

UNIVERSIDADE FEDERAL DO PARÁ INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

Matheus Seribeli Furmigare

Processo de aquisição de conhecimento por meio de técnicas estatísticas: Um estudo de caso sobre descargas parciais em transformadores de alta potência.

Orientador: Prof. Dra. Terezinha Ferreira de Oliveira

Belém 2013

Matheus Seribeli Furmigare

Processo de aquisição de conhecimento através de técnicas estatísticas: Um estudo de caso sobre descargas parciais em transformador de alta potência.

> Dissertação apresentada ao Programa de Pós-Graduação em Matemática e Estatística, da Universidade Federal do Pará, como requisito parcial para a obtenção do grau de Mestre em Estatística.

Orientadora: Prof. Dra. Terezinha Ferreira de Oliveira

Belém 2013

Matheus Seribeli Furmigare

Processo de aquisição de conhecimento através de técnicas estatísticas: Um estudo de caso sobre descargas parciais em transformador de alta potência.

Esta Dissertação foi julgada e aprovada, para a obtenção do grau de Mestre em Estatística, no Programa de Pós-Graduação em Matemática e Estatística, da Universidade Federal do Pará.

Belém, 04 Março de 2013.

Prof. Dra. Rúbia Gonçalves Nascimento Coordenadora do Programa de Pós-Graduação em Matemática e Estatística da UFPA

Banca Examinadora

Prof. Dra. Terezinha Ferreira de Oliveira Universidade Federal do Pará **Orientador** Prof. Dr. Joaquim Carlos Barbosa Queiroz Universidade Federal do Pará **Membro**

Prof. Dr. Valcir João da Cunha Farias Universidade Federal do Pará **Membro** Prof. Dr. Antônio Morais da Silveira Universidade Federal do Pará **Membro**

Dedico esta dissertação à Deus e às pessoas mais importantes na minha vida, meus pais, pelo que me ensinaram e transmitiram, minha esposa e meus filhos, pela força, carinho, incentivo, paciência e confiança.

Agradecimentos

* À Deus por ter me dado saúde, força, coragem e sempre iluminando meu caminho;

* À professora Terezinha Ferreira de Oliveira pela orientação, apoio, incentivo, mas principalmente, compreensão e paciência ao longo do desenvolvimento desta dissertação;

 \star Ao professor Joaquim Carlos Barbosa Queiroz pelo apoi
o e disponibilidade para ajudar na realização desta dissertação;

* Aos professores do Programa de Pós-graduação em Matemática e Estatística pelo conhecimento transmitido durante meu curso de mestrado;

 \star Aos meus pais e irmãos por estarem sempre ao meu lado, incentivando e dando força para chegar até aqui;

* À minha amada esposa pela compreensão e apoio em todos os momentos que passamos juntos;

* Aos participantes da banca examinadora pelas sugestões;

* E, especialmente, aos meus filhos Geovanna e Henrique pelo amor e carinho a mim dedicado.

Resumo

FURMIGARE, Matheus Seribeli Furmigare. Processo de aquisição de conhecimento através de técnicas estatísticas: Um estudo de caso sobre descargas parciais em transformador de alta potência. 2013. Dissertação (Mestrado em Matemática e Estatística), PPGME, UFPA, Belém, Pará, Brasil.

Equipamentos elétricos de alta tensão, especialmente os transformadores de potência, devem ser monitorados permanentemente, de modo a detectar o início de possíveis problemas, evitando que eles se desenvolvam e causem maiores prejuízos. Neste contexto, métodos não destrutivos de diagnósticos, como a análise de descargas parciais, vêm se mostrando adequados em termos de eficácia no monitoramento de equipamentos de alta potência. A detecção e o reconhecimento de Descargas Parciais tornaram-se ferramentas importantes para a avaliação da degradação isolante em dispositivos de alta potência, uma vez que a maior parte das falhas em transformadores é causada por rupturas do meio isolante. Descargas Parciais são fenômenos ocasionados pelo rompimento localizado da rigidez dielétrica do material isolante, caracterizando-se como uma das fontes de defeitos nos isolamentos elétricos.

Este trabalho propõe analisar possíveis influências e relações entre a ocorrência de descargas parciais e variáveis do sistema de energia elétrica, bem como, variáveis de manobras do sistema, utilizando técnicas estatísticas. Foram utilizadas técnicas estatísticas de análise multivariada para avaliar o comportamento das variáveis em estudo, reduzir o número de variáveis e eliminar a multicolinearidade, através da análise fatorial. Foram utilizados os scores dos fatores na análise multivariada de séries temporais, na correlação cruzada para avaliar o comportamento de cada componente e na função de transferência para ajuste do modelo para previsão de descargas parciais. Técnicas estatísticas de análise de contingência e análise de resíduo foram utilizadas para avaliar a relação entre as DP's e as variáveis de manobras do sistema.

Palavras-chave: Descargas Parciais, Análise fatorial, Séries temporais, Modelo de função de transferência.

Abstract

FURMIGARE, Matheus Seribeli. Bayesian estimation in Item Response dichotomous Models on framework Mixed Models. 2012. Dissertation (Master Science in Mathematics and Statistics), PPGME, UFPA, Belém, Pará, Brazil.

High voltage electrical equipment, especially power transformers, should be monitored continuously in order to detect possible problems early, thus preventing them from developing and causing further damage. In this context, non-destructive methods of diagnosis such as partial discharge analysis have shown to be suitable in terms of effectiveness in monitoring high power equipment. The detection and recognition of Partial Discharges have become important tools for assessing the degradation in insulating high-power devices, since most faults in transformers is caused by rupture of the insulating medium. Partial Discharge phenomena are caused by the rupture of localized dielectric strength of insulating material, which is characterized as one of the sources of faults in electrical insulation.

This study aims to analyze possible influences and relationships between the occurrence of partial discharges and system variables of electricity, as well as variables maneuvering system, using statistical techniques. Statistical techniques were used multivariate analysis to evaluate the behavior of the variables under study, reduce the number of variables and eliminate multicollinearity through factor analysis. We used the scores of the factors in the multivariate time series, the cross-correlation to evaluate the behavior of each component and the transfer function for adjusting the model to predict the partial discharges. Statistical techniques of contingency analysis and residual analysis were used to evaluate the relationship between the variables and DP's maneuvering system.

Keywords: Partial Descharge, factor analysis, time series, model transfer function.

Índice

R	\mathbf{esun}	no	\mathbf{vi}
A	\mathbf{bstr}	act	vii
\mathbf{Li}	sta	de Tabelas	x
\mathbf{Li}	sta	de Figuras	xi
1	\mathbf{Int}	rodução	1
	1.1	Aspectos Gerais	1
	1.2	Objetivos	4
		1.2.1 Objetivo geral	4
		1.2.2 Objetivos específicos	4
	1.3	Estrutura da Dissertação	4
2	Me	etodologia	6
	2.1	Introdução	6
	2.2	Análise Fatorial	6
	2.3	Séries temporais	8
		2.3.1 Função de autocorrelação	9
		2.3.2 Função de autocorrelação parcial	10
		2.3.3 Modelos ARIMA(p,d,q) de Box e Jenkins	10
		2.3.4 Análise das funções de autocorrelação	12
		2.3.5 Análise das funções de autocorrelação parcial	12
		2.3.6 Critério de informação de Akaike - AIC	12
		2.3.7 Critério de Informação Bayesiano - BIC	13
		2.3.8 Medidas de erro de previsão	13
	2.4	Função de Transferência para séries temporais multivariadas	14
		2.4.1 Modelo de Função de Transferência	14
		2.4.2 Função de Correlação Cruzada - FCC	15
		2.4.3 Relação entre função de correlação cruzada e função de transferência	16
		2.4.4 Função de correlação cruzada amostral	17
		2.4.5 Identificação do modelo de função de transferência	17
		2.4.6 Identificação do modelo ruído	18
	2.5	K-means	18
	2.6	Análise de contingência	19

3	\mathbf{An}	álise dos dados	21			
	3.1	Localização, sistema de medição e análises descritiva dos dados	21			
	3.2	Análise fatorial	26			
	3.3 Modelagem multivariada com função de transferência $\ldots \ldots \ldots \ldots \ldots \ldots$					
		3.3.1 Ajuste dos modelos para as séries de entrada	30			
		3.3.2 Ajuste do modelo para a série de saída	35			
	3.4	Correlações cruzadas	38			
		3.4.1 Análise da correlação cruzada entre $YvsX_1$	38			
		3.4.2 Análise da correlação cruzada entre $YvsX_2$	38			
		3.4.3 Análise da correlação cruzada entre $YvsX_3$	39			
		3.4.4 Análise da correlação cruzada entre $YvsX_4$	40			
	3.5	Ajuste dos modelos de função de transferência	40			
		3.5.1 Estruturação para a estimativa do modelo	41			
		3.5.2 Análise residual para o Modelo FT	41			
	3.6	Análise da relação dos dados de DP e as manobras no sistema	44			
4	Co	nsiderações Finais e Recomendações	48			
	4.1	Recomendações	49			
\mathbf{A}	An	exo A	50			
	A.1	Tabelas referentes à análise fatorial e análises de séries temporais	50			

Lista de Tabelas

2.1	Forma geral de uma tabela de contingência de duas dimensões	9
3.1	Estatísticas básicas dos dados de descargas parciais.	4
3.2	Cargas fatoriais dos componentes rotacionadas e variância total explicada 2'	7
3.3	Critérios de ajuste e previsão para modelo 3.5	7
3.4	Critérios de ajuste e previsão dos modelos Y e FT	2
3.5	Classificação das ocorrências de manobras	5
3.6	Manobras selecionadas	5
3.7	Manobras selecionadas.	6
3.8	Resultados da análise de contingência.	6
3.9	Análise dos resíduos.	6
A.1	Principais estatísticas das 18 variáveis do sistema.	1
A.2	Correlações entre as variáveis do sistema.	$\overline{3}$
A.3	Teste KMO e de esfericidade de <i>Bartlett</i>	4
A.4	Comunalidades extraída para três fatores com rotação Varimax	4
A.5	Cálculo e Avaliação dos parâmetros da variável X_1	4
A.6	Estudo das autocorrelações dos resíduos para o modelo da variável X_1	5
A.7	Cálculo e Avaliação dos parâmetros da variável X_2	5
A.8	Estudo das autocorrelações dos resíduos para o modelo da variável X_2	5
A.9	Cálculo e Avaliação dos parâmetros da variável X_3	5
A.10	Estudo das autocorrelações dos resíduos para o modelo da variável X_3	6
A.11	Cálculo e Avaliação dos parâmetros da variável X_4	6
A.12	Estudo das autocorrelações dos resíduos para o modelo da variável X_4	6
A.13	Cálculo e Avaliação dos parâmetros da variável Y	6
A.14	Estudo das autocorrelações dos resíduos para o modelo da variável Y	7
A.15	Cálculo dos parâmetros estimados para o modelo de FT	7
A.16	Estudo das autocorrelações dos resíduos do modelo de FT	7
A.17	Estudo das correlações cruzadas dos resíduos $v_s X_1$	8
A.18	Estudo das correlações cruzadas dos resíduos $v_s X_2$	8
A.19	Estudo das correlações cruzadas dos resíduos $v_s X_3$	8

Lista de Figuras

2.1	Construção do modelo de séries temporais (adaptado de Moretin e Toloi, 2004) Sistema dinâmico de função de transferência (adaptado de Wei 1990)	9 15				
3.1	Localização geográfica de Vila do Conde Barcarena - PA	-10 -21				
3.2	Banco de autotransformadores: cada autotransformador representa uma fase	21				
3.3	Mapa estatístico de DP: identificando a quantidade de DP a localização em					
0.0	relação ao ângulo de fase da tensão e a intensidade em pico Coulombs	23				
34	Gráfico de Descargas Parciais - DP - fase 230kV	23				
3.5	Gráficos de Descargas Parciais - DP - fase 500kV.	24				
3.6	Gráficos Box-Plot com linha de referência horizontal em 500 pC: (a) Variáveis DP					
0.0	- fase 230kV e DP - fase 500kV: (b) Variável DP - fase 230 kV a cada hora e (c)					
	Variável DP - fase 500 kV a cada hora.	25				
3.7	Gráfico das Descargas Parciais - DP - fase 230kV (Y).	29				
3.8	Gráfico das Descargas Parciais - DP - fase 500kV (X_1) .	29				
3.9	Gráfico dos Score gerado pela AF para o fator 1 - F1 (X_2)	29				
3.10	Gráfico dos Score gerado pela AF para o fator 2 - F2 (X_3)	30				
3.11	Gráfico dos Score gerado pela AF para o fator 3 - F3 (X_4)	30				
3.12	Função de Autocorrelação da variável (X_1)	31				
3.13	Função de Autocorrelação Parcial da variável (X_1)	31				
3.14	Função de Autocorrelação da variável (X_2)	32				
3.15	Função de Autocorrelação Parcial da variável (X_2)	32				
3.16	Função de Autocorrelação da variável (X_3)	33				
3.17	Função de Autocorrelação Parcial da variável (X_3)	33				
3.18	Função de Autocorrelação da variável (X_4)	34				
3.19	Função de Autocorrelação Parcial da variável (X_4)	34				
3.20	Função de Autocorrelação da variável (Y)	35				
3.21	Função de Autocorrelação Parcial da variável (Y)	35				
3.22	Ajuste do Modelo (Y)	36				
3.23	Função de Autocorrelação Parcial da variável (Y)	37				
3.24	Função de Correlação Cruzada $YvsX_1$	38				
3.25	Função de Correlação Cruzada $YvsX_2$	39				
3.26	Função de Correlação Cruzada $YvsX_3$	39				
3.27	Função de Correlação Cruzada $YvsX_4$	40				
3.28	Modelo F'T ajustado para variável Y	42				
3.29	Previsões obtidas pelo modelo F'I'	43				
3.30	Esquema do sistema de transformação e distribuição	44				

Capítulo 1

Introdução

1.1 Aspectos Gerais

O sistema de energia elétrica é, basicamente, distinguido em três etapas: geração, transmissão e distribuição. Na etapa de transmissão e distribuição a energia elétrica passa por diversas transformações até chegar ao consumidor final. O equipamento que realiza estas transformações são os transformadores ou autotransformadores de potência. Qualquer equipamento, ou arranjo, submetido a um determinado valor de tensão, automaticamente ocorrerá tensões induzidas em objetos próximos, ionização das moléculas de ar nas superfícies condutoras, geram-se sinais eletromagnéticos e forma-se uma distribuição de campo elétrico no espaço em volta do arranjo, diretamente relacionado com a tensão aplicada.

Conforme a intensidade e o mapeamento desse campo elétrico é que surgem os diversos fenômenos, os quais, devido às suas influências e intensidade, podem originar vários problemas nos sistemas de transmissão, tais como: Perdas de Energia; Interferências nas Freqüências Auditivas (Rádio e TV); Corona Visual; Descargas Parciais (DP); e até mesmo Descargas Disruptivas (quando o campo elétrico é suficientemente alto para promover a ionização de todo um percurso até outro elemento na proximidade).

Todos os fenômenos acima são devidos à "ionização" das moléculas de ar em regiões onde o campo elétrico torna-se crítico. A "ionização" apenas representa a separação dos componentes de uma molécula neutra. O termo "efeito corona" é utilizado no caso de descargas ocorrendo no ar em torno de um elemento condutor. Inicialmente, pode-se dizer que o termo "descargas parciais" é utilizado tanto para o caso de descargas nas cavidades ou bolhas de um material isolante quanto para o caso das descargas em superfícies condutoras. As descargas nas cavidades ou bolhas de materiais isolantes podem gerar ondas eletromagnéticas que interfiram nas recepções de rádio, mas o principal inconveniente das descargas nessas cavidades é a deterioração do isolante.

A norma IEC 60270 define descargas parciais como descargas elétricas localizadas na união entre dois condutores, através do isolamento, que pode ou não ocorrer próximo de um condutor. São fenômenos ocasionados pelo rompimento localizado da rigidez dielétrica do material isolante, caracterizando-se como uma das fontes de defeitos nos isolamentos elétricos. Ou seja, são descargas elétricas que não chegam a percorrer todo o caminho dentro de um material isolante colocado entre dois condutores submetidos a uma diferença de potencial. Em dielétricos sólidos essas descargas são produzidas pela ionização de pequenas cavidades de ar no interior do dielétrico; nos líquidos, pela ionização de bolhas de gás no seu interior; e no ar, pela ionização das moléculas de ar que se encontram nos pontos de maior gradiente de potencial. Podem ocorrer em qualquer ponto do dielétrico, na junção de dois dielétricos diferentes ou adjacentes ao condutor, podem também ocorrer seguidamente em vários pontos do dielétrico.

Neste contexto, os transformadores e autotransformadores de potência são equipamentos com todas as características anteriormente apresentadas e, portanto apresentam grande propensão à ocorrência de descargas parciais. Devido à importância, ao alto custo, ao transtorno e às grandes perdas financeiras em caso de falha nesses equipamentos, faz-se necessária o acompanhamento permanente, principalmente da ocorrência e evolução de DP's. Pois, estas são uma fonte contínua de deterioração do material isolante, ou seja, modificam suas propriedades dielétricas. Grande parte dos defeitos nestes equipamentos são devidos às rupturas do meio isolante. Assim, a monitoração permanente pode detectar o início de possíveis problemas, e evitar que eles se desenvolvam e causem prejuízos maiores.

Métodos de medição de DP's são descritos nas Normas IEC-270 e são geralmente adotados pela comunidade técnica. No entanto, para a maioria dos sistemas disponíveis no mercado, a avaliação do estado de isolamento depende muito da decisão de peritos humanos. Além disso, sabe-se que equipamentos elétricos apresentam particularidades quanto ao seu funcionamento, é cauteloso considerar que a análise e o diagnóstico devem utilizar conhecimentos e regras diferentes para cada tipo de equipamento. Daí a importância de fazer uma análise da evolução das descargas em um mesmo grupo de dados e utilizar com cautela as informações históricas obtidas de bancos de dados.

Entretanto, diversos pesquisadores têm estudado técnicas e desenvolvido ferramentas de análise para identificar padrões em dados de DPs. Diferentes técnicas foram desenvolvidas para a escolha das características que vão fornecer a informação relevante para o reconhecimento de padrões em DPs. Algumas delas podem ser agrupadas em: métodos estatísticos; análise do formato dos pulsos; processamento de sinais; processamento de imagens e aproximações de série temporal. Os métodos estatísticos têm sido amplamente utilizados na extração de características das DP's,. Cuenca (1998) e RAHMAN et. al. (2000) utilizaram análise de componentes principais para avaliar o comportamento e distinguir a localização da DP. E, Diniz (2005) e Cuenca (2005) análises de componentes independentes (ICA), estes trabalhos apresentaram bons resultados em relação a identificação de padrões em DP's, por exemplo, em Cuenca (2005), utilizou sete componentes principais, alcançando uma eficiência total de 93

James e Pung (1995) utilizaram filtros casados e momentos estatísticos. Mas, também, técnicas de correlação e regressão simples foi empregada por Jeyabalan (2011). E, técnicas de regressão múltipla por James e Jones (1988).

Técnicas multivariadas, como análise de componentes principais, análise fatorial e análise de cluster (R.-j. LIAO et al., 2011), que propôs avaliar o estado de envelhecimento do meio isolante do transformador com base na análise de descargas parciais, através da análise fatorial, avaliou os fatores característicos extraídos para identificar as amostras de óleo de papel com diferentes graus de envelhecimento. Os padrões foram analisados por meio de análise de componentes principais e fator (PCFA), os fatores extraídos representaram informações sobre os padrões de DP e, através da utilização das novas características extraídas pelo método PCFA, os resultados das amostras pôde discriminar diferentes estágios de envelhecimento, fornecendo informação significativa referenciada na avaliação do estado do meio isolante.

Em Barbieri (2012), técnicas de séries temporais não-lineares são aplicadas para distinguir o tipo de descargas parciais e o nível de degradação do material envolvido. A aproximação por séries temporais foi realizada para determinar o número de características de entrada a ser usado por uma rede neural para a classificação.

Em Salama e Bartinikas (2002) e Mazroua et. al. (1994), modelos de séries temporais autoregressivos foram utilizados para modelar o efeito de histerese da característica das DPs e redes neurais para reconhecimentos de padrões de DP's e padrões de pulsos de DP's.

Este trabalho propõe analisar possíveis influências e relações entre a ocorrência de descargas parciais e variáveis do sistema de energia elétrica, bem como, variáveis de manobras do sistema, utilizando técnicas estatísticas. Foram utilizadas técnicas estatísticas de análise multivariada para avaliar o comportamento das variáveis em estudo, reduzir o número de variáveis e eliminar a multicolinearidade, através da análise de componentes principais e análise fatorial. Foram utilizados os scores dos fatores na análise multivariada de séries temporais, na correlação cruzada para avaliar o comportamento de cada componente e na função de transferência para ajuste do modelo para previsão de descargas parciais. Técnicas estatísticas de análise de contingência e análise de resíduo foram utilizadas para avaliar a relação entre as DP's e as variáveis de manobras do sistema.

1.2 Objetivos

1.2.1 Objetivo geral

• O objetivo geral deste trabalho é aplicar técnicas estatísticas para a aquisição de conhecimento sobre descargas parciais em transformadores de alta potência.

1.2.2 Objetivos específicos

- Aplicar métodos estatísticos multivariados para redução das variáveis envolvidas no sistema de transformação de tensão do sistema de transmissão e distribuição de energia elétrica.
- Analisar através de técnicas de análises de séries temporais o comportamento temporal da série de descargas parciais.
- Propor um modelo dinâmico linear com base na metodologia de Box e Jenkins para séries temporais multivariadas, utilizando função de transferência, avaliando a capacidade de monitoramento do sistema e a contribuição dos fatores gerados pela análise fatorial no ajuste deste modelo.
- Aplicar técnicas não paramétricas para investigar a influências das manobras dos equipamentos da subestação na ocorrência de descargas parciais.

1.3 Estrutura da Dissertação

Esta dissertação encontra-se dividida em 4 capítulos, a saber:

Capítulo 1: Refere-se à introdução do trabalho, onde são abordados os aspectos gerais, a importância, o objetivo geral, os objetivos específicos e a estrutura do trabalho.

Capítulo 2: Apresenta a metodologia seguida, apresentando uma revisão teórica sobre análise fatorial, análise de séries temporais univariadas, análise de séries temporais multivariadas, k-means e análise de contingência.

Capítulo 3: Apresenta a análise dos dados e os resultados obtidos.

Capítulo 4: Considerações finais e recomendações para trabalhos futuros.

Capítulo 2

Metodologia

2.1 Introdução

Neste trabalho utilizou-se a Análise Fatorial (AF), uma ferramenta estatística multivariada para redução da quantidade de variáveis envolvidas no problema e eliminação da multicolinearidade. Técnicas de análise de Séries Temporais (ST) foram utilizadas para avaliar o comportamento da série de Descargas Parciais (DP's) e pré-filtragem para análise da Correlação Cruzada entre as séries de DP's e os fatores extraídos pela AF. Propõe-se um modelo de Séries Temporais Multivariadas, utilizando Função de Transferência (FT) para avaliar a possibilidade de monitoramento das descargas parciais. Técnicas de *K-Means* foram utilizadas para identificar um valor que separa os dados de DP e, por fim, realizam-se Análises de Contingências (AC) e Análise de Resíduos Qui-Quadrado para avaliar a influência das manobras do sistema na ocorrência de DP's.

2.2 Análise Fatorial

A AF é uma das técnicas de análise multivariada que permite realizar redução ou sumarização de dados. Esta técnica de interdependência considera todas as variáveis simultaneamente, assumindo que cada uma delas está relacionada com as demais, a fim de se estudar a inter-relação existente entre as mesmas. De acordo com Johnson e Wichern (1992) e Hair *et al* (1998) a AF é um método estatístico utilizado para descrever a estruturada variabilidade entre variáveis observadas correlacionadas em termos de um número potencialmente menor de variáveis não observadas chamados fatores.

O Coeficiente de Kaiser-Meyer-Olkin - KMO e Teste de esfericidade de Bartlett são medidas utilizadas para avaliar a viabilidade da análise fatorial. O KMO parte do princípio que as correlações parciais entre pares de variáveis devem ser pequenas se o modelo for adequado. Valores do teste abaixo de 0,50 não são adequados (FAVERO, 2009). O Teste de esfericidade de Bartlett testa a hipótese de que a matriz de correlação é uma matriz identidade, ou seja, que não há correlação entre as variáveis. O método utilizado na AF para a extração dos fatores foi o de Componentes Principais, com rotação ortogonal dos eixos via Varimax. Esta análise não só reduz a dimensionalidade de Xcomo também permite a interpretabilidade das novas variáveis e é baseada em um modelo em que as variáveis são expressas como uma função linear de um número reduzido de fatores latentes, conforme modelo (2.1) a seguir:

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon} \tag{2.1}$$

onde

 ${\bf X}$ é a matriz das variáveis

 μ é o centróide

 ${\bf L}$ é a matriz de loading dos fatores

 ${\bf F}$ é o vetor de fatores

 ε é o vetor de resíduos do modelo

Para este modelo assume-se que $E(\mathbf{F}) = 0$, $Cov(\mathbf{F})$, $\mathbf{E}(\varepsilon) = 0$ e $Cov(\varepsilon) = \Psi$, onde Ψ é uma matriz diagonal, $\mathbf{F} \in \varepsilon$ são independentes.

A variância de cada variável X_i pode ser dividida em duas parcelas, aquelas devido a fatores comuns e aquelas devido a fatores individuais, de acordo com o modelo (2.2) que se segue,

$$Var(\mathbf{X}_{i}) = s_{ii} = \lambda_{i1}^{2} + \lambda_{i2}^{2} + \dots + \lambda_{iq}^{2} + \Psi$$
(2.2)

A variância devido aos fatores comuns é chamada de comunalidade. A comunalidade total é dada pela equação (2.3):

$$H = h_1^2 + h_2^2 + \dots + h_p^2 = \sum_{i=1}^p \sum_{j=1}^q \lambda_{ij}^2$$
(2.3)

Assim, a matriz de covariância de \boldsymbol{X} , (2.4) será a soma da matriz cheia de $\boldsymbol{L}L^T$ e da matriz diagonal de Ψ , isto é:

$$Cov(\boldsymbol{X}) = \Sigma = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}$$
(2.4)

onde

 L_{pxm} é a matriz de carga dos fatores, constituída dos coeficientes l_{ij} chamados de carga da i-ésima variável no j-ésimo fator;

 Ψ_{pxp} é a matriz de covariância $Cov(\varepsilon) = E[\varepsilon \varepsilon^T]$, definida em (2.5),

$$\Psi = \begin{bmatrix} \Psi_1 & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \Psi_p \end{bmatrix}$$
(2.5)

Logo, no modelo de Análise Fatorial, os fatores comuns são responsáveis pela estrutura de covariância, enquanto os valores individuais só influenciam a diagonal da matriz de covariância de X. Os fatores comuns são responsáveis pela estrutura de dependência entre as variáveis.

Um recurso para melhorar a interpretabilidade L, é realizar uma rotação conjunta dos eixos. Utilizou-se o método de rotação ortogonal Varimax, o qual gera uma matriz G dos fatores que maximizam a variância do quadrado das colunas de $L^T = GL$.

Pode-se provar que a estimativa de máxima verossimilhança das comunalidades (2.6) é:

$$\hat{h}_{i}^{2} = \tilde{I}_{i1}^{2} + \tilde{I}_{i2}^{2} + \ldots + \tilde{I}_{im}^{2}$$
(2.6)

com a proporção da variância amostral total para o j-ésimo fator dada por (2.7):

$$\frac{\tilde{I}_{i1}^2 + \tilde{I}_{i2}^2 + \ldots + \tilde{I}_{im}^2}{s_{11} + s_{22} + \ldots + s_{pp}}$$
(2.7)

a proporção padronizada da variância amostral total para o j-ésimo fator dada por (2.8)

$$\frac{\tilde{I}_{i1}^2 + \tilde{I}_{i2}^2 + \ldots + \tilde{I}_{im}^2}{pp}$$
(2.8)

O algoritmo básico do Varimax define (2.9) para j = 1, ..., m e i = 1, ..., p:

$$\widetilde{I}_{ij}^* = \frac{\widehat{I}_{ij}^*}{\widehat{h}_i} \tag{2.9}$$

para ser os coeficientes finais de rotação escalados pela raiz quadrada da comunalidade. O procedimento utiliza uma transformação ortogonal que faz com que (2.10) seja a maior possível.

$$V = \frac{1}{p} \sum_{j=1}^{m} \left[\sum_{i=1}^{p} (\widetilde{I}_{ij}^{*})^{4} - \left(\sum_{i=1}^{p} (\widetilde{I}_{ij}^{*})^{2} \right)^{2} / p \right]$$
(2.10)

A AF se diferencia de outras técnicas da estatística multivariada pelo fato de, além de separar as variáveis em fatores não correlacionados (ou correlações não significativas), estes podem ser utilizados em outras técnicas estatísticas, como por exemplo, em análises de regressão (Lattin *et al*, 2011). Neste trabalho os fatores gerados pela AF foram utilizados na análise de séries temporais multivariadas através da técnica de função de transferência.

2.3 Séries temporais

Uma Série Temporal, $Z_t, t = 1, 2, 3, ...$, segundo Moretin e Toloi (2006), é qualquer conjunto de observações ordenadas no tempo. Os modelos utilizados para descrever séries temporais são

processos estocásticos, isto é, processos controlados por leis probabilísticas. Uma metodologia bastante utilizada na análise de modelos paramétricos é conhecida como abordagem de Box e Jenkins (1970). Tal metodologia consiste em ajustar modelos conhecidos como auto-regressivos integrados de médias móveis, ARIMA(p,d,q), a um conjunto de dados. A estratégia para a construção do modelo é baseada em ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos estágios do ciclo iterativo, conforme Figura 2.1.



Figura 2.1 Construção do modelo de séries temporais (adaptado de Moretin e Toloi, 2004).

Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação. A fase crítica do procedimento acima é a identificação. É possível que vários pesquisadores identifiquem modelos diferentes para a mesma série temporal.

As principais ferramentas utilizadas para a análise de séries temporais univariadas são: a Função de Autocorrelação (FAC) e a Função de Autocorrelação Parcial (FACP).

2.3.1 Função de autocorrelação

A Função de Autocorrelação (FAC) representa a correlação simples entre Z_t e $Z_{(t+k)}$ em função da defasagem k. A FAC de uma série temporal pode ser definida por (2.11),

$$\rho_k = \frac{Cov[Z_t, Z_{t+k}]}{\sqrt{Var(Z_t)Var(Z_{t+k})}}$$
(2.11)

O coeficiente de autocorrelação varia entre -1 e 1: Se ρ assume o valor 1, diz-se que as duas variáveis medidas possuem uma autocorrelação positiva absoluta; caso contrário, se $\rho = -1$, diz-se que a autocorrelação negativa é absoluta. Quando ρ assume valor 0 (zero) não existe autocorrelação entre as variáveis, ou seja, autocorrelação nula.

2.3.2 Função de autocorrelação parcial

A Função de Autocorrelação Parcial (FACP) representa a correlação entre $Z_t \in Z_{(t+k)}$ como uma função da defasagem k, filtrado o efeito de todas as outras defasagens sobre $Z_t \in Z_{(t+k)}$.

Box, Jenkins e Reinsel (2006) propõem que a FACP seja calculada como o valor do coeficiente ϕ_{kk} na equação (2.12),

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \ldots + \phi_{kk}\rho_{j-k}, j = 1, \ldots, k$$
(2.12)

e, em geral, ϕ_{kk} é dado por (2.13),

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|} \tag{2.13}$$

onde, P_k é a matriz de autocorrelações e P_k^* é a matriz P_k com a última coluna substituída pelo vetor de autocorrelações.

2.3.3 Modelos ARIMA(p,d,q) de Box e Jenkins

Estes modelos foram propostos por volta da década de 1960, pelos professores George E. P. Box e Gwilym M. Jenkins. Esta é uma metodologia muito utilizada pela comunidade científica para a análise de séries temporais, principalmente devido a alta capacidade de gerar previsões significativas (MORETTIN e TOLOI, 2004).

Segundo Box e Jenkins (1976), uma série temporal é um conjunto de observações de uma dada variável ordenada segundo o parâmetro tempo em intervalos iguais. É importante observar se a série é estacionária, pois uma série estacionária apresenta uma função amostral do processo com mesma forma em todos os instantes do tempo.

Intuitivamente, um processo Z é estacionário quando se desenvolve no tempo de modo que a escolha de uma origem dos tempos não é importante. Em outras palavras, as características de Z(t + k), para todo k, são as mesmas de Z(t).

Uma exigência para a construção de modelos ARIMA é que a série seja estacionária. Uma das formas mais utilizadas para tornar uma série estacionária é por meio da aplicação de diferenças na série, identificadas como, $\Delta Z_t = Z_t - Z_{t-1} = (1-B)Z_t$, onde $BZ_t = Z_{t-1}$, é conhecido como

operador atraso. Deste modo, para uma série Z_t não estacionária, pode-se torná-la estacionária e representá-la por modelos conhecidos como ARMA. Assim, segundo Moretin e Toloi (2004), se $W_t = \Delta^d Z_t$, onde d é a ordem da diferença aplicada, for estacionária, podemos representar W_t por um modelo ARMA(p, q), (2.14),

$$\phi(B)W_t = \theta(B)a_t \tag{2.14}$$

onde, $\phi(B) = 1 - \phi_1(B) - \dots \in \theta(B) = \theta_1(B) - \dots$

sendo:

- $\phi(B)$: Polinômio Auto-regressivo (AR);
- $\theta(B)$: Polinômio Médias Móveis (MA).

Se W_t for obtido pela diferenciação de Z_t , então Z_t é uma integral de W_t , daí dizermos que Z_t segue um modelo auto-regressivo, integrado, de médias móveis, ou modelo ARIMA, dado por (2.15),

$$\phi(B)\Delta^d Z_t = \theta(B)a_t \tag{2.15}$$

de ordem (p, d, q) e escrevemos ARIMA(p, d, q), se $p \in q$ são as ordens de $\phi(B) \in \theta(B)$, respectivamente.

Se $\phi(B) = 1$, têm-se os chamados Modelos Auto-regressivos Puros AR(p, 0) (MORETIN e TOLOI, 2004), onde as estimativas de Z para um tempo t dependem de uma combinação linear de p termos da série observada, incluindo o termo aleatório a_t de ruído branco (erros de estimação com distribuição normal de média zero e variância constante e não correlacionadas). Deste modo se tem (2.16),

$$Z_t = \sum_{i=1}^p \phi_i Z_{t-i} + a_t \tag{2.16}$$

Quando $\theta(B) = 1$, têm-se os chamados Modelos Média Móveis Puros (MA), este modelo assume que a série modelada é gerada através de uma combinação linear de q sinais de ruído a_{t-i} , aleatórios e independentes entre si. Assim Z_t é dado por (2.17),

$$Z_t = -\sum_{i=1}^p \theta_i a_{t-i} + a_t$$
 (2.17)

Os Modelos Auto-regressivos - Médias Móveis Puros ARMA, com AR(p,0) e MA(0,q) são aqueles cujo polinômio $\phi(B) = 1$ e $\theta(B) = 1$. Deste modo Z_t é dado por (2.18),

$$Z_t = \sum_{i=1}^p \phi_i Z_{t-i} - \sum_{i=1}^q \theta_i a_{t-i} + a_t$$
(2.18)

2.3.4 Análise das funções de autocorrelação

Os processos AR(p), MA(q) e ARMA(p,q) apresentam as funções de autocorrelações com as seguintes características:

- Um processo AR(p) apresenta FAC com decaimento de acordo com exponenciais e/ou senóides amortecidas, infinitas em extensão;
- Um processo MA(q) apresenta FAC finita, no sentido que ela apresenta um corte após o lag q;
- Um processo ARMA(p,q) apresenta FAC infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após o lag q p.

2.3.5 Análise das funções de autocorrelação parcial

Os processos AR (p), MA (q) e ARMA (p,q) apresentam as funções de autocorrelações parciais com as seguintes características:

- Um processo AR(p) apresenta FACP $\phi_{kk} \neq 0$, para $k \leq p \in \phi_{kk} = 0$, para k > p;
- Um processo MA(q) apresenta FACP de similar à FAC de um processo AR (p);
- Um processo ARMA(p,q) apresenta FACP com as características da FAC de um processo MA puro e/ou senóides amortecidas apões o lag q - p.

2.3.6 Critério de informação de Akaike - AIC

A partir de 1970 foram propostos vários procedimentos para identificação de modelos ARMA. Um deles foi proposto por Akaike (1973, 1974) apud Moretin e Toloi (2004). A idéia é escolher as ordens $k \in l$ que minimizem a quantidade (2.19),

$$P(k,l) = ln\hat{\sigma}_{k,l}^2 + (k+l)\frac{C(N)}{N}$$
(2.19)

em que $\hat{\sigma}_{k,l}^2$ é uma estimativa da variância residual obtida ajustando um modelo ARMA(k,l) às N observações da série e C(N) é uma função do tamanho da série.

E, então, Akaike propõe escolher o modelo cujas ordens $k \in l$ minimizam o critério (2.20),

$$AIC(k,d,l) = Nln\hat{\sigma}_{a}^{2} + \frac{N}{N-d}2(k+l+1+\delta_{d0}) + Nln2\pi + N$$
(2.20)

em que

$$\delta_{d0} = \begin{cases} 1, & \text{se} \quad d = 0, \\ 0, & \text{se} \quad d \neq 0, \end{cases}$$
(2.21)

e $\hat{\sigma}_a^2$ é o estimador de máxima veros similhança de $\hat{\sigma}_{k,l}^2$

2.3.7 Critério de Informação Bayesiano - BIC

Sugerem Akaike (1977), Rissanen (1978) e Schwars (1978) apud Moretin e Toloi (2004) minimizar o Critério de Informação Bayesiano, dado por (2.22),

$$BIC = -2ln(ML) + (numeros \quad de \quad parametros)lnN$$
(2.22)

No caso de um processo ARMA, essa quantidade é dada por (2.23),

$$BIC(k,l) = ln\hat{\sigma}_{k,l}^2 + (k+l)\frac{lnN}{N}$$
(2.23)

em que $\hat{\sigma}_{k,l}^2$ é a estimativa de máxima verossimilhança da variância residual do modelo ARMA(k,l).

2.3.8 Medidas de erro de previsão

Para se escolher um método de previsão ou outro é importante utilizar uma medida de erro com a finalidade de se encontrar aquele método que melhor atenda os resultados esperado (MORETIN e TOLOI, 2004). Neste trabalho foram utilizadas duas medidas do erro, raiz do erro quadrático médio (RMSE, sigla em inglês) e o erro percentual médio absoluto (MAPE, sigla em inglês), respectivamente, para avaliar nossas previsões. O cálculo das medidas utilizadas pode ser visualizado nas equações (2.24) e (2.25), RMSE e MAPE, respectivamente,

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (a_t - y_t)^2}$$
(2.24)

Na equação (2.24) n é o número de padrões, a_t representa o valor real no instante t, e y_t representa o valor previsto no instante t.

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{a_t - y_t}{a_t} \right|$$
(2.25)

Onde, n é o número de padrões, a_t representa o valor real no instante t, e y_t representa o valor previsto no instante t.

2.4 Função de Transferência para séries temporais multivariadas

Suponha-se que X mede o nível de uma entrada de processo dinâmico e que o nível de X influência no nível da saída de Y do sistema. Mesmo em caso que devido à inércia do sistema, uma mudança de X de um nível para o outro não tenha efeito imediato sobre Y, mas sim, irá produzir uma resposta retardada em Y, chegando ao equilíbrio a um novo nível. Refere-se a um processo de resposta dinâmica. Um modelo que descreve esta resposta dinâmica é chamado de modelo de função de transferência. Vamos supor que observações de entrada e saída são feitas a intervalos equidistantes de tempo, o modelo de função de transferência associada, então, será chamado de um modelo de função de transferência discreta (BOX E JENKINS, 2006).

2.4.1 Modelo de Função de Transferência

Considerar que X_t e Y_t são séries adequadamente transformadas de modo que ambas são estacionárias. Em um sistema linear de entrada e saída simples as séries de entrada X_t e saída Y_t são relacionadas através de um filtro linear (2.26)

$$Y_t = v(B)X_t + N_t \tag{2.26}$$

onde, $v(B) = \sum_{-\infty}^{\infty} v_j B^j$ é referido como a função de transferência de filtro por Box e Jenkins (1976) e N_t é série de ruído do sistema que é independente da série de entrada x_i .

Os coeficientes no modelo da função de transferência (2.26) são frequentemente chamados de pesos da resposta impulso. O modelo de função de transferência é considerado estável se a sequência desses pesos da resposta impulso é finita, i.e., $\sum |v_i| < \infty$. Portanto, em um sistema estável uma entrada limitada sempre produz uma saída limitada. O modelo de função de transferência é dito ser causal se $v_j = 0$ para j < 0, em outras palavras, o sistema não responde à série de entrada até que ela tenha sido realmente aplicada no sistema. Um modelo causal e estável é geralmente descrito como (2.27),

$$Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \ldots + N_t = v(B) X_t + N_t$$
(2.27)

onde $v(B) = \sum_{-\infty}^{\infty} v_j B^j$, $\sum |v_i| < \infty$ e X_t e N_t são independentes, e assim (2.26) é dado por (2.28),

$$Y_t = v(B)X_t + N_t \tag{2.28}$$

Wei (1990) esquematiza o sistema de função de transferência conforme Figura 2.2.



Figura 2.2 Sistema dinâmico de função de transferência (adaptado de Wei, 1990)

O objetivo do modelamento de função de transferência é identificar e estimar a função de transferência v(B) e o modelo de ruído para N_t com base na informação disponível das séries de entrada X_t . Desta forma, v(B) pode ser representada por (2.29),

$$v(B) = \frac{w_s(B)B^b}{\delta_r} \tag{2.29}$$

onde, $w_s(B) = w_0 - w_1B - \ldots - w_sB^s$, $\delta_r(B) = 1 - \delta_1B - \ldots - \delta_rB^r$, e b é um parâmetro de defasagem que representa o *lag* do tempo presente que decorre antes que o impulso da variável de entrada produza um efeito sobre a variável d saída. Para um sistema estável é assumido que as raízes de $\delta_r(B) = 0$ estão fora do círculo unitário.

Uma vez que $w_s(B)$, $\delta_r(B)$ e *b* são encontrados os pesos v_j da resposta impulso podem ser obtidos igualando-se os coeficientes de B^j em ambos os lados da equação (2.30),

$$\delta_r(B)v(B) = w_s(B)B^b \tag{2.30}$$

2.4.2 Função de Correlação Cruzada - FCC

Da mesma forma que a FAC para identificar modelos estocásticos, a ferramenta utilizada para a análise de dados e identificação de modelos de função de transferência é a função de correlação cruzada entre a entrada e a saída (BOX e JENKINS, 2006).

Wei (1990) descreve que a FCC é utilizada para medir a relação existente entre duas variáveis aleatórias. Assim, se X_t e Y_t são processos estocásticos estacionários conjuntamente para t = $0, \pm 1, \pm 2, \ldots$, e se X_t e Y_t são processos univariados estacionários e, a correlação cruzada entre X_t e Y_t , $Cov(X_t, Y_t)$, é uma função somente da diferença de tempo (s - t). Em tais casos, a função de covariância cruzada entre X_t e Y_t é dada por (2.31),

$$\gamma_{xy}(k) = E[(X_t - \mu_x)(Y_{t+k} - \mu_y)]$$
(2.31)

para $k = 0, \pm 1, \pm 2, \dots$ Sob a padronização, se tem a função de correlação cruzada (2.32):

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \tag{2.32}$$

para $k = 0, \pm 1, \pm 2, \ldots$, onde σ_x e σ_y são os desvios padrões de X_t e Y_t , respectivamente.

2.4.3 Relação entre função de correlação cruzada e função de transferência

Para um tempo t + k, o modelo de função de transferência (2.27) pode ser escrito por (2.33),

$$Y_{t+k} = v_0 X_{t+k} + v_1 X_{t+k-1} + v_2 X_{t+k-2} + \dots + N_{t+k}$$
(2.33)

Sem perda de generalidades pode-se assumir que $\mu_x = 0$ e $\mu_y = 0$. Multiplicando ambos os termos em (2.30) por X_t , e calculando a esperança se tem (2.34):

$$\gamma_{xy}(B) = v_0 \gamma_{xx}(k) + v_1 \gamma_{xx}(k-1) + v_2 \gamma_{xx}(k-2) + \dots$$
(2.34)

Lembre-se que ao adotar $\gamma_{xn}(B) = 0$, para todo k, se tem (2.35):

$$\rho_{xy}(k) = \frac{\sigma_x}{\rho\mu} [v_0 \rho_x(k) + v_1 \rho_x(k-1) + v_2 \rho_x(k-2) + \ldots]$$
(2.35)

A relação entre a FCC, $\rho_{xy}(k)$, e a função resposta impulso, v_j , é contaminada pela estrutura de auto-correlação das séries de entrada X_t . Se a série de entrada for um ruído branco, isto é, $\rho_x(k) = 0$ para $k \neq 0$, a equação (2.35), se reduz a (2.36):

$$v_k = \frac{\sigma_y}{\sigma_x} \rho_{xy} \tag{2.36}$$

Observa-se assim que a função de resposta de impulso v_k é diretamente proporcional a função de correlação cruzada $\rho_{xy}(k)$.

No modelo geral de função de transferência, equação (2.26), pode-se assumir que a série de entrada X_t segue um processo ARMA, equação (2.14), e $a_t = \alpha_t$ (ruído branco) definida como (2.37),

$$\alpha_t = \frac{\phi_x(B)}{\theta_x(B)} X_t \tag{2.37}$$

é chamada frequentemente de série de entrada pré ajustada.

Aplicando um processo de transformação de pré ajustamento na série de saída Y_t , obtém-se a série de saída filtrada (2.38).

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)} Y_t \tag{2.38}$$

fazendo, $\epsilon_t = \theta_x^{(-1)}(B)\phi_x(B)n_t$, o modelo de função de transferência (2.24), substituindo (2.26) e (2.37) em (2.38), torna-se (2.39):

$$\beta_t = v(B)\alpha t + \epsilon t \tag{2.39}$$

Daí, os pesos de resposta de impulso v_j , para a função de transferência podem, consequentemente, ser encontrados como (2.40):

$$v_k = \frac{\sigma_\beta}{\sigma_\alpha} \rho_{\alpha\beta}(k) \tag{2.40}$$

onde a partir de então, as etapas fundamentais de identificação dos modelos de função de transferência podem ser conduzidas.

2.4.4 Função de correlação cruzada amostral

Para um conjunto de dados das séries $X_t \in Y_t$, $1 \le t \le n$, a função de correlação cruzada (2.32) passa a ser estimada pela seguinte função de correlação cruzada amostral (2.41),

$$\widehat{\rho}_{xy}(k) = \frac{\widehat{\gamma}_{xy}(k)}{S_x S_y}, \quad k = \pm 1, \pm 2, \dots$$
(2.41)

onde,

$$\widehat{\gamma}_{xy}(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n} -k(X_t - \overline{x})(Y_{t+k} - \overline{y}), & k \ge 0\\ \\ \frac{1}{n} \sum_{t=1-k}^{n} (X_t - \overline{x})(Y_{t+k} - \overline{y}), & k < 0 \end{cases}$$
(2.42)

com, $S_x = \sqrt{\widehat{\gamma}_{xx}(0)}$ e $S_y = \sqrt{\widehat{\gamma}_{yy}(0)}$, e \overline{x} e \overline{y} são as médias amostrais das séries X_t e Y_t , respectivamente.

2.4.5 Identificação do modelo de função de transferência

A função de transferência v(B) pode ser obtida procedendo da seguinte forma:

- 1. Pré ajustando as séries de entrada (2.14), com $a_t = \alpha_t$ (ruído branco), conforme (2.37);
- 2. Calcular a série de saída filtrada. Transformar a série de saída Y_t utilizando o modelo pré ajustado no item 1 para gerar (2.38);

3. Calcular a função de autocorrelação cruzada amostral $\hat{\rho}_{\alpha\beta}(k)$, para estimar v_k , dado por (2.43)

$$\widehat{v}_k = \frac{\widehat{\sigma}_\beta}{\widehat{\sigma}_\alpha} \widehat{\rho}_{\alpha\beta}(k) \tag{2.43}$$

4. Identificar b, δ_r(B) e w_s(B) pelo ajuste padrão de v̂_k com o de v_k, uma vez escolhidos b, r e s, as estimativas preliminares δ̂_j e ŵ_j relacionados com v_k podem ser estimadas como mostrado em (2.30). Portanto, tem-se uma estimativa da função de transferência v(B) como (2.44):

$$\widehat{v}_t = \frac{\widehat{w}_s(B)}{\widehat{\delta}_s(B)} B^b \tag{2.44}$$

2.4.6 Identificação do modelo ruído

Obtido o modelo de função de transferência, pode-se calcular a série de ruídos estimada por (2.46),

$$\widehat{N}_t = Y_t - \widehat{v}(B)X_t = Y_t - \frac{\widehat{w}(B)}{\widehat{\delta}(B)}B^b X_t$$
(2.45)

O modelo apropriado para o ruído pode ser identificado examinando-se sua FAC e FACP ou por outra ferramenta de identificação de séries univariadas de tempo. Então,

$$\phi(B)n_t = \theta(B)a_t \tag{2.46}$$

Combinando (2.45) e (2.46) tem-se o seguinte modelo de função de transferência (2.47):

$$Y_t = \frac{w(B)}{\delta(B)} X_{t-b} + \frac{\theta(B)}{\phi(B)} a_t$$
(2.47)

2.5 K-means

O processo de clusterização utilizado se baseia em um método não hierárquico que utiliza o algoritmo *K-Means* para identificar clusters de conjuntos semelhantes. Na técnica, o critério de semelhança entre dois conjuntos (que define a participação ou não de um conjunto no *cluster*) leva em conta a soma da distância euclidiana entre todas as variáveis destes conjuntos. A solução do processo de clusterização será aquela na qual se minimiza a dispersão intra-*clusters* e maximiza a dispersão inter-*cluster* (Hair *et al*, 2005). Neste trabalho, esta técnica foi utilizada para definir o ponto que separa a amostra de dados de DP's em dois grupos e, em seguida, analisar se há relação entre um dos grupos e as manobras do Sistema elétrica de potência.

FURMIGARE, Matheus Seribeli

2.6 Análise de contingência

A análise de contingência é um teste não paramétrico indicado para determinar a magnitude da associação de variáveis mensuradas em nível ordinal e dispostas em tabelas de contingência nxn. O valor do tamanho de cada amostra deve ser ≥ 2 .

Em uma tabela bi-dimensional (Tabela 2.1) com N observações classificadas com relação a dois métodos de agrupamento, pode-se cruzar diversas características relevantes aos métodos pesquisados com diversas variáveis, tomadas duas a duas. Cada uma das células $n_{11}, n_{12}, \ldots, n_{2j}$, \ldots, n_{ij} representa a associação ou a contagem de grupos em cada um dos métodos aplicado.

Método I			Méte	odo J			Total
	1	2	•••	j	•••	J	-
1	n_{11}	n_{12}	•••	n_{1j}	•••	n_{1J}	$n_{1.}$
2	n_{21}	n_{22}	•••	n_{2j}	•••	n_{2J}	$n_{2.}$
•••	•••	• • •	•••	• • •	• • •	•••	• • •
l	n_{l1}	n_{l2}	•••	n_{lj}	•••	n_{lJ}	$n_{l.}$
•••	•••	• • •	• • •	•••	• • •	• • •	• • •
L	n_{L1}	n_{L2}	•••	n_{Lj}	•••	n_{LJ}	$n_{L.}$
Total	$n_{.1}$	$n_{.2}$	•••	$n_{.j}$	•••	$n_{.J}$	N

Tabela 2.1 Forma geral de uma tabela de contingência de duas dimensões.

Para a detecção de associação entre os métodos, ou seja, saber se as diferenças observadas entre os métodos são significativos, foi aplicado o teste qui-quadrado (χ^2) por meio da equação (2.48),

$$\chi^2 = \sum_{i=1}^{l} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})}{E_j}$$
(2.48)

onde, l = linhas, c = colunas, O_{ij} é a freqüência observada e E_{ij} é a freqüência esperada, dada por (2.49),

$$E_{ij} = \frac{n_{.j}n_{i.}}{N} \tag{2.49}$$

e, o coeficiente de contingência (C) é dado por (2.50),

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} \tag{2.50}$$

onde, N = soma de todas as categorias.

FURMIGARE, Matheus Seribeli

Fonte: Modificado de Everitt (1992)

O método usado para decidir se o teste χ^2 é independente ou se não estão associado, ou seja, se é possível rejeitar a hipótese de nulidade, H_0 , é baseado na distribuição de probabilidade χ^2 sob a pressuposição de que a hipótese nula é verdadeira.

O número de graus de liberdade, gl, para a estatística χ^2 , é definido por (2.51),

$$gl = (L-1)(J-1)$$
(2.51)

Capítulo 3

Análise dos dados

3.1 Localização, sistema de medição e análises descritiva dos dados

O autotransformador (AT) objeto do estudo encontra-se em Vila do Conde, cidade de Barcarena, Pará. Barcarena que está localizada a uma latitude 01°30'21"sul, longitude 48°37'33" oeste e a uma altitude de 15 metros em relação ao nível do mar. A Figura 3.1 destaca no mapa do Brasil a localização da cidade de Barcarena - PA.



Figura 3.1 Localização geográfica de Vila do Conde, Barcarena - PA.

Barcarena se destaca por ser um importante polo industrial do estado do Pará, onde é feita a industrialização, beneficiamento e exportação de caulim, alumina, alumínio e cabos para transmissão de energia elétrica. O banco de autotransformadores, denominado VCAT7 - 01 juntamente, com os bancos VCAT7 - 02 e VCAT7 - 03, ambos trifásicos (fases A, B e V), fornecem energia ao sistema elétrico da ALBRÁS, ALUNORTE, VALE PARAGOMINAS e CELPA (Belém, Barcarena e Santa Maria). A Figura 3.2 mostra o banco de autotransformadores VCAT7 - 01 formado por três AT monofásicos.



Figura 3.2 Banco de autotransformadores: cada autotransformador representa uma fase.

A Figura 3.2 mostra o banco de AT é formado por 3 AT monofásicos. A demanda máxima formada por essas cargas equivale a 1872 MW (Potência Ativa - P), 2080 MVA (Potência Aparente - S) com fator de potência 0,9 (razão entre P e S). Juntos os três bancos possuem 2250 MVA, ou seja, 750 MVA cada. O sistema encontra-se em situação de risco, isto por que, a perda de um ou dois desses AT de 750 MVA e 500/230 kV (recebe tensão de 500 kV e reduz para 230 kV) poderá causar sobrecarga nos autotransformadores remanescentes e gerar graves consequências, como o desligamento geral da subestação de Vila do Conde e, por conseguinte, a interrupção no fornecimento de energia.

É fundamental, portanto, que os transformadores de potência sejam periodicamente submetidos a processos de avaliação do seu sistema isolante para se verificar possíveis defeitos incipientes e, se comprovado por outros ensaios realizar uma manutenção corretiva desses equipamentos.

Neste contexto, a monitoração das DP se mostra eficaz no sentido de não afetar o funcionamento do AT. O sistema que realiza as medições de DP é denominado ICMmonitor, um sistema comercial que funciona como um filtro passa - alta, captando os pulsos de corrente de alta freqüência originados das DP.

Para caracterizar satisfatoriamente a ocorrência de descargas parciais, faz-se necessária a aquisição dos sinais durante vários ciclos da rede. A forma adotada para registrar a ocorrência dessas DP, foi pela construção de um mapa estatístico de descarga parcial (LEVY *et al*, 2006). A técnica de análise de descargas parciais que é conhecida como Phase Resolved Partial Discharge Analysies (PRPDA) ou Análise das Distribuições das Descargas Parciais em Função do Ângulo de Fase consiste em registrar cada descarga parcial com a sua amplitude, que é proporcional à carga aparente (sua unidade de medida é *picoCoulombs - pC*), e ao ângulo de fase (ϕ_i) em relação a um sinal de referência da tensão aplicada à amostra (FILHO, 2000). A Figura 3.3

mostra o mapa estatístico gerado pelo sistema de medição, onde cada ponto representa a amplitude do sinal de descarga parcial (eixo vertical) e a posição em relação ao ângulo de fase do sinal de referência, ilustrada na parte superior da figura.



Figura 3.3 Mapa estatístico de DP: identificando a quantidade de DP, a localização em relação ao ângulo de fase da tensão e a intensidade em pico Coulombs.

Os dados coletados formaram uma série temporal e foram coletados remotamente via sistema ICMmonitor, durante o período de 01 de agosto de 2011 a 22 de Dezembro de 2011, representando valores de DP em intervalos de uma hora. Porém, como veremos adiante, os dados de temperatura e umidade relativa do ar são medidos em intervalos de três horas e, então, optou-se por analisar os dados coletados a cada três horas. Dessa maneira, foram coletados 1152 dados de cada sensor de DP do AT da fase B do banco VCAT7-01. Os quais representados por **DP - fase 230 kV** (Figura 3.4) e **DP - fase 500 kV** (Figura 3.5), respectivamente a saída de tensão 230 kV e entrada de tensão 500 kV do AT.



Figura 3.4 Gráfico de Descargas Parciais - DP - fase 230kV.



Figura 3.5 Gráficos de Descargas Parciais - DP - fase 500kV.

Conforme as Figuras 3.4 e 3.5, todos os dados de DP coletados durante o período de 01 de Agosto de 2011 a 22 de Dezembro de 2011 estão representados, cuja unidade de medida é pico-Coulomb - pC. Notam-se elevadas freqüências, porém sem tendência, podendo, as séries serem consideradas estacionárias. Os dois intervalos com maiores valores de DP, foram da segunda quinzena do mês de Setembro a primeira quinzena do mês de Outubro, e, da segunda quinzena do mês de Novembro a primeira quinzena do mês de Dezembro nas duas séries. A Tabela 3.1 apresenta as estatística descritivas para o comportamento das DP's.

Estatísticas	DP - Fase 230 kV	DP - Fase 500 kV
Dados	1152	1152
Mínimo	$93,\!02$	$53,\!57$
Máximo	$3953,\!50$	$2276,\!80$
Média	577,77	$328,\!95$
Desvio Padrão	$480,\!85$	$356,\!13$
Assimetria	$1,834^{*}$	$2,048^{*}$
Curtose	$4,264^{*}*$	$4,949^{*}*$

Tabela 3.1 Estatísticas básicas dos dados de descargas parciais.

Erro Padrão de 0,72; **Erro Padrão de 0,144.

Conforme a Tabela 3.1 cada série contém 1152 dados de DP, ambas com relação de 42,5 vezes entre o menor e o maior valor, porém em média a série DP - fase 500 kV apresenta valor menor comparado à de DP - fase 230 kV. Em relação à assimetria e curtose, observa-se que ambas apresentam assimetria positiva e são platicúrtica (curtose maior que 0,263, valor de referência de uma distribuição normal).

Para visualizar o comportamento das DP, foram gerados gráficos Box-Plot e verificou-se que os dados de DP têm comportamento parecido, porém os dados DP - fase 500 kV se apresenta menos dispersos e com menores valores em comparação aos de DP - fase 230 kV, Figura 3.6a.



Figura 3.6 Gráficos Box-Plot com linha de referência horizontal em 500 pC: (a) Variáveis DP - fase 230kV e DP - fase 500kV; (b) Variável DP - fase 230 kV a cada hora e (c) Variável DP - fase 500 kV a cada hora.

Conforme os gráficos da Figura 3.6 (b e c) os maiores valores de DP ocorrem das 18h00 às 06h00, principalmente às 03h00 e 06h00. Especialistas da Eletronorte não souberam explicar o motivo dos maiores valores ocorrerem nesses horários.

Para avaliar o comportamento das séries de DP foram coletadas variáveis do sistema e ambientais, relacionadas abaixo:

• Corrente da fase de 230 kV - CB230;

- Corrente da fase de 500 kV CB500;
- Potência Ativa da fase de 230 kV PA230;
- Potência Ativa da fase de 500 kV PA500;
- Potência Reativa da fase de 230 kV PR230;
- Potência Reativa da fase de 500 kV PR500;
- Temperatura do Óleo TO;
- Temperatura de Enrolamento da fase de 13 kV TEB13;
- Temperatura de Enrolamento da fase de 230 kV TEB230;
- Temperatura de Enrolamento da fase de 500 kV TEB500;
- Tensão entre as fases A e B de 230 kV T_AB230;
- Tensão entre as fases A e B de 500 kV T_AB500;
- Tensão entre as fases B e V de 230 kV T_BV230;
- Tensão entre as fases B e V de 500 kV T_BV500;
- Tensão entre as fases A e V de 230 kV T_AV230;
- Tensão entre as fases A e V de 500 kV T_AV500;
- Temperatura de Ar TA;
- Umidade Relativa do Ar UR.

As Tabelas A1.1 e A.1.2 no anexo A apresentam as estatísticas descritivas e análise de correlação dessas variáveis.

3.2 Análise fatorial

A Análise Fatorial (AF) foi utilizada neste estudo para se reduzir o número de variáveis por meio da formação de fatores não correlacionados e utilização dos escores obtidos na análise de série temporal. A verificação da adequabilidade da aplicação da AF foi através do teste de KMO e do teste de esfericidade de Bartlett. O KMO apresentou valor de 0,902 e o teste de esfericidade de Bartlett obteve $\chi^2_{153} = 49433, 22$ com nível descritivo p = 0,000, resultados que indicaram a adequabilidade da AF (Tabela A.1.3, Anexo A), com comunalidades acima de 0,5 (HAIR et al., 2005).

Os três fatores explicaram 84,36% da variabilidade total, estes foram obtidos pelo método de componentes principais e rotação Varimax.

			Fatores	
Variáveis	-	$\mathbf{F1}$	$\mathbf{F2}$	F3
T_AB500		0,915	-0,229	0,090
T_BV500		0,915	-0,217	0,108
T_VA500		0,912	-0,212	0,108
PR230		0,820	-0,082	0,086
PR500		-0,842	0, 127	-0,067
CB230		-0,728	0,429	0, 190
CB500		-0,850	0,397	0,168
PA230		0,738	-0,419	-0,248
PA500		0,735	0,414	0,256
TOB		-0,215	0,947	0,005
TEB13		-0,220	$0,\!942$	-0.008
TEB230		-0.214	$0,\!945$	0.003
TEB500		-0.292	0,919	0,013
TA		-0,249	-0,807	0,082
UR		0,250	-0,730	0,016
T_AB230		-0,008	0,020	0,966
T_BV230		0,041	0,016	0,960
T_VA230		-0,019	0,015	$0,\!974$
Soma das cargas	Total	6,58	5,56	3,04
quadradas	% da Variância	36, 54	30, 89	16,92
rotacionadas	% da Variância			
	Acumulada	36,54	67, 43	84, 36

Tabela 3.2 Cargas fatoriais dos componentes rotacionadas e variância total explicada.

O primeiro fator (F1) contribuiu com 36,54% da variância total. Este representou as tensões de entrada entre fases (A, B e V) de 500 kV, as potências ativas e reativas de 230 kV e 500 kV e soo kV e as correntes 230 kV e 500 kV. Este fator, com maior contribuição, foi representado pelas

variáveis de segurança do sistema. Segundo especialistas, esta maior variação pode ser devido à falta de controle pelo sistema sobre essas variáveis, pois a potências ativas e reativas dependem do consumo dos clientes, e as tensões entre fases na entrada de 500 kV dependem da geração na Usina Hidrelétrica de Tucuruí, em Tucuruí - PA.

O segundo fator (F2), com 30.89% de contribuição à variância total, representou as variáveis de temperatura do equipamento (AT) e a temperatura e umidade relativa do ar. Este fator foi denominado de variáveis de temperatura e esta separação pode ser explicada pelo fato do sistema ser susceptível à temperatura externa, embora haja equipamentos que controlem a temperatura interna, evitando assim, o superaquecimento do sistema.

O terceiro fator representou 16,92% da variância total dos dados e está relacionada às tensões de fase de saída (230 kV), esta separação foi obtida pelo fato das tensões de saída não podem variar muito em relação a 230 kV, existindo manobra internas e comutações de tensão.

3.3 Modelagem multivariada com função de transferência

Uma série temporal, muitas vezes, pode apresentar um componente relacionado ao sinal ou informação e outra relacionada a um ruído. Quando há alguma informação na série, esta é caracterizada pela presença de um padrão não aleatório associado, em geral, à ocorrência de correlações entre as observações em tempos diferentes.

A determinação de quais são as observações que contribuem para a formação do padrão da série pode ser identificada pela função de autocorrelação, autocorrelação parcial e a correlação cruzada, no caso das séries temporais multivariadas. Além disso, a correlação cruzada indica a relação entre duas séries em tempos diferentes.

Para a modelagem multivariada com função de transferência, foram utilizadas 1124 observações e 24 observações para a validação dos modelos. Foram consideradas as variáveis DP - fase 230 kV como resposta, DP - fase 500 kV, os fatores F1, F2 e F3 como variáveis de entrada do modelo de função de transferência. Para simplificar, a seguir a variável DP - fase 230 kV será tratada como Y, e as variáveis DP - fase 500 kV, F1, F2 e F3 serão tratadas como X_1 , X_2 , X_3 e X_4 .

Os gráficos das séries são apresentados nas Figuras 3.7 a 3.11, nas quais observam-se que todas as séries são estacionárias e não há indícios de sazonalidade, não sendo assim, necessário diferenciações ou transformações. Entretanto, observa-se uma acentuada variabilidade relativa presente em todas elas.



Figura 3.7 Gráfico das Descargas Parciais - DP - fase 230kV(Y).



Figura 3.8 Gráfico das Descargas Parciais - DP - fase 500kV (X_1) .



Figura 3.9 Gráfico dos Score gerado pela AF para o fator 1 - F1 (X_2) .



Figura 3.10 Gráfico dos Score gerado pela AF para o fator 2 - $F2(X_3)$.



Figura 3.11 Gráfico dos Score gerado pela AF para o fator 3 - F3 (X_4) .

3.3.1 Ajuste dos modelos para as séries de entrada

Para se ajustar modelos de função de transferência devem-se previamente ajustar as séries de entrada por modelos ARIMA, conforme descrito em ??, pois estes terão como finalidade a préfiltragem para análise das Funções de Correlação Cruzada (FCC). Na tentativa de se ajustar modelos mais significativos para as variáveis explicativas, diversos modelos ARIMA são ajustados até que se encontre aquele que se ajuste mais adequadamente aos dados. A escolha dos modelos a serem ajustados são baseados nas análises das funções de autocorrelação (FAC) e autocorrelação parcial (FACP), Figuras 3.12 a 3.15.

Pode-se verificar analisando os gráficos das Funções de Autocorrelações e de Autocorrelações Parciais das Figuras 3.12 e 3.13, respectivamente, que a variável X_1 apresenta, para a FAC um decaimento lento e com formato senoidal, característica de uma série com sazonalidade, porém, optou-se por não realizar nenhuma transformação aos dados, e para a FACP não se observou um comportamento característico, destacando-se elevadas significâncias para os lag's, ou defasagens, 6, 7, 8 e 9.



Figura 3.12 Função de Autocorrelação da variável (X_1) .



Figura 3.13 Função de Autocorrelação Parcial da variável (X_1) .

Optou-se por não realizar nenhum tipo de transformação, deixando a cargo do modelo representar a característica da informação contida na série. Nesse sentido, após algumas tentativas, foi ajustado o modelo (3.1) abaixo. O modelo ajustado para a série da variável X_1 apresentou, de certo modo, superparametrização, como era esperado conforme se observou nas FAC e FACP da série. Essa superparametrização pode ser atribuída à presença de frequências muito altas. Foi possível observar, entretanto, que as autocorrelações residuais não foram significativas indicando ausência de autocorrelação e um bom ajuste do modelo. E, assim, o modelo ARIMA ajustado para a variável X_1 foi:

$$(1-0, 3B-0, 1B^2)(1-0, 94B^8)X_{1t} = (1+0, 05B^4+0, 05B^6+0, 17B^7-0, 65B^8+0, 10B^9)a_t \quad (3.1)$$

Na Figura 3.14, observa-se que a FAC da variável X_2 apresenta uma queda após o *lag* 2, porém, elevada significância para as defasagens múltipla de 8, e na Figura 3.15 a FACP com vários *lag's* significativos



Figura 3.14 Função de Autocorrelação da variável (X_2) .



Figura 3.15 Função de Autocorrelação Parcial da variável (X_2) .

Do mesmo modo, optou-se por não realizar nenhum tipo de transformação, deixando modelo capturar as características da informação contida na série. Nesse sentido, após algumas tentativas, foi ajustado o modelo (3.2) abaixo. O modelo ajustado para a série da variável X_2 apresentou-se, também, superparametrização, pelo mesmo motivo da série anterior. Observou-se que as autocorrelações residuais não foram significativas indicando ausência de autocorrelação e um bom ajuste do modelo. E, assim, o modelo ARIMA ajustado para a variável X_2 foi:

$$(1-0,69B-0,12B^3-0,08B^5-0,12B^8+0,08B^{1}6)(1-0,99B^8)X_{2t} = (1+0,98B^8)a_t \quad (3.2)$$

A Figura 3.16 e 3.17, FAC e FACP da variável X_3 , respectivamente, observou-se comportamento parecido com o da variável X_1 , porém com alternância entre valores significativos positivos e negativos, também, característica de uma série com sazonalidade. Ressalta-se na FACP os valores significativos nos *lag's* 2, 6, 7, 8 e 9.



Figura 3.16 Função de Autocorrelação da variável (X_3) .



Figura 3.17 Função de Autocorrelação Parcial da variável (X_3) .

Do mesmo modo, não realizou-se nenhum tipo de transformação, a série capturou a caracterísitica da informação contida, pois, as autocorrelações residuais não foram significativas indicando ausência de autocorrelação e um bom ajuste do modelo. Que, após algumas tentativas, foi ajustado conforme o modelo (3.3) abaixo. Também, apresentou, de certo modo, superparametrização, como era esperado conforme se observou nas FAC e FACP da série. Essa superparametrização pode ser atribuída à presença de frequências muito altas. E, assim, o modelo ARIMA ajustado para a variável X_3 foi:

$$(1 - 0,87B - 0,20B^2 - 0,17B^7)(1 - 0,99B^8)X_{3t} = (1 + 0,85B^8)a_t$$
(3.3)

Na Figura 3.18 e 3.19 observa-se um comportamento parecido com a variável X_3 , porém mais atenuado, com baixos valores significativos para as FAC e FACP da variável X_4 .



Figura 3.18 Função de Autocorrelação da variável (X_4) .



Figura 3.19 Função de Autocorrelação Parcial da variável (X_4) .

Não realizou-se nenhum tipo de transformação e, como os modelos anteriores, o modelo representou a característica da informação contida na série. Pois, foi possível observar que as autocorrelações residuais não foram significativas indicando ausência de autocorrelação e um bom ajuste do modelo. E, após algumas tentativas, o modelo ajustado foi o (3.4) abaixo. Também, observou-se uma superparametrização, que pode ser atribuída à presença de frequências muito altas. E, assim, o modelo ARIMA ajustado para a variável X_4 foi:

$$(1 - 0, 26B - 0, 06B^6 - 0, 08B^{12} - 0, 09B^{18} + 0, 15B^{24})(1 - 0, 14B^8)X_{4t} = a_t$$
(3.4)

Para cada uma das séries foram ajustados modelos cujos parâmetros estimados obtiveram estimativas significativas a um nível descritivo de 5%, e verificou-se que os resíduos apresentaram

autocorrelações não significativas, o que caracteriza ajuste aceitável, ou seja, o modelo conseguiu capturar o padrão da série, conforme Tabelas A.2.1 ao A.2.8 (Anexo A.2).

3.3.2 Ajuste do modelo para a série de saída

Para o ajuste da série de saída, Y, foi seguido os mesmos procedimentos para as séries de entrada. Ajustou-se um modelo para a série de saída por modelo ARIMA, com intuito de pré-filtragem para análise das Funções de Correlação Cruzada (FCC). Porém, aproveitou-se para compararmos um modelo univariado com o modelo multivariado em relação a seus parâmetros. Do mesmo modo, a escolha do modelo procedeu observando as análises das funções de autocorrelação (FAC) e autocorrelação parcial (FACP), Figuras 3.20 e 3.21, respectivamente.



Figura 3.20 Função de Autocorrelação da variável (Y).



Figura 3.21 Função de Autocorrelação Parcial da variável (Y).

Pode-se verificar que a FAC e FACP, na Figura 3.20 e 3.21, apresenta para a FAC um decaimento lento e com formato senoidal, característica de uma série com sazonalidade, porém, optou-se por não realizar nenhuma transformação aos dados, e para a FACP não se observou um comportamento característico, destacando-se elevadas significâncias para os lag's, ou defasagens, 6, 7, 8 e 9. Após algumas tentativas foi ajustado o modelo (3.5), abaixo:

$$(1 - 0,96B^8)Y_t = (1 + 0,32B + 0,19B^2 + 0,04B^3 + 0,10B^6 + 0,21B^7 + 0,61B^8)a_t$$
(3.5)

O modelo (3.5), denominado de Modelo Y, é composto por um parâmetro autoregressivo e seis parâmetros de médias móveis. A estimativa do parâmetro autoregressivo de ordem 8 é igual a 0,96. Isso significa que, o valor de DP no período atual (t) é influenciado pelo valor de DP há 8 defasagens (lag's) atrás (t - 8) em cerca de 96%, cada lag representa o valor de DP coletado a cada 3 horas, portanto, quando refere-se a 8 lag's atrás é igual à 24 horas atrás. Nota-se que o modelo capturou a característica sazonal indicada pela FAC, observada na Figura 3.20.

Dentre os parâmetros de médias móveis destaca-se as estimativas $0,19, 0,32 \in 0,61$ de ordens 1, 2 e 8, respectivamente. Isso significa que, a estimativa do parâmetro de médias móveis de ordem um quer dizer que, na média, os erros da variável pessoal admitido são corrigidos em torno de 19% no período atual, em relação a seu próprio valor do lag anterior (3 horas atrás). O mesmo tipo de análise vale para o parâmetro de médias móveis, de ordem 2. Nesse caso, na média, o desequilíbrio no valor de DP é corrigido em torno de 32% no período atual, tendo como base seu respectivo valor passado de dois *lag's* anterior (6 horas atrás). A estimativa de média móvel de ordem 8 reforça a característica sazonal, o valor para a estimativa desse parâmetro mostra que, na média, os erros da variável de DP são corrigidos em 61%, a cada 8 *lag's* (24 horas).





Figura 3.22 Ajuste do Modelo (Y).



Na Figura 3.23 observa-se o ajuste das previsões para 24 lag's a frente, representando 3 dias.

Figura 3.23 Função de Autocorrelação Parcial da variável (Y).

Nota-se nas Figuras 3.22 e 3.23 que o Modelo Y capturou a relação dos valores de DP no curto e longo prazo, especialistas avaliaram este resultado positivamente, pois, era conhecida a influência de uma hora após a ocorrência de DP, ou seja, em curto prazo. As relações representadas pelos parâmetros 6,7 e, principalmente, 8 representaram novo conhecimento e proposta para estudos futuros. Os parâmetros estimados obtiveram estimativas significativas a um nível descritivo de 5% e os resíduos apresentaram autocorrelações não significativas, conforme Tabelas A.2.9 e A.1.10 (Anexo A.2.5), o que caracteriza a adequabilidade do modelo, ou seja, o modelo conseguiu capturar o padrão da série.

Foram utilizados os critérios de Akaike e SBC para avaliar o ajuste do modelo em relação ao ajuste do Modelo de FT adiante, o menor valor nessas estatísticas representa melhor ajuste. A Tabela 3.3 mostra os valores obtidos para os critérios de Akaike e SBC, e os valores de ajuste RMSE e MAPE para o ajuste do modelo Y.

Modelo Y
15889,74
15924, 94
284, 11
40,48%

Tabela 3.3 Critérios de ajuste e previsão para modelo 3.5.

3.4 Correlações cruzadas

Com o objetivo de se verificar a relação de influência das variáveis explicativas na variável resposta, foram construídas e analisadas as funções de autocorrelações cruzadas.

3.4.1 Análise da correlação cruzada entre $YvsX_1$

A Figura 3.24 mostra o gráfico da função de correlação cruzada entre as variáveis de saída Y e a variável de entrada X_1 .



Figura 3.24 Função de Correlação Cruzada $YvsX_1$.

Observa-se na Figura 3.24 que há correlação significativa no *lag* zero entre as séries de entrada e saída, indicando que estão em fase, ou seja, uma relação direta e positiva entre $Y \in X_1$. Portanto, um aumento na série X_1 ocasiona um aumento em Y para o mesmo *lag* comparado. Além disso, observa-se *lag's* positivos e negativos com correlações significativas, indicando serem necessários parâmetros no numerador e no denominador do modelo de função de transferência.

3.4.2 Análise da correlação cruzada entre $YvsX_2$

A Figura 3.25 mostra o gráfico da função de correlação cruzada entre as variáveis de saída Y e a variável de entrada X_2 .

Observa-se na Figura 3.25 que há correlação significativa no *lag* zero entre as séries de entrada e saída, indicando que estão em fase, porém, com uma relação direta e negativa entre Ye X_1 . Portanto, um aumento na série X_1 ocasiona uma diminuição em Y para o mesmo *lag*



Figura 3.25 Função de Correlação Cruzada YvsX₂.

comparado. Além disso, observam-se *lag's* positivos e negativos com correlações significativas, indicando serem necessários parâmetros no numerador e no denominador do modelo de função de transferência.

3.4.3 Análise da correlação cruzada entre $YvsX_3$

A Figura 3.26 mostra o gráfico da função de correlação cruzada entre as variáveis de saída Y e a variável de entrada X_3 .



Figura 3.26 Função de Correlação Cruzada YvsX₃.

Observa-se na Figura 3.26 que não há correlação significativa no lag zero entre as séries de

entrada e saída, indicando que não estão em fase. Porém, observa-se um valor significativo no *lag* 16, característico de um modelo com *feedback*, essa característica indica ser necessário préajustar a série de saída e, pode ser necessários, parâmetros no denominador do modelo de função de transferência.

3.4.4 Análise da correlação cruzada entre $YvsX_4$

A Figura 3.27 mostra o gráfico da função de correlação cruzada entre as variáveis de saída Y e a variável de entrada X_4 .



Figura 3.27 Função de Correlação Cruzada YvsX₄.

Observa-se na Figura 3.27 que há correlação significativa no *lag* zero entre as séries de entrada e saída, indicando que estão em fase, ou seja, uma relação direta e positiva entre $Y \in X_1$. Portanto, um aumento na série X_1 ocasiona um aumento em Y para o mesmo *lag* comparado. Além disso, observam-se *lag's* positivos e negativos com correlações significativas, indicando serem necessários parâmetros no numerador e no denominador do modelo de função de transferência.

3.5 Ajuste dos modelos de função de transferência

A partir das análises feitas sobre as funções de correlação cruzada, um modelo de função de transferência foi testado, sendo que, houve a necessidade de se fazer o ajuste da componente de erro ou ruído no modelo final da função de transferência, o que contribuiu para a grande quantidade de parâmetros no modelo, abaixo o modelo está representado na forma do operador atraso B. Em seguida apresenta-se a estimativa para estruturação do modelo proposto.

$$Y_{t} = \frac{(\omega_{01} + \omega_{11}B^{7})}{(1 - \delta_{11}B)(1 - \delta_{21}B^{8})}X_{1t} + \frac{(\omega_{02} + \omega_{12}B^{3})}{(1 - \delta_{12}B^{5})}X_{2t} + \frac{\omega_{03}}{(1 + \delta_{13}B^{2} - \delta_{13}B^{3})}X_{3t} + \omega_{04}X_{4t} + (1 + \theta_{1}B + \theta_{2}B^{2})(1 + \theta_{6}B^{6})$$
(3.6)
$$(1 + \theta_{7}B^{7})(1 + \theta_{8}B^{8})(1 + \theta_{9}B^{9})(1 + \theta_{15}B^{1}5)a_{t}$$

3.5.1 Estruturação para a estimativa do modelo

Na Tabela A.2.11 (Anexo A) observa-se que as estimativas dos parâmetros de cada variável de entrada e do termo de correção do tipo médias móveis, apresentaram valores significativos ao nível de 5%. Apenas, os parâmetros de média móvel de ordem 2 e o parâmetro da variável X_4 , ω_{04} , foram significativos ao nível de 10%. Portanto, foram utilizados para a formulação do modelo (3.7), denominado de Modelo FT.

$$Y_{t} = \frac{(0,69+0,13B^{7})}{(1-0,29B)(1-0,29B^{8})}X_{1t} + \frac{(6,79+28,97B^{3})}{(1-0,71B^{5})}X_{2t} + \frac{0,055}{(1+0,43B^{2}-0,78B^{3})}X_{3t} + 13,39X_{4t} + (1+0,24B+0,05B^{2})(1+0,1B^{6}) (3.7) (1+0,1B^{7})(1+0,14B^{8})(1+0,1B^{9})(1+0,07B^{1}5)a_{t}$$

onde, a_t é um ruído branco.

O significado físico dos parâmetros do modelo não é claro, porém se compararmos ao Modelo Y, equação (3.5), pode-se perceber que as variáveis de entrada substituíram o parâmetro autoregressivo de ordem 8 do Modelo Y, no sentido em que os parâmetros do termo de correção do erro foram de média móveis. E, aumentou a quantidade de parâmetros de médias móveis, no entanto com valores abaixo de 0,24, observado no *lag* 1.

Para a aquisição do conhecimento esse resultado se mostrou interessante, segundo especialistas, pois não havia conhecimento sobre os valores de DP apresentarem um comportamento típico de sazonalidade, e que este comportamento pode ser substituído por componentes de sistema, representadas pelas variáveis de entrada do Modelo FT

3.5.2 Análise residual para o Modelo FT

Para verificar o ajuste do Modelo FT é importante observar que as autocorrelações dos resíduos sejam não significativas, bem como, não apresente correlação cruzada significativa com as variáveis de entrada. Nesse sentido, verifica-se nas Tabelas A.2.12 à A.2.16 (anexo A) a análise residual para o Modelo FT, equação (3.7).

Observa-se na Tabela A.2.12 que os resíduos do modelo não apresentaram autocorrelações significativas, podendo ser representados por ruído branco. E, nas Tabelas A.2.13 à A.2.16, as quais representam a correlação cruzada entre os resíduos do Modelo FT e as variáveis de entrada X_1 , X_2 , X_3 e X_4 , respectivamente, se observam que não há correlação cruzada significativa evidenciando que as funções de transferências relacionadas a essas variáveis são adequadas. Portanto, esses resultados denotam a adequabilidade do Modelo FT proposto, a Figura 3.28 mostra o ajuste do modelo em relação aos dados de DP.



Figura 3.28 Modelo FT ajustado para variável Y.

Na Figura 3.28 observa-se que o modelo se ajustou bem aos dados observados. Foram utilizados os critérios de Akaike e SBC para avaliar o ajuste do modelo em relação ao ajuste do Modelo Y, equação (3.5). Dessa forma a Tabela 3.4 mostra os valores obtidos para os critérios de Akaike e SBC, e os valores de ajuste RMSE e MAPE para os modelos Y e FT, equações 3.5 e 3.7, respectivamente.

Modelo Y	Modelo FT
15889,74	15430, 23
15924, 94	15520, 59
284, 11	248, 35
40,48%	36,06%
	Modelo Y 15889,74 15924,94 284,11 40,48%

Tabela 3.4 Critérios de ajuste e previsão dos modelos Y e FT.

O modelo FT apresentou melhor ajuste em relação ao modelo Y, pois conforme a Tabela 3.4



os critérios AIC, SBC, RMSE e MAPE apresentaram menores valores para o modelo FT. Na Figura 3.29 observa-se o ajuste das previsões para 24 *lag's* a frente, representando três dias.

Figura 3.29 Previsões obtidas pelo modelo FT.

Na Figura 3.29 verifica-se que o modelo se ajustou bem aos dados para validação, e um valor fora do intervalo de confiança de 95%, o mesmo do modelo Y. Para a validação dos dados realizada pelo modelo FT obteve-se para o MAPE e o RMSE, 31.29% e 602.33, respectivamente. Comparando-se aos valores de MAPE e RMSE do modelo Y, cujos valores foram de 42,69% e 573,56, respectivamente.

Portanto, o modelo de função de transferência, Modelo FT, apresentou melhor ajuste em relação aos dados de DP, mesmo que para as previsões tenho apresentado valor de RMSE maior. Assim, para o processo de aquisição do conhecimento, observa-se que as variáveis do sistema utilizadas na AF se mostraram relevantes para o processo gerador das DP. No entanto, observou-se altas freqüências em alguns períodos da série de DP, podendo ser a causa da superparametrização do modelo. Neste sentido, foi proposto uma análise da relação entre as variáveis de manobras do sistema e as variáveis de DP.

3.6 Análise da relação dos dados de DP e as manobras no sistema

Para avaliar a relação entre os dados de DP e as manobras no sistema foi utilizada a análise de contingência, separando os dados de DP conforme agrupamento proposto pelo *k-means* realizado aos dados.

Na análise de contingência foram utilizados os dados de DP e os dados de manobras no sistema, entre 05 de agosto de 2011 e 31 de outubro de 2011. Foram coletados 250.000 dados de manobras, porém, após consultar os especialistas, reduziu-se para 33.961 dados, filtrando apenas as manobras de abertura e fechamento de disjuntores e seccionadoras, e comutações automáticas do AT. As comutações são manobras realizadas internamente no AT para evitar que a tensão transformada não varie muito em relação a 230 kV, por isso, de acordo com a AF ela não varia conforme a tensão de entrada 500 kV.

Os disjuntores e seccionadoras são equipamentos de segurança do SEP e localizam-se antes e depois do AT. A Figura 3.30 representa a localização e a denominação de cada tipo de manobra.



Figura 3.30 Esquema do sistema de transformação e distribuição.

A Figura 3.30 representa os equipamentos e suas localizações em relação ao AT, o bloco DJ7 representa todos os disjuntores localizados antes do AT, o bloco SC7 as seccionadoras e serão tratados como um bloco conforme entorno tracejado. O bloco AT representa o autotransformador, bem como suas manobras de comutações, e os blocos DJ6 e SC6 os disjuntores e seccionadoras, respectivamente, após o transformador e, também, serão tratados como um bloco apenas.

A energia gerada em Tucuruí - PA chega aos DJ7 e SC7, depois é transformada no AT e, em seguida, passa por DJ6 e SC6 antes de chegar aos consumidores. No entanto, há outros equipamentos envolvidos, porém não foram considerados, ou por não apresentarem manobras ocorridas no período analisado, ou por recomendação dos especialistas.

Observou-se que as manobras ou não ocorriam na mesma hora da coleta dos dados de DP, ou apenas um tipo de manobra ocorria, ou mais de uma ocorriam durante o tempo de coleta da DP. Dessa maneira foram classificadas as manobras de acordo com a Tabela 3.5, a qual o valor "0" representa não ocorrência e "1" ocorrência.

bela 3.5 <i>Classific</i> e	DJ7 e SC7	AT	DJ6 e SC6
Não Ocorrência	0	0	0
Ocorrência	0	0	1
:	:	÷	:
Ocorrência	1	1	1

Conforme visto na Tabela 3.5 as manobras de disjuntores e seccionadoras são representadas indistintamente, pois as seccionadoras são equipamentos que abrem ou fecham o circuito elétrico, assim como um disjuntor. Após avaliar os dados de manobras verificou-se que as comutações ocorrem com maior freqüência, inviabilizando sua separação das outras variáveis. Nesse sentido, optou-se por analisar os dados separando entre disjuntores e comutações de entrada, disjuntores e comutações de saída, apenas comutações e os três tipos de manobras ocorrendo simultaneamente, conforme ilustra a Tabela 3.6.

	DJSC7	AT	DJSC6
Nenhuma manobra.	0	0	0
Manobras nos disjuntores ou seccionadora após			
o AT, juntamente com manobras de comutações.	0	1	1
Apenas manobras de comutações.	0	1	0
Manobras nos disjuntores ou seccionadora antes			
do AT, juntamente com manobras de comutações.	1	1	0
Todas as manobras.	1	1	1

Tabela 3.6 Manobras selectionadas.

Realizando o k-means nos dados de DP dividiu-se a séries em dois grupos com ponto de corte de 345 pC. Dessa forma, a tabela de contingência gerada é apresentada na tabela 3.7.

Observa-se na Tabela 3.7 a maior quantidade de ocorrência da manobra de comutação, rep-

Tabela	3.7 Manobras	selecionadas.
	$DP \le 345pC$	DP > 345pC
000	105	91
011	41	27
010	197	154
110	19	46
111	9	15

resentada por **010**. Com essas freqüências, procedeu-se a análise de contingência e a análise de resíduo qui-quadrado, os resultados são apresentados nas tabelas 3.8 e 3.9.

T 1 1 9 0	ו ער ס	1	11.	1	·· ^ ·
Tapela 3.8	Resultados	aa	anause	ae	contingencia.
100010101010	10000000000	0000	0	~~~	0010001090100000

Tabela de Contingência	2x5
Coef. de Contingência (C)	$0,\!1657$
Qui - quadrado	$19,\!8723$
Graus de Liberdade	4
Р	0,0005

Observa-se na Tabela 3.8 que o resultado para a análise de contingência foi significativo ao nível de 1% e com valor de 0,1657 para o coeficiente de contingência. O resultado do teste quiquadrado mostra que há diferenças significativas entre as freqüências da Tabela 3.7. Portanto, analisando o teste de resíduos qui-quadrado tem-se o seguinte resultado mostrado na Tabela 3.9.

Tabela	3.9 Análise dos	resíduos.
	$DP \le 345pC$	DP > 345pC
000	0,2880	-0,2880
011	1,3198	-1,3198
010	1,8158	-1,8158
110	-3,9777	3,9777
111	-1,5174	1,5174
Nível Alfa	0.05	1.96
Nível Alfa	0.01	2.576

Observa-se na Tabela 3.9 que há relação positiva entre os valores de DP maiores que 345 pC e as manobras de disjuntores e seccionadoras antes do AT. Pois, verificou-se que quando

há a presença das manobras do tipos 1xx, ou seja, disjuntores e/ou seccionadores antes do AT (representadas por 110 ou 111), há uma relação positiva e de, aproximadamente, 3,98 e 1,52 para as manobras 110 e 111, respectivamente. Verifica-se que a relação com a não ocorrência de manobras é inversa, ou seja, há uma relação positiva entre não ocorrência de manobras e valores menores ou iguais à 345 pC.

No entanto, não se verifica uma relação entre os valores de DP maiores que 345 pC e as manobras de comutações, disjuntores e seccionadoras depois do AT, indicando que estas podem não estar influenciando as DP's com valores acima de 345 pC.

Portanto, para o processo de aquisição de conhecimento, esses resultados mostraram que as manobras dos disjuntores e seccionadoras podem estar influenciando nas medições das DP, ou seja, representado ruídos nos dados de DP. Para os especialistas estes resultados se mostraram relevantes e propõem mais estudos para avaliar esta influência (ou interferência). Porém, era esperada maior influência das comutações, pois é conhecido dos especialistas tais influências, no entanto, com estes resultados pode-se supor que as comutações influenciam os valores de DP, porém não superior a 345 pC.

Capítulo 4

Considerações Finais e Recomendações

Nesta dissertação procurou-se, por meio de métodos estatísticos, descrever e estudar o comportamento das séries de descargas parciais (DP's), avaliando-o em relação às variáveis do sistema elétrico de potência (SEP) e às manobras nos equipamentos da subestação.

Para avaliar a relação entre as DP's e as variáveis do SEP foi realizada uma redução das variáveis do SEP por meio da técnica estatística multivariada análise fatorial (AF). Esta técnica se mostrou adequada às variáveis separando-as em três fatores, cada fator coerente com as explicações de especialistas, e conseguiu capturar 84,36

O modelo de série temporal univariado mostrou bom ajuste aos dados de DP's e identificou características relevantes sobre o comportamento da série, como o parâmetro auto-regressivo de ordem 8, pois se conhecia a influência de uma hora atrás, no entanto o modelo mostrou que há uma influencia a cada 24 horas.

Os scores dos fatores gerados pela AF foram utilizados como variáveis de entrada para o modelo de função de transferência. Este modelo apresentou melhor ajuste em relação ao modelo univariado. Observou-se a grande quantidade de parâmetros neste modelo, supõe-se ser devido à grande variabilidade da série de DP's.

Para avaliar se haviam influências externas nos dados de DP's foram analisadas as manobras do sistema, representadas por abertura e fechamento de disjuntores e seccionadoras, e comutações do autotransformador. Esta análise indicou que as manobras de disjuntores e seccionadoras no lado de alta tensão (no lado do autotransformador que recebe a energia gerada em Tucuruí) influenciam em valores de DP's acima de 345 pC. Tal análise se mostrou importante para os especialistas, já que havia uma suposição de que estas manobras afetavam o sistema de medição de DP, porém não se sabia ao certo o quanto. No entanto, estudos mais específicos e testes devem ser realizados para conhecimentos conclusivos.

4.1 Recomendações

Recomenda-se para trabalhos futuros:

- i) Desenvolver outras metodologias para o ajuste da série temporal, que represente a alta variabilidade das séries, como os modelos da família ARCH, ou seja, modelos Autoregressivos com heterocedáticidade condicional generalizado GARCH multivariados, ou modelos ajustados por Redes Neurais Artificiais;
- ii) E, devido a ocorrência de modelos tipo feedback, talvez sejam obtidos melhores resultados com a aplicação aos dados de modelos da classe dos modelos Autoregressivo vetorial VAR;
- *iii*) Aumentar o período de coleta de dados e equipamentos para avaliar a relação das DP's com as manobras dos equipamentos da subestação.

Apêndice A

Tabelas

Neste apêndice serão apresentadas todas as tabelas referentes aos resultados da AF, das análises de séries temporais e das análises de contingência.

A.1 Tabelas referentes à análise fatorial e análises de séries temporais

Para avaliar o comportamento das séries de DP foram coletadas variáveis do sistema e ambientais. denominadas de variáveis de controle. relacionadas abaixo:

- Corrente da fase de 230 kV CB230;
- Corrente da fase de 500 kV CB500;
- Potência Ativa da fase de 230 kV PA230;
- Potência Ativa da fase de 500 kV PA500;
- Potência Reativa da fase de 230 kV PR230;
- Potência Reativa da fase de 500 kV PR500;
- Temperatura do Óleo TO;
- Temperatura de Enrolamento da fase de 13 kV TEB13;
- Temperatura de Enrolamento da fase de 230 kV TEB230;
- Temperatura de Enrolamento da fase de 500 kV TEB500;
- Tensão entre as fases A e B de 230 kV T_AB230;
- $\bullet\,$ Tensão entre as fases A e B de 500 kV T_AB500;
- Tensão entre as fases B e V de 230 kV T_BV230;

- Tensão entre as fases B e V de 500 kV T_BV500;
- Tensão entre as fases A e V de 230 kV T_AV230;
- Tensão entre as fases A e V de 500 kV T_AV500;
- Temperatura de Ar TA;
- Umidade Relativa do Ar UR.

Na Tabela A.1 abaixo estão as principais estatísticas descritivas das variáveis de controle listadas acima.

	Dados	Mínimo	Máximo	Média	Desvio Padrão	Assimetria	Curtose
CB230	1152	873,50	1268,70	1000,76	$54,\!83$	$0,42^{*}$	0,67**
CB500	1152	384,76	$571,\!94$	$454,\!20$	31,92	0,31*	-0,33**
PA230	1152	$353,\!00$	$496,\!68$	$410,\!85$	$22,\!47$	0,16*	-0,26**
PA500	1152	356,70	$521,\!92$	$416,\!32$	$23,\!18$	$0,29^{*}$	$0,\!27^{**}$
PR230	1152	-59,75	$73,\!05$	$7,\!82$	$22,\!65$	$0,03^{*}$	-0,33**
PR500	1152	-52,22	101,78	$17,\!39$	$25,\!68$	$0,70^{*}$	-0,40**
TOB	1152	$34,\!230$	$49,\!350$	$42,\!24$	2,75	-0,02*	-1,02**
TEB13	1152	$35,\!09$	48,70	$42,\!13$	$2,\!80$	$0,03^{*}$	$-1,10^{**}$
TEB230	1152	$36,\!78$	$50,\!65$	43,77	2,85	$0,02^{*}$	$-1,07^{**}$
TEB500	1152	$38,\!63$	$55,\!14$	$46,\!62$	$3,\!13$	$0,02^{*}$	-1,01**
T_AB230	1152	$229,\!89$	$239,\!07$	$234,\!12$	1,07	-0,57*	$0,81^{**}$
T_AB500	1152	$501,\!59$	$546,\!68$	$527,\!22$	8,76	-0,14*	-0,69**
T_BV230	1152	$229,\!61$	$239,\!07$	$234,\!32$	1,06	-0,64*	$0,96^{**}$
T_BV500	1152	$500,\!96$	$547,\!93$	$527,\!20$	8,84	-0,15*	-0,66**
T_VA230	1152	$233,\!78$	$243,\!52$	$238,\!66$	$1,\!14$	-0,60*	$0,95^{**}$
T_VA500	1152	$505,\!35$	$551,\!06$	$531,\!76$	8,71	-0,17*	-0,68**
TA	1152	$18,\!00$	$36,\!00$	$27,\!99$	$2,\!90$	$0,27^{*}$	-0,98**
UR	1152	$26,\!00$	$99,\!00$	$81,\!63$	$17,\!64$	-0,94*	-0,25**

Tabela A.1 Principais estatísticas das 18 variáveis do sistema.

Erro Padrão de 0,072; **Erro Padrão de 0,144.

A Tabela A.3 apresenta o Teste KMO e de esfericidade de Bartlett.

A Tabela A.4 apresenta as comunalidades extraída para três fatores com rotação Varimax.

A Tabela A.5 apresenta o ajuste do modelo ARIMA para a variável X_1 .

A Tabela A.6 apresenta o estudo das autocorrelações dos resíduos para o modelo da variável X_1 .

A Tabela A.7 apresenta o ajuste do modelo ARIMA para a variável X_2 .

A Tabela A.8 apresenta o estudo das autocorrelações dos resíduos para o modelo da variável X_2 .

A Tabela A.9 apresenta o ajuste do modelo ARIMA para a variável X_3 .

A Tabela A.10 apresenta o estudo das autocorrelações dos resíduos para o modelo da variável X_3 .

A Tabela A.11 apresenta o ajuste do modelo ARIMA para a variável X_4 .

A Tabela A.12 apresenta o estudo das autocorrelações dos resíduos para o modelo da variável X_4 .

A Tabela A.13 apresenta o ajuste do modelo ARIMA para a variável Y.

A Tabela A.14 apresenta o estudo das autocorrelações dos resíduos para o modelo da variável Y.

A Tabela A.15 apresenta o resultado da estruturação para a estimativa do modelo de FT, ou seja, o resultado referente à análise de séries temporais com os parâmetros estimados para o modelo de FT.

A Tabelas A.16 apresenta as autocorrelações dos resíduos para o Modelo FT, equação (3.7).

A Tabelas A.17 apresenta as autocorrelações cruzadas entre os resíduos do Modelo FT e a variável X_1 .

A Tabelas A.18 apresenta as autocorrelações cruzadas entre os resíduos do Modelo FT e a variável X_2 .

A Tabelas A.19 apresenta as autocorrelações cruzadas entre os resíduos do Modelo FT e a variável X_3 .

	T_AB500	T_BV500	T_VA500	PR230	PR500	CB230	CB500	PA230
T_AB500	1,000	0,993	0,997	0,791	0,780	-0,633	-0,792	-0,658
T_BV500	0,993	1,000	0,993	0,792	0,783	-0,623	-0,782	-0,651
T_VA500	0,997	0,993	1,000	0,794	0,781	-0,613	-0,776	-0,636
PR230	0,791	0,792	0,794	1,000	0,866	-0,427	-0,593	-0,474
PR500	0,780	0,783	0,781	0,866	1,000	-0,533	-0,672	-0,559
CB230	-0,633	-0,623	-0,613	-0,427	-0,533	1,000	0,967	0,965
CB500	-0,792	-0,782	-0,776	-0,593	-0,672	0,967	1,000	$0,\!956$
PA230	-0,658	-0,651	-0,636	-0,474	-0,559	0,965	$0,\!956$	1,000
TOB	-0,428	-0,418	-0,413	-0,289	0,326	0,528	$0,\!538$	0,509
TEB13	-0,433	-0,424	-0,419	-0,302	0,342	0,519	0,533	0,503
TEB230	-0,426	-0,416	-0,411	-0,292	$0,\!336$	0,521	0,533	0,505
TEB500	-0,487	-0,478	-0,472	-0,340	$0,\!388$	$0,\!582$	$0,\!600$	0,569
TA	-0,429	-0,414	-0,410	-0,319	0,320	$0,\!481$	0,507	$0,\!491$
UR	$0,\!432$	0,421	$0,\!421$	$0,\!306$	-0,336	-0,412	-0,453	-0,408
T_AB230	$0,029^{*}$	$0,047^{*}$	$0,042^{*}$	0,002*	-0,012*	0,101	$0,\!120$	0,163
T_BV230	0,077	$0,\!094$	$0,\!091$	$0,\!055$	-0,055	$0,\!073$	0,082	$0,\!133$
	TOB	TEB13	TEB230	TEB500	ТА	UR	T_AB230	T_BV230
T_AB500	-0,428	-0,433	-0,426	-0,487	-0,429	0,432	0,029*	0,077
T_BV500	-0,418	-0,424	-0,416	-0,478	-0,414	$0,\!421$	$0,047^{*}$	0,094
T_VA500	-0,413	-0,419	-0,411	-0,472	-0,410	$0,\!421$	$0,042^{*}$	0,091
PR230	-0,289	-0,302	-0,292	-0,340	-0,319	0,306	$0,002^{*}$	$0,\!055$
PR500	0,326	$0,\!342$	$0,\!336$	$0,\!388$	0,320	-0,336	-0,012*	-0,055
CB230	0,528	0,519	$0,\!521$	$0,\!582$	$0,\!481$	-0,412	0,101	0,073
CB500	$0,\!538$	0,533	0,533	$0,\!600$	0,507	-0,453	$0,\!120$	0,082
PA230	0,509	0,503	0,505	$0,\!569$	$0,\!491$	-0,408	0,163	$0,\!133$
PA500	0,518	0,509	0,512	$0,\!573$	$0,\!483$	-0,406	$0,\!170$	$0,\!139$
TOB	1,000	$0,\!970$	0,975	0,969	0,752	-0,662	$0,\!041$	0,024
TEB13	$0,\!970$	1,000	$0,\!985$	$0,\!974$	0,737	-0,646	0,033	0,014
TEB230	0,975	$0,\!985$	1,000	0,974	0,736	-0,645	$0,\!041$	0,023
TEB500	0,969	$0,\!974$	0,974	1,000	0,722	-0,632	$0,\!045$	0,023
TA	0,752	0,737	0,736	0,722	$1,\!000$	-0,839	$0,\!113$	$0,\!102$
UR	-0,662	-0,646	-0,645	-0,632	-0,839	$1,\!000$	-0,032	-0,016
T_AB230	0,041*	$0,033^{*}$	0,041*	0,045*	$0,\!113$	-0,032*	1,000	0,940
T_BV230	$0,024^{*}$	0,014*	$0,023^{*}$	0,023*	0,102	-0,016*	$0,\!940$	1,000

Tabela A.2 Correlações entre as variáveis do sistema.

Valores não significativos para nível de 0,05.

Kaiser-Meyer-Olkin Measure of	0,902T	
Bartlett's Test of Sphericity	Approx. Chi-Square df Sig.	$\begin{array}{r} 49433,\!219 \\ 153 \\ 0,\!000 \end{array}$

Tabela A.3 Teste KMO e de esfericidade de Bartlett.

Tabela A.4 Comunalidades extraída para três fatores com rotação Varimax. Inicial Extração

	Inicial	Extração
T_AB500	1,000	0,899
T_BV500	1,000	0,897
T_VA500	1,000	0,890
PR230	1,000	$0,\!688$
PR500	1,000	0,731
CB230	$1,\!000$	0,744
CB500	1,000	0,904
PA230	$1,\!000$	0,783
PA500	$1,\!000$	0,772
TOB	$1,\!000$	0,943
TEB13	$1,\!000$	0,936
TEB230	1,000	0,939
TEB500	1,000	0,930
TA	$1,\!000$	0,722
UR	$1,\!000$	$0,\!596$
T_AB230	$1,\!000$	0,935
T_BV230	$1,\!000$	0,926
T_VA230	$1,\!000$	0,952

Tabela A.5 Cálculo e Avaliação dos parâmetros da variável X_1

Parâmetros	Estimativas	Erro Padrão	t-valor	approx $P_r > t $	Lag
MA1,1	-0,04818	0,01977	-2,44	0,0150	4
MA1,2	-0,04972	0,02064	-2,41	0,0162	6
MA1,3	-0,17295	0,02180	-7,93	< 0,0001	7
MA1,4	$0,\!65234$	0,02943	$22,\!17$	< 0,0001	8
MA1,5	-0,09621	0,02053	$-4,\!69$	< 0,0001	9
AR1,1	0,30062	0,02977	$10,\!10$	< 0,0001	1
AR1,2	0,09877	0,02997	3,30	0,0010	2
AR2,1	0,94026	$0,\!01482$	$63,\!46$	<0,0001 8	

Lag	χ^2	Grau de Liberdade	$P_r > \chi^2$					Autocori	relações
6		0		0,000	-0,001	-0,001	$0,\!031$	-0,060	0,002
12	$6,\!57$	4	0,1603	$0,\!005$	0,002	0,001	-0,011	-0,031	-0,005
18	10,75	10	$0,\!3773$	0,031	0,011	$0,\!045$	0,007	-0,019	-0,012
24	$12,\!37$	16	0,7183	0,004	-0,014	0,016	0,006	-0,023	-0,020

Tabela A.6 Estudo das autocorrelações dos resíduos para o modelo da variável X_1

Tabela A.7 Cálculo e Avaliação dos parâmetros da variável X₂

Parâmetros	Estimativas	Erro Padrão	t-valor	approx $P_r > t $	Lag
MA1,1	0,98543	0,01136	86,71	10,0001	8
AR1,1	$0,\!69555$	0,02341	29,71	j0,0001	1
AR1,2	-0,12351	$0,\!02500$	-4,94	0,0001	3
AR1,3	0,08360	0,02342	3,57	0,0004	5
AR1,4	$0,\!11988$	0,02261	$5,\!30$	0,0001	8
AR1,5	-0,07937	0,02210	$-3,\!59$	0,0003	16
AR2,1	0,99992	0,0001286	7775,88	10,0001	8

Tabela A.8 Estudo das autocorrelações dos resíduos para o modelo da variável X₂

Lag	χ^2	Grau de Liberdade	$P_r > \chi^2$					Autocori	relações
6		0		0,033	-0,023	0,014	-0,005	0,012	0,011
12	$4,\!68$	5	$0,\!4557$	-0,009	0,012	0,002	-0,015	-0,033	0,021
18	$11,\!65$	11	$0,\!3905$	-0,017	-0,008	-0,051	-0,045	0,001	0,034
24	$22,\!55$	17	0,1644	0,021	0,033	-0,006	0,008	0,041	-0,078

Tabela A.9 Cálculo e Avaliação dos parâmetros da variável X₃

Parâmetros	Estimativas	Erro Padrão	t-valor	approx $P_r > t $	Lag
MA1,1	0,84793	0,01965	43,14	10,0001	8
AR1,1	0,87295	0,02877	$30,\!34$	0,0001	1
AR1,2	-0,20468	0,02845	-7,19	0,0001	2
AR1,3	0,16932	0,02091	8,10	0,0001	7
AR2,1	0,99556	0,0045090	$220,\!80$	i0,0001	8

Lag	χ^2	Grau de Liberdade	$P_r > \chi^2$					Autocori	relações
6	$2,\!01$	1	$0,\!1558$	-0,013	0,032	-0,014	-0,003	-0,017	0,012
12	$7,\!57$	7	$0,\!3719$	-0,034	0,036	-0,009	0,018	-0,042	0,017
18	$10,\!84$	13	$0,\!6238$	-0,047	0,007	0,020	0,010	0,000	-0,007
24	$19,\!29$	19	$0,\!4383$	$0,\!024$	-0,004	-0,019	$0,\!003$	0,064	-0,048

Tabela A.10 Estudo das autocorrelações dos resíduos para o modelo da variável X_3

Tabela A.11 Cálculo e Avaliação dos parâmetros da variável X_4

Parâmetros	Estimativas	Erro Padrão	t-valor	approx $P_r > t $	Lag
AR1,1	0,26092	0,02825	9,24	j0,0001	1
AR1,2	0,06344	0,02863	$2,\!22$	0,0269	6
AR1,3	-0,08542	0,02868	-2,98	0,0030	12
AR1,4	-0,08836	0,02865	-3,08	0,0021	18
AR1,5	0,14564	0,02875	5,07	;0,0001	24
AR2,1	$0,\!14297$	0,02980	4,80	;0,0001	8

Tabela A.12 Estudo das autocorrelações dos resíduos para o modelo da variável X_4

Lag	χ^2	Grau de Liberdade	$P_r > \chi^2$					Autocori	relações
6		0		-0,012	0,014	-0,035	-0,014	-0,009	-0,007
12	$5,\!92$	6	$0,\!4325$	0,017	0,001	0,037	-0,012	-0,033	0,020
18	8,84	12	0,7162	-0,027	-0,039	-0,002	-0,008	0,016	0,000
24	$14,\!29$	18	0,7102	$0,\!015$	-0,036	-0,046	-0,008	-0,026	-0,018

Tabela A.13 Cálculo e Avaliação dos parâmetros da variável Y

Parâmetros	Estimativas	Erro Padrão	t-valor	approx $P_r > t $	Lag
MA1,1	-0,32023	0,02414	-13,27	10,0001	1
MA1,2	-0,18746	0,02241	-8,37	j0,0001	2
MA1,3	-0,04290	0,01825	-2,35	0,0189	3
MA1,4	-0,09673	0,01814	-5,33	i0,0001	6
MA1,5	-0,20885	0,02250	-9,28	i0,0001	7
MA1,6	$0,\!60568$	0,02739	$22,\!12$	i0,0001	8
AR1,1	0,96446	$0,\!01171$	$82,\!38$;0,0001	8

Lag	χ^2	Grau de Liberdade	$P_r > \chi^2$					Autocori	relações
6		0		0,027	-0,027	0,036	0,047	-0,050	0,008
12	$11,\!07$	5	$0,\!0501$	0,002	-0,026	0,031	-0,018	-0,013	-0,007
18	$18,\!02$	11	0,0811	0,062	0,025	0,002	-0,005	-0,039	-0,009
24	$20,\!83$	17	$0,\!2339$	0,008	$0,\!033$	$0,\!027$	-0,012	-0,001	$0,\!020$

Tabela A.14 Estudo das autocorrelações dos resíduos para o modelo da variável Y

Tabela A.15 Cálculo dos parâmetros estimados para o modelo de FT

Parâmetros	Estimativas	Erro Padrão	t-valor	approx $P_r > t $	Lag	Variável	Deslocamento
MA1,1(θ_1)	-0,23904	0,03070	-7,79	0,0001	1	Y	0
MA1,2(θ_2)	-0,05235	0,03081	-1,70	0,0893	2	Y	0
MA2,1(θ_{6})	-0,10018	0,03050	-3,28	0,0010	6	Υ	0
MA3,1(θ_7)	-0,09959	0,03039	-3,28	0,0010	7	Υ	0
MA4,1(θ_8)	-0,14013	0,03048	-4,60	;0,0001	8	Υ	0
MA5,1(θ_9)	-0,10507	0,03081	-3,41	0,0006	9	Υ	0
MA6,1(θ_{15})	-0,07285	0,03075	-2,37	0,0178	15	Υ	0
NUM1(ω_{01})	$0,\!69491$	$0,\!03318$	20,94	0,0001	0	X_1	0
NUM1,1(ω_{11})	-0,13200	0,02686	-4,91	0,0001	7	X_1	0
DEN1,1(δ_{11})	$0,\!29436$	$0,\!02667$	$11,\!04$	0,0001	1	X_1	0
DEN2,1(δ_{21})	$0,\!28750$	$0,\!03675$	$7,\!82$	0,0001	8	X_1	0
NUM2(ω_{02})	6,78742	$7,\!34114$	0,92	$0,\!3552$	0	X_2	0
NUM1,1(ω_{21})	-28,9740	7,73029	-3,75	0,0002	3	X_2	0
DEN1,1(δ_{12})	0,71458	$0,\!10861$	$6,\!58$	0,0001	5	X_2	0
NUM3(ω_{03})	$0,\!05485$	0,20257	0,27	0,7866	0	X_3	0
DEN1,1(δ_{13})	-0,42719	0,00448	-95,2	0,0001	2	X_3	0
DEN1,2(δ_{23})	0,77795	0,00801	$97,\!17$;0,0001	3	X_3	0
NUM4(ω_{04})	13,39776	7,60528	1,76	0,0781	0	X_4	0

Tabela A.16 Estudo das autocorrelações dos resíduos do modelo de FT.

Lag	χ^2	Grau de Liberdade	$P_r > \chi^2$					Autocor	relações
6		0		0,001	0,005	$0,\!025$	$0,\!059$	-0,005	-0,000
12	7,70	5	$0,\!1734$	0,006	0,009	0,002	$0,\!050$	0,000	-0,005
18	$20,\!19$	11	0,0428	$0,\!053$	$0,\!058$	$0,\!005$	0,068	0,009	0,006
24	$29,\!21$	17	$0,\!0526$	0,033	$0,\!016$	-0,001	-0,005	$0,\!016$	$0,\!079$

Lag	χ^2	Grau de Liberdade	$P_r > \chi^2$	Autocorrelaçõe					
5	$4,\!92$	2	$0,\!0853$	0,006	-0,042	-0,028	0,038	-0,006	-0,018
11	$14,\!25$	8	$0,\!0755$	0,038	-0,010	-0,068	0,002	-0,045	-0,014
17	$23,\!19$	14	$0,\!0572$	0,022	-0,020	0,028	-0,077	-0,017	0,009
23	$34,\!09$	20	$0,\!0255$	$0,\!001$	-0,014	-0,005	$0,\!074$	0,063	-0,003

Tabela A.17 Estudo das correlações cruzadas dos resíduos vsX_1 .

Tabela A.18 Estudo das correlações cruzadas dos resíduos vsX_2 .

Lag	χ^2	Grau de Liberdade	$P_r > \chi^2$					Autocori	relações
5	$5,\!41$	3	$0,\!1439$	-0,032	$0,\!005$	-0,008	0,008	0,003	0,061
11	$9,\!63$	9	0,3809	-0,002	0,035	-0,028	$0,\!041$	-0,009	0,002
17	$16,\!97$	15	0,3204	-0,033	-0,045	0,028	$0,\!049$	-0,017	-0,004
23	$18,\!93$	21	$0,\!5895$	-0,020	$0,\!013$	0,008	-0,033	-0,007	-0,001

Tabela A.19 Estudo das correlações cruzadas dos resíduos vsX_3 .

Lag	χ^2	Grau de Liberdade	$P_r > \chi^2$					Autocori	relações
5	$8,\!86$	3	$0,\!0512$	-0,008	0,010	0,068	$0,\!031$	0,027	-0,038
11	$13,\!27$	9	$0,\!1507$	-0,004	-0,007	0,009	-0,041	0,028	-0,036
17	20,75	15	$0,\!1452$	0,013	0,023	0,026	-0,048	-0,021	-0,051
23	$24,\!12$	21	$0,\!2875$	0,022	-0,018	-0,034	-0,016	-0,008	-0,027

Referências Bibliográficas

Barbieri, L., Villa, A. e Malgesini, R. (2012) A step forward in the characterization of the partial discharge phenomenon and the degradation of insulating materials through nonlinear analysis of time series. IEEE Electrical Insulation Magazine, 28, 4, 14 - 21.

Box, G.E.P. e Jenkins, G.M. (1970) Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day. Revised edition, 1976.

Box, G. E. P., Jenkins, G. M e Reinsel, G. C. (2008) Time Series Analysis: Forecasting and Control. 4^a Ed. San Francisco: Wiley.

Cuenca, W. M. H. (1998) Aplicação de Sistemas Inteligentes no Reconhecimento de Padrões de DPs em Transformadores de Potência. Dissertação de Mestrado, Programa de Engenharia Elétrica, COPPE/UFRJ, Rio de Janeiro, RJ.

Cuenca, W. M. H. (2005) Caracterização dos Sinais de Descargas Parciais em Equipamentos de Alta Tensão a Partir dos Modelos Experimentais. Tese de Doutorado, Programa de Engenharia Elétrica, COPPE/UFRJ, Rio de Janeiro, RJ.

Everitt, B. S. (1992) The analysis of contingency table. 2^a ed. London: Chapman and Hall.

Fávero, L. P. [et al.] (2009) Análise de Dados: modelagem multivariada para tomada de decisões, 1^a Edição, Editora - Rio de Janeiro : Elsevier.

Filho, V. S. (2000) Imagens de Defeitos por Descarga Parcial Estimulada por Raio-X Pulsado em Materiais Dielétricos Poliméricos, Tese de doutorado, Universidade federal do Paraná.

Hair, J. F. Jr., Anderson, R. E. e Tatham, R. L. (1998) Multivariate Data Analysis. 5^a ed. New Jersey: Prentice Hall.

IEC 60270 (200) High-voltage test techniques - Partial discharge measurements, 3^a ed.

James, R. E. e Jones, S. L. (1988) Some Aspects of the Statistical Modeling of Partial Discharge Inception Conditions, IEEE Transactions on Electrical Insulation, 23, 2, 297 - 306.

James, R. E. e Phung, B. T. (1995) Development of Computer-based Measurements and their Application to PD Pattern Analysis, IEEE Transactions on Electrical Insulation, 2, 838-856.

Jeyabalan, V. e Usa, S. (2011) Statistical Techniques for Partial-Discharge Location in Transformer Windings, IEEE Transactions on Power Delivery, 26, 3, 2064 - 2065.

Johnson, R. A. e Wichern, D. W. (1992) Applied Multivariate Statistical Analysis. Prentice Hall.

Ledolter, J. e Bisgaard, S. (2010) Challenges in constructing time series models from process data, Quality and Reliability Engineering International, 27, 165 - 178.

Levy, A. F. S., Carvalho, A. T., Junior, H. P. A., Filho, O. O. e Nascimento, L. F. A. (2006) Desenvolvimento e Implementação de um Sistema Integrado de Medição de descargas Parciais e Fator de Potência On-Line em Transformadores de Potência na ELETRONORTE, apresentado no International Workshop On Power Transformers, Recife - Brazil.

Liao, R.-j.; Yang, L.-j.; Li, J.; Grzybowski, S. (2011) Aging Condition Assessment of Transformer Oil-paper Insulation Model based on Partial Discharge Analysis. IEEE Transactions on Dielectrics and Electrical Insulation, 18, 1, 303 - 311.

Mazroua, A. A., Bartinikas, R., Salama, M. M. (1994) Discrimination Between PD Pulse Shapes Using Different Neural Network Paradigms, IEEE Trasactions on Electrical Insulation, 1, 6, 1119 - 1131.

Morettin, P. A., Toloi, C. M. C. (2004) Análise de Séries Temporais, Edgard Blücher: São Paulo. Rahman, M. K. A., Arora, R. e Srivastava, S. C. (2000) Partial Discharge Classification Using Principal Component Transformation, IEEE Proc., Part A, v. 147, pp. 7-13, 2000.

Salama, M. M. e Bartinikas, R. (2002) Determination of Neural Network Topology for Partial Discharge Pulse Pattern Recognition, IEEE Trasactions Neural Networks, 13, 446 - 456.

Wei, W. S. (1989) Time Series Analysis: Univariate and Multivariate, Addison-Wesley, 1989.