

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

**DETECÇÃO E DIAGNÓSTICO DE FALHAS BASEADO  
EM MODELOS EMPÍRICOS NO SUBESPAÇO DAS  
VARIÁVEIS DE PROCESSO (EMPVSUB)**

DISSERTAÇÃO DE MESTRADO

*María Eugenia Hidalgo Bastidas*

**Porto Alegre**

**2018**



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

**DETECÇÃO E DIAGNÓSTICO DE FALHAS BASEADO  
EM MODELOS EMPÍRICOS NO SUBESPAÇO DAS  
VARIÁVEIS DE PROCESSO (EMPVSUB)**

*María Eugenia Hidalgo Bastidas*

Dissertação de Mestrado apresentada como  
requisito parcial para obtenção do título de  
Mestre em Engenharia

*Área de concentração:* Pesquisa e  
Desenvolvimento de Processos

*Linha de Pesquisa:* Projeto, Simulação,  
Modelagem, Controle e Otimização de Processo  
Químico.

**Orientador:**  
**Prof. Dr. Jorge Otávio Trierweiler**

**Co-orientador:**  
**Prof. Dr. Marcelo Farenzena**

**Porto Alegre**

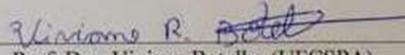
**2018**

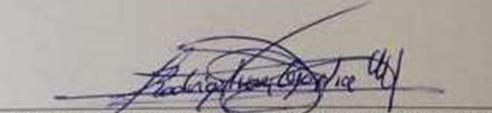


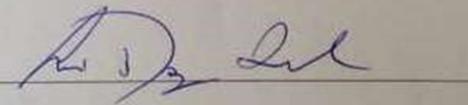
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

A Comissão Examinadora, abaixo assinada, aprova a Dissertação Detecção e diagnóstico de falhas baseado em modelos empíricos no subespaço das variáveis de processo (EMPVSub), elaborada por Maria Eugenia Hidalgo Bastidas, como requisito parcial para obtenção do Grau de Mestre em Engenharia.

Comissão Examinadora:

  
Prof. Dra. Viviane Botelho (UFCSPA)

  
Prof. Dr. Rodrigo Ivan Goytia Mejia (UNISINOS)

  
Dr. André Domingues Quelhas (PETROBRAS)



## Resumo

O escopo desta dissertação é o desenvolvimento de uma metodologia para a detecção e diagnóstico de falhas em processos industriais baseado em modelos empíricos no subespaço das variáveis do processo com expansão não linear das bases. A detecção e o diagnóstico de falhas são fundamentais para aumentar a segurança, confiabilidade e lucratividade de processos industriais. Métodos qualitativos, quantitativos e baseados em dados históricos do processo têm sido estudados amplamente. Para demonstrar as vantagens da metodologia proposta, ela será comparada com duas metodologias consideradas padrão, uma baseada em Análise de Componentes Principais (PCA) e a outra baseada em Mínimos Quadrados Parciais (PLS). Dois estudos de casos são empregados nessa comparação. O primeiro consiste em um tanque de aquecimento com mistura e o segundo contempla o estudo de caso do processo da *Tennessee Eastman*. As vantagens da metodologia proposta consistem na redução da dimensionalidade dos dados a serem usados para um diagnóstico adequado, além de detectar efetivamente a anormalidade e identificar as variáveis mais relacionadas à falha, permitindo um melhor diagnóstico. Além disso, devido à expansão das bases dos modelos é possível trabalhar efetivamente com sistemas não lineares, através de funções polinomiais e exponenciais dentro do modelo. Adicionalmente o trabalho contém uma metodologia de validação dos resultados da metodologia proposta, que consiste na eliminação das variáveis do melhor modelo obtido pelos Modelos Empíricos, através do método *Backward Elimination*. A metodologia proposta forneceu bons resultados na área do diagnóstico de falhas: conseguiu-se uma grande diminuição da dimensionalidade nos sistemas estudados em até 93,55%, bem como uma correta detecção de anormalidades e permitiu a determinação das variáveis mais relacionadas às anormalidades do processo. As comparações feitas com as metodologias padrões permitiram demonstrar que a metodologia proposta tem resultados superiores, pois consegue detectar as anormalidades em um espaço dimensional reduzido, detectando comportamentos não lineares e diminuindo incertezas.

**Palavras Chave:** diagnóstico de falhas, análise de componentes principais, mínimos quadrados parciais, regressão linear com regularização, modelos empíricos, *backward elimination*.



## Abstract

Fault detection and diagnosis are critical to increasing the safety, reliability, and profitability of industrial processes. Qualitative and quantitative methods and process historical data have been extensively studied. This article proposes a methodology for fault detection and diagnosis, based on historical data of processes and the creation of empirical models with the expansion of nonlinear bases (polynomial and exponential bases) and regularization techniques. To demonstrate the advantages of the proposed approach, it is compared with two standard methodologies: Principal Components Analysis (PCA) and the Partial Least Squares (PLS), performed in two case studies: a mixed heating tank and the Tennessee Eastman Process. The advantages of the proposed methodology are the reduction of the dimensionality of the data used, in addition to the effective detection of abnormalities, identifying the variables most related to the fault. Furthermore, the work contains a methodology to validate the diagnosis results consisting of variable elimination from the best empirical models with the Backward Elimination algorithm. The proposed methodology achieved a promising performance, since it can decrease the dimensionality of the studied systems up to 93.55%, reducing uncertainties, and capturing nonlinear behaviors.

**Keywords:** fault diagnosis, linear regression with regularization, empirical models, backward elimination.



*“An analyst knows they have achieved perfection, not when there is nothing left to add to the model, but when there is nothing left to take away”*

*“Um analista conhece que atingiu a perfeição, não quando não se tem nada mais para agregar ao modelo, mas sim quando não se tem nada mais para ser removido dele”.*

*(Antoine de Saint-Exupéry)*



## Agradecimentos

Agradeço ao Programa de Estudante-Convênio de Pós-Graduação (PEC-PG) administrado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de mestrado para a realização dos estudos na IES Brasileira Universidade Federal do Rio Grande do Sul (UFRGS), sem a qual não tivesse sido possível.

Aos meus orientadores.

Aos meus pais, Pedro e Auxiliadora, e ao meu namorado Fidel, por sempre me apoiar e serem minha inspiração.

Aos colegas do GIMSCOP, especialmente a Nelzy, Bruno, Leonardo e Lucas, pela ajuda, amizade e bons momentos compartilhados.

Às colegas do primeiro semestre do mestrado Vitória, Leticia e Adriane, pela amizade em aqueles momentos de tensão durante as disciplinas obrigatórias e depois.

Ao secretário do PPGEQ, Patrício, e a secretária do Departamento de Engenharia Química, Dóris por serem as pessoas mais profissionais da UFRGS que sempre me ajudaram nos momentos que precisei, com muito carinho e paciência.

A minha universidade: Universidad Central de Venezuela (UCV) e ao meu país VENEZUELA por me dar todas as ferramentas necessárias para atingir os meus objetivos durante este mestrado acadêmico apesar dos inconvenientes em geral.

A UFRGS pela oportunidade de me aperfeiçoar como profissional e pessoa.

Em suma, agradeço a todos aqueles que de alguma forma contribuíram com este trabalho.



## SUMÁRIO

<b>Capítulo 1 – Introdução .....</b>	<b>1</b>
1.1 Objetivos do trabalho .....	3
1.2 Estrutura da dissertação.....	4
<b>Capítulo 2 – Revisão Bibliográfica .....</b>	<b>5</b>
<b>Capítulo 3 – Fundamentos.....</b>	<b>10</b>
3.1 Classificação dos Métodos de Diagnóstico de Falhas .....	10
3.2 Métodos baseados em dados históricos do processo para diagnóstico de falhas ....	11
3.2.1 Mínimos Quadrados Ordinários (OLS) .....	11
3.2.2 Mínimos Quadrados Parciais (PLS) .....	11
3.2.3 Análise de Componentes Principais (PCA).....	13
3.2.4 Análise de Contribuições.....	14
3.2.5 Regressão Linear com Regularização.....	14
3.2.6 Regressão Ridge .....	16
3.2.7 Regressão Lasso .....	16
3.2.8 Regressão Lars.....	17
3.2.9 Regularização Lasso Lars .....	17
3.3 Seleção de Variáveis .....	17
3.3.1 Ant Colony Optimization (ACO) .....	17
3.3.2 Forward Selection.....	19
3.3.3 Backward Elimination .....	19
3.3.4 Stepwise procedure.....	19
3.4 Métricas de Ajuste de Dados .....	19
<b>Capítulo 4 – Metodologia .....</b>	<b>21</b>
4.1 Metodologia Padrão .....	21
4.1.1 Análise de Componentes Principais (PCA).....	21
4.1.2 Análise de Contribuições.....	22
4.1.3 Mínimos Quadrados Parciais (PLS) .....	23
4.2 Metodologia Proposta (EMPVSub) .....	24
4.3 Metodologia de validação (Eliminação de Variáveis) .....	26
4.4 Estudos de caso.....	26
4.4.1 Estudo de Caso 01: Tanque de Aquecimento com Mistura.....	27
4.4.2 Estudo de Caso 02: Tennessee Eastman Process (TEP).....	29
4.5 Ferramentas Computacionais .....	33
<b>Capítulo 5 – Resultados .....</b>	<b>34</b>
5.1 Estudo de Caso 01: Tanque de Aquecimento com Mistura .....	34
5.1.1 Metodologia Padrão 01: Análise de Componentes Principais (PCA) .....	34
5.1.2 Metodologia Padrão 02: Partial Least Squares (PLS) .....	38
5.1.3 Metodologia Proposta: Modelos Empíricos .....	40
5.1.4 Metodologia de Validação: Eliminação de Variáveis.....	45
5.2 Estudo de Caso 02: Tennessee Eastman Process (TEP) .....	48
5.2.1 Metodologia Padrão 01: Análise de Componentes Principais (PCA) .....	48
5.2.2 Metodologia Padrão 02: Partial Least Squares (PLS) .....	51
5.2.3 Metodologia Proposta: Modelos Empíricos .....	55
5.2.4 Metodologia de Validação: Eliminação de Variáveis.....	61
<b>Capítulo 6 – Considerações e Trabalhos Futuros .....</b>	<b>63</b>
6.1 Principais conclusões .....	63

6.2 Sugestões para trabalhos futuros.....	64
<b>Referências.....</b>	<b>65</b>
<b>Apêndice A.....</b>	<b>69</b>
Apêndice A.1 Modelo Matemático do Caso de Estudo 01: Tanque de Aquecimento com Mistura .....	69

## LISTA DE FIGURAS

Figura 4.1: Esquema da Metodologia da Análise de Componentes Principais (PCA) e Análise de Contribuições .....	23
Figura 4.2: Esquema da Metodologia dos Mínimos Quadrados Parciais (PLS) .....	24
Figura 4.3: Esquema da Metodologia EMPVSub .....	24
Figura 4.4: Caso de Estudo 01: Sistema de Tanque de Aquecimento com Mistura (Adaptado do GIMSCOP).....	27
Figura 4.5: Processo Industrial Tennessee Eastman Process (Downs e Vogel, 1993) .....	29
Figura 4.6: Reações que envolvem o TEP (Downs e Vogel, 1993) .....	30
Figura 5.1: Metodologia Padrão 01 (PCA) aplicada ao estudo de caso 01 nos cenários 01 (operação normal) e 02 (operação normal com mudanças de setpoint).....	35
Figura 5.2: Metodologia Padrão 01 (PCA) aplicada ao estudo de caso nos cenários 01 e 03 .....	36
Figura 5.3: Análise de Contribuições para o estudo de caso 01: cenários 01 e 02.....	37
Figura 5.4: Análise de Contribuições para o estudo de caso 01: cenários 01 e 03.....	37
Figura 5.5: Metodologia Padrão 02 (PLS) aplicada ao estudo de caso 01 nos cenários 01 e 02 .....	38
Figura 5.6: Metodologia Padrão 02 (PLS) aplicada ao estudo de caso 01 nos cenários 01 e 03 .....	39
Figura 5.7: Modelo Empírico 1: Fs (Cenários 01 e 02) .....	41
Figura 5.8: Modelo Empírico 6: Fs (cenários 01 e 02) .....	42
Figura 5.9: Modelo Empírico 6: T4 (cenários 01 e 02) .....	43
Figura 5.10: Modelo Empírico 6 Fs (Cenários 01 e 03) .....	44
Figura 5.11: Modelo Empírico 7: T4 (Cenários 01 e 03).....	44
Figura 5.12: Metodologia Padrão (PCA) aplicada ao estudo de caso 02 nos cenários 01 e 02 .....	48
Figura 5.13: Análise de Contribuições aplicada ao estudo de caso 02 nos cenários 01 e 02 .....	49
Figura 5.14: Metodologia Padrão (PCA) aplicada ao estudo de caso 02 nos cenários 01 e 03 .....	49
Figura 5.15: Análise de Contribuições aplicada ao estudo de caso 02 nos cenários 01 e 03 .....	50
Figura 5.16: Metodologia Padrão (PCA) aplicada ao estudo de caso 02 nos cenários 01 e 04 .....	50
Figura 5.17: Análise de Contribuições aplicada ao estudo de caso 02 nos cenários 01 e 04 .....	51
Figura 5.18: Metodologia Padrão (PLS) aplicada ao estudo de caso 02 nos cenários 01 e 02 (variável de saída XMEAS_1).....	51
Figura 5.19: Metodologia Padrão (PLS) aplicada ao estudo de caso 02 nos cenários 01 e 03 (variável de saída XMV_10) .....	52
Figura 5.20: Metodologia Padrão (PLS) aplicada ao estudo de caso 02 nos cenários 01 e 04 (variável de saída XMEAS_9).....	53
Figura 5.21: Modelo Empírico 6: XMEAS_1 (Cenários 01 e 02).....	56
Figura 5.22: Modelo Empírico 1: XMEAS_1 (Cenários 01 e 02).....	57
Figura 5.23: Modelo Empírico 1: XMV_10 (Cenários 01 e 03) .....	58
Figura 5.24: Modelo Empírico 6: XMV_10 (Cenários 01 e 03) .....	59
Figura 5.25: Modelo Empírico 1: XMEAS_9 (Cenários 01 e 04).....	60



## LISTA DE TABELAS

Tabela 4.1: Variáveis do caso de estudo 01: Tanque de Aquecimento com Mistura .....	27
Tabela 4.2: Valores das variáveis de saída do processo para o caso de estudo 01 em operação normal (cenário 01).....	28
Tabela 4.3: Valores dos distúrbios não medidos, causantes das falhas no sistema (cenário 02 e 03) para o caso de estudo 01 .....	29
Tabela 4.4: Variáveis Manipuladas do Processo (Adaptado de DOWNS E VOGEL, 1993) .....	30
Tabela 4.5: Variáveis Medidas do Processo Contínuo (Adaptado de DOWNS E VOGEL, 1993) .....	31
Tabela 4.6: Distúrbios do Tennessee Eastman Process (Adaptado de DOWNS E VOGEL, 1993) .....	32
Tabela 5.1: Parâmetros do Modelo para as variáveis de saída Fs e T4 .....	40
Tabela 5.2: Modelos Empíricos para o caso de estudo 01 a partir da ferramenta ACO .....	40
Tabela 5.3: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 1 da variável de saída Fs nos cenários 01 e 02 (i=variável contida no modelo empírico) .....	45
Tabela 5.4: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 da variável de saída Fs (i=variável contida no modelo empírico) .....	46
Tabela 5.5: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 para a variável de saída T4(i=variável contida no modelo empírico).....	46
Tabela 5.6: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 para a variável de saída Fs (i=variável contida no modelo empírico) .....	47
Tabela 5.7: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 7 para a variável de saída T4 (i=variável contida no modelo empírico).....	47
Tabela 5.8: Parâmetros do Modelo para as variáveis de saída XMEAS_1, XMV_10 e XMEAS_9 .....	53
Tabela 5.9: Modelos Empíricos para o caso de estudo 02 a partir da metodologia proposta .....	55
Tabela 5.10: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 1 da variável de saída XMEAS_1 nos cenários 01 e 02 (i=variável contida no modelo empírico) .....	61
Tabela 5.11: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 1 da variável de saída XMV_10 nos cenários 01 e 03 (i=variável contida no modelo empírico) .....	61
Tabela 5.12: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 da variável de saída XMV_10 nos cenários 01 e 03 (i=variável contida no modelo empírico) .....	62
Tabela 5.13: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 da variável de saída XMV_10 nos cenários 01 e 03 (i=variável contida no modelo empírico) .....	62



## NOTAÇÃO E SIMBOLOGIA

<i>CVA</i>	Análise da Variação Canónica
<i>PCA</i>	Análise de Componentes Principais
<i>CCA</i>	Análise de Correlação Canónica
<i>FDA</i>	Análise do Discriminante de Fisher
<i>ACO</i>	Ant Colony Optimization
$\lambda$	Autovalores da matriz S
$R^2$	Coefficiente de Determinação
<i>Cont<sub>ij</sub></i>	Contribuição de cada variável
<i>CONT<sub>j</sub></i>	Contribuição Total
$\Sigma$	Desvio-Padrão Unitário
<i>F<math>\alpha</math></i>	Distribuição de Probabilidade Contínua de Fisher
<i>F(x)</i>	Distribuição Gaussiana ou Normal
$T^2$	Estatística T square – T Hotelling
<i>GIMSCOP</i>	Grupo de Intensificação, Modelagem, Simulação e Controle de Processos
<i>KPI</i>	Indicador de Desempenho Chave
$X_N$	Matriz de Base de Dados Normalizada
$X$	Matriz de Base de Dados Original
$S$	Matriz de Covariância
$T_p$	Matriz de Projeção
$\Lambda$	Matriz que contém os autovalores $\lambda$
$V$	Matriz que contém os vetores de carga (vetores ortogonais)
$P$	Matriz que contém os vetores de carga associado ao número de componentes principais
$\mu$	Média Zero

<i>OLS</i>	Mínimos Quadrados Ordinários
<i>PLS</i>	Mínimos Quadrados Parciais
<i>A</i>	Número de componentes principais
<i>N</i>	Número de observações – Número de Filas da Matriz <i>X</i>
<i>M</i>	Número de variáveis – Número de Colunas da Matriz <i>X</i>
<i>SDV</i>	Singular Decomposition Value - Decomposição em Valores Singulares
<i>PRESS</i>	Soma dos Quadrados de Predição
<i>TEP</i>	Tennessee Eastman Process
<i>UFGRS</i>	Universidade Federal do Rio Grande do Sul
<i>T<math>\alpha</math>2</i>	Valor limite do T Hotelling
<i>Y</i>	Variável de Saída
$\tilde{X}$	Variável Normalizada
<i>X</i>	Variável observada

# Capítulo 1 – Introdução

Uma falha é definida como o desvio não permitido de pelo menos uma das propriedades características de uma variável do comportamento aceitável, bem como um sintoma ou evento anormal no processo, por exemplo, a alta temperatura em um reator ou baixa qualidade em um produto diferente da operação normal (ISERMANN, 1997; VENKATASUBRAMANIAN et al., 2003).

Segundo a UNESCO (FRANK, 2009) as falhas que acontecem em um sistema de controle automático podem ser classificadas em quatro grandes grupos: falhas no sensor, falhas no atuador, falhas de componentes (mudanças não desejadas no sistema da planta) e falhas nas unidades de controle (erros de controle causados por defeitos de hardware ou software do sistema automatizado). A ocorrência de falhas gera uma grande quantidade de riscos operacionais, que são a probabilidade da ocorrência de perdas econômicas, danos à infraestrutura e sanções legais para uma empresa, além da possibilidade de incidentes ou acidentes catastróficos, riscos ao pessoal e meio-ambiente.

É importante saber que as indústrias petroquímicas perdem pelo menos 20 bilhões de dólares a cada ano e tem qualificado os eventos anormais como seu problema principal, o qual precisa ser resolvido para evitar perdas de produtividade assim como situações de perigo (NIMMO, 1995).

Dentro dos acidentes produzidos por um deficiente diagnóstico de falhas se pode citar a explosão da Refinaria BP da Cidade de Texas no ano 2005, que teve como consequência a morte de quinze pessoas e 180 feridos, tal como perdas financeiras que superaram os 1500 milhões de dólares. Outro exemplo foi o vazamento químico tóxico de Dupont La Porte no ano 2014, o qual ocasionou o vazamento de 24 mil libras de mercaptanos por erros de comunicação, causando a morte de quatro pessoas e a penalidade à empresa em 372 mil dólares. Finalmente tem-se a explosão da fábrica de plásticos ICL (stockline plastic) em 2004, que causou a morte de nove trabalhadores e quarenta feridos, bem como a multa de 400 mil dólares (OKOH; HAUGEN, 2007).

Em consequência dos casos indicados anteriormente o diagnóstico de falhas se tornou como uma das atividades mais importantes dentro da engenharia de processos, sendo assim o componente principal da gestão de eventos anormais (AEM), isto é, a detecção oportuna, o diagnóstico e a correção das condições anormais das falhas em um processo ou sistema (VENKATASUBRAMANIAN *et al.*, 2003). A fim de usar corretamente a terminologia no campo do diagnóstico de falhas, ISERMANN (1997) e GERTLER (1991) sugeriram as seguintes definições:

- **Detecção:** tempo de ocorrência da falha.
- **Isolamento:** tipo e localização da falha.
- **Identificação:** tamanho e comportamento no tempo da falha.
- **Diagnóstico:** tipo, tamanho, localização e tempo de detecção da falha.

Nota-se que o diagnóstico faz uma análise inteira da ocorrência da falha, em outras palavras, consegue enxergar a causa da anormalidade do sistema, o qual é peça fundamental na hora de atuar corretiva ou preventivamente. Portanto, aumenta-se a segurança da planta, permitindo a geração de informações para atividades de gerenciamento de ativos, detecção de anormalidades oportunas para a diminuição dos riscos operacionais, redução de custos por atividades não programadas e otimização das atividades de manutenção.

Como consequência, o tema tem sido amplamente estudado e existe uma grande quantidade de informação bibliográfica sobre o diagnóstico de falhas e a gestão de eventos anormais, através da implementação de métodos analíticos, estatísticos e causais. Existindo uma classificação destes métodos: métodos baseados em modelos quantitativos, métodos baseados em modelos qualitativos e métodos baseados em dados históricos do processo. Em primeiro lugar, os modelos quantitativos têm como enfoque modelos gerais entrada/saída e espaço/estado, usam técnicas como redundância analítica para gerar resíduos e isolar falhas do processo; mas são impopulares pela alta complexidade computacional e dimensionalidade e, a não linearidade dos processos (VENKATASUBRAMANIAN *et al.*, 2003).

Seguidamente os modelos qualitativos são desenvolvidos na base de alguma compreensão fundamental da física e química do processo, representando uma abstração do sistema físico real, em função da qualidade das restrições conseguem capturar a informação do sistema, porém geram soluções espúrias. É o caso dos modelos causais e as hierarquias de abstração (VENKATASUBRAMANIAN *et al.*, 2003; FRANK, 1996).

Finalmente os métodos baseados nos dados históricos do processo, onde a única exigência para a geração dos modelos são os dados históricos associados ao sistema em estudo. A transformação desses dados em informação útil para o sistema de diagnóstico é conhecida como extração de características, que pode ser qualitativa (sistemas inteligentes e tendências de análise qualitativa) ou quantitativa (classificadores estatísticos, PCA, PLS). Uma das limitações destes métodos é a extrapolação dos dados fora da sua área de formação (CHIANG, RUSSELL, BRAATZ, 2000; GERTLER, 1991; VENKATASUBRAMANIAN *et al.*, 2003).

O diagnóstico de falhas é realizado geralmente baseado em algum modelo empírico (entrada/saída) ou semi-empírico do processo. As situações onde o processo funciona em malha fechada, em uma região de verossimilhança, perto de um ponto de funcionamento, os modelos lineares são quase sempre adequados (VENKATASUBRAMANIAN *et al.*, 2003).

No entanto, a pesar de que o tema tem sido amplamente estudado e existe uma grande quantidade de métodos para a detecção e o diagnóstico de falhas, cada um deles apresenta tanto limitações quanto vantagens, que se podem aprimorar, com a finalidade de apresentar

uma metodologia interessante, robusta e que não foi estudada até o momento, sendo o caso da metodologia proposta: detecção e diagnóstico de falhas baseado em modelos empíricos no subespaço das variáveis do processo (EMPVSub), apresentada no capítulo 4.

Desta maneira, a EMPVSub consiste na criação de modelos empíricos, construídos a partir de base de dados histórica de processos, aplicando técnicas estatísticas multivariadas e métodos de seleção de variáveis para a diminuição de incertezas e demanda computacional, bem como usando técnicas de regularização para reduzir o espaço dimensional dos dados para aperfeiçoar o diagnóstico de falhas, e aprimorar a previsão de novos resultados. Além disso, a metodologia apresenta um módulo que permite a expansão não linear das bases dos dados (bases exponenciais e polinomiais), com a finalidade de fazer o método mais robusto que permita enxergar comportamentos não lineares de processos reais. Conjuntamente com a metodologia desenvolvida foram revisadas duas metodologias padrões: a Análise de Componentes Principais (PCA) e os Mínimos Quadrados Parciais (PLS), com a finalidade de compará-los com a metodologia proposta. Estas últimas estão baseadas em dados históricas de processos, mas apresentam como desvantagem principal que trabalham em todo momento com a totalidade das variáveis presentes na base de dados do sistema.

As metodologias padrões com base em PCA e PLS foram escolhidas, porque apesar de serem as mais primitivas: (i) são consideradas clássicas para a detecção e diagnóstico de falhas, (ii) são facilmente implementáveis através de ferramentas de cálculo sendo o caso do *Python*, (iii) são usadas em estudo de casos clássicos para monitoramento de processos por exemplo o *Tennessee Eastman Process* permitindo comparar seus resultados com os obtidos a partir de novos métodos e, (iv) são desenvolvidas em sistemas estáticos quanto a metodologia proposta.

Tanto a metodologia proposta quanto as metodologias padrões foram implementadas em dois estudo de casos: o primeiro deles, o sistema de um tanque de aquecimento com mistura, desenvolvido no Departamento de Engenharia Química da UFRGS, e o segundo, o *Tennessee Eastman Process (TEP) benchmark problem*, que consiste de um processo não linear amplamente estudado na área do diagnóstico de falhas e estratégias de controle.

Além da metodologia proposta para o diagnóstico de falhas, é proposta uma metodologia de validação dos resultados da metodologia do diagnóstico, que consiste na eliminação das variáveis contidas nos melhores modelos a partir da técnica *Backward Elimination*, baseada em testes estatísticos como o *F-Test* e a distribuição de Fisher, para corroborar os resultados da variável que está relacionada à falha no sistema.

## 1.1 Objetivos do trabalho

O presente trabalho busca desenvolver uma metodologia para a detecção e o diagnóstico de falhas para sistemas industriais, baseado na criação de modelos empíricos a partir de dados históricos no subespaço das variáveis do processo. Técnicas de regressão linear com regularização serão utilizadas para a diminuição da dimensionalidade de dados, bem como técnicas estatísticas multilíneas, aplicada em sistemas dinâmicos, para a seleção das variáveis dentro do modelo que podem estar relacionadas com a causa da falha (podendo ter a desvantagem de resultados falsos positivos), para permitir a diminuição das medições das variáveis seletivas e, portanto, precisar de um número menor de sensores que tenham que ser mantidos funcionando corretamente. Adicionalmente os modelos serão estendidos visando capturar os possíveis comportamentos não-lineares e desta maneira gerar uma vantagem adicional para os modelos baseados em dados históricos.

Para evidenciar as vantagens da abordagem proposta é realizada a comparação desta com técnicas baseadas em PCA e PLS. As metodologias são testadas em dois estudos de casos: o primeiro um tanque de aquecimento com mistura e o segundo o *Tennessee Eastman Process*, os quais serão brevemente explicados dentro do corpo do trabalho.

Sendo assim, o principal objetivo dessa dissertação é:

- Desenvolver um método para o diagnóstico de falhas que precise de menos variáveis explicativas dentro de um modelo baseado em dados históricos de um processo.

Como objetivos secundários tem-se os seguintes:

- Diagnosticar falhas em processos não lineares.
- Reconhecer e validar padrões em modelos empíricos.

## 1.2 Estrutura da dissertação

O presente trabalho está estruturado em seis capítulos; sendo que no Capítulo 1 é feita uma breve introdução ao tema em estudo, fazendo abordagens dos conceitos básicos, motivação, e os objetivos do trabalho.

No Capítulo 2, é apresentada uma revisão bibliográfica dos trabalhos que foram feitos até agora no Diagnóstico de Falhas em Processos industriais, as vantagens, desvantagens e limitações de cada um deles, e os tópicos de interesse a serem desenvolvidos no presente trabalho a partir deles.

No Capítulo 3, são apresentados os fundamentos teóricos do presente trabalho, tanto das metodologias com base em PCA e PLS, quanto da metodologia proposta EMPVSub.

No Capítulo 4, são apresentadas as metodologias padrões e a metodologia proposta, estudos de casos, bem como a sistemática adotada na validação e comparação dos métodos estudados.

No Capítulo 5, são mostrados os resultados obtidos após a aplicação das metodologias padrões, propostas e de validação, nos casos de estudo apresentados no Capítulo 4.

Finalmente, as principais conclusões e sugestões para trabalhos futuros são apresentadas no Capítulo 6.

## Capítulo 2 – Revisão Bibliográfica

O diagnóstico de falhas é conhecido como uma das atividades mais importantes dentro da engenharia de processos, constituindo o componente principal da gestão de eventos anormais (VENKATASUBRAMANIAN *et al.*, 2003). Assim, existe uma gama de informação bibliográfica sobre o tema, e, sendo que as metodologias desenvolvidas no presente trabalho, estão dentro da classificação dos métodos baseados em dados históricos de processos para a detecção e diagnóstico de falhas, a seguir serão resumidas uma serie de trabalhos de interesse baseadas em essas metodologias.

É importante dizer que as técnicas clássicas que apresentam aplicações bem sucedidas para monitoramento de processos são o PCA e o PLS. Relatos do uso dessas técnicas no gerenciamento de falhas em processos foram feitos por MACGREGOR *et al.* (1991); MACGREGOR *et al.* (1994); MACGREGOR, KOURTI (1995). No entanto, sempre é possível aprimorar o uso de técnicas clássicas e usá-las para desenvolver novas metodologias gerando tanto vantagens como desvantagens, dependendo do objetivo de estudo.

MACGREGOR, NOMIKOS (1994) usaram o PCA *multiway* para ampliar o uso de métodos de projeção múltipla para extrair a informação dos dados de trajetória múltipla através da projeção dos dados em espaços de dimensão reduzida definido por variáveis latentes ou componentes principais.

MACGREGOR, KOURTI, NOMIKOS (1996) trabalharam com procedimentos estatísticos multivariados baseados em várias versões da análise de componentes principais (PCA) e mínimos quadrados parciais (PLS) para monitoramento e diagnóstico de processos industriais, concluindo que esses métodos são capazes de estudar processos com um grande número de variáveis de processo e qualidade altamente correlacionados, indicando que a única informação necessária para explorá-los é uma boa base de dados sobre as operações históricas do processo.

RUSSELL, BRAATZ, RICHARD (2000) desenvolveram um trabalho de detecção e diagnóstico de falhas a partir de métodos baseados em dados históricos do processo em processos químicos, aplicado no *benchmark Tennessee Eastman Process*, abordando técnicas como análise de componentes principais (PCA), a análise do discriminante de Fisher (FDA), mínimos quadrados parciais (PLS), e a análise da variação canônica (CVA). Indicaram que o objetivo do monitoramento do processo foi desenvolver medições que sejam sensíveis e robustas para todas as falhas dentro de um processo. No caso do PCA, é

considerada como uma técnica adequada para redução de dimensionalidade em termos de captura de variância de dados, que contabiliza correlações entre as variáveis, melhora a proficiência para a detecção de falhas e permite o diagnóstico através de técnicas de estatística multivariada. O FDA é apresentado como uma técnica de redução de dimensionalidade desenvolvida para a classificação de padrões determina a parcela do espaço de observação que é mais eficaz em discriminar entre várias classes de dados; os desenvolvimentos teóricos sugerem que deve ser mais eficaz do que o PCA para diagnóstico de falhas, mas na prática não foi assim. Para o caso do PLS, sendo um método de decomposição de dados para maximizar a covariância entre o bloco preditor (independente) e o bloco predito (dependente) para cada componente, foi determinado como um modelo inferencial adequado para a detecção, identificação e diagnóstico de falhas, baseado nos resultados obtidos. No caso do CVA, os resultados obtidos foram bons, mas isto devido a sua estreita relação com o PCA, FDA e PLS. Cada um dos métodos apresentou vantagens e desvantagens dependendo do enfoque do problema, mas, em suma, os melhores resultados obtidos para o diagnóstico de falhas para os diferentes cenários do estudo de caso, foram obtidos através dos métodos PCA e PLS.

MISRA *et al.* (2002) desenvolveram o método análise de componentes principais multiescala MSPCA o qual faz a correlação cruzada entre os sensores (abordagem PCA) e faz uma autocorrelação dentro de um sensor (abordagem *wavelet* ou de sinais individuais). Os sinais individuais do sensor são decompostos em aproximações e detalhes em diferentes escalas. As contribuições de cada escala são coletadas em matrizes separadas e o modelo PCA é então construído para extrair a correlação em cada escala. As contribuições deste método permitem trabalhar com dados de processos que normalmente não são estacionários e representa um efeito cumulativo de muitos fenômenos do processo subjacentes, cada um operando de forma diferente.

As metodologias para monitoramento de processos baseadas em PCA e PLS, têm sido estudadas nas últimas décadas, sendo a principal vantagem desses métodos a capacidade de lidar com um número maior de variáveis altamente correlacionadas e reduzir a medida de processo de alta dimensão em um espaço latente de baixa dimensão (QIN, 2003; MACGREGOR, KOURTI, 1995).

Segundo CHOI *et al.* (2005) apesar de que o PCA ter apresentado bons resultados para vários estudos, é importante se levar em conta que a maioria das vezes, o método assume que o processo monitorado é linear, no entanto para sistemas não lineares o monitoramento baseado em PCA se mostra ineficaz e problemático, o que leva ao desenvolvimento de vários métodos PCA não lineares.

ZHANG (2012) abordou um novo algoritmo o baseado numa versão não linear do PCA, chamada KPCA, o método decompõe tanto o espaço principal KPCA quanto o espaço residual em dois subespaços. Em comparação com as técnicas estatísticas tradicionais, o subespaço da falha é separado com base na influência relevante para falhas. O método encontra as direções principais relevantes para falhas e componentes principais do subespaço sistemático e do subespaço residual para o monitoramento do processo.

DING (2014) realizou um trabalho de projeto de sistemas com dados históricos para o diagnóstico de falhas e sistemas de controle tolerantes a falhas, onde trabalhou com dois *benchmarks* reconhecidos: um sistema de três tanques acoplados e o *Tennessee Eastman Process*. Foram estudadas técnicas estatísticas multivariadas como PCA, PLS, CCA

(análise de correlação canónica) com o intuito de diagnosticar falhas, indicando que as técnicas resultaram adequadas.

Por outro lado, os mínimos quadrados parciais (PLS) servem como uma ferramenta poderosa para o monitoramento do indicador de desempenho chave (KPI) na indústria de processos em grande escala nas últimas duas décadas. No entanto, a abordagem padrão e suas modificações recentes ainda encontram alguns problemas para o diagnóstico de falhas relacionadas ao KPI do processo relacionado (YIN; ZHU; KAYNAK, 2015).

Os Mínimos Quadrados Parciais (PLS) apresentam um melhor diagnóstico de falhas para problemas de classificação em pequena escala em comparação com o PCA. São usados de forma semelhante ao PCA para o monitoramento de processo e são mais eficazes na supervisão das variáveis de processo que são mais influentes nas variáveis de qualidade (CHIANG; RUSSELL; BRAATZ, 2000).

Para lidar com a não linearidade QIN, MCAVOY (1992) propuseram uma abordagem neural PLS que incorporou redes com *feedforward* na modelagem do PLS.

As abordagens estatísticas multivariadas para a detecção de falhas com base em dados operacionais históricos são úteis para processos com um grande número de variáveis medidas e quando os modelos causais não estão disponíveis (YOON; MACGREGOR; 2001).

Segundo o indicado anteriormente as técnicas PCA e PLS, apresentam bons resultados na área do diagnóstico de falhas, sendo assim interessante estudar o seu embasamento teórico, além de se considerar como metodologias padrões clássicas, que permitem detectar anormalidades de processos e junto com técnicas de isolamento e identificação de variáveis, permite determinar as variáveis mais relacionadas à falha, além disso, são de fácil implementação em ferramentas computacionais, permitindo ser comparadas com novas metodologias, sendo uma delas a metodologia desenvolvida no presente trabalho.

Os trabalhos anteriormente mencionados fornecem modelos empíricos a partir de uma base de dados históricos de processos, usando todo o conjunto de variáveis disponíveis para efetuar o diagnóstico de falhas. No entanto, trabalhar com um conjunto reduzido de variáveis fornece vantagens como diminuição de incertezas e demanda computacional, bem como um diagnóstico mais exato. Assim, técnicas como seleção de variáveis e regularização, representam uma otimização dos métodos convencionais para o diagnóstico de falhas baseado em base de dados histórica de processos, além de melhorar o desempenho de generalização dos estimadores de regressão linear não paramétricos (ORR, 2008).

A seleção de variáveis e recursos tornou-se o foco de muitas pesquisas em áreas de aplicação onde conjuntos de dados com dezenas ou centenas de milhares de variáveis, estão disponíveis. Essas áreas incluem processamento de texto de documentos da Internet, análise de matriz de expressão gênica e química combinatória. O objetivo da seleção de variáveis é triplo: (i) melhorar o desempenho de previsão dos preditores, (ii) fornecer preditores mais rápidos e mais econômicos e (iii) fornecer uma melhor compreensão do processo subjacente que gerou os dados. As contribuições desta questão especial cobrem uma ampla gama de aspectos desses problemas: proporcionando uma melhor definição da função objetivo, seleção de características multivariadas, métodos de pesquisa eficientes e métodos de avaliação de validade de características (GUYON, ELISSEFF, 2003).

Com a finalidade do ajuste de modelos de regressão que permitem a seleção de variáveis preditivas de forma automática, existem várias técnicas baseadas em índices estatísticos, dentro desse conjunto de técnicas se encontram: *forward selection*, *backward elimination*, *stepwise procedure*, *ant colony optimization*, entre outras (XU; ZHANG, 2001; RANZAN *et al.*, 2014).

MEHMOOD, *et. al* (2012) indicaram que a seleção de variáveis para a redução de dados e uma melhor interpretabilidade está obtendo cada vez maior importância, pelo qual fizeram uma revisão dos métodos disponíveis para a seleção das variáveis dentro de um dos enfoques de modelado para dados de alto rendimento, como é o caso dos mínimos quadrados parciais. Dentro da revisão está uma classificação dos métodos: métodos de filtrado (coeficientes de regressão, carga de pesos, etc), métodos de envoltura (algoritmos genéticos, *backward elimination*, etc) e métodos de integração (seleção interativa de variáveis, limiar suave, etc).

KING e JACKSON (2014) realizaram um estudo em um conjunto de dados ambientais, aplicando técnicas de seleção de variáveis aos componentes gerados pelo método PCA através do modelo Broken-stick, essas técnicas foram *backward elimination*, *forward selection*, B2 e B4, avaliando os métodos de seleção de variáveis através da análise Procrustes, que é uma medida de semelhança e diagramas bivariados para medir o sucesso dos diferentes métodos de seleção de variáveis. O estudo atingiu resultados positivos para cada uma das técnicas de seleção de variáveis aplicadas.

FERNANDEZ *et al* (2009) propuseram o método de seleção de variáveis *backward elimination* usando PLS. Além usaram o erro de predição de raiz quadrática meia (RMSEP) para quantificar o ganho obtido usando o rango de variáveis selecionado. O método foi aplicado a diferentes conjuntos de dados e os resultados obtidos revelaram que se podem aprimorar ou pelo menos manter constantes os rendimentos de predição dos modelos PLS em comparação com os modelos de espectro inteiro. Além, com a vantagem de reduzir o número de variáveis, se conduz a uma interpretação mais simples da relação entre o modelo e as propriedades da amostra.

Motivado ao fato do importante de fazer uma seleção de variáveis no ajuste de dados do sistema para determinar qual delas gera uma melhor regressão dos dados, também resulta interessante determinar o número de variáveis que estará presente no ajuste para evitar problemas de *overfitting* ou *underfitting*. Sendo assim, a regularização permite contornar esses problemas, permitindo um ajuste adequado dos dados incorporando informações adicionais à função objetivo (eliminando parâmetros de pouca importância, penalidade).

Tanto a seleção de variáveis quanto as técnicas de regularização apresentam um risco considerável na geração de modelos espúrios (BABYAK, 2004). Frequentemente os dados são sobreajustados, desenvolvendo modelos com muitas variáveis de predição, sendo pouco provável que superem as provas de validação (HARRELL *et al.*, 1985) pelo qual é necessário aplicar técnicas de regularização para obter resultados confiáveis.

FRIEDMAN, HASTIE e TIBSHIRANI no ano 2010, desenvolveram um estudo de algoritmos para a estimação de modelos com técnicas de regularização: LASSO, RIDGE e Elastic net, onde os algoritmos usavam a coordenada da descida ao longo do caminho da regularização. Cada passo de descenso de coordenadas foi rápido, com uma fórmula

explicita para cada minimização de coordenadas. A velocidade de cálculo foi bastante notable, obtendo resultados interessantes.

KOH, KIM e BOYD (2007) propuseram um estudo de regressão logística com regularização LASSO, para a seleção de características em problemas de classificação, obtendo resultados prometedores, para uma grande quantidade de dados (dezenas de milhares de características), realizando uma variação do método clássico (que utiliza um método de gradiente conjugado preconditionado para calcular a busca em linha), utiliza técnicas de início em quente para conseguir uma boa aproximação de todo o caminho de regularização, permitindo de uma forma mais eficiente resolver os problemas de overfitting.

Assim, baseado nos resultados dos métodos clássicos, bem como nas modificações de alguns dos trabalhos anteriormente resumidos, o presente trabalho, vai usar técnicas de regressão clássicas como PCA e PLS, bem como técnicas de seleção de variáveis como backward elimination e otimização estocástica (ACO); assim como métodos de regularização LASSO-LARS, na criação de uma metodologia de detecção e diagnóstico de falhas baseado em modelos empíricos, a qual será detalhada nos capítulos posteriores.

## Capítulo 3 – Fundamentos

### 3.1 Classificação dos Métodos de Diagnóstico de Falhas

VENKATASUBRAMANIAN *et al.* (2003) fizeram um estudo bastante completo acerca dos métodos existentes para o diagnóstico de falhas, segmentando-os em três categorias: (i) métodos baseados em modelos quantitativos, (ii) métodos baseados em modelos qualitativos e (iii) métodos baseados em dados históricos do processo; a Figura 3.1 apresenta essas categorias, assim como métodos específicos correspondentes a cada uma delas. Embora cada um dos métodos possua vantagens e desvantagens, os autores concluíram que uma boa metodologia para o diagnóstico de falhas é aquela que apresenta a combinação de diferentes técnicas contidas nas diferentes categorias contempladas na classificação.

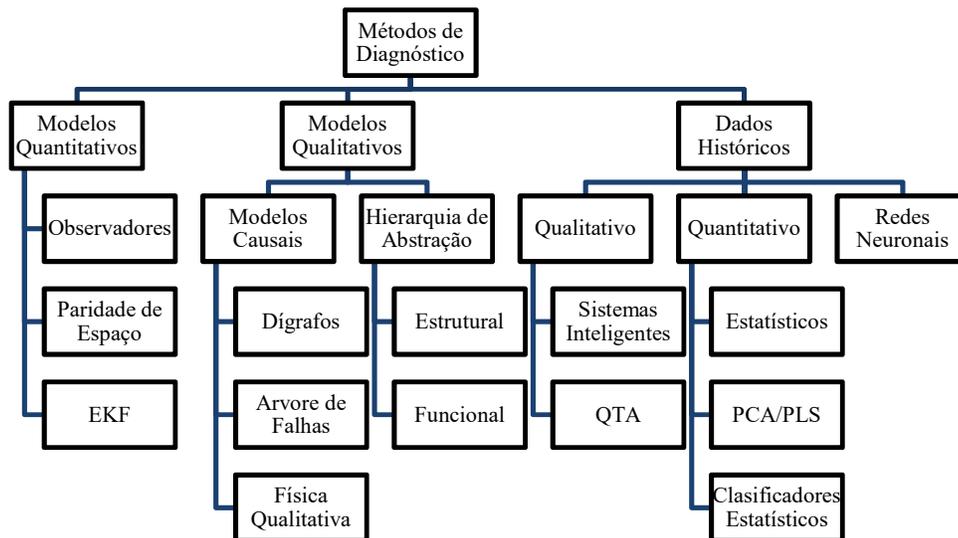


Figura 3.1: Classificação dos métodos de diagnóstico de falhas (Adaptado de VENKATASUBRAMANIAN *et al.*, 2003)

No trabalho do VENKATASUBRAMANIAN *et al.* (2003), é indicado que alguns desses métodos podem se complementar resultando em melhores sistemas de diagnóstico. A integração desses recursos complementares é uma forma de desenvolver sistemas híbridos que podem superar as limitações das estratégias de solução individuais. LAMEDA (2015) indicou desde o ponto de vista da aplicação industrial, que a maioria das aplicações de diagnóstico de falhas em processos industriais está baseada na informação histórica do processo. Isto se deve a fácil implementação, requerendo pouco esforço na modelagem de processos e conhecimentos prévios.

Baseado na classificação de VENKATASUBRAMANIAN *et al.* (2003); GERTLER, (1991), e no indicado por LAMEDA (2015) em relação a vantagem de usar métodos baseados em dados históricos de processos para o diagnóstico de falhas, na seção 3.2 serão especificados uma série de trabalhos feitos na área, bem como o embasamento teórico associado.

### 3.2 Métodos baseados em dados históricos do processo para diagnóstico de falhas

Antes de explicar as referidas técnicas é importante mostrar o modelo geral de regressão multilinear, que corresponde à regressão por Mínimos Quadrados Ordinários (OLS), a partir da qual são descritos os demais modelos de regressão.

#### 3.2.1 Mínimos Quadrados Ordinários (OLS)

A regressão por Mínimos quadrados ordinários (OLS) é a regressão linear adaptada a um modelo linear com coeficientes  $w = (w_1, \dots, w_p)$  que minimiza a soma residual dos quadrados entre as respostas observadas no conjunto de dados e as respostas previstas pela aproximação linear (PEDREGOSA *et al.*, 2011). Foi desenvolvido por Carl Friedrich Gauss e Adrien-Marie Legendre em 1795. Matematicamente é resolvido através da equação (3.1), onde  $X$  corresponde as variáveis explicativas (independentes), representa o fator explicativo na equação e  $y$  a variável explicada (dependente); é o valor que se quer atingir.

$$\min_w \|XW - y\|_2^2 \quad (3.1)$$

É importante mencionar que a regressão linear múltipla ordinária, é usual quando: (i) as variáveis explicativas são poucas, (ii) quando não existem problemas de multicolinearidade e (iii) quando existe uma clara relação entre as variáveis. Se alguma dessas condições falha, então a regressão linear múltipla ordinária não é eficiente. Por outro lado, a OLS é geralmente usada com uma grande quantidade de variáveis explicativas, mas quando o número de variáveis é consideravelmente grande é possível gerar um modelo que ajuste bem os dados, mas que erra na predição de dados novos. Nestes casos, onde existem muitas variáveis explicativas, é possível a existência de poucas variáveis não observáveis que coletam a maior variabilidade das variáveis de resposta. Nesses casos, uma modificação do OLS, denominada Mínimos Quadrados Parciais permite extrair essas variáveis latentes, coletando a maior variação das variáveis explicativas de maneira que sirvam para modelar as variáveis de resposta da melhor maneira possível (GAVIRIA, 2015).

#### 3.2.2 Mínimos Quadrados Parciais (PLS)

A regressão por Mínimos Quadrados Parciais (PLS) ou *Projection to Latent Structures/Partial Least Squares*, como também é conhecida, foi introduzida por Herman Wold em 1950 (WESTLAND, 2015). É uma técnica de redução de dimensionalidade para

maximizar a covariância entre a matriz preditora  $X$  (independente) e a previsão da matriz  $Y$  (dependente) para cada componente do espaço reduzido.

O PLS calcula os vetores de carga e de projeções que estão correlacionados com o bloco previsto, descrevendo uma grande quantidade de variação no bloco preditor. Quando o bloco preditor tem apenas uma variável, o método PLS é conhecido como PLS1 e quando o bloco preditor tem múltiplas variáveis é conhecido como PLS2. O PLS requer os passos de calibração e predição (RUSSELL; BRAATZ; RICHARD, 2000).

É importante saber que os dados (matrizes  $X$  e  $Y$ ) antes de implementar a metodologia, devem ser normalizados: média zero e desvio padrão unitário. A matriz  $X$ , segundo (3.2) é decomposta em uma matriz de projeção  $T \in \mathbb{R}^{n \times a}$  e uma matriz de carga  $P \in \mathbb{R}^{m \times a}$ , onde  $a$  é a componente do PLS (redução da ordem) mais uma matriz residual  $E \in \mathbb{R}^{n \times m}$ .

$$X = TP^T + E \quad (3.2)$$

O produto da matriz  $TP^T$  pode se expressar como a soma do produto dos vetores de projeção  $t_j$  (a  $j$ -ésima coluna de  $T$ ) e os vetores de carga  $p_j$  (a  $j$ -ésima coluna de  $P$ ), matematicamente segundo (3.3).

$$X = \sum_{j=1}^a t_j p_j^T + E \quad (3.3)$$

De maneira similar  $Y$ , segundo (3.4) é decomposta em uma matriz de projeção  $U \in \mathbb{R}^{n \times a}$  e matriz de carga  $Q \in \mathbb{R}^{p \times a}$ , mais a matriz residual  $F \in \mathbb{R}^{n \times p}$ .

$$Y = UQ^T + F \quad (3.4)$$

O produto  $UQ^T$  pode-se expressar como a soma do produto dos vetores de projeção  $u_j$  (a  $j$ -ésima coluna de  $U$ ) e os vetores de carga  $q_j$  (a  $j$ -ésima coluna de  $Q$ ). Matematicamente segundo (3.5).

$$Y = \sum_{j=1}^a u_j q_j^T + F \quad (3.5)$$

Se  $a$  for igual ao  $\min_{\{m,n\}}$ , então  $E$  e  $F$  são zero, e o PLS é reduzido ao método mínimos quadrados ordinários (OLS). O objetivo do PLS é determinar os vetores de carga e de projeção que estejam correlacionados com  $Y$ , enquanto descrevem uma grande variação em  $X$ . O PLS fornece uma estimativa do vetor de projeção  $Y$ ,  $\hat{u}_j$ , ao vetor de projeção de  $X$ ,  $t_j$ , segundo (3.6) e (3.7), respectivamente.

$$\hat{u}_j = b_j t_j \quad (3.6)$$

$$\hat{U} = TB \quad (3.7)$$

onde o  $b_j$  é o coeficiente de regressão.  $U$  corresponde ao indicado em (3.4), mas em forma de matriz. Onde  $B \in \mathbb{R}^{a \times a}$  é a diagonal da matriz de regressão, com  $B_{jj}=b_j$  e  $\hat{U}$  tem  $\hat{u}_j$  como suas colunas. Substituindo  $\hat{U}$  por  $U$ , e levando em conta que isso irá modificar a matriz residual, tem como resultado o mostrado em (3.8).

$$Y = TBQ^T + F \quad (3.8)$$

Sendo  $F$  é a matriz do erro de predição. A matriz  $B$  é selecionada a partir da norma 2 de  $F$  (o maior valor singular de  $F$ ),  $F_{N2}$  é minimizado. Os vetores de projeção  $t_j$  e  $\hat{u}_j$  são calculados para cada fator do PLS ( $j=1,2, \dots, a$ ), de modo que a covariância entre  $X$  e  $Y$  seja maximizada em cada fator.

$$y_i = T_i B_i q_i^T + f_i \quad (3.9)$$

onde  $y_i \in \mathbb{R}^n$  é a  $i$ -ésima coluna de  $Y$ ,  $T_i \in \mathbb{R}^{n \times a}$  é a matriz de projeção,  $B_i \in \mathbb{R}^{a \times a}$  é a matriz de regressão,  $q_i \in \mathbb{R}^{a \times a}$  é o vetor de carga, e  $f_i \in \mathbb{R}^n$  é o vetor do erro de predição. Como  $Y$  tem  $p$  colunas, o range de  $i$  vai de 1 até  $p$ . Isto pode se observar segundo (3.9).

### 3.2.3 Análise de Componentes Principais (PCA)

RUSSELL, BRAATZ, RICHARD (2000) indicaram o PCA como uma técnica de redução de dimensionalidade amplamente usada no monitoramento de processos. Além de ser uma boa técnica em termos de capturar variabilidade nos dados e correlacionar variáveis. A partir de uma matriz de dados  $X$   $n \times m$  (sendo  $n$  o número de observações e  $m$  o número de variáveis do processo) se determina um conjunto de vetores ortogonais chamados vetores de carga, representados pelas colunas da matriz  $V \in \mathbb{R}^{m \times m}$ , ordenados pelo valor da variância nas direções do vetor. Os vetores de carga podem ser calculados via a decomposição dos valores singulares (SVD) uma vez calculada a matriz de covariância  $S$  a partir da matriz de dados  $X$ , segundo (3.10).

$$S = \frac{1}{n-1} X^T X = V \Lambda V^T \quad (3.10)$$

onde a diagonal da matriz  $\Lambda \in \mathbb{R}^{m \times m}$  contém os autovalores não negativos na ordem de magnitude decrescente ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ ) e o autovalor  $i$ -ésimo é igual ao quadrado do  $i$ -ésimo valor singular ( $\lambda_i = \sigma_i^2$ ).

Com a finalidade de capturar adequadamente as variações dos dados e minimizar o efeito do ruído aleatório que interfere na representação do PCA, os vetores de carga correspondem aos maiores valores singulares que são mantidos (onde a variância acumulativa seja  $\geq \phi$ , normalmente  $\phi$  captura valores de 90-95%). O número total dos maiores valores singulares constitui o número de componentes principais do sistema  $a$ .

Selecionando as colunas da matriz de carga  $V$  correspondentes ao número de componentes principais, é obtida a matriz de carga  $P \in \mathbb{R}^{a \times m}$  para corresponder aos vetores de carga associados aos primeiros valores singulares. As projeções das observações em  $X$  no espaço dimensional inferior estão contidas na matriz de projeção  $T_p$ , indicada em (3.11)

$$T_p = X P^T \quad (3.11)$$

sendo  $T_p$  a projeção de  $X$  no espaço de observação  $m$ -dimensional. A partir disto se pode obter um vetor no espaço de projeção dimensional inferior  $t_{pi} = p_i X^T$ , onde  $p_i$  é o vetor de carga e  $t_{pi}$  o  $i$ -ésima componente principal de  $X$ .

Com a finalidade de fazer a detecção de falhas a partir de dados históricos do processo empregando o PCA, é usada a estatística  $T^2$ , também denominada de  $T$  de Hotelling, definida segundo (3.12).

$$T^2 = \sum_{i=1}^a \frac{t_{pi}^2}{\lambda_i} \quad (3.12)$$

O valor obtido a partir de (3.12) vai consolidar a informação mais relevante dos dados proporcionados, ou seja, aos correspondentes componentes principais que apresentem uma

variância cumulativa superior a 90 ou 95% (valor comumente adotado na literatura). Com o intuito de determinar aqueles pontos que estiverem fora dos valores em operação normal (*outliers*) e puderem ser considerados como uma potencial falha no sistema é preciso definir um valor limite  $T_\alpha^2$ , o qual permite conhecer o instante do tempo em que acontece a falha, o valor é calculado segundo a seguinte equação:

$$T_\alpha^2 = \frac{a(n-1)(n+1)}{n(n-a)} F_\alpha(a, n-a) \quad (3.13)$$

Onde  $F_\alpha$  é uma distribuição de probabilidade contínua, denominada distribuição F de Fisher, conforme já definido n como o número de observações da matriz nxm e a o número de componentes principais.

### 3.2.4 Análise de Contribuições

Segundo RUSSELL, BRAATZ, RICHARD (2000), a análise de contribuições é uma abordagem do PCA para o diagnóstico de falhas que leva em conta as correlações espaciais, melhorando assim as técnicas estatísticas univariadas. A abordagem baseia-se na quantificação da contribuição de cada variável de processo para as pontuações individuais da representação PCA e para cada variável de processo somando as contribuições apenas das pontuações responsáveis pelo *status* fora de controle. O procedimento é aplicado em resposta a uma violação do  $T^2$  e é resumido da seguinte forma:

- Inicialmente são verificadas as pontuações normalizadas  $(t_{pi}/\sigma_i)^2$  para a observação x e determinados os resultados do status anormal.
- É calculada a contribuição de cada variável  $x_j$  para cada uma das pontuações anormais (3.14).

$$cont_{i,j} = \frac{t_{pi}}{\sigma_i^2} p_{i,j} (x_j - \mu_j) \quad (3.14)$$

- Onde  $p_{i,j}$  é o (i, j) ésimio elemento da matriz de carga P.
- Quando o  $cont_{i,j}$  é negativo, é preciso ajustá-lo a zero.
- Posteriormente é calculada a contribuição total da j-ésima variável do processo,  $x_j$ . Segundo (3.15).

$$CONT_j = \sum_{i=1}^r cont_{i,j} \quad (3.15)$$

- Finalmente é feita a representação da  $CONT_j$  para todas as m variáveis do processo,  $x_j$  em uma mesma figura.
- As variáveis responsáveis pela falha podem ser priorizadas pelos valores de contribuição total  $CONT_j$ , e os responsáveis do processo podem se concentrar imediatamente nessas variáveis com altos valores de  $CONT_j$  e usar o conhecimento de seus processos para determinar a causa de qualquer anormalidade do processo.

### 3.2.5 Regressão Linear com Regularização

Fazendo uma correta implementação de técnicas de ajustes de dados, tanto para ajustes lineares ou não lineares (dependendo do sistema), é possível conseguir bons resultados,

isto é, quando um modelo estatístico se ajusta adequadamente ao conjunto de dados de origem; mas, às vezes, o modelo se mostra ineficaz para prever novos resultados (CÁRDENAS-MONTES, 2015), isto é conhecido como *overfitting* ou sobreajuste (Figura 3.2).

Além disso, é interessante lembrar o Princípio de Parcimônia, o qual exige o uso de modelos e procedimentos que contenham tudo o que é necessário para a modelagem, mas nada a mais. O *overfitting* é o uso de modelos que violam a parcimônia, que incluem mais termos do que são necessários ou usam abordagens mais complicadas das que se precisam, sendo isto um problema para a integridade e a confiabilidade de um modelo estatístico (HAWKINGS, 2004).

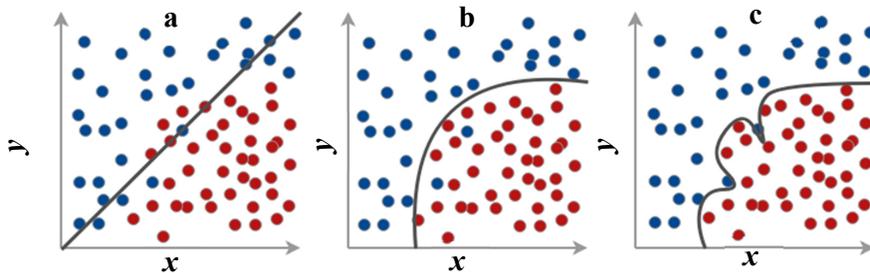


Figura 3.2: Ilustrando *underfitting* (a), ajuste adequado (b) e *overfitting* (c) para o caso de um problema de classificação (BISHOP, 2006)

Uma ferramenta para contornar o problema do *overfitting* é a regularização, a qual é uma técnica aplicada a funções objetivo envolvendo ajuste de dados, que introduz informações adicionais para resolver o problema. Tal adição resulta na eliminação de parâmetros de pouca importância, e, portanto, em um modelo mais convexo, do qual se espera que seja mais representativo da realidade. Em regressão linear, a regularização busca o encolhimento dos coeficientes em direção ao zero, com a finalidade de melhorar a interpretabilidade e diminuir o erro de predição. Segundo BISHOP (2006) essa informação adicional, ou penalidade, como também é conhecida, serve para limitar aos coeficientes de alcançar grandes valores. O termo de penalização mais simples leva a forma de uma soma de quadrados de todos os coeficientes, levando a uma função de erro modificada, como é mostrada na equação (3.16).

$$E(w) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, w) - t_n\}^2 + \frac{\lambda}{2} \|w\|^2 \quad (3.16)$$

Onde  $\|w\|^2 \equiv ww^T = w_0^2 + w_1^2 + \dots + w_M^2$ , e o coeficiente  $\lambda$  modula a importância relativa do termo de regularização em comparação com o termo de erro de soma de quadrados. Segundo BISHOP (2006), esta escolha particular de regulador é conhecida na literatura de aprendizagem de máquinas como *decadência do peso* porque, em algoritmos de aprendizagem sequencial, incentiva os valores de peso a diminuir em direção a zero, a menos que sejam suportados pelos dados.

Uma forma mais geral de regularizador é dada pela seguinte equação:

$$\frac{1}{2} \sum_{n=1}^N \{t_n - w^T \Phi(x_n)\}^2 + \frac{\lambda}{2} \sum_{j=1}^M |w_j|^q \quad (3.17)$$

É possível observar na Figura 3.3 os contornos da função de regularização para diferentes valores de  $q$ . O caso de  $q=1$  é conhecido como o LASSO na literatura estatística, e  $q=2$  corresponde ao regularizador quadrático. Tem a propriedade de que se  $\lambda$  for suficientemente grande, alguns dos coeficientes  $w_j$  são conduzidos para zero, levando a um modelo esparsos no qual as funções base correspondentes não desempenham nenhum papel.

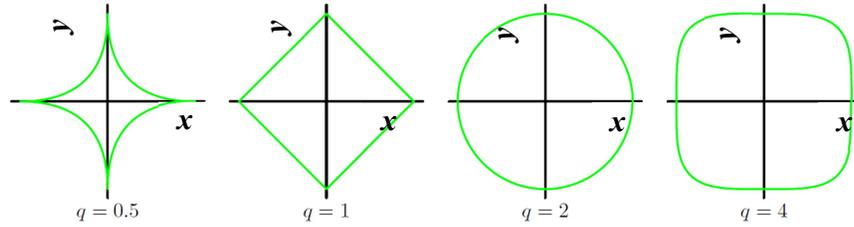


Figura 3.3: Contornos do termo de regularização para valores do parâmetro  $q$  (BISHOP, 2006)

Para visualizar o efeito na regularização, primeiro deve-se observar que a minimização anterior é equivalente à minimização do erro da soma de quadrados não regularizado sujeito a restrição mostrada na equação (3.18).

$$\sum_{j=1}^M |w_j|^q \leq \eta \quad (3.18)$$

Para um valor apropriado do parâmetro  $\eta$ , onde as duas abordagens podem ser relacionadas usando os multiplicadores de Lagrange.

A regularização permite que modelos complexos sejam ajustados utilizando conjuntos de dados de tamanho limitado sem ajuste excessivo, essencialmente limitando a complexidade do modelo.

Entre os métodos de regularização além do regulador quadrático e do LASSO têm-se o Ridge, LARS, LASSO-LARS e Elastic Net.

### 3.2.6 Regressão Ridge

A regressão Ridge aborda alguns dos problemas dos mínimos quadrados comuns, impondo uma penalidade ao tamanho dos coeficientes. Os coeficientes minimizam uma soma residual penalizada de quadrados, segundo (3.19).

$$\min_w \|XW - y\|_2^2 + \lambda \|W\|_2^2 \quad (3.19)$$

Aqui o  $\lambda$  é um parâmetro de complexidade que controla a quantidade de diminuição, quanto maior o valor de  $\lambda$ , maior a penalização da utilização dos coeficientes o que torna mais robustos à colinearidade (PEDREGOSA *et al.*, 2011).

### 3.2.7 Regressão Lasso

O Lasso (*least absolute shrinkage and selection operator*) é uma penalização linear que tende a reduzir o número de coeficientes, uma vez que os que menos contribuem para qualidade do ajuste acabam assumindo o valor nulo. É útil em alguns contextos devido a sua tendência de preferir soluções com menos valores de parâmetros, efetivamente reduzindo o número de variáveis sobre as quais a solução dada é dependente.

Matematicamente, consiste em um modelo linear regularizador onde a função objetivo a minimizar é mostrada em (3.20).

$$\min_w \frac{1}{2n_{samples}} \|XW - y\|_2^2 + \lambda \|W\|_1 \quad (3.20)$$

A estimativa do Lasso resolve assim a minimização da penalidade de mínimos quadrados com o  $\lambda\|w\|_1$  adicionado, onde  $\lambda$  é uma constante e  $\|w\|_1$  é a norma L1 do vetor de parâmetros. É uma técnica boa para seleção de variáveis. Entretanto, tem como desvantagem que não é boa para seleção de grupos para preditores fortemente correlacionados (PEDREGOSA *et al.*, 2011).

### 3.2.8 Regressão Lars

A regressão de menor ângulo (Lars) é um algoritmo de regressão para dados de alta dimensão. Lars é semelhante à regressão passo a passo. Em cada etapa, ele encontra o preditor mais correlacionado à resposta. Quando existem vários preditores com correlação igual, em vez de continuar ao longo do mesmo preditor, ele persegue em uma direção equiangular entre os preditores (PEDREGOSA *et al.*, 2011).

Dentro das vantagens do método Lars se tem que é numericamente eficiente em contextos onde  $p \gg n$  (isto é quando o numero de dimensões é significativamente maior do que o numero de pontos). É computacionalmente tão rápido quanto a seleção direta e tem a mesma ordem de complexidade que os mínimos quadrados comuns. Produz um caminho de solução linear por partes completas, que é útil em validação cruzada, ou tentativas similares para ajustar o modelo. É facilmente modificado para produzir soluções para outros estimadores, como o Lasso (PEDREGOSA *et al.*, 2011).

Uma desvantagem do método corresponde a que o Lars se baseia em uma reposição iterativa dos resíduos, pelo que parece ser especialmente sensível aos efeitos de ruído (PEDREGOSA *et al.*, 2011).

### 3.2.9 Regularização Lasso Lars

Lasso-Lars é a implementação do modelo Lasso usando o algoritmo Lars (PEDREGOSA *et al.*, 2011), e ao contrário da implementação baseada na coordenada da descida, é produzida a solução exata, que é linear por partes como uma função da norma de seus coeficientes, segundo (3.21).

$$\min_w \frac{1}{2n_{samples}} \|XW - y\|_2^2 + \lambda \|W\|_1 \quad (3.21)$$

Este método ajusta o modelo linear eliminando preditores no processo de estimação, e atribui-se os coeficientes de menor módulo, como zero.

## 3.3 Seleção de Variáveis

É interessante apresentar também um conjunto de técnicas para ajuste de modelos de regressão que permitem a seleção de variáveis preditivas de forma automática, fazendo uso de índices estatísticos. Dentro do conjunto de técnicas se têm *forward selection*, *backward elimination*, *stepwise procedure*, *ant colony optimization*, entre outras (XU; ZHANG, 2001; RANZAN *et al.*, 2014).

### 3.3.1 Ant Colony Optimization (ACO)

Dentro do GIMSCOP foi desenvolvida uma ferramenta que utiliza modelos de regressão multilinear associados ao método de otimização estocástica ACO (*Ant Colony Optimization*) que é uma metodologia de otimização global aplicável à seleção de variáveis. Na ferramenta, as variáveis são filtradas, selecionando grupos de componentes com maior correlação. Após a seleção deste grupo de elementos, estes são usados como elementos de entrada de modelos multilineares, e então submetidos a testes para predizer

variáveis de interesse. O algoritmo está baseado no comportamento coletivo de formigas que saem em busca de alimento (RANZAN *et al.*, 2014).

Segundo RANZAN *et al.* (2014) inicialmente as formigas ocupam todos os possíveis caminhos entre o ninho e fonte de alimento, e com o decorrer do tempo, todas as formigas ocupam somente o menor caminho entre os dois pontos. Este comportamento é possível com o uso de feromônios; substâncias químicas que, disseminadas entre seres da mesma espécie, promovem reações específicas em seus indivíduos. Estas substâncias evaporam com o tempo, e funcionam como chamariz para as demais formigas. As formigas que percorrem o menor caminho retornam ao formigueiro mais rapidamente, e assim a trilha percorrida por estes indivíduos apresenta maior concentração de feromônio.

Além disso, como o objetivo de melhorar as predições, a ferramenta foi modificada aplicando o método de regularização Lasso-Lars na construção dos modelos, para selecionar variáveis de interesse.

O esquema do algoritmo implementado para a resolução do problema de seleção do grupo de variáveis que utiliza o ACO pode ser observado na Figura 3.4. Este está dividido em quatro fases, a primeira delas corresponde à inicialização das variáveis necessárias para a solução do problema de otimização, a segunda a inicialização da solução, a terceira a rotina de busca e a última a apresentação dos resultados.

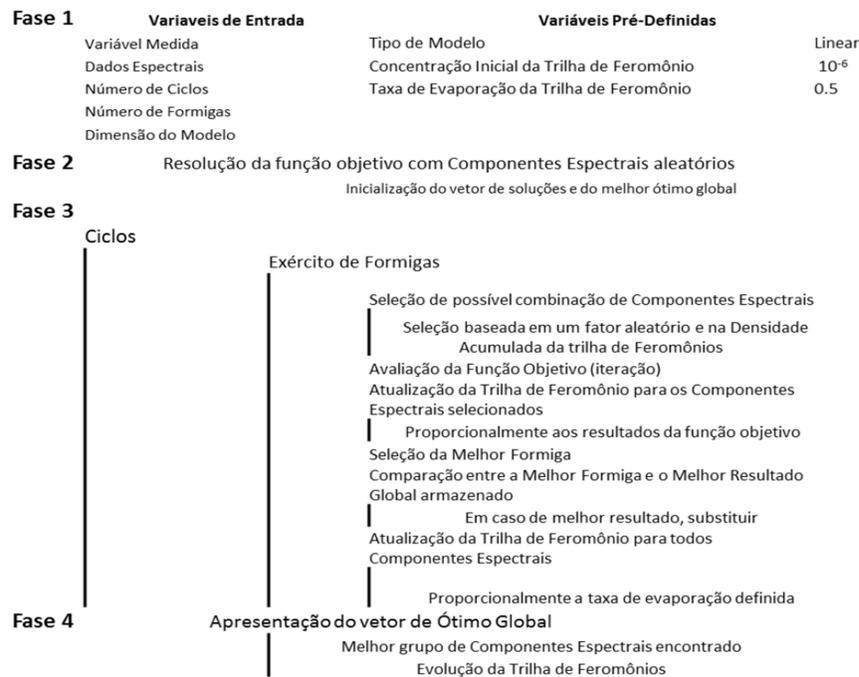


Figura 3.4: Esquema do Algoritmo da Ferramenta ACO (RANZAN *et al.*, 2014)

Na primeira etapa, é dada uma matriz de dados, é escolhido o número de ciclos que o algoritmo realizará para a busca do ótimo e é indicada a dimensão do modelo linear bem como o tipo do modelo utilizado para a avaliação do grupo de elementos. Na segunda etapa é inicializado o vetor de soluções, através da resolução da função objetivo com a seleção aleatória de um conjunto de elementos. Na terceira etapa é determinada a melhor combinação de elementos para prever as variáveis de estado observadas (sendo o núcleo do algoritmo). Na última etapa são apresentados os resultados da melhor solução obtida

durante o processo de otimização, bem como os índices de ajuste de interesse para cada solução obtida (RANZAN *et al.*, 2014).

### 3.3.2 *Forward Selection*

O método *Forward Selection* adiciona variáveis ao modelo uma de cada vez. A primeira variável incluída no modelo é aquela que possui a maior correlação com a variável independente “y”. A variável que entra no modelo como a segunda variável é aquela que possui a correlação mais alta com “y”, depois que y foi ajustado para o efeito da primeira variável. Esse processo termina quando a última variável que entra no modelo tem um coeficiente de regressão insignificante ou todas as variáveis estão incluídas no modelo (XU; ZHANG, 2001).

### 3.3.3 *Backward Elimination*

Em contraste com o *Forward Selection*, o *Backward Elimination* começa com o modelo completo e sucessivamente elimina uma variável de cada vez. A primeira variável excluída é aquela com a menor contribuição para a redução da soma de quadrado de erro preditivo (PRESS) ou dependendo do *F-Test*. Supondo que existam mais variáveis que são insignificantes, o processo opera, eliminando a próxima variável mais insignificante. O processo é encerrado quando todas as variáveis são significativas ou todas, exceto uma variável, foram excluídas (XU; ZHANG, 2001).

### 3.3.4 *Stepwise procedure*

No *stepwise procedure*, uma variável inserida no modelo nos estágios iniciais de seleção pode ser excluída nos estágios posteriores. Os cálculos feitos para inclusão e eliminação de variáveis são os mesmos que os procedimentos *forward selection* e *backward elimination*. Ou seja, o método *stepwise procedure* é essencialmente um procedimento de seleção direta, mas em cada etapa é considerada a possibilidade de excluir uma variável, como no *backward elimination*. O número de variáveis retidas no modelo baseia-se nos níveis de significância assumidos para inclusão e exclusão de variáveis do modelo (XU; ZHANG, 2001).

## 3.4 Métricas de Ajuste de Dados

Com a finalidade de capturar o erro associado a cada um dos modelos gerados, e quantificar a qualidade preditiva de cada um deles, foi usado como índice de ajuste, o coeficiente de determinação ( $R^2$  – *Coefficient of Prediction*).

Para o caso de quantificar a qualidade preditiva de cada modelo em função das bases de dados em operação normal e as bases de dados em operação anormal, foi usada a métrica *Predicted Residual Error Sum of Squares* - PRESS.

Para conferir a melhora do modelo uma vez que seja eliminada uma variável segundo o método *Backward Elimination*, é usado o *F-Test*, o qual é um teste estatístico que tem uma distribuição F sobre a hipótese nula. É usado quando são comparados modelos estáticos a partir de um conjunto de dados, a fim de identificar o modelo que melhor se adequa a população a partir da qual os dados foram mostrados (LOMAX; HAHS-VAUGHN, 2007).

Considerando dois modelos, 1 e 2, onde o modelo 1 é o modelo restrito e o modelo 2 é o modelo irrestrito, o modelo 1 com  $p_1$  parâmetros e o modelo 2 com  $p_2$  parâmetros, onde  $p_2 > p_1$ , e para qualquer escolha de parâmetros do modelo 2. Tipicamente, o modelo 2 dará um PRESS (*Predicted Residual Error Sum of Squares*) menor que o modelo 1. Mas muitas vezes, quer determinar se o modelo 1 dá um ajuste significativamente melhor aos dados.

Assim, se houver  $n$  pontos de dados para estimar os parâmetros de ambos modelos, então pode-se calcular a estatística  $F$ , segundo a equação (3.23), (ZHANG; WANG, 1997).

$$F = \frac{\frac{PRESS_1 - PRESS_2}{p_2 - p_1}}{\frac{PRESS_2}{n - p_2 - 1}} \quad (3.23)$$

Sobre a hipótese nula de que o modelo 2 não oferece um ajuste significativamente melhor do que o modelo 1, o  $F$ -Test terá uma distribuição  $F$ , com  $(p_2 - p_1, n - p_2)$  grau de liberdade. A hipótese nula será rejeitada se o  $F$  calculado a partir dos dados for maior que o valor crítico da distribuição  $F$  (ZHANG; WANG, 1997).

O coeficiente de determinação  $R^2$  é um índice de ajuste utilizado para avaliar a capacidade de ajuste do modelo aos dados (PEDREGOSA *et al.*, 2011), sendo definido pela seguinte expressão:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.24)$$

Onde  $y_i$  corresponde ao valor verdadeiro da variável,  $\hat{y}_i$  corresponde ao valor predito e  $\bar{y}$  é o valor médio. O  $R^2$  é a porcentagem da variação da variável de resposta explicada pela relação com uma ou mais variáveis preditoras. Quanto maior é o  $R^2$ , melhor o modelo ajusta os dados.

Segundo PEDREGOSA *et al.*, (2011) o *Predicted Residual Error Sum of Squares* (PRESS) é utilizado para avaliar a capacidade de predição do modelo, segundo (3.25). Quanto menor o valor PRESS, melhor a capacidade de predição do modelo. Junto com o  $R^2$ , podem evitar um ajuste excessivo do modelo (*overfitting*), pois são calculadas usando observações não incluídas na estimativa do modelo (por exemplo, a base de dados com presença do distúrbio).

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.25)$$

Onde  $y_i$  corresponde ao valor medido da variável e  $\hat{y}_i$  corresponde ao valor predito.

# Capítulo 4 – Metodologia

## 4.1 Metodologia Padrão

### 4.1.1 Análise de Componentes Principais (PCA)

Segundo RUSSELL, BRAATZ, RICHARD (2000), o procedimento para fazer um diagnóstico de falhas usando o método PCA, consiste nas seguintes etapas:

**Etapa 1:** Inicialmente é criada uma base de dados do processo (Base), em função do número total de variáveis de entrada e saída do sistema (variáveis controladas, manipuladas e medidas). Os dados são dispostos em uma matriz  $X_{n \times m}$ , sendo  $n$  o número de observações e  $m$  o número de variáveis de processo.

**Etapa 2:** Logo, é feito um pré-tratamento da matriz  $X$ , quer dizer, uma normalização padrão onde a média seja zero e o desvio-padrão seja unitário, isto é feito usando o comando `StandardScaler` do pacote `sklearn.preprocessing` da ferramenta Python (PEDREGOSA *et al.*, 2011). Este ponto é muito importante, pois a maioria dos dados de qualquer sistema tem diferentes ordens de grandeza, pelo qual, devem-se estandardizar para levar para uma mesma base; o qual é feito ao normalizar a matriz.

**Etapa 3:** A matriz  $X$  pré-tratada é transformada em uma matriz de covariância  $S$ .

**Etapa 4:** Posteriormente são calculados os autovalores e auto vetores do sistema (também chamados de vetores de carga, dispostos em uma matriz  $V$ ) a partir da matriz  $S$  (matriz de covariância). Em função dos autovalores calculados, é determinada a variância de cada um dos componentes do sistema em função do total; a partir disso, é calculada a variância acumulativa dos componentes do sistema. Quando a variância acumulativa seja maior ou igual que 95%, o número de componentes associados à dita variância cumulativa constituirá o número de componentes principais do sistema.

**Etapa 5:** Em função do número de componentes principais determinado a partir da matriz  $V$  obtida na etapa anterior, é preciso selecionar tantas colunas quanto número de componentes principais tenha o sistema. Em função disso, será transformada a matriz  $V$  na matriz  $P$ .

**Etapa 6:** Posteriormente é calculada a matriz de projeção  $T$ , em função da matriz  $P$  e a matriz  $X$ . Dita matriz de projeção contém a informação mais importante do sistema, representada com o número total de variáveis do processo em um espaço dimensional reduzido.

**Etapa 7:** A partir da matriz  $T$  e os autovalores do sistema, é calculada a estatística  $T$  de Hotelling ou  $T^2$ , além do limiar do  $T$  de Hotelling para uma variância acumulativa do 95 e 99%, respetivamente.

**Etapa 8:** É feita um gráfico do  $T^2$  em função do número de observações e da matriz de projeção  $T$ , bem como do limiar de  $T^2$  em função do  $n$ . O gráfico deverá mostrar uma tendência dos dados do  $T^2$ , que se encontre por baixo do limiar de  $T^2$ ; o qual apresenta a condição normal do processo, e que no sistema não está acontecendo nenhuma anormalidade.

**Etapa 9:** Com a finalidade de fazer o diagnóstico de falhas, é calculada uma nova matriz de projeção  $T^*$ , em função da matriz  $P$  determinada na etapa 5 e de uma nova matriz de base de dados  $X^*$ , a qual será a matriz de base de dados que apresenta a anormalidade no processo.

**Etapa 10:** É calculada a estatística do  $T^2$  em função da nova matriz de projeção  $T^*$  e dos autovalores do sistema calculados na etapa 4. A tendência gerada será representada no gráfico feita na etapa 8. O gráfico deverá mostrar o momento (tempo – eixo x) em que a anormalidade acontece no sistema; quer dizer, quando a tendência esteja acima do limiar do  $T^2$ .

A Figura 4.1 ilustra as etapas de 1 a 10 descritas acima.

#### 4.1.2 Análise de Contribuições

Segundo RUSSELL, BRAATZ, RICHARD (2000), para se fazer um diagnóstico de falhas usando o método análise de contribuições, levando em conta os resultados obtidos após a aplicação da metodologia do PCA, e com a finalidade de determinar a variável mais relacionada à falha, devem ser feitas um conjunto de operações que foram resumidas na Figura 4.1 (etapas 11-14) e que são a seguir descritas:

**Etapa 11:** Inicialmente é calculada a pontuação normalizada, a qual é ponto chave no cálculo da contribuição de cada variável no processo.

**Etapa 12:** Logo, é calculada a contribuição de cada variável ( $cont_{ij}$ ) em função do número total de variáveis.

**Etapa 13:** Posteriormente se a  $cont_{ij}$  é um número negativo, este deve virar positivo.

**Etapa 14:** Finalmente é determinada a contribuição total ( $CONT_j$ ) da  $j$ -ésima variável do processo e representada em uma mesma figura com a finalidade de avaliar a variável mais relacionada à falha ou variável latente.

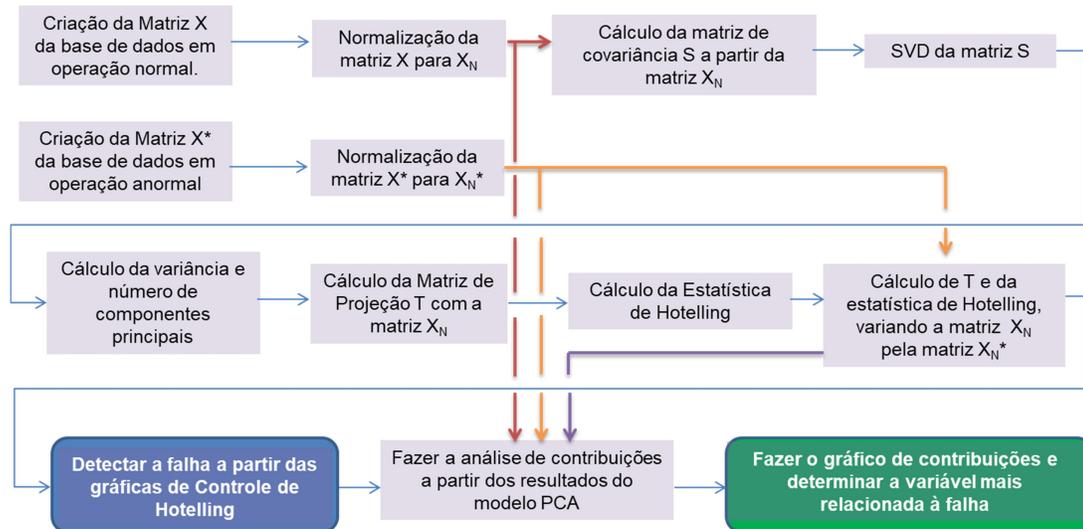


Figura 4.1: Esquema da Metodologia da Análise de Componentes Principais (PCA) e Análise de Contribuições

#### 4.1.3 Mínimos Quadrados Parciais (PLS)

Segundo o trabalho de RUSSELL, BRAATZ, RICHARD (2000) para fazer um diagnóstico de falhas usando o método PLS, devem ser feitas um conjunto de operações que foram resumidas na Figura 4.2 segundo as etapas a seguir:

**Etapa 1:** Inicialmente deve-se criar a Base, composta pelas variáveis de entrada do processo com a maior variabilidade possível (todas as variáveis das quais vai depender o modelo) e dispor a mesma dentro de uma matriz de dados  $X_{n \times m}$ , sendo  $n$  o número de observações e  $m$  o número de variáveis. Assim mesmo, deve ser criado um vetor  $Y$  correspondente a variável de saída do processo, em função da qual ficarão as variáveis de entrada do sistema.

**Etapa 2:** Deve-se fazer a normalização da matriz  $X$ , onde a média seja zero e o desvio-padrão seja unitário (isto será feito usando o comando `StandardScaler` do pacote `sklearn.preprocessing` da ferramenta Python).

**Etapa 3:** É feita uma regressão linear das variáveis de entrada da matriz  $X$  em função do vetor  $Y$ , usando o comando `PLSRRegression` do pacote `sklearn.cross_decomposition` do Python (PEDREGOSA *et al.*, 2011). A partir disto, o produto vai ser um modelo para a variável de saída, composta por coeficientes lineares associados a cada uma das variáveis de entrada do processo.

**Etapa 4:** Para quantificar a qualidade do modelo, se calcula o coeficiente de determinação ( $R^2$ ).

**Etapa 5:** No modelo obtido na Etapa 3 vai ser implementada uma base de dados normalizada em operação normal e outra em operação anormal (distúrbios medidos ou não medidos) do sistema.

**Etapa 6:** Visando se determinar algum desvio indesejado no modelo se faz o gráfico de tendências tanto na base de dados em operação normal quanto em operação anormal, em função do tempo ou número de observações.

**Etapa 7:** Deve ser feito o cálculo do Índice de Ajuste PRESS em função do valor em operação normal da variável de saída (vetor de saída:  $Y$ ) e a predição feita pelo modelo gerado ao implementar a base de dados em operação anormal do processo. A partir de dito índice de ajuste, se poderá quantificar a captura da anormalidade do sistema e, portanto, determinar a variável que esteja mais relacionada com a falha do processo.

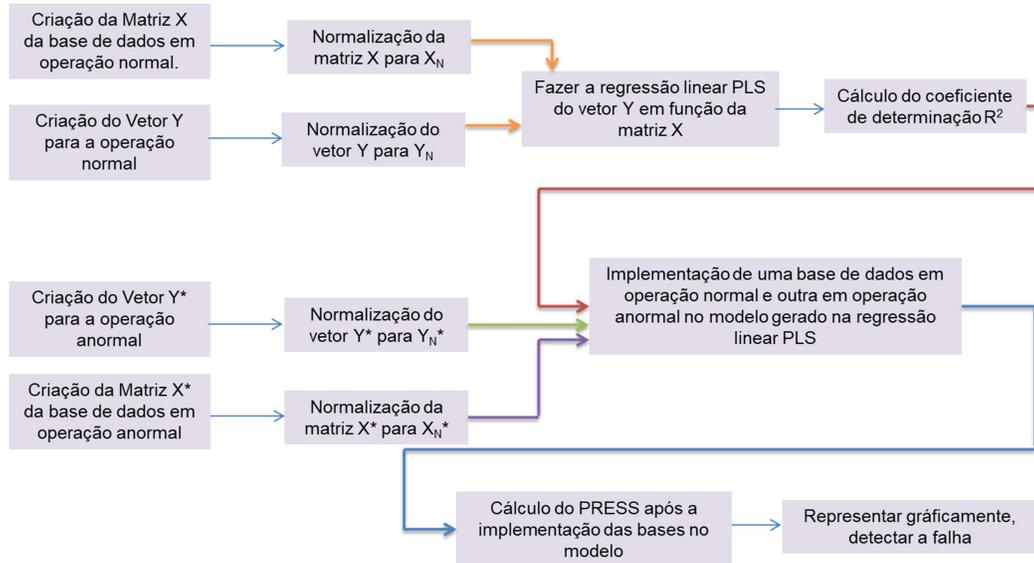


Figura 4.2: Esquema da Metodologia dos Mínimos Quadrados Parciais (PLS)

## 4.2 Metodologia Proposta (EMPVSub)

Como indicado nos objetivos do trabalho, a metodologia proposta tem a finalidade da criação de modelos empíricos a partir de base de dados históricos, para a detecção e diagnóstico de falhas em processos industriais. Os passos da metodologia proposta são mostrados na Figura 4.3 e nas etapas a seguir.

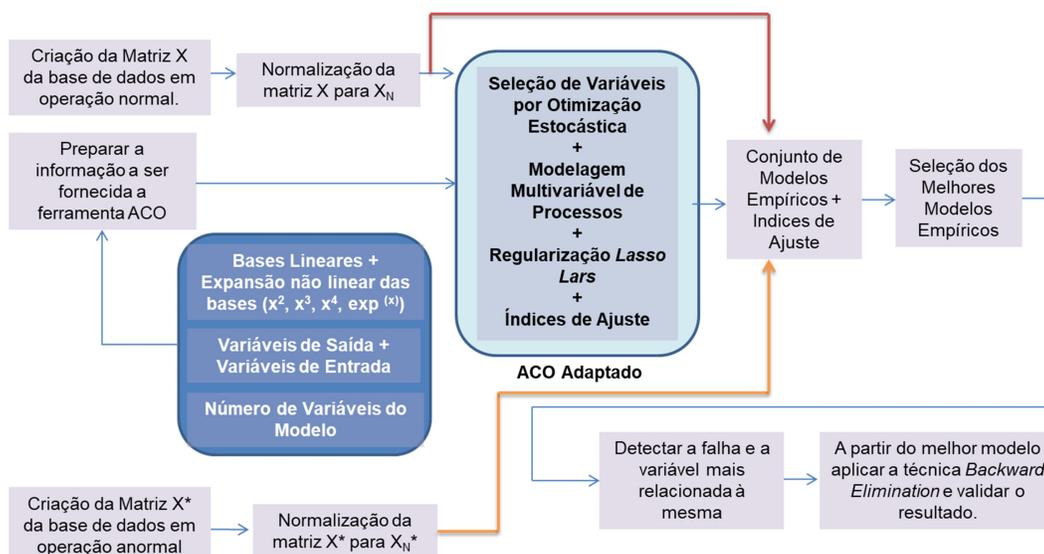


Figura 4.3: Esquema da Metodologia EMPVSub

**Etapa 1:** Inicialmente se deve gerar uma matriz de base de dados  $X$  a partir de um estudo de caso, contendo as variáveis de entrada e as variáveis de saída do sistema (variáveis manipuladas, controladas e medidas). A matriz deverá ter o maior intervalo de variabilidade possível para cada variável, com a finalidade do aumento da probabilidade de captura de dados que estejam fora do intervalo definido como normal.

**Etapa 2:** Montar uma matriz de dados contendo dados correspondentes à operação anormal (quando entra um distúrbio no sistema), seguindo o mesmo procedimento da Etapa 1.

**Etapa 3:** Pré-tratar os dados das Etapas 1 - 2, i.e. fazer uma normalização padrão onde a média seja zero e o desvio-padrão seja unitário, usando o comando `StandardScaler` do Python (PEDREGOSA *et al.*, 2011). Este ponto é fundamental para sistemas cujas variáveis apresentem diferentes ordens de grandezas, pelo qual, convém ter a base de dados normalizada ( $X_N$ ).

**Etapa 4:** A matriz  $X_N$  é processada no algoritmo de otimização ACO modificado (Ranzan *et al.*, 2014) a fim de:

- Determinar as variáveis explicativas mais representativas, as quais serão expandidas através de modelos não lineares através de expansões polinomiais ( $x^2$ ,  $x^3$ ,  $x^4$ ) e exponencial ( $e^x$ ). Após a expansão das bases não lineares, estas deverão ser normalizadas segundo o explicado na Etapa 3.
- A partir da premissa anterior, a matriz  $X_N$ , aumentará o número de colunas em função do número de bases não lineares fornecidas (polinomiais e exponenciais), pois, no momento de fazer a expansão das bases, tem-se uma representação não linear das variáveis contidas na base de dados original, além das demais variáveis de entrada iniciais; desta forma, será ampliado espaço de seleção de variáveis. Por exemplo, tendo  $m$  variáveis na base de dados original,  $i$  expansões polinomiais e  $j$  expansões exponenciais, o número total de variáveis de entrada a serem consideradas pelo ACO na seleção do subconjunto de variáveis a compor cada um dos modelos concorrentes será:  $m_{total} = m + im + jm$ .
- Selecionar os componentes do modelo a ser gerado, ou seja, o conjunto que contém as variáveis de entrada, bem como a variável de saída baseado na matriz  $X_N$ . A partir do conjunto de variáveis de entrada definido, a ferramenta ACO selecionará um subconjunto de variáveis mais relacionadas com a variável de saída.
- Proporcionar o número de modelos empíricos a serem gerados (modelos concorrentes), bem como o número de variáveis desejadas em cada um deles a partir do subconjunto de variáveis de entrada selecionadas pelo ACO na premissa anterior. Isto dependerá dos requerimentos do processo e do analista, fornecendo um bom ajuste de dados com um número reduzido de variáveis.

**Etapa 5:** Uma vez realizadas as etapas anteriores, serão gerados os modelos empíricos para cada uma das variáveis de saída definidas por parte do analista.

**Etapa 6:** A partir dos modelos concorrentes gerados na Etapa 5, são calculados os coeficientes de determinação  $R^2$  para se determinar o melhor modelo.

**Etapa 7:** Para o melhor modelo se calcula o PRESS. Se o valor do PRESS é superior a 1, entre o valor predito pelo modelo ao implementar a base de dados em operação anormal e o valor da variável de saída em operação normal, é detectada a falha.

É importante indicar que nem todos os modelos gerados detectam as falhas, assim, a partir deles é possível descartar as variáveis dependentes deles como possíveis variáveis causadoras do distúrbio no processo. Mas, nos modelos que conseguem detectar a falha, as variáveis dependentes, são consideradas como variáveis latentes ou variáveis que poderiam estar relacionadas com a falha. Se uma variável está presente em todos os modelos concorrentes gerados que capturam a falha, é possível que essa variável seja causadora da falha. No entanto, o anteriormente indicado pode não ser considerado como um resultado concluinte, pois, poderia ser que a variável assumida como mais relacionada à falha só seja peça fundamental para o ajuste do modelo. Sendo assim, tem-se uma metodologia de validação, à metodologia de modelos empíricos, indicada na seguinte seção.

### 4.3 Metodologia de validação (Eliminação de Variáveis)

Uma vez diagnosticada a falha, objetiva-se agora é se determinar qual é a variável mais relacionada à falha do processo, para tanto se adotará um procedimento de eliminação de variáveis baseado no *Backward elimination*, a qual é explicada na Figura 4.3, cujas etapas são detalhadas a seguir:

**Etapa 8:** A partir do melhor/melhores modelos selecionado segundo a metodologia de modelos empíricos, é feita a eliminação de cada uma das variáveis que compõem o ajuste, mantendo as demais com a finalidade de observar as mudanças nas tendências da implementação do modelo nos cenários de estudo; sendo o cenário correspondente à operação anormal do processo o mais importante, pois, se a supressão de uma das variáveis faz que a tendência não capture a anormalidade, quer dizer que a variável eliminada poderia se considerar como a variável mais relacionada à falha.

**Etapa 9:** Simultaneamente à Etapa 8, calcular o *F-Test* e o valor da distribuição F cada vez que seja eliminada uma variável do modelo empírico.

**Etapa 10:** Levando em conta que a hipótese nula consiste em que o modelo onde é eliminada uma variável não oferece um melhor ajuste aos dados do que o modelo original; se o *F-Test* for maior do que a Distribuição F, a hipótese nula é rejeitada, ao rejeitar a hipótese nula, se estaria aceitando que a variável eliminada do modelo é a variável mais relacionada à falha.

**Etapa 11:** Comparar o se o resultado da variável obtida como mais relacionada à falha na metodologia de modelos empíricos é a mesma que para o resultado obtido através da metodologia de validação. Se forem as mesmas, a variável obtida como mais relacionada à falha está correta.

### 4.4 Estudos de caso

As metodologias mostradas nos pontos anteriores são aplicadas em dois estudos de caso. O primeiro consiste em um Tanque de Aquecimento com Mistura e o segundo, trata-se do problema padrão denominado de *Tennessee Eastman Process* (TEP).

## 4.4.1 Estudo de Caso 01: Tanque de Aquecimento com Mistura

O tanque de aquecimento com mistura, representado na Figura 4.4, consiste em um tanque agitado que mistura uma corrente quente e uma corrente fria proveniente de um reservatório de água. A corrente quente é aquecida através de um aquecedor a gás conectado a um controlador *on-off*. Dentro do tanque de mistura, há outro aquecedor elétrico acionado em função do nível do tanque. A tubulação que liga a saída do tanque tem um comprimento significativo o que leva a um elevado tempo morto e a uma perda de calor afetando a temperatura de saída da planta. O objetivo principal é atingir a vazão de saída ( $F_s$ ) e a temperatura de saída ( $T_4$ ) nos valores desejados no processo. Além disso, tem-se que garantir a estabilidade operacional da planta, não permitindo que as resistências sejam queimadas e o tanque transborde ou esvazie.

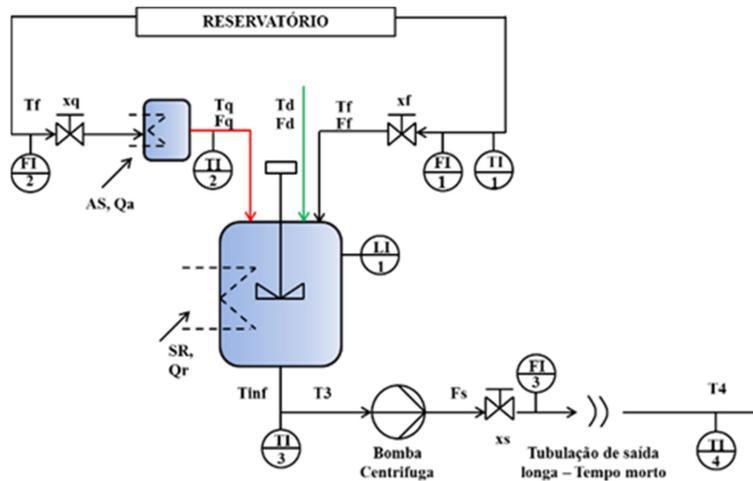


Figura 4.4: Caso de Estudo 01: Sistema de Tanque de Aquecimento com Mistura (Adaptado do GIMSCOP)

Com o objetivo de implementar as metodologias para efetuar diagnóstico de falhas, foi obtida uma base de dados do sistema anteriormente mostrado em operação normal (cenário 01), e posteriormente foram geradas outras duas bases de dados do mesmo sistema, mas em operação anormal (com diversos modos de operação), o qual foi realizado alterando a operação normal do sistema, incluindo dois distúrbios não medidos; sendo o primeiro a entrada de uma corrente fria externa ( $F_d$ ) no tanque de mistura (cenário 02) e o segundo desligando a resistência do aquecedor elétrico ( $Q_r$ ) dentro do tanque de mistura (cenário 03). Isto foi realizado em várias oportunidades para validar a resposta do sistema para distintos modos de operação.

É importante conhecer que o sistema originalmente conta com 12 variáveis (medidas e não medidas), as quais são mostradas na Tabela 4.1.

Tabela 4.1: Variáveis do caso de estudo 01: Tanque de Aquecimento com Mistura

Variável	Descrição	Unidade	Tipo
$X_q$	Abertura da válvula da vazão de água quente	%	Variável Medida
$X_f$	Abertura da válvula da vazão de água fria	%	

Variável	Descrição	Unidade	Tipo
Xs	Abertura da válvula da vazão de saída	%	
Ff	Vazão da água fria	gpm	
Fq	Vazão da água quente	gpm	
Fs	Vazão de saída	gpm	
Tq	Temperatura da água quente	°C	
T3	Temperatura de saída do tanque de aquecimento com mistura	°C	
T4	Temperatura de saída	°C	
H	Altura do Tanque de Aquecimento com mistura	M	
SR	Abertura de válvula	%	
SA	Abertura de válvula	%	
Tinf	Temperatura do Ambiente	°C	
Tf	Temperatura da água fria	°C	
Td	Temperatura da água fria da corrente externa	°C	Variável não Medida
Fd	Vazão da água fria da corrente externa	Gpm	
Qa	Resistencia do Aquecedor Eléctrico do Tanque de água Quente	W	
Qr	Resistencia do Aquecedor Eléctrico do Tanque de Aquecimento com mistura	W	

Os valores da variável de saída Fs e T4, em operação normal, bem como os valores do distúrbio do F<sub>d</sub> e do Q<sub>r</sub>, próprios dos cenários 02 e 03, respetivamente, são mostrados nas Tabelas 4.2 e 4.3, respetivamente.

Tabela 4.2: Valores das variáveis de saída do processo para o caso de estudo 01 em operação normal (cenário 01)

Tempo (s)	0	300	600	900	1200	1500	1800	2100	2400	2700
Fs (gpm) Cenário 01	2,0	3,5	3,0	1,5	4,0	4,5	5,5	2,5	5,0	4,2

T4 (°C)	30								
Cenário 01									

Tabela 4.3: Valores dos distúrbios não medidos, causantes das falhas no sistema (cenário 02 e 03) para o caso de estudo 01

Tempo (s)	150-300	450-600	750-900	1050-1200	1350-1500	1650-1800	1950-2100	2250-2400	2550-2700
Fd (gpm)	1,0	1,7	1,5	0,8	2,0	2,2	2,8	1,2	2,5
Tempo (s)	0-1500 (without disturb)					1500-3000 (disturb)			
Qr (W)	200					0			

#### 4.4.2 Estudo de Caso 02: Tennessee Eastman Process (TEP)

Motivado ao fato de aplicar as metodologias presentes neste trabalho em um estudo de caso mais próximo de um processo real, utilizou-se o estudo de caso da *Tennessee Eastman Process* (TEP), pois é um dos modelos não lineares amplamente utilizados para o estudo dos sistemas de detecção e diagnóstico de falhas (Figura 4.5). O simulador do TEP está disponibilizado em <http://depts.washington.edu/control/LARRY/TE/download.html>. O código é produto da revisão feita por BATHOLT (2015) do TEP original feito por DOWNS, VOGEL (1993).

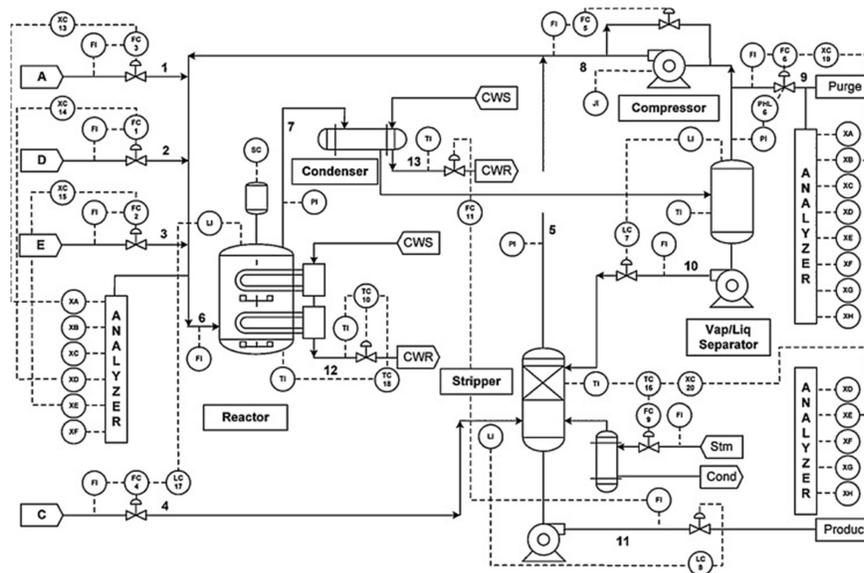


Figura 4.5: Processo Industrial Tennessee Eastman Process (Downs e Vogel, 1993)

O TEP (DOWNS; VOGEL, 1993) é um modelo de processo multivariável, não linear, amplamente adotado na literatura, utilizado para a detecção de falhas usando *thresholds*. O TEP foi criado pela *Eastman Chemical Company* com o intuito de proporcionar um problema baseado em um processo real para avaliar o controle de processos e métodos de monitoramento. Apesar de ser um processo um tanto antigo, o TEP continua a ser uma ferramenta importante para a comparação e/ou validação de algoritmos por ser um modelo não linear de um sistema multicomponente bastante complexo.

O TEP consiste de um processo com os reagentes gasosos A, C, D e E, e um componente inerte B, com dois produtos químicos líquidos, G e H, e um subproduto F, conforme a reação mostrada na Figura 4.6.

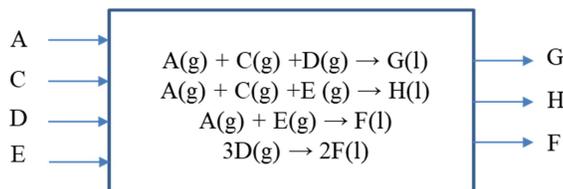


Figura 4.6: Reações que envolvem o TEP (Downs e Vogel, 1993)

As reações são irreversíveis, exotérmicas e aproximadamente de primeira ordem em relação às concentrações dos reagentes. As velocidades de reação são funções de Arrhenius da temperatura, sendo que a reação que tem a maior energia de ativação é a reação de H, resultando em uma maior sensibilidade a temperatura. Este processo tem cinco (5) unidades de operações principais: um reator, um separador líquido-vapor, um *stripper* de produtos, um compressor de reciclo e um condensador de produto.

O processo é iniciado com os reagentes gasosos no reator, que é equipado com uma serpentina de arrefecimento onde circula água e um agitador com velocidade variável. O produto do reator é resfriado por meio de um condensador e, em seguida, alimentado para um separador de vapor líquido. O vapor que sai do separador é reciclado para a alimentação do reator através de um compressor. Uma porção desse fluxo é purgada para evitar que o subproduto e o componente inerte sejam acumulados no processo. Os componentes condensados a partir do separador são bombeados para o *stripper* e são removidos aos reagentes remanescentes. Os produtos G e H que saem da base do separador são enviados para um processo posterior, que não está incluído na planta. O processo possui doze (12) válvulas disponíveis para manipulação (ver Tabela 4.4) e 41 variáveis medidas disponíveis para monitoramento ou controle, das quais na simulação de BATHLET *et al.* (2015) são consideradas só 22 variáveis (ver Tabela 4.5).

Tabela 4.4: Variáveis Manipuladas do Processo (Adaptado de DOWNS E VOGEL, 1993)

Nome da Variável	Número da Variável	Unidade das Variáveis	Nome da Variável	Número da Variável	Unidade das Variáveis
Vazão de alimentação do componente D	XMV_1	Kgh <sup>-1</sup>	Vazão do separador líquido (corrente 10)	XMV_7	m <sup>3</sup> h <sup>-1</sup>
Vazão de alimentação do componente E	XMV_2	Kgh <sup>-1</sup>	Vazão do Separador Líquido do Produto (corrente 11)	XMV_8	m <sup>3</sup> h <sup>-1</sup>
Vazão de alimentação do componente A	XMV_3	Kscmh	Válvula de Vapor do Separador	XMV_9	%

Nome da Variável	Número da Variável	Unidade das Variáveis	Nome da Variável	Número da Variável	Unidade das Variáveis
Vazão de alimentação A e C (corrente 4)	XMV_4	Kscmh	Vazão de água de refrigeração do reator	XMV_10	m <sup>3</sup> h <sup>-1</sup>
Válvula de Reciclo do compressor	XMV_5	%	Vazão de água de refrigeração do condensador	XMV_11	m <sup>3</sup> h <sup>-1</sup>
Válvula de Purga	XMV_6	%	Velocidade do Agitador	XMV_12	RPM

Os objetivos de controle para o TEP são típicos para um processo químico:

- Manter as variáveis de processo nos valores desejados.
- Manter as condições de operação do processo dentro das restrições do equipamento.
- Minimizar a variabilidade da taxa de produto e a qualidade do produto durante a ocorrência dos distúrbios.
- Minimizar o movimento das válvulas que afetam outros processos.

Tabela 4.5: Variáveis Medidas do Processo Contínuo (Adaptado de DOWNS E VOGEL, 1993)

Nome da Variável	Número da Variável	Unidade das Variáveis	Nome da Variável	Número da Variável	Unidade das Variáveis
Alimentação do componente A (corrente 1)	XMEAS_1	Kscmh	Nível do separador do produto	XMEAS_12	%
Alimentação do componente D (corrente 2)	XMEAS_2	Kgh <sup>-1</sup>	Pressão do separador do produto	XMEAS_13	kPa Gauge
Alimentação do componente E (corrente 3)	XMEAS_3	Kgh <sup>-1</sup>	Vazão do separador do produto (corrente 10)	XMEAS_14	m <sup>3</sup> h <sup>-1</sup>
Alimentação da corrente C (corrente 4)	XMEAS_4	Kscmh	Nível do <i>Stripper</i>	XMEAS_15	%
Vazão do reciclo (corrente 8)	XMEAS_5	Kscmh	Pressão do <i>Stripper</i>	XMEAS_16	kPa Gauge
Taxa de alimentação do reator (corrente 6)	XMEAS_6	Kscmh	Vazão do <i>Stripper</i> (corrente 11)	XMEAS_17	m <sup>3</sup> h <sup>-1</sup>

Pressão do reator	XMEAS_7	kPa Gauge	Temperatura do <i>Stripper</i>	XMEAS_18	°C
Nível do reator	XMEAS_8	%	Vazão de vapor do <i>Stripper</i>	XMEAS_19	Kgh <sup>-1</sup>
Temperatura do reator	XMEAS_9	°C	Trabalho do compressor	XMEAS_20	kW
Taxa da purga (corrente 9)	XMEAS_10	Kscmh	Temperatura de saída da água de refrigeração do reator	XMEAS_21	°C
Temperatura do separador de produto	XMEAS_11	°C	Temperatura de saída da água de refrigeração do separador	XMEAS_22	°C

Para testar as metodologias do presente trabalho, com a finalidade da detecção e diagnóstico de falhas, é preciso ter bases de dados que representem alguma anormalidade do processo a ser capturada. O TEP apresenta um conjunto de testes que foram previamente definidos com a finalidade de serem usados para comparar e contrastar abordagens alternativas para operar e controlar automaticamente esse processo. Esses testes são mostrados na Tabela 4.6.

Tabela 4.6: Distúrbios do Tennessee Eastman Process (Adaptado de DOWNS E VOGEL, 1993)

Número de Variável	Variável de Processo	Tipo	Número de Variável	Variável de Processo	Tipo
IDV_1	Relação de alimentação A/C, composição B constante (corrente 6)	Degrau	IDV_8	Composição da alimentação A, B, C (corrente 4)	Variação Aleatória
IDV_2	Composição de B, Taxa constante A/C (corrente 4)	Degrau	IDV_9	Temperatura da alimentação D (corrente 2)	Variação Aleatória
IDV_3	Temperatura da alimentação D (corrente 2)	Degrau	IDV_10	Temperatura da alimentação C (corrente 4)	Variação Aleatória
IDV_4	Temperatura de entrada da água de resfriamento do reator	Degrau	IDV_11	Temperatura de entrada da água de resfriamento do reator	Variação Aleatória

Número de Variável	Variável de Processo	Tipo	Número de Variável	Variável de Processo	Tipo
IDV_5	Temperatura de entrada da água de resfriamento do condensador	Degrau	IDV_12	Temperatura de entrada da água de resfriamento do condensador	Variação Aleatória
IDV_6	Perda da alimentação A (corrente 1)	Degrau	IDV_13	Cinética do Reator	( <i>Slow drift</i> )
IDV_7	Perda de Pressão na alimentação C (corrente 4)	Degrau	IDV_14	Válvula de água de refrigeração do reator	Fixa

Os cenários a serem analisados após aplicar a cada uma das metodologias anteriormente ditas para o estudo de caso 02 são apresentados a seguir:

- Cenário 01: Operação normal para as variáveis XMEAS\_1, XMV\_10 e XMEAS\_9.
- Cenário 02: Ocorrência do distúrbio IDV\_1, após  $t = 8$  hr para a variável XMEAS\_1.
- Cenário 03: Ocorrência do distúrbio IDV\_4, após  $t = 8$  hr para a variável XMV\_10.
- Cenário 04: Ocorrência do distúrbio IDV\_11, após  $t = 8$  hr para a variável XMEAS\_9.

## 4.5 Ferramentas Computacionais

Foi usada a ferramenta Matlab R2012b © para a criação da base de dados dos dois estudos de caso, bem como para estabelecer mudanças nas estruturas de controle no estudo de caso 1. A metodologia foi desenvolvida totalmente através da ferramenta Python 2.7/3.5.

## Capítulo 5 – Resultados

Neste capítulo são mostrados os resultados da metodologia proposta (seção 4.2), a qual consiste na criação de modelos empíricos a partir de dados históricos de processo, onde o número de variáveis do modelo é reduzido a partir de técnicas de regularização. Posteriormente, expande-se em bases não lineares com a finalidade de fazer diagnóstico de falhas em um espaço dimensional reduzido, conseguindo capturar comportamentos não lineares.

Depois feita a análise do diagnóstico de falhas, será corroborado o resultado a partir de uma metodologia de validação (ver seção 4.3), baseada na eliminação das variáveis do melhor modelo obtido na metodologia de modelos empíricos baseado na técnica *Backward Elimination* e técnicas estatísticas como o *F-Test* e a Distribuição de Fisher.

Tanto a metodologia de modelos empíricos quanto a metodologia de validação serão testadas em dois estudos de casos, o primeiro, o tanque de aquecimento com mistura e o segundo o *benchmark Tennessee Eastman Process*.

### 5.1 Estudo de Caso 01: Tanque de Aquecimento com Mistura

#### 5.1.1 Metodologia Padrão 01: Análise de Componentes Principais (PCA)

Aqui são analisados os resultados da aplicação da metodologia PCA para as variáveis de saída selecionadas: vazão de saída (Fs) e Temperatura de Saída (T4) para os cenários 02 e 03. Inicialmente, serão comparados os comportamentos dos cenários 1 e 2, pelo gráfico de  $T^2$  (Figura 5.1).

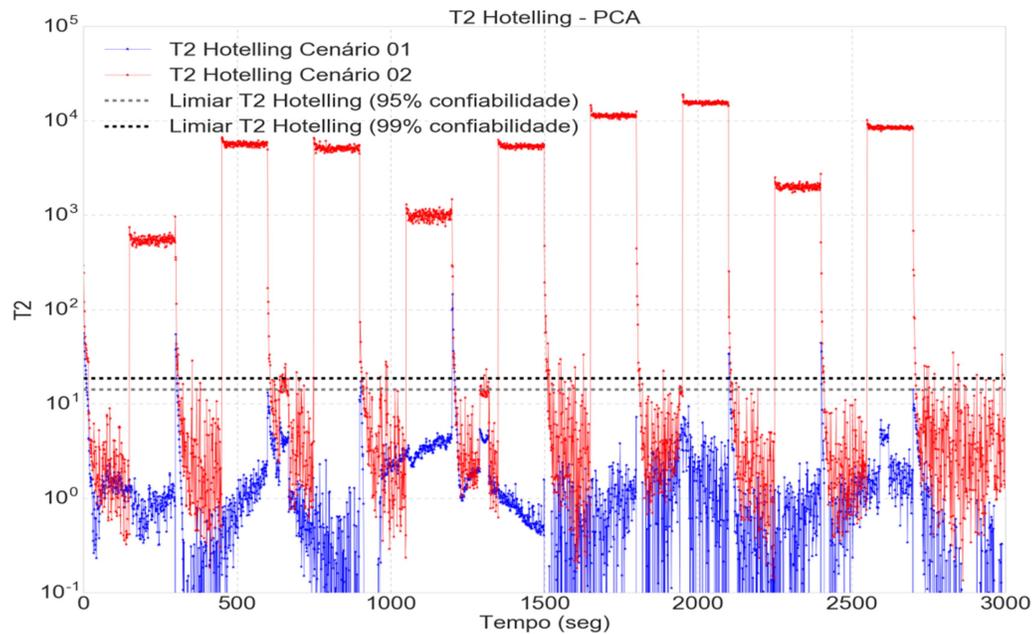


Figura 5.1: Metodologia Padrão 01 (PCA) aplicada ao estudo de caso 01 nos cenários 01 (operação normal) e 02 (operação normal com mudanças de setpoint).

Verifica-se na Figura 5.1 o resultado da aplicação da metodologia do PCA nos cenários 01 e 02, o método captura a tendência da operação normal do sistema do cenário 01, além de enxergar que nenhuma anormalidade está acontecendo (a tendência fica abaixo do limiar  $T_{\alpha}^2$ ). Para o cenário 02, os valores da tendência encontram-se sobre o limiar  $T_{\alpha}^2$ , o que significa que alguma anormalidade está acontecendo, concordando com a natureza da falha e com os resultados mostrados no trabalho de CHIANG, RUSSELL, BRAATZ (2000). Contudo, no momento em que a operação normal do sistema troca de modo de operação, essa mudança é percebida como uma situação anormal (valores intermitentes em azul sobre o limiar do  $T_{\alpha}^2$ ), isto mostra que esta metodologia é sensível as mudanças do set point, apesar de que sejam considerados como uma operação normal do sistema, o que pode ser considerado como uma desvantagem do método.

A aplicação para o cenário 03 é mostrada na Figura 4.2.

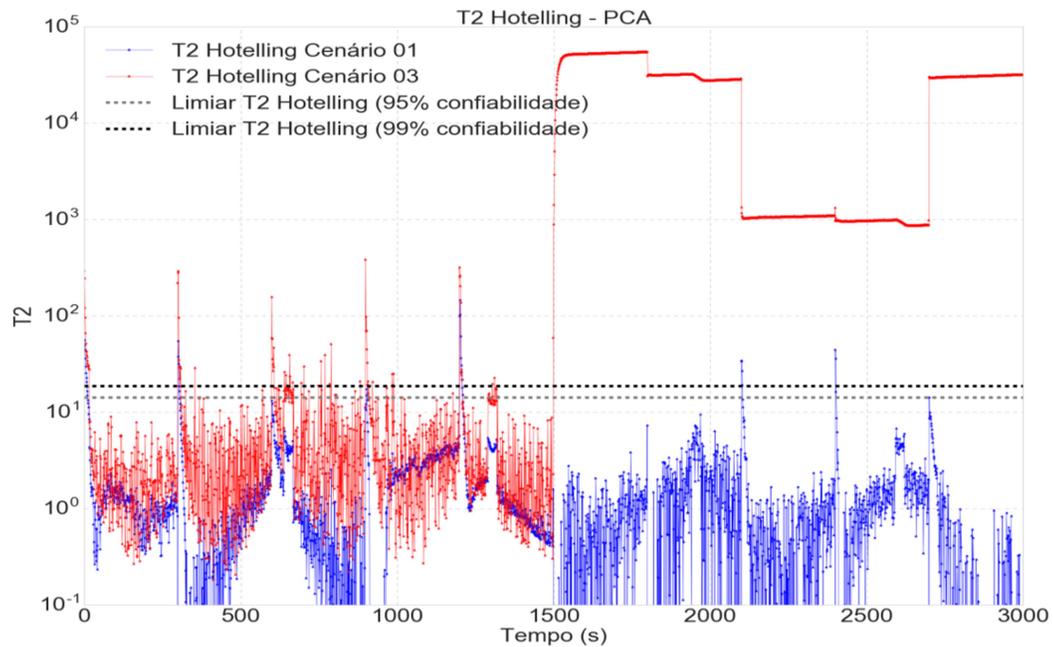


Figura 5.2: Metodologia Padrão 01 (PCA) aplicada ao estudo de caso nos cenários 01 e 03

Na Figura 5.2 verifica-se um comportamento similar ao indicado no cenário anterior: o método consegue detectar que alguma anormalidade está acontecendo no sistema a partir do tempo  $t=1500$  s. Segundo isto, é possível indicar que a metodologia do PCA atinge uma das etapas do diagnóstico de falhas como é o caso da detecção, mas não consegue identificar ou isolar qual variável está mais relacionada com a ocorrência do distúrbio, isto concorda com CHENG (2013) quem indica que o PCA tem uma dificuldade para isolar as várias falhas.

Para isolar e identificar as variáveis, é preciso desenvolver métodos de isolamento e identificação a partir da detecção da falha realizada, como é o caso da análise de contribuições (CHIANG; RUSSELL; BRAATZ, 2000) o qual foi desenvolvido dentro da metodologia padrão segundo o trabalho de CHIANG, RUSSELL e BRAATZ (2000). No entanto, os resultados da aplicação da análise de contribuições, não foram totalmente claros: encontrou-se que para os cenários 01 e 02 representado na Figura 5.3, a variável mais relacionada à falha, foi considerada a  $X_s$  (abertura da válvula de saída). Esta não condiz com o resultado esperado, que seria a diminuição de fluxos de entrada ( $F_f$ ,  $F_q$ ) no primeiro momento para contrapor o fluido novo na entrada ( $F_d$ ), ou alguma modificação de alguma variável relacionada com o aumento da temperatura do sistema, pois a corrente que entra no sistema como distúrbio, é fria; sendo assim o método não é fidedigno.

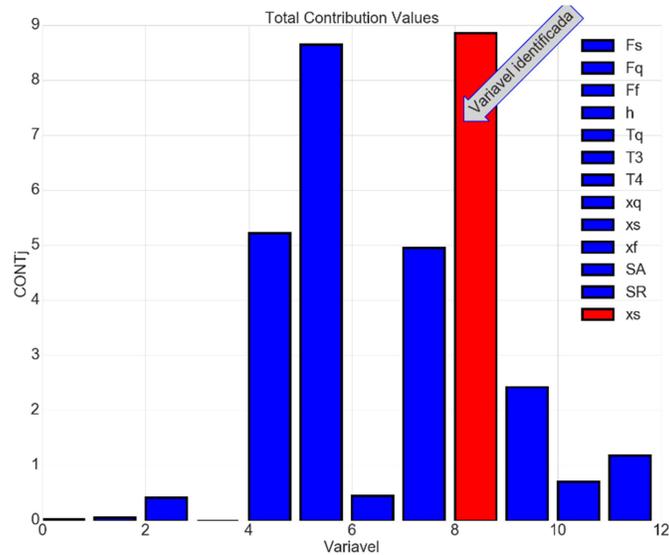


Figura 5.3: Análise de Contribuições para o estudo de caso 01: cenários 01 e 02

Para os cenários 01 e 03, representados na Figura 5.4, os resultados ficam perto do esperado, pois, quando o  $Q_r$  é desligado no  $t=1500$  s, espera-se que as variáveis  $X_q$ ,  $T_3$ ,  $T_4$  ou  $T_q$  reajam para compensar ao distúrbio que está acontecendo no sistema, decorrente de uma mudança na energia adicionada ao sistema, o que se poderia observar nas variáveis associadas ao controle de temperatura. Deste modo, a metodologia indicou que a variável mais relacionada à falha foi a  $X_q$ .

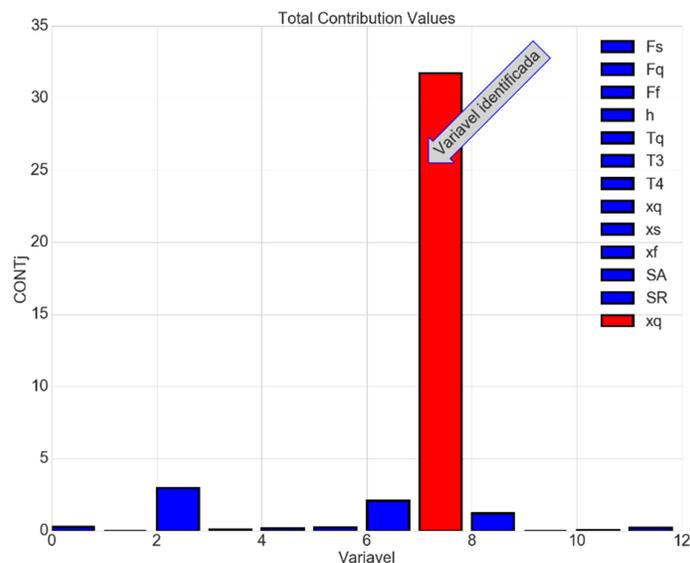


Figura 5.4: Análise de Contribuições para o estudo de caso 01: cenários 01 e 03

Apesar do PCA juntamente com a análise de contribuições conseguirem detectar a falha corretamente, a fase do diagnóstico não é totalmente confiável (como foi possível observar para os cenários 01 e 02). Além disso, o método apresenta como desvantagens o fato que trabalha em todo momento com a totalidade das variáveis do sistema e que enxerga mudanças do set point em operação normal como anormalidades, o que é falso.

### 5.1.2 Metodologia Padrão 02: Partial Least Squares (PLS)

Após a aplicação da metodologia padrão *Partial Least Squares* (PLS), no Estudo de Caso 01 obtiveram-se dois modelos, um para cada uma das variáveis de saída selecionadas: vazão de saída (Fs) e a temperatura de saída (T4), os respectivos parâmetros dos modelos são apresentados na Tabela 5.1.

Na Figura 5.5 são observadas as tendências que segue a variável de saída Fs tanto no cenário 01 e 02, quanto na implementação do modelo obtido pela aplicação da metodologia PLS. Pode se observar que os valores do cenário 01 e do Modelo – Cenário 01, bem como os valores do cenário 02 e do Modelo – Cenário 02 se sobrepõem, visto que ajustam aos dados reais com bastante precisão; fato confirmado através do valor do coeficiente de determinação  $R^2=0,99$  (ver Tabela 5.1). No caso da implementação do Modelo – Cenário 02 é possível observar que os dados seguem uma tendência diferente à tendência da operação normal. Sendo assim, mostra-se que alguma anormalidade está acontecendo no sistema e que o modelo é capaz de capturá-la. Além disso, isto pode ser verificado com o índice de ajuste implementado PRESS, o qual para os casos onde a variável não está sendo influenciada pelo distúrbio no sistema, apresenta o valor de zero (0) e quando a anormalidade afeta a variável apresenta um valor maior que zero.

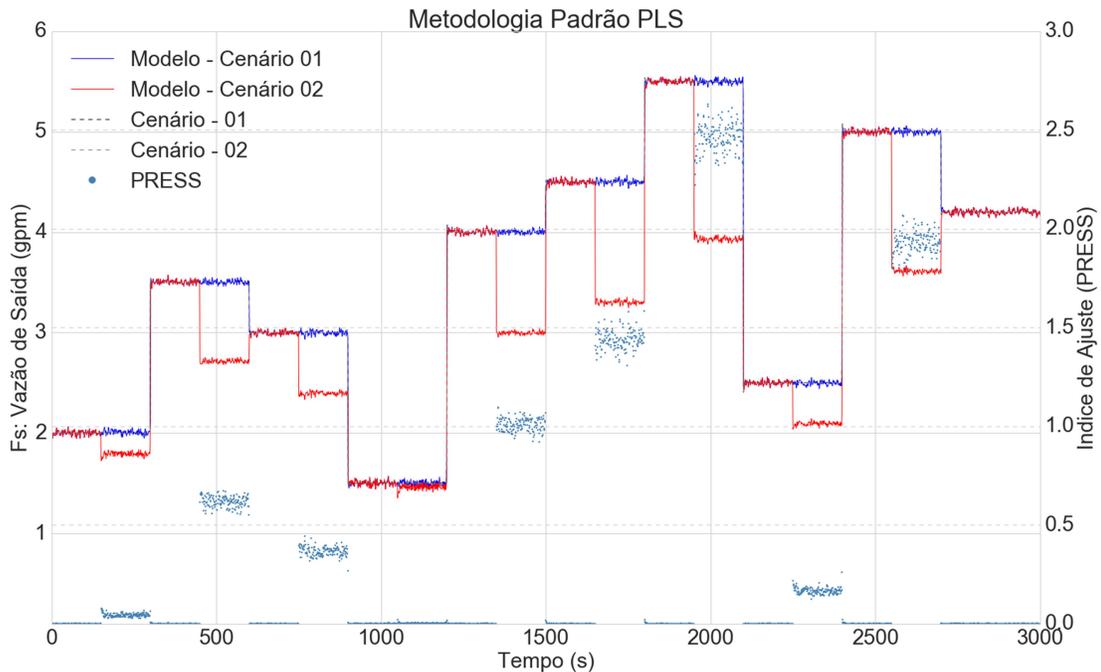


Figura 5.5: Metodologia Padrão 02 (PLS) aplicada ao estudo de caso 01 nos cenários 01 e 02

Na Figura 5.6, pode-se observar que as tendências dos modelos têm um comportamento similar ao caso dos cenários 01 e 02, quer dizer, os cenários 01 e 03 conseguem ser adaptados a partir dos Modelos - Cenário 01 e 03, respectivamente. O Modelo – Cenário 03 consegue capturar o comportamento anormal no processo a partir do tempo  $t=1500$  s, quando a tendência esperada da temperatura muda para um valor menor, a pesar que o modelo não reproduza especificamente o valor real da temperatura T4, o que

pode se validar pelo valor do coeficiente de determinação  $R^2=0,80$  (ver Tabela 5.1) o que significa que a implementação do modelo não mostra um ajuste perfeito do sistema.

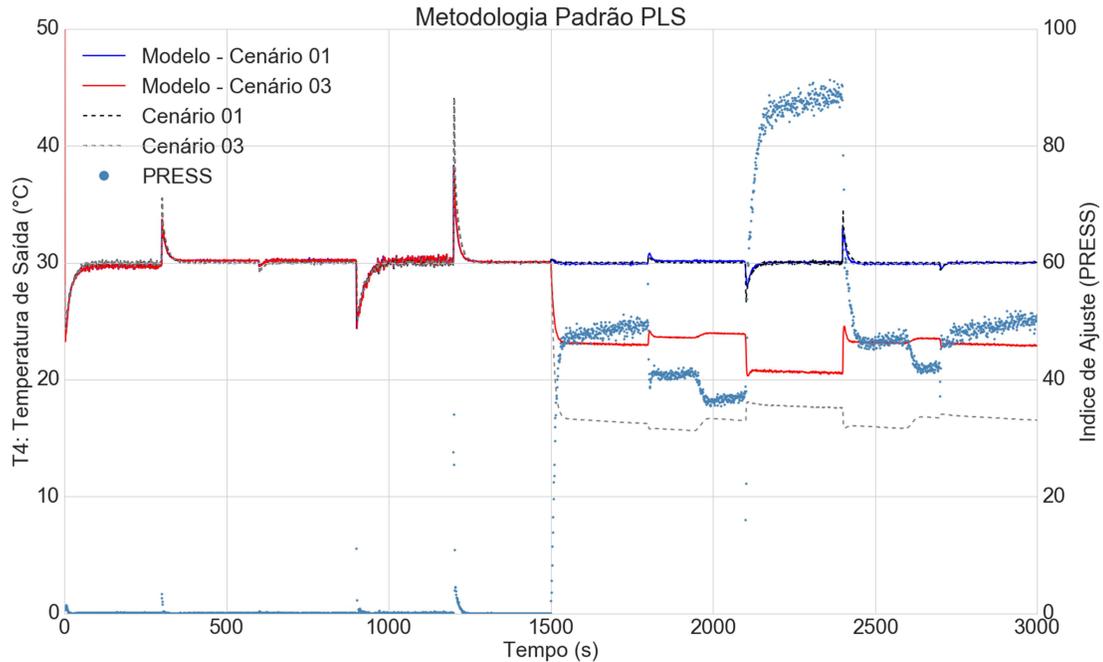


Figura 5.6: Metodologia Padrão 02 (PLS) aplicada ao estudo de caso 01 nos cenários 01 e 03

O anteriormente observado mostra que a metodologia PLS, sendo um método multidimensional de projeção múltipla, consegue ser um bom método para detecção de falhas, pois consegue extrair informação nos dados da trajetória multivariada, detectando a ocorrência de perturbações observáveis, concordando com KOURTI; NOMIKOS; MACGREGOR (1995). Além do diagnóstico de falhas, o modelo gerado pode ser utilizado para outros fins, sendo isto uma vantagem para o método PCA, concordando com CHIANG; RUSSELL; BRAATZ (2000). No entanto, não diminui a dimensionalidade do problema.

Os casos dos cenários 01 & 02 e 01 & 03, dos modelos mostrados na Figura 5.5 e na Figura 5.6, reproduzem a metodologia de CHIANG, RUSSELL, BRAATZ (2000) efetivamente, mas mostram uma das desvantagens do método como é que os modelos gerados estão compostos pela totalidade das variáveis do processo (11 variáveis), sendo isto desfavorável no momento de determinar a/as variável/variáveis que estejam mais relacionadas com a causa do problema.

Tabela 5.1: Parâmetros do Modelo para as variáveis de saída Fs e T4

Variáveis do Modelo	Coeficientes das variáveis do Modelo	
	Vazão de Saída (Fs)	Temperatura de Saída (T4)
Fq	-39,61	1161,16
Ff	10,20	-12,32
H	-14,44	734,01
Tq	-0,00	0,03
T3	0,01	-1,03
Fs	-	1,69
T4	4,38	-
Xq	1,80	-136,87
Xf	0,26	-54,47
Xs	-0,00	44,80
SA	-0,00	-0,01
SR	0,01	0,28
R <sup>2</sup>	0,99	0,80

### 5.1.3 Metodologia Proposta: Modelos Empíricos

Na Tabela 5.2, mostra-se para cada um dos sete (7) modelos empíricos obtidos a partir da metodologia proposta, para cada uma das saídas do processo, suas variáveis dependentes e o R<sup>2</sup> para cada um dos casos.

Tabela 5.2: Modelos Empíricos para o caso de estudo 01 a partir da ferramenta ACO

Variável de Saída/ Coeficiente de Determinação	Modelos Empíricos						
	1	2	3	4	5	6	7
Vazão de Saída (Fs)	Ff, h	xs, xf, h	xs, xf, h, xq <sup>3</sup>	xs, xf, h, xq <sup>3</sup> , T4 <sup>3</sup>	xq, xs, Ff, h, SA <sup>4</sup>	xq, xs, xf, h, xI <sup>2</sup> , T4 <sup>3</sup>	xq, xs, xf, h, exp (xs) xI <sup>2</sup> , FI <sup>3</sup> , T4 <sup>3</sup>
R <sup>2</sup>	0,97	0,97	0,97	0,97	0,97	0,98	0,98
Temperatura de Saída (T4)	exp(Fq), exp(Ff)	Fq, Ff, h	xq, Fq, T3, T3 <sup>2</sup>	xq, Fq, T3, T3 <sup>2</sup> , h <sup>3</sup>	xf, h, T3, xq <sup>2</sup> , T3 <sup>2</sup> , T3 <sup>3</sup>	Fq, h, exp(T3), Fs <sup>2</sup> , T3 <sup>2</sup> , Fq <sup>3</sup> , T3 <sup>4</sup>	Ff, h, T3, exp(SR), exp(Fs), Fs <sup>2</sup> , T3 <sup>2</sup> , T3 <sup>3</sup>
R <sup>2</sup>	0,13	0,28	0,79	0,15	0,90	0,86	0,85

Após obtidos os modelos, foi possível fazer a seleção dos melhores para cada uma das variáveis de saída definidas, em função do valor do coeficiente de determinação ( $R^2$ ). É importante ressaltar que todos os modelos gerados para a variável de saída  $F_s$ , apresentam um valor do  $R^2$  maior do que 0,97 (ver Tabela 5.2), o qual representa um bom ajuste dos dados para qualquer um deles, bem como a garantia de que o modelo deveria reproduzir as tendências dos dados em operação normal e enxergar as anormalidades do sistema quando as variáveis que compõem o modelo estejam relacionadas com a falha, fornecendo uma probabilidade de serem causadoras diretas ou indiretas do distúrbio. Desta forma, foi possível observar a sobreposição dos valores reais versus os valores obtidos do  $F_s$  após a implementação do modelo em operação normal, bem como a detecção da anormalidade após a implementação do modelo na base de dados em operação anormal. Apenas  $F_f$  e  $h$  (i.e., modelo empírico 1) são suficientes para representar o sistema.

Assim, foi selecionado o Modelo Empírico 1 representado na Figura 5.7, como o melhor modelo, o qual consegue um bom ajuste dos dados ( $R^2=0,97$ ) e permite se detectar facilmente a anormalidade do processo e representa a máxima diminuição da dimensionalidade do sistema (diminuição de 11 para 2 variáveis), fornecendo uma otimização do diagnóstico de falhas (uma redução de 81,82%).

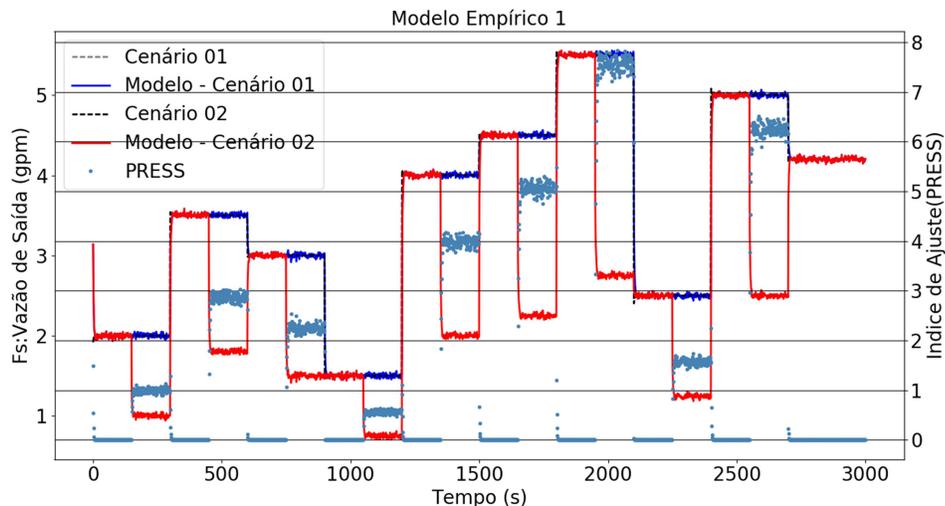


Figura 5.7: Modelo Empírico 1:  $F_s$  (Cenários 01 e 02)

Na Figura 5.7 fica evidente a ocorrência de uma situação anormal no processo no intervalo de tempo indicado na Tabela 5.2, além disso, é possível observar que os valores do  $PRESS \geq 0$ , encontram uma forte diferença entre as tendências da implementação do modelo para os dois cenários, corroborando a anormalidade no processo. Além disso, é possível determinar outra informação relevante dessa figura: o modelo só com duas variáveis consegue detectar a situação anormal, demonstrando que dependendo do modelo concorrente selecionado é possível trabalhar com um número menor de dados e ainda assim capturar a informação de interesse podendo descartar os demais modelos, evitando desta forma violar o Princípio de Parcimônia (BISHOP, 2006), o qual foi possível implementando a metodologia proposta no caso de estudo, resolvendo o problema de *overfitting* usando técnicas de regularização na criação dos modelos empíricos concorrentes.

No entanto, a Figura 5.8 mostra a representação do modelo empírico 6 para a variável de saída  $F_s$ , a qual apresenta um comportamento similar ao mostrado pelo modelo empírico 1, só que este apresenta uma redução da dimensionalidade de 11 para 7 variáveis (redução do 36,36 %), e é possível observar que as variáveis  $h$  e  $F_f$  (baixa a forma de abertura de válvula da vazão da água fria) estão presentes nos parâmetros do modelo, conferindo a análise feita no parágrafo anterior.

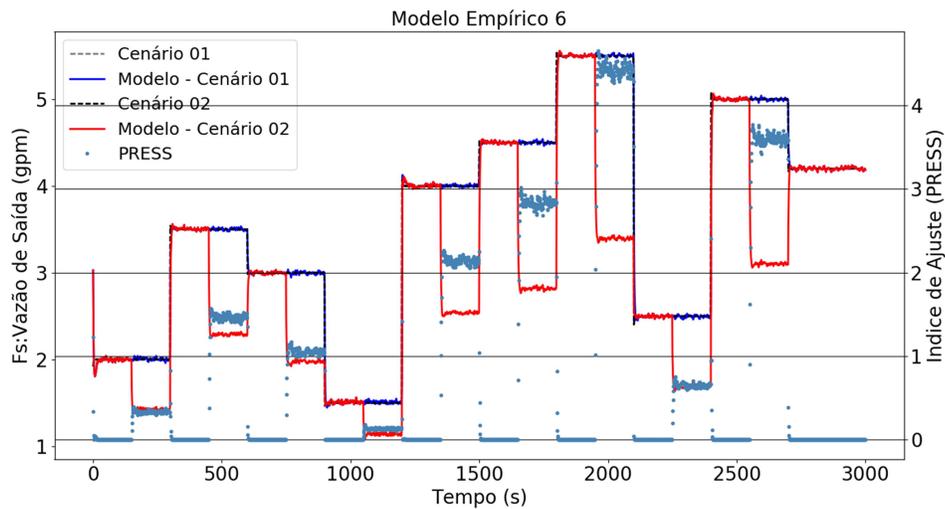


Figura 5.8: Modelo Empírico 6:  $F_s$  (cenários 01 e 02)

No caso da variável de saída  $T_4$ , é possível se observar a partir dos valores do  $R^2$  mostrados na Tabela 5.2, que todos os modelos empíricos gerados a partir da metodologia proposta não foram tão eficientes quanto para o caso da variável de saída  $F_s$ . No entanto, os modelos 5, 6 e 7 (que apresentam o melhor ajuste) foram estudados, com a finalidade de analisar a qualidade de ajuste e a possibilidade de capturar a anormalidade no processo. Sendo assim, escolheu-se o modelo empírico 6 como o melhor modelo, como mostrado na Figura 5.9. Este consegue um bom ajuste dos dados bem como a captura da anormalidade, fornecendo uma redução da dimensionalidade do sistema de 11 para 7 variáveis (uma redução de 36,36%).

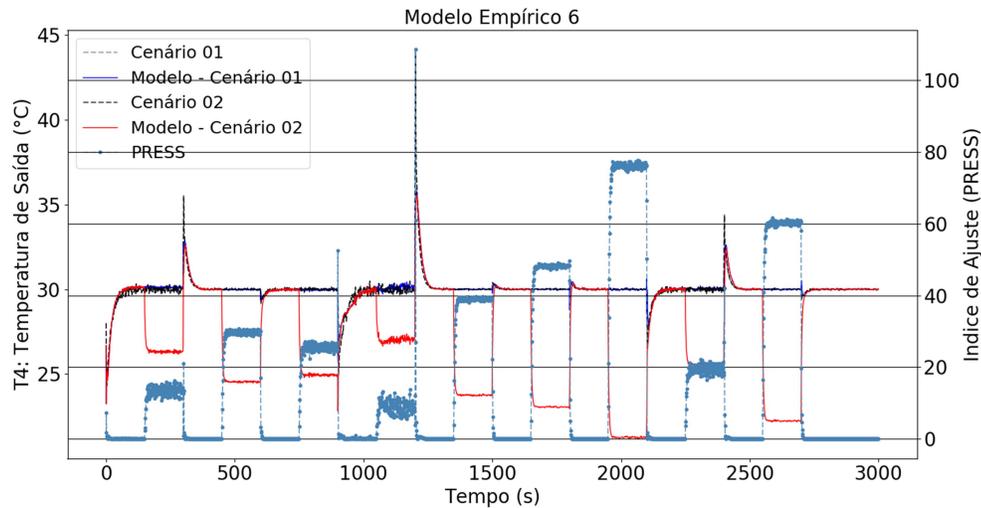


Figura 5.9: Modelo Empírico 6: T4 (cenários 01 e 02)

Nota-se que a partir dos melhores modelos selecionados para as variáveis de saída T4 e  $F_s$  é possível detectar quais variáveis são comuns entre ambos modelos, o que aumenta a relação dessas variáveis comuns com a anormalidade (afetadas pelo mesmo distúrbio, cenário 02). Neste caso, a variável comum entre os dois modelos são a altura (h) e a vazão  $F_f$ , desta forma, h e  $F_f$  são as variáveis mais relacionadas à falha ou tem a ver com uma variável medida ou não medida, relacionada ou causadora da falha, diretamente ou indiretamente. Assim, pode-se centrar a atenção em todas as variáveis e subsistemas relacionados com a altura do tanque de aquecimento com mistura, sendo isto vazões medidas:  $F_f$ ,  $F_q$  (junto com as variáveis  $X_f$  e  $X_q$ ) ou vazões não medidas (vazões externas), concordando isto com a causa conhecida (cenário 02), que consiste na entrada de uma vazão externa no tanque  $F_d$ .

Posteriormente foi aplicada a metodologia proposta, levando em conta a base de dados em operação anormal do cenário 03, no entanto, os resultados obtidos para a variável de saída  $F_s$ , apesar dos modelos empíricos obtidos terem sido bons tendo como base, o valor  $R^2$  da Tabela 5.2, nem todos conseguiram detectar a anormalidade do processo (caso dos modelos 1, 2, 3 e 5). Isto demonstra que as variáveis que compõem ditos modelos não estão relacionadas com o distúrbio. Contrariamente, os modelos 4, 6 e 7 detectam a situação anormal, sendo que, pelo menos uma das variáveis que compõem estes modelos está relacionada com a falha. Este ponto é importante, porque descartando esses quatro modelos e as variáveis que compõem os mesmos, é possível reduzir a dimensionalidade do problema e focar a atenção do diagnóstico nas variáveis dependentes dos modelos que capturam a anormalidade que não estão presentes nos modelos descartados. Neste caso a única variável relacionada à falha é a T4. A Figura 5.10 mostra o modelo empírico 6, o qual foi selecionado como o melhor modelo que representa o processo e enxerga a anormalidade definida no cenário 03. Este modelo fornece uma redução da dimensionalidade no processo de 11 para 7 variáveis (uma redução de 36,36%).

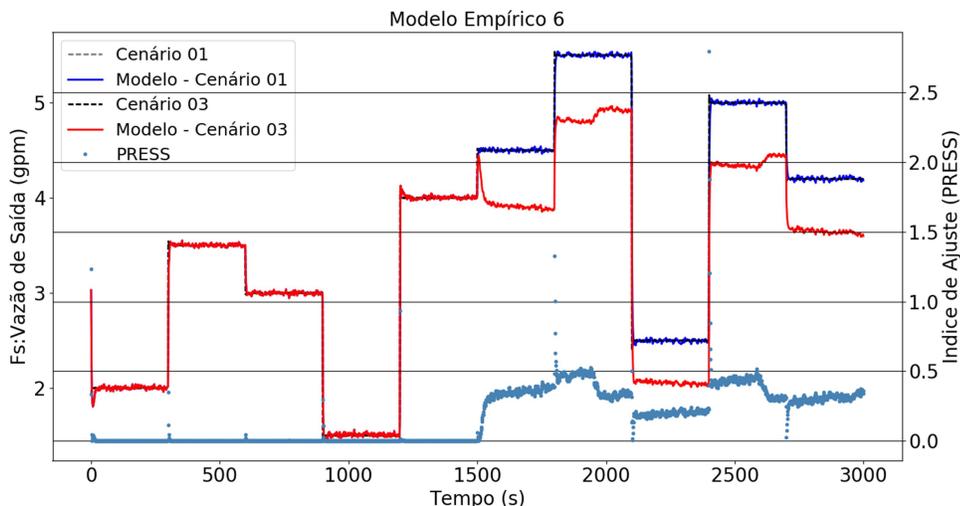


Figura 5.10: Modelo Empírico 6 Fs (Cenários 01 e 03)

De forma similar, mostra-se o resultado do melhor modelo empírico selecionado para a variável T4 (modelo empírico 7, diminuindo a dimensionalidade do sistema de 11 para 8 variáveis, correspondendo uma redução de 27,27%), levando em conta o cenário 03, o qual é representado na Figura 5.11. É possível observar que a partir do tempo  $t = 1500$  s, acontece uma situação anormal no sistema que produz a drástica diminuição da T4. Os modelos que capturam o distúrbio têm como a variável mais relacionada a temperatura T3 (descartando as variáveis não relacionadas: vazões e níveis). Portanto, a falha no sistema é causada por algum fenômeno que altera a temperatura, sendo assim, uma alta probabilidade que a falha tenha origem na energia fornecida ao sistema.

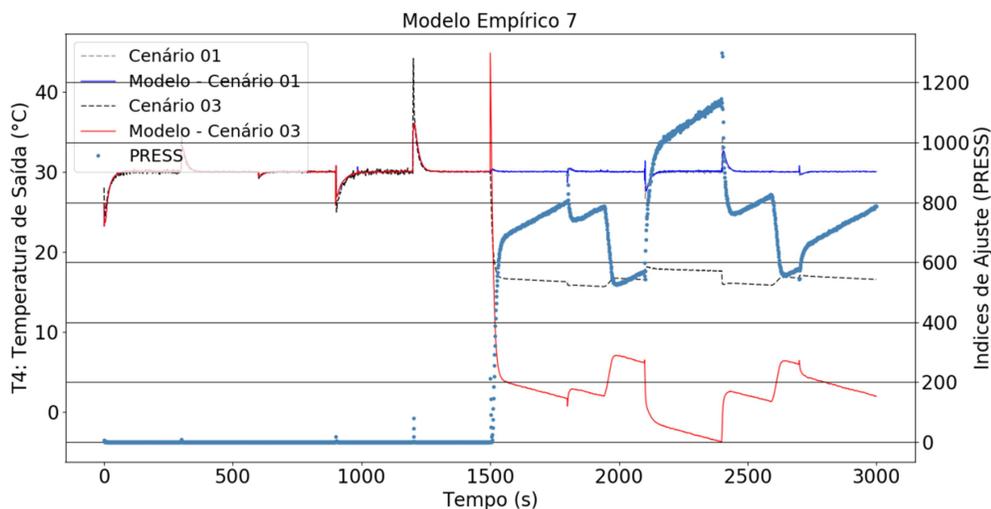


Figura 5.11: Modelo Empírico 7: T4 (Cenários 01 e 03)

Para cada um dos cenários estudados é possível observar que a diminuição da dimensionalidade do processo facilita o diagnóstico de falhas, pois, diminui a incerteza que variáveis que não estejam relacionadas com a causa da anormalidade poderiam trazer para

o diagnóstico. Há também melhora no desempenho computacional, pois se tem menos dados a serem processados, bem como o desempenho dos estimadores de regressão linear (ORR, 1995). O anteriormente indicado é produto do módulo da regularização dentro da metodologia proposta.

#### 5.1.4 Metodologia de Validação: Eliminação de Variáveis

Aplicando a metodologia de validação, baseado na técnica *Backward Elimination* nos melhores modelos obtidos segundo a metodologia proposta, é possível validar o resultado da variável considerada como mais relacionada à falha para a maioria dos cenários, conforme apresentado a seguir.

#### Cenários 01 e 02

Inicialmente é importante ressaltar que a hipótese nula consiste em que o modelo onde é eliminada uma variável não oferece um melhor ajuste aos dados do que o modelo original. A hipótese nula é rejeitada se o valor do F-Test  $> F_0$ . Ao rejeitar a hipótese nula, se estaria aceitando que a variável eliminada do modelo é a variável mais relacionada à falha.

Para o caso da variável de saída  $F_s$ , as variáveis que estão mais relacionadas à falha segundo o modelo empírico 1 são as variáveis  $F_f$  e  $h$  (diretamente dependentes entre si no processo) pois detectam a anormalidade segundo observado na Figura 5.7. Para o caso da variável de saída  $T_4$  o melhor modelo obtido foi o modelo empírico 6, onde as variáveis que o compõem são  $F_q$ ,  $h$ ,  $T_3$ ,  $F_s$  e  $T_3$ , sendo que o modelo captura o comportamento anormal próprio do cenário 02, é possível que alguma de ditas variáveis esteja relacionada com a falha, conforme explicado nos resultados da metodologia de modelos empíricos.

As Tabelas 5.3, 5.4 e 5.5 mostram os resultados da Metodologia de Eliminação de Variáveis para as variáveis de saída  $F_s$  e  $T_4$  no estudo de caso 01, em função do F-Test e a Distribuição de Fisher para cada um dos melhores modelos obtidos após a implementação da metodologia de modelos empíricos.

Tabela 5.3: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 1 da variável de saída  $F_s$  nos cenários 01 e 02 ( $i$ =variável contida no modelo empírico)

Testes Estatísticos	Variável Explícita	
	$i=h$	$i=F_f$
$F_{0,i}$	3,840	3,840
F-Test $i$	0,983	0,987
F-Test $i > F_{0,i}$	-	-

Eliminando qualquer uma das duas variáveis do modelo empírico 1 para a variável de saída  $F_s$ , segundo o resultado do F-Test  $i < F_{0,i}$  mostrado na Tabela 5.3 nenhuma variável deveria ser rejeitada do modelo, sendo assim, a validação do modelo empírico 1, não permite concluir qual das duas poderia ser a variável mais relacionada á falha, mas é possível dizer que o modelo depende das duas variáveis para dar o melhor ajuste aos dados de entrada. O mesmo acontece com o modelo empírico 6, segundo os resultados mostrados

na Tabela 5.4, em função dos quais resulta impossível determinar qual variável deveria ser rejeitada do modelo para melhorar seu ajuste, no momento em que acontece a falha.

Tabela 5.4: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 da variável de saída Fs (i=variável contida no modelo empírico)

Testes Estatísticos	Variável Explícita					
	i=xq	i=xs	i=xf	i=h	i=xf <sup>2</sup>	i=T3 <sup>4</sup>
F <sub>0,i</sub>	2,102	2,102	2,102	2,102	2,102	2,102
F-Test i	0,245	0,767	0,240	0,970	1,103	0,99
F-Test $i > F_{0,i}$	-	-	-	-	-	-

No entanto, é possível usar o outro conjunto de modelos empíricos gerados pela metodologia para validar os resultados dos cenários 01 e 02 correspondentes a variável de saída T4. No caso do modelo empírico 6, não estão presentes ditas variáveis (h e Ff) dentro do mesmo modelo, o qual, melhora a expectativa do diagnóstico, contrariamente aos modelos empíricos 1 e 6 da variável de saída Fs.

A Tabela 5.5 apresenta os resultados da metodologia de validação para a variável de saída T4 nos cenários 01 e 02, onde é possível detectar que eliminando as variáveis Fq ou Fs, o modelo melhora. Sendo assim que estas variáveis poderiam estar relacionadas à falha do sistema, sendo estas vazões do fluido. Deste modo, é possível indicar que ditas variáveis têm problemas ou têm relação com algum problema que esteja ocasionando uma falha no processo. O qual é certo, pois o distúrbio Fd, é um distúrbio não medido e nunca vai ser detectado diretamente pelo modelo, mas sim pelas variáveis mais relacionadas a dito distúrbio, que corresponderia neste caso as vazões Fq ou Fs.

Tabela 5.5: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 para a variável de saída T4(i=variável contida no modelo empírico)

Testes Estatísticos	Variáveis Explícitas						
	i=Fq	i=h	i=exp(T3)	i=Fs <sup>2</sup>	i=T3 <sup>2</sup>	i=Fq <sup>3</sup>	i=T3 <sup>4</sup>
F <sub>0,i</sub>	2,102	2,102	2,102	2,102	2,102	2,102	2,102
F-Test i	2,171	0,989	0,025	2,257	0,152	2,115	0,166
F-Test $i > F_{0,i}$	X	-	-	X	-	X	-

### Cenários 01 e 03

Para os cenários 01 e 03, os resultados da implementação da metodologia de eliminação de variáveis, são mostrados nas Tabelas 5.6 e 5.7. É possível observar que a

variável que precisa ser rejeitada do modelo resulta em T4 para a variável de saída Fs, e a variável T3 para a variável de saída T4, corroborando o resultado obtido após feita a análise na metodologia de modelos empíricos. Deste modo, é validado que a falha IDV\_1, está relacionada com a temperatura do tanque de aquecimento com mistura, desta forma, o analista deveria inspecionar o sistema associado a resistências e sistemas fornecedores de energia.

Tabela 5.6: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 para a variável de saída Fs (i=variável contida no modelo empírico)

Testes Estatísticos	Variáveis Explícitas					
	i=xq	i=xs	i=xf	i=h	i=xf <sup>2</sup>	i=T4 <sup>3</sup>
<b>F<sub>0,i</sub></b>	2,102	2,102	2,102	2,102	2,102	2,102
<b>F-Test i</b>	0,292	0,418	0,041	0,999	1,181	45,229
<b>F-Test <sub>i</sub> &gt; F<sub>0,i</sub></b>	-	-	-	-	-	X

Tabela 5.7: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 7 para a variável de saída T4 (i=variável contida no modelo empírico)

Testes Estatísticos	Variável Explicativa						
	i=Ff	i=h	i=T3	i=exp(Fs)	i=Fs <sup>2</sup>	i=T3 <sup>2</sup>	i=T3 <sup>3</sup>
<b>F<sub>0,i</sub></b>	2,012	2,012	2,012	2,012	2,012	2,012	2,012
<b>F-Test i</b>	0,877	1,000	4,766	0,378	0,914	8,134	2,621
<b>F-Test <sub>i</sub> &gt; F<sub>0,i</sub></b>	-	-	X	-	-	X	X

## 5.2 Estudo de Caso 02: Tennessee Eastman Process (TEP)

### 5.2.1 Metodologia Padrão 01: Análise de Componentes Principais (PCA)

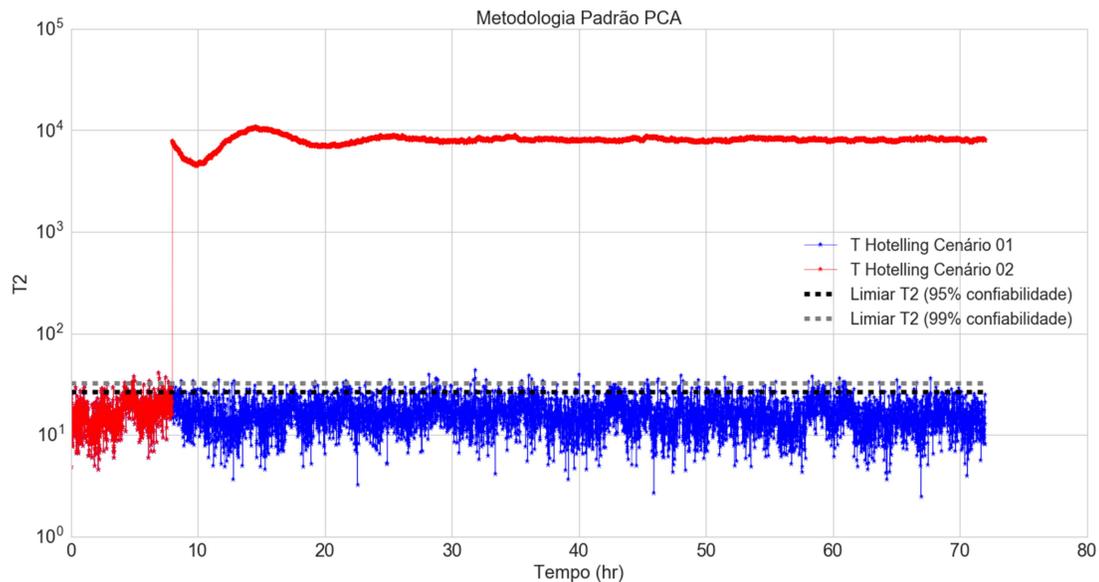


Figura 5.12: Metodologia Padrão (PCA) aplicada ao estudo de caso 02 nos cenários 01 e 02

A Figura 5.12 mostra o resultado da estatística T de Hotelling para o estudo de caso 02, quando o distúrbio que acontece no sistema é aquela do cenário 02, quer dizer, quando acontece uma mudança no degrau na relação de alimentação A/C (IDV\_1). Pode-se observar que a partir do tempo  $t=8$  hr a metodologia consegue capturar que alguma anormalidade está acontecendo no sistema, mas não é possível determinar as variáveis que estão mais relacionadas com a causa de dita anormalidade.

CHIANG, RUSSELL, BRAATZ (2000) fizeram um estudo mais aprofundado no diagnóstico de falhas empregando o PCA como técnica de detecção combinando o PCA tradicional com o método análise de contribuições, onde é possível diagnosticar qual é a variável mais relacionada com a causa da anormalidade. Tal resultado pode se observar na Figura 5.13, onde se mostra corretamente que a variável que está mais relacionada com a causa da falha no sistema é XMV\_3 (vazão de alimentação do componente A). A desvantagem deste método (combinação do PCA tradicional com a análise de contribuições) apesar de ser efetivo, é que ele utiliza em todo momento a totalidade das variáveis da base de dados do sistema, o qual não é ótimo para sistemas de grande quantidade de variáveis.

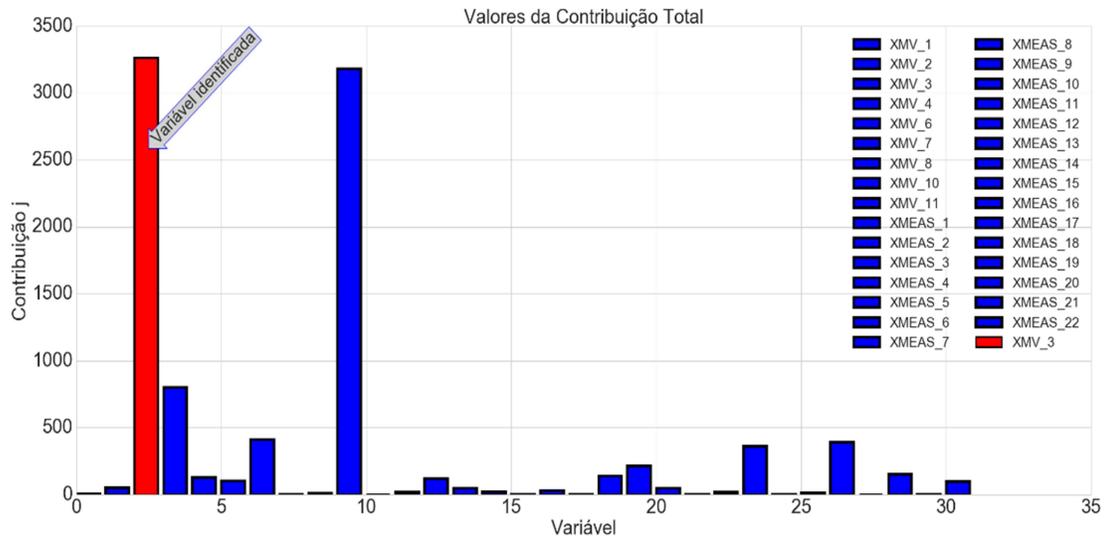


Figura 5.13: Análise de Contribuições aplicada ao estudo de caso 02 nos cenários 01 e 02

Similar ao caso anterior, a Figura 5.14 mostra o resultado da aplicação da metodologia do PCA no caso de estudo 02 no cenário 03, i.e. quando acontece uma mudança tipo degrau na temperatura de entrada da água de resfriamento do reator (IDV\_4). Pode-se observar que no tempo  $t=8$  hr acontece alguma anormalidade, mas não é possível diagnosticar qual é a variável mais relacionada com o distúrbio.

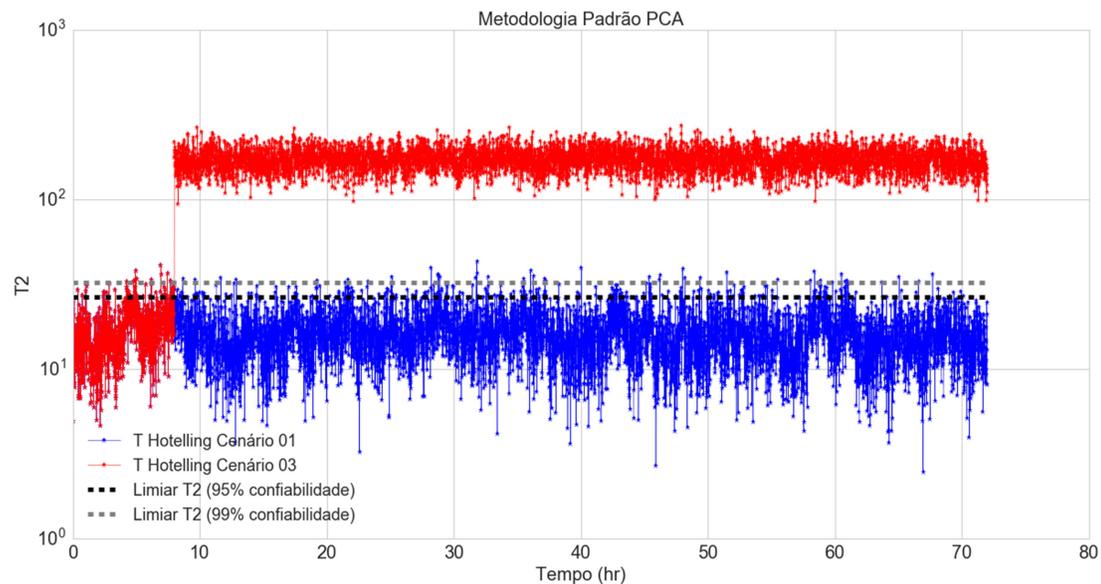


Figura 5.14: Metodologia Padrão (PCA) aplicada ao estudo de caso 02 nos cenários 01 e 03

Aplicando a combinação da técnica PCA tradicional mais a análise de contribuições é obtido que a variável mais relacionada à falha do sistema é a vazão da água de refrigeração do reator (XMV\_10) mostrada na Figura 5.15. Certamente esta não é a variável mais relacionada com a falha, pois o diagnóstico deveria ter mostrado a variável temperatura do reator (XMEAS\_9) ou a temperatura da água de refrigeração do reator propriamente, mas esta última não é uma variável medida no sistema. Não obstante, da totalidade das variáveis do sistema que são medidas e estão relacionadas à falha se tem a XMV\_10 e XMEAS\_9. A metodologia mostra como resultado a variável vazão da água de refrigeração do reator (XMV\_10) como a variável mais relacionada à falha no sistema, a

qual, apesar de não ser a melhor, não está totalmente errada, pois dita vazão da água do reator é regulada dependendo do valor da temperatura do reator.

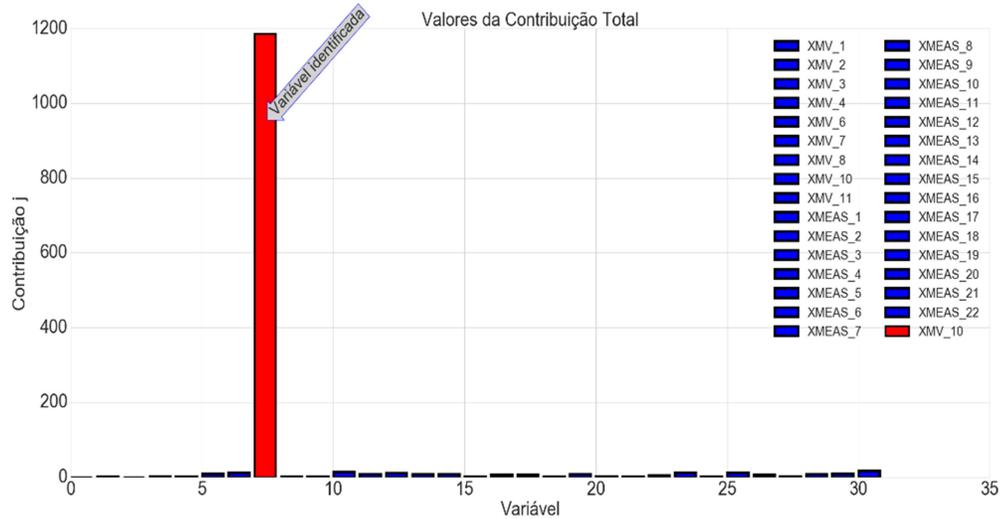


Figura 5.15: Análise de Contribuições aplicada ao estudo de caso 02 nos cenários 01 e 03

O último dos cenários a estudar para o caso de estudo 02 é o cenário 04 (Figura 5.16), constituído por uma variação aleatória da temperatura de entrada da água de resfriamento do reator (IDV\_11). Após a aplicação da metodologia do PCA, é possível capturar o comportamento anormal do sistema no tempo  $t = 8$  hr, mas não é possível determinar a variável mais relacionada com a causa do distúrbio.

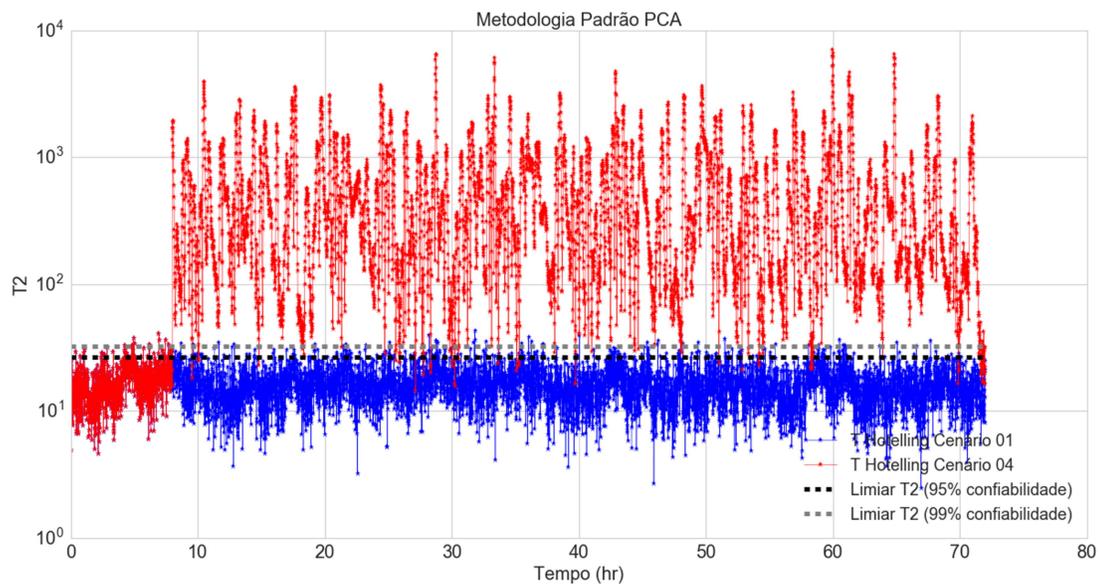


Figura 5.16: Metodologia Padrão (PCA) aplicada ao estudo de caso 02 nos cenários 01 e 04

Na Figura 5.17, pode-se observar o resultado da aplicação da análise de contribuições, para determinar a variável que esteja mais relacionada com a causa da anormalidade no sistema. Para o cenário 04, a variável mais relacionada com a anormalidade é a

temperatura do reator (XMEAS\_9) a qual é a variável esperada conhecendo o distúrbio no sistema em operação normal, pois a variável temperatura da água de resfriamento do reator, não é uma variável medida por nenhum sensor, desta maneira, não pode ser indicada como a verdadeira falha.

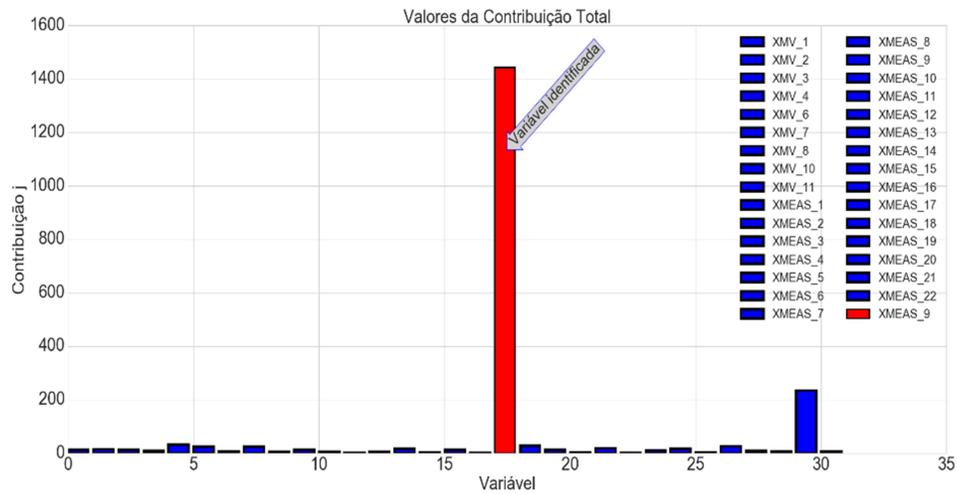


Figura 5.17: Análise de Contribuições aplicada ao estudo de caso 02 nos cenários 01 e 04

### 5.2.2 Metodologia Padrão 02: Partial Least Squares (PLS)

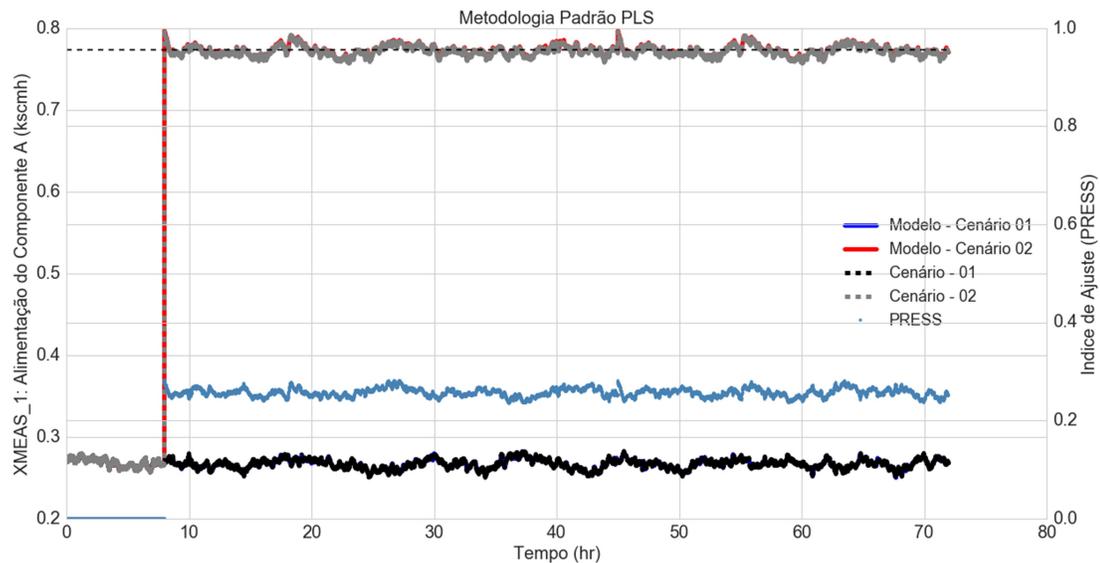


Figura 5.18: Metodologia Padrão (PLS) aplicada ao estudo de caso 02 nos cenários 01 e 02 (variável de saída XMEAS\_1)

Na Figura 5.18 são observadas as linhas de tendências para a variável de saída XMEAS\_1 tanto nos cenários 01 e 02, quanto na implementação do modelo obtido pela aplicação da metodologia PLS em ditos cenários. Pode se observar que os valores do cenário 01 e do Modelo – Cenário 01, bem como os valores do cenário 02 e do Modelo – Cenário 02 quase se superpõem, respectivamente; o que mostra que o modelo se ajusta aos dados reais com bastante precisão; validado pelo valor do coeficiente de determinação  $R^2=0,95$  (ver Tabela 5.3). No caso da implementação do Modelo – Cenário 02, é possível observar que os dados seguem uma tendência diferente do comportamento de operação normal. Sendo assim, mostra-se que alguma anormalidade está acontecendo no sistema e que o modelo é capaz de capturar dita anormalidade. Além disso, isto pode ser verificado

com o índice de ajuste implementado PRESS, o qual para os casos onde a variável não está sendo influenciada pelo distúrbio no sistema, apresenta o valor de zero (0) e quando a anormalidade afeta a variável apresenta um valor maior que zero.

No entanto, o modelo obtido a partir do PLS tem como desvantagem que está composto pela totalidade das variáveis do processo (31 variáveis), sendo isto desfavorável no momento de determinar a/s variável/variáveis que estejam mais relacionadas com a causa do problema.

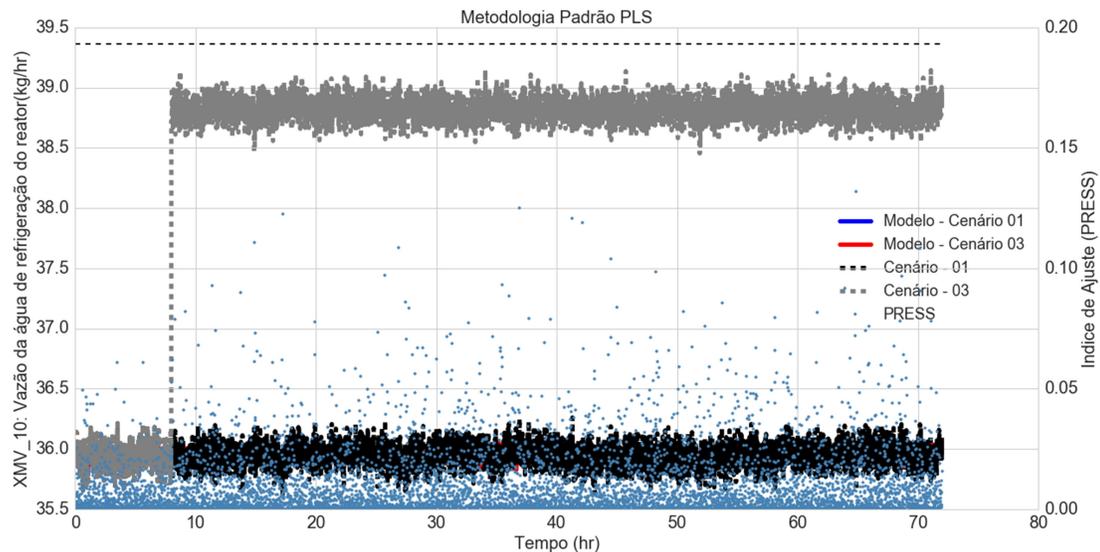


Figura 5.19: Metodologia Padrão (PLS) aplicada ao estudo de caso 02 nos cenários 01 e 03 (variável de saída XMV\_10)

Na Figura 5.19 são observadas as tendências que segue a variável de saída XMV\_10 tanto no cenário 01 e 03, quanto na implementação do modelo obtido pela aplicação da metodologia PLS. Pode-se observar que o cenário 01 e o Modelo – Cenário 01 se superpõem, sendo isto que os dados em operação normal se ajustam bastante bem ao modelo obtido pela metodologia PLS, sendo satisfatório apesar do baixo valor do coeficiente de determinação  $R^2=0,09$  (mostrado na Tabela 5.3). Contrariamente acontece com o Cenário 03 e o Modelo – Cenário 03 (que para o tema em estudo do diagnóstico de falhas é o mais importante), pois é evidente que os dados do cenário 03 não se ajustam ao modelo (a tendência associada ao Modelo – Cenário 03 fica superposta da tendência do Modelo – Cenário 01 na Figura 5.19), sendo isto que o método não funciona para dito conjunto de dados, pois que não detecta nenhuma anormalidade no sistema, tornando impossível determinar a variável mais relacionada com a falha. CHIANG, RUSSELL, BRAATZ (2000), chegaram no mesmo resultado, onde o PLS não mostrou resultados bons para o distúrbio IDV\_4, motivado a que quando acontece a falha IDV\_4 tem-se um aumento repentino da temperatura do reator, que é compensado pelas malhas de controle. As outras variáveis medidas e manipuladas permanecem estáveis depois da ocorrência da falha. A média e o desvio padrão de cada variável é menor do que o 2% entre a falha IDV\_4 e a condição em operação normal, isto faz que as tarefas de detecção e diagnóstico sejam mais desafiantes que para o caso da falha IDV\_1.

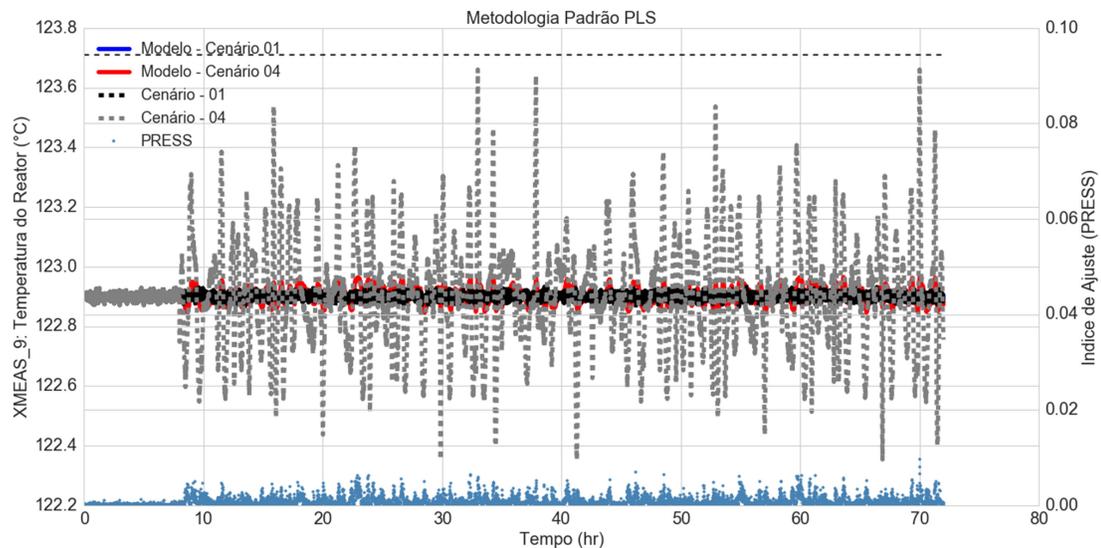


Figura 5.20: Metodologia Padrão (PLS) aplicada ao estudo de caso 02 nos cenários 01 e 04 (variável de saída XMEAS\_9)

Similar à variável de saída XMV\_10 acontece com a variável de saída XMEAS\_9, no momento da ocorrência do distúrbio IDV\_11, pois o cenário 04 e o Modelo- Cenário 04 ficam bem diferentes entre si, pelo mesmo motivo indicado no caso anterior, o qual pode ser verificado pelo baixo valor do coeficiente de determinação  $R^2 = 0,09$  (ver Tabela 4.3) e os valores do PRESS (ver Figura 5.20) depois da ocorrência do distúrbio no sistema.

Apresenta-se na Tabela 5.8 os valores dos coeficientes das variáveis que compõem o modelo obtido a partir da metodologia PLS, para o caso de estudo 02 TEP.

Tabela 5.8: Parâmetros do Modelo para as variáveis de saída XMEAS\_1, XMV\_10 e XMEAS\_9

Variáveis do Modelo	Coeficientes das Variáveis do Modelo		
	XMEAS_1	XMV_10	XMEAS_9
XMV_1	2,48E-04	0,14	-0,02
XMV_2	-1,95E-04	-0,04	-0,02
XMV_3	1,02E-02	0,02	0,00
XMV_4	1,02E-04	0,12	0,01
XMV_6	3,91E-05	-0,01	0,00
XMV_7	-1,49E-04	0,07	0,00
XMV_8	-1,05E-04	-0,25	0,02
XMV_10	-5,30E-05	-	0,01
XMV_11	1,83E-06	0,00	0,00
XMEAS_1	-	-0,26	-0,10

Variáveis do Modelo	Coeficientes das Variáveis do Modelo		
	XMEAS_1	XMV_10	XMEAS_9
XMEAS_2	-6,53E-07	0,00	0,00
XMEAS_3	4,05E-08	0,00	0,00
XMEAS_4	-1,24E-05	-0,04	0,00
XMEAS_5	-4,03E-05	-0,01	0,00
XMEAS_6	7,67E-05	0,00	0,00
XMEAS_7	7,44E-05	0,00	0,00
XMEAS_8	-9,15E-06	0,00	0,00
XMEAS_9	-1,32E-03	0,50	-
XMEAS_10	-6,06E-03	-0,09	-0,05
XMEAS_11	1,99E-04	0,30	0,05
XMEAS_12	3,04E-05	0,00	0,00
XMEAS_13	-3,24E-05	0,00	0,00
XMEAS_14	1,16E-04	0,01	0,00
XMEAS_15	-9,30E-06	0,00	0,00
XMEAS_16	-3,10E-05	0,00	0,00
XMEAS_17	1,07E-05	0,00	0,00
XMEAS_18	-4,80E-04	-0,15	-0,05
XMEAS_19	-2,05E-06	0,00	0,00
XMEAS_20	6,10E-05	0,00	0,00
XMEAS_21	7,03E-04	-0,38	0,03
XMEAS_22	-1,14E-04	0,04	0,00
R2	0,95	0,19	0,09

## 5.2.3 Metodologia Proposta: Modelos Empíricos

A Tabela 5.9 apresenta cada um dos seis (6) modelos empíricos obtidos a partir da metodologia proposta, para cada uma das saídas do estudo de caso 02 (XMEAS\_01, XMEAS\_9, XMV\_10), suas variáveis dependentes e o coeficiente de determinação  $R^2$  para cada um dos modelos.

Tabela 5.9: Modelos Empíricos para o caso de estudo 02 a partir da metodologia proposta

Variável de Saída / Coeficiente de Determinação	Modelos Empíricos					
	1	2	3	4	5	6
<b>Alimentação do componente A (corrente 1) XMEAS_01</b>	XMV_8, XMV_3 <sup>3</sup>	XMEAS_21, XMV_3 <sup>3</sup> , XMV_6 <sup>4</sup>	exp(XMV_10), exp(XMEAS_3), XMV_3 <sup>3</sup> , XMEAS_22 <sup>3</sup>	exp(XMEAS_11), XMV_2 <sup>3</sup> , XMV_3 <sup>3</sup> , XMEAS_22	XMV_10, exp(XMV_2), XMV_3 <sup>3</sup> , XMEAS_22 <sup>3</sup> , XMV_1 <sup>4</sup> , XMEAS_5 <sup>4</sup>	XMV_10, XMEAS_9, exp(XMV_2), XMEAS_21 <sup>2</sup> , XMV_3 <sup>3</sup> , XMEAS_22 <sup>3</sup> , XMEAS_19 <sup>4</sup>
<b>R<sup>2</sup></b>	0.69	0.73	0.77	0.78	0.78	0.8
<b>Temperatura do reator XMEAS_9</b>	XMV_10 <sup>3</sup> , XMEAS_18 <sup>3</sup>	XMEAS_13 <sup>2</sup> , XMV_10 <sup>3</sup> , XMV_11 <sup>3</sup>	XMEAS_13 <sup>2</sup> , XMV_10 <sup>3</sup> , XMV_11 <sup>3</sup> , XMEAS_18 <sup>3</sup>	XMEAS_13 <sup>2</sup> , XMV_10 <sup>3</sup> , XMV_11 <sup>3</sup> , XMEAS_11 <sup>3</sup> , XMEAS_18 <sup>3</sup>	XMV_10, XMEAS_13 <sup>2</sup> , XMV_10 <sup>3</sup> , XMV_11 <sup>3</sup> , XMEAS_15 <sup>3</sup> , XMEAS_18 <sup>3</sup>	exp(XMEAS_16), XMV_10 <sup>2</sup> , XMV_10 <sup>3</sup> , XMV_11 <sup>3</sup> , XMEAS_16 <sup>3</sup> , XMEAS_18 <sup>3</sup> , XMEAS_21 <sup>3</sup>
<b>R<sup>2</sup></b>	0.80	0.8	0.85	0.86	0.86	0.88
<b>Vazão de água de refrigeração do reator XMV_10</b>	XMEAS_16 <sup>3</sup> , XMEAS_21 <sup>3</sup>	XMEAS_5, XMEAS_9 <sup>3</sup> , XMEAS_16 <sup>3</sup>	XMEAS_5, XMEAS_22 <sup>2</sup> , XMEAS_16 <sup>3</sup> , XMEAS_21 <sup>3</sup>	XMEAS_6, exp(XMV_6), XMEAS_16 <sup>3</sup> , XMEAS_21 <sup>3</sup> , XMEAS_20 <sup>4</sup>	XMEAS_19, exp(XMEAS_9), XMV_2 <sup>2</sup> , XMEAS_16 <sup>2</sup> , XMEAS_21 <sup>3</sup> , XMV_2 <sup>4</sup>	exp(XMV_6), exp(XMEAS_18), exp(XMEAS_20), XMEAS_7 <sup>2</sup> , XMEAS_21 <sup>3</sup> , XMEAS_20 <sup>4</sup> , XMEAS_21 <sup>4</sup>
<b>R<sup>2</sup></b>	0.66	0.75	0.78	0.79	0.83	0.83

A Figura 5.21 mostra o melhor modelo adaptado aos cenários 01 e 02 do estudo de caso 02, sendo este o modelo 6, segundo o qual as tendências do Cenário 01 e do Modelo – Cenário 01 bem como do Cenário 02 e do Modelo – Cenário 02, são bastante precisas conforme os valores reais das variáveis tanto em operação normal quanto em operação anormal (observar valor do  $R^2$  na Tabela 5.9). O Modelo – Cenário 02 consegue capturar a anormalidade após a ocorrência do distúrbio no processo, sendo isto a partir do instante de tempo  $t = 8$  hr.

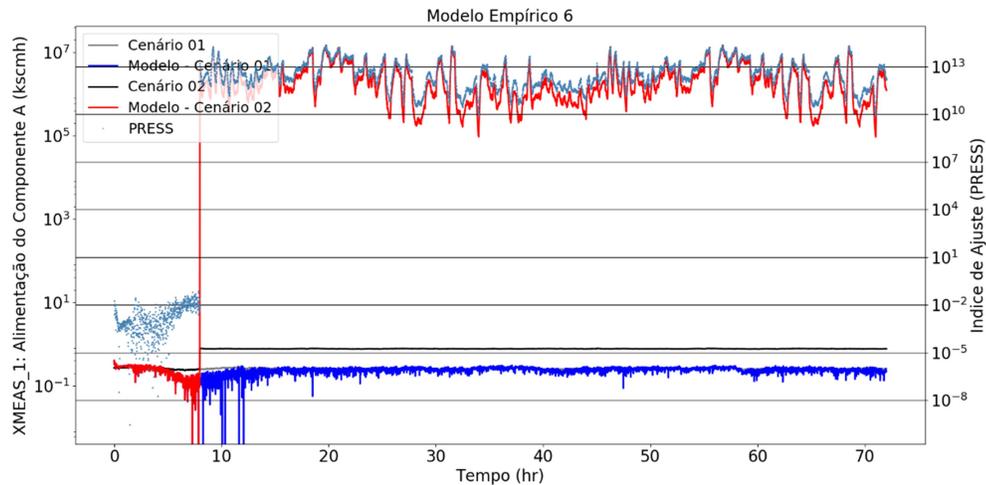


Figura 5.21: Modelo Empírico 6: XMEAS\_1 (Cenários 01 e 02)

Todos os modelos concorrentes gerados para a variável XMEAS\_01 (ver Tabela 5.9) permitem detectar a anormalidade no sistema, só que cada um deles apresenta maior precisão que os outros (como visto na Tabela 5.9). Sendo assim, é possível capturar o comportamento anormal do sistema só com duas variáveis, caso do modelo empírico 01 mostrado na Figura 5.22, o qual permite uma redução da dimensionalidade do sistema a diagnosticar de 31 para duas variáveis (fornecendo uma redução de 93,55%), uma das quais é a variável mais relacionada com a causa da falha, não obstante, a pesar de capturar o comportamento anormal do processo, ser um modelo bom e apresentar a variável que poderia estar mais relacionada à causa da falha em um sistema reduzido, o modelo empírico 6, representado na Figura 5.21, apresenta uma melhor representação do sistema (um ajuste mais preciso dos dados), além de conter a variável mais relacionada à falha. Com isto, o modelo empírico 1 pode diagnosticar com mais facilidade a variável relacionada à falha, mas o modelo empírico 6, além de capturar o comportamento anormal, pode representar melhor a realidade do processo, com fins de predição de dados, devido à expansão das bases fornecidas ao sistema que permite capturar o comportamento não linear próprio do processo, permitindo uma maior precisão do ajuste dos dados com um número reduzido de variáveis, sendo a redução da dimensionalidade de 31 para 7 variáveis (constituindo uma redução de 77,42%).

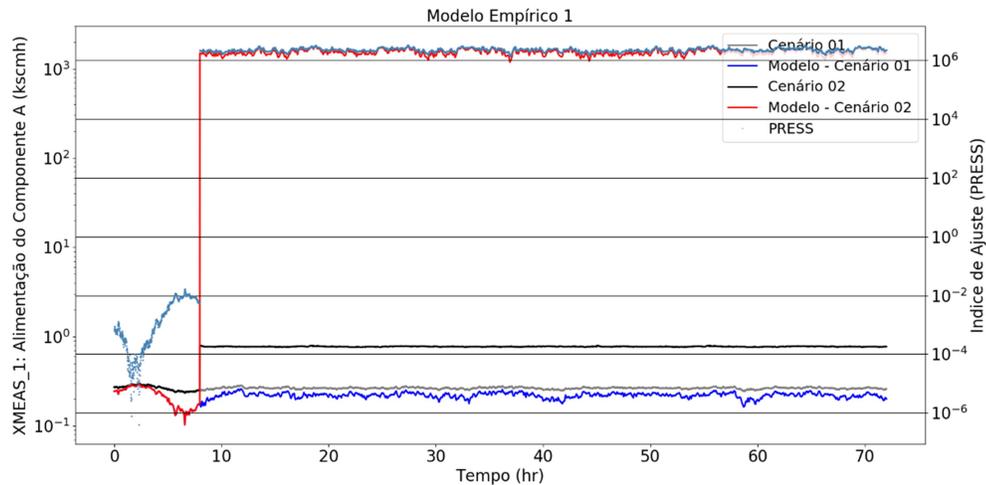


Figura 5.22: Modelo Empírico 1: XMEAS\_1 (Cenários 01 e 02)

Segundo CHIANG, RUSSELL, BRAATZ (2000) no seu estudo do sistema TEP usando diferentes técnicas de diagnóstico de falhas usando base de dados histórica do processo, quando acontece a falha IDV\_1 (Cenário 02) é induzida uma mudança de passo na relação de alimentação dos componentes A/B na corrente 4, o que resulta em um aumento na alimentação do componente B e uma diminuição na alimentação do componente A na corrente 4. Isto acarreta uma diminuição na alimentação do componente A na corrente de reciclo levando a reação da malha de controle na direção de aumentar a alimentação do componente A na corrente 1. Como consequência, esses dois efeitos se contrapõem ao longo do tempo, o que resulta em uma composição de alimentação do componente A constante na corrente 6 após tempo suficiente. As variações dos produtos e composições da corrente 6 para o reator causam variações no nível do reator, que afeta o fluxo 4 através de um *loop* de controle em cascata. O fluxo da corrente 4 eventualmente se estabelece para um valor de estado estacionário inferior ao seu valor nas condições normais de operação. Uma vez que mais da metade das variáveis monitoradas se desviam significativamente do seu comportamento operacional normal, espera-se que essa falha seja facilmente detectada.

Baseado no exposto, a falha IDV\_1 deveria ser facilmente identificável com diferentes métodos de diagnóstico, motivado a que ela influencia a maioria das variáveis do processo, pelo qual uma variação da mesma pode ser capturada através de outras variáveis de saída do sistema. Isto é validado com a metodologia proposta no presente trabalho, pois a falha IDV\_1 (cenário 02) é capturada por todos os modelos concorrentes gerados a partir da metodologia proposta para a variável de saída XMEAS\_1.

Revisando todas as variáveis contidas em cada um dos modelos gerados (ver Tabela 5.9) é possível observar que a variável XMV\_3 (uma das duas variáveis contidas no modelo empírico 1) está presente em cada um desses modelos, contrariamente da variável XMV\_8 contida no modelo empírico 1, mas que não aparece em nenhum dos outros modelos que capturaram a anormalidade do processo), o que mostra que a XMV\_3 poderia ser a variável causadora ou mais relacionada com a falha. Sendo XMV\_3 a vazão de alimentação do componente A, esta é a variável mais relacionada à causa da falha IDV\_1, pois está ligada diretamente com a mudança de passo na relação de alimentação dos componentes A/B, além de ser a mesma variável causante da falha que determinaram as metodologias padrões PCA e PLS, após a implementação da análise de contribuições.

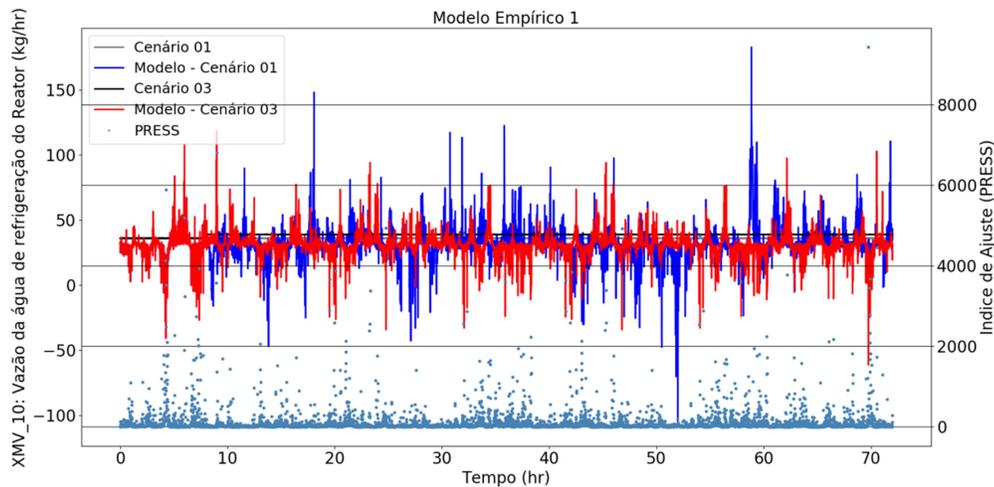


Figura 5.23: Modelo Empírico 1: XMV<sub>10</sub> (Cenários 01 e 03)

A Figura 5.23 mostra o comportamento do melhor modelo (modelo empírico 1) que representa o sistema tanto em operação normal (cenário 01) quanto em operação anormal (cenário 03), que consegue capturar a anormalidade que está acontecendo no processo. Trata-se do distúrbio IDV<sub>4</sub>, i.e. uma mudança de degrau na temperatura de entrada da água de resfriamento do reator, o qual é capturado por cada um dos modelos concorrentes (ver Tabela 5.9) gerados para a variável de saída: vazão da água de refrigeração do reator (XMV<sub>10</sub>).

O resultado do estudo feito por CHIANG, RUSSELL, BRAATZ (2000) indica que a maioria das estatísticas (PCA, PLS1, PLS2, FDA) consegue capturar adequadamente a anormalidade que está acontecendo no sistema. No entanto, para o caso do diagnóstico, muitas das estatísticas falham. O melhor dos resultados foi o caso da estatística PLS2. O PLS tradicional (melhor conhecido como PLS1) não mostrou resultados bons (baixa precisão de ajuste de dados), similar para estatísticas como PCA, FDA e CVA. Uma das causas desta limitante é que quando acontece a falha IDV<sub>4</sub> tem-se um aumento repentino da temperatura do reator, que é compensado pelas malhas de controle. As outras variáveis medidas e manipuladas permanecem estáveis depois da ocorrência da falha, a média e o desvio padrão de cada variável é menor do que o 2% entre a falha IDV<sub>4</sub> e a condição em operação normal, isto faz que as tarefas de detecção e diagnóstico sejam mais desafiadoras que para o caso da falha IDV<sub>1</sub>.

A metodologia proposta no presente trabalho atinge o objetivo, que é detectar a ocorrência da anormalidade no processo. Apesar de que o ajuste dos dados não seja tão fino (pode-se observar pelo valor do coeficiente de determinação  $R^2=0,66$ ), o modelo concorrente mostrado na Figura 5.24 consegue representar as tendências reais da variável de saída ao longo do tempo de maneira satisfatória, bem como capturar do momento em que o distúrbio entrou no processo,  $t = 800$  s.

É importante ressaltar que as variáveis presentes em cada um dos modelos concorrentes mostrados na Tabela 5.9 para a variável de saída XMV<sub>10</sub>, encontram-se em forma não linear. Isto quer dizer, que a melhor maneira de representar a relação entre as variáveis do processo para dita variável de saída através da metodologia proposta foi da forma não

linear das variáveis latentes, o qual é totalmente válido tratando-se do sistema TEP, um sistema bastante não linear.

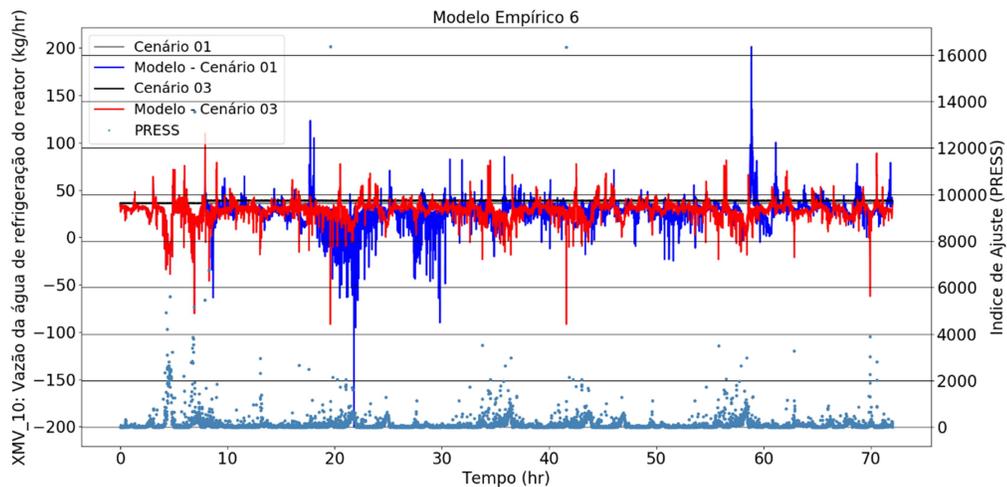


Figura 5.24: Modelo Empírico 6: XMV\_10 (Cenários 01 e 03)

Nota-se na Figura 5.23 que o melhor modelo empírico selecionado foi o modelo 1 (redução do 93,55% da dimensionalidade), pois, apesar que o modelo empírico 6 (redução do 77,42% da dimensionalidade) mostrado na Figura 5.24 apresenta um melhor valor do coeficiente de determinação (ver Tabela 5.9) que permitiria uma melhor predição dos dados, este não apresenta um ajuste dos dados tão fino quanto o modelo empírico 1.

A proposição que a XMV\_10 consiga capturar a anormalidade é válida, pois dita variável é sensível a variações de temperatura. As variáveis explicativas em cada um dos modelos gerados na Tabela 5.9 para a XMV\_10 encontram-se de forma não linear, o qual é justificado pela natureza do sistema. O modelo empírico selecionado consegue capturar o comportamento anormal do sistema só com duas variáveis permitindo uma redução da dimensionalidade de 31 para duas variáveis, constituindo uma redução de 93,55%. A partir das variáveis contidas no modelo empírico 1 é possível assumir que pelo menos uma das duas variáveis está relacionada com a causa da anormalidade, e pois o modelo captura o distúrbio a partir do  $t = 8$  hr. Assim, as variáveis suspeitas de serem causadoras da falha são XMEAS\_21 e XMEAS\_16, no entanto, a variável XMEAS\_16 não está presente no modelo empírico 6. Portanto, a variável XMEAS\_21 é assumida como a variável mais relacionada a falha do processo, e resulta com certeza uma variável causadora da falha, pois está ligada à temperatura da água de resfriamento do reator.

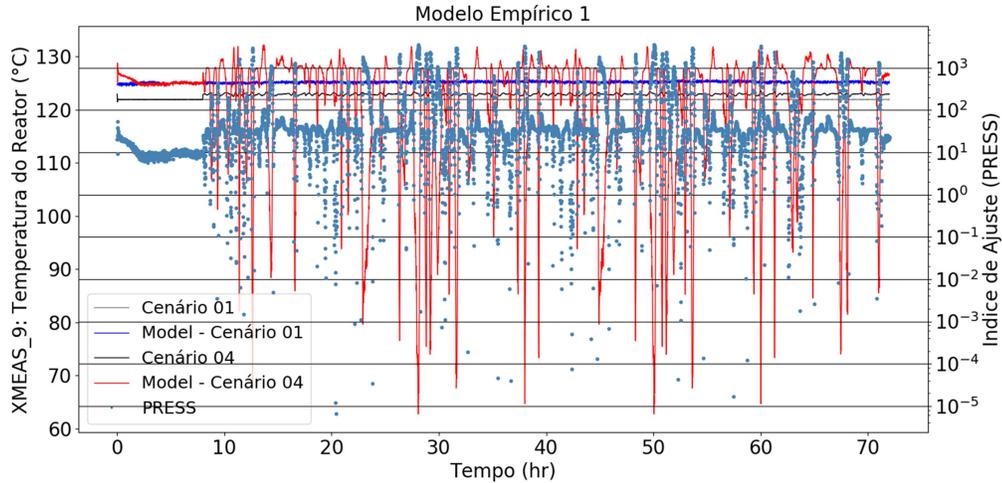


Figura 5.25: Modelo Empírico 1: XMEAS\_9 (Cenários 01 e 04)

A Figura 5.25 mostra o cenário 04 do caso de estudo 02, onde o processo é afetado pelo distúrbio IDV\_11, i.e. a variação aleatória da temperatura da água de resfriamento do reator. A variável de saída que captura de melhor maneira a anormalidade do processo é a temperatura do reator (XMEAS\_9), representada através do modelo empírico 1, o qual reduz a dimensionalidade do sistema em estudo de 31 para duas variáveis (redução do 93,55%). Apesar de que os dados não se ajustem perfeitamente à realidade do processo, devido à alta sensibilidade desta variável às mudanças aleatórias permanentes do distúrbio, o modelo concorrente gerado pela metodologia proposta, consegue capturar a tendência desta variável só com duas variáveis de saída (XMV\_10 e XMEAS\_18), uma das quais é a variável que está mais relacionada com a anormalidade do processo. É importante mencionar que as variáveis que compõem o modelo empírico 1, estão sobre a forma não linear, o qual é compreensível dada a natureza não linear do processo em estudo.

Todos os modelos concorrentes gerados pela metodologia proposta para a variável de saída XMEAS\_9 conseguem capturar a anormalidade do processo, mas nem todas as variáveis que estão presentes em cada um deles estão relacionadas com a falha do sistema produzida pelo distúrbio IDV\_11. Entretanto, através do modelo 1, representado na Figura 5.25, é possível fazer uma análise sobre qual dessas duas variáveis poderia estar mais relacionada com a falha, baseado na frequência de ocorrência nos demais modelos concorrentes gerados. Sendo assim, é possível observar a partir da Tabela 5.9 que a variável XMV\_10 é aquela que está presente em cada um dos modelos gerados que capturam a anormalidade do processo, desta maneira, é possível descartar a variável XMEAS\_18 presente no modelo 1, como uma variável que poderia estar relacionada com a falha. Além disso, pode-se corroborar o resultado obtido como satisfatório, devido a que a variável XMV\_10 é a variável mais relacionada com a falha indicada para o estatístico padrão PCA (análise de contribuições), feita no presente trabalho.

Segundo CHIANG, RUSSELL, BRAATZ (2000) a falha IDV\_11 introduz grandes oscilações na vazão da água de resfriamento do reator, o qual resulta em uma flutuação na temperatura do reator. A característica desta falha, assim como a falha IDV\_4, é que as demais variáveis permanecem ao redor do seu valor do set point seguindo comportamentos

similares as condições em operação normal do processo, o que faz mais complexo determinar a variável que esta causando a anormalidade no processo.

#### 5.2.4 Metodologia de Validação: Eliminação de Variáveis

Aplicando a metodologia de validação, baseada na técnica *Backward Elimination* nos melhores modelos obtidos segundo a metodologia proposta, é possível validar o resultado da variável considerada como mais relacionada à falha para a maioria dos cenários, conforme apresentado a seguir.

#### Cenários 01 e 02

Para o caso da variável de saída XMEAS\_1 o melhor modelo é o modelo empírico 1, composto pelas variáveis XMV\_8 e XMV\_3, fornecendo um bom ajuste dos dados de entrada e detectando a anormalidade IDV\_1 no processo. O resultado da metodologia de modelos empíricos indicou que a variável mais relacionada à falha é a XMV\_3. Aplicando a metodologia de validação, é possível corroborar este resultado, ao rejeitar a hipótese nula para o caso da eliminação da variável XMV\_3, conforme mostra a Tabela 5.10.

Tabela 5.10: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 1 da variável de saída XMEAS\_1 nos cenários 01 e 02 (i=variável contida no modelo empírico)

	i=XMV_8	i=XMV_3 <sup>3</sup>
<b>F<sub>0,i</sub></b>	2,100	2,100
<b>F-Test i</b>	1,002	601317,771
<b>F-Test i &gt; F<sub>0,i</sub></b>	-	X

#### Cenários 01 e 03

Para o caso da variável de saída XMV\_10 o melhor modelo é o modelo empírico 1, composto pelas variáveis XMEAS\_16 e XMEAS\_21, fornecendo um bom ajuste dos dados de entrada e detectando a anormalidade IDV\_4 no processo. O resultado da metodologia de modelos empíricos indicou que a variável mais relacionada à falha é a XMEAS\_21. Aplicando a metodologia de validação, é possível corroborar este resultado, ao rejeitar a hipótese nula para o caso da eliminação da variável XMEAS\_21, conforme mostra a Tabela 5.11.

Tabela 5.11: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 1 da variável de saída XMV\_10 nos cenários 01 e 03 (i=variável contida no modelo empírico)

	i=XMEAS_16 <sup>3</sup>	i=XMEAS_21 <sup>3</sup>
<b>F<sub>0,i</sub></b>	2,100	2,100
<b>F-Test i</b>	1,250	2,191
<b>F-Test i &gt; F<sub>0,i</sub></b>	-	X

Aplicando a metodologia de validação para a mesma variável de saída XMV\_10, no caso do modelo empírico 6, não é possível obter um resultado concluinte sobre a variável mais relacionada à falha no processo, como é possível observar na Tabela 5.12, pois para nenhuma das variáveis eliminadas do modelo, é possível rejeitar a hipótese nula.

Tabela 5.12: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 da variável de saída XMV\_10 nos cenários 01 e 03 (i=variável contida no modelo empírico)

	$i=e^{(XMV_6)}$	$i=e^{(XMEAS_{18})}$	$i=e^{(XMEAS_{20})}$	$i=XMEAS_{7^3}$	$i=XMEAS_{21^3}$	$i=XMEAS_{20^4}$	$i=XMEAS_{21^4}$
<b>F<sub>0,i</sub></b>	2,101	2,101	2,101	2,101	2,101	2,101	2,101
<b>F-Test i</b>	1,807	1,171	0,655	1,324	1,045	1,492	0,965
<b>F-Test i &gt; F<sub>0,i</sub></b>	-	-	-	-	-	-	-

### Cenários 01 e 04

Para o caso da variável de saída XMEAS\_9, o melhor modelo foi o modelo empírico 1, composto pelas variáveis XMV\_10 e XMEAS\_18, fornecendo um bom ajuste dos dados de entrada e detectando a anormalidade IDV\_11 no processo. O resultado da metodologia de modelos empíricos indicou que a variável mais relacionada à falha é a XMV\_10. Aplicando a metodologia de validação, é possível corroborar este resultado, ao rejeitar a hipótese nula para o caso da eliminação da variável XMV\_10, conforme mostra a Tabela 5.13.

Tabela 5.13: Valores da distribuição de Fisher e F-Test para o Modelo Empírico 6 da variável de saída XMV\_10 nos cenários 01 e 03 (i=variável contida no modelo empírico)

	$i=XMV_{10}^3$	$i=XMEAS_{18}^3$
<b>F<sub>0,i</sub></b>	2,099	2,099
<b>F-Test i</b>	2,807	1,1716
<b>F-Test i &gt; F<sub>0,i</sub></b>	X	-

# Capítulo 6 – Considerações e Trabalhos Futuros

## 6.1 Principais conclusões

A metodologia proposta fornece uma considerável redução da dimensionalidade dos dados a serem usados para o diagnóstico de falhas. No estudo de caso 01, no cenário 01, para a  $F_s$  a redução da dimensionalidade foi de 11 para 2 variáveis (uma redução de 81,82%), e para o T4 foi de 11 para 7 variáveis (redução de 36,36%); no cenário 02, para a  $F_s$  a redução da dimensionalidade foi de 11 para 7 variáveis (redução de 36,36%), e para o T4 de 11 para 8 variáveis (27,27% de redução). No caso de estudo 02, para o modelo 1 da variável de saída XMEAS\_1 a redução da dimensionalidade foi de 31 para duas variáveis (93,55% de redução), e para o modelo 6, a redução foi de 31 para 7 variáveis (redução de 77,42%), para a variável de saída XMV\_10 a redução da dimensionalidade foi de 31 para duas variáveis (redução de 93,55%), bem como para a variável de saída XMEAS\_9, onde a redução da dimensionalidade do sistema em estudo, foi de 31 para duas variáveis (93,55% de redução). Isto constitui uma otimização no diagnóstico e uma alta efetividade do método proposto, contrariamente as metodologias padrões PCA e PLS que trabalham em todo momento com a totalidade das variáveis da base de dados histórica do processo.

A metodologia do PCA não aponta para a variável mais relacionada à causa da falha, e além de considerar erroneamente as mudanças de set point como anormalidades do sistema. Assim, é necessária a aplicação de técnicas adicionais para melhorar o diagnóstico e determinar qual é a variável mais relacionada à falha como é o caso da análise de contribuições.

A metodologia PLS trabalha em todo momento com a totalidade das variáveis da base de dados histórico do processo, e depende do intervalo de variabilidade das variáveis na hora de criar o modelo para ter um bom ajuste caso a base de dados com anormalidade esteja muito afastada (maior que o 50%) da base de dados em operação normal. Além disso, não mostrou resultados adequados para todos os estudos de casos. Por exemplo, no estudo de caso 02, para a variável de saída XMV\_10, não conseguiu capturar a anormalidade do processo.

Contrariamente, a metodologia proposta faz uma correta detecção da falha, permite determinar a variável mais relacionada ao distúrbio permitindo fazer um diagnóstico de falhas mais realista e finalmente com a expansão das bases permite trabalhar com sistemas

não lineares. Métodos propostos adicionais como a metodologia de validação baseada no *Backward Elimination*, permite verificar que os resultados da metodologia de modelos empíricos foram corretos.

É importante indicar que tanto as metodologias propostas quanto as padrões têm limitações, pois estão sendo estudados em sistemas dinâmicos aplicando métodos estáticos, no entanto, constitui uma base bastante interessante para trabalhos futuros aplicando metodologias dinâmicas ou em sistemas em tempo real.

## 6.2 Sugestões para trabalhos futuros

Com base no trabalho desenvolvido, recomendam-se os seguintes temas como trabalhos futuros para aprofundar e complementar os trabalhos do diagnóstico de falhas em processos industriais baseado em modelos empíricos:

- Criar uma metodologia onde sejam gerados conjuntos de modelos para uma maior quantidade de variáveis de saída a partir da adaptação da ferramenta ACO, selecionar os melhores de cada conjunto, e avaliar a detecção de falhas a partir das diferentes variáveis de saída, através de métricas de ajuste de dados, e, em função delas selecionar as variáveis presentes em cada um dos modelos que poderiam estar relacionadas com a falha.
- Comparar os resultados obtidos com outras metodologias padrões como FDA, CVA, PCA Dinâmico.
- Modificar a metodologia proposta para trabalhar com sistemas em tempo real e não com base de dados histórica de processos.
- Fazer a comparação ou trabalho conjunto da metodologia de modelos empíricos com a metodologia desenvolvida por CHENG (2013), uma vez que essa elimina as desvantagens do método PCA tradicional, trabalhando com uma base de dados histórica modificada, o qual poderia constituir um bom tema de estudo no diagnóstico de falhas.

## Referências

BATHELT, A. Revision of the Tennessee Eastman Process Model. IFAC-PapersOnLine, v. 48, n. 8, p. 309-314, 2015.

BISHOP, C. Pattern Recognition and Machine Learning. USA: Springer, 2006.

BABYAK, M. What you see may not be what you get: a brief nontechnical introduction to overfitting in regression-type models. Psychosom Med. 2004

CHENG, Y. Data-driven Techniques on Alarm System Analysis and Improvement. 2013. (PhD). Department of Electrical & Computer Engineering, University of Alberta, Canada.

CHIANG, L.; RUSSELL, E.; BRAATZ, R. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. Chemometrics and Intelligent Laboratory Systems, v. 50, n. 2, p. 243-252, 2000.

CHOI, S. et al. Fault detection and identification of nonlinear processes based on kernel PCA. Chemometrics and Intelligent Laboratory Systems, v. 75, n. 1, p. 55-67, 2005.

DING, S. X. Data-driven Design of Fault Diagnosis and Fault-Tolerant Control Systems. Springer-Verlag London, 2014. 300.

DOWNS, J. V., E. A plant-wide industrial process control problem. Computer & Chemical Engineering, v. 17, n. 3, p. 245-255, 1993.

FERNANDEZ et al. A backward variable selection method for PLS regression ( BVSPLS). Anal Chim Acta. 2009.

FRANK, P. M. Fault Diagnosis and Fault-Tolerant Control. Encyclopedia of Life Support Systems. XVI 2009.

GAVIRIA, C. Regresión por Mínimos Cuadrados Parciales PLS Aplicada a Datos Variedad Valuados. 2015. (Magister). Universidad de Colombia, Colombia.

GUYON, I.; ELISSEEFF, A. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, v. 3, p. 1157-1182, 2003.

HAWKINS, D. The Problem of Overfitting. Journal of chemical information and computer sciences, v. 44, 2004.

HARRELL et al. Regression models for prognostic prediction: advantages problems and suggested solutions. Cancer Treat Rep. 1985.

ISERMANN, R. Supervision, fault-detection and fault-diagnosis methods — An introduction. Control Engineering Practice, v. 5, n. 5, p. 639-652, 1997/05/01/ 1997. ISSN 0967-0661. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S09670666197000464> >.

ISERMANN, R. Fault-Diagnosis Systems. 1. Springer-Verlag Berlin Heidelberg, 2006. XVIII, 475.

KOH et al. An interior-point-method for large scale  $l_1$ -regularized logistic regression. Journal of Machine Learning Research. 2007

KOURTI, T.; NOMIKOS, P.; MACGREGOR, J. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. Journal of Process Control, v. 5, n. 4, p. 277-284, 1995.

LAMEDA, C. Métodos Relacionados con Diagnóstico de Fallas con Síntomas Imprecisos Mediante Comparación de Casos. Revista Digital de Investigación y Postgrado de la Universidad Nacional Experimental Politécnica “Antonio José de Sucre”. Venezuela. 5: 851-865 p. 2015.

LOMAX, R.; HAHS-VAUGHN; DEBBIE. Statistical Concepts: A Second Course. 2012.

NIMMO, I. Abnormal Situation Management. Process and Control Engineering, p. 36-45, 1995.

MACGREGOR et al. Multivariate Statistical Process Control of Batch Processes Using PCA. IFAC.1994

MACGREGOR, KOURTI, NOMIKOS. Analysis, monitoring and fault diagnosis of industrial process using multivariate statistical projection methods. IFAC.1996.

MACGREGOR, NOMIKOS. Monitoring batch processes using multiway principal component analysis. *Process Systems Engineering*. 1994.

OKOH, P.; HAUGEN, S. The Influence of Maintenance on Some Selected Major Accidents. *Chemical Engineering Transactions*, v. 31, p. 493-498, 2007.

ORR, M. Regularization in the selection of radial basis function centers. *Neural Computation*, v. 7, n. 3, p. 606-623, 1995.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2011.

RANZAN, C. et al. Wheat flour characterization using NIR and spectral filter based on Ant Colony Optimization. *Chemometrics and Intelligent Laboratory Systems*, v. 132, p. 133-140, 2014.

RUSSELL, E. L. C., LEO H. BRAATZ, RICHARD D. Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes. Springer-Verlag London, 2000. XIII, 192.

VENKATASUBRAMANIAN, V.; RENGASWAMY, R.; KAVURI, S. N. A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Computers & Chemical Engineering*, v. 27, n. 3, p. 313-326, 2003/03/15/ 2003. ISSN 0098-1354. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0098135402001618> >.

VENKATASUBRAMANIAN, V. et al. A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering*, v. 27, n. 3, p. 327-346, 2003/03/15/ 2003. ISSN 0098-1354. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S009813540200162X> >.

VENKATASUBRAMANIAN, V.; RENGASWAMY, R.; KAVURI, S. N. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, v. 27, n. 3, p. 293-311, 2003/03/15/ 2003. ISSN 0098-1354. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0098135402001606> >.

WESTLAND , C. Structural Equation Models. Springer International Publishing, 2015. 175.

XU, L.; ZHANG, W.-J. Comparison of different methods for variable selection. *Analytica Chimica Acta*, p. 477-483, 2001.

YIN, S.; ZHU, X.; KAYNAK, O. Improved PLS Focused on Key-Performance-Indicator-Related Fault Diagnosis. *IEEE Transactions on Industrial Electronics*, v. 62, n. 3, p. 1651 - 1658, 2015.

YOON, S.; MACGREGOR, J. Fault diagnosis with multivariate statistical models part I: using steady state fault signatures. *Journal of Process Control*, v. 11, n. 4, p. 387-400, 2001.

# Apêndice A

## Apêndice A.1 Modelo Matemático do Caso de Estudo 01: Tanque de Aquecimento com Mistura

O estudo de caso 01: Tanque de Aquecimento com Mistura mostrado na Figura A.1 foi apresentado na seção 4.3, no entanto, nessa seção não foram especificados os modelos matemáticos que representam o processo nem a implementação no simulink, portanto, serão mostrados a seguir.

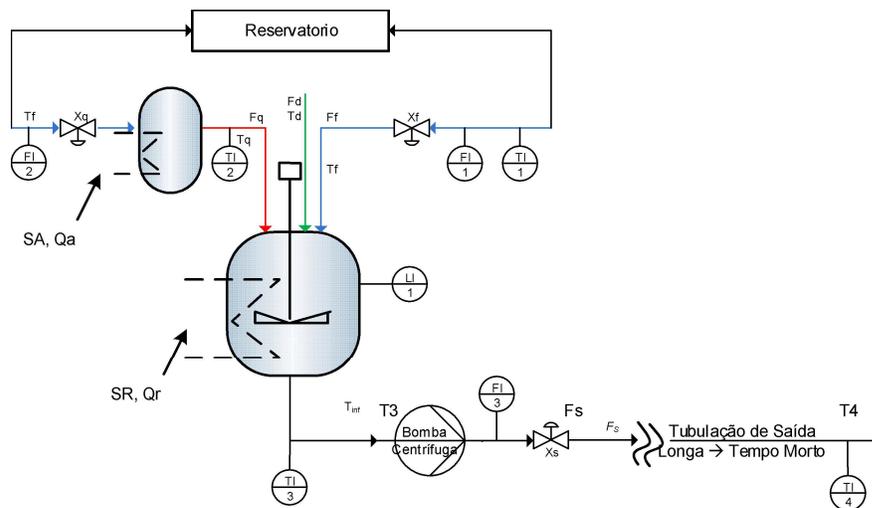


Figura A.1: Estudo de Caso 01 Tanque de Aquecimento com Mistura

O sistema trabalha com uma vazão nominal de 5 gpm com água (considerando  $G_f = 1$ ). Para essa vazão a perda de carga na tubulação de saída é de 5 psi. Tanto o tanque quanto a saída da tubulação estão à mesma pressão. A alimentação do tanque está sujeita a uma pressão estática de 20 psi devido a coluna da água e a perda de carga para vazão nominal de 5 gpm da caixa da água até o ponto de bifurcação entre a corrente quente e fria é de 1 psi. A perda de carga no aquecedor de passagem para uma vazão nominal de 3 gpm é de 2 psi. A unidade deve ser capaz de trabalhar com uma vazão máxima de 10 gpm. A bomba centrífuga disponível para operar a planta tem  $\Delta P_{B,max} = 30 \text{ psi}$  e  $F_{MAX} = 20 \text{ gpm}$ .

Conhecendo que a equação básica que descreve o funcionamento de uma válvula de controle é dada por (A.1)

$$F = f(x)Cv \sqrt{\frac{\Delta P_v}{Gf}} \quad (\text{A.1})$$

Onde:  $F$  : é a vazão em gpm,  $f(x)$  : função que relaciona a abertura da válvula com o tipo de obturador (p.ex., linear, parabólico, igual percentagem, etc.),  $Cv$  : coeficiente de vazão (tamanho característico da válvula) [gpm/psi<sup>0,5</sup>],  $\Delta P_v$  : perda de carga na válvula [psi] e  $Gf$  : que é a razão entre a massa específica do líquido e a massa específica da água @ 60 °F.

Levando em conta que a perda de carga para a linha correspondente a corrente fria, a corrente quente e para a tubulação de saída são dadas pelas equações A.1, A.2 e A.3.

$$\Delta P v_f = 20 - \left( \frac{F_q + F_f}{5} \right)^2 \quad (\text{A.2})$$

$$\Delta P v_q = 20 - \left( \frac{F_q + F_f}{5} \right)^2 - 2 \left( \frac{F_q}{3} \right)^2 \quad (\text{A.3})$$

$$\Delta P v_s = 30 \sqrt{1 - \left( \frac{F_s}{20} \right)^2} - 5 \left( \frac{F_s}{5} \right)^2 \quad (\text{A.3})$$

Com a finalidade de atender as demandas requeridas para as vazões máximas, foram dimensionadas válvulas com  $Cv_f = 2,5$  e obturador linear ( $f(x_f) = x_f$ ) para a corrente fria,  $Cv_q = 5$  e obturador parabólico ( $f(x_q) = x_q^{1,6}$ ) para a válvula instalada na corrente quente e  $Cv_s = 10$  e obturador parabólico ( $f(x_s) = x_s^3$ ) para a válvula instalada na saída do tanque. As equações resultantes para as vazões de alimentação da corrente fria ( $F_f$ ), da corrente quente ( $F_q$ ) e para a saída do tanque ( $F_s$ ) são dadas pelas equações A.4, A.5 e A.6.

$$F_f = 2,5x_f \sqrt{20 - \left( \frac{F_q + F_f}{5} \right)^2} \quad (\text{A.4})$$

$$F_q = 5x_q^{1,6} \sqrt{20 - \left( \frac{F_q + F_f}{5} \right)^2 - 2 \left( \frac{F_q}{3} \right)^2} \quad (\text{A.5})$$

$$F_s = \frac{5x_s^3 \sqrt{30} \sqrt{-15x_s^6 + \sqrt{6625x_s^{12} + 640x_s^6 + 16}}}{20x_s^6 + 1} \quad (\text{A.6})$$

A modelagem fenomenológica da planta de aquecimento e mistura foi feita através do balanço de massa e de energia do sistema, considerando o tanque em formato cilíndrico, mistura perfeita, massa específica e capacidades térmicas constantes e iguais. Além disso, por tratar-se de um tanque fechado, foi considerada a taxa de evaporação do líquido que causa um aumento na pressão permitindo o aumento da temperatura do tanque.

$$A \frac{dh}{dt} = F_f + F_q + F_d - F_s - \frac{W_g}{\rho} \quad (\text{A.7})$$

$$h \frac{dT_s}{dt} + T_s \frac{dh}{dt} = \frac{F_f T_f}{A} + \frac{F_q T_q}{A} + \frac{F_d T_d}{A} - \frac{F_s T_s}{A} - \frac{W_g \lambda_v}{\rho c_{p,a} A} + \frac{\dot{Q}_R}{\rho c_{p,a} A} \quad (\text{A.8})$$

Onde  $F_q, F_f, F_s$  e  $F_d$  são as vazões das correntes quente, fria, de saída e do distúrbio e  $T_q, T_f, T_s$  e  $T_d$  são as temperaturas dessas correntes, respectivamente.  $h$  é a altura de nível de líquido do tanque,  $A$  é a área de seção transversal,  $V_a$  é o volume do aquecedor a gás,  $\dot{Q}_a$  é a potência de aquecimento da resistência elétrica do tanque menor (ver Figura A.1),  $\dot{Q}_R$  é a potência de aquecimento da resistência elétrica do tanque de aquecimento com mistura,  $\rho$  e  $c_{p,a}$  são a massa específica e a capacidade calorífica da água, respectivamente,  $W_g$  é a taxa mássica de vaporização do líquido e  $\lambda_v$  é o calor latente de vaporização. A taxa mássica de vaporização do líquido é dada pela Equação A.9.

$$W_g = K_G (P - P_v) \quad (\text{A.9})$$

Onde  $K_G$  é o coeficiente global de transferência de massa entre fases,  $P$  é a pressão de vapor do líquido, dada pela equação de Antoine (dependente apenas da temperatura em que se encontra), e  $P_v$  é a pressão que o vapor exerce no sistema. A determinação da pressão que o vapor exerce no sistema é obtida através de um balanço de mols para a fase vapor, considerando o comportamento de gás ideal:

$$\frac{dn}{dt} = \frac{W_g}{\bar{M}} \quad (\text{A.10})$$

$$P_v = \frac{nR_s}{A(h_{m\acute{a}x}-h)} \quad (\text{A.11})$$

Onde  $n$  é o número de mols de vapor,  $\bar{M}$  é a massa molar da água,  $R$  é a constante universal dos gases e  $h_{m\acute{a}x}$  é a altura total do tanque cilíndrico.

Para a tubulação de saída, a equação de balanço trata-se de uma perda de calor devido à troca com o meio ambiente e um tempo morto devido ao comprimento.

$$\frac{dT_{4a}}{dt} = \frac{F_s}{V_T} (T_s - T_{4a}) - \frac{\dot{Q}_T}{\rho c_{p,a} V_T} \quad (\text{A.12})$$

$$\dot{Q}_T = \frac{UA_\infty(T_s - T_{4a})}{\ln\left(\frac{T_s - T_{inf}}{T_{4a} - T_{inf}}\right)} \cong UA_\infty \left( (T_s - T_{inf})(T_{4a} - T_{inf}) \left( \frac{1}{2}(T_s - T_{inf}) + \frac{1}{2}(T_{4a} - T_{inf}) \right) \right)^{\frac{1}{3}} \quad (\text{A.13})$$

$$\frac{\Delta T_4(s)}{\Delta T_{4a}(s)} = \exp\left(-\frac{V_T}{F_s} s\right) \quad (\text{A.14})$$

Onde  $UA_\infty$  é o coeficiente global de transferência de calor da tubulação para o ambiente e  $V_T$  o volume total da tubulação.

O modelo desenvolvido no GIMSCOP para a corrente Fq, originalmente tinha um aquecedor a gás, mas isto foi modificado colocando um tanque de menor tamanho, com uma resistência elétrica ativada por um controlador on-off, o balanço de energia associado a dita modificação é o mostrado na equação A.15. Sendo K a constante de condutividade térmica e tendo um valor de 0.25 W e V o volume do tanque de aquecimento, tendo um valor de 0.2 m<sup>3</sup>.

$$\frac{dTq}{dt} = \frac{F}{V} *(Ttf - Tq) + \frac{Q*SA}{Cp*p*V} - \frac{K*(Tq-Tinf)}{Cp*p*V} \quad (\text{A.15})$$

Conforme observado pelo balanço de massa, Equação A.7, a variação do nível do tanque depende apenas das vazões de alimentação e saída, ou seja, independe do nível de líquido presente no tanque, caracterizando um sistema integrador. Sistemas integradores são característicos por possuírem a função de transferência com um pólo na origem ( $s = 0$ ) e o comportamento frente a uma resposta degrau ser linearmente crescente ou decrescente, sendo necessário operar em malha fechada para estabilizá-lo.

No balanço de energia, Equação A.8, a variação da temperatura  $T_s$  é influenciada pelas correntes de alimentação quente e fria e pela resistência elétrica de aquecimento. A temperatura  $T_4$ , que é a variável que entra na serpentina a fim de aquecer o reator, possui a perda de calor devido ao tempo morto e a temperatura externa inferior a temperatura do fluido. As equações que representam as vazões de alimentação da corrente quente e fria possuem como variável dependente além da abertura das válvulas quente e fria, a influência de uma corrente na outra, dificultando o cálculo analítico dessas vazões.

O processo foi implementado no simulink representado nas Figuras A.2, A.3, A.4, A.5 e A.6.

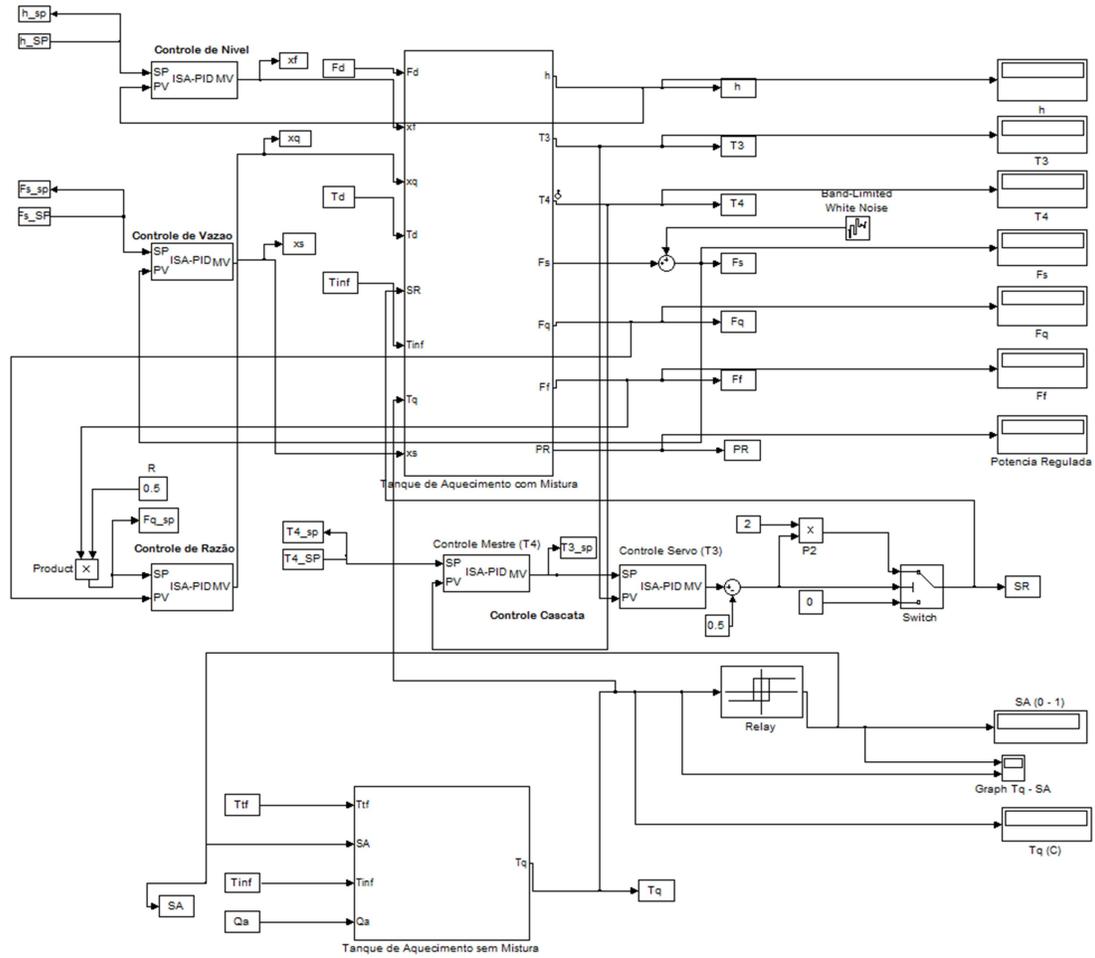


Figura A.2: Implementação no Simulink do Caso de Estudo 01 Tanque de Aquecimento com Mistura (Adaptado do sistema feito no GIMSCOP)

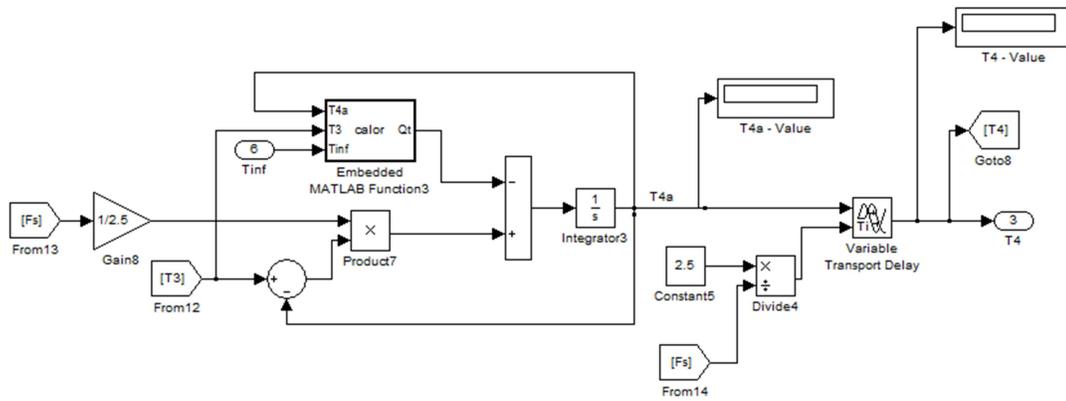


Figura A.3 Implantação do balanço de energia para a saída do Tanque de Aquecimento com Mistura em Simulink (Adaptado do Sistema feito no GIMSCOP)

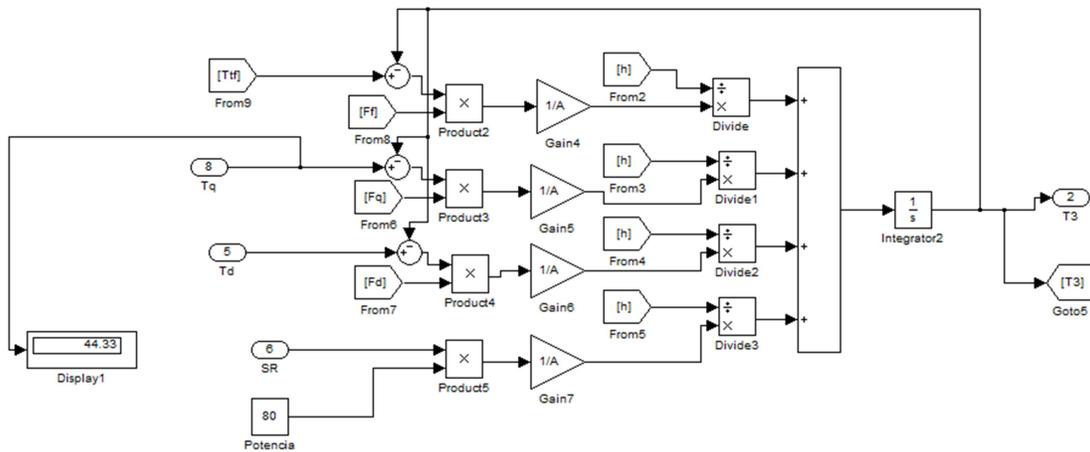


Figura A.4 Implementação no Simulink do Balanço de Energia para as correntes associadas á entrada do Tanque de Aquecimento com Mistura (Adaptado do Sistema feito no GIMSCOP)

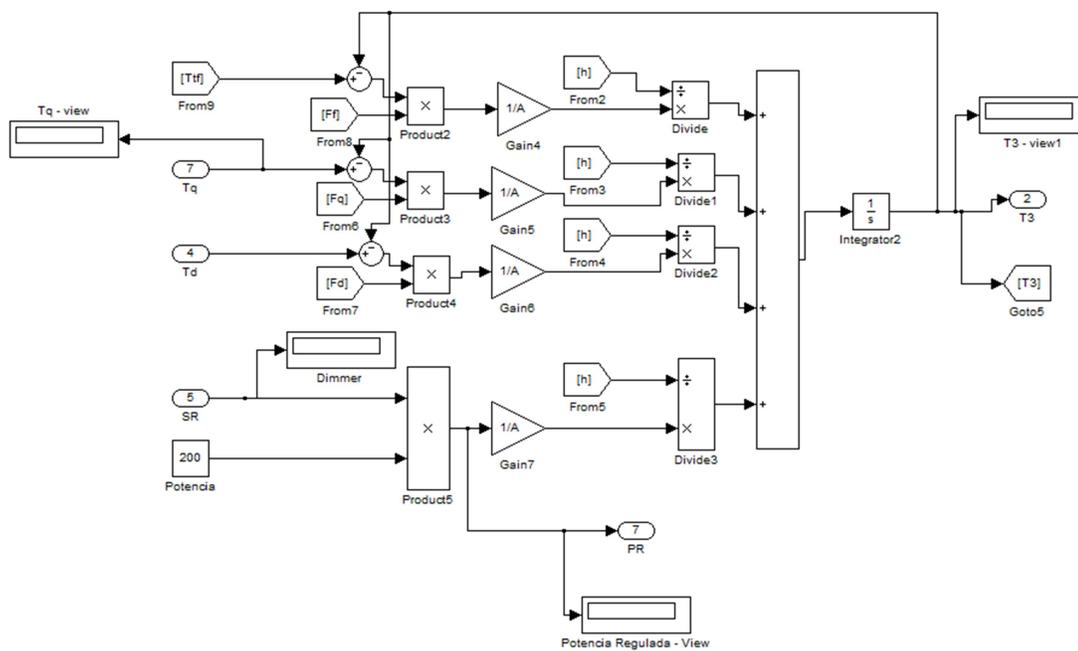


Figura A.5 Implementação no Simulink do Balanço de Energia para o Tanque de Aquecimento com Mistura (Adaptado do Sistema feito no GIMSCOP)

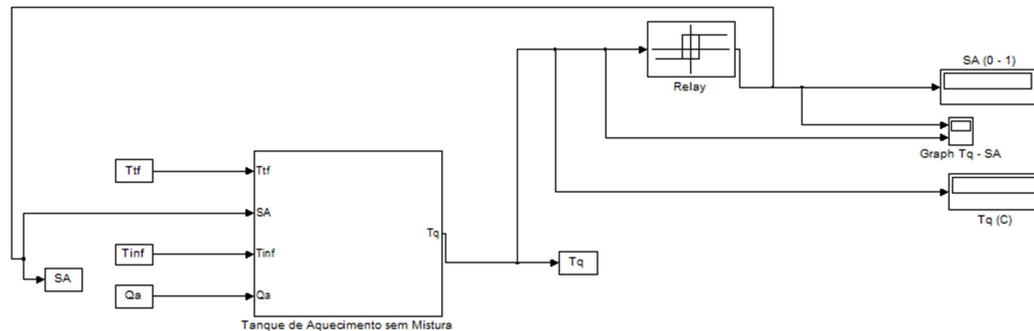


Figura A.6 Implementação no Simulink do Controlador on/off para o sistema de aquecimento da corrente Fq no Tanque de Aquecimento de menor tamanho.