

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Programa de Graduação em Engenharia Elétrica

Trabalho de Conclusão de Curso

**Utilização de um sensor de campo 3D e métodos estatísticos  
para diagnóstico de isoladores**

Haslan J. G. Pedro

Rômulo R. Maranhão do Valle

Orientador

Campina Grande

2014, Novembro

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Programa de Graduação em Engenharia Elétrica

## **Utilização de um sensor de campo 3D e métodos estatísticos para diagnóstico de isoladores**

Trabalho de conclusão de curso submetido ao Programa de Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande como parte dos requisitos necessários para obtenção do grau de Engenheiro Eletricista.

Área de Concentração: Compatibilidade Eletromagnética

Orientador: Rômulo R. Maranhão do Valle

Campina Grande  
Haslan J. G. Pedro

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Programa de Graduação em Engenharia Elétrica

## **Utilização de um sensor de campo 3D e métodos estatísticos para diagnóstico de isoladores**

Trabalho de conclusão de curso submetido ao  
Programa de Graduação em Engenharia Elétrica  
da Universidade Federal de Campina Grande  
como parte dos requisitos necessários para ob-  
tenção do grau de Engenheiro Eletricista.

Trabalho aprovado. Campina Grande, 28 de Novembro de 2014:

---

**Rômulo R. Maranhão do Valle**  
Orientador

---

**Bruno Barbosa Albert**  
Convidado

Campina Grande  
2014, Novembro

*A minha avó Sebastiana Gonçalves, a quem dedico todo meu amor,  
gratidão e eterna saudade*

*“A quem muito foi dado, muito será exigido;  
e a quem muito foi confiado, muito mais será cobrado”.*  
*(Bíblia Sagrada, Lucas 12, 48)*

# Agradecimentos

Primeiramente, gostaria de agradecer aos meus pais, pois toda e qualquer vitória ou conquista na minha vida eu devo aos dois. Muito obrigado pelo apoio dado durante toda minha vida. Agradeço também em especial a minha avó Sebastiana, que não se encontra mais em vida e não pôde desfrutar dessa felicidade comigo.

Aos meus irmãos Jaqueline e Erbson, pelo apoio e incentivo durante os momentos difíceis.

Aos professores Glauco Fontgalland, Rômulo R. Maranhão do Valle e Bruno Barbosa Albert por toda a ajuda oferecida nessa fase de graduação, sanando minhas dúvidas, sejam pessoais ou profissionais e por me fazer despertar para a área de Telecomunicações. Em especial ao professor Rômulo, orientador deste TCC.

A Damião Fernandes pela amizade durante a graduação, uma amizade que se propaga além da sala de aula, a qual levo pra vida toda. À Pedro Venske pela ajuda com o MatLab.

A Pedro Igor, meu querido sobrinho, pelo amor e carinho.

# Resumo

O desempenho de quatro isoladores de linhas de transmissão submetidos a 12 kV foram avaliados considerando a poluição química no qual o isolador foi submetido. Um sistema de processamento e aquisição de dados foi montado, a fim de permitir a classificação do nível de contaminação, identificar a melhor faixa de frequência que contenha a assinatura espectral dos isoladores e investigar qual polarização é mais apropriada para as medições, utilizando uma abordagem estatística. Quando submetido a elevadas tensões, os isoladores irradiam sinais de rádio frequência devido ao efeito corona. Métodos estatísticos como histogramas, teste de normalidade e o coeficiente de correlação foram utilizados para classificar o nível de poluição em isoladores de vidro. Mostra-se que, em determinadas situações os histogramas podem ser usados para distinguir o estado de conservação dos isoladores. Ao utilizar o coeficiente de correlação em conjunto com o teste de normalidade, é possível identificar faixas de frequência de interesse, ou seja, faixas onde a assinatura espectral dos isoladores se encontra. Os testes foram limitados a uma banda de 300 MHz (na de 30 MHz até 330 MHz), pois utilizar bandas maiores aumenta demasiadamente o tempo necessário para aquisição de dados. Foram utilizados dois tipos de antenas: bicônica e isotrópica (sensor 3D). Durante as medições com a antena bicônica, dois tipos de polarização foram empregados, a polarização vertical e horizontal. A escolha de uma antena (bicônica) com um dos tipos de polarização não considera sinais presentes nas demais polarizações, ocasionando perda de informação relevante do sinal irradiado pelos isoladores. Para investigar esse problema uma antena com características especiais foi empregada. Trata-se do sensor 3D (antena isotrópica) que captura o sinal irradiado pelos isoladores nas polarizações segundo os três eixos, fornecendo um único vetor campo elétrico que resume a polarização total incidente. Medições realizadas com o sensor 3D foram confrontadas com as medições realizadas com a antena linear, com o objetivo de verificar qual polarização leva a um melhor casamento com a polarização do sinal irradiado pelos isoladores.

**Palavras-chaves:** Coeficiente de correlação, Alta tensão, Isoladores, Histogramas, Teste de normalidade.

# Abstract

The insulating performance of four transmission lines subjected to 12 kV were evaluated considering the chemical pollution of the insulator which has undergone. A system of data acquisition and processing was mounted to enable the classification of the level of contamination, to identify the best frequency band containing the spectral signature of the insulators and investigate polarization which is more suitable for measurements using a statistical approach. When subjected to high voltages, the insulator radiate radio frequency signals due to the corona effect. Statistical methods such as histograms, test normality and the correlation coefficient were used to rate the level of pollution in glass insulators. It is shown that in certain situations the histograms can be used to distinguish the condition of the insulators. By using the correlation coefficient in conjunction with the normality test can be identified frequency bands of interest, or bands where the spectral signature is the insulators. Tests were limited to a band of 300 MHz (30 MHz to 330 MHz), for use higher bands excessively increases the time required for data acquisition. Two types of antennas were used: biconical and isotropic (3D sensor). During measurements with the biconical antenna, two types of polarization were used, the vertical and horizontal polarization. The choice of antenna (biconical) with one type of polarization does not consider the signals present in the other polarization, causing loss of relevant information of the signal radiated by insulators. To investigate this problem an antenna with special features was used. This is the 3D sensor (isotropic antenna) that captures the radiated signal by insulators in polarizations based on the three, providing a single electric field vector that summarizes the overall bias incident. Measurements with 3D sensor were compared with measurements made with the linear antenna, in order to find what polarization leads to a better match with the polarization of the radiated signal by insulators.

**Key-words:** Correlation coefficient, High Voltage, Insulators, Histograms, Normality test.

# Lista de ilustrações

Figura 1 – Mapeamento de um evento em números reais [29] . . . . .	19
Figura 2 – Correlação linear positiva [29] . . . . .	24
Figura 3 – Correlação não linear [29] . . . . .	25
Figura 4 – Curva Normal com $\mu$ constante [5] . . . . .	28
Figura 5 – Curva Normal com $\sigma$ constante [5] . . . . .	29
Figura 6 – Histograma com um pico centrado próximo a média [12] . . . . .	32
Figura 7 – Histograma com dois picos [12] . . . . .	33
Figura 8 – Histograma com forma de platô [12] . . . . .	33
Figura 9 – Diagrama 3D do irradiador isotrópico ISo [30] . . . . .	34
Figura 10 – Diagrama no plano horizontal, distribuição espacial isotrópica [30] . . . . .	35
Figura 11 – Diagrama no plano vertical, distribuição espacial preferencial [30] . . . . .	36
Figura 12 – Exemplo de diagrama de irradiação para antenas diretivas [30] . . . . .	36
Figura 13 – Onda EM com polarização linear Vertical [42] . . . . .	37
Figura 14 – Ondas EM com polarização linear Vertical e Horizontal respectivamente [42]	38
Figura 15 – Onda EM com polarização oblíqua [42] . . . . .	38
Figura 16 – Onda EM com polarização oblíqua [42] . . . . .	39
Figura 17 – Onda EM com polarização circular [42] . . . . .	39
Figura 18 – Relação entre os scripts utilizados neste trabalho . . . . .	45
Figura 19 – Caixa de diálogo para seleção de arquivos do isolador limpo . . . . .	46
Figura 20 – Caixa de diálogo para seleção de arquivos do isolador sujo . . . . .	47
Figura 21 – Matriz de coeficientes de correlação . . . . .	48
Figura 22 – Histograma do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização horizontal . . . . .	51
Figura 23 – Histograma do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utilizando uma antena bicônica com polarização horizontal . . . . .	51
Figura 24 – Histograma do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utilizando uma antena bicônica com polarização horizontal . . . . .	52
Figura 25 – Espectro do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização horizontal . . . . .	52
Figura 26 – Espectro do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utili- zando uma antena bicônica com polarização horizontal . . . . .	53
Figura 27 – Espectro do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utili- zando uma antena bicônica com polarização horizontal . . . . .	53
Figura 28 – Coeficiente de correlação do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização horizontal . . . . .	54

Figura 29 – Coeficiente de correlação do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utilizando uma antena bicônica com polarização horizontal . . .	54
Figura 30 – Coeficiente de correlação do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utilizando uma antena bicônica com polarização horizontal . . .	55
Figura 31 – Grau de correlação do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização horizontal . . . . .	56
Figura 32 – Grau de correlação do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utilizando uma antena bicônica com polarização horizontal . . . . .	56
Figura 33 – Grau de correlação do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utilizando uma antena bicônica com polarização horizontal . . . . .	57
Figura 34 – Grau de correlação do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização vertical . . . . .	58
Figura 35 – Grau de correlação do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utilizando uma antena bicônica com polarização vertical . . . . .	59
Figura 36 – Grau de correlação do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utilizando uma antena bicônica com polarização vertical . . . . .	59
Figura 37 – Coeficiente de correlação entre os dados dos isoladores poluídos (Conjunto III e V) na faixa de 30 MHz a 330 MHz . . . . .	61
Figura 38 – Coeficiente de correlação entre os dados dos isoladores poluídos (Conjunto IV e V) na faixa de 30 MHz a 330 MHz . . . . .	62
Figura 39 – Faixas onde há sinal irradiado pelos isoladores nas polarizações horizontal e vertical na faixa de 30 MHz a 330 MHz . . . . .	63

# Lista de tabelas

Tabela 1	– Valores Críticos associados ao grau de confiança na amostra . . . . .	21
Tabela 2	– Média e variância do espectro de cada isolador . . . . .	50
Tabela 3	– Faixas de interesse para o Conjunto I . . . . .	55
Tabela 4	– Faixas de interesse para o Conjunto II . . . . .	60
Tabela 5	– Sub-faixas onde o sinal é captado apenas na polarização vertical ou horizontal, na faixa de 30 MHz até 130 MHz . . . . .	60
Tabela 6	– Sub-faixas onde o sinal é captado apenas na polarização vertical ou horizontal, na faixa de 130 MHz até 230 MHz . . . . .	60
Tabela 7	– Sub-faixas onde o sinal é captado apenas na polarização vertical ou horizontal, na faixa de 230 MHz até 330 MHz . . . . .	60

# Lista de abreviaturas e siglas

AE	Análise Exploratória
EM	Eletromagnético
f.p.	Função de Probabilidade
f.d.p.	Função Densidade de Probabilidade
ISo	Irradiador Isotrópico
LAT	Laboratória de Alta Tensão
LT	Linhas de transmissão
MATLAB	Matrix Laboratory
RF	Rádio Frequência
SW	Shapiro-Wilk
v.a.	Variável Aleatória
UFMG	Universidade Federal de Campina Grande

# Lista de símbolos

$COV$	Covariância
$\rho$	Coefficiente de Correlação de Pearson
$\sigma$	Desvio-Padrão
$\sigma_X$	Desvio-Padrão da Variável Aleatória X
$\sigma_Y$	Desvio-Padrão da Variável Aleatória Y
$\hat{\sigma}$	Desvio-padrão estimado
$D$	Diretividade
$\Omega$	Espaço amostral
$f(\mathbf{x})$	Função densidade de probabilidade
$G$	Ganho
$U$	Intensidade de radiação
$V$	Matriz covariância
$\mu$	Média
$\bar{x}$	Média Aritmética
$\mu_X$	Média da Variável Aleatória X
$\mu_Y$	Média da Variável Aleatória Y
$\hat{\mu}$	Média estimada
$N$	Número de Amostras
$p$	Número de Variáveis
$W$	Parâmetro de decisão do critério de Shapiro-Wilk
$P_{irr}$	Potência irradiada
$P$	Probabilidade
$Z_\alpha$	Valor Crítico

<b>E</b>	Valor Esperado
$\sigma^2$	Variância
<b>X</b>	Variável Aleatória X
<b>Y</b>	Variável Aleatória Y
<b>Z</b>	Variável Reduzida
$m'$	Vetor de valores esperados
$y'$	Vetor de observações
$l'$	Vetor unitário

# Sumário

	Introdução . . . . .	16
<b>I</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>17</b>
1	FUNDAMENTAÇÃO TEÓRICA . . . . .	18
1.1	Variável qualitativa e quantitativa . . . . .	18
1.2	Variável Aleatória . . . . .	18
1.3	Determinação do tamanho de uma amostra . . . . .	20
1.4	Alguns parâmetros estatísticos . . . . .	21
1.4.1	Valor esperado . . . . .	21
1.4.2	Desvio-padrão . . . . .	22
1.4.3	Variância . . . . .	22
1.4.4	Covariância . . . . .	23
1.4.5	Coeficiente de correlação . . . . .	24
1.4.5.1	Coeficiente de correlação de Pearson . . . . .	25
1.4.6	Algumas suposições para a utilização do coeficiente de correlação de Pearson	26
1.4.7	Fatores que influenciam o coeficiente de correlação . . . . .	27
1.5	Distribuição Normal . . . . .	27
1.6	Teste de normalidade . . . . .	29
1.7	Histogramas . . . . .	31
1.8	Considerações sobre antenas . . . . .	34
1.8.1	Antena Isotrópica . . . . .	34
1.8.2	Antena Omnidirecional . . . . .	35
1.8.3	Antena Direcional . . . . .	35
1.8.4	Polarização de antenas . . . . .	36
1.8.5	Intensidade de Radiação . . . . .	40
1.8.6	Diretividade . . . . .	40
1.8.7	Ganho . . . . .	41
<b>II</b>	<b>TRATAMENTO DOS DADOS</b>	<b>42</b>
2	TRATAMENTO DOS DADOS . . . . .	43
2.1	Informações gerais sobre a origem dos dados . . . . .	43
2.2	Detalhes do processamento dos dados . . . . .	45
2.2.1	Script fase_1_2.m . . . . .	45

2.2.2	Script importa_dados.m . . . . .	46
2.2.3	Script normalidade_e_correlacao . . . . .	46
2.2.4	Script sw_test.m . . . . .	48
2.2.5	Script coef_diag.m . . . . .	48
<b>III</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>49</b>
<b>3</b>	<b>RESULTADOS E DISCUSSÕES . . . . .</b>	<b>50</b>
3.1	Análise do Conjunto I . . . . .	50
3.2	Análise do Conjunto II . . . . .	58
3.3	Análise dos Conjuntos III, IV e V . . . . .	61
	Conclusões . . . . .	64
	Referências . . . . .	66
	<b>ANEXOS</b>	<b>69</b>
	ANEXO A – FASE_1_2.M . . . . .	70
	ANEXO B – SCRIPT IMPORTA_DADOS.M . . . . .	73
	ANEXO C – SCRIPT NORMALIDADE_E_CORRELACAO.M . . . . .	74
	ANEXO D – SCRIPT SW_TEST.M . . . . .	79
	ANEXO E – SCRIPT COEF_DIAG.M . . . . .	85

# Introdução

A rede elétrica de distribuição de energia enfrenta grandes problemas, dentre eles destacam-se as interrupções de energia causada por defeitos nos isoladores das linhas de transmissão (LT). Durante sua vida de operação, o isolador fica exposto à poluição ambiental e intempéries durante muitos anos. Disto decorre o acúmulo de sujeira, levando à degradação de sua isolação e, portanto, comprometendo seu papel de isolar as LT.

Isoladores submetidos a alta tensão emitem radiação eletromagnética. O fenômeno pode ser interpretado, de maneira simplificada, como resultado da ionização das moléculas do ar no entorno da superfície dos condutores, devido ao elevado campo elétrico presente nessa região, e consequente ressonância em frequências específicas. Este efeito produz som, luz, pulsos eletromagnéticos e em menor escala aquecimento. Os sinais irradiados pelos isoladores submetidos à alta tensão cobrem um amplo espectro de frequência. Para cada geometria e estado físico do isolador têm-se um espectro em particular de frequência de emissão (assinatura do isolador).

A identificação da largura de banda que melhor representa as respostas espectrais dos isoladores é um dos principais focos deste trabalho. Isto é útil para diferenciar isoladores limpos de isoladores poluídos e também na escolha da antena para captação do sinal, além de reduzir o número de medições e o tempo de operação das mesmas.

Neste trabalho, tiramos vantagem de parâmetros estatísticos, tais como coeficiente de correlação e o teste de normalidade para fazer a análise dos dados medidos. Estes dados, foram obtidos por meio de medições realizadas no Laboratório de Alta Tensão - LAT, da Universidade Federal de Campina Grande, utilizando um analisador de espectro em conjunto com o sensor 3D que é capaz de capturar o sinal irradiado pelos isoladores nas polarizações segundo os três eixos, fornecendo um único vetor campo elétrico que resume a polarização total incidente no sensor.

O presente trabalho está dividido em três partes, a primeira parte é referente à fundamentação teórica, onde os principais conceitos são definidos. A segunda parte refere-se à exposição de detalhes do processamento de dados, do algoritmo utilizado e suas relações. Por fim, a terceira parte faz-se uma discussão dos principais resultados bem como das conclusões da análise.

Parte I

Fundamentação Teórica

# 1 Fundamentação teórica

## 1.1 Variável qualitativa e quantitativa

Em pesquisas que envolvem fenômenos aleatórios, a definição de variáveis é importante já que, através delas pode-se aferir questões de interesse. Assim, é possível entender que a variável é uma primeira forma de tornar prática a construção teórica. Uma variável pode se apresentar das seguintes formas:

1. Nominal: é aquela que permite o agrupamento da unidade de observação de acordo com uma classificação qualitativa em categorias definidas, ou seja, consiste simplesmente em nomear ou rotular. Ao se trabalhar com essa escala, cada unidade de observação deve ser classificada em somente uma categoria.
2. Ordinal: permite o agrupamento da unidade de observação de acordo com uma ordem de classificação. Fornece informações sobre a ordenação das categorias, mas não indica a grandeza das diferenças entre os valores.
3. Intervalar: ocorre quando as unidades de observação, além de estarem numa ordem de classificação, possibilitam quantificar as diferenças entre elas.

De acordo com o nível de mensuração, a variável pode ser classificada em qualitativa e quantitativa. Variável qualitativa é aquela cujo nível de mensuração é nominal ou ordinal, enquanto a quantitativa é aquela em que o nível de mensuração é intervalar [22].

A variável quantitativa pode ser discreta ou contínua, sendo a primeira resultante de contagem, assumindo somente valores inteiros. Por outro lado, a última resulta de medições, assumindo qualquer valor no campo dos números reais.

Outra diferença entre os dois tipos de variáveis está na interpretação de seus resultados. A variável discreta assume exatamente o valor a ela atribuído. Já a variável contínua assume um valor aproximado, por não existirem instrumentos de medida capazes de medir com precisão absoluta, ou não haver interesse em se determinar um valor contínuo com tanta precisão, considerando todas as casas decimais.

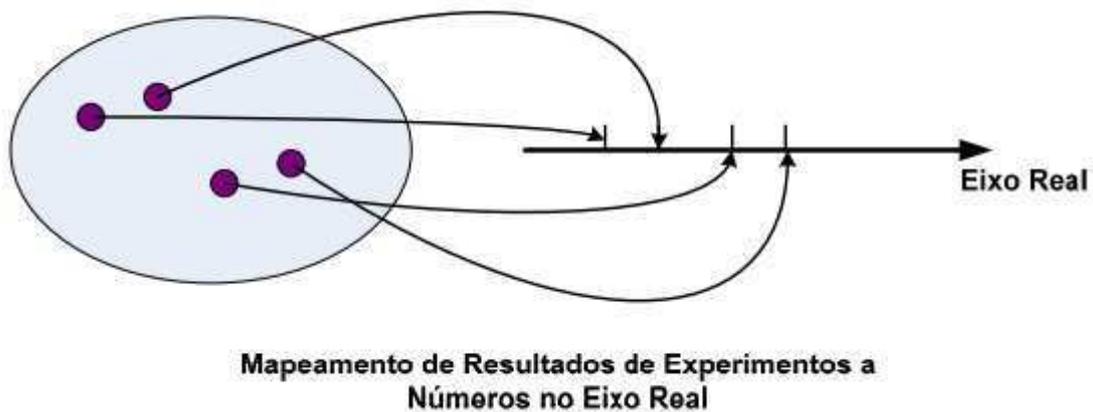
## 1.2 Variável Aleatória

O conceito de variável aleatória está intimamente ligado às condições do experimento. Se um determinado experimento é realizado repetidamente, com todas as condições precisamente mantidas, e se nessas condições obtém-se resultados idênticos, as variáveis avaliadas são ditas

determinísticas. Entretanto, se os resultados numéricos variam, as variáveis avaliadas são ditas aleatórias [18]. Uma variável aleatória (v.a.) é uma função que atribui um valor numérico a cada resultado individual de uma experiência aleatória, em outras palavras, é uma função que associa elementos do espaço amostral a valores numéricos, ou seja:

$$X : \Omega \rightarrow \mathbb{R} \quad (1.1)$$

Dessa forma, pode-se escrever os resultados de um experimento aleatório através de números, ao invés de palavras ou símbolos, possibilitando um tratamento matemático facilitado. Em outras palavras, a variável aleatória mapeia o resultado do experimento em números reais (Figura 1).



**Figura 1** – Mapeamento de um evento em números reais [29]

O espaço amostral ( $\Omega$ ) pode ser considerado como um conjunto de todos os resultados possíveis do experimento, que pode ser finito ou infinito. Nem sempre é fácil definir um espaço amostral de um experimento.

A importância do espaço amostral provém, sobretudo de ser um meio para a definição de eventos, que representa um subconjunto do espaço amostral. Em geral, há muito mais interesse em eventos do que no espaço amostral, devido ao evento representar uma amostra reduzida do espaço amostral, levando a uma considerável redução no tempo de realização do experimento.

As variáveis aleatórias auxiliam no cálculo das probabilidades. Uma medida de probabilidade é associada ao espaço amostral por meio de uma variável aleatória; a essa associação se dá o nome de função de probabilidades (f.p.), no caso de uma v.a. discreta, e uma função densidade de probabilidade (f.d.p.), para uma v.a. contínua.

### 1.3 Determinação do tamanho de uma amostra

Ao se trabalhar com um grande conjunto de dados no qual se deseja extrair informações, torna-se um desafio trabalhar com todos os elementos do conjunto, devido as limitações de tempo e muitas vezes de recursos. Geralmente, é estudado um pequeno grupo de elementos retirados do conjunto, que pode ser rotulado como uma amostra de um conjunto, buscando-se generalizar conclusões referentes ao conjunto inicial. De fato, a amostra de um conjunto não representa perfeitamente uma população, ou seja, a utilização de uma amostra implica na aceitação de uma margem de erro denominado de erro amostral, que é a diferença entre o resultado amostral e o resultado referente ao conjunto no qual foi amostrado.

Apesar de não ser possível evitar um erro amostral, pode-se limitar seu valor através da escolha de uma amostra de tamanho adequado. O erro amostral e o tamanho da amostra apresentam relação inversa, ou seja, quanto maior o tamanho da amostra, menor o erro cometido e vice-versa.

A determinação do tamanho de uma amostra é um problema de relevante importância, porque:

- Amostras desnecessariamente grandes acarretam desperdício de tempo e dinheiro
- Amostras muito pequenas podem levar a resultados não confiáveis.

Em muitos casos é possível determinar o tamanho mínimo de uma amostra para estimar um parâmetro estatístico, como por exemplo, a média ( $\mu$ ).

A fórmula para o cálculo do tamanho da amostra para uma análise confiável da média ( $\mu$ ) é dada por:

$$n = \left( \frac{Z_{\alpha} \sigma}{D} \right)^2 \quad (1.2)$$

Onde  $n$  é o número de indivíduos da amostra,  $Z_{\alpha}$  é o valor crítico que corresponde ao grau de confiança desejado,  $\sigma$  é o desvio-padrão e  $D$  é a diferença entre a média amostral e a média do conjunto amostrado.

Os valores correspondentes aos graus de confiança mais utilizado e os valores de  $Z$  correspondentes podem ser mostrados na tabela 1:

Grau de confiança ( $1 - \alpha$ )	$\alpha$	Valor Crítico $Z_\alpha$
90%	0,10	1,645
95%	0,05	1,96
99%	0,01	2,575

**Tabela 1** – Valores Críticos associados ao grau de confiança na amostra

## 1.4 Alguns parâmetros estatísticos

Parâmetro é uma medida que descreve de forma reduzida uma característica, representada pela variável, da população de amostras. Geralmente, o parâmetro normalmente é desconhecido, e deseja-se estimá-lo através de dados amostrais. Quando se fala em população, faz-se referência a um conjunto de elementos distintos que apresentam pelo menos uma característica em comum, aos quais os resultados de estudo deverão ser inferidos.

Na inferência estatística, parâmetros como valor esperado ( $E(x)$ ), desvio-padrão ( $\sigma$ ), variância ( $\sigma^2$ ), covariância ( $COV$ ) e o coeficiente de correlação ( $\rho$ ), são muito utilizados para caracterizar uma população de amostras.

### 1.4.1 Valor esperado

Em estatística, o valor esperado é a média ponderada de todos os valores possíveis da variável aleatória. Em outras palavras, cada valor possível que a v.a. pode assumir é multiplicada pelo seu peso, e os produtos resultantes são somados para encontrar o valor esperado. Os pesos usados no cálculo do valor esperado são as probabilidades no caso de uma v.a. discreta, ou função densidade de probabilidade (f.d.p.), no caso de uma v.a. contínua [10, 26, 28, 31 e 40].

O valor esperado de uma v.a. discreta  $X$ , é definida por:

$$E(x) = \sum_{i=1}^{\infty} x_i P_x(X = x_i) \quad (1.3)$$

Onde  $x_i$  é o valor do  $i$ -ésimo dado, e  $P_x(X = x_i)$  é a probabilidade da variável aleatória  $X$  ser igual a  $x_i$ .

Já para uma v.a. contínua, o valor esperado é dado por:

$$E(x) = \int_{-\infty}^{+\infty} x f(x) dx \quad (1.4)$$

Onde  $x$  é o valor da variável no intervalo considerado e  $f(x)$  é a função densidade de probabilidade.

Se os eventos forem de igual probabilidade o valor esperado é a média aritmética, que pode ser obtida dividindo-se a soma das observações pelo número delas. Se tivermos uma série

de  $n$  valores de uma variável  $x$ , a média aritmética será determinada pela expressão:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n nx_i \quad (1.5)$$

Tanto a média como o valor esperado indicam o valor em torno do qual há um equilíbrio da distribuição dos dados, sendo calculado apenas para variáveis quantitativas (discretas e contínuas). Uma das limitações de se usar a média é que esta é muito sensível a valores extremos.

#### 1.4.2 Desvio-padrão

O Desvio-padrão ( $\sigma$ ) é um parâmetro amplamente usado na estatística para medir a variabilidade dos dados em relação a média [3]. A dispersão das observações que constituem uma amostra pode ser caracterizada pelos desvios de cada observação em relação à média ( $x_i - \bar{x}$ ), podendo tomar valores positivos ou negativos, e o somatório dos desvios de cada observação em relação à média é igual a zero. Um desvio-padrão baixo indica que o conjunto de dados tende a ser muito próximo da média (valor esperado), enquanto que um alto desvio-padrão indica que os dados estão em uma ampla gama de valores.

A magnitude do desvio-padrão depende da dispersão das observações relativamente à média, não variando com o aumento do tamanho das amostras e possuindo a mesma unidade que a média. O desvio-padrão é dado por:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^2} \quad (1.6)$$

Onde  $\sigma$  é o desvio-padrão,  $\bar{x}$  é a média e  $n$  é número de elementos no conjunto.

O desvio padrão, como medida de dispersão, não deve ser usado quando a população não segue uma distribuição normal ou aproximadamente normal. Nestes casos, o desvio padrão pode não ser uma boa estimativa de dispersão, pelo fato da média ser pouco resistente a observações extremas. Nessas situações, poderá ser mais adequada a descrição da dispersão com outras medidas para melhorar a confiança das conclusões estatísticas.

#### 1.4.3 Variância

Outro método para cálculo do desvio de um grupo de amostras, é a utilização da variância. Ao contrário do desvio padrão, o qual usa o valor absoluto do desvio, a variância ( $\sigma^2$ ) toma o quadrado dos desvios a fim de eliminar os valores negativos. O quadrado dos desvios garante que a medida de variação será nula somente se todos os valores forem iguais [41], tornando-se uma medida adequada para medir a variabilidade das amostras. No entanto, existem duas limitações com a variância. Primeiro, porque os desvios das amostras em relação à média são elevadas ao

quadrado, o que dá mais peso as amostras com valores que estão muito longe da média, isso pode levar ao descarte do grupo de amostras. A variância ( $\sigma^2$ ) é dado por:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x - \bar{x})^2 \quad (1.7)$$

Onde um elevado valor de variância indica que as amostras estão dispersas; um menor valor de variância indica que os dados estão estreitamente distribuídos em torno da média.

#### 1.4.4 Covariância

Em estatística, a covariância é uma medida de como duas variáveis aleatórias variam simultaneamente. Se os maiores valores de uma variável correspondem, principalmente, aos maiores valores de outra variável, e o mesmo ocorrendo para os valores menores, as variáveis tendem a apresentar um comportamento semelhante e a covariância é positiva [15 e 21]. No caso oposto, quando os maiores valores de uma variável correspondem principalmente aos menores valores de outra variável, as variáveis tendem a apresentar um comportamento oposto e a covariância é negativa. O sinal de covariância, portanto, mostra a tendência na relação linear entre as variáveis.

A covariância para duas variáveis aleatórias  $X$  e  $Y$ , cada uma com  $N$  amostras, é definida por:

$$COV(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \quad (1.8)$$

Onde  $\mu_X = E(X)$  e  $\mu_Y = E(Y)$  são as respectivas médias, que podem ser escritas explicitamente, por:

$$COV(X, Y) = \sum_{i=1}^N \frac{(x_i - \bar{x})(y_i - \bar{y})}{N} \quad (1.9)$$

Para variáveis não correlacionadas,

$$COV(X, Y) = E(X)E(Y) - \mu_X\mu_Y = 0, \quad (1.10)$$

No entanto, se as variáveis são correlacionadas, sua covariância será diferente de zero, se as variáveis não são correlacionadas sua covariância é zero. De fato, se  $COV(X, Y) > 0$ , significa que  $Y$  aumenta à medida que  $X$  aumenta, e se  $COV(X, Y) < 0$ ,  $Y$  diminui à medida que  $X$  aumenta.

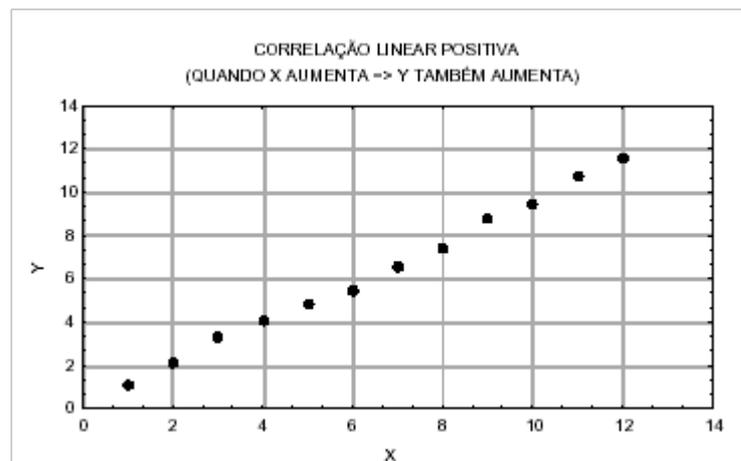
A normalização da covariância (coeficiente de correlação) mostra a sua magnitude e força da relação linear entre as variáveis.

### 1.4.5 Coeficiente de correlação

Em estudos que envolvem duas ou mais variáveis, é importante conhecer o grau de relacionamento entre elas. A medida do grau de relacionamento entre duas variáveis é chamada de coeficiente de correlação, também conhecida como medida de interdependência, de intercorrelação ou de relação entre as variáveis [17].

Diferentes formas de correlação podem existir entre as variáveis. O caso mais simples e mais conhecido é a correlação simples, envolvendo duas variáveis,  $X$  e  $Y$ . A relação entre duas variáveis será linear quando o valor de uma pode ser obtido por meio de uma equação da reta. Assim, é possível ajustar uma reta da forma  $Y = a + bX$  aos dados (sendo  $a$  e  $b$  constantes).

Uma forma simples de verificar o tipo de correlação existente entre duas variáveis e a tendência de variação conjunta entre elas é através do gráfico chamado de “diagrama de dispersão”. Trata-se de um gráfico onde são representados os pares  $(X_i, Y_i)$ ,  $i = 1, 2, \dots, n$ , onde  $n$  é o número total de observações. A figura 2 apresenta um exemplo de uma relação linear entre duas variáveis hipotéticas  $X$  e  $Y$ .

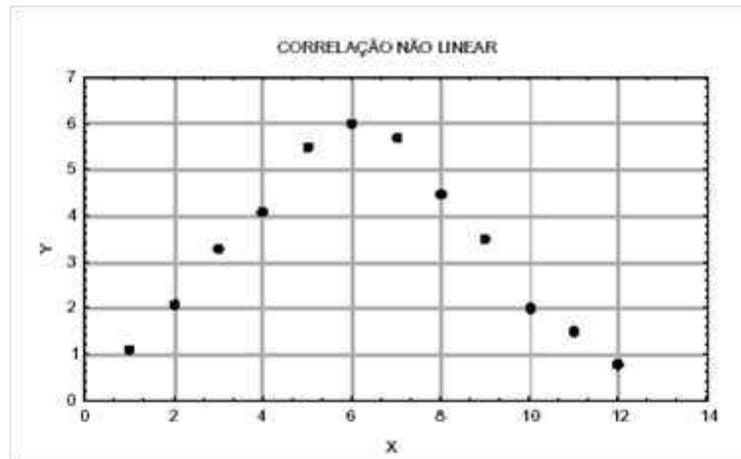


**Figura 2** – Correlação linear positiva [29]

Entretanto, quando não for possível o ajuste por uma reta, não significa que não existe correlação entre elas. Poderá haver uma correlação não-linear entre as mesmas, como mostra a figura 3.

O método que permite estudar as relações ou associações é conhecido como “Análise de Correlação”. Esta análise mostra o grau de relacionamento entre as variáveis, fornecendo um número adimensional que indica como as variáveis variam conjuntamente.

Quando a análise envolve grande número de variáveis e há interesse em conhecer a correlação duas a duas, é comum a construção de uma matriz de correlações. Esta é uma matriz formada pelas correlações entre  $X_i$  e  $X_j$ ,  $i \neq j$ , fora da diagonal e 1 na diagonal, indicando a correlação das variáveis  $X_i$  e  $X_j$ , sendo  $i = j$ .



**Figura 3** – Correlação não linear [29]

Podem ocorrer, ainda, situações onde se tem dois conjuntos de variáveis, um composto por uma variável ( $Y$ ) e outro com  $p$  variáveis ( $X_1, X_2, \dots, X_p$ ), e se deseja analisar a correlação entre a variável  $Y$  e a variável  $X_i$ ,  $i = 1, 2, \dots, n$ . Neste caso a correlação é chamada de múltipla.

Existem situações em que o relacionamento entre duas variáveis não é linear, ou uma delas é contínua, ou o número de pares das medidas é muito pequeno. Então, para cada uma dessas situações há necessidade de uma medida adequada de associação entre as variáveis.

#### 1.4.5.1 Coeficiente de correlação de Pearson

O método usualmente conhecido para medir a correlação entre duas variáveis aleatórias é o Coeficiente de Correlação de Pearson, também conhecido como Coeficiente de Correlação do Momento Produto [14].

O coeficiente de correlação de Pearson é bastante utilizado devido ao seu elevado grau de confiabilidade estatística mesmo para distribuições assimétricas. Além disso, o coeficiente de correlação de Pearson é capaz de fornecer resultados substanciais na detecção da associação linear entre duas variáveis em pequenas amostras ( $n < 30$ ). Na presença de *outliers* (valores discrepantes), o coeficiente de correlação de Pearson apresenta boa estimativa, desde que não sejam frequentes [13].

O coeficiente de correlação de Pearson ( $\rho$ ) é definido por:

$$\rho_{X,Y} = \rho = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \quad (1.11)$$

Onde  $\sigma_X$  e  $\sigma_Y$  são os desvios-padrão das variáveis  $X$  e  $Y$ , respectivamente, e  $COV(X,Y)$  é uma medida que expressa a variação conjunta de duas variáveis.

O coeficiente de correlação de Pearson ( $\rho$ ) é adimensional e varia de -1 a +1. O sinal indica a direção positiva ou negativa e o valor sugere a força da relação entre as variáveis.

Uma correlação perfeita (-1 ou 1) indica que o escore de uma variável pode ser determinado exatamente ao se saber o escore da outra. Por outro lado, uma correlação de valor zero indica que não há relação linear entre as variáveis [2 e 20].

Entretanto, na prática ocorrem diferentes valores de  $\rho$ . A interpretação de  $\rho$  depende muito dos objetivos de sua utilização e as razões pelas quais este é calculado. O coeficiente de correlação pode ser avaliado da seguinte maneira [6, 40 e 43]:

- Se  $0.00 \leq |\rho| < 0.30$ , existe fraca correlação linear;
- Se  $0.30 \leq |\rho| < 0.60$ , existe moderada correlação linear;
- Se  $0.60 \leq |\rho| < 0.90$ , existe forte correlação linear;
- Se  $0.90 \leq |\rho| < 1.00$ , existe correlação linear muito forte.

O coeficiente de correlação de Pearson não diferencia entre variáveis independentes e variáveis dependentes. O valor absoluto do coeficiente  $\rho$  não é afetado por qualquer transformação linear de  $X$  ou  $Y$  [24].

#### 1.4.6 Algumas suposições para a utilização do coeficiente de correlação de Pearson

Uma vez definido o conceito e sua interpretação, é importante entender algumas suposições necessárias para utilização do coeficiente de correlação de Pearson. São elas:

- As variáveis devem ser quantitativas (discretas ou contínuas);
- Inferências realizadas a partir de uma quantidade reduzida de observações devem ser interpretadas com cautela. Isso porque amostras pequenas não fornecem estimativas confiáveis.
- As variáveis devem ter distribuição normal, o que equivale a dizer que para cada  $X$  dado, a variável  $Y$  é normalmente distribuída. Esta hipótese é necessária para fazer inferências estatísticas, sendo imprescindível para amostras pequenas [11], e diminui sua importância à medida que aumenta o tamanho da amostra [19];
- Faz-se necessário a análise de *outliers* (valores discrepantes), pois o coeficiente de correlação é fortemente afetado pela presença deles, levando a erros do tipo I e do tipo II;
- É necessário que as observações sejam independentes, ou seja, a ocorrência de uma observação  $X_1$  não influencia a ocorrência de outra observação  $X_2$ . A violação desta condição implica risco de assumir correlações espúrias [5].

Obedecer às suposições acima é de vital importância durante a análise; caso contrário, pode-se incorrer em erros do tipo I ou tipo II. O erro do tipo I consiste em concluir que a hipótese nula é falsa quando ela é verdadeira. Logo, não existe relação entre as variáveis ( $H_0$  é verdadeira). O erro do tipo II consiste em concluir que a hipótese nula é verdadeira quando ela é falsa. Logo, existe relação entre  $X$  e  $Y$  ( $H_0$  é falsa) [27].

#### 1.4.7 Fatores que influenciam o coeficiente de correlação

Dentre os fatores que afetam a intensidade do coeficiente de correlação, bem como a precisão de sua estimação, podem ser citados o tamanho da amostra (principalmente quando é pequena), os *outliers* (valores discrepantes), a restrição da amplitude de uma das variáveis ou de ambas e erros de medidas.

Os *outliers* afetam o coeficiente de correlação principalmente quando a amostra é pequena, tendendo a distorcer o valor da média e, por consequência, o valor do coeficiente de correlação, levando a erros como concluir que uma hipótese é verdadeira quando é falsa, ou vice-versa. Quando a amostra é grande o efeito de *outliers* sobre o coeficiente de correlação é reduzido. Estes podem ser detectados por meio da análise exploratória (AE) [16] que emprega uma variedade de técnicas gráficas para:

1. Investigar hipóteses sobre um conjunto de dados;
2. Detectar *outliers*;
3. Examinar relações entre as variáveis;
4. Analisar aspectos distributivos.

A restrição de amplitude de uma das variáveis ou de ambas pode ocorrer ao selecionar uma amostra homogênea para estudo. Quando a amostra é homogênea, o valor do coeficiente de correlação tende a aumentar, pois um dos fatores na intensidade da correlação é a variabilidade da amostra. Quando maior a variabilidade das variáveis envolvidas, menor a correlação entre elas.

## 1.5 Distribuição Normal

A distribuição normal é de longe a mais importante distribuição de probabilidade. Esta distribuição é frequentemente uma boa aproximação para a distribuição de muitos conjuntos diferentes de dados na indústria [25]. Além de descrever uma série de fenômenos físicos e financeiros, possui grande uso na estatística inferencial. É inteiramente descrita por sua média e pelo seu desvio padrão, ou seja, conhecendo-se estes valores consegue-se determinar qualquer probabilidade em uma distribuição normal.

A distribuição normal foi descoberta em 1733 por Abraham de Moivre que desenvolveu uma expressão matemática da curva normal [41]. Seu resultado foi estendido por Laplace, em seu livro *Analytical Theory of Probabilities* (1982). Uma das primeiras aplicações foi empregada por Carl Gauss, que considerou a distribuição dos erros obtidos observações reais para modelos astronômicos. Os fundamentos de seu raciocínio estão contidos em sua obra publicada em 1809, com o título *Theoria Motus Corporum Coelestium*, foi neste trabalho que ele fez o primeiro uso da teoria dos mínimos quadrados e a curva normal.

Uma v.a.  $X$  é normalmente distribuída, se sua função densidade de probabilidade [9] é dada por:

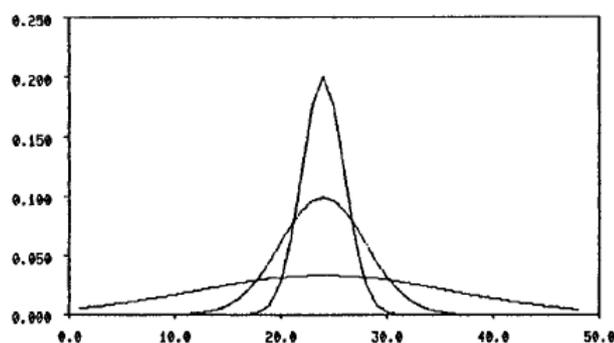
$$f(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)e^{\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]} \quad -\infty \leq x \leq +\infty \quad (1.12)$$

Onde  $\sigma$  é o desvio-padrão e  $\mu$  é a média da variável aleatória  $X$ , respetivamente. Note que, uma vez especificado a média e o desvio-padrão, a curva normal é completamente determinada.

A distribuição normal tem um pico de  $\left(\frac{1}{\sigma\sqrt{2\pi}}\right)$  em  $X = \mu$  e cai simetricamente com uma taxa determinada pelo desvio-padrão ( $\sigma$ ).

Na figura 4, é apresentado a curva normal com a variação do desvio padrão, mantendo a média constante. É possível notar que a variação do desvio padrão afeta tanto a forma da curva como seu ponto de máximo, além de manter sua área constante.

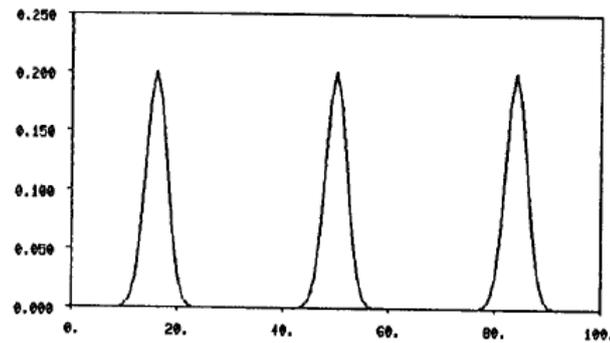
Na figura 5, é apresentada a curva normal com a variação da média, mantendo o desvio padrão constante. Nota-se que a variação da média não afeta a forma da curva normal, apenas a desloca ao longo do eixo horizontal.



**Figura 4** – Curva Normal com  $\mu$  constante [5]

Examinando a primeira e segunda derivadas, bem como o gráfico da distribuição normal, algumas propriedades são encontradas:

1. O máximo ocorre em  $X = \mu$
2. A curva é simétrica sobre o eixo vertical passando pela média



**Figura 5** – Curva Normal com  $\sigma$  constante [5]

3. Os pontos de inflexão estão localizados nos valores da média mais ou menos o desvio-padrão,  $X = \mu \pm \sigma$
4. A medida que se afasta da média, a curva se aproxima do eixo horizontal assintoticamente
5. A área total abaixo da curva e acima do eixo horizontal é igual a 1.

Como existem uma infinidade de distribuições normais (uma para cada média e desvio-padrão), a unidade estudada é transformada (peso, espessura, campo elétrico, potência, etc.), na unidade  $Z$ , que indica o número de desvios que separa a variável aleatória  $X$  da média.

Dessa forma, o cálculo de probabilidades (área sob a curva), pode ser realizado através de uma distribuição Normal padronizada, onde o parâmetro é a variável reduzida  $Z$ .

Para transformar um valor  $X$  no escore  $Z$ , usa-se a seguinte equação:

$$Z = \frac{x - \mu}{\sigma} \quad (1.13)$$

O cálculo da variável reduzida  $Z$  faz uma transformação dos valores reais em valores codificados. Esta transformação é feita descontando-se a média para eliminar o efeito de localização (tendência central) e dividindo-se pelo desvio padrão para eliminar o efeito de escala.

Se cada valor de dados de uma variável aleatória  $X$  normalmente distribuída for transformado em um escore  $Z$ , o resultado será uma curva normal padrão. Podemos utilizar a curva normal padrão e o escore  $Z$  para obter áreas (e portanto, probabilidades) sob qualquer curva normal [8].

O valor do escore  $Z$  é então, alocado em tabelas, que informa a probabilidade do evento.

## 1.6 Teste de normalidade

O coeficiente de correlação de Pearson ( $\rho$ ) é fortemente influenciado pela média da distribuição. Por esse motivo, um dos pressupostos centrais para que essa medida seja adequada-

mente utilizada é de que as observações obedeçam a uma distribuição normal. Existem testes disponíveis para averiguar em que medida as observações estão normalmente distribuídas. A maneira mais fácil é usando métodos gráficos como histogramas, porém não são suficientes para fornecer provas conclusivas sobre a normalidade dos dados. Portanto, para reforçar os métodos gráficos, métodos formais devem ser utilizados antes de fazer qualquer conclusão sobre a normalidade dos dados.

O teste de Shapiro-Wilk [39] é bastante aplicado como teste de normalidade devido as suas propriedades, e sua baixa taxa de erro, fazendo dele um dos mais eficazes e atraentes métodos de teste de normalidade [23]. Originalmente, o teste de Shapiro- Wilk (SW) foi restrito a apenas 50 amostras, sendo o primeiro teste a detectar desvios de normalidade devido a assimetria e achatamento na distribuição.

O método é baseado na estatística  $W$ , no qual atesta se um conjunto de amostras tem uma distribuição normal específica. A estatística  $W$  é calculada da seguinte forma: dado um vetor de valores esperados de ordem estatística normal padrão  $m' = (m_1, m_2, \dots, m_n)$ , e dada uma matriz quadrada de covariância  $V = (v_{ij})$ , isto é, se  $x_1 \leq x_2 \leq \dots x_n$  é uma amostra aleatória de tamanho  $n$  de uma distribuição normal com média 0 e variância 1, então:

$$E(x)_i = m_i \quad (i = 1, 2, \dots, n) \quad (1.14)$$

$$COV(x_i, x_j) = v_{ij} \quad (i, j = 1, 2, \dots, n) \quad (1.15)$$

Dado um vetor ordenado de observações  $y' = (y_1, \dots, y_n)$ . O objetivo é derivar um teste de hipóteses para uma amostra com distribuição normal de média  $\mu$  e variância  $\sigma$ .

Claramente, se  $y_i$  é um conjunto normal de amostras, pode ser expressado por:

$$y_i = \mu + \sigma x_i \quad (i = 1, 2, \dots, n) \quad (1.16)$$

Como resultado das generalizações do teorema dos mínimos quadrados [1], que as melhores estimativas lineares de  $\mu$  e  $\sigma^2$  que minimizam a forma quadrática  $(y - \mu \mathbf{1} - \sigma m)' \mathbf{V}^{-1} (y - \mu \mathbf{1} - \sigma m)$ , onde  $\mathbf{1}' = (1, 1, 1, \dots, 1)$  é um vetor unitário, são respectivamente:

$$\hat{\mu} = \frac{m' V^{-1} (m \mathbf{1}' - \mathbf{1} m') V^{-1} y}{\mathbf{1}' V^{-1} \mathbf{1} m' V^{-1} m - (\mathbf{1}' V^{-1} m)^2} \quad (1.17)$$

$$\hat{\sigma} = \frac{\mathbf{1}' V^{-1} (\mathbf{1} m' - m \mathbf{1}') V^{-1} y}{\mathbf{1}' V^{-1} \mathbf{1} m' V^{-1} m - (\mathbf{1}' V^{-1} m)^2} \quad (1.18)$$

Para uma distribuição simétrica,  $\mathbf{1}'V^{-1}m = 0$ , e portanto:

$$\hat{\mu} = \frac{1}{n} \sum_1^n y_i = \bar{y} \quad (1.19)$$

$$\hat{\sigma} = \frac{m'V^{-1}y}{m'V^{-1}m} \quad (1.20)$$

Dado

$$S^2 = \sum_1^n (y_i - \bar{y})^2 \quad (1.21)$$

que denota uma estimativa imparcial e simétrica de  $(n-1)\sigma^2$ .

A estatística  $W$  para o teste de normalidade é definida como:

$$W = \frac{(\sum_1^n a_i y_i)^2}{\sum_1^n (y_i - \bar{y})^2} \quad (1.22)$$

onde  $a_i = (a_1, \dots, a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}}$  é uma constante gerada pelas médias, variâncias e covariâncias de uma amostra de tamanho  $n$  a partir de uma distribuição normal,  $\bar{y}$  é média das amostras e  $y_i$  são os valores das amostras ordenados.

Pequenos valores de  $W$  ( $W \ll 0.01$ ) indicam que o conjunto de dados não pode ser modelado por uma distribuição normal específica. Grandes valores de  $W$  ( $W \gg 0.01$ ) significa que o conjunto de dados pode ser modelado por uma distribuição normal. O teste de SW foi modificado por Royston [33] para aumentar a restrição do tamanho da amostra para 2000 e o algoritmo AS181 foi fornecido [34-35]. Mais tarde Royston observou que os pesos  $a$  usados no algoritmo eram inadequados para amostras de tamanho superior a 50 [36]. Royston em 1995 forneceu um algoritmo que melhora a aproximação dos pesos, fornecendo o algoritmo AS R94 [37], podendo ser usado amostras de qualquer tamanho  $n$ , desde que esteja na faixa  $3 \leq n \leq 5000$ .

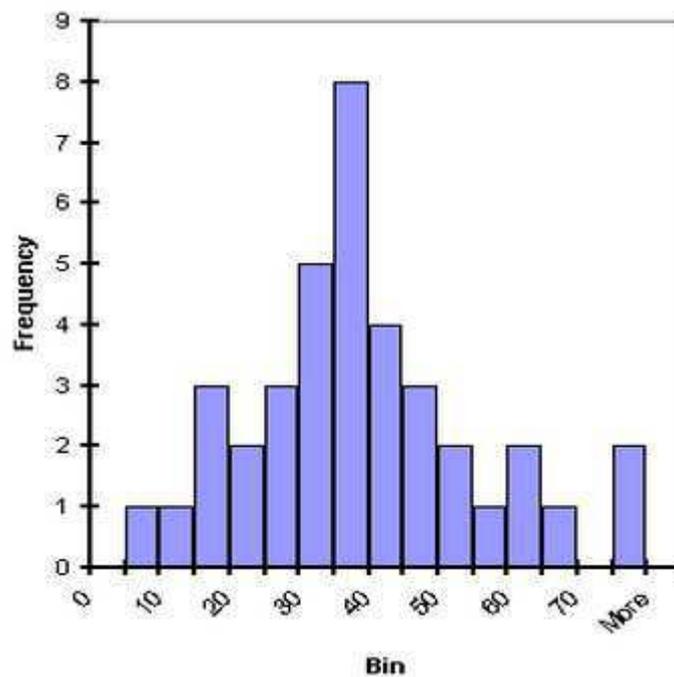
## 1.7 Histogramas

Uma maneira de descrever a distribuição de um conjunto de dados é por meio de histogramas, que permite observar imediatamente uma forma ou padrão específico de sua distribuição. Histogramas fornecem três informações importantes sobre a distribuição dos dados: forma, média e variância. Os histogramas dão informações sobre como os dados estão acumulados, uma vez que alguns valores ocorrem com mais frequência do que outros. Além disso, os picos nos histogramas mostram as semelhanças entre os dados. Este parâmetro pode fornecer uma

informação valiosa sobre o processo, enquanto a média não fornece uma informação completa sobre os dados.

Parte da importância de se usar histogramas é que eles permitem analisar grandes conjuntos de dados, reduzindo-os em um único gráfico que pode mostrar a localização dos picos e o grau de significância deles.

Na figura 6, é possível observar que o histograma apresenta um pico bem definido próximo ao valor da média. Embora existam *outliers*, eles são de frequência relativamente baixa. Assim, pode-se dizer que para este caso o desvio padrão não foi significativamente influenciado pela ocorrência de *outliers*.



**Figura 6** – Histograma com um pico centrado próximo a média [12]

Na figura 7, pode-se notar a presença de dois picos distintos de amplitudes diferentes. Isso pode ocorrer em situações onde há mistura de dados com diferentes médias.

Na figura 8, é apresentado um histograma na forma de platô, onde as classes centrais possuem aproximadamente a mesma frequência, esse aspecto está relacionado possivelmente quando há mistura de várias distribuições com médias diferentes.

Quando faz-se o uso de histogramas para classificação de isoladores, a localização da média nem sempre é crucial para distinguir o isolador limpo do isolador sujo. No entanto, o desvio e a forma do histograma pode indicar se os dados diferem entre si.

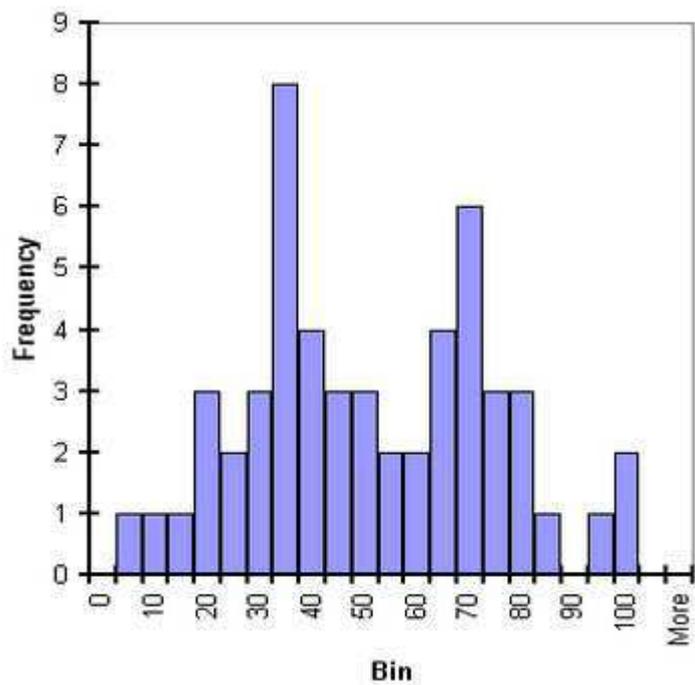


Figura 7 – Histograma com dois picos [12]

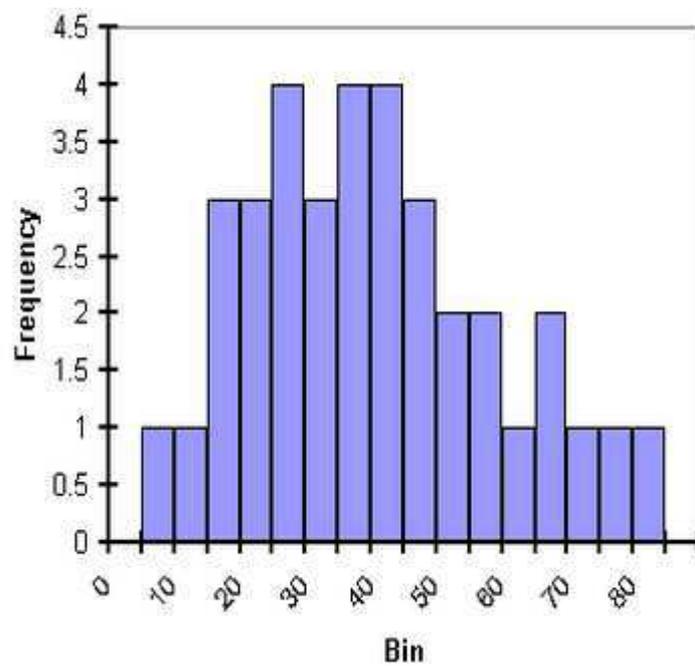


Figura 8 – Histograma com forma de platô [12]

## 1.8 Considerações sobre antenas

As antenas são classificadas conforme sua capacidade de distribuição de energia no espaço como:

- Isotrópicas
- Omnidirecionais
- Diretivas

### 1.8.1 Antena Isotrópica

Uma antena isotrópica é uma estrutura que tem como característica a capacidade de irradiar igualmente em todas as direções. É frequentemente utilizado como referência para expressar as propriedades diretivas de outras antenas. Antenas isotrópicas são também conhecidas como irradiadores isotrópicos (ISo).

A definição teórica de um ISo é a de fonte pontual de ondas eletromagnéticas (EM) que apresenta a mesma amplitude ou propriedades quando medida em qualquer direção (3D), a uma distância constante. Essa fonte de campos deve irradiar uniformemente em todas as direções sobre esferas concêntricas, sendo considerada o irradiador de referência para outros tipos de fontes. O ISo obedece lei do inverso do quadrado da distancia, na qual a potência total irradiada cai com o inverso do quadrado da distância.

O diagrama de irradiação do ISo é semelhante ao de uma casca esférica com seu centro na fonte de ondas EM, como pode ser visto na figura 9.

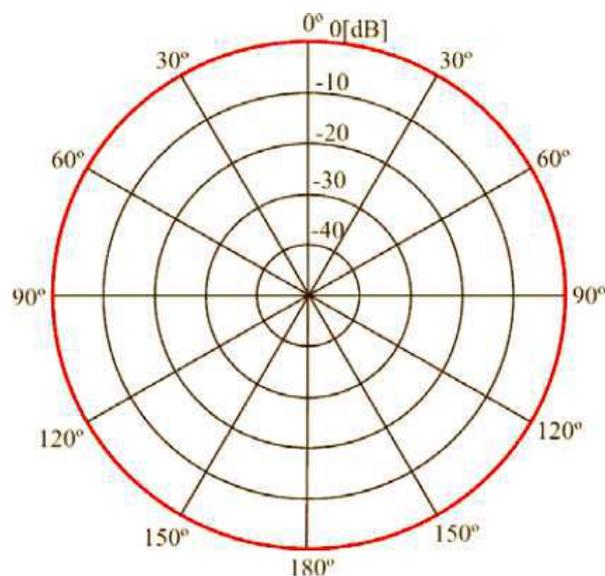


**Figura 9** – Diagrama 3D do irradiador isotrópico ISo [30]

### 1.8.2 Antena Omnidirecional

O irradiador Omnidirecional é uma antena que apresenta em um dos dois planos principais um diagrama de irradiação não-direcional, ou seja, sem preferência de irradiação e nos outros dois planos ortogonais uma direção preferencial de irradiação. Para aplicações com grandes ângulos de abertura essa antena apresenta boa facilidade de uso pelo fato da mesma não precisar ser direcionada, o que facilita sua instalação.

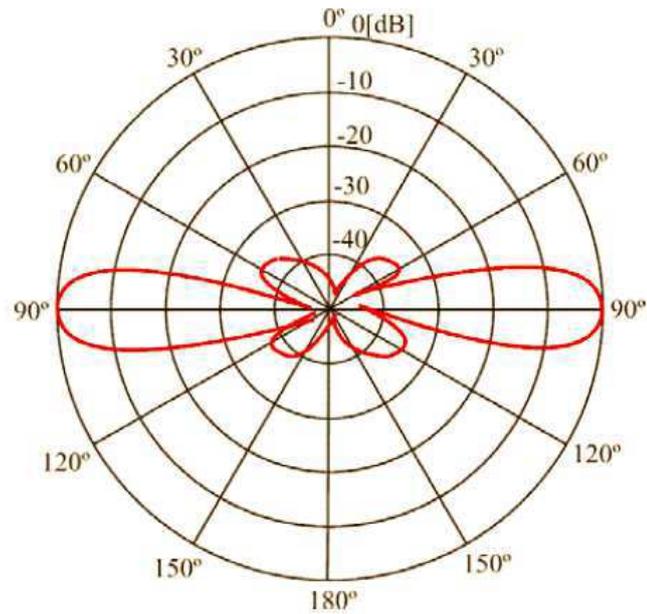
Como exemplo, considere o diagrama de irradiação, no plano horizontal, apresentando a mesma distribuição espacial dos campos encontrada para as antenas isotrópicas (Figura 10). Porém, num plano ortogonal ao plano horizontal, plano vertical, a distribuição espacial do campo se encontra localizada em setores preferenciais, não apresentando uma irradiação uniforme em todas as direções desse plano vertical, como pode ser visto na figura 11.



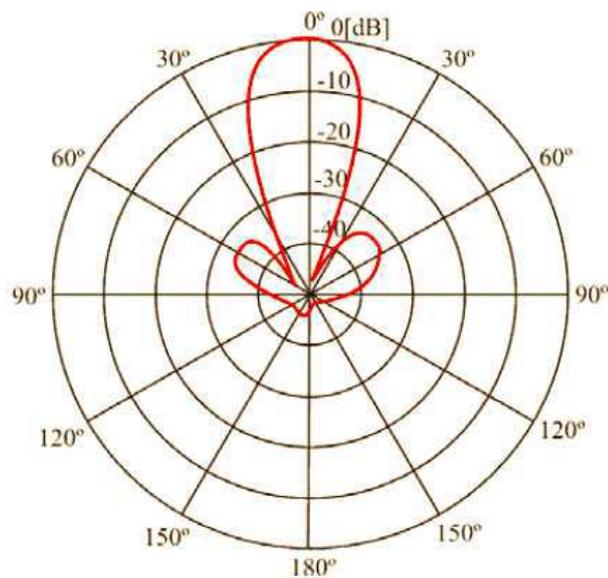
**Figura 10** – Diagrama no plano horizontal, distribuição espacial isotrópica [30]

### 1.8.3 Antena Direcional

O irradiador direcional não apresenta, em qualquer dos planos ortogonais, uma distribuição espacial uniforme de campo em todas as direções, como no caso do ISO. Essa antena pode apresentar uma ou várias direções preferenciais de irradiação, garantindo uma capacidade de irradiar ou receber ondas EM com mais eficiência em uma direção do que em outra, ou seja, concentra a energia de irradiação numa direção específica ou direções específicas. Na figura 12, temos o comportamento do diagrama em apenas um dos planos ortogonais, bastando fazer uma rotação em torno do seu eixo para obtermos a distribuição tridimensional. Antenas com essas características para o diagrama de irradiação são utilizadas, principalmente, em enlaces ponto a ponto e na eliminação de interferências provenientes de direções bem definidas.



**Figura 11** – Diagrama no plano vertical, distribuição espacial preferencial [30]



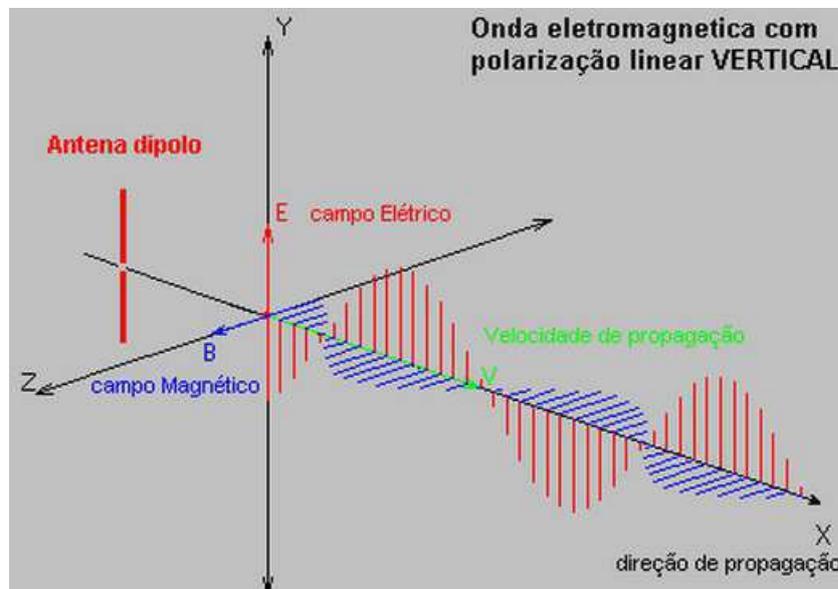
**Figura 12** – Exemplo de diagrama de irradiação para antenas diretivas [30]

#### 1.8.4 Polarização de antenas

A polarização de uma antena é a polarização da onda de campo elétrico por ela irradiado. Dessa forma a polarização da onda EM, é visualizada a partir da curva construída, sobre um plano de referência, normal à direção de propagação, com os pontos resultantes da projeção do extremo do vetor campo elétrico sobre este plano, quando da propagação da onda. Para que os casos de onda polarizada possam ser gerados, é necessário que dois campos em quadratura espacial sejam combinados, podendo serem variados suas amplitudes, suas fases relativas ou ambas.

A polarização pode ser classificada em três tipos independentes segundo a definição acima, como: elíptica, circular ou linear.

Se o vetor que representa o campo elétrico num ponto do espaço como função do tempo apresenta sua variação sempre ao longo de uma linha reta, que é normal à direção de propagação, o campo é dito polarizado linearmente. Como exemplo, considere uma antena dipolo posicionada verticalmente, alimentado por um gerador de frequência, gerando uma onda eletromagnética polarizada verticalmente (a componente do campo elétrico está no plano vertical e, conseqüentemente, o campo magnético está no plano horizontal). Na figura 13, observa-se os vetores  $\mathbf{E}$ ,  $\mathbf{B}$  e  $\mathbf{V}$ , representam respectivamente os vetores campo elétrico, campo magnético e velocidade de propagação. Nota-se que os três vetores citados estão  $90^\circ$  físicos entre si, e os vetores  $\mathbf{E}$  e  $\mathbf{B}$  estão em fase, ou seja, zero graus de defasamento elétrico.



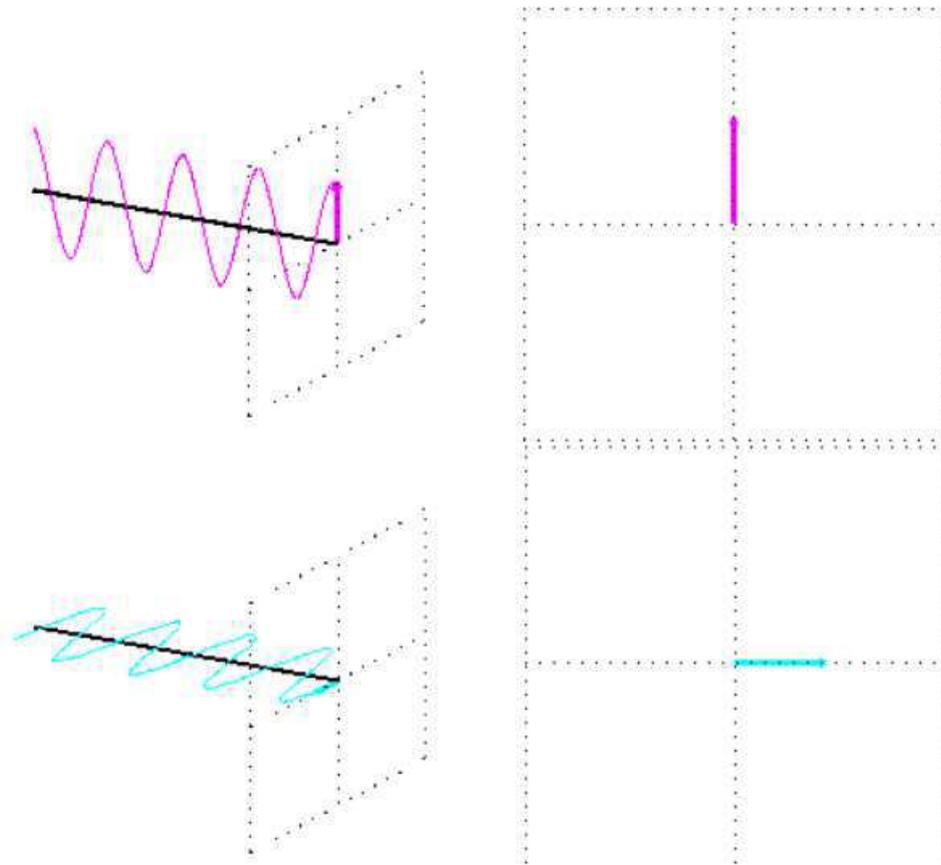
**Figura 13** – Onda EM com polarização linear Vertical [42]

A figura 14, ilustra uma onda com polarização linear vertical e uma onda com polarização linear horizontal (o campo magnético está a sempre a  $90^\circ$  físicos).

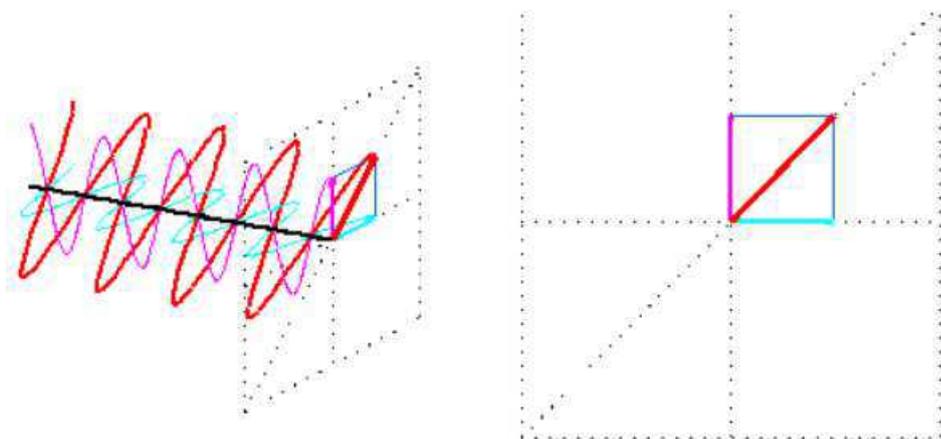
A combinação de duas ondas linearmente polarizadas, uma com polarização vertical e outra com polarização horizontal, defasadas eletricamente de  $0^\circ$ , resulta em uma onda com polarização oblíqua, como pode ser visto na figura 15.

A combinação de duas ondas EM linearmente polarizadas, uma na vertical e outra horizontal, de mesma amplitude e eletricamente defasadas de  $90^\circ$ , resulta em uma onda EM circularmente polarizada, como pode ser visto na figura 16.

Por último, temos a polarização elíptica onde o campo elétrico é dado pela soma vetorial de duas componentes, de amplitudes diferentes, polarizadas horizontalmente e verticalmente, e defasadas de  $90^\circ$  no tempo ou quando a defasagem no tempo é diferente de  $90^\circ$  independente do



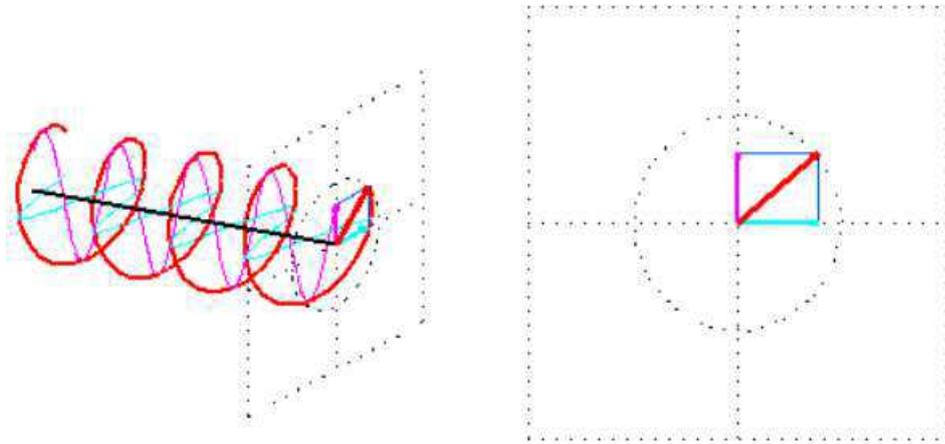
**Figura 14** – Ondas EM com polarização linear Vertical e Horizontal respectivamente [42]



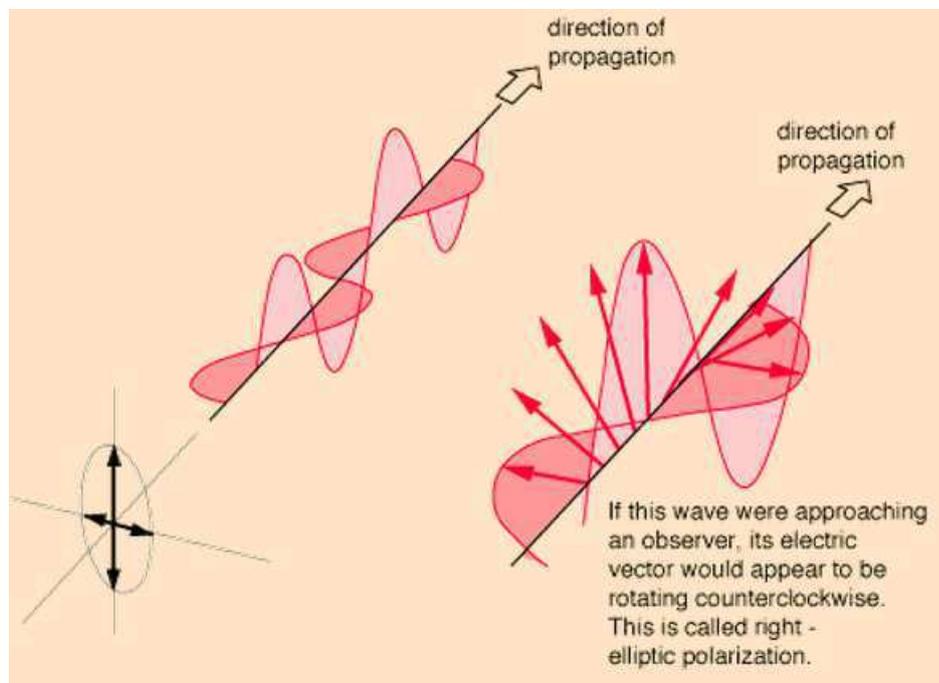
**Figura 15** – Onda EM com polarização oblíqua [42]

valor da amplitude de cada uma, como pode ser mostrado na figura 17.

Sinais irradiados podem conter informação em campos com polarizações desconhecidas. Dependendo do tipo de aplicação pode ser inadequado o uso de antenas receptoras com apenas uma das polarizações lineares, pois quando uma delas é utilizada os sinais com polarização não coincidente não serão captados da forma mais eficiente. Se o sinal e a antena estiverem com polarizações cruzadas o sinal não será recebido.



**Figura 16** – Onda EM com polarização oblíqua [42]



**Figura 17** – Onda EM com polarização circular [42]

No caso específico dos sinais irradiados pelos isoladores, a assinatura espectral do isolador pode estar tanto na polarização vertical como na horizontal. A escolha de um desses tipos de polarização leva ao descarte dos dados referentes a outras polarizações. Para contornar esse problema, uma antena com características especiais é empregada.

O sensor 3D (Isotrópico) empregado captura o sinal irradiado pelos isoladores, combinando-os em um único vetor campo elétrico que resume a polarização total que está incidindo no sensor. O sensor utilizado é do fabricante R&S (Rohde Schwarz) e cobre uma ampla faixa de frequência.

### 1.8.5 Intensidade de Radiação

Uma propriedade importante da antena é a sua característica direcional, propriedade que se refere à orientação espacial da irradiação. A intensidade de irradiação é um parâmetro de campo distante, e pode ser obtida através da multiplicação da densidade de potência com o quadrado da distância. Sendo expresso por:

$$U = r^2 W_{rad} \quad (1.23)$$

A potência total irradiada por uma antena é obtida pela integração da intensidade de irradiação, ao longo de todo ângulo sólido de  $4\phi$ , escrita matematicamente da seguinte forma:

$$P_{irr} = \oint_S U d\Omega \quad (1.24)$$

Onde o elemento diferencial  $d\Omega$  do ângulo sólido é definido como:  $\sin(\theta)d\theta d\phi$ .

A intensidade da irradiação para uma antena ISO é constante em todas as direções do espaço. Portanto, da equação (1.23), podemos calcular a intensidade da radiação média constante  $U_o$  da antena ISO a partir da equação (1.25).

$$U_o = \frac{P_{irr}}{4\pi} \quad (1.25)$$

### 1.8.6 Diretividade

A diretividade  $D(\theta, \phi)$  de uma antena é definida como a razão entre a intensidade da radiação  $U(\theta, \phi)$  em uma dada direção e a intensidade da radiação média  $U_o$  (esse é o caso da intensidade da irradiação de uma antena isotrópica onde  $U = U_o$ ) considerando todas as direções.

$$D(\theta, \phi) = \frac{U(\theta, \phi)}{U_o} \quad (1.26)$$

Para uma antena com componentes de polarização ortogonal, definimos a diretividade parcial numa dada polarização como a parte da intensidade da irradiação correspondente a uma determinada polarização dividida pela intensidade da irradiação média total sobre todas as direções [7]. Para um sistema em coordenadas esféricas, a diretividade máxima total  $D_o$  para as componentes ortogonais  $\theta$  e  $\phi$  de uma antena é escrita através da soma das duas diretividades parciais:

$$D(\theta, \phi)_0 = D(\theta, \phi)_\theta + D(\theta, \phi)_\phi \quad (1.27)$$

Onde as expressões para as diretividades parciais são:

$$D(\theta, \phi)_\phi = \frac{4\pi U_\phi}{P_\theta + P_\phi} \quad (1.28)$$

$$D(\theta, \phi)_\theta = \frac{4\pi U_\theta}{P_\theta + P_\phi} \quad (1.29)$$

### 1.8.7 Ganho

Como visto na seção anterior a diretividade é definida somente a partir da intensidade da irradiação. Para quantificar quão eficiente a antena transforma a potência disponível nos seus terminais de entrada em potência irradiada, considerando ainda suas características diretivas, definimos o parâmetro ganho de potência  $G$ . Ele leva em consideração as perdas da potência de alimentação por efeito Joule (aquecimento) na estrutura irradiante. O ganho absoluto de uma antena é proporcional a relação entre a intensidade da radiação, em uma dada direção, e a potência líquida nos terminais de entrada da antena, desconsiderados os descasamentos de impedância e de polarização:

$$G(\theta, \phi) = 4\pi \frac{U(\theta, \phi)}{P_{in}} \quad (1.30)$$

O ganho de potência reflete o fato das antenas reais não se comportarem como as antenas ideais devido ao fato de que parte da potência de entrada ser perdida (absorvida na antena ou nas estruturas vizinhas). Para antenas com eficiência de transmissão de 100% a característica de diretividade expressa também o seu ganho de potência para uma dada direção.

Embora o ganho esteja relacionado com a diretividade, ela é uma medida que considera a eficiência da antena, assim como sua capacidade direcional. Em muitos casos é comum empregar o ganho relativo, que é a razão entre o ganho de potência em uma determinada direção e o ganho de potência de uma antena de referência correspondente, com a mesma potência. Em muitos casos toma-se como referência a antena isotrópica (ISo). Neste caso, o ganho é expresso em dBi.

$$G(\theta, \phi) = 4\pi \frac{U(\theta, \phi)}{P_{in}(\text{fonte isotropica de baixa perda})} \quad (1.31)$$

O ganho leva em consideração as perdas na antena, estando relacionado à antena de referência. Quando a referência for o ISo indica-se o ganho por  $G_0$ , quando for o dipolo de meia onda por  $G_{\lambda/2}$ . O ganho expresso em decibéis é calculado por:

$$G(dB) = 10 \log_{10} G \quad (1.32)$$

## Parte II

### Tratamento dos dados

## 2 Tratamento dos dados

### 2.1 Informações gerais sobre a origem dos dados

Os dados foram obtidos por meio de testes realizados no Laboratório de Alta Tensão – LAT da Universidade Federal de Campina Grande – UFCG. A configuração dos equipamentos de medição foi composta de um analisador de espectro (*R&S FSL*) e uma antena de referência. Realizaram-se mais de 300 medições, modificando os níveis de tensão, a posição da antena e a frequência de leitura do espectro. Foram utilizados isoladores limpos e isoladores poluídos (com diferentes níveis de sujeira) obtidos no próprio laboratório. Dentre os dados medidos e coletados, escolheu-se cinco conjuntos para análise, possuindo as seguintes características:

#### Conjunto I:

- Nível de tensão aplicado ao isolador: 12 kV
- Antena: Bicônica
- Polarização: Horizontal
- Faixa de frequência: 30 - 330 MHz
- Isolador limpo: C1|1
- Isolador poluído: C2|2
- Número de pontos: 2001

#### Conjunto II:

- Nível de tensão aplicado ao isolador: 12 kV
- Antena: Bicônica
- Polarização: Vertical
- Faixa de frequência: 30 - 330 MHz
- Isolador limpo: C1|1
- Isolador poluído: C2|2
- Número de pontos: 2001

## Conjunto III:

- Nível de tensão aplicado ao isolador: 12 kV
- Antena: Bicônica
- Polarização: Horizontal
- Faixa de frequência: 30 - 330 MHz
- Isolador limpo: 8 MS
- Isolador poluído: 8 MS CM SEGUNDO
- Número de pontos: 601

## Conjunto IV:

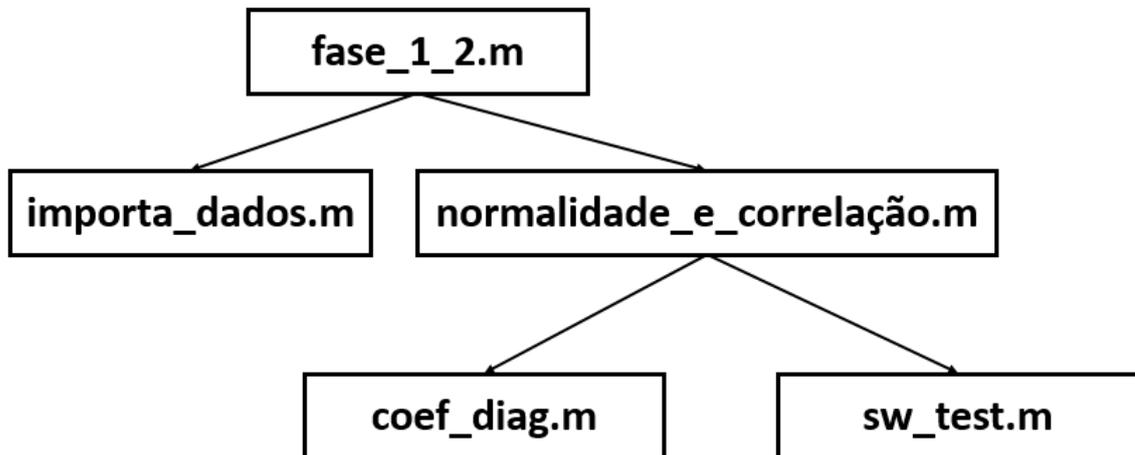
- Nível de tensão aplicado ao isolador: 12 kV
- Antena: Bicônica
- Polarização: Vertical
- Faixa de frequência: 30 - 330 MHz
- Isolador limpo: 8 MS
- Isolador poluído: 8 MS CM SEGUNDO
- Número de pontos: 601

## Conjunto V:

- Nível de tensão aplicado ao isolador: 12 kV
- Antena: Sensor 3D (Isotrópica)
- Polarização: 3D
- Faixa de frequência: 30 - 330 MHz
- Isolador limpo: —
- Isolador poluído: CEAL1
- Número de pontos: 601

## 2.2 Detalhes do processamento dos dados

O processamento de dados está baseado na relação entre quatro *scripts* (escritos utilizando o software MATLAB ©), são eles: *fase\_1\_2.m*, *importa\_dados.m*, *normalidade\_e\_correlacao.m*, *sw\_test.m* e *coef\_diag.m*. A relação entre os *scripts* é mostrada na figura abaixo.



**Figura 18** – Relação entre os scripts utilizados neste trabalho

De posse do conjunto de dados para análise, foi feita uma separação e organização manual entre o conjunto de amostras referente ao isolador limpo e ao isolador poluído. Depois disso, inicia-se o processamento dos dados pelo *script* principal *fase\_1\_2.m*.

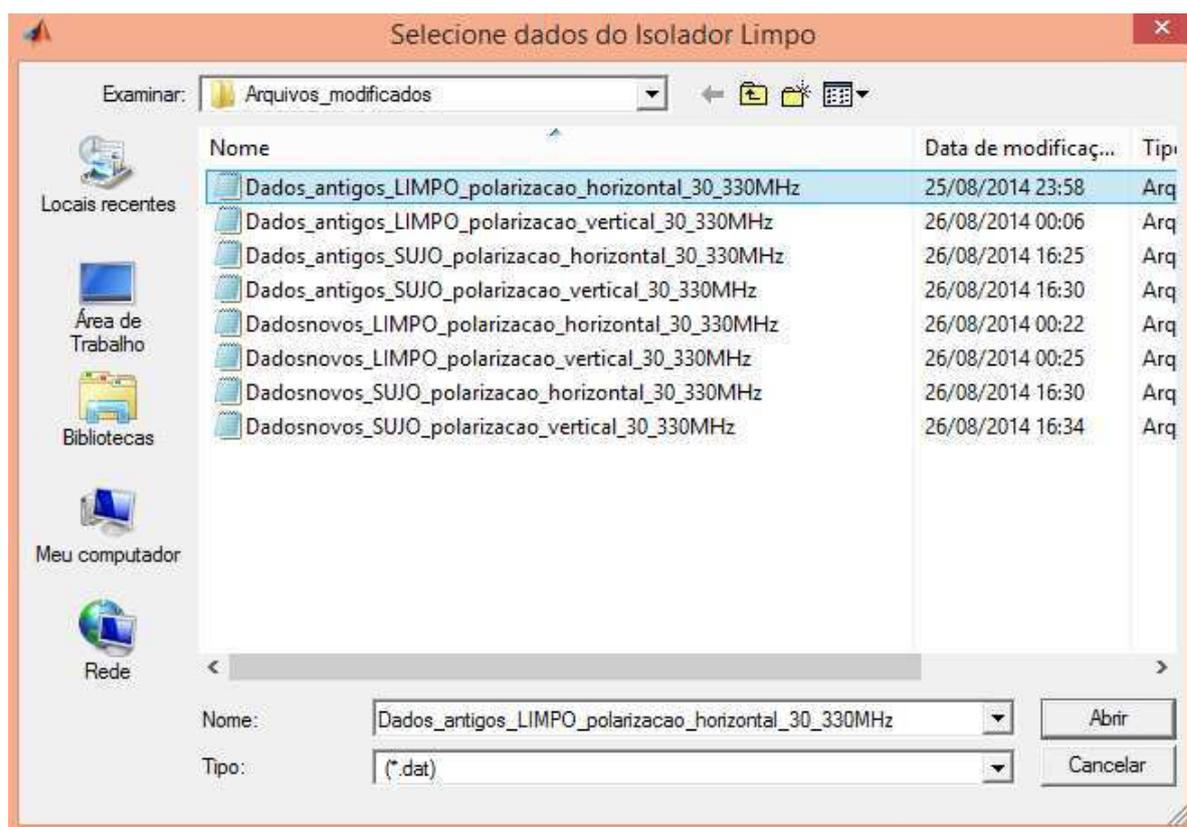
### 2.2.1 Script *fase\_1\_2.m*

Neste *script* os dados são importados e organizados. Obtêm-se como primeiro resultado o histograma e o espectro do sinal irradiado pelos isoladores. Para entender melhor como funciona este *script*, é feita uma breve descrição a seguir.

Para a importação dos dados, é utilizado o comando *uigetfile()*, este comando é responsável por exibir uma caixa de dialogo que permite ao usuário selecionar quais arquivos do isolador limpo e quais arquivos do isolador poluído serão importados. O usuário também pode digitar o nome de cada arquivo, como pode ser visto nas figuras 19 e 20. Se o arquivo é válido o comando retorna o nome do arquivo como uma *string*, caso contrário exibe uma mensagem de erro.

Uma vez escolhidos os arquivos, os dados são carregados no *workspace* do software MATLAB ©. Em seguida, as amostras de frequência e amplitude de cada isolador são separadas em variáveis distintas, facilitando a manipulação dos dados.

O primeiro resultado fornecido pelo *script* *fase\_1\_2.m* são os histogramas (comando *hist()*) sobrepostos do sinal emitido pelo isolador limpo e pelo isolador poluído. Além do histograma, este *script* fornece o gráfico do espectro emitido pelos isoladores (limpo e poluído).



**Figura 19** – Caixa de diálogo para seleção de arquivos do isolador limpo

O *script fase\_1\_2.m* envia os dados da amplitude dos isoladores (limpo e poluído) para o *script normalidade\_e\_correlacao.m*, este *script.m* fornece um dos principais resultados deste trabalho que é o teste de normalidade em conjunto com coeficiente de correlação.

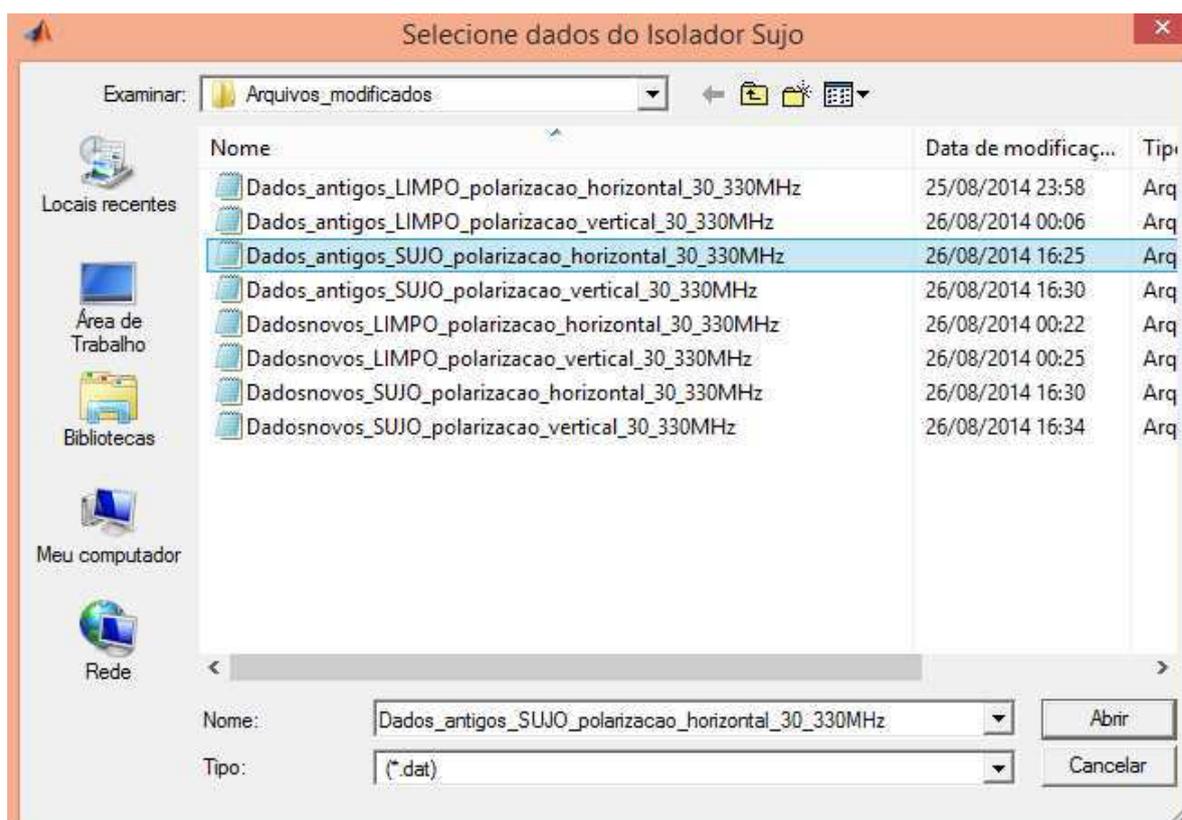
### 2.2.2 Script importa\_dados.m

O *script importa\_dados.m* é responsável por carregar todos os dados - durante a fase de seleção do *script fase\_1\_2.m* - no *workspace* do software utilizado. Este *script* recebe como parâmetro de entrada o nome do arquivo selecionado <sup>1</sup>, dependendo da natureza do arquivo, podem ser necessários alguns ajustes nos parâmetros 'DELIMITER' e 'HEADERLINES'. O parâmetro 'DELIMITER' é interpretado como um separador de colunas em um arquivo. O parâmetro 'HEADERLINES', é interpretado como o número de linhas que compõe o cabeçalho de texto (ignorado pelo programa). O comando *importdata()* carrega os dados no *workspace* do software, retornando uma matriz de células com os dados textuais e numéricos.

### 2.2.3 Script normalidade\_e\_correlacao

O *script normalidade\_e\_correlacao()* é responsável pelos teste de normalidade de SW, pelo cálculo do coeficiente de correlação, eliminação de picos indesejados e diagnóstico da faixa.

<sup>1</sup> **Nota:** tanto o arquivo selecionado como o *script* devem estar na mesma pasta



**Figura 20** – Caixa de diálogo para seleção de arquivos do isolador sujo

Inicialmente, este *script* recebe como parâmetros de entrada os dados de amplitude dos isoladores (limpo x poluído), e em seguida, dois conjuntos de amostras (um atribuído ao isolador limpo e outro atribuído ao isolador poluído) de tamanho definido pelo usuário. São utilizados pelo teste de normalidade. Se a normalidade de apenas um ou nenhum dos conjuntos passar no teste (apresentar distribuição normal), uma amostra será adicionada em ambos os conjuntos e o teste será repetido até que os dois conjuntos passem no teste; em seguida o coeficiente de correlação de Pearson é calculado.

O coeficiente de correlação é calculado utilizando comando *corrcoef()*. Por *default* este comando calcula o coeficiente de correlação de Pearson entre duas variáveis, retornando uma matriz quadrada como pode ser visto na figura 21. A interpretação da matriz de coeficientes de correlação é feita da seguinte forma: os elementos da diagonal principal representam os coeficientes de correlação de uma variável com ela mesma, ou seja, os elementos da diagonal principal sempre apresentarão valores iguais a 1. Por outro lado, os elementos fora da diagonal principal correspondem ao coeficiente de correlação entre as duas variáveis. Sendo assim, é armazenado o coeficiente de correlação fora da diagonal principal.

A fim de melhorar a visualização e análise gráfica do coeficiente de correlação *versus* frequência, faz-se necessário eliminar possíveis picos, que na maioria das vezes se devem à presença de *outliers*. Os parâmetros utilizados para se identificar um pico de correlação são:

$$R = \begin{bmatrix} 1.0000 & 1.0000 \\ 1.0000 & 1.0000 \end{bmatrix}$$

**Figura 21** – Matriz de coeficientes de correlação

largura de faixa e a amplitude do coeficiente de correlação. Esses parâmetros são definidos pelo usuário e dependendo da aplicação do critério podem sofrer alterações. No presente trabalho, para larguras de faixa menores que 0.5 MHz ou 500 kHz e amplitudes de correlação fora da faixa que vai de  $-0.2$  à  $+0.2$  são considerados picos de correlação, logo são eliminados zerando a amplitude do pico. Uma vez eliminados os picos de correlação, o gráfico do coeficiente de correlação *versus* frequência é construído.

#### 2.2.4 Script `sw_test.m`

Este *script* realiza o teste de normalidade de Shapiro-Wilk, retornando a estatística duas hipóteses:  $H_0$  conjunto de dados tem distribuição normal e  $H_1$  conjunto de dados não tem distribuição normal. É importante ressaltar que o teste SW é válido para amostras de tamanho entre 3 e 5000, sendo assim, inicia-se o teste com dois conjuntos de dados (amostras do isolador limpo e do isolador poluído) selecionando as cinco primeiras amostras de cada. Caso a hipótese  $H_0$  seja satisfeita para ambos os conjuntos selecionados, o coeficiente de correlação de Pearson é calculado e armazenado, caso contrário é adicionado uma amostra para cada conjunto, e a hipótese é testada novamente. Como resultado obtém-se coeficientes de correlação cobrindo faixas de frequência de largura variada.

#### 2.2.5 Script `coef_diag.m`

Este *script* classifica cada faixa de frequência em três níveis, são eles: 1 - Fraco; 2 - Moderado; e 3 - Forte. Cada um dos três níveis tem como base o nível de correlação pré-definido na seção Fundamentação Teórica. A classificação é feita a cada 10 MHz com objetivo de apresentar de forma resumida o coeficiente de correlação *versus* frequência.

A classificação é feita da seguinte forma: um vetor com os dados de frequência e amplitude dos coeficiente de correlação é analisado e comparado com os três principais níveis de correlação apresentados neste trabalho (Fraco, Moderado e Forte). Se mais de 10% da faixa, ou seja, 1 MHz apresentar correlação moderada, toda a faixa será classificada com o nível moderado de correlação, da mesma forma se a faixa apresentar correlação forte. Porém, se menos de 10% da faixa apresentar correlação forte ou moderada, a faixa de frequência será classificada com nível fraco de correlação.

## Parte III

### Resultados e Discussões

## 3 Resultados e Discussões

### 3.1 Análise do Conjunto I

Durante a análise do primeiro conjunto de dados (Conjunto I), o melhores resultados encontrados pela análise por histogramas ocorreram em toda a faixa de frequência considerada (30 MHz a 330 MHz), como mostra as figuras 22-24. Nota-se, considerável discrepância entre a variância do isolador limpo e do isolador poluído em toda faixa de frequência considerada. A tabela 2, mostra os valores da média e da variância do espectro de cada isolador em toda a faixa que vai de 30 MHz a 330 MHz.

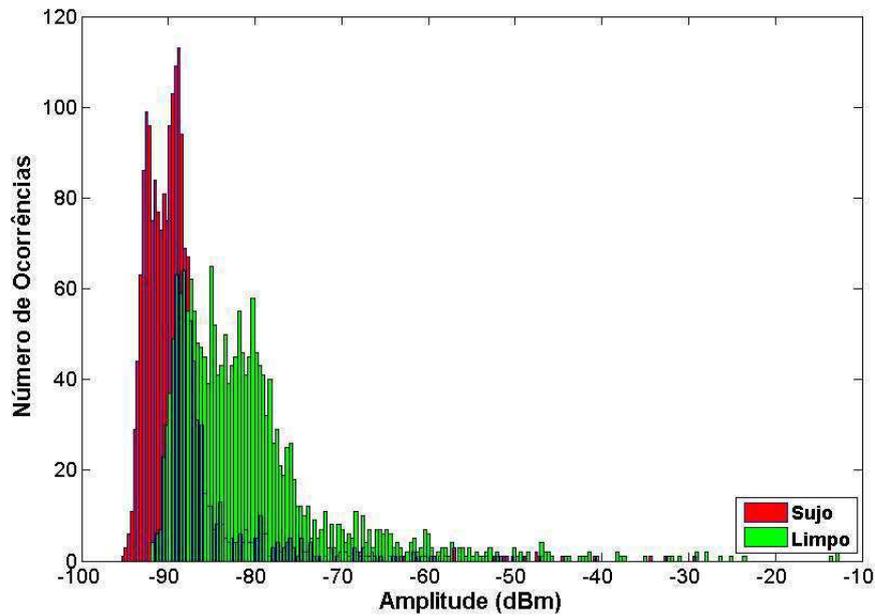
Frequência (MHz)	Isolador Limpo		Isolador Poluído	
	Média	Variância	Média	Variância
30 - 130	-80,190	95,1659	-87,4760	48,5577
130 - 230	-86,7394	46,9595	-92,5394	9,0115
230 - 330	-90,2445	11,9926	-92,9176	0,8494

**Tabela 2** – Média e variância do espectro de cada isolador

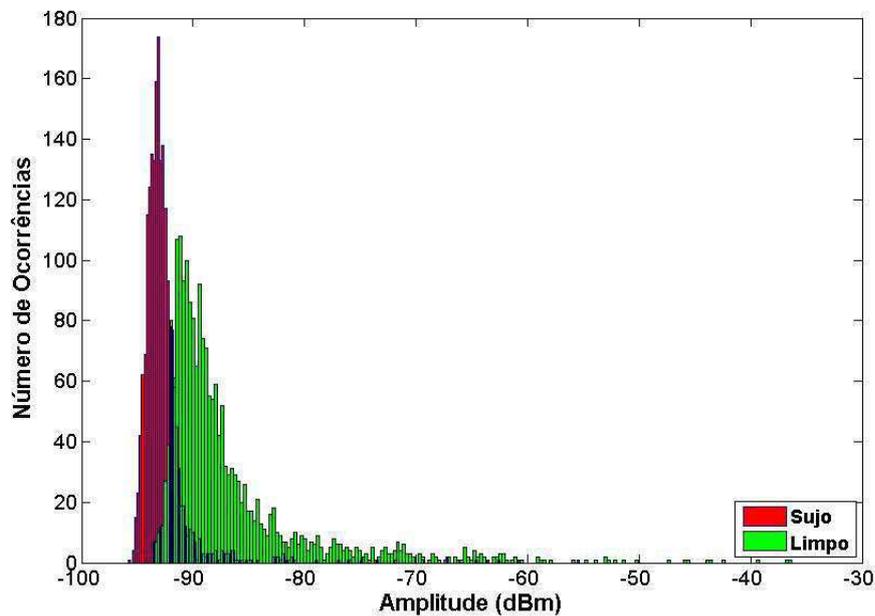
O emprego da análise por histogramas permite distinguir isoladores limpos de isoladores poluídos onde seus respectivos espectros são distintos entre si, como pode ser visto nas figuras 25-27. Isso ocorre devido ao nível de potência e ao grau de variabilidade do sinal irradiado pelo isolador limpo ser superior se comparado com sinal irradiado pelo isolador poluído.

A distinção entre as médias e as variâncias das amostras do espectro emitido pelos isoladores torna o método de análise por histogramas bastante eficiente, produzindo resultados de rápida interpretação. Porém, se os sinais emitidos pelos isoladores apresentarem média e variância aproximadamente iguais, o método por histogramas torna-se inadequado para classificação do estado de conservação dos isoladores, tornando-se necessário o emprego de outros métodos estatísticos em conjunto.

Para compensar a limitação imposta pelo método de análise por histogramas, fez-se necessária a utilização de outros métodos estatísticos como a análise por correlação e o teste de normalidade. O emprego do teste de normalidade em conjunto com o coeficiente de correlação, permite identificar faixas que contenham a assinatura espectral dos isoladores. As figuras (28-30) mostram a aplicação dos métodos nas faixas de 30 MHz a 130 MHz, 130 MHz a 230 MHz e de 230 MHz a 330 MHz respectivamente. É possível observar significativa oscilação do coeficiente de correlação ao longo de toda a faixa de frequência. Para esta aplicação o nível de correlação escolhido está entre fraco e moderado (entre -0.6 a +0.6), sendo assim, oscilações dentro dessa faixa são toleráveis e relevantes, desde que, apresentem menor comportamento oscilatório. A



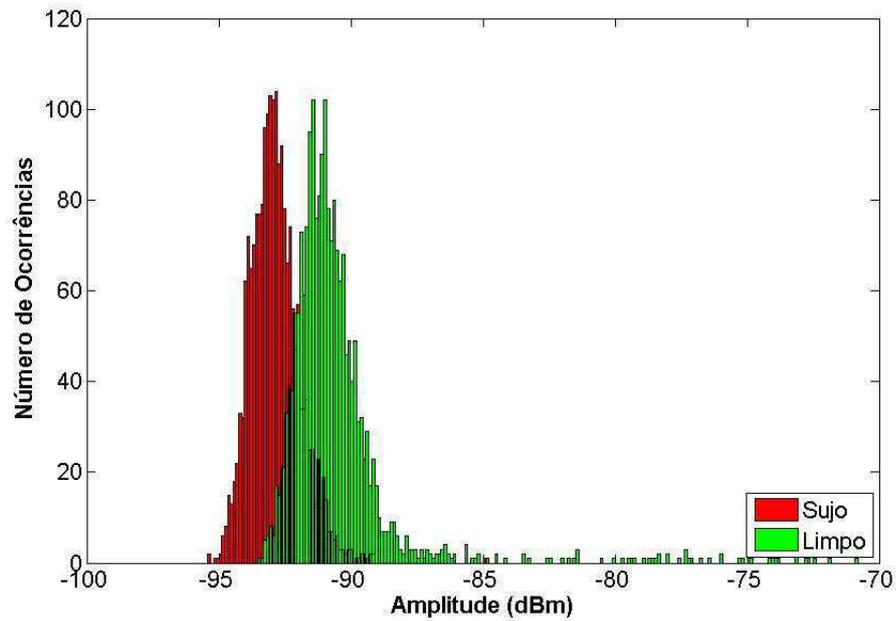
**Figura 22** – Histograma do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização horizontal



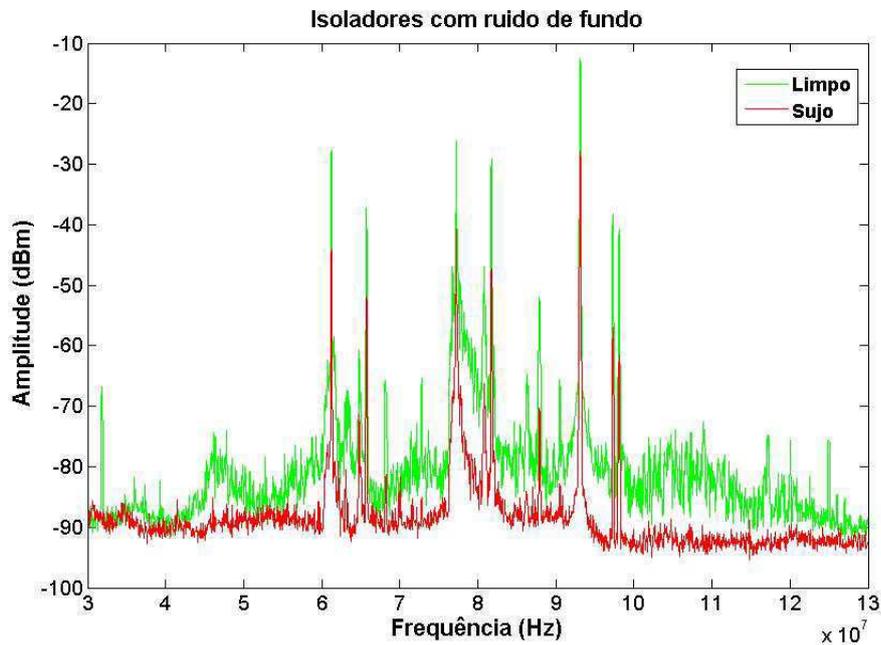
**Figura 23** – Histograma do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utilizando uma antena bicônica com polarização horizontal

tabela 3 apresenta as sub-faixas de frequência selecionadas que detêm as características citadas anteriormente.

Com o objetivo de melhorar a visualização dos resultados exibidos nas figuras (28-30),

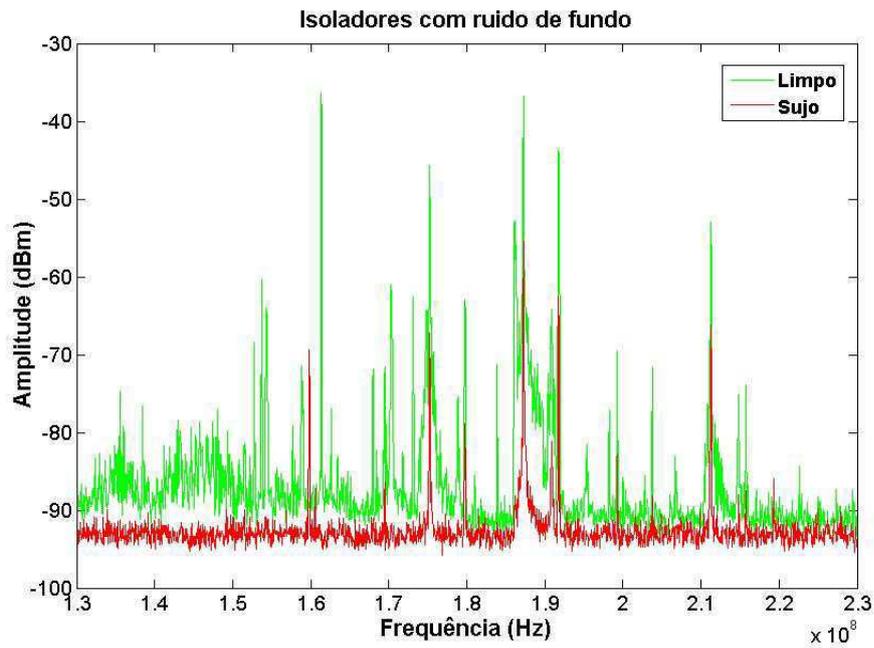


**Figura 24** – Histograma do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utilizando uma antena bicônica com polarização horizontal

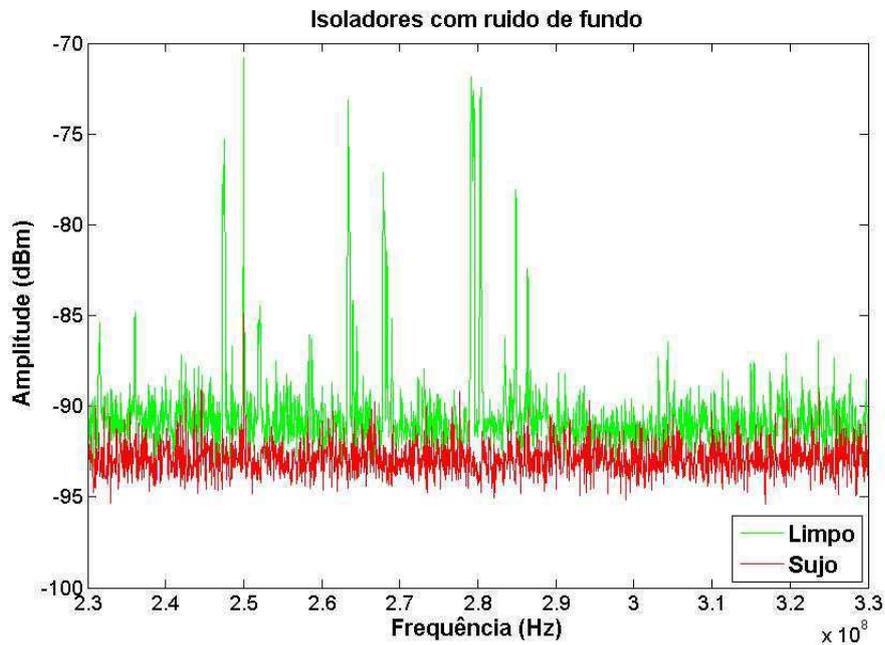


**Figura 25** – Espectro do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização horizontal

buscou-se apresenta-los de uma forma resumida. Para isso classificamos cada faixa em três níveis distintos: Nível I (amplitude igual a 1, onde existe relação fraca entre o isolador limpo x isolador sujo), Nível II (amplitude igual a 2, onde existe relação moderada entre o isolador limpo x isolador sujo) e Nível III (amplitude igual a 3, onde existe relação forte entre o isolador limpo

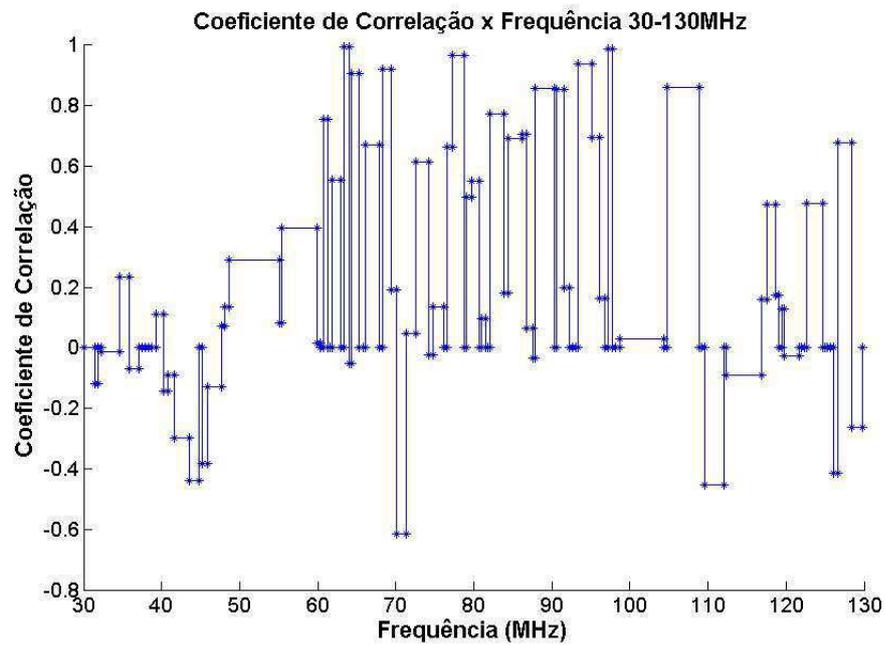


**Figura 26** – Espectro do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utilizando uma antena bicônica com polarização horizontal

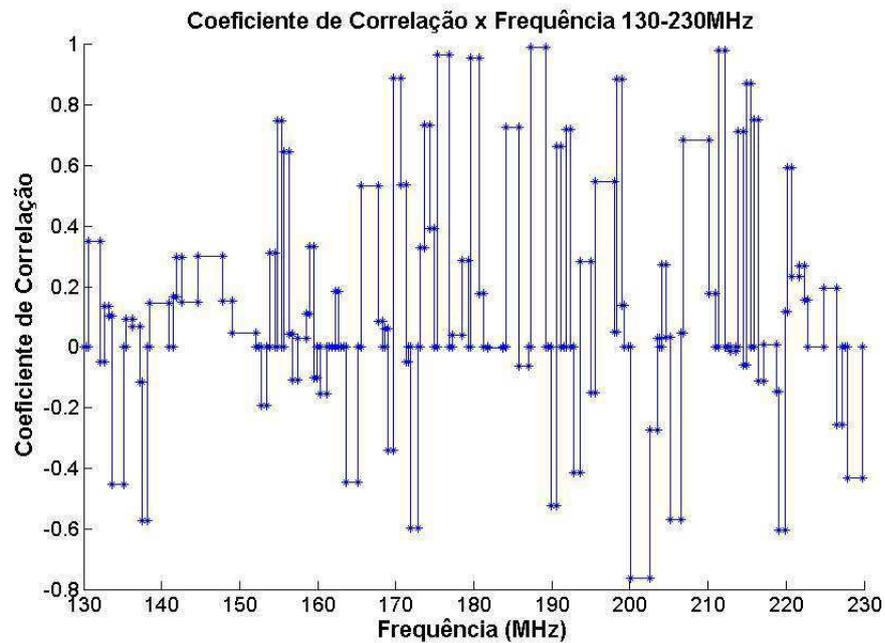


**Figura 27** – Espectro do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utilizando uma antena bicônica com polarização horizontal

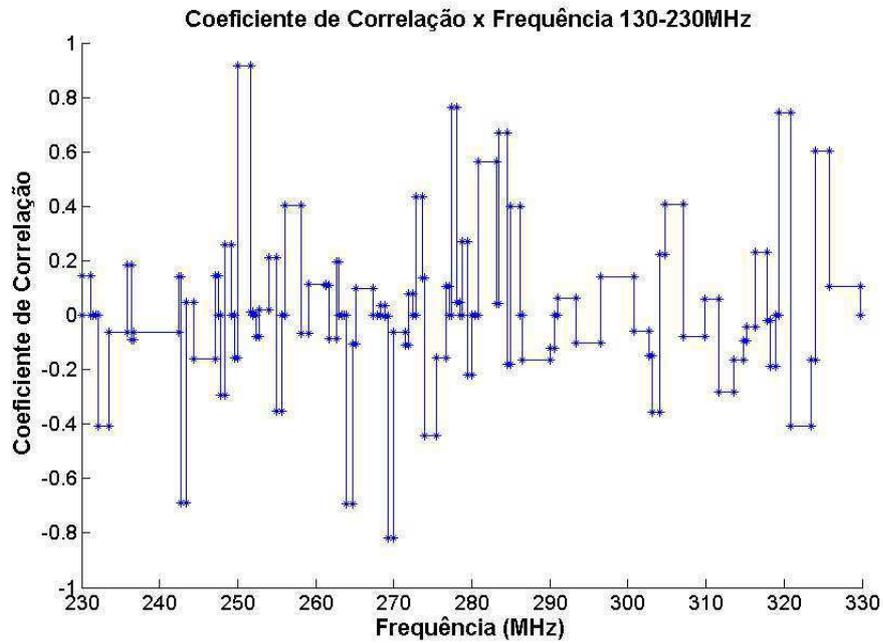
x isolador sujo), cada classificação é feita a cada 10 MHz. As figuras (31-33) apresentam os resultados da classificação feita para a faixa de 30 MHz a 330 MHz, onde é possível identificar faixas de correlação fraca (Nível I), correlação moderada (Nível II) e correlação forte (Nível III).



**Figura 28** – Coeficiente de correlação do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização horizontal



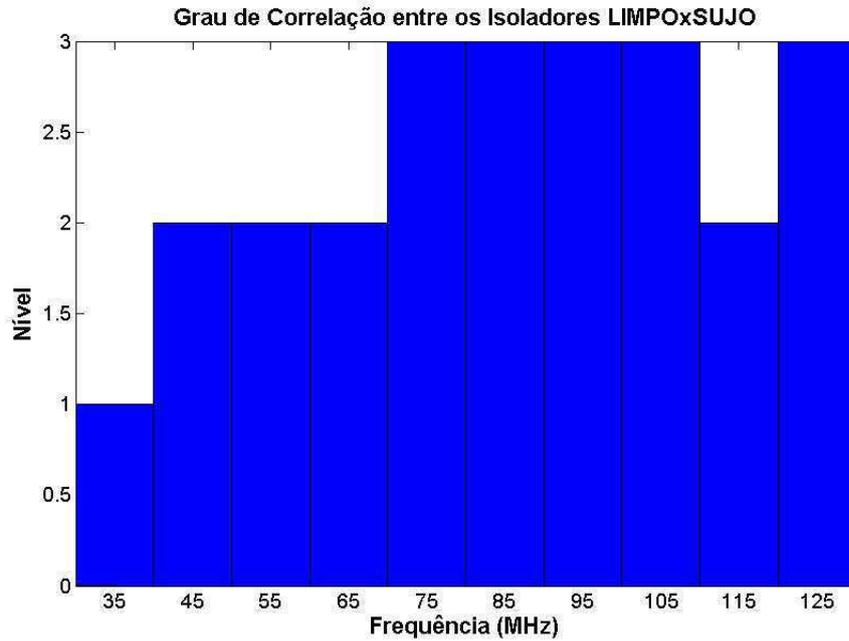
**Figura 29** – Coeficiente de correlação do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utilizando uma antena bicônica com polarização horizontal



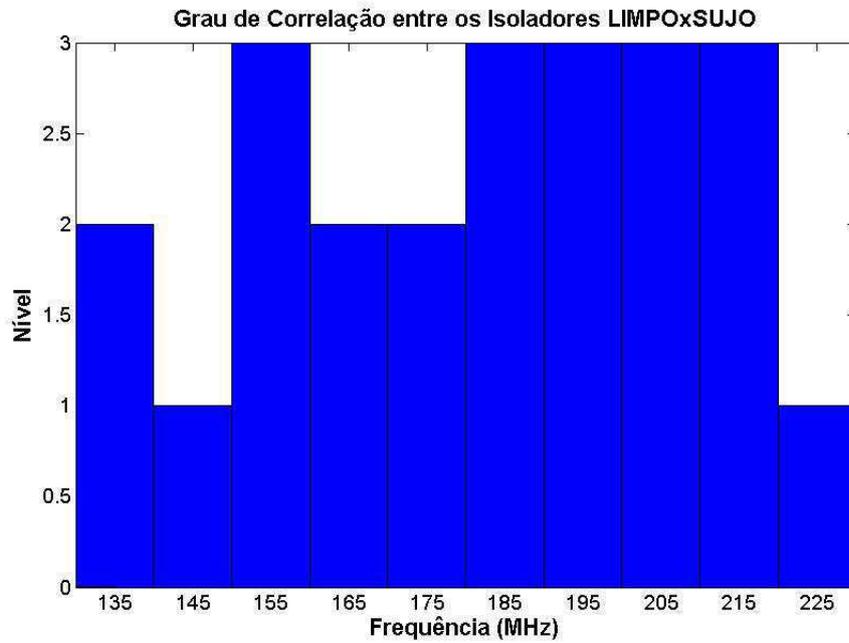
**Figura 30** – Coeficiente de correlação do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utilizando uma antena bicônica com polarização horizontal

Faixa analisada (MHz)	Faixas de interesse (MHz)		
	30 - 130	30 - 70	110 - 120
130 - 230	130 - 150	160 - 180	220 - 230
230 - 330	230 - 250	260 - 280	320 - 330

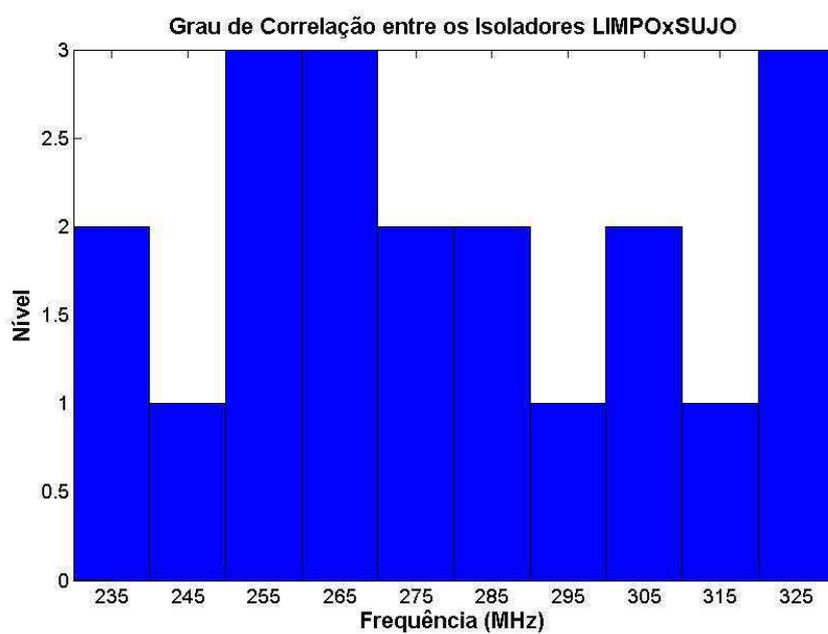
**Tabela 3** – Faixas de interesse para o Conjunto I



**Figura 31** – Grau de correlação do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização horizontal



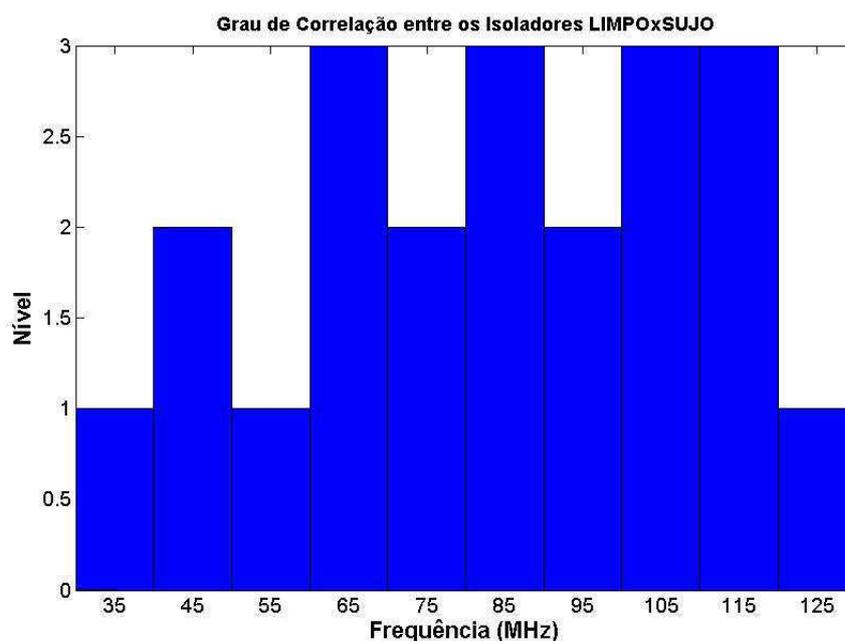
**Figura 32** – Grau de correlação do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utilizando uma antena bicônica com polarização horizontal



**Figura 33** – Grau de correlação do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utilizando uma antena bicônica com polarização horizontal

## 3.2 Análise do Conjunto II

A análise do segundo conjunto de dados (Conjunto II) é feita de forma resumida, utilizando-se a antena bicônica com polarização vertical, onde examinamos o coeficiente de correlação para cada 10 MHz de faixa, os resultados referentes são apresentados nas figuras (34-36).

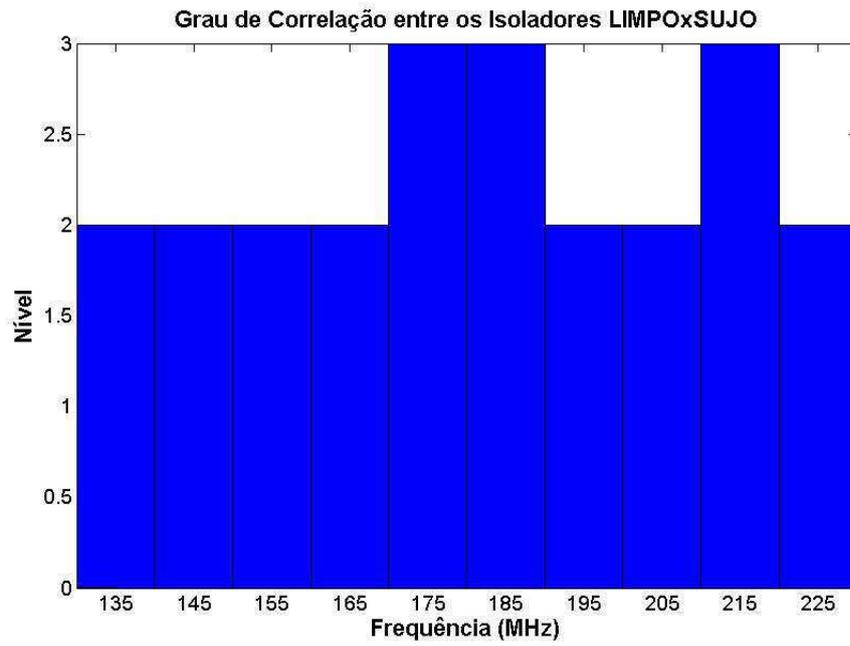


**Figura 34** – Grau de correlação do isolador limpo e do isolador sujo (30 MHz a 130 MHz) utilizando uma antena bicônica com polarização vertical

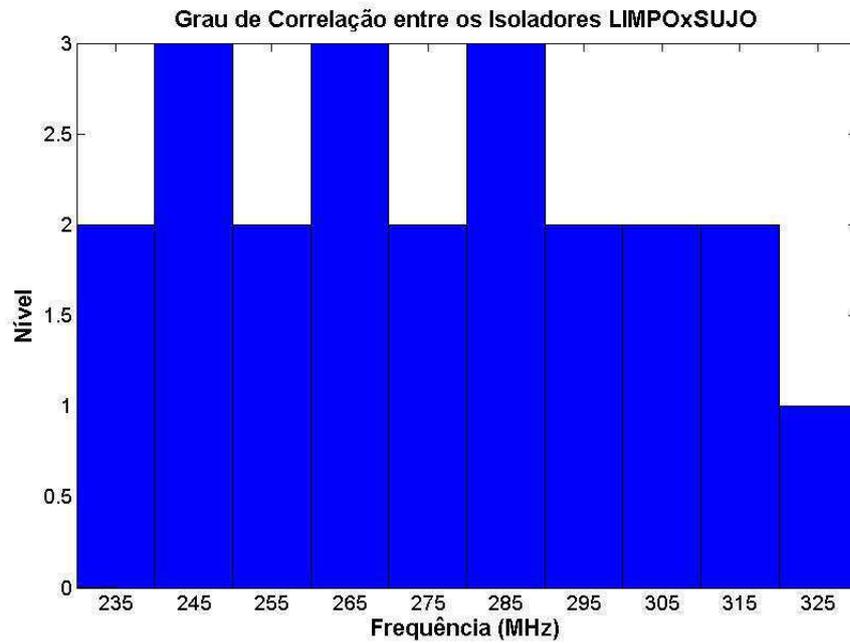
Se compararmos com os resultados apresentados na seção anterior (figuras 31-33), observamos que existem faixas em comum, onde o coeficiente de correlação apresenta-se nos níveis fraco ou moderado, ou seja, é possível distinguir o isolador limpo do isolador sujo nessas faixas, tanto na polarização horizontal como na polarização vertical. Além disso, existem faixas de frequência, como por exemplo, a faixa de 120 MHz a 130 MHz (Figura 34) onde os sinais emitidos pelos isoladores estão fracamente correlacionados, enquanto que para a mesma faixa de frequência (figura 31) os sinais emitidos pelos isoladores estão fortemente correlacionados, isto indica que existem faixas de frequência onde parte do sinal relevante que está sendo emitido pelos isoladores não está sendo captado pela antena.

A tabela 4 apresenta as frequências de interesse, onde comparamos com a tabela 3, com a finalidade de observar em quais faixas de frequência o sinal emitido pelos isoladores é captado em apenas uma das polarizações (horizontal ou vertical), como resultado temos as tabelas 5-7.

O próximo objetivo é comparar qual polarização é mais adequada, isto é, a que oferece a menor rejeição do sinal emitido pelos isoladores. Para atingir esse objetivo comparamos a antena



**Figura 35** – Grau de correlação do isolador limpo e do isolador sujo (130 MHz a 230 MHz) utilizando uma antena bicônica com polarização vertical



**Figura 36** – Grau de correlação do isolador limpo e do isolador sujo (230 MHz a 330 MHz) utilizando uma antena bicônica com polarização vertical

bicônica (polarizada horizontalmente e verticalmente) com o sensor 3D, este sensor é capaz de captar o sinal nas três polarizações:  $x$ ,  $y$  e  $z$ .

Faixa analisada (MHz)	Faixas de interesse (MHz)			
30 - 130	30 - 60	70 - 80	90 - 100	120 - 130
130 - 230	130 - 170	190 - 210	220 - 230	—
230 - 330	230 - 240	250 - 260	270 - 280	290 - 330

**Tabela 4** – Faixas de interesse para o Conjunto II

Horizontal Sub-faixa (MHz)	Vertical Sub-faixa (MHz)
60 - 70	70 - 80
110 - 120	90 - 100
—	120 - 130

**Tabela 5** – Sub-faixas onde o sinal é captado apenas na polarização vertical ou horizontal, na faixa de 30 MHz até 130 MHz

Horizontal Sub-faixa (MHz)	Vertical Sub-faixa (MHz)
170 - 180	190 - 210

**Tabela 6** – Sub-faixas onde o sinal é captado apenas na polarização vertical ou horizontal, na faixa de 130 MHz até 230 MHz

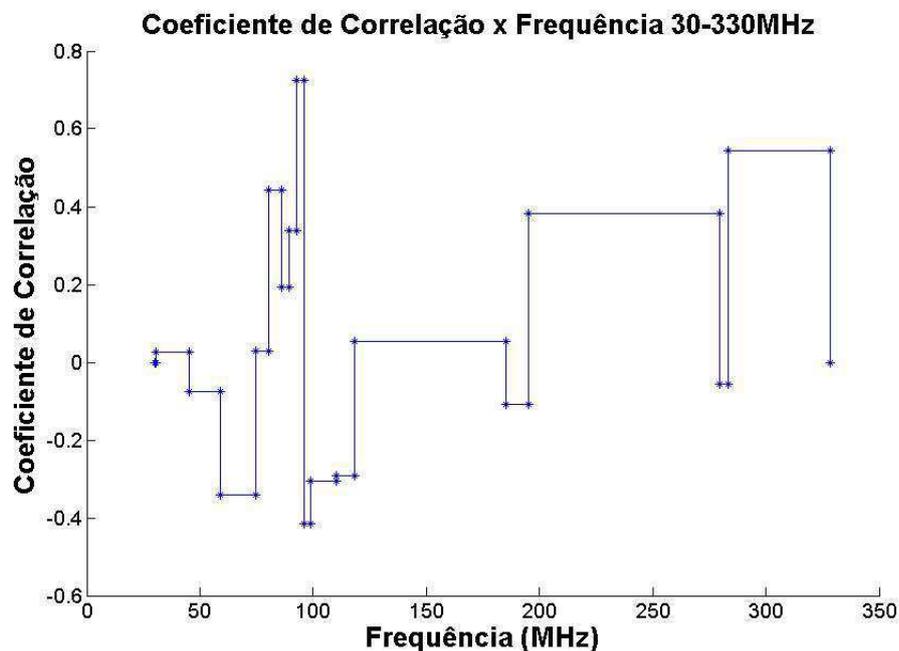
Horizontal Sub-faixa (MHz)	Vertical Sub-faixa (MHz)
240 - 250	250 - 260
260 - 270	290 - 320

**Tabela 7** – Sub-faixas onde o sinal é captado apenas na polarização vertical ou horizontal, na faixa de 230 MHz até 330 MHz

### 3.3 Análise dos Conjuntos III, IV e V

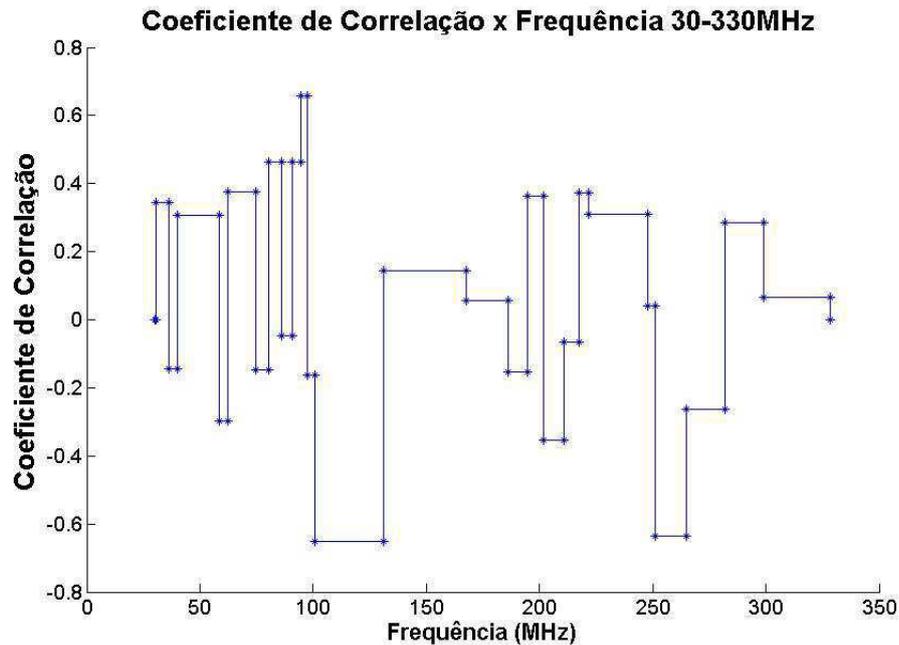
Nesta seção são analisados os conjuntos III, IV e V. Os dados dos conjuntos III e IV foram obtidos utilizando-se a antena bicônica nas polarizações horizontal e vertical respectivamente, enquanto que os dados do conjunto V foram obtidos utilizando-se o sensor 3D. Durante as medições com o sensor 3D apenas os dados do isolador poluído foram armazenados, dessa forma, qualquer comparação com o objetivo de classificação do estado de conservação do isolador utilizando o sensor 3D torna-se inviável. Sendo assim, utilizaremos os dados dos isoladores poluídos dos conjuntos III, IV e V para averiguar qual das polarizações é mais viável (com menor rejeição do sinal emitido pelos isoladores).

Iniciamos a análise observando o gráfico do coeficiente de correlação *versus* frequência entre os dados dos isoladores poluídos (Conjuntos III e V e Conjuntos IV e V), na faixa que vai de 30 MHz a 330 MHz, como apresentado nas figura 37 e 38. É possível notar faixas de frequência que apresentam baixa, média e alta correlação. As faixas que apresentam baixa correlação, indicam que o sinal emitido pelo isolador sujo não foi captado pela antena bicônica na polarização considerada, enquanto que, as faixas de frequência que apresentam média e alta correlação indicam que o sinal emitido pelo isolador sujo foi captado pela antena bicônica na polarização considerada.



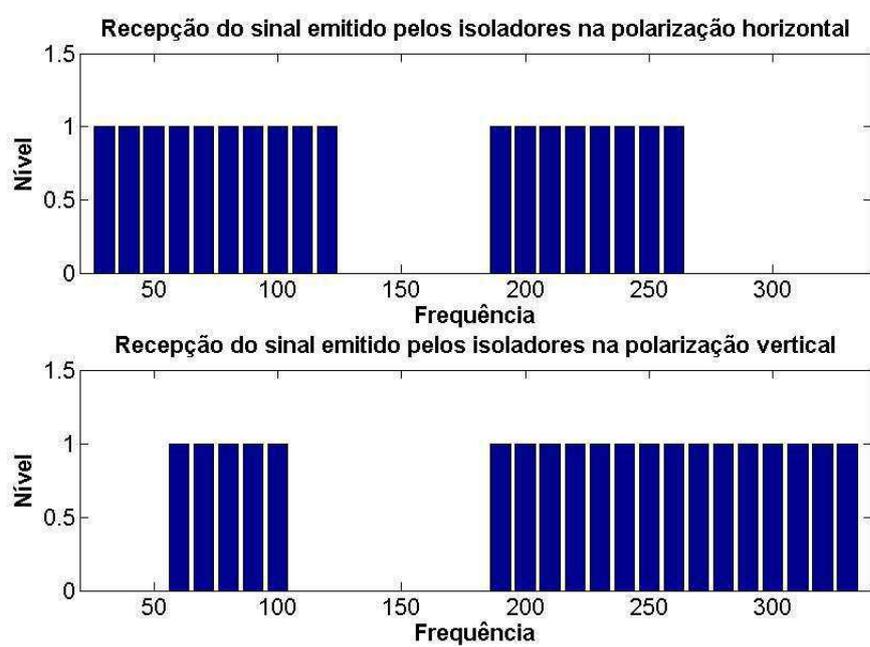
**Figura 37** – Coeficiente de correlação entre os dados dos isoladores poluídos (Conjunto III e V) na faixa de 30 MHz a 330 MHz

A figura 39 mostra em quais faixas de frequência o sinal irradiado pelos isoladores foi recebido. O eixo vertical corresponde ao nível de recepção, isto é, nível 0 significa que nenhum sinal irradiado pelos isoladores foi recebido pela antena e o nível 1 significa que o sinal



**Figura 38** – Coeficiente de correlação entre os dados dos isoladores poluídos (Conjunto IV e V) na faixa de 30 MHz a 330 MHz

irradiado pelos isoladores foi recebido pela antena. É possível observar também, que na faixa de 60 MHz a 100 MHz e na faixa de 190 MHz a 260 MHz, os sinais emitidos pelos isoladores são captados pelas polarizações horizontal e vertical. Porém, ao se observar a faixa de 30 MHz a 60 MHz, verifica-se que o sinal irradiado pelos isoladores é recebido apenas pela antena na polarização horizontal. Já na faixa que vai de 260 MHz a 330 MHz, o sinal irradiado pelos isoladores é captado apenas na polarização vertical. Estes resultados indicam que utilizar a antena na polarização vertical é mais vantajoso do que utilizar a antena na polarização horizontal visto que apenas uma faixa de largura 30 MHz não é coberta pela mesma. Utilizando a antena na polarização horizontal, 70 MHz de largura de faixa do sinal irradiado pelo isolador deixa de ser medido. De fato o uso de uma das polarizações permite negligenciar boa parte do sinal emitido pelos isoladores, porém o uso do sensor 3D permitiu captar os sinais emitidos pelos isoladores nas polarizações horizontal, vertical e azimutal, permitindo comparar com as outras polarizações mencionadas.



**Figura 39** – Faixas onde há sinal irradiado pelos isoladores nas polarizações horizontal e vertical na faixa de 30 MHz a 330 MHz

# Conclusões

Apresentamos um método estatístico relativamente simples para distinguir o estado de conservação (limpo ou poluído) através do sinal irradiado pelos isoladores de vidro quando submetidos à altas tensões. Para o primeiro conjunto de dados (Conjunto I) a análise por histogramas pode diferenciar os sinais emitidos pelos isoladores por meio de suas respectivas variâncias. Em toda faixa de frequência analisada (30 MHz - 330 MHz) não ocorreram sobreposições consideráveis do espectro irradiado pelos isoladores, fato que contribui para utilização do método na distinção do estado de conservação dos isoladores.

A utilização de outros métodos estatísticos como a análise por correlação e o teste de normalidade, contribui para identificação de faixas de frequência específicas que contenham a assinatura espectral dos isoladores. Durante a análise gráfica do coeficiente de correlação *versus* frequência entre os dados do conjunto I e do conjunto II, notou-se que em determinadas faixas de frequência a informação está contida tanto na polarização vertical como na polarização horizontal, como nas faixas de 30 MHz - 60 MHz, 130 MHz - 150 MHz, 160 MHz - 170 MHz, 220 MHz - 240 MHz e 320 MHz - 330 MHz (Conjuntos I e II); contudo, nas faixas de frequências não citadas há informações contidas apenas em uma das polarizações (horizontal ou vertical).

Para verificar qual das polarizações recebe maior parte do sinal irradiado pelos isoladores, utilizou-se como alternativa um sensor 3D. Este sensor é capaz de captar o sinal nas três polarizações:  $x$ ,  $y$  e  $z$ , de maneira que qualquer informação que a antena bicônica polarizada verticalmente ou horizontalmente não consiga recebida, será captada pelo sensor 3D.

Analisando o gráfico do coeficiente de correlação *versus* frequência entre os dados dos isoladores poluídos (Conjuntos III, IV e V), foi possível notar que na faixa de 60 MHz a 100 MHz e na faixa de 190 MHz a 260 MHz, os sinais emitidos pelos isoladores são captados pelas polarizações horizontal e vertical, isto é, a escolha da polarização é indiferente. Porém, ao observar a faixa de 30 MHz a 60 MHz, o sinal emitido pelos isoladores é captado apenas utilizando a antena na polarização horizontal. Já na faixa que vai de 260 MHz a 330 MHz o sinal emitido pelos isoladores é captado apenas na polarização vertical. Nota-se também que na faixa de 130 MHz a 180 MHz não há sinal recebido pela antena nas polarizações vertical e horizontal, ou seja, deve-se evitar fazer medições nesta faixa, pois não haverá sinal irradiado pelos isoladores. Diante disso, deve-se considerar que o uso da antena bicônica na polarização vertical ou horizontal acarreta em determinadas faixas leva a rejeição de do sinal, sendo assim, o emprego do sensor 3D é justificável, pois consegue captar todo o sinal emitido pelos isoladores ao longo da faixa considerada. Comparando as polarizações horizontal e vertical, nota-se com os resultados que utilizar a antena na polarização vertical é mais vantajoso do que utilizar a antena na polarização vertical visto que apenas uma faixa de largura 40 MHz (30 MHz - 40 MHz e 110

MHz - 120 MHz) não é coberta pela mesma, já ao utilizar a antena na polarização horizontal 70 MHz (270 MHz - 330 MHz) de largura de faixa do sinal emitido pelo isolador deixa de ser medido.

## Referências

- [1] AITKEN, A.C., “**On least squares and linear combination of observations**”, Proc. Roy. Soc. Edin. 55, 42-8, 1935.
- [2] ALDRICH, J. H., “**Correlations Genuine and Spurious in Pearson and Yule**”, Statistical Science, pag. 364-376, 1995.
- [3] ALTMAN, D. G., “**Statistics notes: measurement error**”, BMJ Journal, vol. 313, pag. 364-376, 1996.
- [4] ANCOMBE, F. J., “**Graphs in Statistical Analysis**”, The American Statistician, 27: 17-21., 1973.
- [5] ASHENAYI, K.; SINGH, S.; HOBALLAH, I, “**Application of normal distribution in modeling global irradiation**”, Proceedings of the Twentieth Southeastern Symposium, vol., no., pp.470, 474, 1988.
- [6] ASUERO, A. G.; SAYAGO, A.; GONZALEZ, A. G, “**The Correlation Coefficient: An Overview**”, Critical Reviews in Analytical Chemistry, 36:41-59, 2006.
- [7] BALANIS, C. A., “**Antennas Theory: Analisis and Design**”. 3. ed., Wiley, New York, 2005. 45 p.
- [8] BARTKOWIAK, A., “**Should Normal Distribution be Normal? The Student’s T Alternative**”, Computer Information Systems and Industrial Management Applications, 36:41-59, 2006.2007. CISIM '07. 6th International Conference on , vol., no., pp.3,8, 28-30 June 2007.
- [9] BENDAT, J. S.; PIERSOL, A. G, “**Random Data: Analysis and Measurement Procedures**”, Wiley, New York, p. 371, 1971.
- [10] BRYANT, E. C., “**Statistical analysis**”, New York: McGraw-Hill Book, 1960. 303p.
- [11] BUNCHAFT, G.; KELLNER, S. R., “**Estatística sem mistérios**”, 2.ed. Petrópolis: Vozes, 1999, v.2, 303p.
- [12] CECIRE, K., “**Histograms: Construction, Analysis and Understanding**”, 2002. Disponível em: <<http://quarknet.fnal.gov/toolkits/new/histograms.html>>. Acesso em: 01 ago. 2014.
- [13] CHOK, N. S., “**Pearson’s versus Spearman’s and Kendall’s Correlation Coefficients for Continuous Data**”, M.Sc. Thesis, University of Pittsburgh, 53 pp.
- [14] COHEN, J., “**Statistical power analysis for the behavioral sciences**”, Hillsdale, NJ, Erlbaum, 1988, p. 273-406.
- [15] DONALD, E. K., “**The Art of Computer Programming**”, vol. 2: Seminumerical Algo-

rithms, 1998, 3<sup>a</sup> ed, pag.232 Boston: Addison-Wesley.

[16] ELLISON, A. M., “**Exploratory data analysis and graphic display: design and analysis of ecological experiments**”, New York: Chapman & Hall, 1993. pag. 14-41.

[17] GARSON, G. D., “**Statnotes: Topics in Multivariate Analysis**”, Disponível em: <<http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm>>. Acesso em: 14 jul. 2014.

[18] HART, G. C., “**Uncertainty analysis of loads, and safety in structural engineering**”, New Jersey: Prentice-Hall, 1982, 240 p.

[19] JOHNSON, R. A.; WICHERN, D. W., “**Princípios de Análise Multivariada**”, 2. ed. New Jersey: Prentice Hall International, 1988. 607p.

[20] KOZAK, M., “**What is strong correlation?**”, Teaching Statistics, pag. 85-86, 2009.

[21] KRZANOWSKI, W. J., “**Applied multivariate statistical analysis**”, Chap. 7.1, Oxford University Press, New York, 1988.

[22] LIRA, S. A., “**Análise de correlação: abordagem teórica e de construção dos coeficientes com aplicação**”, 2004. 196f. Dissertação (Mestrado) - Setor de Ciências Exatas e Tecnologia, Universidade Federal do Paraná, Curitiba.

[23] MENDES, M.; PALA, A., “**Typer I error rate and power of three normality tests**”, Parkistan Journal of information and tecnologia 2(2), pp; 135-139.

[24] MOORE, D. S.; McCABE, G., “**Introduction to the practice of statistics**”, New York, Freeman.

[25] NAVE, M. R., “**Clasificación de la Polarización**”. 2005. Disponível em: <<http://hyperphysics.phy-astr.gsu.edu/hbasees/phyopt/polclas.html>>. Acesso em: 23 ago. 2014.

[26] OFLYNN, M.; McCABE, G., “**Probabilities, Random Variables, and Random Processes**”, Harper & Row, New York, 1982.

[27] OSBORNE, J.; WATERS, E., “**Four assumptions of multiple regression that researchers should always test**”, Assessment, Research & Evaluation, 8, Disponível em: <<http://PAREonline.net/getvn.asp?v=8&n=2>>. Acesso em: 20 jul. 2014.

[28] PAPOULIS, A., “**Probability, Random Variables, and Stochastic Processes**”, New York: McGraw-Hill, p. pp. 139–152.

[29] PESSOA, G., “**Série Ferramentas de Gestão: Diagrama de Dispersão**, Disponível em: <<http://gerisval.blogspot.com.br/2011/01/serie-ferramentas-de-gestao-diagrama-de.html>>”. Acesso em: 01 ago. 2014.

[30] PIRES, E. S., “**Análise, Projeto e Concepção de um Elemento Irradiador Quase-isotrópico em 3D**”. 2006. Dissertação (Mestrado) - Área de Microondas e Eletromagnetismo Aplicados, Universidade Federal de Campina Grande, Paraíba.

- [31] PORTNOI, M., “**Probabilidade, Variáveis Aleatórias, Distribuição de Probabilidades e Geração Aleatória: Conceitos sob a ótica de Avaliação de Desempenho de Sistemas**”, 2010. Disponível em: <[http://www.eecis.udel.edu/~portnoi/classroom/support\\_material/probabilidade\\_va\\_geracao\\_aleatoria.pdf](http://www.eecis.udel.edu/~portnoi/classroom/support_material/probabilidade_va_geracao_aleatoria.pdf)>. Acesso em: 28 jul. 2014.
- [32] RICHARD, W. H., “**Random variables, mean and the expected value: The art of probability for scientists and engineers**”, Addison-Wesley. p. 64 ff. ISBN 0-201-40686-1.
- [33] ROYSTON, J. P., “**An Extension of Shapiro and Wilk’s W Test for Normality to Large Samples**”, Applied Statistics, 31, pp. 115-124, 1982a.
- [34] ROYSTON, J. P., “**Algorithm AS 177: expected Normal Order Statistics (Exact and Approximate)**”, Applied Statistics, 31, pp. 161-165, 1982b.
- [35] ROYSTON, J. P., “**Algorithm AS 181: The W test for normality**”, Applied Statistics, 31, pp. 176-180, 1982c.
- [36] ROYSTON, J. P., “**Approximating the Shapiro-Wilk W test for Non-normality**” Statistics and Computing, 2, pp. 177-119, 1992.
- [37] ROYSTON, J. P., “**Remark AS R94: A Remark on Algorithm AS181: The W-test for Normality**” Journal of the Royal Statistical Society, Vol. 44, No. 4, pp. 547-551, 1995.
- [38] SCHIELD, M., “**Correlation, Determination And Causality In Introductory Statistics**”, American Statistical Association, Section on Statistical Education, 1995.
- [39] SHAPIRO, S. S.; WILK, M. B., “**A comparative study of various tests for normality**” Journal of the American Statistical Association. Alexandria, v. 63, n. 324, p. 1343-1372, Dec. 1968.
- [40] SHELDON, M. R., “**Introduction to probability models**”, Chap. 2.4, (9th ed.). Academic Press. p. 38 ff. ISBN 0-12-598062-0, 2007.
- [41] SIEGEL, R., HOWELL, J. R., 2002. “**Thermal radiation heat transfer**”, Taylor & Francis, 4<sup>a</sup> ed., New York.
- [42] SZILÁGYI, A.; HIRANO, T.; PY4ZBZ. “**POLARIZAÇÃO DE ONDAS**”. 2008. Disponível em: <<http://www.qsl.net/py4zbz/antenas/polarizacao.htm>>. Acesso em: 20 ago. 2014.
- [43] TAYLOR, R., “**Interpretation of Correlation Coefficient: A Basic Review**”, JMSD 1:35-39, Jan/Feb. 1990.

Anexos



```
42 %QUANDO UTILIZAR A ANTENA 3D, DESCOMENTAR O TRECHO ABAIXO
43 %Arquivos da ANTENA 3D
44 %for a = 1:length(y)
45 %     y_eixo_x(a,1) = y(a,1);
46 %     y_eixo_y(a,1) = y(a,6);
47 %
48 %end
49 %%%CORRECAO%%
50 %y_eixo_x1 = (1e6).*y_eixo_x;
51
52 %ALTERANDO A ESCALA DE 6001 PONTOS PARA 601 PONTOS
53 contador = 1;
54 b = 1;
55 while contador <= 6001
56
57     x_novo_eixo_x(b,1) = x_eixo_x{1,1}(contador,1);
58     x_novo_eixo_y(b,1) = x_eixo_y{1,1}(contador,1);
59
60     b = b + 1;
61     contador = contador + 10;
62 end
63
64
65 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
66
67
68 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
69 %IMPORTANTANDO OS DADOS DO ISOLADOR SUJO
70
71 n2 = length(y);
72
73 for K = 1:n2
74     importa_dados(y{K});
75
76     %DADOS EM dBm
77     y_eixo_x{K,1} = data(:,1);
78     y_eixo_y{K} = data(:,2);
79 end
80
81 contador = 1;
82 b = 1;
83
84 %ALTERANDO A ESCALA DE 6001 PONTOS PARA 601 PONTOS
85 while contador <= 6001
86
87     y_novo_eixo_x(b,1) = y_eixo_x{1,1}(contador,1);
88     y_novo_eixo_y(b,1) = y_eixo_y{1,1}(contador,1);
```

```
89
90     b = b + 1;
91     contador = contador + 10;
92 end
93
94
95 #####TESTE NORMALIDADE#####
96
97 normalidade_e_correlacao(x_novo_eixo_y, y_novo_eixo_y);
98
99 #####
100
101 ##### GRAFICO HISTOGRAMA DADOS ANTIGOS 18/09/2012 #####
102
103     figure (12),
104     hist(y_novo_eixo_y(1:601,1),100),
105     hold on,
106     hist(x_novo_eixo_y(1:601,1),100),
107     xlabel('Amplitude (dBm)'),
108     ylabel('Numero de Ocorrencias'),
109     legend('Sujo', 'Limpo', 'Location', 'SouthEast');
110     hold off;
111
112 #####
113
114 ##### GRAFICO DO ESPECTRO #####
115
116
117     figure(15),
118     plot(x_novo_eixo_x, x_novo_eixo_y, '-g', y_novo_eixo_x, ...
119         y_novo_eixo_y, '-r'),
120     xlabel('Frequencia (Hz)'),
121     ylabel('Amplitude (dBm)'),
122     title('Isoladores com ruido de fundo'),
123     legend('Limpo', 'Sujo', 'Location', 'SouthEast');
124
125 #####
```

## ANEXO B – Script importa\_dados.m

```

1 function importa_dados (Nome_do_arquivo)
2 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
3 %%O script importa_dados.m eh responsavel por carregar todos os dados
4 %%presentes nos arquivos selecionados na workspace do MATLAB.
5 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
6
7 %O ARQUIVO ESPECIFICO DEVE CONTER UM BLOCO DE DADOS, COM OU SEM POSICOES
8
9
10 %EH INTERPRETADO COMO UM SEPARADOR DE COLUNAS EM UM ARQUIVO
11 %DEPENDENDO DA NATUREZA DO ARQUIVO ESTE PARAMETRO DEVE SER ALTERADO
12 DELIMITER = ',';
13
14 %NUMERO DE LINHAS QUE REPRESENTA O CABECALHO DE TEXTO
15 %DEPENDENDO DA NATUREZA DO ARQUIVO ESTE PARAMETRO DEVE SER ALTERADO
16 HEADERLINES = 17;
17
18 %CARREGA DADOS DE Nome_do_arquivo
19 %importdata RETORNA UMA MATRIZ COM OS DADOS DO ARQUIVO
20 %RETORNA UMA MATRIZ DE CELULAS COM OS DADOS TEXTUAIS (IGNORADOS PELOS
21 %DELIMITADORES
22 %ALEM DE UMA MATRIZ DE CELULAR COM O TIPO DE DADOS
23
24 newData1 = importdata(Nome_do_arquivo, DELIMITER, HEADERLINES);
25
26 %RETORNA UMA MATRIZ DE CELULAS CONTENDO OS NOMES DOS CAMPOS QUE SAO
27 %RETORNADOS POR importdata
28
29 vars = fieldnames(newData1);
30
31 assignin('base', vars{1}, newData1.(vars{1}));
32
33 end

```

# ANEXO C – Script

## normalidade\_\_e\_\_correlacao.m

```

1
2 function normalidade_e_correlacao(sinal_x, sinal_y)
3 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
4 %% O script normalidade_e_correlacao() e responsavel pelos teste de      %%
5 %% normalidade de SW, pelo calculo do coeficiente de correlacao,          %%
6 %% eliminacao de picos indesejados e diagnostico da faixa                %%
7 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
8
9 %#PARAMETROS INICIAIS
10 n = 1;
11 k = 1;
12 linha = 2;
13 lim_inferior(1,1) = 1;
14
15 while k+4 < 602
16
17
18     a = sinal_x(n:4+k,1);
19     b = sinal_y(n:4+k,1);
20
21     %APLICANDO O TESTE DE NORMALIDADE
22     h1 = swtest(a);
23     h2 = swtest(b);
24
25     %VERIFICACAO DAS CONDICAOES NECESSARIAS PARA O CALCULO DO COEFICIENTE DE
26     %CORRELACAO DE PEARSON
27
28     if (h1 == 0 && h2 == 0)
29         k = k+1;
30
31     else if (h1 == 1 && h2 == 0)
32         r = corrcoef(a,b);
33         g(linha,1) = r(2,1);
34         %%Conjunto de amostras
35         lim_inferior(linha,1) = k;
36         n = k + 6;
37         k = n;
38         linha = linha + 1;
39

```

```

40     else if(h1 == 0 && h2 == 1)
41         r = corrcoef(a,b);
42         g(linha,1) = r(2,1);
43         %%Conjunto de amostras
44         lim_inferior(linha,1) = k;
45         n = k + 6;
46         k = n;
47         linha = linha + 1;
48
49     else if(h1 == 1 && h2 == 1)
50         r = corrcoef(a,b);
51         g(linha,1) = r(2,1);
52         %%Conjunto de amostras
53         lim_inferior(linha,1) = k;
54         n = k + 6;
55         k = n;
56         linha = linha + 1;
57
58         end
59     end
60 end
61 end
62
63 end
64
65 if (k > 596)
66     lim_inferior(linha,1) = k;
67 end
68
69
70 %%CONVERTENDO AS AMOSTRAS EM SUAS RESPECTIVAS FREQUENCIAS
71 %%SE TEMOS 2001 PONTOS DIVIDIR POR 20, POIS SAO NECESSARIOS 20 PONTOS PARA
72 %1 MHz
73 %%SE TEMOS 601 PONTOS DIVIDIR POR 2, POIS SAO NECESSARIOS 2 PONTOS PARA
74 %1 MHz
75
76 for d = 1:length(lim_inferior)-1
77     fr(d,1) = (lim_inferior(d+1,1) - lim_inferior(d,1))/2;
78 end
79
80 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
81
82 %%VERIFICANDO SE A SOMA ESTA DE ACORDO COM O FAIXA CONSIDERADA
83 soma = sum(fr)
84
85 %##### ANALISE DE PICOS #####
86

```

```
87 contador = 0;
88 faixa_de_picos = 0;
89
90 for t = 1:length(fr)
91     if (g(t, 1) > 0.2 || g(t,1) < -0.2)
92         faixa_de_picos = faixa_de_picos + fr(t,1);
93         contador = contador + 1;
94     end
95 end
96
97 m = faixa_de_picos/contador;
98
99 for r = 1:length(fr)
100     if ((g(r,1) > 0.2 || g(r,1) < -0.2) && fr(r, 1) <= 0.50)
101         g(r,1) = 0;
102     end
103 end
104
105
106 ##### CLASSIFICACAO DE FAIXAS #####
107
108 %Nivel de Classificacao:
109 %Nivel 1 - Fraco
110 %Nivel 2 - Moderado
111 %Nivel 3 - Forte
112
113 %ORGANIZANDO AS OS VETORES PARA ANALISE
114
115 soma = 0;
116 b = 1;
117 c = 1;
118 aux2 = 0;
119
120
121 while b < length(fr)
122     soma = aux2 + soma + fr(b,1);
123
124     if(soma >= 10)
125         aux = soma - fr(b,1);
126         if soma >= 20
127             fr1(c,1) = abs(aux -20);
128         else
129             fr1(c,1) = abs(aux - 10);
130         end
131
132         if (soma - 10 > 10)
133             fr1(c+1,1) = 10;
```

```
134     else
135         fr1(c+1,1) = soma - 10;
136     end
137
138     gmod(c,1) = g(b,1);
139     gmod(c+1,1) = g(b,1);
140     aux2 = soma - 10;
141
142
143     soma = 0;
144
145
146     c = c + 1;
147     else
148         if (soma < 10)
149             gmod(c,1) = g(b,1);
150             fr1(c,1) = fr(b,1);
151             aux2 = 0;
152         end
153         %
154     end
155
156     b = b + 1;
157     c = c + 1;
158 end
159
160 soma_2 = 0;
161 h = 1;
162 z = 1;
163
164 comp = length(fr1);
165
166 for k = 1:length(gmod)
167     soma_2 = soma_2 + fr1(k,1);
168
169     if (soma_2 >= 10.0)
170         soma_2 = 0;
171         coordenada{h,1} = z:k;
172         h = h + 1;
173         z = k+1;
174
175     end
176 end
177
178 if (k == length(gmod))
179     coordenada{h,1} = (z-1):k;
180
```

```
181 end
182
183 if (h < 10)
184     h = h +1;
185     coordenada{h,1} = coordenada{h-1,1};
186 end
187
188 Coef_diag(gmod,fr1, coordenada)
189
190 #####
191 #####Contrucao Grafica#####
192
193
194
195 f1 = 30; %Frequencia Inicial
196
197
198 figure (66),
199 hold on,
200
201 L{1,1}(1,1) = f1;
202 g(length(fr)+1,1) = 0;
203
204 for s = 1:length(fr)
205     f1 = f1 + fr(s,1);
206     L{1,1}(s+1,1) = f1;
207 end
208 soma = sum(fr)
209
210
211 stairs(L{1,1}, g, '-*b');
212
213 title('Coeficiente de Correlacao x Frequencia 30-330MHz '),
214 ylabel('Coeficiente de Correlacao'),
215 xlabel('Frequencia (MHz)'),
216 hold off;
217
218 end
```

## ANEXO D – Script sw\_test.m

```

1 function [H, pValue, W] = sw_test(x, alpha, tail)
2 %SWTEST Shapiro–Wilk parametric hypothesis test of composite normality.
3 % [H, pValue, SWstatistic] = SWTEST(X, ALPHA, TAIL) performs
4 % the Shapiro–Wilk test to determine if the null hypothesis of
5 % composite normality is a reasonable assumption regarding the
6 % population distribution of a random sample X. The desired significance
7 % level, ALPHA, is an optional scalar input (default = 0.05).
8 % TAIL indicates the type of test (default = 1).
9 %
10 % The Shapiro–Wilk hypotheses are:
11 % Null Hypothesis: X is normal with unspecified mean and variance.
12 % For TAIL = 0 (2–sided test), alternative: X is not normal.
13 % For TAIL = 1 (1–sided test), alternative: X is upper the normal.
14 % For TAIL = -1 (1–sided test), alternative: X is lower the normal.
15 %
16 % This is an omnibus test, and is generally considered relatively
17 % powerful against a variety of alternatives.
18 % Shapiro–Wilk test is better than the Shapiro–Francia test for
19 % Platykurtic sample. Conversely, Shapiro–Francia test is better than the
20 % Shapiro–Wilk test for Leptokurtic samples.
21 %
22 % When the series 'X' is Leptokurtic, SWTEST performs the Shapiro–Francia
23 % test, else (series 'X' is Platykurtic) SWTEST performs the
24 % Shapiro–Wilk test.
25 %
26 % [H, pValue, SWstatistic] = SWTEST(X, ALPHA, TAIL)
27 %
28 % Inputs:
29 % X – a vector of deviates from an unknown distribution. The observation
30 % number must exceed 3 and less than 5000.
31 %
32 % Optional inputs:
33 % ALPHA – The significance level for the test (default = 0.05).
34 %
35 % TAIL – The type of the test (default = 1).
36 %
37 % Outputs:
38 % SWstatistic – The test statistic (non normalized).
39 %
40 % pValue – is the p–value, or the probability of observing the given
41 % result by chance given that the null hypothesis is true. Small values

```

```
42 %   of pValue cast doubt on the validity of the null hypothesis.
43 %
44 %   H = 0 => Do not reject the null hypothesis at significance level ALPHA.
45 %   H = 1 => Reject the null hypothesis at significance level ALPHA.
46 %
47
48 %
49 % References: Royston P. "Algorithm AS R94", Applied Statistics (1995)
50 % Vol. 44, No. 4.
51 % AS R94 — calculates Shapiro–Wilk normality test and P-value
52 % for sample sizes 3 <= n <= 5000. Handles censored or uncensored data.
53 % Corrects AS 181, which was found to be inaccurate for n > 50.
54 %
55
56 %
57 % Ensure the sample data is a VECTOR.
58 %
59
60 if numel(x) == length(x)
61     x = x(:);           % Ensure a column vector.
62 else
63     error(' Input sample 'X' must be a vector. ');
64 end
65
66 %
67 % Remove missing observations indicated by NaN's and check sample size.
68 %
69
70 x = x(~isnan(x));
71
72 if length(x) < 3
73     error(' Sample vector 'X' must have at least 3 valid observations. ');
74 end
75
76 if length(x) > 5000
77     warning('Shapiro–Wilk test might be inaccurate due to large sample ...
78             size ( > 5000). ');
79 end
80
81 %
82 % Ensure the significance level, ALPHA, is a
83 % scalar, and set default if necessary.
84 %
85
86 if (nargin >= 2) && ~isempty(alpha)
87     if numel(alpha) > 1
88         error(' Significance level 'Alpha' must be a scalar. ');
```

```

89     end
90     if (alpha <= 0 || alpha >= 1)
91         error(' Significance level ''Alpha'' must be between 0 and 1.');
```

---

```

92     end
93 else
94     alpha = 0.05;
95 end
96
97 %
98 % Ensure the type-of-test indicator, TAIL, is a scalar integer from
99 % the allowable set {-1 , 0 , 1}, and set default if necessary.
100 %
101
102 if (nargin >= 3) && ~isempty(tail)
103     if numel(tail) > 1
104         error('Type-of-test indicator ''Tail'' must be a scalar.');
```

---

```

105     end
106     if (tail ~= -1) && (tail ~= 0) && (tail ~= 1)
107         error('Type-of-test indicator ''Tail'' must be -1, 0, or 1.');
```

---

```

108     end
109 else
110     tail = 1;
111 end
112
113 % First, calculate the a's for weights as a function of the m's
114 % See Royston (1995) for details in the approximation.
115
116 x      = sort(x); % Sort the vector X in ascending order.
117 n      = length(x);
118 mtilde = norminv(((1:n)' - 3/8) / (n + 0.25));
119 weights = zeros(n,1); % Preallocate the weights.
120
121 if kurtosis(x) > 3
122
123     % The Shapiro-Francia test is better for leptokurtic samples.
124
125     weights = 1/sqrt(mtilde'*mtilde) * mtilde;
126
127     %
128     % The Shapiro-Francia statistic W is calculated to avoid excessive
129     % rounding errors for W close to 1 (a potential problem in very large
130     % samples).
131     %
132
133     W = (weights' * x) ^2 / ((x - mean(x))' * (x - mean(x)));
134
135     nu = log(n);
```

```

136     u1      =   log(nu) - nu;
137     u2      =   log(nu) + 2/nu;
138     mu      =   -1.2725 + (1.0521 * u1);
139     sigma   =   1.0308 - (0.26758 * u2);
140
141     newSFstatistic =   log(1 - W);
142
143     %
144     % Compute the normalized Shapiro-Francia statistic and its p-value.
145     %
146
147     NormalSFstatistic =   (newSFstatistic - mu) / sigma;
148
149     % the next p-value is for the tail = 1 test.
150     pValue   =   1 - normcdf(NormalSFstatistic, 0, 1);
151
152 else
153
154     % The Shapiro-Wilk test is better for platykurtic samples.
155
156     c      =   1/sqrt(mtilde'*mtilde) * mtilde;
157     u      =   1/sqrt(n);
158
159     PolyCoef_1   =   [-2.706056 , 4.434685 , -2.071190 , -0.147981 , ...
160         0.221157 , c(n)];
161     PolyCoef_2   =   [-3.582633 , 5.682633 , -1.752461 , -0.293762 , ...
162         0.042981 , c(n-1)];
163
164     PolyCoef_3   =   [-0.0006714 , 0.0250540 , -0.39978 , 0.54400];
165     PolyCoef_4   =   [-0.0020322 , 0.0627670 , -0.77857 , 1.38220];
166     PolyCoef_5   =   [0.00389150 , -0.083751 , -0.31082 , -1.5861];
167     PolyCoef_6   =   [0.00303020 , -0.082676 , -0.48030];
168
169     PolyCoef_7   =   [0.459 , -2.273];
170
171     weights(n)   =   polyval(PolyCoef_1 , u);
172     weights(1)   =   -weights(n);
173
174     % Special attention when n=3 (this is a special case).
175     if n == 3
176         weights(1)   =   0.707106781;
177         weights(n)   =   -weights(1);
178     end
179
180     if n >= 6
181         weights(n-1) =   polyval(PolyCoef_2 , u);
182         weights(2)   =   -weights(n-1);

```

```

183
184     count = 3;
185     phi = (mtilde'*mtilde - 2 * mtilde(n)^2 - 2 * mtilde(n-1)^2)/...
186           (1 - 2 * weights(n)^2 - 2 * weights(n-1)^2);
187 else
188     count = 2;
189     phi = (mtilde'*mtilde - 2 * mtilde(n)^2) / ...
190           (1 - 2 * weights(n)^2);
191 end
192
193 %
194 % The vector 'WEIGHTS' obtained next corresponds to the same
195 % coefficients listed by Shapiro-Wilk in their original test
196 % for small samples.
197 %
198
199 weights(count : n-count+1) = mtilde(count : n-count+1) / sqrt(phi);
200
201 %
202 % The Shapiro-Wilk statistic W is calculated to avoid excessive
203 % rounding errors for W close to 1 (a potential problem in very large
204 % samples).
205 %
206
207 W = (weights' * x) ^2 / ((x - mean(x))' * (x - mean(x)));
208
209 %
210 % Calculate the significance level for W (exact for n=3).
211 %
212
213 newn = log(n);
214
215 if (n > 3) && (n <= 11)
216
217     mu = polyval(PolyCoef_3 , n);
218     sigma = exp(polyval(PolyCoef_4 , n));
219     gam = polyval(PolyCoef_7 , n);
220
221     newSWstatistic = -log(gam-log(1-W));
222
223 elseif n >= 12
224
225     mu = polyval(PolyCoef_5 , newn);
226     sigma = exp(polyval(PolyCoef_6 , newn));
227
228     newSWstatistic = log(1 - W);
229

```

```
230     elseif n == 3
231         mu         = 0;
232         sigma      = 1;
233         newSWstatistic = 0;
234     end
235
236     %
237     % Compute the normalized Shapiro–Wilk statistic and its p-value.
238     %
239
240     NormalSWstatistic = (newSWstatistic - mu) / sigma;
241
242     % The next p-value is for the tail = 1 test.
243     pValue = 1 - normcdf(NormalSWstatistic, 0, 1);
244
245     % Special attention when n=3 (this is a special case).
246     if n == 3
247         pValue = 1.909859 * (asin(sqrt(W)) - 1.047198);
248         NormalSWstatistic = norminv(pValue, 0, 1);
249     end
250
251 end
252
253 % The p-value just found is for the tail = 1 test.
254 if tail == 0
255     pValue = 2 * min(pValue, 1-pValue);
256 elseif tail == -1
257     pValue = 1 - pValue;
258 end
259
260 %
261 % To maintain consistency with existing Statistics Toolbox hypothesis
262 % tests, returning 'H = 0' implies that we 'Do not reject the null
263 % hypothesis at the significance level of alpha' and 'H = 1' implies
264 % that we 'Reject the null hypothesis at significance level of alpha.'
265 %
266
267 H = (alpha >= pValue);
```

## ANEXO E – Script coef\_diag.m

```

1 function coef_diag(gmod,fr1, coordenada)
2
3 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
4 %%Este script classifica cada faixa de frequencia em tres niveis: 1-Fraco;%
5 %%2 – Moderado; e 3 – Forte. Cada um dos tres niveis tem como base o nivel%
6 %%de correlacao pre-definido. As classificacao eh feita a cada 10 MHz.    %
7 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
8
9 %PARAMETROS INICIAIS
10 limitante = 0;
11 gmod;
12 fr1;
13
14 %SEPARANDO OS RESULTADOS PARA CADA FAIXA DE 10 MHz
15 faixa1 = coordenada{1,1}
16 faixa2 = coordenada{2,1}
17 faixa3 = coordenada{3,1}
18 faixa4 = coordenada{4,1}
19 faixa5 = coordenada{5,1}
20 faixa6 = coordenada{6,1}
21 faixa7 = coordenada{7,1}
22 faixa8 = coordenada{8,1}
23 faixa9 = coordenada{9,1}
24 faixa10 = coordenada{10,1}
25
26 cont1 = 0;
27 cont2 = 0;
28
29 % ESTE LOOP FAZ A CLASSIFICACAO E ARMAZENA O RESULTADO EM UM VETOR CHAMADO
30 %vetor_conclusao
31
32 for a = 1:length(coordenada)
33     auxiliar = length(coordenada{a,1})
34
35     for b = coordenada{a,1}(1,1):coordenada{a,1}(1,auxiliar)
36
37         if (gmod(b,1) < +0.3 && gmod(b,1) > -0.3)
38
39             if (b == coordenada{a,1}(1,auxiliar))
40                 vetor_conclusao(a,1) = 1;
41                 break

```

```
42         end
43     else
44         if (gmod(b,1) > +0.6 || gmod(b,1) < -0.6)
45             cont1 = cont1 + 1;
46             limitante = limitante + fr1(b,1);
47
48             if (limitante > 1 && cont1 >= cont2)
49                 vetor_conclusao(a,1) = 3;
50                 limitante = 0;
51                 cont2 = 0;
52                 cont1 = 0;
53                 break
54             else
55                 if (b == coordenada{a,1}(1,auxiliar))
56                     vetor_conclusao(a,1) = 1;
57                     break
58                 end
59             end
60
61     else
62         if ((gmod(b,1)<+0.6&&gmod(b,1)>+0.3) || (gmod(b,1)<-0.3...
63             &&gmod(b,1)>-0.6))
64
65             cont2 = cont2 + 1;
66             limitante = limitante + fr1(b,1);
67
68             if (limitante > 1 && cont2 >= cont1)
69                 vetor_conclusao(a,1) = 2;
70                 limitante = 0;
71                 cont2 = 0;
72                 cont1 = 0;
73                 break
74             else
75                 if (b == coordenada{a,1}(1,auxiliar))
76                     vetor_conclusao(a,1) = 1;
77                     break
78                 end
79             end
80
81         end
82     end
83 end
84 end
85 end
86
87
88 %VETOR COM OS RESULTADOS DA CLASSIFICACAO
```

```
89 vetor = vetor_conclusao(1:10,1)
90
91 ##### COMENTAR AS FAIXAS QUE NAO USAR #####
92 faixas_de_freq = [35,45,55,65,75,85,95,105,115,125];
93 %faixas_de_freq = [135,145,155,165,175,185,195,205,215,225];
94 %faixas_de_freq = [235,245,255,265,275,285,295,305,315,325];
95
96
97 %GRAFICO
98
99 figure(77),
100 colormap(winter(1)),
101 bar(faixas_de_freq, vetor, 1),
102 ylabel('Nivel')
103 xlabel('Frequencia (MHz)'),
104 title('Grau de Correlacao entre os Isoladores LIMPOxSUJO');
105 end
```