

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Programa de Pós-Graduação em Engenharia Elétrica

André Emanuel Lebedev Martinez Moreira

# **Monitoramento de Condição de uma Unidade Geradora Hidrelétrica**

Belo Horizonte

2016

André Emanuel Lebedev Martinez Moreira

## **Monitoramento de Condição de uma Unidade Geradora Hidrelétrica**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica.

Orientadora: Prof.<sup>a</sup> Carmela Maria Polito Braga, Dra.

Belo Horizonte

2016

*Aos meus pais e à minha irmã.*

# Agradecimentos

Agradeço aos meus pais, Maria Natália e Wanderval, à minha irmã, Nádia, e aos meus amigos por todo o apoio, carinho e compreensão.

À minha orientadora, Professora Carmela Maria Polito Braga, pelo conhecimento passado, pelas sugestões e pela participação constante no desenvolvimento deste trabalho.

Ao professor Hugo Michel, aos colegas e amigos Bruno Monteiro e Joyce Ferreira e a todos os professores e colegas do Grupo de Validação de Sistemas, pelos ensinamentos, pela troca de experiências e pelas conversas.

À toda a equipe do Projeto P&D GT338, pela oportunidade e pela participação no desenvolvimento do trabalho.

*“A dúvida é o princípio da sabedoria.”  
(Aristóteles)*

# Resumo

As usinas hidrelétricas são responsáveis pela maior parte da geração de energia elétrica na matriz energética brasileira. Monitorar e atestar o bom funcionamento das unidades turbina-gerador é essencial, especialmente em períodos de seca, quando as condições de operação podem ser mais atípicas. Nessa perspectiva, apresenta-se o desenvolvimento e aplicação de técnicas visando o monitoramento amplo de condição de uma unidade geradora, utilizando ferramentas de controle estatístico de processos multivariados, como a carta de Hotelling  $T^2$ , o índice SPE, o índice combinado  $\varphi$ , além de PCA e *contribution plots*. Essa unidade faz parte de uma usina hidrelétrica que possui modos particulares de operação, que incluem variações constantes de *setpoint*, gerando condições de não-estacionariedade. Portanto, propõe-se uma metodologia de análise e monitoramento de medições pertinentes utilizando transformações de variáveis, um estimador recursivo de variância para identificação do modo de operação, filtros de média na etapa de projeto e um procedimento de adaptação de médias para o monitoramento em tempo certo. Os algoritmos são desenvolvidos em ambiente Matlab, testados em dados históricos da usina, com a análise dos resultados do PCA visando à identificação das variáveis que, provavelmente, causam os alarmes gerados pelas cartas de controle. A metodologia é desenvolvida visando a aplicação posterior em um sistema PIMS instalado na usina, para integrar soluções de monitoramento da planta. Os resultados apresentados mostram que a metodologia proposta permite o monitoramento do processo com diferentes pontos e modos de operação considerados normais, ainda que parte deles não tenham sido observados na análise retrospectiva dos dados.

**Palavras-chave:** controle estatístico de processos. análise de componentes principais. carta de Hotelling  $T^2$ . monitoramento de condição. unidade geradora hidrelétrica.

# Abstract

Hydroelectric power plants are responsible for most of the energy generation of Brazil's electrical grid. Monitoring and attesting the proper functioning of the turbine-generator units is essential, especially in periods of drought, when the operation conditions might be atypical. In that perspective, a development and application of techniques is presented aiming a broad condition monitoring of a generator unit, using multivariate statistical process control tools, such as Hotelling's  $T^2$  chart, SPE and combined  $\varphi$  indexes, besides PCA and contribution plots. This unit is part of a hydroelectric power plant that has particular modes of operation, which include constant variations of setpoint, causing non-stationarity conditions. Therefore, a methodology is proposed for the analysis and monitoring of pertinent measurements using variable transformations, recursive variance estimation for identifying operational modes, mean filters at the desing step and a procedure for mean adaptation for just-in-time monitoring. The algorithms are developed in Matlab, tested with historical data from the plant, along with PCA analysis for identifying the variables that most likely caused the alarms generate by the control charts. The methodology is developed aiming further implementation on a PIMS system, installed in the plant's site, to integrate monitoring solutions for the processes. The results show that the methodology allows monitoring a process with different operational points and modes, even if some of them have not been observed in the retrospective data analysis.

**Key-words:** statistical process. principal component analysis. Hotelling's  $T^2$  chart. condition monitoring. hydroelectric generator unit.

# Lista de ilustrações

Figura 1 – Usina Hidrelétrica de Emborcação . . . . .	14
Figura 2 – Análise Externa . . . . .	18
Figura 3 – Distribuição de Probabilidade Normal Bivariada . . . . .	19
Figura 4 – Carta $T^2$ de Hotelling . . . . .	21
Figura 5 – Exemplo de PCA: (a) Variável $\mathbf{x}_1$ ( $\bar{x}_1 = 3$ e $\sigma_1 = 6$ ). (b) Variável $\mathbf{x}_2$ ( $\bar{x}_2 = -1$ e $\sigma_2 = 3$ ). (c) Gráfico de $\mathbf{x}_1^*$ vs $\mathbf{x}_2^*$ e eixos de maior variabilidade calculados pelo PCA. (d) Gráfico de $\mathbf{x}_1$ vs $\mathbf{x}_2$ e eixos de maior variabilidade calculados pelo PCA . . . . .	27
Figura 6 – Exemplo de representação gráfica de <i>contribution plots</i> . . . . .	30
Figura 7 – Potência Ativa do Gerador em operação mista: CAG e local. (a) Nove dias de operação. (b) Oito horas de operação com CAG . . . . .	33
Figura 8 – Arquitetura de automação da UHE Emborcação . . . . .	35
Figura 9 – <i>CompactRIO cRIO-9082</i> da <i>National Instruments</i> . . . . .	35
Figura 10 – Distribuição dos Sensores do Sistema Online de Monitoramento de Vibração no gerador . . . . .	36
Figura 11 – Acelerômetro do Estator . . . . .	40
Figura 12 – Potência ativa em trecho de setembro de 2015. . . . .	44
Figura 13 – Potência ativa em trecho de setembro de 2015, de operação via CAG. . . . .	44
Figura 14 – Comparação entre o resultado do monitoramento com média fixa na fase de projeto (b) e média móvel na fase de projeto (c). Velocidade da vibração no Estator (a). . . . .	45
Figura 15 – Histogramas do resíduo da velocidade da vibração na amostra original . . . . .	47
Figura 16 – Histogramas do resíduo do logaritmo natural da velocidade da vibração . . . . .	47
Figura 17 – Temperatura da Sapata nº 16 do Mancal de Escora do Gerador 3 . . . . .	48
Figura 18 – Filtros de média sem o teste CUSUM . . . . .	52
Figura 19 – Diagrama do cálculo adaptativo da média . . . . .	53
Figura 20 – Filtros de média com o teste CUSUM . . . . .	53
Figura 21 – Estimaco Recursiva da Varincia da Potncia Ativa: Perodo de 20 dias . . . . .	54
Figura 22 – Definio de modo de operao: Perodo de 20 dias . . . . .	55
Figura 23 – Definio de modo de operao: Perodo de 60 horas . . . . .	55
Figura 24 – Operao Manual: Índices estatsticos na fase de projeto . . . . .	56
Figura 25 – Operao Manual: Índices estatsticos na fase de monitoramento em trecho normal . . . . .	57
Figura 26 – Operao Manual: Índices estatsticos na fase de monitoramento em trecho com muitos alarmes no SPE . . . . .	58



Figura 27 – Operação Manual: <i>Contribution Plots</i> na fase de monitoramento em trecho com muitos alarmes no SPE . . . . .	58
Figura 28 – Operação Manual com dois patamares: Índices estatísticos na fase de projeto em 168 MW . . . . .	59
Figura 29 – Operação Manual com dois patamares: Índices estatísticos na fase de projeto em 138 MW . . . . .	60
Figura 30 – Operação Manual com dois patamares: Índices estatísticos na transição de patamar de operação na fase de monitoramento . . . . .	60
Figura 31 – Operação Manual com dois patamares: Índices estatísticos na fase de monitoramento (exemplo 1) . . . . .	61
Figura 32 – Operação Manual com dois patamares: <i>Contribution Plots</i> na fase de monitoramento (exemplo 1) . . . . .	62
Figura 33 – Operação Manual com dois patamares: Índices estatísticos na fase de monitoramento (exemplo 2) . . . . .	63
Figura 34 – Operação Manual com dois patamares: <i>Contribution Plots</i> na fase de monitoramento (exemplo 2) . . . . .	63
Figura 35 – Operação via CAG: Potência Ativa . . . . .	64
Figura 36 – Operação via CAG: Índices estatísticos na fase de projeto . . . . .	65
Figura 37 – Operação via CAG: Índices estatísticos na fase de monitoramento (exemplo 1) . . . . .	66
Figura 38 – Operação via CAG: <i>Contribution Plots</i> na fase de monitoramento (exemplo 1) . . . . .	66
Figura 39 – Operação via CAG: Índices estatísticos na fase de monitoramento (exemplo 2) . . . . .	67
Figura 40 – Operação via CAG: <i>Contribution Plots</i> na fase de monitoramento (exemplo 2) . . . . .	68

# Lista de tabelas

Tabela 1 – Sensores do monitoramento de vibração . . . . .	37
Tabela 2 – Base de Dados para o PCA . . . . .	49

# Lista de abreviaturas e siglas

ANEEL	Agência Nacional de Energia Elétrica
CAG	Controle Automático da Geração
CEP	Controle Estatístico de Processos
CEPM	Controle Estatístico de Processos Multivariados
CLP	Controlador Lógico Programável
COS	Centro de Operação do Sistema da CEMIG
CuSum	<i>Cumulative Sum</i>
ONS	Operador Nacional do Sistema Elétrico
OPC	<i>The Interoperability Standard for Industrial Automation</i>
PCA	Análise de Componentes Principais ( <i>Principal Component Analysis</i> )
PIMS	<i>Process Information Management System</i>
RBC	<i>Reconstruction-based contribution</i>
SPE	<i>Squared Prediction Error</i>
UHE	Usina Hidrelétrica
VBA	<i>Visual Basic for Applications</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	Objetivos	13
1.2	Estrutura da Dissertação	13
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>15</b>
2.1	Monitoramento de Condição	15
2.1.1	Monitoramento de Condição e Análises Estatísticas	15
2.2	Controle Estatístico de Processos	16
2.2.1	Visão Geral	16
2.3	Controle Estatístico de Processos Multivariados	16
2.3.1	Visão Geral	16
2.3.2	Base estatística para dados multivariados	18
2.3.3	Carta $T^2$ de Hotelling	20
2.4	<i>Análise de Componentes Principais</i>	23
2.4.1	Visão Geral	23
2.4.2	Método PCA	24
2.5	Erro Quadrático de Predição (SPE)	28
2.6	Estatística Combinada ( $\varphi$ )	29
2.7	<i>Contribution Plots</i>	30
2.8	Comentários Finais	31
<b>3</b>	<b>USINA HIDRELÉTRICA DE EMBORCAÇÃO</b>	<b>32</b>
3.1	Informações Gerais	32
3.2	Modos de Operação	32
3.2.1	Características da Geração	32
3.3	Arquitetura de Automação	33
3.4	Sistema <i>Online</i> de Monitoramento de Vibração (SOMV)	35
3.5	Sensores da Rede Digital	37
3.6	Sistema PIMS: PI da Osisoft	37
3.7	Comentários Finais	38
<b>4</b>	<b>METODOLOGIA PARA MONITORAMENTO AMPLO DE UMA UNIDADE GERADORA HIDRELÉTRICA</b>	<b>39</b>
4.1	Processamento para criação da base de dados	39
4.1.1	Filtragem	39
4.1.2	Cálculo recursivo do RMS	40

<b>4.2</b>	<b>Definição do modo de operação</b>	<b>41</b>
4.2.1	Média Aritmética Recursiva	41
4.2.2	Filtro de Primeira Ordem	42
4.2.3	Estimador Recursivo da Variância	43
<b>4.3</b>	<b>Fase de Projeto</b>	<b>43</b>
4.3.1	Média adaptativa para projeto	44
4.3.2	Normalidade das Variáveis	46
4.3.3	Implementação do PCA	48
4.3.4	Implementação da carta $T^2$ de Hotelling e do SPE	49
<b>4.4</b>	<b>Fase de Monitoramento</b>	<b>50</b>
4.4.1	Monitoramento Eletrônico com Média Adaptativa	50
4.4.2	Teste CUSUM	51
<b>4.5</b>	<b>Comentários Finais</b>	<b>53</b>
<b>5</b>	<b>RESULTADOS E ANÁLISES</b>	<b>54</b>
<b>5.1</b>	<b>Modo de Operação</b>	<b>54</b>
<b>5.2</b>	<b>Cenários Operacionais</b>	<b>56</b>
5.2.1	Operação Manual via COS	56
5.2.2	Operação Manual via COS com Mudança de Ponto de Operação	58
5.2.3	Operação via CAG	63
<b>5.3</b>	<b>Comentários Finais</b>	<b>68</b>
<b>6</b>	<b>CONCLUSÕES</b>	<b>69</b>
<b>6.1</b>	<b>Conclusões Gerais</b>	<b>69</b>
<b>6.2</b>	<b>Sugestões de Continuidade</b>	<b>70</b>
	<b>REFERÊNCIAS</b>	<b>71</b>

# 1 Introdução

As usinas hidrelétricas são responsáveis pela maior parte da geração de energia elétrica na matriz energética brasileira. Segundo dados da Agência Nacional de Energia Elétrica, ANEEL, em 2007, dos mais de 100 mil megawatts de potência instalada, 75% eram provenientes de fontes hídricas (ANEEL, 2008). Mesmo depois dos longos períodos de seca dos últimos anos, que levaram a um aumento significativo da geração por usinas termoelétricas para compensar a redução da capacidade das hidrelétricas, estas ainda são responsáveis por cerca de 65% da potência gerada (ANEEL, 2016).

Monitoramento de condição é uma das principais estratégias para garantir o bom funcionamento de uma unidade geradora hidrelétrica. Sistemas de monitoramento de condição de máquinas elétricas buscam a detecção de faltas incipientes, para prover informações úteis para a operação e manutenção. Através da aquisição de dados de variáveis de processo, seguido do processamento dos sinais, diferentes técnicas são utilizadas visando a detecção e classificação de padrões (SILVA, 2013).

As plantas industriais, como as usinas hidrelétricas, em sua maioria, são compostas por dezenas de variáveis de processo. A ocorrência de alguma falta pode ser refletida em diversas variáveis simultaneamente. Essa correlação pode tornar a detecção de faltas e o diagnóstico das causas pela inspeção individual de cada grandeza inviável ou ineficiente. A complexidade de plantas industriais justifica o uso de abordagens que considerem um conjunto de dados de múltiplas dimensões para monitorar o desempenho, detectar eventos anormais e identificar causas. Da mesma forma, o monitoramento de condição pode beneficiar-se de técnicas que levem em consideração as correlações das variáveis de processo a serem monitoradas.

Nesse contexto, são pertinentes as técnicas de Monitoramento e Controle Estatístico Multivariado, que permitem a análise de características estatísticas de um conjunto de variáveis de processos, identificando padrões que possam indicar a ocorrência de faltas e suas possíveis causas. As técnicas de Controle Estatístico de Processos mais comuns, sob uma ótica univariada, supõem apenas uma variável de processo ou uma característica de qualidade. Para problemas multivariados, existe a possibilidade de uso de cartas de controle de cada variável individualmente, porém essa solução é ineficaz e pode induzir resultados equivocados (MONTGOMERY, 2004; MASON; YOUNG, 2002). Justifica-se, portanto, uma abordagem multivariada, com análises simultâneas de diversas variáveis.

Uma das principais técnicas utilizadas para tratar problemas multivariados é a Análise de Componentes Principais ou *Principal Component Analysis*. PCA é um procedimento cuja finalidade é a redução de um conjunto de dados com variáveis correlacionadas

em um conjunto de eixos não-correlacionados (componentes principais), de modo a reter informações relevantes, em termos da variabilidade total dos dados.

O monitoramento de uma unidade geradora de uma usina hidrelétrica deve ser capaz de detectar alterações no comportamento característico que possam ser indicativos de faltas como desalinhamento do eixo e cavitação na turbina, por exemplo. A fim de monitorar as condições de operação, uma série de medições é pertinente, como a vibração em diferentes eixos cartesianos, temperaturas dos mancais, entre outras. Cada falta afeta o sistema globalmente, causando alterações nas diversas variáveis presentes. Dessa forma, a aplicação de técnicas de monitoramento e controle estatístico de processos multivariado, além do PCA, pode prover informações relevantes sobre a presença de faltas na operação de unidades geradoras.

## 1.1 Objetivos

Este trabalho visa o monitoramento de condição de uma unidade geradora hidrelétrica através da aplicação de técnicas de Controle Estatístico de Processos Multivariado. Sobre dados de sensores, definem-se métricas para o desempenho global do sistema buscando a detecção de mudanças no monitoramento de uma unidade geradora, além da diferenciação entre possíveis faltas no sistema, as quais deseja-se alarmar, e mudanças no ponto de operação, intencionalmente aplicadas, quando os alarmes das cartas devem ser silenciados. A utilização de diferentes sistemas de aquisição de dados, com diferentes taxas de amostragem, demanda o processamento prévio, para a construção de uma base de dados apropriada para a aplicação das técnicas multivariadas. Após esta etapa inicial, busca-se descorrelacionar e reduzir a dimensão dos dados por meio da Análise de Componentes Principais e, posteriormente, a construção de cartas de controle multivariadas.

São realizadas implementações do PCA e da carta  $T^2$  de Hotelling, em ambiente Matlab, sobre dados de sensores relativos ao monitoramento de vibração de uma das unidades geradoras da Usina Hidrelétrica de Emborcação, ou UHE Emborcação, Figura 1, administrada pela CEMIG. Os algoritmos são implementados e testados no ambiente Matlab.

## 1.2 Estrutura da Dissertação

Esta dissertação está dividida em seis capítulos. O capítulo atual trata das motivações e objetivos do trabalho, além da estrutura da dissertação.

O capítulo 2, "Revisão Bibliográfica, discute os fundamentos e aplicações de Monitoramento de Condição, Controle Estatístico de Processos Multivariado e da Análise

Figura 1: Usina Hidrelétrica de Emborcação



Fonte: Informações gerais da UHE Emborcação no site da CEMIG. Acessado em 23 de fevereiro de 2016.

de Componentes Principais. Além disso, apresenta exemplos de usos dessas técnicas em problemas diversos.

No capítulo 3, "Usina Hidrelétrica de Emborcação", descreve-se a UHE Emborcação, sua arquitetura de automação, os sensores pertinentes ao trabalho e o sistema PIMS.

No capítulo 4, "Metodologia para Monitoramento de Condição da Unidade Geradora Hidrelétrica" são apresentados os algoritmos de processamento dos dados, o método PCA e o cálculo das cartas de controle, além de detalhes das implementações em ambiente Matlab.

No capítulo 5, apresentam-se os resultados e análises dos testes realizados em ambiente Matlab, para diferentes condições operacionais da usina.

O último capítulo trata das conclusões gerais e das sugestões de continuidade.



## 2 Revisão Bibliográfica

### 2.1 Monitoramento de Condição

Historicamente, fabricantes e usuários de máquinas elétricas utilizavam apenas mecanismos simples como proteção contra sobre-corrente e sobre-tensão para garantir a segurança e o bom funcionamento dos equipamentos. O crescimento da quantidade e complexidade dos processos envolvendo máquinas elétricas, levou a inúmeros desenvolvimentos na área de diagnóstico de faltas (NANDI; TOLIYAT; LI, 2005).

Nessa perspectiva, ganha força o monitoramento de condição das máquinas rotativas elétricas para a detecção de mudanças nas condições operacionais, que podem corresponder a faltas incipientes. Entende-se monitoramento de condição como estratégias para garantir o bom funcionamento de uma máquina, buscando a detecção de faltas incipientes para prover informações úteis para a operação e manutenção. Através da aquisição de dados de variáveis de processo, seguido do processamento dos sinais, diferentes técnicas são utilizadas visando a detecção e classificação de padrões (SILVA, 2013). No caso das unidades geradoras hidrelétricas, o monitoramento de condição é focado nas medições de vibração (SILVA, 2013). No entanto, integrar na análise outras variáveis como potências, correntes e temperaturas permite uma avaliação mais ampla do estado operacional de uma unidade. Nembhard et al. (2014) demonstra um método de diagnóstico de faltas utilizando medições de vibração e de temperatura nos enrolamentos de uma máquina rotativa, visando o monitoramento de condição da mesma.

#### 2.1.1 Monitoramento de Condição e Análises Estatísticas

Diversos estudos já foram divulgados integrando análises estatísticas em soluções de monitoramento de condição. Rutten, Nguyen e Golinval (2011) propõe o uso de diferentes variações de Análise de Componentes Principais para potencializar a capacidade de detecção de métodos utilizados no monitoramento de condição de sistemas industriais. Patel et al. (1996) propõe métodos estatísticos e de redes neurais para o monitoramento de condição de um motor de turbina a gás. He et al. (2009) demonstra o uso do PCA sobre medições de vibração para classificar padrões de falhas em máquinas.

## 2.2 Controle Estatístico de Processos

### 2.2.1 Visão Geral

Controle Estatístico de Processos, CEP, é constituído por técnicas e ferramentas para monitorar o desempenho de processos visando atingir melhorias na qualidade e na produtividade, evitar desperdícios e remanufatura. A aplicação dessas ferramentas é comum em duas situações. A primeira delas se dá na análise retrospectiva dos dados do processo. A segunda delas corresponde ao monitoramento das variáveis de processo ou características de qualidade quando assume-se que o processo encontra-se no chamado **estado de controle estatístico**.

Montgomery (2004) argumenta que qualquer processo, mesmo que rigorosamente projetado e mantido, apresenta uma variabilidade inerente devido a inúmeras pequenas causas inevitáveis e aleatórias. Quando estão presentes apenas causas aleatórias de variação, ditas causas comuns, o processo está em estado de controle estatístico, ou seja, no melhor estado de operação possível sem modificar o processo.

Para plantas industriais, operar nesse estado corresponde a ter um sistema de controle com bom desempenho, com sensores e atuadores em funcionamento correto e sintonia dos controladores adequada. Nessa condição, as saídas e indicadores de qualidade da planta apresentam um comportamento estatístico com média e variabilidade estáveis. A partir dessa condição aplicam-se as ferramentas de CEP para monitorar o processo, buscando identificar quando o mesmo saiu ou está próximo de sair do estado de controle estatístico, auxiliando na definição de paradas de manutenção, além de identificação de faltas e suas causas atribuíveis.

De acordo com Qin (2003), CEP tem sido largamente utilizado em diferentes processos industriais, como na indústria química e na microeletrônica. As funções típicas do CEP incluem detecção, identificação, diagnóstico e estimação de faltas, entre outros.

O monitoramento *just in time* de uma carta de controle pode indicar que um processo saiu do estado de controle estatístico quando novos pontos são localizados fora dos limites de controle. Além disso, pode-se identificar padrões determinísticos, como mudanças no valor médio e padrões cíclicos, indicando a presença de alguma causa atribuível.

## 2.3 Controle Estatístico de Processos Multivariados

### 2.3.1 Visão Geral

As técnicas de Controle Estatístico de Processos mais comuns, sob uma ótica univariada, supõem apenas uma variável de processo ou uma característica de qualidade. No entanto, como citado no Capítulo 1, a análise simultânea de diversas variáveis pode

ser mais eficiente na detecção de eventos anormais.

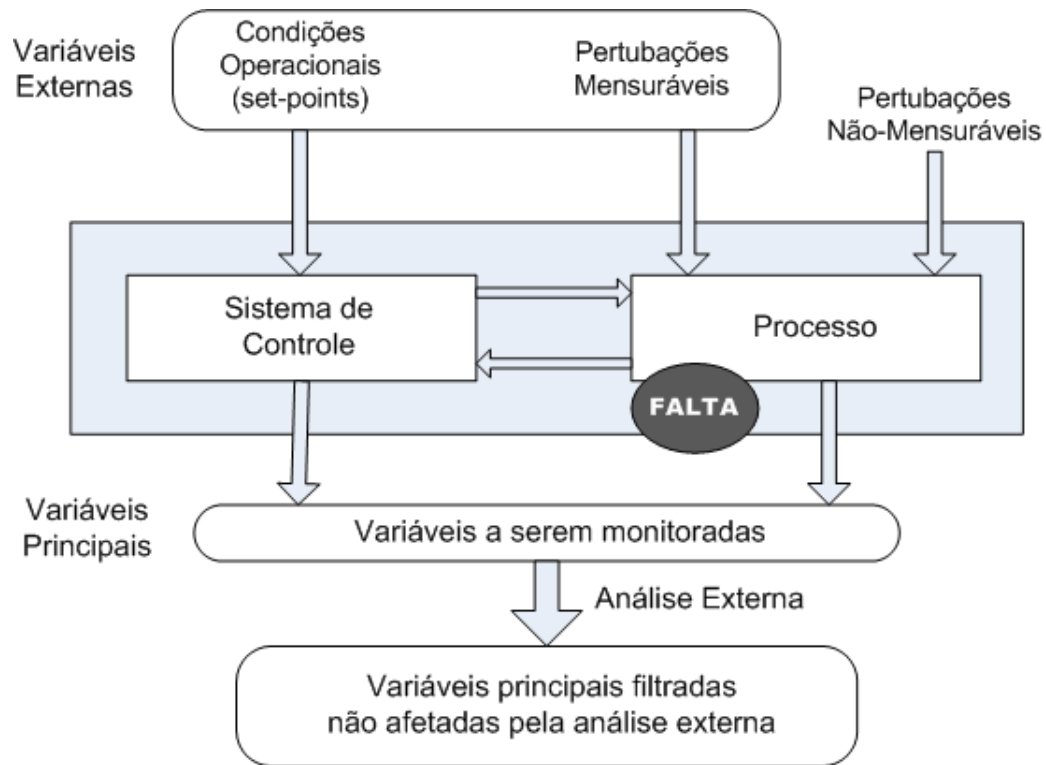
De acordo com Qin (2003), as técnicas de CEP univariado têm sido investigadas no contexto de controle de processos desde a década de 80. Qin (2003) destacam que o uso de CEP multivariado para detecção de situações anormais foi explorado fortemente na área de controle de qualidade. A aplicação dessas técnicas esteve mais associada ao monitoramento de características de qualidade. O uso da estatística  $T^2$  de Hotelling, juntamente com a Análise de Componentes Principais, sobre esses indicadores foca principalmente na detecção, com poucas aplicações na identificação de causas. O uso de CEP multivariado foi, no entanto, ampliado para o monitoramento de variáveis de processo. Com base nas mesmas estatísticas, em especial o índice  $T^2$  de Hotelling, o monitoramento de processos estendeu as funcionalidades do CEP multivariado, inclusive para detecção de faltas. Focando em variáveis de processo, em vez de variáveis de qualidade, os modelos estatísticos multivariados também provêm informações a respeito da correlação das variáveis, empiricamente.

O método PCA e a carta de Hotelling, assim como outras técnicas de CEP multivariado, assumem médias e variabilidades estáveis para a estimação das características estatísticas, no entanto diversos processos industriais apresentam mais de um ponto de operação em condições normais. Dessa forma, podem ocorrer mudanças nas médias, variâncias e covariâncias que não devem ser alarmadas, pois não correspondem a faltas. Kano et al. (2004) tratam justamente do problema de monitorar um processo com operação normal em diferentes pontos. Salienta-se a necessidade de diferenciar faltas de mudanças operacionais. Para tal objetivo, técnicas como o PCA e a carta de Hotelling são aliadas ao que é chamado de análise externa, *external analysis*. Esse procedimento consiste em separar as variáveis monitoradas em dois grupos, aquele cujas mudanças são diretamente explicáveis, como set-point de controladores ou perturbações mensuráveis, e aquele cujas mudanças não são determinadas pelos operadores do processo. Quando um alarme é gerado, verificam-se quais variáveis o causaram, para identificar se ocorreu uma falta. A Figura 2 mostra o diagrama dessa análise externa.

Na literatura, é possível encontrar uma vasta gama de aplicações de Monitoramento e Controle Estatístico Multivariado. Lee, Qin e Lee (2006) apresentam o uso da carta de Hotelling em um processo de tratamento de esgoto e outro de fabricação de dispositivos semicondutores, em conjunto com a Análise de Componentes Principais (PCA, seção 2.4), além de outras técnicas como o ICA, *Independent Component Analysis*, obtendo resultados satisfatórios na detecção de faltas.

Kano et al. (2002) aplicaram ferramentas de CEP multivariado, incluindo a carta de Hotelling e variações do PCA no simulador do *Tennessee Eastman Process*, um sistema de processamento químico com um reator, um condensador, separadores, entre outros equipamentos. Alkaya e Eker (2011) utilizam o PCA e a carta de Hotelling, porém com

Figura 2: Análise Externa



Fonte: adaptada de Kano et al. (2004).

um método de adaptação do limite de controle da carta para contornar situações que gerariam alarme falso, como possíveis regimes transitórios durante a operação normal de um processo.

Outros exemplos de aplicações de CEP Multivariado, utilizando a carta de Hotelling, podem ser verificados em Zhang e Zhang (2010) e Majid et al. (2011).

### 2.3.2 Base estatística para dados multivariados

Segundo Montgomery (MONTGOMERY, 2004), é comum a utilização da distribuição normal como modelo do comportamento estatístico de uma característica de qualidade contínua. A função de densidade de probabilidade normal univariada é dada pela Equação (2.1), onde  $\mu$  é a média de  $x$  e  $\sigma$  é o desvio padrão:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < +\infty \quad (2.1)$$

O termo expoente pode ser escrito como:

$$(x - \mu)(\sigma^2)^{-1}(x - \mu) \quad (2.2)$$

Esse termo refere-se à distância padronizada ao quadrado (em unidades de desvio padrão) de  $x$  à média  $\mu$ .

Para o caso multivariado, com  $p$  variáveis, pode-se representar o conjunto de dados por  $\mathbf{x} = [x_1, x_2, \dots, x_p]'$ , com um vetor de médias  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]'$ . As variâncias e covariâncias das componentes de  $\mathbf{x}$  podem ser representadas por uma **matriz de covariância**  $\Sigma$  de dimensão  $p \times p$ . Dessa forma, a distância padronizada de  $\mathbf{x}$  a  $\boldsymbol{\mu}$  para o caso multivariado é

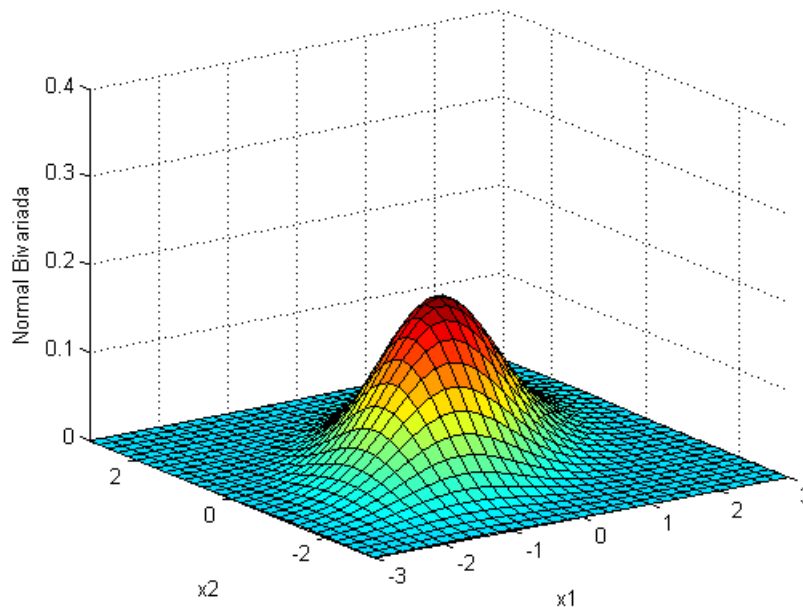
$$(\mathbf{x} - \boldsymbol{\mu})'(\boldsymbol{\Sigma})^{-1}(\mathbf{x} - \boldsymbol{\mu}) \quad (2.3)$$

Generalizando a Equação (2.3), a função densidade de probabilidade normal multivariada é dada pela Equação (2.4). A Figura 3 ilustra o caso bivariado.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'(\boldsymbol{\Sigma})^{-1}(\mathbf{x} - \boldsymbol{\mu})} \quad (2.4)$$

onde  $-\infty < x_i < +\infty, \quad \forall \quad i = 1, 2, \dots, p$

Figura 3: Distribuição de Probabilidade Normal Bivariada



Fonte: elaborada pelo autor.

Supondo uma amostra aleatória de uma distribuição normal multivariada definida por

$$\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_i, \dots, \mathbf{X}_n$$

onde o  $i^o$  vetor amostral corresponde as observações  $x_{i1}, x_{i2}, \dots, x_{ip}$ . A matriz de covariância amostral  $S$  é, portanto, definida pela Equação (2.5):

$$S_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad i, j = 1, 2, \dots, p \quad (2.5)$$

### 2.3.3 Carta $T^2$ de Hotelling

Uma das principais ferramentas do CEP multivariado é a carta de controle  $T^2$  de Hotelling. Análoga ao gráfico  $\bar{x}$  de Shewart para problemas univariados, a carta de Hotelling monitora o vetor de médias do processo. Segundo Mason e YOUNG (2002), essa ferramenta é dotada de considerável flexibilidade, sendo capaz de detectar pequenas mudanças no processo, além de viabilizar o monitoramento *offline* e *online* de processos multivariados.

Montgomery (2004) destaca que para a construção da carta  $T^2$  de Hotelling é necessário estimar  $\boldsymbol{\mu}$  e  $\boldsymbol{\Sigma}$  a partir das amostras do processo, ou seja, calcular o vetor de médias amostral  $\bar{\mathbf{X}}$  e a matriz de covariâncias amostral  $\mathbf{S}$ . Supondo  $m$  amostras de tamanho  $n$  de  $p$  variáveis de processo ou características de qualidade, as médias e variâncias para cada amostra são definidas por:

$$\bar{x}_{jk} = \frac{1}{n} \sum_{i=1}^n x_{ijk} \quad \begin{cases} j = 1, 2, \dots, p \\ k = 1, 2, \dots, m \end{cases} \quad (2.6)$$

$$S_{jk}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})^2 \quad \begin{cases} j = 1, 2, \dots, p \\ k = 1, 2, \dots, m \end{cases} \quad (2.7)$$

onde  $x_{ijk}$  é a  $i$ -ésima observação da  $j$ -ésima variável na  $k$ -ésima amostra e  $\bar{x}_{jk}$  são os elementos do vetor de médias de cada amostra  $\bar{\mathbf{X}}$ . A covariância entre as variáveis  $j$  e  $h$  na  $k$ -ésima amostra é

$$S_{jhk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})(x_{ihk} - \bar{x}_{hk}) \quad \begin{cases} k = 1, 2, \dots, m \\ j \neq h \end{cases} \quad (2.8)$$

Obtendo-se as médias das estatísticas  $\bar{x}_{jk}$ ,  $S_{jk}^2$  e  $S_{jhk}$  de todas as  $m$  amostras, tem-se

$$\bar{\bar{x}}_j = \frac{1}{m} \sum_{k=1}^m \bar{x}_{jk} \quad j = 1, 2, \dots, p \quad (2.9)$$

$$\bar{S}_j^2 = \frac{1}{m} \sum_{k=1}^m S_{jk}^2 \quad j = 1, 2, \dots, p \quad (2.10)$$

$$\bar{S}_{jh} = \frac{1}{m} \sum_{k=1}^m S_{jhk} \quad j \neq h \quad (2.11)$$

$\bar{x}_j$  são elementos do vetor  $\bar{\bar{\mathbf{X}}}$ . A matriz  $\mathbf{S}$  de dimensão  $p \times p$  pode ser utilizada como estimador de  $\Sigma$ , para um processo sob controle, e é definida pela Equação (2.12):

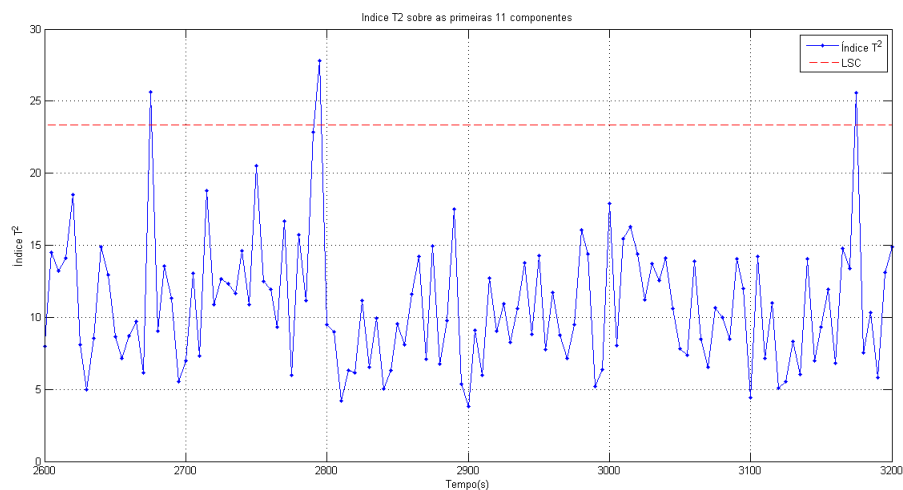
$$\mathbf{S} = \begin{bmatrix} \bar{S}_{11}^2 & \bar{S}_{12}^2 & \bar{S}_{13}^2 & \cdots & \bar{S}_{1p}^2 \\ \bar{S}_{21}^2 & \bar{S}_{22}^2 & \bar{S}_{23}^2 & \cdots & \bar{S}_{2p}^2 \\ \bar{S}_{31}^2 & \bar{S}_{32}^2 & \bar{S}_{33}^2 & \cdots & \bar{S}_{3p}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{S}_{p1}^2 & \bar{S}_{p2}^2 & \bar{S}_{p3}^2 & \cdots & \bar{S}_{pp}^2 \end{bmatrix} \quad (2.12)$$

Utilizando  $\bar{\bar{\mathbf{X}}}$  como a média do processo em estado sob controle e  $\mathbf{S}$  como estimativa da matriz de covariância  $\Sigma$ , pode-se calcular o índice  $T^2$  da carta de controle  $T^2$  de Hotelling a partir da Equação (2.13):

$$T^2 = n(\bar{\mathbf{X}} - \bar{\bar{\mathbf{X}}})' \mathbf{S}^{-1} (\bar{\mathbf{X}} - \bar{\bar{\mathbf{X}}}) \quad (2.13)$$

Dessa forma, o procedimento de uso da carta  $T^2$  de Hotelling consiste em escolher um grupo de amostras no qual o processo encontra-se em estado de controle estatístico, definir o limite de controle da carta, testar se o processo realmente estava sob controle quando foram obtidas essas amostras, ou seja, calcular o índice  $T^2$  para cada amostra e plotar o resultado na carta, verificando se nenhum ponto ultrapassou o limite de controle estabelecido. Esse procedimento é denominado **Fase 1** (MONTGOMERY, 2004). A Figura 4 ilustra um exemplo da carta de Hotelling.

Figura 4: Carta  $T^2$  de Hotelling



Fonte: elaborada pelo autor.

Montgomery (2004) define os limites da **Fase 1** de acordo com as Equações (2.14):

$$\begin{aligned} LSC &= \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \\ LIC &= 0 \end{aligned} \quad (2.14)$$

onde  $p$  corresponde ao número de variáveis de processo ou características de qualidade,  $m$  ao número de amostras,  $n$  ao tamanho de cada amostra e  $\alpha$  corresponde ao intervalo de confiança da carta, ou seja, o percentual esperado de amostras abaixo do limite de controle, em condições normais.  $F_{\alpha,p,mn-m-p+1}$  corresponde ao valor tabelado da distribuição  $F^1$ .

Após a **Fase 1**, parte-se para o monitoramento de novas amostras do conjunto de variáveis. Essa etapa, denominada **Fase 2**, consiste em definir novos limites para a carta de Hotelling, a partir das Equações (2.15), semelhantes aos limites da **Fase 1**, porém com uma pequena modificação no numerador:

$$\begin{aligned} LSC &= \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \\ LIC &= 0 \end{aligned} \quad (2.15)$$

É interessante destacar o caso específico em que o subgrupo amostral possui tamanho  $n = 1$ . Em ambientes industriais é comum lidar com problemas multivariados no qual as amostras de cada variável a ser monitorada são unitárias, como é o caso de variáveis de processo. Considerando  $\bar{\mathbf{x}}$  e  $\mathbf{S}$ , como o vetor média amostral e a matriz de covariância amostral de  $m$  amostras de tamanho 1 de  $p$  variáveis, respectivamente, a estatística  $T^2$  de Hotelling, apresentada anteriormente pela Equação (2.13), reduz-se a Equação (2.16):

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \quad (2.16)$$

Os limites de controle da Fase 2, para observações individuais, passam a ser definidos pelas Equações (2.17):

$$\begin{aligned} LSC &= \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha,p,m-p} \\ LIC &= 0 \end{aligned} \quad (2.17)$$

<sup>1</sup> A razão de duas variáveis com distribuição Qui-Quadrada,  $\chi^2$ , constitui uma nova variável com distribuição de probabilidade F. Sendo  $x_1$  uma variável com distribuição de probabilidade normal, ao obter amostras de  $x_1$ , a variância amostral  $S^2$  possui distribuição de probabilidade  $\chi^2$ . Para processos multivariados, as variâncias conjuntas de duas ou mais variáveis de processo são consideradas, sendo pertinente o uso da distribuição de probabilidade F. (MONTGOMERY, 2004; BRAGA, 2013)



Mason e YOUNG (2002) destacam que a definição do limite superior da carta de Hotelling para a Fase 1 a partir da distribuição Beta<sup>2</sup> conduz a resultados mais adequados no monitoramento de processos multivariados, uma vez que limites baseados na distribuição F ou  $\chi^2$  podem ser imprecisos. Dessa forma, os limites da Fase 1 são definidos pela Equação (2.18):

$$\begin{aligned} LSC &= \frac{(m-1)^2}{m} \beta_{\alpha, p/2, (m-p-1)/2} \\ LIC &= 0 \end{aligned} \quad (2.18)$$

O índice  $T^2$  é utilizado como uma métrica para o desempenho global do sistema. O monitoramento dessa métrica a cada nova amostragem das variáveis de processo pertinentes busca a detecção de mudanças no sistema, caracterizadas por pontos fora do limite da carta de Hotelling.

O uso do controle estatístico multivariado é apropriado para a detecção de mudanças no monitoramento de vibração de uma unidade geradora. Essas mudanças podem corresponder a faltas no sistema, as quais deseja-se alarmar, ou apenas mudanças no ponto de operação, quando os alarmes das cartas devem ser silenciados. Este pode ser o caso, por exemplo, quando a usina opera com potência reduzida, devido a baixa oferta de energia na entrada do processo, ou seja, um nível baixo de água no reservatório da usina. Situação semelhante também pode ocorrer por definição da ONS, Operador Nacional do Sistema Elétrico, ou do COS, centro de operação da CEMIG, visando regular o fluxo de potência nas linhas de transmissão de energia.

## 2.4 Análise de Componentes Principais

### 2.4.1 Visão Geral

A aplicação da carta de Hotelling é mais adequada para um número  $p$  reduzido de variáveis. Caso contrário estas cartas de controle tornam-se ineficientes na detecção de mudanças no processo. Modificações importantes no vetor de média são mais difíceis de serem detectadas para um número  $p$  elevado (MONTGOMERY, 2004). Além disso, variáveis muito correlacionadas geram problemas numéricos no cálculo da estimativa da inversa da matriz de covariância.

Uma alternativa para tratar esse tipo de problema é reduzir a dimensão do espaço de variáveis através do uso da *Análise de Componentes Principais* ou *Principal Compo-*

<sup>2</sup> A distribuição Beta,  $f(\beta)$ , restringe os valores de  $\beta$  para o intervalo (0,1), diferentemente das distribuições F e  $\chi^2$  que permitem quaisquer valores maiores do que zero. Dentro desse intervalo a distribuição Beta pode assumir diferentes formas, próximas a diferentes distribuições, como a normal, a  $\chi^2$  e a F (MASON; YOUNG, 2002).

*ment Analysis*, além de descorrelacionar os dados. PCA é uma técnica que consiste em realizar uma transformação num grupo de variáveis, representando-as em eixos ortogonais definidos como as direções de maior variabilidade dos dados.

O uso do PCA busca transformar as informações de  $p$  variáveis de processo correlacionadas em  $p$  componentes principais independentes. A primeira componente corresponde ao eixo de maior variabilidade dos dados. A segunda componente corresponde ao eixo de maior variabilidade ortogonal a primeira componente e assim por diante. As primeiras componentes principais podem conter grande parte da variabilidade total dos dados, logo é possível reduzir o espaço dimensional, utilizando apenas  $k$  componentes, sendo  $k < p$  (JACKSON, 2003; MASON; YOUNG, 2002). Qin (2003) também salienta que, no monitoramento de processos, parte das componentes principais podem ser consideradas inativas, ou seja, pouco representativas da variabilidade dos dados, de forma que o espaço dimensional possa ser reduzido. PCA visando o monitoramento estatístico multivariado tem sido utilizado em uma série de aplicações na indústria (LEE; QIN; LEE, 2006; KANO et al., 2002; KANO et al., 2004; ALKAYA; EKER, 2011; ZHANG; ZHANG, 2010; MAJID et al., 2011). Nesses exemplos, aplicam-se cartas de controle estatístico multivariado sobre algumas das componentes principais dos dados originais. A literatura contém outras versões do método de PCA, como a mistura probabilística de PCAs (KIM; KIM; BANG, 2003), cuja a finalidade é lidar com bases de dados com características não-estacionárias, como dados históricos de processos com diferentes patamares de operação.

## 2.4.2 Método PCA

Assumindo uma base de dados  $X$  com  $p$  variáveis e  $m$  amostras:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mp} \end{bmatrix} \quad (2.19)$$

O vetor de médias amostrais  $\bar{x}_i$  e o de variâncias amostrais  $\sigma_i^2$  das  $p$  variáveis são definidos por:

$$\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ki} \quad i = 1, 2, \dots, p \quad (2.20)$$

$$\sigma_i^2 = \frac{1}{m-1} \sum_{k=1}^m (x_{ki} - \bar{x}_i)^2 \quad i = 1, 2, \dots, p \quad (2.21)$$

A matriz de covariância amostral da base  $X$  (2.19) é dada por:

$$S = \begin{bmatrix} S_{11}^2 & S_{12}^2 & S_{13}^2 & \cdots & S_{1p}^2 \\ S_{21}^2 & S_{22}^2 & S_{23}^2 & \cdots & S_{2p}^2 \\ S_{31}^2 & S_{32}^2 & S_{33}^2 & \cdots & S_{3p}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{p1}^2 & S_{p2}^2 & S_{p3}^2 & \cdots & S_{pp}^2 \end{bmatrix} \quad (2.22)$$

onde

$$S_{ij}^2 = \frac{1}{m-1} \sum_{k=1}^m (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad i, j = 1, 2, \dots, p \quad (2.23)$$

Aplicar o PCA utilizando diretamente a matriz de covariância amostral é problemático, pois as variáveis podem possuir diferentes ordens de grandeza, diferentes médias e variâncias, logo as direções de maior variabilidade poderão ser polarizadas. Variáveis com diferentes grandezas físicas, por exemplo, ocasionariam um cálculo inadequado do PCA. Para contornar esse problema, uma alternativa é utilizar a matriz de fatores de correlação amostral,  $P$ , calculada pela Equação (2.25):

$$P = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \rho_{23} & \cdots & \rho_{2p} \\ \rho_{31} & \rho_{32} & \rho_{33} & \cdots & \rho_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \rho_{p3} & \cdots & \rho_{pp} \end{bmatrix} \quad (2.24)$$

onde

$$\rho_{ij} = \frac{1}{m-1} \sum_{k=1}^m \frac{(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \quad i, j = 1, 2, \dots, p \quad (2.25)$$

A matriz de fatores de correlação  $P$  é equivalente a matriz de covariância, normalizada pelas variâncias. Dessa forma, cada elemento  $\rho_{ij}$  representa a correlação entre as variáveis  $i$  e  $j$ , variando de 0 a 1. A diagonal dessa matriz é composta por elementos iguais a 1, uma vez que a correlação entre uma variável e ela mesma é máxima.

Outra alternativa seria normalizar a base, ou seja, subtrair a média e dividir pela raiz da variância de cada elemento da matriz  $X$ . Assim, cada uma das  $p$  variáveis possuirá média igual a zero e variância igual a um. Dessa forma, a matriz de covariância amostral da base normalizada será igual à matriz de fatores de correlação da base original, portanto os dois procedimentos são equivalentes.

As direções de maior variabilidade da base  $X$  são determinadas pelos autovetores da matriz  $P$ , que podem ser calculadas, por exemplo, pela decomposição por valores singulares de  $P$ :

$$\Lambda = C'PC \quad (2.26)$$

onde cada coluna de  $C$  é um autovetor de  $P$  e matriz diagonal  $\Lambda$  é composta pelos autovalores  $\lambda_i$  de  $P$ .

A primeira coluna de  $C$  é o autovetor que representa a direção de maior variabilidade dos dados. O segundo autovetor, ortogonal ao primeiro, representa a segunda direção de maior variabilidade e, assim, sucessivamente.

Como descrito por Montgomery (2004), calculada a matriz  $C$  de transformação, podem-se calcular as componentes  $\mathbf{z}$  associadas a cada amostra  $\mathbf{x} = \{x_1, x_2, \dots, x_p\}$ . Inicialmente, normaliza-se a amostra utilizando as médias e variâncias calculadas pelas Equações (2.20) e (2.21). As componentes principais  $\mathbf{z}$ , ou *scores*, são calculadas pelas Equações (2.28):

$$\begin{aligned} \mathbf{x}^* &= \{x_1^*, x_2^*, \dots, x_p^*\} \\ x_j^* &= \frac{x_j - \bar{x}_j}{\sigma_j} = \frac{x_j - \bar{x}_j}{\sqrt{\sigma_j^2}} \end{aligned} \quad (2.27)$$

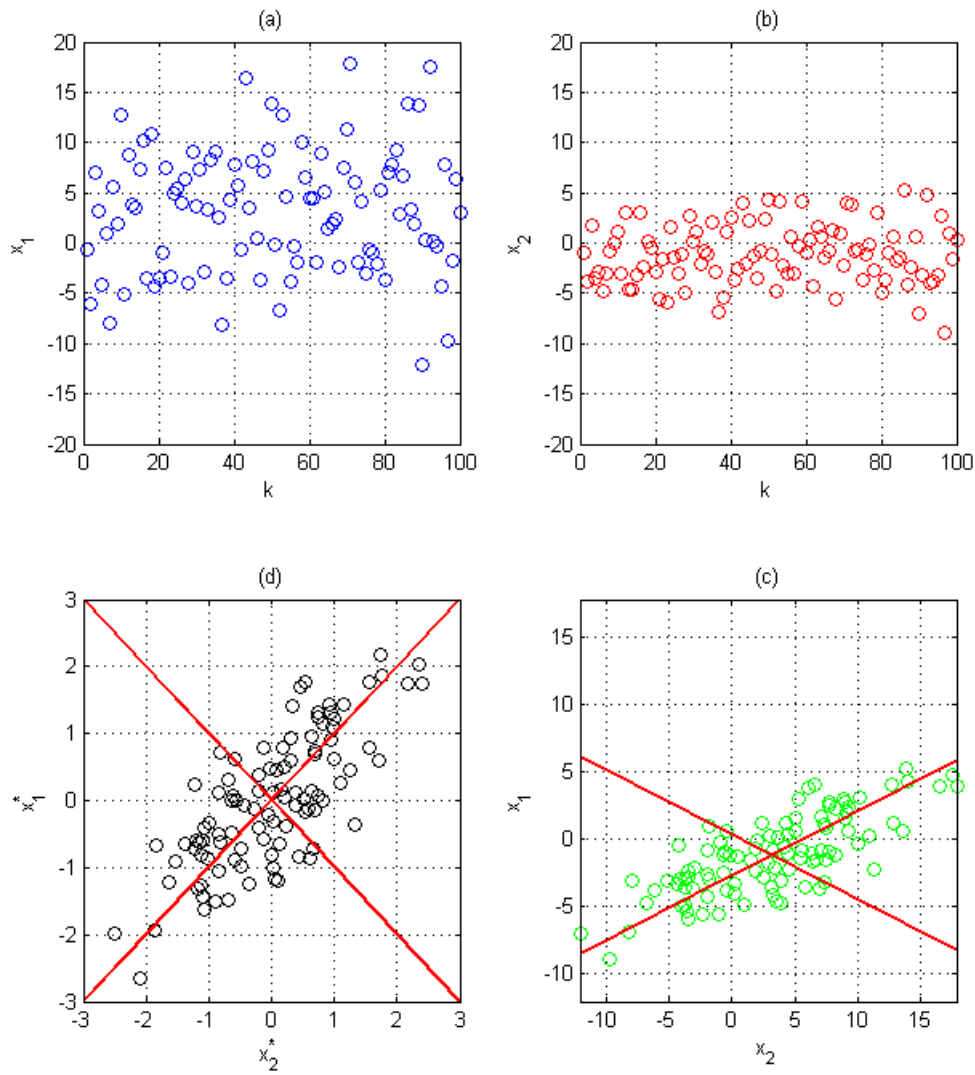
$$\begin{aligned} z_1 &= c_{11}x_1^* + c_{12}x_2^* + \dots + c_{1p}x_p^* \\ z_2 &= c_{21}x_1^* + c_{22}x_2^* + \dots + c_{2p}x_p^* \\ &\vdots \\ z_p &= c_{p1}x_1^* + c_{p2}x_2^* + \dots + c_{pp}x_p^* \end{aligned} \quad (2.28)$$

Os coeficientes  $c_{ij}$  são definidos como os valores do autovetor  $i$  associado ao autovalor  $\lambda_i$  da matriz de covariância ou da matriz de fatores de correlação  $P$ . Os índices dos autovalores obedecem à relação  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . A Figura 5 ilustra o método PCA para uma base de duas dimensões. Os gráficos (a) e (b) mostram as amostras de duas variáveis com distribuição de probabilidade normal. O gráfico (c) ilustra o espaço amostral normalizado bidimensional e, em vermelho, os eixos de maior variabilidade obtidos pelo PCA. O gráfico (d) ilustra o mesmo resultado do PCA, para o espaço amostral original.

Cada coeficiente é proporcional a influência da variável correspondente em cada componente. Este aspecto do PCA pode auxiliar no rastreamento das faltas, uma vez que

identificando-se os *scores* com maior influência no alarme da falta, os coeficientes de maior valor podem indicar as variáveis que, provavelmente, provocaram a mudança do processo para um estado indesejável.

Figura 5: Exemplo de PCA: (a) Variável  $\mathbf{x}_1$  ( $\bar{x}_1 = 3$  e  $\sigma_1 = 6$ ). (b) Variável  $\mathbf{x}_2$  ( $\bar{x}_2 = -1$  e  $\sigma_2 = 3$ ). (c) Gráfico de  $\mathbf{x}_1^*$  vs  $\mathbf{x}_2^*$  e eixos de maior variabilidade calculados pelo PCA. (d) Gráfico de  $\mathbf{x}_1$  vs  $\mathbf{x}_2$  e eixos de maior variabilidade calculados pelo PCA



Fonte: elaborada pelo autor.

Jackson (2003) define a Análise de Componentes Principais de forma semelhante, porém apresenta outra maneira de calcular os autovetores e autovalores. O cálculo das componentes, segundo Jackson (2003), é dado pela Equação (2.29), semelhante à Equação (2.28), porém na forma matricial e normalizando apenas pela média:

$$\mathbf{z}_i = \mathbf{u}_i'[\mathbf{x} - \bar{\mathbf{x}}] \quad (2.29)$$

onde  $\mathbf{z}_i$  são as componentes principais, ou *scores* das componentes principais,  $\mathbf{u}_i$  são os vetores direcionais da transformação (autovetores normalizados de  $\mathbf{S}$ ),  $\mathbf{x}$  é o vetor de  $p$  variáveis originais e  $\bar{\mathbf{x}}$  o vetor de média. Os vetores  $\mathbf{u}_i$  são obtidos pela solução do sistema de duas equações (2.30) e (2.31)

$$[\mathbf{S} - l\mathbf{I}]\mathbf{t}_i = 0 \quad (2.30)$$

e

$$\mathbf{u}_i = \frac{\mathbf{t}_i}{\sqrt{\mathbf{t}_i' \mathbf{t}_i}} \quad (2.31)$$

onde  $\mathbf{S}$  é a matriz de covariância amostral das variáveis originais,  $l$  são as raízes da equação  $|\mathbf{S} - l\mathbf{I}| = 0$  (autovalores de  $\mathbf{S}$ ),  $\mathbf{I}$  é a matriz identidade de dimensão  $p$  e  $\mathbf{t}_i$  são os vetores solução da Equação (2.30) (autovetores de  $\mathbf{S}$ ).

Um dos objetivos do PCA é reduzir a dimensão da base de dados, sem perdas significativas na variabilidade. É necessário, portanto, calcular-se a proporção de variabilidade explicada por cada componente. Quanto maior for o autovalor associado à determinada componente, melhor ela explica a variabilidade total dos dados. A proporção de variabilidade de uma componente  $z_j$  é obtida pela Equação 2.32:

$$\frac{\lambda_j}{\sum_{j=1}^p \lambda_j} \quad (2.32)$$

Dessa forma, é possível escolher um número reduzido de componentes que expliquem a maior parte da variabilidade, diminuindo a dimensão da base de dados e aumentando a eficiência de outras técnicas estatísticas como a carta  $T^2$  de Hotelling. Este é um método de fácil interpretação para auxílio na escolha do número de componentes.

## 2.5 Erro Quadrático de Predição (SPE)

O erro quadrático de predição, SPE, ou estatística Q (JACKSON, 1991), é um índice que computa o erro entre os dados originais e o modelo estatístico proveniente das componentes principais retidas. Esse índice indica quanto o modelo de PCA calculado explica a variabilidade do conjunto de dados inicial.

Uma das formas de calcular o SPE é a seguinte:

$$\text{SPE} = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}), \quad (2.33)$$

$\hat{\mathbf{x}}$  é a predição da amostra computada através das componentes retidas

O cálculo do limite superior de controle do SPE, segundo Jackson (1991) é dado por:

$$\theta_i = \sum_{j=k+1}^p \lambda_j^i, \quad i = 1, 2, 3 \quad (2.34)$$

e

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}. \quad (2.35)$$

o valor crítico  $\alpha$ -superior  $Q_\alpha$  é dado por:

$$Q_\alpha = \theta_1 \left[ \frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1 \right]^{1/h_0}, \quad (2.36)$$

## 2.6 Estatística Combinada ( $\varphi$ )

O índice combinado  $\varphi$  (YUE; QIN, 2001) constitui uma soma dos índices  $T^2$  e do SPE ponderada pelos respectivos limites de controle. O cálculo do índice é dado por:

$$\varphi = \frac{T^2}{T_\alpha^2} + \frac{\text{SPE}}{Q_\alpha}. \quad (2.37)$$

O limite de controle, demonstrado por (YUE; QIN, 2001), para um nível de significância  $\alpha$ , é calculado como:

$$\varphi_\alpha = g \chi_{\alpha; h}^2, \quad (2.38)$$

em que o coeficiente  $g$  é calculado como:

$$g = \left( \frac{k}{\chi_{\alpha; k}^4} + \frac{\sum_{i=k+1}^p \lambda_i^2}{Q_\alpha^2} \right) \left( \frac{k}{\chi_{\alpha; k}^2} + \frac{\sum_{i=k+1}^p \lambda_i}{Q_\alpha} \right)^{-1} \quad (2.39)$$

sendo  $k$  o número de componentes retidas, e o grau de liberdade  $h$  para a distribuição  $\chi^2$  é:

$$h = \left( \frac{k}{\chi_{\alpha; k}^2} + \frac{\sum_{i=k+1}^p \lambda_i}{Q_\alpha} \right)^2 \left( \frac{k}{\chi_{\alpha; k}^4} + \frac{\sum_{i=k+1}^p \lambda_i^2}{Q_\alpha^2} \right)^{-1}. \quad (2.40)$$

O índice combinado visa simplificar a etapa de monitoramento em tempo certo. Inspeccionando apenas um índice, espera-se detectar os alarmes gerados tanto pelo índice  $T^2$  quanto pelo SPE.

## 2.7 Contribution Plots

*Contribution plot* é um método para determinação da contribuição de cada variável em um índice estatístico. É tipicamente utilizado em conjunto com ferramentas como o PCA, o índice  $T^2$  e o SPE. Nesse contexto ele representa uma quantização do impacto de cada variável em cada amostra das componentes retidas do PCA. Na detecção de uma mudança, indicada por uma amostra com índice estatístico acima do limite, esse método ajuda a identificar as variáveis que mais contribuíram com uma condição anormal de operação (ZHU; BRAATZ, 2014).

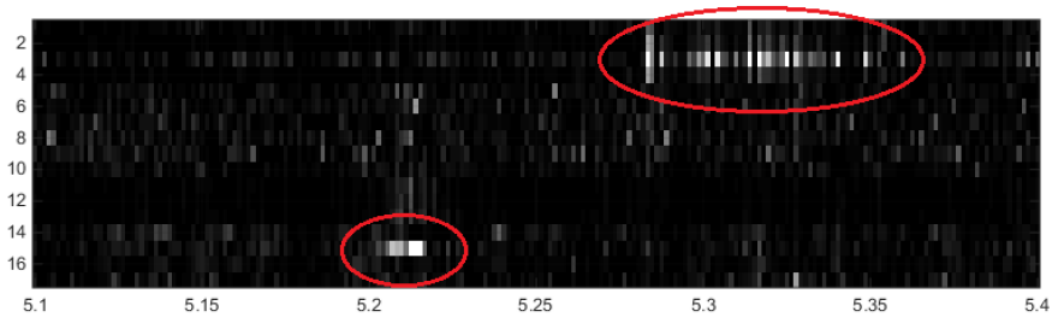
Alcala e Qin (2009) propõem um método para cálculo da contribuição baseada em reconstrução (RBC, de *reconstruction-based contribution*). Em um sistema com  $p$  variáveis normalizadas, na ocorrência de uma falta na variável  $x_i$ , o vetor amostral correspondente é  $\mathbf{x} \in \mathbb{R}^{p \times 1}$  e a direção da falta é  $\xi_i$ , sendo  $\xi_i$  a  $i$ -ésima coluna da matriz identidade. o vetor reconstruído ao longo da direção  $\xi_i$  é dado por:

$$\mathbf{x}_i^{rec} = \mathbf{x} - \xi_i f_i, \quad (2.41)$$

sendo  $f_i$  a componente de amplitude da falta. O valor de  $f_i$  deve minimizar o índice de detecção calculado sobre  $\mathbf{x}_i^{rec}$ . Alcala e Qin (2009) apresentam o restante do cálculo das contribuições.

A Figura 6 ilustra uma forma de representação gráfica de *contribution plots*. O eixo vertical corresponde às variáveis numeradas de 1 a 17 e o eixo horizontal ao tempo. A escala de cinza é utilizada para indicar a contribuição de cada variável em cada amostra. Quanto mais próximo da cor preta menor é a contribuição. Quanto mais próximo da cor branca, maior é a contribuição. Nesse exemplo, a variável de número 15 concentra grande parte da variabilidade dos dados no primeiro trecho em destaque, assim como variável de número 3 no segundo trecho. Essa representação gráfica auxilia na identificação das variáveis mais provavelmente relacionadas aos alarmes gerados pelos índices estatísticos  $T^2$ , SPE e combinado.

Figura 6: Exemplo de representação gráfica de *contribution plots*



Fonte: elaborada pelo autor.



## 2.8 Comentários Finais

No presente capítulo foi apresentada uma revisão de ferramentas de CEP multivariado, em especial a carta  $T^2$  de Hotelling e o PCA. Visando a aplicação dessas técnicas no monitoramento amplo de condição de uma unidade geradora da Usina Hidrelétrica de Emborcação, o capítulo seguinte trata da descrição do processo, detalhando dados gerais da Usina, a arquitetura de automação utilizada e os sensores para monitoramento.

## 3 Usina Hidrelétrica de Emborcação

### 3.1 Informações Gerais

A Usina Hidrelétrica de Emborcação, gerida pela CEMIG, está localizada no Rio Paranaíba no município de Araguari, no Triângulo Mineiro, na divisa entre os estados de Minas Gerais e de Goiás. A usina possui quatro unidades geradoras totalizando uma capacidade de geração de 1192 MW. Estão presentes duas subestações, uma de 500 kV e outra de 138 kV. As unidades turbina-gerador são de grande porte, com constantes de tempo térmicas de até uma hora.

### 3.2 Modos de Operação

Duas das quatro máquinas da Usina Hidrelétrica de Emborcação podem operar como gerador ou como motor síncrono. A unidade geradora utilizada neste trabalho é operada apenas como gerador, por isso os modelos estatísticos estão restritos a este modo de operação.

#### 3.2.1 Características da Geração

A Potência Nominal dos geradores é de 298 MW, no entanto esse patamar de geração só é possível quando o reservatório da usina encontra-se próximo de sua capacidade máxima. Esta condição não foi observada nos últimos anos, logo as máquinas são operadas, tipicamente, entre 130 e 250 MW, dependendo do nível do reservatório e, principalmente, da demanda por energia.

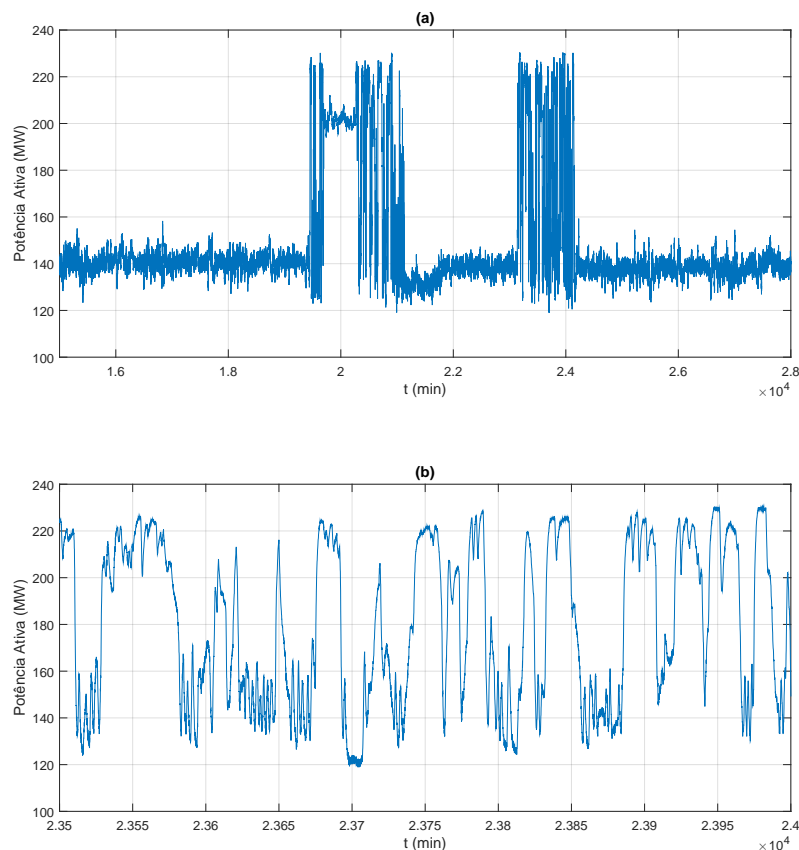
A operação da UHE Emborcação possui uma característica particular que a difere da maior parte das usinas hidrelétricas. Ela faz parte do **Controle Automático da Geração, ou CAG**, que é um mecanismo do **Operador Nacional do Sistema Elétrico, ONS**, responsável pelo equilíbrio entre geração e carga. Para realizar essa tarefa, as concessionárias de energia disponibilizam algumas máquinas de algumas usinas para o CAG determinar, diretamente, o *set-point* de potência ativa. Um *link* de rede entre o CAG e o Centro de Operação da CEMIG, COS, em Belo Horizonte, transmite o valor do *set-point*, que é roteado diretamente para a usina e copiado na memória do CLP correspondente, sendo a referência de potência ativa nos momentos em que a máquina é "operada" pelo CAG.

Nesse estado de operação particular, o *set-point* de potência ativa sofre variações bruscas, de até 80 MW, e é mantido fixo por períodos muito curtos, geralmente, menores

que 30 minutos.

A Figura 7 ilustra essa característica peculiar. Nos momentos em que a máquina tem seu *setpoint* definido pelo CAG, a potência ativa sofre variações bruscas em curto período entre 120 e 230 MW. Por outro lado, quando a usina é operada localmente, sob a supervisão do COS, o *setpoint* é praticamente fixo.

Figura 7: Potência Ativa do Gerador em operação mista: CAG e local. (a) Nove dias de operação. (b) Oito horas de operação com CAG



Fonte: elaborada pelo autor.

### 3.3 Arquitetura de Automação

A Usina Hidrelétrica de Emborcação possui um conjunto de CLP's Siemens S7-400, interligados em uma rede ethernet por meio do protocolo proprietário S7 da Siemens:

- 1 para cada unidade geradora
- 1 para os Sistemas Auxiliares
- 1 para a Tomada d'Água e Vertedouro

- 1 para a subestação de 138kV
- 1 para a subestação de 500kV

Além dos CLP's, existem dois servidores de alarmes, dois módulos de comunicação com o centro de operação da CEMIG, em Belo Horizonte, e dois computadores com um sistema SCADA para supervisão e operação do processo. Os CLP's, interligados por uma rede Ethernet por fibra ótica, estão conectados em um switch, ou estrela ótica. Os servidores de alarmes e os terminais de supervisão constituem outra rede ethernet por fibra ótica.

A usina possui um sistema do tipo PIMS, *Plant Information Management System*. Esses sistemas buscam coletar dados de uma ou várias plantas, armazená-los em um banco de dados e possibilitar o acesso aos mesmos para auxiliar a tomada de decisão em diferentes níveis de gerência. Sistemas PIMS tem como base um banco de dados temporal<sup>1</sup> para armazenamento de dados históricos de processos. O PIMS da usina é o sistema PI, da empresa *Osisoft*, com banco de dados denominado Servidor PI.

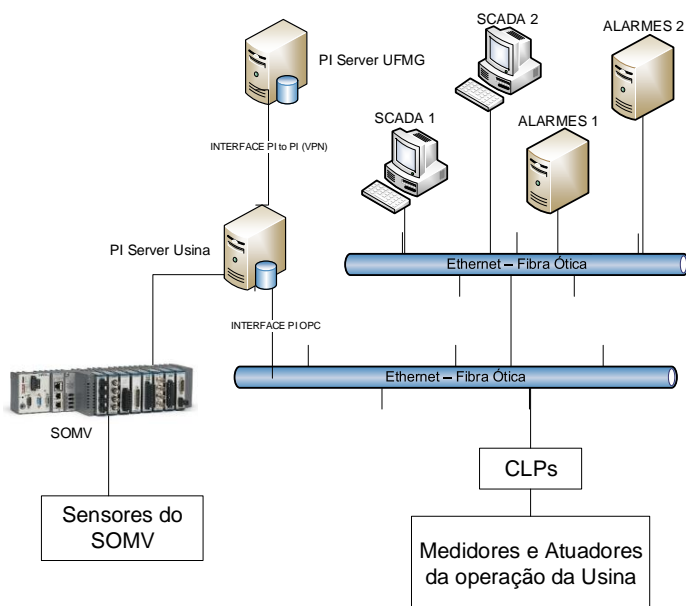
O PIMS coleta dados da rede de CLP's, via interface PI OPC, e envia-os para outro servidor PI localizado na UFMG, via interface *PI to PI*, como mostra a Figura 8. Para possibilitar o envio sem perdas, a taxa de amostragem das variáveis no Servidor PI foi definida experimentalmente em  $0,2Hz$ , ou seja, um novo conjunto de amostras unitárias é armazenado a cada 5 segundos.

A usina também possui um sistema de monitoramento de vibração, descrito na seção 3.4, que envia dados para o Servidor PI com uma taxa de amostragem de  $1,0Hz$ . A visão geral da arquitetura de automação da usina está ilustrada na Figura 8:

---

<sup>1</sup> Banco de dados temporal caracteriza-se por armazenar dados históricos temporalmente. Cada valor de uma variável do banco, ou *tag*, é armazenada sequencialmente com o instante de tempo, ou *Time Stamp*, e uma informação de qualidade do dado.

Figura 8: Arquitetura de automação da UHE Emborcação



Fonte: elaborada pelo autor.

### 3.4 Sistema *Online* de Monitoramento de Vibração (SOMV)

O sistema de monitoramento de vibração de uma unidade geradora de uma usina hidrelétrica objetiva a detecção de mudanças, que possam implicar em faltas, como desalinhamento do eixo e cavitação, para prover informações úteis para as equipes de operação e manutenção.

A UHE Emborcação possui o Sistema *Online* de Monitoramento de Vibração, SOMV, desenvolvido no projeto de Pesquisa e Desenvolvimento “GT338 - Desenvolvimento de Sistema Integrado Para Supervisão Remota Segura de Usinas Hidrelétricas”, parceria entre a UFMG e a CEMIG. Ele é composto por 10 sensores para uma unidade geradora. A aquisição e processamento dos dados dos sensores é feita por meio de uma controladora *CompactRIO cRIO-9082* da *National Instruments* (Figura 9).

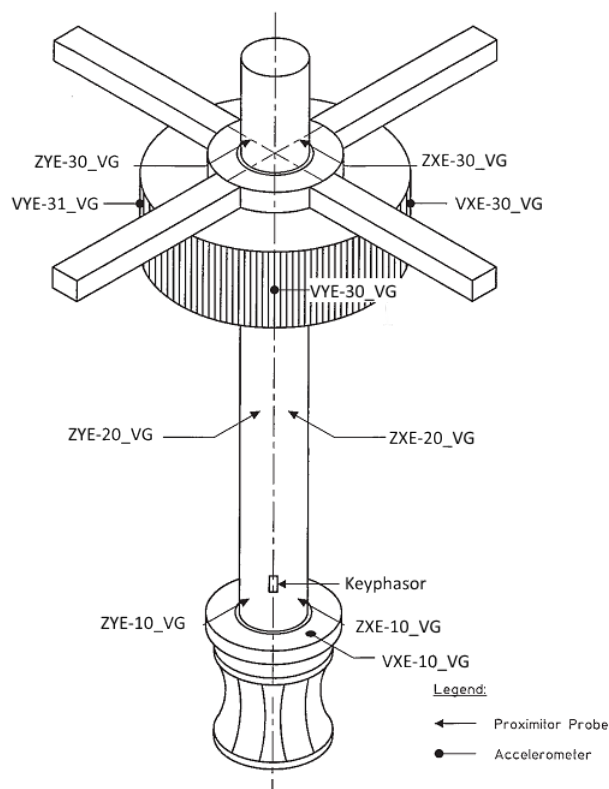
Figura 9: *CompactRIO cRIO-9082* da *National Instruments*

Fonte: elaborada pelo autor.

A instrumentação do SOMV corresponde àquela tipicamente utilizada para o mo-

monitoramento de vibração em uma unidade geradora. Acelerômetros em diferentes eixos são utilizados para monitorar a aceleração e velocidade da vibração e sensores de proximidade, em quadratura, indicam o estado de alinhamento do gerador. A Figura 10 ilustra a distribuição dos sensores na unidade geradora, enquanto a tabela 1 relaciona cada sensor da Figura 10 com sua função.

Figura 10: Distribuição dos Sensores do Sistema Online de Monitoramento de Vibração no gerador



Fonte: elaborada pelo autor.

Há um sensor de proximidade, indicado na Figura 10 como "Keyphasor", cuja função é marcar as voltas do eixo do gerador. Localizado acima da turbina e abaixo do estator, o sensor está na mesma altura de um pequeno furo no eixo da máquina. Dessa forma, no momento em que o furo passa pelo sensor há uma queda abrupta no sinal de saída do Keyphasor, sendo possível determinar, por processamento digital, os instantes de cada revolução do eixo. O uso do Keyphasor é necessário para processar os dados dos proximetros e acelerômetros, uma vez que a velocidade angular de rotação do eixo não é exatamente constante.

A controladora cRIO coleta continuamente os dados brutos dos sensores, processando-os para obter o valor RMS de velocidade da vibração a partir das medidas dos acelerômetros e o deslocamento pico-a-pico dos proximetros. Estas informações são enviadas, via interface PI OPC, para o Servidor PI da usina. Algoritmos de análise espectral para

Tabela 1: Sensores do monitoramento de vibração

<b>TAG</b>	<b>Descrição</b>
VXE-10_VG	Vibração (Valor RMS da velocidade) do Acelerômetro da Tampa da Turbina
VXE-30_VG	Vibração (Valor RMS da velocidade) do Acelerômetro em 0º do Gerador
VYE-30_VG	Vibração (Valor RMS da velocidade) do Acelerômetro em 120º do Gerador
VYE-31_VG	Vibração (Valor RMS da velocidade) do Acelerômetro em 240º do Gerador
ZXE-10_VG	Deslocamento Pico-a-Pico do Proxímetro X da Turbina
ZXE-20_VG	Deslocamento Pico-a-Pico do Proxímetro X do Mancal Intermediário
ZXE-30_VG	Deslocamento Pico-a-Pico do Proxímetro X do Gerador
ZYE-10_VG	Deslocamento Pico-a-Pico do Proxímetro Y Turbina
ZYE-20_VG	Deslocamento Pico-a-Pico do Proxímetro Y do Mancal Intermediário
ZYE-30_VG	Deslocamento Pico-a-Pico do Proxímetro Y do Gerador

detecção de faltas também são computados na controladora, porém seus resultados são disponibilizados apenas localmente devido ao grande volume de dados.

### 3.5 Sensores da Rede Digital

A integração do sistema PIMS e a rede digital da usina possibilita o acesso aos dados de diversos sensores de cada unidade geradora. Dentre eles, alguns foram selecionados como mais pertinentes para o monitoramento de condição da unidade geradora:

- **Temperaturas:** As medições de temperatura dos mancais, dos radiadores e da água do gerador indicam se a máquina está operando sob limites especificados no projeto da usina. Além disso, desalinhamentos podem provocar aquecimentos desiguais na estrutura mecânica da unidade.
- **Grandezas Elétricas:** Potência Ativa, Reativa e Correntes indicam a condição operacional da máquina e mudanças de carga
- **Fluxo de água nos radiadores:** pode sofrer alterações decorrentes de sobreaquecimentos

### 3.6 Sistema PIMS: PI da Osisoft

O sistema PI, baseado na plataforma Windows, possui centenas de interfaces para a coleta de dados encapsulados em vários protocolos de comunicação, arquivos de texto, web, etc. Dessas interfaces, duas fazem parte da coleta dos dados dos sensores da rede digital da usina (seção 3.5): a *PI to PI* e a *PI OPC*. Além disso, o sistema possui uma série

de módulos de software com diferentes funcionalidades. Desses componentes, destacam-se dois: o ACE e o ProcessBook.

A interface PI OPC é composta por um cliente para realizar a comunicação via protocolo OPC, padrão industrial utilizado em larga escala, com diversos tipos de dispositivos. A interface faz a conversão dos dados recebidos e armazena-os no Servidor PI. A interface *PI to PI*, por sua vez, realiza a comunicação entre dois Servidores PI, através de uma VPN, *Virtual Private Network*, copiando as informações de tags de um servidor para tags de outro.

O *PI ACE*, *Advanced Computing Engine*, é um módulo do sistema PI que permite a realização de cálculos avançados com os dados das variáveis armazenadas no Servidor PI. Utilizando um ambiente de desenvolvimento como o Microsoft Visual Studio, com a instalação do *PI ACE*, é possível, via programação em Visual Basic, acessar os dados das tags, efetuar cálculos e armazenar resultados no Servidor PI.

O ProcessBook é uma aplicação cliente do Sistema PI que possibilita a construção de telas para visualização dos valores das tags armazenadas no Servidor PI. Suas ferramentas incluem gráficos de tendência, ferramentas básicas de desenho, inserção de imagens, entre outras. Existe também a possibilidade de estender as funcionalidades do software via scripts em Visual Basic. Através do ProcessBook é possível visualizar as cartas de controle estatístico.

O sistema *PI* da usina possui as ferramentas necessárias para uma futura implementação da metodologia de monitoramento amplo de condição descrita no próximo capítulo

### 3.7 Comentários Finais

Este capítulo tratou do processo sobre o qual são aplicadas as técnicas de Controle Estatístico Multivariado visando o monitoramento de condição. Detalhou-se a UHE Emborcação, os sensores do sistema de monitoramento de vibração e a plataforma de desenvolvimento do sistema PI. O capítulo seguinte corresponde a metodologia para utilização das técnicas descritas no capítulo 2, sobre as variáveis de processo descritas no capítulo 3.



## 4 Metodologia para Monitoramento Amplo de uma Unidade Geradora Hidrelétrica

Apresenta-se a metodologia para desenvolvimento de um sistema de monitoramento amplo de uma unidade geradora hidrelétrica. A primeira etapa consiste no pré-processamento para construção de uma base de dados uniforme e própria para a aplicação de técnicas de CEP multivariado.

A etapa de projeto contém a definição dos trechos em que o processo é considerado em estado de controle estatístico, ou seja, na melhor condição operacional para cada modo de operação. Além disso, são definidos os parâmetros necessários para a realização do monitoramento, como número de componentes principais retidas, matrizes de covariância e limites das cartas de controle.

A etapa final consiste na aplicação de adaptações de médias de modo a possibilitar o monitoramento contínuo em diferentes patamares de potência ativa para cada modo de operação. Em seguida é feita uma validação *offline* com dados históricos do processo.

### 4.1 Processamento para criação da base de dados

Como descrito no capítulo anterior, os dados para a criação da base vem de duas fontes diferentes:

- Rede digital da usina: comunicação OPC entre os CLP's e o Servidor PI. Taxa de amostragem de  $0,2Hz$ .
- Sistema de monitoramento de vibração: Aquisição pela controladora cRIO-9082. Taxa de amostragem de  $10KHz$ .

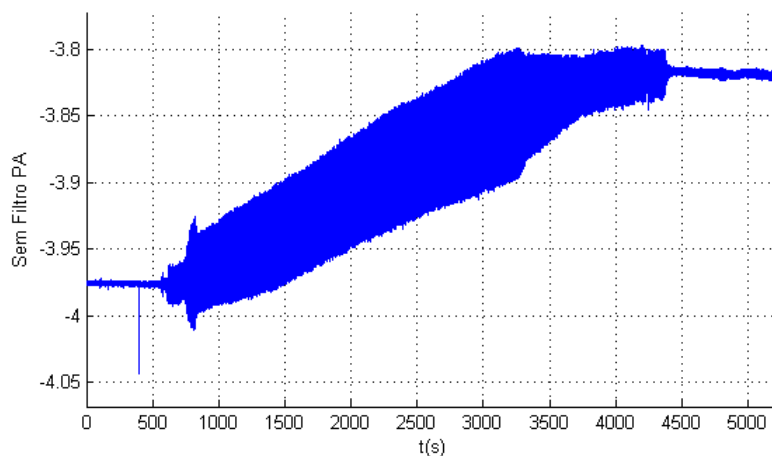
Além de serem duas fontes de dados distintas, fora de sincronia, as taxas de amostragem são muito diferentes. O desafio foi processar os dados dos sensores do sistema de monitoramento, transformando-os em dados com a taxa de amostragem de  $0,2Hz$ , sincronizadas, preservando informações importantes para o controle estatístico, como média e variabilidade.

#### 4.1.1 Filtragem

A Figura 11 mostra o sinal bruto do acelerômetro, sem filtros, obtido através de um teste de campo feito em 2012. A máquina, inicialmente parada, foi partida até ter

uma potência ativa de geração nominal, depois teve a potência reduzida em 40% e passou por um procedimento de frenagem até uma nova parada. A informação de interesse para o monitoramento da vibração é a variabilidade desse sensor, ou seja, as frequências altas do sinal. A componente contínua dos sinais deve ser desprezada. Para isso realiza-se a aquisição com um capacitor para retirar a componente contínua dos sinais, preservando a variabilidade dos mesmos.

Figura 11: Acelerômetro do Estator



Fonte: elaborada pelo autor.

Além disso, foram utilizados filtros *anti-aliasing* na aquisição dos dados dos sensores relativos à vibração.

#### 4.1.2 Cálculo recursivo do RMS

Para reduzir a taxa de amostragem por um fator de 10.000, foi necessário traduzir a variabilidade do sinal em um valor para cada instante de tempo. Utilizou-se, para esse fim, o valor RMS (*Root Mean Square*) do sinal:

$$RMS(x) = \sqrt{\bar{x}^2 + \sigma^2(x)} \quad (4.1)$$

onde  $\bar{x}$  é a média estimada do sinal  $x$  e  $\sigma(x)$  é o desvio padrão estimado do sinal  $x$ . Dessa forma, o valor RMS dos sensores filtrados, que possuem média zero, será igual ao desvio padrão. O valor RMS é, portanto, uma medida da variabilidade dos sensores filtrados.

Considerando a necessidade de esse cálculo ser contínuo e com baixo custo computacional para não sobrecarregar a controladora cRIO-9082, escolheu-se realizar um cálculo recursivo do valor RMS. Como deseja-se obter a **raiz da média do sinal ao quadrado**,

ou *Root Mean Square*, foi utilizada a Equação (4.2) para estimar a média dos sinais ao quadrado, como segue:

$$y_{quad}(k) = y_{quad}(k-1) + \alpha(y^2(k) - y_{quad}(k-1)) \quad (4.2)$$

Extraindo a raiz de  $y_{quad}(k)$ , tem-se uma estimativa do valor RMS do sinal a cada instante de tempo:

$$RMS_y(k) = \sqrt{\frac{2-\alpha}{2(1-\alpha)}y_{quad}(k)} \quad (4.3)$$

## 4.2 Definição do modo de operação

Como discutido no capítulo anterior, a Usina de Emborcação possui dois modos distintos de operação para a geração. Um com *setpoints* fixos para a potência ativa, definidos manualmente pelos operadores, de acordo com determinações do centro de operação do sistema, COS, e outro com *setpoints* definidos automaticamente pelo CAG. Neste caso, as variações de potência gerada são muito frequentes.

Como não há nenhum dado disponível no sistema PIMS a respeito do modo de operação, além da própria potência ativa da máquina, surge a necessidade de extrair a informação sobre o modo de operação computacionalmente. Uma das formas de realizar isso é através da estimação da variância da potência ativa. Com *setpoints* fixos, a variância é muito menor. Além disso, é interessante que essa estimativa seja feita online e com baixo custo computacional. Propõe-se, então, a utilização de um estimador recursivo da variância. Por inspeção visual da estimação da variância de diferentes trechos de potência ativa é possível definir um limiar de variância que diferencia os modos de operação.

### 4.2.1 Média Aritmética Recursiva

A média aritmética e a variância de  $m$  observações de uma variável  $x_k$ , são dadas por:

$$\bar{x}_m = \frac{1}{m} \sum_{k=1}^m x_k \quad (4.4)$$

$$s_{\bar{x}_m}^2 = \frac{1}{m} \sum_{k=1}^m x_k^2 - \bar{x}_m^2 \quad (4.5)$$

O cálculo recursivo da média aritmética é deduzido a seguir:

$$\begin{aligned}
\bar{x}_1 &= \frac{x_1}{1}, \\
\bar{x}_2 &= \frac{x_1 + x_2}{2}, \\
2\bar{x}_2 &= x_1 + x_2, \\
\bar{x}_3 &= \frac{2\bar{x}_2 + x_3}{3}, \\
\bar{x}_m &= \frac{(m-1)\bar{x}_{m-1} + x_m}{m}, \\
\bar{x}_m &= \left(1 - \frac{1}{m}\right)\bar{x}_{m-1} + \frac{1}{m}x_m.
\end{aligned} \tag{4.6}$$

#### 4.2.2 Filtro de Primeira Ordem

A equação 4.6 assemelha-se ao cálculo de um filtro de primeira ordem, dado por:

$$y(k) = \beta y(k-1) + \alpha u(k) \tag{4.7}$$

O coeficiente  $\alpha$  é determinado pela relação  $\alpha = 1 - e^{-N}$ , onde  $N$  é o número de amostras em uma constante de tempo do filtro. O coeficiente  $\beta$  é determinado pela relação  $\beta = (1 - \alpha)$ .

A equação 4.7 possui coeficientes constantes, enquanto a equação 4.6 possui coeficientes que variam, tendendo a reduzir o peso da entrada no valor da média. A variância do valor estimado para a média pelo filtro de primeira ordem é calculada a seguir, para  $u(k)$  independentes e identicamente distribuídas e valor esperado  $E[u(k)] = 0$ :

$$\begin{aligned}
s_{y(k)}^2 &= E[y^2(k)] = E[(\beta y(k-1) + \alpha u(k))^2], \\
s_{y(k)}^2 &= \beta^2 E[y^2(k-1)] + 2\beta\alpha E[y(k-1)u(k)] + \alpha^2 E[u^2(k)], \\
s_{y(k)}^2 &= \beta^2 s_{y(k)}^2 + \alpha^2 s_{u(k)}^2, \\
s_{y(k)}^2 &= \frac{\alpha^2}{1 - \beta^2} s_{u(k)}^2 = \frac{\alpha}{2 - \alpha} s_{u(k)}^2.
\end{aligned} \tag{4.8}$$

Igualando as variâncias dos dois filtros, equações 4.5 e 4.8, chega-se a seguinte relação entre  $m$  e  $\alpha$ :

$$\alpha = \frac{2}{m+1} \tag{4.9}$$

Dessa forma, é possível implementar um filtro de 1ª ordem exponencial para estimar a média recursivamente com a mesma variância da média aritmética.

### 4.2.3 Estimador Recursivo da Variância

A equação da variância amostral pode ser expandida como segue:

$$\begin{aligned}
 s_{x_k}^2 &= \frac{1}{m-1} \sum_{k=1}^m (x_k - \bar{x}_m)^2, \\
 s_{x_k}^2 &= \frac{1}{m-1} \left[ \sum_{k=1}^m x_k^2 - \sum_{k=1}^m 2x_k \bar{x}_m + \sum_{k=1}^m \bar{x}_m^2 \right] \\
 s_{x_k}^2 &= \frac{m}{m-1} \left[ \left( \frac{1}{m} \sum_{k=1}^m x_k^2 \right) - 2\bar{x}_m \left( \frac{1}{m} \sum_{k=1}^m x_k \right) + \bar{x}_m^2 \right], \\
 s_{x_k}^2 &= \frac{m}{m-1} \left[ \overline{x_m^2} - (\bar{x}_m)^2 \right]^2. \tag{4.10}
 \end{aligned}$$

Estimando os valores de  $\bar{x}_m$  e  $\overline{x_m^2}$  com filtros de primeira ordem, tem-se:

$$\hat{x}(k) = (1 - \alpha)\hat{x}(k-1) + \alpha x(k) \tag{4.11}$$

$$\widehat{x^2}(k) = (1 - \alpha)\widehat{x^2}(k-1) + \alpha x^2(k) \tag{4.12}$$

Utilizando as Equações 4.10, 4.11 e 4.12, obtêm-se um estimador recursivo para a variância de  $x(k)$ :

$$s_{x_k}^2 = \frac{2 - \alpha}{2(1 - \alpha)} \left[ \widehat{x^2}(k) - (\hat{x}(k))^2 \right]^2 \tag{4.13}$$

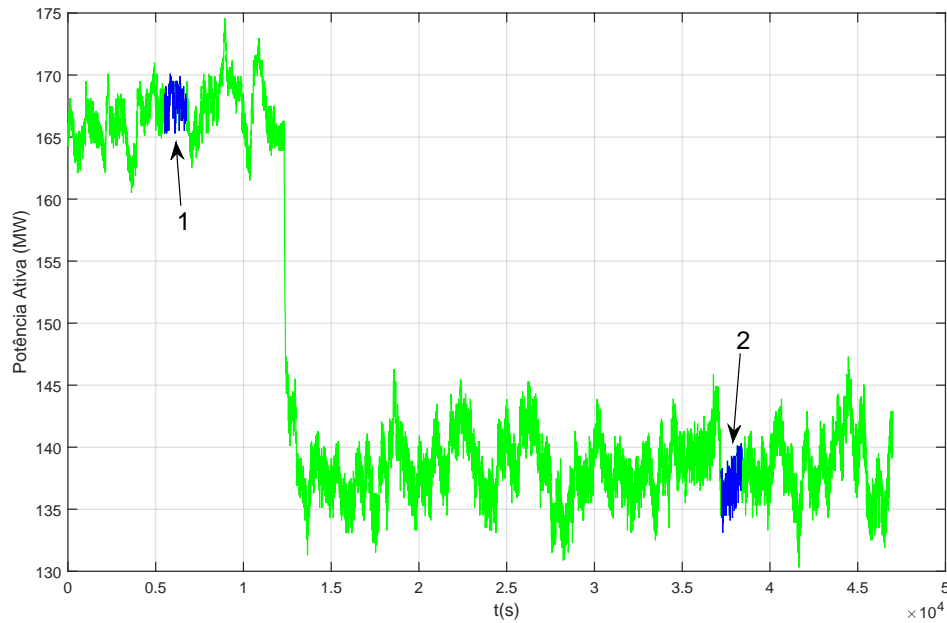
## 4.3 Fase de Projeto

O primeiro passo do projeto é a definição dos patamares de operação. Analisa-se os dados históricos disponíveis e define-se quais são os valores de potência ativa típicos para cada modo de operação. Essa seleção dos trechos para o projeto leva em consideração a experiência dos operadores e engenheiros da usina.

Para o modo de operação manual, ou seja, com *setpoints* estáveis de potência ativa, foram definidos dois trechos, destacados na Figura 12, extraídos de um histórico de dois dias. O primeiro trecho corresponde a uma operação típica em torno de 168 MW, enquanto o segundo trecho corresponde a uma operação em torno de 138 MW.

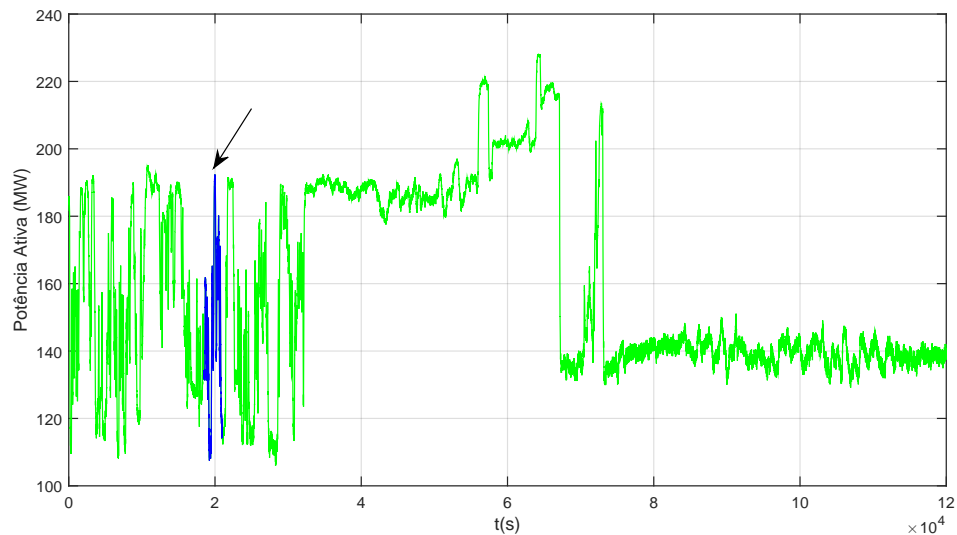
Para o modo de operação via CAG, ou seja, com *setpoints* variáveis de potência ativa, foi definido um trecho, destacado na Figura 13. Este corresponde a uma operação típica com variações de potência ativa entre 110 e 190MW.

Figura 12: Potência ativa em trecho de setembro de 2015.



Fonte: elaborada pelo autor.

Figura 13: Potência ativa em trecho de setembro de 2015, de operação via CAG.



Fonte: elaborada pelo autor.

### 4.3.1 Média adaptativa para projeto

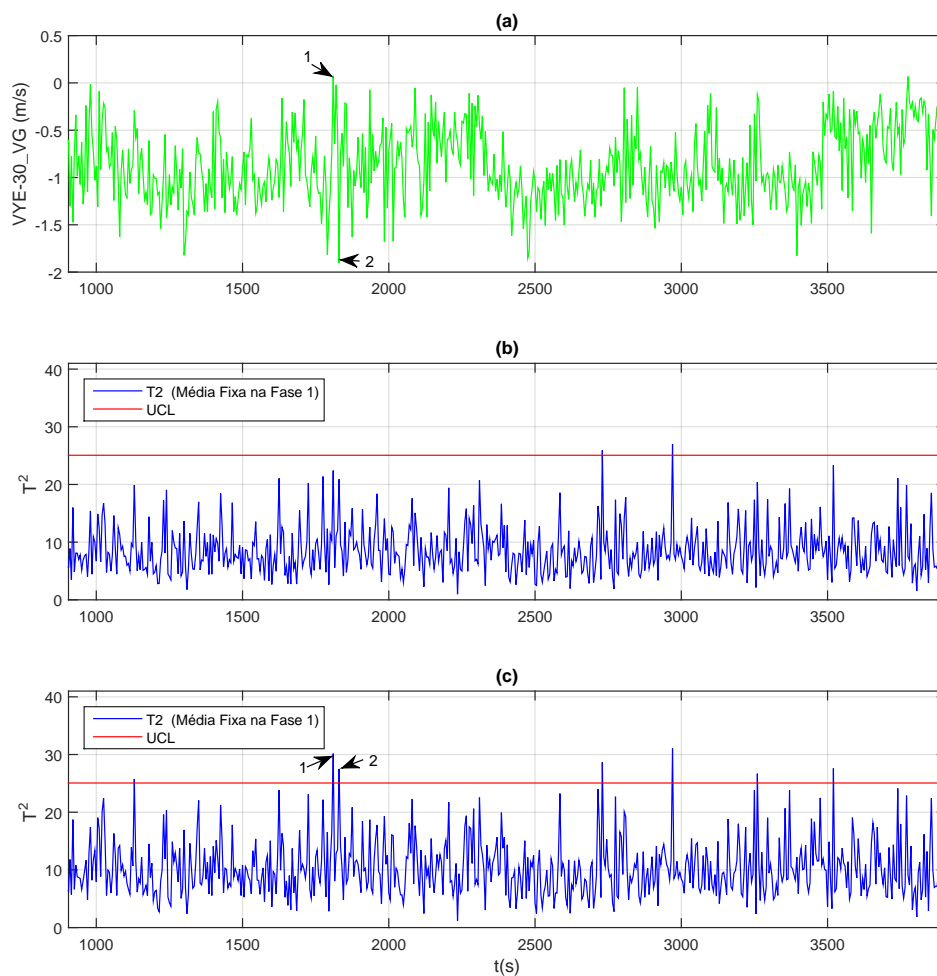
Processos industriais frequentemente apresentam pequenas variações na média das variáveis, mesmo que o *setpoint* de uma variável de referência, como a potência ativa para um gerador hidrelétrico, esteja fixo. Conseqüentemente, ao definir trechos de referência para o projeto de cartas de controle estatístico, é possível que pequenas variações na

média sejam tratadas como parte da variabilidade comum das variáveis, reduzindo, assim, a sensibilidade do monitoramento estatístico.

A alternativa proposta para tratar esse problema é a utilização de média móvel na normalização das variáveis antes da aplicação do PCA. Dessa forma, a fase de projeto leva em consideração pequenos desvios na normalidade das variáveis, aumentando a sensibilidade das cartas na etapa de monitoramento.

A Figura 14 ilustra o resultado da carta de Hotelling na fase de monitoramento para dois projetos distintos, um com média fixa e outro com média móvel. Dois instantes de variações abruptas na vibração absoluta do estator em determinado eixo apenas são alarmados pela carta de Hotelling correspondente ao projeto com média móvel. Além disso, a carta proveniente do projeto com média fixa possui menos pontos acima do limite de controle do que o previsto para um intervalo de confiança de 99%.

Figura 14: Comparação entre o resultado do monitoramento com média fixa na fase de projeto (b) e média móvel na fase de projeto (c). Velocidade da vibração no Estator (a).



Fonte: elaborada pelo autor.

### 4.3.2 Normalidade das Variáveis

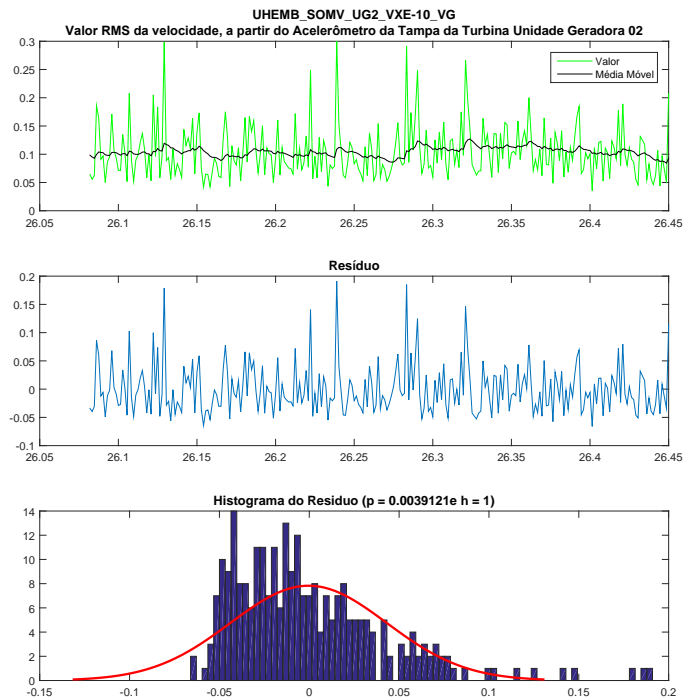
As técnicas estatísticas descritas no capítulo 2 assumem que as variáveis possuem distribuição de probabilidade normal. Na seção 4.3.1, apresentou-se uma forma de reduzir o efeito de pequenos comportamentos não-normais na fase de projeto através do uso de média móvel. Nem sempre isso é suficiente. Algumas variáveis, como o valor RMS da velocidade de vibração proveniente dos acelerômetros, podem apresentar comportamentos não-normais mais significativos. Para contornar este problema pode-se recorrer a transformações de variáveis, mudando assim a distribuição de probabilidade de forma a aproximar-se de uma distribuição normal.

No caso das variáveis relativas aos acelerômetros, utilizou-se o logaritmo natural dos dados adquiridos. As Figuras 15 e 16 ilustram os resultados dessa transformação no histograma dos resíduos de um dos acelerômetros. Para o mesmo parâmetro de cálculo de média móvel, o resíduo da vibração da tampa da turbina foi notavelmente não-normal. Realizando o teste de Kolmogorov-Smirnov, chega-se a um valor de  $p = 0,0039$ , indicando uma forte rejeição da hipótese nula de distribuição normal dos dados.

Utilizando a transformação por logaritmo natural, o histograma do resíduo da vibração deste acelerômetro aproxima-se muito mais de uma distribuição normal. Além disso, o teste de Kolmogorov-Smirnov aponta para a não-rejeição da hipótese nula de normalidade, sendo, assim, mais apropriado para a aplicação das técnicas de controle estatístico multivariado descritas anteriormente.

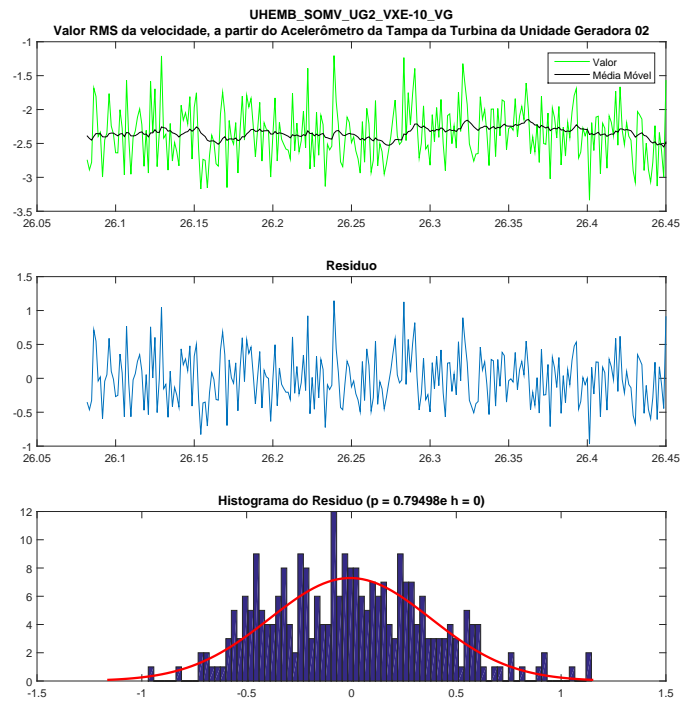


Figura 15: Histogramas do resíduo da velocidade da vibração na amostra original



Fonte: elaborada pelo autor.

Figura 16: Histogramas do resíduo do logaritmo natural da velocidade da vibração



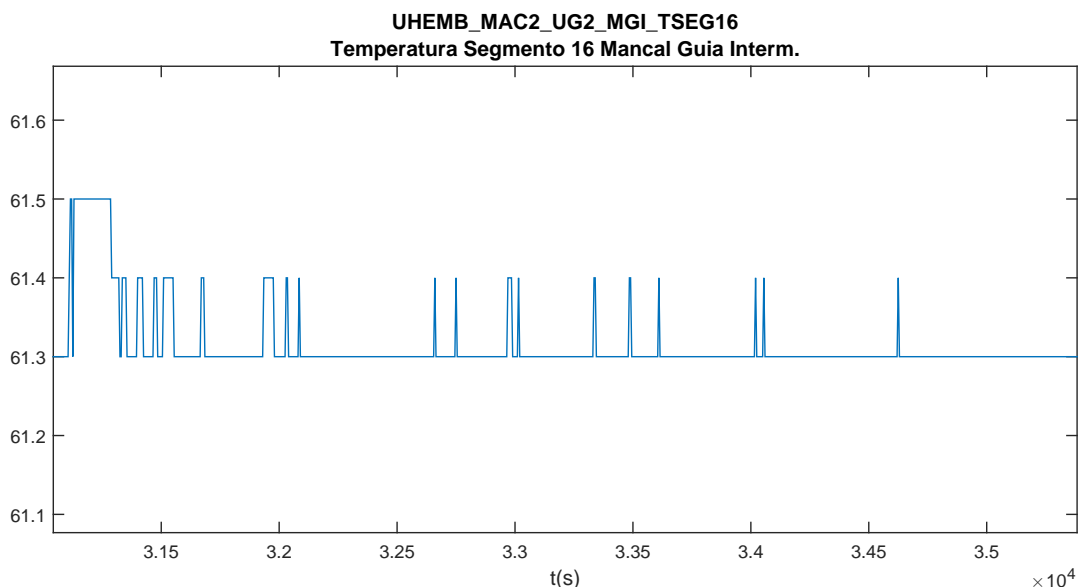
Fonte: elaborada pelo autor.

### 4.3.3 Implementação do PCA

Os dados históricos foram obtidos através do sistema PIMS instalado na usina. Como detalhado na descrição da arquitetura de automação, os dados da rede digital são coletados via protocolo OPC, diretamente dos CLPs. Por decisão da equipe de operação da usina, os dados das medições de temperatura estão representados na memória do CLP com resolução de apenas uma casa decimal, apesar de que as conversões analógico-digital dos cartões de entradas do CLP permitiriam resoluções maiores.

Essa particularidade faz com que as medições das temperaturas, da forma com que estão disponíveis no sistema PIMS, passem longos períodos com valores constantes, sem apresentar nenhuma variabilidade, uma vez que são variáveis de dinâmica lenta e com ruídos com oscilações típicas menores que  $0,1^{\circ}\text{C}$ . A Figura 17 ilustra essa característica.

Figura 17: Temperatura da Sapata nº 16 do Mancal de Escora do Gerador 3



Fonte: elaborada pelo autor.

Como todas as temperaturas variam, em estados de operação mais estável, entre poucos valores possíveis, a distribuição de probabilidade desses dados é muito distante da normal. Considerando que a usina já dispõe de alarmes para valores altos individuais de cada sensor de temperatura e pela inadequação desses dados para as técnicas utilizadas na metodologia descrita nesse capítulo, optou-se por omitir as variáveis de temperatura da análise. A base de dados resultante está apresentada na Tabela 2.

Selecionado o trecho de interesse dos dados, o cálculo do PCA foi realizado seguindo as equações da seção 2.4. Foram implementadas duas funções no Matlab: *parametrosPCA.m* e *calculaScores.m*.

Tabela 2: Base de Dados para o PCA

TAG	Descrição
VXE-10_VG	Vibração (Valor RMS da velocidade) do Acelerômetro da Tampa da Turbina
VXE-30_VG	Vibração (Valor RMS da velocidade) do Acelerômetro em 0º do Gerador
VYE-30_VG	Vibração (Valor RMS da velocidade) do Acelerômetro em 120º do Gerador
VYE-31_VG	Vibração (Valor RMS da velocidade) do Acelerômetro em 240º do Gerador
ZXE-10_VG	Deslocamento Pico-a-Pico do Proxímetro X da Turbina
ZXE-20_VG	Deslocamento Pico-a-Pico do Proxímetro X do Mancal Intermediário
ZXE-30_VG	Deslocamento Pico-a-Pico do Proxímetro X do Gerador
ZYE-10_VG	Deslocamento Pico-a-Pico do Proxímetro Y Turbina
ZYE-20_VG	Deslocamento Pico-a-Pico do Proxímetro Y do Mancal Intermediário
ZYE-30_VG	Deslocamento Pico-a-Pico do Proxímetro Y do Gerador
A_AZ	Corrente de Saída na Fase AZ
A_BR	Corrente de Saída na Fase BR
A_VM	Corrente de Saída na Fase VM
FREQ	Frequência do Gerador
MVAR	Potência Reativa
MW	Potência Ativa
RAD_FLUX	Fluxo de Água nos Radiadores

- *parametrosPCA.m*: recebe como parâmetro uma matriz de amostras  $X$ , cujas linhas são as amostras e as colunas são as variáveis. A função retorna uma estrutura de dados contendo:
  - a matriz de fatores de correlação, calculada pela Equação (2.24)
  - o vetor de médias amostral, calculado pela Equação (2.20)
  - o vetor de variâncias amostral, calculado pela Equação (2.21)
  - os autovetores da matriz de fatores de correlação
  - os autovalores da matriz de fatores de correlação
- *calculaScores.m*: recebe como parâmetros a estrutura de dados retornada pela função *parametrosPCA.m* e um conjunto de amostras. A função retorna a matriz de *scores* para o conjunto de amostras de entrada.

#### 4.3.4 Implementação da carta $T^2$ de Hotelling e do SPE

As cartas  $T^2$  de Hotelling foram calculadas sobre os *scores* das componentes principais retidas. Utilizando a Equação 2.32 foi definido o número de componentes necessárias para explicar acima de 90% da variabilidade total dos dados. Utilizando essas componen-

tes nos mesmos intervalos de cálculo do PCA, obteve-se o Índice  $T^2$  para cada amostra. Foram implementadas duas funções no Matlab: *parametrosT2.m* e *calculaIndiceT2.m*:

- *parametrosT2.m*: recebe como parâmetro uma matriz de scores, cujas linhas são as amostras e as colunas são os scores. A função retorna uma estrutura de dados contendo:
  - a matriz de covariância amostral, calculada pela Equação (2.22)
  - o vetor de médias amostral
  - o vetor de variâncias amostral
- *calculaIndiceT2.m*: recebe como parâmetros a estrutura de dados retornada pela função *parametrosT2.m* e um conjunto de scores. A função retorna o vetor de índices  $T^2$  para o conjunto de amostras de entrada.

Após a obtenção dos índices, calcula-se o limite LSC da Fase 1 pela Equação (2.18).

O cálculo do SPE, do índice combinado e dos limites correspondentes foi realizado via implementação direta das equações apresentadas no Capítulo 2.

## 4.4 Fase de Monitoramento

Definidos os parâmetros para cálculo dos *scores*, o número de componentes retidas de cada patamar e os limites dos índices, a fase de monitoramento consiste numa simulação do monitoramento *online*, feita com dados históricos da planta em trechos diferentes dos de projeto para validação da capacidade de detecção de mudanças do sistema.

### 4.4.1 Monitoramento Eletrônico com Média Adaptativa

O procedimento geral para o monitoramento foi definido como (SOUSA et al., 2015):

- Identificar os pontos de operação típicos;
- Realizar o cálculo dos coeficientes das componentes principais para cada ponto de operação;
- Definir o número de componentes a serem utilizadas para cada ponto de operação;
- Calcular a matriz de covariâncias dos scores e os limites de controle para cada ponto de operação;

- Na fase de monitoramento, verificar continuamente o comando de potência da operação do gerador. Identificar o ponto de operação mais próximo e chavear os parâmetros de cálculo dos scores e dos índices.

Caso a usina opere em um ponto distinto daqueles identificados antes do monitoramento, as diferenças nas médias das variáveis, mesmo que pequenas, poderiam causar grandes desvios no resultado do índice  $T^2$ , uma vez que trata-se de uma métrica multivariada que acumula os erros na estimativa de cada variável. Essas técnicas de estatística multivariada pressupõem estacionariedade, condição muitas vezes difícil de ser alcançada na operação de um processo em que as variáveis são dinâmicas e onde ocorrem transições entre diferentes pontos de operação. É preciso, então, buscar uma alternativa de análise em estado quase estacionário.

Dessa forma, é interessante que se calcule, a cada período de amostragem, as médias atuais das variáveis, utilizando o filtro de média da Equação 4.7. No entanto, o ajuste do coeficiente  $\alpha$  não é trivial. Se o filtro for rápido demais, a média irá oscilar muito, causando desvios indesejados no cálculo do PCA. Se a média for muito lenta, quando ocorrer uma transição entre dois pontos de operação, o cálculo da média pode demorar a estabilizar, causando a necessidade de silenciar os alarmes do monitoramento por períodos muito longos.

Para solucionar esse problema, propõe-se o uso de um cálculo de média adaptativa com teste CUSUM (SOUSA et al., 2015), detalhado na seção a seguir.

#### 4.4.2 Teste CUSUM

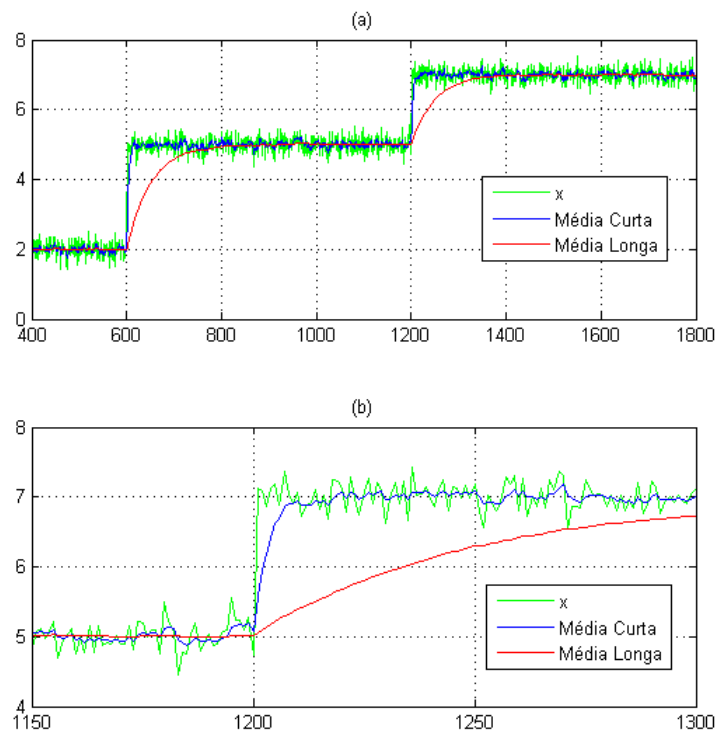
O algoritmo consiste em calcular, inicialmente, duas médias utilizando o filtro recursivo da Equação 4.7. Um dos filtros, denominado “Filtro de média de curta duração” é calculado com um valor de  $M$  pequeno, ou seja, é um filtro rápido que acompanha melhor os desvios da variável. O outro filtro, denominado “Filtro de média de longa duração” é calculado com um valor de  $M$  grande, ou seja, é um filtro lento que resulta numa média com variações suaves.

A saída do filtro longo é mais adequada para ser utilizada como média para a normalização das variáveis, no entanto, na ocorrência de uma mudança de ponto de operação, o filtro demora a estabilizar no novo patamar, como é mostrado na Figura 18:

O teste CUSUM, a cada amostragem, consiste em:

- Calcular a média curta e a média longa no instante atual.
- Calcular o resíduo, ou seja, a diferença entre as duas médias.

Figura 18: Filtros de média sem o teste CUSUM

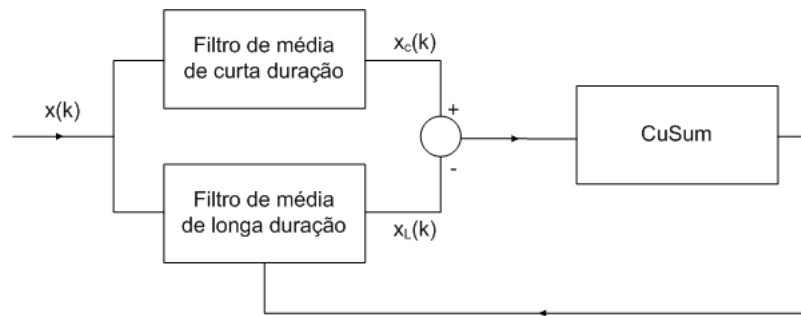


Fonte: elaborada pelo autor.

- Se o resíduo for positivo acumula-se o seu valor numa variável auxiliar  $C^+$ , caso o contrário acumula-se o seu valor numa variável  $C^-$ .
  - Se  $C^+$  estiver acima de um limiar pré-estabelecido, ou *Threshold*, gera-se um alarme positivo e  $C^+$  é zerado.
  - Se  $C^-$  estiver abaixo de um limiar pré-estabelecido, ou *Threshold*, gera-se um alarme negativo e  $C^-$  é zerado.
- Se o resíduo for positivo,  $C^-$  é zerado, caso o contrário,  $C^+$  é zerado.
- Se nas últimas  $k_{MAX}$  amostras, ocorrerem mais do que  $n_A$  alarmes positivos ou mais do que  $n_A$  alarmes negativos, assume-se que ocorreu uma mudança de patamar. Atribui-se, então, o valor atual da média curta à média longa.

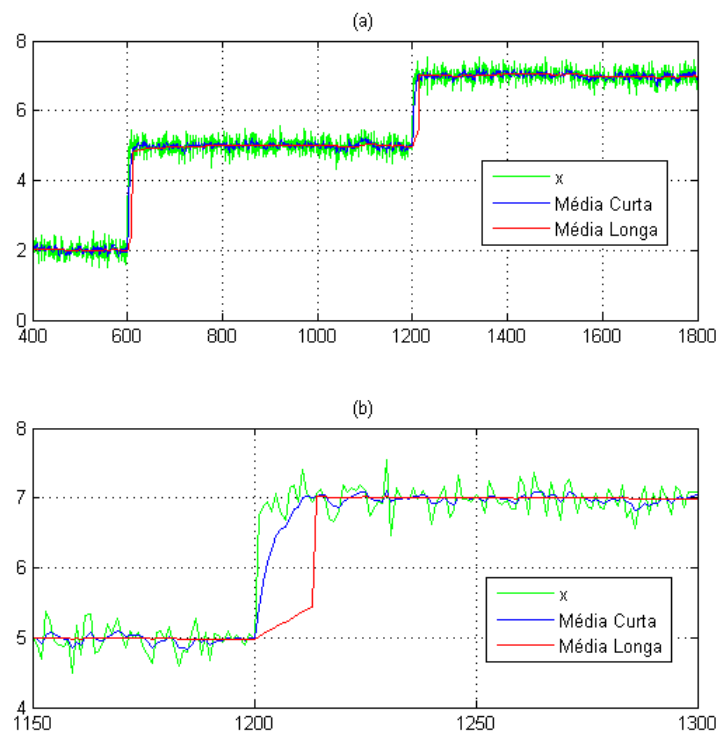
Os parâmetros  $k_{MAX}$ , *Threshold*,  $n_A$  e os fatores de esquecimentos dos filtros para cada variável foram projetados experimentalmente, por tentativa-e-erro. As Figuras 19 e 20 ilustram um diagrama do cálculo adaptativo da média e um exemplo de resultado para a mesma variável da figura anterior, respectivamente. Esse procedimento permite a adaptação rápida a novos patamares, mantendo a média estável em cada patamar.

Figura 19: Diagrama do cálculo adaptativo da média



Fonte: elaborada pelo autor.

Figura 20: Filtros de média com o teste CUSUM



Fonte: elaborada pelo autor.

## 4.5 Comentários Finais

Este capítulo tratou da metodologia desenvolvida para monitoramento amplo usando técnicas de estatística multivariada, detalhando as implementações realizadas. O capítulo seguinte trata da apresentação e análise dos resultados obtidos, sobre dados de operação de uma unidade geradora hidrelétrica.

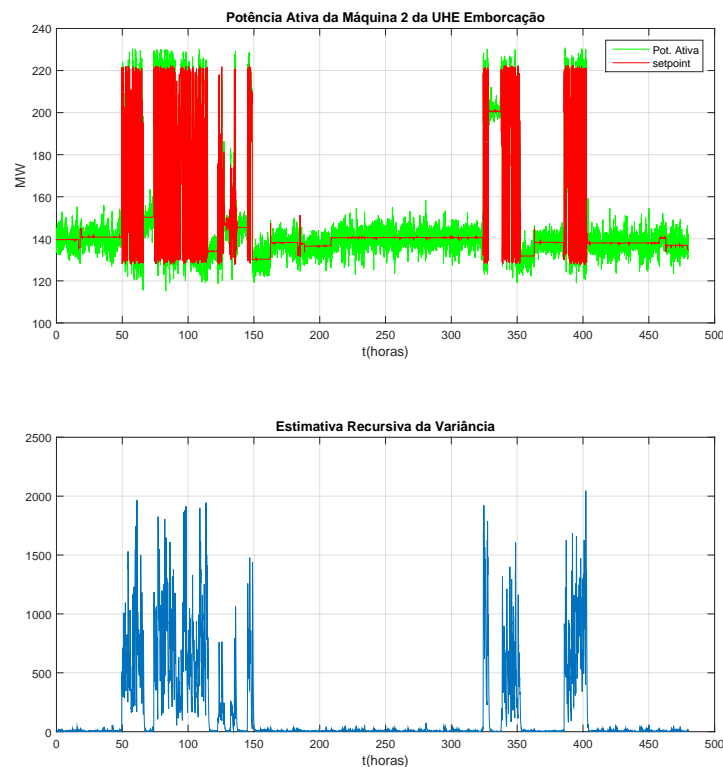
## 5 Resultados e Análises

Apresentam-se os resultados do cálculo computacional do modo de operação de uma unidade geradora da UHE Emborcação, da fase de projeto e de validação do monitoramento amplo de condição sobre dados históricos de diferentes cenários de operação da unidade.

### 5.1 Modo de Operação

Como descrito no capítulo anterior, a definição do modo de operação se deu pela estimação recursiva da variância da potência ativa da máquina. A Figura 21 mostra a variância estimada para um período de 20 dias. Observa-se que nos períodos de operação com *setpoint* definido pelo CAG, a variância estimada é significativamente maior do que na operação manual, com *setpoint* fixo.

Figura 21: Estimação Recursiva da Variância da Potência Ativa: Período de 20 dias



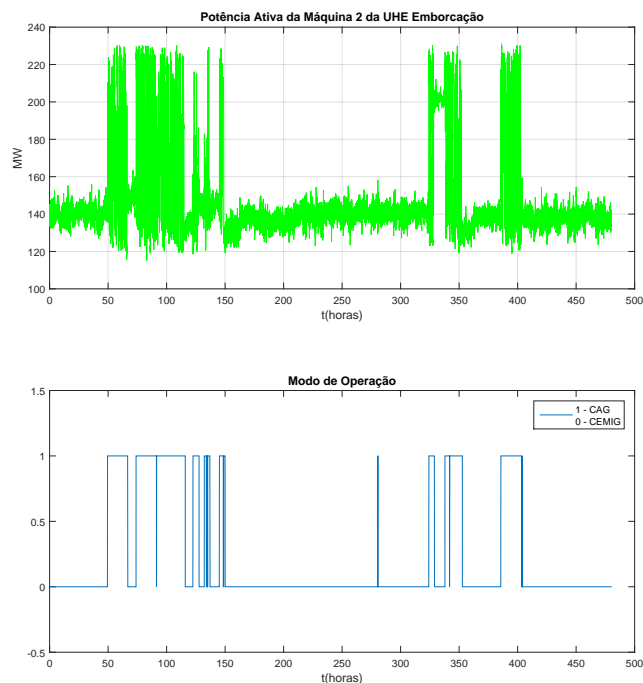
Fonte: elaborada pelo autor.

Estabelecendo, visualmente, um limite de variância acima do qual assume-se que o modo de operação é via CAG, tem-se um método para definição do modo de operação.



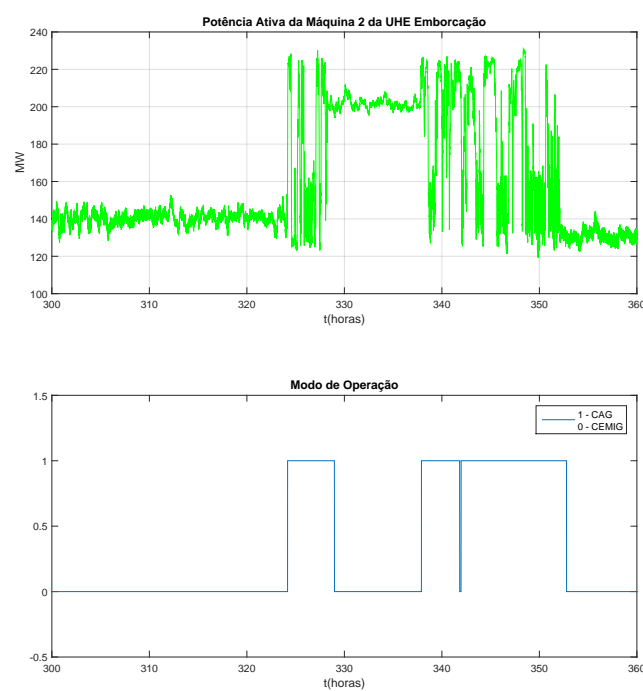
As Figuras 22 e 23 apresentam a potência ativa e o modo de operação para o mesmo período de 20 dias e para um recorte de 60 horas, respectivamente.

Figura 22: Definição de modo de operação: Período de 20 dias



Fonte: elaborada pelo autor.

Figura 23: Definição de modo de operação: Período de 60 horas



Fonte: elaborada pelo autor.

## 5.2 Cenários Operacionais

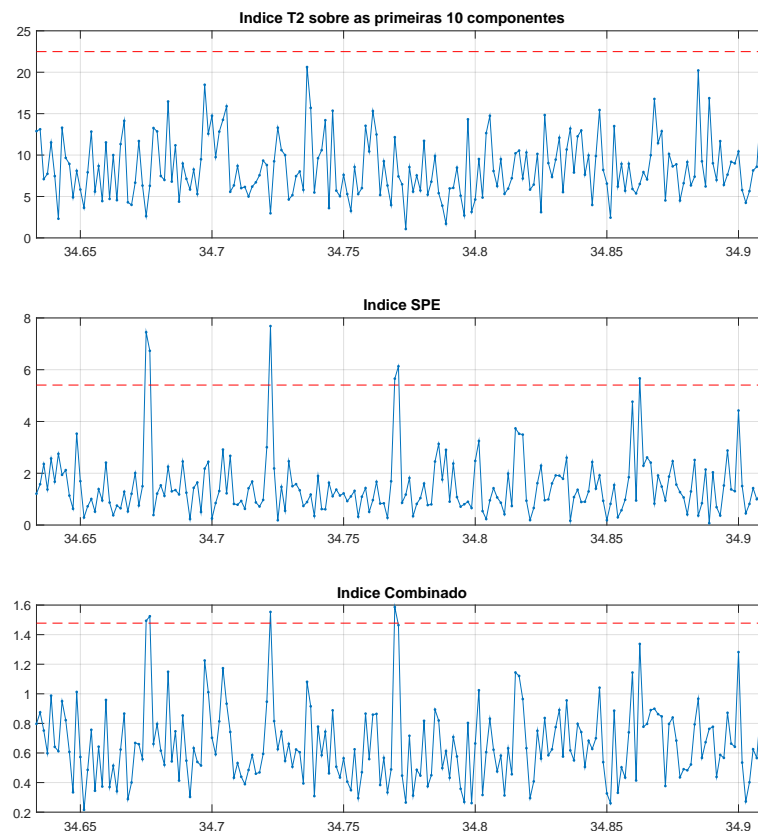
A seguir, apresentam-se os resultados do monitoramento amplo de condição para diferentes cenários operacionais da unidade geradora.

### 5.2.1 Operação Manual via COS

Este cenário corresponde a um longo período de definição manual da potência ativa gerada, quando manteve-se o *setpoint* fixo próximo de 138MW, por atuação de um operador na sala de controle da usina, de acordo com determinação do COS. O cálculo do PCA levou a uma redução da dimensão da base de dados, restando-se 10 componentes principais para cálculo dos índices estatísticos. Estimou-se que estas são suficientes para explicar cerca de 91% da variabilidade total dos dados.

Na fase de projeto, em algumas amostras o índice combinado ficou um pouco acima do limite, no entanto não foram encontradas causas atribuíveis. Nessas amostras, o índice SPE ultrapassou o limite, enquanto a carta de hotelling não gerou alarmes. Pelos modelos estatísticos utilizados, espera-se que alguns pontos fiquem acima do limite superior de controle.

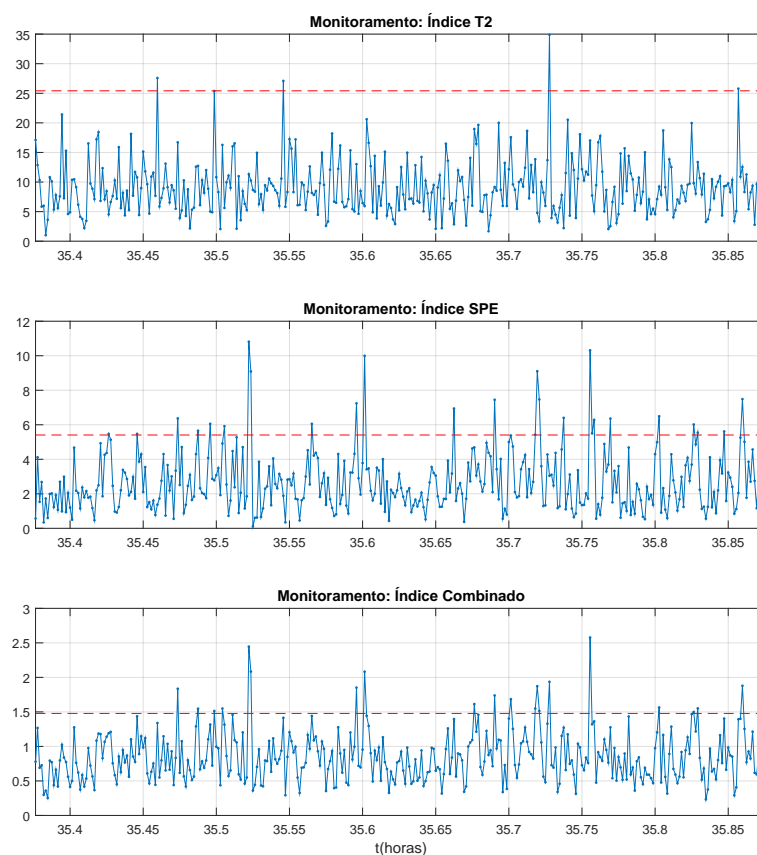
Figura 24: Operação Manual: Índices estatísticos na fase de projeto



Fonte: elaborada pelo autor.

A seguir apresenta-se o resultado do monitoramento para dois trechos distintos. No primeiro deles, os índices estatísticos não indicam mudanças significativas, apenas alguns pontos fora do limite de controle, compatíveis com o intervalo de confiança definido. Pequenos *spikes* na saída de alguns acelerômetros ocasionaram algumas amostras com índice  $T^2$  superior ao limite, observadas na Figura 25. SPE e índice combinado indicam situação operacional semelhante a da fase de projeto.

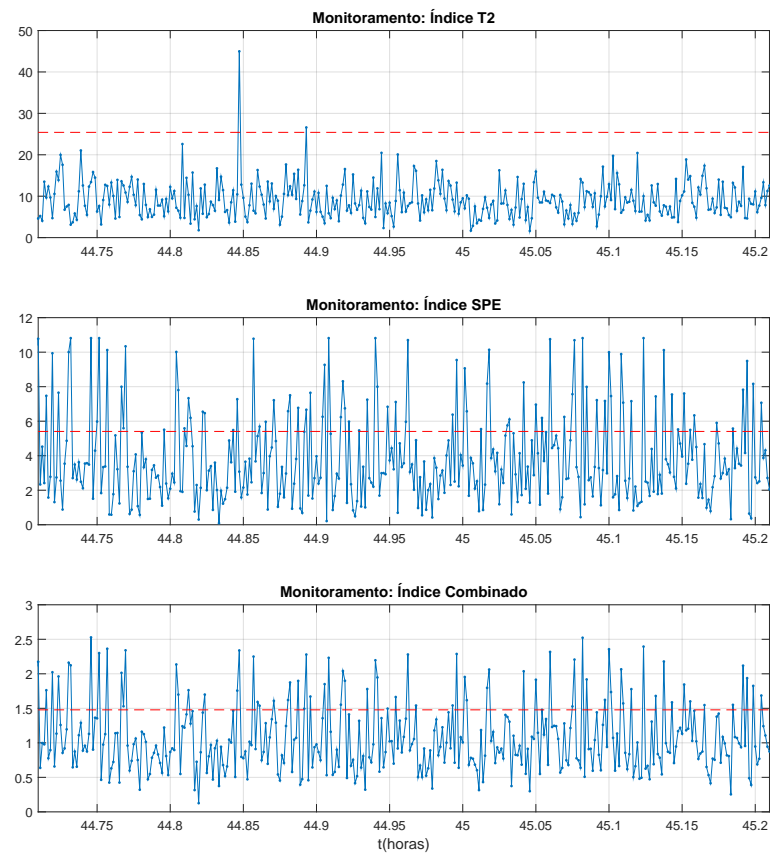
Figura 25: Operação Manual: Índices estatísticos na fase de monitoramento em trecho normal



Fonte: elaborada pelo autor.

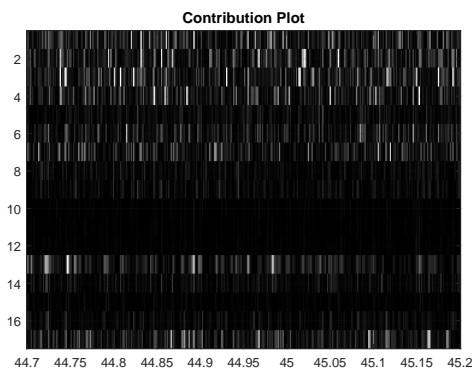
No segundo trecho de monitoramento, apresentado na Figura 26, o índice  $T^2$  foi calculado bem acima do limite em apenas uma amostra, também correspondente a um *spike* numa medição de vibração. Isso indicaria novamente uma situação normal de operação, bem representada pelo modelo estatístico. No entanto, o SPE e, conseqüentemente, o índice combinado foram calculados acima do limite em várias amostras, muito além do esperado. Dessa forma, considera-se que o modelo estatístico não é mais suficientemente representativo. Recorrendo ao *contribution plot*, Figura 27, observa-se que várias variáveis possuem contribuição significativa em diversas amostras, reforçando a necessidade de reprojeter as cartas de controle, uma vez que as variâncias e as correlações não correspondem mais aos valores calculados na fase de projeto.

Figura 26: Operação Manual: Índices estatísticos na fase de monitoramento em trecho com muitos alarmes no SPE



Fonte: elaborada pelo autor.

Figura 27: Operação Manual: *Contribution Plots* na fase de monitoramento em trecho com muitos alarmes no SPE



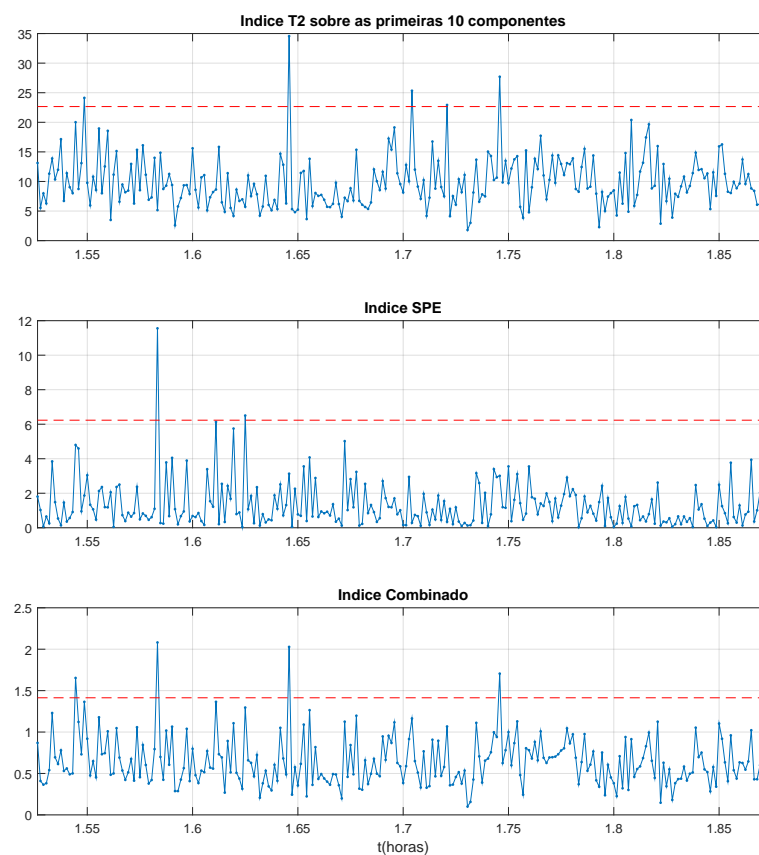
Fonte: elaborada pelo autor.

## 5.2.2 Operação Manual via COS com Mudança de Ponto de Operação

Este cenário corresponde aos dois dias de operação apresentados no capítulo anterior, na Figura 12. São dois patamares de operação manual, em torno de 168 e 138MW,

respectivamente. Pela necessidade de reprojeter as cartas para o patamar de 138MW, verificada no cenário anterior, foram escolhidos dois trechos, um para cada *setpoint* de potência gerada, em que considerou-se que o processo encontrava-se no melhor estado operacional. As Figuras 28 e 29 mostram as cartas de controle dos índices estatísticos no conjunto de amostras utilizadas na fase de projeto de cada patamar de potência. Assim como no cenário anterior, algumas amostras estão com índices um pouco acima do limites, no entanto sem causas atribuíveis identificadas e em número próximo ao esperado pelos intervalos de confiança utilizados.

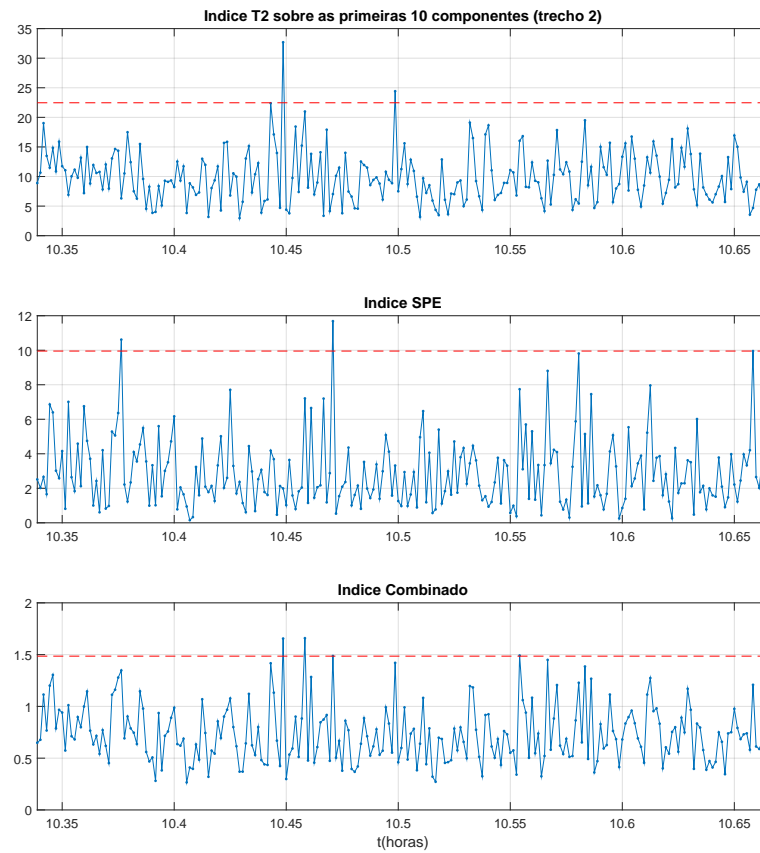
Figura 28: Operação Manual com dois patamares: Índices estatísticos na fase de projeto em 168 MW



Fonte: elaborada pelo autor.

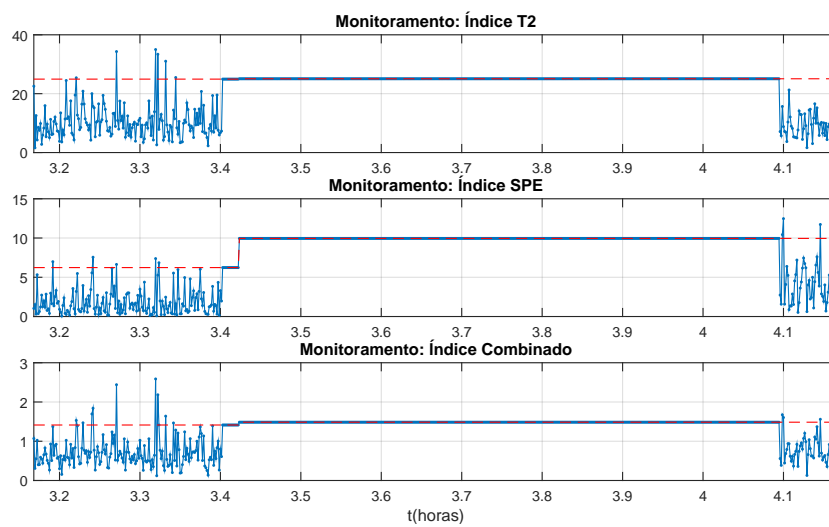
Seguindo a metodologia proposta, quando ocorre uma mudança de *setpoint* pelo operador, os alarmes devem ser silenciados até que as grandezas medidas acomodem-se e volte-se a um estado de controle estatístico. Esse procedimento é realizado saturando-se os índices nos limites correspondentes, como observado na Figura 30.

Figura 29: Operação Manual com dois patamares: Índices estatísticos na fase de projeto em 138 MW



Fonte: elaborada pelo autor.

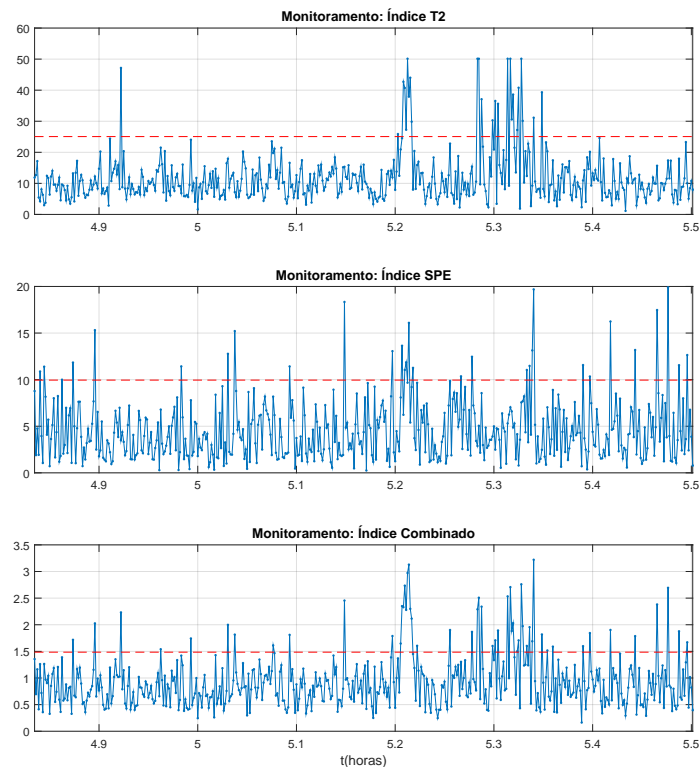
Figura 30: Operação Manual com dois patamares: Índices estatísticos na transição de patamar de operação na fase de monitoramento



Fonte: elaborada pelo autor.

A seguir, assim como para o primeiro cenário, apresenta-se o resultado do monitoramento para dois trechos distintos. No primeiro deles, Figura 31, verificam-se dois momentos com grande número de amostras com índices acima do limites. Em  $t = 5,2 \text{ horas}$ , o índice  $T^2$  mantém-se por algumas amostras acima do limite, o que também é verificado pelo SPE, porém com menos intensidade. Em  $t = 5,3 \text{ horas}$ , a variabilidade do índice  $T^2$  cresce consideravelmente, apresentando várias amostras acima do limite por alguns minutos.

Figura 31: Operação Manual com dois patamares: Índices estatísticos na fase de monitoramento (exemplo 1)



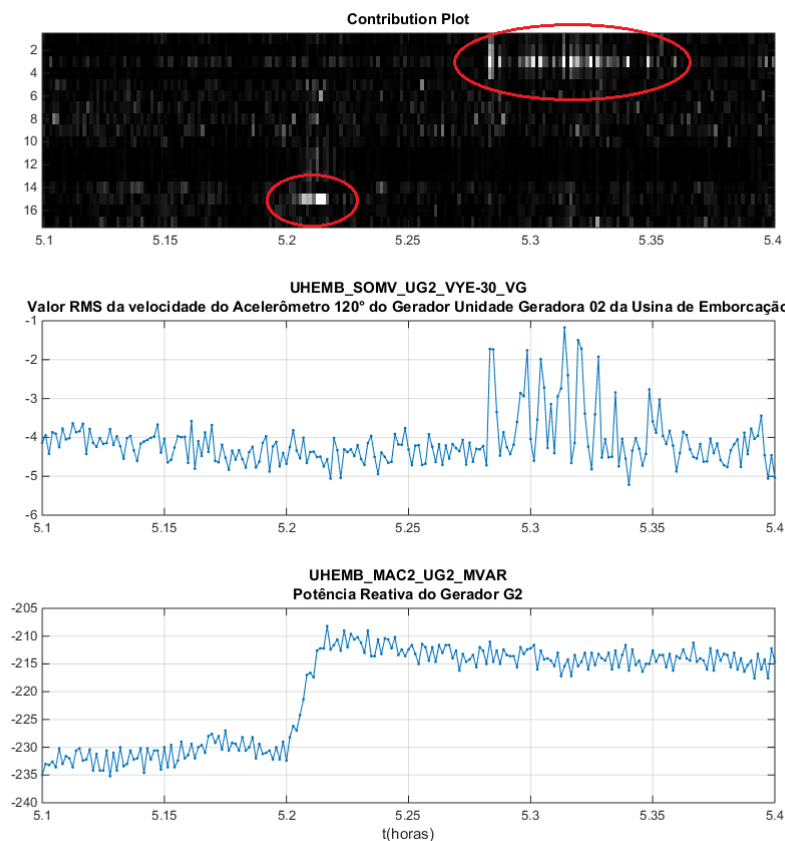
Fonte: elaborada pelo autor.

Recorrendo ao *contribution plot*, primeiro gráfico da Figura 32, verifica-se que próximo ao instante de tempo  $t = 5,2 \text{ horas}$ , a potência reativa é a variável com contribuição mais significativa na variabilidade dos dados. No terceiro gráfico da Figura 32, observa-se uma forte mudança na média da potência reativa nesse instante de tempo, por uma provável mudança no setpoint de potência reativa. O monitoramento foi eficiente para alarmar essa condição. O uso da média adaptativa ajusta o cálculo da normalização da potência reativa, fazendo com que em algumas amostras depois as cartas não alarmem mais. Como não se trata de uma falta, e sim de uma mudança normal dentro da operação da usina, este é o comportamento desejado do sistema de monitoramento.

Próximo ao instante de tempo  $t = 5,3 \text{ horas}$ , o *contribution plot* indica que a

terceira variável da base, uma das velocidades de vibração do estator, é responsável por grande parte da variabilidade dos dados. O segundo gráfico da Figura 32 mostra que essa variável apresentou um crescimento de variabilidade significativo nesse período, indicando alguma condição atípica de vibração no estator, onde o acelerômetro correspondente está instalado.

Figura 32: Operação Manual com dois patamares: *Contribution Plots* na fase de monitoramento (exemplo 1)

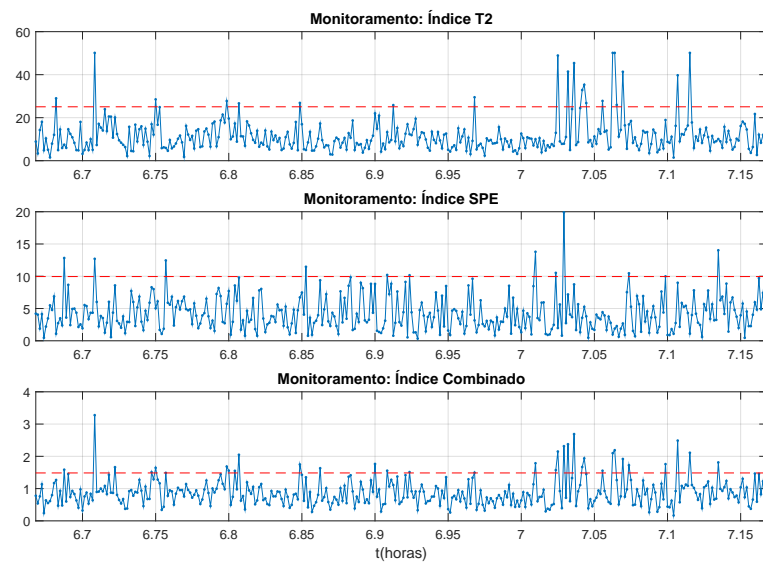


Fonte: elaborada pelo autor.

No segundo trecho de monitoramento, apresentado na Figura 33, observa-se uma amostra, próxima a  $t = 6,71$  horas, com alto valor de índice  $T^2$  e uma concentração de amostras, depois de  $t = 7,01$  horas, com valores dos três índices acima do limite. Pelo *contribution plot*, Figura 34, verifica-se a amostra em  $t = 6,71$  horas fortemente influenciada pelas quatro medições de vibração por acelerômetros. A região depois de  $t = 7,01$  horas também tem variabilidade causada em grande parte por essas variáveis, principalmente a terceira, a medição a partir do acelerômetro no eixo de 120 graus do gerador. O segundo e terceiro gráfico da Figura 34 mostram que essas condições atípicas de vibração são verificadas também diretamente no histórico dessas variáveis.

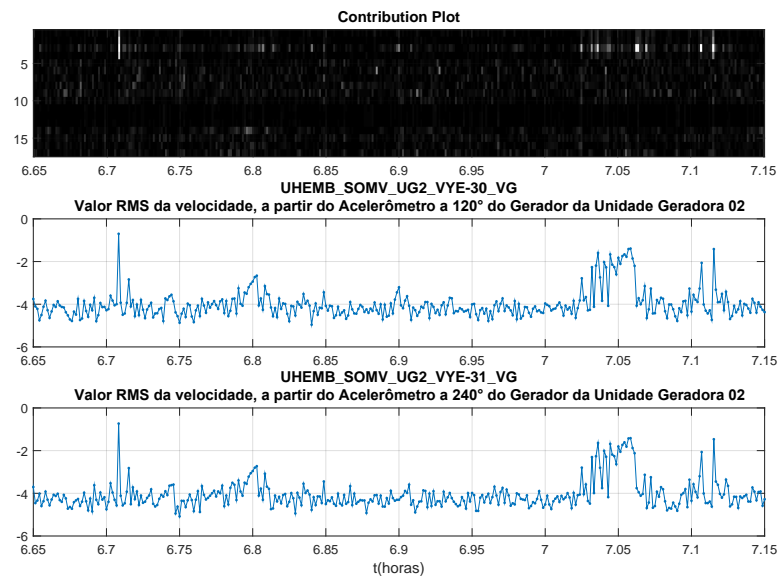


Figura 33: Operação Manual com dois patamares: Índices estatísticos na fase de monitoramento (exemplo 2)



Fonte: elaborada pelo autor.

Figura 34: Operação Manual com dois patamares: *Contribution Plots* na fase de monitoramento (exemplo 2)



Fonte: elaborada pelo autor.

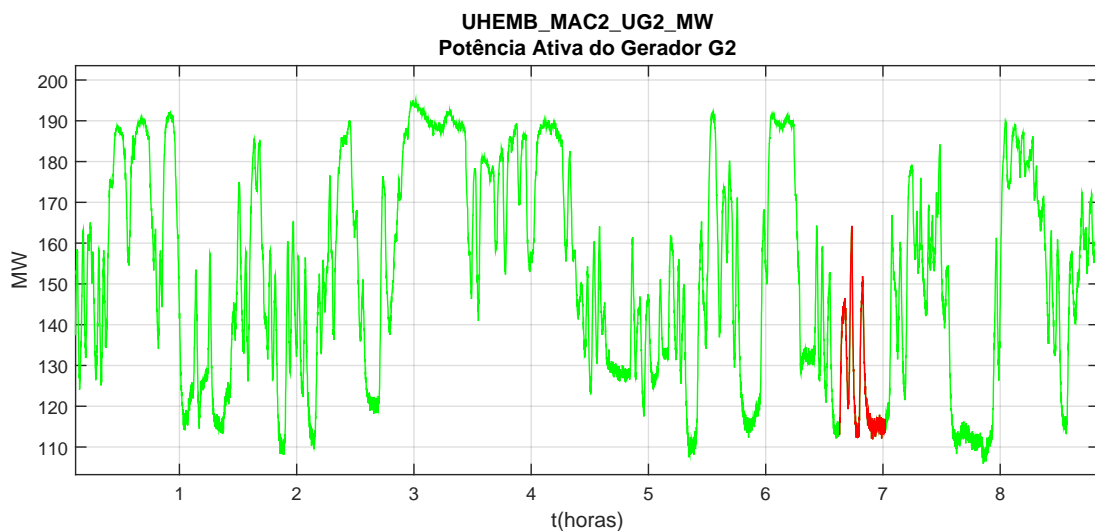
### 5.2.3 Operação via CAG

Este cenário corresponde a operação com *setpoint* de potência ativa definida pelo CAG. Analisando os dados brutos das variáveis para esse caso, verifica-se que os próximos tem saída com maior variabilidade a medida em que a potência gerada reduz. Esse comportamento pode ser explicado pelo fato do gerador ser projetado para ter uma potên-

cia ativa nominal de 298 MW. Uma geração desse porte só é possível com o reservatório da usina quase completamente cheio, quando há pressão suficiente no tubo de sucção e na caixa espiral da turbina para que essa potência seja alcançada. Como esse não é o caso há alguns anos, a máquina é operada com potência abaixo da nominal. Quanto menor for a potência que se espera gerar, menor será a pressão da água na turbina e, conseqüentemente, menor será o torque sobre o eixo da unidade, resultando em um desbalanceamento maior do mesmo. Esse efeito resulta em sinais de vibração relativa dos proxímetros com variabilidade maior.

Devido a esse fenômeno, foi escolhido um trecho para projeto das cartas de controle quando a potência ativa média é mais baixa, em torno de 125MW, como observado na Figura 35. Nesse trecho, a variabilidade dos proxímetros é mais alta, mesmo com um comportamento da unidade considerado normal. Dessa forma, evita-se que os parâmetros do modelo estatístico sejam calculados com variância dos proxímetros baixa, o que ocasionaria muitos alarmes falsos no monitoramento assim que a potência média estivesse abaixo do trecho de projeto.

Figura 35: Operação via CAG: Potência Ativa



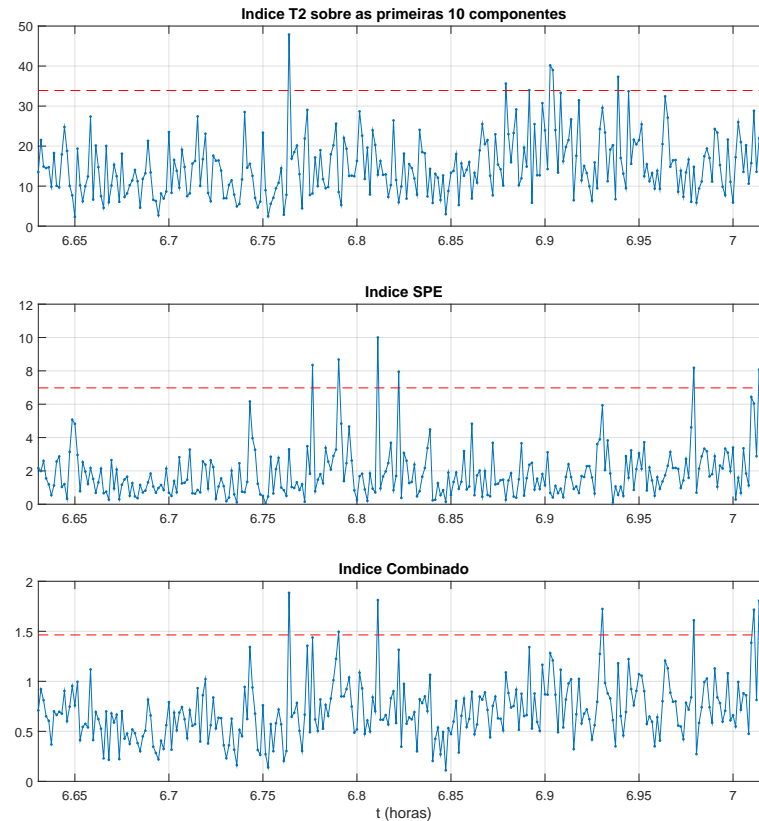
Fonte: elaborada pelo autor.

Outras modificações foram necessárias para realizar um monitoramento adequado nesse modo de operação. A potência ativa foi omitida da base de dados, uma vez que apresenta um comportamento bem distante de uma variável com distribuição normal. Além disso, os parâmetros dos filtros de média móvel da fase de projeto e da fase de monitoramento foram reajustados para cada variável.

A Figura 36 mostra as cartas de controle dos índices estatísticos para conjunto de amostras utilizadas na fase de projeto. Assim como nos cenários anteriores, algumas amostras estão com índices um pouco acima dos limites, no entanto sem causas atribuíveis

identificadas. Assim sendo, os dados responsáveis por estes alarmes foram mantidos, pois assume-se que correspondem a condições normais de operação.

Figura 36: Operação via CAG: Índices estatísticos na fase de projeto

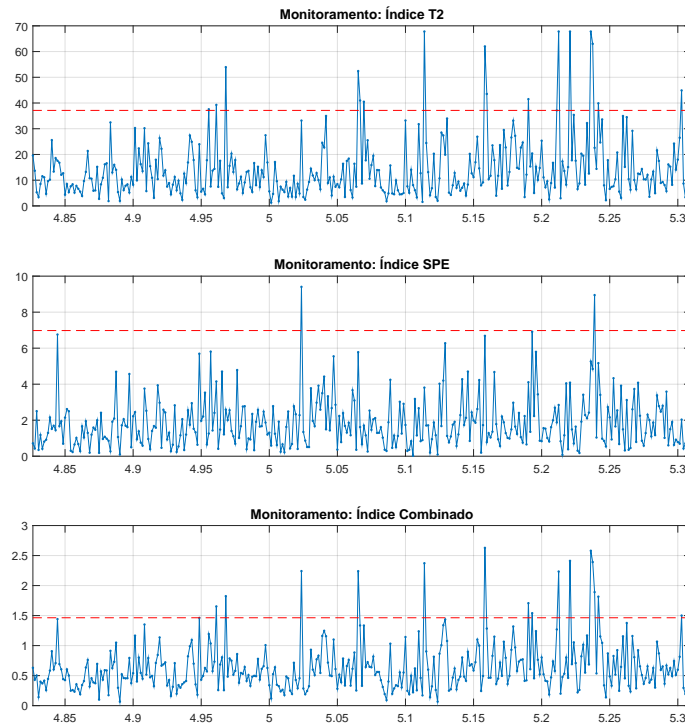


Fonte: elaborada pelo autor.

A Figura 37 corresponde a um trecho de monitoramento, no qual verifica-se um número significativo de amostras com índice  $T^2$  correspondente acima do limite da carta.

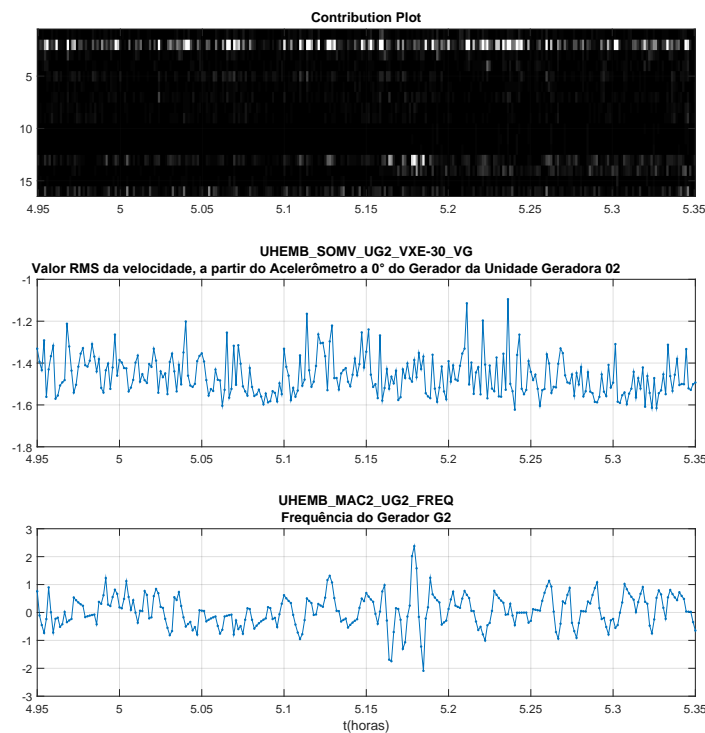
Pelo *contribution plot*, primeiro gráfico da Figura 38, observa-se que a segunda variável da base, uma das velocidades de vibração do estator, é responsável por grande parte da variabilidade dos dados. A frequência do gerador, décima terceira variável da nova base, possui contribuição significativa no período entre  $t = 5,15$  e  $t = 5,2$  horas. O resíduo da frequência, terceiro gráfico da Figura 38, passa por um período de maior oscilação, possivelmente causada por variações mais bruscas no setpoint de potência ativa, uma vez que a frequência do gerador é a variável de atuação para regular a potência ativa. Quanto a vibração, o segundo gráfico mostra diversos trechos de maior vibração absoluta, que ocasionaram a maior parte dos alarmes gerados pela carta de Hotelling nesse trecho.

Figura 37: Operação via CAG: Índices estatísticos na fase de monitoramento (exemplo 1)



Fonte: elaborada pelo autor.

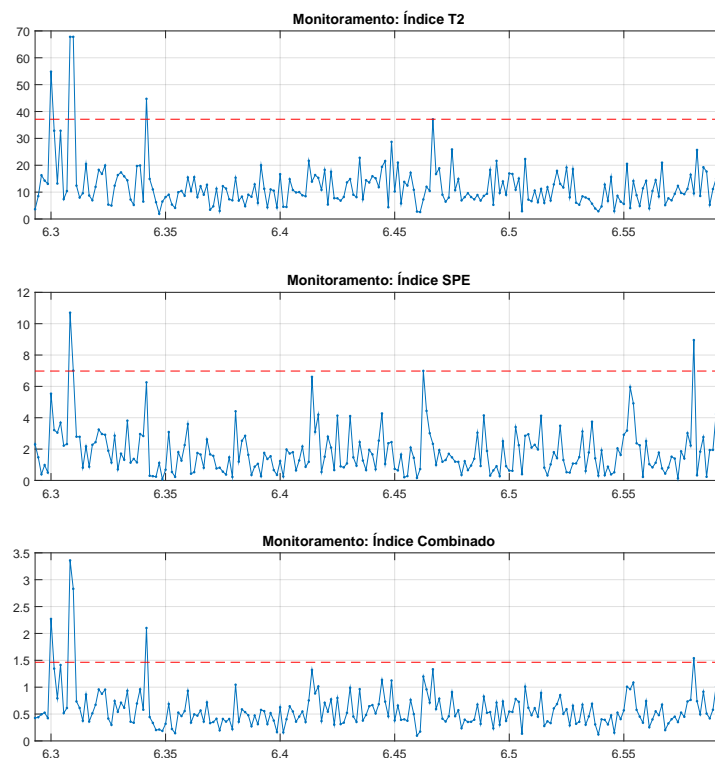
Figura 38: Operação via CAG: *Contribution Plots* na fase de monitoramento (exemplo 1)



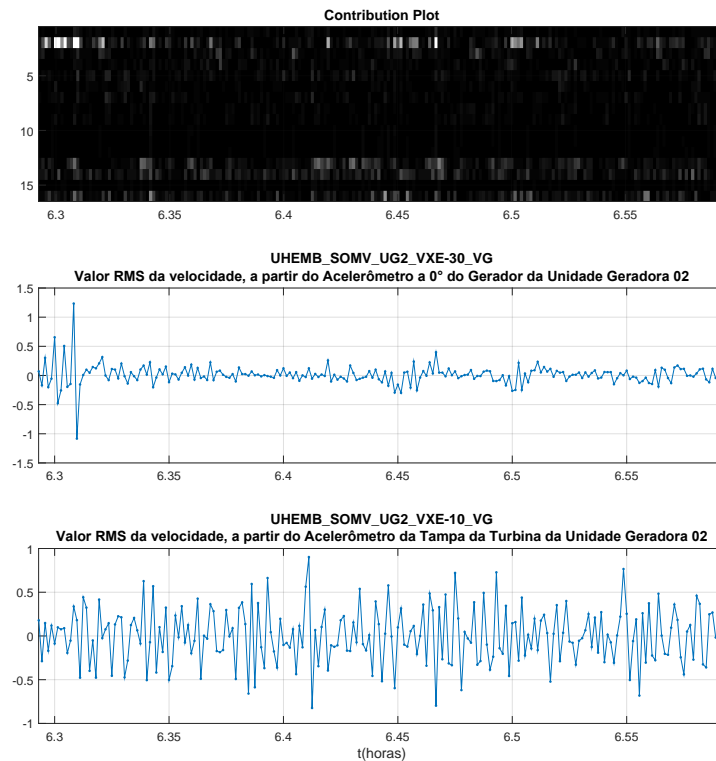
Fonte: elaborada pelo autor.

Em um segundo trecho de monitoramento, apresentado na Figura 39, observam-se algumas amostras, a partir de  $t = 6,3$  horas, com alto valor de índice  $T^2$  e uma delas com alto valor de SPE. Pelo *contribution plot*, Figura 40, são provavelmente causados pela vibração, medida a partir do acelerômetro no eixo 0 graus do gerador, o que é confirmado inspecionando esta variável individualmente (segundo gráfico). A título de comparação, apresenta-se no terceiro gráfico a vibração na tampa da turbina no mesmo trecho, onde não verifica-se nenhum comportamento atípico.

Figura 39: Operação via CAG: Índices estatísticos na fase de monitoramento (exemplo 2)



Fonte: elaborada pelo autor.

Figura 40: Operação via CAG: *Contribution Plots* na fase de monitoramento (exemplo 2)

Fonte: elaborada pelo autor.

### 5.3 Comentários Finais

Este capítulo tratou da apresentação e análise dos resultados obtidos pelos procedimentos descritos anteriormente para as fases de projeto e monitoramento de condição com técnicas de estatística multivariada. Verificou-se a capacidade do sistema de monitoramento implementado de detectar mudanças em diferentes cenários operacionais e indicar as variáveis ou setores da unidade geradora provavelmente responsáveis por cada uma das mudanças. Além disso, o sistema indica, através de um aumento significativo de alarmes, quando o modelo estatístico estimado não corresponde mais ao comportamento da unidade. Nesse caso, pode haver a demanda de manutenção dos equipamentos ou a necessidade de reprojeter o sistema de monitoramento, para adequá-lo a novas referências. O capítulo seguinte trata das conclusões gerais do trabalho e das sugestões de continuidade.

## 6 Conclusões

### 6.1 Conclusões Gerais

Apresentou-se o desenvolvimento de um sistema de monitoramento amplo de condição de uma unidade geradora hidrelétrica com diferentes modos de operação. O objeto de estudo foi a Usina Hidrelétrica de Emborcação, gerida pela CEMIG, localizada no curso do Rio Paranaíba, em Minas Gerais. As técnicas desenvolvidas são aplicadas em variáveis coletadas por um Sistema *Online* de Monitoramento de Vibração, implementado em uma controladora da *National Instruments*, em conjunto com outras medições da rede digital da usina. A metodologia desenvolvida é composta por três etapas: pré-processamento de variáveis de vibração; estimação recursiva de variância para cálculo computacional do modo de operação; uso de técnicas de análise e controle estatístico multivariado, como PCA, Carta de Hotelling  $T^2$ , índice SPE, índice combinado e *contribution plots*.

A UHE Emborcação faz parte de um mecanismo do Operador Nacional do Sistema Elétrico, chamado Controle Automático da Geração, que confere a ela um modo de operação não presente na maioria das usinas hidrelétricas. Nos frequentes períodos em que uma unidade geradora é operada nessa condição, o *setpoint* de potência ativa é variado constantemente dentro de grandes faixas de operação, levando a unidade a uma condição de forte não-estacionariedade. Mesmo quando a operação está fora do CAG, é comum que os operadores da usina façam modificações locais de *setpoint* para atender as demandas definidas pelo Centro de Operação do Sistema (COS) da CEMIG. Para lidar com esse contexto, são propostas adaptações de média e chaveamento de modelos estatísticos para o monitoramento. O uso de média-móvel na etapa de projeto das cartas e na etapa de monitoramento adapta as técnicas multivariadas às condições de não-estacionariedade.

O desenvolvimento de toda a metodologia de monitoramento amplo de condição é realizado e testado em ambiente Matlab sobre dados históricos do processo. Os resultados apresentados atestam a capacidade do sistema de monitoramento de detectar mudanças em diferentes cenários operacionais e indicar as variáveis ou setores da unidade geradora provavelmente responsáveis por cada uma das mudanças. Além disso, o sistema aponta, através de um aumento significativo de alarmes, quando o modelo estatístico estimado não corresponde mais ao comportamento da unidade. Nesse caso, pode haver a demanda pela manutenção dos equipamentos ou a necessidade de reprojeter o sistema de monitoramento, para adequá-lo a novas referências. A metodologia permite o monitoramento do processo com diferentes modos e patamares de operação considerados normais, ainda que parte destes não tenham sido observados na análise retrospectiva dos dados. Os algoritmos foram desenvolvidos antecipando uma futura implementação no sistema PIMS instalado

na usina.

## 6.2 Sugestões de Continuidade

A respeito do procedimento de monitoramento proposto, sugerem-se as seguintes linhas de desenvolvimento:

- ampliar o uso de médias-móveis com teste CUSUM em conjunto com as técnicas de controle estatístico multivariado para outros tipos de processos industriais.
- recalcular os parâmetros de monitoramento para condições mais atuais da usina, uma vez que nos últimos meses o nível do reservatório subiu consideravelmente. As faixas de potência ativa dos diferentes modos de operação da usina são superiores às verificadas nos dados históricos utilizados no desenvolvimento da metodologia.
- Implementar os algoritmos desenvolvidos no sistema PIMS, para a execução *online* dos novos projetos das cartas e índices, para possibilitar que as equipes de operação e manutenção da usina façam uso desse sistema, extraindo informações úteis para realizações de intervenções no processo.



# Referências

- ALCALA, C. F.; QIN, S. J. Reconstruction-based contribution for process monitoring. *Automatica*, v. 45, n. 7, p. 1593 – 1600, 2009. Citado na página 30.
- ALKAYA, A.; EKER, I. Variance sensitive adaptive threshold-based PCA method for fault detection with experimental application. *ISA Transactions*, v. 50, n. 2, p. 287–302, 2011. ISSN 0019-0578. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0019057810001278>>. Citado 2 vezes nas páginas 17 e 24.
- ANEEL. *Atlas de Energia Elétrica do Brasil*. 3. ed. ANEEL, 2008. 236 p. Disponível em: <<http://www.aneel.gov.br/arquivos/PDF/atlas3ed.pdf>>. Citado na página 12.
- ANEEL. *Capacidade de Geração do Brasil*. 2016. Disponível em: <<http://www.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>>. Acesso em: 01 fev. 2016. Citado na página 12.
- BRAGA, C. M. P. *Notas de aula da disciplina Introdução ao Controle Estatístico de Processos do Departamento de Engenharia Eletrônica da UFMG*. Belo Horizonte: [s.n.], 2013. Citado na página 22.
- HE, Q.; YAN, R.; KONG, F.; DU, R. Machine condition monitoring using principal component representations. *Mechanical Systems and Signal Processing*, v. 23, n. 2, p. 446 – 466, 2009. ISSN 0888-3270. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0888327008000769>>. Citado na página 15.
- JACKSON, J. E. *A User's Guide to Principal Components*. 1. ed. [S.l.]: Wiley-Interscience, 1991. 569 p. ISBN 0471622672. Citado 2 vezes nas páginas 28 e 29.
- JACKSON, J. E. *A User's Guide to Principal Component*. 1. ed. New York: Wiley-Interscience, 2003. Citado 2 vezes nas páginas 24 e 27.
- KANO, M.; HASEBE, S.; HASHIMOTO, I.; OHNO, H. Evolution of multivariate statistical process control: application of independent component analysis and external analysis. *Computers & Chemical Engineering*, v. 28, n. 6–7, p. 1157–1166, 2004. ISSN 0098-1354. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0098135403002497>>. Citado 3 vezes nas páginas 17, 18 e 24.
- KANO, M.; NAGAO, K.; HASEBE, S.; HASHIMOTO, I.; OHNO, H.; STRAUSS, R.; BAKSHI, B. R. Comparison of multivariate statistical process monitoring methods with applications to the eastman challenge problem. *Computers & Chemical Engineering*, v. 26, n. 2, p. 161 – 174, 2002. ISSN 0098-1354. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0098135401007384>>. Citado 2 vezes nas páginas 17 e 24.
- KIM, H.-C.; KIM, D.; BANG, S. Y. An efficient model order selection for PCA mixture model. *Pattern Recognition Letters*, v. 24, n. 9-10, p. 1385 – 1393, 2003. Citado na página 24.

- LEE, J.-M.; QIN, S. J.; LEE, I.-B. Fault detection and diagnosis based on modified independent component analysis. *AIChE*, v. 52, n. 10, p. 3501–3514, 2006. Citado 2 vezes nas páginas 17 e 24.
- MAJID, N. A. A.; TAYLOR, M. P.; CHEN, J. J. J.; STAM, M. A.; MULDER, A.; YOUNG, B. R. Aluminium process fault detection by Multiway Principal Component Analysis. *Control Engineering Practice*, v. 19, n. 4, p. 367–379, 2011. ISSN 0967-0661. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0967066110002650>>. Citado 2 vezes nas páginas 18 e 24.
- MASON, R. L.; YOUNG, J. C. *Multivariate Statistical Process Control with Industrial Applications*. 1. ed. Philadelphia: Asa Siam, 2002. Citado 4 vezes nas páginas 12, 20, 23 e 24.
- MONTGOMERY, D. C. *Introdução ao Controle Estatístico da Qualidade*. 4. ed. [S.l.]: LTC, 2004. ISBN 8521614004. Citado 8 vezes nas páginas 12, 16, 18, 20, 21, 22, 23 e 26.
- NANDI, S.; TOLIYAT, H. A.; LI, X. Condition monitoring and fault diagnosis of electrical motors-a review. *IEEE Transactions on Energy Conversion*, v. 20, n. 4, p. 719–729, Dec 2005. ISSN 0885-8969. Citado na página 15.
- NEMBHARD, A. D.; SINHA, J. K.; PINKERTON, A. J.; ELBHBAH, K. Combined vibration and thermal analysis for the condition monitoring of rotating machinery. *Structural Health Monitoring*, SAGE Publications, v. 13, n. 3, p. 281–295, 2014. Citado na página 15.
- PATEL, V. C.; KADIRKAMANATHAN, V.; KULIKOV, G. G.; ARKOV, V. Y.; BREIKIN, T. V. Gas turbine engine condition monitoring using statistical and neural network methods. In: *Modeling and Signal Processing for Fault Diagnosis (Digest No.: 1996/260)*, IEE Colloquium on. [S.l.: s.n.], 1996. p. 1/1–1/6. Citado na página 15.
- QIN, J. S. Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, Wiley Online Library, v. 17, n. 8-9, p. 480–502, 2003. Citado 3 vezes nas páginas 16, 17 e 24.
- RUTTEN, C.; NGUYEN, V. H.; GOLINVAL, J. C. Comparison of output-only methods for condition monitoring of industrial systems. *Journal of Physics: Conference Series*, v. 305, n. 1, p. 012101, 2011. Disponível em: <<http://stacks.iop.org/1742-6596/305/i=1/a=012101>>. Citado na página 15.
- SILVA, D. H. d. J. *Processamento de Sinais para Monitoramento de Vibrações em Unidades Geradoras Hidrelétricas*. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, 2013. Citado 2 vezes nas páginas 12 e 15.
- SOUSA, B. M.; LEBEDEV, A.; FERREIRA, J.; AGUIAR, W. E.; BRAGA, C. M. P.; BRAGA, A. R.; LEMOS, A. P. Multivariate Statistical Process Control of Boiler and Turbine-Generator units of a Thermoelectric Power Plant With Multiple Operating Set Points. In: *IEEE INTERNATIONAL CONFERENCE ON INDUSTRIAL TECHNOLOGY*. Seville, Spain: IEEE, 2015. p. 1907 – 1913. Citado 2 vezes nas páginas 50 e 51.

YUE, H. H.; QIN, S. J. Reconstruction based fault identification using a combined index. *Industrial and Engineering Chemistry Research*, v. 40, n. 20, p. 4403 – 4414, 2001. Citado na página 29.

ZHANG, Y.; ZHANG, Y. Fault detection of non-Gaussian processes based on modified independent component analysis. *Chemical Engineering Science*, v. 65, n. 16, p. 4630–4639, 2010. ISSN 0009-2509. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0009250910003003>>. Citado 2 vezes nas páginas 18 e 24.

ZHU, X.; BRAATZ, R. D. Two-Dimensional Contribution Map for Fault Identification. *IEEE Control Systems Magazine*, v. 34, n. 5, p. 72 – 77, 2014. Citado na página 30.