

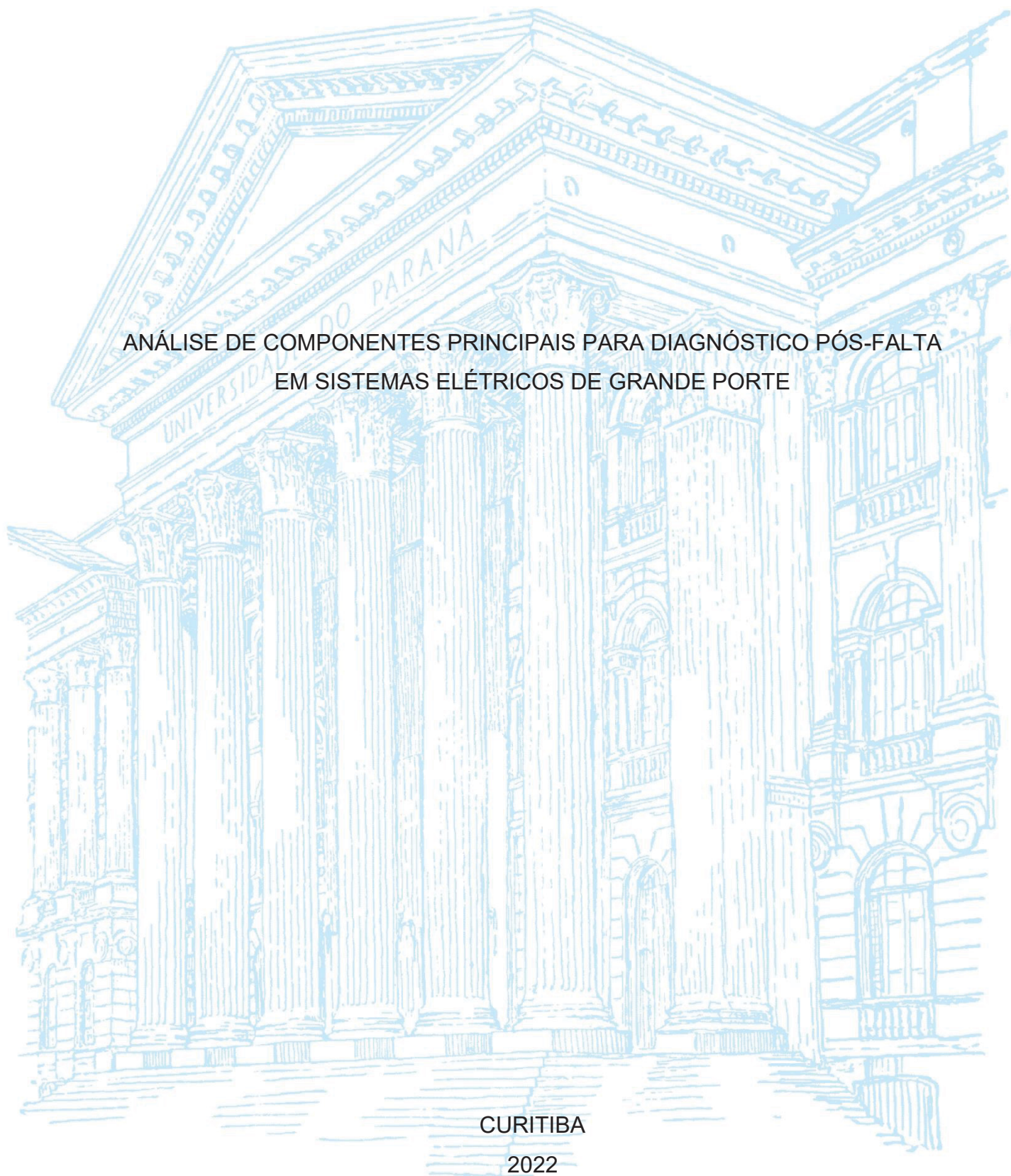
UNIVERSIDADE FEDERAL DO PARANÁ

VICTOR FREDERICO MÜLLER JUNIOR

ANÁLISE DE COMPONENTES PRINCIPAIS PARA DIAGNÓSTICO PÓS-FALTA  
EM SISTEMAS ELÉTRICOS DE GRANDE PORTE

CURITIBA

2022



VICTOR FREDERICO MÜLLER JUNIOR

ANÁLISE DE COMPONENTES PRINCIPAIS PARA DIAGNÓSTICO PÓS-FALTA  
EM SISTEMAS ELÉTRICOS DE GRANDE PORTE

Dissertação apresentada ao curso de Pós-Graduação em Engenharia Elétrica, Setor de Tecnologia, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Alexandre Rasi Aoki

Coorientador: Prof. Dr. Ricardo Schumacher

CURITIBA

2022

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
UNIVERSIDADE FEDERAL DO PARANÁ  
SISTEMA DE BIBLIOTECAS – BIBLIOTECA CIÊNCIA E TECNOLOGIA

Müller Junior, Victor Frederico

Análise de componentes principais para diagnóstico pós-falta em sistemas elétricos de grande porte / Victor Frederico Müller Junior. – Curitiba, 2022.

1 recurso on-line : PDF.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica.

Orientador: Prof. Dr. Alexandre Rasi Aoki.

Coorientador: Prof. Dr. Ricardo Schumacher.

1. Engenharia elétrica. 2. Sistemas elétricos - planejamento. 3. Análise de componentes principais. 4. Estatística. I. Aoki, Alexandre Rasi. II. Schumacher, Ricardo. III. Universidade Federal do Paraná. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

Bibliotecário: Nilson Carlos Vieira Junior CRB-9/1797



## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação ENGENHARIA ELÉTRICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **VICTOR FREDERICO MÜLLER JUNIOR** intitulada: **ANÁLISE DE COMPONENTES PRINCIPAIS PARA DIAGNÓSTICO PÓS-FALTA EM SISTEMAS ELÉTRICOS DE GRANDE PORTE**, sob orientação do Prof. Dr. ALEXANDRE RASI AOKI, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 25 de Fevereiro de 2022.

Assinatura Eletrônica  
25/02/2022 16:02:50.0  
ALEXANDRE RASI AOKI  
Presidente da Banca Examinadora

Assinatura Eletrônica  
25/02/2022 15:39:53.0  
GUSTAVO HENRIQUE DA COSTA OLIVEIRA  
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica  
28/02/2022 10:36:23.0  
FABIO ALESSANDRO GUERRA  
Avaliador Externo (INSTITUTO GNARUS)

Assinatura Eletrônica  
28/02/2022 14:25:43.0  
THELMA SOLANGE PIAZZA FERNANDES  
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Dedico esta dissertação aos meus pais, Victor e Lygia, à minha esposa Andrea e às minhas filhas Letícia e Larissa.

## AGRADECIMENTOS

A Deus pela benção da vida.

À minha esposa Andrea pelo seu amor e por todo apoio e incentivo que sempre me impulsiona.

Às minhas filhas Letícia e Larissa, criaturas maravilhosas cuja existência faz com que eu me inspire e esforce para sempre ser para elas um exemplo digno.

Aos meus pais, Victor e Lygia, por todo amor com que me criaram e por todas as oportunidades de estudo que me proporcionaram.

Ao meu professor orientador Dr. Alexandre Rasi Aoki por toda sua ajuda e apoio ao longo deste trabalho, e principalmente por ter acreditado em minha capacidade para conduzir este trabalho até o final.

Ao meu professor coorientador Dr. Ricardo Schumacher por sua ajuda e contribuições na execução deste trabalho.

Aos meus colegas de mestrado e à equipe do projeto de Pesquisa & Desenvolvimento que contribuíram com seu trabalho, ideias e discussões.

Aos meus colegas da COPEL que com seu conhecimento e boa vontade sempre *colaboraram* e contribuíram fornecendo as informações e esclarecimentos solicitados.

À COPEL pela oportunidade de participar do projeto de Pesquisa & Desenvolvimento no âmbito ANEEL, PD-06491-0531/2019, que viabilizou minha participação neste mestrado.

Em realidade, na escola não devemos aprender para saber, mas devemos aprender para sempre podermos aprender com a vida.

Rudolf Steiner

## RESUMO

Os sistemas de medição fasorial sincronizada estão sendo amplamente utilizados ao redor do mundo, e se tornaram elementos essenciais na evolução da operação dos sistemas elétricos de potência de grande porte. Estes sistemas são baseados em unidades de medição fasorial, chamadas PMUs, as quais são capazes de gravar e comunicar dados da dinâmica dos sistemas elétricos de potência de forma sincronizada por GPS e com alta taxa de amostragem, gerando assim um enorme conjunto de dados. Neste trabalho é apresentada uma metodologia para seleção de dados para análise pós-falta, através da redução da dimensionalidade, utilizando a análise de componentes principais. A análise de componentes principais é uma técnica que permite identificar como diferentes variáveis numéricas variam em conjunto, tendo como princípio a combinação das múltiplas variáveis numéricas em um conjunto menor de variáveis, que são combinações lineares ponderadas do conjunto original. Para validar a metodologia proposta, foram usados dados reais e foram identificados alguns recentes casos reais conhecidos de ocorrências no sistema interligado nacional, para que a análise pós-falta conseguisse detectar uma anomalia. A validação dos resultados obtidos foi feita utilizando um algoritmo de detecção de anomalias baseado em um modelo gaussiano. Com a aplicação da metodologia, comprovou-se a possibilidade de se utilizar um conjunto de dados menor do que o original, porém mantendo as características dos atributos. Demonstrou-se um caso em que foi possível fazer a redução de 190 atributos, para um conjunto de oito componentes principais mantendo a precisão no *F1 Score*. Esta metodologia automática de seleção de dados traz benefícios para análise pós-falta devido ao ganho de desempenho obtido com a redução do conjunto de dados dos atributos elétricos, sem perda significativa na precisão da capacidade de detecção de anomalia.

Palavras-chave: Análise de componentes principais. Estatística multivariada. Monitoramento de área ampla. Unidade de medição fasorial.



## ABSTRACT

Synchronized phasor measurement systems are being widely used around the world, and have become essential elements in the evolution of the operation of large electrical power systems. These systems are based on phasor measurement units, called PMUs, which are able to record and communicate data on the dynamics of the SEPs in a synchronized way by GPS and with a high sampling rate, thus generating a huge dataset. This work presents a methodology for selecting data for post-fault analysis, through dimensionality reduction, using principal component analysis. Principal component analysis is a technique that allows identifying how different numerical variables vary together, based on the principle of combining multiple numerical variables into a smaller variables set, which are weighted linear combinations of the original set. To validate the proposed methodology, real data were used and some recent known real cases of occurrences in the national interconnected system were identified, so that the post-fault analysis could detect an anomaly. The validation of the obtained results was done using an anomaly detection algorithm using a Gaussian model. With the application of the methodology, the possibility of using a smaller data set than the original was proved, while maintaining the characteristics of the attributes. A case was demonstrated in which it was possible to reduce 190 attributes to a set of eight main components while maintaining accuracy in the F1 Score. This automatic data selection methodology benefits post-fault analysis due to the performance gain obtained by reducing the electrical attribute dataset, without significant loss in the accuracy of anomaly detection capability.

Keywords: Principal Component Analysis. Multivariate statistics. Wide-area monitoring. Phasor measurement units.

## LISTA DE FIGURAS

FIGURA 1 - SENOIDAL E A REPRESENTAÇÃO FASORIAL .....	19
FIGURA 2 - ESQUEMA DE PMUs e PDC .....	21
FIGURA 3 - ARQUITETURA EM CAMADAS .....	22
FIGURA 4 - ESTRUTURA DE UMA PMU COM SEUS PRINCIPAIS ELEMENTOS.	24
FIGURA 5 - PASSOS DA ANÁLISE DE COMPONENTES PRINCIPAIS.....	34
FIGURA 6 - ARQUITETURA DO SMFS.....	46
FIGURA 7 - EXEMPLO DE UM DIAGRAMA DE UMA PMU .....	47
FIGURA 8 - FLUXOGRAMA DA METODOLOGIA .....	49
FIGURA 9 - MATRIZ DE CONFUSÃO .....	55
FIGURA 10 - GRÁFICO DE FREQUÊNCIA.....	59
FIGURA 11 - LOCALIZAÇÃO DAS PMUs SELECIONADAS NO PARANÁ.....	61
FIGURA 12 - LOCALIZAÇÃO DAS PMUs EM CURITIBA .....	62
FIGURA 13 - FLUXO DE ANÁLISE DE DADOS .....	85
FIGURA 14 - DADOS PADRONIZADOS .....	99
FIGURA 15 - DADOS DOS COMPONENTES PRINCIPAIS.....	99

## LISTA DE QUADROS

QUADRO 1 - ARTIGOS SELECIONADOS .....	39
QUADRO 2 - COMPARAÇÃO DOS ELEMENTOS OBSERVADOS NOS ARTIGOS .....	43
QUADRO 3 - RELAÇÃO DE ATRIBUTOS DE UMA PMU .....	47

## LISTA DE TABELAS

TABELA 1 - COMPOSIÇÃO DO BANCO DE ARTIGOS.....	37
TABELA 2 - RESUMO DAS MÉTRICAS DOS ARTIGOS .....	38
TABELA 3 - LISTA DE PMUs.....	63
TABELA 4 - GRUPOS DE PMUs .....	63
TABELA 5 - RESULTADOS DA PMU PIL.....	66
TABELA 6 - RESULTADOS DA PMU SMC .....	67
TABELA 7 - RESULTADOS DA PMU GRL.....	68
TABELA 8 - RESULTADOS DA PMU DJP.....	69
TABELA 9 - RESULTADOS DA PMU JGI.....	70
TABELA 10 - RESULTADOS DA PMU SGD.....	71
TABELA 11 - RESULTADOS DA PMU CVN.....	72
TABELA 12 - RESULTADOS DA PMU GUA.....	73
TABELA 13 - RESULTADOS DO CONJUNTO PIL+SMC+GRL .....	74
TABELA 14 - RESULTADOS DO CONJUNTO DJP+JGI+SGD.....	75
TABELA 15 - RESULTADOS DO CONJUNTO PIL+SMC+GRL+DJP+JGI+SGD.....	76
TABELA 16 - RESULTADOS DO CONJUNTO DE TODAS AS PMUs .....	77
TABELA 17 - RESULTADOS DA MÉTRICA F1 SCORE .....	78
TABELA 18 - RESULTADOS DO PCA – DADOS1.....	79
TABELA 19 - RESULTADOS DO PCA – DADOS2.....	80
TABELA 20 - RESULTADOS DO PCA - DADOS3.....	80
TABELA 21 - RESULTADOS DO PCA - DADOS4.....	81
TABELA 22 - QUANTIDADE DE PCs POR CENÁRIO .....	82
TABELA 23 - TAXA DE COMPRESSÃO POR CENÁRIO .....	83
TABELA 24 - DADOS DE EXEMPLO .....	93
TABELA 25 - DADOS PADRONIZADOS .....	94
TABELA 26 - MATRIZ DE CORRELAÇÃO .....	95
TABELA 27 - AUTOVALORES E AUTOVETORES .....	95
TABELA 28 - DADOS PARA CALCULAR O PC1 .....	96
TABELA 29 - DADOS PARA CALCULAR O PC2 .....	97
TABELA 30 - DADOS PARA CALCULAR O PC3 .....	97
TABELA 31 - MATRIZ RESULTANTE .....	98

## LISTA DE ABREVIATURAS OU SIGLAS

EMS	- Energy Management System
GPS	- Global Position System
ICA	- Independent Component Analysis
IEEE	- Institute of Electrical and Electronic Engineers
KNN	- k-nearest neighbors
ONS	- Operador Nacional do Sistema Elétrico
PC	- Componente Principal
PCA	- Principal Component Analysis
PDC	- Phasor Data Concentrator
PMU	- Phasor Measurement Unit
RMS	- Root Mean Square
SCADA	- Supervisory Control and Data Acquisition
SEP	- Sistema Elétrico de Potência
SMFS	- Sistema de Medição Fasorial Sincronizada
SOC	- Second-of-century
SVD	- Singular Value Decomposition
TCP/IP	- Transmission Control Protocol/Internet Protocol
UHE	- Usina Hidroelétrica
UTC	- Universal Time Coordinated
WAC	- Wide Area Control
WAMS	- Wide Area Measurement System

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>14</b>
1.1 CONTEXTO .....	14
1.2 JUSTIFICATIVA .....	15
1.3 OBJETIVOS .....	16
1.4 ESTRUTURA DA DISSERTAÇÃO .....	16
<b>2 FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>18</b>
2.1 SISTEMA DE MEDIÇÃO FASORIAL SINCRONIZADA .....	18
2.1.1 Técnicas de medição.....	19
2.1.1.1 Medição fasorial.....	19
2.1.1.2 Medição Fasorial Sincronizada.....	20
2.1.2 Estrutura física (Arquitetura).....	21
2.1.2.1 Estrutura hierárquica do PDC.....	21
2.1.2.2 Estrutura da PMU .....	23
2.1.3 Estrutura lógica - Dados .....	25
2.1.4 Aplicações .....	28
2.2 ANÁLISE DE COMPONENTES PRINCIPAIS .....	29
2.2.1 Definição de componentes principais.....	30
<b>2.2.2</b> Descrição matemática .....	<b>31</b>
2.2.3 Funcionamento do algoritmo de análise de componentes principais .....	33
2.3 CONSIDERAÇÕES FINAIS .....	36
<b>3 REVISÃO DE LITERATURA</b> .....	<b>37</b>
3.1 PROCESSO DE SELEÇÃO DO PORTFÓLIO BIBLIOGRÁFICO .....	37
3.2 ANÁLISE DA LITERATURA .....	39
3.3 CONSIDERAÇÕES FINAIS .....	42
<b>4 MATERIAL E MÉTODOS</b> .....	<b>45</b>
4.1 MATERIAIS .....	45
4.1.1 Recursos Computacionais.....	45
4.1.2 Dados .....	45
4.1.2.1 Arquitetura do SMFS .....	45
4.1.2.2 Estrutura dos dados das PMUs .....	46
4.2 MÉTODO.....	48
4.2.1 Fluxograma .....	48

4.2.2	Preparação e carga dos dados .....	50
4.2.3	Algoritmo de detecção de anomalias .....	51
4.2.4	Padronização dos dados .....	52
4.2.5	Algoritmo de análise dos componentes principais.....	53
4.2.6	Detecção de anomalias com os componentes principais .....	54
4.2.7	Métricas para comparação dos resultados.....	55
<b>5</b>	<b>APRESENTAÇÃO E ANÁLISE DOS RESULTADOS .....</b>	<b>58</b>
5.1	DESCRIÇÃO DOS CENÁRIOS.....	58
5.1.1	Descrição do Caso .....	58
5.1.2	Abordagem 1 – Quantidade de PMUs.....	62
5.1.3	Abordagem 2 – Alimentação do PCA.....	63
5.2	ANÁLISE DOS RESULTADOS .....	64
5.2.1	Resultados obtidos .....	64
5.2.2	Eixo de análise 1 – Quantidade de PMUs .....	78
5.2.3	Eixo de análise 2 – Alimentação do PCA .....	79
5.2.4	Redução da dimensionalidade .....	81
5.3	DISCUSSÃO .....	83
<b>6</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS .....</b>	<b>86</b>
	<b>REFERÊNCIAS.....</b>	<b>89</b>
	<b>APÊNDICE 1 – EXEMPLO DE EXECUÇÃO DE PCA.....</b>	<b>92</b>

# 1 INTRODUÇÃO

## 1.1 CONTEXTO

Os sistemas de medição fasorial sincronizada são elementos essenciais na evolução da operação dos sistemas elétricos de potência (SEP) de grande porte. Estes sistemas são baseados em unidades de medição fasorial (PMU – Phasor Measurement Unit), as quais são capazes de gravar e comunicar dados da dinâmica dos SEPs de forma sincronizada por GPS (Global Positioning System) e com alta taxa de amostragem, composta de até 60 amostras por segundo.

Os sistemas de medição de área ampla (ou em inglês *WAMS - Wide Area Measurement System*) complementam o arcabouço disponível para o controle de sistemas elétricos de áreas amplas (*WAC - Wide Area Control*), que é composto por outros sistemas como o SCADA (*Supervisory Control and Data Acquisition*) e o EMS (*Energy Management System*).

Os sistemas de medição fasorial sincronizada estão sendo amplamente utilizados ao redor do mundo, com um crescimento exponencial nos últimos anos. Nos Estados Unidos desde a década passada já há uma rede continental com mais de 2.000 PMUs ajudando a melhorar a confiabilidade da rede de energia elétrica da América do Norte (OVERHOLT et al., 2015). Da mesma forma, também na China na década passada já havia uma rede composta por mais de 2.400 PMUs cobrindo todas as subestações de 500 kV no país e uma série de importantes usinas de energia e subestações de 220/110 kV (LU et al., 2015). Na Europa continental atualmente existe um sistema de medição fasorial abrangendo quase toda a área territorial dos diversos países. Apesar dos desafios de trocas de dados das PMUs entre os diferentes operadores dos sistemas de transmissão nacionais interconectados, as aplicações com os dados intercambiados permitem que os operadores gerenciem e operem todo o sistema de forma segura e eficiente (SATTINGER; GIANNUZZI, 2015).

No Brasil, a implantação dos sistemas de medição fasorial sincronizada também vem crescendo cada vez mais, seguindo a coordenação e os padrões estabelecidos pelo ONS (Operador Nacional do Sistema Elétrico). E neste aspecto o estado do Paraná, através da COPEL Geração e Transmissão S.A., foi pioneiro na implantação do maior sistema do país (OPERADOR..., 2017).



## 1.2 JUSTIFICATIVA

Os sistemas de medição fasorial sincronizada são fundamentais para a operação do sistema elétrico de grande porte, que consiste na programação, coordenação, supervisão, controle e análise, com a finalidade de garantir seu funcionamento de forma otimizada, confiável, segura e com qualidade (ONS, 2020).

Porém, o sistema elétrico não é imune a falhas, e quando elas acontecem devem ser analisadas detalhadamente para que o problema que a causou não se repita e nem se propague. A etapa de análise após uma falha corresponde a uma pesquisa bastante profunda sobre as causas da mesma. É uma tarefa bastante trabalhosa e que exige grande precisão, onde são utilizados todos os tipos de dados disponíveis (CARDOSO, ROLIM e ZÜRN, 2004). Com o crescimento exponencial dos dados de monitoramento, o diagnóstico de faltas enfrenta enormes desafios ao lidar com os grandes volumes de dados (XU et al., 2017).

Os avanços tecnológicos provocaram um aumento exponencial na disponibilidade de dados, conforme o conceito de *Big Data*. Grandes quantidades de dados estão sendo geradas nas formas estruturadas, não estruturadas e semiestruturadas, o que implica na adoção de novas técnicas de armazenamento, processamento e tratamento dos dados (MOHARM, 2019).

A chamada abordagem analítica de dados (ou *data analytics*, do inglês) é um conceito que se refere ao processo de coletar e examinar conjuntos de dados para tirar conclusões sobre as informações que eles contêm com o auxílio de sistemas e *softwares* inteligentes. Sendo assim, trata-se de uma metodologia que permite integrar os dados de vários sistemas computacionais de operação do sistema elétrico, provendo informações importantes para validar ou refutar modelos, teorias e hipóteses acerca de um problema (GE et al., 2017).

Contudo, apesar deste enorme potencial do conjunto composto pelo aumento crescente na coleta de dados no sistema elétrico de potência com o desenvolvimento da abordagem analítica de dados, pouco se encontrou na literatura pesquisada até o momento (conforme detalhado na seção 3.2 do presente documento), sobre trabalhos que abordam a junção de dados fasoriais sincronizados para seleção e tratamento de múltiplos atributos dos dados de várias PMUs interligadas sistemicamente, de forma a disponibilizá-los para análises posteriores.

### 1.3 OBJETIVOS

O principal objetivo deste trabalho é contribuir para a análise sistêmica pós-falta em sistemas elétricos de grande porte, através da coleta de dados do sistema de medição fasorial sincronizado e do desenvolvimento de uma metodologia de seleção de dados relevantes.

Os objetivos específicos são:

- Mapear e avaliar as bases de dados do sistema de medição fasorial sincronizado disponível na empresa COPEL Geração e Transmissão S.A.;
- Desenvolver uma metodologia automática de seleção de dados para análise pós-falta, utilizando análise de componentes principais;
- Analisar casos de falta em sistemas elétricos de grande porte;
- Analisar os dados selecionados usando um algoritmo de detecção de anomalias; e
- Validar os resultados com casos reais de faltas no sistema na empresa COPEL Geração e Transmissão S.A.

### 1.4 ESTRUTURA DA DISSERTAÇÃO

A estrutura deste trabalho está dividida em seis capítulos, sendo o primeiro, o capítulo introdutório, onde se contextualiza o problema, descreve-se a justificativa da pesquisa e os objetivos do trabalho. No capítulo dois, são discutidos os aspectos do sistema de medição fasorial sincronizada, abordando a técnica de medição fasorial, a forma como ela é sincronizada e a arquitetura da solução do sistema de medição fasorial sincronizada. Nesta solução, é explorada a estrutura lógica e os aspectos dos dados que são obtidos. Ainda no capítulo dois, é apresentada a técnica de análise de componentes principais, que será utilizada para realização da redução da dimensionalidade do conjunto de dados obtidos com o sistema de medição fasorial sincronizada, e assim otimizar o tratamento analítico destes dados.

O capítulo três apresenta a revisão da literatura recente sobre o assunto, em especial a aplicação de técnicas de análise de dados e análise de componentes principais em aplicações com o sistema de medição fasorial sincronizada.

No capítulo quatro são apresentados os materiais e os métodos utilizados para a execução do trabalho. Entre os materiais são descritos os recursos computacionais, bem como os dados e suas características principais. Na sequência o método de execução é exposto através de um fluxograma, o qual é descrito e explicado.

No capítulo cinco os resultados obtidos pela execução do método de seleção de dados são apresentados e analisados. O caso real de falta que forneceu os dados utilizados para realização das simulações é descrito. Os cenários de simulações avaliados são apresentados, assim como os resultados obtidos para todas as simulações. Além disso, é realizada a análise destes resultados e apresentada uma discussão baseada nestes resultados obtidos. Por fim, o capítulo seis apresenta as conclusões do trabalho desenvolvido e algumas proposições para trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 SISTEMA DE MEDIÇÃO FASORIAL SINCRONIZADA

O SCADA (*Supervisory Control And Data Acquisition*) e o EMS (*Emergency Management System*) são ferramentas que auxiliam o agente operador do sistema de potência em seu esforço para otimizar a operação do sistema elétrico, com relação à economia, segurança operacional e robustez, além da segurança humana e dos materiais. As ações do agente operador são preferencialmente preventivas, ou seja, são ações tomadas para ajustar as condições operacionais do sistema de potência para o presente e uma situação esperada para um futuro próximo. Ações preventivas, com base em critérios simples, podem ser implementadas no SCADA ou no EMS e ser executadas automaticamente ou servirem de sugestão para uma tomada de decisão por parte do agente operador. Porém, os sistemas SCADA e EMS normalmente não são rápidos o suficiente para capturar toda a dinâmica do sistema de potência (KARLSSON et al., 2004).

O controle de ângulo, por exemplo, é mais preciso se for baseado em PMUs, pois sem estas o fluxo de potência é um método indireto de medir e controlar o ângulo. Uma PMU é um dispositivo para medição sincronizada de tensões e correntes alternadas, com uma referência de tempo (ângulo) comum. A referência de tempo mais comum é o sinal do sistema de posicionamento global (GPS), que tem uma precisão melhor que 1  $\mu$ s. Desta forma, as grandezas de corrente alternada podem ser medidas, convertidas em fasores (números complexos representados por sua magnitude e ângulo de fase) e com marcação de tempo (KARLSSON et al., 2004).

A tecnologia de fasores sincronizados no tempo tem a sua história se iniciando a partir de um componente chave, o desenvolvimento da tecnologia de unidade de medição fasorial (PMU) na década de 1980. Este trabalho iniciou-se na American Electric Power Service Corporation, com a liderança principal de Arun Phadke e Jim Thorp, tendo posteriormente a equipe se transferido para a Virginia Tech. A tecnologia em si envolve o cálculo de medições fasoriais em tempo real sincronizadas com uma referência de tempo absoluta, fornecida pelo Sistema de Posicionamento Global (GPS), para fornecer informações instantâneas e precisas sobre a condição do sistema elétrico devidamente monitorado, permitindo assim que ações corretivas imediatas ocorram (PHADKE, 2015).

## 2.1.1 Técnicas de medição

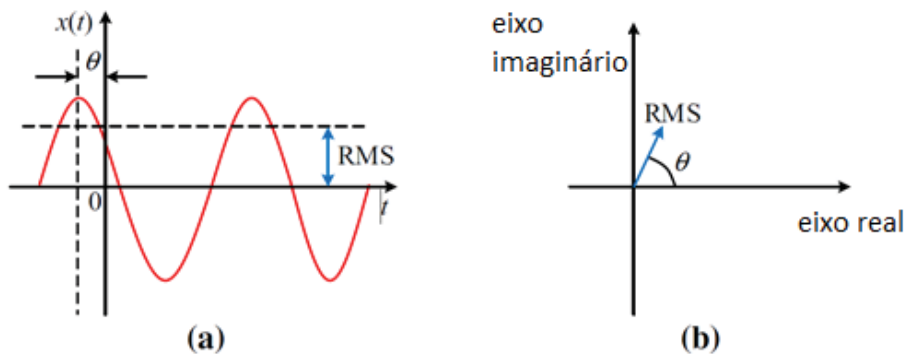
### 2.1.1.1 Medição fasorial

O objetivo da técnica de estimação fasorial é simplesmente obter a representação fasorial. Amostras de dados em forma de onda são coletadas em um subconjunto de dados que normalmente corresponde a um período da frequência fundamental do sistema de energia. Uma resposta senoidal fornecida pela equação (1), com sua respectiva representação fasorial fornecida pela equação (2) é graficamente demonstrada na FIGURA 1.

$$x(t) = X_m \cos(\omega t + \theta) \quad (1)$$

$$X = \left( \frac{X_m}{\sqrt{2}} \right) e^{j\theta} = \left( \frac{X_m}{\sqrt{2}} \right) (\cos \theta + j \operatorname{sen} \theta). \quad (2)$$

FIGURA 1 - SENOIDAL E A REPRESENTAÇÃO FASORIAL



FONTE: Adaptado de PHADKE (2018)

Utilizando-se as amostras ao longo de um período da frequência nominal do sistema de potência, o fasor é estimado pela Transformada Discreta de Fourier, expressa na equação (3) (PHADKE; BI, 2018):

$$\bar{X} = X_r + jX_i = \frac{\sqrt{2}}{N} \sum_{n=1}^N \left( x_n \cos \frac{2n\pi}{N} + jx_n \sin \frac{2n\pi}{N} \right) \quad (3)$$

Onde  $X_r$  e  $X_i$  são partes reais e imaginárias do fasor real;  $x_n$  é a amostra de dados; e  $N$  é o número de amostras em um período. O fasor tem a magnitude do valor da raiz quadrada média (RMS) do sinal de entrada, e o ângulo de fase é o ângulo entre o instante em que o sinal é medido e o pico do sinal de entrada.

Se a frequência do sinal de entrada for diferente da frequência nominal, uma correção de magnitude e ângulo deve ser aplicada para levar em conta o desvio de frequência. Além disso, presume-se que para determinar a estabilidade o sinal de entrada foi filtrado de acordo com o requisito do critério de *Nyquist* para a taxa de amostragem selecionada (PHADKE; BI, 2018).

#### 2.1.1.2 Medição Fasorial Sincronizada

Fasor é uma representação complexa do sinal senoidal puro, que contém amplitude ( $A$ ) e ângulo ( $\theta$ ). O ângulo de fase depende da escolha do eixo  $t = 0$ , que indica a referência de tempo. Ao sincronizar os processos de amostragem para diferentes sinais que podem estar separados por centenas de quilômetros, é possível colocar seus fasores no mesmo diagrama fasorial, atendendo ao propósito de comparação e análise (PHADKE; BI, 2018).

O uso mais comum do GPS (Sistema de Posicionamento Global) é determinar as coordenadas geográficas, porém, para a PMU o elemento mais importante é o pulso por segundo transmitido pelo GPS. Este pulso que é recebido por qualquer receptor na terra é coincidente com todos os outros pulsos recebidos dentro de 1us. Esses pulsos são fornecidos pelos sinais de relógio precisos que são mantidos pelo GPS e esses sinais de relógio são conhecidos como a hora do GPS, que não leva em consideração a rotação da Terra. As correções na hora do GPS são feitas em seus respectivos receptores para compensar essa diferença, de modo que forneçam uma hora universal coordenada. Deve-se notar que por padrão as PMUs usam a base de tempo UNIX com um contador de SOC (*second-of-century*) que começou a contar à meia-noite de 1º de janeiro de 1970 (PHADKE; BI, 2018).

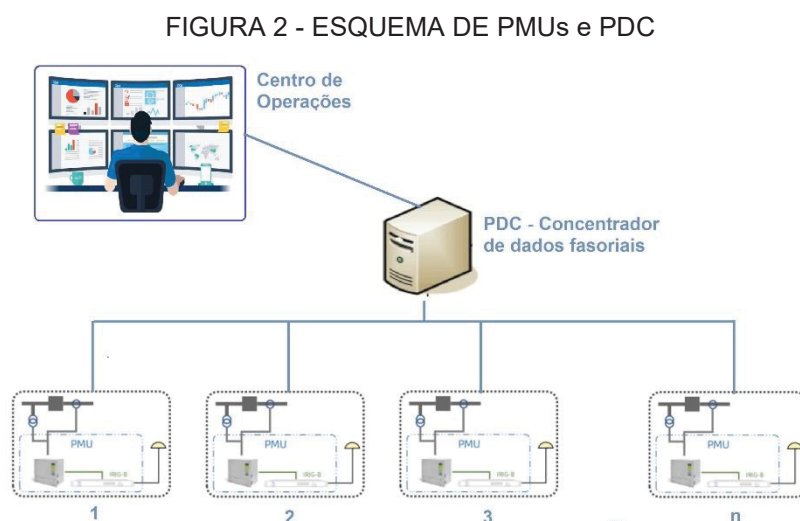
## 2.1.2 Estrutura física (Arquitetura)

As unidades de medição fasorial são instaladas em subestações do sistema de potência. Na maioria das aplicações, os dados fasoriais são usados em locais remotos das PMUs. Portanto, uma arquitetura envolvendo PMUs, *links* de comunicação e concentradores de dados deve existir para obter todos os benefícios do sistema de medição de uma PMU (PHADKE; THORP, 2017).

### 2.1.2.1 Estrutura hierárquica do PDC

Uma vez que os requisitos para monitoramento e controle de uma área ampla podem variar de uma concessionária de energia para outra, a arquitetura para tal sistema deve ser projetada de acordo com as tecnologias que a concessionária de energia possui no momento. Além disso, para evitar se tornar obsoleto, o *design* deve ser escolhido para se ajustar ao direcionamento tecnológico que a concessionária em questão planeja seguir (KARLSSON et al., 2004).

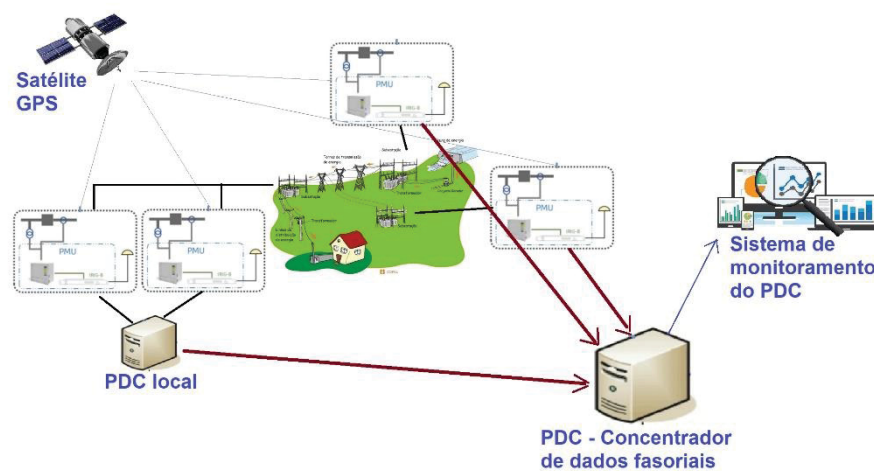
A aplicação mais comum baseada em PMUs é o WAMS (Wide Area Measurement System), ou sistema de medição em área ampla. Sua finalidade é melhorar a estimativa de estado, a análise pós-falha e as informações do agente operador. Nas aplicações WAMS, várias PMUs são conectadas a um concentrador de dados fasoriais (PDC), com um armazenamento em massa acessível a partir da central de controle, conforme a FIGURA 2 (KARLSSON et al., 2004).



FONTE: Adaptado de KARLSSON (2004)

Um PDC é um dispositivo (*hardware e software*) que pode receber fluxos de dados de PMUs ou de outros PDCs, calcular novos sinais, arquivar dados e enviar fluxos de dados para outros PDCs. Uma série de PDCs locais podem ser integrados, juntamente com PMUs em uma solução abrangente de um sistema maior com um sistema de PDC no topo, localizado, por exemplo, no centro de controle; conforme a FIGURA 3 (KARLSSON et al., 2004).

FIGURA 3 - ARQUITETURA EM CAMADAS



FONTE: Adaptado de KARLSSON (2004)

A FIGURA 3 mostra as PMUs situadas em subestações fornecendo medições de tensões e correntes de sequência positiva com registro de tempo de todos os barramentos e alimentadores monitorados (bem como a frequência e a taxa de mudança de frequência). As medições são armazenadas em dispositivos de armazenamento de dados locais, que podem ser acessados de locais remotos para fins de *post mortem* ou diagnóstico. A capacidade de armazenamento local é necessariamente limitada, e os dados armazenados pertencentes a um evento interessante do sistema de energia devem ser sinalizados para armazenamento permanente para que não sejam sobrescritos quando a capacidade de armazenamento local se esgotar. Os dados fasoriais também estão disponíveis para as aplicações em tempo real em um fluxo constante assim que as medições são feitas. Pode haver algumas tarefas de aplicativos locais que requeiram dados de PMU; nesse caso, eles podem ser disponibilizados localmente (PHADKE; THORP, 2017).



Os dispositivos no próximo nível da hierarquia são comumente conhecidos como PDCs. A função típica dos PDCs é reunir dados de várias PMUs, rejeitar dados inválidos, alinhar os registros de tempo e criar um registro coerente de dados gravados simultaneamente de uma parte mais ampla do sistema de energia. Existem instalações de armazenamento locais nos PDCs, bem como funções de aplicativos que precisam dos dados de PMU disponíveis no PDC. Isso pode ser disponibilizado pelos PDCs para os aplicativos locais em tempo real. Perceptivelmente, a comunicação e o gerenciamento de dados nos PDCs criam uma maior latência nos dados em tempo real, mas a experiência prática mostra que isso pode ser controlado (PHADKE; THORP, 2017).

Dependendo da abrangência do sistema sendo monitorado, pode-se considerar mais um nível na hierarquia: um super concentrador de dados. As funções neste nível são semelhantes às aquelas nos níveis de PDC, ou seja, há facilidade para armazenamento de dados alinhados com marcas de tempo (em uma latência de dados um pouco aumentada), bem como um fluxo constante de dados quase em tempo real. Trata-se de um PDC que centraliza dados de outros PDCs ou até mesmo diretamente de algumas PMUs (PHADKE; THORP, 2017).

Portanto, de forma resumida pode-se entender a estrutura hierárquica como sendo composta por três camadas:

- Camada de base: PMUs nas subestações;
- Camada intermediária: PDC concentrando os dados de um conjunto de PMUs; e
- Camada superior: Super PDC concentrando dados de outros PDCs ou diretamente de PMUs.

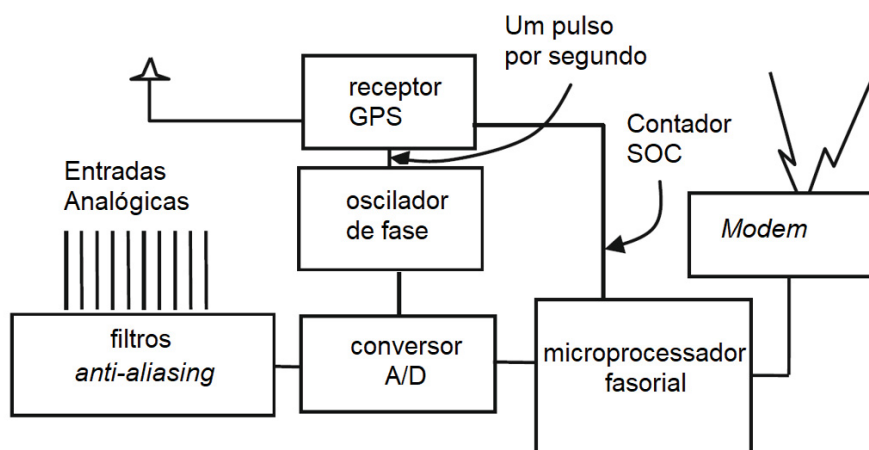
#### 2.1.2.2 Estrutura da PMU

As unidades de medição fasorial feitas por diferentes fabricantes diferem umas das outras em muitos aspectos importantes. Portanto, é difícil discutir a configuração do *hardware* de uma PMU de uma forma que seja universalmente aplicável. No entanto, é possível avaliar a estrutura de uma PMU genérica, que irá capturar a essência de seus componentes principais (PHADKE; THORP, 2017).

A FIGURA 4 é baseada na configuração das primeiras PMUs construídas na Virginia Tech, que mostra os principais componentes de uma PMU. Considerando-se

que estas PMUs evoluíram a partir do desenvolvimento do relé de distância de componente simétrico, conseqüentemente, a estrutura mostrada na FIGURA 4 é similar a de um relé de computador. As entradas analógicas são correntes e tensões obtidas dos enrolamentos secundários dos transformadores de corrente e tensão. Todas as correntes e tensões trifásicas são usadas para que a medição de sequência positiva possa ser realizada. Em contraste com um relé, uma PMU pode ter correntes em vários alimentadores originados na subestação e tensões pertencentes a vários barramentos na subestação (PHADKE; THORP, 2017).

FIGURA 4 - ESTRUTURA DE UMA PMU COM SEUS PRINCIPAIS ELEMENTOS



FONTE: Adaptado de PHADKE; THORP (2017)

Os sinais de corrente e tensão são convertidos em transformadores para tensões na faixa apropriada, normalmente na faixa de  $\pm 10$  V, para que correspondam aos requisitos dos conversores analógico-digital. A taxa de amostragem escolhida para o processo de amostragem determina a resposta de frequência dos filtros *anti-aliasing*, também conhecidos como filtro analógico passa-baixas, necessários quando a frequência da amostragem não atende o critério de Nyquist, eles limitam a máxima frequência do sinal à metade da frequência de amostragem. Como em muitos projetos de relé, pode-se usar uma alta taxa de amostragem (chamada de sobreamostragem) em conjunto com a alta frequência de corte correspondente dos filtros *anti-aliasing* analógicos. Esta etapa é então seguida por um filtro de decimação digital que converte os dados amostrados para uma taxa de amostragem mais baixa, fornecendo assim um filtro *anti-aliasing* digital concatenado com os filtros *anti-aliasing* analógicos. A vantagem de tal esquema é que os filtros *anti-aliasing* eficazes compostos de um

*frontend* analógico e um filtro de decimação digital são muito mais estáveis no que diz respeito ao envelhecimento e variações de temperatura. Isso garante que todos os sinais analógicos tenham o mesmo deslocamento de fase e atenuação, garantindo assim que as diferenças de ângulo de fase e magnitudes relativas dos diferentes sinais não sejam alteradas (PHADKE; THORP, 2017).

Como um benefício adicional da técnica de sobreamostragem, se houver a possibilidade de armazenar dados brutos de amostras dos sinais analógicos, eles podem ser de grande utilidade como gravadores digitais de falha de largura de banda alta. O relógio de amostragem é travado em fase com o pulso do relógio GPS. As taxas de amostragem têm aumentado constantemente ao longo dos anos, começando com uma taxa de 12 amostras por ciclo da frequência de potência nominal nas primeiras PMUs, até 96 ou 128 amostras por ciclo em dispositivos mais modernos, como conversores analógico-para-digital mais rápidos e microprocessadores, que tem se tornado cada vez mais comuns (PHADKE; THORP, 2017).

O microprocessador calcula estimativas de sequência positiva de todos os sinais de corrente e tensão. Algumas outras estimativas de interesse são a frequência e a taxa de variação da frequência medida localmente e que também estão incluídas na saída da PMU. A estampa de tempo é criada a partir de dois dos sinais derivados do receptor GPS, e identifica a identidade dos segundos do UTC (*Universal Time Coordinated*) e o instante que define o limite de um dos períodos de frequência de potência conforme definido na norma IEEE (PHADKE; THORP, 2017).

Finalmente, a principal saída da PMU é a medição com estampa de tempo a ser transferida pelos *links* de comunicação por meio de *modems* adequados para um nível superior na hierarquia do sistema de medição (PHADKE; THORP, 2017).

### 2.1.3 Estrutura lógica - Dados

Os dados da PMU se originam no ponto de medição e passam por um processo e caminho de comunicação até o seu ponto de uso. Problemas que surgem ao longo deste caminho podem afetar o desempenho de vários aplicativos que fazem uso dos dados de fasores sincronizados e, portanto, comprometem a capacidade desses aplicativos de atender às necessidades e objetivos dos usuários de forma eficaz (NASPI, 2017).

O termo “qualidade de dados” é abrangente e deve ser interpretado como sendo composto pela: precisão; conteúdo; adequação para uso; e disponibilidade de dados. O termo “disponibilidade do fluxo de dados” deve ser interpretado como um termo que diz respeito a se os dados relevantes chegam ao aplicativo, mesmo que não a tempo para um uso funcional (NASPI, 2017).

A definição de quais são os atributos de dados mais importantes depende do problema de interesse a ser resolvido e do aplicativo projetado para resolvê-lo. Na maioria dos casos, o analista deseja realizar uma análise específica, e para isto identifica os dados necessários para realizá-la e determina se os dados disponíveis são adequados para esse fim. Em alguns casos, o analista recebe um novo conjunto de dados e busca quais conhecimentos podem ser adquiridos, considerando a qualidade dos dados disponíveis como um dado (NASPI, 2017).

Um aplicativo que recebe dados precisos, de alta qualidade e altamente disponíveis provavelmente produzirá respostas analíticas mais informativas e utilizáveis. Por outro lado, um aplicativo que recebe dados com falhas devido a imprecisões, lacunas e atrasos provavelmente produzirá respostas limitadas por maior incerteza ou deixará de produzir respostas (NASPI, 2017).

Adequação para uso significa que os dados fornecem informações e conteúdo suficientes para que um aplicativo forneça resultados significativos e confiáveis; se um aplicativo produz resultados questionáveis devido a deficiências nos dados de entrada, os dados não são adequados para uso (NASPI, 2017).

As deficiências de dados podem ser categorizadas com base na fonte e no tipo. Esta é apenas uma dessas categorizações; outras podem ser usadas e são encontradas na literatura. As seis categorias a seguir se alinham bem com o que os usuários experimentam e que são fáceis de contabilizar na prática. Essas categorias de deficiências de dados são (MARTIN, CHEN, 2018):

- perda de dados: é definida como dados planejados e esperados por um determinado aplicativo, mas não recebidos. Isso inclui perdas na criação de dados (medições de PMU), comunicação de dados e dispositivos de armazenamento de dados. A perda de dados afeta todos os aspectos do uso de dados;
- desordenação de dados: refere-se aos dados que foram recebidos pelo aplicativo final, mas não representam a medição de uma forma significativa. Normalmente, os dados são corrompidos no subsistema de comunicação, mas

a corrupção pode ocorrer no ponto de medição ou também no aplicativo de uso final. A corrupção pode fazer com que os dados mudem para qualquer valor aleatoriamente e, portanto, pode produzir erros imprevisíveis;

- representação imprecisa: inclui todos os erros relacionados à precisão que fazem com que o valor de engenharia represente incorretamente a quantidade de engenharia. A fonte da imprecisão pode variar do transdutor original ao dispositivo de exibição / gravação. Escala, linearidade, algoritmos de estimativa, tempo, interferência, ruído e erros de fator de correção podem criar imprecisão. Dados imprecisos podem ser difíceis de detectar, pois podem resultar de dispositivos de transformação, fatores de escala inadequados e erros de processamento que são pequenos o suficiente para passar como dados bons, mas grandes o suficiente para serem mal interpretados;
- falta de precisão: refere-se à precisão com que a medição pode ser resolvida. Está diretamente relacionado ao número de dígitos numéricos que o representam. Se houver mais dígitos, a representação é mais precisa, mesmo que representem apenas ruído. A característica mais óbvia da medição de baixa precisão é uma aparência de forma de onda granular. Isso acontece porque as etapas discretas nos valores de medição são visíveis dentro da faixa do gráfico. A faixa combinada com a resolução pode mostrar ou ocultar as etapas de medição discretas. Por exemplo, um gráfico de uma forma de onda de um sinal de 500 kV pode parecer irregular quando traçado com uma faixa de 2 kV, mas parece suave quando traçado com uma faixa de 200 kV. Embora isso possa ocorrer com qualquer representação de sinal digital que seja ampliada o suficiente, só se torna um problema quando as etapas ocorrem dentro da faixa de análise;
- identificação incorreta de dados: os sinais são medidos em uma subestação e os valores são enviados para uma central de controle onde são processados e registrados. Em muitos casos, os valores são posteriormente enviados para outros centros de controle ou monitoramento. Os dados só têm significado quando são identificados com a quantidade de engenharia que representam. A identificação correta deve ser aplicada em muitos lugares, como exibições em um diagrama do sistema, gráficos de dados de série temporal e dispositivos de armazenamento de dados que serão usados para análise posterior. Dados mal identificados levam a erros em cada um desses processos. A identificação

incorreta geralmente é resolvida durante o processo de instalação e validação. No entanto, em um caso típico, as alterações em outros sistemas, atualizações de *software* e outras manutenções do sistema podem resultar em alterações inadvertidas no sistema de fasor sincronizado. Isso pode resultar em erros de identificação de sinal; e

- latência excessiva ou inconsistente: os dados dos fasores sincronizados são medidos e relatados sem demora para permitir seu uso em aplicativos em tempo real. Alguns aplicativos, como telas de quadro de mapas, podem não exigir atualizações muito rapidamente, mas os controles baseados em medições geralmente requerem muito pouco atraso para operar com eficácia. Atraso excessivo ou variável, denominado latência, pode degradar significativamente o valor da medição. A primeira atividade que a latência afeta é a coleta de dados. Um PDC que coleta dados para alinhamento de tempo terá que esperar pelo último quadro correspondente a cada marcação de horário antes de completar o quadro agregado e encaminhá-lo. Todos os dados com latência excessiva são perdidos se o atraso exceder o tempo de espera do PDC. Da mesma forma, mesmo sem um PDC, os dados enviados para um monitor ou aplicativo analítico ficam inutilizáveis se o aplicativo não esperar o tempo suficiente por eles. Os aplicativos do centro de controle devem apresentar as informações aos operadores em um intervalo de tempo que seja útil para eles, que geralmente é de segundos a alguns minutos em alguns casos.

#### 2.1.4 Aplicações

Com o surgimento e o desenvolvimento da tecnologia de fasores sincronizados, as aplicações potenciais de PMUs e WAMS no monitoramento, operação, proteção e controle de sistemas de potência passaram a ser exploradas. As aplicações incluem validação de modelo, estimativa de estado, proteção e controle de malha fechada com medições de fasores sincronizados, entre outros (PHADKE; BI, 2018).

A visualização pura e simples de dados brutos de fasores sincronizados em tempo real pode ser uma maneira eficaz de fornecer a consciência situacional de uma área ampla. No entanto, os aplicativos e análises para uso em tempo real e *offline*

permitem expandir o valor dos fasores sincronizados (JONES et al., 2015).

Há muitas aplicações existentes e potenciais da tecnologia de medição sincronizada, variando de ferramentas de monitoramento e consciência situacional a esquemas de proteção da integridade do sistema. As aplicações incluem linha de base de desempenho do sistema, análise de eventos, validação de modelo, detecções de instabilidade e oscilação, gerenciamento de energia reativa, restauração, determinação de harmônicos, localização de falha e análise de transientes rápida, todos oferecendo confiabilidade significativa e benefícios financeiros para o sistema elétrico de potência de grande porte (MANDANI et al., 2015).

A análise *offline* normalmente se refere a aplicações que não são sensíveis ao tempo, incluindo, mas não se limitando à análise pós-evento, análise estatística, análise de qualidade de dados, validação de modelo de sistema, redução de modelo, impacto social, estudos forenses e outros. Essas aplicações geralmente requerem extração de grande volume de dados do historiador, análise de dados no domínio temporal ou espacial e comparação ou integração com dados de outras fontes. Para análises estatísticas, técnicas de análise de *big data* e processamento paralelo são introduzidos para acelerar a velocidade de computação (ZHOU et al., 2016)

## 2.2 ANÁLISE DE COMPONENTES PRINCIPAIS

O PCA (*Principal Components Analysis*) é uma das ferramentas mais amplamente utilizadas e úteis no campo da análise exploratória de dados. Oferece uma visão geral do assunto em questão, mostrando a relação que existe entre objetos, bem como entre objetos e variáveis. O ponto forte do PCA é que ele fornece a oportunidade de visualizar dados em referência a objetos descritos por mais de três variáveis (MONFREDA, 2012). A utilização do PCA permite estudar e compreender tais sistemas, viabilizando uma abordagem visual por permitir mostrar em sistemas de apenas duas ou três dimensões que, de outra forma, necessariamente teriam que ser vistos em mais de três dimensões para serem estudados. O PCA permite que os dados mantenham sua estrutura original, fazendo apenas uma rotação ortogonal das variáveis, o que ajuda a simplificar a visualização de todas as informações já contidas nos dados. Consequentemente, PCA pode ser considerada a melhor técnica para começar a abordar qualquer problema multivariado qualitativo, seja ele não supervisionado ou supervisionado (MONFREDA, 2012)

### 2.2.1 Definição de componentes principais

A técnica de análise de componentes principais foi inicialmente descrita por Karl Pearson (1901). Ele propôs um método prático de cálculo para duas ou três variáveis apenas. Uma descrição de métodos computacionais práticos veio muito mais tarde, feita por Hotelling (1933). Mesmo então, os cálculos eram extremamente desafiadores, pois apesar de simples eram trabalhosos mesmo que para mais do que poucas variáveis, visto que tinham que ser feitos à mão. Somente após os computadores eletrônicos terem se tornado disponíveis de uma forma mais acessível é que ampliou-se o uso da técnica de componentes principais (MANLY; ALBERTO, 2019).

A análise de componentes principais é um dos métodos multivariados mais simples. O objetivo da análise é tomar  $p$  variáveis  $X_1, X_2, \dots, X_p$  e encontrar combinações destas para produzir índices  $Z_1, Z_2, \dots, Z_p$ , que sejam não correlacionados na ordem de sua importância, e que descrevam a variação nos dados. A falta de correlação significa que os índices estão medindo diferentes “dimensões” dos dados, e a ordem é tal que  $\text{Var}(Z_1) \geq \text{Var}(Z_2) \geq \dots \geq \text{Var}(Z_p)$ , em que  $\text{Var}(Z_i)$  denota a variância de  $Z_i$ . Os índices  $Z$  são, então, os componentes principais. Ao fazer uma análise de componentes principais, há sempre a esperança de que as variâncias da maioria dos índices serão tão baixas a ponto de serem desprezíveis. Neste caso, a maior parte da variação no conjunto de dados completos pode ser descrita adequadamente pelas poucas variáveis  $Z$  com variâncias que não são desprezíveis, obtendo-se a possibilidade de uma economia, seja de armazenamento, processamento ou tempo, pela simples redução da quantidade de variáveis a serem manipuladas (MANLY; ALBERTO, 2019).

A análise de componentes principais nem sempre funciona, no sentido de que um grande número de variáveis originais é reduzido a um pequeno número de variáveis transformadas. De fato, se as variáveis originais são não correlacionadas, então a análise não chega a nada. Os melhores resultados são obtidos quando as variáveis originais são altamente correlacionadas, positiva ou negativamente. Se este é o caso, então é bastante concebível que vinte ou mais variáveis originais possam ser adequadamente representadas por duas ou três componentes principais. Se este estado desejável de relações de fato ocorre, então os componentes principais



importantes realmente serão boas representações do conjunto de dados original. Será também uma indicação de que há uma boa quantidade de redundância nas variáveis originais, com a maioria delas medindo coisas semelhantes (MANLY; ALBERTO, 2019).

Um exemplo de aplicação da análise de componentes principais pode ser descrito a partir de um conjunto de dados de grandezas elétricas. As  $p$  variáveis  $X_1, X_2, \dots, X_p$  podem ser representadas por grandezas como frequência, tensão nas fases A, B e C, corrente nas fases A, B e C de vários equipamentos diferentes. O objetivo é encontrar as combinações destas variáveis, para produzir novos índices  $Z_1, Z_2, \dots, Z_p$ , que descrevam a variação nos dados. Estes novos índices, nada mais são do que novas variáveis calculadas para substituir as variáveis originais. Neste exemplo, se houver grande correlação entre estas grandezas, elas poderão ser melhor representadas por uma quantidade menor de novas variáveis transformadas.

### 2.2.2 Descrição matemática

Algebricamente, componentes principais são combinações lineares particulares das  $p$  variáveis randômicas  $X_1, X_2, \dots, X_p$ . Geometricamente, estas combinações lineares representam a seleção de um novo sistema de coordenadas obtido pela rotação do sistema original com  $X_1, X_2, \dots, X_p$  como os eixos de coordenadas. Os novos eixos representam as direções com variabilidade máxima e fornecem uma descrição mais simples e parcimoniosa da estrutura de covariância (JOHNSON; WICHERN, 2007).

Os componentes principais dependem exclusivamente da matriz de covariância  $\Sigma$  (ou da matriz de correlação) de  $X_1, X_2, \dots, X_p$ .

Seja o vetor  $X' = [X_1, X_2, \dots, X_p]$  com matriz de covariância  $\Sigma$ , cujos autovalores são  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  e os autovetores associados  $a_1, a_2, \dots, a_p$ .

Considerem-se as combinações lineares:

$$Z_1 = a'_1 X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad (4)$$

$$Z_2 = a'_2 X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \quad (5)$$

⋮

$$Z_p = a'_p X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p. \quad (6)$$

Aplicando-se os conceitos de variância e covariância, tem-se:

$$Var(Z_i) = a_i' \Sigma a_i \quad i = 1, 2, \dots, p \quad (7)$$

$$Cov(Z_i, Z_k) = a_i' \Sigma a_k \quad i, k = 1, 2, \dots, p. \quad (8)$$

Os componentes principais são as combinações lineares  $Z_1, Z_2, \dots, Z_p$  não correlacionadas, cujas variâncias são a maior possível.

O primeiro componente principal é a combinação linear com variância máxima. Isto é, aquela que maximiza  $Var(Z_i) = a_i' \Sigma a_i$ . Fica claro que  $Var(Z_i) = a_i' \Sigma a_i$ , pode ser incrementado multiplicando-se qualquer  $a_i$  por alguma constante.

Para eliminar essa indeterminação, é conveniente restringir a atenção ao vetor de coeficientes de comprimento unitário. Portanto, define-se que:

o primeiro componente principal é a combinação linear  $a_1'X$  que maximiza  $Var(a_1'X)$  sujeito a restrição da equação (10)

$$a_1' a_1 = 1 \quad (9)$$

o segundo componente principal é a combinação linear  $a_2'X$  que maximiza  $Var(a_2'X)$  sujeito à restrição das equações (10) e (11) .

$$a_2' a_2 = 1 \quad (10)$$

$$Cov(a_1'X, a_2'X) = 0. \quad (11)$$

Os componentes principais seguintes são definidos continuando da mesma maneira. Se existem  $p$  variáveis, então existirão no máximo  $p$  componentes principais. Para se usar os resultados de uma análise de componentes principais, não é necessário saber como as equações, para os componentes principais, são obtidas. Entretanto, é útil entender a natureza das equações. De fato, uma análise de componentes principais envolve encontrar os autovalores de uma matriz de covariâncias amostral (MANLY; ALBERTO, 2019).

A matriz de covariâncias é simétrica e tem a forma:

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{13} \\ c_{21} & c_{22} & \dots & c_{23} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_{p1} & c_{p2} & \dots & c_{p3} \end{bmatrix}. \quad (12)$$

Em que o elemento  $c_{ii}$  na diagonal é a variância de  $X_i$ , e os termos fora da diagonal  $c_{ij} = c_{ji}$  são a covariância entre as variáveis  $X_i$  e  $X_j$ .

As variâncias dos componentes principais são os autovalores da matriz  $C$ . Existem  $p$  destes autovalores, alguns dos quais podem ser zero. Autovalores negativos não são possíveis para uma matriz de covariâncias. Assumindo que os autovalores estão ordenados como  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , então  $\lambda_1$  corresponde ao

i-ésimo componente principal

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p. \quad (13)$$

Em particular,  $Var(Z_i) = \lambda_i$ , e as constantes  $a_{i1}, a_{i2}, \dots, a_{ip}$  são os elementos do correspondente autovetor, escalonado de modo que

$$a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2 = 1. \quad (14)$$

Uma propriedade importante dos autovalores é que a soma deles é igual à soma dos elementos da diagonal (o traço) da matriz C. Isto é,

$$\lambda_1 + \lambda_2 + \dots + \lambda_p = c_{11} + c_{22} + \dots + c_{pp}. \quad (15)$$

porque  $c_{ii}$  é a variância de  $X_i$  e  $\lambda_i$  é a variância de  $Z_i$ , isso significa que a soma das variâncias dos componentes principais é igual à soma das variâncias das variáveis originais. Portanto, os componentes principais contam com toda a variação nos dados originais.

A fim de evitar uma ou duas variáveis tendo uma indevida influência nos componentes principais, é usual codificar as variáveis  $X_1, X_2, \dots, X_p$ , para terem médias zero e variâncias um no início de uma análise. A matriz C então toma a forma

$$C = \begin{bmatrix} 1 & c_{12} & \dots & c_{13} \\ c_{21} & 1 & \dots & c_{23} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_{p1} & c_{p2} & \dots & 1 \end{bmatrix}. \quad (16)$$

Em que  $c_{ij} = c_{ji}$  é a correlação entre  $X_i, X_j$ . Em outras palavras, a análise de componentes principais é feita sobre a matriz de correlação. Neste caso, a soma dos termos da diagonal, e portanto, a soma dos autovalores, é igual p, o número de variáveis X.

### 2.2.3 Funcionamento do algoritmo de análise de componentes principais

A ideia central da análise baseia-se na redução do conjunto de dados a ser analisado, principalmente quando os dados são constituídos de um grande número de variáveis inter-relacionadas.

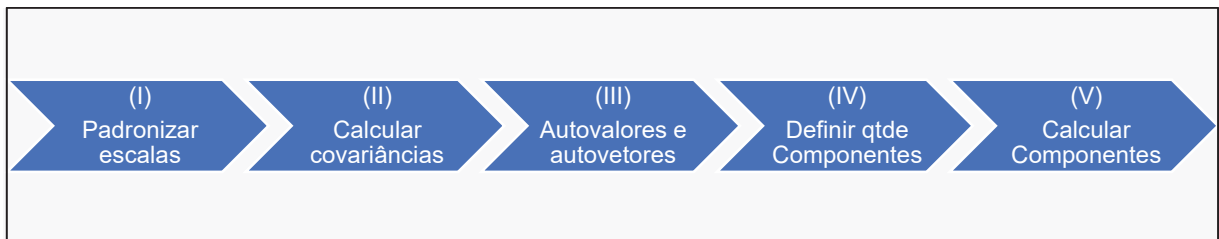
Geralmente as variáveis variam em conjunto (covariam), e algumas variações em uma delas são, na verdade, duplicadas pela variação da outra. A análise de componentes principais é uma técnica para descobrir o modo como as variáveis numéricas covariam. A ideia na PCA é combinar múltiplas variáveis predictoras numéricas em um conjunto menor de variáveis, que são combinações lineares

ponderadas do conjunto original. O menor conjunto de variáveis, os componentes principais, “explica” a maior parte da variabilidade do conjunto completo de variáveis, reduzindo a dimensão dos dados (BRUCE; BRUCE, 2019).

Esta variação em conjunto de algumas variáveis pode ser observada em conjuntos de dados de grandezas elétricas, ao se concatenar dados de leituras de equipamentos diferentes ligados no mesmo sistema elétrico. A concatenação coloca lado a lado as diferentes grandezas de diferentes equipamentos, observados no mesmo instante de tempo. Desta forma, as variáveis de frequência medidas em cada um destes equipamentos são um exemplo de candidatas que poderão vir a ser identificadas como tendo uma variação em conjunto (covariação). Ressalta-se aqui que este exemplo é apresentado tendo-se um conhecimento prévio dos dados, mas uma característica da análise de componentes principais é que esta identificação da covariação é feita sem que se tenha nenhum conhecimento prévio dos dados.

Os passos em uma análise de componentes principais podem agora ser estabelecidos, conforme ilustrado na FIGURA 5.

FIGURA 5 - PASSOS DA ANÁLISE DE COMPONENTES PRINCIPAIS



FONTE: O autor (2022)

Na primeira etapa (I) é feita uma verificação das variáveis do conjunto de dados original, para identificar discrepâncias nos valores, devido ao fato das diferenças grandezas estarem em diferentes escalas de valor. Esta discrepância poderia causar distorções na análise, influenciando o algoritmo a estabelecer relações equivocadas entre as variáveis pelo fato de estarem sendo representadas em escalas diferentes. Portanto, para evitar erros nos resultados deve-se executar uma técnica de padronização, em estatística também chamada de padronização, que coloca todas as variáveis do conjunto em escalas semelhantes através da subtração da média e divisão pelo desvio-padrão.

Na etapa seguinte (II) é efetuado o cálculo da matriz de covariâncias  $C$ , conforme descrito na seção 2.2.2.

Após ter-se obtido a matriz de covariâncias, passa-se para a etapa III do processo, onde deve-se identificar os autovalores  $\lambda_1, \lambda_2, \dots, \lambda_p$  e os correspondentes autovetores  $a_1, a_2, \dots, a_p$ . Os coeficientes do  $i$ -ésimo componente principal são então os elementos de  $a_i$ , enquanto que  $\lambda_i$  é sua variância (MANLY; ALBERTO, 2019). Os pesos usados para formar os componentes principais revelam as contribuições relativas das variáveis originais para os novos componentes principais. Os pesos ( $a_{i1}, a_{i2}$ ) são conhecidos como as cargas dos componentes. Elas transformam as variáveis originais nos componentes principais. O primeiro componente principal,  $Z_1$ , é a combinação linear que melhor explica a variação total (BRUCE; BRUCE, 2019).

Para reduzir a quantidade de variáveis a ser trabalhada no novo conjunto de dados, na etapa IV efetua-se a decisão do número total de componentes que melhor representarão o conjunto de dados original. Esta seleção pode ser feita por diferentes critérios, em que se destacam dois:

- Seleção dos primeiros componentes cuja soma da variância cumulativa exceda um determinado limite, geralmente adotado como 80%;
- Mediante os autovalores, pelo critério sugerido por Kaiser (KAISER, 1960 apud MARDIA, 1979), que consiste em incluir somente aquelas componentes cujos autovalores sejam superiores ou iguais a 1 (um);

Para qualquer que seja o critério adotado, o importante é que após a seleção, devem ser descartados quaisquer componentes que expliquem somente uma pequena proporção da variação de dados. Por exemplo, começando com 20 variáveis, pode ser obtido que os primeiros três componentes expliquem 90% da variância total. Com base nisso, os outros 17 componentes podem ser razoavelmente ignorados (MANLY; ALBERTO, 2019).

A validação cruzada oferece um método mais formal para escolher o número de componentes significativos. A ideia de validação “fora da amostra” não é nova, mas não foi muito aproveitada até que os conjuntos de dados passaram a ter uma maior quantidade de amostras. Normalmente seria usada a maioria dos dados para treinar o modelo, e uma proporção menor para testá-lo (BRUCE; BRUCE, 2019).

Por fim, tendo sido identificados os autovalores e os autovetores e a quantidade de componentes principais (denominada aqui como  $q_{cp}$ ), na última etapa (V) efetua-se o cálculo dos componentes principais propriamente dito. Deve-se

executar os cálculos nas equações algébricas (4),(5) e (6) que fornecem os valores dos componentes principais selecionados na etapa IV, de  $Z_1$  até  $Z_{qcp}$ , para cada uma das amostras do conjunto de dados. Assim obtém-se uma lista de  $qcp$  componentes principais, as novas variáveis do conjunto.

O Apêndice 1 apresenta um exemplo numérico de aplicação do algoritmo de análise de componentes principais.

### 2.3 CONSIDERAÇÕES FINAIS

Este capítulo apresentou alguns aspectos do sistema de medição fasorial sincronizada, com ênfase maior nos aspectos relacionados ao grande volume de dados disponibilizados através da tecnologia de PMUs. As oportunidades de utilização destes dados, descritas na seção 2.1.4, estão intimamente vinculadas com a qualidade dos dados, conforme descrito na seção 2.1.3. Além disso, o grande volume de dados exige a adoção de novas abordagens e diferentes técnicas para sua utilização de forma mais eficaz.

Seguindo este viés, a seção 2.2, aborda o funcionamento de uma técnica com o potencial de viabilizar a redução da dimensionalidade do conjunto de dados, sem perder a essência de suas informações. Considerando que um dos diferenciais dos dados das PMUs é a sua capacidade de coletar muitos dados em um pequeno intervalo de tempo, para que esta característica não seja perdida, a técnica de PCA permite explorar a possibilidade de redução do tamanho da matriz de dados abordando o outro eixo, que refere-se à quantidade de diferentes atributos disponíveis.

### 3 REVISÃO DE LITERATURA

#### 3.1 PROCESSO DE SELEÇÃO DO PORTFÓLIO BIBLIOGRÁFICO

Para formar um portfólio bibliográfico, foi utilizado um processo de revisão sistemática da literatura inspirado em Ensslin et al. (2010), sendo que o primeiro passo é definir dois eixos de pesquisa e suas respectivas palavras-chave associadas para formação do banco de dados bruto.

Para abranger as publicações mais relevantes sobre a abordagem analítica de dados em sistemas elétricos de grande porte, utilizou-se no eixo 1 o termo *Wide Area Monitoring* e no eixo 2 *Data Science*. As palavras-chave escolhidas no primeiro eixo foram: *Phasor Measurement Units; Synchrophasor; PMU e Wide-area Monitoring*. No segundo eixo foram: *Big Data Applications, Multivariate Statistics, Multivariate Analysis, e Data Analytics*. A seguir, a busca é realizada combinando cada palavra-chave do primeiro eixo com as palavras-chave do segundo eixo, utilizando o *software Publish or Perish*, que obtém os metadados das citações brutas no banco de dados do *Google Scholar*. Foi aplicado um filtro para selecionar somente os artigos publicados entre 2010 e 2020, sendo que, no total, o primeiro banco de dados bruto foi composto por 3.890 documentos, cuja composição quantitativa pode ser observada na TABELA 1.

TABELA 1 - COMPOSIÇÃO DO BANCO DE ARTIGOS

	<b>Big data applications</b>	<b>multivariate statistics</b>	<b>Multivariate Analysis</b>	<b>data analytics</b>	<b>Total</b>
<b>Phasor measurement units</b>	66	20	59	809	954
<b>synchrophasor</b>	54	11	42	720	827
<b>PMU</b>	157	57	312	992	1518
<b>Wide-area monitoring</b>	37	26	51	477	591
<b>Total</b>	314	114	464	2998	3890

FONTE: O autor (2022)

A partir deste ponto, para refinamento da lista de artigos, adota-se a seguinte sequência de filtros:

- Eliminação das duplicidades. Identificação dos artigos que contêm no título os termos: *Big Data; Data Analytics; Data Mining; Multivar...; Phasor Measurement; PMU*. Avaliação dos artigos de toda a lista pelos

respectivos títulos, com destaque para os identificados na seleção prévia. Avaliação da publicação dos artigos e eliminação daqueles que têm origem em conferências. Descarte dos artigos que não são de alguma das seguintes Editoras: IEEE, IET, Elsevier, Springer, Scopus, Science Direct, MDPI e Wiley;

- Após esta etapa a lista foi reduzida a 158 artigos. Na sequência foi feita a análise dos resumos de cada um dos artigos, para se chegar numa seleção final de 26 artigos para compor o portfólio bibliográfico, referente a *Wide Area Monitoring* e *Data Science*;
- A última etapa consistiu na leitura de introdução e conclusão dos 26 artigos reduzindo o portfólio a 10.

TABELA 2 - RESUMO DAS MÉTRICAS DOS ARTIGOS

Métrica	Valor
Anos de publicação	5 (2015-2019)
Quantidade de anos com citações	6
Quantidade de artigos	26
Total de citações	880
Citações por ano	146,6667
Citações por artigo	33,84615
Autores por artigo	3,692308

FONTE: O autor (2022)

A TABELA 2 mostra os resultados das métricas obtidos através do *software Publish or Perish*, para a lista dos 26 artigos selecionados. O QUADRO 1 mostra a lista dos 10 artigos finais selecionados.



QUADRO 1 - ARTIGOS SELECIONADOS

Número	Título	Ano	Autores
1	Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis	2014	XIE, L.; CHEN, Y.; KUMAR, P. R.
2	Principal component analysis of wide-area phasor measurements for islanding detection—A geometric view	2015	LIU, X.; LAVERTY, D. M.; BEST, R. J.; LI, K.
3	Data mining and analytics in the process industry: The role of machine learning	2017	GE, Z.; SONG, Z.; DING, S. X.; HUANG, B.
4	Wide-Area Monitoring of Power Systems Using Principal Component Analysis and -Nearest Neighbor Analysis	2018	CAI, L.; THORNHILL, N. F.; KUENZEL, S.
5	Real-Time Application of an Indirect Power-System Contingency Screening Method Based on Adaptive PCA	2019	DIMITROVSKA, T.; RUDEŽ, U.
6	Corruption-Resilient Detection of Event-Induced Outliers in PMU Data: A Kernel PCA Approach	2019	CHATTERJEE, K.; CHAUDHURI, N. R.
7	Massive streaming PMU data modelling and analytics in smart grid state evaluation based on multiple high-dimensional covariance test	2017	CHU, L.; QIU, R.; HE, X.; LING, Z.; LIU, Y.
8	Robust ensemble data analytics for incomplete PMU measurements-based power system stability assessment	2017	ZHANG, Y.; XU, Y.; DONG, Z. Y.
9	A data imputation model in phasor measurement units based on bagged averaging of multiple linear regression	2018	LE, N. T.; BENJAPOLAKUL, W.
10	Statistical Characterization of PMU Error for robust WAMS based analytics	2019	T AHMAD, N SENROY

FONTE: O autor (2022)

### 3.2 ANÁLISE DA LITERATURA

XIE et al. (2014) apresentam um estudo que explora a dimensionalidade implícita nos dados de PMU, propondo as justificativas teóricas para um algoritmo de detecção de eventos, baseado em análise de componentes principais (PCA). A técnica apresentada é um método *online*, orientado por dados, que não requer conhecimento do modelo ou topologia do sistema, implementa a redução de dimensionalidade durante o estágio de treinamento para extração das características principais dos dados da PMU; realiza a detecção de eventos utilizando um conjunto reduzido de PMUs como pilotos para ter melhor performance computacional em

aplicações de tempo real e para realizar a detecção de eventos *online*, não requer grandes *buffers* de dados. Neste estudo a redução de dimensionalidade é realizada levando em consideração apenas dados de frequência e tensão.

Liu et al. (2015) apresentam a implementação de um método de detecção de ilhamento baseado em PCA, cuja proposta envolve duas etapas: 1) modelagem de PCA *offline* usando dados históricos para obter os componentes principais e limites de controle e 2) monitoramento *online* para determinar se um evento de ilhamento ocorre. O trabalho focou na interpretação geométrica de T2 e Q, e mostrou que usando medições de frequência T2 detecta eventos de incompatibilidade de geração e Q detecta eventos de ilhamento.

De acordo com Ge et al. (2017) o método PCA têm múltiplas aplicações, os autores destacam a aplicação do PCA para monitoramento de processos, devido à adequação da técnica como uma ferramenta de modelagem básica. Outro destaque de aplicação do PCA é a redução de dimensionalidade das variáveis, possível quando há uma grande correlação entre as mesmas, podendo gerar uma compactação das informações e consequentemente facilitando outras análises realizadas a partir destes dados, inclusive de forma visual.

Cai et al. (2018) propuseram a integração do método kNN (vizinho mais próximo) com o método de monitoramento estatístico baseado em PCA para que um grande número de variáveis possa ser analisado em tempo real para monitoramento de área ampla de sistemas de potência e, ao mesmo tempo, o efeito de mascaramento de as tendências oscilatórias e o ruído nas medições elétricas dos distúrbios podem ser reduzidos. Mais especificamente, kNN é aplicado em T2 e Q para obter duas novas estatísticas de monitoramento para detectar distúrbios. Porém, os casos de uso apresentados não utilizaram dados de PMUs.

Dimitrovska et al. (2019) apresentam uma proposta de um método aplicável em tempo real para triagem de contingências, através do reconhecimento de condições recorrentes e perigosas, da operação do sistema de potência. O método proposto utiliza o PCA para redução da dimensionalidade e reconhecimento de padrões em uma base de dados extensa, com amostras anteriores à falta. Posteriormente, a partir desta base de exemplos reduzida através do PCA é feita a procura de similaridade do evento corrente, para identificar uma situação fora do normal.

Chatterjee e Chaudhuri (2019) apresentam uma proposta de adaptação do método PCA para tratar dados de baixa qualidade obtidos em PMUs, que podem estar induzindo à interpretação de que se trata de um evento altamente não linear, prejudicando a confiabilidade de detecção de eventos. A mudança no quadrado da norma de pontuação do componente principal ao longo da primeira direção principal no espaço de recursos entre duas janelas sucessivas é insensível a dados ruins e pode ser usada como uma métrica para detectar *outliers* induzidos por evento.

Chu et al. (2017) apresentam um método orientado a dados em tempo real para indicar a avaliação do estado de dados de PMU de um fluxo massivo. Na proposta, os dados de PMU são inicialmente representados com uma sequência de grandes matrizes aleatórias. A técnica realiza o cálculo de componente principal nas várias matrizes modeladas.

Zhang et al. (2017) propõem um modelo de análise de dados de conjunto para avaliação de estabilidade pré-contingência baseada em PMU, considerando medições de dados incompletos. A motivação da proposta é trazer mais robustez para esta avaliação, em função da possibilidade de casos de baixa qualidade dos dados fornecidos por uma determinada PMU.

Le e Benjapolakul (2018) demonstram preocupação com a dependência que o uso de dados de PMU tem da rede de comunicação. Os autores apresentam uma proposta de método de recuperação de dados para preencher a medição de frequência sincronizada.

De acordo com Ahmad e Senroy (2020) a qualidade das medições afeta muito o monitoramento de segurança do sistema de energia e outros aplicativos de análise de dados de missão crítica que dependem de dados de PMU. Por isso, os autores realizaram um estudo do erro introduzido pelo canal de instrumentação a montante da PMU, caracterizando-o como não gaussiano. No artigo é proposta uma nova técnica de agrupamento baseada no modelo de mistura gaussiana, para caracterizar e relacionar os erros presentes nos dados de medição da PMU com a saturação no transformador de corrente, comprimento do cabo e carga da PMU. Ao final é proposto um emulador de erro para que possa ser usado em testes e algoritmos de análise de dados de PMUs.

### 3.3 CONSIDERAÇÕES FINAIS

Sistemas de medição de fasores sincronizados em conjunto com a análise de seus respectivos dados está se tornando tema frequente para diferentes estudos em todo o mundo.

Várias abordagens diferentes podem ser encontradas na literatura e é possível encontrar informações em diferentes fontes. Portanto, o desenvolvimento de um processo estruturado de levantamento do material bibliográfico é fundamental para garantir a análise do material mais relevante.

Com o processo de revisão da literatura realizado neste trabalho, foi possível obter acesso aos artigos com maior relevância, porém devido à dinâmica do assunto certamente novos artigos continuaram a ser produzidos.

Da leitura dos artigos selecionados é possível afirmar que a redução de dimensionalidade é bastante utilizada para a compressão de dados com o objetivo de otimizar o desempenho de análises destes dados.

Para a comparação dos trabalhos selecionados no processo de revisão de literatura, o QUADRO 2 foi definido a partir da definição de alguns critérios, sendo que para cada artigo foi avaliado se o estudo desenvolvido apresenta (S - Sim) ou não (N - Não) o elemento ou abordagem que estão sendo procurados. Quando não foi possível definir se o elemento é abordado no artigo foi utilizado a nomenclatura S.I. (Sem Informação) no QUADRO 2.

O primeiro critério utilizado verifica se foi utilizado um sistema real de grande porte com o uso de PMUs. O segundo critério verifica se foram utilizadas múltiplas PMUs simultaneamente. O terceiro critério verifica se o trabalho foi realizado com a utilização de múltiplas grandezas simultaneamente. O quarto critério identifica se é abordada a questão de redução de dimensionalidade. O quinto critério identifica se foi utilizada a técnica de análise de componentes principais. O sexto critério tem por objetivo verificar se o trabalho efetua alguma análise para identificar qual pode ser o conjunto de dados mais adequado para executar o algoritmo de análise de componentes principais e obter coeficientes para montagem das equações que sejam mais representativos do conjunto de variáveis, sem precisar utilizar necessariamente todo o conjunto de amostras disponível. Esta abordagem visa verificar se há possibilidades de otimizar o processamento de uma forma mais eficaz, utilizando um conjunto de amostras que seja significativo na representação das variáveis originais.

QUADRO 2 - COMPARAÇÃO DOS ELEMENTOS OBSERVADOS NOS ARTIGOS

	1	2	3	4	5	6	7	8	9	10	Método Proposto
Sistema Real de Grande Porte com PMUs	S	S	N	N	S	N	S	N	N	N	S
Múltiplas PMUs simultâneas	S	S	N	N	N	S	S	S	N	N	S
Múltiplas grandezas	N	N	N	S	N	S.I.	N	S.I.	N	N	S
Redução de Dimensionalidade	S	S	S	S	S	S	S	N	N	N	S
PCA	S	S	S	S	S	S	N	N	N	N	S
Análise do conjunto de dados para o PCA	N	N	N	N	N	N	N	N	N	N	S

FONTE: O autor (2022)

Contemplando o QUADRO 2 observa-se que os trabalhos que mais se aproximam são os de número 1 e 2, porém ambos não têm abordagem para tratamento de múltiplas grandezas e nem uma abordagem para analisar qual o melhor conjunto de dados a ser utilizado de forma mais eficaz para alimentação do PCA.

No primeiro artigo, XIE et al. (2014) apresentam um trabalho consistente, porém utilizam um conjunto de dados reduzido, sem explorar as possibilidades de trabalhar com múltiplas grandezas simultaneamente. Neste trabalho são apresentados dois estudos de casos em que são utilizados apenas os dados de tensão e frequência, e em ambos os casos a análise de componentes principais é tratada em separado. No primeiro caso é feita uma redução de sete variáveis de frequência para apenas duas e de sete variáveis de tensão para três. No segundo caso é feita uma redução de 14 variáveis de frequência para apenas duas e de oito variáveis de tensão para três. Além disso, o trabalho não avalia as possibilidades de diferenças dos resultados obtidos aplicando-se subconjuntos de dados diferentes na execução do PCA, para obtenção dos coeficientes que compõem as equações do algoritmo.

No segundo artigo, Liu et al. (2015) apresentam um trabalho focado na detecção de condições de ilhamento utilizando análise de componentes principais. Porém para este trabalho, apresenta um estudo de caso baseado em seis PMUs, utilizando apenas a frequência. Portanto, o uso do algoritmo de análise de componentes principais é feito com apenas seis variáveis de frequência que são reduzidas para apenas uma, que será usada na detecção de condições de ilhamento. Assim, da mesma forma que em XIE et al. (2014), não são avaliadas outras

possibilidades de composição do conjunto de dados para execução do algoritmo do PCA.

A partir desta comparação é possível visualizar que nenhum dos artigos desenvolveu um estudo como o que é apresentado neste trabalho, que consiste em trabalhar com sistema real de grande porte, utilizando múltiplas PMUs e diferentes grandezas simultaneamente, para reduzir a dimensionalidade utilizando a análise de componentes principais e principalmente, realizar a análise do conjunto de dados mais adequado para execução do algoritmo de análise de componentes principais.

## 4 MATERIAL E MÉTODOS

### 4.1 MATERIAIS

#### 4.1.1 Recursos Computacionais

Para o desenvolvimento deste trabalho foi utilizado o *software* MATLAB versão 9.4.0.813654 (R2018a), da MathWorks, Inc. Além das bibliotecas padrão da ferramenta, foi utilizado também a *toolbox* “Statistics and Machine Learning”, versão 11.3, que disponibiliza funções baseadas em análise de componentes principais.

Para obtenção dos dados das PMUs foi utilizado o *software* PhasorPoint, fornecido pela Psymetrix, licenciado para a COPEL Geração e Transmissão S.A.

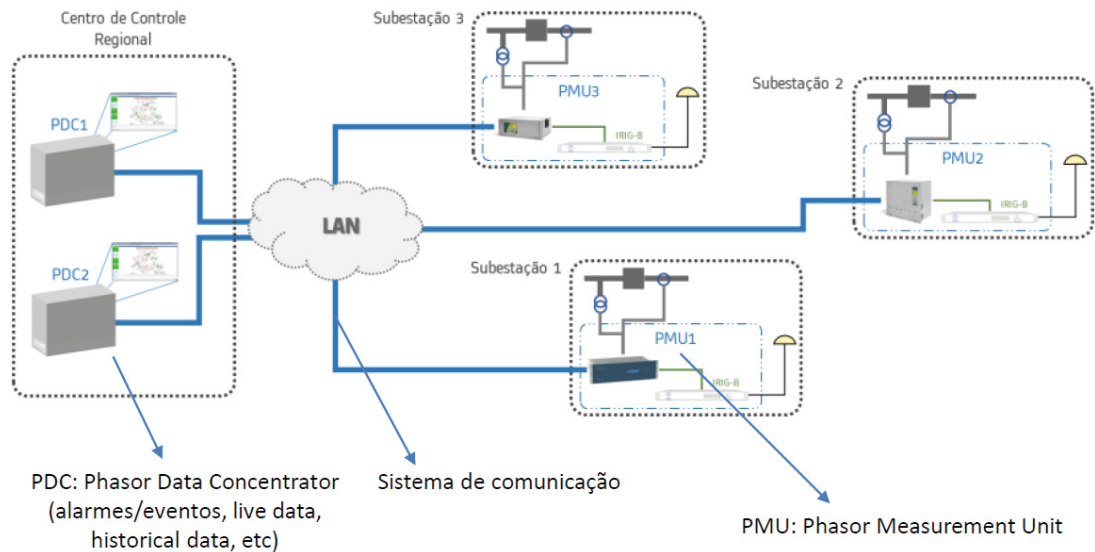
#### 4.1.2 Dados

Os dados utilizados foram obtidos do sistema de medição fasorial sincronizada, da empresa COPEL Geração e Transmissão S.A. A seleção das PMUs que estão sendo analisadas levou em consideração algumas das características apresentadas na seção 2.1.3, em especial a categoria de deficiência de dados relacionada com a perda de dados. Assim, as PMUs que apresentam lacunas na coleta de qualquer grandeza são descartadas, sendo privilegiado o uso de PMUs que apresentam o conjunto completo de dados.

##### 4.1.2.1 Arquitetura do SMFS

A arquitetura do SMFS utilizado para o fornecimento de dados pode ser visualizada na FIGURA 6.

FIGURA 6 - ARQUITETURA DO SMFS



FONTE: Adaptado de COPEL(2021)

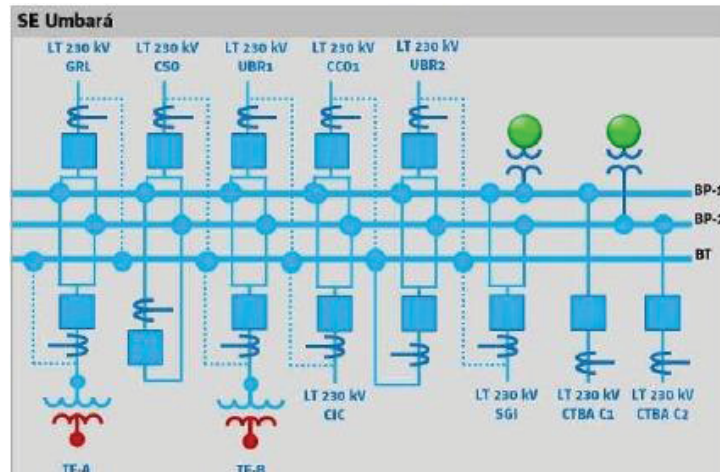
A FIGURA 6 mostra que a solução é composta por diversas PMUs, instaladas em cada uma das subestações da COPEL Geração e Transmissão. Estas PMUs se comunicam pela rede com dois servidores, um principal e um de *backup*, os quais concentram os dados de todas as PMUs. E, a partir deste servidor, é possível acessar os dados de todas as PMUs. Atualmente o sistema de Medição Fasorial Sincronizada da COPEL Geração e Transmissão é composto de 56 PMUs.

#### 4.1.2.2 Estrutura dos dados das PMUs

Cada PMU gera o seu respectivo conjunto de dados, contendo as grandezas relacionadas às medições do local onde está instalada.



FIGURA 7 - EXEMPLO DE UM DIAGRAMA DE UMA PMU



FONTE: PhasorPoint (2021)

A FIGURA 7 mostra o exemplo do diagrama esquemático de uma PMU. As medições que são coletadas são relativas às grandezas de tensão, frequência e taxa de variação da frequência ( $df/dt$ ) do barramento e das correntes em todos os pontos de conexão com outras linhas de transmissão. Além disso, para a tensão e corrente é feita a medição da magnitude e do ângulo para cada uma das três fases.

Assim, a estrutura de dados fornecida por cada PMU segue um padrão conforme a relação de atributos descrita no QUADRO 3.

QUADRO 3 - RELAÇÃO DE ATRIBUTOS DE UMA PMU

Item	Atributo	Unidade
1	Estampa de Tempo	Data e Hora
2	Status	Texto
3	Frequência	Hertz
4	$df/dt$ – taxa de variação da frequência	
5	Tensão Barra magnitude Fase A	Volt
6	Tensão Barra ângulo Fase A	Graus
7	Tensão Barra magnitude Fase B	Volt
8	Tensão Barra ângulo Fase B	Graus
9	Tensão Barra magnitude Fase C	Volt
10	Tensão Barra ângulo Fase C	Graus
11	Corrente Linha magnitude Fase A	Ampere
12	Corrente Linha ângulo Fase A	Graus
13	Corrente Linha magnitude Fase B	Ampere
14	Corrente Linha ângulo Fase B	Graus
15	Corrente Linha magnitude Fase C	Ampere
16	Corrente Linha ângulo Fase C	Graus

FONTE: O autor (2022)

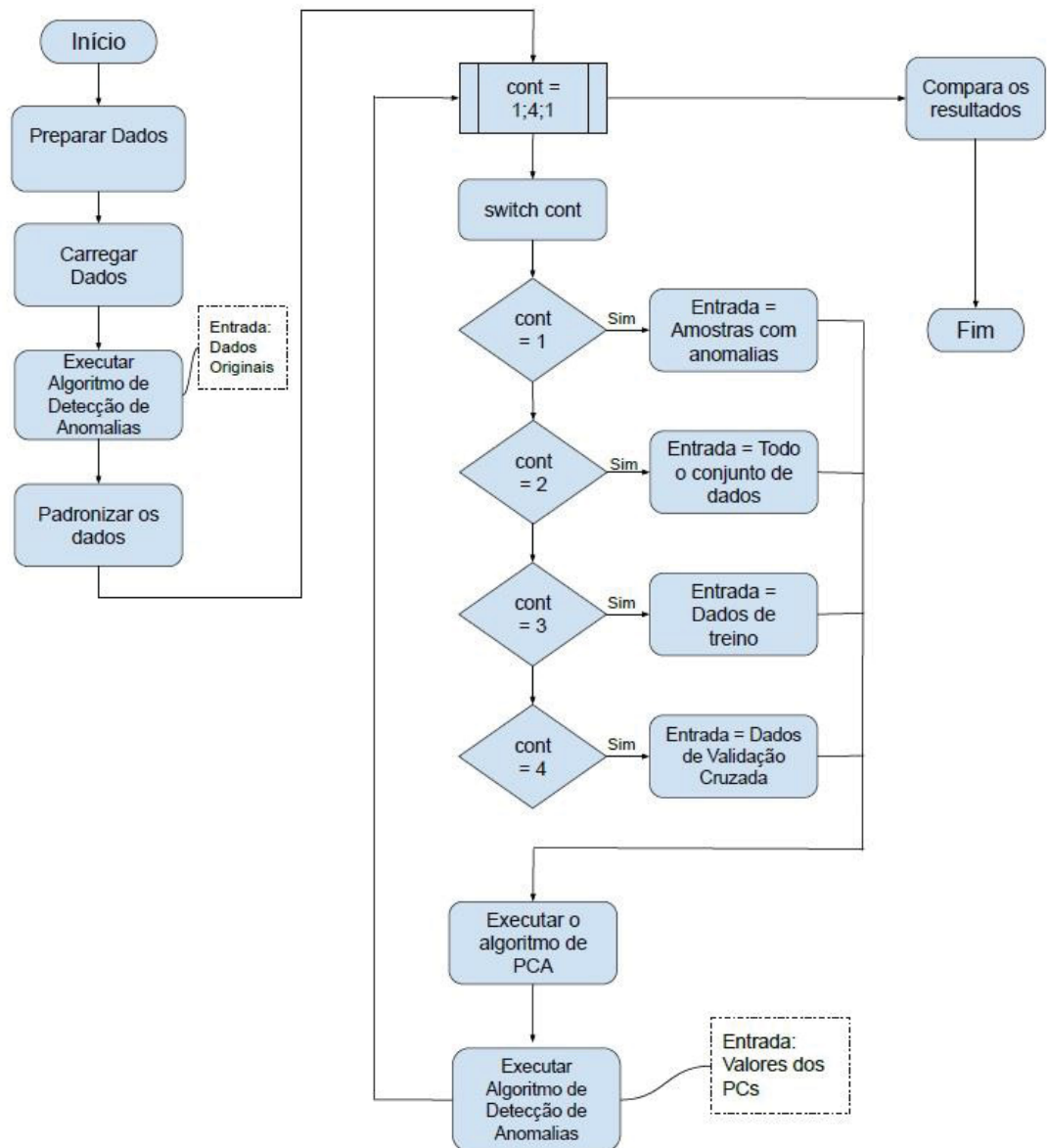
O QUADRO 3 contém o padrão da lista de atributos que fazem parte do conjunto de dados de uma PMU. Os atributos relacionados com a corrente (11 a 16) se repetem para cada uma das linhas que estiverem conectadas no respectivo barramento. Por exemplo, o diagrama da FIGURA 7 mostra nove linhas conectadas no barramento, o que implica que o conjunto de medições de corrente se repetirá nove vezes.

## 4.2 MÉTODO

### 4.2.1 Fluxograma

A FIGURA 8 mostra a representação do fluxo de trabalho executado para a metodologia de análise dos dados das PMUs.

FIGURA 8 - FLUXOGRAMA DA METODOLOGIA



FONTE: O autor (2022)

A sequência de atividades representada na FIGURA 8 inicia com a preparação dos dados das PMUs e sua carga no programa computacional. Na sequência é executado o método de detecção de anomalias com este conjunto de dados original, para que sirva de referência comparativa. Para executar o algoritmo de análise de componentes principais, antes é necessário executar um procedimento de padronização dos dados.

Em seguida inicia-se um conjunto de ações que é executado quatro vezes dentro de uma estrutura de repetição. Para cada iteração há uma variação na seleção

dos dados de entrada que alimentam a execução do procedimento de seleção de componentes principais. A seleção de dados para cada uma das quatro iterações é:

- somente as amostras com anomalias;
- todo o conjunto de dados;
- somente o conjunto de dados de treino; e
- somente o conjunto de dados de validação cruzada e teste.

Então, a partir do conjunto de dados selecionado, executa-se o algoritmo para identificação dos componentes principais. Na sequência, seleciona-se os componentes principais de acordo com a sua contribuição para explicar a variância. A partir desta seleção, efetua-se o cálculo dos componentes principais selecionados. Com a relação completa de componentes principais selecionados já calculados, efetua-se então uma nova execução do método de detecção de anomalias, agora alimentado por este conjunto de componentes principais.

Após encerrar-se este conjunto de análises, efetua-se uma comparação dos resultados obtidos em cada uma das iterações, encerrando assim o procedimento.

#### 4.2.2 Preparação e carga dos dados

O procedimento de preparação e carga dos dados consiste basicamente na leitura dos arquivos com as leituras de dados das PMUs e sua respectiva preparação. O procedimento permite a leitura de mais de um arquivo, permitindo assim fazer a composição através da concatenação dos dados de várias PMUs.

No caso de concatenação de vários arquivos, a premissa é de que todos referem-se ao mesmo intervalo de tempo. Assim a quantidade de registros em todos eles é a mesma. Desta forma, seleciona-se uma única vez a lista com a estampa de tempo, que deve ser a mesma em todos os arquivos, e que equivale ao item 1 do QUADRO 3. Para cada arquivo seleciona-se os atributos equivalentes aos itens de 3 em diante deste mesmo quadro. Estes atributos são concatenados lado a lado, o que significa que, analisando sob o aspecto de uma tabela esta concatenação de dados aumenta a quantidade de colunas e mantém a quantidade de linhas original.

Após concluir a montagem do conjunto de dados, o procedimento separa o conjunto em três partes:

- dados de treino;

- dados de teste;
- dados de validação cruzada.

A separação dos dados é feita com base em variáveis pré-configuradas que determinam a proporção, e conseqüentemente a quantidade de registros que compõem cada um dos três conjuntos. Por exemplo, pode-se usar uma configuração que determina que 70% dos dados compõem o conjunto de treino, 15% dos dados compõem o conjunto de teste e 15% dos dados formam o conjunto de validação cruzada.

#### 4.2.3 Algoritmo de detecção de anomalias

Neste processo utiliza-se um modelo gaussiano para detectar exemplos anômalos no conjunto de dados. Nesse conjunto de dados, encontra-se uma distribuição gaussiana e, em seguida, encontra-se valores que tenham probabilidade muito baixa e, portanto, podem ser considerados anomalias.

Para realizar a detecção de anomalias, primeiro é necessário ajustar um modelo para a distribuição dos dados.

Dado um conjunto de treinamento  $\{x^{(1)}; \dots; x^{(m)}\}$  (onde  $x^{(i)} \in \mathbb{R}^n$ ), estima-se a distribuição gaussiana para cada uma das características  $x_i$ . Para cada característica  $i = 1 \dots n$ , encontra-se os parâmetros  $\mu_i$  e  $\sigma_i^2$  que se ajustam aos dados na  $i$ -ésima dimensão  $\{x^{(1)}; \dots; x^{(m)}\}$  (a  $i$ -ésima dimensão de cada exemplo).

A distribuição gaussiana é dada por:

$$p(x; \mu; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (17)$$

onde  $\mu$  é a média e  $\sigma^2$  controla a variância.

São estimados os parâmetros,  $(\mu_i, \sigma_i^2)$ , do  $i$ -ésimo atributo usando as seguintes equações.

- média:

$$\mu_i = \frac{1}{m} \sum_{j=1}^m x_i^{(j)} \quad (18)$$

- variância:

$$\sigma_i^2 = \frac{1}{m} \sum_{j=1}^m (x_i^{(j)} - \mu_i)^2. \quad (19)$$

Após estimarem-se os parâmetros gaussianos, pode-se investigar quais amostras têm uma probabilidade muito alta dada essa distribuição e quais amostras têm uma probabilidade muito baixa. As amostras de baixa probabilidade são mais prováveis de serem as anomalias no conjunto de dados. Neste caso, a maneira de determinar quais exemplos são anomalias é selecionar um limite com base em um conjunto de validação cruzada. Para isto, utiliza-se um algoritmo para selecionar o limite  $\epsilon$  usando a pontuação F1 Score no conjunto de validação cruzada, conforme será apresentado na seção 4.2.7.

#### 4.2.4 Padronização dos dados

Nesta etapa é feita uma padronização dos dados, ou seja, é feito um escalonamento de cada dimensão para que fiquem todas no mesmo intervalo. Isto é feito para que as variáveis com grande escala não dominem o processo sobre as demais.

No caso, os atributos como frequência, magnitude e ângulo de tensão e corrente, são representados por grandezas diferentes em suas respectivas escalas. Na análise comparativa dos atributos, é necessário que todos os atributos sejam transformados para que possam ser observados dentro do mesmo intervalo de escala. Isto é feito através da subtração da média e divisão pelo desvio padrão em cada atributo do conjunto de dados. A padronização é feita usando a fórmula *z-score*, para que se tenha média zero e desvio padrão unitário:

$$z = \frac{x - \mu}{\sigma} \quad (20)$$

Onde  $\mu$  é a média e  $\sigma$  é o desvio padrão.

#### 4.2.5 Algoritmo de análise dos componentes principais

Este processo executa a análise de componentes principais dos dados recebidos, efetua a seleção dos componentes principais com base na quantidade mínima e calcula e monta a lista de valores dos componentes principais.

Para execução da análise de componentes principais, propriamente dita, utiliza-se o algoritmo que retorna duas informações essenciais: a matriz de coeficientes e o vetor de explicação por componente principal.

Cada coluna da matriz de coeficientes contém os coeficientes para um componente principal e as colunas estão em ordem decrescente de variância do componente. Por padrão, a função `pca` centraliza os dados e usa o algoritmo de decomposição de valor singular (SVD - Singular Value Decomposition).

O vetor de explicação por componente principal contém a porcentagem da variância total explicada para cada um dos componentes principais identificados.

Uma vez que o objetivo neste processo é diminuir a dimensionalidade do conjunto de dados, é necessário definir quantos componentes principais serão utilizados. Para isto utiliza-se um critério de importância cumulativa, ou seja, seleciona-se os  $n$  primeiros componentes cuja soma cumulativa da porcentagem de variância excedam o limite de 80%. Este valor é obtido a partir do vetor de explicação por componente principal.

Tendo obtido a quantidade de componentes principais selecionados, é este número que define a nova dimensão do conjunto de dados que se tem como resultado, ou seja, o equivalente à quantidade de colunas da tabela.

Para definir os novos valores de cada amostra do conjunto de dados, identifica-se a matriz de autovetores, que nada mais é que a matriz de coeficientes retornada pelo algoritmo de análise de componentes principais.

Para cada amostra, em cada um dos componentes principais selecionados, faz-se a soma cumulativa do autovetor, multiplicado pelo valor original.

Ou seja, tem-se a seguinte equação:

$$PC_{z,j} = \sum_{i=1}^{\text{atributos}} \text{autovetores}_{ji} \cdot \text{dados}_{zi} \quad (21)$$

O que também equivale a dizer que, para cada amostra, calcula-se a nova relação de atributos, obtidos a partir do cálculo de componentes principais, aplicando as seguintes equações lineares:

$$PC_{z,1} = a_{11} \cdot X_1 + a_{12} \cdot X_2 + \dots + a_{1i} \cdot X_i \quad (22)$$

$$PC_{z,2} = a_{21} \cdot X_1 + a_{22} \cdot X_2 + \dots + a_{2i} \cdot X_i \quad (23)$$

⋮

$$PC_{z,j} = a_{j1} \cdot X_1 + a_{j2} \cdot X_2 + \dots + a_{ji} \cdot X_i \quad (24)$$

onde:

- $i$  = quantidade de atributos do conjunto de dados originais;
- $j$  = quantidade de componentes principais selecionados;
- $z$  = quantidade de amostras do conjunto de dados originais;
- $a$  = autovetores;
- $X$  = dados originais,

Com isto, este procedimento monta um novo conjunto de dados, mantendo a mesma quantidade de amostras do conjunto original, porém reduzindo a quantidade de atributos, para tantos quanto foram definidos na seleção dos componentes principais. Os valores do novo conjunto de dados são definidos através da aplicação das equações lineares de montagem de cada um dos componentes principais, que na realidade equivalem aos novos atributos deste novo conjunto de dados.

#### 4.2.6 Detecção de anomalias com os componentes principais

Na sequência da montagem do novo conjunto de dados, com dimensões reduzidas, aplica-se novamente o mesmo algoritmo de detecção de anomalias utilizando um modelo gaussiano para detectar exemplos anômalos no conjunto de dados.

O conjunto de dados de treino, teste e validação cruzada é remontado com os novos valores. Mas continua utilizando-se as mesmas variáveis pré-configuradas que determinam a proporção, e conseqüentemente a quantidade, de registros que compõem cada um dos três conjuntos.



Os critérios e a metodologia de detecção de anomalias continuam sendo exatamente os mesmos utilizados na etapa anterior, para que possa ser feita uma comparação fidedigna entre os resultados obtidos.

#### 4.2.7 Métricas para comparação dos resultados

Na última etapa da metodologia, executa-se uma comparação dos resultados obtidos na detecção de anomalias a partir dos dados originais, com os resultados obtidos a partir dos dados com dimensionalidade reduzida.

A técnica adotada para fazer esta análise é a matriz de confusão, uma métrica que nada mais é do que uma tabela que mostra o número de previsões corretas e incorretas, sendo que neste caso, considera-se como uma previsão a detecção de anomalia baseada no conjunto de dados com dimensionalidade reduzida. A tabela tem uma estrutura como a demonstrada na FIGURA 9.

FIGURA 9 - MATRIZ DE CONFUSÃO

		Resposta Prevista	
		yPrev = 1	yPrev = 0
Resposta Real	y = 1	Verdadeiro POSITIVO	Falso NEGATIVO
	y = 0	Falso POSITIVO	Verdadeiro NEGATIVO

FONTE: Adaptado de BRUCE e BRUCE (2019)

A imagem representada na FIGURA 9, tem por objetivo mostrar que em cada quadrante da tabela são calculadas as quantidades de amostras que atendem aos respectivos critérios (BRUCE; BRUCE, 2019), que são:

- verdadeiro POSITIVO: são computadas todas as amostras do conjunto de anomalias de dimensionalidade reduzida que foram identificadas como anomalias e que também foram identificadas como anomalias no conjunto de dados original;
- falso POSITIVO: são computadas todas as amostras do conjunto de anomalias de dimensionalidade reduzida que foram identificadas como

anomalias, porém não foram identificadas como anomalias no conjunto de dados original;

- falso NEGATIVO: são computadas todas as amostras do conjunto de anomalias de dimensionalidade reduzida que não foram identificadas como anomalias, porém foram identificadas como anomalias no conjunto de dados original; e
- verdadeiro NEGATIVO: são computadas todas as amostras do conjunto de anomalias de dimensionalidade reduzida que não foram identificadas como anomalias e que também não foram identificadas como anomalias no conjunto de dados original.

A partir dos números obtidos para preenchimento desta tabela, efetua-se o cálculo das seguintes métricas: acurácia, precisão, revocação e *F1 Score*.

A equação da acurácia é (BRUCE; BRUCE, 2019):

$$acurácia = \frac{\sum VerdadeiroPositivo + \sum VerdadeitoNegativo}{\sum amostras}. \quad (25)$$

A equação da precisão é (BRUCE; BRUCE, 2019):

$$precisão = \frac{\sum VerdadeiroPositivo}{\sum VerdadeiroPositivo + \sum FalsoPositivo}. \quad (26)$$

A equação da revocação é (BRUCE; BRUCE, 2019):

$$revocação = \frac{\sum VerdadeiroPositivo}{\sum VerdadeiroPositivo + \sum FalsoNegativo}. \quad (27)$$

E a equação do *F1 Score* é (BRUCE; BRUCE, 2019):

$$F1\ score = 2 \cdot \frac{precisao \cdot revocação}{precisao + revocação}. \quad (28)$$

O *F1 Score* é conhecido como a média harmônica entre a precisão e a revocação e portanto, permite obter um resultado combinando estas duas métricas.

O processo permite também a comparação visual dos resultados de anomalias obtidos, utilizando um gráfico que plota os pontos identificados como anomalias nos dados originais e os pontos identificados como anomalias no conjunto de dimensionalidade reduzida.

## 5 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Adotou-se duas abordagens distintas para realizar a análise da aplicação do PCA. Na primeira abordagem, relacionada à quantidade de PMUs utilizada, avalia-se a eficiência da solução aplicada individualmente por PMU ou aplicada em conjuntos de PMUs agrupadas. O objetivo é verificar e comparar os resultados obtidos com o olhar individual sobre cada PMU ou o olhar mais sistêmico, juntando dados do maior número possível de PMUs. Na segunda abordagem, relacionada à forma de alimentar o PCA, verifica-se qual conjunto de dados pode ser mais eficaz para executar o algoritmo de PCA. O objetivo é verificar os resultados obtidos utilizando-se todo o conjunto de dados e comparar com os resultados obtidos usando subconjuntos de dados para execução do PCA.

### 5.1 DESCRIÇÃO DOS CENÁRIOS

#### 5.1.1 Descrição do Caso

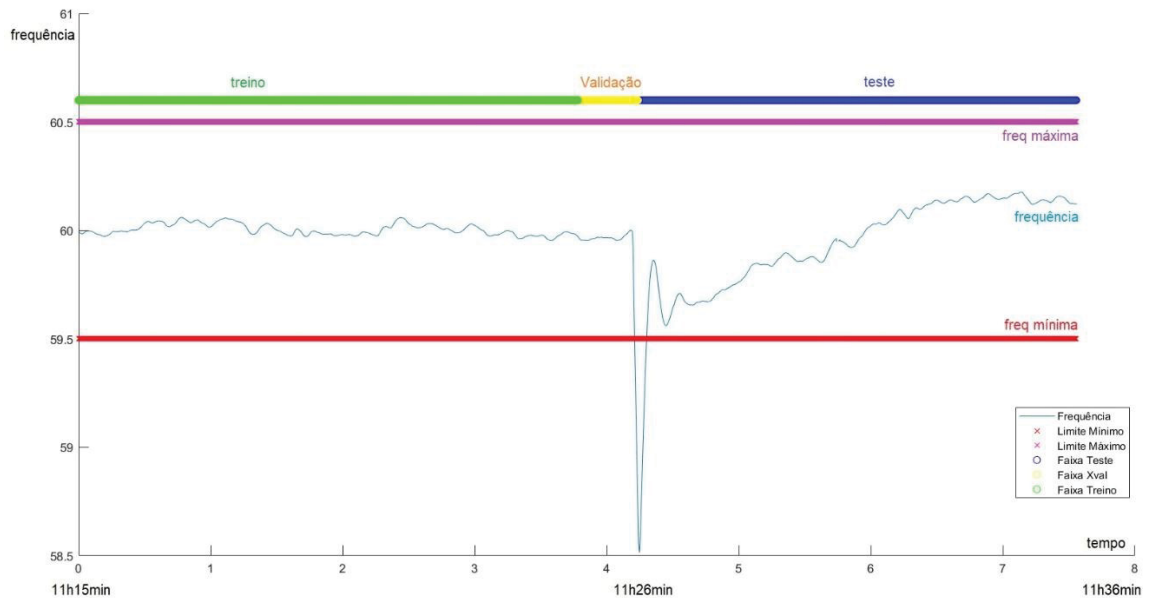
Para o desenvolvimento deste trabalho foi utilizado um caso real. Foram identificados dois casos reais conhecidos do sistema interligado nacional ocorridos no último ano, tendo sido o critério de seleção, o caso em que verificou-se haver a melhor qualidade dos dados disponíveis. Neste caso, o conceito utilizado para a determinação da qualidade dos dados foi o da categoria de perda de dados, conforme explicado na seção 2.1.3. Para realização desta análise foram inspecionados todos os 41 arquivos disponibilizados com as leituras das PMUs dos dois casos, tendo tido como resultado desta inspeção a constatação de que os arquivos do caso selecionado são os que apresentavam a maior quantidade de atributos preenchidos.

Selecionou-se um evento ocorrido no dia 28/05/2021, às 11h26, em que ocorreu o desligamento automático do polo 2 do Elo CC 800 kV Xingu / Estreito (BMTE) e de sete unidades geradoras da UHE Belo Monte (Norte Energia S.A.), as quais estavam gerando no momento um montante de aproximadamente 4.050 MW.

Para o estudo do caso selecionou-se o intervalo de tempo compreendido entre 11h15 e 11h36. Considerando que os dados coletados pelas PMUs são fornecidos em uma taxa de 60 leituras por segundo, isso significa dizer que este conjunto de dados é composto por 75.600 (setenta e cinco mil e seiscentos) registros de leitura.

Neste intervalo de tempo o comportamento da frequência lida pelas PMUs que foram selecionadas teve a variação conforme demonstrado na FIGURA 10.

FIGURA 10 - GRÁFICO DE FREQUÊNCIA



FONTE: O autor (2022)

A FIGURA 10 mostra um gráfico de variação da frequência ao longo do tempo. O eixo das ordenadas representa a variação do valor da frequência em Hz (Hertz) e o eixo das abscissas representa a variação das leituras ao longo do tempo, iniciando em 11h15 e indo até 11h36.

No gráfico, seguindo uma sequência de observação das linhas de baixo para cima, tem-se a seguinte interpretação:

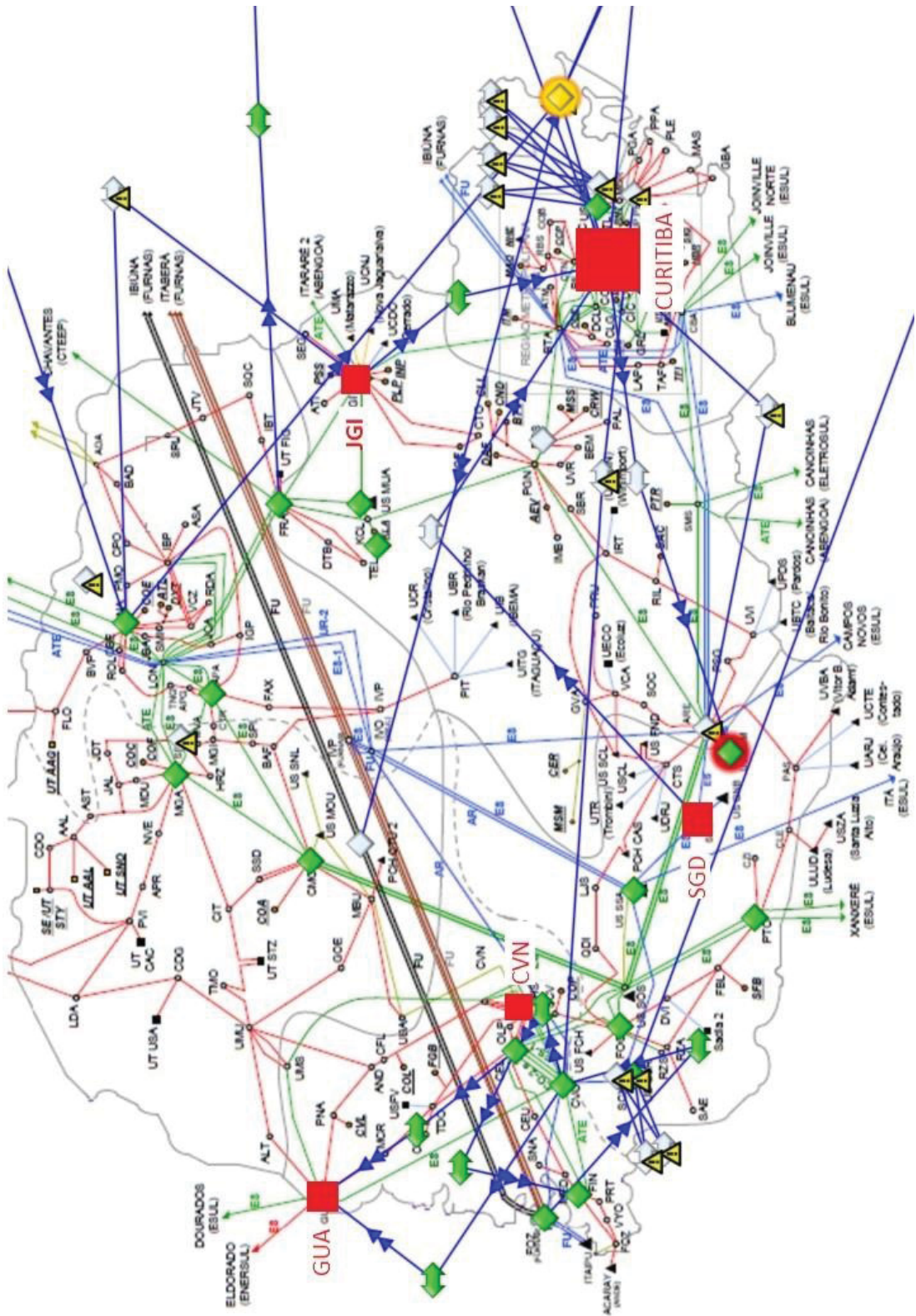
- Linhas verde, amarela e azul: representa a identificação da delimitação dos conjuntos de dados para treinamento, teste e validação cruzada. Sendo que a faixa em verde representa os dados para treinamento, a faixa em amarela representa os dados para validação cruzada e a faixa em azul representa os dados para teste;
- Linha magenta: representa o limite superior aceitável para a oscilação da frequência, 60,5 MHz, conforme submódulo 2.9 (Requisitos mínimos de qualidade de energia elétrica para acesso ou integração à Rede Básica do ONS);

- Linha vermelha grossa: representa o limite inferior aceitável para a oscilação da frequência, que é 59,5 MHz, conforme submódulo 2.9 (Requisitos mínimos de qualidade de energia elétrica para acesso ou integração à Rede Básica do ONS); e
- Linha em azul claro: representa a grandeza da frequência. Pode-se observar que a frequência apresenta uma oscilação que ultrapassa a fronteira do limite mínimo estabelecido pelo ONS.

Assim, pode-se observar que no instante citado anteriormente, 11h26, é possível identificar uma variação na frequência que extrapola o limite mínimo regulamentado pelo ONS.

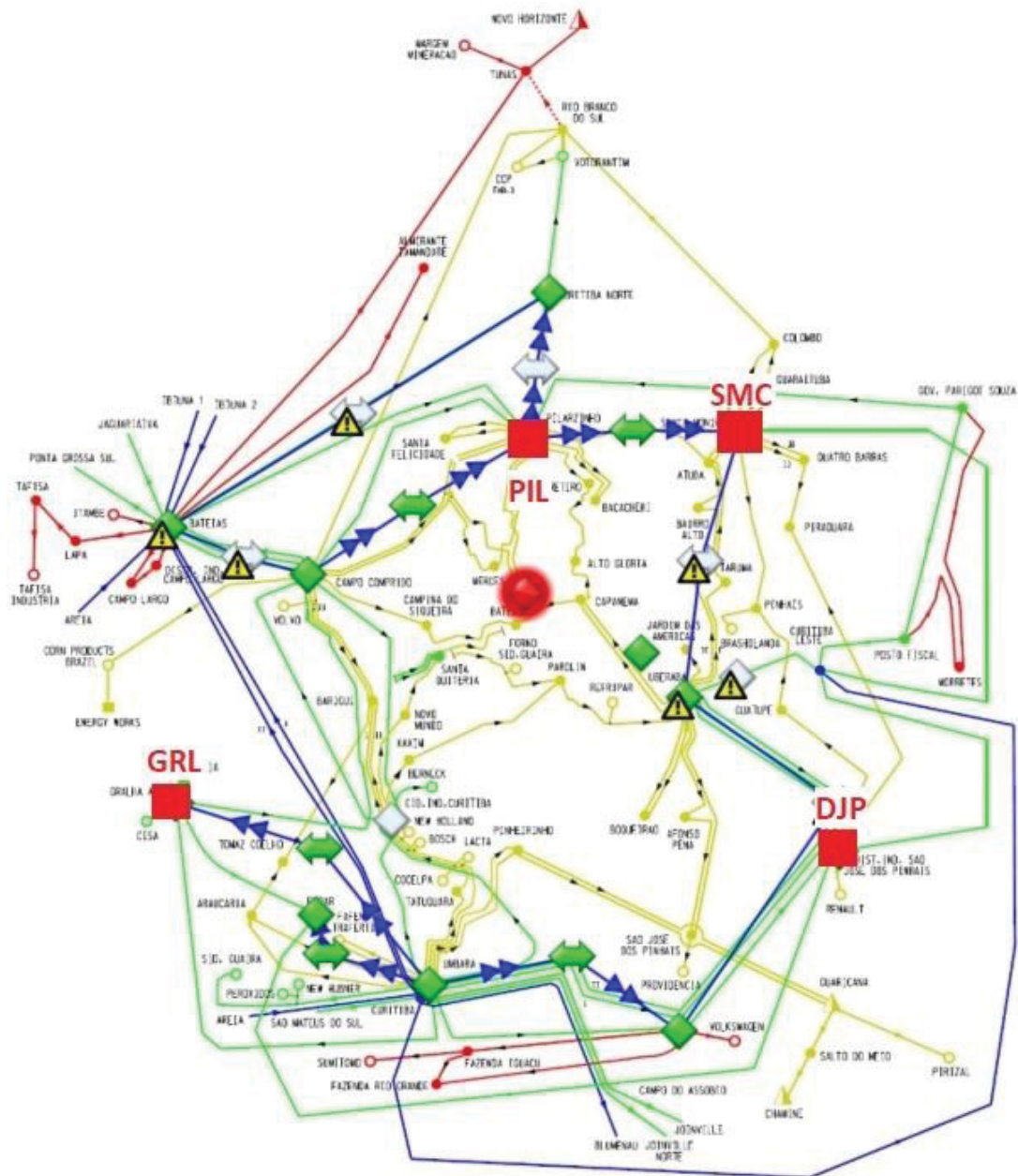
Para análise do caso, entre os arquivos das 41 PMUs, foram selecionadas oito PMUs localizadas em subestações da COPEL Geração e Transmissão representando 20% do conjunto disponibilizado, tendo como critério de seleção a qualidade dos dados. Foi feita uma verificação nos arquivos coletados para identificar as PMUs que não apresentassem nenhuma falha de leitura ou coleta no intervalo de tempo previamente selecionado. Estas PMUs estão geograficamente distribuídas conforme a FIGURA 11. Quatro delas estão na região da cidade de Curitiba, conforme a FIGURA 12, e quatro estão no interior do estado do Paraná, mais especificamente nas regiões das cidades de Jaguariaíva, Pinhão, Cascavel e Guaira.

FIGURA 11 - LOCALIZAÇÃO DAS PMUs SELECIONADAS NO PARANÁ



FONTE: PhasorPoint (2021)

FIGURA 12 - LOCALIZAÇÃO DAS PMUs EM CURITIBA



FONTE: PhasorPoint (2021)

### 5.1.2 Abordagem 1 – Quantidade de PMUs

A primeira abordagem relaciona-se com a quantidade de PMUs sendo analisada, podendo ser tanto uma análise individual para cada PMU, quanto uma análise por agrupamentos de PMUs.

No cenário de análise individual, o método é aplicado individualmente em



cada uma das PMUs, as quais apresentam uma caracterização inicial conforme exibida na TABELA 3:

TABELA 3 - LISTA DE PMUs

PMU	Quantidade de atributos	Quantidade de anomalias
PIL	32	215
SMC	20	687
GRL	26	548
DJP	26	3982
JGI	32	290
SGD	26	965
CVN	14	537
GUA	14	2432

FONTE: O autor (2022)

No cenário de análise agrupada, o método é aplicado em grupos de PMUs, montados aleatoriamente, os quais apresentam uma caracterização inicial conforme exibida na TABELA 4.

TABELA 4 - GRUPOS DE PMUs

PMU	Quantidade de atributos	Quantidade de anomalias
PIL+SMC+GRL	78	692
DJP+JGI+SGD	84	1735
PIL+SMC+GRL+DJP+JGI+SGD	162	962
Todas as oito PMUs	190	857

FONTE: O autor (2022)

### 5.1.3 Abordagem 2 – Alimentação do PCA

A segunda abordagem relaciona-se com o conteúdo do conjunto de dados que é utilizado para executar o algoritmo de análise de componentes principais. São experimentadas quatro configurações diferentes.

Na primeira configuração, o algoritmo de análise de componentes principais é alimentado somente com os dados das amostras que foram identificadas como

anomalias, durante a etapa do processo em que é feita a detecção de anomalias no conjunto de dados completo (Dados1).

Na segunda configuração, o algoritmo de análise de componentes principais é alimentado com todo o conjunto completo de amostras (Dados2).

Na terceira configuração, o algoritmo de análise de componentes principais é alimentado somente com as amostras que foram selecionadas para compor o conjunto de dados de treino, para o algoritmo de detecção de anomalias (Dados3).

Na quarta configuração, o algoritmo de análise de componentes principais é alimentado somente com as amostras que foram selecionadas para compor o conjunto de dados de validação cruzada e testes, para o algoritmo de detecção de anomalias (Dados4).

## 5.2 ANÁLISE DOS RESULTADOS

### 5.2.1 Resultados obtidos

A combinação destas duas abordagens resulta em 48 diferentes cenários, que são apresentados da TABELA 5 até a TABELA 16.

Cada tabela mostra o resultado de uma das 12 configurações de agrupamentos de PMUs, conforme descrito na seção 5.1.2. Em cada tabela, as quatro colunas de valores mostram os resultados obtidos com cada uma das quatro configurações de alimentação do PCA, conforme a seção 5.1.3. Esta combinação das abordagens faz com que cada tabela apresente os resultados de quatro cenários diferentes.

Todas as 12 tabelas são montadas com a mesma estrutura. As linhas apresentam as seguintes informações:

- Matriz de confusão: mostra os resultados da matriz de confusão desta análise, conforme a estrutura da FIGURA 9 e as definições na seção 4.2.7:
  - acurácia: resultado da acurácia, conforme equação (25);
  - precisão: resultado da precisão, conforme equação (26);
  - revocação: resultado da revocação, conforme equação (27);
  - *F1 Score*: resultado do *F1 Score*, conforme equação (28).

- PCs Seleccionados: relação dos componentes principais que foram selecionados de acordo com o critério de soma cumulativa da variância e as respectivas contribuições.

Na TABELA 5 até a TABELA 12 são mostrados os resultados obtidos com a abordagem de análise individual das PMUs e na TABELA 13 até a TABELA 16 são mostrados os resultados obtidos com a análise agrupada das PMUs. Além disso, cada uma das tabelas mostra os resultados obtidos com a abordagem de diferentes configurações para a alimentação do PCA. Nestas tabelas o valor destacado em vermelho refere-se ao(s) conjunto(s) de dados que alcançou(çaram) o maior valor de *F1 Score*.

A TABELA 5 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados da PMU PIL, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 3, esta configuração da primeira abordagem originalmente apresenta 32 atributos e 215 anomalias.

TABELA 5 - RESULTADOS DA PMU PIL

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   2656  30564   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  30391   2829   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   1445  31775   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  20653  12567   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	19,14%	92,52%	15,94%	66,75%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	8,00%	91,48%	4,35%	62,17%
<b>F1 Score</b>	<b>0,148</b>	<b>0,956</b>	<b>0,083</b>	<b>0,767</b>
<b>PCs Selecionados</b>	PC 1 = 47,58 % PC 2 = 17,75 % PC 3 = 13,22 % PC 4 = 3,73 %	PC 1 = 38,45 % PC 2 = 17,22 % PC 3 = 12,15 % PC 4 = 9,30 % PC 5 = 3,23 %	PC 1 = 47,05 % PC 2 = 17,68 % PC 3 = 13,34 % PC 4 = 3,78 %	PC 1 = 35,30 % PC 2 = 17,31 % PC 3 = 11,93 % PC 4 = 10,87 % PC 5 = 3,80 % PC 6 = 3,46 %

Fonte: O autor (2022)

Na TABELA 5, observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados2 alimentando o algoritmo de análise de componentes principais (destacado em vermelho). E neste caso, obtém-se o critério de seleção através da acumulação de 80% da contribuição com apenas cinco componentes principais.

A TABELA 6 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados da PMU SMC, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 3, esta configuração da primeira abordagem originalmente apresenta 20 atributos e 687 anomalias.

TABELA 6 - RESULTADOS DA PMU SMC

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   345  32875   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  11565  21655   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   128  33092   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  32018   1202   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	13,03%	42,71%	12,46%	96,82%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	1,04%	34,81%	0,39%	96,38%
<b>F1 Score</b>	<b>0,021</b>	<b>0,516</b>	<b>0,008</b>	<b>0,982</b>
<b>PCs Selecionados</b>	PC 1 = 44,84 % PC 2 = 15,02 % PC 3 = 12,68 % PC 4 = 6,33 % PC 5 = 5,14 %	PC 1 = 31,08 % PC 2 = 16,32 % PC 3 = 14,82 % PC 4 = 13,00 % PC 5 = 5,02 %	PC 1 = 44,68 % PC 2 = 14,90 % PC 3 = 12,90 % PC 4 = 6,23 % PC 5 = 5,15 %	PC 1 = 27,83 % PC 2 = 15,49 % PC 3 = 14,27 % PC 4 = 12,29 % PC 5 = 7,78 % PC 6 = 4,21 %

Fonte: O autor (2022)

Na TABELA 6, observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados4 alimentando o algoritmo de análise de componentes principais. E neste caso, obtém-se o critério de seleção através da acumulação de 80% da contribuição com seis componentes principais.

A TABELA 7 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados da PMU GRL, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 3, esta configuração da primeira abordagem originalmente apresenta 26 atributos e 548 anomalias.

TABELA 7 - RESULTADOS DA PMU GRL

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   779  32441   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   3102  30118   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	14,18%	100,00%	20,32%	100,00%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	2,34%	100,00%	9,34%	100,00%
<b>F1 Score</b>	<b>0,046</b>	<b>1,000</b>	<b>0,171</b>	<b>1,000</b>
<b>PCs Selecionados</b>	PC 1 = 42,91 % PC 2 = 15,98 % PC 3 = 14,69 % PC 4 = 4,63 % PC 5 = 4,02 %	PC 1 = 31,95 % PC 2 = 15,92 % PC 3 = 15,44 % PC 4 = 12,43 % PC 5 = 4,89 %	PC 1 = 42,34 % PC 2 = 16,02 % PC 3 = 14,95 % PC 4 = 4,53 % PC 5 = 4,12 %	PC 1 = 23,77 % PC 2 = 20,43 % PC 3 = 16,39 % PC 4 = 15,78 % PC 5 = 4,48 %

Fonte: O autor (2022)

Na TABELA 7, observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados2 e os Dados4 alimentando o algoritmo de análise de componentes principais. E em ambos os casos, obtém-se o critério de seleção através da acumulação de 80% da contribuição com cinco componentes principais.

A TABELA 8 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados da PMU DJP, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 3, esta configuração da primeira abordagem originalmente apresenta 26 atributos e 3.982 anomalias.

TABELA 8 - RESULTADOS DA PMU DJP

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   0  33220   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  28395   4825   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   22  33198   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  30155   3065   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	12,12%	87,24%	12,17%	91,89%
<b>Precisão</b>	NaN %	100,00%	100,00%	100,00%
<b>Revocação</b>	0,00%	85,48%	0,07%	90,77%
<b>F1 Score</b>	0,00%	<b>0,922</b>	<b>0,001</b>	<b>0,952</b>
<b>PCs Selecionados</b>	PC 1 = 45,32 % PC 2 = 16,37 % PC 3 = 14,34 % PC 4 = 4,55 %	PC 1 = 21,91 % PC 2 = 17,26 % PC 3 = 15,77 % PC 4 = 13,70 % PC 5 = 9,22 % PC 6 = 4,14 %	PC 1 = 44,94 % PC 2 = 16,10 % PC 3 = 14,50 % PC 4 = 4,46 % PC 5 = 4,28 %	PC 1 = 31,83 % PC 2 = 15,32 % PC 3 = 14,79 % PC 4 = 13,19 % PC 5 = 4,22 % PC 6 = 3,86 %

Fonte: O autor (2022)

Na TABELA 8, observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados4 alimentando o algoritmo de análise de componentes principais. E neste caso, obtém-se o critério de seleção através da acumulação de 80% da contribuição com seis componentes principais.

A TABELA 9 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados da PMU JGI, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 3, esta configuração da primeira abordagem originalmente apresenta 32 atributos e 290 anomalias.

TABELA 9 - RESULTADOS DA PMU JGI

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   1817  31403   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   4000  29220   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	16,92%	100,00%	22,70%	100,00%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	5,47%	100,00%	12,04%	100,00%
<b>F1 Score</b>	<b>0,104</b>	<b>1,000</b>	<b>0,215</b>	<b>1,000</b>
<b>PCs Selecionados</b>	PC 1 = 34,25 % PC 2 = 13,62 % PC 3 = 11,64 % PC 4 = 4,29 % PC 5 = 3,84 % PC 6 = 3,40 % PC 7 = 3,33 % PC 8 = 3,17 % PC 9 = 3,09 %	PC 1 = 31,85 % PC 2 = 12,66 % PC 3 = 11,98 % PC 4 = 7,76 % PC 5 = 3,53 % PC 6 = 3,33 % PC 7 = 3,23 % PC 8 = 3,20 % PC 9 = 3,16 %	PC 1 = 34,37 % PC 2 = 13,63 % PC 3 = 11,53 % PC 4 = 4,23 % PC 5 = 3,84 % PC 6 = 3,51 % PC 7 = 3,25 % PC 8 = 3,23 % PC 9 = 3,10 %	PC 1 = 26,67 % PC 2 = 12,93 % PC 3 = 12,25 % PC 4 = 11,61 % PC 5 = 3,68 % PC 6 = 3,67 % PC 7 = 3,38 % PC 8 = 3,31 % PC 9 = 3,17 %

Fonte: O autor (2022)

Na TABELA 9, observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados2 e os Dados4 alimentando o algoritmo de análise de componentes principais. E em ambos os casos, obtém-se o critério de seleção através da acumulação de 80% da contribuição com cinco componentes principais.



A TABELA 10 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados da PMU SGD, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 3, esta configuração da primeira abordagem originalmente apresenta 26 atributos e 965 anomalias.

TABELA 10 - RESULTADOS DA PMU SGD

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   1581  31639   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  11402  21818   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	16,30%	100,00%	42,28%	100,00%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	4,76%	100,00%	34,32%	100,00%
<b>F1 Score</b>	<b>0,091</b>	<b>1,000</b>	<b>0,511</b>	<b>1,000</b>
<b>PCs Selecionados</b>	PC 1 = 41,05 % PC 2 = 16,52 % PC 3 = 14,69 % PC 4 = 6,60 % PC 5 = 4,09 %	PC 1 = 32,81 % PC 2 = 16,93 % PC 3 = 16,47 % PC 4 = 13,44 % PC 5 = 3,99 %	PC 1 = 38,40 % PC 2 = 16,28 % PC 3 = 14,79 % PC 4 = 9,33 % PC 5 = 4,12 %	PC 1 = 35,10 % PC 2 = 17,72 % PC 3 = 17,03 % PC 4 = 11,85 %

Fonte: O autor (2022)

Na TABELA 10, observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados2 e os Dados4 alimentando o algoritmo de análise de componentes principais. Porém, com os Dados2 obtém-se o critério de seleção através da acumulação de 80% da contribuição com cinco componentes principais, e com a configuração de Dados4 atinge-se o critério com apenas quatro componentes principais.

A TABELA 11 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados da PMU CVN, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 3, esta configuração da primeira abordagem originalmente apresenta 14 atributos e 537 anomalias.

TABELA 11 - RESULTADOS DA PMU CVN

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   260  32960   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   7563  25657   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	12,80%	100,00%	32,12%	100,00%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	0,78%	100,00%	22,77%	100,00%
<b>F1 Score</b>	<b>0,016</b>	<b>1,000</b>	<b>0,371</b>	<b>1,000</b>
<b>PCs Selecionados</b>	PC 1 = 33,90 % PC 2 = 17,90 % PC 3 = 13,41 % PC 4 = 11,36 % PC 5 = 6,27 %	PC 1 = 39,35 % PC 2 = 14,53 % PC 3 = 13,95 % PC 4 = 8,87 % PC 5 = 7,21 %	PC 1 = 33,13 % PC 2 = 17,90 % PC 3 = 13,68 % PC 4 = 10,85 % PC 5 = 6,38 %	PC 1 = 32,47 % PC 2 = 15,83 % PC 3 = 14,38 % PC 4 = 14,03 % PC 5 = 6,29 %

Fonte: O autor (2022)

Na TABELA 11, observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados2 e os Dados4 alimentando o algoritmo de análise de componentes principais. E em ambos os casos, obtém-se o critério de seleção através da acumulação de 80% da contribuição com cinco componentes principais.

A TABELA 12 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados da PMU GUA, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 3, esta configuração da primeira abordagem originalmente apresenta 14 atributos e 2.432 anomalias.

TABELA 12 - RESULTADOS DA PMU GUA

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   3011  30209   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   3508  29712   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  32926   294   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	20,08%	100,00%	21,40%	99,22%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	9,06%	100,00%	10,56%	99,11%
<b>F1 Score</b>	<b>0,166</b>	<b>1,000</b>	<b>0,191</b>	<b>0,996</b>
<b>PCs Selecionados</b>	PC 1 = 30,69 % PC 2 = 20,98 % PC 3 = 13,84 % PC 4 = 12,28 % PC 5 = 6,89 %	PC 1 = 24,71 % PC 2 = 22,28 % PC 3 = 15,32 % PC 4 = 14,47 % PC 5 = 6,20 %	PC 1 = 30,66 % PC 2 = 20,25 % PC 3 = 14,00 % PC 4 = 12,21 % PC 5 = 6,88 %	PC 1 = 32,54 % PC 2 = 16,86 % PC 3 = 14,75 % PC 4 = 14,50 % PC 5 = 4,53 %

Fonte: O autor (2022)

Na TABELA 12, observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados2 alimentando o algoritmo de análise de componentes principais. E neste caso, obtém-se o critério de seleção através da acumulação de 80% da contribuição com cinco componentes principais.

A TABELA 13 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados do conjunto composto pelas PMUs PIL, SMC e GRL, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 4, esta configuração da primeira abordagem originalmente apresenta 78 atributos e 692 anomalias.

TABELA 13 - RESULTADOS DO CONJUNTO PIL+SMC+GRL

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   454  32766   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  18159  15061   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   476  32744   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  22145  11075   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	13,32%	60,16%	13,38%	70,70%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	1,37%	54,66%	1,43%	66,66%
<b>F1 Score</b>	<b>0,027</b>	<b>0,707</b>	<b>0,028</b>	<b>0,800</b>
<b>PCs Selecionados</b>	PC 1 = 45,08 % PC 2 = 15,57 % PC 3 = 12,57 % PC 4 = 4,63 % PC 5 = 3,95 %	PC 1 = 33,13 % PC 2 = 15,07 % PC 3 = 12,49 % PC 4 = 11,79 % PC 5 = 4,16 % PC 6 = 3,69 %	PC 1 = 44,62 % PC 2 = 15,48 % PC 3 = 12,67 % PC 4 = 4,51 % PC 5 = 4,04 %	PC 1 = 26,90 % PC 2 = 16,26 % PC 3 = 15,07 % PC 4 = 12,72 % PC 5 = 4,55 % PC 6 = 3,61 % PC 7 = 3,47 %

Fonte: O autor (2022)

Na TABELA 13 observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados4 alimentando o algoritmo de análise de componentes principais. E neste caso, obtém-se o critério de seleção através da acumulação de 80% da contribuição com sete componentes principais.

A TABELA 14 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados do conjunto composto pelas PMUs DJP, JGI e SGD, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 4, esta configuração da primeira abordagem originalmente apresenta 84 atributos e 1.735 anomalias.

TABELA 14 - RESULTADOS DO CONJUNTO DJP+JGI+SGD

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   177  33043   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   516  32704   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	12,58%	100,00%	13,48%	100,00%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	0,53%	100,00%	1,55%	100,00%
<b>F1 Score</b>	<b>0,011</b>	<b>1,000</b>	<b>0,031</b>	<b>1,000</b>
<b>PCs Selecionados</b>	PC 1 = 37,56 % PC 2 = 13,78 % PC 3 = 12,98 % PC 4 = 4,78 % PC 5 = 4,19 % PC 6 = 3,17 % PC 7 = 3,14 % PC 8 = 2,22 %	PC 1 = 22,71 % PC 2 = 17,09 % PC 3 = 13,41 % PC 4 = 12,70 % PC 5 = 3,97 % PC 6 = 3,73 % PC 7 = 3,10 % PC 8 = 2,64 % PC 9 = 1,96 %	PC 1 = 36,61 % PC 2 = 13,60 % PC 3 = 13,04 % PC 4 = 5,55 % PC 5 = 4,12 % PC 6 = 3,26 % PC 7 = 3,17 % PC 8 = 2,28 %	PC 1 = 24,19 % PC 2 = 16,64 % PC 3 = 13,39 % PC 4 = 13,07 % PC 5 = 3,77 % PC 6 = 2,98 % PC 7 = 2,95 % PC 8 = 2,32 % PC 9 = 1,89 %

Fonte: O autor (2022)

Na TABELA 14 observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados2 e os Dados4 alimentando o algoritmo de análise de componentes principais. E em ambos os casos, obtém-se o critério de seleção através da acumulação de 80% da contribuição com nove componentes principais.

A TABELA 15 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais exclusivamente com os dados do conjunto composto pelas PMUs PIL, SMC, GRL, DJP, JGI e SGD, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 4, esta configuração da primeira abordagem originalmente apresenta 162 atributos e 962 anomalias.

TABELA 15 - RESULTADOS DO CONJUNTO PIL+SMC+GRL+DJP+JGI+SGD

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   345  32875   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33095   125   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   627  32593   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	13,03%	99,67%	13,78%	100,00%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	1,04%	99,62%	1,89%	100,00%
<b>F1 Score</b>	<b>0,021</b>	<b>0,998</b>	<b>0,037</b>	<b>1,000</b>
<b>PCs Selecionados</b>	PC 1 = 40,95 % PC 2 = 14,52 % PC 3 = 12,70 % PC 4 = 4,67 % PC 5 = 4,01 % PC 6 = 3,11 % PC 7 = 3.06 %	PC 1 = 26,07 % PC 2 = 14,81 % PC 3 = 13,99 % PC 4 = 12,45 % PC 5 = 4,87 % PC 6 = 3,74 % PC 7 = 3.04 % PC 8 = 2.84 %	PC 1 = 40,24 % PC 2 = 14,39 % PC 3 = 12,77 % PC 4 = 4,84 % PC 5 = 4,06 % PC 6 = 3,17 % PC 7 = 3.14 %	PC 1 = 24,07 % PC 2 = 17,40 % PC 3 = 13,92 % PC 4 = 12,92 % PC 5 = 4,24 % PC 6 = 3,19 % PC 7 = 3.15 % PC 8 = 2.23 %

Fonte: O autor (2022)

Na TABELA 15 observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados4 alimentando o algoritmo de análise de componentes principais. E neste caso, obtém-se o critério de seleção através da acumulação de 80% da contribuição com oito componentes principais.

A TABELA 16 mostra o resultado obtido ao se executar o algoritmo de detecção de anomalias após alimentar o algoritmo de análise de componentes principais com os dados de todas as PMUs, em suas quatro configurações de alimentação do PCA. Conforme mostrado na TABELA 4, esta configuração da primeira abordagem originalmente apresenta 190 atributos e 857 anomalias.

TABELA 16 - RESULTADOS DO CONJUNTO DE TODAS AS PMUs

Itens	Dados1	Dados2	Dados3	Dados4
<b>Matriz de Confusão</b>	+-----+-----+   566  32654   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+	+-----+-----+   600  32620   +-----+-----+   0   4580   +-----+-----+	+-----+-----+  33220   0   +-----+-----+   0   4580   +-----+-----+
<b>Acurácia</b>	13,61%	100,00%	13,70%	100,00%
<b>Precisão</b>	100,00%	100,00%	100,00%	100,00%
<b>Revocação</b>	1,70%	100,00%	1,81%	100,00%
<b>F1 Score</b>	<b>0,034</b>	<b>1,000</b>	<b>0,035</b>	<b>1,000</b>
<b>PCs Selecionados</b>	PC 1 = 37,09 % PC 2 = 14,52 % PC 3 = 12,91 % PC 4 = 6,15 % PC 5 = 4,45 % PC 6 = 3,28 % PC 7 = 3,13 %	PC 1 = 26,10 % PC 2 = 14,38 % PC 3 = 13,94 % PC 4 = 12,51 % PC 5 = 5,11 % PC 6 = 4,02 % PC 7 = 3,09 % PC 8 = 2,83 %	PC 1 = 36,44 % PC 2 = 14,36 % PC 3 = 12,93 % PC 4 = 6,29 % PC 5 = 4,54 % PC 6 = 3,38 % PC 7 = 3,16 %	PC 1 = 23,99 % PC 2 = 16,99 % PC 3 = 13,79 % PC 4 = 12,91 % PC 5 = 4,91 % PC 6 = 3,22 % PC 7 = 3,18 % PC 8 = 2,28 %

Fonte: O autor (2022)

Na TABELA 16 observa-se que tanto o melhor resultado de acurácia, quanto do *F1 Score*, são obtidos com os Dados2 e os Dados4 alimentando o algoritmo de análise de componentes principais. E em ambos os casos, obtém-se o critério de seleção através da acumulação de 80% da contribuição com oito componentes principais.

Em todas as tabelas pode-se observar que os melhores resultados obtidos para o *F1 Score* acompanham os melhores resultados obtidos para a acurácia. Pelo fato do *F1 Score* ser o resultado combinado de outras métricas, este será adotado para fazer as análises subsequentes. A TABELA 17 mostra uma compilação dos valores obtidos pela métrica do *F1 Score* para todas as configurações das duas abordagens.

TABELA 17 - RESULTADOS DA MÉTRICA F1 SCORE

	Dados1	Dados2	Dados3	Dados4
PIL	0,148	0,956	0,083	0,767
SMC	0,021	0,516	0,008	0,982
GRL	0,046	1	0,171	1
DJP	0,000	0,922	0,001	0,952
JGI	0,104	1	0,215	1
SGD	0,091	1	0,511	1
CVN	0,016	1	0,371	1
GUA	0,166	1	0,191	0,996
PIL+SMC+GRL	0,027	0,707	0,028	0,8
DJP+JGI+SGD	0,011	1	0,031	1
PIL+SMC+GRL+DJP+JGI+SGD	0,021	0,998	0,037	1
Todas as oito PMUs	0,034	1	0,035	1

Fonte: O autor (2022)

A TABELA 17 mostra em destaque na cor vermelha os melhores resultados da *F1 Score* obtidos com as diferentes configurações na alimentação do PCA para cada agrupamento de PMUs.

### 5.2.2 Eixo de análise 1 – Quantidade de PMUs

A primeira linha de análise que será feita, tem por base a primeira abordagem, relacionada à quantidade de PMUs. Para cada PMU individualizada e para cada agrupamento de PMUs, serão analisados os resultados obtidos na execução do procedimento.

Foi feita uma comparação, linha a linha, dos resultados. Pode-se observar na TABELA 17 que para oito dos doze conjuntos, o melhor resultado obtido foi com a alimentação do PCA através do conjunto Dados2. Isto significa que para 66% dos casos esta é a forma mais efetiva de alimentar o algoritmo do PCA. Porém para dez dos doze conjuntos, o melhor resultado obtido foi com a alimentação do PCA através dos Dados4, ou seja, para 83% dos casos esta é a forma mais efetiva de alimentar o algoritmo do PCA.

Observa-se ainda que no caso da configuração de Dados2, existe uma variação maior entre os resultados. É o caso específico da PMU SMC que possui uma



variação de 94% para os melhores resultados e do conjunto PIL+SMC+GRL que possui uma variação de 41%. No caso da configuração de Dados4, observa-se que o maior valor de variação é 30%, que é o caso da PMU PIL que é a de pior desempenho.

### 5.2.3 Eixo de análise 2 – Alimentação do PCA

A segunda linha de análise feita, tem por base a abordagem relacionada à fonte de alimentação do algoritmo de análise de componentes principais, verificando para cada uma das quatro configurações, os resultados obtidos em cada um dos conjuntos de dados de PMUs (individualizados e agrupados).

Para isto, foi feita uma comparação, coluna a coluna, dos resultados. Neste caso, não foi facilmente identificado um padrão, como no caso da análise anterior. Para cada configuração diferente para a fonte de entrada de dados, os resultados obtidos em cada conjunto de dados de PMUs foram bastante diversos. Os dados da TABELA 18 até a TABELA 21 mostram os resultados obtidos para o *F1 Score* em cada um dos quatro casos, exibindo em ordem decrescente do maior para o menor, tendo na coluna “Variação” o percentual de variação do item para o melhor resultado.

TABELA 18 - RESULTADOS DO PCA – DADOS1

Conjuntos de PMUs	F1 Score	Variação
GUA	0,166	
PIL	0,148	12%
JGI	0,104	60%
SGD	0,091	82%
GRL	0,046	261%
Todas as 8 PMUs	0,034	388%
PIL+SMC+GRL	0,027	515%
SMC	0,021	690%
PIL+SMC+GRL+DJP+JGI+SGD	0,021	690%
CVN	0,016	938%
DJP+JGI+SGD	0,011	1409%
DJP	0	--

Fonte: O autor (2022)

TABELA 19 - RESULTADOS DO PCA – DADOS2

Conjuntos de PMUs	F1 Score	Variação
Todas as 8 PMUs	1	
SGD	1	0%
JGI	1	0%
GUA	1	0%
GRL	1	0%
DJP+JGI+SGD	1	0%
CVN	1	0%
PIL+SMC+GRL+DJP+JGI+SGD	0,998	0%
PIL	0,956	5%
DJP	0,922	8%
PIL+SMC+GRL	0,707	41%
SMC	0,516	94%

Fonte: O autor (2022)

TABELA 20 - RESULTADOS DO PCA - DADOS3

Conjuntos de PMUs	F1 Score	Variação
SGD	0,511	
CVN	0,371	38%
JGI	0,215	138%
GUA	0,191	168%
GRL	0,171	199%
PIL	0,083	516%
PIL+SMC+GRL+DJP+JGI+SGD	0,037	1281%
Todas as 8 PMUs	0,035	1360%
DJP+JGI+SGD	0,031	1548%
PIL+SMC+GRL	0,028	1725%
SMC	0,008	6288%
DJP	0,001	51000%

Fonte: O autor (2022)

TABELA 21 - RESULTADOS DO PCA - DADOS4

Conjuntos de PMUs	F1 Score	Variação
Todas as 8 PMUs	1	
SGD	1	0%
PIL+SMC+GRL+DJP+JGI+SGD	1	0%
JGI	1	0%
GRL	1	0%
DJP+JGI+SGD	1	0%
CVN	1	0%
GUA	0,996	0%
SMC	0,982	2%
DJP	0,952	5%
PIL+SMC+GRL	0,8	25%
PIL	0,767	30%

Fonte: O autor (2022)

A TABELA 19 e a TABELA 21 apresentam os melhores resultados do *F1 Score* e compartilham uma característica com relação ao tratamento individual comparado ao agrupamento de PMUs. Observa-se na TABELA 19 que o pior resultado é obtido pela SMC, e este resultado influencia uma performance ruim também para o conjunto PIL+SMC+GRL, porém constata-se que a abordagem agrupada destas PMUs traz um resultado melhor do que a abordagem individual da PMU SMC, compensando o desempenho ruim desta. Na TABELA 21 é possível fazer a mesma constatação, porém com relação ao resultado da PMU PIL.

Observa-se também que nestas duas tabelas o conjunto de dados composto por todas as PMUs traz um ótimo desempenho, compensando desempenhos inferiores proporcionados pelas análises individuais. Estas observações apontam para resultados mais consistentes ao se adotar uma análise agrupada.

#### 5.2.4 Redução da dimensionalidade

Outro aspecto importante a ser observado nos resultados obtidos é com relação à redução de dimensionalidade obtida com a execução do algoritmo de análise de componentes principais. A TABELA 22 mostra a quantidade original de atributos existente em cada configuração de conjuntos de PMUs e relacionado ao lado a quantidade de componentes principais obtida na execução de cada cenário.

TABELA 22 - QUANTIDADE DE PCs POR CENÁRIO

Conjuntos de PMUs	Qtde. atributos original	Quantidade de PCs			
		Dados1	Dados2	Dados3	Dados4
PIL	32	4	5	4	6
SMC	20	5	5	5	6
GRL	26	5	5	5	5
DJP	26	4	6	5	6
JGI	32	9	9	9	9
SGD	26	5	5	5	4
CVN	14	5	5	5	5
GUA	14	5	5	5	5
PIL+SMC+GRL	78	5	6	5	7
DJP+JGI+SGD	84	8	9	8	9
PIL+SMC+GRL+DJP+JGI+SGD	162	7	8	7	8
Todas as 8 PMUs	190	7	8	7	8

Fonte: O autor (2022)

Na TABELA 22 observa-se que não existe muita diferença na quantidade de componentes principais obtidos ao se variar a abordagem do eixo de análise relacionado à alimentação do PCA. Em apenas três casos a variação chega ser de dois componentes principais, nos outros nove casos a variação é de apenas um componente principal.

A TABELA 23 mostra a taxa de compressão obtida em cada um dos 48 cenários, calculada através da simples divisão da quantidade original de atributos pela quantidade de componentes principais obtida.

TABELA 23 - TAXA DE COMPRESSÃO POR CENÁRIO

Conjuntos de PMUs	Taxa de Compressão			
	Dados1	Dados2	Dados3	Dados4
PIL	800,0%	640,0%	800,0%	533,3%
SMC	400,0%	400,0%	400,0%	333,3%
GRL	520,0%	520,0%	520,0%	520,0%
DJP	650,0%	433,3%	520,0%	433,3%
JGI	355,6%	355,6%	355,6%	355,6%
SGD	520,0%	520,0%	520,0%	650,0%
CVN	280,0%	280,0%	280,0%	280,0%
GUA	280,0%	280,0%	280,0%	280,0%
PIL+SMC+GRL	1560,0%	1300,0%	1560,0%	1114,3%
DJP+JGI+SGD	1050,0%	933,3%	1050,0%	933,3%
PIL+SMC+GRL+DJP+JGI+SGD	2314,3%	2025,0%	2314,3%	2025,0%
Todas as 8 PMUs	2714,3%	2375,0%	2714,3%	2375,0%

Fonte: O autor (2022)

Ao se observar a taxa de compressão na TABELA 23 pelo aspecto do eixo de abordagem relacionado à quantidade de PMUs, observa-se que quanto maior a quantidade de atributos maior é a taxa de compressão obtida. Isto implica em dizer que uma quantidade maior de atributos originais obtida pela junção de várias PMUs pode ser representada quase que pela mesma quantidade de componentes principais obtida pela análise de uma única PMU. Porém, neste caso, deve-se observar que as equações que formam cada componente principal serão compostas por uma quantidade de coeficientes maior.

### 5.3 DISCUSSÃO

A análise dos resultados no eixo de análise 1, que aborda a quantidade de PMUs, indica que independente da quantidade de PMUs utilizada, a estratégia mais eficaz para alimentação do algoritmo de análise de componentes principais é com o conjunto de Dados4, que leva uma pequena vantagem sobre o conjunto de Dados2.

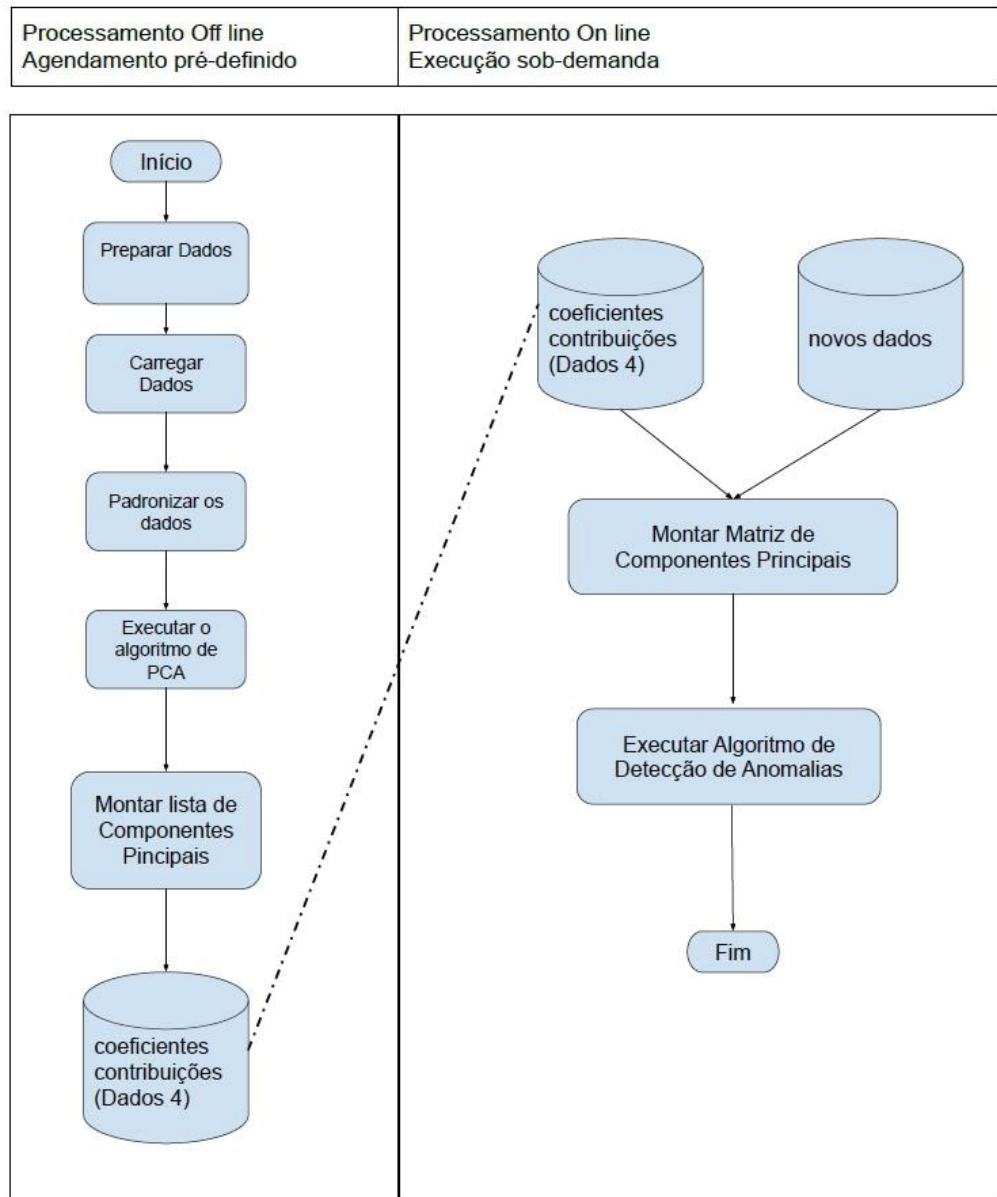
A análise dos resultados no eixo de análise 2, que aborda a perspectiva da alimentação do algoritmo de análise de componentes principais, indica que os melhores resultados são obtidos quando se utiliza dados da maior quantidade possível de PMUs de forma agrupada.

A análise da redução de dimensionalidade, através da quantidade de componentes principais, indica que é possível obter taxas de compressão melhores utilizando-se mais atributos obtidos através do maior agrupamento possível de PMUs.

Ao combinar os resultados destas análises, infere-se que o melhor cenário é quando da utilização de dados de validação cruzada e teste (Dados4), da maior quantidade possível de PMUs, para alimentação do algoritmo de análise de componentes principais. Este é o cenário representado na opção 3 do fluxo apresentado na FIGURA 8, da seção 4.2.1.

Este cenário permite interpretar que pode ser adotada uma metodologia em que é possível realizar um pré-processamento *offline* dos dados, para obtenção dos parâmetros de execução do algoritmo de análise de componentes principais. Este pré-processamento vai gerar a base de treino para o algoritmo de análise de dados, no caso deste trabalho para detecção de anomalias, e por ser *offline* pode ser executado periodicamente, de acordo com as necessidades e possibilidades da infraestrutura existente. Após a geração desta base de treino, o algoritmo de análise de dados pode ser executado a qualquer momento que for necessário, bastando acrescentar ao conjunto de dados de treino o conjunto de dados que se deseja verificar. Esta sequência de ações é observada na FIGURA 13.

FIGURA 13 - FLUXO DE ANÁLISE DE DADOS



FONTE: O autor (2022)

A proposta de XIE et al. (2014) foi de implementar a redução de dimensionalidade durante o estágio de treinamento, o que foi demonstrado neste trabalho como sendo a melhor alternativa entre as quatro opções avaliadas.

Conforme proposto por Ge et al. (2017), o método PCA mostrou-se eficaz para a redução de dimensionalidade dos atributos, caracterizando uma grande correlação entre os atributos das PMUs utilizadas, fazendo com que sejam facilitadas as análises posteriores realizadas a partir destes dados.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Os sistemas de medição fasorial sincronizada estão sendo amplamente utilizados ao redor do mundo, e se tornaram elementos essenciais na evolução da operação dos sistemas elétricos de potência (SEP) de grande porte. Estes sistemas são baseados em unidades de medição fasorial, chamadas PMUs, as quais são capazes de gravar e comunicar dados da dinâmica dos SEPs de forma sincronizada por GPS e com alta taxa de amostragem (até 60 amostras/segundo), gerando assim um enorme conjunto de dados. A abordagem analítica de dados é um conceito que se refere ao processo de coletar e examinar conjuntos de dados para tirar conclusões sobre as informações que eles contêm com o auxílio de sistemas e softwares inteligentes.

Neste contexto, este trabalho propôs uma abordagem para a junção de dados fasoriais sincronizados para seleção e tratamento dos mais relevantes integrados sistemicamente, de forma a disponibilizá-los para análises posteriores mais otimizadas.

Primeiramente foi realizado um levantamento da arquitetura do sistema de medição fasorial da empresa COPEL Geração e Transmissão S.A., tendo sido identificado que se trata de um sistema composto por 56 PMUs. Para este trabalho foram disponibilizados os arquivos de dados de 41 PMUs, tendo sido analisada a estrutura e qualidade dos dados do sistema, através da verificação de falha de leitura ou coleta dos dados.

Foi desenvolvida uma metodologia para seleção de dados para análise pós-falta, através da redução da dimensionalidade, utilizando a análise de componentes principais. Com a aplicação da metodologia, comprovou-se a possibilidade de se utilizar um conjunto de dados menor do que o original, porém mantendo as características dos atributos. Demonstrou-se um caso em que foi possível fazer a redução de 190 atributos, para um conjunto de oito componentes principais com uma precisão de 100% no *F1 Score*. Esta metodologia automática de seleção de dados traz benefícios para análise pós-falta devido ao ganho de desempenho obtido com a redução do conjunto de dados dos atributos elétricos, sem perda significativa na precisão da capacidade de detecção de anomalia.

Para validar a metodologia proposta, usando os dados reais que estavam disponíveis, foram identificados alguns recentes casos reais conhecidos de



ocorrências no sistema interligado nacional, para que a análise pós-falta conseguisse detectar a anomalia. O critério de seleção do caso para análise foi com relação à qualidade dos dados, portanto selecionou-se o caso que apresentou a melhor qualidade dos dados disponíveis considerando o conceito da perda de dados, ou seja, foi selecionado o caso em que se constatou a menor perda de dados. Com este critério foi feita a seleção de oito PMUs para utilização nos casos de uso, as quais estão espalhadas geograficamente pelo estado do Paraná.

Os cenários foram montados com base em dois eixos de abordagem. Foi feita uma análise pelo eixo de abordagem relacionado à quantidade de PMUs utilizada, a qual indicou que a estratégia mais eficaz para alimentação do algoritmo de análise de componentes principais é com os dados de validação cruzada e teste. Foi feita também uma análise por um eixo de abordagem relacionado ao conjunto de dados que alimentaria o algoritmo de análise de componentes principais, a qual indicou que os melhores resultados são obtidos quando se utiliza dados da maior quantidade possível de PMUs de forma agrupada, situação em que também obtém-se a maior taxa de compressão da dimensionalidade. A combinação dos resultados destas duas abordagens indica que o melhor cenário é quando da utilização de dados de validação cruzada e teste, da maior quantidade possível de PMUs, para alimentação do algoritmo de análise de componentes principais.

A validação dos resultados obtidos foi feita utilizando-se um algoritmo de detecção de anomalias utilizando um modelo gaussiano. Este algoritmo primeiro foi executado através do programa computacional com o conjunto de dados original, tendo em seguida sido executado com o conjunto de dados de dimensionalidade reduzida. O F1 Score foi a métrica adotada para a comparação destes resultados. Este procedimento foi repetido para todos os cenários.

Desta forma, entende-se que a aplicação da metodologia de análise de componentes principais para redução de dimensionalidade dos dados coletados a partir de um grande conjunto de PMUs contribui para a análise sistêmica pós-falta em sistemas elétricos de grande porte.

O método desenvolvido permite a utilização com qualquer conjunto de dados de PMUs, independente de fabricante ou fornecedor destes equipamentos, ou da quantidade disponível. Foi demonstrada a aplicação da metodologia em cenários variando de apenas uma PMU a oito PMUs diferentes.

Observa-se como limitação do presente trabalho que foram usados apenas dados de boa qualidade e que foi feita uma análise pós-falta através de um único algoritmo de detecção de anomalias.

Como trabalhos futuros, sugere-se: automatizar a análise da qualidade dos dados de entrada, utilizando-se técnicas específicas para este fim, para garantir maior precisão no processo e maior volume de dados; utilizar os dados selecionados para mais aplicações de análise pós-falta além de detecção de anomalias, tais como localização de falta e estabilidade de tensão.

Sugere-se também utilizar outra técnica como a análise de componentes independentes (ICA) que tenta decompor um sinal multivariado em sinais não gaussianos independentes, a qual pode ser utilizada para remoção de ruídos no sinal de entrada e também para a previsão de séries temporais multivariadas.

## REFERÊNCIAS

AHMAD, T.; SENROY, N. Statistical Characterization of PMU Error for Robust WAMS Based Analytics. **IEEE Transactions on Power Systems**, v. 35, n. 2, p. 920–928, 2020. IEEE.

BRUCE, P.; BRUCE, A. **Estatística prática para cientistas de dados: 50 conceitos essenciais**. Rio de Janeiro: Alta Books, 2019.

CAI, L.; THORNHILL, N. F.; KUENZEL, S.; Wide-Area Monitoring of Power Systems Using Principal Component Analysis and -Nearest Neighbor Analysis. **IEEE Transactions on Power Systems**, 2018. IEEE. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8272321/>>. .

CARDOSO JR, G.; ROLIM, J. G.; ZÜRN, H. H. Diagnóstico de faltas em sistemas de potência: definição do problema e abordagens via inteligência artificial. **Sba: Controle & Automação Sociedade Brasileira de Automatica**, v. 15, n. 2, p. 215–229, 2004.

CHATTERJEE, K.; CHAUDHURI, N. R. Corruption-Resilient Detection of Event-Induced Outliers in PMU Data: A Kernel PCA Approach. **2019 IEEE Power & Energy Society General Meeting**, 2019. IEEE. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8973424/>>. .

CHU, L.; QIU, R.; HE, X.; LING, Z.; LIU, Y. Massive Streaming PMU Data Modelling and Analytics in Smart Grid State Evaluation based on Multiple High-Dimensional Covariance Test. **IEEE Transactions on Big Data**, v. 4, n. 1, p. 55–64, 2017.

DIMITROVSKA, T.; RUDEŽ, U.; ... Real-Time Application of an Indirect Power-System Contingency Screening Method Based on Adaptive PCA. **IEEE Transactions on Power Systems**, 2019. IEEE. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8723513/>>. .

ENSSLIN L, ENSSLIN SR, LACERDA RTO, TASCA JE. Processo de seleção de portfólio bibliográfico, 2010.

GE, Z.; SONG, Z.; DING, S. X.; HUANG, B. Data mining and analytics in the process industry: The role of machine learning. **IEEE Access**, 2017. IEEE. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8051033/>>. .

JOHNSON. R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6. Ed. New Jersey: Pearson Prentice Hall, 2007.

JONES, K. D.; CANO, E. B.; CHEN, H.; ROBINSON, F.; THOMAS, K.; GARDNER, R. M. Strategies for success with synchrophasors. **IEEE Power & Energy Magazine**, v. 13, n. 5, p. 29-35, setembro/outubro 2015

KARLSSON, D.; HEMMINGSSON, M.; LINDAHL. S. Wide Area System Monitoring and Control. **IEEE Power & Energy Magazine**, p. 68-76, setembro/outubro 2004

LE, N. T.; BENJAPOLAKUL, W. A data imputation model in phasor measurement units based on bagged averaging of multiple linear regression. **IEEE Access**, 2018. IEEE. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8412191/>>. .

LIU, X.; LAVERTY, D. M.; BEST, R. J.; LI, K.; Principal component analysis of wide-area phasor measurements for islanding detection—A geometric view. **Transactions on Power Delivery**, 2015. IEEE. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/7041222/>>. .

LU, C.; SHI, B.; WU, X.; SUN, H. Advancing China's Smart Grid: Phasor Measurement Units in a Wide-Area Management System. **IEEE Power and Energy Magazine**, v. 13, n. 5, p. 60–71, 2015. IEEE.

MADANI, V.; GIRI, J.; KOSTEREV, D.; NOVOSEL, D.; BRANCACCIO, D. Challenging changing landscapes. **IEEE Power & Energy Magazine**, v. 13, n. 5, p. 18-28, setembro/outubro 2015

MANLY, B. F. J.; ALBERTO, J. A. N. **Métodos estatísticos multivariados: uma introdução**. 4. ed. Porto Alegre: Bookman, 2019.

MARDIA, K. V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. London: Academic Press, 1979.

MARTIN, K.; CHEN, K. Impact of Phasor Measurement Data Quality in Grid Operations. In: NUTHALAPATI, S. (Ed.). **Power System Grid Operation Using Synchrophasor Technology**. Springer, 2018. P. 13-40.

MOHARM, K. State of the art in big data applications in microgrid: A review. **Advanced Engineering Informatics**, 2019. Elsevier. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S147403461830702X>>. .

MONFREDA, M. A Powerful Interpretative Tool at the Service of Analytical Methodology. In: SANQUANSAT, P. (Ed.). **Principal Component Analysis**. InTech, 2012. p. 49-49.

North American Synchrophasor Initiative (NASPI). **PMU Data Quality: A framework for the attributes of PMU data quality and a methodology for examining data quality impacts to synchrophasor applications**. 2017. *White Paper*.

ONS. O Sistema Interligado Nacional, 2020. Disponível em: <<http://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>>. Acesso em: novembro de 2020.

Operador Nacional do Sistema vai adotar monitoramento da Copel. **Agência Estadual de Notícias**, Curitiba, 10 mar. 2017. Disponível em: <<http://www.historico2.aen.pr.gov.br/modules/noticias/article.php?storyid=93029&tit=Operador-Nacional-do-Sistema-vai-adotar-monitoramento-da-Copel>>

- OVERHOLT, P.; ORTIZ, D.; SILVERSTEIN, A. Synchrophasor technology and the DOE: Exciting opportunities lie ahead in development and deployment. **IEEE Power and Energy** 2015. [ieeexplore.ieee.org](https://ieeexplore.ieee.org). Disponível em: <<https://ieeexplore.ieee.org/abstract/document/7203260/>>. .
- PHADKE, A. G. PMU memories: looking back over 40 years. **IEEE Power & Energy Magazine**, v. 13, n. 5, p. 93-96, setembro/outubro 2015
- PHADKE, A. G.; BI, T. Phasor measurement units, WAMS, and their applications in protection and control of power systems. **Journal of Modern Power Systems and Clean Energy**, volume 6, p. 619-629, julho 2018.
- PHADKE, A. G.; THORP, J. S. **Synchronized Phasor Measurements and Their Applications**. 2. ed. Springer, 2017
- SATTINGER, W.; GIANNUZZI, G. Monitoring Continental Europe: An Overview of WAM Systems Used in Italy and Switzerland. **IEEE Power and Energy Magazine**, v. 13, n. 5, p. 41–48, 2015. IEEE.
- XIE, L.; CHEN, Y.; KUMAR, P. R. Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis. **IEEE Transactions on Power Systems**, 2014. IEEE. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/6808416/>>. .
- XU, Y.; SUN, Y.; WAN, J.; LIU, X.; SONG, Z. Industrial Big Data for Fault Diagnosis: Taxonomy, Review, and Applications. **IEEE Access**, v. 5, p. 17368–17380, 2017.
- ZHANG, Y.; XU, Y.; DONG, Z. Y. Robust Ensemble Data Analytics for Incomplete PMU Measurements-Based Power System Stability Assessment. **IEEE Transactions on Power Systems**, v. 33, n. 1, p. 1124–1126, 2017. IEEE.
- ZHOU, D.; GUO, J.; ZHANG, Y.; CHAI, J.; LIU, H.; LIU, Y.; HUANG, C.; GUI, X. LIU, Y. Distributed Data Analytics Platform for Wide-Area Synchrophasor Measurement Systems. **IEEE Transactions on smart grid**, v. 7, n. 5, p. 2397-2405, setembro 2016.

## APÊNDICE 1 – EXEMPLO DE EXECUÇÃO DE PCA

O exemplo de execução do algoritmo de análise de componentes principais utiliza a simulação de um arquivo de uma PMU com apenas algumas grandezas elétricas e uma quantidade reduzida de registros para facilitar a visualização dos cálculos executados. Na sequência o procedimento é descrito passo a passo.

### Passo 1 – Carregar os dados

A primeira etapa é coletar e preparar o arquivo de dados. Neste exemplo é utilizada a simulação de um arquivo de dados coletados de uma PMU. Este arquivo é composto de nove atributos:

- Frequência;
- VA1 V\_BARRA\_230kV Magnitude;
- VA1 V\_BARRA\_230kV Angle;
- IA1 I\_PIL\_230kV Magnitude;
- IA1 I\_PIL\_230kV Angle;
- IB1 I\_PIL\_230kV Magnitude;
- IB1 I\_PIL\_230kV Angle;
- IC1 I\_PIL\_230kV Magnitude;
- IC1 I\_PIL\_230kV Angle.

Na estrutura do arquivo, cada atributo significa uma coluna, portanto o arquivo do exemplo é composto por 9 colunas.

O arquivo é composto por 20 linhas. Cada linha representa uma leitura realizada pela PMU em um instante de tempo diferente, ou seja, neste arquivo há 20 amostras de leituras diferentes.

Portanto, o arquivo de dados original corresponde a uma matriz (ou tabela) de 20 linhas e 9 colunas, conforme pode ser observado na TABELA 24.

TABELA 24 - DADOS DE EXEMPLO

Frequencia	VA1 V_BARRA_230k Magnitude	VA1 V_BARRA_230k Angle	IA1 I_PIL_230kV Magnitude	IA1 I_PIL_230kV Angle	IB1 I_PIL_230kV Magnitude	IB1 I_PIL_230kV Angle	IC1 I_PIL_230kV Magnitude	IC1 I_PIL_230kV Angle
59,995884	234721,5938	-38,759216	161,140137	159,813065	139,792236	39,040302	154,202255	-74,056831
59,995747	234877,7344	64,358185	149,724594	-96,880051	128,657852	142,375687	143,169586	29,541449
60,05061	235344,7813	-77,449036	119,078575	122,090202	99,535378	1,28456	113,481361	-110,706078
60,053352	235397,6875	140,2966	104,759247	-20,015644	85,713394	-140,985519	99,389198	107,804649
59,999168	235469,4375	-137,807205	87,41449	63,610638	69,163269	-57,113621	82,906342	-167,537582
59,98307	235508,9531	-70,09877	78,678459	131,726166	60,583057	11,035629	74,419205	-98,683884
59,976151	235658,4688	-146,062454	70,982048	55,90802	53,25399	-64,745255	67,05703	-173,751419
60,030491	235703,4844	59,807262	67,660645	-98,019814	49,989178	141,325089	63,851284	32,735207
60,01461	235551,8438	-41,577637	66,341576	161,972397	48,845463	41,807598	62,877396	-67,197731
59,974556	235695,125	-56,206402	63,633713	148,201538	46,239735	28,298553	60,303825	-80,58638
59,972183	235613,3281	108,528534	63,853165	-47,41291	46,592445	-167,353882	60,655853	83,648186
59,987183	235813,3125	-14,015858	64,960487	-169,927765	47,637699	70,033165	61,667675	-39,002537
59,699997	240969,2188	-166,601776	26,156544	26,514076	13,424844	-103,878334	23,005125	165,314285
59,732357	237955,5313	-92,14994	30,719246	49,670177	25,829597	-96,88205	22,00359	174,592133
59,86142	237254,8594	-113,617455	31,062504	32,025307	25,334633	-114,473694	22,450531	158,057007
59,890411	234620,4688	-154,466049	26,274044	16,147686	15,527028	-131,566849	19,338787	153,000412
60,030895	234530,2031	67,977905	32,195587	-116,331085	19,745014	105,236145	25,529186	18,614689
60,117641	234404,6406	-135,073654	32,875774	43,866547	19,830507	-92,211609	26,41782	179,066956
60,134007	234156,5781	77,454285	37,557301	-100,556084	23,808165	128,151642	31,426981	33,357956
60,149174	233845,5938	34,270061	39,339825	-142,386871	25,35988	87,872101	33,362576	-8,989517

FONTE: O autor (2022)

### Passo 2 – Padronizar os dados

Os dados representados na TABELA 24 estão em escalas diferentes. Por exemplo, as informações relativas à frequência (Frequência) estão entre 59,70 e 60,15, as informações relativas à tensão (VA1 V\_BARRA\_230k Magnitude) estão entre 233.845,59 e 240.969,22 e as informações relativas à corrente (IA1 I\_PIL\_230kV Magnitude) estão entre 26,16 e 161,14. O objetivo da etapa de padronização é minimizar estas diferenças de escala e fazer com que todos os atributos sejam representados através de uma mesma escala, permitindo assim que possam ser comparados entre si sem causar distorções pelas enormes diferenças de valores.

A análise de componentes principais pode ser feita tanto usando a padronização, quanto a normalização, porém neste exemplo utiliza-se a padronização. Após aplicação da equação (20), a tabela de dados padronizada é representada como na TABELA 25.

TABELA 25 - DADOS PADRONIZADOS

Frequencia	VA1 V_BARRA_ 230k Magnitude	VA1 V_BARRA_ 230k Angle	IA1 I_PIL_230k V Magnitude	IA1 I_PIL_230k V Angle	IB1 I_PIL_230k V Magnitude	IB1 I_PIL_230k V Angle	IC1 I_PIL_230k V Magnitude	IC1 I_PIL_230k V Angle
0,116988	-0,591696	-0,043613	2,336612	1,410241	2,402146	0,467634	2,287528	-0,777329
0,115795	-0,492679	1,027263	2,051087	-1,022348	2,096644	1,481168	2,012689	0,119262
0,593395	-0,196499	-0,445407	1,284569	1,052755	1,29759	0,097318	1,273113	-1,09451
0,617265	-0,162949	1,815885	0,926415	-0,293932	0,918346	-1,298096	0,922057	0,796591
0,145576	-0,117448	-1,072228	0,492588	0,498565	0,464249	-0,475464	0,511446	-1,586358
0,005438	-0,092389	-0,369075	0,274083	1,144071	0,228827	0,192958	0,300019	-0,990463
-0,054794	0,002427	-1,157959	0,08158	0,42557	0,027735	-0,550316	0,116617	-1,640135
0,418252	0,030973	0,980002	-0,001495	-1,033149	-0,061844	1,470864	0,036757	0,146903
0,280003	-0,06519	-0,072882	-0,034487	1,430704	-0,093225	0,494776	0,012496	-0,717967
-0,068679	0,025672	-0,224802	-0,102216	1,300203	-0,16472	0,362277	-0,051615	-0,833839
-0,089337	-0,026199	1,485973	-0,096727	-0,553566	-0,155043	-1,556722	-0,042846	0,587529
0,041243	0,100621	0,213347	-0,069031	-1,714595	-0,126364	0,771618	-0,01764	-0,473951
-2,458801	3,370258	-1,37126	-1,039594	0,147014	-1,065086	-0,934141	-0,980779	1,294308
-2,177097	1,459117	-0,598076	-0,925472	0,366456	-0,724728	-0,86552	-1,005729	1,374603
-1,053563	1,014783	-0,821017	-0,916886	0,199242	-0,738308	-1,038063	-0,994595	1,2315
-0,801188	-0,655825	-1,24523	-1,036655	0,048775	-1,007407	-1,205716	-1,072113	1,187738
0,421769	-0,713067	1,064854	-0,888546	-1,206679	-0,891675	1,116896	-0,917901	0,024697
1,17692	-0,792693	-1,04384	-0,871533	0,311457	-0,889329	-0,819712	-0,895764	1,413331
1,319392	-0,950003	1,163267	-0,754439	-1,057185	-0,780191	1,341656	-0,770979	0,152292
1,451425	-1,147215	0,714798	-0,709854	-1,4536	-0,737616	0,946586	-0,72276	-0,214204

FONTE: O autor (2022)

### Passo 3 – Obter os Autovalores e Autovetores

Calcula-se a matriz de correlação da tabela de dados padronizada. Como esta tabela possui 9 atributos, então a matriz de correlação é uma matriz 9x9, conforme pode-se observar na TABELA 26.



TABELA 26 - MATRIZ DE CORRELAÇÃO

	Frequencia	VA1 V_BARRA_230k Magnitude	VA1 V_BARRA_230k Angle	IA1 I_PIL_230kV Magnitude	IA1 I_PIL_230kV Angle	IB1 I_PIL_230kV Magnitude	IB1 I_PIL_230kV Angle	IC1 I_PIL_230kV Magnitude	IC1 I_PIL_230kV Angle
Frequencia	1	-0,86953	0,499264	0,255662	-0,24484	0,218675	0,486089	0,258989	-0,38927
VA1 V_BARRA_230k Magnitude	-0,86953	1	-0,41512	-0,27018	0,188579	-0,25004	-0,41144	-0,26251	0,343626
VA1 V_BARRA_230k Angle	0,499264	-0,41512	1	0,255697	-0,56239	0,239409	0,409807	0,258019	0,036833
IA1 I_PIL_230kV Magnitude	0,255662	-0,27018	0,255697	1	0,250964	0,997224	0,256822	0,999022	-0,50186
IA1 I_PIL_230kV Angle	-0,24484	0,188579	-0,56239	0,250964	1	0,260089	-0,34425	0,25175	-0,28654
IB1 I_PIL_230kV Magnitude	0,218675	-0,25004	0,239409	0,997224	0,260089	1	0,237348	0,993393	-0,46688
IB1 I_PIL_230kV Angle	0,486089	-0,41144	0,409807	0,256822	-0,34425	0,237348	1	0,259227	-0,41168
IC1 I_PIL_230kV Magnitude	0,258989	-0,26251	0,258019	0,999022	0,25175	0,993393	0,259227	1	-0,51977
IC1 I_PIL_230kV Angle	-0,38927	0,343626	0,036833	-0,50186	-0,28654	-0,46688	-0,41168	-0,51977	1

FONTE: O autor (2022)

A partir da matriz de correlação de ordem 9x9, calculam-se os nove autovalores e os seus respectivos autovetores.

TABELA 27 - AUTOVALORES E AUTOVETORES

	Auto Valor	% de contribuição	autoVETORES								
			1	2	3	4	5	6	7	8	9
PC1	4,030302	44,78	0,316158	-0,30789	0,231678	0,438749	0,021662	0,428645	0,274072	0,439766	-0,32613
PC2	2,382743	26,47	-0,37781	0,336654	-0,39735	0,259828	0,520597	0,273281	-0,29955	0,260357	-0,1237
PC3	1,140874	12,68	-0,32497	0,344209	0,461743	0,222392	-0,38545	0,245169	-0,08833	0,215566	0,500296
PC4	0,745903	8,288	-0,31302	0,434978	-0,11405	-0,04337	-0,21143	-0,05229	0,712425	-0,02924	-0,37694
PC5	0,304541	3,384	0,000362	-0,23656	-0,37946	0,064756	0,21984	0,120689	0,521691	0,01904	0,678558
PC6	0,277235	3,08	0,035912	0,187411	0,621245	-0,11818	0,696849	-0,1289	0,216959	-0,10538	0,057313
PC7	0,115098	1,279	0,740357	0,62396	-0,17876	0,047318	-0,03545	-0,01989	-0,01677	0,085602	0,138163
PC8	0,003282	0,036	0,065336	0,061024	0,017997	-0,13041	-0,00707	0,747214	-0,01682	-0,64257	-0,0559
PC9	0,000021	2E-04	-0,00513	0,007222	0,003419	0,806996	0,000816	-0,29805	-0,00176	-0,50972	-0,00418

FONTE: O autor (2022)

A TABELA 27 mostra a relação de nove autovalores, com os seus respectivos autovetores. Os autovalores estão classificados em ordem decrescente e apresentam também o seu percentual de contribuição para explicar a variância. Por exemplo, na primeira linha é mostrado o autovalor de maior contribuição (44,78%) cujo valor é 4,030302, e ao lado são mostrados os nove valores que compõem o seu respectivo autovetor associado.

Com esta relação, a partir do percentual de contribuição, selecionam-se as três primeiras linhas para montagem de três Componentes Principais, pois a soma

destes percentuais (44,78% + 26,47% + 12,68%) equivale a 83,93%, atendendo assim ao critério de selecionar os componentes principais cuja soma acumulada de contribuição atingisse 80%.

#### Passo 4 – Cálculo de cada Componente Principal

Aplica-se a equação (21) com  $j$  (quantidade de componentes principais) igual a três e  $i$  (quantidade de atributos) igual a nove ( $j = 3$ ;  $i = 9$ ).

Para o cálculo do PC1 do primeiro registro ( $z = 1$ ), tem-se os dados exibidos na TABELA 28

TABELA 28 - DADOS PARA CALCULAR O PC1

	1	2	3	4	5	6	7	8	9
autoVetor	-0,377812	0,336654	-0,39735	0,259828	0,520597	0,273281	-0,299554	0,260357	-0,123702
Dado padronizado	0,116988	-0,591696	-0,043613	2,336612	1,410241	2,402146	0,467634	2,287528	-0,777329
autoVetor * ( (Dado padronizado - Media) / DesvioPadrao)	-0,0441995	-0,1991968	0,0173296	0,6071172	0,7341672	0,6564609	-0,1400816	0,5955739	0,0961572

FONTE: O autor (2022)

$$\begin{aligned}
 PC_{1,1} &= 0,316158 \cdot (0,116988) + \\
 &-0,307886 \cdot (-0,591696) + \\
 &0,231678 \cdot (-0,043613) + \\
 &0,438749 \cdot (2,336612) + \\
 &0,021662 \cdot (1,410241) + \\
 &0,428645 \cdot (2,402146) + \\
 &0,274072 \cdot (0,467634) + \\
 &0,439766 \cdot (2,287528) + \\
 &-0,326128 \cdot (-0,777329) \\
 \mathbf{PC_{1,1}} &= \mathbf{3,682111}
 \end{aligned}$$

Para o cálculo do PC2 do primeiro registro ( $z = 1$ ), tem-se os dados exibidos na TABELA 29.

TABELA 29 - DADOS PARA CALCULAR O PC2

	1	2	3	4	5	6	7	8	9
autoVetor	-0,377812	0,336654	-0,39735	0,259828	0,520597	0,273281	-0,299554	0,260357	-0,123702
Dado padronizado	0,116988	-0,591696	-0,043613	2,336612	1,410241	2,402146	0,467634	2,287528	-0,777329
autoVetor * ( (Dado padronizado - Media) / DesvioPadrao)	-0,0441995	-0,1991968	0,0173296	0,6071172	0,7341672	0,6564609	-0,1400816	0,5955739	0,0961572

FONTE: O autor (2022)

$$\begin{aligned}
 PC_{1,2} = & -0,377812 \cdot (0,116988) + \\
 & 0,336654 \cdot (-0,591696) + \\
 & -0,397350 \cdot (-0,043613) + \\
 & 0,259828 \cdot (2,336612) + \\
 & 0,520597 \cdot (1,410241) + \\
 & 0,273281 \cdot (2,402146) + \\
 & -0,299554 \cdot (0,467634) + \\
 & 0,260357 \cdot (2,287528) + \\
 & -0,123702 \cdot (-0,777329)
 \end{aligned}$$

$$PC_{1,2} = 2,323328$$

Para o cálculo do PC3 do primeiro registro ( $z = 1$ ), tem-se os dados exibidos na TABELA 30.

TABELA 30 - DADOS PARA CALCULAR O PC3

	1	2	3	4	5	6	7	8	9
autoVetor	-0,324972	0,344209	0,461743	0,222392	-0,385451	0,245169	-0,088333	0,215566	0,500296
Dado padronizado	0,116988	-0,591696	-0,043613	2,336612	1,410241	2,402146	0,467634	2,287528	-0,777329
autoVetor * ( (Dado padronizado - Media) / DesvioPadrao)	-0,0380178	-0,2036671	-0,020138	0,5196438	-0,5435788	0,5889317	-0,0413075	0,4931133	-0,3888946

FONTE: O autor (2022)

$$\begin{aligned}
 PC_{1,3} = & -0,324972 \cdot (0,116988) + \\
 & 0,344209 \cdot (-0,591696) + \\
 & 0,461743 \cdot (-0,043613) + \\
 & 0,222392 \cdot (2,336612) + \\
 & -0,385451 \cdot (1,410241) + \\
 & 0,245169 \cdot (2,402146) + \\
 & -0,088333 \cdot (0,467634) + \\
 & 0,215566 \cdot (2,287528) + \\
 & 0,500296 \cdot (-0,777329)
 \end{aligned}$$

$$PC_{1,3} = 0,366087$$

Esta sequência de cálculos deve ser executada para cada um dos 20 registros do arquivo.

### Passo 5 – Matriz Final Resultante

Após efetuar os cálculos para os 20 registros, obtém-se o novo conjunto de dados com três variáveis (PC1; PC2 e PC3), conforme observa-se na TABELA 31.

TABELA 31 - MATRIZ RESULTANTE

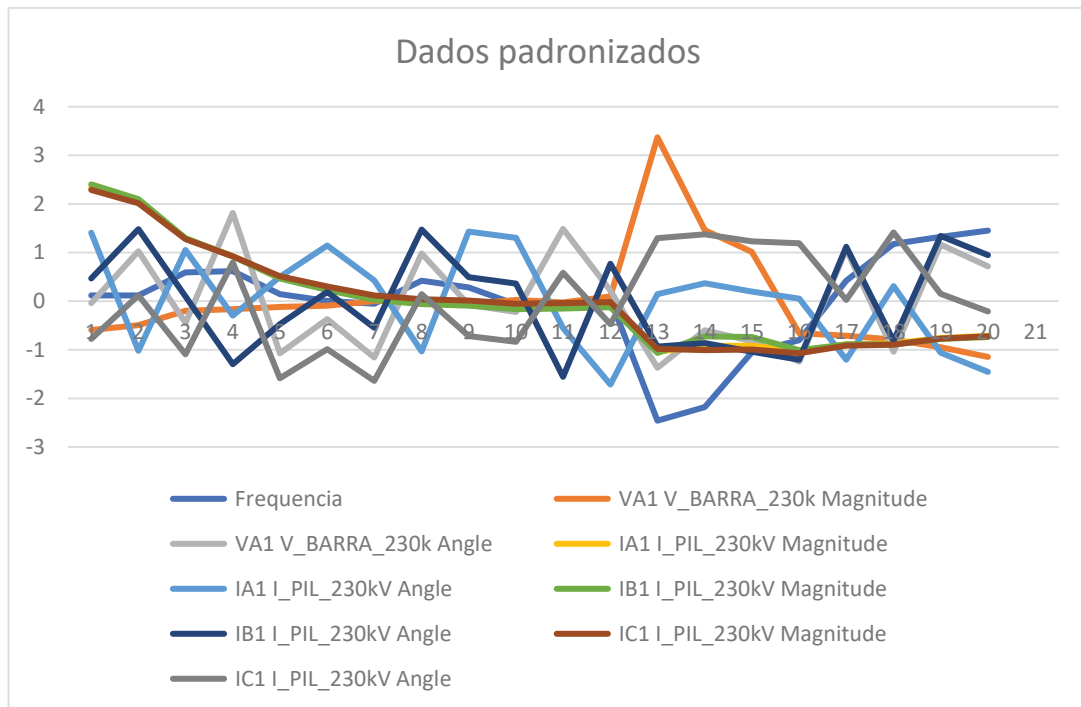
PC1	PC2	PC3
3,682111	2,323329	0,366087
3,454939	0,021452	1,99406
2,231023	1,560779	-0,54985
1,249691	-0,14058	1,838225
0,871655	1,317741	-1,193032
0,695621	0,985677	-0,97581
0,205907	1,13006	-1,401935
0,671575	-1,541347	0,661477
0,44286	0,555464	-1,129284
0,179541	0,710412	-1,09722
-0,433906	-0,538718	1,282214
0,268183	-1,188669	0,425227
-4,151642	1,988161	1,295584
-3,11267	1,131035	0,95619
-2,673701	0,629304	0,344268
-2,414756	-0,007441	-0,567072
-0,304213	-2,501874	-0,126091
-1,462039	-0,766593	-1,082821
0,269791	-2,861598	-0,37872
0,330028	-2,806593	-0,661499

FONTE: O autor (2022)

## Passo 6 – Visualização Gráfica

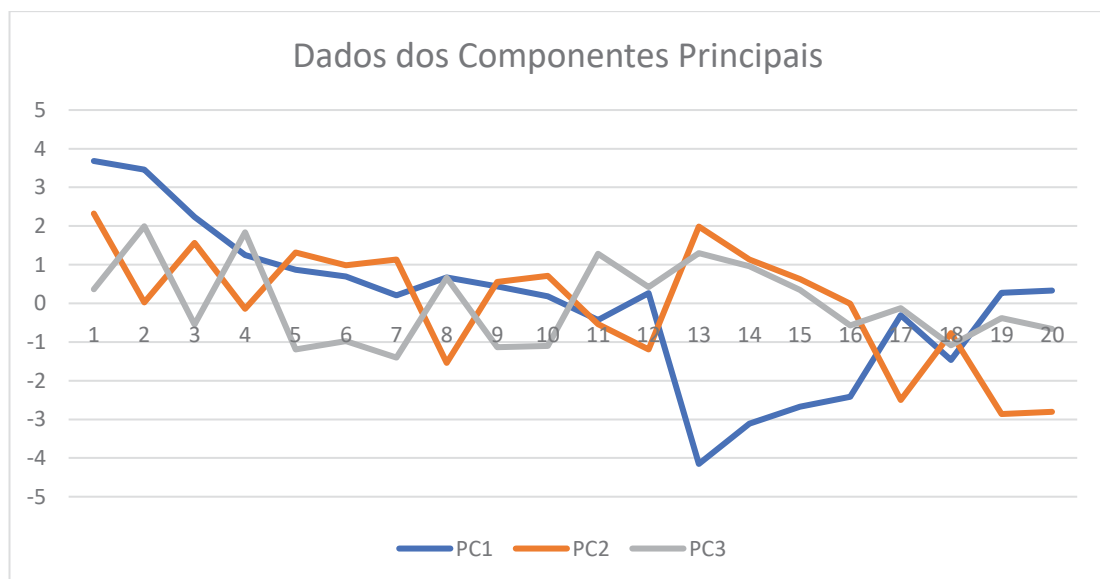
Ao final do processo de cálculo, através dos gráficos da FIGURA 14 e da FIGURA 15 é possível observar a similaridade dos valores obtidos com os valores originais, apesar da redução na quantidade de variáveis observadas.

FIGURA 14 - DADOS PADRONIZADOS



FONTE: O autor (2022)

FIGURA 15 - DADOS DOS COMPONENTES PRINCIPAIS



FONTE: O autor (2022)