



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

**FÁBIO DE JESUS DA SILVA GOMES**

**UM MÉTODO PARA ANÁLISE DE COMPORTAMENTO DE  
DESCARGAS PARCIAIS EM HIDROGERADORES POR MEIO DE  
MODELOS LINEARES GENERALIZADOS**

Belém-PA  
2016

FÁBIO DE JESUS DA SILVA GOMES

**UM MÉTODO PARA ANÁLISE DE COMPORTAMENTO DE  
DESCARGAS PARCIAIS EM HIDROGERADORES POR MEIO DE  
MODELOS LINEARES GENERALIZADOS**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Matemática e Estatística na Universidade Federal do Pará (PPGME-UFPA) como requisito parcial à obtenção do título de Mestre em Estatística.

Orientadora:

Dra. Terezinha Ferreira de Oliveira

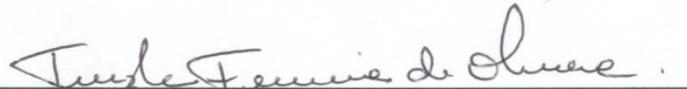
Belém-PA  
2016

FÁBIO DE JESUS DA SILVA GOMES

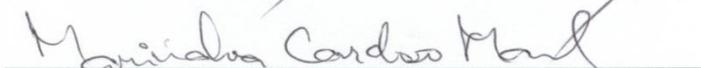
**UM MÉTODO PARA ANÁLISE DE COMPORTAMENTO DE  
DESCARGAS PARCIAIS EM HIDROGERADORES POR MEIO DE  
MODELOS LINEARES GENERALIZADOS**

Dissertação de mestrado apresentado ao Programa de Pós-Graduação em Matemática e Estatística na Universidade Federal do Pará (PPGME-UFPA) como requisito parcial à obtenção do título de Mestre em Estatística.

Banca Examinadora:

  
\_\_\_\_\_  
Prof. Dr. Terezinha F. Oliveira (PPGME – UFPA)  
(Orientadora)

  
\_\_\_\_\_  
Prof. Dr. Antônio M. Silveira (PPGCC – UFPA)

  
\_\_\_\_\_  
Prof. Dr. Marinalva C. Maciel (Estatística – UFPA)

  
\_\_\_\_\_  
Prof. Dr. Valcir João C. Farias (PPGME – UFPA)

Data da avaliação: 26/01/2016.

## AGRADECIMENTOS

A Deus, em primeiro lugar, por ter me concedido forças para continuar cursando ao longo destes dois anos mesmo em meio as circunstâncias adversas da vida cotidiana.

Aos professores do Programa de Mestrado em Matemática e Estatística, pelos conhecimentos compartilhados, proporcionando a mim e a meus colegas de sala o aprendizado e condições para a busca de novos conhecimentos.

À minha orientadora Profa. Dra. Terezinha Ferreira de Oliveira pelas orientações que me permitiram melhorar a estrutura e valor desta pesquisa, pela paciência e credibilidade que me foram concedidas, e ainda, por disponibilizar sua sala, na ausência de outro espaço para que este trabalho fosse desenvolvido.

À minha mãe que sempre acreditou em mim, orando constantemente para que Deus me abençoasse em meus estudos.

À minha mãe e irmã.

É triste pensar que a natureza fala e que o  
gênero humano não a ouve.

Vitor Hugo

## RESUMO

Este trabalho propõe um método para a análise de descargas parciais em hidrogeradores por meio dos modelos lineares generalizados. Os dados são de origem da base de dados do Centro de Pesquisa da Eletronorte, coletados em um dos hidrogeradores da Usina Hidroelétrica de Tucuruí a partir de acopladores capacitivos. Estes acopladores são associados às fases A, B e V que alimentam o hidrogerador, e distribuídos em ambos os lados do estator, sendo quatro grupos de cada lado. Considerou-se como variável resposta o número de ocorrências de descargas parciais, sendo esta explicada pelas variáveis amplitude e ângulo. Assim, foram utilizados os modelos de Poisson e Binomial Negativo para representar os dados do lado esquerdo e do lado direito do hidrogerador, sendo o modelo Binomial Negativo o que melhor se ajustou. Por fim, é feito a interpretação dos parâmetros do modelo ajustado fazendo as devidas análises do comportamento deste fenômeno.

**Palavras-Chave:** Descargas parciais, tríades, dados de contagem, número de ocorrência, modelo de Poisson, modelo Binomial Negativo.

## ABSTRACT

This paper proposes a method for the analysis of partial discharges in hydro through generalized linear models. The data is sourced from the Research Center database Eletronorte collected in one of the hydro Hydroelectric Tucuruí from capacitive couplers. These couplers are associated with phases A, B and V which fuel hydro and distributed on both sides of the stator, four groups of each side. It was considered as the dependent variable the number of partial discharge events, this being explained by the variable amplitude and angle. Thus, the models used were Poisson and negative binomial to represent the data of the left and right side of the hydro, and the Negative Binomial model the best fit. Finally, it made the interpretation of the parameters of the fitted model making the necessary analysis of the behavior of this phenomenon.

**Keywords:** Partial discharges, triads, count data, event number, the Poisson model, Negative Binomial model.

## LISTA DE FIGURAS

<b>Figura 2.1</b> – Sinal de DP gerado através de pulso de corrente no tempo.....	15
<b>Figura 2.2</b> – Corte transversal do isolamento epóxi-mica da barra estatórica de um hidrogerador.....	16
<b>Figura 2.3</b> - Padrão para a descarga interna.....	17
<b>Figura 2.4</b> – Padrão para a descarga de ranhura.....	18
<b>Figura 2.5</b> – Representação da descarga parcial superficial No material dielétrico polimérico.....	19
<b>Figura 2.6</b> – Padrão para a descarga de superfície.....	19
<b>Figura 2.7</b> – Delaminação do compósito epóxi-mica no isolamento de uma barra estatórica de hidrogerador.....	20
<b>Figura 2.8</b> – Padrão para a descarga de delaminação.....	20
<b>Figura 2.9</b> – Descarga entre bobinas de fases diferentes.....	21
<b>Figura 4.1</b> – Distribuição dos acopladores capacitivos no hidrogerador.....	41
<b>Figura 4.2</b> – Gráfico do kernel de densidade das amplitudes das descargas das tríades do lado esquerdo.....	44
<b>Figura 4.3</b> – Gráfico do kernel de densidade dos ângulos das descargas das tríades do lado esquerdo.....	46
<b>Figura 4.4</b> – Gráfico do kernel de densidade das amplitudes das descargas das tríades do lado direito.....	47
<b>Figura 4.5</b> – Gráfico do kernel de densidade dos ângulos das descargas das tríades do lado direito.....	49
<b>Figura 4.6</b> – Histograma das ocorrências de descargas das tríades do lado esquerdo no intervalo de amplitude de -300mV a 300mV.....	50
<b>Figura 4.7</b> – Histograma das ocorrências de descargas das tríades do lado direito no intervalo de amplitude de -300mV a 300mV.....	51
<b>Figura 4.8</b> – Gráficos Boxplots. (a) tríade TR1E, (b) tríade TR2E, (c) tríade TR3E e (d) tríade TR4E no intervalo de amplitude de -300mV a 300mV.....	52
<b>Figura 4.9</b> – Gráficos Boxplots. (a) tríade TR1D, (b) tríade TR2D, (c) tríade TR3D e (d) tríade TR4D no intervalo de amplitude de -300mV a 300mV.....	53
<b>Figura 4.10</b> – Gráfico normal de probabilidade do modelo de Poisson para as tríades do lado esquerdo.....	55
<b>Figura 4.11</b> – Gráfico normal de probabilidade do modelo de Poisson para as tríades do lado direito.....	56

<b>Figura 4.12</b> – Gráfico normal de probabilidade do modelo Binomial Negativo para as tríades do lado esquerdo.....	59
<b>Figura 4.13</b> – Gráficos para diagnóstico do modelo Binomial Negativo para tríade TR1E. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.....	60
<b>Figura 4.14</b> – Gráfico normal de probabilidade do modelo Binomial Negativo para as tríades do lado direito.....	63
<b>Figura 4.15</b> – Gráficos para diagnóstico do modelo Binomial Negativo para tríade TR1D. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.....	64
<b>Figura 4.16</b> – Aumento percentual nas ocorrências de descargas parciais considerando a amplitude de referência amp1 para as tríades esquerdas.....	66
<b>Figura 4.17</b> – Aumento percentual nas ocorrências de descargas parciais considerando a amplitude de referência amp1 para as tríades direitas.....	67

## LISTA DE TABELAS

<b>Tabela 4.1</b> – Dados da tríade TR1E para as amplitudes de -375mV a +749mV.....	42
<b>Tabela 4.2</b> – Organização dos dados da tríade TR1E variando de -300mV a +300mV.....	43
<b>Tabela 4.3</b> – Medidas estatísticas para cada tríade do hidrogerador no intervalo de amplitude de -300mV a 300mV.....	51
<b>Tabela 4.4</b> – Valores da <i>deviance</i> residual, graus de liberdade, razão, AIC e <i>p-value</i> das tríades para o modelo de Poisson.....	54
<b>Tabela 4.5</b> – Valores da <i>deviance</i> residual, graus de liberdade, razão, AIC e <i>p-value</i> das tríades para o modelo Binomial Negativo.....	57
<b>Tabela 4.6</b> – Estimativa dos parâmetros do modelo Binomial Negativo para as tríades do lado esquerdo do hidrogerador.....	58
<b>Tabela 4.7</b> – Estimativa dos parâmetros do modelo Binomial Negativo para os dados do lado direito do hidrogerador.....	61

## SUMÁRIO

1. Introdução.....	11
2. Descargas Parciais.....	14
2.1. Conceito.....	14
2.2. Tipos de Descargas Parciais.....	16
2.2.1. Descargas Internas.....	16
2.2.2. Descargas de Ranhura.....	17
2.2.3 Descargas de Superfície.....	18
2.2.4. Descargas de Delaminação.....	19
2.2.5. Descargas Corona.....	21
2.2.6. Descargas entre Fases.....	21
2.3. Detecção e Análise de Descargas Parciais.....	21
3.1. Modelos Lineares Generalizados.....	23
3.1.1 Família Exponencial.....	24
3.1.2. Estrutura do Modelo Linear Generalizado.....	25
3.1.3. Estimacão dos Parâmetros Desconhecidos.....	25
3.2. Modelos para Dados de Contagem.....	26
3.2.1. Modelo de Poisson.....	26
3.2.2. Modelo Binomial Negativo.....	28
3.2.3. Adequabilidade do Modelo.....	29
3.2.4. Seleção do Modelo.....	29
3.2.4.1. Teste de Razão de Verossimilhança (Estatística de Wilks).....	30
3.2.4.2. Teste $F$ .....	31
3.2.4.3. Teste Wald.....	31
3.2.4.4. Teste Rao (Estatística Score).....	32
3.2.4.5. Método de Akaike.....	32
3.2.5. Ajuste do Modelo.....	32
3.2.5.1. <i>Deviance</i> .....	33
3.2.5.2. Estatística de Person Generalizada.....	33
3.2.6. Superdispersão.....	34
3.2.7. Análise de Resíduos.....	35
3.2.7.1. Resíduo de Pearson.....	35
3.2.7.2. <i>Deviance</i> Residual.....	36
3.2.8. Análise de Diagnóstico.....	36

3.2.8.1. Pontos de Alavanca.....	37
3.2.8.2. Pontos Influentes.....	37
3.2.9. Análise Gráfica.....	38
3.2.9.1. Gráfico da Normalidade para $r_i^{*D}$ com Envelope.....	38
3.2.9.2. Gráfico de $r_i^{*D}$ versus Valores Ajustados $\hat{\mu}_i$ .....	39
3.2.9.3. Gráfico de Índices.....	39
3.3. Estimativa da Densidade Kernel.....	39
4. Aplicação em Descargas Paciais.....	41
4.1. Organização dos Dados.....	42
4.2. Análise Exploratória dos Dados.....	43
4.2.1. Kernel da Densidade de Ocorrência de Descargas Parciais nas Amplitudes e Ângulos nas Fases do Hidrogerador de Todas as Descargas.....	43
4.2.2. Histograma de todas as Tríades para as Descargas no Intervalo de Amplitude de -300mV a 300mV.....	49
4.2.3. Boxplot das Classes de Amplitudes das Tríades do Hidrogerador no Intervalo de Amplitude de -300 a 300mV.....	52
4.3. Ajuste de modelos.....	53
4.3.1. Ajuste pelo Modelo Poisson.....	54
4.3.2. Ajuste pelo Modelo Binomial Negativo.....	56
4.3.2.1. Lado esquerdo.....	57
4.3.2.2. Lado direito.....	61
4.6. Interpretação do Modelo Final.....	65
5. Conclusão.....	68
6. Referências Bibliográficas.....	70
APÊNDICE A.....	73

# Capítulo 1

## 1. Introdução

Preservar a vida útil e manter uma atividade sem interrupções prolongadas dos hidrogeradores nas usinas hidroelétricas é um problema que tem sido investigado e estudado nos últimos anos com o intuito de elaborar estratégias para se realizar a manutenção preditiva, ou seja, baseada nas condições do equipamento reduzindo assim o número de manutenções preventivas. Como consequência, após anos de estudos constatou-se que uma das principais causas do mau funcionamento de máquinas rotativas elétricas de grande porte são as Descargas Parciais (DPs). As mesmas são concebidas como fontes de defeitos nos isolamentos elétricos e sua presença pode provocar uma falha futura do isolamento, pois, inicialmente o fenômeno ocorre no isolamento elétrico de forma parcial, porém com o tempo pode se tornar total causando um curto-circuito, danificando o equipamento elétrico.

Os estudos com o objetivo de compreender melhor este fenômeno têm sofrido avanços consideráveis possibilitando assim identificar padrões nos dados de descargas parciais, relacionando-os ao mesmo tempo com as suas causas. Normas específicas, como as elaboradas pela *International Electrotechnical Commission* (IEC), estabelecem parâmetros e padrões para que os processos de detecção e medição das descargas parciais sejam realizados, possibilitando assim, a criação de uma base de dados confiável para o desenvolvimento dos estudos.

Campbell *et al.* (1994) analisa a utilização do sistema de monitoramento de descargas parciais em hidrogeradores para distinguir interferências do ruído das DPs, descreve também o aparelho de medição e o método para gravar com melhor precisão a forma de onda dos pulsos de descargas, e mostra os resultados das medições em vários hidrogeradores ativos, encontrando uma grande variedade de formas de onda de DPs.

Warren e Stone (2000) realizaram uma análise estatística de aproximadamente 19.000 resultados de teste, classificando as descargas parciais como alta e baixa a partir da comparação das descargas de no enrolamento de máquinas semelhantes. Posteriormente, os autores utilizaram um banco de dados de mais de 60.000 resultados de teste obtidos a partir de centenas de máquinas, onde o objetivo é o reconhecimento de quais máquinas apresentam deterioração no isolamento do estator e constataram

diferença na atividade de DP tanto em função da idade do enrolamento, como do fabricante da máquina (STONE; WARREN, 2005).

Maqrashi (2004) analisa os sinais produzidos por certos defeitos por meio da aplicação de técnicas de análise de dados estatísticos de descarga parcial. O resultado desta análise é então utilizado no processo de reconhecimento de tipos de defeitos.

Venkatesh *et al.* (2011) propõem uma nova abordagem da rede neural probabilística utilizando Base Radial (RBPNN) com algoritmo de seleção para a classificação de múltiplas fontes de descarga parciais.

Mas'ud *et al.* (2014) propõem uma técnica para a classificação de padrões de descarga parcial baseada na aprendizagem por conjunto de rede neural (ENN) a partir de parâmetros estatísticos associados a diversos tipos de falhas.

Percebe-se então uma extensa quantidade de estudos que tem contribuído para compreensão das descargas parciais, bem como a utilização de variadas ferramentas científicas que tem proporcionado a elaboração de métodos, normas e padrões que permitem extrair características consistentes deste fenômeno. Assim, dentro deste contexto, uma ferramenta bastante útil e amplamente utilizada nas mais diversas áreas de estudo são os modelos estatísticos de regressão, e especificamente os Modelos Lineares Generalizados (MLG). Estes modelos representam uma extensão dos modelos lineares clássicos e têm sido empregados em estudos diversos, apresentando resultados satisfatórios que justificam sua relativa importância no campo da pesquisa científica.

Desta forma, com a proposta de alargar o conjunto de ferramentas utilizadas no estudo das DPs, este trabalho possui como objetivo geral propor um método de organização dos dados e a utilização da metodologia dos Modelos Lineares Generalizados que permita a extração de características das descargas parciais ocorridas em hidrogeradores.

Os dados utilizados neste estudo são provenientes da base de dados do Centro de Pesquisas da Eletronorte, sendo estes coletados por acopladores capacitivos conectados a um dos hidrogeradores da Usina Hidroelétrica de Tucuruí. Assim, inicialmente foi usado um aplicativo para converter o formato do arquivo para uma extensão compatível com o bloco de notas e posterior utilização no *software* RStudio. Este programa foi utilizado por ser gratuito e possuir pacotes com funções apropriadas que permitem uma análise rápida e fácil dos dados, possibilitando a exibição de gráficos, estimação dos parâmetros do modelo, bem como uma análise diagnóstica dos resíduos.

O presente trabalho está estruturado em cinco capítulos, sendo o capítulo um a introdução. No capítulo dois inicia-se uma revisão literária sobre descargas parciais. Já o terceiro capítulo segue a revisão abordando os Modelos Lineares Generalizados, sua estrutura, alguns parâmetros que permitem verificar a qualidade do ajuste, bem como aborda especificamente sobre os Modelos de Poisson e Binomial Negativo. No quarto capítulo faz-se o ajuste destes modelos aos dados das descargas parciais, considerando o melhor ajuste e fazendo-se as devidas considerações. Por fim, no capítulo cinco é feita a conclusão deste estudo apresentando considerações sobre os resultados obtidos. Desta forma espera-se que este estudo possa contribuir de forma significativa para o estudo das DPs e possa possibilitar continuidade na utilização dos MLGs no complexo estudo deste fenômeno.

## Capítulo 2

### 2. Descargas Parciais

As descargas parciais são concebidas como fontes de defeitos nos sistemas de isolamentos elétricos e sua presença pode ser um indicativo de uma falha futura destes, pois, inicialmente o fenômeno ocorre no isolamento de forma parcial e com o tempo existe a tendência de se tornar total, o que causaria um curto-circuito entre as partes energizadas e o terra, danificando o equipamento elétrico.

Diversos pesquisadores tais como Fruth *et. al.* (1994), Gross *et. al.* (1998), Suwarno *et. al.* (2008), entre outros têm realizado estudos para identificar padrões em dados de descargas parciais, relacionando-os com as suas causas, bem como determinar características que viabilizem uma maior compreensão desses fenômenos. Normas específicas estabelecem parâmetros para que os processos de detecção e medição das DPs sejam realizados. E, dessa forma, possibilitará a criação de uma base de dados para o desenvolvimento de estudos.

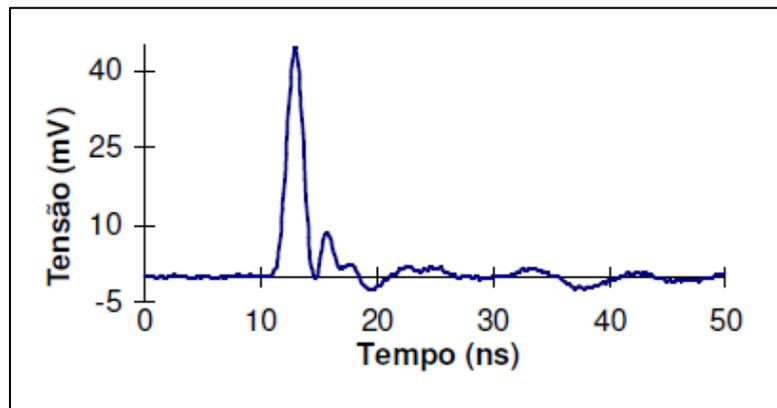
A medição de descargas parciais tem sido utilizada há décadas tanto por fabricantes de máquinas rotativas de grande porte, como por utilitários como uma forma eficiente de avaliar a qualidade de um novo sistema de isolamento elétrico, bem como um meio de detectar fontes de DP em isolamento elétrico envelhecido do enrolamento, resultantes de altas tensões de funcionamento de uma máquina. Em comparação com outros testes dielétricos<sup>1</sup>, o caráter diferenciador das medições de descargas parciais localizadas é que este permite identificar os pontos fracos do sistema de isolamento.

#### 2.1. Conceito

As DPs são sucessões de descargas elétricas incompletas, rápidas e intermitentes da ordem de nano segundos, que ocorrem pela proximidade entre duas partes condutoras de eletricidade e um meio isolante. Elas são incompletas, pois ocorrem numa região definida do espaço sujeita a um campo elétrico e não chegam a completar um circuito fechado (CUENCA, 2005; FAIER, 2006). Na Figura 2.1 se observa o formato de uma descarga parcial, medida através do pulso de corrente ao longo do tempo, medido em nanosegundos.

<sup>1</sup> Os materiais que se caracterizam por oferecer uma considerável resistência à passagem da corrente elétrica são denominados isolantes ou dielétricos. Testes dielétricos permitem identificar os pontos do material que apresentam maior fragilidade quanto à resistência a formação de corrente elétrica.

**Figura 2.1** – Sinal de DP gerado através de pulso de corrente no tempo.

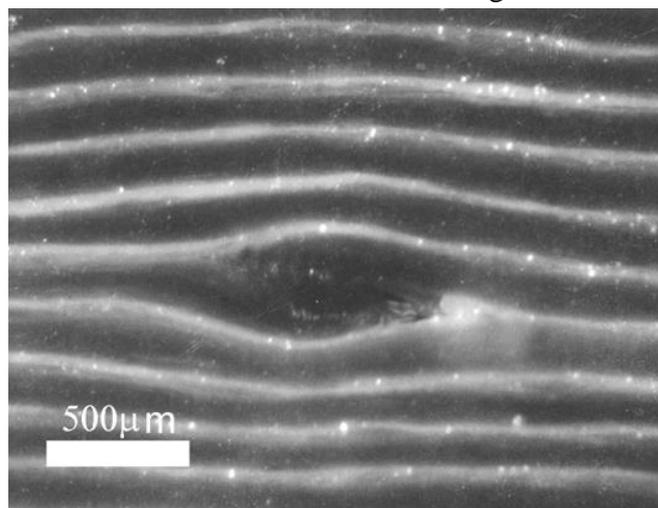


**Fonte:** Levy (2001)

Em materiais sólidos essas descargas são produzidas pela ionização de pequenas cavidades de ar no interior do material isolante, podendo ocorrer em qualquer ponto do dielétrico, ou ainda, na junção de dois dielétricos diferentes ou adjacentes ao condutor. De acordo com Vitoldo Filho (2000), a formação destas cavidades de ar na estrutura de materiais isoladores pode ocorrer por diversos motivos, dependendo da natureza do material e de seu processo de fabricação. Para materiais como as resinas epóxi, a formação dos vazios pode ocorrer durante o processo de cura do material, podendo ser devido à infiltração do ar atmosférico ou devido à formação de gases residuais produzidos por reações químicas durante o processo de cura da resina.

De acordo com o referido autor, em sistemas de isolamento de barras estatóricas de hidrogeradores se utiliza o compósito epóxi e mica, cujo objetivo é aumentar a rigidez dielétrica, já que a mica apresenta alta rigidez dielétrica quando o campo elétrico é aplicado perpendicularmente aos seus planos atômicos. Os vazios podem aparecer na matriz de epóxi decorrentes do processo de cura do epóxi ou na interface entre o epóxi e a mica devido à falta de adesão entre as duas fases. A Figura 2.2 mostra um corte da seção transversal do isolamento de uma barra estatórica de um hidrogerador, onde aparece um grande vazio formado na região de epóxi. A região escura corresponde a matriz de epóxi e as linhas claras são as lâminas de mica.

**Figura 2.2** – Corte transversal do isolamento epóxi-mica da barra estatórica de um hidrogerador.



**Fonte:** Vitoldo Filho (2000).

As descargas parciais são extremamente prejudiciais tanto para o material dielétrico quanto para o equipamento que o envolve. Segundo Kawakatsu (2009) e Diniz (2005) elas são acompanhadas por diversas manifestações físicas tais como geração de calor, ondas eletromagnéticas interferindo nas recepções de rádio, pulsos elétricos, pulsos acústicos, luz, reações químicas em gases como o ar ou o hidrogênio e, além de outros, a deterioração do material dielétrico, sendo uma das principais falhas no isolamento. Conseqüentemente, em função da intensidade das descargas o material pode ter uma vida útil precoce.

## **2.2. Tipos de Descargas Parciais**

As DPs apresentam certas características que servem para determinar a natureza das mesmas, bem como a sua origem. De acordo com essas características foram elaboradas normas que apresentam alguns padrões de DPs e suas respectivas fontes originadoras. Os padrões reconhecidos até o momento são correspondentes às descargas internas, descargas de ranhura, descargas de superfícies, descargas de delaminação e descargas de corona.

### **2.2.1. Descargas Internas**

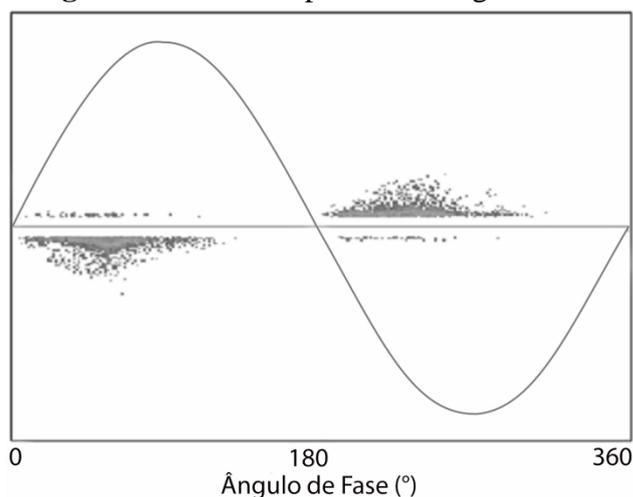
De acordo com a IEC 60034, embora os processos de fabricação sejam projetados para minimizar vazios internos, inevitavelmente, há algum conteúdo vazio em um sistema de isolamento de mica e fita de resina impregnada, que é normalmente usado em máquinas rotativas de alta tensão. Na verdade, a mica no sistema de isolamento impede as descargas parciais de desenvolverem-se provocando um colapso

completo. Enquanto os vazios internos forem pequenos e não aumentarem significativamente, a fiabilidade operacional não é reduzida.

Segundo Brasil (2013), associado às descargas internas está um processo de envelhecimento do isolamento em que pequenos caminhos se propagam através do epóxi em torno das barreiras de mica. Este fenômeno é denominado de arborescência elétrica e pode ocasionar a falha do isolamento principal. A ocorrência contínua de descargas internas em vazios preenchidos por gases, em geral, é o responsável pela iniciação da arborescência elétrica.

A Figura 2.3 apresenta um padrão gráfico para a descarga interna, definido pelo Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE) através da norma IEEE P1434 (2011). De acordo com esta norma o número de descargas positivas é aproximadamente igual ao de descargas negativas, possuindo uma forma arredondada semelhante para cada semiciclo.

**Figura 2.3** – Padrão para a descarga interna.



Fonte: Adaptado de IEEE P1434

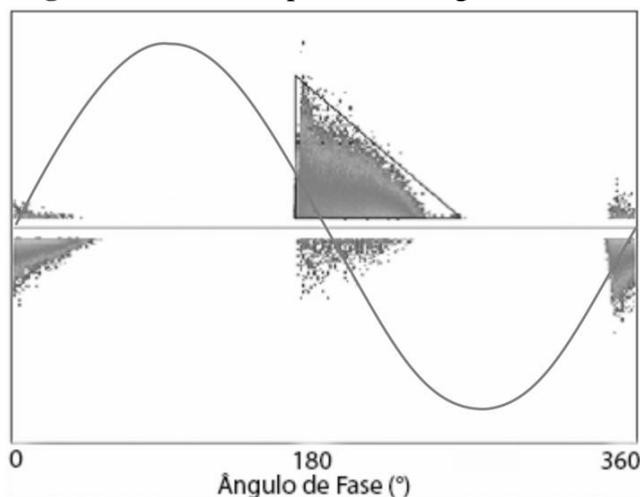
O gráfico da Figura 2.3, assim como os demais que serão apresentados aqui, exibe nuvens de pontos que representam as descargas parciais ocorridas durante o período de medição, considerando o ângulo de fase (medido em graus) no eixo horizontal e a amplitude da descarga (sem escala nos gráficos apresentados pela norma) no eixo vertical.

### 2.2.2. Descargas de Ranhura

Descargas de ranhura em máquinas de alta tensão irão desenvolver-se quando o revestimento do condutor está danificado devido a certos fatores como ao movimento da barra/bobina, a corrosão do material, desgaste, ataque químico ou deficiências de fabricação. DPs de alta intensidade irão se desenvolver podendo resultar em danos

adicionais ao isolamento principal e, eventualmente, em uma falha de isolamento. Embora o tempo absoluto entre a detecção deste fenômeno e a falha de isolamento final seja desconhecido, sua detecção em um estágio inicial é importante para se iniciar a ação corretiva apropriada (IEC 60034). A Figura 2.4 apresenta um padrão gráfico para a descarga de ranhura presente na norma IEEE P1434 (2011). Assim, percebe-se que o número de descargas positivas é maior que o de descargas negativas, possuindo uma forma triangular característica no semiciclo positivo para este tipo de descarga.

**Figura 2.4** – Padrão para a descarga de ranhura.



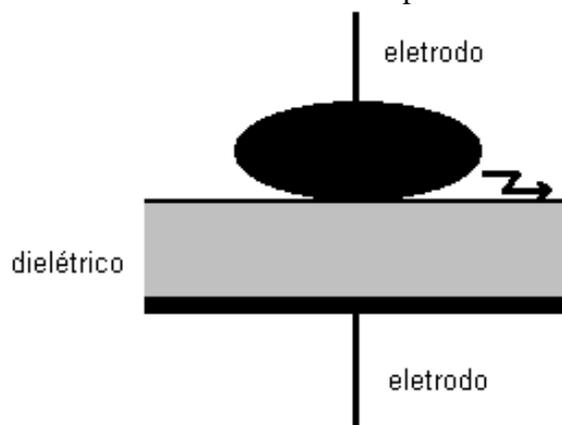
Fonte: Adaptado de IEEE P1434

### 2.2.3. Descargas de Superfície

Descargas de superfície são as descargas que ocorrem na superfície de um material dielétrico, normalmente partindo de um eletrodo para a superfície (VITOLDO FILHO, 2000). As descargas de superfície geralmente se iniciam quando o campo elétrico ao longo da superfície excede o valor da rigidez dielétrica do gás ou líquido que está ao redor. Esse fenômeno pode acontecer quando o revestimento aplicado se tornar ineficiente devido à contaminação por sujeira ou umidade, entre outros (Brasil, 2013). Descargas de superfície podem também resultar de uma inadequada distância de isolamento entre fase ou por utilização do sistema de suporte da cabeça da bobina inadequado.

Assim como as descargas internas, as descargas superficiais ocasionam alterações na superfície do isolamento iniciando caminhos condutores que se propagam ao longo da direção do campo elétrico (VITOLDO FILHO, 2000). Estes caminhos condutores conhecidos como “trilhamento” também podem levar o isolamento à ruptura total, danificando o equipamento. A Figura 2.5 apresenta um esquema simplificado de uma descarga parcial superficial ocorrida num material dielétrico polimérico.

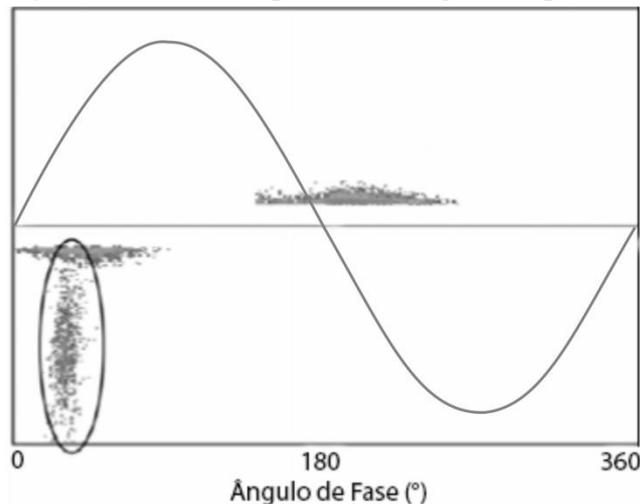
**Figura 2.5** – Representação da descarga parcial superficial  
No material dielétrico polimérico.



**Fonte:** Vitoldo Filho (2000)

A Figura 2.6 exibe o padrão gráfico para a descarga de superfície de acordo com a norma IEEE P1434 (2011). Segundo esta norma, esse padrão apresenta uma nuvem de pontos vertical durante o meio ciclo positivo, sendo que o número de descargas positivas é menor que o de descargas negativas e com valores bastante altos para a amplitude.

**Figura 2.6** – Padrão para a descarga de superfície.



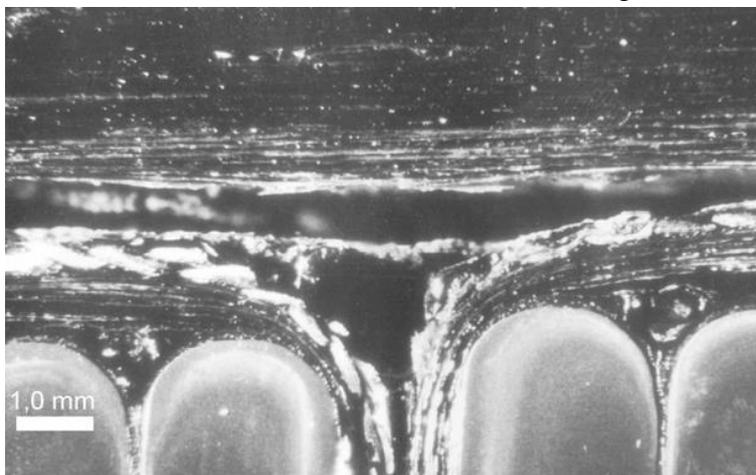
**Fonte:** Adaptado de IEEE P1434

#### 2.2.4. Descargas de Delaminação

De acordo com Vitoldo Filho (2000) uma delaminação é considerada pelos especialistas como os vazios que podem aparecer na interface dos elementos que constituem um compósito tal como o epóxi-mica, devido a falta de adesão entre as duas fases, e cuja principal característica é a formação de vazios plano de grande extensão. Estas regiões são denominadas pela norma IEC 60034-27 como delaminação interna. A Figura 2.7 mostra a imagem de um corte da seção transversal do isolamento de uma

barra estatórica de hidrogerador, onde aparece um grande vazio devido à delaminação do epóxi e da mica na região próxima aos condutores.

**Figura 2.7** – Delaminação do compósito epóxi-mica no Isolamento de uma barra estatórica de hidrogerador.

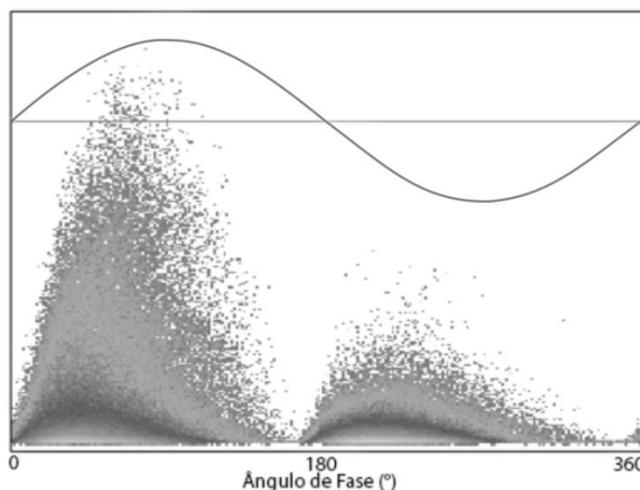


**Fonte:** Vitoldo Filho (2000)

Essas delaminações resultam em descargas de energia relativamente elevada, que pode atacar significativamente o isolamento. Em particular, a delaminação irá reduzir a condutividade térmica do isolamento, o que pode levar a um envelhecimento acelerado, ou mesmo uma fuga térmica (IEC 60034-27).

A Figura 2.8 exibe o padrão gráfico para a descarga de delaminação de acordo com a norma IEEE P1434 (2011). Esta norma estabelece que a frequência de descargas geralmente é baixa, com o número de DPs positivas menor que o das negativas, apresentando ainda uma nuvem de pontos de forma arredondada no meio ciclo positivo.

**Figura 2.8** – Padrão para a descarga de delaminação.



**Fonte:** Adaptado de IEEE P1434

### 2.2.5. Descargas Corona

A descarga corona é uma descarga que ocorre no ar e partem de pontas agudas em eletrodos metálicos. A geometria dos eletrodos influencia diretamente a distribuição da intensidade do campo elétrico (CUENCA, 2005). Por conseguinte, estas pontas, com pequenos raios de curvatura, geram regiões nas vizinhanças do condutor com campo elétrico elevado, pois o gradiente de tensão é alto. Assim, as proximidades da ponta o campo é maior, diminuindo à medida que se afasta da ponta. Logo, a descarga corona ocorre numa região limitada a uma pequena distância da ponta do eletrodo.

### 2.2.6. Descargas entre Fases

Segundo Stone (2009), bobinas operando em alta tensão e colocadas adjacentes a outras bobinas de alta tensão alimentadas por outra fase requerem uma distância mínima para evitar DP no espaço existente entre elas. Esta DP vai gradualmente corroer o isolamento externo da bobina (*groundwall*) e pode levar à falha do estator. Quanto maior for a tensão da máquina e quanto mais fino o isolamento *groundwall*, maior deve ser o espaçamento entre as bobinas. Infelizmente, em muitos casos observa-se espaçamento inadequado, e, conseqüentemente, ocorrências de DPs de alta intensidade e reações químicas com formação de ozônio. Na Figura 2.9 mostra-se o resíduo branco causado pelo ozônio resultante da DP entre duas bobinas em diferentes fases que foram instalados muito perto de um outro.

**Figura 2.9** – Descarga entre bobinas de fases diferentes.



Fonte: Stone (2009)

## 2.3. Detecção e Análise de Descargas Parciais

As técnicas de medida de descargas parciais sofreram nos últimos anos uma evolução substancial devido ao aumento da velocidade dos sistemas de aquisição de dados (VITOLDO FILHO, 2000). De acordo com Cuenca (2005), essas técnicas são

baseadas em vários fenômenos produzidos pela descarga tal como luminescência, calor, ruídos, transformações químicas e pulsos de descargas elétricas. Quanto à análise das DPs, Vitoldo Filho (2000) diz que a técnica empregada consiste em registrar cada descarga parcial com a sua amplitude e o ângulo de fase em relação à tensão aplicada à amostra. Esta técnica é conhecida como *Phase-Resolved Partial Discharge Analysis* (PRPDA).

Segundo Vitoldo Filho (2000), a partir das medições de DPs feitas, é possível gerar algumas distribuições estatísticas, as quais são utilizadas para identificar os defeitos geradores de descarga parcial, dentre as quais citam-se

- Número de Ocorrências em Função do Ângulo de Fase - Esta distribuição fornece o número de ocorrência das descargas em função do ângulo de fase, e apresenta formas distintas para os dois semiciclos devido à irregularidade na superfície interna do vazio;
- Distribuições das Ocorrências das Amplitudes - A distribuição das ocorrências das amplitudes também traz informações relevantes a respeito da evolução das descargas parciais em sistemas de isolamento. Esta técnica está sendo utilizada atualmente para monitoração em linha da evolução das descargas parciais em sistemas de isolamento de grandes geradores.

Alguns fabricantes de analisadores de descargas parciais fornecem software para a identificação de prováveis tipos de defeitos através da utilização de redes neurais treinadas com os parâmetros estatísticos das distribuições geradas por defeitos conhecidos (VITOLDO FILHO, 2000). Os parâmetros estatísticos dessas distribuições, tais como valor médio, desvio padrão, assimetria e curtose, são utilizados em correlações com os tipos de defeitos.

## Capítulo 3

### 3.1. Modelos Lineares Generalizados

Uma ferramenta muito utilizada para se estudar e compreender diversos fenômenos físicos, sociais, químicos, dentre outros é a modelagem estatística. Assim, um modelo estatístico constitui-se de uma equação matemática que relaciona os diversos fatores (covariáveis) que influenciam de forma significativa os resultados observados (variável resposta) no fenômeno em estudo. O modelo linear normal, “criado” no início do século XIX por Legendre e Gauss, dominou a modelação estatística até meados do século XX (TURKMAN, 2000).

O modelo de regressão normal linear, por muitas décadas atendeu satisfatoriamente aos objetivos pretendidos nessas análises. Entretanto uma das condições impostas por este modelo era que o conjunto de valores observados seguisse uma distribuição normal. Naturalmente surgiram outras situações que não puderam ser explicadas adequadamente pelo modelo pelo fato de sua variável resposta seguir a uma distribuição não normal, ou ainda, por possuir um caráter não linear.

Desta forma outros modelos surgiram na tentativa de explicar melhor o fenômeno estudado tais como o modelo *probit*, os modelos *log-lineares*, o modelo complementar *log-log* e outros. Esses modelos apresentam uma estrutura de regressão linear e sua variável resposta segue uma distribuição pertencente a um conjunto de distribuições com propriedades muito específicas, denominado família exponencial.

De acordo com Cordeiro (2013), os Modelos Lineares Generalizados correspondem a uma síntese destes acima citados e de outros modelos, unificando a teoria da modelação estatística da época. Assim, a variável resposta passa a seguir uma distribuição, membro da família exponencial de distribuições, sendo a sua estrutura linear descrita por um conjunto de variáveis independentes. A relação entre a média de todas as observações e a estrutura linear é estabelecida por uma função de ligação adequadamente escolhida para este fim.

As distribuições de probabilidade Normal, Gamma, Poisson, Binomial, Binomial Negativa, Normal Inversa, e outras, fazem parte da família exponencial e os modelos *logístico*, *log-linear*, *probit* e *normal linear* passam a ser particularidades suas.

### 3.1.1. Família exponencial

De acordo com Marciano (2009) a família exponencial pode ser generalizada por se incluir uma constante denominada parâmetro de dispersão. Assim, no contexto dos MLG's diz-se que uma variável aleatória  $Y$  segue uma distribuição pertencente à família exponencial de distribuição (ou família de dispersão exponencial de distribuição) se possuir função densidade de probabilidade (*f. d. p.*) dada por

$$f(y; \theta, \phi) = \exp\{\phi[y\theta - b(\theta)] + c(y; \phi)\}, \quad (3.1)$$

onde  $\theta$  e  $\phi$  são parâmetros escalares,  $c(\cdot)$  é uma função real conhecida e  $b(\cdot)$ , uma função conhecida e diferenciável.

Além disso,  $\theta$  representa o parâmetro de localização, na sua forma canônica, indicando a posição onde a distribuição varia dentro do intervalo correspondente aos valores de  $Y$ , sendo este o que caracteriza a distribuição. O parâmetro  $\phi$  denomina-se parâmetro de dispersão, assumindo sempre valores positivos. Desta forma, as distribuições cuja (*f. d. p.*) puder ser escrita conforme a (3.1) farão parte da família exponencial, sendo uma distribuição de natureza uni paramétrica se  $\phi$  for um parâmetro conhecido e  $\theta$  desconhecido, ou bi paramétrica se ambos forem desconhecidos.

A seguir são mostradas algumas distribuições que obedecem a (3.1), pertencendo assim, à família em discussão.

#### Distribuição de Poisson

Se  $Y$  segue uma distribuição de Poisson de média  $\mu$  e variância  $\sigma^2$ , ou seja,  $Y \sim P(\mu, \sigma^2)$ , sua (*f. d. p.*) é dada por

$$\begin{aligned} f(y; \mu) &= \frac{e^{-\mu} \mu^y}{y!} & (3.2) \\ &= \exp[\log(e^{-\mu}) + \log(\mu^y) + \log 1 - \log(y!)] \\ &= \exp\{y \log(\mu) - \mu + [-\log(y!)]\} \end{aligned}$$

Desse modo, com  $y = 0, 1, 2, \dots$  se tem

$$\begin{aligned} \theta = \log(\mu) &\Rightarrow \mu = e^\theta ; \phi = 1 ; b(\theta) = \mu = e^\theta ; b'(\theta) = e^\theta = \mu \\ c(y; \phi) &= -\log(y!) ; b''(\theta) = e^\theta = \mu ; V(\mu) = e^\theta = \mu \end{aligned}$$

#### Distribuição Binomial Negativa

Seja  $Y$  uma variável aleatória que segue uma distribuição Binomial Negativa de parâmetros  $t$  e  $\pi$ , ou seja,  $Y \sim BN(t, \pi)$ . A sua (*f. d. p.*) é dada por

$$\begin{aligned} f(y; t, \pi) &= \frac{\Gamma(y+t)}{\Gamma(y+1)\Gamma(t)} \pi^t (1-\pi)^y & (3.3) \\ &= \exp \left[ y \log(1-\pi) + t \log(\pi) + \log \left( \frac{\Gamma(y+t)}{\Gamma(y+1)\Gamma(t)} \right) \right], \end{aligned}$$

onde  $0 < \pi < 1$  e  $t > 0$ . Fazendo  $t = \frac{1}{\alpha}$  e  $\pi = \frac{1}{1+\alpha\mu}$  se tem que

$$= \exp \left[ y \log \left( \frac{\alpha\mu}{1+\alpha\mu} \right) + \frac{1}{\alpha} \log \left( \frac{1}{1+\alpha\mu} \right) + \log \left( \frac{\Gamma(y+\frac{1}{\alpha})}{\Gamma(y+1)\Gamma(\frac{1}{\alpha})} \right) \right]$$

Assim,

$$\theta = \log \left( \frac{\alpha\mu}{1+\alpha\mu} \right) ; \phi = 1 ; b(\theta) = -\frac{1}{\alpha} \log \left( \frac{1}{1+\alpha\mu} \right) = -\frac{1}{\alpha} \log(1 - e^\theta)$$

$$b'(\theta) = \frac{1}{\alpha} \frac{e^\theta}{(1-e^\theta)} = \mu ; c(y; \phi) = \log \left( \frac{\Gamma(y+\frac{1}{\alpha})}{\Gamma(y+1)\Gamma(\frac{1}{\alpha})} \right) ; b''(\theta) = \mu(1 + \alpha\mu)$$

Então  $V(\mu) = \mu(1 + \alpha\mu)$ .

### 3.1.2. Estrutura do Modelo Linear Generalizado

Como dito anteriormente o MLG é constituído de uma variável resposta que segue uma distribuição pertencente à família exponencial, um conjunto de variáveis explicativas que descreve sua estrutura linear e uma função de ligação responsável por relacionar a média da variável resposta com a sua estrutura linear. Assim, a ideia básica consiste em abrir opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear (PAULA, 2010).

Uma importante característica dos MLGs é a suposição de independência, ou pelo menos de não correlação, entre as observações (OLIVEIRA, 2013). A escolha da distribuição para a variável resposta deve ser subsequente a uma análise exploratória de dados que tem como finalidade verificar suas características como assimetria, intervalo de variação, natureza e outras.

Apesar de algumas restrições inerentes a esta classe de modelos como a sua estrutura de linearidade, ao fato de suas distribuições pertencerem à família exponencial e a exigência de independência para as respostas, estes assumem notadamente grande importância na elaboração de modelos estatísticos.

### 3.1.3. Estimação dos Parâmetros Desconhecidos

Cordeiro (2013) e Paula (2010) apresentam o vetor score  $U(\beta)$ , também conhecido como função score, e através do sistema  $U(\beta) = 0$  e utilizando-se da matriz de informação de Fisher os parâmetros desconhecidos podem ser estimados. Entretanto, as equações de verossimilhança do sistema geralmente não possuem solução analítica, pois são equações não lineares. Neste caso deve-se recorrer a métodos matemáticos que possam gerar soluções numéricas para as mesmas. Alguns métodos, apresentados por

vários autores como Cordeiro e Paula, por exemplo, são usados para encontrar os estimadores de máxima verossimilhança tais como o de *Newton-Raphson*, *Scores* de Fisher e o dos mínimos Quadrados Ponderados, sendo este último o mais utilizado.

### 3.2. Modelos Para Dados de Contagem

Os dados na forma de contagem são provenientes de uma simples contagem de eventos (CORDEIRO, 2010). Segundo (SANTOS, 2013) e (MONTROYA, 2009), esta pode ser definida como o número de eventos que ocorrem numa mesma unidade de observação. A contagem resultante da observação pode ser realizada num intervalo espacial definido (número por unidade de área, volume e outros) ou temporal sendo, por exemplo, o número de ligações telefônicas que chegam a uma central por minuto (intervalo temporal), ou ainda, o número de insetos que podem aparecer numa determinada plantação por quadrante (intervalo espacial).

O modelo de regressão de Poisson é gerado a partir da distribuição de Poisson que possui uma importante característica, ou seja, sua variância é igual a sua média (MARCIANO, 2009). Devido ao fato da variável resposta dos dados de contagem assumir apenas valores inteiros não negativos, este é o modelo mais utilizado para estes tipos de dados, pois teoricamente é mais adequado. Entretanto é comum acontecerem situações em que o modelo de Poisson apresenta problemas, não se ajustando adequadamente aos dados. Uma destas situações ocorre quando há fortes indícios de superdispersão, sendo isto confirmado com os resíduos plotados fora da banda de confiança gerada.

Para se tratar o problema da superdispersão recorre-se então a outros modelos e técnicas desenvolvidos para esta finalidade. O modelo Binomial Negativo, por exemplo, consegue incorporar satisfatoriamente a superdispersão exprimindo resultados que permitem uma boa análise do fenômeno em estudo.

#### 3.2.1. Modelo de Poisson

O modelo de Poisson baseia-se na distribuição de Poisson, assim suponha-se que  $Y_1, \dots, Y_n$  são variáveis aleatórias independentes tais que  $Y_i \sim P(\mu_i)$ . A função de probabilidade de  $Y_i$  com  $i = 1, \dots, n$ , é dada por

$$f(y_i; \mu_i) = P(Y_i = y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, \quad (3.4)$$

onde  $y_i = 0, 1, \dots$  e  $\mu_i$  ( $\mu_i > 0$ ) representa o número médio de ocorrências de um determinado evento.

A média  $E(Y)$  e a variância  $VAR(Y)$  são dadas por

$$E(Y) = VAR(Y) = \mu \quad (3.5)$$

Agora seja  $Y$  uma variável aleatória que representa o número de ocorrências de determinado evento num dado intervalo de tempo ou espaço definido. Seja ainda  $X$  um vetor de covariáveis dado por  $X = (X_1, \dots, X_p)$  e  $x_i^T$  um vetor que corresponde a uma determinada observação do indivíduo  $i$  expresso por  $x_i^T = (x_{i1}, \dots, x_{ip})$ . Considera-se então que a observação  $x_i^T$  segue uma distribuição de Poisson, ficando definida da seguinte maneira

$$Y|X = x_i^T \sim P(\mu_i)$$

onde  $\mu_i = f(x_i^T, \beta)$  corresponde ao número médio de ocorrências de um evento dada a observação  $x_i^T$ . Desta forma (RAMALHO, 1996) diz que o modelo de regressão de Poisson resulta do fato de se colocar do parâmetro  $\mu_i$  como função do conjunto de covariáveis, sendo  $\beta$  um vetor de parâmetros desconhecidos.

Como as covariáveis integram a estrutura linear do modelo e estão associadas à média da variável resposta, então a maneira mais simples de ligação entre esses dois elementos é a forma aditiva representada do seguinte modo

$$\begin{aligned} f(x_i^T, \beta) &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \\ \mu_i &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \end{aligned} \quad (3.6)$$

Verifica-se, entretanto, uma inconsistência no modelo já que a componente sistemática, neste caso, pode assumir qualquer valor real, enquanto  $\mu_i$  pode assumir somente valores reais positivos, ou seja,  $\mu_i > 0$ . Assim, para associar essas componentes sem violar suas características pode-se usar como função de ligação a função logarítmica para  $\mu_i$ , obtendo-se

$$\begin{aligned} \log(\mu_i) &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \\ \log(\mu_i) &= \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \\ \mu_i &= \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}), \end{aligned} \quad (3.7)$$

onde  $\beta_0$  e  $\beta_j$ , com  $j = 1, 2, \dots, p$ , são os vetores de parâmetros desconhecidos que podem ser interpretados como o efeito esperado pela mudança no logaritmo da média por unidade de variação na covariável  $X_j$  do vetor de covariáveis  $X$ . Desse modo podem-se estimar os parâmetros do modelo pelo método de máxima verossimilhança.

### 3.2.2. Modelo Binomial Negativo

Para o modelo Binomial Negativo assume-se que a variável aleatória  $Y_i$  segue uma distribuição Binomial Negativa de parâmetros  $t$  e  $\pi$ , ou seja,  $Y_i \sim BN(t, \pi)$ . Assim a função de probabilidade de  $Y_i$  com  $i = 1, \dots, n$ , é dada por

$$f(y_i; t, \pi) = \frac{\Gamma(y_i+t)}{\Gamma(y_i+1)\Gamma(t)} \pi^t (1-\pi)^{y_i}, \quad (3.8)$$

onde  $0 < \pi < 1$  e  $t > 0$ .

Considerando  $t = \frac{1}{\alpha}$  e  $\pi = \frac{1}{1+\alpha\mu_i}$  obtêm-se a seguinte expressão

$$f(y_i; \mu_i, \alpha) = \frac{\Gamma(y_i+\frac{1}{\alpha})}{\Gamma(y_i+1)\Gamma(\frac{1}{\alpha})} \left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}, \quad (3.9)$$

onde  $y_i = 0, 1, \dots$ ,  $0 < \mu_i < 1$  e  $\alpha$  representa o parâmetro de heterogeneidade.

A média  $E(Y)$  e a variância  $VAR(Y)$  são dadas por

$$E(Y) = \mu \quad (3.10)$$

$$VAR(Y) = \mu + \alpha\mu^2 \quad (3.11)$$

Como  $\alpha > 0$ , observa-se facilmente que  $VAR(Y) > E(Y)$ . Isto significa que a distribuição binomial negativa pode ser utilizada para modelar dados com superdispersão.

Seja  $Y$  uma variável aleatória que representa o número de ocorrências de determinado evento para um número finito de observações  $n$ , onde  $X$  é um vetor de variáveis explicativas dado por  $X = (X_1, \dots, X_p)$  e  $x_i^T = (x_{i1}, \dots, x_{ip})$  um vetor que corresponde a uma determinada observação do indivíduo  $i$ . Então se a observação  $x_i^T$  segue uma distribuição Binomial Negativa, a mesma fica definida assim

$$Y|X = x_i^T \sim BN(\mu_i, \alpha)$$

onde  $\mu_i = f(x_i^T, \beta)$  representa o número médio de ocorrências de um evento dada a observação  $x_i^T$ .

Desse modo, define-se a seguinte equação para  $\mu_i$

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

Assim, para eliminar a violação do domínio da média,  $0 < \mu_i < 1$ , utiliza-se a função logística como função de ligação, obtendo-se

$$\begin{aligned} \log\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right) &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \\ \mu_i &= \frac{1}{\alpha} \frac{\exp\{\beta_0 + \sum_{j=1}^p \beta_j x_{ij}\}}{1 - \exp\{\beta_0 + \sum_{j=1}^p \beta_j x_{ij}\}} \end{aligned} \quad (3.12)$$

onde  $\beta_0$  e  $\beta_j$ , com  $j = 1, 2, \dots, p$ , são os vetores dos coeficientes de regressão que indicam a variação esperada da média por unidade de variação na covariável  $X_j$ . Os parâmetros do modelo Binomial Negativo também podem ser estimados pelo método de máxima verossimilhança.

### 3.2.3. Adequabilidade do Modelo

A escolha de um modelo linear generalizado para representar um dado fenômeno em estudo com um grau de representatividade satisfatório passa necessariamente pelas etapas de formulação do modelo, ajuste do modelo e inferência. A formulação do modelo parte da premissa de que o modelo deve representar as principais características da variável resposta dos dados observados. Assim, esta corresponde a uma combinação que envolve a escolha da distribuição da variável aleatória, das variáveis explanatórias que farão parte da matriz modelo e da função de ligação compatível com a distribuição proposta.

A etapa de ajuste representa o processo de estimação dos parâmetros lineares dos modelos e de determinadas funções das estimativas desses parâmetros, que representam medidas de adequação dos valores estimados (CORDEIRO, 2010). Desta forma o ajuste do modelo corresponde em determinar aquele mais parcimonioso, ou seja, com o menor número de variáveis explanatórias, entretanto, que ofereça uma boa interpretação e análise do problema de forma que a informação perdida não seja significativa. Em termos estatísticos, isto equivale a comparar o modelo ajustado com o saturado e verificar se a discrepância é significativa (OLIVEIRA, 2013).

Por fim, a etapa da inferência consiste em verificar a adequação do modelo em sua totalidade. Além disso, deve-se fazer a análise das discrepâncias resultantes do modelo, pois se estas se revelarem significativas as mesmas podem resultar de um modelo não representativo ou da presença de observações influentes nos dados. Segundo Cordeiro (2013) deve-se, nessa etapa, verificar a precisão e a interdependência das estimativas, testar os parâmetros de interesse, analisar estatisticamente os resíduos e realizar previsões.

### 3.2.4. Seleção do Modelo

O processo de escolha de um modelo que se adeque bem ao fenômeno em estudo deve ser feito mediante uma estratégia que permita testá-los formalmente com vista a ser obter o modelo mais apropriado para o caso. Isto se deve ao interesse de se verificar os efeitos que as variáveis consideradas e suas interações provocam nos

valores ajustados. Desta forma são naturalmente gerados vários modelos que provavelmente refletem de forma significativa o fenômeno em questão.

Em geral, o algoritmo de ajuste deve ser aplicado não a um MLG isolado, mas a vários modelos de um conjunto bem amplo que deve ser, realmente, relevante para a natureza das observações que se pretende analisar (PAULA, 2010). Ainda segundo o autor o fator escala não é tão crucial, pois a constância da variância e a normalidade não são essenciais para a distribuição da variável resposta. Entretanto, são comuns os casos em que primeiramente devem-se transformar os dados para um MGL produzindo um bom ajuste.

De acordo com Cordeiro (2010) outro ponto a ser considerado quanto à seleção do modelo refere-se às covariáveis a serem inseridas na sua estrutura linear. Desse modo se os dados apresentam  $m$  covariáveis, define-se então, um conjunto de  $2^m$  modelos. Assim, faz-se necessário selecionar um modelo com  $p$  variáveis explanatórias, sendo  $p \leq m$ , e que explique satisfatoriamente os dados observados. Após ser selecionado um modelo é importante fazer uma ponderação considerando a qualidade do ajuste do modelo e o seu grau de complexidade.

Alguns critérios para seleção de modelos usados para os MLGs são apresentados a seguir. Estes têm como funcionalidade comparar a qualidade do ajustamento de vários modelos, além de indicarem se os parâmetros do vetor  $\beta$  são significativamente diferentes de zero.

#### 3.2.4.1. Teste de Razão de Verossimilhança (Estatística de Wilks)

O teste da razão da verossimilhança compara os valores maximizados do logaritmo da função de verossimilhança de dois modelos. De acordo com Santos (2013) como a função de verossimilhança  $L(\beta)$  é inferior a 1, e geralmente muito pequena, é usual usar  $l(\beta) = \ln(L(\beta))$ , que multiplicado por  $-2$  o torna positivo, maior e com distribuição conhecida, a distribuição Qui-quadrado.

O teste é aplicado sob a condição de hipótese nula, ou seja,

$$H_0: \text{os dois modelos tem a mesma qualidade de ajuste}$$

A estatística de teste da razão de verossimilhança é definida por

$$RV = -2[l(\beta_0, y) - l(\hat{\beta}, y)] = 2[l(\hat{\beta}, y) - l(\beta_0, y)] \quad (3.13)$$

onde  $l(\hat{\beta}, y)$  é a função log-verossimilhança sem restrição, e  $l(\beta_0, y)$  é a função log-verossimilhança sob a hipótese nula.

Tem-se que  $H_0$  é rejeitado para um determinado nível de significância se,

$$RV > \chi_{p,1-\alpha}^2 \quad (3.14)$$

onde  $(1 - \alpha)$  é o número de graus de liberdade e  $p$  é a dimensão do vetor  $\beta$ .

### 3.2.4.2. Teste $F$

A estatística  $F$  para o caso de hipótese simples, é dada por

$$F = \frac{\frac{\{D(y;\mu_0) - D(y;\hat{\mu})\}}{p}}{\frac{D(y;\hat{\mu})}{(n-p)}} \quad (3.15)$$

Considerando-se o parâmetro de dispersão  $\phi \rightarrow \infty$  e sob a estatística nula  $H_0$  esta estatística segue uma distribuição  $F_{p,(n-p)}$ . De outra forma, a estatística  $F$  não depende do parâmetro de dispersão, como se vê na equação acima, sendo dependente apenas da função desvio.

### 3.2.4.3. Teste Wald

A estatística de Wald, por usar a estimativa não restrita de máxima verosimilhança, é útil na comparação de modelos quando se começa por formar o modelo maximal e se consideram modelos alternativos pela exclusão de covariáveis (TURKMAN, 2000).

O teste de Wald segue uma distribuição normal assintótica de  $\hat{\beta}$  e é dado por

$$W = (\hat{\beta} - \beta_0)^T K_0 (\hat{\beta} - \beta_0) \quad (3.16)$$

onde  $K_0$  representa a matriz de informação de Fisher dada por  $K_0 = (X^T W_0 X)$ . Assim, para grandes amostras rejeita-se  $H_0: \beta = \beta_0$  se

$$W > \chi_{p,1-\alpha}^2, \quad (3.17)$$

onde  $(1 - \alpha)$  é o número de graus de liberdade e  $p$  é a dimensão do vetor  $\beta$ .

Desta forma, segundo Santos (2013) se é pretendido testar a hipótese nula  $H_0: \beta_j = 0$  contra a hipótese alternativa  $H_1: \beta_j \neq 0$ , com  $j = 1, 2, \dots, p$ , a estatística de Wald resume-se a seguinte expressão

$$W = \frac{\hat{\beta}_j}{\sigma(\hat{\beta}_j)} \sim N(0,1) \quad (3.18)$$

Este teste mostra que se  $\beta_j = 0$ , para algum  $j$ , o coeficiente independente em questão é irrelevante para o modelo, e caso  $\beta_j \neq 0$ , a respectiva covariável é estatisticamente significativa para o modelo.

#### 3.2.4.4. Teste Rao (Estatística Score)

A estatística de Rao, ou teste score, pelo contrário, é útil na escolha de modelos, quando se parte do modelo nulo, i.e., o modelo sem covariáveis, ou de um modelo minimal e se consideram modelos alternativos pela inclusão de covariáveis (TURKMAN, 2000). A função score é definida por

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta}$$

Se definirmos o teste score para  $U(\beta) = 0$ , ele é calculado da seguinte forma

$$S = U(\beta_0)^T [K_0]^{-1} U(\beta_0) \quad (3.19)$$

onde  $K_0$  é a matriz de informação de Fisher dada por  $K_0 = (X^T W_0 X)$ , onde  $W_0$  é calculado sob a hipótese nula.

#### 3.2.4.5. Método de Akaike

Outro critério para a seleção do melhor modelo é o Critério de Informação de Akaike (AIC), o qual se baseia na função log-verossimilhança. Segundo Marciano (2009) este critério fornece uma medida relativa da informação perdida na adoção de um determinado modelo. Desse modo a escolha do modelo é feita a partir do menor valor apresentado por este critério. A estatística correspondente é dada por

$$AIC = -2l(\hat{\theta}) + 2p \quad (3.20)$$

onde  $\hat{\theta}$  é o estimador de máxima verossimilhança,  $l$  é a função de log-verossimilhança e  $p$  representa o número de parâmetros do modelo.

#### 3.2.5. Ajuste do Modelo

Geralmente quando se investiga um determinado fenômeno consideram-se todas as covariáveis  $X_p$ , com  $p = 1, 2, \dots, p$ , que teoricamente influenciam a variável resposta  $Y_i$ , com  $i = 1, 2, \dots, n$ , a fim de se fazer análises corretas do caso. No entanto, quando há uma quantidade consideravelmente grande de variáveis explicativas, o interesse é encontrar um modelo que proponha uma boa interpretação do fenômeno, dado que o mesmo apresente um número reduzido destes parâmetros ajustando-se bem aos dados. Sabe-se que o modelo nulo apresenta apenas ao parâmetro constante  $\beta_0$  correspondente a média  $\hat{\mu}$  de todas as observações do vetor  $Y$ . De outra forma, o modelo saturado possui  $n$  parâmetros, um para cada observação  $y_i$ , com  $i = 1, 2, \dots, n$ , isto significa que as estimativas são as próprias observações,  $\hat{\mu}_i = y_i$ . Oliveira (2013) diz que o modelo adequado tem resumir os dados parcimoniosamente, de forma que a informação perdida

não seja significativa. O autor ainda diz que, estatisticamente, isto significa comparar o modelo ajustado com o saturado e verificar se a discrepância é significativa.

Algumas medidas são usadas para verificar a bondade do ajuste tais como a *deviance*, com tradução encontrada na literatura como “desvio”, e a estatística generalizada de Pearson, as quais serão descritas adiante.

### 3.2.5.1. Deviance

A *deviance* avalia a discrepância entre os valores ajustados pelo modelo saturado e os valores ajustados pelo modelo proposto. Sua expressão é dada por

$$D(y; \hat{\mu}) = 2(l_s - l_a) \quad (3.21)$$

onde  $l_a = l(\hat{\mu}; y)$  representa a função log-verossimilhança do modelo proposto e  $l_s = l(y; y)$ , a função log-verossimilhança do modelo saturado.

A *deviance* assume sempre valores maiores ou iguais a zero, i.e.,  $D(y; \hat{\mu}) \geq 0$ . É demonstrado pela equação acima, que seu valor diminui à medida que se acrescenta covariáveis ao modelo, chegando ser igual a zero quando  $l_a = l_s$ . Para avaliar se o modelo proposto se ajusta bem aos dados, recorre-se ao teste de hipótese, sob a seguinte hipótese nula

$H_0$ : o ajustamento do modelo proposto é igual ao do modelo saturado

Tem-se então, que para amostras consideravelmente grandes, e sob  $H_0$ , que a *deviance* apresenta distribuição assintótica Qui-quadrado,

$$D \sim \chi^2_{(J-(p+1))} \quad (3.22)$$

onde  $J$  é o número de covariáveis e  $(p + 1)$ , o número de parâmetros do modelo proposto.

Assim a *deviance* para os modelos que são tratados neste trabalho fica expressa das seguintes maneiras,

#### Modelo de Poisson

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{y_i}{\mu_i} \right) - (y_i - \mu_i) \right] \quad (3.23)$$

#### Modelo Binomial Negativo

Para o Modelo Binomial Negativo a *deviance* é obtida da seguinte forma

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{y_i}{\mu_i} \right) - \left( y_i + \frac{1}{\alpha} \right) \log \left( \frac{1 + \alpha y_i}{1 + \alpha \mu_i} \right) \right] \quad (3.24)$$

### 3.2.5.2. Estatística de Pearson Generalizada

A estatística de Pearson generalizada é dada por

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad (3.25)$$

onde  $V(\hat{\mu}_i)$  é a função de variância para a distribuição do modelo em estudo, sendo  $V(\hat{\mu}_i) = \hat{\mu}$  para a distribuição de Poisson e  $V(\hat{\mu}_i) = \hat{\mu} + \alpha\hat{\mu}^2$  para a distribuição binomial negativa.

### 3.2.6. Superdispersão

A superdispersão é um fenômeno que frequentemente ocorre em dados de contagem, sendo indicada quando se observa que a variância é maior que a média. Um modelo linear generalizado que segue a distribuição de Poisson, por exemplo, onde tem-se que a variância e a média são iguais, muito provavelmente não representará bem os dados, pois a variância da variável resposta observada é maior do que a prevista por este modelo. Assim, o modelo de Poisson pode conduzir a uma análise distante do que de fato é o fenômeno estudado. De acordo com Kabacoff (2011), se a superdispersão está presente e não levada em consideração no modelo proposto, poderá ocorrer que os erros padrão e os intervalos de confiança sejam muito pequenos e o teste de significância, demasiado liberal, ou seja, serão encontrados efeitos que não estão realmente presentes.

Cordeiro (2013) diz que no ajuste de um modelo espera-se que o desvio residual seja aproximadamente igual ao seu número de graus de liberdade. A não ocorrência disso pode estar estritamente relacionada à componente aleatória do modelo, pela presença de uma variabilidade não percebida pelo modelo. Pode ainda estar ligada ao seu preditor linear, pela ausência de uma estrutura ótima, ou ainda, a uma função de ligação inadequada. Segundo Kabacoff (2011) a superdispersão é sugerida quando a razão entre o desvio residual e o seu grau de liberdade é muito maior que 1. Esta razão é assim expressa

$$\text{razão} = \frac{\text{desvio residual}}{\text{grau de liberdade residual}}$$

Se esta razão for igual a 1, então conclui-se que a variância é igual a média da variável resposta dos dados, propriedade chamada por Kleiber *et. all.* (2008) de equidispersão. Santos (2013) diz que alguns autores recomendam verificar a presença de superdispersão caso esta razão seja maior que 2. Logo, se seu valor for próximo de 1, pode-se assumir a não existência de superdispersão.

Em caso de haver uma quantidade substancial de superdispersão para a qual o modelo proposto não está totalmente especificado, como no caso do modelo baseado na distribuição de Poisson, uma solução possível é considerar uma distribuição mais flexível que não impõe a igualdade de média e variância. A distribuição mais amplamente utilizada neste contexto é a Binomial Negativo (KLEIBER *et. all.*, 2008).

### 3.2.7. Análise de Resíduos

A análise de resíduos permite avaliar se o nível de representatividade da realidade em estudo, dada pelo modelo, é adequado ou se este se afasta drasticamente do fenômeno observado. Para os modelos lineares generalizados assume-se que os erros  $\epsilon_i$ , com  $i = 1, 2, \dots, n$ , satisfazem os seguintes pressupostos: são independentes entre si, e possuem distribuição normal com média zero ( $\mu = 0$ ) e variância constante  $\sigma^2$ , ou seja,  $\epsilon_i \sim N(0, \sigma^2)$ .

A qualidade de um ajuste pode ser, portanto, verificada através da análise de resíduos. Esta permite avaliar se a função de distribuição da variável resposta, a função de ligação e as variáveis que compõem o preditor linear são os mais adequados possíveis. Além disso, os resíduos são úteis para determinar a presença de pontos aberrantes nos resultados do modelo proposto, e se estes são ou não influentes.

Entende-se por resíduos a discrepância entre os valores observados  $y_i$  e os valores ajustados  $\hat{\mu}_i$ . Para os modelos lineares generalizados são definidos três tipos diferentes de resíduos capazes de avaliar o ajuste do modelo: os resíduos de Pearson, a *deviance* residual e os resíduos de Ascombe. No presente trabalho usaremos apenas os dois primeiros, os quais serão apresentados a seguir.

#### 3.2.7.1. Resíduo de Pearson

O resíduo de Pearson é dado pela seguinte expressão

$$r_i^P = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\sqrt{\text{Var}(\hat{Y}_i)}} \quad (3.26)$$

onde  $\hat{\mu}_i$  é a média ajustada de  $\hat{Y}_i$  e  $\text{Var}(\hat{Y}_i)$  é a sua função de variância.

O resíduo  $r_i^P$ , definido acima, corresponde à contribuição de cada observação para o cálculo da estatística de Pearson generalizada. De acordo com Cordeiro (2013) esta quantidade é uma componente da estatística de Pearson generalizada  $X_p^2 = \sum_{i=1}^n r_i^{P^2}$ . Ainda segundo o autor, a desvantagem deste tipo de resíduo é que a sua distribuição é, geralmente, bastante assimétrica para modelos não-normais.

Segundo Turkman (2000) têm-se  $\text{Var}(Y_i - \hat{\mu}_i) \approx \text{Var}(Y_i)(1 - h_{ii})$ . Desse modo, o resíduo de Pearson padronizado (studentizado) é dado por

$$r_i^{*P} = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\sqrt{\text{Var}(Y_i)(1 - h_{ii})}} \quad (3.27)$$

onde  $h_{ii}$  é o  $i$ -ésimo elemento da diagonal principal da matriz de projeção  $H$ . Os resíduos de Pearson padronizados tem, aproximadamente, variância igual a um quando o parâmetro de dispersão  $\phi \rightarrow 0$  (CORDEIRO, 2013).

### 3.2.7.2. Deviance Residual

A *deviance* residual é obtida a partir da função desvio. Esta representa uma parcela correspondente à contribuição da  $i$ -ésima observação para a medida total da discrepância da *deviance*  $D(y; \hat{\mu})$ . Desse modo, tomando-se a Equação 4.21 como referência, cada uma dessas parcelas pode ser escrita da seguinte forma

$$\begin{aligned} d_i &= 2(l_s - l_a) \\ d_i &= 2[l(y_i; y_i) - l(\hat{\mu}_i; y_i)] \end{aligned} \quad (3.28)$$

Assim, considerando o parâmetro de dispersão  $\phi = 1$ , a *deviance* pode ser escrita em termos de suas componentes do desvio, sendo expressa como

$$D(y; \hat{\mu}) = \sum_{i=1}^n d_i \quad (3.29)$$

Portanto a *deviance* residual é dada pela seguinte expressão

$$r_i^D = \delta_i \sqrt{d_i}, \quad (3.30)$$

onde  $\delta_i = \text{sign}(y_i - \hat{\mu}_i)$ . Entretanto, a *deviance* residual padronizada é a mais utilizada para os modelos lineares generalizados e é definido como segue abaixo

$$r_i^{*D} = \frac{r_i^D}{\sqrt{\hat{\phi}(1 - h_{ii})}} \quad (3.31)$$

Além destes resíduos há também os resíduos de Ascombe sobre os quais facilmente se encontra abordagens na literatura, porém este não é mencionado aqui, pois para o presente trabalho os resíduos definidos acima foram suficientes.

### 3.2.8. Análise de Diagnóstico

Durante a etapa de análise do ajuste do modelo deve-se considerar a verificação de possíveis afastamentos das suposições propostas para modelo tais como a suposição de independência para os erros aleatórios  $\epsilon_i$ 's, com  $i = 1, 2, \dots, n$ , e que  $\epsilon_i \sim N(0, \sigma^2)$ , pois se estas pressuposições não forem satisfeitas, a análise dos resultados podem resultar em erros grosseiros. Além disso, de acordo com Cordeiro (2013), outro fato bastante comum é a presença de pontos atípicos (falhas isoladas), que podem influenciar, ou não, no ajuste do modelo.

Estas observações discrepantes podem surgir por falhas humanas como medidas erradas das observações, erro de transcrição, modelo mal especificado, escala usada de forma errada, além daquelas situações que fogem ao controle humano, gerando assim, um peso desproporcional nas estimativas dos parâmetros do modelo. Para avaliar a

presença destes pontos atípicos e verificar se possuem influência significativa no modelo serão apresentadas algumas técnicas desenvolvidas para este fim.

### 3.2.8.1. Pontos de Alavanca

Os pontos de alavanca, também encontrado na literatura como *leverage*, são observações discordantes assim designadas por possuírem um perfil diferente dos demais pontos no que diz respeito aos valores das variáveis explicativas, produzindo um peso desproporcional nos respectivos valores ajustados. De acordo com Turkman (2000) a *leverage* mede o efeito que a observação tem nos valores preditos, sendo um indicativo de quão influente uma observação é.

Para os MLG's as observações conhecidas como pontos de alavanca podem ser identificadas pelos elementos  $h_{ii}$  da matriz de projeção generalizada cuja expressão é dada a seguir como

$$H = W^{\frac{1}{2}}X(X^TWX)^{-1}X^TW^{\frac{1}{2}} \quad (3.32)$$

onde  $W$  é a matriz de pesos e  $h_{ii}$  são os elementos da diagonal principal de  $H$ .

A matriz  $H$  é simétrica e idempotente de tal forma que  $\text{posto}(H) = \sum_{i=1}^n h_{ii} = p$ . Assim, em média, pode-se esperar que  $h_{ii} \approx \frac{p}{n}$ , sendo  $0 \leq h_{ii} \leq 1$ . Segundo Paula (2010), em particular, se  $h_{ii} = 1$  implica em  $\hat{\mu}_i = y_i$ , entretanto a recíproca não é verdadeira. Além disso, de acordo com o autor, para valores altos de  $h_{ii}$  predomina a influência de  $y_i$  sobre o correspondente valor ajustado. Desta forma consideram-se pontos de alavanca aqueles cuja medida de  $h_{ii}$  é tal que

$$h_{ii} \geq \frac{2p}{n} \quad (3.33)$$

### 3.2.8.2. Pontos Influentes

Uma observação é considerada realmente influente se a sua omissão no conjunto de dados resulta em mudanças substanciais nas estatísticas de diagnósticos do modelo (CORDEIRO, 2013). Assim, uma proposta interessante para se averiguar a real influência de um ponto discrepante, como é o caso de um ponto de alavanca, consiste na retirada (ou deleção) da observação considerada influente, e assim, comparar o ajuste do modelo escolhido sem esta observação particular, com o ajuste do mesmo modelo incluído todas as observações.

Segundo Turkman (2000) um indicador da  $i$ -ésima observação no vetor estimado  $\hat{\beta}$ , pode ser calculado pela diferença  $\hat{\beta} - \hat{\beta}_{(i)}$ , onde  $\hat{\beta}_{(i)}$  representa a estimativa

de máxima verossimilhança do vetor  $\hat{\beta}$  sem a observação  $y_i$ , e  $\hat{\beta}_{(i)}$  representa a mesma estimativa, porém com todas as observações.

Para os modelos lineares generalizados a Distância entre verossimilhanças, usada para calcular esta diferença é dada por

$$LD_i = \frac{2}{p} [l(\hat{\beta}) - l(\hat{\beta}_{(i)})] \quad (3.34)$$

onde  $l(\cdot)$  é a função de log-verossimilhança.

Expandindo-se  $LD_i$  em série de Taylor, obtém-se

$$\hat{\beta}_{(i)} = \hat{\beta} - w_i^{\frac{1}{2}} (1 - h_{ii})^{\frac{1}{2}} r_i^{*P} (X^T W X)^{-1} x_i \quad (3.35)$$

onde  $w_i$  é elemento da matriz de pesos  $W$ . Segundo Turkman (2000), William (1987) propõe que esta é uma fórmula mais simples para aproximar  $\hat{\beta}_{(i)}$  do seu valor real. Assim, Oliveira (2013) diz que a expressão acima pode ser aproximada pela Distância generalizada de Cook, definida como

$$D_i = \frac{h_{ii}}{p(1-h_{ii})} (r_i^{*P})^2, \text{ com } r_i^{*P} = \frac{(y_i - \hat{\mu}_i)}{\sqrt{V(\hat{\mu}_i)(1-h_{ii})}} \quad (3.36)$$

onde  $p$  é o posto da matriz modelo  $X$  e  $r_i^{*P}$  é o resíduo de Pearson padronizado.

Portanto,  $D_i$  será grande quando a  $i$ -ésima observação for aberrante. Uma proposta para avaliar uma observação é considera-lo influente se  $D_i > \frac{\chi_p^2(\alpha)}{p}$ .

### 3.2.9. Análises Gráficas

A análise gráfica corresponde a técnicas usadas para auxiliar na verificação do ajuste de um modelo. Estas são de âmbito informal, pois se baseia em análise visual de gráficos para detectar padrões ou pontos atípicos. Alguns tipos de gráficos são importantes para se verificar a presença de pontos aberrantes ou influentes no modelo proposto os quais serão apresentados a seguir. Para a geração destes gráficos utilizou-se o software R, pois mesmo possui pacotes apropriados que permite a criação destes com certa facilidade.

#### 3.2.9.1. Gráfico Normal de Probabilidade para $r_i^{*D}$ com Envelope

O gráfico normal de probabilidade possibilita identificar a distribuição originária dos dados, bem como pontos atípicos. Desse modo, este recurso permite detectar o afastamento da suposição de normalidade dos resíduos. Espera-se ainda, que ausência de pontos discordantes, seu aspecto seja o de uma reta, sendo colinear da reta ajusta. Entretanto, percebe-se que dificilmente os pontos do gráfico seguem exatamente a disposição da reta ajustada mantendo certa variação em torno desta.

Para a construção do gráfico seminormal de probabilidade utilizam-se os valores de  $|r_i^{*P}|$  versus  $z_i$ , onde  $r_i^{*P}$  é a *deviance* residual padronizada e  $z_i$  é dado pela seguinte expressão

$$z_i = \frac{\phi^{-1}(i+n-0,125)}{(2n+0,5)} \quad (3.37)$$

onde  $i = 1, 2, \dots, n$ .

Em virtude da disposição irregular dos resíduos em torno da reta ajustada tornar a avaliação visual um tanto complexa, criou-se uma espécie de banda de confiança para estes pontos possibilitando verificar se este afastamento da normalidade realmente é significativo. Para a construção desta região Paula (2010) diz que Atkinson (1985) sugere a construção de um tipo de banda de confiança através de simulações, a qual denominou de envelope. Estes são construídos com resíduos gerados a partir do modelo ajustado.

### 3.2.9.2. Gráfico de $r_i^{*D}$ versus Valores Ajustados $\hat{\mu}_i$

Este gráfico permite verificar a heterocedasticidade da variância e a presença de pontos discrepantes. Segundo Marciano (2009) o que se espera é que o gráfico apresente a distribuição dos resíduos em torno de zero com amplitude constante, onde desvios sistemáticos podem ter algum tipo de curvatura ou uma amplitude muito diferente do valor ajustado.

### 3.2.9.3. Gráfico de Índices

Este tipo de gráfico é gerado assumindo-se que no eixo horizontal estão as ordens das observações. Sua análise visual permite verificar a presença de pontos atípicos que poderão ser classificados como influentes, não influentes ou como pontos aberrantes (*outlier*). Para o presente trabalho destacam-se os seguintes gráficos

- i. *deviance* residual padronizada ( $r_i^{*D}$ ): útil para detectar pontos aberrantes;
- ii. pontos de alavanca  $h_{ii}$  (*leverage*): utilizado para observar pontos influentes;
- iii. distância de Cook ( $D_i$ ): usado para confirmar se os pontos tidos como *leverage* são realmente influentes.

## 3.3. Estimativa da Densidade Kernel

Esta seção não trata especificamente de um tópico dos MLGs, mas foi inserida neste capítulo por se tratar de uma ferramenta muito útil para o método proposto neste

trabalho, pois permitirá verificar a concentração das descargas parciais em torno das variáveis explicativas.

Guidoum (2015) diz que o kernel é um tipo especial de função de ponderação  $K(x)$  que possui as seguintes propriedades: possui valores reais, é uma função não-negativa  $K(x) \geq 0$ , é simétrica  $\int_{\mathbb{R}} xK(x)dx = 0$  e a integral definida sobre seu conjunto suporte é igual a 1,  $\int_{\mathbb{R}} K(x)dx = 1$ . Alguns tipos de função kernels são definidos sendo os mais comuns o Gaussiano, Retangular, Triangular, Biweight e Epanechnikov.

Segundo o autor a estimativa de densidade de kernel é um método não-paramétrico de estimar a função densidade de probabilidade de uma variável aleatória, pois não assume qualquer distribuição subjacente para a variável. Essencialmente, para cada dado é criada uma função de kernel com o ponto de referência no seu centro. A função densidade de probabilidade é então estimada pela adição de todas estas funções do kernel e dividindo o resultado pelo número de dados. Assim, a estimativa pode ser dada pela seguinte expressão

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right), \quad (3.38)$$

onde  $K(\cdot)$  é a função de kernel devidamente escolhida,  $n$  é a quantidade de dados,  $h$  é a largura de banda (*bandwidth*) ou parâmetro de suavização e  $x_i$  representa cada dado.

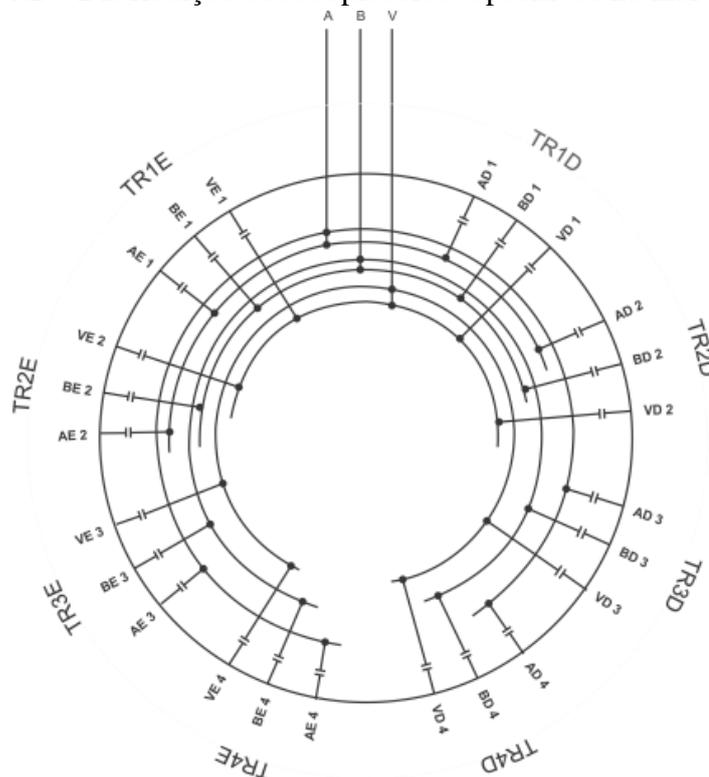
A *bandwidth* é um parâmetro que controla o grau de suavização aplicado aos dados. Em consequência deve ser escolhida de forma eficiente, pois se a largura de banda é pequena, obtém-se um estimador pouco suave, com alta variabilidade. Pelo contrário, se o valor de  $h$  é grande, o estimador resultante será mais suave e mais longe da função que se pretende estimar. O R possui pacotes de funções que facilitam o cálculo das estimativas e, por padrão utilizam o kernel gaussiano escolhendo automaticamente o melhor valor para  $h$  de acordo com os dados analisados.

## Capítulo 4

### 4. Aplicação em Descargas Parciais

Para a realização desta pesquisa foram utilizados os registros de descargas parciais da base de dados do centro de pesquisa da Eletronorte. Os dados processados são provenientes de um dos hidrogeradores da Usina Hidroelétrica de Tucuruí, que foram coletados no mês de setembro de 2014 a partir de 24 acopladores capacitivos conectados em ambos os lados do estator, sendo este alimentado pelas fases A, B e V. Esses acopladores foram distribuídos em oito grupos nas três fases, sendo quatro grupos em cada lado do hidrogerador. A Figura 4.1 mostra a forma como os acopladores foram distribuídos ao longo de sua estrutura.

**Figura 4.1** – Distribuição dos acopladores capacitivos no hidrogerador.



Baseado na forma como os acopladores foram dispostos na estrutura do hidrogerador, e pelo fato dos dados serem analisados por grupo de acopladores, foi utilizada uma notação própria neste trabalho com o intuito de facilitar a análise pretendida, a qual poderá ser adotada em estudos semelhantes para outros hidrogeradores com arquitetura similar. Assim, a denominação Tríade se referiu a cada agrupamento de acopladores (fases A, B e V), os índices de 1 a 4 indicaram a ordem do grupo em cada lado do hidrogerador e a terminação D e E indicou o lado direito e

esquerdo respectivamente. Por exemplo, a notação TR3E, se referiu à terceira tríade do lado esquerdo, ou terceiro grupo de acopladores do lado esquerdo do hidrogerador.

Para viabilizar a utilização desses dados em programas que permitem uma análise estatística satisfatória para este caso, foi utilizado um aplicativo desenvolvido em *Delphi* para converter o formato gerado pelo equipamento de detecção de descargas parciais para a extensão “.txt” reconhecida pelo *software* R para posterior análise de dados e formulação dos modelos.

#### 4.1. Organização dos dados

Os dados brutos, obtidos a partir de cada acoplador, foram agrupados de acordo com as suas respectivas tríades, e em seguida organizados em tabelas considerando o número de ocorrências de descargas parciais com sua respectiva amplitude e ângulo de fase. As ocorrências de descargas foram organizadas por tríade (fases A, B e V) e lado considerando a amplitude e o ângulo. Foi considerada a ocorrência de descargas em intervalos de classe de 74 mV para amplitude, e, para os ângulos em intervalos de classe de 45°. A Tabela 4.1 exibe as ocorrências de descargas parciais nos intervalos de amplitude e ângulo para a primeira tríade (nas fases A, B e V) do lado esquerdo denotada por TR1E. Salienta-se ainda, que não foram utilizados os dados obtidos por três acopladores, sendo um do lado esquerdo conectado à fase B e dois do lado direito acoplados à fase V.

**Tabela 4.1** – Dados da tríade TR1E para as amplitudes de -375mV a +749mV.

Amplitude (mV)	Ângulo (°)								Total
	0-45	45-90	90-135	135-180	180-225	225-270	270-315	315-360	
-375 – -301	1	-	1	-	-	1	-	-	3
-300 – -226	2	-	-	-	2	2	-	-	6
-225 – -151	3	2	3	4	5	13	-	2	32
-150 – -76	9	34	48	48	23	22	-	34	218
-75 – -1	211	210	243	249	189	208	123	235	1668
0 – 74	268	203	191	187	218	272	195	268	1802
75 – 149	334	116	185	171	183	355	186	365	1895
150 – 224	125	22	23	34	9	41	25	119	398
225 – 299	64	12	-	3	-	4	-	72	155
300 – 374	28	4	-	-	-	-	-	21	53
375 – 449	17	-	-	-	-	-	-	14	31
450 – 524	1	1	-	-	-	-	-	4	6
525 – 599	1	-	-	-	-	-	-	2	3
600 – 674	2	-	-	-	-	-	-	2	4
675 – 749	2	-	-	-	-	-	-	-	2
<b>Total</b>	1068	604	694	696	629	918	529	1138	6276

Também foi considerado que existe uma estrutura de correlação das medições nos vinte e dois sensores, assim os dados das tríades foram analisados isoladamente, embora se tenha gerado um modelo englobando as ocorrências de descargas parciais de todas as tríades, considerando o lado e as tríades como fatores.

A amplitude teve grande variação em todas as tríades, porém foram considerados neste estudo os intervalos variando de -300mV a +300mV, por apresentarem a maior massa de dados.

**Tabela 4.2** – Organização dos dados da tríade TR1E variando de -300mV a +300mV.

Amplitude (mV)	Ângulo (°)								Total
	0-45	45-90	90-135	135-180	180-225	225-270	270-315	315-360	
-300 – -226	2	-	-	-	2	2	-	-	6
-225 – -151	3	2	3	4	5	13	-	2	32
-150 – -76	9	34	48	48	23	22	-	34	218
-75 – -1	211	210	243	249	189	208	123	235	1668
0 – 74	268	203	191	187	218	272	195	268	1802
75 – 149	334	116	185	171	183	355	186	365	1895
150 – 224	125	22	23	34	9	41	25	119	398
225 – 299	64	12	-	3	-	4	-	72	155
<b>Total</b>	1016	599	693	696	629	917	529	1095	6174

## 4.2. Análise exploratória dos dados

A análise exploratória consistiu no estudo dos *kernels* considerando todas as ocorrências de descargas por tríade (fases A, B e V) em intervalos de classe de 74 mV para amplitude, e, para os ângulos em intervalos de classe de 45°.

E dos histogramas e dos *boxplots* considerando as ocorrências de descargas de -300mV a 300mV por tríade (fases A, B e V) em intervalos de classe de 74 mV para amplitude, e para os ângulos em intervalos de classe de 45°, tendo em vista à modelagem estatística.

### 4.2.1. Kernel da densidade de ocorrência de descargas parciais nas amplitudes e ângulos nas fases do hidrogerador de todas as descargas.

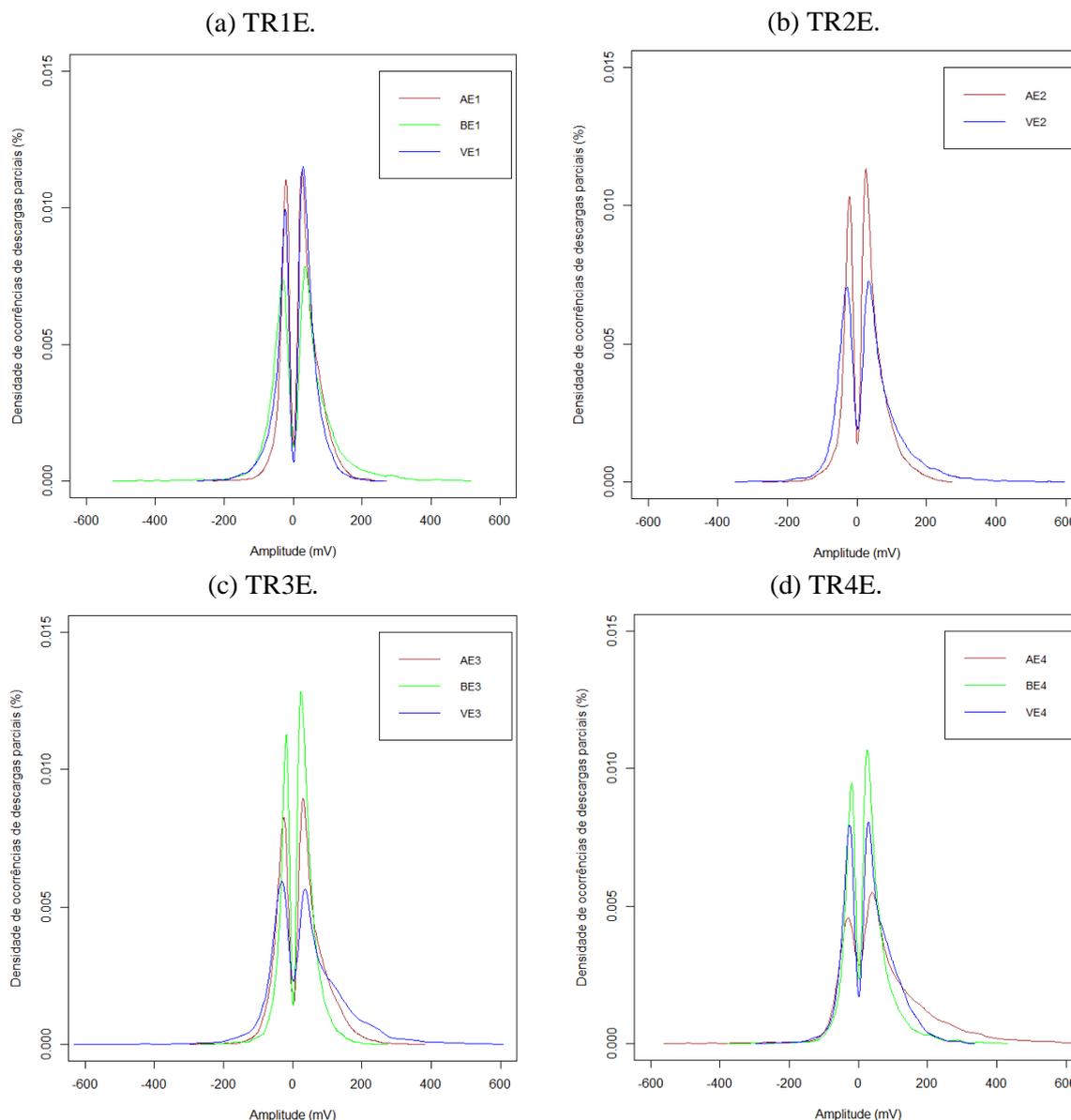
#### Lado esquerdo

No lado esquerdo, mais de 96% das descargas parciais em todas as tríades ocorreram no intervalo de amplitude de -300mV a 300mV. E, o número de ocorrências de descargas em amplitudes positivas foi superior ao número de descargas em amplitudes negativas.

A Figura 4.2 mostra os gráficos dos *kernels* das densidades das ocorrências de DPs nas amplitudes das tríades do lado esquerdo do hidrogerador. Desse modo, foi

observado que as densidades apresentaram comportamento simétrico, com a maior densidade nos intervalos  $(-100,0)$  e  $(0,+100)$ . A exceção ocorre para a fase V da tríade TR3E e fase A da TR4E, pois para o mesmo intervalo apresentam menor densidade.

**Figura 4.2** – Gráfico do kernel de densidade das amplitudes das descargas das tríades do lado esquerdo.



A TR1E apresentou 0,05% das ocorrências em amplitudes menores que -300mV, 98,38% no intervalo de -300mV a 300mV e 1,58% acima de 300mV. Do total de 6276 descargas, 30,70% foram em amplitudes negativas e 69,30% em amplitudes positivas, com o maior número de ocorrências no intervalo de ângulo 315 a 360° representando 18,13% seguido do intervalo de 0 a 45° com 17,02% das ocorrências.

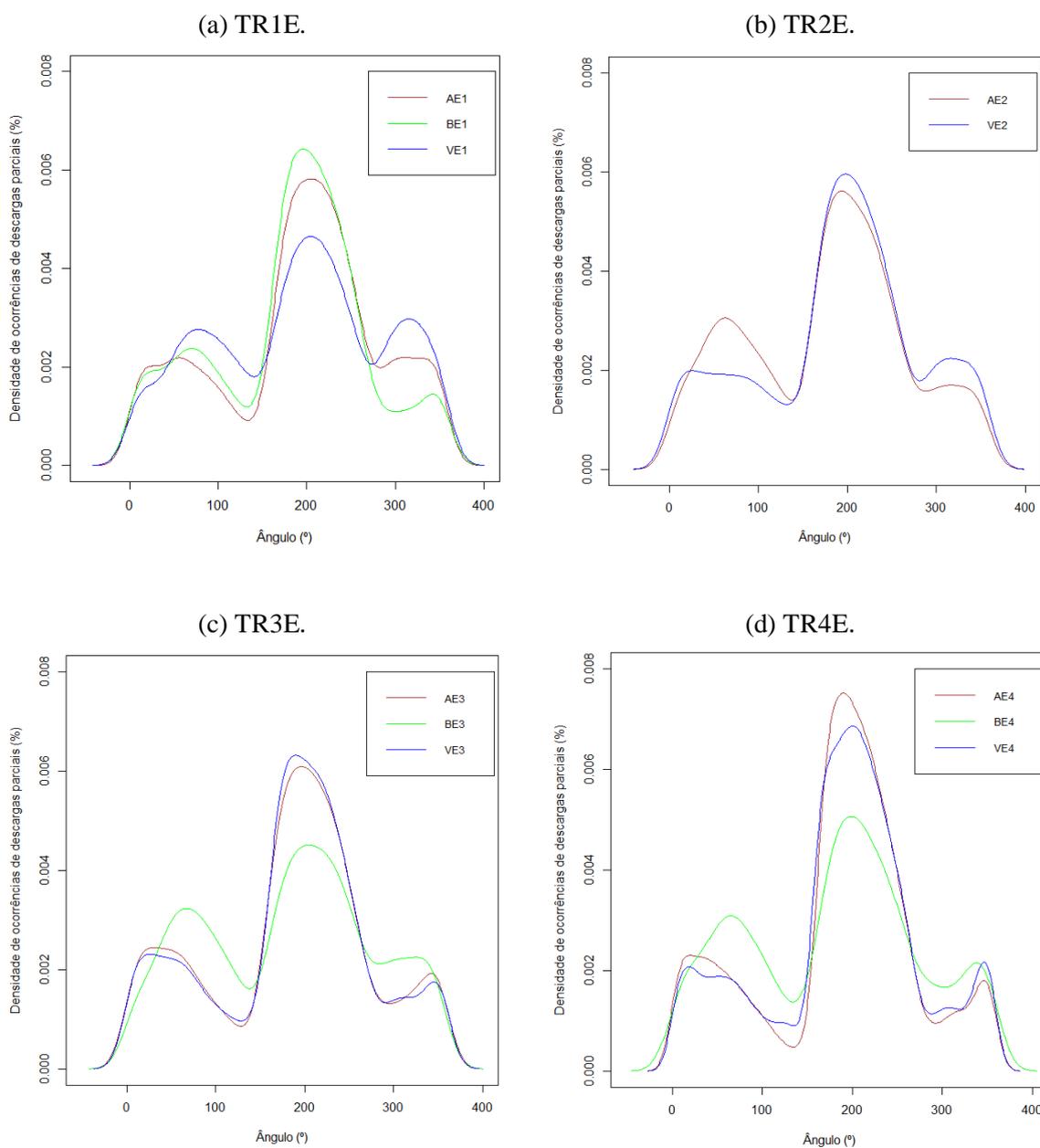
A TR2E apresentou 0,06% das ocorrências em amplitudes menores que -300mV, 98,79% no intervalo de -300mV a 300mV e 1,15% acima de 300mV. Do total de 4695 descargas, 32,12% foram em amplitudes negativas e 67,88% em amplitudes positivas, com o maior número de ocorrências no intervalo de ângulo 90 a 135° com 23,41% das ocorrências.

A TR3E apresentou 0,07% das ocorrências em amplitudes menores que -300mV, 97,15% no intervalo de -300mV a 300mV e 2,78% acima de 300mV. Do total de 6872 descargas, 34,08% foram em amplitudes negativas e 65,92% em amplitudes positivas, com o maior número de ocorrências no intervalo de ângulo 90 a 135° representando 19,78% das ocorrências.

A TR4E apresentou 0,05% das ocorrências em amplitudes menores que -300mV, 96,99% no intervalo de -300mV a 300mV e 2,96% acima de 300mV. Do total de 6013 descargas, 28,29% foram em amplitudes negativas e 71,71% em amplitudes positivas, com o maior número de ocorrências no intervalo de ângulo 225 a 270° representando 17,16% das ocorrências.

A Figura 4.3 mostra os gráficos dos *kernels* das densidades de ocorrências nos ângulos, sendo observado que as densidades se concentram em três faixas de intervalo em cada tríade, sendo elas de (0, +150), (+150, +250) e (+250, +360), onde a maior concentração de descargas parciais se encontra na segunda faixa intervalar. Este comportamento é o mesmo para todas as tríades do lado esquerdo.

**Figura 4.3** – Gráfico do kernel de densidade dos ângulos das descargas das tríades do lado esquerdo.



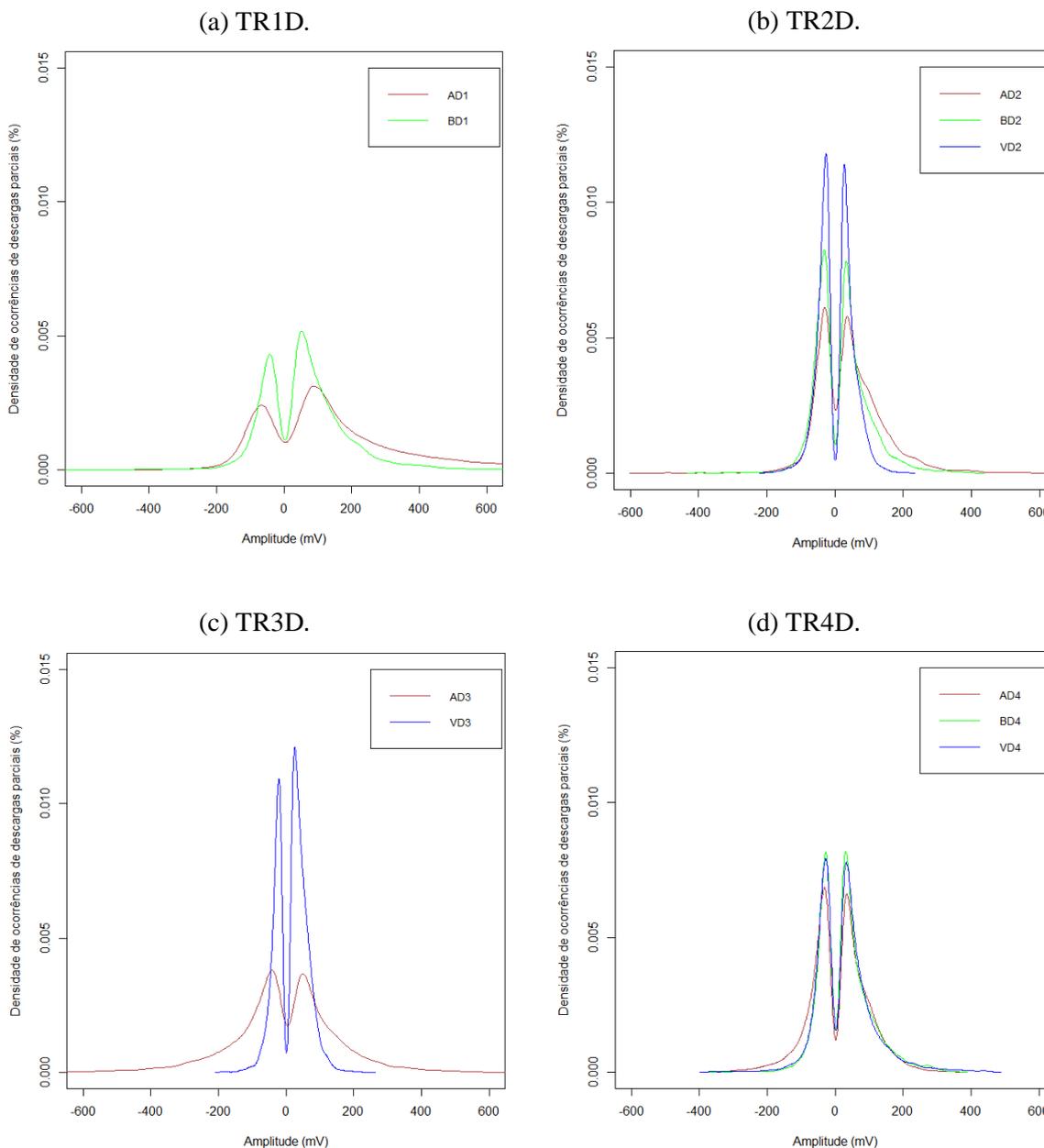
### Lado Direito

No lado direito, mais de 96% das descargas parciais em todas as tríades ocorreram no intervalo de amplitude de -300mV a 300mV. Neste lado também, o número de ocorrências de descargas em amplitudes positivas foi superior ao número de descargas em amplitudes negativas.

De acordo com a Figura 4.4, correspondente ao lado direito do hidrogerador, foi verificado que o *kernel* das densidades das amplitudes das descargas da tríade TR2D também gerou gráficos semelhantes, acontecendo o mesmo com as tríades TR1D e TR4D. No grupo TR3D o acoplador da fase V apresentou alta densidade no intervalo

(-100,0) e (0,+100), enquanto o acoplador da fase A exibiu baixa densidade para o mesmo intervalo.

**Figura 4.4** – Gráfico do kernel de densidade das amplitudes das descargas das tríades do lado direito.



A TR1D apresentou o maior número ocorrências de descargas de todas as tríades apesar de se utilizar somente dois acopladores. Esta apresentou 0,26% das ocorrências em amplitudes menores que -300mV, 88,09% no intervalo de -300mV a 300mV e 11,65% acima de 300mV. Do total de 8214 descargas, 29,17% foram em amplitudes negativas e 70,83% em amplitudes positivas, com o maior número de ocorrências no intervalo de ângulo 225 a 270° representando 23,09% das ocorrências.

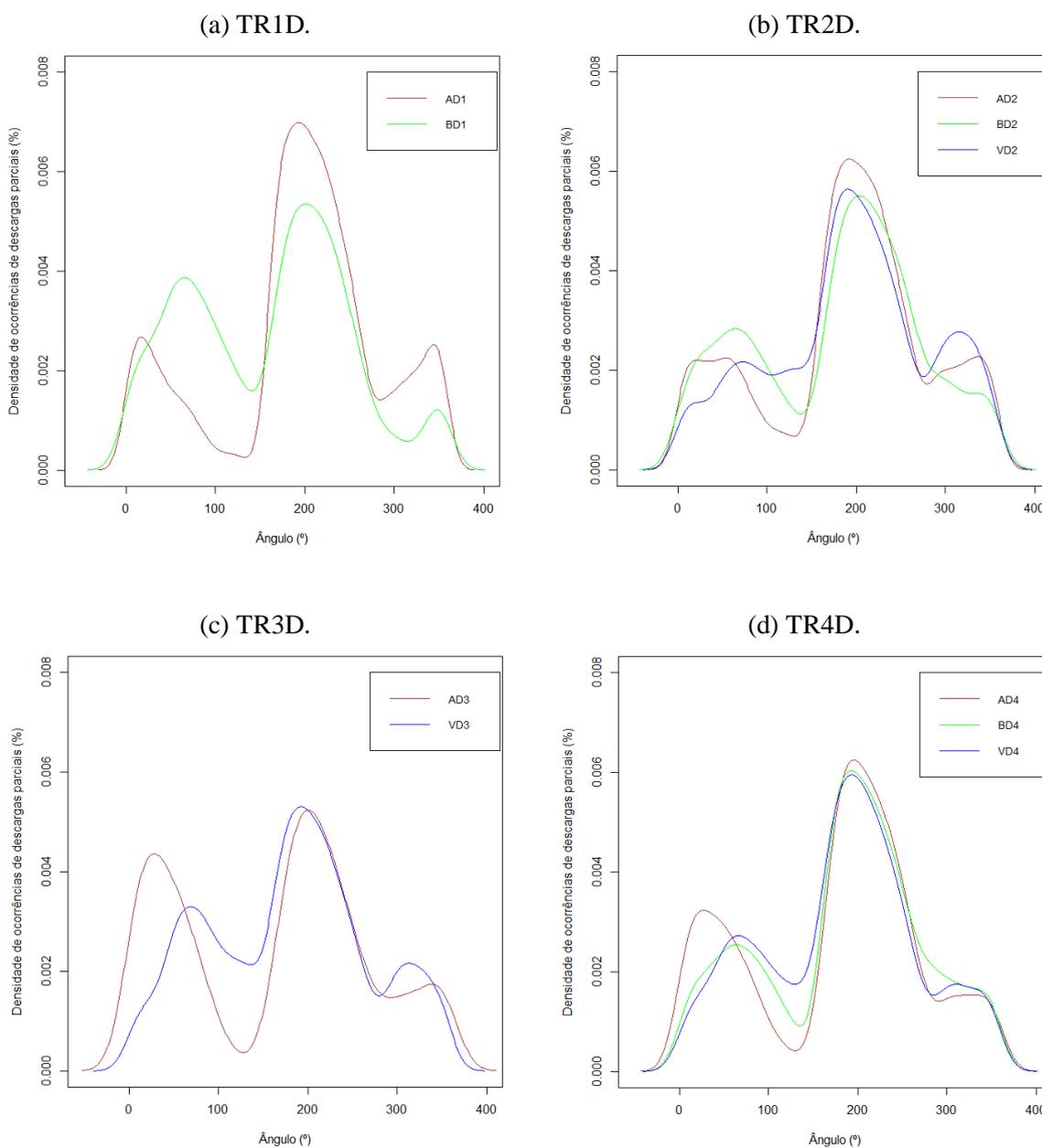
Das ocorrências em amplitudes menores que -300mV, a tríade TR2D apresentou 0,12%, no intervalo de -300mV a 300mV apresentou 98,18% e 1,70% acima de 300mV. Do total de 6873 descargas, 36,62% foram em amplitudes negativas e 63,38% em amplitudes positivas, com o maior número de ocorrências no intervalo de ângulo 225 a 270° com 18,26% das ocorrências.

A TR3D apresentou 4,31% das ocorrências em amplitudes menores que -300mV, 92,65% no intervalo de -300mV a 300mV e 3,04% acima de 300mV. Do total de 6353 descargas, 45,44% foram em amplitudes negativas e 54,56% em amplitudes positivas, com o maior número de ocorrências no intervalo de ângulo 45 a 90° representando 20,95%, seguido pelo intervalo de 225 a 270° representando 18,83% das ocorrências.

A TR4D apresentou 0,50% das ocorrências em amplitudes menores que -300mV, 97,94% no intervalo de -300mV a 300mV e 1,57% acima de 300mV. Do total de 7466 descargas, 41,27% foram em amplitudes negativas e 58,73% em amplitudes positivas, com o maior número de ocorrências no intervalo de ângulo 45 a 90° representando 17,21% das ocorrências.

Os gráficos dos *kernels* das densidades de ocorrências nos ângulos para as tríades do lado direito são exibidos na Figura 4.5. Observa-se um comportamento similar ao das tríades esquerdas, pois o maior número de descargas parciais se concentra nos intervalos de (0,+150), (+150,+250) e (+250,+360), exibindo maior densidade na segunda faixa intervalar. Na tríade TR3D, porém, verificou-se um aspecto diferente das demais, apresentando curvas distintas para os dois acopladores (AD3 e VD3) e com densidades mais homogêneas nas duas primeiras faixas intervalares.

**Figura 4.5** – Gráfico do kernel de densidade dos ângulos das descargas das tríades do lado direito.



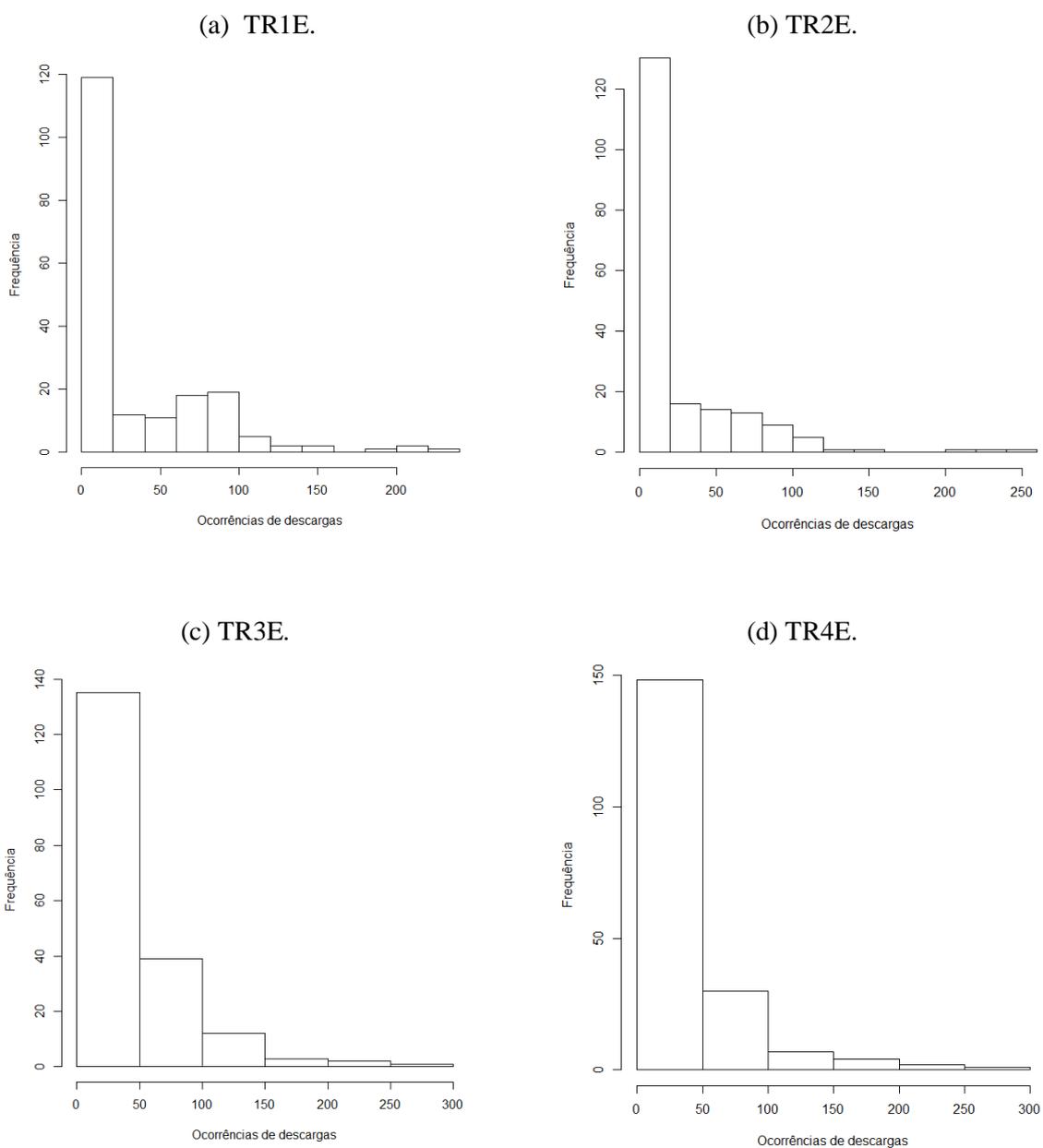
#### 4.2.2. Histogramas de todas as tríades para as descargas no intervalo de amplitude de -300mV a 300mV

Os histogramas das Figuras 4.6 e 4.7 foram feitos considerando as ocorrências de descargas de -300mV a 300mV por tríade (fases A, B e V) em intervalos de classe de 74 mV para amplitude, e para os ângulos em intervalos de classe de 45° já que em todas as tríades a maior concentração dos dados se encontra nestas amplitudes. Na Tabela 4.2 são exibidos algumas medidas estatísticas: média, mediana, desvio padrão, assimetria e curtose, que auxiliarão na interpretação gráfica.

É possível observar nas figuras e tabela citadas que todas as tríades apresentam assimetria positiva, com uma maior concentração dos dados no lado esquerdo. Este pode ser um indício que os modelos de Poisson e Binomial Negativo se ajustam aos dados.

Quanto à curtose são platicúrticas ( $C_p < 0,263$ ) as distribuições das tríades 2 e 4 do lado esquerdo e as demais leptocúrticas ( $C_p > 0,263$ ).

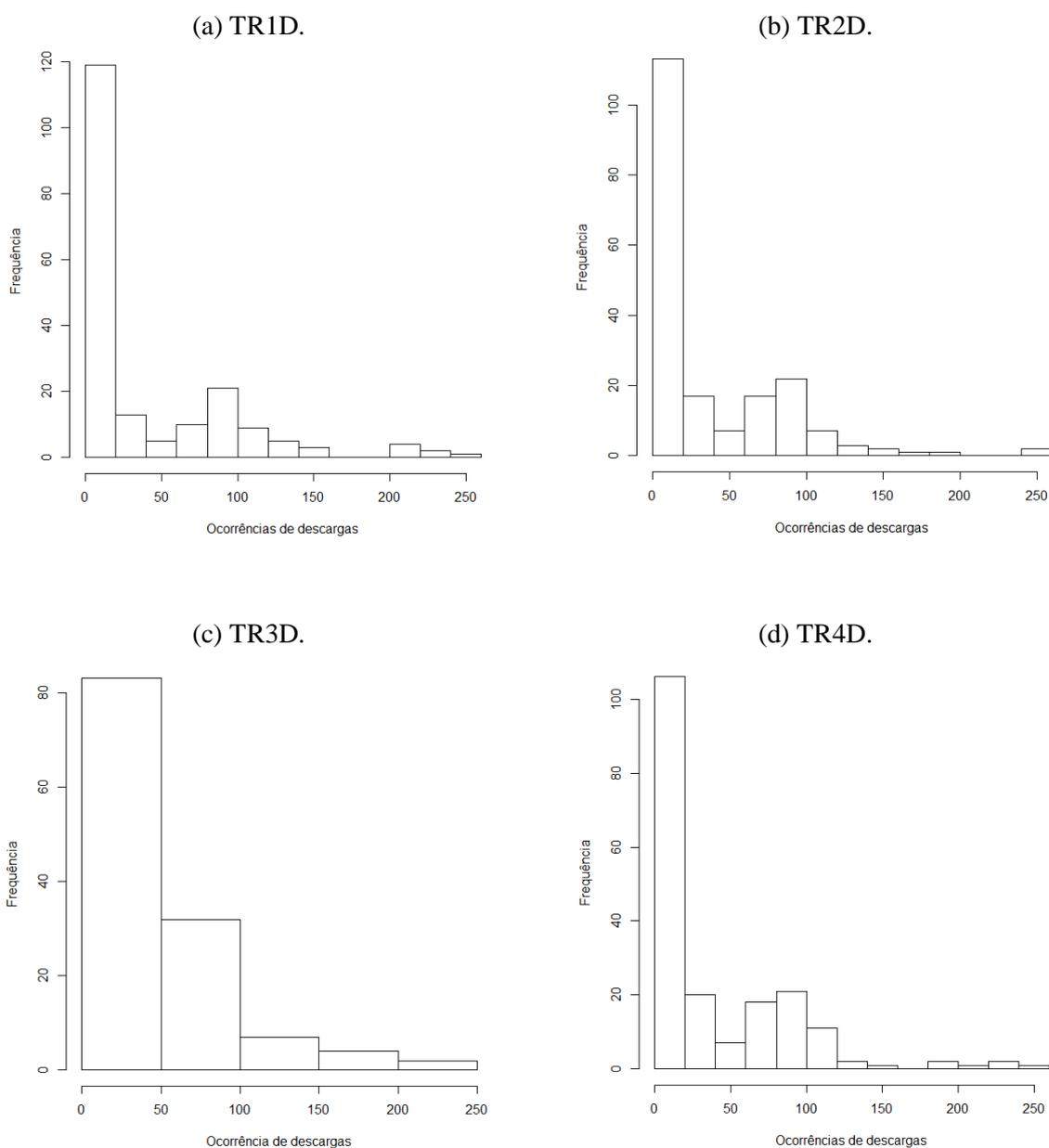
**Figura 4.6** – Histograma das ocorrências de descargas das tríades do lado esquerdo no intervalo de amplitude de -300mV a 300mV



**Tabela 4.3** – Medidas estatísticas para cada tríade do hidrogerador no intervalo de amplitude de -300mV a 300mV.

Estadística	TR1E	TR2E	TR3E	TR4E	TR1D	TR2D	TR3D	TR4D
Média	32,16	24,16	34,77	30,37	37,69	35,15	30,66	38,08
Mediana	4,50	1,00	6,50	2,50	5,00	9,00	1,00	9,00
Desv. Pad.	46,29	41,80	50,24	48,65	55,79	48,09	47,02	51,56
Assimetria	1,79	1,66	1,69	1,72	1,76	1,63	1,89	1,69
Curtose	0,32	0,25	0,30	0,24	0,30	0,38	0,29	0,36

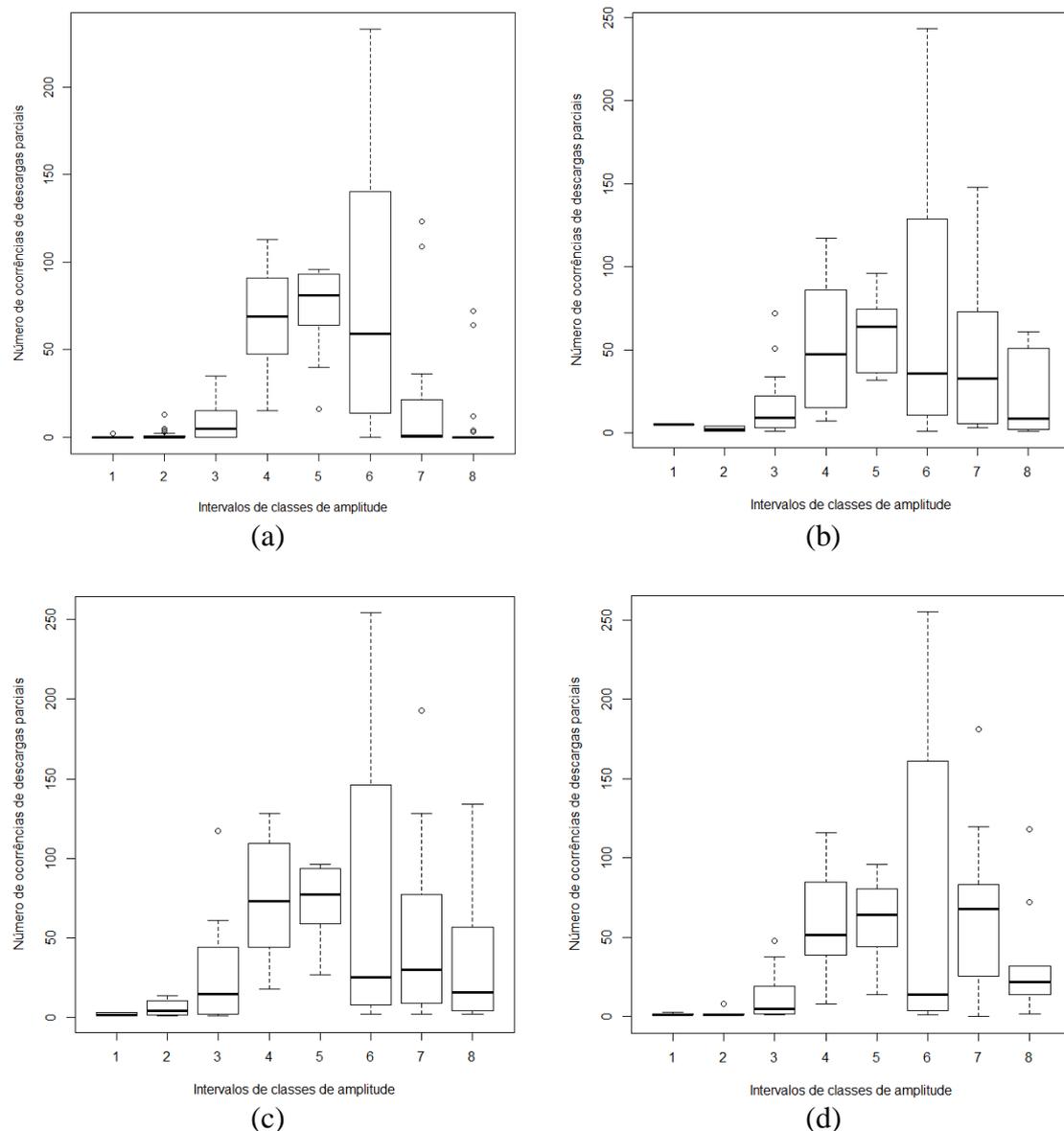
**Figura 4.7** – Histograma das ocorrências de descargas das tríades do lado direito no intervalo de amplitude de -300mV a 300mV



### 4.2.3. Boxplot das Classes de Amplitudes das Tríades do Hidrogerador no intervalo de amplitude de -300mV a 300mV

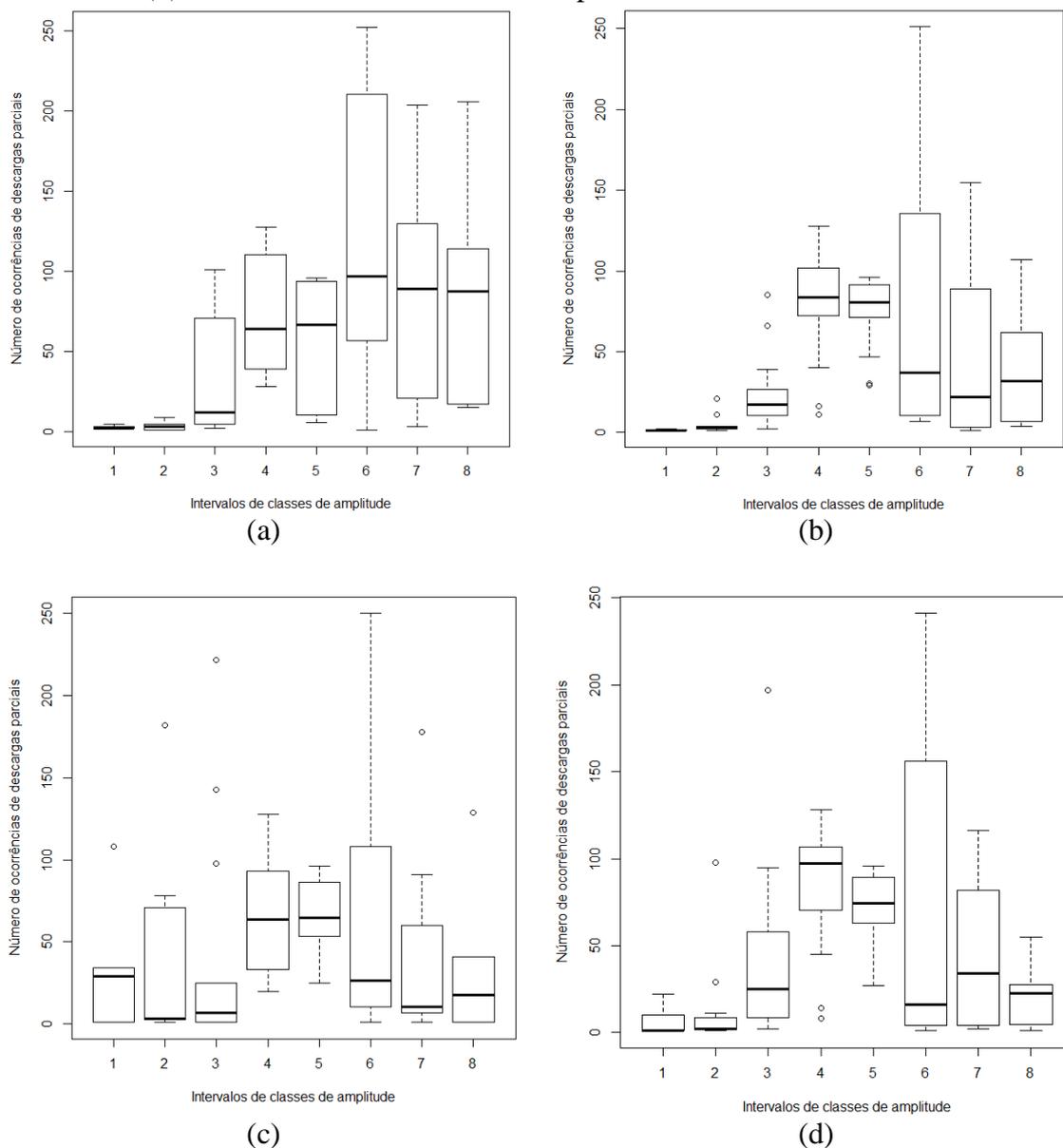
Pode-se observar nas Figuras 4.8 a semelhança de ocorrência de descargas parciais nos oito intervalos de classe de amplitude e que a maior massa de dados se concentrou nos intervalos 4, 5 e 6 que representaram os intervalos de amplitude -75mV a 1mV, 0mV a 74 mV e 75mV a 149mV, respectivamente. Já nos demais intervalos o número de ocorrências diminuiu consideravelmente exibindo menor valor nos intervalos de classes das extremidades. Foi observado também que o total de ocorrências de descargas parciais nos intervalos de classe de ângulos que não houve um comportamento padrão para todas as tríades.

**Figura 4.8** – Gráficos Boxplots. (a) tríade TR1E, (b) tríade TR2E, (c) tríade TR3E e (d) tríade TR4E no intervalo de amplitude de -300mV a 300mV.



Observa-se pelos gráficos mostrados na Figura 4.9 que para a maioria das tríades, a maior quantidade de massa de dados também se concentrou nos intervalos de classes próximos de zero (intervalos de classe 4, 5 e 6), porém com maior dispersão que para o lado esquerdo, como nas tríades TR2D, TR3D e TR4D, já a tríade TR1D apresenta uma quantidade considerável de dados nas classes 7 e 8 (150 a 224 mV e 225 a 299 mV).

**Figura 4.9** – Gráficos Boxplots. (a) tríade TR1D, (b) tríade TR2D, (c) tríade TR3D e (d) tríade TR4D no intervalo de amplitude de -300mV a 300mV.



### 4.3. Ajuste de modelos

Para o ajuste de modelos foi considerado a que existe uma estrutura de correlação das medições nos vinte e dois sensores, de modo que as tríades foram analisadas isoladamente, embora se tenha gerado um modelo englobando as ocorrências

de descargas parciais de todas as tríades, considerando o lado e as tríades como fatores. Neste modelo mais geral as estimativas das interações não foram significativas

#### 4.3.1. Ajuste pelo Modelo Poisson

Como a variável resposta de cada tríade é referente aos dados de contagem, pois diz respeito ao número de ocorrência de descargas parciais captadas por acopladores capacitivos conectados as fases A, B e V do hidrogerador. Logo, como visto na Seção 3.3 podemos supor inicialmente que a variável aleatória de cada tríade segue uma distribuição de Poisson. A princípio foram testados modelos considerando as interações entre os fatores, porém as estimativas de seus parâmetros não foram significativas, após outras especificações, foi obtido o modelo (4.2), onde *amp* refere-se às classes de amplitudes e *ang* à dos ângulos, com os intervalos de classes *amp1* e *ang1* considerados como referências.

$$\log(\mu) = \beta_0 + \beta_1 amp2 + \dots + \beta_7 amp8 + \beta_8 ang2 + \dots + \beta_{14} ang8 \quad (4.2)$$

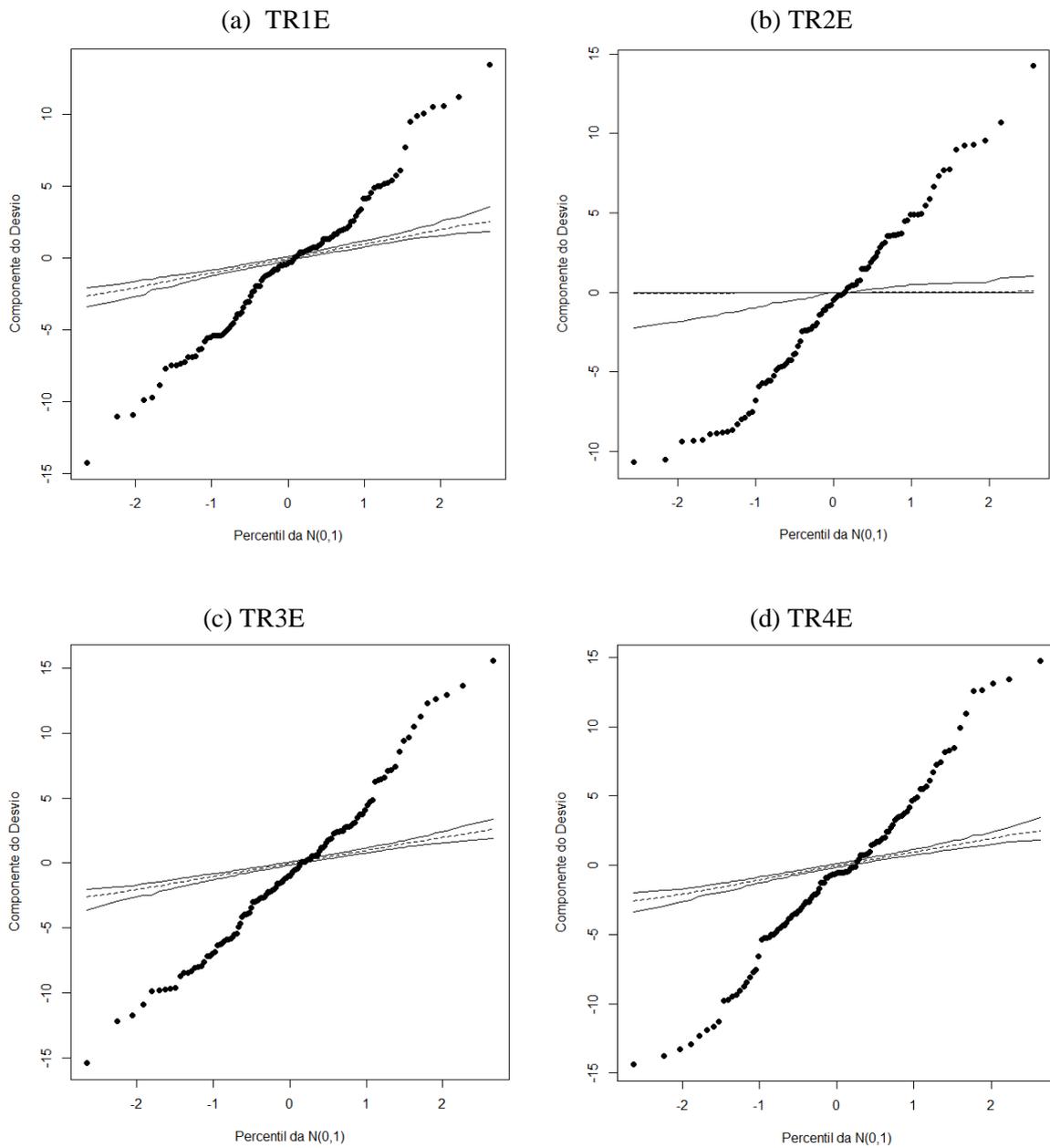
A *deviance* residual, seu grau de liberdade, a razão entre ambos, a medida de Akaike (AIC) e o valor de *p* para o teste do modelo estão dispostos na Tabela 4.3. É possível observar que a *deviance* residual para cada tríade apresenta valores altos quando comparados com o respectivo grau de liberdade, gerando assim valores bem maiores que dois para a razão indicadora de superdispersão.

**Tabela 4.4** – Valores da *deviance* residual, graus de liberdade, Razão, AIC e *p-value* das tríades para o modelo de Poisson.

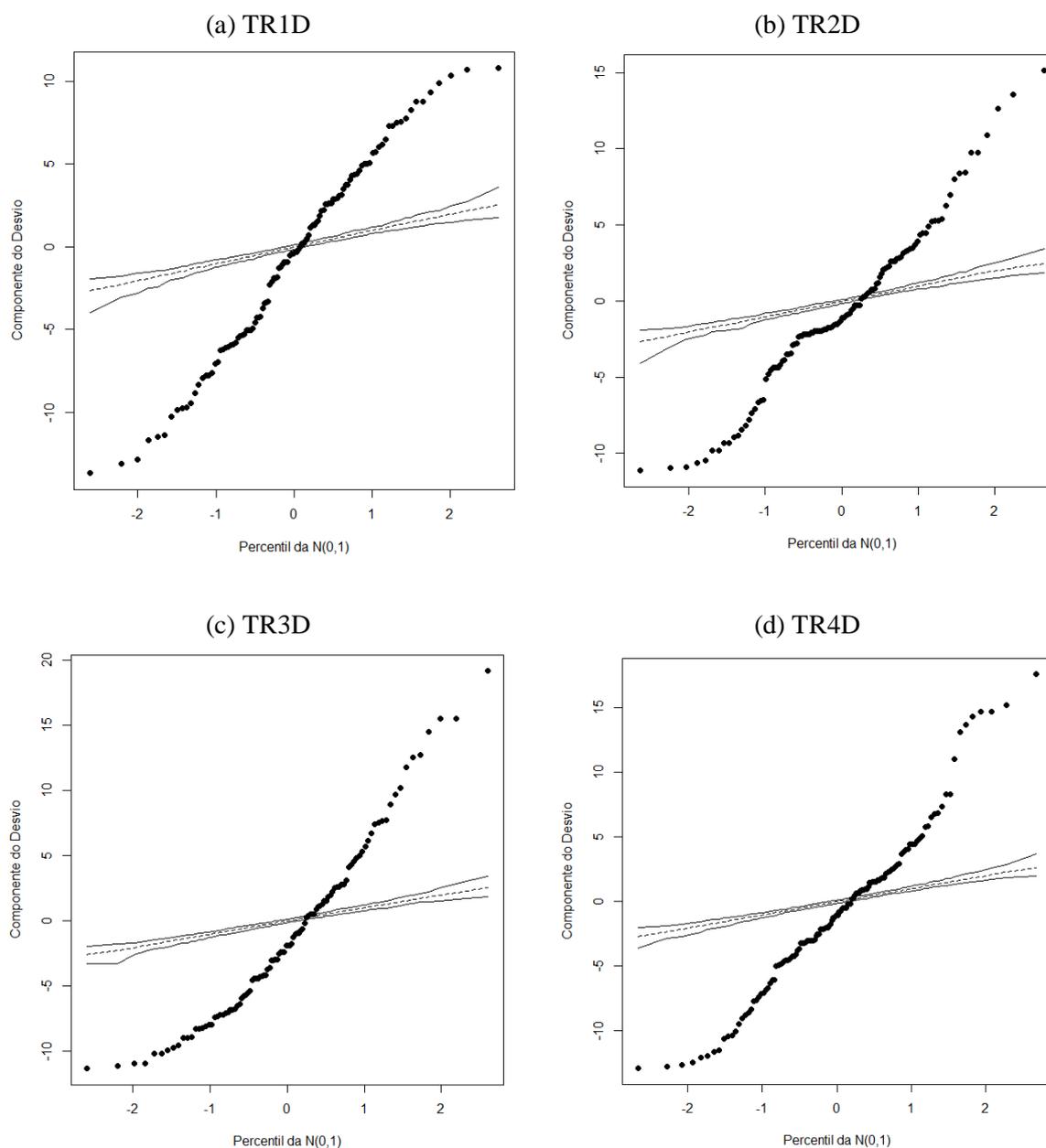
Tríades	Deviance Residual	G. Liberdade	Razão	AIC	<i>p-value</i>
<i>TR1E</i>	2593,60	105	24,70	3224,00	0,0000
<i>TR2E</i>	2425,10	82	29,57	2944,20	0,0000
<i>TR3E</i>	3848,40	110	34,99	4507,50	0,0000
<i>TR4E</i>	3770,80	103	36,61	4377,70	0,0000
<i>TR1D</i>	3482,30	97	35,90	4104,10	0,0000
<i>TR2D</i>	2968,00	106	28,00	3629,60	0,0000
<i>TR3D</i>	4154,70	91	45,66	4729,50	0,0000
<i>TR4D</i>	4662,70	117	39,85	5363,10	0,0000

Os gráficos normais de probabilidade de todas as tríades do hidrogerador, com seus respectivos envelopes simulados são mostrados nas Figuras 4.10 e 4.11, onde é possível observar que praticamente a totalidade dos resíduos está fora das bandas de confiança, sendo dispostos transversalmente a estes.

**Figura 4.10** – Gráfico normal de probabilidade dos resíduos do modelo de Poisson para as tríades do lado esquerdo.



**Figura 4.11** – Gráfico normal de probabilidade dos resíduos do modelo de Poisson para as tríades do lado direito.



### 4.3.2. Ajuste pelo Modelo Binomial Negativo

Como o modelo de Poisson não se ajustou aos dados, apresentando fortes indícios de superdispersão, com resíduos fora da banda de confiança, sendo ainda os modelos, pode-se recorrer então ao modelo Binomial Negativo, já que este consegue incorporar satisfatoriamente o fenômeno da superdispersão. Do mesmo modo, como ocorreu com o modelo de Poisson, foi observado que a interação entre as variáveis não foi significativa e apresentou resíduos não normais, de forma que se chegou ao mesmo modelo aditivo (4.2) proposto para o modelo de Poisson.

Na Tabela 4.4 são exibidos, para todas as tríades do hidrogerador, os valores da *deviance* residual, seu grau de liberdade, a razão entre ambas, bem como a medida de Akaike (AIC) e o valor de  $p$  para o teste do modelo. Nota-se assim, que a razão entre a *deviance* residual e o seu respectivo grau de liberdade é aproximadamente igual a 1, sugerindo que o modelo incorporou bem a superdispersão. Os valores de  $p$  também sugerem que existem evidências a um nível de significância entre e 0,01 e 0,05 que os modelos se ajustam razoavelmente aos conjuntos de dados.

**Tabela 4.5** – Valores da *deviance* residual, graus de liberdade, Razão, AIC e *p-value* das tríades para o modelo Binomial Negativo.

Tríades	Deviance Residual	G. Liberdade	Razão	AIC	<i>p-value</i>
TR1E	130,90	105	1,25	1117,50	0,0442
TR2E	105,55	82	1,29	917,92	0,0411
TR3E	136,12	110	1,24	1192,80	0,0462
TR4E	127,64	103	1,24	1102,40	0,0503
TR1D	121,31	97	1,25	1115,50	0,0480
TR2D	130,15	106	1,23	1162,40	0,0557
TR3D	121,30	91	1,32	1079,30	0,0186
TR4D	148,38	117	1,27	1301,80	0,0265

Foram obtidas as estimativas dos modelos para cada tríade do hidrogerador, assim como foram gerados os gráficos normais de probabilidade e os gráficos diagnósticos que avaliam os pontos considerados atípicos ocorrentes no modelo. Inicialmente serão exibidos os resultados obtidos para o lado esquerdo e posteriormente, os do lado direito.

#### 4.3.2.1. Lado esquerdo

A Tabela 4.5 apresenta as estimativas dos parâmetros para as quatro tríades do lado esquerdo. Vê-se que praticamente todas os intervalos de classe de amplitudes são significativas e positivas, quando comparadas com o intervalo de classe de referência *amp1* (-300 a 226mV), excetuando o *amp2* (-225 a -151mV) para todas as tríades e os *amp3* (-150 a -76 mV) e *amp8* (225 a 299mV) da tríade TR2E. Além disso, as estimativas dos parâmetros apresentam um comportamento semelhante, aumentando gradativamente seu valor até o *amp6* (75 a 149 mV) diminuindo posteriormente até a última classe. Este comportamento está relacionado ao número de ocorrências existente em cada intervalo de amplitude, pois aumentam em direção os intervalos centrais e diminuem em direção aos intervalos das extremidades.

Quanto aos ângulos, as estimativas dos parâmetros, na sua maioria, não são significativas quando comparados com a classe de 0-45°, tomada como referência, ocorrendo poucas exceções, sendo significativa apenas nas tríades TR2E , o intervalo

$ang3$  (90-135°) e o intervalo  $ang4$ (135-180°) e,  $ang3$  (90-135°) de TR3E, apresentando valores positivos.

**Tabela 4.6** – Estimativa dos parâmetros do modelo Binomial Negativo para os dados das tríades do lado esquerdo do hidrogerador.

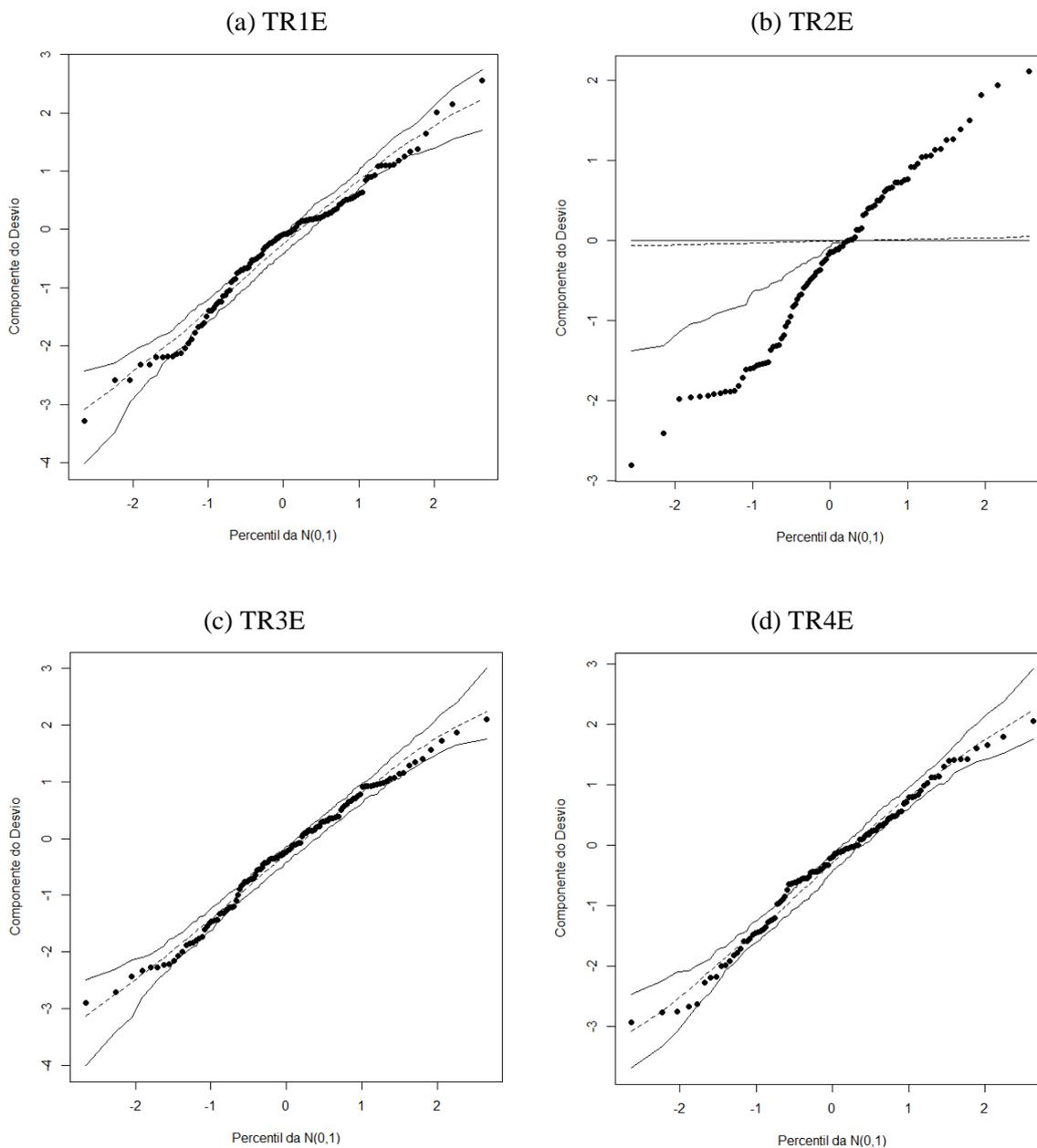
(a) TR1E.					(b) TR2E.				
Coef.	Est.	Er. Pad.	Z	Pr(> z )	Coef	Est.	Er. Pad.	Z	Pr(> z )
<i>Const</i>	0,89	0,63	1,40	0,1614	<i>const</i>	0,26	0,99	0,26	0,7936
<i>amp2</i>	0,49	0,70	0,70	0,4851	<i>amp2</i>	-0,22	1,04	-0,21	0,8296
<i>amp3</i>	1,96	0,65	3,00	<b>0,0027</b>	<i>amp3</i>	1,90	0,98	1,95	0,0513
<i>amp4</i>	3,60	0,64	5,61	<b>0,0000</b>	<i>amp4</i>	3,14	0,96	3,26	<b>0,0011</b>
<i>amp5</i>	3,68	0,64	5,73	<b>0,0000</b>	<i>amp5</i>	3,33	0,96	3,45	<b>0,0005</b>
<i>amp6</i>	3,77	0,64	5,88	<b>0,0000</b>	<i>amp6</i>	3,38	0,97	3,48	<b>0,0005</b>
<i>amp7</i>	2,45	0,65	3,75	<b>0,0002</b>	<i>amp7</i>	2,78	0,99	2,80	<b>0,0051</b>
<i>amp8</i>	2,59	0,72	3,59	<b>0,0003</b>	<i>amp8</i>	1,88	1,01	1,86	0,0629
<i>ang2</i>	-0,59	0,29	-2,07	0,0380	<i>ang2</i>	0,56	0,37	1,51	0,1316
<i>ang3</i>	-0,16	0,31	-0,54	0,5917	<i>ang3</i>	1,35	0,36	3,71	<b>0,0002</b>
<i>ang4</i>	-0,21	0,30	-0,70	0,4855	<i>ang4</i>	0,87	0,36	2,37	<b>0,0178</b>
<i>ang5</i>	-0,45	0,30	-1,52	0,1295	<i>ang5</i>	0,15	0,36	0,42	0,6767
<i>ang6</i>	-0,13	0,28	-0,45	0,6494	<i>ang6</i>	0,51	0,37	1,40	0,1619
<i>ang7</i>	-0,50	0,33	-1,53	0,1259	<i>ang7</i>	0,63	0,39	1,61	0,1077
<i>ang8</i>	-0,07	0,28	0,24	0,8120	<i>ang8</i>	0,23	0,37	0,53	0,5328

(c) TR3E.					(d) TR4E.				
Coef.	Est.	Er. Pad.	Z	Pr(> z )	Coef.	Est.	Er. Pad.	Z	Pr(> z )
<i>Const</i>	0,20	0,53	0,38	0,7010	<i>const</i>	0,14	0,67	0,21	0,8360
<i>amp2</i>	1,23	0,56	2,21	<b>0,0270</b>	<i>amp2</i>	0,31	0,72	0,42	0,6714
<i>amp3</i>	2,79	0,52	5,34	<b>0,0000</b>	<i>amp3</i>	2,23	0,65	3,41	<b>0,0006</b>
<i>amp4</i>	3,80	0,51	7,46	<b>0,0000</b>	<i>amp4</i>	3,76	0,64	5,86	<b>0,0000</b>
<i>amp5</i>	3,82	0,51	7,51	<b>0,0000</b>	<i>amp5</i>	3,79	0,64	5,90	<b>0,0000</b>
<i>amp6</i>	3,82	0,51	7,45	<b>0,0000</b>	<i>amp6</i>	4,03	0,64	6,25	<b>0,0000</b>
<i>amp7</i>	3,37	0,55	6,15	<b>0,0000</b>	<i>amp7</i>	3,90	0,68	5,76	<b>0,0000</b>
<i>amp8</i>	2,86	0,57	4,97	<b>0,0000</b>	<i>amp8</i>	3,15	0,69	4,60	<b>0,0000</b>
<i>ang2</i>	0,10	0,33	0,31	0,7597	<i>ang2</i>	0,11	0,34	0,31	0,7572
<i>ang3</i>	0,81	0,33	2,44	<b>0,0146</b>	<i>ang3</i>	0,36	0,35	1,05	0,2950
<i>ang4</i>	0,63	0,35	1,80	0,0717	<i>ang4</i>	0,27	0,36	0,76	0,4484
<i>ang5</i>	0,14	0,33	0,41	0,6801	<i>ang5</i>	0,24	0,36	0,67	0,5044
<i>ang6</i>	0,23	0,33	0,70	0,4852	<i>ang6</i>	0,34	0,35	0,99	0,3213
<i>ang7</i>	0,41	0,34	1,22	0,2234	<i>ang7</i>	-0,10	0,38	-0,26	0,7969
<i>ang8</i>	0,12	0,34	0,35	0,7224	<i>ang8</i>	0,26	0,35	0,75	0,4558

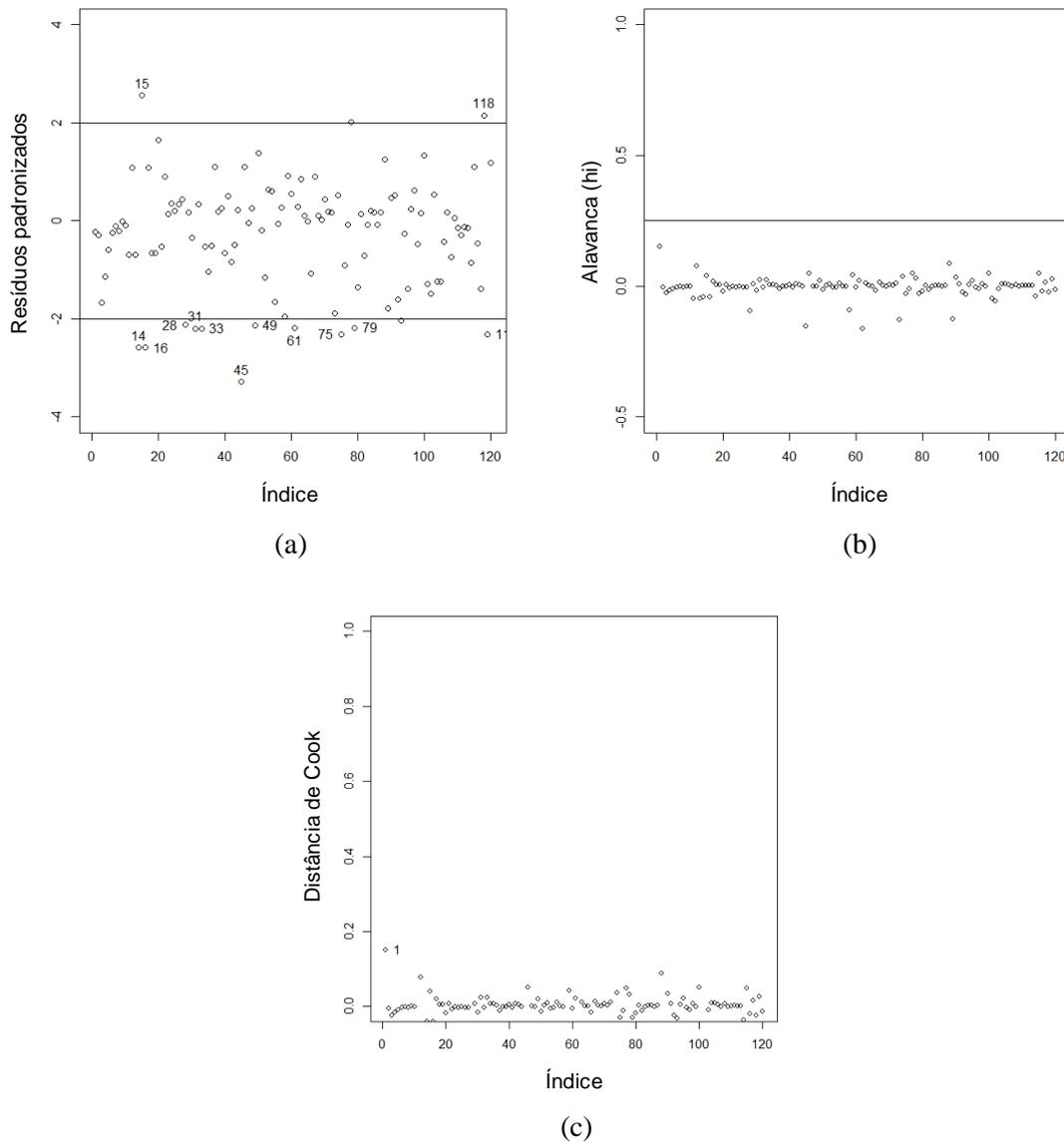
A Figura 4.12 mostra que para as tríades do lado esquerdo, excetuando a tríade TR2E, o gráfico normal de probabilidade apresenta a maioria dos resíduos dentro da banda de confiança ou sobre os limites desta, com poucos pontos fora do envelope.

**Figura 4.12** – Gráfico normal de probabilidade dos resíduos do modelo Binomial Negativo para os dados para as tríades do lado esquerdo.



A seguir a Figuras 4.13 mostra alguns gráficos para se analisar a presença de pontos atípicos. No gráfico (a) aparecem as ocorrências #14, #15, #16, #28, #31, #33, #45, #49, #61, #75, #79, #118 e #119 como possíveis pontos aberrantes. Já no gráfico (b) não há ocorrências apontadas como alavancas. No gráfico (c) o ponto #1 é indicado como influente. Essas observações pertencem às fases A (#14, #31, #49, #75, #79, todos com 1 ocorrência e #28 com 6 ocorrências), B (#1 – 2 ocorrências, #15 – 123 ocorrências, #45 – 1 ocorrência, #118 – 109 ocorrências) e V (#16 – 1 ocorrência, #33 – 1 ocorrência, #61 – 3 ocorrências, #119 – 2 ocorrências).

**Figura 4.13** – Gráficos para diagnóstico do modelo Binomial Negativo para a tríade TR1E. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.



Para se verificar o nível de influências desses pontos e se realmente provocam alterações consideráveis nas estimativas dos parâmetros, foram retirados os pontos considerados aberrantes e notou-se que não houve mudanças significativas nas estimativas dos parâmetros, permanecendo inalterados o desvio residual, graus de liberdade e AIC. Em adição a isto, o fato de não serem apontados pontos de alavanca, como mostra o gráfico (c), pode indicar que estes pontos não prejudicam a análise deste estudo.

Seguindo este processo foi verificado, para as demais tríades do lado esquerdo (ver APÊNDICE A), o nível de influências dos pontos considerados aberrantes pela exclusão destes e foi observado que não houve mudanças significativas nas estimativas

dos parâmetros. Além disso, para a tríade TR2E ocorreu erros na geração do gráfico normal de probabilidade não sendo possível gerá-lo. Outro fato ocorrido na tríade TR4E foi que, apesar da retirada dos pontos sugeridos como aberrantes não ter provocado uma mudança significativa nas estimativas dos parâmetros, os pontos residuais passaram a se concentrar dentro ou sobre os limites do envelope simulado.

#### 4.3.2.2. Lado direito

A Tabela 4.6 mostra que os parâmetros estimados para as classes de amplitude e ângulo da tríade TR3D não são significativos, quando comparadas com o intervalo de classe de referência. Quanto as demais tríades os parâmetros são significativos apenas para as amplitudes, excetuando-se a *amp2* de todas as tríades e *amp8* para a tríade TR4D. Observa-se ainda, que o valor dos parâmetros das amplitudes oscilam seu valores aumentando seu valor até a classe 4 (*amp4*), diminuindo na classe 5 (*amp5*), aumentando na classe 6 (*amp6*) e diminuindo a partir daí até a classe 8 (*amp8*). As estimativas dos ângulos não são significativas em nenhuma das tríades.

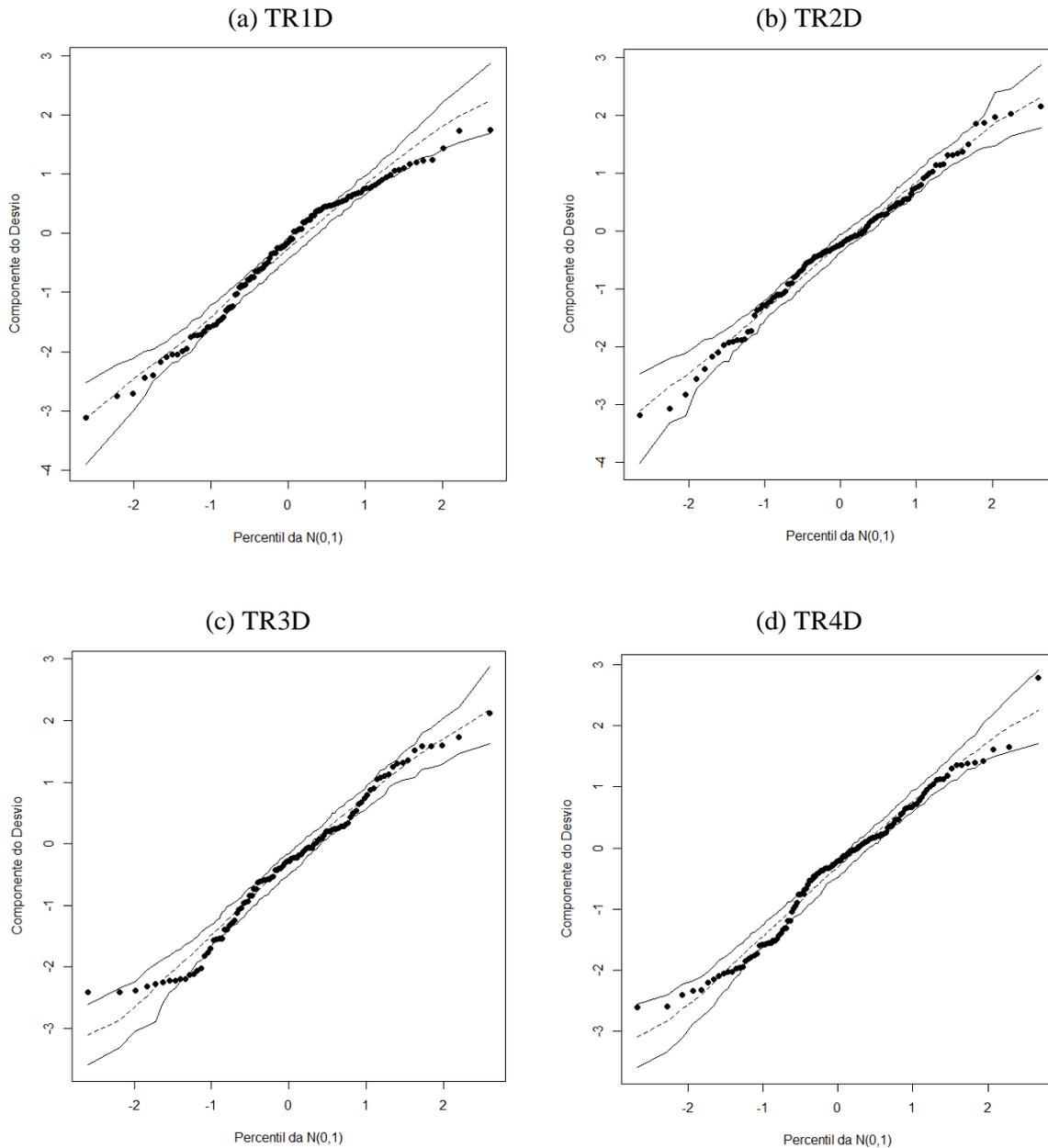
**Tabela 4.7** – Estimativa dos parâmetros do modelo Binomial Negativo para os dados das tríades do lado direito do hidrogerador.

(a) TR1D.					(b) TR2D.				
Coef.	Est.	Er. Pad.	Z	Pr(> z )	Coef	Est.	Er. Pad.	Z	Pr(> z )
<i>const</i>	0,95	0,58	1,64	0,1000	<i>const</i>	0,04	0,70	0,06	0,9551
<i>amp2</i>	0,29	0,63	0,47	0,6410	<i>amp2</i>	1,55	0,73	2,13	0,0333
<i>amp3</i>	2,74	0,57	4,82	<b>0,0000</b>	<i>amp3</i>	3,01	0,69	4,35	<b>0,0000</b>
<i>amp4</i>	3,46	0,56	6,21	<b>0,0000</b>	<i>amp4</i>	4,36	0,69	6,32	<b>0,0000</b>
<i>amp5</i>	3,24	0,56	5,80	<b>0,0000</b>	<i>amp5</i>	4,33	0,69	6,27	<b>0,0000</b>
<i>amp6</i>	3,91	0,57	6,83	<b>0,0000</b>	<i>amp6</i>	4,34	0,69	6,26	<b>0,0000</b>
<i>amp7</i>	3,51	0,58	6,00	<b>0,0000</b>	<i>amp7</i>	3,85	0,71	5,42	<b>0,0000</b>
<i>amp8</i>	3,45	0,60	5,73	<b>0,0000</b>	<i>amp8</i>	3,51	0,73	4,78	<b>0,0000</b>
<i>ang2</i>	-0,21	0,32	0,64	0,5220	<i>ang2</i>	-0,14	0,28	0,49	0,6258
<i>ang3</i>	-0,48	0,36	1,34	0,1790	<i>ang3</i>	0,01	0,29	0,06	0,9545
<i>ang4</i>	-0,27	0,37	0,74	0,4600	<i>ang4</i>	-0,38	0,30	1,29	0,1980
<i>ang5</i>	-0,02	0,31	0,07	0,9440	<i>ang5</i>	0,19	0,27	0,69	0,4915
<i>ang6</i>	0,20	0,31	0,65	0,5150	<i>ang6</i>	0,32	0,28	1,16	0,2463
<i>ang7</i>	-0,36	0,32	1,13	0,2570	<i>ang7</i>	-0,49	0,29	1,71	0,0868
<i>ang8</i>	0,07	0,31	0,24	0,8110	<i>ang8</i>	0,24	0,28	0,86	0,3890

(c) TR3D.					(d) TR4D.				
Coef.	Est.	Er. Pad.	Z	Pr(> z )	Coef.	Est.	Er. Pad.	Z	Pr(> z )
<i>const</i>	3,16	0,53	5,99	<b>0,0000</b>	<i>const</i>	1,83	0,52	3,51	<b>0,0004</b>
<i>amp2</i>	0,34	0,61	0,56	0,5736	<i>amp2</i>	0,64	0,56	1,15	0,2513
<i>amp3</i>	0,20	0,55	0,36	0,7203	<i>amp3</i>	1,93	0,53	3,66	<b>0,0002</b>
<i>amp4</i>	0,95	0,52	1,83	0,0675	<i>amp4</i>	2,82	0,52	5,39	<b>0,0000</b>
<i>amp5</i>	1,04	0,52	2,01	0,0442	<i>amp5</i>	2,64	0,52	5,06	<b>0,0000</b>
<i>amp6</i>	0,87	0,54	1,62	0,1051	<i>amp6</i>	2,71	0,52	5,16	<b>0,0000</b>
<i>amp7</i>	0,29	0,57	0,50	0,6181	<i>amp7</i>	2,16	0,56	3,90	<b>0,0000</b>
<i>amp8</i>	0,10	0,64	0,16	0,8769	<i>amp8</i>	1,22	0,56	2,18	0,0296
<i>ang2</i>	0,51	0,38	1,37	0,1708	<i>ang2</i>	0,30	0,32	0,93	0,3500
<i>ang3</i>	0,43	0,40	1,08	0,2806	<i>ang3</i>	-0,12	0,33	0,37	0,7081
<i>ang4</i>	-0,54	0,41	1,31	0,1902	<i>ang4</i>	-0,33	0,35	0,93	0,3533
<i>ang5</i>	-0,10	0,40	0,25	0,8047	<i>ang5</i>	-0,50	0,34	1,49	0,1369
<i>ang6</i>	0,44	0,39	1,14	0,2529	<i>ang6</i>	-0,30	0,33	0,90	0,3657
<i>ang7</i>	0,17	0,42	0,39	0,6936	<i>ang7</i>	-0,50	0,39	1,27	0,2026
<i>ang8</i>	-0,33	0,42	0,79	0,4287	<i>ang8</i>	-0,04	0,35	0,12	0,9056

A Figura 4.14 mostra que a maioria dos resíduos oscila entre os limites do envelope simulado. Entretanto, a tríade TR1D apresenta um ou dois pontos fora da banda de confiança, enquanto a tríade TR3D exibe a cauda inferior com alguns resíduos ultrapassando as margens do envelope.

**Figura 4.14** – Gráfico normal de probabilidade dos resíduos do modelo Binomial Negativo para as tríades do lado direito.

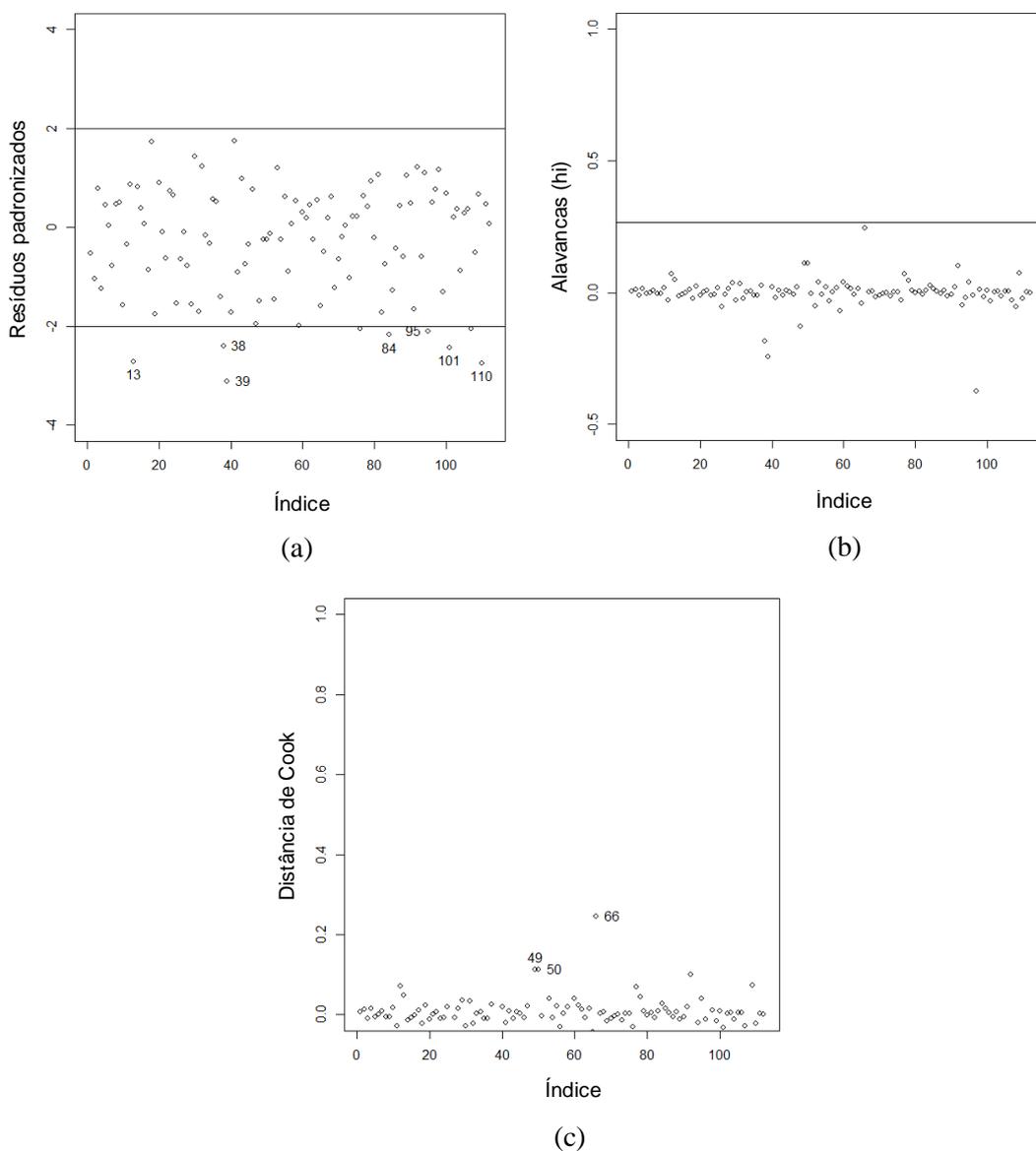


Seguindo o mesmo procedimento que foi realizado para o lado esquerdo do hidrogerador, foram retirados os pontos sugeridos como atípicos pelos gráficos de diagnóstico e testado o modelo para se verificar seu grau de influencia. Contatou-se que não houve mudanças nas estimativas que pudessem confirmar estes pontos como capazes de prejudicar a análise que se pretende fazer no presente trabalho. A tríade TR3D apresentou todos resíduos dentro ou sobre os limites do envelope simulado, mas a estimativa dos parâmetros do modelo continuaram sendo não significativa. A seguir

são exibidos os gráficos diagnósticos da tríade TR1D do lado direito do hidrogerador, sendo os demais exibidos no Apêndice A.

No diagnóstico gráfico exibido pela Figura 4.15, o gráfico (a) apresenta os pontos #13, #38, #39, #84, #95, #101 e #110 como aberrantes, (b) não indica pontos como possíveis alavancas e (c) aponta como prováveis pontos influentes as observações #49, #50 e #66, sendo relacionados às fases A (#13 – 3 ocorrências, #38 – 5 ocorrências, #49 – 1 ocorrência, #66 – 2 ocorrências, #110 – 3 ocorrências), B (#39 – 1 ocorrência, #50 – 2 ocorrências, #95 – 6 ocorrências) e V (#101 – 2 ocorrências).

**Figura 4.15** – Gráficos para diagnóstico do modelo Binomial Negativo para a tríade TR1D. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.



#### 4.4. Interpretação do Modelo Final

A interpretação dos parâmetros do modelo pode ser feita analisando o efeito que cada variável provoca no número de ocorrências de descargas parciais, podendo ser obtido pela expressão  $|1 - e^{\beta_j}| \times 100$ .

O modelo que melhor se ajustou aos dados em estudo foi o modelo Binomial Negativo a partir de (4.2). Será feita, para cada modelo, a interpretação dos seus respectivos parâmetros através do efeito que provocam na ocorrência de descargas parciais. Ressalta-se que a interpretação feita será sempre em relação aos parâmetros de referência, *amp1* para a variável amplitude e *ang1* para a variável ângulo.

Para a tríade TR1E considerando o intervalo *amp6*, por exemplo (Tabela 4.5 (a)), o número provável de ocorrências aumenta  $|1 - e^{3,77}| \times 100 = 4238,01\%$ . De outra forma, se for considerada o intervalo *amp3*, é provável que ocorra um aumento de  $|1 - e^{1,96}| \times 100 = 609,93\%$  no número esperado de ocorrências.

Apesar das estimativas dos parâmetros da tríade TR2E ter sido significativas, o gráfico normal de probabilidade apresentou um comportamento não formal, e persistindo o problema mesmo com a retirada dos pontos considerados aberrantes, optou-se aqui por não fazer referência à interpretação do modelo para esta tríade, e sugerir que se faça uma análise mais rigorosa desta tríade com a finalidade de levantar suas particularidades.

Já para a tríade TR3E, percebe-se inicialmente que todos os parâmetros também são positivos como na tríade anterior, logo todos os intervalos tendem a aumentar as ocorrências de descargas parciais, sendo o maior efeito causado pelos intervalos *amp5* e *amp6* (Tabela 4.5 (c)), aumentando o número de DPs em  $|1 - e^{3,82}| \times 100 = 4460,42\%$ . De outra forma, o intervalo *ang3* provoca um aumento esperado de  $|1 - e^{0,81}| \times 100 = 124,79\%$ , sendo este o menor efeito para este modelo.

Para a tríade TR4E, assim como na tríade TR1E, somente as amplitudes influenciam efetivamente nas ocorrências de descargas, onde o *amp3* (Tabela 4.5 (d)), provoca um aumento de  $|1 - e^{2,23}| \times 100 = 829,99\%$ , crescendo este percentual gradativamente até o *amp6* ( $|1 - e^{4,03}| \times 100 = 5526,09\%$ ) e diminuindo novamente em direção ao *amp8* ( $|1 - e^{3,15}| \times 100 = 2233,61\%$ ).

Quanto às tríades do lado direito a interpretação é análoga às da esquerda considerando as estimativas dos parâmetros nas tabelas 4.6(a) à 4.6(d). Assim, para a TR1D se observa que o *amp4* gera um aumento na probabilidade de ocorrência de

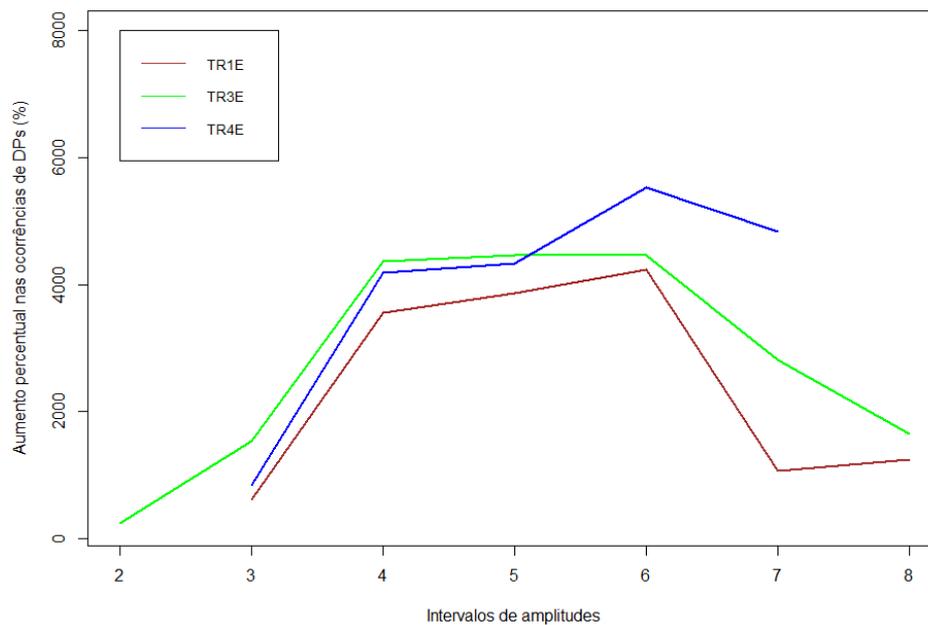
descargas de  $|1 - e^{3,46}| \times 100 = 3081,70\%$ . O mesmo acontece para os demais parâmetros positivos de acordo com o seu respectivo valor.

A mesma análise é feita para a tríade TR2D. Observa-se que os seus parâmetros são positivos, logo em seus intervalos tendem a aumentar a probabilidade de ocorrências de descargas parciais. Por exemplo, para o intervalo  $amp8$ , o aumento provável é de  $|1 - e^{3,51}| \times 100 = 3244,83\%$  em relação ao  $amp1$ .

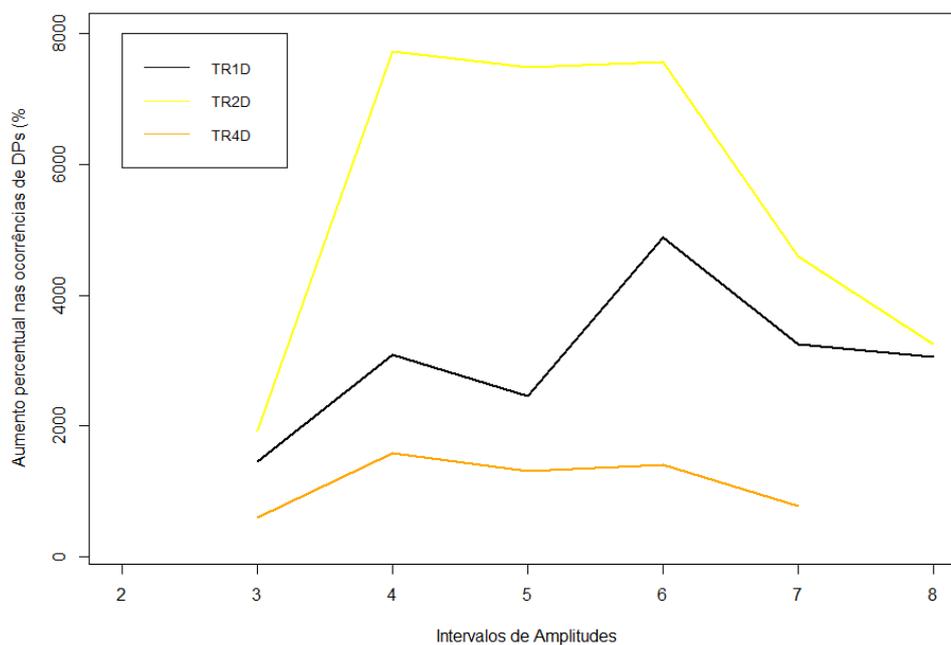
As estimativas dos parâmetros da tríade TR3D não foram significativas para nenhuma das variáveis explicativas consideradas. Por fim, para tríade TR4D, analogamente aos modelos anteriores, percebe-se que as amplitudes, por serem todas positivas, aumentam a ocorrências de DPs em maior ou menor probabilidade de acordo com seus respectivos valores.

É possível perceber então que em todos os modelos, o número de ocorrências de descargas parciais é influenciado pela amplitude que pelo ângulo, sendo os efeitos desse fator vistos nas Figura 4.16 e 4.17.

**Figura 4.16** – Aumento percentual nas ocorrências de descargas parciais considerando a amplitude de referência  $amp1$  para as tríades esquerdas.



**Figura 4.17** – Aumento percentual nas ocorrências de descargas parciais considerando a amplitude de referência *amp1* para as tríades direitas.



Desta forma, percebe-se que para as tríades do lado esquerdo o maior aumento de percentual de ocorrência está nos intervalos de amplitudes mais próximos de zero, ou seja, as classes 4,5 e 6 que representam, respectivamente, as faixas de -75mV a 1mV, 0mV a 74 mV e 75mV a 149mV, diminuindo à medida que estas se afastam deste valor. Quanto às tríades do lado direito, embora apresentem uma oscilação no intervalo de amplitude 5, no geral se comportam de forma similar as tríades esquerdas. Esta característica indica como se distribui as ocorrências de descargas parciais em relação à variável de maior efeito.

## Capítulo 5

### 5. Conclusão

As pesquisas referentes a descargas parciais têm sido alargadas nos últimos anos. Estas buscam uma maior compreensão deste fenômeno objetivando seu correto reconhecimento e identificação de suas fontes. Desta forma, neste trabalho, apresentou-se um método de organização e de análise de descargas parciais, permitindo extrair características consideráveis que contribuirão para uma melhor compreensão do comportamento das DP's em hidrogeradores.

A primeira etapa deste método foi a estruturação dos dados brutos, que foram agrupados a suas respectivas tríades e posteriormente organizados em tabelas, exibindo o número de ocorrências de descargas parciais por amplitude e ângulo de fase. A amplitude foi organizada em intervalos de 74 mV, enquanto o ângulo, em intervalos de 45°. Em seguida, foi realizada a análise exploratória dos dados que consistiu no estudo dos *kernels* de cada tríade, dos histogramas e dos *boxplots*. Esta análise permitiu constatar a possibilidade da utilização dos Modelos Lineares Generalizados bem como perceber o comportamento das descargas, ou seja, como elas se distribuem ao longo dos intervalos de amplitude e de ângulo de ocorrência.

Na etapa da aplicação do modelo inicialmente foi ajustado o modelo de Poisson para os dados, porém apresentou valores altos para a razão de dispersão, sendo este fato interpretado como havendo a presença de superdispersão nos dados. Isto foi confirmado pela disposição praticamente transversal dos resíduos em relação às bandas de confiança, estando a quase totalidade destes fora do envelope simulado, sugerindo que este modelo não se ajustou bem aos dados. Posteriormente se ajustou o modelo Binomial Negativo aos dados e foi verificado que este incorporou melhor a superdispersão para praticamente todas as tríades.

Assim, através do método aplicado aqui, foi observado que para os intervalos de amplitude que contiveram valores próximos de zero, o aumento percentual de ocorrência de descargas parciais é maior do que para aquelas que possuem valores bem maiores que zero. A interpretação deste comportamento sugere que isto se dá pelo fato da maioria das descargas parciais possuírem baixas amplitudes, concentrando assim os dados numa grande massa próxima deste valor. Considerando o ângulo, este não apresentou grande significância nos resultados do modelo, com exceção de alguns

intervalos de ângulo, pois as ocorrências de descargas estão bem distribuídas ao longo de todas as classes de ângulo, não apresentando um padrão semelhante em todas as tríades.

Pode-se dizer que o estudo de descargas parciais através de Modelos Lineares Generalizados para dados de contagem se mostrou bastante viável para o propósito deste trabalho, pois foi possível descrever satisfatoriamente os dados deste estudo, podendo ser estendido para análise de dados de descargas parciais de outros hidrogeradores, contribuindo para a formulação de um método de avaliação de DP's em hidrogeradores.

## 6. Referências Bibliográficas

- [1] BRASIL, F. S. *Estudo Experimental de Técnicas de Medição de Descargas Parciais nos Enrolamentos Estatóricos de Hidrogeradores*. Dissertação de Mestrado – Programa de Pós-Graduação em Engenharia Elétrica. UFPA. 2013.
- [2] CAMPBELL, S. R *et al.* *Characteristics of Partial Discharge Pulses from Operating Rotating Machines*. Conference Record of the IEEE International Symposium on Electrical Insulation. Pittsburgh-USA. 1994.
- [3] CORDEIRO, G. M.; DEMÉTRIO C. G. B. *Modelos Lineares Generalizados e Extensões*. Departamento de Estatística e Informática – UFRPE/Departamento de Ciências Exatas – USP. 2013.
- [4] CUENCA, W. M. H. *Caracterização dos Sinais de Descargas Parciais em Equipamentos de Alta Tensão a Partir de Modelos Experimentais*. Tese – Engenharia Elétrica – Universidade do Rio de Janeiro, COPPE. 2005.
- [5] CUENCA, W. M. H.; LEVY A. F. S. *Remoção do Ruído Gaussiano e Outliers para Medição de Descarga Parcial em Equipamentos de Alta Tensão Utilizando a Análise de Componentes Independentes*. Anais do IX Congresso Brasileiro de Redes Neurais/Inteligência computacional. Ouro Preto. 2009.
- [6] DINIZ, F. C. C. B. *Supressão de Ruído, Detecção e Classificação de Sinais de Descargas Parciais em Transformadores de Potência*. Tese – Engenharia Elétrica UFRJ. 2005.
- [7] FAIER, J. M. *Curvas Principais Aplicadas na Identificação de Descargas Parciais em Equipamentos de Potência*. Dissertação de Mestrado – Programa de Pós-Graduação de Engenharia da Universidade do Rio de Janeiro. 2006.
- [8] FILHO, V. S. *Imagens de Defeitos de Descarga Parcial Estimulada por Raio X Pulsados em Materiais Dielétricos Poliméricos*. Tese de Doutorado – Programa Interdisciplinar de Pós-Graduação em Engenharia da Universidade do Paraná. 2000.
- [9] FRUTH, B. A; GROSS D. W. *Partial discharge signal generation transmission and acquisition*. IEE Proc. Sci. Meas. Technol., Vol. 142, No. I, January 1995.
- [10] FUHR J. *Procedure for Identification and Localization of Dangerous PD Sources in Power Transformers*. IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 12 – No. 05. 2005.
- [11] GROSS, D. W.; FRUTH, B. A. *Characteristics of Phase Resolved Partial Discharge Pattern in Spherical Voids*. 1998.

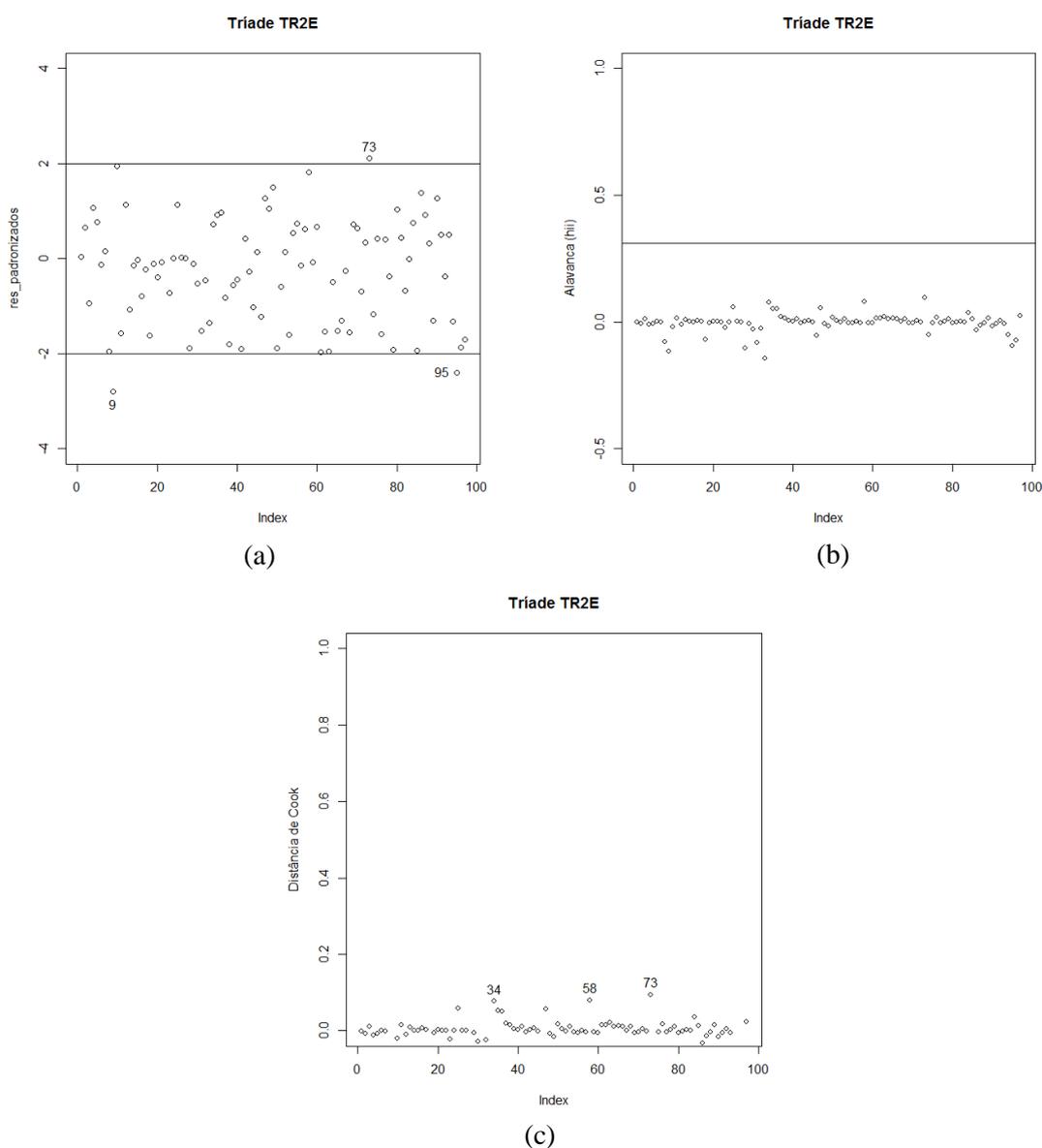
- [12] GUIDOUM, A. C. *Kernel Estimator and Bandwidth Selection for Density and its Derivatives*. Department of Probabilities & Statistics - Faculty of Mathematics - University of Science and Technology Houari Boumediene. 2015.
- [13] LÉVESQUE, M. et. al. *Insulation Degradation Analysis of Stator Bars Subjected to Slot Partial Discharges*. Electrical Insulation Conference. Canada. 2013.
- [14] IEC 60034-27. *Rotating electrical machines-Part 27: Off-line partial discharge measurements on the stator winding insulation of rotating electrical machines*. Ed.1. 2006.
- [15] IEEE P1434. *IEEE Draft Guide for the Measurement of Partial Discharges in AC Electric Machinery*. 2011.
- [16] KAWAKATSU W. M. *Descargas Parciais em Barras Estatóricas de Hidrogeradores: Modelagem por Guia de Onda e Análise de um Acoplador*. Dissertação. Engenharia Elétrica – UFPA. 2009.
- [17] LEAL, L. B. *Descargas Parciais em Barras/Bobinas do Circuito do Estator de Máquinas Síncronas*. Trabalho de Conclusão de Curso da Faculdade de Engenharia Elétrica – Universidade Federal de Itajubá. 2009.
- [18] LEVY A. F. S. et. all. *Técnicas Integradas para Localização e Identificação de Descargas Parciais em Subestações Blindadas Isoladas a Gás SF<sub>6</sub>*. CEPEL. 2001.
- [19] MAQRASHI, A. A. *Pattern Recognition of Partial Discharges using Matlab Tools*. Sultan Qaboos University, Oman. 2004.
- [20] MARCIANO F. W. P. *Principais tipos de resíduos utilizados na análise de diagnóstico em MLG com aplicações para os modelos: Poisson, ZIP e ZINB*. Projeto de Iniciação Científica. Departamento de Estatística e Matemática Aplicada – UFCE. 2009.
- [21] MAS'UD, A. A. et. al. *Application of an ensemble neural network for classifying partial discharge patterns*. Electric Power Systems Research. 2014.
- [22] MONTOYA A. G. M. *Inferência e diagnóstico em modelos para dados de contagem com excesso de zeros*. Dissertação de Mestrado. Instituto de Matemática, Estatística e Computação Científica – UNICAMP. 2009.
- [23] NAGAMINE C. M. L. *Modelos para Dados de Contagem: um estudo sobre o número de ovos do mosquito Aedes Aegypti*. Dissertação. Departamento de Pós-Graduação em Estatística – Universidade Federal de São Carlos – UFSCar. 2007.
- [24] OLIVEIRA S. *Inferência e análise de resíduos e de diagnóstico em modelos lineares generalizados*. Monografia. Graduação em Estatística – UFJF. 2013.

- [25] PAULA G. A. *Modelos de regressão com apoio computacional*. Instituto de Matemática e Estatística – USP. 2010.
- [26] RAMALHO J. J. S. *Modelos de regressão para dados de contagem*. Mestrado em Matemática Aplicada a Economia. Instituto Superior de Economia e Gestão – Universidade de Lisboa. 1996.
- [27] RODRIGUES C. *Distribuições em series de potências modificadas inflacionadas e distribuição Weibull binomial negativa*. Dissertação. Estatística e Experimentação Agronômica – USP. 2011.
- [28] SANTOS J. H. P. *Modelos para dados de contagem com excesso de zeros*. Dissertação de Mestrado em Estatística. Escola de Ciências – Universidade do Minho. 2013.
- [29] STONE, G. C. *et al. Recent Problems Experienced with Motor and Generator Windings*. Copyright Material IEEE. 2009.
- [30] SUWARNO; MIZUTANI T. *Diagnosis of Insulation Conditions: Interpretation of Partial Discharges from  $\phi$ -q-n pattern, Pulse-Sequence and Pulse Waveform*. International Conference on Condition Monitoring and Diagnosis. China. 2008.
- [31] TADANO, Y. S. *et. all. Método de regressão de Poisson: metodologia para avaliação do impacto da poluição atmosférica na saúde populacional*. Engenharia Mecânica e de Materiais – Universidade Tecnológica do Paraná - UTFPR. 2009.
- [32] TURKMAN, M. A. A, SILVA G. L. *Modelos Lineares Generalizados – da teoria à prática*. Universidade de Lisboa. 2000.
- [33] VENKATESH, S. *et. al. Orthogonal least square center selection technique – A robust scheme for multiple source Partial Discharge pattern recognition using Radial Basis Probabilistic Neural Network*. Expert Systems with Applications. 2011.
- [34] VIEIRA, G. A. B. *Estudo da Variabilidade de Utilização da Técnica de Ultra-Som Associada ao Processamento Digital de Sinais (PDS) para Detecção da Ocorrência de Pontos Quentes em Equipamentos Elétricos de Alta Tensão*. Dissertação de Mestrado. CEFET/RJ. 2008.
- [35] VOGELSANG, R. *et al. Performance testing of high voltage generator – and motor insulation systems*. 5th WSEAS/IASME International Conference on Electric Power Systems. 2005.
- [36] WARREM, V; STONE, G. C. *Advancements in Partial Discharge Analysis to Diagnose Stator Winding Problems*. Conference Record of the IEEE International Symposium on Electrical Insulation. Anaheim-USA. 2000.

## APÊNDICE A – Gráficos diagnósticos para o Modelo Binomial Negativo para as tríades de ambos os lados do hidrogerador.

Em A1, pra o modelo na tríade TR2E, os gráficos (a) e (c) sugerem, respectivamente, as ocorrências #9, #73 e #95 como pontos aberrantes e as ocorrências #34, #58 e #73 como influentes. O diagnóstico não aponta pontos de alavanca. Esses pontos têm os seguintes números de ocorrências: #9 (fase V – 1 ocorrência), #34 (fase V – 243 ocorrências), #58 (fase A – 137 ocorrências), #73 (fase A – 226 ocorrências) e #95 (fase B – 3 ocorrências).

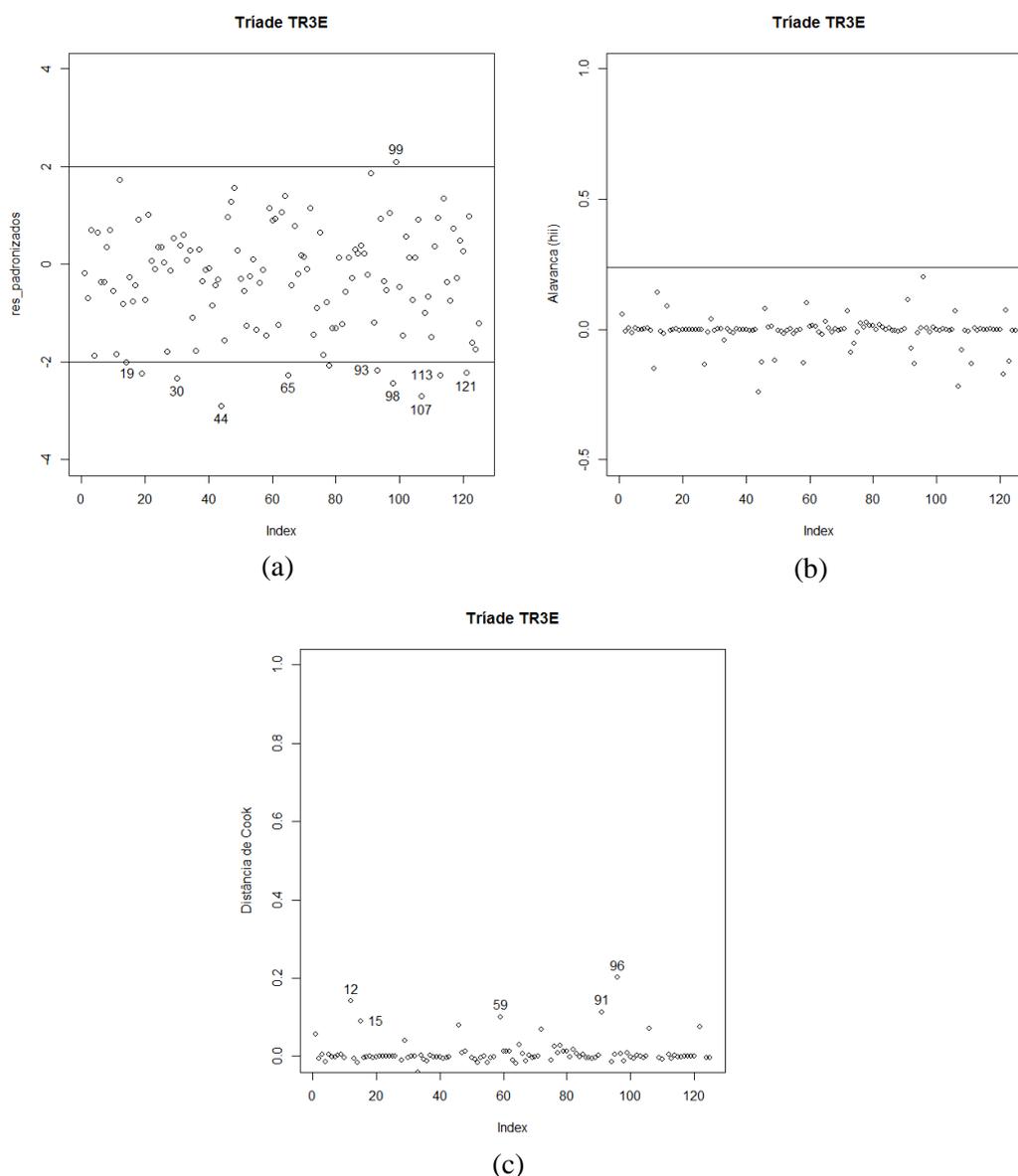
**A1** – Gráficos para diagnóstico do modelo Binomial Negativo para a tríade TR2E. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.



Percebe-se em A2 que para a tríade TR3E, os gráficos (a) e (c) sugerem as ocorrências #19, #30, #44, #65, #93, #98, #99, #107, #113 e #121 como pontos aberrantes, e #12, #15, #59, #91 e #96 como pontos influentes. Esses pontos estão

associados à fase A (#44 – 2 ocorrências, #91 – 241 ocorrências, #98 – 1 ocorrência, #121 – 4 ocorrências), à fase B (#12 – 177 ocorrências, #19 – 1 ocorrência, #30 – 2 ocorrências, #107 – 2 ocorrências) e à fase V (#15 – 1 ocorrência, #59 – 236 ocorrências, #65 – 1 ocorrência, # – 5 ocorrências, #96 – 1 ocorrência, #99 – 117 ocorrências).

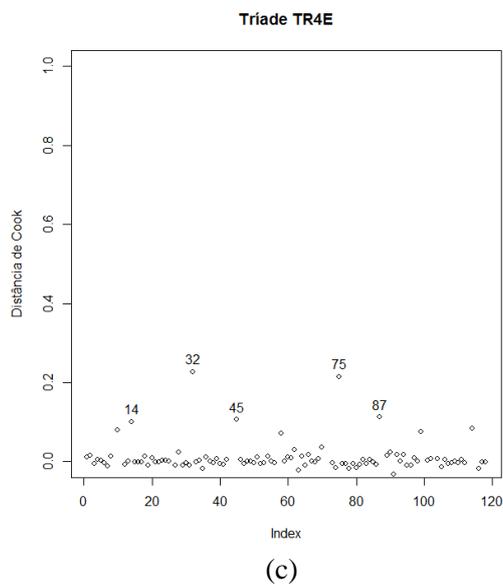
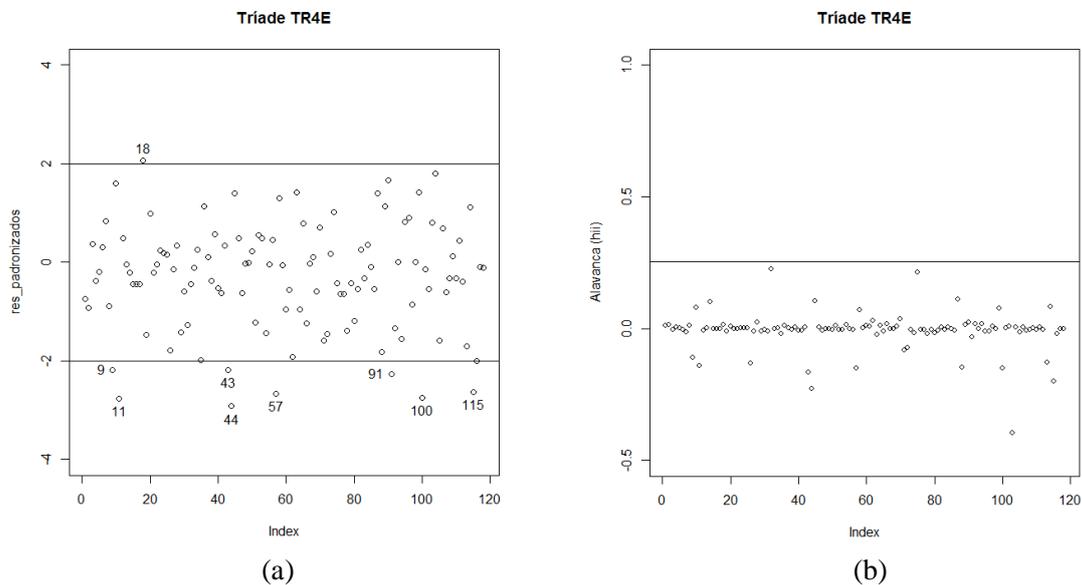
**A2** – Gráficos para diagnóstico do modelo Binomial Negativo para a tríade TR3E. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.



Verifica-se em A3 que para a tríade TR4E, o gráfico (b) não sugeriu pontos de alavanca, o gráfico (a) aponta como pontos aberrantes as ocorrências #9, #11, #18, #43, #44, #57, #91, #100 e #115 e o gráfico (c) indica os pontos #14, #32, #45, #75 e #87 como influentes, sendo os mesmos relacionados às fases A (#9 – 4 ocorrências, #18 – 48 ocorrências, #43 – 6 ocorrências, #75 – 1 ocorrência, #87 – 250 ocorrências), B (#14

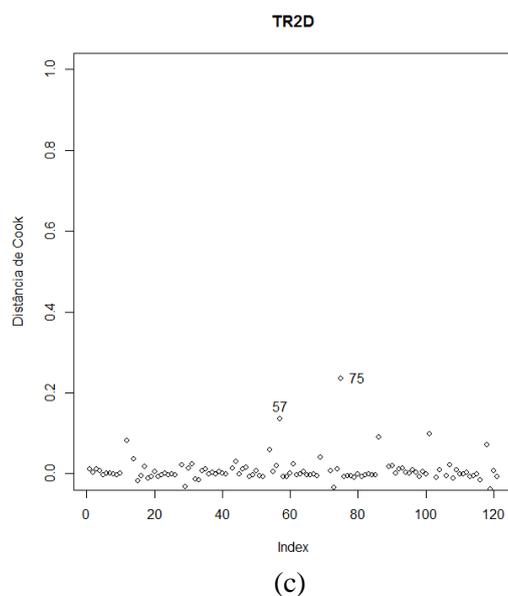
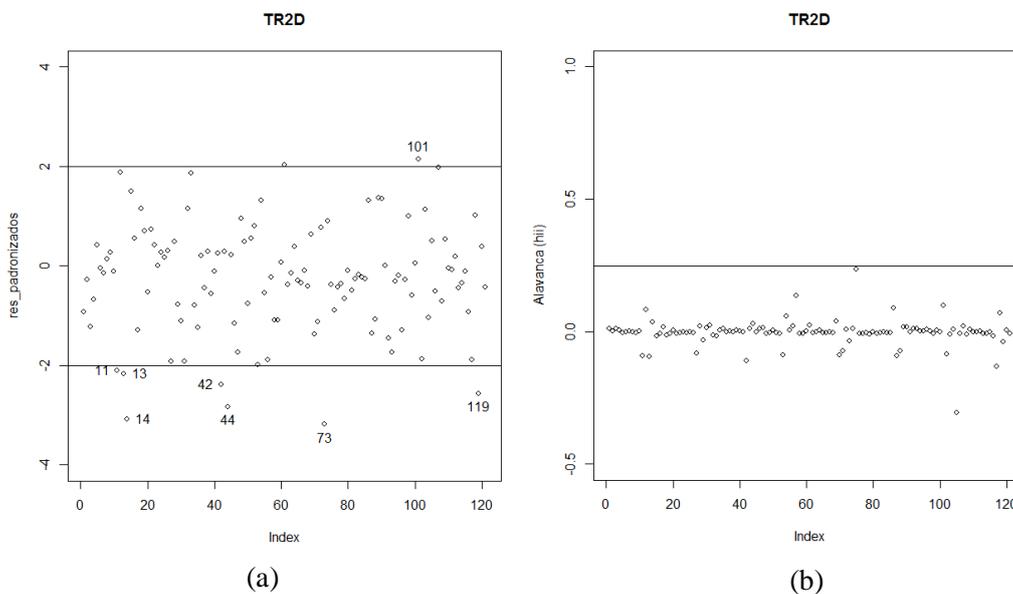
– 1 ocorrência, #32 – 1 ocorrência, #44 – 1 ocorrência, #57 – 2 ocorrências, #91 – 2 ocorrências, #100 – 1 ocorrência) e V (#11 – 1 ocorrência, #45 – 255 ocorrências).

**A3** – Gráficos para diagnóstico do modelo Binomial Negativo para a tríade TR4E. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.



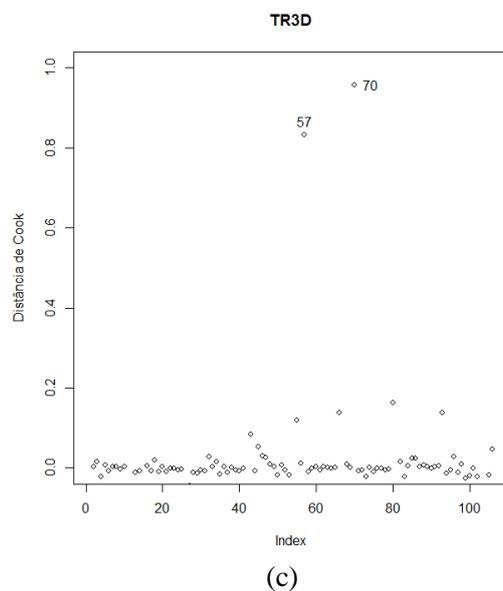
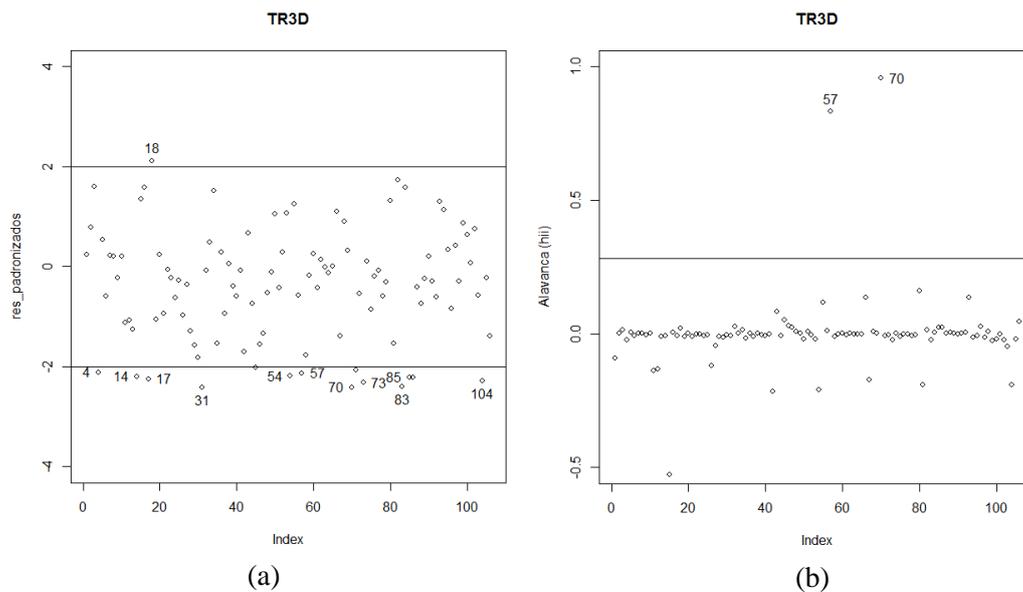
Os gráficos de A4 apresentam os seguintes diagnósticos para o modelo: (a) indica os pontos #11 (fase A – 10 ocorrências), #13 (fase V – 9 ocorrências), #14 (fase A – 1 ocorrência), #42 (fase B – 7 ocorrências), #44 (fase V – 2 ocorrências), #73 (fase V – 1 ocorrência), #101 (fase A – 168 ocorrências) e #119 (fase A – 4 ocorrências) como pontos aberrantes; (b) não apresenta pontos de alavanca e (c) aponta os pontos #57 (fase A – 1 ocorrência) e #75 (fase A – 1 ocorrência) como influentes.

**A4** – Gráficos para diagnóstico do modelo Binomial Negativo para a tríade TR2D. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.



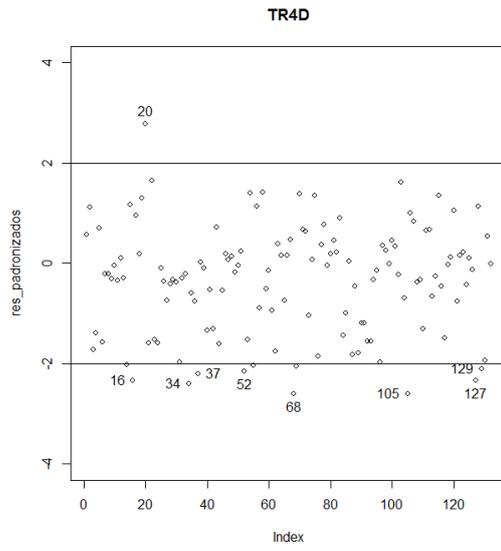
No diagnóstico da tríade TR3D, exibido em A5 tem-se em (a) os pontos #4 (fase V – 1 ocorrência), #14 (fase A – 1 ocorrência), #17 (fase V – 2 ocorrências), #18 (fase A – 222 ocorrências), #31 (fase V – 1 ocorrência), #54 (fase A – 1 ocorrência), #57 (fase A – 1 ocorrência), #70 (fase A – 1 ocorrência), #73 (fase V – 1 ocorrência), #83 (fase V – 1 ocorrência), #85 (fase A – 1 ocorrência) e #104 (fase B – 1 ocorrência) suspeitos de serem pontos aberrantes. Em (b) os pontos #57 e #70 são considerados como alavancas e em (c), estes dois são apontados como influentes.

**A5** – Gráficos para diagnóstico do modelo Binomial Negativo para a tríade TR3D. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.

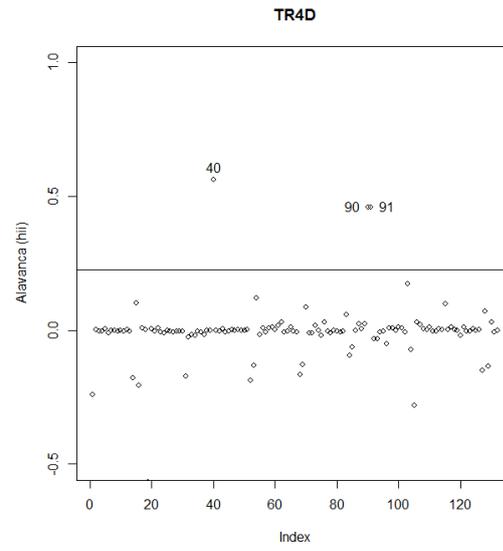


Por fim, em A6 mostra os gráficos diagnósticos da tríade TR4D e apresenta em (a) os índices de ocorrências #16 (fase V - 3 ocorrências), #20 (fase A - 98 ocorrências), #34 (fase A - 2 ocorrências), #37 (fase A - 1 ocorrência), #52 (fase A - 4 ocorrências), #68 (fase A - 1 ocorrência), #105 (fase V - 1 ocorrência), #127 (fase A - 3 ocorrências) e #129 (fase V - 5 ocorrências) como possíveis valores aberrantes. Em (b) verifica-se os índices #40 (fase V - 1 ocorrência), #90 (fase A - 1 ocorrência) e #91 (fase B - 1 ocorrência) como prováveis ocorrências com alavancagem, sendo que em (c) todos estes pontos são tidos como influentes.

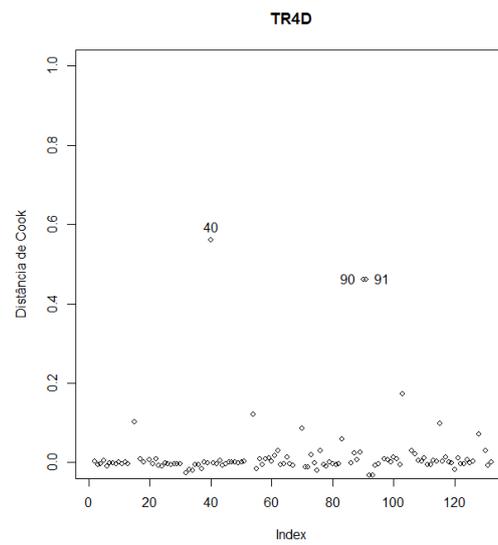
**A6** – Gráficos para diagnóstico do modelo Binomial Negativo para a tríade TR4D. (a) pontos aberrantes, (b) pontos de alavanca e (c) pontos influentes.



(a)



(b)



(c)