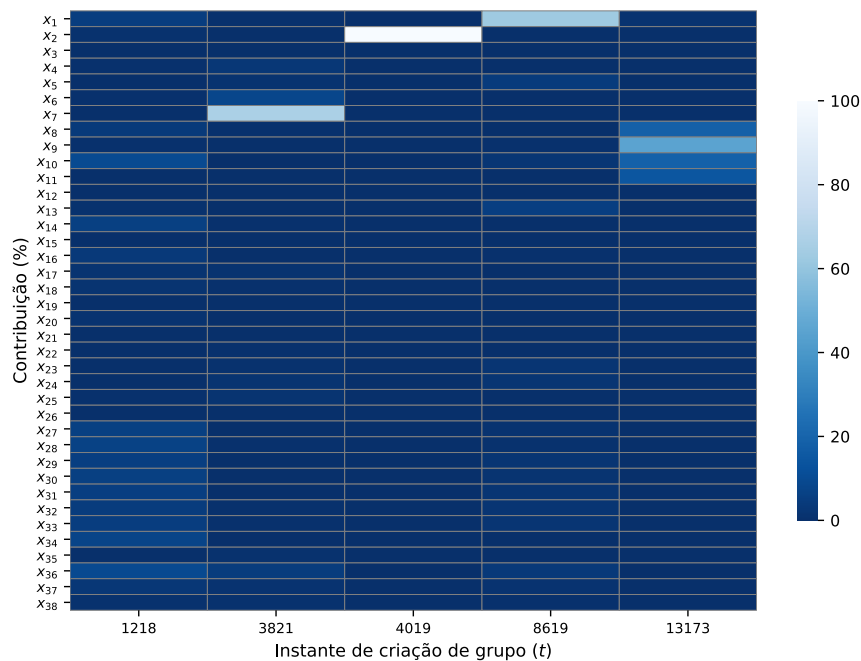


Figura 6.12 – Contribuição de cada variável para a criação de grupos pelo eSBM+.

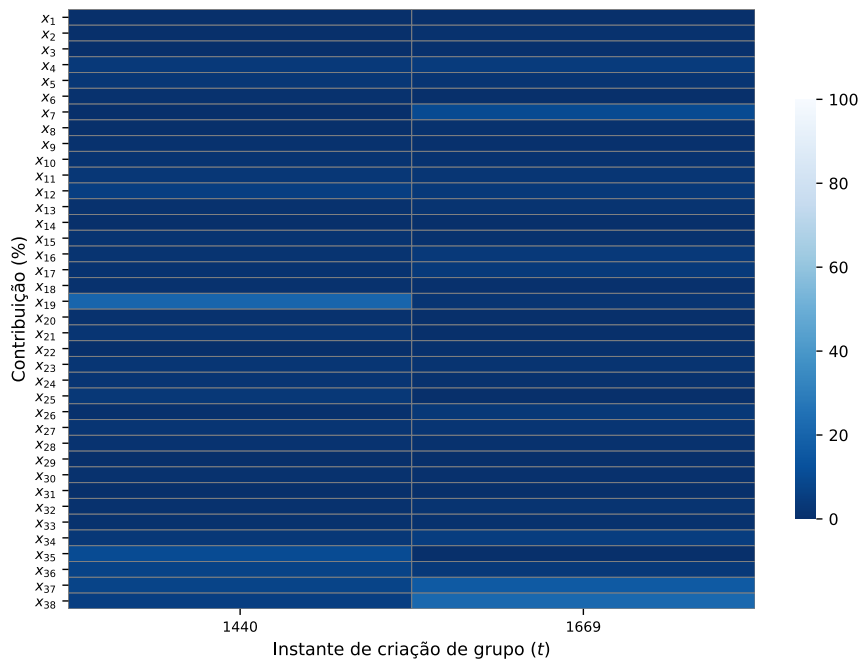


Figura 6.13 – Contribuição de cada variável para a criação de grupos pelo OEC.

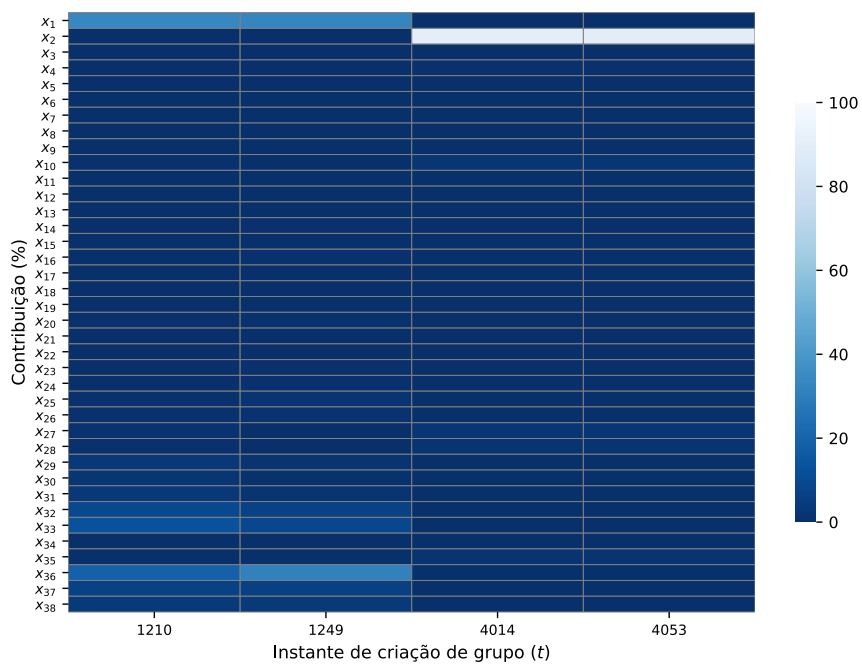


Figura 6.14 – Contribuição de cada variável para a criação de grupos pelo OEC modificado.

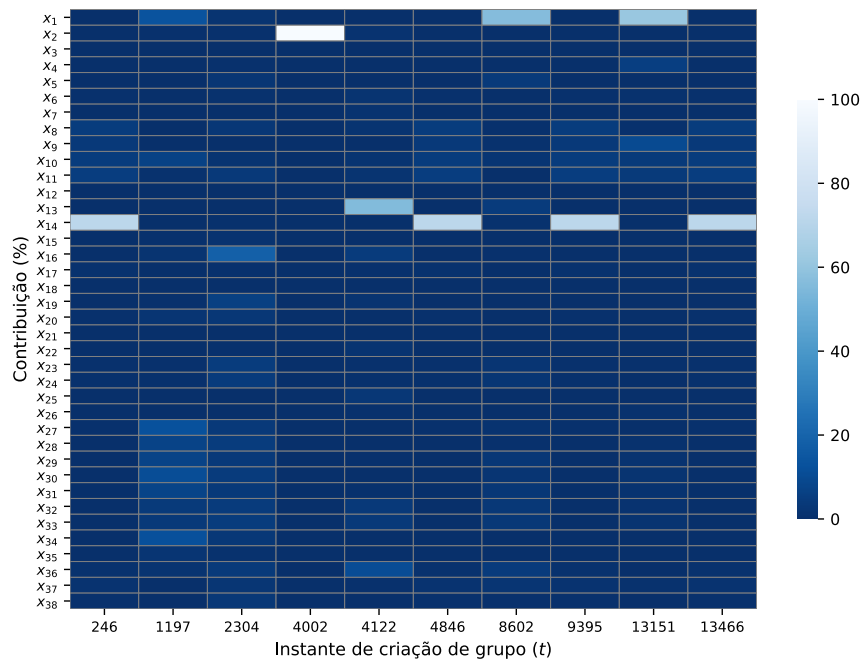


Figura 6.15 – Contribuição de cada variável para a criação de grupos pelo AutoCloud.

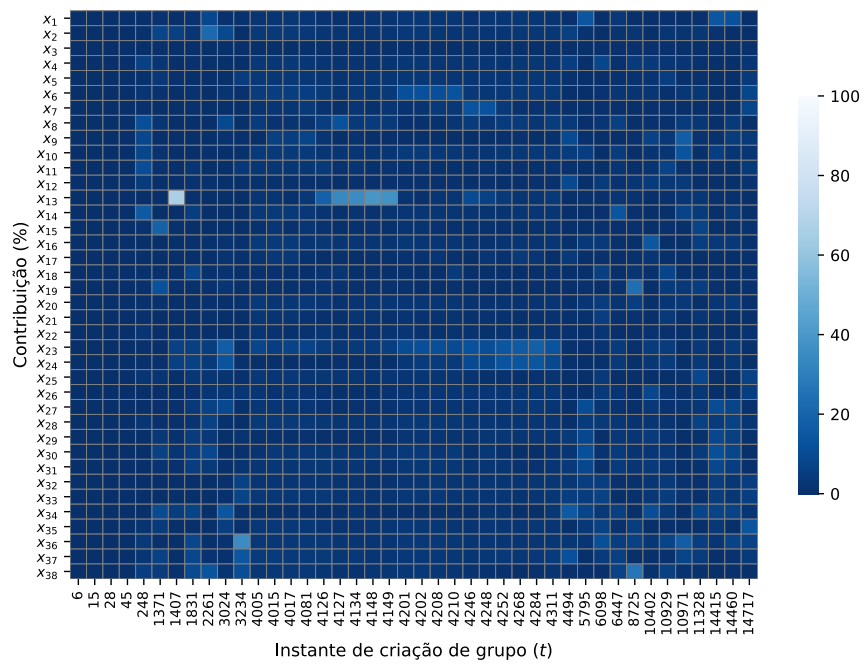


Figura 6.16 – Contribuição de cada variável para a criação de grupos pelo MicroTEDAClus.



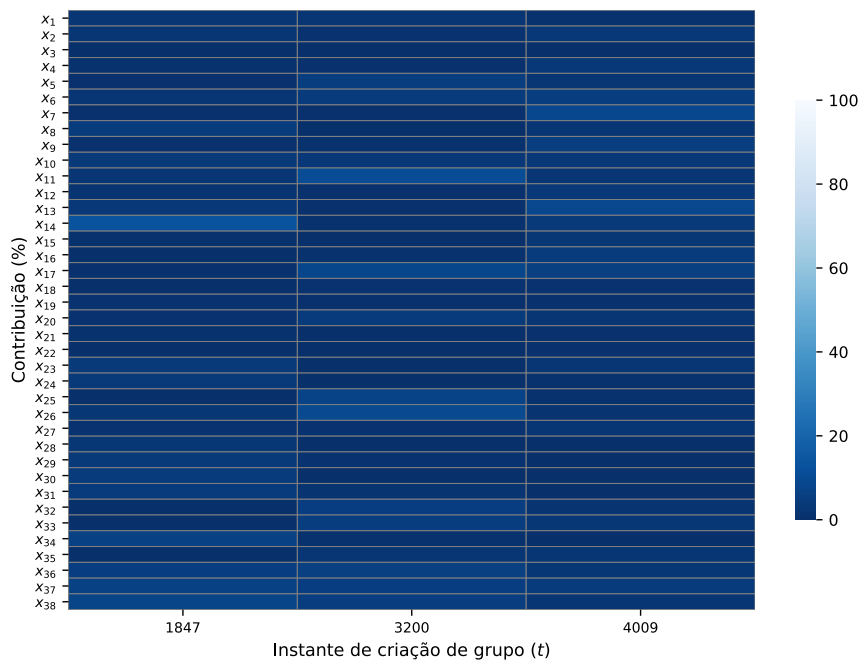


Figura 6.17 – Contribuição de cada variável para a criação de grupos pelo MicroTEDAClus modificado.



# Capítulo 7

## Conclusão

Nesse trabalho foi proposto um *framework* portátil e escalável para uso no monitoramento de processos industriais. A estrutura proposta unifica o uso das diferentes abordagens de métodos evolutivos e justifica as novidades capturadas, de forma a auxiliar o supervisor no entendimento do comportamento do processo por meio do método.

Também buscou-se implementar melhorias ao método evolutivo eSBM que maximizassem seu desempenho, reduzisse o custo computacional e facilitassem sua parametrização. A maior dificuldade no uso do método está em sua parametrização, pois é diretamente dependente do processo alvo e não há uma metodologia de fácil execução que direcione a definição dos valores de forma a maximizar o desempenho do método, mesmo com informações técnicas do processo ainda há um certo grau de dificuldade. Foi proposto então um conjunto de melhorias a fase de Estimção, Classificação, Evolução e um novo procedimento de composição da matriz memória ao eSBM.

Por meio dos experimentos executados, demonstrou-se que as melhorias propostas ao eSBM, que resultaram no eSBM+, beneficiaram significativamente o desempenho do método, tanto no que diz respeito a detecção, classificação e consumo de recursos computacionais quanto a facilidade de parametrização. Nessa última ainda houve uma redução expressiva no número de parâmetros que necessitam de serem ajustados pelo supervisor do processo. Embora o método crie um número elevado de grupos para modelar mudanças incipientes, esse comportamento demonstra-se útil em termos de supervisão, pois pode alertar o supervisor sobre um problema que está ocorrendo no processo e causando sua degradação, detectando-o de forma prematura ou ao menos antes de agravar-se.

Também pelos experimentos, foi possível observar, o uso do EMPF com quatro métodos evolutivos baseados em diferentes abordagens. O método de causa raiz indicou, em todos os métodos que houve criação de grupos nos instantes que era esperado, variáveis que pertenciam ao conjunto de variáveis que foram indicadas pela equipe de supervisão como as principais responsáveis pela falha, justificando a criação de um novo grupo. A informação de contribuição reduz significativamente o conjunto de variáveis a serem investigadas pelo

supervisor, de forma a minimizar o esforço necessário na investigação da causa real da mudança.

## 7.1 Trabalhos Futuros

Para trabalhos futuros propõe-se estudos tanto ao eSBM+ quanto ao EMPF.

### eSBM+

- Utilizar regiões que delimitem os grupos que não possuam formato de hiper-elipsoide para melhorar a qualidade da estimativa.
- Investigar novos critérios para a junção de grupos.
- Avaliar operadores de similaridade diferentes dos apresentados aqui.

### EMPF

- Adicionar um nível de pré-processamento de dados (*e.g.* PCA, média móvel, etc.).
- Compartilhamento de informações internas dos métodos evolutivos com o RBC, principalmente em métodos que utilizam modelos gaussianos, nesse caso específico, teria-se uma conversação com alta confiabilidade entre as partes. No entanto, cabe investigação, pois são necessárias ajustes devido a natureza do PCA, que necessita das variáveis na mesma escala para estimativa correta dos autovetores. Já adiantando uma solução, normalizar os dados na entrada do método evolutivo, contudo, essa abordagem pode não ser adequada a todos os métodos, o melhor seria fazer isso no RBC de forma independente.
- Fazer com que o RBC tenha sua própria base de conhecimento sobre os métodos que é atualizada constantemente, de forma similar a um método evolutivo.

# Referências

- M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1948.
- C. F. Alcalá and S. J. Qin. Reconstruction-based contribution for process monitoring. *Automatica*, 45(7):1593–1600, 2009. doi: 10.1016/j.automatica.2009.02.027. URL <https://doi.org/10.1016/j.automatica.2009.02.027>.
- C. F. Alcalá and S. J. Qin. Reconstruction-based contribution for process monitoring with kernel principal component analysis. *Industrial & Engineering Chemistry Research*, 49(17):7849–7857, 2010. doi: 10.1021/ie9018947. URL <https://dx.doi.org/10.1021/ie9018947>.
- P. Angelov. Outside the box: an alternative data analytics framework. *Journal of Automation Mobile Robotics and Intelligent Systems*, 8(2):29–35, 2014. URL [https://doi.org/10.14313/JAMRIS\\_2-2014/16](https://doi.org/10.14313/JAMRIS_2-2014/16).
- P. Angelov, D. P. Filev, and N. Kasabov. *Evolving intelligent systems: methodology and applications*, volume 12. John Wiley & Sons, 2010.
- M. Bartyś, R. Patton, M. Syfert, S. de las Heras, and J. Quevedo. Introduction to the damadics actuator fdi benchmark study. *Control engineering practice*, 14(6):577–596, 2006.
- C. G. Bezerra, B. S. J. Costa, L. A. Guedes, and P. P. Angelov. An evolving approach to data streams clustering based on typicality and eccentricity data analytics. *Information Sciences*, 2020.
- H. P. Bloch and F. K. Geitner. *Compressors: How to Achieve High Reliability & Availability*. McGraw Hill Professional, 2012.
- G. Box. Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. *The annals of mathematical statistics*, 25(2):290–302, 1954. doi: 10.1214/aoms/1177728786. URL <https://dx.doi.org/10.1214/aoms/1177728786>.

- K. P. Burnham and D. R. Anderson. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304, 2004. doi: 10.1177/0049124104268644. URL <https://dx.doi.org/10.1177/0049124104268644>.
- H. Cardot, P. Cénac, P.-A. Zitt, et al. Efficient and fast estimation of the geometric median in hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19(1):18–43, 2013. doi: 10.3150/11-bej390. URL <https://dx.doi.org/10.3150/11-bej390>.
- L. H. Chiang, E. L. Russell, and R. D. Braatz. *Fault detection and diagnosis in industrial systems*. Springer Science & Business Media, 2000.
- H.-J. Cho and J.-K. Park. An expert system for fault section diagnosis of power systems using fuzzy relations. *IEEE transactions on power systems*, 12(1):342–348, 1997.
- R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1):1–18, 2000.
- R. Dunia and S. Joe Qin. Subspace approach to multidimensional fault identification and reconstruction. *AIChE journal*, 44(8):1813–1831, 1998. doi: 10.1002/aic.690440812. URL <https://doi.org/10.1002/aic.690440812>.
- L. M. Elshenawy and T. A. Mahmoud. Fault diagnosis of time-varying processes using modified reconstruction-based contributions. *Journal of Process Control*, 70:12–23, 2018. doi: 10.1016/j.jprocont.2018.07.017. URL <https://dx.doi.org/10.1016/j.jprocont.2018.07.017>.
- Z. Ge. Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemometrics and Intelligent Laboratory Systems*, 171:16–25, 2017.
- Z. Ge, M. Zhang, and Z. Song. Nonlinear process monitoring based on linear subspace and bayesian inference. *Journal of Process Control*, 20(5):676–688, 2010. doi: 10.1016/j.jprocont.2010.03.003. URL <https://dx.doi.org/10.1016/j.jprocont.2010.03.003>.
- G. H. Golub and C. F. Van Loan. Matrix computations, johns hopkins u. *Math. Sci., Johns Hopkins University Press, Baltimore, MD*, 1996.
- B. He, T. Chen, and X. Yang. Root cause analysis in multivariate statistical process monitoring: Integrating reconstruction-based multivariate contribution analysis with fuzzy-signed directed graphs. *Computers & Chemical Engineering*, 64:167–177, 2014. doi: 10.1016/j.compchemeng.2014.02.014. URL <https://dx.doi.org/10.1016/j.compchemeng.2014.02.014>.
- S. Heo and J. H. Lee. Fault detection and classification using artificial neural networks. *IFAC-PapersOnLine*, 51(18):470–475, 2018.

- H. Hotelling. The generalization of student's ratio. *Ann. Math. Statist.*, 2(3):360–378, 08 1931. doi: 10.1214/aoms/1177732979. URL <https://doi.org/10.1214/aoms/1177732979>.
- F. Iglésias. *A Revolução Industrial*. Editora Brasiliense S.A., 1981.
- J. E. Jackson. *A user's guide to principal components*. John Wiley & Sons, 1991. URL <https://doi.org/10.1002/0471725331>.
- J. E. Jackson and G. S. Mudholkar. Control procedures for residuals associated with principal component analysis. *Technometrics*, 21(3):341–349, 1979. doi: 10.1080/00401706.1979.10489779. URL <https://dx.doi.org/10.1080/00401706.1979.10489779>.
- Q. Jiang, X. Yan, and B. Huang. Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes. *Industrial & Engineering Chemistry Research*, 58(29):12899–12912, 2019.
- S. Joe Qin. Statistical process monitoring: basics and beyond. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(8-9):480–502, 2003.
- R. A. Johnson, D. W. Wichern, et al. *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ, 2002.
- P. M. Kelly. An algorithm for merging hyperellipsoidal clusters. *Los Alamos National Laboratory, Tech. Rep*, 1994.
- W. Ku, R. H. Storer, and C. Georgakis. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 30(1): 179–196, 1995. doi: 10.1016/0169-7439(95)00076-3. URL [https://doi.org/10.1016/0169-7439\(95\)00076-3](https://doi.org/10.1016/0169-7439(95)00076-3).
- A. Lemos, W. Caminhas, and F. Gomide. Adaptive fault detection and diagnosis using an evolving fuzzy classifier. *Information Sciences*, 220:64–85, 2013.
- G. Li, C. F. Alcalá, S. J. Qin, and D. Zhou. Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the tennessee eastman process. *IEEE transactions on control systems technology*, 19(5):1114–1127, 2010. doi: 10.1109/TCST.2010.2071415. URL <https://dx.doi.org/10.1109/TCST.2010.2071415>.
- G. Li, S. J. Qin, and T. Chai. Multi-directional reconstruction based contributions for root-cause diagnosis of dynamic processes. In *2014 American Control Conference*, pages 3500–3505. IEEE, 2014. doi: 10.1109/ACC.2014.6859002. URL <https://dx.doi.org/10.1109/ACC.2014.6859002>.
- E. Lughofer. *Evolving fuzzy systems-methodologies, advanced concepts and applications*, volume 53. Springer, 2011.

- J. Maia, C. A. S. Junior, F. G. Guimarães, C. L. de Castro, A. P. Lemos, J. C. F. Galindo, and M. W. Cohen. Evolving clustering algorithm based on mixture of typicalities for stream data mining. *Future Generation Computer Systems*, 106:672–684, 2020. URL <https://doi.org/10.1016/j.future.2020.01.017>.
- D. Mariano and W. M. Caminhas. Desenvolvimento e aplicação de uma técnica de detecção de falhas baseada em predição por similaridade de modelos em processos e equipamentos, 2019. Monografia (Graduação em Engenharia de Controle e Automação), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte-MG, Brazil.
- M. A. Marins, F. M. Ribeiro, S. L. Netto, and E. A. da Silva. Improved similarity-based modeling for the classification of rotating-machine failures. *Journal of the Franklin Institute*, 355(4):1913–1930, 2018.
- L. F. Mendonça, J. Sousa, and J. S. da Costa. An architecture for fault detection and isolation based on fuzzy methods. *Expert systems with applications*, 36(2):1092–1104, 2009.
- G. Monkman and S. Wegerich. Similarity-based modeling of vibration features for fault detection and identification. *Sensor Review*, 2005.
- D. C. Montgomery. *Introduction to statistical quality control*. John Wiley & Sons, 2007.
- M. Moshtaghi, C. Leckie, S. Karunasekera, J. C. Bezdek, S. Rajasegarar, and M. Palaniswami. Incremental elliptical boundary estimation for anomaly detection in wireless sensor networks. In *2011 IEEE 11th international conference on data mining*, pages 467–476. IEEE, 2011. doi: 10.1109/icdm.2011.80. URL <https://dx.doi.org/10.1109/icdm.2011.80>.
- M. Moshtaghi, C. Leckie, and J. C. Bezdek. Online clustering of multivariate time-series. In *Proceedings of the 2016 SIAM international conference on data mining*, pages 360–368. SIAM, 2016.
- M. Namdari and H. Jazayeri-Rad. Incipient fault diagnosis using support vector machines based on monitoring continuous decision functions. *Engineering Applications of Artificial Intelligence*, 28:22–35, 2014.
- A. Perez, F. Jaramillo, V. Quintero, and M. Orchard. Characterizing the degradation process of lithium-ion batteries using a similarity-based-modeling approach. In *PHM Society European Conference*, volume 4, 2018.
- A. Perez, H. Rozas, F. Jaramillo, V. Quintero, and M. Orchard. A simulation engine for the characterization of capacity degradation processes in lithium-ion batteries undergoing heterogeneous operating conditions. In *Proceedings of the Annual Conference of the PHM Society*, volume 11, 2019.



- R. M. Pipke. Residual based monitoring of human health, Aug. 5 2014. US Patent 8,795,170.
- B. Samanta. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical systems and signal processing*, 18(3): 625–644, 2004.
- R. Singer, K. Gross, J. Herzog, R. King, and S. Wegerich. Model-based nuclear power plant monitoring and fault detection: Theoretical foundations. In *Proceedings of the International Conference on Intelligent Systems Applications to Power Systems*, 6 1997.
- D. M. Tax and R. P. Duin. Data description in subspaces. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 2, pages 672–675. IEEE, 2000.
- F. A. Tobar, L. Yacher, R. Paredes, and M. E. Orchard. Anomaly detection in power generation plants using similarity-based modeling and multivariate analysis. In *Proceedings of the 2011 American Control Conference*, pages 1940–1945. IEEE, 2011.
- G. Z. Vachtsevanos. *Intelligent fault diagnosis and prognosis for engineering systems*. John Wiley & Sons, 2006.
- V. Venkatasubramanian, R. Vaidyanathan, and Y. Yamamoto. Process fault detection and diagnosis using neural networks—i. steady-state processes. *Computers & Chemical Engineering*, 14(7):699–712, 1990.
- V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin. A review of process fault detection and diagnosis: Part iii: Process history based methods. *Computers & chemical engineering*, 27(3):327–346, 2003a.
- V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers & chemical engineering*, 27(3):293–311, 2003b.
- L. Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.
- S. Wegerich, A. Wilks, and R. Pipke. Nonparametric modeling of vibration signal features for equipment health monitoring. In *Proceedings of the IEEE Aerospace Conference*, volume 7, pages 3113–3121. Citeseer, 2003.
- S. W. Wegerich. Similarity based modeling of time synchronous averaged vibration signals for machinery health monitoring. In *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No. 04TH8720)*, volume 6, pages 3654–3662. IEEE, 2004.
- H. Yue and S. Qin. Fault reconstruction and identification for industrial processes. In *AIChE Annual Meeting, Miami, FL*, 1998.

- H. H. Yue and S. J. Qin. Reconstruction-based fault identification using a combined index. *Industrial & engineering chemistry research*, 40(20):4403–4414, 2001. doi: 10.1021/ie000141+. URL <https://dx.doi.org/10.1021/ie000141+>.
- L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan, and M. Wei. A review on deep learning applications in prognostics and health management. *IEEE Access*, 7:162415–162438, 2019.
- R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 115:213–237, 2019.
- J. Zhu, Z. Ge, and Z. Song. Distributed parallel pca for modeling and monitoring of large-scale plant-wide processes with big data. *IEEE Transactions on Industrial Informatics*, 13(4):1877–1885, 2017. doi: 10.1109/TII.2017.2658732. URL <https://dx.doi.org/10.1109/TII.2017.2658732>.