

Gilmar Rosa

Robust Estimation of Functional Model
Three Dimensional (3D) for Identification
Faults in Transmission Lines

**Belo Horizonte
2016**

Gilmar Rosa

Estimação Robusta de Modelo Funcional Tridimensional (3D), para Identificação de Faltas em Linhas de Transmissão

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Minas Gerais, para a obtenção de Título de Doutor em Engenharia de Produção, na Linha de Pesquisa Modelagem Estocástica e Simulação.

Orientador: Prof. Dr. Marcelo Azevedo Costa.

**Belo Horizonte
2016**

R788e

Rosa, Gilmar.

Estimação robusta de modelo funcional tridimensional (3D), para identificação de faltas em linhas de transmissão [manuscrito] / Gilmar Rosa. – 2016.

Ix, 109 f., enc.: il.

Orientador: Marcelo Azevedo Costa.

Tese (doutorado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Apêndices: f. 77-109.

Bibliografia: f. 71-76.

1. Engenharia de produção - Teses. 2. Análise funcional - Teses.
3. Linhas de telecomunicação - Teses. I. Costa, Marcelo Azevedo.
II. Universidade Federal de Minas Gerais. Escola de Engenharia.
III. Título.

CDU: 658.5(043)



FOLHA DE APROVAÇÃO


Robust Estimation of Functional Model Three Dimensional (3D) for Identification Faults in Transmission Lines

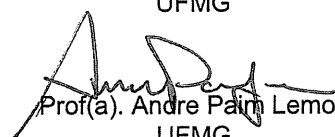
GILMAR ROSA


Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA DE PRODUÇÃO, como requisito para obtenção do grau de Doutor em ENGENHARIA DE PRODUÇÃO, área de concentração PESQUISA OPERACIONAL E ENGENHARIA DE MANUFATURA, linha de pesquisa Modelagem Estocástica e Simulação.

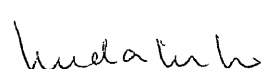
Aprovada em 08 de agosto de 2016, pela banca constituída pelos membros:


Prof(a). Marcelo Azevedo Costa - Orientador
UFMG


Prof(a). Walmir Matos Caminhas
UFMG


Prof(a). Andre Palm Lemos
UFMG


Prof(a). Fernando Luiz Pereira de Oliveira
DEST/UFOP


Prof(a). Linda Lee Ho
Universidade de São Paulo

Belo Horizonte, 8 de agosto de 2016.

Agradecimentos

Agradecimento especial ao *Prof. Dr. Marcelo Azevedo Costa* pela orientação, dedicação, paciência e, principalmente, pela amizade, respeito e consideração durante todo período de desenvolvimento deste trabalho acadêmico, o qual sem a sua presença efetiva não tornaria este trabalho viável.

Agradeço aos meus colaboradores, à Fundação de Amparo à Pesquisa de Minas Gerais - **FAPEMIG**, pelo suporte financeiro a esta pesquisa, aos meus amigos e especialmente à minha família.

Agradeço ao **Departamento de Engenharia de Produção - DEP/UFMG** pela oportunidade de frequentar o programa de *Doutorado em Engenharia de Produção*, desse distinto e conceituado Departamento.

Finalmente agradeço a todos que direta e indiretamente contribuíram para o desenvolvimento deste trabalho.

Resumo

Este trabalho apresenta uma metodologia inovadora de detecção e classificação de faltas em 2D e sua extensão para 3D, ou seja, levando-se em consideração as três fases da linhas de transmissão (LT), baseada na forma funcional da elipse. São apresentados algoritmos de implementação da metodologia, bem como testes em dados reais e simulados. Os resultados foram satisfatórios com índices de exatidão similares aos encontrados na literatura em que se utiliza a *transformada de wavelet* padrão. É proposta construção de limites de confiança elípticos, através de regressão quantílica, os quais são utilizados para monitoramento da linha de transmissão com a finalidade de detecção de faltas. Assim após a detecção da falta, características dos sinais sob falta são extraídas, via distância métrica e utilizadas como entrada em classificadores com a finalidade de determinar o tipo de falta que ocorreu. Dessa forma, a simplicidade da metodologia e sua implementação inteligível são características inerentes, o que pode tornar seu uso bem mais extensivo.

Palavras-chave: Linhas de Transmissão, Análise Funcional, Regressão Quantílica, Classificadores, Wavelet.

Abstract

This thesis presents a novel method for detection and fault classification 2D and its extension to 3D, that is, taking into account the three phases of the transmission lines (LT) based on the functional form of the ellipse. the methodology implementation of Algorithms are presented, as well as tests in real and simulated data. The results were satisfactory indices with similar accuracy to those found in the literature which uses the *wavelet* transform pattern. It is proposed construction elliptical confidence limits through quantile regression, which are used for monitoring the transmission line to the fault detection purpose. Thus upon detection of the fault characteristics of the signals are extracted faulted via distance metric as input and used in binders in order to determine the type of fault that has occurred. Thus, the simplicity of the methodology and implementation are intelligible inherent characteristics, which can make your most extensive use

Keywords: Transmission lines, Functional analysis, Quantile Regression, Classifiers, Wavelet.

Lista de Figuras

| | | |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Sistema elétrico: geração, transmissão e distribuição. | 6 |
| 2.2 | Gerador de Energia elétrica | 6 |
| 2.3 | Sinais de tensão(KV) e corrente(KA). | 7 |
| 2.4 | Defasagem $\varphi=10$ graus | 14 |
| 2.5 | Defasagem $\varphi=30$ graus | 14 |
| 2.6 | Fluxograma do método de Gomes et al. (2013). | 22 |
| 3.1 | Algoritmo para obter os parâmetros da elipse. | 27 |
| 3.2 | CDF da Distribuição Normal Padrão | 28 |
| 3.3 | Valores de $\sum [p_p(y-m)]$ para $m=x$, x pertencente a uma amostra de 1000 observações de uma distribuição uniforme $[0,1]$, com $p=0,25;0,50$ e $0,75$ | 30 |
| 3.4 | Fluxograma dos métodos de estimação em regressão quantílica. | 34 |
| 3.5 | Tensão e corrente após operação de padronização, rotação e translação. | 36 |
| 3.6 | Valores absolutos dos sinais padronizados, rotacionados e transladados. | 36 |
| 3.7 | Fluxograma da detecção de faltas em duas dimensões usando elipses | 36 |
| 3.8 | Sistema de monitoramento <i>online</i> em 2D | 37 |
| 3.9 | Sinais de tensão e corrente antes e depois da falta e ponto da falta na vertical | 38 |
| 3.10 | Região de controle dos sinais antes e depois da falta e o ponto de detecção da falta. | 38 |
| 3.11 | Defasagem $\frac{\pi}{6}$ | 40 |
| 3.12 | Defasagem $\frac{\pi}{3}$ | 40 |
| 3.13 | Defasagem $\frac{4\pi}{9}$ | 40 |
| 3.14 | Defasagem $\frac{5\pi}{9}$ | 40 |
| 3.15 | Projeção 3D | 41 |
| 3.16 | Proposta de algoritmo modificado para correção de Bonferroni | 44 |
| 3.17 | Algoritmo para o monitoramento tridimensional de faltas. | 45 |
| 3.18 | Algoritmo para determinar a menor distância entre o ponto e a elipse estimada | 46 |

| | | |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.19 | Distância do ponto à Elipse estimada. | 47 |
| 3.20 | k- vizinhos mais próximos com k=6 | 51 |
| 3.21 | Exemplo de árvore de decisão | 54 |
| 4.1 | Sinais de corrente e tensão para a linha de transmissão (fases A, B e C) antes e após a falta por queimadas (AQ) - Banco de dados 04aq0201. | 58 |
| 4.2 | Sinais de Tensão e Corrente através do tempo(a), limite de Confiança Gomes et al (2013) (b), limite de Confiança não gaussiano (c) e limite de Confiança por regressão quantílica(d) | 61 |
| 4.3 | Corrente e tensão (a) e Corrente e suas fases A, B e C e os respectivos pontos de falta | 63 |
| 5.1 | Ângulos de defasagem: $\varphi = \frac{2\pi}{3}$, $\varphi = \frac{\pi}{2}$ e $\varphi = \frac{\pi}{3}$, respectivamente | 71 |
| B.1 | Ponto A ortogonal à curva | 81 |
| B.2 | Ponto B externo à curva | 81 |
| B.3 | Comparação de Curvas | 82 |

Lista de Tabelas

| | | |
|------|------------------------------------------------------------------------------------|----|
| 3.1 | Nível de significância global estimado via algoritmo | 44 |
| 4.1 | Base de dados | 56 |
| 4.2 | Análise de Correlação para corrente nas fases A,B e C | 57 |
| 4.3 | Estimativas do fator de potência e valores de pico da tensão e corrente. | 59 |
| 4.4 | Percentual (%) de pontos no interior dos intervalos Construídos (dados simulados). | 60 |
| 4.5 | Aplicação de algoritmo de detecção de faltas 2D em dados reais | 62 |
| 4.6 | Aplicação de algoritmo de detecção de faltas 3D em dados reais | 64 |
| 4.7 | Tabela de Classificação | 64 |
| 4.8 | Matriz de confusão | 65 |
| 4.9 | Percentuais de exatidão e matriz de confusão | 66 |
| 4.10 | Percentuais de exatidão e matriz de confusão | 67 |
| 4.11 | Validação Cruzada - <i>leave-one-out</i> | 68 |
| 4.12 | Validação Cruzada | 68 |
| 4.13 | Percentuais de exatidão dos melhores atributos | 69 |
| A.1 | Análise de Correlação e testes de normalidade - Fase A | 78 |
| A.2 | Análise de Correlação e testes de normalidade - Fase B | 79 |
| A.3 | Análise de Correlação e testes de normalidade - Fase C | 79 |
| C.1 | Descrição dos Bancos de Dados | 86 |

Sumário

| | | |
|----------|----------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 1 | Introdução | 1 |
| 1.1 | Objetivos da tese | 3 |
| 1.2 | Publicações decorrentes | 3 |
| 1.3 | Estrutura/organização da tese | 3 |
| 2 | Sistema elétrico de potência (SEP) e linhas de transmissão (LT). | 5 |
| 2.1 | Sistema elétrico de potência - SEP | 5 |
| 2.2 | Faltas em linhas de transmissão | 7 |
| 2.2.1 | Definição de faltas | 7 |
| 2.3 | Técnicas de detecção e classificação de <i>faltas</i> baseadas em análise de sinais e sistemas inteligentes | 8 |
| 2.4 | Técnicas de detecção e classificação de <i>faltas</i> em LT através da análise funcional e ajuste da elipse (<i>fit ellipse</i>) | 11 |
| 2.4.1 | Revisão bibliográfica relacionada a análise funcional | 11 |
| 2.4.2 | Técnica de detecção e classificação de faltas baseada em análise funcional | 13 |
| 2.4.3 | Técnica de extração dos atributos de sinais elétricos via ajuste da elipse. | 14 |
| 2.4.4 | Rotação e translação de seção cônica – sinais padronizados | 17 |
| 2.4.5 | O modelo funcional estocástico de energia elétrica | 18 |
| 2.4.6 | Classificação de falta através dos atributos extraídos da análise funcional e ajuste da elipse (<i>fit ellipse</i>) | 21 |
| 2.5 | Análise e detecção de sinais elétricos via <i>Transformada Wavelet</i> | 23 |
| 3 | Detecção e classificação de faltas em LT através de análise funcional. | 25 |
| 3.1 | Método de Mínimos quadrados modificado para ajuste da elipse. | 25 |
| 3.2 | Estimação robusta: Modelos de Regressão Quantílica (MRQ). | 27 |
| 3.2.1 | Distribuição acumulada e definição de quantis. | 28 |

| | | |
|----------|---------------------------------------------------------------------------------------------------------------|-----------|
| 3.2.2 | Modelo de regressão quantílica condicional | 30 |
| 3.2.3 | Estimação pontual dos parâmetros dos Modelos de Regressão Quantílica (MRQ) | 31 |
| 3.2.4 | Estimação Intervalar em MRQ | 33 |
| 3.2.5 | Formulação do limite de confiança elíptico via regressão quantílica | 35 |
| 3.3 | Monitoramento estatístico de linhas de transmissão: técnica de detecção de faltas em 2D | 36 |
| 3.4 | Técnica de análise funcional para as três fases da LT (3D) , através do ajuste da elipse | 39 |
| 3.4.1 | Introdução | 39 |
| 3.4.2 | Estimação da corrente de pico I_0 em três dimensões | 40 |
| 3.4.3 | Metodologia em três dimensões - 3D | 40 |
| 3.5 | Nível de significância em limites de Confiança simultâneos | 42 |
| 3.6 | Correção de Bonferroni modificada, aplicada em comparações simultâneas na ausência de independência | 43 |
| 3.7 | Detecção de faltas por limites de Confiança simultâneos-3D | 44 |
| 3.8 | Extração de atributos em três dimensões | 45 |
| 3.8.1 | Distância métrica do ponto a elipse funcional | 45 |
| 3.8.2 | Atributos ou características extraídas dos sinais em 3D | 46 |
| 3.8.3 | Seleção de atributos | 47 |
| 3.8.4 | Índice de exatidão Global - G | 50 |
| 3.8.5 | Classificadores | 50 |
| 4 | Experimentos e resultados | 55 |
| 4.1 | Descrição da base de dados obtida da Companhia Energética de Minas Gerais S.A – CEMIG. | 55 |
| 4.1.1 | Da correlação e teste de normalidade para Corrente e Tensão | 56 |
| 4.1.2 | Da correlação dos resíduos da corrente para as três fases A, B e C | 56 |
| 4.2 | Da maior sensibilidade relativas as faltas da corrente elétrica | 58 |
| 4.3 | Da metodologia 2D melhorada | 58 |
| 4.3.1 | Avaliação da melhoria proposta na metodologia de Gomes et al (2013)-Dados simulados | 58 |
| 4.3.2 | Aplicação da Metodologia 2D melhorada em dados reais: detecção de faltas | 61 |
| 4.4 | Metodologia 3D | 63 |
| 4.4.1 | Detecção de faltas em 3D | 63 |
| 4.5 | Classificação de faltas | 64 |
| 4.5.1 | Definição das Amostras de treinamento e validação | 65 |
| 4.5.2 | Resultados dos classificadores | 66 |

| | | |
|----------|------------------------------------------------------------------------------|------------|
| 4.5.3 | Validação cruzada: <i>leave-one-out</i> | 68 |
| 4.6 | Redução da dimensionalidade dos atributos: busca exaustiva | 68 |
| 5 | Comentários e trabalhos futuros | 70 |
| 5.1 | Comentários finais | 70 |
| 5.2 | Proposta de continuidade. | 70 |
| | Referências Bibliográficas | 72 |
| A | Análise dos resíduos. | 78 |
| A.1 | Correlação e teste de normalidade | 78 |
| B | Demonstrações Matemáticas | 80 |
| B.1 | Comparação do IC proposto por Gomes et al. (2013) e a elipse padrão. | 80 |
| B.2 | Equações para estimação da corrente de pico I_0 em 2D e 3D | 83 |
| C | Banco de Dados | 85 |
| C.1 | Bancos de dados utilizados | 85 |
| D | Artigo decorrente | 87 |
| E | Programas em Software <i>R</i> | 100 |

Lista de Abreviaturas

| | |
|--------------------|----------------------------------------------------------------------------------|
| $\omega = 2\pi f$ | <i>Velocidade angular</i> |
| $\psi(t)$ | <i>mother wavelet</i> |
| φ | <i>Ângulo de defasagem entre os sinais</i> |
| $\vartheta(t)$ | <i>Sinal da tensão ou voltagem no tempo t</i> |
| $i(t)$ | <i>Sinal da corrente ou amperagem no tempo t</i> |
| 2D | <i>Bidimensional ou duas dimensões</i> |
| 3D | <i>Tridimensional ou três dimensões</i> |
| ARL | <i>Average Run Length</i> |
| BayesNet | <i>Modela de uma Rede Bayesiana como um classificador</i> |
| BE | <i>Busca exaustiva</i> |
| BPNN | <i>Back-Propagation Neural Networks</i> |
| BTNB | <i>Classificador que combina uma tabela de decisão (DT) com Naive Bayes (NB)</i> |
| CA | <i>Corrente alternada</i> |
| CART | <i>Classification And Regression Tree</i> |
| CC | <i>Corrente contínua</i> |
| CDF | <i>Função distribuição acumulada</i> |
| CEMIG | <i>Companhia Energética de Minas Gerais S.A</i> |
| CWT | <i>Transformada wavelete contínua</i> |
| D_e | <i>Distância euclidiana</i> |
| DWT | <i>Transformada wavelete discreta</i> |
| EPS | <i>Sistema de Distribuição de Energia elétrica</i> |
| f | <i>Frequência de uma onda</i> |
| FLS | <i>Lógica fuzzy</i> |
| FT | <i>Transformada de fourier</i> |
| G | <i>Índice de exatidão global</i> |
| GA | <i>Algoritmo genético</i> |
| Hz | <i>Hertz</i> |
| I_0 | <i>Corrente de pico</i> |
| I_a, I_b e I_c | <i>Corrente nas fases a,b e c respectivamente</i> |
| i.i.d | <i>Independentes e identicamente distribuídos</i> |
| kNN | <i>Classificador vizinho mais próximo</i> |
| LDA | <i>Linear Discriminant Analysis</i> |
| LLO | <i>Fase-fase aberta (line-line-open)</i> |
| LO | <i>Fase aberta (line-open)</i> |
| LT | <i>Linha de transmissão</i> |
| LTP | <i>Linha de transmissão primária</i> |
| MCMB | <i>Markov Chain Marginal Bootstrap</i> |
| MRA | <i>Transformada Wavelet multi-resolução</i> |
| MRL | <i>Modelo de regressão linear</i> |
| MRQ | <i>Modelo de regressão quantílica</i> |
| PCA | <i>Análise de componentes principais</i> |
| PF | <i>Fator de potência</i> |
| RDP | <i>Registrador Digital de Perturbação</i> |
| RNA | <i>Rede neural artificial</i> |
| RPM | <i>Revoluções por minuto</i> |
| SBS | <i>Busca Seqüencial Regressiva</i> |
| SFS | <i>Busca Seqüencial Progressiva</i> |
| STFT | <i>Short Time Fourier Transform</i> |
| V_0 | <i>Tensão de pico</i> |
| V_a, V_b e V_c | <i>Tensão nas fases a,b e c respectivamente</i> |
| WT | <i>Transformada wavelete</i> |

Capítulo 1

Introdução

Em geral detecção de faltas em algum componente do sistema elétrico de potência (SEP), Oliveira et al (2000), em sua fase inicial é essencial uma vez que tais faltas podem comprometer o fornecimento de energia elétrica a diversos consumidores. Uma falta em um SEP é uma condição anormal que envolve uma falha elétrica em um elemento do sistema, operando em uma das tensões primárias dentro do próprio sistema.

Em um sistema elétrico de potência (SEP), a linha de transmissão (LT) é o elemento mais vulnerável a faltas, uma vez que, devido à sua dimensão física, fica exposta a vários fatores, tais como curtos-circuitos causados por árvores, isoladores danificados, descargas atmosféricas, dentre outros. Para a proteção das linhas de transmissão, o EPS conta com um sistema composto basicamente de disjuntores, transdutores e relés. Atualmente, os relés do sistema de proteção são totalmente digitais, baseados em microprocessadores e têm sua funcionalidade controlada por software. Os relés digitais efetuam várias funções, tais como: proteção, supervisão, transmissão de sinais, dentre outras.

Apesar de todo o desenvolvimento alcançado, detecção de faltas em um SEP não é uma abordagem trivial em linha de transmissão, dados os inúmeros fatores que afetam a eficácia dos diversos métodos existentes, tais como: resistência da falta, distância da falta, ângulo de incidência da falta, ruídos nos sinais, dentre outros. Métodos de detecção de faltas em linhas de transmissão de energia têm o objetivo de detectar desvios dos sinais elétricos a partir do comportamento esperado de tais sinais em condições normais de funcionamento.

As técnicas baseadas em inteligência computacional têm se mostrado as mais eficientes, devido à sua capacidade de aprendizado, generalização e robustez, o que contribui para superar as deficiências dos

outros métodos de proteção de linhas de transmissão. Sinteticamente, são baseadas na utilização de redes neurais artificiais (RNAs), Transformada de Fourier (FT), Wavelets ou uma combinação destas técnicas.

A abordagem, que é objeto desta tese, é modelar, com a maior precisão possível, o comportamento esperado dos sinais elétricos em condições normais de operação. Mesmo sob condições normais, os sinais elétricos estão sujeitos a ruídos aleatórios, dessa forma, tem-se a motivação de se estabelecerem limites superiores e inferiores, com a finalidade de monitorar o comportamento dos sinais elétricos numa rede de transmissão de energia. Quanto maiores forem os limites, mais difícil a detecção de faltas, contrariamente quanto mais estreita forem os limites mais provável para detectar falsas faltas, outra característica importante que o processo estocástico gerador dos dados amostrais não possui uma distribuição gaussiana.

Pode ser mostrado que o comportamento dos sinais de tensão e corrente de uma linha de transmissão sob condições normais ou nominais de funcionamento pode ser representado como uma elipse bidimensional (2D), utilizando-se os sinais elétricos em suas respectivas fases. Estimando-se os limites superior e inferior da elipse, tem-se um gráfico de controle não-linear bidimensional. Portanto, as condições de falha são detectadas se os sinais de corrente e tensão se afastam significativamente das condições operacionais normais, ou seja, fora do intervalo elíptico. Essa abordagem é denominada *análise funcional*.

Na modelagem 3D, sugere-se utilizar a corrente elétrica em suas três fases, neste caso, o monitoramento da rede elétrica é feita através de três intervalos de confiança simultâneos.

Detectada a falta, é possível extrair atributos dos resíduos dos sinais elétricos, utilizando-se a distância euclidiana e utilizar esses atributos como entrada de classificadores a fim de determinar o tipo de falta presente na amostra de dados disponíveis. Essa abordagem se mostrou bastante eficaz com índices de exatidão superiores aos encontrados na literatura. Ademais, sua simplicidade e facilidade computacional podem ser apresentadas como justificativa para o uso desse método.

Este trabalho propõe novas abordagens para estimar os parâmetros do *modelo funcional* e os limites superior e inferior. Esses limites são determinados usando regressão quantílica não-linear, além disso propõe também o uso da tensão em suas três fases. Novos atributos para entrada nos diversos classificadores são apresentados e, por fim, uma comparação da metodologia proposta com a técnica tradicional de análise via *wavelet*.

1.1 Objetivos da tese

Neste trabalho, o foco está na análise robusta de faltas que ocorrem em linhas de transmissão. Basicamente compõe-se da melhoria da proposta de Gomes et al (2013) na identificação e classificação por meio do ajuste de Mínimos Quadrados aplicado à elipse (*Direct Least Squares Fitting of Ellipses*), proposição de intervalos de confiança elípticos robustos para fins de monitoramento do comportamento dos sinais de tensão e corrente (2D) e utilizando a corrente elétrica em três fases (3D), através de métodos de regressão quantílica, considerando que os resíduos dos sinais elétricos indicam não possuir distribuição normal.

Diferentemente de Gomes et al (2013) e dos métodos existentes na literatura (*transformada wavelet*) e *transformada de fourier*, propomos uma metodologia de detecção de faltas em 3D utilizando simultaneamente os três sinais de corrente elétrica. Essa proposta mostrou-se sensível a pequenas perturbações no sinal elétrico, no entanto, tal observação pode ser contornada, aumentando o nível de confiança do intervalo elíptico ou aumentando o número de pontos de controle fora do intervalo.

Por fim, propomos um conjunto de atributos dos sinais de corrente que poderão ser utilizados como entrada nos diversos classificadores, baseado em distâncias métricas, no nosso caso, distância euclidiana, que comparados à proposta de Gomes et al (2013) apresentou melhores resultados, e comparados aos métodos tradicionais apresentou resultados próximos.

1.2 Publicações decorrentes

ROSA, G.; COSTA, M. A.. Estimação robusta de modelo funcional para identificação de faltas em linhas de transmissão, (2015). (Apresentação de Trabalho/Congresso). XIV EMR - Escola de Modelos de Regressão-(Brazilian School of Regression Models).

ROSA, G.; COSTA, M. A.. Robust functional analysis for fault detection in power transmission lines, (2016). Applied Mathematical Modelling (aceito para publicação).

1.3 Estrutura/organização da tese

O presente trabalho está organizado como descrito a seguir. O Capítulo 2 apresenta o sistema elétrico de potência (SEP), linhas de transmissão e tipos de faltas associadas. Neste capítulo apresentamos uma revisão da modelagem funcional dos sinais elétricos e uma revisão da literatura relacionada. O Capítulo 3 apresenta a teoria de base utilizada para extensão da metodologia proposta por Gomes et al (2013), para detecção e classificação de faltas. O Capítulo 4 apresenta os resultados obtidos utilizando-se a

metodologia proposta no capítulo 3 e a comparação com outras metodologias tradicionais. No Capítulo 5, encontram-se os comentários finais e, sugestões para trabalhos futuros. E finalmente, no Apêndice, encontramos as demonstrações matemáticas resultantes da pesquisa, bem como um detalhamento dos bancos de dados de casos reais utilizados, o artigo aceito para publicação e os programas em software **R** utilizados.

Capítulo 2

Sistema elétrico de potência (SEP) e linhas de transmissão (LT).

Neste Capítulo apresentamos o sistema de elétrico de potência (SEP), uma revisão dos trabalhos publicados relacionados ao assunto e uma ampla apresentação de análise de sinais elétricos baseados na forma funcional da elipse, bem como apresentamos a tradicional metodologia de análise de sinais *wavelet*.

2.1 Sistema elétrico de potência - SEP

Um **SEP** é composto de muitos componentes que podem danificar-se ao longo do tempo, seja pelo desgaste natural ou pelo mau uso dos equipamentos. Conforme Figura 2.1 um SEP pode ser dividido em cinco partes: *geração, transmissão, distribuição e consumidores finais*.

Cada componente do sistema elétrico realiza uma função específica e bem definida. A *geração* da energia é obtida através da transformação de uma determinada forma de energia (mecânica, térmica, eólica ou solar) em energia elétrica. Uma Fonte Trifásica é constituída de três fontes iguais defasadas 120° uma da outra. Supondo um rotor¹, Figura 2.2, girando no sentido anti-horário com (*3600 rpm revoluções por minuto*), ($f = 60 \text{ Hz}$), seu campo magnético produz tensão em formas senoidais ilustradas na Figura 2.2. Essa tensão atinge seu valor máximo e mínimo com uma distância de $1/3$ de um período, ou seja, com uma defasagem de 120° , e isto devido ao deslocamento espacial de 120° dos enrolamentos do rotor. Como resultado, visto que as bobinas são iguais (mesma seção e mesmo número de espiras), o alternador produz 3 tensões de mesmo valor com uma defasagem angular de 120° entre elas. Estas tensões atingem seus valores máximos e mínimos com uma distância de $1/3$ de um período, ou seja, com uma defasagem

¹Rotor é tudo que gira em torno de seu próprio eixo produzindo movimentos de rotação

de 120°. Essa energia elétrica produzida é transmitida por meio de *linhas de transmissão*.



Figura 2.1: Sistema elétrico: geração, transmissão e distribuição.

Fonte: <http://www.abradee.com.br>

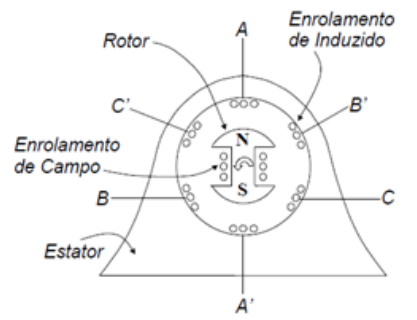


Figura 2.2: Gerador de Energia elétrica

As *linhas de transmissão (LT)* são responsáveis por ligar as grandes usinas de geração às áreas de grande consumo, realizando o transporte de energia elétrica. As estações de *distribuição* alimentam os *consumidores* industriais de médio e pequeno porte, os consumidores comerciais, de serviços e residenciais. Fuchs (1979) diferencia linhas de transmissão (**LT**) e linhas de transmissão primária (**LTP**).

As *linhas de transmissões primárias* são as que operam com tensões suficientemente baixas, para ocupar vias públicas e, suficientemente elevadas, para assegurar boa regulação mesmo para potências razoáveis. Desempenham o papel de conector entre o consumidor final e as estações de distribuições (FUCHS,1979, p. 6). Representa a venda a granel da energia elétrica.

As *linhas de transmissão* são os elementos do sistema elétrico de potência que transportam toda a energia gerada até as estações de distribuição e operam com as tensões mais elevadas do sistema, tendo como função principal o transporte da energia entre os centros de geração e as estações de distribuição. Também têm a função de interligação das estações de geração de energia elétrica. Em geral, apenas poucos consumidores com um alto consumo de energia elétrica são conectados às linhas de transmissão onde predomina a estrutura de linhas aéreas. A linha de transmissão é o elemento do sistema elétrico mais vulnerável, principalmente pela sua dimensão física, pois fica exposta a vários fatores como descargas atmosféricas, queimadas, desgaste natural, etc. Desse modo, a detecção de falha total ou parcial na continuidade do fornecimento de energia elétrica numa **LT** é fundamental para o restabelecimento rápido da sua condição normal de operação, ou seja, o restabelecimento do transporte de energia elétrica.

A transmissão da energia elétrica pode ocorrer em corrente alternada (CA) ou em corrente contínua (CC). Dentre os sistemas em CA, o trifásico tornou-se o mais conveniente, por razões técnicas, econômicas e,

por reduzir a perda de energia no transporte de energia, passou a ser o padrão para a geração, transmissão e distribuição de energia em CA. O sistema trifásico é formado pelas fases A, B e C, com as respectivas tensões V_a , V_b e V_c e correntes induzidas I_a , I_b e I_c . As tensões e correntes apresentam um comportamento senoidal e as fases são defasadas entre si, por um ângulo de 120 graus, tornando o sistema balanceado. A Figura 2.3 mostra os sinais de tensão e corrente de uma linha de transmissão com três fases A, B e C.

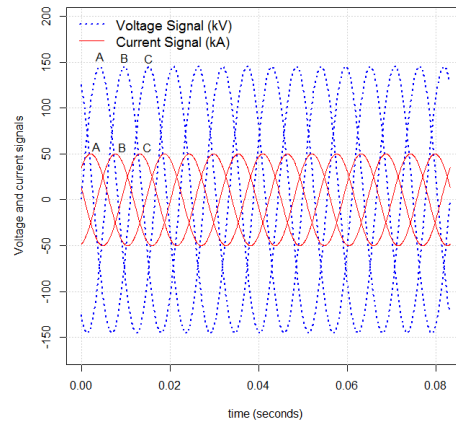


Figura 2.3: Sinais de tensão(KV) e corrente(KA).

2.2 Faltas em linhas de transmissão

2.2.1 Definição de faltas

A falha total ou parcial no fornecimento de energia elétrica numa LT é denominada de *falta*. A ocorrência de uma falta pode ser um fenômeno interno ou externo ao sistema. De acordo com Tleis (2008) dois tipos de faltas podem ocorrer em um sistema: curto-circuito e circuito aberto.

Faltas tipo curto-circuito podem ser causadas por fatores meteorológicos como, descargas atmosféricas, chuva torrenciais, ventos fortes, ou ainda por outros fatores como poluição depositada em isoladores, inundações e incêndios debaixo de linhas de transmissão, contato com árvores, vandalismos, acidentes com veículos, entre outros. Faltas do tipo circuito aberto podem ser causadas por falhas em junções de cabos ou falha em um dispositivo de proteção instalado na linha, como por exemplo, um disjuntor. Em faltas entre fases, a falta ocorre devido a um arco elétrico entre dois condutores de alta-tensão, enquanto que para faltas entre fases e terra, a falta será devido a um arco elétrico entre o condutor de alta-tensão e um objeto aterrado, como a própria torre em si. (INÁCIO, 2010, p. 13, apud Tleis, 2008).

As faltas tipo curto-circuito podem ocorrer entre fases, entre fases e terra ou ambos os casos, podendo ser classificadas como:

- 1) fase-terra (line-ground – LG);
- 2) fase-fase (line-line – LL);

- 3) fase-fase-terra (line-line-ground – LLG);
- 4) trifásico (line-line-line – LLL);

As faltas tipo circuito aberto podem ocorrer em uma ou duas fases simultaneamente. Apesar de possível, o caso de três fases abertas normalmente não é uma falta, mas uma operação de comutação normal do sistema, tal como a abertura das três fases de um disjuntor. Assim, os casos de circuito aberto são normalmente classificados como:

- 1) fase aberta (line-open – LO);
- 2) fase-fase aberta (line-line-open – LLO).

A detecção de faltas em linhas de transmissão é baseada na identificação das mudanças ocorridas nos sinais de tensão e corrente, quando da ocorrência da mesma. Dessa forma, se o espectro dos sinais de falta for extraído, a detecção e diagnóstico podem ser realizados. (MAKMIG, 2002).

Os métodos de análise e diagnóstico de faltas em SEP podem ser classificados de acordo como métodos de detecção, classificação e localização da falta, e basicamente são métodos baseados em análise de sinais e com uso de sistemas inteligentes e conhecimento estatístico.

Os métodos baseados em análise de sinais e sistemas inteligentes utilizam as mudanças de parâmetros da linha de transmissão para a tomada de decisão sobre a ocorrência da falta, tais como, impedância e magnitude dos sinais de corrente e tensão. Normalmente utilizam-se de ferramentas como Transformada de Fourier ou a Transformada Wavelet (WT), para extrair informações de frequência dos sinais de corrente e tensão, para a detecção das faltas. Com estes atributos, diferentes sistemas inteligentes, são capazes de detectar e classificar os diversos tipos de faltas envolvidas. (INÁCIO, 2010).

2.3 Técnicas de detecção e classificação de *faltas* baseadas em análise de sinais e sistemas inteligentes

Diferentes técnicas têm sido usadas para a detecção, classificação e localização de *faltas* em SEP. Upen-
dar, Gupta e Singh (2012) propuseram um algoritmo de detecção e classificação de faltas em LT que primeiramente detecta a falta, usando a transformada wavelet e, posteriormente, utiliza a Regression Tree para classificar a falta detectada. Os autores compararam o algoritmo com a classificação por redes neurais com aprendizado *back-propagation* (BPNN) concluindo, através de simulações de vários cenários, que a técnica de classificação proposta é simples e pode alcançar uma precisão elevada.

Han et al. (2012) descrevem um método bastante simples de detecção e classificação de faltas baseados em transformada de fourier de curto prazo (*Short Time Fourier Transform-STFT*). Basicamente o algoritmo consiste em detectar a falta usando a diferença do cálculo da STFT entre dois pontos (S_1 e S_2), no período de 0,5 segundos. S_1 é definido como STFT de cada tensão A, B e C no estado normal de operação. S_2 é a STFT de cada tensão A, B e C obtido com um atraso (*delay*) de 0,5 segundo. Se essa diferença, ΔS_2 , ou seja, $S_2[n] - S_2[n - 1]$, for maior que determinado limite (*threshold*), diz-se que a falta ocorreu. A magnitude da taxa de variação de S_2 em relação ao valor de S_2 imediatamente anterior, é usado para detectar a falha na LT. Detectada a falta, dois indicadores são ajustados para determinar o tipo de falta, utilizando-se as medidas NS_1 (Máximo normalizado ΔS_1) e NS_2 (Máximo normalizado ΔS_2). Segundo os autores, as simulações mostraram que o algoritmo sugerido pode detectar vários tipos de faltas em linhas de transmissão.

Um importante e abrangente levantamento das técnicas empregadas na detecção e classificação de faltas em LT, e que tem evoluído ao longo da última década, está contido em Sundaravaradan et al. (2014). Esta revisão (*survey*) centra-se principalmente sobre a aplicação da transformada discreta *Transformada Wavelet multi-resolução (MRA)*, juntamente com Redes Neurais Artificiais (*RNA*) e/ou lógica *fuzzy* na análise de faltas em LT.

Jwad e Lefley (2014) apresentam uma avaliação criteriosa da combinação de Transformada Wavelet e técnicas de redes neurais (*RNA*) em linhas de transmissão (*LT*), com o objetivo de detectar e classificar faltas.

Aggarwal et al. (2015) apresentam um estudo de detecção e localização de faltas baseados em redes neurais artificiais (*RNA*) e algoritmo genético (*GA*), usando unicamente a informação da corrente. O método não determina o tipo de falta, apenas fornece a informação da presença da falta e sua localização. Nos estudos realizados em diferentes tipos de falta, os autores observaram que a energia espectral e o pico máximo da corrente são consideravelmente diferentes.

Rafinia e Moshtagh (2014) descrevem um método de localização de faltas em linhas subterrâneas usando a Transformada Wavelet e lógica fuzzy (*FLS*). No seu estudo, utilizou simulações de LT subterrânea com 20 KV, utilizando o software SMTP (*Electromagnetic Transients Program*). Este software permite simular fenômenos transitórios (que duram poucos microsegundos) de natureza eletromagnética, eletromecânica e sistemas de controle associados a sistemas elétricos polifásicos. A falha foi analisada a partir de diferentes cenários das características da corrente e tensão simuladas. A informação tratada no algoritmo de

localização de falhas, segundo o autor, foram muito próximas de situações reais. Concluem que a abordagem proposta, baseada na combinação na transformada wavelet (WT) e lógica fuzzy (FLS), é robusta, segundo resultados obtidos em diferentes estudos de caso. A vantagem da técnica pode ser diretamente atribuída ao fato de que WT extrai de forma eficaz as características cruciais de tempo e frequência de sinais transitórios e a abordagem FLS é capaz fornecer uma precisão muito elevada na classificação e localização de falhas.

Shaik e Pulipaka (2015) utilizam a Transformada Wavelet (WT) com Redes Neurais Artificiais (RNA) para detectar, classificar e localizar a falta em linhas de transmissão. A técnica se baseia na utilização de GPS, para sincronizar as amostras de tensão e corrente de ambas as extremidades da linha de transmissão em análise. Esses sinais são utilizados para o cálculo do índice de falhas. Esses índices são comparados com valores limites para detectar a falha. Redes Neurais Artificiais (RNA) são empregadas para localizar a falta. A técnica se baseia na análise conjunta de sinais transitórios da corrente e a falha associada.

De maneira geral, há uma predominância das técnicas baseadas em Transformada Wavelet (WT) e Transformada Fourier (FT) na extração de atributos ou características dos sinais elétricos. Esses atributos são usados como entradas para os diversos classificadores, especialmente as Redes Neurais Artificiais (RNA), Lógica Fuzzy (FLS) e suas variações.

Embora os métodos baseados em conhecimentos *estatísticos* de detecção, classificação e localização de faltas têm avançado de maneira tímida, pode-se citar os trabalhos de Morales (2015) onde compara-se a Transformada Wavelet com Análise de Componentes Principais (PCA), sob a ótica da classificação de faltas. Samantaray (2009) e Morales et al. (2013) utilizam PCA na classificação de faltas e Yusuffa et al. (2011) propõem um método de localização de faltas baseado em transformada wavelet e métodos de regressão. De maneira geral, modelos estatísticos usando funções trigonométricas simples, em vez de equações matemáticas complexas, têm sido aplicados com sucesso para detectar e classificar faltas em linhas de transmissão.

Extração de atributos dos sinais elétricos, considerando a forma funcional do gráfico bidimensional da tensão e da corrente elétrica como uma elipse (Awan e Ramos, 2010), (Gomes et al., 2013) e (Saribuluti, 2014) tem sido a referida análise bastante estudada devido à sua simplicidade matemática.

2.4 Técnicas de detecção e classificação de *faltas* em LT através da análise funcional e ajuste da elipse (*fit ellipse*)

Alguns estudos de modelagem via ajuste da elipse, *fit ellipse*, existentes na literatura serão apresentados, a seguir.

2.4.1 Revisão bibliográfica relacionada a análise funcional

Ramos et al. (2009) propõem um algoritmo para medir a impedância de processadores de sinais digitais. Esse algoritmo é baseado no ajuste da elipse com a finalidade de extração de atributos da magnitude da impedância e a estimativa da diferença de fases. A metodologia permite trabalhar com um grande número de amostras por processamento não consecutivos de dados, ou seja, segmentados, propiciando um uso reduzido da memória computacional.

Basicamente o algoritmo assume possuir dois sinais no plano XY, denominado de v_1 e v_2 .

$$v_1(t) = D_1 \cos(2\pi ft + \varphi_1) + C_1 \quad (2.1)$$

$$v_2(t) = D_2 \cos(2\pi ft + \varphi_2) + C_2 \quad (2.2)$$

em que D_i , $i \in \{1,2\}$, é a amplitude, C_i a componente estocástica, φ_i é o ângulo de fase e f é a frequência. A diferença do ângulo de fase é dada por: $\Delta\varphi \doteq \varphi_2 - \varphi_1$.

Após manipulações algébricas das Equações (2.1) e (2.2), tem-se a equação quadrática que governa os sinais conjuntamente dada por:

$$\left(\frac{v_1 - C_1}{D_1}\right)^2 + \left(\frac{v_2 - C_2}{D_2}\right)^2 - 2\frac{(v_1 - C_1)(v_2 - C_2)}{D_1 D_2} \cos(\Delta\varphi) - \sin^2(\Delta\varphi) = 0 \quad (2.3)$$

Assim sendo, a Equação (2.3), pode ser reescrita como uma equação geral da elipse:

$$F(v_1, v_2) = av_1^2 + bv_1v_2 + cv_2^2 + dv_1 + ev_2 + f = 0 \quad (2.4)$$

Seguindo a metodologia, definiram-se os seguintes Equações:

a) taxa de amplitude:

$$\frac{D_2}{D_1} = \sqrt{\frac{a}{c}} \quad (2.5)$$

b) As componentes C_1 e C_2

$$\begin{aligned} C_1 &= \frac{2cd-be}{b^2-4ac} \\ C_2 &= \frac{2ae-bd}{b^2-4ac} \end{aligned} \tag{2.6}$$

c) A diferença entre os ângulos de fase.

$$\cos(\Delta\varphi) = \frac{\text{sign}(a)b}{2\sqrt{ac}} \tag{2.7}$$

d) Equação magnitude da impedância $|Z|$.

$$|Z| = \frac{D_2}{D_1} |Z_R| \tag{2.8}$$

e) E finalmente o ângulo de fase φ_z é dado por:

$$\varphi_Z = \Delta\varphi + \varphi_{Z_R} \tag{2.9}$$

Onde Z_R é a fase de referência, φ_{Z_R} é o ângulo de fase de referência e $\text{sign}(x)$ é uma função matemática que, se aplicada aos elementos do vetor x , retorna (1, 0 ou -1) se o elemento do vetor x for positivo, nulo ou negativo, respectivamente.

Awan e Ramos (2010) aperfeiçoaram a metodologia proposta por Ramos et al. (2009), à medida em que introduziram a interpolação da transformada da Fourier (DFT) (Schoukens, Pintelon Van Kamme, 1992), com o objetivo de estimar a frequência do sinal (f), com a finalidade de amenizar os efeitos de ruídos na estimação dos parâmetros da elipse. Ainda acrescentaram um algoritmo recursivo em que não há necessidade do ângulo de fase de referência, (φ_{Z_R}) para o cálculo da variação do ângulo de fase $\Delta\varphi \doteq \varphi_2 - \varphi_1$.

Gomes et al. (2013), utilizando-se da mesma linha de raciocínio, implementaram os estimadores Fator de Potência (PF), representado por $\cos(\varphi)$, de cálculo de tensão de pico (V_0) e de corrente de pico (I_0). Propuseram ainda uma metodologia de detecção e classificação de faltas baseadas na análise funcional e no ajuste da elipse. Essa metodologia será abordada com maiores detalhes na seção 2.4.2,

sendo seu aprimoramento umas das propostas desta tese.

2.4.2 Técnica de detecção e classificação de faltas baseada em análise funcional

A detecção e classificação de faltas, utilizando análise funcional, através do ajuste da elipse descrita nessa seção, é baseada no trabalho de Gomes et al. (2013). Preliminarmente, é descrita a modelagem matemática utilizada pelos autores do comportamento de sinais elétricos de tensão e corrente.

A dinâmica teórica dos sinais de tensão e corrente em uma fase de uma linha de transmissão é representada a partir de funções cosseno na forma:

$$v(t) = V_0 \cos(\omega t) \quad (2.10)$$

$$i(t) = I_0 \cos(\omega t - \varphi) \quad (2.11)$$

Onde (V_0) e (I_0) representam os valores de pico dos sinais de tensão e corrente, respectivamente $\cos(\varphi)$ é o fator de potência (PF), φ é o ângulo de defasagem entre os sinais de tensão e corrente e $\omega = 2\pi f$ é a velocidade angular. Define-se por frequência f a menor parte que não se repete de uma forma de onda periódica como o ciclo. O tempo de duração de um ciclo é o período ou T. A frequência f de uma onda no Sistema Internacional de Unidades (SI) é dada por **hertz** (Hz) e definida pelo inverso do período T.

$$f = \frac{1}{T}$$

No sistema elétrico brasileiro, a frequência utilizada é de **60 Hz**, o que representa um período de onda de aproximadamente 0,0167 segundos. A frequência f e a frequência angular ω estão relacionadas por $\omega = 2\pi f$.

Sem perda de generalidade, os sinais de tensão e corrente podem ser representados a partir de funções do tipo seno, conforme Equações (2.12) e (2.13). Esta é a representação matemática utilizada no desenvolvimento da modelagem funcional.

$$v(t) = V_0 \sin(\omega t) \quad (2.12)$$

$$i(t) = I_0 \sin(\omega t - \varphi) \quad (2.13)$$

De forma alternativa, conhecidos os valores de pico, os sinais de tensão e corrente podem ser padronizados na forma:

$$x(t) = \frac{v(t)}{V_0} = \text{sen}(\omega t) \quad (2.14)$$

$$y(t) = \frac{i(t)}{I_0} = \text{sen}(\omega t - \varphi) \quad (2.15)$$

Os valores instantâneos da tensão e corrente podem ser projetados em um espaço bidimensional. A forma do comportamento bidimensional esperada, dos sinais de tensão e de corrente senoidal para uma fase de uma linha de transmissão em operação normal, (isto é, sem ruído), é fortemente afetada pelo valor do ângulo φ . A variação no fator de potência ($\cos(\varphi)$) do sistema, provoca mudanças na geometria da elipse, alterando o tamanho dos raios e ângulo de rotação da elipse em relação ao eixo das abscissas. A Figura 2.4 e a Figura 2.5 mostram a mudança na forma da elipse, com a mudança do valor do ângulo φ e, consequentemente, do fator de potência, $\cos(\varphi)$.

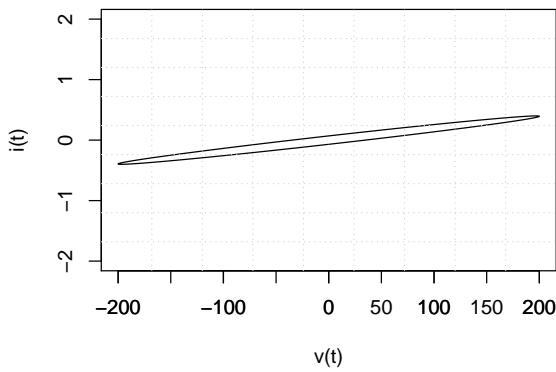


Figura 2.4: Defasagem $\varphi=10$ graus

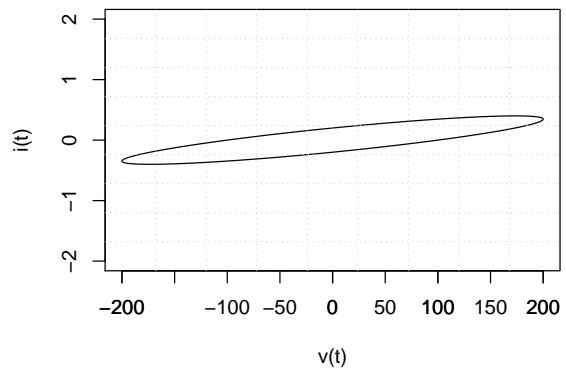


Figura 2.5: Defasagem $\varphi=30$ graus

2.4.3 Técnica de extração dos atributos de sinais elétricos via ajuste da elipse.

Considerando as propriedades trigonométricas elementares, Equações (2.16) e (2.17):

$$\text{sen}^2(\omega t) + \text{cos}^2(\omega t) = 1 \quad (2.16)$$

$$\text{sen}(\omega t + \varphi) = \text{sen}(\omega t)\text{cos}(\varphi) + \text{cos}(\omega t)\text{sen}(\varphi) \quad (2.17)$$

é possível representar a dinâmica dos sinais de tensão e corrente em um espaço bidimensional definido pelos sinais mencionados, a partir de uma equação de uma elipse:

$$V_0^2 \vartheta_{(t)}^2 - 2V_0 I_0 \cos(\omega t) \vartheta_{(t)} i_{(t)} + I_0^2 i_{(t)}^2 - V_0^2 I_0^2 \sin^2(\varphi) = 0 \quad (2.18)$$

As Equações (2.14) e (2.15) representam o comportamento teórico dos sinais de tensão e corrente padronizados, assumindo um fator de potência constante e supondo ausência de eventuais ruídos em condições normais de operação. Em particular, a operação de padronização dos sinais de tensão e corrente permite a comparação desses sinais utilizando a mesma escala unitária de medida, ou seja, ambos os sinais padronizados apresentam valores de pico iguais à unidade.

Na ausência de valores para V_0 , I_0 e $\cos(\varphi)$, os mesmos podem ser estimados a partir de uma amostra dos sinais de tensão e corrente sob condições normais de operação. Gomes et al (2013) propõem estimadores a partir da solução de um sistema de equações lineares. Essas estimativas podem ser calculadas diretamente dos dados, usando-se conhecimentos básicos de equações cônicas, para derivar os estimadores dos valores de pico e de potência, conforme descrito a seguir:

Uma seção cônica em \mathbb{R}^2 é um conjunto de pontos cujas coordenadas em relação à base canônica satisfazem a equação geral, (LAY et al, 2015):

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (2.19)$$

Onde $[a, b, c, d, e, f]$ é o vetor dos coeficientes da seção cônica definida na Equação (2.19).

Das Equações (2.18) e (2.19), tem-se:

$$a = V_0^2, \quad b = -2V_0 I_0 \cos(\omega t), \quad c = I_0^2, \quad d = 0, \quad e = 0, \quad f = V_0^2 I_0^2 \sin^2(\varphi)$$

Pode-se redefinir um novo vetor de coeficientes da seção cônica na forma $[A, B, C, D, E, F]$, onde:

$$A = \frac{a}{-f} = \frac{a}{I_0^2 \sin^2(\varphi)}, \quad B = \frac{b}{-f} = \frac{-2\cos(\varphi)}{V_0 I_0 \sin^2(\varphi)}, \quad C = \frac{c}{-f} = \frac{1}{V_0^2 \sin^2(\varphi)}, \quad D = \frac{d}{-f} = 0, \quad E = \frac{e}{-f} = 0 \quad e$$

$$F = \frac{f}{-f} = -1$$

Gomes et al (2013) propõem reescrever a Equação (2.19) da seguinte forma:

$$AX^2 + BXY + CY^2 + DX + EY = 1 \quad (2.20)$$

A Equação (2.20) pode ser representada na forma matricial $X\beta = Y$, onde:

$$\mathbf{X} = \begin{bmatrix} x_{(t)}^2 & x_{(t)}y_{(t)} & y_{(t)}^2 & x_{(t)} & y_{(t)} \end{bmatrix}, \quad \beta = \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix}, \quad \text{e} \quad \mathbf{Y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix}_{1 \times n}$$

Desta forma, os valores dos coeficientes são estimados da seguinte forma:

$$X^T X \beta = X^T Y$$

$$\beta = (X^T X)^{-1} X^T Y$$

Onde:

$$\beta^T = \begin{bmatrix} A & B & C & D & E \end{bmatrix}$$

Assim, os valores de pico estimados para V_0 , I_0 e fator de potência, $\cos(\varphi)$, são obtidos a partir das expressões:

$$\cos(\varphi) = \frac{-A}{2\sqrt{A}\sqrt{C}} \tag{2.21}$$

$$I_0 = \frac{1}{\sqrt{A \sin^2(\varphi)}} \tag{2.22}$$

$$V_0 = \frac{1}{\sqrt{C \sin^2(\varphi)}} \tag{2.23}$$

Ou seja, uma vez estimados os valores de pico (V_0 e I_0), os sinais poderão ser padronizados, utilizando-se novamente as Equações (2.14) e (2.15).

A estimativa dos coeficientes $[a,b,c,d,e,f]$, nesta tese, foi através do algoritmo *Direct Least Square Fitting of Ellipses*, proposto por Fitzgibbon, Pilu e Fischer (1996) e melhorado por Halíř e Flusser (1998), apresentado com maiores detalhes na Seção 3.1.

A utilização desse método de estimação, proposto por Gomes et al. (2013), não garante que os coeficientes estimados sejam uma elipse. Pode ocorrer que os parâmetros estimados gerem uma cônica

degenerada. Um estudo, por simulação, da qualidade de estimação dos valores de pico V_0 , I_0 e fator de potência ($\cos(\varphi)$), via método proposto por Gomes et al. (2013) e pelo numérico de mínimos quadrados proposto por Pilu, Fitzgibbon e Fischer (1996), é apresentado na Seção 4.3.1.

2.4.4 Rotação e translação de seção cônica – sinais padronizados

A rotação e a translação da elipse têm a finalidade de centrar a cônica na origem dos eixos cartesianos, no ponto (0,0). Assim, após a padronização dos dados originais, é necessário rotacionar e transladar a cônica obtida a fim de que sejam construídos intervalos de confiança. Para rotacionar, o procedimento é simples: bastam conhecimentos básicos de seções cônicas. Partindo desse princípio, a equação geral de uma seção cônica é dada na Equação (2.19) e tem sua forma reduzida quando os termos b , d e e forem iguais a zero, para tal, basta rotacionar por um ângulo $\alpha = 45^\circ$ ou $\alpha = 135^\circ$, a fim de que a elipse obtida fique na posição padrão. Essa operação pode ser realizada conforme indicado na Equação (2.24), a seguir:

$$\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \quad (2.24)$$

Somente a operação de rotação não garante que o centro da elipse estará centrada no ponto de coordenadas (0,0). Para isso torna-se necessário proceder a translação da cônica. Nesse caso a operação é um pouco mais complicada. Assim, realizada a rotação com ângulo α indicado, a seguinte expressão é obtida:

$$a'x'^2 + c'y'^2 + d'x' + e'y' + f' = 0 \quad (2.25)$$

Através da rotação o termo $b'x'y'$ foi eliminado da Equação (2.25). Para obter a equação reduzida da cônica, ainda é necessário eliminar os termos d' e e' , fazendo $d' = 0$ e $e' = 0$. Desta forma, uma translação dos pontos é realizada, obtendo x'' e y'' , podendo reescrever a Equação (2.25) da seguinte forma:

$$a' \left(x' + \frac{d'}{2a'} \right)^2 + c' \left(y' + \frac{e'}{2c'} \right)^2 + f' - a' \left(\frac{d'}{2a'} \right)^2 - c' \left(\frac{e'}{2c'} \right)^2 = 0$$

E assumindo que:

$$x'' = \left(x' + \frac{d'}{2a'} \right), y'' = \left(y' + \frac{e'}{2c'} \right) \text{ e } f'' = f' - a' \left(\frac{d'}{2a'} \right)^2 - c' \left(\frac{e'}{2c'} \right)^2$$

Após a rotação e translação, a equação reduzida da cônica é dada por:

$$a'x''^2 + c'y''^2 + f'' = 0 \quad (2.26)$$

$$a'x''^2 + c'y''^2 = -f'' \text{ ou } \left(-\frac{a'}{f''}x''^2\right) + \left(-\frac{c'}{f''}y''^2\right) = 1$$

fazendo:

$$a_r = -\frac{a'}{f''}x''^2, b_r = -\frac{c'}{f''}, x_r = x'' \text{ e } y_r = y''$$

temos:

$$a_rx_r^2 + b_ry_r^2 = 1 \quad (2.27)$$

2.4.5 O modelo funcional estocástico de energia elétrica

A representação na forma de uma elipse da dinâmica dos sinais de tensão e corrente, descrita nas Seções 2.4.2 e 2.4.3, não considera a existência de ruído, o que é inerente aos processos de geração, transmissão e consumo da energia. Gomes et al. (2013) propõem um modelo funcional que considera duas componentes estocásticas associadas respectivamente aos sinais de tensão e de corrente.

$$\begin{aligned} \vartheta_{(t)} &= V_0 \text{sen}(\omega t) + V_0 \xi_V \\ i_{(t)} &= I_0 \text{sen}(\omega t - \varphi) + I_0 \xi_I \end{aligned} \quad (2.28)$$

Onde ξ_V e ξ_I são duas variáveis aleatórias, independentes, normalmente distribuídas, com média zero e variâncias σ_V^2 e σ_I^2 , respectivamente, $0 \leq (\sigma_V^2, \sigma_I^2) < \infty$. O modelo estocástico assume que as variáveis aleatórias apresentem distribuição normal. De maneira similar ao modelo sem a componente estocástica, no modelo estocástico, as estimativas dos parâmetros de pico e fator de potência também são calculadas utilizando a solução do sistema de equações lineares, conforme Equações (2.21), (2.22) e (2.23).

Para o ajuste do modelo estocástico, é necessário estimar os parâmetros de variância σ_V^2 e σ_I^2 . Segundo Gomes et al. (2013), no cálculo do modelo proposto, é considerado o tempo (t^*) em que se minimiza a distância quadrática. Para tal, é utilizado o método de otimização unidimensional *Golden Search*:

$$t^* = \underset{x}{\operatorname{argmin}} [(x_i - a \sin(\omega t))^2 + (y_i - b \cos(\omega t))^2]$$

Obtido o valor de t^* , é possível obter os resíduos marginais:

$$r_{x_i} = x_i - a \sin(\omega t^*) \quad (2.30)$$

$$r_{y_i} = y_i - b \cos(\omega t^*) \quad (2.31)$$

A partir dos resíduos estimados, os parâmetros de variância são estimados a partir da soma dos quadrados dos resíduos marginais:

$$\hat{\sigma}_V^2 = \frac{\sum_{i=1}^N r_{x_i}^2}{n-1} \quad (2.32)$$

$$\hat{\sigma}_I^2 = \frac{\sum_{i=1}^N r_{y_i}^2}{n-1} \quad (2.33)$$

O objetivo final do cálculo de σ_V^2 e σ_I^2 é a definição de limite superior e limite inferior para o comportamento dos sinais de tensão e corrente em condições normais de operação. A forma funcional desses limites são elipses estimadas a partir dos sinais padronizados e rotacionados, $x_{r(t)}$ e $y_{r(t)}$ conforme Equações (2.34) e (2.35).

a) Limite externo ou limite superior:

$$\frac{x_{r(t)}^2}{(a + z_{\frac{\alpha}{2}} \sigma_V)^2} + \frac{y_{r(t)}^2}{(b + z_{\frac{\alpha}{2}} \sigma_I)^2} = 1 \quad (2.34)$$

b) Limite interno ou limite inferior:

$$\frac{x_{r(t)}^2}{(a - z_{\frac{\alpha}{2}} \sigma_V)^2} + \frac{y_{r(t)}^2}{(b - z_{\frac{\alpha}{2}} \sigma_I)^2} = 1 \quad (2.35)$$

Esta abordagem apresenta algumas limitações pois supõe que os ruídos são independentes. Além disso, a forma funcional da elipse, gerada pelos intervalos de confiança, não são uma elipse em sua forma padrão, a demonstração matemática encontra-se no Apêndice B.1. Como consequência, ao selecionar um valor $z_{\frac{\alpha}{2}} = 3$, que corresponde teoricamente a um intervalo 99,7% de confiança. No nível de confiança real do intervalo, obtido pela metodologia de Gomes et al (2013), os limites de controle fornecem nível de confiança menor do que o especificado, logo haverá uma taxa de detecção de falsas faltas maior que o esperado.

A obtenção do intervalo de confiança menor do que o esperado é uma limitação inerente à metodolo-

gia proposta por Gomes et al. (2013), uma vez que os limites construídos não são uma elipse na sua forma padrão. Ainda assim, apresenta grandes vantagens pois, estimados os parâmetros do modelo, cada par de pontos de sinais de tensão e corrente, x_p e y_p , podem ser rapidamente caracterizados como pertencente ou não aos limites de controle, a partir da verificação das seguintes condições:

$$\text{a) } \frac{x_p^2}{\left(a - z_{\frac{\alpha}{2}} \sigma_V\right)^2} + \frac{y_p^2}{\left(b - z_{\frac{\alpha}{2}} \sigma_I\right)^2} \geq 1 \quad (2.36)$$

$$\text{b) } \frac{x_p^2}{\left(a + z_{\frac{\alpha}{2}} \sigma_V\right)^2} + \frac{y_p^2}{\left(b + z_{\frac{\alpha}{2}} \sigma_I\right)^2} \leq 1 \quad (2.37)$$

onde os pontos x_p e y_p são obtidos a partir da translação e rotação dos dados, após a padronização dos dados $\vartheta_{(t)}$ e $i_{(t)}$. Caso as condições não se verifiquem, o ponto será classificado como uma observação fora das condições normais de operação.

Na construção de intervalos de confiança paramétrico, a suposição de normalidade e independência dos resíduos é extremamente conveniente. Na metodologia proposta por Gomes et al. (2013), além da normalidade e independência, a suposição adicional de variâncias iguais para as componentes estocásticas, σ_V^2 e σ_I^2 , também apresenta grande vantagem para a simplificação do modelo e estimação do parâmetro de dispersão.

Na prática, além de não se verificar a normalidade dos resíduos, os mesmos apresentam correlação que compromete o desempenho do modelo proposto. De maneira alternativa, caso a suposição de normalidade dos resíduos não seja verificada, em Gomes et al. (2013), propõe-se o uso dos percentis dos resíduos para definir os limites inferior e superior dos limites de confiança (limites de controle), conforme descrito a seguir. Vale ressaltar que a metodologia de limites de confiança, via percentis, não foi testada e nem avaliada no artigo original de Gomes et al. (2013), somente sugerida.

a) Limite externo ou limite superior por percentis:

$$\frac{x_{r(t)}^2}{\left(a + r_{x;1-\frac{\alpha}{2}}\right)^2} + \frac{y_{r(t)}^2}{\left(b + r_{y;1-\frac{\alpha}{2}}\right)^2} = 1 \quad (2.38)$$

b) Limite interno ou limite inferior por percentis:

$$\frac{x_{r(t)}^2}{\left(a - r_{x;\frac{\alpha}{2}}\right)^2} + \frac{y_{r(t)}^2}{\left(b - r_{y;\frac{\alpha}{2}}\right)^2} = 1 \quad (2.39)$$

Onde $r_{x;\frac{\alpha}{2}}$ e $r_{x;1-\frac{\alpha}{2}}$, são os percentis de ordem $(\frac{\alpha}{2})$ e $(1 - \frac{\alpha}{2})$, respectivamente.

Gomes et al (2013) propuseram um sistema de monitorimento para uma única fase de uma linha de transmissão, na qual os comportamentos médios dos sinais de tensão e de corrente, em condições normais de funcionamento, são representados como uma elipse bidimensional. Além disso, ao assumir resíduos gaussianos independentes para sinais de tensão e corrente, limites superior e inferior para a elipse são estimados. Esse método pode ser visto como um gráfico de controle não-linear de duas dimensões e que permite a monitorização do sinal eléctrico.

Os gráficos de controle visam determinar se um processo de fabricação está no estado de controle estatístico; isto é, se a média e a variabilidade estão próximas da média e da variabilidade em condições normais de operação. Esses gráficos podem ser estendidos para detectar falhas em linhas de transmissão. Pode-se pensar em gráficos de controle multivariados, tanto para sinais de tensão e corrente. Neste caso, o gráfico de controle T^2 de *Hotelling* é a ferramenta mais usada (Hotelling, 1992 ; Montgomery, 2007). Alternativas robustas à carta de controle T^2 de Hottelling também tem sido proposta em Abu Shawiesh et al (2014).

2.4.6 Classificação de falta através dos atributos extraídos da análise funcional e ajuste da elipse (*fit ellipse*)

De acordo com a análise funcional, após detectada a falta, inicia-se a fase da sua classificação. Novos dados de entrada são padronizados utilizando, como referência, as estimativas dos parâmetros obtidas da amostra coletada anteriormente à falta. Ou seja, usa-se V_0 e I_0 estimados com os dados em condições normais de operação (na ausência da falta). Em seguida, após a padronização, aplicam-se as operações de rotação e translação, nos pontos amostrais obtidos após a falta. Nessas operações, também usam-se os parâmetros estimados pela amostra inicial de tamanho N , assumindo condições de funcionamento normais.

Como padrões de entrada do classificador, foram utilizadas as seguintes estimativas das características dos sinais, após a ocorrência da falta:

1. V_0^f e I_0^f ;
 2. O fator de potência, $\cos(\varphi)^f$ da amostra obtido após a ocorrência da falta;
-

3. Raio ($R_{(x)}^f$) da elipse, raio de maior projeção sobre o eixo das *abcissas*;
4. Raio ($R_{(y)}^f$) da elipse, raio de maior projeção sobre o eixo das *ordenadas*;
5. O ângulo θ^f de inclinação da elipse em relação ao eixo das *abcissas*.

Para classificação, foram testados diversos classificadores, sendo os que apresentaram melhores resultados foram: BTNB, BayesNet e kNN.

BTNB é um classificador híbrido que combina uma tabela de decisão (DT) com Naive Bayes (NB), (PANDA e PATRA, 2009). O algoritmo divide os atributos em dois grupos distintos, no qual um é modelado pelo Naive Bayes e o outro é modelado pela tabela de decisão.

BayesNet corresponde à modelagem de uma Rede Bayesiana como um classificador. Uma Rede Bayesiana é um Grafo Direcionado Acíclico (*Directed Acyclic Graph - DAG*) cujos nós representam variáveis aleatórias.

O kNN baseia-se na classificação a partir da análise da classe predominante entre os k-vizinhos mais próximos desse padrão. Os k-vizinhos mais próximos são definidos a partir de uma função de distância. O padrão é realizar o cálculo via distância Euclidiana.

A Figura 2.6 resume o fluxograma do método de classificação das faltas utilizando a análise funcional.

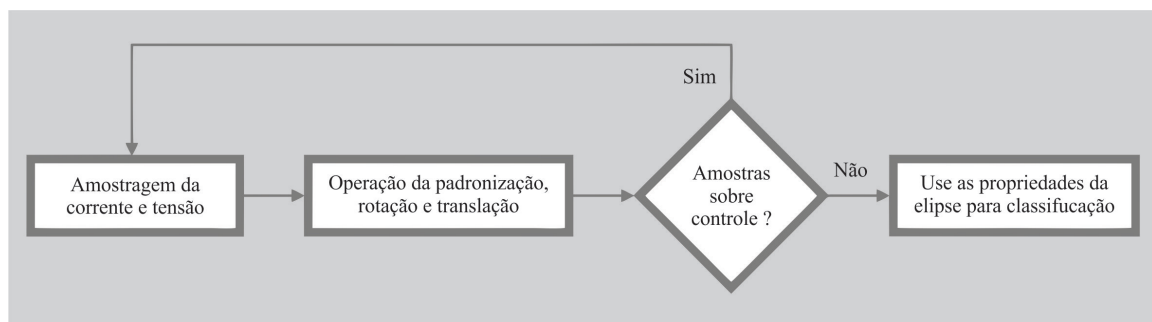


Figura 2.6: Fluxograma do método de Gomes et al. (2013).

2.5 Análise e detecção de sinais elétricos via *Transformada Wavelet*

Recentemente, vários estudos têm sido realizados sobre a detecção e classificação de faltas em linhas de transmissão usando várias técnicas de processamento de sinais discretos, especialmente *wavelet* e *transformada de Fourier*, (HUA et al,2008).

O termo *wavelet* foi cunhado originalmente por Grossmann, A., Morlet, J.,(1984), sendo a base matemática formalizada pelo físico Alex Grossmann. As pesquisas sísmicas estudadas por Morlet exibiam frequências que mudavam rapidamente ao longo do tempo para as quais a transformada de Fourier (FT) não era uma ferramenta adequada para análise.

Segundo Santoso et al (1996), a análise via *wavelet* é uma metodologia para detectar, localizar e investigar a viabilidade de classificar vários tipos de distúrbios que afetam a qualidade do sinal. A idéia fundamental subjacente da abordagem consiste em decompor um determinado sinal de perturbação para os outros sinais que representam uma versão suavizada e detalhada do sinal original. A decomposição é realizada utilizando técnicas de decomposição do sinal. Análise wavelet envolve uma função "*wavelet*" adequada, denominada "*mother wavelet*". Duas versões da transformada wavelet são utilizadas largamente, a contínua (wavelet contínua (CWT)) e a discreta (wavelet discreta (DWT)).

Segundo Malathi, Marimuthu , Baskar (2010), Upendar et all (2012) e Eissa(2013), a *transformada wavelet contínua* (CWT) de um sinal contínuo $x(t)$, é definido como:

$$CWT_{\psi}x(a,b) = \int_{-\infty}^{\infty} x(t) \psi_{a,b}^*(t) dt \quad (2.40)$$

Onde $\psi(t)$ é a *mother wavelet* ou função base e $\psi_{a,b} = \frac{1}{\sqrt{a}} \psi_{a,b}^* \left(\frac{t-b}{a} \right)$ são suas versões dilatadas e transladadas, os parâmetros a e b são dilatação e de translação, respectivamente, com $(a,b) \in \mathfrak{R}, a \neq 0$. O asterisco denota o complexo conjugado. A CWT em diferentes escalas e locais fornece informações da variável tempo-frequência do sinal. A contrapartida digitalmente implementável da CWT é conhecida como transformada wavelet discreta (DWT). A DWT de um sinal $x(t)$ é definida como:

$$DWT(x,m,n) = \frac{1}{\sqrt{a_0^m}} \sum_m \sum_n x(k) \psi^* \left(\frac{k - nb_0 a_0^m}{a_0^m} \right) \quad (2.41)$$

Onde $a = a_0^m$ e $b = nb_0 a_0^m$, são constantes fixas cuja generalização é $a_0 = 2$ e $b_0 = 1$. Notadamente k ,

m e n são variáveis inteiras. A implementação efetiva da DWT é feita através da análise multirresolução (MRA), (Mallat, 1989). O sinal original é analisado em bandas de frequência diferentes, com diferentes resoluções. O sinal é decomposto e suavizado numa versão mais detalhada com repetição do processo. Essa decomposição do sinal original é obtida por meio de aplicação de filtros passa-alto e passa-baixo sucessivas vezes no sinal. As fases sucessivas da decomposição são conhecidas como níveis. Os detalhes da MRA, através dos vários níveis, contêm atributos para a detecção e classificação de defeitos.

A *Transformada de Fourier (FT)*, que é uma das mais simples ferramentas de processamento de sinais discretos, é frequentemente utilizada para propósito de classificação. A metodologia de FT utiliza as mudanças do comportamento do sinal para a classificação de falta. A FT dos sinais sobre todas as frequências apresentadas, não nos fornece qualquer informação sobre o tempo (t) em que as frequências foram coletadas. Daí a superioridade do uso de Wavelet que é uma ferramenta que ajuda na análise do sinal, com informação de tempo, no domínio da frequência.

Capítulo 3

Detecção e classificação de faltas em LT através de análise funcional.

Neste Capítulo, apresentamos a metodologia de estimação dos parâmetros da elipse através de um procedimento de mínimos de quadrados modificado, bem como uma revisão dos métodos de regressão quantílica. Propõem-se também métodos de detecção e classificação de faltas em 2D e 3D, uma metodologia inovadora de estimação da corrente de pico I_0 , utilizando-se as três fases conjuntamente e, em seguida, os métodos de classificação utilizados neste trabalho.

3.1 Método de Mínimos quadrados modificado para ajuste da elipse.

A curva cônica (elipse, parábola ou hipérbole) é definida por um polinômio de segunda ordem:

$$F(\vec{\Psi}, \vec{x}) = ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (3.1)$$

Onde $\vec{\Psi} = (a, b, c, d, e, f)^T$ é um vetor de parâmetros da curva cônica e $\vec{x} = (x^2, xy, y^2, x, y, 1)$, com (x, y) representando a coordenada de um ponto ao longo da curva. $F(\vec{\Psi}, \vec{x}_i) = d$ é a *distância algébrica ou distância euclidiana* de um ponto \vec{x}_i à cônica $F(\vec{\Psi}, \vec{x}) = 0$. Uma maneira de minimizar a distância algébrica utilizando um conjunto N de pontos é através de mínimos quadrados padrão, (PILU, FITZGIBBON e FISHER, 1996).

$$\hat{\vec{\Psi}} = \arg \min_{\vec{\Psi}} \left\{ \sum_{i=1}^N F(\vec{\Psi}, \vec{x}_i)^2 \right\} \quad (3.2)$$

Para assegurar que a solução seja a especificação de uma elipse, a restrição da Equação (3.3) deve ser agregada, conforme indicado por Pilu, Fitzgibbon e Fischer (1996). Essa restrição garante uma solução única, e que os parâmetros estimados sejam uma elipse não degenerada.

$$b^2 - 4ac = 1 \quad (3.3)$$

A solução da Equação (3.2) pode ser obtida por métodos de otimização da seguinte forma:

$$\begin{cases} \hat{\Psi} = \arg \min_{\Psi} \left\{ \sum_{i=1}^N F(\vec{\Psi}, \vec{x}_i)^2 \right\} \\ \text{Sujeito a: } \vec{\Psi}^T C \vec{\Psi} = 1 \end{cases}$$

Onde a matriz de planejamento x_i é dada por:

$$x_i = \begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^2 & x_i y_i & y_i^2 & x_i & y_i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 & x_n y_n & y_n^2 & x_n & y_n & 1 \end{bmatrix}$$

A matriz C, de tamanho 6x6, expressa a restrição $4ac - b^2 = 1$:

$$C = \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Uma solução alternativa para a Equação (3.2) foi implementada por Halř e Flusser (1998) em que, basicamente, utilizam-se os procedimentos de mínimos quadrados, dividindo a matriz x_i , que é uma matriz $n \times 6$, em duas matrizes $n \times 3$. Uma somente com as n linhas e as 3 primeiras colunas, $\vec{x}_1 = (x^2, xy, y^2)$, e a outra matriz com as n linhas e as 3 últimas colunas, $\vec{x}_2 = (x, y, 1)$. De maneira similar, o vetor $\vec{\Psi}$ é

particionado em $\vec{\Psi}_1 = (a,b,c)$ e $\vec{\Psi}_2 = (d,e,f)$.

A matriz C pode ser simplificada como: $C_1 = \begin{bmatrix} 0 & 0 & 2 \\ 0 & -1 & 0 \\ 2 & 0 & 0 \end{bmatrix}$

Em seguida, atravs de procedimentos de lgebra elementar em conjunto com procedimentos numricos, so calculados os parmetros da elipse. A Figura 3.1 apresenta o algoritmo proposto por Halr e Flusser (1998).

 importante frisar que a implementaco computacional do mtodo proposto por Halr e Flusser (1998)  bem simples e seu custo  bem inferior em relato  soluo pelo mtodo de otimizao convencional.

-
-
- Passo 1.** *Obtenha os vetores de coordenadas \vec{x}, \vec{y} que definem a elipse ou parte da elipse;*
- Passo 2.** *Defina \vec{x}_1 vetor de tamanho $3 \times n$ formada por (x^2, xy, y^2) e \vec{x}_2 vetor de tamanho $3 \times n$ formada por $(x, y, 1)$;*
- Passo 3.** *Calcule $S_1 = \vec{x}_1^T \vec{x}_1$; $S_2 = \vec{x}_1^T \vec{x}_2$; $S_3 = \vec{x}_2^T \vec{x}_2$;*
- Passo 4.** *Obtenha o valor de $T = S_3^{-1} S_2^T$;*
- Passo 5.** *Obtenha a matriz M , fazendo $M = S_1 + S_2 T$;*
- Passo 6.** *Aplique a restrio $C_1 = \begin{bmatrix} 0 & 0 & 2 \\ 0 & -1 & 0 \\ 2 & 0 & 0 \end{bmatrix}$, com a seguinte operao $M^{c1} = C_1^{-1} M$;*
- Passo 7.** *Obtenha autovetores da Matriz M^{c1} , onde M^{c1}  uma matriz 3×3 ;*
- Passo 8.** *Calcule a condio de existncia da elipse, $cond = 4ac - b^2 > 0$, da seguinte forma:*
 $cond = 4 \times M^{c1}[1,*] \times M^{c1}[3,*] - (M^{c1}[2,*])^2$;
- Passo 9.** *$\vec{\Psi}_1$ so os autovalores da Matriz M^{c1} em que, $cond > 0$, (calculada em 8), seja satisfeita;*
- Passo 10.** *Obtenha o vetor, $\vec{\Psi}_2 = T \vec{\Psi}_1$;*
- Passo 11.** *Os coeficientes da elipse so: $(a,b,c,d,e,f) = (\vec{\Psi}_1, \vec{\Psi}_2)$.*
-
-

Figura 3.1: Algoritmo para obter os parmetros da elipse.

3.2 Estimaco robusta: Modelos de Regresso Quantlica (MRQ).

A regresso quantlica  uma tcnica utilizada na anlise estatstica, produzindo estimativas da varivel dependente, condicionadas na mediana condicional ou em qualquer outro quantil (percentil) definido. Apesar de ser uma tcnica muito recente, sua origem data da segunda metade do sculo XVIII (HAO e NAIMAN, 2007). O uso de regresso quantlica permite analisar os efeitos de uma varivel sobre a outra em quantis especficos.

3.2.1 Distribuição acumulada e definição de quantis.

Para descrever a distribuição de uma variável aleatória Y , pode-se utilizar a Função de Distribuição Acumulada (CDF), Hao e Naiman(2007). A CDF é uma função F_y , que para cada valor de y , retorna uma proporção da população em $Y \leq y$.

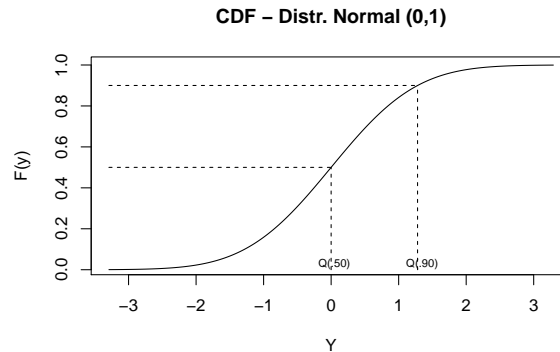


Figura 3.2: CDF da Distribuição Normal Padrão

A Figura 3.2 mostra a CDF de uma distribuição normal padrão, ou seja, $Y \sim N(0,1)$. A CDF pode ser usada para qualquer intervalo de valores de y . A CDF pode ser usada para calcular todas as outras probabilidades envolvendo Y , em particular $P(Y > y) = 1 - F_y(y)$. No caso da Figura 3.2, podemos perceber que $F_y(0) = 0,50$ e, bem como $P(Y > 1,28) = 1 - F_y(1,28) = 1 - 0,90 = 0,10$ e $P(a < Y < b) = F_y(b) - F_y(a)$. Analisando novamente a Figura 3.2, temos: $F_y(1,28) - F_y(0) = 0,90 - 0,50 = 0,40$. As três propriedades que caracterizam uma CDF são as mais importantes para o desenvolvimento da teoria de regressão quantílica:

1. $y_1 \leq y_2 \Rightarrow F_y(y_1) \leq F_y(y_2), \forall y_1, y_2 \in \mathfrak{R}$, isto é F_y é não decrescente (*propriedade da monotonicidade*);
2. $\lim_{x \rightarrow -\infty} F_y(y) = 0$ e $\lim_{x \rightarrow +\infty} F_y(y) = 1$;
3. $\lim_{x \rightarrow x_0^+} F_y(y) = F_y(y_0)$, isto é F_y é contínua a direita.

O quantil pode ser definido de três formas:

1. Pelo valor de m tal que $100p\%$ dos valores populacionais ou amostrais são inferiores a ele, com $0 < p < 1$;
2. Pela CDF de uma variável aleatória Y , em que $F_y = P(Y)$, e utilizando a função inversa da CDF no ponto p , tal que: $F^{-1}(p) = \inf\{y : F(y) \geq p\}$. Logo se têm que $F^{-1}(1/2)$ é a mediana e $F^{-1}(1/4)$ é o primeiro quartil e $F^{-1}(3/4)$ é o terceiro quartil.

3. Finalmente, podemos definir o quantil de ordem p de uma terceira maneira, essencial ao entendimento dos modelos de regresso quantlica, o qual  descrito a seguir.

Considerando o problema de se estimar um valor da varivel aleatria X , com uma distribuio de probabilidade F , com a seguinte funo custo dada por:

$$p_p(u) = \begin{cases} p & \text{se } u > 0 \\ (p-1) & \text{se } u \leq 0 \end{cases}$$

$$= [p - I(u < 0)] \text{ , } 0 < p < 1 \quad (3.7)$$

Em que $p_p(u)$  a funo custo e $I(\bullet)$  a funo indicadora usual.

Considere o problema de estimar m , que minimize a funo custo mdio (KOENKER, 2005, pg. 5-7):

$$E[p_p(y - m)] = (p-1) \int_{-\infty}^m (y - m) dF(y) + p \int_m^{\infty} (y - m) dF(y) \quad (3.8)$$

Diferenciando esta expresso em relao m e igualando a zero temos:

$$(1-p) \int_{-\infty}^m dF(y) - p \int_m^{\infty} dF(y) = F(m) - p = 0 \quad (3.9)$$

Como F  montona, qualquer elemento do conjunto $\{y : F(y) = p\}$ minimiza a funo custo. Quando a soluo  nica $m = F^{-1}(p)$. Em particular, $p = \frac{1}{2}$, temos:

$$E[p_{1/2}(y - m)] = -\frac{1}{2} \int_{-\infty}^m dF(y) + \frac{1}{2} \int_m^{\infty} dF(y) = \frac{1}{2} E[|y - m|] \quad (3.10)$$

Portanto, a minimizao de $E[p_{1/2}(y - m)]$  equivalente  minimizao de $\frac{1}{2} E[|y - m|]$, que resulta em m ser igual  mediana.

A partir destas definies de quantis  desenvolvida a teoria de regresso quantlica, a qual  bastante semelhante  teoria de regresso tradicional. A literatura relacionada  regresso quantlica faz um paralelo entre modelo de regresso linear simples, como forma de desenvolvimento do modelo de regresso quantlica. Nessa linha, suponha o seguinte:

$$\min_{\mu \in \mathbb{R}} = \sum_{i=1}^n (y_i - \mu)^2 \quad (3.11)$$

Logo o valor $\hat{\mu}$, que miniza o somatrio, tendo $\mu(x) = x^T\beta$,  obtido por Mnimos Quadrados. No caso da regresso quantlica, o objetivo  solucionar o seguinte problema:

$$\hat{m} = \min_{m \in \mathfrak{R}} = \sum_{i=1}^n p_p(y_i - p) \quad (3.12)$$

Em particular, quando $p = \frac{1}{2}$, m  exatamente a mediana da amostra. Numa demonstrao simples, considere uma amostra de 1000 observaes com distribuo uniforme $[0,1]$. Calcula-se o valor da Equao (3.12), para cada m igual a cada valor de x presente na amostra para os trs percentis diferentes p . No exemplo, utilizamos os percentis 0,25 0,50 e 0,75. A Figura 3.3, mostra os resultados.

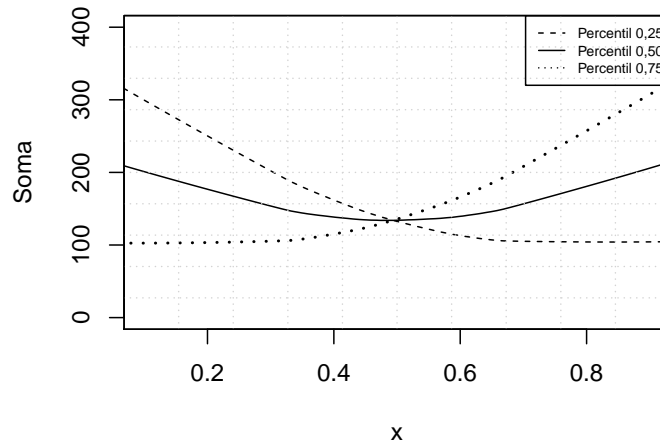


Figura 3.3: Valores de $\sum [p_p(y-m)]$ para $m=x$, x pertencente a uma amostra de 1000 observaes de uma distribuo uniforme $[0,1]$, com $p=0,25;0,50$ e $0,75$.

 fcil demonstrar que para a Distribuo Uniforme no intervalo $[a,b]$, a inversa  dada por $F^{-1}(p) = a + p(b - a)$. E se considerarmos o intervalo $[0,1]$, temos a seguinte relao: $F^{-1}(p) = p$, onde $0 < p < 1$. Logo,  possvel verificar que o valor que torna a soma da Figura 3.3 mnima  $p = 0,50$, ou seja, pode-se concluir que  exatamente a mediana de distribuo uniforme $[0,1]$. Tal concluso pode ser generalizada para outros valores de p , utilizados no exemplo simulado anteriormente.

3.2.2 Modelo de regresso quantlica condicional

De acordo com Hao e Naiman (2007), para uma distribuo assimtrica, a mediana pode tornar-se a mais adequada medida de tendncia central, portanto, a regresso condicional da mediana, em detrimento  regresso condicional da mdia, deve ser considerada para objetivo de modelagem. Regresso condicional

da mediana foi proposta por Boscovich¹ em meados do sculo XVIII e foi posteriormente investigada por Laplace² e Edgeworth³. A regresso da mediana estima o efeito de uma co-varivel na mediana condicional, representando assim a localizao central, mesmo quando a distribuo  assimtrica.

Para modelar tanto a localizao (*percentil*) quanto a forma (assimetria), Koenker e Bassett (1978) propuseram uma forma mais ampla que o modelo de regresso mediana, que so conhecidos como Modelos de Regresso Quantlica (MRQ). O MRQ estima o potencial efeito de uma co-varivel em vrios quantis na distribuo condicional. Desta forma, o MRQ modela o efeito de uma co-varincia sobre a distribuo completa e acomoda a heterocedasticidade. Koenker e Bassett (1978) propuseram o MRQ linear simples da seguinte forma:

Suponha que:

$$y_i^{(p)} = \beta_0^{(p)} + \beta_1 x_i^{(p)} + \epsilon_i^{(p)} \quad (3.13)$$

Em que $0 < p < 1$ indica a proporo da populao que representa valores (escores) abaixo do quantil p .

Nos Modelos de Regresso Linear (**MRL**), a mdia condicional de y_i dado x_i  expresso por: $E(y_i|x_i) = \beta_0 + \beta_1 x_i + \epsilon_i$, isso  equivalente a requerer que os erros tenham esperana zero, ou seja, mdia zero. Em contraste, para o correspondente **MRQ**, especifica-se o **p-simo** quantil condicional de y_i dado x_i como: $E(y_i^{(p)}|x_i^{(p)}) = \beta_0^{(p)} + \beta_1 x_i^{(p)} + \epsilon_i^{(p)}$. Logo o **p-simo** quantil  determinado pelos parmetros especficos, $\beta_0^{(p)}$ e $\beta_1^{(p)}$, e um especfico valor da co-varivel $x_i^{(p)}$.  importante notar que para diferentes valores do quantil p de interesse, os erros $\epsilon_i^{(p)}$ para um valor fixo i so correlacionados.

3.2.3 Estimaco pontual dos parmetros dos Modelos de Regresso Quantlica (MRQ)

Sendo os primeiros defensores do mtodo Erros Absolutos, (Boscovitch , Laplace e Edgeworth) sugeriram mtodos bastante engenhosos para minimizar somas de erros absolutos para problemas de regresso bivariadas at a introduo do algoritmo *simplex* no final de 1940 e a formulao do problema como programao linear. Essa formulao  uma ferramenta indispensvel para compreender o comportamento estatstico. Situaoes que buscam otimizar uma funo linear, sujeita a restrioes lineares so chamadas de **problemas de Programao Linear**.

¹Ruggiero Giuseppe Boscovich,(1711-1787), foi um jesuta, fsico, astrnomo, matemtico, filsofo e diplomata. Nascido na extinta Repblica de Ragusa, posteriormente viveu na Inglaterra, Frana e finalmente na Itlia

²Pierre Simon Marquis de Laplace, (1749-1827), matemtico, astrnomo e fsico francs.

³Francis Ysidro Edgeworth,(1845-1926), economista britnico e Professor da *Universidade de Oxford*

A primeira formulaco bsica do problema de regresso quantlica :

$$\min_{m \in \mathbb{R}} \sum_{i=1}^n p_p \left(y_i - x_i^T \beta^{(p)} \right) \quad (3.14)$$

Koenker (2005) mostra que a estimativa $\hat{\beta}^{(p)}$ pode ser obtida reformulando-se a equaco anterior e transformando-a em um problema de programaco linear, da seguinte forma:

$$\min \{ p \mathbf{1}_n^T \mu + (1-p) p \mathbf{1}_n^T v \mid y - X\beta = \mu - v \} \quad (3.15)$$

$\beta \in \mathbb{R}^p$; $(\mu, v) \in \mathbb{R}_+^{2n}$

Onde: p  o percentil $0 < p < 1$ e $\mathbf{1}_n^T$  um vetor de $1 \times n$ de valores iguais a 1, μ e v so vetores $n \times 1$, e μ_i e v_i , seus respectivos termos.

Estes valores so definidos da seguinte forma:

$$\mu_i = \begin{cases} y_i - \hat{y}_i & \text{se } y_i - \hat{y}_i > 0 \\ 0 & \text{C. C.} \end{cases} \quad v_i = \begin{cases} \hat{y}_i - y_i & \text{se } y_i - \hat{y}_i < 0 \\ 0 & \text{C. C.} \end{cases}$$

Com $\hat{y}_i = x_i^T \hat{\beta}$.

Dessa forma, pode-se enunciar o problema de programaco linear de forma usual:

$$\begin{cases} \min d^T \theta \\ \text{Sujeito a : } B\theta = Y \text{ e } \theta \geq 0 \end{cases}$$

Em que $\theta (\phi^T, \varphi^T, \mu^T, v^T)$, $\phi [\beta]_+$, $\varphi = [-\beta]_+$, v e μ so os mesmos citados anteriormente. $[Z]_+$  a parte negativa de um conjunto z , ou seja, os termos β_{i+} de $[\beta]_+$, que podem ser definidos como:

$$\beta_{i+} = \begin{cases} \beta_i & \text{se } \beta_i > 0 \\ 0 & \text{C. C.} \end{cases}$$

A transposta da matriz \mathbf{B}  definida por:

$$\begin{bmatrix} X \\ -X \\ I_n \\ -I_n \end{bmatrix}$$

Em que I_n é a matriz identidade de ordem n . Além disso, o vetor \mathbf{d} é definido como:

$$\mathbf{d} = (0^T, 0^T, p1_n^T, 1 - p1_n^T)$$

onde $0^T = (0 \ 0 \dots 0)$. Esta formulação (\mathbf{P}) representa um problema de programação linear.

As soluções dos problemas acima mais difundidas são: o algoritmo proposto por Barrodale e Roberts (1973) e os algoritmos de programação linear conhecidos como ponto interior e algoritmo de ponto interior com pré-processamento proposto por Portnoy e Koenker (1997).

O algoritmo proposto por Barrodale e Roberts (1973) é uma adaptação do método simplex para o problema de minimização de desvios absolutos. Possui bom desempenho para bancos de dados com 5.000 observações e 50 variáveis conforme estudos de Chen e Wei(2005), no entanto, perde sua capacidade para banco de dados maiores. Esse algoritmo, adaptado para regressão quantílica, pode ser consultado em Koenker e d'Orey(1987).

No procedimento proposto por Portnoy e Koenker(1997), o qual utiliza o algoritmo de programação linear conhecido como de ponto interior, os autores demonstraram que o procedimento é superior ao método algoritmo simplex, sendo preferível na presença de grandes bancos de dados. Na melhoria do algoritmo ponto interior, os próprios autores propuseram um pré-processamento antes do procedimento. Com essa melhoria, os autores obtiveram, em algumas situações, *performances* semelhantes a do método de mínimos quadrados.

3.2.4 Estimação Intervalar em MRQ

A estimação intervalar em MRQ é realizada de forma linear e não linear.

Estimação intervalar em MRQ linear

A estimação intervalar em MRQ está centrada nos erros padrões dos coeficientes estimados. Existem três técnicas devidamente solidificadas, de estimação de parâmetro: estimação assintótica (*estimação direta*), estimação por métodos de reamostragem (*bootstrap*) e estimação pelo método dos escores ordinais

(*Rank*). A Figura 3.4 mostra o fluxograma da divisão dos métodos para ajuste de limites de confiança. A

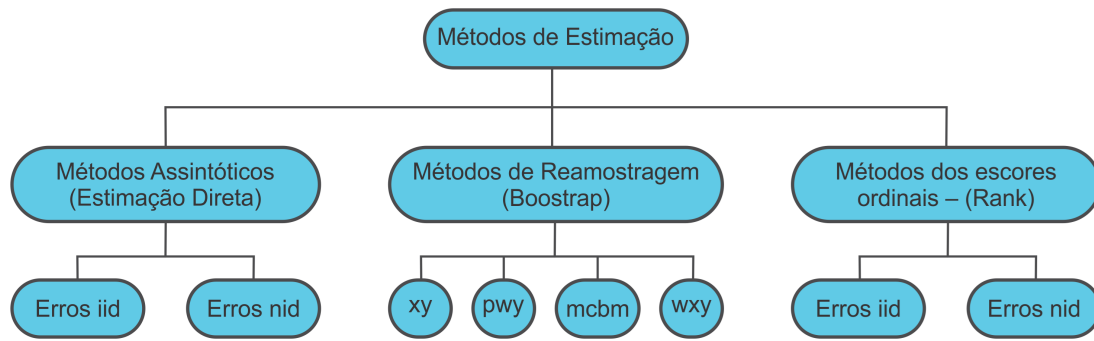


Figura 3.4: Fluxograma dos métodos de estimação em regressão quantílica.

estimação direta, ou estimação assintótica, considera os erros independentes e identicamente distribuídos (*iid*) ou erros não i.i.d (*nid*).

A estimação pelos métodos de reamostragem (bootstrap) é implementada por quatro diferentes maneiras: bootstrap convencional (*xy*), bootstrap por cadeia de markov marginal (*mcbm*), aproximação por Parzer Wei Ying (*pwy*) e bootstrap generalizado (*wxy*).

A estimação pelo método dos escores ordinais (*Rank*) pode ser obtida por duas maneiras: por erros independentes e identicamente distribuídos (*iid*) ou erros não i.i.d (*nid*).

Estimação intervalar para MRQ não linear

Embora haja uma extensa experiência computacional com modelos de regressão quantílica quando os parâmetros são lineares, a experiência de regressão quantílica não linear é muito mais limitada. O comportamento assintótico do estimador de regressão quantílica não linear se aproxima bastante da teoria bem sedimentada para o caso de mínimos quadrados não-linear, e assim, o arcabouço de inferência para regressão linear quantílica não linear pode ser adaptado diretamente de métodos existentes. Nesta seção será considerado o algoritmo de ponto interior para regressão quantílica não linear.

A primeira formulação básica do problema de regressão quantílica é:

$$\min_{q \in \mathbb{R}} \sum_{i=1}^n p_p (y_i - q) \quad (3.16)$$

$$\min_{q \in \mathbb{R}} \frac{1-p}{n} \sum_{y_i < q} |y_i - q| + \frac{p}{n} \sum_{y_i > q} |y_i - q| \quad (3.17)$$

O p -ésimo quantil chamado de q pode ser obtido por métodos de otimização.

De (3.17), podemos derivar:

$$\left(a^{(p)}, b^{(p)}\right) = \arg \min_q \frac{1-p}{n} \sum_{y_{r(i)} < q_i} |y_{r(i)} - q(x_{r(i)}, a, b)| + \frac{p}{n} \sum_{y_{r(i)} > q_i} |y_{r(i)} - q(x_{r(i)}, a, b)| \quad (3.18)$$

Onde: $q(x_{r(i)}, a, b) = \frac{b}{a} \sqrt{a^2 - x_{r(i)}^2}$

3.2.5 Formulação do limite de confiança elíptico via regressão quantílica

Assumindo que os dados amostrais estejam padronizados, rotacionados e transladados, a estimação dos parâmetros das bandas de confiança elíptica é dada por:

$$\frac{x_{r(t)}^2}{\left[a\left(1-\frac{\alpha}{2}\right)\right]^2} + \frac{y_{r(t)}^2}{\left[b\left(1-\frac{\alpha}{2}\right)\right]^2} = 1 \quad (3.19)$$

$$\frac{x_{r(t)}^2}{\left[a\left(\frac{\alpha}{2}\right)\right]^2} + \frac{y_{r(t)}^2}{\left[b\left(\frac{\alpha}{2}\right)\right]^2} = 1$$

Para um $\alpha = 0,05$, temos um percentil p para o limite superior de 0,975 e para o limite inferior de 0,025. Dessa forma têm-se uma banda de confiança de 95%.

Para estimar o vetor de parâmetros $\theta = (a, b)$ da Equação (3.19), via regressão quantílica, considere a Equação (??), reescrita como:

$$\left(a^{(p)}, b^{(p)}\right) = \arg \min_{\theta} \frac{1-p}{n} \sum_{g_i(\theta) < J_i(\theta) \delta_{\theta}} |g_i(\theta) - J_i(\theta) \delta_{\theta}| + \frac{p}{n} \sum_{g_i(\theta) > J_i(\theta) \delta_{\theta}} |g_i(\theta) - J_i(\theta) \delta_{\theta}| \quad (3.20)$$

Onde:

$$\theta = (a, b),$$

$$g_i(\theta) = y_{r(t)},$$

$$J_i(\theta) = J(x_{r(t)}, a, b) = \frac{b}{a} \sqrt{a^2 - x_{r(t)}^2} \quad e,$$

$$\delta_{\theta} \in \Re$$

A partir dos valores estimados, os sinais de tensão e corrente são normalizados e rotacionados, o que gera dados de seguimento de uma equação da elipse reduzida, como mostrado na Equação (2.18). Uma vez que os dados são centrados na origem e igualmente dispersos com respeito aos eixos, consideram-se os valores absolutos dos sinais, como mostrado nas Figuras 3.5 e 3.6 para estimação dos parâmetros.

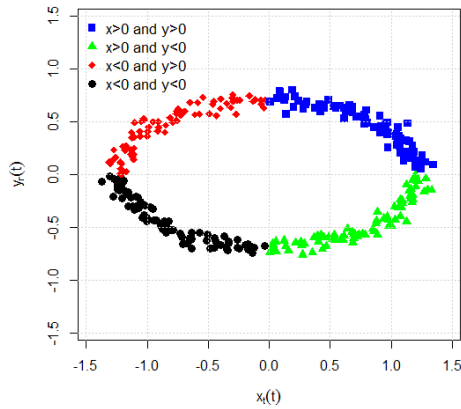


Figura 3.5: Tensão e corrente após operação de padronização, rotação e translação.

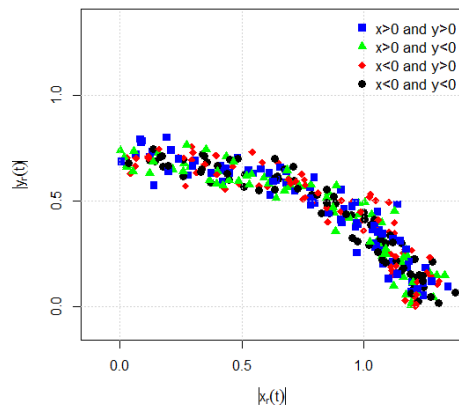


Figura 3.6: Valores absolutos dos sinais padronizados, rotacionados e transladados.

3.3 Monitoramento estatístico de linhas de transmissão: técnica de detecção de faltas em 2D

Sugere-se o uso de um critério de sensibilidade suplementar de três pontos consecutivos que encontram-se fora da região de controle, a fim de detectar verdadeiras faltas e minimizar as taxas de detecção de falsas faltas.

A Figura 3.7 (Flowchart) ilustra o método de detecção de faltas proposto usando as equações da elipse de limite inferior e superior.

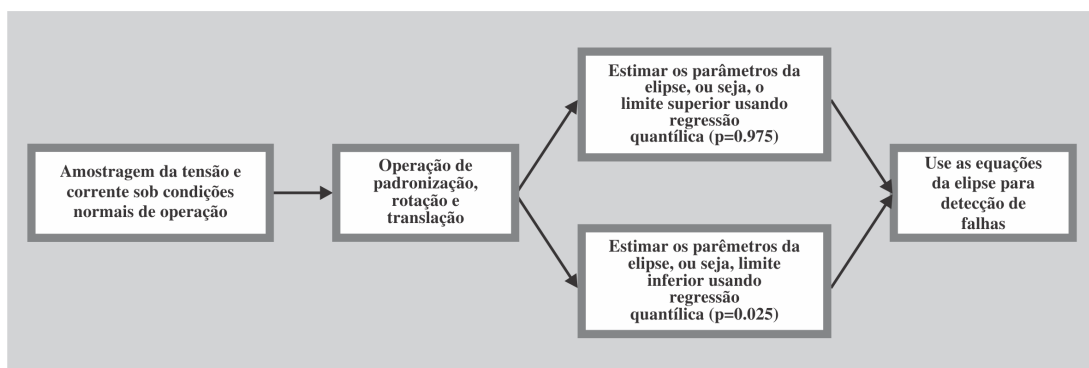


Figura 3.7: Fluxograma da detecção de faltas em duas dimensões usando elipses

O critério de decisão em relação ao ponto pertencer ou não ao limite de confiança, conforme Equações (2.36) e (2.37) podem ser reescritos, usando os procedimentos de regressão quantílica, da seguinte forma:

$$\text{a) } \frac{x_p^2}{[a^{(0.025)}]^2} + \frac{y_p^2}{[b^{(0.025)}]^2} \geq 1 \quad (3.21)$$

$$\text{b) } \frac{x_p^2}{[a^{(0.975)}]^2} + \frac{y_p^2}{[b^{(0.975)}]^2} \leq 1 \quad (3.22)$$

Logo o ponto x_p^2 e y_p^2 pertencerá ao limite de confiança se atender, simultaneamente, os critérios definidos pelas Equações (3.21) e (3.22).

Em adição ao fluxograma apresentado na Figura 3.7, o algoritmo computacional para o sistema de monitoramento *online* em duas dimensões, apresentado na Figura 3.8, a seguir, é uma alternativa para implementação.

Limites de controle

Passo 1. *Obtenha uma amostra de dados de tamanho n dos sinais de corrente e tensão sob condições normais de operação e estime os seguintes parâmetros: V_0 , I_0 e $\cos(\varphi)$*

Passo 2. *Padronize os sinais pelos valores de pico da tensão e corrente, $x_{(t)} = \frac{v_{(t)}}{V_0}$ e $y_{(t)} = \frac{i_{(t)}}{I_0}$*

Passo 3. *Rotacione os sinais padronizados por um ângulo de $\theta = \frac{\pi}{4}$ ou $\theta = \frac{3\pi}{4}$*

Passo 4. *Calcule os limites superiores e inferiores de confiança para $a^{(p)}$ e $b^{(p)}$. Use $p = 0,975$ para o limites superiores $(a^{(0,975)}, b^{(0,975)})$ e $p = 0,025$ para os limites inferiores $(a^{(0,025)}, b^{(0,025)})$, assumindo limite de confiança de 95%.*

Sistema de monitoramento *online*

Passo 5. *Para novos valores dos sinais de tensão e corrente, primeiramente aplique as operações de padronização e rotação usando as estimativas de pico estimadas sob condições normais de operação (passo 1 a 3).*

Passo 6. *Use as equações 3.21 e 3.22, com os parâmetros estimados no passo 4, para testar se os sinais estão no interior dos limites de controle*

Passo 7. *Se três ou mais pontos consecutivos estão fora dos limites de controle, é assumido que o sistema está fora das condições normais de operação.*

Figura 3.8: Sistema de monitoramento *online* em 2D

Para melhor entendimento e como exemplo de aplicação em dados reais, apresentamos a Figura 3.9 em que os sinais de corrente e tensão, antes e depois da falha, além do momento de detecção da falta pelo método. A Figura 3.10 mostra que os sinais de corrente e de tensão estão fora da região de controle, que representa uma condição anormal de operação. Se os critérios de três pontos consecutivos no exterior das regiões de controle é aplicado, então a condição de falta é detectada dentro de um tempo 0,208

milissegundos.

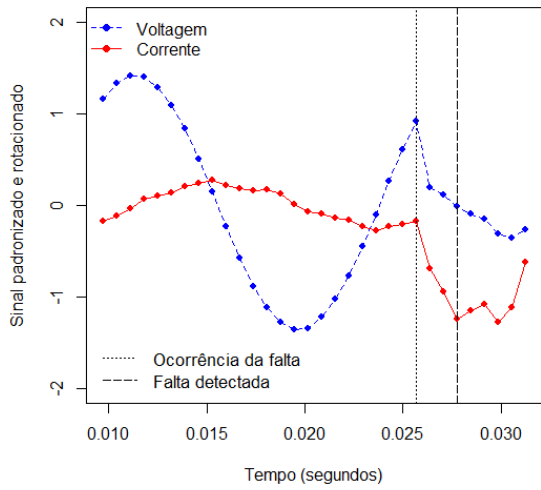


Figura 3.9: Sinais de tensão e corrente antes e depois da falta e ponto da falta na vertical

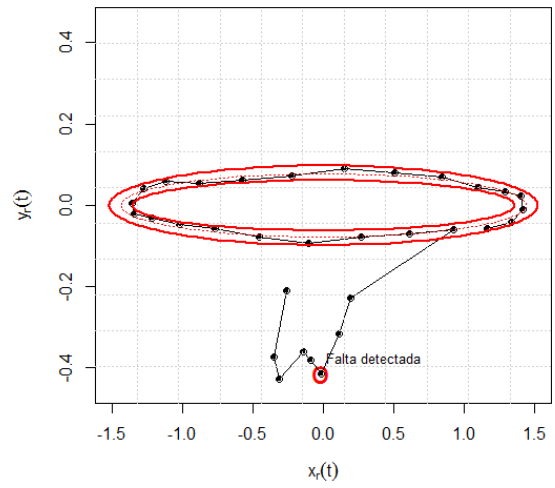


Figura 3.10: Região de controle dos sinais antes e depois da falta e o ponto de detecção da falta.

3.4 Técnica de análise funcional para as três fases da LT (3D) , através do ajuste da elipse

3.4.1 Introdução

O modelo proposto por Gomes et al (2013) tem como finalidade analisar os sinais provenientes de: duas fases de tensão ou duas de corrente ou uma combinação de uma fase da tensão e outra da corrente. É objetivo deste trabalho propor um modelo matemático que leva em consideração, simultaneamente, três sinais elétricos, ou seja, as tensões $\vartheta_{a(t)}$, $\vartheta_{b(t)}$ e $\vartheta_{c(t)}$ ou as correntes induzidas $i_{a(t)}$, $i_{b(t)}$ e $i_{c(t)}$. Para desenvolvimento do modelo em três dimensões, considerando as correntes induzidas, definem-se os seguintes sinais não padronizados:

$$i_{a(t)} = I_0(a) \sin(\omega t) \quad (3.23)$$

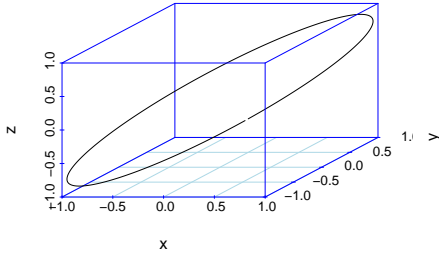
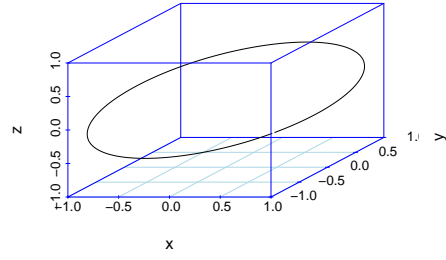
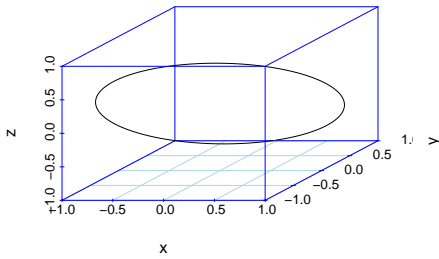
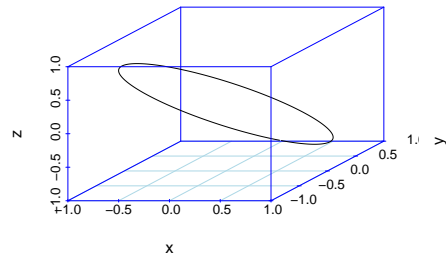
$$i_{b(t)} = I_0(b) \sin\left(\omega t + \frac{2\pi}{3}\right) \quad (3.24)$$

$$i_{c(t)} = I_0(c) \sin\left(\omega t - \frac{2\pi}{3}\right) \quad (3.25)$$

Os valores instantâneos da corrente, projetados em um espaço tridimensional, (3D), perfazem a forma tridimensional dos sinais elétricos.

As Figuras 3.11, 3.12, 3.13 e 3.14 mostram o comportamento da elipse no espaço 3D, com a variação do ângulo de defasagem $\frac{2\pi}{3}$, em diferentes valores.

Analisando o gráfico em três dimensões, percebemos que a inserção da terceira fase na análise funcional provoca uma mudança na direção do eixo (z) da elipse gerada pelos dados. Assim, a análise funcional proposta se baseia na análise funcional das projeções bidimensionais, ou seja, dois-a-dois utilizando uma abordagem aprimorada à proposta de Gomes et al (2013).

Figura 3.11: Defasagem $\frac{\pi}{6}$ Figura 3.12: Defasagem $\frac{\pi}{3}$ Figura 3.13: Defasagem $\frac{4\pi}{9}$ Figura 3.14: Defasagem $\frac{5\pi}{9}$

3.4.2 Estimação da corrente de pico I_0 em três dimensões

A estimação da corrente de pico é uma importante etapa para as análises propostas. Pois é através dela que padronizamos os sinais da corrente. A proposta apresentada por Gomes et al (2013) é limitada à estimação apenas em duas dimensões. Nesta tese propomos a seguir a extensão para três dimensões cuja demonstração completa da derivação desta equação de estimação para 3D é dada no Apêndice B.2. A estimativa de final I_0 , obtida da Equação (3.26), através de procedimentos de otimização, como por exemplo *Golden Search*. Para o caso 3D, a estimativa da corrente de pico, I_0 , considerando conjuntamente as fases A, B e C é dada por:

$$i_{a(t)}^2 - i_{b(t)} \times i_{c(t)} - I_0^2 \sin(\varphi)^2 = 0 \quad (3.26)$$

3.4.3 Metodologia em três dimensões - 3D

Para fins de melhoramento da proposta em duas dimensões, foi introduzido neste trabalho a estimação dos parâmetros da elipse usando o método numérico de mínimos quadrados, proposta por Pilu, Fitzgibbon e Fisher (1996), e melhorado por Halíč e Flusser (1998). Para estimação dos limites de confiança foi, introduzida a regressão quantílica não linear, por ser robusta à presença de *ouliers* e não requerer a

especificação da distribuição estocástica dos ruídos dos sinais elétricos.

Este trabalho também contribui para uma extensão do conceito dado na Seção 3.2.5, para o caso de três dimensões. No caso 3D as *bandas de confiança* são dadas por combinações de duas-a-duas das fases do sinal, da seguinte forma:

$$\left\{ \begin{array}{l} \frac{x_{r(t)}^2}{\left[a \left(1 - \frac{\alpha}{2} \right) \right]^2} + \frac{y_{r(t)}^2}{\left[b \left(1 - \frac{\alpha}{2} \right) \right]^2} = 1 \\ \frac{x_{r(t)}^2}{\left[a \left(\frac{\alpha}{2} \right) \right]^2} + \frac{y_{r(t)}^2}{\left[b \left(\frac{\alpha}{2} \right) \right]^2} = 1 \end{array} \right. \quad (3.27)$$

$$\left\{ \begin{array}{l} \frac{x_{r(t)}^2}{\left[a \left(1 - \frac{\alpha}{2} \right) \right]^2} + \frac{z_{r(t)}^2}{\left[c \left(1 - \frac{\alpha}{2} \right) \right]^2} = 1 \\ \frac{x_{r(t)}^2}{\left[a \left(\frac{\alpha}{2} \right) \right]^2} + \frac{z_{r(t)}^2}{\left[c \left(\frac{\alpha}{2} \right) \right]^2} = 1 \end{array} \right. \quad (3.28)$$

$$\left\{ \begin{array}{l} \frac{y_{r(t)}^2}{\left[b \left(1 - \frac{\alpha}{2} \right) \right]^2} + \frac{z_{r(t)}^2}{\left[c \left(1 - \frac{\alpha}{2} \right) \right]^2} = 1 \\ \frac{y_{r(t)}^2}{\left[b \left(\frac{\alpha}{2} \right) \right]^2} + \frac{z_{r(t)}^2}{\left[c \left(\frac{\alpha}{2} \right) \right]^2} = 1 \end{array} \right. \quad (3.29)$$

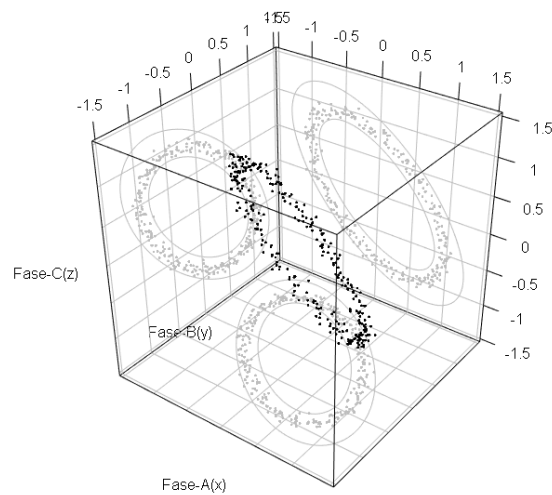


Figura 3.15: Projeção 3D

Uma dificuldade a ser superada é definir um valor do nível de significância (α), que seja comum a todos os limites construídos, de forma a garantir um nível de confiança global desejado. Para o caso de variáveis de aleatórias independentes e identicamente distribuídas, *i.i.d*, têm-se a solução proposta por Bonferroni (1981) que não pode ser aplicada diretamente ao nosso caso pois as fases de sinais elétricos não são independentes, conforme demonstradas anteriormente.

3.5 Nível de significância em limites de Confiança simultâneos

A correção de Bonferroni, Miller (1981), Simens (1986) e Winker, Helmut e Staszewska-Bystrova (2014), é aplicável em qualquer situação onde se deseja efetuar vários testes de hipóteses (limites de confiança) e se pretende controlar o nível de significância/grau de confiança global. Para r limites simultâneos e mesmo nível de significância (α) em cada j -ésimo limite, o nível significância global é:

$$\alpha^* = \alpha \times r. \quad (3.30)$$

Por exemplo, para $r = 3$ comparações, um nível global $\alpha^* = 0,05$ só se garante se cada j -ésimo limite de confiança é construído ao nível de significância de $\alpha^*/3 = 0,016667$.

No caso de r limites de confiança individuais com diferentes níveis de significância, a significância global é dada por $\alpha^* = \sum_{j=1}^r \alpha_j$.

Esse resultado vem diretamente da desigualdade de Bonferroni que, resumidamente, pode ser descrita como:

Seja A_j o evento aleatório, “o j -ésimo limite de confiança populacional”, e $1 - \alpha_j$ a respectiva probabilidade.

Deseja-se estimar a $P \left[\bigcap_{j=1}^r A_j \right]$ que exceda a quantidade global $1 - \alpha^*$, dessa forma pode-se definir:

$$P \left[\bigcap_{j=1}^r A_j \right] = 1 - P \left[\bigcup_{j=1}^r \bar{A}_j \right] \geq 1 - \sum_{j=1}^r P [\bar{A}_j] = 1 - \sum_{j=1}^r \alpha_j \quad (3.31)$$

Desde que se escolha $\sum_{j=1}^r \alpha_j = \alpha^*$, tem-se a garantia de um grau de confiança global de pelo menos $1 - \alpha^*$.

Um inconveniente é que o procedimento de Bonferroni assume independência entre os limites de confiança. No problema tratado nesta pesquisa, os limites de confiança não são independentes.

3.6 Correção de Bonferroni modificada, aplicada em comparações simultâneas na ausência de independência

Devido à quebra da suposição de independência entre os j -ésimos limites de confiança individuais estimados, não se pode aplicar diretamente a correção de Bonferroni dada na Equação (3.30). Dessa forma, é necessário estimar o valor de significância (α_j), comum a cada j -ésimo limite de confiança, que garanta o nível de significância global (α^*).

Assim sendo, nesta tese foi necessário desenvolver uma correção para o procedimento de Bonferroni original, conforme descrevemos a seguir:

Dada uma banda de confiança global, com um nível de significância global α^* associado, o objetivo é determinar o nível de significância α_j em cada j -ésimo limite de confiança construído, de forma a garantir o nível de significância global desejado. Observe que α_j é uma variável aleatória a ser estimada em função do nível de significância global fixado, α^* .

No caso tridimensional, são os pontos padronizados, transladados e rotacionados de coordenada $(x_{r(t)}, y_{r(t)}, z_{r(t)})$, avaliados considerando as combinações duas-a-duas, das bandas de confiança A_1 , A_2 e A_3 . Para estimar a significância global ($\hat{\alpha}^*$), avalia-se, preliminarmente, se o ponto pertence simultaneamente ao j -ésimo limite de confiança estimado, utilizando a condição requerida na Equação (3.32). Como definido anteriormente, o ponto pertence ao interior da banda de confiança, satisfaz-se a regra definida na Equação (3.19). Dessa forma têm-se que:

$$L^{(t)} = \begin{cases} 1 & \text{se } \{(x_{r(t)}, y_{r(t)}) \in A_1\} \cap \{(x_{r(t)}, z_{r(t)}) \in A_2\} \cap \{(y_{r(t)}, z_{r(t)}) \in A_3\} \\ 0 & \text{C. C.} \end{cases} \quad (3.32)$$

Onde $L^{(t)}$ é uma função dicotômica assumindo 1, se o ponto estiver contido simultaneamente nos três intervalos, e zero caso contrário.

O nível de significância global α^* , é dado por:

$$\hat{\alpha}^* = 1 - \sum_{i=1}^n \frac{L_i^{(t)}}{n} \quad (3.33)$$

A correção da significância global α^* , para um valor pré-fixado, depende de procedimento iterativo, em que a significância de cada j -ésimo limites de confiança α_j é atualizada.

Para obter esta estimativa de α_j , que garante um significância global α^* , pré-fixada, com base numa amostra de dados, o seguinte algoritmo pode ser utilizado.

Passo 1. Obtenha uma amostra de dados padronizados, transladados e rotacionados, $\{(x_{r(t)}, y_{r(t)}, z_{r(t)})\}$;

Passo 2. Para inicialização do processo iterativo, use $\alpha_j^{(0)} = \alpha^*/3$;

Passo 3. Utilizando o $\alpha_j^{(0)}$, ajuste as bandas de confiança A_j , onde $j = 1, 2$ e 3 , via regressão quantílica, dos sinais dos sinais padronizados;

Passo 4. Use Equação (3.32) e Equação (3.33) para estimar a significância global $\hat{\alpha}^*$;

Passo 5. Se $\hat{\alpha}^* \neq \alpha^*$, faça:

Passo 5.1. Faça uma correção no nível de significância comum da seguinte forma: $\alpha_j^{(1)} = \alpha_j^{(0)} \left(\frac{\hat{\alpha}^*}{\alpha^*}\right)^2$;

Passo 5.2. retorne ao passo 3 e faça $\alpha_j^{(0)} = \alpha_j^{(1)}$;

Passo 6. O procedimento se encerra quando $\hat{\alpha}^* \simeq \alpha^*$.

Figura 3.16: Proposta de algoritmo modificado para correção de Bonferroni

Para ilustrar a modificação na correção de Bonferroni, através de simulações com amostras de $n = 10.000$, com 32 ciclos por segundo, com as respectivas variâncias, $\sigma^2 = (0,005^2; 0,01^2; 0,05^2$ e $0,10^2)$

Tabela 3.1: Nível de significância global estimado via algoritmo

| Variância | α^* | $\hat{\alpha}^*$ | $\alpha_j = \frac{\alpha^*}{3}$ Correção de Bonferroni | $\hat{\alpha}_j$ Bonferroni Modificado |
|-----------|------------|------------------|--------------------------------------------------------------|----------------------------------------------|
| $0,005^2$ | 0,05 | 0,05 | 0,016667 | 0,02269782 |
| $0,01^2$ | 0,05 | 0,05 | 0,016667 | 0,02277454 |
| $0,05^2$ | 0,05 | 0,05 | 0,016667 | 0,02233740 |
| $0,10^2$ | 0,05 | 0,05 | 0,016667 | 0,02293718 |

3.7 Detecção de faltas por limites de Confiança simultâneos-3D

O monitoramento *online* dos sinais de elétricos para detecção de faltas pode ser obtido usando os princípios de gráfico de controle. O monitoramento consiste em acompanhar cada amostra e avaliar se a mesma se encontra, simultaneamente, dentro dos limites de controle ajustados dois-a-dois. A Figura 3.17 mostra o algoritmo computacional proposto para o monitoramento tridimensional sistema.

-
-
- Passo 1.** Usando uma amostra de tamanho n de corrente em condições normais de funcionamento, aplicar a Eq. (3.26) nos dados amostrais, a fim de estimar o parâmetro I_0 .
- Passo 2.** Calcule os sinais padronizados da corrente em cada fase: $x_{(t)} = \frac{i_{(t)}(a)}{I_0}$, $y_{(t)} = \frac{i_{(t)}(b)}{I_0}$ e $z_{(t)} = \frac{i_{(t)}(c)}{I_0}$;
- Passo 3.** Rotacione os sinais padronizados por um ângulo $= \frac{\pi}{4}$ ou $= \frac{3\pi}{4}$;
- Passo 4.** Calcule os limites superior e inferior para $a^{(p)}$, $b^{(p)}$ e $c^{(p)}$ usando Eq. (3.20), dois-a-dois. Use $p=0.975$ para limite superior ($a^{(0.975)}$, $b^{(0.975)}$ e $c^{(0.975)}$) e $p=0.025$ para limite inferior ($a^{(0.025)}$, $b^{(0.025)}$ e $c^{(0.025)}$). Assumindo nível de confiança de 95%.
- Passo 5.** Para o monitoramento do sistema, obtenha novas amostras de sinais da corrente. Primeiro aplique as operações padronização e rotação, utilizando os parâmetros estimados em condições normais de operação. (etapas 1 a 3).
- Passo 6.** Use as Equações, (3.27), (3.28) e (3.29) e os parâmetros estimados na etapa 4 para testar se sinais de corrente estão dentro dos limites de controle;
- Passo 7.** Se três ou mais pontos consecutivos se encontram fora das regiões de controle, em pelo menos uma das bandas de confiança, assume-se que o sistema não está em condições de funcionamento normal, isto é, há evidências de que ocorreu uma falha eléctrica na linha de transmissão.
-
-

Figura 3.17: Algoritmo para o monitoramento tridimensional de faltas.

3.8 Extração de atributos em três dimensões

Detectada a falta, conforme descrito na Seção 3.7, a fase seguinte é a extração de atributos para entrada no classificador, para discriminar o tipo de falta ocorrido.

3.8.1 Distância métrica do ponto a elipse funcional

Para calcular a distância métrica do ponto sob análise, normalmente sob falta, à elipse funcional é necessário especificar na elipse um ponto que seja a menor distância entre a elipse estimada pelas amostras em condições normais e o ponto sob análise. Para tal propõe-se estimar na elipse este respectivo ponto e posteriormente usar a distância euclidiana para obter esta distância.

O procedimento para calcular o ponto da curva, que seja a menor distância, de um ponto qualquer fora da elipse estimada a ser avaliado, é uma função do ângulo ω na Equação (3.34), para tal, é utilizado o método de otimização unidimensional *Golden Search*:

$$\omega^i = \operatorname{argmin} \left[x_{r(i)}^* - a \operatorname{sen}(\omega) \right]^2 + \left[y_{r(i)}^* - b \cos(\omega) \right]^2 \quad (3.34)$$

Onde: (a,b) são os parâmetros da elipse estimada em condições normais, (x_r^*,y_r^*) é o ponto externo a elipse e ω o ângulo a ser calculado, de forma a minimizar a distância entre a elipse e o ponto dado.

A distância mínima do ponto à elipse estimada, de forma bidimensional, é dada por:

$$D_i = \left[x_{r(i)}^* - a \operatorname{sen}(\omega^i) \right]^2 + \left[y_{r(i)}^* - b \operatorname{cos}(\omega^i) \right]^2 \quad (3.35)$$

Para o cálculo da distância do ponto externo, (x_r^*,y_r^*) , à elipse estimada, faça o seguinte:

-
-
- Passo 1** Usando os sinais da fase A e fase B, rotacionados e transladados (x_r,y_r) ;
- Passo 2** Ajuste a elipse funcional, com um percentil 50%. Nesse caso é a elipse que passa pelos pontos medianos do gráfico bidimensional;
- Passo 3** Estime os parâmetros (a,b) da elipse estimada em 2. Use algoritmo de Fitzgibbon;
- Passo 4** Calcule o ponto na elipse, usando a estimativa ω^i , obtida pela Equação 3.34, ou seja, $(a \operatorname{sen}(\omega^i), b \operatorname{cos}(\omega^i))$;
- Passo 5** Calcule a distância métrica, usando a Equação 3.35, dado um ponto sob análise ou sob falta (x_r^*,y_r^*) e o ponto estimado na elipse, $(a \operatorname{sen}(\omega^i), b \operatorname{cos}(\omega^i))$
-
-

Figura 3.18: Algoritmo para determinar a menor distância entre o ponto e a elipse estimada

O algoritmo da Figura 3.18, pode replicado para as fases (A,B), (A,C) e (C,B) da corrente elétrica.

A Figura 3.19, a seguir, mostra a distância perpendicular do ponto à elipse estimada.

Para o caso 3D, a distância métrica, é calculada utilizando-se a Equação (3.35), em suas combinações duas-a-duas das fases da corrente elétrica da seguinte forma.

$$D_i^{(a,b,c)} = \sqrt{D_i^{(a,b)} + D_i^{(a,c)} + D_i^{(b,c)}} \quad (3.36)$$

Para cada ponto (i) da amostra a ser avaliada, ou seja amostra, em tese, sob falta, é possível determinar um ponto perpendicular na elipse bidimensional, e dessa forma obter uma distância global para as três simultaneamente, conforme Equação (3.36).

3.8.2 Atributos ou características extraídas dos sinais em 3D

Os atributos extraídos em três dimensões(3D) dos distâncias euclidianas dos sinais da corrente, são basicamente medidas estatísticas descritivas enumeradas a seguir:

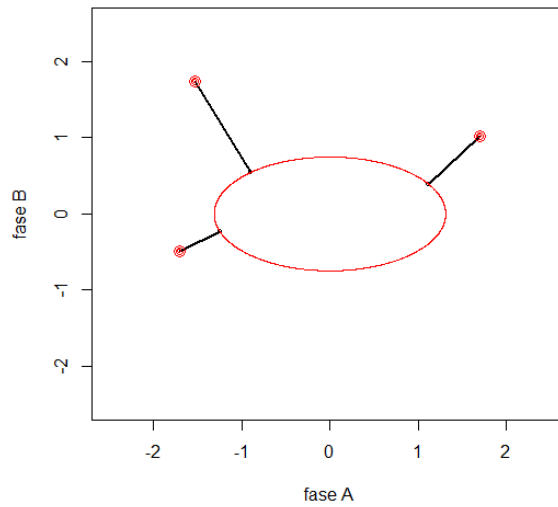


Figura 3.19: Distância do ponto à Elipse estimada.

1. Mínimo (Min);
2. Máximo (Máx);
3. Quartil 1 (Q1);
4. Quartil 3 (Q3);
5. Média (\bar{x});
6. Mediana (\tilde{x});
7. Variância (s^2);
8. Desvio Padrão (s);
9. Assimetria (Skewness);
10. Curtose (Kurtosis).

3.8.3 Seleção de atributos

Métodos automáticos de seleção de atributos são importantes em muitas situações em que se tem disponível um conjunto grande de atributos e deseja-se selecionar um subconjunto adequado, para ser usado como entrada em um classificador. Esse problema é tratado como redução de dimensionalidade.

O termo dimensionalidade é associado ao número de características de uma representação, ou seja, a

dimensão do espaço de características (atributos). As principais razões para que a dimensão seja a menor possível são duas, o custo de medição e a melhoria da precisão do classificador.

Os algoritmos de seleção de atributos são desenvolvidos principalmente para reduzir a dimensionalidade do espaço de atributos de forma que ocorra a menor queda possível no poder de distinção das classes por um classificador no espaço de características. A seleção de atributos pode também gerar um ganho no poder de distinção, por exemplo, quando se elimina informação redundante que atrapalhe a estimativa dos parâmetros do modelo, ou seja, problemas numéricos. Uma consequência da aplicação de um bom algoritmo de seleção de atributos é a redução do número necessário de amostras de treinamento para se obter bons resultados com um classificador, ou seja, a redução do problema da maldição da dimensionalidade (*Curse of dimensionality*), ou seja, adicionar características não significa sempre melhora no desempenho de um classificador (BELLMAN, R., 1961).

A seleção automática de características é uma técnica de otimização que, dado um conjunto de \mathbf{M} atributos, tenta selecionar um subconjunto de tamanho requerido \mathbf{P} ($\mathbf{P} < \mathbf{M}$) que maximiza uma função critério. É desejável que a função critério seja maior quanto menor for a redundância entre os atributos e quanto maior a facilidade de discriminar padrões de classes diferentes.

Os métodos determinísticos de solução única são baseados em buscas conforme critérios previamente estabelecidos. A maioria possui duas abordagens: progressiva (*forward*) e regressiva (*backward*): a abordagem progressiva inicia-se com um conjunto de avaliação (temporário) vazio e, conforme o algoritmo é executado, são inseridas características nesse conjunto, até que esse fique com tamanho requerido. Já na abordagem regressiva, inicia-se com um conjunto de avaliação contendo todas as características disponíveis e, nas interações do algoritmo, são excluídas características até que esse conjunto fique com o tamanho requerido. Os métodos de seleção de atributos consistem na avaliação de todos os atributos tomados individualmente. Os métodos mais comuns de seleção de atributos são:

1. Busca Sequencial Progressiva (SFS),
2. Busca Sequencial Regressiva (SBS),
3. Busca Exaustiva (BF).

A seguir descrevemos com mais detalhes aos métodos de **SFS**, **SBS** e **BE**:

Busca Seqüencial Progressiva (SFS)

O algoritmo SFS (*Seqüencial Forward Feature Selection*) inicia-se com um conjunto vazio. A cada iteração é selecionada a característica que, unida ao conjunto determinado pela iteração anterior, produz o melhor resultado da função critério. Essa característica é adicionada ao conjunto de características anterior e uma nova iteração é realizada. De acordo com Aha e Bankert (1996), Alamedine, Khalil e Marque (2013), o procedimento consiste em acrescentar sequencialmente o atributo possível de discriminação, quando combinado com as características que já foram selecionados.

Esse método tem menor custo computacional quando se deseja obter conjuntos pequenos em relação ao total de atributos (Jain e Zongker, 1997). Uma vez que um atributo tenha sido selecionado, ele não pode ser descartado do subconjunto.

Busca Seqüencial Regressiva (SBS)

A diferença entre SBS e SFS é que o SBS é iniciado com o conjunto de características completo (contendo todos os \mathbf{X} atributos), eliminando *uma a uma*, as menos importantes, ou seja, remove-se o atributo que maximiza a função critério para os restantes.

O algoritmo SBS, (Theodoridis e Koutrumbas, 2003), inicia-se com um conjunto de atributos completo ($X_0 = X$) e, nas iterações do algoritmo, remove-se o atributo com o mínimo valor da função critério $J(X_k + X^-)$, onde X_k é o conjunto de atributos já selecionados e X^- é o atributo a ser removido. O método SBS tem menor custo computacional, quando se deseja obter conjuntos grandes em relação ao total de atributos. Uma característica desse método é que uma vez eliminado o atributo, ele não retornará ao subconjunto (YAGHOUBI, 2014).

Busca exaustiva (BE)

O algoritmo de busca exaustiva ou força bruta é um algoritmo trivial, mas de uso muito geral, que consiste em enumerar todos os possíveis candidatos de uma solução e verificar se cada um satisfaz o problema. A Busca exaustiva varre o espaço de busca, fazendo combinações entre as possíveis características (atributos), buscando o melhor resultado.

Esse algoritmo possui uma implementação muito simple, e sempre encontrará uma solução se ela existir. Entretanto, seu custo computacional é proporcional ao número de candidatos à solução, que, em problemas reais, tende a crescer exponencialmente. Portanto, a força bruta é tipicamente usada em problemas

cujos tamanhos são limitados ou quando há uma heurística usada para reduzir o conjunto de candidatos para um espaço aceitável. Também pode ser usado quando a simplicidade da implementação é mais importante que a velocidade de execução, como nos casos de aplicações críticas em que os erros de algoritmo possuem sérias consequências.

3.8.4 Índice de exatidão Global - G

Um dos parâmetros utilizados na comparação de classificadores é o Índice de exatidão Global - G . É obtido por meio da matriz de erro ou matriz de confusão.

A matriz de erro ou matriz de confusão é uma matriz quadrada de números que expressam a quantidade de unidades amostrais, associada a uma dada categoria durante o processo de classificação efetuado, e a categoria real a que pertencem essas unidades.

A exatidão global é obtida dividindo a soma dos elementos da diagonal principal, x_{ii} , da matriz de confusão pelo número total de amostras n , ou seja:

$$G = \frac{\sum_{i=1}^c x_{ii}}{n} \quad (3.37)$$

Onde G é uma estimativa do índice de concordância global;

x_{ii} é o valor na linha i e coluna i ;

n é o número total de amostras e c o número total de classes.

3.8.5 Classificadores

O objetivo dos classificadores é classificar corretamente amostras futuras numa classe específica. Objetos ou indivíduos são caracterizados por um conjunto de atributos e através dessas características consegue-se determinar a classe a qual pertence o indivíduo ou objeto.

Classificador vizinho mais próximo: k -NN

Dado um conjunto de amostras (vetores de atributos ou padrões), já classificados nas classes $C = \{c_1, c_2, c_3, \dots, c_n\}$. Esse será o conjunto de referência para o método do *vizinho mais próximo*. Dado um padrão x cuja classificação é desconhecida, a regra k -NN classifica x na mesma classe que a do vetor $c_i \in C$ mais próximo de x . O objetivo é encontrar os K -vizinhos mais próximos de x que sejam similares para determinar o rótulo de classe de x . Sem qualquer conhecimento prévio, o classificador k NN utiliza

geralmente a distância euclidiana como a métrica de distância ou similaridade. No entanto, este método é simples e de fácil de implementação, podendo produzir resultados competitivos mesmo em comparação aos métodos de aprendizado de máquina mais sofisticadas. (ZHANG, H. et al, 2006)

O algoritmo de busca dos K -vizinhos mais próximos, $kNN(s_c, k)$, recebe como parâmetros o elemento central da consulta s_c e uma quantidade K de vizinhos. Formalmente, dado um sub-conjunto de dados $S = \{s_1, s_2, \dots, s_n\}$, contido em \mathbb{S} e uma função de similaridade d , os K -vizinhos mais próximos, centrados em um elemento $s_c \in \mathbb{S}$, com número de elementos em torno de s_c igual a K , temos:

$$kNN(s_c, k) = A = \{s_j | s_j \in S, \forall s_i \in (S - A) : d(s_j, s_c) \leq d(s_j, s_i), |A| = k\} \quad (3.38)$$

A Figura 3.20, ilustra uma consulta aos k -vizinhos mais próximos em uma base de dados bidimensional, onde s_c é o elemento central de busca, ou seja, $kNN(s_c, 6)$.

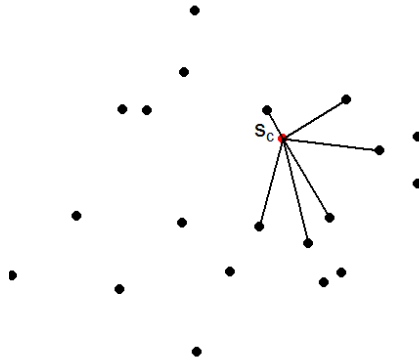


Figura 3.20: k - vizinhos mais próximos com $k=6$

O desempenho de um classificador kNN é determinado principalmente pelo valor K , bem como a métrica de distância aplicada. No entanto, tem sido demonstrado que quando os pontos não estão uniformemente distribuídos, predeterminar valor de K se torna mais difícil. De um modo geral, maiores valores de K são mais robustos em relação ao erro de classificação, não sendo isto uma verdade absoluta.

Bayesian Networks-BayesNet

O algoritmo *Bayesian Networks* ou *BayesNet* corresponde a uma modelagem de uma rede bayseana como um classificador. Redes Bayesianas são algoritmos de aprendizado de máquinas capazes de fornecer previsões associadas a valores de probabilidades. O estudo de Redes Bayesianas é complexo e exige um conhecimento razoável sobre probabilidade.

Considere uma Rede Bayesiana (**RB**), (Vijayarani e Muthulakshmi, 2013), contendo n nós. Seja $X = x_1, \dots, x_n$, com $n \geq 1$, um conjunto de variáveis.

Uma particular distribuição conjunta dos valores é representada por $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ ou de maneira simplificada $P(x_1, x_2, \dots, x_n)$. Utilizando a *regra da cadeia*, podemos reescrever como:

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i | x_1 \dots x_{i-1}) \quad (3.39)$$

Assumindo a propriedade de **Markov**⁴, a estrutura de uma **RB** implica que o valor de um determinado **nó** está *condicionado* apenas aos valores dos seus *pais*. $Pais(x)$ são todos os nós que possuem arcos que vão em direção ao nó da variável (x). Logo a Equação (3.39) pode ser reescrita como:

$$P(X) = P(x_1, x_2, \dots, x_n) = \prod_i P(x_i | Pais(x_i)) \quad (3.40)$$

$P(X)$ é a probabilidade conjunta para a rede e $P(x | Pais(x))$ são as probabilidades condicionais a x em relação aos seus pais.

Linear Discriminant Analysis-LDA

LDA é um método de classificação que encontra uma combinação linear dos atributos de dados que melhor separarão os dados em classes. É uma generalização do discriminante linear de Fisher, um método utilizado em estatística, reconhecimento de padrões e aprendizado de máquina (*machine learning*) para encontrar uma combinação linear de atributos que caracterizam ou separam duas ou mais classes de objetos ou eventos.

Para o problema *C-classes*, LDA tem como objetivo buscar um conjunto de vetores ideais, denotado por $W = [w_1, w_2, \dots, w_n]$, tal que a função discriminante de *Fisher*, Equação (3.41), seja maximizada.

$$J(W) = tr \left(\frac{W^T S_b W}{W^T S_w W} \right) \quad (3.41)$$

onde a matriz de dispersão intra-classe S_w e matriz de espalhamento entre-classes S_b são definidas por:

$$S_w = \sum_{i=1}^C \sum_{j=1}^{n_i} p_i (x_{ij} - \bar{x}_i) (x_{ij} - \bar{x}_i)^T \quad (3.42)$$

⁴Um processo de Markov é um processo estocástico em que a probabilidade de o sistema estar no estado j no período $(n+1)$ depende somente do estado em que o sistema está no período n . Ou seja, para os processos de Markov, só interessa o estado imediato, logo $P_{ij} = P(X_{n+1} = j | X_n = i)$

$$S_b = \sum_{i=1}^C p_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (3.43)$$

Onde p_i é a probabilidade a priori da i -ésima classe, \bar{x}_i é a média da classe i , x_{ij} é a j -ésima amostra da classe i e n_i é o número de amostras da classe i . E finalmente \bar{x} é a média global das amostras. O objetivo da LDA é minimizar a variância intraclasse e maximizar a variância entre as classes. A solução é obtida pela decomposição dos autovalores de $S_w^{-1}S_b$ e tomando os autovetores correspondentes aos maiores autovalores. (ZHAO et al, 2012), (YAO, et al, 2014).

Classificador árvore de decisão - Decision Trees

Decision Trees ou árvores de decisão são sistemas de decisão de vários estágios em que as classes são sequencialmente rejeitadas até que chegamos a uma classe finalmente aceita. Para esse efeito, o espaço de características é dividido em regiões únicas, correspondente às classes, *de uma forma sequencial* (ROKACH e MAIMON, 2014). Mediante um vetor de atributos, a busca da região em que vector de características será atribuído é conseguido através de uma sequência de decisões ao longo de um caminho de nós de uma árvore apropriadamente construídos. Esses esquemas oferecem vantagens quando um grande número de classes é envolvido. A sequência de decisões é aplicada a atributos individuais e as questões a serem respondidas são da forma "este atributo $x_i \leq \gamma$?", onde γ é um valor de corte, (*threshold*). A Figura 3.21 mostra a árvore com seus nós de decisão t_i e os respectivos estágios de classificação w_i . Note que é possível chegar a uma decisão sem ter testado todas as características disponíveis.

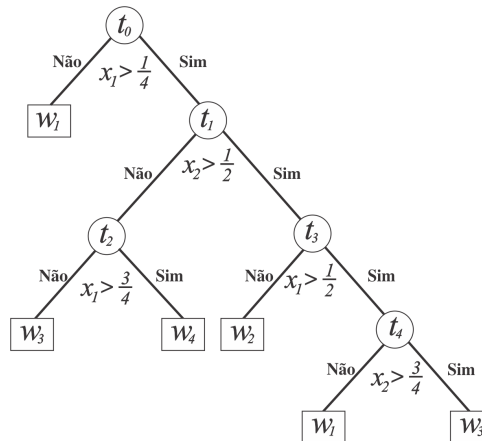


Figura 3.21: Exemplo de árvore de decisão

Estes classificadores não refletem o estado da arte, no entanto, a escolha foi em função dos trabalhos de Gomes et al (2013), a fim que os resultados obtidos possam ser comparados.

Capítulo 4

Experimentos e resultados

Este Capítulo apresenta os resultados relacionados a dados reais e simulados, conforme metodologia apresentada no Capítulo 3.

4.1 Descrição da base de dados obtida da Companhia Energética de Minas Gerais S.A – CEMIG.

O número de informações documentados em relação às faltas ocorridas em linhas de transmissão, presente no sistema da Cemig é muito grande, entretanto, existem apenas uma pequena parcela de registros oscilográficos para essas faltas, obtidas através de Registradores Digital de Perturbações (RDPs). Os tipos de faltas selecionados foram as provocadas por queda de árvore (W1), descarga elétrica (ND), entrelaçamento de cabos (K6) e queimada (AQ).

O banco de dados foi formado por amostras oscilográficas coletadas no período de 2001 a 2003. Em alguns casos, foi possível obter mais de um registro de falta para uma mesma ocorrência de falta em uma linha de transmissão. Isso ocorreu pelo fato de na mesma linha existirem dois RDPs, um em cada estação(fonte) da linha.

Ocorreu também do RDP registrar para a mesma falta mais de uma oscilografia referente à uma falta recorrente. Para o primeiro caso, foi selecionada apenas a oscilografia que apresentou a maior sobrecorrente em uma das fases, considerando que essa seja o mais representativo para caracterizar a falta. No segundo caso, foi selecionado apenas o primeiro registro das faltas recorrentes.

A oscilografia para cada fase da linha de transmissão é formada por um período de tempo de sinal pré-falta (operação normal - OP) e período de falta, caso tenha ocorrido falta na fase. Para cada oscilografia foram extraídas amostras de sinal em operação normal das 3 fases. Os dados foram amostrados com uma taxa de amostragem de 24 pontos por ciclo, ou um ponto a cada 0,07 milissegundos. Após o tratamento da base, excluindo registros considerados redundantes, a base ficou caracterizada de acordo com a Tabela 4.1.

Tabela 4.1: Base de dados

| Tipo de falta | Qtd de Bancos |
|------------------------------|---------------|
| OP- Operação Normal | 41 |
| AQ- Queimadas | 15 |
| K6 - Entrelaçamento de Cabos | 4 |
| ND - Descarga Elétrica | 15 |
| W1 - Queda de Arvore | 7 |
| Total | 82 |

Neste estudo foram utilizados 41 bancos de dados, com 6 colunas (três primeiras fases da corrente (A, B e C) e as três últimas fases da tensão (A, B e C)) e totalizando 3.000 linhas. Esses bancos contêm informações antes e após a ocorrência de faltas.

A sumarização dos resíduos em termos de análise de normalidade e correlação entre a corrente e tensão nos permitirá uma melhor compreensão do fenômeno elétrico. A análise de correlação e normalidade, nos 41 bancos de dados selecionados, amostraram-se 96 amostras ($n=96$), correspondendo a quatro ciclos, ou seja, cada ciclo possui 24 realizações do processo estocástico sob estudo. Essas amostras iniciais estão sob condições normais de operação. A análise foi desenvolvida em cada banco de dados, e suas respectivas combinações de fases: A , B , C .

4.1.1 Da correlação e teste de normalidade para Corrente e Tensão

Nesta seção analisamos a correlação e teste de normalidade entre a corrente (Fases A, B e C da Corrente) e tensão (Fases A, B e C da tensão). As Tabelas A.1, A.2 e A.3, do Apêndice, mostram as respectivas análises dos resíduos. De maneira geral, podemos verificar que há correlação estatisticamente significativa entre as fases da corrente e tensão, bem como a hipótese de normalidade dos resíduos mostrou-se rejeitada. Assim sendo, esses resultados nos dá suporte para análises com uso de ferramentas estatísticas robustas.

4.1.2 Da correlação dos resíduos da corrente para as três fases A, B e C

A tensão de pico, estimada para a análise de correlação da corrente nas fases A , B e C , foi obtida utilizando-se a Equação (3.26), que leva em consideração os valores dos sinais elétricos nas fases A , B e

C, dada na Seção 3.4.2. De maneira geral, através da Tabela 4.2, verificamos que em todos os bancos de dados, a correlação é altamente significativa entre as fases da corrente.

Tabela 4.2: Análise de Correlação para corrente nas fases A,B e C

| Ord | Banco | Matriz de Correlação | | | Ord | Banco | Matriz Correlação | | | | |
|-----|----------|----------------------|--------|--------|--------|-------|-------------------|---|--------|--------|--------|
| | | A | B | C | | | A | B | C | | |
| 1 | 01aq0101 | A | 1,0000 | 0,7537 | 0,6415 | 22 | 33ND0401 | A | 1,0000 | 0,7443 | 0,7405 |
| | | B | | 1,0000 | 0,7112 | | | B | | 1,0000 | 0,7333 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 2 | 02aq0102 | A | 1,0000 | 0,6584 | 0,6385 | 23 | 36ND0502 | A | 1,0000 | 0,6770 | 0,6768 |
| | | B | | 1,0000 | 0,6463 | | | B | | 1,0000 | 0,6910 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 3 | 04aq0201 | A | 1,0000 | 0,5787 | 0,5284 | 24 | 37ND0601 | A | 1,0000 | 0,6063 | 0,6017 |
| | | B | | 1,0000 | 0,5355 | | | B | | 1,0000 | 0,6166 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 4 | 06aq0203 | A | 1,0000 | 0,6532 | 0,6719 | 25 | 39ND0701 | A | 1,0000 | 0,8207 | 0,8120 |
| | | B | | 1,0000 | 0,7175 | | | B | | 1,0000 | 0,8091 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 5 | 08aq0301 | A | 1,0000 | 0,7372 | 0,7212 | 26 | 40ND0801 | A | 1,0000 | 0,6478 | 0,6608 |
| | | B | | 1,0000 | 0,7248 | | | B | | 1,0000 | 0,6397 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 6 | 09aq0302 | A | 1,0000 | 0,7373 | 0,7274 | 27 | 41ND0901 | A | 1,0000 | 0,6537 | 0,6530 |
| | | B | | 1,0000 | 0,7201 | | | B | | 1,0000 | 0,6715 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 7 | 10aq0401 | A | 1,0000 | 0,6314 | 0,6658 | 28 | 42ND1001 | A | 1,0000 | 0,6589 | 0,6755 |
| | | B | | 1,0000 | 0,6568 | | | B | | 1,0000 | 0,6951 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 8 | 12aq0501 | A | 1,0000 | 0,7106 | 0,6467 | 29 | 43ND1101 | A | 1,0000 | 0,7190 | 0,7058 |
| | | B | | 1,0000 | 0,6875 | | | B | | 1,0000 | 0,7165 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 9 | 13aq0601 | A | 1,0000 | 0,7165 | 0,7634 | 30 | 45ND1201 | A | 1,0000 | 0,5774 | 0,5632 |
| | | B | | 1,0000 | 0,7648 | | | B | | 1,0000 | 0,5570 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 10 | 14aq0701 | A | 1,0000 | 0,7165 | 0,7634 | 31 | 46ND1301 | A | 1,0000 | 0,5774 | 0,5632 |
| | | B | | 1,0000 | 0,7648 | | | B | | 1,0000 | 0,5570 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 11 | 15aq0801 | A | 1,0000 | 0,6773 | 0,6495 | 32 | 47ND1401 | A | 1,0000 | 0,6578 | 0,6367 |
| | | B | | 1,0000 | 0,6495 | | | B | | 1,0000 | 0,6683 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 12 | 17aq0901 | A | 1,0000 | 0,6506 | 0,6396 | 33 | 49ND1502 | A | 1,0000 | 0,5820 | 0,5539 |
| | | B | | 1,0000 | 0,6517 | | | B | | 1,0000 | 0,5342 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 13 | 18aq0902 | A | 1,0000 | 0,7725 | 0,7312 | 34 | 51ND1602 | A | 1,0000 | 0,6027 | 0,6068 |
| | | B | | 1,0000 | 0,7348 | | | B | | 1,0000 | 0,6067 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 14 | 22aq1001 | A | 1,0000 | 0,6425 | 0,6438 | 35 | 53W10102 | A | 1,0000 | 0,7820 | 0,6904 |
| | | B | | 1,0000 | 0,6554 | | | B | | 1,0000 | 0,7097 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 15 | 23aq1101 | A | 1,0000 | 0,6320 | 0,6301 | 36 | 56W10105 | A | 1,0000 | 0,8654 | 0,8631 |
| | | B | | 1,0000 | 0,6318 | | | B | | 1,0000 | 0,8476 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 16 | 25K60101 | A | 1,0000 | 0,7010 | 0,6981 | 37 | 59W10301 | A | 1,0000 | 0,8193 | 0,8245 |
| | | B | | 1,0000 | 0,6577 | | | B | | 1,0000 | 0,8105 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 17 | 26K60201 | A | 1,0000 | 0,6491 | 0,6464 | 38 | 60W10501 | A | 1,0000 | 0,6240 | 0,6967 |
| | | B | | 1,0000 | 0,6518 | | | B | | 1,0000 | 0,6941 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 18 | 27K60202 | A | 1,0000 | 0,6771 | 0,6617 | 39 | 61W10701 | A | 1,0000 | 0,7953 | 0,8007 |
| | | B | | 1,0000 | 0,6768 | | | B | | 1,0000 | 0,8037 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 19 | 28K60301 | A | 1,0000 | 0,6771 | 0,6617 | 40 | 62W10801 | A | 1,0000 | 0,7953 | 0,8007 |
| | | B | | 1,0000 | 0,6768 | | | B | | 1,0000 | 0,8037 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 20 | 29ND0201 | A | 1,0000 | 0,7848 | 0,7837 | 41 | 63W10901 | A | 1,0000 | 0,6261 | 0,6405 |
| | | B | | 1,0000 | 0,7846 | | | B | | 1,0000 | 0,5770 |
| | | C | | | 1,0000 | | | C | | | 1,0000 |
| 21 | 32ND0302 | A | 1,0000 | 0,7537 | 0,6415 | | | | | | |
| | | B | | 1,0000 | 0,7112 | | | | | | |
| | | C | | | 1,0000 | | | | | | |

4.2 Da maior sensibilidade relativas as faltas da corrente elétrica

O uso da energia espectral da corrente fornece resultados bem superiores, na análise, do que o uso da tensão elétrica. Tal comportamento pode ser explicado pela maior sensibilidade a faltas da corrente elétrica (AGGARWAL et al, 2015). Observe que após a ocorrência da falta, os sinais de corrente possuem maior amplitude, Figura 4.1a, do que os sinais da tensão, Figura 4.1b.

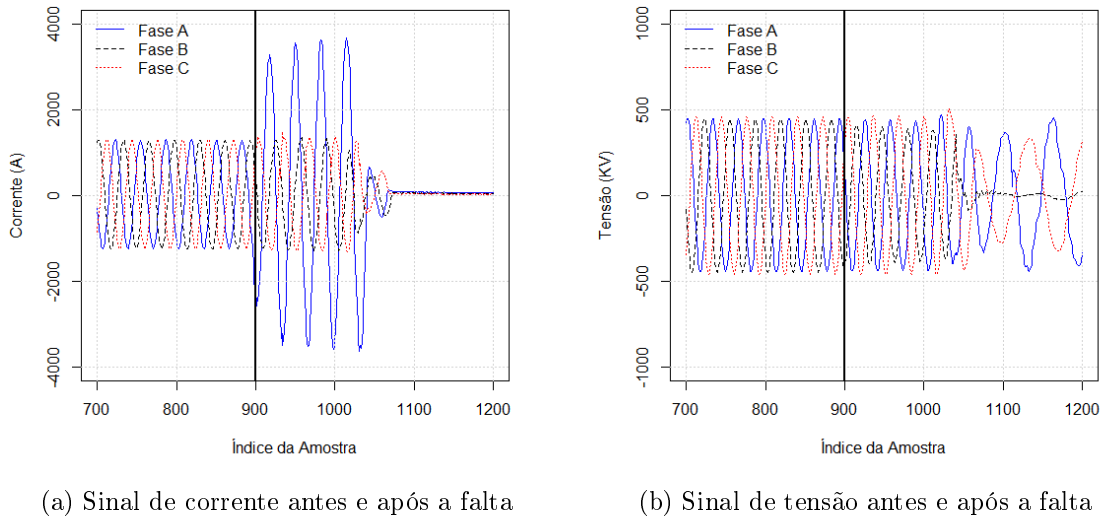


Figura 4.1: Sinais de corrente e tensão para a linha de transmissão (fases A, B e C) antes e após a falta por queimadas (AQ) - Banco de dados 04aq0201.

4.3 Da metodologia 2D melhorada

4.3.1 Avaliação da melhoria proposta na metodologia de Gomes et al (2013)- Dados simulados

Nesta seção, comparamos a proposta de construção de limites de confiança sugerida por Gomes et al (2013) e a melhoria proposta nesta tese através da construção dos intervalos via regressão quantílica e estimação dos parâmetros da elipse via algoritmo de Fitzgibbon. A comparação se restringe em avaliar se os intervalos construídos conseguem retornar o nível de confiança determinado a priori na sua construção, para tal utilizaram-se dados simulados.

Foram usadas amostras de tamanho $n=320$, com 10 ciclos. Valores aleatórios para ξ_V e ξ_I foram gerados usando distribuição gaussiana com média zero e variâncias iguais, ou seja, $\xi_V \sim N(0, \sigma_V^2)$ e $\xi_I \sim N(0, \sigma_I^2)$. Foram utilizados os seguintes valores de variância: $\sigma_V^2 = \sigma_I^2 = (0,005^2; 0,01^2; 0,05^2 \text{ e } 0,10^2)$. Foram

avaliadas 10.000 repetições, sendo que cada repetição é uma amostra de $n=320$, gerada, utilizando as Equações dadas em (2.28).

As estimativas de V_0 e I_0 e $\cos(\varphi)$ foram calculadas utilizando o método proposto por Gomes et al (2013) e também utilizando a melhoria introduzida nesta tese através de procedimento numérico de Mínimos Quadrados, conforme proposto por Pilu, Fitzgibbon e Fischer (1996). O que se percebe de maneira geral é que à medida que o ruído dos resíduos aumenta, a melhoria introduzida neste trabalho apresenta melhores resultados (ver Tabela 4.3).

Tabela 4.3: Estimativas do fator de potência e valores de pico da tensão e corrente.

| $\sigma_V^2 = \sigma_I^2$ | | Método 1 | | | Método 2 | | |
|---------------------------|------------|-----------------|--------|----------|-----------------|--------|----------|
| | | $\cos(\varphi)$ | V_0 | I_0 | $\cos(\varphi)$ | V_0 | I_0 |
| 0,005 ² | Valor Real | -0,50 | 770 | 8760 | -0,50 | 770 | 8760 |
| | Média | -0,50 | 770,05 | 8.760,60 | -0,50 | 770,00 | 8.760,02 |
| | Variância | 0,00 | 0,39 | 4,49 | 0,00 | 0,39 | 4,49 |
| | Erro(%) | 0,02 | 0,01 | 0,01 | 0,02 | 0,00 | 0,00 |
| 0,01 ² | Média | -0,50 | 770,22 | 8762,48 | -0,50 | 770,01 | 8760,18 |
| | Variância | 0,00 | 0,79 | 9,00 | 0,00 | 0,79 | 9,00 |
| | Erro(%) | 0,08 | 0,03 | 0,03 | 0,08 | 0,00 | 0,00 |
| 0,05 ² | Média | -0,49 | 775,11 | 8.817,77 | -0,49 | 770,19 | 8.761,80 |
| | Variância | 0,01 | 3,93 | 45,58 | 0,01 | 3,87 | 44,84 |
| | Erro(%) | 1,93 | 0,66 | 0,66 | 1,90 | 0,02 | 0,02 |
| 0,10 ² | Média | -0,46 | 790,38 | 8993,12 | -0,47 | 772,03 | 8784,32 |
| | Variância | 0,01 | 7,94 | 90,81 | 0,01 | 7,52 | 85,70 |
| | Erro(%) | 7,30 | 2,65 | 2,66 | 6,83 | 0,26 | 0,28 |

Nota: $\text{Erro}(\%) = |(\text{ValorReal} - \text{Media}) / \text{ValorReal} \times 100|$

Método 1: Procedimento de Gomes et al. (2013)

Método 2: Procedimento numérico de mínimos quadrados, Pilu, Fitzgibbon e Fischer (1996)

A Tabela 4.4 apresenta a proporção de pontos dentro dos limites de controle, utilizando os mesmos dados simulados que foram aplicados para estimar os parâmetros. Ao fazer isso, espera-se encontrar a proporção de pontos próximos $(1 - \alpha)\%$, em que foi escolhida como $\alpha = 0,05$, (5%). Os resultados mostram que os limites de controle, assumindo variáveis aleatórias gaussianas independentes, detêm menos de 95% dos pontos dentro dos limites de controle. Em média, 92,5% de pontos simulados estão dentro dos limites de controle, embora a simulação e o método original de Gomes et al (2013) seja baseada em variáveis aleatórias gaussianas independentes. Como consequência, 7,5% dos pontos, em média, estão fora dos limites de controle. Usando o percentil dos resíduos (suposição não-gaussiana) a proporção de pontos dentro dos limites de controle é de aproximadamente 45%. Esse é o menor valor entre os métodos avaliados e representa quase a metade do nível de confiança proposto (95%). Assim, se este método é aplicado, em seguida, aproximadamente 64% dos pontos irão situar entre os limites de controle. Na prática, não recomendamos essa abordagem. Finalmente, a abordagem de regressão quantílica não-linear assegura 94,6% dos pontos dentro dos limites de controle, o que é muito mais perto de 95% do que as abordagens

anteriores (ver Tabela 4.4) .

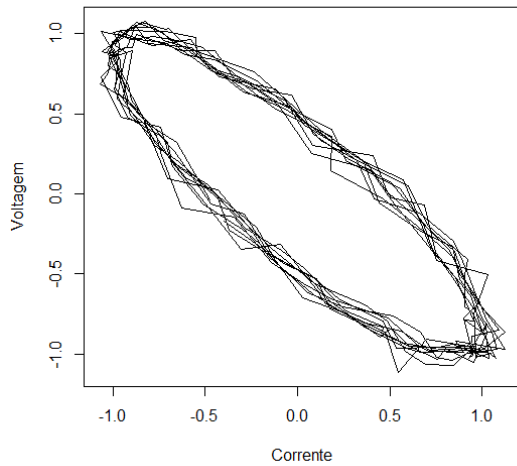
Tabela 4.4: Percentual (%) de pontos no interior dos intervalos Construídos (dados simulados).

| Variância $\sigma_V^2 = \sigma_I^2$ | Estatísticas | Método Gaussiano | Método Não Gaussiano | Regressão Quantílica |
|----------------------------------------|--------------|---------------------|-------------------------|-------------------------|
| 0,005 ² | Mínimo | 89,0625 | 37,8125 | 92,1875 |
| | Média | 92,5620 | 45,9676 | 94,6291 |
| | Máximo | 95,9375 | 54,0625 | 96,2500 |
| 0,01 ² | Mínimo | 89,0625 | 37,1875 | 91,8750 |
| | Média | 92,5672 | 45,9618 | 94,6194 |
| | Máximo | 96,2500 | 54,0625 | 95,9375 |
| 0,05 ² | Mínimo | 89,0625 | 36,5625 | 91,2500 |
| | Média | 92,5797 | 45,4936 | 94,6096 |
| | Máximo | 96,2500 | 53,4375 | 95,9375 |
| 0,10 ² | Mínimo | 89,3750 | 37,1875 | 91,8750 |
| | Média | 92,6403 | 44,8731 | 94,6658 |
| | Máximo | 95,9375 | 52,5000 | 95,9375 |

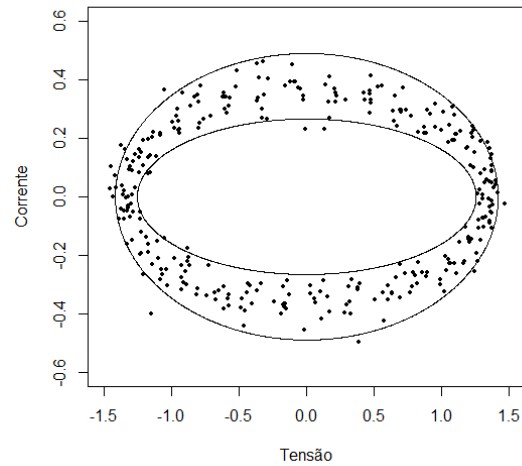
Na prática, a utilização dos métodos de resíduos gaussianos e não gaussianas geram limites de controle estreitos, com taxas de falhas maiores do que o nível de significância de α , pré-estabelecido. Logo há uma maior probabilidade de unidades amostrais serem classificadas como exterior à elipse (ver Figura 4.2)

De maneira visual, podemos perceber que as elipses construídas com base na metodologia de Gomes et al (2013) e resíduos não gaussianos não são elipses em sua forma padrão, uma vez que nas suas extremidades a amplitude intervalar é muito menor que em outras regiões da elipse.

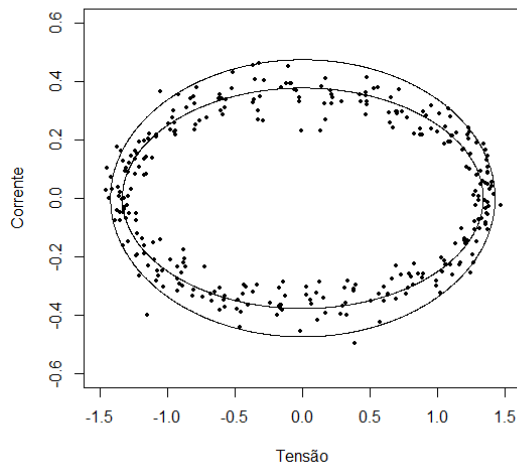
Em complemento no Apêndice B.1 tem-se uma demonstração matemática comparando o limite de confiança proposto por Gomes et al (2013), e a elipse padrão, concluindo que o limite de confiança proposto pelos autores não é uma elipse na sua forma padrão.



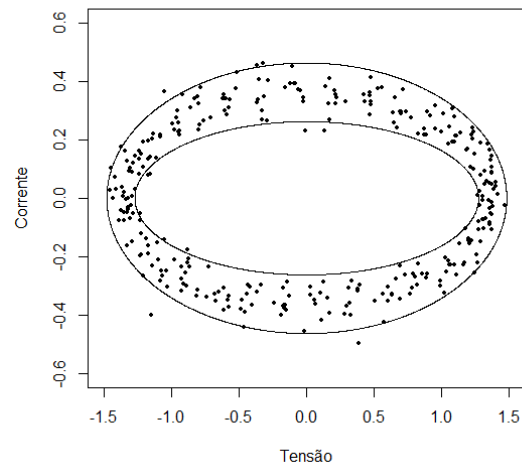
(a) Sinal da Tensão e Corrente no espaço bidimensional



(b) Limite de confiança por Gomes et al (2013)



(c) Limite de confiança por resíduos não gaussianos



(d) Limite de confiança por regressão quantílica.

Figura 4.2: Sinais de Tensão e Corrente através do tempo(a), limite de Confiança Gomes et al (2013) (b), limite de Confiança não gaussiano (c) e limite de Confiança por regressão quantílica(d)

4.3.2 Aplicação da Metodologia 2D melhorada em dados reais: detecção de faltas

Utilizando o sistema de monitoramento *online* 2D, conforme algoritmo dado na Figura 3.8. Utilizaram-se os ciclos iniciais de cada banco de dados, ou seja 96 observações, para ajustar os limites de controle, e posteriormente, todos os pontos amostrais para testar o algoritmo proposto. Assim, tem-se os seguintes resultados para os banco de dados avaliados.

A Tabela 4.5 mostra os resultados considerando as fases A, B e C da corrente e tensão, separadamente, sendo que a coluna "Detecção da Fase A", significa que foram utilizadas a fase A da Corrente e a respectiva fase A da tensão e, assim sucessivamente, para as demais fases B e C.

Utilizando a fase A da corrente e da tensão temos 12,19% dos bancos analisados, o algoritmo não detectou corretamente o ponto de falta, especialmente nas amostras relacionadas às faltas por descarga elétrica (ND) 9,75% e por queda de árvore na linha de transmissão (W1) 2.44%.

Com relação à fase B, verificamos que em 14,63% dos bancos, o algoritmo não detectou corretamente. E com relação à fase C da corrente e tensão, em 53,66% dos bancos, o algoritmo não detectou corretamente a falta. Isso pode ser explicado pela menor variação do sinal após a ocorrência da falta, conforme Figura 4.3a e Figura 4.3b, a seguir, obtida do banco de ordem 1, (01aq0101).

Em uma análise mais acurada do banco de ordem 1, verificamos que o índice da amostra em que ocorreu a falta é 691, logo o algoritmo detecta de maneira precisa o ponto de falta.

Tabela 4.5: Aplicação de algoritmo de detecção de faltas 2D em dados reais

| Ord | Banco | Ponto da Falta | Detecção da Falta A | Detecção da Falta B | Detecção da Falta C | Ord | Banco | Ponto da Falta | Detecção da Falta A | Detecção da Falta B | Detecção da Falta C |
|-----|----------|----------------|---------------------|---------------------|---------------------|-----|----------|----------------|---------------------|---------------------|---------------------|
| 1 | 01aq0101 | 695 | 691 | 220 | 46 | 22 | 33ND0401 | 716 | 716 | 404 | 5 |
| 2 | 02aq0102 | 231 | 244 | 229 | 501 | 23 | 36ND0502 | 814 | 691 | 686 | 820 |
| 3 | 04aq0201 | 905 | 897 | 1086 | 919 | 24 | 37ND0601 | 1031 | 1 | 1 | 1 |
| 4 | 06aq0203 | 1242 | 1242 | 1429 | 1292 | 25 | 39ND0701 | 265 | 175 | 260 | 75 |
| 5 | 08aq0301 | 254 | 256 | 1 | 1 | 26 | 40ND0801 | 1070 | 783 | 1071 | 29 |
| 6 | 09aq0302 | 215 | 123 | 230 | 1 | 27 | 41ND0901 | 1042 | 1 | 1 | 1 |
| 7 | 10aq0401 | 1120 | 255 | 762 | 138 | 28 | 42ND1001 | 1275 | 1266 | 495 | 381 |
| 8 | 12aq0501 | 1169 | 1161 | 425 | 311 | 29 | 43ND1101 | 254 | 254 | 254 | 368 |
| 9 | 13aq0601 | 791 | 791 | 283 | 808 | 30 | 45ND1201 | 1214 | 1 | 1 | 1 |
| 10 | 14aq0701 | 721 | 712 | 213 | 123 | 31 | 46ND1301 | 258 | 458 | 1 | 1 |
| 11 | 15aq0801 | 845 | 843 | 1035 | 849 | 32 | 47ND1401 | 1353 | 1399 | 1361 | 1541 |
| 12 | 17aq0901 | 854 | 845 | 851 | 39 | 33 | 49ND1502 | 173 | 1 | 1 | 1 |
| 13 | 18aq0902 | 243 | 243 | 441 | 247 | 34 | 51ND1602 | 877 | 877 | 888 | 1030 |
| 14 | 22aq1001 | 1251 | 1246 | 192 | 393 | 35 | 53W10102 | 744 | 815 | 743 | 4 |
| 15 | 23aq1101 | 1209 | 1209 | 1332 | 1328 | 36 | 56W10105 | 746 | 977 | 1123 | 8 |
| 16 | 25K60101 | 257 | 209 | 191 | 8 | 37 | 59W10301 | 626 | 626 | 626 | 23 |
| 17 | 26K60201 | 1172 | 1163 | 1165 | 27 | 38 | 60W10501 | 857 | 401 | 928 | 15 |
| 18 | 27K60202 | 258 | 256 | 265 | 258 | 39 | 61W10701 | 286 | 289 | 278 | 3 |
| 19 | 28K60301 | 516 | 516 | 516 | 22 | 40 | 62W10801 | 1424 | 1 | 1425 | 3 |
| 20 | 29ND0201 | 274 | 274 | 274 | 6 | 41 | 63W10901 | 246 | 102 | 247 | 251 |
| 21 | 32ND0302 | 695 | 691 | 220 | 46 | | | | | | |

Uma limitação importante é que este procedimento só leva em conta uma das fases da corrente e da tensão. Logo se a falta ocorreu na fase B ou C da corrente ou da tensão, não afetando as demais, o algoritmo não detecta tal situação. Dessa forma, é conveniente uma metodologia que leva em consideração uma análise conjunta das três fases, conforme proposto neste trabalho.

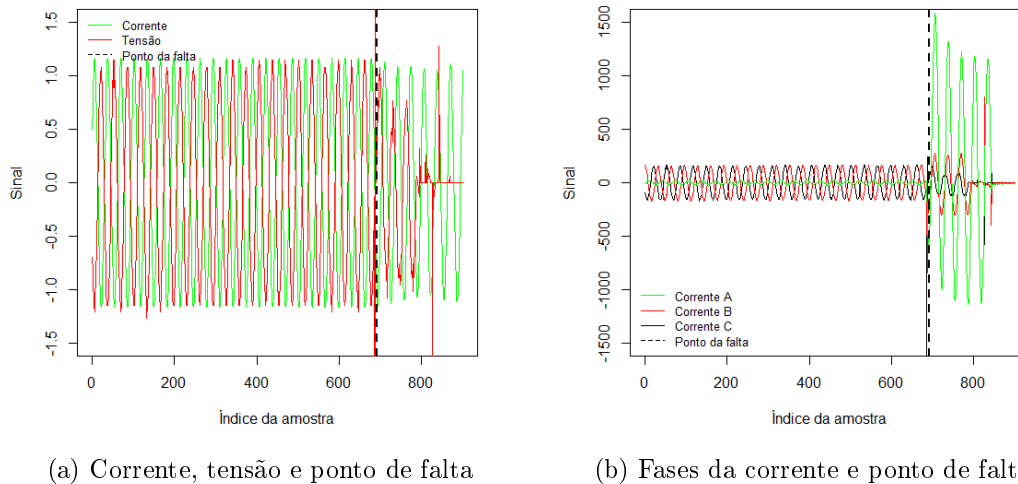


Figura 4.3: Corrente e tensão (a) e Corrente e suas fases A, B e C e os respectivos pontos de falta

4.4 Metodologia 3D

Nesta Seção apresentamos a metodologia de detecção 3D, onde simultaneamente utilizaram-se as fases A, B e C da corrente elétrica, para monitoramento *Online 3D* de ocorrências de faltas em Linhas de Transmissão.

4.4.1 Detecção de faltas em 3D

Utilizando o algoritmo de detecção de faltas 3D, proposto Seção 3.7 e Figura 3.17, e ainda informação da corrente elétrica, nas suas respectivas fases A, B e C, conforme Tabela 4.6, temos que, em 34,15% dos bancos de dados, a falta foi detectada exatamente no terceiro ponto após a ocorrência e em 65,85% dos casos, a detecção ocorreu nos dados em situação de pré-falta. Em todos os casos testados, verificou-se que o procedimento detectou a presença de falta.

Observa-se ainda que nas amostras, sob condições normais, utilizadas para ajustes do modelos (as 96 amostras iniciais), não houve caso de detecção de falso positivo, ou seja, o algoritmo não acusou a presença de faltas nas amostras iniciais sob condições normais.

A abordagem em 3D apresentou melhores resultados em relação à abordagem 2D, pois detectou de maneira satisfatória a presença de falta em linhas de transmissão. A sensibilidade à falta da abordagem 3D é muito superior pois detectou a falta antes da ocorrência do *pico* da falta em 65,85% dos casos.

Tabela 4.6: Aplicação de algoritmo de detecção de faltas 3D em dados reais

| Ord. | Banco | Ponto de detecção da falta | Ponto da Falta | Ponto da falta ou pré-falta | Ord. | Banco | Ponto de detecção da falta | Ponto da Falta | Ponto da falta ou pré-falta |
|------|----------|----------------------------|----------------|-----------------------------|------|----------|----------------------------|----------------|-----------------------------|
| 1 | 01aq0101 | 698 | 695 | Ponto Falta | 22 | 33ND0401 | 444 | 716 | Pré-falta |
| 2 | 02aq0102 | 158 | 231 | Pré-falta | 23 | 36ND0502 | 215 | 814 | Pré-falta |
| 3 | 04aq0201 | 908 | 905 | Ponto Falta | 24 | 37ND0601 | 1034 | 1031 | Ponto Falta |
| 4 | 06aq0203 | 170 | 1242 | Pré-falta | 25 | 39ND0701 | 109 | 265 | Pré-falta |
| 5 | 08aq0301 | 109 | 254 | Pré-falta | 26 | 40ND0801 | 128 | 1070 | Pré-falta |
| 6 | 09aq0302 | 108 | 215 | Pré-falta | 27 | 41ND0901 | 161 | 1042 | Pré-falta |
| 7 | 10aq0401 | 1123 | 1120 | Ponto Falta | 28 | 42ND1001 | 1278 | 1275 | Ponto Falta |
| 8 | 12aq0501 | 1172 | 1169 | Ponto Falta | 29 | 43ND1101 | 113 | 254 | Pré-falta |
| 9 | 13aq0601 | 119 | 791 | Pré-falta | 30 | 45ND1201 | 1032 | 1214 | Pré-falta |
| 10 | 14aq0701 | 724 | 721 | Ponto Falta | 31 | 46ND1301 | 124 | 258 | Pré-falta |
| 11 | 15aq0801 | 848 | 845 | Ponto Falta | 32 | 47ND1401 | 128 | 1353 | Pré-falta |
| 12 | 17aq0901 | 857 | 854 | Ponto Falta | 33 | 49ND1502 | 176 | 173 | Ponto Falta |
| 13 | 18aq0902 | 129 | 243 | Pré-falta | 34 | 51ND1602 | 297 | 877 | Pré-falta |
| 14 | 22aq1001 | 1254 | 1251 | Ponto Falta | 35 | 53W10102 | 125 | 744 | Pré-falta |
| 15 | 23aq1101 | 773 | 1209 | Pré-falta | 36 | 56W10105 | 143 | 746 | Pré-falta |
| 16 | 25K60101 | 108 | 257 | Pré-falta | 37 | 59W10301 | 481 | 626 | Pré-falta |
| 17 | 26K60201 | 1175 | 1172 | Ponto Falta | 38 | 60W10501 | 860 | 857 | Ponto Falta |
| 18 | 27K60202 | 164 | 258 | Pré-falta | 39 | 61W10701 | 119 | 286 | Pré-falta |
| 19 | 28K60301 | 134 | 516 | Pré-falta | 40 | 62W10801 | 112 | 1424 | Pré-falta |
| 20 | 29ND0201 | 125 | 274 | Pré-falta | 41 | 63W10901 | 192 | 246 | Pré-falta |
| 21 | 32ND0302 | 698 | 695 | Ponto Falta | | | | | |

4.5 Classificação de faltas

Na Seção 2.4.6, descrevemos os atributos e os classificadores utilizados nos trabalhos de Gomes et al (2013), para classificação de faltas em linhas de transmissão de energia e em síntese, os resultados divulgados, por Gomes et al (2013) em artigo *Detection and Classification of Faults in Power Transmission Lines Using Functional Analysis and Computational Intelligence*, *IEEE Transactions on Power Delivery*, para os classificadores estudados estão na Tabela 4.7, a seguir. Estes resultados se referem a um ajuste de modelos utilizando a ferramenta Weka, versão 3.6.4. Em seguida, os autores optaram pelo classificador *Bayesian Networks* para aplicação em dados reais.

Tabela 4.7: Tabela de Classificação

| Classificador | Acurácia (%) |
|-------------------|--------------|
| Bayesian Networks | 98,0 |
| kNN k=1 | 97,4 |
| DTNB | 97,2 |
| kNN k=3 | 95,5 |
| Decision Trees | 95,2 |

Fonte: Gomes et al (2013)

Restringindo-se ao classificador *Bayesian Networks* ou *BayesNet*, com aplicação na base de dados reais, previamente descrita, a matriz de confusão associada é dada na Tabela 4.8, em que 63 exemplos foram classificados corretamente, o que corresponde a uma exatidão global (G) de 76,83% de acerto. O classificador é capaz de classificar corretamente todas as condições normais de operação (OP), no entanto,

as outras situações que são, de expressiva importância, o classificador usado pelos autores não conseguiu distinguir com grande precisão as situações apresentadas.

Tabela 4.8: Matriz de confusão

| OP | AQ | K6 | ND | W1 | |
|----|----|----|----|----|----|
| 41 | 0 | 0 | 0 | 0 | OP |
| 0 | 11 | 0 | 4 | 2 | AQ |
| 0 | 0 | 2 | 3 | 1 | K6 |
| 0 | 4 | 0 | 7 | 2 | ND |
| 0 | 0 | 2 | 1 | 2 | W1 |

Fonte: *Gomes et al (2013)*

4.5.1 Definição das Amostras de treinamento e validação

Em cada banco por conter informações de dados em condições normais e sob condições de falta, obteve-se o ponto de falta usando o algoritmo proposto na Figura 3.17.

Em seguida dividiu-se em dois cada banco de dados, ficando então 41 em condições normais e 41 sob condições de falta, totalizando 82 bancos.

Posteriormente foi obtido as amostras de treinamento e validação, com a finalidade de calcular a distância do ponto à elipse funcional, da seguinte forma:

(a) Amostras de treinamento

- **Para os bancos em condições normais:** usou-se as 43 amostras iniciais (da amostra 1 a 43) para estimar a elipse funcional e os 43 pontos seguintes, (da amostra 44 a 86), para calcular a distância do ponto em condições normais à elipse estimada;
- **Para os bancos em condições de falta:** usou-se as 43 amostras (da amostra 1 a 43), em condições normais, para estimar a elipse funcional e 48 amostras iniciais após o ponto de falta, ou seja, após o ponto de falta f , ou seja $f+48$, para obter o distância do ponto sob falta à elipse estimada sob condição normal.

(b) Amostras de validação.

- **Para os bancos em condições normais:** usou-se 43 amostras (da amostra 87 a 129) para estimar a elipse funcional e outros 43 pontos em condições normais, (da amostra 130 à 172), para calcular a distância do ponto à elipse estimada.

- **Amostra em condições de falta:** usou-se as 43 amostras em condições normais (da amostra número 87 a 129) para estimar a elipse funcional e outras 48 amostras sob falta, após ($f+48$), ou seja, após o ponto de falta f , da amostra ($f+49$) a ($f+96$), para calcular a distância do ponto à elipse estimada.

Assim sendo, obteve-se duas tabelas: uma com os atributos de *treinamento* e outra com os atributos de *validação*. Cada tabela com 82 linhas e 10 colunas. As colunas correspondem aos atributos, (ver Seção 3.8.2), calculados sobre as distâncias métricas.

4.5.2 Resultados dos classificadores

Neste trabalho, utilizando-se os atributos discutidos na Seção 3.8, como entrada nos classificadores, e ampliando a gama de classificadores usados por Gomes et al (2013), temos os resultados constantes da Tabela 4.9, a seguir. As matrizes de confusão dadas, têm nos rótulos das linhas, os valores reais, e nos rótulos da coluna os valores estimados.

Tabela 4.9: Percentuais de exatidão e matriz de confusão

| Classificador | Exatidão G (%) | Matriz de Confusão | | | | | |
|----------------|----------------|--------------------|-----------|-----------|-----------|-----------|-----------|
| | | <i>OP</i> | <i>AQ</i> | <i>K6</i> | <i>ND</i> | <i>W1</i> | |
| kNN k=1 | 67,08 | 41 | 2 | 0 | 7 | 1 | <i>OP</i> |
| | | 0 | 6 | 0 | 3 | 1 | <i>AQ</i> |
| | | 0 | 0 | 1 | 0 | 0 | <i>K6</i> |
| | | 0 | 3 | 2 | 2 | 0 | <i>ND</i> |
| | | 0 | 4 | 1 | 3 | 5 | <i>W1</i> |
| | | | | | | | |
| kNN k=2 | 65,85 | 39 | 5 | 0 | 8 | 2 | <i>OP</i> |
| | | 1 | 8 | 0 | 1 | 1 | <i>AQ</i> |
| | | 0 | 0 | 1 | 0 | 0 | <i>K6</i> |
| | | 0 | 1 | 2 | 4 | 1 | <i>ND</i> |
| | | 1 | 1 | 1 | 2 | 1 | <i>W1</i> |
| | | | | | | | |
| kNN k=3 | 64,63 | 41 | 5 | 0 | 9 | 2 | <i>OP</i> |
| | | 0 | 6 | 1 | 3 | 2 | <i>AQ</i> |
| | | 0 | 0 | 2 | 0 | 1 | <i>K6</i> |
| | | 0 | 3 | 1 | 2 | 0 | <i>ND</i> |
| | | 0 | 1 | 0 | 1 | 2 | <i>W1</i> |
| | | | | | | | |
| BayesNet | 63,41 | 41 | 1 | 2 | 4 | 6 | <i>OP</i> |
| | | 0 | 6 | 0 | 0 | 0 | <i>AQ</i> |
| | | 0 | 5 | 2 | 6 | 1 | <i>K6</i> |
| | | 0 | 0 | 0 | 3 | 0 | <i>ND</i> |
| | | 0 | 3 | 0 | 2 | 0 | <i>W1</i> |
| | | | | | | | |
| LDA | 56,09 | 39 | 3 | 1 | 6 | 3 | <i>OP</i> |
| | | 1 | 4 | 2 | 1 | 0 | <i>AQ</i> |
| | | 0 | 1 | 1 | 1 | 1 | <i>K6</i> |
| | | 1 | 2 | 0 | 1 | 2 | <i>ND</i> |
| | | 0 | 5 | 0 | 6 | 1 | <i>W1</i> |
| | | | | | | | |
| Decision Trees | 43,90 | 30 | 0 | 0 | 4 | 0 | <i>OP</i> |
| | | 11 | 2 | 0 | 4 | 2 | <i>AQ</i> |
| | | 0 | 4 | 2 | 0 | 0 | <i>K6</i> |
| | | 0 | 2 | 2 | 2 | 5 | <i>ND</i> |
| | | 0 | 7 | 0 | 5 | 0 | <i>W1</i> |
| | | | | | | | |

O classificador *BayesNet*, (Vijayarani e Muthulakshmi, 2013), utilizado por Gomes et al (2013), que for-

neceu melhores resultados, com índice de exatidão global de 76,83%, foi o melhor resultado obtido em dados reais, considerando os atributos de entrada definidos pelos autores, conforme Seção 2.4.6. Um caso particular de oito (08) situações que eram verdadeiramente descarga elétrica (ND) foram classificadas incorretamente como outras classes (OP, AQ, K6 e W1).

Utilizando os novos atributos propostos conforme nesta tese, conforme Seção 3.8.2, o classificador *BayesNet* obteve índice de exatidão global de 63,41%. Apenas uma situação que verdadeiramente era uma descarga elétrica (ND) foi classificada incorretamente como sendo uma condição normal de operação (OP), assim sendo, demonstra que os novos atributos propostos melhoram de maneira significativa o desempenho do classificador *BayesNet*.

O classificador *kNN*, com $k=1$ foi o classificador que apresentou o melhor resultado com índice de exatidão global (G) de 67,07%, utilizando dados reais. Observamos que à medida que o número de vizinhos aumento $K = 2,3$, o índice de classificação global (G) torna-se menos preciso.

Pontuamos que tal classificador *kNN* ($k=1$) apresentou desempenho próximo ao classificador *BayesNet*, tanto nos trabalhos de simulação apresentados por Gomes et al (2013), quanto em aplicações em dados reais.

O classificador *kNN* é um dos classificadores mais simples na área de mineração de dados, logo o custo computacional é extremamente baixo.

Para fins comparativos, avaliou-se a classificação utilizando a transformada de *wavelete* do sinal de corrente, (ver Seção 2.5), e utilizando apenas a fase A da corrente, extraíram-se os coeficientes (atributos) utilizando-os como entrada nos classificadores utilizados neste trabalho. A Tabela 4.10, a seguir, mostra os resultados comparativos: De maneira geral, a proposta desta tese em relação ao uso consagrado da

Tabela 4.10: Percentuais de exatidão e matriz de confusão

| Classificador | Exatidão G (%) Tabela 4.9 | Exatidão G (%) <i>wavelete</i> |
|----------------|------------------------------|-----------------------------------|
| kNN k=1 | 67,07 | 67,07 |
| kNN k=2 | 65,85 | 63,41 |
| kNN k=3 | 62,19 | 58,54 |
| BayesNet | 63,41 | 52,44 |
| LDA | 56,09 | 52,44 |
| Decision Trees | 43,90 | 50,00 |

transformada de *wavelete* apresenta resultados bastante próximos. Verificamos que metodologia proposta

apresenta bons resultados em relação aos classificadores mais simples (*kNN* e *LDA*) e como também aos mais complexos (*BayesNet* e *Decision Tree*).

Esses resultados podem ser considerados iguais aos melhores resultados obtidos pelos diversos métodos de classificação e diagnóstico de faltas em linhas de transmissão encontrados na literatura, especialmente $kNN=1$ (PATEL e PATEL,2012)

4.5.3 Validação cruzada: *leave-one-out*

Tabela 4.11: Validação Cruzada - *leave-one-out*

| Classificador | Exatidão G (%) | Kappa (%) |
|---------------|----------------|-----------|
| KNN 1 | 70,73 | 0,5617 |
| KNN 2 | 63,41 | 0,4592 |
| KNN 3 | 74,39 | 0,6078 |
| Decision Tree | 71,95 | 0,5638 |
| BayesNet | 62,19 | 0,3705 |
| LDA | 59,76 | 0,3503 |

Embora seja um método validação com vício inerente, é comum testar classificadores, considerando os atributos de *treinamento* e atributos de teste como sendo as mesmas informações. De maneira geral espera-se que se o classificador está classificando de maneira correta, então estes valores serão próximos de 100% de exatidão. A Tabela 4.12, a seguir mostra os resultados, em que podemos observar que o classificador que apresenta melhores resultados é o *Knn*.

Tabela 4.12: Validação Cruzada

| Classificador | Exatidão G (%) | Exatidão G (%) |
|---------------|----------------------|--------------------|
| | Dados de Treinamento | Dados de Validação |
| KNN 1 | 100,00 | 98,78 |
| KNN 2 | 86,59 | 84,14 |
| KNN 3 | 84,15 | 78,05 |
| Decision Tree | 78,05 | 73,17 |
| BayesNet | 74,39 | 73,17 |
| LDA | 52,44 | 57,32 |

4.6 Redução da dimensionalidade dos atributos: busca exaustiva

Percebemos que dependendo do classificador, o número de atributos influencia significativamente na exatidão da classificação. Para os classificadores Decision Tree, BayesNet e LDA sugerimos manter todos os

Tabela 4.13: Percentuais de exatidão dos melhores atributos

| Classificador | Estatística | Cobinação de melhores atributos | | | | | | | | | |
|---------------|-------------|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|------------------|
| | | $\binom{10}{1}$ | $\binom{10}{2}$ | $\binom{10}{3}$ | $\binom{10}{4}$ | $\binom{10}{5}$ | $\binom{10}{6}$ | $\binom{10}{7}$ | $\binom{10}{8}$ | $\binom{10}{9}$ | $\binom{10}{10}$ |
| kNN=1 | Mínimo | 58,54 | 57,32 | 58,54 | 59,76 | 59,76 | 64,63 | 65,85 | 65,85 | 65,85 | |
| | Média | 65,12 | 65,53 | 66,60 | 67,22 | 67,47 | 67,49 | 67,41 | 67,37 | 67,32 | 67,07 |
| | Máximo | 71,95 | 71,95 | 71,95 | 71,95 | 70,73 | 70,73 | 69,51 | 69,51 | 68,29 | |
| kNN=2 | Mínimo | 54,88 | 53,66 | 53,66 | 56,10 | 56,10 | 57,32 | 54,88 | 58,54 | 58,54 | |
| | Média | 62,20 | 62,98 | 64,27 | 64,08 | 64,20 | 64,25 | 64,20 | 63,66 | 62,56 | 65,85 |
| | Máximo | 69,51 | 69,51 | 70,73 | 70,73 | 70,73 | 70,73 | 69,51 | 69,51 | 65,85 | |
| kNN=3 | Mínimo | 46,34 | 54,88 | 54,88 | 57,32 | 57,32 | 60,98 | 60,98 | 60,98 | 62,20 | |
| | Média | 60,98 | 62,60 | 63,53 | 63,98 | 63,93 | 64,20 | 63,78 | 63,77 | 63,66 | 64,63 |
| | Máximo | 70,73 | 67,07 | 68,29 | 69,51 | 69,51 | 69,51 | 68,29 | 68,29 | 65,85 | |
| Decision Tree | Mínimo | 58,54 | 57,32 | 58,54 | 56,10 | 54,88 | 50,00 | 51,22 | 53,66 | 58,54 | |
| | Média | 85,98 | 80,60 | 76,43 | 73,49 | 71,47 | 67,89 | 65,57 | 63,63 | 63,29 | 63,41 |
| | Máximo | 95,12 | 92,68 | 90,24 | 85,37 | 82,93 | 79,27 | 78,05 | 70,73 | 68,29 | |
| Bayes Net | Mínimo | 48,78 | 48,78 | 48,78 | 48,78 | 48,78 | 48,78 | 50,00 | 50,00 | 53,66 | |
| | Média | 51,71 | 52,49 | 52,80 | 52,32 | 52,49 | 53,20 | 53,98 | 54,31 | 55,37 | 56,09 |
| | Máximo | 53,66 | 56,10 | 60,98 | 59,76 | 58,54 | 57,32 | 57,32 | 58,54 | 59,76 | |
| LDA | Mínimo | 20,73 | 37,80 | 37,80 | 41,46 | 41,46 | 41,46 | 41,46 | 41,46 | 41,46 | |
| | Média | 47,93 | 54,50 | 51,97 | 49,18 | 47,29 | 46,25 | 45,63 | 45,09 | 44,51 | 43,90 |
| | Máximo | 64,63 | 67,07 | 67,07 | 67,07 | 65,85 | 63,41 | 52,44 | 52,44 | 52,44 | |

atributos extraídos e para o classificador kNN, as estatísticas da Tabela 4.13 indicam que um subconjunto de 5 (cinco) atributos já fornecem uma boa classificação. Numa busca exaustiva, os atributos que formam o subconjunto de melhores atributos para classificação são: *mínimo*, *máximo*, *variância*, *desvio padrão* e *assimetria*

Por outro lado, percebemos que os classificadores Decision Tree, BayesNet e LDA, na presença de poucos atributos, conseguem distinguir com menor precisão as situações de falta.

Capítulo 5

Comentários e trabalhos futuros

Este Capítulo contém as considerações finais e uma proposta para trabalho futuro baseado em curvas principais (*Principal Curves*).

5.1 Comentários finais

Neste trabalho propomos métodos simples em termos de complexidade matemática, rápido do ponto de vista computacional e eficiente para detecção e classificação de faltas em linhas de transmissão. A contribuição significativa é a possibilidade de avaliação conjunta das três fases da linha de transmissão, uma vez, que determinada falta pode atingir somente umas das fases, logo, os métodos baseados somente em algumas fases não conseguiriam detectar tal anomalia. Todos os métodos propostos estão associados a algoritmos para sua implementação, logo, trabalhos futuros poderão ser realizados em outras linguagens de programação.

Outro comentário importante é que nas simulações e análises dos bancos de reais, realizadas no R, os algoritmos foram bastante rápidos, podendo a metodologia ser utilizada como um monitoramento *online*. Isso permitirá ao usuário final monitorar em tempo real a transmissão de energia elétrica.

5.2 Proposta de continuidade.

Como proposta de trabalhos futuros, sugerimos que a construção dos limites de confiança para monitoramento *online* seja avaliado utilizando curvas principais. Curvas principais foram originalmente definidas por Hastie (1984) e Hastie e Stuetzle (1989), cujo objetivo é capturar a noção de uma *curva suave*,

passando pelo "meio" de uma distribuição de dados. O uso de curvas principais eliminaria uma etapa importante no processamento dos dados, que é a rotação e a translação da curva cônica.

Percebam que a metodologia de curvas principais, independentemente da posição que se encontra a elipse, consegue passar uma curva pelo centro dos dados (ver Figura 5.1), logo, construir limites de confiança, partindo dessa premissa é uma sugestão desta tese.

Outra observação importante é que a análise não ficaria confinada à forma funcional da elipse padrão, pois curvas principais conseguem captar as deformações que porventura existam na elipse padrão.

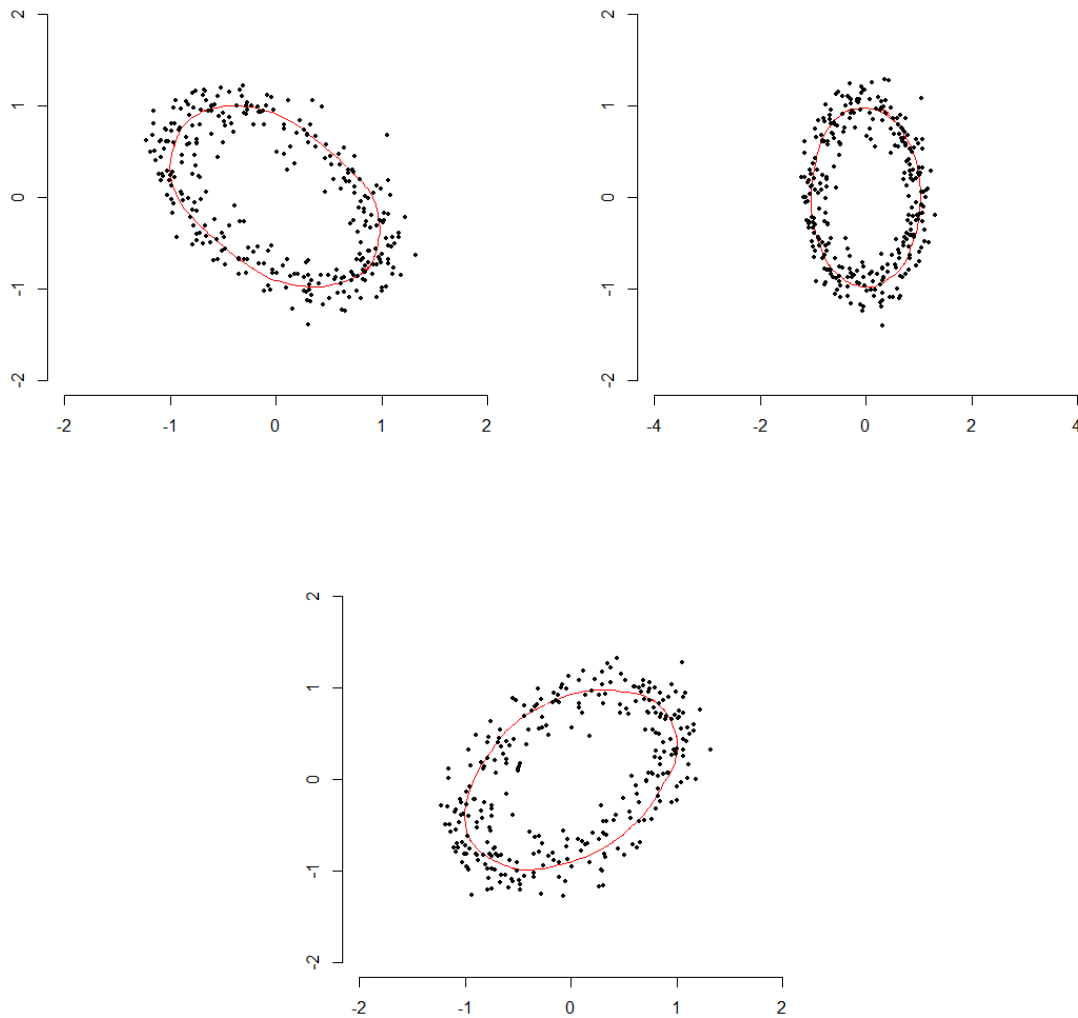


Figura 5.1: Ângulos de defasagem: $\varphi = \frac{2\pi}{3}$, $\varphi = \frac{\pi}{2}$ e $\varphi = \frac{\pi}{3}$, respectivamente

Referências Bibliográficas

- [1] Abu Shawiesh, M. O. A., Kibria, G., George, F. (2014). A Robust Bivariate Control Chart Alternative to the Hotelling's T2 Control Chart. *Quality and Reliability Engineering International*, 30(1), 25-35.
 - [2] Aggarwal, R. K., Blond, S. L., Beaumont, P., Baber, G., Kawano, F., and Miura, S., (2012). "High frequency fault location method for transmission lines based on artificial neural network and genetic algorithm using current signals only". *Developments in Power Systems Protection. 11th International Conference on*.
 - [3] Aha, D., W., Bankert, R. L. (1996). A comparative evaluation of sequential feature selection algorithms, *Learning from Data*. Springer New York, pp. 199-206.
 - [4] Alamedine, D., Khalil, M., Marque, C. (2013). Comparison of different EHG feature selection methods for the detection of preterm labor. *Computational and mathematical methods in medicine*, v. 2013.
 - [5] Awan Z. A., Ramos P., M., (2010). "Improvement of phase difference estimation using modified ellipse fit method". *IEEE Instrum Meas Technol Conf*; 10(15):78-81.
 - [6] Bassett Jr, G., and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, 73(363), 618-622.
 - [7] Barrodale, I., Roberts, F. D. (1973). An improved algorithm for discrete L1 linear approximation. *SIAM Journal on Numerical Analysis*, 10(5), 839-848.
 - [8] Bellman, R. (1961). *Adaptative control processes*.
 - [9] Casella, G. and Berger, R. L., (2001), "Statistical Inference", 2nd ed. Duxbury Press.
 - [10] Chamié Filho, R. H., Carvalho, L. F., Machado, P. L., de Oliveira, R. M. (2013). Analysis of voltages induced on power outlets due to atmospheric discharges on Radio Base Stations. *Applied Mathematical Modelling*, 37(9), 6530-6542.
-

-
- [11] Chen, C., Wei, Y. (2005). Computational issues for quantile regression. *Sankhy: The Indian Journal of Statistics*, 399-417.
- [12] David, H. A. and Nagaraja, H. N. (2003). "Order Statistics". *Wiley Series in Probability and Statistics*. Doi:10.1002/0471722162.
- [13] Eissa, M. M. (2013). A new digital busbar protection technique based on frequency information during CT saturation. *International Journal of Electrical Power Energy Systems*, 45(1), 42-49.
- [14] Fisz, M. (1963). "Probability theory and mathematical statistics". New York: Wiley.
- [15] Fitzgibbon, A. W., Pilu, M and Fischer, R. B.: (1996). "Direct least squares fitting of ellipses". In *Proc. of the 13th International Conference on Pattern Recognition*, 253–257.
- [16] Fuchs, Rubens Dario, (1979). *Transmissão de Energia Elétrica – Linhas Aéreas*, LTC Editora S.A., 2ª edição, ISBN: 85-216-0082-8.
- [17] Gomes, A. D S., Costa, A., M., Faria, T. G. A, Caminhas, W. M.,(2013). Detection and classification of faults in Power Transmission Lines Using Functional Analysis and Computational Inteligence. *IEEE Transactions on Power Delivery*.
- [18] Grossmann, A., Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.* 15, 723-736
- [19] G. Rosa, M.A. Costa, Robust functional analysis for fault detection in power transmission lines, *Applied Mathematical Modelling* (2016), <http://dx.doi.org/10.1016/j.apm.2016.05.042>.
- [20] Han, J., Kim, W. K., Lee, J. W., and Kim, C. H., (2012). "Fault type classification in transmission line using STFT." *Developments in Power Systems Protection, 2012. DPSP 2012. 11th International Conference on. IET, 2012.*
- [21] Halíř, R., Flusser, J., (1998). "Numerically stable direct least squares fitting of ellipses", *Proc. WSCG' 98*, pp. 125-132, University of West Bohemia, Campus Bory, Czech Republic.
- [22] Hastie, T. J. (1984). *Principal Curves and Surfaces*, unpublished Ph. D thesis.
- [23] Hastie, T., Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, 84(406), 502-516.
- [24] Hao, L., Naiman, D. Q. (2007). *Quantile regression* (No. 149). Sage.
- [25] Hotelling, H. (1992). The generalization of Student's ratio (pp. 54-65). Springer New York.
-

-
- [26] Hua, L., Baoqun, Z., Hong, Z. (2008, July). Recognition and classification of power quality event in power system using wavelet transformation. In Control Conference, 2008. CCC 2008. 27th Chinese (pp. 43-46). IEEE.
- [27] Inácio, M. J. (2010). Detecção e classificação de faltas em linhas de transmissão utilizando transformada wavelet e rede lógica neurofuzzy com aprendizado participativo. Engenharia elétrica, Universidade Federal de Minas Gerais. Tese de Doutorado.
- [28] Jwad, D. M., and Lefley., P. W. (2014). "Evaluation Studies of Combined Wavelet and Neural Network Applications in High Voltage Transmission Line Protection."12-96.
- [29] Koenker, R. (2005). Quantile regression (No. 38). Cambridge university press.
- [30] Koenker, R., Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50.
- [31] Koenker, R. W., d'Orey, V. (1987). Algorithm AS 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3), 383-393.
- [32] Lay, D., C., Lay, S., R., McDonald, J., J., (2015) "Linear Algebra and Its Applications", (5th Edition), Hardcover.
- [33] Malathi, V., Marimuthu, N. S., Baskar, S. (2010). Intelligent approaches using support vector machine and extreme learning machine for transmission line protection. *Neurocomputing*, 73(10), 2160-2167.
- [34] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7), 674-693.
- [35] Makming, P., Bunjongjit, S., Kunakorn, A., Jiriwibhakorn, S., and Kando, M., (2002). Fault diagnosis in transmission lines using wavelet transform analysis. In *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES (Vol. 3, pp. 2246-2250)*. IEEE
- [36] Miller, R. G. (1981). Simultaneous statistical inference. Springer.
- [37] Montgomery, D. C. (2007). Introduction to statistical quality control. John Wiley Sons.
- [38] Morales, J. A., Orduna, E., Rehtanz, C., Cabral, R. J., and Bretas, A. S., (2015). Comparison between Principal Component Analysis and Wavelet Transform 'Filtering Methods for Lightning Stroke Classification on Transmission Lines. *Electric Power Systems Research*, 118, 37-46.
-

-
- [39] Morales, J. , Orduna, E., (2013). Patterns extraction for lightning transmission lines protection based on principal component analysis, *IEEE Latin America Trans.* 11(1).
- [40] Oliveira, C. C. B., Schmidt, H. P., Kagan, N., Robba, E. J. (2000). *Introdução a sistemas elétricos de potência: componentes simétricas*. São Paulo: Edgard Blücher, 2a edição.
- [41] Osborne, M. R., Watson, G. A. (1971). On an algorithm for discrete nonlinear L1 approximation. *The Computer Journal*, 14(2), 184-188.
- [42] Panda, M., and Patra, M. R. (2009, January). Semi-Naïve Bayesian method for network intrusion detection system. In *Neural information processing* (pp. 614-621). Springer Berlin Heidelberg.
- [43] Patel, M., Patel, R. N. (2012). Fault detection and classification on a transmission line using wavelet multi resolution analysis and neural network. *International Journal of Computer Applications*, 47(22), 27-33.
- [44] Pilu, M., Fitzgibbon, A. W., and Fisher, R. B. (1996). Ellipse-specific direct least-square fitting. In *Image Processing, 1996. Proceedings., International Conference on* (Vol. 3, pp. 599-602). IEEE.
- [45] Portnoy, S., Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science*, 12(4), 279-300.
- [46] Provete, D., B., Silva, F., R., Souza, T., G., (2011), *Estatística aplicada à ecologia usando o R*, URL https://cran.r-project.org/doc/contrib/Provete-Estatistica_aplicada.pdf.
- [47] R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- [48] Ramos, P. M., Janeiro, F. M., Tlemçani, M., and Serra, A. C. (2009). Recent developments on impedance measurements with DSP-based ellipse-fitting algorithms. *Instrumentation and Measurement, IEEE Transactions on*, 58(5), 1680-1689.
- [49] Rafinia, A. and Moshtagh, J., (2014). "A new approach to fault location in three-phase underground distribution system using combination of wavelet analysis with ANN and FLS." *International Journal of Electrical Power Energy Systems* 55 : 261-274.
- [50] Rokach, L., Maimon, O. (2014). *Data mining with decision trees: theory and applications*. World scientific.
- [51] Samantaray, S.R., (2009). "Decision tree based fault zone identification and fault classification in flexible AC transmissions-based transmission line." *IET-Gener Trans Distrib* ;3(5):425-36.
-

-
- [52] Santoso, S., Powers, E. J., Grady, W. M., Hofmann, P. (1996). Power quality assessment via wavelet transform analysis. *Power Delivery, IEEE Transactions on*, 11(2), 924-930.
- [53] Saribulut, L. A., "simple power factor calculation for electrical power systems", (2014). *International Journal of Electrical Power Energy Systems*, v. 62, p. 66-71.
- [54] Schoukens, J., Pintelon, R., Van Kamme, H. (1991, May). The interpolated fast Fourier transform: a comparative study. In *Instrumentation and Measurement Technology Conference, 1991. IMTC-91. Conference Record.*, 8th IEEE (pp. 358-364). IEEE.
- [55] Shaik, A. G., Pulipaka, R. R. V., (2015). A new wavelet based fault detection, classification and location in transmission lines. *International Journal of Electrical Power Energy Systems*, 64, 35-40.
- [56] Simens, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3), 751-754.
- [57] Sundaravaradan, N. A., Rajaraman, P., Reddy, M., and Mohanta, D. K. (2014). Wavelet based transmission line fault analysis: A literature survey. In *Environment and Electrical Engineering (EEEIC), 2014 14th International Conference on* (pp. 254-259). IEEE.
- [58] Theodoridis, S., Koutroumbas, K. (2003). *Pattern recognition*.
- [59] Tleis, N. D. (2008). *Power Systems Modelling and Fault Analysis - Theory and Practice*. Elsevier Ltd, Oxford, UK.
- [60] Upendar, J., C. P. Gupta, and G. K. Singh., (2012). "Statistical decision-tree based fault classification scheme for protection of power transmission lines." *International Journal of Electrical Power Energy Systems* 36.1: 1-12.
- [61] Vijayarani, S., Muthulakshmi, M. (2013). Comparative analysis of bayes and lazy classification algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(8), 3118-3124.
- [62] Winker, P., Helmut, L., e Staszewska-Bystrova, A. (2014). Confidence Bands for Impulse Responses: Bonferroni versus Wald. In *Annual Conference 2014 (Hamburg): Evidence-based Economic Policy* (No. 100597). Verein für Socialpolitik/German Economic Association.
- [63] Yaghoobi, R. (2014). EEG signal classification using Bayes and Naïve Bayes Classifiers and extracted features of Continuous Wavelet Transform. *Majlesi Journal of Multimedia Processing*, 3(1).
- [64] Yao, C., Lu, Z., Li, J., Xu, Y., Han, J. (2014). A subset method for improving linear discriminant analysis. *Neurocomputing*, 138, 310-315.
-

- [65] Yusuffa, A.A., Fei, C., Jimoha, A.A., and Munda, J.L., (2011). "Fault location in a series compensated transmission line based on wavelet packet decomposition and support vector regression", *Electric Power Systems Research* Volume 81, Issue 7, Pages 1258–1265
- [66] Zhang, H., Berg, A.C., Maire, M., Svm-knn, J.M.: Discriminative nearest neighbor classification for visual category recognition. In: *CVPR '06*, pp. 2126–2136. IEEE Computer Society, Los Alamitos, CA, USA (2006).
- [67] Zhao, J., Philip, L. H., Shi, L., Li, S. (2012). Separable linear discriminant analysis. *Computational Statistics Data Analysis*, 56(12), 4290-4300.
-

Apêndice A

Análise dos resíduos.

A.1 Correlação e teste de normalidade

As Tabelas A.1, A.2 e A.3 mostram a análise dos resíduos da Corrente e tensão nas Fases A, B e C .

Tabela A.1: Análise de Correlação e testes de normalidade - Fase A

| Ord. | Banco | Correlação | Teste Correlação valor p | Teste Nor. Corrente valor p | Teste Nor. Tensão valor p | Média Corrente | Desvio Padrão Corrente | Média Tensão | Desvio Padrão Tensão |
|------|----------|------------|-----------------------------|-----------------------------------|---------------------------------|-------------------|------------------------------|-----------------|----------------------------|
| 1 | 01aq0101 | 0,390486 | 0,000084 | 0,000000 | 0,530793 | 0,044857 | 0,089959 | -0,006906 | 0,053710 |
| 2 | 02aq0102 | 0,486570 | 0,000000 | 0,150867 | 0,000000 | 0,001630 | 0,017919 | 0,011080 | 0,031264 |
| 3 | 04aq0201 | 0,013751 | 0,894217 | 0,393148 | 0,000000 | -0,000066 | 0,007945 | -0,000887 | 0,004329 |
| 4 | 06aq0203 | 0,039885 | 0,699628 | 0,013013 | 0,000000 | -0,000035 | 0,007434 | -0,001249 | 0,005198 |
| 5 | 08aq0301 | -0,039219 | 0,70441 | 0,626018 | 0,000000 | 0,001111 | 0,008695 | -0,003281 | 0,009820 |
| 6 | 09aq0302 | 0,245231 | 0,016032 | 0,000000 | 0,017388 | -0,000961 | 0,005465 | 0,001194 | 0,010448 |
| 7 | 10aq0401 | 0,283265 | 0,005165 | 0,000000 | 0,419301 | 0,010493 | 0,028485 | 0,000681 | 0,011335 |
| 8 | 12aq0501 | 0,265683 | 0,008890 | 0,000000 | 0,056851 | -0,000616 | 0,003718 | 0,000018 | 0,006206 |
| 9 | 13aq0601 | 0,220345 | 0,030987 | 0,000000 | 0,007630 | 0,015515 | 0,033890 | 0,000821 | 0,012040 |
| 10 | 14aq0701 | 0,437926 | 0,000008 | 0,000000 | 0,542982 | 0,087957 | 0,158380 | -0,001965 | 0,021251 |
| 11 | 15aq0801 | 0,207967 | 0,042028 | 0,058528 | 0,000000 | 0,000095 | 0,006633 | -0,000009 | 0,006314 |
| 12 | 17aq0901 | 0,234707 | 0,021347 | 0,000000 | 0,053263 | -0,002282 | 0,009277 | 0,000827 | 0,011927 |
| 13 | 18aq0902 | 0,439516 | 0,000007 | 0,774701 | 0,000000 | 0,000795 | 0,012589 | 0,000586 | 0,005275 |
| 14 | 22aq1001 | -0,332975 | 0,000917 | 0,327689 | 0,000000 | 0,001232 | 0,013848 | -0,010389 | 0,023725 |
| 15 | 23aq1101 | 0,244935 | 0,016164 | 0,000966 | 0,000000 | -0,000512 | 0,015924 | 0,000003 | 0,009614 |
| 16 | 25K60101 | -0,054254 | 0,599581 | 0,000000 | 0,244495 | -0,005934 | 0,013193 | 0,000766 | 0,003601 |
| 17 | 26K60201 | 0,603425 | 0,000000 | 0,000147 | 0,000627 | 0,009588 | 0,029507 | 0,001786 | 0,066195 |
| 18 | 27K60202 | 0,347866 | 0,000515 | 0,000060 | 0,000000 | -0,002564 | 0,023712 | 0,024640 | 0,057269 |
| 19 | 28K60301 | 0,046183 | 0,655015 | 0,000000 | 0,000000 | 0,001315 | 0,014885 | -0,008973 | 0,025349 |
| 20 | 29ND0201 | 0,547610 | 0,000000 | 0,001904 | 0,000000 | 0,000411 | 0,013555 | -0,000115 | 0,013629 |
| 21 | 32ND0302 | 0,390486 | 0,000084 | 0,000000 | 0,530793 | 0,044857 | 0,089959 | -0,006906 | 0,053710 |
| 22 | 33ND0401 | 0,139959 | 0,173812 | 0,218489 | 0,000000 | 0,001860 | 0,019962 | -0,012154 | 0,028198 |
| 23 | 36ND0502 | 0,538918 | 0,000000 | 0,117619 | 0,000000 | -0,000837 | 0,023335 | 0,034762 | 0,063641 |
| 24 | 37ND0601 | -0,326052 | 0,001187 | 0,66368 | 0,000000 | 0,000036 | 0,011173 | -0,007154 | 0,013175 |
| 25 | 39ND0701 | 0,230622 | 0,023785 | 0,000000 | 0,024455 | -0,001689 | 0,005627 | 0,001265 | 0,010179 |
| 26 | 40ND0801 | 0,682553 | 0,000000 | 0,000066 | 0,000518 | -0,000437 | 0,028659 | -0,001491 | 0,024672 |
| 27 | 41ND0901 | 0,165633 | 0,106795 | 0,025707 | 0,000001 | 0,000233 | 0,011358 | -0,004778 | 0,011369 |
| 28 | 42ND1001 | 0,378049 | 0,000146 | 0,000000 | 0,079449 | 0,016736 | 0,037751 | -0,000043 | 0,015923 |
| 29 | 43ND1101 | 0,304706 | 0,002540 | 0,202702 | 0,000000 | 0,001180 | 0,011732 | 0,006852 | 0,022167 |
| 30 | 45ND1201 | -0,186867 | 0,068300 | 0,019793 | 0,000000 | 0,000616 | 0,012767 | -0,005246 | 0,014276 |
| 31 | 46ND1301 | 0,176465 | 0,085455 | 0,073088 | 0,000000 | 0,001793 | 0,015115 | -0,001302 | 0,013284 |
| 32 | 47ND1401 | 0,337278 | 0,000778 | 0,023020 | 0,000000 | 0,000883 | 0,009476 | 0,001731 | 0,009841 |
| 33 | 49ND1502 | -0,364517 | 0,000261 | 0,34799 | 0,000000 | 0,000439 | 0,011202 | -0,008917 | 0,014763 |
| 34 | 51ND1602 | 0,270130 | 0,007774 | 0,002721 | 0,000000 | 0,000440 | 0,009377 | 0,002605 | 0,010323 |
| 35 | 53W10102 | 0,481916 | 0,000001 | 0,000000 | 0,055382 | 0,027009 | 0,043188 | -0,003807 | 0,024647 |
| 36 | 56W10105 | 0,167817 | 0,102187 | 0,799703 | 0,000000 | 0,002034 | 0,021674 | -0,001487 | 0,022343 |
| 37 | 59W10301 | 0,503645 | 0,000000 | 0,000035 | 0,000000 | -0,007410 | 0,080729 | 0,045820 | 0,099817 |
| 38 | 60W10501 | 0,788066 | 0,000000 | 0,001604 | 0,000007 | -0,001930 | 0,070342 | -0,006067 | 0,086522 |
| 39 | 61W10701 | 0,722500 | 0,000000 | 0,000000 | 0,004275 | 0,003143 | 0,017843 | -0,001164 | 0,019193 |
| 40 | 62W10801 | -0,013621 | 0,895209 | 0,001361 | 0,000000 | 0,001602 | 0,012033 | -0,005412 | 0,018371 |
| 41 | 63W10901 | 0,636497 | 0,000000 | 0,000000 | 0,000000 | -0,002094 | 0,068608 | -0,005530 | 0,068744 |

Tabela A.2: Análise de Correlação e testes de normalidade - Fase B

| Ord. | Banco | Correlação | Teste Correlação valor p | Teste Nor. Corrente valor p | Teste Nor. Tensão valor p | Média Corrente | Desvio Padrão Corrente | Média Tensão | Desvio Padrão Tensão |
|------|----------|------------|-----------------------------|-----------------------------------|---------------------------------|-------------------|------------------------------|-----------------|----------------------------|
| 1 | 01aq0101 | 0,469341 | 0,000001 | 0,000000 | 0,218181 | 0,007539 | 0,032634 | -0,000005 | 0,025500 |
| 2 | 02aq0102 | -0,027280 | 0,79191 | 0,178476 | 0,000000 | 0,000170 | 0,009533 | -0,002865 | 0,005465 |
| 3 | 04aq0201 | 0,162044 | 0,114716 | 0,167888 | 0,000000 | -0,000168 | 0,007618 | 0,000504 | 0,006876 |
| 4 | 06aq0203 | 0,166400 | 0,10516 | 0,290349 | 0,000000 | -0,000706 | 0,007377 | -0,000148 | 0,007136 |
| 5 | 08aq0301 | 0,045551 | 0,659438 | 0,92327 | 0,000000 | 0,001060 | 0,008447 | -0,001112 | 0,012416 |
| 6 | 09aq0302 | 0,421379 | 0,000019 | 0,000000 | 0,026470 | 0,003039 | 0,017242 | 0,000500 | 0,012063 |
| 7 | 10aq0401 | 0,301212 | 0,002862 | 0,000000 | 0,020576 | 0,002392 | 0,010138 | 0,000508 | 0,009230 |
| 8 | 12aq0501 | -0,012907 | 0,900674 | 0,000000 | 0,239728 | -0,001873 | 0,003582 | -0,000026 | 0,005983 |
| 9 | 13aq0601 | 0,250761 | 0,013729 | 0,000000 | 0,586452 | 0,003022 | 0,014704 | 0,001087 | 0,010350 |
| 10 | 14aq0701 | 0,422925 | 0,000018 | 0,000000 | 0,099927 | 0,065182 | 0,143138 | -0,000207 | 0,022030 |
| 11 | 15aq0801 | -0,110288 | 0,28475 | 0,157668 | 0,000000 | -0,000432 | 0,007337 | -0,001145 | 0,003179 |
| 12 | 17aq0901 | 0,147751 | 0,150834 | 0,000000 | 0,699752 | -0,002694 | 0,013282 | 0,001131 | 0,010643 |
| 13 | 18aq0902 | 0,439169 | 0,000008 | 0,708776 | 0,000000 | 0,001213 | 0,012187 | 0,002218 | 0,015212 |
| 14 | 22aq1001 | -0,240360 | 0,018330 | 0,310551 | 0,000000 | 0,000093 | 0,013375 | -0,008267 | 0,019296 |
| 15 | 23aq1101 | 0,139456 | 0,175379 | 0,000654 | 0,000000 | -0,000519 | 0,011808 | 0,001168 | 0,013982 |
| 16 | 25K60101 | -0,319581 | 0,001504 | 0,000000 | 0,038361 | -0,008131 | 0,018354 | 0,001016 | 0,004085 |
| 17 | 26K60201 | 0,550001 | 0,000000 | 0,000008 | 0,006613 | 0,000145 | 0,022975 | 0,001962 | 0,046645 |
| 18 | 27K60202 | 0,337681 | 0,000766 | 0,001356 | 0,000000 | -0,000835 | 0,023312 | 0,024175 | 0,054304 |
| 19 | 28K60301 | -0,184638 | 0,071719 | 0,006654 | 0,000000 | 0,000488 | 0,010919 | -0,004897 | 0,018118 |
| 20 | 29ND0201 | 0,524598 | 0,000000 | 0,15492 | 0,000105 | 0,000537 | 0,012117 | -0,000186 | 0,012011 |
| 21 | 32ND0302 | 0,469341 | 0,000001 | 0,000000 | 0,218181 | 0,007539 | 0,032634 | -0,000005 | 0,025500 |
| 22 | 33ND0401 | -0,231975 | 0,022952 | 0,021235 | 0,000000 | 0,001669 | 0,011538 | -0,008806 | 0,029087 |
| 23 | 36ND0502 | 0,433992 | 0,000010 | 0,864619 | 0,000000 | -0,001140 | 0,024994 | 0,001612 | 0,064268 |
| 24 | 37ND0601 | 0,002039 | 0,984268 | 0,591893 | 0,000000 | -0,000162 | 0,011178 | -0,006682 | 0,013132 |
| 25 | 39ND0701 | 0,412024 | 0,000030 | 0,000000 | 0,048113 | 0,006465 | 0,020993 | 0,001182 | 0,011827 |
| 26 | 40ND0801 | 0,596591 | 0,000000 | 0,000002 | 0,000038 | -0,000417 | 0,023176 | -0,001711 | 0,023497 |
| 27 | 41ND0901 | 0,029421 | 0,77599 | 0,016795 | 0,000005 | -0,000042 | 0,012672 | -0,003536 | 0,010771 |
| 28 | 42ND1001 | 0,382379 | 0,000120 | 0,000000 | 0,035532 | 0,015676 | 0,034403 | 0,000559 | 0,014349 |
| 29 | 43ND1101 | 0,268347 | 0,008206 | 0,024222 | 0,000000 | 0,001104 | 0,010601 | 0,007282 | 0,024517 |
| 30 | 45ND1201 | -0,390124 | 0,000085 | 0,392697 | 0,000000 | 0,001191 | 0,011807 | -0,009534 | 0,017258 |
| 31 | 46ND1301 | -0,005731 | 0,955809 | 0,213719 | 0,000000 | 0,001568 | 0,013017 | -0,003062 | 0,012031 |
| 32 | 47ND1401 | 0,337014 | 0,000786 | 0,362878 | 0,000000 | 0,001137 | 0,008216 | -0,000643 | 0,007051 |
| 33 | 49ND1502 | -0,512180 | 0,000000 | 0,000000 | 0,000000 | 0,001150 | 0,010634 | -0,007731 | 0,014913 |
| 34 | 51ND1602 | 0,334739 | 0,000858 | 0,196253 | 0,000000 | 0,001004 | 0,009008 | 0,002246 | 0,009277 |
| 35 | 53W10102 | 0,596958 | 0,000000 | 0,000000 | 0,006728 | 0,002588 | 0,023829 | -0,000038 | 0,048790 |
| 36 | 56W10105 | 0,011456 | 0,911789 | 0,073728 | 0,000000 | 0,001085 | 0,016050 | -0,009526 | 0,022718 |
| 37 | 59W10301 | 0,569074 | 0,000000 | 0,176248 | 0,000004 | 0,002452 | 0,048720 | 0,000142 | 0,026723 |
| 38 | 60W10501 | 0,612913 | 0,000000 | 0,000000 | 0,000000 | -0,001780 | 0,073807 | -0,005342 | 0,084796 |
| 39 | 61W10701 | 0,297999 | 0,003190 | 0,000000 | 0,478339 | -0,004898 | 0,023871 | 0,001047 | 0,020751 |
| 40 | 62W10801 | 0,168693 | 0,100385 | 0,134173 | 0,000000 | 0,001490 | 0,013343 | -0,006478 | 0,024152 |
| 41 | 63W10901 | 0,664786 | 0,000000 | 0,000004 | 0,000678 | -0,002343 | 0,078033 | -0,003730 | 0,074986 |

Tabela A.3: Análise de Correlação e testes de normalidade - Fase C

| Ord. | Banco | Correlação | Teste Correlação valor p | Teste Nor. Corrente valor p | Teste Nor. Tensão valor p | Média Corrente | Desvio Padrão Corrente | Média Tensão | Desvio Padrão Tensão |
|------|----------|------------|-----------------------------|-----------------------------------|---------------------------------|-------------------|------------------------------|-----------------|----------------------------|
| 1 | 01aq0101 | 0,337257 | 0,000779 | 0,000000 | 0,180102 | 0,024227 | 0,055716 | -0,001356 | 0,040148 |
| 2 | 02aq0102 | 0,044252 | 0,668571 | 0,771129 | 0,000005 | 0,000588 | 0,011277 | -0,001048 | 0,022220 |
| 3 | 04aq0201 | 0,207728 | 0,042269 | 0,184834 | 0,000000 | -0,000228 | 0,007356 | 0,000924 | 0,007675 |
| 4 | 06aq0203 | 0,208261 | 0,041732 | 0,078595 | 0,000000 | 0,000004 | 0,008052 | 0,000836 | 0,009566 |
| 5 | 08aq0301 | 0,222368 | 0,029440 | 0,743257 | 0,000000 | 0,001082 | 0,008035 | -0,000938 | 0,011080 |
| 6 | 09aq0302 | 0,319504 | 0,001508 | 0,000000 | 0,664863 | 0,010640 | 0,033393 | 0,000320 | 0,011151 |
| 7 | 10aq0401 | 0,421764 | 0,000019 | 0,000000 | 0,831477 | 0,009180 | 0,024158 | 0,001308 | 0,012519 |
| 8 | 12aq0501 | 0,103372 | 0,316219 | 0,000000 | 0,013848 | -0,001057 | 0,005109 | -0,000117 | 0,005663 |
| 9 | 13aq0601 | 0,380774 | 0,000130 | 0,000000 | 0,559551 | 0,000967 | 0,007829 | 0,001312 | 0,008931 |
| 10 | 14aq0701 | 0,485195 | 0,000001 | 0,000000 | 0,195014 | 0,049898 | 0,083374 | -0,000825 | 0,007234 |
| 11 | 15aq0801 | 0,251078 | 0,013606 | 0,016456 | 0,000000 | 0,000069 | 0,008144 | 0,001900 | 0,010909 |
| 12 | 17aq0901 | 0,428084 | 0,000014 | 0,000000 | 0,134899 | -0,000641 | 0,015631 | 0,001454 | 0,014769 |
| 13 | 18aq0902 | 0,584858 | 0,000000 | 0,252865 | 0,000000 | -0,000518 | 0,016178 | 0,030041 | 0,057469 |
| 14 | 22aq1001 | -0,503720 | 0,000000 | 0,082103 | 0,000000 | 0,001295 | 0,013601 | -0,013473 | 0,025698 |
| 15 | 23aq1101 | 0,194344 | 0,05778 | 0,000118 | 0,000000 | -0,000454 | 0,013764 | -0,001183 | 0,011320 |
| 16 | 25K60101 | -0,010078 | 0,922366 | 0,000000 | 0,873118 | -0,002471 | 0,014510 | 0,000404 | 0,007903 |
| 17 | 26K60201 | 0,449773 | 0,000004 | 0,000000 | 0,001717 | 0,014491 | 0,034417 | 0,000113 | 0,055546 |
| 18 | 27K60202 | 0,310041 | 0,002111 | 0,022635 | 0,000000 | 0,000539 | 0,017826 | 0,010257 | 0,028945 |
| 19 | 28K60301 | -0,206392 | 0,043643 | 0,002305 | 0,000000 | 0,000737 | 0,009770 | -0,005429 | 0,018863 |
| 20 | 29ND0201 | 0,626026 | 0,000000 | 0,000493 | 0,000004 | 0,000445 | 0,015205 | -0,000240 | 0,012918 |
| 21 | 32ND0302 | 0,337257 | 0,000779 | 0,000000 | 0,180102 | 0,024227 | 0,055716 | -0,001356 | 0,040148 |
| 22 | 33ND0401 | 0,095114 | 0,356632 | 0,031098 | 0,000000 | 0,001648 | 0,017726 | -0,010651 | 0,025022 |
| 23 | 36ND0502 | 0,540357 | 0,000000 | 0,055072 | 0,000000 | -0,000882 | 0,020496 | 0,039654 | 0,071662 |
| 24 | 37ND0601 | -0,050617 | 0,624305 | 0,180352 | 0,000000 | -0,000006 | 0,011850 | -0,005994 | 0,014666 |
| 25 | 39ND0701 | 0,311878 | 0,001980 | 0,000000 | 0,404867 | 0,010495 | 0,027575 | 0,000702 | 0,010564 |
| 26 | 40ND0801 | 0,701666 | 0,000000 | 0,000000 | 0,000000 | 0,000371 | 0,021508 | -0,002342 | 0,035328 |
| 27 | 41ND0901 | 0,304290 | 0,002577 | 0,000016 | 0,000007 | -0,000344 | 0,011437 | -0,004947 | 0,011867 |
| 28 | 42ND1001 | 0,379513 | 0,000137 | 0,000000 | 0,661508 | 0,020498 | 0,044202 | -0,000095 | 0,017318 |
| 29 | 43ND1101 | 0,350254 | 0,000468 | 0,179548 | 0,000000 | 0,001093 | 0,011842 | 0,012168 | 0,031527 |
| 30 | 45ND1201 | -0,208726 | 0,041267 | 0,036325 | 0,000000 | 0,000806 | 0,010164 | -0,007832 | 0,015665 |
| 31 | 46ND1301 | 0,186247 | 0,069237 | 0,517622 | 0,000000 | 0,001308 | 0,013603 | 0,000462 | 0,011310 |
| 32 | 47ND1401 | 0,004879 | 0,962374 | 0,890096 | 0,000001 | 0,001068 | 0,007877 | -0,001778 | 0,004938 |
| 33 | 49ND1502 | -0,210524 | 0,039512 | 0,278595 | 0,000000 | 0,000452 | 0,007499 | -0,006171 | 0,012293 |
| 34 | 51ND1602 | 0,314168 | 0,001826 | 0,000899 | 0,000000 | 0,001137 | 0,013484 | 0,004413 | 0,016579 |
| 35 | 53W10102 | 0,444658 | 0,000006 | 0,000000 | 0,002196 | 0,046773 | 0,069972 | -0,005766 | 0,031151 |
| 36 | 56W10105 | 0,463160 | 0,000002 | 0,083735 | 0,000000 | 0,001327 | 0,015410 | 0,000166 | 0,014704 |
| 37 | 59W10301 | 0,395930 | 0,000065 | 0,004468 | 0,000000 | -0,008503 | 0,067473 | 0,050299 | 0,087254 |
| 38 | 60W10501 | 0,511770 | 0,000000 | 0,000000 | 0,000026 | -0,008005 | 0,096274 | -0,002024 | 0,084506 |
| 39 | 61W10701 | 0,225166 | 0,027407 | 0,000000 | 0,175531 | -0,003426 | 0,022462 | 0,001068 | 0,018163 |
| 40 | 62W10801 | 0,497440 | 0,000000 | 0,024424 | 0,000000 | 0,000309 | 0,014329 | 0,005236 | 0,031842 |
| 41 | 63W10901 | 0,662276 | 0,000000 | 0,000000 | 0,021799 | -0,002035 | 0,061594 | -0,001697 | 0,053084 |

Apêndice B

Demonstrações Matemáticas

Neste Apêndice B, apresentamos as demonstrações matemáticas sendo que na Seção B.1, demonstramos que o IC, proposto por Gomes et al (2013), não é uma elipse na sua forma padrão, e na Seção B.2 demonstramos a validade da Equação (3.26), apresentada na Seção 3.4.2.

B.1 Comparação do IC proposto por Gomes et al. (2013) e a elipse padrão.

Com o intervalo de confiança construído na forma proposta por Gomes et al. (2013), pode-se afirmar que todos os pontos da elipse encontram-se no intervalo obtido nas proximidades dos pontos $[x=0, y=b]$ ou $[x=a, y=0]$, ou seja, a proposta dos autores só garante a existência do intervalo de confiança sob as retas suporte que definem o plano cartesiano, sendo que fora do plano do ponto cartesiano, simplesmente assume-se que o intervalo é verdadeiro.

Assim sendo, não se provou que todos os pontos de elipse padrão, que estão a uma distância ortogonal dos limites construídos do intervalo, também estejam contidos no intervalo, ou seja, a figura obtida pelos intervalos de confiança não são, a priori, uma elipse na forma padrão.

A demonstração a seguir mostra que existem pontos externos ao intervalo de confiança, mas que estão no interior de uma elipse padrão.

Seja $x_\omega = a \cos(\omega)$, $y_\omega = b \sin(\omega)$ e $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

O cálculo da reta *tangente* ao ponto **A** :

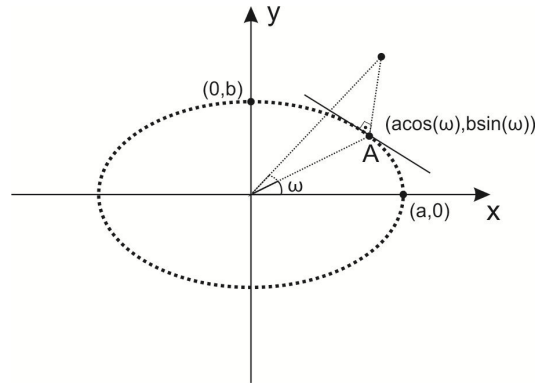


Figura B.1: Ponto **A** ortogonal à curva

Suponha $x > 0$ e $y > 0$, então $\frac{y^2}{b^2} = 1 - \frac{x^2}{a^2}$ e $y^2 = b^2 \left(1 - \frac{x^2}{a^2}\right)$, logo $y = b\sqrt{\left(1 - \frac{x^2}{a^2}\right)}$

Ou seja, $y = b \left(1 - \frac{x^2}{a^2}\right)^{\frac{1}{2}}$, logo a derivada primeira é:

$$\frac{dy}{dx} = -\frac{b}{a} \frac{x}{\sqrt{a^2 - x^2}} \text{ e } \frac{dy}{dx}|_A = -\frac{b}{a} \frac{a}{\sqrt{a^2 - x^2}} = -\frac{b \cos(\omega)}{a \sin(\omega)}$$

A reta perpendicular no ponto **A** possui inclinação dada por $m = \frac{1}{\frac{dy}{dx}|_A} = \frac{a \cos(\omega)}{b \sin(\omega)} = \frac{a}{b} \tan(\omega)$.

Seja o ponto **B** localizado na reta ortogonal ao ponto **A**, então $\tan(\theta) = \frac{a}{b} \tan(\omega)$.

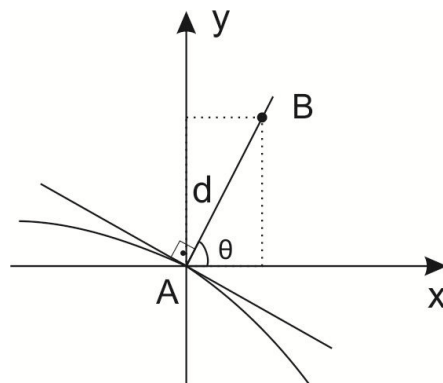


Figura B.2: Ponto **B** externo à curva

O par de coordenadas (x_B, y_B) pode ser definido como:

$$x_B = x_A + d \cos(\theta)$$

$$y_B = y_A + d \sin(\theta)$$

Então:

$$x_B = a \cos(\omega) + d \cos(\theta)$$

$$y_B = b \sin(\omega) + d \cos(\theta)$$

Análise: $\tan(\theta) = \frac{a}{b} \tan(\omega)$, logo $\frac{\tan(\theta)}{\tan(\omega)} = \frac{a}{b}$, ou seja, (x_B, y_B) não representam pontos em uma elipse de eixos $x_B = (a + d) \cos(\omega)$ e $y_B = (b + d) \sin(\omega)$, conforme proposto no artigo de Gomes (2013).

A figura a seguir mostra uma comparação com $0 \leq \omega \leq \frac{\pi}{2}$, ou seja, o primeiro quadrante de uma elipse em sua forma padrão.

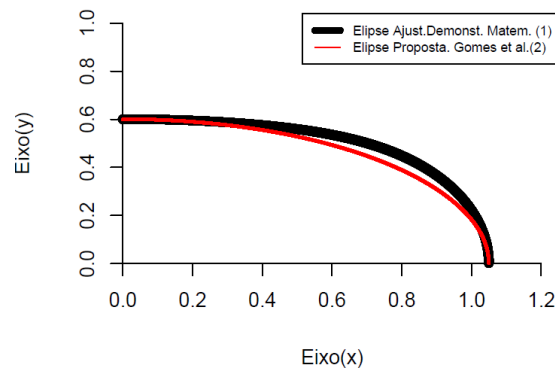


Figura B.3: Comparação de Curvas

Pela análise matemática e gráficos comparativos, verifica-se que o intervalo de confiança proposto pelos autores não é uma elipse na sua forma padrão, pois a curva, na região fora das proximidades dos eixos da elipse, está deslocada em relação ao eixo original da elipse padrão. No entanto, verificamos que as diferenças são mínimas e que o intervalo proposto pelo autor é mais exigente em relação à demonstração matemática.

B.2 Equações para estimação da corrente de pico I_0 em 2D e 3D

$$x = I_0 \text{sen}(\omega t) \quad (1)$$

$$y = I_0 \text{sen}(\omega t + \varphi) = I_0 \{ \text{sen} \omega t \cos \varphi + \sin \varphi \cos \omega t \} \quad (2)$$

$$z = I_0 \text{sen}(\omega t - \varphi) = I_0 \{ \text{sen} \omega t \cos \varphi - \sin \varphi \cos \omega t \}$$

Sejam as equações trigonométricas elementares:

$$\sin^2 \omega t + \cos^2 \omega t = 1 \text{ e } x^2 + \cos^2 \omega t = 1$$

Substituindo (1) em (2)

$$y = x \cos \varphi + I_0 \text{sen} \varphi \cos \omega t \Rightarrow \cos \omega t = \frac{1}{I_0 \text{sen} \varphi} (y - x \cos \varphi)$$

$$\begin{aligned} \cos^2 \omega t &= (I_0 \text{sen}^2 \varphi)^{-1} (y - x \cos \varphi)^2 \\ &= (y^2 - 2xy \cos \varphi + x^2 \cos^2 \varphi) (I_0 \text{sen}^2 \varphi)^{-1} \end{aligned}$$

$$\cos \omega t = \frac{y - x \cos \varphi}{I_0 \text{sen} \varphi}$$

Como $\cos^2 \omega t + \text{sen}^2 \omega t = 1$ e $\text{sen}^2 \omega t = \frac{x^2}{I_0^2}$
temos $\left(\frac{y - x \cos \varphi}{I_0 \text{sen} \varphi} \right)^2 + \frac{x^2}{I_0^2}$, logo $\frac{y^2 - 2xy \cos \varphi + x^2 \cos^2 \varphi}{I_0^2 \text{sen}^2 \varphi} + \frac{x^2}{I_0^2} = 1$
assim: $y^2 - 2xy \cos \varphi + x^2 \cos^2 \varphi + x^2 \text{sen}^2 \varphi = I_0^2 \text{sen}^2 \varphi$
ou seja assim: $y^2 - 2xy \cos \varphi + x^2 = I_0^2 \text{sen}^2 \varphi$

logo:

$$x^2 + y^2 - 2xy \cos \varphi - I_0^2 \text{sen}^2 \varphi = 0 \quad (3)$$

De maneira equivalente:

$$i_{a(t)}^2 + i_{b(t)}^2 - 2 i_{a(t)} i_{b(t)} \cos \varphi - I_0^2 \text{sen}^2 \varphi = 0$$

A seguir derivamos a equação para as três fases, ou seja, três dimensões-3D.

$$\text{De: } y = x \cos \varphi + I_0 \text{sen} \varphi \cos \omega t, \quad \text{onde: } \cos \omega t = \frac{1}{I_0 \text{sen} \varphi} (y - x \cos \varphi),$$

Logo:

$$z = x \cos \varphi - I_0 \text{sen} \varphi \frac{1}{I_0 \text{sen} \varphi} (y - x \cos \varphi)$$

$$\text{Então: } z = 2x \cos \varphi - y, \quad \text{segue que: } \cos \varphi = \frac{z + y}{2x}$$

$$\text{Substituindo } \cos \varphi = \frac{z + y}{2x} \quad \text{em } \textcircled{3}$$

Temos:

$$y^2 - 2xy \left(\frac{z + y}{2x} \right) + x^2 - I_0^2 \text{sen}^2 \varphi = 0$$

Logo

$$x^2 - xy - I_0^2 \text{sen}^2 \varphi = 0$$

ou de maneira equivalente

$$i_{a(t)}^2 - i_{b(t)} \times i_{c(t)} - I_0^2 \text{sen}^2(\varphi) = 0$$

Apêndice C

Banco de Dados

Neste Apêndice C descrevemos os bancos de dados reais utilizados nesta tese.

C.1 Bancos de dados utilizados

Cada arquivo contém uma matriz de dados de 3000×7 , ou seja três mil linhas por sete colunas, na seguinte ordem das colunas: Tensão A, Tensão B, Tensão C, Corrente A, Corrente B e Corrente C, com os 3 mil pontos coletados das oscilografias. As amostras utilizadas foram: 1 ; 2 ; 4 ; 6 ; 8 ; 9 ; 10 ; 12 ; 13 ; 14 ; 15 ; 17 ; 18 ; 22 ; 23 ; 25 ; 26 ; 27 ; 28 ; 29 ; 32 ; 33 ; 36 ; 37 ; 39 ; 40 ; 41 ; 42 ; 43 ; 45 ; 46 ; 47 ; 49 ; 51 ; 53 ; 56 ; 59 ; 60 ; 61 ; 62 ; 63.

Outros 22 bancos foram descartados, devido ao fato de que em alguns casos foi possível obter mais de um registro de falta para uma mesma ocorrência de falta em um linha de transmissão. Isto ocorreu pelo fato de na mesma linha existirem dois RDPs, um em cada estação(fonte) da linha. Ocorreu também do RDP registrar para a mesma falta mais de uma oscilografia referente a uma falta recorrente.

Para o primeiro caso, foi selecionado apenas a oscilografia que apresentou a maior sobrecorrente em uma das fases, considerando que esse seja o mais representativo para caracterizar a falta. No segundo caso, foi selecionado apenas o primeiro registro para as faltas recorrentes. Por isso, os bancos assinalados com (#) não foram utilizados, foram considerados redundantes.

Foram utilizadas as seguintes convenções, conforme Tabela C.1, a seguir:

Tabela C.1: Descrição dos Bancos de Dados

| Sigla | Descrição | Bancos | Qtd de Banco |
|-------|-------------------------|---------|--------------|
| OP | Operação Normal | 1 a 63 | 41 |
| AQ | Queimada | 1 a 24 | 15 |
| K6 | Entrelaçamento de Cabos | 25 a 28 | 4 |
| ND | Descarga Elétrica | 29 a 51 | 15 |
| W1 | Queda de Árvore | 52 a 63 | 7 |

| | | |
|--------------------|--------------------|--------------------|
| 1. 01aq0101.txt | 22. 22aq1001.txt | 43. 43ND1101.txt |
| 2. 02aq0102.txt | 23. 23aq1101.txt | 44. 44ND1102.txt # |
| 3. 03aq0103.txt # | 24. 24aq1102.txt # | 45. 45ND1201.txt |
| 4. 04aq0201.txt | 25. 25K60101.txt | 46. 46ND1301.txt |
| 5. 05aq0202.txt # | 26. 26K60201.txt | 47. 47ND1401.txt |
| 6. 06aq0203.txt | 27. 27K60202.txt | 48. 48ND1501.txt # |
| 7. 07aq0204.txt # | 28. 28K60301.txt | 49. 49ND1502.txt |
| 8. 08aq0301.txt | 29. 29ND0201.txt | 50. 50ND1601.txt # |
| 9. 09aq0302.txt | 30. 30ND0202.txt # | 51. 51ND1602.txt |
| 10. 10aq0401.txt | 31. 31ND0301.txt # | 52. 52W10101.txt # |
| 11. 11aq0402.txt # | 32. 32ND0302.txt | 53. 53W10102.txt |
| 12. 12aq0501.txt | 33. 33ND0401.txt | 54. 54W10103.txt # |
| 13. 13aq0601.txt | 34. 34ND0402.txt # | 55. 55W10104.txt # |
| 14. 14aq0701.txt | 35. 35ND0501.txt # | 56. 56W10105.txt |
| 15. 15aq0801.txt | 36. 36ND0502.txt | 57. 57W10106.txt # |
| 16. 16aq0802.txt # | 37. 37ND0601.txt | 58. 58W10107.txt # |
| 17. 17aq0901.txt | 38. 38ND0602.txt # | 59. 59W10301.txt |
| 18. 18aq0902.txt | 39. 39ND0701.txt | 60. 60W10501.txt |
| 19. 19aq0903.txt # | 40. 40ND0801.txt | 61. 61W10701.txt |
| 20. 20aq0904.txt # | 41. 41ND0901.txt | 62. 62W10801.txt |
| 21. 21aq0905.txt # | 42. 42ND1001.txt | 63. 63W10901.txt |

Apêndice D

Artigo decorrente

Robust functional analysis for fault detection
in power transmission lines

Aceito para publicação pela Applied Mathematical Modelling

Accepted Manuscript

Robust functional analysis for fault detection in power transmission lines

Gilmar Rosa , Marcelo Azevedo Costa

PII: S0307-904X(16)30301-8

DOI: 10.1016/j.apm.2016.05.042

Reference: APM 11195

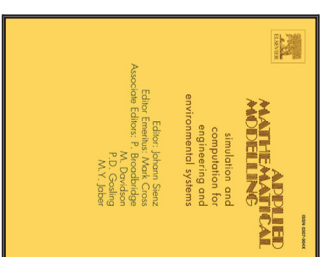
To appear in: *Applied Mathematical Modelling*

Received date: 25 August 2015

Revised date: 7 April 2016

Accepted date: 25 May 2016

Please cite this article as: Gilmar Rosa , Marcelo Azevedo Costa , Robust functional analysis for fault detection in power transmission lines, *Applied Mathematical Modelling* (2016), doi: 10.1016/j.apm.2016.05.042



Highlights

- A model for transmission lines under normal operating condition is revisited.
- Improved robust estimates are presented using constrained optimization.
- Upper and lower limits are proposed using non-linear quantile regression.
- Faults are detected using statistical control theory.
- Results shows improved performance for the new proposed method.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Robust functional analysis for fault detection in power transmission lines

Gilmar Rosa, Marcelo Azevedo Costa*

Abstract

Fault detection methods in power transmission lines aim to detect deviations of the electrical signals from the expected behavior of such signals under normal operating conditions. One approach is to model, as accurately as possible, the expected behavior of the electrical signals under normal operating conditions. Furthermore, even under normal conditions, electrical signals are subject to random noises. Therefore, upper and lower limits must be established. The larger the limits, the harder the fault detection. On the contrary, the narrower the limits the more likely to detect false faults. Functional analysis of power transmission lines was originally proposed to represent the behavior of the electrical signals and to estimate the upper and lower limits under normal operating conditions. Nonetheless, the originally proposed estimates are biased and rely on statistical assumptions that do not hold in practice. This work proposes new methods to estimate the parameters of the functional model and new upper and lower limits that do not rely on specific statistical assumptions. Simulated and real case results show that the proposed robust functional analysis reduces bias and provides more accurate false fault detection rates, as compared to the previous method.

Keywords: functional analysis, non-linear quantile regression, fault detection, power transmission lines.

1. Introduction

An Electric Power System (EPS) consists of many components that can be damaged over time, either by natural deterioration or misuse of equipment. In general, detection of faults in any EPS component in early stages is crucial, since faults may affect the supply of energy to consumers. Moreover, a fault in an EPS represents an abnormal condition such as an electrical failure in one or more elements in the electrical system.

Electricity is transmitted from generation stations to consumers along the transmission lines, which traverse great distances as illustrated in Fig. 1 (a). The transmission lines typically operate in a three phase system (A, B and C), conducting alternating currents. This configuration reduces energy loss in power transmission. In general, there is one transmission line for each phase, and each phase is represented by its voltage and current signals. Under normal

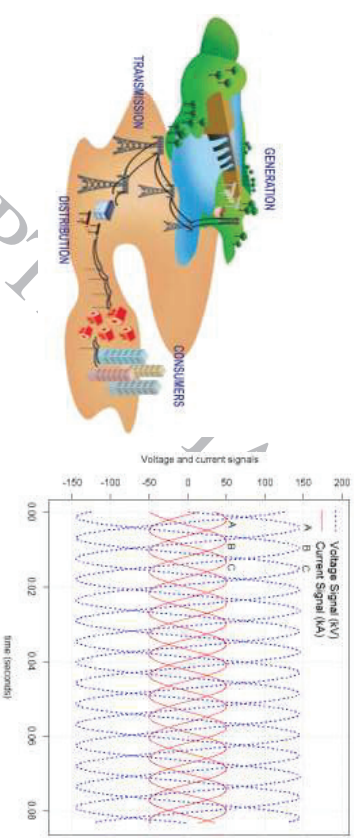
2

operating conditions, the expected signals of voltage and current are represented by cosine functions, shown in Eq.(1) and Eq.(2).

$$v(t) = V_0 \cdot \cos(\omega t) \quad (1)$$

$$i(t) = I_0 \cdot \cos(\omega t - \varphi) \quad (2)$$

where φ is the delay between the current signal $i(t)$ and the voltage signal $v(t)$. V_0 and I_0 are the peak values of the voltage and current signals, respectively, and $\omega = 2\pi f$ is the angular velocity. In the Brazilian system, $f = 60 \text{ Hz}$ and $\cos(\varphi)$ is also known as the power factor (PF). Eq. (1) and Eq. (2) show that the current and voltage signals are time dependent. Fig. 1 (b) shows the voltage and current signals of a three phase transmission line with phases A, B and C. Faults represent conditions in which signal behavior deviates significantly from expected behavior.



(a) The electrical power system is composed of generation, transmission and distribution systems. (b) The voltage and current signals of a three phase transmission line are represented by cosine functions.

Figure 1. Energy distribution system (a) and the expected behavior of the voltage and current signals for each phase (A, B and C) of a transmission line.

In general, wavelet transform and Fourier transform are the dominant methods for feature extraction of electrical signals in transmission lines. The features are used as inputs for different classification methods, in which Artificial Neural Networks (ANN) and Fuzzy Logic (FL) are the most common methods. For example, Rafinha and Moshagh [1] present a method for fault location in underground lines using wavelet transform and FL. Using simulated scenarios, the

3

authors concluded that wavelet transform effectively represent time characteristics of the line signals, whereas FL can locate and classify faults accurately. Aggarwal et al. [2] present a method for detection and location of faults based on ANN and genetic algorithm (GA), using only current signal information. The authors observed that the spectral energy and peak current are considerably different among different fault types. Sundaravaradan et al. [3] presents a survey of the most recent methods applied in fault detection in transmission lines.

Nonetheless, different statistical models have been presented. Morales [4], Samantafay [5] and Morales et al. [6] use Principal Component Analysis (PCA) for fault classification. Yussuffa et al. [7] use wavelet transform and statistical regression methods. Upendar et al. [8] proposed a detection and fault classification method using wavelet transform and regression trees (CART). The authors compared the proposed algorithm with the classification results using ANN. The simulation study shows that the proposed classification method is simpler than ANN and achieves high accuracy. Furthermore, statistical models using simple trigonometric functions, instead of complex mathematical equations, have been successfully applied to detect and classify faults in transmission lines [9,10,11].

In particular, Gomes et al. [10] introduced a novel monitoring system for a single phase of a transmission line in which the mean behaviors of the voltage and current signals under normal or nominal operating conditions are represented as a two dimensional ellipse. Furthermore, by assuming additive independent Gaussian noises for voltage and current signals, upper and lower limits for the ellipse are estimated. This method can be seen as a non-linear two-dimensional control chart and it allows monitoring of the electrical signal.

Control charts aim at determining if a manufacturing process is in a state of statistical control; i.e., the process is following its expected mean and expected variability under normal operating conditions. These charts can be extended to detect faults in transmission lines, as long as the expected mean and variability under normal operating conditions are known. Therefore, one may think of multivariate control charts for both voltage and current signals. In this case, the Hotelling's T^2 control chart is the most common used tool [12,13]. Robust alternatives to Hotelling's T^2 control chart also have been proposed [14]. In brief, the T^2 control chart is equivalent to the squared Mahalanobis distance, as shown in Eq. (3)

$$\chi_0^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \quad (3)$$

where n is the sample size, $\bar{\mathbf{x}}$ is the observed sample mean, $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix. For example, using current and voltage signals, $\mathbf{x}_i^T = [v_i(t), i_i(t)]$ and $\bar{\mathbf{x}}^T = [\bar{v}, \bar{i}] = \frac{1}{n} \sum_{t=1}^n v_i(t) / n$ and $\bar{i} = \frac{1}{n} \sum_{t=1}^n i_i(t) / n$. The parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ must be known in advance.

As shown in Eq. (3), it is commonly assumed that the mean vector is constant under normal operating conditions; i.e., does not change over time. On the contrary, the expected behavior of voltage and current signals under normal operating conditions does change over time, as

previously shown. In addition, the method is computationally simpler than most commonly applied fault detection methods in Electric Power Systems, such as Wavelets [1,2,4,7], Fourier Transforms [15] and models of Artificial Intelligence [2,16]; and, it achieves superior fault detection rates. Furthermore, mathematical and statistical modelling of transmission lines under normal operating conditions are much simpler than modelling specific fault conditions, such as transitory effects produced by atmospheric current discharges [17].

Functional analysis of power transmission lines aims to represent voltage and current signals under normal operating conditions using stochastic models, as accurately as possible. Consequently, faults can be detected as fast as possible, since upper and lower bounds for the signals under normal operating conditions can be statistically estimated. Furthermore, false faults can be controlled. Briefly, the faster a fault is detected the more likely the occurrence of false faults. This is a standard problem in statistical signal analysis, and often neglected by standard fault detection methods. Furthermore, estimates of the functional model may present bias because some of the statistical assumptions do not hold in practice, such as independent Gaussian noise errors. Therefore, the functional model proposed by Gomes et al. [10] has major limitations: it produces both higher false fault rates than originally expected and also biased estimates of the current and voltage signal parameters.

This work proposes improvements to the functional analysis model of transmission lines. A constrained linear optimization problem achieves more accurate estimates of current and voltage signal parameters. Using quantile regression and elliptical equations, new upper and lower bounds are estimated achieving a more precise false fault rate, as compared to the previous approach. Thus, this work evaluates the main limitations of the method proposed by Gomes et al. [10], which are: (i) the estimated confidence intervals are inconsistent with the α -level chosen by the user; and, (ii) the estimates of the parameters are biased. In addition, we propose new estimates for the parameters of the ellipse, and new upper and lower limit estimates using non-linear quantile regression [18,19]. Therefore, we do not rely on any statistical assumptions about noise distribution. In brief, to estimate the mean parameter using a sample of size n , one can solve the following optimization problem, $\min_{\hat{\mu}} \sum_{i=1}^n (x_i - \hat{\mu})^2$, which leads to $\hat{\mu} = \bar{x}$ where \bar{x} is the sample mean. Similarly, the estimated median is the solution of the following optimization problem: $\min_m \sum_{i=1}^n |x_i - m|$. In general, the p th quantile, hereafter named q , can be estimated as the solution of an optimization problem presented in Eq. (4).

$$\min_{\hat{\mu}} \sum_{x_i < q} (x_i - \hat{\mu})^2 + \sum_{x_i > q} (x_i - \hat{\mu})^2 \quad (4)$$

Eq. (4) can be applied to estimate quantile regression models, for example, a linear equation model can be estimated using Eq. (4) and $q(\beta_0, \beta_1) = \beta_0 + \beta_1 x$ (see [18,19]). In the present work, a non-linear equation is applied. It is assumed that both the upper and lower limits are represented by ellipse equations.

This paper is organized as follows: section 2 reviews the functional analysis for power transmission lines, presents new estimates for the functional model and proposes new upper and lower limits using non-linear quantile regression. Section 3 presents results using simulated and real case data sets. Discussion and conclusion are presented in section 4.

2. Materials and Methods

2.1 Monitoring statistic for transmission lines

Although the transmission line has three phases, a monitoring system can be applied to each phase individually, as if they were independent. For each phase, a control chart can be applied to voltage and current signals, separately. Gomes et al. [10] proposed a new approach in which both current and voltage signals are simultaneously controlled using a two-dimensional control region. Following Gomes et al. [10], under normal operating conditions and assuming no noise, voltage and current signals can be represented by the ellipse given by Eq. (5)

$$V_{r(t)}^2 - 2V_0I_0\cos(\varphi)v_{r(t)}i_{r(t)} + I_0^2i_{r(t)}^2 - V_0^2v_{r(t)}^2\sin^2(\varphi) = 0 \quad (5)$$

Eq. (5) can be applied directly to any point $P(v_{r(t)}, i_{r(t)})$. If the result is zero, then the point $(v_{r(t)}, i_{r(t)})$ belongs to the ellipse, regardless the value of time t . Furthermore, a control region is created assuming additive random variables for both voltage and current signals, as shown in Eq. (6).

$$\begin{aligned} v_{r(t)} &= V_0 \cdot \cos(\omega t) + \xi_V \cdot V_0 \\ i_{r(t)} &= I_0 \cdot \cos(\omega t - \varphi) + \xi_I \cdot I_0 \end{aligned} \quad (6)$$

Where ξ_V and ξ_I are two random variables, with means of zero and variances of σ_V^2 and σ_I^2 , respectively. They represent noise components associated to voltage and current signals, respectively.

The monitoring statistics are the standardized values of voltage and current. First, instant values of voltage and current signals are divided by their peak values: $X_{r(t)} = \frac{v_{r(t)}}{V_0}$, and $Y_{r(t)} = \frac{i_{r(t)}}{I_0}$. Second, the points are rotated using an angle of $\theta = \frac{\pi}{4}$, or $\theta = \frac{3\pi}{4}$. The standardized and rotation operations further simplify the equation of the ellipse, Eq. (6), which is reduced to Eq. (7).

$$\frac{x_{r(t)}^2}{a^2} + \frac{y_{r(t)}^2}{b^2} = 1 \quad (7)$$

Where $x_{r(t)}$ and $y_{r(t)}$ are the voltage and current signals after standardization and rotation, respectively; and a and b are constant values related to V_0 , I_0 and φ . An alternative representation of Eq. (7) is presented in Eq. (8).

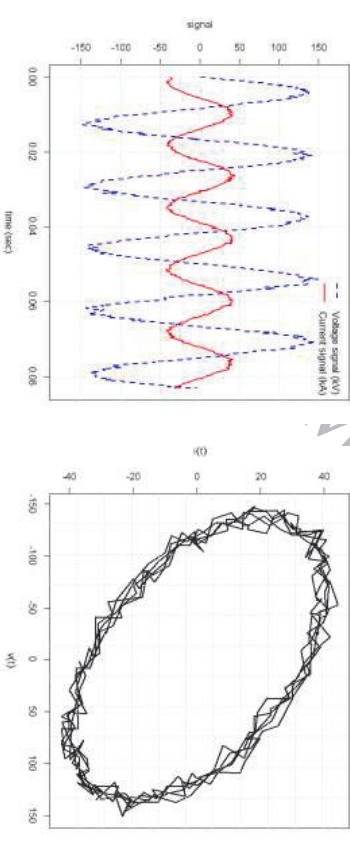
$$(8)$$

$$\begin{aligned} x_{r(t)} &= a \cdot \sin(\omega t) + \xi_V \\ y_{r(t)} &= b \cdot \cos(\omega t) + \xi_I \end{aligned}$$

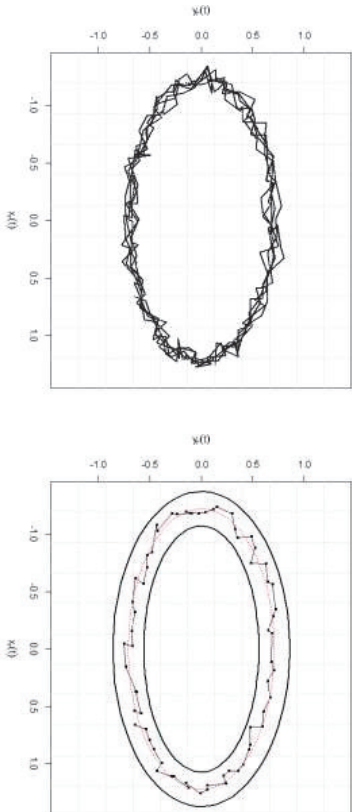
From Eq. (8), and assuming independent Gaussian variables for ξ_V and ξ_I , upper and lower limits for the control region are given by Eq. (9).

$$\begin{aligned} \frac{x_{r(t)}^2}{\left(a + Z_{\frac{\alpha}{2}} \cdot \sigma_V\right)^2} + \frac{y_{r(t)}^2}{\left(b + Z_{\frac{\alpha}{2}} \cdot \sigma_I\right)^2} &\leq 1 \\ \frac{x_{r(t)}^2}{\left(a - Z_{\frac{\alpha}{2}} \cdot \sigma_V\right)^2} + \frac{y_{r(t)}^2}{\left(b - Z_{\frac{\alpha}{2}} \cdot \sigma_I\right)^2} &\geq 1 \end{aligned} \quad (9)$$

Where $Z_{\frac{\alpha}{2}}$ is the Z -score statistic with $1 - \frac{\alpha}{2}$ confidence level. For instance, if $\alpha = 0.3\%$ then $Z_{\frac{\alpha}{2}} \approx 3$. Fig. 2 (a) shows voltage and current signals over time assuming stochastic error components. Fig. 2 (b) shows the elliptical behavior of voltage and current signals in one phase of a transmission line. Fig. 2 (c) shows the ellipse after standardization and rotation. Fig. 2 (d) shows upper and lower limits, assuming independent and Gaussian noises, under normal operating conditions.



(a) Current and voltage signals over time assuming stochastic noise components. (b) Voltage and current signals projected into the bi-dimensional space (voltage versus current).



(c) Voltage and current signals after standardization and rotation operations.

(d) Estimated upper and lower limits assuming stochastic error components under normal operating conditions.

Figure 2. Voltage and current signals over time (a). Elliptical behavior of voltage and current signals of a phase in a transmission line (b). Voltage and current signals after standardization and rotation (b). Upper and lower limits, assuming independent Gaussian noises under normal operating conditions.

2.1.1 Upper and lower limits estimates

Estimates for the V_0 , I_0 , and $\cos(\varphi)$ parameters have been proposed [10] based on the solution of a linear equation system, which is a simplified version of Eq. (5). In this case, all terms are divided by $V_0^2 \sin^2(\varphi)$. Let $t = 1, \dots, N$ be the time period in which the voltage and current signals are under normal operating conditions. Let $\mathbf{v} = [v^{(1)}, v^{(2)}, \dots, v^{(N)}]^T$ be the voltage signal column vector and $\mathbf{i} = [i^{(1)}, i^{(2)}, \dots, i^{(N)}]^T$ the current signal vector. Let $\mathbf{X} = [V^2 \mathbf{v} \mathbf{i}^T \mathbf{v} \mathbf{i}]$ be a matrix of dimensions $N \times 5$, and $\boldsymbol{\beta}^T = [\beta_1 \beta_2 \beta_3 \beta_4 \beta_5]$ be the vector of parameters ($\beta_{3 \times 1}$). The estimates of the parameters are the solution of the following linear equation: $\mathbf{X}\boldsymbol{\beta} = \mathbf{1}$, where $\mathbf{1}$ is the unitary vector. In this case, the solution can be written as $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{1}$. The estimates of V_0 , I_0 , and $\cos(\varphi)$ are shown in Eq. (10).

$$\begin{aligned} \cos(\varphi) &= \frac{-\beta_1}{\sqrt{\beta_1 \cdot \beta_3}} \\ I_0 &= \frac{1}{\sqrt{\beta_1 \cdot \sin^2(\varphi)}} \\ V_0 &= \frac{1}{\sqrt{\beta_3 \cdot \sin^2(\varphi)}} \end{aligned} \quad (10)$$

8

The estimated parameters shown in Eq. (10) are used to standardize both the voltage and current signals, and to estimate parameters a and b (see Eq. (7)). To estimate the variance parameters of the stochastic components, σ_v^2 and σ_i^2 , we first estimate the residuals of the standardized and rotated signals using a one dimensional optimization procedure for each point in the sample, as shown in Eq. (11). Thus, for each value of $t = 1, \dots, N$, voltage and current residuals are calculated by solving one non-linear optimization problem.

$$t_i^* = \arg \min_t \left\{ (x_{r(t)} - \hat{a} \cdot \sin(\omega t))^2 + (y_{r(t)} - \hat{b} \cdot \cos(\omega t))^2 \right\} \quad (11)$$

$$r_{x(t)} = x_{r(t)} - \hat{a} \cdot \sin(\omega t_i^*) \quad (12)$$

$$r_{y(t)} = y_{r(t)} - \hat{b} \cdot \cos(\omega t_i^*)$$

From the estimated residuals, the estimated variances are

$$\begin{aligned} \hat{\sigma}_v^2 &= \frac{\sum_{t=1}^N r_{x(t)}^2}{N-1} \\ \hat{\sigma}_i^2 &= \frac{\sum_{t=1}^N r_{y(t)}^2}{N-1} \end{aligned} \quad (13)$$

As will be shown in the simulation study, the proposed estimates have major limitations, such as assuming independence between the voltage and current residuals. Furthermore, the estimated values present some bias, and the Gaussian noise assumption does not hold in practice. Therefore, the confidence level of the control limits is much lower than the α -level chosen by the user, which leads to a higher level of false detection rates. Gomes et al. [10] proposed the use of residual percentiles in the upper and lower equations in order to overcome the Gaussian assumption. Nevertheless, this solution further compromises the α -level, which significantly deviates from the original value chosen by the user. We provide a robust approach to estimate the parameters of the ellipse and to estimate upper and lower limits.

2.2 Robust upper and lower limits using non-linear quantile regression

Our proposal does not rely on any stochastic assumption about the voltage and current noises. We do rely on a parametric form for upper and lower limits, which is the ellipse equation. Nevertheless, before estimating upper and lower limits, we propose a more efficient procedure to estimate the parameters of the ellipse equation.

Using the voltage and current signals, the general equation of a second order conic section is a second order polynomial equation, as shown in Eq. (14).

$$\beta_1 v_{(t)}^2 + \beta_2 v_{(t)} i_{(t)} + \beta_3 i_{(t)}^2 + \beta_4 v_{(t)} + \beta_5 i_{(t)} + \beta_6 = 0 \quad (14)$$

9

to submit to Applied Mathematical Modelling

Following Fitzgibbon et al. [20], Eq. (14) represents an ellipse if the following constraint is applied: $\beta_2^2 - 4\beta_1\beta_3 < 0$. Therefore, the parameters of the ellipse can be estimated solving the optimization problem shown in Eq. (15) (see Halir et al. [21]).

$$\min (\mathbf{X}\beta)^T \mathbf{X}\beta \quad (15)$$

subject to: $\beta^T \mathbf{C}\beta = 1$

Where $\mathbf{X} = [v^2 \ v \ i^2 \ v \ i \ 1]$, $\beta^T = [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5 \ \beta_6]$, and

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 2 & & & \\ 0 & -1 & 0 & \dots & & 0 \\ 2 & 0 & 0 & & & \\ \vdots & & & \ddots & & \vdots \\ \dots & & & & \dots & 0 \end{bmatrix}$$

Eq. (15) provides robust estimates for the parameters of the ellipse even, if partial data is available. Halir et al. [21] provide further information about computing aspects of the estimates.

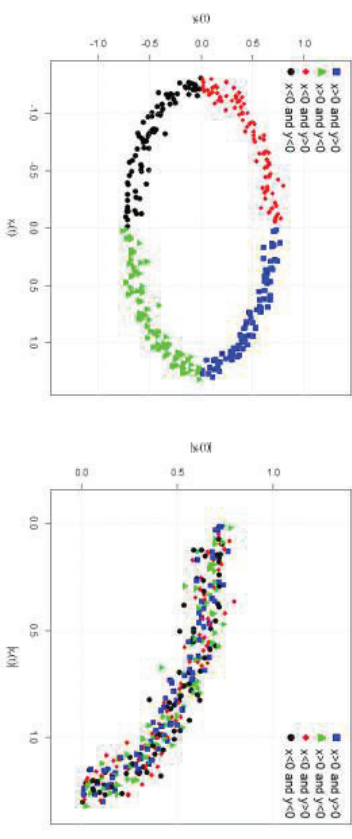
From Eq. (15) and using Eq. (10), new estimates for V_0 , I_0 and φ are achieved. A simulation study presented in section 3, shows that this approach provides more reliable estimates than the previous one.

From the estimated values, the voltage and current signals are standardized and rotated, which generates data that follows a reduced ellipse equation, as shown in Eq. (7) and Eq. (8). Since the data is centered at the origin and symmetrically dispersed with respect to the axis, we consider the absolute values of the signals, as shown in Fig. 3. Thus, for positive values of voltage and current signals, i.e., $|x_{r(t)}|$ and $|y_{r(t)}|$, Eq. (7) can be rewritten as:

$$y_{r(t)} = \frac{b}{a} \sqrt{a^2 - x_{r(t)}^2} \quad (16)$$

10

to submit to Applied Mathematical Modelling



(a) Voltage and current signals after standardization and rotation operators. (b) Absolute values of the standardized and rotated voltage and current signals. **Figure 3.** The non-linear quantile regression equation (Eq. (16)) is applied to the absolute values of the standardized and rotated signals (b), since they are symmetric with respect to the axis (a).

Different from the previous approach, the radii of the upper and lower limits are no longer estimated from stochastic components, i.e., the residuals. Nevertheless, our approach also assumes that both upper and lower limits are represented as ellipses. We propose to estimate separate values for the parameters a and b in Eq. (17); i.e., $a^{(p)}$ and $b^{(p)}$ are the parameters of the ellipse that leaves $100p\%$ of the points below the ellipse, $p \in [0, 1]$. Thus, for the upper limit $p = 0.975$, and for the lower limit $p = 0.025$. In this case, estimated upper and lower limits using quantile regression hold approximately 95% of the sample points under normal operating conditions. Upper and lower limits are estimated using Eq. (17).

$$(a^{(p)}, b^{(p)}) = \arg \min_{a,b} \frac{1-p}{n} \sum_{y_{r(t)} \leq q} |y_{r(t)} - q(x_{r(t); a, b})| + \frac{p}{n} \sum_{y_{r(t)} > q} |y_{r(t)} - q(x_{r(t); a, b})| \quad (17)$$

Where $q_i = q(x_{r(t); a, b}) = \frac{b}{a} \sqrt{a^2 - x_{r(t)}^2}$, and p represents the p th quantile selected by the user. Thus, using non-linear quantile regression, upper and lower limit equations with 95% of statistical confidence can be written as:

$$\frac{x_{r(t)}^2}{[a(0.975)]^2} + \frac{y_{r(t)}^2}{[b(0.975)]^2} \leq 1 \quad (18)$$

$$\frac{x_{r(t)}^2}{[a(0.025)]^2} + \frac{y_{r(t)}^2}{[b(0.025)]^2} \geq 1$$

11

to submit to Applied Mathematical Modelling

Eq. (18) shows that upper and lower limits are estimated by means of optimization problems rather than stochastic assumptions for the noise distribution.

2.2.1 Design of the control region

In practice, assuming a regular sample period of 32 points per cycle, a single sample is collected every 0.52 milliseconds, or 1,920 samples per second. In this case, due to the large number of samples in such short time, the monitoring of the signals is online. Gomes et al. [10] suggest using an additional sensitivity criteria of three consecutive points lying outside the control region, in order to detect true faults and minimize false fault detection rates. Thus, assuming a 3-sigma confidence level (99.7%) a false fault occurs with every 40,000,000 sampled points, or at every 6 hours, on average. Furthermore, the confidence level and the number of points lying outside the control region can be chosen properly in order to adjust the average run length (ARL) [22,23]. Fig. 4 illustrates the proposed fault detection method using the proposed lower and upper bound ellipse equations.

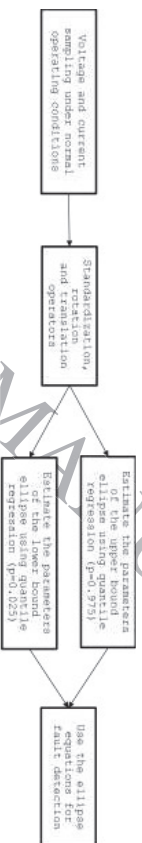


Figure 4. Flowchart of the proposed fault detection method using ellipse equations.

2.2.2 Computational algorithm for the two-dimensional monitoring system.

In addition to Fig. 4, the algorithm shown below provides detailed information of the estimates of the parameters of the upper and lower bound ellipses under normal operating conditions, and the applications of the model to detect faults.

Design of the control limits

- **Step 1:** Using a sample of size n of voltage and current signals under normal operating conditions, solve the linear optimization problem presented in Eq. (15). Then, apply Eq. (10) to the estimated coefficients in order to estimate parameters V_0 , I_0 and $\cos(\varphi)$.
- **Step 2:** Calculate the standardized voltage and current signals using the estimated voltage and current peak values, $X_{(i)} = \frac{v_{(i)}}{V_0}$, and $Y_{(i)} = \frac{i_{(i)}}{I_0}$.
- **Step 3:** Rotate the standardized signals by an angle $\theta = \frac{\pi}{4}$ or $\theta = \frac{3\pi}{4}$.
- **Step 4:** Calculate upper and lower values for $a^{(p)}$ and $b^{(p)}$ using Eq. (17). Use $p = 0.975$ for upper limit ($a^{(0.975)}$, $b^{(0.975)}$) and $p = 0.025$ for lower limit ($a^{(0.025)}$, $b^{(0.025)}$). Thus, a confidence level of 95% is assumed.

12

to submit to Applied Mathematical Modelling

The online monitoring system

- **Step 5:** For new values of voltage and current signals, first apply the standardized and rotation operations using the estimated parameters under normal operating conditions (steps 1 to 3).
- **Step 6:** Use Eq. (18) and the estimated parameters in **step 4** to test whether the voltage and current signals are within the control limits.
- **Step 7:** If three or more consecutive points lie outside the control region then it is assumed that the system is not under normal operating conditions; i.e., there is evidence that an electric fault occurred in the transmission line.

Next, we describe a simulation study to evaluate our proposed method.

2.3 The Monte-Carlo simulation study

We propose the following simulation study to compare the original estimates of V_0 , I_0 and φ , and the method which we are presenting. Based on a real case scenario, we set $V_0 = 770kV$, $I_0 = 8.76 kA$, and $\varphi = \frac{2\pi}{3}$ ($\cos(\varphi) = -0.5$). Current and voltage signals were generated using a standard sampling scheme for transmission lines with 32 points per cycle, as shown in Eq. (19).

$$\begin{aligned}
 v_{(i)} &= V_0 \cdot \cos\left(\frac{k}{32 \times T} + \xi_V \cdot V_0\right) \\
 i_{(i)} &= I_0 \cdot \cos\left(\frac{k}{32 \times T} - \varphi + \xi_I \cdot I_0\right)
 \end{aligned} \quad (19)$$

Where $k = \{0, 2, \dots, n-1\}$ and $T = \frac{1}{f}$. The sample size, n , was set as $n = 320$, or 10 cycles. Random values for ξ_V and ξ_I were generated using independent Gaussian distributions with means of zero and equal variance. We used the following variance values: $\sigma_V^2 = \sigma_I^2 = (0.005^2, 0.01^2, 0.05^2 \text{ and } 0.10^2)$. The total number of Monte Carlo simulations was set at 10,000. For each Monte Carlo simulation voltage and current signals of size $n = 320$ were generated using Eq. (19). Then, estimates for V_0 , I_0 and $\cos(\varphi)$ were calculated using methods proposed by Gomes et al. [10] and Fitzgibbon et al. [20]. Finally, upper and lower limits were calculated based on Gaussian and independence assumptions of the residuals, non-Gaussian residuals and using non-linear quantile regression. The proportion of points within the control limits were calculated for each method.

2.4 The Brazilian energy distribution company data set

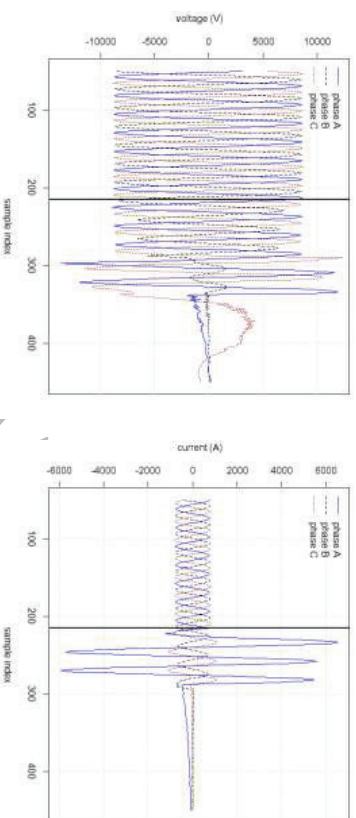
A real case scenario was obtained from CEMIG, a Brazilian energy distribution company. CEMIG is one of the largest energy companies in Brazil, located in the southeast region. The data represent records of voltage and current signals of a transmission line a few moments before an electrical lightning event which caused a phase-ground fault. The lightning fault happened in 2003 and the data base has 650 observations, which represent a time range of 0.45 seconds. The data were sampled using a sampling rate of 24 points per cycle, or one point per 0.07 milliseconds. During the initial 215 observations, or 0.15 seconds, the line is under normal operating conditions. At observation number 216 the lightning fault happens. Due to the opening time of the electrical breaker, which usually starts, on average, within

13

to submit to Applied Mathematical Modelling

three sine cycles after the fault, or 0.05 seconds, observations number 217 to 289 represent the voltage and current signals during the fault. We used voltage and current signals from phase A.

Figs 5 (a) and 5 (b) show voltage and current signals for each phase before and after the occurrence of the fault. It can be seen that the current signal is more sensitive to the fault as compared to the voltage signal.



(a) voltage signals before and after the fault.

(b) current signals before and after the fault.

Figure 5. Voltage and current signals for the transmission line (phases A, B and C) after and before a lightning fault.

3. Results

Using the simulated data, Table 1 shows the results for the parameters, estimated using the previous approach [10] and the approach proposed by Fitzgibbon et al. [20]. For smaller values of variances, both methods achieved very similar results. The larger the variances the better are the estimates using Fitzgibbon et al. [20], as expected. In practice, under normal operating conditions the noise variances for both voltage and current signals are very small and close to 0.01^2 . Under these circumstances, both methods provide similar results.

to submit to Applied Mathematical Modelling

Table 1. Estimated peak values of voltage and current signals; and estimated power factor, using simulated data. Best results are shown in bold.

| Variance (values) | Statistics | Gomes et al. [10] | | | Fitzgibbon et al. [20] | | |
|-------------------|-------------------|-------------------|-----------------|------------------|------------------------|-----------------|------------------|
| | | $\cos(\varphi)$ | V_0 | I_0 | $\cos(\varphi)$ | V_0 | I_0 |
| 0.05^2 | True value | -0.50 | 770 | 8760 | -0.50 | 770 | 8760 |
| | Mean | -0.4999 | 770.0521 | 8760.5986 | -0.4999 | 770.0016 | 8760.0239 |
| | SD | 0.0007 | 0.3909 | 4.4885 | 0.0007 | 0.3908 | 4.4884 |
| | Error (%) | 0.0178 | 0.0068 | 0.0068 | 0.0178 | 0.0002 | 0.0003 |
| 0.01^2 | Mean | -0.4996 | 770.2150 | 8762.4826 | -0.4996 | 770.0130 | 8760.1844 |
| | SD | 0.0014 | 0.7943 | 9.0024 | 0.0014 | 0.7939 | 8.9950 |
| | Error (%) | 0.0754 | 0.0279 | 0.0283 | 0.0753 | 0.0017 | 0.0021 |
| 0.05^2 | Mean | -0.4903 | 775.1050 | 8817.7664 | -0.4905 | 770.1851 | 8761.8014 |
| | SD | 0.0069 | 3.9318 | 45.5769 | 0.0068 | 3.8706 | 44.8428 |
| | Error (%) | 1.9349 | 0.6630 | 0.6594 | 1.8986 | 0.0240 | 0.0206 |
| 0.10^2 | mean | -0.4635 | 790.3784 | 8993.1207 | -0.4659 | 772.0260 | 8784.3151 |
| | SD | 0.0142 | 7.9376 | 90.8065 | 0.0135 | 7.5240 | 85.6990 |
| | Error (%) | 7.2993 | 2.6465 | 2.6612 | 6.8297 | 0.2631 | 0.2776 |

Absolute error (%): $100 \times |(\text{real value} - \text{estimated value}) / \text{real value}|$

Using previous estimated values, upper and lower limits were calculated assuming Gaussian random variables and using the percentiles of the residuals (non-Gaussian approach). In addition, upper and lower limits were calculated using non-linear quantile regression. Table 2 presents the proportion of points within the control limits using the same simulated data that were applied to estimate the parameters. By doing so, it is expected to find the proportion of points close to $(1 - \alpha)\%$, where α was chosen as 0.05 (5%). Results show that control limits, assuming independent Gaussian random variables hold fewer than 95% of points within the control limits. On average, 92.5% of simulated points are within the control limits, even though the simulation and the original method were based on independent Gaussian random variables. As a consequence, 7.5% of the points, on average, are outside the control limits. Using the percentiles of the residuals (non-Gaussian assumption), the proportion of points within the control limits is approximately 45%. This is the smallest value among the evaluated methods and represents almost half of the proposed confidence level (95%). Thus, if this method is applied, approximately 64% of the points will lie outside the control limits. Consequently, this method will frequently detect faults (false faults) even though the transmission line is operating under normal conditions. In practice, we do not recommend such an approach. Finally, the non-linear quantile regression approach holds 94.6% of the points within the control limits, which is much closer to 95% than previous approaches.

to submit to Applied Mathematical Modelling

Table 2. Proportion of points within the control limits for each method in 10,000 simulations.

| Variance values | Statistics | Independent Gaussian assumption | non-Gaussian assumption | Non-linear quantile regression |
|--------------------|------------|---------------------------------|-------------------------|--------------------------------|
| 0.005 ² | minimum | 89.0625 | 37.8125 | 92.1875 |
| | mean | 92.5620 | 45.9676 | 94.6291 |
| | maximum | 95.9375 | 54.0625 | 96.2500 |
| 0.01 ² | minimum | 89.0625 | 37.1875 | 91.8750 |
| | mean | 92.5672 | 45.9618 | 94.6194 |
| | maximum | 96.2500 | 54.0625 | 95.9375 |
| 0.05 ² | minimum | 89.0625 | 36.5625 | 91.2500 |
| | mean | 92.5797 | 45.4936 | 94.6096 |
| | maximum | 96.2500 | 53.4375 | 95.9375 |
| 0.10 ² | minimum | 89.3750 | 37.1875 | 91.8750 |
| | mean | 92.6403 | 44.8731 | 94.6658 |
| | maximum | 95.9375 | 52.5000 | 95.9375 |

Using the Brazilian energy distribution company data set, the parameters of the transmission line were estimated using the first 215 observations under normal operating conditions, as shown in Fig. 6 (a). Table 3, shows estimated parameters of voltage and current using the linear equation approach proposed by Gomes et al. [10] and using the penalized linear equation approach [20]. Results show differences with respect to voltage and current peak values. The parameters were estimated using the observations before the fault, i.e., under normal operating conditions. The final estimated voltage and current peak values are $V_0 = 768.53$ V and $I_0 = 8743.06$ A. The voltage and current standard deviations, for the standardized signals under normal operating conditions, are $s_V = 0.0074$ and $s_I = 0.0107$. It is worth mentioning that the current standard deviation is larger than the voltage standard deviation. As shown in Figure 5 (b), this is because current signals are more sensitive than voltage signals to disturbances in transmission lines, even under normal operating conditions. It is also worth mentioning that the estimated values are similar to the values used in the simulation study.

Table 3. Estimates of peak values parameters

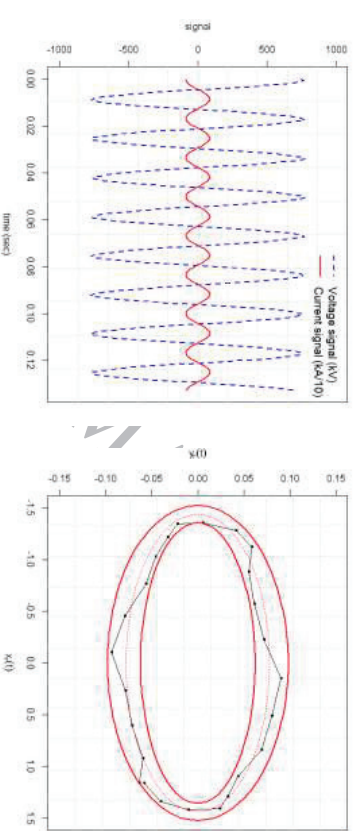
| Phase | Parameters | Gomes et al [10] | Fitzgibbon [20] |
|-------|-----------------|------------------|-----------------|
| A | $\cos(\varphi)$ | 0.7376 | 0.7376 |
| | V_0 | 8.621 kV | 8.743 kV |
| | I_0 | 757.8 A | 768.5 A |
| B | $\cos(\varphi)$ | 0.7377 | 0.7377 |
| | V_0 | 8.624 kV | 8.723 kV |
| | I_0 | 763.5 A | 772.3 A |
| C | $\cos(\varphi)$ | 0.7384754 | 0.7384752 |
| | V_0 | 8.520 kV | 8.841 kV |

16

to submit to Applied Mathematical Modelling

| I_0 | 767.9 A | 796.8 A |
|-------|---------|---------|
|-------|---------|---------|

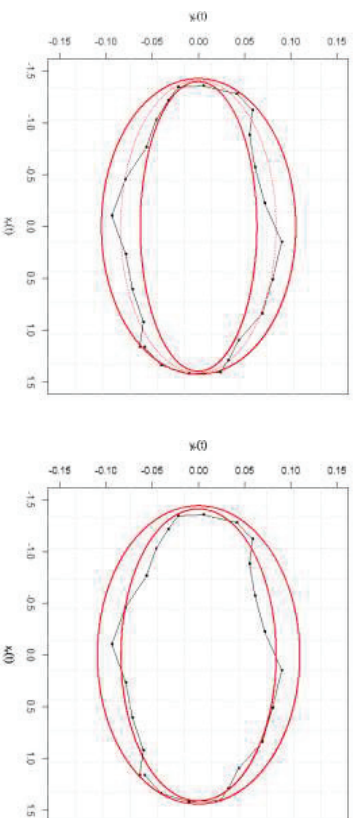
Upper and lower limits with confidence level of 99.7% were estimated using our proposed method, shown in Fig. 6(b), using the Gaussian residuals approach [24,25], shown in Fig. 6 (c), and using the method of non-Gaussian residuals [25,26], shown in Fig. 6(d). As previously described in the simulation study, the methods using Gaussian and non-Gaussian residuals generate narrow control limits with false fault rates greater than the selected parameter which is 0.3%. This is illustrated in Fig. 6(c) and Fig. 6(d), which show observations lying outside the control region. Our proposed approach, shown in Fig. 6(b), provides a proper fit of the observations under normal operating conditions, as expected using a 99.7% confidence level.



(a) Current and voltage signals under normal operating conditions.

(b) Voltage and current signals under normal operating conditions within the control limits, estimated using quantile regression.

17

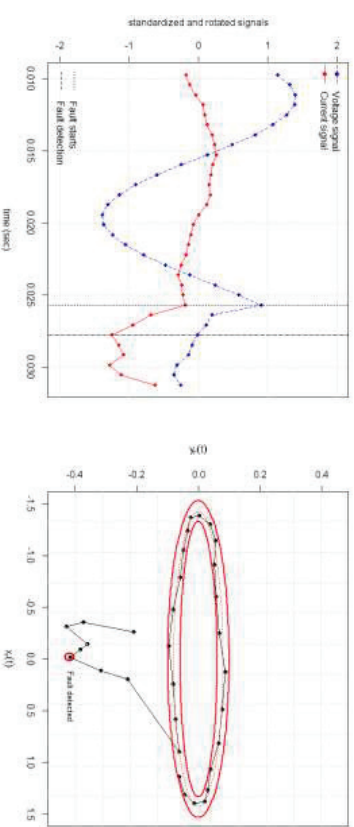


(c) Voltage and current signals under normal operating conditions within the control limits, estimated assuming gaussian residuals [10]

(d) Voltage and current signals under normal operating conditions within the control limits, estimated using non-gaussian residuals [10].

Figure 6. Real case scenario using voltage and current signals from CEMIG. The signals under normal operating conditions are shown in (a). Control limits with 99.7% confidence level estimated using our proposal are presented in (b). The gaussian and non-gaussian control limits estimates are shown in (c) and (d), respectively.

Fig. 7(a) shows current and voltage signals before and after the lightning fault, which is represented by the dotted vertical line. Fig. 7(b) shows that the current and voltage signals leave the control region, representing an abnormal condition. If the criterion of three consecutive points lying outside the control region is applied, then the lightning fault condition is detected within 0.208 milliseconds after the fault, or in one eighth of a cycle. This value is much smaller than standard fault detection approaches.



(a) Voltage and current signals before and after the lightning fault (vertical dotted line).

(b) Control region and voltage and current signals before and after the lightning fault.

Figure 7. Voltage and current signals, and the control region, a few moments both before and after a lightning fault. Fault is detected after three consecutive points lying outside the control region.

Using the data under normal operating conditions, control regions were estimated using the models proposed in [10] and the proposed non linear quantile regression approach. For each method the proportion of points lying inside the control regions were calculated. All methods use a confidence level of 99.7%, i.e., it is expected that, theoretically, 99.7% of the points will lie inside the control region. Table 4 shows the proportion of points lying inside the control regions. The independent Gaussian assumption and non-gaussian assumption models, proposed in [10], achieved a much smaller proportions of points within the control region as compared to the value of 99.7%. Our proposed non linear quantile regression control region achieved the proportion of points much closer to 99.7% as compared to the previous methods.

Table 4. Proportion of points within the control limits for each method under normal operating condition (i.e. the first 215) observation, using a confidence level of 99.7%.

| Phase | Independent Gaussian assumption | Non-Gaussian assumption | Non linear quantile regression |
|-------|---------------------------------|-------------------------|--------------------------------|
| A | 84.19% | 45.12% | 98.14% |
| B | 86.98% | 43.26% | 98.14% |
| C | 85.12% | 41.86% | 98.14% |

4. Discussion and conclusion

This paper proposes a new approach to estimate the parameters of the elliptical model which is applied to detect faults in transmission lines. In addition, it successfully applies non-linear

to submit to Applied Mathematical Modelling

quantile regression to estimate upper and lower control limits. The simulation study and the real case scenario show that our proposal achieves more accurate values for the parameters of the model and provides false detection rates closer to the α -level chosen by the user. Therefore, it provides more reliable control regions than previous approaches.

It is worth mentioning that in energy transmission fault detection methods, voltage and current signals are successively evaluated in order to detect faults as soon as possible. Faults are deviations of the signals from normal operating conditions. Extreme deviations from normal operating conditions are detected quickly and easily. On the contrary, subtle deviations from normal operating conditions are difficult to detect because they approximate normal operating conditions. Therefore, it is crucial to represent normal operating conditions as accurately as possible, thereby establishing a proper boundary between normal and fault conditions. Furthermore, the larger the boundary the more difficult to detect faults. On the contrary, the narrower the boundary, the faster a fault is detected but the greater the chance of false faults. This work proposes new estimate procedures to create a more accurate mathematical representation of the current and voltage signals under normal operating conditions. In addition, new parametric upper and lower bounds are proposed and estimated using quantile regression. This approach does not require any statistical assumption about data uncertainties and create an accurate region under normal operating conditions. The user can change the width of the region in order to adjust the expected number of false faults. A real case scenario shows that this approach can detect faults in one eighth of a cycle, on average.

The main limitation of our proposal, and previous proposals, is the fact that each phase of the transmission line is controlled independently. In a real case scenario, faults may affect more than one phase simultaneously. Therefore, future studies aim at developing a three dimensional control region which controls, simultaneously, the current signals for all three phases of the transmission line. Furthermore, it is of interest to classify fault conditions, i.e., to use signal information under fault conditions to identify the types of faults. Such conditions may include a fire close to the transmission line, or cable entanglement, or electrical lightning, or falling tree, or any other condition. This information is crucial for preparing maintenance teams for line repair. For instance, statistics calculated from the residuals under fault conditions can be used as input variables for fault type classification models.

Finally, the proposed approach can be easily employed in different research areas as control schemes.

Acknowledgements

The authors thank CAPES, CNPq e FAPEMIG for financial support.

20

to submit to Applied Mathematical Modelling

References

- [1] A. Rafinia, J. Moshtagh. A new approach to fault location in three-phase underground distribution system using combination of wavelet analysis with ann and fls. *International Journal of Electrical Power & Energy Systems*, 55:261-274, 2014.
- [2] Aggarwal, R. K., et al. "High frequency fault location technique for transmission lines based on artificial neural networks and genetic algorithms using current signals only." 11th International Conference on Developments in Power Systems Protection, 2012, DPSP 2012. University of Bath, 2012.
- [3] N. Sundaravaran, P. Rajaraman, D. Suzith, M. D. Kumar, G. Abilash, M. J. B. Reddy, D. K. Mohanta. Wavelet based transmission line fault analysis: A literature survey. In *Environment and Electrical Engineering (E3E3)*, 2014 14th International Conference on, pages 254-259. IEEE, 2014.
- [4] Morales, John A., et al. "Comparison between Principal Component Analysis and Wavelet Transform 'Filtering' Methods for Lightning Stroke Classification on Transmission Lines." *Electric Power Systems Research* 118 (2015): 37-46.
- [5] Samantaryay, S. R. "Decision tree-based fault zone identification and fault classification in flexible AC transmission-based transmission line." *Generation, Transmission & Distribution, IET* 3.5 (2009): 425-436.
- [6] Morales, Jose Andre, and Eduardo Orduna. "Patterns extraction for lightning transmission lines protection based on principal component analysis." *Latin America Transactions, IEEE (Revista IEEE America Latina)* 11.1 (2013): 518-524.
- [7] Yusuff, A. A., et al. "Fault location in a series compensated transmission line based on wavelet packet decomposition and support vector regression." *Electric Power Systems Research* 81.7 (2011): 1258-1265.
- [8] Uppendar, J., C. P. Gupta, and G. K. Singh. "Statistical decision-tree based fault classification scheme for protection of power transmission lines." *International Journal of Electrical Power & Energy Systems* 36.1 (2012): 1-12.
- [9] Awan, Zeeshan A., and Pedro M. Ramos. "Improvement of phase difference estimation using modified ellipse fit method." *Instrumentation and Measurement Technology Conference (I2MTC)*, 2010 IEEE. IEEE, 2010.
- [10] Gomes, A. J. D. S., Costa, A., M., Faria, T. G. A., Caminhas, W. M., (2013). Detection and classification of faults in Power Transmission Lines Using Functional Analysis and Computational Intelligence. *IEEE Transactions on Power Delivery*.
- [11] Sarbulut, Lütfü. "A simple power factor calculation for electrical power systems." *International Journal of Electrical Power & Energy Systems* 62 (2014): 66-71.
- [12] D. C. Montgomery. *Introduction to statistical quality control*. John Wiley & Sons, 2007.
- [13] H. Hotelling. *The generalization of Students ratio*. Springer, 1992.

21

to submit to Applied Mathematical Modelling

- [14] M. O. A. Abu-Shawiesh, G. Kibria, and F. George. A robust bivariate control chart alternative to the hotelling's T^2 control chart. *Quality and Reliability Engineering International*, 30(1):25-35, 2014.
- [15] J. Han, W.-K. Kim, J.-W. Lee, C.-H. Kim. Fault type classification in transmission line using stft. In *Developments in Power Systems Protection*, 2012. DPSP 2012. 11th International Conference on, pages 1-5. IET, 2012.
- [16] D. Jwad, P. Leey. Evaluation studies of combined wavelet and neural network applications in high voltage transmission line protection. 2014.
- [17] R. H. Chanié Filho, L. F. Carvalho, P. L. Machado, R. M. de Oliveira. Analysis of voltages induced on power outlets due to atmospheric discharges on radio base stations. *Applied Mathematical Modelling*, 37(9):6530-6542, 2013.
- [18] G. Bassett Jr, R. Koenker. Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, 73(363):618-622, 1978.
- [19] R. Koenker. *Quantile regression*. Number 38. Cambridge university press, 2005.
- [20] A. Fitzgibbon, M. Pilu, R. B. Fisher. Direct least square fitting of ellipses. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(5):476-480, 1999.
- [21] R. Halir, J. Flusser. Numerically stable direct least squares fitting of ellipses. In *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG*, volume 98, pages 125-132. Citeseer, 1998.
- [22] P. Goldsmith, H. Whitfield. Average run lengths in cumulative chart quality control schemes. *Technometrics*, 3(1):11-20, 1961.
- [23] M. B. Khoo, V. Wong, Z. Wu, P. Castagliola. Optimal designs of the multivariate synthetic chart for monitoring the process mean vector based on median run length. *Quality and Reliability Engineering International*, 27(8):981-997, 2011.
- [24] G. Casella, R. L. Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.
- [25] M. Fisz. *Probability theory and mathematical statistics*. Technical report, 1963.
- [26] H. A. David, H. N. Nagaraja. *Order statistics*. Wiley Online Library, 1970.

Apêndice E

Programas em Software *R*

As análises e simulações foram realizadas no software R, disponível em <https://cran.r-project.org/>. O R também é chamado de programa “orientado ao objeto” (*object oriented programming*), o que significa que utilizar o R envolve basicamente a criação e manipulação de objetos em uma tela branca em que o usuário tem de dizer exatamente o que deseja que o programa execute ao invés de simplesmente pressionar um botão. E vem daí uma das grandes vantagens em se usar o R: o usuário tem total controle sobre o que está acontecendo e também tem de compreender totalmente o que deseja antes de executar uma análise.



Com o R é possível manipular e analisar dados, visualizar gráficos e escrever desde pequenas linhas de comando até programas inteiros. O R é a versão em código aberto de uma linguagem de programação inventada nos anos 1980 no Bell Labs chamada de S. Essa linguagem tornou-se bastante popular e vários produtos comerciais que a usam estão disponíveis, como o S-PLUS (R Core Team, 2014). Um aspecto interessante é que a linguagem R, ao contrário de outras linguagens como Fortran e C, é uma linguagem interpretada, o que a faz ser mais fácil de programar, pois processa linhas de comando e as transforma em linguagem de máquina (código binário que o computador efetivamente lê), mas isso diminui a velocidade de processamento.

Como o R é um software livre, não existe a possibilidade de o usuário entrar em contato com um serviço de suporte de usuários, muito comuns em softwares pagos. Ao invés disso, existem várias listas de correio eletrônico que fornecem suporte à comunidade de usuários (<http://www.r-project.org/mail.html>) (PROVETE et al, 2011).

```

## DETECTA FALTA EM 3D
# Carrega a Base de Dados
rm(list=ls(all=TRUE))
library(ellipse)
require(misc3d)
require(scatterplot3d)
require(shape)
require(quantreg)
require(tcltk)
require(shape)
require(plot3D)
require(akima)
require(rgl)
require(grDevices)

M1 <- matrix(0, 123, 5)
M2 <- matrix(0, 41, 8)
M3 <- matrix(0, 41, 2)
for (ik in 1:41){
  temp<-lerarquivo(ik)
  A<- data.frame(temp[1:7])
  B<- data.frame(temp[8:14])
  C<- data.frame(temp[15:21])
  temp_base <-data.frame(lerarquivobase(ik))

#####
xyA<-A
xyB<-B
xyC<- C
N<-dim(xy)[1]

#####
#CALCULO PARA 3D.
x<-xy[,2]
y<-xy[,3]
z<-xy[,4]

phi1<-2*pi/3
basin <- function(I) {
  D <- sum(x^2-y^2-I^2*sin(phi1))^2
  return(D)
}
result <- optimize(basin, c(0,10000), maximum=FALSE, tol=1e-8)
result$minimum
I<-result$minimum
#plot(x, type="l")
#lines(y)
#lines(z)
IOxy<-IOxz<-IOyz<-I

#####
#PADRONIZAÇÃO DO SINAL#####

```

```

#####
xpxy<- (x)/IOxy
xpxz<- (x)/IOxz
ypxy<- (y)/IOxy
#####
xpxz<- (x)/IOxz
zpxz<- (z)/IOxz
#####yz
ypyzy<- (y)/IOyz
zpyz<- (z)/IOyz
r_IO<-cbind(IOxy, IOxz, IOyz)##
#print(r_IO)
#####ROTAÇÃO DO SINAL PADRONIZADO#####
xy<-rotatexy(cbind(xpxy, ypxy), angle=45, mid = c(0, 0))
xz<-rotatexy(cbind(xpxz, zpxz), angle=45, mid = c(0, 0))
yz<-rotatexy(cbind(ypyz, zpyz), angle=45, mid = c(0, 0))

#####TRANSLAÇÃO DA ELIPSE#####
ff<-fit.ellipse(xy[,1],xy[,2])
xyt_x<-xy[,1]+ff$coef[4]/(2*ff$coef[1])
xyt_y<-xy[,2]+ff$coef[5]/(2*ff$coef[3])
fxy<-ff$coef[6]-ff$coef[1]*(ff$coef[4]/(2*ff$coef[1]))^2-
ff$coef[3]*(ff$coef[5]/(2*ff$coef[3]))^2

ff<-fit.ellipse(yz[,1],yz[,2])
yzt_y<-yz[,1]+ff$coef[4]/(2*ff$coef[1])
yzt_z<-yz[,2]+ff$coef[5]/(2*ff$coef[3])
fyz<-ff$coef[6]-ff$coef[1]*(ff$coef[4]/(2*ff$coef[1]))^2-
ff$coef[3]*(ff$coef[5]/(2*ff$coef[3]))^2

ff<-fit.ellipse(xz[,1],xz[,2])
xzt_x<-xz[,1]+ff$coef[4]/(2*ff$coef[1])
xzt_z<-xz[,2]+ff$coef[5]/(2*ff$coef[3])
fxz<-ff$coef[6]-ff$coef[1]*(ff$coef[4]/(2*ff$coef[1]))^2-
ff$coef[3]*(ff$coef[5]/(2*ff$coef[3]))^2

#####CALCULO DOS PARAMENTROS#####
#xy<- (cbind(xyt_x, xyt_y))
#yz<- (cbind(yzt_y, yzt_z))
#xz<- (cbind(xzt_x, xzt_z))

#p*=p/3
p<-0.05
pk<- (p)/3
pkc<- (1-p)

percent_aux <- matrix(0, 75, 3)

for (l in 1:75) {
  resultxy<-aux1(xyt_x,xyt_y,pk)

```

```

resultxz<-aux1(xzt_x,xzt_z,pk)
resultyz<-aux1(yzt_y,ztt_z,pk)

result.final<-cbind(resultxy,resultxz,resultyz)

percent.final<-rep(0, N)
for (j in 1:N) {
  percent.final[j] <- sum(resultt.final[j,])
  if (sum(resultt.final[j,])==3) {percent.final[j]<-1} else {percent.final[j]<-0}
}

percent.estimado<-sum(percent.final)/N*100
percent_aux[1,1]<-pk
percent_aux[1,2]<-percent.estimado
percent_aux[1,3]<-(abs(95-percent.estimado))

#print(percent.estimado)
pk<-pk*(percent.estimado/95)^2
#print(cbind(pk,percent.estimado))
}

#####
for (l in 2:75) {
  if ((percent_aux[1,3] <= percent_aux[l-1,3]))
  {
    perf<-percent_aux[l,2]
    pkf<-percent_aux[l,1]
  }
}

#####r_percent_aux<-percent_aux[percent_aux[,1]==min(percent_aux[,1])&
percent_aux[,2]==perf]
percent_aux1<-percent_aux[percent_aux[,3]==min(percent_aux[,3]),]
percent_aux2<-percent_aux1[percent_aux1[,1]==min(percent_aux1[,1]),]

perf<-percent_aux2[2]
pkf<-percent_aux2[1]*2
r_percent_Inic<- cbind(perf,pkf)

deltaxp<-aux2(xyt_x,xyt_y,pkf)
deltaxzp<-aux2(xzt_x,xzt_z,pkf)
deltayzp<-aux2(yzt_y,ztt_z,pkf)

#####Estimativa Final
#Estim_Final <- aux4(C,I0xy,I0xz, I0yz, deltaxp, deltaxzp, deltaxzp)
r_percent<-cbind(perf,pkf)
#####Deltas#####
r_deltas<-cbind(deltaxp,deltaxzp,deltayzp) ###
colnames(r_deltas)<-cbind("axy", "bxy", "axz", "bxz", "ayz", "byz") ###
#print(r_deltas)

#####
#####Arquivo Após falta de três pontos#####

```

```

arq_detecta_falta <-aux5(temp_base,I0xy,I0xz, I0yz, deltaxp, deltaxzp,
deltaxp,ik)
M3[ik,1]<-ik
M3[ik,2]<-arq_detecta_falta

#####Verificação Prê-Falha#####
perc_pre_falta <- aux4(C,I0xy,I0xz, I0yz, deltaxp, deltaxzp, deltaxzp)

#####Verificação Pos-Falha#####
perc_pos_falta <- aux4(xyB,I0xy,I0xz, I0yz, deltaxp, deltaxzp, deltaxzp)

#####Estadísticas#####
r_pos<-cbind(dim(xyB) [1], perc_pos_falta*100)
r_pre<-cbind(dim(xyA) [1], perc_pre_falta*100)

r_pre_pos<- cbind(r_pre, r_pos) ###
colnames(r_pre_pos) <- cbind("N_pre", "Perc_pre", "N_pos", "perc_pos")
#print(r_pre_pos)
#print(r_deltas)
#print(r_percent)
#print(r_pre_pos)
#print(cbind(I0xy,I0xz, I0yz))

#Parametro Interno
M1[3*ik-2,1]<-r_deltas[1,1]
M1[3*ik-1,1]<-r_deltas[1,3]
M1[3*ik,1]<-r_deltas[1,5]

M1[3*ik-2,2]<-r_deltas[1,2]
M1[3*ik-1,2]<-r_deltas[1,4]
M1[3*ik,2]<-r_deltas[1,6]

#Parametro Externo
M1[3*ik-2,3]<-r_deltas[2,1]
M1[3*ik-1,3]<-r_deltas[2,3]
M1[3*ik,3]<-r_deltas[2,5]

M1[3*ik-2,4]<-r_deltas[2,2]
M1[3*ik-1,4]<-r_deltas[2,4]
M1[3*ik,4]<-r_deltas[2,6]

M1[3*ik-2,5]<-I0xy
M1[3*ik-1,5]<-I0xz
M1[3*ik,5]<-I0yz

#####
M2[ik,1]<-r_percent[1]
M2[ik,2]<-r_percent[2]
M2[ik,3]<-r_pre_pos[1]
M2[ik,4]<-r_pre_pos[2]
M2[ik,5]<-100-r_pre_pos[2]
M2[ik,6]<-r_pre_pos[3]
M2[ik,7]<-r_pre_pos[4]
M2[ik,8]<-100-r_pre_pos[4]
print(ik)

```

```

}
colnames(M1)<-c("ai", "bi", "ae", "be", "I1")
colnames(M2)<-
c("Perce_Inte", "Alfa_Ajustado", "N_Pre_Falta", "Pontos_In_Pre", "Pontos_Ext_Pre",
  "N_Pos_Falta", "Pontos_In_Pos_falta", "Pontos_Ext_Pos_falta")
write.table(M3, "E:/Base_Cemig/resultados/ponto3qdfaltas.txt", sep=";", dec = ",")

##Tabela parametros
write.table(M1, "E:/Base_Cemig/resultados/tabela1_parametros1.txt", sep=";", dec
= ",") ##Tabela Parametros
write.table(M2, "E:/Base_Cemig/resultados/tabela2_parametros2.txt", sep=";", dec
= ",")
##Tabela parametros
#####
#Função detecta a falta com três pontos consecutivos#
aux5<-function(temp_base, I0xy, I0xz, I0yz, deltaxyp, deltaxzp, deltaxyz, ik)
{
  xy <- temp_base[1:1500, ]
  dim(xy)
  N<-dim(xy) [1]
  x<-xy[, 2]
  y<-xy[, 3]
  z<-xy[, 4]

  xpxy<-(x)/I0xy
  ypyz<-(y)/I0yz
  zpzx<-(z)/I0xz

  ##xz
  xpxz<-(x)/I0xz
  zpxz<-(z)/I0xz

  ##yz
  ypyz<-(y)/I0yz
  zpyz<-(z)/I0yz

  xy<-rotatexy(cbind(xpxy, ypyz), angle=45, mid = c(0,0))
  xz<-rotatexy(cbind(xpxz, zpxz), angle=45, mid = c(0,0))
  yz<-rotatexy(cbind(ypyz, zpyz), angle=45, mid = c(0,0))

  xyt_x<-xy[, 1]+ff$coef[4]/(2*ff$coef[1])
  xyt_y<-xy[, 2]+ff$coef[5]/(2*ff$coef[3])
  fxy<-ff$coef[6]-ff$coef[1]*(ff$coef[4]/(2*ff$coef[1]))^2-
ff$coef[3]*(ff$coef[5]/(2*ff$coef[3]))^2

  yzt_y<-yz[, 1]+ff$coef[4]/(2*ff$coef[1])
  yzt_z<-yz[, 2]+ff$coef[5]/(2*ff$coef[3])
  fyz<-ff$coef[6]-ff$coef[1]*(ff$coef[4]/(2*ff$coef[1]))^2-
ff$coef[3]*(ff$coef[5]/(2*ff$coef[3]))^2

```

```

xzt_x<-xz[, 1]+ff$coef[4]/(2*ff$coef[1])
xzt_z<-xz[, 2]+ff$coef[5]/(2*ff$coef[3])
fxz<-ff$coef[6]-ff$coef[1]*(ff$coef[4]/(2*ff$coef[1]))^2-
ff$coef[3]*(ff$coef[5]/(2*ff$coef[3]))^2

pxy<-aux3(xyt_x, xyt_y, N, deltaxyp)
pxz<-aux3(xzt_x, xzt_z, N, deltaxzp)
pyz<-aux3(yzt_y, yzt_z, N, deltaxyz)

vetor.final<-cbind(pxy, pxz, pyz)

k<-0
jj<-3
while(jj<3000) {
  jj<-jj+1;
  if( sum(vetor.final[jj, ])! =3) & (sum(vetor.final[jj-1, ])! =3) &
(sum(vetor.final[jj-2, ])! =3) )
  {
    #NomeArg=paste(c("E:/Base_Cemig/amostraf/amostra"), ik, sep="")
    #NomeArg=paste(c(NomeArg), ".txt", sep="")
    write.table(temp_base[(jj+1):(jj+96), ], NomeArg, sep=";", col.names=TRUE)
    #jj<-jj+96
    #if(k=1){return(jj);break}
    #k<-k+1
    return(jj)
  }
}
#print(temp_base[(jj+1):(jj+20), ])
#return(xy[(jj-2:jj+194), ])
write.table(M3, "E:/Base_Cemig/amostraf/posicao.txt", sep=";", col.names=TRUE)

#Fim Programa#
}

#####
lerarquivobase<-function(k){
...
if(k=1){k<-read.table("E:/Base_Cemig/Dad/01_aq_0101.txt", sep=";", header=TRUE)}
if(k=41){A<-read.table("E:/Base_Cemig/Dad/63_M1_0901.txt", sep=";", header=TRUE)}
return(result)
}

#####
lerarquivov<-function(k){
  if(k=1){A<-
read.table("E:/Base_Cemig/amostraa/A01_aq_0101.txt", sep=";", header=TRUE)}
  if(k=41){A<-
read.table("E:/Base_Cemig/amostrac/C63_M1_0901.txt", sep=";", header=TRUE)}
}

```

```

result<-c(A)
return(result)
)
#
#####
aux4 <- function(xy, I0xy, I0xz, I0yz, deltaxyp, deltaxzp, deltayzp)
{
  dim(xy)
  N<-dim(xy)[1]
  x<-xy[,2]
  y<-xy[,3]
  z<-xy[,4]

  #####xy
  xpxy<-(x)/I0xy
  ypxy<-(y)/I0xy

  #####xz
  xpxz<-(x)/I0xz
  zpxz<-(z)/I0xz

  #####yz
  ypyz<-(y)/I0yz
  zpyz<-(z)/I0yz

  xy<-rotatexy(cbind(xpxy, ypxy), angle=45, mid = c(0,0))
  xz<-rotatexy(cbind(xpxz, zpxz), angle=45, mid = c(0,0))
  yz<-rotatexy(cbind(ypyz, zpyz), angle=45, mid = c(0,0))

  ff<-fit.ellipse(xy[,1], xy[,2])
  xyt_x<-xy[,1]+ff$coef[4]/(2*ff$coef[1])
  xyt_y<-xy[,2]+ff$coef[5]/(2*ff$coef[3])
  fxy<-ff$coef[6]-ff$coef[1]*(ff$coef[4])/(2*ff$coef[1])^2-
  ff$coef[3]*(ff$coef[5])/(2*ff$coef[3])^2

  ff<-fit.ellipse(yz[,1], yz[,2])
  yzt_x<-yz[,1]+ff$coef[4]/(2*ff$coef[1])
  yzt_z<-yz[,2]+ff$coef[5]/(2*ff$coef[3])
  fyz<-ff$coef[6]-ff$coef[1]*(ff$coef[4])/(2*ff$coef[1])^2-
  ff$coef[3]*(ff$coef[5])/(2*ff$coef[3])^2

  ff<-fit.ellipse(xz[,1], xz[,2])
  xzt_x<-xz[,1]+ff$coef[4]/(2*ff$coef[1])
  xzt_z<-xz[,2]+ff$coef[5]/(2*ff$coef[3])
  fxz<-ff$coef[6]-ff$coef[1]*(ff$coef[4])/(2*ff$coef[1])^2-
  ff$coef[3]*(ff$coef[5])/(2*ff$coef[3])^2

  pxy<-aux3(xyt_x, xyt_y, N, deltaxyp)
  pxz<-aux3(xzt_x, xzt_z, N, deltaxzp)
  pyz<-aux3(yzt_y, yzt_z, N, deltayzp)

  result.final<-cbind(pxy, pxz, pyz)
  percent.final<-rep(0, N)

```

```

for (j in 1:N) {
  percent.final[j] <- sum(result.final[j,])
  if (sum(result.final[j,])==3) {percent.final[j]<-1} else {percent.final[j]<-0}
}
percent.estimado<- sum(percent.final)/N
#print(percent.estimado)
return(percent.estimado)
}
#####
aux3<-function(xy, xz, N, deltas)
{
  xy<-cbind(xy, xz)
  krq<-0
  ponto.interior<-rep(0, N)

  for (j in 1:N) {
    L1<-(xy[j,1])^2/deltas[1,1]^2 + (xy[j,2])^2/deltas[1,2]^2
    L2<-(xy[j,1])^2/deltas[2,1]^2 + (xy[j,2])^2/deltas[2,2]^2
    if ((L1>=1) & (L2 <=1)) {krq<-krq+1; ponto.interior[j]<-1}
  }
  percent.pontos<-(krq/N*100)
  return(ponto.interior)
}
#####
aux2<-function(x1, y1, p)
{
  us <- cbind(p/2, (1-p/2))
  xy<-cbind(x1, y1)
  #####Regressão Quantílica#####
  Dat <- NULL
  Dat$x <- xy[,1]
  Dat$y <- xy[,2]
  deltas <- matrix(0, 2, length(us))
  ajustes<-ajustes1 <- matrix(0, 2, length(us))
  for (i in 1:length(us)) {
    tau = us[i]
    fit <- nlrq(abs(y^2) ~ (bq^2*(1-(x^2)/aq^2))), data = Dat, tau =tau,
    start = list(aq=0.5, bq=0.1), trace = FALSE)
    deltas[i, ] <- coef(fit)
  }
  return(deltas)
}
#####
aux1<-function(x1, y1, p)
{
  us <- cbind(p/2, (1-p/2))

```

```

xy<-cbind(x1,y1)
#####
Dat <- NULL
Dat$x <- xy[,1]
Dat$y <- xy[,2]
deltas <- matrix(0, 2, length(us))
ajustes<-ajustes1 <- matrix(0, 2, length(us))
for (i in 1:length(us)) {
  tau = us[i]
  fit <- nlrq(abs(y^2) ~ (bq^2*(1-(x^2)/aq^2))), data = Dat, tau =tau,
start = list(aq=0.5,bq=0.1),trace = FALSE)
  deltas[i, ] <- coef(fit)
}
# print(deltas)
#####
krq<-0
ponto.interior<-rep(0,N)
for (j in 1:N) {
  L1<-(xy[j,1])^2/deltas[1,1]^2 + (xy[j,2])^2/deltas[1,2]^2
  L2<-(xy[j,1])^2/deltas[2,1]^2 + (xy[j,2])^2/deltas[2,2]^2
  if (L1>=1) & (L2 <=1){krq<-krq+1;ponto.interior[j]<-1}
}
percent.pontos<-(krq/N*100)
return(ponto.interior)
# print("Fim do PROGRAMA")
#####FTM DO
PROGRAMA#####
}
fit.ellipse <- function (x, y = NULL) {
  A <- matrix(c(2*f[1], f[2], f[2], 2*f[3]), nrow=2, ncol=2, byrow=T )
  b <- matrix(c(-f[4], -f[5]), nrow=2, ncol=1, byrow=T )
  soln <- solve(A) %*% b
  b2 <- f[2]^2 / 4
  center <- c(soln[1], soln[2])
  names(center) <- c("x", "y")
  num <- 2 * (f[1] * f[5]^2 / 4 + f[3] * f[4]^2 / 4 + f[6] * b2 -
f[2]*f[4]*f[5]/4 - f[1]*f[3]*f[6])
  den1 <- (b2 - f[1]*f[3])
  den2 <- sqrt((f[1] - f[3])^2 + 4*b2)
  den3 <- f[1] + f[3]
  semi.axes <- sqrt(c( num / (den1 * (den2 - den3)), num / (den1 * (-den2 -
den3)) ))
# calculate the angle of rotation
term <- (f[1] - f[3]) / f[2]
angle <- atan(1 / term) / 2

```

```

list(coef=f, center = center, major = max(semi.axes), minor = min(semi.axes),
)
angle = unname(angle)
}

```

#CONSTROI ELLIPSE 3D.

```
library(ellipse)
require(misc3d)
require(scatterplot3d)
require(shape)
require(quantreg)
require(tcltk)
require(shape)
require(plot3d)
require(akima)
require(rgl)
require(ggDevices)

N <- 300
psi <- 0 #Onde começa a amostragem.
phi <- 2*pi/3 #ê determinado
slambda <- seq(0, N-1)
tempo <- slambda/(N*60)
result <- rep(0, N)

e1 <-rnorm(N,0,0.05)
e2 <-rnorm(N,0,0.05)
e3 <-rnorm(N,0,0.05)
x1 <- sin(2*pi*60*tempo + psi)+e1
y1 <- sin(2*pi*60*tempo + psi + phi)+e2
z1 <- sin(2*pi*60*tempo + psi - phi)+e3
xy<-rotatexy(cbind(x1,y1),angle=45)
xz<-rotatexy(cbind(x1,z1),angle=45)
yz<-rotatexy(cbind(y1,z1),angle=-135)

x0<-rnorm(N,-1.5,0)
y0<-rnorm(N,1.5,0)
z0<-rnorm(N,-1.5,0)

open3d()
plot3d(x1,y1,z1, col="black", xlim=c(-1.5,1.5), ylim=c(-1.5,1.5),
       zlim=c(-1.5,1.5), xlab = "Fase-A(x)", ylab = "Fase-B(y)", zlab = "Fase-C(z)",
       font=1)
grid3d(c("x", "y+", "z"), at = NULL, col = "gray", n = 5)

points3d(x1,y1,z0,col="gray", add=T)
points3d(x1,y0,z1,col="gray", add=T)
points3d(x0,y1,z1,col="gray", add=T)

vv<-ellipse(.60, scale = c(.52, .52), centre = c(0, 0))
vv<-rotatexy(vv,angle=-90)
lines3d(vv[,1],vv[,2],z0,col="gray", add=T)
vv<-ellipse(.60, scale = c(.35, .35), centre = c(0, 0))
vv<-rotatexy(vv,angle=-90)
```

#FIGURA KNN

```
lines3d(vv[,1],vv[,2],z0,col="gray", add=T)

vv<-ellipse(.60, scale = c(.52, .52), centre = c(0, 0))
vv<-rotatexy(vv,angle=-90)
lines3d(x0,vv[,1],vv[,2],col="gray", add=T)
vv<-ellipse(.60, scale = c(.35, .35), centre = c(0, 0))
vv<-rotatexy(vv,angle=-90)
lines3d(x0,vv[,1],vv[,2],col="gray", add=T)

vv<-ellipse(.60, scale = c(.52, .52), centre = c(0, 0))
vv<-rotatexy(vv,angle=-90)
lines3d(vv[,1],y0,vv[,2],col="gray", add=T)
vv<-ellipse(.60, scale = c(.35, .35), centre = c(0, 0))
vv<-rotatexy(vv,angle=-90)
lines3d(vv[,1],y0,vv[,2],col="gray", add=T)

xax<-rnorm(30,0,1)
xsb<-rnorm(30,0,1)

pdf("teste.pdf", width = 20, height = 14)
#xlim=c(-2.5,2),ylim=c(-2,2),
plot(Xax,xsb,axes=F,xlim=c(-1.5,0.85),ylim=c(-1.5,.62),xla="", ylab="",pch=20,
     cex=2)
mean(Xax)
mean(Xsb)
points(mean(Xax),mean(Xsb),pch=20, cex=2, col="red")
vet<-rep(0,30)
di<-rep(0,30)
vet<-cbind(xa,xsb)
for(i in 1:30){
  di[i]<-sqrt((mean(Xax)-xa[i])^2+(mean(Xsb)-xb[i])^2)
}
vet<-data.frame(xa,xsb,di)
library(plyr)
vet_final<-arrange(vet,desc(di))
n<-30
segments(vet_final[n,1],vet_final[n,2],mean(Xax), mean(Xsb),lwd=2)
segments(vet_final[n-1,1],vet_final[n-1,2],mean(Xax), mean(Xsb),lwd=2)
segments(vet_final[n-2,1],vet_final[n-2,2],mean(Xax), mean(Xsb),lwd=2)
segments(vet_final[n-3,1],vet_final[n-3,2],mean(Xax), mean(Xsb),lwd=2)
segments(vet_final[n-4,1],vet_final[n-4,2],mean(Xax), mean(Xsb),lwd=2)
segments(vet_final[n-5,1],vet_final[n-5,2],mean(Xax), mean(Xsb),lwd=2)
legend(locator(1), xpd=TRUE, ncol=2, legend=expression(s[c]),cex=1.5,bty="n")
#savePlot("fi_knn_cap3.png", type="png", width = 20, height = 14)
```