



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA METALÚRGICA E DE MATERIAIS
GRADUAÇÃO EM ENGENHARIA METALÚRGICA

RAPHAELLA HERMONT FONSECA MURTA

**ANÁLISE DE AGRUPAMENTO DE SINAIS OBTIDOS POR SIMULAÇÃO DO
ENSAIO ULTRASSÔNICO PELA TÉCNICA TOFD PARA RECONHECIMENTO DE
DEFEITOS DE SOLDAGEM MODELADOS.**

FORTALEZA

2015

RAPHAELLA HERMONT FONSECA MURTA

**ANÁLISE DE AGRUPAMENTO DE SINAIS OBTIDOS POR SIMULAÇÃO DO
ENSAIO ULTRASSÔNICO PELA TÉCNICA TOFD PARA RECONHECIMENTO DE
DEFEITOS DE SOLDAGEM MODELADOS.**

Monografia apresentada ao Curso de Engenharia Metalúrgica do Departamento de Engenharia Metalúrgica e de Materiais da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharel em Engenharia Metalúrgica.

Orientador: Prof. Dr. Elineudo Pinho de Moura

**FORTALEZA
2015**

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca de Ciências e Tecnologia

-
- M965a Murta, Raphaella Hermont Fonseca.
Análise de agrupamento de sinais obtidos por simulação do ensaio ultrassônico pela técnica TOFD para reconhecimento de defeitos de soldagem modelados. / Raphaella Hermont Fonseca Murta. – 2015.
81 f. : il., color.
- Monografia (Graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Departamento de Engenharia Metalúrgica e de Materiais, Curso de Engenharia Metalúrgica, Fortaleza, 2015.
Orientação: Prof. Dr. Elineudo Pinho de Moura
1. Materiais - Testes. 2. Soldagem. 3. Metalurgia. I. Título.

CDD 620.11

RAPHAELLA HERMONT FONSECA MURTA

**ANÁLISE DE AGRUPAMENTO DE SINAIS OBTIDOS POR SIMULAÇÃO DO
ENSAIO ULTRASSÔNICO PELA TÉCNICA TOFD PARA RECONHECIMENTO DE
DEFEITOS DE SOLDAGEM MODELADOS.**

Monografia apresentada ao Curso de Engenharia Metalúrgica do Departamento de Engenharia Metalúrgica e de Materiais da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharel em Engenharia Metalúrgica.

Aprovado em: ____ de _____ de _____.

BANCA EXAMINADORA

Prof. Elineudo Pinho Moura; D. Sc.(Orientador)
Universidade Federal do Ceará (UFC)

Prof. Lindberg Lima Gonçalves; D. Phil.
Universidade Federal do Ceará (UFC)

Filipe Francisco Rocha Damasceno; M.Sc.
Universidade Federal do Ceará (UFC)

A Deus.

À minha família.

AGRADECIMENTOS

Aos meus pais, Ana Lúcia e Breno Murta, por sempre me apoiarem em cada passo da minha vida.

Ao meu orientador, Professor Dr. Elineudo Pinho de Moura, pela oportunidade e pela grande ajuda no desenvolvimento deste trabalho.

Aos integrantes do Centro de Ensaios não Destrutivos - CENDE/UFC, Flávison de Almada e Carlos Spínola, pelo grande apoio na parte programacional.

À minha família, em especial Ariana Salim, Arthur Murta e Thiago Ribeiro, pelo apoio e companheirismo.

Aos meus professores, em especial Natália Barroso, Jorge Brandão e Ernani Ribeiro, por acreditarem em mim e terem me incentivado e ensinado muito durante a graduação.

Aos meus amigos, Isabelle Maria, Larissa Almeida, Nadja Nogueira e Thiago Andrade, por se fazerem presente na minha vida.

Aos meus colegas de curso, em especial Amanda Bonfim, Gabriel Gonczarowska, Giovani Gonçalves e Luiza Guedes, pela amizade, paciência e convivência nesses cinco anos.

"Ne sachant pas que c'était impossible, il y a été
et l'a fait" (Jean Cocteau)

RESUMO

O ensaio não destrutivo por ultrassom é amplamente utilizado para verificar a existência de descontinuidades no interior dos materiais. A técnica TOFD (*Time Of Flight Diffraction* [do inglês]) é uma técnica ultrassônica muito utilizada e que aliada com a técnica *Phased Array* detecta bem pequenas descontinuidades no material. Entretanto, a classificação do tipo de descontinuidade a partir dos sinais ultrassônicos obtidos durante uma inspeção requer um profissional altamente capacitado e, considerando o volume de dados a ser analisado, um longo tempo. Essa tarefa pode ser realizada por algoritmos de classificação e/ou agrupamento capazes de processar rapidamente uma grande quantidade de dados. Neste trabalho, foram simulados com o auxílio do programa COMSOL três tipos de defeitos de soldagem (falta de penetração, poro e trinca). Em seguida foram aplicados dois métodos de reconhecimento de padrões: Análise de componentes principais (PCA) e k-médias, para verificar como esses dados são separados. Este trabalho é um passo inicial para o emprego de simulação numérica do ensaio ultrassônico para produzir sinais simulados e utilizá-los no treinamento de ferramentas de reconhecimento de padrão, a serem posteriormente testados com sinais ultrassônicos experimentais.

Palavras-chave: Ultrassom, TOFD, Simulação, Defeitos de soldagem, PCA, K-médias.

ABSTRACT

The non-destructive testing by ultrasound is widely used to check for discontinuities inside the material. The Time Of Flight Diffraction is an ultrasonic technique widely used and combined with the phased array detects far smaller discontinuities in the material. However, the classification of the type of discontinuity from the ultrasonic signals obtained during an inspection requires a highly skilled professional, and a long time if considering the volume of data to be analyzed. This task may be performed by classification and *clustering* algorithms able to quickly process a large amount of data. In this work was simulated discontinuities in software COMSOL three types of welding defects (lack of penetration, pore and crack); then it was applied two pattern recognition methods: Principal component analysis and the k-means, to check how the data is separated. This work is a first step for the numerical simulation of use of ultrasonic testing to produce simulated signals and use them in pattern recognition tools training, to be later tested with experimental ultrasonic signals.

Keywords: Ultrasound, TOFD, Simulation, welding defects, PCA, k-means

LISTA DE FIGURAS

FIGURA 1 – REPRESENTAÇÃO DE UMA DESCONTINUIDADE DO TIPO FALTA DE PENETRAÇÃO.	19
FIGURA 2 – REPRESENTAÇÃO DE UMA DESCONTINUIDADE DO TIPO PORO.....	19
FIGURA 3 – REPRESENTAÇÃO DE UMA DESCONTINUIDADE DO TIPO TRINCA.	20
FIGURA 4 – CONFIGURAÇÃO DA TÉCNICA TOFD.....	21
FIGURA 5 – REPRESENTAÇÃO DO PHASED ARRAY.....	22
FIGURA 6 – DIRECIONAMENTO DO FEIXE POR PHASED ARRAY.....	22
FIGURA 7 – DIRECIONAMENTO DO FEIXE COM A VARIAÇÃO DO NÚMERO DE TRANSDUTORES.	23
FIGURA 8 – MATRIZ DE DADO DE DIMENSÃO ‘N X P’	27
FIGURA 9 – MATRIZ DE COVARIÂNCIA DE ORDEM ‘P X P’	28
FIGURA 10 – (A) REPRESENTAÇÃO DOS DADOS ANTES DO PCA; (B) REPRESENTAÇÃO DOS DADOS DEPOIS DO PCA.....	29
FIGURA 11 – SIMILARIDADE	30
FIGURA 12 – REPRESENTAÇÃO DO FUNCIONAMENTO DO K-MÉDIAS.	31
FIGURA 13 – FUNÇÃO DE CUSTO DO K-MÉDIAS.....	31
FIGURA 14 – REPRESENTAÇÃO DA RELAÇÃO ENTRE A DISTÂNCIA ENTRE OS TRANSDUTORES E A ESPESSURA DO MATERIA.....	36
FIGURA 15 – REPRESENTAÇÃO DA PEÇA NO COMSOL.....	38
FIGURA 16 – CAMPO DE DESLOCAMENTOS EM UMA PEÇA CONTENDO DEFEITO DO TIPO FALTA DE PENETRAÇÃO (2MM X 4MM).....	40
FIGURA 17 – A-SCAN GERADO DURANTE SIMULAÇÃO DA PROPAGAÇÃO DA ONDA POR UM MEIO CONTENDO DEFEITO DO TIPO FALTA DE PENETRAÇÃO (2MM X 4MM).....	40
FIGURA 18 – CAMPO DE DESLOCAMENTOS EM UMA PEÇA CONTENDO UMA DESCONTINUIDADE DO TIPO PORO (DIÂMETRO = 2MM).....	41

FIGURA 19 – A-SCAN GERADO DURANTE SIMULAÇÃO DA PROPAGAÇÃO DA ONDA POR UM MEIO CONTENDO DEFEITO DO TIPO PORO (DIÂMETRO = 2MM).	41
FIGURA 20 – CAMPO DE DESLOCAMENTOS EM UMA PEÇA CONTENDO DEFEITO DO TIPO TRINCA (0.1MM)	42
FIGURA 21 – A-SCAN GERADO DURANTE SIMULAÇÃO DA PROPAGAÇÃO DA ONDA POR UM MEIO CONTENDO DEFEITO DO TIPO TRINCA (0.1MM X 1.0MM).	42
FIGURA 22 – (A) RESULTADO DO PCA DOS 108 SINAIS ULTRASSÔNICOS SIMULADOS; (B) CLASSIFICAÇÃO CONHECIDA DOS DADOS.	45
FIGURA 23 – (A) APRESENTA A PROJEÇÃO DAS DUAS PRIMEIRAS COMPONENTES PRINCIPAIS DA ENVOLTÓRIA DOS DADOS SIMULADOS; (B) CLASSIFICAÇÃO CONHECIDA DOS DADOS.....	47
FIGURA 24 – HISTOGRAMA DO NÚMERO ÓTIMO DE AGRUPAMENTOS SUGERIDO PARA OS SINAIS ULTRASSÔNICOS SIMULADOS: (A) BASEADO NO ÍNDICE SILHUETA; (B) BASEADO NO ÍNDICE; (C) BASEADO NO MÍNIMO ENTRE OS ÍNDICES	50
FIGURA 25 – (A) RESULTADO DA ANÁLISE DE AGRUPAMENTO; (B) RESULTADO DA PCA; (C) CLASSIFICAÇÃO CONHECIDA DOS SINAIS SIMULADOS.....	51
FIGURA 26 – HISTOGRAMA DO NÚMERO ÓTIMO DE AGRUPAMENTOS SUGERIDO PARA AS ENVOLTÓRIAS DOS SINAIS ULTRASSÔNICOS SIMULADOS: (A) BASEADO NO ÍNDICE SILHUETA; (B) BASEADO NO ÍNDICE DAVIES-BOULDIN; (C) BASEADO NO MÍNIMO ENTRE OS ÍNDICES.	54
FIGURA 27 – (A) RESULTADO DA ANÁLISE DE AGRUPAMENTO; (B) RESULTADO DA PCA; (C) CLASSIFICAÇÃO CONHECIDA DAS ENVOLTÓRIAS DOS SINAIS SIMULADOS.....	55
FIGURA 28 – HISTOGRAMA DO NÚMERO ÓTIMO DE AGRUPAMENTOS SUGERIDO PARA AS 2(DUAS) PRIMEIRAS COMPONENTES PRINCIPAIS DOS SINAIS ULTRASSÔNICOS SIMULADOS:(A) BASEADO NO ÍNDICE SILHUETA;(B) BASEADO NO ÍNDICE DAVIES-BOULDIN;(C) BASEADO NO MÍNIMO ENTRE OS ÍNDICES.....	57
FIGURA 29 – HISTOGRAMA DO NÚMERO ÓTIMO DE AGRUPAMENTOS SUGERIDO PARA AS 20 (VINTE) PRIMEIRAS COMPONENTES PRINCIPAIS DOS SINAIS ULTRASSÔNICOS SIMULADOS: (A) BASEADO NO ÍNDICE SILHUETA; (B) BASEADO NO ÍNDICE DAVIES-BOULDIN; (C) BASEADO NO MÍNIMO ENTRE OS ÍNDICES.....	58

FIGURA 30 – (A) RESULTADO DA ANÁLISE DE AGRUPAMENTO DAS 2 PRIMEIRAS COMPONENTES PRINCIPAIS; (B) RESULTADO DA PCA; (C) CLASSIFICAÇÃO CONHECIDA DOS SINAIS SIMULADOS.....	60
FIGURA 31 – HISTOGRAMA DO NÚMERO ÓTIMO DE AGRUPAMENTOS SUGERIDO PARA AS 2 (DUAS) PRIMEIRAS COMPONENTES PRINCIPAIS DAS ENVOLTÓRIAS DOS SINAIS ULTRASSÔNICOS SIMULADOS.	61
FIGURA 32 – (A) RESULTADO DA ANÁLISE DE AGRUPAMENTO DAS 2 PRIMEIRAS COMPONENTES PRINCIPAIS DAS ENVOLTÓRIAS DOS SINAIS SIMULADOS; (B) RESULTADO DA PCA; (C) CLASSIFICAÇÃO CONHECIDA DAS ENVOLTÓRIAS DOS SINAIS SIMULADOS.	62
FIGURA 33 – HISTOGRAMA DO NÚMERO DE AGRUPAMENTOS SUGERIDOS PARA AS 20 (VINTE) PRIMEIRAS COMPONENTES PRINCIPAIS DAS ENVOLTÓRIAS DOS SINAIS ULTRASSÔNICOS SIMULADOS: (A) BASEADO NO ÍNDICE SILHUETA; (B) BASEADO NO ÍNDICE DAVIES-BOULDIN; (C) BASEADO NO MÍNIMO ENTRE OS ÍNDICES.	64
FIGURA 34 – (A) RESULTADO DA ANÁLISE DE AGRUPAMENTO DAS 20 PRIMEIRAS COMPONENTES PRINCIPAIS DAS ENVOLTÓRIAS DOS SINAIS SIMULADOS; (B) RESULTADO DA PCA; (C) CLASSIFICAÇÃO CONHECIDA DO CONJUNTO FORMADO PELAS ENVOLTÓRIAS DOS SINAIS SIMULADOS.....	65
FIGURA 35 – HISTOGRAMA DO NÚMERO ÓTIMO DE AGRUPAMENTOS SUGERIDO PARA O CONJUNTO DE SINAIS ULTRASSÔNICOS CAPTURADOS EXPERIMENTALMENTE: (A) BASEADO NO ÍNDICE SILHUETA; (B) BASEADO NO ÍNDICE DAVIES-BOULDIN; (C) BASEADO NO MÍNIMO ENTRE OS ÍNDICES.	68
FIGURA 36 - (A) RESULTADO DA ANÁLISE DE AGRUPAMENTO; (B) RESULTADO DA PCA; (C) CLASSIFICAÇÃO CONHECIDA DOS SINAIS EXPERIMENTAIS.	70
FIGURA 37 – HISTOGRAMA DO NÚMERO ÓTIMO DE AGRUPAMENTOS SUGERIDO PARA O CONJUNTO DAS ENVOLTÓRIAS DOS SINAIS ULTRASSÔNICOS CAPTURADOS EXPERIMENTALMENTE: (A) BASEADO NO ÍNDICE SILHUETA; (B) BASEADO NO ÍNDICE DAVIES-BOULDIN; (C) BASEADO NO MÍNIMO ENTRE OS ÍNDICES.....	72
FIGURA 38 – (A) RESULTADO DA ANÁLISE DE AGRUPAMENTO; (B) RESULTADO DA PCA; (C) CLASSIFICAÇÃO CONHECIDA DAS ENVOLTÓRIAS DOS SINAIS EXPERIMENTAIS. ...	73

LISTA DE TABELAS

TABELA 1 – PROPRIEDADES DO MATERIAL	35
TABELA 2 – DEFINIÇÃO DAS VARIÁVEIS DO PROBLEMA.....	37
TABELA 3 – FUNÇÕES RELACIONADAS COM A FONTE	38
TABELA 4 – MATRIZ DE CONFUSÃO OBTIDA PELA APLICAÇÃO DA REGRA DA MÉDIA DA CLASSE MAIS PRÓXIMA CONSIDERANDO APENAS AS 2 PRIMEIRAS COMPONENTES PRINCIPAIS DOS SINAIS SIMULADOS.	46
TABELA 5 – MATRIZ DE CONFUSÃO OBTIDA PELA APLICAÇÃO DA REGRA DA MÉDIA DA CLASSE MAIS PRÓXIMA CONSIDERANDO AS 20 (VINTE) PRIMEIRAS COMPONENTES DOS SINAIS SIMULADOS.....	47
TABELA 6 – MATRIZ DE CONFUSÃO OBTIDA PELA APLICAÇÃO DA REGRA DA MÉDIA DA CLASSE MAIS PRÓXIMA CONSIDERANDO APENAS AS 2 (DUAS) PRIMEIRAS COMPONENTES PRINCIPAIS DA ENVOLTÓRIA DOS SINAIS ULTRASSÔNICOS SIMULADOS.	48
TABELA 7 – MATRIZ DE CONFUSÃO OBTIDA PELA APLICAÇÃO DA REGRA DA MÉDIA DA CLASSE MAIS PRÓXIMA CONSIDERANDO AS 20 (VINTE) PRIMEIRAS COMPONENTES DA ENVOLTÓRIA DOS SINAIS ULTRASSÔNICOS SIMULADOS.	48
TABELA 8 – AGRUPAMENTO, OU PSEUDO-CLASSIFICAÇÃO, DOS SINAIS SUGERIDA PELA K-MÉDIAS.	52
TABELA 9 – AGRUPAMENTO, OU PSEUDO-CLASSIFICAÇÃO, DAS ENVOLTÓRIAS DOS SINAIS SUGERIDA PELO ALGORITMO K-MÉDIAS.	56
TABELA 10 – AGRUPAMENTO, OU PSEUDO-CLASSIFICAÇÃO, PROPOSTO PELO ALGORITMO K-MÉDIAS APÓS PROCESSAMENTO DAS PRIMEIRAS COMPONENTES PRINCIPAIS DOS SINAIS.	59
TABELA 11 – AGRUPAMENTO, OU PSEUDO-CLASSIFICAÇÃO, PROPOSTO APÓS ANÁLISE DE AGRUPAMENTO DAS 2 PRIMEIRAS COMPONENTES PRINCIPAIS DAS ENVOLTÓRIAS DOS SINAIS.....	63
TABELA 12 – AGRUPAMENTO, OU PSEUDO-CLASSIFICAÇÃO, PROPOSTO PELO ALGORITMO K-MÉDIAS APÓS PROCESSAMENTO DAS 20 PRIMEIRAS COMPONENTES PRINCIPAIS DAS ENVOLTÓRIAS DOS SINAIS SIMULADOS.	63

TABELA 13 – MATRIZ DE CONFUSÃO OBTIDA PELO EMPREGO DA REGRA DA MÉDIA DA CLASSE MAIS PRÓXIMA APLICADA ÀS 2 PRIMEIRAS COMPONENTES PRINCIPAIS DOS SINAIS REAIS.	66
TABELA 14 – MATRIZ DE CONFUSÃO OBTIDA PELO EMPREGO DA REGRA DA MÉDIA DA CLASSE MAIS PRÓXIMA APLICADA ÀS 20 PRIMEIRAS COMPONENTES PRINCIPAIS DOS SINAIS REAIS.	67
TABELA 15 – AGRUPAMENTO, OU PSEUDO-CLASSIFICAÇÃO, SUGERIDA PELA K-MÉDIAS PARA OS SINAIS CAPTURADOS EXPERIMENTALMENTE.	67
TABELA 17 – MATRIZ DE CONFUSÃO OBTIDA PELO EMPREGO DA REGRA DA MÉDIA DA CLASSE MAIS PRÓXIMA APLICADA ÀS 2 PRIMEIRAS COMPONENTES PRINCIPAIS DAS ENVOLTÓRIAS DOS SINAIS REAIS.	69
TABELA 18 – MATRIZ DE CONFUSÃO OBTIDA PELO EMPREGO DA REGRA DA MÉDIA DA CLASSE MAIS PRÓXIMA APLICADA ÀS 20 PRIMEIRAS COMPONENTES PRINCIPAIS DAS ENVOLTÓRIAS DOS SINAIS REAIS.	71
TABELA 19 – AGRUPAMENTO, OU PSEUDO-CLASSIFICAÇÃO, DOS SINAIS CAPTURADOS EXPERIMENTALMENTE SUGERIDA PELA K-MÉDIAS.	74

SUMÁRIO

1 INTRODUÇÃO	16
2 REVISÃO BIBLIOGRÁFICA	18
2.1 DEFEITOS DE SOLDAGEM	18
2.1.1 <i>Falta de penetração</i>	18
2.1.2 <i>Poros</i>	19
2.1.3 <i>Trinca</i>	20
2.2 <i>TIME OF FLIGHT DIFFRACTION - TOFD</i>	20
2.3 <i>PHASED ARRAY</i>	22
2.4 MÉTODOS NUMÉRICOS	24
2.4.1 <i>Método dos Elementos Finitos</i>	24
2.4.2 <i>Métodos das Diferenças Finitas</i>	24
2.4.3 <i>Método dos Volumes Finitos</i>	25
2.5 COMSOL.....	25
2.6 CONDIÇÕES DE CONTORNO	25
2.6.1 <i>Free</i>	25
2.6.2 <i>Low-reflection boundary</i>	25
2.6.3 <i>Prescribed displacement</i>	26
2.7 ANÁLISE DE COMPONENTES PRINCIPAIS - PCA	26
2.8 K-MÉDIAS	29
2.8.1 <i>Silhueta</i>	32
2.8.2 <i>Davies-Bouldin</i>	32
3 MATERIAIS E MÉTODOS.....	34
3.1 COMSOL.....	34
3.1.1 <i>Falta de penetração</i>	39
3.1.2 <i>Poros</i>	41
3.1.3 <i>Trinca</i>	42
4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	44
4.1 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA).....	44
4.2 K-MÉDIAS	49
4.2.1 <i>Análise de agrupamento dos sinais ultrassônicos simulados</i>	49
4.2.2 <i>Análise de agrupamento das envoltórias dos sinais ultrassônicos simulados</i>	53
4.2.3 <i>Análise de agrupamento das componentes principais dos sinais ultrassônicos simulados</i>	56
4.2.4 <i>Análise de agrupamento das componentes principais das envoltórias dos sinais ultrassônicos simulados</i>	61
4.3 ANÁLISE DE COMPONENTES PRINCIPAIS E ANÁLISE DE AGRUPAMENTOS DOS SINAIS ULTRASSÔNICOS CAPTURADOS EXPERIMENTALMENTE	66
5 CONCLUSÃO.....	75
6 SUGESTÕES PARA TRABALHOS FUTUROS	77
7 REFERÊNCIAS	78

1 INTRODUÇÃO

Ensaaios não destrutivos são técnicas utilizadas para verificar a existência de descontinuidades nos materiais, sem danificar ou impedir seu uso posterior. Tais técnicas são amplamente utilizadas nos setores petroquímico, químico, aeroespacial, siderúrgico, naval, entre outros.

Além de detectar a presença de descontinuidades, algumas técnicas permitem também o seu dimensionamento. Ensaio visual, correntes parasitas, líquidos penetrantes e ultrassom são exemplos de métodos não destrutivos existentes que contribuem para garantir a qualidade do material, prevenir acidentes e reduzir custos, sendo considerado fator de competitividade entre as empresas que os utilizam. Neste trabalho, será abordado o ensaio não destrutivo por ultrassom.

O ensaio por ultrassom utiliza ondas mecânicas de alta frequência para detectar descontinuidades existentes no interior dos materiais. Ao encontrar uma descontinuidade estas ondas são refletidas de volta ao sensor piezoelétrico, onde são convertidas em impulsos elétricos. A partir da variação da amplitude da onda registrada pelo sensor é possível determinar a espessura do material, detectar e dimensionar descontinuidades. Tal ensaio possui alta sensibilidade na detecção de pequenas descontinuidades internas e exige um inspetor de bom nível técnico (Andreucci,2014).

Uma das técnicas ultrassônicas, a chamada técnica do tempo de percurso da onda difratada ou TOFD (*Time Of Flight Diffraction* [do inglês]), vem sendo cada vez mais utilizada e superando técnicas tradicionais. Essa técnica baseia-se na difração da onda causadas pelas extremidades (superior e inferior) da descontinuidade presente no interior do material, possui alta probabilidade de detecção de defeitos, e uma vasta área de aplicação (Silva,1999).

Entretanto, a classificação do tipo de descontinuidade a partir dos sinais ultrassônicos obtidos durante uma inspeção requer um profissional altamente capacitado e, considerando o volume de dados a ser analisado, um longo tempo. Essa tarefa pode ser realizada por algoritmos de classificação e/ou agrupamento capazes de processar rapidamente uma grande quantidade de dados.

Um classificador pode ser ensinado a reconhecer padrões a partir de exemplos conhecidos (já classificados) e, posteriormente, ser utilizado para prever as classes de dados desconhecidos (ou novos). Existem dois tipos de aprendizado: o supervisionado e o não

supervisionado (Kaufman e Rousseeu, 1990). No aprendizado supervisionado, o classificador é informado sobre a classificação real dos dados durante o treinamento, apresentar cada um dos dados de entrada e a qual classe pertence.

Diferentemente, no aprendizado não supervisionado não é requerido uma informação prévia sobre a classificação real dos dados. Na análise de agrupamento, uma abordagem não supervisionada, dado um conjunto de dados não classificado, deseja-se descobrir o número de grupos (ou *clusters*) existentes a partir das características dos dados de entrada. Espera-se que elementos de um mesmo grupo sejam similares e elementos de grupos diferentes sejam tão diferentes quanto possível. O conceito de similaridade entre dois pontos pertencente a um conjunto de dados multidimensionais pode ser construído com base na distância entre os pontos. Assim, quanto mais perto os pontos, mais semelhantes são eles. O algoritmo k-médias é uma técnica de agrupamento. Ao representar cada grupo por um centróide, comparar as distâncias entre os dados de entrada e os centróides e avaliar quão bem diferentes aglomerados estão separados uns dos outros e quão bem os elementos de um mesmo conjunto estão agrupados, o algoritmo decide a que grupo pertence cada dado de entrada.

Neste trabalho o algoritmo k-médias foi utilizado para agrupar três conjuntos de sinais ultrassônicos produzidos durante a simulação da propagação de onda, segundo a configuração do ensaio pela técnica TOFD, em um meio modelado contendo descontinuidades como falta de penetração, poro e trinca. Foi utilizado, também, a análise de componentes principais (PCA), para remover a correlação e reduzir a dimensionalidade dos dados.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta uma revisão bibliográfica que consiste de informações essenciais ao entendimento do tema estudado e técnicas empregadas ao longo do trabalho. Na seção 2.1 serão apresentados os tipos de descontinuidades considerados nesse trabalho. As seções 2.2 e 2.3, introduzem os princípios da técnica ultrassônica TOFD e *Phased Array*, respectivamente. A seção 2.4, define os métodos de elementos finitos, diferença finita e volumes finitos, possíveis de serem utilizados na solução numérica da equação da onda. Na seção 2.5 será abordado o programa COMSOL, um ambiente para simulação multifísica por meio de elementos finitos. A seção 2.6 apresenta algumas das condições de contorno possíveis de serem utilizadas no COMSOL. As seções 2.7 e 2.8 apresentarão a técnica de análise de componentes principais e a análise de agrupamento K-médias, respectivamente.

2.1 Defeitos de soldagem

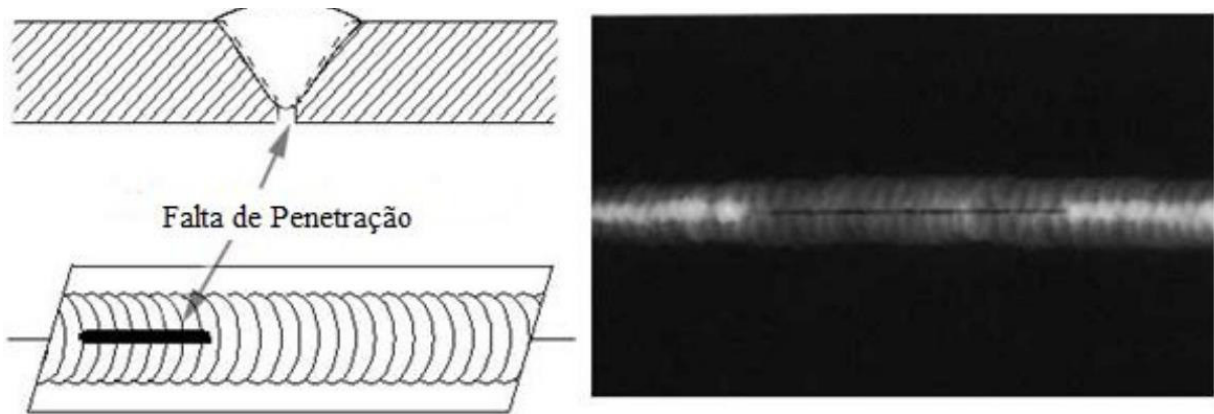
Uma descontinuidade é definida como uma interrupção das estruturas típicas da junta de solda, ou seja, a falta de homogeneidade de características físicas, mecânicas ou metalúrgica do material. A existência de descontinuidades em uma junta não significa que a mesma seja defeituosa, o que dependerá dos critérios de aceitação aplicáveis. Caso sua presença comprometa a utilização do material, peça ou estrutura, a descontinuidade é considerada um defeito, o que pode exigir ações corretivas tais como reparo ou substituição (Charles et al, 2002).

Três tipos de descontinuidades foram considerados nesse estudo, são elas: Falta de Penetração, Poro e Trinca.

2.1.1 Falta de penetração

A falta de penetração é uma descontinuidade ocasionada pelas falhas ao fundir e encher completamente a raiz de solda. As causas mais frequentes desta descontinuidade são a manipulação incorreta do eletrodo, juntas mal projetadas, corrente de soldagem insuficiente, velocidade de soldagem muito alta e diâmetro de eletrodo muito grande. Como consequência, tem-se a redução da seção útil da solda e concentração das tensões (Marques et al, 2009). Este tipo de descontinuidade é ilustrada na Figura 1.

Figura 1 – Representação de uma descontinuidade do tipo falta de penetração.



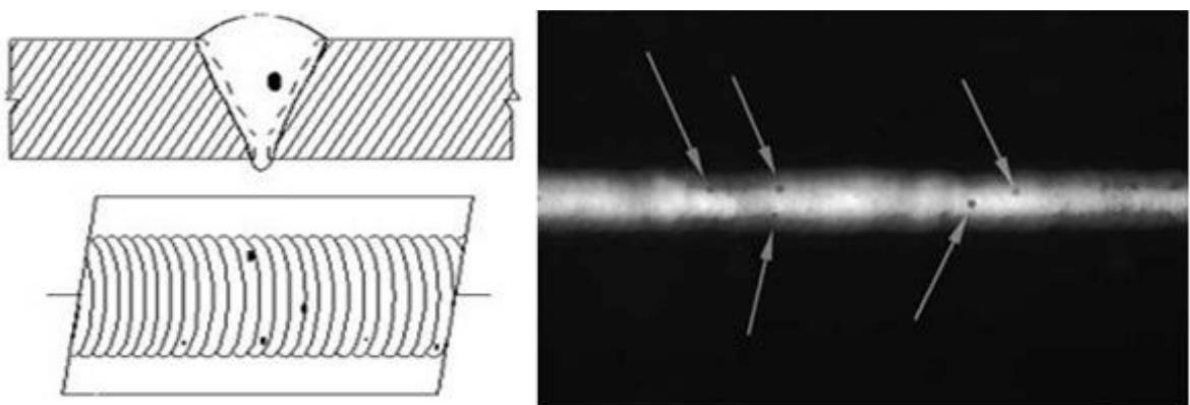
Fonte: NDT Resource Center (2012).

2.1.2 Poro

O poro representa uma descontinuidade ocasionada pela evolução de gases durante a solidificação da solda. As bolhas de gás podem ser aprisionadas pelo metal solidificado à medida que a poça de fusão é deslocada. As causas desta descontinuidade são umidade ou contaminações, por óleo, graxa, ferrugem, etc., na região de juntas, e corrente ou tensão de soldagem excessivas.

Pequenas quantidades de poros não são consideradas prejudiciais. Já em quantidades acima de um determinado limite pré-estabelecido (que varia para cada material), a porosidade pode afetar as propriedades mecânicas e reduzir a seção efetiva de junta (Marques et al, 2009). A representação da descontinuidade do tipo poro está ilustrada na Figura 2.

Figura 2 – Representação de uma descontinuidade do tipo poro.



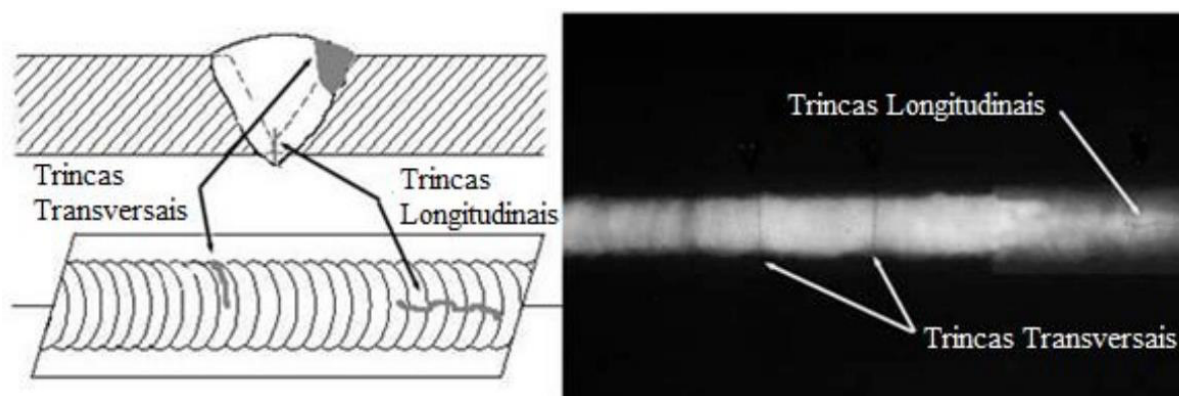
FONTE: NDT RESOURCE CENTER (2012).

2.1.3 Trinca

A trinca é uma das descontinuidades mais graves que podem ocorrer em uma solda, sendo considerada como resultado da incapacidade do material em responder às solicitações impostas localmente pela tensão decorrentes do processo de soldagem.

Quando se é aplicado calor em um material, são acarretadas expansões e contrações localizadas no próprio material, o que causa o aparecimento de tensões e trações na região de solda. Este estado de tensão juntamente com a fragilização, associada às mudanças microestruturais e/ou a presença de certos elementos, particularmente o hidrogênio, podem resultar em formação de trincas (Marques et al, 2009). A descontinuidade do tipo trinca é representada na Figura 3.

Figura 3 – Representação de uma descontinuidade do tipo trinca.



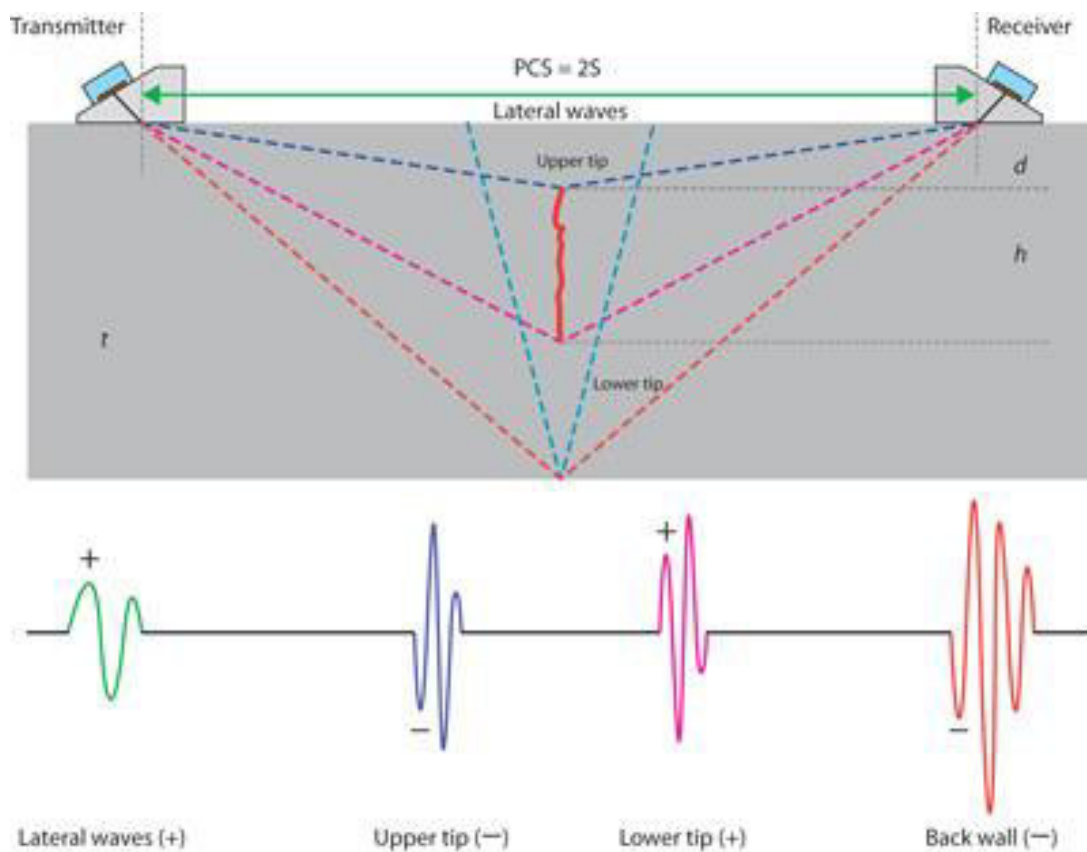
Fonte: NDT Resource Center (2012)

2.2 Time of Flight Diffraction - TOFD

A Técnica do Tempo de Percurso de Onda Difrata (Time Of Flight Diffraction [do ingl] - TOFD) descrita por M.G. Silk e B.H. Lidington Silk em 1975, é utilizada para localizar descontinuidades através da diferença entre os tempos de percurso dos sinais ultrassônicos que são difratados pelas extremidades (superior e inferior) das falhas.

Calcula-se a profundidade da descontinuidade utilizando o tempo de fuga da onda difratada a partir da ponta superior de tal falha. Já para calcular o comprimento subtrai-se a profundidade da ponta inferior da profundidade da ponta superior. A técnica TOFD é representada na Figura 4.

Figura 4 – Configuração da técnica TOFD.



Fonte: Olympus Corporation

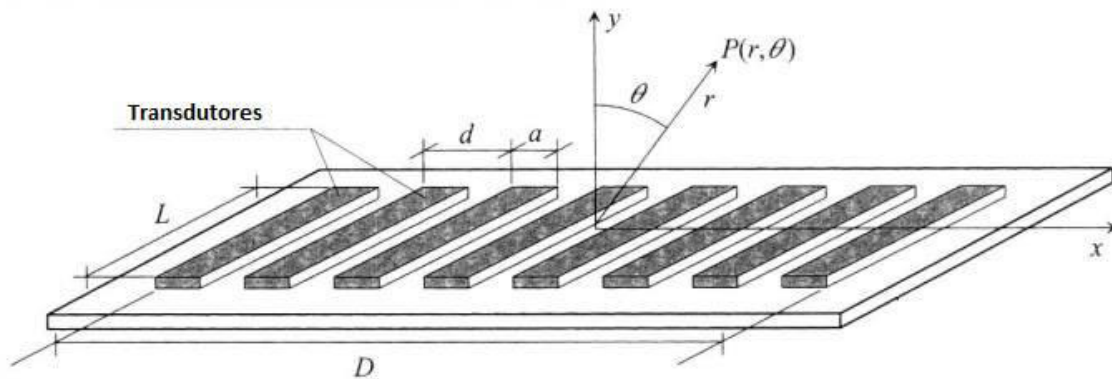
A primeira indicação produzida através dessa técnica é o sinal correspondente à onda lateral, formado pela propagação do som logo abaixo da superfície da peça. Caso os transdutores estejam sobre uma região contendo uma descontinuidade, o segundo e o terceiro sinais são resultado das difrações das pontas superior e inferior do defeito, respectivamente. Por fim tem-se o sinal resultante da reflexão de ondas longitudinais, tipo de onda em que a movimentação das partículas do meio ocorre na mesma direção da propagação da onda, na parede oposta da peça (Silva, 1999).

Uma análise simples desses sinais fornece as dimensões da descontinuidade. Dependendo da experiência do inspetor e da complexidade do caso é possível ainda identificar o tipo de descontinuidade que a peça apresenta. Sendo assim, o TOFD é uma importante técnica para a detecção de descontinuidade e sua aplicação é permitida com o equipamento ainda em operação, não sendo necessárias paradas do equipamento.

2.3 Phased Array

A interferência construtiva e destrutiva de ondas geradas por várias fontes pontuais estimuladas em instantes ligeiramente diferentes pode ser utilizada para produzir um feixe direcionado. Transdutores compostos por múltiplos elementos vibrantes que se utilizam desse princípio de interferência construtiva e destrutiva são chamados *Phased Array* (Bevan e Copson, 1987). A Figura 5 representa um esboço do princípio da técnica *Phased Array* utilizando 8 cristais de comprimento L e largura a , que estão a uma distância d .

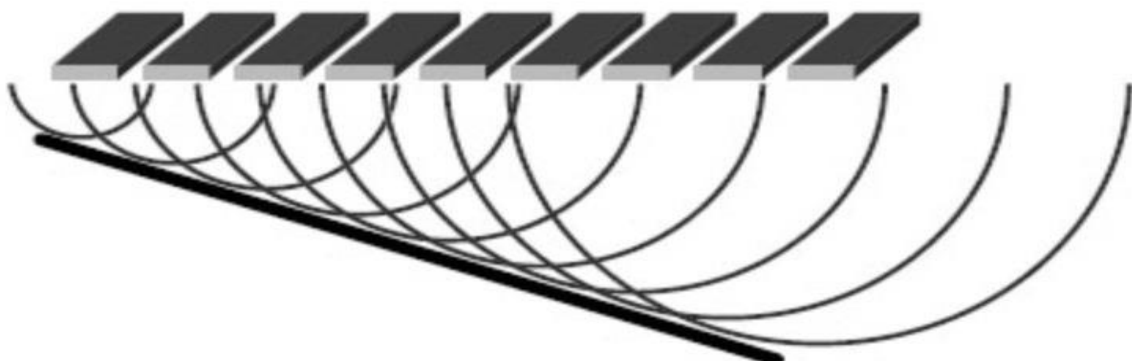
Figura 5 – Representação do Phased Array.



Fonte: Adaptado de WOOH S e SHI Y (1999).

O princípio de funcionamento do transdutor *Phased Array* é muito útil na simulação da geração do feixe sônico incidindo obliquamente. Ao colocar os transdutores para vibrar com um pequeno *delay* em relação ao transdutor vizinho, conseguimos direcionar o feixe. Como está sendo mostrado na Figura 6.

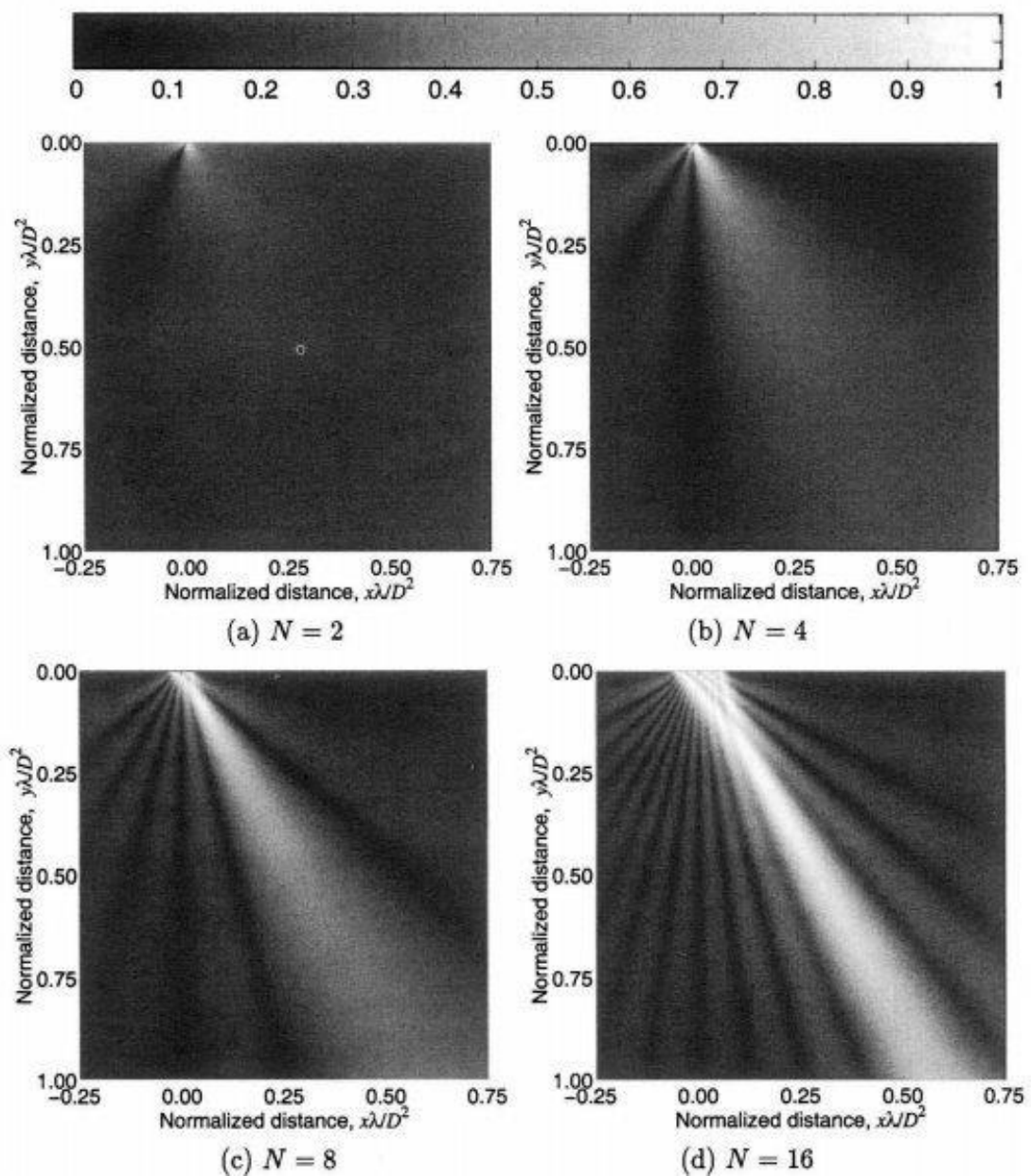
Figura 6 – Direcionamento do feixe por Phased Array.



Fonte: NDT-KITS BLOG (2013).

O número de cristais utilizados no direcionamento do feixe é uma variável importante e quanto maior esse número mais estreito é o feixe, implicando em feixe melhor direcionado e com maior pressão sônica na direção do feixe (Woo S e Shi Y, 1999). Por outro lado, maior será também o esforço computacional necessário à simulação. Desta forma, visando reduzir o esforço computacional, otimizar a simulação e produzir um feixe bem direcionado, faz-se necessário escolher adequadamente o número de pontos vibrantes. Na Figura 7, é representado o efeito da variação do número transdutores e o direcionamento do feixe.

Figura 7 – Direcionamento do feixe com a variação do número de transdutores.



Fonte: WOOH S e SHI Y (1999).

2.4 Métodos numéricos

O conjunto de ferramentas ou métodos utilizados para se obter soluções aproximadas de problemas matemáticos são denominados de Métodos Numéricos. Esses métodos são aplicados em problemas que não apresentam uma solução de forma exata, precisando assim, ser resolvidos numericamente (Buffoni, 2002).

Um importante tipo de Método Numérico é o método de discretização, o qual aproxima as equações por um sistema de funções algébricas em pontos discretos no espaço e tempo, cujo a idéia consiste em transformar um problema complexo na soma de diversos problemas simples.

As três metodologias mais utilizadas são: Método dos Elementos Finitos, Método das Diferenças Finitas E Método dos Volumes Finitos, a serem brevemente explicados no presente capítulo.

2.4.1 Método dos Elementos Finitos

O Método dos Elementos Finitos consiste em se dividir o domínio do problema em sub-regiões de geometria simples. Utiliza-se a forma integral das equações como ponto de partida, para isso multiplica-se os elementos por uma função peso antes de serem integradas em todo o domínio.

Neste método, o domínio é dividido em elementos discretos, que podem ser quadriláteros ou triângulos, e as malhas são facilmente refinadas em geometrias complicadas (Steven e Raymond, 2009).

2.4.2 Métodos das Diferenças Finitas

O Métodos das Diferenças Finitas é o método mais antigo para resolução de Equações Diferenciadas Parciais. Neste método, utilizam-se as derivadas obtidas através da expansão em série de Taylor ou aproximação polinomial como ponto de partida e, como nos métodos anteriores, o domínio de solução é subdividido em números finitos. Este método é o mais fácil de ser utilizado em geometria imersa (Steven e Raymond, 2009).

2.4.3 Método dos Volumes Finitos

O Método dos Volumes Finitos utiliza a forma integral das equações como ponto de partida, transformando integrais de volume em equações diferenciais parciais que contenham termos divergentes em integrais de superfície (Ferziger e Peric, 1997).

O domínio de solução é subdividido em um número finito de volumes de controle adjacentes entre si, onde as equações de conservação são aplicadas, pois o fluxo entrando num elemento é idêntico ao fluxo saindo do elemento adjacente. Este método é facilmente formulado e aplicado em malhas não estruturadas (Steven e Raymond, 2009).

2.5 COMSOL

O COMSOL é um software com base em métodos numéricos avançados para modelar e simular problemas baseados em física. Através deste software é possível adicionar equações que podem descrever uma propriedade de material, alterar as condições de contorno, reproduzir materiais com diferentes formatos e criar novas interfaces físicas. O COMSOL utiliza a técnica de elementos finitos.

2.6 Condições de contorno

As condições de contorno são exigências dadas para o extremo da peça. Para uma simulação adequada, determinam-se as condições de contorno: *Free*, *Low-Reflection Boundary* e *Prescribed Displacement*.

2.6.1 Free

A condição de contorno *Free* representa uma condição que não possui restrições e nenhum carregamento. Os sinais que são encontrados fora dos limites da peça são interpretados como tendo o valor zero. (Landau et al, 1986)

2.6.2 Low-reflection boundary

A condição de contorno *Low-Reflection Boundary* permite que as ondas não sejam refletidas nos contornos selecionados, criando-se uma variação na impedância (Filho, 2002).

Esta condição de contorno é necessária e suficiente para garantir que haverá reflexão total da onda na parede oposta, como acontece na interface aço/ar.

2.6.3 Prescribed displacement

A condição de contorno *Prescribed Displacement* é uma condição referente ao carregamento nas linhas que representam as vibrações dos cristais dos transdutores (dispositivo que recebe o sinal). A condição é utilizada também quando não se quer que ocorra o deslocamento (Filho, 2002). Essa condição de contorno dispensa a necessidade de modelar um sistema maior do que o necessário.

2.7 Análise de componentes principais - PCA

A Análise de Componentes Principais (*Principal Components Analysis [do ingl] - PCA*) é um procedimento matemático que utiliza uma transformação ortogonal para converter um conjunto de observações de p dados/amostras possivelmente correlacionadas (x_1, x_2, \dots, x_p) a um conjunto de valores de p dados/amostras linearmente descorrelacionadas (z_1, z_2, \dots, z_p) que descrevam a variação nos dados. As variáveis descorrelacionadas são chamadas de componentes principais, que é o arranjo que melhor representa a distribuição dos dados (Varella, 2008).

Esta transformação é definida de forma que cada componente principal é uma combinação linear de todas as variáveis originais, os mesmos são independentes entre si e estimados com o propósito de conservar o máximo de informação possível, em termos da variação total contida nos dados. Desta forma, a técnica é utilizada para a redução de massa de dados, com menor perda possível da informação.

O primeiro passo para calcular a componente principal é colocar os dados em forma de matriz de dados de dimensão ' $n \times p$ ', onde ' p ' o número de dados amostrais e ' n ' são os números de atributos (Varella, 2008). A Figura 8 representa esta matriz.

Figura 8 – Matriz de dado de dimensão ‘n x p’

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix}$$

Fonte: Varella, 2008.]

Depois de obtida a matriz, é necessário padronizar os dados, pois normalmente as características são observadas em unidades de medidas diferentes entre si. A padronização realizada neste trabalho teve a média igual a 0 e a variância igual a 1, obtida a partir da equação 1.

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s(x_j)}, \quad i = 1, 2, \dots, p \quad \text{e} \quad j = 1, 2, \dots, n \quad (1)$$

\bar{x}_j e $s(x_j)$ são, respectivamente, a estimativa da média e o desvio padrão da característica j. O desvio padrão, indicado pela equação 2, é dado pela raiz da variância, a qual é uma variável aleatória que indica quão longe em geral os seus valores se encontram do valor esperado.

$$s = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (x_i - \bar{x})^2} \quad (2)$$

Para verificar a dispersão, levando em conta duas variáveis aleatória utiliza-se uma medida bidimensional chamada de covariância, e calculada pela equação 3.

$$Cov(X, Y) = \frac{\sum_{i=1}^p (x_i - \bar{x})(y_i - \bar{y})}{p} \quad (3)$$

Para realizar-se a análise da componente principal, é necessário calcular uma matriz de covariância, uma matriz simétrica e de ordem ‘p x p’, representada na figura 9.

Figura 9 – Matriz de covariância de ordem ‘p x p’

$$S = \begin{bmatrix} \hat{V}ar(x_1) & \hat{C}ov(x_1x_2) & \hat{C}ov(x_1x_3) & \cdots & \hat{C}ov(x_1x_p) \\ \hat{C}ov(x_2x_1) & \hat{V}ar(x_2) & \hat{C}ov(x_2x_3) & \cdots & \hat{C}ov(x_2x_p) \\ \hat{C}ov(x_3x_1) & \hat{C}ov(x_3x_2) & \hat{V}ar(x_3) & \cdots & \hat{C}ov(x_3x_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{C}ov(x_px_1) & \hat{C}ov(x_px_2) & \hat{C}ov(x_px_3) & \cdots & \hat{V}ar(x_p) \end{bmatrix}$$

Fonte: Varella, 2008.

Após construída a matriz de covariância, é necessário calcular os seus autovalores e auto vetores (Boldrini,1986). Os autovalores λ da matriz (p x p) são as raízes da equação 4.

$$\det(S - \lambda I) = 0 \quad (4)$$

Os auto vetores v desta matriz são as soluções da equação para cada λ , como pode ser observado na equação 5 (Boldrini,1986).

$$Sv = \lambda v \quad \text{ou} \quad (S - \lambda I)v = 0 \quad (5)$$

Em seguida, arranja-se os p auto vetores de modo decrescente de acordo com os valores dos p auto valores relacionados. Isto é, será chamado de v_1 o auto vetor correspondente ao maior auto valor λ_1 , de v_2 o auto vetor correspondente ao segundo maior auto valor λ_2 , e assim sucessivamente de modo que será chamado de v_p o auto vetor correspondente ao menor auto valor λ_p .

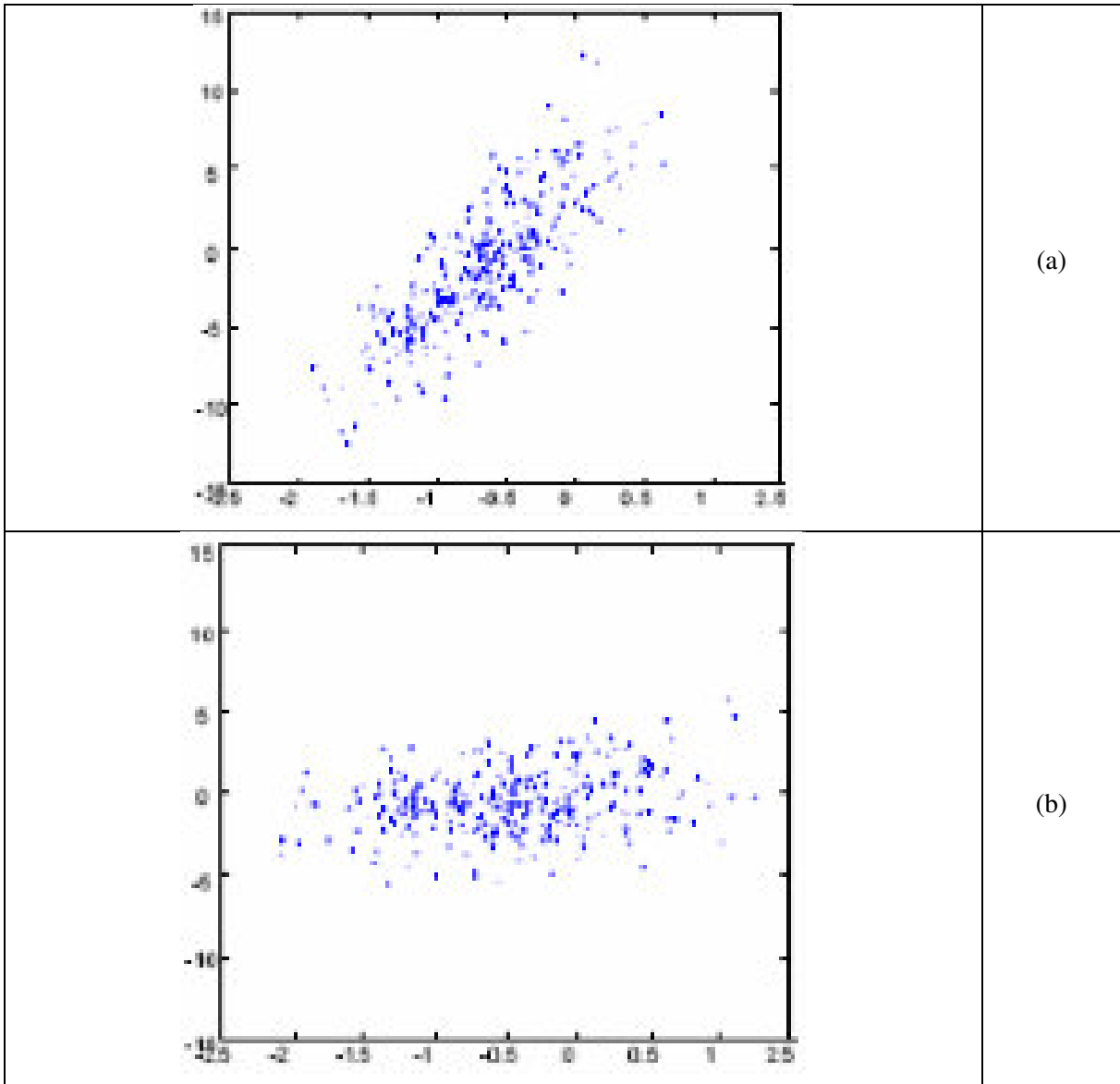
Esta é a matriz de transformação utilizada para o cálculo da PCA. Multiplicam-se os dados por ela para se obter novos dados com uma menor dimensão.

Os auto valores correspondem a variância, ou seja, quanto maiores os auto valores maior a proporção de dados que os auto vetores relacionados irão explicar. Com isso, serão descartadas quaisquer componentes que explicam somente uma pequena proporção da variação nos dados.

Ao dividir o auto valor pela soma de todos auto valores, dará um número menor que um, que representa a proporção de dados explicados pelo o auto vetor associado, este dado irá nos ajudar a escolher quais e quantos auto vetores representará o conjunto de dados total (Varella, 2008).

Na Figura 10, está a representação das duas primeiras dimensões dos dados antes (a) e depois (b) da aplicação da técnica PCA, é importante lembrar que conseguimos representar apenas as duas primeiras dimensões em um plano.

Figura 10 – (a) representação dos dados antes do PCA; (b) representação dos dados depois do PCA.



Fonte: VASCONCELOS, 2007.

2.8 K-médias

Um desafio recorrente enfrentado na engenharia é a classificação de dados. Como agrupar os dados observados em estruturas que façam sentido ou como desenvolver algoritmos capazes de classificar esses dados em diferentes classes? (Wangenheim, 2006).

É necessário, então, agrupar os dados da forma mais similar possível, contudo a similaridade é difícil de ser definida. Quanto maior for a medida de dissimilaridade menor será a semelhança entre os indivíduos. Na Figura 11 é possível observar esta dificuldade. Dependendo do critério escolhido, podem-se agrupar as imagens em diferentes conjuntos.

Figura 11 – Similaridade



Fonte: XCITEFUN (2007).

O algoritmo K-médias agrupa os dados de acordo com a Distância Euclidiana. A distância Euclidiana entre os pontos $P = (p_1, p_2, \dots, p_n)$ e $Q = (q_1, q_2, \dots, q_n)$, em um espaço euclidiano n-dimensional, é definida pela equação 6 (Deza E e Deza M, 2009).

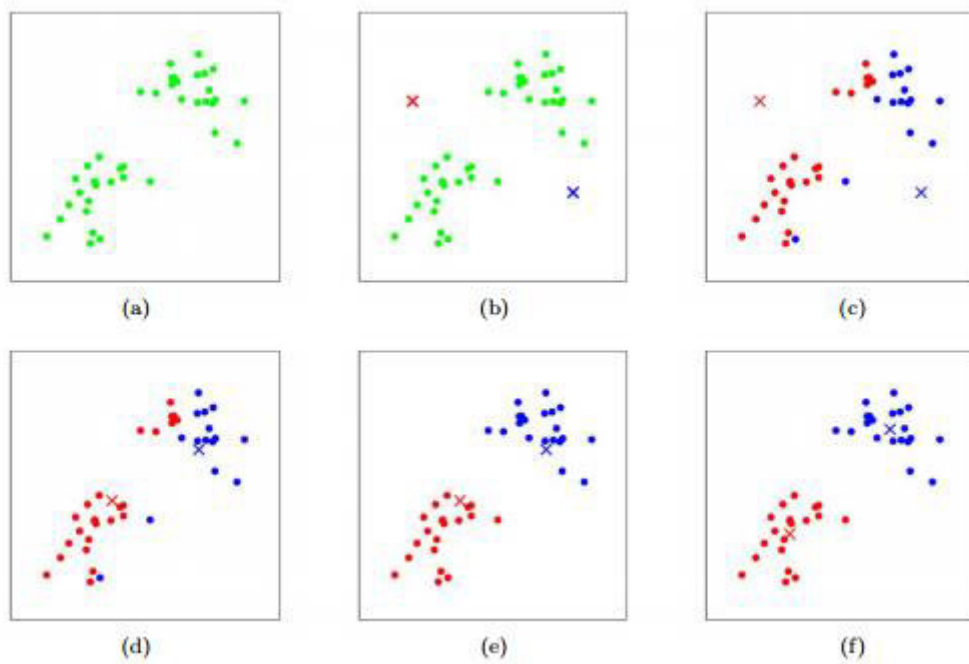
$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}. \quad (6)$$

O objetivo do algoritmo é agrupar os dados em uma quantidade predefinida k de classes. Para isso, o algoritmo experimenta números de grupos, k, que varia de 1 e o valor dado pela raiz quadrada do número total de dados de entrada.

Normalmente utilizando a distância Euclidiana, cada um dos elementos (dados de entrada) é associado ao centróide mais próximo. Outros tipos de distância podem ser utilizadas. A posição dos k centróides é então recalculada utilizando a média dos objetos do agrupamento correspondente resultante do passo anterior. Esse procedimento é repetido até que a mudança na posição dos centróides seja inferior a um limiar, ou um número pré-definido de iterações seja alcançado (Anil, 2009).

Na Figura 12 é possível observar o esquema de funcionamento do algoritmo K-médias.

Figura 12 – Representação do funcionamento do K-médias.



Fonte: PIECH (2013).

Considere um conjunto de dados p -dimensional com n elementos, o algoritmo k -média basicamente avalia uma função de custo calculada com base na distância de cada elemento ao centróide do grupo a que está associado. Como mostrado na Figura 13.

Figura 13 – Função de Custo do K-médias.

$$\text{Função de custo} \leftarrow J = \sum_{j=1}^k \sum_{i=1}^n \underbrace{\|x_i^{(j)} - c_j\|}_{\text{Distância Euclidiana}}^2$$

The diagram includes the following annotations:

- k : Número de Classes
- n : Número de Dados
- $x_i^{(j)}$: Dado i
- c_j : Centroide j
- $\|x_i^{(j)} - c_j\|$: Pertencente ao grupo j
- $\|x_i^{(j)} - c_j\|^2$: Distância Euclidiana

Fonte: Adaptado de Frankenstein, 2009.

Modificando o números de grupos, k , e utilizando alguns índices tais como: Silhueta e Davies Bouldin, descritos a seguir, o algoritmo surge um número ótimo de agrupamentos. Ou

seja, qual é a melhor forma de dividir os dados entre as divisões feita pelo o k-médias. A qualidade do agrupamento é tão melhor quanto mais próximo da unidade for o índice Silhueta, e/ou quanto mais próximo de zero for o índice Davies-Bouldin.

2.8.1 Silhueta

Neste critério, cada ponto é representado por uma silhueta e o número de conjuntos k , que melhor agrupa os dados, é aquele que se posiciona melhor dentro do ponto.

Este critério combina a ideia de coesão e separação, no qual verifica quão próximo os elementos estão do mesmo conjunto e quão longe os elementos estão do grupo mais próximo.

Para obter o índice Silhueta, calcula-se primeiro a distância de um ponto para todos os elementos do mesmo grupo e depois calcula-se a média dessas distâncias (a média dessas distâncias será chamada de a_i). (Bolshakova N, 2002)

Em seguida, calcula-se a distância de um ponto para todos os elementos do grupo mais próximo e depois calcula-se a média dessas distâncias (a média dessas distâncias será chamada de b_i). O índice é dado pela equação 7.

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (7)$$

O valor de s_i está no intervalo $[-1, 1]$, sendo que os melhores agrupamentos possuem o índice Silhueta mais próximos de 1 e os piores mais próximo de -1. Portanto, para os melhores agrupamentos o a_i é próximo de zero e b_i próximo de 1. O índice silhueta é a média de todos os s_i para todos os i .

2.8.2 Davies-Bouldin

O índice de Davies-Bouldin, assim como o Silhueta, é utilizado para a avaliação da partição dos dados, de tal forma que através da comparação relativa entre diferentes divisões do conjunto definimos qual é a melhor divisão para os dados. (Petrovic , 2002)

Para obter o índice Davies-Bouldin, calculam-se todas as k distâncias entre o centroide do *cluster* i para todos os elementos do grupo i . (chamadas de d_i). Em seguida, calcula-se a

distância entre cada par de centroide i e j . (chamadas de $d(c_i, c_j)$). O índice é dado pela equação 8.

$$DB = \frac{1}{k} \sum_{i=1, i \neq j}^k \max \left(\frac{d_i + d_j}{d(c_i, c_j)} \right) \quad (8)$$

Para valores de DB mais próximos de zeros temos os melhores agrupamentos. O índice Davies-Bouldin é considerado mais complexo quando comparado com o Silhueta, mas simplifica os cálculos ao utilizar centróides em vez de pontos no seu cálculo.

3 MATERIAIS E MÉTODOS

A modelagem de um meio bidimensional contendo uma descontinuidade interna e a simulação numérica do ensaio ultrassônico foi realizada com auxílio do ambiente COMSOL. O meio modelado reproduz as propriedades mecânicas de um aço carbono comum e a descontinuidade inserida é escolhida entre falta de penetração, poro e trinca. Cada simulação produz um sinal de deslocamento, lido na posição do receptor, no tempo. A análise de componentes principais combinada com o algoritmo k-médias foram empregados de modo a extrair informações e classificar automaticamente os sinais produzidos durante a simulação.

As próximas seções desse capítulo definem as variáveis do problema, como: informações sobre a fonte, as propriedades do meio, os tipos e dimensões das descontinuidades inseridas neste, bem como os detalhes da configuração adotadas no ambiente COMSOL durante a simulação.

3.1 COMSOL

Simulações são úteis no estudo das variáveis de um problema e ajudam na sua compreensão. Podem ser empregados para prever resultados, principalmente quando a realização de experimentos é cara e difícil de conduzir. Neste trabalho, o programa COMSOL foi utilizado para simular o ensaio não destrutivo por ultrassom em um meio bidimensional e isotrópico (meio cujas propriedades físicas independem da direção) de modo a produzir sinais para um estudo de caracterização de defeitos. Isso requer a escolha adequada do modelo geométrico (análise bidimensional) definida na aba *Model Wizard* do ambiente COMSOL. Também foi adicionado a interface *Solid Mechanics* que possibilita computar dados sobre deslocamento, força e a deformação a partir das equações de Navier-Stokes. Adicionalmente, o problema abordado é do tipo transiente, o que exige habilitar a opção *Time-Dependent*.

O material foi modelado como elástico e isotrópico (*Linear Elastic e Isotropic*) com as propriedades da peça mostrada na Tabela 1. Em seguida, a geometria do sistema foi definida como sendo um retângulo de comprimento igual a C (0.10 metro) e largura igual a L (0.019 metro).

Tabela 1 – Propriedades do material

Propriedades	Valores Estabelecidos
Material	Aço
Velocidade da onda longitudinal	5900 m/s
Velocidade da onda transversal	3200 m/s
Densidade	7900 kg/m ³

Fonte : Tabela do próprio autor.

A configuração para realização do ensaio ultrasônico empregando a técnica TOFD envolve a utilização de dois transdutores, um emissor e um receptor, um de cada lado da região inspecionada, na qual o emissor produz e incide um feixe sônico oblíquo. Na prática isso é alcançado montando os transdutores emissor e receptor sobre cunhas confeccionadas em plexiglas com ângulos calculados de acordo com o ângulo de incidência desejado. Para reproduzir computacionalmente essa configuração é necessário adicionar nós computacionais, o que implica no aumento do esforço computacional. A solução adotada para gerar um feixe sônico incidente inclinado foi utilizar o princípio de funcionamento da técnica *Phased Array*. O ângulo de incidência igual a 60° foi escolhido com base em trabalhos publicados que apontam esse como sendo o ângulo de melhor desempenho na inspeção por TOFD (Temple, 1986)

A distância entre os múltiplos elementos que compõem um transdutor phased array está relacionada com ângulo máximo de varredura que esse transdutor pode alcançar e o comprimento de onda no meio inspecionado e pode ser calculada de acordo com a equação 9.

$$d = \frac{\lambda}{1 + \sin(\theta)_{\text{máx}}}$$

(9)

Considerando que o transdutor será utilizado na inspeção de chapas confeccionadas em aço, cuja velocidade de propagação do som é 5900 m/s, e fixando sua frequência em 5 MHz (frequência comumente utilizada na inspeção por ultrassom), o comprimento da onda produzida por esse transdutor, dado pela razão entre a velocidade e a frequência da onda, é 1.18×10^{-3} m.

Definindo o ângulo máximo de incidência de nosso transdutor como sendo igual a 90° , o espaçamento, d , entre os elementos que o compõe é de $\lambda/2$ (0.59×10^{-3} m).

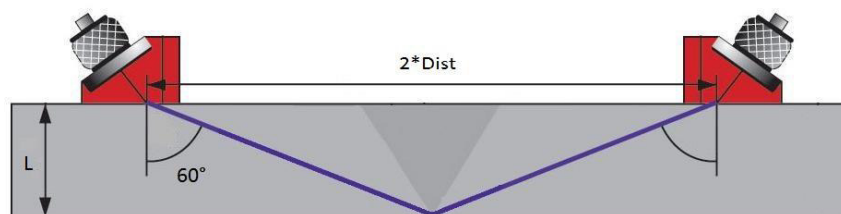
O cálculo do atraso no tempo de excitação de cada cristal que compõe o transdutor faz uso das relações trigonométricas do Princípio de Huygens, um método de representação da frente de onda, onde cada ponto de uma frente se comporta como uma nova fonte de ondas elementares, que se propagam para além da região já atingida pela onda original e com a mesma frequência que ela. A relação trigonométrica utilizada é dada pela equação 10.

$$\sin \theta = \frac{C \cdot dt}{d} \quad (10)$$

onde C é a velocidade com que a onda propaga no material inspecionado [5900 m/s para o aço carbono comum], dt é o atraso da ativação das fontes (*delay*), e d representa o espaçamento entre as múltiplas fontes. Como descrito anteriormente, com base no trabalho de Temple, J.A.G em 1986, o valor de θ foi definido como sendo igual a 60° , e a distância entre os elementos do transdutor, calculada de acordo com a equação 9, é igual a metade do comprimento de onda ($\lambda/2$). A equação 10 fornece um atraso no tempo de excitação entre as fontes igual a 8.6603×10^{-8} segundo.

A distância entre os transdutores é $2 \cdot \text{Dist}$, como mostra a Figura 14. Sendo $\tan(60^\circ)$ igual a L (espessura da peça) dividido por Dist (metade da distância entre os transdutores), temos que dist é igual a 0.034641 m ($L \cdot \tan(60^\circ)$).

Figura 14 – Representação da relação entre a distância entre os transdutores e a espessura do material



Fonte: Adaptado de NTD resource center - Ultrasound

Com essas informações é possível definir algumas das variáveis do problema. No ambiente COMSOL tais variáveis são definidas em *Parameters* na opção de *Global Definitions*. A Tabela 2 ilustra os valores das variáveis.

Tabela 2 – Definição das variáveis do problema

Nome	Expressão	Valor	Descrição
C	100[mm]	0.10000 m	Comprimento da peça
L	19[mm]	0.01900 m	Espessura da peça
f	5e6[Hz]	5.000E6 Hz	Frequencia do transdutor
Lamb	5900[m/s]/f	0.001180 m	Comprimento de onda
ang	pi/3[rad]	1.0472	Ângulo de incidência
dt	$((\text{lamb}/2) * \sin(\text{ang}))/5900[\text{m/s}]$	8.6603E-8 s	Atraso na excitação dos elementos transdutores (Delay)
dist	$L * \tan(\text{ang})$	0.034641 m	Metade da distância entre emissor e receptor
tamcrist	$9 * (\text{lamb}/2)$	0.005310 m	Dimensão lateral do transdutor
time	$(3 * (L / \cos(\text{ang}))) / 5900[\text{m/s}]$	2.0339E-5 s	Tempo de propagação da onda

Fonte : Tabela do próprio autor.

Após definidas tais variáveis deve-se especificar acerca do transdutor, ou seja, das funções relacionadas com a fonte. Como apresentado na seção 2.3, quanto maior o número de cristais que compõe o transdutor melhor a qualidade do feixe sônico produzido. No entanto, maior será o esforço computacional necessário à simulação. Estabeleceu-se, com base em trabalhos anteriores (Wooh e Shi, 1999) que o transdutor reproduzido nesse trabalho seria composto por oito elementos vibrantes (ver Figura 7).

Com o objetivo de modelar um transdutor *phased array* composto por 8 cristais (ou fontes vibrantes), é necessário fixar 9 pontos equidistantes. O primeiro ponto criado na superfície encontra-se a $(0.05 \cdot \text{dist} - \text{lamb}/4)$ metros da extremidade esquerda do sistema e um novo ponto é criado a cada $\text{lamb}/2$ de distância. Esses 9 pontos serão utilizados para formar oito linhas que constituirão os cristais. Procedimento análogo foi repetido para criar oito linhas horizontais com comprimento de $\text{lamb}/2$ na superfície do lado direito da peça, que representará o receptor. As linhas foram criadas uma do lado da outra com distância da primeira linha de $(0.05 + \text{dist} - \text{lamb}/4)$ metros da superfície.

Devemos também criar nove funções em relação ao tempo, a variável independente do problema. Cada uma dessas funções está relacionado com um ponto, que forma o transdutor, onde o momento inicial é quando cada cristal começar a vibrar e o final é o instante em que ele irá parar. A Tabela 3 mostrar cada função e os seus respectivos valores de tempo inicial e final.

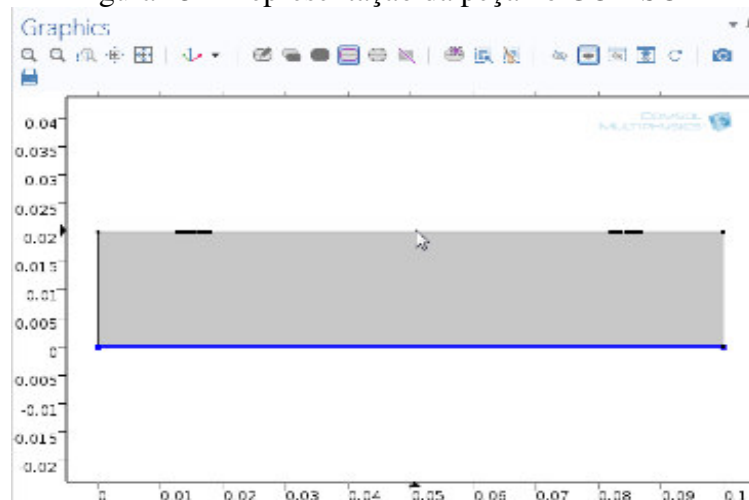
Tabela 3 – Funções relacionadas com a fonte

Início	Fim	Função
0	3/f	$(1-\cos(2\pi f t/3))\cos(2\pi f t)$
1*dt	3/f+1*dt	$(1-\cos(2\pi f (t-1*dt)/3))\cos(2\pi f (t-1*dt))$
2*dt	3/f+2*dt	$(1-\cos(2\pi f (t-2*dt)/3))\cos(2\pi f (t-2*dt))$
3*dt	3/f+3*dt	$(1-\cos(2\pi f (t-3*dt)/3))\cos(2\pi f (t-3*dt))$
4*dt	3/f+4*dt	$(1-\cos(2\pi f (t-4*dt)/3))\cos(2\pi f (t-4*dt))$
5*dt	3/f+5*dt	$(1-\cos(2\pi f (t-5*dt)/3))\cos(2\pi f (t-5*dt))$
6*dt	3/f+6*dt	$(1-\cos(2\pi f (t-6*dt)/3))\cos(2\pi f (t-6*dt))$
7*dt	3/f+7*dt	$(1-\cos(2\pi f (t-7*dt)/3))\cos(2\pi f (t-7*dt))$
8*dt	3/f+8*dt	$(1-\cos(2\pi f (t-8*dt)/3))\cos(2\pi f (t-8*dt))$

Fonte : Tabela do próprio autor.

As posições do emissor e do receptor foram definidas de modo que a região onde serão modelados as descontinuidades, representando uma junta soldada defeituosa, fique localizada no centro dos transdutores e seja totalmente sonificada. A esquematização da geometria modelada é mostrada na Figura 15.

Figura 15 – Representação da peça no COMSOL



Fonte : Figura do próprio autor.

Para solução numérica exige a escolha de uma tipo de malha. Entre os tipos disponibilizados pelo ambiente COMSOL, foi escolhida uma malha triangularizada (referida como FreeTri). A malha triangularizada foi escolhida devido a sua adaptatividade em geometrias diferentes. Apenas o tamanho máximo da maior dimensão triângulo é requerida. Nesse estudo, cujo tamanho da malha está diretamente relacionado com o comprimento da onda, o tamanho máximo do lado do triangulo foi fixado em $1/9$ do comprimento de onda.

Para essa simulação foram utilizadas as condições de contorno de superfície livre (na face onde é colocado o transdutor), reflexão (na face oposta ao transdutor) e absorção (nas extremidades da peça) que o COMSOL define, respectivamente, por Free, Prescribed Displacement e Low-Reflection Boundary. Estas condições estão sendo definidas na seção 2.6.

Por fim, foi escolhido um método padrão do COMSOL (*solver*) para resolução de equações e o algoritmo foi executado para simular a onda ultrassônica propagando por $2.0339 \cdot 10^{-5}$ segundos. Dependendo da capacidade da máquina cada simulação pode demorar entre 30 minutos e 1 hora. Os sinais A-scan (amplitude de deslocamento no tempo) foram registrados com 4831 pontos. Visando reduzir o esforço computacional envolvido nas próximas etapas do trabalho (PCA e K-médias) e para permitir uma possível comparação entre sinais simulados e sinais experimentais, todos os sinais foram posteriormente reamostrados para uma quantidade de 512 pontos.

Foram modelados três tipos de descontinuidades comumente encontradas em juntas soldadas: poro, trinca e falta de penetração.

Para cada tipo de descontinuidade foram feitas várias simulações, variando, dependendo do tipo, suas dimensões e sua posição. No caso de falta de penetração sua posição é constante em relação ao centro do cordão. Alterações foram efetuadas apenas nas suas dimensões (largura e altura), enquanto que para os outros dois tipos de descontinuidades há alteração tanto no tamanho como na sua posição. A seguir, serão detalhadas as simulações realizadas para o presente trabalho.

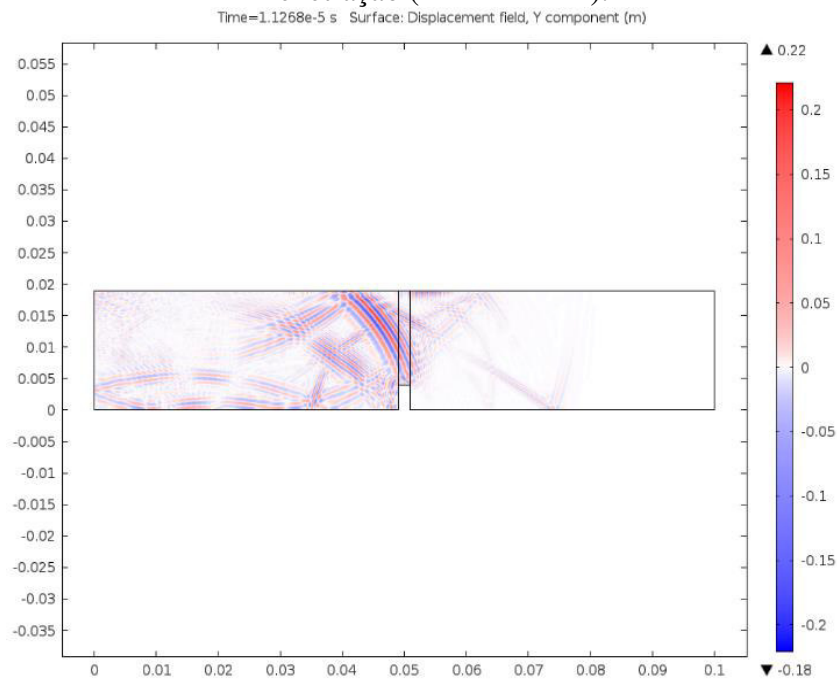
3.1.1 Falta de penetração

Foi simulada a propagação de onda em um meio contendo falta de penetração com diferentes dimensões. Descontinuidades produzidas pela combinação da largura, que variou de

2 mm a 7 mm, e altura, que variou de 4 mm a 9 mm, totalizando 36 simulações para esse tipo de defeito.

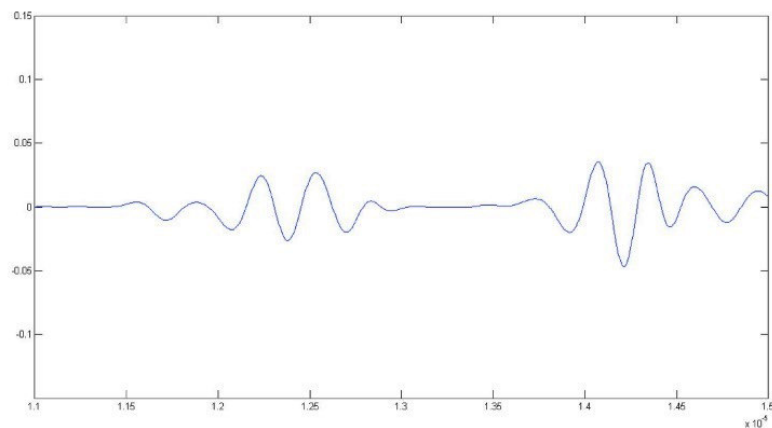
A Figura 16 apresenta o campo de deslocamentos verificado no instante em que a onda alcança uma falta de penetração com largura de 2 mm e altura de 4 mm. A Figura 17 apresenta o A-scan (amplitude de deslocamento no tempo) produzido após a simulação da propagação da onda em um meio que contém a respectiva descontinuidade.

Figura 16 – Campo de deslocamentos em uma peça contendo defeito do tipo Falta de Penetração (2mm X 4mm).



Fonte: Figura do próprio autor

Figura 17 – A-scan gerado durante simulação da propagação da onda por um meio contendo defeito do tipo Falta de Penetração (2mm X 4mm).



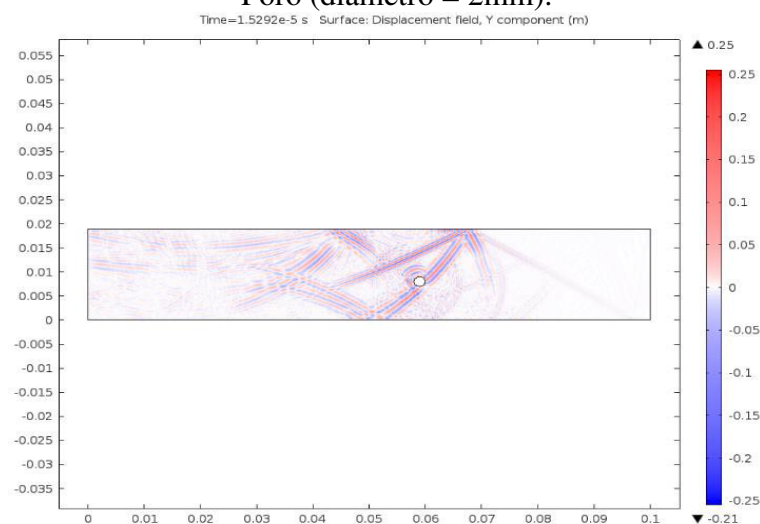
Fonte: Figura do próprio auto

3.1.2 Poro

Também foi modelado um sistema que apresenta as propriedades mecânicas do aço carbono comum e contém uma descontinuidade do tipo poro. A posição do poro foi escolhida aleatoriamente em torno da região da peça localizada entre os transdutores. Os poros podem assumir dois tamanhos: 1 mm ou 2 mm de diâmetro. Foram realizadas 18 simulações da propagação da onda para cada um dos possíveis tamanhos de poro, totalizando 36 simulações.

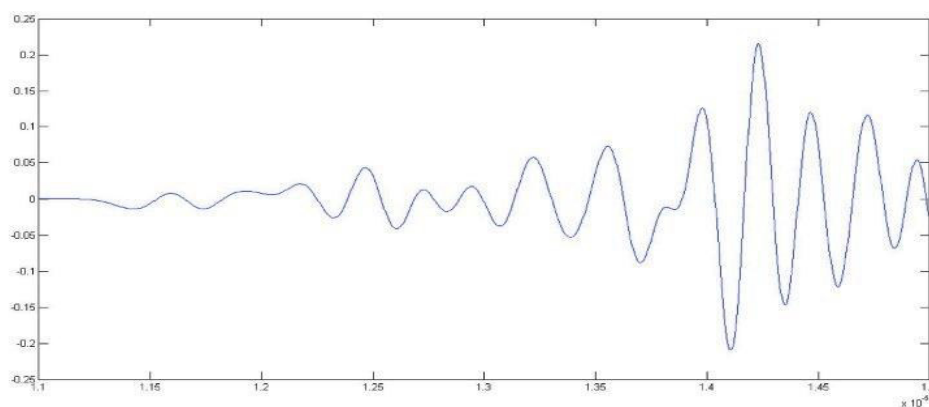
Nas figuras 18 e 19, temos uma representação do campo de deslocamentos de uma das simulações com 2 mm de diâmetro do tipo poro, a Figura 18 mostra o momento que a onda alcança a descontinuidade e a 19 o A-scan gerado por esse defeito.

Figura 18 – Campo de deslocamentos em uma peça contendo uma descontinuidade do tipo Poro (diâmetro = 2mm).



Fonte: Figura do próprio autor

Figura 19 – A-scan gerado durante simulação da propagação da onda por um meio contendo defeito do tipo Poro (diâmetro = 2mm).

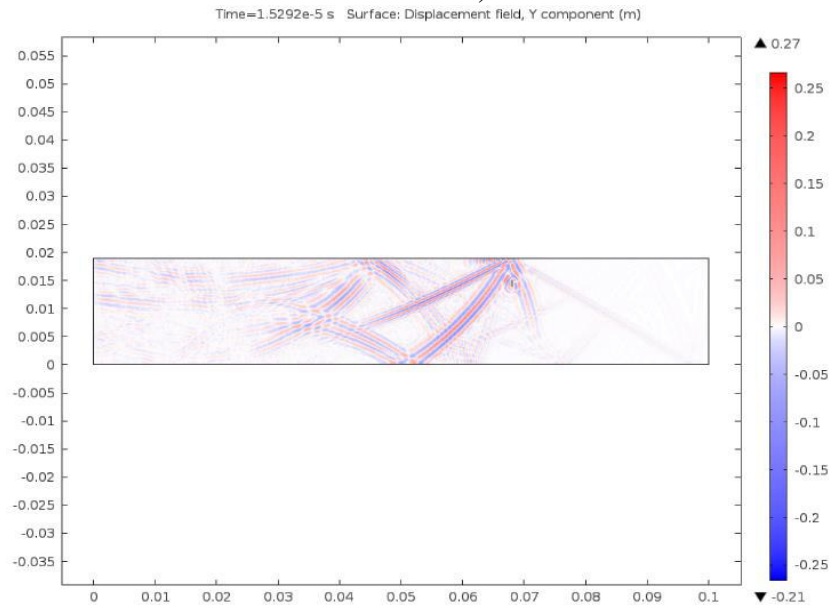


Fonte: Figura do próprio autor

3.1.3 Trinca

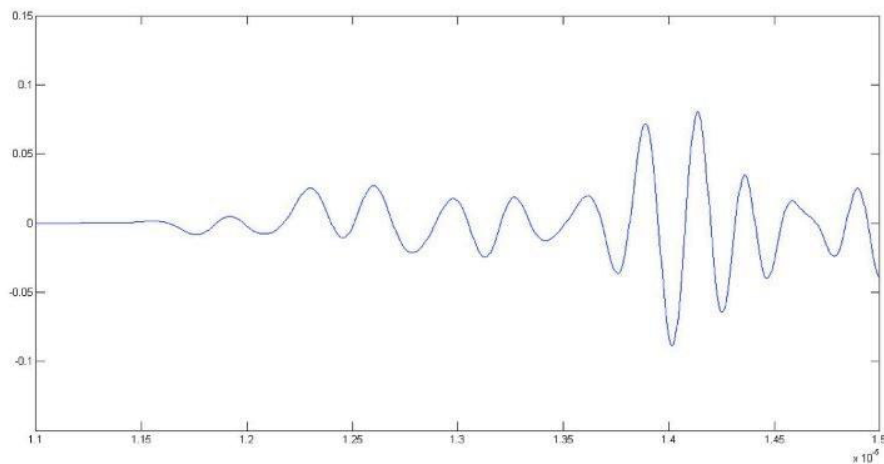
Foram realizadas um total de 36 simulações da trinca, alterando tanto as dimensões quanto a posição da descontinuidade no interior da peça. Foi estabelecido que a altura da trinca é sempre 10 vezes sua largura, e as dimensões estudadas são: 0.1 mm x 1.0 mm; 0.2 mm x 2.0 mm; 0.3 mm x 3.0 mm e 0.4 mm x 4.0 mm. As Figuras 20 e 21 representam o instante em que a onda alcança uma trinca de 0.1 mm por 1.0 mm e o respectivo A-scan registrado pelo receptor posicionado na superfície da peça.

Figura 20 – Campo de deslocamentos em uma peça contendo defeito do tipo Trinca (0.1mm X 1.0mm).



Fonte: Figura do próprio auto

Figura 21 – A-scan gerado durante simulação da propagação da onda por um meio contendo defeito do tipo Trinca (0.1mm X 1.0mm).



Fonte: Figura do próprio autor

Foram modeladas 36 descontinuidades para cada um dos três tipos de defeitos de soldagem, totalizando 108 simulações, e produzido um A-scan para cada simulação.

O resultado da análise de componentes principais dos sinais simulados, bem como a comparação entre a classificação dos sinais obtida pela regra da média da classe mais próxima aplicada sobre o resultado da análise de componentes principais e da classificação obtida pela análise de agrupamento, baseada na técnica k-médias, são apresentados na seção 4 – APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS, a seguir.

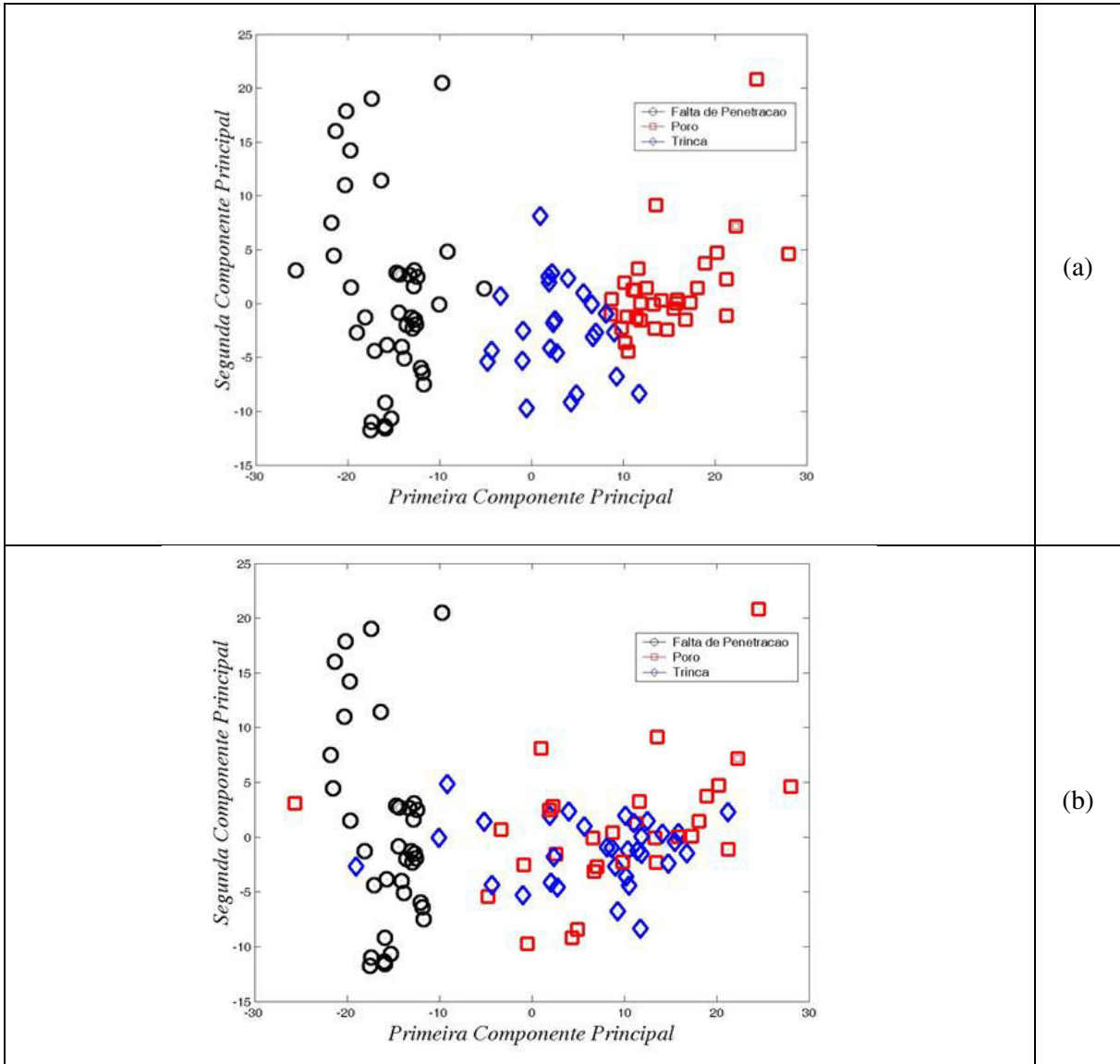
4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

4.1 Análise de Componentes Principais (PCA)

A Análise de Componentes Principais foi utilizada com três propósitos principais: reduzir a dimensionalidade de cada sinal mantendo o máximo de informações possível; facilitar a visualização dos dados e revelar alguma separação dos mesmos. A técnica foi aplicada com o auxílio do programa Matlab.

Primeiramente é feita a normalização do tipo Média-Zero nos dados, também chamada normalização Unidade-Variante, obtida através da subtração da média do conjunto de ' p ' amostras (sinais) de dimensão ' n ' (número de atributos da amostra ou tamanho do sinal), seguida da divisão pelo desvio padrão do conjunto. Em seguida é necessário construir a matriz de covariância (com dimensão ' $p \times p$ '). Posteriormente, calcula-se os auto valores e os auto vetores dessa matriz de covariância. Por fim, cria-se uma nova matriz, chamada matriz de transformação, em que os auto vetores estão reordenados de acordo com a ordem decrescente dos auto valores associados. É a partir da multiplicação dos dados de entrada pela matriz de transformação que se obtém os novos dados com dimensão menor. A quantidade de informações preservadas no conjunto de dados redimensionado depende do número de componentes principais consideradas e é dada pela razão entre o somatório do número de auto valores considerados pelo somatório de todos os auto valores da matriz de covariância dos dados de entrada. A Figura 22(a) exibe o resultado da Análise de Componentes Principais dos 108 sinais ultrassônicos simulados (3 classes x 36 sinais por classe) em comparação com a classificação conhecida desses mesmos dados (figura 22(b)). Evidentemente, apenas as duas primeiras componentes podem ser exibidas num gráfico bidimensional.

Figura 22 – (a) resultado do PCA dos 108 sinais ultrassônicos simulados; (b) classificação conhecida dos dados.



Fonte: Figura do próprio autor

Matrizes de confusão foram calculadas pela aplicação da regra da média da classe mais próxima sobre um número prédefinido de componentes principais. A regra da média da classe mais próxima consiste em calcular a média dos valores de todos os dados de entrada que compõe cada classe, medir a distância entre as médias das classes a todos os dados de entrada, e associar cada um dos dados de entrada à classe cuja média está mais próxima.

A tabela 4 exibe a matriz de confusão obtida pela aplicação da regra da média da classe mais próxima nas duas primeiras componentes. Neste caso, o quociente entre a soma dos dois maiores auto valores e o somatório de todos os auto valores da matriz de covariância dos sinais simulados é igual a 0.454. Isso significa que as duas primeiras componentes principais contêm 45.4% das informações do conjunto original.

Tabela 4 – Matriz de confusão obtida pela aplicação da regra da média da classe mais próxima considerando apenas as 2 primeiras componentes principais dos sinais simulados.

	Falta de Penetração	Poro	Trinca
Classificado como Falta de Penetração	100.0000	2.7778	13.8889
Classificado como Poro	0	52.7778	52.7778
Classificado como Trinca	0	44.4444	33.3333

Fonte: Tabela do próprio autor.

Observa-se pela matriz de confusão que 100% dos dados associados à falta de penetração foram corretamente classificados como falta de penetração. Aproximadamente 53% dos dados pertencentes a classe poro foram reconhecidos como poro, e apenas cerca de 33% dos dados referentes a trincas foram corretamente identificados como trinca. A taxa média de acerto na classificação foi de 62.037%.

Caso sejam considerados os 20 (vinte) maiores auto valores, cuja soma alcança 505.88084 [adimensional], tem-se 97.86% do somatório de todos os auto valores. Deste modo, esta transformação reduz a dimensionalidade dos dados de 512 para 20 enquanto a maioria das informações (97.86%) contidas no conjunto de entrada é preservada.

Uma nova matriz de confusão foi calculada pela aplicação da regra da média da classe mais próxima sobre as vinte principais componentes e é apresentada pela Tabela 5. A nova matriz de confusão revela que 100% dos dados relacionados a falta de penetração continuam sendo corretamente classificados. A taxa de acerto na classificação das classes poro e trinca alcançou 63.88% e 41.67, respectivamente, e a taxa média de acerto passou de 62.0370% para 68.5185%.

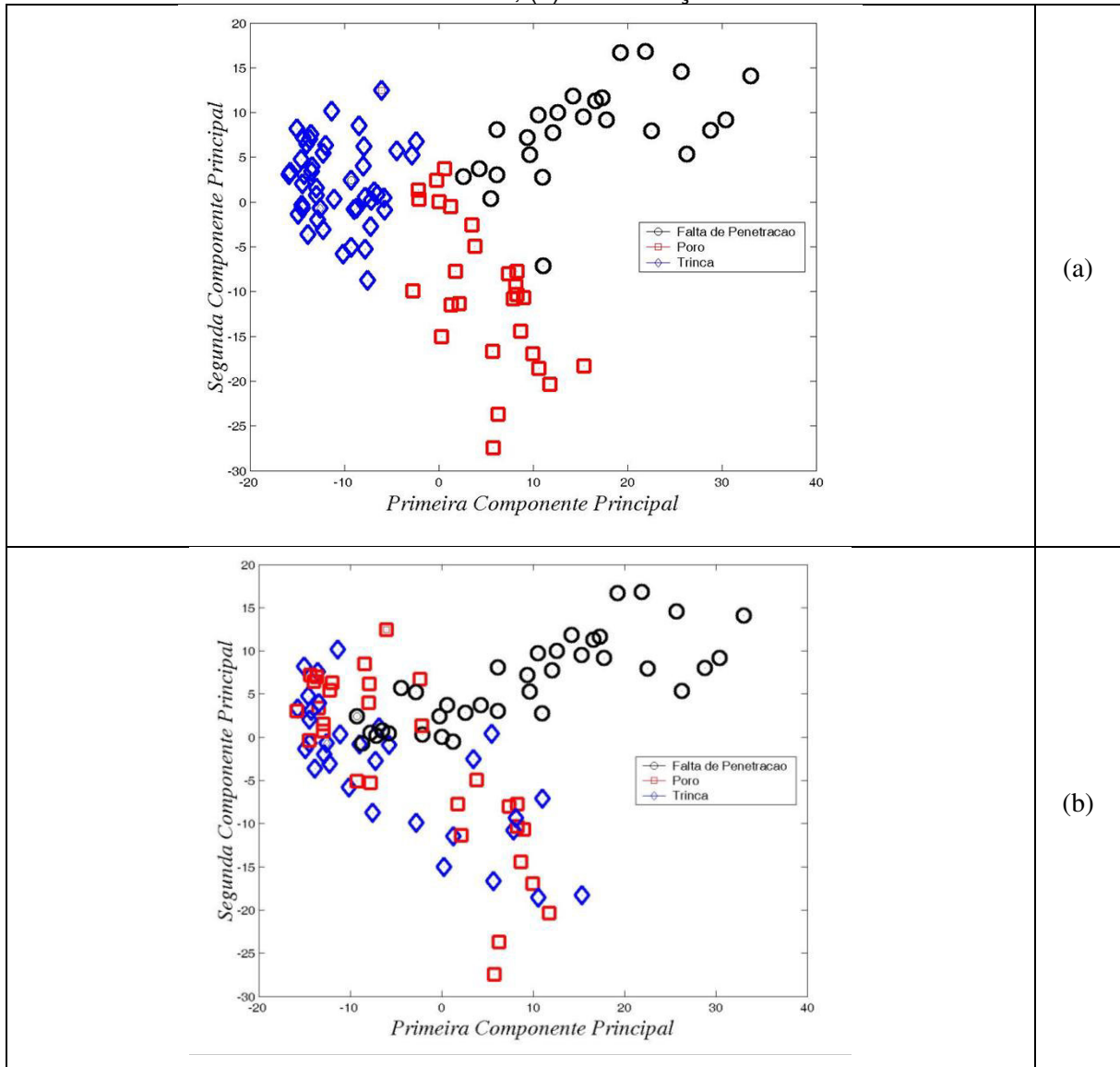
Inspirado em um trabalho de reconhecimento de padrão de defeitos de soldagem utilizando redes neurais (Moura et al, 2005) que concluiu que o simples cálculo da envoltória do sinal de TOFD melhora a taxa de acerto na classificação dos sinais, experimentou-se o uso da envoltória dos sinais simulados como uma etapa de pré-processamento. A envoltória foi obtida pela aplicação do filtro *Savitzky-Golay*. A Figura 23(a) apresenta a projeção das duas primeiras componentes principais da envoltória dos dados simulados e a classificação conhecida desses mesmos dados, Figura 23(b).

Tabela 5 – Matriz de confusão obtida pela aplicação da regra da média da classe mais próxima considerando as 20 (vinte) primeiras componentes dos sinais simulados.

	Falta de Penetração	Poro	Trinca
Classificado como Falta de Penetração	100.0000	2.7778	13.8889
Classificado como Poro	0	63.8889	44.4444
Classificado como Trinca	0	33.3333	41.6667

Fonte: Tabela do próprio autor.

Figura 23 – (a) apresenta a projeção das duas primeiras componentes principais da envoltória dos dados simulados; (b) classificação conhecida dos dados.



Fonte: Tabela do próprio autor.

As tabelas 6 e 7 apresentam as matrizes de confusão obtidas pela aplicação da regra da média da classe mais próxima aos resultados da PCA considerando, respectivamente, 2 (duas) e 20 (vinte) primeiras componentes principais da envoltória dos sinais ultrassônicos simulados.

Tabela 6 – Matriz de confusão obtida pela aplicação da regra da média da classe mais próxima considerando apenas as 2 (duas) primeiras componentes principais da ENVOLTÓRIA DOS SINAIS ultrassônicos simulados.

	Falta de Penetração	Poro	Trinca
Classificado como Falta de Penetração	63.8889	0	8.3333
Classificado como Poro	13.8889	47.2222	25.0000
Classificado como Trinca	22.2222	52.7778	66.6667

Fonte: Tabela do próprio autor.

Tabela 7 – Matriz de confusão obtida pela aplicação da regra da média da classe mais próxima considerando as 20 (vinte) primeiras componentes da ENVOLTÓRIA DOS SINAIS ultrassônicos simulados.

	Defeito do tipo Falta de Penetração	Defeito do tipo Poro	Defeito do tipo Trinca
Classificado como Falta de Penetração	83.3333	0	8.3333
Classificado como Poro	0	58.3333	16.6667
Classificado como Trinca	16.6667	41.6667	75.0000

Fonte: Tabela do próprio autor,

Em ambos os casos percebe-se que a utilização da envoltória dos sinais ocasionou uma piora significativa na classificação dos sinais pertencentes a classe falta de penetração e uma grande melhora na classificação dos sinais relacionados a trinca. Observa-se também que esse pré-processamento ocasionou uma queda de 62.04% para 59.26% na taxa média de acerto quando apenas as 2 primeiras componentes são utilizadas. Esse resultado parece ser incoerente com os obtidos no trabalho que inspirou o uso da envoltória dos sinais. No entanto, é importante lembrar que o trabalho citado empregou classificadores não-lineares implementados por uma rede neural artificial, uma ferramenta mais complexa que a PCA. Por outro lado, a utilização

de 20 primeiras componentes elevou de 68.52% para 72.22% a taxa média de acerto. Vale ressaltar que as duas primeiras componente principais contêm apenas 45.18% das informações do conjunto de entrada, contra 98.55% preservadas nas vinte primeiras componentes.

4.2 K-médias

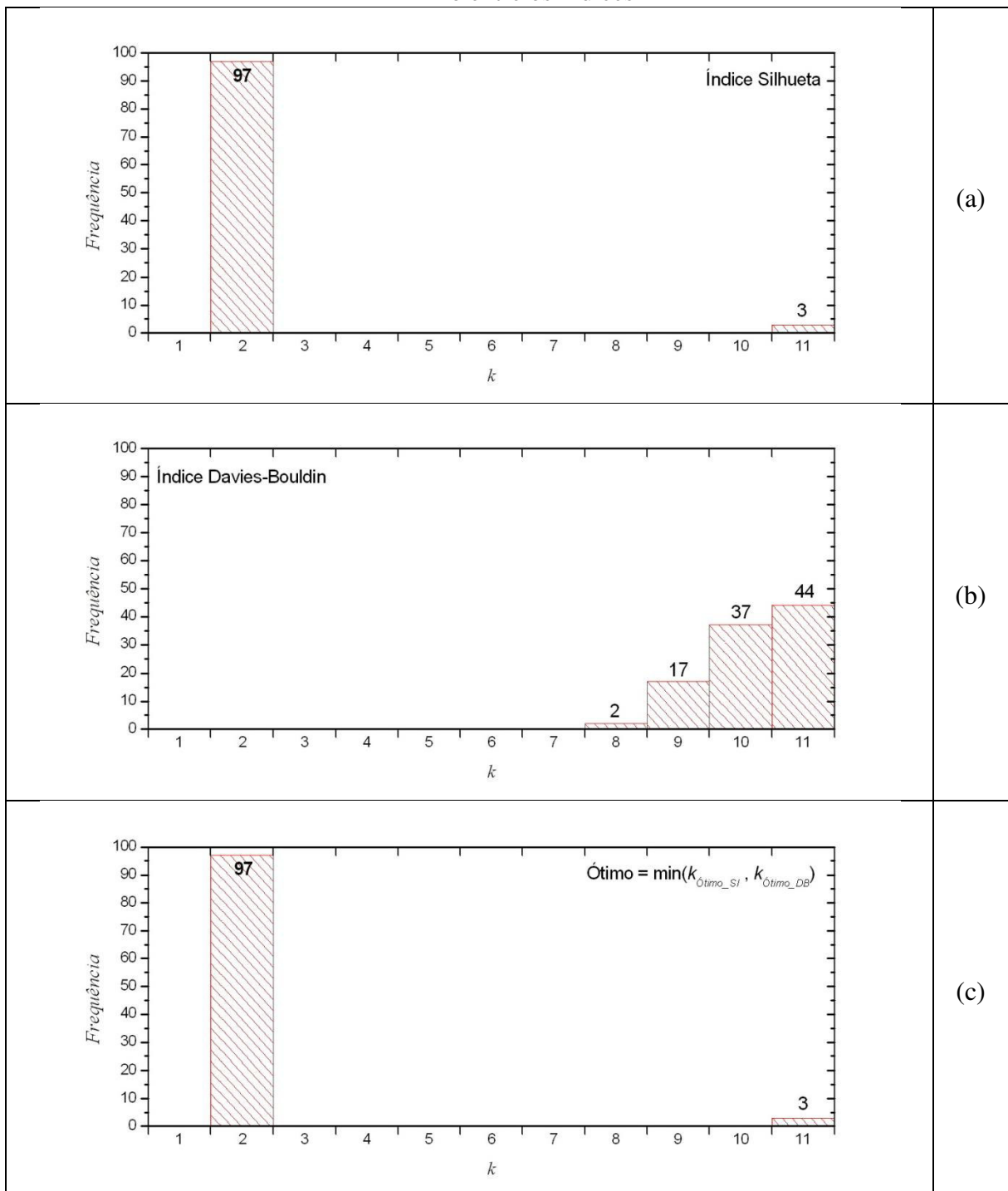
A técnica de agrupamento k-médias, apresentada na seção 2.8 da revisão bibliográfica, foi utilizada para agrupar os dados em k grupos (onde k varia de 1 até a raiz do número de dados) e os índices Silhueta e Davies-Bouldin podem nos informar em quantas classes é melhor agrupar esses dados.

Inicialmente são apresentados os resultados obtidos pelo agrupamento dos sinais ultrassônicos simulados. Em seguida, utilizando a mesma metodologia, são apresentados os resultados da análise de agrupamento aplicada às envoltórias desses sinais e finalmente os resultados obtidos pelo agrupamento das componentes principais dos sinais.

4.2.1 Análise de agrupamento dos sinais ultrassônicos simulados

Para determinar o número ótimo de grupos, k, o algoritmo testou todos os valores entre 1 e o valor dado pela raiz quadrada do número total de dados de entrada, ou seja, o tamanho do conjunto. Lembrando que neste trabalho foram simulados 36 sinais para cada uma das três discontinuidades estudadas, totalizando 108 sinais. Assim, o algoritmo experimentou todos os valores entre 1 e 11. A Figura 24 apresenta os histogramas resultantes da análise de agrupamento, pela utilização de um algoritmo k-médias, que surgem o número ótimo de agrupamentos, baseado no máximo valor do índice Silhueta e no mínimo valor do índice Davies-Bouldin dos agrupamentos. Após uma centena de execuções, e com base no máximo valor do índice Silhueta, o algoritmo sugeriu 97 vezes o número ótimo de grupos é 2. (Figura 24(a)). Após uma centena de execuções, o índice Davies-Bouldin sugeriu: 2 vezes 8 *clusters*; 17 vezes 9 *clusters*; 37 vezes 10 *clusters*; e 44 vezes 11 *clusters* (Figura 24(b)). Entretanto, seguindo o comumente utilizado e descrito na literatura, adotou-se o menor valor de k ($k = \min(k_{SI_ótimo}, k_{DB_ótimo})$), apenas 3 vezes o número igual a 11 foi sugerido como ótimo e o número de grupos sugeridos mais frequentemente é 2. Percebe-se que nenhuma vez foi sugerido um único grupo, pois não faz sentido criar um único grupo com todo o conjunto de dados, ou outro número de grupos.

Figura 24 – Histograma do número ótimo de agrupamentos sugerido para os sinais ultrassônicos simulados: (a) baseado no índice Silhueta; (b) baseado no índice; (c) baseado no mínimo entre os índices

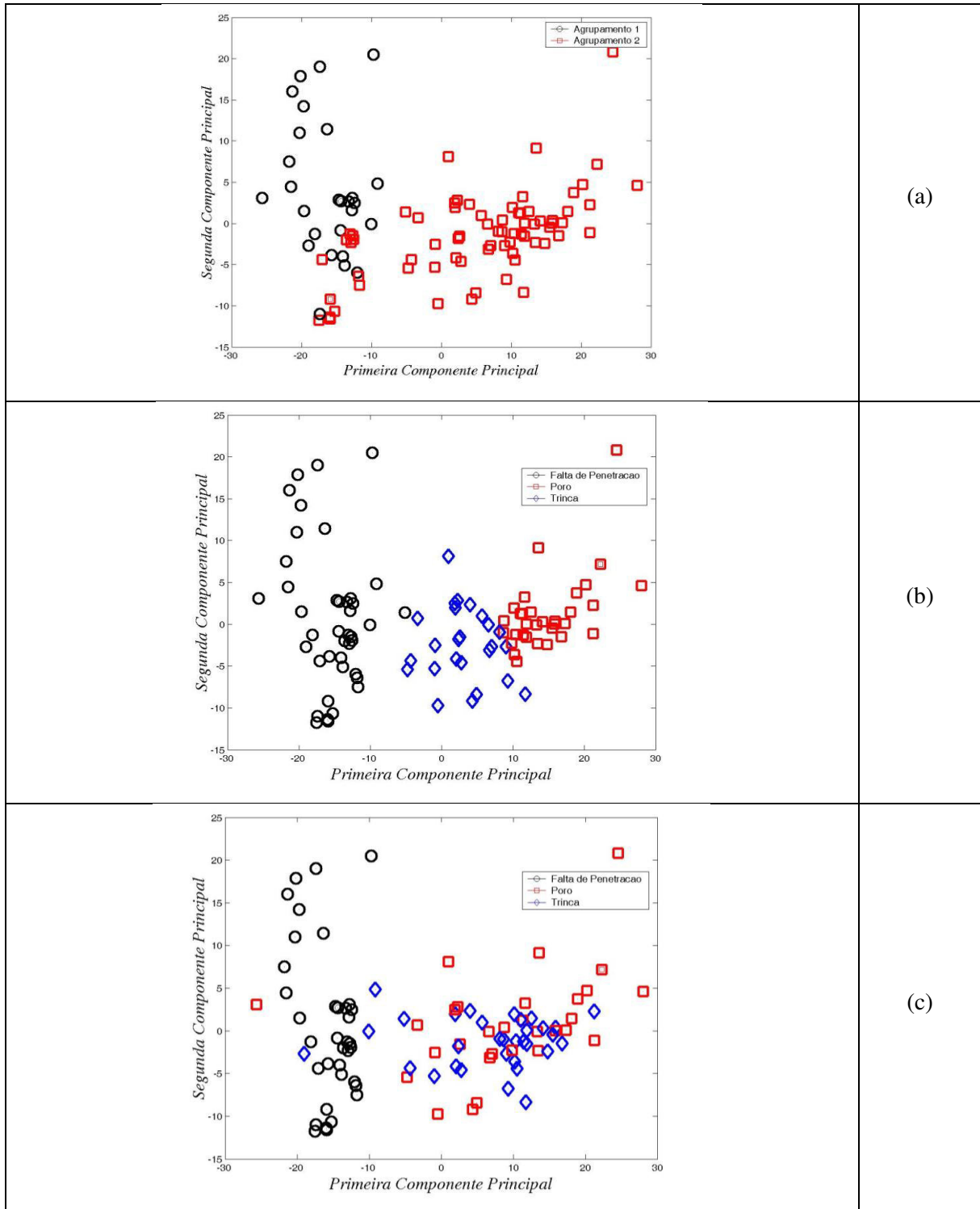


Fonte: Figura do próprio autor.

A Figura 25(a) mostra como o k-médias sugere agrupar os sinais ultrassônicos simulados. Para facilitar a avaliação, o resultado da PCA (exibida pela Figura 22) desses mesmos sinais é rerepresentado pela Figura 25(b). É preciso salientar que o k-médias não utilizou os dados da PCA, apenas foi mostrado na figura para facilitar a comparação dos

resultados. A figura 25(a) apenas destaca quais dados foram associados a cada *cluster*. A classificação conhecida (gabarito) dos sinais é rerepresentada na figura 25(c).

Figura 25 – (a) resultado da análise de agrupamento; (b) resultado da PCA; (c) classificação conhecida dos sinais simulados.



Fonte: Figura do próprio autor.

A Figura 25(a) revela que um dos agrupamentos é constituído basicamente pelos sinais de falta de penetração, e o segundo agrupamento reuni os sinais das classes poro, trinca e alguns do tipo falta de penetração. A maioria dos sinais ($23/36 = 64\%$) pertencentes a falta de penetração foram associados ao agrupamento 1. Um único sinal pertencente a classe poro e 4 sinais pertencentes a classe trinca foram também atribuídos ao grupo 1, que acumulou 28 elementos. O restante dos sinais pertencentes a falta de penetração ($13/36 = 36\%$), juntamente com a grande maioria dos sinais pertencentes a poro (35 elementos) e trinca (32 elementos) foram atribuídos ao agrupamento 2, que acumulou 80 elementos. Esse resultado encontra-se também sintetizado na tabela 8.

Tabela 8 – Agrupamento, ou pseudo-classificação, dos sinais sugerida pela k-médias.

	Falta de Penetração	Poro	Trinca
Grupo 1	23	1	4
Grupo 2	13	35	32

Fonte: Tabela do próprio autor.

A falta de penetração, introduzida na seção 2.1.1, é uma descontinuidade caracterizada pela insuficiência de metal na raiz da solda e, conseqüentemente, sua posição no centro do cordão de solda facilita sua localização. Evidentemente, a modelagem desse tipo de descontinuidade levou em consideração a posição de sua ocorrência em relação ao centro do cordão. Por outro lado, não existem sítios específicos ou regiões preferenciais para formação de poros (seção 2.1.2) e trincas (seção 2.1.3), que podem aparecer em posições aleatórias. Adicionalmente, as descontinuidades do tipo falta de penetração foram modeladas com altura entre 4.0 mm e 9.0 mm e largura variando de 2.0 mm a 7.0 mm, ao passo que a maior descontinuidade do tipo trinca considerada nesse trabalho apresenta dimensão de 4.0 mm de altura por 0.4 mm de largura, enquanto o diâmetro do poro alterna entre 1.0 mm e 2.0 mm. Diante disso, podemos supor que os fatos da falta de penetração sempre ocorrer no centro do cordão e sido modelada com dimensões significativamente maiores, podem explicar o motivo do algoritmo k-médias reunir a maioria dessas descontinuidades em um grupo diferente do tipo poro e trinca, que por sua vez tem dimensões parecidas e a aleatoriedade de suas posições.

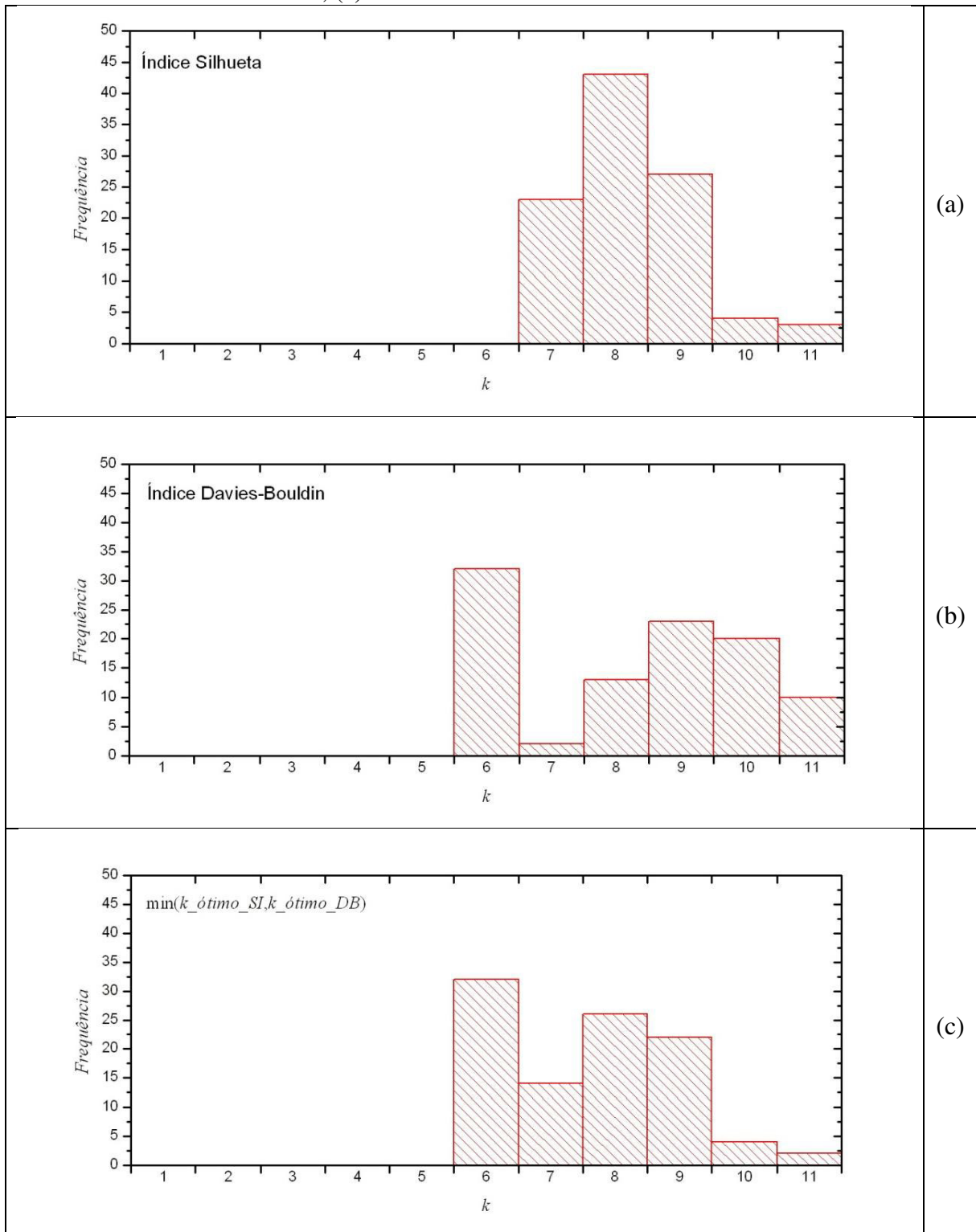
4.2.2 Análise de agrupamento das envoltórias dos sinais ultrassônicos simulados

A metodologia utilizada na análise dos sinais ultrassônicos simulados foi repetida para analisar as envoltórias desses sinais, e os resultados produzidos mostram que o algoritmo sugere um número de grupos diferente para subdividir o conjunto de entrada. Tais resultados são apresentados a seguir de forma mais detalhada.

Os histogramas das Figuras 26(a) e 26(b) mostram, após uma centena de repetições do k-médias, a frequência com que diferentes quantidades de agrupamentos é sugerida como ótimo pelos índices Silhueta e Davies-Bouldin, respectivamente, para que o conjunto de dados formados pelas envoltórias dos sinais simulados sejam subdivididos. De acordo com a figura 26(c), que mostra a frequência com que cada número de grupos é sugerida como ótimo a partir da função $\text{mim}(k_{SI_ótimo}, k_{DB_ótimo})$, a quantidade de agrupamentos sugerida mais frequentemente é seis.

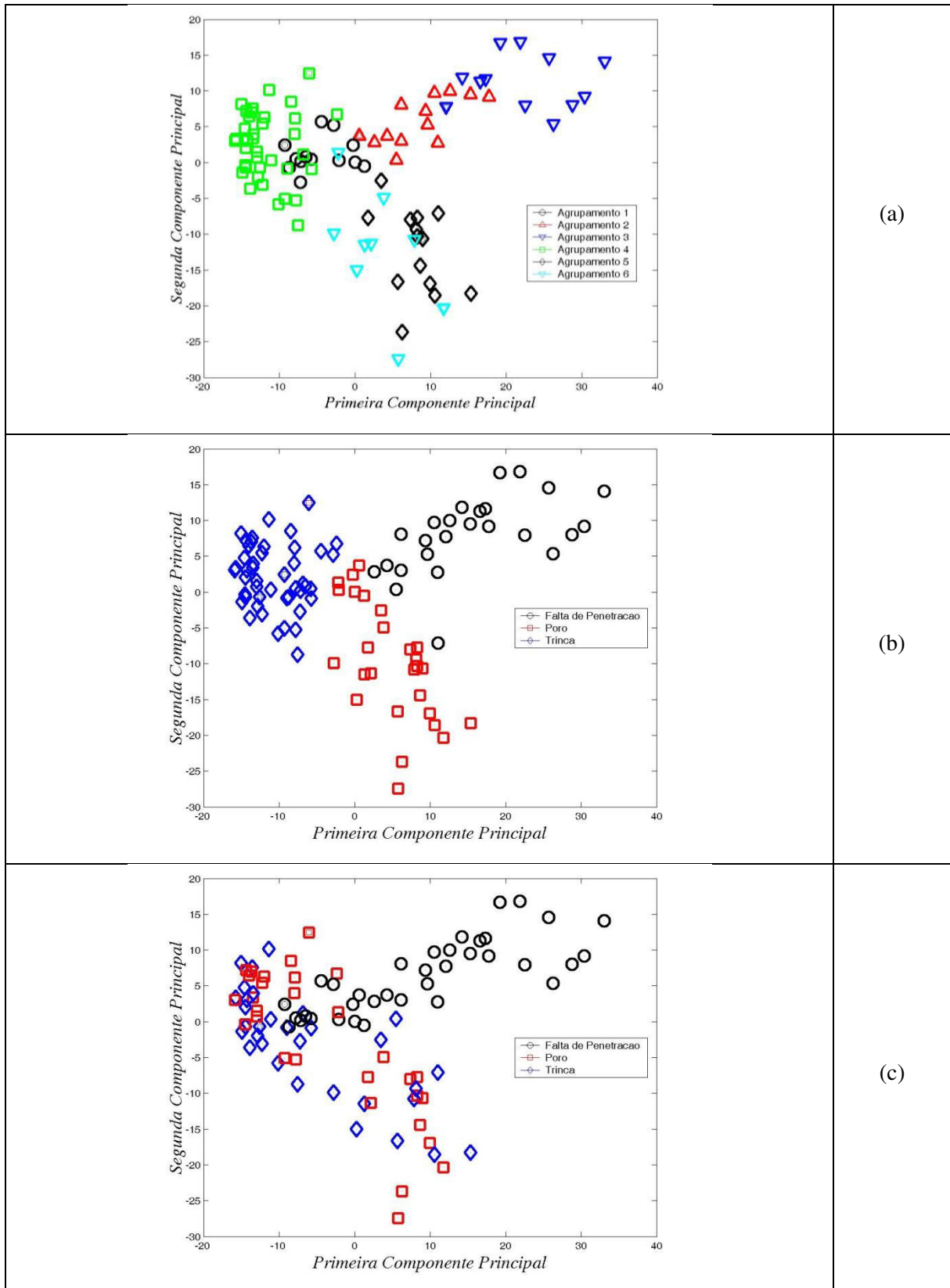
Uma vez determinada a quantidade de subconjuntos recomendada pela análise de agrupamento para dividir o conjunto de entrada ($k = \text{seis}$), cada um dos dados do conjunto foi atribuído a um dos grupos. Um exame detalhado da distribuição dos dados pertencentes a classe falta de penetração revelou a existência de um padrão que corresponde a $\{1;1;2;2;3;3;1;1;2;2;3;3;1;1;2;2;3;3;1;1;2;2;3;3;1;1;2;2;3;3;1;1;2;2;3;3\}$, de modo que as envoltórias dos sinais de falta de penetração foram subdivididos em três subconjuntos de acordo com a profundidade da descontinuidade (pouco profunda; média e profunda) e independente de sua largura. Assim, o primeiro agrupamento reúne todas as envoltórias dos sinais de falta de penetração com 2.0 e 3.0 mm de profundidade, enquanto aquelas com 4.0 e 5.0 mm de profundidade foram juntadas no grupo 2. Adicionalmente, a cada um desses grupos foi atribuído um sinal pertencente a classe trinca. Todas as envoltórias dos sinais de falta de penetração com 6.0 e 7.0 mm de profundidade foram associadas ao 3º agrupamento. As outras duas classes de envoltórias, poro e trinca, foram subdivididas em 3 agrupamentos. Ao grupo 4 foram atribuídas 19 envoltórias de sinais pertencentes a classe poro e 23 correspondentes a trinca. 10 envoltórias de poro mais 7 de trinca compõem o grupo 5. Por fim, o sexto e menor grupo é composto por 7 das envoltórias correspondentes a poro mais 4 pertencentes a classe trinca. Esse resultado está sintetizado pela tabela 9 e retratado pela Figura 27.

Figura 26 – Histograma do número ótimo de agrupamentos sugerido para as envoltórias dos sinais ultrassônicos simulados: (a) baseado no índice Silhueta; (b) baseado no índice Davies-Bouldin; (c) baseado no mínimo entre os índices.



Fonte: Figura do próprio autor.

Figura 27 – (a) resultado da análise de agrupamento; (b) resultado da PCA; (c) classificação conhecida das envoltórias dos sinais simulados.



Fonte: Figura do próprio autor.

Tabela 9 – Agrupamento, ou pseudo-classificação, das envoltórias dos sinais sugerida pelo algoritmo k-médias.

	Falta de Penetração	Poro	Trinca
Grupo 1	12	0	1
Grupo 2	12	0	1
Grupo 3	12	0	0
Grupo 4	0	19	23
Grupo 5	0	10	7
Grupo 6	0	7	4

Fonte: Tabela do próprio autor.

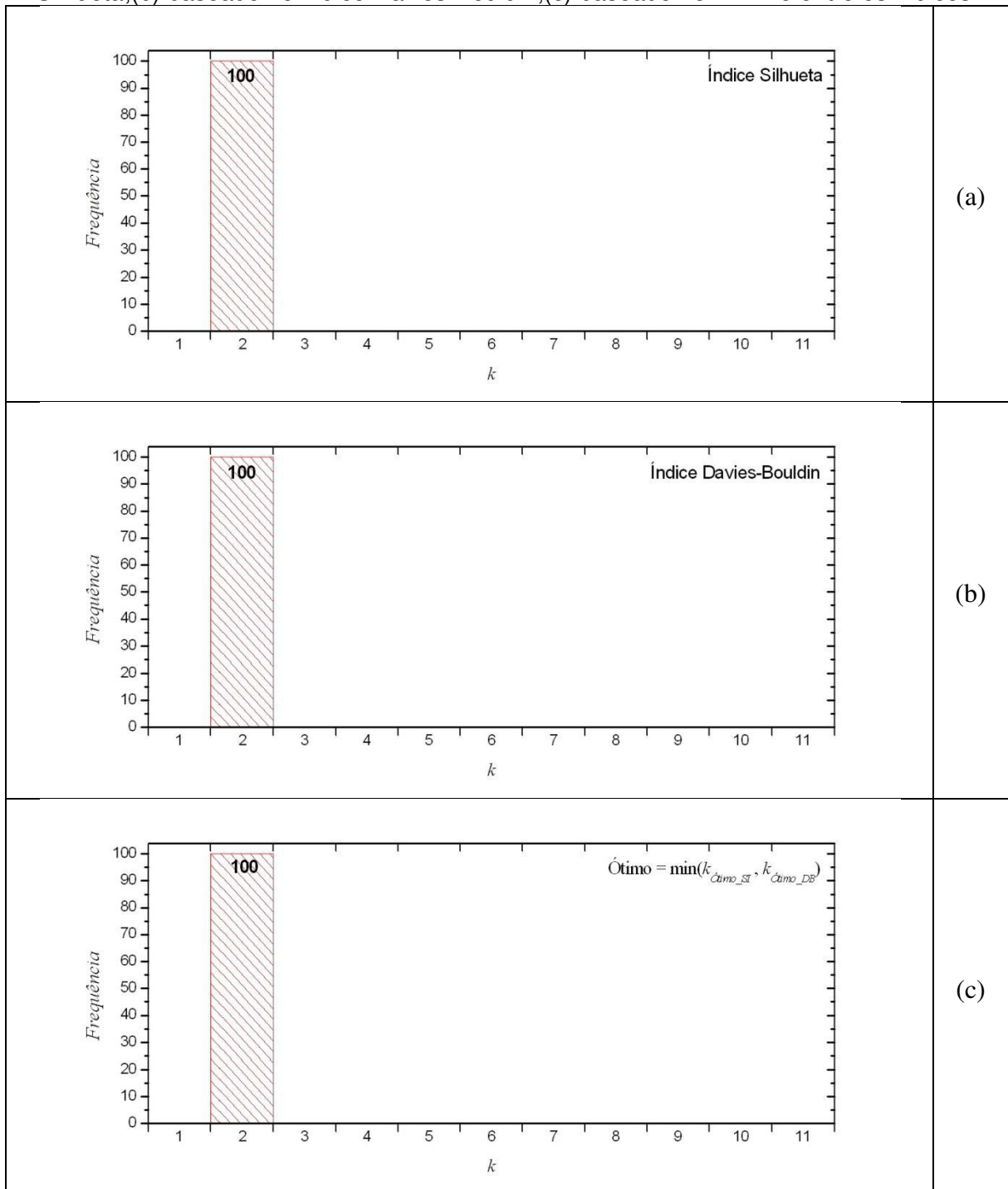
É importante destacar novamente que a análise de agrupamento não foi realizada sobre as componentes principais das envoltórias dos sinais, e que o resultado da análise de agrupamento (Figura 27(a)) é apresentado juntamente com o resultado da PCA (Figura 27(b)) e a classificação conhecida (Figura 27(c)) das envoltórias dos sinais apenas para facilitar a visualização de quais dados foram associados a cada *cluster*.

4.2.3 Análise de agrupamento das componentes principais dos sinais ultrassônicos simulados

Após a análise de agrupamento realizada sobre os sinais simulados e suas envoltórias, propôs-se verificar o efeito da combinação do algoritmo k-médias com a análise de componentes principais na taxa de acerto de classificação das três classes de sinais ultrassônicos de defeitos modelados, ou seja, o algoritmo analisou um número pré-definido de componentes principais dos sinais.

Os histogramas produzidos após 100 repetições completas da análise k-médias realizada sobre apenas as duas primeiras componentes principais dos sinais ultrassônicos simulados revelam que a quantidade de grupos sugerida pelos índices Silhueta (Figura 28(a)) e Davies-Bouldin (Figura 28(b)) é a mesma e igual a dois. Conseqüentemente, o número de agrupamentos sugerido como ótimo, selecionado a partir do $\min(k_{SI_ótimo}, k_{DB_ótimo})$, também é dois (Figura 28(c)).

Figura 28 – Histograma do número ótimo de agrupamentos sugerido para as 2(duas) primeiras componentes principais dos sinais ultrassônicos simulados:(a) baseado no índice Silhueta;(b) baseado no índice Davies-Bouldin;(c) baseado no mínimo entre os índices.

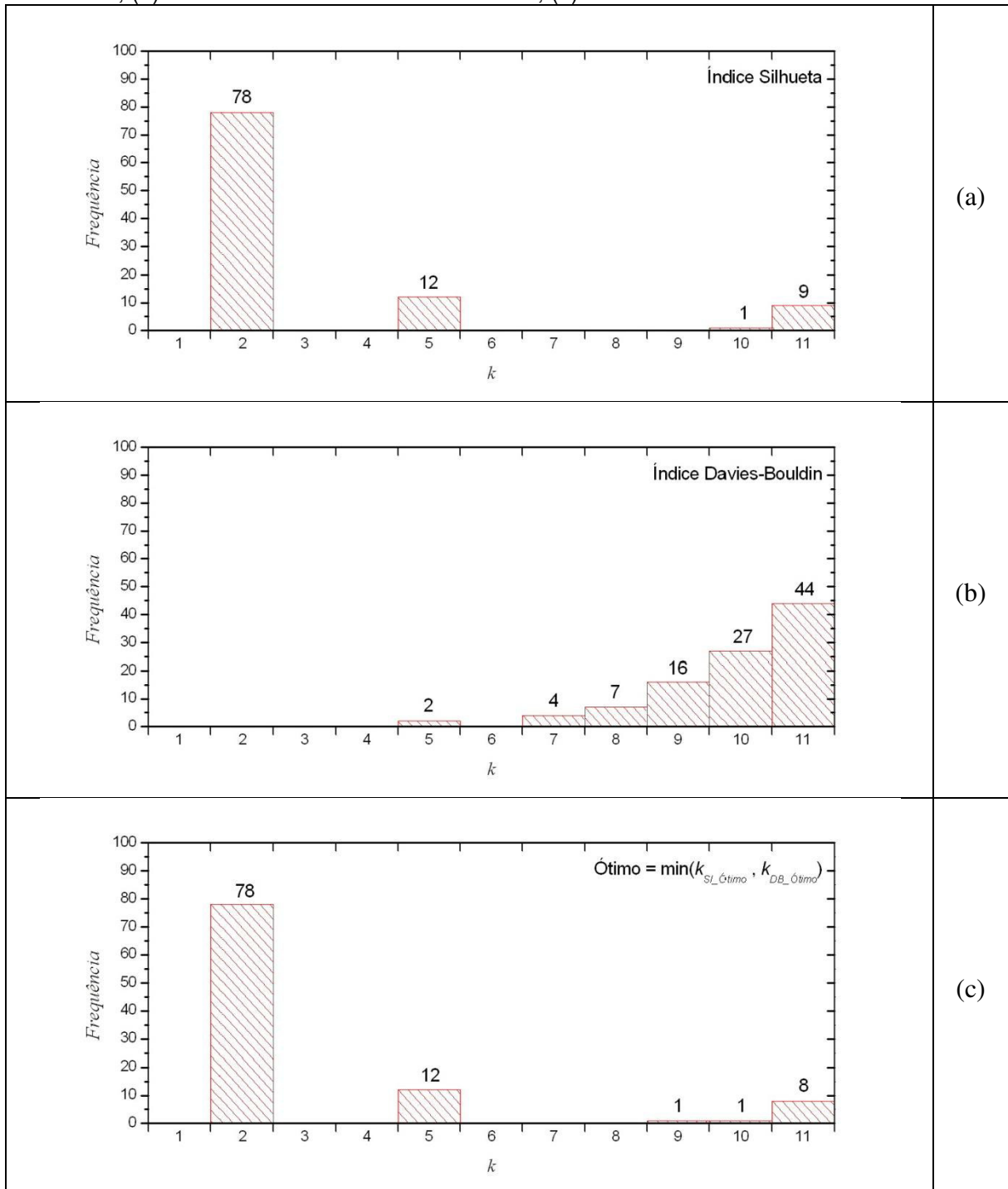


Fonte: Figura do próprio autor

Igualmente, a Figura 29 exibe a frequência com que número ótimo de agrupamentos é sugerido quando o mesmo procedimento foi utilizado para analisar as 20 (vinte) primeiras componentes principais. Existe uma diferença significativa entre o k ótimo sugerido pelos

índices Silhueta e Davies-Bouldin. Ainda assim, em 78% dos casos, o algoritmo sugere, baseado no $\min(k_{SI_ótimo}, k_{DB_ótimo})$, que a melhor forma de reorganizar os dados também é alcançada utilizando 2 grupos (ver Figura 29(c)).

Figura 29 – Histograma do número ótimo de agrupamentos sugerido para as 20 (vinte) primeiras componentes principais dos sinais ultrassônicos simulados: (a) baseado no índice Silhueta; (b) baseado no índice Davies-Bouldin; (c) baseado no mínimo entre os índices.



Fonte: Figura do próprio autor

Verificando, simultaneamente, o grupo ao qual cada elemento do conjunto de entrada foi associado e sua respectiva classificação conhecida foi possível construir a tabela 10 e constatar que: i) todos os 36 sinais relativos à falta de penetração, juntamente com 4 sinais da classe poro e 6 sinais da classe trinca, foram associados ao primeiro grupo; ii) a maioria dos sinais relativos à poro (88.9% = 32/36) e trinca (83.3% = 30/36) foram reunidos no segundo grupo, que não teve nenhum sinal de falta de penetração acrescentado. É importante destacar que, independente do número de componentes principais considerados, duas ou vinte primeiras componentes principais, o agrupamento proposto foi exatamente o mesmo no que se refere especificamente a identificação ($i = \{1,2,\dots,36\}$) dos sinais. A distribuição obtida pela análise de agrupamento das duas primeiras componentes principais dos sinais ultrassônicos simulados pode ser visualizada na Figura 30. Esse resultado é idêntico ao alcançado com a utilização das vinte primeiras componentes principais.

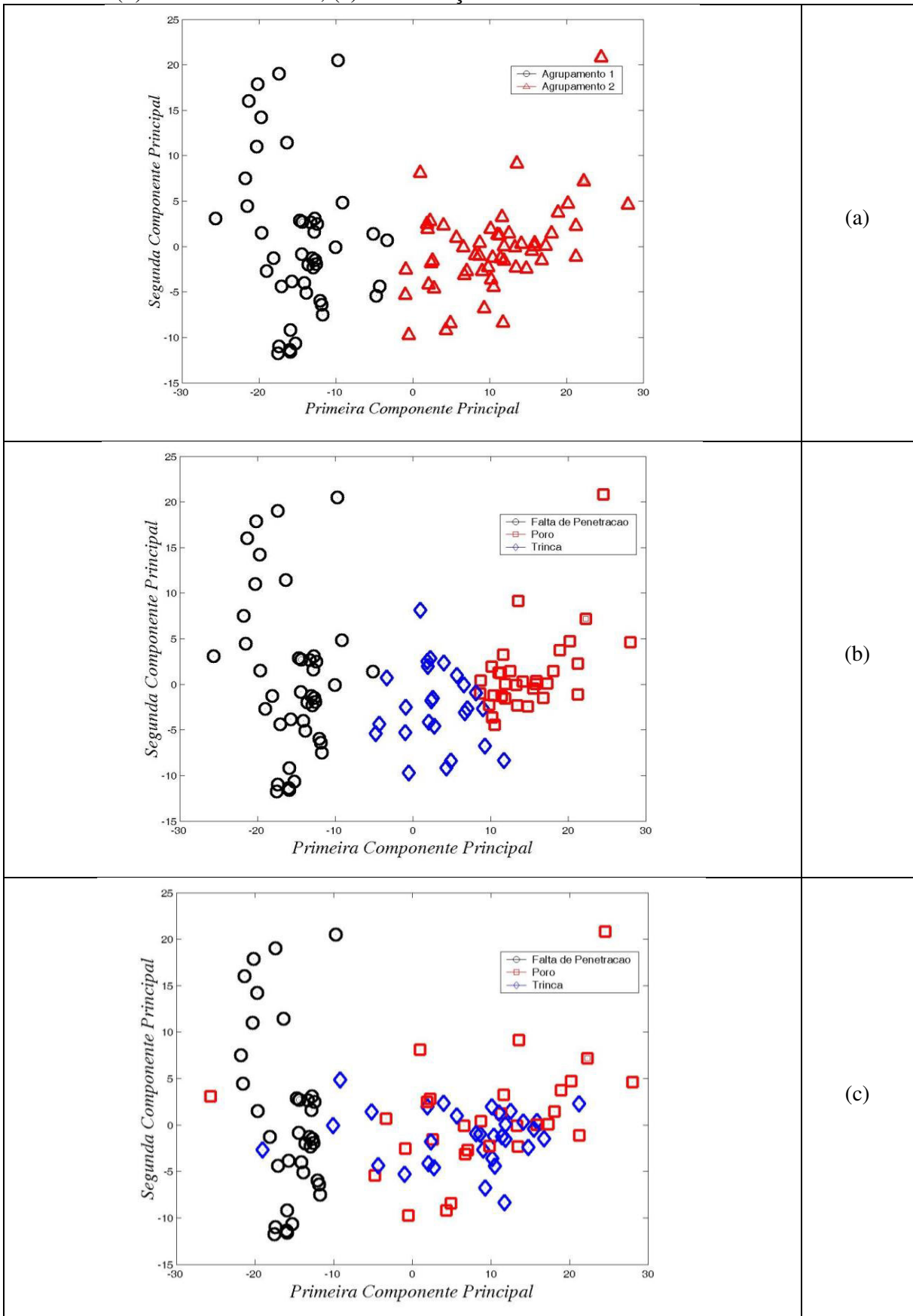
Tabela 10 – Agrupamento, ou pseudo-classificação, proposto pelo algoritmo k-médias após processamento das primeiras componentes principais dos sinais.

	Falta de Penetração	Poro	Trinca
Grupo 1	36	4	6
Grupo 2	0	32	30

Fonte: Tabela do próprio autor.

Acreditamos que a mesma hipótese levantada para justificar o agrupamento sugerido para os sinais ultrassônicos simulados pode ser empregada para explicar o agrupamento das duas (ou vinte) primeiras componentes principais desses sinais: i) a similaridade entre as pequenas dimensões de poros e trincas versus as dimensões maiores das falta de penetração; ii) a aleatoriedade das posições e poros e trincas versus o determinismo da posição da falta de penetração em relação ao centro do cordão.

Figura 30 – (a) resultado da análise de agrupamento das 2 primeiras componentes principais; (b) resultado da PCA; (c) classificação conhecida dos sinais simulados.

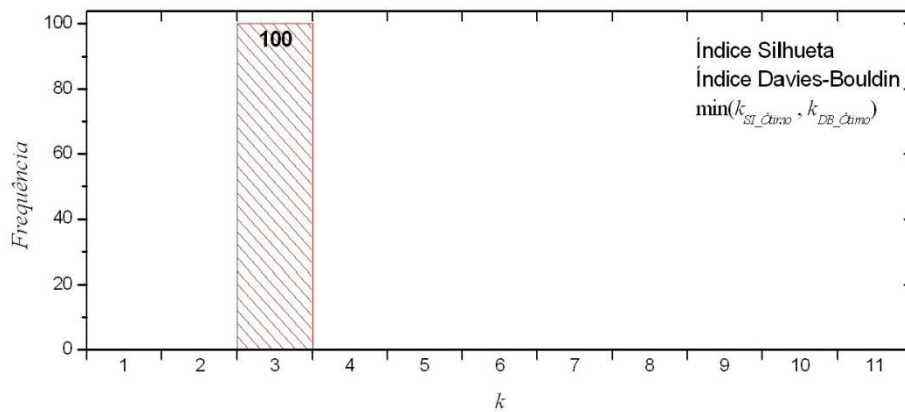


Fonte: Figura do próprio autor

4.2.4 Análise de agrupamento das componentes principais das envoltórias dos sinais ultrassônicos simulados

Quando o algoritmo k-médias realiza a análise das duas primeiras componentes principais das envoltórias dos sinais ultrassônicos de defeitos modelados, o número de grupos e a distribuição sugerida para os elementos do conjunto examinado divergem dos resultados anteriores. Em todas as cem execuções do algoritmo, os dois índices empregados estimaram que a melhor forma de distribuir os dados é dividi-los em três categorias (ver Figura 31).

Figura 31 – Histograma do número ótimo de agrupamentos sugerido para as 2 (duas) primeiras componentes principais das envoltórias dos sinais ultrassônicos simulados.

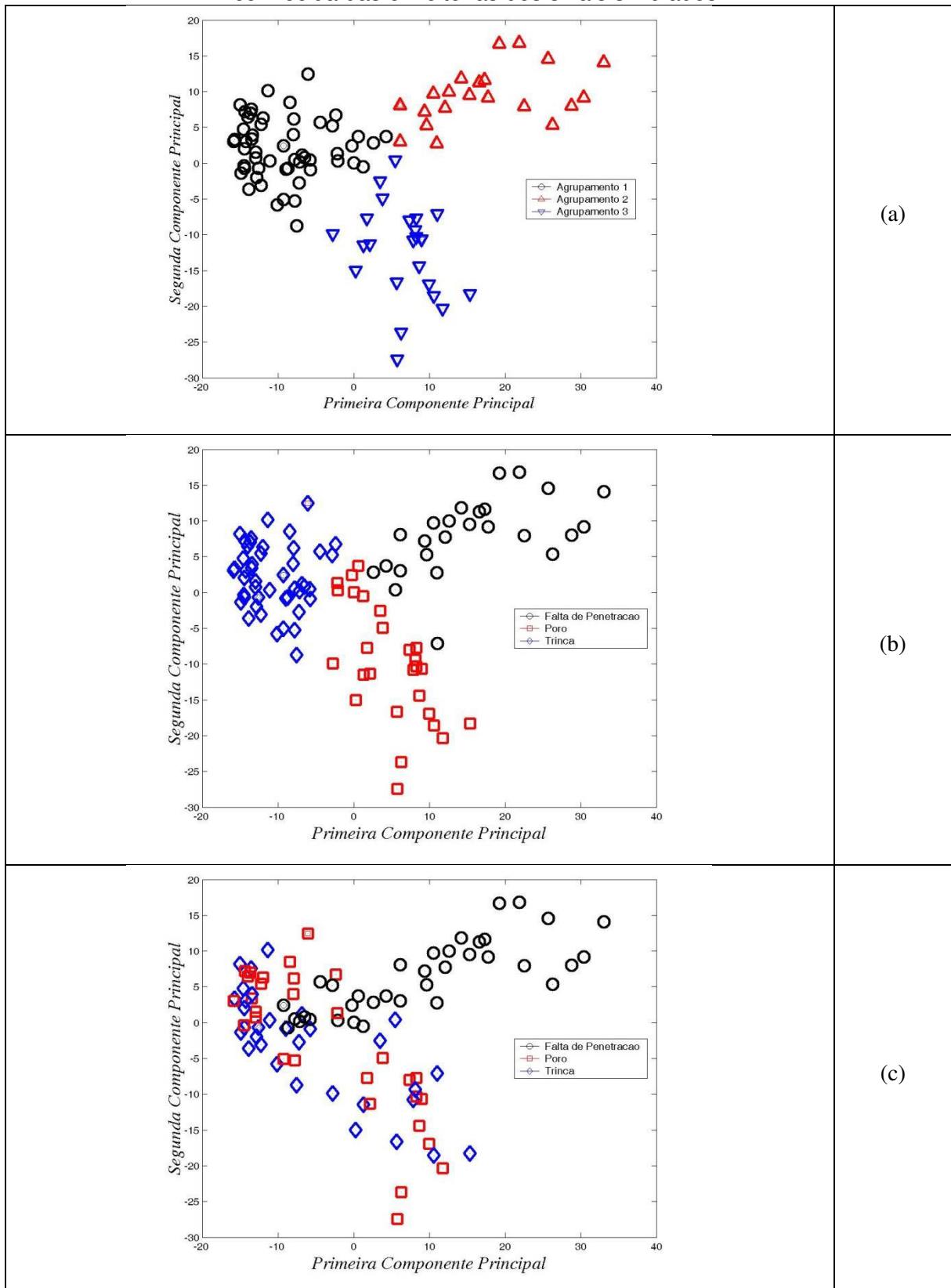


Fonte: Figura do próprio autor

Durante a segunda etapa, o algoritmo k-médias distribui os dados de entrada (duas primeiras componentes principais das envoltórias dos sinais ultrassônicos de defeitos modelados) em um número de agrupamentos sugerido na etapa anterior. Um exame sobre os índices dos dados, $i = \{1,2,\dots,36\}$, aparenta revelar a existência de um certo padrão na distribuição dos dados $\{1,1,1,2,2,2,1,1,1,2,2,2,1,1,1,2,2,2,1,1,2,2,2,1,1,2,2,2,2,1,1,2,2,2,2\}$ que está relacionado com a profundidade e largura da falta de penetração. Nenhum padrão, entretanto, foi detectado na distribuição dos dados referentes às classes poro e trinca. A tabela 11 exhibe o resultado dessa classificação.

A Figura 32 apresenta (a) o resultado da análise de agrupamento feito sobre as 2 primeiras componentes principais das envoltórias dos sinais simulados em comparação com o (b) resultado da própria análise de componentes principais e a (c) classificação conhecida dos respectivos sinais simulados.

Figura 32 – (a) resultado da análise de agrupamento das 2 primeiras componentes principais das envoltórias dos sinais simulados; (b) resultado da PCA; (c) classificação conhecida das envoltórias dos sinais simulados.



Fonte: Figura do próprio autor.

Tabela 11 – Agrupamento, ou pseudo-classificação, proposto após análise de agrupamento das 2 primeiras componentes principais das envoltórias dos sinais.

	Falta de Penetração	Poro	Trinca
Grupo 1	15	21	24
Grupo 2	21	0	0
Grupo 3	0	15	12

Fonte: Figura do próprio autor

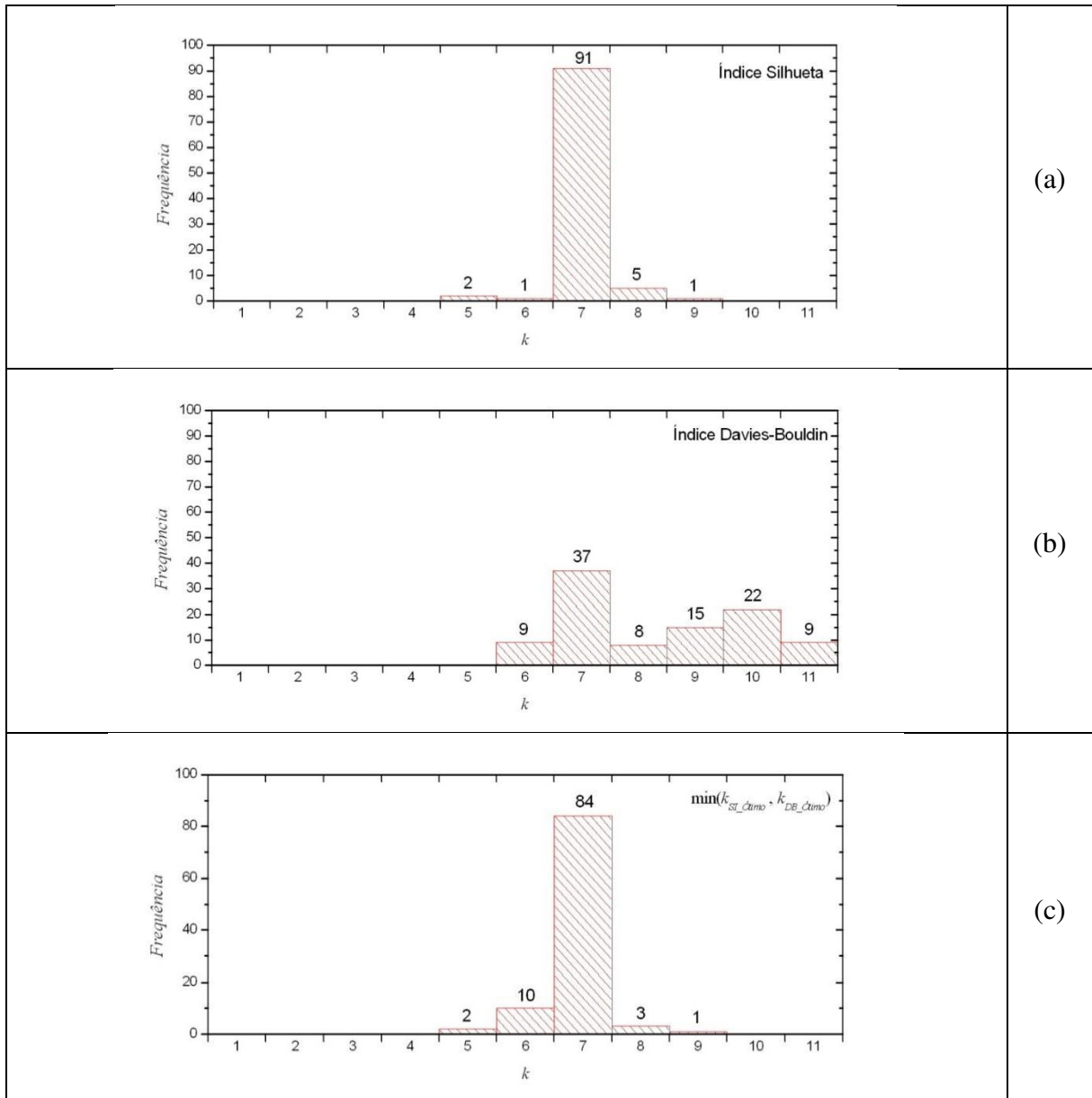
Posteriormente, quando o número de componentes principais analisadas é igual a 20 (vinte), a quantidade de *clusters* recomendada pelo algoritmo k-médias é diferente. Com ampla maioria das vezes (84%) o algoritmo recomenda a existência de sete *clusters*. Cinco (2%), seis (10%), oito (3%) e nove (1%) *clusters* também foram indicados (ver Figura 33). Novamente, a comparação do resultado da análise de agrupamento com a classificação conhecida dos sinais simulados possibilitou a identificação de um padrão na distribuição dos dados correspondentes a classe falta de penetração {1;1;2;2;3;3;1;1;2;2;3;3;1;1;2;2;3;3;1;1;2;2;3;3;1;1;2;2;3;3;1;1;2;2;3;3}, semelhante ao já observado durante a análise de agrupamento das envoltórias dos sinais ultrassônicos simulados (seção 4.2.2), e que também aparenta depender da profundidade da descontinuidade. Nenhum padrão foi identificado na distribuição das demais classes. A tabela 12 exibe um resumo dessa classificação, e a figura 33 permite comparação mútua dos resultados encontrados pelas análises K-médias e PCA com a classificação gerada durante a simulação.

Tabela 12 – Agrupamento, ou pseudo-classificação, proposto pelo algoritmo k-médias após processamento das 20 primeiras componentes principais das envoltórias dos sinais simulados.

	Falta de Penetração	Poro	Trinca
Grupo 1	12	0	1
Grupo 2	12	0	0
Grupo 3	12	0	0
Grupo 4	0	19	23
Grupo 5	0	10	8
Grupo 6	0	3	0
Grupo 7	0	4	4

Fonte: Tabela do próprio autor.

Figura 33 – Histograma do número de agrupamentos sugeridos para as 20 (vinte) primeiras componentes principais das envoltórias dos sinais ultrassônicos simulados: (a) baseado no índice Silhueta; (b) baseado no índice Davies-Bouldin; (c) baseado no mínimo entre os índices.

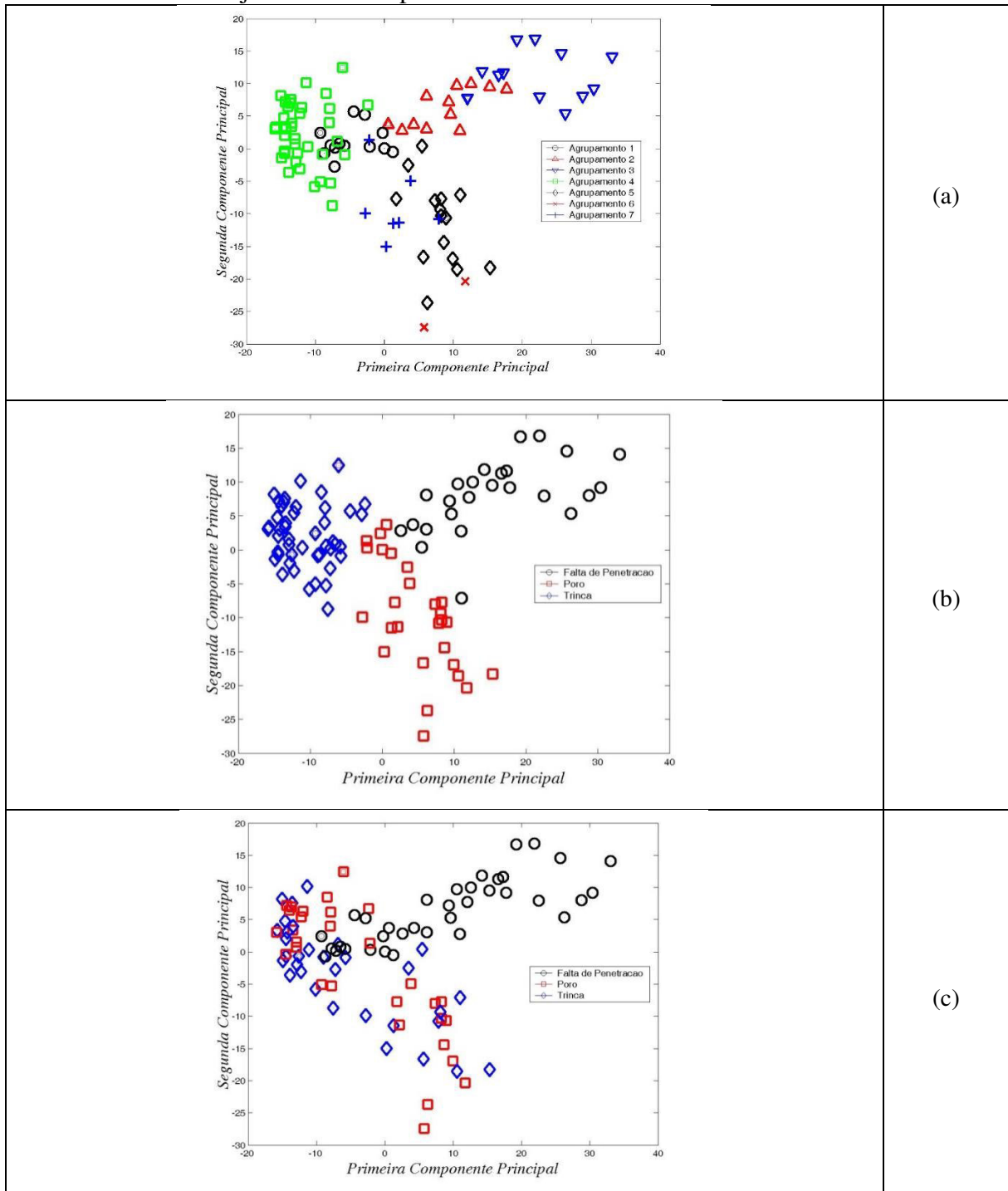


Fonte: Figura do próprio autor

Uma possível explicação para a diferença entre os resultados obtidos com o uso de duas e vinte componentes principais pode se apoiar no fato de que a medida de similaridade adotada pelo algoritmo k-médias utiliza a distância Euclidiana. É sabido que a vigésima componente representa uma quantidade muito menor de informações que a primeira. No entanto, para um dado de entrada n-dimensional em que as dimensões significam as componentes principais das envoltórias dos sinais ($n = 2$ ou 20), o cálculo da distância Euclidiana desconsidera que cada uma das dimensões (componentes principais) contém uma

quantidade diferente de informação sobre a envoltória do sinal. A Figura 34 apresenta (a) o resultado da análise de agrupamento das 20 primeiras componentes principais das envoltórias dos sinais simulados (b) resultado da própria análise de componentes principais e a (c) classificação conhecida dos respectivos sinais simulados

Figura 34 – (a) resultado da análise de agrupamento das 20 primeiras componentes principais das envoltórias dos sinais simulados; (b) resultado da PCA; (c) classificação conhecida do conjunto formado pelas envoltórias dos sinais simulados.



Fonte: Figura do próprio autor.

4.3 Análise de Componentes Principais e Análise de Agrupamentos dos sinais ultrassônicos capturados experimentalmente

Todos os resultados apresentados até o momento foram obtidos a partir da análise de sinais ultrassônicos simulados. Com o intuito de validá-los, investigou-se como o algoritmo k-médias agruparia sinais ultrassônicos reais. Para isso, chapas de aço foram inicialmente unidas por cordões de solda contendo descontinuidades propositalmente inseridas e posteriormente submetidas à inspeção ultrassônica automatizada. Os mesmos três tipos de defeitos de soldagem foram considerados: Poro, Trinca e Falta de Penetração. Após a inspeção ultrassônica, e com auxílio do resultado radiográfico realizado nos mesmos cordões de solda, 60 A-scans foram selecionados para cada um dos três tipos de descontinuidades citados, totalizando 180 A-scans. Em seguida, o procedimento aplicado anteriormente aos dados simulados foi repetido sobre os dados reais e os resultados de todas as análises estão exibidos nesta seção.

Os sinais capturados experimentalmente foram primeiramente submetidos a uma análise de componentes principais e matrizes de confusão produzidas a partir do emprego da regra da média da classe mais próxima aplicada as duas (e vinte) primeiras componentes principais desses sinais são exibidas nas tabelas 12 e 13, respectivamente. Uma taxa média de acerto na classificação de 45% foi alcançada com o uso das duas primeiras componentes principais, que mantêm 17.73% das informações. As vinte primeiras componentes principais mantêm 74.77% das informações e o uso dessas 20 componentes contribuiu para elevar a taxa média de acerto na classificação para 66.11%.

Tabela 13 – Matriz de confusão obtida pelo emprego da regra da média da classe mais próxima aplicada às 2 primeiras componentes principais dos sinais reais.

	Falta de Penetração	Poro	Trinca
Classificado como Falta de Penetração	85.00	55.00	28.33
Classificado como Poro	5.00	1.67	23.33
Classificado como Trinca	10.00	43.33	48.33

Fonte: Tabela do próprio autor.

Tabela 14 – Matriz de confusão obtida pelo emprego da regra da média da classe mais próxima aplicada às 20 primeiras componentes principais dos sinais reais.

	Falta de Penetração	Poro	Trinca
Classificado como Falta de Penetração	85.00	33.33	3.33
Classificado como Poro	5.00	46.67	30.00
Classificado como Trinca	10.00	20.00	66.67

Fonte: Tabela do próprio autor.

Prosseguindo, os sinais reais foram submetidos à análise de agrupamento para determinar o número ótimo de grupos, k , em que os sinais capturados podem ser divididos e a que grupo pertence cada sinal. Durante a primeira etapa o algoritmo testou os valores inteiros entre 1 e 14 (dado pela raiz quadrada do tamanho do conjunto de entrada composto por 180 A-scans). O número recomendado como ótimo pelos 2 índices empregados para dividir o conjunto de entrada variou significativamente e o valor ótimo mais citado foi sete. (Figura 35). Em seguida, dar-se início a etapa para determinar a que grupo pertence cada sinal.

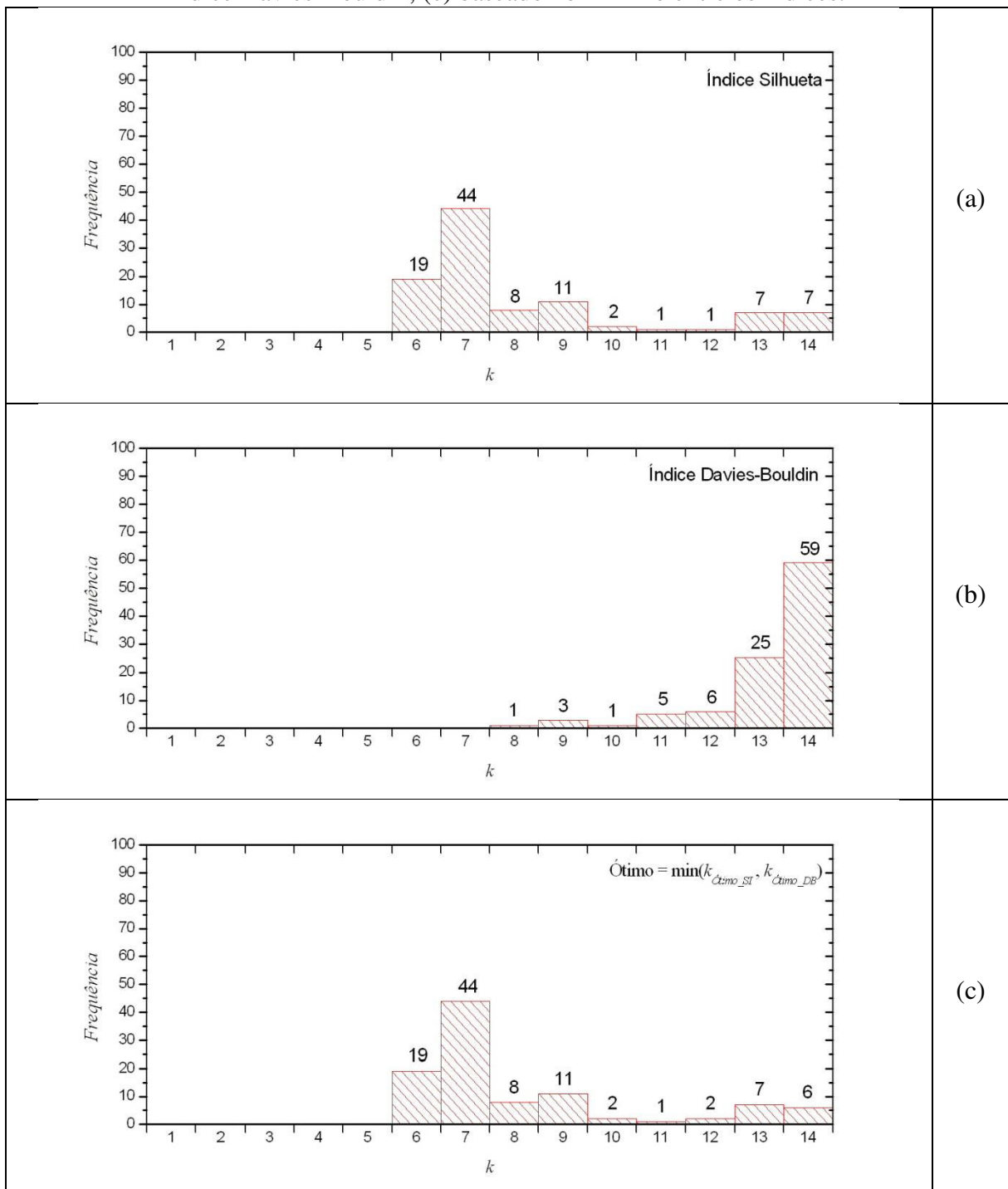
Tendo sido determinado a quantidade ótima para dividir o conjunto de entrada, inicia-se a etapa de determinar a que grupo pertence cada sinal. A tabela 15 exhibe os resultados obtidos nesta etapa da análise.

Tabela 15 – Agrupamento, ou pseudo-classificação, sugerida pela k -médias para os sinais capturados experimentalmente.

	Falta de Penetração	Poro	Trinca
Grupo 1	12	24	1
Grupo 2	47	23	7
Grupo 3	1	0	22
Grupo 4	0	4	5
Grupo 5	0	8	2
Grupo 6	0	1	9
Grupo 7	0	0	14

Fonte: Tabela do próprio autor.

Figura 35 – Histograma do número ótimo de agrupamentos sugerido para o conjunto de sinais ultrassônicos capturados experimentalmente: (a) baseado no índice Silhueta; (b) baseado no índice Davies-Bouldin; (c) baseado no mínimo entre os índices.



Fonte: Figura do próprio autor.

Ao examinar como os A-scans foram divididos percebe-se que os sinais referentes à falta de penetração estão distribuídos majoritariamente em dois grupos (*cluster 1* e *2*), e com apenas um sinal em um terceiro grupo (*cluster 3*). Os sinais referentes a trinca foram

distribuídos em todos os grupos, enquanto os referentes a poro foram distribuídos em cinco dos sete grupos sugeridos. A classificação dos sinais simulados é definida na modelagem da descontinuidade. Diferente dos sinais simulados, os sinais analisados nessa seção foram capturados durante a inspeção de juntas soldadas que apresentam descontinuidades internas reais. Estas não possuem tamanhos pré-definidos e sua classificação foi realizada a partir da análise da imagem radiográfica da respectiva descontinuidade, o que impede o dimensionamento de sua profundidade. Assim, não é possível a identificação de padrões que justifiquem o modo como foi realizada a partição desses sinais. A Figura 36 permite a comparação entre (a) o resultado da análise de agrupamento; (b) o resultado da análise de componentes principais; e (c) a classificação conhecida dos 180 sinais ultrassônicos capturados experimentalmente.

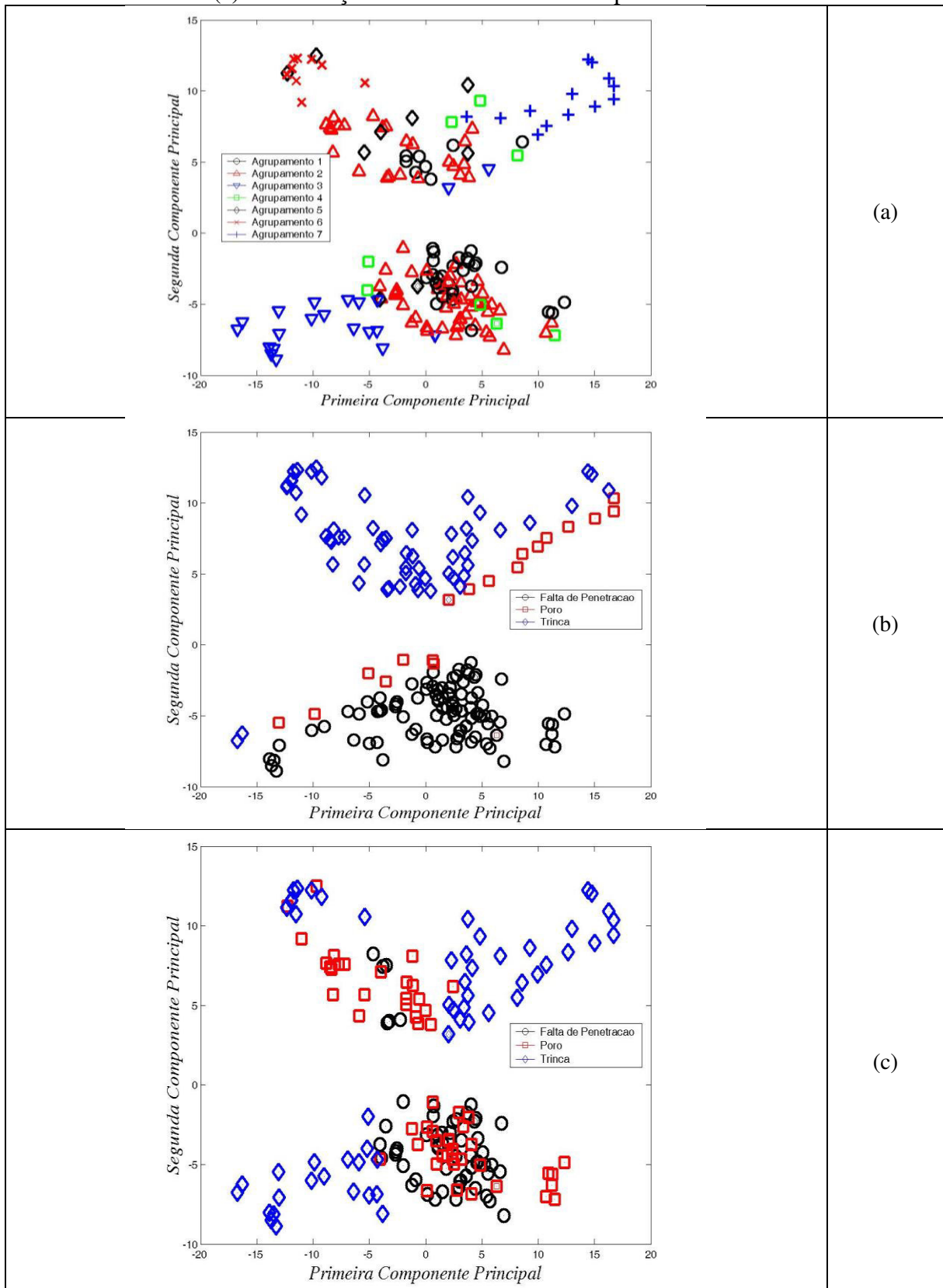
Por fim, foram analisadas as envoltórias dos sinais capturados experimentalmente. A tabela 16 exibe a matriz de confusão calculada pela regra da média da classe mais próxima aplicada às duas primeiras componentes principais dessas envoltórias. As duas primeiras componentes principais armazenam 45.05% das informações e foram responsáveis por 73.89% de acerto na classificação. As vinte primeiras componentes principais registram 94.96% das informações, e a tabela 17 calculada pela regra da média da classe mais próxima aplicada às vinte componentes das envoltórias mostra um acerto na classificação de 76.11%.

Tabela 166 – Matriz de confusão obtida pelo emprego da regra da média da classe mais próxima aplicada às 2 primeiras componentes principais das envoltórias dos sinais reais.

	Falta de Penetração	Poro	Trinca
Classificado como Falta de Penetração	68.33	5.00	0.00
Classificado como Poro	21.67	61.67	8.33
Classificado como Trinca	10.00	33.33	91.67

Fonte: Tabela do próprio autor.

Figura 36 - (a) resultado da análise de agrupamento; (b) resultado da PCA; (c) classificação conhecida dos sinais experimentais.



Fonte: Figura do próprio autor.

Tabela 177 – Matriz de confusão obtida pelo emprego da regra da média da classe mais próxima aplicada às 20 primeiras componentes principais das envoltórias dos sinais reais.

	Falta de Penetração	Poro	Trinca
Classificado como Falta de Penetração	68.33	3.33	0.00
Classificado como Poro	21.67	66.67	6.67
Classificado como Trinca	10.00	30.00	93.33

Fonte: Tabela do próprio autor.

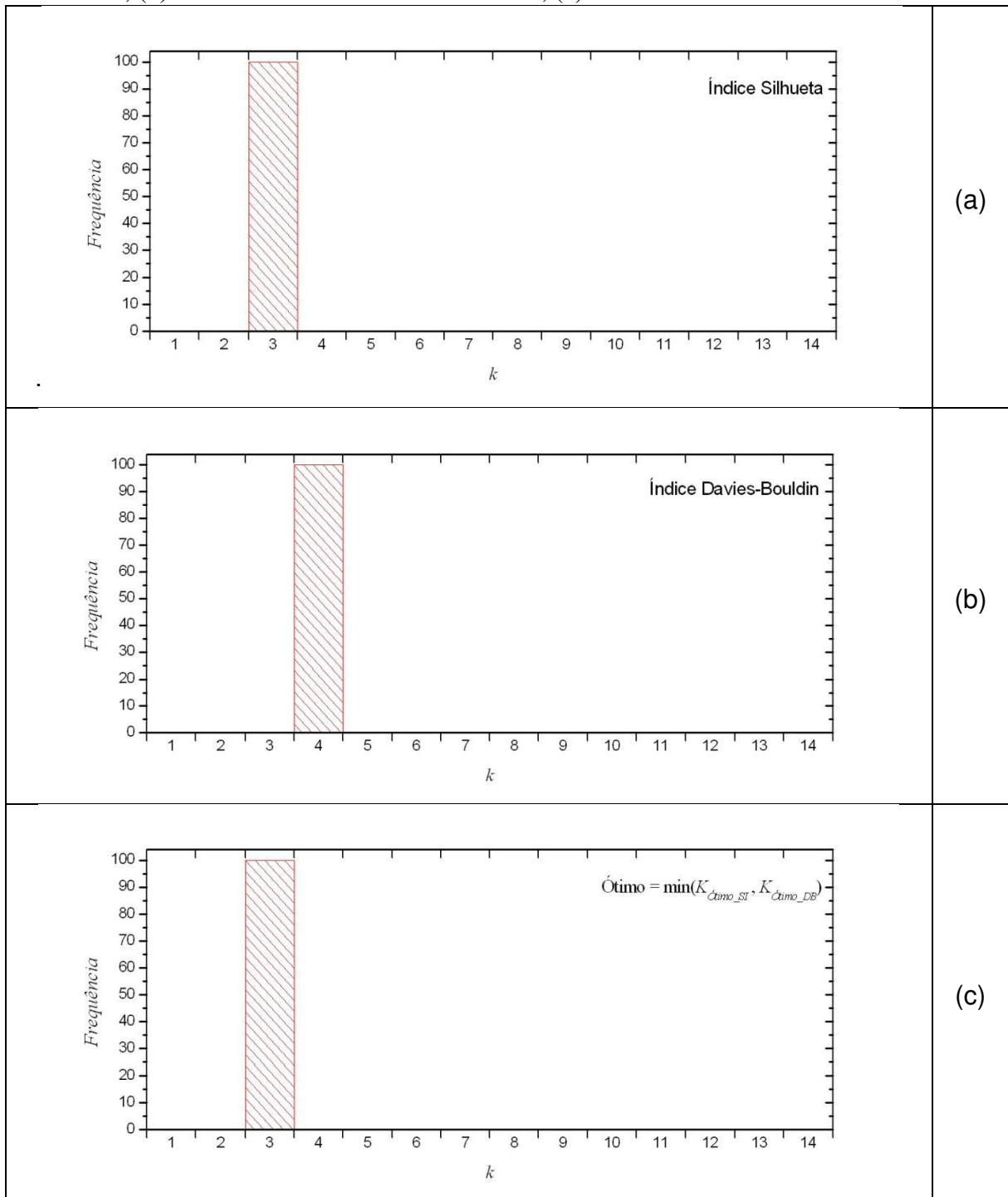
Prosseguindo, os sinais reais foram submetidos à análise de agrupamento para determinar o número ótimo de grupos, k , em que os sinais capturados podem ser divididos e a que grupo pertence cada envoltória de sinal real.

Em uma centena de execuções que o algoritmo k -médias testou valores de k entre 1 e 14, o índice silhueta recomendou $k = 3$ todas as vezes, o índice Davies-Bouldin recomendou sempre $k = 4$, e o número obtido pela função $k_{\text{ótimo}} = \text{mínimo}(k_{\text{ótimo_SI}}, k_{\text{ótimo_DB}})$ sugere a existência de 3 *clusters* é o valor ótimo (Figura 37).

A tabela 18 mostra o resultado obtido quando o conjunto de entrada formado pelas envoltórias dos sinais capturados durante a inspeção dos corpos de prova contendo defeitos nos cordões de solda é reagrupado em três grupos. Vemos que a maioria (57%) dos sinais que compõe o agrupamento 1 são relativos a poro. O restante (43%) corresponde à falta de penetração. Nenhum sinal de trinca foi associado ao grupo 1. Todos os 60 sinais relativos à trinca foram associados ao grupo 2, de modo que 62.5% desse grupo é composto por essa categoria de descontinuidade. Poros e falta de penetração respondem respectivamente por 29.2% e 8.3% do grupo 2. O 3º agrupamento é formado exclusivamente por alguns dos sinais de falta de penetração.

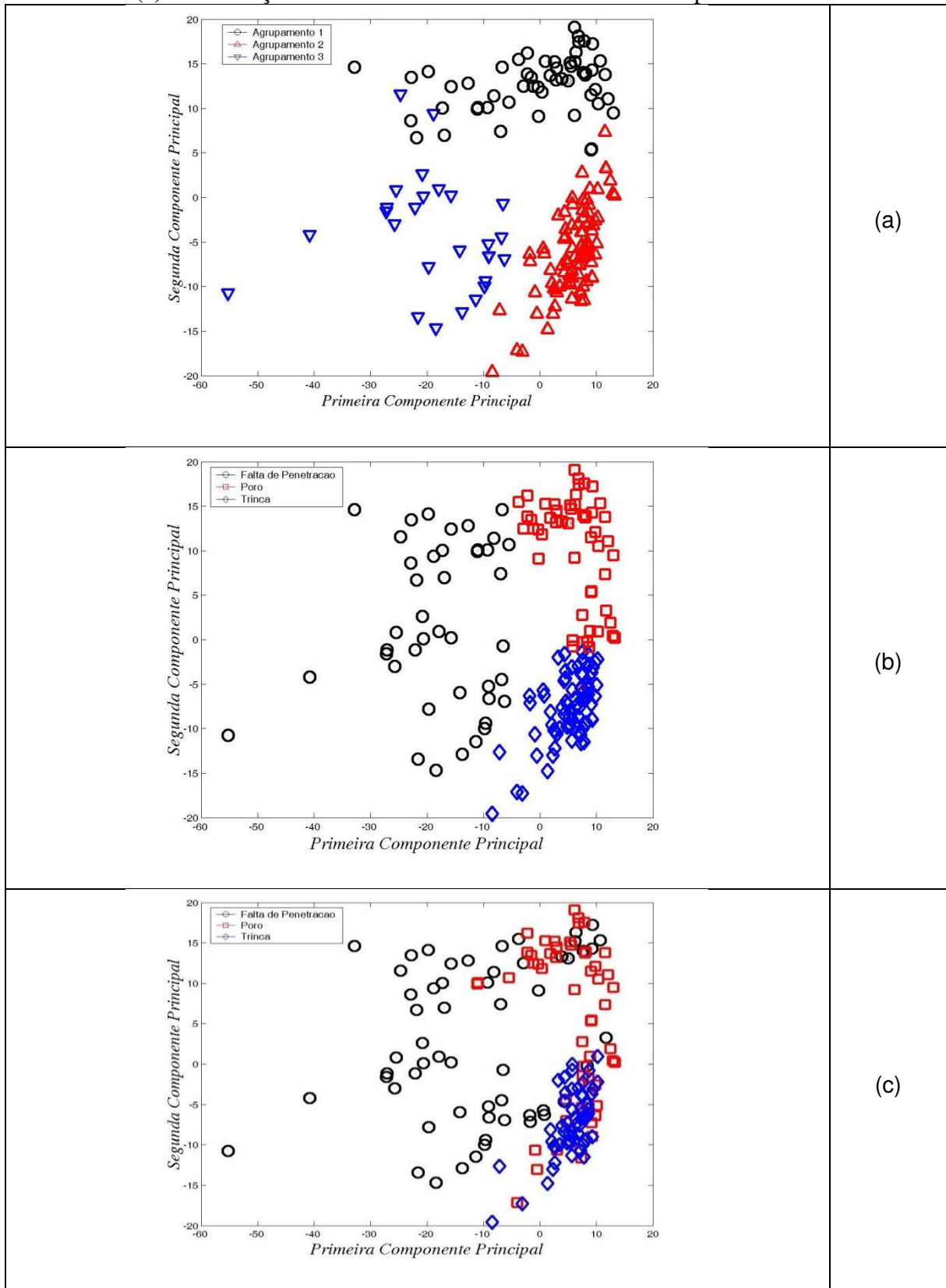
A Figura 38 permite a comparação entre o (a) resultado da análise de agrupamento; o (b) resultado da análise de componentes principais; e a (c) classificação conhecida dos 180 sinais ultrassônicos capturados experimentalmente.

Figura 37 – Histograma do número ótimo de agrupamentos sugerido para o conjunto das envoltórias dos sinais ultrassônicos capturados experimentalmente: (a) baseado no índice Silhueta; (b) baseado no índice Davies-Bouldin; (c) baseado no mínimo entre os índices.



Fonte: Figura do próprio autor.

Figura 38 – (a) resultado da análise de agrupamento; (b) resultado da PCA; (c) classificação conhecida das envoltórias dos sinais experimentais.



Fonte: Figura do próprio autor.

Tabela 188 – Agrupamento, ou pseudo-classificação, dos sinais capturados experimentalmente sugerida pela k-médias.

	Falta de Penetração	Poros	Trinca
Grupo 1	24	32	0
Grupo 2	8	28	60
Grupo 3	28	0	0

Fonte: Tabela do próprio autor.

5 CONCLUSÃO

- 1) A análise de componentes principais dos sinais ultrassônicos simulados, e de suas envoltórias, revelou que a classe falta de penetração é mais facilmente separável. As duas outras classes (poro e trinca) parecem não ser linearmente separáveis.
- 2) A análise de componentes principais das envoltórias dos sinais ultrassônicos simulados ocasionou uma piora significativa da classificação dos sinais pertencentes a classe falta de penetração e uma grande melhora na classificação dos sinais relacionados a trinca.
- 3) A análise de agrupamento realizada sobre os sinais simulados revelou que a melhor maneira de agrupar os dados é em duas partes. Um dos agrupamentos é constituído basicamente pelos sinais de falta de penetração, e o segundo agrupamento reúne os sinais do tipo poro, trinca e alguns do tipo falta de penetração. O fato da falta de penetração sempre ocorrer no centro do cordão e terem sido modeladas com dimensões significativamente maiores, provavelmente explica o motivo do algoritmo k-médias reunir a maioria dessas descontinuidades em um grupo diferente do tipo poro e trinca, que por sua vez tem dimensões parecidas e a aleatoriedade de suas posições.
- 4) A análise de agrupamento realizada sobre a envoltória dos sinais simulados revelou que a melhor maneira de agrupar os dados é em seis partes, de modo que as envoltórias dos sinais de falta de penetração foram subdivididos em três subconjuntos de acordo com a profundidade da descontinuidade (pouco profunda; média e profunda) e independente de sua largura, já as outras duas classes de envoltórias, poro e trinca, foram subdivididas em 3 agrupamentos, que aparentemente não fizeram algum sentido lógico.
- 5) A análise de agrupamento das componentes principais dos sinais ultrassônicos simulados revelou que a melhor maneira de agrupar os dados é em duas partes, todos os sinais do tipo falta de penetração ficaram agrupados no primeiro grupo, já o segundo grupo reuniu os outros dois tipos de descontinuidades (poro e trinca). A mesma hipótese levantada para justificar o agrupamento sugerido para os sinais ultrassônicos simulados pode ser empregada para explicar o agrupamento das duas (ou vinte) primeiras componentes principais desses sinais: i) a similaridade entre as pequenas dimensões de poros e trincas versus as dimensões maiores das falta de penetração; ii) a aleatoriedade das posições e poros e trincas versus o determinismo da posição da falta de penetração em relação ao centro do cordão.
- 6) É importante notar que tanto a análise de agrupamento realizada sobre os sinais simulados quanto a análise de agrupamento das componentes principais dos sinais ultrassônicos simulados foi revelado que a melhor maneira de agrupar os dados é em duas classes.

7) A análise de agrupamento das duas componentes principais das envoltórias dos sinais ultrassônicos simulados revelou que a melhor maneira de agrupar os dados é em três partes, a divisão dos dados foi semelhante ao já observado durante a análise de agrupamento das envoltórias dos sinais ultrassônicos simulados (seção 4.21), e que também aparenta depender da profundidade da descontinuidade; já para vinte componentes principais o agrupamento ótimo foi em sete classes, no qual todos as descontinuidades do tipo falta de penetração foram agrupados na classe 1, 2 e 3 e as descontinuidade do tipo trinca e poro não foram bem separadas e foram classificadas na 4,5,6 e 7 classe, sem uma divisão que parecesse ter sentido lógico.

8) A análise do agrupamento das duas e de vinte componentes principais, e das envoltórias, dos sinais ultrassônicos simulados, revelou que os sinais são melhores separados quando se utiliza duas componentes, neste caso cem por cento das vezes disse que o k ótimo é dois, uma possível explicação para esse resultado pode se apoiar no fato de que o agrupamento realizado pelo algoritmo k -médias leva em consideração a distância Euclidiana entre os elementos de entrada e os centroides. Desta forma, desconsidera-se que cada uma das componentes leva consigo uma quantidade diferente de informações. Por exemplo, a vigésima componente representa uma quantidade muito menor de informações que a primeira, o que não é considerado no cálculo da distância.

9) E quanto aos sinais reais, verificou-se na análise de duas e vinte componentes principais que a descontinuidade do tipo falta de penetração foi 85% das vezes reconhecida, e as outras duas descontinuidade não foram bem reconhecidas; já para a análise do agrupamento dos sinais reais, o número ótimo sugerido para divisão dos dados foi em sete classes, na qual todos os defeitos do tipo falta de penetração foram agrupados na primeira, segunda e terceira classe, e os defeitos do tipo trinca e poro não foram agrupados de uma forma que parecesse fazer sentido.

10) No presente trabalho, mesmo sendo simulado três tipos de defeito de soldagem, o algoritmo k -médias não retornou três como valor ótimo para o agrupamento. A melhor separação ocorreu no agrupamento das duas componentes principais dos sinais ultrassônicos simulados, onde em cem por cento das vezes, o programa disse que o número ótimo de classes é dois, separando bem o defeito falta de penetração em um grupo e confundindo o defeito do tipo trinca e poro em outro.

6 SUGESTÕES PARA TRABALHOS FUTUROS

- 1) Realizar um estudo de refino de malha antes da simulação;
- 2) Considerar a modelagem de trincas com altura maiores que 5.0 mm;
- 3) Utilizar um conjunto maior de dados de entrada;
- 4) Reproduzir um transdutor *phased array* composto por um número maior de cristais para melhorar a qualidade do feixe ultrassônico simulado;
- 5) Emprego de outras ferramentas de reconhecimento de padrão, tais como: redes neurais artificiais; transformada Karhunen–Loève; discriminador Gaussiano;
- 6) Comparar com sinais produzidos por algoritmos que utilize um método computacional diferente dos elementos finitos empregados pelo COMSOL;
- 7) Controle na posição das descontinuidades do tipo poro e trinca.

7 REFERÊNCIAS

ANDREUCCI R, **Ensaio por Ultrassom**, Apostila da Associação Brasileira de Ensaios não Destrutivos e Inspeção- ABENDI, Maio de 2014.

ANIL K.J; **Data Clustering: 50 Years Beyond k-means**, Article, Department of Computer Science & Engineering; Michingan State University, 2009.

BEVAN B., COPSON B.; **The Mathematical Theory of Huygens' Principle**, Book, 3 ed, New York, 1987.

BOLDRINI, C. ;**Álgebra Linear**; Livro; 3 ed. Editora Harbra, 1986.

BOLSHAKOVA N; **Cluster validation techniques for genome expression data**; Article, Department of Computer Science, Trinity College, Dublin, June 2002.

BUFFONI S. O.; **Apostila de Introdução Aos Métodos Numéricos - Parte I**. Apostila, Universidade Federal Fluminense, 2002.

CAMURÇA P; **Simulação do ensaio ultrassônico pela técnica TOFD utilizando o método dos volumes finitos**; Dissertação de mestrado, Universidade Federal do Ceará, 2014.

CATUNDA C; **Implementação de modelagem elástica para detecção de defeitos por ultrassom em juntas soldadas dissimilares**; Dissertação de mestrado, Universidade Federal do Rio de Janeiro, 2004.

CHARLES, J. H. et al., Radiographic Interpretation. In: Moore, P.O. (editor) **Nondestructive Testing Handbook**, Radiographic Testing, 3 ed, vol. 4, pp. 207–218. ASTM, Columbus 2002.

COMSOL; **Acoustic Module**. Disponível em: <<https://br.comsol.com/support/releasenotes/4.3/acoustics/>> Acesso em: 01 set. 2015.

DEZA E; DEZA M; **Encyclopedia of Distances**, Springer, p. 94, 2009.

FILHO F; **Modelagem sísmica de ondas elásticas e migração reversa no tempo em meios transversalmente isotrópicos**. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, 2002.

FERZIGER J. H.; PERIC M.; **Computational Methods for Fluid Dynamics**. Book, SpringerVerlg Berlin Heidelberg New York, 1997.

FRANKENSTEIN Y; Aprendizado de Máquina (II);**K-means Clustering**; article, ago 2009.

JUNIOR, A.; CAMARINHA, D.; **Algoritmo K-MEANS**; Instituto de Matemática e Estatística. Disponível em: <<http://www.ime.usp.br/~arcjr/machine/>> Acesso em: 20 ago. 2015. (Junior e Camarinha, 2015)

KAUFMAN, L.; ROUSSEU, P.J., **Finding Groups in Data: An Introduction to Cluster Analysis**. Book, John Wiley & Sons, Inc., New York, 1990.

LANDAU, L D; PITAEVSKII, L. P.; KOSEVICH, A. M.; LIFSHITZ, E.M; **Theory of Elasticity, Third Edition: Volume 7 (Course of Theoretical Physics)** ;International Edition; 1986.

PETROVIC S. **A Comparison Between the Silhouette Index and the Davies-Bouldin Index in Labelling IDS Clusters**. Article, Department of Computer Science and Media Technology, Gjøvik University College, Norway, 2002.

PIECH C.; **K Means- The Basic Idea**. 2013. Disponível em: <<http://stanford.edu/~cpiech/cs221/handouts/kmeans.html>> Acesso em: 20 ago. 2015.

MARQUES, P.; MODENESI, P.; BRACARENSE, A.; **Fundamentos e Tecnologia. Soldagem**. livro, v. 1, 3 ed atualizada, 2009.

M.G. Silk, B.H. Lidington. **The potential of scattered or diffracted ultrasound in the determination of crack depth**. Article, Non-Destructive Testing, Volume 8, Issue 3, June 1975.

MOURA E. P.; SIQUEIRA M H S; SILVA R R; RABELLO J M A; **Welding defect pattern recognition in TOFD signals Part 2. Non-linear classifiers**, Article, Insight (47), 2005.

NDT-KITS BLOG; **What's ultrasonic thickness measurement ?**. 2013. Disponível em: <<http://www.ndt-kits.com/blog/?paged=2>> Acesso em: 10 set. 2015.

NDT RESOURCE CENTER; **Radiograph interpretation – Welds**. Disponível em: <<https://www.nde-ed.org/EducationResources/CommunityCollege/Radiography/TechCalibrations/RadiographInterp.htm>> Acesso ago.15 out. 2015.

NDT RESOURCE CENTER; **Ultrasound– Weldments (Welded Joints)**. Disponível em:<<https://www.ndeed.org/EducationResources/CommunityCollege/Ultrasonics/SelectedApps/weldments.htm>> Acesso ago.15 Dez. 2015.

OLYMPUS. **Ultrasonic time-of-flight-diffraction (TOFD) examination of butt-fusion joints of high-density polyethylene (HDPE)**. Disponível em:<<http://www.olympus-ims.com/en/applications/ultrasonic-tofd-butt-fusion/>> Acesso em: 10 nov. 2015.

SILVA, I. C., **Avaliação da Técnica do Tempo de Percurso da Onda Difrata no Dimensionamento de Descontinuidades**, Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, março 1999.

STEVEN C. C ; RAYMOND P.; **Numerical Methods for Engineers**, Book, 6 ed, McGraw-Hill Education , 2009.

TEMPLE, J.A.G., **Predicted ultrasonic resposner for pulse-echo inspections**, Article, British Journal of NDT, pp. 145 - 154, May, 1986.

VASCONCELOS S.; **Análise de componentes principais**, Dissertação de Mestrado, Universidade Federal do Fluminense, 2007.

VARELLA C. A. A.; **Análise de componentes principais**, Dissertação de Mestrado, Universidade Federal Rural do Rio de Janeiro, novembro 2008.

WANGENHEI R.; **Análise de Agrupamentos**. Última atualização out 2006; Disponível em: <<http://www.inf.ufsc.br/~patrec/agrupamentos.html>> Acesso em: 28 ago. 2015.

WOOH S.; SHI Y.; **A Simulation Study of the Beam Steering Characteristics for Linear Phased Arrays**. Article, Journal of Nondestructive Evaluation, June 1999, Volume 18, Issue 2, p 39-57.

XCITEFUN. **Animals & humans face similarity**. Disponível em: <<http://forum.xcitefun.net/animals-and-humans-face-similarity-t24618.html>> Acesso em: 12 set. 2015.