



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA  
UNIDADE ACADÊMICA DE ENGENHARIA QUÍMICA

**JOSE EDIVALDO DA SILVA JUNIOR**

**AVALIAÇÃO ESTATÍSTICA DE ALGORITMOS DE EXCEÇÃO E  
COMPRESSÃO DE DADOS APLICADOS EM SISTEMA DE GESTÃO DE DADOS  
INDUSTRIAIS**

Campina Grande – PB

2021

**JOSE EDIVALDO DA SILVA JUNIOR**

**AVALIAÇÃO ESTATÍSTICA DE ALGORITMOS DE EXCEÇÃO E  
COMPRESSÃO DE DADOS APLICADOS EM SISTEMA DE GESTÃO DE DADOS  
INDUSTRIAIS**

Dissertação em Engenharia Química apresentado ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Campina Grande, como parte dos requisitos para obtenção do título de mestre em Engenharia Química.

Orientador: Prof. Dr. Heleno Bispo da Silva Júnior

Campina Grande – PB

2021

S586a

Silva Júnior, José Edivaldo da.

Avaliação estatística de algoritmos de exceção e compressão de dados aplicados em sistema de gestão de dados industriais / José Edivaldo da Silva Júnior. – Campina Grande, 2022.

144 f. : il. : color.

Dissertação (Mestrado em Engenharia Química) – Universidade Federal de Campina Grande, Centro de Ciências e Tecnologia, 2021.

"Orientação: Prof. Dr. Heleno Bispo da Silva Júnior".

Referências.

1. Compressão de Dados. 2. Análise de Sensibilidade. 3. Cartas de Controle. I. Silva Júnior, Heleno Bispo da. II. Título.

CDU 004.65(043)

JOSE EDIVALDO DA SILVA JUNIOR

**AVALIAÇÃO ESTATÍSTICA DE ALGORITMOS DE EXCEÇÃO E  
COMPRESSÃO DE DADOS APLICADOS EM SISTEMA DE GESTÃO DE  
DADOS INDUSTRIAIS**

Dissertação em Engenharia Química apresentado ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Campina Grande, como parte dos requisitos para obtenção do título de mestre em Engenharia Química.

Avaliado em: 29/07/2021

Banca Examinadora:



---

Prof. Dr. Heleno Bispo da Silva Júnior – UAEQ/CCT/UFCG

Orientador



Prof. Dr. Antônio Tavernard P. Neto

Avaliador Interno



---

Prof. Dr. Sidinei Kleber da Silva – UAEQ/CCT/UFCG

Avaliador Externo

## DEDICATÓRIA

Dedico este trabalho a minha família, por todo o amor e suporte durante todos estes anos. Estudo e dedicação só foram possíveis por tê-los em minha vida. Ao meu pai Edivaldo, que em vida me mostrou que ter sonhos é viver, e trabalhar é buscar tornar estes sonhos realidade. A minha mãe Maria que me mostra que a alegria é sempre o melhor caminho para enfrentar a jornada, assim como o amor a Cristo é o maior tesouro.

## AGRADECIMENTOS

A Deus por todos os momentos que ele tem me proporcionado, sustentado, guiado e direcionado. Por ter me ouvido e respondido por todos estes anos.

Aos meus pais por todo o amor, carinho, dedicação, exemplo, amor, honestidade e trabalho.

Aos meus tios e tias que me apoiaram nos momentos difíceis e que me mostram o caminho da educação.

Ao professo Heleno que me apoiou no desenvolvimento deste trabalho e que deu os melhores conselhos e direcionamentos para que eu chegasse a concluir esta etapa da minha carreira profissional. Obrigado por acreditar que eu conseguiria fazer um bom trabalho.

## RESUMO

A definição dos parâmetros de exceção e compressão de dados para o PI System tem por objetivo otimizar a compressão e armazenamento de dados industriais. Com objetivo de ajustar as configurações de ambos os algoritmos, estes foram replicados e validados em ambiente de análises de dados e utilizados como ambiente de simulação. Neste ambiente foram realizados experimentos numéricos para o processo de compressão de dados aplicado no PI System. Análises de sensibilidade para ambos os algoritmos de exceção e compressão de dados foram realizadas, adotando-se métricas estatísticas para avaliação da reconstrução das informações a partir dos dados comprimidos. Foi construída uma simulação de tanque flash para uma mistura de hidrocarbonetos em simulador de processos SimCentral. Uma malha de controle para a fração de nível foi definida e dados coletados desta variável passaram por análise do processo de compressão. Cartas de controle estatístico para média e amplitude foram construídas para o nível do tanque em duas condições operacionais. Estas cartas foram utilizadas para verificação dos impactos da exceção e compressão de dados sobre conclusões obtidas a partir das cartas de controle. A análise do desempenho de ambos os algoritmos em ambiente de simulação e análise possibilitou a definição de parâmetros ajustados com base nos intervalos definidos pelos indicadores utilizados. O resultado obtido para a compressão de dados do nível do tanque foi satisfatório, os pares *ExcDev* e *ExcMax* para exceção e *CompDev* e *CompMax* para compressão foram definidos, alcançando erro médio quadrado de 0,000638% e reduzindo 80% dos pontos originais. A partir da utilização dos 20% dos pontos, os dados reconstruídos foram capazes de manter as conclusões estatísticas demonstradas a partir dos dados brutos. Constatou-se que os indicadores estatísticos foram cruciais na definição dos parâmetros, pois delimitaram os limites de aceitáveis de exclusão de dados, mantendo-se as informações necessárias a manutenção das conclusões estatísticas.

Palavras chaves: Compressão de dados, Análise de Sensibilidade, Cartas de Controle.

## ABSTRACT

The definition of exception parameters and data compression for the PI System is intended to optimize the compression and storage of industrial data. In order to adjust the settings of both algorithms, they were replicated and validated in a data analysis environment and used as a simulation environment. In this environment, numerical experiments were performed for the data compression process applied in the PI System. Sensitivity analyzes for both exception and data compression algorithms were performed, adopting statistical metrics to assess the reconstruction of data from compressed data. A flash tank simulation for a mixture of hydrocarbons was built in SimCentral process simulator. A control loop for the level fraction was defined and data collected from this variable underwent compression process analysis. Statistical control charts for mean and amplitude were constructed for tank level under two operating conditions. These charts were used to verify the impacts of the exception and data compression on conclusions obtained from the control charts. The analysis of the performance of both algorithms in a simulation and analysis environment allowed the definition of adjusted parameters based on the intervals defined by the indicators used. The result obtained for the tank level data compression was satisfactory, the pairs ExcDev and ExcMax for exception and CompDev and CompMax for compression were defined, reaching mean square error of 0.000638% and reducing 80% of the original points. From the use of 20% of the points, the reconstructed data were able to maintain the statistical conclusions shown from the raw data. It was found that the statistical indicators were crucial in defining the parameters, as they delimited the acceptable limits of data exclusion, keeping the necessary information to maintain the statistical conclusions

Keywords: Data compression, Swinging Door Trending, Statistical Analysis, Statistical Process Control, Control Charts.



## LISTA DE FIGURAS

<i>Figura 1 Pirâmide de Automação.</i>	7
<i>Figura 2 Evolução da automação e suas tecnologias.</i>	8
<i>Figura 3 Infraestrutura do PI System.</i>	10
<i>Figura 4 Infraestrutura do PI System e suas conexões.</i>	11
<i>Figura 5 Fluxo de dados do PI System.</i>	13
<i>Figura 6 Linha do tempo de determinantes sobre compressão de dados.</i>	15
<i>Figura 7 Seleção de pontos para execução da compressão e exceção.</i>	18
<i>Figura 8 Abordagem de equilíbrio ou controle, quando causas atribuíveis são eliminadas, portanto reduzindo a necessidade de inspeção.</i>	23
<i>Figura 9 As propriedades do Big Data e suas nuances.</i>	24
<i>Figura 10 Definição de subgrupos racionais.</i>	25
<i>Figura 11 Exemplo de carta de controle.</i>	27
<i>Figura 12 Melhoramento do processo a partir de cartas de controle.</i>	28
<i>Figura 13 Parâmetros do algoritmo de exceção de dados.</i>	30
<i>Figura 14 Condições de envio de dados pelo Algoritmo de exceção de dados.</i>	31
<i>Figura 15 Conceito geral do Swinging Door Trending – SDT.</i>	32
<i>Figura 18 Início de Avaliação das Inclinações Máxima e Mínima, primeiro ponto avaliado após o último armazenamento.</i>	34
<i>Figura 19 Avaliação da inclinação do segundo ponto.</i>	35
<i>Figura 20 Avaliação das Inclinações Máxima e Mínima do ponto atual.</i>	36
<i>Figura 21 Detecção de ponto com inclinação fora do intervalo aceitável.</i>	36
<i>Figura 22 Armazenamento do ponto anterior e definição de novas inclinações máxima e mínima.</i>	37
<i>Figura 23 Condições de armazenamento de dados pelo algoritmo de compressão de dados</i>	38
<i>Figura 24 Entidades envolvidas no processo de compressão e consulta de dados armazenados.</i>	46
<i>Figura 25 Fluxograma desenvolvido para implementação do SDT.</i>	48
<i>Figura 26 Dados brutos e Referência para avaliação da exceção.</i>	56
<i>Figura 27 Janela de dados brutos e Referência para avaliação da exceção.</i>	57
<i>Figura 28 Ampliação sobre segmentos de dados aplicados à validação da exceção.</i>	58
<i>Figura 29 Dados brutos e Referência para avaliação da compressão.</i>	62
<i>Figura 30 Dados brutos e Referência para avaliação da compressão.</i>	63
<i>Figura 31 Zoom aplicado aos dados usado na validação do algoritmo de compressão.</i>	64
<i>Figura 32 Erros gerados pelo processo de exceção.</i>	67
<i>Figura 33 Erros gerados somente pelo SDT.</i>	69
<i>Figura 34 Erros gerados pela combinação dos ambos os algoritmos.</i>	71
<i>Figura 35 Indicadores PDM, RVC e RVE obtidos variando-se Exdev Algoritmo de exceção.</i>	73
<i>Figura 36 Indicadores ISE, IAE e COMPRATIO obtidos variando-se Exdev Algoritmo de exceção.</i>	74
<i>Figura 37 Cenário 1, caso 2, Indicadores de PDM, RVC RVE avaliando-se ExcMax.</i>	75

Figura 38 Cenário 1, caso 2, Indicadores de mse, mae, CompRation avaliando-se ExcMax. _____	75
Figura 39 Avaliação do indicador PDM para o algoritmo de exceção. _____	77
Figura 40 Avaliação do indicador MAE para o algoritmo de exceção. _____	78
Figura 41 Avaliação do indicador MSE para o algoritmo de exceção. _____	78
Figura 42 Avaliação do indicador RVC para o algoritmo de exceção. _____	80
Figura 43 Avaliação do indicador RVE para o algoritmo de exceção. _____	80
Figura 44 Avaliação do indicador Índice de correlação para o algoritmo de exceção. _____	82
Figura 45 Avaliação do indicador Razão de Compressão para o algoritmo de exceção. _____	83
Figura 46 Indicadores PDM, RVC e RVE obtidos variando-se CompMax algoritmo de compressão SDT. _____	84
Figura 47 Indicadores ISE, IAE e COMPRATIO obtidos variando-se CompMax algoritmo de compressão SDT. _____	85
Figura 48 Indicadores PDM, RVC e RVE obtidos variando-se CompMax algoritmo de compressão SDT. _____	87
Figura 49 Caso 2, Indicadores de MSE, MAE, CompRatio avaliando-se CompDev. _____	87
Figura 50 Avaliação do indicador PDM para o algoritmo de compressão. _____	89
Figura 51 Avaliação do indicador RVE para o algoritmo de compressão. _____	90
Figura 52 Avaliação do indicador RVC para o algoritmo de compressão. _____	91
Figura 53 Avaliação do indicador MAE para o algoritmo de compressão. _____	92
Figura 54 Avaliação do indicador MSE para o algoritmo de compressão. _____	92
Figura 55 Avaliação do indicador Índice de Correlação para o algoritmo de compressão. _____	93
Figura 56 Avaliação do indicador Razão de Compressão para o algoritmo de compressão. _____	94
Figura 57 Diagrama de representação da simulação de tanque flash. _____	95
Figura 58 Especificações do controle de nível do tanque. _____	97
Figura 59 Cenários de operação do tanque e suas respectivas cartas de controle. _____	100
Figura 60 Dados para construção dos limites de controle. _____	103
Figura 61 Construção de carta de controle para o cenário 1. _____	104
Figura 62 Dados brutos de simulação de processo de flasheamento. _____	105
Figura 63 Estruturação de avaliação da compressão considerando as fases de análise. _____	106
Figura 64 Avaliação do processo de compressão de dados nas regiões Z1 e Z3. _____	107
Tabela 20 Parâmetros de construção de cartas XBarra-R. _____	107
Figura 65 Carta X-BarraR para o estado estacionário 1 (Zona 1). _____	108
Figura 66 Carta X-BarraR para o estado estacionário 2 (Zona 3). _____	108
Figura 67 Fluxo de avaliação do processo de compressão de dados. _____	111
Figura 68 Valores padrão da compressão e exceção fornecidos pelo PI System _____	112
Figura 69 Processo de compressão de dados com completo com parâmetros padrão – Dados arquivados. _____	113
Figura 70 Processo de compressão de dados com completo com parâmetros padrão – Dados reconstruídos (A) e Erros individuais de cada valor reconstruído (B). _____	113
Figura 71 Indicadores estatísticos para avaliação dos parâmetros do algoritmo de exceção de dados. _____	114
Figura 72 Avaliação do Algoritmo de exceção: ExcDev= 0.1, ExcMax=10. _____	118
Figura 74 Avaliação do Algoritmo de exceção: ExcDev= 0.1, ExcMax=40. _____	120

<i>Figura 75 Avaliação do Algoritmo de exceção: ExcDev= 0.2, ExcMax=20.</i>	121
<i>Figura 76 Avaliação do Algoritmo de exceção: ExcDev= 0.2, ExcMax=40.</i>	122
<i>Figura 77 - Indicadores estatísticos para avaliação dos parâmetros do algoritmo de compressão de dados.</i>	123
<i>Figura 86 Avaliação do Algoritmo de Compressão: CompDev= 0.05, CompMax=10.</i>	127
<i>Figura 87 - Avaliação do Algoritmo de Compressão: CompDev= 0.05, CompMax=20.</i>	128
<i>Figura 88 - Avaliação do Algoritmo de Compressão: CompDev= 0.05, CompMax=40.</i>	129
<i>Figura 89 Avaliação do Algoritmo de Compressão: CompDev= 0.1, CompMax=20.</i>	130
<i>Figura 90 Avaliação do Algoritmo de Compressão: CompDev= 0.1, CompMax=40.</i>	131
<i>Figura 91 Processo de compressão de dados com parâmetros de algoritmos de exceção e compressão.</i>	133
<i>Figura 92 Cartas de controle a partir de dados comprimidos.</i>	138

## LISTA DE TABELAS

<i>Tabela 1 Benefícios da aplicação do processo de compressão de dados no PI System.</i>	29
<i>Tabela 2 Significados dos parâmetros do algoritmo de exceção.</i>	31
<i>Tabela 3 Fórmulas para construção das cartas de controle da média e amplitude.</i>	43
<i>Tabela 4 Informações básicas da tag coletada e configuração da exceção de dados.</i>	54
<i>Tabela 5 Comparativo das informações básicas entre os resultados para validação do algoritmo de exceção em janela de dados.</i>	55
<i>Tabela 6 Comparação numérica entre os algoritmos de exceção de referência (PI System) e o desenvolvido (Simulador).</i>	59
<i>Tabela 7 Comparativo de resultados das Informações básicas para validação do algoritmo de compressão na massa total de dados.</i>	60
<i>Tabela 8 Indicadores estatísticos aplicados na validação do algoritmo de compressão de dados.</i>	61
<i>Tabela 9 Descrição da configuração do algoritmo de exceção e erros obtidos pelo processo de exceção de dados.</i>	65
<i>Tabela 10 Descrição da configuração do algoritmo de Compressão e erros obtidos pelo processo de Compressão de dados.</i>	68
<i>Tabela 11 Descrição da configuração do algoritmo de Exceção e Compressão e erros obtidos pela combinação de ambos os algoritmos.</i>	70
<i>Tabela 12 Parâmetros de avaliação aplicados na avaliação da exceção de dados variando-se o parâmetro ExcDev.</i>	73
<i>Tabela 13 Parâmetros de avaliação aplicados na avaliação da exceção de dados variando-se o parâmetro ExcMax.</i>	74
<i>Tabela 14 Parâmetros de avaliação aplicados na avaliação da compressão de dados variando-se o parâmetro CompMax.</i>	84
<i>Tabela 15 Parâmetros de avaliação aplicados na avaliação da compressão de dados variando-se o parâmetro CompDev.</i>	86
<i>Tabela 16 Composição inicial da mistura de hidrocarbonetos.</i>	96
<i>Tabela 17 Modelagem termodinâmica para mistura de hidrocarbonetos.</i>	96
<i>Tabela 18 Especificações de entrada definidas em SRC1.</i>	96
<i>Tabela 19 Especificações de sink e válvulas.</i>	98
<i>Tabela 20 Parâmetros de construção de cartas XBarra-R.</i>	107
<i>Tabela 21 Verificação dos intervalos de interesse dos indicadores estatísticos para algoritmo de exceção.</i>	116
<i>Tabela 22 Resultados da exceção de dados, apresentação de resultados para os intervalos estabelecidos.</i>	116
<i>Tabela 23 Verificação dos intervalos de interesse dos indicadores estatísticos para algoritmo de exceção.</i>	125
<i>Tabela 24 Resultados da compressão de dados, apresentação de resultados para os intervalos estabelecidos.</i>	126
<i>Tabela 25 Parâmetros ajustados para os algoritmos de exceção e compressão de dados.</i>	132
<i>Tabela 26 Resultado do ajuste do processo de compressão de dados.</i>	134

*Tabela 27 Resultados obtidos para cartas de controle  $\bar{X}$ , considerando a utilização de dados cumprido com o processo de compressão ajustado..... 137*

## LISTA DE ABREVIATURAS E SIGLAS

tag	Ponto de armazenamento exclusivo para dados do Data Archive
PDM	Diferença Percentual Entre médias
RVC	Razão entre Variância dos Dados Reconstruídos e Variância dos Dados Originais
RVE	Razão entre Variância dos Erros Reconstruídos e Variância Dados Originais
LCS	Limite de Controle Superior
LC	Linha de Controle
LCI	Limite de Controle Inferior
PDM	Diferença Percentual entre Médias (PDM)
RVC	Razão entre Variância dos Dados Reconstruídos e Variância dos Dados Originais
RVE	Razão entre Variância dos Erros Reconstruídos e Variância Dados Originais
MSE	Erro Quadrado Médio
MAR	Erro Absoluto Médio
CompRatio	Razão de compressão

## SUMÁRIO

1. INTRODUÇÃO .....	1
2. OBJETIVOS .....	4
2.1. Objetivo Geral .....	4
2.2. Objetivos Específicos .....	4
2.3. Motivação .....	5
3. REVISÃO BIBLIOGRÁFICA .....	6
3.1. Sistema de Gerenciamento de Informação de Plantas - PIMS .....	6
3.2. PI System .....	9
3.3. Compressão de dados .....	15
3.4. Avaliação do desempenho da compressão de dados segundo a dinâmica .....	17
3.5. Controle Estatístico de Processos .....	21
3.5.1. Subgrupos Racionais .....	25
3.5.2. Cartas de controle: Conceitos e Princípios .....	26
4. FUNDAMENTAÇÃO TEÓRICA .....	29
4.1. Processo de compressão de dados .....	29
4.1.1. Algoritmo de exceção de dados .....	30
4.1.2. Algoritmo de compressão - <i>Swinging Door Trending</i> (SDT) .....	32
4.2. Construção de Indicadores Estatísticos .....	39
4.2.1. Diferença Percentual Entre médias (PDM) .....	40
4.2.2. Razão entre Variância dos Dados Reconstruídos e Variância dos Dados Originais (RVC) .....	40
4.2.3. Razão entre Variância dos Erros Reconstruídos e Variância Dados Originais (RVE) .....	41
4.2.4. Erro Quadrado Médio (MSE) .....	41

4.2.5. Erro Médio Absoluto (MAE) .....	41
4.2.6. Razão de compressão (CompRatio) .....	42
4.2.7. Coeficiente de Correção de Pearson .....	42
4.3. Controle Estatístico de Processos e o Monitoramento de Ativos Industriais.....	43
5. METODOLOGIA .....	45
5.1. Compressão de dados .....	45
5.1.1. Fluxo de Dados na Fase de Compressão.....	45
5.1.2. Fluxograma do Algoritmo de Compressão de Dados.....	46
5.1.3. Validação do algoritmo desenvolvido.....	49
5.2. Análise de indicadores de compressão de dados .....	50
5.3. Cartas de controle a partir de dados comprimidos.....	51
6. RESULTADOS E DISCUSSÕES .....	52
6.1. Validação do Algoritmo de exceção e compressão .....	53
6.1.1. Validação do Algoritmo de Exceção de dados .....	54
6.1.2. Validação do Algoritmo de Compressão .....	59
6.2. Erros gerados pelos algoritmos de exceção e compressão de dados	65
6.2.1. Erros originados pelo processo de exceção .....	65
6.2.2. Erros originados pelo processo de compressão .....	68
6.2.3. Erros originados pelos processos de exceção e compressão combinados.....	70
6.3. Análise de compressão de dados pelo algoritmo Swinging Door Trending	72
6.3.1. Estratégia de Avaliação Aplicada.....	72
6.3.2. Avaliação do algoritmo de Exceção de dados .....	72
6.3.3. Avaliação do algoritmo de Compressão de dados .....	83
6.4. Estudo de caso .....	95



6.4.1. Processo Químico.....	95
6.5. Controle Estatístico de processos.....	101
6.5.1. Cartas de controle a partir de serie de dados temporais.....	102
6.5.2. Cartas de controle para monitoramento de nível de tanque .....	104
6.6. Avaliação dos parâmetros de exceção e compressão .....	110
6.7. Processo de Compressão de Dados em todo o intervalo de dados .	112
6.8. Avaliação dos parâmetros do algoritmo de exceção.....	114
6.9. Avaliação dos parâmetros do algoritmo de compressão.....	123
6.10. Resultados da avaliação dos algoritmos de exceção e compressão de dados .....	132
6.11. Aplicação de Cartas de controle para o tanque flash.....	136
7. CONCLUSÃO .....	141
8. SUGESTÃO DE TRABALHOS FUTURO.....	143
REFERÊNCIAS.....	144

## 1. INTRODUÇÃO

É inquestionável a importância dos dados industriais, assim como seu poder em definir o direcionamento operacional e gerencial sobre os sistemas que os geram. Nesse quesito, a operação da planta se detém atingir as metas de produção segundo metas estabelecidas pela esfera gerencial. A possibilidade de integrar estas camadas das organizações adiciona aos negócios em competitividade e versatilidade, principalmente quando mais de um produto faz parte dos produtos finais.

A presença da instrumentação madura, responsável por medir grandezas e por atuar sobre o processo através de atuadores variados, descreve o avanço em direção a linhas produtivas com maior eficiência. Diretamente impactada, a qualidade tem agora meios de garantir a maior reprodutibilidade da especificação do produto final. A existência de sensores, tais como medidores de fluxo volumétrico, pressão, temperatura, corrente elétrica, não se resumem a controlar o processo. Hoje, sensores e atuadores são base para a construção de novos conhecimentos sobre o processo, quando combinados a sistema de armazenamento e análise de dados de planta.

O aprimoramento gerencial com base em dados coletados e armazenados do processo produtivo é o avanço esperado quando se investe em tais tipos de equipamento e estrutura de análise tecnologicamente sofisticada. A determinação de metas de produção e estimativa de tempo e de custo de manufatura já não é mais puramente empírica. As metas podem ser estimadas a partir de dados temporais anteriores e modelos de produção validados, simulações e avaliação de cenários tornam-se possíveis.

Em conjunto com a equipe operacional, a possibilidade de realização de análises, simples e complexas, previsão e planejamento da produção, auditorias de estados anteriores do processo produtivo tornam-se fácil executáveis. Em casos reais, tais análise são capazes de determinar lucros consideráveis, devido a racionalização dos recursos de produção e aprimoramentos decorrente ao tratamento dos dados segundo metodologias clássicas ou inovadoras. Dentre as análises clássicas que correlaciona a camada gerencial com o a camada operacional, estão as cartas de controle, ferramentas estatísticas que auxiliam na detecção da presença de

produto não conforme. Sendo assim, o estado de instabilidade torna-se inaceitável do processo industrial, quando fora de uma margem aceitável pré-definida.

Para que tais metodologias sejam aplicadas industrialmente, sistemas historiadores dedicados são necessários, os quais são responsáveis por manter o histórico dos dados coletados. Comercialmente estes sistemas oferecem, além disso, toda uma estrutura de integração com sistemas de automação, OPC-server, CLP (controlador lógico programável) e bancos de dados relacionais (MySQL, MariaDB, Oracle) e também vasta quantidade de interfaces com protocolos industriais e formatos de arquivos diferentes.

A integração de dados de várias fontes considera que o sistema de gerenciamento performe o armazenamento adequado dos dados coletados. Para este fim, o processo de compressão é adotado. Este confere ao sistema a capacidade de redução dos custos de armazenamento, redução de redundância nos dados e custo de infraestrutura de rede. Em sistemas de gerenciamento de dados industriais como PI System da Osisoft, Historian da General Eletrics e InfoPlus 21 da Aspentech, o processo de compressão desempenha função central em termos de manutenção informação nos dados coletados.

Para o PI System, o processo de compressão de dados foi baseado no algoritmo denominado *Swinging Door Trending (SDT)* (BRISTO, 1990). Este é integrado a uma fase anterior ao algoritmo original e chamada de boxcar, ou amortizador, em tradução livre. A fase de boxcar, denominada como *exception*, traduzido por exceção, é responsável por reduzir a presença de ruídos nos dados (ALSMEYER, 2006).

Segundo os critérios de cada variável a ser coletada, o ajuste dos parâmetros dos algoritmos de exceção e compressão se faz necessário. Tal ajuste evita a perda ou distorção das informações contidas intrinsecamente em cada valor coletado de planta. Considerando que o mau desempenho do algoritmo de compressão pode vir a inviabilizar a aplicação de diversas análises, a partir dos dados armazenados. Quando estes são consultados a reconstrução é realizada considerando os dados resultantes do processo de compressão. Para a tomada de decisão assertiva, os dados presente em banco deve suportar a reconstrução linear do perfil equivalente ao presente nos dados (SINGHAL e SEBORG, 2005).

Partindo da aplicação da estatística descritiva, algumas métricas de avaliação são aplicadas como ferramenta de análise de desempenho da compressão de dados. Tal análise avalia os dados em ambos estados disponíveis: bruto e reconstruído (THURNHILL, CHOUDHULL, SHAH, 2004). É proposta a análise de sensibilidade em pares dos parâmetros de *ExcMax* e *ExcDev* para o algoritmo de exceção, além dos pares *CompMax* e *CompDev* para o algoritmo de compressão. A seleção dos pares de parâmetros considera os intervalos de maior relevância dos indicadores de compressão, arbitrários para os dados em análises.

O estudo de caso corresponde a coleta de dados de fração de nível de um tanque flash de uma mistura de hidrocarbonetos. A esta variável é aplicada a avaliação de dois estados estacionários, para os quais duas cartas de controle são construídas. O monitoramento do nível do tanque é realizado a partir dos dados reconstruídos aplicados a cartas de controle  $\bar{X}$  e  $R$ . É esperado que, para os parâmetros ajustados de ambos os algoritmos, os resultados do monitoramento estatístico sejam os mesmos quando comparando resultados entre cartas aplicando dados brutos e dados reconstruídos.

Como resultado sugere-se uma metodologia de avaliação dos parâmetros de configuração dos algoritmos, com objetivo de extrair a melhor reconstrução dos dados originais, ajustando os intervalos aceitáveis para os indicadores estatísticos utilizados. Esta abordagem prove ao usuário uma alternativa ao arbitramento destes parâmetros, recaindo somente a definição dos intervalos aceitáveis dos indicadores estatísticos de processo de compressão.

## 2. Objetivos

Nesta seção são apresentados os objetivos do presente trabalho. Estes são divididos entre objetivo geral e objetivos específicos. Por fim, a motivação do trabalho é apresentada.

### 2.1. Objetivo Geral

Avaliar a influência do processo de compressão e exceção de dados aplicados sobre dados históricos colhidos de plantas industriais sobre as informações contidas nos dados.

### 2.2. Objetivos Específicos

Implementar algoritmo de compressão e exceção em ambiente de desenvolvimento;

- Validar algoritmo de exceção aplicado com dados de referências processados pelo algoritmo aplicado pela Osisoft;
- Implementar algoritmo de compressão *Swinging Door Trending* (SDT) em ambiente de desenvolvimento;
- Validar algoritmo de compressão implementado com dados de referências processados pelo algoritmo aplicado pela Osisoft;
- Determinar indicadores estatísticos descritivos para avaliar o impacto da compressão de dados sobre as medidas centrais e de variabilidade de cada conjunto de dados;
- Aplicar indicadores estatísticos para comparação das características dos conjuntos de dados antes e após o processo de compressão de dados;
- Traçar perfil dos indicadores estatísticos dentro de um intervalo fixo de parâmetros de compressão para cada conjunto de dados;
- Comparar os resultados do algoritmo de compressão puro com os resultados do algoritmo compressão quando combinado com algoritmo de exceção;
- Avaliar os resultados de compressão quando os parâmetros de compressão de dados são configurados a partir dos indicadores apresentados;
- Aplicar cartas de controle  $\bar{X}$  e R para verificar se dados reconstruídos podem ser aplicados ao monitoramento da saúde variáveis chave em processos químicos.

### 2.3. Motivação

O presente trabalho visa expandir a quantidade de análises do impacto do processo de compressão de dados. A contribuição está na determinação de indicadores de qualidade estatística para os dados comprimidos, considerando a manutenção das informações contidas nos dados.

Para tanto, dados brutos e comprimidos em um intervalo definido de parâmetros de compressão foram analisados. Os experimentos foram conduzidos com base no algoritmo SDT, este aplicado industrialmente no sistema de PIMS da Osisoft, PI System (THURNHILL *et al.* 2004).

Este sistema é aplicado nos mais diversos setores, classificado como sistema PIMS, do inglês, *Plant Information Management System*. Sua característica primordial é o armazenamento de dados em tempo real para a posterior consulta. Nesse contexto, devido a grande quantidade de dados, algoritmos de compressão de dados desempenham papel importante para o armazenamento de dados, que quando otimizados reduzem a redundância nos dados historiados e os custos direcionados ao hardware e equipamentos aplicados aos sistemas PIMS.

Portanto, a exploração de técnicas de avaliação do processo de compressão torna-se justificada e desejada, considerando a possibilidade de diminuir custos de armazenamento de dados, redução quantidade de dados em tráfico na rede industrial, diminuir a redundância nos dados, extrair a melhor relação dados brutos / dados comprimidos e manter ao máximo as informações contidas nos dados, as quais são o motivo real de todo o sistema de gerenciamento de dados industriais.

### **3. Revisão Bibliográfica**

A presente seção apresenta trabalhos científicos correlacionados ao tema apresentado. Com teor diversificado, cada documento apresentado contribui com o presente tema, auxiliando ao leitor a consulta outras fontes científicas relacionadas. É apresentado o conceito do sistema de gerenciamento de informação de Plantas e sua integração com os demais sistemas de gerenciamento e de operação industriais. O PI system é destaque na seção, por ser conhecido como o representante mais popular e por aplicar o algoritmo de compressão de dados SDT. São apresentados também os conceitos de compressão de dados os trabalhos relevantes que contribuíram para o desenvolvimento para o desenvolvimento dos algoritmos de compressão de dados. Por fim são introduzidas aplicações de controle estatístico de processos correlatos aos presentes tema.

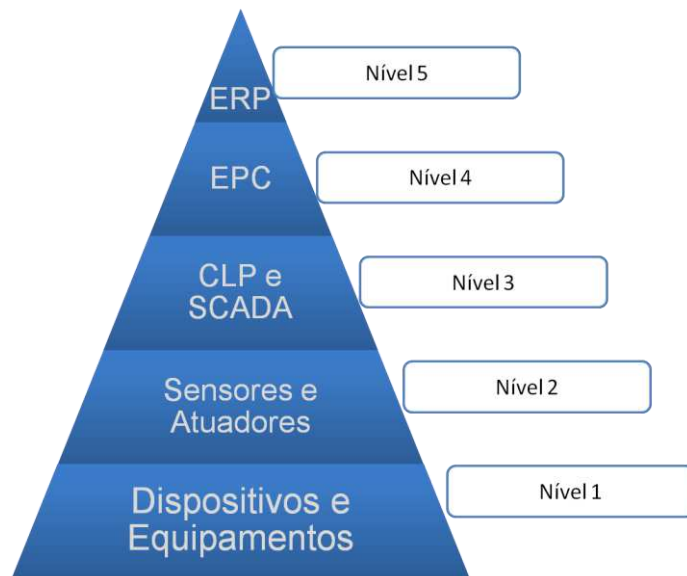
#### **3.1. Sistema de Gerenciamento de Informação de Plantas - PIMS**

Considera-se como principal característica de um sistema PIMS, a integração entre elementos de coleta e armazenamento através de uma rede industrial de comunicação. O PIMS possibilita a monitoramento e a supervisão on-line de sistemas de produção, segurança e sua própria estrutura de computadores e de rede.

Dentre os sistemas de gerenciamento de dados industriais estão o PI System®, da Osisoft, e o InfoPlus.21® da AspenTech. Ambos com foco no armazenamento de dados e disponibilização de um ambiente propício para análise dos dados.

Tomando-se o cenário global industrial, as camadas do sistema de automação são descritas segundo demonstrado na Figura 1 (SOUZA et al, 2005).

Figura 1 Pirâmide de Automação.



Fonte: (GROOVER, 2015)

Na hierarquia demonstrada na pirâmide de automação é possível partir do nível mais básico ao mais avançado. Critérios variados como nível de tecnologias e informações colhidas ou estimadas do processo são considerados. É possível elencar principais características de cada fase presente na pirâmide da Figura 1:

**Nível 1** – Na base da pirâmide encontram-se os equipamentos e dispositivos aplicados no processo produtivos. Estes são os principais elementos da pirâmide de automação, os quais necessitam suportar as demais camadas tanto em relação ao hardware quanto aos softwares aplicados.

**Nível 2** – Encontram-se os sensores e atuadores. Estes elementos são cruciais na coleta de dados e no controle das condições de operação das plantas industriais. Exemplos dos primeiros são sensores de nível, pressão, temperatura e dos segundos são inversores de frequência;

**Nível 3** – Os controladores lógicos programáveis (CLPs) e os sistemas supervisórios, que combinados são denominados SCADA, inglês (*Supervisory Control and Data Acquisition*);

**Nível 4** – Capazes de integrar informações de mais de uma área de supervisão, os sistemas de gestão da informação surgem com papéis complementares: O *Plant Information Management System (PIMS)* e *Manufacturing Execution System (MES)*. A

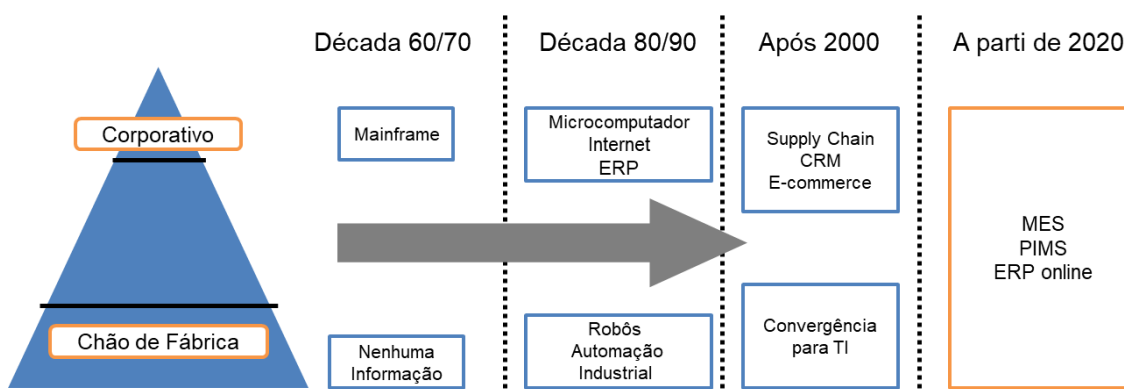


combinação de ambas recebe a denominação de *Enterprise Production System*(EPS). Neste nível ocorre a centralização da informação

**Nível 5** – O último nível da pirâmide corresponde ao nível corporativo e estratégico do processo de manufatura. Neste, os dados de processo são transformados em informações de cunho financeiro. Denominado de *Enterprise Resource Planning (ERP)*, tem por característica auxiliar na tomada de decisão no âmbito global de negócio.

A estrutura demonstrada tem o foco principal em desenvolver a conexão entre dados colhidos em “chão-de-fábrica”, o a camada de inteligência operacional e estratégico/financeiro. Os níveis extremos, base e topo, da pirâmide podem ser nomeados como níveis chão-de-fábrica e corporativo, respectivamente. Na Figura 2, é demonstrada uma linha do tempo com relação aplicação de novas técnicas, metodologias e conceitos aos sistemas de automação.

Figura 2 Evolução da automação e suas tecnologias.



Fonte: Venturelli (2011)

Acima é demonstrada o progresso da implantação dos conceitos da pirâmide de controle/automação. Para a realidade brasileira, o início é considerado na década de 60/70 com a adição de computadores de grande porte, também conhecidos como *mainframe*, com seu foco de aplicação voltado para as necessidades industriais.

Nas duas décadas posteriores, a possibilidade de possuir computadores de mesa com estrutura dedicada ao consumidor final tornou-se intensificada. A popularidade destes equipamentos adicionou a indústria uma realidade mais flexível, agora a compacidade de processamento e cálculos começou a ser distribuída. O

controle de gerencial/industrial passou a ter um direcionamento rumo a integração destas duas esferas.

A automação seguiu o mesmo caminho, introduzindo a inteligência e estrutura de sensoriamento e atuação sobre o parque industrial. Esse quesito tornou-se evidente à medida que tarefas desempenhadas anteriormente por humanos, passaram a ser exercidas por uma estrutura mecanizada programável.

Como produto desta fase, a virada do século deu início a uma onda de novas ferramentas e possibilidades de integração de vários setores. A presença do computador no dia-a-dia das pessoas culminou no desenvolvimento de ferramentas computadorizadas, *sites* e no desenvolvimento de Tecnologias da Informação. Foi através dos avanços em telecomunicação que surgiram melhoras na integração dos setores produtivos e consumidores. Consideram a intensificação do comercio internacional e o maior estreitamento das cadeias produtivas, o comercio tornou-se ainda mais competitivo e descentralizado.

O aumento de competitividade e integração industrial tornou-se uma necessidade global. Após a virada do milênio, a dinâmica mundial foi aumentara drasticamente. A logística combinadas a novos mecanismos de busca tornou o comercio de produtos manufaturados em uma disputa por qualidade e preços cada vez menores.

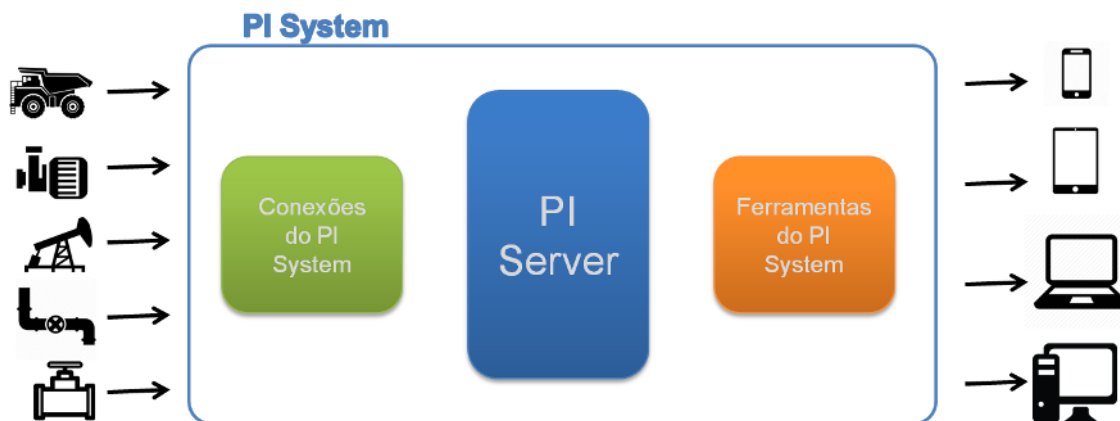
O impacto desse comportamento foi absorvido pela ampliação nas estratégias de negócios. Sistemas de gerenciamento de dados/informação tornaram-se um diferencial competitivo ou até mesmo pré-requisitos para manutenção da posição de mercado. Como Venturelli (2011) estima, os sistemas MES, PIMS e ERPS tornaram-se os sistemas que direcionam ações gerenciais rumo ao patamar da concorrência global.

### **3.2. PI System**

O PI system é constituído de um conjunto de softwares integrados, dedicados ao gerenciamento de dados em tempo-real. O seu foco está em oferecer acesso a ferramentas para historiar, consultar, analisar, visualizar e compartilhar dados temporais de processo. A Figura 3 representa o sistema e sua infraestrutura básica,

na qual cada elemento demonstrado exerce um papel específico e que define as camadas a que pertencem (OSISOFT, 2019).

Figura 3 Infraestrutura do PI System.



Fonte: Adaptada de Osisoft® (2019)

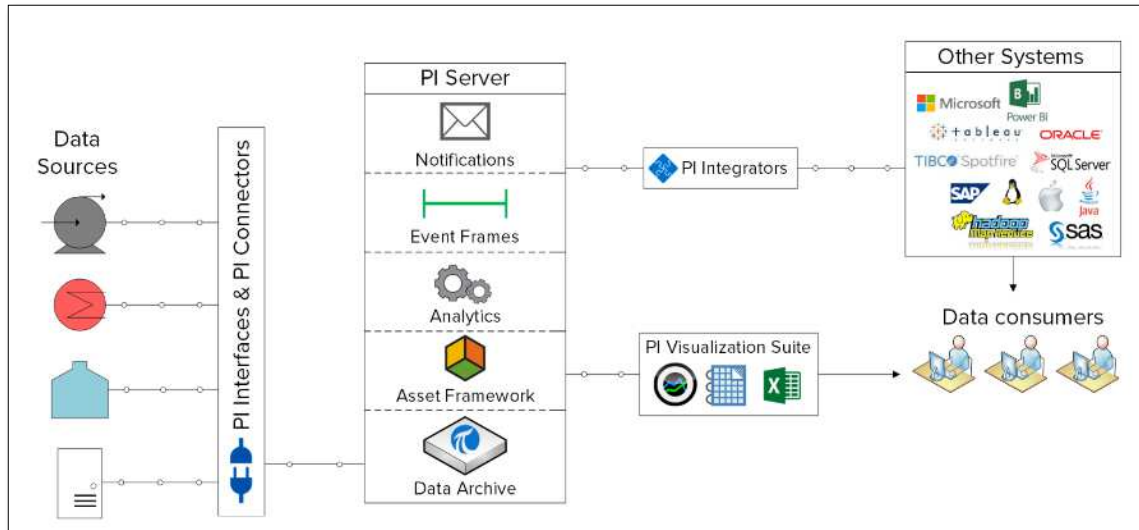
As conexões do PI System são os elementos essenciais para a coleta de dados dos processos. A quantidade de interfaces, programas dedicados a comunicação entre sistemas diferentes, é vasta e que chega a ultrapassar a quantidade de 450 interfaces diferentes. Isto permite que o PI System seja conectado aos mais variados sistemas aplicados em coletas em nível de chão de fábrica, tais como SCADA (Controle Supervisório e Aquisição de Dados), SDCDs (Sistema Digital de Controle Distribuído) e RDBS (Gerenciador de Dados Relacional).

Dentre todos os elementos do PI System, o PI Server merece o maior destaque. É através deste que todos os dados são arquivados e organizados, em formato temporal, ou seja, ele o servidor do banco de dados temporal. Diferente de um banco de dados convencional, no qual os dados são armazenados em tabelas e colunas e relacionados por IDs, o PI Server organiza e armazena os dados por meio das Tags, em base temporal nos arquivos denominados *Data Archive*. O PI Server possui funções tais como, segurança de dados, cálculos em tempo real, definição de eventos de processo, notificações automáticas, monitoramento do sistema e administração do sistema (*PI System Management Tools e PI System Directory*) (OSISOFT, 2019).

As ferramentas do PI sistema são aplicadas a necessidades específicas, a Figura 4 demonstra a sua variedade. O contexto geralmente define a ferramenta adequada para análise ou para apoio a tomada de decisão estratégica, frente à operação do processo em gerenciamento. Estas ferramentas possibilitam, por

exemplo, acesso aos dados em diversos dispositivos, visualização de processos, consulta de dados pelo Excel, *Dashboards* com *SharePoint* e registro manual de dados.

Figura 4 Infraestrutura do PI System e suas conexões.



Fonte: Osisoft® (2019)

O fluxo de dados adotado pelo *PI System* pode ser observado na Figura 4. Esta demonstra o processo pelo qual os dados coletados pelo *PI interface* passam até serem gravados pelo *PI Server* no Arquivo de dados Primário. A lista de etapas tem o seguinte caráter:

1 – Nesta etapa os dados são coletados por interfaces do *PI system*. Os dados podem vir de servidores *OPC* ou *SDCDs*, correspondendo a fonte de dados, por exemplo. Nesta camada ocorre a integração da camada de automação e controle com a camada *PIMS*.

2 - Quando os dados estão na *PI Interface* eles passam pelo processo de exceção de dados, forma pela qual acontece a exclusão de partes dos dados. Acontece desta forma a exclusão de ruídos decorrente do sensoriamento. Depois disso, os dados são armazenados em *Buffer*, o qual é responsável por garantir a preservação dos dados caso uma falha na comunicação como o *PI Server*.

3 - A chegada dos dados no *PI Server* é através do *PI Snapshot*. Nessa fase os dados são tratados novamente. Os dados passam então pelo algoritmo de compressão de dados. Nessa fase, o algoritmo *SDT* é responsável por excluir os

dados que não passam pelos seus critérios de processamento. Em direção a fila de eventos, os quais passam pela Memória *Cache* e aguardam o processamento.

4 - A última fase do fluxo de dados está na gravação dos dados que estavam na fila de gravação no arquivo primário. À medida que os Arquivos Primários alcançam a quantidade de memória estabelecida, um novo arquivo primário é criado e o anterior torna-se um arquivo secundário.

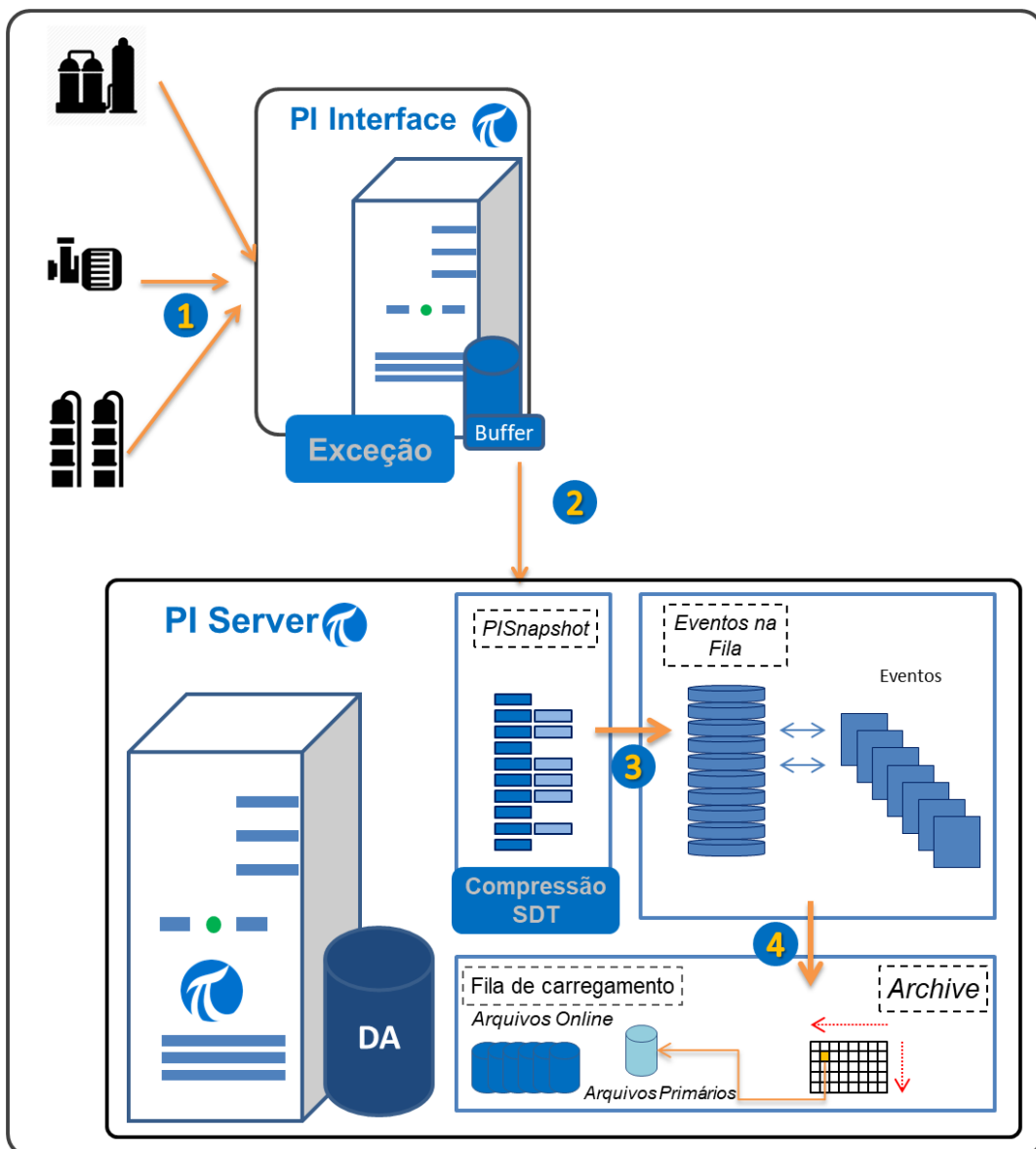
Para os casos nos quais a redundância está presente, ou seja, mais de um *PI Server* está em operação, a fase dois acontece em ambos de forma simultânea. Esse processo define uma opção de manutenção em caso de inoperabilidade das máquinas que compõem o sistema. Em casos deste tipo, torna-se possível restabelecer tanto os dados quanto a infraestrutura do sistema com as mesmas condições e configurações anteriores ao desmantelamento.

No mesmo sentido, a manutenção dos dados é garantida pelas opções de backup dos dados. Para a operação em situações normais do sistema, a configuração do backup em disco específico, em máquina local ou remota em reação a maquia de operação do *PI Server*, é estimulado como forma de garantir a integridade de arquivos secundários, que não recebem novos dados do sistema.

A

Figura 5 demonstra detalhadamente o fluxo de dados, considerando agora os processos que ocorrem entre a entrada de dados nas interfaces de coleta até o armazenamento de dados nos arquivos do *PI Data Archive*.

Figura 5 Fluxo de dados do PI System.



Para esclarecimento da relação entre cada etapa do processo e os algoritmos de compressão de dados, a seguir são listadas todas as etapas assim como suas respectivas descrições:

**1. Coleta de dados pela interface:** O PI System possui mais de 450 interfaces de coleta de dados, a escolha de uma interface depende do tipo de dado a ser coletado, origem e protocolo de comunicação. Após coletado de dados, estes passam pelo algoritmo de exceção de dados, caso este seja de tipo numérico e não definido como *step point*, o atributo *step* da do PI Point deve ser igual a 0 (Zero). Os pontos são enviados para o *PI Buffer Subsystem*. Caso este seja ativado, o algoritmo de compressão também é executado na interface. O Buffer Subsystem, no entanto, não deleta os dados a serem arquivados, ele os marcar com dados comprimidos. Todos os pontos são enviados para o *PI Snapshot*.

**2. Entrada no *Snapshot*:** O snapshot tem a função de prover dados em tempo real para das ferramentas clientes do PI System assim como reter os dados que passarão pelo algoritmo de compressão. Há dois comportamentos previstos a depender dos status de configuração do *PI Buffer Subsystem*. Caso este esteja ativado para a referida PI tag, os dados marcados como arquivados são enviados diretamente para a fila de Eventos, caso contrário, os dados recebidos pelo snapshot são processados pelo algoritmo de compressão e os dados que passam por este são enviados a fila de eventos.

**3. Entrada na *Event Queue*** (Fila de Eventos): A fila de eventos é um buffer de um ou mais arquivo de eventos destinados a serem arquivados pelo PI Data Archive. Os eventos são constituídos por dados que são armazenados integralmente nos PI Archives. A fila de eventos existe para que o serviço de armazenamento tenha controle de prioridade durante a escrita de dados no PI Archive primário.

**4. Envio dos dados para o *PI Archive Primário*:** O *PI Data Archive Subsystem* executa a escrita de dados nos eventos. A escrita é realizada em ordem de prioridade. A fila de eventos portanto, evita que o congestionamento deste serviço ocasione perda de dados. A escrita ocorre até que o *archive* primário alcance tamanho definido após isso passa-se à criação de outro archive file, processo denominado *archive shift*. Por fim um novo archive file é definido como primário e anterior é definido com secundário,

disponível para ser consultado em caso de consultada de dados histórico e não recebe novos valores.

### **3.3. Compressão de dados**

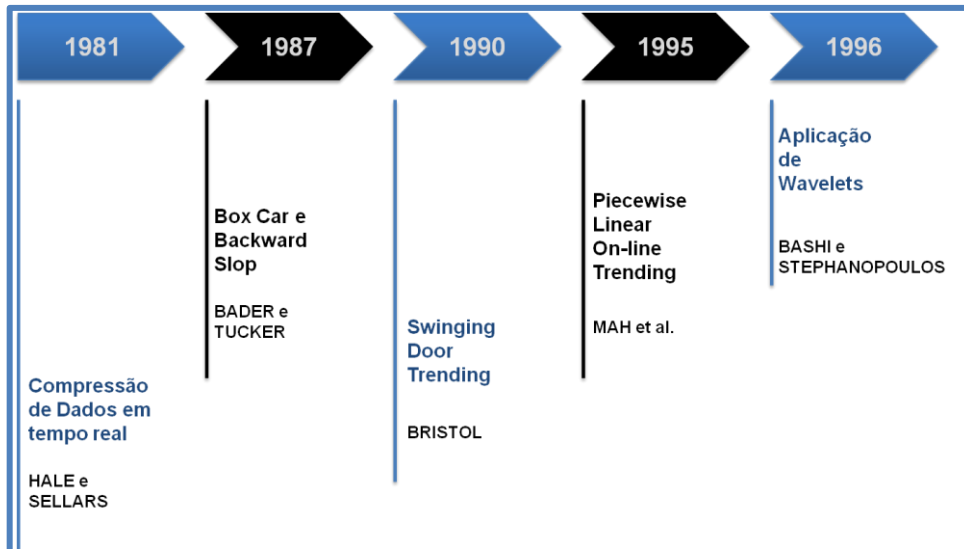
À medida que a coleta de dados industrial se intensificou, o volume de dados destinados ao armazenamento ganhou destaque. A massa de dados em transição em redes de comunicação também cresceu por consequência. Desse modo, determinar alternativas de controle ou diminuição destes aspectos na base de dados ganhou destaque.

No entanto, a análise do processo de compressão de dados aplicado a dados de processos recebeu atenção tardia. Quando se compara com as aplicações outras técnicas de automação, tais como redes industriais, protocolos de comunicação, esta diferença torna-se evidente.

Uma linha do tempo é apresentada a seguir na Figura 6. Composta por trabalhos de relevância para o desenvolvimento dos algoritmos de compressão de dados aplicados pelos sistemas historiadores mais comumente aplicados nos mais diversos setores, *PI System* (Osisoft®) e *InfoPlus21* (AspenTech®)

Figura 6 Linha do tempo de determinantes sobre compressão de dados.





Fonte: Próprio autor

Na vanguarda da análise dos historiadores, Hale e Sallars (1981) focaram na avaliação das estruturas de computadores e rede. Constatou-se na época o crescimento em nas estruturas de controle, monitoramento em empresas de setores de químico e de refino de petróleo. Através da rotina de coletas de dados e relatórios diários, a responsabilidade de lidar e gerenciar os dados industriais das fabricas da *DuPont®* foram tratados concluindo-se a vantagem de se aplicar os sistemas historiados nas plantas industriais.

Bader e Tucker (1987) iniciou análise dos métodos compressão de dados aplicados aos sistemas historiadores. Foi responsável por propor o método *Box Car e Backward Slop*, método hoje aplicado no sistema PIMS do Historiador comercial *InfoPLus da Aspentech®*.

Bristol (1990) foi o responsável por introduzir outro método de compressão de dados, o *Swinging Door Trending*. Este método consiste na avaliação dos dados ponto a ponto, segundo os seus limites de limites superiores e inferiores, construídos sobre o primeiro ponto arquivado e último ponto em análise.

Este método propõe avaliar se os dados estão dentro um paralelogramo de validação. Neste caso os parâmetros como altura, e comprimento do paralelogramo são definidos pelo usuário, correspondendo a soma dos desvios e a diferença máxima de tempo entre o último ponto armazenado e último ponto analisado.

Como o método proposto por Bader e Tucker (1987), este método é do tipo de compressão por perda de dados. Os dados excluídos são os considerados redundantes, presentes nos dados brutos. O algoritmo aplicado ao historiador *PI System* possui bases firmadas nesse algoritmo, porém, com a adição de uma etapa anterior de linearização segmentada, denominada como exceção (ALSMEYER, 2005).

Mah et al.(1995) introduziu outro método de compressão de dados por perdas de dados *Piewise Linear On-line Trendin (PLOT)*. Este é proposto como alternativa ao *Box Car* e *Backward Slop* e ao *Swinging Door Trending*. O argumento no qual fundamenta este método contraria a utilização de ambos por alterarem a variabilidade dos dados e gerarem muitos pontos desviantes, do inglês *outlier*. O método PLOT possui embasamento estatístico e desempenho superior aos demais métodos propostos. Outra característica do PLOT é capacidade de adaptação à variabilidade de processos e a presença de ruídos.

Backlsh e Stephan (1996) foram além no desenvolvimento de métodos de compressão de dados. Baseados em na Teoria das *Wavelets* o método promete a compressão de dados sem a exclusão de dados e com altas taxas de compressão de dados. Neste trabalho, algumas análises dos parâmetros de operação de novo algoritmo foram realizadas. Pode-se constatar, segundo os autores, que os métodos tradicionais aplicados comercialmente possuem baixo desempenho quando comparados com o método apresentado baseado em *wavelet*.

### **3.4. Avaliação do desempenho da compressão de dados segundo a dinâmica**

Na realidade industrial, as variáveis que estão em coleta por sistemas historiadores não possuem todas as dinâmicas iguais. Existem, no entanto, grupos de variáveis que possuem dinâmicas diferentes entre si e que devem ter tratamento correspondente.

A consideração da dinâmica de associada é inerente ao processo que é objeto de mensuração (SEBORG, 2005). Ao considerar a amostragem de temperatura, fluxo volumétrico e a pressão, por exemplo, faz-se necessária a avaliação prévia por profissionais que são capazes de definir o tempo de resposta necessário para atuação

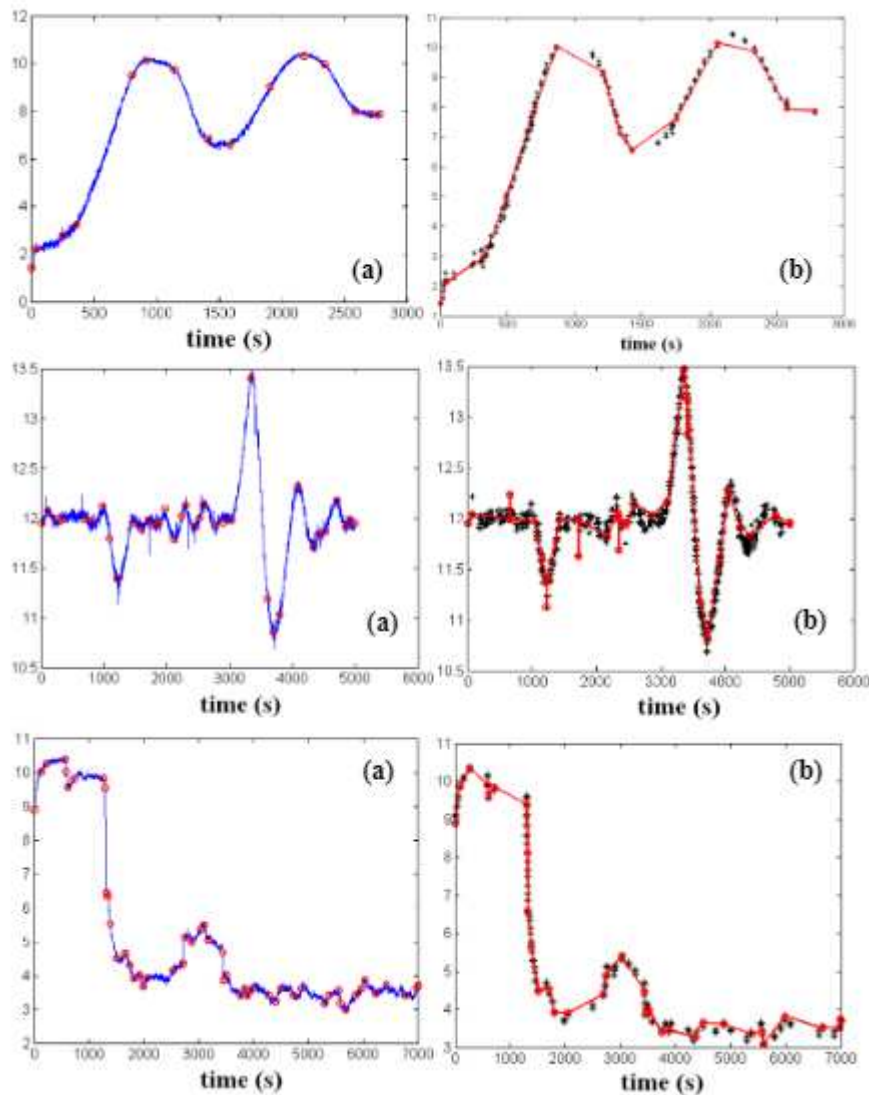
ou ajuste sobre o sistema. Nesse caso, para que haja a possibilidade de ajuste dos parâmetros de compressão de maneira sistemática, lança-se mão das médias estatísticas de centralidade e variabilidade. Essas grandezas possibilitam a aplicação da mesma metodologia, independente da dinâmica da variável a ser mensurada e arquivada (SEBORG, 2005).

Silveira et al (2012) desenvolveram uma metodologia de avaliação dos parâmetros do algoritmo de compressão considerando pontos característicos de dinâmica de processo do qual os dados eram coletados. Nesta metodologia, o parâmetro *CompDev* é calculado por um processo de otimização numérica com restrições. As restrições eram impostas pela seleção manual de pontos considerados importantes e que deveriam ser mantidos após o processo a execução do algoritmo de compressão.

A vantagem desta metodologia é que este garante que pontos selecionados são armazenados ao fim da compressão, possibilitando que a dinâmica do processo seja reconstruída a partir destes e demais que não foram selecionados manualmente. A desvantagem é que esta seleção manual se torna impraticável em sistemas reais, pois a seleção manual de pontos em sistema com grande quantidade de PI Points torna-se onerosa.

Na Figura 7, são apresentados (a) dados brutos com ponto pré-selecionados em vermelho, (b) ponto em preto após exceção e pontos arquivados em vermelho, linha vermelha são os pontos reconstruídos:

Figura 7 Seleção de pontos para execução da compressão e exceção.



Fonte: Adaptado de Silveira et al, 2012.

Para a metodologia aplicada por Silveira et al (2012), os gráficos (a) correspondem a seleção de pontos chave para a manutenção da dinâmica dos dados da variável em análise. Os eixos X e Y correspondem respectivamente ao tempo (segundos) e aos sinais de variáveis de processo, (SILVEIRA et al, 2002). Um algoritmo de otimização foi aplicado na combinação de ambos os algoritmos de exceção e compressão de dados. A função objetivo foi definida como o erro quadrado entre os dados originais e dados reconstruídos.

No entanto, a metodologia adotada não define métricas estatísticas de desempenho como forma de avaliação da compressão de dados. A quantidade de pontos definida antes da aplicação da otimização é definida manual e empiricamente, de acordo com a experiência do responsável pelo processo.



### 3.5. Controle Estatístico de Processos

Considerada um elemento importante para a gestão da qualidade total, o controle estatístico de processo (CEP) é reconhecido em âmbito global como um caminho para manter a cultura do engajamento na melhoria da qualidade. Em um ciclo contínuo de melhoria e manutenção da qualidade, o CEP define o grau de competitividade em mercado global do processo produtivo. Tendo-se o foco tanto no âmbito da operação quanto na esfera dos produtos manufaturados (CAULCUTT, 1995).

Objetivo do controle estatístico de processos é a detecção de cenários nos quais ocorre a transição de um estado desejado para um estado indesejado de um processo sob monitoramento. A detecção da mudança é realizada baseada em conceitos e técnicas estatísticas que desempenham tarefas como coleta, classificação, análise e interpretação dos dados.

Quando detectado um cenário de transição para condição de fora de controle estatístico, a necessidade de diagnóstico de processo é evidenciada, por definir a causa raiz da perda do estado desejado e possibilitar ao operador do processo tomar as ações necessárias para corrigir a situação especial e redirecionar o processo para o estado de operação desejado (CINART et al, 2005).

Foi durante os anos de 1920 e 1930 que os fundamentos básicos do CEP e das cartas de controle foram propostos por Walter A. Shewhart, até meados de 1970, alguns avanços foram realizados. Porém, a pesquisa científica nesta área era escassa. No entanto, com a mudança de cenário mercadológico, devido ao progressivo aumento em competitividade cada vez mais globalizada, no qual a melhorias na qualidade fundamentais para a sobrevivência de muitas indústrias no ambiente mercadológico.

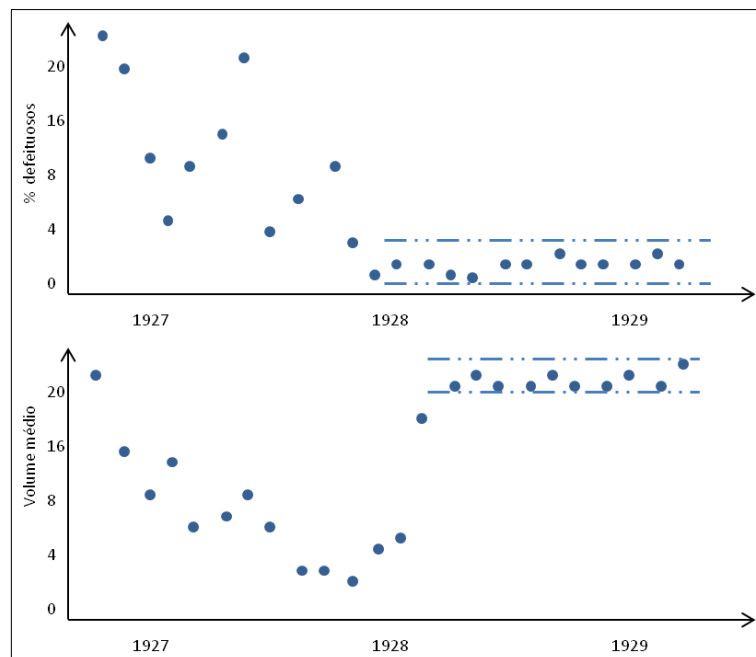
Historicamente a análise de variáveis de processo a partir de cartas de controle teve um papel importante para o desenvolvimento de equipamentos e produtos com qualidade e baixa de taxa de defeitos. Introduzido o conceito de cartas de controle por Walter A. Shewhart, quando na *Bell Laboratories*, estas só vieram a ter sua aplicação consolidada a partir da Segunda Guerra, direcionadas a produção em larga escala de produtos manufaturados nesse período. A sua estruturação partiu do objetivo de

definir uma metodologia voltada a manutenção dos produtos sob a condição de menos variabilidade possível. A aplicação desta metodologia foi direcionada as variáveis chaves para o controle da qualidade dos processos de manufatura (Montgomery, 2012).

A introdução da análise estatística mostrou-se uma estratégia valiosa. Shewhart definiu com os gráficos de controle univariados. Este corresponde ao estabelecimento de elementos essenciais a avaliação da qualidade dos processos produtivos. Além disso, possibilitou a integração entre a área gerencial e operacional de maneira sinérgica, contribuindo em direção à otimização e detecção de falhas operacionais.

Em seu livro introdutório dos conceitos sobre cartas de controle, Shewhart (1980) estabeleceu a relação do controle estatístico com a necessidade de inspeção dos processos foi realizada. A comparação entre os momentos em que o processo esteve sem ou com avaliação estatística pode ser enxergada na Figura 8 a seguir. É nítida a diferença entre os dois momentos, quando se considera os valores: quando o controle estatístico de processo foi estabelecido, a partir de janeiro de 1926, a variabilidade mudou, estabelecendo-se entre os limites inferior e superior definidos para o percentual de defeitos e volume médio das variáveis analisadas. Produto dessa redução de estabilidade pode-se perceber a baixa necessidade investigação e inspeção, estabelecendo-se uma condição de redução dos custos do processo e melhoria da qualidade dos produtos manufaturados.

Figura 8 Abordagem de equilíbrio ou controle, quando causas atribuíveis são eliminadas, portanto reduzindo a necessidade de inspeção.



Fonte: Adaptado de Shewhart (1980)

Considerando o caso de cartas de controle univariadas, as variáveis de processo tomadas para o monitoramento são tais como variáveis de controladas ou variáveis manipuladas. Por ter uma grande familiaridade com o processo, os operadores selecionam estas variáveis segundo critérios relevantes. A frequência de amostragem, intervalo de variação da medida e a dinâmica são críticos na determinação de restrições do processo em análise (MONTGOMERY, 2012).

Outras técnicas de controle estatístico podem ser aplicadas além da análise cartas de controle univariadas, as denominadas técnicas multivariadas. Neste caso, a autocorrelação entre variáveis deve estar presente, o que afeta diretamente os resultados das cartas de controle univariadas, caso sejam aplicadas em variáveis desta natureza e o surgimento de erros tipo I e tipo II. Exemplos de cartas controle multivariada são cartas de controle de X-barra com limites de Bonferroni, gráficos  $T^2$  de Hotterling, Análise de Componentes principais (SOUZA, RIGÃO, 2005).

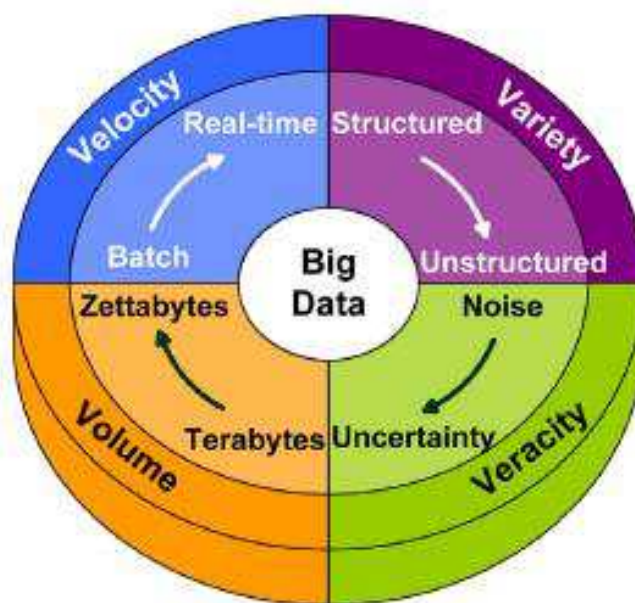
O impacto da compressão de dados foi avaliado por Imitez, Choudhury, Shah (2006). Em seus estudos, a comparação entre os algoritmos de compressão *Swinging Door Trending* e *Wavelet* foi realizada sob o contexto da análise multivariada de



dados. Foi demonstrado que o impacto do processo de compressão pode chegar a inviabilizar análises multivariadas de dados a partir dos dados reconstruídos, por não manterem a correção entre as variáveis comprimidas.

O interesse em avaliar em tempo real os ativos industriais é crescente. Em busca da alta performance, a combinação entre eficiência e inovações tecnológicas vem à tona, rotulada como manufatura inteligente, do inglês *Smart Manufacturing*. Surge então a oportunidade de ampliar a aplicação de técnicas clássicas de avaliação e monitoramento. Como sugerido por He e Wang (2017), a aplicação dos dados em tempo real para o monitoramento de ativos promete ter sua expansão e evolução graças as 4 V's característicos do Conceito presentes hoje em sistema de informação de plantas industriais, como representado na Figura 9.

Figura 9 As propriedades do Big Data e suas nuances.



FONTE: WANG (2017)

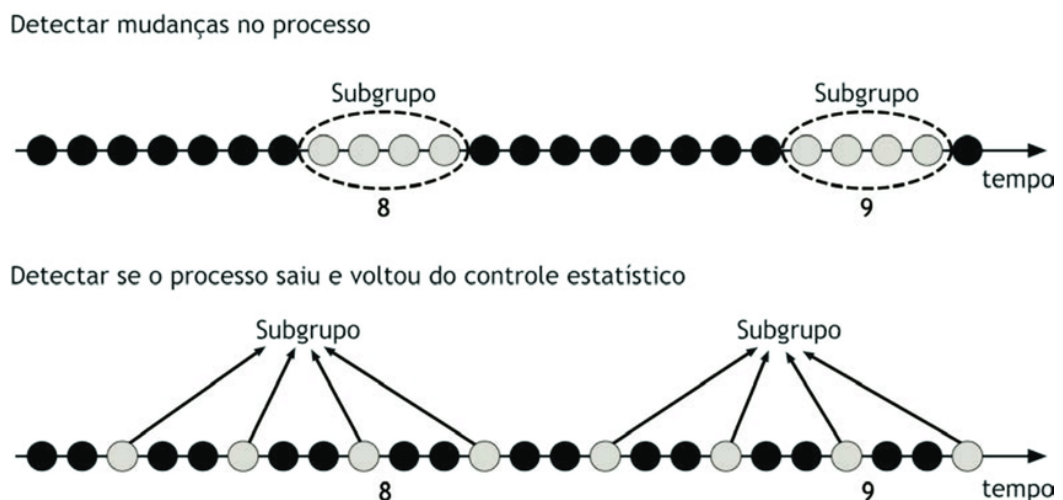
Surge a partir das considerações de Choudhury et al (2004) e Wang (2017) a possibilidade de avaliar algo que é fundamental nos dados industriais, sua veracidade das informações contidas nos dados, mesmo após o processo de compressão. Nesta avaliação, as cartas de controle surgem como um estudo de caso direcionado para averiguar a existência de algum ruído ou modificação de resultados à medida que o grau de compressão, ou até mesmo construir para o condicionamento de limites em variáveis chave de processo. Considerando tal necessidade, aplicar as cartas de

controle clássicas, como X-Barra e R-Barra, e X-Barra e S-Barra se torna uma opção plausível. Considerando a aceitação de desempenho robusto destas cartas, pode-se averiguar que a validação da qualidade dos dados é também um fator que pode ser tratado como prioritário, caso o interesse em desenvolver uma estrutura de monitoramento.

### 3.5.1. Subgrupos Racionais

Baseando-se na disponibilidade de coleção de dados ou amostras, as cartas de controle são construídas para a detecção da presença de causas especiais no processo. A partir das coleções de dados ou amostras são construídos os subgrupos racionais, selecionados de maneira a evidenciar a possível presença de variabilidade nos dados originais, caso estejam presentes. A Figura 10 demonstra a criação de dois subgrupos racionais, 8 e 9. Estão definidos pela coleta de 4 valores assumidos pela variável monitorada. Portanto, o valor n é o assumido como a quantidade de pontos necessário para representar o estado do sistema.

Figura 10 Definição de subgrupos racionais.



Fonte: (MONTGOMERY e RUNGER, 1992)

Impacto que o agrupamento ou amostragem dos dados amostrados pode ser grande sobre o comportamento dos gráficos de controle. Na condição de observações correlação temporal positiva, os limites de controle serão mais próximos um do outro, conduzindo a grande quantidade de alarmes falsos, devido a estruturação inadequada dos dados observados.

Os subgrupos racionais são uma maneira de rearranjar os dados coletados em subgrupos homogêneos internamente, ou seja, assume-se que as causas assinaláveis são constantes, compostos por itens os mais homogêneos possíveis, em cada subgrupo, de maneira a evidenciar a diferença entre grupos (MONTGOMERY e RUNGER, 1992).

O tamanho mínimo de quatro itens por cada subgrupo é justificado aplicando-se o teorema do limite central, definindo que a distribuição das médias amostrais ( $\bar{X}$ -barra) se aproxima da distribuição normal, quando as amostras são de tamanho igual ou maior a quatro, mesmo considerando as suas origens numa população não normal. Esta consideração demonstra-se um útil para garantir a distribuição normal das médias e ajuda na interpretação e definição dos limites das cartas de controle (MONTGOMERY e RUNGER, 1992).

### 3.5.2. Cartas de controle: Conceitos e Princípios

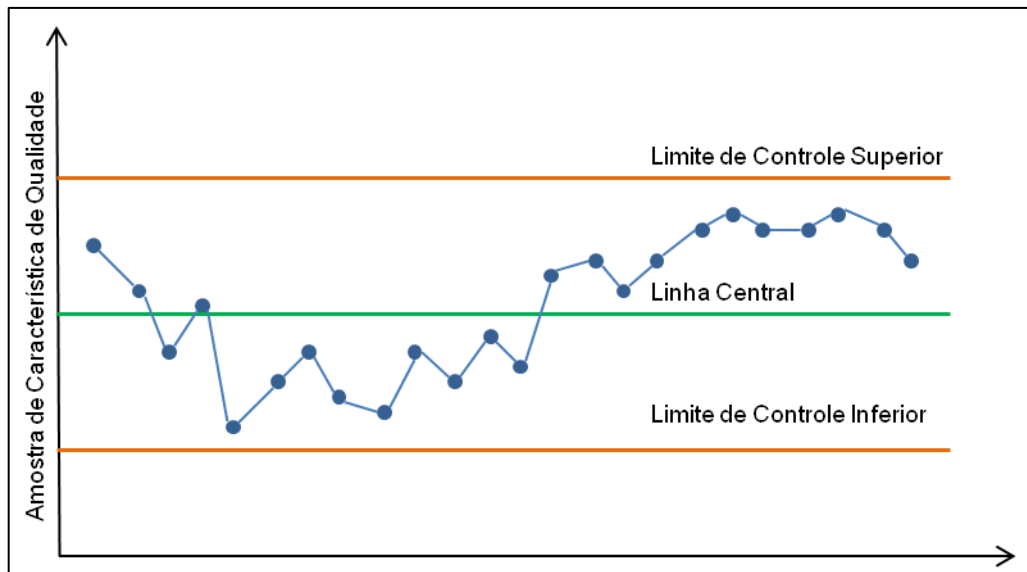
Cartas de controle são gráficos de visualização de características de medidas ou características de alguma grandeza a ser monitorada. A sua aplicação é direcionada a verificação da existência de causas de variação do processo, as quais são classificadas em dois grupos:

Causas comuns: são aleatórias e inerentes ao processo, inevitáveis. Tidas como impossíveis de serem previstas, identificadas ou corrigidas.

Causas especiais: são causadas por algum distúrbio de processo, falha em equipamento ou motivo de fácil identificação, como erros de procedimentos. Estas causas podem ser corrigidas, de modo a evitar distúrbios e instabilidades de processo.

Na Figura 11, há um exemplo genérico de carta de controle. Nela, estão presentes os Limites de controle inferior e superior assim como a linha do centro.

Figura 11 Exemplo de carta de carta de controle.



Fonte: Montgomery (2012)

O uso desta ferramenta de monitoramento se dá pela verificação da posição de cada ponto situado no gráfico, os quais podem assumir duas condições:

1. Entre os limites de controle superior e inferior:

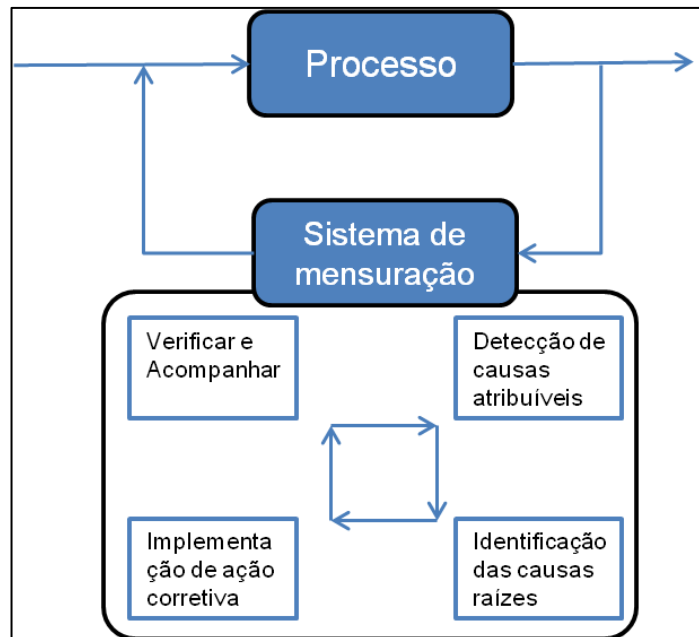
Nesta condição, se diz que o sistema monitorado está sob controle estatístico de processo. Ou seja, só existem causas comuns no sistema monitorado, nenhuma ação sobre este é necessária.

2. Acima ou abaixo dos limites de controle superior e inferior

Este estado do sistema é dito fora de controle estatístico de processo, pela existência de causas especiais no processo. Portanto, faz-se necessária avaliação das condições de operação ou investigação de funcionamento anormal de componentes do sistema monitorado, causadoras de variabilidade estatística no processo.

A Figura 12 a seguir demonstra a atuação das castas de controle. Nesse ciclo, o principal objetivo é definir a estabilidade do sistema, se há causas especiais no processo que causem variabilidade no processo. Este tipo de análise é realizada dentro de um ciclo de avaliação contínuo de monitoramento (MONTGOMERY, 2012).

Figura 12 Melhoria do processo a partir de cartas de controle.



Fonte: Montgomery (2012)

Os quatro elementos para o aprimoramento da qualidade fazem parte de um processo contínuo e de pode requerer um plano de ação para a manutenção da ausência de causas especiais controle de processo estatístico. Pode-se enumerar:

1. Detectar as causas especiais
2. Identificar o causa raiz do problema
3. Programar ação corretiva
4. Verificar por melhorias e dar continuidade da operação

## 4. Fundamentação Teórica

Esta seção apresenta os objetos de estudo, a sequência sugerida a diante. Primeiro aborda-se o processo de compressão aplicado no PI System, definido pela combinação entre os algoritmos de exceção e compressão de dados. Em seguida, os indicadores estatísticos são apresentados e, por fim, os conceitos aplicados de controle estatístico de processos.

### 4.1. Processo de compressão de dados

O processo de compressão de dados aplicado no PI System é constituído pela combinação de dois algoritmos denominados de exceção e compressão de dados. Estes são responsáveis pela filtragem de dados coletados a serem escritos no banco de dados de base temporal os *PI Archives*, gerenciados pelo *PI Data Archive*.

Ambos os algoritmos são tidos como mecanismos de filtragem de dados, os quais definem quais dados são mantidos e excluídos. Este processo é conhecido por compressão de dados por exclusão de dados, existindo outras técnicas de aderentes a outros tipos de dados, como a compressão realizada sobre vídeos e imagens. Para o contexto de dados de base temporal os algoritmos por exclusão de dados têm sua aplicação consolidada.

É possível listar os benefícios da aplicação dos ambos os algoritmos em sistemas historiadores de dados industriais (Osisoft, 2021):

Tabela 1 Benefícios da aplicação do processo de compressão de dados no PI System.

Benefícios	Descrição
Armazenamento	Redução da quantidade de espaço necessário para o armazenamento de dados, mantendo mais espaço disponível para utilização.
Velocidade de Transmissão	O tempo necessário para o envio de dados pela rede depende da quantidade de dados enviada. A redução da massa de dados minimiza a o tempo necessário para o fluxo de dados entre os PI Componentes disponíveis dentro da rede.
Arquivamento e Backup	Pela redução da quantidade de dados, processos como o arquivamento e backup ganham eficiência.
Performance do PI System	A redução da quantidade de dados a ser processada pelo PI System permite que o sistema armazene mais dados para uma mesma capacidade de armazenamento de dados, permitindo a minimização do tempo de consulta assim como a diminuição do tempo necessário para realizar um processo de backup de dados.

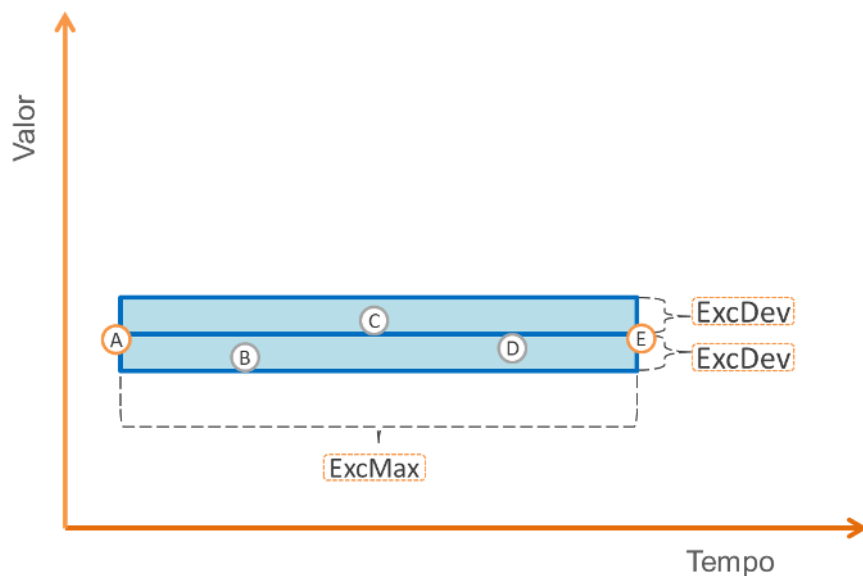
Fonte: Osisoft, 2021

#### 4.1.1. Algoritmo de exceção de dados

O algoritmo de exceção de dados é responsável por remover valores que não mudam com o tempo ou valores que mudam dentro dos limites de acurácia do instrumento. Para o primeiro caso, não é necessário aguardar valores que não tem alteração alguma em seus valores com o passar do tempo. Enquanto o segundo define o que é denominado ruído nos dados, essencialmente valores que não tem significância, originados por erro instrumental.

A exceção de dados aplica uma banda morta com relação ao primeiro ponto, os pontos que ficam dentro desta são excluídos. Outra condição que ocorre é o limite de tempo para que o algoritmo armazene um valor. A Figura 13 demonstra todos os parâmetros que definem o comportamento do algoritmo de exceção de dados.

Figura 13 Parâmetros do algoritmo de exceção de dados.



Fonte: Adaptado Osisoft® (2019)

Cada tag do PI System contém os parâmetros presentes na Figura 13 anterior, tratados como atributos de *PI Point*. O significado de cada um assume é apresentado detalhadamente na Tabela 2.

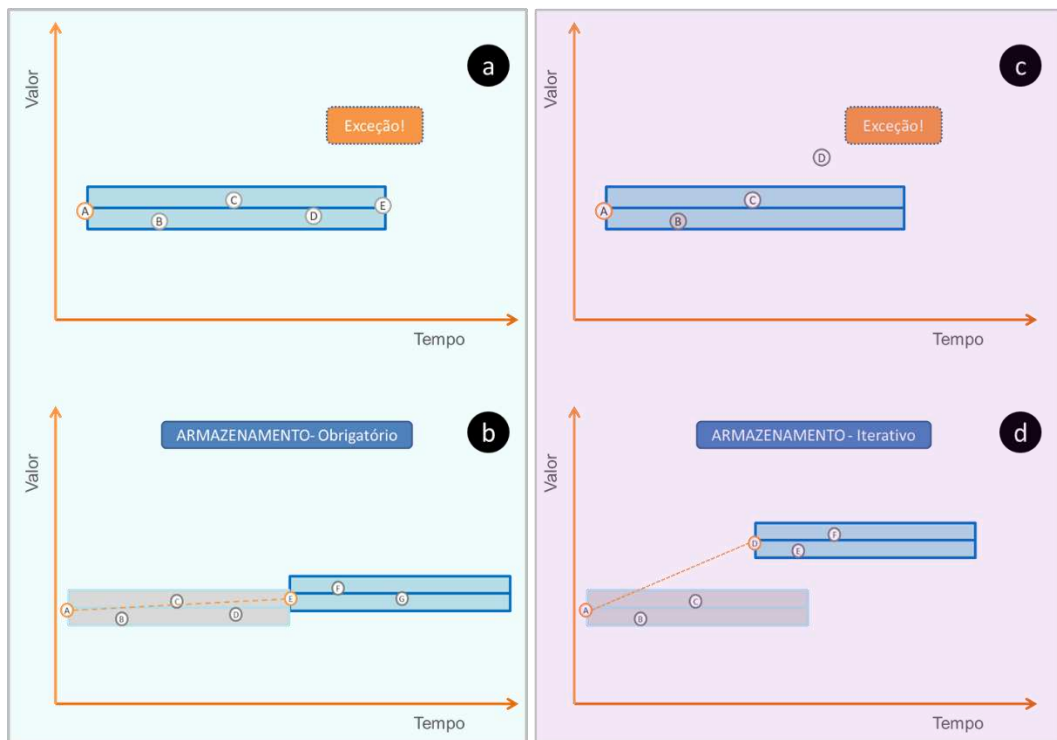
Tabela 2 Significados dos parâmetros do algoritmo de exceção.

Parâmetro	Descrição
ExcDev	Quanto o valor de cada tag precisa mudar antes que a <i>PI Interface</i> envie o valor para o <i>PI Data Archive</i> .
ExcMax	Define o limite de tempo que a <i>PI Interface</i> pode ficar sem reportar um valor para o <i>PI Data Archive</i> .
ExcMin	Define a frequência que a <i>PI Interface</i> deve reportar valores para o <i>PI Data Archive</i> .

Fonte: Adaptado Osisoft® (2019)

Para demonstrar duas condições que podem ocorrer quando os valores de ExcDev ou ExcMax acontecem, a Figura 14 é apresentada. Das quatro imagens, duas condições são apresentadas, imagens a e b definem a condição 1 e imagens c e d a condição 2. A imagem **a** define o processo de exceção, a imagem **b** define o começo de uma nova banda morta começando do ponto **E** definida verticalmente por  $+ExcDev$  e  $-ExcDev$ . Entre **a** e **b** aconteceu o envio do ponto **E** e exclusão dos demais pontos. A imagem **c** define também o momento de exceção de dados. A partir do ponto **D** uma nova banda morta é criada, como demonstrado na imagem **d**, na qual evidencia-se que os pontos **b** e **c** são excluídos.

Figura 14 Condições de envio de dados pelo Algoritmo de exceção de dados.



Fonte: Próprio autor



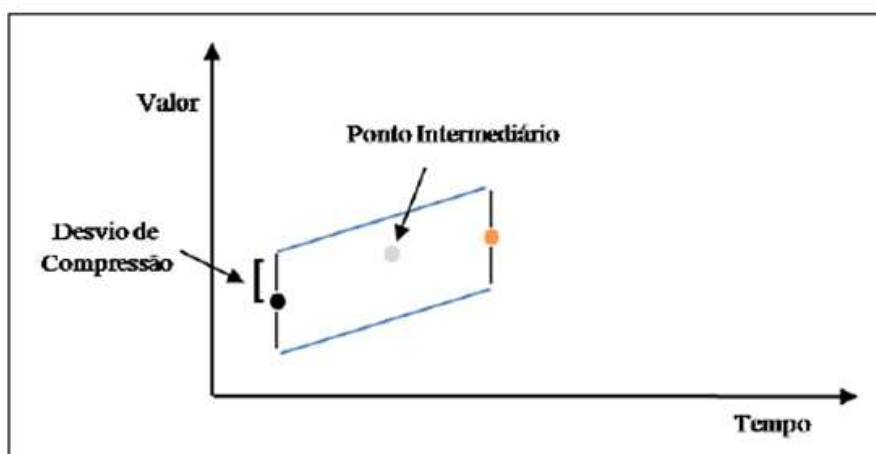
#### 4.1.2. Algoritmo de compressão - *Swinging Door Trending* (SDT)

Apresentado por Bristol (1990), o algoritmo é apresentado e demonstra as potencialidades quanto a sua robustez. Possuindo dois parâmetros de sintonia, a tempo máximo de compressão e desvio de compressão, e um taxa como saída, a taxa de compressão. A relação entre estes é a seguinte, quanto maior o tempo máximo de compressão e desvio de compressão, maior será a quantidade de pontos excluídos, por conseguinte, maior a taxa de compressão de dados, a qual é definida após o processo.

A análise do algoritmo pode ser realizada graficamente. Tomando-se um gráfico, no qual o eixo das abscissas corresponde ao tempo e o eixo das ordenadas os valores das grandezas a serem medidas. O acompanhamento deste gráfico com o tempo supõe a introdução de novo ponto à medida que novas amostras são colhidas por sensores de acordo com o tempo de processamento do sistema de coleta de dados, típica função e PLC's, do inglês *Programmable Logic Controllers*.

Na Figura 15, é possível observar a presença de dois pontos. Ao considerar desvios superior e inferior para ambos os pontos, é possível verificar um paralelogramo que é redefinido a cada novo ponto em análise. O ponto preto corresponde ao ponto armazenado, o ponto cinza é um ponto intermediário e o ponto laranja é o ponto recém-colhido, ou ponto atual. As arestas verticais do paralelogramo presente na figura são definidas como o dobro do Desvio de Compressão (DC), o primeiro parâmetro de entrada a ser analisado.

Figura 15 Conceito geral do *Swinging Door Trending* – SDT.

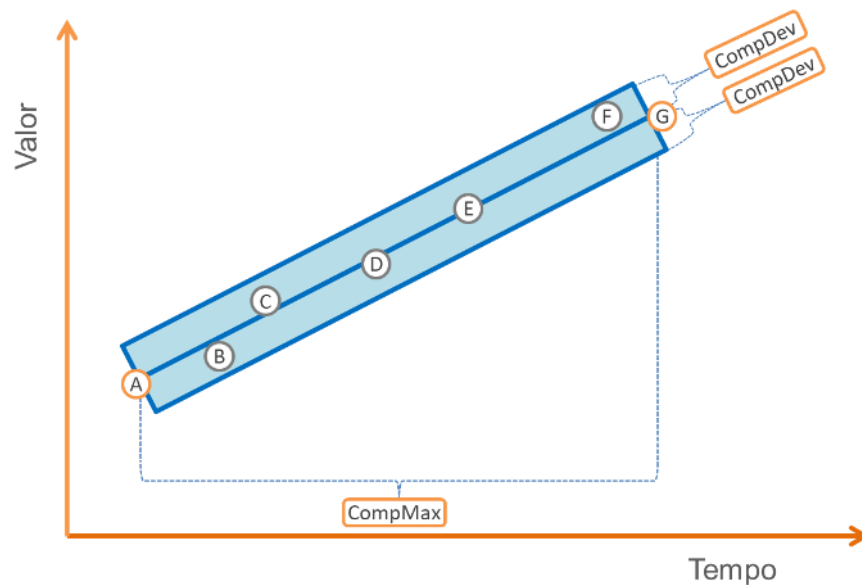


Fonte: Adaptado de NETO (2015)

À medida que a coleta de novos dados se processa novos paralelogramos são formados entre o último ponto armazenado e o ponto atual. Esta é uma maneira didática de demonstrar geometricamente o processo desempenhado pelo algoritmo de compressão de dados. A diferença entre este e o algoritmo de exceção de dados está na forma geométrica aplicada em ambos os processos. O algoritmo de exceção é explicado por um retângulo enquanto o algoritmo de compressão *Swinging Door Trending* é exemplificado por um paralelogramo.

Ao considerar o mesmo algoritmo no ambiente do PI System, é possível ilustrar o mesmo paralelogramo com demonstrado na

Figura 16 Parâmetros do algoritmo de compressão de dados.



Fonte: Adaptado Osisoft® (2019)

Como para o algoritmo de exceção, os parâmetros da compressão de dados são armazenados como atributos de cada *PI Point*. As definições dos parâmetros relacionados a este algoritmo são apresentadas na Tabela 17.

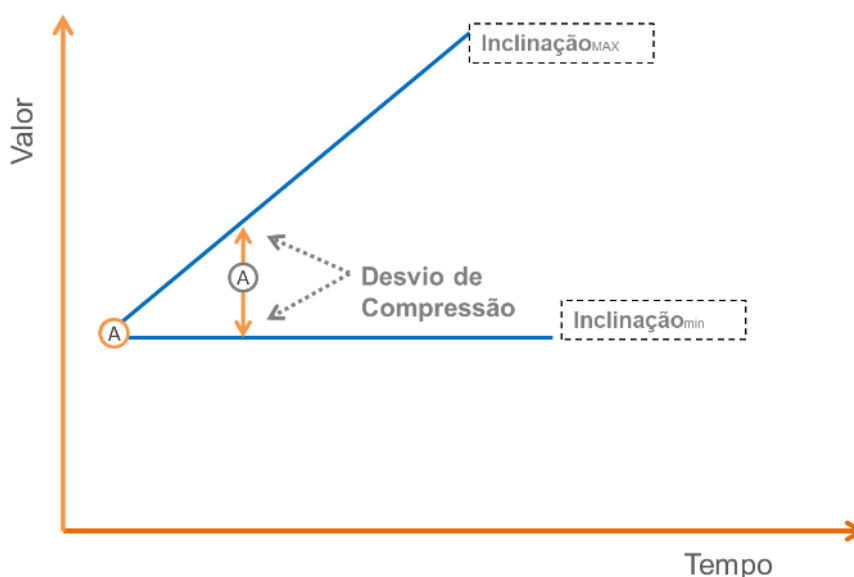
Figura 17 Significados dos parâmetros do algoritmo de compressão.

Parâmetros	Descrição
CompDev	Define quanto os valores do <i>PI Point</i> precisam mudar para que o <i>PI Data Archive</i> os armazene.
CompMax	Define quanto tempo o algoritmo avalia os valores do <i>PI Point</i> considerando o <i>CompDev</i> estabelecido.
CompMin	Define a frequência que um valor deve ser armazenado.

Fonte: Adaptado Osisoft® (2019)

Outra forma de observar o algoritmo é através da avaliação retas traçadas entre estes mesmo pontos. Na Figura 18, as retas a partir do último ponto armazenado até o limite superior ( $S_M$ ) e Limite Inferior ( $S_m$ ) em relação ao ponto atual. Tais linhas delimitam os limites de avaliação do próximo ponto a ser processado durando a compressão.

Figura 18 Início de Avaliação das Inclinações Máxima e Mínima, primeiro ponto avaliado após o último armazenamento.



Fonte: Adaptado Osisoft® (2019)

Com a coleta de outro ponto, a inclinação deste em relação ao último ponto armazenado é calculada. Caso a inclinação do ponto atual esteja entre as  $S_m$  e  $S_M$ , o ponto passa a ser pivô e a gerar as retas que definem os novos limites inferior ( $PS_m$ ) e superior ( $PS_M$ ), os quais diferem no eixo y o valor do *CompDev* para mais e para menos.

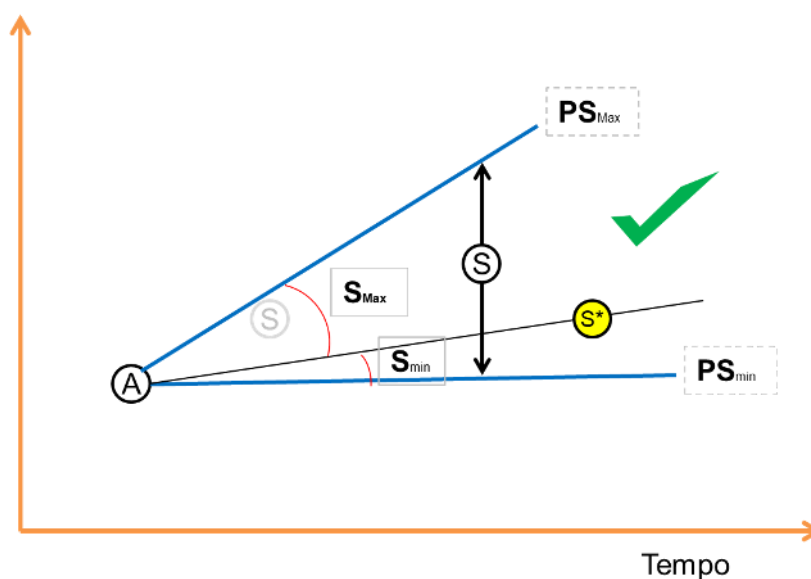
A partir destes, retas  $PS_{max}$  e  $PS_{min}$  com inclinações temporárias máxima ( $S_{Max}$ ) e mínima ( $S_{min}$ ) são traçadas respectivamente até o último ponto avaliado. Compare-se os valores:

1. Se  $S_m$  menor que  $PS_m$ ,  $S_m$  torna-se o valor de  $PS_m$ , do contrário o valor de  $S_m$  permanece o mesmo;
2. Se  $S_M$  menor que  $PS_M$ ,  $S_M$  torna-se o valor de  $PS_M$ , do contrário o valor de  $S_M$  permanece o mesmo

Portanto as inclinações de  $S_{Max}$  e  $S_{min}$  nunca aumentam, devido a essa assimilação de valores de inclinação as inclinações são sempre menores em relação aos pontos atuais. Na

Figura 19 o processo descrito anterior, a inclinação do ponto atual é avaliada, novos  $PS_{Max}$  e  $PS_{min}$  são definidos.

Figura 19 Avaliação da inclinação do segundo ponto.

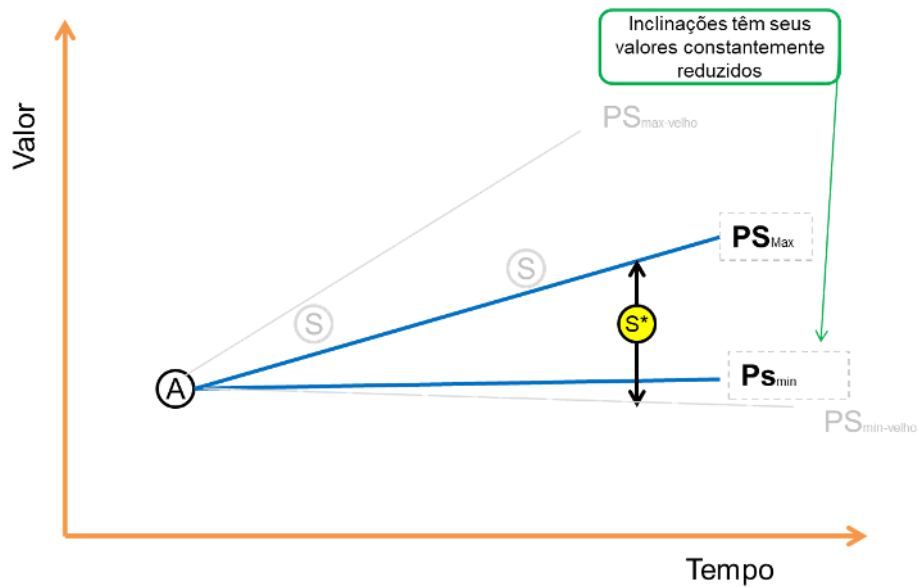


Fonte: Adaptado Osisoft® (2019)

Partindo do último ponto armazenado, traçam-se novas inclinações e avalia-se  $S_m$  e  $S_M$ . Na

Figura 19 a  $S_{Max}$  sofre atualização e  $S_{min}$  permaneceu constante, portanto, o ponto anterior não foi arquivado. A Figura 20 apresenta a avaliação de mais um ponto a ser avaliado.

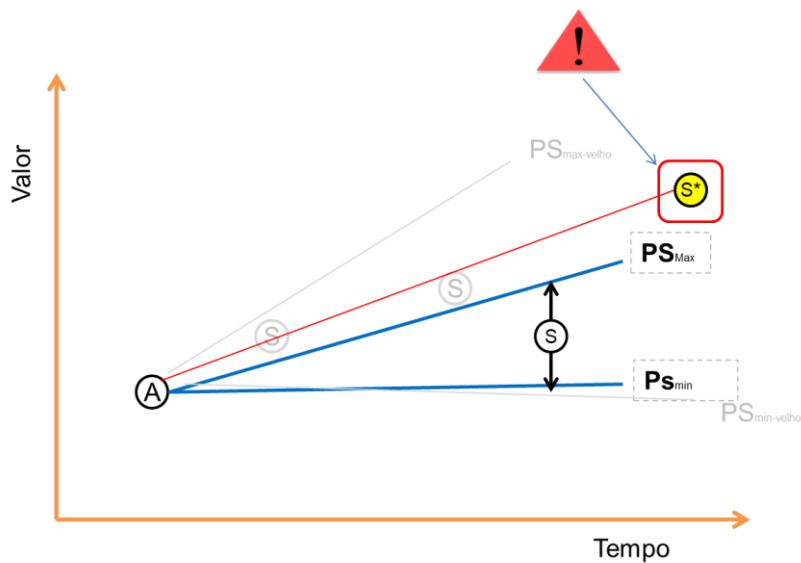
Figura 20 Avaliação das Inclinações Máxima e Mínima do ponto atual.



Fonte: Adaptado Osisoft® (2019)

Na Figura 21, um novo ponto passa a ser avaliado. Uma reta a partir do último ponto armazenado até este novo ponto foi traçada, tendo sua inclinação verificada como fora do intervalo formado pelas das retas  $PS_{Max}$  e  $PS_{min}$  cujas inclinações são respectivamente  $S_{Max}$  e  $S_{min}$ .

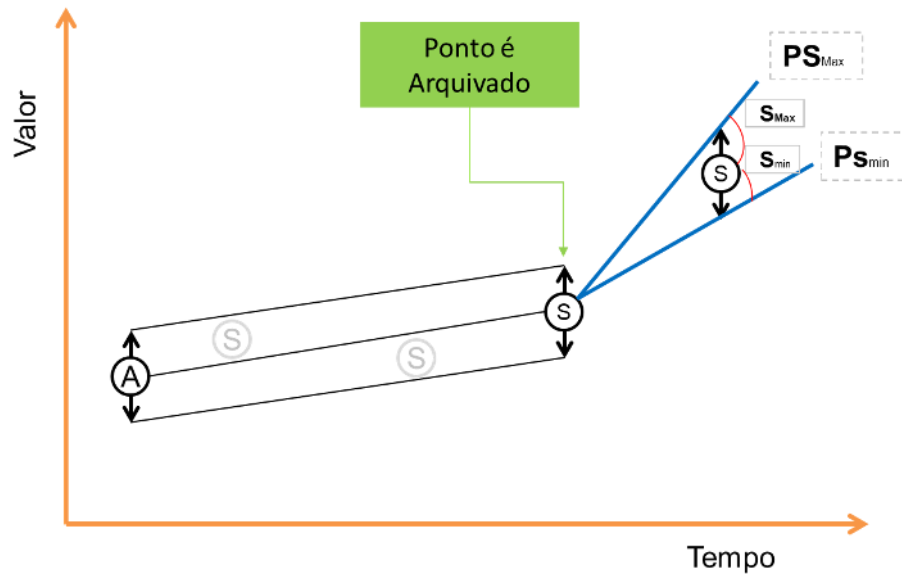
Figura 21 Detecção de ponto com inclinação fora do intervalo aceitável.



Fonte: Adaptado Osisoft® (2019)

Neste cenário, ilustrado na Figura 22, do ponto anterior ao atual é arquivado e do ponto atual toma lugar na definição dos  $PS_{Max}$  e  $PS_{min}$ , que em relação a estes serão traçadas as retas com inclinação  $S_{Max}$  e  $S_{min}$ .

Figura 22 Armazenamento do ponto anterior e definição de novas inclinações máxima e mínima.

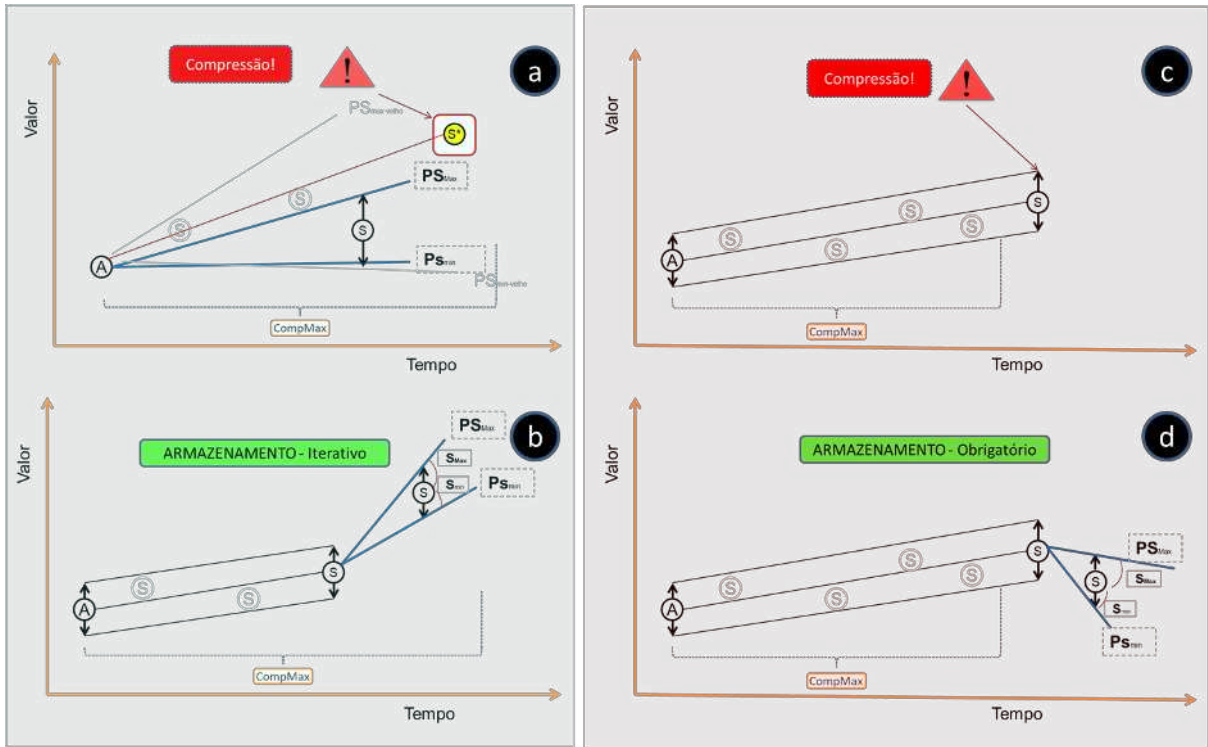


Fonte: Adaptado Osisoft® (2019)

Há duas possibilidades quanto ao armazenamento de pontos durante a execução do algoritmo de compressão de dados, armazenamento iterativo e obrigatório. O armazenamento iterativo se dá pela avaliação das inclinações das retas ponto a ponto, como descrito anteriormente, dependente do valor de  $CompDev$ . Enquanto o armazenamento obrigatório se dá pela avaliação do  $CompMax$ .

A Figura 23 apresenta ambos as possíveis condições de armazenamento, as imagens **a** e **b** correspondem ao armazenamento iterativo e as imagens **c** e **d** o armazenamento obrigatório.

Figura 23 Condições de armazenamento de dados pelo algoritmo de compressão de dados



Fonte: Próprio autor.

O algoritmo de compressão (SDT) tem suas vantagens e desvantagens segundo levantados por Xiadong (2002) avaliou os fatores de desempenho do algoritmo de compressão segundo os indicadores apresentados abaixo:

### 1. Razão de compressão

Razão entre a quantidade de pontos comprimidos e quantidade de pontos originais, demonstrado na Eq (4.1). Esta define o percentual de pontos que foram armazenados com relação aos dados originais. Definida pela fórmula abaixo

$$CR = \frac{N^{\circ} \text{ pontos\_Arquivados}}{N^{\circ} \text{ pontos\_Originais}} \quad (4.1)$$

### 2. Soma do quadrado do Erro de compressão

Considera a soma dos quadrados dos erros, Eq 4.2, justificando-se a potência do erro como uma maneira computar tanto erros positivos quanto erros negativos.

$$CE = \sum (y_i - \bar{y}_i)^2 \quad (4.2)$$

A partir desses Xiaodong (2002), pode elencar os pontos positivos do algoritmo de compressão de dados. Como pontos negativos do *Swinging Door Trending*, têm-se:

1. Quando em presença de ruídos, seu desempenho é reduzido. Este fato justifica a fase de exceção de dados, a qual contorna esse comportamento através de um algoritmo aproximação de segmentos lineares (ELSMeyer, 2006);

3. A depender dos parâmetros de configuração do algoritmo, os resultados podem variar. Neste caso, a taxa de compressão de dados varia de acordo como os parâmetros adotados, chegando ao limite que não deve ser transpassado: os dados passam a sofrer distorção devido a exclusão em excesso de dados;

Como pontos positivos do *Swinging Door Trending*, têm-se:

1. O algoritmo é confiável, partindo-se de parâmetros adequados aos dados a serem comprimidos. Portanto, a depender das configurações do algoritmo de compressão, este pode conferir aos dados resultantes do processo de compressão as características presentes antes do processo.

2. O algoritmo de compressão é capaz de comprimir a massa de dados com taxas de compressão elevadas e manter os erros reduzidos. Neste caso, a presença de erros mínimos aceitáveis é considerada, pois, por natureza os algoritmos de compressão de dados geram erros. Devido ao processo reconstrução consistir na interpolação linear entre os pontos arquivados para a reconstrução dos pontos excluídos, é natural a presença de erros.

#### **4.2. Construção de Indicadores Estatísticos**

Considerados abaixo, cada um dos indicadores é avaliado segundo a relação entre os dados brutos e comprimidos. Estes indicadores mostram-se importantes para a comparação a ser realizada entre os parâmetros de compressão em avaliação.

Visando contribuir no sentido da experimentação e avaliação estatística, o trabalho de Thurnhill, et al (2004), abordaram as análises de Estatística Descritiva, avaliação de não linearidade, no plano da frequência e introduziu o índice de Harris como métricas de avaliação do processo de compressão. Para tanto, dados brutos e comprimidos em um intervalo definido de parâmetros de compressão foram



analisados. Os experimentos foram conduzidos com base no algoritmo *Swinging Door Trending* (SDT), este aplicado industrialmente no sistema de PIMS da Osisoft, PI System (THURNHILL et al, 2004).

#### 4.2.1.Diferença Percentual Entre médias (PDM)

Avaliar a média dos dados originais e a média dos dados reconstruídos determina em que ponto os dados estão se concentrado. Por ser uma medida central, os dados reconstruídos devem ter uma medida central condizente com os dados originais, indicando à medida que os dados reconstruídos preservam a relação com os dados de originais. O percentual entre médias (PDM) torna a comparação de distanciamento mais fácil de ser percebida. (THURNHILL, 2004). A seguir pode-se perceber a sua fórmula:

$$PDM = 100 \frac{média(y) - média(\hat{y})}{\rho_y} \quad (4.3)$$

Quando em análise, quanto maior o PDM, maior o distanciamento entre as médias. Ou seja, maior foi o distanciamento em relação à média entre os dados originais e os reconstruídos.

#### 4.2.2.Razão entre Variância dos Dados Reconstruídos e Variância dos Dados Originais (RVC)

A medida de variabilidade amostral para os dados originais determina o distanciamento dos dados com relação à média amostral, resultando em uma característica da amostra. Assim, avaliar a variância tanto dos dados originais quanto dos dados reconstruídos é importante para análise da variabilidade. A razão entre as variâncias demonstra a diminuição ou aumento da variabilidade nos dados reconstruídos. Para RVC igual a 1, os dados reconstruídos não foram impactados pelo processo de compressão de dados devido a variabilidade dos dados permanecer a mesma, quanto menor o valor de RVC, maior é a variabilidade dos dados reconstruídos em relação dados originais(THURNHILL, 2004).

$$RVC = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2} \quad (4.4)$$

#### 4.2.3. Razão entre Variância dos Erros Reconstruídos e Variância Dados Originais (RVE)

Relaciona a variância dos erros, obtidos pela diferença entre os dados originais e os reconstruídos, com a variância dos dados originais. Este indicador resulta na indicação do quanto a compressão afeta a variabilidade dos erros. Para fins de análise, quanto mais próximo de 1 for o valor do indicador RVE, menos o processo de compressão gera dispersão nos dados, devido a menor dispersão do erro (THURNHILL, 2004). Sua fórmula é como se segue:

$$e_i = y_i - \hat{y}_i \quad (4.5)$$

$$RVE = \frac{e_i}{\sigma_y^2} \quad (4.6)$$

#### 4.2.4. Erro Quadrado Médio (MSE)

Este indicador compara o resultado da compressão com os dados brutos, é uma maneira de verificar o quanto de informação é perdida pela exclusão de dados e reconstrução destes dados para mesmo timestamp. Este é definido pela média do erro quadrado entre o sinal original e o sinal reconstruído:

$$MSE = \frac{\sum_{i=0}^N (y_i - \hat{y}_i)^2}{N} \quad (4.7)$$

Onde  $y_i$  corresponde ao  $i$ -ésimo ponto do sinal original,  $\hat{y}_i$  é o  $i$ -ésimo ponto do sinal reconstruído,  $N$  é o número de pontos do sinal original.

#### 4.2.5. Erro Médio Absoluto (MAE)

Semelhantemente ao MSE, o erro quadrado médio é uma medida do distanciamento entre o sinal bruto e sinal reconstruído. É definido como está apresentado em (4.8)

$$MAE = \frac{\sum_{i=0}^N |y_i - \hat{y}_i|^2}{N} \quad (4.8)$$

Onde  $y_i$  corresponde ao i-ésimo ponto do sinal original,  $\hat{y}_i$  é o i-ésimo ponto do sinal reconstruído, N é o número de pontos do sinal original.

#### 4.2.6. Razão de compressão (CompRatio)

A razão de compressão utilizada representa qual o percentual que os dados comprimidos são em relação aos dados originais. Idealmente, quando menor este número maior a taxa de compressão de dados, mais a quantidade de pontos excluídos.

$$CompRatio = \frac{N \text{ Dados Comprimidos}}{N \text{ Dados Brutos}} \quad (4.8)$$

O CompRatio define o quanto de dados permanecem em relação a quantidade original. Para o processo de compressão aplicado no PI System, esta razão não é um valor de entrada. Pode ser calculada quando se sabe o valor a quantidade pontos está presente no sinal original.

#### 4.2.7. Coeficiente de Correção de Pearson

Este indicador é uma medida estatística que mede a existência e força correlação linear entre dois conjuntos de dados. A existência é verificada pelo cálculo do valor-p e a força pelo cálculo de "r", definido pela razão entre a covariâncias de duas variáveis e o produto de seus desvios padrão, apresentada abaixo:

$$r_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (4.8)$$

A correlação pode assumir valores entre -1 e 1, oscilando entre correção linear positiva e negativa. Segundo Cohen (1988) o valor absoluto de "r" é classificado como pequeno entre 0 e 0,1, médio entre 0,1 e 0,3 e grande para valores acima de 0,5. Para o presente trabalho, os valores para o índice de correlação de Pearson devem ser o mais próximo possível de 1.

### 4.3. Controle Estatístico de Processos e o Monitoramento de Ativos Industriais

As cartas de controle permitem a estimação da variabilidade do processo, gerando informações sobre as variáveis chaves do sistema, importantes parâmetros do processo, assim como também a sua estabilidade com o decorrer da operação.

Cartas de controle para variáveis tem o direcionamento em dois sentidos estatístico: análise da dispersão (variabilidade da variável) e da tendência central (média da variável). Portanto, de maneira a averiguar o comportamento estatístico do processo, faz-se necessário tanto a aplicação quanto a construção serem conduzidas aos pares, combinando uma carta para tendência central e outra para dispersão (GRANT,1964)

Com o objetivo de verificar o impacto do processo de compressão de dados sobre o monitoramento da qualidade do processo. O par de cartas de controle da média e amplitude,  $\bar{X}$  e R é utilizado para este fim. A construção e aplicação das cartas de controle se dá a partir das fórmulas apresentadas na Tabela 3.

Tabela 3 Fórmulas para construção das cartas de controle da média e amplitude.

Carta de controle	LIC	LC	LSC
X-barra	$\bar{\bar{X}} - A_2 \bar{R}$	$\bar{\bar{X}}$	$\bar{\bar{X}} + A_2 \bar{R}$
R	$D_3 \bar{R}$	$\bar{R}$	$D_4 \bar{R}$

Fonte: Montgomery (2012)

As constantes  $A_2$ ,  $D_3$  e  $D_4$  são tabelas e definidas ao considerar a quantidade de observações(n) utilizado em cada amostra utilizada para construção das cartas de controle. As variáveis  $\bar{\bar{X}}$  e  $\bar{R}$  correspondem respectivamente a média das médias e amplitude das amostras. Estas amostras são coletadas do sistema a ser monitorado quando causas especiais estão ausentes, portanto, o estado representado por estas amostras é tomado como referência de qualidade do sistema. O detalhamento e demonstração das fórmulas acima podem ser consultadas em literatura especializada como MONTGOMERY (2012).

Ao considerar o monitoramento de ativos industriais, a definição da variável de processo a ser monitorada é o primeiro passo. Após isso, faz-se necessário coletar as amostras do sistema em um intervalo de tempo no qual se apresenta estável, sem efeitos de causas especiais. Cada amostra é definida por conjunto de  $n$  de subgrupos racionais, medidas retiradas do sistema sob condições semelhantes, porém independentes entre si.

A utilização das cartas é então restrita as configurações estabelecidas na fase de construção das cartas de controle. Em sistemas industriais cada condição de operação possui uma carta de controle estabelecida sobre as condições normais a cada ponto operacional.

## 5. METODOLOGIA

A presente seção objetiva demonstrar os conceitos aplicados para o desenvolvimento dos resultados. A partir desta é possível averiguar as considerações e abordagem adotadas para a análise do tema proposto. Estão presentes demonstração da aplicação dos conceitos sobre compressão de dados, demonstrase o processo de desenvolvimento adotado para o desenvolvimento do algoritmo e processamento dos dados, assim como a sua validação. Os indicadores de desempenho são tratados logo em seguida. São elencados os fatores de impacto na análise de cada um deles. Por fim, as cartas de controle estatístico são abordadas e os detalhes de sua aplicação para avaliação do processo de compressão.

### 5.1. Compressão de dados

Para o desenvolvimento da análise de compressão de dados, faz-se necessário a avaliação de estatísticas dos dados antes e após o processo de compressão de dados e, então, pode-se determinar os indicadores estatísticos e avaliar a configuração mais adequada. Para a determinação dos indicadores estatísticos de avaliação, faz-se necessário o desenvolvimento de uma biblioteca de funções dedicadas. A linguagem de produção escolhida foi Python®.

#### 5.1.1. Fluxo de Dados na Fase de Compressão

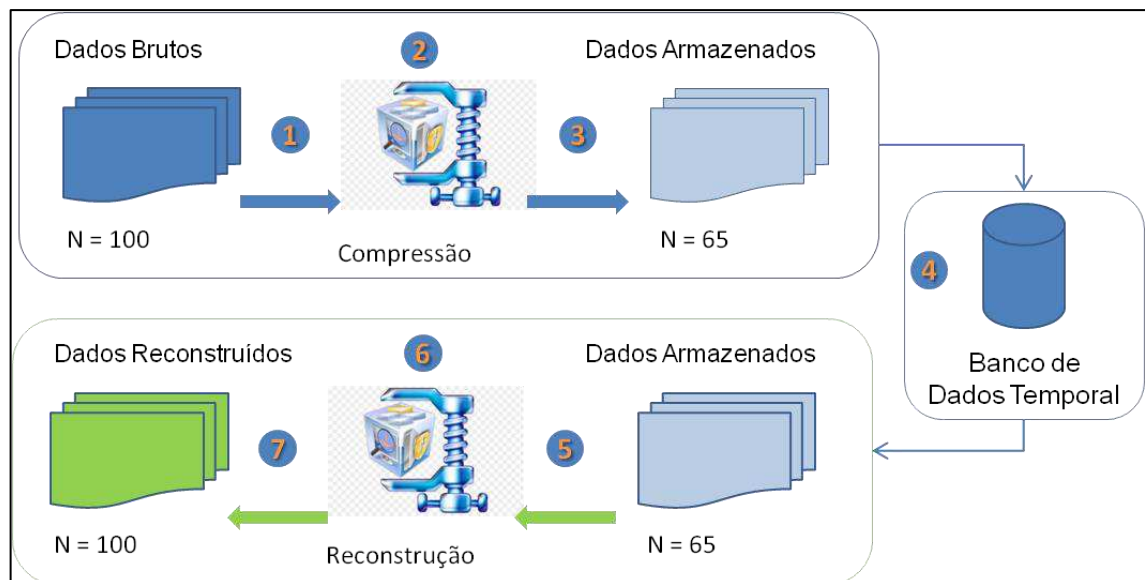
O processo de coleta e armazenamento de dados industriais possuem uma determinada quantidade de etapas em sequencias, e cada uma com o fim específico, em conjunto contribuindo para a operação ocorra satisfatoriamente em quesitos como segurança, facilidade de manutenção e confiabilidade para transmissão dos dados.

A Figura 24 demonstra uma estrutura simplificada dos principais processos presentes no arquivamento e posterior consulta de dados. Como entrada têm-se os dados brutos (1) que passam pelo processo de compressão (2). No presente trabalho é composto pelo algoritmo *Swinging Door Trending*, no qual parte dos dados é apagada, classificado como compressão com perdas de dados. O resultado é denominado de dados arquivados (3), seguindo para o armazenamento no banco de dados temporal dedicado (4).

Caso a necessidade de construir os pontos que foram excluídos, o processo inverso ao da compressão é realizado, a reconstrução (6). O método matemático aplicado nessa reconstrução é interpolação linear, capaz de reconstruir os pontos excluídos para o *timestemp* correspondente. A partir da comparação entre dados brutos e armazenados (3), os *timestemps* são definidos e aplicado ao algoritmo de reconstrução por interpolação. O resultado corresponde aos dados reconstruídos (7) por interpolação, correspondentes aos pares excluídos.

Este mesmo processo de reconstrução também é aplicado para a consulta dos valores em tempos que não existem os valores originais, correspondendo a dados interpolados, não existentes nos dados brutos.

Figura 24 Entidades envolvidas no processo de compressão e consulta de dados armazenados.



Fonte: Próprio Autor

Portanto, as grandezas dados brutos, arquivados e reconstruídos desempenham papel fundamental para avaliação do desempenho da compressão de dados. Aplicando-se os indicadores estatísticos, pode-se avaliar a influência e os impactos desse processo sobre as informações contidas nos dados.

### 5.1.2. Fluxograma do Algoritmo de Compressão de Dados

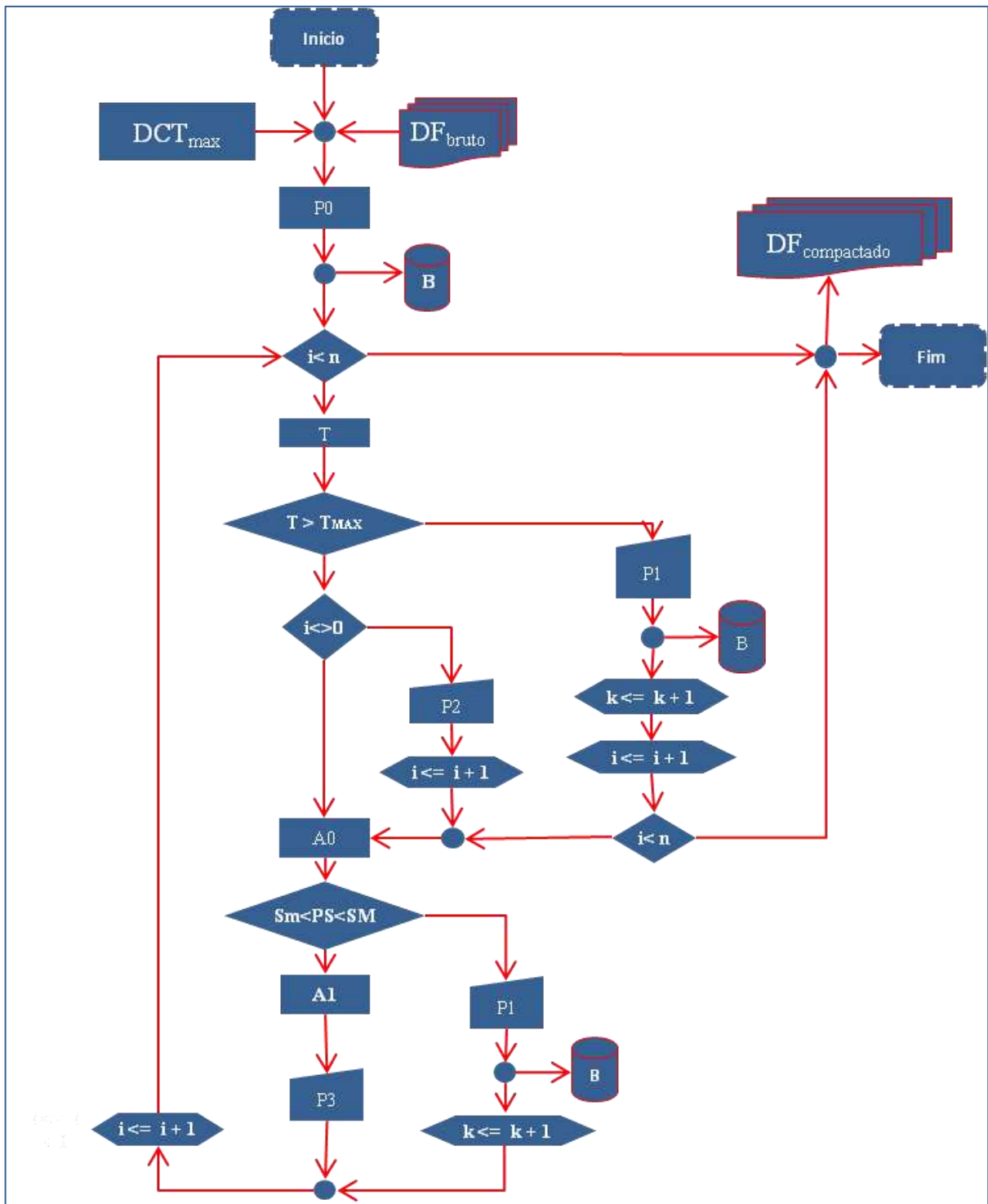
A partir da análise do artigo de Bristol (1990) é possível construir o fluxograma do algoritmo de compressão *Swinging Door Trending*. Em seguida, na Figura 25, o

fluxograma é apresentado. Os passos do algoritmo podem ser descritos a se seguinte sequência:

- 1- Entrada dos dados Brutos;
- 2- Definição de ambos os parâmetros Desvio de Compressão (DC) e Tempo máximo de processamento ( $T_{max}$ )
- 3- Armazenamento do primeiro ponto;
- 4- Interação em todos os pontos até que o espaço de tempo entre o tempo atual e o tempo anterior do último ponto armazenado seja maior que o valor de  $T_{max}$ , ocasionando um armazenamento forçado seguida de atualização dos pontos armazenado, último avaliado e atual;
- 5- Caso o espaço de tempo entre o tempo atual e o tempo anterior do último ponto armazenado seja menor que  $T_{max}$ , ocorre avaliação das inclinações dos pontos subsequentes;
- 6- Caso a inclinação do ponto atual esteja entre as  $S_m$  e  $S_M$ , estas sofreram atualização caso a Inclinação do  $PS_m$  e  $PS_M$  menores que as anteriores, caso contrário o último ponto avaliado é armazenado e todos os pontos sofrem atualização novamente;
- 7- O ponto será armazenado caso a  $PS$  seja maior que  $S_m$  e  $S_M$ , neste caso, o último ponto avaliado é armazenado e todos os pontos sofrem atualização novamente.



Figura 25 Fluxograma desenvolvido para implmentação do SDT.



Fonte: Próprio Autor

### 5.1.3. Validação do algoritmo desenvolvido

Considerando que o presente trabalho de visa avaliar o processo de compressão de dados aplicado industrialmente no PI system, é de importância máxima a validação do algoritmo de compressão aplicado para a avaliação aqui proposta. Diretrizes para esta validação foram tomadas de Alsmeyer (2006).

A validação consiste em alguns passos, listados a seguir:

1. Seleção de um conjunto de dados a partir de dados de planta de processamento de minério de bauxita;
2. Para cada tag do tipo contínua selecionada, arbitrar os parâmetros de compressão e exceção para três condições:
  - a. Sem compressão;
  - b. Somente com Exceção;
  - c. Semente com Compressão;
  - d. Com exceção e compressão.
3. Coletar dados contido nas tags para cada caso especificado no ponto 2 acima, em espaço fixo de tempo;
4. Processar cada caso de acordo com o especificado em 2 acima, em comparação com os respectivos dados coletados para o referido caso;
5. Avaliar ponto a ponto e o valor total dos dados resultantes do processamento e os dados do referido caso;

O critério principal para a validação dos algoritmos de compressão e exceção consiste na comparação de ponto a ponto entre os dados resultantes do processamento, pelo algoritmo desenvolvido para o presente trabalho, com os dados colhidos como referência.

Para que esta comparação seja fidedigna, os parâmetros de cada algoritmo de exceção e compressão do PI system são colhidos e aplicados em seus casos respectivos para comparação de resultados sem nenhum tipo de conversão. Essa versatilidade é proposital para a redução de qualquer interferência do usuário sobre os resultados e para redução de esforço na aplicação de análises aqui propostas.

## 5.2. Análise de indicadores de compressão de dados

Os indicadores aplicados no presente trabalho possuem fundamentação estatística, como descrita na fundamentação teórica anterior. Os seguintes indicadores são utilizados:

1. Diferença Percentual Entre médias (PDM)
2. Razão entre Variância dos Dados Reconstruídos e Variância dos Dados Originais (RVC)
3. Razão entre Variância dos Erros Reconstruídos e Variância Dados Originais (RVE)
4. Taxa de compressão de dados

É intencional a aplicação de diversos indicadores. Como o processo tanto de exceção como de compressão são responsáveis por excluir dados redundantes antes de serem arquivados, torna-se vantajoso ampliar as opções de análises visando elencar a aspectos estatísticos presentes nos dados.

A estabilidade e amplitude alcançadas pelos valores das variáveis podem ser diversas. Portanto, considerar estas características nas análises dos indicadores pode revelar uma maneira de classificar os tipos de variáveis em grupo.

A estacionariedade é um fator determinante no desenvolvimento de análises temporais. A ausência desta característica pode impactar consideravelmente o comportamento resultados. Faz-se necessário a investigação deste critério na análise da compressão através de cada indicador apresentado.

Os indicadores serão aplicados visando:

1. Determinar quais deles tem maior sensibilidade em detectar uma maior alteração nos dados;
2. Verificar para cada configuração de compressão e exceção se existe a correlação entre eles;
3. Avaliar a existência de pontos de inflexão entre os indicadores, quando avaliados em paralelo;

O procedimento adotado para análise de sensibilidade da compressão em relação a cada indicador será feita de três formas, para cada *tag*:

1. Os valores dos parâmetros de exceção serão varridos do estado sem exceção até o estado de exceção mediana, mantendo sem compressão de dados;
2. Os valores dos parâmetros de compressão serão varridos do estado sem compressão até o estado de compressão mediana, mantendo com exceção de dados;
3. Ambos os parâmetros de exceção e compressão são varridos de um estado mínimo até o estado de mediano.

A avaliação estatística seguirá o desenvolver de cada indicador em cada avaliação fases 4, 5 e 6 do processo de análise de desempenho estatístico para cada cenário.

### **5.3. Cartas de controle a partir de dados comprimidos**

A aplicação de cartas de controle na avaliação do impacto da compressão de dados através do algoritmo de SDT se dará em duas condições

1. Construção das cartas de controle a partir de dados comprimidos historiados;
2. A partir dos dados não comprimidos, construir a carta de controle, passando para avaliar o monitoramento e possíveis distorções devido ao processo de compressão.

Para as cartas em análise, o estudo de caso apresentado adiante será responsável por gerar os dados através de simulações numéricas. Este dados será considerado tanto para a ajuste da performance do processo de compressão de dados para pares definidos de parâmetros, assim como também para verificar a viabilidade de sua utilização quando os dados estão comprimido de forma adequada ou inadequada.

## 6. RESULTADOS E DISCUSSÕES

Nesta seção são apresentados os resultados das análises executadas de acordo com a metodologia adotada. Espera-se que a construção das cartas de controle a partir de dados comprimidos possa não demonstrar alguma alteração até certo ponto da compressão de dados. A investigação desta hipótese requer os dados brutos sejam comprimidos com diferentes parâmetros de exceção, *ExcMax* e *ExcDev*, e diferentes parâmetros compressão, *CompMax* e *CompDev*. Para isto a avaliação dos dados brutos deve ser feita a partir de uma ampla variedade de configurações de exceção e compressão de dados.

No entanto, este processamento não é factível industrialmente, por restrições financeiras, pois o custo de manutenção de um sistema PIMS está intimamente ligado com a quantidade de hardware e modalidade de licenciamento do sistema instalado.

Por este motivo, foi desenvolvido um algoritmo dedicado a reprodução fidedigna do algoritmo aplicado industrialmente, de maneira que possibilite a construção de um ambiente de experimentação numérica, aqui chamado de simulador. Este tem por finalidade ser um ambiente de simulação do processo de compressão de dados. Com este simulador é possível avaliar os resultados obtidos para cada conjunto de parâmetros dos algoritmos de exceção e compressão. Este simulador nada mais é que uma biblioteca Python que possibilita a criação de análises numéricas e estatísticas em ambiente de análise de dados.

Neste, há a possibilidade de obter os mesmos resultados que os obtidos industrialmente, além de não requerer a mesma quantidade de hardware ou licença. Porém é imprescindível garantir que ambos o algoritmo de exceção e o SDT presentes no simulador sejam idênticos em resultados ao PI System, ou seja, que gerem os mesmos pontos após a execução dos algoritmos de compressão e exceção.

Garantida esta reprodução fidedigna dos algoritmos, parte-se a verificação da existência de parâmetros ótimos de compressão de dados. A abordagem exploratória, também chamada de análise de sensibilidade, foi definida como o método para alcançar este objetivo. Os instrumentos desta exploração são os indicadores estatísticos já apresentados, aplicados para avaliação do comportamento das características estatísticas dos dados tais como média, variância e erro.

Por fim, são aplicados os conceitos de cartas de controle. Os limites de controle são definidos a partir dos dados brutos de processo. A estes dados brutos são então aplicados os algoritmos de exceção e compressão de dados e seus resultados aplicados a cartas de controle, com o objetivo de verificar a hipótese que o processo de compressão pode vir a inviabilizar análises estatísticas, quando esta compressão não é configurada adequadamente.

Durante a avaliação dos dados, cada configuração dos algoritmos gera resultados, os quais são posteriormente analisados. Neste trabalho, foram adotadas as orientações da Osisoft ao que diz respeito à configuração dos parâmetros de configuração do algoritmo de exceção, correlacionando-os aos parâmetros do algoritmo de compressão SDT. Esta adequação tem por objetivo diminuir a complexidade inicial associada ao problema.

### **6.1. Validação do Algoritmo de exceção e compressão**

Como forma de avaliar as considerações em torno do algoritmo de compressão proposto, um conjunto de dados foi escolhido para ser armazenados no PI System, Osisoft. Os dados correspondem ao par de valores de corrente elétrica de motores de bomba e o *timestamp* correspondente.

Estes são coletados em formato de *tag* (PI Point) com os algoritmos de exceção e compressão desligados. A partir desta *tag*, três cópias foram criadas para análise dos algoritmos desenvolvidos. Ao todo, quatro *tags* são utilizadas para validação dos algoritmos. As características de cada *tag* correspondem a lista a seguir:

1. Dados brutos, com exceção e compressão desativados;
2. Dados para validação do processo de exceção, *ExcDev* e *ExcMax* especificados;
3. Dados para validação do processo de compressão, *CompDev* e *CompMax* especificados;
4. Dados para validação dos processos de exceção e compressão combinados, *ExcDev*, *ExcMax*, *CompDev* e *CompMax* especificados.

A procedimento de validação do algoritmo de compressão é apresentado a seguir.

### 6.1.1. Validação do Algoritmo de Exceção de dados

Em posse dos valores gerados em 4 dias de coleta de dados, valores das *tag* com as quatro características acima foram armazenados em formatos *csv* específico para processamento. Os dados brutos e os dados processados pelo algoritmo do *PI System*, passam com os valores de *ExcDev* e *ExcMax*. Os dados brutos têm sua origem na coleta de valores de correntes de motores de bombas centrífuga.

É perceptível que os dados estão dispostos em degrau, considerando o período de coleta dos dados, é possível inferir que a *tag* coleta dados em situações diferentes do processo. A Tabela 4 Informações básicas da *tag* coletada e configuração da exceção de dados. contém informações sobre a *tag* de coleta, tais como quantidade de pontos e parâmetros de configuração do algoritmo de exceção, estas utilizadas na validação do algoritmo de exceção de dados.

Tabela 4 Informações básicas da *tag* coletada e configuração da exceção de dados.

<b>Nº Ponto Brutos</b>	<b>7648</b>	-
<b>Nº Pontos Exc-Osisoft</b>	398	-
<b>Nº Pontos Exc-Simulador</b>	397	-
<b>ExcDev</b>	3	Unidade Eng.
<b>ExcMax</b>	1800	Segundos

Fonte: Próprio Autor

Para esta quantidade de dados, a Figura 26 não facilita a visualização dos pontos gerados como resultado. Ao ampliar os dados numa janela de tempo menor, a visualização torna-se possível nas Figura 27 e Figura 28, cujas informações são apresentadas na

Tabela 5. É possível visualizar os resultados do processo de exceção a partir da que demonstra os dados brutos, assim como os dados de referência e processados pelo algoritmo de exceção padrão (*PI System*) e o do presente trabalho.

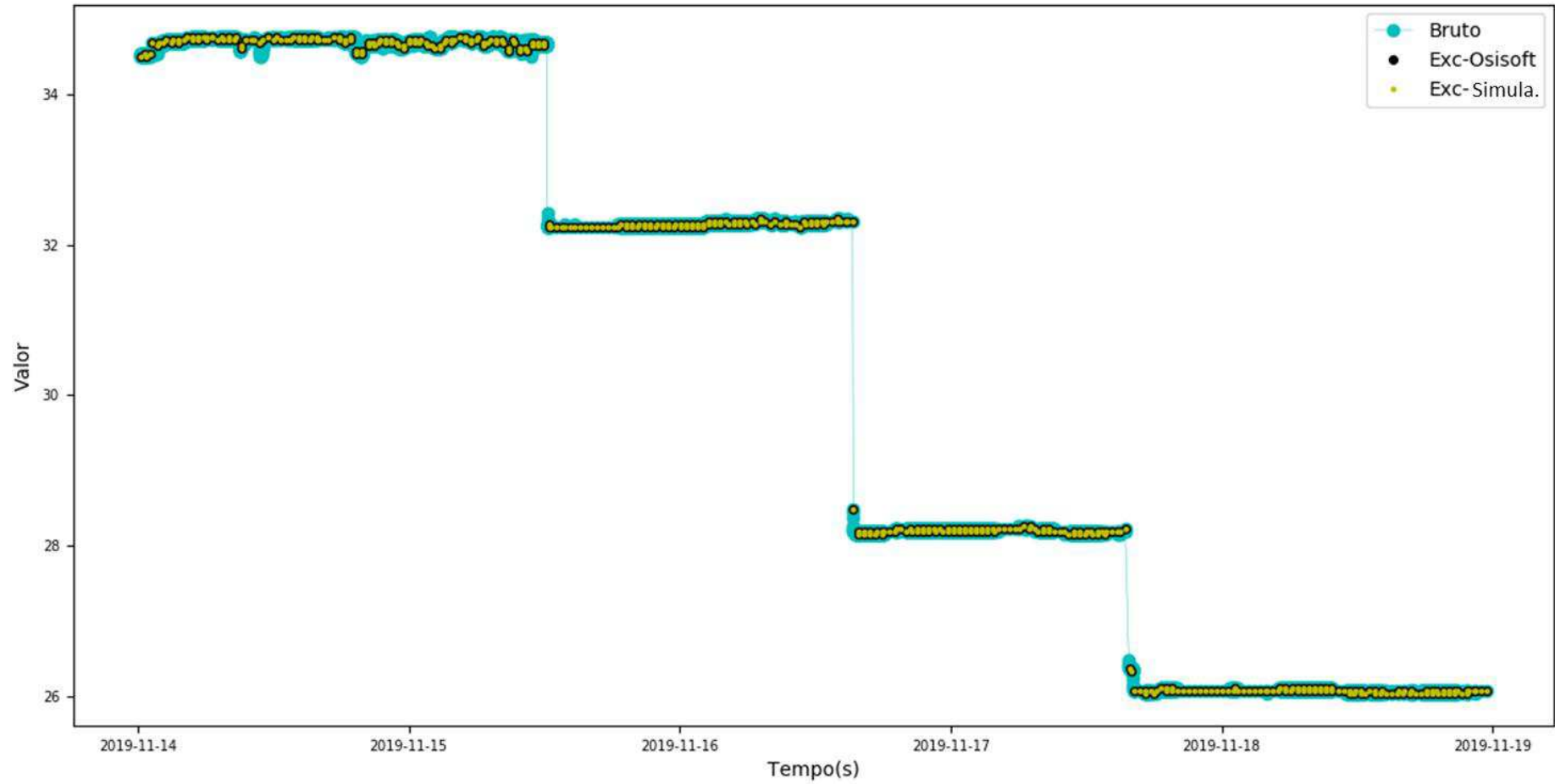
Tabela 5 Comparativo das informações básicas entre os resultados para validação do algoritmo de exceção em janela de dados.

<b>Nº Pontos Brutos</b>	<b>1517</b>	-
<b>Nº Pontos Exc-Osisoft</b>	88	-
<b>Nº Pontos Exc-Simulador</b>	88	-
<b>ExcDev</b>	3	Unidade Eng.
<b>ExcMax</b>	1800	Segundos

Fonte: Próprio Autor

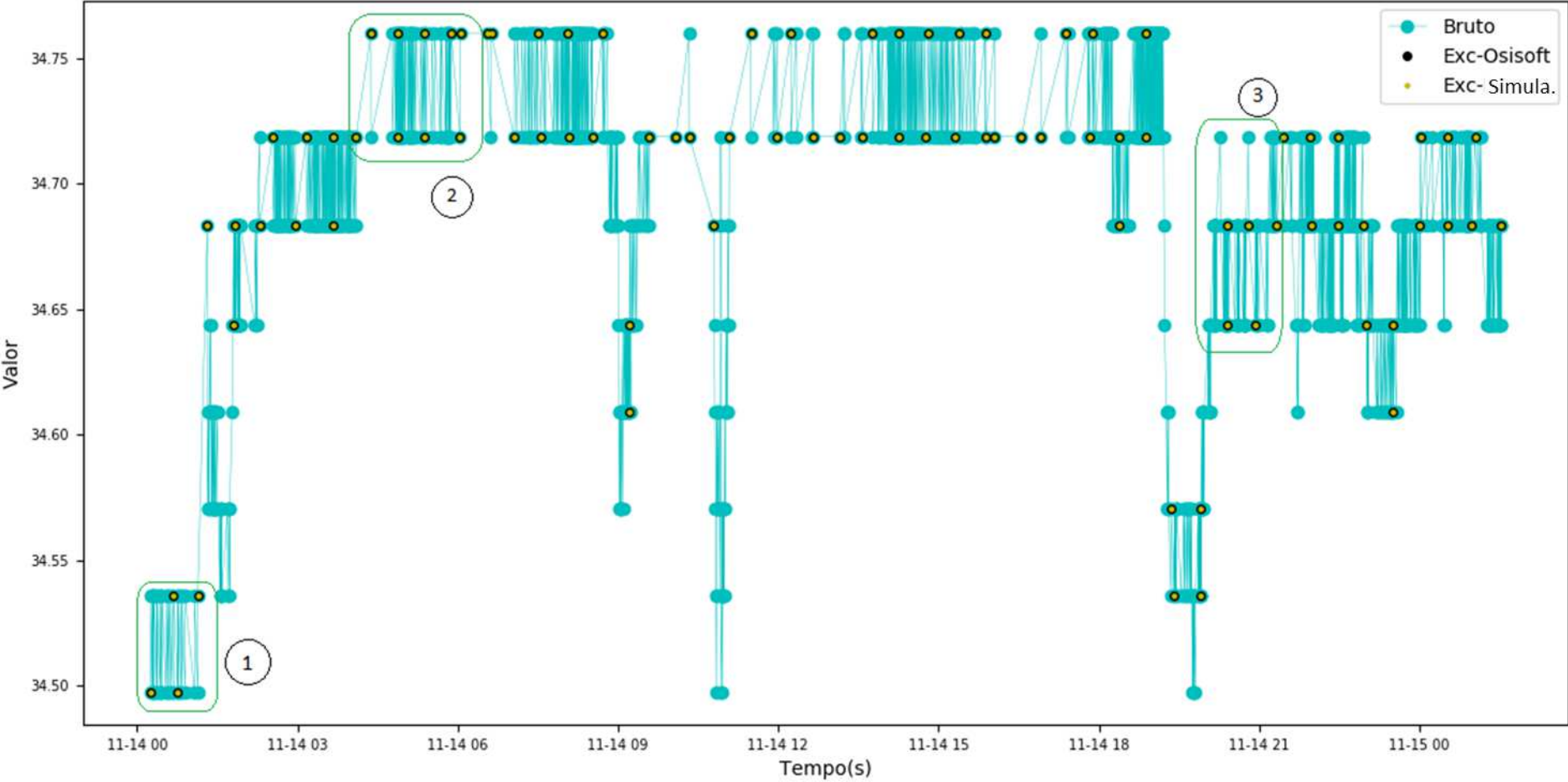


Figura 26 Dados brutos e Referência para avaliação da exceção.



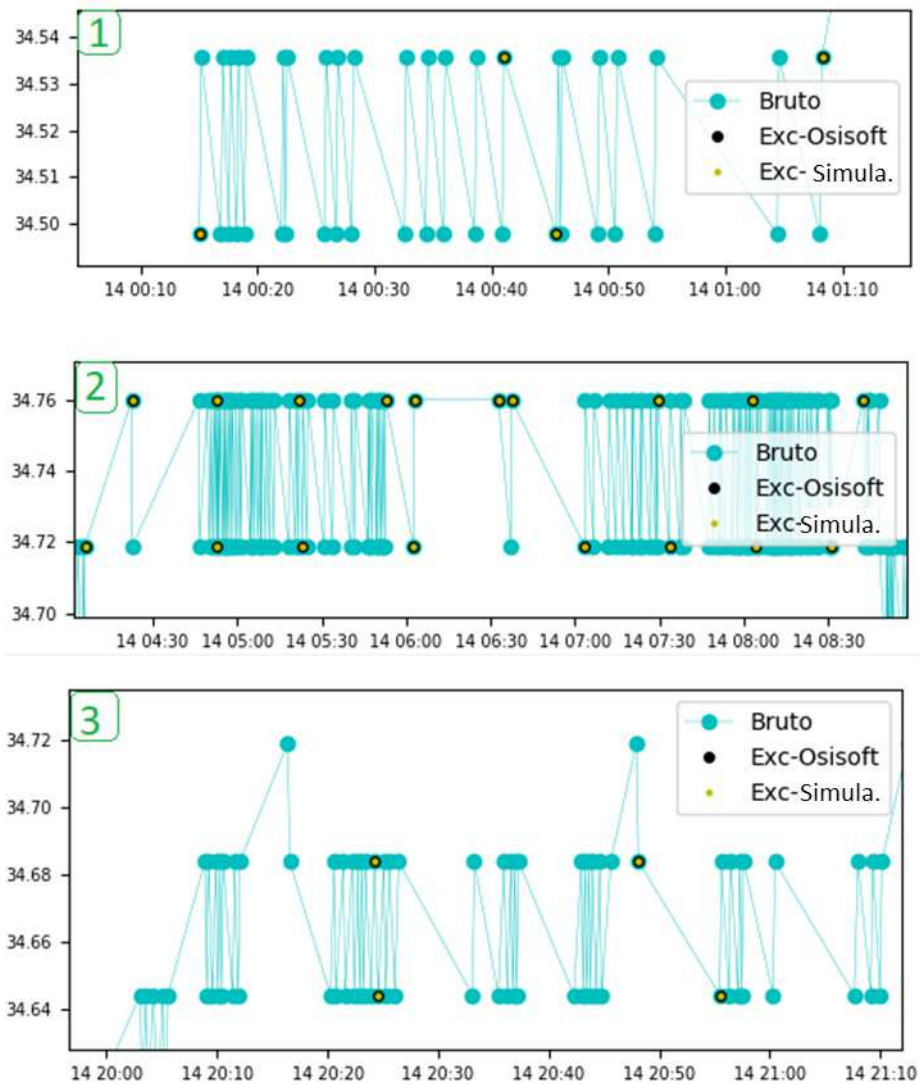
Fonte: Próprio Autor

Figura 27 Janela de dados brutos e Referência para avaliação da exceção.



Fonte: Próprio Autor

Figura 28 Ampliação sobre segmentos de dados aplicados à validação da exceção.



Fonte: Próprio Autor

A partir da Tabela 6, é possível verificar igualdade nos resultados produzidos por ambos os algoritmos de compressão de dados. Os indicadores adotados para esta avaliação numérica são os mesmos a serem aplicados a diante, na avaliação estatística da compressão de dados.

Portanto, considera-se validado o algoritmo de exceção de dados, tendo como comparativo os dados referentes obtidos do processamento do algoritmo de exceção aplicado no *PI System* da Osisoft®.

Tabela 6 Comparação numérica entre os algoritmos de exceção de referência (PI System) e o desenvolvido (Simulador).

Indicadores Estatísticos	Exceção (PI System)	Exceção (Simulador)
PDM	-7.972690273142183	-7.972690273142183
RVC	1.3167542891622104	1.3167542891622104
RVE	0.001483315112588503	0.001483315112588503
MSE	15.132031655609136	15.132031655609136
MAE	0.02639669188295052	0.02639669188295052
Ind. Correlação	0.7964838500855029	0.7964838500855029

Fonte: Próprio Autor

### 6.1.2. Validação do Algoritmo de Compressão

O procedimento para a validação do algoritmo de compressão aplicado no presente trabalho é muito semelhante ao aplicado à validação do algoritmo de exceção. Os dados brutos são os mesmos, alterando-se somente os dados considerados como referência, ou seja, dados armazenados quando somente o processo de compressão no *PI System* está ativado.

Aplicando a compressão nesta mesma janela de dados, os valores obtidos de compressão corresponderam os obtidos pelo processo de compressão configurando com os mesmos parâmetros. Visualmente e numericamente a partir dos resultados apresentados nas Figura 29, Figura 30, Figura 31 e Tabela 8.

Segundo os resultados da Tabela 7 abaixo, o algoritmo de compressão de dados aplicado no presente trabalho foi capaz de arquivar os pontos correspondentes aos pontos gerados pelo algoritmo padrão, aplicado no *PI System*. No entanto, faz-se necessário verificar se estes pontos são realmente os mesmo um a um. Para isso, foram aplicados os indicadores estatísticos apresentados como formas de análise de desempenho de compressão de dados.

Por inferência estatística, a equivalência dos conjuntos de dados é refletida na equivalência entre cada um dos indicadores estatísticos apresentados, tendo em vista que são fundamentalmente avaliação da dispersão e variabilidade dos nos dados. Então é justificada a aplicação dos algoritmos em todos os dados brutos disponíveis.

A avaliação destes indicadores foi realizada após a compressão de dados ser aplicada integralmente a massa de dados disponível da *tag*, ou seja, sem a restrição da janela de dados anterior. A Figura 31 demonstra os resultados visualmente e a Tabela 7 os resultados numéricos da aplicação da compressão de dados sobre todos os dados.

Tabela 7 Comparativo de resultados das Informações básicas para validação do algoritmo de compressão na massa total de dados.

	Valor Numérico	Unidades
Nº Pontos Brutos da Janela	1517	-
CompDev	3	Unidade Eng.
CompMax	1800	Segundos
	Osisoft	Simulador
Nº Pontos	1193	1193

Fonte: Próprio Autor

Os resultados da aplicação dos indicadores estatísticos sobre todos os dados presentes da *tag* de estudo, é apresentada a seguir na Tabela 8. É verificada a correspondência entre os resultados obtidos comparando-se os algoritmos padrão Osisoft® e o desenvolvido no presente trabalho. Esse resultado implica na equivalência entre os algoritmos.

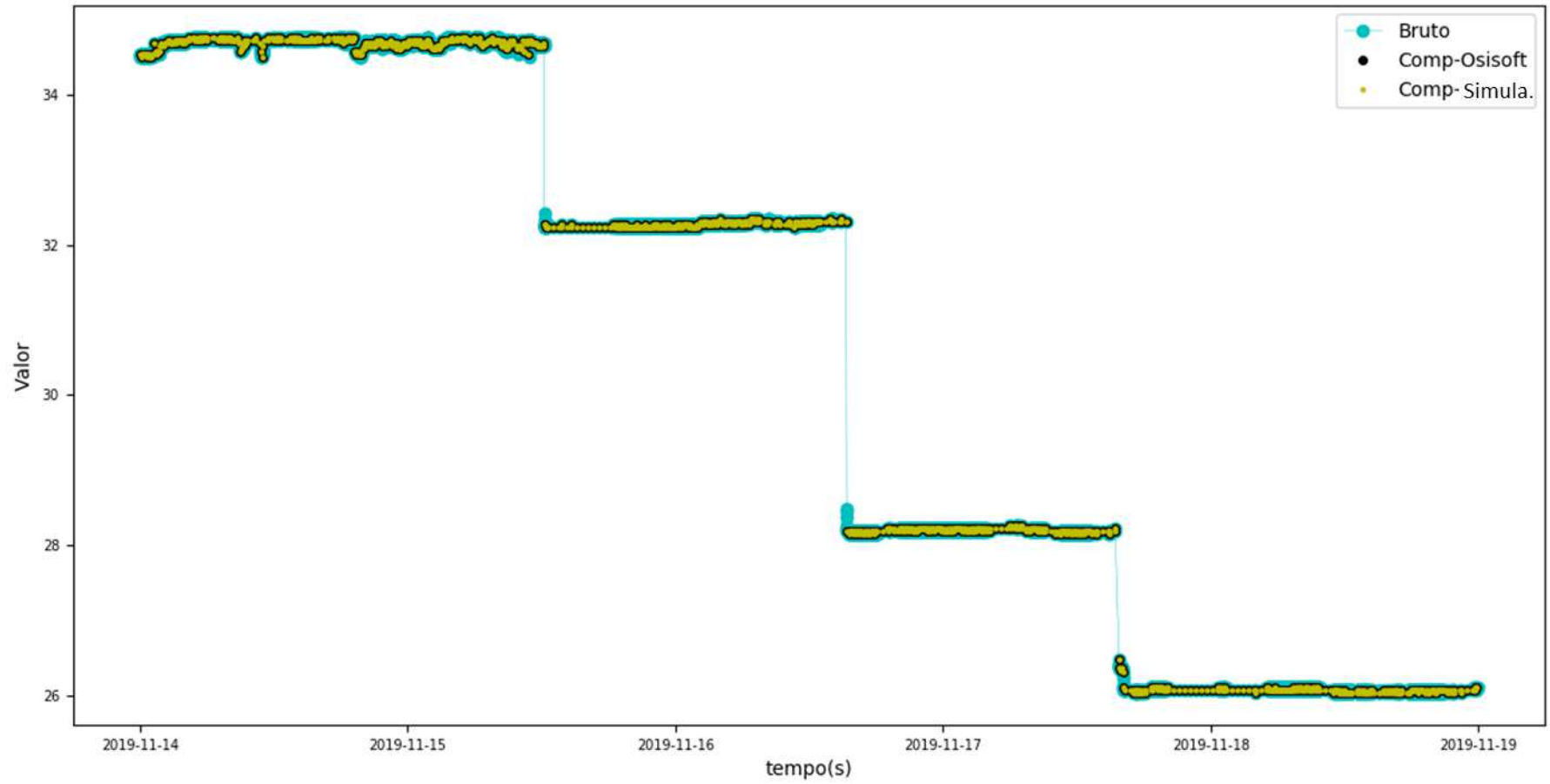
Esse resultado é primordial para a análise estatística da compressão de dados, pois a partir deste algoritmo, construídos em ambiente dedicado a experimentação numérica, será possível explorar dados de diferentes naturezas e definir estratégia para a construção de banco de dados consiste, essa consistência averiguada estatisticamente entre dados brutos e arquivados.

Tabela 8 Indicadores estatísticos aplicados na validação do algoritmo de compressão de dados.

<b>Indicadores Estatísticos</b>	<b>Compressão (PI System)</b>	<b>Compressão (Simulador)</b>
PDM	0.01677255094820765	0.01677255094820765
RVC	1.001740190301243	1.001740190301243
RVE	0.0017796484158100446	0.0017796484158100446
MSE	0.0025403891049495795	0.0025403891049495795
MAE	0.02062448724467673	0.02062448724467673
Ind. Correlação	0.9998925258853149	0.9998925258853149

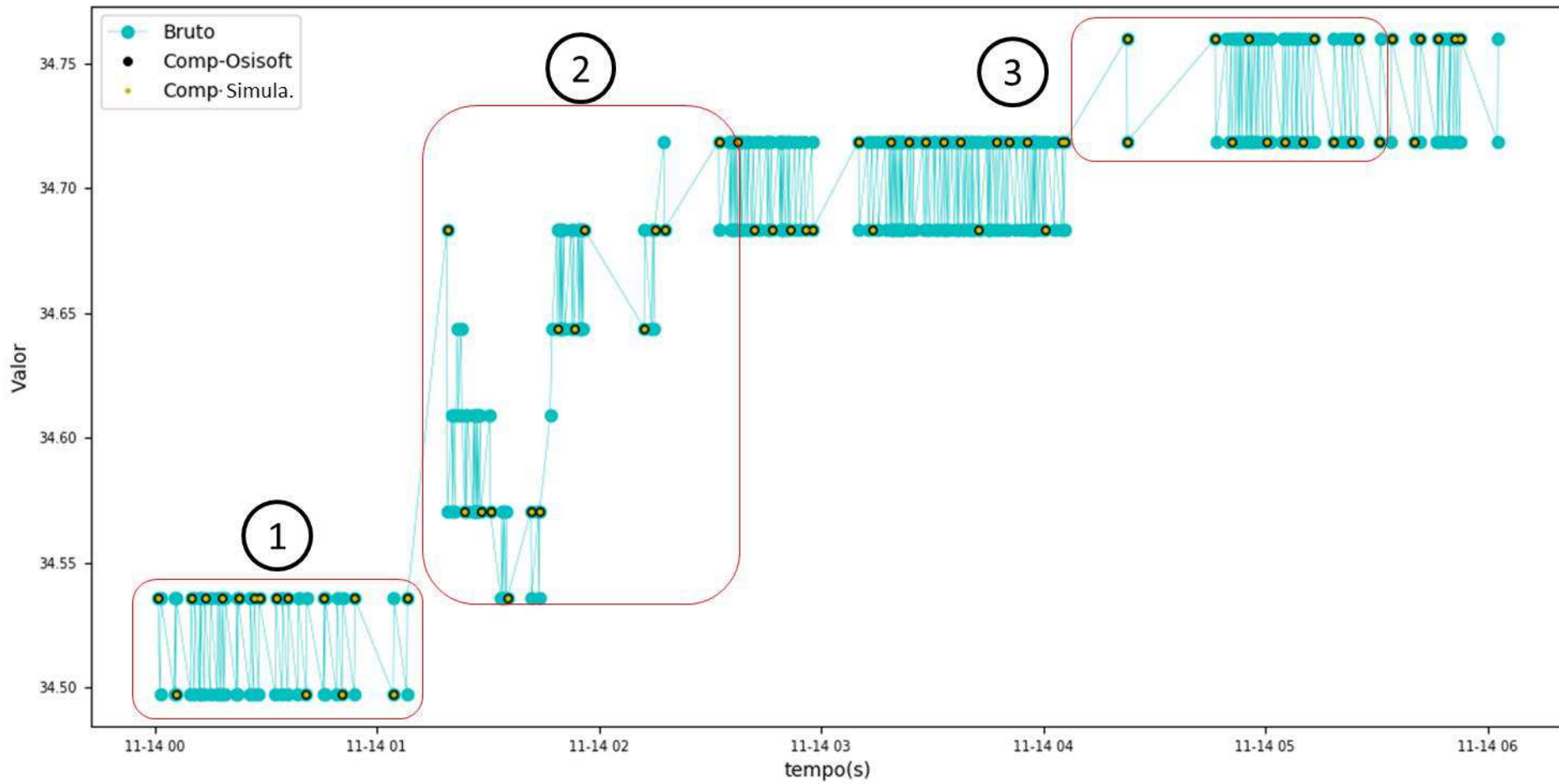
Fonte: Próprio Autor

Figura 29 Dados brutos e Referência para avaliação da compressão.



Fonte: Próprio Autor

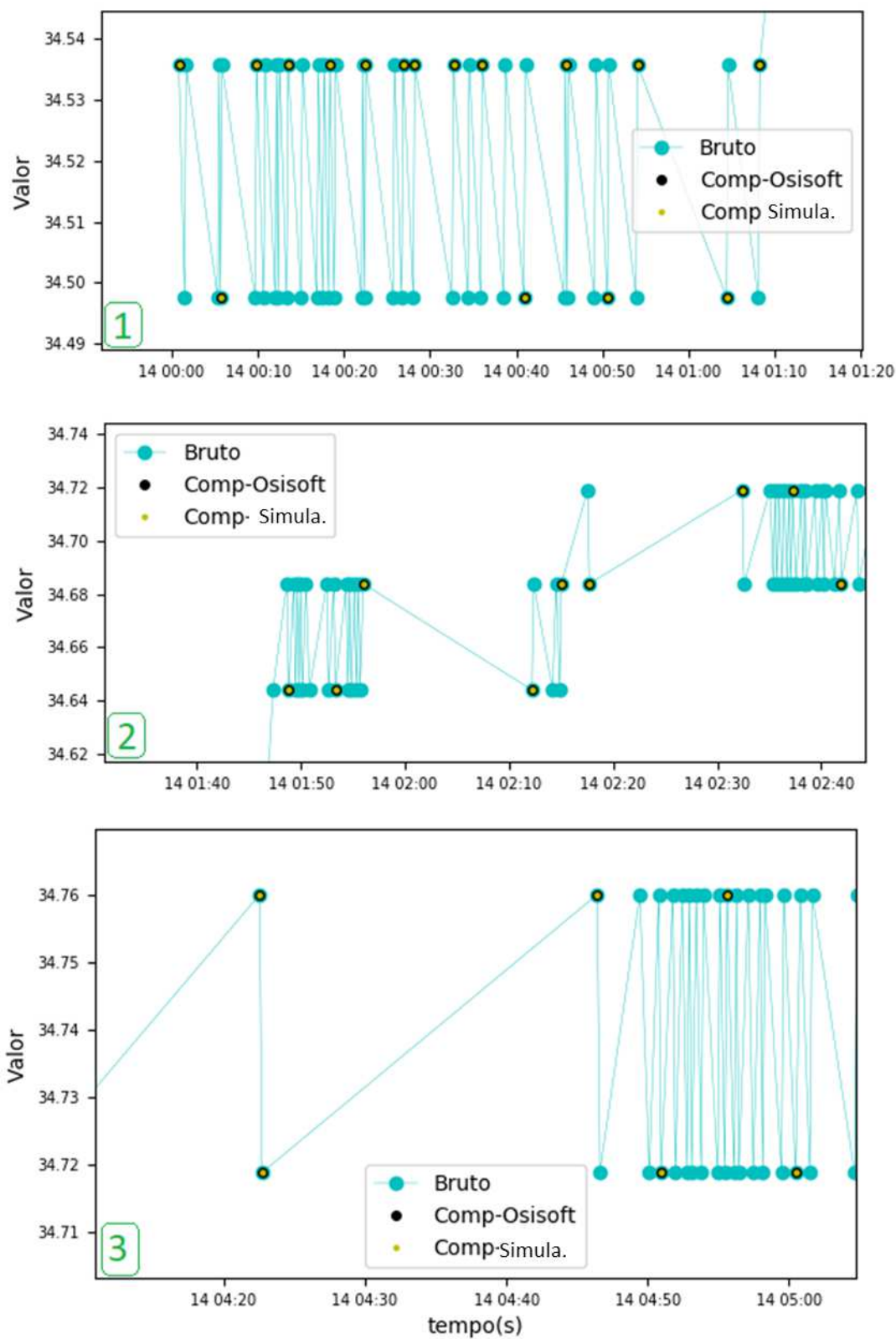
Figura 30 Dados brutos e Referência para avaliação da compressão.



Fonte: Próprio Autor



Figura 31 Zoom aplicado aos dados usado na validação do algoritmo de compressão.



Fonte: Próprio Autor

## 6.2. Erros gerados pelos algoritmos de exceção e compressão de dados

A avaliação e verificação dos erros associados ao processo de exceção e compressão são aqui demonstradas. Os parâmetros para a demonstração dos erros são apresentados assim como os resultados visuais da exclusão dos pontos por ambos os algoritmos.

Os erros foram definidos ponto a ponto para cada valor recuperado a partir dos dados de reconstrução de dados. Essa reconstrução de dados foi realizada por interpolação linear, como é realizada pelo *PI System*, assim como a comparação com dados gerados por estes. E comparação com os algoritmos padrão do PI. O conjunto de dados utilizado para estas análises corresponde ao mesmo conjunto aplicado para a validação de ambos os algoritmos de compressão e exceção de dados.

### 6.2.1. Erros originados pelo processo de exceção

Os erros associados ao processo de exceção puro foram obtidos a partir do erro absoluto da diferença entre o valor bruto e o valor recuperado na janela escolhida para análise, originada da tag de validação. Os erros são obtidos para os algoritmos da Osisoft e do presente trabalho.

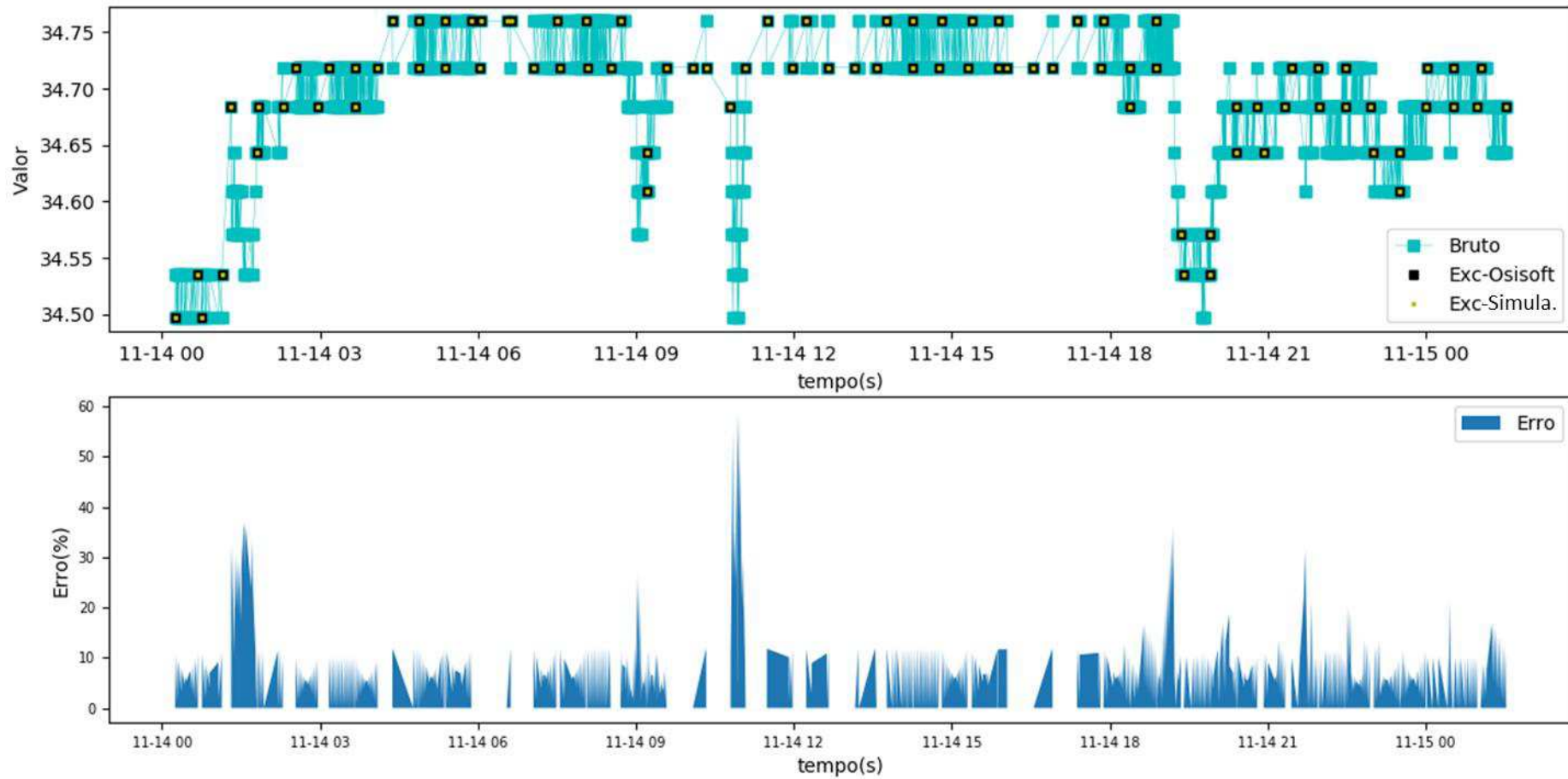
Tabela 9 Descrição da configuração do algoritmo de exceção e erros obtidos pelo processo de exceção de dados.

	Valor Numérico	Unidades
<b>Nº Pontos Brutos</b>	1517	-
<b>ExcDev</b>	3	Unidade Eng.
<b>ExcMax</b>	1800	Segundos
	Osisoft	Simulador
<b>Nº Pontos</b>	88	88
<b>Soma dos Erros Absolutos</b>	0.02639669188295052	0.02639669188295052
<b>Soma dos Esso Quadrados</b>	0.001483315112588503	0.001483315112588503

Fonte: Próprio Autor

Na Figura 32 são apresentados dois gráficos. O primeiro demonstra os dados brutos, de referência e resultados obtidos, tanto pelo algoritmo exceção nativo do PI System quanto o desenvolvido. O segundo gráfico corresponde aos erros associados a cada ponto excluído. No gráfico dos erros, é possível notar a presença de um pico sobressalente. Devido à exclusão de um ponto discrepante aos pontos próximos. Portanto, o erro elevado nesta região é produto da interpolação dos pontos não excluídos, ou seja, a partir dos pontos arquivados no processo de exceção.

Figura 32 Erros gerados pelo processo de exceção.



Fonte: Próprio Autor

### 6.2.2. Erros originados pelo processo de compressão

Pela aplicação do SDT em dados brutos é fácil perceber que muitos pontos são excluídos mesmo considerando o grande distanciamento dos pontos anteriores não excluídos. A Tabela 10 evidencia a configuração do algoritmo SDT, ou seja, os parâmetros de *CompDev* e *CompMax*.

Os parâmetros de compressão definidos são listados na Tabela 10 a seguir. Em comparação ao caso anterior, onde se avaliou unicamente o algoritmo de exceção de dados, o presente caso demonstra que o algoritmo SDT pode vir a armazenar mais pontos, a depender da configuração adotada.

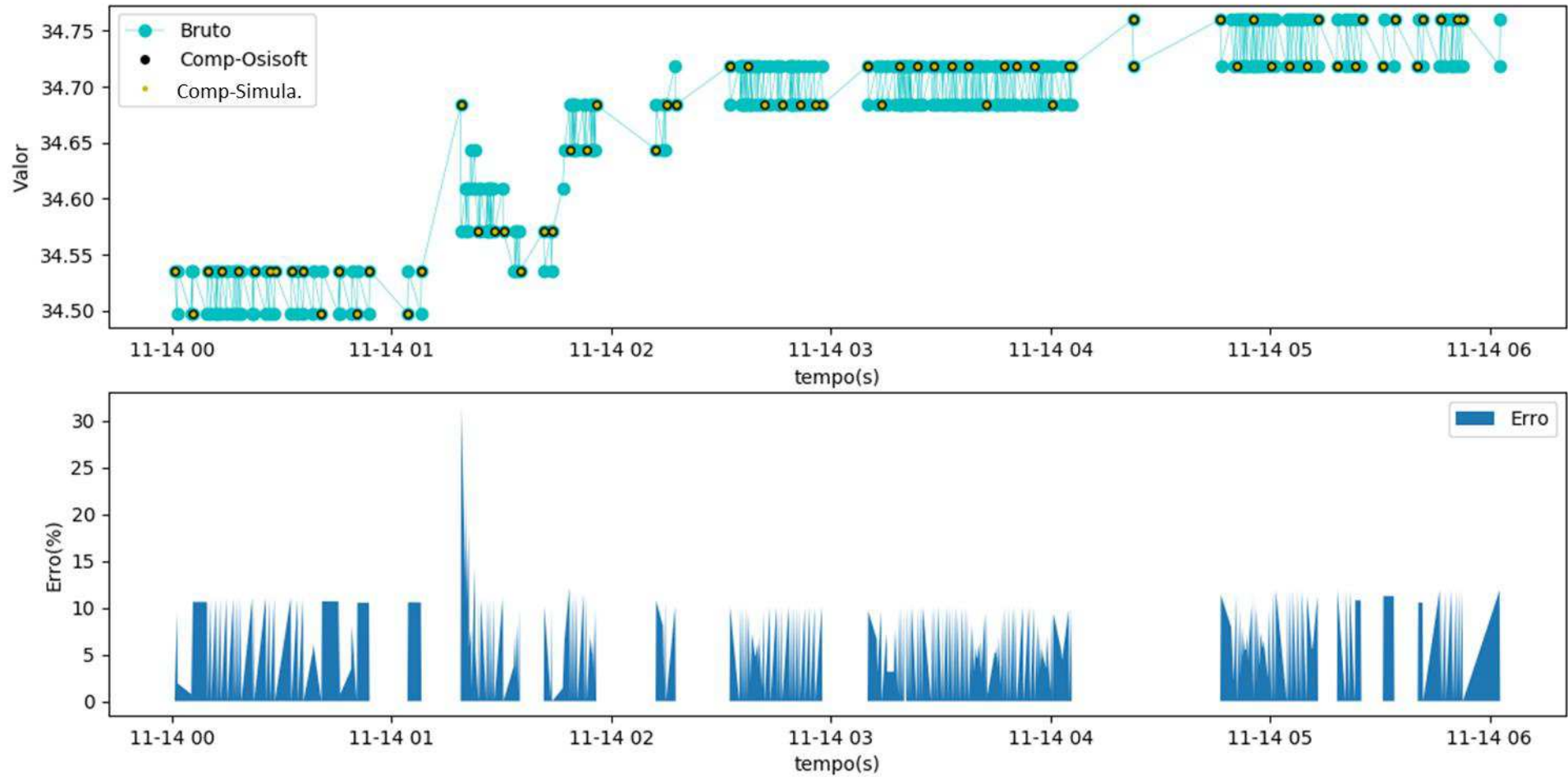
Tabela 10 Descrição da configuração do algoritmo de Compressão e erros obtidos pelo processo de Compressão de dados.

	Valor Numérico	Unidades
<b>Nº Pontos Brutos da Janela</b>	1517	-
<b>CompDev</b>	3	Unidade Eng.
<b>CompMax</b>	1800	Segundos
	Osisoft	Simulador
<b>Nº Pontos</b>	88	88
<b>Soma dos Erros Absolutos</b>	0.0006619377432387497	0.0006619377432387497
<b>Soma dos Esso Quadrados</b>	0.018765165171126596	0.018765165171126596

Fonte: Próprio Autor

Esta tabela nos permite confirmar que o objetivo de reproduzir o algoritmo padrão aplicado pelo PI System foi alcançado, pois em ambos os resultados gerados os valores dos indicadores equivalem entre si.

Figura 33 Erros gerados somente pelo SDT.



Fonte: Próprio Autor

### 6.2.3. Erros originados pelos processos de exceção e compressão combinados

Combinar os dois algoritmos de exceção e compressão foi o terceiro passo rumo a validação de ambos. Como esta combinação é o modo comumente aplicado no *PI System*, assegurar que ambos reproduzem o mesmo comportamento é crucial para o trabalho.

Foi aplicada à tag de referência uma janela de dados. Os dados da janela e dos parâmetros de configuração dos algoritmos de exceção e compressão de dados são apresentados na Tabela 11.

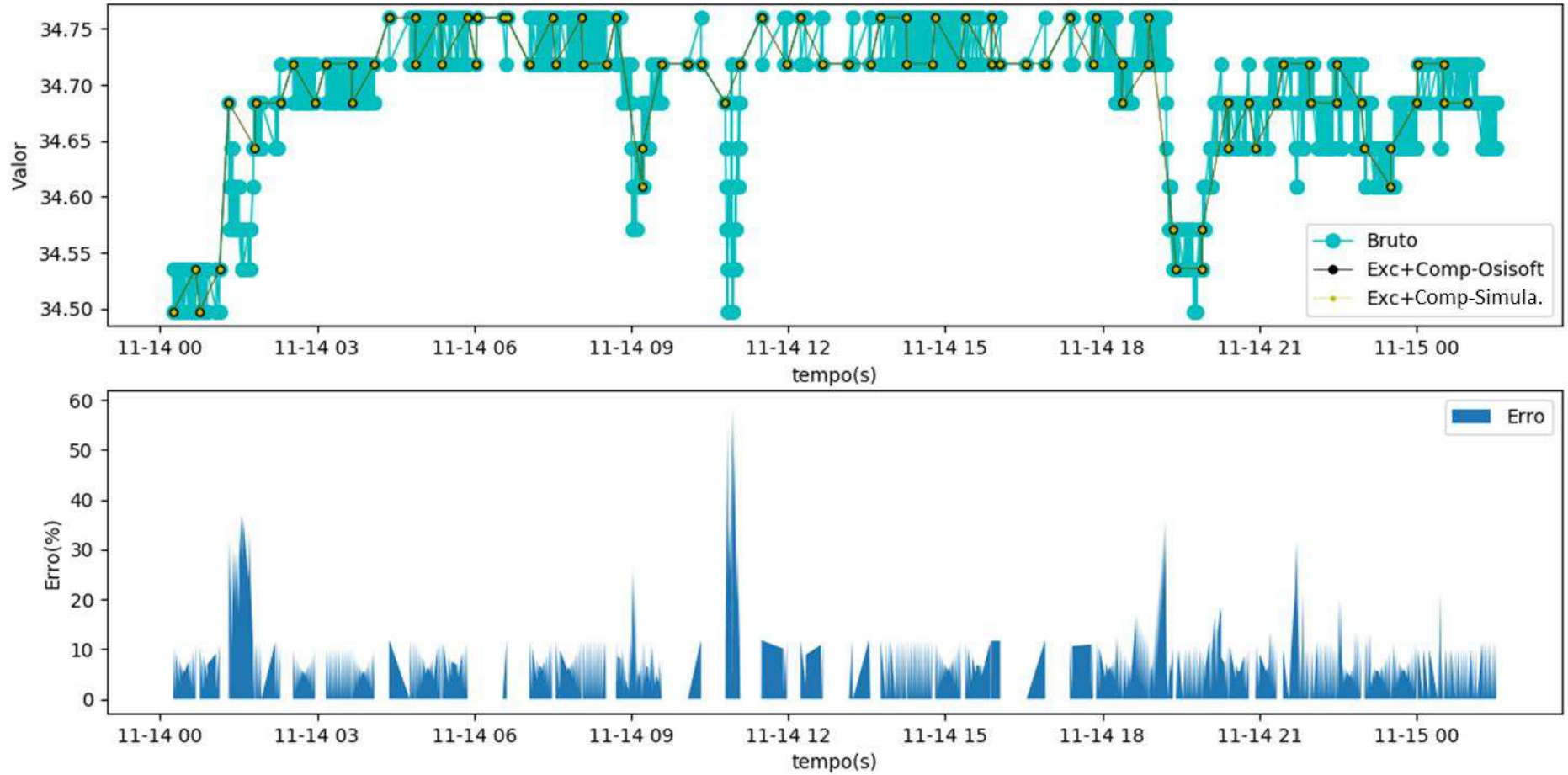
Portanto, a avaliação dos erros para mais de um conjunto de dados foi realizada. Tornou-se interessante avaliar nesse ponto os erros obtidos por ambos os algoritmos de exceção e compressão combinados. Os conjuntos de dados possuem as características descritas na abaixo.

Tabela 11 Descrição da configuração do algoritmo de Exceção e Compressão e erros obtidos pela combinação de ambos os algoritmos.

	Valor Numérico	Unidades
<b>Nº Pontos Brutos da Janela</b>	1517	-
<b>ExcDev</b>	0.5	Unidade Eng.
<b>ExcMax</b>	300	Segundos
<b>CompDev</b>	0.5	Unidade Eng.
<b>CompMax</b>	300	Segundos
	Osisoft	Simulador
<b>Nº Pontos</b>	86	86
<b>Soma dos Erros Absolutos</b>	0.0006619377432387497	0.018765165171126596
<b>Soma dos Erros Quadrados</b>	0.0006619377432387497	0.018765165171126596

Fonte: Próprio Autor

Figura 34 Erros gerados pela combinação dos ambos os algoritmos.



Fonte: Próprio Autor



### **6.3. Análise de compressão de dados pelo algoritmo Swinging Door Trending**

É realizada a análise dos parâmetros de compressão aplicado industrialmente, pelo sistema historiador da Osisoft, através da abordagem estatística dos dados antes e após o processo de compressão de dados. Esta análise se baseia em medida de centralidade e variabilidade dos dados.

Aplicar o algoritmo em dados reais obtidos em planta industrial tornará possível a avaliar o desempenho do parâmetro de compressão, desvio de compressão, e do parâmetro de exceção, desvio de compressão frente aos indicadores propostos. Para cada intervalo de configuração dos algoritmos de compressão e exceção, será possível a definição de um perfil de desempenho de compressão comparando-se a taxa de compressão obtida.

Avaliar as curvas destes indicadores frente à taxa de compressão possibilita estimar qual o conjunto de parâmetros de compressão e exceção que surtem efeito mínimo na alteração das informações contidas nos dados brutos.

#### **6.3.1. Estratégia de Avaliação Aplicada**

A avaliação de sensibilidade do algoritmo se deu através da compressão dos dados obtidos de uma variável industrial. A coletada destes dados foi realizada a partir de uma PI tag selecionada, cujas configurações de compressão foram desabilitadas.

Portanto, através dos dados brutos industriais foram construídos cenários de compressão e exceção diferentes em si. Estes permitem avaliar os indicadores estatísticos para configuração de ambos os processos de exceção e compressão de dados. Seguir são demonstrados os 3 cenários de avaliação adotados.

#### **6.3.2. Avaliação do algoritmo de Exceção de dados**

Neste cenário são apresentados os resultados da análise de sensibilidade sobre os dados industriais. Os impactos da exceção de dados são avaliados a partir dos indicadores adotados para tal avaliação.

Como parâmetros do algoritmo de exceção têm-se dois: *ExcDev* e *ExcMax*. Na avaliação dos efeitos destes parâmetros sobre os dados, análise foram individuais foram realizadas.

## Caso 01 – Avaliação de *ExcDev*

Em primeiro momento, foi possível estabelecer as seguintes condições para os parâmetros de entrada no algoritmo de exceção:

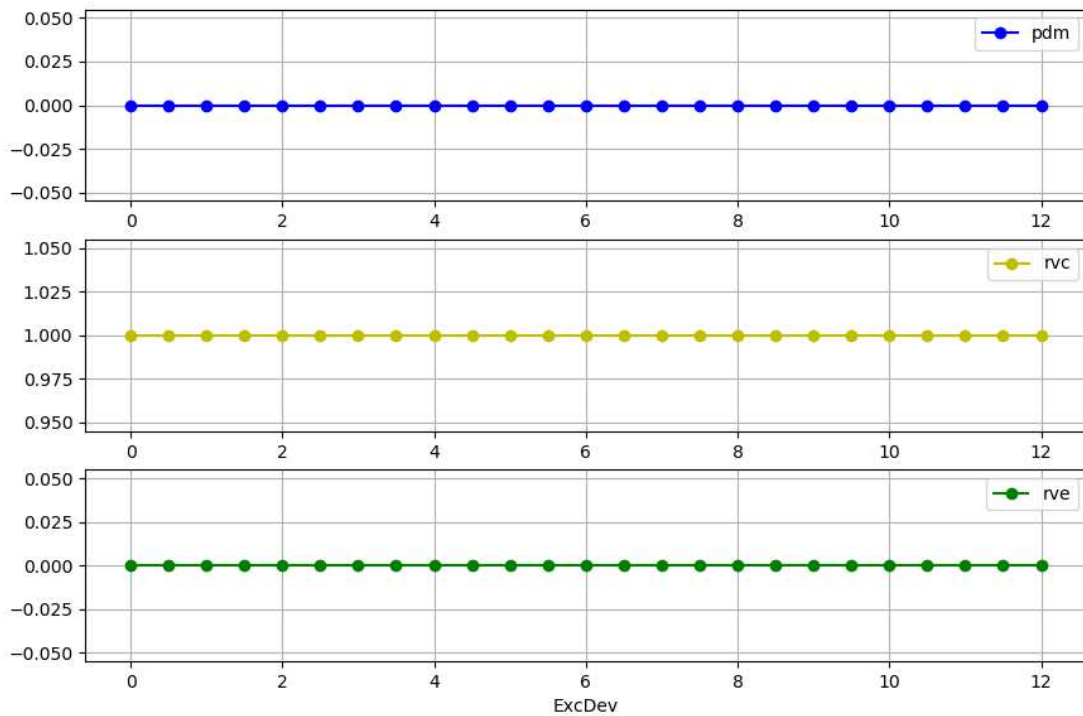
Tabela 12 Parâmetros de avaliação aplicados na avaliação da exceção de dados variando-se o parâmetro *ExcDev*.

	Início	Incremento	Nº de Iterações
<i>ExcDev</i>	0	0.05	25
<i>ExcMax</i>	1800s	0	0

Fonte: Próprio Autor

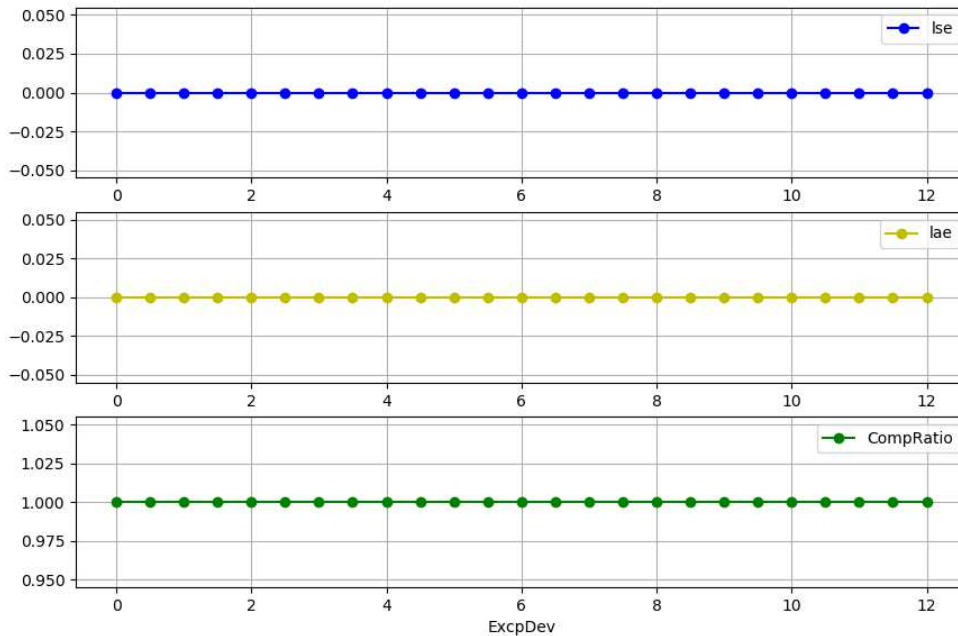
Para a base de dados adotada, contendo 7647 pontos, os resultados obtidos são os demonstrados a seguir:

Figura 35 Indicadores PDM, RVC e RVE obtidos variando-se *Excdev* Algoritmo de exceção.



Fonte: Próprio Autor

Figura 36 Indicadores ISE, IAE e COMPRATIO obtidos variando-se *Exdev* Algoritmo de exceção.



Fonte: Próprio Autor

É demonstrado através da Figura 35 que não houve alterações nas propriedades *PDM*, *RVC*, *RVE*. Igual comportamento foi percebido na e Figura 36, onde os indicadores *ISE*, *IAE* e *CompRatio*. Devido a este comportamento é verificado que o valor de *ExcDev* não exerce efeito sobre estes indicadores e nem alterou os dados de maneira alguma, nem na quantidade de pontos excluída da análise. Este comportamento é devido a não exclusão de dados, de modo que a quantidade de pontos finais é igual a quantidade de pontos iniciais.

### Caso-02 – Avaliação de *ExcMax*

A seguir foi realizada a avaliação do algoritmo de exceção de dados segundo a análise do parâmetro *ExcMax*. Seguindo no mesmo sentido da avaliação anterior, os valores do *ExcMax* serão assumidos entre demonstrados na Tabela 13.

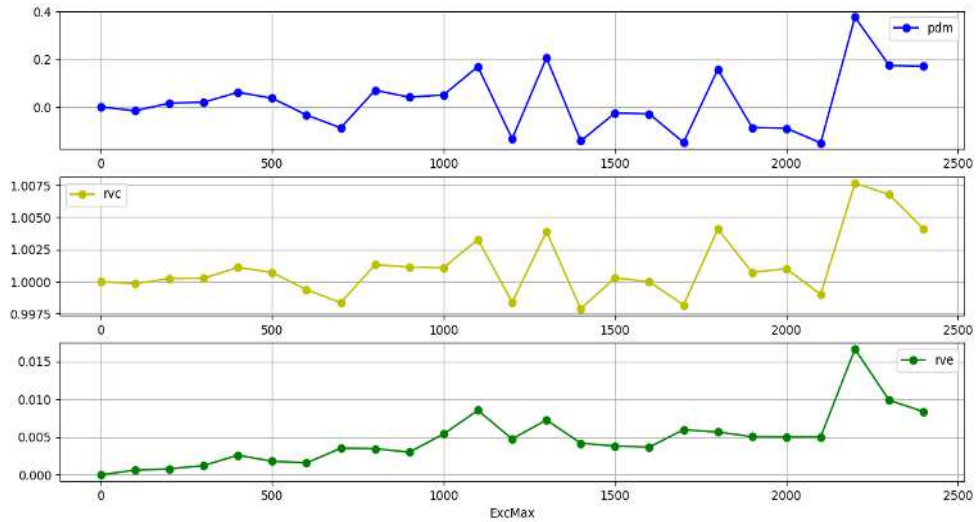
Tabela 13 Parâmetros de avaliação aplicados na avaliação da exceção de dados variando-se o parâmetro *ExcMax*.

	Início	Incremento	Nº de
<b>ExcMax</b>	0	100	25
<b>ExcDev</b>	3	0	0

Fonte: Próprio Autor

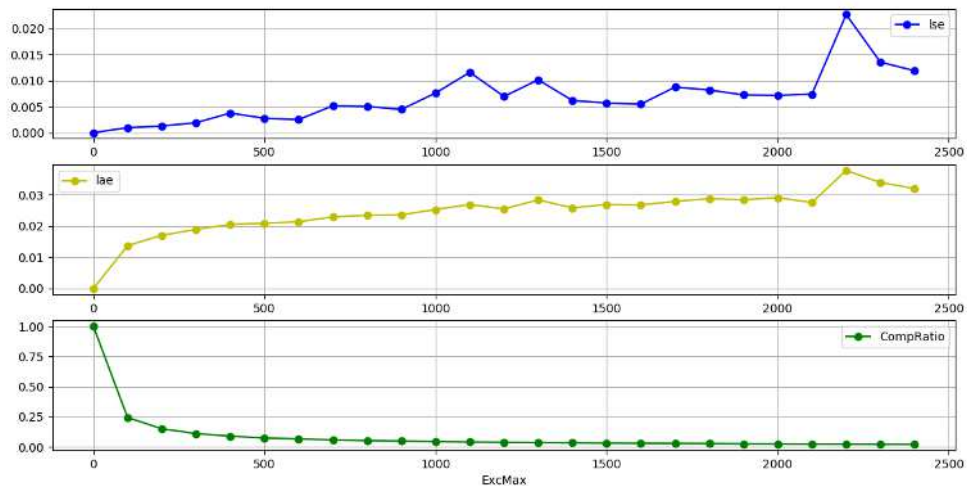
Como resultado de cada uma das avaliações realizadas, os indicadores de desempenho foram calculados. Desta forma, foi possível definir o perfil de cada um deles, como demonstrado abaixo nas Figura 37 e Figura 38.

Figura 37 Cenário 1, caso 2, Indicadores de PDM, RVC RVE avaliando-se *ExcMax*.



Fonte: Próprio Autor

Figura 38 Cenário 1, caso 2, Indicadores de mse, mae, CompRatio avaliando-se *ExcMax*.



Fonte: Próprio Autor

O perfil de demonstrado para os indicadores de PDM, RVC e RVE são demonstrados na Figura 37. Verifica-se a existência de instabilidade de nos indicadores a partir de um valor de *ExcMax* específico, 500. A partir deste valor no eixo x, os indicadores no eixo y têm seus valores alterados sem seguir uma tendência específica.

O comportamento oscilatório apresentado é causado pela exclusão de pontos específicos, os quais são diferentes entre si, mesmo que a razão de compressão não se altere. Ou seja, os pontos excluídos em cada configuração do algoritmo são diferentes entre si, mesmo que a quantidade de pontos finais seja igual. Este comportamento exerce influência sobre os indicadores.

### **Caso-03 Avaliação conjunta de *ExcMax* e *ExcDev***

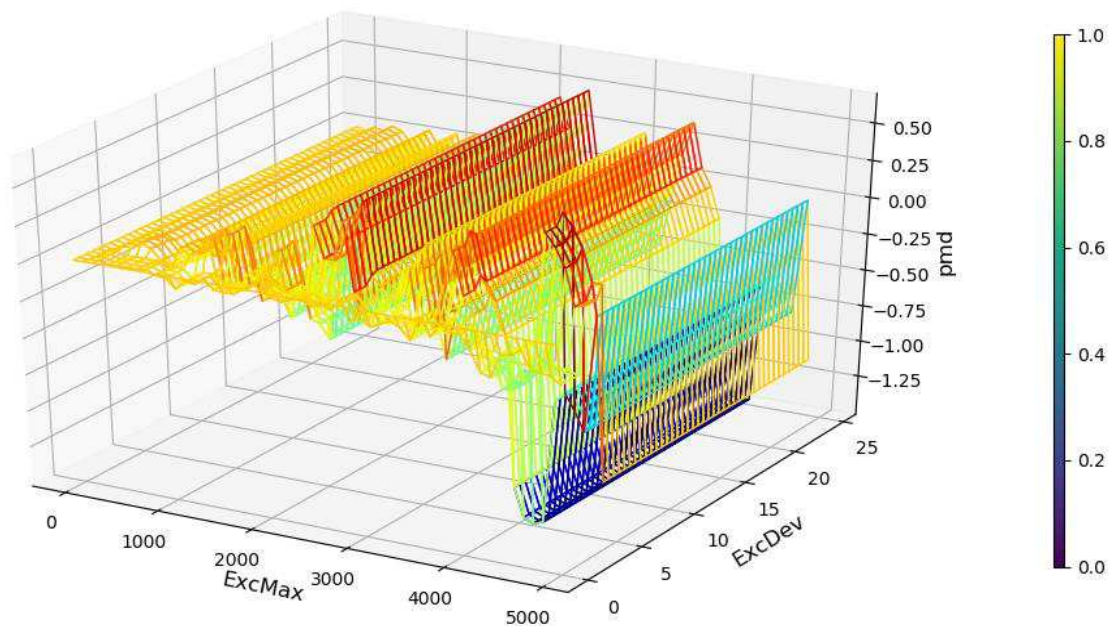
Como sugerido acima os dois parâmetros de configuração do algoritmo de exceção, *ExcMax* e *ExcDev*, exercem influência sobre cada característica elucidada por cada indicador avaliada. Portanto, faz-se necessário verificar o grau de impacto destas duas variáveis sobre cada um dos indicadores.

Para esta verificação, lançou-se mão da análise de cada indicador dentro de um intervalo definido. Ambos os parâmetros *ExcMax* e *ExcDev* tiveram seus valores assumidos entre os intervalos que contemplaram os valores padrão adotados pelo PI System na criação de tags padrão do tipo numérica. A seguir são demonstrados cada um dos indicadores adotados no presente trabalho.

Da Figura 39 à Figura 45 demonstra-se o comportamento dos indicadores frente a valores específicos dos parâmetros *ExcMax* e *ExcDev* de configuração do algoritmo de exceção de dedos. A análise de cada gráfico é realizada comparando-se com o gráfico representante da Razão de Compressão, considerando que este há de definir até que ponto houve melhora na redução no número de pontos armazenados.

A Figura 39 demonstra o comportamento oscilante do indicador PMD. É possível verificar que existe uma faixa na qual há maior oscilação neste indicador é maior. Para valores de *ExcDev* na faixa de 0 a 5 e *ExcMax* entre 0 e 1000, PDM não sofre grande alteração. Devido a isso, estas faixas são consideradas como boas estimativas de configuração a tag em análise. É perceptível que à medida que *ExcMax* aumenta, PDM sofre cada vez mais oscilação, culminando numa grande perda de tendência entre os valores de 4000 e 5000.

Figura 39 Avaliação do indicador PDM para o algoritmo de exceção.

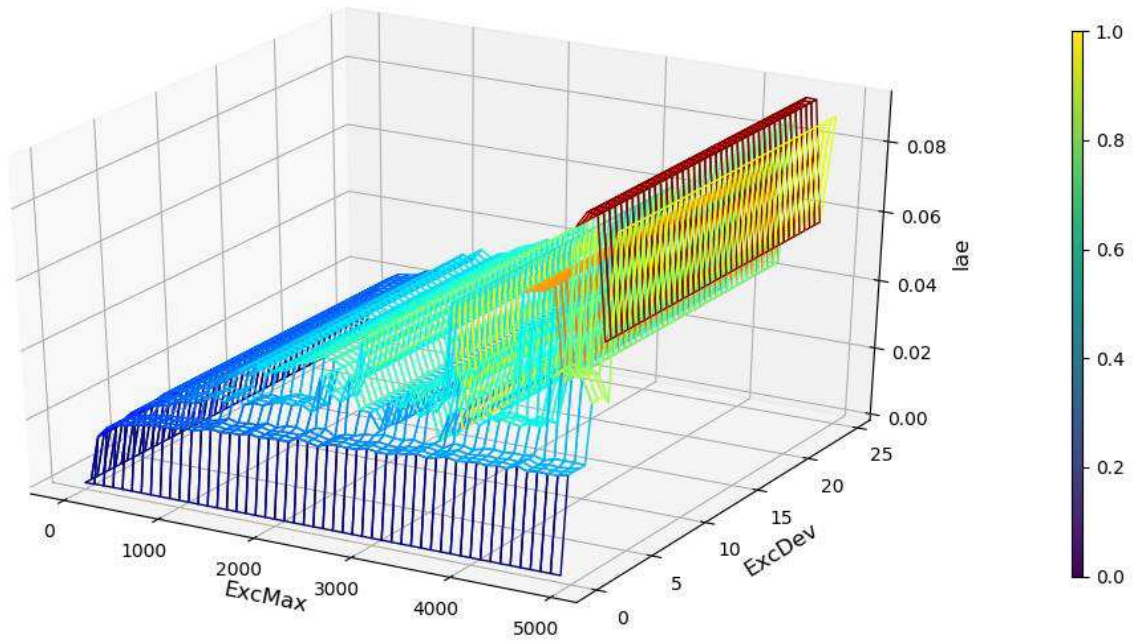


Fonte: Próprio Autor

Nas Figura 40 e Figura 41 encontram-se representados os perfis de comportamento dos indicadores MAE e MSE, respectivamente. Em ambos é possível verificar a existência de uma faixa de configuração dos parâmetros *ExcMax* e *ExcDev* que determina a transição de um estado menos caótico em termos de variação de valores para um a mais caótico, no qual oscilações cada vez maiores estão presentes.

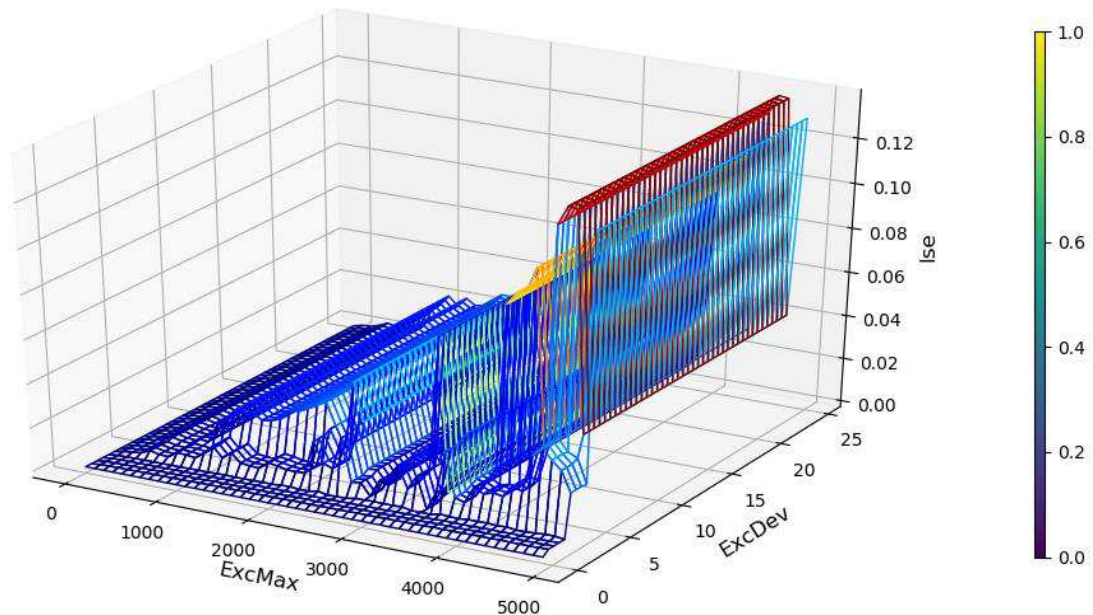
Para estes a faixa é definida entre os valores 0 e 5 para o *ExcMax* tanto para o *ExcMax* quanto para o *ExcDev*. É perceptível que o *ExcMax* rege os valores tanto do MAE quanto do MSE. No entanto, as escalas de ambos os indicadores são diferentes por definição, o que possibilita ressignificar as pequenas variações em MAE em termos de curvas evidenciadas em MSE. Em termos de análise, evidencia-se que ambos os parâmetros do algoritmo de exceção podem ser ajustados, de maneira a prevenir que erros maiores sejam obtidos.

Figura 40 Avaliação do indicador MAE para o algoritmo de exceção.



Fonte: Próprio Autor

Figura 41 Avaliação do indicador MSE para o algoritmo de exceção.



Fonte: Próprio Autor

Considerando a avaliação da variância antes e após o processo de exceção de dados, é possível verificar alteração com o variar de ambos os parâmetros *ExcMax* e

*ExcDev*. As figuras a seguir demonstram os perfis da variância dos dados e dos erros de reconstrução com relação à variância dos dados antes do processo de exceção.

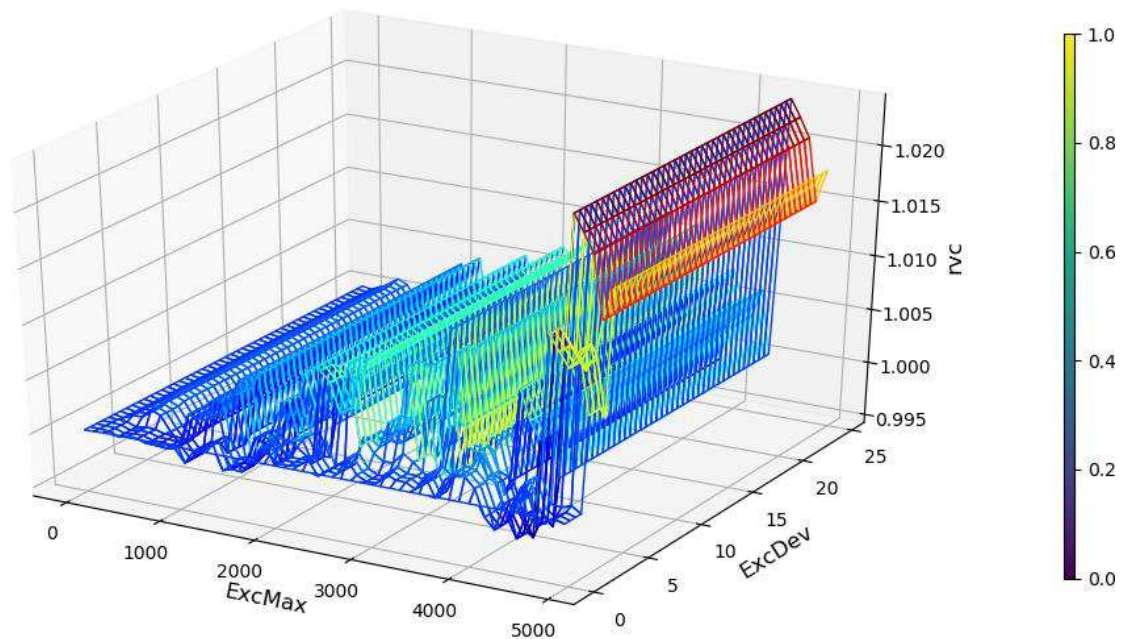
Para verificar o efeito do processo de exceção de dado sobre a variância dos dados reconstruídos, a Figura 42 apresenta o gráfico do indicador RVC. Este indicador demonstra a deformação causada sobre a variância levando-se em conta todos os pontos reconstruídos. A variância dos dados é igual a unidade, como esperado.

Quando ambos os parâmetros *ExcMax* e *ExcDev* são definidos incrementalmente é possível verificar cada avaliação do RVC para cada par de configuração do algoritmo de exceção. Conforme o valor de ambos os parâmetros aumenta, maior a é o valor de RVC. Porém, o maior contribuidor para o aumento de variância é o *ExcMax*, pois à medida que seu valor aumenta, maior é de RVC para o mesmo *ExcDev*. O mesmo comportamento não é observado para o *ExcDev*, que não exerce mesmo efeito sobre o indicador.

É constatado que o delta de variação do RVC não alcança ordem de grandeza maior que centésimos de unidade. Encontra partida, temos que verificada novamente, a existência de uma região na qual o valor do RVC não é alterado significativamente. Nesta, a variação dos valores assumidos pelo indicador é reduzida, evidenciando a existência de pares de *ExcMax* e *ExcDev* nos quais o valor assumido por RVC é mínimo. Ou seja, é possível definir estes pares de maneira a alcançar redução consideráveis na quantidade pontos brutos, assegurando ainda a variância reduzida.



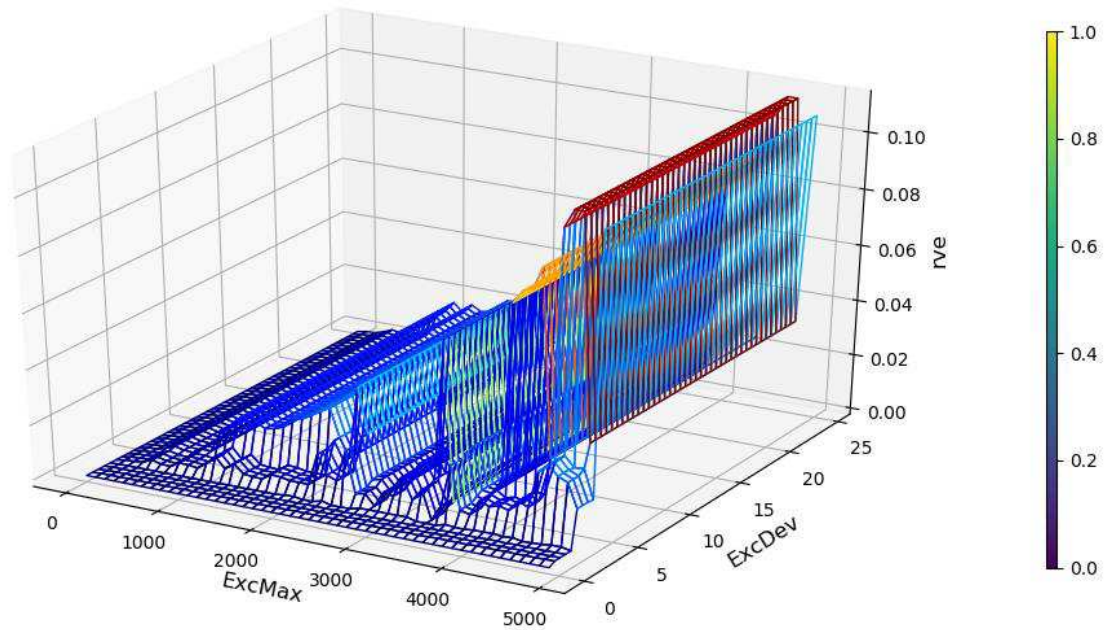
Figura 42 Avaliação do indicador RVC para o algoritmo de exceção.



Fonte: Próprio Autor

Avaliação da variância dos erros objetiva verificar a quão acentuada é a geração dos erros em cada par de parâmetros *ExcMax* e *ExcDev*. Semelhantemente as considerações do indicador RVC, o presente indicador possui uma variação reduzida, porém partindo do ponto zero, onde há erro gerado, por não haver nenhum ponto excluído, ou seja, os pontos reconstruídos coincidirem com os pontos brutos. Na Figura 43 é verificada a existência de uma região onde ocorre a transição de um estado com melhores valores de RVE e para valores cada vez mais oscilantes. Portanto, é mantida a consideração da existência de pares *ExcMax* e *ExcDev* associados a manutenção dos valores reduzidos de RVE concomitantemente a manutenção da qualidade dos dados e razão de compressão em nível aceitável.

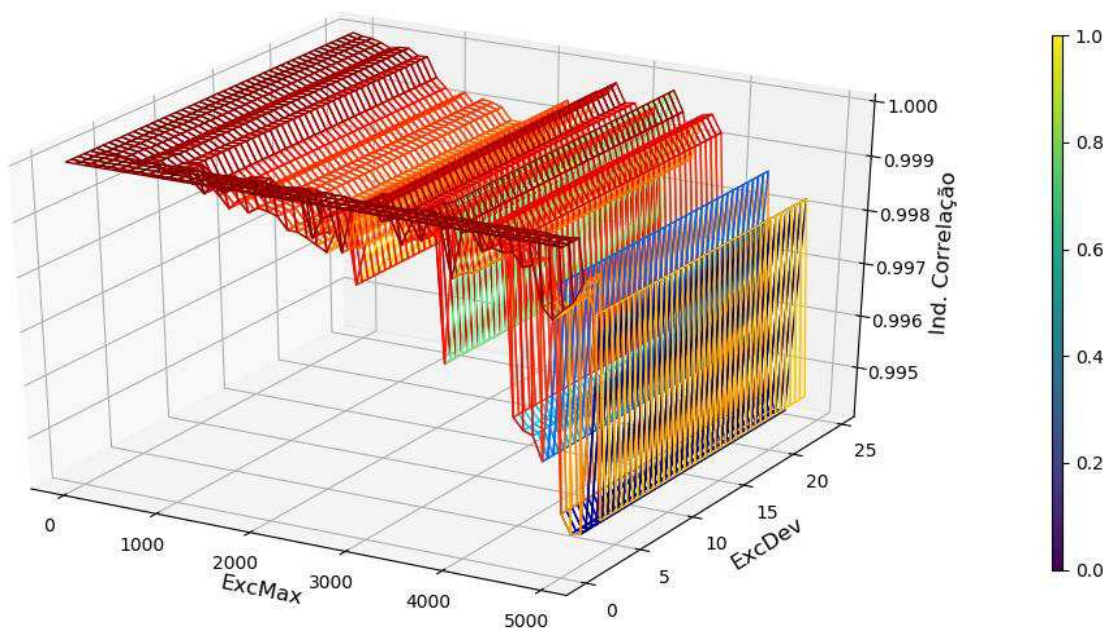
Figura 43 Avaliação do indicador RVE para o algoritmo de exceção.



Fonte: Próprio Autor

A avaliação da linearidade entre os dados brutos e dados reconstruídos é realizada pelo índice de correlação de Pearson, ilustrada na Figura 44. Para cada par *ExcMax* e *ExcDev* foi realizada a avaliação deste indicador. A medida que o ambos os parâmetros sofrem incrementos sucessivos, verifica-se a degradação da linearidade nos dados reconstruídos, fortemente indicada pela variação deste indicador em ordem de grande considerada alta, tendo em vista que este indicador assume valores entre 0 e 1, onde zero indica nenhuma correlação linear entre os dados e 0 zero indica máxima correlação linear entre os dados. Ao observar o perfil deste indicador, evidencia-se a existência da região de conservação da característica de linearidade entre os dados, estabelecida para valores de *ExcMax* próximo de 100 e *ExcDev* em torno 4.

Figura 44 Avaliação do indicador Índice de correlação para o algoritmo de exceção.

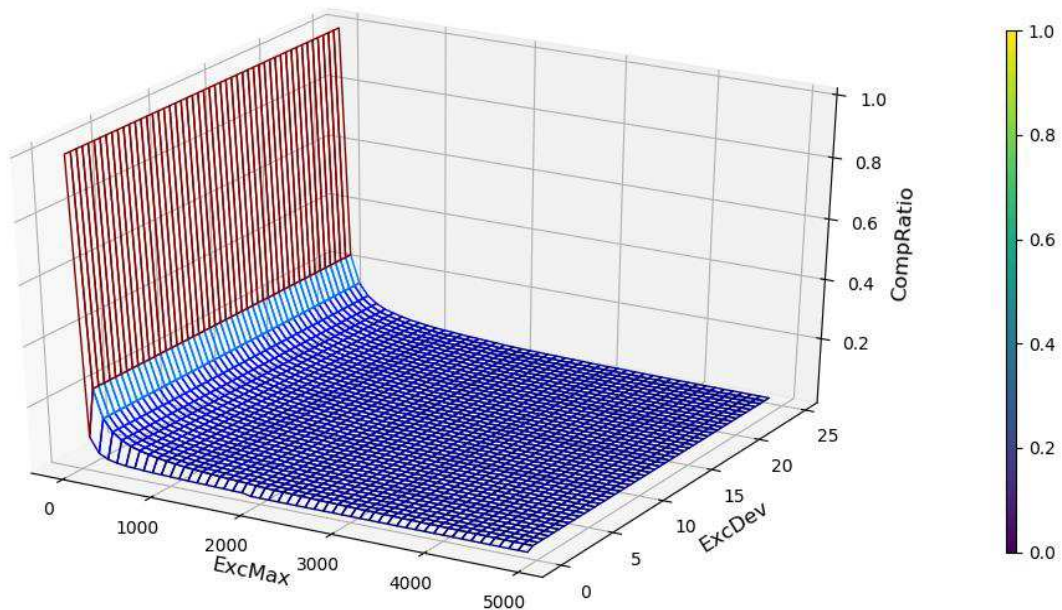


Fonte: Próprio Autor

A Figura 45 demonstra a curva do Indicador Razão de Compressão para o algoritmo de exceção, o qual não pode ser definido como configuração ao algoritmo, mas sendo este calculado com os dados resultantes de saída. No entanto, sua importância está em definir a o percentual de pontos foram mantidos e enviados para o processo compressão de dados. A curva gerada foi originada pela configuração do algoritmo de exceção para cada par de *ExcMax* e *ExcDev* até os limites demonstrados graficamente.

Evidencia-se através desta curva a existência de uma região de parâmetros ótimos para este algoritmo. É percepção é alcançada ao observar que a razão de compressão é máxima quando, igual à unidade, quando não existe exceção, e mínima quando os valores de ambos os parâmetros são possuem valores incrementados. No entanto, o incremento de ambos os parâmetros não exerce o mesmo impacto no indicador *CompRatio*. Ao avaliar individualmente os parâmetros verifica-se que o valor *CompRatio* é sensível a alterações de *ExcMax*, enquanto permanece inalterável em função de alterações em *ExcDev*. Tal comportamento reafirma a existência de uma configuração ótima de operação do algoritmo de exceção de dados para presente tag, neste caso, 6 em torno de 1000 a 500, e *ExcDev* próximo aos valores em torno de 0,5.

Figura 45 Avaliação do indicador Razão de Compressão para o algoritmo de exceção.



Fonte: Próprio Autor

### 6.3.3. Avaliação do algoritmo de Compressão de dados

O presente cenário visa determinar os indicadores de compressão adotados para o algoritmo de compressão SDT. Como parâmetros de avaliação tomaram-se os dois identificados como chave, *CompMax* e *CompDev*. Estes são vistos como restritivos ao armazenamento de todos os dados de entrada, ou seja, a configuração destes definidos como valor zero, implica no armazenamento total dos dados de entrada.

#### Caso 01 – Avaliação de *CompMax*

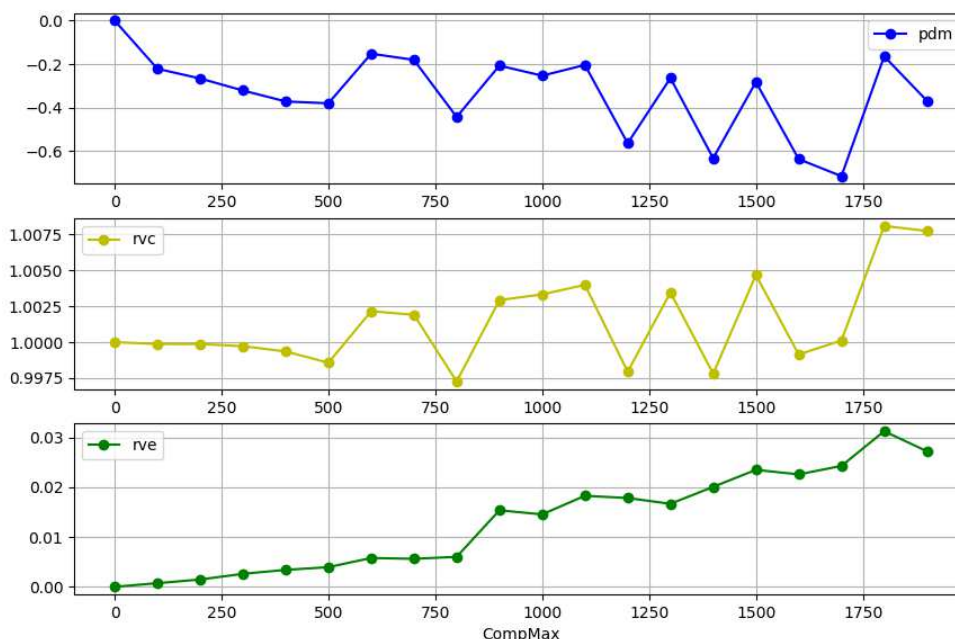
O primeiro parâmetro tomado para a avaliação é o *CompMax*. A sua avaliação se dá entre duas condições do algoritmo. A primeira condição se é sem compressão, no qual o valor de *CompMax* é igual a zero. A segunda é um estado no qual o valor de *CompMax* seja elevado, passando até do valor 1800, tomado como padrão de início de pela Osisoft. Para este caso, a Figura 11 demonstra os respectivos valores assumidos por ambos os parâmetros. A Figura 46 demonstra os perfis dos indicadores produzidos para cada valor de *CompMax* adotado.

Tabela 14 Parâmetros de avaliação aplicados na avaliação da compressão de dados variando-se o parâmetro *CompMax*.

	Início	Incremento	Final
<b>CompMax</b>	0	100	10
<b>CompDev</b>	3	0	3

Fonte: Próprio Autor

Figura 46 Indicadores PDM, RVC e RVE obtidos variando-se *CompMax* algoritmo de compressão SDT.



Fonte: Próprio Autor

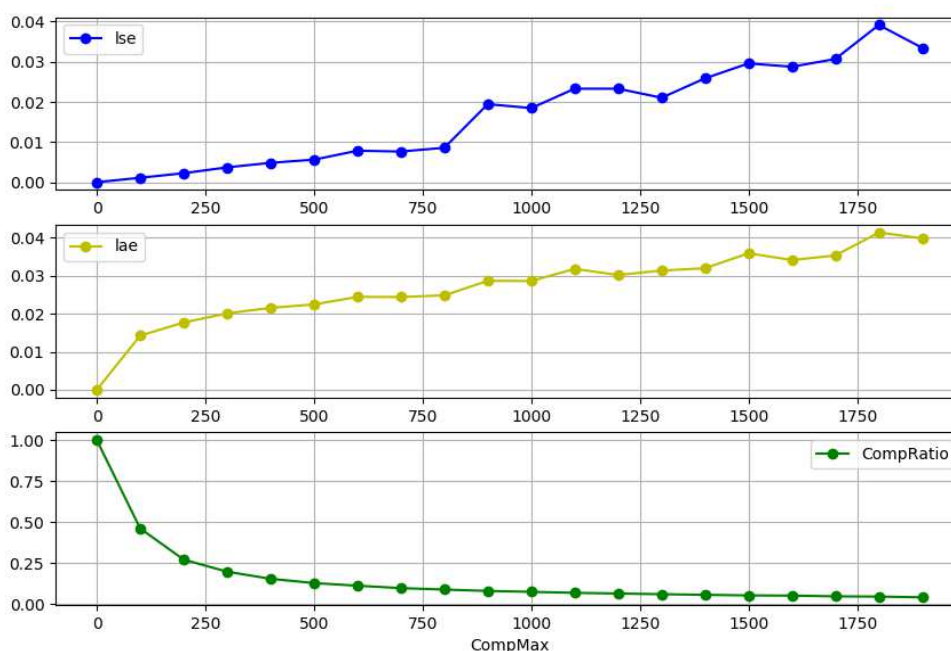
Pode-se verificar que o valor o indicador mais afetado em tal análise é o PDM. Observa-se que seu valor varia em forma decrescente na ordem de décimos de percentuais. Tal comportamento sugere que este seja o indicador mais sensível nesta avaliação.

Em detrimento do primeiro indicador, PDM, os outros dois indicadores tiveram sua amplitude de avaliação menos que o anterior. Ambos em ordem de grandeza na casa de unidades percentuais ou menor.

É verificado que para os três gráficos, presentes na figura acima, tiveram oscilações originadas a partir do valor de 500s assumido por *CompMax*. As tendências destes perfis não se alteram a partir deste ponto, mesmo quando sugerindo a mudança de direção abrupta.

Portanto, a variância do erro, RVC, é definida como decrescente antes deste ponto e depois é verificada a presença de oscilação. O mesmo comportamento se dá com relação ao indicador da média, PDM, porém em ordem de maior. Na Figura 47, os demais indicadores de compressão têm seus perfis determinados para as mesmas condições de *CompMax* e *CompDev*.

Figura 47 Indicadores ISE, IAE e COMPRATIO obtidos variando-se *CompMax* algoritmo de compressão SDT.



Fonte: Próprio Autor

Ao considerar a avaliação dos erros quadrados e absolutos, IAE e IAS, pode-se verificar o comportamento semelhante ao anteriormente verificado na Figura 47, em torno do valor de 500s assumido por *CompMax*. As oscilações posteriores a este valor são menores quando comparados como indicadores anteriores. A ordem de grande é nas unidades percentuais.

Ao verificar que quando o valor de *CompMax* assume o valor de 500s, o valor da razão de compressão já assume valores próximos do recomendado pela Osisoft, entre 10% e 1%. Isto sugere que o valor de *CompMax* aceitável seria de 500s.

#### Caso 02 - Avaliação de *CompDev*

O presente caso refere-se à avaliação do algoritmo de compressão quando o *CompDev* assume valores variantes entre o estado de inexistência de compressão e

de valores além do estipulado como padrão pela Osisoft, definidos no momento de criação de tag.

A Tabela 15 contém os valores assumidos pelos parâmetros de compressão adotados para análise. Neste caso, mantém-se o valor de *CompMax* fixo enquanto, varia-se os valores de *CompDev*.

Tabela 15 Parâmetros de avaliação aplicados na avaliação da compressão de dados variando-se o parâmetro *CompDev*.

	Início	Incremento	Final
CompDev	0	0.05	10
CompMax	1800s	0	1800s

Fonte: Próprio Autor

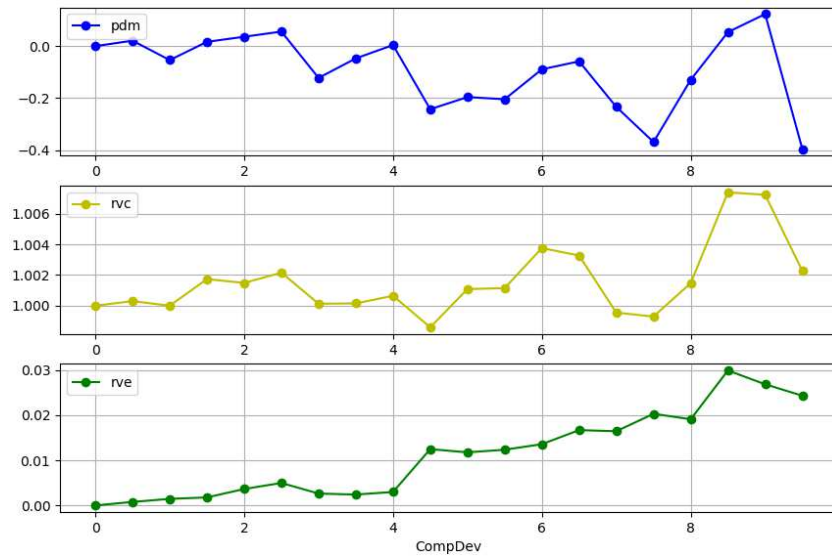
Ambas as Figura 48 e Figura 49, ilustram o comportamento de todos os indicadores estatísticos propostos. É perceptível que as escalas são diferentes entre si. No entanto, é possível observar as oscilações com relação ao mesmo eixo Y.

Os indicadores de PDM, RVC e RVE são tratados na Figura 48. Dentre estes, o indicador PDM possui maior oscilação a partir do valor 2, assumido pelo *CompDev*. Este mesmo valor 2 não é percebido como valor anterior a oscilação assumida por estes indicadores. Seguindo na direção do eixo X, é notada oscilação a partir do valor 4, assumido por *CompDev*, para o indicador RVE e RVC. Essa oscilação, no entanto, não define o ganho em relação a razão de compressão de dados, situada na Figura 49.

Ao comparar-se PDM, RVC e RVE com os ganhos em relação razão de compressão, verifica-se que a partir do valor 2 de *CompDev* não existe ganho sobre a compressão de dados. A razão de compressão também demonstra que o ganho não há vantagem em elevar o valor de *CompDev* além do valor 4. Os indicadores de erro, ISE e IAE, demonstram a elevação dos erros à medida que valor de *CompDev* aumenta.

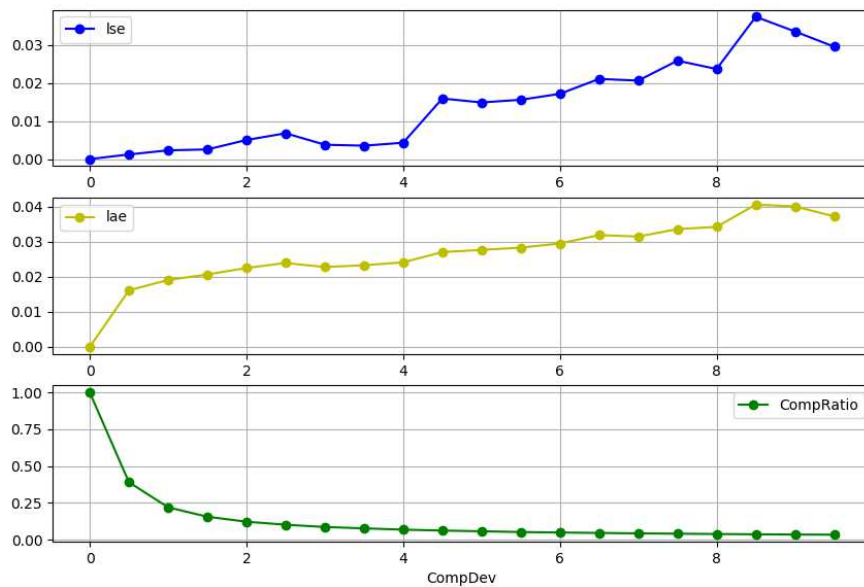
Portanto, faz-se necessária a avaliação destes indicadores em dados comprimidos a partir de outros valores de *CompMax*. Estas novas análises possibilitam a investigação da existência de uma região de compressão máxima aceitável estatisticamente.

Figura 48 Indicadores PDM, RVC e RVE obtidos variando-se CompMax algoritmo de compressão SDT.



Fonte: Próprio Autor

Figura 49 Caso 2, Indicadores de MSE, MAE, CompRatio avaliando-se CompDev.



Fonte: Próprio Autor



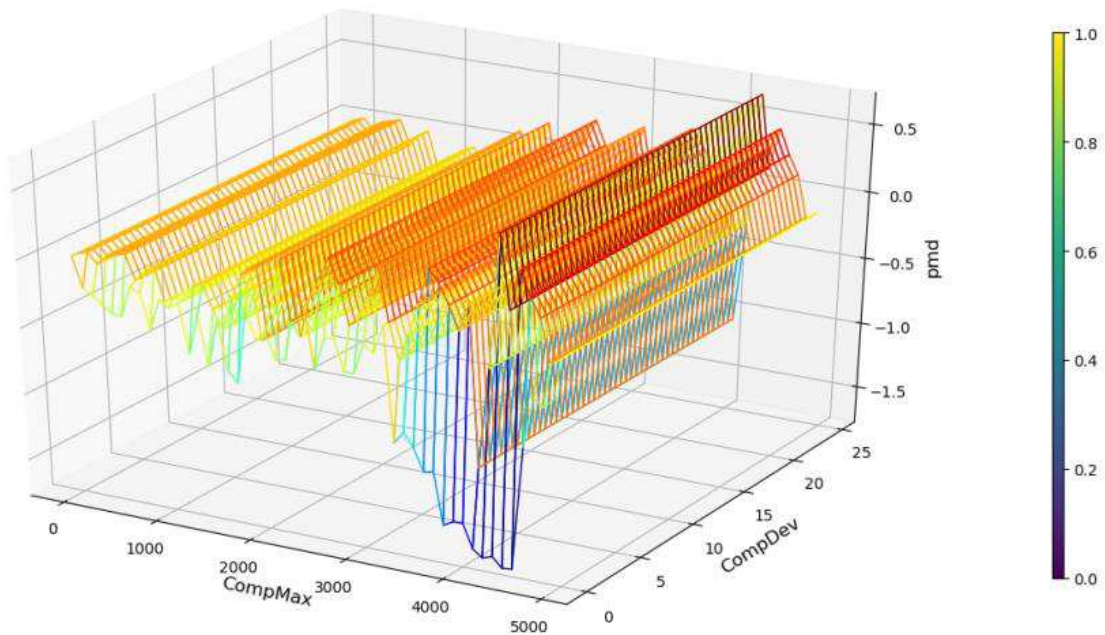
### Caso 03 – Avaliando ambos *CompMax* e *CompDev*

Considerando ambos os parâmetros de configuração do algoritmo de compressão como variável, pode-se construir gráficos em 3d, nos quais ambos os parâmetros são avaliados. Com os resultados, segue-se ao cálculo de cada indicador estatístico apresentado como indicador estatístico.

Todos os gráficos apresentados da Figura 50 até Figura 56 são obtidos pela plotagem dos indicadores avaliados em cada par *CompMax* e *CompDev*. Semelhantemente a análise realizada para o algoritmo de exceção, aqui são apresentados todos os indicadores.

A Figura 50 demonstra o indicador PMD, destinado a apresentar os efeitos da compressão sobre a medida de tendência central. Diferente do comportamento apresentado por este indicador quando avaliados sobre os dados resultantes da exceção de dados, verifica-se que oscilações estão presentes, mesmo quando valores de *CompMax* e *CompDev* são pequenos. Este comportamento é devido à natureza não linear do algoritmo compressão de dados. E notada a escala do eixo Y como altera minimamente entre -1.5 e 0.5. À medida que o valor *CompMax* é aumentado, o PDM decai gradativamente, independentemente do valor que o *CompDev* assume. Ao atingir este mínimo local, a oscilação de torna-se maior, de maneira não estabelecer nenhum padrão identificável, ou seja, o distanciamento da média atual distancia-se da média original.

Figura 50 Avaliação do indicador PDM para o algoritmo de compressão.



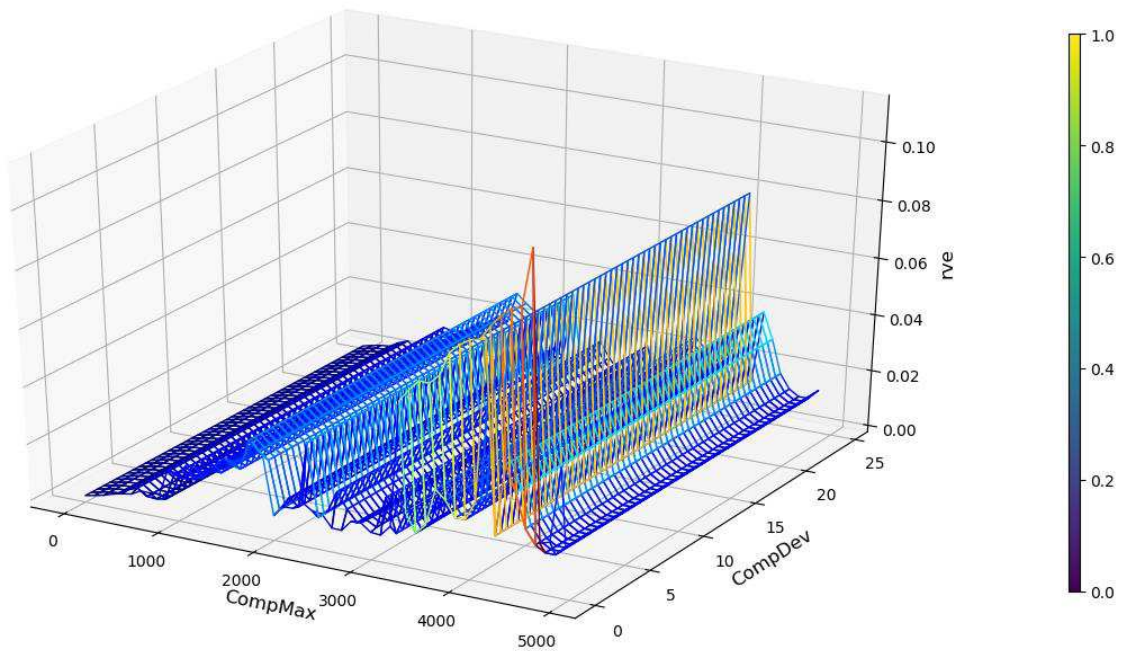
Fonte: Próprio Autor

A Figura 51 e Figura 52 demonstram o comportamento dos indicadores RVE e RVC, que avaliam a dispersão dados com relação variância dos erros gerados e a variância dos dados reconstruídos, respectivamente. As escalas de ambos são diferentes entre si, pois para o erro o valor 0 (zero) corresponde ao estado de inexistência de compressão. Porém para a razão das variâncias, o estado inicial é 1 (um), pois a variância dos dados final e inicial são iguais entre si.

O RVE demonstra aderência à conclusão tirada do gráfico de PDM, a existência da região onde ocorre a inversão de tendência destes indicadores.

É possível verificar que a tendência neste gráfico também foi quebrada entre os valores de 1800 e 2000 para o *CompMax*, evidenciando que o valor de *ComDev* não exerceu forte influência sobre os indicadores. A partir deste valor de *CompMax*, a existência de oscilações no valor do RVE. Antes deste ponto, é consistente a o aumento de RVE com relação ao aumento de *CompMax*, porém o mesmo comportamento é presente para o aumento de *ComDev*, evidenciando que com valores em torno de 2 e 5, o RVE não se altera. Este comportamento, define que o *CompMax* permite a avaliação de mais pontos durante o processo de compressão, sem que haja o armazenamento obrigatório de pontos devido a este parâmetro, adicionando-se maior dispersão no erro de reconstrução dos dados

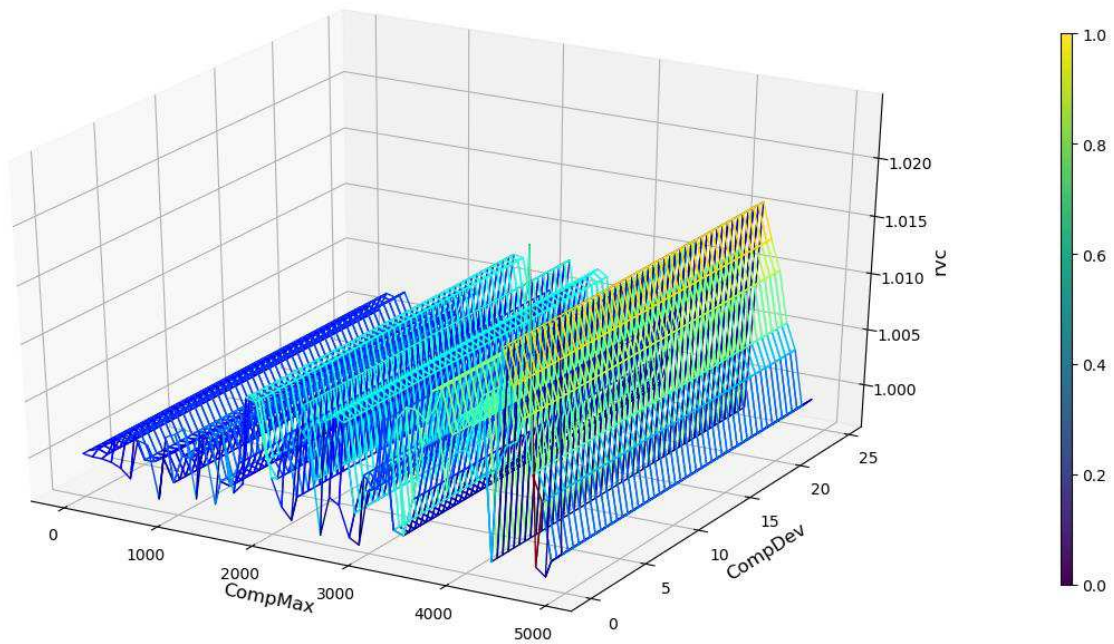
Figura 51 Avaliação do indicador RVE para o algoritmo de compressão.



Fonte: Próprio Autor

O comportamento apresentando por RVC é semelhante ao comportamento de RVE, porém não é verificada a transição tão facilmente como na figura anterior. A dependência de RVE com o *CompDev* é verificada. As oscilações no indicador são perceptíveis para valores de *CompMax* em torno de 1000, após isso ocorre um decaimento e posterior oscilação e em seus valores.

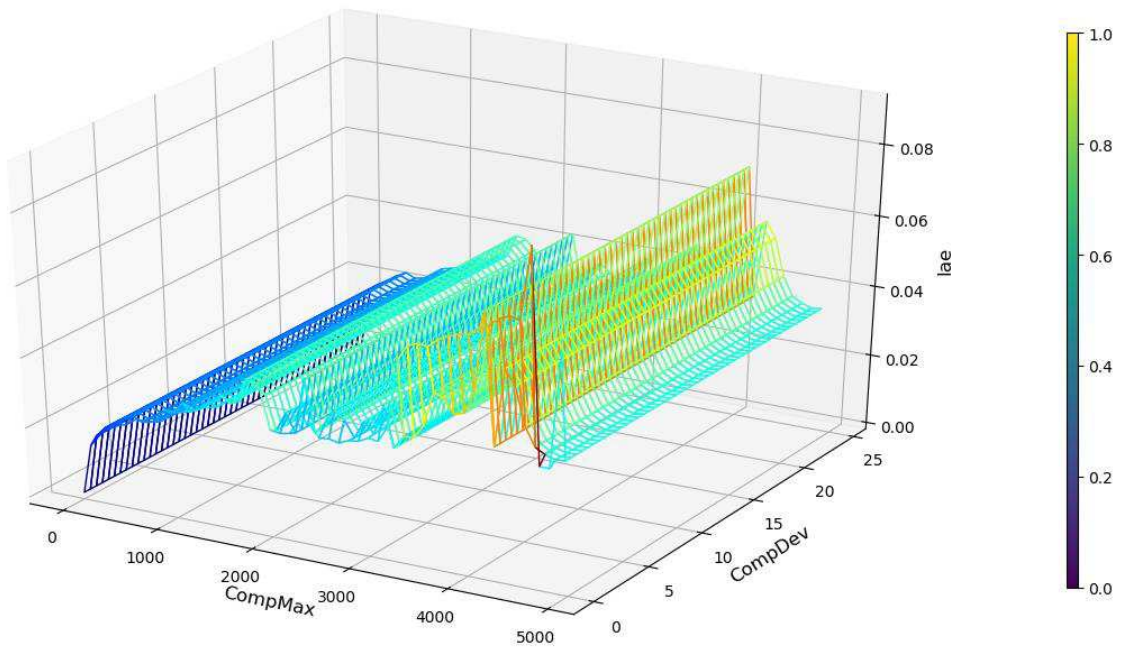
Figura 52 Avaliação do indicador RVC para o algoritmo de compressão.



Fonte: Próprio Autor

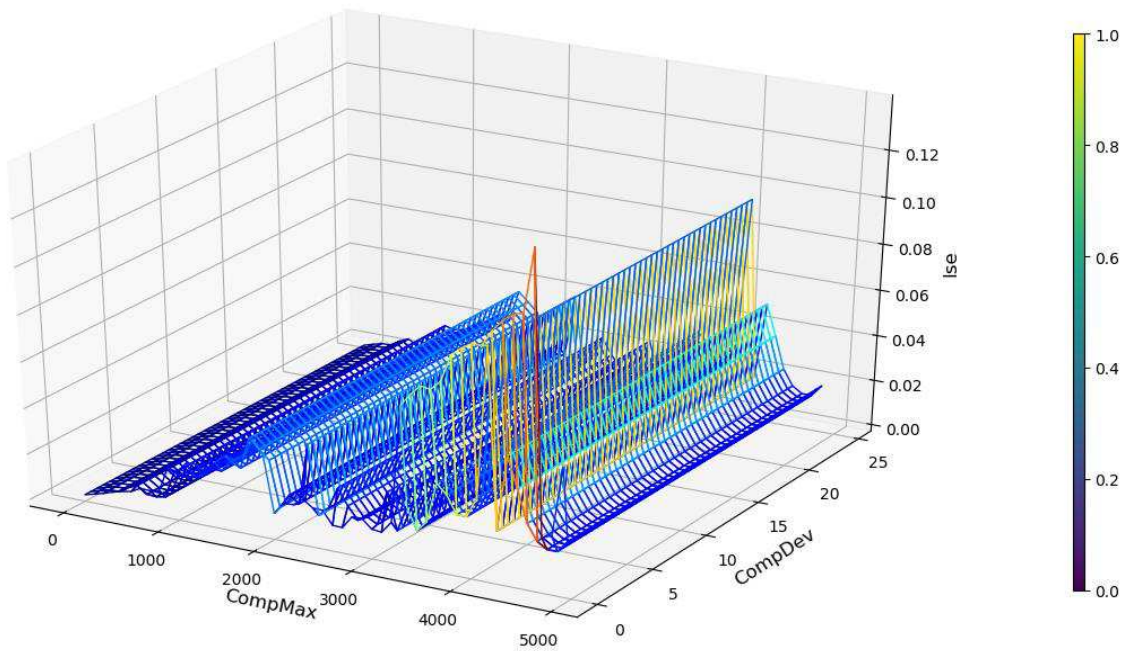
A Figura 53 e Figura 54 demonstram a avaliação do erro quadrado médio e erro absoluto, respectivamente. Ao seguir os eixos no sentido crescente, percebe-se que existe o aumento do erro com relação a *CompMax*, o mesmo não pode ser dito para *CompDev*, excesso para valores reduzidos, na transição do estado com e sem compressão de dados. Observa-se por ambos os indicadores que o valor de erro possui tendência consistente com até o valor 1000 de *CompMax*, após este valor tanto o ISE quanto o IAE assumem valores não consistentes e oscilantes.

Figura 53 Avaliação do indicador MAE para o algoritmo de compressão.



Fonte: Próprio Autor

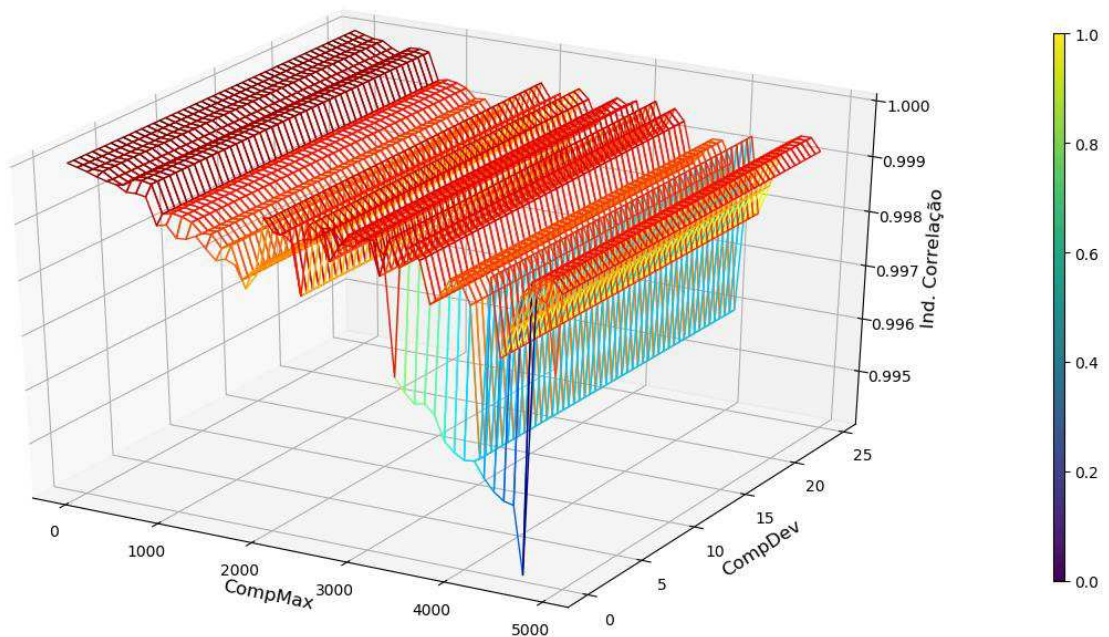
Figura 54 Avaliação do indicador MSE para o algoritmo de compressão.



Fonte: Próprio Autor

Os valores assumidos pelo índice de correlação linear são apresentados na Figura 55. A curva do indicador possui comportamento de tendência consistente e consistente. Vale ressaltar a região na qual o valor de *CompMax* equivalente a 1000, ocorre o primeiro decaimento no valor destes dados. Ao considerar este resultado com os anteriores para esta mesma região, evidencia-se que existe forte indicativo que nesta região ocorre um ponto que maior compressão de dados em conjunto pouca variação estatística.

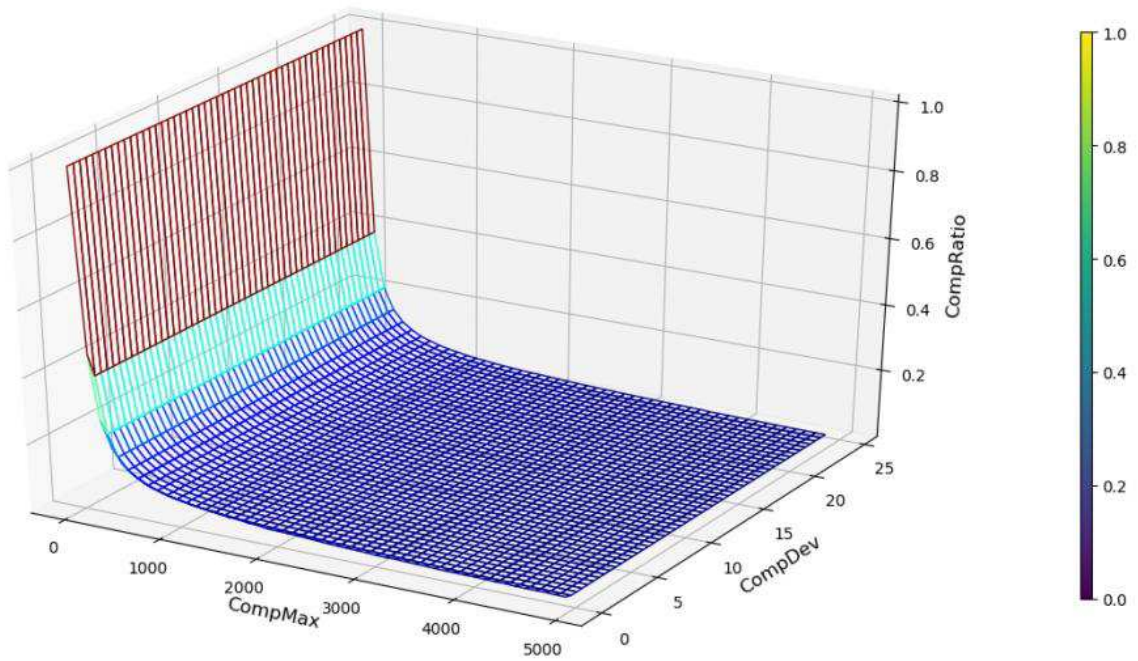
Figura 55 Avaliação do indicador Índice de Correlação para o algoritmo de compressão.



Fonte: Próprio Autor

O último indicador e mais importante para avaliação da compressão é apresentado na Figura 56. Ao verificar o comportamento da razão compressão frente aos valores assumido por *CompDev* e *CompMax*, é possível verificar que na região identificada pelos indicadores anteriores, existe a consolidação do valor máximo de compressão, correspondendo ao valor mais de 80% de compressão dos dados. Este resultado reafirma que os indicadores estatísticos apresentados são capazes de auxiliar na identificação de desta região de configuração de dados, na qual o valor a compressão alcançada mantém a consistência estatística.

Figura 56 Avaliação do indicador Razão de Compressão para o algoritmo de compressão.



Fonte: Próprio Autor

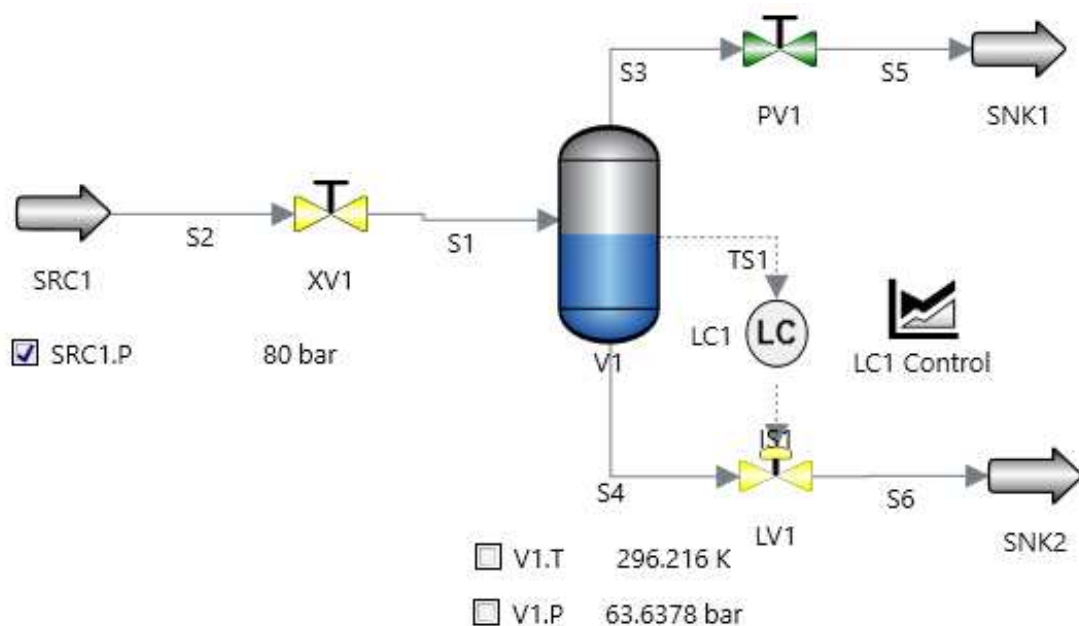
## 6.4. Estudo de caso

O presente estudo de caso visa demonstrar a aplicabilidade da compressão de dados no contexto da engenharia química. O impacto do processo de compressão sobre conclusões do dia a dia também é apresentado, ao se considerar a aplicação de cartas de controle estatístico de processo à malhas de controle de processos químicos.

### 6.4.1. Processo Químico

Neste trabalho, o processo considerado corresponde a um tanque flash multifásico. Os parâmetros de design e operação do sistema foi configurado a partir da documentação padrão da plataforma de simulação Simcentral®. A Figura 57 demonstra os equipamentos e instrumentos envolvidos nessa simulação.

Figura 57 Diagrama de representação da simulação de tanque flash.



Fonte: Próprio Autor

O fluido de entrada é composto por uma mistura de hidrocarbonetos que estabelecem entre um mistura líquido-vapor. A composição inicial da mistura de hidrocarbonetos é apresentada na



Tabela 16 abaixo.

Tabela 16 Composição inicial da mistura de hidrocarbonetos.

Nome completo do componente	Composição Inicial
Metano	1
Etano	0.5
Propano	0.3
n-Butano	0.1

Fonte: Adaptado da documentação do Simcentral

A modelagem termodinâmica da mistura líquido-vapor é realizada aplicando-se o modelo de Soave-Redlich-Kwong (SRK) e para definir as interações binária entre fases utiliza-se os parâmetros disponíveis no banco SIMCI do Simcentral. A Figura 18 apresenta os campos de preenchimento dentro do simulador.

Tabela 17 Modelagem termodinâmica para mistura de hidrocarbonetos.

Fonte: Próprio autor

A condição de entrada do fluido definido anteriormente é configurada no componente *source* SRC1. São consideradas as variáveis pressão (P), temperatura (T) e fluxo mássico (W), como apresentado na Tabela 18.

Tabela 18 Especificações de entrada definidas em SRC1.

Variável	Composição Inicial
W	40000 kg/h
T	27 °C
P	8000 kPa (gauge)

Fonte: Adaptado da documentação do Simcentral.

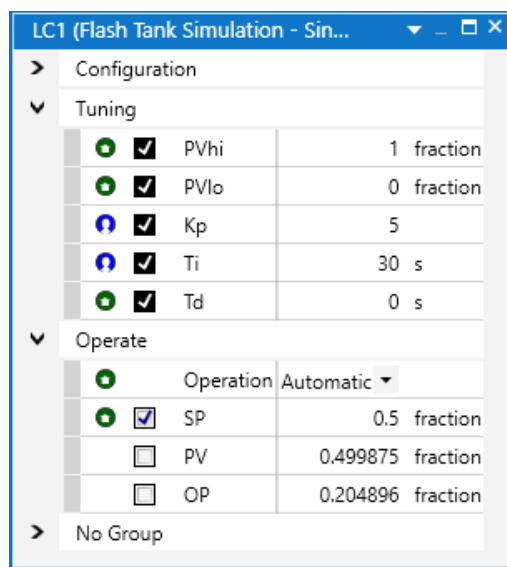
A definição do sistema apresentado segue o fluxo de usual de configuração de uma simulação no SimCentral. Inicialmente resolve-se ambos os balanços de massa e energia no *Process Mode*. Considerando que todas as condições necessárias essas à solução são inseridas de tal forma que temos obtém-se o estado da simulação como especificado e resolvidas, segue-se para o *Fluid Flow model*.

Neste são definidas as condições de escoamento dos fluidos por balanços de pressão, neste caso instrumentos tais como válvulas devem ser especificados. Assim como anteriormente, as informações necessárias são introduzidas, obtendo-se um estado de simulação com especificado e resolvido.

Considerando que a simulação teve suas condições que massa, energia e pressão especificadas, parte-se a verificação da dinâmica de processos. Posicionadas as válvulas de controle, a sintonia do controlador é segundo passo, logo após o sistema alcançar o estado estacionário.

O sistema de controle aplicado ao tanque flash controla a sua fração de nível ao manipular a abertura de válvula. A ação deste controlador é do tipo direta, pois para um distúrbio positivo da fração de nível, a ação realizada sobre a válvula é também positiva, ou seja, aumenta a vazão de saída do tanque por sua abertura ser aumentada. As especificações do controle de nível do tanque são apresentadas na Figura 58. O controle aplicado é do tipo Proporcional e Integral (PI) e tem por *setpoint* a fração de 0,5 para a primeira condição operacional e 0,75 para segunda condição operacional.

Figura 58 Especificações do controle de nível do tanque.



Fonte: Próprio autor.

Antes de passar-se a simulação dos pontos operacionais, é necessário calcular o gradiente de pressão da simulação utilizando-se do estágio *Fluid Flow*. Este é definido pelas condições de contorno de pressão, definidas pelas pressões de fluido na fonte de escoamento, no *source* SCR1 e no *sink* SKN1 e SKN2, onde ocorre a absorção do fluido especificado.

Além das pressões em posições extremas com *sources* e *sinks* há ainda os fluxos intermediários entre estas condições de contorno de pressão que interferem no perfil de pressão do sistema. Estes fluxos sofrem interferência por conta da resistência em seus escoamentos, fatores como o diâmetro e comprimento de tubo, CV de válvulas são exemplos de fatores que devem ser considerados quando se define o perfil de pressão e fluxo em simulações como a do tanque flash. Portanto, ao mudar de modo de simulação *Process* para o *Fluid Flow* algumas variáveis são especificadas, os valores de pressão para os SINKs e CVs de válvulas conforme a Tabela 19. O fluxo de entrada é então calculado, a análise do comportamento da dinâmica do processo é a próxima etapa.

Tabela 19 Especificações de sink e válvulas.

Model na simulação	Variável	Valores
XV1	CV	100
XV1	ManPos	0.5
PV1	CV	75
PV1	ManPos	0.25

LV1	Cv	50
SNK1	P	100 kPa (gauge)
SNK2	P	100 kPa (gauge)

Fonte: Adaptado da documentação do Simcentral.

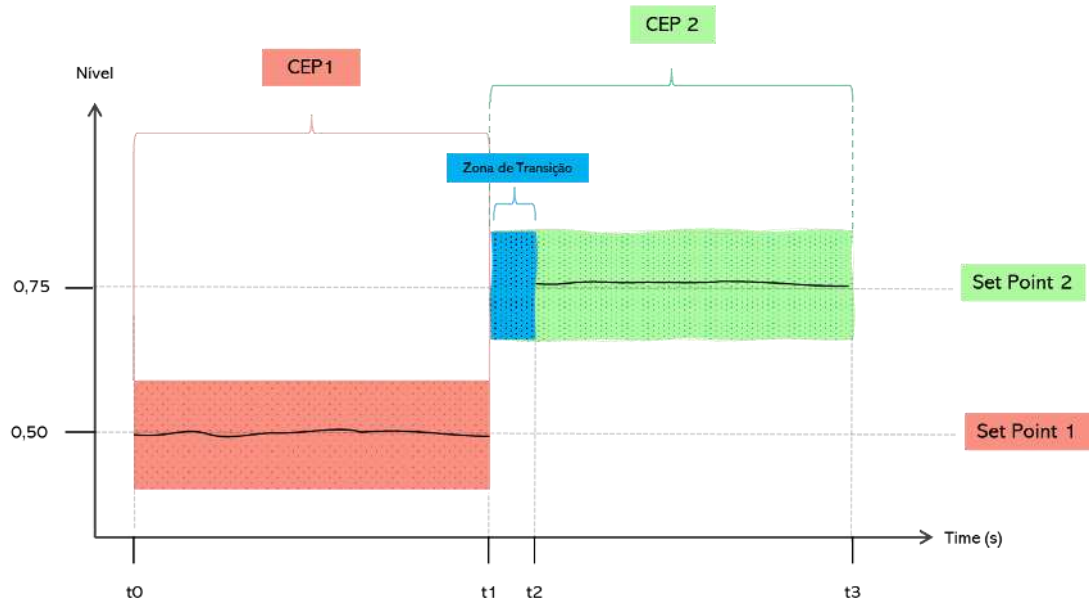
Ao selecionar o modo *Dynamics* no SimCentral, o processo passa a gerar dados em escala temporal para as variáveis de processo, tendo como condições iniciais os valores obtidos pelas etapas anteriores. Neste ponto é possível simular as duas condições operacionais previstas, ambas correspondem a especificação de *setpoints* diferentes entre si para a fração de nível do tanque flash.

O primeiro cenário tem por *setpoint* o nível de 50% do tanque flash. O segundo ponto de operação possui *setpoint* de 75% de nível do mesmo tanque. Entre estes dois pontos operacionais há a chamada zona de transição, que depende do quanto tempo o controlador de nível é capaz de levar o sistema até a nova condição de operação, ou seja, se este é capaz de prevenir *overshoots* na variável controlada ou instabilidade operacional. Esta zona, portanto, está na região entre estados estacionários de processos diferente.

São adicionados ruídos em ambos os cenários do sistema, ao se alterar a posição da válvula de alimentação dos tanques flash em desvio de até 10% da posição de origem de seu ponto de operação normal, correspondente a 80%. Portanto, o processo possui variabilidade suficiente para que o sistema possa ser monitorado através da técnica de controle estatístico de processos, que é aplicável para detecção da presença de causas especiais no sistema.

A Figura 59 demonstra a aplicação do controle estatístico de processos para a fração do nível do tanque pra duas condições operacionais especificadas. Devido a natureza diversa entre ambos os pontos é necessário construir dois pares de cartas de control, um par para cada condição operacional. Para o caso ótimo, onde o controle de nível está bem sintonizado, a zona de transição é mínima ou inexistente, pois o controle levaria o sistema ao ponto operacional rapidamente e sem comprometer a estabilidade operacional.

Figura 59 Cenários de operação do tanque e suas respectivas cartas de controle.



Fonte: Próprio Autor

## **6.5. Controle Estatístico de processos**

A aplicação do controle estatístico de processos visa demonstrar o resultado que pode ser obtido quando o processo de compressão é aplicado a dados oriundos de processos químicos. Ao se aplicar o processo de compressão de dados em faixa de configurações possíveis, obtendo-se para cada configuração uma taxa de compressão específica, portanto, pode-se verificar o quão importante é definir uma estratégia de configuração deste processo.

Como instrumento para avaliação em processos químicos, o controle estatístico de processos tem aplicação clássica e consolidada, justificando sua aplicação como forma de evidenciar e constatar possíveis distorções sobre dados fornecido para os tomadores de decisão. Devido a seu base estatística as distorções podem ser percebidas da alteração dispersão dos pontos presentes em cartas de controle, alteração da localização dos pontos em relação aos limites de controle ou até mesmo.

Toma-se com o adequado que a fase 1 do monitoramento, que corresponde a construção das cartas de controle, seja realizada a partir de dados brutos, ou seja, sem que o processo de compressão de dados seja aplicado. Na fase do monitoramento, os dados reconstruídos oriundos do processo de compressão serão comparados aos dados brutos. Por fim, métricas estatísticas são avaliadas as cartas de controle de processo aplicadas a malha de controle selecionada.

### 6.5.1. Cartas de controle a partir de série de dados temporais

Os dados industriais de controle de processos e instrumentação são classificados como dados de séries temporais. Para estes dados cada valor está associado um dado estado ou medida em um momento específico, ou seja, uma medida possui um por si só duas dimensões, o valor e o tempo no qual este valor está atribuído.

A construção das cartas de controle a partir de dados de séries temporal deve ser estruturada de maneira específica. A Figura 60 auxilia o entendimento do processo de construção das cartas de controle estatístico de processos. Seguindo-se os índices desta figura, tem-se as seguintes definições:

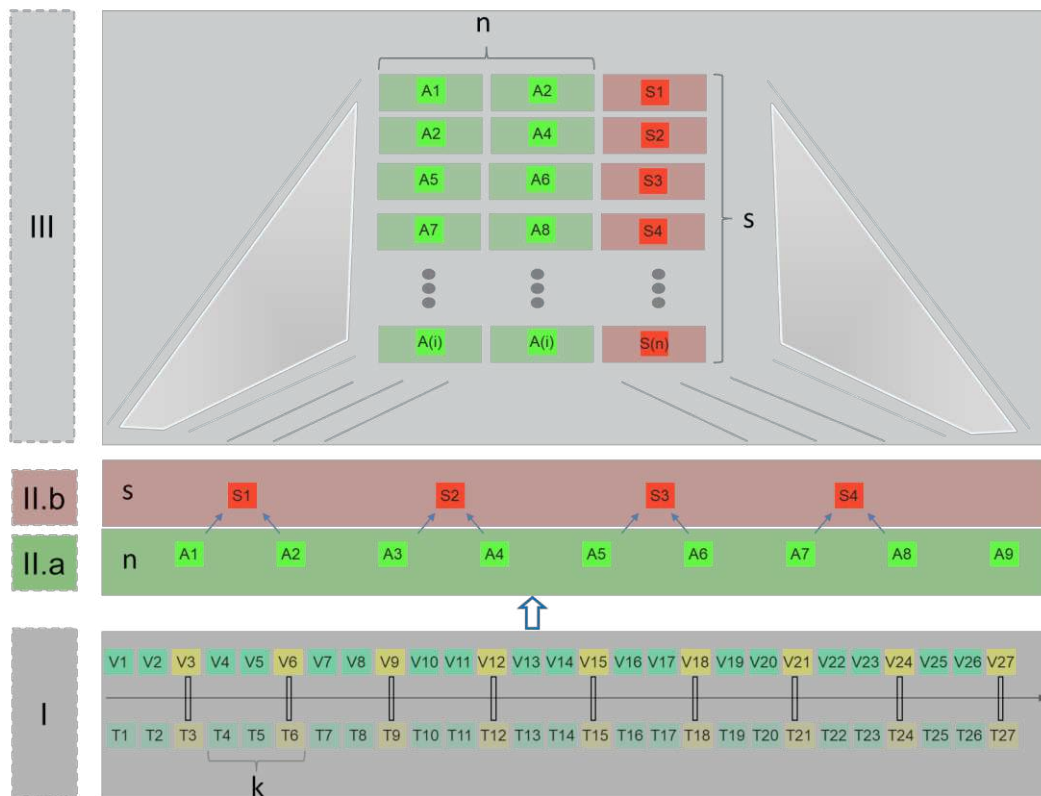
1. *Séries de dados temporais:* As séries carregam dados de tempo e valor de cada tag originados pelo processo de coleta de dados. O valor  $k$  corresponde ao intervalo de coleta de pontos a serem utilizados no processo de construção dos limites controle, os demais pontos não contribuem neste processo;

2. *Construção das amostras:* As amostras correspondem aos pontos coletados. Na seção II.a estes pontos são denominados como  $A(i)$ , sendo  $n$  o total de amostras por subgrupo  $s(i)$ . Na seção II.b Cada subgrupo possui o mesmo número de amostras  $n$ , amostra podem ser descartadas caso a quantidade de subgrupos necessários já tenham sido alcançadas;

3. *Dados para construção dos Limites de controle:* as amostras e seus respectivos subgrupos são estabelecidos como conjunto de dados finais, e disponíveis para a construção dos limites de controle.

Portanto, a construção dos limites de controle deve considerar a definição destes três parâmetros de construção de conjunto de dados finais que correspondem  $k$ ,  $n$  e  $s$ , intervalo de tempo de coleta entre pontos, quantidade de amostras por subgrupo e quantidade de subgrupos.

Figura 60 Dados para construção dos limites de controle.



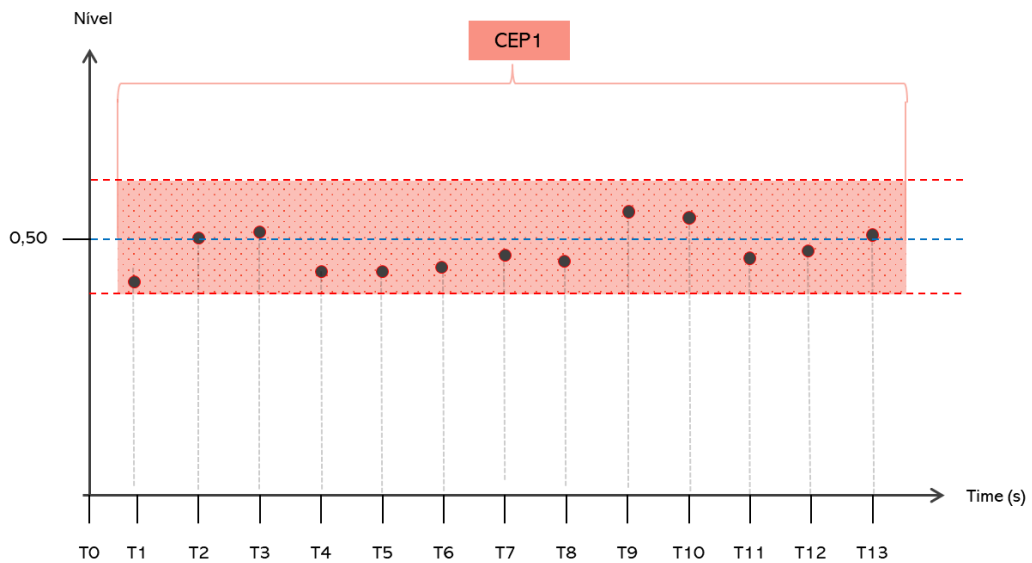
Fonte: Próprio Autor

Para o problema apresentado, a construção das cartas de controle se dá a partir de uma parte dos dados obtidos em cada cenário apresentado. Portanto, são definidos os parâmetros  $k$ ,  $s$  e  $n$  para necessários a esta tarefa, considerando que a simulação está sob efeito unicamente de causas comuns ao processo, ou seja, ruídos que tem amplitudes de na ordem de 5% dados apresentados. Na realidade de chão de fábrica, as cartas de controle são definidas previamente, através de dados coletados em momentos que há evidências da inexistência de causas especiais no processo monitorado.

Portanto, a Figura 61 demonstra a seleção final de 13 amostras, direcionados para a construção das cartas de controle para o cenário 1. Na área avermelhada, as amostras utilizadas na construção são apresentadas, caso alguns destas amostras estejam acima ou abaixo dos limites superiores e inferiores de controle, respectivamente, é necessária a reconstrução da carta, desconsiderando a amostra fora do padrão.



Figura 61 Construção de carta de controle para o cenário 1.



Fonte: Próprio Autor

### 6.5.2. Cartas de controle para monitoramento de nível de tanque

Com citado anteriormente, os dados coletados a partir da simulação da Plataforma SimCentral foram segmentados em 3 zonas.

A Figura 62 apresenta a variável controlada do tanque flash, o nível. Nesta são apresentadas tanto o valor do *setpoint* (em laranja) quanto a variável de processo (em azul), as quais são apresentadas em 4 representações: (i) todo o perfil do nível é plotado em gráfico que aborda todas as zonas discutidas na

Figura 59; (ii) a primeira zona corresponde os dados do cenário 1; (ii) a segunda aos dados da zona de transição e (iii) a zona terceira aos dados do cenário 2.

Figura 62 Dados brutos de simulação de processo de flasheamento.

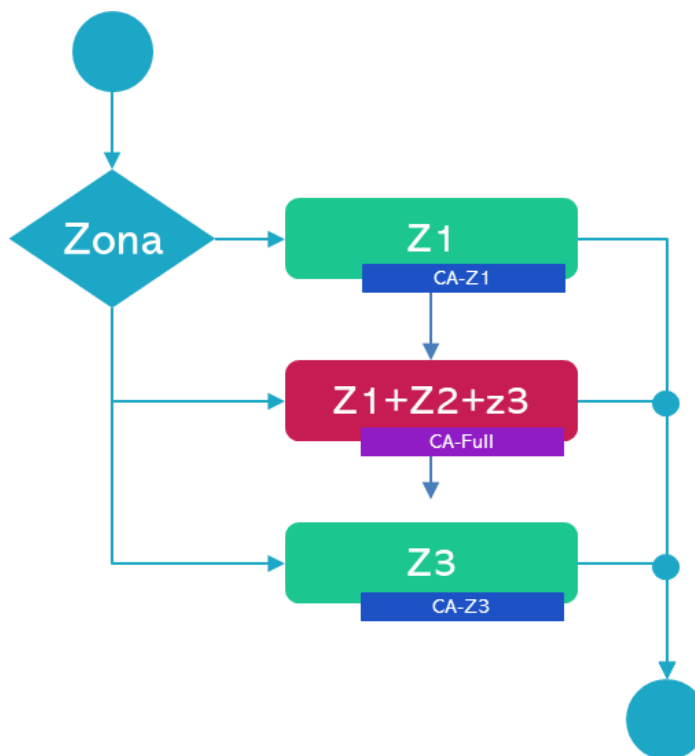


Fonte: Próprio Autor

A análise aqui proposta, demonstra que o controle do sistema não está adequado a transição do primeiro estado estacionário para o segundo estado estacionário. Esta conclusão é evidenciada pela existência do *overshoot*, ocasionando também o aumento do tempo necessário para que o sistema alcance o segundo estado estacionário.

Esta característica dos dados reflete uma condição indesejada no sistema físico, porém, importante para a avaliação do processo de compressão de dados, pois este comportamento pode vir a ser omitido, caso os parâmetros de compressão de dados sejam configurados incorretamente. Portanto, a análise seguirá em três fases, como demonstrado na Figura 63.

Figura 63 Estruturação de avaliação da compressão considerando as fases de análise.



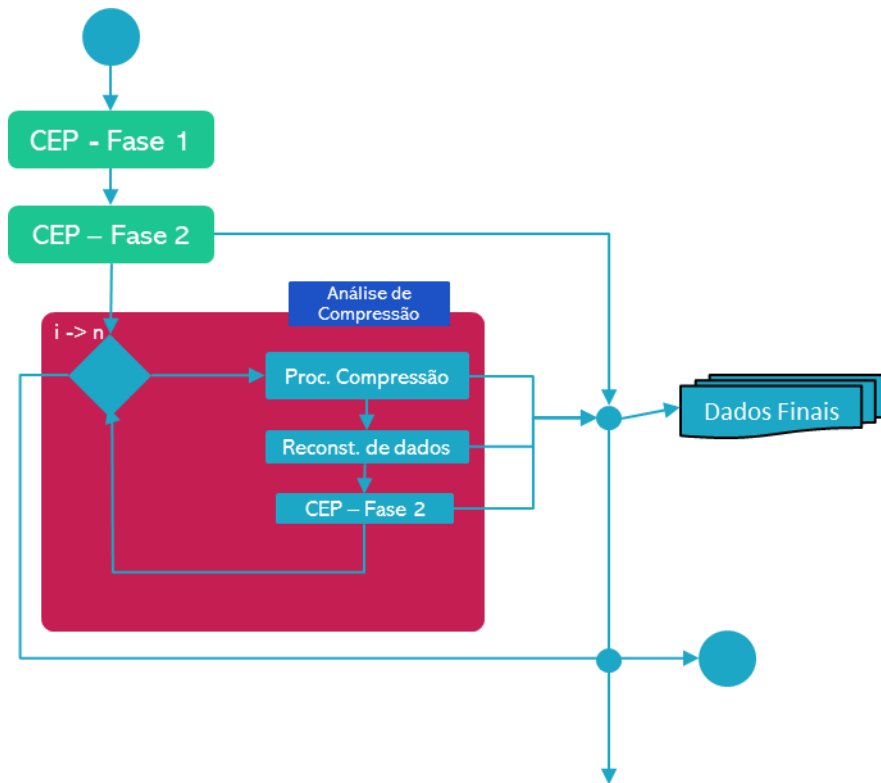
Fonte: Próprio Autor

O bloco CA-Full corresponde a avaliação do processo de compressão de dados, considerando a combinação de todas as zonas. Esta abordagem visa verificar o impacto do processo de compressão sobre a região de transição de estados estacionários. Para o caso ótimo, os dados recuperados devem manter o perfil final o mais próximo possível do perfil original, antes do processo de compressão de dados. Os blocos CA-Z1 e CA-Z3 correspondem a análise de compressão realizada respectivamente as zonas Z1 e Z3. Nestes blocos são aplicados a construção das cartas de controle estatístico.

A Figura 64 apresenta diagrama de avaliação adotado para as zonas Z1 e Z3. É considerada a entrada de dados brutos e saída de uma lista de gráficos e dados utilizados em avaliações estatísticas. São construídos os gráficos de CEP unicamente com dados brutos (CEP-Fase 1), os dados comprimidos e reconstruídos são aplicados aos limites de controle definidos a partir dos dados

brutos. Por fim são colhidos resultados de dados brutos, comprimidos, reconstruídos e dados das amostras plotados CEP.

Figura 64 Avaliação do processo de compressão de dados nas regiões Z1 e Z3.



Fonte: Próprio Autor.

Para avaliação das cartas de controle de processos do tipo Xbarra-R são aplicados os parâmetros apresentados na Tabela 20.

Tabela 20 Parâmetros de construção de cartas XBarra-R.

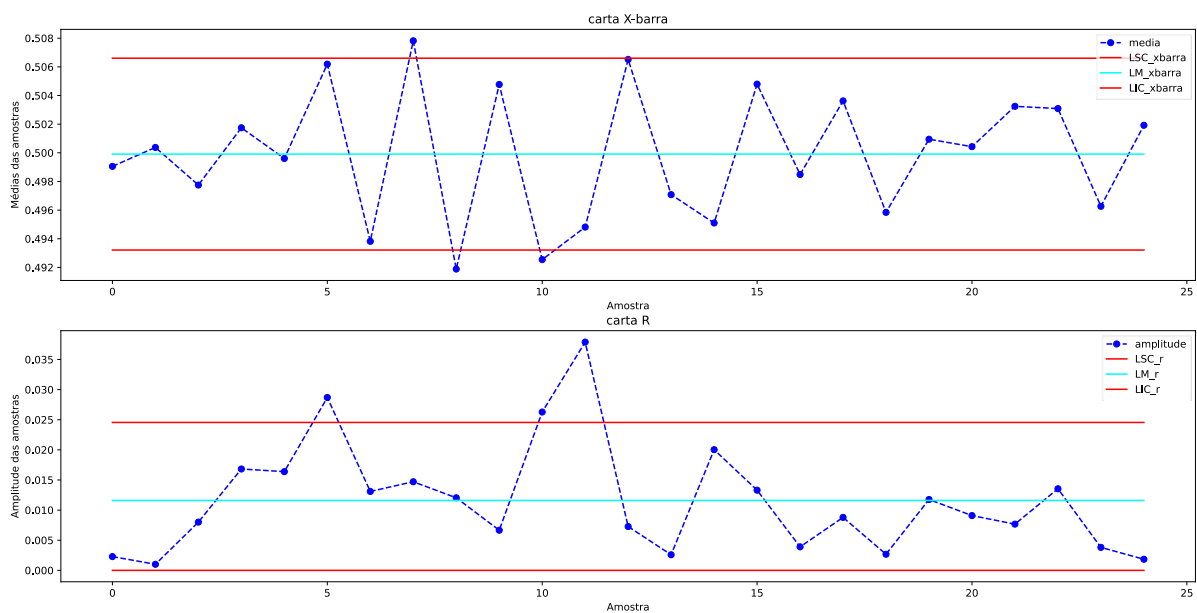
Parâmetros	Valores
k	10
n	5
s	25

Fonte: Próprio autor

Com estes parâmetros de construção de carta de controle e com dados brutos de nível, foram definidos os limites de controle para os dois estados estacionários.

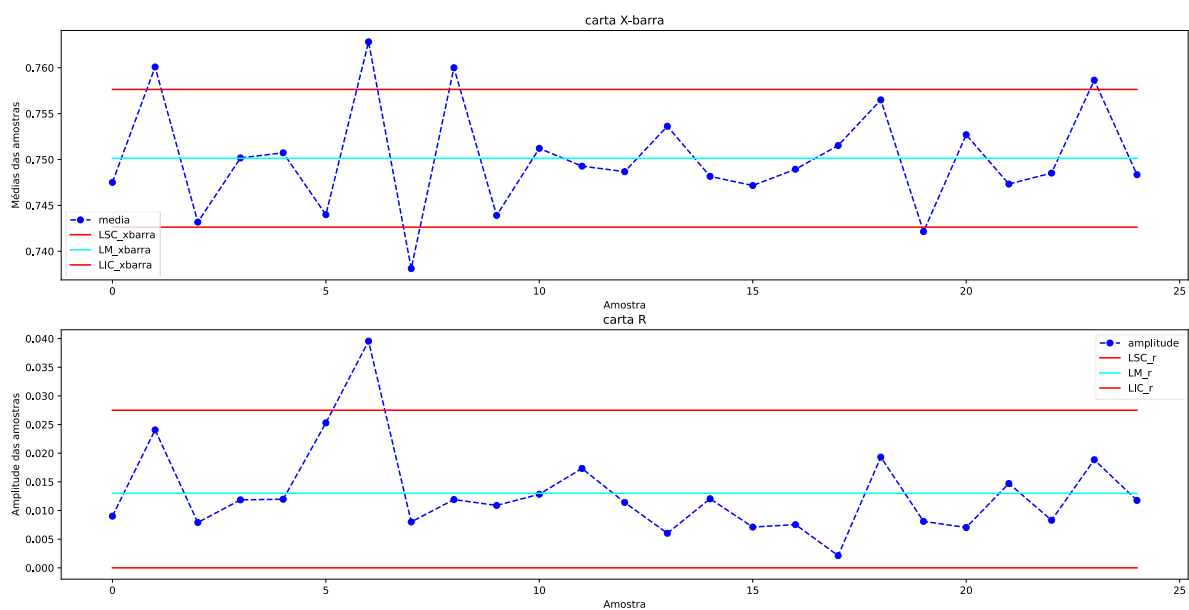
São demonstrados dos dois limites de controle nas Figura 65 e Figura 66, correspondendo aos estados estacionários de 1 e 2.

Figura 65 Carta X-BarraR para o estado estacionário 1 (Zona 1).



Fonte: Próprio autor

Figura 66 Carta X-BarraR para o estado estacionário 2 (Zona 3).



Fonte: Próprio autor

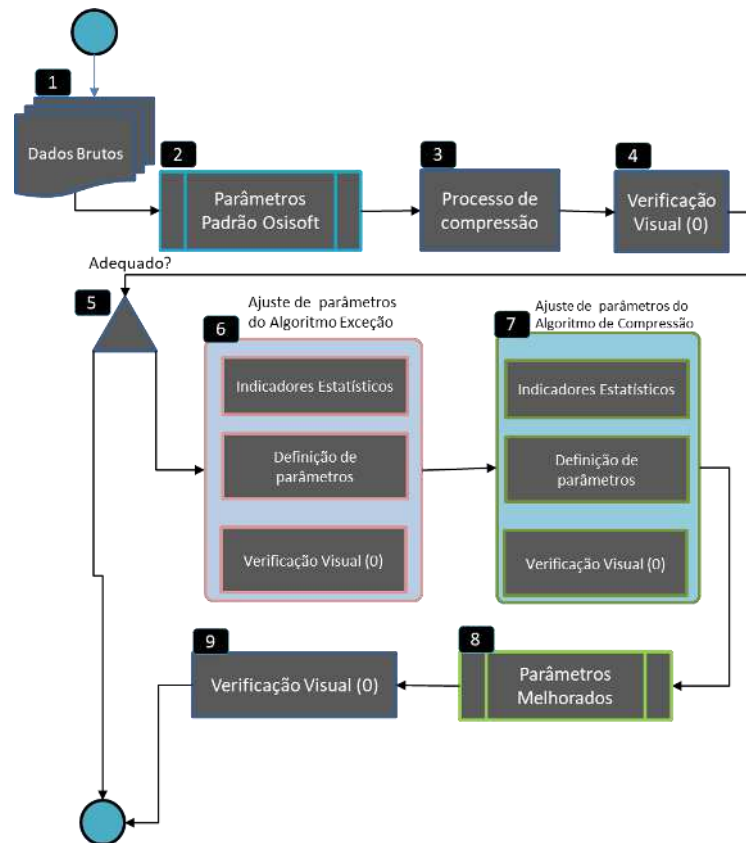


## **6.6. Avaliação dos parâmetros de exceção e compressão**

Para avaliação dos parâmetros de exceção e compressão de dados, foi adotado o processo apresentado na Figura 67 abaixo. A ingestão dos dados brutos acontece como primeiro processamento (1). Ocorre a definição dos parâmetros (2) dos algoritmos de exceção e compressão de dados de acordo com os parâmetros padrão sugeridos pela Osisof e sua execução em seguida (3), considerando tags do tipo dupla precisão (Float32 ou Float64).

A visualização dos resultados do processo de compressão de dados é realizada, de maneira a verificar se há a manutenção do perfil dos dados brutos (4). Caso o perfil identificado nos dados reconstruídos seja semelhante aos dados brutos segue ao fim do processo, mantendo-se os parâmetros padrão como os parâmetros adequados aos dados. Para o caso o processo de compressão altere o perfil dos dados reconstruídos, segue-se para as etapas de ajuste dos parâmetros de exceção (6) e compressão (7) de dados. Na etapa 8 os parâmetros ajustados são aplicados como atuais para a tag em avaliação. A verificação visual (9) é etapa final, na qual são apresentados os dados reconstruídos.

Figura 67 Fluxo de avaliação do processo de compressão de dados.



Fonte: Próprio Autor

Para a avaliação do impacto do processo de compressão de dados sob as condições de operação do tanque de nível, os dados dos coletados do nível do tanque foram coletados e submetidos ao processo de compressão, considerando ambos os algoritmos de exceção e compressão de dados.

Os parâmetros listados são os valores padrão fornecidos pelo PI System no momento de criação da tag do tipo *Double*, equivalente ao *Float64* no PI Data Archive, cuja amplitude está em  $\pm 5.0e-324$  to  $\pm 1.7e308$ , com 15 dígitos de precisão. Este tipo de tag utilizada 64-bit para armazenamento de cada um de seus pontos numéricos. Abaixo a Figura X apresenta a região onde o *PI System Management Tools (SMT)* disponibiliza a configuração dos parâmetros de compressão e exceção por parte do usuário, assim com os valores padrão que são endereçados a tag recém-criada.



Figura 68 Valores padrão da compressão e exceção fornecidos pelo PI System

The screenshot shows the 'System' tab of a configuration window. At the top, there are tabs for 'General', 'Archive', 'Classic', 'Security', and 'System'. The 'System' tab is active. Below the tabs, there are several input fields and radio button groups. The 'Typical Value' is set to 50, 'Zero' is 0, and 'Span' is 100. There are five radio button groups: 'Scan' (On), 'Archiving' (On), 'Step' (Off), 'Shutdown' (Off), and 'Compressing' (On). Below these are two 'Exception Deviation' and 'Compression Deviation' sections. Each has a value field (0.1 and 0.2 respectively) and a unit dropdown set to 'Eng. Units'. Below each deviation section is a time selection table with columns for Day, Hr, Min, and Sec. The 'Exception Deviation' table has values 0, 0, 10, 0. The 'Compression Deviation' table has values 0, 8, 0, 0.

Day	Hr	Min	Sec
0	0	0	0
0	0	10	0

Day	Hr	Min	Sec
0	0	0	0
0	8	0	0

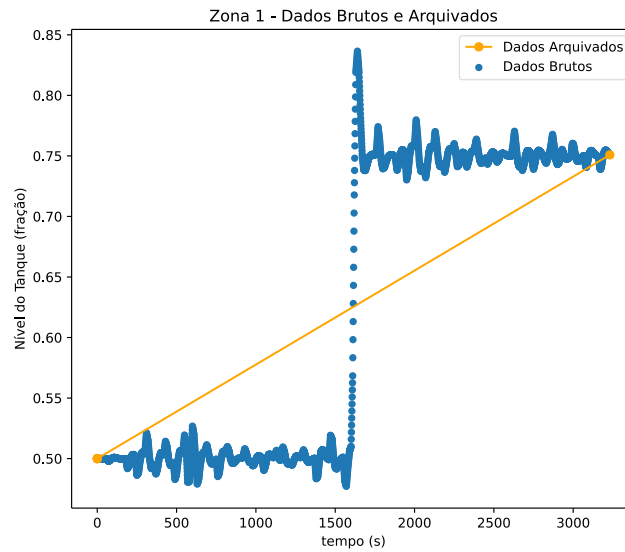
Fonte: Adaptado de Osisof®

### 6.7. Processo de Compressão de Dados em todo o intervalo de dados

Aqui foi o processo de compressão de dados é realizado de acordo com os parâmetros definidos como padrão pela osisoft. Estes são definidos assim que a criação de uma nova tag é realizada, como demonstrados na Figura 68. A avaliação do processo de compressão segue o fluxograma admitido na Figura 67. Este tópico cobre os pontos de (1) até (5) do referido fluxograma.

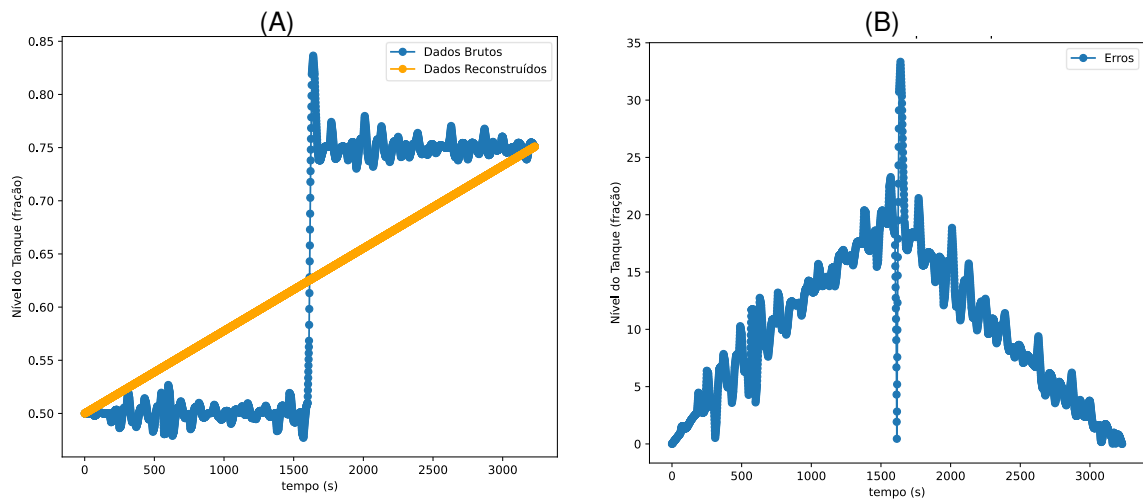
A seguir são apresentados os resultados do processo de compressão de dados. Nestas são apresentados os resultados do processo de compressão com os dados padrão destacando-se dados brutos vs dados arquivados (Figura 69), dados brutos vs dados recuperados (Figura 70.A) e por fim os erros associados a reconstrução de dados (Figura 70.B), realizado por interpolação linear entre os dados arquivados.

Figura 69 Processo de compressão de dados com completo com parâmetros padrão – Dados arquivados.



Fonte: Próprio Autor

Figura 70 Processo de compressão de dados com completo com parâmetros padrão – Dados reconstruídos (A) e Erros individuais de cada valor reconstruído (B).



Fonte: Próprio Autor

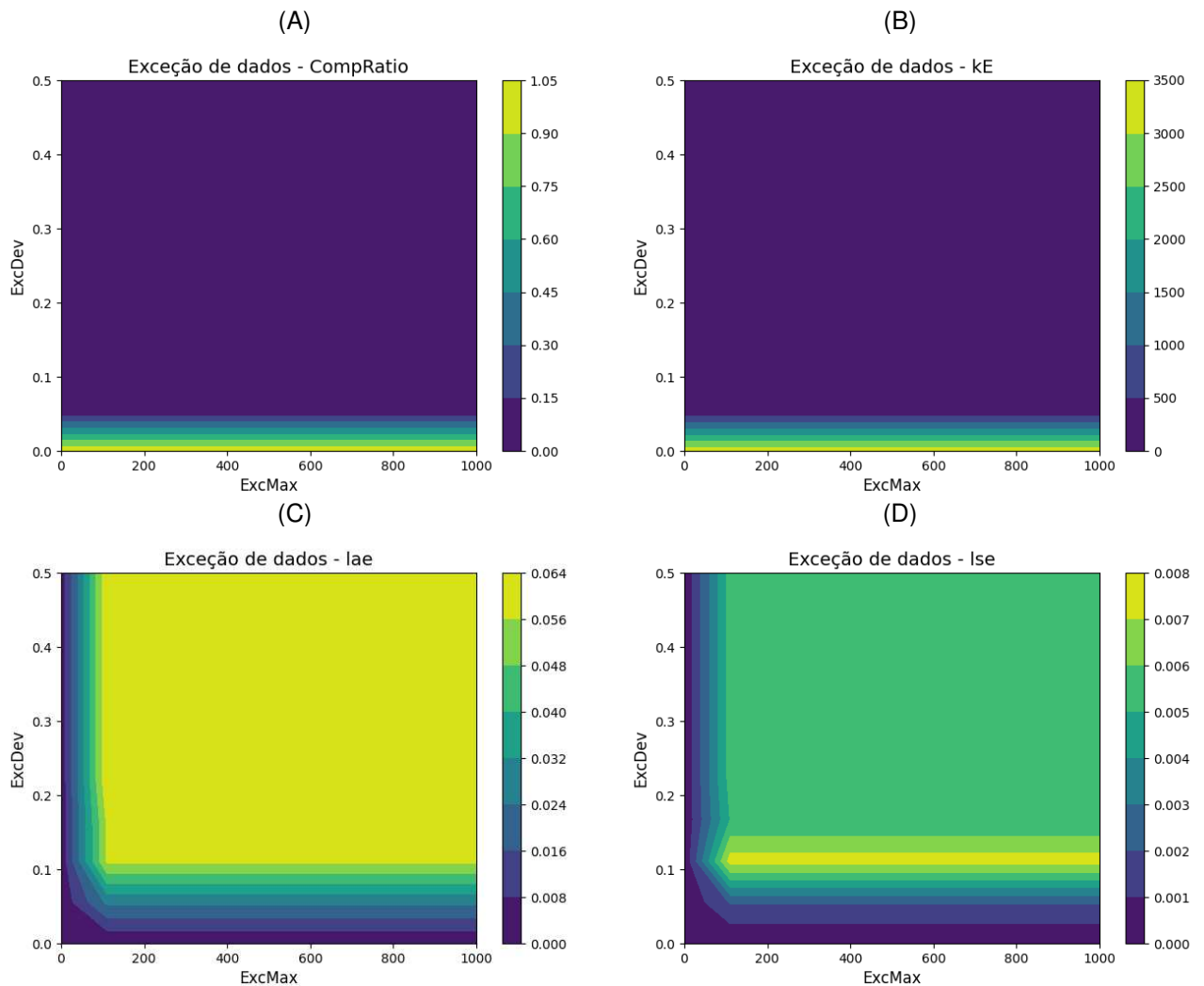
É possível perceber que os resultados alcançados são inadequados com os parâmetros padrão, para ambos os algoritmos de exceção e compressão. Os dados reconstruídos não mantiveram o perfil presente nos dados originais. Portanto, aqui verificamos que a necessidade de realizar os ajustes dos parâmetros de exceção e compressão de dados. Tomando-se como base o diagrama da Figura 67, passamos para os passos (6) e (7).

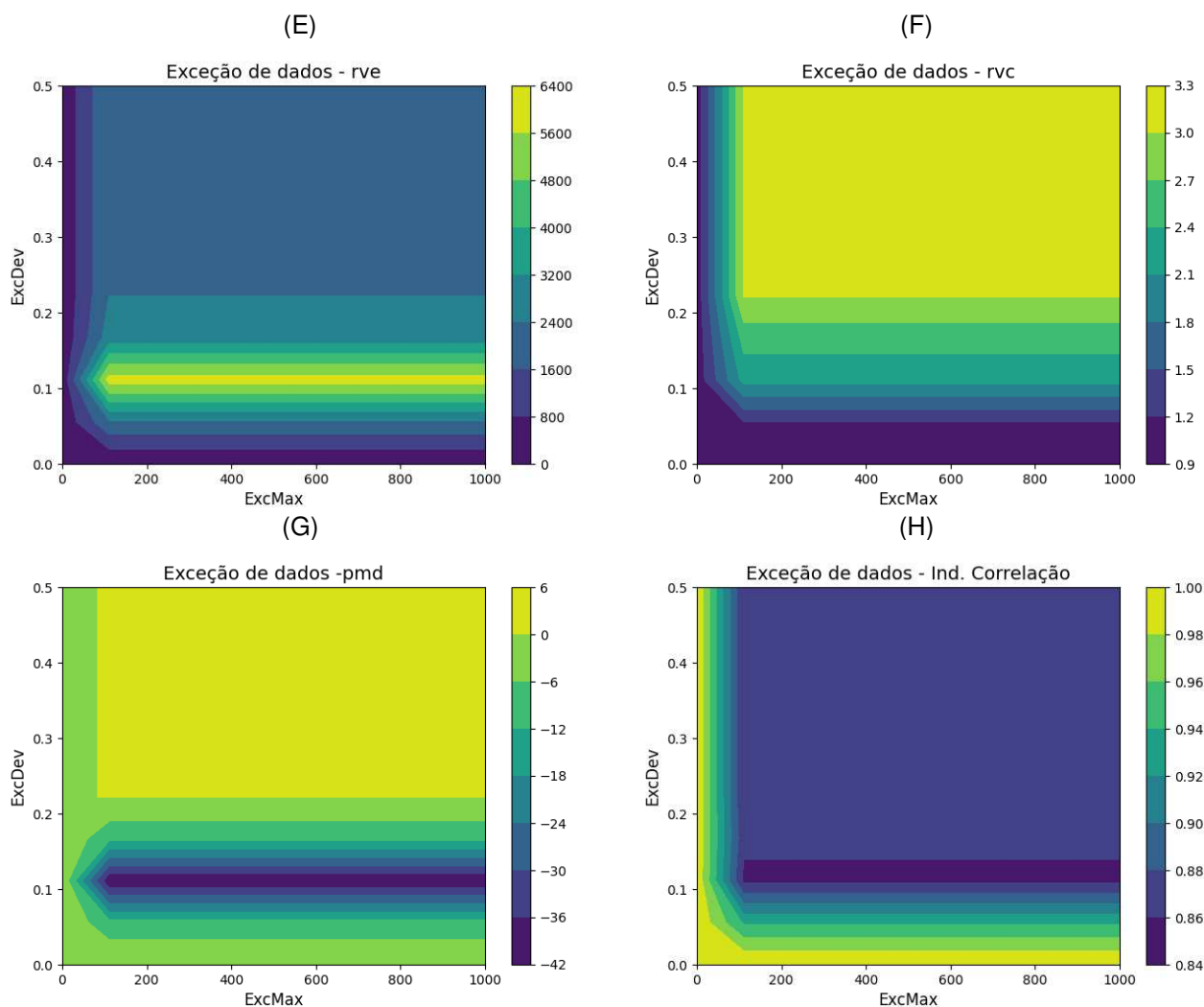
## 6.8. Avaliação dos parâmetros do algoritmo de exceção

A definição dos parâmetros de exceção inicia-se ao avaliar os indicadores estatísticos a partir dos dados brutos. Estas visualizações fornecem o intervalo possível para se encontrar o par de parâmetros mais adequados do algoritmo de exceção, *ExcMax* e *ExcDev* neste caso.

Na Figura 71 são apresentados os oito gráficos que demonstram os perfis dos indicadores estatísticos com base nas duas variáveis em análise, com intervalos entre [0 – 0,5] para *ExcDev* e [0 – 100] para *ExcMax*. Cada um com contribuição na definição dos valores dos parâmetros do algoritmo.

Figura 71 Indicadores estatísticos para avaliação dos parâmetros do algoritmo de exceção de dados.





Fonte: Próprio Autor

Em Figura 71.A e Figura 71.B evidencia-se que há uma região em relação a *ExcDev* no qual a quantidade de ponto não é mais reduzidas, entre [0- 0,05], não havendo mudança em seu valor independente deis deste limite. Por outro lado, nada ocorre em relação a *ExcMax*, este não exercendo influência nestes indicadores. Nas Figura 71.C e Figura 71.D demonstram-se que ambos os parâmetros exercem a influência sobre as somas dos erros médio absoluto e quadrado, respectivamente. Demonstrando o intervalo de influência está entre [0-0,05] para *ExcDev* e [0-100] para *ExcMax*. Nas Figura 71.E e Figura 71.F contribuem com a verificação razão entre variâncias e entre erro e variância dos dados. Novamente é possível definir uma faixa de interesse na qual há variação destes indicadores: entre [0-0,05] para *ExcDev* e [0-100] para *ExcMax*. Na Figura 71.G a alteração na média dos dados é evidenciada, o intervalo de interesse fica em [0-0,1] para *ExcDev* e indefinido para *ExcMax*. Na Figura

71.H considera-se a relação linear entre os dados através do índice de correlação de Pearson, o intervalo de interesse fica em [0-0,08] para *ExcDev* e [0-110] para *ExcMax*.

Como resultado destas avaliações, tem-se o intervalo no qual os parâmetros de exceção adequados podem ser encontrados. Todas estas verificações são listadas na Tabela 21 abaixo.

Tabela 21 Verificação dos intervalos de interesse dos indicadores estatísticos para algoritmo de exceção.

Indicador	Figura 65	ExcMax	ExcDev	Indicador	Figura 65	ExcMax	ExcDev
CompRatio	A	-	[0 - 0,05]	RVE	E	[0 - 100]	[0 - 0,05]
KE	B	-	[0 - 0,05]	RVC	F	[0 - 100]	[0 - 0,05]
MAE	C	[0 - 100]	[0 - 0,05]	PMD	G	-	[0 - 0,10]
MSE	D	[0 - 100]	[0 - 0,05]	Ind. Corre.	H	[0 - 110]	[0 - 0,08]

Fonte: Próprio Autor

Os intervalos que englobam por todos os indicadores são o seguinte: [0-0,1] para *ExcDev* e [0-100] para *ExcMax*, em incrementos de 0,05 e 10 respectivamente. A Tabela 22 a seguir, apresenta-se as avaliações realizadas para cada par de configuração de *ExcDev* e *ExcMax* e os respectivos resultados de CompRatio, MSE as figuras nas quais os dados são plotados, das Figura 71 a Figura 76.

Tabela 22 Resultados da exceção de dados, apresentação de resultados para os intervalos estabelecidos.

Figura	ExcDev	ExcMax	CompRatio	Erro de Reconstrução (MSE)
Figura 71	0.05	10	19.9257%	9.600000E-05
Figura 72	0.05	20	9.9628%	1.430000E-03
Figura 74	0.05	40	5.0123%	4.517000E-03
Figura 75	0.1	20	10.0247%	1.430000E-03
Figura 76	0.1	40	5.012%	4.517000E-03

Fonte: Próprio Autor

Para cada figura entre Figura 71 a Figura 76, foram traçadas três avaliações gráficas, (1) dados brutos versus dados arquivados (2) dados brutos reconstruídos e (3) erros associados. A seleção do par de parâmetros é realizada ao ver verificar o MSE.

Para cada figura são apresentados os erros de reconstrução ponto a ponto, como forma de auxiliar a análise da distorção do algoritmo de exceção sobre a dinâmica dos dados, considerando somente os dois estados estacionários. É verificado que para menor MSE da Tabela 22 a maior *CompRatio* é alcançada, porém com resultado suficiente, representado por 80% redução dos pontos de originais.

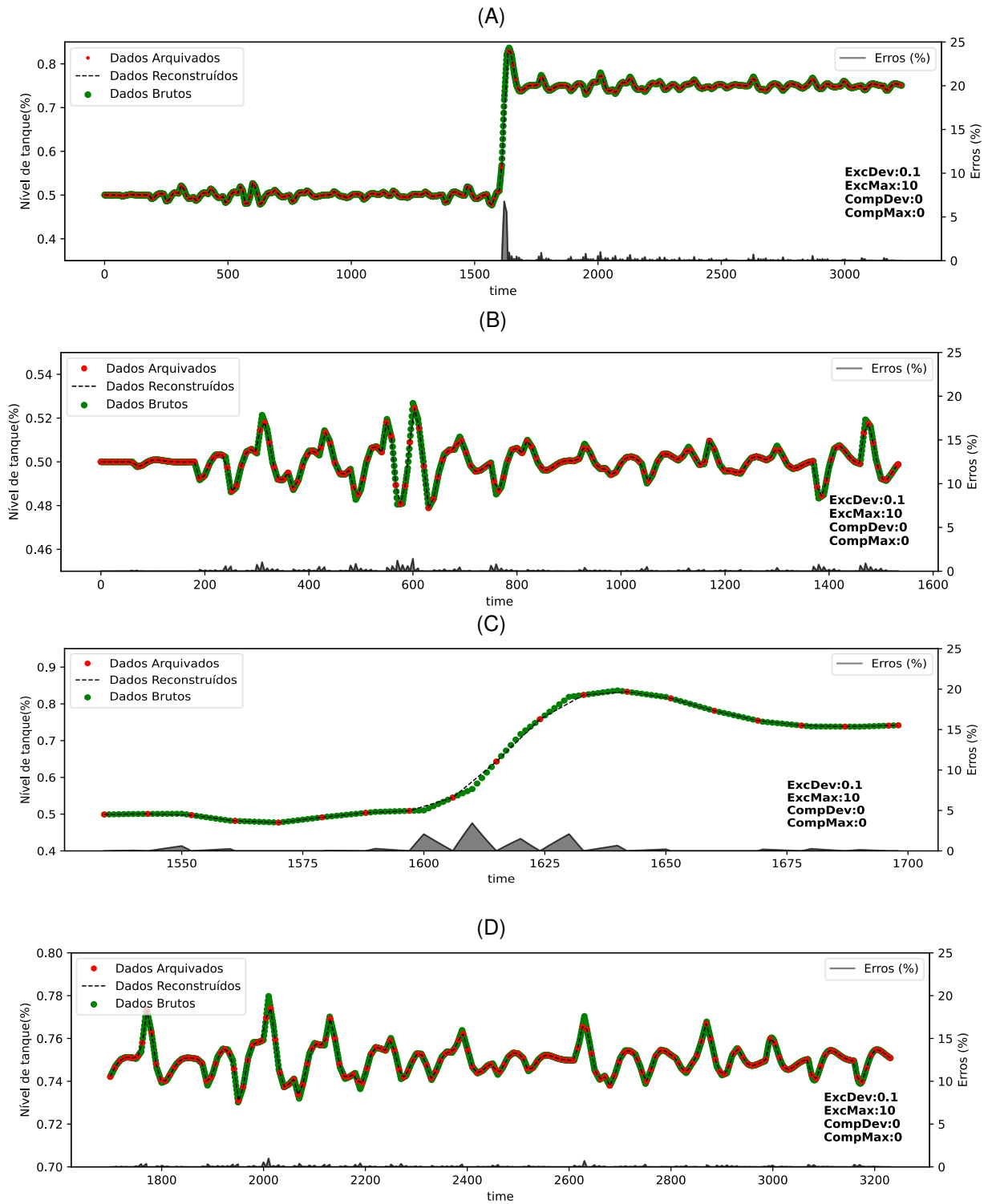
Na parte C das figuras no intervalo de Figura 71 até Figura 76, é percebido que a dinâmica do processo é alterada à medida que o erro aumenta. Este resultado demonstra a necessidade de manter-se o erro o menor possível, mesmo que a *CompRatio* não seja a máxima. Caso esta premissa não seja seguida para o algoritmo de exceção, parte considerável da dinâmica de processo é perdida, como fica evidente na Figura 76. A dinâmica de processo considerada é a estabelecida pela mudança de estado estacionário devido a mudança de *setpoint* do sistema de controle de nível.

É fácil perceber que o sistema controle não se encontra sintonizado adequadamente, devido ao *overshoot* presente na mudança de *setpoint* de 0,5 para 0,75 de nível. Este comportamento deve ser mantido no histórico na tag de armazenamento dos dados, pois a consulta posterior dos dados armazenados possibilitará auditar os resultados de performance do sistema de controle em execução.

Outro fator importante a se considerar é que a exceção exerce prioritariamente o papel de remoção de ruídos presentes nos dados de coleta, portanto, não é exigido um valor alto de *CompRatio* para este algoritmo. Pelo contrário, é mandatário obter uma redução atenuada da quantidade de pontos objetivando a manutenção de dados que venham ser determinantes na manutenção do perfil dinâmico da variável.

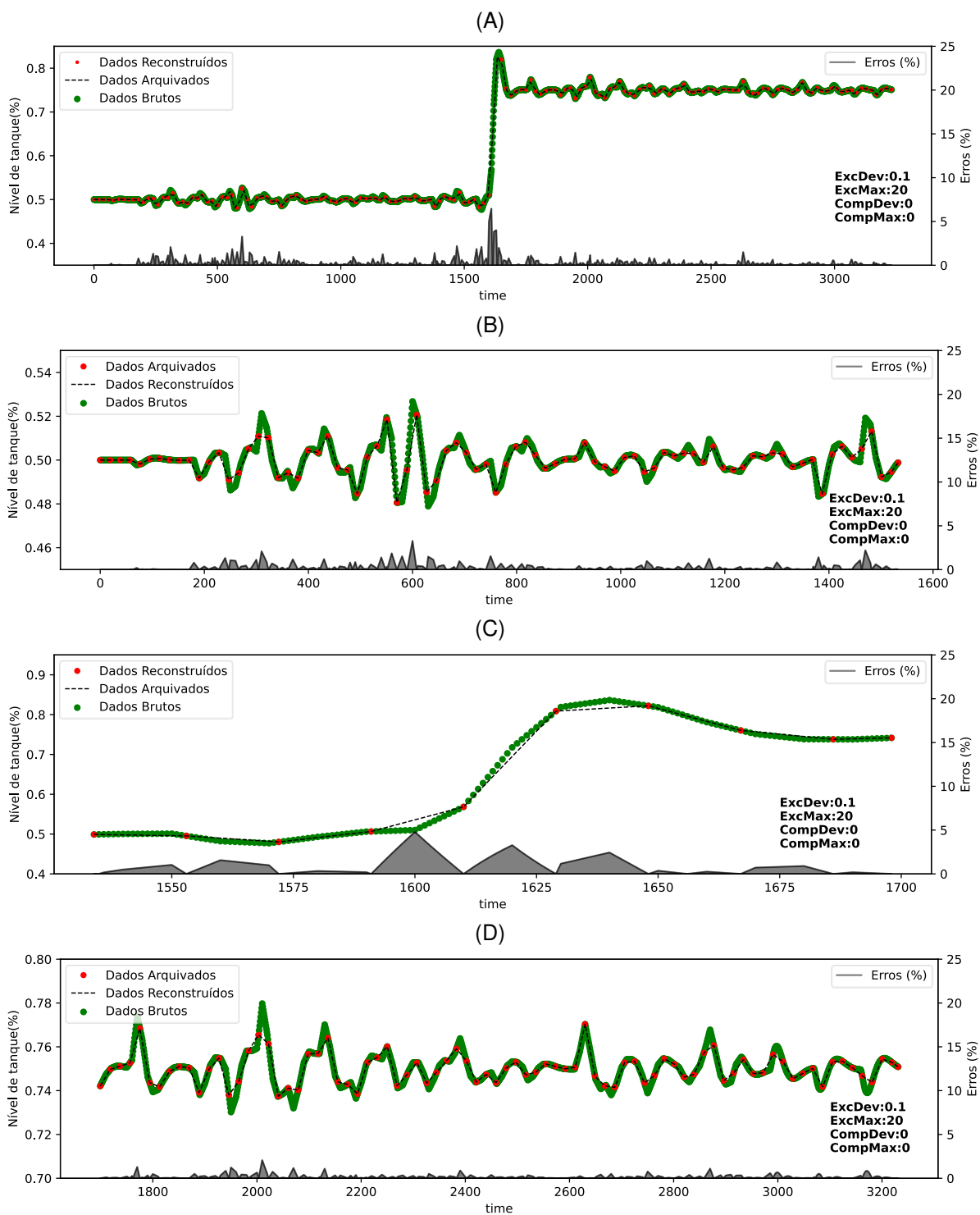
Portanto, o par de pontos escolhido é o apresentado na Figura 71. Para o par de *ExcMax* e *ExcDev* iguais a 10 e 0,1, respectivamente, alcança-se resultado suficiente seguindo-se as premissas estabelecidas para esta fase do processo de compressão de dados do PI System.

Figura 72 Avaliação do Algoritmo de exceção: ExcDev= 0.1, ExcMax=10.



Fonte: Próprio Autor.

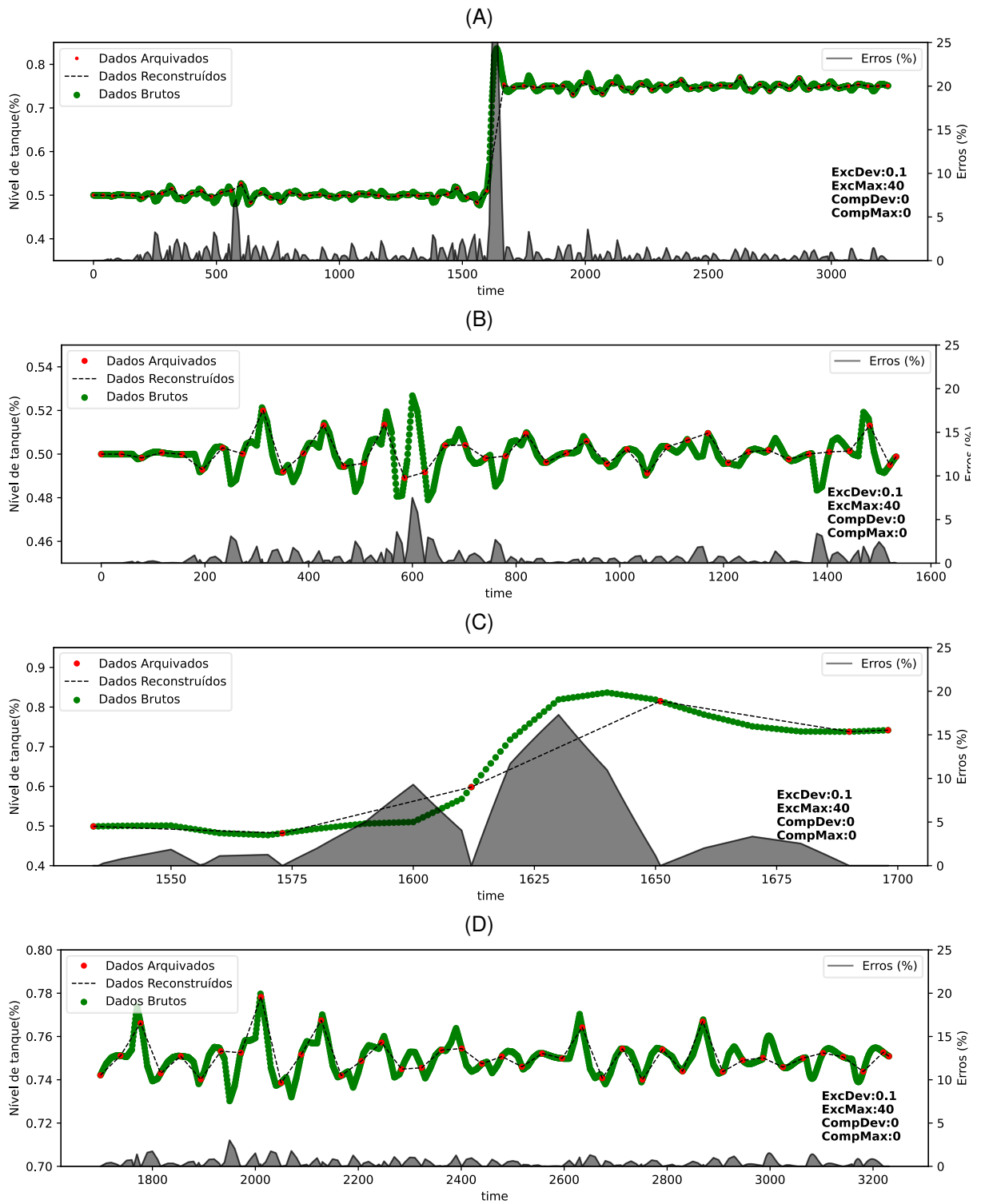
Figura 73 Avaliação do Algoritmo de exceção: ExcDev= 0.1, ExcMax=20.



Fonte: Próprio Autor

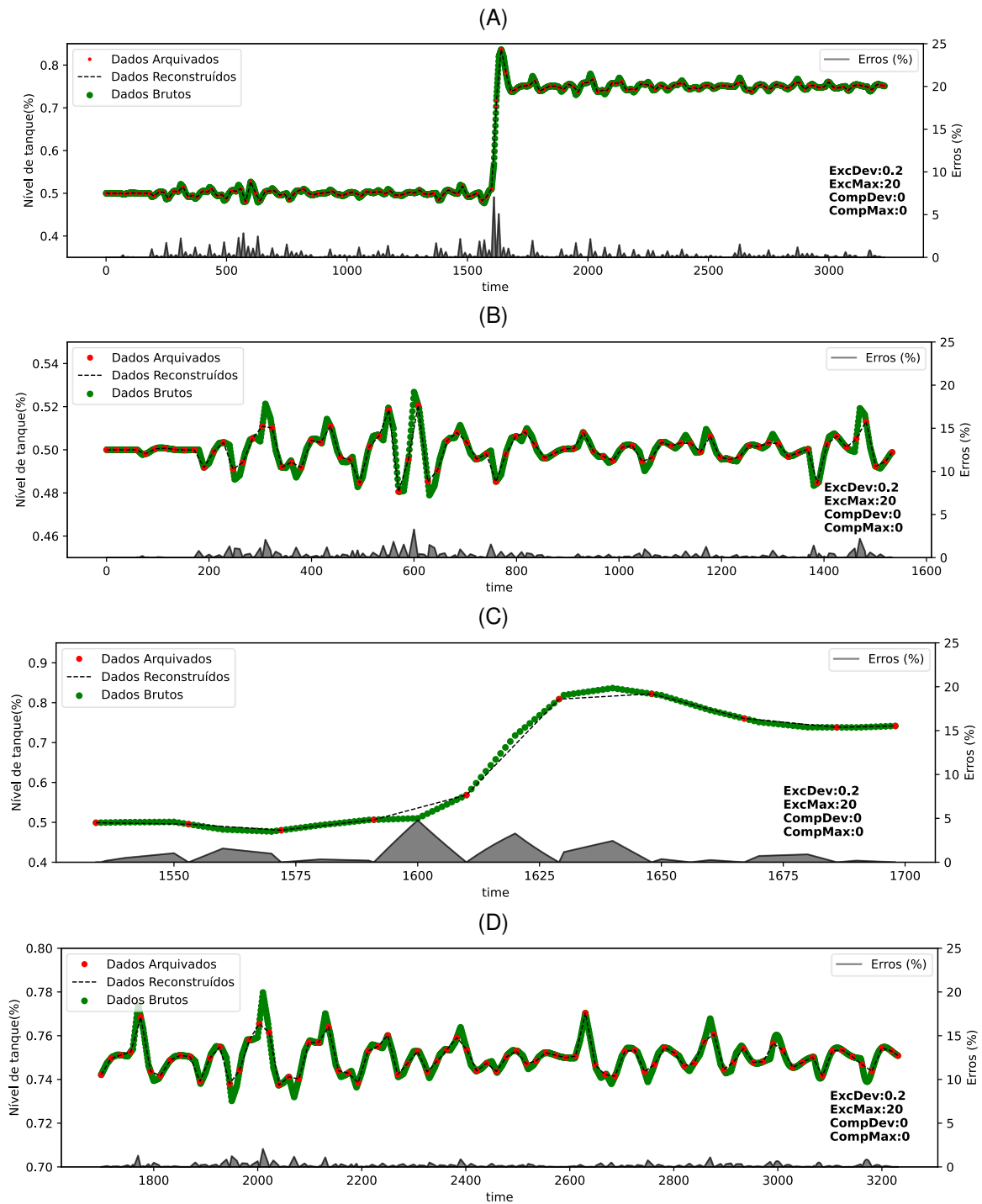


Figura 74 Avaliação do Algoritmo de exceção: ExcDev= 0.1, ExcMax=40.



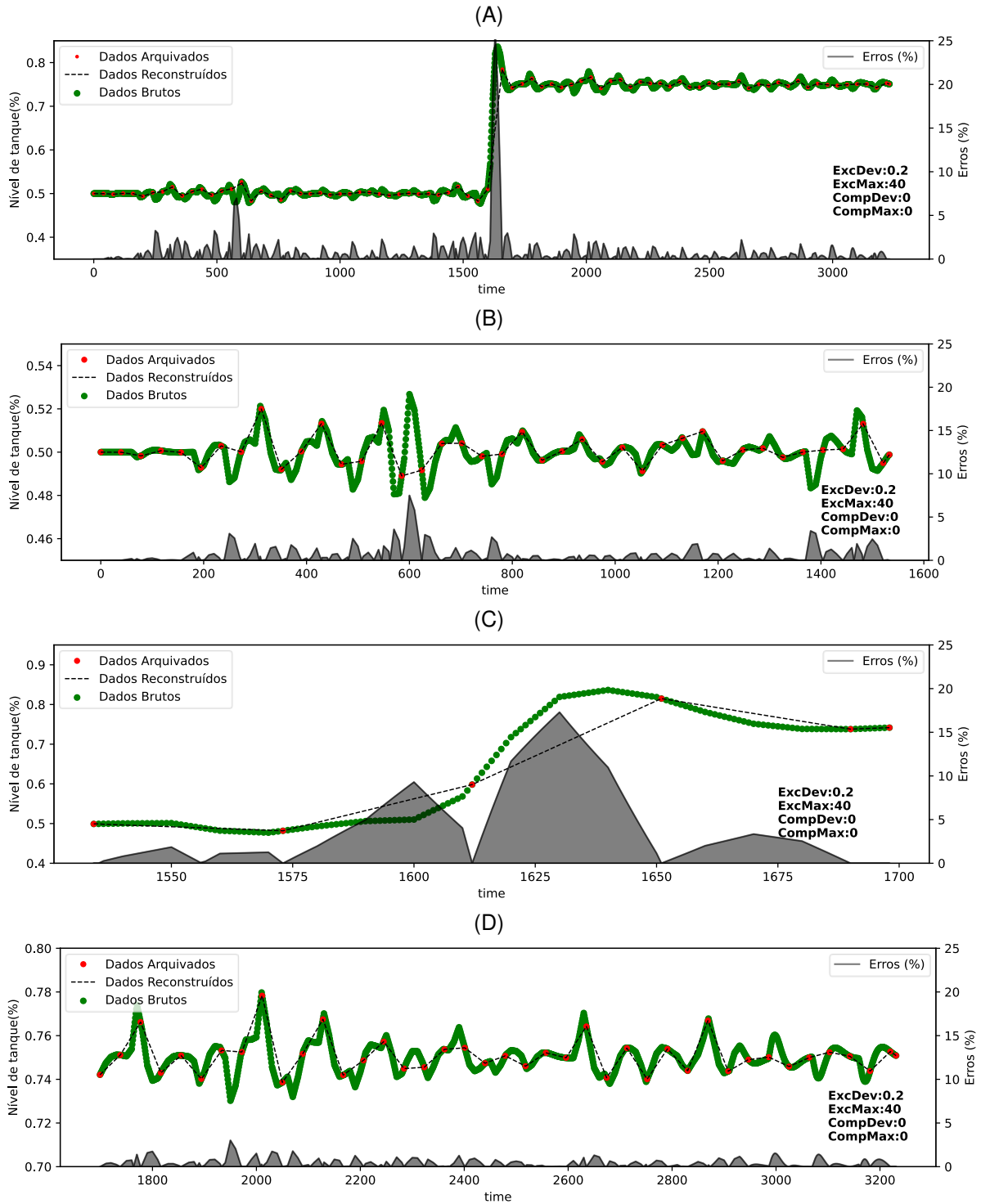
Fonte: Próprio Autor

Figura 75 Avaliação do Algoritmo de exceção: ExcDev= 0.2, ExcMax=20.



Fonte: Próprio Autor

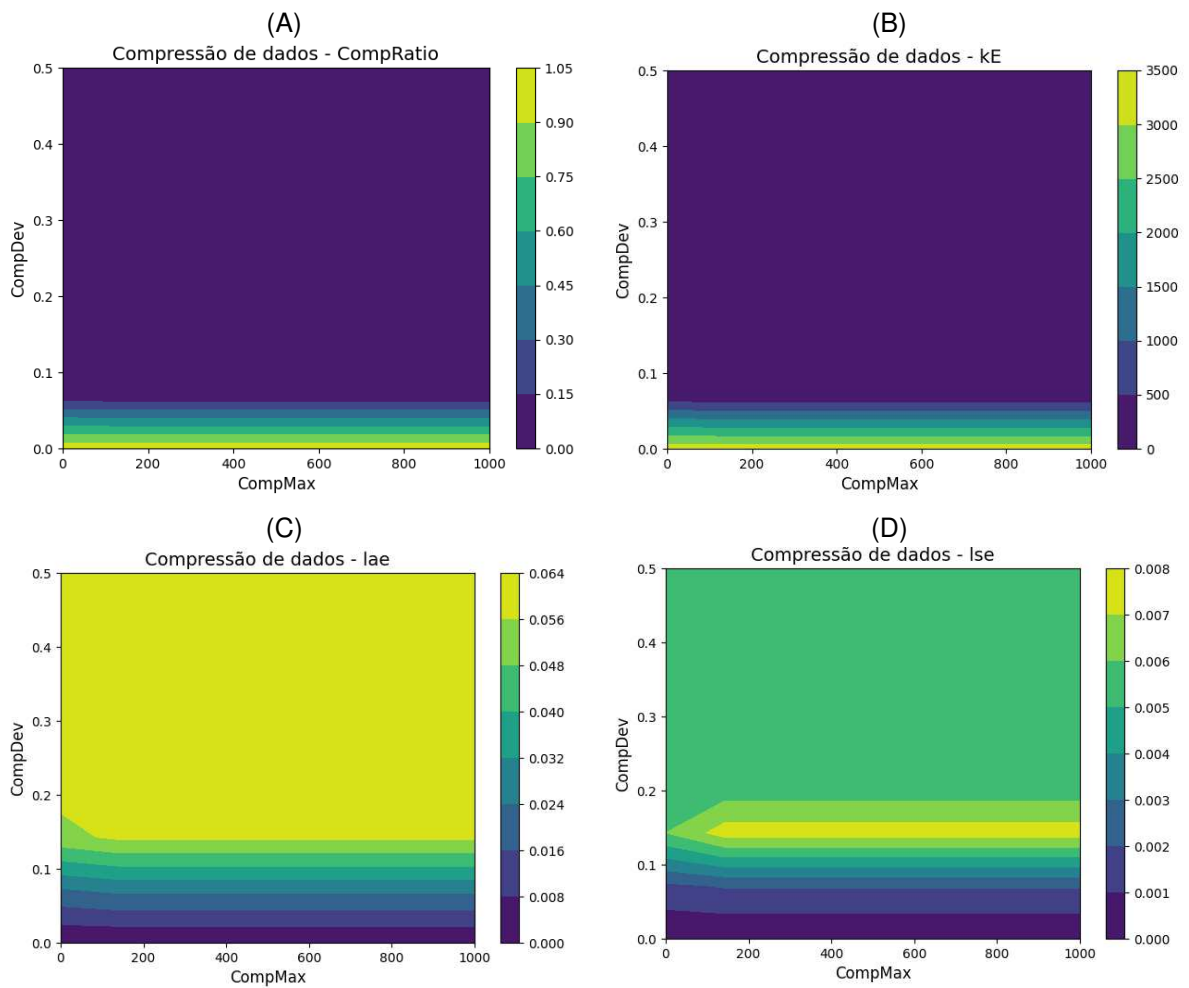
Figura 76 Avaliação do Algoritmo de exceção: ExcDev= 0.2, ExcMax=40.

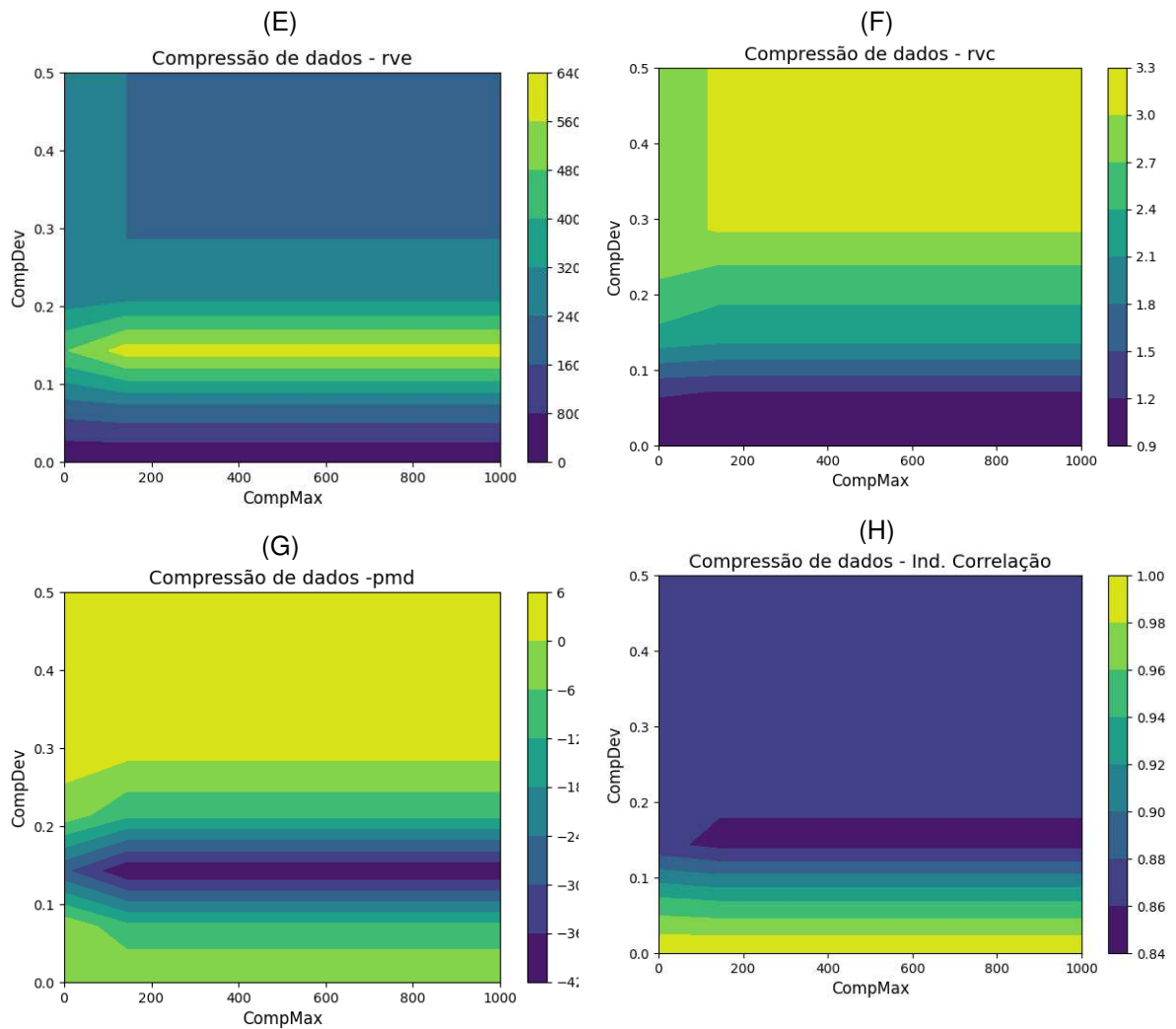


## 6.9. Avaliação dos parâmetros do algoritmo de compressão

Para análise dos parâmetros do algoritmo de compressão de dados, foi realizada a avaliação dos indicadores propostos em intervalo definido. A Figura 77 apresenta os oito gráficos de nível, cada um dos indicadores propostos plotados nos intervalos [0-0,5] para *CompDev* e [0-1000] para *CompMax*.

Figura 77 - Indicadores estatísticos para avaliação dos parâmetros do algoritmo de compressão de dados.





Fonte: Próprio Autor

Em Figura 78.A e Figura 79.B evidenciam que há um limite que compressão natural, ou seja, a exclusão de todos os pontos. Em A, a fração de dados restantes chega a zero, assim como em B, que evidencia a quantidade de pontos arquivada. No entanto, ao considerar o eixo em relação a *CompDev*, evidencia-se que este dita a quantidade de pontos que são arquivados, especificamente entre os valores 0 e 0.6. Nas Figura 80.C e Figura 81.D o comportamento semelhante é identificado, porém, sinais da influência de *CompMax*, sobre a qualidade da compressão são evidenciados. Os intervalos de interesse ficam entre [0-0,15] para *CompDev* e [0-100] para *CompMax*.

Na Figura 82.E e Figura 83.F demonstra que a região especificada é influenciada por ambos parâmetros, com maior influência em relação a *CompDev*. Estes indicadores apontam o impacto do algoritmo de compressão sobre a variância do erro e covariância dos dados. Para ambos os intervalos, de interesse são os seguintes [0

– 0,3] para *CompDev* e [0 – 100] para *CompMax*, por estes demonstrarem o gradiente numérico entre estes valores.

Na Figura 84 G, apresenta avaliação do indicador PMD. Nesta figura evidencia-se a distorção aplicada pelo algoritmo de compressão sobre a média dos dados. Devido a reversão dos valores em torno de *CompDev* igual a 1.5, o intervalo de interesse é estabelecido entre [0 –0,15]. A influência de *CompMax* é presente no intervalo [0-180]. Na Figura 85.H é apresentado a índice de correlação linear de pearson. O gradiente apresentado demonstra que o valores deste indicador são influenciados prioritariamente por *CompDev*, cujo intervalo de interesse se estabelece entre [0 – 0,15], enquanto para os parâmetros *CompMax* o intervalo por ser verificado entre [0 -100], porém sem demonstrar grandes mudanças.

Como resultado destas avaliações, tem-se o intervalo no qual os parâmetros de compressão adequados podem ser encontrados. Todas estas verificações são listas na Tabela 23abaixo.

Tabela 23 Verificação dos intervalos de interesse dos indicadores estatísticos para algoritmo de exceção.

Indicador	Figura 77	CompMax	CompMax	Indicador	Figura 77	CompMax	CompMax
CompRatio	A	-	[0 - 0,06]	RVE	E	[0 - 100]	[0 - 0,30]
KE	B	-	[0 - 0,06]	RVC	F	[0 - 100]	[0 - 0,30]
MAE	C	[0 - 100]	[0 - 0,15]	PMD	G	[0 - 180]	[0 - 0,15]
MSE	D	[0 - 100]	[0 - 0,15]	Ind. Corre.	H	[0 - 100]	[0 - 0,15]

Fonte: Próprio Autor

Os intervalos que englobam por todos os indicadores são o seguinte: [0-0,1] para *CompDev* e [0-100] para *CompMax*, em incrementos de 0,05 e 10 respectivamente. Ao considerar o ganho de em *CompRatio* o indicador que dita o máximo de compressão e os demais como restrições contra perda informações de dados. Portanto, para de *CompMax* maiores que 0.06, os ganhos em *CompRatio* não previstos, porém, como demonstrados por demais indicadores este acréscimo tende impactar a qualidade dos dados armazenados. Foi mantido o intervalo que abrange além deste valor, adiantes serão apresentados o intervalo considerado será [0-0,1] para *CompDev* para fins de análise.

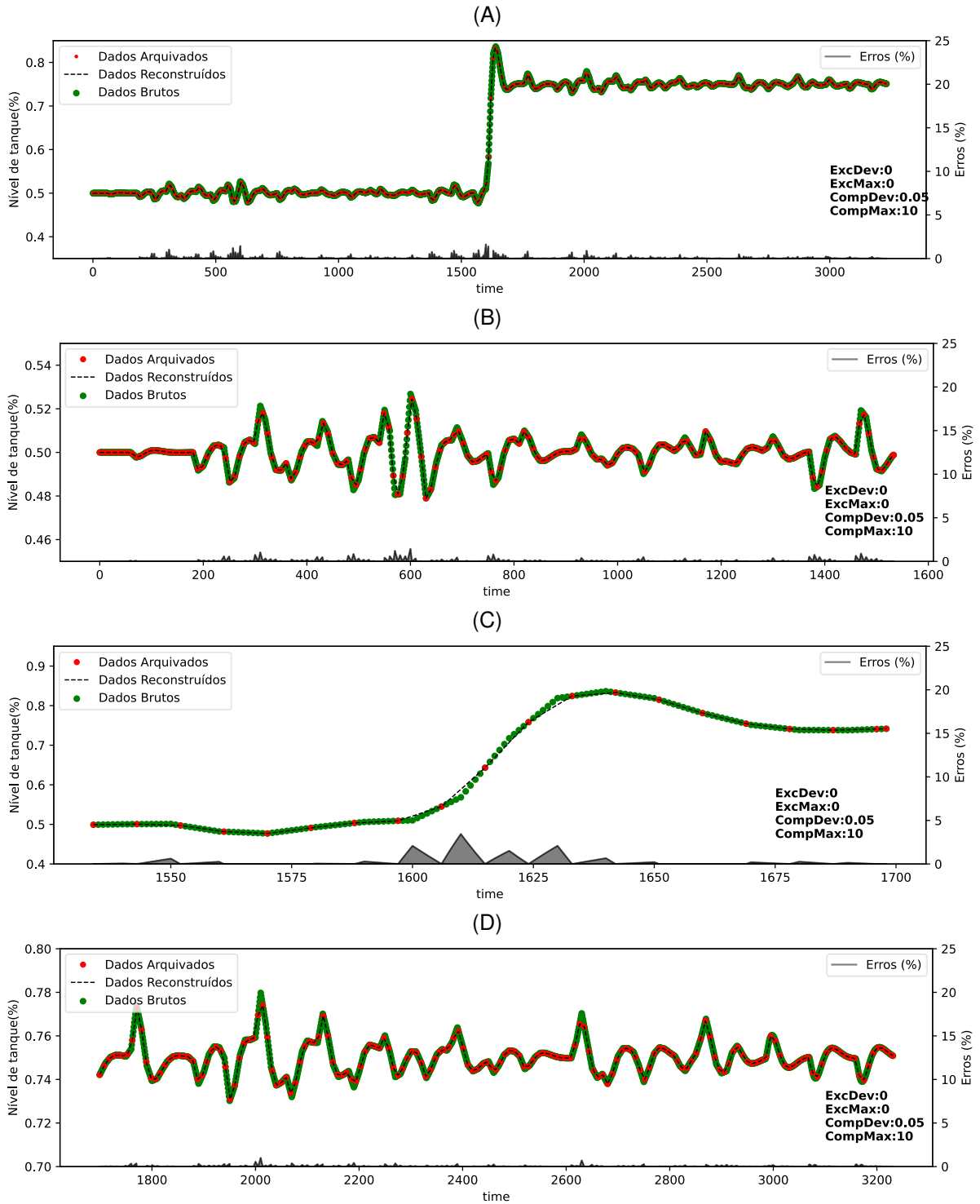
Para os intervalos definidos para parâmetros *CompDev* e *CompMax*, estão disponíveis as análises da zona completa (A), zona 1(B), zona 2(C) e zona 3(D). São apresentados os dados brutos, arquivados e reconstruídos assim como a avaliação dos erros individuais entre dados brutos e reconstruídos. A Tabela 24 a seguir demonstra os parâmetros aplicados em cada figura, assim como os resultados dos dois indicadores *CompRatio* e *MSE*:

Tabela 24 Resultados da compressão de dados, apresentação de resultados para os intervalos estabelecidos.

Figura	CompDev	CompMax	CompRatio	Erro de Reconstrução (MSE)
Figura 86	0.05	10	11,139%	0,00960%
Figura 87	0.05	20	5,291%	0,14300%
Figura 88	0.05	40	2,599%	0,45170%
Figura 89	0.1	20	5,291%	0,14300%
Figura 90	0.1	40	2,599%	0,45170%

Fonte: Próprio Autor

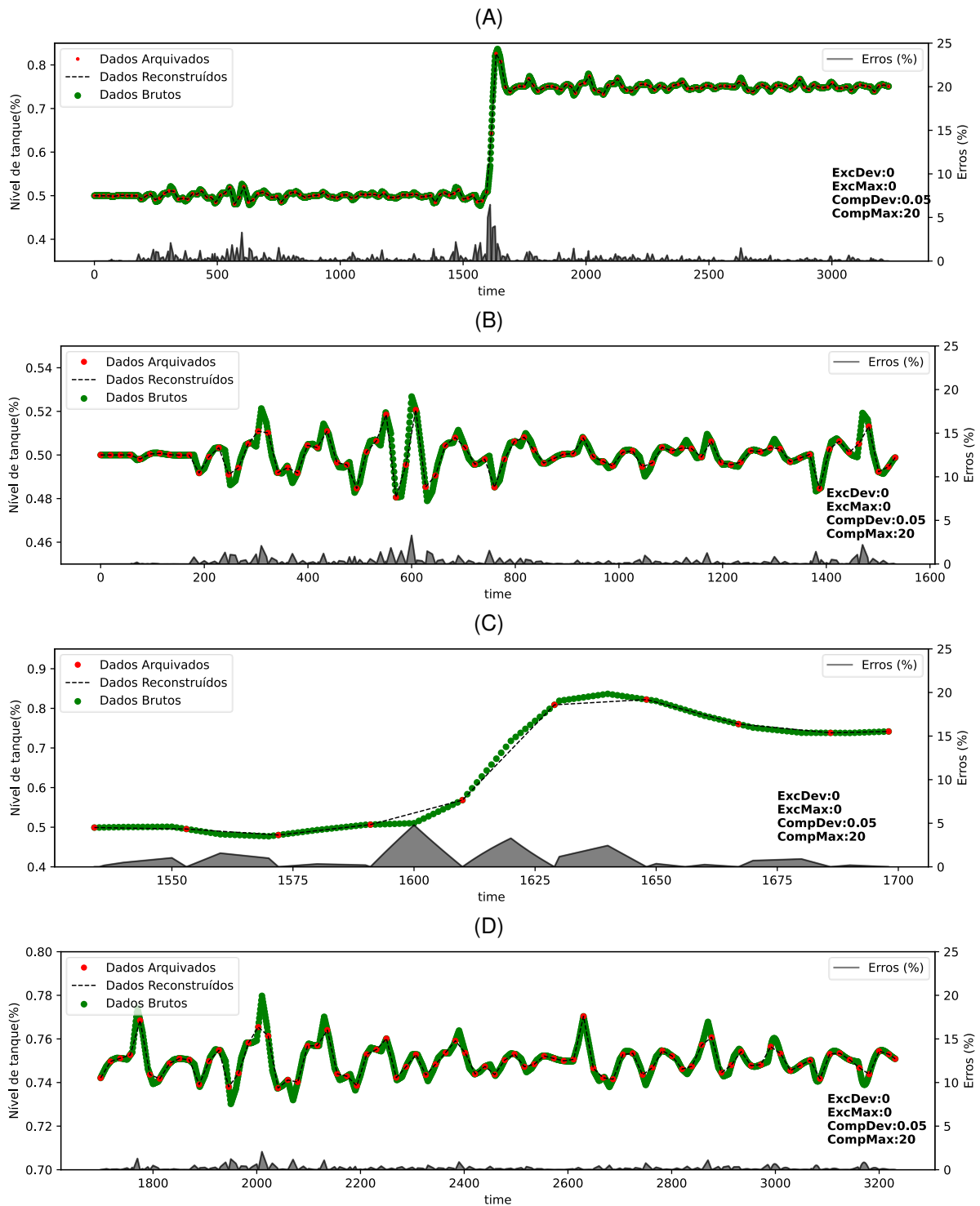
Figura 86 Avaliação do Algoritmo de Compressão: CompDev= 0.05, CompMax=10.



Fonte: Próprio Autor

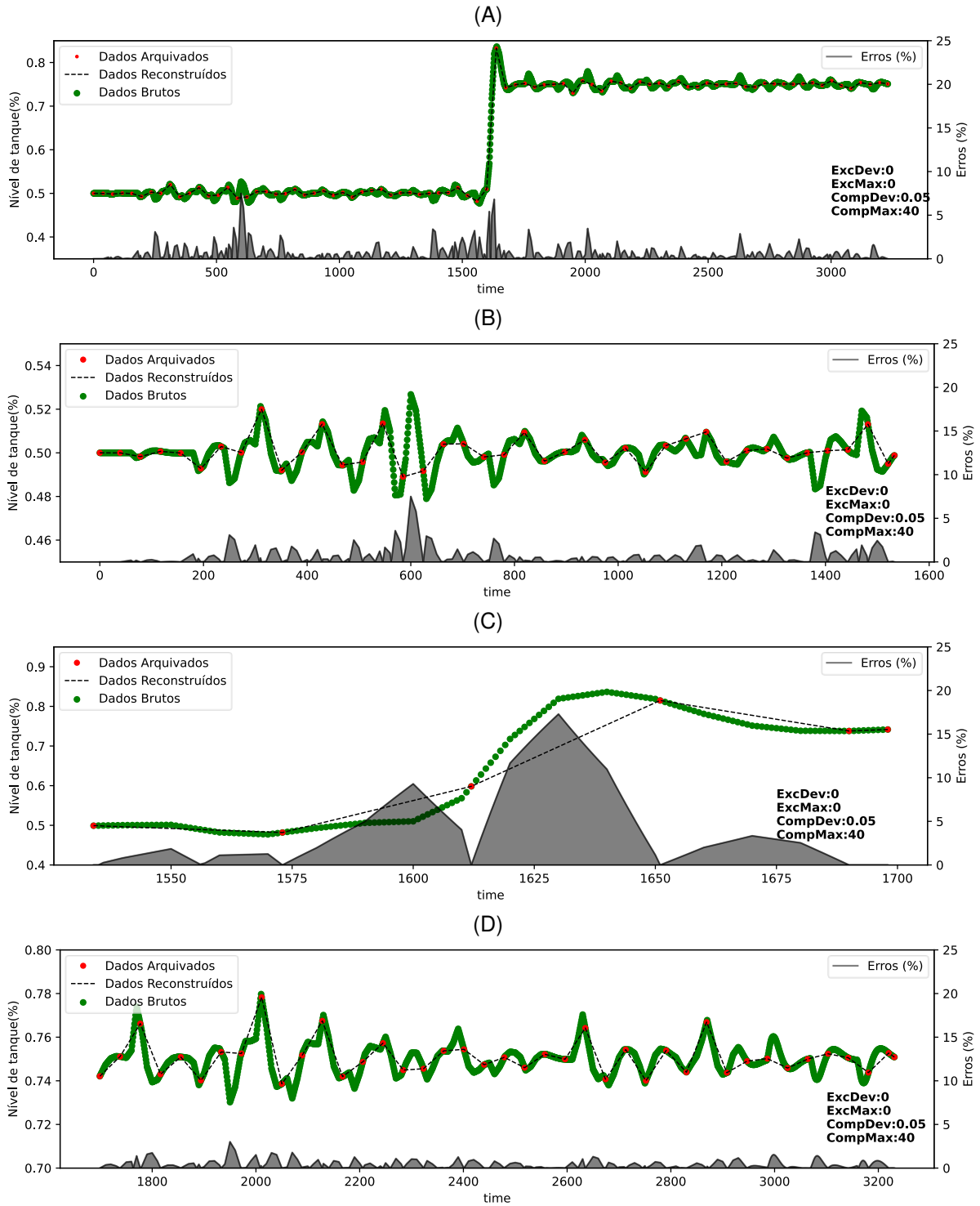


Figura 87 - Avaliação do Algoritmo de Compressão: CompDev= 0.05, CompMax=20.



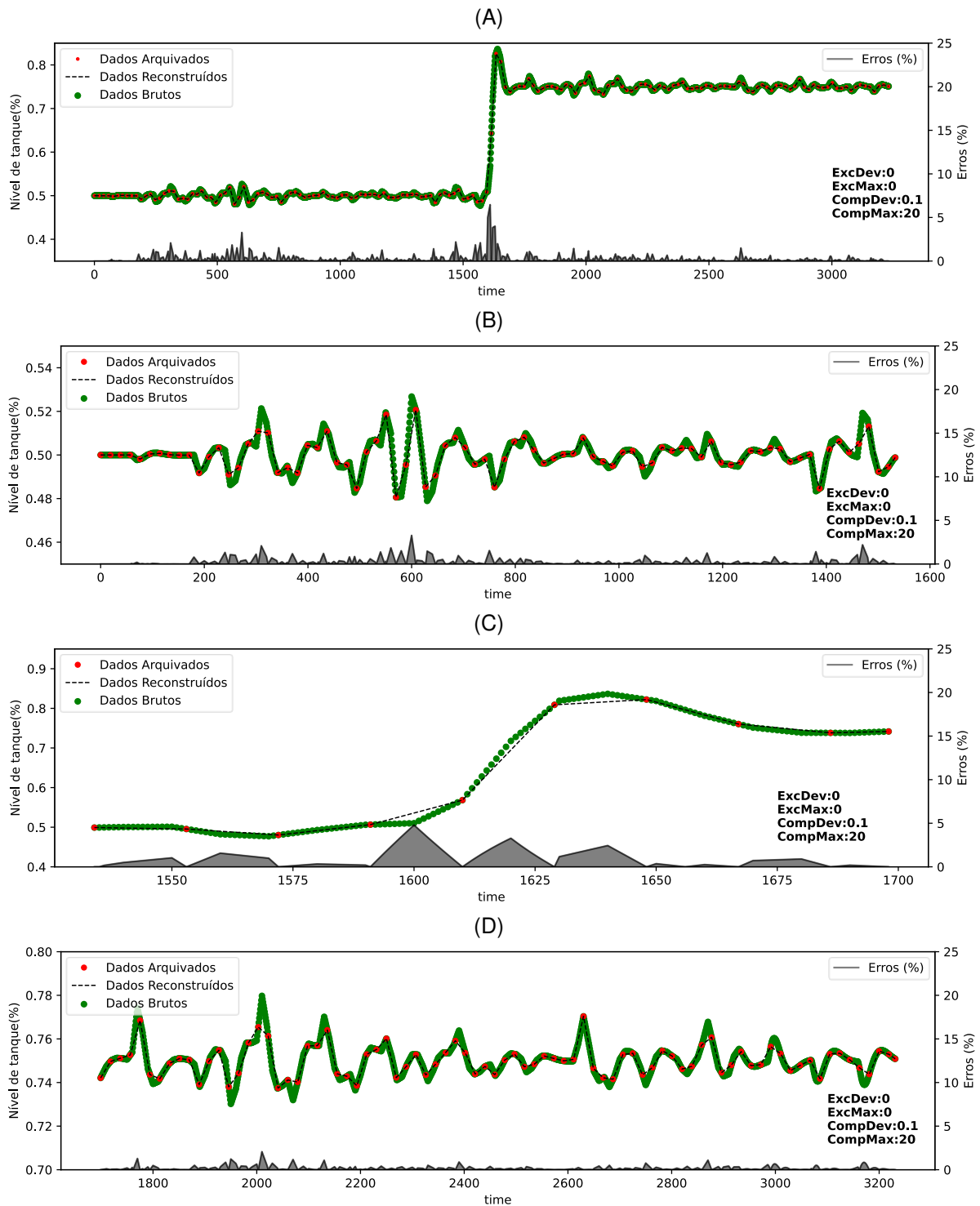
Fonte: Próprio Autor

Figura 88 - Avaliação do Algoritmo de Compressão: CompDev= 0.05, CompMax=40.



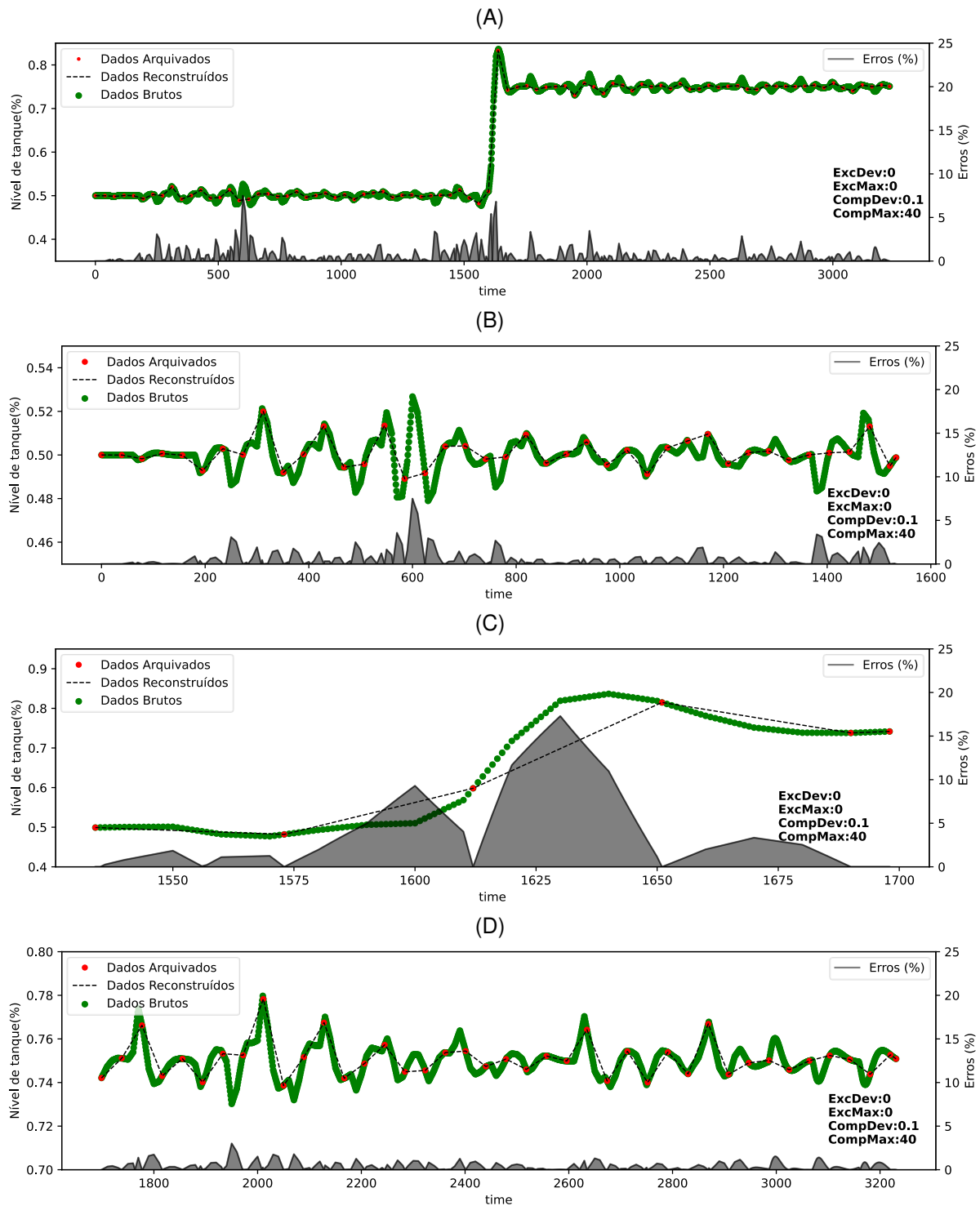
Fonte: Próprio Autor

Figura 89 Avaliação do Algoritmo de Compressão: CompDev= 0.1, CompMax=20.



Fonte: Próprio Autor

Figura 90 Avaliação do Algoritmo de Compressão: CompDev= 0.1, CompMax=40.



## 6.10. Resultados da avaliação dos algoritmos de exceção e compressão de dados

Como resultado das avaliações anteriores, têm-se os parâmetros de ambos os algoritmos de exceção e compressão apresentados na Tabela 25.

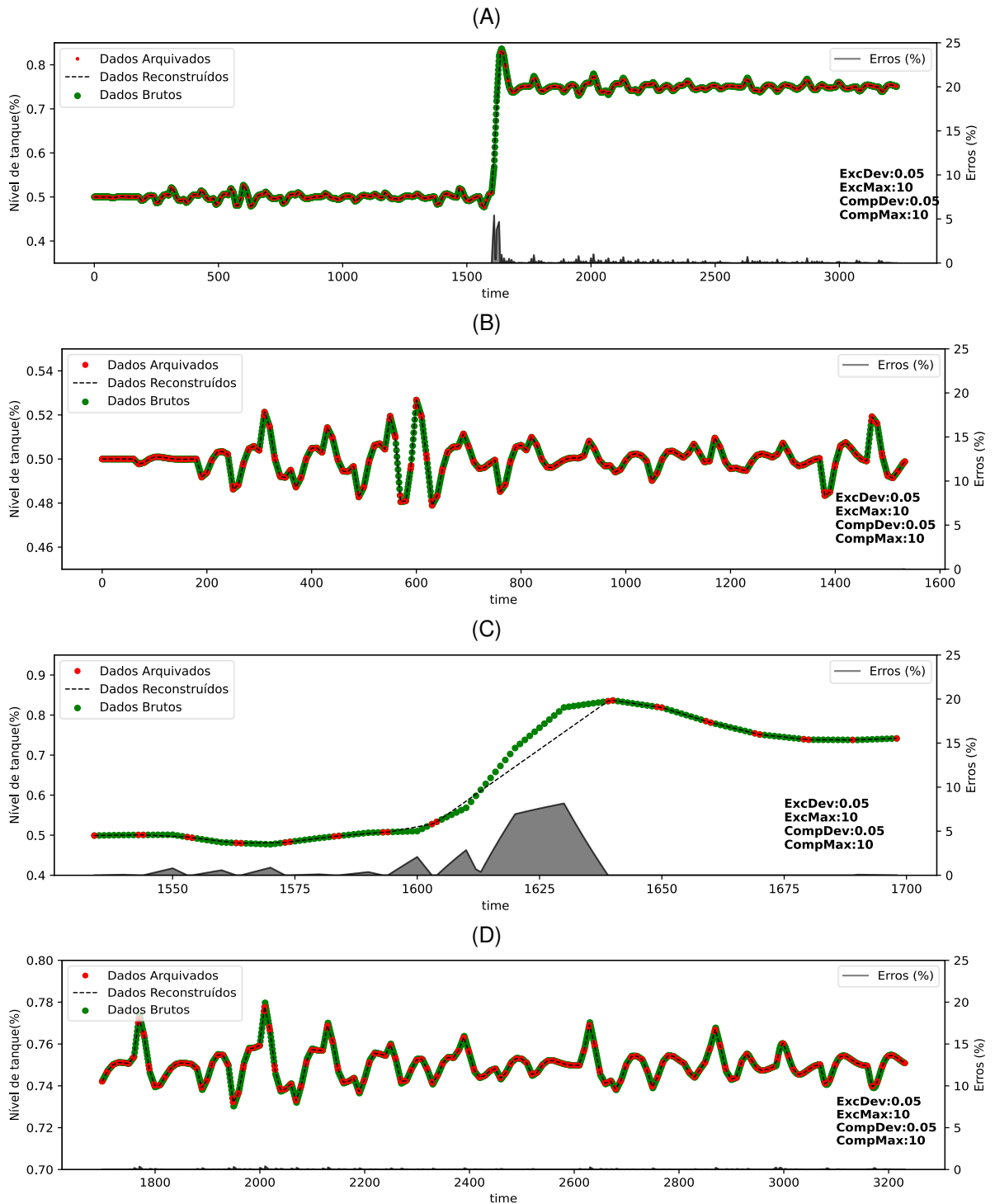
Tabela 25 Parâmetros ajustados para os algoritmos de exceção e compressão de dados.

Parâmetros	Valores
ExcDev	0.05
ExcMax	10
CompDev	0.05
CompMax	10

Fonte: Próprio Autor

Segue-se então na Figura 91 os resultados de todo o processo de compressão de dados. Em Figura 91.A são apresentados todo o conjunto de dados. Em Figura 91.B e Figura 91.C são apresentados os resultados para ambos os estados estacionários de operação do flash tanque. A mudanças de estado estacionário é apresentada Figura 91.C, cuja presença pode ser verificada a partir dos dados historiados.

Figura 91 Processo de compressão de dados com parâmetros de algoritmos de exceção e compressão.



Fonte: Próprio Autor

Os resultados obtidos com os melhores parâmetros são apresentados na mantendo-se a correlação linear em 99%. Para a zona 2, os resultados foram satisfatórios, considerando CompRatio de 16% e erro quadrado médio de  $1,9 \times 10^{-5}$  % para a reconstrução dos dados. Foi observado que o PDM nesta região foi alterado

positivamente, indicando alteração residual na média dos dados reconstruído. É também indicada a mudança em RVE, o maior valor entre as 3 regiões observadas, o que evidencia a presença de maiores distorções nos dados reconstruídos. Estes são divididos em suas respectivas zonas e por fim apresentados os resultados para o total dos dados.

Tabela 26 Resultado do ajuste do processo de compressão de dados.

Indicador	Zone-1	Zone-2	Zone-3	Completa
pdm	0,001599867	3,8029337	-0,040256	0,0605632
rvc	0,99999181	1,0827293	1,062441	1,0034067
rve	0,008221499	241,48379	48,68112	8,089218735
MSE	9,7903e-12	0,0002891	1,8925365e-07	6,38037355e-06
MAE	1,11008945e-07	0,0071273132	0,00023449047	0,000503387
nC	1534	165	1533	3232
kC	308	28	308	642
CompRatio	0,200782268578	0,16969696969	0,200913242	0,19863861386138615
Ind, Corr,	0,999999898	0,99359576	0,998438989851	0,999800054

Fonte: Próprio Autor

O processo de compressão executado alcançou resultados semelhantes nas zonas 1 e 3, que correspondem as regiões operacionais do sistema, os valores do erro quadrado médio foram reduzidos e *CompRatio* obtido de 20%, mantendo-se a correlação linear em 99%.

Para a zona 2, os resultados foram satisfatórios, considerando *CompRatio* de 16% e erro quadrado médio de 1,9 e-5 % para a reconstrução dos dados. Foi observado que o PDM nesta região foi alterado positivamente, indicando alteração residual na média dos dados reconstruído. É também indicada a mudança em RVE, o maior valor entre as 3 regiões observadas, o que evidencia a presença de maiores distorções nos dados reconstruídos.

Considerando os resultados obtido sobre o conjunto total dos dados, verifica-se que os ambos os algoritmos de exceção e compressão de dados foram ajustados com sucesso, sendo definidos como apresentado na Tabela 25. Os indicadores estatísticos utilizados demonstram que 19,86% dos dados foram arquivados, com erro

mélio quadrado de 0,000638%. Os indicadores PDM, RVC, RVE indicam a presença de distorções mínimas sobre os dados reconstruídos.



### 6.11. Aplicação de Cartas de controle para o tanque flash

Como resultado podemos verificar o uso das cartas de controle para o tanque flash em seus dois estados estacionários, Considerando que antes do processo de ajustes dos parâmetros dos algoritmos de exceção e compressão, esta aplicação seria inviável, Ao considerar os parâmetros padrão em tags tipo *Float32* ou *Float34*, os dados não eram armazenados, portanto sem dados suficientes para a aplicação de cartas de controle, como apresentado anteriormente nas Figura 69 e Figura 70,

Ao utiliza-se os parâmetros de compressão de dados alcançados através da verificação no intervalo especificado através dos indicadores, é possível obter dados reconstruídos que representam o perfil observado nos dados brutos. Para este fim são utilizados os limites de controle apresentados em Figura 65 para a zona 1 e Figura 66 para a zona 3, respectivamente o primeiro e segundo estados estacionários,

A seguir são apresentados as Figura 92.A e Figura 92.B, Ambas as cartas foram construídas a partir dos dados reconstruídos. Como resultados práticos, a discrepância entre os resultados das cartas de controle para ambas as zonas, comparando-se os subgrupos racionais obtidos a partir dos dados brutos e dados reconstruídos,

A Tabela 27 apresenta um comparativo quantitativo de pontos fora dos limites de controle para ambas as zonas, é esperada que a quantidade de pontos fora dos limites de controle seja igual entre subgrupos gerados a partir de dados brutos e reconstruídos, ou seja, quantitativamente os mesmos valores de subgrupos fora dos limites de controle. Para zona 1 foi observada uma diferença de 0% e para zona 3 diferença de 2,532%.

Tabela 27 Resultados obtidos para cartas de controle  $\bar{X}$ , considerando a utilização de dados cumprido com o processo de compressão ajustado.

	Zona 1		Zona 3	
	Dados brutos	Dados Reconstituídos	Dados brutos	Dados Reconstituídos
Acima de LSC	31	31	34	33
Abaixo de LIC	40	40	45	44
total	71	71	79	77
Diferença	0%		2,532%	

Fonte: Próprio Autor

Estes resultados demonstram que mesmo com o processo de compressão ajustados, os resultados da obtidos pelas cartas de controle foram minimamente alterados. Porém, esta diferença não leva a tomada de decisão imprecisa, pois a quantidade de falsos negativos ou falso positivos obtido ambos apresentados na Tabela 27,

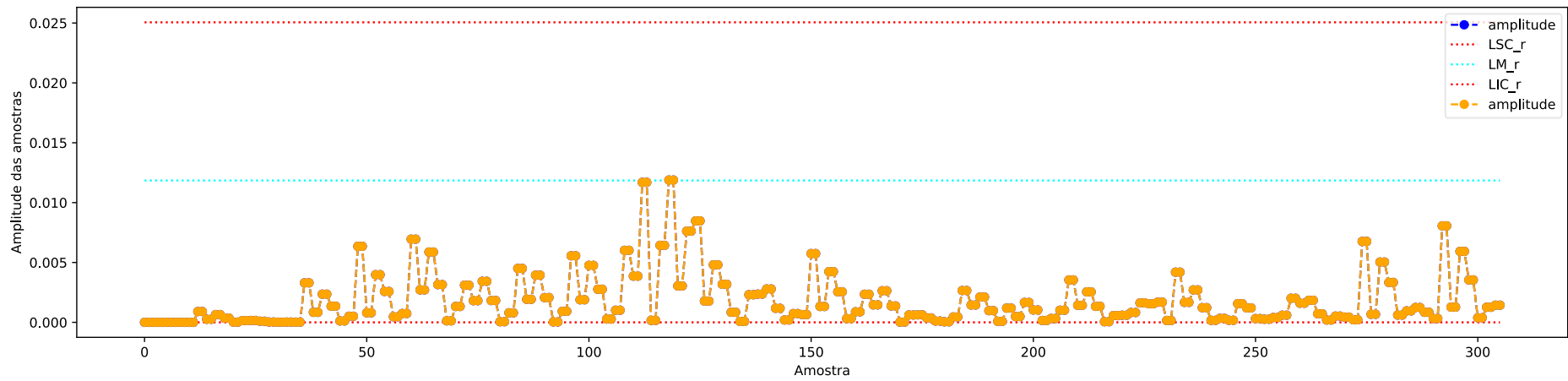
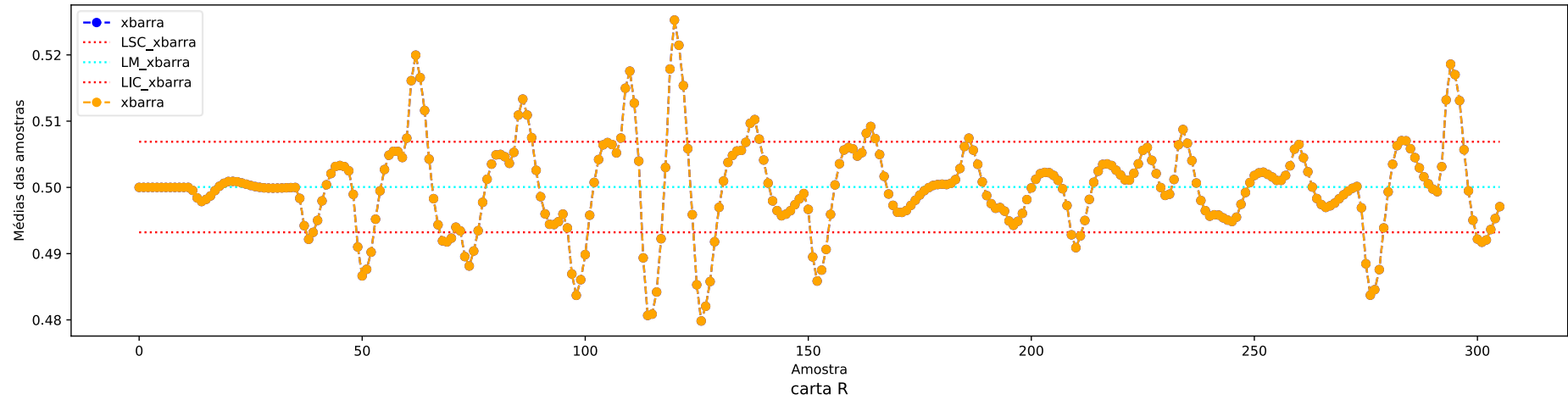
Análise semelhante não é possível ser realizada sobre a carta R, Devido as características do processo em estudo, o qual não possui variabilidade oscilante nas zonas 1 e 3. No entanto, é possível avaliar o comportamento desta carta, ao considerar a zona 2, na qual o corre a mudança do estado estacionário 1 para o estado estacionário 2,

A Figura 92,C traça toda as curvas de subgrupos racionais, construídos com dados brutos e dados comprimidos para zona completa, combinando as zonas 1, 2 e 3. É possível verificar uma mudança na variabilidade do sistema em torno na zona 2, porém sem alcançar o rompimento dos limites de controle definidos, destoando do comportamento observado quando a carta R é operada com dados brutos. No entanto, há uma forte mudança no comportamento da média, ao combinar ambos os comportamentos, é possível concluir que houve uma mudança de especificação do sistema monitorado.

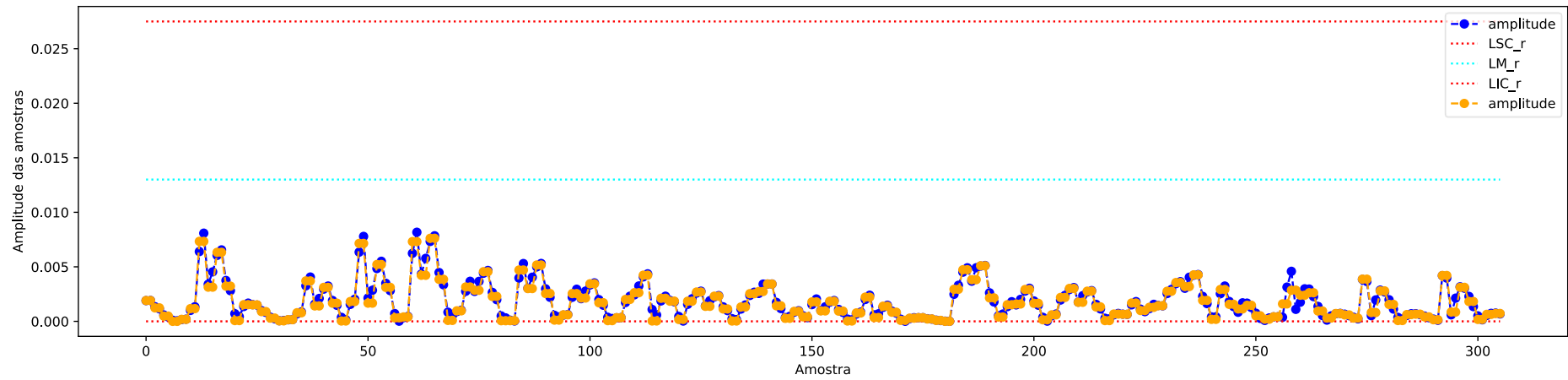
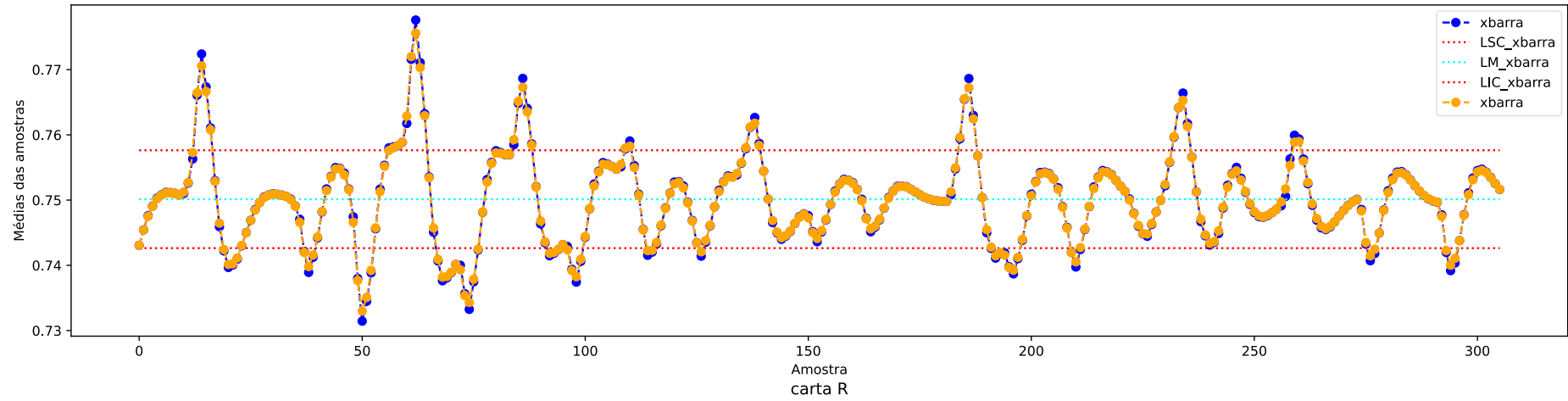
Figura 92 Cartas de controle a partir de dados comprimidos.

(A) Zona 1

carta X-barra

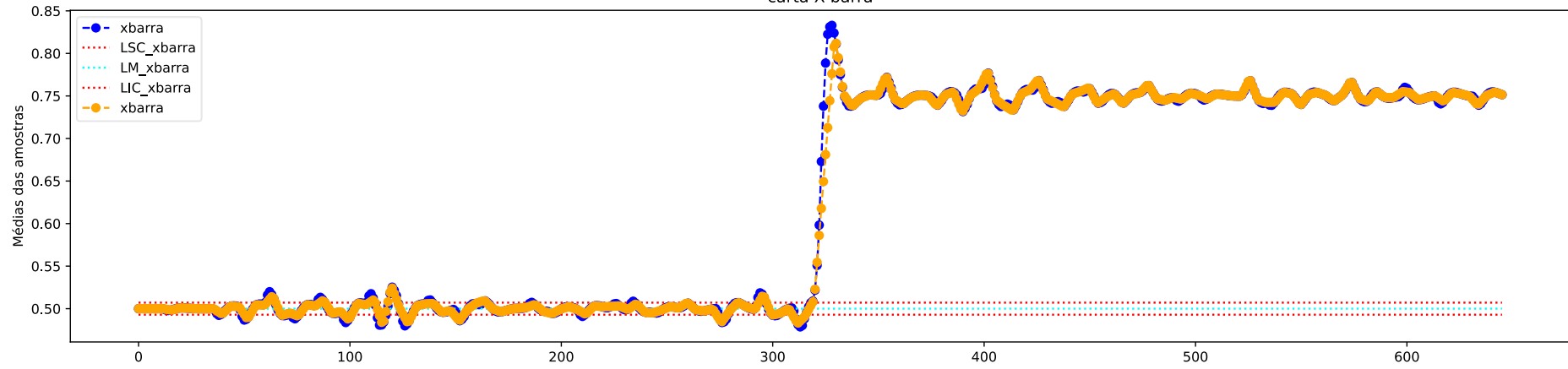


(B) Zona3  
carta X-barra

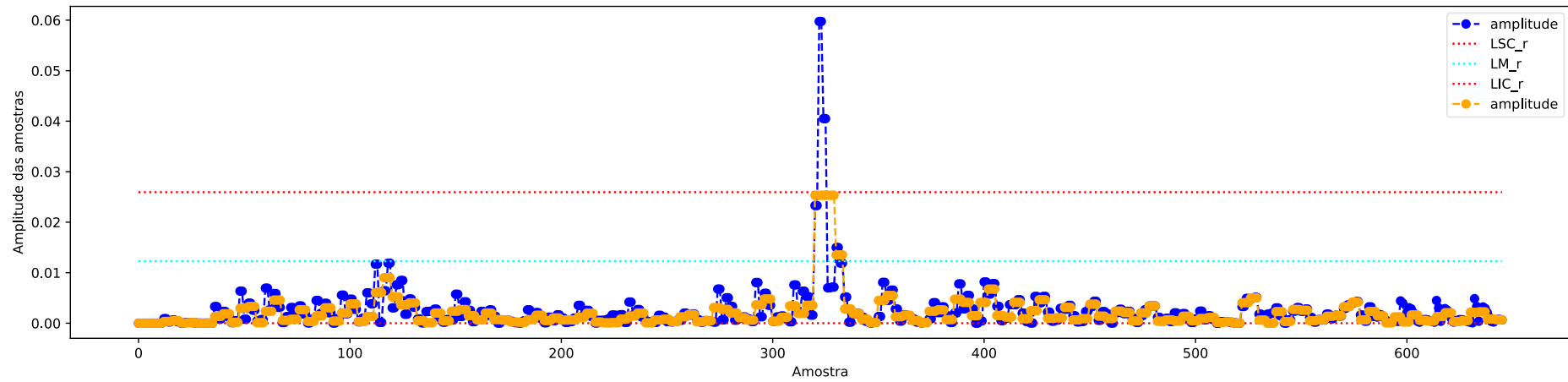


(C) Zona Completa

carta X-barra



carta R



Fonte: Próprio Autor

## 7. Conclusão

O presente trabalho alcançou o objetivo proposto de avaliar o processo de compressão de dados aplicado em sistema industrial de armazenamento de dados, comporto pela combinação dos algoritmos de exceção e compressão de dados,

O processo de compressão de dados foi encapsulado em biblioteca de análise e simulação, construída em linguagem Python®, A validação da biblioteca foi realizada a partir da coleta de dados de corrente elétrica de motores de bombas industriais, A coleta de dado foi realizada em um conjunto de tags com propriedades diferentes, de maneira a possibilita a validação dos algoritmos em estudo,

Indicadores estatísticos foram aplicados para analisar medidas de centralidade, dispersão e similaridade nos dados brutos e reconstruídos, Estes foram utilizados em duas situações, validação dos algoritmos e na busca por parâmetros ótimos de exceção e compressão de dados,

O estudo de caso foi aplicado ao problema clássico de simulação dinâmica de equilíbrio líquido vapor de hidrocarbonetos em tanque flash. A plataforma SimCentral possibilitou a definição dos balanços de massa, energia além de possibilitar a construção da malha de controle para a fração de nível. Esta variável foi coleta juntamente com o timestamp de simulação para dois estados operacionais do nível de tanque. Para a fração do nível, o processo de avaliação dos parâmetros de exceção e compressão foi realizado com sucesso e alcançou resultados adequados. O perfil definido pela dinâmica da variável foi mantido, assim como o erro médio quadrado da reconstrução de dados fico em níveis significativamente baixos, na ordem de 0,000638% para o conjunto completo de dados, alcançando a redução de 80% quantidade de pontos originais.

Como resultado foram definidos os parâmetros de compressão (*CompDev*, *ComMax*) e de exceção (*ExcDev*, *ExcMaxI*) para o armazenamento adequado destes dados, caso esta variável venha a ser armazenada no PI System. Estes resultados foram alcançados aplicando-se a metodologia proposta, tomando-se os indicadores em intervalos que demonstram menor variação em medidas de variabilidade e centralidade entre os dados bruto e dados reconstruídos,

Para o monitoramento de causas especiais foi aplicado a controle de estatístico de processos  $\bar{X}$  e R. Como resultado, foi observado que os dados reconstruídos a partir do resultado do processo de compressão ajustado foram capazes de manter as informações de processo. Sem o ajuste dos parâmetros de exceção e compressão, os dados não transmitiam as informações estatísticas necessárias à aplicação das cartas utilizadas,

A equivalência entre as cartas de controle construídas a partir de dados comprimidos foi verificada. Para as zonas 1 e 2, foram encontradas diferenças de 0% e 2,53% em quantidade de subgrupos fora dos limites de controle. Portanto, após o ajuste dos parâmetros do processo de compressão de dados é possível a utilização dos reconstruídos para o monitoramento de causas especiais em processos químicos. Foi possível habilitar os dados reconstruído fossem empregados em cartas de controle de processo, antes impossível devido a inexistência de dados armazenados na tag de coleta sistema monitorado, quando o processo de compressão é definido com os valores padrão.

Portanto, a aplicação dos indicadores utilizados demonstra-se como vantajosa para definição de intervalos de busca dos valores de parâmetros de exceção e compressão, chegando a reduzir em 80% a quantidade de ponto armazenados após o processo de compressão.

## **8. Sugestão de trabalhos futuro**

Em função da análise de sensibilidade ter aplicada no presente trabalho, é sugerida a automatização da seleção dos intervalos de interesse para os indicadores aplicados assim como a definição dos pares de parâmetros *Excdev*, *ExcMax* e *CompDev*, *CompMax*.

É sugerida a verificação da metodologia apresentadas em variáveis diversas de outros processos químicos com dinâmicas variadas, na tentativa de mapear os intervalos típicos assumido pelos indicadores estatísticos utilizados. As avaliações de outras viáveis de processo estabelecer estratégia de automatização da metodologia adotada. Outras métricas de avaliação padronizadas de erros podem ser utilizadas, de forma a adequar a comparação entre as variáveis diferentes.



## REFERÊNCIAS

MONTGOMERY; David, **The Use of Statistical Process Control and Design of Experiments in Product and Process Improvement**, 1992,

CINAR, ali; PALAZOGLU, Ahmet; KAYIHAN, Ferhan, **Chemical Process Performance Evaluation** , 2007,

BRISTOL, E,H, **Swinging Door Trending: adaptive trend recording**, pp, 749-753, 1990,

SOUZA; Rigão, **Identificação de Variáveis Fora de Controle em Processos Produtivos Multivariados**, 2005,

PYTHON – (<https://wiki.python.org/moin/>,) Acessado 30 de3 Outubro

SILVEIRA; TRIERWEILER; FARENZENA; TEXEIRA, **Systematic Approaches for PI System Data Compression Tuning**, 2012,

HE; WANG, **Statistical Process Monitoring as a Biga Analytics tool for Smart Manufacturing**, 2017,

THORNHILL, N,F; CHOUDHURY, M, A, A.; SHA, S,L, **The impact of compression on data driven process analyses**, 2004,

SHEWHART, Walter A, **Economic Control of Quality Manufactured Product**, American Society for Quality (ASQ), 1980,

NETO, AUGUSTO; SOUZA; AFONSO, **Adaptative Swinging Door Trending - Um Algoritmo Adaptativo para Compressão de Dados em Tempo Real**, 2014,

CAULCUTT, R, **The right and wrong of control charts**, *Applied Statistics*,[S, 1]: v, 44,n, 3, p, 279-288, 1992,

MASTRANGELO, CHRISTINAM; RUNGER; GEORGE C.; DOUGLAS, C MONTGOMERY **Statistical Process Monitoring with Principal Components**, *Quality and Reliability Engineering international*, vol 12, p 203-2010, 1996,

XIAODONG, Feng; CHANGLIN, Chenge; CHANGLIN, Liu; HUIHE, Shao - **Improved Process Data Compression Algorithm** 4<sup>th</sup>World Congress of Control and Automation, p,2190-2193, 2002,

ALSMEYER - **Automatic Adjustment of Data Compression in Process Information Management Systems**, 2006,

SOUZA, Alessandro J, ; Rafael **Gerência de informação de informação de processos industriais: Um estudo de caso na produção de petróleo e Gás**, 2005,

BusissWire- **Osisoft Announces \$135 Million Minority Investment by Technology Crossover Ventures and Kleiner Perkins Caufield & Byers** <<https://www.businesswire.com/news/home/20110105006161/en/OSIsoft-Announces-135-Million-Minority-Investment-Technology>> acessado em Dezembro de 2019

DANTE, Jeferson Roberto; SETT, João Antônio Palma; LAUS, Luis Paulo; FERRAZZA, Paulo Humberto; MAGATÃ, Leandro, **Monitoring failure on a FMS using PIMS: A case study**,

OSISOFT - **Use of the PI System and Statistical Analysis for Quality Control and Performance Monitoring** <<https://www.osisoft.com/Presentations/Use-of-the-PI-System-and-Statistical-Analysis-for-Quality-Control-and-Performance-Monitoring/>> acessado em dezembro de 2019

OSISOFT – **PI System Administration Version 2018 SP3 Patch 3** < [Sample Class Workbook v2.1 \(windows.net\)](#)> acessado em junho de 2021.

STEPHNOPOULOS, Backshi - **Compression of Chemical Process Data by Functional Approximation And Feature Extraction**, 1996

OSISOFT - **What are recommended Exception and Compression** < - <https://customers.osisoft.com/s/knowledgearticle?knowledgeArticleUrl=3226OSI8> > Acessado em Março de 2020

COHEN, Jacob - **Statistical Power Analysis for Behavioral Sciences**, edição 2, 1988.