



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E
DE COMPUTAÇÃO



Análise de desempenho de algoritmos de compressão de dados com perda para aplicações industriais

Edson Jackson de Medeiros Neto

Orientador: Prof. Dr. Luiz Affonso Henderson Guedes de Oliveira (DCA/UFRN)

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica e de Computação da UFRN (área de concentração: Engenharia de Computação) como parte dos requisitos para obtenção do título de Mestre em Ciências.

Número de Ordem do PPgEEC: M458
Natal, RN, Novembro de 2015

UFRN / Biblioteca Central Zila Mamede
Catalogação da Publicação na Fonte

Medeiros Neto, Edson Jackson de.

Análise de desempenho de algoritmos de compressão de dados com perda para aplicações industriais. / Edson Jackson de Medeiros Neto. – Natal, RN, 2014.

74 f. : il.

Orientador: Luiz Affonso Henderson Guedes de Oliveira, Dr.

Dissertação (Mestrado) – Universidade Federal do Rio Grande do Norte. Centro de Tecnologia. Programa de Pós-Graduação Engenharia Elétrica e da Computação.

1. Automação industrial - Dissertação. I. Oliveira, Luiz Affonso Henderson Guedes de. II. Universidade Federal do Rio Grande do Norte. III. Título.

RN/UF/BCZM

CDU 681.51

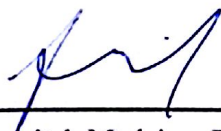
Análise de desempenho de algoritmos de compressão de dados com perda para aplicações industriais

Edson Jackson de Medeiros Neto

Dissertação de Mestrado aprovada em 9 de Novembro de 2015 pela banca examinadora composta pelos seguintes membros:



Prof. Dr. Luiz Affonso H. Guedes de Oliveira (orientador) DCA/UFRN



Prof. Dr. Ivanovitch Medeiros Dantas da Silva IMD/UFRN



Dr. Fábio Soares de Lima Petrobras

*Aos meus pais, Isac e Célia, pelo
apoio incondicional em todos os
momentos de minha vida.*

Agradecimentos

Agradeço primeiramente a Deus, por me guiar ao longo de toda a minha trajetória no desenvolvimento deste trabalho.

Ao meus pais e irmãos André e Lucas pelo grande apoio.

Aos companheiros da LogAp Sistemas. Em especial Joilson Abrantes e Danilo Barros.

Aos colegas do Laboratório de Informática Industrial (LII) pela contribuição seja de forma direta ou indireta no desenvolvimento deste trabalho.

Ao meu orientador Luiz Affonso H. Guedes, sou muito grato pela orientação.

Aos demais membros da banca Ivanovitch Medeiros e Fábio Lima agradeço a participação e contribuição para este trabalho.

A todos os colegas e professores do Programa de Pós-graduação em Engenharia Elétrica e de Computação, que contribuíram para minha formação acadêmica. Em especial Anderson Santos e Daniel Macedo, pelas dicas e ajudas durante toda a trajetória neste mestrado.

Resumo

O grande volume de dados gerados como resultado da supervisão de processos de automação na indústria gerou como consequência um vasto espaço de armazenamento em disco, assim como dificuldade na transmissão destes dados por *links* de telecomunicações. Os algoritmos de compressão com perda de dados surgiram na década de 90 com intuito de solucionar estes problemas, passando a serem utilizados em sistemas de supervisão industrial para a compressão de dados em tempo real. Para isso estes foram projetados para eliminar informações redundantes e indesejadas de forma simples e eficiente. No entanto, os parâmetros destes algoritmos necessitam de serem configurados para cada variável de processo, tornando inviável a configuração manual em caso de sistemas que supervisionam milhares de variáveis. Nesse contexto este trabalho propõe o algoritmo *Adaptive Swinging Door Trending*, que consiste numa adaptação do *Swinging Door Trending*, em que seus principais parâmetros são ajustados dinamicamente através da análise de tendências do sinal. Propõe-se também uma análise comparativa de desempenho dos algoritmos de compressão com perda de dados aplicados sobre variáveis de processo de séries temporais e cartas dinamométricas de fundo de poço. Os algoritmos abordados para efeito comparativos foram os lineares por partes e os de transformadas.

Palavras-chave: Adaptive Swinging Door Trending, PIMS, Swing Door Trending, Descritores de Fourier, Automação Industrial, Cartas Dinamométricas.

Abstract

The great amount of data generated as the result of the automation and process supervision in industry implies in two problems: a big demand of storage in discs and the difficulty in streaming this data through a telecommunications link. The lossy data compression algorithms were born in the 90's with the goal of solving these problems and, by consequence, industries started to use those algorithms in industrial supervision systems to compress data in real time. These algorithms were projected to eliminate redundant and undesired information in a efficient and simple way. However, those algorithms parameters must be set for each process variable, becoming impracticable to configure this parameters for each variable in case of systems that monitor thousands of them. In that context, this paper propose the algorithm *Adaptive Swinging Door Trending* that consists in a adaptation of the *Swinging Door Trending*, as this main parameters are adjusted dynamically by the analysis of the signal tendencies in real time. It's also proposed a comparative analysis of performance in lossy data compression algorithms applied on time series process variables and dynamometer cards. The algorithms used to compare were the piecewise linear and the transforms.

Keywords: Adaptive Swing Door Trending, PIMS, Swing Doors Trending, Fourier Descriptors, Industrial Automation, Dynamometer Cards.

Sumário

Sumário	i
Lista de Figuras	iii
Lista de Tabelas	vii
Lista de Artigos Aceitos e/ou Publicados	ix
1 Introdução	1
1.1 Gerenciamento de processos automatizados	1
1.1.1 PIMS (Plant Information Management Systems)	3
1.1.2 Compressão de dados	5
1.2 Objetivos	6
1.3 Contribuições	6
1.4 Divisão da Dissertação	7
2 Fundamentação Teórica	9
2.1 Método de compressão linear por partes	10
2.1.1 Algoritmo <i>Boxcar/Backslope</i>	10
2.1.2 <i>Swinging Door Trending</i> (SDT)	11
2.1.3 <i>Improved Swinging Door Trending</i> (ISDT)	12
2.2 Transformadas	15
2.3 Comparações entre métodos	18
2.4 Sistemas Relacionados a Compressão de Dados na Indústria	19
2.4.1 BR-PlantHistorian	19
2.4.2 BR-PlantViewer	22
2.4.3 BR-Collector	24
3 Proposta	25
3.1 Algoritmo de compressão Adaptativo	25
3.1.1 <i>Adaptive Swinging Door Trending</i> (ASDT)	25

3.2	Compressão aplicada a Cartas Dinamométricas	27
3.2.1	Carta Dinamométrica	28
3.2.2	Aquisição de Dados	29
3.2.3	Modos de Operação	29
3.2.4	Algoritmo de compressão de carta com base em Descritores de Fourier	31
3.2.5	<i>Swinging Door Trending</i> aplicado a carta dinamométrica	33
3.2.6	Métrica de similaridade entre curvas	34
4	Resultados	37
4.1	Ferramentas Utilizadas	37
4.2	Compressão aplicada a séries temporais	37
4.2.1	Estudo de Caso 1	38
4.2.2	Estudo de Caso 2	39
4.3	Compressão aplicada a cartas dinamométricas de fundo de poço	41
4.3.1	Métricas de comparação	42
4.3.2	Estudo de caso 1	42
4.3.3	Resultados Gerais do Estudo de Caso 1	58
4.3.4	Estudo de Caso 2	58
4.3.5	Resultados Gerais do Estudo de Caso 2	63
4.3.6	Considerações Finais dos Estudos de Casos	64
4.4	Suíte BR-Plant	65
5	Conclusão e Trabalhos Futuros	67
5.1	Trabalhos Futuros	68
	Referências Bibliográficas	69
A	Bombeio Mecânico	73

Lista de Figuras

1.1	Pirâmide da automação - modelo hierárquico representando os diversos níveis da automação industrial.	2
1.2	Sistema PIMS como centralizador de informações.	3
1.3	Transformando dados em conhecimento [Filho 1993].	4
1.4	Gráfico de crescimento do consumo de espaço em disco ao longo de 4 dias.	5
1.5	Diagrama de blocos de compressão de dados.	7
2.1	Algoritmo Boxcar/Backslope.	11
2.2	Algoritmo Swinging Door Trending.	12
2.3	Algoritmo Swinging Door Trending.	13
2.4	Análise multiresolução de uma transformada <i>wavelet</i>	16
2.5	Exemplos de <i>wavelet</i> mãe.	17
2.6	Arquitetura da suíte BR-Plant	20
2.7	Estatísticas de configuração das variáveis historiadas	21
2.8	Etapas de compressão do BR-PlantHistorian	22
2.9	Exemplo de tela de análise em tempo real do BR-PlantViewer	23
2.10	Arquitetura do BR-Collector.	24
3.1	Exemplo de carta dinamométrica.	28
3.2	Assinaturas Padrões de Cartas Dinamométricas	30
3.3	Espectro carta dinamométrica	32
3.4	Alinhamento entre curvas usando o algoritmo DTW	35
3.5	Exemplo código algoritmo DTW.	35
4.1	Sinal original sem compressão; Sinal comprimido através do ASDT; Sinal reconstruído.	38
4.2	Erro dos algoritmos de compressão ASDT e SDT a medida que se aumenta a taxa de compressão.	39
4.3	Análise algoritmo SDT aplicado a variável FI-240051	40
4.4	Análise algoritmo SDT aplicado a variável fração molar etano	40

4.5	Análise algoritmo ASDT aplicado a variável FI-240051	41
4.6	Análise algoritmo ASDT aplicado a variável fração molar etano	41
4.7	<i>Swinging Door</i> aplicado carta de operação normal.	44
4.8	FFT aplicado a cartas de operação normal.	44
4.9	Análise de erro de compressão para carta de operação Normal.	45
4.10	<i>Swinging Door</i> aplicado sobre carta de operação Pancada de Fluido.	45
4.11	FFT aplicado sobre carta de operação Pancada de Fluido.	46
4.12	Análise de erro de compressão para carta de operação Pancada de Fluido.	46
4.13	<i>Swinging Door</i> aplicado a carta de operação Válvula de Passeio.	47
4.14	FFT aplicado a carta de operação Válvula de Passeio.	48
4.15	Análise de erro de compressão para carta de operação Válvula de Passeio.	48
4.16	<i>Swinging Door</i> aplicado a carta de operação Válvula de Pé.	49
4.17	FFT aplicado a carta de operação Válvula de Pé.	50
4.18	Análise de erro de compressão para carta de operação Válvula de Pé.	50
4.19	<i>Swinging Door</i> aplicado a carta de operação Coluna Móvel.	51
4.20	FFT aplicado a carta de operação Coluna Móvel.	51
4.21	Análise de erro de compressão para carta de operação Coluna Móvel.	52
4.22	<i>Swinging Door</i> aplicado a carta de operação Batidas na Bomba de Fundo.	53
4.23	FFT aplicado a carta de operação Batidas na Bomba de Fundo.	53
4.24	Análise de erro de compressão para carta de operação Batidas na Bomba de Fundo.	54
4.25	<i>Swinging Door</i> aplicado a carta de operação Interferência de Gás.	55
4.26	FFT aplicado a carta de operação Interferência de Gás.	55
4.27	Análise de erro de compressão para carta de operação Interferência de Gás.	56
4.28	<i>Swinging Door</i> aplicado a carta de operação Pistão Preso	56
4.29	FFT aplicado a carta de operação Pistão Preso	57
4.30	Análise de erro de compressão para carta de operação Pistão Preso.	57
4.31	Boxplot (Diagrama de Caixa)	59
4.32	Análise da distribuição do erro para um conjunto de cartas de modo de operação Normal	61
4.33	Análise da distribuição do erro para um conjunto de cartas de modo de operação Pancada de Fluido	61
4.34	Análise da distribuição do erro para um conjunto de cartas de modo de operação Interferência de Gás	62
4.35	Análise da distribuição do erro para um conjunto de cartas de modo de operação de Válvula Passeio	63

4.36	Análise da distribuição do erro para um conjunto de cartas de modo de operação de Válvula de Pé	64
4.37	Análise de cartas dinamométricas através do BR-PlantViewer	66
A.1	Unidade de Bombeio Mecânico	73

Lista de Tabelas

4.1	Taxa de compressão por variável para o algoritmo SDT.	40
4.2	Taxa de compressão por variável para o algoritmo ASDT.	41

Lista de Artigos Aceitos e/ou Publicados

MEDEIROS, EDSON; AUGUSTO, L.; C. SOUZA, D.; GUEDES, L. A., ADAPTIVE SWING DOORS TRENDING: UM ALGORITMO ADAPTATIVO PARA COMPRESSÃO DE DADOS EM TEMPO REAL, *em* Congresso Brasileiro de Automática (CBA), 2014, Belo Horizonte.

Capítulo 1

Introdução

A utilização em larga escala dos sistemas de automação e controle e a expansão no desenvolvimento de sensores tem aumentado o número de pontos de aquisição, que juntamente com a alta amostragem utilizada em sistemas de controle, contribuíram para gerar uma grande quantidade de dados provenientes do campo industrial. Consequentemente, salvar esses dados demanda um vasto espaço para armazenamento e dificulta a transmissão dos dados na rede, devido ao alto tráfego demandado [Huang et al. 2006].

Esses dados possuem importantes informações que podem ser posteriormente utilizadas pelos sistemas corporativos de gestão da planta, *Enterprise Resource Planning* (ERP), responsáveis pela transformação desses dados em informações de negócio [Souza et al. 2005]. Diante da extrema importância desses dados e de seu grande volume, a necessidade de possuir um eficiente sistema de armazenamento se torna cada vez mais presente a nível de sistemas de gerência de informação de processos, englobados pelo termo geral *Enterprise Production Systems* (EPS).

O objetivo desse capítulo é contextualizar a necessidade de compactação da grande quantidade de dados provenientes de sensores de chão de fábrica. Ao final do capítulo serão descritos as motivações e os objetivos deste trabalho, além da organização do respectivo documento.

1.1 Gerenciamento de processos automatizados

Os processos de automação industrial podem ser representados de forma hierarquizada, por meio da subdivisão em níveis conforme mostrado na figura 1.1. No primeiro nível, encontram-se os sensores e atuadores, como sensores de nível, pressão, temperatura, de fim de curso, válvulas, inversores de frequência, etc. No nível seguinte, encontram-se os Controladores lógicos programáveis (CLPs) e os sistemas supervisórios - SCADA (*Supervisory Control and Data Acquisition*), que se comunicam com os CLPs. Até o início da

década de 1990, os sistemas de controle formavam ilhas de automação, onde cada sistema controlava o seu parque sem possibilitar a integração das informações. Com a disponibilização dos dados da produção, desde o chão-de-fábrica até o produto final, pode-se subir mais um nível na pirâmide da automação, onde se encontram os sistemas de gerência de informação de processos, que são englobados com o termo geral de *Enterprise Production Systems* (EPS), onde estão incluídos os *Plant Information Management System* (PIMS), e os *Manufacturing Execution Systems* (MES). No nível mais alto da pirâmide encontram-se os sistemas corporativos de gestão da planta, *Enterprise Resource Planning* (ERP), responsáveis pela transformação desses dados em informações de negócio.

A figura 1.1 também representa o fluxo das informações do processo automatizado, mostrando que os dados provenientes da planta industrial, por meio dos equipamentos e dos sensores, os quais fazem parte da base da pirâmide, são requisitados e processados por ferramentas em nível de supervisor (CLP, SCADA), estas ferramentas disponibilizam estes dados para sistemas pertencentes à camada EPS, estes sistemas centralizam e disponibilizam as informações do processo para o nível corporativo da pirâmide de automação.

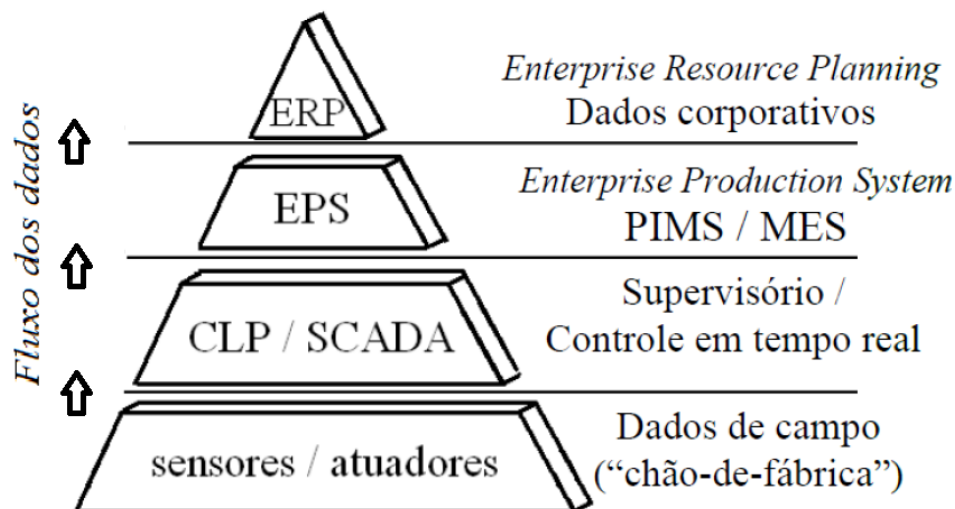


Figura 1.1: Pirâmide da automação - modelo hierárquico representando os diversos níveis da automação industrial.

1.1.1 PIMS (Plant Information Management Systems)

Os PIMS são sistemas que adquirem dados de processo de diversas fontes (através da aquisição em nível de supervisor), os armazenam num banco de dados históricos e os disponibilizam através de diversas formas de representação [Carvalho et al. 2005]. Os PIMS nasceram na indústria de processos contínuos mais propriamente na indústria química e petroquímica para resolver o problema da fragmentação de dados e proporcionar uma visão unificada do processo. Numa primeira fase, os PIMS passam a se constituir na ferramenta fundamental para o engenheiro de processos. A partir de uma estação ele pode visualizar tanto os dados de tempo real como históricos da planta. Pode montar tabelas, gráficos de tendência e sinóticos e eliminar as ilhas de informação, concentrando em uma única base de dados informação sobre todos os aspectos de uma planta. A figura 1.2 ilustra o esquemático do sistema PIMS, que pode ser visualizado como um centralizador de dados.

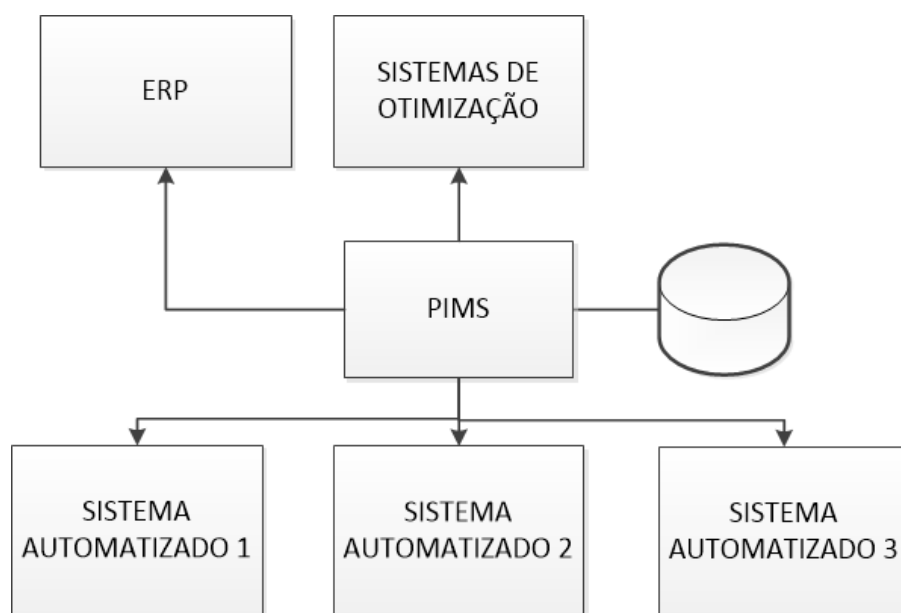


Figura 1.2: Sistema PIMS como centralizador de dados.

Embora os PIMS tenham origem na indústria química e petroquímica, rapidamente foram expandidos para outros mercados como: papel e celulose, siderúrgica, cimento, mineração, ou seja, praticamente todas as indústrias de processos [Filho 1993].

A boa gerência dos dados oriundos de um processo desempenha um papel importante na indústria. Deles depende a eficiência de suas operações, que é sustentada pela prática de diversos elementos, entre os quais estão [Bakshi & Stephanopoulos 1996]:

- Acompanhamento estatístico e o planejamento da produção;
- Melhoria contínua dos processos;
- Controle de qualidade;
- Validação e identificação de modelos;
- Monitoramento, detecção e diagnóstico de falhas;
- Auditoria de desempenho em malhas de controle.

Os constituintes principais de um PIMS são o historiador de processos, a interface gráfica para recuperação e visualização dos dados armazenados e as aplicações clientes complementares [Filho 1993]. Em um histórico, por exemplo, a base de dados é registrada, geralmente, como uma lista contendo os seguintes atributos:

- Timestamp: armazena data e tempo do correspondente valor;
- Identificação do dado (tag);
- Valor;
- Qualidade do dado: informa se o dado não é considerado confiável.

Um sistema de gerenciamento eficaz é capaz de concentrar a massa de dados e permitir transformá-los em informação útil, e essa informação em conhecimento. A Figura 1.3 ilustra essa ideia. Dados com elevado valor agregado são aqueles que, em pouca quantidade, permitem que deles se extraia conhecimento. No outro extremo, os dados brutos estão em grande quantidade e possuem pouco valor, pois a informação relevante não pôde ser isolada, e o conhecimento torna-se mais difícil de ser obtido. Assim, entre outros benefícios, um bom sistema de gerenciamento possibilita ao engenheiro de processos entender as situações operacionais que se apresentam comparando-as com padrões previamente arquivados.



Figura 1.3: Transformando dados em conhecimento [Filho 1993].

A centralização das informações em uma única base de históricos traz como desafios a necessidade de se formar um *big data* de dados de variáveis de processos temporais, capaz de armazenar dados de milhares a milhões destas variáveis por um longo período de tempo. Com isso, o armazenamento deste grande volume dados em disco rígido, assim como o tráfego de rede dos mesmos, passou a se tornar um problema real. Uma boa solução para esse problema é a utilização de um bom sistema de compressão de dados em tempo real, onde os dados são compactados na medida em que são recebidos [Shu & Shu 2012].

1.1.2 Compressão de dados

A compressão de dados tem como principal objetivo eliminar informações redundantes, a fim de minimizar o espaço necessário para o armazenamento. Consequentemente, diminuindo os custos com armazenamento e transmissão de dados, além de melhorar a eficiência na análise das informações. A Figura 1.4 apresenta um gráfico comparativo entre o espaço ocupado por dados sem compressão e dados com compressão.

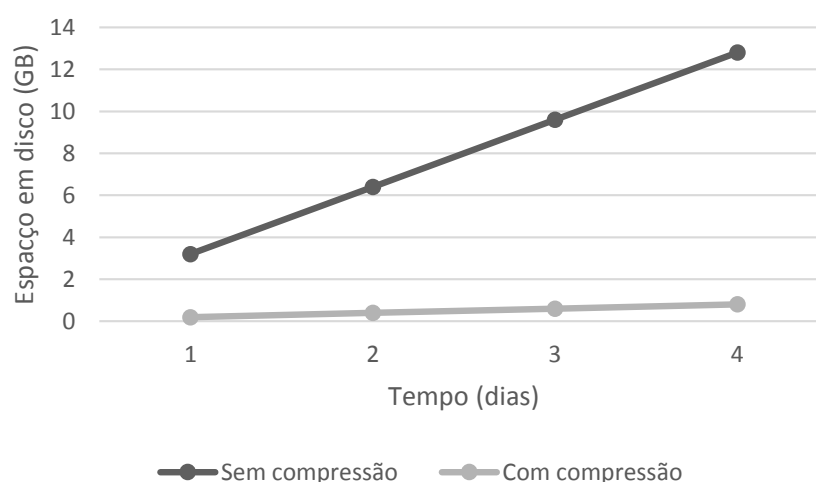


Figura 1.4: Gráfico de crescimento do consumo de espaço em disco ao longo de 4 dias.

Os algoritmos de compressão se dividem em dois grupos: com e sem perda de dados. Os algoritmos sem perda de dados permitem que os dados, quando descompactados, sejam idênticos aos dados originais. Esses algoritmos costumam proporcionar uma baixa taxa de compressão. Os algoritmos de compressão com perda possibilitam uma alta taxa de compressão, por outro lado, podem acarretar alta perda de dados, permitindo apenas uma estimativa dos dados originais [Souza et al. 2005].

É desejável que um algoritmo de compressão de dados com perda, permita uma análise similar entre os dados comprimidos e os originais sem eliminar informações relevantes e com a mais alta taxa de compressão possível. Esses algoritmos eliminam dados considerados redundantes ou próximos de uma determinada tendência linear, ou seja, aqueles dados que poderiam ser recuperados posteriormente com baixo erro, utilizando uma técnica de interpolação adequada, são descartados.

1.2 Objetivos

Considerando a real importância da compactação dos dados no que diz respeito ao gerenciamento da informação dos processos industriais. Este trabalho tem por objetivo realizar um estudo e análise de desempenho dos algoritmos de compactação com perdas de dados em tempo real para aplicações industriais. Também se pretende propor uma metodologia de ajustes adaptativo de parâmetros do algoritmo que obtenha melhor custo benefício em termos de compressão de dados aplicados a variáveis de processos.

Um outro objetivo é realizar um estudo de algoritmos de compressão aplicados a cartas dinamométricas para aplicações em sistemas de tempo real, assim como incorporar ao sistema BR-PlantHistorian, explicado no capítulo 3, a capacidade de armazenar dados de cartas dinamométricas de forma otimizada utilizando tais algoritmos. Visando contribuição para o desenvolvimento desta ferramenta, assim como uma validação destes algoritmos em um sistema de tempo real aplicado na indústria.

1.3 Contribuições

Em resumo as contribuições dessa dissertação são:

1. Proposição de um algoritmo adaptativo denominado *Adaptive Swinging Door Trending* aplicado a dados de variáveis de processo temporais.
2. Análise comparativa entre os algoritmos *Adaptive Swinging Door Trending* e o *Swinging Door Trending* em termos de compressão de dados aplicados a séries temporais.
3. Análise comparativa entre o algoritmo *Swinging Door* e o algoritmo de Descritores de Fourier aplicados a cartas dinamométricas de fundo de poço.
4. Implementação na suíte BR-Plant de compressão de dados aplicado a cartas dinamométricas.

A figura 1.5 mostra um diagrama de blocos que retrata os tipos de compressão de dados e suas ramificações, assim como seus principais algoritmos. Os blocos destacados em azul representam os algoritmos relacionados as contribuições desta dissertação.

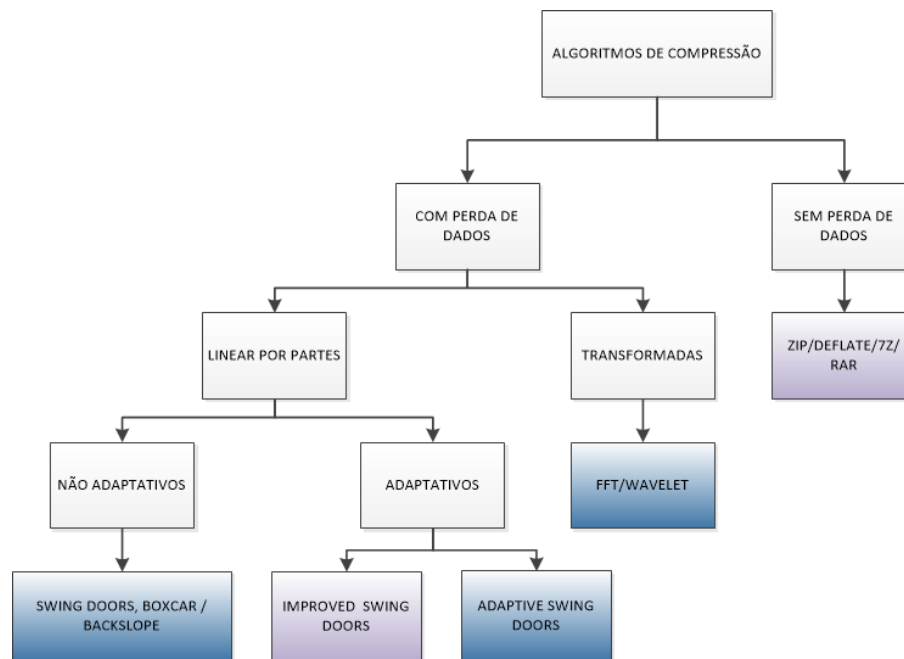


Figura 1.5: Diagrama de blocos de compressão de dados.

1.4 Divisão da Dissertação

O presente trabalho está organizado da seguinte forma: No capítulo 2 é feito um embasamento teórico sobre os algoritmos de compressão propostos na literatura com foco na aplicação em sistemas de automação, juntamente com uma revisão bibliográfica dos principais trabalhos relacionados; no capítulo 3 será feita a proposição de uma adaptação do algoritmo *Swinging Door*, assim como o embasamento sobre cartas dinamométricas e os algoritmos de compressão a serem aplicados nas mesmas; em seguida no capítulo 4 são apresentados os resultados obtidos com relação a compressão de dados aplicado a séries temporais e cartas dinamométricas; por fim no capítulo 5 será feita uma análise conclusiva com base nos resultados apresentados.

Capítulo 2

Fundamentação Teórica

A compressão de dados é realizada com o objetivo de eliminar informações redundantes e/ou tornar possível o armazenamento de um grande volume de informações em um espaço de disco reduzido. Uma estratégia de compressão de dados eficaz é importante para que a performance do banco de dados não degrade rapidamente ao longo do tempo, para o caso de sistemas que necessitam armazenar um grande volume de dados [Andrade et al. 2006].

Uma estratégia de compressão de dados é considerada boa, quando possui as seguintes características: não eliminar informações relevantes para o sistema; ter uma alta taxa de compressão (ou o suficiente para os recursos de armazenamento disponíveis); os dados reconstruídos devem ser os mais próximos possíveis dos dados originais; o algoritmo deve ser simples, rápido e implicar em baixo overhead para a máquina que realiza a compressão.

Porém, estas características são difíceis de serem obtidas por uma mesma estratégia, visto que são inversamente proporcionais, ou seja, quanto maior a taxa de compressão, menor a qualidade do sinal reconstruído. Desta forma, deve haver um equilíbrio nessas duas características.

Existem vários tipos de algoritmos de compressão de dados, por isso deve-se escolher o que melhor satisfaz a cada tipo de aplicação. Os algoritmos se classificam em com e sem perda de informação conforme descrito na subseção 1.1.2.

Entre os principais algoritmos de compressão estão:

- Os algoritmos de codificação de repetição (sem perda de informação): basicamente uma palavra muito frequente, ou uma repetição contínua de um determinado caractere, é substituído por uma codificação que ocupa menos espaço. É mais utilizado em compressões de documentos de texto. Como exemplo temos a codificação *run-length*;

- Algoritmos de amostragem a intervalos fixos (com perda de informação): são salvas amostras periódicas dos dados, porém para aumentar a taxa de compressão o período de amostragem deve ser aumentado, causando assim perda de informações intermediárias relevantes;
- Algoritmos de amostragem variável (com perda de informação): são mais eficientes. É um aprimoramento do algoritmo anterior porque salva amostras apenas quando há mudanças significativas dos dados, mantendo a qualidade dos dados, além de permitir uma alta taxa de compressão.

As próximas seções 2.1 e 2.2 representam os dois principais tipos de métodos de compressão com perda mais difundidos na literatura para aplicações a dados de processos industriais.

2.1 Método de compressão linear por partes

Os métodos "lineares por partes"(ou *piecewise linear*, do termo em inglês) são aqueles utilizados atualmente na indústria, pelo fato de poderem ser aplicados em tempo real no momento da aquisição dos dados. Estes métodos são assim chamados, pois consideram em seu processo de compressão que o sinal segue uma linha reta enquanto os pontos amostrados estiverem dentro de uma tolerância especificada. Comercialmente podem-se destacar os algoritmos *Boxcar/Backslope* [Hale & Sellars 1981] e o *Swinging Door Trending* [Bristol 1990], dos quais serão descritos nas próximas subseções.

2.1.1 Algoritmo *Boxcar/Backslope*

O algoritmo *Boxcar/Backslope* [Hale & Sellars 1981] contém dois critérios que são consultados para decidir se um dado será armazenado ou não. Os critérios são o *Boxcar* e o *Backslope*. Esse algoritmo possui basicamente três parâmetros, que devem ser definidos de acordo com a natureza da variável de processo envolvida, cujas informações serão comprimidas: O desvio de compressão, o tempo máximo de compressão e a qualidade do dado.

O desvio de compressão é um parâmetro utilizado para definir o limite superior e inferior dos dois critérios de *Boxcar* e de *BackSlope*, conforme pode ser observado na figura 2.1. O tempo máximo de compressão especifica o maior tempo permitido entre o último valor armazenado e o último valor recebido. E a qualidade do dado define se o dado medido pelo instrumento é bom ou ruim, ou seja, se é confiável ou não, tomando como base a calibração do instrumento.

O processo de decisão se um dado novo deve ser armazenado ou não, é baseado nos critérios citados anteriormente, na qualidade do dado e no tempo transcorrido entre o último valor armazenado e o último valor recebido. O critério BackSlope verifica se o último valor recebido se apresenta dentro do limite do desvio de compressão, em relação a reta definida entre os dois últimos valores armazenados. O critério de Boxcar verifica se o último valor recebido apresentou um desvio maior que o limite do desvio de compressão, em relação à reta paralela ao eixo do tempo que passa pelo último valor armazenado. O novo valor é armazenado somente se violar os dois critérios.

Em caso de o tempo transcorrido entre o último valor armazenado e o último valor recebido for maior que o parâmetro de tempo máximo de compressão definido, o valor recebido é forçado a ser armazenado. Se a qualidade do último dado recebido mudar, o valor anterior ao último recebido será armazenado [Souza et al. 2005].

A figura 2.1 ilustra o funcionamento deste algoritmo, onde as bolas pretas representam os dois últimos dados armazenados no banco de dados, a bola cinza representa o dado que não foi armazenado no banco de dados, e a bola vermelha representa um dado que violou os dois critérios do algoritmo, e portanto será armazenado no banco de dados.

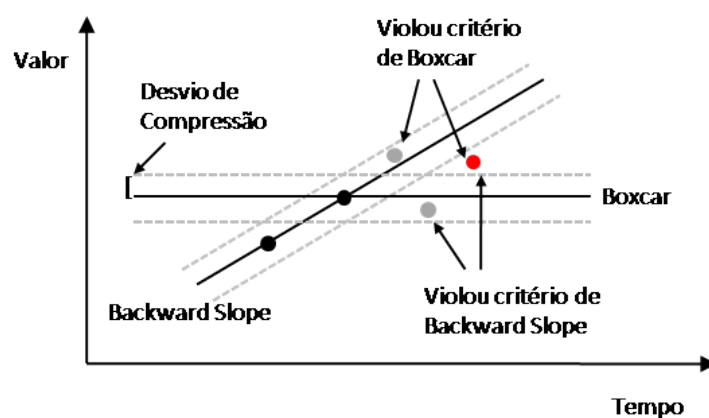


Figura 2.1: Algoritmo Boxcar/Backslope.

2.1.2 *Swinging Door Trending* (SDT)

O algoritmo *Swinging Door Trending* [Bristol 1990], contém três parâmetros que são definidos de acordo com a variável cujas informações serão comprimidas. São eles: tempo máximo de compressão e desvio de compressão. Quanto maior for o tempo máximo de compressão e o desvio de compressão, maior será a taxa de compressão.

O algoritmo cria uma área de cobertura no formato de um paralelogramo entre o último valor armazenado e o valor atual, este paralelogramo possui largura igual ao dobro do desvio de compressão, mostrado na figura 2.2. Caso um dos pontos presentes entre o último valor armazenado e o valor atual estiver fora da área do paralelogramo, todos os pontos são descartados, com exceção do penúltimo ponto recebido, que é armazenado no banco de dados. O último valor recebido é forçado a ser armazenado sempre que o tempo transcorrido desde o último valor armazenado for superior ao tempo máximo de compressão.

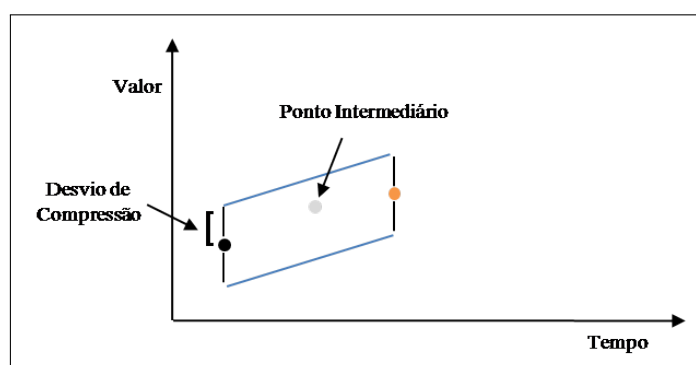


Figura 2.2: Algoritmo Swinging Door Trending.

A figura 2.2 ilustra o funcionamento do algoritmo. A bola preta representa um dado que foi armazenado, a bola cinza representa um dado que não foi armazenado e a bola laranja representa o último valor recebido.

A figura 2.3a apresenta um paralelogramo formado pelo último valor armazenado e o último valor recebido. Nenhum dos valores intermediários ficaram fora do paralelogramo, portanto nada será armazenado.

A figura 2.3b mostra outra situação, onde um dos valores intermediários ficou fora da área do paralelogramo. Logo, o valor anterior ao último valor recebido (em vermelho), será armazenado no banco de dados.

2.1.3 Improved Swinging Door Trending (ISDT)

O algoritmo *Swinging Door* é simples mas possui algumas deficiências: Não detecta e trata bordas, não tem desempenho muito efetivo na presença de ruídos e os resultados obtidos são dependentes do valor definido para o desvio de compressão [Xiaodong et al. 2002]. No entanto, o algoritmo é confiável e pode obter uma taxa de Compressão com

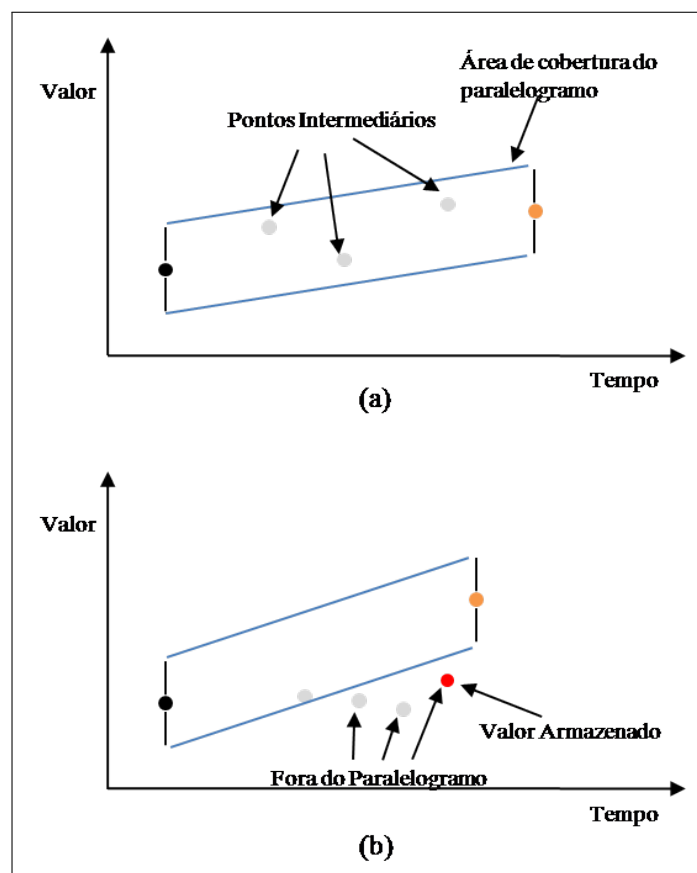


Figura 2.3: Algoritmo Swinging Door Trending.

erros bem reduzidos e sem perda de dados relevantes, além de possuir baixa complexidade computacional.

Baseado na análise de desempenho do *Swinging Door*, [Xiaodong et al. 2002], propõe um melhoramento do algoritmo. O objetivo do algoritmo é: 1. Detectar e eliminar dados redundantes; 2. Um desvio adaptativo (E), que varia com o resultado da compressão anterior, para alcançar uma melhor compressão. Portanto o melhoramento do algoritmo consiste no ajuste de parâmetros de forma dinâmica a medida que novos dados são adquiridos.

Os parâmetros do algoritmo são: O desvio de compressão (E), tempo máximo de compressão ($FSRL$, Forced Storage-Recording Limit), F_{adj} ($0 \leq F_{adj} \leq 1$), parâmetro que determina a variância do desvio de compressão, de acordo com o resultado da compressão anterior), o limite superior do desvio e o limite inferior do desvio.

A seguir será descrito o passo a passo do algoritmo visando uma melhor compressão do mesmo:

1. Setar o valor de E (desvio de compressão) com base em resultados experimentais. Setar o valor de $FSRL$ com base na taxa de compressão. Setar o valor para F_{adj} .
2. O desenvolvimento do algoritmo funciona da mesma forma do SDT, com apenas algumas modificações.
3. Quando o tamanho do intervalo entre o último valor armazenado é maior que o parâmetro $FSRL$ (Forced Storage-Recording Limit), significa que o valor de desvio de compressão E pode ter sido muito alto, ou o processo corrente está em um estado estável. Portanto o desvio de compressão é multiplicado também por um parâmetro descrito anteriormente, que é o f_{adj} , no intuito de diminuir o desvio de compressão, e é atribuído o valor zero ao CIL (compress length interval).
4. Em caso do valor atual ter de ser armazenado no banco de dados segundo os mesmo critérios de armazenamento do SDT e o valor do CIL ser menor que o $FSRL$, o valor de E pode ser muito pequeno ou o processo pode estar no estado de flutuação. Então E é substituído por E/F_{adj} , tendendo o intervalo de compressão a atingir uma maior taxa de compressão.
5. Obviamente, se E continua a crescer, pode causar armazenamento de dados indesejados. Ou ao contrário, se E continua a decrescer, podendo resultar numa baixa taxa de compressão quando o processo estiver em um estado dinâmico. Portanto é definido um limite máximo e um limite mínimo para o valor de E , que em caso de

atingir valor fora dos limites definidos o parâmetro atribui-se a E o valor de desvio de compressão definido no início do algoritmo.

2.2 Transformadas

Os métodos de compressão de dados que se baseiam em transformadas, caracterizam-se por converter o sinal do domínio do tempo para o domínio da frequência, como as transformadas de Cosseno, de *Laplace* e de *Fourier*, ou simultaneamente nos domínios de frequência e tempo, como na mais recente transformada *wavelet*. Essa última tem sido foco de estudo por vários grupos de pesquisa ([Misra et al. 2000], [Bakshi & Stephanopoulos 1996], [Watson et al. 1995]; [Karim & Ismail 2009]), sendo aplicada a compressão de sinais de processos.

Uma transformada linear usualmente compacta a informação original dos dados em um número pequeno de componentes. Por exemplo, a transformada discreta de Cosseno realiza o mapeamento do domínio do tempo para o domínio da frequência, e geralmente o vetor de sinal tem pouca energia nas regiões de alta frequência da banda espectral. Esta propriedade, conhecida como compactação, implica que uma grande parte dos componentes de frequência do vetor são muito próximos de zero e podem ser descartados [Gersho & Gray 1992]. Setar esses coeficientes para zero é conhecido como *thresholding* (limiar) e é a base para compressão através de uma aproximação funcional. Ao descartar esses componentes, setando para zero os seus coeficientes, e aplicar transformada inversa para retornar o sinal para o domínio do tempo, este será apenas uma aproximação do sinal original, caracterizando um método de compressão com perda.

A seguir serão mostradas as transformadas consideradas mais relevantes para o presente trabalho. A transformada discreta de *Fourier* é dada pela equação 2.1, onde x_n , $n = 0, 1, \dots, N - 1$ corresponde ao sinal discreto no tempo e X_k refere-se ao coeficiente da transformada discreta de *Fourier*.

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-2\pi jkn/N} \quad (2.1)$$

A transformada discreta de Cosseno é dada pela equação 2.2 e 2.3, sendo f_j o sinal no domínio de tempo discreto e F_k é o k -ésimo coeficiente da transformada discreta do Cosseno. Algoritmos para computar esta transformada podem ser encontrados em outras referências ([Ahmed et al. 1974]; [Rao & Yip 1990]).

$$F_k = \sqrt{\frac{2}{N}} v_k \sum_{j=0}^{N-1} f_j \cos \left[\frac{(2j+1)k\pi}{2N} \right] \quad (2.2)$$

$$v_k = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } k = 0 \\ 1 & \text{otherwise} \end{cases} \quad (2.3)$$

Transformadas como as de *Fourier* e de Cosseno fornecem um espectro de frequências, porém não informam onde essas frequências ocorrem [Watson et al. 1995]. Uma transformada wavelet pode ser representada simultaneamente em escala (ou frequência) e tempo, a chamada análise multirresolução [Mallat 1989]. A Figura 2.4 mostra um exemplo desse tipo de análise, fornecendo informações sobre a magnitude das frequências em diferentes intervalos de tempo. A transformada *wavelet* na forma discreta é dada pela aproximação 2.4, onde $\psi_{j,k}(t)$ são as funções de base ortonormais, provenientes da dilatação e translação da função $\psi(t)$, esta denominada *wavelet* mãe.

$$f(t) = \sum_{j,k=-\infty}^{\infty} c_{j,k} \psi_{j,k}(t) \quad (2.4)$$

$$c_{j,k} = 2^{j/2} \sum_{l=-\infty}^{\infty} f(l) \psi(2^j l - k) \quad (2.5)$$

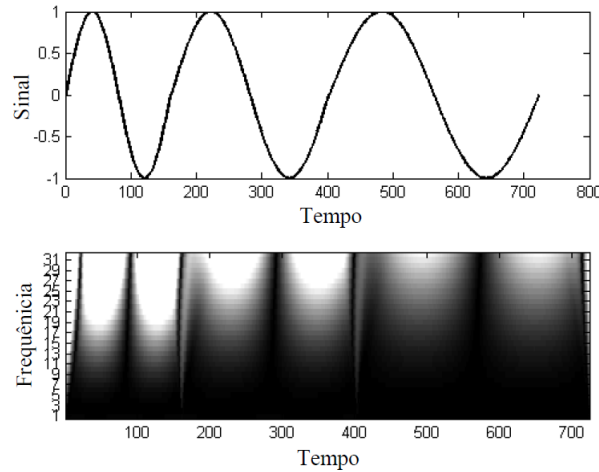


Figura 2.4: Análise multirresolução de uma transformada *wavelet*.

Os coeficientes $c_{j,k}$ são calculados conforme a equação 2.5, onde $f(l)$ é o sinal discreto no domínio do tempo, j é o fator de escala ou dilatação, e k o parâmetro de translação. A primeira e mais simples *wavelet* mãe é a *wavelet* de Haar. A *wavelet* mais

comumente utilizada é a de Daubechies. A figura 2.4 exibe estas *wavelets*, assim como outros exemplos.

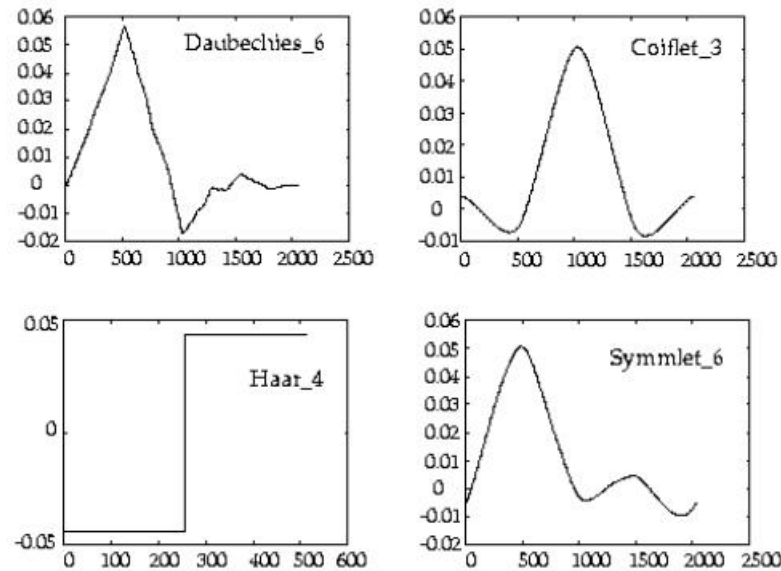


Figura 2.5: Exemplos de *wavelet* mãe.

A análise multirresolução consiste na divisão do espaço de resolução em subespaços representados por uma função de escala. Estas funções são ortogonais. Assim, um sinal em determinado nível de resolução é projetado em um subespaço de *aproximação* e outro de *detalhe*. A compressão de dados pela análise multirresolução é obtida através de restrições aos coeficientes dessas transformações a diferentes níveis de resolução. Componentes relativos a porções de alta frequência do sinal (coeficientes de detalhe) podem ser distinguidos daqueles relativos a tendências mais lentas (coeficientes de aproximação). Dessa forma, desconsiderando os coeficientes de menor magnitude, isto é, de detalhe, é possível gravar um número menor de informação ao mesmo tempo em que se preserva a informação essencial, relativa às tendências mais lentas.

Inicialmente, as transformadas *wavelets* foram concebidas para sequências de dados em batelada [Watson et al. 1995], baseando-se nas aplicações já existentes a outros tipos de sinais. Entretanto, algumas abordagens para aplicações online desse método foram propostas por [Bakshi & Stephanopoulos 1996] e [Misra et al. 2000].

Na próxima seção 2.3, será descrito os principais trabalhos que fizeram uma análise comparativas entre algoritmos de compressão com perda de dados aplicados a processos industriais e o impacto dessas técnicas na análise da planta.

2.3 Comparações entre métodos

Em [Watson et al. 1998] foram feitas comparações entre os métodos lineares, de quantização vetorial e de transformadas (Fourier, Cosseno e *Wavelets*), considerando dois conjuntos de dados reais de planta. Para critério de avaliação foram levados em consideração a razão de compressão 2.6 (*RC*) e o Erro Global Relativo 2.7 (*EGR* em %). Para os testes efetuados em métodos lineares (BC, BS, BC/BS e SDT), o algoritmo SDT mostrou desempenho superior aos demais. Apesar das transformadas *Wavelet* e de Cosseno apresentarem desempenho similar em diversos testes, a transformada *Wavelet* ainda apresenta desempenho superior para sinais que apresentam mudanças do tipo degrau ou variações do tipo impulso. Também foi feita uma avaliação mais geral comparando o SDT com a transformada *Wavelet*, mostrando que a transformada *Wavelet* apresenta melhor resultado quando se considera a relação *RC* vs *EGR*.

$$RC = \frac{n^{\circ} \text{ de dados originais}}{n^{\circ} \text{ de dados armazenados} \times 2} \quad (2.6)$$

$$\%EGR = 100 \times \frac{\sum (f_i - f'_i)^2}{\sum f_i^2} \quad (2.7)$$

[Singhal & Seborg 2005] realizaram estudos sobre o efeito da compressão de dados no casamento de padrões sobre histórico de dados. Os métodos de compressão utilizados na análise foram: O método BC, um algoritmo de compressão de média de dados, o algoritmo usado no PI System e transformadas *Wavelets*. Para o casamento de padrões foram utilizadas técnicas de *Fator de Similaridade por PCA* e *Fator de Similaridade por Distância*. Os testes foram feitos com dados simulados de um processo de um reator CSTR não-isotérmico. As transformadas *Wavelets* forneceram os melhores resultados.

[Thornhill 2004] fez um estudo sobre o impacto da compressão em análise de processos baseada em dados operacionais. O foco deste trabalho foi na compressão pelos métodos lineares por partes, aplicado a dados de processos industriais. Para isso foi implementado o algoritmo SDT descrito por [Bristol 1990]. Três exemplos distintos de sinais reais foram utilizados: Um conjunto de dados que apresenta oscilação permanente; outro que possui variações bruscas de tendência; e o último apenas corrompido por ruído randômico. Como medidas de desempenho foram utilizadas as seguintes estatísticas: Diferença percentual entre médias; razão entre variâncias do sinal original e reconstruído, e entre erros de reconstrução; medidas de não-linearidades e índice de Harris. Os testes demonstraram que todas as propriedades estatísticas utilizadas como medidas sofrem alterações com a compressão dos dados. Com isso o autor faz alertas com base no valor de

RC de modo a evitar a baixa fidelidade dos dados comprimidos em determinados tipos de sinais.

2.4 Sistemas Relacionados a Compressão de Dados na Indústria

Conforme descrito na subseção 1.1.1, os sistemas PIMS são de grande importância, pois permitem a extração de informações de chão de fábrica da indústria, possibilitando a otimização dos processos e por consequência o aumento da produtividade, assim como maior assertividade na realização de manutenção preventiva e corretiva de máquinas, reduzindo custos e riscos de produção. Dos PIMS existentes no mercado, os mais consolidados são o Plant Information (PI) da OSIsoft e o InfoPlus da AspenTech. Dos nacionais se destacam o Blue Plant da Altus e o Elipse Plant Manager da Elipse Software.

O Laboratório de Informática Industrial da UFRN juntamente com uma parceria com a Petrobras desenvolveram o BR-Collector, o BR-PlantViewer e o BR-PlantHistorian que em conjunto formam um sistema PIMS, cuja suíte é denominada BR-Plant. Um dos objetivos do presente trabalho é contribuir no desenvolvimento dessas ferramentas, incorporando as técnicas de compressão mais adequadas analisadas neste trabalho, assim como análise de cartas dinamométricas.

Para o entendimento do fluxo de comunicação e do papel de cada módulo do sistema a figura 2.6 mostra a arquitetura do BR-Plant. A captura dos dados é feita através do coletor (BR-Collector), um serviço denominado Interface de Coleta é o responsável por buscar os dados no coletor, realizar *buffer* e cálculo de desvio de exceção para otimização da comunicação e repassar os dados para o serviço BRHistorian-Server, este responsável pelo processamento, armazenamento e disponibilização dos dados. As ferramentas BR-PlantHistorian e BR-PlantViewer são responsáveis pela configuração do historiador e das análises dos dados, respectivamente.

Uma breve descrição de cada sistema que compõe a suíte BR-Plant é feita nas próximas subseções.

2.4.1 BR-PlantHistorian

A ferramenta BR-PlantHistorian, desenvolvida através de uma parceria entre a UFRN e a Petrobras, é um sistema criado com o intuito de armazenar e disponibilizar dados de processos industriais de forma otimizada por um tempo determinado pelo usuário do

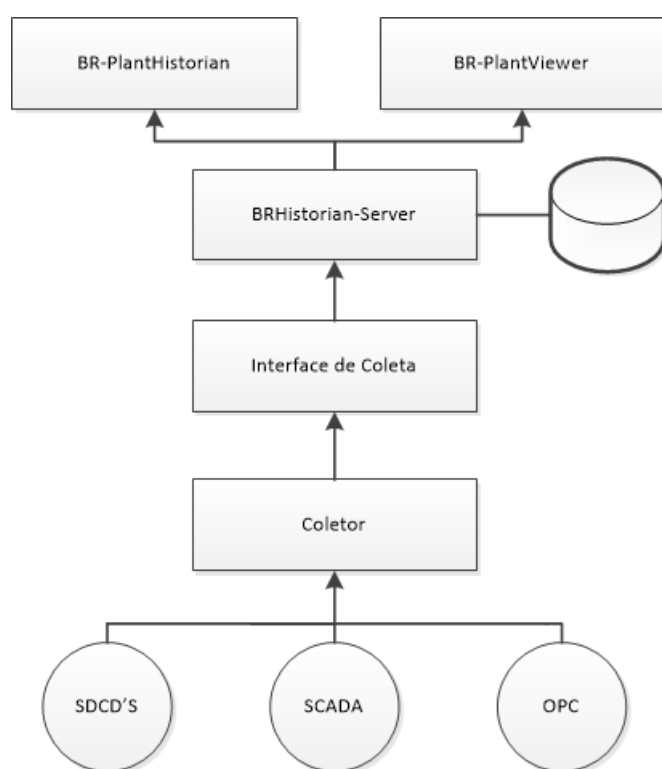


Figura 2.6: Arquitetura da suíte BR-Plant

2.4. SISTEMAS RELACIONADOS A COMPRESSÃO DE DADOS NA INDÚSTRIA21

sistema. Trata-se de um historiador de processos industriais. A figura 2.7 ilustra a tela do BR-PlantHistorian que mostra informações estatísticas das variáveis de processos que estão sendo historiadas.

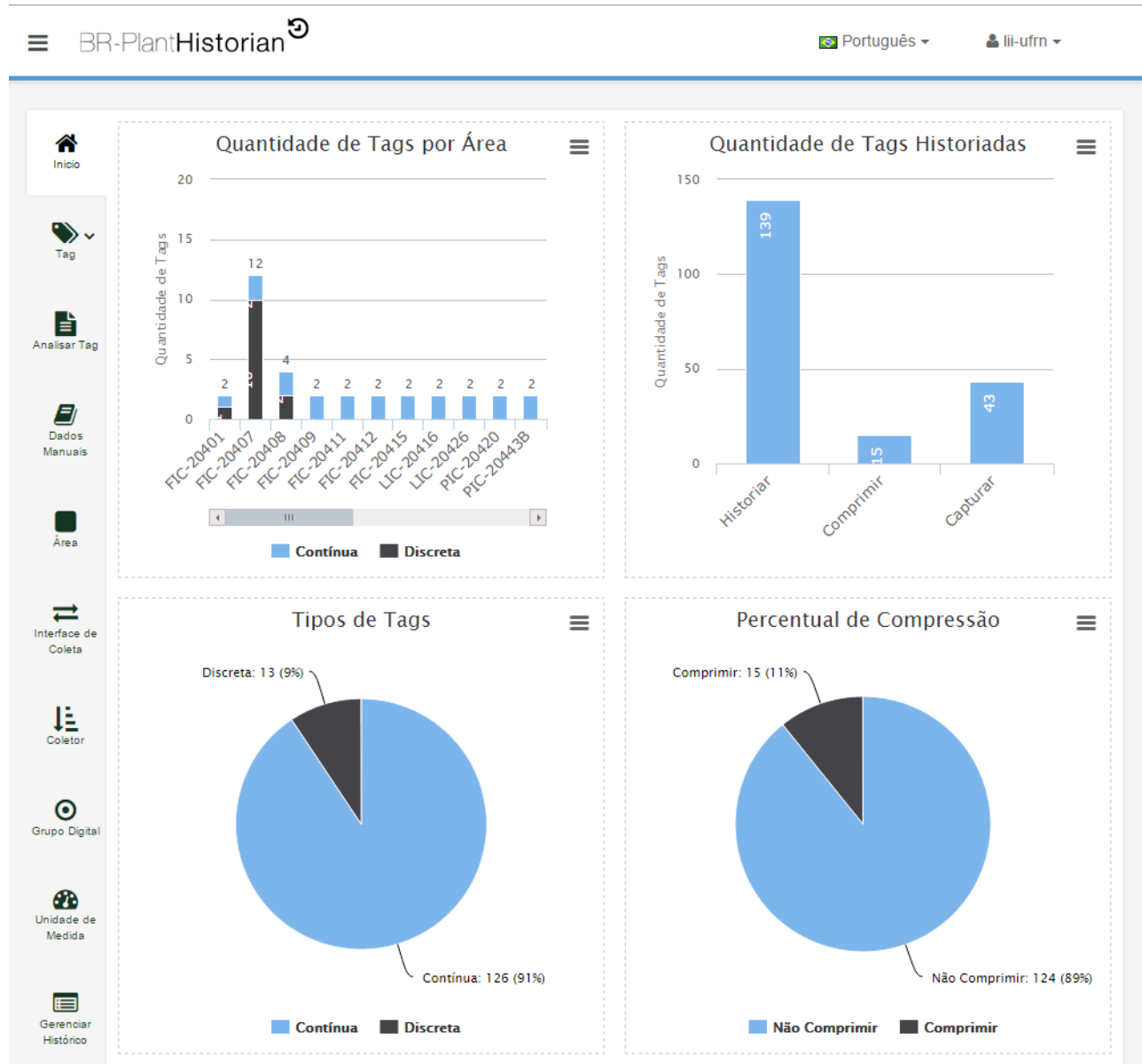


Figura 2.7: Estatísticas de configuração das variáveis historiadas

O BR-PlantHistorian é totalmente distribuído e modularizado, utilizando tecnologias que possibilitam a distribuição do historiador de forma escalável, maximizando o desempenho do sistema simplesmente adicionando novos servidores ao *cluster*. O sistema foi projetado com o intuito de ter capacidade de historiar mais de 80 mil registros por segundo

com apenas um servidor.

O sistema é subdividido em três módulos principais: Módulo de captura dos dados (BR-Collector); módulo de compressão e armazenamento (com perda e sem perda, que podem ser utilizados simultaneamente); módulo de visualização e análise (BR-PlantViewer pode ser enxergado como um módulo).

Para o presente trabalho, a principal contribuição está relacionada ao módulo de compressão de dados. Este módulo tem grande importância, principalmente quando se trata do armazenamento acima de 80 mil variáveis de processo, pois ao armazenar os dados de forma eficiente e organizada, possibilita como consequência uma alta eficiência na disponibilização dos dados.

O módulo de compressão de dados possui três níveis de compressão conforme a figura 2.8, o primeiro nível é o desvio de compressão, que visa a redução de tráfego de rede, basicamente só envia os dados para o servidor caso ultrapasse uma tolerância com relação ao último valor enviado; o segundo nível é o de compressão linear com perdas que possui os seguintes algoritmos implementados: Swing Doors Trending, Boxcar/Backslope e Improved Swing Doors Trending; o terceiro e último nível é o de compressão de dados sem perdas que utiliza o algoritmo *Deflate*, possibilitando ainda uma taxa de compressão de até 30%. Todos os níveis de compressão podem ser ativados ou não, isso fica a critério do usuário.

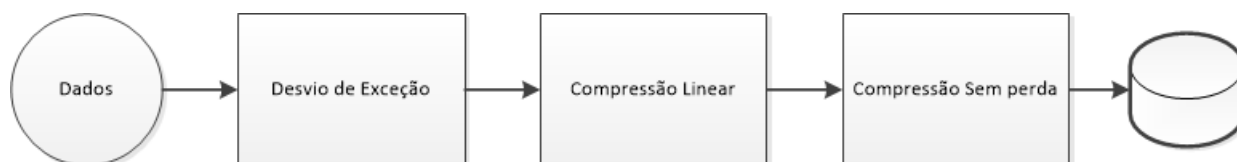


Figura 2.8: Etapas de compressão do BR-PlantHistorian

2.4.2 BR-PlantViewer

O BR-PlantViewer é um sistema desenvolvido através de uma parceria entre a UFRN e a Petrobras, que inicialmente foi projetado para permitir a visualização e análise dos dados do BR-PlantHistorian de forma simples e intuitiva através de *browsers* (tecnologia *web*). Com um tempo a ferramenta evoluiu possibilitando o cruzamento de informações de diversas fontes de dados de forma centralizada na aplicação.

Atualmente o sistema possibilita o acesso a informações do sistema PI da OSIsoft, de Bancos de dados relacionais (Oracle, PostgreSQL, Microsoft SQL Server, MySql), do

2.4. SISTEMAS RELACIONADOS A COMPRESSÃO DE DADOS NA INDÚSTRIA23

BR-PlantHistorian e do BR-Collector. Com isso é possível realizar cruzamento e mineração de dados, exibindo informações através de diversos tipos de componentes gráficos. A visualização das informações pode ser feita de forma personalizada e restrita para cada nível organizacional.

Abaixo segue alguns benefícios que o sistema proporciona:

- Criação e monitoramento em tempo real de indicadores de desempenho (KPI's);
- Criação de interfaces para monitoramento e visualização dos dados;
- Controle de permissão de acesso a relatórios;
- Acessível via dispositivos móveis.

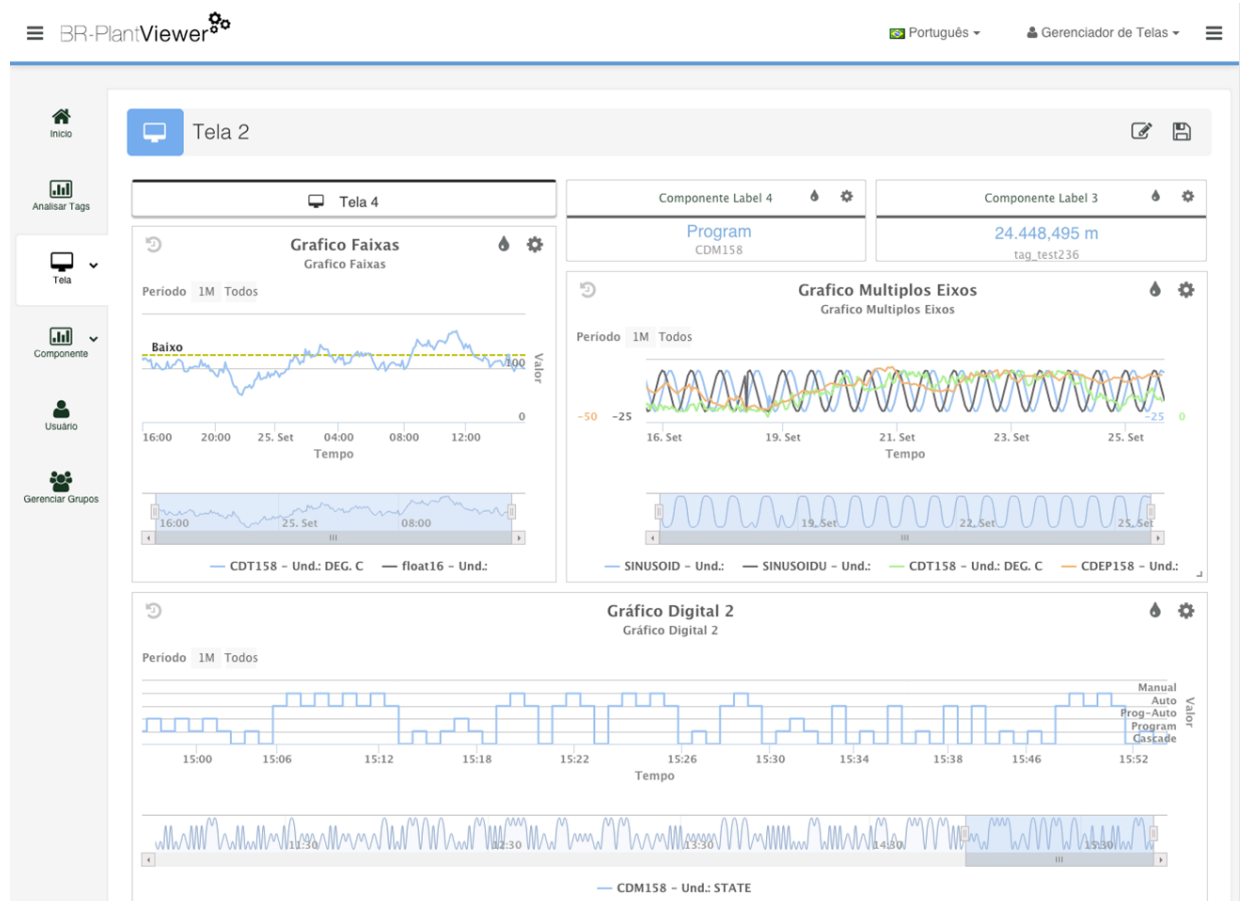


Figura 2.9: Exemplo de tela de análise em tempo real do BR-PlantViewer

2.4.3 BR-Collector

A ferramenta BR-Collector, desenvolvida através de uma parceria entre a UFRN e a Petrobras, é um produto criado para integrar a captura de dados em um ambiente industrial e disponibilizá-los de forma simplificada. Atualmente o BR-Collector suporta captura de dados de processo histórico e em tempo real, captura de dados de alarmes e eventos históricos [Machado 2014].

Padronizado em módulos, o BR-Collector possui uma arquitetura dividida em *slots* e *drivers*. Os *drivers* são responsáveis por implementar os detalhes de cada protocolo de comunicação e servem como interfaces para o BR-Collector através dos *slots*. Os *slots*, por sua vez, têm como função padronizar a forma como o BR-Collector entende cada tipo de dado. Atualmente são disponibilizados quatro *slots* e quatro *drivers* conforme demonstra a Figura 2.10. Estes drivers possibilitam a comunicação com servidores de dados OPC DA, OPC AE, PI da OSIsoft e o BR-AlarmExpert.

À medida que necessário, novos *drivers* podem ser acoplados a essa arquitetura sem qualquer alteração nas camadas superiores.



Figura 2.10: Arquitetura do BR-Collector.

Capítulo 3

Proposta

Neste capítulo é apresentada a proposta de dissertação que tem como contribuições principais: a proposição de um algoritmo adaptativo denominado *Adaptive Swinging Door Trending* para aplicação a dados de variáveis de processo temporais; o embasamento teórico sobre cartas dinamométricas de fundo de poço e a descrição do algoritmo de compressão linear por partes *Swinging Door* e o algoritmo baseado em transformada de *Fourier*, aplicados a cartas dinamométricas, para posterior comparação de desempenho.

3.1 Algoritmo de compressão Adaptativo

O principal problema dos algoritmos de compressão com perda não adaptativos com parâmetros fixos analisados neste trabalho, consiste da necessidade de conhecimento prévio do sinal, pois fatores como a amplitude e a dinâmica do sinal, interferem diretamente na qualidade da compressão dos dados. Em termos de aplicação destes algoritmos em sistemas que monitoram milhares de variáveis de processo, sintonizar esses parâmetros de forma manual para cada variável é inviável.

Com isso, propõe-se um algoritmo adaptativo capaz de se adequar a uma grande quantidade de variáveis de processo, podendo estas, apresentarem sinais de natureza diferente, de modo a garantir um bom nível de compressão e um mínimo de erro na reconstrução do sinal, evitando a necessidade de uma configuração manual de cada variável ou até mesmo uma configuração indevida que leve a perda de informações relevantes do sinal.

3.1.1 *Adaptive Swinging Door Trending* (ASDT)

O algoritmo proposto consiste de uma variação do algoritmo *Swinging Door Trending*, permanecendo todas as suas características básicas, mas utilizando análises de tendências

em tempo real como forma de possibilitar a variação dos principais parâmetros do algoritmo de forma adaptativa. Para isso foi utilizado a média móvel exponencial, que é uma técnica bastante utilizada em séries temporais, que funciona como um filtro, suavizando o sinal e permitindo realizar previsões e análise de tendências do mesmo.

Abaixo temos a fórmula da média móvel exponencial (S):

$$S_n = (X - S_{n-1})\alpha + S_{n-1} \quad (3.1)$$

Onde $\alpha = 2/(N + 1)$ é o chamado parâmetro adaptativo (entre 0 e 1), N representa o tamanho do período a ser considerado na média móvel, X é o valor corrente do sinal, S_n corresponde a média móvel exponencial referente a iteração atual do cálculo e S_{n-1} corresponde a última.

O principal parâmetro do algoritmo SDT é o desvio de compressão, este possui relação direta com a taxa de compressão do sinal, ou seja, quanto maior o desvio de compressão, muito provavelmente maior será a taxa de compressão do sinal.

O algoritmo proposto foi denominado de *Adaptive Swinging Door Trending*, que consiste basicamente em adaptar o parâmetro desvio de compressão do SDT de forma dinâmica tendo como base a média móvel exponencial do sinal.

Com isso a fórmula do desvio de compressão (E_n) ao longo das iterações do algoritmo segue abaixo:

$$E_n = \gamma S_n \quad (3.2)$$

O algoritmo ASDT possui os seguintes parâmetros: γ (valor entre 0 e 1), representando o percentual de influência de S_n (média móvel exponencial do sinal) na adaptação do desvio de compressão, *MaxTime* (tempo máximo de compressão, mesmo do STD), e N que consiste do número de períodos considerados no cálculo da média móvel.

Em termos de funcionamento do algoritmo, este funciona da mesma forma que o SDT só que a cada iteração é feito um cálculo da média móvel exponencial do sinal como demonstrado na Equação 3.1, em seguida é efetuado o cálculo do desvio de compressão com base na Equação 3.2, depois é dada continuidade a toda a lógica do algoritmo SDT descrita na subseção 2.1.2.

A principal vantagem do uso do algoritmo ASDT se deve ao fato de proporcionar um ajuste adaptativo do desvio de compressão em tempo real a cada iteração, utilizando a análise de tendência do sinal através da média móvel exponencial. Com isso, variações bruscas na amplitude do sinal possuem menor influência sobre o desempenho do algoritmo se comparado com o SDT.

3.2 Compressão aplicada a Cartas Dinamométricas

Devido à alta concorrência e a necessidade de cumprir prazos, as indústrias modernas buscam alta disponibilidade e confiabilidade de seus equipamentos para atender as demandas do mercado. Com essa visão, a atividade de manutenção sofreu várias alterações nos últimos anos que levaram a uma evolução no ponto de vista da organização e planejamento de sua execução. De acordo com [Pinto & Nascif 1998], as causas diretas para este desenvolvimento são:

- O rápido aumento da quantidade e diversidade de elementos físicos que compõem equipamentos variados das plantas de processo que devem ser mantidos disponíveis;
- Projetos de engenharia mais complexos;
- Novos métodos para as atividades de manutenção;
- Novas abordagens sobre a organização da manutenção e suas responsabilidades.

O conceito de manutenção preditiva surgiu como um dos resultados dessas demandas. A manutenção preditiva é a monitoração regular das condições de operação (variáveis e parâmetros) de um dispositivo ou processo que fornecem os dados necessários para assegurar o intervalo máximo possível entre os reparos e o melhor planejamento entre as intervenções.

Historicamente, o primeiro método de elevação artificial utilizado pela indústria do petróleo foi o bombeio mecânico [Johnson-Fagg 1958]. Sua importância é refletida no número de instalações encontradas na indústria, estabelecendo o bombeio mecânico como o método mais utilizado no mundo [Lima et al. 2009]. O Apêndice A descreve em mais detalhes a justificativa de uso deste método.

Na prática, o acompanhamento do estado do sistema de bombeio mecânico é realizado através da leitura de sua carta dinamométrica. Através dessa carta é possível saber as condições da bomba localizada no fundo do poço. A carta dinamométrica consiste em um gráfico que relaciona a carga e a posição da bomba, refletindo as condições atuais de bombeio [Rogers et al. 1990]. Dessa forma, a carta pode assumir vários formatos durante a produção do poço, podendo representar situações de funcionamento normal ou indicar irregularidades no sistema.

O processo de identificação de situações de funcionamento anormal do sistema de bombeio mecânico se transforma, nesse caso, em um problema de interpretação visual [Dickinson & Jennings 1990]. Essa abordagem pode ser influenciada por vários fatores, como o pró-

prio comportamento complexo do sistema, diversidade de formas de cartas dinamométricas, além do conhecimento e experiência do engenheiro responsável.

3.2.1 Carta Dinamométrica

Uma carta dinamométrica nada mais é do que um gráfico bidimensional que representa os efeitos gerados pela carga atuante na bomba durante um ciclo de bombeio. A abscissa do gráfico representa a posição da haste polida e a ordenada representa a carga aplicada no sistema 3.1. Existem dois tipos de cartas dinamométricas: a carta de superfície e a de fundo. As cartas são registradas na superfície através de dinamômetros e no fundo do poço através de dispositivos especiais ou modelos matemáticos [Lima et al. 2009].

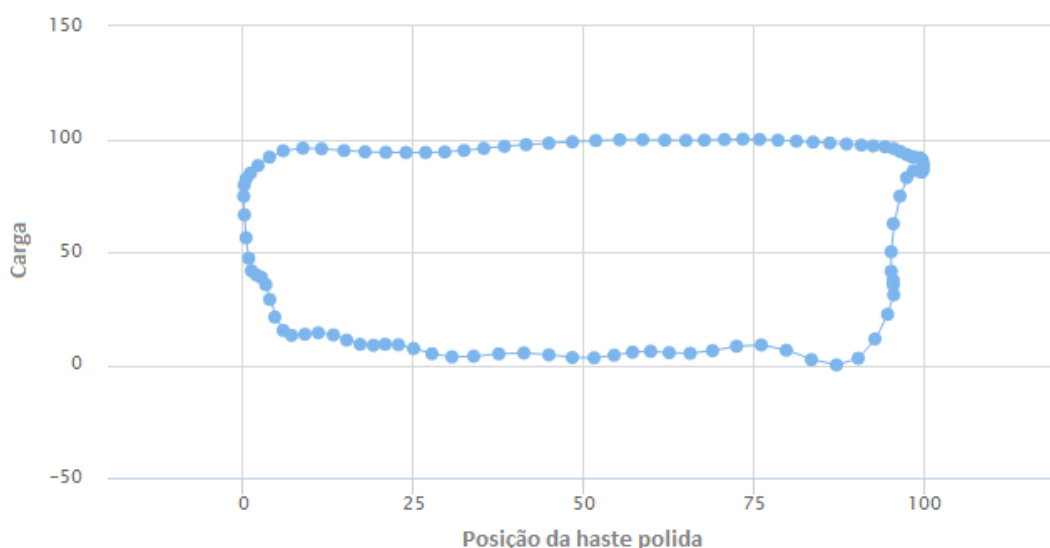


Figura 3.1: Exemplo de carta dinamométrica.

As cartas dinamométricas estão entre as principais ferramentas de análise e avaliação das condições de bombeio, registrando as cargas na haste polida ou no fundo em função do curso das hastes. É possível observar diversas condições de bombeio através da carta, incluindo a possibilidade de detecção de condições de falha.

Devido aos efeitos dinâmicos, alongação das hastes e cargas inerciais a carta dinamométrica apresenta deformação. Desta forma, a interpretação da carta para uma correta avaliação das condições e do desempenho do sistema se torna uma tarefa árdua. Com

isso, a utilização das cartas dinamométricas de fundo de poço oferece uma interpretação mais direta do funcionamento do sistema de bombeio do que o uso de cartas de superfície.

Neste estudo foram usadas as cartas dinamométricas de fundo de poço, resultantes da aplicação de algoritmos que eliminam fatores de deformação da carta de superfície. Quando se utiliza as cartas de fundo, é possível observar que para os mesmos padrões de operação, as cartas de fundo são as mesmas [Filho 1993].

Com a carta de fundo, é mais fácil distinguir modos de operação do sistema de bombeio, tais como:

- Normal;
- Pancada de fluido;
- Vazamento de válvulas de passeio e pé;
- Escorregamento na bomba;
- Furo na coluna de produção;
- Interferência de gás.

3.2.2 Aquisição de Dados

Os dados de cartas dinamométricas de fundo de poço utilizados nas análises desta dissertação foram os mesmos utilizados na tese [Lima 2014]. Cada carta é representada por 100 pontos, sendo resultante da simplificação de uma carta de superfície através de algoritmos de filtragem de ruído [Filho et al. 1996].

3.2.3 Modos de Operação

Cada modo de operação é distinguido por uma carta dinamométrica com um formato característico. A seguir serão apresentados referências de cartas que representam alguns dos principais modos de operação dos sistemas de bombeio mecânico em campos de petróleo segundo [Lima 2014]. A figura 3.2 ilustra as assinaturas padrões das cartas dinamométricas de cada modo de operação.

Normal

O padrão de bombeio normal caracteriza a forma ideal de funcionamento e proporciona as seguintes características:

- Alta eficiência energética;
- Baixa interferência de gás;

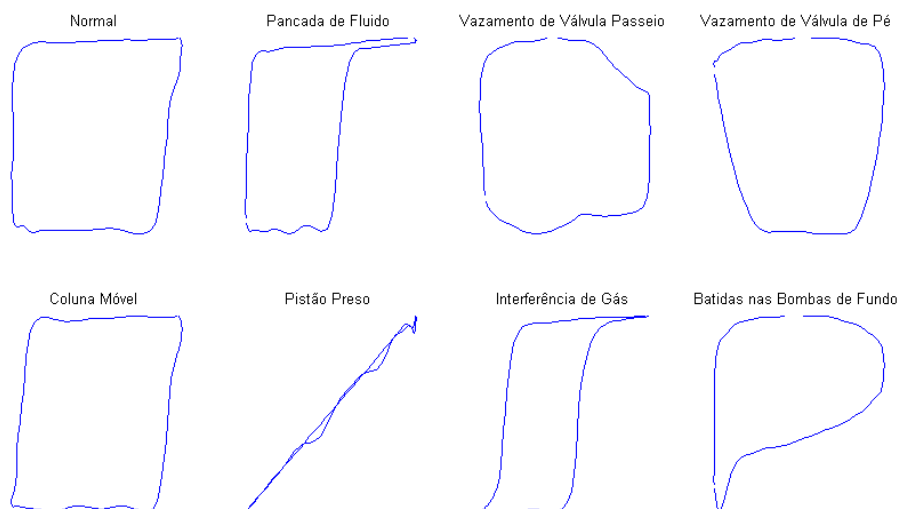


Figura 3.2: Assinaturas Padrões de Cartas Dinamométricas

- Baixa ou média pressão de sucção.

Pancada de Fluido

Este padrão está associado com as seguintes características:

- Baixa pressão de sucção;
- Baixa interferência de gás;
- Sucção da bomba bloqueada.

Vazamento de Válvulas

A condição de vazamento de válvulas ocorre quando existe vazamento em algumas das válvulas de fundo (válvula passeio ou válvula de pé).

Coluna Móvel

Este padrão está associado à falta de ancoramento do tubo de produção. Este fenômeno ocorre em poços profundos em que a âncora de tubo não foi instalada ou foi solta durante a operação. Devido ao movimento recíproco do bombeio mecânico, a coluna de tubo de produção tende a encurtar e alongar, fazendo a carta de fundo inclinar, porém matendo a forma de paralelogramo.

Pistão Preso

Este modo de operação ocorre quando uma coluna de hastes se parti devido a esforços excessivos decorrentes da atuação de cargas variáveis, consequentemente sujeitas à fadiga.

Interferência de Gás

Este padrão está associado à presença significativa de gás na sucção da bomba. Quando este modo de operação ocorre, a bomba realiza uma compressão do gás entre as válvulas, reduzindo o espaço para o óleo e interferindo no funcionamento das válvulas de pé e de passeio. A carta, neste modo de operação, apresenta uma suave diminuição da carga ao decorrer de todo o curso do pistão.

Batidas na Bomba de Fundo

A bomba de fundo é uma bomba de deslocamento positivo. Em outras palavras, quando o fluido é succionado pela bomba, o mesmo não retorna. As batidas na bomba de fundo podem ocorrer na parte superior ou inferior da camisa do inserto.

3.2.4 Algoritmo de compressão de carta com base em Descritores de Fourier

O algoritmo de compressão de carta dinamoétrica com base em descritores de frequência utilizado neste trabalho foi o de limiarização dos Descritores de Fourier. Para isso a carta dinamoétrica original é adaptada para que suas coordenadas sejam representadas por um número complexo. Depois aplica-se a transformada Discreta de Fourier (DFT/FFT). A compressão deste algoritmo é feita com base na limiarização, ou seja, zerando os componentes de magnitude de frequência que possuem valores próximos de zero. A figura 3.3 mostra a magnitude do espectro de uma carta dinamoétrica, observa-se que os componentes da posição 20 a 83 estão muito próximos de zero e podem ser descartados de forma a não prejudicar a visualização da carta quando esta for reconstruída.

Uma vantagem deste algoritmo de compressão consiste no maior controle quanto a taxa de compressão que se deseja aplicar sobre a carta. Devido este possuir diretamente o controle no descarte de componentes de frequência, ou seja, caso se deseje uma compressão de 80% para uma carta original que contém 100 pontos por exemplo, serão descartados 80 componentes de frequências mais próximos do centro da figura 3.3. Portanto neste algoritmo é possível dizer exatamente a taxa de compressão que se deseja aplicar.

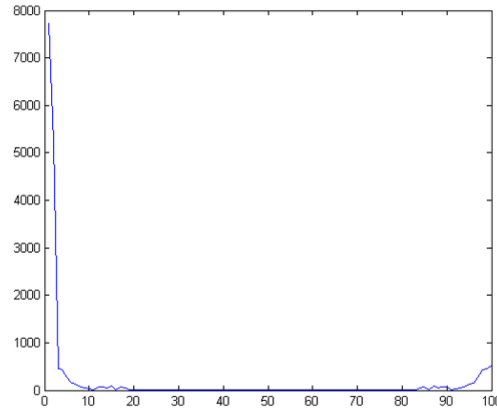


Figura 3.3: Espectro carta dinamométrica

Nas próximas subseções será descrito como funciona o processo de compressão da carta dinamométrica, assim como a reconstrução da carta simplificada resultante da compressão.

Descritores de Fourier

O descritor de Fourier é um algoritmo compacto e de baixa complexidade [Kunttu & Lepisto 2005]. Para implementar este algoritmo, considere os seguintes pontos: (x_k, y_k) , que representa as coordenadas do contorno do objeto, onde $k = 0, 1, 2, \dots, N - 1$ e N é a quantidade total de pontos. A equação 3.3 indica a função complexa das coordenadas do contorno do objeto.

$$z_k = x_k + j(y_k) \quad (3.3)$$

As coordenadas de cada ponto que representa a carta dinamométrica são adaptadas para serem representadas por um número complexo. Onde cada ponto de coordenada (x, y) será representado por um número complexo z , sendo x corresponde a posição da haste polida e y é a carga do sistema para cada posição da haste polida, ou seja, a matriz bidimensional de dimensão $2 \times N$ que representa a carta é transformada para um vetor complexo de tamanho N .

Os descritores de Fourier são obtidos aplicando a transformada Discreta de Fourier conforme a equação 3.4 sobre o vetor complexo que representa a carta, descrito pela equação 3.3. A transformada Discreta de Fourier pode ser aplicada através de dois algoritmos que são: o *Discrete Fourier Transform* (DFT) e o *Fast Fourier Transform* (FFT). Recomenda-se utilizar a FFT devido possuir menor complexidade computacional.

$$F_n = \sum_{k=0}^{N-1} z_k \cdot e^{-j2\pi nk/N} \quad (3.4)$$

Reconstrução da carta dinamométrica

Como demonstrado, o processo de compressão da carta dinamométrica é feito no domínio da frequência. Os descritores de Fourier são armazenados no banco de dados e somente são reconstruídos quando necessário visualizar a carta simplificada. Para reconstrução do sinal que representa a carta dinamométrica é necessário saber quantos pontos existiam originalmente na carta.

A reconstrução é feita preenchendo com zeros exatamente nas posições dos componentes de frequência que foram descartadas no processo de limiarização e realizando a transformada discreta de Fourier inversa conforme a equação 3.5. Como resultado da transformada inversa será um sinal no domínio do tempo representado por um vetor de números complexos. Para obter a reconstrução da carta simplificada basta transformar o vetor de números complexos obtidos novamente para uma matriz bidimensional, convertendo cada número complexo do vetor na coordenada de um ponto (x, y) da carta.

$$z_n = \frac{1}{N} \sum_{k=0}^{N-1} F_k \cdot e^{j2\pi nk/N} \quad (3.5)$$

3.2.5 *Swinging Door Trending* aplicado a carta dinamométrica

Para aplicar o algoritmo *Swinging Door* a carta dinamométrica não foi necessário realizar nenhuma adaptação na estrutura do algoritmo, permanecendo os mesmos passos descritos na subseção 2.1.2. A única modificação é que quando este algoritmo é aplicado a séries temporais o eixo horizontal (x) do gráfico é o tempo, e o eixo vertical (y) é o valor da série correspondente e no caso de cartas dinamométricas a coordenada x representa a posição das hastes e o y a carga. Embora as cartas dinamométricas apresentem representação bidimensional (forma) diferente das séries temporais, de acordo com os testes feitos o *Swinging Door* se apresentou perfeitamente aplicável.

A principal vantagem do *Swinging Door* está relacionada as operações de compressão serem feitas no domínio do tempo. E os dados das cartas dinamométricas resultantes da compressão armazenados no banco de dados poderem ser recuperados sem necessidade de realizar nenhuma transformação, ou seja, os dados das cartas são recuperados exatamente como são exibidos.

3.2.6 Métrica de similaridade entre curvas

Como forma de medir a similaridade entre cartas dinamométricas originais e simplificadas (resultantes da compressão) será utilizado o algoritmo *Discrete Dynamic Time Warping Distance*. Pois o mesmo é uma ferramenta adequada para medir similaridades de curvas de formato bidimensional representadas por uma diferente quantidade de pontos (representativos das coordenadas dos vértices da curva), pois as cartas simplificadas resultantes da compressão utilizando o algoritmo *Swinging Door* apresentam uma quantidade de pontos diferentes da original. No caso das cartas simplificadas resultantes da compressão utilizando os Descritores de Fourier apresentam a mesma quantidade de pontos da carta original, devido as transformações feitas.

Discrete Dynamic Time Warping Distance

O algoritmo Dynamic Time Warping (DTW) foi primeiramente proposto na década de 60 como métrica de similaridade entre sinais de voz [Rabiner & Juang 1993]. Desde então, tem sido bastante utilizado como métrica de similaridade entre séries temporais que podem variar em tempo e velocidade. Outras aplicações estão relacionadas ao reconhecimento de voz e o reconhecimento de assinaturas e escritas manuais, assim como medição de similaridades entre formas bidimensionais. No geral, qualquer dado que pode ser representado como uma sequência linear pode ser analisado com o DTW.

O funcionamento do DTW como métrica de similaridade entre curvas se baseia no cálculo representativo da distância entre as curvas, realizando o alinhamento entre as mesmas. Ou seja, dado duas curvas f e g o algoritmo realiza o alinhamento e cálculo da distância entre as curvas $d(f, g)$. Observando a figura 3.4 é possível visualizar um exemplo de alinhamento entre curvas discretas no tempo. Basicamente, o algoritmo realiza o alinhamento entre as curvas com base na combinação dos pontos mais próximos entre as curvas, através de uma métrica de distância. Após feito o alinhamento considerando a minimização das distâncias entre os pontos das curvas, o algoritmo realiza um somatório dessas distâncias, retornando um valor representativo de similaridade, ou seja, quanto maior o valor, menos similar são as curvas entre si. A equação 3.6 mostra a representação matemática do cálculo de distância entre duas curvas f e g , da qual se baseia o algoritmo. A distância euclidiana foi utilizada como métrica de distância entre pontos.

O uso deste método de alinhamento de curvas possibilita comparar curvas com diferentes dinâmicas, números de pontos e taxas de variação.

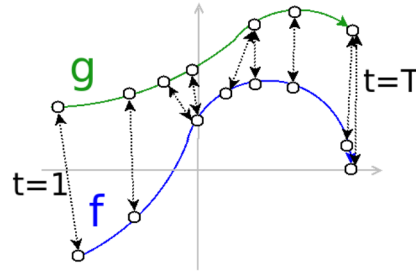


Figura 3.4: Alinhamento entre curvas usando o algoritmo DTW

```

int DTWDistance(s: array [1..n], t: array [1..m]) {
    DTW := array [0..n, 0..m]

    for i := 1 to n
        DTW[i, 0] := infinity
    for i := 1 to m
        DTW[0, i] := infinity
    DTW[0, 0] := 0

    for i := 1 to n
        for j := 1 to m
            cost := d(s[i], t[j])
            DTW[i, j] := cost + minimum(DTW[i-1, j],
                                         DTW[i, j-1],
                                         DTW[i-1, j-1])

    return DTW[n, m]
}

```

Figura 3.5: Exemplo código algoritmo DTW.

$$d(f, g) = \sum_{i=1}^N \min \left(\sum_{j=1}^M |f_i - g_j| \right) \quad (3.6)$$

Para mais detalhes do funcionamento do algoritmo DTW, a figura 3.5 mostra o código fonte simplificado do mesmo. Este código consiste de uma função que recebe como parâmetros as curvas s e t e retorna um valor de distância representando a métrica de similaridades entre as duas curvas. A função interna $d(s[i], t[j])$ retorna a distância euclidiana entre o ponto da curva $s[i]$ e o ponto da curva $t[j]$, e a função $\text{minimum}(x1, x2, x3)$ retorna o menor valor entre os parâmetros $x1$, $x2$ e $x3$.

Capítulo 4

Resultados

Neste capítulo são apresentados os resultados da dissertação no que diz respeito a três contribuições principais: a análise comparativa entre os algoritmos *Adaptive Swinging Door Trending* e o *Swinging Door Trending*, em termos de compressão de dados aplicados a séries temporais; um estudo de compressão aplicada a cartas dinamométricas de fundo de poço e a implementação do suporte a análise de cartas dinamométricas pela suíte BR-Plant.

4.1 Ferramentas Utilizadas

Para os experimentos da seção 4.2, os algoritmos foram implementados na linguagem de programação *Java*, utilizando como IDE de desenvolvimento o *NetBeans*. Para os experimentos efetuados nos estudos de casos da seção 4.3, foi utilizado a ferramenta *Matlab R2008a*.

4.2 Compressão aplicada a séries temporais

Com o intuito de se validar a eficiência e usabilidade do algoritmo *Adaptive Swinging Door Trending* foram destacados determinados cenários onde o algoritmo clássico *Swinging Door Trending* é utilizado como referência comparativa. Tais cenários foram abordados em dois diferentes estudos de casos, explanados detalhadamente em conjunto com seus resultados nas subseções 4.2.1 e 4.2.2. Estes cenários foram feitos com base em experimentos com dados reais representativos de séries temporais obtidos da indústria de petróleo e gás.

Foram usados como critérios de avaliação de desempenho dos algoritmos duas métricas, que são: o erro médio quadrático do sinal reconstruído da compressão com o sinal original, calculado conforme a Equação 4.1 e a taxa de compressão.

$$e = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (4.1)$$

4.2.1 Estudo de Caso 1

O primeiro estudo de caso analisado foi realizado com o intuito de medir a eficiência do algoritmo proposto ao comprimir sinais com bruscas mudanças no ponto de operação (*setpoint*). Para análise deste estudo de caso foram utilizados dados de volume de uma variável de processo real não comprimida ao longo do tempo, ocorrendo mudanças de *setpoint* por meio de um controlador.

Para o algoritmo SDT, foi variado o parâmetro de desvio de compressão, aumentando o mesmo gradativamente, aplicando o algoritmo aos dados da variável de processo e observando a taxa de compressão. Assim, este parâmetro inicia em 0 e é incrementado até atingir a taxa de compressão de 93,33%, convergindo para o valor de 14. Com isso, obteve-se para este parâmetro e taxa de compressão um erro de 1160.

O mesmo experimento foi feito para o algoritmo ASDT considerando o parâmetro de janela de 10, e variando o desvio de compressão até se atingir uma taxa de compressão próxima do SDT. Com isso obteve-se para o parâmetro adaptativo 0,1 uma taxa de compressão de 93,64% com erro de 121,9. O gráfico da figura 4.1 mostra uma análise da compressão, comparando o sinal original e o reconstruído do algoritmo ASDT. Este gráfico é bastante similar ao resultado proveniente do algoritmo SDT, não sendo perceptível diferenças do ponto de vista visual, por isso não foi disponibilizado para este algoritmo.

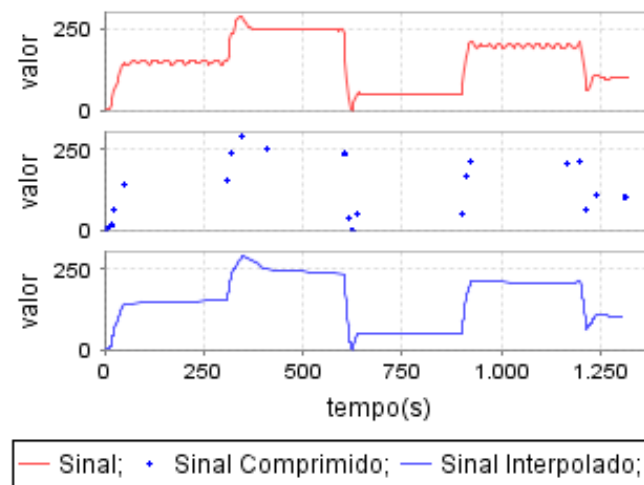


Figura 4.1: Sinal original sem compressão; Sinal comprimido através do ASDT; Sinal reconstruído.

Comparando os dois algoritmos percebe-se que o ASDT mostra-se mais adequado para este experimento com variações no *setpoint* em caso de altas taxas de compressão, pois este apresenta um erro 9,5 vezes menor que o SDT e praticamente mesma taxa de compressão.

Para uma análise mais detalhada foi realizado uma variação no parâmetro de desvio de compressão do algoritmo SDT e no parâmetro adaptativo do algoritmo ASDT, aumentando gradativamente estes parâmetros com o intuito de avaliar o comportamento destes algoritmos para esta mesma variável de volume com variações de *setpoint*.

Observando a figura 4.2 resultante da variação dos parâmetros descritos no parágrafo anterior, percebe-se a variação do erro a medida que aumenta a taxa de compressão, concluindo-se que para taxas de compressão acima de 90% o algoritmo SDT apresenta maior erro, mostrando-se que o ASDT apresenta melhor desempenho para altas taxas de compressão com erro menor em relação ao SDT.

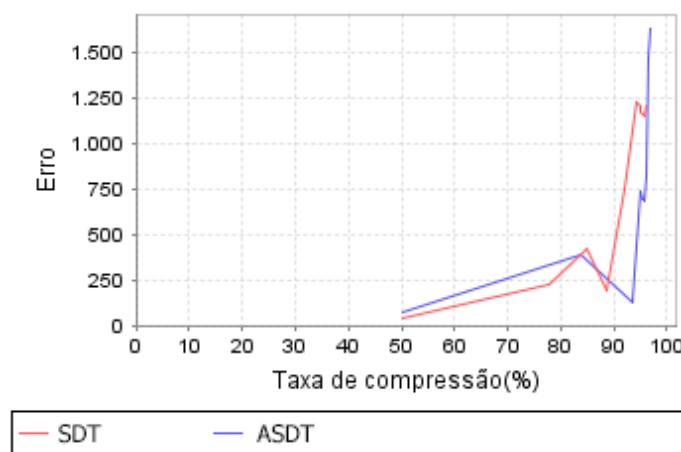


Figura 4.2: Erro dos algoritmos de compressão ASDT e SDT a medida que se aumenta a taxa de compressão.

4.2.2 Estudo de Caso 2

Por fim, o segundo estudo de caso foi realizado com o objetivo de avaliar o algoritmo proposto em termos de capacidade de generalização na eficiência de compactação, quando aplicado a sinais de diferentes naturezas com os mesmos parâmetros de configuração. Como o algoritmo clássico *Swinging Door* não é adaptativo, a tendência é que o mesmo não se comporte bem quando sinais com características diferentes são compactados utilizando o mesmo parâmetro fixo.

Para este estudo de caso foram utilizados dois sinais de naturezas diferentes do ponto

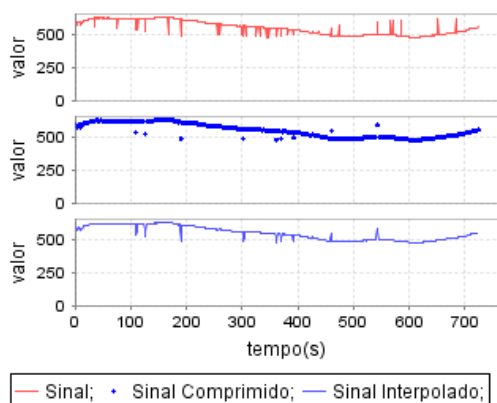


Figura 4.3: Análise algoritmo SDT aplicado a variável FI-240051

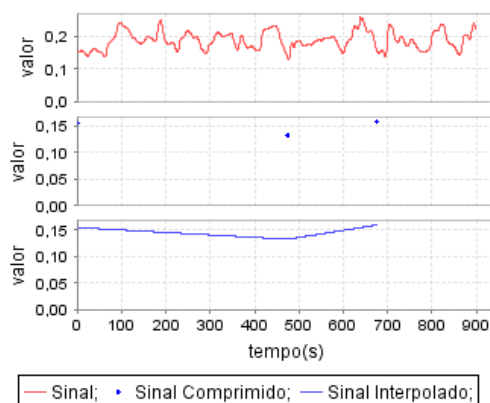


Figura 4.4: Análise algoritmo SDT aplicado a variável fração molar etano

de vista de amplitude e variação do sinal, que no caso foi a variável de processo simulada *fração molar do etano no GLP* (Gás Liquefeito de Petróleo), e a variável real FI-240051.

A tabela 4.1 mostra a taxa de compressão do algoritmo SDT para as duas variáveis de processo consideradas, utilizando o mesmo desvio de compressão fixo de 0,1.

Tabela 4.1: Taxa de compressão por variável para o algoritmo SDT.

Variável	Compressão (%)
Fração molar	99,67%
FI-240051	53,52%

A figura 4.3 mostra o sinal real, comprimido e reconstruído da variável de processo FI-240051 aplicada ao algoritmo SDT utilizando desvio de compressão de 0,1.

A figura 4.4 mostra o sinal real, comprimido e reconstruído da variável de processo simulada *fração molar do etano*, aplicada ao algoritmo SDT utilizando desvio de compressão de 0,1.

Comparando as figuras 4.3 e 4.4, percebe-se que o desvio de compressão utilizado no sinal FI-240051 é inadequada na variável simulada *fração molar do etano*, visto que o sinal reconstruído da figura 4.4 não é capaz de representar o sinal real, demonstrando que o parâmetro fixo de desvio de compressão do SDT adequado para uma variável de compressão pode ser inadequado para uma outra variável.

A tabela 4.2 a seguir mostra a taxa de compressão do algoritmo ASDT para as duas variáveis de processo consideradas, utilizando o parâmetro adaptativo 0,001 e período de 10 amostras.

As figuras 4.5 e 4.6 são resultantes da aplicação do algoritmo ASDT nas variáveis de processo FI-240051 e *fração molar de etano*, respectivamente. Embora estes sinais

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO41

Tabela 4.2: Taxa de compressão por variável para o algoritmo ASDT.

Variável	Compressão (%)
Fração molar	79,78%
FI-240051	63,03%

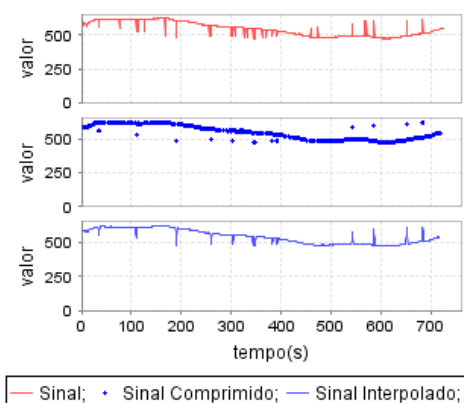


Figura 4.5: Análise algoritmo ASDT aplicado a variável FI-240051

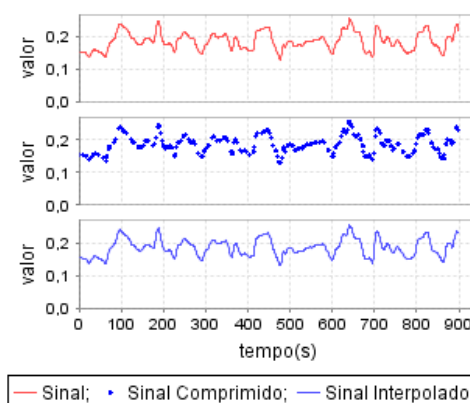


Figura 4.6: Análise algoritmo ASDT aplicado a variável fração molar etano

sejam diferentes, o ASDT consegue realizar uma taxa de compressão menor que o SDT, mas com maior facilidade de adaptação, visto que os dois sinais apresentaram boa taxa de compressão com aproximação do sinal reconstruído bastante próxima do real. Não ocorrendo perda significativa do sinal.

4.3 Compressão aplicada a cartas dinamométricas de fundo de poço

Esta seção está dedicada a apresentar os resultados da comparação de desempenho dos algoritmos de compressão *Swinging Door Trending* e Limiarização com base nos Descritores de Fourier (DFT/FFT), utilizando as métricas descritas na subseção 4.3.1.

O propósito é analisar e identificar qual algoritmo proporciona melhor custo benefício em termos de taxa de compressão e erro de reconstrução da carta dinamométrica, em quais cenários estes são mais adequados e quais as vantagens e desvantagens de cada um deles.

4.3.1 Métricas de comparação

Taxa de compressão

Métrica representada neste trabalho em percentual (de 0 a 100) conforme a equação 4.2, utilizada com o intuito de medir o grau de compressão da carta simplificada (resultante da compressão) com relação a carta original. Na equação 4.2, T_x representa a taxa de compressão, N_t o número total de pontos da carta original e N_c o número total de pontos da carta comprimida.

$$T_x = \frac{N_t - N_c}{N_t} \times 100 \quad (4.2)$$

Erro de reconstrução

Como forma de medição de similaridade entre curvas poligonais representadas por um conjunto de vértices, neste caso particular serão as cartas dinamométricas, será utilizado o algoritmo *Discrete Dynamic Time Warping Distance*. Este algoritmo retorna um valor referente a distância entre duas curvas poligonais, neste caso a carta original e a carta simplificada, quanto maior este valor, menos similar são as cartas entre si, permitindo se quantificar como uma métrica de erro.

4.3.2 Estudo de caso 1

Este estudo de caso tem por objetivo principal realizar uma comparação entre os algoritmos de compressão *Swinging Door* e os Descritores de Fourier (FFT) aplicados a cartas dinamométricas de fundo de poço, analisando a máxima taxa de compressão que se pode atingir sem gerar comprometimento visual da carta simplificada resultante da compressão. Com isso será feita uma análise de erro e taxa de compressão para as principais referências de cartas que representam os modos de operação dos sistemas de bombeio mecânico em campos de petróleo. Visto que cada modo de operação possui uma carta com formato característico, tais formatos foram levados em consideração na análise.

Os modelos das cartas dinamométricas de referências utilizados estão sendo representados por gráficos bidimensionais onde o eixo x é a posição da haste e o eixo y o torque aplicado. Os valores que representam as coordenadas de cada ponto na carta foram normalizados de 0 a 100, para facilitar a análise.

Para cada modelo de referência de carta, este foi submetido aos algoritmos variando seus parâmetros de modo a obter taxas de compressão de 50%, 60%, 70%, 80% e 90%. Para cada taxa de compressão foi medido o erro considerando a carta simplificada e a

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO43

carta original utilizando como métrica o algoritmo *Discrete Dynamic Time Warping*. As próximas subseções mostram os resultados dos experimentos feitos para cada carta representativa do modo de operação característico.

Normal

Observando a figura 4.7 é possível visualizar gráficos da carta original e a carta simplificada (submetida ao algoritmo *Swinging Door*), para taxas de compressão variando de 50 a 90% para o modelo de carta caracterizando o modo de operação Normal. A taxa de compressão máxima atingida sem gerar comprometimento visual foi de 80%, a partir de 90% a carta original comparada com a simplificada apresenta mudanças significativas que podem comprometer um diagnóstico futuro.

Para efeito de comparação a figura 4.8 mostra os resultados da compressão aplicando o algoritmo de Descritores de Fourier, sendo possível observar que se atingiu uma taxa de compressão de até 90% sem gerar comprometimento visual.

A figura 4.9 mostra a variação do erro com relação a taxa de compressão, para o algoritmo *Swinging Door* (SDT, com legenda em azul) comparando com o algoritmo de Descritores de Fourier (FFT, com legenda em vermelho). Para o algoritmo SDT observa-se que a medida que se aumenta a taxa de compressão ocorre variação do erro de 3,66 a 5,36. Já o algoritmo FFT mostra variação de erro de 0 a 4,64.

Com base nos resultados expostos percebe-se que o algoritmo de Descritores de Fourier para carta modelo de modo de operação Normal apresenta melhor desempenho, conseguindo atingir uma taxa de compressão superior (90%), sem gerar comprometimento visual, com erro significativamente menor (em todas as taxas de compressão).

Pancada de Fluido

Para a carta representativa do modo de operação de Pancada de Fluido, o algoritmo *Swinging Door* apresenta máxima taxa de compressão de 80% sem proporcionar comprometimento visual, conforme a figura 4.10.

Para efeito de comparação a figura 4.11 mostra os resultados da compressão aplicando o algoritmo de Descritores de Fourier, sendo possível observar que se atingiu uma taxa de compressão de até 80% sem gerar comprometimento visual.

A figura 4.12 mostra que o algoritmo SDT para cada taxa de compressão apresenta erro variando de 8,98 a 9,26. Para o algoritmo FFT o erro varia de 0 a 11,09.

Com base nos resultados expostos, os dois algoritmos atingem uma taxa de compressão máxima de 80% sem gerar comprometimento visual que possa afetar a interpretação.

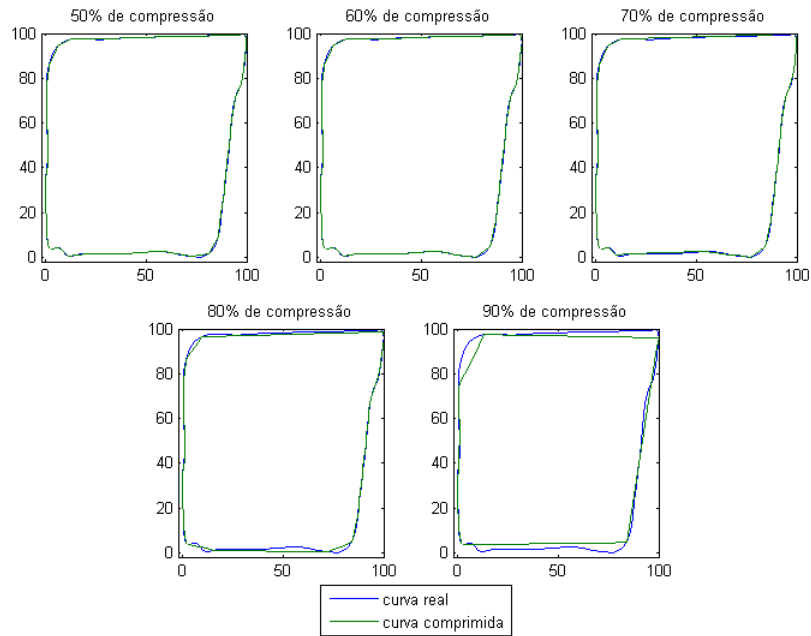


Figura 4.7: *Swinging Door* aplicado carta de operação normal.

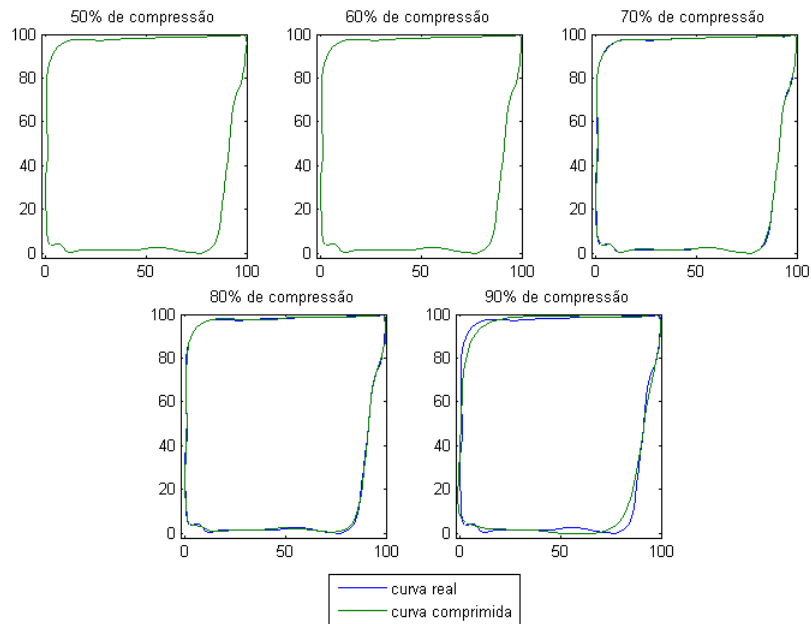


Figura 4.8: FFT aplicado a cartas de operação normal.

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO45

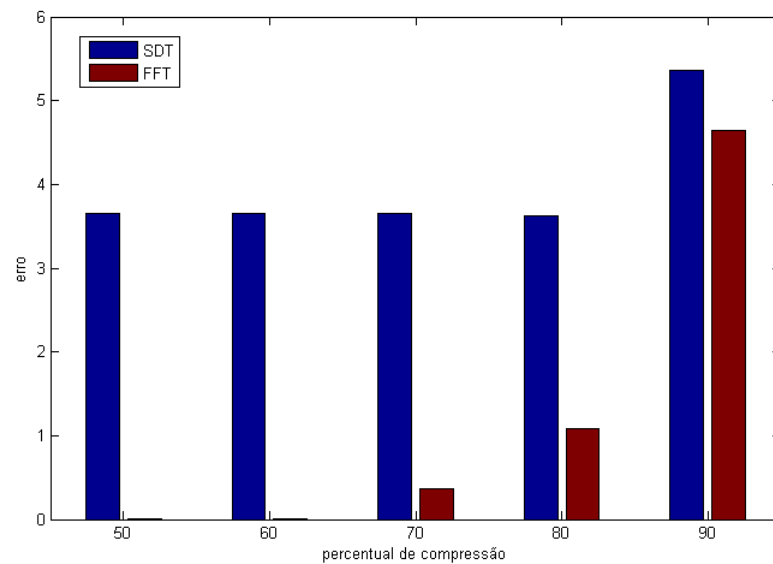


Figura 4.9: Análise de erro de compressão para carta de operação Normal.

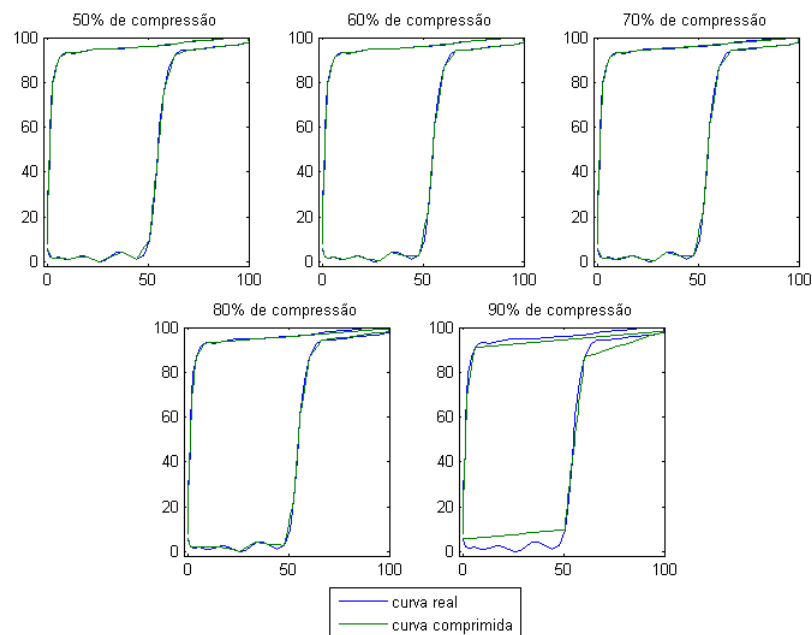


Figura 4.10: *Swinging Door* aplicado sobre carta de operação Pancada de Fluido.

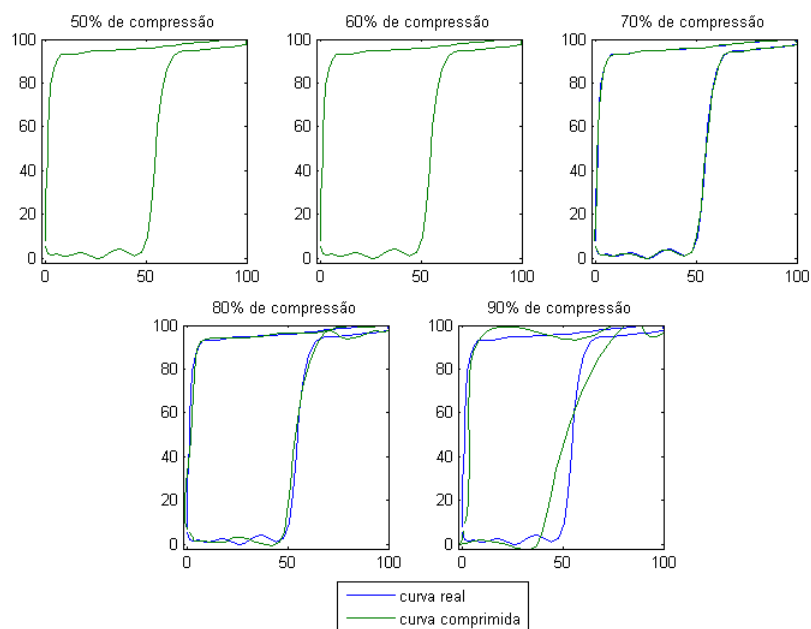


Figura 4.11: FFT aplicado sobre carta de operação Pancada de Fluido.

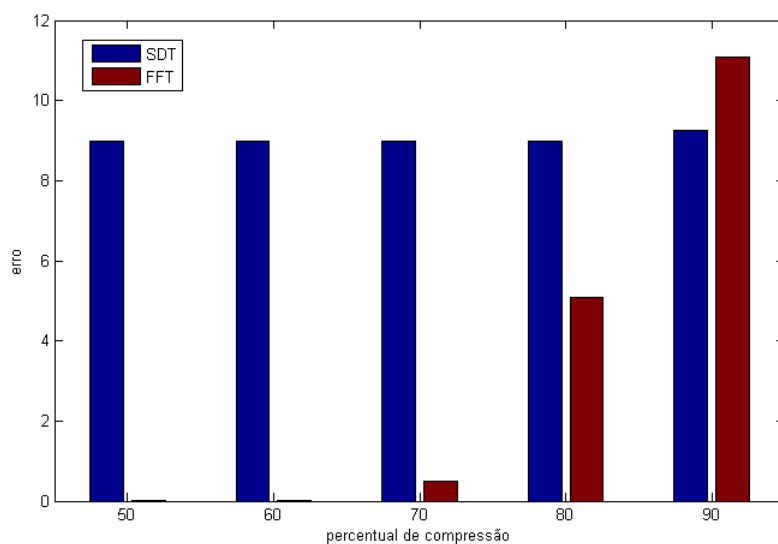


Figura 4.12: Análise de erro de compressão para carta de operação Pancada de Fluido.

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO47

Vazamento na Válvula de Passeio

Para a carta representativa do modo de operação de Válvula de Passeio, o algoritmo *Swinging Door* apresenta máxima taxa de compressão de 80% sem proporcionar comprometimento visual significativo, conforme a figura 4.13.

Para efeito de comparação a figura 4.14 mostra os resultados da compressão aplicando o algoritmo de Descritores de Fourier, sendo possível observar que se atingiu uma taxa de compressão de até 90% sem gerar comprometimento visual.

A figura 4.15 mostra que para o algoritmo SDT a variação do erro para cada taxa de compressão é de 3,53 a 9,25. Para o algoritmo FFT o erro varia de 0 a 3,27.

Com base nos resultados expostos, o algoritmo FFT apresenta taxa de compressão de 90%, com erro significativamente menor em todas as taxas de compressão com relação ao algoritmo *Swinging Door*. Portanto os Descritores de Fourier (FFT) para a carta do modo de operação do tipo Válvula de Passeio apresenta melhores resultados.

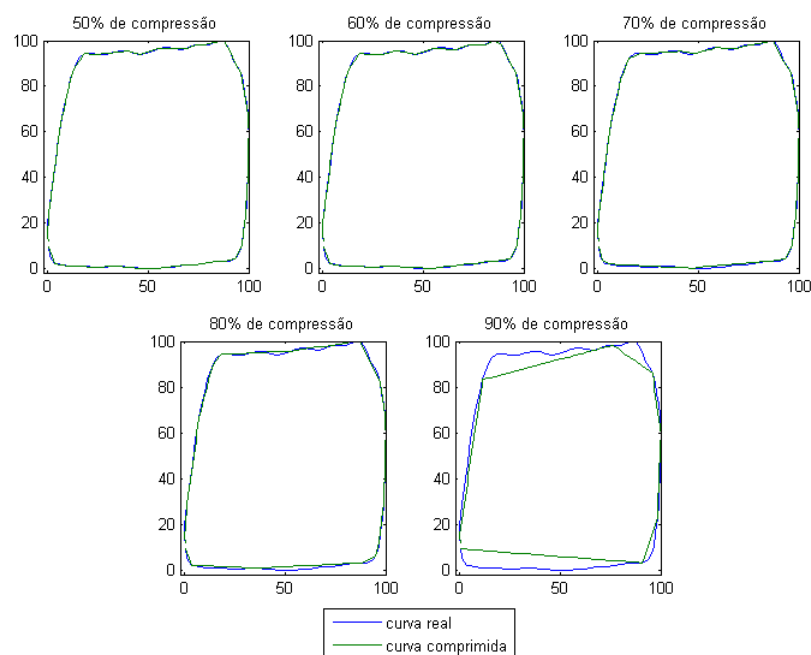


Figura 4.13: *Swinging Door* aplicado a carta de operação Válvula de Passeio.

Vazamento na Válvula de Pé

Para a carta representativa do modo de operação de Válvula de Pé, o algoritmo *Swinging Door* apresenta máxima taxa de compressão de 80% sem proporcionar comprometimento visual significativo, conforme a figura 4.16.

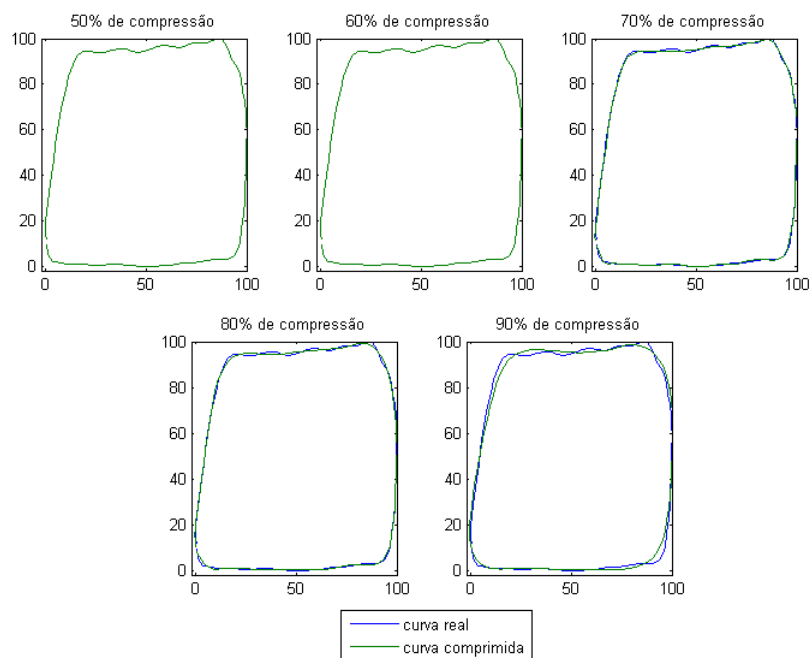


Figura 4.14: FFT aplicado a carta de operação Válvula de Passeio.

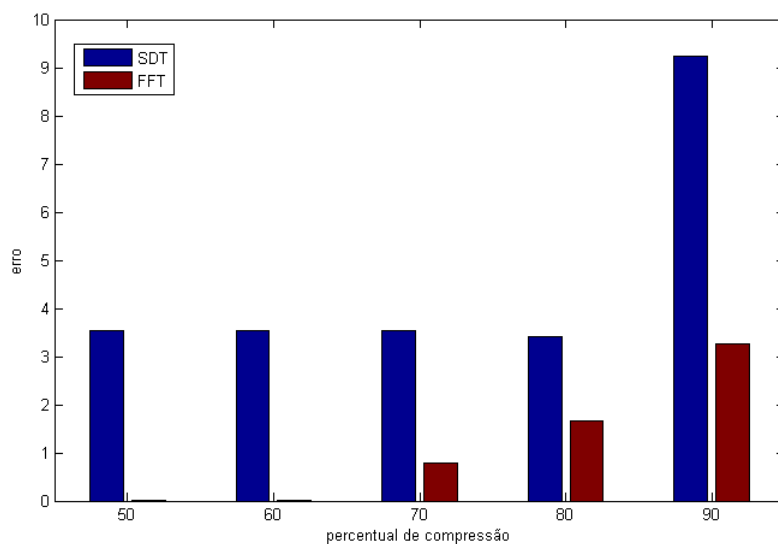


Figura 4.15: Análise de erro de compressão para carta de operação Válvula de Passeio.

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO49

Para efeito de comparação a figura 4.17 mostra os resultados da compressão aplicando o algoritmo de Descritores de Fourier, sendo possível observar que se atingiu uma taxa de compressão de até 90% sem gerar comprometimento visual.

A figura 4.18 mostra que o algoritmo SDT para cada taxa de compressão apresenta erro variando de 3,38 a 9. Para o algoritmo FFT o erro varia de 0 a 3,65.

Com base nos resultados expostos, o algoritmo FFT para a carta do modo de operação do tipo Válvula de Pé, apresenta melhores resultados, pois atingiu uma taxa de compressão máxima de 90% sem gerar comprometimento visual.

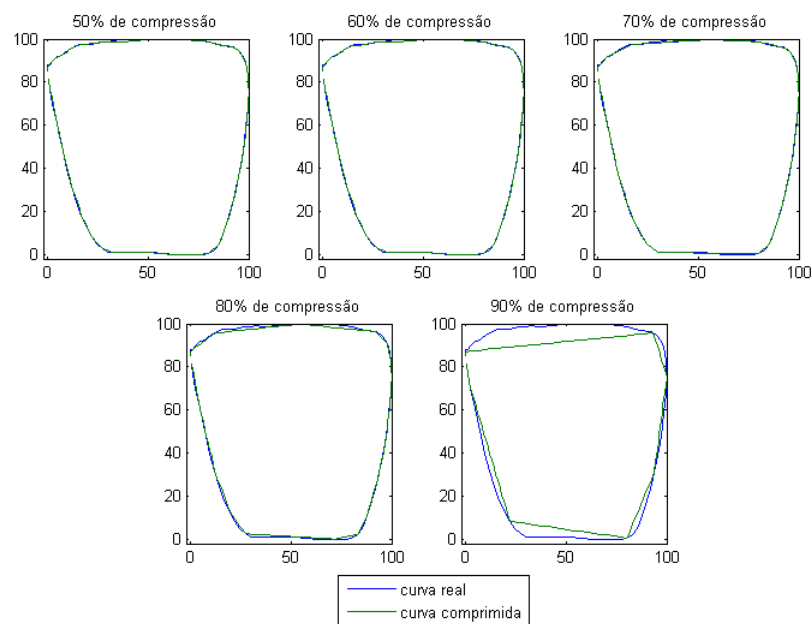


Figura 4.16: *Swinging Door* aplicado a carta de operação Válvula de Pé.

Coluna Móvel

Para o padrão de carta representativa do modo de operação Coluna Móvel, o algoritmo *Swinging Door* apresenta máxima taxa de compressão de 80% sem proporcionar comprometimento visual significativo, conforme a figura 4.19.

Para efeito de comparação a figura 4.20 mostra os resultados da compressão aplicando o algoritmo de Descritores de Fourier, sendo possível observar que se atingiu uma taxa de compressão de até 90% sem gerar comprometimento visual.

A figura 4.21 mostra que o algoritmo SDT para cada taxa de compressão apresenta erro variando de 3,65 a 7,23. Para o algoritmo FFT o erro varia de 0 a 4,73.

Com base nos resultados expostos, para o padrão de carta Coluna Móvel, o algoritmo

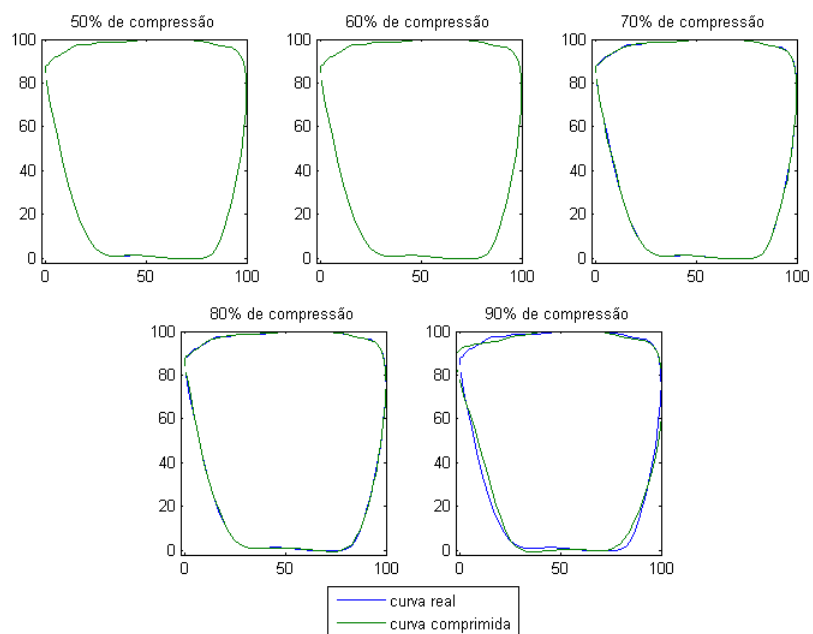


Figura 4.17: FFT aplicado a carta de operação Válvula de Pé.

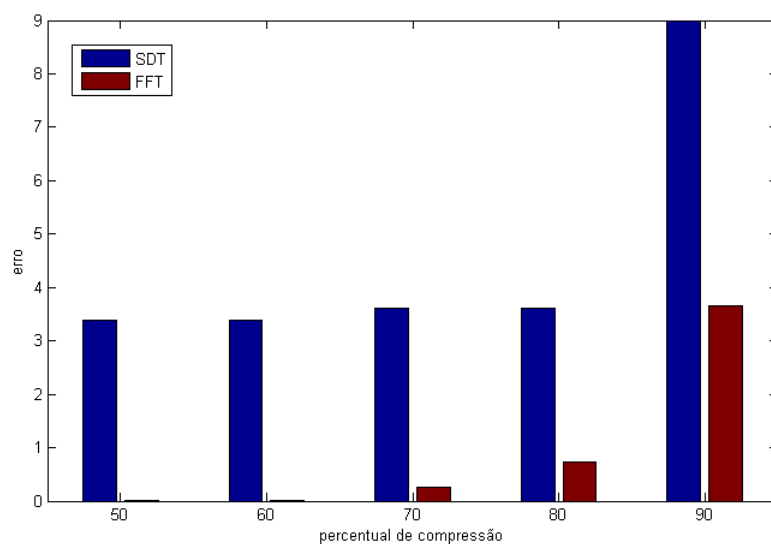


Figura 4.18: Análise de erro de compressão para carta de operação Válvula de Pé.

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO51

FFT demonstra melhor desempenho, apresentando máxima taxa de compressão de 90%, contra a de 80% do algoritmo SDT.

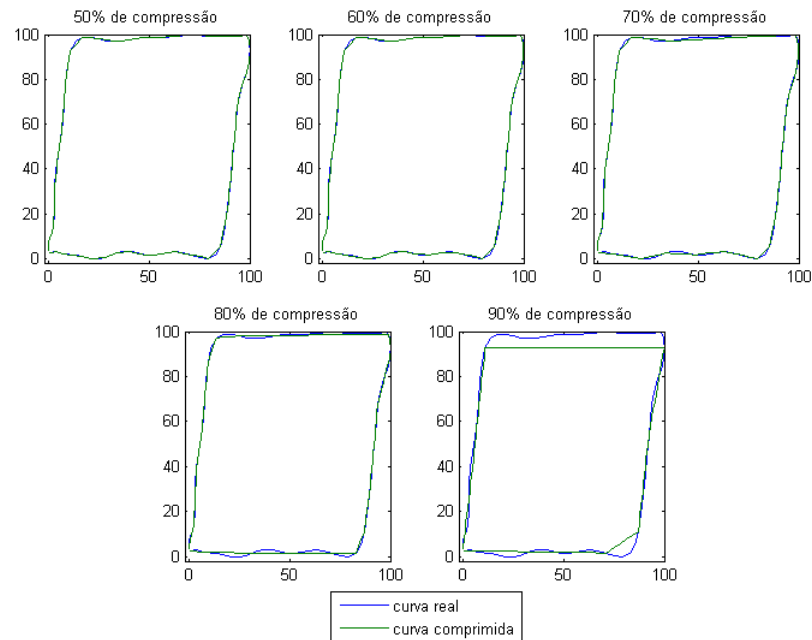


Figura 4.19: *Swinging Door* aplicado a carta de operação Coluna Móvel.

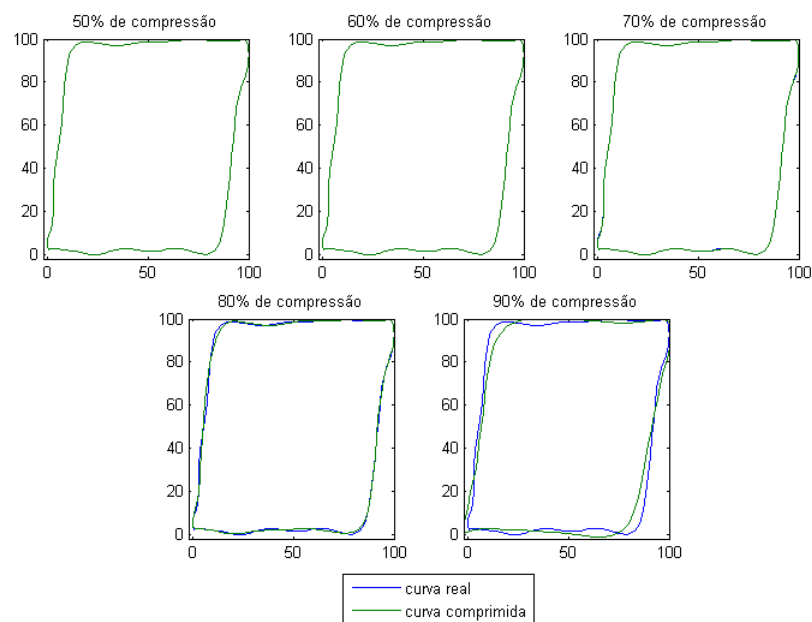


Figura 4.20: FFT aplicado a carta de operação Coluna Móvel.

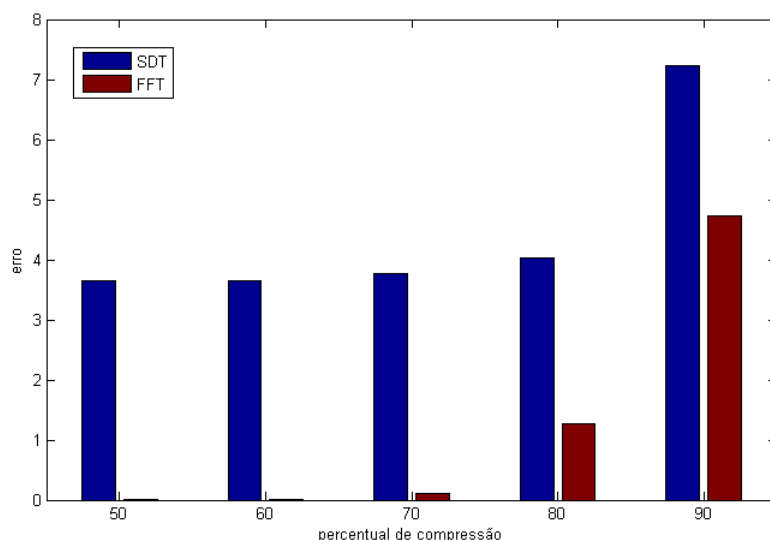


Figura 4.21: Análise de erro de compressão para carta de operação Coluna Móvel.

Batidas na Bomba de Fundo

Para o padrão de carta representativa do modo de operação de Batidas na Bomba de Fundo, o algoritmo *Swinging Door* apresenta máxima taxa de compressão de 80% sem proporcionar comprometimento visual significativo, conforme a figura 4.22.

Para efeito de comparação a figura 4.23 mostra os resultados da compressão aplicando o algoritmo de Descritores de Fourier, sendo possível observar que se atingiu uma taxa de compressão de até 90% sem gerar comprometimento visual.

A figura 4.24 mostra que o algoritmo SDT para cada taxa de compressão apresenta erro variando de 6,39 a 14,62. Para o algoritmo FFT o erro varia de 0 a 6,51.

Com base nos resultados expostos, para o padrão de carta Batidas na Bomba de Fundo, o algoritmo FFT demonstra melhor desempenho, apresentando máxima taxa de compressão de 90%, contra a de 80% do algoritmo SDT.

Interferência de Gás

Para o padrão de carta representativa do modo de operação de Interferência de Gás, o algoritmo *Swinging Door* apresenta máxima taxa de compressão de 80% sem proporcionar comprometimento visual significativo, conforme a figura 4.25.

Para efeito de comparação a figura 4.26 mostra os resultados da compressão aplicando o algoritmo de Descritores de Fourier, sendo possível observar que se atingiu uma taxa de compressão de até 90% sem gerar comprometimento visual.

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO53

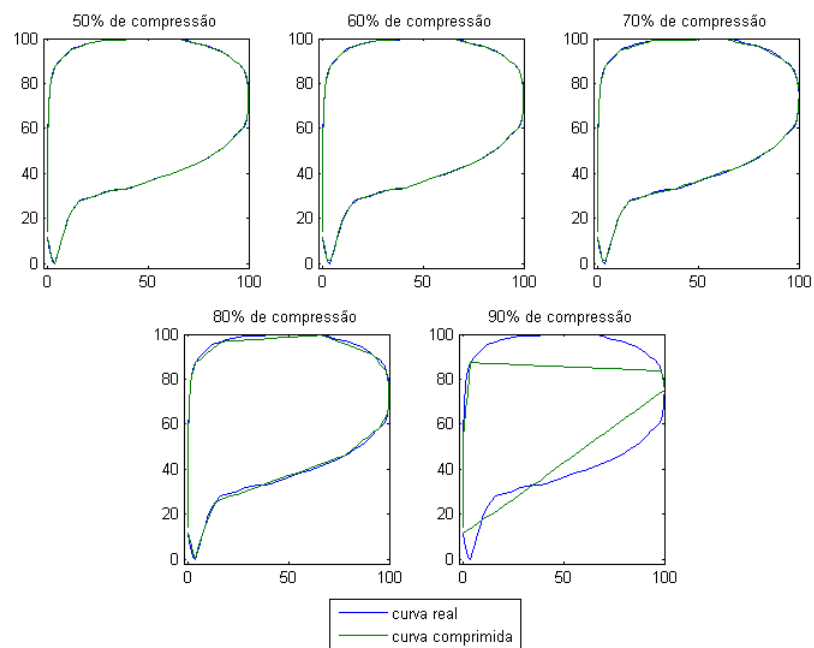


Figura 4.22: *Swinging Door* aplicado a carta de operação Batidas na Bomba de Fundo.

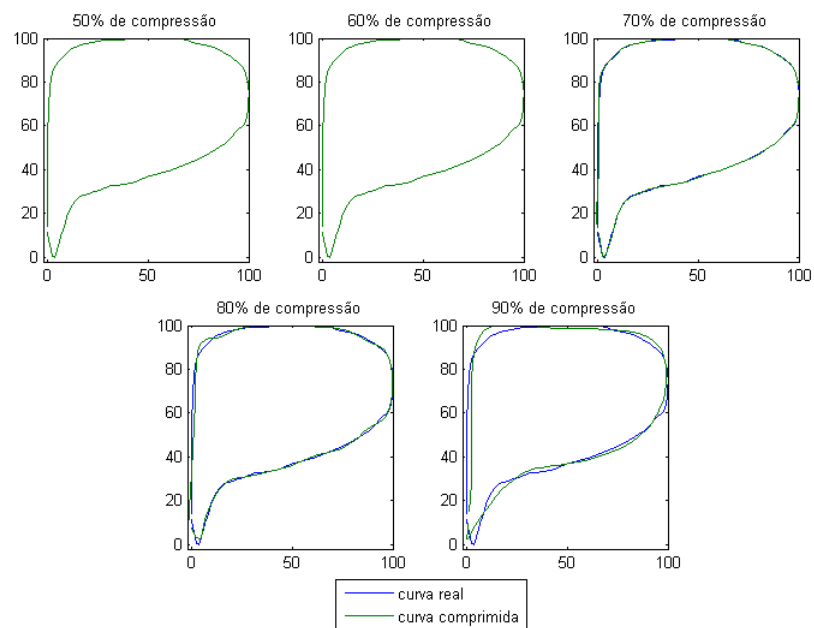


Figura 4.23: FFT aplicado a carta de operação Batidas na Bomba de Fundo.

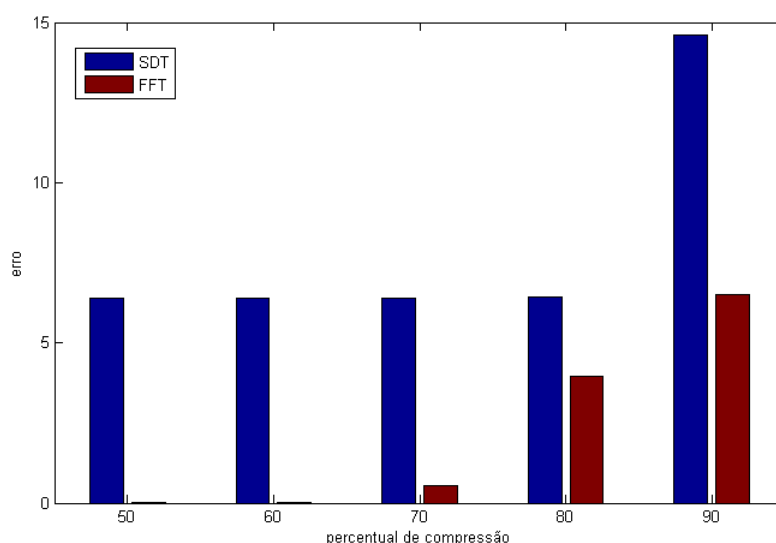


Figura 4.24: Análise de erro de compressão para carta de operação Batidas na Bomba de Fundo.

A figura 4.27 mostra que o algoritmo SDT para cada taxa de compressão apresenta erro variando de 5,36 a 8,5. Para o algoritmo FFT o erro varia de 0 a 6,45.

Com base nos resultados expostos, para o padrão de carta Interferência de Gás, o algoritmo FFT demonstra melhor desempenho, apresentando máxima taxa de compressão de 90%, contra a de 80% do algoritmo SDT.

Pistão Preso

Para o padrão de carta representativa do modo de operação de Pistão Preso, o algoritmo *Swinging Door* apresenta máxima taxa de compressão de 90% sem proporcionar comprometimento visual significativo, conforme a figura 4.28.

Para efeito de comparação a figura 4.29 mostra os resultados da compressão aplicando o algoritmo de Descritores de Fourier, sendo possível observar que se atingiu uma taxa de compressão de até 90% sem gerar comprometimento visual.

A figura 4.30 mostra que o algoritmo SDT para cada taxa de compressão apresenta erro variando de 2,94 a 3,94. Para o algoritmo FFT o erro varia de 0 a 3,65.

Com base nos resultados expostos, para o padrão de carta Pistão Preso, os dois algoritmos apresentaram excelente desempenho com taxa de compressão máxima de 90% sem comprometer a interpretação da carta.

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO55

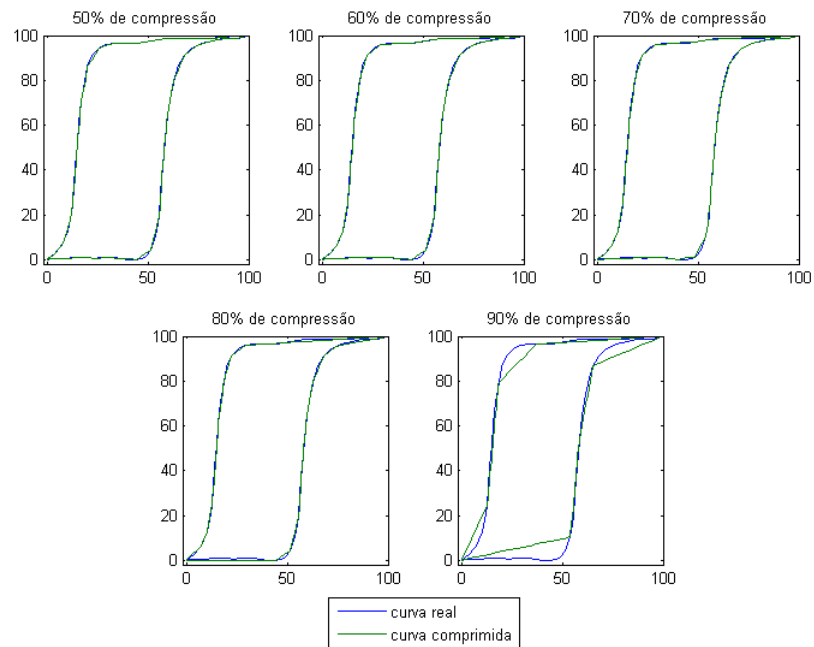


Figura 4.25: *Swinging Door* aplicado a carta de operação Interferência de Gás.

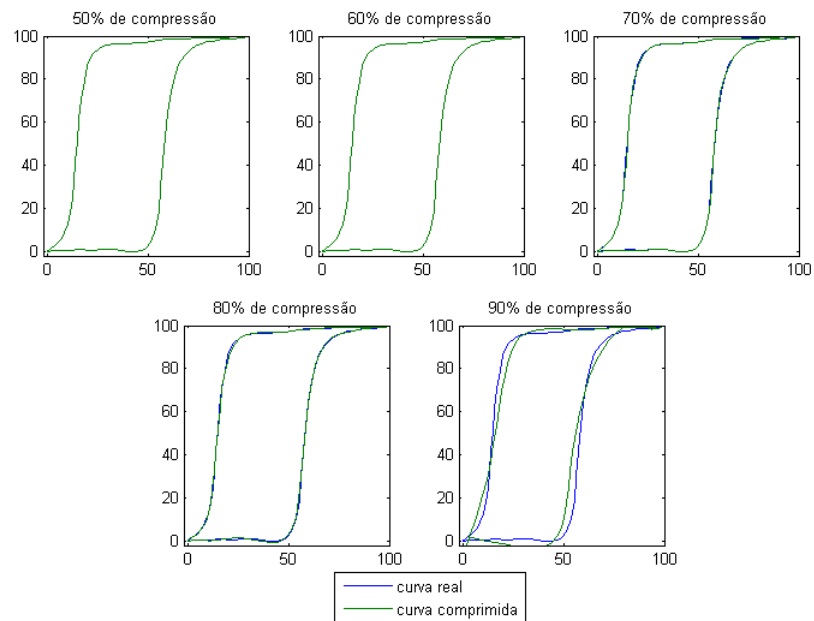


Figura 4.26: FFT aplicado a carta de operação Interferência de Gás.

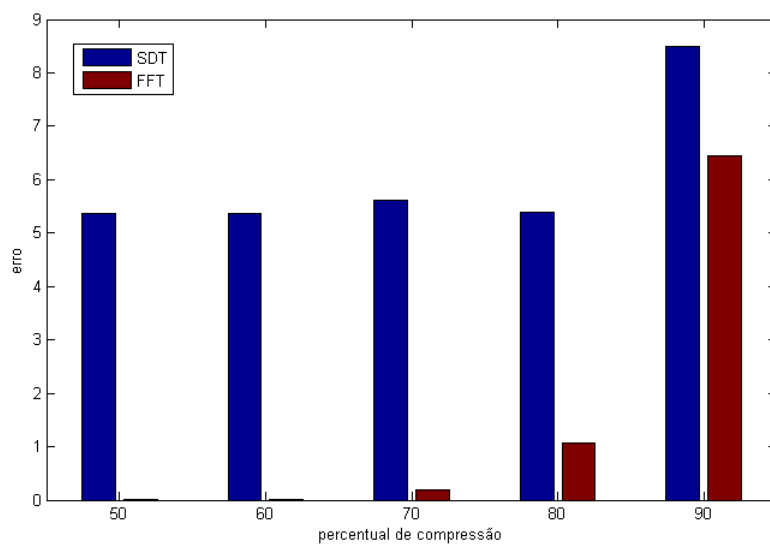


Figura 4.27: Análise de erro de compressão para carta de operação Interferência de Gás.

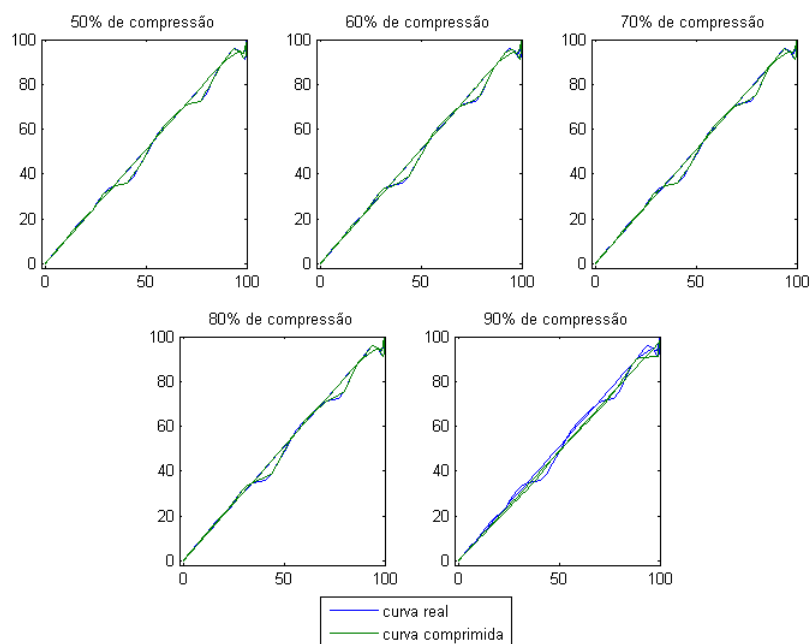


Figura 4.28: *Swinging Door* aplicado a carta de operação Pistão Preso

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO57

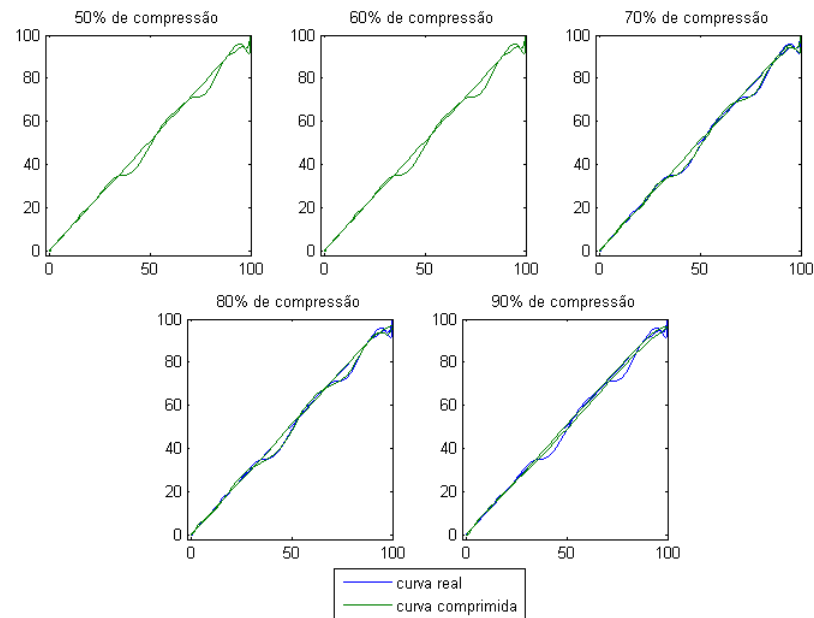


Figura 4.29: FFT aplicado a carta de operação Pistão Preso

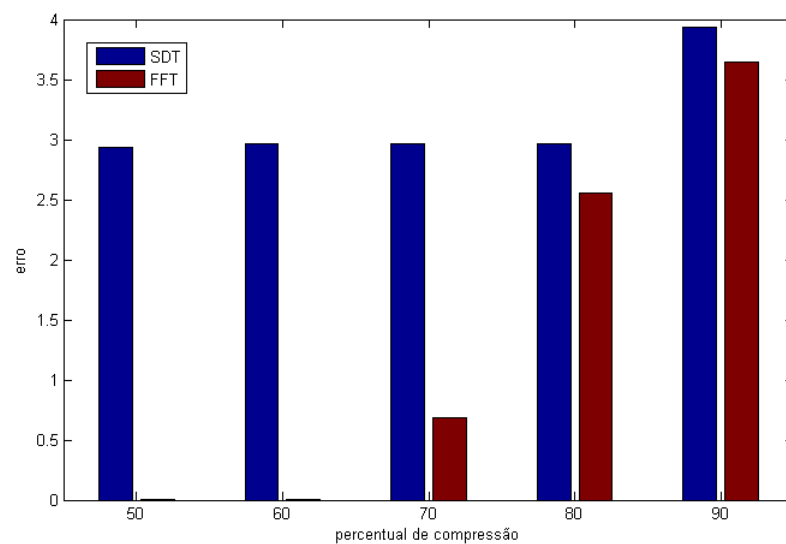


Figura 4.30: Análise de erro de compressão para carta de operação Pistão Preso.

4.3.3 Resultados Gerais do Estudo de Caso 1

Com base nos experimentos descritos na subseção 4.3.2, pode-se chegar as seguintes conclusões:

- Para todas as cartas representativas dos padrões de operação, a máxima de taxa de compressão dos algoritmos foi de 90% sem gerar comprometimento visual;
- Os dois algoritmos apresentaram excelente desempenho em todos os casos, para taxas de compressão menores ou igual a 80%;
- O algoritmo de Descritores de Fourier apresenta melhor desempenho com relação ao *Swinging Door* em quase todos os modos de operação, empatando somente com os padrões de Pancada de Fluido e Pistão Preso, sob o ponto vista de comprometimento visual.
- O algoritmo de Descritores de Fourier apresenta menor erro em todos os padrões de operação, para todas as taxas de compressão. Com exceção do padrão Pancada de Fluido que para a taxa de compressão de 90% os Descritores de Fourier apresentam erro maior.

4.3.4 Estudo de Caso 2

Este estudo de caso tem por objetivo realizar uma comparação entre os algoritmos de compressão de Descritores de Fourier e o *Swinging Door* para um conjunto de cartas dinamométricas de fundo de poço, analisando a estatística do erro das cartas originais com relação as cartas simplificadas (resultantes da compressão) para subconjuntos subdivididos por formato característico da carta, ou seja, por modo de operação.

Os experimentos foram feitos com base na análise utilizando estatística descritiva, observando a variação da variável erro para diferentes grupos de cartas (classificados por modo de operação), estes submetidos a diferentes taxas de compressão. Para isto foi utilizado o Diagrama de Caixa como ferramenta gráfica, detalhado na subseção 4.3.4, analisando a variável erro em função do fator de interesse taxa de compressão variando de 50 a 90% com amostragem de 5%.

Diagrama de Caixa (Diagrama de Extremos e Quartis)

Box plots [Chambers et al. 1983] ou diagrama de caixa é uma ferramenta para localizar e analisar a variação de uma variável dentre diferentes grupos de dados. Esta ferramenta é um gráfico em que o eixo vertical representa a variável a ser analisada e o eixo horizontal o fator de interesse.

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO59

A figura 4.31 representa um exemplo de gráfico de Diagrama de Caixa. Este gráfico representa a análise estatística de vários grupos. Cada grupo é analisado através de cinco medidas estatísticas: o limite inferior (LI), o quartil inferior (Q1) correspondente a 25% dos menores valores, a mediana (Q2), o quartil superior (Q3) correspondente a 75% dos valores observados e o limite superior (LS). A caixa é constituída pelo limite inferior (Q1), a mediana (Q2) e o quartil superior (Q3), esta caixa representa 50% dos valores observados.

O limite inferior (LI) é calculado como o maior valor entre a menor amostra e o valor x resultante da equação $x = Q1 - 1,5(Q3 - Q1)$. O limite superior (LS) é calculado como o menor valor entre a maior amostra e o valor y resultante da equação $y = Q3 + 1,5(Q3 - Q1)$. Os pontos fora destes limites são considerados valores discrepantes (*outliers*) e podem ser observados na figura como pequenos círculos.

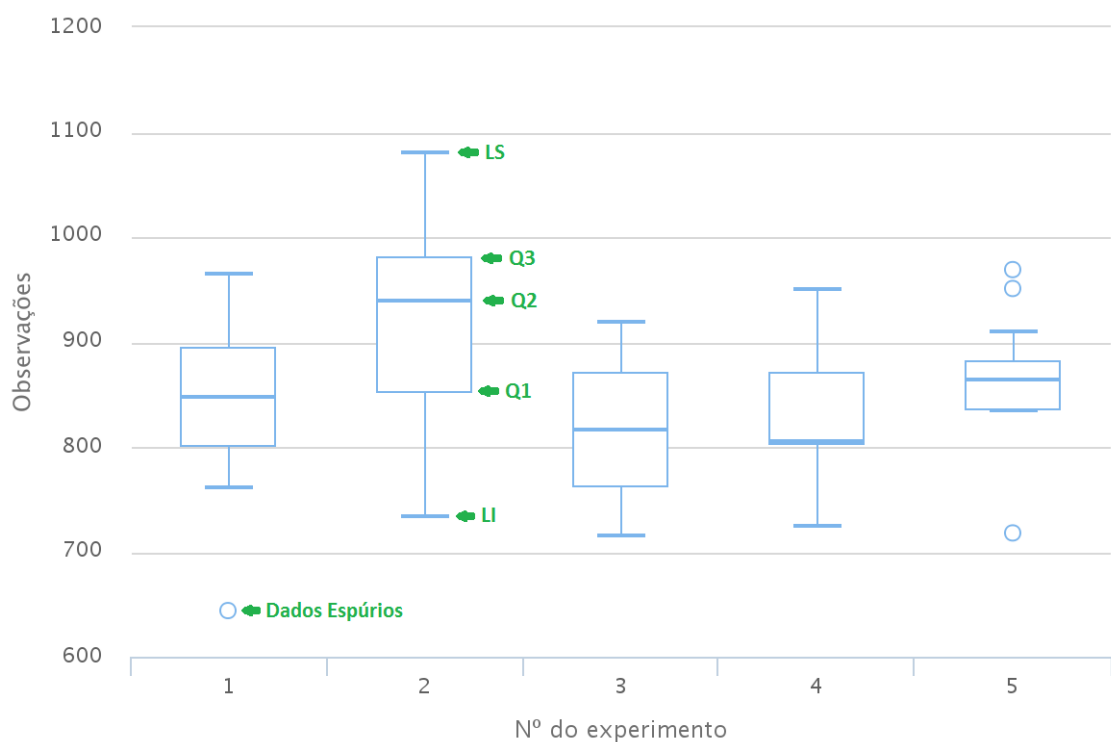


Figura 4.31: Boxplot (Diagrama de Caixa)

As próximas subseções retratam os experimentos feitos para cada subconjunto de cartas dinamométricas que caracteriza um modo de operação de bombeio mecânico.

Normal

A figura 4.32, mostra um diagrama de caixa representando a distribuição estatística da variável erro (eixo vertical) em função das taxas de compressão 50, 55, 60, 65, 70, 75, 80, 85 e 90% (eixo horizontal). Este diagrama foi construído tendo como base um conjunto de 100 cartas dinamométricas representativas do modo de operação normal, submetidas aos algoritmos de Descritores de Fourier e *Swinging Door* nas diferentes taxas de compressão mencionadas. As caixas em azul se referem ao resultado da compressão aplicando o algoritmo *Swinging Door* (SDT), e as caixas pretas se referem a aplicação do algoritmo de Descritores de Fourier (FFT).

Observando a figura, percebe-se que as caixas do algoritmo *Swinging Door* apresentam para taxas de compressão até 85% erro máximo em torno de 12, mínimo em torno de 3,5, quartil superior em torno de 8, mediana em torno de 7, quartil inferior em torno de 5. Isso implica que 50% das cartas apresentam erro entre 5 e 8, com mediana em torno de 7.

As caixas do algoritmo de Descritores de Fourier começam a apresentar erros diferentes de zero ao atingir taxa de compressão de 60%, apresentando resultados superiores ao *Swinging Door* até a taxa de 80%, a partir dos 90% o erro máximo da caixa apresenta valores maiores que o *Swinging Doors*, embora 50% das cartas incluídas entre os quartis inferior e superior apresentem erros inferiores em todos os casos.

Pancada de Fluido

Para as cartas representativas do modo de operação Pancada de Fluido, foram utilizadas no experimento um conjunto de 55 cartas. Observando a figura 4.33 é possível perceber que o algoritmo *Swinging Door* apresenta caixas similares até a taxa de compressão de 80%, a partir dos 90% este apresenta maior variação na distribuição do erro.

Para o algoritmo de Descritores de Fourier todas as caixas desconsiderando o erro mínimo e máximo apresentam desempenho superior ao *Swinging Door*. A partir dos 85% o erro máximo torna-se maior.

Interferência de Gás

Para as cartas representativas do modo de operação Interferência de Gás, foram utilizadas no experimento um conjunto de 28 cartas. Observando a figura 4.34 é possível perceber que o algoritmo *Swinging Door* apresenta caixas similares até a taxa de compressão de 85%, a partir de 90% este apresenta maior variação na distribuição do erro.

Para o algoritmo de Descritores de Fourier todas as caixas apresentam desempenho superior ao *Swinging Door*.

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO61

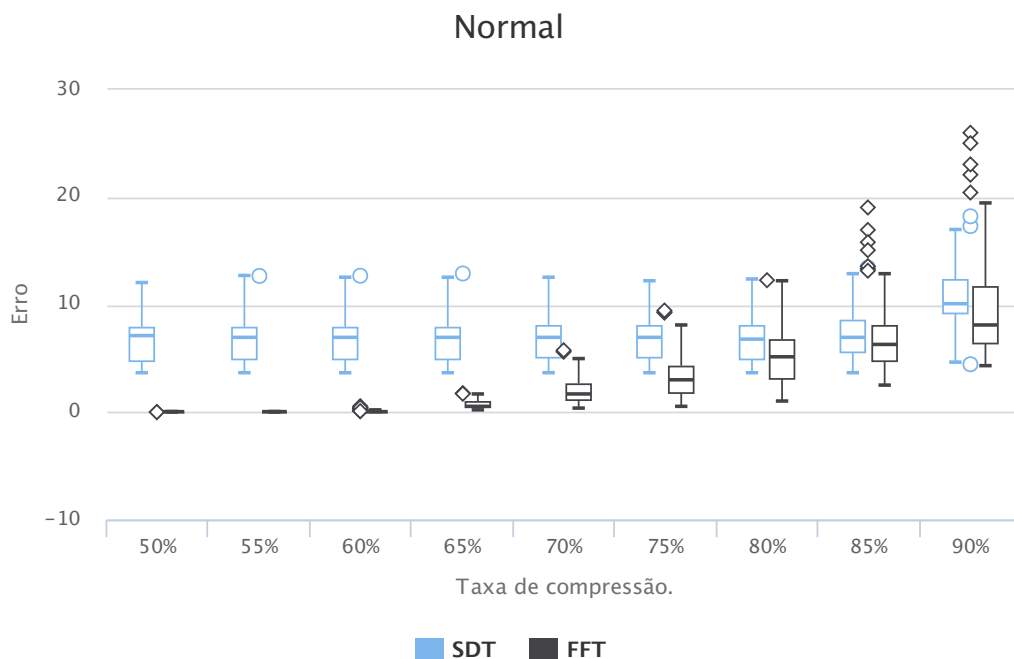


Figura 4.32: Análise da distribuição do erro para um conjunto de cartas de modo de operação Normal

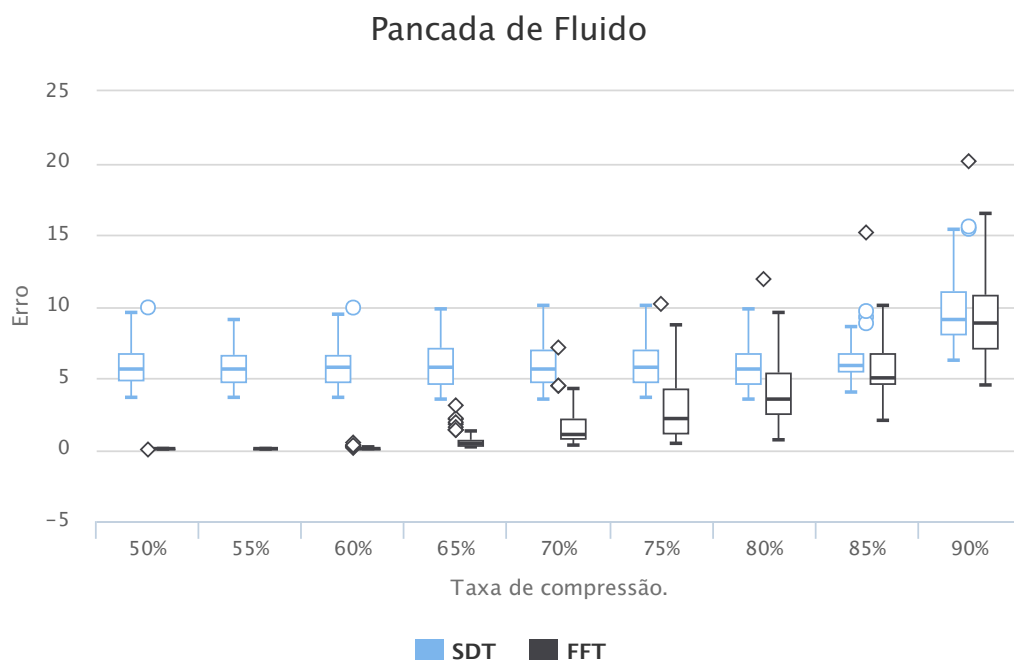


Figura 4.33: Análise da distribuição do erro para um conjunto de cartas de modo de operação Pancada de Fluido

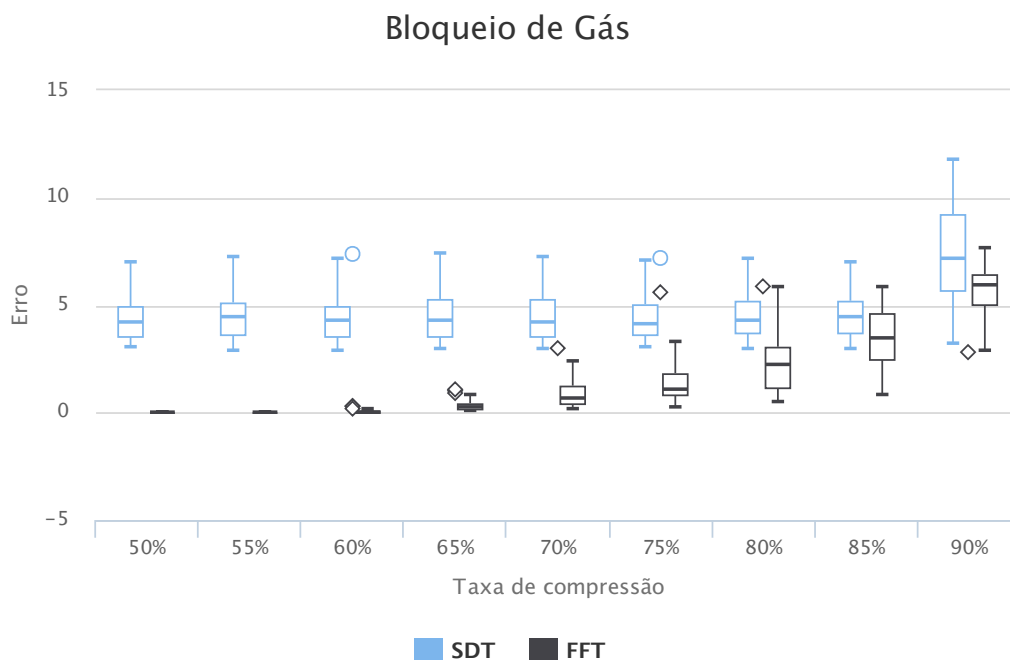


Figura 4.34: Análise da distribuição do erro para um conjunto de cartas de modo de operação Interferência de Gás

Vazamento de Válvula Passeio

Para as cartas representativas do modo de operação Vazamento de Válvula Passeio, foram utilizadas no experimento um conjunto de 13 cartas. Observando a figura 4.35 é possível perceber que o algoritmo *Swinging Door* apresenta caixas similares até a taxa de compressão de 85%, a partir de 90% este apresenta maior variação na distribuição do erro.

Para o algoritmo de Descritores de Fourier todas as caixas apresentam desempenho superior ao *Swinging Door* até 85% de compressão. A partir de 90% apresentou maior variação do erro com valor máximo de erro superior.

Vazamento de Válvula de Pé

Para as cartas representativas do modo de operação Vazamento de Válvula de Pé, foram utilizadas no experimento um conjunto de 20 cartas. Observando a figura 4.35 é possível perceber que o algoritmo *Swinging Door* apresenta caixas similares até a taxa de compressão de 85%, a partir de 90% este apresenta maior variação na distribuição do erro, com erro máximo bem maior que o erro apresentado aos 85%.

Para o algoritmo de Descritores de Fourier todas as caixas apresentam desempenho

4.3. COMPRESSÃO APLICADA A CARTAS DINAMOMÉTRICAS DE FUNDO DE POÇO63

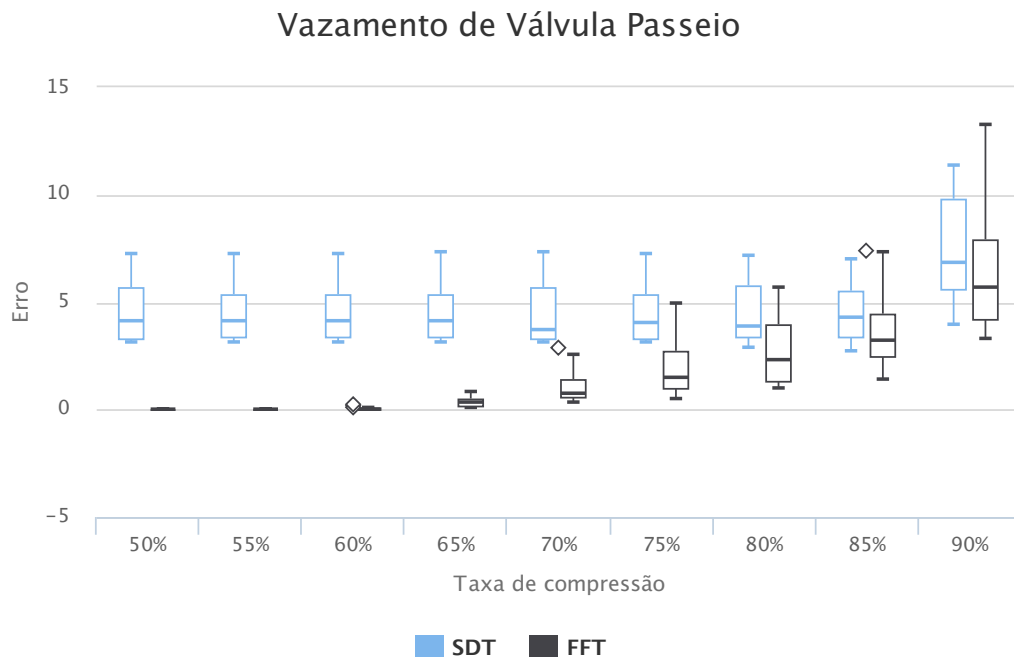


Figura 4.35: Análise da distribuição do erro para um conjunto de cartas de modo de operação de Válvula Passeio

superior ao *Swinging Door* até 80% de compressão. A partir de 85% apresentou maior variação do erro com valor máximo de erro superior.

4.3.5 Resultados Gerais do Estudo de Caso 2

Com base nos experimentos apresentados na subseção 4.3.4, pode-se concluir que para todos os modos de operação analisados, o algoritmo de Descritores de Fourier demonstra melhores resultados até a taxa de compressão de 80%, erros próximos de zero até a taxa de 65% e curva de erro da mediana apresenta comportamento exponencial com valores entre zero e 10 a medida que varia as taxas de 50% a 90%.

Com relação ao algoritmo *Swinging Door*, observa-se para todos os modos de operação que as caixas apresentam comportamento similar em termos de variação de erro para as taxas de compressão de 50% a 85%, mantendo a mediana com pouca variação nesse intervalo.

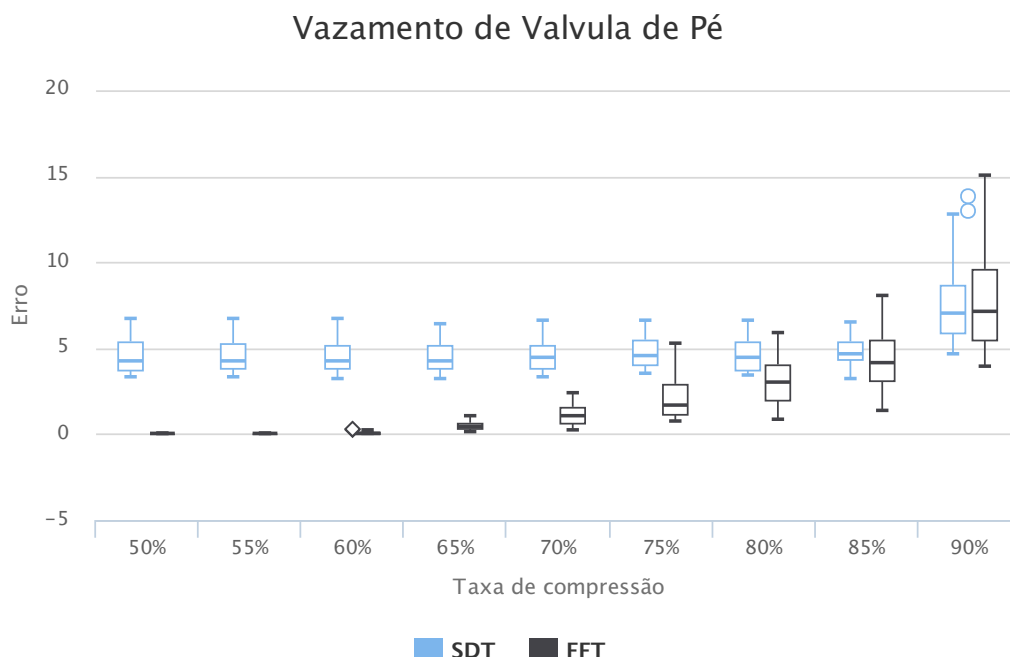


Figura 4.36: Análise da distribuição do erro para um conjunto de cartas de modo de operação de Válvula de Pé

4.3.6 Considerações Finais dos Estudos de Casos

Conforme demonstrado nos dois estudos de casos descritos nesta seção, estes possibilitam relacionar a máxima taxa de compressão que se pode atingir, sem gerar comprometimento visual, com a variação do erro, de tal modo que para cada taxa de compressão é possível medir qual o erro máximo aceitável. Isto possibilita que seja possível identificar quando o algoritmo gerou perda significativa de informação em tempo real.

Os dois algoritmos apresentam excelentes desempenhos do ponto de vista da relação custo benefício entre máxima taxa de compressão, com o erro mínimo entre a carta simplificada e a original, atingindo uma taxa de compressão de 80% sem gerar comprometimento visual significativo.

O algoritmo de Descritores de Fourier apresentou melhores resultados em quase todos os testes feitos, mas mesmo assim a escolha do algoritmo de compressão a ser utilizado vai depender do problema. Visto que existe um outro fator importante a ser considerado, que se trata da complexidade computacional. O algoritmo *Swinging Door* apresenta complexidade computacional no pior caso de $O(n)$, e possibilita que a carta simplificada resultante de sua aplicação seja armazenada exatamente como será visualizada.

O algoritmo de Descritores de Fourier dependendo de qual algoritmo será utilizado

apresenta maior complexidade computacional que o *Swinging Door*. Caso seja utilizado o algoritmo Discrete Fourier Transform (DFT), este apresenta complexidade no pior caso de $O(n^2)$. Se for utilizado o algoritmo Fast Fourier Transform (FFT), este possui complexidade computacional de $O(n \cdot \log(n))$. Por isso recomenda-se a utilização do algoritmo FFT, pois sua complexidade computacional é menor que a do DFT. Um outro fato é que a carta simplificada resultante da aplicação do algoritmo de Descritores de Fourier é armazenada no domínio da frequência, sendo necessário realizar a transformada inversa para visualizar a carta simplificada no domínio do tempo.

Os algoritmos também apresentam características diferentes quanto a configuração dos parâmetros de compressão. Pois o *Swinging Door* não garante que os parâmetros configurados vão atingir a taxa de compressão que o usuário deseja. Já o algoritmo de Descritores de Fourier possui o controle da taxa de compressão exata que se deseja atingir por meio do parâmetro de configuração.

Portanto o algoritmo de compressão mais adequado depende do problema a ser aplicado. Pois terá que ser ponderado a taxa de compressão, o erro, a complexidade computacional e o nível de controle na taxa de compressão desejada.

4.4 Suíte BR-Plant

As cartas dinamométricas de fundo de poço conforme descrito no capítulo 3 são elementos cruciais para o acompanhamento do funcionamento dos poços equipados com o método de elevação por Bombeio Mecânico. O aumento do parque tecnológico de automação nos campos de petróleo *onshore* gerou como consequência o crescimento significativo do número de cartas dinamométricas na mesma proporção [Gomes 2009].

Com isso, devido a sua importância e necessidade de armazenamento em larga escala, foi adicionado na suíte BR-Plant o suporte ao tipo de dados carta dinamométrica, necessitando de desenvolvimento e customização das ferramentas BR-PlantHistorian, BR-Collector e BR-PlantViewer mencionados na seção 2.4.

No BR-Collector foi implementado o suporte a captura de cartas dinamométricas. Na ferramenta BR-PlantHistorian foi adicionado o tipo carta dinamométrica e para otimizar o desempenho no armazenamento e disponibilização dos dados, foi adicionado a possibilidade de compressão aplicada a cartas através de uma adaptação do algoritmo *Swinging Door Trending*. No BR-PlantViewer foi criado um componente que permite a visualização e análise de cartas dinamométricas armazenadas no BR-PlantHistorian.

A figura 4.37 mostra uma análise de cartas dinamométricas no BR-PlantViewer, comparando as cartas originais com suas versões simplificadas resultantes da compressão.

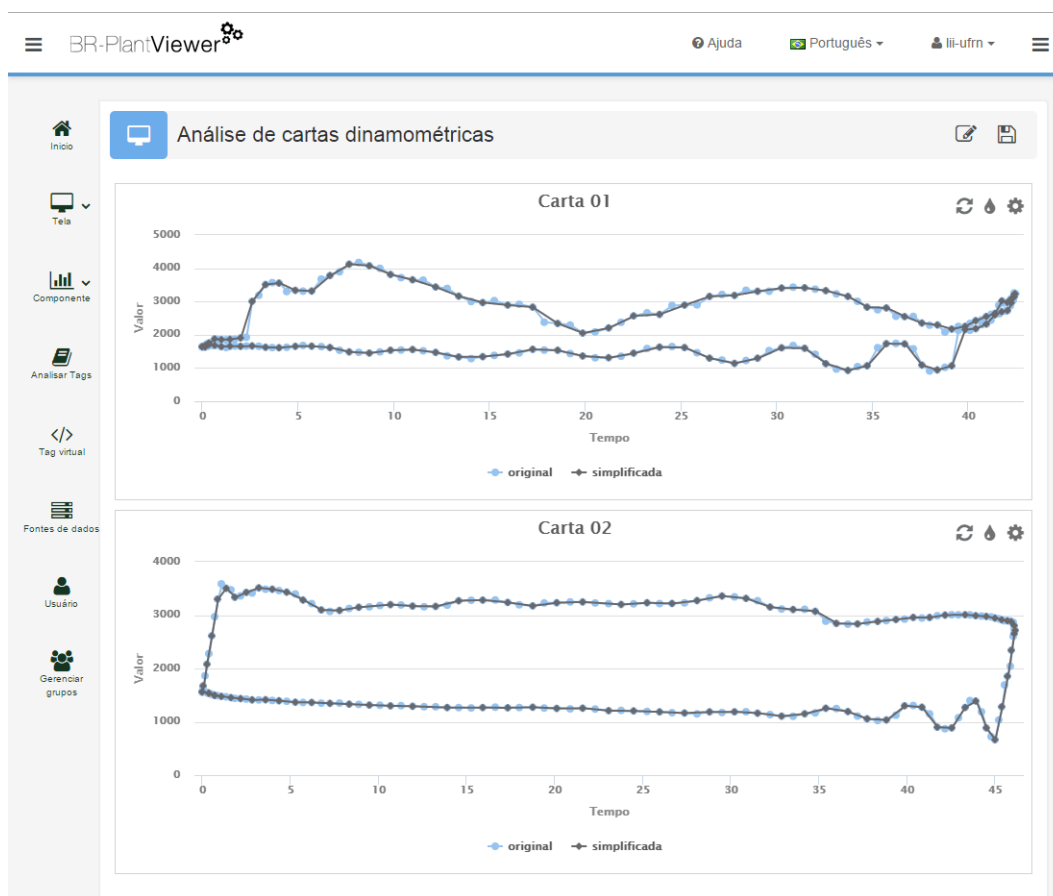


Figura 4.37: Análise de cartas dinâmométricas através do BR-PlantViewer

Capítulo 5

Conclusão e Trabalhos Futuros

Os avanços tecnológicos na área de automação industrial proporcionaram o aumento significativo do número de sensores e atuadores no chão de fábrica da indústria. Consequentemente este avanço contribuiu na geração de um grande volume de dados provenientes dos equipamentos instalados no campo industrial. Com isso os sistemas digitais responsáveis por gerenciar estes dados passaram a ter de lidar com problemas relacionados a necessidade de expansão de memória computacional e velocidade na recuperação e transferência de informação. Nesse contexto surgiram os algoritmos de compressão com perda, visando reduzir este grande volume de dados gerado e minimizar estes problemas.

Esta dissertação teve como foco a análise de algoritmos de compressão com perda, aplicados a dados industriais, com base em três contribuições principais. A primeira contribuição consistiu na proposição de uma adaptação do algoritmo *Swinging Door Trending*, alterando seu principal parâmetro de compressão de forma adaptativa com base na análise de tendência do sinal em tempo real, para aplicação em sinais de variáveis de processos temporais. A segunda proposição visou a comparação entre o método de compressão linear por partes *Swinging Door* com o de transformada de Fourier com base em limiarização, aplicados a cartas dinamométricas de fundo de poço. A terceira proporcionou a evolução do sistema BR-Plant para suportar o armazenamento e a análise de cartas dinamométricas de superfície e de fundo de poço, bem como também foi adicionado no módulo de compressão do sistema a opção de compressão para cartas dinamométricas utilizando o algoritmo *Swinging Door*.

Na análise de compressão do algoritmo *Adaptive Swinging Door Trending* (ASDT) comparado com o *Swinging Door Trending* (SDT) tradicional, concluiu-se que para os experimentos realizados com sinais que apresentam variações no *setpoint*, o algoritmo ASDT apresenta menor erro para as mesmas taxas de compressão atingidas pelo algoritmo SDT. Para o cenários em que estes algoritmos são submetidos a sinais de diferentes amplitudes e dinâmicas com os mesmos parâmetros de compressão, o algoritmo ASDT

apresenta melhor desempenho, demonstrando maior capacidade de generalização quando aplicado a sinais diferentes.

Para a análise de compressão aplicada a cartas dinamométricas de fundo de poço foi feita a comparação entre os algoritmos *Swinging Door Trending* e o de Descritores de Fourier. Esta comparação foi feita com base em dois estudos de casos.

No primeiro estudo de caso foi feita uma análise comparativa da aplicação dos algoritmos visando atingir a máxima taxa de compressão sem gerar comprometimento visual entre a carta original e a simplificada (resultante da compressão), para cada carta padrão representativa de um modo de operação do bombeio mecânico. Neste estudo de caso os dois algoritmos apresentam excelentes resultados até a taxa de compressão de 80% sem gerar comprometimento visual. No geral o algoritmo de Descritores de Fourier apresenta melhor desempenho em termos de erro de compressão e máxima taxa de compressão atingida para quase todos os modos de operação.

No segundo estudo de caso, a comparação entre os algoritmos foi feita utilizando estatística descritiva sobre um conjunto de cartas dinamométricas através do uso de diagramas de caixa. Este estudo foi feito para um subconjunto de cartas subdivididos por modo de operação. Em geral o algoritmo de Descritores de Fourier apresenta melhores resultados em termos de variação do erro de compressão até a taxa de compressão de 80%.

5.1 Trabalhos Futuros

Para futuras contribuições recomenda-se um estudo que possibilite sugerir as configurações dos parâmetros mais adequados para o algoritmo *Adaptive Swinging Door Trending* para cada tipo de processo. Com o intuito destes serem utilizados como configuração padrão.

Uma outra contribuição interessante seria realizar testes de compressão aplicada a cartas dinamométricas utilizando a transformada *Wavelet*. Assim como acrescentar novas contribuições ao sistema BR-Plant, adicionando o algoritmo de Descritores de Fourier aplicado a cartas, com recomendações das configurações de parâmetros de compressão mais adequadas.

Uma outra abordagem para ser trabalhada seria utilizar uma estratégia adaptativa para os parâmetros de configuração do algoritmo de Descritores de Fourier aplicados a cartas dinamométricas.

Referências Bibliográficas

- Ahmed, N., T. Natarajan & K.R. Rao (1974), 'Discrete cosine transform', *Computers, IEEE Transactions on* **C-23**(1), 90–93.
- Andrade, Wany Leydiane S., Alessandro J. de Souza, Rafael H. Feijó, Clauber Gomes Bezerra, André Laurindo Maitelli, Adelardo A. Dantas de Medeiros & Luiz Affonso H. G. de Oliveira (2006), Gerência e armazenamento de dados em tempo real em ambientes industriais, *em* 'VII Induscon', pp. 1–6.
- Bakshi, Bhavik R. & George Stephanopoulos (1996), 'Compression of chemical process data by functional approximation and feature extraction', *AIChE Journal* **42**(1), 477–492.
- Bristol, E. H. (1990), Swinging door trending: adaptive trend recording, *em* 'Proceedings of the ISA National Conf.', pp. 749–753.
- Carvalho, Fábio Barros, Bernardo Soares Torres, Marcos de Oliveira Fonseca & Constantino Seixas Filho (2005), 'On-line data compression and error analysis using wavelet technology', *Tecnologia em Metalurgia e Materiais* **1**(4), 1–5.
- Chambers, J.M., W.S. Cleveland, B. Kleiner & P.A. Tukey (1983), 'Graphical Methods for Data Analysis', *The Wadsworth Statistics/Probability Series. Boston, MA: Duxury*.
- Dickinson, Roderick R. & James W. Jennings (1990), 'Use of pattern-recognition techniques in analyzing downhole dynamometer cards', *SPE Production Engineering* **5**, 187–192.
- Filho, Constatino Seixas (1993), *Notas de Aula - Capítulo 6, PIMS - Process Information Management System*.
- Filho, M.A. Barreto, M. Tygel, A.F. Rocha & C.K. Morooka (1996), Automatic downhole card generation and classification, *em* 'Society of Petroleum Engineers'.

- Filho, Manuel Almeida Barreto (1993), Geração de carta dinamométrica de fundo para diagnóstico do bombeio mecânico em poços de petróleo, Dissertação de mestrado, Universidade Estadual de Campinas - UNICAMP, Brasil.
- Gersho, Allen & Robert M. Gray (1992), *Vector Quantization and Signal Compression*, springer^a edição, Norwell, MA.
- Gomes, Heitor Penalva (2009), Desenvolvimento de um sistema inteligente para a análise de cartas dinamométricas no método de elevação por bombeio mecânico, Dissertação de mestrado, Universidade Federal do Rio Grande do Norte - UFRN, Brasil.
- Hale, J. C. & H. L. Sellars (1981), 'Historical Data Recording for process computers', *Chemical Engineering Progress* .
- Huang, Wenjun, Weimin Wang & Hui Xu (2006), A lossless data compression algorithm for real-time database, *em* 'Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on', Vol. 2, pp. 6645–6648.
- Johnson-Fagg (1958), *Sucker Rod Handbook: Handbook 489*, Bethlehem Steel.
- Karim, Samsul Ariffin Abdul & Mohd Tahir Ismail (2009), 'Compression of chemical signal using wavelet transform', *European Journal of Scientific Research* **36**(1), 513–520.
- Kunttu, Iivari & Leena Lepisto (2005), Efficient fourier shape descriptor for industrial defect images using wavelets, *em* 'Society of Photo-Optical Instrumentation Engineers', Vol. 44.
- Lima, Fábio Soares (2014), Detecção e Classificação de Modos de Operação do Bombeio Mecânico Via Cartas Dinamométricas, Tese de doutorado, Universidade Federal do Rio Grande do Norte - UFRN.
- Lima, Fábio Soares, Luiz Affonso H. Guedes & Diego R. Silva (2009), 'Detecção de falhas no sistema de bombeio mecânico utilizando descritores de fourier e ferramentas estatísticas', *Simpósio Brasileiro de Automação Inteligente* .
- Machado, Rivaldo (2014), Desenvolvimento de um middleware para comunicação via web services e sua aplicação em sistemas de aquisição de dados industriais. Dissertação de Mestrado.

- Mallat, S.G. (1989), 'A theory for multiresolution signal decomposition: the wavelet representation', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **11**(7), 674–693.
- Misra, Manish, S. Joe Qin, Shailesh Kumar & Dick Seemann (2000), 'On-line data compression and error analysis using wavelet technology', *AIChE Journal* **46**(1), 119–132.
- Nascimento, Joao Maria (2005), Simulador computacional para poços de petróleo com método de elevação artificial por bombeio mecânico, Dissertação de mestrado, Universidade Federal do Rio Grande do Norte, Brasil.
- Pinto, Alan Kardec & Júlio Nascif (1998), *Manutenção - Função Estratégica*, Quality-Mark Editora.
- Rabiner, Lawrence & Biing-Hwang Juang (1993), *Fundamentals of Speech Recognition*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Rao, K. R. & P. Yip (1990), *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press Professional, Inc., San Diego, CA, USA.
- Rogers, J.D., C.G. Guffey & W.J.B. Oldham (1990), Artificial neural networks for identification of beam pump dynamometer load cards, *em* 'Society of Petroleum Engineers', Vol. 2.
- Schirmer, P., J.C. Gay & P. Toutain (1991), 'Use of advanced pattern-recognition and knowledge-based system in analyzing dynamometer cards', *SPE Computer Applications* **3**, 21–24.
- Shu, Sihui & Yi Shu (2012), A two-stage data compression method for real-time database, *em* 'System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2012 3rd International Conference on', Vol. 2, pp. 204–207.
- Singhal, A & D.E. Seborg (2005), 'Effect of data compression on pattern matching in historical data', *Industrial & Engineering Chemistry Research* **44**(9), 3203–3212.
- Souza, Alessandro J., Rafael H. Feijo, Gustavo B. P. Leitao, Adelardo A. D. Medeiros, Clauber G. Bezerra, Wany L. S. Andrade, Luiz Affonso Guedes & Andre L. Maitelli (2005), 'Gerência de informação de processos industriais: Um estudo de caso na produção de petróleo e gás', *VII Simposio Brasileiro de Automacao Inteligente*.

Thornhill, N. (2004), 'The impact of compression on data-driven process analyses', *Journal of Process Control* **14**(4), 389–398.

URL:<http://dx.doi.org/10.1016/j.jprocont.2003.06.003>

Watson, Matthew J., Antonios Liakopoulos, Dragana Brzakovic & Christos Georgakis (1998), 'A practical assessment of process data compression techniques', *Industrial & Engineering Chemistry Research* **37**(1), 267–274.

URL:<http://dx.doi.org/10.1021/ie970401w>

Watson, M.J., A. Liakopoulos, D. Brzakovic & Christos Georgakis (1995), Wavelet techniques in the compression of process data, *em* 'American Control Conference, Proceedings of the 1995', Vol. 2, pp. 1265–1269 vol.2.

Xiaodong, Feng, Cheng Changling, Liu Changling & Shao Huihe (2002), An improved process data compression algorithm, *em* 'Intelligent Control and Automation, 2002. Proceedings of the 4th World Congress on', Vol. 3, pp. 2190–2193 vol.3.

Apêndice A

Bombeio Mecânico

O bombeio mecânico é a forma mais comum de elevação artificial [Schirmer et al. 1991]. Sua importância se reflete no número de instalações existentes, que correspondem a 80% dos poços produtores mundiais e brasileiros [Nascimento 2005]. A Figura A.1 apresenta uma instalação de bombeio mecânico.

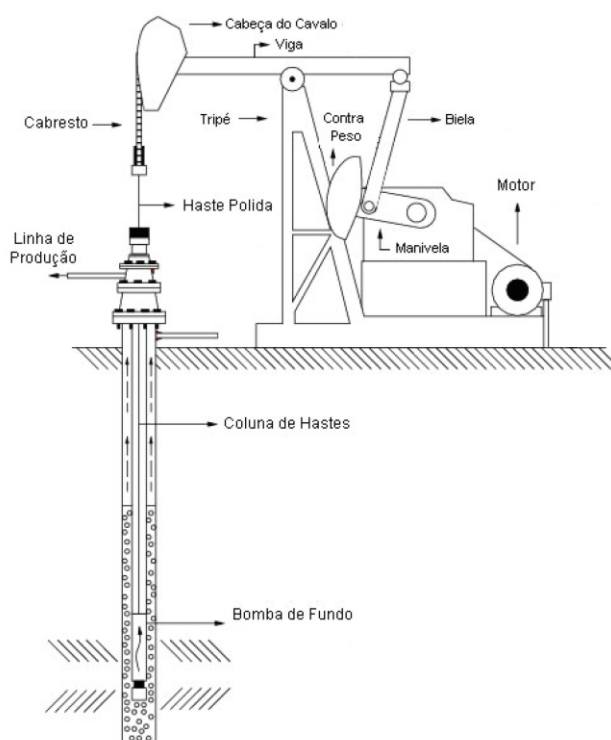


Figura A.1: Unidade de Bombeio Mecânico

O bombeio mecânico tem sua popularidade relacionada ao baixo custo com investimentos e manutenção, flexibilidade de vazão e profundidade, boa eficiência energética e a possibilidade de operar com fluidos de diferentes composições e viscosidades em larga

faixa de temperatura.

As principais vantagens do bombeio mecânico são: sua simplicidade de operação, manutenção e projeto de novas instalações. A partir de condições normais, pode ser utilizado até o fim da vida produtiva de um poço e sua capacidade de bombeio pode ser modificada em função das mudanças de comportamento do poço. Porém, a principal vantagem deste método diz respeito ao custo/produção ao longo da vida produtiva do poço.